小·时百科

专业易用的数理百科

http://wuli.wiki



致乐于分享知识的人

版权声明

小时百科项目的官方网站为 wuli.wiki，网站上免费提供本书的网页版和 pdf 版的下载，仅供**个人学习使用**．为了维持项目不断发展，强烈建议每位读者试读超过 1 小时后在网站上捐款，我们在此表示衷心感谢．若未经同意，请勿以任何方式转载本项目内容（包括插图，代码，网页等）．版权所有，保留一切权利．

本书不定期更新，为保证内容质量请及时下载最新版．当前版本编译于：2021 年 4 月 18 日．

小时百科

简介

小时百科（以下简称百科）从结构上尝试将教材和百科这两种不同形式的文本融合到一起，使其既适合初学者自学，又可按照非常灵活的顺序阅读．百科计划涵盖物理专业本科课程中的大部分内容，适用于具有普通高中及以上数学物理水平的读者．也正因为如此，百科是一个庞大的工程，短时间内很难被完成，所以百科将长期处于更新状态．为保证阅读质量，请定期从项目网站下载最新版．

在介绍百科的特点以前，我们先来看一般数理教材的不足：

1. 需要按顺序学习，不适合初学者快速了解或查找某个话题或知识点．例如某高中生需要了解角动量的概念，直接翻开力学书的相关章节发现看不懂，却又不知道需要先学什么，也没时间从头先看完高等数学和线性代数的教材再开始学习．

2. 读者不能自己掌握所学内容的深度和严谨性．例如许多高等数学教材在读者对微积分还没有一个大概的了解时就介绍极限的 $\varepsilon-\delta$ 定义，微分/积分中值定理，可微，可积等．这些内容对物理的初步学习来说显得过于严谨，会极大加重学习成本．

3. 不够自洽（self-contained）．一本教材的自洽性指目标读者在学习前是否还需要学习其他教材．大部分本科物理教材对高中生都是不自洽的，因为它们往往假设读者具有一定的微积分和线性代数基础．

再来看一般网络百科（如百度百科或维基百科）的不足：

1. 每个词条都大而全，涵盖词条标题的所有相关内容．

2. 读者同样不能自己掌握所学内容的深度和严谨程度．

3. 容易出现公式定理的堆积，缺乏知识点导入和讲述，缺乏例题，习题等．

为了克服上述困难，百科采用以下形式：

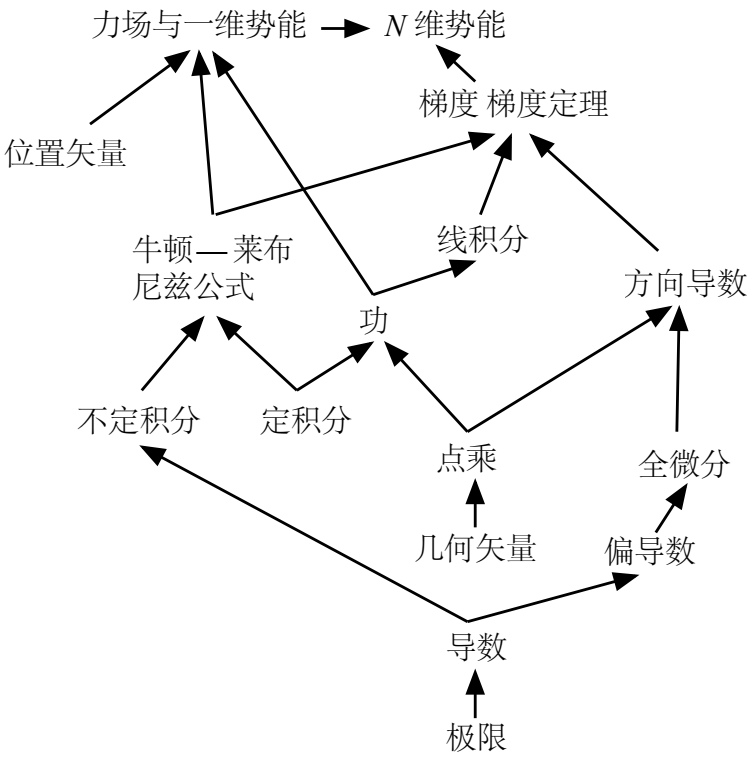


图 1: 由“预备知识”画出的知识树（目标词条为“力场 势能[905]”）

1. 将知识点划分为词条，且在每个词条中列出学习该词条前需要先学习哪些词条．这样相当于建立了一个知识树（如图 1）．项目网站 wuli.wiki 上可以自动生成任意目标词条的知识树．

2. 采用词条分级，把同一个话题以不同深度，严谨度和适用范围等划分成若干个等级的同名词条．这样读者可以选择螺旋式学习（例如初中，高中，大学物理中所学的话题几乎相同，但程度不同）．暂定初级词条从科普开始，尽量少使用数学．随着词条级别升高，会使用适用范围更广的定义，更严谨的表述和更抽象的数学等．

词条

百科内容繁多，不同词条的重要性相去甚远，不建议初学者按照词条的排列顺序依次学习，而是应该以初级词条给出的主线来学习，再根据兴趣和需要阅读其余词条．

理论上来说，读者可以直接跳到最感兴的词条，如果“预备知识”中列出

的词条都已经掌握，就可以开始学习该词条，否则就先掌握“预备知识”中的词条．如果“预备知识”出现在词条开始，则必须先掌握，如果出现在正文中，则只有阅读该部分时需要掌握．如果正文中引用了没有出现在“预备知识”中的内容，则读者可自行决定是否阅读．

为了便于书内的跳查，词条之间进行了大量的交叉引用，例如“导数简介[158]”右上角中括号中的数字代表被引用词条的页码．由于每个词条的公式编号都从 1 开始，引用其他词条中的公式有时会用类似“式 2 [158]”的格式，右上角的方括号中是公式所在词条的起始页码．在本书的 PDF 电子版中，点击该页码即可自动跳转到对应的页面．在电脑上阅读，推荐使用 Adobe Reader 阅读器，在苹果® 的 iOS 设备上推荐使用 GoodReader 应用（两个软件都可以在不同的面板中打开同一本书的不同页码）．在 Adobe Reader 中，使用快捷键组合“Alt + 左箭头”即可返回跳查前的位置，在移动设备的阅读软件中通常也有相应的返回按钮．由于本书的电子版是原生 PDF（区别于扫描版），还具有占用设备存储空间小，便于分享，便于查找关键字等种种优势．

目录

第一部分　科普

第一章　经典力学

经典力学及其他物理理论 [3]　经典力学 [4]　动量和能量一维势能曲线 [10]

第二章　电动力学

电动力学 [15]　静电的基本规律和性质 [17]　荷质比的测定 [18]　右手定则 [20]

第三章　量子力学

原子的观念 [23]　从天球的音乐到玻尔模型 [36]　量子力学 [60]

第四章　其他

天文学常识 [74]　时间的计量 [80]　时间的计量 2 [83]　加密货币简介 [85]

第二部分　微积分

第一章　基础

公理系统 [91]　集合 [92]　整数 [95]　映射 [98]　抽象 [102]　函数 [103]　反函数 [103]　无穷的概念 [104]　阶乘 [105]　排列 [106]　组合 [106]　克罗内克 delta 函数 [107]　二项式定理 [107]　二项式定理（非整数幂） [108]　直线和平面的交点 [109]　点到直线的距离 [110]　海伦公式 [110]　直线和球的交点 [110]　等比数列 [112]　三角恒等式 [112]　双曲函数 [114]　sinc 函数 [116]　充分必要条件 [116]　四象限 Arctan 函数 [118]　极坐标系 [119]　阿基米德螺线 [121]　柱坐标系 [121]　球坐标系 [123]　抛物线坐标系 [124]　一元函数的对称与周期性 [127]　球坐标与直角坐标的转换 [128]　圆锥曲线的极坐标方程 [131]　椭圆的三种定义 [133]　双曲线的三种定义 [134]　抛物线的三种定义 [136]　圆锥曲线的光

学性质 [137] 复数 [142]

第二章 一元微积分

微积分导航 [148] 极限 [149] 小角正弦极限 [152] 自然对数底 [153] 切线与割线 [155] 函数的连续性 [156] 导数 [158] 求导法则 [160] 反函数求导 [163] 基本初等函数的导数 [165] 高阶导数 [169] 莱布尼兹公式 [169] 导数与函数极值 [170] 用极值点确定函数图像 [172] 一元函数的微分 [173] 复合函数求导 链式法则 [174] 曲率 曲率半径（平面） [179] 泰勒级数 [181] 泰勒级数 2 [184] 导数与差分 [185] 不定积分 [187] 换元积分法 [188] 分部积分法 [190] 积分表 [192] 定积分 [198] 牛顿—莱布尼兹公式 [202] 幂级数 [205] 幂级数 2 [206] 极坐标中的曲线方程 [208] 曲线的长度 [209] 函数的算符 [210] 常微分方程 [211] 一阶线性微分方程 [212] 二阶常系数齐次微分方程 [215] 一维齐次亥姆霍兹方程 [217] 二阶常系数非齐次微分方程 [218] 正交函数系 [221] 傅里叶级数（三角） [222] 傅里叶级数（指数） [228] 狄拉克 delta 函数 [230] 多元狄拉克 delta 函数 [236] 傅里叶变换（三角） [236] 傅里叶变换（指数） [237] 连续正交归一基底与傅里叶变换 [240] 傅里叶变换与矢量空间 [243] 拉普拉斯变换 [246] 拉普拉斯变换的性质 [248] 柯西—阿达玛公式 [253] Gamma 函数 [254] Euler-Mascheroni 常数 [256] Gamma 函数 2 [257] 双 Gamma 函数 [257] 渐近展开 [258] 拉普拉斯方法 [261]

第三章 多元微积分与矢量分析

偏导数 [266] 最小二乘法 [267] 全微分 [269] 复合函数的偏导 链式法则 [270] 全导数 [273] 矢量的导数 求导法则 [274] 矢量的偏导数 [276] 偏微分算符 [277] 一元矢量函数的积分 [277] 方向导数 [279] 二元函数的极值（充分条件） [281] 重积分 [283] 立体角 [286] 极坐标系中单位矢量的偏导 [287] 正交曲线坐标系 [288] 多元函数的傅里叶级数 [291] 正交曲线坐标系中的重积分 [291] 矢量场 [294] 线积分 [295] 曲面积分 通量 [297] 证明闭合曲面的法向量面积分为零 [299] 矢量算符 [300] 拉普拉斯算符 [301] 一种矢量算符的运算方法 [302] 矢量算符运算法则 [304] 分部积分的高维拓展 [306] 梯度 梯度定理 [307] 散度 散度定理 [311] 牛顿—莱布尼兹公式的高维拓展 [314] 旋度 [315] 斯托克斯定理 [318] 调和场 [319] 矢量分

析总结 [319] 散度的逆运算 [324] 旋度的逆运算 [325] 亥姆霍兹分解 [328] 亥姆霍兹分解 2 [330] 拉格朗日乘数法 [331] 一阶线性常微分方程组 [333] 齐次函数的欧拉定理 [336] 多元泰勒展开 [338] 雅可比行列式 [338] 高斯积分 [339] 多维球体的体积 [340] 多元函数积分和宇称 [343]

第四章 复变函数

复变函数 [346] 幂函数（复数） [350] 指数函数（复数） [351] 三角函数（复数） [352] 复变函数的导数 柯西—黎曼条件 [354] 解析函数与散度旋度 [356] 复变函数的积分 [357] 牛顿—莱布尼兹公式（复变函数） [364] 柯西积分定理 [366]

第三部分 线性代数

第一章 几何矢量

几何矢量 [370] 几何矢量的运算 [375] 线性相关性 [378] 几何矢量的基底和坐标 [379] 几何矢量的内积 [383] 正交归一基底 [386] 矢量叉乘 [387] 矢量叉乘分配律的几何证明 [391] 连续叉乘的化简 [392] 三矢量的混合积 [392]

第二章 矩阵和行列式

矩阵 [396] 平面旋转矩阵 [402] 三维旋转矩阵 [404] 逆矩阵 [405] 叉乘的矩阵形式 [406] 矩阵的秩 [407] 厄米共轭算符的映射结构 [408] 矩阵的迹 [410] 高斯消元法求逆矩阵 [412] 克拉默法则 [413] 过渡矩阵 [414] 绕轴旋转矩阵 [416] 旋转矩阵的导数 [418] 四元数与旋转矩阵 [419] 对称矩阵 [421] 厄米矩阵 [422] 矩阵的本征方程 [423] 对称矩阵的本征问题 [425] 厄米矩阵的本征问题 [426] 相似变换和相似矩阵 [429] 逆序数 [430] 列维—奇维塔符号 [431] 行列式 [432] 行列式的性质 [435] 行列式与体积 [438] 克莱姆法则 [438] 海森矩阵 [439] 正交矩阵酉矩阵 [440] 幺正变换 [442]

第三章 线性空间

矢量空间 [444] 张成空间 [447] 线性映射 [451] 线性映射的坐标表示 [454]

矩阵与线性映射 [456] 矢量空间的表示 [458] 对偶空间 [460] 子空间 [464] 子空间的直和补空间 [465] 线性映射的结构 [468] 狄拉克符号 [470] 内积 内积空间 [471]

第四章 线性方程组

线性方程组 [475] 高斯消元法解线性方程组 [477] 线性方程组解的结构 [482] 线性映射的结构 2 [484] 超定线性方程组 [485]

第五章 张量代数

张量 [488] 爱因斯坦求和约定 [498] 张量的坐标变换 [501] 协变和逆变 [504] 张量积 [508] 张量的分类 [510]

第六章 未归类

张量积空间 [513] 线性代数导航 [521] 施密特正交归一化 [522] 平面旋转变换 [525] 对易算符 [526] 正交子空间 [527] 矩阵指数 [529]

第四部分 代数学进阶

第一章 群论

群 [534] 置换的奇偶性 [539] 子群 [541] 陪集和同余 [543] 正规子群 [547] 群的直积和直和 [551] 单群 [553] 群同态 [555] 群作用 [557] 商空间 [560] 群的表示 [561] 自由群 [562] 群的自由积 [564] 群论中的证明和习题解答 [565] 李群 [566]

第二章 环论

环 [569] 环的理想和商环 [571] 环同态 [573] 整环 [574] 素理想与极大理想 [577] 爱森斯坦判别式 [579] 真因子树 [581] 唯一析因环 [583] 主理想整环 [584] 欧几里得环 [584]

第三章 域论

域 [588] 域上的代数 [590] 四元数 [593] 域的扩张 [598]

第四章　外代数

外代数 [603]　外导数 [604]　霍奇星算子 [608]

第五章　复半单李代数

李代数 [611]　李代数的子代数、理想与商代数 [614]

第五部分　拓扑学

第一章　点集拓扑

拓扑空间 [619]　点集的内部、外部和边界 [622]　连续映射和同胚 [625]　紧致性 [628]　连通性 [632]　道路连通性 [634]　分离性 [636]　乘积拓扑 [640]　商拓扑 [643]　映射空间 [648]　空间偶和带基点空间 [649]　拓扑群 [651]

第二章　同伦论

映射的同伦和空间的同伦 [654]　可缩空间 [657]　基本群 [658]　高阶同伦群 [663]　球面的同伦群 [668]　覆叠空间 [668]　基本群的计算 [670]

第六部分　微分几何

第一章　欧几里得空间

三维欧几里得空间中的曲线 [677]　光滑映射 [681]　切空间（欧几里得空间） [683]

第二章　流形

流形 [688]　子流形 [691]　积流形 [691]　流形上的切空间 [692]　抽象指标 [694]　纤维丛 [696]　向量丛和切丛 [698]　微分形式 [700]　费罗贝尼乌斯定理 [703]

第三章　联络

联络 (向量丛) [708]　曲率 (向量丛) [711]　平行性 (向量丛) [713]　和乐群（向量丛） [715]

第四章　黎曼几何

黎曼度量与伪黎曼度量 [719]　Levi-Civita 联络 [720]

第七部分　概率与统计

第一章　基础

随机变量 概率分布函数 [723]　随机变量的变换 [726]　多变量分布函数 [728]　高斯分布（正态分布） [731]　二项分布 [732]　中心极限定理 [736]　二维随机走动 [737]　平均值的不确定度 [738]

第八部分　偏微分方程和特殊函数

第一章　基础

分离变量法 [742]　拉普拉斯方程 [742]　球坐标系中的矢量算符 [742]　球坐标系中的拉普拉斯方程 [744]　柱坐标系中的矢量算符 [746]　柱坐标系中的拉普拉斯方程 [747]　泊松方程 [748]　球坐标系中的亥姆霍兹方程 [751]　柱坐标中的亥姆霍兹方程 [752]　勒让德多项式 [753]　连带勒让德多项式 [755]　贝塞尔函数 [757]　球贝塞尔函数 [759]　球谐函数 [761]　球谐函数表 [765]　Wigner D 矩阵 [766]　平面波的球谐展开 [767]　球谐波的归一化 [769]　分离变量法与张量积空间 [770]　广义球谐函数 [771]　误差函数 [774]　虚误差函数 [776]　超几何函数 [776]　椭圆积分 [777]　库仑函数 [778]

第九部分　数学分析

第一章　基础

集合的极限 [783]　度量空间 [785]　度量空间中的概念 [787]　柯西序列 完备度量空间 [790]　完备空间 [792]　范数赋范空间 [794]　里斯引理（赋范空间） [798]　投影算符 [799]　柯西—施瓦茨不等式 [800]　巴拿赫空间 [801]　巴拿赫定理 [803]　一致收敛 [805]　黎曼积分与勒贝格积分 [805]　数学分析笔

记 [806]

第十部分　泛函分析

第一章　笔记

希尔伯特空间 [816]　泛函分析笔记 1 [816]　泛函分析笔记 2 [824]　泛函分析笔记 3 [830]　泛函分析笔记 4 [832]　泛函分析笔记 5 [833]　装备希尔伯特空间 [840]　宇称算符 [842]

第二章　其他

堆放排列组合 [844]　范德蒙恒等式 [844]　解三棱锥顶角 [846]　足球顶点坐标的计算 [847]　Clebsch－Gordan 系数 [852]　Wigner 3j 符号 [855]　Wigner 6j 符号 [857]　Wigner 9j 符号 [857]　二元关系 [858]　群论笔记 [859]　范畴论 [861]　余元公式 [866]　代数学基本定理 [869]

第十一部分　经典力学

第一章　质点运动学

物理量和单位转换 [873]　无量纲的物理公式 [876]　位置矢量 位移 [878]　速度 加速度（一维） [879]　速度 加速度 [881]　圆周运动的速度 [883]　圆周运动的加速度 [885]　匀加速直线运动 [887]　匀加速运动 [889]　匀速曲线运动 [890]　极坐标中的速度和加速度 [891]　速度的坐标系变换 [892]　加速度的坐标系变换 [894]

第二章　质点动力学

力的分解与合成 [898]　牛顿运动定律 惯性系 [899]　圆周运动的向心力 [900]　重力 重量 [902]　功 功率 [903]　动能 动能定理（单个质点） [904]　力场 势能 [905]　机械能守恒（单个质点） [909]　动量 动量定理（单个质点） [910]　角动量 角动量定理 角动量守恒（单个质点） [911]　简谐振子 [913]　受阻落体 [915]　单摆 [917]　圆锥摆 [918]　傅科摆 [919]　惯性力 [920]　滑块和运动斜面问题 [922]　离心力 [924]　科里奥利力 [925]　旋转参考系的“机械能守恒”

[926] 地球表面的科里奥利力 [927]

第三章 质点系与刚体

自由度 [930] 质点系 [931] 质心 质心系 [931] 质点系的动量 [939] 刚体 [939] 轻杆模型 [941] 动量定理 动量守恒 [944] 质点系的动能 柯尼希定理 [946] 力矩 [947] 刚体的静力平衡 [950] 角动量 [952] 角动量定理 角动量守恒 [954] 二体系统 [957] 二体碰撞 [960] 刚体定轴转动 转动惯量 [962] 平行轴定理与垂直轴定理 [966] 常见几何体的转动惯量 [967] 刚体的平面运动方程 [970] 惯性张量 [973] 刚体的瞬时转轴 [977] 刚体定轴转动 2 [978] 刚体的动能定理 [978] 刚体的运动方程 [979] 刚体运动方程（四元数） [981]

第四章 软体和流体力学

绳的法向压力 [984] 杨氏模量 [986] 流 流密度 [986] 浮力 阿基米德原理 [988] 伯努利方程 [989] 黏度 [990] 流体运动的描述方法 [990] 物质导数（实质导数） [992]

第五章 振动与波动

振动的指数形式 [995] 能量法解谐振动问题 [997] 拍频 [999] 受阻简谐振子 [999] 简谐振子的品质因数 [1001] 简谐振子受迫运动 [1004] 共振 [1006] 平面波 [1008] 波包 [1011] 高斯波包 [1012] 群速度 [1013] 多普勒效应（一维匀速）[1013] 多普勒效应 [1016] 一维波动方程 [1017] 二维波动方程 [1020] 波的能量 [1021] 波的强度 [1023] 冲击波 [1024]

第六章 中心力场问题

万有引力 引力势能 [1027] 壳层定理 [1028] 开普勒三定律 [1033] 中心力场问题 [1034] 开普勒问题 [1037] 开普勒问题的运动方程 [1040] 拉普拉斯—龙格—楞次矢量 [1041] 轨道方程 比耐公式 [1042] 开普勒第一定律的证明 [1043] 开普勒第二和第三定律的证明 [1045] 散射 [1047] 卢瑟福散射 [1047] 闭合轨道的条件 [1048] Bohr-Sommerfeld 氢原子模型 [1049]

第七章 分析力学

拉格朗日方程 [1052] 拉氏方程和极值问题 [1057] 贝尔特拉米等式 [1058]

最速降线问题 [1058] 单摆（大摆角） [1061] 双摆和三摆 [1062] 达朗贝尔定理 [1063] 最小作用量哈密顿原理 [1065] 运动积分 [1067] 勒让德变换 [1073] 哈密顿正则方程 [1075] 泊松括号 [1078] 正则变换 2 [1079]

第八章 轨道力学

二体问题综述 [1083] 轨道参数 时间变量 [1088] 限制性三体问题 [1091] 雅可比常量 [1093] 雅可比常量 2 [1094] 拉格朗日点 [1096] 光子火箭 [1098] 太空电梯 [1100]

第十二部分 光学

第一章 几何光学

惠更斯原理 [1105] 光的折射 斯涅尔定律 [1107] 薄透镜 [1108]

第二章 波动光学

可见光谱 [1112] 双缝干涉中一个重要极限 [1112] 杨氏双缝干涉实验 [1114] 劳埃德镜实验 [1116] 普通光源的发光机理 [1118] 单色光 [1118] 高斯光束 [1119]

第十三部分 电动力学

第一章 基础

电流 [1123] 电流密度 [1123] 库仑定律 [1123] 电场 [1124] 磁场 [1129] 电势 电势能 [1130] 电偶极子 [1137] 电偶极子 2 [1141] 导体 [1142] 电压 [1143] 电容 [1145] 电介质的简单模型 [1148] 电介质的微观结构 [1149] 电极化强度 [1152] 电极化强度与极化电荷的关系 [1153] 极化电流 [1154] 介质中的静电场 [1155] 电阻 欧姆定律 [1157] 电感 [1159] 电场的高斯定律 [1161] 电场的高斯定律证明 [1166] 导体的静电平衡 [1168] 磁场的高斯定律 [1169] 电场的能量 [1169] LC 振荡电路 [1170] 力电振动类比 [1173] LC 受迫振荡电路 [1173] 比奥萨伐尔定律 [1175] 安培环路定理 [1179] 洛伦兹力 [1180] 霍尔效应 [1182] 磁场的能量 [1183] 磁通量 [1185] 安培力 [1186] 磁矩 [1187] 磁场中闭合电流的合力 [1188] 磁场中闭合电流的力矩 [1189] 法拉第电磁感应定律 [1191]

位移电流 [1192] 分子电流和分子磁矩 [1192] 磁介质 [1194] 磁化强度 [1195] 顺磁质的磁化 [1198] 抗磁质的磁化 [1198] 有磁介质时的安培环路定理 [1199] 电路 [1201] 基尔霍夫电路定律 [1202] 电容—电阻电路充放电曲线 [1208] 惠斯通电桥 [1209]

第二章 电动力学 2

电荷守恒 电流连续性方程 [1212] 电多极展开 [1212] 电磁场标势和矢势 [1213] 磁标势 [1214] 磁矢势 [1215] 磁多极矩 [1216] 规范变换 [1217] 洛伦兹规范 [1218] 库仑规范 [1219] 电磁场的能量守恒 坡印廷矢量 [1220] 麦克斯韦方程组 [1223] 麦克斯韦方程组（介质） [1224] 电磁场推迟势 [1225] 磁单极子 [1226] 非齐次亥姆霍兹方程 推迟势 [1227] 电场波动方程 [1230] 真空中的平面电磁波 [1230] 平面电磁波的能量叠加 [1231] 电磁波包的能谱 [1232] 电偶极子辐射 [1233] 介质中的波动方程 [1235] 导体中的电磁波 [1236] 菲涅尔公式 布儒斯特角 [1237] 盒中的电磁波 [1238] 电磁场的动量守恒 动量流密度张量 [1240] 磁旋比 玻尔磁子 [1243]

第三章 电动力学 3

电磁场的参考系变换 [1246] 拉格朗日电磁势 [1247] 电磁场角动量分解 [1249]

第十四部分 相对论

第一章 狭义相对论

狭义相对论的基本假设 [1253] 事件与尺缩效应 [1256] 时间的变换与钟慢效应 [1260] 洛伦兹变换 [1263] 斜坐标系 [1265] 斜坐标系表示洛伦兹变换 [1268] 约化光速 [1268] 洛伦兹变换的代数推导 [1271] 相对论速度变换 [1274] 相对论加速度变换 [1277] 时空的四维表示 [1279] 闵可夫斯基空间 [1282] 光的多普勒效应 [1283] 洛伦兹群 [1286] 托马斯进动 [1291]

第二章 相对论动力学

相对论动力学假设 [1296] 尘埃云的能动张量 [1298]

第三章　广义相对论

双生子佯谬 [1304]　因果结构 [1308]

第十五部分　量子力学

第一章　量子力学 1

玻尔原子模型 [1313]　原子单位制 [1315]　自然单位制 [1320]　指数衰减 [1321]　精细结构常数 [1321]　量子力学与矩阵 [1322]　算符和本征问题 [1325]　定态薛定谔方程 [1328]　薛定谔方程 [1331]　不确定原理 [1332]　多维空间中的量子力学 [1333]　无限深势阱 [1335]　无限深势阱中的高斯波包 [1337]　升降算符 [1339]　简谐振子（升降算符）[1340]　简谐振子升降算符归一化 [1343]　简谐振子（级数）[1344]　一维自由粒子（量子）[1345]　一维自由高斯波包（量子）[1346]　平面波的的正交归一化 [1348]　无限深圆形势阱 [1350]　有限深球势阱 [1351]　一维散射（量子）[1351]　一维散射态的正交归一化 [1354]　一维 delta 势能散射 [1359]　一维阶梯势能 [1361]　有限深势阱 [1362]　方势垒的定态波函数 [1365]　方势垒散射数值计算 [1369]　动量表象的薛定谔方程 [1372]　拉比频率 [1373]　二维无限深势阱 [1374]　平均值（量子力学）[1376]　守恒量（量子力学）[1377]　好量子数 [1379]　薛定谔方程的分离变量法 [1380]　对易厄米算符与共同本征矢 [1381]　矩阵对易与共同本征矢 [1384]　类氢原子的约化质量 [1385]　概率流密度 [1386]　算符的矩阵表示 [1388]　角动量（量子）[1391]　球坐标系中的角动量算符 [1393]　轨道角动量升降算符归一化 [1394]　自旋角动量 [1395]　自旋角动量矩阵 [1396]

第二章　量子力学 2

算符的指数函数 波函数传播子 [1399]　平移算符 [1400]　旋转算符 [1401]　角动量加法（量子力学）[1403]　氢原子基态的波函数 [1407]　球坐标系中的定态薛定谔方程 [1409]　柱坐标系中的薛定谔方程 [1411]　球坐标中的薛定谔方程 [1412]　量子力学中的变分法 [1413]　含时微扰理论 [1414]　几种含时微扰 [1417]　含连续态的微扰理论 [1419]　量子散射 [1420]　量子散射的延迟 [1425]　波恩近似（散射）[1426]　质心系中的多粒子问题 [1427]　三维简谐振子（球坐标）[1428]　球面散射态与平面散射态的转换 [1429]　库仑散射

（量子）[1431] 量子力学的基本假设[1435] 电磁场中的薛定谔方程及规范变换[1436] 库仑规范（量子）[1437] 长度规范[1437] 速度规范[1438] 加速度规范[1439] Volkov 波函数[1441] 电磁场中的类氢原子[1443] 跃迁概率（一阶微扰）[1443] 单电子跃迁截面（一阶微扰）[1446] 氢原子电离计算（一阶微扰）[1447] 氢原子的选择定则[1449] 跃迁偶极子矩阵的三种形式[1451] 康普顿散射[1453] 全同粒子[1454] 泡利不相容原理[1457] 粒子交换算符[1458] 对称化与反对称化[1463] 密度矩阵[1466] Hartree-Fork 方法[1468] 多通道散射[1472] Adiabatic 笔记[1473] 晶体衍射[1474] 超导唯象解释——伦敦方程[1475]

第三章 量子力学与量子场论

前言[1479] 基本概念[1479] 全同粒子的统计[1493] 近似理论：微扰[1496] 角动量 2[1502] 冷原子基本知识[1507] 两个原子间的相互作用[1516] Feshbach 共振[1523] BCS-BEC Crossover 的平均场描述[1525] BEC 超流[1529]

第十六部分 原子分子物理

第一章 原子

类氢原子的定态波函数[1536] 氢原子波函数分析[1539] 电子轨道与元素周期表[1540] 能项符号[1542] 氢线（21 厘米线）[1543] 兰姆位移[1543] Keldysh 参数[1544] ADK 电离率[1545] 抛物线坐标系中的类氢原子定态波函数[1545] FROG[1546]

第二章 分子

双原子分子势能曲线[1555]

第十七部分 热力学和统计力学

第一章 热力学

气体分子对容器壁的压强[1558] 理想气体状态方程[1560] 温度 温标[1562] 理想气体的内能[1563] 压强体积图[1563] 热平衡 热力学第零定律[1564]

热传导定律 [1565] 热力学第一定律 [1566] 等压过程 [1566] 等体过程 [1570] 等温过程 [1572] 热容量 [1573] 绝热过程 [1574] 熵 [1576] 准静态过程 [1580] 卡诺热机 [1581] 分子平均碰壁数 [1585] 气体分子的速度分布 [1586] 麦克斯韦—玻尔兹曼分布 [1588] 理想气体分压定律 [1589] 饱和蒸汽压 [1589] 大气密度和压强 [1592] 黑体辐射定律 [1594] 维恩位移定律 [1594] 斯特藩—玻尔兹曼定律 [1595]

第二章 统计力学

相空间 [1598] 理想气体的状态密度（相空间） [1598] 理想气体单粒子能级密度 [1599] 理想气体（微正则系综法） [1600] 正则系宗法 [1601] 理想气体（正则系宗法） [1601] 理想气体（巨正则系综法） [1604] 等间隔能级系统（正则系宗） [1606] 巨正则系综法 [1608] 量子气体（单能级巨正则系综法） [1610] 量子气体（巨正则系宗） [1611] 光子气体 [1612] 理想气体的熵：纯微观分析 [1613] 伊辛模型 [1615] 统计力学公式 [1616]

第十八部分 近代物理

第一章 量子场论

标量场 [1622] 洛伦兹群覆盖群 SL(2,C) 的不可约表示 [1626]

第二章 高能物理

基本粒子 [1633]

第三章 弦理论

引力量子化 [1635] 弦论概述 [1635] 弦论的种类 [1637] BRST 量子化 [1638] RNS 超弦 [1645]

第四章 宇宙学

宇宙的演化 [1654] 宇宙学红移 [1656] 宇宙中的距离 [1658] Friedmann-Robertson-Walker (FRW) 度规 [1659]

第五章 宇宙学扰动

标量扰动 [1661] 张量扰动 [1662]

第六章 引力波

引力波的几何描述 [1667] TT 规范 [1670] 引力波和测试质量的相互作用 [1674]

第十九部分 科学计算

第一章 Matlab 简介

计算物理导航 [1681] Matlab 简介 [1682] Matlab 的变量与矩阵 [1685] Matlab 的判断与循环 [1696] Matlab 的函数 [1700] Matlab 画图 [1704] Matlab 画箭头矢量场图 [1706] Matlab 球坐标中的分布图 [1708] Matlab 的程序调试及其他功能 [1710]

第二章 Python 简介

Python 简介 [1715] Python 基本变量类型 [1719] Python 数据类型 [1720] Numpy 库 [1721] Python 字符串处理 [1726] Python 文件读写 [1732] Python 判断与循环 [1736] Python 函数 [1738] Python 画图 [1743] Python 的类 [1747] Python 模块 [1748] Python 异常处理 [1751] SciPy 数值微分与积分 [1752] SciPy 最小二乘法 [1755]

第三章 Linux 简介

Linux 基础 [1761] 在 Linux 上编译 C/C++ 程序 [1764] Makefile 简介 [1766] Makefile 笔记 [1769]

第四章 C++ 简介

C++ 基础 [1773] C++ 的整数 [1775] C/C++ 多文件编译笔记 [1777] C++ 的 namespace [1778] C++ 异常处理 [1779] 调试 C++ 程序 [1780] 数据结构：密矩阵 [1782] 数据结构：带对角矩阵 [1783] C++ 矩阵类的实现 [1785] SLISC 库简介 [1788] SLISC 的密矩阵类 [1790] SLISC 矩阵的基本运算 [1793] SLISC 密矩阵的切割 [1795] SLISC 的 Mcoo 矩阵类 [1798] SLISC 的 band 类 [1800]

SLISC 的文件读写 [1801] SLISC 的计时工具 [1803] SLISC 的 matt/matb 文件格式 [1804] C++ 中的 SFINAE 技巧 [1807]

第五章 其他

BLAS 简介 [1813] GNU Scientific Library [1817] 文本文件与字符编码 [1818] 正则表达式 [1820] LaTeX 结构简介 [1823] 校验和 [1826] GitHub Desktop 的简单使用 [1828] cuBLAS 库 [1831] Julia 分形 [1837]

第六章 数值验证及常用算法

二项式定理（非整数幂）的数值验证 [1841] 二分法 [1841] 多区间二分法 [1843] 冒泡法 [1845] 高斯消元法程序 [1847] Nelder-Mead 算法 [1851] 最小二乘法的数值计算 [1855] 数值积分（梯形法）[1857] 稀疏矩阵 [1858] 函数求值 [1859] 离散傅里叶变换 [1860] 离散正弦变换 [1867]

第七章 计算机图形学

图像坐标系 [1870] 三维投影 [1870] 相机模型 [1872] 由图像坐标计算射线 [1874] 计算 3D 艺术画 [1874] 相机的定位 [1877] 长方形相机定位法 [1878] 解三棱锥 1 （Matlab） [1879]

第八章 微分方程数值解

简谐振子受迫运动的简单数值计算 [1882] 天体运动的简单数值计算 [1883] 常微分方程（组）的数值解 [1886] 中点法解常微分方程（组） [1888] 四阶龙格库塔法 [1891] 刚体转动数值模拟 [1894]

第九章 偏微分方程数值解

一维波动方程的数值解 [1901]

第十章 一维薛定谔方程数值解

一维势能束缚态数值解（试射法）[1907] 无限深势阱中的高斯波包 Matlab 代码 [1915] 方势垒定态波函数程序 [1917]

第十一章 氢原子薛定谔方程数值解

氢原子波函数画图程序 [1923] 球谐函数数值计算 [1924] Crank-Nicolson 算法（一维） [1925] Gauss-Lobatto 积分 [1926] FEDVR 网格 [1928] Lanczos 算法 [1931] 指数格点 [1936] 虚时间法求基态波函数 [1936] 电磁场中氢原子薛定谔方程数值解 [1937] 氢原子球坐标数值解 TDSE 2 [1940]

第十二章 氦原子薛定谔方程数值解

氦原子数值解 TDSE 笔记 [1944] 氦原子波函数数值分析 [1954]

第二十部分 附录

第一章 附录

小时百科符号与规范 [1967] 常见物理量 [1969] 国际单位制 [1970] 物理学常数 [1973] 国际单位制词头 [1975] 厘米—克—秒单位制 [1975] 高斯单位制 [1976] 小时百科图标 [1978]

第一部分

科普

第一章

经典力学

经典力学及其他物理理论

科学理论的可证伪性

著名的奥地利哲学家波普尔（Popper）对科学的划界是：一个命题是科学的，当且仅当它是**可证伪**的．如果有人提出一个物理理论，那么既可以尝试用它来计算已有的实验结果，也可以用它来预言一些没有做过的实验结果．如果在实验误差范围内，所有实验与理论计算得到的结果一致，那么就还没有证据表明这个理论是错误的，但也不能说它是绝对正确的．毕竟我们不可能把一个理论的每一种实验，每一套参数都做一遍．然而一旦有一个实验与该理论的计算结果不相符，那么就可以证明这个理论是错误的[1]，这就是科学理论的**可证伪性**．

然而可惜的是，在物理学中目前还没有一个理论可以在任意范围内解释实验或观测结果，所有的理论（如经典力学，相对论，量子力学，量子场论）都只在一定的范围内成立．我们能做的仅仅是不断创造与实验符合得更精确，且适用范围更广的理论．这样一来，给一个曾经普遍接受的理论打上“错误”的标签似乎有些不妥，于是我们一般称其为**在适用范围内成立**．

物理理论的适用范围

本科物理学课程一般从经典力学开始，我们将其作为例子来讨论．经典力学在**宏观低速**的范围内适用．粗略而言，“宏观”要求物体的质量远大于原子的质量，“低速”要求物体的运动速度远小于光速．事实上还有一个条件是“弱引力场”，例如由于水星离太阳较近，引力场较强，导致其轨道与经典力学的计算出现偏差（轨道进动）．所以严格来说，经典力学是一个被证伪的理论．

若上述中只有“低速”条件不满足，我们就需要使用狭义相对论，若“弱引力场”条件不满足，就需要广义相对论（狭义相对论是广义相对论的一部分），若“宏观”条件不满足，就需要量子力学，若都不满足，那么现在还没有非常完善的理论可以计算（我们预先把它叫做**万物之理，theory of everything**）．

以相对论（狭义和广义的统称）为例，它所适用的范围既包含了经典力学

[1] 当然首先要考虑是否存在计算错误，实验操作失误，或者存在未考虑到的因素

适用的范围，又包含了“高速”和“强引力场”，所以原则上相对论可以完全取代经典力学. 由于经典力学在适用的范围内已经得到几百年来大量的实验验证，那么如果相对论是正确的，在经典力学适用的范围内，用相对论计算问题就应该得到同样的结果[2]. 值得注意的是，相对论提出的一些物理概念与经典力学大相径庭. 经典的万有引力定律提出任何两个物体之间都存在万有引力，而相对论却指出并不存在引力，而是有质量的物体扭曲了周围的时空，使周围物体的运动方式不同. 既然相对论的适用范围更广，那么至少从目前看来相对论对物理现象的解释才是更可信的，而经典力学的解释是错误的.

既然如此，为什么我们还要先学习经典力学呢？首先无论在概念上还是数学上，经典力学比相对论简单得多. 其次在日常生活或生产中我们接触的绝大部分运动都在经典力学的适用范围内. 第三，相对论中同样会出现“参考系”，“速度”，“能量”，“动量”，等概念，这些概念只有先学习经典力学才会有一个初步的认识，才能继续学习相对论. 最后，通过学习经典力学可以了解物理中常见的数学工具，包括一些基础的微积分，矢量分析，线性代数等，这些数学在物理的其他领域更是无处不在.

以上论述同样适用于量子力学与经典力学的关系. 量子力学除了在经典力学的范围适用，还描述了微观粒子的运动. 总而言之经典力学在现代的物理学中只是一个简单的近似模型，提出的一些原理并不正确，公式也只是一种近似. 一些民间科学家（俗称**民科**）时常企图推翻牛顿定律，显然是还不了解这点.

另一方面，即使是相对论和量子力学也并非完美无瑕，通常所说的量子力学是指“非相对论量子力学”，即同样要求“低速”和“弱引力场”. 量子场论（**Quantum Field Theory**）结合了量子力学和狭义相对论，但仍然没有解决“强引力场”的问题.

经典力学

- 本词条存在未完成的内容.

[2]准确来说，二者计算结果的误差需要在实验的测量误差范围内.

预备知识 物理理论[3]，高中物理力学部分

现在开始我们假设读者已经具有高中物理中的基本力学知识，如果你觉得没有把握，不妨复习一下.

经典力学

经典力学描述的是满足宏观低速弱引力场条件下，若干的粒子（即**质点**）受若干力以后的运动情况．质点没有转动的概念，只能移动，任何大小不可忽略的物体都可以看作由相互作用的质点组成，所以经典力学可以描述任何宏观低速物体的运动.

牛顿三定律[899] 可以作为经典力学的**公设**．公设在这里指的是某个理论体系内的基本假设．由这些假设再加上数学推导可以得出理论体系中的所有推论，例如动量定理动量定理[944]，角动量定理角动量定理[954]，动能定理[904]，开普勒三定律[1033] 等．公设本身无法被推导而来[3]，只能通过实验验证.

我们可以将牛顿三定律想象成一个懂经典力学的电脑，只需输入初始时刻所有粒子的状态（位置和速度/动量），以及每个粒子的受力情况，就可以得到接下来每个粒子的运动方式，具体来说就是粒子位置关于时间的函数．我们在研究某个问题时，往往把讨论范围内的粒子叫做**系统**.

例如我们要研究太阳系中行星的运动，就把太阳系内的天体看成是系统内，把宇宙中其他所有天体视为系统外．另外我们往往把这些行星看作是粒子（质点），这是因为比起他们的距离来，他们本身的大小可以忽略不计[4]．这个问题中，如果我们选择合适的参考系，系统外对系统内的作用也可以忽略不计.

我们只要知道某时刻（初始时刻）它们的位置和动量[5]，将它们输入“牛顿三定律”这台电脑中，就可以得到接下来所有天体的运动情况，即任意时刻任意粒子的位置和动量.

---

[3] “无法被推导”是在这个理论内部而言的．一个理论的基本假设可以由更高级，使用范围更广的理论的基本假设推导出来．例如牛顿第二定律可以由狭义相对论或量子力学的基本假设推导而来.

[4] 在物理中，我们经常使用这种手段简化问题，并将其称为**近似**

[5] 在大学物理中，你会发现我们更多地使用动量而不是速度，在已知质量的情况下，它们所包含的信息是一样的，即一一对应

刚体

质点的模型忽略了物体的实际大小，如果要考虑一个大小不可忽略的物体的转动，以及若干个大小不可忽略的物体之间的相互作用该怎么办？事实上，这些物体也是由原子和分子构成的[6]．在固体中，原子之间的作用力使它们相对位置较难改变，例如可以想象每个原子和与它相邻的原子都以较硬的弹簧相连．所以，我们同样可以把这些物体看作是由质点组成得．这样，我们就可以用质点的力学来描述有形状有质量分布的物体了．在力学中我们经常使用一种理想模型叫做**刚体**[939]，即假设物体中连接这些质点的弹簧的劲度系数为无限大，或者把他们看成不可伸缩的棍子．所以如果我们近似认为一个物体无限硬，忽略其自身的形变，我们就可以用刚体的模型来描述它．

受力决定运动

经典力学中最简单的一类问题是：已知初始时刻若干粒子的运动状态，以及接下来每一个粒子的受力．以直线运动为例，F 随**时间** t 的变化（即函数 $F(t)$）．根据牛顿第二定律，可以求出每个粒子任意时刻的加速度[879] a（正比于 F），进而求出速度 v 随时间的变化，也就是求“加速度—时间”曲线下面在某段时间的面积（这在数学上叫做**定积分**[198]）．有了速度随时间的变化，我们又可以通过面积求位置随时间的变化（同样是定积分，求“速度—时间”曲线下面的面积）．例如我们知道一个沿直线运动的粒子受一个恒力（例如自由落体），它的加速度是恒定的，这个过程如图 1．这种方法适用于任何形状的 $F(t)$ 曲线．

简谐振子

为了方便下文举例，我们先来介绍力学中的一个经典模型即**简谐振子**[913]．假设无摩擦的光滑水平直轨道上有一根质量不计的弹簧一端固定不动，而另一端上固定了一个质点．如果给这个质点一个初速度或初位移，那么它将会沿着轨道往复运动，我们把这种运动叫做**简谐运动**．顺带一提，我们以后会看到位置关于时间的函数恰好是 $\cos(\omega t)$，ω 是**角频率**，代表振动的快慢．

[6]经典力学不能准确描述原子的结构和运动，要使用量子力学，但我们这里讨论的是组成物体后的宏观运动，所以即使用经典力学来描述原子得到的宏观运动也是正确的

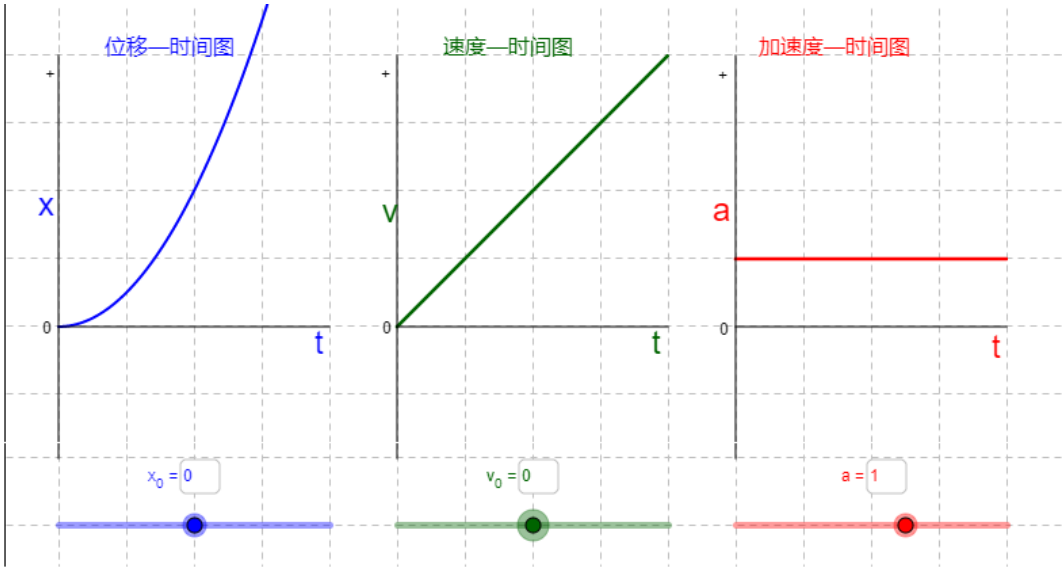


图 1: 匀加速直线运动（见“互动演示”页面），加速度曲线下方从 0 到 t 的面积是速度，速度曲线下方从 0 到 t 的面积是位移

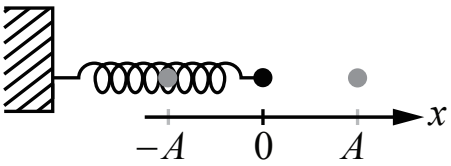


图 2: 简谐振子

更困难的问题

事实上我们很少会遇到“已知每个质点的 $F(t)$”这么简单的条件．即使在简谐振子这样简单的模型中，我们也不可能事先就知道质点受力关于时间的变化情况．因为受力分析得到的是力**关于位置**的变化（$F(x) = -kx$，k 是劲度系数），而粒子的位置变化（函数 $x(t)$）又取决于受力．这样一来，我们就没有办法像上面一样直接用定积分来解决这个问题了．

为了解决这样的问题，我们需要使用**常微分方程**[211]．但现在我们可以用以下的简单思路进行一个近似的数值计算：例如在简谐振子的模型中，假设质点初始时刻处于弹簧的平衡位置 $x = 0$（受力为零）并具有一个初速度 v_0，在很小一段时间 Δt 内，由于 x 仍然很小，我们可以继续认为质点不受力，做匀速运动．这样当它移动到 $x = v\Delta t$ 时，我们重新计算它的受力得到 $F = -kx = -kv\Delta t$，然后假设在下一段很短的时间 Δt 内仍然假设质点受力不变，计算出位置和速

度的变化．如此循环，就可以数值计算出 $x(t)$ 的一系列散点，连起来就得到了位移—时间曲线．当 Δt 取得越小，这种近似就越精确．

事实上力学中（或者物理中）只有在很少的情况下微分方程的解才能用公式表达出来（解析解），其他情况我们只能用计算机进行数值解[1886]．

材料力学

未完成：研究可以发生形变的固体，在建筑和制造业有广泛应用．

未完成：波动在哪里介绍？

流体力学

气体液体同样可以用理想化的力学模型来描述，我们把它叫做**流体**，流体力学是力学中十分重要的分支，轮船，高铁，飞机和火箭的设计都要使用它．

经典力学只能精确描述固体和液体的一些宏观力学性质，而另一些性质，例如导电性（包括超导），颜色，是否透明等，需要涉及到原子分子的性质，以及他们的相互作用．物理学的一大分支——凝聚态主要就是在研究固体和液体．

力的分类

牛顿三定律只给出了运动和力的关系以及作用力和反作用力，却没有限定这些力是什么力以及该如何具体计算．现代物理告诉我们，宇宙中所有已知的力（也叫相互作用）都可以归为四种，分别是**万有引力（gravitational force）**，**电磁力（electromagnetic force）**，**弱相互作用（weak interaction）**，和**强相互作用（strong interaction）**．后两种是短程力，通常只存在于原子核内部，不适用于经典力学，先不讨论．

你可能会问，我们高中学了那么多摩擦力，支持力，弹力，浮力，为什么都没有出现呢？事实上这些力都只是微观电磁力的宏观表现，都可以归类到电磁力中．我们知道日常生活中的物质是由原子构成的，原子由原子核与电子构成，前者带正电，后者带负电，整体呈电中性．除了核反应以外，任何原子之间的作用都可以看成是由许多电子和原子核构成的系统内部的电磁作用力．由于

经典力学是宏观的理论，所以我们不能用经典力学直接计算原子内部的电磁作用带来的运动规律，但可以计算摩擦力，支持力等宏观的力带来的宏观运动.

在经典力学中万有引力是一个十分重要的研究对象，我们将在万有引力和天体运动中（未完成）详细讨论.

分析力学

分析力学的数学形式比牛顿定律复杂，但和牛顿力学是等效的，即任何分析力学计算的问题用牛顿定律同样能计算并得到完全相同的结果，反之亦然. 分析力学常见的形式有两种，一种是拉格朗日力学，使用拉格朗日方程，另一种叫哈密顿力学，使用哈密顿方程. 量子力学就是在哈密顿力学的基础上建立起来的，其重要性可见一斑.

相比与牛顿力学，分析力学的优势主要有两点. 一是当计算的系统越来越复杂的时候（例如蒸汽机等复杂的机械结构）分析力学一般能更快捷地列出微分方程组，而牛顿力学所需要的受力分析会变得非常复杂. 另一方面，分析力学能使我们站在更高的高度看问题（例如可以分析出系统中的守恒量）.

为什么复杂的数学形式反而在一些情况下使列方程变得简单呢？如果把牛顿定律比作电脑，那么分析力学就相当于一个更智能的电脑，对于系统中的某些物体，无需告诉电脑它的受力，只需告诉电脑它物体之间的运动的约束即可. 约束简单来说就是怎样的运动是不可能的：例如单摆中的质点就“不可能”沿着绳的方向运动，两个咬合的齿轮“不可能”一个转动一个不转. 用约束条件代替受力分析，在适当的时候可以大大提高列方程的效率.

例如在以下活塞结构中，活塞，连杆，飞轮被约束起来，活塞的运动状态（位置和速度）决定了整个系统的运动状态（未完成）

狭义相对论

（未完成）同样讨论质点组成的系统受力后的运动情况. 牛顿三定律只适用于宏观低速弱引力场条件，如果粒子的速度相对光速不可忽略，那就需要使用“更精确”的牛顿第二定律，且惯性参考系切换也变得更复杂（洛伦兹变换）. 当速度越低，狭义相对的计算结果越接近牛顿力学. 相对论同样并不适用微观.

广义相对论

未完成：加上强引力场和非惯性系.

混沌

（未完成）二体运动是令人愉悦的简单曲线（可参考词条“二体运动”），但三体问题不是．事实上，三体问题除去一些简单近似周期解（可参考“限制性三体问题”），之外没有解析解.

还有更复杂的问题，例如：**湍流**.

动量和能量一维势能曲线

预备知识 经典力学[4]

动量和动能

动量描述了了物体运动的物体的惯性．惯性可以简单理解为让一个运动的粒子停下来有多困难．这个困难程度可以用力乘以时间表示．例如一个在光滑水平面有动量为 $p = mv$ 的箱子，要使它停下来，人就要沿运动反方向以恒力 F 推箱子，若一段时间 t 后箱子停了下来，那么有 $p = Ft$．同理，如果要使箱子从静止达到动量 p，就需要以恒力 F 作用在箱子上同样的时间．所以动量是力在时间上的累加.

相比动量，能量可以看作是力在空间上的累加．同样是箱子的例子，如果想让静止的箱子达到动能 $E_k = mv^2/2$（角标 k 表示 kinetic energy），就需要用恒力 F 推箱子，在力的方向移动距离 s，使 $E_k = Fs$.

理论上，有了牛顿定律我们就可以求出所有物体的运动情况，那为什么还需要动量和动能（能量）？因为动量和能量在封闭系统中是守恒的，可以只根据初末状态就得到一些结论而不需要知道具体过程．例如两个箱子碰撞，我不需要知道碰撞用了多久，缓冲距离是多少，材料的性质等等就可以根据碰撞前的状态求出碰撞后的状态．又例如物体从静止开始延光滑轨道滚落一定高度，

不需要知道过程就可以直到末速度. 由此可见守恒量在物理中具有十分重要的意义.

势能曲线

在更高级的物理理论中我们往往不讨论力，而是势能. 一维直线运动中，如果受力只是关于位置的函数（如简谐振子），那么这个力（也可以称为**力场**）就叫**保守力**，保守力在高维的情况下有更复杂的定义[905] 先不讨论. 之所以叫做保守力，是因为其对质点做功只与质点的初末位置有关而与运动过程无关. 对于保守力，我们可以计算出每个位置的**势能**，也叫**势能函数**或**势能曲线**. 一维直线运动情况下，坐标为 x，势能函数可以记为 $V(x)$. 某个位置力的大小就是 $V(x)$ 曲线的斜率（曲线与水平方向夹角的 tan 值），方向就是曲线下降的方向. 粒子沿受力方向运动，动能增加，势能减小，总能量/机械能（动能加势能）不变.

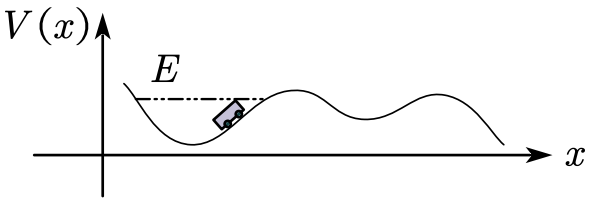


图 1: 轨道小车模型，小车只能在其所在的 $V(x) \leqslant E$ 区间内往返运动，每个位置的速度都是确定的.

作为一个形象但不准确的比喻，想象高低不平的光滑“直”轨道上的小车，我们可以让轨道的高度为 $h(x)$，小车在 x 处时重力势能为 $V(x) = mgh(x)$，这样具有一定总机械能（动能加势能）E 的小车就会在轨道上运动. 机械能是守恒的，所以小车高度越高，势能越大，动能就越小. 动能为零时势能最大，小车达到最高点，满足 $V(x) = E$，然后掉头返回. 如果小车开始时在某 $E \leq V(x)$ 的区间运动，那么它将一直在该区间往返运动，某点的速度为[7]

$$v(x) = \sqrt{\frac{2}{m}[E - V(x)]} \tag{1}$$

注意这个速度是沿轨道方向的速度而不是 x 方向的速度，这就是为什么说这个模型“不准确”. 在真正的直线运动中该式同样成立.

[7] 推导：$\frac{1}{2}mv(x)^2 = E - V(x)$

根据势能曲线，我们可以给一些特殊的点分类．当势能曲线的切线水平，即质点受力为零时，这样的点叫做**平衡点**．但平衡点也分为两种情况，如果平衡点处的势能曲线是凸的，即处于“山峰”，那么它是**非稳定**的平衡点；当平衡点处势能曲线是凹的，即处于“山谷”，那么它是**稳定**的平衡点．请读者找出图 1 中的稳定和非稳定平衡点．

简谐振子

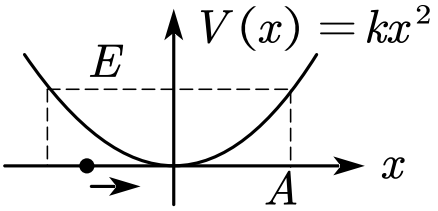


图 2: 简谐振子的势能曲线

作为势能曲线的一个经典例子，我们来看简谐振子[913]．若一个质点通过质量不记的理想弹簧固定在 x 轴的某点，把质点的平衡位置作为原点，那么势能曲线为 $V(x) = kx^2$（图 1 ）．若质点的总能量为 E，那么我们根据曲线可以马上确定质点的运动范围，即简谐运动的振幅 A．当质点经过原点时，势能为零，动能取最大值 E；当质点达到 $x = \pm A$ 时，动能为零，势能取最大值 E．

势阱和势垒

未完成：未完成

在量子力学中，有几个特殊的势能曲线较为常见，我们先看看他们的经典力学版本．

未完成：有限深/无限深势阱/势垒，散射

仍然考虑一维运动，假设光滑水平面上有两面墙，粒子来两面墙之间来回反弹，我们该用什么样的势能函数呢？如果墙是软的，例如把两面墙比作弹簧，得到的势能将如（未完成）

中的公式不会出现力，只会出现势能．常见的势能（三角势垒见下文，方势垒，方势阱，边界处斜率无穷大怎么办？类比碰撞时的冲力．初始动能（总

能量），大于，小于最大势能的时候分别会沿原方向运动，反弹．更理想的情况：无限深势阱（小球在两面墙之间无限反弹），delta 势垒/势阱（受到一个微小扰动，相当于无限窄的方势垒/势阱），对经典粒子运动没有任何影响（物理上不存在，但是作为模型有计算简单的优点）.

第二章

电动力学

电动力学

预备知识 经典力学[4]

本科阶段通常会有几门不同深度关于电磁场场的课程，第一门往往叫做**电磁学（electromagnetism）**，进阶的课程叫做**电动力学（electrodynamics）**[1]. 从课程内容上来说，电磁学就是简单的电动力学. 严格来说，电动力学强调场和电荷的运动规律，如果研究静止的电荷和电磁场，可以将其称为**静电学（electrostatics）**或**静磁学（magnetostatics）**. 我们以下统一使用电动力学.

经典力学研究若干粒子（质点）受若干力后的运动情况，这里的粒子可以是任何有质量的粒子，力也可以是任何力. 电动力学研究的是一堆带电荷的质点（既有质量又有电荷）受一堆电磁力后的运动情况. 所以唯一需要做的是，弄清如何计算这些电磁力，再用经典力学就可以得出粒子的运动.

电磁力和电磁场

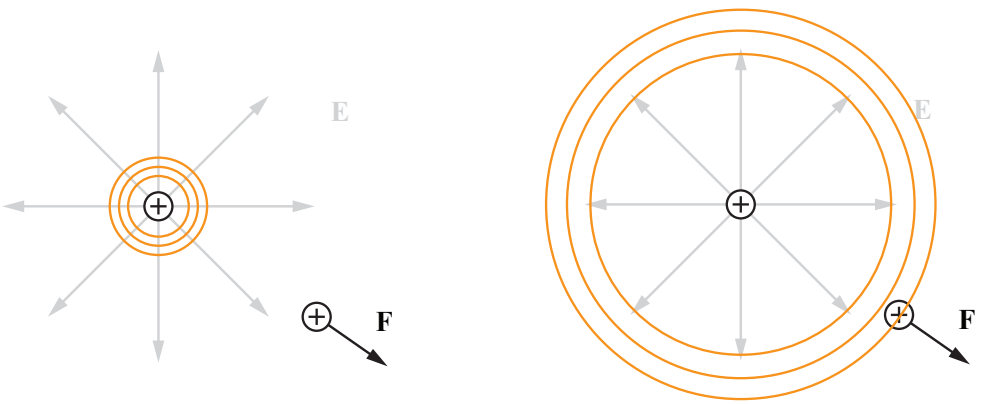


图 1: 静止的电荷突然振动几下又停下来，扰动就会以电磁波（橙色部分）的形式向外传播，到达另一电荷的位置时，才会有力的变化. 图中只画出了一个电荷产生的场.

[1] 电动力学字面上就是“电的动力学”而不是“电动的力学”，动力学英文就是 dynamics，又例如热力学的英文是 thermodynamics.

经典力学中万有引力是超距的，就是说只要两个质点存在，不管他们相隔多远都会立即有作用力（与质量和距离有关），如果它们之间的距离发生改变，那么这个作用力也会立即改变．这是不合理的（现在我们知道任何信息的传递都不可能超过光速，如果一个粒子动一下，另一个粒子马上就能感到力的变化，那就可以超过光速传递信息）．电动力学不同，我们需要先计算电场和磁场（场就像波一样，传播需要时间），再根据粒子所在位置的场计算粒子的受力（其他地方的场如何与粒子受力无关）．例如库仑定理的形式虽然和万有引力一样，但是在电动力学中我们先计算一个粒子在另一个粒子处产生的电场，再计算处于该电场中的粒子所受电场力（库仑力）．当两个粒子都静止时，这么做似乎和直接由距离计算力有没有区别，但如果某时刻其中一个粒子抖动了一下，电场（先不提磁场）就会像扰动的水波一样将这个扰动沿各个方向以一定的速度传播，直到传播到另一个粒子所在的地方，另一个粒子才能从电场中感觉到受力的变化．

磁场虽然产生的方式和对粒子的作用与电场不同，但也会像波一样传播．事实上，电场和磁场并不是独立传播的，上面说的电场扰动的传播时必须同时借助磁场．

电荷

电荷在电动力学中扮演了两个角色．一是电磁场是由电荷产生的．二是只有带电荷的粒子在电磁场中才会受力，其他粒子不会．

麦克斯韦方程组

麦克斯韦方程组是一组写描述场变化规律的数学公式．可以想象成一个懂电动力学的计算机，只要告诉它所有的带电粒子在哪里，以及如何运动（位置关于时间的函数），它就能计算出空间中任何一点的电磁场．和机械波（如水波）不同，电场和磁场在每个位置都既有方向也有大小（有方向和大小的量叫做矢量，这种场叫做矢量场）．

麦克斯韦方程组是电动力学的公设之一．

电磁力

任何一个带电粒子某时刻受到的电磁力（也叫广义洛伦兹力），都可以由它所在的位置处的电磁场（两个矢量）以及它当时运动速度（也是矢量）通过公式计算出来．这是电动力学的另一个公设．

可见，在经典力学的基础上加上麦克斯韦方程和广义洛伦兹力的公式后，如果我们知道一堆带电粒子在某个时刻的运动状态（位置和速度），我们就可以知道接下来每个粒子如何运动．

静电的基本规律和性质

两种电荷

[2]高中时候我们就已经了解到摩擦带电，并且人们将用绸子摩擦过的玻璃棒所带电荷称为正电荷，将用毛皮摩擦过的硬橡胶棒所带电荷叫做负电荷．（注意这种命名一开始是任意的，但一直沿用至今）

这两种电荷相互接触能够互相中和，也就是正负电荷完全抵消．

性质同性相斥，异性相吸．（也就是同种电荷相互排斥，异种电荷相互吸引）

电荷守恒定律

电荷既不能被创造，也不能被消灭，它们只能从一个物体转移到另一个物体，或者从物体的一部分转移到另一部分．换句话说，就是任何物理过程中电荷的代数和是守恒的．（这是物理学中的普遍的基本定律）

$$\sum Q_i = C \tag{1}$$

电荷的量子化（charge quantization）

通过密里根油滴实验，密里根发现油滴所带的总电荷是一个数值的整数倍．这个数值就是电荷量 e，也就是量子化的．在这里稍微提一下其实验内容，

[2]参考：新概念物理教程电磁学（赵凯华，陈熙谋）

对其实验感兴趣的读者可以去查找相关资料作为补充阅读.

具体实验图如下：

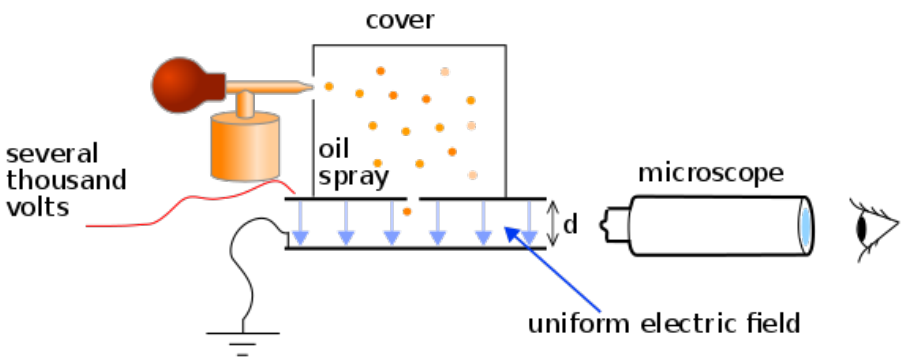


图 1: Simplified scheme of Millikan's oil drop experiment(来自维基百科)

这个实验的原理是根据油滴的受力平衡原理（粗略的理解可以认为：油滴的重力 mg 和电场力 qE 平衡，但其实具体实验过程中还需考虑空气的粘滞性对油滴产生的阻力等因素），计算出油滴的带电量，从而得出电荷量.）

点电荷

在研究带电体的时候，当带电体的形状大小，和电荷分布等因素可以忽略不考虑的情况下，我们为了研究的方便，经常把这样的物体抽象的看成一个几何点，也就是点电荷.

荷质比的测定

预备知识 静电的基本规律和性质[17]

荷质比

[3]原理：利用电子（或其他带电粒子）在磁场中偏转性的特点，测得粒子电荷与质量之比（即荷质比）说明：荷质比是带电微观粒子的基本参量之一.

典型的测量荷质比的方式有两种

[3]参考文献：电磁学（赵凯华，陈熙谋）.

汤姆孙测量电子荷质比的方法

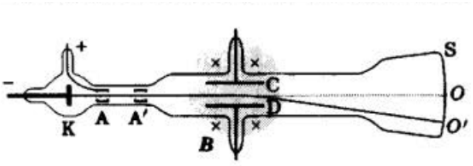


图 1: 汤姆孙法测荷质比

玻璃管内抽成真空，在阳极 A 与阴极 K 之间维持数千伏特的电压，靠管内残存的气体的离子在阴极引起的二次发射产生电子流．阳极 A 和第二个金属屏 A' 中央各有一个小孔，K、A 之间被加速了电子流，只有很窄一束能够通过两孔．玻璃管的中部 C、D 为电容板的两极板，在其间可产生一竖直方向的电场．图中阴影部分，是由管外的电磁铁产生一方向垂直纸面的磁场．适当的调节电场和磁场的强度，可使它们作用在电子上的力达到平衡，即：

$$eE = evB \tag{1}$$

然后，将电场切断，电子束在磁场区域内将沿圆弧运动，此圆弧半径可得：

$$R = \frac{mv}{eB} \tag{2}$$

因此，电子的荷质比为：

$$\frac{e}{m} = \frac{v}{RB} = \frac{E}{RB^2} \tag{3}$$

通过测得 R 之后，我们就能求出荷质比了．

磁聚焦法

抽真空的玻璃管中装有热阴极 K 和有小孔的阳极 A．在 A，K 之间加电压 ΔU 时，由阳极的小孔射出的电子动能为：

$$\frac{1}{2}mv^2 = e\Delta U \tag{4}$$

从得到其速率为：

$$v = \sqrt{\frac{2e\Delta U}{m}} \tag{5}$$

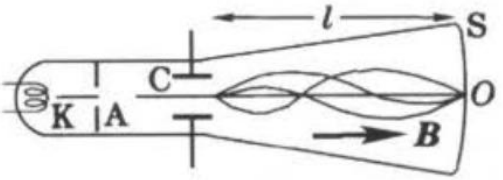


图 2: 磁聚焦法测荷质比

在电容器 C 上加一个不大的横向交变电场，使不同时刻通过这里的电子发生不同程度的偏转．在电容器 C 和荧光屏 S 之间加一均匀的纵向磁场，电子从 C 出来后将沿螺旋线运动，到达距离 $h = \frac{2\pi mv}{eB}$ 的地方聚焦．适当的调节磁感应强度 B 大小，使得电子流的焦点落在荧光屏 S 上．其中 h 是 C 到 S 间的距离．将上述方程中的 v 消去，我们有：

$$\frac{e}{m} = \frac{8\pi^2 \Delta U}{h^2 B^2} \tag{6}$$

（上式中方程右边的各量都能够测出来，因此我们能够确定 e/m）

右手定则

预备知识 几何矢量[370]

物理中常用右手定则来判断方向，右手定则分为以下两种．

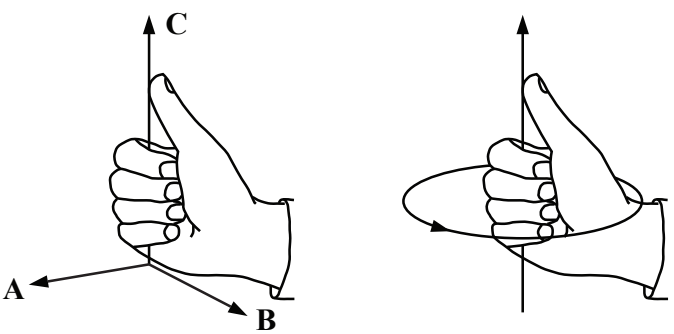


图 1: 两种右手定则

第一种右手定则

假设有两个不共线的矢量，第一个为 **A**，第二个为 **B**，它们可以定义一个平面．现在我们想根据 **A** 和 **B** 确定该平面的法向量 **C**，但由于平面的法向量有两个，我们如何区分它们呢？我们可以通过图 1（左）所示的右手定则来确定其中一个法向量：首先将右手的四指指向 **A**，再将四指弯向 **B**，这时伸出拇指，拇指的方向即为右手定则定义的方向．

我们通常使用的空间直角坐标系被称为**右手系**，是因为我们可以通过右手定则从 x 轴和 y 轴的方向定义 z 轴的方向（图 2 右）．同理，我们也可以定义所谓的**左手系**．

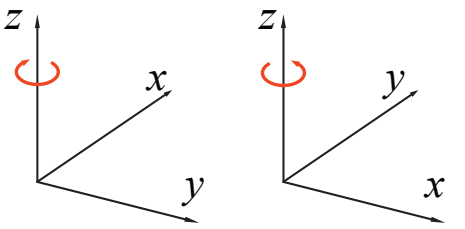


图 2: 左手系（左）与右手系（右）

第二种右手定则

假设空间中有一圆环，且圆环上有一正方向，我们可以用右手定则指定圆环所在平面的一个法向量．如图 1（右），用右手握住该圆环，手指与圆环平行，且指尖指向圆环的正方向，伸出拇指，则拇指所指的方向就是右手定则判断的方向．

第三章

量子力学

原子的观念

[1]在《生活大爆炸》的第 3 季第 10 集中，莱纳德邀请伯纳黛特参观他正在做的验证 AB 效应（全称是阿哈朗诺夫-玻姆效应）的实验．佩妮不想落于人后，也想在聚餐的时候谈论莱纳德的实验，于是央求谢尔顿教她物理．

但物理是没法速成的，要讲就得从古希腊开始．

“假想在一个炎热的夏季的夜晚，你刚刚在阿戈拉（集市）买完东西……”

“但，这与莱纳德的研究有什么关系呢？”

谢尔顿的回答是：“科学是个 2600 年的旅程，从古希腊开始，经由牛顿，到玻尔（旧量子论），然后薛定谔（波动力学），到哥本哈根学派”，最后我们才能谈论莱纳德的实验．

我们也采用类似的路径，首先让我们回到 2600 年前雅典的阿戈拉，一个炎热的夏季夜晚，那里正有人在发表关于原子的理论．

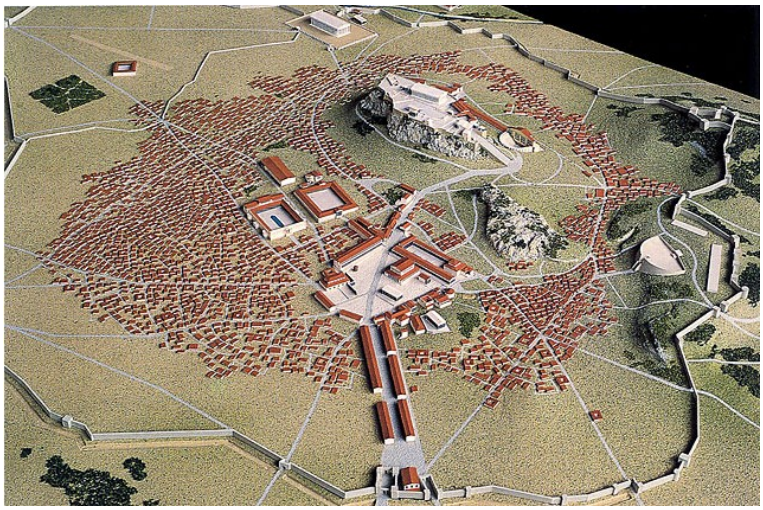


图 1: 公元前 2 世纪雅典鸟瞰图

柏拉图的四元素说

关于物质的理论，自古就有，比如古希腊的泰勒斯曾说“万物皆水”，后来又有人说万物是四种元素“水、气、土、火”构成．这就是所谓四元素说．

[1]本文根据 CC-BY 协议转载自季燕江的《量子序曲》，进行了重新排版和少量修改

四元素说本来是古希腊人的常识（common sense），但柏拉图给这种关于物质的学说理论化，系统化了．这些内容被记载在柏拉图（427 BC —347 BC）的宇宙论《蒂迈欧篇》中．

这里我们可以给出一个论证的概要：

- 万物或者是可见的，或者是可以触摸的．可见是因为光，可以触摸是因为坚硬．光是火的性质，坚硬是土的性质，这样我们就有了“火和土”两种元素．
- 万物是三维的，我们需要像“胶水”一样的元素把“火和土”按比例混合起来成为三维的物体．这里柏拉图是通过构造如下数列来论证的：

$$1 = 1^2 = 1^3, 2, 4 = 2^2, 8 = 2^3, 16 = 4^2, 32, 64 = 4^3, ... \tag{1}$$

  这里 $1 = 1^3$，$8 = 2^3$，$64 = 4^3$，...，叫做立方数．每两个立方数之间正好是两个数，比如 1 和 8 之间是 2 和 4；而 8 和 64 之间是 16 和 32．柏拉图因此论证说需要两种元素在“火和土”之间调和使生成万物，这两种元素就是水和气．
- 不论是可见，还是可以触摸，都可以归结为形状，而形状可以归结为多边形的拼合，多边形则可以归结为三角形的拼合．三角形是研究形的基础，或说三角形是研究几何学的基础．
- 柏拉图提出了两种基本的直角三角形：（1）等腰直角三角形，记做 T_{45}；（2）一个锐角为 30^o 的直角三角形，记做 T_{30}．
- 利用两种基本的直角三角形可以拼成四种正多面体，即正四面体、正八面体、正六面体和正二十面体．

1. 柏拉图认为立方体，即正六面体最稳定，因此把土的形定为正六面体．
2. 正四面体有最锐利的尖角，因此是最活跃的．火的形就是正四面体．
3. 剩下的还有气和水．因为气比水活跃，因此气的形就是正八面体．
4. 还剩下最后一个，正二十面体，是水的形．

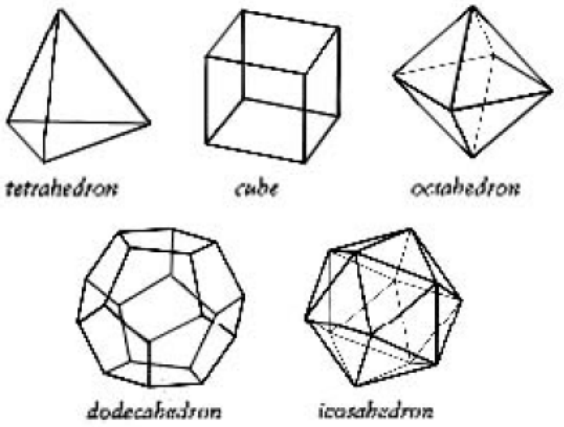


图 2: 五种正多面体，也称柏拉图多面体．这里面的正十二面体没有出现在对元素的构造中，这是柏拉图理论中欠缺美感的地方．

表 1: 柏拉图的“元素说”

元素	正 n 面体	等腰直角三角形	30^o 直角三角形
土	正六面体	12	0
火	正四面体	0	8
气	正八面体	0	16
水	正二十面体	0	40

这些形都是很小的，用今天的话说就是“水气土火”的原子．并且这些原子还可以相互转化．

- 只有土是由等腰直角三角形围成的，因此土最稳定，它会被火溶解，可以被气或水分解，但不会转变成其它元素．
- “火、气、水”都是由 30^o 的直角三角形围成的，因此可相互转换．比如，我们可以写出如下的反应方程式：

$$1Water \to 2Gas + 1Fire \tag{2}$$

反应前有 40 个 T_{30}，而反应后有 $2 \times 16 + 8 = 40$ 个 T_{30}．我们可以把上式与电解水的反应方程式比较：

$$2H_2O \to 2H_2 \uparrow + O_2 \uparrow \tag{3}$$

从思维的角度这是属于同一母型（Prototype）的．如果我们说古代思想会对近代科学有启迪作用，或我们说人总是在几种思维母型里打转转就毫不奇怪了．实际上量子力学的创建者，比如海森堡自幼就熟读《蒂迈欧篇》，如果说这些这些直观的图像会对他有潜移默化的影响可能并不夸张．

以上是柏拉图元素论的概要，当然这些都是非常粗糙的论证，而且经不起实验的定量检测，但元素之间可以互相转化，这个想法还是启迪了后来的炼金术，并最终导致了近代化学的出现．

伊壁鸠鲁的原子论

原子在古希腊语中不可再分的意思，这个不可再分固可以做物理上的拆分理解，亦可作逻辑上的不可再分（析）理解．比如刚才介绍的柏拉图关于“水气火土”四元素的理论就是逻辑上无法再分析的范例．

柏拉图的理论在古代是显学，2000 多年来一直有稳定的传承，但它在古代并非没有对手，比如从德谟克利特、留基伯到伊壁鸠鲁、卢克莱修的原子论．但说实话这两种理论区别并不大，他们争执不休，以至于柏拉图都准备带弟子去焚烧德谟克利特的著作，并非是他们对自然真有什么本质的看法不同，毕竟都是一个时代对自然的看法．

关键的分歧是他们对伦理学和政治学的观点不一样，而古人的学问是个整体，伦理学的基础是哲学和科学，要驳倒对方，争取听众，釜底抽薪，攻击对方的科学观点是一个潜在的选择．

古代原子论者是今天所说的唯物论者，他们不相信灵魂，把生命看做是一堆原子具有功能性、活性或协调一致运动的集合．就好像是一把调谐良好的里拉琴，只要弦不断就能奏出美妙的音乐，对应于人的生命状态，而弦断则对应人亡，灵魂在这里是无所栖身的．没有灵魂，自然就没有死后世界，所有传统的道德说理就被架空了，Polis 将处于危机之中．柏拉图派对唯物论者（或自然哲学家）的反感和敌对就在这里．

当然这并非我们现在的主题，我也不再继续展开讨论，而仅仅强调一点，即当我们讲到古代原子论的时候，我们应把柏拉图派的很多观点、理论也置于古代原子论的范畴内，而不仅仅是介绍自然哲学家的原子论．实际上很多柏拉图派的理念，比如天球的理念和近代原子模型是很接近的．对玻尔、海森堡这些哲学倾向很强的物理学家，很难想象他们没有阅读过《理想国》和《蒂迈欧

篇》，而如果读过的话，他们一定对柏拉图的“数学-几何学”的宇宙模型印象深刻.

下面我将扼要地介绍伊壁鸠鲁的原子论，伊壁鸠鲁（341 BC —270 BC）的《致希罗多德书信》是现存最早的关于原子论的记录，再早的比如德谟克利特和留基伯的就都已经失传了.

- 在视觉上存在最小的点，再小我们就看不见了，这个视觉上的最小的点是有大小的. 由此类比：物质也是由最小的不可再分的最小单元构成的，再小是不存在的. 就这一点而言，伊壁鸠鲁的方案和柏拉图的方案是不同的，柏拉图的基础三角形没有明确地说存在最小、最基础的三角形. 今天的人大多会认为物质存在最小的不可再分的单元是很抽象的，或很难想象，但在古人那里也许未必. 比如在中国古代是通过计量“肥而美”的“黑小米”来建立度量衡制度的，而在古代西方也有类似的比喻，比如卢克莱修在《物性论》中把水想象为大量罂粟籽的集合，比如阿基米德在《数沙者》中假想整个宇宙都充满了沙子，然后计算了整个宇宙中有多少粒沙子. 古人的数学观念是由“1，2，3，……”出发开始建立的，对他们来说想象一个由原子构成的世界其实是更简单的. 这个原子就好比构成宇宙、万物的基础砖石，我们先寻求生命的结构和功能，进而问生命的可能的意义是什么？

- 原子很小，因为我们谁也没见过原子. 但我们没法假设原子的大小是绝对的 0，因为如果那样的话，很多原子就无法构成万物了，而万物在我们的眼里是有体积的.

- 原子如果有大小，那它就有部分，有形状. 所以这里我们说的“不可再分”其实是物理意义下的，因为从几何学的角度，原子既然有部分，有形状，那它就可以在想象中继续分解. 这里区分了物理意义下的“分”，和数学意义下的“分”. 前者对应今天的原子物理和量子力学，而后者对应的则是数学分析.

- 原子既然有大小，有形状，它们之间就可以发生机械的相互作用，甚至结合成几个原子构成的具有特定功能的结构单元——分子.

  没错，这都是伊壁鸠鲁天才的猜想. 从字面上说已经很类似今天化学、甚至生物化学对结构和功能的解释.

当然，这里我们无需把这一天才的猜想神秘化，古代世界是个机械的世界，各种机械装置，从三列战舰、到攻城者德米特里乌斯的巨型攻城器械，古希腊人生活在这样一个机械无所不在的航海和贸易的时代，把原子想象成简单机械，把分子想象成具有特定功能的机械组合，进而把生命想象成极复杂的巨型机械都不是不可能的.

小结一下：柏拉图的四元素说和伊壁鸠鲁的原子论都是基于“形”的学说，但伊壁鸠鲁不能把他的理论数学化同时他的兴趣也不在于研究诸如——既然原子是有大小的，那么原子的大小到底是多少呢？——这类问题.

因此虽然古代原子论，无论是柏拉图的学说还是伊壁鸠鲁等自然哲学家的学说对人们想象原子有启发，但都不能说古希腊人已经接近了一个近代的原子论. 近代原子论正如海森堡所说，是建立在一系列实验事实基础上的，当然还有概念体系和计算方法.

以下摘录海森堡在《物理学和哲学》第四章的一段讨论作为本小节的结束.

“在将原子物理学中的现代观点和希腊哲学作了类比之后，我们必须补充一个警告，即对这种类比不应有所误解. 乍看起来，似乎希腊哲学家由于某种天才直觉而得到了与我们现代相同或很相似的结论，而我们的结论却是经过几个世纪的实验和数学方面的艰苦劳动才得到的. 对我们的类比的这种解释无论如何是一种完全的误解.

在现代科学和希腊哲学之间有着巨大的差别，那就是现代科学的经验主义态度. 自从伽利略和牛顿的时代以来，现代科学就已奠基于对自然的详细研究之上，奠基于这样一个假设之上，这就是：只有已被实验证实的或至少能被实验证实的陈述才是容许作出的. 为了研究细节并在连续不断的变化中找到经久不变的定律，人们可用一个实验在自然中隔离出若干事件，这种观念希腊哲学家是没有想到过的.

由此可见，现代科学在一开始就立足于一个比古代哲学更谨慎同时也更巩固得多的基础之上. 因此，现代物理学的陈述在某种意义上比希腊哲学更严肃得多.”

三角学

角度

首先由圆出发，假设圆的半径是 R，角 θ 张成的圆弧是 l.

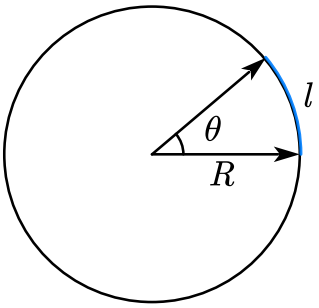


图 3: 圆

$$l = R\theta \tag{4}$$

这其实构成了对角度 θ 的定义，

$$\theta = \frac{l}{R} \tag{5}$$

假想这个圆以半径 R 完整地转一圈，我们把这个角度定义为 2π，在平面上的张角只能取 $0 \to 2\pi$，π 是个无理数，它近似地等于：

$$\pi = 3.14159265... \tag{6}$$

这样定义的角度叫“弧度制”. 但通常我们会把一个完整的圆周分成完全相等的 360 份，每一份称为一度，一度再等分为 60 份，每份叫一分，一分再等分为 60 份，每份叫 1 秒，这就是所谓“度分秒”制.

“度分秒制”的好处是，60 很容易被 2、3、4、5、6 等分，这对古人精确地计算星星在天空中的方位是很重要的. 比如：17 度 25 分 30 秒可记为：

$$17^o 25' 30'' \tag{7}$$

我们可以把一个圆周对应的张角分成完全相等的四份，每一份都是个直角，一个直角在弧度制下是 $\pi/2$，而在度分秒制下则是 90^o.

正弦和余弦函数

现在来考虑一个直角三角形，即一个角是直角的三角形：

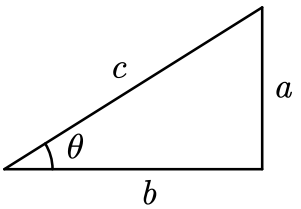


图 4: 直角三角形

假设其中一个锐角是 θ，θ 所对的直角边是 a，另一条直角边是 b，斜边是 c，我们有毕达哥拉斯定理：“斜边的平方等于两直角边的平方和”.

$$c^2 = a^2 + b^2 \tag{8}$$

我们可定义正弦和余弦函数如下：

$$\sin\theta = \frac{a}{c} \qquad \cos\theta = \frac{b}{c} \tag{9}$$

即正弦函数被定义为对边 a 与斜边 c 之比，而余弦函数被定义为邻边 b 和斜边 c 之比.

类似地我们还可定义正切函数（$\tan\theta$）和余切函数（$\cot\theta$）：

$$\begin{aligned} \tan\theta &= \frac{a}{b} \\ \cot\theta &= \frac{b}{a} \end{aligned} \tag{10}$$

现在考虑一个单位圆，即半径为 1 的圆，假设圆上一点逆时针地围绕圆心转动起来．这一点向 x 轴的投影是 x，向 y 轴的投影是 y.

据说第一个三角函数表（Trigonometric table）是古希腊天文学家伊巴古（又译作喜帕恰斯，190-125 BC）编纂的.

根据上表，我们可以计算：

$$\begin{aligned} \sin(\theta + \pi/2) &= \sin\theta\cos\pi/2 + \cos\theta\sin\pi/2 = \cos\theta \\ \cos(\theta + \pi/2) &= \cos\theta\cos\pi/2 - \sin\theta\sin\pi/2 = -\sin\theta \end{aligned} \tag{11}$$

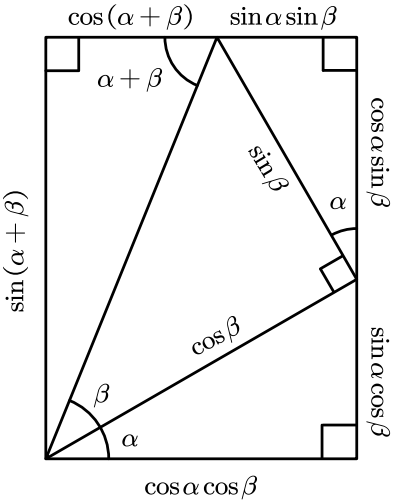


图 5: 两角和公式

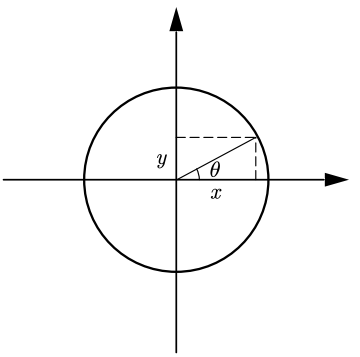


图 6: 投影

于是得到正弦/余弦函数的性质：

$$\begin{aligned}\sin(\theta+\pi/2)&=\cos\theta\\ \cos(\theta+\pi/2)&=-\sin\theta\end{aligned}\tag{12}$$

当 θ 从 $0\to 2\pi$ 变化时，x 的取值将从 $1\to 0\to -1\to 0\to 1$，然后重复，而 y 的取值将从 $0\to 1\to 0\to -1\to 0$，然后重复.

这种周期性的运动很类似月相的变化，望是满月，朔是新月，由于月球围绕地球的运动，我们在地球上观察将看到月亮由月初的新月，到月中（十五）的满月，然后到月末又回到新月，然后重复.

表 2: 常见三角函数的取值

θ	$\sin\theta$	$\cos\theta$
0	0	1
$\pi/4$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$
$\pi/2$	1	0
π	0	-1

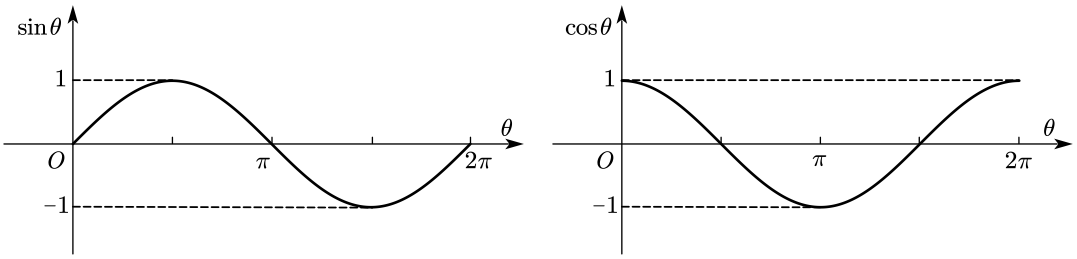


图 7: $\sin\theta$ 和 $\cos\theta$

在此意义下我们称 θ 为相位（phase）.

正弦和余弦函数在物理学中的用处很多，比如它是弹簧振子的解，是波动方程的解等.

振动

对振动而言，相位是：

$$\theta = \omega t \tag{13}$$

这里 ω 叫角频率，当 $\omega t = 2\pi$ 时，振动正好重复一个周期 T，

$$T = \frac{2\pi}{\omega} \tag{14}$$

描述振动的方程是：

$$A\cos\omega t \tag{15}$$

A 叫做振幅，振动将在 $A \to 0 \to -A \to 0 \to A$ 重复发生.

表 3: $A\cos\omega t$ 描述的运动

时间	相位	位置
0	0	A
T/4	$\pi/2$	0
T/2	π	-A
3T/4	$3\pi/2$	0
T	2π	A

但有时候当 $t=0$ 时，我们就已经有了一个非零的相位 θ_0 了，这时描述振动的方程将变成：

$$A\cos(\omega t+\theta_0) \tag{16}$$

这里 θ_0 叫做初相位（initial phase），我们管用正弦/余弦函数描述的振动叫简谐振动.

波动

振动是在时间变量上的重复. 波动是在时间和空间上的重复. 我们所处的世界是一维的时间再加上三维的空间. 为了简单，我们先考虑空间也是一维的情况.

假设我们在体育场里造人浪（Mexican wave），开始的时候也许只有一个人重复地做“站起-欢呼-坐下”，再“站起-欢呼”再“坐下”，然后重复的动作，他旁边紧挨着的人注意到他的动作，也会跟着模仿他的动作，但因为距离的存在，旁边的人要反应过来跟着他一起做需要过一段时间，即有个时间的延迟.



图 8: 墨西哥人浪起源于 1986 年的墨西哥世界杯. 在那届世界杯上，热情的球迷自发地运用一种交替起立欢呼的方式自娱自乐，为球队加油，从远处看就像一阵阵的海浪.

假设人与人之间的间距是 Δx，时间的延迟是 Δt，波动在空间上传播的速度就是：

$$v = \frac{\Delta x}{\Delta t} \tag{17}$$

假设第一个人正好完成了一个振动的周期 T：“坐着、站起、欢呼、坐下、坐着”

这时如果他抬头看的话，他会发现距离他 λ 远处某个人正处于和他一样“坐着”的状态，而且他马上就要站起来了，因为他紧邻的人已经在做“站起”的动作了.

可以设想，这两个相距 λ 的人将做完全同步的振动，在我们外人看来，他们一起“坐着”，又一起“站起”，一起“欢呼”，……

如果用振动的语言说的话，它们的相位是相同的. 但这么说让我们不舒服，因为毕竟后一个人的运动要晚于第一个人，我们说，“空间相距 λ 远的两个点，它们运动的相位正好差 2π”.

这样说就舒服了.

现在我们假设体育场里已经造起了人浪，我们应该如何描述它呢？这是一个以空间位置 x 和时间 t 为变量的函数.

假设 t 固定，比如令 $t = 0$ 我们得到一个随 x 起伏变化的函数，这就相当于是照相机照的快照，在某个时间运动冻结了，但仍会在空间 x 上有个分布，假设这个函数是：

$$\psi(x, t = 0) \tag{18}$$

这个分布其实就是个波形，它围绕 $x = 0$ 附近展开. 单独看这个随 x 起伏的波形，它也是周期性的，是空间调制意义下的周期性，每增加波长 λ，ψ 就重复.

假设 ψ 可以表示为一个正弦/余弦函数，这意味着：x 每增加 λ，相位的变化是 2π. 即：

$$\psi(x, t = 0) = A\cos\frac{2\pi x}{\lambda} \tag{19}$$

现在假设时间是 t，并且 $t > 0$，波形会向前传播，时间 t 波形向前移动了 vt，这意味着波函数是在 $x = vt$ 附近展开的.

$$\psi(x, t) = A\cos\frac{2\pi}{\lambda}(x - vt) \tag{20}$$

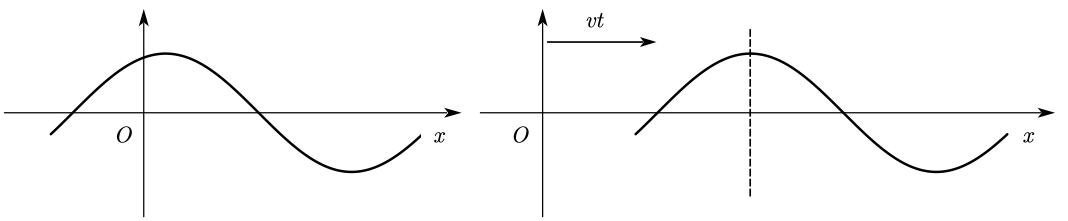


图 9: 经过时间 t，波形会向前移动 vt，或者说是函数沿 x 方向发生了平移，平移前函数是 $\psi(x)$，而平移后函数是 $\psi(x-vt)$.

假设波形移动 λ 需要花费时间 T，波形移动的速度 v 是：

$$v = \frac{\lambda}{T} \tag{21}$$

波函数 $\psi(x,t)$ 变成：

$$\psi(x,t) = A\cos 2\pi\left(\frac{x}{\lambda} - \frac{t}{T}\right) \tag{22}$$

定义波矢 k，和角频率 ω 分别为：

$$\begin{aligned} k &= \frac{2\pi}{\lambda} \\ \omega &= \frac{2\pi}{T} \end{aligned} \tag{23}$$

波函数取常见的形式：

$$\psi(x,t) = A\cos\left(kx - \omega t\right) \tag{24}$$

现在相位是：

$$kx - \omega t \tag{25}$$

固定时间，我们会得到波动随空间的分布，x 越大，相位也越大. 固定位置，我们会得到波动就在这个位置随时间的振动.

习题 1

假设人眼瞳孔大小是 2 mm，可见光波长是 550 nm，利用公式：

$$\theta = 1.22 \cdot \frac{\lambda}{D} \tag{26}$$

估算人眼的角分辨本领，并把结果换算为“度分秒”为单位.

据说第谷（1546-1601）对星星观察的准确度能够达到二弧分，已经逼近肉眼的理论极限，当然夜晚观星瞳孔会放大，这有利于提高他的精度.

从天球的音乐到玻尔模型

[2]形或古希腊人所说的“idea”，有多种含义，比如：形状，这是和视觉有关的；比如风格、分类，这可以是和视觉有关的，也可以无关，比如音乐也可以有风格，这就是和声音有关的了.

和视觉有关的“形”是直观的，我们无须论证，纠结于如何用语言表达，仅凭图形——或者是静态的，或者是想象中动态的——直接给出结果. 对形的研究会导向几何学，几何本身是视觉的，而视觉是偏好静的，偏好不动的，但一加“学”，几何“学”或“学”几何就动起来了.

我们如何学呢？或者演示，用圆规和直尺，或者像毕达哥拉斯那样拿根木棍面对沙土，世界是一步一步地被展现出来的，一笔一划本身就是个动态的过程. 我们努力说：“首先如何，其次如何，然后，又然后……”

所谓动态就是次序，我们首先只关注首先要解决的，其次，带着对刚刚过去的对首先的记忆，探讨紧接着要解决的问题，我们的思想没法分叉. 人在专注的状态下，视觉也需要一个焦点. 当我们的视觉遭遇挑战，看不清某物的时候，我们凝眼观瞧，把视线使劲聚焦于某物，凝眼就是凝神，不受诱惑地专注于某物，看清楚一点再继续看下一点.

这个结构很像自然数：“0，1，2，3，……”，一步一步地展示给你看，比如“我是如何用直尺和圆规作图的”，这种线性展开的结构就是时间，“学”的过程，对学的人是学，在 Challenge，对展示的人来说是在“证”，在说服，这个过程是世界次第展开的过程，是叙事、是 Chronicle.

形与声

“学”依赖语言，语言是一种声音现象.

据说人能够发出一个八度再加一个四度的声音.

[2]本文根据 CC-BY 协议转载自季燕江的《量子序曲》，进行了重新排版和少量修改

古代世界，天和地很近，音乐和人也很近．孔子闻韶乐“三月不知肉味”，这种沉浸在声音里的境界和我们今天听流行音乐，把音乐当做一种背景噪音，是完全不同的两种声音技术．今天我们听音乐往往是为了抑制我们心中的背景噪音．

古代的音乐都很简单．简单到好比就是敲击单音音叉发出的声音，单音音叉是校音用的，它在古代世界的对应物是中国的黄钟律管或希腊的单弦琴（Monochord）．它们发出很纯的音，基本上就是一个频率．孔子一生关心礼，礼与乐相联，乐就是音及音的混杂与排列．

我们用音高，频率，响度，音色等来描述声音．音高就是频率，是描述“音”诸参数中最重要的一个．人天生就是一个感知音高的灵敏动物，高音激越，使人振奋，低音呜咽，让人伤感．简单的音乐庄重使人入静，而复杂多变的音乐也如一场“视觉的盛宴”，它使我们好奇和沉迷．

听觉和视觉一样，是感觉，同时也是思维，我们的眼睛和耳朵接受信息，同时也处理、歪曲信息以为我们所用．古代的政治传统，古代的教育家都注重音乐教育，这其中最重要的就是对音乐体系的保留和传承．

比如唱歌的时候要先定调，调可以定低点，显得庄重，也可以定高点，显得轻快．定好调后，一系列的声音次第展开，它们的相对音高保持一个固定的结构，比如：

“低，低低，高，高高，低，中中，……”

在给定乐谱的前提下．基准音高的选取，或所谓定调是任意的．我们可以定高点，无非大家唱不上去而已．但因为有人唱不上去，这个定调就也不是完全主观任意的了．

古代政治秩序大多由推崇勇猛进取精神的战士集团建立，对战士共同体而言，最重要的是要保持这种勇猛进取的精神，能够保持这种精神的音乐会与特定音高有关，这是人群的共同经验．比如柏拉图在《理想国》中就说，要摒弃悲伤和软绵绵的吕底亚调和伊奥尼亚调，而推崇多利亚调和佛里吉亚调．

保持这种对声音的共同经验在古代政治传统中是非常重要的，其中之一就是确定音调，或基准音的频率，然后在此基础上给出其他音的定义，其他音是相对于基准音而言的，可以更高，也可以更低，构成一个阶梯状的结构．

原子的“idea”是无所不在的，这里由人的听觉经验，我们再次得到了原子的概念，即存在着“音高”的原子，进一步细分不同音高的原子是没有必要的．

保存音乐制度最简单的方法就是造一套标准的乐器，然后后人反复向这些标准的乐器学习，第一套自然是由城邦的缔造者“铸造”的.

考虑到弦乐器与弦绷紧的程度有关，受湿度、温度影响较大，青铜器制造的发音器会是理想的选择，这是为什么“钟”会成为“国家”符号的原因，塔可夫斯基电影《安德烈 • 卢布廖夫》再现的是俄罗斯帝国创旦的精神基础，在影片的结尾就出现了工匠之子铸钟的奇迹.



图 1: 工匠之子铸钟

钟是要发音的，音高是有标准的，“音高”高一些，低一些，很微妙，但人的耳朵，或某些人的耳朵天生就是辨别音高的灵敏仪器．只有能发出特定音高的钟才是可以被接受，一只发音不准的钟在敲响的时候不嘹亮，不能激发人民激越的精神，这对城邦是不利的.

这里有个似是而非但很有趣的讨论，人有时间感，但人的时间感是非常内在的，几乎不存在什么可以相互交流的基础．这是妨碍人产生运动观念，并研究运动的重要原因．音高即频率，频率就是时间的倒数，人没法准确标记时间的流逝，但人却是辨别频率（时间倒数）的精密仪器．同时我们的发音器官，还能娴熟地对不同音高的声音进行模仿，这是我们具有语言和音乐能力的生物学基础.

类似地，我们还可以讨论视觉，讨论视觉对位置和速度的分辨．人天生就能在相当精确的意义下辨别位置，但我们对速度的判断就要差许多．我们说 A 比 B 快，其实是通过位置下的判断，即 AB 同时出发，但 A 先撞线，所以 A 更快．这是亚里士多德无法得到“正确”的落体规律的原因，他受人本身的局限，速度是很难直接被看的.

古代实验技术还没有充分发展起来，而实验技术的充分发展与资本主义的生产方式兴盛有关，近代自然科学与资本主义生产方式同步爆发并非巧合．回

顾二者，科学史和资本主义发展史，两者讲的是同一个故事，只是叙事的角度生了变化.

由“造钟”故事，我们得到一个新洞见，即：“音与形有关”. 对钟来说这是大大地简单化了，因为材质也很重要，但形状确实决定了钟振动的频率. 这意味着：“听音可以定形，定形可以定音”.

形既是形状也是模型，还是形式. 在毕达哥拉斯和柏拉图的传统里，形是与数紧密相连的. 比如钟的形由何而定呢？长、宽、高、是数字，钟的厚度也是数字，但这一堆数字的集合又有什么意义呢？

当我滔滔不绝地罗列一堆数字的时候，这是没有意义的. 我们需要给出数字和数字之间的关系，才有意义. 而且最好是只给出一个关系（或最少关系），就能让所有的数字各就各位. 找到这样的规律自然是对思维的奖励，是可以向众人夸耀的；同时这也是技术，有了技术我们就能铸钟，小孩的父亲是会铸钟的，但他把技术带到坟墓里去了.

《安德烈 • 卢布廖夫》中的小孩是幸运的，他必须试试，他也只能试试. 在拜占庭衰败之后，东正教来到了俄罗斯与当地的土豪、愚民混合，文明在绝望中重新开始，这就是俄罗斯的宿命. 卢布廖夫受不会铸造但却造出钟的小孩的激励，重新拿起画笔开始画注定会塑造俄罗斯民族精神的那些很平、很抽象圣像画.



图 2: 卢布廖夫的《三圣像》

画是形（idea），音是声（logos）. 形和声都能塑造性格，前提是我们生活

在某种生活中，或我们生活在某种历史中.

数字与和谐

“几何学”（Geometry）是对形的规定，而“和声学”（Harmonics）是对音的规定. 所谓规定就是数字之间的联系，最简单的数字和数字间的联系是“相等”，稍微高级点的是比例，是合乎比例.

比如人脸，人脸上五官的位置和尺寸是需要合乎比例的，这种合乎比例是我们天生可以判断的，但很难用数字说清楚. 近一二十年随着计算机对数据处理能力的提高和神经科学的进步，在这方面有了很多具体技术的进展.

比例或合乎比例会产生美，这是某种审美观念下的模式识别. 比如古代东夷部族以扁头为美，甚至不惜把小孩的头骨弄扁以合乎比例. 这个习俗在今天还有遗存，不少地方有端正小孩睡姿以把头睡扁的说法.

在音乐中我们很容易发现音高与数字的关系. 这是毕达哥拉斯的贡献. 音乐的历史一定很古老. 在毕达哥拉斯之前人类就有音乐了，不但有音乐还有规定音高的一套体系，即有一套术语来说清楚“不同音高”的音之间的关系.

比如当我发出一个音后，让你发出一个高四度的音，你就能发出这样一个音，并得到我的认同. 这套语言游戏能够玩儿的起来. 这些当然都是基于感官经验讲的，本来和数字没啥关系. 传说毕达哥拉斯在路过铁匠铺时，受到叮叮当当声音的启发，回去研究各种乐器的音高，比如弦乐.

所谓弦乐器就是一根绷紧的弦，两端固定，中间可以快速振动起来，扰动空气发出声音，弦乐的频率自然就是琴弦发出的声音. 这是典型的机械振动的问题，弦上会有波动，但因琴弦两端是固定的，所以波传播不出去，它只能被限制在琴弦上振动，并整体具有一个轮廓，琴弦就在这个轮廓内振动，这种振动叫驻波.

琴弦上的振动是波动，当一列波从左向右传播时碰到弦的端点会反射回来，驻波就是两列相向传播的波的叠加：

$$\frac{A}{2}\left(\cos(kx-\omega t)+\cos(kx+\omega t)\right)=A\cos kx\cos\omega t \tag{1}$$

这里 A 是振动的幅度，振动的轮廓线是：

$$A\cos kx \tag{2}$$

$k = \frac{2\pi}{\lambda}$ 是波矢，λ 是波长，因为琴弦的两端已经被限制住了，琴弦的长度 L 可以取半波长，一个波长，一个半波长，……，简单说就是半波长的整数倍：

$$L = \frac{n\lambda}{2}, n = 1, 2, 3, ... \tag{3}$$

波长可以表示为：

$$\lambda = \frac{2L}{n} \tag{4}$$

这就是合乎比例.

考虑到弦上波速 v 是个常量，频率可以表示为：

$$\nu = \frac{nv}{2L} \tag{5}$$

给定弦长 L，只有这样的波动，或这样波动的叠加才可以存在. 进一步讲，如果我们考虑一个符合两端被限制住的琴弦的一般运动，这个一般运动总是可以被分解为一系列不同 n 取值的，波长为 $\lambda_n = \frac{2L}{n}$，频率为 $\nu_n = \frac{nv}{2L}$ 的振动的叠加.

$$\sum_n A_n \cos\frac{n\pi x}{L} \cos\frac{n\pi v t}{L} \tag{6}$$

我们管 $n = 1$ 的音叫做基音，这个频率 $\frac{v}{2L}$ 的声音是最主要的，但弦上也会有 $n = 2, 3, ...$ 的成分，这些音叫做泛音.

拨动长度 L 的琴弦，我们听到的是基因和泛音的混合，最主要的是基因，频率为 $\nu_1 = \frac{v}{2L}$，其次是第一个泛音，频率为 $\nu_2 = 2\nu_1$，它们之间是 1: 2 的关系.

假如我们把琴弦的长度减半，其实就是用手在弦长的一半按住琴弦，此时我们会有新的弦长 $L/2$，同时新的基因频率 $2\nu_1$，但此时，因为弦长只剩下一半了，我们拨动琴弦发出的声音里就没有 ν_1 的成分了.

听起来的感觉是这样的，首先 $L/2$ 琴弦发出的音和 L 琴弦发出的音很像，其次 $L/2$ 琴弦发出的音当然要比 L 琴弦发出的音要高，这就好比是一个人沿螺旋形的楼梯升高，每个台阶都对应一个特定音高的音，在螺旋式升高了几个音之后我们又回到了起始位置，只是高了一些，我们还可以继续螺旋升高，每提升一个台阶都会感觉和曾经的某个台阶很像，只是更高了.

在音乐理论里，我们管这个结构叫“八度”，当音高由 ν_0 提高一倍到 $2\nu_0$ 的时候，我们就说“升了八度”. 类似地，当音高由 nu_0 降一倍到 $\nu_0/2$ 时，我们就说“降了八度”. 我们一般能发出一个八度再加一个四度的音.

毕达哥拉斯的和声学

毕达哥拉斯研究了音和形的关系，并发现这个关系可以被数字精确地描述. 比如我们刚刚讨论过的，当弦乐器的弦长比是 1：2 时，频率比是 2：1，正好对应音乐理论中的“八度音程”.

八度关系本来就存在于音乐实践中，属于人的日常经验. 现在发现“一个八度”可以表示为精确的数字比 1：2，这个数字比其实是对形的描述. 只是因为这里弦是一维的，我们对形的描述比较简单.

一个日常经验可以对应于一个数字的比例关系是足够让人兴奋的，毕达哥拉斯讲“万物皆数”，其实讲的是“万物皆合乎比例”. 只有合乎比例，万物才能存在，只是这些比例有待我们的发现.

合乎比例是个静态的世界观.

除了 1：2，毕达哥拉斯还发现当弦长比是 2：3 时，音的关系是音乐理论中的五度音程. 而弦长比是 3：4 时是音乐理论中的四度音程. 毕达哥拉斯只发现了这几个关系. 它足够优美，但还不足以解释音乐理论中所有的音. 但这已经足够他喟瑟的了.

更重要的是他开辟了一个用数字、用比例关系去研究音乐的方法，进而是研究整个宇宙万物的方法，可以说今天的理论家都是毕达哥拉斯的信徒.

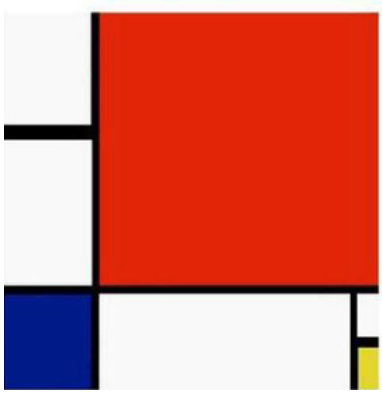


图 3: 蒙德里安的作品是“万物皆数”、“整体和谐”观念在绘画领域的实践.

毕达哥拉斯方案的缺陷是他被简单数字迷住了，1：2，2：3，3：4 确实解释了八度音程、五度音程、和四度音程．但再要想把人对声音的感官经验——极其灵敏的感官经验——和简单数字比（m：n）建立关系就很困难了．

根据近代的十二平均律，我们在八度音程里面做 12 均分，这个均分是合乎比例地分——作为人，我们当然是凭我们的耳朵来分，这里我们必须赞叹人听觉器官的精密——我们要找到某个合适的比例因子 q，使得：

$1\nu_0$，$q\nu_0$，$q^2\nu_0$，……$q^{12}\nu_0 = 2\nu_0$

因此：

$$q = 2^{\frac{1}{12}} \tag{7}$$

这里难的是对 2 开 12 次方，$\sqrt{2}$ 就已经是无理数了，即 $\sqrt{2}$ 就已经没办法表示成一个简单数字的比例了！这是毕达哥拉斯方案失败的原因．

我们解出：$q \approx 1.059463$，并制表：

表 1:　　平均律

n	q^n
0	1
1	1.059463
2	1.122462
…	…
11	1.8877
12	2

四度音程对应的弦长比是 3：4，计算出来的频率比是：$\frac{4}{3} = 1.33333$，对应十二平均律表格中是 $n = 5$ 的情形，可见：1.33484 和 1.33333 相当接近．五度音程对应的弦长比是 3：2，频率比是：$\frac{2}{3} = 1.5$，对应十二平均律是 $n = 7$，1.49831 和 1.5 也很接近．

天体音乐

音乐与舞蹈相联系，古人总是载歌载舞，而载歌载舞是对“天”，对想象中“绝对秩序”的模仿，通过模仿来表达对“天”和人格化的“天”——神的亲近和虔敬．

根据古人的观念，天是天球，有几重天球，离地球最远的是恒星天，它们构成了一个背景，一个不动的背景．还有行星，“金木水火土，太阳和月亮”，它们相对于不动的背景穿行．每个行星都有自己的天球，以自己独特的方式运动．

天体运行的很慢，在没有灯光污染的古代，天体运行是很合适的研究对象，对恒星而言就是绘制星表，把所有可见的，相对而言都不运动的那些恒星的方位表达出来，所谓方位就是方向，所有恒星离我们是一样远的，它们居于最外层的天球．

在这种叙述下，每个恒星对应一个倾角 θ 和一个方位角 ϕ，我们需要某种制图技术把天球上的恒星投影到平面上，这种制图技术和制作世界地图的技术没有什么区别．我们得到的星图，简单说就是诸星座．

恒星天以下还有土星天球，木星天球，火星天球，太阳天球，金星天球，水星天球和月亮天球．这是按照由外到内的次序，月亮天球离我们最近，月亮之下就是凡俗世界了（月下世界），万物变化不定，没有规律．但自月亮天球及以上就是神圣的所在，天球庄严地运转，超脱于朽坏和变化，被神圣的数学描述．

数字关系是不朽的，诸天也是不朽的，研究天体运行是研究数学，即像毕达哥拉斯在音乐中曾经找到的，找到简单的比例，天球的运行需要合乎比例，并作为一个整体和谐地存在，所谓和谐就是和声（Harmonics）．

这是“万物皆数”观念在天文学中的运用，古希腊的哲学家们已经能够计算太阳的大小，月球的大小，太阳和地球的距离，以及月亮到地球的距离，各个行星运转的周期等等．在他们的眼里，宇宙整体是一个和谐的存在：诸天各有各的半径，这是宇宙的形，而诸天各以不同的速度运转将会发出声音，速度越快音高就越高，月音低沉，土星离地球最远，运行最快因此土星音也是最激昂的．

传说乐器是阿波罗神给人的礼物，它是理性的象征，乐器因形的合乎比例而发出和谐的声音，和谐的声音使人的心灵柔和、敏感，function well 成为一架理性的机器．

据柏拉图的《蒂迈欧篇》，宇宙是造物主理性的设计，在比喻的意义下，我们把宇宙想象为一把里拉琴，诸天的位置对应弦上不同的位置．我们无法想象这天体的音乐是不合乎比例的，虽然我们谁都没有听过天体的音乐（天籁之声），但诸天发出的音乐，有的如男低音，有的如男高音，又有的如女低音，有

的如女高音．并整体符合某种比例，某种和谐关系．就好像毕达哥拉斯发现的弦乐中的 1：2：3：4．

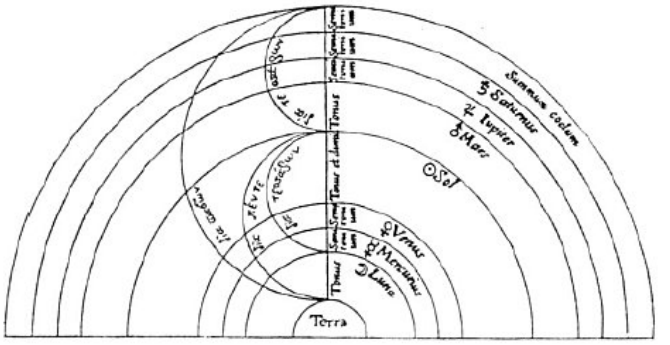


图 4: 天体音乐的一种构想：最大圆弧表示八度音程，向下是两个四度音程，从地球向上还有个五度音程，并与从恒星天向下的四度音程正好重合．频率比由高到低是：$2:\frac{3}{2}:\frac{9}{8}:1$.（根据一本中世纪手抄本的插图绘制，收入 O. Pederson, M. Pihl: Early Physics and Astronomy, 1974．图 6.1）

这里整体和谐的思想是首要的，它或者体现为音乐的悦耳清晰（比如孔子的“三月不知肉味”），或者干脆就体现为一种数学关系的简单和优美（比如毕达哥拉斯的“1：2：3：4”）．人对音乐的欣赏和想象是可以闭上眼睛进行的，任随自己的思绪伴随着音乐的节奏奔跑，这就摆脱了日常经验对思维的限制，成为一种纯内在的，只与不朽的形式相关的理性思维．

西塞罗在《国家篇》中让西庇阿梦见自己身处宇宙之中，听到：

“由各个天体自身的运动和冲击产生出声音，这种声音是那些按恰当比率严格区别开来的各个不相等的音程划分出来的；它由高音和低音混合而成，将各种不同的和音造成统一的音程；……在处于最高点的星天（恒星天）历程上，那里的运动无比地迅速，就发生尖锐的快速的声音；而月球的历程（那是最低的）则以厚重的声音运动着；”

我们谁都没有听到过天体发出的音乐，西塞罗说这是因为我们从小听习惯了，反而听不见了．毕达哥拉斯的说法更高明，他说除了他自己谁也听不见天体的音乐，毕达哥拉斯说：

“他既不创作也不演奏任何人类演奏的那种竖琴或歌的旋律，而只使用一种神秘的、莫测高深的神圣方法，全神贯注于他的听觉和心灵，使他自己沉浸在流动的宇宙谐音之中．……只有他才能听到并理解这种谐音，以及由这些天

体激发起来的和声.”

毕达哥拉斯的高明之处在于点明依靠感官——耳朵——是听不见“天体音乐”的，他需要的沉思，即：全神贯注于听觉和心灵，使自己沉浸在流动的宇宙谐音之中．类似地，柏拉图在《理想国》中嘲笑“仰望星空者”是观星迷，并摆明他自己研究天文学的方法是几何学.

研究天文学也并非是简单地应用数学-几何学，按照柏拉图的说法，研究是要发现理念，现象被理念（光）照亮，新的理念就是新的形式，就是新的类.换言之就是要发现具有表现力的新的数学-几何学.

西庇阿之梦是西方艺术中的常见母题，往前自然是毕达哥拉斯的“天体音乐”和柏拉图的“厄尔神话”，往后则是比如库布里克的《2001 太空漫游》.

影片开始的时候，节奏非常缓慢，人（猿）生活在自然中，直到他们凭视觉洞见了一个抽象的几何形体从天而降，这是对几何学起源的神话式描写．在“数学-几何学”光芒的照耀下，镜头一转人类就进入了太空时代.



图 5: 2001 太空漫游

这就是理性的力量，但首先你需要像那只人（猿）一样被理想的几何打动，为之着迷，这就像毕达哥拉斯发现 1：2：3：4 可以解释音程一样，瞬间被理性的力量击中并宣称“万物皆数”!

而当影片即将结束的时候，我们看到类似西庇阿之梦的梦境，变换的色彩，抽象的几何形体，流动冲撞，这其实是对天体音乐的视觉再现．而随之进入的是更为具象的人类生活，所有美好或可以激发起美好与和谐感觉的画面，西方历史中值得尊敬和记录的种种视觉切片.

向伟大的西方文明致敬!

从毕达哥拉斯和柏拉图始到太空时代终，影片推出的 1968 年正是人类进军太空的最高潮，一年后的 1969 年人类首次登月成功.

在影片即将结束的时候我们听到的是《蓝色多瑙河》.

（从几何形体到《蓝色多瑙河》都是最常见的西方文明的符号.）

开普勒

开普勒是位承前启后的人物，一方面他和托勒密、西塞罗一样醉心于天体音乐的概念，希望能够发现宇宙整体和谐的规律，另一方面他关于行星运动的三个定律直接导致了牛顿的经典力学.

在牛顿的体系里万有引力 $F = GMm/r^2$ 和牛顿第二定律 $F = ma$ 取代了整体和谐，微分和积分取代了数字之间的合乎比例，运动的轨迹取代了静态的天球.

托勒密是古代天文学的集大成者，托勒密天文学的核心是匀速圆周运动，他把行星的运动分解为数个匀速圆周运动的叠加，并很好地与当时的天文学观测数据吻合. 这是一种描述性的理论，表面看起来简单，但进入细节后就会觉得很繁复. 即便是今天，我们也很难凭脑子去想哪怕是几个匀速圆周运动的叠加.

开普勒的天球理论

从整体和谐的观念出发，就需要找到更简单清晰的数学规律，具体说就是某种比例关系. 开普勒的出发点和毕达哥拉斯很类似，都是整数.

在开普勒的年代，哥白尼的体系已经逐渐为天文学家接受，太阳不再是行星，地球取代了它的位置. 已知行星按距离太阳由近到远排列是：水星、金星、地球、火星、木星和土星.

它们的轨道半径比是：

8：15：20：30：115：195

为什么太阳有 6 颗行星，不多不少正好 6 颗，而它们的半径比又正好是以上的整数比？

今天看来开普勒的问题是完全没有意义的，因为根据万有引力定律，行星实际上可以出现在距离太阳的任何距离上，而今天已知的大小行星的数目也远远超出 6 颗. 换句话说，开普勒的问题只有放到“整体和谐”观念下才有意义.

柏拉图在《蒂迈欧篇》中曾用四种正多面体与“水气土火”四种元素对应，但实际上有五种正多面体，这让人感到很不完美.

现在开普勒把五种正多面体与行星所在的天球对应，具体过程是这样的：

1. 水星天球在最里面；
2. 在水星天球之外构造一个正 8 面体，使之与水星天球相切，在正 8 面体外再构造一个外接球，这个球就是金星天球；
3. 在金星天球外构造一个正 20 面体，地球天球就在这个正 20 面体的外接球上；
4. 在地球天球外构造一个正 12 面体，火星天球就位于这个正 12 面体的外接球上；
5. 在火星天球外构造一个正 4 面体，木星天球就位于这个正四面体的外接球上；
6. 最后在木星天球外构造一个正立方体，土星天球就位于这个正立方体的外接球上.

这样我们就用 5 种正多面体，得到了 6 个行星天球．根据立体几何，我们可以严格证明只有 5 种正多面体，现在不多不少各用一次，外接内切得到了正好 6 个行星天球．而在当时人的知识里，太阳只有 6 颗行星，不多不少 6 个天球，每个天球上镶嵌上 1 颗行星.

更加令人赞叹的是，根据开普勒的天球套天球模型，我们能精确地计算出 6 个天球的半径比，它们正好是：

8：15：20：30：115：195

误差有，但不大（不超过 5%）.

这个结果太完美了，可谓是毕达哥拉斯“万物皆数”纲领下的巅峰之作．日后开普勒虽然有更为人称道的行星运动三定律，但他本人仍然最钟爱这个“整体和谐”观念下的理论.

开普勒的行星运动定律

开普勒的这个“古典理论”并没有受到当时学术圈的重视，但他树立了他作为优秀数学家的名声．第谷·布拉赫是当时最了不起的实验家，第谷读过开普勒关于天球的著作，虽然不以为然，但主动寻求与他合作.

第谷积累了当时最丰富的对行星观测的资料（也有恒星的），但仅仅是观测就已经耗费了他一生的精力，现在他预感到他的人生不久了，于是把资料留给一位优秀的数学家——开普勒——期待他能把资料整理出来并有所发现，给他带来名声.

第谷的数据中尤其以火星的数据特别详尽，但当开普勒试图用哥白尼的体系对这些数据进行处理时，却碰到了麻烦.

假设火星沿一个匀速圆周的轨道围绕太阳运动，基于哥白尼理论的计算和第谷的实验数据差别较大. 本来这个差距可以通过假设更复杂的圆周运动体系来处理——即假设火星同时参与几个匀速圆周运动——这是托勒密和哥白尼体系中都允许的技巧.

这么做带来的是概念上的简单，即只使用更容易让人理解的匀速圆周运动来模仿行星的运动，但从技术的角度，当面对越来越精确的观测数据的时候就会显得太繁琐，并且是难以想象的. 开普勒对此非常不满意，于是开始尝试用更多曲线来模仿行星的运动，而不仅仅是限于匀速圆周运动.

1. 开普勒关于行星运动的第一个定律说：行星按椭圆轨道围绕太阳运动，太阳在椭圆的一个焦点上.

   轨道当然是一种静态的观点，因为它并不涉及快慢.

   我们必须要承认想象一个椭圆轨道，可比要想象哪怕两个匀速圆周运动的叠加容易多了. 开普勒的椭圆把我们的想象力从哥白尼和托勒密的几十个正圆中解放出来了.

   （哥白尼的学说用 34 个正圆解释了托勒密需要 77 个正圆才能解释的天体运动，而开普勒现在只需要 7 个椭圆[3].）

2. 开普勒关于行星运动的第二个定律说：行星在离太阳最近的时候运动速度最快，离太阳最远的时候运动速度最慢. 并且可以表示为一个比例关系：

$$rv = RV \tag{8}$$

[3]第 7 个应该是月球的，月球按椭圆围绕地球运动.

这里 r 表示行星离太阳最近时候的距离，v 表示此时行星运动的速度；而 R 表示行星离太阳最远时候的距离，V 表示此时行星运动的速度.

这里有快慢，但仍然采取“合乎比例”这一静态观点下的语言（和杠杆定律采用的是相同的语言）. 如果不看 v，而看角动量（定义为 $\mathbf{J}=\mathbf{r}\times\mathbf{p}$）的话，角动量是不随时间变化的.

3. 开普勒关于行星运动的第三个定律说：行星做轨道运动的半径 R ——严格说应该是“行星离太阳最近距离”加“行星离太阳最远距离”之和的一半——的立方与行星做轨道运动周期 T 的平方之比是个常数.

   即：$\frac{R^3}{T^2}$ 是个常数.

   这其实也是在讲运动要合乎比例，只是这个比例更复杂，涉及了立方和平方，考虑到它对所有的行星都适用，这是个强大的、令人耳目一新的定律.

如此优美普适比例关系的背后一定存在着个解释，就好像毕达哥拉斯的“1：2：3：4”关系的背后是关于琴弦振动的理论.

牛顿的经典力学

开普勒定律的背后是牛顿的经典力学. 我们现在来勾勒其轮廓：

1. 物体不受外力时，物体将保持匀速直线运动或静止状态.

2. 物体运动状态的改变正比于物体所受的外力之和，反比于物体本身的质量. 即：

$$F=ma \tag{9}$$

   这里质量 m 被定义为物体维持运动状态难易程度的量度.

   “物体运动状态的改变”就是“物体的加速度”：

$$a=\frac{dv}{dt}=\frac{d^2x}{dt^2} \tag{10}$$

3. 两个具有质量的物体之间会有万有引力，万有引力正比于两个物体质量的乘积，同时反比于两物体间距离的平方.

$$F = \frac{GMm}{r^2} \tag{11}$$

假设 A、B 之间存在引力相互作用，A 给 B 多大的力，B 就给 A 多大的力，只是方向反了.

仔细读的话，这里有两种定义质量的方式，一种是通过运动定义的，物体保持原运动状态的难易程度，叫惯性质量；另一种是通过引力定义的，叫引力质量. 我们假设引力质量和惯性质量是相同的，这里并没有太多道理可讲，或者看做是实验（比如落体实验）的结果（同时落地），或者干脆讲这就是个假设，一个迄今不会给理论带来麻烦的假设，不但不会带来麻烦，还会带来好处，比如它是研究广义相对论的出发点.

牛顿的体系和古典的“天球音乐”模型相比差距很大. 在牛顿的体系里力是核心概念，力驱动行星运动，相比于太阳，行星很小，我们可进一步把行星抽象为具有质量的点（mass point，质点），它在万有引力的驱动下运动.

我们要想了解行星的运动，需要发展求解微分方程的技巧，即如何求解：

$$F = ma \tag{12}$$

这是一个关于位置 $x(t)$ 的二阶微分方程，所谓二阶就是这里出现了对时间 t 的二阶微分 $\frac{d^2}{dt^2}$. 我们需要积分一次求出速度如何随时间变化 $v(t)$，然后再积分一次求出位置如何随时间变化 $x(t)$，最后再根据 $x(t)$ 得到行星运动的轨迹，比如一个椭圆.

轨迹是整体、静止观念下的，但为了得到轨迹，我们必须由力 F 而 $v(t)$，再 $x(t)$. 从数学的角度，这当然要比列等式，加减乘除、乘方、开方要难. 并且这里真正具有运动的概念了，或者说变化，时时刻刻的变化是个逃不掉的概念.

这可以从对速度的定义看出：

$$v = \frac{dx}{dt} = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} \tag{13}$$

如果仅仅把速度定义为：

$$v = \frac{\Delta x}{\Delta t} = \frac{x(t_2) - x(t_1)}{t_2 - t_1} \tag{14}$$

这还是一个静态的图像，即我们在时刻 t_2 和时刻 t_1 各拍摄一张快照，分别凝神观瞧，用尺子做测量，然后代进公式里计算．我们怎么说这个速度 v 才不过分？它不属于 t_2，也不属于 t_1，它是 t_1 到 t_2 之间的平均效果．

那我们还能说时刻 t_0 时的速度 $v(t_0)$ 吗？如果不能加速度 a 的定义就成了空中楼阁．

在牛顿的体系里，速度必须对每一个点都有意义，但如果我们把眼光只聚焦在一点上是不可能有速度的，速度是变化，对一个点怎么能说变化呢？此时我们考虑的是一个点，但这一点的近邻也必须包括进来，否则就不会有变化，不会有速度．

记号：$\lim\limits_{\Delta t \to 0} \frac{x(t+\Delta t) - x(t)}{\Delta t}$ 表示的是一系列的操作，我们先测 $\Delta t = 1$ 秒，然后 0.1 秒，0.01 秒，0.001 秒……

如此构造出一个无穷的序列，就像我们曾经讨论过的 0，1，2，3……，这是一个用自然数标记的序列，它是可数的（countable），但无限延伸，没头儿．从技术的角度，我们会发现这个序列往往会很快收敛在某个稳定值上，这个就叫极限．

某时刻 t_0 的速度 $v(t_0)$ 因此就有了定义，它是在极限下得到定义的，这个极限可能存在，可能不存在，但我们物理上只讨论那些极限存在的情况．所谓微积分就是要发展出一套这么做的技巧，更重要的是逻辑体系，把它说严谨，用公理、定义和定律的体系．

现在我们就有了经典力学．

原子光谱的玻尔理论

经典电磁学

经典力学里有一条不太让我们放心，这里面似乎只有引力，而引力是个太弱的力．两个人面对面站着，吹口气的力都比他们之间的引力大．

在我们的生活中，除重力外，其他力基本上都不是引力，比如弹簧的弹性回复力，比如我们俩亲切地抱着的压力，比如摩擦力……

这些力的来源是电磁相互作用．

电现象、磁现象和光现象都是人类很早就发现并研究的现象．其规律被麦克斯韦总结成一组非常优美也抽象的数学公式（麦克斯韦方程组）：

$$\begin{aligned} \nabla \cdot E &= \frac{\rho}{\epsilon_0} \\ \nabla \cdot B &= 0 \\ \nabla \times E &= -\frac{\partial B}{\partial t} \\ \nabla \times B &= \mu_0 j + \mu_0 \epsilon_0 \frac{\partial E}{\partial t} \end{aligned} \tag{15}$$

1. 这里第一个式子说的事情和引力很类似，写成力的形式：

$$F = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r^2} \tag{16}$$

即两个电荷之间的力与电量的乘积 $q_1 \cdot q_2$ 成正比，与两个电荷之间距离的平方 r^2 成反比．

这个结果和万有引力几乎一模一样，有两点不同：（1）我们这里讨论的静电力（也叫库仑力）比引力要强的多；（2）有两种电荷，相同电荷之间是斥力，而相异电荷之间是引力．

2. 第二个式子说的是，在自然界中不存在磁单极子，但物理学家早就准备好了一套磁单极子存在的理论了，只等哪天找到它，就在方程的右侧加上一项．

3. 第三个式子和第四个式子说的是变化的磁场会感生电场，而变化的电场也会感生磁场；前者是发电机的原理，而后者是电磁铁的原理．它们在一起可以解释电磁辐射或光波的存在．

在电磁学的研究中，由于电磁相互作用太强了，力反而不是重点，重点是场，是电场和磁场在时空中的分布和传播．比如对一个电的振子，能量会以电磁波的形式向外辐射，这是必须考虑的物理过程．而对引力，我们就根本不需要考虑引力波．

经典理论的困难

假设一个质子和一个电子，相距 0.5×10^{-10} 米，这个距离就是氢原子中电子和质子的距离．我们可以先计算他们之间的电磁相互作用，电子和质子都带一个单位的电荷 e，但符号相反电子带负电，质子带正电．

$$F_e = -\frac{1}{4\pi\epsilon_0}\frac{e^2}{r^2} \tag{17}$$

这里：真空电容率，$\epsilon_0 = 8.854\times10^{-12}F\cdot m^{-1}$；电子电荷，$e = 1.602\times10^{-19}C$．代入计算得：

$$F_e = 8.25 \times 10^{-8}(N) \tag{18}$$

看起来很小，但要看和谁比．

现在计算电子和质子之间的万有引力，还是这个间距，

$$F_g = \frac{Gm_em_p}{r^2} \tag{19}$$

这里：引力常数，$G = 6.673 \times 10^{-11}N \cdot m^2/kg^2$；电子质量，$m_e = 9.109 \times 10^{-31}kg$；质子质量，$m_p = 1.673 \times 10^{-27}kg$．代入计算得：

$$F_g = 3.63 \times 10^{-47}(N) \tag{20}$$

电磁相互作用和引力的比值是：

$$\frac{F_e}{F_g} = 2.27 \times 10^{39} \tag{21}$$

电磁相互作用要远远大于引力．

这意味着引力对研究原子尺寸的物理问题是完全可以忽略不计的．

现在只考虑电磁相互作用，电磁相互作用使电子围绕质子运动，但电子会向外辐射电磁波，它损失能量的速度太快了，电子会飞快地撞向质子．估算的结果是只需要 10^{-11} 秒数量级的时间电子就会掉到质子上，即氢原子是不稳定的．

它只能在这个世界上存在一瞬，但经验告诉我们，我们身体里有大量的氢原子，而我们是稳定的．

这就是把经典理论应用到原子现象时碰到的困难．

氢原子光谱

一个成功的原子理论应该能够描述原子物理学中的典型现象，原子是稳定的存在，这当然是其中很重要的一个现象．但除此之外还有更独特地属于原子的现象——光谱现象．

光谱现象分为两类，发射光谱和吸收光谱．

所谓发射光谱就是炙热原子发射的光通过三棱镜分光形成的谱分布，吸收光谱是当热光源发出的光通过冷原子气体时，部分光被原子吸收后形成的谱分布．所谓谱分布就是光强相对于波长的分布．

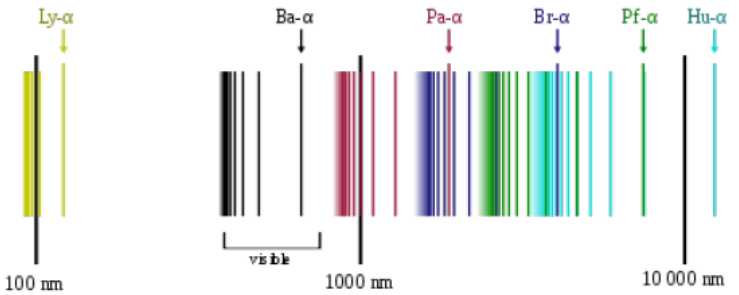


图 6: 氢原子光谱

我们发现发射光谱中原子发出的特定波长的光，在吸收光谱中也出现，只不过发射变成了吸收．

通俗地说原子就是一个爱戴戒指的人，但她只戴特定尺寸的戒指，戴腻了她就扔，扔掉戒指的尺寸和她拿来戴的尺寸完全一样．对这个现象的解释倒也简单，因为她有 5 个手指，每个手指粗细不同，但都有确定的尺寸．

我们有理由猜测光谱与原子的本性有关，实验也确实支持我们的这种想法，每种原子都有不同的光谱，它们的谱线出现在不同波长的位置上，就好像是指纹，人人不同，成为我们的标识．

即便是对最简单原子的光谱，比如氢原子的光谱，乍看起来都是很复杂的，但感觉它们是有规律的，或用老话讲，看起来它们是合乎比例的，只是这个比例有待我们的发现．

就好像毕达哥拉斯发现和声学里的 1：2：3：4，原子物理早期的突破也来自于人们找到了一个简单、优美的数学式子，这个式子解释了氢原子光谱中可见光区域里的谱线位置（即巴尔末系）：

$$\frac{1}{\lambda}=\frac{4}{B}\left(\frac{1}{2^2}-\frac{1}{n^2}\right) \tag{22}$$

其中 $n=3,4,...$，这就是巴尔末公式，它解释了氢原子光谱中最显著的几条线，很快被推广为里德堡公式：

$$\frac{1}{\lambda}=R_H\left(\frac{1}{n^2}-\frac{1}{n'^2}\right) \tag{23}$$

这里 $n=1,2,...$，$n'=n+1,n+2,....$

里德堡公式可以解释氢原子光谱中所有的谱线位置.

原子的稳定性仍然是很重要的问题，但现在首先要解释为什么会有谱线的规律，这个简单的公式强烈地提示我们在氢原子的问题里存在着更为简单清晰的概念体系和数学结构.

玻尔理论

某种意义上说，玻尔理论从牛顿的经典力学重新退回到了毕达哥拉斯的“天球音乐”图像. 天球之所以只奏响这些特定的音，是因为整体的和谐，是形的制约，使天球只能发出特定音高的音.

现在氢原子就是个小天球，它只发射（或吸收）特定波长的光，特定波长的光就是特定频率的光，频率（音）是“形”制约的结果，形的制约就是几何关系，两端固定的弦就是一维振动的形. 现在我们需要发现的是氢原子的形.

首先这里有个概念需要澄清，我们说是说氢原子，但其实这里我们研究的是电子，因为质子比电子质量大太多了，质子运动的速度比电子运动的速度要小很多，或者说质子很难跟得上电子的运动，所以我们这里只需要研究电子的运动就可以了，而质子则作为固定的背景最后考虑.

现在假设有一个毕达哥拉斯的信徒来研究原子中电子的运动，他会怎么说呢？

首先电子仍然会受质子的吸引，这个力是：

$$F=\frac{1}{4\pi\epsilon_0}\frac{e^2}{r^2} \tag{24}$$

假设电子处在某个半径为 r 的正圆轨道上，电子的势能是：

$$V(r) = -\frac{1}{4\pi\epsilon_0}\frac{e^2}{r} \tag{25}$$

电子的动能是：

$$K = \frac{1}{8\pi\epsilon_0}\frac{e^2}{r} \tag{26}$$

电子的总能量是：

$$E = K + V = -\frac{1}{8\pi\epsilon_0}\frac{e^2}{r} \tag{27}$$

电子只能在特定的轨道上运动，这就好像毕达哥拉斯派把宇宙想象为一把里拉琴，只允许在能奏响整体和谐乐音的位置上放置天球，行星在天球上沿正圆轨道呼啸而过，发出特定频率的声音，行星的速度越快，频率也越高……

现在我们只需要把这幅图像套用过来即可，电子也只能出现在特定半径的轨道上，到底是哪些半径允许，这是有标准的，类似于弦上驻波的整体和谐的标准.

但电子怎么能和波联系起来呢？假如要联系起来又应该怎么联系呢？

硬要往下讲就是假想的历史了．因为玻尔确实不是按这个思路思维的，而德布罗意提出物质波又在玻尔之后，受玻尔模型的启发.

我们现在要求自己有如神助，假想一个能代表电子和谐运动的波沿着电子的轨道运行，运转一圈正好是波长的整数倍，首尾相接形成圆轨道上的驻波.

$$2\pi r = n\lambda \tag{28}$$

我们又想到行星运转越快对应发出的声音就越高，频率 $\omega = 2\pi/T$ 是时间上的调制，还有波矢 $k = 2\pi/\lambda$，反映的是空间上的调制.

假如我们让电子运行的速度乘以质量（即动量）正比于波矢 k 会有什么结果呢？（纯属猜测）

假设比例因子是 $\hbar$，这个比例因子是研究原子尺度物理问题必须出现的.

$$p = mv = \hbar k = \hbar\frac{2\pi}{\lambda} \tag{29}$$

因此：

$$mvr = \hbar\frac{2\pi r}{\lambda} = \hbar\frac{n\lambda}{\lambda} = n\hbar \tag{30}$$

即：

$$mvr = n\hbar \tag{31}$$

这就是我们猜测出的对氢原子而言，整体和谐的条件.

电子只能处在由 $mvr = n\hbar$（角动量量子化条件）规定的 r 上，并满足：

$$\frac{mv^2}{2} = \frac{1}{2}m\left(\frac{n\hbar}{mr}\right)^2 = \frac{1}{8\pi\epsilon_0}\frac{e^2}{r} \tag{32}$$

求得：

$$r_n = n^2\frac{4\pi\epsilon_0\hbar^2}{me^2}, n = 1, 2, 3, ... \tag{33}$$

定义玻尔半径 a_0 为：

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{me^2} \tag{34}$$

$r_n = n^2 a_0$（代入物理常数，可计算得 $a_0 = 0.529 \times 10^{-10}m$），由这一系列 r_n，我们可以得到一系列的能量 E_n：

$$E_n = -\frac{me^4}{32\pi^2\epsilon_0^2\hbar^2}\cdot\frac{1}{n^2} \tag{35}$$

或代入具体数值：

$$E_n = \frac{-13.6eV}{n^2} \tag{36}$$

因为整体和谐条件的限制，电子只能占据那一系列轨道 r_n，因此能量的取值也是一系列分立的取值 E_n，这一系列分立取值的电子能量就叫做能级．电子离质子越近（n 小），电子的能量越低，反之电子离质子越远（n 大），电子的能量就越高.

氢原子里只有一个电子，假设这一个电子处在比较高能量的轨道上，它可以向下跃迁，电子的能量将降低，多余的能量将以光子的形式释放出去，假设较高能级用 n' 标记，较低能级用 n 标记，能量差为：

$$h\nu = \frac{hc}{\lambda} = E_{n'} - E_n = 13.6 \times \left(\frac{1}{n^2} - \frac{1}{n'^2}\right) eV \tag{37}$$

于是我们得到里德堡公式.

$$\frac{1}{\lambda} = \frac{13.6eV}{hc} \times \left(\frac{1}{n^2} - \frac{1}{n'^2}\right) \tag{38}$$

原子的稳定性还是问题吗？在整体和谐的观念下其实已经没有问题了．我们急需澄清的是那个与电子的运动状态相联系的波到底是什么？此刻——玻尔模型的出现——说明替代范式已经出现，与其苦苦执着于老范式，不如发展新范式，而在新范式下，很多老问题是没有意义的，它们被更急迫的问题所替代.

习题 1

考虑一个平面版的开普勒“圆环套圆环模型”，最内层是个圆轨道，比如水星，然后在外面套一个正方形，使正方形的四条边与水星轨道相切，然后再在正方形的外面再外接一个圆，作为金星的轨道．我们在金星的轨道外做一个正三角形，使正三角形的三条边与金星的轨道相切，最后在金星的轨道外面做一个外接圆，是地球的轨道.

现在我们可以求出水星轨道、金星轨道和地球轨道的比值了，并与观测数据进行比较.

习题 2

下表罗列了开普勒时代六大行星的轨道和周期，你能否把六大行星的轨道比表示为简单整数比，并计算这么做的误差是多少.

表 2: 六大行星的轨道和周期

行星	轨道半径（天文单位）	周期（年）
水星	0.387	0.24
金星	0.723	0.615
地球	1.0	1.0
火星	1.524	1.88
木星	5.2	11.86
土星	9.539	29.46

习题 3

下图是古希腊时期的音阶图，包含两个八度和一个四度，试计算出各个频率比．同时我们也发现某些音是无法根据毕达哥拉斯的“1：2：3：4”规律计算的．这构成了音乐理论中的一个问题，即如何把剩下音的规律用数字表达出来．

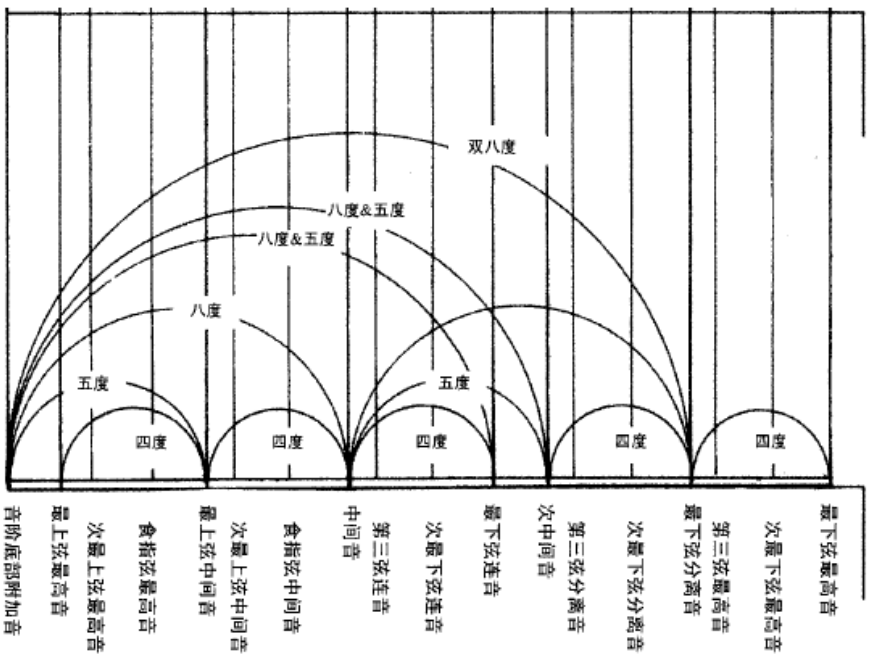


图 7: 和声的基本原理：音阶.（维特鲁威，《建筑十书》，北京大学出版社，2012. 图 81）

量子力学

预备知识 动量和能量[10]

我们试图在科普部分以通俗易懂的语言简单介绍量子力学的基本假设，有了这些基本假设，一切其他结论都可以被推导出来．注意这里介绍的是非相对论量子力学，只适用于低速微观粒子．另外量子力学中有不同的**诠释**，本书使用常见的**哥本哈根诠释**.

波函数

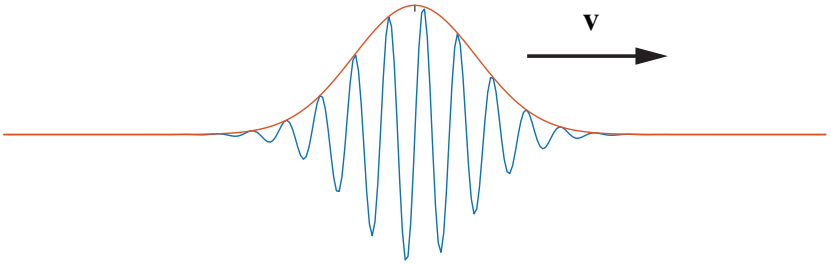


图 1: 蓝线是一维波函数/绳子的形状，橙线是振幅．某个位置的振幅越大，粒子就越有可能出现在该位置（与振幅平方成正比）．图中的形状叫做波包，一般来说波函数可以是几乎任何形状．

这里只讨论一维的情况，即粒子沿直线运动．经典力学中用位置关于时间的函数 $x(t)$ 描述一个粒子的运动，而量子力学中用波函数描述．波函数是一个关于位置和时间的函数，一般记为 $\Psi(x,t)$，例如一条紧绷的橡皮绳（或弦）上的波动，绳子的高度 y 是关于位置 x 和时间 t 的函数[4]．某个位置振动的幅度的平方就是粒子在这个位置出现的概率[5]．经典力学中，给出粒子初始的位置和速度和势能，它将按照牛顿第二定律运动．量子力学中，给出粒子的初始波函数，波函数将按照含时薛定谔方程变化，把势能和某个时刻的波函数代入薛定谔方程，就可以解出任意时刻的波函数（这类方程并不是数的方程，而是函数的方程，所以解出的是函数，而不是数）．薛定谔方程是量子力学的公设之一，就像牛顿三定律是牛顿力学的公设．

量子力学用波函数表示粒子的状态是因为许多实验发现微观粒子（如电子，质子，中子）在运动时具有一些经典物理中波的性质（例如声波，水波，光波），这就是著名的**波粒二象性**．

波包

即使粒子的位置一般不能精确确定，我们往往也能确定其大致范围（例如实验物理学家往往知道某粒子在实验室而不在月球），所以波函数往往只有在

[4]波函数和绳子的运动方式相似但不完全相同

[5]量子力学中的波函数的值可以是复数，粒子出现在某点的概率是复数模长的平方．这里为了简单暂时使用实数，可以看做是复数的实部

一定范围内不为零（我们只能在一定空间范围内测到检测到粒子）．将振幅关于位置的函数画出来，其形状往往像一个包（越靠近中间，检测到粒子的概率越大），我们称为**波包**（图 1）．

自由粒子

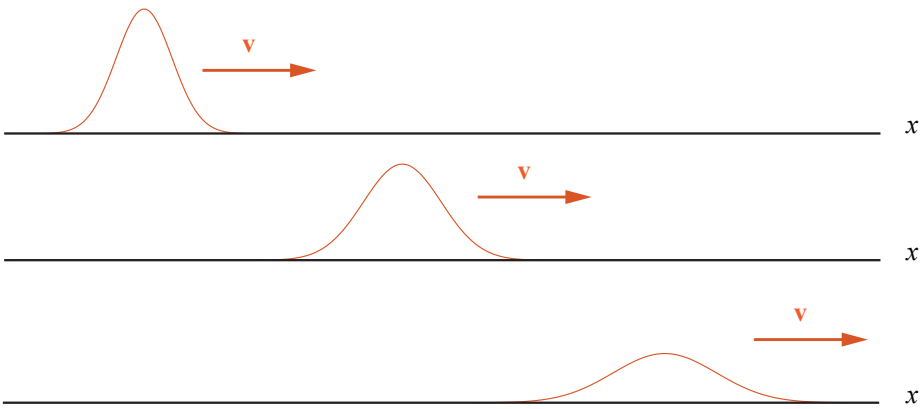


图 2: 对应自由粒子的波包．注意图中波包仅画出了图 1 中的振幅，没画出波函数．

先看不受力的粒子（叫自由粒子），在经典力学中，粒子不受力时静止或者做匀速运动．量子力学中，若粒子不受力（势能处处为常数）且初始时波函数为波包且具有初速度，波包中心就会匀速运动，但同时波包会扩散（越变越宽，越变越矮）．想象一根紧绷的绳子的一端突然迅速上下抖了几下又停下来，那么将产生一个波包将从绳的一段传到另一端（不同的是绳子上的波包形状不会变化）．如果波包开始是静止，那么波包只会在原地扩散（这也与绳子的运动方式不同）[6]．

在一般问题中，波包的长度非常小（原子尺度）．当我们回到宏观中，波包的大小就可以忽略不计，把波包近似成质点，那波包的位置就是质点的位置，速度就是质点的速度，这样就得到了匀速运动的宏观质点．

[6] 小时百科的图标就是自由高斯波包，详见“小时百科图标[1978]”

一维势能

以下通过几个例子来定性说明在一些特定势能曲线下粒子如何运动（即非自由粒子）.

散射

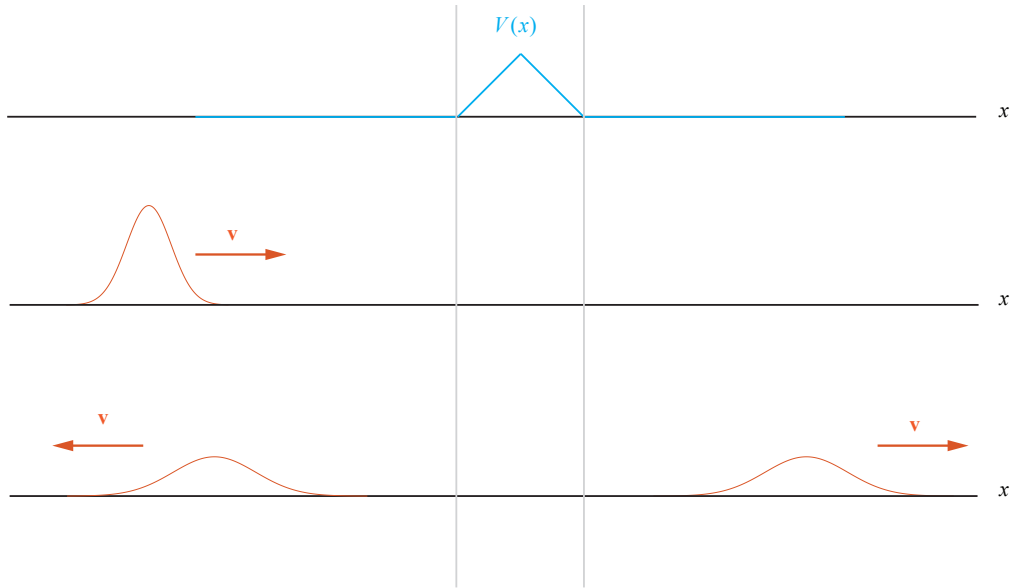


图 3: 波包遇到方势垒后的散射.

经典力学中，若粒子匀速入射到如图 3 所示的势能 $V(x)$（叫做势垒）上，如果粒子的初始动能小于势垒的最大值，则粒子将会原路返回，若粒子初始动能大于势垒的最大值，则粒子将继续前进. 在量子力学中，无论波包以什么初速度入射，波包都会被分为两个不同方向运动的波包，但两个波包的相对大小取决于初始波包的形状和动能. 定性来说，入射波包的能量越大，粒子就越有可能穿过势垒，所以向右运动的波包就相对越大. 反之，如果入射波包的能量越小，粒子就越有可能被反弹，所以向左运动的波包就越大.

这个问题中最不可思议的地方在于当粒子的初始动能小于势垒的最大值时，它仍会有一定的概率穿过势垒，这在经典力学中时不可能的. 顾名思义，我们把这种量子力学中特有的现象叫做**隧道效应**. 隧道效应带给我们的一个

显而易见的疑惑是：如果我们在势垒内部检测到粒子，那它的动能岂不是小于零？如果再求速度或动量（需要开方），岂不是会得到虚数？毫无疑问，虚数是没有物理意义的，但思考一下可以发现在这个问题隐藏了一个假设就是我们可以同时准确地测量粒子的位置和动能，也就是同时准确测量位置和动量．而根据下面将要介绍的不确定性原理，这是不可能的．

定性来说，粒子进入或穿过势垒的概率与两个因素有关：势垒的高度和宽度．势垒比动能大得越多，概率就越小，势垒越宽，概率也越小（波函数的振幅呈指数下降）．

有限深势阱

经典力学中的有限深势阱见“动量和能量[10]”．在量子力学中，被困在势阱中的波包在有限深势阱中同样一边来回碰撞一边扩散．但由于隧道效应，波函数会稍微渗透到势阱的外部

对经典的粒子（经典力学中的粒子）来说，当粒子被困在势阱内时，两种势阱带来的效果是一样的．

> 未完成：无限深势阱并没有提到！

波的叠加

为什么要用波函数来描述粒子？因为波的一个重要特点就是具有空间分布且可以叠加．如果一个波包 $\Psi_1(x,t)$ 是薛定谔方程的解，且另一个波包 $\Psi_2(x,t)$ 也是薛定谔方程的解，那这两个波包叠加（即把两个波函数相加）后 $\Psi(x,t)=\Psi_1(x,t)+\Psi_2(x,t)$ 仍然是薛定谔方程的解．例如，两个不同方向运动的波包相遇然后远离．注意这里两个波包并不代表两个粒子，而是代表一个粒子有可能在一个波包的位置被观测到也有可能在另一个波包的位置被观测到．

另一个关于波的可叠加性的经典实验就是双缝干涉，当两个点状波源各自发出相同频率的圆形波时，在远处放一个观测屏，就可以观测到干涉条纹（见图 4 和图 5）．这类干涉现象在量子物理被发现很久以前就已经有了详细的研究．然而当物质波的概念提出后，物理学家竟然也观测到了微观粒子产生的干涉（如电子的双缝干涉），进而产生了类似波函数和波动方程（即薛定谔方程）这类的数学工具．

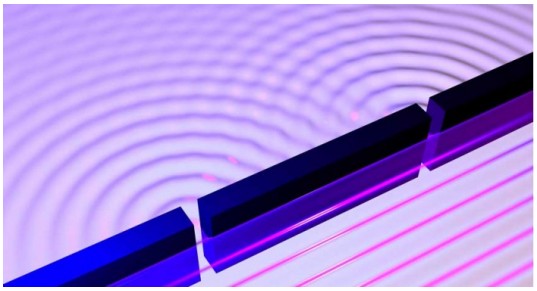


图 4: 水波的干涉

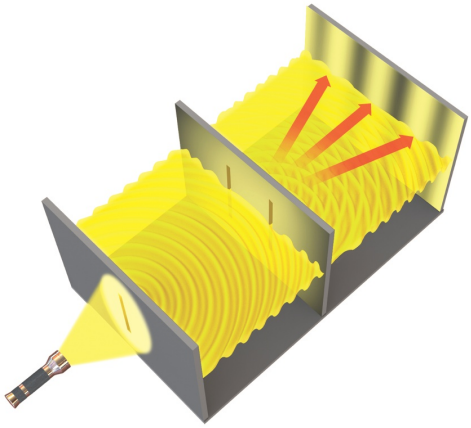


图 5: 光波的干涉

能量本征态

在经典力学中，要测量一个粒子的能量，我们只需要先观测其速度，则能量等于动能（$mv^2/2$）加上该位置的势能．量子力学中，测量波包的能量会得到什么结果呢？这时候我们需要薛定谔方程的另一个用途：对某个势能，我们能解出一些（往往是无穷多个）特殊的波函数，称为能量的**本征函数**或者**本征态**．这些波函数特殊在哪里呢？一般来说，我们对一个粒子测某个物理量（位置，动量，能量），测得的值都是不确定的（这与仪器的误差无关，而是本质上的不确定），但如果粒子的波函数恰好是一个能量本征态，那么测量其能量会得到唯一确定的值（先不讨论用什么方法测）叫能量的**本征值**．每个本征态对应唯一一个本征值．

要测量任意波包（波函数）的能量，我们可以想办法用不同的能量本征态叠加来凑出所需波函数，例如第一个本征态乘以常数 C_1 加上第二个本征态乘

以常数 C_2 等，这叫波函数的线性叠加．如果对这个波包测量能量，测到第 i 个能量本征值的概率等于 $|C_i|^2$（之前提到波函数一般是复数，这里的系数事实上也同样是复数，而这里的绝对值符号代表复数的模长）．也就是说，某个能量本征态占总波函数的比例越多，越有可能测出对应的能量本征值．

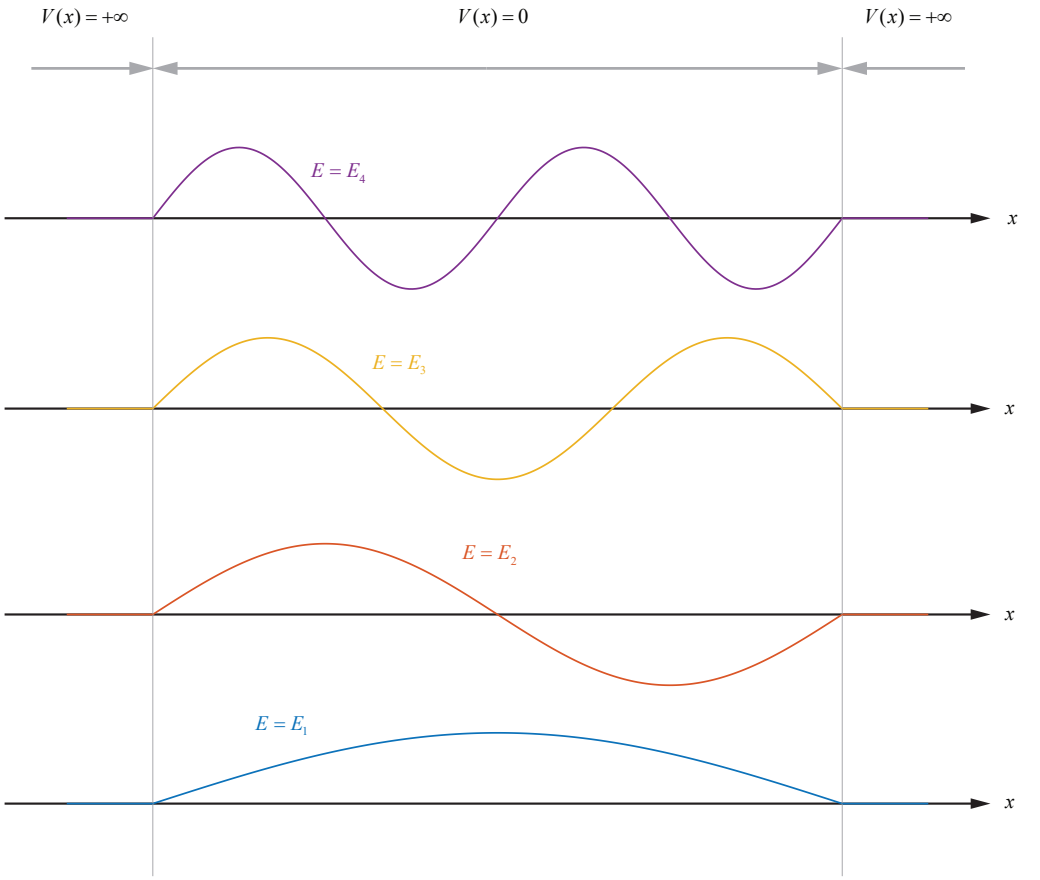


图 6: 无限深势阱的一些波函数能量本征态，每个能量 E_i 叫做一个能级，共有无穷多个能级，将这些本征态叠加就能获得任何波函数例如波包．

测量理论

我们介绍量子力学的另一组重要公设：这里将其统称为**测量理论**．这里的**测量**是一个抽象的概念，我们先不讨论用什么仪器进行测量，也不讨论现实中这种测量可不可行，只是假设存在“测量某个物理量”这种操作．

上面介绍的能量测量的步骤就是测量理论的一个特例．量子力学中的任何

物理量的测量都可以由类似的方法求出，只是不同的物理量都有各自的一组本征态，可以通过物理量对应的**本征方程**求出．以下我们先不讨论如何列出本征方程以及解出本征态，而是直接给出本征态．如果要对一个粒子测量某物理量，就先计算出这个物理量的各个本征态，再计算它们占波包的比例 C_i，测量到第 i 个本征值的可能性就是 $|C_i|^2$．

这里有一个条件是任何可能存在的波包都可以由任何测量量的本征态叠加而成，这叫做本征态的**完备性**．

例 1　位置和动量的本征态

位置的本征态是一个无穷窄无限高的波包，称为 δ 函数．δ 函数描述的粒子只可能出现在某点．

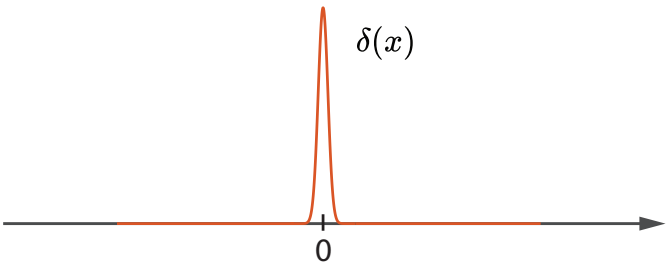


图 7: δ 函数示意图．注意其宽度为无穷窄．$\delta(x)$ 的中心在坐标原点，如果需要任意 x_0 处的 δ 函数，可以使用 $\delta(x-x_0)$．

动量的本征态是一个无限长的波，即某个频率的 sin 函数，称为平面波（区别于波包），见图 8．在量子力学中，波也叫德布罗意波．

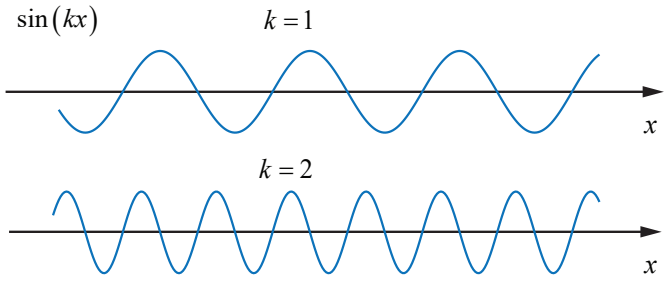


图 8: 一维平面波 $\sin(kx)$，由负无穷延伸至正无穷．

注意图中画出的是波动关于位置 x 的函数，而不是关于时间 t 的函数，所

以我们把 k 叫做空间频率[7]（或者波数），与波长[8]（即空间周期）λ 的关系是 $k = 2\pi/\lambda$．动量的本征值为

$$p = \hbar k = \frac{h}{\lambda} \tag{1}$$

其中 $\hbar = h/(2\pi)$ 是约化普朗克常数，h 是普朗克常数．所以和空间频率成正比，和波长成反比．

测量改变粒子状态

在经典力学中，我们完全可以假设测量不会对粒子的状态产生影响（例如给粒子拍照不会改变它的运动状态）．而在量子力学中，由于考虑的粒子质量非常小，测量会不可避免地改变粒子的状态．

测量理论假设，在对某个物理量进行测量后，如果得到第 i 个本征值，粒子的状态会立即**坍缩**为这个本征值对应的本征态，然后继续按照薛定谔方程演化．

那么我们应该如何理解测到某个本征值的概率 $|C_i|^2$ 呢？这里显然不是指对处于某状态的粒子连续测量 N 次，就大概有 $|C_i|^2 N$ 次会测到本征值．因为由于测量带来的坍缩，只有第一次测量的概率是 $|C_i|^2$，之后由于波函数会坍缩到第 i 个本征态并继续按薛定谔方程继续演化，再次测量时，我们就要根据新的波函数重新计算每个 C_i，而不是使用第一次测量前波函数．

正确的理解应该是，如果我们有 N 个处于同样状态（具有相同波函数）的相同粒子，对这些粒子进行测量，大概会有 $|C_i|^2 N$ 个次测量得到第 i 个本征值．或者是，我们把同一个粒子测量 N 次，但每次测量后都进行某种操作，将其恢复到测量前的状态，再进行下一次测量．

这就意味着，除非粒子在测量前就已经处于该本征态，测量操作必定会改变粒子的状态．例如粒子的波函数如为图 1 中的波包，当我们测量其位置，如果得到的值为 x_0，那么波函数将会马上变为 x_0 处的一个无限窄的波包（δ 函数）．

[7] 严格来说应该是**角频率**而不是频率

[8] 也叫**德布罗意波长**

连续本征值与离散本征值

上面我们提到一个物理量往往有无穷多个本征波函数. 一些情况下我们求出的本征波函数对应的值是不连续的（如无限深势阱中的能量本征态），而另一些情况下则是连续的（如位置和动量的本征值总是连续的，即它们可以取任意实数）. 经典物理世界中这些物理量都是连续的，而量子中的物理量会出现离散的情况，这就是**量子**这个词的来源（表示一份一份的）.

离散本征值的情况在数学上是容易处理的，我们上面使用的符号 C_i 就是假设了本征值是离散的（因为这里 i 是整数），连续的情况有类似的理解，但数学形式有所不同（比如求和变为积分）. 我们这里再举一个例子.

例 2　角动量

量子力学中，另一个著名的具有离散本征值的物理量就是角动量. 例如一个电子围绕固定的原子核旋转，它就可能具任意方向的角动量. 注意这是一个三维空间中的问题，在直角坐标系中，我们可以将角动量的三个分量看作三个不同的物理量分别为 L_x，L_y 和 L_z. 为了简单起见我们假设这三个量各自只有两个本征值，即[9] $L_x=-1/2$ 或 $+1/2$，$L_y=-1/2$ 或 $+1/2$，$L_z=-1/2$ 或 $+1/2$. 对应的 6 个状态记为[10] $|x-\rangle$，$|x+\rangle$，$|y-\rangle$，$|y+\rangle$，$|z-\rangle$，$|z+\rangle$. 三维空间中的波函数具有 $\Psi(x,y,z,t)$ 的形式，由于波函数的数学形式比较复杂，这里不具体给出.

为了简单起见，我们下面只讨论 x 和 z 两个方向. L_x 本征态与 L_z 本征态之间的关系为

$$\begin{cases} |x+\rangle = \dfrac{1}{\sqrt{2}}|z+\rangle + \dfrac{1}{\sqrt{2}}|z-\rangle \\ |x-\rangle = \dfrac{1}{\sqrt{2}}|z+\rangle - \dfrac{1}{\sqrt{2}}|z-\rangle \end{cases} \tag{2}$$

现在，如果我们对处于 $|x+\rangle$ 态的粒子测量 x 方向的角动量，根据测量理论中对本征态的定义，得到的只可能是其对应的本征值 $L_x=1/2$. 同理，如果对 $|z-\rangle$ 态的粒子测量 z 方向的角动量，只可能得到 $L_z=-1/2$.

---

[9] 以后我们会知道这个例子描述的是自旋为 1/2 的粒子的自旋角动量，这里为了书写方便使用约化普朗克常数 $\hbar$ 作为角动量的单位

[10] 这里我们使用 $|\ldots\rangle$ 来表示波函数，这种符号叫做**狄拉克符号**，如果你不习惯，完全可以把它们记为 $\Psi_{x+}(\ldots)$，$\Psi_{x-}(\ldots)$ 等形式. 事实上，本例讨论的自旋态，并不能用波函数表示，但为了方便理解，我们姑且假设它们可以.

但如果我们对 $|x+\rangle$ 态的粒子测量 z 方向的角动量会发生什么呢？根据测量理论，我们首先需要将 $|x+\rangle$ 表示成 $|z+\rangle$ 和 $|z-\rangle$ 的线性组合，而式 2 中的第一条已经给出了这个关系．于是我们马上得到 $C_1 = 1/\sqrt{2}$，$C_2 = -1/\sqrt{2}$．所以测得 $L_z = -1/2$ 和 $L_z = 1/2$ 的概率分别是 $|C_1|^2 = 1/2$，$|C_2|^2 = 1/2$，即二者概率相同．

根据同样的过程，我们也从式 2 中的第二条得出，若粒子处于 $|x-\rangle$ 态，测量 z 方向角动量得到 $L_z = 1/2$ 和 $L_z = -1/2$ 的概率同样也都是 $1/2$．

薛定谔的猫

注意我们在测量理论中并没有指明什么样的事件会构成测量．还是用上面的例子，假设一个粒子的波函数是两个反方向运动的波包，且两个波包的形状相同．我们在一个方向放一个屏幕用于测量位置，检测到粒子后，屏幕会发光．那么根据测量理论，这个粒子有 $1/2$ 的概率会落到屏幕上（朝着屏幕方向运动的波包）．按照通常的理解，这似乎就对粒子构成了测量．

但是这并不严谨，因为屏幕也是由微观粒子（质子，电子等）构成的，也可以由波函数描述．如果将屏幕与粒子作为考察的系统，我们可以用一个多粒子的波函数来描述这整个系统（多粒子的波函数这里作不介绍）．当一个波包传播到屏幕的位置时，波函数描述下的屏幕的也变为包含发光的本征态与不发光的本征态的叠加，正如粒子的初始状态是向左传播的波函数和向右传播的波函数的叠加．

如果假设“人眼看屏幕”这个操作对屏幕的波函数进行了测量，那么只有当实验者去看屏幕的瞬间，波函数才会坍缩到屏幕发光或不发光的本征态．

这个论述可以一直进行下去，人眼也是由微观粒子构成的，那么如果也纳入考察的系统（用波函数描述），那人眼的状态就会是“看到发光屏幕的人眼”和“没有看到发光屏幕”的人眼这两种本征态的叠加，直到人脑对人眼进行测量后才会得到确定的结果．实验者最终也会处于“看到发光屏幕的人”和“没看到发光屏幕的人”这两种本征态的叠加态．

薛定谔的猫就是这类佯谬中的一个具体例子．在这个假想实验中，一瓶放射性物质旁边放有一个粒子检测器（盖革计数器），当检测器检测到放射性出的粒子时，就会触发一个释放毒药的开关，将箱子里的猫毒死．由于放射出的粒子是由波函数描述的，对这个实验的理解取决于这一系列过程中的哪一步构

成测量．如果我们认为检测器对放射粒子构成了测量，那么从测量到结果的那一刻，一切都是确定的，也就不存在佯谬．但如果我们认为只有当人打开箱子看到猫的那一刻才构成测量，那么在打开箱子前猫都处于死和活两个本征态的叠加态．更不可思议的理解是，根本不存在测量，当人打开箱子检查后，人也成了“看到活猫”的人和“看到死猫”的人的叠加态．

关于什么构成测量这个问题至今仍然没有公认的解释，但在现实中研究者往往假设测量仪器对粒子构成了测量[11]，且得到了实验与计算的高度吻合．

不确定性原理

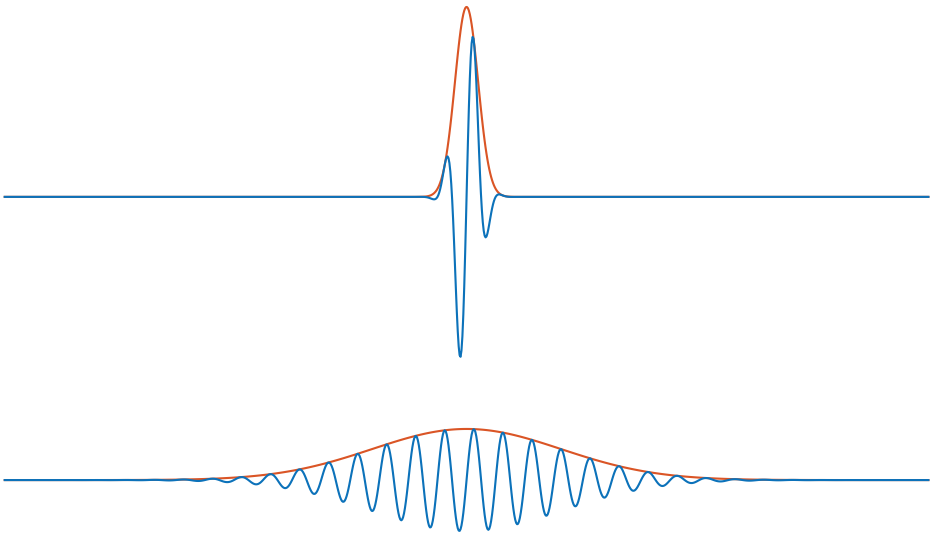


图 9: 波包越短，频率就变得越模糊，波包越长，频率就变得更精确．

首先注意**不确定原理**并不是量子力学的基本假设，而是从测量理论推导出的一些结论．最常见的一个不确定性原理说的是，我们无法既准确测量粒子的位置又准确测量粒子的动量，还有一种指的是无法既准确测量能量又准确测量时间．这里只讨论前者．

对不确定原理的一个理解是，我们准备许多处于相同状态的粒子，将其分

[11]这一方面也是由于多粒子波函数的计算极其复杂，目前我们还无法精确计算多于两个粒子的波函数的薛定谔方程，即使是用超级计算机做数值解．量子力学中所有涉及多粒子波函数的理论几乎都是用了不同程度的近似．

为两组，第一组测量位置，第二组测量动量．然后计算所有测得位置的不确定度（例如标准差）Δx，以及所有测得动量的不确定度 Δp．然后我们会发现，无论初始时粒子处于什么状态，二者的乘积 $\Delta x \Delta p$ 总会大于一个常数．

$$\Delta x \Delta p \geqslant \frac{\hbar}{2} \approx 5.3 \times 10^{-35}\ \mathrm{Js} \tag{3}$$

上文提到位置的本征态是一个 δ 函数，对应本征值为该波包的位置（坐标）．动量的本征态恰恰相反，是平面波．动量的本征值为平面波的空间频率乘以一个常数．如果粒子的波包很窄（但不是无限窄），那么要测量位置，我们只需要使用波包中心附近的一些 δ 函数就可以叠加出待测的波包，所以测量结果也只可能离波包中心较近．但在测量动量时，由于这个波包和平面波一点都不像，我们需要许多不同频率的平面波叠加才能得到这个波包（为什么无限长的平面波相加能得到有限长的波包？这是数学上一个非常有趣的结果，叫做傅里叶变换）．

另一种情况是，如果波包很长但不是无限长（想象绳子的一段连续上下抖了许多下才停下来），那么将波包看起来与某个频率的平面波十分相似．在测量位置时，我们需要将许多不同坐标的位置本征函数叠加得到波包，那测得的坐标就可能在很大的范围内出现．而测量动量时，我们只需要使用某个频率附近的一些平面波，所以测得的动量（与频率成正比）也都很接近中心频率对应的动量．

对不确定性原理的另一种理解来自测量理论中的波函数坍缩．假设我们先对处于某个状态的粒子准确地测量位置，那么根据测量理论，无论测到什么值，测量完后波函数都会坍缩为该位置的本征态，即 δ 函数．如果这时再马上测量动量，我们测到的就不是粒子原来状态的动量了，所以即使测到一个精确的动量值也并没有意义．反之如果我们先精确地测量某状态下的粒子的动量，那么测量完后波函数将会坍缩为某空间频率的平面波，再马上测量其位置，得到的值也同样没有意义了．

事实上对不确定性原理的这种理解对任何两个本征函数不相同的物理量都适用，因为当测量完第一个物理量后，波函数发生坍缩，再测量第二个物理量就没有意义了．但幸运的是有一些物理量有一组相同的本征函数（例如动量和动能的本征函数都是平面波），这样我们只需要测量其中一个，无需再次测量就可以直接计算出另一个物理量的值．

第四章

其他

天文学常识

太阳系模型

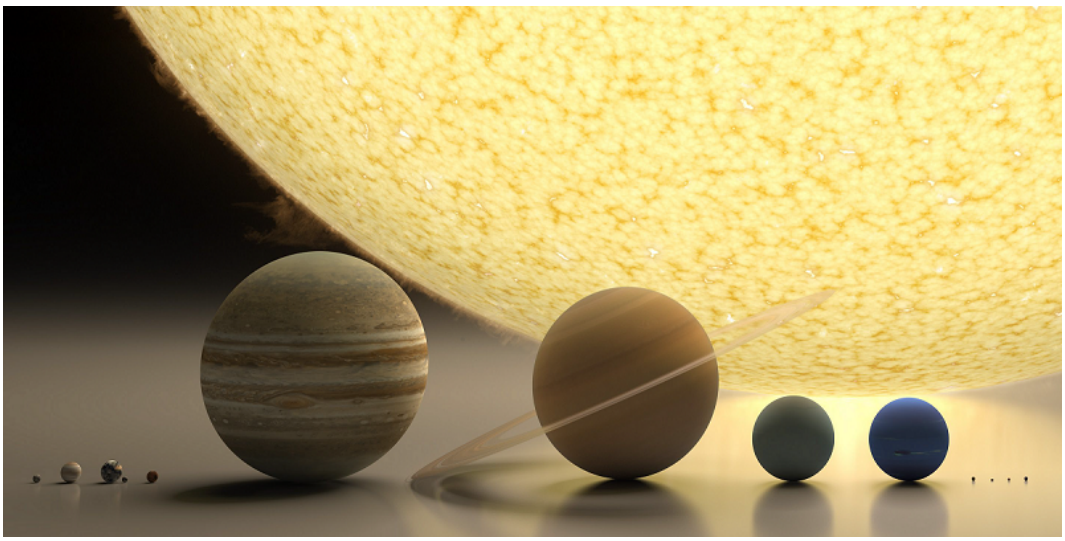


图 1: 真实比例下太阳与行星的大小（来自 pinterest.com），行星从左到右依次为水星，金星，地球（及月亮），火星，木星，土星，天王星，海王星，其他小行星

恒星和行星一个显著的区别就是恒星很大，而且会发光．太阳是太阳系中唯一一个恒星，其他天体都不发光，而是反射太阳的光．

关于太阳系模型一个误区是行星的尺寸和距离，几乎所有（实物的）太阳系模型的比例都是完全错误的，原因是如果按照真实比例制作模型，要么就是行星小得看不到，要么就是模型大得不切实际．我们这里来构建一个符合比例的太阳系模型，即用乒乓球来代表太阳，再按比例计算其他长度和距离，结果见表 1．

注意最后一栏中，即使按照模型的比例，最近的恒星离我们也有 1000 多公里，可见宇宙是相当空旷的．

表 1: 太阳系的一些数据

	真实长度	模型长度
太阳半径	6.96×10^5 km	20.0 mm
地球半径	6.37×10^3 km	0.183 mm
地球到太阳	1.46×10^8 km	4.20 m
月球半径	1.74×10^3 km	0.0500 mm
月球到地球	3.84×10^5 km	1.10 cm
木星半径	7.15×10^4 km	2.05 mm
木星到太阳	7.79×10^8 km	22.4 m
海王星到太阳	4.50×10^9 km	129 m
最近的恒星到太阳	3.99×10^{13} km	1.15×10^3 km



图 2: 星星的轨迹（来自 burro.case.edu），注意如果测量图中每条弧线的张角就可以计算曝光的时间

星空

除了太阳系中的天体，肉眼或望远镜能看到的其他星星都属于恒星（注意有一些亮点是星云，即星际尘埃而不是单个恒星），离太阳最近的恒星叫做

Alpha Centauri（4.22 光年）. 由于这些恒星的距离太远，即使用天文望远镜也看不到它们的行星. 恒星在空中的相对位置是几乎不变的（即使太阳系以惊人的速度绕星系中心旋转，但转一圈仍需要 2.3 亿年，所以与其他恒星的相对位置基本保持不变）.

那我们看到的星空是如何变化的呢？由于地球每 24 小时由西向东自转一圈，所以我们在地球上看到的星空由东向西围绕地球旋转. 长时间曝光的照片可以很好地展示星星的轨迹（图 2）. 如果站在北极点，那么星空的旋转中心会在正上方，如果站在赤道，旋转中心会在地平线上，且南北各有一个. 如果站在南极点，则旋转中心同样在正上方但旋转方向却与北极的相反.

一个常见的问题是为什么白天看不到星星和月亮. 这是因为白天阳光照亮了大气中的尘埃和云，相比之下星星的亮度太暗了所以肉眼不容易看到. 许多时候月亮也会在白天出现，只是因为阳光太亮不容易注意到.

季节与昼夜

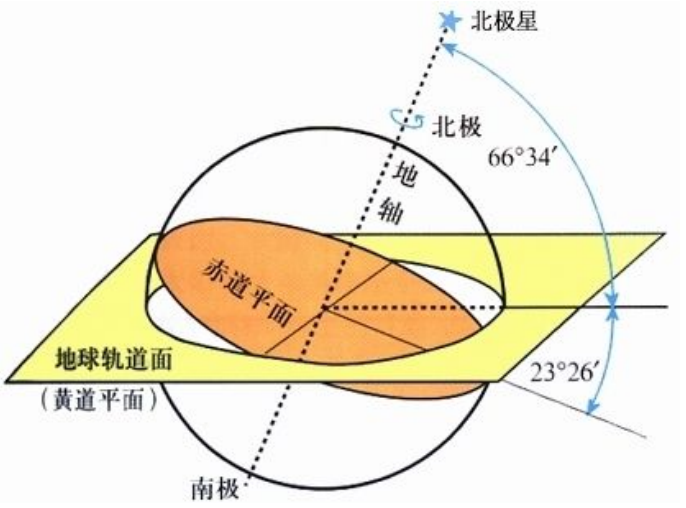


图 3: 黄赤交角示意图

如图 3，地球绕太阳公转轨道所在的平面叫**黄道平面**，而地轴与黄道平面有大约 23.5° 的倾角，所以当地球运动到太阳的一侧时，太阳直射到**北回归线**（北纬 23.5° 的纬线），这一天就叫做**夏至**，当地球绕太阳旋转 90° 后，太阳直射到赤道上，这一天就叫做**秋分**，而地球再公转 90° 到另一侧时，太阳直射到南回归线，这一天就是**冬至**，再公转 90°，阳光再次直射到赤道上，这一天就

是**春分**．注意这是对北半球而言，南半球相反．因此季节的变化本质上是太阳直射点的变化，与地球与太阳的距离关系不大，事实上在北半球的冬季地球到太阳的距离要小于夏季的距离．

回想第一节的太阳系模型，由于地球到太阳的距离远大于二者的半径，所以可以近似认为太阳照到地球的光都是平行的，这样，任何时刻只有半个地球被照亮，这半个地球就是白天，而另外半个是晚上．由于地轴的倾斜，北半球某条纬线以上的地区在夏天会被全部被照亮，这样无论地球自转多少，这些地区都是白天，这就形成了**极昼**．与此同时，南半球某条纬线以上的地区全部没有阳光，就形成了**极夜**．同理，在冬天（也就是南半球的夏天）南极附近会出现极昼而北极附近出现极夜．另外在赤道，**昼长**（白天的长度）和**夜长**（夜晚的长度）一年四季都相等．

天球坐标系

下面我们来更具体地描述星空．由于天体离我们足够遥远，来自天体的光线已近似于平行光，这时候我们已经失去了确定远近的能力（当然事实上我们有其他的办法确定天体的视向距离），而仅仅知道天体的方位，所有的天体都彷佛分布在一个球面上．假定围绕着我们有一个大圆球，它的半径比我们所能看到的最远的天体的距离还要大得多，这个假想的球体称为**天球**．观测者总是处于天球的中心，这是因为观测者的任意位移（包括地球的公转）都比天球半径小得多，可以忽略不计．天体和观测者眼睛连成的直线与天球面的交点，也就是天体在天球面上的投影，称为天体的**视位置**．

为了定量地表述天体在天球上的位置，我们需要建立球坐标系（正如我们描述地球上某个地点需要使用经度和纬度一样），最常用的两个天球坐标系是**地平坐标系**和**赤道坐标系**．

地平坐标系（图 5）是以观测者的位置为基础建立的，是最自然的坐标系．观测者立直站在地平面上，通过观测者和天球心的直线在天球上有两个交点，头顶 Z 称为**天顶**，脚下 Z' 称为**天底**，与该直线垂直的平面叫做天球地平面，与天球相交的大圆称为**地平圈**，也叫**真地平**．过天顶和北天极（地轴北极与天球的交点）的大圆称为**天子午圈**，天子午圈和真地平相交于 N 和 S 点，N 点靠近北天极，称为**南点**，其实和我们通常所说的东南西北的概念是一致的．天球上天体的坐标用**地平经度** A 和**地平纬度** h 来描述，具体的度量方法已在图

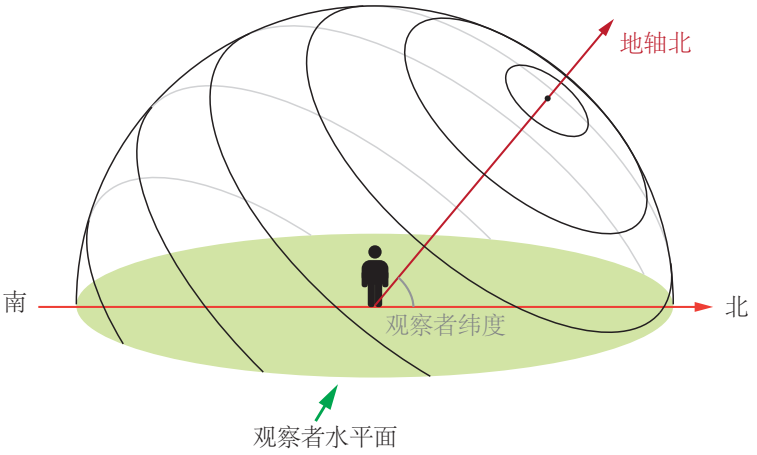


图 4: 地平坐标系

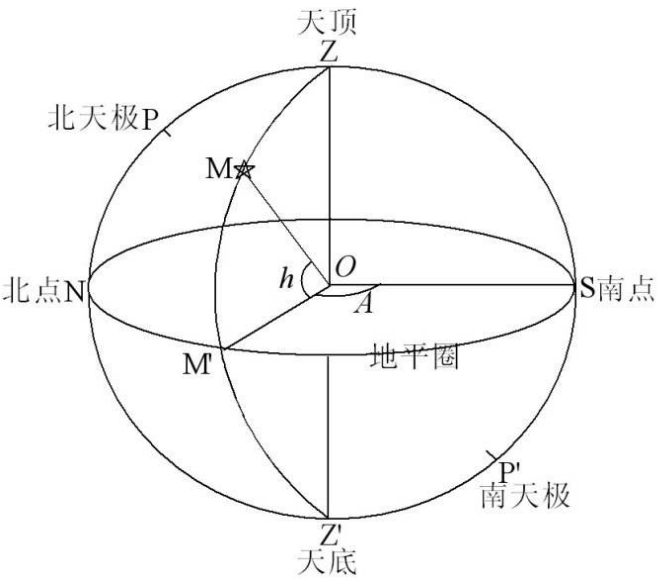


图 5: 地平坐标系

上标注，有时候也称**方位角**和**高度**．对于北天极，很容易证明它的地平高度恰好等于该地的纬度．

尽管地平坐标系在日常观星时很常用，但是同一个天体的坐标会随着观测地点的不同而不同，在实际的天文观测中，最常使用的坐标系称为**赤道坐标系**（图 6）．我们将地球的赤道面无限延展，与天球相交的大圆称为**天赤道**．地球公转轨道面延伸与天球相交的大圆称为**黄道**，黄道事实上是太阳相对于星空背景运动的轨迹，一年运动一圈，人们也常常把这种运动叫做太阳的周年视运

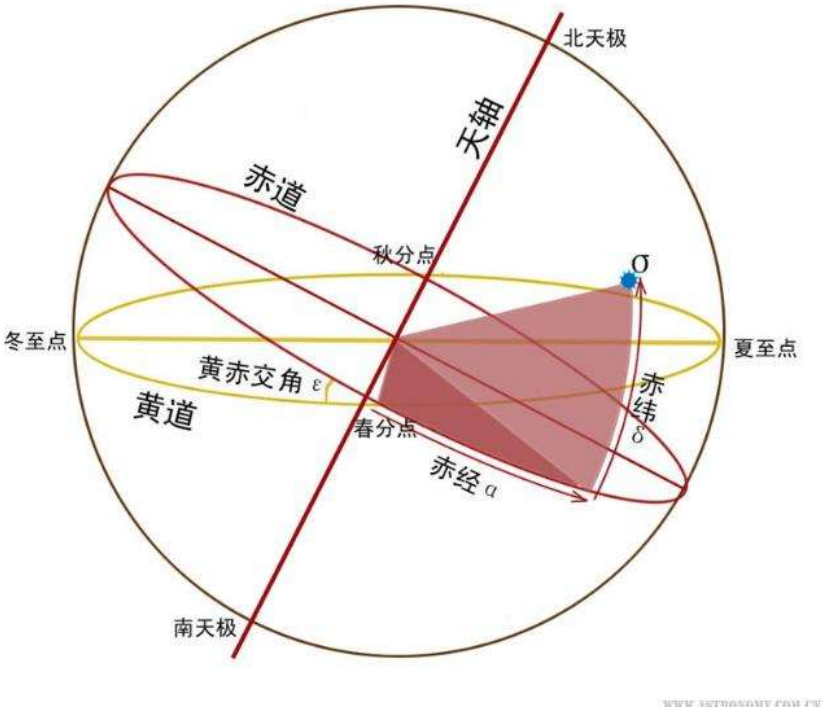


图 6: 赤道坐标系

动．天赤道与黄道有两个交点，分别称为**春分点**和**秋分点**，其中春分点是太阳视运动沿着黄道从天赤道以南穿过天赤道以北的交点，秋分点则相反．赤道坐标系中天体的坐标用**赤经**和**赤纬**来度量，具体的度量方法在图上标出．天体的赤经和赤纬是不会随观测地点、观测时间的改变而改变的．

天体的周日视运动

在地平坐标系中，人们站在地球上不觉得地球在自转，只看到天球带动所有天体绕着天轴在反方向旋转，只有北天极和南天极两点不动，这种运动称为天球的**周日视运动**．恒星因为周日视运动在天球上画出了一系列平行的圈（图 7），我们称天体高度角达到最大的时候为**上中天**，相反为下中天，事实上就是天体通过天子午圈的时候上/下中天．容易发现，有一些天体永远不会落下，有些天体也永远不会升起，这是由天体的赤纬和观测地点的纬度共同决定的．

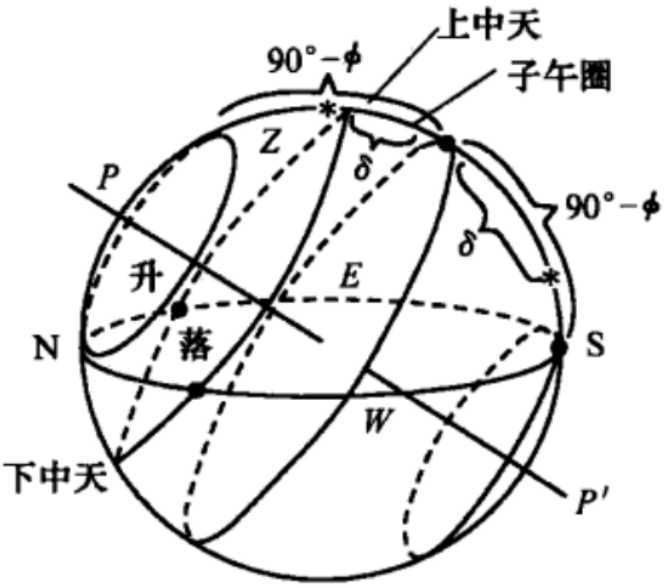


图 7: 周日视运动

时间的计量

- 这只是一个综述，需要链接到具体页面

秒的定义

[1]秒在历史上有过三种定义，16 世纪末开始出现以秒为最小单位的钟表. 当时秒的定义是将一个太阳日划分为 $24 \times 60 \times 60 = 86400$ 等分，并定义每等分一秒. 而一个**太阳日（solar day）**可以定义为地球上某处观察到太阳从一天的最高点到第二天最高点的时间间隔.

未完成：太阳日也有不同的定义

第一个定义的问题在于潮汐力的作用下，地球的自转速度并不恒定，1940 年左右石英钟的精确度已经超过了地球自转所定义的秒. 科学家发现用地球的回归年定义秒更精确，于是从 1956 年，科学家使用回归年来定义秒（定义略）.

[1]参考 Wikipedia 页面：Second，Leap Second，Coordinated Universal Time，Time Zone，Solar time，Julian Day.

从 1967，国际单位制[1970] 将秒重新定义为：一秒等于铯（Cs）原子 133 基态的超精细能级之间的跃迁辐射的电磁波周期的 9,192,631,770 倍．该定义一直沿用至今．本书中如无特别说明，“秒”一律指国际单位的定义．

协调世界时（UTC）

当前世界上最广泛的一套时间标准叫做**协调世界时（Coordinated Universal Time）**，注意作为英语和法语的一种妥协，它的缩写是 **UTC** 而不是 CUT（类似于国际标准单位是 SI 而不是 IS）．UTC 规定了日历，世界时区的划分，以及闰年，闰秒（见下文）等规则．

UTC 中的时区是不随季节改变的（见“夏令时”），所有的时区都以**格林尼治时间（Greenwich mean time，或 GMT）**为参考，格林尼治天文台位于英国，它所在的时区 GMT+0 都使用格林尼治天文台的当地时间．从该时区开始，向东的时区依次为 GMT+1，GMT+2，……，GMT+12；向西的时区依次为 GMT-1，GMT-2，……，GMT-12．中国通用的北京时间处于 GMT+8．虽然 UTC 是一个标准，但是大家往往把 UTC 和 GMT 当作同义词使用，例如北京时间也可以记为 UTC+8．

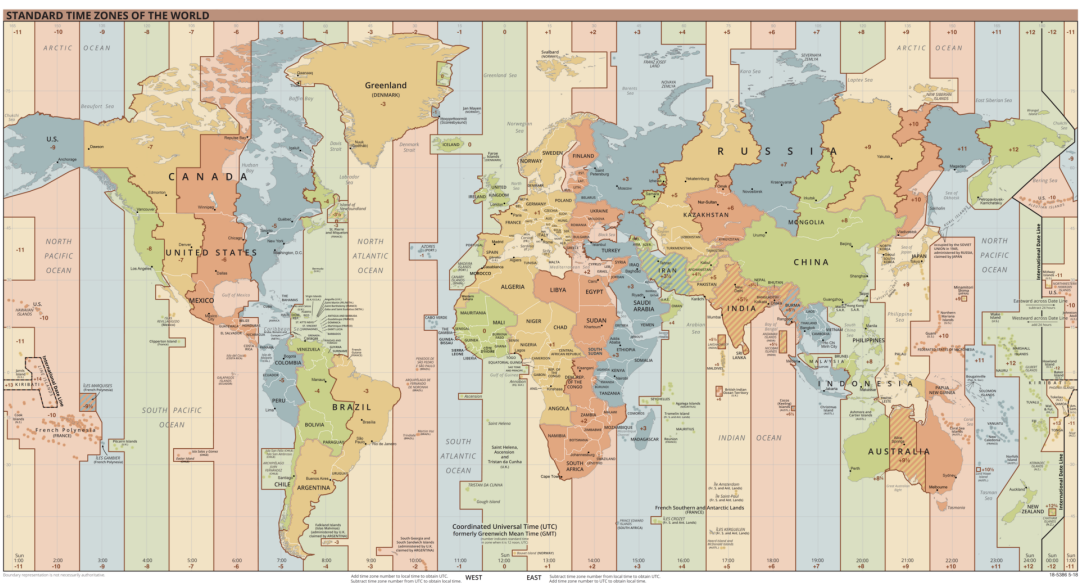


图 1: 世界时区划分（来自 Wikipedia）

回归年

回归年（tropical year）也叫**太阳年（solar year）**，是在地球上观察到太阳直射点经度变化一个周期的时间．回归年可以近似认为是地球绕太阳一周的时间．

一个回归年约为 365 天 5 小时 48 分 45 秒，比 365 天多出约 0.2422 天，约每四年多出一天，于是 UTC 规定**闰年（leap year）**为：年数不为 100 的倍数且能被 4 整除的年，以及年数能被 400 整除的年．如果一年不是闰年，就是平年，闰年平年比额外增加一个**闰日（leap day）**即 2 月 29 日．闰年有一个记忆口诀：四年一润，百年不润，四百年再润．

夏令时

夏令时也称为**日光节约时间（daylight saving time）夏时制（summer time）**．在一些维度较高的国家和地区由于夏天太阳升起更早，当地政府出于节能等考虑会在协调世界时的基础上把时间调快一个小时，即时区 +1．例如图 1 中美国[2]纽约的时区是 GMT-5，但夏天则会变为 GMT-4．

闰秒

由于地球日在潮汐力的作用下逐渐变长，目前一个地球约为 86400.002 秒（国际单位的定义），而我们使用的钟表只有 86400 秒．UTC 规定钟表的一秒需要符合国际单位，所以规定在每个月的最后如果它和 UT1 时间（暂不介绍）之差超过 0.6 秒，就插入一个闰秒．到目前为止，所有的闰秒都只在 6 月 30 日和 12 月 31 日最后插入过，例如截止到 2020 年底，最近的 3 次闰秒分别在 2016/12，2015/6，2012/6 的最后插入过．当闰秒发生时，理论上人们需要把钟表暂停一秒（智能手机等设备会自动调整）．

儒略日

（未完成）

[2]美国除了亚利桑那和夏威夷州都采用夏令时．

Unix 时间戳

在计算机编程中，为了方便计算时间差会使用 Unix 时间（或时间戳，time stemp），也称为 Epoch 时间或者 POSIX 时间．Unix 时间是一个数（可以为负），定义为从 1970-1-1 的 0 时到某一时刻的秒数（国际单位的定义）减去这段时间内所有插入的闰秒．所以每当闰秒被插入时，Unix 时间同样需要暂停一秒．所以如果给 Unix 时间加上小数点后面的部分，除了闰秒外的任意时刻都可以用其 Unix 时间戳唯一表示．闰秒使得 Unix 时间变得不完美，所以在一些编程语言或程序库中，也定义了包含闰秒的 Unix 时间．

时间的计量 2

预备知识 时间的计量[80]

世界时

通过遥远天体的观测等方法决定一个参考系，即 **ICRF（International Celestial Reference Frame）**，决定地球在 ICRF 中的转角**地球旋转角（Earth Rotation Angle，ERA）** θ，并以此定义 **UT1 儒略日**为

$$\theta = 2\pi(0.7790572732640 + 1.00273781191135448T_u) \tag{1}$$

其中 T_u = UT1 儒略日 $-$ 2451545.0，且 2.7378×10^{-3} 约等于 $1/365.25$ 即一回归年的太阳日个数．

把 UT1 儒略日的小数部分除以 86400（$24 \times 60 \times 60$），就可以定义 UT1 时间，即上一个 UT1 正午经过的 UT1 秒数（注意儒略日是从中午开始计算的）．注意这里的 UT1 秒是由式 1 定义的而不符合国际单位．由于地球自转速度会在潮汐力的作用下发生改变，所以 UT1 中一秒的长度也会在几天到几十天的周期内变化．UT2 就是在 UT1 的基础上把周期性潮汐力对地球自转的影响过滤掉了．测量得出，现在地球自转一圈的平均用时（UT2）约比 100 年前的 1 天慢 1.7 毫秒．

两种太阳时

地方视太阳时（local apparent solar time）可以定义为一个理想日晷（把太阳看作一个点光源）显示的时间．视太阳时的问题在于它是不均匀的而是随季节周期性变化．太阳相对于地球的行经角速度由两部分组成，一个是上文的 ERA 角速度，另一个是在 ICRF 参考系中太阳直射点相对地轴的角速度．前者在变化非常缓慢几乎可以忽略不记，而后者会发生显著的季节性变化．改变化由两个因素决定：一是地球的公转旋转轨道是椭圆的，当地球远离太阳时公转角速度变慢，靠近太阳时角速度变快；二是由于赤道平面相对于公转平面（黄道平面）是倾斜的，当太阳直射点纬度大时行经角速度快（行纬角速度慢），接近赤道时行经角速度慢（行纬角速度快）．

如果把一年中的视太阳时做一个线性拟合，就得到了**平太阳时（mean solar time）**．平太阳时和 UT1 几乎相等（除了每百年增加 1.7ms 误差）．当前的平太阳日约为国际单位的 86400.002s．视太阳时和平太阳时之间的差距叫做**均时差（equation of time）**，由图 1 可得在一月中旬慢 15 分钟左右，而在 11 月初则快 17 分钟左右．

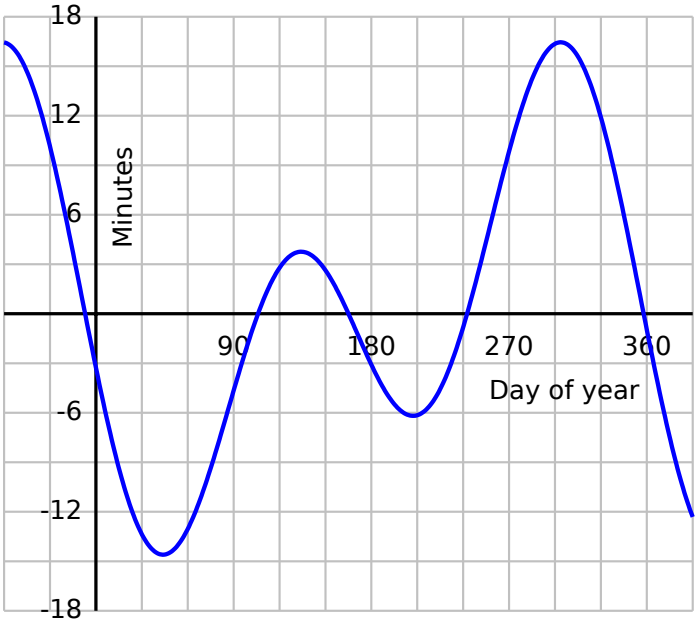


图 1: 均时差，横轴表示日期，纵轴为分钟（来自 Wikipedia）

加密货币简介

本文试图在尽量不涉及计算机专业知识的前提下讲解加密货币的意义和原理．注意为了方便理解，一些细节并不十分严谨．

黄金

相对于数字货币，各国政府发行的货币称为法币．法币的出现最早在 1694 年的英国，在此之前的漫长岁月中，人类基本在使用一些天然的物质充当货币，如早期的贝壳，直到黄金．直到今天，黄金仍然承担着货币的角色，各个国家的国库中依然存有大量的黄金作为法币的信用保证．

黄金能作为等价物是大家的共识，但注意它没有任何政府背书或发行，且总量基本固定．那么它作为货币时，价值体现在哪里呢？简单来说就体现在它的记账功能．如果直接以商品交换商品，效率非常低，因为很难配对到合适的人．黄金或者一般等价物的使用，相当于产生了不可篡改账本，安全地记录每个人所拥有的资源．

黄金使交易更便捷，从而对古代经济发展起到了重大的助推作用．为什么不能用纸和笔记账？因为不安全——陌生人难以辨认真伪，很可能造假，也可能被大量增发．可见要让大家都信任一个中立的记账系统，它必须具有以下特点

1. **安全性**：不能伪造，不能被篡改，不能被发行者大量增发，不会变质．黄金在这点上的优势无疑是巨大的，黄金几乎无法人工合成（只有核反应才能合成极少量），金矿勘探和开采的难度非常大，在自然条件下可以稳定存放．但有一个缺点就是鉴别真伪成本较高，需要彻底熔化或者使用现代仪器．

2. **便携**：由于黄金总量相对较少，所以用作货币单价很高，只需携带少量即可满足日常需求．但这比起当代电子设备还是不够方便．

3. **转账方便**：黄金交易在古代的确是算非常方便，但在比起比起电子转账同样差很多．

综上，黄金的价值就在于它利用其出色的物理特性，在历史上相当长的时间里充当账本的角色，极大地促进了人类社会的早期发展．需要指出的是这个“账本”只记录每人的余额而不记录转账历史，这是又一缺点．

数字货币

在互联网高度发达的今天，我们是否可以在互联网上建立一个像黄金一样数字账本呢？只要把以上三点实现得更好，它就具有比黄金更大的利用价值．这看起来容易，实则非常困难，要付出巨大的代价．我们的银行网银，支付软件，游戏币，充值卡等，本质上是政府或者公司发行的，政府（或公司）对这些系统的全部服务器和数据库有绝对的控制权．理论上他们可以对其进行任意修改，且其他人几乎不可能从外部破解．政府为了刺激经济会经常增发货币，但也有可能导致通货膨胀．例如美国在 2020-2021 年因新冠疫情大量增发美元使其他国家的美元储备贬值；发行者有能力冻结或没收账户中的财产[3]；在一些小国家也可能会发生政权不稳定使法币失去信用．所以只从以上的第 1 点来看，所有被政府或公司控制的中心化数字记账方式都远不如黄金．所以即使第 2 和第 3 点做得再好也还是有缺憾．

那么创造数字黄金的难点就在于如何使这个系统像黄金一样不受任何小团体的控制，即**去中心化（decentralize）**．只有这样才能使第 1 点的安全性成立．这就要求没有个体或者少部分人能够篡改（或毁灭）该网络账本．直到中本聪在 2009 年提出区块链的概念并发行了比特币后，这种机制才慢慢成为可能．比特币网络是一个去中心化网络，也就是没有中心服务器，人人都可以搭建服务器并接入该网络，共同完成记账．账本在每个服务器中都有一个完整的备份，这包含有史以来每一个账户发生的每一笔交易记录，根据这些交易记录可以得到每个账户的余额．

用户可以匿名地注册任意多的账户，注册时无需提供任何身份信息，每个账户都有一个独特的公开的地址，称为**公钥（public key）**，以及与之匹配的密码，称为**私钥（private key）**．转账的操作和微信等类似，使用一个账户和密码就可以联网后请求比特币服务器把余额转到任意指定的地址[4]．只有一笔交易

[3] 设想若两国开战，理论上一个国家可以冻结它在另一个国家的法币．

[4] 服务器并不储存密码，但是可以轻易地验证转账请求中用于签名的密码是否和账户的地址匹配．转账请求中同样不包含密码，只包含用密码生成的数字签名，这涉及到密码学中的非对称加密，超出了本文范围．

取得网络中大部分服务器的**共识（concensus）**，确认一致后才能生效．由于任何人都可以搭建服务器进行记账，所以所有的交易记录也都是只能是公开的，但由于账户是匿名注册的且可以注册任意多个，甚至每笔交易都使用一个不同的账号，所以用户的隐私仍然可以得到保障．

安全性的代价

既然任何人都可以参与到比特币网络中，那么将其恶意破坏也会变得比中心化网络更容易．对于中心化服务器，控制者只需要设置一个密码就可以防止恶意篡改，而如果有人要恶意破坏比特币网络，他只需要购买大量服务器，在网络中的数量达到多数就可以通过所谓的“共识”来做假账，例如花一笔钱然后事后抹掉记录再花一次（即**双花，double spending**）[5]．所以为了提高整个网络的安全性，我们必须使得“买下大部分服务器”这件事变得异常艰难．即大大提高每台服务器所需的成本．比特币网络的做法就是把记账权和服务器的**算力**（计算能力）挂钩，服务器必须完成一些很难（但没有意义）的数学题才能获取更多的记账权．理想状况下是整个网络的总算力极高，且服务器的所有权非常分散，使得少部分人几乎不可能取得该网络的控制权从而对数据进行恶意篡改．这些无意义的计算最终会消耗大量资源，但这却是维持去中心化账本的安全性的必要代价．在 2016 年，据估计比特币网络的总算力已经达到世界前 500 的超级计算机集群的总和．这使得比特币比其他的货币更为安全．

在另一些加密货币中，为了不过多消能源和硬件，也有采用把服务器所有者的加密货币余额作为比重获取记账权．比特币的方法叫做**工作证明（proof of power）**，后者叫做**持有量证明（proof of stake）**．但无论如何，为了使网络具有安全性，网络中的服务器必须以某种珍贵的资源作为记账权的分配标准．

挖矿奖励

既然服务器需要那么多资源，为什么还有人愿意搭建呢？为了补偿这些维护比特币网络安全性的服务器所有者，比特币协议规定每十分钟就会对成功解出数学题并获得记账权的账户进行奖励，所以人们就形象地把提供服务器算力并获取奖励的行为称为**挖矿（mining）**，挖矿的人称为**矿工（miner）**[6]．而所

[5]但他仍然不能把其他任意账户中的钱转给自己，因为他没有其他账户的密码（密钥）．

[6]miner 也可以指挖矿所用的软件．

有的比特币都是通过挖矿产生的．比特币协议规定每 4 年挖矿的收益就会减半，从而使其总量保持不变．除了挖矿获得奖励外，矿工还会获得一些交易手续费．当所有比特币都被挖完后，理想情况下此时比特币的价格已经昂贵到足以让交易费支持所有矿工的开支，这样该系统就能自给自足地一直运行下去．

第二部分

微积分

第一章

基础

公理系统

公理，定义和定理

一个**命题（proposition）**就是一个陈述句．我们讨论问题时会大量使用命题来作描述，这些命题有的是成立的，称为**真命题（true proposition）**；反之，有些命题不成立，称为**假命题（false proposition）**．比如说，我们通常认为命题“水是剧毒的”是一个假命题．但是这个命题是对于我们人类的生理学而言才假的，对于某些生物，水确实是致命毒药．也就是说，一个命题是真是假，要看该命题被放在什么样的环境下讨论的；讨论问题的环境，被称为**理论（theory）**．在理论限定的框架下讨论问题，本质上是在判断一个个命题在给定框架下的真假性，判断这一动作被称为**证明（prove）**．特别地，在汉语中，有时也会使用“证伪”一词，来简略表达“证明为假”的意思．

我们是不可能证明所有的命题的，所以任何理论必须有一个出发点，也就是一些基础命题．这些基础命题本身不可证明，而是被默认成立；它们决定了理论的样貌，理论中一切其它命题都是由这些命题根据逻辑推演得到的．这样的基础命题被称为一个理论的**公理（axiom）**．合格的公理系统中，公理之间彼此不能矛盾，比如命题“1 是整数”和“1 不是整数”不能作为同一个公理系统的两条公理．

有了公理系统以后，我们还需要明确所讨论的对象是什么．比如我用了皮亚诺公理来定义小学四则运算，那么为了讨论“$1+1$ 等于几”这样的问题，我首先需要明确“1”和“$+$”具体指什么．用来明确概念的命题，被称为**定义（definition）**．如果说公理系统是创建了一个宇宙的基础参数的话，那么定义就是在给这个宇宙里已经自然存在的事物进行命名，这样才能讨论这些事物．当我们说“把……称为”时，也是一种下定义的方式．

最后，任何一个理论的绝大部分内容都是在使用基础命题来进行推演，看哪些命题为**真**．**真命题**被称为**定理（theorem）**．有时候，根据定理作用的不同，我们也可能称其中一些为**引理（lemma）**、**推论（corollary）**等．所有定理加在一起就构成了整个理论．

不同的公理系统可能推演出相同的命题，也可能推演出彼此矛盾的命题，

更可能存在一些无法判断是否成立的命题．一个公理系统中所无法判断是否成立的命题，就叫做独立于这个公理系统的命题．这就好比我在描述一个多面体，使用公理“这个多面体有六个相同的面，且六个面彼此要么垂直要么平行”，那么就可以推演出定理“六个面都是正方形”，但是命题“这个多面体是蓝色的”就是独立于所给公理的命题，无法被证明或证伪．

如果两个公理系统能够推演出完全一样的命题（定理），那么这两个公理系统就是等价的．如果公理系统 A 能够推演出公理系统 B 的一切定理，但是 B 不能推出 A 中的一切定理，即 A 能推演出的一些定理实际上是独立于 B 的命题，那么可以认为是公理系统 A 包含了公理系统 B．**在阅读本段话时，请注意命题和定理的区别：定理是在给定公理体系下的真命题．**

特别地，**哥德尔第一不完备性定理（Gödel's first incompleteness theorem）**说明，一切含有小学算术的公理系统中，总存在一些定理，如果它们为真，那么永远无法证明它们为真．这意味着系统里有一些真命题，但是我们没有任何能力意识到它们是真命题．由于还存在很多很难证明的命题，在一个命题被证明前，我们很难知道它究竟是假命题、可以证明的真命题，还是永远证不出来的真命题．前两类假以时日都可以通过证明得知命题的真假，但最后一类永远都会是吊人胃口的难题，因为你永远都没法知道它到底能不能被证出来．

集合

预备知识 公理系统[91]

集合

对于物理学习而言，集合论没必要从公理角度来严格理解，所以在此不会给出用于划定集合论讨论范围的公理系统，而是**朴素集合论（naive set theory）**的解释，即比较接近自然语言的表达方式．

集合（set）是由**元素（element）**组成的．任何事物和概念都可以成为元素，任何不同的元素都可以放在一起，构成一个集合．可以说，如果我们划定一个讨论的范围，那么这个范围就是一个集合，范围涉及到的事物和概念就是这个集合当中的元素．公理系统的作用，也就是在所有可能讨论的话题所构成

的集合中，限定一个子集作为讨论范围．对于集合 A，定义 $|A|$ 为 A 中元素的数量，称为集合 A 的**基数（cardinal number）**或**势（cardinality）**.

表达一个集合的方式有多种．最简单的方式是列出所有集合中的元素，在数学中规定的语法规范是用大括号“{}”来列举集合中的一切元素，以逗号隔开彼此. 比如，$\{猪, 牛, 狗, 羊, 猫\}$ 是了一个具有五个元素的集合，$\{1, 2, 3, 4, \dots\}$ 则是全体正整数的集合．第二个例子并没有显然地列举出所有正整数，只是用省略号表达了这个意思；也就是说，表达一个集合的方式并没有死板的规定，只要能让读者理解就可以了.

另一种常见的表达集合的方式是确定一个规则，语法规范是“$\{x|x$需要满足的条件$\}$”. 比如全体正整数的集合，也可以写为 $\{x|x$ 是一个正整数$\}$. 如果有多个条件，也可以列在一起，比如全体正整数的集合：$\{x|x$是一个正数，且 x 是一个整数$\}$. 特别地，如果某条规则是“x 属于某集合”，我们通常会将这个条件写到单竖线的前面，如全体正整数的集合：$\{x \in \mathbb{Z}|x$ 是一个正数$\}$．这里，$\in$ 是一个简写的符号，$A \in B$ 等于说“A 是 B 的元素”.

有一个特殊的集合，它不含有任何元素，被称为**空集（empty set）**，记作 $\varnothing$，有时也写作 ϕ．$\varnothing$ 是一切集合的子集.

元素和子集

若 a 是集合 A 的元素，我们就说 a **属于（in）** A，记为 $a \in A$ 或者 $A \ni a$.

如果集合 A 的元素都是集合 B 的元素，那么称 A 是 B 的**子集（subset）**，本书中记为 $A \subset B$ 或者 $A \subseteq B$，也可以反过来写为 $B \supset A$ 和 $B \supseteq A$．一切集合都是自身的子集．如果 A 是 B 的子集但又和 B 不同，也就是说 A 没有包含 B 的所有元素，那么称 A 是 B 的**真子集（proper subset）**，本书中记为 $A \subsetneq B$，$A \subsetneqq B$，$B \supsetneqq A$ 或 $B \supsetneq A$[1]．B 的真子集一定是 B 的子集，但是 B 本身是 B 的子集而非真子集.

注意区分这两个情况，前一个情况中 A 是 B 的元素，后一个情况中 A 是 B 的子集．另外，集合本身也可以是别的集合的元素，元素的概念没有限定，任何事物和概念都可以成为元素，包括集合．不过，集合的集合我们一般用**集族（family of sets）**或者简单地用**族（family）**来称呼.

[1]有的地方会用 $\subset$ 来表示“真子集”，和我们这里的定义矛盾；本书中 $\subset$ 就表示子集，但是尽量避免使用这个符号，以尽力避免读者的混淆.

集合运算

集合间可以互相操作，生成新的集合，这种操作被称为集合间的**运算（operation）**.

$\cap$ 表示两个集合的交，意思是将两个集合中共有的元素提取出来，组成一个新的集合. 比如说，$\mathbb{N}^+$ 表示全体自然数的集合，$\mathbb{R}^+$ 表示全体正实数的集合，$\mathbb{Z}$ 表示全体整数的集合，那么显然我们可以有 $\mathbb{N}^+ = \mathbb{R}^+ \cap \mathbb{Z}$. 多个集合 A_i 的交集，可以写为 $A_0 \cap A_1 \cap A_2 \cap A_3 \ldots$，也可以用一个大号的交集符号简记为 $\bigcap A_i$，表示“所有形式为 A_i 的集合的交集”.

类似地，将两个集合中都有的元素提取出来，组成一个新的集合的操作，被称为集合的并，用符号 $\cup$ 和 $\bigcup$ 表示. 注意，如果两个元素中有相同元素，那么在并集中这个元素只出现一次. 这是因为我们关心的是每个元素是否出现在集合中，计算集合元素数量时也不会重复计算同一个元素. 这是一个并集的例子：

$$\{猪, 牛, 狗, 羊, 猫\} \cup \mathbb{N}^+ = \{猪, 狗, 猫, 牛, 羊, 1, 2, 3, 4, \ldots\} \tag{1}$$

注意，列举时元素的顺序也不影响集合的本质.

对于集合 A 和 B，$A \backslash B$ 或者 $A - B$ 表示他们的**差集**. 差集所包含的元素是 A 中全体元素中减去 B 中元素，如果 B 还含有 A 中所没有的元素，那么这部分元素可以忽略掉. 例如，如果令 $A = \{0, 1, 2, 3\}$，$B = \{2, 3, 4\}$，那么 $A - B = \{0, 1\}$.

如果我们划定了一个讨论范围，被讨论的元素构成的集合称为 U，那么 U 中的任意一个子集 A 都可以进行**取补集**运算，得到 $A^C = U - A$，称为 A 的**补集（complement）**.

笛卡尔积

对于集合 A 和 B，$A \times B$ 表示集合间的**笛卡尔积（Cartesian product）**，得到一个新的集合. $A \times B$ 中的元素表示为 (a, b)，其中 $a \in A, b \in B$. 用集合论的术语表达就是

$$A \times B = \{(a, b) | a \in A, b \in B\} \tag{2}$$

例如，如果令 $A=\{0,1,2,3\}, B=\{a,b,c\}$，那么

$$A\times B=\{(0,a),(0,b),(0,c),(1,a),(1,b),(1,c),(2,a),(2,b),(2,c),(3,a),(3,b),(3,c)\} \tag{3}$$

可以看到，集合 A 有 4 个元素，集合 B 有 3 个元素，而集合 $A\times B$ 有 $3\times 4=12$ 个元素.

集合运算的规律

de Morgan 公式（de Morgan's laws）：$(\bigcup A_i)^C=\bigcap A_i^C$，$(\bigcap A_i)^C=\bigcup A_i^C$. 用自然语言表达，就是：交集的补等于补集的并，并集的补等于补集的交.

分配律（distribution laws）：$A\cup(\bigcap B_i)=\bigcap(A\cup B_i)$，$A\cap(\bigcup B_i)=\bigcup(A\cap B_i)$. 用自然语言表达，就是：并集运算对交集运算满足分配律，交集运算也对并集运算满足分配律，就像小学所学的乘法对加法满足分配律一样.

拓展

目前为止我们只介绍了集合之间的运算以及集合中元素的数目，但没有对集合本身建立一个结构. 从基数的角度来看，只要两个集合之间存在一一对应，那么就可以把它们看作同一个集合，因为在我们的讨论范围里它们都是完全一致的（只讨论了基数）. 但是这样很无聊，没有太多研究的意义，因此数学家们开始给集合赋予各种各样的结构，进一步细分集合的分类，由此诞生了拓扑学、代数学等分支. 现代数学的绝大部分分支都是通过给集合赋予结构来描述的，可以说集合论是现代数学的基石，几乎每一条数学定理都是集合论的定理.

整数

预备知识 逻辑量词，二元关系[858]

整数的概念是大家所熟悉的. 严格的数学中对于整数的定义过于抽象，主要是为了逻辑基础的严谨性；严格定义依然是建立在我们熟知的整数概念上的，所以在物理学习中没必要深入到整数的定义上，读者只需要按照通常的理解

方式来认识整数就可以．我们把整数构成的集合称为**整数集（Set of Integers）**，记为 $\mathbb{Z}$．

整数通常涉及的运算有加法和乘法．加法的逆运算被称为减法，且减法在整数集上是封闭的，即两个整数相减仍为整数；乘法的逆运算被称为除法，除法在整数集上就不封闭，例如 $2/3 \notin \mathbb{Z}$，尽管 2 和 3 都属于 $\mathbb{Z}$．

在中小学中我们学过带余除法，用现代数学语言可以记成如下形式：当被除数为 a，除数为 b，**余数**为 c 的时候，存在一个整数 k，使得 $a = kb + c$．作为练习，用逻辑量词来表达时，这句话可以写为：当被除数为 a，除数为 b，余数为 c 的时候，$\exists k \in \mathbb{Z}, \text{s.t.} a = kb + c$．在**群论（group theory）**[534]、**环论（ring theory）**和**域论（field theory）**中，我们会遇到大量基于模运算概念的运算．本词条中，我们会使用钟表模型来阐释如何进行模运算．

将除数固定为 b，那么如果两个整数 a_1 和 a_2 除以 b 所得到的余数相同，我们就称 a_1 和 a_2 **模** b **同余**（$a_1 \equiv a_2 \bmod b$）．这里**模**的意思类似于“除以”，**同余**的意思是“余数相同”．在这种情况下，我们也把 a_1 和 a_2 称为彼此的**模** b **的同余数（congruence mod** b**）**．为了简化表达，当默认了 b 的取值以后，我们也可以把 $\bmod b$ 省略，简单地称 a_1 和 a_2 **同余（congruent）**，并记为 $a_1 \equiv a_2$．

由词条**二元关系**[858] 中对**等价关系**的讨论可以推知，给定整数 b 以后，模 b 同余的关系是一个整数集上的等价关系．也就是说，我们可以利用 b 把整数集划分为若干个等价类，称作**整数集模** b **的同余类（congruence class）**．由等式 $a = kb + c$ 可知，a 就在模 b 余 c 的同余类中．在词条**二元关系**[858] 中所讨论的“两数的差是 3 的倍数”这一关系，其实就是一个模 3 同余的关系．

钟表和模算数

钟表的模型可以帮助你直观地理解模运算，以及初步感受“集合间的运算”是什么样的．

我们所熟知的加法、减法和乘法运算是在一个无穷集合上讨论的，即整数集．如果把“等于”的概念替换为“模同余”的概念，我们就可以在有限集合上进行加法、减法和乘法运算．最常见的例子，是钟表上整点的计算，这是一个模 12 同余的运算．

模运算的概念是，给定用来作除数的 b 以后，将模 b 同余的整数都看成同一个．在钟表的例子里，$b = 12$，那么我们把 $0, 12, 24, 72$ 等都看成同一个元素，

$3, 15, 27, 147$ 等也都看成同一个元素．判断整数 a_1 和 a_2 是否模 12 同余的方法很简单，就是看 $a_1 - a_2$ 是否是 12 的倍数．这样，我们可以**把整数轴卷起来**，让所有彼此同余的整数都重合起来，此时 12 点就是 0 点，14 点就是 2 点，28 点就是 4 点．为了方便讨论，我们只取 $0, 1, 2, \cdots, 11$ 分别作为 12 个同余类的代表，然后进行运算．

我们要讨论的是，从两个同余类中随便取一个数字，将它们进行加法，减法或者乘法运算后，结果落入哪一个同余类．在这样的运算规则下，用上一段所规定的代表元来进行讨论的话，我们可以得到以下运算的例子：

$$\begin{aligned} &1+1\equiv 2 \quad 4+8\equiv 0 \quad 11+8\equiv 7 \\ &3-2\equiv 1 \quad 5-7\equiv 10 \\ &2\times 3\equiv 6 \quad 3\times 4\equiv 0 \quad 4\times 7\equiv 4 \end{aligned} \tag{1}$$

模运算有一个美妙的性质．如果 a_1 和 a_2 同余，那么我们可以将它们的关系表达为 $a_1 = a_2 + 12k$，其中 k 是唯一存在的整数．利用这个关系容易证明，对于任意的整数 a_3，我们都有 $a_1 + a_3 \equiv a_2 + a_3$，$a_1 - a_3 \equiv a_2 - a_3$ 以及 $a_1 \times a_3 \equiv a_2 \times a_3$．这三个式子实际上说明了，在进行同余类的加法、减法和乘法的时候，无论取谁作为代表元素，结果都还是在同一个同余类里．因此我们才可以放心地把同余类的运算看作是任意代表元之间的运算，比如钟表整点的计算．

在通常的钟表中，我们把整数轴卷成周长为 12 的钟表．我们也可以把整数轴按其它周长来卷曲．当把整数轴卷曲成周长为 b 的钟表时，在钟表上进行的加减和乘法运算就是数学家们所讨论的模 b 的加减和乘法运算．

一般地，我们把一个如上定义的、含 n 个钟点、定义了加法和乘法的结构，记为 $\mathbb{Z}_n$．这是“环”的一个例子，参见词条**环**[569] 以及其后的内容．

素数

根据小学所学定义，**素数（prime numbers）**是指那些除了 1 和自身以外没有别的因子的整数．

在代数学中，整数集是**交换环（commutative ring）**的一个特例，而素数是更广义的**交换环的素元素（prime element of a commutative ring）**的一个特例．根据交换环中素元素的定义，整数中的素数 p 是满足如下需求的整数：如

果 p 能整除 ab，其中 a 和 b 是任意满足条件的整数组合，那么 p 必能整除 a 和 b 中的一个．用逻辑量词来简洁地表达，就是：$\forall a,b\in\mathbb{Z}$，如果 $p|ab$，那么要么 $p|a$，要么 $p|b$.

映射

预备知识 集合[92]

给定集合 A 和 B，我们可以假想从 A 中**每一个**元素上拉**一根**有方向的线连接到 B 中的一个元素，这些线的连接方式就被称为一个从 A 到 B 的**映射（mapping）**，也叫**算符（operator）**．将这个映射记为 f，A 叫做 f 的**定义域（domain）**，B 叫做**到达域（codomain）**[2]，B 中被线连接到的元素的集合叫做 f 的**值域（range）**[3]或**像（image）**，记为 $f(A)$.

我们一般将“f 是从 A 到 B 的映射”记为

$$f:A\to B \tag{1}$$

也就是说从 A 的元素上拉线到 B 的元素上．有时候，为了表示映射的定义域 A 或到达域 B 是另一个集合的 S 的子集，我们也会将映射记为

$$f:A\subseteq S\to B \quad 或 \quad f:A\to B\subseteq S \tag{2}$$

注意映射是有方向区分的，A 中每个元素都**有且只有**一根线拉出去，但是 B 中的元素可以同时被一根或多根线连接，也可以没有连接（即不在值域中）.

映射的类型

定义 1

如果映射 $f:A\to B$ 中每个 B 中元素只被 1 根或者 0 根线连接，那么称 f 是一个单射（**injection**）．如果 $f:A\to B$ 中每个 B 中元素都被至少 1 根线连接，那么称 f 是一个满射（**surjection**）．如果 f 既是单射又是满射，那么称它为一个双射（**bijection**），或者叫一一对应（**one-to-one correspondence**）.

[2]也叫陪域、上域、目标集（target set）

[3]值域在一些文献中指的是到达域.

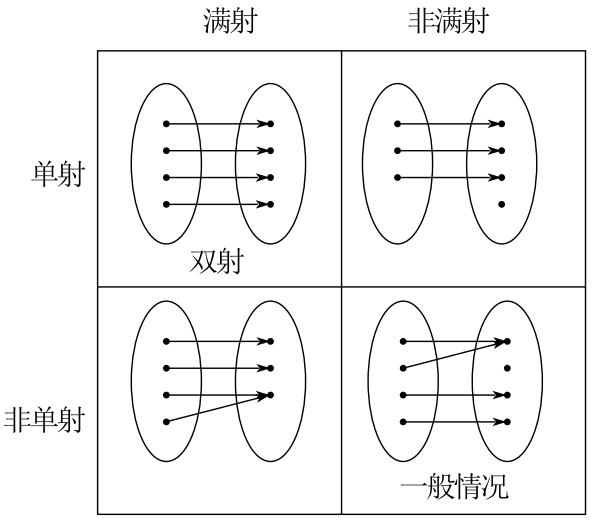


图 1: 映射的分类

如果 $f : A \to B$ 是一个双射，那么 A 中每一个元素都唯一地连接到 B 中某一个元素，并且 B 中每一个元素也都唯一被 A 中某一个元素所连接，因此很明显可以将这个过程反过来，从 B 中向 A 中拉连接线．另外，如果 A 和 B 存在双射，意味着 A 和 B 的元素数量应该一致[4]．

函数是一种常见的映射，例如 $f(x) = 2x$ 可以看作映射 $f : \mathbb{R} \to \mathbb{R}$．但是映射可以从任意集合到任意集合．例如将整数映射到正多边形，将函数的映射到函数或实数（一般把这种映射称为**算符**）等．

注意当一个集合中有无限个元素时，我们有可能在它的子集和它本身之间建立一一映射，例如函数 $\tan(x)$ 可以从实轴的开区间 $(-\pi/2, \pi/2)$ 一一映射到整个实轴 $\mathbb{R}$，又例如我们可以将全体整数 $\mathbb{Z}$ 乘以二后一一映射到全体偶数 $2\mathbb{Z}$ 上．这时我们仍然认为这两个集合的元素一样多，虽然直觉上可能不容易接受．

Cantor-Bernstein 定理显示，如果集合 A 到集合 B 上存在一个单射 f 和一个满射 g，那么总可以利用 f 和 g 来构造出一个双射．

多元运算

有时候我们需要将两个集合 A, B 中任意各取一个元素，然后映射另一个集合 C 中的元素，称为**二元运算（binary operation）**．我们可以使用笛卡尔积

[4]本书中统一使用这种定义．一些其他教材中也把我们的“单射”称为“一一映射”，把“满射”称为“到上”，把“双射”称为“一一到上”，需要特别小心．

（式 2 [92]）将这个映射表示为

$$A \times B \to C \tag{3}$$

一个简单的例子就是两个实数的的加法减法或乘法可以表示为 $\mathbb{R}\times\mathbb{R} \to \mathbb{R}$（或简记为 $\mathbb{R}^2 \to \mathbb{R}$），但除法不可以，因为除数的集合不是 $\mathbb{R}$ 而是 $\mathbb{R}$ 去掉 0.

同理，**多元运算**可以用多个卡氏积表示为

$$A_1 \times \cdots \times A_N \to C \tag{4}$$

例如含有 N 个自变量的函数就是一个 N 元运算. 特殊地，N 个相同集合 A 做卡氏积可以简单表示为 A^N，例如 N 个有序复数的集合为 $\mathbb{C}^N$.

相等和拓展

定义 2

当映射（算符）$f : A \to B$ 和 $g : C \to D$ 的定义域相等（$A = C$）且对任意 $x \in A$ 都有 $fx = gx$，那么我们就说两个映射（算符）相等，记为 $f = g$，否则它们就不相等.

注意定义中的 B, D 不需要相等.

定义 3

若 A 是 C 的子集（$A \subseteq C$），我们就说 g 是 f 的拓展（**extension**），记为 $f \subseteq g$. 特殊地，当 A 是 C 的真子集（$A \subset C$），就记为 $f \subset g$.

恒等映射

定义 4

若一个集合到它子集的映射 $f : X \to X$ 把任意 $x \in X$ 映射到 x 本身，我们就叫它恒等映射（**identity map**）或者单位算符（**unit operator**），通常用 I 或 E 表示.

注意对不同集合 X，它们的单位算符定义域并不相等，所以它们的单位算符也不相等.

复合映射

定义 5

给定两个映射 $f : A \to B$ 和 $g : C \to D$，如果 f 的值域 R 是 g 的定义域 C 的一个子集，则可以定义复合映射（**composition of maps**）$g \circ f : A \to D$，即先将 A 中的元素通过 f 映射到 $R \subseteq C$，再通过 g 映射到 D 的元素.

在没有歧义的情况下也可以将“$\circ$”省略，尤其是将映射称为**算符**时.

复合映射常见的例子是复合函数，令 $\mathbb{R}$ 上的函数 $f(x) = \sin x$，$g(x) = x^2$，则复合函数 $g \circ f : \mathbb{R} \to [0, 1]$ 为 $(g \circ f)(x) = g(f(x)) = \sin^2 x$.

根据定义，复合映射满足**分配律**，令 f, g, h 为映射，则

$$h \circ (g \circ f) = (h \circ g) \circ f \tag{5}$$

映射的逆与逆映射

先来看简单的情况，如果 $f : A \to B$ 是单射，那么可以把它的逆映射记为 $f^{-1} : B' \to A'$，其中 $B' \subseteq B$ 是 f 的值域，A' 可以是 A 的任意父集. 就必须要求 f 为单射，否则若从 f^{-1}

未完成：幂集最好另开词条，这里只介绍单射的情况，以及简单介绍一下逆像的概念

给定集合 A, B，定义 B^A 为“从 A 到 B 的所有可能的映射所构成的集合”. 如果 B 是一个二元集合，即它只有两个元素，不妨记为 $B = \{0, 1\}$，那么 B^A 可以用来表示 A 的**幂集**，即由 A 的所有子集所构成的集合. 这是因为对于任意的 $f \in B^A$，我们可以把这个 f 对应到 A 的子集 S，其中 S 的元素全都被 f 映射到 1 上，$A - S$ 的元素全都被 f 映射到 0 上. 当然，0 和 1 的地位反过来也可以. 由于这个特点，我们简单地把 A 的幂集记为[5] 2^A.

利用映射 $f : A \to B$，可以导出一个映射 $f^{-1} : 2^B \to 2^A$，称为映射 f 的**逆**（**inverse**）. 对于 B 的任意子集 C，有

$$f^{-1}(C) = \{x | x \in A, f(x) \in C\} \tag{6}$$

此时，$f^{-1}(C)$ 称为 C 在 f 下的**逆像**（**inverse image**）或**原像**（**preimage**）.

[5] 当 A，B 都是有限集的时候，$|B^A| = |B|^{|A|}$. 特别地，$|2^A| = 2^{|A|}$.

特别地，和 f 的值域中不相交的 C 被 f^{-1} 映射到空集上，而空集也是 A 的一个子集．如果 f 是一个双射，那么对于任意 $y \in B$，单元素子集 $\{y\}$ 都被 f^{-1} 映射在 A 的某个单元素子集上，那么我们也可以认为此时 f^{-1} 实际上是单个元素映射在单个元素上，也就是从 B 到 A 的映射．

如果 $f : A \to B$ 是**单射**，那么 f^{-1} 总是把单点集映射到单点集上，因此这时我们可以定义 $f^{-1} : B \to A$，使得 $\forall x \in A$ 都有 $f^{-1}(f(x)) = x$．特别地，此时我们将 f^{-1} 称为 f 的**逆映射（inverse map）**[6]．

对于双射 $f : A \to B$，显然 $f \circ f^{-1}$ 和 $f^{-1} \circ f$ 都是单位算符（恒等映射）．注意两者的定义域分别为 A 和 B，当 $A \neq B$ 时不能写成 $f \circ f^{-1} = f^{-1} \circ f$．如果把 A，B 到自身的恒等映射分别记为 I_A 和 I_B，那么 $f \circ f^{-1} = I_B$，$f^{-1} \circ f = I_A$．

例 1

如果取正弦函数 $y = \sin x$ 的值域为 $R = [-1, 1]$ 如果取定义域为 $\mathbb{R}$，那么它不是一个单射，因为每一个 $y \in R$ 都对应无穷个 x，所以不存在反函数．但如果取定义域为 $[-\pi/2, \pi/2]$，那么它是一个单射，存在反三角函数 $\sin^{-1} : [-1, 1] \to [-\pi/2, \pi]$．

根据以上定义，$\sin^{-1}(\sin(x))$ 是定义在 $[-\pi/2, \pi/2]$ 上的恒等函数，而 $\sin\big(\sin^{-1}(x)\big)$ 是定义在 $[-1, 1]$ 上的恒等函数，所以有 $\sin \circ \sin^{-1} \subseteq \sin^{-1} \circ \sin$．

抽象

- 本词条存在未完成的内容．

我们这里要讨论的是**抽象（abstract）**一词，意为“抽出一个对象的部分特性做研究而忽略其它特性”或者“抽出若干对象的共同特征做研究而抛弃它们的特有特征”，其中“象”取“类比”、“相似”之意．当你在脑海中想象“树”的概念时，只关注树共有的一些关键特质，比如有树叶有枝干有根系等，至于具体多少树叶、根系如何分布则被忽略了，因此抽象的结果往往是不能画出来的，因为画出来的树都有了具体的特征了，不再是抽象的树．抽象的对立面，是具象．abstract 一词是由 ab-（向外的）和-tract（拉、拔）构成的，词义和抽

[6]注意“映射的逆”和“逆映射”的区别．映射的**逆**是子集到子集的映射，而**逆映射**是点到点的映射．任何映射都有**逆**，但是只有单射才有**逆映射**．

象完全一样．当我们说一个事物更抽象时，我们其实也在说它更为“一般”或者更“广义”.

未完成：例子：从位移，速度，加速度等中抽象出矢量

函数

预备知识 映射[98]

- 本词条存在未完成的内容.

广义来说，任何映射都可以叫做**函数（function）**．所以我们也可以用映射的记号表示函数．例如 $f:\mathbb{R}\to\mathbb{R}$ 表示定义域为实数集，值为实数的函数，通常记为 $f(x)$．又例如 $f:\mathbb{R}^2\to\mathbb{C}$ 表示实变量和复值的二元函数 $f(x_1,x_2)$．注意其中 $\mathbb{R}^2$ 表示笛卡尔积（式 2 [92]）$\mathbb{R}\times\mathbb{R}$.

复合函数

未完成：复合函数即复合映射

函数的性质

以后我们会看到一些用极限[149] 和导数[158] 描述的性质．例如连续性[156]，一致连续，可导.

反函数

- 本词条处于草稿阶段.

预备知识 函数[103]

由于函数本质上是映射[98]，那么如果一个函数的映射存在逆映射，那么我们就把逆映射对应的函数成为**反函数（inverse function）**.

我们知道不是所有的映射都有反映射．只有单射存在逆映射，而多对一映射不存在逆映射.

$$x = f^{-1}[f(x)] \tag{1}$$

我们往往可以通过缩小定义域的方式来使一个函数具有反函数．例如 $\sin^{-1} x$ 是 $\sin(x)$ 在 $[-\pi/2, \pi/2]$ 区间上的反函数.

图像

如果一个实函数 $y = f(x)$（$f : \mathbb{R} \to \mathbb{R}$）可以使用图像描述，那么反函数就是 $f(x)$ 关于直线 $y = x$ 的镜像对称.

例如 x^2 和 $\sqrt{x}$ 在 $(0, \infty)$ 区间的函数图像．关于 $y = x$ 对称．S

无穷的概念

预备知识 映射[98]

显然由于空集不含任何元素，$|\varnothing| = 0$. 对两个集合求并集时，由于相同元素不重复计数，所以对于集合 A 和 B，我们有不等式：$|A \cap B| \leqslant |A| + |B|$.

不同的集合有可能有一样多的元素，这个时候它们的基数相等．对于无穷集合，我们没法一个个数出集合中的元素，所以靠数元素数目是没法研究无穷集合的基数的．康托尔（Cantor）注意到，可以通过将两个集合的元素一一对应，从而比较集合的大小．能够建立双射的集合彼此大小一样，也就是说基数一样；对于集合 A 和 B，如果 A 到 B 只能建立单射而无法建立满射，那么就认为 A 的基数严格小于 B 的基数；反之，如果 B 到 A 可以建立满射但不能建立双射，那么就认为 B 的基数严格大于 A 的基数．由 **Cantor-Bernstein 定理**，这两个关于严格大于和小于的判断是等价的.

对于有限集合 A，$|A| = n$，这里 n 是某个非负整数．显然，如果 A 和 B 都是有限集合，那么它们的基数之间的大小关系和对应的 n 的大小关系是一致的.

无限集合分为**可数集（countable set）**和**不可数集（uncountable set）**．对于有限集合，如果 A 是 B 的真子集，那么 $|A| < |B|$ 严格成立；但是如果 A 是 B 的真子集且 A 是无穷集合，那么我们最多只能认为 $|A| \leqslant |B|$．

可数集是指大小和整数集 $\mathbb{Z}$ 相同的集合，也就是说，可以和全体整数进行一一对应[98]．比如，全体偶数的集合 $2\mathbb{Z}$ 就是一个可数集，只需要令映射 $f \to \mathbb{Z} = 2\mathbb{Z}$ 满足对于任何 $\mathbb{Z}$ 中的元素 n，$f(n) = 2n$ 就可以．由于全体整数的数量和全体非负整数的数量一样，可数集里的元素和非负整数也可以建立双射，这样我们就可以按照双射的连接“挨个数出”可数集里的元素，这就是可数集名称的来源．有理数集是可数的，但无理数集和实数集是不可数的．

可数集是最小的一类无穷集合，也就是说，任何无穷集合都可以满射到可数集上．

不可数集是指无法和可数集建立双射的无穷集合，它们都严格大于可数集．由 **Cantor 定理**可知，对于任何集合（不论有限无限）A，$|A| < |2^A|$[7]严格成立．我们把可数集的基数记为 $\aleph_0$；如果集合 A 的基数为 $\aleph_n$，那么定义 $|2^A| = \aleph_{n+1}$．实数集的基数被定义为 $\aleph$，利用 Cantor-Bernstein 定理可以证明 $\aleph = \aleph_1$．

集合论中的基本难题，连续统假设，猜测在 $\aleph_0$ 和 $\aleph$ 之间没有别的无穷基数，即不存在一个集合，严格大于整数集而又严格小于实数集．很不幸的是，连续统假设对于目前公认的集合论公理系统（ZFC 公理）是一个独立命题，即无法证明．

阶乘

- 本词条处于草稿阶段．

对正整数 $n > 0$，**阶乘（factorial）**定义为小于等于 n 且大于零的所有整数相乘

$$n! = 1 \cdot 2 \cdot 3 \ldots (n-2)(n-1)n = \prod_{k=1}^{n} k \tag{1}$$

其中 $\prod$ 符号的用法与求和符号 $\sum$ 相似，只是把加法变为乘法．

[7] 2^A 指 A 的全体子集构成的集合，称作 A 的幂集．至于为什么用这个表示方法，请参考[98]．

特殊地，我们定义

$$0! = 1 \tag{2}$$

阶乘可以拓展到非整数，见“Gamma 函数[254]”.

排列

预备知识 集合[92]，阶乘[105]

[8]我们讨论含有 N 个元素的任意集合，由于集合中元素的名称不重要，我们以下将它记为 $\{1, 2, \ldots, N\}$. 注意集合的是没有顺序的，例如 $\{1, 2, 3\}$ 和 $\{1, 3, 2\}$ 是同一个集合. 当我们把集合 S 中的的元素按照某种顺序排列成一个序列时，就称为它是集合 S 的一种**排列（permutation）**.

那么 N 个元素的集合一共有几种不同的排列呢？第 1 个位置有 N 种不同的可能，确定之后第 2 个位置有 $N-1$ 种不同的可能，第 3 个位置有 $N-2$ 种……最后一个位置只有 1 种. 所以可能性的种数可以用阶乘[105] 表示，记为 A_N

$$A_N = N! = N(N-1)(N-2)\ldots 1 \tag{1}$$

我们可以把第 i 种排列记为 p_i，该排列的元素按照顺序分别记为 $p_{i,1}, p_{i,2}, \ldots, p_{i,N}$.

组合

- 本词条处于草稿阶段.

预备知识 阶乘[105]

[9]数学上，**组合（combination）**用于计算从 n 个不同的物体中选取 $0 \leqslant r \leqslant n$ 个，有几种方法. 国内高中课本常用 C_n^r 来表示，但我们使用 $\binom{n}{r}$.

$$\binom{n}{r} = \frac{n!}{(n-r)!r!} \tag{1}$$

[8]参考 Wikipedia 相关页面.

[9]参考 Wikipedia 相关页面.

克罗内克 delta 函数

- 补充一些例题

克罗内克 δ 函数（Kronecker delta function），定义为

$$\delta_{ij} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \tag{1}$$

易证 δ 函数的两个常用性质

$$\sum_j a_j \delta_{ij} = a_i \tag{2}$$

$$\sum_{ij} a_i b_j \delta_{ij} = \sum_k a_k b_k \tag{3}$$

二项式定理

预备知识 组合[106]

二项式展开公式为

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i} \quad (n\text{为正整数}) \tag{1}$$

其中表示**组合（combination）**，定义为

$$\binom{n}{i} = \frac{n(n-1)\dots(n-i+1)}{i!} = \frac{n!}{i!(n-i)!} \tag{2}$$

推导

若展开多项式的时候先不合并同类项（每项前面的系数都是 1 ）则

- $(a+b)^0 = 1$ 有 1 项

- $(a+b)^1 = a+b$ 有 2 项
- $(a+b)^2 = aa+ab+ba+bb$ 有 4 项
- $(a+b)^3 = aaa+aab+aba+abb+baa+bab+bba+bbb$ 有 8 项
- $(a+b)^n$ 有 2^n 项（若不合并相同项）

这就相当于用 a 和 b 填满 n 个有序的位置，每个位置都可以取 a 或 b，共有 2^n 种排列，每种排列就是一项，所以共有 2^n 项.

下面把 2^n 项中的相同项进行合并，把其中出现了 i 个 a 及 $n-i$ 个 b 的项都记为 a^ib^{n-i}，那么共有 $\binom{n}{i}$ 个这样的项. 把它们相加得 $\binom{n}{i}a^ib^{n-i}$. 所以

$$(a+b)^n = \sum_{i=0}^{n}\binom{n}{i}a^ib^{n-i} \tag{3}$$

证毕.

二项式定理（非整数幂）

预备知识 二项式定理[107]

当 a，b 为实数，u 为非零实数时，有

$$(a+b)^u = \sum_{i=0}^{+\infty}\frac{u(u-1)\dots(u-i+1)}{i!}a^ib^{u-i} \tag{1}$$

容易看出，当 u 为整数时，$i>u$ 的所有项为 0，得到整数指数的二项式定理[107].

数值验证

在学习微积分之前，这里只给出一个数值验证的方法（而不是证明）. 在微积分中，这个定理可以用泰勒展开[181] 推导出来.

首先化简上式，不妨令 $|a|<|b|$，把 b^u 提出括号，再令 $x\equiv a/b$，有 $|x|<1$.

$$(a+b)^u = b^u(1+x)^u = b^u\sum_{i=0}^{+\infty}\frac{u(u-1)\dots(u-i+1)}{i!}x^i \tag{2}$$

所以只要用数值验证

$$(1+x)^u=\sum_{i=0}^{+\infty}\frac{u(u-1)\dots(u-i+1)}{i!}x^i \tag{3}$$

即可．接下来，可以用计算器或程序对式 3 的前 N 项进行求和．如果增加 N 使结果趋近精确值，则验证成功（见下表）．这里给出计算下表的 Matlab 程序[1841]．

表 1: 数值验证二项式定理（非整数幂）

$(1+x)^u$	$N=5$	$N=20$	$N=100$	精确值前 8 位
$x=0.5,u=-0.3$	0.88445640	0.88546751	0.88546749	0.88546749
$x=0.6,u=3.1$	4.2930453	4.2931093	4.2931093	4.2931093

直线和平面的交点

预备知识 高中解析立体几何，几何矢量[370]

若平面上任意一点为 $\mathbf{p}=(p_x,p_y,p_z)$，法向量为 $\mathbf{n}=(n_x,n_y,n_z)$．直线上一点为 $\mathbf{s}=(s_x,s_y,s_z)$，方向为 $\mathbf{v}=(v_x,v_y,v_z)$，求射线与平面的交点．注意 $\mathbf{n}$ 和 $\mathbf{v}$ 不必是单位矢量．

平面方程为

$$(\mathbf{r}-\mathbf{p})\cdot\mathbf{n}=0 \tag{1}$$

直线的参数方程为

$$\mathbf{r}=\lambda\mathbf{v}+\mathbf{s} \tag{2}$$

式 2 代入式 1 解得

$$\lambda=\frac{(\mathbf{p}-\mathbf{s})\cdot\mathbf{n}}{\mathbf{v}\cdot\mathbf{n}} \tag{3}$$

再代入式 2 得交点为

$$\mathbf{r}=\frac{(\mathbf{p}-\mathbf{s})\cdot\mathbf{n}}{\mathbf{v}\cdot\mathbf{n}}\mathbf{v}+\mathbf{s} \tag{4}$$

（未完成：给出 Matlab 代码）

点到直线的距离

- 本词条处于草稿阶段.

一点 (x_1, y_1) 到直线 $ax + by + c = 0$ 的最短距离为

$$\frac{|ax_1 + by_1 + c|}{\sqrt{a^2 + b^2}} \tag{1}$$

最近点的坐标为

$$\left(\frac{b^2 x_1 - aby_1 - ac}{a^2 + b^2}, \quad \frac{a^2 y_1 - abx_1 - bc}{a^2 + b^2} \right) \tag{2}$$

未完成：推导

海伦公式

[10]若已知三角形的边长，其面积可以用海伦公式计算

$$A = \sqrt{s(s-a)(s-b)(s-c)} \tag{1}$$

其中 $s = (a + b + c)/2$.

未完成：证明

直线和球的交点

预备知识 Matlab 的函数[1700]，一元二次方程

[10]参考 Wikipedia 相关页面.

直线参数方程可以用几何矢量[370]表示为 $\mathbf{P}_0+\lambda\hat{\mathbf{v}}$，其中 $\hat{\mathbf{v}}$ 是单位矢量．球心为 $\mathbf{P}_c$，半径为 R，那么直线和球相交的点满足

$$(\mathbf{P}_0+\lambda\hat{\mathbf{v}}-\mathbf{P}_c)^2=R^2 \tag{1}$$

其中左边的平方表示矢量和自身的内积[383]，即矢量模长的平方．展开后得到关于 λ 的二次方程

$$\lambda^2+2\hat{\mathbf{v}}\cdot(\mathbf{P}_0-\mathbf{P}_c)\lambda+(\mathbf{P}_0-\mathbf{P}_c)^2-R^2=0 \tag{2}$$

直接使用求根公式得

$$\lambda=\frac{-b\pm\sqrt{b^2-4c}}{2} \tag{3}$$

其中

$$b=2\mathbf{v}\cdot(\mathbf{P}_0-\mathbf{P}_c)\qquad c=(\mathbf{P}_0-\mathbf{P}_c)^2-R^2 \tag{4}$$

以下给出 Matlab 代码

未完成：应该支持多条直线而不是多个球……

```
% 计算直线和球的焦点
% numel(L)=6, size(P)=[N,3], numel(R)=1
function [P1,P2] = LSph2P(L6, P, R)
L6 = L6(:)';
P0 = L6(1:3); v0 = L6(4:6)/norm(L6(4:6));
P0_P = P0-P;
b = 2*P0_P*v0';
c = dot(P0_P, P0_P, 2) - R^2;
d = b.^2 - 4*c;
if d < 0
    P1 = []; P2 = []; return;
end
d = sqrt(d);
P1 = 0.5*(-b -d).*v0 + P0;
if d == 0
    P2 = []; return;
end
P2 = 0.5*(-b +d).*v0 + P0;
end
```

等比数列

- 本词条需要更多讲解，便于帮助理解.
- 本词条需要更多参考文献.

求和公式（从 1 开始 n 项求和）

$$1+q+q^2+\cdots+q^{n-1}=\frac{q^n-1}{q-1} \tag{1}$$

证明

使用数学归纳法，当 $n=1$ 时

$$\frac{q^n-1}{q-1}=1 \tag{2}$$

恰好是第 1 项．第 $n+1$ 项为

$$\frac{q^{n+1}-1}{q-1}-\frac{q^n-1}{q-1}=q^n \tag{3}$$

证毕.

三角恒等式

预备知识 三角函数

这里列出几个高中常见的三角函数恒等式，推导从略．以下用到的两个高中不常见的三角函数分别为 $\csc x=1/\sin x$，$\sec x=1/\cos x$，分别读作 cosecant 和 secant

勾股定理

$$\sin^2 x+\cos^2 x=1 \tag{1}$$

等式两边同除 $\cos^2 x$ 得

$$\tan^2 x+1=\sec^2 x \tag{2}$$

两角和公式

$$\sin(x \pm y) = \sin x \cos y \pm \cos x \sin y \tag{3}$$

$$\cos(x \pm y) = \cos x \cos y \mp \sin x \sin y \tag{4}$$

二倍角公式

令式 3 中 $y = x$ 取上号得

$$\sin 2x = 2 \sin x \cos x \tag{5}$$

$$\cos 2x = \cos^2 x - \sin^2 x \tag{6}$$

降幂公式

结合式 6 和 $\sin^2 x + \cos^2 x = 1$ 可以得到

$$\sin^2 x = \frac{1}{2}(1 - \cos 2x) \tag{7}$$

$$\cos^2 x = \frac{1}{2}(1 + \cos 2x) \tag{8}$$

和差化积公式

$$\sin x + \sin y = 2 \sin\left(\frac{x+y}{2}\right) \cos\left(\frac{x-y}{2}\right) \tag{9}$$

$$\sin x - \sin y = 2 \sin\left(\frac{x-y}{2}\right) \cos\left(\frac{x+y}{2}\right) \tag{10}$$

$$\cos x + \cos y = 2 \cos\left(\frac{x+y}{2}\right) \cos\left(\frac{x-y}{2}\right) \tag{11}$$

$$\cos x - \cos y = -2 \sin\left(\frac{x+y}{2}\right) \sin\left(\frac{x-y}{2}\right) \tag{12}$$

这里介绍一种推导方法可方便记忆. 以式 11 为例，$\cos x, \cos y$ 和 $\cos x + \cos y$ 分别等于图 1 中矢量 $\mathbf{A}, \mathbf{B}$ （令它们的模长为 1）和 $\mathbf{A} + \mathbf{B}$ 在水平方向的投影长度，而 $\mathbf{A} + \mathbf{B}$ 在水平方向的投影长度等为 $|\mathbf{A} + \mathbf{B}| \cos[(x + y)/2]$，其中

$|\mathbf{A}+\mathbf{B}|=2\cos[(y-x)/2]$，代入可得式 11．利用 $\mathbf{A}+\mathbf{B}$ 在竖直方向的投影可得式 9，把式 9 和式 11 中的 y 分别替换成 $-y$ 和 $y+\pi$ 可推导出式 10 和式 12．

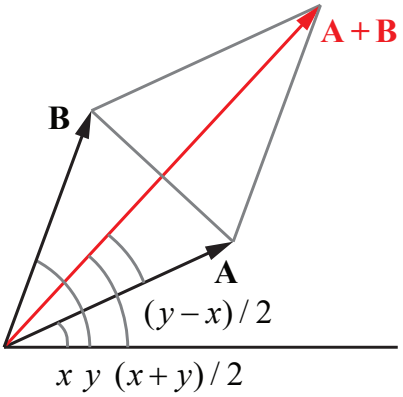


图 1: 和差化积公式推导

积化和差公式

根据和差化积公式，我们也可以直接写出积化和差公式

$$\sin x\sin y=\frac{1}{2}[\cos(x-y)-\cos(x+y)] \tag{13}$$

$$\cos x\cos y=\frac{1}{2}[\cos(x+y)+\cos(x-y)] \tag{14}$$

$$\sin x\cos y=\frac{1}{2}[\sin(x+y)+\sin(x-y)] \tag{15}$$

双曲函数

预备知识 指数函数

[11]这里介绍三种**双曲函数（hyperbolic function）**：双曲正弦函数，双曲余弦函数和双曲正切函数．本文只讨论实数自变量的情况．他们的定义分别为

$$\sinh x=\frac{\mathrm{e}^{x}-\mathrm{e}^{-x}}{2} \tag{1}$$

[11]参考 Wikipedia 相关页面．

$$\cosh x = \frac{\mathrm{e}^{x} + \mathrm{e}^{-x}}{2} \tag{2}$$

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{\mathrm{e}^{x} - \mathrm{e}^{-x}}{\mathrm{e}^{x} + \mathrm{e}^{-x}} \tag{3}$$

其中 e 是一个高等数学中常见的常数，叫做自然对数底[153]．这三个函数的图像如图 1．

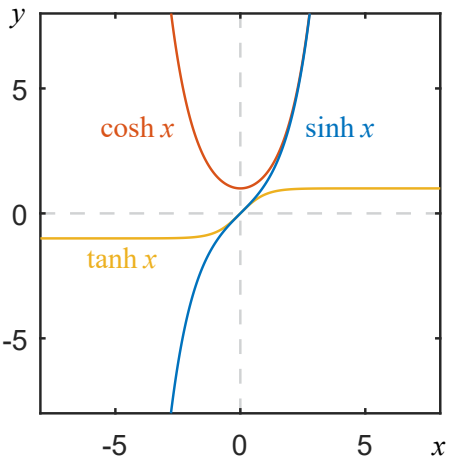


图 1: 三种双曲函数的图像

注意 $\sinh x$ 和 $\tanh x$ 是奇函数，$\cosh x$ 是偶函数．

例 1　反双曲正弦函数

要求 $\sinh x$ 的反函数，我们令

$$x = \sinh y = \frac{\mathrm{e}^{y} - \mathrm{e}^{-y}}{2} \tag{4}$$

整理成关于 e^{y} 的二次方程，得

$$(\mathrm{e}^{y})^{2} - 2x\mathrm{e}^{y} - 1 = 0 \tag{5}$$

解出 e^{y} 为

$$\mathrm{e}^{y} = x \pm \sqrt{x^{2} + 1} \tag{6}$$

由于 $\mathrm{e}^{y} > 0$，上式取正号．两边取自然对数，得

$$y = \sinh^{-1} x = \ln\left(x + \sqrt{x^{2} + 1}\right) \tag{7}$$

一些性质

$$\cosh^2 x - \sinh^2 x = 1 \tag{8}$$

$$\sinh(x+y) = \sinh x \cosh x + \cosh x \sinh x \tag{9}$$

$$\cosh(x+y) = \cosh x \cosh x + \sinh x \sinh x \tag{10}$$

sinc 函数

预备知识 连续函数[156]

sinc 函数的定义为

$$\operatorname{sinc} x = \begin{cases} \dfrac{\sin x}{x} & (x \neq 0) \\ 1 & (x = 0) \end{cases} \tag{1}$$

sinc 函数的图像如图 1，可以证明，该函数在 $x = 0$ 处是连续的，即

$$\lim_{x \to 0} \frac{\sin x}{x} = \operatorname{sinc} 0 = 1 \tag{2}$$

积分性质

$$\int_{-\infty}^{+\infty} \operatorname{sinc}(x)\,\mathrm{d}x = \pi \tag{3}$$

$$\int_{-\infty}^{+\infty} \operatorname{sinc}^2(x)\,\mathrm{d}x = \pi \tag{4}$$

充分必要条件

若由命题 A 能推导出命题 B，则 A 是 B 的充分条件，B 是 A 的必要条件．如何理解这个定义呢？下面举两个例子．

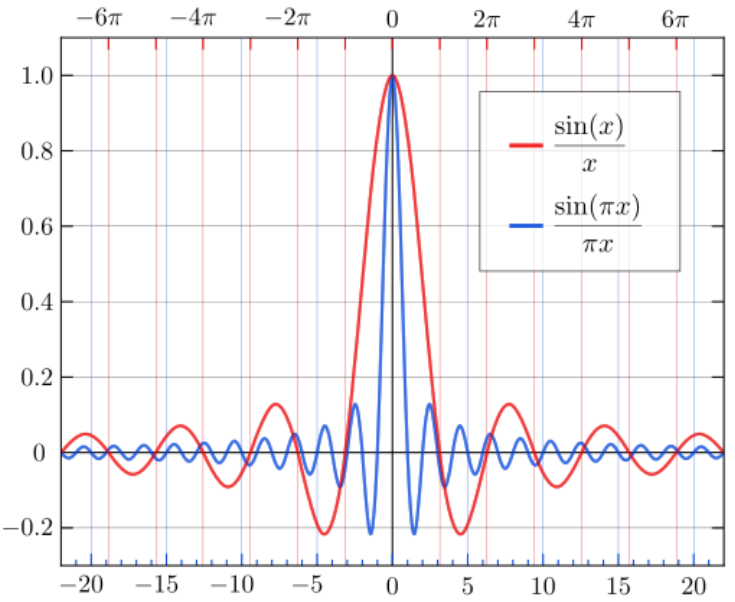


图 1: sinc 函数（来自维基百科）

例 1

命题 A：四边形 $ABCD$ 是一个正方形.

命题 B：四边形 $ABCD$ 的四条边相等.

首先我们考虑 A 对 B 的关系. 显然，由 A 可以推出 B，说明 A 中有充分的信息能得到 B，所以叫做 B 的充分条件. A 中包括得到 B 所必要的信息，还可能包括一些其他信息，例如由命题 A 可以得出四边形任意两条临边垂直. 这些多出来的信息并不一定是得到 B 所必须的，因为还有许多其他的四边形四条边相等但并不是正方形.

那如何判断 A 中有没有多余的信息呢？我们可以反过来试图用 B 推导命题 A，若原则上得不出 A（而不是因为我们逻辑水平不够），则证明 A 中有多余的条件. 这时我们说 A 不是 B 的必要条件，因为 A 中的一些信息是多余的，也就是没有必要的. 综上，A 是 B 的充分非必要条件.

现在我们从 B 的角度考虑. 虽然由条件 B 不能推导出条件 A，但是 B 是 A 中信息的一部分，B 必须要成立才有可能使 A 成立，也就是说如果 B 不成立 A 就不可能成立（四条边不全相等的四边形一定不是正方形）. 所以说 B 是 A 的必要条件. 另外，由 B 中的少量信息不能得到 A，所以 B 不是 A 的充分条件. 综上，B 是 A 的必要非充分条件.

例 2

命题 A：三角形 X 的其中两内个角分别为 90° 和 45°.

命题 B：三角形 X 有两个 45° 的内角.

利用三角形三个内角和为 180° 的事实，可以从 A 推出 B，说明 A 是 B 的充分条件，B 是 A 的必要条件. 但也可以从 B 推出 A，说明 B 是 A 的充分条件，A 是 B 的必要条件. 所以 A 和 B 既是彼此的充分条件也是彼此的必要条件. 所以我们说 A 和 B 互为充分必要条件. 若 A 是 B 的充分必要条件，B 一定也是 A 的充分必要条件. 因为两种表述都意味着 A，B 命题等效，所提供的信息都是一样的，两者都没有任何多余的或者缺失的信息.

需要注意的是

1. 充分/必要条件是两个命题之间的关系，若直说一个命题是充分/必要条件没有意义.

2. 讨论充分/必要条件需要在一定的前提下进行. 以上两个例子中的前提如：我们讨论的是欧几里得几何中的平面四边形和三角形. 当然，我们也可以把这个前提直接写在每个命题中.

3. 在证明 A 是 B 的充分必要条件时，需要分别证明 A（相对于 B）的充分性和必要性. 充分性需要由 A 证明 B，必要性需要由 B 证明 A.

4. 在证明 A 是 B 的充分非必要条件时，除了需要证明 A 的充分性，还需非必要性，即 B 不能推出 A. 只要我们可以举出一个 B 成立 A 不成立的反例，就立刻证明了不可能由 B 推出 A.

四象限 Arctan 函数

我们经常会遇到这样一个问题：已知平面直角坐标系上一点 P，坐标为 (x, y)，求射线 OP 与 x 轴正方向的夹角 θ. 首先我们要给这个夹角取一个范围，一般来说既可以取 $[0, 2\pi)$ 也可以取 $(-\pi, \pi]$，但如无特殊说明，我们统一使用后者.

一些教材中直接用 $\theta = \arctan(y/x)$，$\theta \in (-\pi/2, \pi/2)$ 来表示这一关系，这是不严谨的．我们下面来定义一个符合要求的新函数，记为 $\operatorname{Arctan}(y, x)$ [12]．其中定义为 $x, y \in \mathbb{R}$，即任意实数，值域为 $(-\pi, \pi]$．

$$\operatorname{Arctan}(y,x) \equiv \begin{cases} \arctan(y/x) & (x>0) \\ \arctan(y/x)+\pi & (x<0,\ y\geqslant 0) \\ \arctan(y/x)-\pi & (x<0,\ y<0) \\ \pi/2 & (x=0,\ y>0) \\ -\pi/2 & (x=0,\ y<0) \\ 0 & (x=0,\ y=0) \end{cases} \tag{1}$$

本书统一使用该定义，但也有一些其他文献将其定义为上式加 π，使值域为 $(0, 2\pi]$，或者认为 $x = 0, y = 0$ 无定义（不属于定义域）．

偏导数

函数在除了在原点和 x 轴的负半轴，在其它定义域都是连续且光滑的，即存在连续的无穷阶偏导[266]．一阶偏导为

$$\frac{\partial}{\partial x}\operatorname{Arctan}(y,x) = \frac{-y}{x^2+y^2} \qquad \frac{\partial}{\partial y}\operatorname{Arctan}(y,x) = \frac{-x}{x^2+y^2} \tag{2}$$

极坐标系

预备知识 平面直角坐标系，位置矢量[878]，四象限 Arctan 函数[118]

在平面上上取一个点作为原点，过原点作一条轴称为**极轴**，并选定极轴的正方向，规定单位长度．该平面上某点与原点连成的线段叫做**极径**，其长度一般用 r （或 ρ ）表示．若 r 为负值，则表示反方向的长度．极径与极轴的夹角叫做极角（规定逆时针旋转极角增加，顺时针旋转则减少），用 θ 表示．θ 的值

[12] 在许多编程语言中 arctan 被记为 atan，Arctan 被记为 atan2．也有一些文献将 Arctan 记为 Tan^{-1} 或 atan2

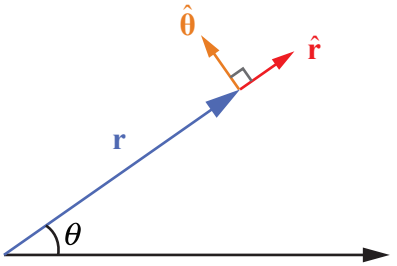


图 1: 极坐标系和两个单位矢量

通常表示成弧度，取值范围一般选 $(-\pi,\pi]$ 或 $[0,2\pi)$．于是任何一点都可以用两个有序实数 (r,θ) 来表示其在该平面上的位置，这就是一个点的**极坐标**.

为了表示一个坐标对应的单位矢量，我们一般把坐标变量名记为粗体并在上方加一个标记．例如直角坐标系中，$\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}$（有时也记为 $\hat{\mathbf{i}},\hat{\mathbf{j}},\hat{\mathbf{k}}$）代表 x,y,z 轴方向的单位矢量．在极坐标中，定义 $\hat{\mathbf{r}}$ 为 r 增加的方向的单位矢量，$\hat{\boldsymbol{\theta}}$ 为 θ 坐标增加方向的单位矢量（即 $\hat{\mathbf{r}}$ 逆时针旋转 $\pi/2$ 的方向）．$\hat{\mathbf{r}}$ 与 $\hat{\boldsymbol{\theta}}$ 互相垂直，构成一对**单位正交基底**，平面上的任意矢量都可以正交分解到这两个方向上．我们通常把 $\hat{\mathbf{r}}$ 的方向叫做**径向**，把 $\hat{\boldsymbol{\theta}}$ 的方向叫做**法向**．要注意极坐标中的两个单位矢量是 θ 的函数，对于不同的 θ，它们的方向也不同.

习题 1

试证明极坐标方程 $r=r_0/\cos(\theta-\theta_0)$ 和 $r=2R\cos(\theta-\theta_0)$ 分别表示一条直线和一个圆.

与直角坐标的转换

要在极坐标系的基础上建立一个直角坐标系，习惯的做法是取原点相同，且令 x 轴与极轴重合，y 轴取 $\theta=\pi/2$ 的方向．这样将 $\mathbf{r}$ 用 $\hat{\mathbf{x}}$ 和 $\hat{\mathbf{y}}$ 展开，就得到

$$\mathbf{r}=r\cos\theta\,\hat{\mathbf{x}}+r\sin\theta\,\hat{\mathbf{y}} \tag{1}$$

即

$$\begin{cases} x=r\cos\theta \\ y=r\sin\theta \end{cases} \tag{2}$$

这样就从极坐标转换成为直角坐标.

要从直角坐标转换为极坐标，首先由勾股定理有 $r^2=x^2+y^2$．使用反正

切函数，我们可以表示 $x>0$ 或 $\theta\in(-\pi/2,\pi/2)$ 时的 θ，即 $\theta=\arctan(y/x)$. 为了表示任意情况我们可以使用 Arctan 函数式 1 [118]. 这样，从直角坐标转到极坐标的转换就可以表示为

$$\begin{cases} r=\sqrt{x^2+y^2} \\ \theta=\mathrm{Arctan}(y,x) \end{cases} \tag{3}$$

根据 Arctan 的定义（式 1 [118]），θ 的范围是 $(-\pi,\pi]$.

阿基米德螺线

- 本词条需要更多参考文献.
- 本词条处于草稿阶段.

预备知识 极坐标中的曲线方程[208]

阿基米德螺线（图 1）的极坐标方程为

$$r=a+b\theta \qquad (\theta>-a/b) \tag{1}$$

注意虽然从原点发出的任意射线都与曲线有无穷多个交点，但一个 θ 仍然只对应一个 r，函数仍然是单值的.

柱坐标系

预备知识 极坐标系[119]

若在原有的直角坐标系上定义柱坐标系（图 1），可用三个变量 (r,θ,z) 描述三维空间中任意一点. 其中 r 代表该点到 z 轴的距离（$r\geqslant 0$），θ 代表与 x 轴的夹角，z 与直角坐标系相同. 柱坐标系相当于在极坐标系的基础上增加了一个垂直轴.

柱坐标系中的单位矢量如图 1 中的 $\hat{\mathbf{r}},\hat{\boldsymbol{\theta}},\hat{\mathbf{z}}$ 所示. 其中 $\hat{\mathbf{r}},\hat{\boldsymbol{\theta}}$ 与极坐标系中的定义相同，$\hat{\mathbf{z}}$ 是直角坐标系 z 轴的单位矢量，注意三个单位矢量两两垂直，构

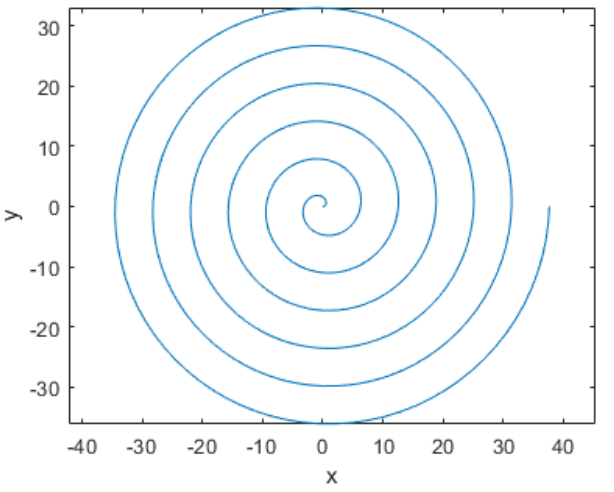


图 1: 阿基米德螺线，$r = \theta$

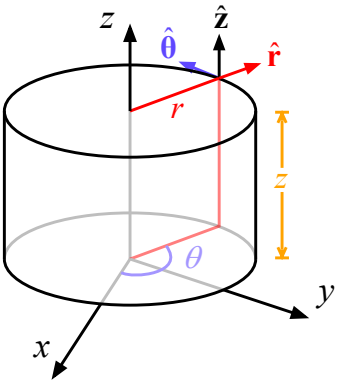


图 1: 定义柱坐标系

成一组单位正交基底，任何矢量可以在这组基底上展开．再次强调，与直角坐标系不同的是，$\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}$ 并不是常矢量，而是坐标 θ 的函数．

柱坐标与直角坐标间的转换类比式 2 [119] 和式 3 [119] 即可

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \\ z = z \end{cases} \qquad \begin{cases} r = \sqrt{x^2 + y^2} \\ \theta = \mathrm{Arctan}(y, x) \\ z = z \end{cases} \tag{1}$$

球坐标系

预备知识 位矢[878]

球坐标

三维直角坐标系中的一点 P 的位置可以用 (r,θ,ϕ) 这 3 个有序实数来表示，称为该点的**球坐标**（图 1）．其中 r 表示该点到原点的距离（$r\geqslant 0$），即位矢[878] 的**模长**；θ 表示该点的位矢与 z 轴的夹角（$\theta\in[0,\pi]$），即**极角**；ϕ 表示该点的位矢在 $x-y$ 平面上的投影与 x 轴的夹角（$\phi\in[0,2\pi]$或$[-\pi,\pi]$），即**方位角**．注意有些教材中用 θ 表示方位角，ϕ 表示极角，或者将 ϕ 记为 φ，r 记为 ρ 等，需要通过上下文判断每个坐标符号的具体含义．

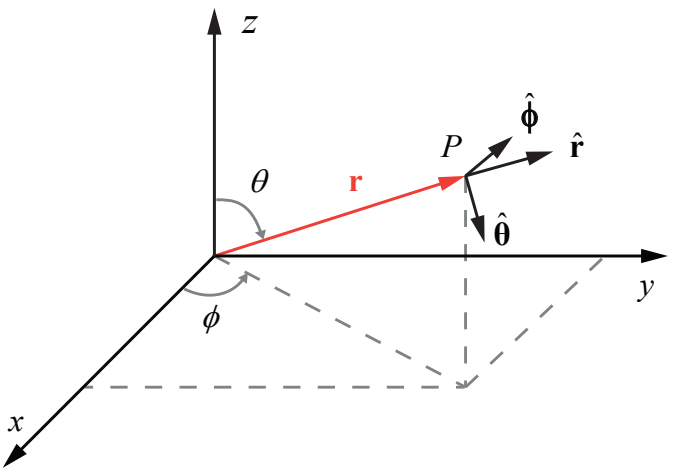


图 1: 球坐标系

球坐标系中的单位矢量

三个球坐标分别有对应的单位矢量 $\hat{\mathbf{r}},\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\phi}}$（如图）．定义它们的方向分别指向对应坐标增加的方向，例如 r 增加时，点 $P(r,\theta,\phi)$ 就向 $\hat{\mathbf{r}}$ 的方向移动．三个单位矢量两两垂直，形成一组正交归一基底，任意三维矢量都可以表示成它

们的线性组合．即

$$\mathbf{v}=(\mathbf{v}\cdot\hat{\mathbf{r}})\,\hat{\mathbf{r}}+(\mathbf{v}\cdot\hat{\boldsymbol{\theta}})\,\hat{\boldsymbol{\theta}}+(\mathbf{v}\cdot\hat{\boldsymbol{\phi}})\,\hat{\boldsymbol{\phi}}=v_r\,\hat{\mathbf{r}}+v_\theta\,\hat{\boldsymbol{\theta}}+v_\phi\,\hat{\boldsymbol{\phi}} \tag{1}$$

与直角坐标系不同的是，按照定义，球坐标的三个单位矢量是关于 θ 和 ϕ 的函数．即 $\hat{\mathbf{r}}(\theta,\phi)$，$\hat{\boldsymbol{\theta}}(\theta,\phi)$，$\hat{\boldsymbol{\phi}}(\phi)$．例如 P 的球坐标为 $(1,\pi/2,0)$，直角坐标为 $(1,0,0)$ 时，$\hat{\mathbf{r}}=\hat{\mathbf{x}}$，$\hat{\boldsymbol{\theta}}=-\hat{\mathbf{z}}$，$\hat{\boldsymbol{\phi}}=\hat{\mathbf{y}}$．但是球坐标为 $(1,\pi/2,\pi/2)$，直角坐标为 $(0,1,0)$ 时，$\hat{\mathbf{r}}=\hat{\mathbf{y}}$，$\hat{\boldsymbol{\theta}}=-\hat{\mathbf{z}}$，$\hat{\boldsymbol{\phi}}=-\hat{\mathbf{x}}$．一般地，对于球坐标为 (r,θ,ϕ) 的点 $\hat{\mathbf{r}},\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\phi}}$ 与 $\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}$ 的关系见球坐标与直角坐标的转换[128]．另外注意改变 r 时 $\hat{\mathbf{r}},\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\phi}}$ 都保持不变，且 $\hat{\boldsymbol{\phi}}(\phi)$ 仅由坐标 ϕ 决定．

三个坐标按照 (r,θ,ϕ) 排序，是为了使对应的单位矢量满足 $\hat{\mathbf{r}}\times\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\phi}}$（类比直角坐标系的三个单位矢量必须满足 $\hat{\mathbf{x}}\times\hat{\mathbf{y}}=\hat{\mathbf{z}}$，见矢量的叉乘[387]）．这也是所有**正交曲线坐标系**的要求．

球坐标系中矢量的两种表示方法

球坐标系中，矢量可以用球坐标 (r,θ,ϕ) 表示，即矢量以原点为起点，以终点的球坐标表示该矢量．

更常见的方法，是将矢量投影到 3 个单位矢量上（当然，要说明是关于哪个点的单位矢量），用单位矢量的线性组合来表示．在矢量分析中，这种方法常用于表示**矢量场**．

例如任意一点 $P(r,\theta,\phi)$ 的位矢[878] 都可以表示为 $r\,\hat{\mathbf{r}}$．又如原点处电荷 q 产生的电场为 $\mathbf{E}=kq\,\hat{\mathbf{r}}/r^2$．又如一个绕 z 轴逆时针旋转（角速度 ω）的圆柱，在 P 点的线速度为

$$\mathbf{v}=\omega r\sin\theta\,\hat{\boldsymbol{\phi}} \tag{2}$$

抛物线坐标系

预备知识 抛物线的三种定义[136]

[13]抛物线的极坐标方程为（式 1 [136]）

$$r=\frac{\eta}{1-\cos\theta} \tag{1}$$

[13]参考 [19] Chap 3.5 和 Wikipedia 相关页面．

若选用不同的半通径 ξ （$\xi > 0$），将得到一系列缩放的抛物线（图 1 中的绿色）．我们也可以把这些抛物线旋转 180 度，得到

$$r = \frac{\xi}{1 + \cos\theta} \tag{2}$$

这时我们把半通径记为 η（$\eta > 0$），当它取不同的值也得到一系列抛物线（图 1 中的红色）．另外我们再定义 $\eta = 0$ 和 $\xi = 0$ 分别代表极轴的正半轴和负半轴（包括原点）．综上，通过 ξ, η 两个坐标，我们就能确定平面上的唯一一点．

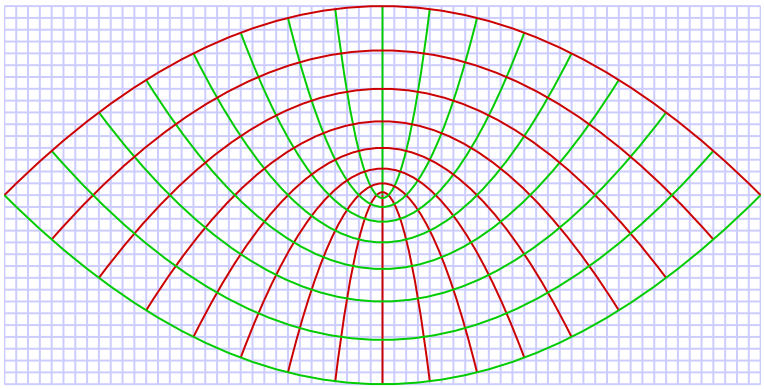


图 1: 抛物线坐标系，极轴向上（来自 Wikipedia）

若把这些曲线绕极轴旋转一周变为一系列抛物面，那么我们只需要再指定一个方位角 ϕ 就可以用坐标 (ξ, η, ϕ) 确定空间中的任意一点．

与直角坐标和球坐标的转换

一般令极轴与直角坐标的 z 轴重合，则根据定义有

$$\xi = r(1 + \cos\theta) = \sqrt{x^2 + y^2 + z^2} + z \tag{3}$$

$$\eta = r(1 - \cos\theta) = \sqrt{x^2 + y^2 + z^2} - z \tag{4}$$

$$\phi = \mathrm{Arctan}(y, x) \tag{5}$$

其中 Arctan 见“四象限 Arctan 函数[118]”．可以解得

$$z = \frac{\xi - \eta}{2} \tag{6}$$

$$x = \sqrt{\xi\eta}\cos\phi \qquad y = \sqrt{\xi\eta}\sin\phi \tag{7}$$

正交曲线坐标系

预备知识 正交曲线坐标系[288]

可以证明抛物线坐标系是一个正交曲线坐标系，即图 1 中过任意一点的两条坐标曲线都垂直：

在某点 $\mathbf{r}$，延 η 曲线的切线方向为 $\partial\mathbf{r}/\partial\xi$，延 ξ 曲线的切线方向为 $\partial\mathbf{r}/\partial\eta$

$$\mathrm{d}x = \sqrt{\frac{\eta}{\xi}}\frac{\cos\phi}{2}\,\mathrm{d}\xi + \sqrt{\frac{\xi}{\eta}}\frac{\cos\phi}{2\sqrt{\xi}}\,\mathrm{d}\eta - \sqrt{\xi\eta}\sin\phi\,\mathrm{d}\phi \tag{8}$$

$$\mathrm{d}y = \sqrt{\frac{\eta}{\xi}}\frac{\sin\phi}{2}\,\mathrm{d}\xi + \sqrt{\frac{\xi}{\eta}}\frac{\sin\phi}{2}\,\mathrm{d}\eta + \sqrt{\xi\eta}\cos\phi\,\mathrm{d}\phi \tag{9}$$

$$\mathrm{d}z = \frac{1}{2}\,\mathrm{d}\xi - \frac{1}{2}\,\mathrm{d}\eta \tag{10}$$

所以三个归一化单位矢量分别为（上式的每一列归一化）

$$\hat{\boldsymbol{\xi}} = \sqrt{\frac{\eta}{\xi+\eta}}\cos\phi\,\hat{\mathbf{x}} + \sqrt{\frac{\eta}{\xi+\eta}}\sin\phi\,\hat{\mathbf{y}} + \sqrt{\frac{\xi}{\xi+\eta}}\,\hat{\mathbf{z}} \tag{11}$$

$$\hat{\boldsymbol{\eta}} = \sqrt{\frac{\xi}{\xi+\eta}}\cos\phi\,\hat{\mathbf{x}} + \sqrt{\frac{\xi}{\xi+\eta}}\sin\phi\,\hat{\mathbf{y}} + \sqrt{\frac{\eta}{\xi+\eta}}\,\hat{\mathbf{z}} \tag{12}$$

$$\hat{\boldsymbol{\phi}} = -\sin\phi\,\hat{\mathbf{x}} + \cos\phi\,\hat{\mathbf{y}} \tag{13}$$

容易证明它们两两正交．另外有

$$\mathrm{d}\mathbf{r} = \frac{1}{2}\sqrt{\frac{\xi+\eta}{\xi}}\hat{\boldsymbol{\xi}}\,\mathrm{d}\xi + \frac{1}{2}\sqrt{\frac{\xi+\eta}{\eta}}\hat{\boldsymbol{\eta}}\,\mathrm{d}\eta + \sqrt{\xi\eta}\,\hat{\boldsymbol{\phi}}\,\mathrm{d}\phi \tag{14}$$

体积元等于式中三个分量相乘

$$\mathrm{d}V = \frac{1}{4}(\xi+\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta\,\mathrm{d}\phi \tag{15}$$

矢量算符

结合式 14 和式 3 [288] 到式 8 [288] 得

$$\boldsymbol{\nabla} u = 2\sqrt{\frac{\xi}{\xi+\eta}}\frac{\partial u}{\partial \xi} + 2\sqrt{\frac{\eta}{\xi+\eta}}\frac{\partial u}{\partial \eta} + \frac{1}{\sqrt{\xi\eta}}\frac{\partial u}{\partial \phi} \tag{16}$$

$$\boldsymbol{\nabla}\cdot\mathbf{A} = \frac{2}{\xi+\eta}\left[\frac{\partial}{\partial \xi}\left(\sqrt{\xi(\xi+\eta)}A_\xi\right) + \frac{\partial}{\partial \eta}\left(\sqrt{\eta(\xi+\eta)}A_\eta\right)\right] + \frac{1}{\sqrt{\xi\eta}}\frac{\partial A_\phi}{\partial \phi} \tag{17}$$

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{A} =& \left[\frac{2}{\sqrt{\xi+\eta}}\frac{\partial}{\partial \eta}(\sqrt{\eta}A_\phi) - \frac{1}{\sqrt{\xi\eta}}\frac{\partial A_\eta}{\partial \phi}\right]\hat{\boldsymbol{\xi}} \\ &+ \left[\frac{1}{\sqrt{\xi\eta}}\frac{\partial A_\xi}{\partial \phi} - \frac{2}{\sqrt{\xi+\eta}}\frac{\partial}{\partial \xi}\left(\sqrt{\xi}A_\phi\right)\right]\hat{\boldsymbol{\eta}} \\ &+ \left[\frac{2\sqrt{\xi}}{\xi+\eta}\frac{\partial}{\partial \xi}\left(\sqrt{\xi+\eta}A_\eta\right) - \frac{2\sqrt{\eta}}{\xi+\eta}\frac{\partial}{\partial \eta}\left(\sqrt{\xi+\eta}A_\xi\right)\right]\hat{\boldsymbol{\phi}}\end{aligned} \tag{18}$$

$$\nabla^2 u = \frac{4}{\xi+\eta}\left[\frac{\partial u}{\partial \xi}\left(\xi\frac{\partial}{\partial \xi}\right) + \frac{\partial u}{\partial \eta}\left(\eta\frac{\partial}{\partial \eta}\right)\right] + \frac{1}{\xi\eta}\frac{\partial^2 u}{\partial \phi^2} \tag{19}$$

连带拉盖尔多项式（associated Laguerre polynomial）.

一元函数的对称与周期性

轴对称，即当横坐标到对称轴的距离相等时，函数值也相等

$$f(a+x) = f(a-x) \tag{1}$$

中心对称，即当横坐标到对称轴的距离相等时，函数值互为相反数

$$f(a+x) = -f(a-x) \tag{2}$$

周期性，即横坐标经过一定长度后的函数值相等，

$$f(a+x) = f(x) \tag{3}$$

其中，a 分别为对称轴、对称中心和周期 T，现在来看同时含有参数 a 和参数 b 的情况

$$\text{轴对称：}\ f(a+x) = f(b-x) \tag{4}$$

$$\text{中心对称：}\ f(a+x)=-f(b-x) \tag{5}$$

$$\text{周期性：}\ f(a+x)=f(b+x) \tag{6}$$

其对称轴、对称中心、周期分别为：$x=(a+b)/2$，$((a+b)/2,0)$ 和 $T=|b-a|$.

若 $f(x)$ 关于直线 $x=a$ 与直线 $x=b$ 对称，则 $f(x)$ 的一个周期为 $2|b-a|$

若 $f(x)$ 关于点 $(a,0)$ 与点 $(b,0)$ 对称，则 $f(x)$ 的一个周期为 $2|b-a|$

若 $f(x)$ 关于直线 $x=a$ 与点 $(b,0)$ 对称，则 $f(x)$ 的一个周期为 $4|b-a|$

球坐标与直角坐标的转换

预备知识 球坐标系的定义[123]，四象限 Arctan 函数[118]

当我们讨论球坐标和直角坐标的转换时，通常令两个原点重合，取极轴（$\theta=0$）为 z 轴的正方向，$\theta=\pi/2$，$\phi=0$ 为 x 轴的正方向，$\theta=\pi/2$，$\phi=\pi/2$ 为 y 轴的正方向．这时两种坐标之间的变换关系为.

$$\begin{cases} x=r\sin\theta\cos\phi \\ y=r\sin\theta\sin\phi \\ z=r\cos\theta \end{cases} \tag{1}$$

$$\begin{cases} r=\sqrt{x^2+y^2+z^2} \\ \theta=\operatorname{Arctan}\left(\sqrt{x^2+y^2},z\right) \\ \phi=\operatorname{Arctan}(y,x) \end{cases} \tag{2}$$

其中 Arctan 函数（也记为 atan2）的定义见式 1 [118]．注意根据式 1，同一个直角坐标可以对应不同的极坐标，例如将 ϕ 增加 2π 的整数倍，由例如对 z 轴上的点 ϕ 可以取任意值．但根据式 2，我们可以找到两种坐标间的一一对应[92]关系.

矢量的变换

两组基底之间的变换关系为

$$\begin{cases} \hat{\mathbf{r}} = R_{11}\hat{\mathbf{x}} + R_{12}\hat{\mathbf{y}} + R_{13}\hat{\mathbf{z}} \\ \hat{\boldsymbol{\theta}} = R_{21}\hat{\mathbf{x}} + R_{22}\hat{\mathbf{y}} + R_{23}\hat{\mathbf{z}} \\ \hat{\boldsymbol{\phi}} = R_{31}\hat{\mathbf{x}} + R_{32}\hat{\mathbf{y}} + R_{33}\hat{\mathbf{z}} \end{cases} \tag{3}$$

$$\begin{cases} \hat{\mathbf{x}} = R_{11}\hat{\mathbf{r}} + R_{21}\hat{\boldsymbol{\theta}} + R_{31}\hat{\boldsymbol{\phi}} \\ \hat{\mathbf{y}} = R_{12}\hat{\mathbf{r}} + R_{22}\hat{\boldsymbol{\theta}} + R_{32}\hat{\boldsymbol{\phi}} \\ \hat{\mathbf{z}} = R_{13}\hat{\mathbf{r}} + R_{23}\hat{\boldsymbol{\theta}} + R_{33}\hat{\boldsymbol{\phi}} \end{cases} \tag{4}$$

其中 $\mathbf{R}$ 是关于两个角度的三维旋转矩阵[404]

$$\mathbf{R} = \begin{pmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \end{pmatrix} \tag{5}$$

若任意矢量 $\mathbf{v}$ 在直角坐标系和球坐标系中分别表示为

$$\mathbf{v} = v_x\hat{\mathbf{x}} + v_y\hat{\mathbf{y}} + v_z\hat{\mathbf{z}} \tag{6}$$

$$\mathbf{v} = v_r\hat{\mathbf{r}} + v_\theta\hat{\boldsymbol{\theta}} + v_\phi\hat{\boldsymbol{\phi}} \tag{7}$$

则坐标变换关系可以用矩阵乘法表示

$$\begin{pmatrix} v_r \\ v_\theta \\ v_\phi \end{pmatrix} = \mathbf{R} \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} \tag{8}$$

$$\begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = \mathbf{R}^{\mathrm{T}} \begin{pmatrix} v_r \\ v_\theta \\ v_\phi \end{pmatrix} \tag{9}$$

推导

把空间中一点 P 的位矢 $r\,\hat{\mathbf{r}}$ 分解为垂直于 xy 平面的分量 $z = r\cos\theta$ 和 xy 平面的分量 $r\sin\theta$．后者又可以进而分解成 x 分量和 y 分量 $x = r\sin\theta\cos\phi$，$y = r\sin\theta\sin\phi$，这就得到了式 1．

在直角坐标系中，有 $r = \sqrt{x^2+y^2+z^2}$，代入式 1 中的三条关系，就可以很容易解出式 2 中的三条关系．

现在推导变换关系（式 3）．由于 $\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}$ 都是关于 (r,θ,ϕ) 的函数，所以在考察一点 (r,θ,ϕ) 时，$\hat{\mathbf{r}}$ 的球坐标是 $(1,\theta,\phi)$，根据式 1 变换到直角坐标为

$$(\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta) \tag{10}$$

写成矢量的形式，就是

$$\hat{\mathbf{r}} = \sin\theta\cos\phi\,\hat{\mathbf{x}} + \sin\theta\sin\phi\,\hat{\mathbf{y}} + \cos\theta\,\hat{\mathbf{z}} \tag{11}$$

至于式 3 的第二条式子，在同一个球坐标 (r,θ,ϕ) 处，$\hat{\boldsymbol{\theta}}$ 的球坐标为 $(1,\theta+\pi/2,\phi)$，根据式 1 变换到直角坐标再化简就得到直角坐标和对应的矢量形式为

$$(\cos\theta\cos\phi,\ \cos\theta\sin\phi,\ -\sin\theta) \tag{12}$$

$$\hat{\boldsymbol{\theta}} = \cos\theta\cos\phi\,\hat{\mathbf{x}} + \cos\theta\sin\phi\,\hat{\mathbf{y}} - \sin\theta\,\hat{\mathbf{z}} \tag{13}$$

同理可得式 5．将基底变换式 3 和式 4 分别代入式 7 和式 6 得坐标变换式 9 和式 8．

两方向的夹角

预备知识 内积[383]

若已知球坐标系中两个方向分别为 $(1,\theta_1,\phi_1)$ 和 $(1,\theta_2,\phi_2)$ 如何求它们之间的夹角 α 呢？我们可以先计算两个单位矢量的直角坐标，然后对它们进行内积即可得到两矢量夹角的余弦值．由式 1，两矢量的直角坐标分别为

$$(\sin\theta_1\cos\phi_1,\ \sin\theta_1\sin\phi_1,\ \cos\theta_1) \qquad (\sin\theta_2\cos\phi_2,\ \sin\theta_2\sin\phi_2,\ \cos\theta_2) \tag{14}$$

利用三角恒等式（式 4 [112]），得

$$\begin{aligned}\cos\alpha &= \sin\theta_1\cos\phi_1\sin\theta_2\cos\phi_2 + \sin\theta_1\sin\phi_1\sin\theta_2\sin\phi_2 + \cos\theta_1\cos\theta_2 \\ &= \sin\theta_1\sin\theta_2(\cos\phi_1\cos\phi_2 + \sin\phi_1\sin\phi_2) + \cos\theta_1\cos\theta_2 \\ &= \sin\theta_1\sin\theta_2\cos(\phi_2-\phi_1) + \cos\theta_1\cos\theta_2\end{aligned} \tag{15}$$

圆锥曲线的极坐标方程

预备知识 极坐标的定义[119]

圆锥曲线的极坐标方程是 r 关于 θ 的函数

$$r(\theta) = \frac{p}{1 - e\cos\theta} \tag{1}$$

其中 e 是离心率，p 是半通径.

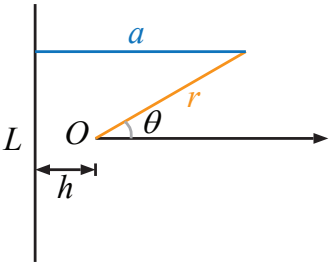


图 1: 由离心率定义圆锥曲线

推导

圆锥曲线的一种定义（与其他定义等效）为（图 1）：平面上有一点 O 和一条直线 L，相距为 h. 平面上某一点到 O 的距离为 r，到 L 的（垂直）距离为 a，令常数 $e > 0$，则所有满足

$$r/a = e \tag{2}$$

的点组成的曲线就是圆锥曲线．e 是常数，叫做**离心率**，O 是**焦点**，L 是**准线**．当 $e=0$ 时曲线是圆[14]，$0<e<1$ 时是椭圆，$e=1$ 时是抛物线，$e>1$ 时是双曲线．

以 O 点为原点，使极轴垂直于准线（如上图）．则 $a=h+r\cos\theta$，根据式 2 得

$$\frac{r}{h+r\cos\theta}=e \tag{3}$$

变形，得

$$r(\theta)=\frac{eh}{1-e\cos\theta} \tag{4}$$

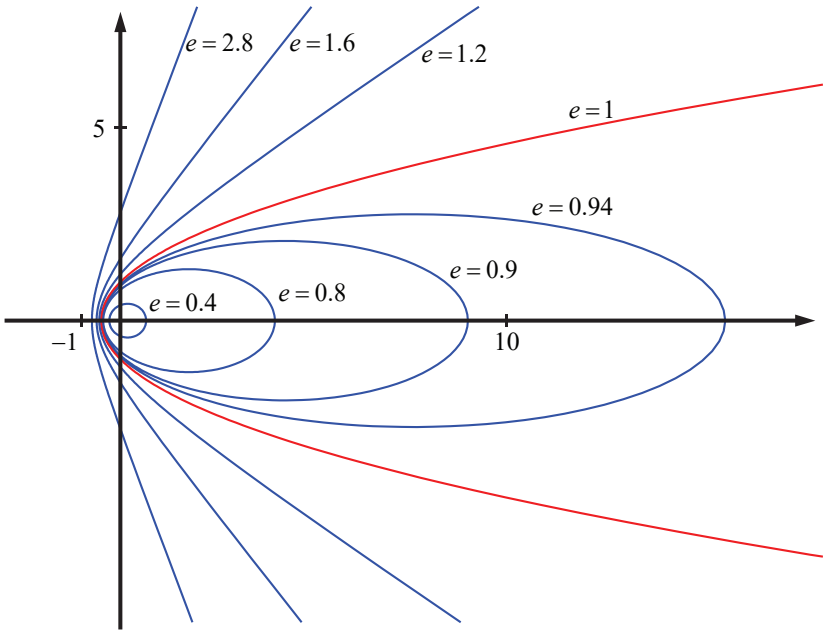


图 2: 不同离心率 e 的圆锥曲线（$h=1$）

若定义圆锥曲线的**通径**为过焦点且平行于准线的直线被圆锥曲线截出的线段长度．记**半通径**为 p，则通径为 $2p$，那么有 $r(\pi/2)=p$．代入式 4 得 $p=eh$．所以式 4 又可以写为

$$r(\theta)=\frac{p}{1-e\cos\theta} \tag{5}$$

注意 p 和 e 分别控制圆锥曲线的大小和形状．由于抛物线的 $e=1$ 不变，所以所有抛物线的形状都相同．

[14]注意根据定义，圆的准线为无穷远，所以只能使用式 1 而不能用式 4．所以在图 2 中，圆的半径为无穷小．

式 5 中一种比较特殊的情况是当圆锥曲线为双曲线（$e>1$）且 $1-e\cos\theta<0$ 时 r 取负值，会产生双曲线的左半支（即离焦点较远的一支，图中未画出）. 左半支上的任意一点同样满足式 2. 若只需要在极坐标中表示较远的一支，我们可以将式 5 中的 r 替换为 $-r$，θ 替换为 $\theta+\pi$，这样，就得到了这支双曲线的正常极坐标方程（$r>0$）

$$r(\theta)=-\frac{p}{1+e\cos\theta} \tag{6}$$

椭圆的三种定义

预备知识 圆锥曲线的极坐标方程[131]

第二种定义

我们已经知道用焦点和准线如何定义椭圆，但从椭圆的极坐标公式难以看出椭圆的对称性，这里用相同的定义推导直角坐标的表达式. 我们不妨先以一个焦点为原点定义直角坐标系，且令 x 轴指向另一个焦点，则有

$$r=\sqrt{x^2+y^2}\qquad \cos\theta=\frac{x}{\sqrt{x^2+y^2}} \tag{1}$$

代入椭圆的极坐标方程式 1 [131] 得

$$\sqrt{x^2+y^2}=p+ex \tag{2}$$

两边平方并整理得

$$(1-e^2)\left(x-\frac{ep}{1-e^2}\right)^2+y^2=\frac{p^2}{1-e^2} \tag{3}$$

由此可见，如果我们把椭圆左移 $ep/(1-e^2)$，椭圆将具有

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \tag{4}$$

的形式. 其中 a 为**半长轴**，b 为**半短轴**. 这就是椭圆的第二种定义，即把单位圆沿两个垂直方向分别均匀拉长 a 和 b. 下面来看系数的关系. 首先定义椭圆的焦距为焦点到椭圆中心的距离（即以上左移的距离）为

$$c=\frac{ep}{1-e^2} \tag{5}$$

式 3 和式 4 对比系数得

$$a = \frac{p}{1-e^2} \qquad b = \frac{p}{\sqrt{1-e^2}} \tag{6}$$

以上两式可以将椭圆的极坐标方程转为直角坐标方程．另外易证

$$a^2 = b^2 + c^2 \tag{7}$$

若要从直角坐标方程变回极坐标方程，将式 5 式 6 逆转得

$$e = \frac{c}{a} \qquad p = \frac{b^2}{a} \tag{8}$$

第三种定义

椭圆的第三种定义是，椭圆上任意一点到两焦点的距离之和等于 $2a$．现在我们来证明前两种定义下的椭圆满足这个条件．由直角坐标方程可知对称性，可在椭圆的两边做两条准线，令椭圆上任意一点到两焦点的距离分别为 r_1 和 r_2，到两准线的距离分别为 d_1 和 d_2，则有

$$e = \frac{r_1}{d_1} = \frac{r_2}{d_2} = \frac{r_1 + r_2}{d_1 + d_2} \tag{9}$$

所以

$$r_1 + r_2 = e(d_1 + d_2) = 2e(c + h) = 2\frac{c}{a}\left(c + \frac{b^2}{c}\right) = 2a \tag{10}$$

证毕．

双曲线的三种定义

预备知识 圆锥曲线的极坐标方程[131]

第二种定义

我们已经知道用焦点和准线如何定义双曲线，双曲线的极坐标方程为（$e > 1$）

$$r = \frac{p}{1 - e\cos\theta} \tag{1}$$

以与极坐标系相同的原点建立直角坐标系，要把以上方程变到直角坐标系中，将 $r=\sqrt{x^2+y^2}$，$\cos\theta = x/\sqrt{x^2+y^2}$ 代入得

$$\sqrt{x^2+y^2} = p + ex \tag{2}$$

两边平方且化简得

$$\frac{(e^2-1)^2}{p^2}\left(x+\frac{ep}{e^2-1}\right)^2 - \frac{e^2-1}{p^2}y^2 = 1 \tag{3}$$

把双曲线沿 x 轴正方向移动 c，可得以下形式

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \tag{4}$$

这就是双曲线的第二种定义，从上式容易看出，双曲线的两支是左右对称的. 以上两式对比系数得

$$a = \frac{p}{e^2-1} \quad b = \frac{p}{\sqrt{e^2-1}} \quad c = \frac{ep}{e^2-1} \tag{5}$$

$$c^2 = a^2 + b^2 \tag{6}$$

用 a,b,c 表示 e,p 有

$$e = \frac{c}{a} \qquad p = \frac{b^2}{a} \tag{7}$$

由离心率的定义，双曲线的焦点到准线的距离为 $p/e = b^2/c$，准线的坐标为 $c - p/e = a^2/c$. 由对称性，双曲线有两个焦点和两条准线，任意一个焦点到双曲线两支的任意一点比上该点到焦点同侧准线的距离都等于离心率.

第三种定义

双曲线的另一种定义是，曲线上任意一点到两个焦点距离之差等于 $2a$. 这里证明前两种定义满足该性质. 由对称性，我们不妨只考虑右支上的某点，令其到右焦点和右准线的距离分别为 r_1 和 d_1，到左焦点和左准线的距离分别为 r_2 和 d_2. 由离心率的定义，有

$$e = \frac{r_1}{d_1} = \frac{r_2}{d_2} = \frac{r_2 - r_1}{d_2 - d_1} \tag{8}$$

由于两准线之间的距离恒为 $2a^2/c$，上式变为

$$r_2 - r_1 = e(d_2 - d_1) = 2a \tag{9}$$

证毕.

渐近线

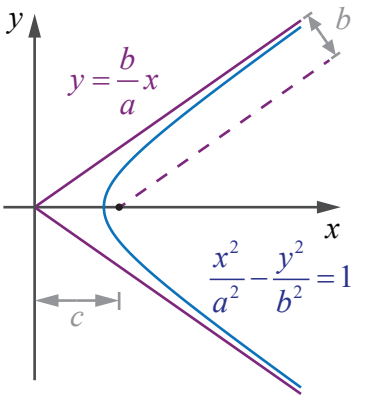


图 1: 双曲线的渐近线

当 x, y 都无穷大时，式 4 中的 1 可以忽略不计，有 $y/x = \pm b/a$，渐近线与 x 轴夹角为

$$\theta_0 = \arctan(b/a) \tag{10}$$

两条渐近线到两个焦点的距离都为

$$c \sin \theta_0 = c \cdot b/c = b \tag{11}$$

事实上这么推导渐近线并不严谨，在学习了高数的相关内容（见“泰勒展开[181]”）后，由式 4 得

$$y = \frac{bx}{a}\sqrt{1-\frac{a^2}{x^2}} \tag{12}$$

把根号部分关于 a^2/x^2 进行泰勒展开，有

$$y = \frac{b}{a}x - \frac{ab}{2x} + \mathcal{O}\left(\frac{1}{x^3}\right) \tag{13}$$

所以当 $x \to \infty$ 时，就有渐进线 $y = bx/a$. 之所以要这样做，是为了防止式 13 右边出现常数项. 如果存在常数项 λ，那么双曲线的渐近线就是 $y = bx/a + \lambda$ 了.

抛物线的三种定义

预备知识 圆锥曲线的极坐标方程[131]

第二种定义

我们已经知道用焦点和准线如何定义抛物线和其他圆锥曲线（式 1 [131]），抛物线的离心率 $e=1$，所以极坐标方程为

$$r=\frac{p}{1-\cos\theta} \tag{1}$$

以与极坐标系相同的原点建立直角坐标系，要把以上方程变到直角坐标系中，将 $r=\sqrt{x^2+y^2}$，$\cos\theta=x/\sqrt{x^2+y^2}$ 代入得

$$\sqrt{x^2+y^2}=p+x \tag{2}$$

两边平方并化简得到

$$y^2=2p\left(x+\frac{p}{2}\right) \tag{3}$$

把双曲线沿 x 轴正方向移动 $p/2$，可得标准抛物线方程

$$y^2=2px \tag{4}$$

所以抛物线的焦距为 $f=p/2$．与椭圆和双曲线不同的是，所有的抛物线的形状都相似（形状相同，大小不同），这是因为抛物线有固定的离心率（离心率决定圆锥曲线的形状，焦距或准线决定大小）．

第三种定义

在 x 轴正半轴作一条与准线平行的直线 L，则抛物线上一点 P 到其焦点的距离 r 与 P 到 L 的距离之和不变．

如图 1，要证明由焦点和准线定义的抛物线满足该性质，只需过点 P 作从准线到直线 L 的垂直线段 AB，由于 r 等于线段 PA 的长度，所以 r 加上 PB 的长度等于 AB 的长度，与 P 的位置无关．证毕．

圆锥曲线的光学性质

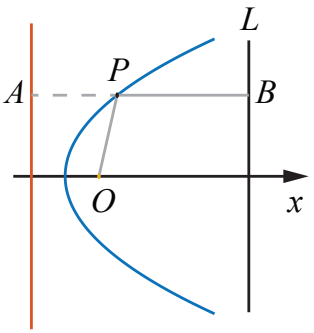


图 1: 抛物线的第三种定义

预备知识 椭圆的三种定义[133]，抛物线的三种定义[136]，双曲线的三种定义[134]

首先，对于本词条涉及到的部分几何概念作出如下规定：

- **角度**：所有角大小均取弧度制 $[0,\pi]$ 范围内的值（不区分始边终边，不区分正负角）；
- **补角**：角的其中一条边的反向延长线与该角的另一条边所形成的角，称为该角的补角（如图 1，$\angle A'BC$ 是 $\angle ABC$ 的补角，$\angle A'BC+\angle ABC=\pi$）；

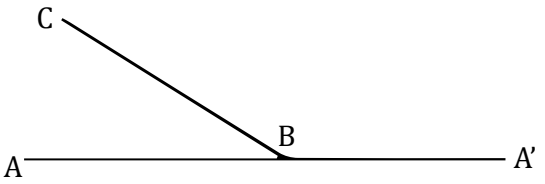


图 1: 补角

- **角平分线**：到角两边的距离相等的点的集合（把角分为大小相等的两部分的直线），称为该角的角平分线.

对两种特殊角的情形作出如下补充：

- 零角（如图 2，$\angle ABC=0$）的补角为平角（$\angle ABA'=\pi$），零角的平分线与角两边重合（直线 AA' 即 $\angle ABC$ 的平分线）；

- 平角（如图 2，$\angle ABA' = \pi$）的补角为零角（$\angle ABC = 0$），平角的平分线与角两边垂直（AA' 的垂线 BN 即 $\angle ABA'$ 的平分线）.

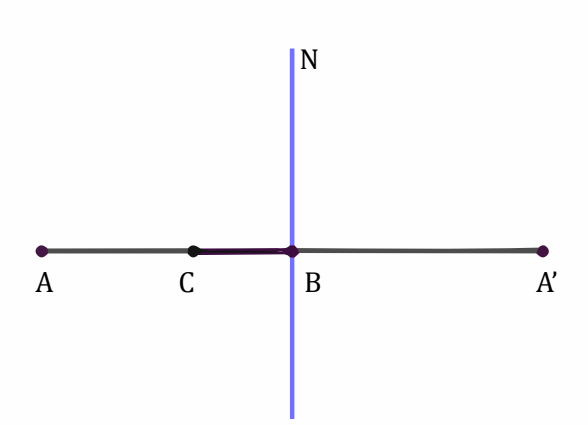


图 2: 零角和平角

显然，对所有角而言，**角平分线与该角的补角平分线互相垂直**.

椭圆的光学性质

F_1，F_2 是椭圆的两个焦点，P 是椭圆上任意一点，则 $\angle F_1PF_2$ 的补角平分线 PT 是椭圆的切线.

几何法证明

作点 F_2 关于直线 PT 的对称点 F_2'. 由于 PT 是 $\angle F_1PF_2$ 的补角平分线，则 F_2' 在 F_1P 的延长线上.

记 Q 是直线 PT 上的任意一点. 于是

$$F_1Q + F_2Q = F_1Q + F_2'Q \geqslant F_1F_2' = F_1P + F_2P \tag{1}$$

当且仅当 P，Q 重合时，不等式（式 1）为等式.

椭圆的一种定义[133] 为：平面上到两焦点的距离之和为定值的点的集合. 显然，椭圆内任意一点到两焦点距离之和小于该定值，而椭圆外任意一点到两焦点距离之和大于该定值.

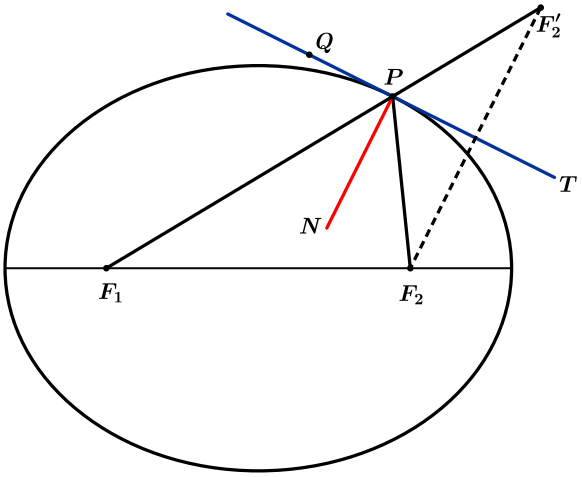


图 3: 椭圆的光学性质

所以，直线 PT 上有且仅有点 P 在椭圆上，其他点都在椭圆外．这就证明了 PT 是椭圆的切线．

等价的命题

结合光在同一介质中直线传播的性质，以及光的反射定律——反射角等于入射角，不难推知，上述几何命题等价于物理命题“真空或同种均匀介质中，从椭圆一个焦点处射出的光线经过在椭圆曲线上的反射后，反射光线都汇聚于另一个焦点”.

解析法推导

记椭圆的直角坐标方程为

$$b^2x^2 + a^2y^2 = a^2b^2 \tag{2}$$

其中 $a > b > 0, c > 0$ 且 $c^2 = a^2 - b^2$，椭圆焦距为 $2c$.

将 y 视作 x 的函数，对方程式 2 等号两边关于 x 求导，可得

$$b^2x + a^2yy' = 0 \tag{3}$$

$P(x_0, y_0)$ 是椭圆上任意一点，则点 P 处的椭圆切线方程为

$$a^2y_0(y - y_0) = -b^2x_0(x - x_0) \quad \Rightarrow \quad b^2x_0x + a^2y_0y = a^2b^2 \tag{4}$$

由切线方程可得椭圆在点 P 处的一个法向量（与切线垂直的向量）为 $\mathbf{n}=(b^2x_0,a^2y_0)$.

向量 $\overrightarrow{F_1P}=(x_0+c,y_0)$, $\overrightarrow{F_2P}=(x_0-c,y_0)$.

记两个任意向量（$\mathbf{s}$ 和 $\mathbf{t}$）的夹角为 $<\mathbf{s},\mathbf{t}>$. 则

$$\cos<\mathbf{n},\overrightarrow{F_1P}>=\frac{\mathbf{n}\cdot\overrightarrow{F_1P}}{|\mathbf{n}|\,|\overrightarrow{F_1P}|}=\frac{b^2x_0(x_0+c)+a^2y_0^2}{\sqrt{b^4x_0^2+a^4y_0^2}\sqrt{(x_0+c)^2+y_0^2}} \tag{5}$$

$$\cos<\mathbf{n},\overrightarrow{F_2P}>=\frac{\mathbf{n}\cdot\overrightarrow{F_2P}}{|\mathbf{n}|\,|\overrightarrow{F_2P}|}=\frac{b^2x_0(x_0-c)+a^2y_0^2}{\sqrt{b^4x_0^2+a^4y_0^2}\sqrt{(x_0-c)^2+y_0^2}} \tag{6}$$

式 5 和式 6 作比，并代入式 2 及 $c^2=a^2-b^2$ 进行化简

$$\begin{aligned}\frac{\cos<\mathbf{n},\overrightarrow{F_1P}>}{\cos<\mathbf{n},\overrightarrow{F_2P}>}&=\frac{a^2b^2+b^2cx_0}{a^2b^2-b^2cx_0}\sqrt{\frac{(x_0-c)^2+y_0^2}{(x_0+c)^2+y_0^2}}\\&=\frac{a^2+cx_0}{a^2-cx_0}\sqrt{\frac{a^2(x_0-c)^2+a^2b^2-b^2x_0^2}{a^2(x_0+c)^2+a^2b^2-b^2x_0^2}}\\&=\frac{a^2+cx_0}{a^2-cx_0}\sqrt{\frac{c^2x_0^2+a^4-2a^2cx_0}{c^2x_0^2+a^4+2a^2cx_0}}\\&=\frac{a^2+cx_0}{a^2-cx_0}\sqrt{\frac{(a^2-cx_0)^2}{(a^2+cx_0)^2}}\end{aligned} \tag{7}$$

由于点 P 在椭圆上，则 $-a\leqslant x_0\leqslant a$，则 $a^2-cx_0>0$，因此

$$\frac{\cos<\mathbf{n},\overrightarrow{F_1P}>}{\cos<\mathbf{n},\overrightarrow{F_2P}>}=1 \tag{8}$$

所以 $<\mathbf{n},\overrightarrow{F_1P}>=<\mathbf{n},\overrightarrow{F_2P}>$，满足入射角等于反射角的反射定律. 由此用解析几何的方法推导出了椭圆的光学性质.

抛物线的光学性质

F 是抛物线的焦点，l 是准线，P 是抛物线上的任意一点，作 $PP'\perp l$，垂足为 P'，则 $\angle FPP'$ 的角平分线 PT 是抛物线的切线.

等价命题：真空或同种均匀介质中，从抛物线焦点射出的光线，经过抛物线曲线的反射后，反射光线平行于抛物线对称轴.

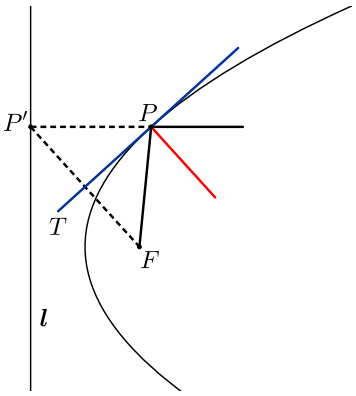


图 4: 抛物线的光学性质

双曲线的光学性质

F_1，F_2 是双曲线的两个焦点，P 是双曲线上的任意一点，则 $\angle F_1PF_2$ 的角平分线 PT 是双曲线的切线.

等价命题：真空或同种均匀介质中，从双曲线一个焦点射出的光线，经过双曲线的反射后，反射光的反向延长线汇聚于另一个焦点.

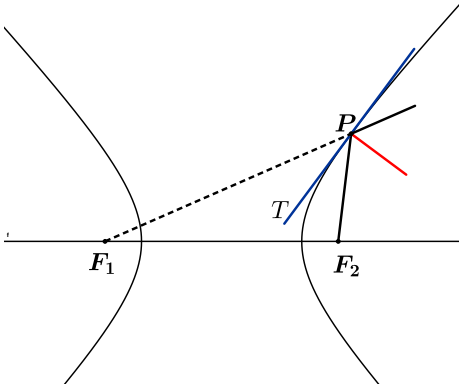


图 5: 双曲线的光学性质

习题 1　证明题

仿照椭圆光学性质的证明和推导过程，证明抛物线和双曲线的光学性质.

复数

预备知识 几何矢量的运算[375]，三角恒等式[112]，四象限 Arctan 函数[118]

定义 1　复数

复数（**complex number**）是一对满足以下运算的有序实数[15]．令 z 为复数，x, y 为实数，则可以表示为 $z = (x, y)$．其中 x, y 分别被称为复数 z 的实部（**real part**）和虚部（**imaginary part**），可以记为 $\mathrm{Re}[z]$ 和 $\mathrm{Im}[z]$．特殊地，我们把复数 $(0, 1)$ 称为虚数单位，用 i 表示[16]．最后我们定义虚部为零的复数 $(x, 0)$ 就是实数 x 本身．把所有复数的集合记为 $\mathbb{C}$，那么全体实数的集合 $\mathbb{R}$ 就是它的一个子真子集[92]，即 $\mathbb{R} \subset \mathbb{C}$.

定义两个复数的加法为实部和虚部分别相加

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2) \tag{1}$$

定义复数和实数 s 相乘为（满足交换律）

$$s(x, y) = (x, y)s = (sx, sy) \tag{2}$$

可见任意一个复数可以表示为一个加法和一个乘法：$(x, y) = (x, 0) + y(0, 1)$，即熟悉的

$$z = x + \mathrm{i}y \tag{3}$$

复平面

由此可以看到，复数跟二维平面上的几何矢量[370] 是十分相似的．如图 1，一个复数可以看做**复平面**上的一个点（或矢量），该矢量在复平面的**实轴**和**虚轴**方向的分量分别等于其实部和虚部．复数的**模**定义为对应矢量的模，即

$$|z| = \sqrt{\mathrm{Re}[z]^2 + \mathrm{Im}[z]^2} \tag{4}$$

另外我们把矢量与实轴的夹角称为**幅角**，记为 $\arg(z)$．我们可以通过 Arctan 函数（式 1 [118]）计算幅角

$$\arg(z) = \mathrm{Arctan}(\mathrm{Im}[z], \mathrm{Re}[z]) \qquad (\arg z \in (-\pi, \pi]) \tag{5}$$

[15] 一些教材常常先定义虚数单位 $\mathrm{i} = \sqrt{-1}$ 或 $\mathrm{i}^2 = -1$，这种定义往往不易理解．我们这里直接将复数定义为服从某种运算规则的实数对，更能揭示复数的本质 [8].

[16] 为了与变量 i 区分，本书中虚数单位使用正体的 i.

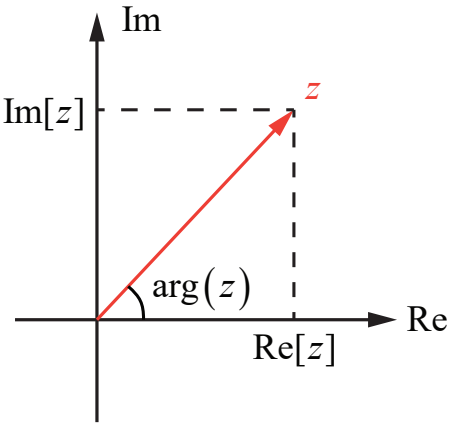


图 1: 复平面与复数

也可以通过模和幅角来计算实部与虚部

$$\mathrm{Re}[z] = |z|\cos(\arg z) \qquad \mathrm{Im}[z] = |z|\sin(\arg z) \tag{6}$$

在“指数函数（复数）[351]”中我们将看到，任意复数也可以通过欧拉公式表示为以下形式

$$z = A(\cos\theta + \mathrm{i}\sin\theta) = A\mathrm{e}^{\mathrm{i}\theta} \tag{7}$$

其中 $A = |z|$，$\theta = \arg z$.

基本运算

共轭

一个复数的共轭等于与其实部相同，虚部相反的复数[17]

$$z^* = \mathrm{Re}[z] - \mathrm{i}\ \mathrm{Im}[z] \tag{8}$$

所以共轭运算不改变复数的模，但将其幅角变为相反数．在复平面上，这相当于把一个点关于 x 轴取镜像对称．

加和减

式 1 中已经定义了加法．与实数相同，定义减法为 $z_1 - z_2 = z_1 + (-z_2)$，有

$$(x_1 + \mathrm{i}y_1) \pm (x_2 + \mathrm{i}y_2) = (x_1 \pm x_2) + \mathrm{i}(y_1 \pm y_2) \tag{9}$$

[17]一些教材也使用 $\bar{z}$ 表示 z 的共轭.

在复平面上，这相当于把两个复数对应的矢量进行矢量相加减[375]．显然，复数的加法满足**交换律**，**分配律**和**结合律**．

特殊地，将一个复数与其复共轭加减可得其实部和虚部

$$\mathrm{Re}[z]=\frac{1}{2}(z+z^*)\qquad \mathrm{Im}[z]=\frac{1}{2}(z-z^*) \tag{10}$$

乘法

两个复数相乘定义为（注意式 2 是该定义的一种特殊情况）

$$z_1z_2=(x_1+\mathrm{i}y_1)(x_2+\mathrm{i}y_2)=(x_1x_2-y_1y_2)+\mathrm{i}(x_1y_2+x_2y_1) \tag{11}$$

可以证明，乘积的模等于两复数模之积，乘积的幅角等于两复数的幅角之和，即

$$|z_1z_2|=|z_1||z_2| \tag{12}$$

$$\arg(z_1z_2)=\arg(z_1)+\arg(z_2) \tag{13}$$

证明：令 $A_i=|z_i|$，$\theta_i=\arg z_i$，则

$$\begin{aligned} z_1z_2&=(A_1\cos\theta_1+\mathrm{i}A_1\sin\theta_1)(A_2\cos\theta_2+\mathrm{i}A_2\sin\theta_2)\\ &=A_1A_2(\cos\theta_1\cos\theta_2-\sin\theta_1\sin\theta_2)+\mathrm{i}A_1A_2(\cos\theta_1\sin\theta_2+\cos\theta_2\sin\theta_1)\\ &=A_1A_2[\cos(\theta_1+\theta_2)+\mathrm{i}\sin(\theta_1+\theta_2)] \end{aligned} \tag{14}$$

其中最后一步用到了两角和公式（式 3 [112]）．容易看出，最后得到的是一个模为 A_1A_2，幅角为 $\theta_1+\theta_2$ 的复数．证毕．

不难证明复数的乘法满足**交换律**和**结合律**．容易证明，一个复数模的平方可以用它和复共轭的乘积表示．

$$x^2+y^2=|z|^2=zz^* \tag{15}$$

除法

和实数一样，复数的除法定义为乘法的逆运算．令 $z_1=zz_2$（$z_2\neq0$），则两个复数相除可以记为

$$z=\frac{z_1}{z_2}=\frac{x_1+\mathrm{i}y_1}{x_2+\mathrm{i}y_2} \tag{16}$$

但我们希望可以将结果的实部与虚部分开，于是我们可以对 z_1, z_2 同时乘以 z_2^*，即 $z_1 z_2^* = z z_2 z_2^*$，或

$$z = \frac{z_1 z_2^*}{z_2 z_2^*} = \frac{(x_1 + \mathrm{i}y_1)(x_2 - \mathrm{i}y_2)}{(x_2 + \mathrm{i}y_2)(x_2 - \mathrm{i}y_2)} = \frac{x_1 x_2 + y_1 y_2}{x_2^2 + y_2^2} + \mathrm{i}\frac{x_2 y_1 - x_1 y_2}{x_2^2 + y_2^2} \tag{17}$$

这个步骤叫做**分母有理化**.

与乘法同理，两个复数相除相当于把它们的模相除，幅角相减，即

$$|z_1/z_2| = |z_1|/|z_2| \tag{18}$$

$$\arg(z_1/z_2) = \arg(z_1) - \arg(z_2) \tag{19}$$

定理 1

两个复数进行任意次加减乘除后再取共轭，等于它们分别取共轭后再进行运算.

根据定义易证. 例如

$$\frac{2z_1 z_2}{(z_3 + z_4)^2} = \frac{2z_1^* z_2^*}{(z_3^* + z_4^*)^2} \tag{20}$$

根据式 15，式 10 和定理 1 易得

$$\begin{aligned} |z_1 + z_2|^2 &= |z_1|^2 + |z_2|^2 + z_1^* z_2 + z_2^* z_1 \\ &= |z_1|^2 + |z_2|^2 + 2\,\mathrm{Re}[z_1^* z_2] \end{aligned} \tag{21}$$

在复平面中，该式可以表示余弦定理，即计算两矢量之和的模.

第二章

一元微积分

微积分导航

从物理学巨人牛顿发明了微积分以来[1]，微积分就在物理学的各个方面被大量使用．高中的物理教学有意避开了使用微积分，但从本科的学习开始，微积分与物理将形影不离．不夸张地说，不懂微积分，就几乎不懂物理．微积分最核心的内容就是极限，求导/微分，积分，微分方程和无穷级数．

本章介绍后面物理部分需要使用的微积分内容，并以帮助读者理解为主而不求严谨详尽．对于一些定理如自然对数底极限和非整数幂的二项式定理，本章只给出数值验证的方法．

极限

极限[149] 的概念是微积分的基础，大致可以理解为“某个表达式在某个量为无穷小或无穷大时所趋近的值”，例如 $1/x$ 在 $x \to \infty$ 时的极限为零，$(1+x)/(2+x)$ 在 $x \to 0$ 时的极限为 $1/2$．

导数

理解极限了以后，导数[158] 便是一个首要的应用．事实上高中物理的许多物理量都使用了导数的概念，只是没有提出“导数”这个词．例如（瞬时）速度的定义就是 $\Delta\mathbf{s}/\Delta t$ 在 $\Delta t \to 0$（趋近于 0）时的极限，而这恰好是导数的定义，即速度是位置矢量（关于时间的函数）$\mathbf{r}(t)$ 对时间的导数．同理，加速度矢量是速度矢量（关于时间的函数）对时间求导．又例如，高中对匀速圆周运动的向心加速度的推导过程中就运用了几何微元法，在微小时间 Δt 内计算圆周运动速度矢量的微小变化[885]．学习了矢量求导[274] 以后，就不必再使用这种不成熟的“几何微元法”，而是直接按照矢量求导法则[885] 即可严谨而轻易地得出向心加速度的公式 $\mathbf{a} = -\omega^2\mathbf{R}$，甚至可以计算非匀速圆周运动乃至任意变速曲线运动的加速度．

[1] 一般认为牛顿和莱布尼兹都分别在十七世纪中独立地发明了微积分，然而他们都声称对方窃取了自己的成果，并为此争执了一生．

积分

高中物理中，位移 $\mathbf{s}$ 等于速度 $\mathbf{v}$ 乘以时间 t，功 W 等于力 F 乘以位移 x 等概念都已经耳熟能详．然而如果速度随时间变化或者力随位置变化时，就不能用简单的乘法来计算这些问题．这时一个基本的思想就是把时间或位移分成许多小份，每份中的速度或力都近似为恒定不变，然后再把所有小份的位移或做功加起来即可．这时用极限的思想，求出当这些小份为无穷小（或者说分成无穷多份）时求和的极限，就得到了总位移和总功，这个过程叫做定积分[198]．

微分方程

大量的物理定律和问题都是通过微分方程（组）来描述的．最简单的微分方程是线性常微分方程，是函数 $y(x)$ 及不同阶导 $y'(x)$，$y''(x)$ 以及自变量 x 组成的等式．例如力学中著名的弹簧振子[913]（又称简谐振子）模型就是通过二阶线性常微分方程（二阶代表方程中出现的最高阶导数为 2）来描述的．

极限

数列的极限

先来看一个数列的例子．

例 1

我们都知道 π 是一个无理数，所以 π 的小数部分是无限多的．目前用计算机，已经可以将 π 精确地计算到小数点后数亿位．然而在实际应用中，往往只用取前几位小数的近似即可．下面给出一个数列，定义第 n 项是 π 的前 n 位小数近似（不考虑四舍五入），即

$$a_0 = 3,\ a_1 = 3.1,\ a_2 = 3.14,\ a_3 = 3.141,\ \ldots. \tag{1}$$

这个数列显而易见的性质，就是当 n **趋于无穷**时，a_n 趋（近）于 π．无穷通常用符号 ∞ 来表示（像“8”横过来写）．我们把这类过程叫做**极限**．以上这种情况，用极限符号表示，就是

$$\lim_{n\to\infty} a_n = \pi \tag{2}$$

这里 lim 是极限（limit）的意思，下方用箭头表示某个量变化的趋势．$\lim\limits_{n\to\infty}$ 在这里相当于一个“操作”，叫**算符（operator）**，a_n 是其作用的对象（相当于函数的自变量）．算符的“输出”就是一个数（a_n 的极限值）．所以不要误以为这条式子是说当 $n\to\infty$ 时，$a_n=\pi$（a_n 是有理数，π 是无理数，等式恒不成立），而要理解成 a_n 经过算符 $\lim\limits_{n\to\infty}$ 的作用以后，得出其极限是 π．类比函数 $\sin x=y$，并不是说 $x=y$，而是说 x 经过正弦函数作用后等于 y．

所以从概念上来说，极限中的“趋于”和“等于”是不同的．趋于更强调变化的过程．趋于的意思可以粗略理解为

- 越来越接近，但不一定相等
- （在不相等的情况下）只有更近，没有最近

对极限来说，第 2 点成立是非常必要的．但是怎样能说明“没有最近”呢？可以看出，当 n 越大，a_n 越接近 π，它们的“距离”，可以用 $|a_n-\pi|$ 来表示．也就是说，对任何一个 a_n，如果所对应的距离 $|a_n-\pi|\neq 0$，总能找到一个更大的数 $m>n$，使 $|a_m-\pi|<|a_n-\pi|$（更近），并且要求之后的所有项都能满足这一条件．只有这样，才能从数学上说明上面两个意思．这就是极限思想的精髓．根据这个思想，下面可以写出数列极限的定义．

数列极限的定义：对于任意给定的 ε（无论它有多么小），总存在 N，当 $n>N$ 时，就有 $|a_n-A|<\varepsilon$（A 为常数）成立，那么数列 a_n 的极限就是 A．

在命题中，通常把“任意”用“$\forall$”（for all）表示，把“存在”用“$\exists$”（there exist(s), there is (are)）表示．即“对 $\forall\varepsilon$，$\exists N$，当 $n>N$ 时，有 $|a_n-A|<\varepsilon$”．

由于以上讨论中 lim 作用的对象是数列，那么箭头右边只能是 ∞（准确来说应该是正无穷 $+\infty$，但是由于数列的项一般是正的，所以正号省略了）．

把定义套用到上面的例一中，如果要求 $|a_n-\pi|<10^{-3}$（给定 $\varepsilon=10^{-3}$），只要令 $N=3$（当然也可以令 $N=4,N=5$，等）就可以保证第 N 项后面所有的项都满足要求．一般地如果给定 $\varepsilon=b\times 10^{-q}(b>1)$，就令 $N=q$，第 N 项以后的项就满足要求．这就从定义直接证明了 $\lim\limits_{n\to\infty}a_n=\pi$．

函数的极限

类比数列的极限，我们也可以定义**函数在正无穷的极限** $\lim\limits_{x\to+\infty}f(x)=A$，即“对 $\forall\varepsilon$，$\exists X$，当 $x>X$ 时，有 $|f(x)-A|<\varepsilon$”．

与数列不同的是，对于函数我们还可以定义**函数在负无穷的极限** $\lim\limits_{x\to-\infty} f(x)$（把以上定义的 $>$ 号改成 $<$ 号即可）.

另外可以定义 $f(x)$ **在** x_0 **处的极限** A，即“对 $\forall\varepsilon>0$，$\exists\delta>0$，当 $|x-x_0|<\delta$ 时，有 $|f(x)-A|<\varepsilon$”. 注意 $f(x)$ 不需要在 x_0 处有定义.

例 2

求函数在某个值处的极限时，通常可以直接代入数值计算，如

$$\lim_{x\to1} 2x+1=3 \qquad \lim_{x\to2}\frac{x+1}{x+2}=\frac{3}{4} \tag{3}$$

当无穷大与常数相加时，可以忽略常数，如

$$\lim_{x\to+\infty}\frac{x+1}{2x+2}=\lim_{x\to+\infty}\frac{x}{2x}=\frac{1}{2} \tag{4}$$

无穷小的阶

如果令 $x\to0$，我们就说 x 是**无穷小**. 但一些无穷小会更快地趋近于 0，若 x 的某个函数 $\alpha(x)$ 满足

$$\lim_{x\to0}\frac{\alpha(x)}{x}=0 \tag{5}$$

那 $\alpha(x)$ 就是 x 的**高阶无穷小**. 若

$$\lim_{x\to0}\frac{\alpha(x)}{x^n}\neq0 \tag{6}$$

则称 $\alpha(x)$ 为 x 的 n 阶无穷小. 例如，cx^n（c 为常数）就是 x 的 n **阶无穷小**，记为 $\mathcal{O}(x^n)$.

在求极限时，若高阶无穷小与低阶无穷小相加，通常可以忽略高阶无穷小. 另外由定义不难推出

$$\mathcal{O}(x^n)x^m=\mathcal{O}\big(x^{n+m}\big) \qquad (m>-n) \tag{7}$$

在物理中，当我们用一个函数 $g(x)$ 来近似另一个函数 $f(x)$ 并记为 $f(x)=g(x)+\mathcal{O}(h^n)$ 时（这里 x 是函数的自变量，h 是函数表达式中一个较小的常数），就说 $g(x)$ 的误差为 $\mathcal{O}(h^n)$.

小角正弦极限

预备知识 极限[149]

这里要介绍的是一个很显然的几何问题，然而它在高等数学和物理中却非常频繁地出现.

设平面上 O 点为圆心，以 R 作为半径画圆．取一段的圆心角为 θ 的圆弧 AB （令长为 l），并作线段 AB （如图 1）.

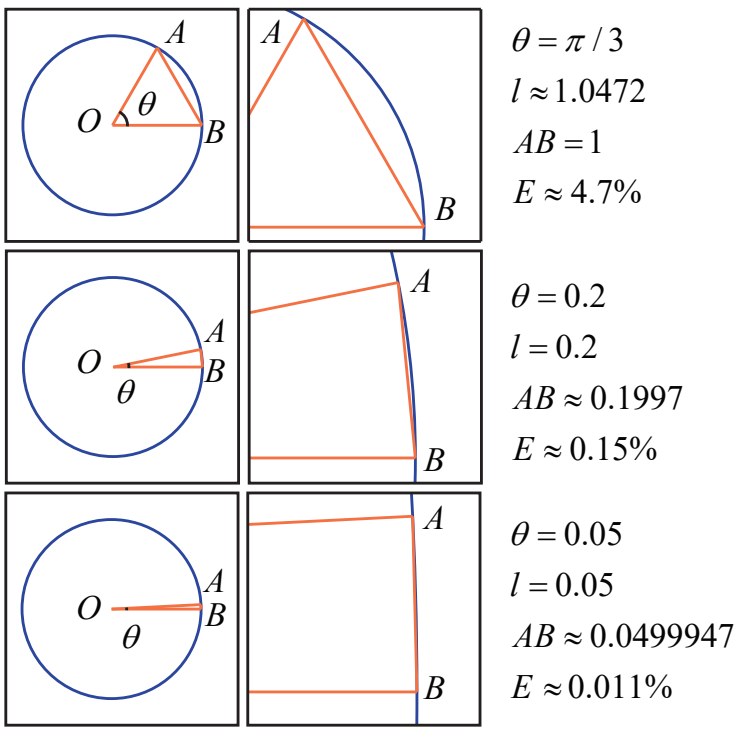


图 1: 单位圆中，随着角度不断减小，弧长与线段长度的相对误差也不断减小

由弧长公式得

$$l = R\theta \tag{1}$$

线段 AB 的长度为

$$AB = 2R\sin\frac{\theta}{2} \tag{2}$$

显然弧长 l 大于线段长度 AB （两点之间直线最短），但从图中可以看出随着 θ 越来越小，二者的相对误差（E）越来越小．用极限[149] 的语言来说，就是当 θ **趋近于** 0 时，它们的比值**趋近于** 1．注意这只是一个经验上的总结，证明参考任意高等数学教材．

所以有

$$1=\lim_{\theta\to0}\frac{AB}{l}=\lim_{\theta\to0}\frac{2R\sin(\theta/2)}{R\theta}=\lim_{\theta\to0}\frac{\sin(\theta/2)}{\theta/2} \tag{3}$$

令 $x=\theta/2$，有

$$\lim_{\theta\to0}\frac{\sin x}{x}=1 \tag{4}$$

这是一个非常重要的极限．在物理中，我们常常会就某个小角使用近似 $\sin x\approx x$，例如“单摆[917]”以及“双缝干涉[1114]”．严格来说，这是一个一阶近似，x 就是 $\sin x$ 的泰勒展开（式 3 [181]）的第一项．

自然对数底

预备知识 极限[149]

微积分中有一个重要的极限，极限值是一个无理数，叫做**自然对数底**，记为[2] e．

$$\mathrm{e}\equiv\lim_{x\to0}(1+x)^{\frac{1}{x}}=2.71828\ldots \tag{1}$$

e 也可以用无穷级数表示为

$$\mathrm{e}=\sum_{n=0}^{\infty}\frac{1}{n!}=1+1+\frac{1}{2}+\frac{1}{6}+\ldots \tag{2}$$

数值验证

这里先用数值的方法验证式 1 ，首先我们可以画出 $(1+x)^{1/x}$ 在原点附近的函数图，注意当 $x=0$ 时，该函数无定义，但这并不妨碍极限的存在．可以看到，无论 x 从左边还是右边趋近于原点（即左极限和右极限），结果都相等．

表 1 用数值计算验证式 1 的右极限．

[2] 为了与其他变量区分，本书使用正体字母表示自然对数底．

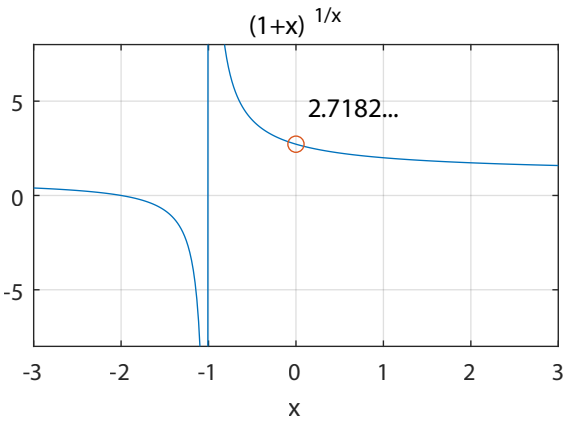


图 1: $(1+x)^{1/x}$ 的函数图

表 1: 极限 e 数值验证（保留 6 位有效数字）

x	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
$(1+x)^{1/x}$	2.59374	2.70481	2.71692	2.71815	2.71827	2.71828

自然对数函数

以 e 为底的对数函数 $\log_{\mathrm{e}} x$ 叫做**自然对数**，通常记为

$$\ln x \qquad 或 \qquad \log x \tag{3}$$

函数图如图 2 .

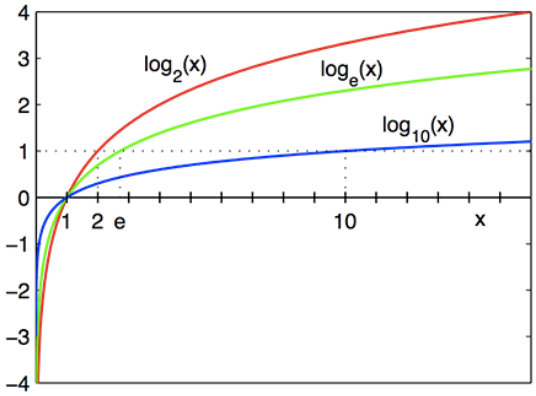


图 2: 几种不同底的对数函数

切线与割线

预备知识 极限[149]

如图一，在一段光滑曲线上任取两点，过这两点做直线，就是曲线过 A 点与 B 点的割线（当然直线与曲线还可以有其他交点）．当 A，B 两点逐渐向 C

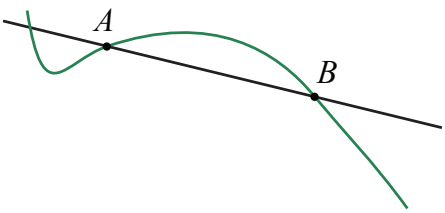


图 1: 割线

点靠近，割线的位置逐渐趋于不变，割线位置的极限[149] 就叫做曲线在 C 点的切线．

以上对切线的定义中，假设在 A，B 两点靠近 C 点的过程中，割线位置的极限存在．如果这个极限不存在，那么 C 点没有极限．下面举一个简单的例子说明．

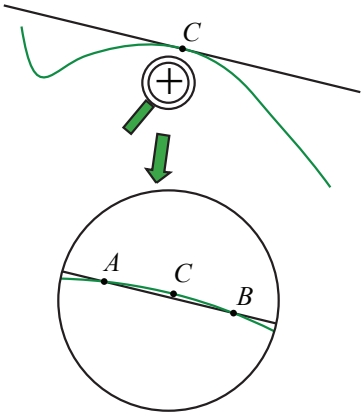


图 2: 割线的极限是切线

例如要求正方形一角的切线，用 A，B 两点接近 C 点，则无论 AB 点有多么靠近 C，切线的位置还要取决于 AB 点的具体位置（如右图）若 B 更接近 C，则直线就更接近竖直方向．反之直线就更接近水平方向．

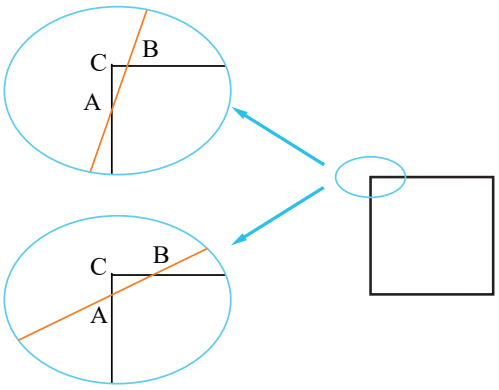


图 3: C 点不存在切线

而真正的极限，只取决于点 A，B 都趋于点 C 的事实，而不要求它们谁更趋近．所以这个极限不存在．

函数的连续性

预备知识 极限[149]

简单来说，**连续函数**定义为：在某个区间内，函数曲线是连续的．例如常见的 $\sin x$，$\exp x$，x^2 都在整个实数域上连续，又例如 $\ln x$ 和 $1/x$ 在区间 $(0,\infty)$ 上连续，$\tan x$ 在所有 $x_n=(1/2+n)\pi$（n 为自然数）处不连续，$1/x$ 在 $x=0$ 处不连续．但这只是一些简单的情况．在一些情况下这种判断方法则显得不严谨，例如 $\sin(1/x)$ 在原点处的连续性（不连续）根据这个定义不好判断．所以我们需要一个更严谨的定义．

首先我们要讨论函数在一个点附近是否是连续的．这个概念的思想核心是，在函数曲线的某一点附近 $(x_0,f(x_0))$，无论我们要求 $f(x)$ 有多接近 $f(x_0)$，只要 x 足够靠近 x_0，就一定能满足条件．比如说，如果定义函数 f 为当 $x<0$ 的时候，$f(x)=0$，其它时候 $f(x)=1$，那么在 $x=0$ 这一点处 f 就是跳跃的．如果我们要求的接近程度小于 1，那么无论 x 多接近 0，只要 $x<0$，$f(x)$ 和 $f(0)=1$ 的距离就永远满足不了需要．

准确地描述以上“连续”的概念，如下所示：

定义 1　函数在一点的连续性

函数 $f(x)$ 在某点 $x=x_0$ 处连续的定义是：函数 $f(x)$ 在某点 $x=x_0$ 处连续，当且仅当对于任何精度要求 $\epsilon>0$，我们都可以找到一个对应的范围 $\delta>0$，使得只要 $|x-x_0|<\delta$，就有 $|f(x)-f(x_0)|<\epsilon$. 如果一个函数在某区间的所有点都连续，我们就说它在这个区间连续.

另一个等价说法是常见的：

$$\lim_{x\to x_0} f(x)=f(x_0) \tag{1}$$

注意这里要求 x 从左边和右边趋近于 x_0 时的极限（即左极限和右极限）都成立.

一致连续

以上所定义的连续性是针对一个个点 x_0 而言的，就算函数在每一个点都连续，我们也只能说这个函数是**逐点连续的（pointwise continuous）**. 事实上，还有一种更强的连续性，它着眼于整体的性质，这就是**一致连续（uniformly continuous）**. 它的准确定义如下：

定义 2　一致连续

如果函数 $f(x)$ 在区间 S 上，对于任意精度 $\epsilon>0$，都存在对应的范围 $\delta>0$，使得只要 $|x_1-x_2|<\delta$，那么 $|f(x_1)-f(x_2)|<\epsilon$.

一致连续着眼于整个区间的性质，而不是一个个点. 显然，一致连续的函数肯定逐点连续，但是逐点连续的函数不一定一致连续，我们举一个反例.

例 1

例如考虑函数 $f(x)=1/x$，那么在区间 $(0,\infty)$ 上，f 就是逐点连续的，但它并不一致连续；对于同样的精度要求 $\epsilon>0$ 和任何范围 $\delta>0$，只要 $0<x_1<\delta$，那么就总有一个足够小的 x_2 使得 $|f(x_1)-f(x_2)|>\epsilon$，毕竟当 x_2 接近 0 的过程中，$f(x_2)$ 的斜率绝对值会趋近于无穷大.

习题 1　连续但不一致连续的函数

试证明 $1/x$ 在区间 $(0,+\infty]$ 以及 x^2 在 $\mathbb{R}$ 都是连续的，但不是一致连续的.

定理 1

设 S 是 $\mathbb{R}$ 的一个区间. 函数 f 在 S 上逐点连续的充分必要条件是，对于

任何开区间 $A \subset \mathbb{R}$，满足 $f(x) \in A$ 的所有 x 构成的集合，是若干开区间的并集．用紧凑的写法来表达就是，$f^{-1}(A)$ 是开区间的并集．

这个定理还可以等价地用闭区间来表达：f 在 S 上逐点连续的充分必要条件是，任何闭区间的逆映射是闭区间的并集．

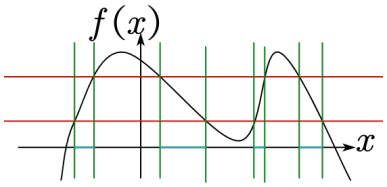


图 1: 如图，红色水平线在 $f(X)$ 轴上划分出了一个开/闭区间，绿色垂直线是取反函数的过程，x 轴上的靛蓝色线段就是取反函数的结果．从这个图可以直观地看出定理 1 的意义．

在实数轴上，开集被定义为任何开区间的并集，而闭集是开集的补集．如果 S 是 $\mathbb{R}$ 的一个子集，那么 S 上的开集被定义为 $\mathbb{R}$ 的开集和 S 的交集．这样一来，定理 1 就可以扩展为：f 在 S 上逐点连续，等价于对于任何开集 A，$f^{-1}(A) \cap S$ 是 S 上的开集，等价于对于任何闭集 B，$f^{-1}(B) \cap S$ 是 S 上的闭集．

用逆映射来刻画连续性是一个非常好用的方法．

导数

预备知识 切线与割线[155]，函数[103]

导数的几何理解

一个一元函数 $y = f(x)$，在直角坐标系中表示为一条曲线．在这个曲线的光滑部分取一点 A，并作其切线[155]．

若切线存在，该切线与 x 轴的夹角的正切值 θ 就叫点 A 的导数．当函数在 A 点递增时，可能的取值为 $\theta \in (0, \pi/2)$，即 $\tan\theta \in (0, +\infty)$．递减时，取 $\theta \in (-\pi/2, 0)$，即 $\tan\theta \in (-\infty, 0)$．当切线水平时，$\theta = \tan\theta = 0$．

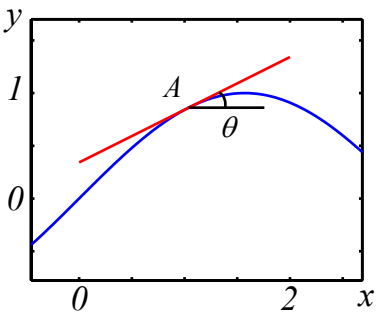


图 1: 点 A 的切线

若函数曲线在 x 的某一开区间内的每一点都可导，则这个区间上每一个 x 对应一个导数. 将其写成关于 x 的函数 $g(x)$，$g(x)$ 就是该区间上的 **导函数**. 通常将导函数记为以下的一种（后 3 种记号的来源见下文）

$$f'(x), \quad [f(x)]', \quad \frac{\mathrm{d}y}{\mathrm{d}x}, \quad \frac{\mathrm{d}f}{\mathrm{d}x}, \quad \frac{\mathrm{d}}{\mathrm{d}x}f(x) \tag{1}$$

在物理中，常常在物理量上方加一点表示对时间求导（注意仅限于对时间求导），例如 $\dot{f}(t) = \mathrm{d}f(t)/\mathrm{d}t$.

若切线不存在（例如折线的棱角处，但也有其他更复杂的情况），我们说点 A 不可导. 如果某区间内是“光滑”的，那么的该区间内处处可导.

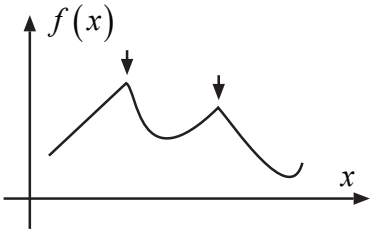


图 2: 棱角处不可导

若函数曲线在某一点附近是光滑的，那么在这点附近取一小段，当这一段取得足够小，可以近似认为它是线段且与切线重合（如下图）. 以这条线段为斜边，作一直角三角形，令其底边长为 $\mathrm{d}x$ （在微积分中，通常把非常小的一段 Δx 记为 $\mathrm{d}x$，$\mathrm{d}x$ 是一不能分割的整体符号，而不是两个量相乘），竖直边的边长为 $\mathrm{d}y$ （当函数递增时，$\mathrm{d}y$ 取正值，反之取负值）. 根据上面导数的定义，$\mathrm{d}y/\mathrm{d}x = \tan\theta$ 就是函数的导数. 所以导数通常表示为 $\mathrm{d}y/\mathrm{d}x$，导数的倒数则为 $\mathrm{d}x/\mathrm{d}y$.

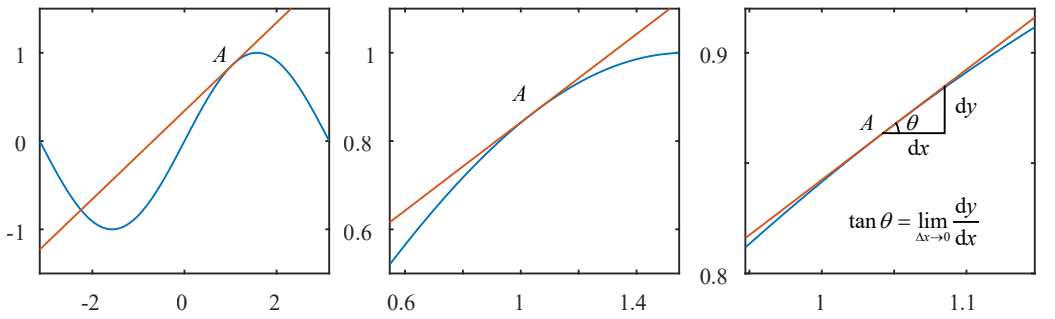


图 3: 将切点放大，会发现切线和曲线在切点附近“重合”

由上面的讨论可得，当 x 增加一小段 Δx 时，y 轴的增量约为 $\Delta y \approx f'(x)\Delta x$，且当 Δx 越小，这条式子就越精确成立，记为 $\mathrm{d}y = f'(x)\,\mathrm{d}x$．这个关系就叫函数的微分．

导数的代数理解

导数的代数理解就是：一个量关于另一个量的变化率．例如质点直线运动时，速度的大小就是其路程对时间的导数．把这种描述用极限[149] 表达出来就是

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \tag{2}$$

在图 3 的右图中，Δx 的始末位置并不非常重要，既可以从 x 取到 $x+\Delta x$，也可以从 $x-\Delta x$ 取到 x 等等（因为当 Δx 非常小的时候，x 附近的曲线基本处处跟切线重合，它们的斜率都是一样的）．所以导数的定义也有其他类似的形式

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x) - f(x-\Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x-\Delta x)}{2\Delta x} \tag{3}$$

虽然上面用到了诸如“近似”等词，但根据定义，极限都是精确的．

例子：速度加速度（一维）[879]．

求导法则

预备知识 基本初等函数的导数[165]

如果需要求导的函数可以看做若干个已知导函数的函数（如基本初等函数）经过四则运算或复合得到的，那么我们可以直接使用一系列求导法则对其求导

四则运算

$$[f(x)\pm g(x)]'=f'(x)\pm g'(x) \tag{1}$$

$$[f(x)g(x)]'=f'(x)g(x)+f(x)g'(x) \tag{2}$$

$$\left[\frac{f(x)}{g(x)}\right]'=\frac{f'(x)g(x)-g'(x)f(x)}{g(x)^2} \tag{3}$$

复合函数

$$f[g(x)]'=f'[g(x)]g'(x) \tag{4}$$

详见“一元复合函数求导（链式法则）[174]”

反函数

$$[f^{-1}(x)]'=\frac{1}{f'[f^{-1}(x)]} \tag{5}$$

详见“反函数求导[163]”.

线性

我们先来证明式 1．对求导而言，**线性**是指若干**函数线性组合**（即把若干个函数分别乘以常数再相加）[3]的求导等于对这些函数先分别求导再进行同样的线性组合．由于函数加减法属于函数线性组合的两种简单情况，这里只需要

[3]线性组合是线性代数中的概念，线性代数中，线性空间都可以做线性组合，例如几何矢量（式 6 [375]）．特定函数的集合可以看作线性空间（例 2 [444]）．

证明求导是一种**线性运算**即可．令若干常数为 c_i，若干可导函数为 $f_i(x)$，根据导数的定义，这些函数线性组合的导数为

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\sum_i c_i f_i(x) &= \lim_{h\to 0}\left[\sum_i c_i f_i(x+h) - \sum_i c_i f_i(x)\right]/h \\ &= \sum_i c_i \lim_{h\to 0}[f_i(x+h) - f_i(x)]/h \\ &= \sum_i c_i f_i'(x)\end{aligned} \tag{6}$$

证毕．

例 1　对函数 $f(x) = 5\sin x + 3x^2$ 求导

这里的 $f(x)$ 可以看做三角函数 $\sin x$ 函数和幂函数 x^2 的线性组合，二者都是基本初等函数，导数分别为 $\cos x$ 和 $2x$，由于求导是线性运算，我们只需要对两个函数各自的导函数进行同样的线性组合即可

$$f'(x) = 5\sin' x + 3(x^2)' = 5\cos x + 3(2x) = 5\cos x + 6x \tag{7}$$

乘积法则

现在证明式 2，令两函数分别为 $f(x)$ 和 $g(x)$，现在求 $f(x)g(x)$ 的导函数．由导数的定义式 2 [158] 得

$$[f(x)g(x)]' = \lim_{h\to 0}\frac{f(x+h)g(x+h) - f(x)g(x)}{h} \tag{8}$$

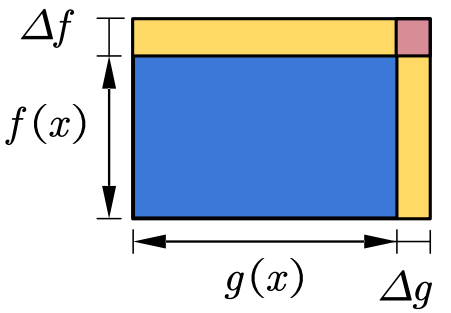


图 1: 乘积法则的几何理解

从几何上来看（图 1 ），我们可以把 $f(x), g(x)$ 看做一个矩形的两条边长，它们的乘积 $f(x)g(x)$ 就是矩形的面积，而 $f(x+h)g(x+h)$ 不妨看做是另一个

矩形的面积．当 x 增加 h 后，令 f 和 g 分别增加 Δf 和 Δg（图中假设他们大于零，其他情况同理），那么矩形面积的增量可以分解为三部分

$$f(x+h)g(x+h)-f(x)g(x)=\Delta fg+f\Delta g+\Delta f\Delta g \tag{9}$$

代入式 8 得

$$[f(x)g(x)]'=\lim_{h\to0}\frac{\Delta f}{h}g+f\lim_{h\to0}\frac{\Delta g}{h}+\lim_{h\to0}\frac{\Delta f\Delta g}{h} \tag{10}$$

容易看出前两项中的两个极限分别是 f,g 的导数，而第三个极限中，两个增量的乘积变小的速度远比前两个极限中要快（我们把它叫做二阶无穷小），所以第三个极限为零．从几何上来看，这是因为随着 h 变小，右上角小矩形的面积 $\Delta f\Delta g$ 比起两条长矩形的面积可以忽略不计．

商法则

要证明式 3，我们可以把 $f(x)/g(x)$ 看成 $f(x)$ 乘以 $1/g(x)$，这样就可以直接使用乘积法则了．但是如何对 $1/g(x)$ 求导呢？我们可以将其看做 $h(x)=1/x$ 和 $g(x)$ 的复合函数 $h[g(x)]$．其中 $h'(x)=-1/x^2$（式 1 [165]）．根据式 4，有

$$\left[\frac{1}{g(x)}\right]'=h'[g(x)]g'(x)=-\frac{g'(x)}{g^2(x)} \tag{11}$$

最后使用乘积法则

$$\begin{aligned}\left[f(x)\frac{1}{g(x)}\right]'&=f'(x)\frac{1}{g(x)}-f(x)\frac{g'(x)}{g^2(x)}\\&=\left[\frac{f(x)}{g(x)}\right]'=\frac{f'(x)g(x)-g'(x)f(x)}{g(x)^2}\end{aligned} \tag{12}$$

证毕．

反函数求导

预备知识 基本初等函数的导数[165]

若已知 $f(x)$ 的导函数为 $f'(x)$，则 $f(x)$ 的反函数 $f^{-1}(x)$ 的导函数为

$$[f^{-1}(x)]' = \frac{1}{f'[f^{-1}(x)]} \tag{1}$$

为了消除上式可能产生的歧义，记 $f(x)$ 的导函数为 $h(x)$，$f(x)$ 的反函数为 $g(x)$．上式变为

$$g'(x) = \frac{1}{h[g(x)]} \tag{2}$$

例 1

在区间 $[0,\infty)$ 上，令函数 $f(x) = x^2$，那么反函数为 $f^{-1}(x) = \sqrt{x}$．已知 $f'(x) = 2x$ （式 7 [165]），带入式 1 得反函数的导数为

$$[f^{-1}(x)]' = \frac{1}{2\sqrt{x}} \tag{3}$$

我们也可以通过直接对 $\sqrt{x} = x^{1/2}$ 求导来验证这一结果的正确性．

反函可导的条件

只有函数 $y = f(x)$，在某个区间 (x_1, x_2) 内连续且单调时．因为如果一个 y 有多个 x 对应，反函数中将会出现一个 x 对应多个 y 的情况．

反函数的定义

令满足上述条件的某函数和反函数分别为 $f(x)$，$g(x)$，在有定义的区间内的任何一对满足 $y = f(x)$ 的 x，y 都满足 $g(y) = x$，则 $g(x)$ 是 $f(x)$ 的反函数．

证明

根据导数和微分的关系，$y = f(x)$ 在曲线上的某点 (x_0, y_0)，有

$$\mathrm{d}y = f'(x_0)\,\mathrm{d}x \tag{4}$$

同一点也满足 $g(y_0) = x_0$，且

$$g'(y_0)\,\mathrm{d}y = \mathrm{d}x \tag{5}$$

对比式 4 和式 5，得

$$g'(y_0) = \frac{\mathrm{d}x}{\mathrm{d}y} = \frac{1}{f'(x_0)} \tag{6}$$

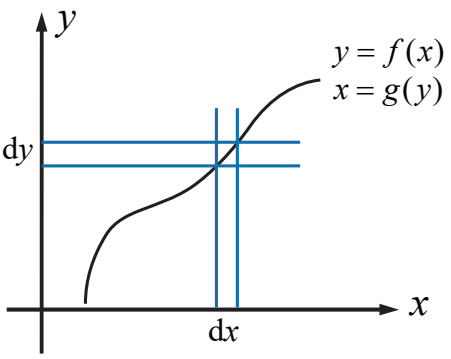


图 1: 在同一点处，$f' = \mathrm{d}y/\mathrm{d}x$，$g' = \mathrm{d}x/\mathrm{d}y$，互为倒数

可见图 1 曲线上同一点处 f' 和 g' 互为倒数．把 $x_0 = g(y_0)$ 代入上式，得

$$g'(y_0) = \frac{\mathrm{d}x}{\mathrm{d}y} = \frac{1}{f'[g(y_0)]} \tag{7}$$

上式中，y_0 可以是 g 函数定义域的任意一点，所以

$$g'(y) = \frac{1}{f'[g(y)]} \tag{8}$$

或者用习惯上的 x 作为自变量，得

$$g'(x) = \frac{1}{f'[g(x)]} \tag{9}$$

证毕.

基本初等函数的导数

预备知识 导数[158]

基本初等函数

基本初等函数由以下五类函数构成（a 是常数）

- **幂函数**

$$x^a \quad (a \in R) \tag{1}$$

- **指数函数**

$$a^x \quad (a>0,\ a\neq 1) \tag{2}$$

- **对数函数**

$$\log_a x \quad (a>0,\ a\neq 1) \tag{3}$$

当底为 $a=\mathrm{e}$ 时，叫做**自然对数函数**，记为 $\ln x$.

- **三角函数**

$$\sin x \qquad \cos x \qquad \tan x \tag{4}$$

- **反三角函数**

$$\arcsin x \qquad \arccos x \qquad \arctan x \tag{5}$$

由以上函数经过有限次四则运算和有限次函数复合所构成并可用一个式子表示的函数，称为**初等函数**. 例如

$$y=\sqrt{1-x^2} \qquad y=\sin^2 x \qquad y=\sqrt{\cot\frac{x}{2}} \tag{6}$$

基本初等函数的导数

基本初等函数在其定义域内都是可导的，导函数如下

- **幂函数**

$$(x^a)'=ax^{a-1} \quad (a\in R) \tag{7}$$

- **三角函数**

$$\sin' x=\cos x \qquad \cos' x=-\sin x \qquad \tan' x=1/\cos^2 x=\sec^2 x \tag{8}$$

- **指数函数**

$$(a^x)'=\ln(a)a^x \tag{9}$$

特殊地，$(\mathrm{e}^x)'=\mathrm{e}^x$

- **对数函数**

$$(\log_a x)'=\frac{1}{\ln(a)x} \tag{10}$$

特殊地，$\ln' x=1/x$.

幂函数证明

由导数的代数定义，$f'(x)=\lim_{h\to0}[f(x+h)-f(x)]/h$，而

$$(x+h)^a-x^a=x^a[(1+h/x)^a-1] \tag{11}$$

由于 $h\to0$，$h/x\to0$．令 $\varepsilon=h/x$，由非整数二项式定理[107]，

$$(1+\varepsilon)^a=1+a\varepsilon+\frac{a(a-1)}{2!}\varepsilon^2+\frac{a(a-1)(a-2)}{3!}\varepsilon^3\dots \tag{12}$$

所以

$$\begin{aligned}(x^a)'&=x^a\lim_{\varepsilon\to0}\frac{(1+\varepsilon)^a-1}{\varepsilon x}\\&=x^{a-1}\lim_{\varepsilon\to0}\left(a+\frac{a(a-1)}{2!}\varepsilon+\frac{a(a-1)(a-2)}{3!}\varepsilon^2\dots\right)=ax^{a-1}\end{aligned} \tag{13}$$

正弦函数证明

使用三角函数和差化积公式化简极限

$$\sin' x=\lim_{h\to0}\frac{\sin(x+h)-\sin x}{h}=\lim_{h\to0}\frac{\sin(h/2)}{h/2}\cos\left(x+\frac{h}{2}\right) \tag{14}$$

由小角正弦值极限[152] 中的结论，其中

$$\lim_{h\to0}\frac{\sin(h/2)}{h/2}=1 \tag{15}$$

所以

$$\sin' x=\lim_{h\to0}\cos\left(x+\frac{h}{2}\right)=\cos x \tag{16}$$

余弦函数证明

若 $f'(x)=g(x)$，且 b 为任意常数，根据导数的定义 $f'(x+b)=g(x+b)$ 同样成立（证明略）．所以 $\sin'(x+\pi/2)=\cos(x+\pi/2)$．而 $\sin(x+\pi/2)=\cos x$，$\cos(x+\pi/2)=-\sin x$ 所以 $\cos' x=-\sin x$

正切函数证明

根据求导法则[160]，因为 $\tan x = \sin x / \cos x$，所以

$$\tan' x = \frac{\sin' x \cos x - \cos' x \sin x}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x \tag{17}$$

对数函数证明

先证明 $\ln' x = 1/x$．$\ln(x+h) - \ln x = \ln(1 + h/x)$，所以

$$\ln' x = \lim_{h\to 0} \frac{\ln(x+h) - \ln x}{h} = \frac{1}{x} \lim_{h\to 0} \frac{\ln(1+h/x)}{h/x} \tag{18}$$

令 $\varepsilon = h/x$，则

$$\ln' x = \frac{1}{x} \lim_{\varepsilon\to 0} \frac{\ln(1+\varepsilon)}{\varepsilon} = \frac{1}{x} \lim_{\varepsilon\to 0} \ln(1+\varepsilon)^{\frac{1}{\varepsilon}} \tag{19}$$

由自然对数底的定义，$\lim\limits_{\varepsilon\to 0}(1+\varepsilon)^{\frac{1}{\varepsilon}} = \mathrm{e}$，所以

$$\ln' x = \frac{\ln \mathrm{e}}{x} = \frac{1}{x} \tag{20}$$

再证明 $\log_a' x = 1/(x \ln a)$．由对数函数的性质 $\log_a b = \ln b / \ln a$

$$\log_a' x = \left(\frac{\ln x}{\ln a}\right)' = \frac{1}{\ln a} \ln' x = \frac{1}{x \ln a} \tag{21}$$

指数函数证明

先证明 $(\mathrm{e}^x)' = \mathrm{e}^x$．由于上面已经证明了 $\ln' x = 1/x$，而 e^x 是 $\ln x$ 的反函数．所以令 $f(x) = \ln x$，$f'(x) = 1/x$，$f^{-1}(x) = \mathrm{e}^x$，代入反函数的求导法则[163]

$$[f^{-1}(x)]' = \frac{1}{f'[f^{-1}(x)]} \tag{22}$$

得

$$(\mathrm{e}^x)' = \frac{1}{1/\mathrm{e}^x} = \mathrm{e}^x \tag{23}$$

再证明 $(a^x)' = a^x \ln a$．$(a^x)' = \left[\left(\mathrm{e}^{\ln a}\right)^x\right]' = \left(\mathrm{e}^{(\ln a)x}\right)'$．把 $\mathrm{e}^{(\ln a)x}$ 看成是 e^u 和 $u = (\ln a)x$ 的复合函数，根据复合函数的求导法则[160]，$(a^x)' = (\ln a)a^x$

高阶导数

- 未完成：说明越高阶可导越光滑，无穷阶可导简称为“光滑”，或“解析”

预备知识 求导法则[160]

一个一元函数 $y=f(x)$ 可导，如果导函数 $y'=f'(x)$ 仍可导，则称 $y'=f'(x)$ 的导数为函数 $y=f(x)$ 的二阶导数，记为

$$y'',\quad f''(x),\quad \frac{\mathrm{d}^2y}{\mathrm{d}x^2},\quad \frac{\mathrm{d}^2f}{\mathrm{d}x^2},\quad y''(x) \tag{1}$$

即

$$y''=(y')',\quad \frac{\mathrm{d}^2y}{\mathrm{d}x^2}=\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}x}{\mathrm{d}y}\right) \tag{2}$$

由导数中的定义式 2 [158]，可得二阶导数的计算公式

$$f''(x)=\lim_{\Delta x\to 0}\frac{f'(x+\Delta x)-f'(x)}{\Delta x} \tag{3}$$

若把 $y=f(x)$ 的导数 $y'=f'(x)$ 称为 $y=f(x)$ 的**一阶导数**，那么，一阶导数的导数就称为**二阶导数**. 若二阶导数 $y''=f''(x)$ 仍然可导，我们就把二阶导数的数 $y''=f''(x)$ 的导数称为**三阶导数**，记为

$$y''',\quad f'''(x),\quad \frac{\mathrm{d}^3y}{\mathrm{d}x^3},\quad \frac{\mathrm{d}^3f}{\mathrm{d}x^3},\quad y'''(x) \tag{4}$$

一般地，如果 $y=f(x)$ 的 $n-1$ 阶导数是可导的，我们就把 $n-1$ 阶导数的导数称为果 $y=f(x)$ 的 **n 阶导数**，记为

$$y^{(n)},\quad f^{(n)}(x),\quad \frac{\mathrm{d}^ny}{\mathrm{d}x^n},\quad \frac{\mathrm{d}^nf}{\mathrm{d}x^n},\quad y^{(n)}(x) \tag{5}$$

二阶和二阶以上的导数统称为**高阶导数**. 利用求导公式和求导法则就可以求出高阶导数.

莱布尼兹公式

预备知识 高阶导数[169]

设函数 $u=u(x), v=v(x)$ 均有 n 阶导数，形如 $y=uv$ 的函数的 n 阶导数，可由**莱布尼兹（Leibniz）公式**求出.

不加推导的给出莱布尼兹公式，可用数学归纳法证明.

$$\begin{aligned}(uv)^{(n)} =& u^{(n)}v + nu^{(n-1)}v' + \frac{n(n-1)}{2!}u^{(n-2)}v'' \\ &+ \cdots + \frac{n(n-1)\cdots(n-k+1)}{k!}u^{(n-k)}v^{(k)} + \cdots + uv^{(n)}\end{aligned} \tag{1}$$

规定一个函数的零阶导数等于函数本身，即 $u^{(0)}=u$，于是莱布尼兹公式可以写成如下的形式

$$(uv)^{(n)} = \sum_{k=0}^{n} C_n^k u^{(u-k)}v^{(k)} \tag{2}$$

例 1　求函数 $y=xe^{-x}$ 的 n 阶导数

可以通过莱布尼兹方程对该函数直接求 n 阶导数.

$$\begin{aligned}(x\mathrm{e}^{-x})^{(n)} &= \sum_{k=0}^{n} C_n^k (\mathrm{e}^{-x})^{(n-k)} \\ &= x(\mathrm{e}^{-x})^{(n)} + n(\mathrm{e}^{(-x)})^{(n-1)} \\ &= x(-1)^n\mathrm{e}^{-x} + n(-1)^{n-1}\mathrm{e}^{-x} \\ &= (-1)^{n-1}(n-x)\mathrm{e}^{-x}\end{aligned} \tag{3}$$

导数与函数极值

预备知识 二阶导数[169]

如图 1，若一个一元函数 $y=f(x)$ 在某区间内处处可导（即对区间内的任何 x 导数 $f'(x)$ 都存在），若区间内存在某些 x_i 能使 $f'(x_i)=0$（即在这些点处函数曲线的斜率为零），这样的点被称为**驻点**.

而从函数曲线来看，驻点又分为三类：**极大值**，**极小值**，**鞍点**. 我们以 x_i 为中心取一个小区间，如果这个区间足够小，那么容易看出对于极大值点，$f'(x)$ 在小区间内递减，对于鞍点，$f'(x)$ 在小区间内恒为非负或恒为非正，对于极小

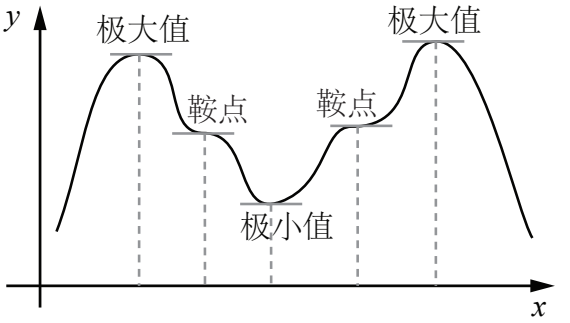


图 1: 导数为零的三种点

值点，$f'(x)$ 在小区间内递增．所以为了判断驻点的类型，我们可以在驻点处求函数的二阶导数 $f''(x_i)$．假设二阶偏导存在，如果 $f''(x_i) < 0$，那么 x_i 是极大值点，如果 $f''(x_i) > 0$，x_i 是极小值点．要注意的是，如果 $f''(x_i) = 0$，不能直接判断 x_i 鞍点，需要进一步分析：例如我们可以判断驻点左边和右边的一阶导数符号，如果同号则是驻点，左正右负则是极大值，左负右正则是极小值．

另外，若某个极小值点是整个考察区间中函数值最小的点，它就被称为**最小值点**，若某个极大值点是该区间中函数值最大的点，它就被称为**最大值点**．

例 1

二次函数 $f(x) = ax^2 + bx + c$ 的导函数为 $f'(x) = 2ax + b$，所以唯一的驻点为 $-b/(2a)$．函数的二阶导数是一个常数 $f''(x) = 2a$，所以当 $a > 0$ 时驻点是唯一的极小值点，即最小值点．同理，当 $a < 0$ 时驻点是最大值点．

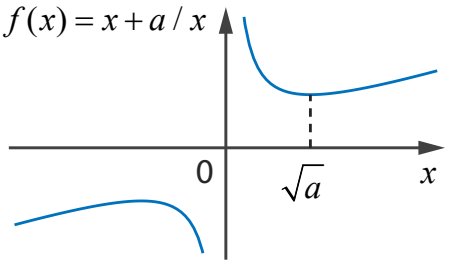


图 2: 例 2 函数图

例 2

函数 $f(x) = x + a/x$ $(a > 0)$ 的一阶导函数为 $f'(x) = 1 - a/x^2$，若我们只考察区间 $(0, +\infty)$，唯一的驻点为 $x = \sqrt{a}$．函数的二阶导函数 $f''(x) = 2a/x^3$ 在驻点处的值为 $2/\sqrt{a} > 0$，所以该驻点为当前区间的最小值点（图 2）．

例 3

函数 $f(x)=x^3$ 的一阶导函数为 $f'(x)=3x^2$，唯一的驻点为 $x=0$．函数的二阶导函数 $f''(x)=6x$ 在驻点处的值为 0．由于 $f'(x)$ 在原点左侧和右侧都大于 0，所以这是一个鞍点．

用极值点确定函数图像

预备知识 导数与函数的极值[170]

许多时候，我们可以通过函数极值点的位置以及种类确定函数的大致图像．

例 1

求以下函数的极值点，并判断该函数的大致图像

$$f(x)=a+\frac{b}{2}x^2-\frac{c}{4}x^4+\frac{d}{6}x^6 \quad (a,b,c,d>0) \tag{1}$$

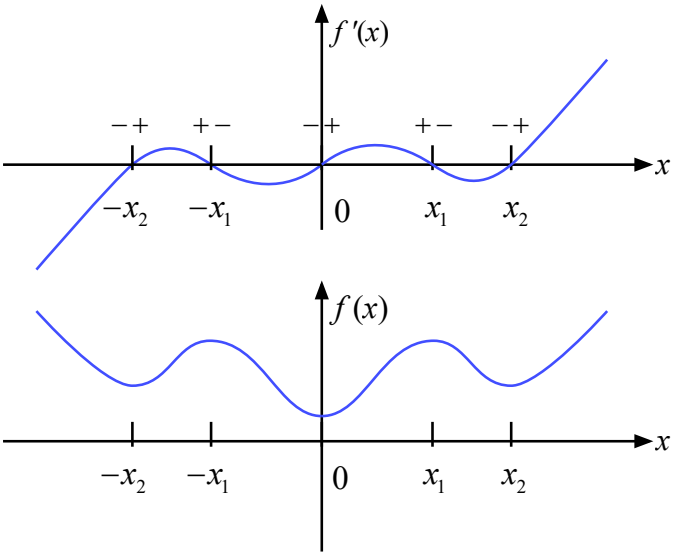


图 1: 导数 $f'(x)$ 与原函数 $f(x)$

先对函数求导，得

$$f'(x)=bx-cx^3+dx^5 \tag{2}$$

这是一个 5 次多项式，最多可能有五个零点．但由于多项式只有奇数项，不难解出可能的根．

$$bx + cx^3 + dx^5 = x[b + c(x^2) + d(x^2)^2] \tag{3}$$

所以其中一个解是 $x = 0$，而 $b + c(x^2) + d(x^2)^2$ 是关于 x^2 的二次方程，当判别式 $\Delta = c^2 - 4bd$ 大于零时，x^2 存在两个大小不等的正根，姑且记为 x_1^2 和 x_1^2，$x_1 < x_2$．此时五个根分别为 $0, \pm x_1, \pm x_2$．

$$f'(x) = d \cdot x(x^2 - x_2^2)(x^2 - x_2^2) = d \cdot x(x + x_1)(x - x_1)(x + x_1)(x - x_2) \tag{4}$$

由于 $d > 0$，可以大致画出 $f'(x)$ 图像如图 1 （下）．

用二阶导数判断分类．若二阶导数为正，则是极小值，若为负，则是极大值，若为零，则是鞍点．

一元函数的微分

预备知识 导数[158]

考察一个连续光滑的函数 $y = f(x)$，在 x 处函数值为 y，若此时函数增加一个无穷小量 $\mathrm{d}x$，函数值会相应增加无穷小量 $\mathrm{d}y$．根据导数的定义[158] $f'(x) = \mathrm{d}y/\mathrm{d}x$，我们将 $\mathrm{d}y$ 与 $\mathrm{d}x$ 的关系记为

$$\mathrm{d}y = f'(x)\,\mathrm{d}x \tag{1}$$

这就是一元函数的**微分**．注意一元函数的求导和微分除了表达方式不同外并无太大区别．从形式上来看，微分是微小变化量之间的线性关系，而导数则强调变化率．

微分近似

严格来说，类似式 1 的微分关系式默认取极限 $\mathrm{d}x \to 0$ 才能使等号成立，但只要在一定范围 Δx 内导函数 $f'(x)$ 的变化非常小，就可以将函数值的变化量 $\Delta y = f(x + \Delta x) - f(x)$ 近似为

$$\Delta y \approx f'(x)\Delta x \tag{2}$$

注意在近似式中不能出现微分符号 d，也不能使用等号．

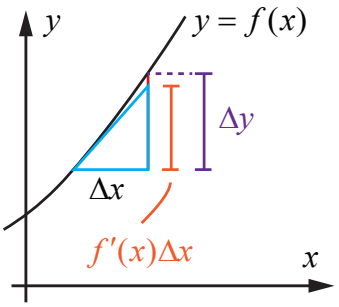


图 1: 微分近似用函数曲线的切线增量 $f'(x)\Delta x$ 来近似函数增量 Δy

例 1　测量误差

若测得立方体的边长为 a，测量的最大可能误差为 σ_a（可以假设 $\sigma_a \ll a$），估计立方体体积的最大误差 σ_V.

立方体的体积与边长的关系为 $V(a) = a^3$，根据微分近似，有

$$\sigma_V \approx V'(a)\sigma_a = 3a^2\sigma_x \tag{3}$$

例 2　细圆环的面积和薄球壳的体积

圆的面积关于其半径的函数为 $A(r) = \pi r^2$，对该式进行微分得 $\mathrm{d}A = 2\pi r\,\mathrm{d}r$. 注意到 $2\pi r$ 为 r 对应的周长，所以该微分式的意义就是，半径为 r，宽度为 $\mathrm{d}r$ 的圆环的面积等于该圆环的周长乘以圆环的宽度.

球的体积关于其半径的函数为 $V(r) = 4\pi r^3/3$，求微分得 $\mathrm{d}V = 4\pi r^2\,\mathrm{d}r$. 注意到 $4\pi r^2$ 为 r 对应的球表面积，所以该微分式的意义是，半径为 r，厚度为 $\mathrm{d}r$ 的球壳的体积等于该球壳的表面积乘以球壳厚度.

复合函数求导 链式法则

预备知识　微分[173]

若有两个一元函数 $f(x)$ 和 $g(x)$，我们可以把 g 的函数值作为 f 的自变量，得到一个新的函数称为 $f(x)$ 和 $g(x)$ 的**复合函数**，记为 $f[g(x)]$.

如果我们已知两个函数 $f(x)$ 和 $g(x)$ 的导函数 $f'(x)$ 和 $g'(x)$，那么我们可以通过以下公式求复合函数 $f[g(x)]$ 的导数.

$$f[g(x)]' = f'[g(x)]g'(x) \tag{1}$$

对于多个函数的复合函数，我们也有类似的公式，例如

$$f[g(h(x))]' = f'[g(h(x))]g'[h(x)]h'(x) \tag{2}$$

例 1　基本初等函数的复合函数求导

我们已经知道基本初等函数的导数[165] 的导函数，下面对它们的一些常见的复合函数进行求导.

$\sin^2 x$ 可以看作幂函数 $f(x) = x^2$ 和 $g(x) = \sin x$ 的复合函数，已知 $f'(x) = 2x$，$g'(x) = \cos x$，代入式 1 得

$$(\sin^2 x)' = 2\sin x \cos x \tag{3}$$

几何理解

为了方便表示，我们把 g 的函数值和 f 的自变量记为 u，把 f 的函数值记为 y.

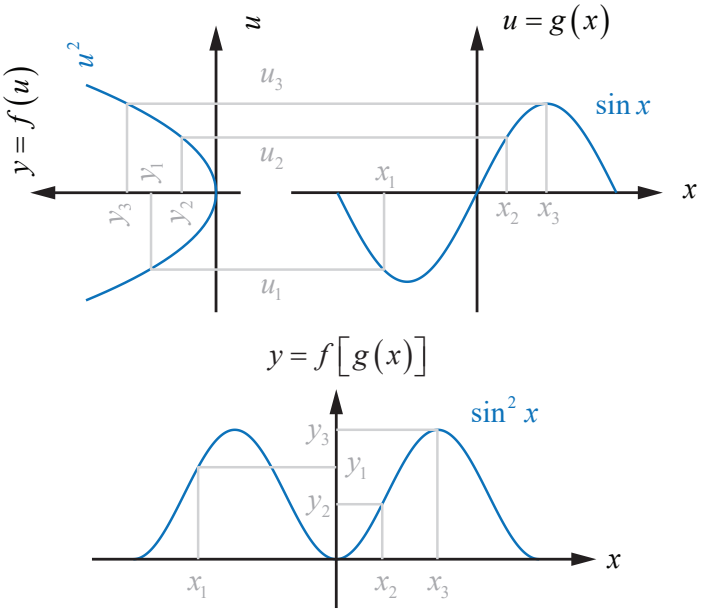


图 1: 可以将 $\sin^2 x$ 看做 $f(u) = u^2$ 和 $g(x) = \sin x$ 的复合函数

我们可以用类似图 1 的图像来直观地理解复合函数. 先画出 $y = f(u)$ 和 $u = g(x)$ 的图像，并将 $g(u)$ 的图像逆时针旋转 90° 使得两图的 u 轴对齐. 这

样对于任何定义域中的自变量 x，我们只需要先在 $g(x)$ 的图中画出 u 的位置，再对应到 $f(u)$ 的图像中求出 y 的位置即可.

现在我们要讨论的问题是，若已知两函数的导函数 $f'(x)$ 和 $g'(u)$（假设它们在定义域内处处可导）如何求复合函数 $f[g(x)]$ 的导数.

对于给定的 x，我们先来看当 x 增加 Δx 时 y 的增量 Δy 的大小. 我们可以使用与图 1 类似的方法画出图 2，然后只需要令 $\Delta x \to 0$，就可以根据定义求出复合函数的导数

$$f[g(x)]' = \frac{\mathrm{d}}{\mathrm{d}x} f[g(x)] = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \tag{4}$$

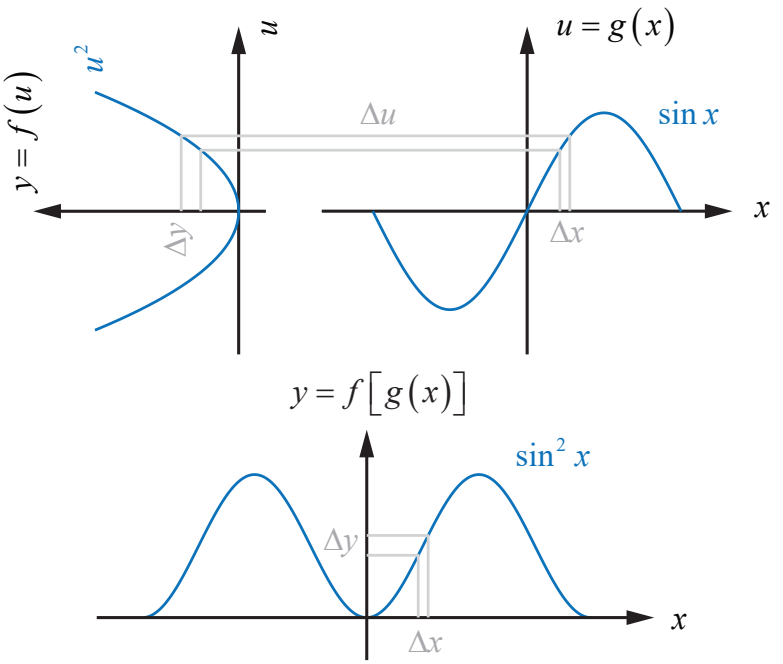


图 2: 用图 1 中的方法求出任意 Δx 对应的 Δy

在这个过程中，我们在得到 Δy 之前先得到了 u 的增量 Δu. 当 Δx 较小时有微分近似（式 2 [173]）

$$\Delta u \approx g'(x)\Delta x \qquad \Delta y \approx f'(u)\Delta u \tag{5}$$

当 $\Delta x \to 0$ 时对应的微分关系（式 1 [173]）为

$$\mathrm{d}u = g'(x)\,\mathrm{d}x \qquad \mathrm{d}y = f'(u)\,\mathrm{d}u \tag{6}$$

将上式中的左边代入右边得

$$\mathrm{d}y = f'(u)g'(x)\,\mathrm{d}x = f'[g(x)]g'(x)\,\mathrm{d}x \tag{7}$$

而复合函数的微分是

$$\mathrm{d}y = f[g(x)]'\,\mathrm{d}x \tag{8}$$

对比以上两式（微分和导数的关系）得

$$f[g(x)]' = f'[g(x)]g'(x) \tag{9}$$

这就是复合函数的求导公式.

在上面的例子中

$$g(x) = \sin x \qquad g'(x) = \cos x \qquad f(u) = u^2 \qquad f'(u) = 2u \tag{10}$$

代入上式得

$$\frac{\mathrm{d}}{\mathrm{d}x}\sin^2 x = 2\sin x\cos x \tag{11}$$

复合函数的求导公式也叫**链式法则**，原因是我们可以把以上推导过程用导数的另外一种符号表示如下.

$$\mathrm{d}y = \frac{\mathrm{d}y}{\mathrm{d}u}\,\mathrm{d}u = \frac{\mathrm{d}y}{\mathrm{d}u}\frac{\mathrm{d}u}{\mathrm{d}x}\,\mathrm{d}x \tag{12}$$

得

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}u}\frac{\mathrm{d}u}{\mathrm{d}x} \tag{13}$$

这种书写方式让人不禁想把 $\mathrm{d}y/\mathrm{d}x$ 看做是 $\mathrm{d}y$ 和 $\mathrm{d}x$ 相除，这样的符号分割是错误的，尤其是在以后学习高阶导数和偏导数时.

多重复合函数

要对多重复合函数如 $f[g(h(x))]$ 求导，可以先对 $g[h(x)]$ 求导得 $g'[h(x)]h'(x)$ 再得到

$$f[g(h(x))]' = f'[g(h(x))]g'[h(x)]h'(x) \tag{14}$$

令 $v = h(x)$，用微分符号可以表示为

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}u}\frac{\mathrm{d}u}{\mathrm{d}v}\frac{\mathrm{d}v}{\mathrm{d}x} \tag{15}$$

任意多重的复合函数求导同理可得.

例 2　对函数求导

$$\frac{1}{\sqrt{x^2+a^2}} \tag{16}$$

首先令 $f(x)=1/\sqrt{x}$ 再令 $g(x)=x^2+a^2$，上式等于 $f[g(x)]$. 由基本初等函数的导数[165]，

$$f'(x)=-\frac{1}{2\sqrt{x^3}} \qquad g'(x)=2x \tag{17}$$

代入式 9，得

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{\sqrt{x^2+a^2}}=f'[g(x)]g'(x)=-\frac{x}{\sqrt{(x^2+a^2)^3}} \tag{18}$$

一种较灵活的情况是，当三个变量只有一个自由度[4]时，任何一个变量都可以看做任何另外两个变量的函数[5]，这时可以根据需要灵活运用链式法则，如例 3.

例 3　加速运动公式

假设质点做一维运动，位移，速度和加速度分别记为 $x(t)$，$v(t)=\mathrm{d}x/\mathrm{d}t$，$a(t)=\mathrm{d}v/\mathrm{d}t$，但若把速度 v 看做复合函数 $v[x(t)]$，根据链式法则有

$$a=\frac{\mathrm{d}v}{\mathrm{d}t}=\frac{\mathrm{d}v}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t}=\frac{\mathrm{d}v}{\mathrm{d}x}v \tag{19}$$

写成微分表达式，有 $a\,\mathrm{d}x=v\,\mathrm{d}v$. 注意到 $\mathrm{d}(v^2)=2v\,\mathrm{d}v$，代入得

$$\mathrm{d}\left(v^2\right)=2a\,\mathrm{d}x \tag{20}$$

若质点做匀加速运动，该式的物理意义是在任何一段微小时间内，速度平方的增量正比于这段时间内的位移增量. 在一段时间 $[t_1,t_2]$ 内把这些增量累加起来，就得到高中熟悉的运动学公式

$$v_2^2-v_1^2=2a(x_2-x_1) \tag{21}$$

其中 x_1,v_1 和 x_1,v_1 分别是 t_1,t_2 时刻的位置和速度.

[4]即任何一个变量值确定后，另外两个变量也随之确定

[5]姑且假设不会出现一个自变量对应两个函数值的情况

曲率 曲率半径（平面）

预备知识 切线[155]，极坐标系[119]

[6]我们来看一个平面上的一个光滑曲线（即处处存在切线），我们如何描述它某点处的弯曲程度呢？一种常用方法是在这点附近取曲线的一小段，然后做一个尽量与它吻合的圆，当这小段的长度趋近于 0 时，这个圆可以唯一确定．我们把这个圆叫做**密切圆（osculating circle）**，把密切圆的半径叫做曲线在该点的**曲率半径（radius of curvature）**，曲率半径的倒数 $1/R$ 叫做**曲率（curvature）**．

我们先来看一个半径为 R 的圆的一小段圆弧，令其长度为 Δl．作这段圆弧两端的切线，令它们的夹角为 $\Delta\theta$，那么显然满足 $R\Delta\theta = \Delta l$．同理，对于任意光滑曲线上长度为 Δl 的一段，我们也可以做相同的方法定义曲率半径 R，但需要令 $\Delta l \to 0$

$$R = \lim_{\Delta l \to 0} \frac{\Delta l}{\Delta \theta} \tag{1}$$

曲率的具体的计算公式取决于使用什么方式定义曲线，最常见描述方式就是在直角坐标系中通过函数 $y(x)$ 来定义，点 (x, y) 处的曲率半径为（$\dot{y}, \ddot{y}$ 分别表示导数和二阶导数）

$$R = \frac{(1+\dot{y}^2)^{3/2}}{\ddot{y}} \tag{2}$$

如果通过极坐标[119] 函数 $r(\theta)$ 定义曲线，则点 (r, θ) 处的曲率半径为

$$R = \frac{(r^2+\dot{r}^2)^{3/2}}{r^2+2\dot{r}^2-r\ddot{r}} \tag{3}$$

直角坐标系的推导

预备知识 高阶导数[169]，一元函数的微分[173]

平面上曲线的最常见描述方式就是通过定义函数 $y(x)$．我们可以通过导数[158] 计算曲线上某点切线关于 x 轴的夹角 θ．

$$\dot{y} = \frac{\mathrm{d}y}{\mathrm{d}x} = \tan\theta \tag{4}$$

[6]参考 Wikipedia 相关页面

曲线长度的微分为

$$\mathrm{d}l = \frac{\mathrm{d}x}{\cos\theta} \tag{5}$$

其中

$$\cos\theta = \frac{1}{\sqrt{1+\tan^2\theta}} = \frac{1}{\sqrt{1+\dot{y}^2}} \tag{6}$$

为了得到 $\mathrm{d}\theta$，我们对式 4 两边做微分得

$$\ddot{y}\,\mathrm{d}x = \frac{1}{\cos^2\theta}\,\mathrm{d}\theta \tag{7}$$

所以曲率半径为

$$R = \frac{\mathrm{d}l}{\mathrm{d}\theta} = \frac{1}{\ddot{y}\cos^3\theta} = \frac{(1+\dot{y}^2)^{3/2}}{\ddot{y}} \tag{8}$$

极坐标系的推导

极坐标系中，同样可以用函数 $r(\theta)$ 描述曲线．令

$$\frac{\mathrm{d}r}{r\,\mathrm{d}\theta} = \tan\alpha \tag{9}$$

即

$$\dot{r} = r\tan\alpha \tag{10}$$

长度微分为

$$\mathrm{d}l = \frac{r\,\mathrm{d}\theta}{\cos\alpha} \tag{11}$$

微分

$$\ddot{r}\,\mathrm{d}\theta = \dot{r}\tan\alpha\,\mathrm{d}\theta + \frac{r}{\cos^2\alpha}\,\mathrm{d}\alpha \tag{12}$$

即

$$\frac{\mathrm{d}\alpha}{\mathrm{d}\theta} = (\ddot{r} - \dot{r}\tan\alpha)\frac{\cos^2\alpha}{r} = \frac{\ddot{r} - \dot{r}\tan\alpha}{r(t+\tan^2\alpha)} \tag{13}$$

注意切线方向的微分是 $\mathrm{d}\theta - \mathrm{d}\alpha$．所以

$$R = \frac{\mathrm{d}l}{\mathrm{d}\theta - \mathrm{d}\alpha} = \frac{\mathrm{d}l\,/\,\mathrm{d}\theta}{1 - \mathrm{d}\alpha\,/\,\mathrm{d}\theta} \tag{14}$$

把式 1 和式 13 代入，再使用式 10 消去 α 得

$$R = \frac{(r^2+\dot{r}^2)^{3/2}}{r^2 + 2\dot{r}^2 - r\ddot{r}} \tag{15}$$

泰勒级数

预备知识 阶乘[105]，高阶导数[169]，分部积分法[190]

若函数 f 在开区间 I 内可以求任意阶的导数（例如多项式，幂函数，三角函数，指数函数，对数函数等），那么这个函数可以用多项式近似，且在某种意义下，总项数 N 越多，近似得越精确．确切地说，对于任何 $x_0 \in I$, 存在唯一一个数列 $\{c_n\}$, 使得对于任何正整数 N, 皆有（其中 $(x-x_0)^0$ 始终视为 1，即使 $x = x_0$）

$$f(x) - \sum_{n=0}^{N} c_n (x-x_0)^n = \mathcal{O}\big(|x-x_0|^{N+1}\big). \tag{1}$$

每一个系数 c_n 由函数在 x_0 处的第 n 阶导数求得

$$c_n = \frac{1}{n!} f^{(n)}(x_0); \tag{2}$$

注意其中 0 的阶乘为 $0! = 1$．另外由式 1 得，当 $x = x_0$ 时，函数值等于多项式值．当项数 N 有限时，通常 $|x-x_0|$ 越小多项式就越接近函数．以上这种把函数展开成多项式的方法就叫**泰勒展开（Taylor expansion）**，得到的多项式叫做**泰勒级数（Taylor series）**．我们先来看一个例子：

例 1　正弦函数

我们在 $x_0 = 0$ 处展开 $\sin x$，由式 1 和式 2 得

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \dots \tag{3}$$

取不同的项数 N 求和，画图如图 1．可见随着项数增加，多项式慢慢趋近正弦函数．

习题 1　一些常见函数关于原点的泰勒展开

关于原点的泰勒展开又叫麦克劳林展开，请验证以下的麦克劳林展开：

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 \dots \qquad (x \in \mathbb{R}) \tag{4}$$

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 \dots \qquad (x \in \mathbb{R}) \tag{5}$$

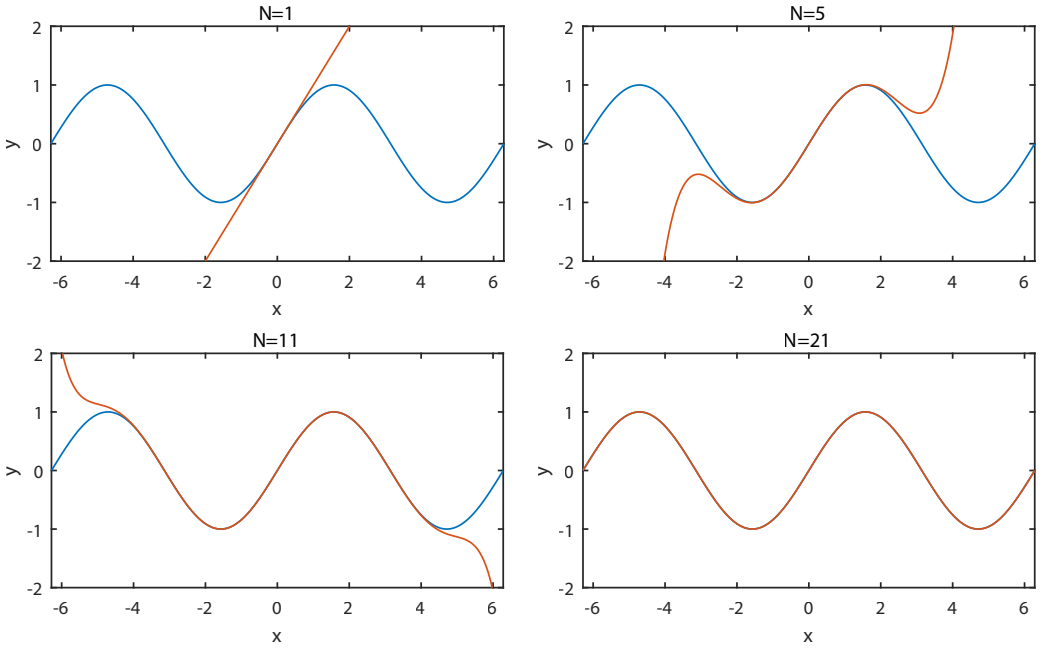


图 1: $\sin x$ 在原点处的泰勒展开的前 N 项求和．容易看出，求和的项数越多，多项式（橙）与 $\sin x$ （蓝）吻合得越好．

$$\mathrm{e}^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 \dots \qquad (x \in \mathbb{R}) \tag{6}$$

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 \dots \qquad (-1 < x < 1) \tag{7}$$

$$\frac{1}{1 \pm x} = 1 \mp x + x^2 \mp x^3 + x^4 \dots \qquad (-1 < x < 1) \tag{8}$$

$$\sqrt{1 \pm x} = 1 \pm \frac{1}{2}x - \frac{1}{8}x^2 \pm \frac{1}{16}x^3 - \frac{5}{128}x^4 \dots \qquad (-1 < x < 1) \tag{9}$$

收敛半径

我们把形如 $\sum_{n=0}^{\infty} c_n(x-x_0)^n$ 的表达式叫做**幂级数（power series）**[205]．泰勒展开就是在用幂级数表示函数．如果某幂级数收敛的点不止 x_0 一点，那么必定存在一个**收敛半径（radius of convergence）** $0 < r < 1$，使得当 $|x-x_0| < r$ 时级数必定收敛，而 $|x-x_0| > r$ 是必定发散（不收敛）．

所以在进行泰勒展开时，如果函数在除了若干发散点[7]外都无穷阶可导，令离 x_0 最近的发散点为 a，那么 $f(x)$ 关于 x_0 泰勒展开后，幂级数的收敛半径就

[7]如式 7 中的 $x = -1$，式 8 和式 9 中的 $x = \mp 1$

是 $r = |a - x_0|$.

例如在式 7 到式 9 中，函数的发散点距离原点都是 1，所以幂级数收敛半径为 1，所以即使在 x_0 的另一侧 $f(x)$ 处处无穷可导，仍然不能取 $|x| > 1$.

幼稚的推导

这里给出一个比较直观的对比系数推导方法，可能对初学者有一定启发或者帮助记忆．但以后会看到这是不严谨的.

我们假设当项数 $N \to \infty$ 时，存在唯一的多项式在某区间内处处趋于无穷可导函数 $f(x)$，即

$$f(x) = \sum_{n=0}^{\infty} c_n (x - x_0)^n \tag{10}$$

首先代入 $x = x_0$，可得第一个系数 $c_0 = f(x_0)$．现在我们对上式两边在 x_0 处求导，得

$$f'(x_0) = c_1 + \sum_{n=2}^{\infty} n c_n (x - x_0)^{n-1} \Bigg|_{x=x_0} = c_1 \tag{11}$$

如果对式 10 两边在 x_0 处求二阶导数，得

$$f''(x_0) = 2c_2 + \sum_{n=3}^{\infty} n(n-1) c_n (x - x_0)^{n-2} \Bigg|_{x=x_0} = 2c_2 \tag{12}$$

即 $c_2 = f''(x_0)/2!$．以此类推，如果对式 10 两边在 x_0 处求 m 阶导数得

$$f^{(m)}(x_0) = m! c_m + \sum_{n=m+1}^{\infty} \frac{n!}{(n-m)!} c_n (x - x_0)^{n-m} \Bigg|_{x=x_0} = m! c_m \tag{13}$$

所以系数公式为

$$c_m = \frac{1}{m!} f^{(m)}(x_0) \tag{14}$$

泰勒展开的存在说明了一些函数（下文的**解析函数**）具有这样的性质：任何一点的性质都能决定完整的函数曲线，这可以类比生物中用一个细胞克隆出一个完整生物体.

推导

如果假设函数 f 在开区间 I 上无穷次可微，那么可以利用分部积分公式[190]：

$$\begin{aligned} f(x) &= f(x_0) + \int_{x_0}^{x} f'(t)\,\mathrm{d}t \\ &= f(x_0) - \int_{x_0}^{x} f'(t)\frac{\mathrm{d}}{\mathrm{d}t}(x-t)\,\mathrm{d}t \\ &= f(x_0) + f'(x_0)(x-x_0) + \int_{x_0}^{x} f''(t)(x-t)\,\mathrm{d}t\,. \end{aligned} \tag{15}$$

由于 $(x-t) = -\frac{\mathrm{d}}{\mathrm{d}t}(x-t)^2/2!$, 故可以再次分部积分，得到

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \frac{1}{2!}\int_{x_0}^{x} f^{(3)}(t)(x-t)^2 dt. \tag{16}$$

如此续行，即得到

$$\begin{aligned} f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \ldots + \frac{f^{(N)}(x_0)}{N!}(x-x_0)^N \\ + \frac{1}{N!}\int_{x_0}^{x} f^{(N+1)}(t)(x-t)^N\,\mathrm{d}t\,. \end{aligned} \tag{17}$$

最后的误差可利用定积分估值估计为

$$\left|\frac{1}{N!}\int_{x_0}^{x} f^{(N+1)}(t)(x-t)^N\,\mathrm{d}t\right| \le \frac{\sup_{t\in I}|f^{(N+1)}(t)|}{N!}|x-x_0|^{N+1}. \tag{18}$$

显然，这就给出了唯一一个满足开头要求的多项式近似.

泰勒级数 2

预备知识 泰勒级数[181]

泰勒展开与近似

事实上，泰勒展开可以看成是微分近似[173] 的一种高阶拓展. 微分近似中，在某点 x_0 附近有

$$f(x) \approx f(x_0) + f'(x_0)(x-x_0) \tag{1}$$

而这恰好是泰勒展开的前两项.然而,这只是函数曲线在 x_0 处的切线(见图 1 [181] 中 $N=1$ 的情况),显然没有高阶的泰勒展开那么精确. 如果我们将 $f(x)$ 近似到其泰勒展开的 x^n 项，我们称这个近似精确到第 n 阶，因为它的误差小于或等于 $n+1$ 阶无穷小[149] $O(x^{n+1})$. 在近似计算中，可以使用式 18 [181] 来精确地估计近似多项式给出的误差. 例如，图 1 [181] 中画出的正弦函数与其近似多项式的误差即可按此公式计算.

泰勒展开的问题

尽管泰勒展开式给出了函数在某点处的多项式近似，但这里很可能会产生一些微妙的问题. 从余项公式式 18 [181] 并不能推出余项一定随着 N 的增大而趋于零，因为函数的 N 阶导数可能会随着 N 的增长而快速增长; 具体来说，有博雷尔（E. Borel）的一个定理：

定理 1　博雷尔定理

对于任何复数序列 $\{c_n\}$, 都存在一个光滑函数 f 使得其在 $x=0$ 处的泰勒展开式为 $\sum_{n=0}^{\infty} c_n x^n$.

因此，显然不收敛的级数 $\sum_{n=0}^{\infty} n!x^n$ 实际上也是某个光滑函数的泰勒展开! 这说明由泰勒展开截断得到的多项式仅仅是一个渐近展开[258], 仅仅在展开点那一处给出了最佳的多项式近似，却并不能说明任何其它性质. 也就是说，对于**固定的** N, 当 x 越来越接近 x_0 时，$f(x)$ 同 N 泰勒多项式之间的差是高于 N 阶的高阶无穷小，但对于固定的 x，却不能断言当 N 趋于无穷时 $f(x)$ 同其泰勒多项式的误差趋于零. 当然，在许多情况下，这对于近似计算已经足够用.

进一步，即便泰勒级数收敛，也不一定收敛到展开的函数本身，例如函数 $f(x)=e^{-1/x^2}$ 在点 $x=0$ 处的所有导数都是零，因此它在这点处的泰勒展开是 $0+0x+0x^2+\ldots$，它与函数 $f(x)$ 的误差永远是函数本身!

由此可见，泰勒级数收敛且收敛到本身的函数实在是非常特殊的. 这样的函数称为**解析函数（analytic function）**. 详见幂级数[206].

导数与差分

预备知识 泰勒展开[181]

一阶导数

我们在导数的定义[158]中已经知道[8]

$$f'(x)=\lim_{h\to 0}\frac{f(x+h/2)-f(x-h/2)}{h} \tag{1}$$

在一些应用（如数值计算）中，我们只能把 h 取一个很小的数值（如 10^{-10}）而并非无穷小，这就需要我们估计用上式右边的**差分**来代替 $f'(x)$ 有多精确．为了估算误差，我们可以将 $f(x\pm h/2)$ 展开为关于 h 的泰勒级数

$$f(x\pm h/2)=f(x)\pm f'(x)\frac{h}{2}+\frac{1}{2}f''(x)\left(\frac{h}{2}\right)^2+\mathcal{O}\left(h^3\right) \tag{2}$$

代入式 1 得

$$\lim_{h\to 0}\frac{f'(x)h+\mathcal{O}(h^3)}{h}=f'(x)+\mathcal{O}\left(h^2\right) \tag{3}$$

所以用差分代替一阶导数可以精确到 h 的二阶无穷小 $\mathcal{O}(h^2)$.

二阶导数

能否用类似的方法来表示二阶导数呢？根据二阶导数的定义，我们需要用双重极限来表示

$$\begin{aligned}f''(x)&=\lim_{l\to 0}\frac{f'(x+l/2)-f'(x-l/2)}{l}\\&=\lim_{l\to 0}\lim_{h\to 0}\frac{1}{lh}[f(x+l/2+h/2)-f(x+l/2-h/2)\\&\qquad -f(x-l/2+h/2)+f(x-l/2-h/2)]\end{aligned} \tag{4}$$

但我们希望只用一个极限来表示二阶导数．然而我们不确定 h 是否需要是 l 的高阶无穷小．我们不妨来试试令 $l=h$，即

$$f''(x)=\lim_{h\to 0}\frac{f(x+h)-2f(x)+f(x-h)}{h^2} \tag{5}$$

[8] 以下假设 $f(x)$ 在某区间内处处可导.

要验证该式成立与否，将 $f(x\pm h)$ 关于 h 做泰勒展开得

$$f(x\pm h)=f(x)\pm f'(x)h+\frac{1}{2}f''(x)h^2\pm\frac{1}{6}f'''(x)h^3+\mathcal{O}\left(h^4\right) \tag{6}$$

代入式 5 右边得

$$\lim_{h\to0}\frac{f''(x)h^2+\mathcal{O}(h^4)}{h^2}=f''(x)+\mathcal{O}\left(h^2\right) \tag{7}$$

这就验证了式 5 的正确性．另外我们得知用差分来近似二阶导数 $f''(x)$ 同样是精确到二阶无穷小 $\mathcal{O}(h^2)$．

不定积分

预备知识 基本初等函数的导数[165]

一个实数函数 $f(x)$ 的**不定积分（indefinite integral）**是另一个函数 $F(x)$，叫做 $f(x)$ 的**原函数（primitive function）**．不定积分记为

$$F(x)=\int f(x)\,\mathrm{d}x \tag{1}$$

注意积分符号 $\int$ 和 $\mathrm{d}x$ 是一个整体算符，作用在他们中间的函数上[9]．

不定积分被定义为**求导的逆运算**．即若能找到 $F(x)$ 使其导数为 $f(x)$，那么 $F(x)$ 就是 $f(x)$ 的一个原函数．

$$F'(x)=f(x) \tag{2}$$

给出一个 $f(x)$，可以找到许多不同的原函数，且这些原函数都只相差一个常数．也就是说，给 $f(x)$ 的任意一个原函数加上一个常数 C，就可以得到 $f(x)$ 的另一个原函数．C 叫做**积分常数（constant of integration）**．

证明：由于常数导数[158] 为 0，给原函数加上常数后式 2 仍然成立

$$\frac{\mathrm{d}}{\mathrm{d}x}[F(x)+C]=f(x) \tag{3}$$

我们可以从几何上来理解该式：将函数曲线 $y=F(x)$ 整体在 y 方向平移并不影响某个 x 坐标处函数曲线的斜率．

[9]有时候为了方便也会记为 $\int\mathrm{d}x\,f(x)$

不定积分的基本性质

由于求导是线性运算，不定积分也是线性运算．即若干函数的线性组合的积分等于分别对这些函数积分再线性组合．令 a_n 为常数，有

$$\int [a_1 f_1(x) + a_2 f_2(x) \dots] \,\mathrm{d}x = a_1 \int f_1(x) \,\mathrm{d}x + a_2 \int f_2(x) \,\mathrm{d}x \dots \tag{4}$$

不定积分计算方法

与求导不同，计算不定积分没有特定的步骤，这里介绍几种方法

1. 最简单直接的方法是把已知的各种常见函数的导数写成积分的形式，例如已知 $\sin x$ 的导数是 $\cos x$，$\cos x$ 的积分就是 $\sin x$ 加任意常数．

2. 换元积分法[188]，包括第一类换元法和第二类换元法．

3. 分部积分法[190]

4. 查表法．许多高等数学教材（包括本书）都会给出一个积分表．当然，在信息技术发达的今天这种方法几乎已经被计算软件和网站取代．

5. 计算软件和网站．常见的符号计算软件有 Mathematica，Maple 等，数学网站有 Wolfram Alpha 等（建议先把积分技巧练熟再使用这些方法）．其中 Wolfram Alpha 对许多积分还会给出详细的计算步骤．

对于一些常用积分，一般要求能熟记或快速推出．见积分表[192] 中的**常用积分**部分．

换元积分法

预备知识 不定积分[187]，复合函数求导[174]

第一类换元积分法

由复合函数的求导法则，令 $F'(x)=f(x)$，则

$$\frac{\mathrm{d}}{\mathrm{d}x}F[u(x)]=f[u(x)]u'(x) \tag{1}$$

由于求导的逆运算是积分，有

$$\int f[u(x)]u'(x)\,\mathrm{d}x=F[u(x)]+C \tag{2}$$

所以如果某个积分可以看成 $\int f[u(x)]u'(x)\,\mathrm{d}x$ 的形式，且 $F(x)$ 较容易求出，即可根据式 2 写出结果．这种方法叫做**第一类换元积分法**．这类换元积分法的技巧就在于如何看出被积函数的的结构是 $\int f[u(x)]u'(x)\,\mathrm{d}x$，只有多练习才能熟能生巧．

例 1

计算

$$\int a\sin(ax+b)\,\mathrm{d}x \tag{3}$$

令 $f(x)=\sin(x)$，$u(x)=ax+b$，则上式刚好是 $\int f[u(x)]u'(x)\,\mathrm{d}x$ 的形式．从基本初等函数积分表已知 $\sin x$ 的一个原函数是 $F(x)=-\cos x$，那么答案就是

$$F[u(x)]+C=-\cos(ax+b)+C \tag{4}$$

总结到更一般的情况，根据换元积分法，若已知 $\int f(x)\,\mathrm{d}x=F(x)+C$，则对于任意常数 a 和 b，必有 $\int a\,f(ax+b)\,\mathrm{d}x=F(ax+b)$．根据积分的基本性质，两边同除 a，得

$$\int f(ax+b)\,\mathrm{d}x=\frac{1}{a}F(ax+b)+C \tag{5}$$

该式使用频率很高，需要熟练掌握．类似例子还有

$$\int\frac{1}{ax+b}\,\mathrm{d}x=\frac{1}{a}\ln(ax+b)+C\qquad\int\mathrm{e}^{ax+b}\,\mathrm{d}x=\frac{1}{a}\mathrm{e}^{ax+b}+C \tag{6}$$

等等．

积分变量替换

换元积分法的过程在形式上可以记为（见微分[173]）

$$\begin{aligned}\int f[u(x)]u'(x)\,\mathrm{d}x &= \int f[u(x)]\,\mathrm{d}[u(x)] = \int f(u)\,\mathrm{d}u = F(u)+C\\ &= F[u(x)]+C\end{aligned} \tag{7}$$

该式把积分变量由 x 换成了 u，故称为**换元积分法**.

第二类换元积分法

第二类换元积分法从某种意义上和第一类换元积分相反. 若要对一个函数积分，先把它的自变量看做另一个变量的函数，再逆向使用式 7，即可化简积分.

$$\int f(x)\,\mathrm{d}x = \int f[x(t)]\,\mathrm{d}[x(t)] = \int f[x(t)]x'(t)\,\mathrm{d}t \tag{8}$$

这个积分看似复杂了，但是如果 $x(t)$ 选取适当，反而可以使计算化简.

例 2

计算

$$\int \frac{\mathrm{d}x}{\sqrt{1-x^2}} \tag{9}$$

显然 $x\in(-1,1)$，选取 $x(t)=\sin t$. 替换后的定义域为 $t\in(-\pi/2,\pi/2)$，函数单调递增[10]. 上面积分变为

$$\int \frac{\mathrm{d}x}{\sqrt{1-x^2}} = \int \frac{\mathrm{d}\sin t}{\sqrt{1-\sin^2 t}} = \int \frac{\cos t\,\mathrm{d}t}{\cos t} = \int 1\,\mathrm{d}t = t+C = \arcsin x + C \tag{10}$$

验证：根据反函数求导法则[163]

$$\arcsin' x = \frac{1}{\cos(\arcsin x)} = \frac{1}{\sqrt{1-x^2}} \tag{11}$$

分部积分法

[10]注意任何积分换元法中的两个变量必须有一一对应的关系，即相互的函数关系在定义域内都为单调.

预备知识 牛顿—莱布尼兹公式[202]

若积分中的被积函数可以表示为两个函数的乘积，则我们可以使用分部积分公式，$f(x)$ 和 $g(x)$ 分别为 $F(x)$ 和 $G(x)$ 的导函数，有

$$\int F(x)g(x)\,\mathrm{d}x = F(x)G(x) - \int f(x)G(x)\,\mathrm{d}x \tag{1}$$

$$\int_a^b F(x)g(x)\,\mathrm{d}x = F(x)G(x)\Big|_a^b - \int_a^b f(x)G(x)\,\mathrm{d}x \tag{2}$$

用 $f^{(n)}(x)$ 表示 n 阶导数，$f^{[n]}(x)$ 表示 n 次不定积分[11]，连续使用 n 次分部积分公式，有

$$\begin{aligned}\int f(x)g(x)\,\mathrm{d}x = {} & f(x)g^{[1]}(x) - f^{(1)}(x)g^{[2]}(x) + \cdots + (-1)^{n-1}f^{(n-1)}(x)g^{[n]}(x) \\ & + (-1)^n \int f^{(n)}(x)g^{[n]}(x)\,\mathrm{d}x\end{aligned} \tag{3}$$

推导

令 $f(x) = F'(x)$，$g(x) = G'(x)$，根据乘法的求导公式

$$[F(x)G(x)]' = f(x)G(x) + F(x)g(x) \tag{4}$$

即

$$F(x)g(x) = [F(x)G(x)]' - f(x)G(x) \tag{5}$$

两边不定积分（积分常数可任取）得

$$\int F(x)g(x)\,\mathrm{d}x = F(x)G(x) - \int f(x)G(x)\,\mathrm{d}x \tag{6}$$

所以如果被积函数等于两个函数的乘积，则可选择其中一个（F ）为“求导项”进行求导，另一个（g）为“积分项”进行不定积分（积分常数可任取），然后代入该式即可.

若要计算定积分，既可以先计算不定积分然后使用牛顿—莱布尼兹公式，也可以直接对式 5 进行定积分得

$$\int_a^b F(x)g(x)\,\mathrm{d}x = [F(x)G(x)]_a^b - \int_a^b f(x)G(x)\,\mathrm{d}x \tag{7}$$

[11]这是笔者自己发明的符号

例 1　求 xe^{-x} 的不定积分和从 0 到 $+\infty$ 的定积分

令 x 项为“求导项”，导数为 1，e^{-x} 为“积分项”，积分为 $-e^{-x}$．代入式 6 得

$$\int xe^{-x}\,dx = x(-e^{-x}) - \int 1\times(-e^{-x})\,dx = -xe^{-x} - e^{-x} + C \tag{8}$$

如果直接计算定积分，把“求导项”和“积分项”直接代入式 7 得

$$\int_0^{+\infty} xe^{-x}\,dx = x(-e^{-x})\Big|_0^{+\infty} - \int_0^{+\infty} 1\times(-e^{-x})\,dx = 0 - e^{-x}\Big|_0^{+\infty} = 1 \tag{9}$$

多次分部积分

由于 $f(x)$ 的 n 次导数可以记为 $f^{(n)}(x)$，不妨把 $g(x)$ 的 n 次不定积分（n 个积分常数任取）记为 $g^{[n]}(x)$．则分部积分式 6 可记为

$$\int f(x)g(x)\,dx = f(x)g^{[1]}(x) - \int f^{(1)}(x)g^{[1]}(x)\,dx \tag{10}$$

再对第二项利用分部积分，仍然将 $f^{(1)}$ 作为“求导项”，$g^{[1]}$ 作为“积分项”，得

$$\int f(x)g(x)\,dx = f(x)g^{[1]}(x) - f^{(1)}(x)g^{[2]}(x) + \int f^{(2)}(x)g^{[2]}(x)\,dx \tag{11}$$

再把 $f^{(2)}$ 作为“求导项”，$g^{[2]}$ 作为“积分项”，分布积分得

$$\begin{aligned}\int f(x)g(x)\,dx = {} & f(x)g^{[1]}(x) - f^{(1)}(x)g^{[2]}(x) + f^{(2)}(x)g^{[3]}(x) \\ & - \int f^{(3)}(x)g^{[3]}(x)\,dx\end{aligned} \tag{12}$$

可以发现若要使用 N 次分部积分，第 $i \leqslant N$ 项等于第 $i-1$ 项中的“求导项”求导，“积分项”积分，再取相反数，最后不定积分中只需把“求导项”额外求一次导即可．

积分表

预备知识 不定积分[187]

这里给出一个基本积分表和一个常用积分表，前者建议熟记．部分积分有的给出计算步骤，没有给出则是由基本初等函数的导数[165]直接逆向得出．所有的不定积分公式都可以通过求导验证．

应用换元积分法[188]，表中任何积分都可以拓展为

$$\int f(ax+b)\,\mathrm{d}x = \frac{1}{a}F(ax+b)+C \tag{1}$$

基本积分表

$$\int x^a\,\mathrm{d}x = \frac{1}{a+1}x^{a+1}+C \quad (a\in R, a\neq -1) \tag{2}$$

$$\int \frac{1}{x}\,\mathrm{d}x = \ln|x|+C \quad （例 6） \tag{3}$$

$$\int \cos x\,\mathrm{d}x = \sin x + C \tag{4}$$

$$\int \sin x\,\mathrm{d}x = -\cos x + C \tag{5}$$

$$\int \tan x\,\mathrm{d}x = -\ln|\cos x|+C \quad （例 2） \tag{6}$$

$$\int \cot x\,\mathrm{d}x = \ln|\sin x|+C \quad （例 3） \tag{7}$$

$$\int \frac{1}{\cos^2 x}\,\mathrm{d}x = \tan x + C \tag{8}$$

$$\int \frac{1}{1+x^2}\,\mathrm{d}x = \arctan x + C \tag{9}$$

$$\int \mathrm{e}^x\,\mathrm{d}x = \mathrm{e}^x + C \tag{10}$$

$$\int x\mathrm{e}^x\,\mathrm{d}x = \mathrm{e}^x(x-1)+C \quad （例 7） \tag{11}$$

$$\int a^x\,\mathrm{d}x = \frac{1}{\ln a}a^x + C \quad （例 1） \tag{12}$$

常用积分表

$$\int \sin^2 x \,\mathrm{d}x = \frac{1}{2}(x - \sin x \cos x) + C \quad （例 4）\tag{13}$$

$$\int \cos^2 x \,\mathrm{d}x = \frac{1}{2}(x + \sin x \cos x) + C \quad （例 5）\tag{14}$$

$$\int \sec x \,\mathrm{d}x = \ln|\tan x + \sec x| + C \quad （例 11）\tag{15}$$

$$\int \ln x \,\mathrm{d}x = x \ln x - x + C \quad （例 8）\tag{16}$$

$$\int \sqrt{a^2 - x^2} \,\mathrm{d}x = \frac{1}{2}\left(x\sqrt{a^2 - x^2} + a^2 \arcsin \frac{x}{a}\right) + C \quad （例 9）\tag{17}$$

$$\int \frac{1}{\sqrt{1 - x^2}} \,\mathrm{d}x = \arcsin(x) + C \quad （例 10）\tag{18}$$

$$\int \frac{1}{\sqrt{1 + x^2}} \,\mathrm{d}x = \ln\left(x + \sqrt{1 + x^2}\right) + C = \sinh^{-1} x + C \quad （例 12）\tag{19}$$

定积分

$$\int_0^{\pi/2} \cos^n \theta \,\mathrm{d}\theta = \frac{\sqrt{\pi}}{2} \frac{\Gamma(n/2 + 1/2)}{\Gamma(n/2 + 1)}\tag{20}$$

其中 n 是满足 $\mathrm{Re}[n] > -1$ 的任意复数[12].

例 1

$$\int a^x \,\mathrm{d}x\tag{21}$$

我们已经知道如何算 e^x 的积分，而 $a = \mathrm{e}^{\ln a}$，再根据式 1 就有

$$\int \mathrm{e}^{\ln(a)x} \,\mathrm{d}x = \frac{1}{\ln a}\mathrm{e}^{\ln(a)x} + C = \frac{1}{\ln a}a^x + C\tag{22}$$

[12]结果来自 Wolfram Alpha

例 2

$$\int \tan x \,\mathrm{d}x \tag{23}$$

这个积分用第一类换元积分法（式 2 [188]）

$$\int f[u(x)]u'(x)\,\mathrm{d}x = F[u(x)] + C \tag{24}$$

首先 $\tan x = \sin x/\cos x$，令 $u(x) = \cos x$，则 $\sin x = -u'(x)$，对比得 $f(x) = -1/x$ 其原函数为 $F(x) = -\ln|x|$，所以

$$\int \tan x \,\mathrm{d}x = \int f[u(x)]u'(x)\,\mathrm{d}x = F[u(x)] + C = -\ln|\cos x| + C \tag{25}$$

例 3

类似例 2，$\cot x = \cos x/\sin x$，令 $u(x) = \sin x$，则 $\cos x = u'(x)$，对比得 $f(x) = 1/x$，原函数为 $F(x) = \ln|x|$（式 3），所以

$$\int \cot x \,\mathrm{d}x = F[u(x)] + C = \ln|\sin x| + C \tag{26}$$

例 4

$$\int \sin^2 x \,\mathrm{d}x \tag{27}$$

用降幂公式（式 7 [112]）和不定积分的线性（式 4 [187]）把上式变为常数的积分和 $\cos 2x$ 的积分，再利用式 4 和式 1 计算后者即可

$$\begin{aligned}\int \sin^2 x \,\mathrm{d}x &= \int \frac{1}{2}\,\mathrm{d}x - \frac{1}{2}\int \cos 2x \,\mathrm{d}x \\ &= \frac{x}{2} - \frac{1}{4}\sin(2x) = \frac{1}{2}(x - \sin x \cos x) + C\end{aligned} \tag{28}$$

例 5

$$\int \cos^2 x \,\mathrm{d}x \tag{29}$$

与例 4 类似，用三角恒等式 $\cos^2(x) = [1 + \cos(2x)]/2$ 得

$$\begin{aligned}\int \cos^2 x \,\mathrm{d}x &= \int \frac{1}{2}\,\mathrm{d}x + \frac{1}{2}\int \cos(2x)\,\mathrm{d}x \\ &= \frac{x}{2} + \frac{1}{4}\sin(2x) = \frac{1}{2}(x + \sin x \cos x) + C\end{aligned} \tag{30}$$

例 6

$$\int \frac{1}{x}\,\mathrm{d}x \tag{31}$$

首先在区间 $(0,+\infty)$ 内，由于 $\ln x$ 的导数是 $1/x$，所以积分结果为 $\ln x + C$. 现在再来考虑区间 $(-\infty,0)$，注意 $\ln x$ 在这里没有定义，不妨看看 $\ln(-x)$，由复合函数求导，其导数恰好为 $1/x$. 所以在除去原点的实数范围内，有

$$\int \frac{1}{x}\,\mathrm{d}x = \ln|x| + C \tag{32}$$

事实上，由于 $1/x$ 在 $x=0$ 没有定义，更广义的原函数可以取

$$\int \frac{1}{x}\,\mathrm{d}x = \begin{cases} \ln x + C_1 & (x>0) \\ \ln(-x) + C_2 & (x<0) \end{cases} \tag{33}$$

其中 C_1 和 C_2 是两个不相同的待定常数.

例 7

$$\int x\mathrm{e}^x\,\mathrm{d}x \tag{34}$$

使用用分部积分式 1 [190]

$$\int F(x)g(x)\,\mathrm{d}x = F(x)G(x) - \int f(x)G(x)\,\mathrm{d}x \tag{35}$$

令 $F(x)=x$，求导得 $f(x)=1$，令 $g(x)=\mathrm{e}^x$，由式 10，$G(x)=\mathrm{e}^x$. 代入分部积分得

$$\int x\mathrm{e}^x\,\mathrm{d}x = x\mathrm{e}^x - \int 1\cdot\mathrm{e}^x\,\mathrm{d}x = \mathrm{e}^x(x-1) + C \tag{36}$$

例 8

$$\int \ln x\,\mathrm{d}x \tag{37}$$

方法一：使用第二类换元法式 8 [188]

$$\int f(x)\,\mathrm{d}x = \int f[x(t)]\,\mathrm{d}[x(t)] = \int f[x(t)]x'(t)\,\mathrm{d}t \tag{38}$$

令[13] $x = \mathrm{e}^t$，求导得 $x'(t) = \mathrm{e}^t$，换元得

$$\int \ln x \,\mathrm{d}x = \int \ln\left(\mathrm{e}^t\right)\mathrm{e}^t \,\mathrm{d}t = \int t\mathrm{e}^t \,\mathrm{d}t \tag{39}$$

由例 7 中的分部积分得

$$\int \ln x \,\mathrm{d}x = \mathrm{e}^t(t-1) + C = \mathrm{e}^{\ln x}(\ln x - 1) + C = x(\ln x - 1) + C \tag{40}$$

方法二：直接使用分部积分法式 1 [190]，对常数 1 积分，对 $\ln x$ 求导，得

$$\int \ln x \,\mathrm{d}x = x\ln x - \int x \cdot \frac{1}{x} \,\mathrm{d}x = x\ln x - x + C \tag{41}$$

例 9

$$\int \sqrt{a^2 - x^2} \,\mathrm{d}x \tag{42}$$

使用第二类换元法式 8 [188]，令 $x = a\sin t$ 得

$$\int a\cos t \,\mathrm{d}(a\sin t) = a^2 \int \cos^2 t \,\mathrm{d}t \tag{43}$$

将例 5 的结论代入得 $a^2(t + \sin t\cos t) + C$，再将 $t = \arcsin(x/a)$ 代入得

$$\int \sqrt{a^2 - x^2} \,\mathrm{d}x = \frac{1}{2}\left(x\sqrt{a^2 - x^2} + a^2 \arcsin\frac{x}{a}\right) + C \tag{44}$$

例 10

$$\int \frac{1}{\sqrt{1 - x^2}} \,\mathrm{d}x \tag{45}$$

使用第二类换元法式 8 [188]，令 $x = \sin t$ 得

$$\int \frac{1}{\sqrt{1 - \sin^2 t}} \,\mathrm{d}(\sin t) = \int \mathrm{d}t = t + C = \arcsin x + C \tag{46}$$

例 11

$$\int \sec x \,\mathrm{d}x \tag{47}$$

[13]注意被积函数只在 $x > 0$ 区间有定义，否则使用 $x = \mathrm{e}^t$ 将会自动忽略 $x \leqslant 0$ 的情况.

分子分母同时乘以 $\sec x+\tan x$，可以发现分子是分母的导数．再用第一类换元积分法（式 2 [188]），令 $u(x)=\sec x+\tan x$，再使用式 3 即可

$$\begin{aligned}\int \sec x\,\mathrm{d}x &= \int \frac{\sec^2 x+\sec x\tan x}{\sec x+\tan x}\,\mathrm{d}x = \int \frac{u'(x)}{u}\,\mathrm{d}x = \int \frac{1}{u}\,\mathrm{d}u \\ &= \ln|u|+C = \ln|\sec x+\tan x|+C\end{aligned} \tag{48}$$

例 12

$$\int \frac{1}{\sqrt{1+x^2}}\,\mathrm{d}x \tag{49}$$

使用第二类换元法式 8 [188]，令 $x=\tan t$，再利用“三角恒等式[112]”的式 2 和式 3 得

$$\int \frac{1}{\sqrt{1+\tan^2 t}}\,\mathrm{d}(\tan t) = \int \frac{1}{\sec t}\sec^2 t\,\mathrm{d}t = \ln|\tan t+\sec t|+C \tag{50}$$

由同一三角恒等式，$\sec t=\sqrt{1+\tan^2 t}=\sqrt{1+x^2}$，所以

$$\int \frac{1}{\sqrt{1+x^2}}\,\mathrm{d}x = \ln\left(x+\sqrt{1+x^2}\right)+C \tag{51}$$

注意上式中 ln 后面的绝对值符号消失是因为 $x+\sqrt{1+x^2}\geqslant 0$ 恒成立．另外由“双曲函数[114]”中例 1 可知上式又等于 $\sinh^{-1}x+C$.

定积分

预备知识 导数[158]

首先以不均匀细绳的质量为例，引入定积分的思想

例 1　不均匀细绳的质量

一条密度不均匀的绳子长为 L，横截面积是 S，细绳距离 O 端 $x(x<L)$ 处的密度为 $\rho(x)$．求绳子的质量．

如果题目中，密度是恒定的，那么直接可以写出绳子的质量为 $m=LS\rho$. 但是题中 $\rho(x)$ 是关于 x 的函数，所以我们要寻找另外的做法．假设绳子的密度变化是连续且“平滑”的，我们可以通过把绳子分割成 n 小节（注意这些小

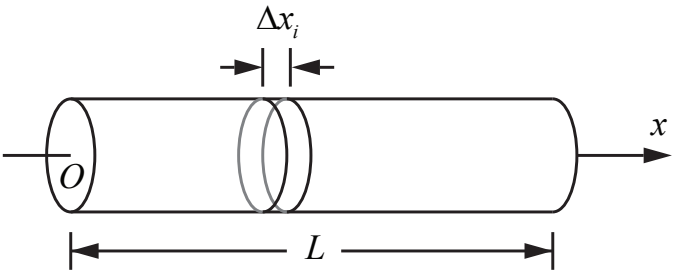


图 1: 密度不均匀的绳子

节必须严格地首尾相接，不能有重合或者空隙）. 第 i 节取 x_i 到 x_{i+1}，令其长度为 $x_{i+1}-x_i=\Delta x_i$ 使每一个小节内，密度可以近似看成是恒定的，这样我们可以用 $\rho(\xi_i)$ $(x_i\leqslant\xi_i\leqslant x_{i+1})$ 来代替第 i 节的密度，当每一节足够小时，可以认为 ξ_i 在 $x_i\leqslant\xi_i\leqslant x_{i+1}$ 约束下的取值并不会影响结果. 第 i 小节的质量为

$$\Delta m_i=\rho(\xi_i)\Delta x_i S \tag{1}$$

所以总的质量用求和符号来表示，就是

$$m=\sum_{i=1}^{n}\Delta m_i\approx\sum_{i=1}^{n}\rho(\xi_i)\Delta x_i S=S\sum_{i=1}^{n}\rho(\xi_i)\Delta x_i \tag{2}$$

由于当 n 取有限值时，上式并不精确成立，所以只能使用约等号，但是 n 越大，约等号两边就越精确成立. 这是极限的思想，用极限符号来表，就是

$$m=\lim_{n\to\infty}\sum_{i=1}^{n}\Delta m_i=\lim_{n\to\infty}\sum_{i=1}^{n}\rho(\xi_i)\Delta x_i S=S\lim_{n\to\infty}\sum_{i=1}^{n}\rho(\xi_i)\Delta x_i \tag{3}$$

这种表达式在物理中反复出现，所以使用积分符号 $\int$ 用于代替极限和求和符号. 另外把 ξ_i 写成 x（当 n 趋近于无穷大时，参量 i 和 Δx_i 具体是多少就不重要了），把表示增量的 Δ 变为表示微分的 d，上式就写为

$$m=\int \mathrm{d}m=\int S\rho(x)\,\mathrm{d}x=S\int\rho(x)\,\mathrm{d}x \tag{4}$$

下面先看看 $\int\rho(x)\,\mathrm{d}x$，即 $\lim\limits_{n\to\infty}\sum\limits_{i=1}^{n}\rho(\xi_i)\Delta x_i$ 的另一种理解. 画出 $\rho(x)$ 图像. 例如 $\rho(x)=x+1$，则 $\rho(\xi_i)\Delta x_i$ 可以表示左图的第 i 个小长方形的面积，$\sum\limits_{i=1}^{n}\rho(\xi_i)\Delta x_i$ 表示长方形面积之和. 如果 n 非常大且每个 Δx_i 取得非常小，左图看起来就会像右图. 所以 $\int\rho(x)\,\mathrm{d}x$ 可以用来表示右图阴影部分的面积.

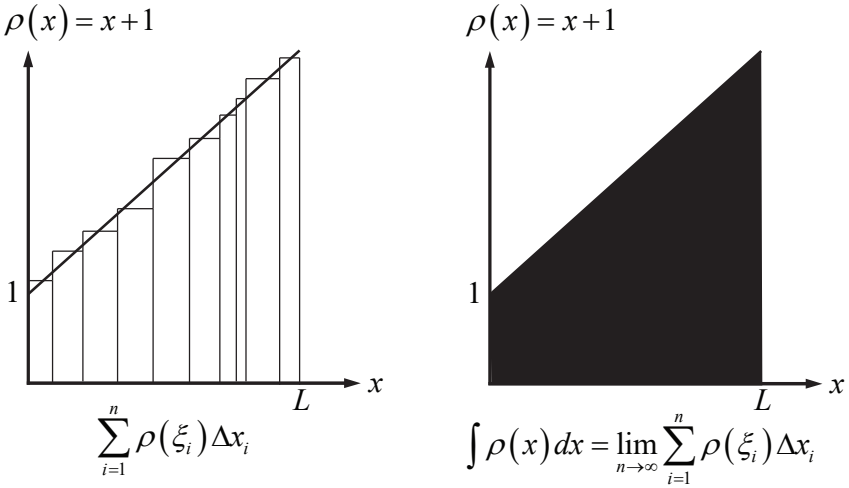


图 2: 定积分可以理解为曲线下面的面积，并看做由无限多个无穷窄的矩形组成.

但 $\int \rho(x)\,\mathrm{d}x$ 里面显然不包含 0 和 L 的信息，我们根据题目中的情况，说这个积分是“从 0 积到 L”，其中 0 是积分下限，L 是积分上限. 为了表示这个信息，把它写到积分号右边变为

$$\int_0^L \rho(x)\,\mathrm{d}x \tag{5}$$

这就是定积分的标准形式，但有时候为了书写方便，在不混淆的情况下可以把积分上下限省略.

这样的写法是很形象的，可以想象，积分号就是函数的曲线需要积分的部分，下标的位置代表曲线的起点，上标代表曲线的终点. 这样，物理中很多问题就可以用积分表示了.

要注意的是，根据上面积分的定义，如果曲线在 x 轴的下方，面积应该表示成负值. 但根据例 1 的物理情景，可知密度不可能是负值.

至于计算积分的具体方法，比求导要复杂得多，甚至很多积分的结果不能用初等函数表示，只能表示为级数等形式. 然而对于基本初等函数的积分，用牛顿－莱布尼兹公式[202] 即可马上求解.

例 2　求圆的面积

在直角坐标系中，圆的方程为 $x^2 + y^2 = R^2$，上半圆的方程可看做 y 关于

x 的函数

$$y = f(x) = \sqrt{R^2 - x^2} \quad (x \in [-R, R]) \tag{6}$$

将该式定积分再乘以 2 即可得到圆的面积

$$S = 2\int_{-R}^{R} f(x)\,\mathrm{d}x = 2\int_{-R}^{R} \sqrt{R^2 - x^2}\,\mathrm{d}x \tag{7}$$

我们还可以用另一种方法验证圆的面积公式. 把原划分成许多微小圆环, 由例 2 [173], 每个微小圆环的面积为 $2\pi r\,\mathrm{d}r$, 所以圆的面积可以用定积分表示为

$$S = \int_0^R 2\pi r\,\mathrm{d}r \tag{8}$$

以上两个定积分的结果都为 πR^2, 过程见“牛顿—莱布尼兹公式[202]”的例 2.

例 3　球体的表面积

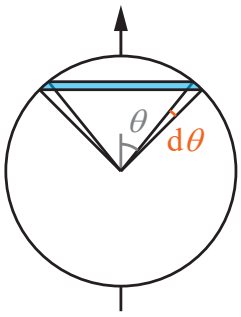


图 3: 将球的表面划分成许多细圆环, 每个对应的极角为 $\mathrm{d}\theta$

以球心为原点建立球坐标系, 我们可以把球体的表面根据不同的 θ 划分成许多细圆环（如图 3), 每个圆环的面积等于周长乘以宽度, 即

$$\mathrm{d}S = 2\pi R\sin\theta \cdot R\,\mathrm{d}\theta \tag{9}$$

所以球的表面积可以用定积分记为

$$S = \int_0^{\pi} 2\pi R\sin\theta \cdot R\,\mathrm{d}\theta = 2\pi R^2 \int_0^{\pi} \sin\theta\,\mathrm{d}\theta \tag{10}$$

虽然我们还不会计算这个定积分（见“牛顿—莱布尼兹公式[202]”), 但我们现在可以用一种巧妙的方法来简化问题. 让我们来计算每个细圆环在极轴方

向投影的长度．我们不妨把极轴叫做 z 轴，则对某个细圆环有 $z = R\cos\theta$，微分得 $\mathrm{d}z = -R\sin\theta\,\mathrm{d}\theta$，将该式消去式 9 中的 $\mathrm{d}\theta$ 得

$$\mathrm{d}S = -2\pi R\,\mathrm{d}z \tag{11}$$

这说明无论细圆环的位置如何，其面积与其在 z 轴投影的长度的比值恒为 2π．至于上式中的负号，是因为我们假设了正的 $\mathrm{d}\theta$ 对应正的面积，而正的 $\mathrm{d}\theta$ 却对应负的 $\mathrm{d}z$．由于面积恒为正值，我们可以取绝对值将负号去掉．这样，球的表面积就可以用定积分表示为

$$S = \int_{-R}^{R} 2\pi R\,\mathrm{d}z \tag{12}$$

由于被积函数是一个常数，定积分的结果就是该常数乘以积分区间的长度即 $4\pi R^2$．

例 4　球的体积

要计算一个半径为 R 的球体的体积，可以将球划分为无限个薄球壳，每个薄球壳的体积等于该球壳的表面积乘以厚度（见例 2 [173]），即 $\mathrm{d}V = 4\pi r^2\,\mathrm{d}r$．所以球的体积可用定积分表示为

$$V = \int_{0}^{R} 4\pi r^2\,\mathrm{d}r \tag{13}$$

同样由“牛顿—莱布尼兹公式[202]”可得积分结果为 $4\pi R^3/3$．

牛顿—莱布尼兹公式

预备知识　不定积分[187]，定积分[198]

牛顿—莱布尼兹公式描述了定积分和不定积分的关系．我们已知不定积分是求导的逆运算，而定积分是函数曲线与 x 轴之间的面积，二者乍看起来没什么联系，但牛顿—莱布尼兹公式却揭示了了二者之间的重要关系．

若 $F(x)$ 是 $f(x)$ 的一个原函数[187]，则

$$\int_{a}^{b} f(x)\,\mathrm{d}x = F(b) - F(a) \tag{1}$$

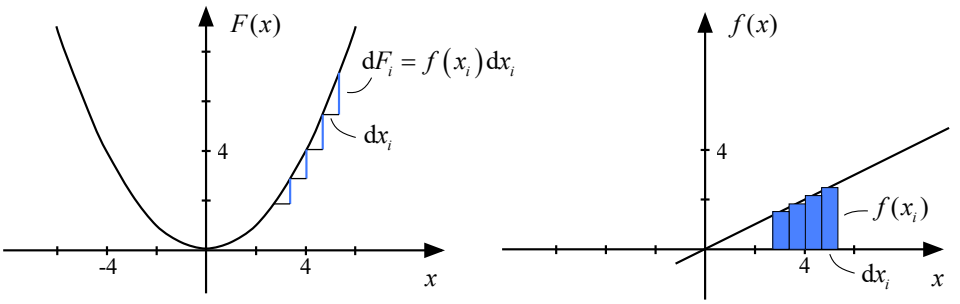


图 1: 右图中 $f(x)$ 的原函数为左图中的 $F(x)$，当步长趋近 0 时，右图中的长方形面积趋近于左图中小竖线的长度.

推导

如图 1，根据定积分[198] 的定义，有[14]

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{\Delta x_i \to 0} \sum_i f(x_i)\Delta x_i \tag{2}$$

其中 $f(x_i)\Delta x_i$ 可看成是右图中第 i 个小矩形的面积，求和是对从 a 到 b 的所有小矩形求和. 现在不妨把 x_i 设为第 i 个小矩形左端的 x 坐标. 考虑到求导是不定积分的逆运算，有 $f(x_i) = F'(x_i)$，所以小矩形的面积变为

$$f(x_i)\Delta x_i = F'(x_i)\Delta x_i \approx \Delta F_i = F(x_{i+1}) - F(x_i) \tag{3}$$

最后一步使用了微分近似. 该式可以理解成，右图中的小矩形面积约等于左图中的小竖线长度，即原函数 $F(x)$ 在 x_i 到 x_{i+1} 间的增量. 当取极限 $\Delta x_i \to 0$ 时，上式取等号. 代回式 1，有

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{\Delta x_i \to 0} \sum_i [F(x_{i+1}) - F(x_i)] = F(b) - F(a) \tag{4}$$

该式可理解为，如果把左图中每一段 Δx_i 所对应的微小增量 ΔF_i 都加起来，再取极限 $\Delta x_i \to 0$，就是 $F(x)$ 从 a 到 b 的总增量. 在计算定积分的过程中，为了书写简洁，我们往往将上式中的 $F(b) - F(a)$ 记为 $F(x)\big|_a^b$.

例 1　计算定积分

$$\int_{-l}^{l} \sin^2\left(\frac{n\pi}{l}x\right)\mathrm{d}x \tag{5}$$

[14] 这里假设极限存在.

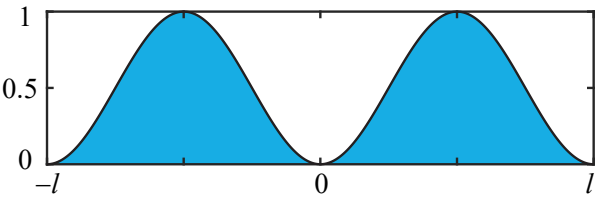


图 2: $y=\sin^2(\pi x/l)$ 的定积分

先计算对应的不定积分．由积分表[192]中的式 13 结合式 1 得不定积分为

$$\int \sin^2(\frac{n\pi}{l}x)\,\mathrm{d}x=\frac{l}{2n\pi}\Big[\frac{n\pi}{l}x-\sin\Big(\frac{n\pi}{l}x\Big)\cos\Big(\frac{n\pi}{l}x\Big)\Big] \tag{6}$$

再利用牛顿—莱布尼兹公式求定积分结果为 1．计算该定积分还有另一种更简单的几何方法（见图 2），由于被积函数的对称性，函数曲线可将区间 $[-l,l]$ 内高为 1 的长方形（面积为 $2l$ ）划分成等面积的上下两部分，曲线下方的面积 l 就是定积分的结果．

例 2　圆的面积

现在我们可以用例 2[198]中列出的两个定积分计算圆的面积．先看第一个定积分，由积分表式 17[192]得

$$\int \sqrt{R^2-x^2}\,\mathrm{d}x=\frac{1}{2}\Big(x\sqrt{R^2-x^2}+R^2\arcsin\frac{x}{R}\Big)+C \tag{7}$$

由牛顿—莱布尼兹公式，$-R$ 到 R 的定积分为 $\pi R^2/2$，所以圆的面积为 πR^2．

第二个定积分要简单得多，由幂函数的积分式 2[192]和牛顿—莱布尼兹公式得

$$\int_0^R 2\pi r\,\mathrm{d}r=\pi\,r^2\Big|_0^R=\pi R^2 \tag{8}$$

例 3　球壳与球盖的面积

现在我们可以直接求例 3[198]中的积分

$$S=2\pi R^2\int_0^\pi \sin\theta\,\mathrm{d}\theta=2\pi R^2\,(-\cos\theta)\Big|_0^\pi=4\pi R^2 \tag{9}$$

我们还可以将积分上下限任意改变，得到球面上一个环形曲面的面积

$$S=2\pi R^2\int_{\theta_1}^{\theta_2}\sin\theta\,\mathrm{d}\theta=2\pi R^2(\cos\theta_1-\cos\theta_2) \tag{10}$$

当 $\theta_1=0$ 且 $\theta_2=\alpha$ 时，我们就得到了球盖的面积

$$S=2\pi R^2(1-\cos\alpha) \tag{11}$$

对定积分上下限求导

有时候我们会需要对定积分的上下限求导，例如

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x f(t)\,\mathrm{d}t \tag{12}$$

我们可以先对定积分用牛顿莱布尼兹公式，令原函数为 $F(x)$，有

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x f(t)\,\mathrm{d}t = \frac{\mathrm{d}}{\mathrm{d}x}[F(x)-F(a)] = f(x) \tag{13}$$

类似地，对积分下限求导如

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_x^a f(t)\,\mathrm{d}t = \frac{\mathrm{d}}{\mathrm{d}x}[F(a)-F(x)] = -f(x) \tag{14}$$

或者对上下限同时求导如

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_{-x}^x f(t)\,\mathrm{d}t = \frac{\mathrm{d}}{\mathrm{d}x}[F(x)-F(-x)] = f(x)+f(-x) \tag{15}$$

幂级数

- 本词条处于草稿阶段.

预备知识 极限[149]，复数[142]

我们把形如

$$s = \sum_{n=0}^{\infty} c_n(x-x_0)^n \tag{1}$$

的表达式叫做**幂级数（power series）**. 其中 $(x-x_0)^0$ 始终视为 1，即使 $x-x_0=0$.

幂级数是无穷级数[??] 的一种，是一个极限. 如果我们把前有限项的求和记为

$$s_m = \sum_{n=0}^{m} c_n(x-x_0)^n \tag{2}$$

那么式 1 幂级数就是

$$s = \lim_{m\to\infty} s_m \tag{3}$$

的简写. 对给定的 x，当极限存在时，我们就说级数**收敛（converge）**，反之就说级数**不收敛**或**发散（diverge）**. 广义来说，幂级数中的 c_n, x, x_0 都可以是复数复数[142].

收敛半径

一种极端的情况是幂级数式 1 只在 $x = x_0$ 一点处收敛（例如 $c_n = n^n$）. 除此之外，必定存在一个**收敛半径（radius of convergence）** $0 < r < 1$，使得当 $|x - x_0| < r$ 时级数总是收敛，而 $|x - x_0| > r$ 时总是发散（不收敛）. 当 x 是复数时，复平面上收敛的区域就是以 x_0 为圆心的一个开圆盘. 当 $|x - x_0| = r$ 时，幂级数可能收敛也可能不收敛.

计算收敛半径的一种简单方法是使用比值判别法

$$r = \lim_{n\to\infty} \left| \frac{c_n}{c_{n+1}} \right| \tag{4}$$

但前提是该极限存在，普适的方法见柯西—阿达玛公式[253].

幂级数 2

预备知识 泰勒级数[181]

在复数域上，形如

$$\sum_{n=0}^{\infty} c_n (z-a)^n \tag{1}$$

的级数称为**幂级数（power series）**，这里 c_n 皆为复数，未定元 z 一般也视为复数.

幂级数的收敛域

定理 1

如果由式 1 给出的幂级数在某点 $z_0 \neq a$ 处收敛，那么它一定在开圆盘 $|z-a| < |z_0 - a|$ 上绝对收敛且内闭一致收敛.

证明很简单：如果 $\sum_{n=0}^{\infty} c_n (z-a)^n$ 在 $z = z_0$ 时收敛，那么 $c_n (z_0 - a)^n \to 0$，从而有一 M 使得 $|c_n (z_0 - a)^n| \leq M$ 对任何 n 都成立. 故对于任何固定的

$0<q<1$，当 $|z-a|<q|z_0-a|$ 时就有

$$\sum_{n=0}^{\infty}|c_n(z-a)^n|=\sum_{n=0}^{\infty}|c_n(z_0-a)^n|\frac{|z-a|^n}{|z_0-a|^n}<\sum_{n=0}^{\infty}Mq^n.$$

于是幂级数诸项绝对值由收敛的几何级数控制.

由此可见，式 1 给出的幂级数的收敛域或者只是一个点 a，或者至少包含某个以 a 为圆心的开圆盘. 这样的开圆盘中最大者叫做幂级数的**收敛圆（disk of convergence）**，其半径称为**收敛半径（radius of convergence）**. 幂级数的收敛半径由柯西-阿达玛公式[253] 给出. 在收敛圆的边界上，无法作出幂级数是否收敛的一般论断，例如幂级数 $\sum_{n=0}^{\infty}z^n$ 在点 $z=\pm1$ 处皆发散，但其和函数 $1/(1-z)$ 在 $z=1$ 时为奇异，在 $z=-1$ 时表现正常.

幂级数的运算

幂级数的四则运算与一般级数的四则运算无异.

定理 2　幂级数的微分与积分

幂级数在进行逐项微分与逐项积分后，其收敛圆不变. 因此，幂级数的和函数可以在收敛圆内逐项微分，也可以逐项积分. 幂级数的和函数在收敛圆内是无穷可微的.

这是柯西-阿达玛公式[253] 的直接推论.

解析函数

由收敛幂级数表示的复变函数称为**解析函数（analytic function）**. 它与用复可微性定义的**全纯函数（holomorphic function）**[354] 是等价的对象，尽管这个事实的证明并不平凡（需要用到柯西积分公式）. 如果限制自变量取实数，那么得到的是**实解析函数（real analytic function）**. 等价地，实解析函数是泰勒级数收敛到其自身的函数. 有如下定理：

定理 3

开区间 I 上的实函数 f 为解析函数，当且仅当对于 I 的任何闭子区间 K，

都有常数 M_K 使得

$$\max_{x\in K}|f^{(n)}(x)|\le n!M_K^n.$$

证明是直接的计算：如果 f 满足此条件，那么可以估算出其泰勒展开式的余项趋于零; 反过来，如果 f 由收敛幂级数表征，那么可以估计其逐项微分得到的幂级数的上界而得到 f 的高阶导数所满足的条件.

这个定理说明：实解析函数的高阶导数随着其阶数的提升不能增长得太快. 举例来说，函数 $f(x)=e^{-1/x^2}$ 在任何不包含原点的开区间上是解析函数; 在原点处它的各阶导数都是零，但是对于任何正数 M (不管有多么大)，都可以找到趋于零的序列 x_n，使得对于充分大的 n 有

$$|f^{(n)}(x_n)|>n!M^n.$$

因此 f 在任何包含原点的开区间上都不是解析函数. 用复变函数论的语言，$z=0$ 是函数 e^{-1/z^2} 的本性奇点.

由此可见，泰勒展开式收敛到其自身的函数是非常特殊的.

极坐标中的曲线方程

- 本词条处于草稿阶段.

预备知识 极坐标[119]

极坐标系中，我们可以用一元函数

$$r=f(\theta)\tag{1}$$

表示一条曲线. 简单的例子如阿基米德螺线[121] 或圆锥曲线[131]. 以下我们讨论如何从函数中计算曲线的一些特征.

计算某点切线的方向

预备知识 导数[158]

在 x-y 平面直角坐标系中，我们可以通过求导（$\mathrm{d}y/\mathrm{d}x$）计算曲线某点切线，那么式 1 表示的极坐标中如何求某点切线的方向呢？可以证明，切线与 $\hat{\boldsymbol{\theta}}$ 方向的夹角为

$$\alpha = \frac{1}{r}\frac{\mathrm{d}r}{\mathrm{d}\theta} = \frac{f'(\theta)}{f(\theta)} \tag{2}$$

或者说与 $\hat{\mathbf{r}}$ 方向的夹角为 $\pi/2 - \alpha$.

（推导未完成）

曲线长度

预备知识 定积分[198]

若用 $\theta \in [\theta_1, \theta_2]$，来表示曲线的一段，那么其长度为

$$l = \int_{\theta_1}^{\theta_2} \sqrt{f(\theta)^2 + f'(\theta)^2}\,\mathrm{d}\theta \tag{3}$$

（推导未完成）

曲线的长度

预备知识 定积分[198]

直角坐标系中，曲线通常用函数 $y(x)$ 描述．我们在曲线上某点 (x, y) 附近取一小段，根据勾股定理，它的长度为

$$\mathrm{d}l = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2} = \sqrt{1 + \dot{y}^2}\,\mathrm{d}x \tag{1}$$

两边积分得

$$l = \int_{x_1}^{x_2} \sqrt{1 + \dot{y}^2}\,\mathrm{d}x \tag{2}$$

这就是曲线在区间 $[x_1, x_2]$ 的长度.

（例题未完成）

函数的算符

预备知识 导数[158]

未完成：定积分算符

算符（也叫**算子**）可以理解为“函数的函数”，即一个函数经过算符作用可以得到另一个算符．例如将一个函数乘以一个数 λ 或另一个函数，又例如对一个函数求导（**微分算符**）等等．

例 1

令算符 $\hat{A}$ 为

$$\hat{A}=2\sin(x) \tag{1}$$

那么它作用在函数 $f(x)$ 上就是

$$\hat{A}f(x)=\sin(x)f(x) \tag{2}$$

令算入 $\hat{B}$ 为

$$\hat{B}=\frac{\mathrm{d}}{\mathrm{d}x}+1 \tag{3}$$

那么

$$\hat{A}f(x)=\left(\frac{\mathrm{d}}{\mathrm{d}x}+1\right)f(x)=\frac{\mathrm{d}f(x)}{\mathrm{d}x}+f(x) \tag{4}$$

令算符 $\hat{C}$ 为

$$\hat{C}=\hat{A}\hat{B}=\sin(x)\left(\frac{\mathrm{d}}{\mathrm{d}x}+1\right) \tag{5}$$

那么

$$\hat{C}f(x)=\hat{A}[\hat{B}f(x)]=\sin(x)\frac{\mathrm{d}f(x)}{\mathrm{d}x}+\sin(x)f(x) \tag{6}$$

令算符 $\hat{D}$ 为

$$\hat{D}=\hat{A}+\hat{B} \tag{7}$$

那么

$$\hat{D}f(x)=\hat{A}f(x)+\hat{B}f(x) \tag{8}$$

令算符 $\hat{E}$ 为

$$\hat{E} = \hat{B}^2 \tag{9}$$

相当于 $\hat{B}$ 对某函数作用两次，即

$$\begin{aligned} \hat{E}f(x) = \hat{B}[\hat{B}f(x)] &= \left(\frac{\mathrm{d}}{\mathrm{d}x} + 1\right)\left[\frac{\mathrm{d}f(x)}{\mathrm{d}x} + f(x)\right] \\ &= \frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} + 2\frac{\mathrm{d}f(x)}{\mathrm{d}x} + f(x) \end{aligned} \tag{10}$$

所以也可以将这个过程简写为

$$\hat{B}^2 = \left(\frac{\mathrm{d}}{\mathrm{d}x} + 1\right)^2 = \frac{\mathrm{d}^2}{\mathrm{d}x^2} + 2\frac{\mathrm{d}}{\mathrm{d}x} + 1 \tag{11}$$

线性

若一个算符 $\hat{A}$ 对两个不同的函数 f, g 和两个不同的常数满足

$$\hat{A}(f + g) \tag{12}$$

是**线性**的

常微分方程

- 本词条处于草稿阶段.

预备知识 简谐振子[913]

作为一个引入的例子，我们首先看“简谐振子[913]”中的式 1. 一般来说，含有函数 $y(x)$ 及其高阶导数 $y^{(n)}$，和自变量 x 的等式叫做**常微分方程**（简称微分方程[15]），即

$$f(y^{(N)}, y^{(N-1)}, \ldots, y, x) = 0 \tag{1}$$

上式中的最高阶导数为 N 阶，所以可以把上式叫做 N 阶微分方程. 注意方程中必须出现 $y^{(N)}$，剩下的 $y^{(N-1)}, \ldots, y, x$ 可以只出现部分或不出现. 所

[15]这里的“常”强调未知函数只有一个变量，用于区别多元微积分中的“偏微分方程”.

有能使微分方程成立的函数 $f(x)$ 都是方程的**解**，如果能找到含有参数的函数 $f(x, C_1, \ldots, C_N)$，使所有可能的解都可以通过给 C_i 赋值来表示，那么这就是函数的**通解**.

有一些微分方程的解法是显然的，例如描述自由落体运动[889] 的微分方程为 $\mathrm{d}^2y/\mathrm{d}t^2 = g$ （假设 y 轴竖直向下）. 要解这个方程，只需对等式两边进行两次不定积分即可得到通解为 $y = C_1 + C_2t + gt^2/2$. 一般来说，如果 N 阶微分方程具有 $y^{(N)} = f(x)$ 的形式，只需进行 N 次积分即可得到通解.

另一些方程是**可以分离变量**的，我们来看“受阻落体[915]”这个例子. 若方程可分离变量，只需先分离变量，再对等式两边求不定积分即可找到通解.

一阶线性微分方程

一阶线性微分方程[212].

二阶线性微分方程

未完成：链接一下 Wolfram Alpha 和 Mathematica 解常微分方程

一阶线性微分方程

- 本词条需要更多参考文献.
- 需要补充例子

预备知识 常微分方程[211]

[16]具有以下形式的微分方程叫做**一阶线性微分方程（first order linear differential equation）**

$$\frac{\mathrm{d}y}{\mathrm{d}x} + p(x)y = f(x) \tag{1}$$

其中函数 $p(x)$ 和 $f(x)$ 是已知的，函数 $y(x)$ 是未知的，称为方程的解. 这三个函数的自变量都是实数，函数值可以是实数或者复数. 若 $f(x) \equiv 0$，则称方程为**齐次的（homogeneous）**，否则就是**非齐次的（inhomogeneous）**. 所以 $f(x)$ 叫做方程的**非齐次项（inhomogeneous term）**.

[16]本文参考 Wikipedia 相关页面和 [2]

线性

若干函数 $y_1(x), y_2(x), \dots$ 的线性组合[17]定义为他们分别乘以一个常数再相加

$$c_1 y_1(x) + c_2 y_2(x) + \dots \tag{2}$$

由于求导操作是线性（式 6 [160]）的，容易证明若 $y_i(x)$（$i = 1, 2, \dots$）都是其次方程的解，那么它们的线性组合也必定是齐次方程的解．微分方程的这种性质就叫做**线性（linear）**.

一阶常系数线性微分方程

当式 1 中的 $p(x) \equiv b$ 是常数时，我们说式 1 的方程是常系数的．先来看齐次的情况，方程为

$$\frac{\mathrm{d}y}{\mathrm{d}x} + by = 0 \tag{3}$$

也就是说我们想要求一个函数 $y(x)$，使得它的导数和 $y(x)$ 自己成正比．一个最显然的解就是 $y(x) \equiv 0$，但它太简单了没什么意义，我们说这个解是**平凡的（trivial）**．对于非零解，一个直接的猜测就是自然指数函数 $y(x) = \mathrm{e}^{ax}$ 其中 a 为常数．把它代入式 3 得

$$ay(x) + by(x) = 0 \tag{4}$$

所以消去 $y(x)$ 得 $b = -a$．再考虑到方程的线性，所以我们把这个解乘以任意常数 C 同样是方程的解，所以式 3 的通解为

$$y(x) = C\mathrm{e}^{-bx} \tag{5}$$

例 1　稀释酒精

一个容器中装有质量为 M（千克）的酒精，从某时刻开始以速率 I（千克每秒）加水，并以同样的速率从底部排水，保持容器中液体的总质量不变．假设容器中有某种快速搅拌装置，使两种液体时刻保持充分混合，求容器中酒精浓度 x（酒精质量比总质量）随时间的变化关系．

[17]线性组合是线性代数中的概念，线性代数中，线性空间都可以做线性组合，例如几何矢量（式 6 [375]）．特定函数的集合可以看作线性空间（例 2 [444]）.

解：令容器中酒精的质量为 m_1，水的质量为 $M-m_1$．流出容器的液体速率可以分为两部分：流出酒精的速率为 xI，流出水的速率为 $I-xI$．所以容器中酒精质量的变化率为

$$m_1' = -xI \tag{6}$$

酒精浓度定义为 $x = m_1/M$，M 是常数，所以

$$x' = \left(\frac{m_1}{M}\right)' = \frac{m_1'}{M} = -\frac{I}{M}x \tag{7}$$

该微分方程符合式 3 的结构，可以直接写出方程的解为

$$x(t) = C\mathrm{e}^{-It/M} \tag{8}$$

但这里的 C 到底是什么呢？如果我们令开始的时间为 $t=0$，那么根据题目条件必须满足 $x(0)=1$，代入得 $C=1$．

未完成：非齐次

一般情况

我们先来看式 1 对应的齐次方程

$$\frac{\mathrm{d}y}{\mathrm{d}x} + p(x)y = 0 \tag{9}$$

这是一个可分离变量的方程，分离变量得

$$\frac{\mathrm{d}y}{y} = -p(x)\,\mathrm{d}x \tag{10}$$

两边积分得

$$\ln|y| = -\int p(x)dx + C \tag{11}$$

两边取自然指数得

$$y = \pm\mathrm{e}^{C}\mathrm{e}^{-\int p(x)\mathrm{d}x} \tag{12}$$

把 $\pm\mathrm{e}^C$ 整体看做一个任意常数 C，上式变为．

$$y = C\mathrm{e}^{-\int p(x)\mathrm{d}x} \tag{13}$$

这就是一阶线性齐次微分方程式 9 的通解，也叫式 1 的**齐次解**．

常数变易法

现在我们用**常数变易法**来解非齐次方程式 1．为书写方便，式 13 中令 $y_0(x)=\exp\big(-\int p(x)\,\mathrm{d}x\big)$．假设上式中的 C 是一个函数 $C(x)$ 而不是常数，代入式 1 得

$$C'y_0+C[y_0'+p(x)y_0]=f(x) \tag{14}$$

由于 y_0 是齐次解，上式方括号中求和为 0，分离变量得

$$\mathrm{d}C=\frac{f(x)}{y_0}\,\mathrm{d}x \tag{15}$$

两边积分得

$$C(x)=\int\frac{f(x)}{y_0}\,\mathrm{d}x \tag{16}$$

所以一阶线性非齐次微分方程的通解为

$$y=y_0\int\frac{f(x)}{y_0}\,\mathrm{d}x \tag{17}$$

其中

$$y_0(x)=\mathrm{e}^{-\int p(x)\mathrm{d}x} \tag{18}$$

注意待定常数包含在式 17 的不定积分中，式 18 中的不定积分产生的待定常数在代入式 17 后可消去．

二阶常系数齐次微分方程

预备知识 常微分方程[211]，指数函数（复数）[351]

- 需要补充例题

[18]二阶常系数齐次微分方程形式如下

$$ay''+by'+cy=0 \tag{1}$$

[18]参考 [2]

注意到指数函数 $y = C\mathrm{e}^{rx}$ 第 n 阶导数为 $r^n\mathrm{e}^{rx}$，不妨尝试把指数函数代入方程，得

$$(ar^2 + br + c)\mathrm{e}^{rx} = 0 \tag{2}$$

由于 $\mathrm{e}^{rx} \neq 0$，必有 $ar^2 + br + c = 0$．把这个二次函数叫做**特征方程**，解特征方程，就可以得到方程的解．根据根的分布，有如下四种情况

1. 有两个不同的实根 r_1 和 r_2（$b^2 - 4ac > 0$），方程的通解为

$$y = C_1\mathrm{e}^{r_1x} + C_2\mathrm{e}^{r_2x} \tag{3}$$

2. 有一个重根 r （$b^2 - 4ac = 0$），方程的通解为

$$y = C_1\mathrm{e}^{rx} + C_2x\mathrm{e}^{rx} \tag{4}$$

3. 有两个纯虚数根 $\pm\mathrm{i}\omega_0$（$b = 0,\ b^2 - 4ac < 0$），方程的通解为

$$y = C_1\cos(\omega_0x) + C_2\sin(\omega_0x) \tag{5}$$

或

$$y = C_1\cos(\omega_0x + C_2) \tag{6}$$

其中 $\omega_0 = \sqrt{c/a}$.

4. 有两个复数根 $r \pm \mathrm{i}\omega$ （$b \neq 0,\ b^2 - 4ac < 0$），方程的通解为

$$y = \mathrm{e}^{rx}[C_1\cos(\omega x) + C_2\sin(\omega x)] \tag{7}$$

或

$$y = C_1\mathrm{e}^{rx}\cos(\omega x + C_2) \tag{8}$$

其中

$$r = -\frac{b}{2a} \qquad \omega = \frac{1}{2a}\sqrt{4ac - b^2} \tag{9}$$

详细推导

情况 1 的结论是显然的，我们先来看情况 3．根据 $y = Ce^{rx}$ 的假设，通解应该是

$$y = C_1 e^{i\omega x} + C_2 e^{-i\omega x} \tag{10}$$

如果这里的 C_1 和 C_2 取任意复数，那么上式就是方程在复数域的通解，其中包含了实数域的通解．这个通解还有另一种等效的形式，令

$$C_1 = \frac{C_3}{2} + \frac{C_4}{2i} \qquad C_2 = \frac{C_3}{2} - \frac{C_4}{2i} \tag{11}$$

代入上式得

$$\begin{aligned} y &= C_3 \frac{e^{i\omega x} + e^{-i\omega x}}{2} + C_4 \frac{e^{i\omega x} - e^{-i\omega x}}{2i} \\ &= C_3 \cos(\omega x) + C_4 \sin(\omega x) \end{aligned} \tag{12}$$

注意如果 C_3, C_4 取任意复数，该式仍然是复数域的通解（因为任何 C_1, C_2 都可以找到对应的 C_3, C_4），但只要把 C_3, C_4 限制在实数域中，该式就是实数域的通解．

情况 4 的结论可以类比情况 3 得出，最后我们来看情况 2．我们可以把情况 2 看做情况 4 的一个极限，即 $\omega \to 0$ 时的情况．如果式 7 中的 C_1, C_2 都是普通常数，则取该极限时可以得到式 4 的第一项 $C_1 e^{rx}$．那如何得到第二项呢？我们不妨令式 7 中的 $C_1 = 0$，$C_2 = C_3/\omega$，再来取极限，得

$$\lim_{\omega \to 0} C_3 e^{rx} \frac{\sin(\omega x)}{\omega x} x = C_3 x e^{rx} \tag{13}$$

这里用到了“小角正弦值极限[152]”中的结论．

一维齐次亥姆霍兹方程

- 本词条需要更多参考文献.
- 参考 docx 版百科

预备知识 二阶常系数微分方程[215]

一维齐次亥姆霍兹方程可以记为

$$\frac{\mathrm{d}^2y}{\mathrm{d}t^2}+\omega^2y=0 \tag{1}$$

这里 ω 为实数.

通解

这个方程属于二阶常系数线性齐次方程，可以假设 e^{rt} 为方程的解，代入原方程得特征方程

$$r^2+\omega^2=0 \tag{2}$$

解得 $r=\pm\mathrm{i}\omega$，即方程在复数域的通解为

$$y=C_1\mathrm{e}^{\mathrm{i}\omega t}+C_2\mathrm{e}^{-\mathrm{i}\omega t} \tag{3}$$

其中 C_1,C_2 是复常数.

如果选择恰当的 C_1 和 C_2，可以使通解变为实数函数．令

$$C_1=C_{1R}+\mathrm{i}C_{1I}\qquad C_2=C_{2R}+\mathrm{i}C_{2I} \tag{4}$$

把 $y(t)$ 分解为实部和虚部，令虚部为零，可得所有可能的实数解

$$\begin{aligned}y(t)=&[(C_{1R}+C_{2R})\cos\omega t+(C_{2I}-C_{1I})\sin\omega t]\\&+\mathrm{i}[(C_{1R}-C_{2R})\sin\omega t+(C_{1I}+C_{2I})\cos\omega t]\end{aligned} \tag{5}$$

令虚部为 0，则 $C_{1R}=C_{2R}$，$C_{1I}=-C_{2I}$.

二阶常系数非齐次微分方程

预备知识 二阶常系数齐次微分方程[215]

结论

在二阶常系数齐次微分方程的右端加上一个函数 $f(x)$，就得到了二阶常系数非齐次微分方程

$$y'' + by' + cy = f(x) \tag{1}$$

这就是**二阶常系数非齐次微分方程**．其解为

$$y(x) = C_1 y_1 + C_2 y_2 - y_1 \int \frac{y_2 f}{W}\,\mathrm{d}x + y_2 \int \frac{y_1 f}{W}\,\mathrm{d}x \tag{2}$$

其中 W 可以写成二阶行列式

$$W = \begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix} = y_1 y_2' - y_1' y_2 \tag{3}$$

其中 y_1, y_2, W, f 都是 x 的函数，后面的括号和自变量被省略．$y_1(x)$ 和 $y_2(x)$ 是对应齐次方程

$$y'' + by' + cy = 0 \tag{4}$$

的两个线性无关的解．

应用

推导

下面介绍的方法叫常数变易法，其主要思想可参考一阶线性非齐次微分方程的通解

设通解的形式为

$$y = v_1 y_1 + v_2 y_2 \tag{5}$$

其中，v_i 也是关于 x 的函数．对该式两边求导，得

$$y' = v_1' y_1 + v_2' y_2 + v_1 y_1' + v_2 y_2' \tag{6}$$

为了接下来计算方便，我们规定 v_1，v_2 满足关系[19]

$$v_1' y_1 + v_2' y_2 = 0 \tag{7}$$

[19]这么规定会不会丢失一部分解呢？或许会，但是由于我们已经有了式 1 对应的齐次解 y_1 和 y_2，根据线性微分方程解的结构（见同济大学的《高等数学》），只需要找到式 1 的任意一个解，就可以找到它的通解．

把式 7 代入式 6，得到

$$y' = v_1 y_1' + v_2 y_2' \tag{8}$$

继续对求导，得到

$$y'' = v_1' y_1' + v_2' y_2' + v_1 y_1'' + v_2 y_2'' \tag{9}$$

把式 5 式 8 式 9 代回原方程式 1 得

$$(v_1' y_1' + v_2' y_2' + v_1 y_1'' + v_2 y_2'') + b(v_1 y_1' + v_2 y_2') + c(v_1 y_1 + v_2 y_2) = f \tag{10}$$

化简，得

$$(v_1' y_1' + v_2' y_2') + v_1(a y_1'' + b y_1' + c y_1) + v_2(a y_2'' + b y_2' + c y_2) = f \tag{11}$$

由于 y_1 和 y_2 都是式 4 的解，式（9）化为

$$v_1' y_1' + v_2' y_2' = f \tag{12}$$

总结一下，刚刚的推导说明，和在（5）的假设条件下，只要满足（10）即可满足（1）式．联立（5）和（10）式，得到关于 v_1' 和 v_2' 的方程组

$$\begin{cases} y_1 v_1' + y_2 v_2' = 0 \\ y_1' v_1' + y_2' v_2' = f \end{cases} \tag{13}$$

解得

$$\begin{cases} v_1' = -y_2 f / W \\ v_2' = y_1 f / W \end{cases} \tag{14}$$

其中

$$W = y_1 y_2' - y_2 y_1' = \begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix} \tag{15}$$

对（13）的两条式子积分，即可得到

$$v_1 = -\int \frac{y_2 f}{W}\,\mathrm{d}x + C_1 \tag{16}$$

$$v_2 = \int \frac{y_1 f}{W}\,\mathrm{d}x + C_2 \tag{17}$$

（15）（16）代入（5）式，得到方程（1）的解为

$$y(x) = C_1 y_1 + C_2 y_2 - y_1 \int \frac{y_2 f}{W}\,\mathrm{d}x + y_2 \int \frac{y_1 f}{W}\,\mathrm{d}x \tag{18}$$

由于上式满足线性微分方程解的结构，所这已经是通解了．但是必须注意，根据常数变易法，我们只能在没有零点的区间内找到方程式 1 的通解．

正交函数系

预备知识 定积分[198]

函数值为实数

定义给出一组函数（有限或无限多个），$f_i(x)$ $(i=1,2\dots)$，如果满足

$$\int_a^b f_i(x)f_i(x)\,\mathrm{d}x \neq 0 \tag{1}$$

当整数 $m \neq n$ 时

$$\int_a^b f_m(x)f_n(x)\,\mathrm{d}x = 0 \tag{2}$$

那么这一组函数就是区间 $[a,b]$ 内的一个正交函数系.

这一组函数的性质可以类比矢量的正交，“两个函数相乘再积分”这个步骤可以类比矢量的内积. 如果两个不同的矢量正交（垂直），则它们的内积为零. 如果它们的模长不为零，则一个矢量与自身内积不为零.

特殊地，若给正交函数系中的每个函数的平方进行归一化，使得

$$\int_a^b f_i(x)f_i(x)\,\mathrm{d}x = 1 \tag{3}$$

那么该正交函数系就是**归一**的. 其性质可以表示为

$$\int_a^b f_m(x)f_n(x)\,\mathrm{d}x = \delta_{mn} \tag{4}$$

其中 δ_{mn} 是克罗内克 δ 函数（Kronecker Delta Function）.

函数值为复数

若函数系中 $f_i(x)$ 的自变量为实数，函数值为复数，则正交的定义变为

$$\int_a^b f_i^*(x)f_j(x)\,\mathrm{d}x \neq 0 \qquad (i \neq j) \tag{5}$$

其中星号表示复共轭．归一化的定义变为

$$\int_a^b f_i^*(x) f_i(x)\,\mathrm{d}x = 1 \tag{6}$$

正交归一条件可以统一写成

$$\int_a^b f_i^*(x) f_j(x)\,\mathrm{d}x = \delta_{ij} \tag{7}$$

傅里叶级数（三角）

预备知识 定积分[198]，狄拉克符号[470]

满足**狄利克雷条件**[20]的周期函数 $f(x)$（周期为 $2l$）可以使用以下三角函数展开

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi}{l}x\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi}{l}x\right) \tag{1}$$

其中

$$a_n = \frac{1}{l}\int_{-l}^{l} f(x)\cos\left(\frac{n\pi}{l}x\right)\mathrm{d}x \tag{2}$$

$$b_n = \frac{1}{l}\int_{-l}^{l} f(x)\sin\left(\frac{n\pi}{l}x\right)\mathrm{d}x \tag{3}$$

说明

注意式 1 的所有 a_n 项为偶函数项，所有 b_n 项为奇函数项．若 $f(x)$ 是偶函数，所有 b_n 项为零，若是奇函数，则所有 a_n 项为零[21]．如果 $f(x)$ 不具有奇偶性，可以表示为偶函数和奇函数之和，分别对应所有 a_n 项和所有 b_n 项

$$f(x) = \frac{1}{2}[f(x) + f(-x)] + \frac{1}{2}[f(x) - f(-x)] \tag{4}$$

$f(x)$ 的函数值既可以是实数也可以是复数．实函数的展开系数 a_n, b_n 也必须取实数，复函数的展开系一定不全是实数．

[20] 函数值有限，存在有限个间断点和有限个极值点．如无特殊说明，本书中做傅里叶级数的展开的函数都满足狄利克雷条件

[21] 证明：如果 $f(x)$ 是偶函数，那么式 3 中的被积函数就是奇函数，所以在区间 $[-l, l]$ 的积分为零．奇函数的证明类似．

完备性

傅里叶级数最奇妙的地方大概就是它能展开任意满足狄利克雷条件的函数，这个性质叫做**完备性**. 我们先不证明完备性，只给出几个例子说明随着求和项数增加，三角函数如何逼近不连续或不光滑的函数.

例 1　方波

首先定义一个方波为

$$f(x)=\begin{cases}1 & (2k\pi<x\leqslant(2k+1)\pi)\\ -1 & ((2k-1)\pi<x\leqslant 2k\pi)\end{cases} \tag{5}$$

注意 $f(x)$ 在 $x=k\pi$ 处存在间断点. 进行傅里叶级数展开得

$$f(x)=\sum_{k=0}^{\infty}\frac{4}{\pi(2k+1)}\sin(2k+1)x \tag{6}$$

注意由于 $f(x)$ 是奇函数，求和只有正弦项. 取级数的前 m 项求和并画图如图 1.

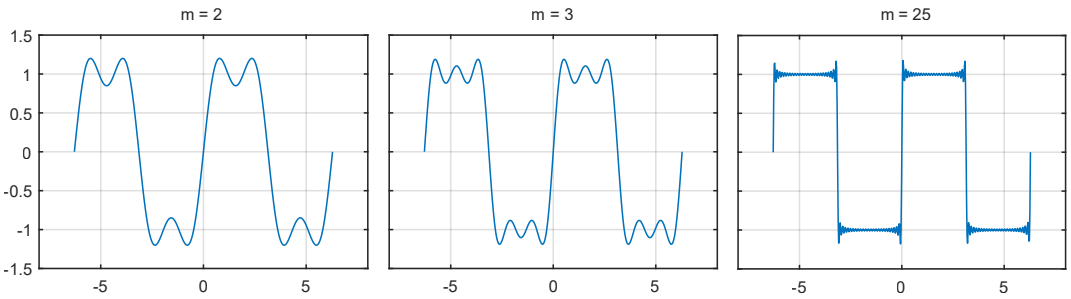


图 1: 有限项傅里叶级数逼近方波

例 2　正弦函数的绝对值

偶函数 $f(x)=|\sin x|$ 存在不光滑的点，展开成傅里叶级数为

$$f(x)=|\sin x|=\frac{2}{\pi}-\frac{4}{\pi}\sum_{k=1}^{+\infty}\frac{1}{4k^2-1}\cos 2kx \tag{7}$$

前 m 项和画图如图 2.

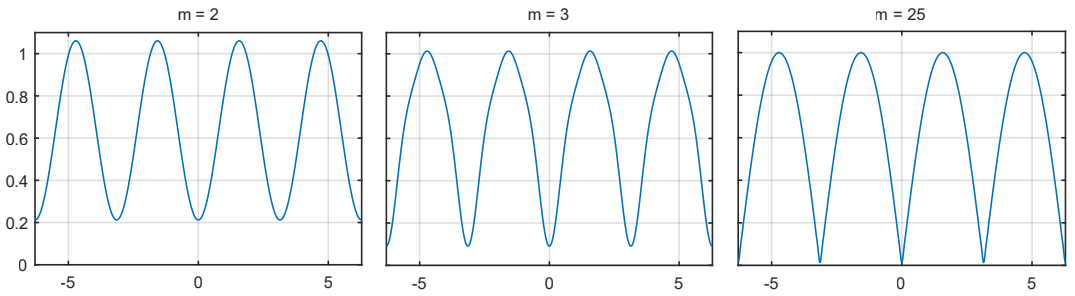


图 2: 有限项傅里叶级数逼近 $|\sin(x)|$

习题 1

根据傅里叶级数的定义用定积分计算式 6 和式 7.

系数公式推导

预备知识 正交归一基底[386]

在泰勒级数[181] 中，我们通过求 n 阶导数的方式来“过滤”出第 n 阶系数. 这里我们用积分的方式来“过滤”系数. 我们在推导中还要介绍一个重要的类比来讲解，即把级数中的 sin 和 cos 的性质对比几何矢量[370] 在正交（不归一）基底上的展开（式 2 [386]）.

给出一组无穷多个函数

$$\begin{aligned}&\frac{1}{2},\ \sin\left(\frac{\pi}{l}x\right),\ \cos\left(\frac{\pi}{l}x\right),\ \sin\left(\frac{2\pi}{l}x\right),\ \cos\left(\frac{2\pi}{l}x\right),\ \dots\\&\sin\left(\frac{n\pi}{l}x\right),\ \cos\left(\frac{n\pi}{l}x\right)\ \dots\end{aligned}\tag{8}$$

其中 n 是正整数. 我们把任意满足狄利克雷条件的函数（以下简称任意函数）比作几何矢量，把上面这组函数（式 8）比作矢量基底（称为**函数基底**），任意函数都可以表示成这组基底的线性组合（即式 1）. 现在把两个任意函数（矢量）$f(x)$ 和 $g(x)$ 的**内积**定义为它们的乘积在 $[-l,l]$ 内积分

$$\langle f|g\rangle=\int_{-l}^{l}f(x)g(x)\,\mathrm{d}x\tag{9}$$

可以证明这组基底**正交**（即任意两个不同的基底内积为 0，证明见下文）但不**归一**[22]. 与“几何矢量在正交但不归一的基底上展开”一样，我们只要把函数

[22] 归一化的矢量也叫单位矢量，与自身内积等于 1，详见“矢量内积”[383]

分别与各个基底内积，再除以基底的模长平方（模方）即可获得线性组合的系数 a_n 和 b_n．可以证明所有基底的模方为 l，这样我们就得到了系数公式（式 2 式 3）.

正弦基底

若我们只需要在一个区间 $[0,l]$ 上展开函数 $f(x)$ 而不在意其他地方，那么我们可以假想 $f(x)$ 是以 $2l$ 为周期的奇函数，这样，我们只需要用正弦基底展开 $f(x)$ 即可[23]．于是我们可以说，式 8 中给出的所有正弦基底在区间 $[0,l]$ 具有完备性.

例 3　用正弦基底展开闭区间内的函数

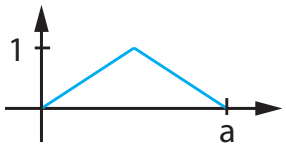


图 3: 三角形函数

定义区间 $[0,a]$ 内的一个三角形函数如下（图 3），先把函数归一化，再用正弦基底做傅里叶展开.

$$f(x)=\begin{cases}\dfrac{2}{a}x & \left(0\leqslant x\leqslant\dfrac{a}{2}\right)\\ 2-\dfrac{2}{a}x & \left(\dfrac{a}{2}<x\leqslant a\right)\end{cases}\tag{10}$$

首先假想 $f(x)$ 为奇函数，原则上可以直接使用式 3 计算展开系数，但由于被积函数 $f(x)\sin(x)$ 为偶函数，可以先把式 3 化简为

$$b_n=\frac{2}{a}\int_0^a f(x)\sin\left(\frac{n\pi}{a}x\right)\mathrm{d}x\tag{11}$$

又注意 $f(x)$ 关于区间中点的对称性，我们可以进一步判断出 n 为偶数时 b_n 项都为零，而 n 为奇数项的 b_n 在 $[0,a/2]$ 的积分等于在 $[a/2,a]$ 的积分，所

[23] 显然我们也可以假想 $f(x)$ 是以 $2l$ 为周期的偶函数，用余弦基底展开 $f(x)$．选择正弦或余弦还应该考虑到函数在该区间的边界条件：正弦可以保证两端函数值为 0，余弦可以保证区间的导数值为 0.

以

$$\begin{aligned} b_n &= \frac{4}{a}\int_0^{a/2} f(x)\sin\Big(\frac{n\pi}{a}x\Big)\,\mathrm{d}x \\ &= \frac{8}{a^2}\int_0^{a/2} x\sin\Big(\frac{n\pi}{a}x\Big)\,\mathrm{d}x \end{aligned} \tag{12}$$

这个积分既可以见例 *，也可以用 Wolfram Alpha 或 Mathematica 完成．结果是

$$b_n = (-1)^{\frac{n-1}{2}}\frac{8}{\pi^2 n^2} \quad (n \text{ 为奇数}) \tag{13}$$

进一步说，我们可以用正弦基底在任意区间 $[x_0, x_0+l]$ 上展开任意函数．要这样做，我们只需要把所有正弦基底平移 x_0 即可．

$$\sin\frac{\pi}{l}(x-x_0),\ \sin\frac{2\pi}{l}(x-x_0),\ \dots\ \sin\frac{n\pi}{l}(x-x_0)\dots \tag{14}$$

证明函数基底正交

现在证明式 8 中任意两个不同的基底内积等于 0．当 $m \neq n$ 时，有

$$\begin{aligned} &\int_{-\pi}^{\pi}\sin(mx)\sin(nx)\,\mathrm{d}x \\ =&\frac{1}{2}\int_{-\pi}^{\pi}\cos[(m-n)x]\,\mathrm{d}x - \frac{1}{2}\int_{-\pi}^{\pi}\cos[(m+n)x]\,\mathrm{d}x = 0 \end{aligned} \tag{15}$$

$$\begin{aligned} &\int_{-\pi}^{\pi}\cos(mx)\cos(nx)\,\mathrm{d}x \\ =&-\frac{1}{2}\int_{-\pi}^{\pi}\cos[(m-n)x]\,\mathrm{d}x + \frac{1}{2}\int_{-\pi}^{\pi}\cos[(m+n)x]\,\mathrm{d}x = 0 \end{aligned} \tag{16}$$

$$\begin{aligned} &\int_{-\pi}^{\pi}\sin(mx)\cos(nx)\,\mathrm{d}x \\ =&\frac{1}{2}\int_{-\pi}^{\pi}\sin[(m+n)x]\,\mathrm{d}x - \frac{1}{2}\int_{-\pi}^{\pi}\sin[(m-n)x]\,\mathrm{d}x = 0 \end{aligned} \tag{17}$$

$$\int_{-\pi}^{\pi} 1\cdot\sin x\,\mathrm{d}x = 0 \tag{18}$$

$$\int_{-\pi}^{\pi} 1\cdot\cos x\,\mathrm{d}x = 0 \tag{19}$$

其中任意一个函数与自己内积都等于 π（除了常函数 1 积分为 2π）.

$$\int_{-\pi}^{\pi} \sin^2(nx)\,\mathrm{d}x = \pi \tag{20}$$

$$\int_{-\pi}^{\pi} \cos^2(nx)\,\mathrm{d}x = \pi \tag{21}$$

$$\int_{-\pi}^{\pi} 1^2\,\mathrm{d}x = 2\pi \tag{22}$$

使用归一化的基底

我们也可以将式 8 的每个基底乘以一个常数进行归一化，即让每个基底的模长为1.我们将归一化后的正弦项和余弦项分别记为 $|S_n\rangle$,$|C_n\rangle$（$n = 1, 2, \dots$），将常数项记为 $|C_0\rangle$.

$$\langle S_n|S_n\rangle = \int_{-l}^{l} \left[B_n \sin\left(\frac{n\pi}{l}x\right)\right]^2 \mathrm{d}x = 1 \tag{23}$$

$$\langle C_n|C_n\rangle = \int_{-l}^{l} \left[A_n \cos\left(\frac{n\pi}{l}x\right)\right]^2 \mathrm{d}x = 1 \tag{24}$$

解得 $A_n = B_n = 1/\sqrt{l}$．最后将常数项归一化变为 $1/\sqrt{2l}$ 即可．所以正交归一得傅里叶级数基底为

$$\frac{1}{\sqrt{2l}},\ \frac{1}{\sqrt{l}}\sin\left(\frac{\pi}{l}x\right),\ \frac{1}{\sqrt{l}}\cos\left(\frac{\pi}{l}x\right),\ \dots\ \frac{1}{\sqrt{l}}\sin\left(\frac{n\pi}{l}x\right),\ \frac{1}{\sqrt{l}}\cos\left(\frac{n\pi}{l}x\right)\ \dots \tag{25}$$

易证将基底乘以常数不影响它们的正交性，我们可以将正交归一基底的性质用克罗内克 δ 函数（式 6 [386]）表示（常数基底可以记为 $|C_0\rangle$）

$$\langle S_m|S_n\rangle = \delta_{m,n} \qquad \langle C_m|C_n\rangle = \delta_{m,n} \qquad \langle S_m|C_n\rangle = 0 \tag{26}$$

如果将任意函数 $f(x)$ 记为 $|f\rangle$，傅里叶级数的公式可以记为（下面默认 $|C_n\rangle$ 求和时 $n = 0, 1, \dots$，对 $|S_n\rangle$ 求和时 $n = 1, 2, \dots$）

$$|f\rangle = \sum_n \left(a_n |C_n\rangle + b_n |S_n\rangle\right) \tag{27}$$

$$a_n = \langle f|C_n\rangle \qquad b_n = \langle f|S_n\rangle \tag{28}$$

这比上文的定义要更简洁．我们还可以证明两函数 $f(x)$ 和 $g(x)$ 的内积为

$$\begin{aligned}\langle f|g\rangle &= \sum_m \left(a_m \langle C_m| + b_m \langle S_m|\right) \sum_n \left(c_n |C_n\rangle + d_n |S_n\rangle\right) \\ &= \sum_{m,n} a_m c_n \delta_{m,n} + b_m d_n \delta_{m,n} \\ &= \sum_n (a_n c_n + b_n d_n)\end{aligned} \tag{29}$$

该式可以类比几何矢量的坐标计算（式 9 [383]），即把两个函数在所有基底上的展开系数分别相乘再相加．

特殊地，一个任意函数的模方为

$$\langle f|f\rangle = \sum_n (a_n^2 + b_n^2) \tag{30}$$

即将所有展开系数分别平方再相加．这与几何矢量模方公式也相同．

傅里叶级数（指数）

预备知识 傅里叶级数（三角）[222]，欧拉公式[351]

$f(x)$ 是自变量为实数的复变函数，若满足狄利克雷条件，则可展在区间 $[-l,l]$ 展开成复数的傅里叶级数

$$f(x) = \sum_{n=-\infty}^{+\infty} c_n \exp\left(\mathrm{i}\frac{n\pi}{l}x\right) \tag{1}$$

其中

$$c_n = \frac{1}{2l}\int_{-l}^{l} f(x)\exp\left(-\mathrm{i}\frac{n\pi}{l}x\right)\mathrm{d}x \tag{2}$$

当 $f(x)$ 为实函数时，c_n 与 c_{-n} 互为复共轭．当 $f(x)$ 为偶函数或奇函数时，分别有 $c_{-n} = c_n$ 或 $c_{-n} = -c_n$．

推导

类比三角傅里叶级数[222] 的情况．这时，完备正交归一的函数基底变为

$$f_n(x) = \exp\left(\mathrm{i}\frac{n\pi}{l}x\right) \quad n \in N \tag{3}$$

定义复函数 $f(x)$ 与 $g(x)$ 的内积为

$$\langle f|g\rangle = \int_{-l}^{l} f(x)^* g(x)\,\mathrm{d}x \tag{4}$$

可证明函数基底（式 3 ）正交且模长为 $2l$，用克罗内克 δ 函数表示为

$$\langle f_m|f_n\rangle = 2l\delta_{mn} \tag{5}$$

与三角傅里叶级数同理，可得式 1 和式 2.

与三角傅里叶级数的关系

考虑到正余弦函数和复指数函数的关系

$$\cos x = \frac{\mathrm{e}^{\mathrm{i}x} + \mathrm{e}^{-\mathrm{i}x}}{2} \qquad \sin x = \frac{\mathrm{e}^{\mathrm{i}x} - \mathrm{e}^{-\mathrm{i}x}}{2\mathrm{i}} \tag{6}$$

三角傅里叶级数的系数式 2 [222] 和式 3 [222] 可以用指数傅里叶级数的系数表示

$$\begin{aligned} a_n &= \frac{1}{l}\int_{-l}^{l} f(x)\cos\left(\frac{n\pi}{l}x\right)\mathrm{d}x \\ &= \frac{1}{2l}\int_{-l}^{l} f(x)\exp\left(\mathrm{i}\frac{n\pi}{l}x\right)\mathrm{d}x + \frac{1}{2l}\int_{-l}^{l} f(x)\exp\left(-\mathrm{i}\frac{n\pi}{l}x\right)\mathrm{d}x \\ &= c_{-n} + c_n \end{aligned} \tag{7}$$

同理，

$$b_n = \frac{c_{-n} - c_n}{\mathrm{i}} \tag{8}$$

注意这里全都有 $n \geqslant 0$. 由以上两式，也可以解得

$$c_n = \frac{a_n - \mathrm{i}b_n}{2} \qquad c_{-n} = \frac{a_n + \mathrm{i}b_n}{2} \tag{9}$$

实函数，奇函数，和偶函数的情况

特殊地，当 $f(x)$ 为实函数时，由于 a_n 和 b_n 必定是实数，根据式 9 可知

$$c_{-n} = c_n^* \tag{10}$$

即正负系数互为复共轭. 当 $f(x)$ 为偶函数或奇函数时，三角傅里叶级数分别只有 a_n 或 b_n 不为零[222]，同样根据式 9 可得，两种情况分别对应

$$c_{-n} = c_n = \frac{a_n}{2} \qquad c_{-n} = -c_n = \mathrm{i}\frac{b_n}{2} \tag{11}$$

由以上两式可得，如果 $f(x)$ 既是实函数又是偶函数时，c_n 和 c_{-n} 是相等的实数，如果既是实函数又是奇函数，c_n 和 c_{-n} 是相反的纯虚数.

性质

$$\int_{-l}^{l} \left|f^2(x)\right| \mathrm{d}x = 2l \sum_n |c_n|^2 \tag{12}$$

证明：用狄拉克符号记为 $|f\rangle = \sum_n c_n |n\rangle$，利用基底的正交性（式 5）

$$\langle f|f\rangle = \sum_i c_i^* \langle i| \sum_j c_j |j\rangle = 2l \sum_{i,j} |c_i|^2 \delta_{i,j} = 2l \sum_n |c_n|^2 \tag{13}$$

如果把基底为正交归一化（每个基底 $|n\rangle$ 除以 $\sqrt{2l}$，使得 $\langle i|j\rangle = \delta_{i,j}$，则有更简洁的关系

$$\int_{-l}^{l} \left|f^2(x)\right| \mathrm{d}x = \sum_n |c_n|^2 \tag{14}$$

狄拉克 delta 函数

预备知识 定积分[198]，sinc 函数[116]

[24]在物理中我们经常会遇到一些模型，如质点和点电荷等，这类模型使用了极限的思想（如令体积趋于无穷小）．如果考察质点的密度或点电荷的电荷密度，将得到无穷大，然而将其密度（电荷密度）在空间中积分却又能得到有限的质量与电荷．为了描述这样的密度（电荷密度）分布，我们引入**狄拉克 δ 函数（Dirac delta function）**.

$\delta(x)$ 并不是数学中一个严格意义上的函数，而是在泛函分析中被称为**广义函数（generalized function）**或**分布（distribution）**，详见泛函分析教材如 [9].

我们来考虑一个函数（图 1 左）

$$f_h(x) = \begin{cases} h & \left(|x - x_0| \leqslant \frac{1}{2h}\right) \\ 0 & \left(|x - x_0| > \frac{1}{2h}\right) \end{cases} \tag{1}$$

其中 h, x_0 是常数．由函数图像易得函数曲线下面的面积为 $\int_{-\infty}^{+\infty} f_h(x)\,\mathrm{d}x = 1$．现在我们令 $h \to \infty$，长方形的高将趋于无穷大，宽将趋于零，而定积分结果不变．

[24]参考 [7] 相关内容.

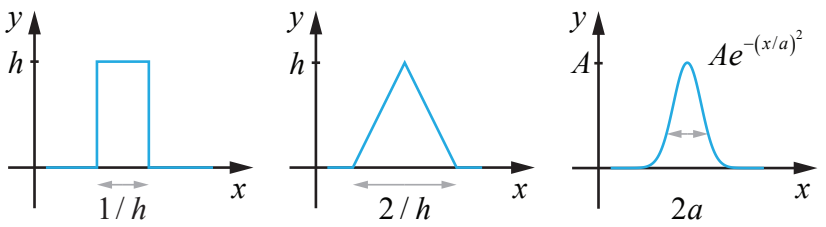


图 1: $\delta(x-x_0)$ 函数列的简单例子

这样，上面的 $f_h(x)$ 就可以表示为 $\delta(x-x_0)$．当然，我们还可以选取其他含有参数的 $f(x)$ 来逼近 δ 函数，如图 1 中的另外两种情况．

用函数列严格定义

本文中我们给出另一个严谨的定义：把 δ 函数看作是满足一定条件的**函数序列**，即无穷个函数按一定顺序排列．

定义 1　狄拉克 δ 函数

令 $\delta_1(x),\delta_2(x),\dots$ 为一个连续实函数的序列．若 $\delta_n(x)$ 满足以下两个条件，那么我们把该函数列称为狄拉克 δ 函数（列）：

$$\lim_{n\to\infty}\int_{-\infty}^{+\infty}\delta_n(x)\,\mathrm{d}x=1 \tag{2}$$

对任意给定的不包含 0 的区间 (a,b)（$a,b\neq 0$，可取 $\pm\infty$），有

$$\lim_{n\to\infty}\int_{a}^{b}\delta_n(x)\,\mathrm{d}x=0 \tag{3}$$

事实上，教材上（如 [7]）常使用以下定义，但笔者认为这两种定义是等价的[25]，且前者更容易理解

定义 2　狄拉克 δ 函数 2

令 $\delta_1(x),\delta_2(x),\dots$ 为一个连续实函数的序列．若 $\delta_n(x)$ 满足以下两个条件，那么我们把该函数列称为狄拉克 δ 函数（列）：

对所有性质良好（例如在 $x=0$ 连续）的 $f(x)$，都有

$$\lim_{n\to\infty}\int_{-\infty}^{+\infty}\delta_n(x)f(x)\,\mathrm{d}x=f(0) \tag{4}$$

[25]容易证明定义 1 是定义 2 的必要条件（只需要令定义 2 中的 $f(x)=1$ 证明式 2；再令 $f(x)$ 在区间 (a,b) 为 1，否则为 0，证明式 3.）．充分条件笔者不会证明，欢迎指正．如果有误以定义 2 为准．

一些物理教材会把 $\delta(x)$ 的性质简单记为

$$\delta(x)=\begin{cases}+\infty & (x=0)\\ 0 & (x\neq 0)\end{cases} \tag{5}$$

$$\int_{-\infty}^{+\infty}\delta(x)\,\mathrm{d}x=1 \tag{6}$$

这是十分不严谨的，$\delta(x)$ 不是函数，更不可积，$x\neq 0$ 时极限 $\lim_{n\to\infty}\delta_n(x)$ 也未必存在（例 1 ）.

习题 1　高斯函数

证明高斯分布函数[731] 可以构成以下 δ 函数列

$$\delta_n(x)=\frac{n}{\sqrt{\pi}}\mathrm{e}^{-n^2x^2}\qquad(n=1,2,\dots) \tag{7}$$

习题 2　sinc 函数

证明 sinc 函数[116] 可以构成以下 δ 函数列（提示：式 4 [116]）.

$$\delta_n(x)=\frac{n}{\pi}\operatorname{sinc}^2(nx)\qquad(n=1,2,\dots) \tag{8}$$

例 1　sinc 函数

证明 sinc 函数[116] 可以构成以下 δ 函数列

$$\delta_n(x)=\frac{n}{\pi}\operatorname{sinc}(nx)\qquad(n=1,2,\dots) \tag{9}$$

该式在傅里叶分析和量子力学中有重要应用 [7]，但证明起来比较困难只好从略（参考式 3 [116]）. 注意即使对于 $x\neq 0$ 上式也不存在 $n\to\infty$ 的极限而是做简谐振动，可见式 5 是十分不严谨的.

常见性质

性质 1

若一个等式中出现了所谓的 δ 函数 $\delta(x)$，那么其严格的定义是先将 $\delta(x)$ 替换为符合定义 1 的任意函数列 $\delta_n(x)$，令**等式在 $n\to\infty$ 的极限时成立**.

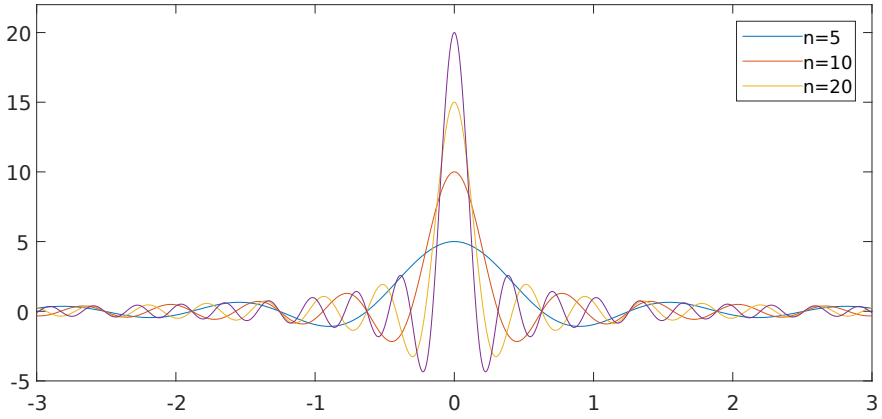


图 2: 例 1 中的 $n\,\mathrm{sinc}(nx)$

例如 $\delta(x)$ 一个重要的性质是：对任意在 $x = x_0$ 处连续函数 $f(x)$，有

$$\int_{-\infty}^{+\infty} f(x)\delta(x - x_0)\,\mathrm{d}x = f(x_0) \tag{10}$$

使用定义 1 中的 $\delta_n(x)$ 表示，该等式的严格意义是（注意极限和积分不可交换，且极限必须在最外面）

$$\lim_{n\to\infty} \int_{-\infty}^{+\infty} f(x)\delta_n(x - x_0)\,\mathrm{d}x = f(x_0) \tag{11}$$

由于我们假设定义 1 和定义 2 等效，这是成立的.

性质 2

$$\delta(ax) = \frac{1}{|a|}\delta(x) \tag{12}$$

我们从式 1 的几何上来不严谨地证明这个性质：与 $\delta(x)$ 相比较，$\delta(ax)$ 的图像在 x 方向变窄了 $|a|$ 倍，所以函数曲线下的面积变为原来的 $1/|a|$ 倍，故 $|a|\delta(ax)$ 下的面积是 1，证毕.

性质 3

作为式 12 的拓展，令 $f(x)$ 的根为 $x_1, x_2, \ldots$，在这些根处的导数为 $f'(x_i)$，那么

$$\delta[f(x)] = \sum_i \frac{1}{|f'(x_i)|}\delta(x - x_i) \tag{13}$$

证明和式 12 类似.

性质 4

对性质良好的函数 $g(x)$ 有

$$g(x)\delta(x) = g(0)\delta(x) \tag{14}$$

证明：对于性质良好的 $f(x)$，$\int f(x)g(x)\delta(x)\,\mathrm{d}x = f(0)g(0) = \int f(x)g(0)\delta(x)\,\mathrm{d}x$. 证毕.

其他应用

再次强调我们不能“按字面意思”理解任何含有 $\delta(x)$ 的等式.

例 2

在傅里叶分析中，时常会看到

$$\int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x = 2\pi\delta(k) \tag{15}$$

然而严格来说，左边的积分并不收敛，所以不能“按字面意思”理解该式. 要严格证明，上式两边除以 2π，令有限区间 $[-n, n]$ 内的积分为

$$\delta_n(k) = \frac{1}{2\pi}\int_{-n}^{n} \mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x = \frac{\sin(nk)}{\pi k} = \frac{n}{\pi}\operatorname{sinc}(nk) \tag{16}$$

由习题 2 中的结论得

$$\lim_{n\to\infty} \delta_n(k) = \delta(k) \tag{17}$$

证毕.

习题 3

证明

$$\begin{aligned}&\int_{-\infty}^{+\infty}\sin(k'x)\sin(kx)\,\mathrm{d}x=\int_{-\infty}^{+\infty}\cos(k'x)\cos(kx)\,\mathrm{d}x\\&=\pi\delta(k'-k)-\pi\delta(k'+k)\end{aligned}\tag{18}$$

提示：使用式 1 [352]，式 2 [352] 和式 15 .

例 3

请证明

$$\int_{-\infty}^{+\infty}\delta(x-x_1)\delta(x-x_2)\,\mathrm{d}x=\delta(x_1-x_2)\tag{19}$$

注意由此可得积分 $\int_{-\infty}^{+\infty}\delta(x)^2\,\mathrm{d}x=+\infty$，即不收敛.

证明：考虑和上文一样的含参函数 $\delta_n(x)$，令 $I_n=\int_{-\infty}^{+\infty}\delta_n(x)\,\mathrm{d}x$，有 $\lim_{n\to\infty}I_n=1$. 再令

$$f_n(x_1,x_2)=\int_{-\infty}^{+\infty}\delta_n(x-x_1)\delta_n(x-x_2)\,\mathrm{d}x\tag{20}$$

未完成：以下证明不严谨，待更正

我们希望证明 $\lim_{n\to\infty}f_n(x_1,x_2)=\delta(x_1-x_2)$. 首先对于给定的 $x_1\neq x_2$ 显然有 $\lim_{n\to\infty}f_n(x_1,x_2)=0$. 所以只需证明

$$\lim_{n\to\infty}\int_{-\infty}^{+\infty}f_n(x_1,x_2)\,\mathrm{d}x_2=1\tag{21}$$

交换积分顺序得

$$\begin{aligned}&\lim_{n\to\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\delta_n(x-x_1)\delta_n(x-x_2)\,\mathrm{d}x\,\mathrm{d}x_2\\&=\lim_{n\to\infty}\int_{-\infty}^{+\infty}\delta_n(x-x_1)\int_{-\infty}^{+\infty}\delta_n(x-x_2)\,\mathrm{d}x_2\,\mathrm{d}x\\&=\lim_{n\to\infty}I_n\int_{-\infty}^{+\infty}\delta_n(x-x_1)\,\mathrm{d}x\\&=\lim_{n\to\infty}I_n^2=1\end{aligned}\tag{22}$$

证毕.

多元狄拉克 delta 函数

- 本词条处于草稿阶段.

预备知识 狄拉克 delta 函数[230]

在三维直角坐标系中，

$$\delta(\mathbf{r}) = \delta(x)\delta(y)\delta(z) \tag{1}$$

这样就有

$$\int \delta(\mathbf{r})\,\mathrm{d}V = 1 \tag{2}$$

其他坐标系

若要把 $\delta(\mathbf{r}-\mathbf{r}_0)$ 转换到其他正交曲线坐标系中，直接使用

$$\delta(\mathbf{r}-\mathbf{r}_0) = \frac{1}{|\mathbf{r}_u(u_0)\mathbf{r}_v(v_0)\mathbf{r}_w(w_0)|}\delta(u-u_0)\delta(v-v_0)\delta(w-w_0) \tag{3}$$

其中 $\mathbf{r}_u(u_0)\mathbf{r}_v(v_0)\mathbf{r}_w(w_0)$ 分别是 $\mathbf{r}$ 在 u_0, v_0, w_0 对 u, v, w 的偏导数.

例如球坐标系中

$$\mathbf{r}_r = 1 \qquad \mathbf{r}_\theta = r \qquad \mathbf{r}_\phi = r\sin\theta \tag{4}$$

有

$$\delta(\mathbf{r}-\mathbf{r}_0) = \frac{1}{r_0^2|\sin\theta_0|}\delta(r-r_0)\delta(\theta-\theta_0)\delta(\phi-\phi_0) \quad (r_0, \theta_0 \neq 0) \tag{5}$$

但当 r_0 或 θ_0 为零时怎么办呢？用 δ^2 来表示！但如何证明？

未完成：……

傅里叶变换（三角）

- 本词条处于草稿阶段.

预备知识 傅里叶级数（三角）[222]

平方可积的函数都可以用正弦和余弦函数来展开

$$f(x)=\frac{1}{\sqrt{\pi}}\int_0^{+\infty}A(k)\cos(kx)+B(k)\sin(kx)\,\mathrm{d}k \tag{1}$$

其中

$$A(k)=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty}\cos(kx)f(x)\,\mathrm{d}x \tag{2}$$

$$B(k)=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty}\sin(kx)f(x)\,\mathrm{d}x \tag{3}$$

为了把整个实数范围的非周期函数用三角函数展开，在三角傅里叶级数中我们取极限 $l\to\infty$ 这时角频率的间隔趋于 0.

$$\Delta k=k_{n+1}-k_n=\frac{(n+1)\pi}{l}-\frac{n\pi}{l}=\frac{\pi}{l}\to 0 \tag{4}$$

傅里叶变换（指数）

- 本词条需要更多讲解，便于帮助理解.
- 本词条存在未完成的内容.

预备知识 傅里叶级数（指数）[228]，傅里叶变换（三角）[236]

用三角傅里叶变换[236] 中同样的方法可把指数傅里叶级数[228] 的区间长度 l 取极限后拓展为指数傅里叶变换

$$g(k)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}f(x)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x \tag{1}$$

$$f(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}g(k)\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}k \tag{2}$$

当 $f(x)$ 为实函数时，$g(k)$ 的实部是偶函数，虚部是奇函数.

实数函数的情况

如果实函数 $f(x)$ 的复数傅里叶变换为 $g(k)$，即 $g(k)$ 需要满足什么条件才能使 $f(x)$ 是实数呢？我们从式 2 开始入手

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}} \int_0^\infty [g(k)\mathrm{e}^{\mathrm{i}kx} + g(-k)\mathrm{e}^{-\mathrm{i}kx}]\,\mathrm{d}k \\ &= \frac{1}{\sqrt{2\pi}} \int_0^\infty [g(k) + g(-k)]\cos(kx)\,\mathrm{d}k \\ &\quad + \frac{\mathrm{i}}{\sqrt{2\pi}} \int_0^\infty [g(k) - g(-k)]\sin(kx)\,\mathrm{d}k \end{aligned} \tag{3}$$

从傅里叶变换（三角函数）我们已知对实数函数，方括号项都必须为实函数，即上式第一个方括号中的虚部为零，第二个方括号中的实部为零，即

$$\begin{aligned} g_{Re}(-k) &= g_{Re}(k) \\ g_{Im}(-k) &= -g_{Im}(k) \end{aligned} \tag{4}$$

所以结论是，当 $f(x)$ 为实函数时，$g(k)$ 的实部是偶函数，虚部是奇函数．或用复共轭记为

$$g(-k) = g(k)^* \tag{5}$$

所以对于实数函数的傅里叶变换，往往只需要 k 的正半轴．

例 1　高斯分布的傅里叶变换

要计算高斯函数

$$f(x) = \exp\left(-ax^2\right) \quad (a > 0) \tag{6}$$

的傅里叶变换，代入式 1 并使用例 2 [774] 有

$$g(k) = \sqrt{\frac{\pi}{a}} \exp\left(-\frac{k^2}{4a}\right) \tag{7}$$

一个方便的记忆法是 x^2 前的系数乘以 k^2 前的系数相乘等于 $1/4$.

例 2

$$f(x) = \begin{cases} \exp(\mathrm{i}k_0 x)\cos^2(ax) & (|x| < \frac{\pi}{2a}) \\ 0 & (\text{其他}) \end{cases} \tag{8}$$

则傅里叶变换为

$$g(k)=\frac{\sqrt{2\pi}a}{4a^2-(k-k_0)^2}\operatorname{sinc}\left[\frac{\pi(k-k_0)}{2a}\right] \tag{9}$$

其中 sinc 函数见相关页面[116].

证明

预备知识 狄拉克 delta 函数[230]

以下的证明可以用矢量空间和基底的概念得到更深刻的理解，详见“傅里叶变换与连续正交归一基底[240]”.

我们把式 2 看作定义，用狄拉克 δ 函数[230] 来证明式 1，反之同理. 把式 2 代入式 1 得

$$g(k)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty}g(k')\mathrm{e}^{\mathrm{i}k'x}\,\mathrm{d}k'\right]\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x \tag{10}$$

交换积分顺序，先对 x 积分得

$$g(k)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty}\mathrm{e}^{\mathrm{i}(k'-k)x}\,\mathrm{d}x\right]g(k')\,\mathrm{d}k' \tag{11}$$

由例 2 [230]，里面的积分变为 $2\pi\delta(k'-k)$，现在只需证明

$$g(k)=\int_{-\infty}^{+\infty}\delta(k'-k)g(k')\,\mathrm{d}k' \tag{12}$$

而这就是式 10 [230]. 证毕.

性质

为了书写方便我们用**算符** $\mathcal{F}$ 和 $\mathcal{F}^{-1}$ 表示傅里叶变换和反变换，即 $\mathcal{F}f=g$ 以及 $\mathcal{F}^{-1}g=f$. 算符在这里可以看作“函数的函数”，即自变量和函数值都是函数.

$$\mathcal{F}[f(x)\mathrm{e}^{\mathrm{i}k_0x}]=g(k-k_0) \tag{13}$$

$$\mathcal{F}^{-1}[g(k)\mathrm{e}^{-\mathrm{i}kx_0}]=f(x-x_0) \tag{14}$$

也就是说，给函数乘以 $\mathrm{e}^{\mathrm{i}k_0x}$ 因子再做傅里叶变换，等于先对函数做傅里叶变换，再向右平移 k_0. 给函数乘以 $\mathrm{e}^{-\mathrm{i}kx_0}$ 因子再做反傅里叶变换，等于先对函数做反傅里叶变换，再向右平移 x_0. 证明显然，留做习题.

变换前后模长不变

$$\int_{-\infty}^{+\infty} g(k)^* g(k)\,\mathrm{d}k = \int_{-\infty}^{+\infty} f(x)^* f(x)\,\mathrm{d}x \tag{15}$$

导数的傅里叶变换

$$\mathcal{F}[f'(x)] = \mathrm{i}kg(k) \tag{16}$$

同理

$$\mathcal{F}^{-1}[g'(k)] = -\mathrm{i}xf(x) \tag{17}$$

作为式 15 的拓展，有

$$\int_{-\infty}^{+\infty} f_1(x)^* f_1(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} g_1(k)^* g_2(k)\,\mathrm{d}k \tag{18}$$

如果 $f_1(x)$ 可以在 $x=0$ 泰勒展开，有

$$\mathcal{F}[f_1(x)f_2(x)] = f_1\left(\mathrm{i}\frac{\partial}{\partial k}\right)g_2(k) \tag{19}$$

性质的证明

证明式 15 ：把傅里叶变换看成傅里叶级数在 $l \to \infty$ 时的极限，使用式 14 [228]，右边的求和在极限下变为积分即可证明．详细过程留做习题．

证明式 16 ：对式 2 关于 x 求导得

$$f'(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} [\mathrm{i}kg(k)]\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}k \tag{20}$$

把方括号看作一整个 k 的函数，那么上式对应的反变换为

$$\mathrm{i}kg(k) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f'(x)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x = \mathcal{F}[f'(x)] \tag{21}$$

其中 $g(k) = F[f(x)]$，证毕．

未完成：证明其他性质

连续正交归一基底与傅里叶变换

预备知识 傅里叶变换（指数）[237]

离散的函数基底

本文使用狄拉克符号[470]．在“傅里叶级数（三角）[222]”中，我们介绍了正交归一函数基底的概念，即把满足一定条件的一元函数的集合看作一个矢量空间[444]，两个函数（矢量）的内积[471] 定义为

$$\langle f|g\rangle = \int_{-\infty}^{+\infty} f(x)^* g(x)\,\mathrm{d}x \tag{1}$$

其中 $*$ 表示复共轭，如果空间中的函数都是实函数则可忽略．

该空间中的一组正交归一基底用狄拉克符号[470] 表示为 $|x_i\rangle$（$i=1,2,\dots$），基底的个数可以是有限个或无限个，空间的维数就是基底的个数．

基底满足正交归一条件（式 1 [386]）

$$\langle x_i|x_j\rangle = \delta_{i,j} \tag{2}$$

若这组正交归一基底是完备的，那么如果一个函数可以分解为它们的线性组合：

$$|f\rangle = \sum_j c_j\,|x_j\rangle \tag{3}$$

两边左乘 $\langle x_i|$，则有

$$\langle x_i|f\rangle = \sum_j c_j\,\langle x_i|x_j\rangle = \sum_j c_j\delta_{i,j} = c_i \tag{4}$$

即

$$c_i = \langle x_i|f\rangle \tag{5}$$

式 4 的过程相当于用正交归一性把 $|x_i\rangle$ 项从式 3 的求和中筛选了出来．我们得到几何矢量中一个熟悉的结论：一个矢量关于一组正交归一基底的坐标等于它在每个基底上的投影．

连续的函数基底

我们接下来用类似的方法来理解傅里叶变换（式 1 [237]）.

$$g(k)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}f(x)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x \tag{6}$$

$$f(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}g(k)\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}k \tag{7}$$

我们令所有可以做傅里叶变换的函数构成的空间为 X，从傅里叶变换的公式，我们猜想该空间的正交归一“基底”为

$$|k\rangle=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\mathrm{i}kx}\qquad(k\in\mathbb{R}) \tag{8}$$

严格来说，X 空间的函数必须要满足 $\langle x|x\rangle$ 为有限值，而式 8 中的函数显然不满足这点，所以它们并不属于 X 空间，而是一个包含 X 的更大的空间，所以这个“基底”只是一个形象的说法，需要加上引号.

显然，式 8 中的任意两个“基底”的内积都不收敛，而且 k 的取值是**连续**的，所以我们不可能用式 2 表示它们的正交归一关系. 但通过（式 15 [230]）

$$\int_{-\infty}^{+\infty}\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x=2\pi\delta(k) \tag{9}$$

可以得到一个和式 2 类似的关系

$$\begin{aligned}\langle k'|k\rangle&=\int_{-\infty}^{+\infty}\frac{\mathrm{e}^{-\mathrm{i}k'x}}{\sqrt{2\pi}}\frac{\mathrm{e}^{\mathrm{i}kx}}{\sqrt{2\pi}}\,\mathrm{d}x\\&=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{\mathrm{i}(k'-k)x}\,\mathrm{d}x=\delta(k'-k)\end{aligned} \tag{10}$$

即

$$\langle k'|k\rangle=\delta(k'-k) \tag{11}$$

这可以看作是**连续基底的正交归一条件**.

现在，把式 6 和式 7 用狄拉克符号表示为

$$g(k)=\langle k|f\rangle \tag{12}$$

$$f(x)=\int_{-\infty}^{+\infty}g(k')\,|k'\rangle\,\mathrm{d}k' \tag{13}$$

它们可以分别看作是把式 5 和 式 3 拓展到连续基底的情况．根据定义，任何能做傅里叶（反）变换的 $f(x)$ 必定能展开成式 13 的形式．再来证明式 12，过程和式 4 类似：式 13 两边左乘 $\langle k|$，使用 δ 函数的性质式 10 [230] 把积分中 $|k\rangle$ 基底的系数“筛选”出来

$$\begin{aligned}\langle k|f\rangle &= \int_{-\infty}^{+\infty} g(k)\,\langle k|k'\rangle\,\mathrm{d}k \\ &= \int_{-\infty}^{+\infty} g(k)\delta(k-k')\,\mathrm{d}k' = g(k)\end{aligned} \tag{14}$$

证毕．

注意该证明并不仅限于傅里叶变换一种情况，任何连续的基底 $|k\rangle$ 只要满足正交归一条件式 11 ，且可以展开某函数 $f(x)$，就都能使式 12 成立．

傅里叶变换与矢量空间

- 本词条存在未完成的内容．

预备知识 傅里叶变换与连续正交归一基底[240]

在“傅里叶变换与连续正交归一基底[240]”中，我们已经用矢量空间和连续基底的概念来理解傅里叶变换，但我们把函数看作矢量本身．而在本文中，我们再次用另一种方法理解傅里叶变换，即把函数看作是矢量的坐标而不是矢量本身．注意这种理解和“连续基底”的概念一样，并不十分严谨，但对于理解量子力学中的“位置表象和动量表象（链接未完成）”却十分有用．

类比傅里叶级数，我们仍然可以将傅里叶变换看作是矢量空间中两组**正交归一基底**之间的变换，我们分别把他们叫做 x 基底和 k 基底．若使用 x 基底，就说在 x **表象**下，若用 k 基底就说在 k 表象下．每个实数 x_0 对应一个基底 $|x_0\rangle$，所有的 $|x_0\rangle$ 构成 x 基底．每个实数 k_0 对应一个基底 $|k_0\rangle$，所有的 $|k_0\rangle$ 构成 k 基底．就像有限维空间中用 $(0,\ldots,1,\ldots,0)$ 表示一组基底的第 i 个关于这组基底的坐标，可以用狄拉克 delta 函数[230] $\delta(x-x_0)$ 表示 $|x_0\rangle$ 关于所有 x 基底的坐标，用 $\delta(k-k_0)$ 表示 $|k_0\rangle$ 关于所有 k 基底的坐标．

现在，函数 $f(x)$ 可以看作是某个矢量 $|v\rangle$ 关于 x 基底的坐标（有限维空间中的求和在无穷维空间中变为积分），而 $g(k)$ 可以看作是 k 基底的坐标[26].

$$|v\rangle = \int f(x_0)\,|x_0\rangle\,\mathrm{d}x_0 = \int g(k_0)\,|k_0\rangle\,\mathrm{d}k_0 \tag{1}$$

矢量空间中两个矢量的内积，在 x 表象下为

$$\begin{aligned}\langle v_1|v_2\rangle &= \left(\int f_1(x_1)\,|x_1\rangle\,\mathrm{d}x_1\right)^{\dagger}\int f_2(x_2)\,|x_2\rangle\,\mathrm{d}x_2\\ &= \iint f_1(x_1)^* f_2(x_2)\delta(x_2-x_1)\,\mathrm{d}x_1\,\mathrm{d}x_2\\ &= \int f_1(x)^* f_2(x)\,\mathrm{d}x\end{aligned} \tag{2}$$

k 表象同理.

将某矢量 $|v\rangle$ 投影到基底 $|x_0\rangle$ 上，可以验证其系数为

$$\langle x_0|v\rangle = \int f(x)\delta(x-x_0)\,\mathrm{d}x = f(x_0) \tag{3}$$

x 表象下，可以验证 x 基底的正交归一化

$$\langle x_1|x_2\rangle = \int \delta(x-x_1)\delta(x-x_2)\,\mathrm{d}x = \delta(x_2-x_1) \tag{4}$$

k 表象同理.

基底变换

将傅里叶变换对应的算符记为 $\mathcal{F}$，反变换记为 $\mathcal{F}^{-1}$. 傅里叶变换可以看成一个无穷维且连续的**酉矩阵**，将同一个矢量从 x 表象坐标变换到 k 表象坐标. 反变换从 k 表象变回 x 表象. 傅里叶变换和反变换并不改变矢量 $|v\rangle$ 本身，$\mathcal{F}$ 是一个单位算符，即 $\mathcal{F}\,|v\rangle = |v\rangle$.

令 $|k_0\rangle$ 基底在 x 表象下的函数（即坐标）为 $\mathrm{e}^{\mathrm{i}k_0x}/\sqrt{2\pi}$，即

$$|k_0\rangle = \int \frac{\mathrm{e}^{\mathrm{i}k_0x_0}}{\sqrt{2\pi}}\,|x_0\rangle\,\mathrm{d}x_0 \tag{5}$$

[26]从该式可以看出**表象**这个词的由来，同一个矢量，使用不同基底后，其坐标（即函数）$f(x)$ 和 $g(x)$ 表面上看起来不同.

则基底变换矩阵的“矩阵元”为

$$\begin{aligned}\langle x_0|k_0\rangle &= \langle x_0| \int \frac{\mathrm{e}^{\mathrm{i}k_0x_1}}{\sqrt{2\pi}} |x_1\rangle \,\mathrm{d}x_1 \\ &= \int \frac{\mathrm{e}^{\mathrm{i}k_0x_1}}{\sqrt{2\pi}} \delta(x_1 - x_0)\,\mathrm{d}x_1 = \frac{\mathrm{e}^{\mathrm{i}k_0x_0}}{\sqrt{2\pi}}\end{aligned} \tag{6}$$

我们将矩阵元记为 $F^*(x_0, k_0)$，也叫做**核（kernel）**．两个变量就相当于矩阵的两个角标．

该矩阵乘以 k 表象的坐标，就是 x 表象的坐标，矩阵的求和同样需要写成积分的形式

$$f(x_0) = \int F^*(x_0, k_0) g(k_0)\,\mathrm{d}k_0 = \int \frac{\mathrm{e}^{\mathrm{i}k_0x_0}}{\sqrt{2\pi}} g(k_0)\,\mathrm{d}k_0 \tag{7}$$

这就是反傅里叶变换．

由于酉矩阵的厄米共轭就是它的逆矩阵，所以逆矩阵的矩阵元为

$$\langle k_0|x_0\rangle = \langle x_0|k_0\rangle^* = \frac{\mathrm{e}^{-\mathrm{i}k_0x_0}}{\sqrt{2\pi}} \tag{8}$$

核为 $F(k_0, x_0)$．该矩阵乘以 x 表象的坐标，就是 k 表象的坐标

$$g(k_0) = \int F(k_0, x_0) f(x_0)\,\mathrm{d}x_0 = \int \frac{\mathrm{e}^{-\mathrm{i}k_0x_0}}{\sqrt{2\pi}} f(x_0)\,\mathrm{d}x_0 \tag{9}$$

这就是傅里叶变换．

x 表象下，k 基底的正交归一化可以记为

$$\langle k_1|k_2\rangle = \int \frac{\mathrm{e}^{\mathrm{i}k_1x}}{\sqrt{2\pi}} \frac{\mathrm{e}^{-\mathrm{i}k_2x}}{\sqrt{2\pi}}\,\mathrm{d}x = \delta(k_2 - k_1) \tag{10}$$

x 基底在 k 表象下的正交归一也同理．之所以说这时一种“幼稚”的做法，是因为式 10 的积分（从 $-\infty$ 到 ∞）并不收敛，δ 函数也不是一个严格意义上的函数．

我们就可以用以下过程在 x 表象下简洁地“证明”$f(x)$ 依次经过傅里叶变换和反变换后，仍然可以得到 $f(x)$．证明的矢量形式为

$$|v\rangle = \int |k\rangle \langle k|v\rangle \,\mathrm{d}k \tag{11}$$

将各个矢量替换为 x 表象的坐标，得

$$\begin{aligned}\mathcal{F}^{-1}[\mathcal{F}f(x)] &= \int \frac{\mathrm{e}^{\mathrm{i}kx}}{\sqrt{2\pi}}\left(\int \frac{\mathrm{e}^{-\mathrm{i}kx'}}{\sqrt{2\pi}} f(x')\,\mathrm{d}x'\right)\mathrm{d}k \\ &= \int f(x')\left(\int \frac{\mathrm{e}^{-\mathrm{i}kx'}}{\sqrt{2\pi}}\frac{\mathrm{e}^{\mathrm{i}kx}}{\sqrt{2\pi}}\,\mathrm{d}k\right)\mathrm{d}x' \\ &= \int f(x')\delta(x-x')\,\mathrm{d}x' \\ &= f(x)\end{aligned} \tag{12}$$

拉普拉斯变换

拉普拉斯变换常用于微分方程的初始值问题，即已知某个物理量在初始时刻 $t=0$ 的值 $f(0)$，而求解它在初始时刻之后的变化情况 $f(t)$．至于它在初始时刻之前的值，我们就让它都等于 0，也就是说

$$f(t)=0 \quad (t<0) \tag{1}$$

为了获得较宽的变换条件，构造一个函数 $g(t)$，

$$g(t)=\mathrm{e}^{-\sigma t}f(t) \tag{2}$$

这里 $e^{-\sigma t}$ 为收敛因子．我们需要选一个充分大的正实数 σ，用来保证 $g(t)$ 在区间 $(-\infty,+\infty)$ 上绝对可积．于是，可以对 $g(t)$ 做傅里叶变换：

$$G(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty} g(t)\mathrm{e}^{-\omega t}\mathrm{d}t=\frac{1}{2\pi}\int_{0}^{\infty} f(t)\mathrm{e}^{-(\sigma+\mathrm{i}\omega)t}\mathrm{d}t \tag{3}$$

将 $\sigma+i\omega$ 记作 p，并将 $G(\omega)$ 改记作 $\bar{f}(p)/2\pi$ 则

$$\bar{f}(p)=\int_{0}^{\infty} f(t)\mathrm{e}^{-pt}\mathrm{d}t \tag{4}$$

其中积分

$$\int_{0}^{\infty} f(t)\mathrm{e}^{-pt}\mathrm{d}t \tag{5}$$

称为**拉普拉斯积分**．$\bar{f}(p)$ 称为 $f(t)$ 的**拉普拉斯函数**．式 4 称为**拉普拉斯变换**（简称**拉氏变换**），e^{-pt} 是拉普拉斯变换的**核**．

$G(\omega)$ 的傅里叶逆变换是：

$$g(t)=\int_{-\infty}^{\infty}G(\omega)\mathrm{e}^{\mathrm{i}\omega t}\mathrm{d}\omega=\frac{1}{2\pi}\int_{-\infty}^{\infty}\bar{f}(\sigma+\mathrm{i}\omega)\mathrm{e}^{\mathrm{i}\omega t}\mathrm{d}\omega \tag{6}$$

即

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\bar{f}(\sigma+\mathrm{i}\omega)\mathrm{e}^{(\sigma+\mathrm{i}\omega)}\mathrm{d}\omega \tag{7}$$

由于 $\sigma+\mathrm{i}\omega=p$，可得 $\mathrm{d}\omega=1/\mathrm{i}\mathrm{d}p$．所以有

$$f(t)=\frac{1}{2\pi\mathrm{i}}\int_{\sigma-\mathrm{i}\infty}^{\sigma+\mathrm{i}\infty}\bar{f}(p)\mathrm{e}^{\mathrm{i}p}\mathrm{d}p \tag{8}$$

$\bar{f}(p)$ 又称为**像函数**，而 $f(t)$ 称为**原函数**．它们之间的关系通常简单地用记号表示为：

$$\begin{aligned}\bar{f}(p)&=\mathscr{S}[f(t)]\\ f(t)&=\mathscr{L}^{-1}[\bar{f}(p)]\end{aligned} \tag{9}$$

我们来看几个例题熟悉一下它的运算．

例 1

求 $\mathscr{L}[1]$.

在 $\mathrm{Re}\,p>0$（即 $\sigma>0$）的半平面上，

$$\int_0^{\infty}1\cdot e^{-pt}dt=\frac{1}{p} \tag{10}$$

所以

$$\mathscr{L}[1]=\frac{1}{p}\quad(\mathrm{Re}\,p>0) \tag{11}$$

例 2

求 $\mathscr{L}[t]$.

在 $\mathrm{Re}\,p>0$ 的半平面上

$$\begin{aligned}\int_0^{\infty}t\mathrm{e}^{-pt}\mathrm{d}t&=-\frac{1}{p}\int_0^{\infty}t\mathrm{d}\left(\mathrm{e}^{-pt}\right)\\ &=-\frac{1}{p}\left[t\mathrm{e}^{-pt}\right]_0^{\infty}+\frac{1}{p}\int_0^{\infty}\mathrm{e}^{-pt}\mathrm{d}t\\ &=\frac{1}{p}\int_0^{\infty}\mathrm{e}^{-pt}\mathrm{d}t=\frac{1}{p^2}\end{aligned} \tag{12}$$

所以

$$\mathscr{L}[t]=\frac{1}{p^2}\quad(\mathrm{Re}\,p>0) \tag{13}$$

例 3

求 $\mathscr{L}[\mathrm{e}^{st}]$，$s$ 为常数.

在 $\operatorname{Re} p > \operatorname{Re} s$ 的半平面上，

$$\begin{aligned}\int_0^{\infty} \mathrm{e}^{st}\mathrm{e}^{-pt}\mathrm{d}t = \int_0^{\infty} \mathrm{e}^{-(p-s)t}\mathrm{d}t &= -\frac{1}{p-s}\left[\mathrm{e}^{-(p-s)t}\right]_0^{\infty} \\ &= \frac{1}{p-s}\end{aligned} \tag{14}$$

所以

$$\mathscr{L}\left[\mathrm{e}^{st}\right] = \frac{1}{p-s} \quad (\operatorname{Re} p > \operatorname{Re} s) \tag{15}$$

例 4

求 $\mathscr{L}[t\mathrm{e}^{st}]$，$s$ 为常数.

在 $\operatorname{Re} p > \operatorname{Re} s$ 的半平面上，

$$\begin{aligned}\int_0^{\infty} t\mathrm{e}^{st}\mathrm{e}^{-pt}\mathrm{d}t &= -\frac{1}{p-s}\int_0^{\infty} t\mathrm{d}\left[\mathrm{e}^{-(p-s)t}\right] \\ &= -\frac{1}{p-s}\left\{\left[t\mathrm{e}^{-(p-s)t}\right]_0^{\infty} - \int_0^{\infty}\mathrm{e}^{-(p-s)t}\mathrm{d}t\right\} \\ &= \frac{1}{(p-s)^2}\end{aligned} \tag{16}$$

所以

$$\mathscr{L}\left[t\mathrm{e}^{st}\right] = \frac{1}{(p-s)^2} \quad (\operatorname{Re} p > \operatorname{Re} s) \tag{17}$$

同理

$$\mathscr{L}\left[t^n\mathrm{e}^{n}\right] = \frac{n!}{(p-s)^{n+1}} \tag{18}$$

拉普拉斯变换的性质

预备知识 拉普拉斯变换[246]

前面我们介绍了拉普拉斯变换，那么这个变换到底给我们带来了什么特别的好处呢？我们下面就来看一下它到底具有哪些性质.

线性性质

若 $\mathscr{L}[f_1(t)]=\bar{f}_1(p)$，$\mathscr{L}[f_2(t)]=\bar{f}_2(p)$，则

$$\mathscr{L}[c_1f_1(t)+c_2f_2(t)]=c_1\bar{f}_1(p)+c_2\bar{f}_2(p) \tag{1}$$

证明起来也十分容易.

$$\begin{aligned}\mathscr{L}[c_1f_1(t)+c_2f_2(t)]&=\int_0^\infty[c_1f_1(t)+c_2f_2(t)]\,\mathrm{e}^{-\rho t}\mathrm{d}t\\&=\int_0^\infty c_1f_1(t)\mathrm{e}^{-pt}\mathrm{d}t+\int_0^\infty c_2f_2(t)\mathrm{e}^{-\rho t}\mathrm{d}t\\&=c_1\bar{f}_1(p)+c_2\bar{f}_2(p)\end{aligned} \tag{2}$$

它能让我们干什么呢？我们可以把未知的函数拆成已知函数的线性组合，相加它们的拉普拉斯变换来得到所要的未知函数的拉普拉斯变换．例如：

例 1　线性性质的应用

求 $\mathscr{L}[\sin\omega t]$，$\omega$ 为常数.

我们知道

$$\sin\omega t=\frac{1}{2\mathrm{i}}\left(\mathrm{e}^{\mathrm{i}\omega t}-\mathrm{e}^{-\mathrm{i}\omega t}\right) \tag{3}$$

所以有

$$\begin{aligned}\mathscr{L}[\sin\omega t]&=\mathscr{L}\left[\frac{1}{2\mathrm{i}}\left(\mathrm{e}^{\mathrm{i}\omega t}-\mathrm{e}^{-\mathrm{i}\omega t}\right)\right]=\frac{1}{2\mathrm{i}}\mathscr{L}\left[\mathrm{e}^{\mathrm{i}\omega t}\right]-\frac{1}{2\mathrm{i}}\mathscr{L}\left[\mathrm{e}^{-\mathrm{i}\omega t}\right]\\&=\frac{1}{2\mathrm{i}}\left[\frac{1}{p-\mathrm{i}\omega}-\frac{1}{p+\mathrm{i}\omega}\right]\\&=\frac{\omega}{p^2+\omega^2}\quad(\mathrm{Re}\,p>0)\end{aligned} \tag{4}$$

同理可得

$$\mathscr{L}[\cos\omega t]=\frac{p}{p^2+\omega^2}\quad(\mathrm{Re}\,p>0) \tag{5}$$

导数定理

拉普拉斯变换的导数定理是说：

$$\mathscr{L}[f'(t)]=p\bar{f}(p)-f(0) \tag{6}$$

证明同样直接积分就可以了.

$$\begin{aligned}\mathscr{L}[f'(t)] &= \int_0^\infty f'(t)\mathrm{e}^{-pt}\mathrm{d}t = \int_0^\infty \mathrm{e}^{-pt}\mathrm{d}f \\ &= \left[\mathrm{e}^{-pt}f(t)\right]_0^\infty - \int_0^\infty f(t)\mathrm{d}\left(\mathrm{e}^{-pt}\right)\end{aligned} \tag{7}$$

取 $\mathrm{Re}\,p > \sigma$，有 $\lim_{t\to\infty}\mathrm{e}^{-pt}f(t) = 0$. 于是可得

$$\begin{aligned}\mathscr{L}[f'(t)] &= -f(0) - \int_0^\infty f(t)\mathrm{d}\left(\mathrm{e}^{-\rho t}\right) = p\int_0^\infty f(t)\mathrm{e}^{-pt}\mathrm{d}t - f(0) \\ &= p\bar{f}(p) - f(0) \quad (\mathrm{Re}\,p > \sigma_0)\end{aligned} \tag{8}$$

我们还可以推广到高阶导数的情形：

$$\mathscr{L}[f^{(n)}(t)] = p^n f(p) - p^{n-1}f(0) - p^{n-2}f'(0) - \cdots - pf^{(n-2)}(0) - f^{(n-1)}(0) \tag{9}$$

有了导数定理，让我们来想想导数的逆运算——积分！有没有积分定理呢？有！

积分定理

$$\mathscr{L}\left[\int_0^t \psi(\tau)\mathrm{d}\tau\right] = \frac{1}{p}\mathscr{L}[\psi(t)] \tag{10}$$

证明其实也一样简单. 设 $f(t) = \int_0^t \psi(\tau)\mathrm{d}\tau$，对 $f(t)$ 应用导数定理，可知：

$$[f'(t)] = p\mathscr{L}[f(t)] - f(0) = p\mathscr{L}[f(t)] \tag{11}$$

所以也就有

$$\frac{1}{p}\mathscr{L}[\psi(t)] = \mathscr{L}[f(t)] = \mathscr{L}\left[\int_0^t \psi(\tau)\mathrm{d}\tau\right] \tag{12}$$

这也就证明了定理.

相似性定理

$$\mathscr{L}[f(at)] = \frac{1}{a}\bar{f}\left(\frac{p}{a}\right) \tag{13}$$

证明留作练习.

位移定理

$$\mathscr{L}[\mathrm{e}^{-\lambda t}f(t)]=\bar{f}(p+\lambda) \tag{14}$$

这个证明也十分简单，留作练习.

延迟定理

$$\mathscr{L}\left[f\left(t-t_0\right)\right]=\mathrm{e}^{-pt_0}\bar{f}(p) \tag{15}$$

用换元积分即可证明. 具体来说：

$$\mathscr{L}\left[f\left(t-t_0\right)\right]=\int_0^{\infty}f\left(t-t_0\right)\mathrm{e}^{-pt}\mathrm{d}t \tag{16}$$

我们知道在 $t<0$ 时，默认 $f(t)$ 都是 0. 那么我们做积分的时候，时间可以改为从 t_0 开始积分. 也就是说：

$$\mathscr{L}\left[f\left(t-t_0\right)\right]=\int_{t_0}^{\infty}f\left(t-t_0\right)\mathrm{e}^{-pt}\mathrm{d}t \tag{17}$$

作换元 $\xi=t-t_0$，那么就有

$$\begin{aligned}\mathscr{L}\left[f\left(t-t_0\right)\right]&=\int_0^{\infty}f(\xi)\mathrm{e}^{-p(\xi+t_0)}\mathrm{d}\xi=\mathrm{e}^{-pt_0}\int_0^{\infty}f(\xi)\mathrm{e}^{-p\xi}\mathrm{d}\xi\\&=\mathrm{e}^{-pt_0}\bar{f}(p)\end{aligned} \tag{18}$$

卷积定理

卷积定理是说，若 $\mathscr{L}[f_1(t)]=\bar{f}_1(p),\mathscr{L}[f_2(t)]=\bar{f}_2(p)$，则

$$\mathscr{L}[f_1(t)*f_2(t)]=\bar{f}_1(p)\bar{f}_2(p) \tag{19}$$

下面我们来证明它. 这个证明稍复杂一些，毕竟它是二重积分：

$$\begin{aligned}&\mathscr{S}\left[f_1(t)*f_2(t)\right]\\&=\int_0^{\infty}f_1(t)*f_2(t)\mathrm{e}^{-pt}\mathrm{d}t\\&=\int_0^{\infty}\left[\int_0^{t}f_1(\tau)f_2(t-\tau)\mathrm{d}\tau\right]\mathrm{e}^{-pt}\mathrm{d}t\end{aligned} \tag{20}$$

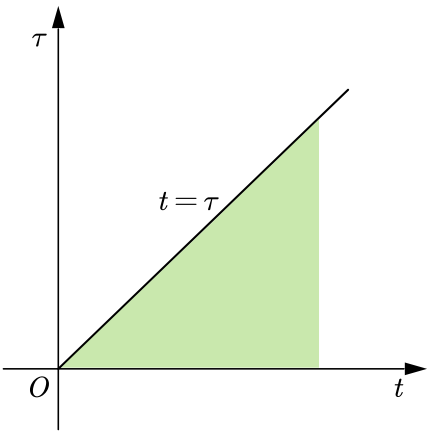


图 1: 积分限

我们需要交换积分次序，首先我们来看一下积分区域，它就是图 1 中的绿色区域．由此可见，现在改变积分次序，就是说

$$\begin{aligned}&\mathscr{L}\left[f_1(t) * f_2(t)\right]\\&=\int_0^{\infty}\left[\int_{\tau}^{\infty} f_2(t-\tau)\mathrm{e}^{-pt}\mathrm{d}t\right] f_1(\tau)\mathrm{d}\tau\end{aligned}\tag{21}$$

作换元，$\xi = t - \tau$，那么就有

$$\begin{aligned}\mathscr{L}\left[f_1(t) * f_2(t)\right] &= \int_0^{\infty}\left[\int_0^{\infty} f_2(\xi)\mathrm{e}^{-p\xi}\mathrm{d}\xi\right] f_1(\tau)\mathrm{e}^{-p\tau}\mathrm{d}\tau\\&= \int_0^{\infty} f_1(\tau)\mathrm{e}^{-p\tau}\mathrm{d}\tau \int_0^{\infty} f_2(\xi)\mathrm{e}^{-p\xi}\mathrm{d}\xi\\&= \bar{f}_1(p)\bar{f}_2(p)\end{aligned}\tag{22}$$

习题 1

求下列函数的拉普拉斯变换．

1.

$$\operatorname{sh}\omega t, \quad \operatorname{ch}\omega t \tag{23}$$

2.

$$\mathrm{e}^{-\lambda t}\sin\omega t, \quad \mathrm{e}^{-\lambda t}\cos\omega t \tag{24}$$

3.

$$\frac{1}{\sqrt{\pi t}} \tag{25}$$

4.

$$\delta(t-\tau) \tag{26}$$

柯西—阿达玛公式

预备知识 幂级数[206]

表述与证明

[27]**柯西—阿达玛公式（Cauchy-Hadamard formula）**是计算幂级数收敛半径的一般公式.

定理 1　柯西—阿达玛公式

设有幂级数

$$\sum_{n=0}^{\infty} c_n(z-a)^n. \tag{1}$$

则幂级数的收敛半径 R 由公式

$$\frac{1}{R}=\limsup_{n\to\infty}|c_n|^{1/n}$$

计算.

证明并不困难. 如果 $|z-a|>R$，那么可取 $\delta>0$ 如此之小，使得 $|z-a|>R(1+\delta)$，从而有

$$|c_n||z-a|^n>(|c_n|^{1/n}R(1+\delta))^n.$$

而按照上极限的定义，有无穷多个 n 使得

$$|c_n|^{1/n}>\frac{1}{R(1+\delta)}.$$

于是式 1 的一般项不趋于零，从而这级数不可能收敛. 而如果 $|z-a|<R$，那么可取 $\delta>0$ 如此之小，使得 $|z-a|<R(1+2\delta)^{-1}$; 而按照上极限的定义，从某个 n 开始总有 $|c_n|^{1/n}<(1+\delta)/R$，因此对于充分大的 n 就有

$$|c_n||z-a|^n\leq\frac{(1+\delta)^n}{(1+2\delta)^n}.$$

[27]参考 Wikipedia 相关页面.

从而幂级数的一般项由收敛的几何级数控制.

应用举例

对于幂级数

$$\sum_{n=0}^{\infty} c_n(z-a)^n,$$

逐项微分和逐项积分的级数分别是

$$\sum_{n=1}^{\infty} nc_n(z-a)^{n-1}, \sum_{n=0}^{\infty} \frac{c_n}{n+1}(z-a)^{n+1}.$$

按照柯西—阿达玛公式，这两个幂级数的收敛半径都与原幂级数相同. **因此逐项微分或逐项积分不改变幂级数的收敛半径.**

对于 $\lambda > 0$，按照柯西—阿达玛公式，幂级数

$$\sum_{n=0}^{\infty} \frac{z^n}{n^{n/\lambda}}$$

的收敛半径是无穷大，而且实际上存在 $A, B > 0$ 使得

$$\left|\sum_{n=0}^{\infty} \frac{z^n}{n^{n/\lambda}}\right| \leq Ae^{B|z|^\lambda}.$$

以 $d_1(n)$ 表示 n 的非 1 因子个数. 幂级数

$$\sum_{n=1}^{\infty} d_1(n)z^n$$

的系数涨落很没有规律：当 n 为素数时 $d_1(n) = 1$，而一般的正整数却可能有很多个因子. 但显然 $1 \leq d_1(n) < n$，因此

$$\lim_{n\to\infty} d_1(n)^{1/n} = 1,$$

所以按照柯西—阿达玛公式，此幂级数的收敛半径是 1.

Gamma 函数

预备知识 定积分[198]

[28]Γ 函数（**gamma function**）可以看成是阶乘在实数或复数域的拓展，该函数有多种定义方法，这里先讨论实数域上的定积分定义．该方法可以定义 $(-1,\infty)$ 区间的阶乘

$$x!\equiv\Gamma(x+1)=\int_0^{+\infty}t^x\mathrm{e}^{-t}\,\mathrm{d}t\qquad(x>-1)\tag{1}$$

即 Γ 函数在区间 $(0,\infty)$ 被定义为

$$\Gamma(x)\equiv\int_0^{+\infty}t^{x-1}\mathrm{e}^{-t}\,\mathrm{d}t\qquad(x>0)\tag{2}$$

可以证明新定义的阶乘的递推关系仍为

$$(x+1)!=(x+1)x!\qquad(x>-1)\tag{3}$$

且 $0!=1$．所以当 x 取正整数 N 时，式 1 的结果仍然是熟悉的 $N!=N(N-1)\ldots,1$．

另外能证明 $(-1/2)!=\sqrt{\pi}$，由此我们可以直接写出半整数的阶乘为

$$\frac{N}{2}!=\frac{N}{2}\left(\frac{N}{2}-1\right)\ldots\frac{1}{2}\sqrt{\pi}\qquad(N>0)\tag{4}$$

推导

预备知识 分部积分法[190]

首先当 $x\leqslant 0$ 时该积分在 $x=0$ 处不收敛，以下仅讨论 x 为正实数的情况[29]．

我们现在验证当 x 取正整数时，新定义的阶乘 $x!=\Gamma(x+1)$ 与原来的定义 $x!=x(x-1)\ldots1$ 相同．首先

$$0!=\Gamma(1)=\int_0^{+\infty}\mathrm{e}^{-t}\,\mathrm{d}t=0-(-1)=1\tag{5}$$

[28]参考 Wikipedia 相关页面 以及 [7] 相关章节．

[29]事实上，自变量为负实数（非整数）时，Γ 函数有另一种定义，这里不讨论．

使用分部积分法[190]，令 t^x 为“求导项”，e^{-t} 为“积分项”，可得递推公式[30]（式 3）

$$\begin{aligned} x! = \Gamma(x+1) &= \int_0^{+\infty} t^x \mathrm{e}^{-t}\,\mathrm{d}t = -\,t^x \mathrm{e}^{-t}\Big|_0^{+\infty} + \int_0^{+\infty} x t^{x-1}\mathrm{e}^{-t}\,\mathrm{d}t \\ &= x\int_0^{+\infty} t^{x-1}\mathrm{e}^{-t}\,\mathrm{d}t = x\Gamma(x) = x(x-1)! \end{aligned} \tag{6}$$

由递推式 6 和初值式 5，对任意正整数 n 有

$$n! = n(n-1)! = n(n-1)(n-2)!... = n(n-1)...1 \tag{7}$$

再来看半整数的阶乘，我们讨论范围内的最小半整数的阶乘为

$$\left(-\frac{1}{2}\right)! = \int_0^{+\infty} \frac{\mathrm{e}^{-x}}{\sqrt{x}}\,\mathrm{d}x = 2\int_0^{+\infty} \mathrm{e}^{-t^2}\,\mathrm{d}t = \int_{-\infty}^{+\infty} \mathrm{e}^{-t^2}\,\mathrm{d}t = \sqrt{\pi} \tag{8}$$

其中使用了换元法令 $x = t^2$ 将定积分变为高斯积分[339].

渐近公式

对于大的 x, 有**斯特林公式（Stirling formula）**：

$$\Gamma(x+1) = \sqrt{2\pi x}\left(\frac{x}{\mathrm{e}}\right)^x \left(1 + \frac{1}{12x} + \frac{1}{288x^2} - \frac{139}{51840x^3} - \frac{571}{2488320x^4} + \cdots\right).$$

这是一个渐近展开，右边的级数是发散的．它的推导可见拉普拉斯方法[261]

Euler-Mascheroni 常数

- 本词条处于草稿阶段.

Euler-Mascheroni 常数定义为

$$\gamma = \lim_{n\to\infty}\left(\sum_{m=1}^{n}\frac{1}{m} - \ln n\right) = 0.5772156619\ldots \tag{1}$$

其中可以把 $1/m$ 看成区间 $[m, m+1]$ 内高为 $1/m$ 矩形的面积，而 $\ln n$ 是函数 $1/x$ 在区间 $[1, n]$ 的定积分（函数曲线下方的面积）.

[30] 该证明仅对 $x > 0$ 适用，这样才有 $0^x\mathrm{e}^{-0} = 0$，使第四个等号成立.

Gamma 函数 2

极限定义

[31]定义复数域 $\mathbb{C}$ 中的 Γ 函数为

$$\Gamma(z) \equiv \lim_{n\to\infty} \frac{n!}{z(z+1)\dots(z+n)} n^z \qquad (z \neq 0, -1, -2, \dots) \tag{1}$$

该定义包括了 Γ 函数的整个定义域.

积分定义

预备知识 复变函数的积分[357]

Gamma 函数的积分定义式 2 [254] 也可以轻易拓展到复数域上，但同样不包含整个定义域

$$z! \equiv \Gamma(z+1) = \int_0^{+\infty} t^z \mathrm{e}^{-t}\,\mathrm{d}t \qquad (\mathrm{Re}[z] > -1) \tag{2}$$

在 $\mathrm{Re}[z] \leqslant -1$ 的区域，积分不收敛.

Weierstrass 定义

$$\frac{1}{\Gamma(z)} \equiv z\mathrm{e}^{\gamma z} \prod_{n=1}^{\infty} \left(1 + \frac{z}{n}\right) \mathrm{e}^{-z/n} \tag{3}$$

其中 $\gamma = 0.57722\dots$ 是 Euler-Mascheroni 常数[256].

双 Gamma 函数

[31]参考 [7] 相关章节.

- 本词条处于草稿阶段.

[32]定义双 **Gamma 函数（digamma function）**为

$$\psi(x)=\ln[\Gamma(x)]'=\frac{\Gamma'(x)}{\Gamma(x)} \tag{1}$$

渐近展开

预备知识 泰勒级数[181]

定义辨析

在数学中，**渐近展开（asymptotic expansion）**是用一列较简单的函数来逐次逼近给定的函数的办法. 它的形式定义如下：

定义 1　渐近展开

设自变量 x 趋于某点 a (有限或无限) 时，函数序列 $\{\phi_n(x)\}$ 满足

$$\phi_{n+1}(x)=o(\phi_n(x)),\ x\to a.$$

则对于给定的函数 f，称 $f(x)$ 在 $x\to a$ 时有渐近展开式

$$f(x)\simeq c_0\phi_0(x)+c_1\phi_1(x)+...+c_n\phi_n(x)+...,\ x\to a \tag{1}$$

是指：任何给定的 n，皆有

$$f(x)=c_0\phi_0(x)+c_1\phi_1(x)+...+c_n\phi_n(x)+o(\phi_n(x)),\ x\to a. \tag{2}$$

式 1 中的级数称为 $x\to a$ 时 $f(x)$ 的渐近级数.

一般常用的序列是单项式序列或者一般的幂函数序列. 渐近展开的四则运算性质是容易验证的.

有如下注意事项：

[32]参考 Wikipedia 相关页面.

1. 一般来说，从渐近展开不能将函数还原：当 $x \to 0$ 时，函数 $f(x) \equiv 0$ 和函数 $g(x) = \mathrm{e}^{-1/x^2}$ 都有渐近展开

$$0 + 0x + 0x^2 + \ldots$$

2. 渐近级数可以收敛也可以发散．式 1 只是一个形式等式，它的真正含义是式 2，而式 2 只表示对于**固定的** n，当 $x \to a$ 时，$f(x)$ 与渐近级数的第 n 项部分和相差一个高阶无穷小．

3. 通过直接计算，可以看出渐近展开式可以逐项积分．但一般来说渐近展开式不可以逐项微分．例如，$f(x) = e^{-x}\sin(e^{2x})$ 当 $x \to +\infty$ 时趋于零，但它的导数却根本没有极限．

基本例子

例 1　泰勒展开

泰勒展开[181] 就是渐近展开的例子．泰勒级数不必收敛，即便收敛也不必收敛到函数本身．在函数 e^{-1/x^2} 的例子中，原点处的泰勒级数与函数本身的差永远是函数本身，但显然这个差在 $x \to 0$ 时衰减得比 x 的任意正幂都快．

例 2　发散级数部分和

如果 $f(x)$ 在 $x \geq 1$ 时是单调不减函数，那么有

$$\sum_{k=1}^{n} f(k) = \int_1^n f(x)dx + O(f(n)) + O(1),\ n \to \infty.$$

例如，如果取 $f(x) = 1/x$，那么有熟知的公式

$$\sum_{k=1}^{n} \frac{1}{k} = \log n + O(1).$$

实际上当然还可以借助更复杂的分析技巧写得再精确些，例如我们知道 $-\log n + \sum_{k=1}^{n} 1/k$ 的极限是存在的，也就是欧拉常数 $\gamma = 0.57721566\ldots$．更精细的渐近展开式可以由欧拉-麦克劳林公式给出．

欧拉的例子

考察非初等的函数

$$f(x)=\int_0^{\infty}\frac{e^{-t}}{x+t}dt$$

当 $x\to+\infty$ 时的行为．欧拉将 $1/(x+t)$ 展开为几何级数 $\sum_{k=0}^{\infty}(-1)^k t^k/x^{k+1}$，代入并计算得到如下的形式等式：

$$f(x)=\sum_{k=0}^{\infty}\frac{(-1)^k k!}{x^{k+1}}. \tag{3}$$

当然，欧拉的时代还没有收敛性的观念．式 3 右边的级数对于任何 x 都不收敛，之所以出现这样的问题是因为 t 的几何级数收敛半径是有限的，于是将几何级数逐项积分的计算违反分析学的准则．

然而，从今天的观点看，这个等式仍然在渐近展开的意义下成立．实际上，通过换元可得

$$f(x)=e^x\int_x^{\infty}\frac{e^{-t}}{t}dt.$$

反复进行分部积分，得到

$$f(x)=\sum_{k=0}^{n}\frac{(-1)^k k!}{x^{k+1}}+(n+1)!e^x\int_x^{\infty}\frac{e^{-t}}{t^{n+2}}dt.$$

最后的这个积分可估算如下：当 $t>x$ 时 $1/t^{n+2}<1/x^{n+2}$，于是

$$\int_x^{\infty}\frac{e^{-t}}{t^{n+2}}dt\le\frac{1}{x^{n+2}}\int_x^{\infty}e^{-t}dt=\frac{e^{-x}}{x^{n+2}}. \tag{4}$$

这样就有

$$f(x)=\sum_{k=0}^{n}\frac{(-1)^k k!}{x^{k+1}}+O\left(\frac{1}{x^{n+2}}\right),\ x\to+\infty.$$

这表示式 3 在渐近展开的意义下成立．

特别地，这个渐近级数尽管发散，但对于大的 x 却可以很好地计算 $f(x)$ 的值，因为根据估计式 4，$f(x)$ 同渐近级数 n 项部分和之差的绝对值不超过

$$\frac{(n+1)!}{x^{n+2}},$$

而对于大的 x，当 $n<x$ 时这个误差项随着 n 的增大而减小，只有当 n 大致超过 x 时误差才会重新开始增大．因此，如果将渐近级数截断到大约 x 项，则部

分和与 $f(x)$ 之差将不超过 $1/2^x$．对于大的 x 来说，这足以给出相当精确的近似值．

由这个例子可以看出渐近展开的意义：可以用一个发散级数去很好地逼近一个收敛的对象．

拉普拉斯方法

预备知识 极值[170]，积分换元[188]，高斯积分[339]，渐近展开[258]，Gamma 函数[254]

在分析数学中，拉普拉斯方法是一种计算含参数积分的渐近展开式的办法．所考察的积分一般具有如下形式：

$$I(t)=\int_a^b \phi(x)\mathrm{e}^{tf(x)}dx, \tag{1}$$

其中 λ 是正的实参数．要考察当 $t\to+\infty$ 时积分的行为．拉普拉斯方法背后的想法很简单：如果函数 f 在某点处达到极大值，那么当 t 很大时，只有极大值点附近的贡献才比较可观，其余部分相比起来都要小得多．

Watson 引理

拉普拉斯方法基于 Watson 引理．它本身也是很有用的．

引理 1　Watson 引理

设 $\alpha>0$，函数 $\phi(x)$ 在区间 $[0,b]$ 上连续，而且 $x\to 0$ 时有渐近展开

$$\phi(x)\simeq c_0+c_1x+c_2x^2+\ldots$$

则当 $t\to+\infty$ 时，含参数的积分

$$I(t)=\int_0^b \phi(x)\mathrm{e}^{-tx^\alpha}dx$$

有渐近展开式

$$I(t)\simeq\sum_{n=0}^{\infty}\frac{c_n}{\alpha}\Gamma\left(\frac{n+1}{\alpha}\right)t^{-(n+1)/\alpha}.$$

证明是直接的计算．固定一个 n．按照渐近展开的定义，存在 $\delta > 0$ 使得 $0 \leq x < \delta$ 时 $\phi(x) - (c_0 + ... + c_n x^n) = O(x^{n+1})$．远离 0 的 x 对积分的贡献是可以忽略的：

$$\left| \int_\delta^b \phi(x) \mathrm{e}^{-tx^\alpha} dx \right| \leq \max_{x \in [0,b]} |\phi(x)| \mathrm{e}^{-t\delta^\alpha}.$$

而在区间 $[0, \delta]$ 上，通过换元可得

$$\begin{aligned} \int_0^\delta x^k \mathrm{e}^{-tx^\alpha} dx &= \frac{t^{-(k+1)/\alpha}}{\alpha} \int_0^{t\delta^\alpha} y^{-(k+1)/\alpha - 1} \mathrm{e}^{-y} dy \\ &= \frac{t^{-(k+1)/\alpha}}{\alpha} \Gamma\left(\frac{k+1}{\alpha}\right) + O(\mathrm{e}^{-t\delta^\alpha/2}). \end{aligned}$$

于是就有

$$\begin{aligned} I(t) &= \int_0^\delta \phi(x) \mathrm{e}^{-tx^\alpha} dx + O(\mathrm{e}^{-t\delta^\alpha}) \\ &= \sum_{k=0}^n \frac{c_n}{\alpha} \Gamma\left(\frac{k+1}{\alpha}\right) t^{-(k+1)/\alpha} + O(t^{-(n+2)/\alpha}). \end{aligned}$$

这就是所求的结果．

拉普拉斯方法

拉普拉斯方法是将 Watson 引理用于式 1 得到的．作出如下假设：

1. 函数 f, ϕ 都是光滑函数．
2. 函数 $f(x)$ 在积分区间内部仅有一个严格的极大值点 $x_0 \in (a, b)$，而且 x_0 是整个区间上的最大值点．
3. 式 1 对于 $t = t_0$ 绝对收敛．

于是 $f'(x_0) = 0$．不妨设 $f''(x_0) < 0$．从泰勒公式知道，有一 $\delta > 0$ 使得在区间 $(x_0 - \delta, x_0 + \delta)$ 上，

$$f(x) = f(x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + (x - x_0)^3 h(x),$$

其中函数 h 有界．当 t 很大时，区间 $(x_0 - \delta, x_0 + \delta)$ 之外的积分的贡献比 t 的任何负幂衰减得都快：根据假定 2，当 $|x - x_0| > \delta$ 时有一 $c > 0$ 使得

$f(x) < f(x_0) - c$，从而区间 $(x_0 - \delta, x_0 + \delta)$ 之外的积分估计为

$$\left| \int_{|x-x_0|>\delta} \phi(x) \mathrm{e}^{tf(x)} dx \right| \leq \mathrm{e}^{-c(t-t_0)} \int_a^b |\phi(x)| \mathrm{e}^{t_0 f(x)} dx$$
$$= O(\mathrm{e}^{-ct}).$$

而在区间 $(x_0 - \delta, x_0 + \delta)$ 上，可作如下换元：$y^2 = f(x_0) - f(x)$，即

$$y = \sqrt{\frac{-f''(x_0)}{2}} (x - x_0) \left(1 - \frac{2}{f''(x_0)} (x - x_0) h(x) \right)^{1/2}.$$

命 $\beta_\pm = \pm\sqrt{f(x_0) - f(x_0 \pm \delta)}$，又设有泰勒展开

$$\psi(y) := \phi(x(y)) x'(y) \simeq c_0 + c_1 y + ..., y \to 0 \tag{2}$$

（这里显然有 $c_0 = \phi(x_0)\sqrt{-2/f''(x_0)}$），则

$$\int_{x_0-\delta}^{x_0+\delta} \phi(x) \mathrm{e}^{tf(x)} dx = \mathrm{e}^{tf(x_0)} \int_{\beta_-}^{\beta_+} \psi(y) \mathrm{e}^{-ty^2} dy$$
$$\simeq \mathrm{e}^{tf(x_0)} \int_0^c [\psi(y) + \psi(-y)] \mathrm{e}^{-ty^2} dy.$$

应用 Watson 引理，最后终于得到式 1 的渐近展开：

$$I(t) \simeq \mathrm{e}^{tf(x_0)} \sum_{n=0}^{\infty} c_{2n} \Gamma\left(n + \frac{1}{2}\right) t^{-n-1/2}, \tag{3}$$

其中 c_n 由式 2 给出．特别地，由于 $\Gamma(1/2) = \sqrt{\pi}$，式 1 渐近公式的首项是

$$I(t) = \mathrm{e}^{tf(x_0)} \left(\phi(x_0) \sqrt{-\frac{2\pi}{tf''(x_0)}} + O(t^{-3/2}) \right).$$

斯特林公式

拉普拉斯方法最基本的应用就是导出 Γ 函数的**斯特林公式（Stirling formula）**．按照定义，

$$\Gamma(t+1) = \int_0^\infty \mathrm{e}^{-x} x^t dx = t^{t+1} \int_0^\infty \mathrm{e}^{-t(x - \log x)} dx.$$

最后一步用到换元 $x \to tx$．最后这个积分刚好符合上一小节所要求的三条假设：这里 $\phi(x) \equiv 1$，$f(x) = -(x - \log x)$．于是 $x = 1$ 是 $f(x)$ 唯一的极大值点，同时也是唯一的最大值点．代入式 3，就得到渐近公式

$$\Gamma(t+1) \simeq \sqrt{2\pi t}\left(\frac{t}{\mathrm{e}}\right)^t.$$

式 3 还给出了更精细的渐近级数：

$$\Gamma(t+1) = \sqrt{2\pi t}\left(\frac{t}{\mathrm{e}}\right)^t\left(1 + \frac{1}{12t} + \frac{1}{288t^2} - \frac{139}{51840t^3} - \frac{571}{2488320t^4} + \cdots\right).$$

毫无疑问这是一个可以对大的 n 来近似计算 $n!$ 的公式．

第三章

多元微积分与矢量分析

偏导数

预备知识 导数[158]

对一个多元函数 $y=f(x_1,x_2\dots x_i\dots)$，如果求导时只把 x_i 看成自变量，剩下的 $x_{j\neq i}$ 都看做常数，得到的导数就叫函数（关于 x_i）的**偏导数**. 以二元函数 $z=f(x,y)$ 为例，对 x 的偏导数常记为

$$\frac{\partial z}{\partial x}\qquad \frac{\partial f}{\partial x}\qquad f_x\qquad \left(\frac{\partial f}{\partial x}\right)_y \tag{1}$$

最后一种记号在括号右下角声明了保持不变的自变量，这在许多情况下能避免混淆.

例 1

对于函数 $f(x,y)=x^2+2y^2+2xy$，两个偏导数分别为

$$\frac{\partial f}{\partial x}=2x+2y\qquad \frac{\partial f}{\partial y}=4y+2x \tag{2}$$

例 2

对于函数 $z=\sin(y\cos x)+\cos^2 x$

$$\frac{\partial z}{\partial x}=-y\cos(y\cos x)\sin x-2\cos x\sin x=-y\cos(y\cos x)\sin x-\sin 2x \tag{3}$$

$$\frac{\partial z}{\partial y}=\cos(y\cos x)\cos x \tag{4}$$

几何意义

类比导数的几何意义（曲线的斜率），若在三维直角坐标系中画出曲面 $f(x,y)$，则 $\partial f/\partial x$ 和 $\partial f/\partial y$ 分别是是某点处曲面延 x 方向和 y 方向的斜率. 所以从某点 (x_0,y_0) 延 x 方向移动一个微小量 Δx，假设曲面平滑，则函数值增加

$$\Delta f\approx\frac{\partial f}{\partial x}\Delta x \tag{5}$$

写成微分关系就是

$$\mathrm{d}f=\frac{\partial f}{\partial x}\,\mathrm{d}x\quad (y\text{ 不变}) \tag{6}$$

高阶偏导

与一元函数的高阶导数类似，多元函数也可以求高阶偏导数，不同的是，由于每求一次偏导都需要指定对哪个变量．例如二元函数 $f(x,y)$ 的二阶偏导有

$$\frac{\partial^2 f}{\partial x^2} \qquad \frac{\partial^2 f}{\partial x \partial y} \qquad \frac{\partial^2 f}{\partial y \partial x} \qquad \frac{\partial^2 f}{\partial y^2} \tag{7}$$

若高阶偏导的分母中出现不止一个变量，我们就称其为**混合偏导**．混合偏导的一个重要性质就是偏导的顺序可以任意改变，例如上式中有 $\partial^2 f/\partial x \partial y = \partial^2 f/\partial y \partial x$．这点本书不做证明，可以通过以上的例子验证．

最小二乘法

预备知识 二元函数的极值[281]

直线拟合

在许多情况下，我们需要将一组散点数据 $x_i, y_i \ \ (i = 1 \ldots N)$ 拟合成特定形式的函数曲线．其中最常见的情况之一是拟合一条直线，形式为 $y = ax + b$．例如给出（图 1），如何确定直线方程的两个最佳系数 a 和 b 呢？

我们可以通过方差来计算拟合的误差，方差越小则说明拟合得越好．

$$S_2 = \sum_{i=1}^{N} (ax_i + b - y_i)^2 \tag{1}$$

对于给定的 N 个 x_i, y_i，方差是 a 和 b 的二元函数．我们只需找到这个二元函数的最小值点即可．由于函数处处光滑，最小值点必定满足 $\partial S_2/\partial a = \partial S_2/\partial b = 0$．将式 1 代入，得到一个线性方程组．

$$\begin{cases} \left(\sum_i x_i^2\right) a + \left(\sum_i x_i\right) b = \sum_i x_i y_i \\ \left(\sum_i x_i\right) a + Nb = \sum_i y_i \end{cases} \tag{2}$$

我们来计算系数行列式，如果能证明行列式恒大于零，则方程组必有唯一解．若令 $\bar{x}$ 为 N 个 x_i 的平均值，考虑到 x_i 互不相等，容易证明

$$N\sum_i x_i^2-\left(\sum_i x_i\right)^2=N\sum_i(x_i-\bar{x})^2>0 \tag{3}$$

证毕．由于方差恒大于零，可知其必定存在最小值，所以系数的唯一解必定是方差函数的极小值点．接下来解线性方程组即可得到系数 a,b．将解出的 a,b 重新代入式 1 可以得出最小方差的值，用于判断拟合结果的好坏．

多项式拟合

以二次多项式 $c_2x^2+c_1x+c_0$ 为例，方差为

$$S_2=\sum_{i=1}^{N}(c_2x_i^2+c_1x_i+c_0-y_i)^2 \tag{4}$$

分别令方差对 c_2,c_1,c_0 的偏导为零，得线性方程组

$$\begin{cases}\left(\sum_i x_i^4\right)c_2+\left(\sum_i x_i^3\right)c_1+\left(\sum_i x_i^2\right)c_0 &=\sum_i y_ix_i^2\\ \left(\sum_i x_i^3\right)c_2+\left(\sum_i x_i^2\right)c_1+\left(\sum_i x_i\right)c_0 &=\sum_i y_ix_i\\ \left(\sum_i x_i^2\right)c_2+\left(\sum_i x_i\right)c_1+\quad Nc_0 &=\sum_i y_i\end{cases} \tag{5}$$

观察系数矩阵可以看出，每条斜线上 x_i 的指数相等，且相邻斜线上 x_i 的指数依次递减．按照此规律容易写出 N 次多项式拟合的方程组．这里不证明系数行列式大于零，但实际应用中方程组几乎都有唯一解．

简谐波拟合

若要拟合 $A\cos(x+\varphi_0)+C$ 形式的函数，可以先利用两角和公式把函数化为 $c_1\cos x+c_2\sin x+c_3$ 的等效形式（因为前者并不是待定系数的线性组合，得到的方程组也不是线性方程组），方差公式为

$$S_2=\sum_i(c_1\cos x_i+c_2\sin x_i+c_3-y_i)^2 \tag{6}$$

分别令方差对 c_1, c_2, c_3 的偏导为零，得线性方程组

$$\begin{cases} \left(\sum_i \cos^2 x_i\right) c_1 + \left(\sum_i \sin x_i \cos x_i\right) c_2 + \left(\sum_i \cos x_i\right) c_3 = \sum_i y_i \cos x_i \\ \left(\sum_i \sin x_i \cos x_i\right) c_1 + \left(\sum_i \sin^2 x_i\right) c_2 + \left(\sum_i \sin x_i\right) c_3 = \sum_i y_i \sin x_i \\ \left(\sum_i \cos x_i\right) c_1 + \left(\sum_i \sin x_i\right) c_2 + N c_3 = \sum_i y_i \end{cases} \tag{7}$$

实际应用中，该方程组同样几乎都有唯一解.

全微分

预备知识 偏导数[266]

以二元函数为例，在偏微分的几何意义中，若 $z = f(x, y)$ 在某点 (x_0, y_0) 附近的曲面光滑[1]，那么如果考虑一个足够小的区域，可以把曲面近似为平面. 设平面方程为

$$z = c_0 + c_x(x - x_0) + c_y(y - y_0) \tag{1}$$

当 $x = x_0$，$y = y_0$ 时显然有 $c_0 = f(x_0, y_0)$，求两个偏导，又有

$$c_x = \frac{\partial f}{\partial x} \qquad c_y = \frac{\partial f}{\partial y} \tag{2}$$

令坐标增量为 $\Delta x \equiv x - x_0$，$\Delta y \equiv y - y_0$，$\Delta z \equiv z - c_0$，则平面方程变为

$$\Delta z = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y \tag{3}$$

令增量为无穷小，即

$$\mathrm{d}z = \frac{\partial f}{\partial x}\,\mathrm{d}x + \frac{\partial f}{\partial y}\,\mathrm{d}y \tag{4}$$

这就是**全微分**关系. 全微分的意义是，从某一点开始向任意方向移动 $(\mathrm{d}x\,, \mathrm{d}y)$，函数的增量等于只向 x 方向移动 $\mathrm{d}x$ 的增量加上只向 y 方向移动 $\mathrm{d}y$ 的增量. 类

[1] 光滑的数学定义是，各一阶偏导数在区域内连续.

似地，N 元函数的全微分关系为

$$\mathrm{d}z=\sum_{i=1}^{N}\frac{\partial f}{\partial x_i}\,\mathrm{d}x_i \tag{5}$$

事实上，偏微分也可以理解为是由该式定义的.

全微分近似

类比一元函数的微分近似[173] $\Delta y\approx \mathrm{d}f/\mathrm{d}x\cdot\Delta x$，若 N 元函数各个变量的一阶偏导在一小块区域内变化不大，那么函数值的变化可近似为

$$\begin{aligned}\Delta z&=f(x_1+\Delta x_1,\ldots,x_N+\Delta x_N)-f(x_1,\ldots,x_N)\\&\approx\frac{\partial f}{\partial x_1}\Delta x_1+\ldots+\frac{\partial f}{\partial x_N}\Delta x_N\end{aligned} \tag{6}$$

例 1　测量误差

测量一个边长各不相同的长方体的体积，若三边的测量值和最大测量误差分别为 $a,\sigma_a,b,\sigma_b,c,\sigma_c$（假设不确定度远小于边长），求体积的最大测量误差 σ_V 及最大相对误差 σ_V/V.

类比“一元函数微分”中的例 1，长方体的体积为 $V(a,b,c)=abc$，由全微分近似得

$$\sigma_V\approx\frac{\partial V}{\partial a}\sigma_a+\frac{\partial V}{\partial b}\sigma_b+\frac{\partial V}{\partial c}\sigma_c=bc\sigma_a+ac\sigma_b+ab\sigma_c \tag{7}$$

相对不确定度为

$$\frac{\sigma_V}{V}\approx\frac{\sigma_a}{a}+\frac{\sigma_b}{b}+\frac{\sigma_c}{c} \tag{8}$$

复合函数的偏导 链式法则

预备知识　全微分[269]，复合函数求导[174]

先来看一个二元函数的例子：若已知二元函数 $z=f(u,v)$，z 是 u,v 的函数，但若 u 和 v 都又是 x 和 y 的函数，则 z 最终是 x 和 y 的函数，即

$$z(x,y)=f[u(x,y),v(x,y)] \tag{1}$$

那如何求 $z(x,y)$ 的偏导数呢？我们先来看全微分关系[269]

$$\mathrm{d}z = \frac{\partial f}{\partial u}\,\mathrm{d}u + \frac{\partial f}{\partial v}\,\mathrm{d}v \tag{2}$$

而 u 和 v 作为 x,y 的函数，它们的微小变化又都是由 x 和 y 的微小变化引起的，所以有全微分

$$\mathrm{d}u = \frac{\partial u}{\partial x}\,\mathrm{d}x + \frac{\partial u}{\partial y}\,\mathrm{d}y \quad \mathrm{d}v = \frac{\partial v}{\partial x}\,\mathrm{d}x + \frac{\partial v}{\partial y}\,\mathrm{d}y \tag{3}$$

代入上式得

$$\begin{aligned}\mathrm{d}z &= \frac{\partial f}{\partial u}\left(\frac{\partial u}{\partial x}\,\mathrm{d}x + \frac{\partial u}{\partial y}\,\mathrm{d}y\right) + \frac{\partial f}{\partial v}\left(\frac{\partial v}{\partial x}\,\mathrm{d}x + \frac{\partial v}{\partial y}\,\mathrm{d}y\right)\\ &= \left(\frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x}\right)\mathrm{d}x + \left(\frac{\partial f}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial y}\right)\mathrm{d}y\end{aligned} \tag{4}$$

这就是 z 关于 x 和 y 的全微分关系．根据偏导数的定义

$$\frac{\partial z}{\partial x} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x} \tag{5}$$

$$\frac{\partial z}{\partial y} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial y} \tag{6}$$

这也叫偏导的**链式法则**．为了方便我们也会把 $\partial z/\partial x$ 和 $\partial z/\partial y$ 分别记为 $\partial f/\partial x$ 和 $\partial f/\partial y$，但有可能产生歧义．

我们可以把以上的例子拓展到任意多个变量的情况，即令

$$z(x_1,\ldots,x_N) = f[u_1(x_1,\ldots,x_N),\ldots,u_M(x_1,\ldots,x_N)] \tag{7}$$

这时链式法则可以记为

$$\frac{\partial z}{\partial x_i} = \sum_j \frac{\partial f}{\partial u_j}\frac{\partial u_j}{\partial x_i} \tag{8}$$

例 1

令 $u(x,y)=x+y$，$v(x,y)=x-y$，$f(u,v)=u^2+v^2$，求 $\partial f/\partial x$ 和 $\partial f/\partial y$．

解：直接套用式 1 得

$$\frac{\partial f}{\partial x} = 2u\cdot 1 + 2v\cdot 1 = 4x \tag{9}$$

$$\frac{\partial f}{\partial y} = 2u\cdot 1 + 2v\cdot(-1) = 4y \tag{10}$$

为了验证，我们也可以直接写出

$$f(x,y)=(x+y)^2+(x-y)^2=2x^2+2y^2 \tag{11}$$

再直接求偏导得 $\partial f/\partial x=4x$，$\partial f/\partial y=4y$，可见结果也是相同的.

通用函数名

物理中常常会出现一种容易混淆的情况，就是当一个因变量可以有几套自变量（例如上面的 $z(u,v)$ 和 $z(x,y)$）时，通常直接用因变量（z）作为函数名而另外不定义函数名（f）. 然而 $z(u,v)$ 与 $z(x,y)$ 中的 z 并不是同一个函数. 以下举例说明

例 2

在二维直角坐标系中，定义势能函数为

$$V=f(x,y)=x^2+y^2+2x \tag{12}$$

而若用极坐标[119] 描述该势能，则函数变为

$$V=g(r,\theta)=f(r\cos\theta,r\sin\theta)=r^2+2r\cos\theta \tag{13}$$

但许多物理书为了表述方便并不用 f 和 g 区分两个不同的函数，而是使用 $V(x,y)$ 表示式 12 和 $V(r,\theta)$ 表示式 13. 这样后者就有可能被误解为

$$V(r,\theta)=r^2+\theta^2+2r\quad\text{(错)} \tag{14}$$

这就需要从语境中判断是否使用了通用函数名[2].

使用通用函数名时，要注意判断偏导数使用的是哪一套变量，例如 $\partial V/\partial x$ 默认使用 $V(x,y)$ 求偏导，即把 y 看成常数；$\partial V/\partial r$ 默认使用 $V(r,\theta)$ 求偏导，即把 θ 看成常数. 一种更复杂的情况如 $(\partial V/\partial x)_\theta$. 按照定义[3]，应该是仅用 x 和 θ 表示 V，然后求偏导. 考虑极坐标的定义，θ 不变意味着 y 与 x 成正比即 $y=x\tan\theta$，代入式 12 得

$$V(x,\theta)=x^2(1+\tan^2\theta)+2x \tag{15}$$

现在再对 x 求偏导即可（略）.

[2] “通用函数名”是笔者起的名字，不清楚是否有其他叫法.

[3] 见偏导数[266] 中的式 1.

显含

在物理中，尤其在分析力学中，我们通常会遇见**显含（explicitly depends on）**的概念．当

$$f(u_1,\dots,u_M) \tag{16}$$

中的某个 u_i 满足 $\partial f/\partial u_i$ 不恒为零时，我们说 f **显含** u_i．特殊地，当某个 $u_i(x_1,\dots,x_N)=x_j$ 时，我们就说函数 f **显含** x_j，否则就说 f **不显含** x_j．

若 f 不显含 x_j，但某个 u_i 满足 $\partial u_i/\partial x_j$ 不恒为零，我们就说 f **隐含（implicitly depends on）** x_j．

例 3

一个质点延着一条静止轨道 $y=0$ 运动，它的动能[904] 为 $E(v_x)=mv_x^2$．虽然速度 v_x 是时间 t 的函数，但我们说 $E(v_x)$ 不显含时间．

若轨道在 y 方向具有随时间变化的速度 $v_y(t)=at^2$，如果把动能记为 $E(v_x,v_y)=m(v_x^2+v_y^2)$，那么该函数仍然不显含 t．

仍然令 $v_y(t)=at^2$，但把动能记为 $E(v_x,t)=mv_x^2+m(at^2)^2$，那么该函数就显含 t．

虽然在后两种情况中物理情景都是一样的，但 $E(v_x,v_y)$ 和 $E(v_x,t)$ 在数学上却是两个不同的函数，使用了通用函数名 E．这两个函数分别是 $f(x,y)=m(x^2+y^2)$ 和 $g(x,y)=mx^2+ma^2y^4$，显然不是同一个函数．它们都显含 x,y，不显含其他任何变量．

全导数

特殊地，当所有的 u_i 都只是一个变量 x 的函数时，我们可以求全导数[273] $\mathrm{d}f/\mathrm{d}x$．这时可以分为 f 隐含 x 和显含 x 来讨论．

全导数

- 需要修改，使用“复合函数的偏导[270]”作为预备知识

预备知识 全微分[269]

若多元函数包含若干个变量（以下以 $f(x,y,t)$ 为例），我们知道它可以对其中任意一个变量求偏导，即 $\partial f/\partial x$，$\partial f/\partial y$，$\partial f/\partial t$，注意求偏导时其余两个变量不变．现在若把 x,y 看做 t 的函数，那么 f 归根结底也是 t 的函数 $f[x(t),y(t),t]$，我们可以将其对 t 求导．为了强调这与对 t 求偏导有所不同，我们把得到的函数叫做**全导数**.

与偏微分中的链式法的推导类似，我们先来看函数的全微分[269]

$$\mathrm{d}f=\frac{\partial f}{\partial x}\,\mathrm{d}x+\frac{\partial f}{\partial y}\,\mathrm{d}y+\frac{\partial f}{\partial t}\,\mathrm{d}t \tag{1}$$

而根据 x,y 与 t 的微分关系

$$\mathrm{d}x=\frac{\mathrm{d}x}{\mathrm{d}t}\,\mathrm{d}t\qquad \mathrm{d}y=\frac{\mathrm{d}y}{\mathrm{d}t}\,\mathrm{d}t \tag{2}$$

代入上式得

$$\mathrm{d}f=\left(\frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t}+\frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}+\frac{\partial f}{\partial t}\right)\mathrm{d}t \tag{3}$$

而若把 f 看做 t 的一元函数，又应该有全微分关系

$$\mathrm{d}f=\frac{\mathrm{d}f}{\mathrm{d}t}\,\mathrm{d}t \tag{4}$$

对比以上两式可得 f 关于 t 的全导数为

$$\frac{\mathrm{d}f}{\mathrm{d}t}=\frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t}+\frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}+\frac{\partial f}{\partial t} \tag{5}$$

矢量的导数 求导法则

预备知识 几何矢量的运算[375]，导数[158]，矢量函数

矢量的导数

若几何矢量 $\mathbf{v}$ 只是一个标量 t 的函数，记为 $\mathbf{v}(t)$，则 $\mathbf{v}$ 对 t 的导数可记为以下的一种

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}\qquad\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v}\qquad\dot{\mathbf{v}} \tag{1}$$

其定义为（类比式 2 [158]）

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}=\lim_{\Delta t\to 0}\frac{\mathbf{v}(t+\Delta t)-\mathbf{v}(t)}{\Delta t} \tag{2}$$

唯一与实函数（$f:\mathbb{R}\to\mathbb{R}$）[103] 的导数不同的是，这里的减法是**矢量相减**，结果还是矢量. 除以 Δt 相当于矢量的数乘[370] $1/\Delta t$，结果也是矢量. 所以 $\mathrm{d}\mathbf{v}/\mathrm{d}t$ 也是一个矢量关于标量 t 的函数.

从直角坐标的角度来看，N 维矢量可以用 N 个实数表示，而两个矢量相减则是它们的各个坐标分别相减，易得式 2 得到的矢量导函数的各个分量等于原矢量函数的各个分量分别求导（详见式 4 ）.

例：速度和加速度（矢量）[881]，匀速圆周运动的速度[883] 和加速度[885].

矢量的求导法则

与标量函数一样，由定义不难证明矢量函数求导也是线性算符（c_i 为常数）[4]

$$\frac{\mathrm{d}}{\mathrm{d}t}[c_1\mathbf{v}_1(t)+c_2\mathbf{v}_2(t)+\dots]=c_1\frac{\mathrm{d}\mathbf{v}_1}{\mathrm{d}t}+c_2\frac{\mathrm{d}\mathbf{v}_2}{\mathrm{d}t}\dots \tag{3}$$

直角坐标中，矢量函数可以看做三个分量上的标量函数且矢量基底不变，所以由上式可得矢量求导就是对每个标量函数求导.

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}=\frac{\mathrm{d}}{\mathrm{d}t}[v_x(t)\hat{\mathbf{x}}]+\frac{\mathrm{d}}{\mathrm{d}t}[v_y(t)\hat{\mathbf{y}}]+\frac{\mathrm{d}}{\mathrm{d}t}[v_z(t)\hat{\mathbf{z}}]=\dot{v}_x(t)\hat{\mathbf{x}}+\dot{v}_y(t)\hat{\mathbf{y}}+\dot{v}_z(t)\hat{\mathbf{z}} \tag{4}$$

要特别注意该式成立的条件是三个基底不随 t 改变，这在其他坐标系中并不成立，例如“极坐标中单位矢量的偏导[287]”.

例：匀速圆周运动的速度[883] 和加速度[885]（求导法）.

矢量数乘，内积或叉乘的求导在形式上都与标量函数的情况类似.

$$\frac{\mathrm{d}}{\mathrm{d}t}[f(t)\mathbf{v}(t)]=\frac{\mathrm{d}f}{\mathrm{d}t}\mathbf{v}+f\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} \tag{5}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathbf{u}(t)\cdot\mathbf{v}(t)]=\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t}\cdot\mathbf{v}+\mathbf{u}\cdot\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} \tag{6}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathbf{u}(t)\times\mathbf{v}(t)]=\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t}\times\mathbf{v}+\mathbf{u}\times\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} \tag{7}$$

[4] 以下法则虽然以导数为例，但对偏导也同样适用.

由定义出发，不难证明以上三式，这里以式 6 为例进行证明．根据内积定义以及标量函数的求导法则[160] 有

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{u}\cdot\mathbf{v}) &= \frac{\mathrm{d}}{\mathrm{d}t}(u_xv_x+u_yv_y+u_zv_z)\\ &= \left(\frac{\mathrm{d}u_x}{\mathrm{d}t}v_x+u_x\frac{\mathrm{d}v_x}{\mathrm{d}t}\right)+\left(\frac{\mathrm{d}u_y}{\mathrm{d}t}v_y+u_y\frac{\mathrm{d}v_y}{\mathrm{d}t}\right)+\left(\frac{\mathrm{d}u_z}{\mathrm{d}t}v_z+u_z\frac{\mathrm{d}v_z}{\mathrm{d}t}\right)\\ &= \left(\frac{\mathrm{d}u_x}{\mathrm{d}t}v_x+\frac{\mathrm{d}u_y}{\mathrm{d}t}v_y+\frac{\mathrm{d}u_z}{\mathrm{d}t}v_z\right)+\left(u_x\frac{\mathrm{d}v_x}{\mathrm{d}t}+u_y\frac{\mathrm{d}v_y}{\mathrm{d}t}+u_z\frac{\mathrm{d}v_z}{\mathrm{d}t}\right)\\ &= \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t}\cdot\mathbf{v}+\mathbf{u}\cdot\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}\end{aligned} \tag{8}$$

例：动量定理[944]，角动量定理（单个质点）[911]．

矢量的高阶导数

与标量函数的高阶导数类似，对某个矢量连续求 N 次导数，就得到该函数的 N 阶导数．上面在求圆周运动的加速度时，事实上我们已经计算了位置矢量的导数（速度）的导数，即位置矢量关于时间的二阶导数．

矢量的偏导数

- 本词条处于草稿阶段．

预备知识 矢量的导数求导法则[274]，偏导数[266]

与标量函数的偏导类似，对一个多元的矢量函数 $\mathbf{v}(x_1,x_2\dots x_N)$，如果把其他自变量都看做常数而对 x_i 求导，那么就得到矢量函数关于 x_i 的**偏导数**．

$$\frac{\partial\mathbf{v}}{\partial x_i}=\lim_{\Delta x_i\to0}\frac{\mathbf{v}(x_1\dots x_i+\Delta x_i\dots x_N)-\mathbf{v}(x_1\dots x_i\dots x_N)}{\Delta x_i} \tag{1}$$

直角坐标系中对多变量矢量函数求偏导就是对矢量的各个分量分别求偏导（见下文“求导法则”）．

未完成：高阶混合偏导．同样与顺序无关．

与标量函数的偏导[266] 类似，多元矢量函数的高阶导数也要声明各阶导数是对哪个变量进行的．

偏微分算符

（未完成）

预备知识 算符[210]

例如

$$\left(\frac{\partial}{\partial x}+1\right)^2=\frac{\partial^2}{\partial x^2}+2\frac{\partial}{\partial x}+1 \tag{1}$$

$$\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^2=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+2\frac{\partial^2}{\partial x\partial y} \tag{2}$$

一元矢量函数的积分

预备知识 矢量的导数[274]

单变量不定积分

令 $\mathbf{f}(t)$ 为只有一个自变量的矢量函数，则与标量函数类似，定义其不定积分为求导的逆运算．也就是说，若能找到 $\mathbf{F}(t)$，使得 $\mathbf{F}(t)$ 对 t 求导就是 $\mathbf{f}(t)$，那么 $\mathbf{F}(t)+\mathbf{C}$（$\mathbf{C}$ 为任意常矢量）就是定积分的结果，都是 $\mathbf{f}(t)$ 的原函数．

在直角坐标系中，我们已经知道对矢量函数 $\mathbf{F}(t)$ 求导就是对它的每个分量函数分别求导，即

$$\mathbf{F}'(t)=\mathbf{f}(t) \tag{1}$$

$$F_x'(t)=f_x(t)\qquad F_y'(t)=f_y(t)\qquad F_z'(t)=f_z(t) \tag{2}$$

考虑到标量函数的不定积分是标量函数求导的逆运算，所以对 $\mathbf{f}(t)$ 不定积分，只需对它的各个分量分别进行不定积分即可．注意每个分量函数在不定积分后都会出现一个待定常数，三个分量中的待定常数相加就得到一个待定常矢量

C.

$$\begin{aligned}\int \mathbf{f}(t)\,\mathrm{d}t &= \hat{\mathbf{x}}\int f_x(t)\,\mathrm{d}t + \hat{\mathbf{y}}\int f_y(t)\,\mathrm{d}t + \hat{\mathbf{z}}\int f_z(t)\,\mathrm{d}t \\ &= [F_x(t)+C_x]\hat{\mathbf{x}} + [F_y(t)+C_y]\hat{\mathbf{y}} + [F_z(t)+C_z]\hat{\mathbf{z}} \\ &= \mathbf{F}(t)+\mathbf{C}\end{aligned} \tag{3}$$

根据式 1 式 2，显然有 $[\mathbf{F}(t)+\mathbf{C}]' = \mathbf{f}(t)$.

单变量定积分

类比一元标量函数定积分[198] 的定义，要计算一元矢量函数 $\mathbf{f}(t)$ 从 t_1 到 t_2 的定积分，就先把区间 $[t_1, t_2]$ 分为 N 个小区间，长度分别为 Δt_i，且令 t_i 为第 i 个区间内的任意一点. 当我们取极限令所有区间长度 Δt_i 都趋近于 0 （这时 $N\to\infty$）时，如果以下极限存在，得到的矢量就是定积分的结果.

$$\int_{t_1}^{t_2}\mathbf{f}(t)\,\mathrm{d}t = \lim_{\Delta t_i\to 0}\sum_{i=0}^{N}\mathbf{f}(t_i)\Delta t_i \tag{4}$$

唯一与标量函数的定积分不同的是，这里的求和是矢量求和. 但在直角坐标系中，我们可以把上式对矢量的求和表示成对各个分量分别求和，而每个分量的极限就是一个标量定积分.

$$\begin{aligned}\int_{t_1}^{t_2}\mathbf{f}(t)\,\mathrm{d}t &= \hat{\mathbf{x}}\lim_{\Delta t_i\to 0}\sum_{i=0}^{N} f_x(t_i)\Delta t_i + \hat{\mathbf{y}}\lim_{\Delta t_i\to 0}\sum_{i=0}^{N} f_y(t_i)\Delta t_i + \hat{\mathbf{y}}\lim_{\Delta t_i\to 0}\sum_{i=0}^{N} f_z(t_i)\Delta t_i \\ &= \hat{\mathbf{x}}\int_{t_1}^{t_2} f_x(t)\,\mathrm{d}t + \hat{\mathbf{y}}\int_{t_1}^{t_2} f_y(t)\,\mathrm{d}t + \hat{\mathbf{z}}\int_{t_1}^{t_2} f_z(t)\,\mathrm{d}t\end{aligned} \tag{5}$$

所以 $\mathbf{f}(t)$ 的定积分就是把直角坐标的各个分量分别进行定积分. 现在对三个定积分分别运用牛顿—莱布尼兹公式[202]，$\mathbf{f}(t)$ 的原函数为 $\mathbf{F}(t)$，各分量的原函数为 $F_x(t), F_y(t), F_z(t)$，则上式等于

$$\begin{aligned}\int_{t_1}^{t_2}\mathbf{f}(t)\,\mathrm{d}t &= \hat{\mathbf{x}}[F_x(t_2)-F_x(t_1)] + \hat{\mathbf{y}}[F_y(t_2)-F_y(t_1)] + \hat{\mathbf{z}}[F_z(t_2)-F_z(t_1)] \\ &= \mathbf{F}(t_2)-\mathbf{F}(t_1)\end{aligned} \tag{6}$$

这就是矢量函数的牛顿—莱布尼兹公式.

例 1　加速度，速度和位移的积分关系

由于质点的速度—时间函数 $\mathbf{v}(t)$ 是位移—时间函数 $\mathbf{r}(t)$ 的导函数，后者就是前者的原函数．所以根据牛顿—莱布尼兹公式式 6 有

$$\mathbf{r}(t)-\mathbf{r}(t_0)=\int_{t_1}^{t_2}\mathbf{v}(t)\,\mathrm{d}t \tag{7}$$

即

$$\mathbf{r}(t)=\mathbf{r}(t_0)+\int_{t_1}^{t_2}\mathbf{v}(t)\,\mathrm{d}t \tag{8}$$

这是一维情况式 7 [879] 的拓展．

同理，由于质点的加速度函数 $\mathbf{a}(t)$ 是速度函数 $\mathbf{v}(t)$ 的导函数，后者可以通过前者定积分得到

$$\mathbf{v}(t)=\mathbf{v}(t_0)+\int_{t_1}^{t_2}\mathbf{a}(t)\,\mathrm{d}t \tag{9}$$

应用举例　匀加速运动[889]

方向导数

预备知识　全微分[269]，正交归一基[386]

先来看一幅等高线图（图 1）．令高度 z 为位置的函数 $z=f(\mathbf{r})$．这里 $\mathbf{r}$ 是二维平面上的位矢[878]，即 $f(\mathbf{r})=f(x,y)$．当位矢沿着等高线移动时，z 不变，而当位矢沿垂直于等高线的方向移动时，z 变化得最快．位置沿其他方向运动，z 的变化速度介于两者之间．

那么如何衡量位置向各个方向移动时 z 变化的快慢呢？我们先规定一个方向 $\hat{\mathbf{n}}=(n_x,n_y)$（平面单位矢量，满足 $n_x^2+n_y^2=1$），然后用**方向导数**来衡量变化率，其定义如下

$$\frac{\mathrm{d}f}{\mathrm{d}n}\equiv\lim_{\Delta s\to 0}\frac{f(\mathbf{r}+\hat{\mathbf{n}}\Delta s)-f(\mathbf{r})}{\Delta s}=\frac{\mathrm{d}f}{\mathrm{d}s} \tag{1}$$

其中 $\hat{\mathbf{n}}\Delta s$ 代表沿 $\hat{\mathbf{n}}$ 方向的微小位移．从几何上来讲，二维函数 $f(\mathbf{r})$ 表示一个曲面，曲面上某点的方向导数就是曲面在该方向的斜率．

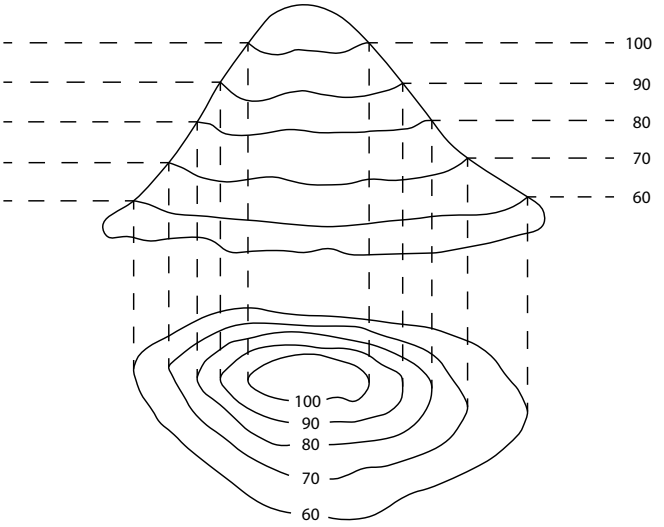


图 1: 等高线

由“全微分[269]”中的结论

$$\mathrm{d}f = \frac{\partial f}{\partial x}\,\mathrm{d}x + \frac{\partial f}{\partial y}\,\mathrm{d}y \tag{2}$$

而现在我们往 $\hat{\mathbf{n}} = (n_x, n_y)$ 方向移动 $\mathrm{d}s$，所以

$$\mathrm{d}x = n_x\,\mathrm{d}s \qquad \mathrm{d}y = n_y\,\mathrm{d}s \tag{3}$$

代入上式，得

$$\mathrm{d}f = \left(\frac{\partial f}{\partial x}n_x + \frac{\partial f}{\partial y}n_y\right)\mathrm{d}s \tag{4}$$

根据导数与微分的关系（也可以通俗地说“两边同除 $\mathrm{d}s$”），就得到方向导数

$$\frac{\partial f}{\partial n} = \frac{\mathrm{d}f}{\mathrm{d}s} = \frac{\partial f}{\partial x}n_x + \frac{\partial f}{\partial y}n_y \tag{5}$$

如果使用平面的正交归一基[386] $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ 写成矢量内积[383] 的形式，就是

$$\frac{\mathrm{d}f}{\mathrm{d}n} = \left(\frac{\partial f}{\partial x}\hat{\mathbf{x}} + \frac{\partial f}{\partial y}\hat{\mathbf{y}}\right)\cdot\hat{\mathbf{n}} \tag{6}$$

定义**二维直角坐标系中的 Del 算符**为

$$\boldsymbol{\nabla} = \hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} \tag{7}$$

其作用在函数上表示

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial x}\hat{\mathbf{x}} + \frac{\partial f}{\partial y}\hat{\mathbf{y}} \tag{8}$$

则方向导数可以写成相当简洁的形式，即

$$\frac{\partial f}{\partial n} = \boldsymbol{\nabla} f \cdot \hat{\mathbf{n}} \tag{9}$$

多元函数的方向导数

通过和以上类似的分析，可以得出 N 元函数 $f(\mathbf{r}) = f(x_1, x_2 \dots x_N)$ 在单位方向矢量 $\hat{\mathbf{n}} = (n_{x1}, n_{x2} \dots n_{x_N})$ 的方向上的微分关系为

$$\mathrm{d}f = \frac{\partial f}{\partial x_1}\mathrm{d}x_1 + \frac{\partial f}{\partial x_2}\mathrm{d}x_2 \dots = \left(\frac{\partial f}{\partial x_1}n_{x1} + \frac{\partial f}{\partial x_2}n_{x2} \dots\right)\mathrm{d}s \tag{10}$$

方向导数为

$$\frac{\partial f}{\partial n} = \frac{\mathrm{d}f}{\mathrm{d}s} = \left(\frac{\partial f}{\partial x_1}\hat{\mathbf{x}}_1 + \frac{\partial f}{\partial x_2}\hat{\mathbf{x}}_2 \dots \frac{\partial f}{\partial x_N}\hat{\mathbf{x}}_N\right) \cdot \hat{\mathbf{n}} = \boldsymbol{\nabla} f \cdot \hat{\mathbf{n}} \tag{11}$$

形式与式 9 相同．这里定义了 N **维直角坐标系的 Del 算符**为

$$\boldsymbol{\nabla} = \hat{\mathbf{x}}_1\frac{\partial f}{\partial x_1} + \hat{\mathbf{x}}_2\frac{\partial f}{\partial x_2} \dots \hat{\mathbf{x}}_n\frac{\partial f}{\partial x_N} \tag{12}$$

二元函数的极值（充分条件）

预备知识 导数与函数极值[170]，方向导数[279]

[5]类似一元函数，二元函数的极值与其偏导数密切相关．以下讨论中，我们假设在某区域内二元函数的一阶偏导处处存在（即函数曲面处处光滑）．如果二元函数 $f(x, y)$ 在某点 (x_i, y_i) 处对 x, y 的偏导数都为零，那么 (x_i, y_i) 就叫做函数 $f(x, y)$ 的**驻点**．根据式 9 [279]，驻点处各个方向的方向导数也都为零．

我们先来定义二元函数的**极值点**，以驻点为圆心在 xy 平面上作一个圆形区域，若当半径足够小时，$f(x_i, y_i)$ 是该圆形区域的最大值或最小值，那么该

[5]本文参考：[2] 下册的“多元函数的极值及其求法”一节．

驻点就是极大值点或极小值点．与一元函数类似，驻点不一定是极值点．例如 $f(x,y)=xy$ 在坐标原点的两个一阶偏导都为零，但原点并不是极值点．为了判断驻点是不是极值点，也需要用到二阶偏导（假设驻点处的各个二阶偏导都存在）．如果满足

$$\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2}-\left(\frac{\partial^2 f}{\partial x\partial y}\right)^2>0 \tag{1}$$

则驻点是极值点．如果 $\partial^2 f/\partial x^2$ 和 $\partial^2 f/\partial y^2$ 都大于零[6]，则极值为极小值，若都小于零，则极值为极大值．

注意式 1 只是存在极值的充分非必要条件．也就是说存在一些极值点不满足式 1．例如 $f(x,y)=x^4+y^4$ 在原点处的极值点．当式 1 左边小于零时，必定不是极值点，等于零时可能是也可能不是，需要用高阶导数进一步判断，这里暂时不讨论．

证明

类比一元函数的证明，要证明二元函数的某点是极值点，就要证明该点的任意二阶方向导数都大于零或都小于零[7]．令某方向为 $\hat{\mathbf{n}}=\hat{\mathbf{x}}\cos\theta+\hat{\mathbf{y}}\sin\theta$，由式 9 [279] 得该方向的方向导数为

$$\left(\cos\theta\frac{\partial}{\partial x}+\sin\theta\frac{\partial}{\partial y}\right)f \tag{2}$$

再次求方向导数得二阶方向导数为

$$\left(\cos\theta\frac{\partial}{\partial x}+\sin\theta\frac{\partial}{\partial y}\right)^2 f=\frac{\partial^2 f}{\partial x^2}\cos^2\theta+2\frac{\partial^2 f}{\partial x\partial y}\sin\theta\cos\theta+\frac{\partial^2 f}{\partial y^2}\sin^2\theta \tag{3}$$

如果你还不习惯看算符的平方，可以把上式的括号项平方看做两个括号项，依次作用在函数上．以极小值为例，令上式恒大于零，并除以 $\cos^2\theta$ 得

$$\frac{\partial^2 f}{\partial y^2}\tan^2\theta+2\frac{\partial^2 f}{\partial x\partial y}\tan\theta+\frac{\partial^2 f}{\partial x^2}>0 \tag{4}$$

上式左边是关于 $\tan\theta$ 的二次函数，若要恒大于零，则二次项系数要大于零，且判别式需小于零，立即可得式 1 ．同理可得极大值条件．

[6]根据式 1，$\partial^2 f/\partial x^2$ 和 $\partial^2 f/\partial y^2$ 的乘积大于零，所以只需要任意一个大于零，另外一个就必定大于零．一个小于零，另一个也必小于零．

[7]否则延一个方向前进函数值会越来越大，而延另一个方向前进函数值会越来越小，这个点就不是极值点

当判别式小于零时，必然存在不同方向的二阶方向导数具有相反的符号，所以必定不是极值点．而当判别式等于零时，存在某些方向的二阶导数为零，无法判断是否为极值点．

重积分

- 本词条存在未完成的内容．

预备知识 定积分[198]

面积分

面积分也叫**二重积分**，可以看做一元函数定积分[198]的一种拓展．从几何上来理解，如果一元函数的定积分是计算一定区间内被积函数曲线与坐标轴之间的面积，那么二重积分就是计算一定二维区域内被积函数曲面与坐标轴平面之间的体积（式 1 ）．

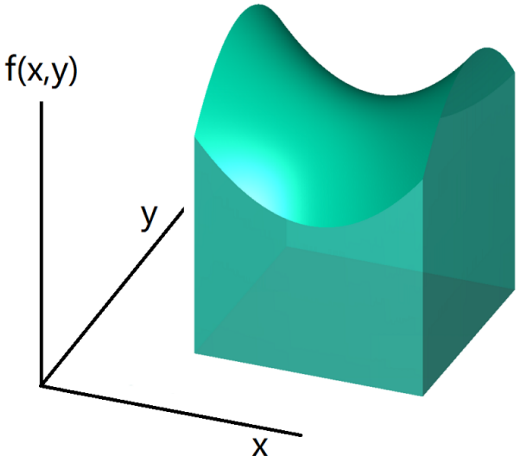


图 1: $f(x,y)$ 曲面下的体积

对一元函数的定积分，我们需要指定积分的下限和上限，那么对于面积分，我们需要指定 x-y 平面上的一块定积分的区域 $\mathcal{S}$．例如式 1 表示的面积分的区域就是一个正方形．与一元函数的情况相同，我们把积分区域划分成许多小长方形，每个小长方形的面积为 $\Delta x_i \Delta y_j$．把二重积分对应的体积用一个个柱体

来近似，每个柱体的体积为 $\Delta x_i \Delta y_j f(x_i, y_i)$ 其中 (x_i, y_i) 是长方形中任意一点. 当这些小长方形趋近无穷小时，就得到了精确的体积. 用双重极限表示为

$$V = \lim_{\substack{\Delta x_i \to 0 \\ \Delta y_j \to 0}} \sum_{i,j} f(x_i, y_j) \Delta x_i \Delta y_j \tag{1}$$

用积分符号可以把上式简记为

$$V = \iint_{\mathcal{S}} f(x, y)\, \mathrm{d}x\, \mathrm{d}y \tag{2}$$

注意积分区域 $\mathcal{S}$ 可以是任何形状的. 我们通过一个例子来说明具体如何计算面积分

例 1　抛物线下的体积

二元函数 $f(x, y) = x^2 + y^2$，可以在直角坐标系 xyz 代表一个抛物面. 现在我们指定一个二维区域，是由 $y_1(x) = -x$，$y_2(x) = x$ 以及 $x = 1$ 所围成的三角形. 现在我们以这个三角形为底取一个无限高的三棱柱，求三棱柱被 $f(x, y)$ 截去的有限部分的体积 V.

那如何计算这个积分呢？我们可以先计算任意 x 处的横截面积 $S(x)$. 这个横截面与曲面相交的曲线方程为 $f(x, y)$（把 x 看做常数，y 看做变量），横截面积为

$$S(x) = \int_{y_1(x)}^{y_2(x)} f(x, y)\, \mathrm{d}y = \frac{1}{3}[y_2^3(x) - y_1^3(x)] + x^2[y_2(x) - y_1(x)] = \frac{8}{3}x^3 \tag{3}$$

现在根据一元函数定积分的定义，体积等于横截面积 $S(x)$ 在 x 方向的定积分

$$V = \lim_{\Delta x_i \to 0} \sum_i S(x_i) \Delta x_i = \int_0^1 S(x)\, \mathrm{d}x = \frac{2}{3}(1^4 - 0^4) = \frac{2}{3} \tag{4}$$

要从定义上证明这种方法可行，我们可以把式 1 写成

$$V = \lim_{\Delta x_i \to 0} \sum_i \left[\lim_{\Delta y_j \to 0} \sum_j f(x_i, y_j) \Delta y_j \right] \Delta x_i \tag{5}$$

可见中括号内的极限就是定积分式 3，外面的极限就是定积分式 4. 把上式记为一条积分式就是

$$V = \int_0^1 \left[\int_{y_1(x)}^{y_2(x)} f(x, y)\, \mathrm{d}y \right] \mathrm{d}x \tag{6}$$

另一种等效的书写方法是，把 $\int \mathrm{d}x\,, \int \mathrm{d}y$ 看成像 $\mathrm{d}/\mathrm{d}x\,, \mathrm{d}/\mathrm{d}y$ 一样的整体算符，写在被积函数前面，使公式看起来更为简洁

$$V = \int_0^1 \mathrm{d}x \int_{y_1(x)}^{y_2(x)} \mathrm{d}y\, f(x,y) \tag{7}$$

总结起来，要在直角坐标系中计算二重积分，就先选取一个变量 y（也可以是 x）进行定积分，积分的上下限分别是二重积分区域 $\mathcal{S}$ 的两条边界线 $y_1(x)$ 和 $y_2(x)$．积分的结果中只含有 x，这时再对 x 进行定积分即可．

体积分

体积分也叫**三重积分**，类比面积分不难理解其概念

例 2　不均匀物体的质量

一个几何体由 $x=0, y=0, z=0$ 三个平面和 $x+y+z=1$ 平面包围而成，其密度为 $f(x,y,z)=x+y+z$，求几何体的总质量 M．

用定积分的思想，我们可以延三个方向把几何体划分成许多小块 (i,j,k)，每个小块的长宽高分别为 $\Delta x_i, \Delta y_j, \Delta z_k$．$x_i, y_j, z_k$ 分别为区间 $\Delta x_i, \Delta y_j, \Delta z_k$ 中的任意一点．物体的总质量可以用极限表示为

$$V = \lim_{\substack{\Delta x_i \to 0 \\ \Delta y_j \to 0 \\ \Delta z_k \to 0}} \sum_{i,j,k} f(x_i, y_j, z_k) \Delta x_i \Delta y_j \Delta z_k \tag{8}$$

用三重积分可以把这个三重极限表示为

$$V = \iiint_{\mathcal{V}} f(x,y,z)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \tag{9}$$

其中 $\mathcal{V}$ 表示几何体所占的空间．我们先把密度函数 $f(x,y,z)$ 沿 z 方向积分得到“面密度” $\sigma(x,y)$（可以想象把几何体在 z 方向压成一个薄板后所具有的面密度关于 x, y 的函数），积分上下限分别为 $z_2(x,y)=1-x-y$ 和 $z_1(x,y)=0$

$$\begin{aligned}
\sigma(x,y) &= \int_0^{1-x-y} (x+y+z)\, \mathrm{d}z \\
&= (x+y)[z_2(x,y) - z_1(x,y)] + \frac{1}{2}[z_2^2(x,y) - z_1^2(x,y)] \\
&= (x+y)(1-x-y) + \frac{1}{2}(1-x-y)^2 \\
&= \frac{1}{2}(1-x^2-y^2) - xy
\end{aligned} \tag{10}$$

再沿 y 方向积分得到“线密度” $\lambda(x)$（可以想象把薄板沿 y 方向压缩成一条线后所具有的线密度关于 x 的函数），积分的上下限分别为 $y_2(x)=1-x$ 和 $y_1(x)=0$

$$\begin{aligned}\lambda(x) &= \int_0^{1-x} \sigma(x,y)\,\mathrm{d}y \\ &= \frac{1}{2}(1-x^2)(1-x) - \frac{1}{6}(1-x)^3 - \frac{1}{2}x(1-x)^2 \\ &= \frac{x^3}{6} - \frac{x}{2} + \frac{1}{3}\end{aligned} \tag{11}$$

最后把线密度从 1 和 0 积分得到总质量

$$M = \int_0^1 \lambda(x)\,\mathrm{d}x = \left[\frac{1}{24}x^4 - \frac{1}{4}x^2 + \frac{1}{3}x\right]_0^1 = \frac{1}{8} \tag{12}$$

与二重积分类似，要从定义上来证明这种算法可行，就把式 8 改写成

$$M = \lim_{\Delta x_i\to 0}\sum_i \left[\lim_{\Delta y_j\to 0}\sum_j \left(\lim_{\Delta z_k\to 0}\sum_k f(x_i,y_j,z_k)\Delta z_k\right)\Delta y_j\right]\Delta x_i \tag{13}$$

可见从内到外的三个极限分别为三个变量的一元定积分，即

$$M = \int_0^1 \mathrm{d}x \int_0^{1-x} \mathrm{d}y \int_0^{1-x-y} \mathrm{d}z\, f(x,y,z) \tag{14}$$

未完成：链接一下柱坐标，球坐标系中的体积分，面积分等

立体角

- 本词条存在未完成的内容.

预备知识 重积分[283]

[8]如果我们以某种锥体（例如圆锥，三棱锥，假设其无限高）的顶点作为圆心作一个半径为 1 的球（**单位球**），那么这个锥体的**立体角（solid angle）**就是单位球的表面被锥体截出的体积，通常用 Ω 表示.

我们知道半径为 R 的球体的表面积为 $4\pi R^2$，所以立体角的取值范围是 $[0,4\pi]$.

[8]参考 Wikipedia 相关页面.

例 1　圆锥的立体角

顶角为 2α 的圆锥在单位球面上可截出一个球盖．在例 3 [202] 中我们知道球盖的面积，所以该圆锥的立体角为

$$\Omega = 2\pi(1-\cos\alpha) \tag{1}$$

对立体角积分

未完成：例如一个质量分布不均匀的球壳质量

极坐标系中单位矢量的偏导

预备知识　极坐标系[119]，矢量的偏导[274]

与直角坐标系不同的是，极坐标系中的 $\hat{\mathbf{r}}$ 与 $\hat{\boldsymbol{\theta}}$ 都是坐标的函数，即 $\hat{\mathbf{r}} = \hat{\mathbf{r}}(r,\theta)$，$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(r,\theta)$，它们对坐标的偏导如下

$$\begin{cases} \dfrac{\partial\hat{\mathbf{r}}}{\partial r} = 0 \\ \dfrac{\partial\hat{\mathbf{r}}}{\partial\theta} = \hat{\boldsymbol{\theta}} \end{cases} \qquad \begin{cases} \dfrac{\partial\hat{\boldsymbol{\theta}}}{\partial r} = 0 \\ \dfrac{\partial\hat{\boldsymbol{\theta}}}{\partial\theta} = -\hat{\mathbf{r}} \end{cases} \tag{1}$$

这是容易理解的，若一个单位矢量绕着它的起点逆时针转动，那么它的终点的速度的方向必然是它本身逆时针旋转 90 度的方向，而大小等于矢量模长乘以角速度．

证明

如果令极轴方向的单位矢量为 $\hat{\mathbf{x}}$，令其逆时针旋转 $\pi/2$ 的矢量为 $\hat{\mathbf{y}}$，则

$$\hat{\mathbf{r}} = \cos\theta\,\hat{\mathbf{x}} + \sin\theta\,\hat{\mathbf{y}} \tag{2}$$

$$\hat{\boldsymbol{\theta}} = \cos(\theta+\pi/2)\,\hat{\mathbf{x}} + \sin(\theta+\pi/2)\,\hat{\mathbf{y}} = -\sin\theta\,\hat{\mathbf{x}} + \cos\theta\,\hat{\mathbf{y}} \tag{3}$$

所以

$$\begin{cases} \dfrac{\partial \hat{\mathbf{r}}}{\partial r} = 0 \\ \dfrac{\partial \hat{\mathbf{r}}}{\partial \theta} = -\sin\theta\,\hat{\mathbf{x}} + \cos\theta\hat{\mathbf{y}} = \hat{\boldsymbol{\theta}} \end{cases} \qquad \begin{cases} \dfrac{\partial \hat{\boldsymbol{\theta}}}{\partial r} = 0 \\ \dfrac{\partial \hat{\boldsymbol{\theta}}}{\partial \theta} = -\cos\theta\,\hat{\mathbf{x}} - \sin\theta\,\hat{\mathbf{y}} = -\hat{\mathbf{r}} \end{cases} \tag{4}$$

事实上，由于 $\hat{\mathbf{r}}$ 与 $\hat{\boldsymbol{\theta}}$ 都只是 θ 的函数，也可以把偏导符号改成导数符号

$$\frac{\mathrm{d}\hat{\boldsymbol{\theta}}}{\mathrm{d}\theta} = -\hat{\mathbf{r}} \qquad \frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}\theta} = \hat{\boldsymbol{\theta}} \tag{5}$$

正交曲线坐标系

- 本词条存在未完成的内容.

预备知识 球坐标系[123]，柱坐标系[121]

[9]如果 u, v, w 是三维空间中某曲线坐标系的三个坐标，空间任意一点的位置矢量[878] $\mathbf{r}$ 都是它们的函数 $\mathbf{r}(u, v, w)$. 那么定义任意一点处三个单位矢量为

$$\hat{\mathbf{u}} = \frac{\partial \mathbf{r}/\partial u}{|\partial \mathbf{r}/\partial u|} \qquad \hat{\mathbf{v}} = \frac{\partial \mathbf{r}/\partial v}{|\partial \mathbf{r}/\partial v|} \qquad \hat{\mathbf{w}} = \frac{\partial \mathbf{r}/\partial w}{|\partial \mathbf{r}/\partial w|} \tag{1}$$

注意一般来说，这三个矢量会随着 $\mathbf{r}$ 改变. 形象地说：当我们分别只把 u, v, w 增加一点时，$\mathbf{r}$ 会分别沿 $\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{w}}$ 方向移动（请以球坐标系和柱坐标系为例思考）.

若对空间中任意一点，式 1 中的三个矢量都两两正交，那么这个曲线坐标系就是**正交曲线坐标系（orthogonal curvilinear coordinate system）**.

习题 1

练习：试着用式 1 计算球坐标和柱坐标中的单位矢量，例如球坐标中

$$\mathbf{r} = r\sin\theta\cos\phi\,\hat{\mathbf{x}} + r\sin\theta\sin\phi\,\hat{\mathbf{y}} + r\cos\theta\,\hat{\mathbf{z}} \tag{2}$$

然后用点乘[383]证明它们的单位矢量总是两两垂直. 即球坐标系和柱坐标系都是正交曲线坐标系.

[9]本文参考 Wikipedia 相关页面.

正交曲线坐标系中的矢量算符

预备知识 梯度[307]，旋度[315]

令式 1 中三个分母为 $f(u,v,w), g(u,v,w), h(u,v,w)$，则

$$\begin{aligned}\mathrm{d}\mathbf{r} &= \frac{\partial \mathbf{r}}{\partial u}\,\mathrm{d}u + \frac{\partial \mathbf{r}}{\partial v}\,\mathrm{d}v + \frac{\partial \mathbf{r}}{\partial w}\,\mathrm{d}w \\ &= f\hat{\mathbf{u}}\,\mathrm{d}u + g\hat{\mathbf{v}}\,\mathrm{d}v + h\hat{\mathbf{w}}\,\mathrm{d}w\end{aligned} \tag{3}$$

令 $s(u,v,w)$ 为标量函数，$\mathbf{A}(u,v,w)$ 为矢量函数，且

$$\mathbf{A}(u,v,w) = A_x(u,v,w)\hat{\mathbf{u}} + A_y(u,v,w)\hat{\mathbf{v}} + A_z(u,v,w)\hat{\mathbf{w}} \tag{4}$$

那么该坐标系中的梯度[307]，散度[311]，旋度[315] 算符分别为

$$\boldsymbol{\nabla} s = \frac{1}{f}\frac{\partial s}{\partial u}\hat{\mathbf{u}} + \frac{1}{g}\frac{\partial s}{\partial v}\hat{\mathbf{v}} + \frac{1}{h}\frac{\partial s}{\partial w}\hat{\mathbf{w}} \tag{5}$$

$$\boldsymbol{\nabla}\cdot\mathbf{A} = \frac{1}{fgh}\left[\frac{\partial}{\partial u}(ghA_u) + \frac{\partial}{\partial v}(fhA_v) + \frac{\partial}{\partial w}(fgA_w)\right] \tag{6}$$

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{A} = {}& \frac{1}{gh}\left[\frac{\partial}{\partial v}(hA_w) - \frac{\partial}{\partial w}(gA_v)\right]\hat{\mathbf{u}} \\ & + \frac{1}{fh}\left[\frac{\partial}{\partial w}(fA_u) - \frac{\partial}{\partial u}(hA_w)\right]\hat{\mathbf{v}} + \frac{1}{fg}\left[\frac{\partial}{\partial u}(gA_v) - \frac{\partial}{\partial v}(fA_u)\right]\hat{\mathbf{w}}\end{aligned} \tag{7}$$

$$\nabla^2 s = \frac{1}{fgh}\left[\frac{\partial}{\partial u}\left(\frac{gh}{f}\frac{\partial s}{\partial u}\right) + \frac{\partial}{\partial v}\left(\frac{fh}{g}\frac{\partial s}{\partial v}\right) + \frac{\partial}{\partial w}\left(\frac{fg}{h}\frac{\partial s}{\partial w}\right)\right] \tag{8}$$

未完成：推导参考 [12] 附录.

未完成：以下内容合并到相关词条

柱坐标系

我们先来分析柱坐标系[10]，位置矢量 $\hat{\mathbf{r}}$ 在直角坐标系中展开为

$$\mathbf{r}(r,\theta,z) = r\cos\theta\,\hat{\mathbf{x}} + r\sin\theta\,\hat{\mathbf{y}} + z\hat{\mathbf{z}} \tag{9}$$

[10]由于极坐标系可以看做柱坐标系 $z=0$ 的情况，我们不单独讨论

柱坐标系中三个单位矢量 $\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{z}}$ 的方向被定义为每个坐标增加时 $\mathbf{r}$ 增加的方向，即以下偏导数的方向

$$\begin{cases} \dfrac{\partial \mathbf{r}}{\partial r} = \cos\theta\,\hat{\mathbf{x}} + \sin\theta\,\hat{\mathbf{y}} \\ \dfrac{\partial \mathbf{r}}{\partial \theta} = -r\sin\theta\,\hat{\mathbf{x}} + r\cos\theta\,\hat{\mathbf{y}} \\ \dfrac{\partial \mathbf{r}}{\partial z} = \hat{\mathbf{z}} \end{cases} \tag{10}$$

将这三个矢量归一化，就得到三个单位矢量

$$\begin{cases} \hat{\mathbf{r}} = \cos\theta\,\hat{\mathbf{x}} + \sin\theta\,\hat{\mathbf{y}} \\ \hat{\boldsymbol{\theta}} = -\sin\theta\,\hat{\mathbf{x}} + \cos\theta\,\hat{\mathbf{y}} \\ \hat{\mathbf{z}} = \hat{\mathbf{z}} \end{cases} \tag{11}$$

可见柱坐标系和直角坐标系中的 $\hat{\mathbf{z}}$ 相同，而 $\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}$ 分别是 $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ 绕 z 轴逆时针旋转 θ 角所得．所以尽管柱坐标系中的三个单位矢量的方向取决于坐标，但它们始终两两垂直．

我们把单位矢量始终保持两两垂直的坐标系叫做**正交曲线坐标系**，或简称为**曲线坐标系**．我们熟知的直角坐标系显然就是一个正交曲线坐标系，稍后我们会看到球坐标系也是正交曲线坐标系．

现在我们可以将式 9 和式 10 用柱坐标中的三个单位矢量来表示．

$$\mathbf{r} = r\hat{\mathbf{r}} + z\hat{\mathbf{z}} \tag{12}$$

$$\frac{\partial \mathbf{r}}{\partial r} = \hat{\mathbf{r}} \qquad \frac{\partial \mathbf{r}}{\partial \theta} = r\hat{\boldsymbol{\theta}} \qquad \frac{\partial \mathbf{r}}{\partial z} = \hat{\mathbf{z}} \tag{13}$$

与极坐标的情况[287] 类似，将式 11 对 θ 求偏导可以得到单位矢量的偏导

$$\frac{\partial \hat{\mathbf{r}}}{\partial \theta} = \hat{\boldsymbol{\theta}} \qquad \frac{\partial \hat{\boldsymbol{\theta}}}{\partial \theta} = -\hat{\mathbf{r}} \qquad \frac{\partial \hat{\mathbf{z}}}{\partial \theta} = \mathbf{0} \tag{14}$$

根据式 13 和矢量函数的全微分，柱坐标系中一段微小位移可记为

$$\mathrm{d}\mathbf{r} = \frac{\partial \mathbf{r}}{\partial r}\,\mathrm{d}r + \frac{\partial \mathbf{r}}{\partial \theta}\,\mathrm{d}\theta + \frac{\partial \mathbf{r}}{\partial z}\,\mathrm{d}z = \mathrm{d}r\,\hat{\mathbf{r}} + r\,\mathrm{d}\theta\,\hat{\boldsymbol{\theta}} + \mathrm{d}z\,\hat{\mathbf{z}} \tag{15}$$

多元函数的傅里叶级数

预备知识 傅里叶级数[228]，重积分[283]

我们知道有限长区间中一个性质足够良好（满足迪利克雷条件）函数可以展开为三角函数的线性组合，同理，一个性质足够良好的 N 元函数 $f(x_1,\dots,x_N)$ 也可以展开为 N 个三角函数乘积的线性组合

$$f(x_1,\dots,x_N)=\sum_{n_1,\dots,n_N} C_{n_1,\dots,n_N}\exp\left(\mathrm{i}\frac{n_1\pi}{l_1}x_1\right)\dots\exp\left(\mathrm{i}\frac{n_N\pi}{l_N}x_N\right) \tag{1}$$

其中 x_i 的区间长度为 l_i，每个指标求和时取负无穷到正无穷的所有整数．系数 $C_{n_1,\dots,n_N}$ 可以由 N 重积分得到

$$C_{i_1,\dots,i_N}=\int\exp\left(-\mathrm{i}\frac{n_1\pi}{l_1}x_1\right)\dots\exp\left(-\mathrm{i}\frac{n_N\pi}{l_N}x_N\right)f(x_1,\dots,x_N)\,\mathrm{d}x_1\dots\mathrm{d}x_N \tag{2}$$

为了方便讨论，我们取 $N=2$，更高元的情况类比可得．区间 $x\in[0,a],y\in[0,b]$ 内的二元函数 $f(x,y)$ 的傅里叶展开为

$$f(x,y)=\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}C_{m,n}\exp\left(\mathrm{i}\frac{m\pi}{a}x\right)\exp\left(\mathrm{i}\frac{n\pi}{b}y\right) \tag{3}$$

系数由二重积分计算

$$C_{m,n}=\int_0^b\int_0^a\exp\left(-\mathrm{i}\frac{m\pi}{a}x\right)\exp\left(-\mathrm{i}\frac{n\pi}{b}y\right)f(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{4}$$

正交曲线坐标系中的重积分

预备知识 正交曲线坐标系[288]

在计算一些多重积分时，选取合适的坐标系往往可以大大化简问题．

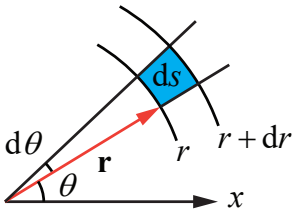


图 1: 极坐标中的面积元

极坐标系中的二重积分

我们来看如何在极坐标系中进行二重积分. 我们先把积分区域划分为无数个小面元，点 $\mathbf{r}$ 处面元的形状如图 1 所示，即把两个坐标 r, θ 坐标分别在原来的基础上增加一个微小值，并围成一块小区域. 由于 $\mathrm{d}r\,, \mathrm{d}\theta$ 都是无穷小，该面元的形状趋近于长方形，其面积为两边长相乘

$$\mathrm{d}s = \mathrm{d}r \cdot r\,\mathrm{d}\theta = r\,\mathrm{d}r\,\mathrm{d}\theta \tag{1}$$

类比例 1 [283]，我们可以将 $f(r,\theta)$ 的面积分记为

$$\iint_{\mathcal{S}} f(r,\theta) r\,\mathrm{d}r\,\mathrm{d}\theta = \int_{\theta_1}^{\theta_2} \mathrm{d}\theta \int_{r_1(\theta)}^{r_2(\theta)} \mathrm{d}r\, r f(r,\theta) \tag{2}$$

其中 $r_1(\theta)$ 与 $r_2(\theta)$ 是区域 $\mathcal{S}$ 的两条边界（类比式 7 [283] 中的 $y_1(x), y_2(x)$）

例 1

求 $f(r,\theta) = ar$ 在内外半径为 R_1, R_2 的圆环区域的面积分.

先来看积分上下限，对于圆环区域，显然有 $r_1(\theta) = R_1$，$r_2(\theta) = R_2$，$\theta_1 = 0$，$\theta_2 = 2\pi$. 直接使用式 2 得

$$\begin{aligned}\iint_{\mathcal{S}} f(r,\theta) r\,\mathrm{d}r\,\mathrm{d}\theta &= \int_0^{2\pi} \left[\int_{R_1}^{R_2} ar^2\,\mathrm{d}r\right] \mathrm{d}\theta = \int_0^{2\pi} \frac{a}{3}(R_2^3 - R_1^3)\,\mathrm{d}\theta \\ &= \frac{2\pi a}{3}(R_2^3 - R_1^3)\end{aligned} \tag{3}$$

如果使用直角坐标系计算该积分，过程将会变得十分复杂.

曲线坐标系中的体积分

在曲线坐标系中，令

$$\frac{\partial \mathbf{r}}{\partial x_i} \equiv f_i(\mathbf{r})\hat{\mathbf{x}}_i \quad (i = 1,2,3) \tag{4}$$

则位矢的全微分为

$$\mathrm{d}\mathbf{r} = \sum_{i=1}^{3} f_i(\mathbf{r})\,\mathrm{d}x_i\,\hat{\mathbf{x}}_i \tag{5}$$

所以空间中的一个体积元（每个 x_i 都分别增加 $\mathrm{d}x_i$ 所围成的长方体）可以表示为

$$\mathrm{d}V = f_1(\mathbf{r})f_2(\mathbf{r})f_3(\mathbf{r})\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3 \tag{6}$$

这是因为根据式 5，单独将坐标 x_i 增加 $\mathrm{d}x_i$ 会导致 $\mathbf{r}$ 在 $\hat{\mathbf{x}}_i$ 方向增加 $f_i(\mathbf{r})\,\mathrm{d}x_i$，这相当于长方体在 $\hat{\mathbf{x}}_i$ 方向的边长，而长方体的体积等于三条边长之积．为了方便书写我们以后将 $\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3$ 记为 d^3x 或 d^3r．

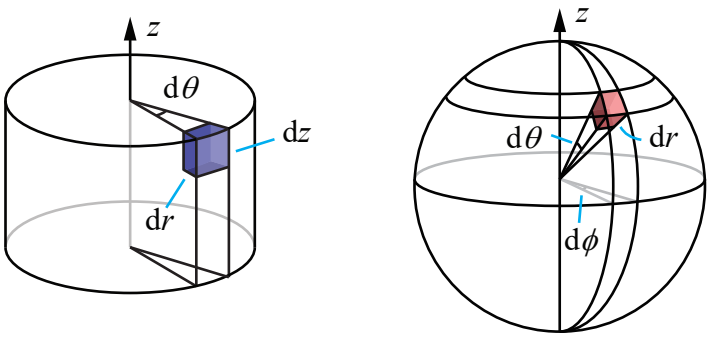


图 2: 柱坐标（左）和球坐标（右）中的体积元

我们已知直角坐标系中 $\partial\mathbf{r}/\partial x_i = 1$，所以体积元为 $\mathrm{d}^3r = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$．对于柱坐标系（图 2 左），由式 13 [288] 得体积元为

$$\mathrm{d}V = \mathrm{d}r \cdot r\,\mathrm{d}\theta \cdot \mathrm{d}z = r\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}z \tag{7}$$

类似地，对于球坐标系（图 2 右），由式 11 [742] 得体积元为

$$\mathrm{d}V = \mathrm{d}r \cdot r\,\mathrm{d}\theta \cdot r\sin\theta\,\mathrm{d}\phi = r^2\sin\theta\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi \tag{8}$$

例 2　球体的体积

在例 4 [198] 中我们用一元函数的定积分得到了球体的体积，现在我们也可

以直接在球坐标中由体积分得到.

$$\begin{aligned} V &= \int 1 \,\mathrm{d}^3 r = \int_0^{2\pi} \int_0^{\pi} \int_0^{R} r^2 \sin\theta \,\mathrm{d}r \,\mathrm{d}\theta \,\mathrm{d}\phi \\ &= \int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \sin\theta \,\mathrm{d}\theta \int_0^{R} r^2 \,\mathrm{d}r \\ &= 2\pi \cdot 2 \cdot \frac{1}{3} R^3 = \frac{4}{3}\pi R^3 \end{aligned} \tag{9}$$

矢量场

预备知识 球坐标系的定义[123]，矢量的求导法则[274]

对空间中指定范围的每一点 P 赋予一个矢量 $\mathbf{v}$，就在该空间中形成了一个**矢量场**. 例如，电荷附近的任意一点都存在一个电场矢量，这就构成了一个矢量场. 管道中任意一点的水流都存在一个速度矢量，它们也构成一个矢量场. 其他例子如力场[905]，电场[1124]，磁场[1129].

直角坐标系

矢量场在不同的参考系中有不同的表示方法. 在空间直角坐标系中，矢量场可以用矢量的三个分量关于 x, y, z 三个坐标的函数表示. 点 $P(x, y, z)$ 处的矢量分量为

$$\begin{cases} v_x(x, y, z) = \mathbf{v} \cdot \hat{\mathbf{x}} \\ v_y(x, y, z) = \mathbf{v} \cdot \hat{\mathbf{y}} \\ v_z(x, y, z) = \mathbf{v} \cdot \hat{\mathbf{z}} \end{cases} \tag{1}$$

也可以作为单位正交基[386] 的线性组合写成一个整体

$$\begin{aligned} \mathbf{v} &= (\mathbf{v} \cdot \hat{\mathbf{x}})\,\hat{\mathbf{x}} + (\mathbf{v} \cdot \hat{\mathbf{y}})\,\hat{\mathbf{y}} + (\mathbf{v} \cdot \hat{\mathbf{z}})\,\hat{\mathbf{z}} \\ &= v_x(x, y, z)\,\hat{\mathbf{x}} + v_y(x, y, z)\,\hat{\mathbf{y}} + v_z(x, y, z)\,\hat{\mathbf{z}} \end{aligned} \tag{2}$$

球坐标系

在球坐标系[123] 中，也可以把每个点的矢量根据该点处的三个单位矢量 $\hat{\mathbf{r}}$，$\hat{\boldsymbol{\theta}}$，$\hat{\boldsymbol{\phi}}$ 分解为三个分量．基底的线性组合为

$$\mathbf{v} = v_r(r,\theta,\phi)\,\hat{\mathbf{r}} + v_\theta(r,\theta,\phi)\,\hat{\boldsymbol{\theta}} + v_\phi(r,\theta,\phi)\,\hat{\boldsymbol{\phi}} \tag{3}$$

需要特别注意，$\hat{\mathbf{r}}$，$\hat{\boldsymbol{\theta}}$，$\hat{\boldsymbol{\phi}}$ 也是关于 (r,θ,ϕ) 的函数，所以对 $\mathbf{v}$ 求导（或偏导）时必须根据矢量的求导法则[274] 进行．

场线

场线是矢量场的一种可视化工具．我们可以顺着矢量场的方向画出多条有方向的不相交曲线，使得曲线上任意一点的切线方向都等于该点处矢量场的方向．对于无散场，场线没有起点和终点；对于无旋场，场线不会闭合．

未完成：链接，说明

线积分

预备知识 矢量的内积[383]，定积分[198]

我们先来通过一个物理的例子引入**曲线积分**的概念，见“功 功率[903]”.

下面讨论如何在直角坐标系中具体计算线积分．为书写方便，以下省略积分路径 C_{ab}.

将被积曲线的参数方程表示为[11] $x(t),y(t),z(t)$，则曲线上任意一点都唯一对应一个 t 值．根据微分关系，当 t 增加 $\mathrm{d}t$ 时，曲线上的一小段位移矢量 $\mathrm{d}\mathbf{r} = (\mathrm{d}x\,,\mathrm{d}y\,,\mathrm{d}z)$ 中

$$\mathrm{d}x = x'(t)\,\mathrm{d}t \qquad \mathrm{d}y = y'(t)\,\mathrm{d}t \qquad \mathrm{d}z = z'(t)\,\mathrm{d}t \tag{1}$$

这样，对曲线上任意一点（对应参数 t），$\mathbf{F}$ 可表示成 t 的矢量函数 $\mathbf{F}(t) = \mathbf{F}[x(t),y(t),z(t)]$．$\mathbf{F}$ 的三个分量[12] 则表示为关于 t 的单变量标量函数

$$F_{x_i}(t) = F_{x_i}[x(t),y(t),z(t)] \quad (i=1,2,3) \tag{2}$$

[11] 注意这里的 t 不一定代表时间，可以是任意参数，甚至可以是 x,y,z 中的一个.

[12] 为了书写简洁，这里定义 $x_1 \equiv x, x_2 \equiv y, x_3 \equiv z$.

下面将三维空间的线积分转换为三个一元定积分

$$\begin{aligned}\int \mathbf{F}(\mathbf{r})\cdot \mathrm{d}\mathbf{r} &= \lim_{n\to\infty}\sum_{i=1}^{n}\mathbf{F}(\mathbf{r}_i)\cdot\Delta\mathbf{r}_i \\ &= \lim_{n\to\infty}\sum_{i=1}^{n}F_x(\mathbf{r}_i)\Delta x_i + \lim_{n\to\infty}\sum_{i=1}^{n}F_y(\mathbf{r}_i)\Delta y_i + \lim_{n\to\infty}\sum_{i=1}^{n}F_z(\mathbf{r}_i)\Delta z_i \\ &= \int F_x(\mathbf{r})\,\mathrm{d}x + \int F_y(\mathbf{r})\,\mathrm{d}y + \int F_z(\mathbf{r})\,\mathrm{d}z\end{aligned} \tag{3}$$

设积分路径 C_{ab} 的起点对应 $t=a$，终点对应 $t=b$．结合式 1，上面每一项积分可以表示为

$$\int F_{x_i}(\mathbf{r})\,\mathrm{d}x_i = \int_a^b F_{x_i}[\mathbf{r}(t)]x_i'(t)\,\mathrm{d}t \quad (i=1,2,3) \tag{4}$$

计算这三个关于 t 的定积分再相加，就可以得出线积分结果．

例 1　计算力场对质点的做功

令力场为 $\mathbf{F}=\alpha r\,\hat{\mathbf{r}}$，一质点从原点出发，沿轨迹 $(x-a)^2+y^2=a^2$ 的上半部分移动到 $(a,0)$，求力对质点做的功．若起点终点不变，轨迹改为延 x 轴，结果又如何？

我们先来建立运动轨迹的参数方程．由于运动是一个圆，我们可以使用圆的参数方程．把角度作为参数 t，$t\in[0,\pi]$．

$$\begin{cases}x(t)=a(1-\cos t)\\ y(t)=a\sin t\end{cases}\qquad\begin{cases}x'(t)=a\sin t\\ y'(t)=a\cos t\end{cases} \tag{5}$$

把力场在直角坐标系中表示为 $\mathbf{F}(x,y)=\alpha(x\,\hat{\mathbf{x}}+y\,\hat{\mathbf{y}})$，两个分量分别为 $F_x=\alpha x, F_y=\alpha y$．由式 4 $(i=1,2)$，力场对质点做功等于两个定积分之和

$$W=\int\mathbf{F}\cdot\mathrm{d}\mathbf{r}=\int_0^{\pi}\alpha a(1-\cos t)\cdot a\sin t\,\mathrm{d}t+\int_0^{\pi}\alpha a\sin t\cdot a\cos t\,\mathrm{d}t \tag{6}$$

注意到第一个积分中的第二项恰好是第二个积分的相反数，所以上式变为

$$\int_0^{\pi}\alpha a^2\sin t\,\mathrm{d}t=2\alpha a^2 \tag{7}$$

现在来计算延 x 轴的直线轨迹运动的情况．由于轨迹上处处都有 $y=0$，$F_y=0$，积分只有 F_x 一项．另外 x 本身就可以作为轨道参数，即 $x(t)=t, y(t)=$

$0, x \in [0, 2a]$. 代入式 4 得做功为

$$W = \int_0^{2a} \alpha x \, \mathrm{d}x = 2\alpha a^2 \tag{8}$$

在上例中，我们发现对于给定的矢量场，即使路径不同，当起点和终点相同时，线积分的结果也相同（虽然我们只计算了两条路径，但这个结论是正确的）. 具有这样性质的矢量场叫做**保守场**，并总存在一个势能函数.

曲面积分 通量

预备知识 矢量场[294]，二重积分[283]，曲面的法向量

我们先来看一个例子

例 1　匀速水流场的流量

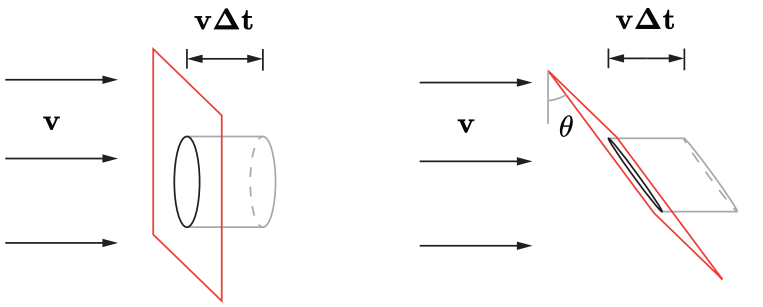


图 1: 匀速水流场的流量

水流中各点的速度可以看做一个矢量场 $\mathbf{v}(\mathbf{r})$，假设 $\mathbf{v}(\mathbf{r}) = \mathbf{v}_0$ 是一个常矢量. 在矢量场中取一个平面，平面的法向量与 $\mathbf{v}_0$ 的夹角为 θ，在平面上选一个面积为 S 的区域，求单位时间通过该区域的水流的体积.

先来考虑 $\theta = 0$ 的情况，在 Δt 时间内，流过区域的水是一个柱体，其底面积为 S，高为 $v_0 \Delta t$，所以体积为 $v_0 S \Delta t$，所以单位时间的体积等于流速乘以面积 $v_0 S$.

我们再来考虑 $\theta \neq 0$ 的情况，Δt 时间内流过区域的水是一个斜柱体，由图 1 可知其体积等于横截面积 $S\cos\theta$ 乘以斜边长度 $v_0\Delta t$ 即 $V = v_0 S\cos\theta\Delta t$，所以单位时间流过的体积为 $v_0 S\cos\theta$．可见随着 θ 增大，流量变小，直到 $\theta = \pi/2$ 时流量为 0．

现在试想速度场 $\mathbf{v}(\mathbf{r})$ 不是常矢量，且有限的平面改为有限的曲面，要计算单位时间流过曲面的体积，我们可以先将曲面划分成无穷多个小曲面，每个曲面上的流速看做常矢量，再按照上述的办法计算并求和即可．

现在我们来定义一个矢量场在一个曲面上的**曲面积分**[13]（或**通量**）．假设矢量场与曲面都处处光滑，令矢量场为 $\mathbf{F}(\mathbf{r})$，给曲面定义一个正方向，并将曲面划分为许多面积为 ΔS_i 的小面元，令其法向量为 $\hat{\mathbf{n}}_i$（与曲面正方向同侧），则一块面元可以表示为 $\Delta\mathbf{S}_i = \Delta S_i\hat{\mathbf{n}}_i$．当面元很小时可以假设其内部的矢量场为常矢量 $\mathbf{F}(\mathbf{r}_i)$，则面积分被定义为

$$\int \mathbf{F}(\mathbf{r})\cdot \mathrm{d}\mathbf{S} = \lim_{\Delta S_i\to 0}\sum_i \mathbf{F}(\mathbf{r}_i)\cdot\Delta\mathbf{S}_i \tag{1}$$

直角坐标系中的曲面积分

直角坐标系中的曲面可以表示为 $f(x,y,z) = 0$ 其法向量 $\hat{\mathbf{n}}(x,y,z)$ 等于 $\pm\boldsymbol{\nabla} f(x,y,z)$ 归一化．令矢量场为 $\mathbf{F}(x,y,z)$，应如何具体计算曲面积分呢？

以下我们来讨论曲面可以表示为 $z = g(x,y)$ 的情况[14]．将曲面沿 x,y 方向划分成许多面元，使每个面元在 x,y 平面上的投影都是一个小长方形，面积为 $\Delta x_i, \Delta y_j$，面元上任意一点的坐标为 $[x_i, y_j, g(x_i,y_j)]$．面元面积与投影面积的关系为 $\Delta S_{ij}\cos\theta = \Delta x_i\Delta y_j$，其中 θ 是面元的法向量与 z 轴的夹角，所以 $\cos\theta = \hat{\mathbf{n}}\hat{\mathbf{z}} = n_z$．所以，矢量场在每个面元上的通量为

$$\mathbf{F}(\mathbf{r}_{ij})\cdot\Delta\mathbf{S}_{ij} = \left(\frac{n_x}{n_z}F_x + \frac{n_y}{n_z}F_y + F_z\right)\Delta x_i\Delta y_j \tag{2}$$

其中 $\hat{\mathbf{n}}, \mathbf{F}$ 的个分量在 $[x_i, y_j, g(x_i,y_j)]$ 处取值．

---

[13] 简称**面积分**，这时需要通过语境与“二重积分[283]”区分开．

[14] 如果不能表示为 $z = g(x,y)$，但可以表示为 $x = g(y,z)$ 或 $y = g(x,z)$，以下过程也类似．

由式 1 的定义，曲面积分为

$$\begin{aligned}\int \mathbf{F}(\mathbf{r})\cdot \mathrm{d}\mathbf{S} &= \lim_{\substack{\Delta x_i\to 0\\ \Delta y_i\to 0}} \sum_{ij} \mathbf{F}(\mathbf{r}_{ij})\cdot \Delta \mathbf{S}_{ij} \\ &= \iint \frac{n_x}{n_z} F_x \,\mathrm{d}x\,\mathrm{d}y + \iint \frac{n_y}{n_z} F_y \,\mathrm{d}x\,\mathrm{d}y + \iint F_z \,\mathrm{d}x\,\mathrm{d}y\end{aligned} \tag{3}$$

这样，我们就把曲面积分转换成了三个二重积分．这三个二重积分分别等于矢量场的三个分量对曲面通量的贡献．特殊地，若曲面方程可以记为 $z = g_1(x,y), x = g_2(y,z), y = g_3(x,z)$ 中的任意一种形式，我们也可以将曲面积分记为

$$\int \mathbf{F}(\mathbf{r})\cdot \mathrm{d}\mathbf{S} = \iint F_x \,\mathrm{d}y\,\mathrm{d}z + \iint F_y \,\mathrm{d}x\,\mathrm{d}z + \iint F_z \,\mathrm{d}x\,\mathrm{d}y \tag{4}$$

例 2

求场 $\mathbf{F}(\mathbf{r}) = \mathbf{r}$ 在曲面（正方向向上）$z = x^2 + y^2 \quad (x, y \in [-1, 1])$ 上的通量.

我们先求曲面的法向量，令 $f(x,y,z) = x^2 + y^2 - z$，则法向量的方向为 $\nabla f = (2x, 2y, -1)$ 由于曲面的正方向向上，我们将 ∇f 取相反数然后归一化得到法向量

$$\hat{\mathbf{n}} = \frac{-2x}{\sqrt{4x^2+4y^2+1}}\hat{\mathbf{x}} + \frac{-2y}{\sqrt{4x^2+4y^2+1}}\hat{\mathbf{y}} + \frac{1}{\sqrt{4x^2+4y^2+1}}\hat{\mathbf{z}} \tag{5}$$

将上式和 $\mathbf{F}(\mathbf{r}) = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}$ 代入式 3 得

$$\begin{aligned}\int \mathbf{F}(\mathbf{r})\cdot \mathrm{d}\mathbf{S} &= -\iint 2x^2 \,\mathrm{d}x\,\mathrm{d}y - \iint 2y^2 \,\mathrm{d}x\,\mathrm{d}y + \iint (x^2+y^2)\,\mathrm{d}x\,\mathrm{d}y \\ &= -\frac{4}{3}x^3\Big|_{-1}^{1} - \frac{4}{3}y^3\Big|_{-1}^{1} + \frac{2}{3}x^3\Big|_{-1}^{1} + \frac{2}{3}y^3\Big|_{-1}^{1} = -\frac{8}{3}\end{aligned} \tag{6}$$

证明闭合曲面的法向量面积分为零

预备知识 散度散度定理[311]

对于闭合曲面 Ω，证明 $\oint_\Omega \mathrm{d}\mathbf{s} = \mathbf{0}$，其中 $\mathrm{d}\mathbf{s}$ 是曲面上的面元矢量，其模长 $\mathrm{d}s$ 为面元的面积，方向为沿着该面元的法线向外．把矢量积分分解成三个分

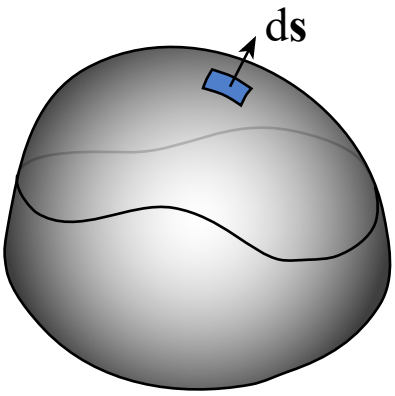


图 1: 闭合曲面上的一个面元

量，有

$$\begin{aligned}\oint_\Omega \mathrm{ds} &= \oint_\Omega [(\hat{\mathbf{x}}\,\mathrm{ds})\hat{\mathbf{x}} + (\hat{\mathbf{y}}\,\mathrm{ds})\hat{\mathbf{y}} + (\hat{\mathbf{z}}\,\mathrm{ds})\hat{\mathbf{z}}] \\ &= \hat{\mathbf{x}}\oint_\Omega \hat{\mathbf{x}}\,\mathrm{ds} + \hat{\mathbf{y}}\oint_\Omega \hat{\mathbf{y}}\,\mathrm{ds} + \hat{\mathbf{z}}\oint_\Omega \hat{\mathbf{z}}\,\mathrm{ds}\end{aligned} \tag{1}$$

下面以第一项为例，证明积分结果为 0，后两项的证明类似. 根据散度定理，矢量场在闭合曲面上的通量等于该矢量场散度在曲面内空间的体积分，所以

$$\oint_\Omega \hat{\mathbf{x}}\,\mathrm{ds} = \int_V (\boldsymbol{\nabla}\cdot\hat{\mathbf{x}})\,\mathrm{d}V = \int_V 0\,\mathrm{d}V = 0 \tag{2}$$

矢量算符

预备知识 叉乘[387]，偏微分算符[277]

标量函数与矢量函数

我们先区分两种函数，第一种是普通的多元函数 $f(x,y,z)$，也叫**标量函数**，即自变量 x,y,z 是实数，函数值也是实数[15]. 另一种是**矢量函数**，一般用粗体加以区分（手写的时候在上方加箭头），如 $\mathbf{f}(x,y,z)$，即函数值是一个 3 维矢量[16]. 矢量函数也可以记为三个分量的形式

$$\mathbf{f}(x,y,z) = f_x(x,y,z)\hat{\mathbf{x}} + f_y(x,y,z)\hat{\mathbf{y}} + f_z(x,y,z)\hat{\mathbf{z}} \tag{1}$$

[15]一些情况下也可以是复数

[16]一些情况下也可以是 $N = 1,2,\ldots$ 维

矢量算符

定义三维的**矢量算符（vector operator）**为（也叫 nabla 或 del 算符）

$$\boldsymbol{\nabla} = \hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z} \tag{2}$$

$\boldsymbol{\nabla}$ 作用在标量函数 $f(x,y,z)$ 上的结果称为函数的**梯度**，是一个矢量函数

$$\boldsymbol{\nabla} f(x,y,z) = \left(\hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z}\right) f = \frac{\partial f}{\partial x}\hat{\mathbf{x}} + \frac{\partial f}{\partial y}\hat{\mathbf{y}} + \frac{\partial f}{\partial z}\hat{\mathbf{z}} \tag{3}$$

这可以类比矢量与标量的乘法.

$\boldsymbol{\nabla}$ 与矢量函数 $\mathbf{f}(x,y,z)$ 的作用通常有两种定义，第一是“点乘”，结果称为函数的**散度（divergence）**，是一个标量函数

$$\begin{aligned}\boldsymbol{\nabla}\cdot\mathbf{f}(x,y,z) &= \left(\hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z}\right)\cdot(\hat{\mathbf{x}}f_x + \hat{\mathbf{y}}f_y + \hat{\mathbf{z}}f_z) \\ &= \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}\end{aligned} \tag{4}$$

可以类比两个矢量的点乘（内积）[383].

另一种情况是“叉乘”，结果称为函数的**旋度（curl）**，是一个矢量函数

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{f}(x,y,z) &= \left(\hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z}\right)\times(\hat{\mathbf{x}}f_x + \hat{\mathbf{y}}f_y + \hat{\mathbf{z}}f_z) \\ &= \begin{vmatrix}\hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ f_x & f_y & f_z\end{vmatrix} \\ &= \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right)\hat{\mathbf{x}} + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right)\hat{\mathbf{y}} + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right)\hat{\mathbf{z}}\end{aligned} \tag{5}$$

可以类比两个矢量的叉乘[383].

另见拉**普拉斯算符**[301].

拉普拉斯算符

预备知识 矢量算符[307]

我们令一个标量函数 $u(x,y,z)$ 的梯度的散度为它的**拉普拉斯（Laplacian）**，合成的算符（类比复合函数）叫做**拉普拉斯算符**，记为 ∇^2.

$$\nabla^2 u = \boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}u) = \frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2} \tag{1}$$

也可以记

$$\begin{aligned}\nabla^2 = \boldsymbol{\nabla}\cdot\boldsymbol{\nabla} = (\boldsymbol{\nabla})^2 &= \left(\hat{\mathbf{x}}\frac{\partial}{\partial x}+\hat{\mathbf{y}}\frac{\partial}{\partial y}+\hat{\mathbf{z}}\frac{\partial}{\partial z}\right)^2 \\ &= \frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\end{aligned} \tag{2}$$

这些定义也容易拓展到 $N=1,2,\ldots$ 元函数上.

一种矢量算符的运算方法

预备知识 矢量算符[300]，连续叉乘的化简[392]

声明：本文中的方法是笔者原创，使用的下标符号也是笔者自己定义的.

在用 $\boldsymbol{\nabla}$ 算符计算梯度，散度和旋度时，我们几乎可以将其看作一个矢量进行运算，唯一的区别就是我们需要明确每一项中的偏微分是对哪些变量进行的. 例如

$$\boldsymbol{\nabla}\cdot(U\mathbf{A}) = \frac{\partial}{\partial x}(UA_x)+\frac{\partial}{\partial y}(UA_y)+\frac{\partial}{\partial z}(UA_z) \tag{1}$$

$$(\mathbf{A}\cdot\boldsymbol{\nabla})U = A_x\frac{\partial U}{\partial x}+A_y\frac{\partial U}{\partial y}+A_z\frac{\partial U}{\partial z} \tag{2}$$

如果以上两式中把 $\boldsymbol{\nabla}$ 符号替换成一个普通的矢量，两式将没有任何区别. 可见 $\boldsymbol{\nabla}$ 符号包含了另一层信息，这个信息通过 $\boldsymbol{\nabla}$ 所在的位置来体现，但我们希望能定义一种新的符号 $[\ldots]_{\ldots}$，把偏导算符的作用对象在方括号的角标中声明，而在方括号内的 $\boldsymbol{\nabla}$ 可以像普通矢量一样进行运算，例如

$$[\boldsymbol{\nabla}\cdot(U\mathbf{A})]_{A\partial U} \equiv [\mathbf{A}\cdot\boldsymbol{\nabla}U]_{A\partial U} \equiv [U\boldsymbol{\nabla}\cdot\mathbf{A}]_{A\partial U} \equiv \mathbf{A}\cdot\boldsymbol{\nabla}U \tag{3}$$

又如，利用矢量公式 $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=\mathbf{B}(\mathbf{A}\cdot\mathbf{C})-\mathbf{C}(\mathbf{A}\cdot\mathbf{B})$，有

$$[\boldsymbol{\nabla}\times(\mathbf{A}\times\mathbf{B})]_{\partial(AB)}=[\mathbf{A}(\boldsymbol{\nabla}\cdot\mathbf{B})+\mathbf{B}(\boldsymbol{\nabla}\cdot\mathbf{A})]_{\partial(AB)} \tag{4}$$

另外，由乘法的求导法则，有

$$[\dots]_{\partial(AB)}=[\dots]_{B\partial A}+[\dots]_{A\partial B} \tag{5}$$

使用这个新符号，我们可以化简许多常用的矢量公式.

例 1

证明 $\boldsymbol{\nabla}\times(U\mathbf{A})=(\boldsymbol{\nabla}U)\times\mathbf{A}+U\boldsymbol{\nabla}\times\mathbf{A}$.

$$\begin{aligned}[\boldsymbol{\nabla}\times(U\mathbf{A})]_{\partial(UA)}&=[\boldsymbol{\nabla}\times(U\mathbf{A})]_{A\partial U}+[\boldsymbol{\nabla}\times(U\mathbf{A})]_{U\partial A}\\&=[(\boldsymbol{\nabla}U)\times\mathbf{A}]_{A\partial U}+[U\boldsymbol{\nabla}\times\mathbf{A}]_{U\partial A}\\&=(\boldsymbol{\nabla}U)\times\mathbf{A}+U\boldsymbol{\nabla}\times\mathbf{A}\end{aligned} \tag{6}$$

证毕.

例 2

化简 $\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{E})$.

$$[\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{E})]_{\partial^2E}=[\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{E})-\nabla^2\mathbf{E}]_{\partial^2E}=\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{E})-\nabla^2\mathbf{E} \tag{7}$$

例 3

证明 $\boldsymbol{\nabla}(\mathbf{F}\cdot\mathbf{G})=\mathbf{F}\times(\boldsymbol{\nabla}\times\mathbf{G})+\mathbf{G}\times(\boldsymbol{\nabla}\times F)+(\mathbf{F}\cdot\boldsymbol{\nabla})\mathbf{G}+(\mathbf{G}\cdot\boldsymbol{\nabla})\mathbf{F}$.

从右向左证明，上式等于

$$\begin{aligned}&[\mathbf{F}\times(\boldsymbol{\nabla}\times\mathbf{G})]_{F\partial G}+[\mathbf{G}\times(\boldsymbol{\nabla}\times F)]_{G\partial F}+[(\mathbf{F}\cdot\boldsymbol{\nabla})\mathbf{G}]_{F\partial G}+[(\mathbf{G}\cdot\boldsymbol{\nabla})\mathbf{F}]_{G\partial F}\\&=[\boldsymbol{\nabla}(\mathbf{F}\cdot\mathbf{G})-(\mathbf{F}\cdot\boldsymbol{\nabla})\mathbf{G}]_{F\partial G}+[\boldsymbol{\nabla}(\mathbf{F}\cdot\mathbf{G})-(\mathbf{G}\cdot\boldsymbol{\nabla})\mathbf{F}]_{G\partial F}\\&\quad+[(\mathbf{F}\cdot\boldsymbol{\nabla})\mathbf{G}]_{F\partial G}+[(\mathbf{G}\cdot\boldsymbol{\nabla})\mathbf{F}]_{G\partial F}\\&=[\boldsymbol{\nabla}(\mathbf{F}\cdot\mathbf{G})]_{F\partial G}+[\boldsymbol{\nabla}(\mathbf{F}\cdot\mathbf{G})]_{F\partial G}\\&=[\boldsymbol{\nabla}(\mathbf{F}\cdot\mathbf{G})]_{\partial(FG)}=\boldsymbol{\nabla}(\mathbf{F}\cdot\mathbf{G})\end{aligned} \tag{8}$$

证毕.

矢量算符运算法则

- 本词条需要更多讲解，便于帮助理解.

预备知识 梯度[307]，旋度[315]，列维—奇维塔符号[431]

[17]这里列举一些常用的矢量算符恒等式.

一阶导数

梯度

$$\boldsymbol{\nabla}(fg) = f\boldsymbol{\nabla}g + g\boldsymbol{\nabla}f \tag{1}$$

$$\boldsymbol{\nabla}(\mathbf{A}\cdot\mathbf{B}) = \mathbf{A}\times(\boldsymbol{\nabla}\times\mathbf{B}) + \mathbf{B}\times(\boldsymbol{\nabla}\times\mathbf{A}) + (\mathbf{A}\cdot\boldsymbol{\nabla})\mathbf{B} + (\mathbf{B}\cdot\boldsymbol{\nabla})\mathbf{A} \tag{2}$$

散度

$$\boldsymbol{\nabla}\cdot(f\mathbf{A}) = f(\boldsymbol{\nabla}\cdot\mathbf{A}) + \mathbf{A}\cdot(\boldsymbol{\nabla}f) \tag{3}$$

$$\boldsymbol{\nabla}\cdot(\mathbf{A}\times\mathbf{B}) = \mathbf{B}\cdot(\boldsymbol{\nabla}\times\mathbf{A}) - \mathbf{A}\cdot(\boldsymbol{\nabla}\times\mathbf{B}) \tag{4}$$

旋度

$$\boldsymbol{\nabla}\times(f\mathbf{A}) = f(\boldsymbol{\nabla}\times\mathbf{A}) + (\boldsymbol{\nabla}f)\times\mathbf{A} \tag{5}$$

$$\boldsymbol{\nabla}\times(\mathbf{A}\times\mathbf{B}) = (\mathbf{B}\cdot\boldsymbol{\nabla})\mathbf{A} - (\mathbf{A}\cdot\boldsymbol{\nabla})\mathbf{B} + \mathbf{A}(\boldsymbol{\nabla}\cdot\mathbf{B}) - \mathbf{B}(\boldsymbol{\nabla}\cdot\mathbf{A}) \tag{6}$$

[17]参考 [12] 相关章节.

二阶导数

$$\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}f)=\nabla^2 f \tag{7}$$

$$\nabla^2(f\mathbf{v})=(\nabla^2 f)\mathbf{v}+f\nabla^2\mathbf{v}+2(\boldsymbol{\nabla}f\cdot\boldsymbol{\nabla})\mathbf{v} \tag{8}$$

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla}f)=\mathbf{0} \tag{9}$$

$$\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\mathbf{v})=0 \tag{10}$$

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{v})=\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{v})-\nabla^2\mathbf{v} \tag{11}$$

复合函数求导

使用链式法则[270] 容易证明以下关系

$$\boldsymbol{\nabla}f[g(\mathbf{r})]=f'[g(\mathbf{r})]\boldsymbol{\nabla}g(\mathbf{r}) \tag{12}$$

$$\boldsymbol{\nabla}\cdot\mathbf{f}[g(\mathbf{r})]=\mathbf{f}'[g(\mathbf{r})]\cdot\boldsymbol{\nabla}g(\mathbf{r}) \tag{13}$$

$$\boldsymbol{\nabla}\times\mathbf{f}[g(\mathbf{r})]=[\boldsymbol{\nabla}\cdot g(\mathbf{r})]\times\mathbf{f}'[g(\mathbf{r})] \tag{14}$$

$$\nabla^2 f[g(\mathbf{r})]=f''[g(\mathbf{r})](\boldsymbol{\nabla}g)^2+f'[g(\mathbf{r})]\nabla^2 g(\mathbf{r}) \tag{15}$$

证明

理论上，我们可以直接根据定义，将各个矢量记为分量的形式证明，但直接写出来非常繁琐．一种简单的记号是使用狄拉克 delta 函数 $\delta_{i,j}$ 和 Levi-Civita 符号[431] $\epsilon_{i,j,k}$，再结合爱因斯坦求和约定证明．

我们另外推荐一种不需要写出分量的推导方法，见“一种矢量算符的运算方法[302]”．

分部积分的高维拓展

预备知识 矢量算符常用公式[304]，牛顿—莱布尼兹公式的高维拓展[314]

以下同一公式中所有体积分 $\int \dots \mathrm{d}V$ 都是在某个有限区域 $\mathcal{V}$ 进行，所有的曲面积分[297] $\int \dots \mathrm{d}\mathbf{s}$ 都在 $\mathcal{V}$ 的边界 $\mathcal{S}$ 上进行，正方向向外.

标量分部积分

这是最接近一元函数分部积分的高维拓展

$$\int (\boldsymbol{\nabla} f) g \,\mathrm{d}V = \oint f g \,\mathrm{d}\mathbf{s} - \int f (\boldsymbol{\nabla} g) \,\mathrm{d}V \tag{1}$$

对于二维情况，面积分变为线积分，方向沿逆时针．一维情况就一元函数分部积分.

证明

把式 1 [304] 两边积分，移项得

$$\int (\boldsymbol{\nabla} f) g \,\mathrm{d}V = \int \boldsymbol{\nabla}(f g) \,\mathrm{d}V - \int f (\boldsymbol{\nabla} g) \,\mathrm{d}V \tag{2}$$

现在只需证明

$$\int \boldsymbol{\nabla}(f g) \,\mathrm{d}V = \oint f g \,\mathrm{d}\mathbf{s} \tag{3}$$

这由式 1 [314] 可证．证毕.

矢量分部积分

$$\int f (\boldsymbol{\nabla} \cdot \mathbf{A}) \,\mathrm{d}V = \oint f \mathbf{A} \cdot \mathrm{d}\mathbf{s} - \int \mathbf{A} \cdot (\boldsymbol{\nabla} f) \,\mathrm{d}V \tag{4}$$

参考 [12]，在电动力学中有应用（见“电场的能量[1169]”）.

证明

把式 3 [304] 两边做体积分

$$\int \boldsymbol{\nabla}\cdot(f\mathbf{A})\,\mathrm{d}V=\int f(\boldsymbol{\nabla}\cdot\mathbf{A})\,\mathrm{d}V+\int \mathbf{A}\cdot(\boldsymbol{\nabla}f)\,\mathrm{d}V \tag{5}$$

由散度定理，左边的体积分变为面积分 $\oint f\mathbf{A}\cdot\mathrm{d}\mathbf{s}$，移项得式 4 . 证毕.

梯度 梯度定理

预备知识 方向导数[279]

在方向导数[279] 中，我们推出方向导数为

$$\frac{\partial f}{\partial n}=\boldsymbol{\nabla}f\cdot\hat{\mathbf{n}} \tag{1}$$

其中 $\boldsymbol{\nabla}f$ 就叫标量函数 f 的**梯度**[18]. 要注意当且仅当 Del 算符 $\boldsymbol{\nabla}$ 作用在标量函数（即函数值是一个数而不是矢量）上时，可以称其为**梯度算符**. 这里的 f 叫做**势函数**. 对于 N 维直角坐标系中的 N 元函数 $f(x_1,x_2\dots x_N)$，其梯度是一个**矢量**函数

$$\boldsymbol{\nabla}f=\sum_{i=1}^{N}\frac{\partial f}{\partial x_i}\hat{\mathbf{x}}_i \tag{2}$$

其中所有的 $\hat{\mathbf{x}}_i$ 组成直角坐标系的正交归一基[386]，现在来看全微分[269] 关系

$$\mathrm{d}f=\sum_{i=1}^{N}\frac{\partial f}{\partial x_i}\,\mathrm{d}x_i \tag{3}$$

在直角坐标系中，位置矢量 $\mathbf{r}$ 的微分为

$$\mathrm{d}\mathbf{r}=\sum_{i=1}^{N}\mathrm{d}x_i\,\hat{\mathbf{x}}_i \tag{4}$$

式 3 可用势函数的梯度和 $\mathrm{d}\mathbf{r}$ 的内积表示

$$\mathrm{d}f=\boldsymbol{\nabla}f\cdot\mathrm{d}\mathbf{r} \tag{5}$$

[18]这里假设 f 在某区域内处处光滑，即所有一阶偏导数处处连续. 这个性质也叫**可微**.

由内积的几何定义[383]可知，从某点出发，若微位移 $\mathrm{d}\mathbf{r}$ 的大小不变，那么当其方向与梯度方向相同时函数增量 $\mathrm{d}f$ 最大；二者方向垂直时，函数增量为零；二者夹角为 θ 时，函数增量等于最大值乘以 $\cos\theta$．所以梯度矢量的方向是函数 f 增加最快的方向，梯度的大小等于该方向的方向导数．注意式 5 和式 1 的关系可以类比一元函数的导数[158]和微分[173]的关系，当函数可微时，二者等效．

现在我们也可以把全微分近似（“全微分[269]”式 6）记为矢量的形式

$$\Delta f \approx \boldsymbol{\nabla} f \cdot \Delta\mathbf{r} \tag{6}$$

用梯度计算曲线（面）的法向量

先以 xy 平面的曲线为例，任意曲线可以用函数 $f(x,y)$ 的**等值线**来表示，即 $f(x,y)=C$（C 为常数）．若从曲线上的某点出发，沿曲线的切线方向取一个微位移 $\mathrm{d}\mathbf{r}=\mathrm{d}x\,\hat{\mathbf{x}}+\mathrm{d}y\,\hat{\mathbf{y}}$，由于 $(x+\mathrm{d}x\,,y+\mathrm{d}y)$ 仍然在等值线上，函数增量 $\mathrm{d}f=0$．代入式 5 得

$$\boldsymbol{\nabla} f \cdot \mathrm{d}\mathbf{r} = 0 \tag{7}$$

即 $f(x,y)$ 的梯度与 $\mathrm{d}\mathbf{r}$ 垂直．所以 $\boldsymbol{\nabla} f(x,y)$ 必定是 (x,y) 点所在等值线的法向量，且指向函数值 C 更大的等值线（因为函数值在梯度方向增加最快）．

极坐标，柱坐标和球坐标中的梯度算符

> **预备知识**　正交曲线坐标系[288]

我们先写出极坐标中函数 $f(r,\theta)$ 的全微分为

$$\mathrm{d}f = \frac{\partial f}{\partial r}\,\mathrm{d}r + \frac{\partial f}{\partial \theta}\,\mathrm{d}\theta \tag{8}$$

再写出极坐标中的微位移为

$$\mathrm{d}\mathbf{r} = \mathrm{d}r\,\hat{\mathbf{r}} + r\,\mathrm{d}\theta\,\hat{\boldsymbol{\theta}} \tag{9}$$

所以为了满足梯度的定义式 5，我们可以把式 8 写为

$$\mathrm{d}f = \frac{\partial f}{\partial r}\,\mathrm{d}r + \frac{1}{r}\frac{\partial f}{\partial \theta}\cdot r\,\mathrm{d}\theta \tag{10}$$

对比式 5，式 9 和式 10 可以得出**极坐标中的梯度算符**为

$$\boldsymbol{\nabla} = \hat{\mathbf{r}}\frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial \theta} \tag{11}$$

同理，柱坐标中的微位移（式 15 [288]）与函数 $f(r,\theta,z)$ 的全微分可以分别表示为

$$\mathrm{d}\mathbf{r} = \mathrm{d}r\,\hat{\mathbf{r}} + r\,\mathrm{d}\theta\,\hat{\boldsymbol{\theta}} + \mathrm{d}z\,\hat{\mathbf{z}} \tag{12}$$

$$\mathrm{d}f = \frac{\partial f}{\partial r}\,\mathrm{d}r + \frac{1}{r}\frac{\partial f}{\partial \theta}\cdot r\,\mathrm{d}\theta + \frac{\partial f}{\partial z}\,\mathrm{d}z \tag{13}$$

所以**柱坐标中的梯度算符**为

$$\boldsymbol{\nabla} = \hat{\mathbf{r}}\frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial \theta} + \hat{\mathbf{z}}\frac{\partial}{\partial z} \tag{14}$$

球坐标也类似，球坐标中的微位移（式 12 [742]）与 $f(r,\theta,\phi)$ 的全微分可以分别表示为

$$\mathrm{d}\mathbf{r} = \mathrm{d}r\,\hat{\mathbf{r}} + r\,\mathrm{d}\theta\,\hat{\boldsymbol{\theta}} + r\sin\theta\,\mathrm{d}\phi\,\hat{\boldsymbol{\phi}} \tag{15}$$

$$\mathrm{d}f = \frac{\partial f}{\partial r}\,\mathrm{d}r + \frac{1}{r}\frac{\partial f}{\partial \theta}\cdot r\,\mathrm{d}\theta + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\cdot r\sin\theta\,\mathrm{d}\phi \tag{16}$$

所以**球坐标中的梯度算符**为

$$\boldsymbol{\nabla} = \hat{\mathbf{r}}\frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial}{\partial \phi} \tag{17}$$

梯度定理

预备知识 线积分[295]，牛顿—莱布尼兹公式[202]

梯度定理：一个标量函数 $f(\mathbf{r})$ 的梯度延任何路径从起点 $\mathbf{r}_i$ 到终点 $\mathbf{r}_f$（角标 i 表示 initial，f 表示 final）线积分的结果等于该函数在末位置的函数值减去初位置的函数值．可以用下式表示

$$\int_{\mathbf{r}_i}^{\mathbf{r}_f} \boldsymbol{\nabla} f(\mathbf{r})\cdot \mathrm{d}\mathbf{r} = f(\mathbf{r}_f) - f(\mathbf{r}_i) \tag{18}$$

梯度定理可以看做是牛顿—莱布尼兹公式[202]

$$\int_a^b f'(x)\,\mathrm{d}x = f(b) - f(a) \tag{19}$$

的拓展，即把一元函数拓展为多元函数，把导函数拓展为梯度函数（在一维情况下，式 18 变为式 19）．所以前者的证明也可以类比后者的证明．

梯度定理的证明

我们先把式 18 路径分为许多首尾相接的小段曲线，则整段曲线的线积分等于所有小曲线的线积分之和. 假设曲线处处光滑，如果每段小曲线都足够短，就可以把它们近似看做线段，且梯度值在上面近似为常矢量. 令第 i 小段的起点和终点分别为 $\mathbf{r}_{i0}, \mathbf{r}_{i1}$，则第 i 段的线积分可近似为

$$\int_{\mathbf{r}_{i0}}^{\mathbf{r}_{i1}} \nabla f(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} \approx \nabla f(\mathbf{r}_i) \cdot \Delta \mathbf{r}_i \tag{20}$$

再利用全微分近似（式 6），上式等于

$$\int_{\mathbf{r}_{i0}}^{\mathbf{r}_{i1}} \nabla f(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} \approx f(\mathbf{r}_{i1}) - f(\mathbf{r}_{i0}) \tag{21}$$

将所有小段的线积分求和得到总的线积分得（注意 $\mathbf{r}_{i1} = \mathbf{r}_{(i+1)0}$）

$$\begin{aligned} \int_C \nabla f(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} &= \sum_{i=1}^{n} \int_{\mathbf{r}_{i0}}^{\mathbf{r}_{i1}} \nabla f(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} \\ &\approx \sum_{i=1}^{n} [f(\mathbf{r}_{i1}) - f(\mathbf{r}_{i0})] = f(\mathbf{r}_f) - f(\mathbf{r}_i) \end{aligned} \tag{22}$$

最后取极限 $n \to \infty$，可使上式精确成立. 证毕.

由梯度求势函数

我们通常把上面的标量函数 $f(\mathbf{r})$ 叫做**势函数**，其地位相当于牛顿—莱布尼兹公式中的原函数. 在这个类比中，既然“对原函数求导”对应“对势函数求梯度”，那么不定积分对应的“通过梯度函数求势函数”又该如何实现呢？

以二维的情况为例，我们可以先指定势函数在某点 $\mathbf{r}_0(x_0, y_0)$ 的值，然后根据式 18，要求势函数任意一点 $\mathbf{r}(x, y)$ 的值，只需从 $\mathbf{r}_0$ 点出发由任意路径线积分到点 $\mathbf{r}$ 即可得到势函数 $f(\mathbf{r})$.

$$f(\mathbf{r}) = f(\mathbf{r}_0) + \int_{\mathbf{r}_0}^{\mathbf{r}} \nabla f(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} \tag{23}$$

计算该线积分一般选取一种简单的路径：即先延从 $\mathbf{r}_0(x_0, y_0)$ 到 $\mathbf{r}_1(x, y_0)$ 的水平线段，再延从 $\mathbf{r}_1(x, y_0)$ 到 $\mathbf{r}(x, y)$ 的竖直线段（当然也可以取中间点为 (x_0, y)）. 若把 ∇f 的两个分量 $\partial f/\partial x$, $\partial f/\partial y$ 简写为 $f_x(x, y), f_y(x, y)$，分关于 x 和 y 的

不定积分记为 $F_x(x,y), F_y(x,y)$，延两个线段的线积分[295]（分别把 x 和 y 作为线积分的参数）分别为

$$\int_{\mathbf{r}_0}^{\mathbf{r}_1} \boldsymbol{\nabla} f(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = \int_{x_0}^{x} f_x(x, y_0)\,\mathrm{d}x + 0 = F_x(x, y_0) - F_x(x_0, y_0) \tag{24}$$

$$\int_{\mathbf{r}_1}^{\mathbf{r}} \boldsymbol{\nabla} f(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = \int_{y_0}^{y} f_y(x, y)\,\mathrm{d}y + 0 = F_y(x, y) - F_y(x, y_0) \tag{25}$$

代回式 23 得势函数为

$$\begin{aligned} f(x,y) &= f(x_0,y_0) + F_x(x,y_0) - F_x(x_0,y_0) + F_y(x,y) - F_y(x,y_0) \\ &= F_y(x,y) - F_y(x,y_0) + F_x(x,y_0) + C \end{aligned} \tag{26}$$

其中 C 为待定常数.

散度 散度定理

预备知识 全微分[269]，曲面积分[297]，流密度[986]

我们在矢量场中取一个闭合曲面 $\mathcal{S}$，其内部空间记为 $\mathcal{V}$. 以向外为正方向，矢量场 $\mathbf{F}(\mathbf{r})$ 在闭合曲面的通量 Φ 可以用以下面积分表示，积分范围默认为 $\mathcal{S}$

$$\Phi = \oint \mathbf{F} \cdot \mathrm{d}\mathbf{s} \tag{1}$$

现在我们把该曲面以其内部一点 $\mathbf{r}$ 为中心按比例不断缩小，若通量与体积 V 的比值存在极限，就把该极限叫做该点的**散度（divergence）**，用 $\boldsymbol{\nabla}\cdot$ 算符[19]记为（下文将介绍）

$$\boldsymbol{\nabla} \cdot \mathbf{F}(\mathbf{r}) \equiv \lim_{V \to 0} \frac{\Phi}{V} \tag{2}$$

若场的分布连续且光滑，则该极限处处存在且与曲面的形状无关[20]，我们就得到了矢量场的散度场（注意是标量场）.

例 1　匀速水流

假设密度不变的水以匀速流动，质量的流密度场 $\mathbf{j}(\mathbf{r})$ 为恒定场. 对于任何一个闭合曲面，流入的流量（负值）和流出的流量（正值）相等，总通量为零. 所以该场的散度处处为零.

[19] 符号 $\boldsymbol{\nabla}$ 的名字为 nabla，作为算符时读作 del，一些教材也会在上方加矢量箭头，原因见下文.

[20] 本书不作证明

例 2　变速水流

在例 1 中，流密度场随 x 坐标线性增加，$\mathbf{j}=(j_0+\alpha x)\hat{\mathbf{x}}$（$\alpha$ 为常数），那么取一个边长为 h 的立方体表面作为闭合曲面，从左侧的流量为 $\Phi_L=-(j_0+\alpha x_0)h^2$，右侧的流量为 $\Phi_R=[j_0+\alpha(x_0+h)]h^2$，其余四个面与 $\mathbf{j}$ 平行，没有流量．闭合曲面的总流量为 $\Phi=\alpha h^3=\alpha V$．根据定义，水流的散度处处为 α．分析可发现该水流中单位体积单位时间必然会凭空产生质量为 α 的水（虽然实际中不可能）．所以散度也叫源密度（**source density**）．

直角坐标系中的散度

若在直角坐标系中给出矢量场

$$\mathbf{F}(x,y,z)=F_x(x,y,z)\hat{\mathbf{x}}+F_y(x,y,z)\hat{\mathbf{y}}+F_z(x,y,z)\hat{\mathbf{z}} \tag{3}$$

令闭合曲面为立方体 $[x,y,z]$-$[x+h,y+h,z+h]$ 的表面．先来考虑 x 方向两个正方形的通量 Φ_x，在点 $\mathbf{r}(x,y,z)$ 附近对 $F_x(\mathbf{r})$ 使用全微分近似[269] 得（为简便书写，以下的函数值和偏导都默认在 $\mathbf{r}$ 处取值）

$$F_x(x+x',y+y',z+z')\approx F_x+\frac{\partial F_x}{\partial x}x'+\frac{\partial F_x}{\partial y}y'+\frac{\partial F_x}{\partial z}z' \tag{4}$$

由于只有 x 方向的场分量对 Φ_x 有贡献，

$$\begin{aligned}\Phi_x&\approx\int_0^h\int_0^h \mathrm{d}y'\,\mathrm{d}z'\times\\&\quad\left[\left(F_x+\frac{\partial F_x}{\partial x}h+\frac{\partial F_x}{\partial y}y'+\frac{\partial F_x}{\partial z}z'\right)-\left(F_x+\frac{\partial F_x}{\partial x}0+\frac{\partial F_x}{\partial y}y'+\frac{\partial F_x}{\partial z}z'\right)\right]\\&=\frac{\partial F_x}{\partial x}h\int_0^h\int_0^h \mathrm{d}y'\,\mathrm{d}z'=\frac{\partial F_x}{\partial x}h^3=\frac{\partial F_x}{\partial x}V\end{aligned} \tag{5}$$

同理可以得到另外四个正方形的通量．六个正方形的总通量为

$$\Phi\approx\left(\frac{\partial F_x}{\partial x}+\frac{\partial F_y}{\partial y}+\frac{\partial F_z}{\partial z}\right)V \tag{6}$$

根据定义式 2，可得直角坐标中的散度公式

$$\boldsymbol{\nabla}\cdot\mathbf{F}=\lim_{V\to0}\frac{\Phi}{V}=\frac{\partial F_x}{\partial x}+\frac{\partial F_y}{\partial y}+\frac{\partial F_z}{\partial z} \tag{7}$$

从形式上，我们可以引入一个 $\boldsymbol{\nabla}$ 算符，在直角坐标系中的形式为

$$\boldsymbol{\nabla} = \hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z} \tag{8}$$

那么 $\boldsymbol{\nabla}\cdot\mathbf{F}(\mathbf{r})$ 从形式上可以看做矢量算符 $\boldsymbol{\nabla}$ 与某点场矢量 $\mathbf{F}(\mathbf{r})$ 的“内积”. 根据式 7，显然散度是一个**线性算符**，即多个矢量场的线性组合的散度等于它们分别求散度再线性组合

$$\boldsymbol{\nabla}\cdot[C_1\mathbf{F}_1(\mathbf{r}) + C_2\mathbf{F}_2(\mathbf{r}) + ...] = C_1\boldsymbol{\nabla}\cdot\mathbf{F}_1(\mathbf{r}) + C_2\boldsymbol{\nabla}\cdot\mathbf{F}_2(\mathbf{r}) + \dots \tag{9}$$

例 3　质点引力场的散度

令质点 m 在坐标原点，则它的引力场（式 2 [1027]）为

$$\mathbf{g}(\mathbf{r}) = -\frac{Gm}{r^3}\mathbf{r} \tag{10}$$

我们在直角坐标系中计算该场的散度. 直角坐标系中，有 $\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}$ 和 $r = \sqrt{x^2+y^2+z^2}$，代入上式再求散度，得

$$\begin{aligned}\boldsymbol{\nabla}\cdot\mathbf{g} &= \frac{Gm}{(x^2+y^2+z^2)^{5/2}}[(2x^2-y^2-z^2) + (2y^2-z^2-x^2) + (2z^2-x^2-y^2)] \\ &= 0\end{aligned} \tag{11}$$

可见引力场的散度为 0. 然而需要注意的是，在坐标原点处式 10 的各个方向的偏导都不存在，所以不能用该公式计算散度. 以上的结论只适用于原点之外的点.

散度定理

我们来考虑一个有限大的闭合曲面并计算通量 Φ. 我们先把曲面内的空间划分成许多体积足够小的微元，第 i 个的体积微元为 V_i，通量为 $\Phi_i \approx \boldsymbol{\nabla}\cdot\mathbf{F}(\mathbf{r}_i)V_i$. 现在来证明所有小曲面的通量之和等于大曲面的通量. 图 1 中所有微元的曲面可划分为两部分，一是相邻两个小曲面的边界（红色），二是小曲面与大曲面重合的部分（黑色）. 前者产生的通量之和为零，因为这些边界都是由正方向相反的两块小曲面重合而成，它们产生的通量等大反向，互相抵消. 后者产生的通量等于大曲面的通量，这是因为每块黑色边界都是由正方向相同的小曲面和大曲面重合而成，产生的通量等大同向. 所以总通量等于

$$\Phi = \sum_i \Phi_i \approx \sum_i \boldsymbol{\nabla}\cdot\mathbf{F}(\mathbf{r}_i)V_i \tag{12}$$

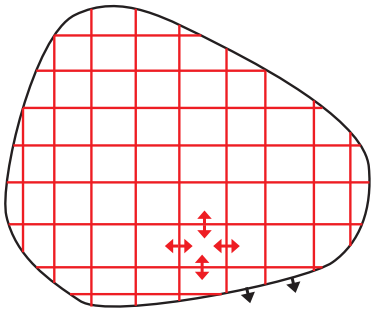


图 1: 证明所有小曲面的通量之和等于大曲面的通量

令微元趋近无穷小，上面的求和变为定积分（积分范围默认为 $\mathcal{V}$ ）

$$\Phi = \oint \mathbf{F} \cdot \mathrm{d}\mathbf{s} = \int \boldsymbol{\nabla} \cdot \mathbf{F}(\mathbf{r})\, \mathrm{d}V \tag{13}$$

所以**散度定理**就是，矢量场在任意闭合曲面的通量等于矢量场的散度在曲面所围空间的体积分.

牛顿—莱布尼兹公式的高维拓展

预备知识 散度散度定理[311]

作为牛顿—莱布尼兹公式[202] 的一个高维拓展，有

$$\int \boldsymbol{\nabla} f \, \mathrm{d}V = \oint f \, \mathrm{d}\mathbf{s} \tag{1}$$

该公式类似于散度定理[311]，但被积函数变为标量而不是矢量. 对于一维情况，该式就是牛顿—莱布尼兹公式.

事实上梯度定理（式 18 [307]）也可以看作是另一种拓展高维拓展.

证明

我们可以对每个分量依次证明. 两边乘以第 i 个分量的单位矢量 $\hat{\mathbf{x}}_i$ 得

$$\int \frac{\partial}{\partial x_i} f \, \mathrm{d}V = \oint (f\hat{\mathbf{x}}_i) \cdot \mathrm{d}\mathbf{s} \tag{2}$$

由散度定理得

$$\oint (f\hat{\mathbf{x}}_i) \, \mathrm{d}\mathbf{s} = \int \boldsymbol{\nabla} \cdot (f\hat{\mathbf{x}}_i) \, \mathrm{d}V = \int \frac{\partial}{\partial x_i} f \, \mathrm{d}V \tag{3}$$

证毕.

旋度

预备知识 圆周运动的速度[883]，散度[311]，线积分[295]

我们在三维空间的矢量场 $\mathbf{F}(\mathbf{r})$ 中取一个闭合回路 $\mathcal{L}$ 并规定一个正方向，并定义该回路的**环流量**为矢量场在回路上的线积分

$$\oint_{\mathcal{L}} \mathbf{F}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} \tag{1}$$

以下假设这个矢量场在某个区域内具有处处具有一阶偏导数.

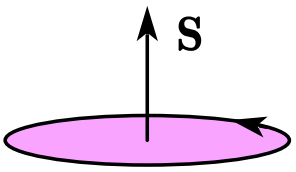


图 1: 面元，法向量为 $\mathbf{S}$，面积为 $|\mathbf{S}|$

现在来定义该矢量场的**旋度（curl）**. 旋度是一个矢量，记为 $\boldsymbol{\nabla} \times \mathbf{F}$. 如图 1，在空间某点 (x, y, z) 处选取一个小面元 $\mathbf{S}$（模长为面元的面积，方向为面元的一个法向量），令面元边界构成的回路为 $\mathcal{L}$，正方向由右手定则[20] 判断. 要定义旋度 $\boldsymbol{\nabla} \times \mathbf{F}$ 在 x 方向的分量，就取 $\mathbf{S}$ 与 x 轴单位矢量 $\hat{\mathbf{x}}$ 同向，再计算换流量除以面积的极限，即

$$\hat{\mathbf{x}} \cdot (\boldsymbol{\nabla} \times \mathbf{F}) = \lim_{S \to 0} \frac{1}{S} \oint_{\mathcal{L}} \mathbf{F}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} \tag{2}$$

同理，要计算旋度的 y, z 分量就把 $\mathbf{S}$ 分别指向单位矢量 $\hat{\mathbf{y}}, \hat{\mathbf{z}}$ 的方向，再讲上式的 $\hat{\mathbf{x}}$ 分别替换为 $\hat{\mathbf{y}}, \hat{\mathbf{z}}$. 这种定义方法叫做“积分—比—极限”.

可以证明，旋度处处存在且与回路的形状选取无关[21]. 所以选取任意方向的面元 $\mathbf{S}$，都有

$$(\boldsymbol{\nabla} \times \mathbf{F}) \cdot \hat{\mathbf{S}} = \lim_{S \to 0} \frac{1}{S} \oint_{\mathcal{L}} \mathbf{F}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} \tag{3}$$

[21] 本书不作证明

直角坐标系中的旋度

在直角坐标系中给出矢量场

$$\mathbf{F}(x,y,z)=F_x(x,y,z)\hat{\mathbf{x}}+F_y(x,y,z)\hat{\mathbf{y}}+F_z(x,y,z)\hat{\mathbf{z}} \tag{4}$$

在点 (x,y,z) 附近，我们可以对场使用微分近似（式 6 [269]）

$$F_i(x+x',y+y',z+z')=F_i(x,y,z)+\frac{\partial F_i}{\partial x}x'+\frac{\partial F_i}{\partial y}y'+\frac{\partial F_i}{\partial z}z' \tag{5}$$

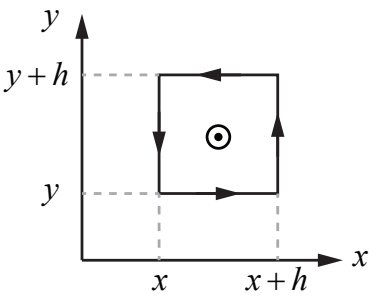


图 2: 直角坐标系中旋度的 z 分量

要求 z 方向的旋度，令闭合回路为图 2 所示的正方形，延 x 方向的两条边的线积分仅由 F_x 贡献，延 y 方向的两条边的线积分仅由 F_y 贡献，所以整个环路的线积分为

$$\begin{aligned}\oint_{\mathcal{L}}\mathbf{F}\cdot\mathrm{d}\mathbf{r}&=\int_0^h\left(F_x+\frac{\partial F_x}{\partial x}x'\right)\mathrm{d}x'-\int_0^h\left(F_x+\frac{\partial F_x}{\partial x}x'+\frac{\partial F_x}{\partial y}h\right)\mathrm{d}x'\\&\quad+\int_0^h\left(F_y+\frac{\partial F_y}{\partial x}h+\frac{\partial F_y}{\partial y}y'\right)\mathrm{d}y'-\int_0^h\left(F_y+\frac{\partial F_y}{\partial y}y'\right)\mathrm{d}y'\\&=h^2\left(\frac{\partial F_y}{\partial x}-\frac{\partial F_x}{\partial y}\right)\end{aligned} \tag{6}$$

所以旋度的 z 分量为

$$G_z=\lim_{h^2\to0}\frac{1}{h^2}\oint_{\mathcal{L}}\mathbf{F}\cdot\mathrm{d}\mathbf{r}=\frac{\partial F_y}{\partial x}-\frac{\partial F_x}{\partial y} \tag{7}$$

类似地，我们可得 x,y 分量

$$G_x=\frac{\partial F_z}{\partial y}-\frac{\partial F_y}{\partial z}\qquad G_y=\frac{\partial F_x}{\partial z}-\frac{\partial F_z}{\partial x} \tag{8}$$

所以类似叉乘的行列式表示（式 15 [387]），我们可以将旋度记为

$$\boldsymbol{\nabla}\times\mathbf{F}=\begin{vmatrix}\hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}}\\ \partial/\partial x & \partial/\partial y & \partial/\partial z\\ F_x & F_y & F_z\end{vmatrix} \tag{9}$$

现在我们知道为什么旋度要记为 $\boldsymbol{\nabla}\times\mathbf{F}$ 了，类比散度，旋度可以从形式上理解为矢量算符 $\boldsymbol{\nabla}$ 与矢量场 $\mathbf{F}$ 的叉乘.

与梯度和散度不同的是，以上定义的旋度运算只能对三维空间的矢量场作用.

例 1　旋转体速度场的旋度

一个物体绕 z 轴旋转，角速度矢量为 $\boldsymbol{\omega}=\omega\hat{\mathbf{z}}$，物体上任意一点的位矢为 $\mathbf{r}$，则速度关于位置的函数 $\mathbf{v}(\mathbf{r})$ 构成一个矢量场（式 5 [883]）

$$\mathbf{v}(\mathbf{r})=\boldsymbol{\omega}\times\mathbf{r}=\omega\hat{\mathbf{z}}\times(x\hat{\mathbf{x}}+y\hat{\mathbf{y}})=-\omega y\hat{\mathbf{x}}+\omega x\hat{\mathbf{y}} \tag{10}$$

使用式 9 计算 $\mathbf{v}(\mathbf{r})$ 的散度，得

$$\boldsymbol{\nabla}\times\mathbf{v}=\begin{vmatrix}\hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}}\\ \partial/\partial x & \partial/\partial y & \partial/\partial z\\ -\omega y & \omega x & 0\end{vmatrix}=2\omega\hat{\mathbf{z}} \tag{11}$$

可见该场的旋度是一个 $\hat{\boldsymbol{\omega}}$ 方向的常矢量. 从这个例子也可以看出，如果一个（三维）矢量场在某个方向没有分量（即平面场），则其旋度必然延该方向（即平面的法向量）.

例 2　无旋度的旋转场

现在我们来看另一个旋转场 $\mathbf{F}(\mathbf{r})=\hat{\mathbf{z}}\times\hat{\mathbf{r}}/r$，写成分量的形式就是

$$\mathbf{F}(\mathbf{r})=\hat{\mathbf{z}}\times\left(\frac{x}{x^2+y^2}\hat{\mathbf{x}}+\frac{y}{x^2+y^2}\hat{\mathbf{y}}\right)=-\frac{y}{x^2+y^2}\hat{\mathbf{x}}+\frac{x}{x^2+y^2}\hat{\mathbf{y}} \tag{12}$$

由于这个场也是一个 xy 平面场，旋度 $\hat{\mathbf{z}}$ 共线，可以直接使用式 7 计算

$$\boldsymbol{\nabla}\times\mathbf{F}=\left(\frac{\partial F_y}{\partial x}-\frac{\partial F_x}{\partial y}\right)\hat{\mathbf{z}}=\left(\frac{1}{r^2}-\frac{2x^2}{r^4}+\frac{1}{r^2}-\frac{2y^2}{r^4}\right)\hat{\mathbf{z}}=\mathbf{0} \tag{13}$$

要注意的是，在原点处由于矢量场不连续（而是出现了无限大的奇点），以上计算在原点处并不成立.

斯托克斯定理

- 本词条存在未完成的内容.

预备知识 旋度[315]

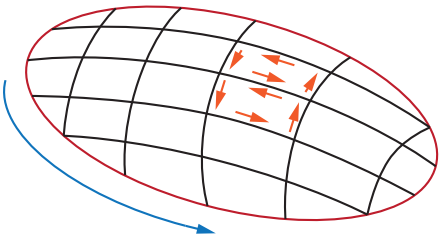


图 1: 斯托克斯定理

如图 1，我们选取一块曲面，并规定一个正方向．使用右手定则[20]，我们也可以定义曲面边界的正方向．空间中存在连续光滑的矢量场 $\mathbf{F}(\mathbf{r})$，则**斯托克斯定理**可以将矢量场在曲面边界上的环流量和矢量场的旋度在曲面上通量等同起来

$$\oint \mathbf{F}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = \int \boldsymbol{\nabla} \times \mathbf{F}(\mathbf{r}) \cdot \mathrm{d}\mathbf{s} \tag{1}$$

要证明这个定理，我们将曲面划分为许多小面元 $\Delta\mathbf{s}_i$，其正方向与曲面一致，边界的正方向同样由右手定则定义．这样，矢量场在曲面上的通量就等于在每个小面元上的通量之和．当面元的面积趋于零时，我们可以认为场的旋度在面元上是常矢量 $\mathbf{F}(\mathbf{r}_i)$，$\mathbf{r}_i$ 为 $\Delta\mathbf{s}_i$ 上任意一点．由式 3 [315] 可知面元的环流量为 $\boldsymbol{\nabla} \times \mathbf{F}(\mathbf{r}_i) \cdot \Delta\mathbf{s}_i$（可类比式 2 [173]），所以根据积分的思想，所有面元的环流量之和为

$$\lim_{\Delta s_i \to 0} \sum_i \boldsymbol{\nabla} \times \mathbf{F}(\mathbf{r}_i) \cdot \Delta\mathbf{s}_i = \int \boldsymbol{\nabla} \times \mathbf{F}(\mathbf{r}) \cdot \mathrm{d}\mathbf{s} \tag{2}$$

最后，如何证明所有面元的环流量之和等于曲面边界的环流量呢？类比散度定理（式 13 [311]）的证明，考虑任意两块相邻的小面元，矢量场在它们共同边界的线积分对一个面元的环流量贡献为正，而对另一个面元的环流量贡献大小相同但符号为负，所以在上式的求和中相加为零．所以，求和中唯一没有被抵消

的环流量来自于曲面边界处的面元，这些面元的边界与曲面边界重合且正方向一致，对求和的贡献恰好等于曲面边界的环流量．证毕．

未完成：例题

调和场

- 本词条处于草稿阶段．

预备知识 调和函数（未完成）

我们把散度和旋度都为零的场称为**调和场**．注意这并不是一个常用的数学名词，笔者只在个别中文教材中见过．由于旋度为零，积分与路径无关，必定可以定义势函数 u，而调和场就是其梯度

$$\mathbf{f}(\mathbf{r}) = \boldsymbol{\nabla} \cdot u \tag{1}$$

要保证散度 $\boldsymbol{\nabla} \cdot f$ 为零，就要求 u 是一个调和函数：

$$\nabla^2 u = 0 \tag{2}$$

所以调和场的充分必要条件是它可以表示为一个调和函数的梯度．注意对于同一个 $\mathbf{f}(\mathbf{r})$，$u(\mathbf{r})$ 并不是唯一的，可以给 u 加上一个任意常数．

矢量分析总结

- 待整合

矢量分析公式列表

令 $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ 为 $\mathbb{R}^3$ 中的可微矢量场，ϕ 为 $\mathbb{R}^3$ 中的可微标量场，∇ 为矢量微分算子 $\hat{\mathbf{x}}\partial/\partial x + \hat{\mathbf{y}}\partial/\partial y + \hat{\mathbf{z}}\partial/\partial z$．

$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})=\mathbf{b}\cdot(\mathbf{c}\times\mathbf{a})=\mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}) \tag{1}$$

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c})=(\mathbf{a}\cdot\mathbf{c})\mathbf{b}-(\mathbf{a}\cdot\mathbf{b})\mathbf{c} \tag{2}$$

$$(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{c}\times\mathbf{d})=(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})-(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c}) \tag{3}$$

$$\nabla\times(\nabla\phi)=\mathbf{0} \tag{4}$$

$$\nabla\cdot(\nabla\times\mathbf{a})=0 \tag{5}$$

$$\nabla\times(\nabla\times\mathbf{a})=\nabla(\nabla\cdot\mathbf{a})-(\nabla\cdot\nabla)\mathbf{a} \tag{6}$$

$$\nabla\cdot(\phi\mathbf{a})=(\nabla\phi)\cdot\mathbf{a}+\phi\nabla\cdot\mathbf{a} \tag{7}$$

$$\nabla\times(\phi\mathbf{a})=(\nabla\phi)\times\mathbf{a}+\phi\nabla\times\mathbf{a} \tag{8}$$

$$\nabla(\mathbf{a}\cdot\mathbf{b})=(\mathbf{a}\cdot\nabla)\mathbf{b}+\mathbf{b}\times(\nabla\times\mathbf{a})+(\mathbf{b}\cdot\nabla)\mathbf{a}+\mathbf{a}\times(\nabla\times\mathbf{b}) \tag{9}$$

$$\nabla\cdot(\mathbf{a}\cdot\mathbf{b})=\mathbf{b}\cdot(\nabla\times\mathbf{a})-\mathbf{a}\cdot(\nabla\times\mathbf{b}) \tag{10}$$

$$\nabla\times(\mathbf{a}\times\mathbf{b})=(\mathbf{b}\cdot\nabla)\mathbf{a}-(\mathbf{a}\cdot\nabla)\mathbf{b}+\mathbf{a}(\nabla\cdot\mathbf{b})-\mathbf{b}(\nabla\cdot\mathbf{a}) \tag{11}$$

标量场和矢量场

场（**field**）是指实线性空间 V[22]

标量场

$$\Phi=\Phi(x,y,z) \tag{12}$$

Φ 的数值是空间位置的函数等值面

$$\Phi=C \tag{13}$$

例如气压场、温度场.

[22]或者更一般地，流形 M. 实线性空间是**流形**[688] 的特例.

矢量场（详见??）

$$\mathbf{A} = \mathbf{A}(x, y, z) \tag{14}$$

$\mathbf{A}$ 的大小、方向是空间位置的函数．例如速度场 $\mathbf{v}$、电场 $\mathbf{E}$.

场线：有方向的曲线，其上每一点切线方向都与 $\mathbf{A}$ 的方向一致.

场管：由一束场线围城的管状区域.

标量场的梯度

方向微商

$$\frac{\partial \Phi}{\partial l} = \lim_{\Delta l \to 0} \frac{\Delta \Phi}{\Delta l} \tag{15}$$

标量场 Φ 在 P 点沿 $\Delta \mathbf{l}$ 方向的方向微商.

标量场 Φ 的梯度

沿方向微商最大的方向（即 $\Delta \mathbf{n}$ 方向）.

$$\boldsymbol{\nabla} \Phi = \frac{\partial \Phi}{\partial n} \tag{16}$$

$\Delta \Phi$ 方向总于 Φ 的等值面垂直.

标量场的梯度是矢量场

电势 U 是标量场，其负梯度 $\mathbf{E}$ 是矢量场.

矢量场的通量和散度 高斯定理

定义

通量

$$\Phi_A = \iint_{(S)} \mathbf{A}\, \mathrm{d}\mathbf{S} = \iint_{(S)} A \cos\theta dS \tag{17}$$

流速场、流量、电通量、磁通量.

散度

$$div\mathbf{A} = \nabla \cdot \mathbf{A} = \lim_{\Delta V \to 0} \frac{\Phi}{\Delta V} \tag{18}$$

矢量场的散度是标量场.

坐标表示

$$\nabla \cdot \mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \tag{19}$$

$$\oint \mathbf{A} = \iiint_V \nabla \cdot \mathbf{A} dV \tag{20}$$

矢量场通过任意闭合曲面 S 的通量等于它向包围体积 V 内的散度积分.

矢量场的环量和旋度 Stokes 定理

定义

环量 Γ

$$\Gamma = \oint_L \mathbf{A} \cdot d\mathbf{l} \tag{21}$$

矢量场 $\mathbf{A}$ 沿闭合回路线积分.

δS 为闭线 L 包围面积，$\mathbf{n}$ 为右旋单位法向量.

旋度 $\boldsymbol{\nabla} \times \mathbf{A}$

$$rot\mathbf{A} = \nabla \times \mathbf{A} = \lim_{\Delta S \to 0} \frac{\oint_L \mathbf{A} d\mathbf{l}}{\Delta S} \tag{22}$$

矢量场的旋度仍是矢量场.

坐标表示

$$\nabla \times \mathbf{A} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix} \tag{23}$$

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial}y + \mathbf{k}\frac{\partial}{\partial z} \tag{24}$$

stokes 定理

$$\oint_L \mathbf{A} d\mathbf{l} = \iint_S (\nabla \times \mathbf{A}) \cdot d\mathbf{S} \tag{25}$$

矢量场在任意闭合回路 $\mathbf{L}$ 上的环量等于它为边界的曲面 $\mathbf{S}$ 上旋度的积分.

矢量场的分类

有散场和无散场

散度为 0，即无源，为无散场；散度不为 0，即有源，有散场.

$$\nabla \cdot \nabla \times \mathbf{A} = 0 \tag{26}$$

知，任何矢量场的旋度永远是无散场.

任何无散场 $\mathbf{B}$ 可表达成某矢量场的旋度

$$\mathbf{B} = \nabla \times \mathbf{A}, \nabla \cdot \mathbf{B} = 0 \tag{27}$$

有旋场和无旋场

旋度为 0，为无旋场; 反之为有旋场.

$$\nabla \times \nabla \Phi = 0 \tag{28}$$

任何标量场的梯度永远是无旋场.

任何无旋场 $\mathbf{A}$ 可表示为某个标量场 Φ 的梯度

$$\mathbf{A} = \nabla \Phi, \nabla \times \mathbf{A} = 0 \tag{29}$$

谐和场

谐和场为某一矢量 $\mathbf{A}$ 在某空间内既无散又无旋，由于其无旋，所以可以由势场表示：

$$\mathbf{A} = \nabla \Phi, \nabla \times \mathbf{A} = 0 \tag{30}$$

同样由于其为无散场，所以有：

$$\mathbf{B} = \nabla \times \mathbf{A}, \nabla \cdot \mathbf{B} = 0 \tag{31}$$

故可以导出 Laplace 方程：

$$\nabla \cdot \nabla = \nabla^2 \tag{32}$$

谐和场的势函数满足 Laplace 方程.

散度的逆运算

预备知识 矢量算符运算法则[304]

定理 1

令 $V(\mathbf{r})$ 为任意标量场，则 $V(\mathbf{r})$ 总能表示为一个矢量场 $\mathbf{F}(\mathbf{r})$ 的散度，即

$$V(\mathbf{r}) = \boldsymbol{\nabla} \cdot \mathbf{F}(\mathbf{r}) \tag{1}$$

且 $\mathbf{F}(\mathbf{r})$ 可以通过以下公式计算：

$$\mathbf{F}(\mathbf{r}) \equiv \frac{1}{4\pi} \int \frac{V(\mathbf{r}')\mathbf{R}}{R^3} \, \mathrm{d}V' \tag{2}$$

其中 $\mathbf{r}, \mathbf{r}'$ 分别是坐标原点指向三维直角坐标 (x, y, z) 和 (x', y', z') 的位置矢量，$\mathbf{R} = \mathbf{r}' - \mathbf{r}$，$R = |\mathbf{R}|$，体积分 $\int \mathrm{d}V' = \int \mathrm{d}x' \, \mathrm{d}y' \, \mathrm{d}z'$ 的区域是整个三维空间或者空间中 $\mathbf{F}$ 不为零的区域，$\times$ 表示矢量叉乘[387].

若在式 2 右边加上任意无散场 $\mathbf{H}(\mathbf{r})$，定理同样成立.

式 2 可以看做是散度运算的逆运算，类似于不定积分[187] 是求导的逆运算. $\mathbf{H}(\mathbf{r})$ 则相当于不定积分中的任意常数. 旋度也有类似的逆运算[325].

该定理在物理中可应用于电场和电荷的关系，定理中的 $\mathbf{F}$ 可看做电场 $\mathbf{E}$，而 $V(\mathbf{r})$ 可看做电荷密度 $\rho(\mathbf{r})$ 除以真空介电常数 ϵ_0. 于是式 1 就是电场的高斯定律（式 2 [1161]），式 2 就是库伦定律（式 6 [1124]）.

定理 2

令 $V(\mathbf{r})$ 为一个标量场，则式 2 得到的 $\mathbf{F}(\mathbf{r})$ 是一个无旋场，即 $\boldsymbol{\nabla} \times \mathbf{F} = \mathbf{0}$.

作为静电学的一个应用，如果仍然将 $\mathbf{F}$ 看做电场，那么该定理说明由静止电荷产生的电场是无旋场.

证明定理 1

对式 2 求散度，得

$$\boldsymbol{\nabla}\cdot\mathbf{F}(\mathbf{r})=\frac{1}{4\pi}\int\boldsymbol{\nabla}\cdot\frac{V(\mathbf{r}')\mathbf{R}}{R^3}\,\mathrm{d}V' \tag{3}$$

使用式 3 [300]，得

$$\boldsymbol{\nabla}\cdot\frac{V(\mathbf{r}')\mathbf{R}}{R^3}=[\boldsymbol{\nabla}\cdot V(\mathbf{r}')]\frac{\mathbf{R}}{R^3}+V(\mathbf{r}')\left(\boldsymbol{\nabla}\cdot\frac{\mathbf{R}}{R^3}\right) \tag{4}$$

由于 $V(\mathbf{r}')$ 是 x',y',z' 而不是 x,y,z 的函数，右边第一项中 $\boldsymbol{\nabla}\cdot V(\mathbf{r}')=0$．第二项中（链接未完成）

$$\boldsymbol{\nabla}\cdot\frac{\mathbf{R}}{R^3}=4\pi\delta(\mathbf{R}) \tag{5}$$

所以式 4 代入式 3 得

$$\boldsymbol{\nabla}\cdot\mathbf{F}(\mathbf{r})=\int\boldsymbol{\nabla}\cdot V(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}')\,\mathrm{d}V'=V(\mathbf{r}) \tag{6}$$

证毕．

证明定理 2

对式 2 求旋度，得

$$\boldsymbol{\nabla}\times\mathbf{F}(\mathbf{r})\equiv\frac{1}{4\pi}\int\boldsymbol{\nabla}\times\frac{V(\mathbf{r}')\mathbf{R}}{R^3}\,\mathrm{d}V' \tag{7}$$

由式 5 [304] 得

$$\boldsymbol{\nabla}\times\frac{V(\mathbf{r}')\mathbf{R}}{R^3}=\boldsymbol{\nabla}V(\mathbf{r}')\times\frac{\mathbf{R}}{R^3}+V(\mathbf{r}')\boldsymbol{\nabla}\times\frac{\mathbf{R}}{R^3} \tag{8}$$

由于 $V(\mathbf{r}')$ 是 x',y',z' 而不是 x,y,z 的函数，右边第一项中 $\boldsymbol{\nabla}\times V(\mathbf{r}')=\mathbf{0}$，第二项中（链接未完成）

$$\boldsymbol{\nabla}\times\frac{\mathbf{R}}{R^3}=\mathbf{0} \tag{9}$$

所以式 7 中积分为零．证毕．

旋度的逆运算

- 本词条存在未完成的内容.

预备知识 矢量算符运算法则[304]

定理 1

[23]令 $\mathbf{F}$ 为无散场，即

$$\boldsymbol{\nabla}\cdot\mathbf{F}=0 \tag{1}$$

则 $\mathbf{F}(\mathbf{r})$ 总能表示为另一个矢量场 $\mathbf{G}(\mathbf{r})$ 的旋度，即

$$\mathbf{F}=\boldsymbol{\nabla}\times\mathbf{G} \tag{2}$$

且 $\mathbf{G}(\mathbf{r})$ 可以通过以下公式计算：

$$\mathbf{G}(\mathbf{r})=\frac{1}{4\pi}\int\mathbf{F}(\mathbf{r}')\times\frac{\mathbf{R}}{R^3}\,\mathrm{d}V' \tag{3}$$

其中 $\mathbf{G}$ 通常被称为矢势（**vector potential**），$\mathbf{r},\mathbf{r}'$ 分别是坐标原点指向三维直角坐标 (x,y,z) 和 (x',y',z') 的位置矢量，$\mathbf{R}=\mathbf{r}'-\mathbf{r}$，$R=|\mathbf{R}|$，体积分 $\int\mathrm{d}V'=\int\mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z'$ 的区域是空间中 $\mathbf{F}$ 不为零的区域，$\times$ 表示矢量叉乘[387].

该定理在电动力学中有两个重要的应用：一个是证明比奥萨法尔定律满足安培环路定理[1179]，另一个是证明磁矢势[1215] $\mathbf{A}$ 必定存在（因为磁场是无散场[1169]）.

推论 1

在定理 1 中，给 $\mathbf{G}(\mathbf{r})$ 加上任意一个无旋场 $\mathbf{H}(\mathbf{r})$（满足 $\boldsymbol{\nabla}\times\mathbf{H}=\mathbf{0}$），也能使式 2 成立. 通过这种方法可以得到式 2 中所有可能的 $\mathbf{G}(\mathbf{r})$.

注意 $\mathbf{H}$ 也可以表示为任意标量函数 $V(\mathbf{r})$ 的梯度 $\boldsymbol{\nabla}V$（引用未完成）.

我们可以认为式 3 是旋度运算的逆运算，这可以类比不定积分[187] 是求导的逆运算. 而无旋场 $\mathbf{H}(\mathbf{r})$ 可以类比不定积分中的任意常数. 散度也有类似的逆运算[324].

定理 2

令 $\mathbf{F}(\mathbf{r})$ 为无散场，则式 3 得到的 $\mathbf{G}(\mathbf{r})$ 仍然是一个无散场.

[23]参考 [12] 相关章节.

显然，给 $\mathbf{G}(\mathbf{r})$ 加上一个任意的无散场 $\mathbf{H}(\mathbf{r})$ 后仍然是一个无散场．若需要满足式 2，则 $\mathbf{H}(\mathbf{r})$ 必须也是无旋的，即 $\mathbf{H}(\mathbf{r})$ 是调和场．

我们已经知道一个矢量场无论求几次旋度，都一直是无散场．而这个定理告诉我们任意无散场无论求几次“逆旋度”也都可以是无散场．

证明定理 1

我们只需要证明式 3 右边第一项求旋度等于 $\mathbf{F}$，使用式 6 [304]（展开的四项中对 $\mathbf{F}$ 微分的两项为零，因为 $\mathbf{F}$ 是 $\mathbf{r}'$ 的函数）：

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{G}&=\frac{1}{4\pi}\int\boldsymbol{\nabla}\times\left(\mathbf{F}\times\frac{\mathbf{R}}{R^3}\right)\mathrm{d}V'\\&=\frac{1}{4\pi}\int\mathbf{F}\left(\boldsymbol{\nabla}\cdot\frac{\mathbf{R}}{R^3}\right)\mathrm{d}V'-\frac{1}{4\pi}\int(\mathbf{F}\cdot\boldsymbol{\nabla})\frac{\mathbf{R}}{R^3}\,\mathrm{d}V'\end{aligned}\tag{4}$$

其中第一个等号是因为“对一个变量积分”再“对另一个变量求导”这两个操作可以交换．

先来证明上式第二项为零：被积函数中的 x 分量为（$\boldsymbol{\nabla}'$ 意味着对 x',y',z' 求偏导，另外注意 $\mathbf{F}$ 是 $\mathbf{r}'$ 的函数）

$$\begin{aligned}(\mathbf{F}\cdot\boldsymbol{\nabla})\frac{x'-x}{|x'-x|^3}&=\mathbf{F}\cdot\left(\boldsymbol{\nabla}\frac{x'-x}{|x'-x|^3}\right)=-\mathbf{F}\cdot\left(\boldsymbol{\nabla}'\frac{x'-x}{|x'-x|^3}\right)\\&=-\boldsymbol{\nabla}'\cdot\left(\mathbf{F}\frac{x'-x}{|x'-x|^3}\right)+(\boldsymbol{\nabla}'\cdot\mathbf{F})\frac{x'-x}{|x'-x|^3}\end{aligned}\tag{5}$$

这里使用了式 3 [304]．由于 $\mathbf{F}(\mathbf{r}')$ 是无散场，最后一项为零．y,z 分量同理．对上式做体积分得（使用散度定理式 13 [311]）

$$-\int\boldsymbol{\nabla}'\cdot\left(\mathbf{F}\frac{x'-x}{|x'-x|^3}\right)\mathrm{d}V'=-\oint\mathbf{F}\frac{x'-x}{|x'-x|^3}\,\mathbf{ds}\tag{6}$$

积分曲面是体积分区域的边界曲面．由于我们假设 $\mathbf{F}$ 只有在体积分内部（不包括边界）不为零，所以该式为零．

再来看式 4 右边第一项，有（引用未完成）

$$\boldsymbol{\nabla}\cdot\frac{\mathbf{R}}{R^3}=4\pi\delta^3(\mathbf{R})\tag{7}$$

代入得

$$\boldsymbol{\nabla}\times\mathbf{G}=\int\mathbf{F}(\mathbf{r}')\delta^3(\mathbf{r}-\mathbf{r}')\,\mathrm{d}V'=\mathbf{F}(\mathbf{r})\tag{8}$$

证毕．

证明定理 2

使用式 4 [304]，有

$$\begin{aligned}\boldsymbol{\nabla}\cdot\mathbf{F} &= \frac{1}{4\pi}\int \boldsymbol{\nabla}\cdot\left(\mathbf{F}\times\frac{\mathbf{R}}{R^3}\right)\mathrm{d}V' \\ &= \frac{1}{4\pi}\int \frac{\mathbf{R}}{R^3}\cdot(\boldsymbol{\nabla}\times\mathbf{F})\,\mathrm{d}V' - \frac{1}{4\pi}\int \mathbf{F}\cdot\left(\boldsymbol{\nabla}\times\frac{\mathbf{R}}{R^3}\right)\mathrm{d}V'\end{aligned} \tag{9}$$

注意 $\mathbf{F}$ 是 $\mathbf{r}'$ 的函数而不是 $\mathbf{r}$ 的函数，所以第一项中 $\boldsymbol{\nabla}\times\mathbf{F}=\mathbf{0}$. 另外由于 $\boldsymbol{\nabla}\times(\mathbf{R}/R^3)=\mathbf{0}$（引用未完成），上式恒为零. 证毕.

亥姆霍兹分解

- 本词条存在未完成的内容.

预备知识 散度的逆运算[324]，旋度的逆运算[325]

[24]任何矢量场都可以分解为一个无旋场 $\mathbf{F}_d(\mathbf{r})$，一个无散场 $\mathbf{F}_c(\mathbf{r})$，和一个调和场（链接未完成）$\mathbf{H}(\mathbf{r})$.

$$\mathbf{F}(\mathbf{r}) = \mathbf{F}_d(\mathbf{r}) + \mathbf{F}_c(\mathbf{r}) + \mathbf{H}(\mathbf{r}) \tag{1}$$

一种分解的计算方法是，令

$$f(\mathbf{r}) \equiv \boldsymbol{\nabla}\cdot\mathbf{F}(\mathbf{r}) \qquad \mathbf{G}(\mathbf{r}) \equiv \boldsymbol{\nabla}\times\mathbf{F}(\mathbf{r}) \tag{2}$$

定义

$$\mathbf{F}_d(\mathbf{r}) \equiv \frac{1}{4\pi}\int \frac{f(\mathbf{r}')\mathbf{R}}{R^3}\,\mathrm{d}V' \tag{3}$$

$$\mathbf{F}_c(\mathbf{r}) \equiv \frac{1}{4\pi}\int \frac{\mathbf{G}(\mathbf{r}')\times\mathbf{R}}{R^3}\,\mathrm{d}V' \tag{4}$$

$$\mathbf{H}(\mathbf{r}) \equiv \mathbf{F}(\mathbf{r}) - \mathbf{F}_d(\mathbf{r}) - \mathbf{F}_c(\mathbf{r}) \tag{5}$$

[24]参考 [12] 相关章节.

其中 $\mathbf{r},\mathbf{r}'$ 分别是坐标原点指向三维直角坐标 (x,y,z) 和 (x',y',z') 的位置矢量，$\mathbf{R}=\mathbf{r}'-\mathbf{r}$，$R=|\mathbf{R}|$，体积分 $\int \mathrm{d}V'=\int \mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z'$ 的区域是整个三维空间或者其中 $\mathbf{F}$ 不为零的区域，$\times$ 表示矢量叉乘[387].

注意式 1 的分解不是唯一的，我们也可以给式 3 和式 4 右边分别加上一个任意调和场，分解依然成立. 可以证明如此定义的 $\mathbf{F}_d$，$\mathbf{F}_c$ 和 $\mathbf{H}$ 分别是无旋场，无散场和调和场.

式 3 和式 4 可以分别类比静电学中的库伦定律（式 6 [1124]）和静磁学中的比奥萨法尔定律（式 2 [1175]）：把 $\mathbf{F}_c$ 看作静电场，满足高斯定律（式 2 [1161]）；把 $\mathbf{F}_c$ 看作静磁场，满足安培环路定理（式 6 [1179]）.

证明

根据定理 2 [324] 以及定理 2 [325]，式 3 和式 4 中定义的 $\mathbf{F}_d(\mathbf{r})$ 和 $\mathbf{F}_c(\mathbf{r})$ 分别是无旋场和无散场，且满足

$$\boldsymbol{\nabla}\cdot\mathbf{F}_d(\mathbf{r})=f(\mathbf{r})=\boldsymbol{\nabla}\cdot\mathbf{F}(\mathbf{r}) \tag{6}$$

$$\boldsymbol{\nabla}\times\mathbf{F}_c(\mathbf{r})=\mathbf{G}(\mathbf{r})=\boldsymbol{\nabla}\times\mathbf{F}(\mathbf{r}) \tag{7}$$

容易证明 $\mathbf{H}(\mathbf{r})$ 必定是调和场（无散且无旋）：

$$\boldsymbol{\nabla}\cdot\mathbf{H}=\boldsymbol{\nabla}\cdot(\mathbf{F}-\mathbf{F}_d)=0 \tag{8}$$

$$\boldsymbol{\nabla}\times\mathbf{H}=\boldsymbol{\nabla}\times(\mathbf{F}-\mathbf{F}_d)=\mathbf{0} \tag{9}$$

证毕.

另一种表示

上述分解采用的是积分的方法，我们也可以用微分来描述. 无旋场总能表示为某个标量函数 $V(\mathbf{r})$ 的梯度（证明见势能[905]），而无散场总能表示为另一个矢量场 $\mathbf{A}$ 的旋度[325]，所以

$$\mathbf{F}_d=\boldsymbol{\nabla}V\qquad \mathbf{F}_c=\boldsymbol{\nabla}\times\mathbf{A} \tag{10}$$

所以亥姆霍兹分解也可以记为

$$\mathbf{F}=\boldsymbol{\nabla}V+\boldsymbol{\nabla}\times\mathbf{A}+\mathbf{H} \tag{11}$$

但是求解 $V(\mathbf{r})$ 和 $\mathbf{A}$ 并不那么直接，一般还是需要先用上一节中的办法进行分解才能求得.

亥姆霍兹分解 2

- 本词条处于草稿阶段.

预备知识 矢量分析总结[319]

在矢量分析中，三维空间中的**亥姆霍兹分解（Helmholtz decomposition）**可在给定边界条件的情况下，将一个矢量场唯一地分解为无旋场和无散场的和．它在流体力学和电磁学中皆有应用.

全空间的亥姆霍兹分解（傅里叶形式）

设 $F(x)$ 是三维空间 $\mathbb{R}^3$ 上的光滑矢量场，满足

$$\int_{\mathbb{R}^3} |F(x)|^2 dx < \infty.$$

则它的傅里叶变换 $\hat{F}$ 是良好定义的．由此可定义一个标量函数

$$\begin{aligned}\Phi(x) &= \int_{\mathbb{R}^3} \frac{i\xi \cdot \hat{F}(\xi)}{|\xi|^2} e^{i\xi\cdot x} d\xi \\ &= \frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{\nabla_y \cdot F(y)}{|x-y|} dy \\ &= -\frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{(x-y)\cdot F(y)}{|x-y|} dy.\end{aligned}$$

和一个矢量场

$$\begin{aligned}A(x) &= \int_{\mathbb{R}^3} \frac{i\xi \times \hat{F}(\xi)}{|\xi|^2} e^{i\xi\cdot x} d\xi \\ &= \frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{\nabla_y \times F(y)}{|x-y|} dy \\ &= -\frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{(x-y)\times F(y)}{|x-y|} dy.\end{aligned}$$

则有

$$F = -\nabla\Phi + \nabla \times A.$$

这就是矢量场 F 的亥姆霍兹分解.

区域的亥姆霍兹分解（强形式）

区域的亥姆霍兹分解（弱形式）

拉格朗日乘数法

预备知识 导数与函数极值[170]，梯度[307]

若要求函数 $f(x_1,\ldots,x_N)$ 在 m 个约束条件 $g_i(x_1,\ldots,x_N)\ (i=1,\ldots,M)$ 下的极值，可用拉格朗日乘数法，令拉格朗日函数为

$$\Lambda(x_1,\ldots,x_N)=f+\sum_{i=1}^{M}\lambda_i g_i \tag{1}$$

其中 λ_i 为未知任意常数．同时满足约束条件和

$$\frac{\partial\Lambda}{\partial x_i}=0\quad(i=1,\ldots,N) \tag{2}$$

的点，就是函数 $f(x_1,\ldots,x_N)$ 的极值点或稳定点．

几何理解

下面举一个二元函数的例子以说明拉格朗日乘数法的几何意义．

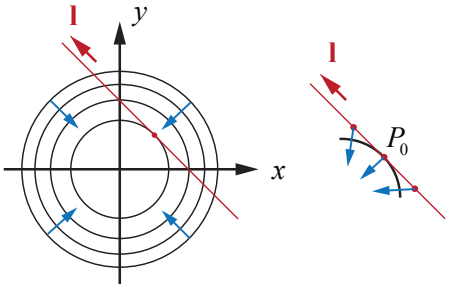


图 1: 例 1

例 1

求函数 $f(x,y)=-a(x^2+y^2)$ 在轨迹 $x+y=c$ 约束下的最值（a,c 均为常数）．

从几何上来理解，函数 $f(x,y)$ 是一个倒置的抛物面，其等高线的俯视图如图 1 蓝色的箭头代表函数值增加最大的方向，即梯度的方向．考察点沿着同一条等高线的位置变化时，函数值不变．曲面上任意一点的梯度为

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial x}\hat{\mathbf{x}} + \frac{\partial f}{\partial y}\hat{\mathbf{y}} = -2ax\hat{\mathbf{x}} - 2ay\hat{\mathbf{y}} = -2a\mathbf{r} \tag{3}$$

其中 $\mathbf{r}$ 为位矢．这说明，梯度的方向总是延径向向内，与距离成正比．

约束条件如图 1 中红线所示，我们要求的就是红线上面函数 f 的极值．在该题的情况下，所求的极值显然在红线离原点最近的地方出现．

但是如何总结出约束条件下多元函数极值点的一般的判断条件并用梯度表示呢？根据梯度的性质（式 5 [307]），如果考察点沿红线的切向移动 $\mathrm{d}\mathbf{l}$，函数值的微分为

$$\mathrm{d}f = \boldsymbol{\nabla} f \cdot \mathrm{d}\mathbf{l} \tag{4}$$

使用导数和极值的关系[170]，我们要找的极值点满足：函数在极值点处延红线切向的方向导数为零．即

$$\boldsymbol{\nabla} f \cdot \mathrm{d}\mathbf{l} = 0 \tag{5}$$

这说明 $\boldsymbol{\nabla} f$ 只可能具有极值点处红线的法向量分量而切向分量为 0，即 $\boldsymbol{\nabla} f$ 与法向量共线．令约束条件即红线的方程为

$$g(x,y) = x + y - c = 0 \tag{6}$$

$g(x,y)$ 的梯度就是其法向量．由矢量共线的定义，存在常数 λ，使

$$\boldsymbol{\nabla} f + \lambda \boldsymbol{\nabla} g = 0 \qquad \text{即} \qquad \boldsymbol{\nabla}(f + \lambda g) = 0 \tag{7}$$

这就是拉格朗日乘数法．其中 $f + \lambda g$ 就是拉格朗日函数．下面根据该条件，对本题进行求解．本题中拉格朗日函数梯度的两个分量分别为

$$\begin{aligned} 0 &= \frac{\partial}{\partial x}(f + \lambda g) = -2ax + \lambda \\ 0 &= \frac{\partial}{\partial y}(f + \lambda g) = -2ay + \lambda \end{aligned} \tag{8}$$

两式消去 λ，得 $x = y$．再结合约束条件 $x + y - c = 0$，得极值点为 $(c/2, c/2)$．可见拉格朗日乘数法求出的偏导条件并没有包含约束条件，而是需要再次联立约束条件才可以得出最后答案．

以上的推导可以很容易地拓展到多元函数和多个限制条件中去．例如求三元函数 $f(x,y,z)$ 中（可看作空间的标量场），在两个约束条件 $g_1(x,y,z)=0, g_2(x,y,z)=0$ 下的极值．注意两个约束条件分别代表一个曲面，同时成立代表两个曲面相交的空间曲线．我们要在该曲线上求函数 f 的极值．同样，若曲线上的某点是极值，那必须满足曲线在该点的切线方向 f 的方向导数为零，即 $\boldsymbol{\nabla} f$ 与该点法向量共线．然而这是个三维问题，空间曲线的任意一点可以有两条不共线的法向量（不一定互相垂直），该点的任何其他法向量可以由这两个法线的线性组合来表示．即存在常数 λ_1 和 λ_2，使得

$$\boldsymbol{\nabla} f+\lambda_1\boldsymbol{\nabla} g_1+\lambda_2\boldsymbol{\nabla} g_2=0 \tag{9}$$

即

$$\boldsymbol{\nabla}(f+\lambda_1 g_1+\lambda_2 g_2)=0 \tag{10}$$

其中 $f+\lambda_1 g_1+\lambda_2 g_2$ 就是拉格朗日函数．当然，限制条件的数目不一定是 2，若只有一个限制条件 $g_1=0$，就是求曲面 $g_1=0$ 上函数 f 取得极值的点，条件是函数的梯度 $\boldsymbol{\nabla} f$ 与曲面的法向量 $\boldsymbol{\nabla} g_1$ 共线．

例 2

椭圆方程为 $x^2/a^2+y^2/b^2=1$，求其内接长方型的最大面积（长方形的边平行于 x 轴和 y 轴）．

令长方型的右上角与椭圆的切点为 (x,y)，则长方型的面积为 $S=4xy$．约束函数为 $g=x^2/a^2+y^2/b^2-1$．拉格朗日函数为

$$\Lambda(x,y)=S+\lambda g=4xy+\lambda\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}-1\right) \tag{11}$$

令 x,y 偏导都为 0，得

$$\begin{cases}4y+2\lambda x/a^2=0\\ 4x+2\lambda y/b^2=0\end{cases} \tag{12}$$

两式消去 λ，得 $x/y=a/b$．再联立椭圆方程，解得 $(x,y)=(a/\sqrt{2},b/\sqrt{2})$，面积最大值为 $2ab$．

一阶线性常微分方程组

预备知识 一阶线性微分方程[212]

解析解

形式为

未完成：一般一阶线性微分方程组如何化成该形式？

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}=\mathbf{A}\mathbf{v} \tag{1}$$

的一阶偏微分方程组的解析解（先假设 $\mathbf{A}$ 为常矩阵）为

$$\mathbf{v}(t)=\exp(\mathbf{A}t)\mathbf{v}(0) \tag{2}$$

其中矩阵的指数函数由矩阵的泰勒级数[181] 定义

$$\exp(\mathbf{A})=1+\mathbf{A}+\frac{1}{2!}\mathbf{A}^2+\dots \tag{3}$$

将式 2 写成矩阵级数的形式即可验证式 1.

形式解

当式 1 中的 $\mathbf{A}$ 是 t 的函数 $\mathbf{A}(t)$ 时，我们可以取微小时间步长 Δt，在每个 Δt 内近似认为 $\mathbf{A}(t_i)$ 为常数，再取极限

$$\mathbf{v}(t)=\lim_{\Delta t\to 0}\prod\exp\left[\mathbf{A}(t_i)\Delta t\right]\mathbf{v}(0) \tag{4}$$

如果两个矩阵 $\mathbf{P},\mathbf{Q}$ 对易，就有

$$\exp(\mathbf{P})\exp(\mathbf{Q})=\exp(\mathbf{P}+\mathbf{Q}) \tag{5}$$

但一般来说 $\mathbf{A}(t_i)$ 之间不对易，所以我们定义一个时间排序算符 $\hat{\mathcal{T}}$ 使例如

$$\hat{\mathcal{T}}[\mathbf{A}(t_1)\mathbf{A}(t_3)\mathbf{A}(k_2)]=\mathbf{A}(t_3)\mathbf{A}(t_2)\mathbf{A}(t_1)\qquad(t_1<t_2<t_3) \tag{6}$$

这样通解在形式上就可以记为

$$\mathbf{v}(t)=\hat{\mathcal{T}}\exp\left(\int_0^t\mathbf{A}(t')\mathrm{d}t'\right)\mathbf{v}(0) \tag{7}$$

然而这么做对于数值计算并没有太大意义.

厄米矩阵

如果 $\mathbf{H}$ 是厄米矩阵，那么 $\mathrm{e}^{\mathrm{i}\mathbf{H}t}$ （$t \in \mathbb{R}$）是一个酉矩阵（幺正矩阵）[440]. 所以 $\mathbf{v}(t)$ 的模 $|\mathbf{v}(t)| = \sum |v_i|^2$ 将不随时间变化.

证明：把式 3 两边取厄米共轭得

$$(\mathrm{e}^{\mathrm{i}\mathbf{H}t})^\dagger = \mathrm{e}^{-\mathrm{i}\mathbf{H}^\dagger t} = \mathrm{e}^{-\mathrm{i}\mathbf{H}t} \tag{8}$$

由于 $[H, H] = 0$，有 $(\mathrm{e}^{\mathrm{i}\mathbf{H}t})^\dagger \mathrm{e}^{\mathrm{i}\mathbf{H}t} = \mathrm{e}^{-\mathrm{i}\mathbf{H}t}\mathrm{e}^{\mathrm{i}\mathbf{H}t} = \mathbf{I}$. 证毕.

数值计算

如果矩阵 $\mathbf{M}$ 是厄米矩阵，则可以先做对角化 $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$，其中 $\mathbf{\Lambda}$ 是对角矩阵，$\mathbf{U}$ 是酉矩阵. 这样就有

$$\exp(\mathbf{M}) = \mathbf{U}\mathbf{U}^\dagger + \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger + \mathbf{U}\frac{1}{2!}\mathbf{\Lambda}^2\mathbf{U}^\dagger + \cdots = \mathbf{U}\exp(\mathbf{\Lambda})\mathbf{U}^\dagger \tag{9}$$

由于对角矩阵相乘等于每个对角元分别相乘，把 $\mathbf{\Lambda}$ 的每个矩阵元求指数函数就可以得到 $\exp(\mathbf{\Lambda})$. 这样做可以减少计算量.

事实上，以上做法相当于分离变量，当 $\mathbf{A}$ 是厄米矩阵时，令 $\mathbf{v}(t) = f(t)\mathbf{u}$，代入方程得

$$\frac{f'(t)}{f(t)}\mathbf{u} = \mathbf{A}\mathbf{u} \tag{10}$$

由于 $\mathbf{A}$ 和 $\mathbf{u}$ 都不含时，所以可以令

$$\begin{aligned} \mathbf{A}\mathbf{u} &= \lambda\mathbf{u} \\ f(t)' &= \lambda f(t) \end{aligned} \tag{11}$$

其中第一个方程是 $\mathbf{A}$ 的本征方程，解为 N 个本征矢 $\mathbf{u}_i$（即 $\mathbf{U}$ 的第 i 列）和 N 个本征值 λ_i（即 $\mathbf{\Lambda}$ 的第 i 个对角元）. 第二条方程的解为 $f(t) = \exp(\lambda t)$. 所以方程的通解为

$$\mathbf{v}(t) = \sum_{i=1}^{N} c_i f_i(t)\mathbf{u}_i = \sum_{i=1}^{N} \mathbf{u}_i^\dagger \mathbf{v}(0)\exp(\lambda_i t)\mathbf{u}_i = \mathbf{U}\exp(\Lambda t)\mathbf{U}^\dagger\mathbf{v}(0) \tag{12}$$

在此基础上使用 Lanczos 算法[1931] 可以进一步提高效率.

算符拆分

有时候我们希望可以在上述计算中把 $\mathbf{A}$ 写成几个矩阵的和的形式（以两个为例）$\mathbf{A}=\mathbf{B}+\mathbf{C}$. 当 $\mathbf{B}$ 和 $\mathbf{C}$ 对易时显然有

$$\exp(\mathbf{A}t)=\exp(\mathbf{B}t)\exp(\mathbf{C}t) \tag{13}$$

在程序中这么做可能可以进一步提高速度[25]. 如果 $\mathbf{B}$ 和 $\mathbf{C}$ 不对易，严格来说上式不成立，但可以证明 $t\to 0$ 时近似成立

$$\begin{aligned}
&\exp(\mathbf{B}t)\exp(\mathbf{C}t)\\
&=\left(1+\mathbf{B}t+\frac{1}{2!}\mathbf{B}^2t^2+\dots\right)\left(1+\mathbf{C}t+\frac{1}{2!}\mathbf{C}^2t^2+\dots\right)\\
&=1+(\mathbf{B}+\mathbf{C})t+\frac{1}{2!}\big(\mathbf{B}^2+\mathbf{C}^2+2\mathbf{BC}\big)t^2+\dots\\
&=\exp(\mathbf{A}t)+\mathcal{O}\big(t^2\big)
\end{aligned} \tag{14}$$

这里的 $\mathcal{O}(t^2)$ 是由于第二个等号后面是 $2\mathbf{BC}$ 而不是 $\mathbf{BC}+\mathbf{CB}$.

齐次函数的欧拉定理

预备知识 矢量空间[444]，复合函数求导链式法则[174]

首先介绍一下什么是齐次函数.

定义 1　齐次函数

假设 $f:V\to W$ 是域 F 内的两个向量空间之间的函数.

我们说 f 是 k 次齐次函数，如果对于所有非零的 $\alpha\in F$ 和 $\mathrm{v}\in V$，都有：

$$f(\alpha\mathrm{v})=\alpha^k f(\mathrm{v}) \tag{1}$$

即是，在欧几里得空间，$f(\alpha\mathrm{v})=f(k)\,f(\mathrm{v})$，其中 $f(k)$ 为指数函数.

例 1

$f(x,y,z)=x^5y^2z^3$ 是 10 次齐次函数，因为 $(\alpha x)^5(\alpha y)^2(\alpha z)^3=\alpha^{10}x^5y^2z^3$.

$f(x,y)=x^5+2x^3y^2+9xy^4$ 是 5 次齐次函数.

[25] 例如二维波函数的动能算符 $T=T_x+T_y$

齐次函数的欧拉定理表述如下：

定理 1　齐次函数的欧拉定理

若 k 齐次函数 $f:\mathbb{R}^n\to\mathbb{R}$ 是可导的，那么

$$\mathrm{x}\cdot\nabla f(\mathrm{x})=kf(\mathrm{x}) \tag{2}$$

证明：记 $f=f(x_1,\ldots,x_n)=f(\mathrm{x})$，把以下等式两端对 α 求导：

$$f(\alpha\mathrm{x})=\alpha^k f(\mathrm{x}) \tag{3}$$

利用复合函数求导法则，可得：

$$\frac{\partial}{\partial x_1}f(\alpha\mathrm{x})\frac{\mathrm{d}}{\mathrm{d}\alpha}(\alpha x_1)+\cdots+\frac{\partial}{\partial x_n}f(\alpha\mathrm{x})\frac{\mathrm{d}}{\mathrm{d}\alpha}(\alpha x_n)=k\alpha^{k-1}f(\mathrm{x}) \tag{4}$$

因此：

$$x_1\frac{\partial}{\partial x_1}f(\alpha\mathrm{x})+\cdots+x_n\frac{\partial}{\partial x_n}f(\alpha\mathrm{x})=k\alpha^{k-1}f(\mathrm{x}) \tag{5}$$

即

$$\mathrm{x}\cdot\nabla f(\alpha\mathrm{x})=k\alpha^{k-1}f(\mathrm{x}) \tag{6}$$

令 $\alpha=1$ 得证．证毕．

类似上面的推导过程，我们还可以得到如下推论：

推论 1

若 $f:\mathbb{R}^n\to\mathbb{R}$ 是可导的，且是 k 阶齐次函数．则它的一阶偏导数 $\partial f/\partial x_i$ 是 $k-1$ 阶齐次函数．

证明：记 $f=f(x_1,\ldots,x_n)=f(\mathrm{x})$，并把以下等式两端对 x_i 求导：

$$f(\alpha\mathrm{x})=\alpha^k f(\mathrm{x}) \tag{7}$$

利用复合函数求导法则，可得：

$$\frac{\partial}{\partial x_i}f(\alpha\mathrm{x})\frac{\mathrm{d}}{\mathrm{d}x_i}(\alpha x_i)=\alpha^k\frac{\partial}{\partial x_i}f(\mathrm{x})\frac{\mathrm{d}}{\mathrm{d}x_i}(x_i) \tag{8}$$

因此：

$$\alpha\frac{\partial}{\partial x_i}f(\alpha\mathrm{x})=\alpha^k\frac{\partial}{\partial x_i}f(\mathrm{x}) \tag{9}$$

所以

$$\frac{\partial}{\partial x_i}f(\alpha\mathrm{x})=\alpha^{k-1}\frac{\partial}{\partial x_i}f(\mathrm{x}) \tag{10}$$

证毕．

多元泰勒展开

多元泰勒展开公式如下

$$f(\mathbf{x})=f(\mathbf{x}_0)+[(\mathbf{r}-\mathbf{r}_0)\boldsymbol{\nabla}]f(\mathbf{x}_0)+\frac{1}{2!}[(\mathbf{r}-\mathbf{r}_0)\boldsymbol{\nabla}]^2f(\mathbf{x}_0)+\dots \\ +\frac{1}{n!}[(\mathbf{r}-\mathbf{r}_0)\boldsymbol{\nabla}]^nf(\mathbf{x}_0)+\mathcal{O}\left(x^{N+1}\right) \tag{1}$$

雅可比行列式

预备知识 行列式[432]，全微分[269]

若有坐标系变换

$$\begin{cases}x=x(u,v,w)\\y=y(u,v,w)\\z=z(u,v,w)\end{cases} \tag{1}$$

根据全微分关系

$$\begin{pmatrix}\mathrm{d}x\\\mathrm{d}y\\\mathrm{d}z\end{pmatrix}=\begin{pmatrix}\partial x/\partial u&\partial x/\partial v&\partial x/\partial w\\\partial y/\partial u&\partial y/\partial v&\partial y/\partial w\\\partial z/\partial u&\partial z/\partial v&\partial z/\partial w\end{pmatrix}\begin{pmatrix}\mathrm{d}u\\\mathrm{d}v\\\mathrm{d}w\end{pmatrix} \tag{2}$$

其中 $\mathbf{J}$ 叫做雅可比矩阵.

考虑 uvw 坐标系中的一个体积元 (u,v,w)-$(u+\mathrm{d}u\,,v+\mathrm{d}v\,,w+\mathrm{d}w)$，一般情况下（不需要是正交曲线坐标系），体积元为平行六面体，起点为 (u,v,w) 的三条棱对应的矢量分别为

$$\begin{pmatrix}\mathrm{d}x_1\\\mathrm{d}y_1\\\mathrm{d}z_1\end{pmatrix}=\mathbf{J}\begin{pmatrix}\mathrm{d}u\\0\\0\end{pmatrix}=\begin{pmatrix}J_{11}\\J_{21}\\J_{31}\end{pmatrix}\mathrm{d}u \tag{3}$$

$$\begin{pmatrix}\mathrm{d}x_2\\\mathrm{d}y_2\\\mathrm{d}z_2\end{pmatrix}=\mathbf{J}\begin{pmatrix}0\\\mathrm{d}v\\0\end{pmatrix}=\begin{pmatrix}J_{12}\\J_{22}\\J_{32}\end{pmatrix}\mathrm{d}v \tag{4}$$

$$\begin{pmatrix} \mathrm{d}x_3 \\ \mathrm{d}y_3 \\ \mathrm{d}z_3 \end{pmatrix} = \mathbf{J} \begin{pmatrix} 0 \\ 0 \\ \mathrm{d}w \end{pmatrix} = \begin{pmatrix} J_{13} \\ J_{23} \\ J_{33} \end{pmatrix} \mathrm{d}w \tag{5}$$

由于平行六面体的体积是同一起点三条矢量的混合积

$$\mathrm{d}V = \begin{vmatrix} \mathrm{d}x_1 & \mathrm{d}y_1 & \mathrm{d}z_1 \\ \mathrm{d}x_2 & \mathrm{d}y_2 & \mathrm{d}z_2 \\ \mathrm{d}x_3 & \mathrm{d}y_3 & \mathrm{d}z_3 \end{vmatrix} = \begin{vmatrix} \mathrm{d}x_1 & \mathrm{d}x_2 & \mathrm{d}x_3 \\ \mathrm{d}y_1 & \mathrm{d}y_2 & \mathrm{d}y_3 \\ \mathrm{d}z_1 & \mathrm{d}z_2 & \mathrm{d}z_3 \end{vmatrix} = |\mathbf{J}| \,\mathrm{d}u \,\mathrm{d}v \,\mathrm{d}w \tag{6}$$

其中 $|\mathbf{J}|$ 叫做**雅可比行列式**.

高斯积分

预备知识 极坐标中的二重积分[291]

以下定积分被定义为**高斯积分**

$$I = \int_{-\infty}^{+\infty} \mathrm{e}^{-x^2} \,\mathrm{d}x \tag{1}$$

求解高斯积分最简单的方法是在极坐标中求解以下面积分

$$I^2 = \int_{-\infty}^{+\infty} \mathrm{e}^{-x^2} \,\mathrm{d}x \int_{-\infty}^{+\infty} \mathrm{e}^{-y^2} \,\mathrm{d}y = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathrm{e}^{-(x^2+y^2)} \,\mathrm{d}x \,\mathrm{d}y \tag{2}$$

在极坐标系中，$r^2 = x^2 + y^2$，上式变为

$$I^2 = \int_0^{2\pi} \int_0^{+\infty} \mathrm{e}^{-r^2} r \,\mathrm{d}r \,\mathrm{d}\theta = 2\pi \int_0^{+\infty} r\mathrm{e}^{-r^2} \,\mathrm{d}r \tag{3}$$

用换元积分法，令 $t = r^2$，$\mathrm{d}t = 2r \,\mathrm{d}r$，得

$$I^2 = \pi \int_0^{+\infty} \mathrm{e}^{-t} \,\mathrm{d}t = \pi \left(-\mathrm{e}^{-t}\right)\Big|_0^{+\infty} = \pi \tag{4}$$

最后开方即可得到高斯积分为

$$I = \int_{-\infty}^{+\infty} \mathrm{e}^{-x^2} \,\mathrm{d}x = \sqrt{\pi} \tag{5}$$

更一般地，由换元积分法式 5 [188] 可得

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-ax^2} \,\mathrm{d}x = \sqrt{\frac{\pi}{a}} \tag{6}$$

多维球体的体积

预备知识 gamma 函数[254]

半径为 R 的 n 维欧几里得空间中的球体的体积可以用 Γ 函数表示为

$$V_n = \frac{\pi^{n/2}}{\Gamma(n/2+1)} R^n = \frac{\pi^{n/2}}{(n/2)!} R^n \tag{1}$$

其中阶乘和 Gamma 函数的关系见式 1 [254]．若定义 n 维球体的表面满足方程 $\sum_{i=1}^{n} x_i^2 = R_n^2$，其中 x_i 为 n 维直角坐标系中第 i 个坐标．所有满足 $\sum_{i=1}^{n} x_i^2 \leqslant R_n^2$ 的坐标点都定义为球内的点，且定义 n 维直角坐标系中的体积为 $V_n = \int \mathrm{d}x_1\, \mathrm{d}x_2 \ldots \mathrm{d}x_n$，积分是对所有球内的点积分．

如果这些定义看起来很抽象，不妨代入到三维空间中考虑．三维直角坐标系中，x_1, x_2, x_3 分别是 x, y, z，R_3 是球的半径，球表面上任意一点都满足 $x^2 + y^2 + z^2 = R_3^2$，且球的体积分为 $\int \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z$ 是对球内部的所有点积分．另外，若把上述定义代入到 1 维和 2 维，不难发现所谓的“1 维球”和“2 维球”分别是半径为 R_1 的线段和半径为 R_2 的圆．

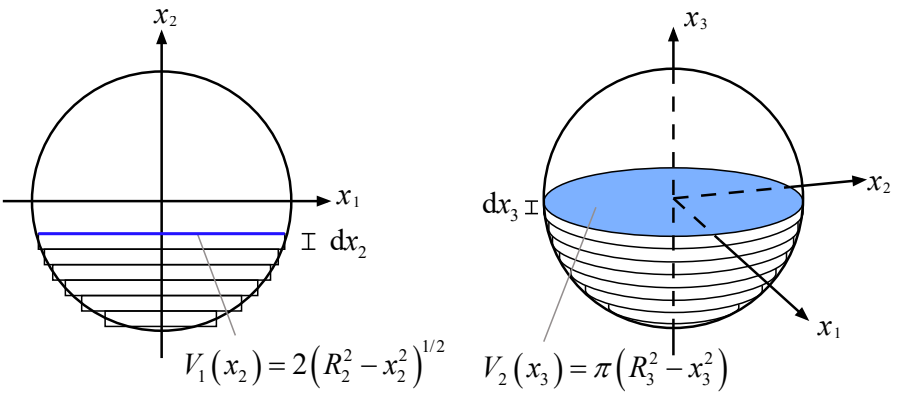
推导

由于正常人的空间想象力最高是 3 维，我们先由 3 维以内的球体总结出体积的递推公式，这样即使我们无法想象高维球的形状，也可以计算其体积．下面在推导前 3 个维度时，请把所有 x_1, x_2, x_3 想象成 xyz．

1 维球

这是一条线段，满足 $x_1^2 \leqslant R_1^2$，“体积”就是线段长度

$$V_1 = \int \mathrm{d}x_1 = 2R_1 \tag{2}$$



二维球是一个圆　　三维球是一个球体

图 1: 二维和三维球的体积

2 维球

这是一个圆，满足 $x_1^2+x_2^2\leqslant R_2^2$，在计算体积 $V_2=\int \mathrm{d}x_1\,\mathrm{d}x_2$ 时，可以先对 x_1 积分再对 x_2 积分

$$V_2=\int\left(\int \mathrm{d}x_1\right)\mathrm{d}x_2=\int V_1(x_2)\,\mathrm{d}x_2 \tag{3}$$

在几何上，这就是说把圆从沿 x_1 轴切成许多一维球（线段），由 $x_1^2\leqslant R_2^2-x_2^2$，一维球的半径为 $R_1(x_2)=(R_2^2-x_2^2)^{1/2}$．代入式 2，得 x_2 处切出的一维球的体积（线段的长度）为

$$V_1(x_2)=\int \mathrm{d}x_1=2R_1=2(R_2^2-x_2^2)^{1/2} \tag{4}$$

再代入式 3，得二维球的体积为（注意 $-R_2<x_2<R_2$ ）

$$V_2=\int V_1\,\mathrm{d}x_2=\int 2(R_2^2-x_2^2)^{1/2}\,\mathrm{d}x_2=\pi R_2^2 \tag{5}$$

3 维球

这是一个球体，满足 $x_1^2+x_2^2+x_3^2\leqslant R_3^2$，计算体积 $V_3=\int \mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3$ 时，可以先对 x_1x_2 积分

$$V_3=\int\left(\int \mathrm{d}x_1\,\mathrm{d}x_2\right)\mathrm{d}x_3=\int V_2(x_3)\,\mathrm{d}x_3 \tag{6}$$

在几何意义上，这是说把球沿 x_1x_2 平面切成许多二维球（圆），然后把球的体积（面积）沿 x_3 轴积分．由 $x_1^2+x_2^2 \leqslant R_3^2-x_3^2$，得 x_3 处二维球半径为 $R_2=(R_3^2-x_3^2)^{1/2}$．由式 5 得体积为

$$V_2(x_3)=\pi R_2^2=\pi(R_3^2-x_3^2) \tag{7}$$

代入式 6 得三维球体积（注意 $-R_2<x_2<R_2$）

$$V_3=\int V_2(x_3)\,\mathrm{d}x_3=\int \pi(R_3^2-x_3^2)\,\mathrm{d}x_3=\frac{4}{3}\pi R^3 \tag{8}$$

n 维球

由以上两个推导，可以在代数上总结出递推的规律．把 n 维球在 $n+1$ 维积分，得[26]

$$V_4=\int_{-R}^{R}\frac{4}{3}\pi(R_4^2-x_4^2)^{3/2}\,\mathrm{d}x_4=\frac{1}{2}\pi^2R^4 \tag{9}$$

令 n 维球体体积维 C_nR^n，那么可以总结出递推公式为

$$V_{n+1}=\int_{-R}^{R}C_n(R^2-x^2)^{n/2}\,\mathrm{d}x \tag{10}$$

使用换元积分法[188]，令 $x=R\sin\theta$，有

$$\begin{aligned}\int_{-R}^{R}(R^2-x^2)^{n/2}\,\mathrm{d}x&=R^{n+1}\int_{-\pi/2}^{\pi/2}\cos^{n+1}\theta\,\mathrm{d}\theta\\&=\sqrt{\pi}\frac{\Gamma(n/2+1)}{\Gamma[(n+1)/2+1]}R^{n+1}\end{aligned} \tag{11}$$

用阶乘表示递归公式（式 1 [254]），就是

$$C_{n+1}=C_n\frac{\sqrt{\pi}(n/2)!}{[(n+1)/2]!} \tag{12}$$

已知 $C_1=2$，易得

$$C_n=\frac{\pi^{n/2}}{(n/2)!} \tag{13}$$

这样就得到了式 1．

[26]可使用 Mathematica 软件计算积分，见 Mathematica 积分

多元函数积分和宇称

预备知识 定积分[198]

如果一个 N 维欧几里得空间中的函数 $f(\mathbf{r}) = f(x_1, x_2, \ldots, x_N)$ 满足 $f(-\mathbf{r}) = f(\mathbf{r})$，我们就说它具有**奇宇称（odd parity）**，如果满足 $f(-\mathbf{r}) = -f(\mathbf{r})$ 我们就说它有**偶宇称（even parity）**.

我们以下要说明的结论是：在中心对称的定义域（即如果 $\mathbf{r}$ 在定义域中，$-\mathbf{r}$ 也在定义域中），具有奇宇称的函数的定积分为 0，具有偶宇称的函数的定积分可能不为 0.

对于 $N = 1, 2, 3$，这是容易理解的，例如一元函数 $\sin x$ 具有奇宇称，所以在任意对称的区间 $[-a, a]$ 做定积分都为 0. 又例如二元函数 $x^3 + y^3$ 和三元函数 $\sin(x + y + z)$ 也具有奇宇称. 证明的思想很简单，做定积分时每个 $\mathbf{r}$ 处的“微元” $\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ 都会在 $-\mathbf{r}$ 处有一个函数值为相反数的“微元”，使两个微元的积分互相抵消.

例 1　极坐标中的函数的宇称

极坐标中的函数

$$f(\mathbf{r}) = f(r, \theta) = r\sin(\theta) \tag{1}$$

具有奇宇称（将 $\mathbf{r}$ 变为 $-\mathbf{r}$，只需要将 θ 加上 π 即可，而 $\sin(\theta + \pi) = -\sin\theta$）. 如果在一个圆环形区域上做定积分，就有

$$\int_0^{2\pi}\int_b^a f(r, \theta)\,\mathrm{d}r\,\mathrm{d}\theta = 0 \tag{2}$$

另一种情况如三维空间中的二维曲面上的积分，如球谐函数满足（见式 11[761]）

$$Y_{l,m}(-\hat{\mathbf{r}}) = (-1)^l Y_{l,m}(\hat{\mathbf{r}}) \tag{3}$$

也就是说当 l 为奇数时球谐函数具有奇宇称，偶数时具有偶宇称. 所以在对整个球面积分时，前者必为零.

宇称函数相乘

如果两个函数 $f(\mathbf{r})$ 和 $g(\mathbf{r})$ 分别可能具有奇宇称或偶宇称，它们相乘所得的函数 $h(\mathbf{r}) = f(\mathbf{r})g(\mathbf{r})$ 的宇称如何呢？从定义不难证明，如果二者宇称相同，那么 $h(\mathbf{r})$ 具有偶宇称，如果一奇一偶，则是奇宇称.

习题 1

求球坐标系中函数 $f(\mathbf{r}) = r^{100}\sin(99\theta)Y_{l,m}(\hat{\mathbf{r}})$ 的宇称.

习题 2

求单位球面上的积分（星号表示复共轭）

$$\int Y^*_{99,77}(\hat{\mathbf{r}})Y_{71,-30}(\hat{\mathbf{r}})Y_{53,20}(\hat{\mathbf{r}})\,\mathrm{d}\Omega \tag{4}$$

第四章

复变函数

复变函数

- 本词条存在未完成的内容.

预备知识 函数[103]，复数[142]

复变函数（complex function）一般是指自变量和函数值都在复数域内取值的函数，即 $f:\mathbb{C}\to\mathbb{C}$，通常表示为

$$w=f(z) \tag{1}$$

更一般地，函数值为复数的函数都可以叫做复变函数．例如 $f:\mathbb{R}\to\mathbb{C}$（自变量是一个实数）或者 $f:\mathbb{R}^N\to\mathbb{C}$（自变量是 N 个有序实数）.

若把自变量 z 拆分成实部 x 和虚部 y，即 $z=x+\mathrm{i}y$，且函数 f 也拆分成实部函数 $u(x,y)$ 和虚部函数 $\mathrm{i}v(x,y)$，则复变函数可记为

$$w=u(x,y)+\mathrm{i}v(x,y) \tag{2}$$

其中 $u,v:\mathbb{R}^2\to\mathbb{R}$ 都是二元实函数．例如，复指数函数[351] 被定义为

$$w=\mathrm{e}^z=\mathrm{e}^x\cos y+\mathrm{i}\mathrm{e}^x\sin y\qquad(z\in\mathbb{C}) \tag{3}$$

即实部和虚部函数分别为 $u(x,y)=\mathrm{e}^x\cos y$，$v(x,y)=\mathrm{e}^x\sin y$.

例如物理中质点的简谐振动的复数形式[995] 为

$$f(t)=\mathrm{e}^{-\mathrm{i}\omega t}=\cos(\omega t)+\mathrm{i}\sin(\omega t)\qquad(t\in\mathbb{R}) \tag{4}$$

它的实部表示振动的位移，而虚部没有物理意义.

复变函数与矢量场

当我们把复变函数记为式 2 的形式后，可以把它看作一个平面矢量场[294]，即复平面上的每一点对应一个几何矢量[370]．令实轴和虚轴分别为 x 轴和 y 轴，其单位矢量分别为 $\hat{\mathbf{x}}$ 和 $\hat{\mathbf{y}}$，令位置矢量[878] $\mathbf{r}=x\hat{\mathbf{x}}+y\hat{\mathbf{y}}$，那么这个矢量场就可以表示为

$$\mathbf{w}(\mathbf{r})=u(\mathbf{r})\hat{\mathbf{x}}+v(\mathbf{r})\hat{\mathbf{y}} \tag{5}$$

这相当于把映射 $f:\mathbb{C}\to\mathbb{C}$ 看成 $f:\mathbb{R}^2\to\mathbb{R}^2$．其中 u,v 分别是二维平面上的标量场．

正因如此，研究复变函数时可以借助许多矢量分析的工具，如散度[311] 和旋度[315] 等．所以以后讲解时我们也会使用相应的矢量分析内容作为预备知识．

复变函数的可视化

使用矢量场的概念可以用于把复变函数可视化，这里以复指数函数（式 3 ）为例．第一种方法是用箭头表示平面矢量场，如图 1．也可以把复变函数的

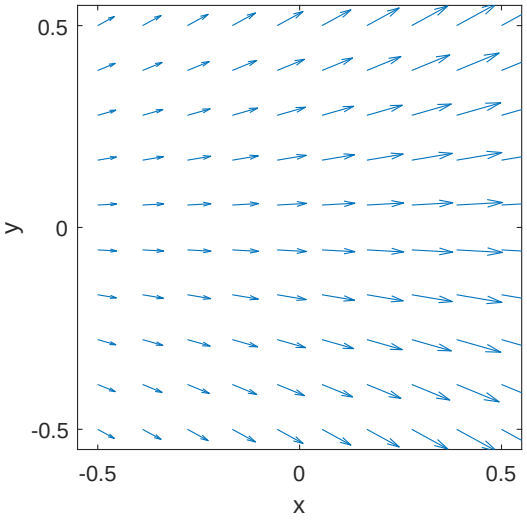


图 1: e^z 对应的二维矢量场（式 5 ），图中的箭头与 $\mathbf{w}(\mathbf{r})$ 成正比．

实部 $u(x,y)$ 和虚部 $v(x,y)$ 分别画成二维标量场（图 2 ）．式 2 也可以画成 3D 的曲面图

还有一种方式是使用变形网格的方式表示平面到平面的映射（图 4 ），由该图可以看出 x 决定 e^z 的模长，y 决定 e^z 的幅角．

作为实函数的拓展

复变函数中很多函数与我们原来我们学过的函数同名，只是定义域和值域从实数拓展到了复数．例如三角函数，对数函数，指数函数等．拓展后，复变函数的定义必须要与原来的实函数保持“兼容”，即当自变量取实数时，函数值也要与原来相同．

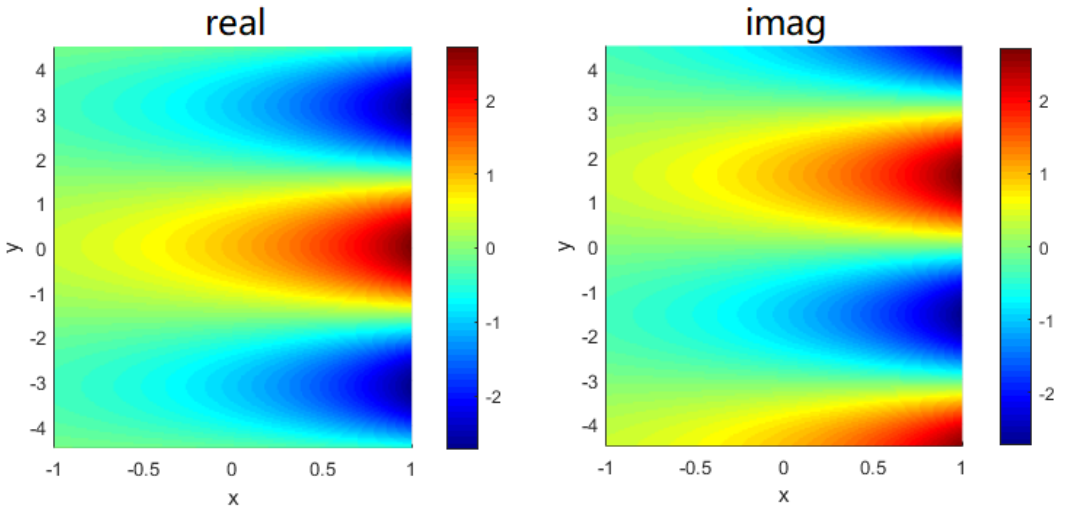


图 2: $\exp(z)$ 的实部和虚部函数

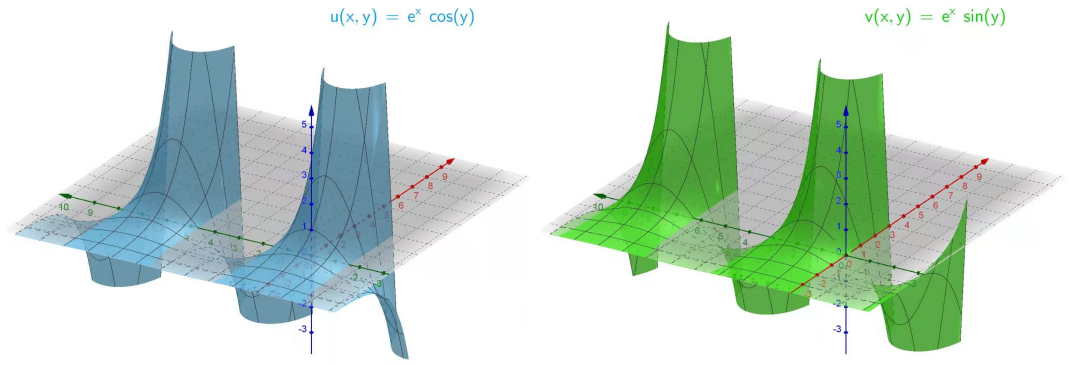


图 3: $\exp(z)$ 的实部和虚部函数

例如，当复指数函数式 3 的自变量只在实轴上取值（即 $y = 0$）时，该函数变为我们原来所熟悉的实函数 e^x．又如，复正弦函数[352] 的定义为

$$\sin z = \sin(x + \mathrm{i}y) = \sin x \cosh y + \mathrm{i} \cos x \sinh y \tag{6}$$

其中 sinh 和 cosh 是双曲正弦和双曲余弦函数双曲函数[114]．当 $y = 0$ 时（$\cosh 0 = 1$，$\sinh 0 = 0$），又重新得到实函数 $\sin x$．

如果只需满足“兼容”的条件，那么拓展的方式可以有无限多种，所以为什么一定要拓展成式 3 或式 6 呢？这是因为我们通常还需要复变函数满足另一个条件，就是拓展后的函数需要是**解析**的，这叫做**解析拓延**．解析拓延是唯一的，例如实函数 $\sin x$ 进行解析拓延后，只有式 6 一种可能的定义．

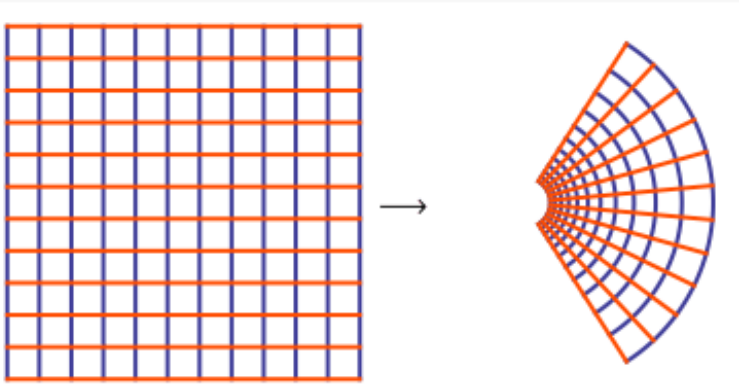


图 4: $\exp(z)$ 函数把复平面上的方形网格映射成另一种形状的网格（来自 Wolfram Alpha）

我们研究的复变函时几乎都是在讨论解析函数（至少在复平面的某个区域内解析），简单来说解析函数要求它可以在复平面的某个区域内关于某点 z_0 展开为（复）无穷级数（类比实函数的泰勒级数[181]）

$$f(z)=\sum_{n=0}^{\infty}c_n(z-z_0)^n \tag{7}$$

我们说，如果复变函数在某个区域解析，那么它在该区域就是**光滑（smooth）**的，即无穷阶连续可导（类比实函数的高阶导数[169]）.

复变函数的导数

复变函数的导数定义与实函数的导数[158] 十分相似：

$$f'(z)=\lim_{\Delta z\to 0}\frac{\Delta f}{\Delta z} \tag{8}$$

只是这里的除法变为复数的除法（式 17 [142]）. 讨论一般复变函数的导数较为复杂，因为它不但取决于 z 的值，还取决于求导的方向（即 Δz 趋近于 0 的方向），这有点类似于方向导数[279]. 但我们通常研究的解析函数有一个非常好的性质：如果复变函数在某个区域上解析，那么在该区域上其导数与求导方向无关. 详见“柯西—黎曼条件[354]”.

对于复数域的基本初等函数（几乎都是解析函数），它们的导函数也大多和实数域基本初等函数的导函数具有一样的表达式[165]. 例如 $(\mathrm{e}^z)'=\mathrm{e}^z$，$\sin' z=\cos z$，等. 四则运算的求导法则也与实函数的情况[160] 相同.

复积分

复变函数的积分需要在复平面上选取一个路径．从矢量场的角度考虑，这可以类比矢量场的线积分[295]．

未完成：未完成，旋度为零，积分结果与路径无关

幂函数（复数）

我们先来看实参数的幂函数 $f(x)=x^a$，$x>0$ 时函数曲线如图 1 所示．注意 $x^{1/a}$ 是 x^a 的反函数．

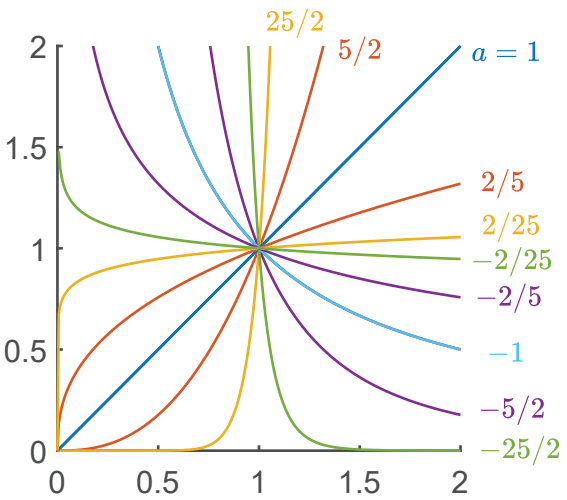


图 1: 实参数的幂函数（相同颜色的函数互为反函数）

当 $x<0$ 时，可得

$$x^a=|x|^a(-1)^a=|x|^a\mathrm{e}^{\mathrm{i}\pi a} \tag{1}$$

当 a 为偶整数时，$x^a=(-x)^a$ 是偶函数，a 为奇整数时，$x^a=-(-x)^a$ 是奇函数．当 a 为非整数时，x^a 必为复数，其模长仍为 $|x|^a$，幅角为常数 $\mathrm{e}^{\mathrm{i}\pi a}$．

复参数的幂函数

我们再来将复数的幂函数分解为模长和相位的形式（令 $z=|z|\mathrm{e}^{\mathrm{i}\phi(z)}$，$a=a_I+\mathrm{i}a_R$ ）

$$z^a=|z|^{a_R}\mathrm{e}^{-\phi(z)a_I}\mathrm{e}^{\mathrm{i}[\ln|z|a_I+\phi(z)a_R]} \tag{2}$$

可见 z^a 的模长和幅角都分别与 z 和 a 有关．一般情况下，这是一个比较复杂的函数，含有不同的分支（因为 $\phi(z)$ 可以加整数个 2π）．当且仅当 a 为整数时才不会出现分支．在数值计算中，分支切割线出现在 $\phi(z)=\pm\pi$ 处，这是因为数值计算通常取 $\phi(z)\in(-\pi,\pi]$．

指数函数（复数）

预备知识 指数函数，复数[142]，基本初等函数的导数[165]

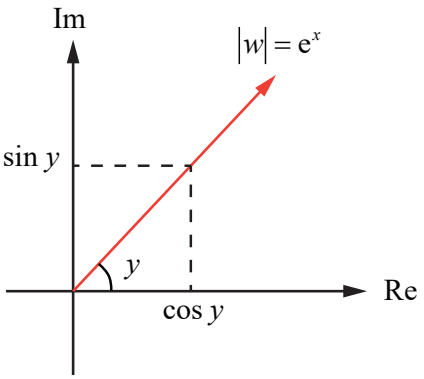


图 1: 复数域中的指数函数

[1]复数域中的指数函数被定义为

$$w=\mathrm{e}^z=\mathrm{e}^{x+\mathrm{i}y}=\mathrm{e}^x(\cos y+\mathrm{i}\sin y) \tag{1}$$

在复平面上表示这个函数，则指数的实部 x 控制函数值 w 的模长，虚部 y 控制 w 的幅角，如图 1

$$|w|=\mathrm{e}^x \qquad \arg(w)=y \tag{2}$$

[1]参考 Wikipedia 相关页面.

当指数为纯虚数时，式 1 变为著名的**欧拉公式（Euler's formula）**

$$\mathrm{e}^{\mathrm{i}x} = \cos x + \mathrm{i}\sin x \tag{3}$$

虽然这里的 x 一般是实数（物理中应用得最多的情况），但根据复数域三角函数的定义[352]，对于任何复数 z，都有欧拉公式

$$\mathrm{e}^{\mathrm{i}z} = \cos z + \mathrm{i}\sin z \tag{4}$$

将“三角函数（复数）[352]”中的式 1 和式 2 代入即可证明.

根据式 1 的定义结合两角和公式（式 3 [112]），容易证明 e^z 同样满足

$$\mathrm{e}^{z_1+z_2} = \mathrm{e}^{z_1}\mathrm{e}^{z_2} \tag{5}$$

虽然我们还没有系统地学习复变函数求导的概念，但我们可以根据式 3 求出一个物理中常见的导数公式

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{\mathrm{i}ax} &= -a\sin(ax) + \mathrm{i}a\cos(ax)\\ &= \mathrm{i}a[\cos(ax) + \mathrm{i}\sin(ax)]\\ &= \mathrm{i}a\mathrm{e}^{\mathrm{i}ax}\end{aligned} \tag{6}$$

进一步拓展，令复常数 $z = a + \mathrm{i}b$ 得

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{zx} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\mathrm{e}^{ax}\mathrm{e}^{\mathrm{i}bx}\right) = (a+\mathrm{i}b)\mathrm{e}^{(a+\mathrm{i}b)x} = z\mathrm{e}^{zx} \tag{7}$$

可见 e^z 的求导与实数域的 e^x 类似（式 9 [165]）.

三角函数（复数）

预备知识 指数函数（复数）[351]

定义

定义复数域的正弦函数为

$$\sin z = \frac{\mathrm{e}^{\mathrm{i}z} - \mathrm{e}^{-\mathrm{i}z}}{2\mathrm{i}} \tag{1}$$

定义复数域的余弦函数为

$$\cos z = \frac{\mathrm{e}^{\mathrm{i}z} + \mathrm{e}^{-\mathrm{i}z}}{2} \tag{2}$$

为什么三角函数要这么定义？因为只有这么定义，才能既“兼容”实数范围内的三角函数，同时满足解析的要求（暂不介绍解析的概念）.

与实数函数的“兼容性”

“兼容性”在这里指若将一个复变函数的自变量取实数，那么结果与使用同名的实数函数相同. 例如将式 1 中的复数 z 取实数 x，得

$$\sin z = \frac{\mathrm{e}^{\mathrm{i}x} - \mathrm{e}^{-\mathrm{i}x}}{2\mathrm{i}} \tag{3}$$

根据欧拉公式[351]，

$$\mathrm{e}^{\mathrm{i}x} = \cos x + \mathrm{i}\sin x \tag{4}$$

$$\mathrm{e}^{-\mathrm{i}x} = \cos x - \mathrm{i}\sin x \tag{5}$$

代入得

$$\sin z = \frac{(\cos x + \mathrm{i}\sin x) - (\cos x - \mathrm{i}\sin x)}{2\mathrm{i}} = \sin x \tag{6}$$

同理可证 $\cos z = \cos x$. 证毕.

两角和公式

利用欧拉公式，容易证明，复数范围内的正余弦函数同样满足两角和公式

$$\sin(z_1 + z_2) = \sin z_1 \cos z_2 + \cos z_1 \sin z_2 \tag{7}$$

$$\cos(z_1 + z_2) = \cos z_1 \cos z_2 - \sin z_1 \sin z_2 \tag{8}$$

实部和虚部

利用两角和公式，令 z_1 等于实数 x，z_2 等于虚数 $\mathrm{i}y$，则有

$$\sin z = \sin(x + \mathrm{i}y) = \sin x \cos \mathrm{i}y + \cos x \sin \mathrm{i}y \tag{9}$$

$$\cos z = \cos(x + \mathrm{i}y) = \cos x \cos \mathrm{i}y - \sin x \sin \mathrm{i}y \tag{10}$$

其中

$$\cos(\mathrm{i}y)=\frac{\mathrm{e}^{-y}+\mathrm{e}^{y}}{2}=\cosh y \tag{11}$$

$$\sin(\mathrm{i}y)=\frac{\mathrm{e}^{-y}-\mathrm{e}^{y}}{2\mathrm{i}}=\mathrm{i}\frac{\mathrm{e}^{y}-\mathrm{e}^{-y}}{2}=\mathrm{i}\sinh y \tag{12}$$

代入得

$$\sin z=\sin(x+\mathrm{i}y)=\sin x\cosh y+\mathrm{i}\cos x\sinh y \tag{13}$$

$$\cos z=\cos(x+\mathrm{i}y)=\cos x\cosh y-\mathrm{i}\sin x\sinh y \tag{14}$$

这样，就把正余弦的实部和虚部分开来了（当然也可以根据定义直接得到两式）.

复变函数的导数 柯西—黎曼条件

- 本词条存在未完成的内容.

预备知识 复变函数[346]，全微分[269]

定义 1

类似实函数的导数[158]，定义复变函数 $w=f(z)$（$f:\mathbb{C}\to\mathbb{C}$）的导数为

$$f'(z)=\lim_{h\to 0}\frac{f(z+h)-f(z)}{h} \tag{1}$$

其中 h 也是一个复数. 极限 $h\to 0$ 在这里是指 h 可以在复平面上以任意方式趋近于 0 都得到同一个极限值，否则极限不存在.

对于实变量的函数，只有正负两个方向趋于零，所以复变函数可导的条件更复杂.

定理 1　柯西—黎曼条件

令 $z=x+y\mathrm{i}$ 且

$$f(z)=u(x,y)+\mathrm{i}v(x,y) \tag{2}$$

那么 $f(z)$ 在该点可导的充分必要条件是：实函数 u,v 在复平面的某点 z 可微，且

$$\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y}\qquad\frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x} \tag{3}$$

该式被称为柯西—黎曼条件（**Cauchy-Riemann condition**）.

可见若柯西—黎曼条件成立，函数 u, v 的四个偏导数中只有两个是独立的.

在复平面的一个开集（链接未完成）D 上，如果函数 $f(z)$ 处处复可微，那么它就是一个**全纯函数（holomorphic function）**也叫做**解析函数（analytical function）**；如果除了一些孤立点外处处复可微，就叫**亚纯函数（meromorphic function）**.

推导

根据全微分[269]

$$\mathrm{d}w = \mathrm{d}u + \mathrm{i}\,\mathrm{d}v = \left(\frac{\partial u}{\partial x}\,\mathrm{d}x + \frac{\partial u}{\partial y}\,\mathrm{d}y\right) + \left(\frac{\partial v}{\partial x}\,\mathrm{d}x + \frac{\partial v}{\partial y}\,\mathrm{d}y\right)\mathrm{i} \tag{4}$$

如果直接由此计算 $\mathrm{d}w/\mathrm{d}z = (\mathrm{d}u + \mathrm{i}\,\mathrm{d}v)/(\mathrm{d}x + \mathrm{i}\,\mathrm{d}y)$ 会发现结果和 $\mathrm{d}y\,/\,\mathrm{d}x$ 有关，即与式 1 中 h 趋近于零点的方向有关. 所以我们换一种思路，直接令导数为

$$f'(z) = a(z) + b(z)\mathrm{i} \tag{5}$$

写成微分形式

$$\mathrm{d}w = f'(z)\,\mathrm{d}z \tag{6}$$

即

$$\mathrm{d}u + \mathrm{i}\,\mathrm{d}v = (a + b\mathrm{i})(\mathrm{d}x + \mathrm{i}\,\mathrm{d}y) = (a\,\mathrm{d}x - b\,\mathrm{d}y) + \mathrm{i}(b\,\mathrm{d}x + a\,\mathrm{d}y) \tag{7}$$

该式对比式 4 得

$$a = \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad b = -\frac{\partial u}{\partial y} = \frac{\partial v}{\partial x} \tag{8}$$

这样，不仅得到了柯西—黎曼条件，也得到了导数的表达式

$$f'(z) = \frac{\partial u}{\partial x} - \frac{\partial u}{\partial y}\mathrm{i} = \frac{\partial v}{\partial y} + \frac{\partial v}{\partial x}\mathrm{i} \tag{9}$$

可见我们仅需要 u, v 中的一个就可以求出导数，因为它们并不是独立的.

未完成：举例子，例如 $\exp, z^a$ 等都满足柯西—黎曼

解析函数与散度旋度

预备知识 复变函数的导数柯西—黎曼条件[354]

解析函数和矢量分析有紧密的关联，如果把复平面对应到 x-y 直角坐标系，那么 $f(z)$ 的实部和虚部 $u(z), v(z)$ 看作矢量的 x, y 分量，那么 $f(z)$ 就定义了一个复平面上的矢量场[294]

$$\mathbf{f}(x,y) = u(x,y)\hat{\mathbf{x}} + v(x,y)\hat{\mathbf{y}} \tag{1}$$

而 $f(z)$ 的复共轭 $f^*(z)$ 对应的二维矢量场为

$$\mathbf{g}(x,y) = u(x,y)\hat{\mathbf{x}} - v(x,y)\hat{\mathbf{y}} \tag{2}$$

定理 1

函数 $f:\mathbb{C}\to\mathbb{C}$ 在某区域解析的充分必要条件是：$\mathbf{g}(x,y)$ 是二维调和场，即散度[311] 和旋度[315] 都为零.

证明留做习题. 由于 $\mathbf{g}$ 的旋度为零，必定存在势函数 ϕ，使

$$u = \frac{\partial\phi}{\partial x} \qquad -v = \frac{\partial\phi}{\partial y} \tag{3}$$

且满足

$$\nabla^2\phi = 0 \tag{4}$$

即 $\mathbf{g}$ 必定可以表示为调和函数的梯度.

我们再来看 $\mathrm{i}f^*(z)$ 代表的矢量场 $\mathbf{h}(x,y)$.

$$\mathrm{i}f^*(z) = \mathrm{i}(u - v\mathrm{i}) = v + u\mathrm{i} \tag{5}$$

所以

$$\mathbf{h}(x,y) = v(x,y)\hat{\mathbf{x}} + u(x,y)\hat{\mathbf{y}} \tag{6}$$

由柯西公式易得 $\boldsymbol{\nabla}\times\mathbf{h} = \mathbf{0}$.

定理 2

函数 $f:\mathbb{C}\to\mathbb{C}$ 在某区域解析的充分必要条件是：$\mathbf{g}(x,y), \mathbf{h}(x,y)$ 旋度都为零.

证明留做习题.

复变函数的积分

- 整合中

预备知识 线积分[295]

我们之前已经接触过了实函数的积分，那么我们如何推广到复数上呢？实际上，同高等数学一样，也采用“分割”、“求和”、“取极限”的步骤来定义积分.

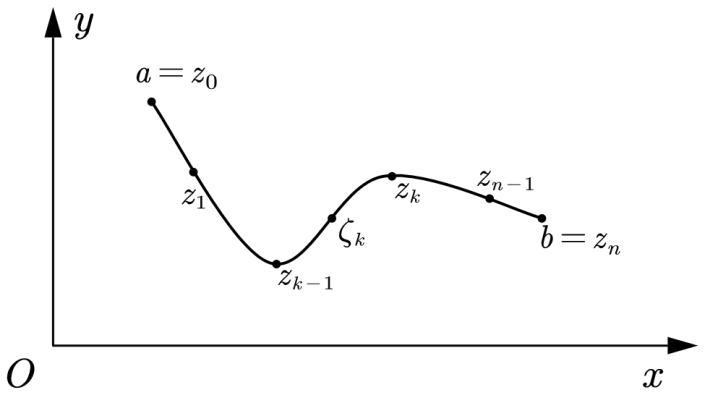


图 1: 复积分的定义

定义 1　复积分的定义

设 C 为复平面上一条起点在 a，终点在 b（$a,b\in\mathbb{C}$）的有向曲线，其方程为

$$z(t)=x(t)+\mathrm{i}y(t)\qquad(t_a\leqslant t\leqslant t_b)\tag{1}$$

函数 $f(z)$ 定义在 C 上．用一组点 $z_0,z_1,\ldots,z_n$ 沿曲线从 a 到 b 对曲线 C 进行分割．当 n 不断增加，设 $\Delta z_k=z_k-z_{k-1}$，ζ_k 为 z_{k-1},z_k 之间任意一点，分割 C 所得各个弧段长度中的最大值 d，不论对 C 的分法及 ζ_k 的取法如何，若存在极限

$$S=\lim_{d\to0}\sum_{k=1}^{n}f\left(\zeta_k\right)\Delta z_k\tag{2}$$

则称 $f(z)$ 沿 C（从 a 到 b）可积，称 S 为 $f(z)$ 沿 C（从 a 到 b）的积分，记作

$$S=\int_C f(z)\mathrm{d}z \tag{3}$$

其中 $f(z)$ 称为被积函数，C 称为积分路径.

现在来看一下用定义计算复积分的一个简单例子.

例 1

设 C 是一条起点为 a 终点为 b 的曲线，试计算 $S=\int_C \mathrm{d}z$.

解：用定义来计算该积分. 由定义，将 $f(z)=1$ 代入，得：

$$\int_C \mathrm{d}z=\lim_{d\to 0}\sum_{k=1}^{n}(z_k-z_{k-1})=b-a \tag{4}$$

通过这道例题，我们发现对于函数 $f(z)=1$，它的积分值只依赖于积分路径 C 的起点 a 和终点 b，而与积分路径的形状是无关的. 这个性质对于更一般的 f 是否成立？在后面的柯西积分定理[366] 中，我们将给出答案.

有了积分定义后，最先关心的问题是：积分存在的条件，积分的性质与积分的计算. 下面就来讨论这几个问题.

复积分与线积分

定理 1

令复平面的实轴为 x 轴，虚轴为 y 轴，用两个二元实函数 u,v 来表示 $f(z)$

$$f(z)=u(x,y)+\mathrm{i}v(x,y) \tag{5}$$

令两个矢量场为（$\mathbf{r}$ 为 xy 平面上的位置矢量）

$$\mathbf{f}(\mathbf{r})=u\hat{\mathbf{x}}-v\hat{\mathbf{y}}\qquad \mathbf{g}(\mathbf{r})=v\hat{\mathbf{x}}+u\hat{\mathbf{y}} \tag{6}$$

那么积分 $\int_C f(z)\mathrm{d}z$ 的实部和虚部分别可以看作两个矢量场 $\mathbf{f}(\mathbf{r}),\mathbf{g}(\mathbf{r})$ 在曲线 C 上的线积分

$$\begin{aligned}\int_C f(z)\,\mathrm{d}z&=\int_C \mathbf{f}(\mathbf{r})\cdot\mathrm{d}\mathbf{r}+\mathrm{i}\int_C \mathbf{g}(\mathbf{r})\cdot\mathrm{d}\mathbf{r}\\&=\int_C u\mathrm{d}x-v\mathrm{d}y+\mathrm{i}\int_C v\mathrm{d}x+u\mathrm{d}y\end{aligned} \tag{7}$$

研究复变函数的积分问题，可以转化为研究实变量的二元实值函数沿曲线 C 的线积分问题.

证明

$$\begin{aligned}\int_C f(z)\,\mathrm{d}z &= \int_C (u+iv)(\mathrm{d}x+\mathrm{i}\,\mathrm{d}y)\\ &= \int_C u\,\mathrm{d}x - v\,\mathrm{d}y + \int_C v\,\mathrm{d}y + u\,\mathrm{d}x\end{aligned} \tag{8}$$

证毕.

复积分的性质与计算

上面我们看到了将线积分和复积分联系起来的. 容易想到，线积分的一些性质可移到复变函数的积分上来.

若 $f(z)$ 与 $g(z)$ 沿曲线 C, C^-（C^- 表示与曲线 C 方向相反的同一条曲线）可积，则有

$$\int_C Af(z)\mathrm{d}z = A\int_C f(z)\mathrm{d}z \quad (A \text{ 为复常数}) \tag{9}$$

$$\int_C [Af(z)+Bg(z)]\mathrm{d}z = A\int_C f(z)\mathrm{d}z + B\int_C g(z)\mathrm{d}z \tag{10}$$

$$\int_C f(z)\mathrm{d}z = -\int_{C^-} f(z)\mathrm{d}z \tag{11}$$

$$\int_C f(z)\mathrm{d}z = \int_{C_1} f(z)\mathrm{d}z + \int_{C_2} f(z)\mathrm{d}z \quad \big(C \text{ 由} C_1 \text{ 与} C_2 \text{ 首尾相接而成}\big) \tag{12}$$

设 L 为曲线 C 的长度，若 $f(z)$ 沿 C 可积，且在 C 上满足 $|f(z)| \leqslant M$，则

$$\left|\int_c f(z)\mathrm{d}z\right| \leqslant \int_c |f(z)|\mathrm{d}s \leqslant ML \tag{13}$$

该式提供了一种估计复变函数积分的模的方法.

未完成：以下待审阅

到现在为止，计算复变函数积分只有两种方法，一是定义，二是. 有无其他方法呢？

我们注意到，由于积分路径常取光滑曲线（或逐段光滑曲线），所以 $f(z)$ 沿曲线 C 的积分可归结为 $f[z(t)]$ 关于曲线 C 的参数的积分.

事实上，若 C 为光滑曲线（或逐段光滑曲线）

$$\begin{aligned} z &= z(t) \\ &= x(t) + \mathrm{i}y(t), (t_a \leqslant t \leqslant t_b) \end{aligned} \tag{14}$$

则 $z'(t)$ 在 $[t_a, t_b]$ 上连续，且 $z'(t) = x'(t) + \mathrm{i}y'(t) \neq 0$. 再设 $f(z)$ 在 C 上连续及

$$\begin{aligned} f[z(t)] &= u[x(t), y(t)] + \mathrm{i}v[x(t), y(t)] \\ &= u_1(t) + \mathrm{i}v_1(t) \end{aligned} \tag{15}$$

则有

$$\begin{aligned} \int_C f(z)\mathrm{d}z &= \int u\mathrm{d}x - v\mathrm{d}y + \mathrm{i}\int_C v\mathrm{d}x + u\mathrm{d}y \\ &= \int_{t_a}^{t_b} \{u[x(t), y(t)]x'(t) - v[x(t), y(t)]y'(t)\}\,\mathrm{d}t + \\ &\mathrm{i}\int_{t_a}^{t_b} \{v[x(t), y(t)]x'(t) + u[x(t), y(t)]y'(t)\}\,\mathrm{d}t \\ &= \int_{t_a}^{t_b} [u_1(t) + \mathrm{i}v_1(t)]\,[x'(t) + \mathrm{i}y'(t)]\,\mathrm{d}t \\ &= \int_{t_a}^{t_b} f[z(t)]\mathrm{d}t \end{aligned} \tag{16}$$

这样一来，便将 $f(z)$ 沿曲线 C 的积分归结为 $f(z)$ 关于曲线 C 的参数 t 的积分.

用式 16 来计算积分 $\int_C f(z)\mathrm{d}z$ 包含三个步骤：

1. 写出曲线 C 的方程 $z = z(t) = x(t) + iy(t), t_a \leqslant t \leqslant t_b$
2. 将 $z = z(t)$ 与 $\mathrm{d}z = z'(t)\mathrm{d}t$ 代入所求积分 $\int_c f(z)\mathrm{d}z$ 中
3. 计算式 16 右端的关于参数 t 的积分.

有了上述求复变函数的积分的方法之后，我们一起来看几个例子，并尝试用不同的方法来求解.

例 2

计算 $\int_C z\mathrm{d}z$，其中 C 为起点在 a 终点在 b 的一条逐段光滑曲线.

我们用定义来计算一下.

由题知，$f(z)=z$．取 $\zeta_k=z_{k-1}$，则 $f(\zeta_k)=z_{k-1}$，作和：

$$A_n=\sum_{k=1}^{n} f(\zeta_k)\,\Delta z_k=\sum_{k=1}^{n} z_{k-1}(z_k-z_{k-1}) \tag{17}$$

因 $f(z)=z$ 在 C 上连续，所以积分一定存在，于是

$$\int_C f(\mathrm{z})\mathrm{dz}=\lim_{\substack{n\to\infty\\ d\to 0}} A_n \tag{18}$$

另外，若取 $\zeta_k=z_k$，则 $f(\zeta_k)=z_k$，作和：

$$B_n=\sum_{k=1}^{n} f(\zeta_k)\,\Delta z_k=\sum_{k=1}^{n} z_k(z_k-z_{k-1}) \tag{19}$$

由之前的讨论知，

$$\int_C f(\mathrm{z})\mathrm{dz}=\lim_{\substack{n\to\infty\\ d\to 0}} B_n \tag{20}$$

由式 18 和式 20 知，

$$\begin{aligned}\int_C f(z)\mathrm{d}z&=\frac{1}{2}\lim_{\substack{n\to\infty\\ d\to 0}}(A_n+B_n)\\ &=\frac{1}{2}\left(b^2-a^2\right)\end{aligned} \tag{21}$$

我们再次发现：$f(z)=z$ 也有一个和 $f(z)=1$ 一样好的性质：它沿曲线 C 的积分只依赖于起点与终点，而与 C 的形状无关！

例 3

试计算 $\int_C x\mathrm{d}z$，其中 C 为起点在 $z=0$，终点在 $z=1+\mathrm{i}$ 的直线段．

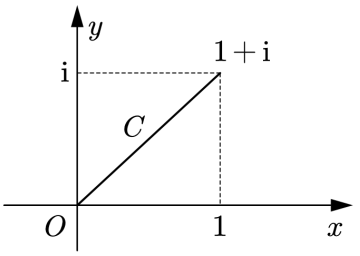


图 2: 积分路径

我们用式 16 计算．首先，写出 C 的方程为

$$z=z(t)=(1+i)t\qquad(0\leqslant t\leqslant 1) \tag{22}$$

由 $x+\mathrm{i}y=z=z(t)=(1+\mathrm{i})t$ 得 $\mathrm{d}z=z'(t)\mathrm{d}t=(1+\mathrm{i})\mathrm{d}t$，将 $x=t$ 和 $\mathrm{d}z$ 代入所求积分，得

$$\int_C x\mathrm{d}z=\int_0^1(1+\mathrm{i})t\mathrm{d}t \tag{23}$$

最后，计算上式右端的积分得

$$\int_C x\mathrm{d}z=\frac{1+\mathrm{i}}{2} \tag{24}$$

例 4

设 Γ 为由 C_1 与 C_2 首尾相接而成的起点在 $z=0$，终点在 $z=1+\mathrm{i}$ 的曲线，如图 3 所示．计算 $\int_\Gamma x\mathrm{d}z$．

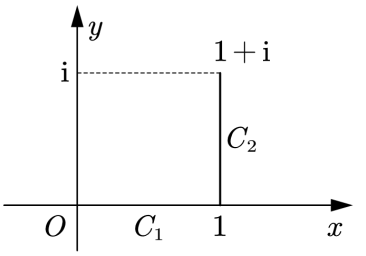


图 3: 积分路径

首先写出 C_1 和 C_2 的方程为

$$\begin{aligned}&C_1:z=z(t)=t(0\leqslant t\leqslant 1)\\&C_2:z=z(t)=1+\mathrm{i}\,t(0\leqslant t\leqslant 1)\end{aligned} \tag{25}$$

下面分别计算 C_1 和 C_2．

在 C_1 上有 $x=t,\mathrm{d}z=\mathrm{d}t$，代入 $\int_{C_1}x\mathrm{d}z$ 得 $\int_{C_1}x\mathrm{d}z=\int_0^1 t\mathrm{d}t$．在 C_2 上有 $x=\mathrm{Re}\,z(t)=1,\mathrm{d}z=\mathrm{i}\mathrm{d}t$，代入 $\int_{C_2}x\mathrm{d}z$ 得到 $\int_{C_2}x\mathrm{d}z=\int_0^1\mathrm{i}\mathrm{d}t$．

最后，计算积分得

$$\int_\Gamma x\mathrm{d}z=\int_{C_1}x\mathrm{d}z+\int_{C_2}x\mathrm{d}z=\int_0^1 t\mathrm{d}t+\int_0^1\mathrm{i}\mathrm{d}t=\frac{1}{2}+\mathrm{i} \tag{26}$$

比较例 3 和例 4，我们发现，对于复变函数 $f(z)=x$，积分的结果却与积分路径的选取很有关系．

再来介绍一个十分重要的例题．

例 5

试证：

$$\oint_C \frac{1}{(z-a)^n} = \begin{cases} 2\pi\mathrm{i}, & n=1 \\ 0, & n\neq 1 \text{ 且为整数.} \end{cases} \tag{27}$$

其中 C 是为以 a 为圆心，以 R 为半径的圆周.

证明：这里的 C 是一条围线，对于沿围线的积分，若无特殊声明，则今后总理解为沿围线的正向积分.

用计算积分的方法来证明本题.

首先，写出 C 的方程为

$$\begin{aligned} z &= z(t) \\ &= a + R\mathrm{e}^{\mathrm{i}\theta} \quad (0 \leqslant \theta \leqslant 2\pi) \end{aligned} \tag{28}$$

其次，$\mathrm{d}z = \mathrm{i}R\mathrm{e}^{\mathrm{i}\theta}\mathrm{d}\theta$，被积函数当 $n=1$ 时为 $(z-a)^{-1} = R^{-1}\mathrm{e}^{-\mathrm{i}\theta}$，当 $n\neq 1$ 且为整数时，为 $(z-a)^{-n} = R^{-n}\mathrm{e}^{-\mathrm{i}n\theta}$

再次，计算积分，$n=1$ 时，有

$$\begin{aligned} \oint_c \frac{1}{z-a}\mathrm{d}z &= \int_0^{2\pi} \frac{\mathrm{i}R\mathrm{e}^{\mathrm{i}\theta}}{R\mathrm{e}^{\mathrm{i}\theta}}\mathrm{d}\theta \\ &= 2\pi\mathrm{i} \end{aligned} \tag{29}$$

当 $n\neq 1$ 且为整数时，有

$$\begin{aligned} \oint_C \frac{1}{(z-a)^n}\mathrm{d}z &= \int_0^{2\pi} \frac{\mathrm{i}R\mathrm{e}^{\mathrm{i}\theta}}{R^n\mathrm{e}^{\mathrm{i}n\theta}}\mathrm{d}\theta \\ &= \frac{\mathrm{i}}{R^{n-1}} \int_0^{2\pi} \mathrm{e}^{-\mathrm{i}(n-1)\theta}\mathrm{d}\theta \\ &= \frac{\mathrm{i}}{R^{n-1}} \left[\int_0^{2\pi} \cos(n-1)\theta\mathrm{d}\theta - \mathrm{i}\int_0^{2\pi} \sin(n-1)\theta\mathrm{d}\theta \right] = 0 \end{aligned} \tag{30}$$

综上所述，问题得证.

例 6

计算积分

$$\oint_{|z-1|=1} \frac{z}{z-1}\mathrm{d}z \tag{31}$$

由积分路径为 $|z-1|=1$，得 $z-1=\mathrm{e}^{\mathrm{i}\theta}(0\leqslant\theta\leqslant 2\pi)$．于是，积分路径的方程为

$$z=z(\theta)=1+\mathrm{e}^{\mathrm{i}\theta},(0\leqslant\theta\leqslant 2\pi) \tag{32}$$

故

$$\begin{aligned}\oint_{z-1}\frac{z}{z-1}\mathrm{d}z&=\int_0^{2\pi}\frac{1+\mathrm{e}^{\mathrm{i}\theta}}{\mathrm{e}^{\mathrm{i}\theta}}\mathrm{i}\mathrm{e}^{\mathrm{i}\theta}\mathrm{d}\theta\\&=\int_0^{2\pi}\mathrm{i}\mathrm{d}\theta+\int_0^{2\pi}(\cos\theta+\mathrm{i}\sin\theta)\mathrm{d}\theta\\&=2\pi\mathrm{i}\end{aligned} \tag{33}$$

例 7

计算积分

$$\oint_{|z-2|=1}\frac{3z^2}{(z-2)^3}\mathrm{d}z \tag{34}$$

由 $|z-2|=1$ 得积分路径的方程为

$$z=z(\theta)=2+\mathrm{e}^{\mathrm{i}\theta},(0\leqslant\theta\leqslant 2\pi) \tag{35}$$

于是

$$\begin{aligned}\oint_{|z-2|=1}\frac{3z^2}{(z-2)^3}\mathrm{d}z&=\int_0^{2\pi}\frac{3\left(2+\mathrm{e}^{\mathrm{i}\theta}\right)^2}{\left(\mathrm{e}^{\mathrm{i}\theta}\right)^3}\mathrm{i}\mathrm{e}^{\mathrm{i}\theta}\mathrm{d}\theta\\&=3\mathrm{i}\int_0^{2\pi}\left(4\mathrm{e}^{-\mathrm{i}2\theta}+4\mathrm{e}^{\mathrm{i}\theta}+1\right)\mathrm{d}\theta\\&=6\pi\mathrm{i}\end{aligned} \tag{36}$$

牛顿—莱布尼兹公式（复变函数）

- 本词条处于草稿阶段.

预备知识 柯西—黎曼条件[354]，复变函数的积分[357]

我们来回顾定理 1 [357]：复积分的实部和虚部可以分别表示为两个二维矢量场的线积分

$$\int_C f(z)\,\mathrm{d}z=\int_C\mathbf{f}(\mathbf{r})\cdot\mathrm{d}\mathbf{r}+\mathrm{i}\int_C\mathbf{g}(\mathbf{r})\cdot\mathrm{d}\mathbf{r} \tag{1}$$

其中

$$\mathbf{f}(\mathbf{r}) = u\hat{\mathbf{x}} - v\hat{\mathbf{y}} \qquad \mathbf{g}(\mathbf{r}) = v\hat{\mathbf{x}} + u\hat{\mathbf{y}} \tag{2}$$

什么情况下，该积分只和积分曲线的起点和终点有关呢？根据斯托克斯定理[318]，充分必要条件是

如果其中的函数在某区域 D 解析，即处处满足柯西—黎曼条件，那么

定理 1

令复平面的实轴为 x 轴，虚轴为 y 轴，用两个二元实函数 u, v 来表示 $f(z)$

$$f(z) = u(x, y) + \mathrm{i}v(x, y) \tag{3}$$

令两个矢量场为（$\mathbf{r}$ 为 xy 平面上的位置矢量）

那么积分 $\int_C f(z)\mathrm{d}z$ 的实部和虚部分别可以看作两个矢量场 $\mathbf{f}(\mathbf{r}), \mathbf{g}(\mathbf{r})$ 在曲线 C 上的线积分

$$\begin{aligned} \int_C f(z)\,\mathrm{d}z &= \int_C \mathbf{f}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} + \mathrm{i}\int_C \mathbf{g}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} \\ &= \int_C u\mathrm{d}x - v\mathrm{d}y + \mathrm{i}\int_C v\mathrm{d}x + u\mathrm{d}y \end{aligned} \tag{4}$$

而柯西—黎曼条件[354] 恰好规定了这两个矢量场的旋度为零，所以如果 $f(z)$ 在考虑的区域上解析，线积分结果只和起点和终点有关，与路径无关（所有的路径必须在解析的区域内）. 于是可以得到类似于牛顿—莱布尼兹公式[202]，令 $\mathbf{f}_R, \mathbf{f}_I$ 的势函数分别为 F_R, F_I，即

$$\boldsymbol{\nabla} F_R = f_R \qquad \boldsymbol{\nabla} F_I = f_I \tag{5}$$

再令

$$F(\mathbf{r}) = F_R(\mathbf{r}) + \mathrm{i}F_I(\mathbf{r}) \tag{6}$$

有

$$\int_{z_1}^{z_2} f(z)\,\mathrm{d}z = F(z_2) - F(z_1) \tag{7}$$

对于任意路径成立. 容易证明 $F(z)$ 就是 $f(z)$ 的原函数，即

$$F'(z) = f(z) \tag{8}$$

这相当于对式 7 实部和虚部分别使用梯度定理[307].

柯西积分定理

预备知识 复变函数的积分[357]

在复变函数的积分[357]里的例子可以发现，有的函数的积分只依赖于积分路径的起点与终点，而与积分路径的形状无关，而有的函数，其积分不仅与积分路径的起点与终点有关，而且与积分路径的形状也有关．深入观察后，可知，前一类函数是解析函数．由此，可提出猜想：解析函数的积分只依赖于积分路径的起点与终点，而与积分路径的形状无关．柯西在 1825 年给出此定理对猜想作了回答．也就是我们现在要介绍的**柯西积分定理（Cauchy's integral theorem）**，也叫**柯西—古萨定理（Cauchy-Goursat theorem）**．

定理 1　柯西积分定理

设 G 为复平面上的单连通区域，C 为 G 内的任意一条围线，如图所示，若 $f(z)$ 在 G 内解析，则

$$\oint_C f(z)\mathrm{d}z = 0 \tag{1}$$

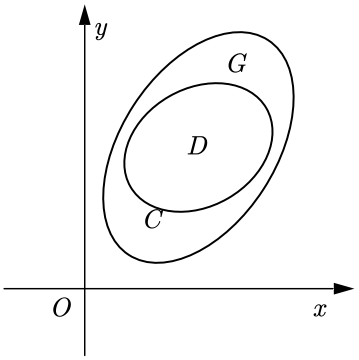


图 1: 积分回路

下面为了简单起见，在假设 $f'(z)$ 连续的情况下证明该定理，也就是黎曼在 1851 年给出的证明方法．

令 $z = x + \mathrm{i}y, f(z) = u(x,y) + \mathrm{i}v(x,y)$，我们有

$$\oint_C f(z)\mathrm{d}z = \oint_C u\mathrm{d}x - v\mathrm{d}y + \mathrm{i}\oint_C v\mathrm{d}x + u\mathrm{d}y \tag{2}$$

由于 $f'(z)=u_x+\mathrm{i}v_x=v_x-\mathrm{i}u_y$ 在 G 内连续，所以，u_x,u_y,v_x,v_y 在 G 内连续，从而，这四个偏导数在由围线 C 及其内部构成的闭区域 D 上连续．又因 C 为光滑或逐段光滑的闭曲线，且 u 与 v 在 D 上连续是显然的．于是，由高等数学中的格林公式或斯托克斯定理[315] 得

$$\begin{aligned}\oint_C u\mathrm{d}x-v\mathrm{d}y&=\iint_D(-v_x-u_y)\,\mathrm{d}x\mathrm{d}y\\ \oint_C v\mathrm{d}x+u\mathrm{d}y&=\iint_D(u_x-v_y)\,\mathrm{d}x\mathrm{d}y\end{aligned}\tag{3}$$

而由 $f(z)$ 在 G 内解析知道，u 与 v 满足柯西—黎曼条件（式 3 [354]）

$$u_x=v_y,\quad u_y=-v_x\tag{4}$$

由此得式 3 中两个积分都为零，即

$$\oint_C f(z)\mathrm{d}z=0\tag{5}$$

这个定理其实也提供了一种计算解析函数沿围线积分的方法．下面来看一道例题．

例 1

计算积分

$$\oint_{|z+1|=\frac{1}{4}}\frac{3z}{(z-2)(z-1)}\mathrm{d}z\tag{6}$$

作区域 $G:|z+1|<1$，显然，积分路径 $|z+1|=1/4$ 位于 G 内，由柯西积分定理，积分结果为 0．

第三部分

线性代数

第一章

几何矢量

几何矢量

预备知识 集合[92]，充分必要条件[116]

我们来回顾高中学的**几何矢量（geometric vector）**，也叫**欧几里得矢量（Euclidean vector）**．本文中简称为“矢量”．粗略地说，矢量是空间中的一些有长度有方向的箭头．我们对它的位置不感兴趣，所有长度和方向相同的矢量都视为同一矢量．本书中矢量用正黑体表示，如 $\mathbf{a}$．在手写时，可以在字母上方加箭头表示，如 $\vec{a}$．特殊地，如果一个矢量的长度等于 1，那么它就是一个**单位矢量**，本书中在矢量上面加上“^”符号表示单位矢量，如 $\hat{\mathbf{a}}$．为了与矢量区分，我们把单个的实数或复数称为**标量**．

几何矢量

矢量（vector）又被翻译为**向量**．这个概念最初来自直观的几何矢量．早在民国时期，学者们引入 vector 的概念时就将它译成了汉语．由于当时的物理学家和数学家没有太多交集，物理学家将它翻译为向量，而数学家翻译为矢量．90 年代时，国家名词委员会商议确定一个统一的 vector 译名，但却无法轻易割舍两个译名中的任何一个，因为它们都非常信达雅地表明了 vector 的含义：向量即有方向的量，矢量即像箭矢一样的量．大概是出自物理学家和数学家的互相尊重，最后确定的方案是双方互换译名，从此物理学界称矢量，数学界称向量[1]．**本书中不区分两个译名的使用**．

本词条中讨论的几何矢量是经典物理学中最长见的一类矢量，是一种具有**长度**和**方向**的量，因此可以画成箭头来表示．生活中这样具有长度和方向的量十分常见：速度有大小有方向，因此可以表示为箭头，箭头的长度代表速度的大小；加速度[881] 也有大小，有方向；我从一个地方运动到另一个地方，那么

[1] 以下加粗部分为力学家朱照宣教授的回忆．**在 20 世纪 90 年代初，国家名词委为此（vector）召开会议，想协调双方，由主任钱三强亲自主持．我曾戏称这是个“一字会”．当时的情况是，学科有分支，术语有派生，犹如家族有后裔．祖宗互相谦让，但子孙繁多，已无法协调．钱先生在会上没有说倾向于哪方面的话．矢量、向量的分歧，一直维持到今．力学这学科，和数学、物理同样有“亲”，力学中 vector 用什么？当年我在“一字会”后还有情绪，埋怨钱先生作为领导“不表态”．过了好些年，才懂得这类事，最多只能因势利导，不能靠行政命令或专家拍板．事实上，台湾物理界至今用的是还“向量”．**

从起点到终点可以画一根箭头，这箭头就是位移矢量[878]．不止在生活中，一切领域里具有方向和大小概念的量都被称为几何矢量．

为什么要强调是“几何”矢量呢？在数学中，矢量的含义要比几何矢量更广泛，也就是说，几何矢量虽然是数学家所研究的矢量的一种，但不是唯一的一种，广义来说任何矢量空间[444] 中的元素都叫做矢量．以后会看到，本文介绍的矢量是一个“实数域上的赋范线性空间”中的元素．这个名字听起来很吓人，但它其实只是三个概念的结合：实数域，范数和线性空间．实数域即我们所熟悉的实数，它作为一个域使得我们可以用它来构建线性空间，而范数则是线性空间上一种表达向量的“绝对值”的函数．这些概念我们都会在对应词条中深入讨论．

零矢量

特别地，长度为零的矢量称为**零矢量（zero vector）**，不作方向区分（可以认为它与任何矢量平行）．零矢量依然是矢量，要注意和数字零进行区分．

几何矢量与起点无关

需要注意的是，为了避免引入过多的要素从而导致概念过于复杂，我们目前讨论的几何矢量只有两个本质属性，长度和方向．将任一矢量进行任意平移后，它仍然是同一个矢量[2]．

当然，如果矢量平移并不改变矢量本身，我们完全可以把所有几何矢量的起点都挪到一起，以这个公共起点为**原点**，那么空间中**矢量的终点所在点**就和**矢量本身**一一对应．这可能会带来一个问题：我们知道，杠杆原理中力的作用效果和力的作用点相关，而力又是矢量，这似乎和“平移不变”矛盾．实际上，在描述杠杆原理等规律时，我们使用两个矢量，一个是力的矢量，一个则是力的作用点的位置矢量[878]，位置矢量的原点根据问题需要来选取．

当把所有矢量的起点都挪到原点后，我们就可以用终点的几何坐标来描述几何矢量．

[2]在微分几何中，我们讨论一种叫“切向量”的对象，这就是一种和所在位置有关的抽象的向量．

坐标和维度

从现在开始，如无特别说明，我们默认几何矢量起点在原点．这样，我们就可以把一个矢量理解为空间中的一个点，即它的终点．本节仅从直觉上引入坐标的概念，基本上是符合多数人的几何直觉的．要更准确地描述相关概念，请参考“基底和坐标[379]”.

我们知道，实数轴可以看成一个一维的几何空间．如果以数字 0 所在的点为原点，那么我们也可以把这根轴本身看成一个几何矢量的集合，每个数字所在的点都是一个几何矢量．如图 1 所示，点 P 表示一个长度为 3.14 的矢量，而点 Q 表示一个长度为 6 的矢量．这样一来，每个实数 x 都可以表示一个矢量，其长度为 $|x|$；当 $x>0$ 时，对应的矢量指向正方向，当 $x<0$ 时指向负方向，而对于 $x=0$，它对应的是零矢量，而零矢量没有方向之分．**任何方向的零矢量都是同一个矢量**.

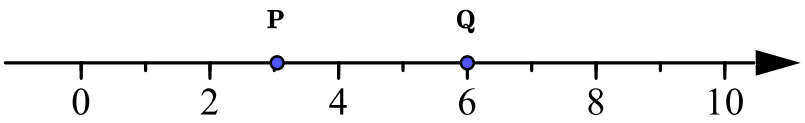


图 1: 实数轴上的几何矢量.

实数轴上的矢量只有一个方向可以选择，虽然有正负之分，但都沿着一条线．因此只需要用一个实数就可以唯一地表示一个矢量．这样的实数，就是对应矢量在实数轴上的一个坐标，该实数的绝对值就是矢量的“长度”.

这里所说的矢量的长度和日常经验可能有所不同．当我们讨论一个长度为 1 m 的位移矢量时，我们可以有不同的坐标来描述这一个矢量，或者说，可以把 1 m 对应到不同的实数．如果我们用每格长 1 m 的坐标轴去度量，那么这个矢量的“长度”就是 1；如果用每格长 1 mm 的坐标轴去度量，那么其“长度”就是 1000．因此我们所说的长度并不具有绝对的意义，而只有相对的意义[3]：同一个矢量的长度在不同坐标表示下可能不一样，但是无论在什么坐标表示下，两条“共线”的矢量的长度之比都是一样的.

实数轴上的矢量没有方向区分（最多是正负之分），因此不用考虑方向，方便我们专心讨论长度问题．实数轴被称为一个**一维**的几何矢量空间，因为所有

[3] 严格来说，这是因为对于同一个矢量空间，我们可以赋予不同的范数/内积[471].

矢量只需要**一个数字**就能**唯一**确定.

是否存在需要更多数字才能确定一个矢量的空间呢？当然有，二维平面就是这样空间，其上的矢量就需要两个数字来描述. 这个时候矢量就有方向之分了，如图 2 中矢量 P 的方向“北偏东 $45°$”和 Q 的方向“正东”就有所区别了.

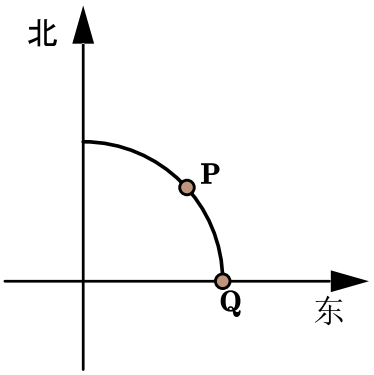


图 2: 二维平面上的几何矢量.

因为矢量是“具有长度和方向”的量，我们可以用一个数字来表示矢量的长度，一个数字来表示矢量的方向（从给定轴算起逆时针旋转的角度，即辐角），这就是极坐标[119] 的表示方法. 当然，我们更熟悉的是用直角坐标系来表示，如图 2 所示，如果矢量 P 和 Q 的长度都是 1，但 P 的辐角为 $\pi/4$，Q 的辐角为 0，那么直角坐标系下就可以把 P 表示为 $(\sqrt{2}/2, \sqrt{2}/2)$，把 Q 表示为 $(1, 0)$.

同理，三维空间中的矢量需要用三个数字来确定.

矢量的坐标表示，不一定直接体现了矢量的两个要素：“方向”和“长度”，但是一定蕴含了这两个要素. 比如说，图 3 中矢量 P 的长度并没有直接出现在其坐标 (a_x, a_y, a_z) 中，但是可以用坐标计算出来：$\sqrt{a_x^2 + a_y^2 + a_z^2}$.

我们容易想象出 1 到 3 维的几何矢量空间以及它们的两种运算，但却很难想象更高维的情况. 作为本章的主要目的，当你理解了本章的所有甚至只是部分词条后，应当可以熟悉该如何讨论任意维度的空间. 讨论的核心在于，n 维空间中的每个矢量都需要 n 个数字来确定，我们可以“定义”一个直角坐标系，把 3 维直角坐标系中的性质推广开来.

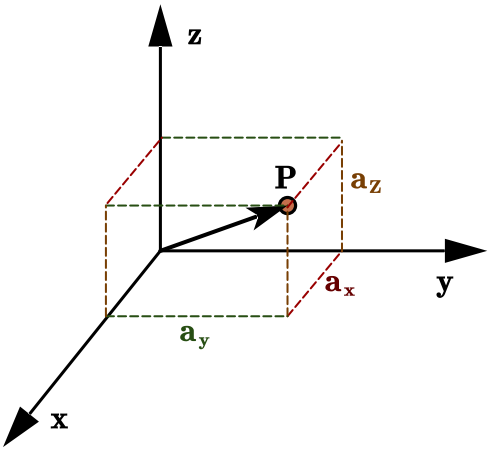


图 3: 三维空间里的几何矢量.

拓展：斜坐标系

当笛卡尔首次提出“坐标”的概念时，他并没有指定必须是“直角坐标”. 实际上，他认为任何两根不平行的直线都可以用来刻画二维平面上的坐标，具体方式详见**斜坐标系**[1265]. 不只是二维平面，任意维度的空间中都可以使用斜坐标系来描述空间中的点.

使用直角坐标系的好处是高度对称，因为当你选定相互垂直的 x 轴和 y 轴以后，y 轴的两边到 x 轴的角度都是 $\pi/2$；另外，几何学中的勾股定理依赖于直角；虽然我们也可以用余弦定理来描述任意三角形的边长关系，但比起勾股定理要麻烦得多，因为表达式中多了一项. 使用直角坐标系能让讨论简洁很多.

参考书推荐

对于几何矢量在高中数学以及几何学中的用途，张景中院士的《绕来绕去的向量法》一书的内容通俗易懂且内容丰富，单墫的《向量与立体几何（数学奥林匹克命题人讲座）》也是一本实用的小册子. 感兴趣的读者可用这两本读物作参考. 另外，数学界也借用物理中“质心”的概念，发展出了“质点几何学”分支，其本质仍然和向量几何一模一样，只不过换了个观点看问题，实质上是仿射几何的一种优化版表述；在上述张景中院士的书中提到了质点几何学，莫绍揆教授也出版了专门讨论此话题的《质点几何学》，但并不建议读者

特意花太多精力学习[4].

几何矢量的运算

预备知识 几何矢量[370]

把所有几何矢量收集到一起，构成一个集合，本身没什么可以研究的. 在集合[92] 中我们说过，没有任何附加结构的集合中，元素叫什么都不重要，只有元素数量重要.

幸运的是，矢量集合有讨论的意义，因为我们可以引入了矢量的运算，这样，矢量集合就有了可以研究的运算结构.

矢量的加减和数乘

矢量的加法

两个矢量相加，结果是另一个矢量，具体定义如下：如图 1，两个矢量相

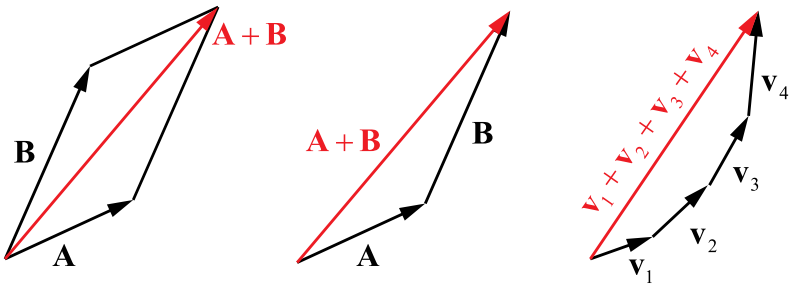


图 1: 矢量的加法

加，既可以使用平行四边形法则，也可以用三角形法则. **平行四边形法则**是指先将两个矢量移动到共同的起点，然后以它们为边做一个平行四边形，再由对角线得到相加后的矢量. **三角形法则**是指将第二个矢量的起点移动到第一个矢量的终点，然后作出从第一个矢量起点指向第二个矢量终点的矢量. 容易证明，二者的结果是一样的.

[4]如果你好奇的话，只需要知道向量法中的**定比分点公式**就是连接质点几何和向量几何的桥梁，就可以把几何问题在这两个观点之间互相转化了.

若有多个矢量连续相加，我们既可以依次使用平行四边形法则，也可以分别把它们依次首尾相接，结果就是由起点指向终点的矢量．二者结果也是一样的，证明留作习题．

容易证明矢量的加法满足**交换律（commutative property）**

$$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A} \tag{1}$$

以及**结合律（associative property）**

$$(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C}) \tag{2}$$

矢量的数乘 共线

第二个矢量运算，是一个矢量和一个数字的乘积，得到一个矢量，称为数乘，我们用例子定义如下．

如图 2，一个矢量与一个正实数相乘，则方向不变，把长度乘以这个实数．若这个数是负数，则把矢量取反方向再把长度乘以这个实数数的绝对值即可．若 λ,μ 表示实数，容易证明分配律 $\lambda(\mathbf{A}+\mathbf{B})=\lambda\mathbf{A}+\lambda\mathbf{B}$ 和 $(\lambda+\mu)\mathbf{A}=\lambda\mathbf{A}+\mu\mathbf{A}$，结合律 $\lambda(\mu\mathbf{A})=(\lambda\mu)\mathbf{A}$．

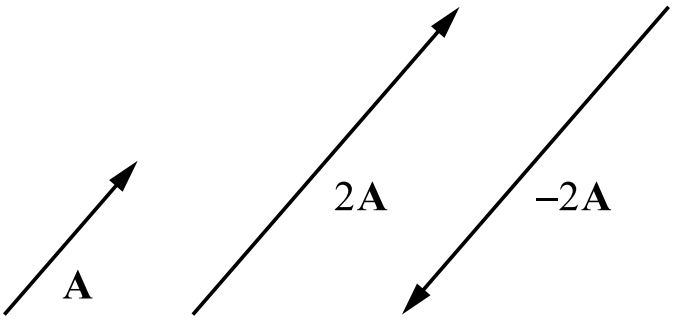


图 2: 矢量的数乘

如果两个矢量的关系可以用 $\mathbf{A}=\lambda\mathbf{B}$ 表示，那么它们就是**共线**的．共线的充分必要条件[116] 是，两矢量方向相同或相反．

特殊地，如果把一个矢量乘以 -1，那么它的长度不变，而方向相反．我们把这个新矢量叫做原矢量的**逆矢量（inverse vector）**．矢量 $\mathbf{A}$ 的逆记为 $-\mathbf{A}$．

矢量的减法

由了加法和数乘，我们并不需要另外定义所谓的**矢量减法**，只需要把 $\mathbf{A}-\mathbf{B}$ 看成 $\mathbf{A}$ 加上 $\mathbf{B}$ 的逆即可，即

$$\mathbf{A}-\mathbf{B}\equiv\mathbf{A}+(-\mathbf{B}) \tag{3}$$

几何上来说，容易证明（留做习题）要计算 $\mathbf{A}-\mathbf{B}$，就先把它们的起点移动到一起，然后画出从 $\mathbf{B}$ 的终点指向 $\mathbf{A}$ 的终点的矢量即可.

移项

矢量的加减法和我们熟知的标量加减法有许多相似之处，我们甚至可以像标量移项一样对几何矢量的表达式进行移项. 例如我们有表达式

$$\mathbf{A}+\mathbf{B}=\mathbf{C}+\mathbf{D} \tag{4}$$

显然在等式两边同时加上或减去相同的矢量仍然可以使等式成立，于是两边同时减 $\mathbf{C}$ 再同时减 $\mathbf{D}$ 得

$$\mathbf{A}+\mathbf{B}-\mathbf{C}-\mathbf{D}=\mathbf{0} \tag{5}$$

这看起来就像我们把等号右边得项移动到了等号左边，并添加了负号.

一个小技巧是，在画矢量减法 $\mathbf{A}-\mathbf{B}=\mathbf{C}$ 时，如果你忘记了 $\mathbf{C}$ 是从 $\mathbf{A}$ 指向 $\mathbf{B}$ 还是 $\mathbf{B}$ 指向 $\mathbf{A}$，那么可以检查一下画出来的三个矢量是否满足 $\mathbf{A}=\mathbf{B}+\mathbf{C}$.

矢量的线性组合

把有限个矢量 $\mathbf{v}_i$ 分别与若干实数 c_i 相乘再相加就得到了这些矢量的一个**线性组合**

$$\sum_i^N c_i\mathbf{v}_i=c_1\mathbf{v}_1+c_2\mathbf{v}_2+\cdots+c_N\mathbf{v}_N \tag{6}$$

注意若无特别说明，线性组合仅指**有限个**矢量的数乘和加法.

根据矢量加法和数乘的定义，容易得知任何有限个矢量的任何线性组合仍然是一个矢量.

习题 1

试说明任意两个不共线的矢量的所有线性组合都会落在同一个平面上.

线性相关性

预备知识 几何矢量的运算[375]

如果存在至少一组**不全为零**系数 c_i 使几个矢量的线性组合等于零，这些矢量就被称为**线性相关（linearly dependent）**的

$$\sum_{i}^{N} c_i \mathbf{v}_i = \mathbf{0} \tag{1}$$

这是因为对于任何一个 c_j 不为零的项，矢量 $\mathbf{v}_j$ 都可以表示为其他矢量的线性组合．只需把上式除以 c_j 即可

$$\mathbf{v}_j = -\sum_{i\neq j} \frac{c_i}{c_j} \mathbf{v}_i \tag{2}$$

如果不存在满足式 1 的非零系数 c_i，这些矢量就是**线性无关（linearly independent）**的，即任何矢量都不可能被其他矢量的线性组合表示．

例 1

我们来看在三维几何矢量空间中，线性无关有什么几何意义．若两个矢量 $\mathbf{v}_1$ 和 $\mathbf{v}_2$ 线性相关，意味着存在不全为零实数 c_1, c_2 使

$$c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 = \mathbf{0} \tag{3}$$

假设 c_1 不为零，则 $\mathbf{v}_1 = c_2 \mathbf{v}_2 / c_1$．这个推导可逆，所以两个几何矢量线性相关当且仅当它们共线，或者说两个几何矢量线性无关当且仅当它们不共线．

再来看三个矢量的情况．类比两个矢量的情况，则线性相关意味着

$$\mathbf{v}_3 = \frac{c_1}{c_3} \mathbf{v}_1 + \frac{c_2}{c_3} \mathbf{v}_2 \tag{4}$$

由几何矢量加法的（几何）定义，要么这三个矢量都共线，要么 $\mathbf{v}_3$ 落在 $\mathbf{v}_1$ 和 $\mathbf{v}_2$ 所在的平面上．该过程的逆过程也成立，所以三个几何矢量线性相关当且仅当它们都共线或者共面，或者说三个几何矢量线性无关当且仅当它们不共面且两两不共线．

如果一个矢量集合中的矢量是线性相关的，那么这个集合被称为一个**线性相关组**；反之，若线性无关，则称为一个**线性无关组**.

如果一组矢量之间线性相关，那么至少有一个矢量是“冗余”的，也就是说，它可以被其它矢量的线性组合表示出来. 这样一来，对于线性相关的矢量组，如果用它们的线性组合来表示其它矢量，那么表示方式都不是唯一的. 线性无关的矢量组，最重要的性质就是它们的线性组合表达式是唯一的，由此引入了基底和坐标[379] 等概念.

显然，给定一个非零的线性相关组，通过逐个移除这些“冗余”的矢量，我们总可以得到一个线性无关组.

以后我们会看到，若将 M 个 N 维空间的几何矢量的坐标表示为 $N \times M$ 的矩阵，可以用这个矩阵的秩[407] R 来判断其中线性无关矢量的个数. 当且仅当 $R = M \leqslant N$ 时，这 N 个矢量才是线性无关的.

几何矢量的基底和坐标

预备知识 线性相关性[378]

在“几何矢量[370]”中，我们对坐标有了一个简单的认识. 本文中，我们将从基底的角度进一步讨论坐标的定义. 本文同样只讨论几何矢量，简称矢量.

沿一条直线的所有矢量都是共线的，所以在一条直线上最多不超过一个矢量线性无关，所有这些共线的矢量的集合[92] 以及它们的加法和数乘运算组成一个**一维矢量空间**. 注意这里先不需要了解矢量空间的一般定义[444]. 一维矢量空间中的任意矢量都可以通过某个固定的矢量乘以某个数得到. 同理，一个平面上的所有矢量的集合以及它们的加法和数乘运算，组成一个**二维矢量空间**，二维矢量空间中最多能找到两个线性无关的矢量，确定它们以后，它们的线性组合可以得到平面上任意其他矢量（例如直角坐标系的两个单位矢量）. 一般地，我们把最多只包含 N 个线性无关矢量的矢量空间叫做 N 维的.

我们容易想象出 1 到 3 维的几何矢量空间以及它们的两种运算，但却很难想象更高维的情况，我们只能试着用公式定理以及低维的类比来理解它们[5].

[5]例如在狭义相对论的闵可夫斯基空间[1282] 中，我们把时间看成空间的第四个维度，但在画图时，我们往往把空间简化为二维的，这样就可以用三维示意图来表示四维时空.

N 维空间中的任意一组线性无关的 N 个矢量 $\boldsymbol{\beta}_1 \dots \boldsymbol{\beta}_N$ 可以作为一组**基底（basis，复数 bases）**，记为 $\{\boldsymbol{\beta}_i\}$，简称**基**．基是有序的，即使是同一组矢量，如果顺序不同，也要视为不同的基．

如果在这组基底中加入该空间中任意一个矢量 $\mathbf{v}$，这组 $N+1$ 个矢量必定线性相关（否则空间就是 $N+1$ 维的），即存在不全为 0 的实数 $c_1 \dots c_{N+1}$ 使下式成立

$$\sum_{i=1}^{N} c_i \boldsymbol{\beta}_i + c_{N+1}\mathbf{v} = \mathbf{0} \tag{1}$$

所以我们可以把等式两边除以[6] c_{N+1}，将 $\mathbf{v}$ 用 $\{\boldsymbol{\beta}_i\}$ 的线性组合表示．令 $a_i = -c_i/c_{N+1}$，该空间中任意矢量 $\mathbf{v}$ 都有

$$\mathbf{v} = \sum_{i=1}^{N} a_i \boldsymbol{\beta}_i \tag{2}$$

这里的 N 个有序实数 $(a_1, \cdots, a_N)$ 就是 $\mathbf{v}$ 关于基底 $\{\boldsymbol{\beta}_i\}$ 的**坐标（coordinates）**．我们说该坐标属于集合 $\mathbb{R}^N$，即实数集 $\mathbb{R}$ 自身的 N 次笛卡尔积（式 2 [92]）．注意不要把这种表示和行矢量（$1 \times N$ 的矩阵）混淆．

例 1　直角坐标

直角坐标是我们最常用的坐标之一．在三维矢量空间中，首先确定三个互相垂直的单位矢量（通常还要求符合右手定则[20]），这里分别记为 $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$（也有教材记为 $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$）．显然，它们是线性无关的，可以作为基底．要确定从坐标原点到空间中任意一点的矢量 $\mathbf{v}$，我们可以画出一个长方体（图 3 [370]），它沿三个方向的边长 x, y, z 就是矢量 $\mathbf{v}$ 的三个坐标（三个有序实数），满足线性组合

$$\mathbf{v} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \tag{3}$$

例 2　斜坐标系

事实上三维矢量空间中任何三个长度不为零且不共线不共面的矢量都可以作为一组基底．当我们要根据坐标画出矢量 $\mathbf{v}$，就画一个平行六面体（图 1），坐标轴上的三条边长就是 $\mathbf{v}$ 的三个坐标．这样的坐标系叫做斜坐标系．这里同样有类似式 3 的线性组合关系，但每个坐标并不是矢量在对应坐标轴上的（垂直）投影．

[6] 两边除以 c_{N+1} 要求 $c_{N+1} \neq 0$．这可以通过反证法证明：如果式 1 中 $c_{N+1} = 0$，则可得出基底线性相关，不成立，证毕．

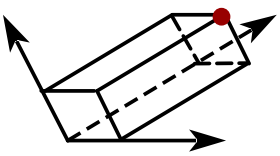


图 1: 斜坐标系

定理 1　坐标的唯一性

有限维空间中，任意给定一个矢量 $\mathbf{v}$ 和一组基底 $\{\boldsymbol{\beta}_i\}$，那么 $\mathbf{v}$ 关于 $\{\boldsymbol{\beta}_i\}$ 的坐标是唯一确定的.

我们可以用反证法证明坐标的唯一性. 假设有两组不全相同的系数都可以使式 2 成立，分别记为 x_i 和 x_i'. 那么分别代入上式再把两式相减得到

$$\sum_{i=1}^{N}(x_i - x_i')\boldsymbol{\beta}_i = \mathbf{0} \tag{4}$$

由于 $(x_i - x_i')$ 不全为零，得基底 $\{\boldsymbol{\beta}_i\}$ 线性相关，而这与基底的定义矛盾. 证毕.

显然，基底不同时，同一个矢量的坐标也不同，因此坐标不能简单地等同于矢量本身. 只有在讨论中固定了基的选择时，才可以把坐标和矢量本身等同.

我们通过一个例子来说明如果用一组线性相关的矢量的线性组合来表示另一个矢量，那么系数是不唯一的.

例 3　线性相关组的表示不唯一

考虑一维几何矢量空间的两个不同几何矢量 $\mathbf{a}$ 和 $\mathbf{b}$，模长分别为 a 和 b. 矢量组 $\{\mathbf{a}, \mathbf{b}\}$ 是线性相关的，因为两个向量可以互相表示：$\mathbf{a} = (a/b)\mathbf{b}$.

对于任意一个矢量 $\mathbf{c}$，模长为 c，如果用 $\mathbf{a}, \mathbf{b}$ 来表示，都有无穷种组合

$$\mathbf{c} = \frac{\lambda}{a}\mathbf{a} + \frac{c-\lambda}{b}\mathbf{b} \tag{5}$$

其中 λ 可以是任意实数.

习题 1　计算坐标

以上所说的坐标不一定是直角坐标系的坐标. 例如平面上两个基底 $\boldsymbol{\beta}_1$ 与 $\boldsymbol{\beta}_2$ 的长度分别为 1 和 2. 夹角为 $\pi/3$，矢量 $\mathbf{v}$ 恰好落在两个基底的角平分线上，长度为 3. 求 $\mathbf{v}$ 的坐标. 答案：$1/\sqrt{3}$，$1/(2\sqrt{3})$.

习题 2　计算坐标

三维几何矢量空间中，建立直角坐标系，基底为 $\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}$．请证明直角坐标（即关于基底 $\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}$ 的坐标）为 $(2,1,1)$，$(1,3,1)$，$(1,1,4)$ 的三个矢量线性无关，并用这三个矢量作为基底，求直角坐标为 $(1,1,1)$ 的矢量关于这组基底的坐标（提示：代入式 3 解方程组）．

矢量运算的坐标形式

除了使用 $\mathbb{R}^N$，我们也常常把坐标记为**列矢量**的形式．列矢量是 N 行 1 列的特殊矩阵[396]．

$$\mathbf{v}=\begin{pmatrix}x\\y\\z\end{pmatrix}_{\{\boldsymbol{\beta}_i\}} \tag{6}$$

由于坐标取决于基底，在列矢量的右下角声明基底是较为严谨的做法，但在不至于混淆的情况下我们可以将其省略，这时我们默认所有列矢量都是用同一组基底．在正文中，为了节约空间，我们将式 6 记为 $\mathbf{v}=(x,y,z)^{\mathrm{T}}_{\{\boldsymbol{\beta}_i\}}$，其中 $^{\mathrm{T}}$ 表示矩阵的转置（见“矩阵[396]”式 2 ）．同样，我们时常省略 $\{\boldsymbol{\beta}_i\}$．

当我们说两个矢量**相等**时，意味着同一基底下两矢量的坐标全都相等．若两矢量在不同基底下的列矢量，则需要先将它们变换到同一基底下再判断是否相等（我们以后再讨论如何进行基底变换）．

当我们确定基底后，之前介绍的加法和数乘[375] 都有对应的坐标运算．由式 2 及矢量加减和数乘的交换律和结合律得

$$\mathbf{v}_1\pm\mathbf{v}_2=\begin{pmatrix}x_1\\y_1\\z_1\end{pmatrix}_{\{\boldsymbol{\beta}_i\}}\pm\begin{pmatrix}x_2\\y_2\\z_2\end{pmatrix}_{\{\boldsymbol{\beta}_i\}}=\begin{pmatrix}x_1\pm x_2\\y_1\pm y_2\\z_1\pm z_2\end{pmatrix}_{\{\boldsymbol{\beta}_i\}} \tag{7}$$

$$\lambda\mathbf{v}=\lambda\begin{pmatrix}x\\y\\z\end{pmatrix}_{\{\boldsymbol{\beta}_i\}}=\begin{pmatrix}\lambda x\\\lambda y\\\lambda z\end{pmatrix}_{\{\boldsymbol{\beta}_i\}} \tag{8}$$

要特别注意，当定义了多组基底时，只有基底相同的两个列矢量按照式 7 相加才有意义．

几何矢量的内积

预备知识 几何矢量的运算[375]

几何定义

我们先来看内积的几何定义．注意该定义不需要任何坐标系的概念．

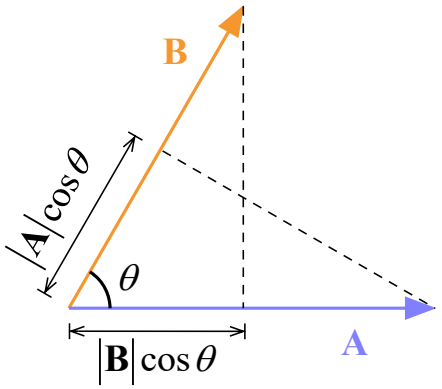


图 1: 内积的几何定义

如图 1，两个几何矢量的**内积（inner product）**[7]就是把它们的模长相乘，再乘以它们的夹角的余弦值．一般用一个实心圆点表示几何矢量的内积（不可省略）：

$$\mathbf{A}\cdot\mathbf{B}=|\mathbf{A}||\mathbf{B}|\cos\theta \tag{1}$$

其中 θ 是两个矢量的夹角．注意两个矢量内积得到的是一个标量．几何定义中（图 1），既可以把内积理解为 $\mathbf{A}$ 投影在 $\mathbf{B}$ 上的模长乘以 $\mathbf{B}$ 的模长，也可以理解为 $\mathbf{B}$ 投影在 $\mathbf{A}$ 上的模长乘以 $\mathbf{A}$ 的模长[8]．可见当两矢量模长不变时，若方向相同，内积取最大值 $|\mathbf{A}||\mathbf{B}|$；若方向相反，内积取最小值 $-|\mathbf{A}||\mathbf{B}|$；若相互垂直，则内积为 0．

我们说两个内积为 0 的矢量互相**垂直（perpendicular）**，或者说**正交（orthogonal）**．几何矢量与自身内积可得该矢量模长的平方．单位矢量与自己的

[7]也叫**点积**或 **点乘**，**标量积（scalar product）**

[8]在这种理解下，若量矢量的夹角为钝角，投影长度取负值

内积等于 1．把一个矢量除以自身模长得到同方向单位矢量的过程叫做矢量的**归一化（normalization）**.

内积的性质

1. **交换律** [9]

$$\mathbf{A}\cdot\mathbf{B}=\mathbf{B}\cdot\mathbf{A} \tag{2}$$

2. **分配律** [10]

$$\mathbf{A}\cdot(\mathbf{B}+\mathbf{C})=\mathbf{A}\cdot\mathbf{B}+\mathbf{A}\cdot\mathbf{C} \tag{3}$$

注意内积不满足结合律，即

$$(\mathbf{A}\cdot\mathbf{B})\mathbf{C}\neq\mathbf{A}(\mathbf{B}\cdot\mathbf{C}) \tag{4}$$

前者是 $\mathbf{C}$ 方向的矢量，后者是 $\mathbf{A}$ 方向的矢量，显然不相等.

内积的坐标运算

若已知 $\mathbf{A},\mathbf{B}$ 在平面直角坐标系 xy 中坐标分别为 (A_x,A_y) 和 (B_x,B_y)，那么如何用坐标表示内积运算的结果呢？先用正交归一基[386] 将两矢量展开

$$\mathbf{A}=A_x\,\hat{\mathbf{x}}+A_y\,\hat{\mathbf{y}}\qquad\mathbf{B}=B_x\,\hat{\mathbf{x}}+B_y\,\hat{\mathbf{y}} \tag{5}$$

所以

$$\mathbf{A}\cdot\mathbf{B}=(A_x\,\hat{\mathbf{x}}+A_y\,\hat{\mathbf{y}})\cdot(B_x\,\hat{\mathbf{x}}+B_y\,\hat{\mathbf{y}}) \tag{6}$$

根据分配律式 3，我们可以把两个括号拆开，变为 4 个内积之和.

$$\mathbf{A}\cdot\mathbf{B}=A_xB_x\,\hat{\mathbf{x}}\cdot\hat{\mathbf{x}}+A_yB_y\,\hat{\mathbf{y}}\cdot\hat{\mathbf{y}}+A_xB_y\,\hat{\mathbf{x}}\cdot\hat{\mathbf{y}}+A_yB_x\,\hat{\mathbf{y}}\cdot\hat{\mathbf{x}} \tag{7}$$

其中 $\hat{\mathbf{x}}\cdot\hat{\mathbf{y}}=\hat{\mathbf{y}}\cdot\hat{\mathbf{x}}=0$（相互垂直），而 $\hat{\mathbf{x}}\cdot\hat{\mathbf{x}}=\hat{\mathbf{y}}\cdot\hat{\mathbf{y}}=1$（相互平行且模长都为 1）．所以最后结果为

$$\mathbf{A}\cdot\mathbf{B}=A_xB_x+A_yB_y \tag{8}$$

[9] 由式 1 易证

[10] 证明见词条最后.

同理，可以在三维直角坐标系 xyz 中把内积结果用坐标表示

$$\mathbf{A}\cdot\mathbf{B}=A_xB_x+A_yB_y+A_zB_z \tag{9}$$

注意内积的代数定义也可以拓展到更高维的情况甚至复数的情况，即对于复数域的 $u_1,u_2,\ldots,u_N$ 和 $v_1,v_2,\ldots,v_N$，

$$\mathbf{u}\cdot\mathbf{v}=\sum_k u_kv_k \tag{10}$$

注意虽然上式中的坐标取决于正交归一基底的选取，但内积的结果却与基底的选取无关. 这是因为内积的几何定义是两个几何矢量间的几何性质，与基底无关.

证明内积的分配律

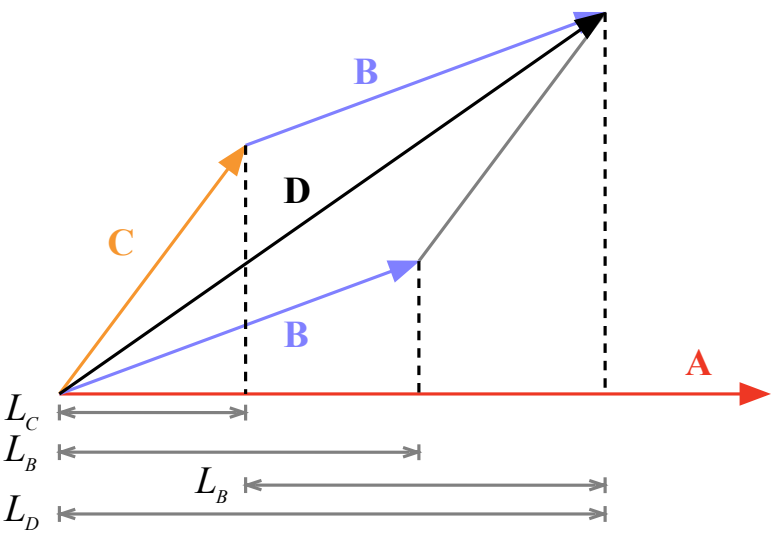


图 2: 内积分配律的证明

如图 2，令 $\mathbf{D}\equiv\mathbf{B}+\mathbf{C}$，把 $\mathbf{A}\cdot\mathbf{B}$，$\mathbf{A}\cdot\mathbf{C}$，$\mathbf{A}\cdot\mathbf{D}$ 分别用几何定义理解为 $\mathbf{B}$，$\mathbf{C}$，$\mathbf{D}$ 在 $\mathbf{A}$ 上的投影乘 $|\mathbf{A}|$，且令投影长度分别为 L_B,L_C,L_D. 那么要证明 $\mathbf{A}\cdot(\mathbf{B}+\mathbf{C})=\mathbf{A}\cdot\mathbf{D}=\mathbf{A}\cdot\mathbf{B}+\mathbf{A}\cdot\mathbf{C}$，只需证明 $L_D=L_B+L_C$ 即可. 现在把 $\mathbf{B}$ 平移使其起点与 $\mathbf{C}$ 的终点对接（投影长度不变）. 从图中立即得出 $L_D=L_B+L_C$.

正交归一基底

预备知识 几何矢量的基底和坐标[379]，几何矢量内积[383]

我们已经知道了矢量基底的概念，如果 N 维空间中一组矢量基底中的每个矢量模长都为 1 且每两个矢量都正交，则我们把这组基底称为**正交归一基（orthonormal bases）**，也叫**单位正交基**. 若把这组正交归一基记为 $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 \dots \hat{\mathbf{x}}_N$，则正交归一可以用内积表示为

$$\hat{\mathbf{x}}_i \cdot \hat{\mathbf{x}}_j = \delta_{ij} \qquad (i, j = 1, \dots, N) \tag{1}$$

其中 δ_{ij} 是克罗内克 delta 函数[107]. 简单来说就是任意两个不同的基底点乘为零，以及任意基底与自身的点成为 1.

单位正交基是某个空间的基底[379]，所以该空间中任意矢量 $\mathbf{v}$ 在单位正交基上的展开

$$\mathbf{v} = \sum_{i=1}^{N} c_i \hat{\mathbf{x}}_i \qquad (i = 1, \dots, N) \tag{2}$$

其中 c_i 就是 $\mathbf{v}$ 的各个坐标. 由正交归一性可以证明（见下文）

$$c_i = \mathbf{v} \cdot \hat{\mathbf{x}}_i \qquad (i = 1, \dots, N) \tag{3}$$

或者把式 2 记为

$$\mathbf{v} = \sum_{i=1}^{N} (\mathbf{v} \cdot \hat{\mathbf{x}}_i) \hat{\mathbf{x}}_i \qquad (i = 1, \dots, N) \tag{4}$$

最常见的例子就是几何矢量在直角坐标系的 $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ 三个单位正交矢量上的展开：

$$\mathbf{v} = (\mathbf{v} \cdot \hat{\mathbf{x}}) \hat{\mathbf{x}} + (\mathbf{v} \cdot \hat{\mathbf{y}}) \hat{\mathbf{y}} + (\mathbf{v} \cdot \hat{\mathbf{z}}) \hat{\mathbf{z}} = x \hat{\mathbf{x}} + y \hat{\mathbf{y}} + z \hat{\mathbf{z}} \tag{5}$$

证明

我们来证明式 3：用 $\hat{\mathbf{x}}_k$ 同时乘以式 2 两边，得

$$\mathbf{v} \cdot \hat{\mathbf{x}}_k = \sum_{i=1}^{N} c_i \hat{\mathbf{x}}_i \cdot \hat{\mathbf{x}}_k = \sum_{i=1}^{N} c_i \delta_{ik} = c_k \qquad (k = 1, \dots, N) \tag{6}$$

最后一步使用了式 2 [107]．证毕．

式 6 的过程可以看作是用“点乘 $\hat{\mathbf{x}}_k$”的操作把式 6 求和中需要的一项筛选出来．注意只有使用正交归一基底才可以进行这样的筛选．一个反例是斜坐标系（例 2 [379]）不能使用式 2 计算坐标，因为它的基底不正交．

矢量叉乘

预备知识 三阶行列式[432]

矢量叉乘在物理定律中十分常见，例如在讨论力学中的力矩[947]，角动量[952]，以及电磁学中的洛伦兹力[1180]，安培力[1186] 时都会使用．以下我们讨论的矢量都是三维空间中的几何矢量[370]，在讨论它们的坐标时，我们默认取正交归一基底．

叉乘的几何定义

两个几何矢量 $\mathbf{A}$，$\mathbf{B}$ 的**叉乘（cross product）**也叫**叉积（cross product）**，**向量积（vector product）**或**矢量积**．叉乘的结果是一个矢量 $\mathbf{C}$．叉乘用“$\times$”表示，且不可省略，即 $\mathbf{A}\times\mathbf{B}=\mathbf{C}$

定义 1　矢量叉乘

要确定一个几何矢量，只需分别确定模长和方向：

1. $\mathbf{C}$ 的模长等于 $\mathbf{A},\mathbf{B}$ 的模长之积与夹角 θ（$0\leqslant\theta\leqslant\pi$）的正弦值相乘．

$$|\mathbf{C}|=|\mathbf{A}||\mathbf{B}|\sin\theta \tag{1}$$

2. $\mathbf{C}$ 的方向垂直于 $\mathbf{A},\mathbf{B}$ 所在的平面，且由右手定则[20] 决定．

与内积和数乘不同，叉乘**不满足交换律**．根据几何定义，$\mathbf{B}\times\mathbf{A}$ 与 $\mathbf{A}\times\mathbf{B}$ 模长相同，方向却相反．表示某个矢量的反方向，就是在前面加负号，所以有

$$\mathbf{B}\times\mathbf{A}=-\mathbf{A}\times\mathbf{B} \tag{2}$$

由式 1 可知，对于两个模场相同的矢量，当它们平行时，叉乘为零矢量；当它们垂直，叉乘的模长等于两模长直接相乘．

等效的叉乘

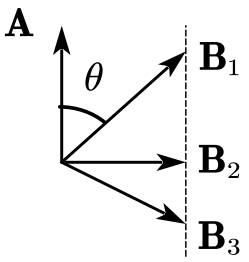


图 1: $\mathbf{A}$ 与 $\mathbf{B}_i$ 相乘结果相同

根据定义 1，当两矢量叉乘时，若把**一个矢量延着与另一个矢量垂直的方向投影，叉乘结果不变**．例如在图 1 中有

$$\mathbf{A}\times\mathbf{B}_1=\mathbf{A}\times\mathbf{B}_3=\mathbf{A}\times\mathbf{B}_2 \tag{3}$$

这是因为 $|\mathbf{B}_i|\sin\theta_i$ 就是投影后的 $\mathbf{B_2}$ 的模长．另一种证明方法：所有指向虚线的 $\mathbf{B}_i$ 都可以表示为 $\mathbf{B}_2+\lambda\mathbf{A}$（$\lambda$ 为任意实数）．根据分配律（见下文）有

$$\mathbf{A}\times(\mathbf{B}_2+\lambda\mathbf{A})=\mathbf{A}\times\mathbf{B}_2+\mathbf{A}\times(\lambda\mathbf{A})=\mathbf{A}\times\mathbf{B}_2 \tag{4}$$

叉乘与数乘的分配律

在 $\mathbf{A}\times\mathbf{B}=\mathbf{C}$ 中，$\mathbf{C}$ 的方向仅由 $\mathbf{A}$ 和 $\mathbf{B}$ 的方向决定．当 $\mathbf{A}$ 和 $\mathbf{B}$ 的方向不变时，$\mathbf{C}$ 的模长正比于 $\mathbf{A}$ 和 $\mathbf{B}$ 的模长之积．假设 λ 为常数（标量），显然有

$$(\lambda\mathbf{A})\times\mathbf{B}=\mathbf{A}\times(\lambda\mathbf{B})=\lambda(\mathbf{A}\times\mathbf{B}) \tag{5}$$

即标量的位置可以任意变换，但矢量与乘号的位置关系始终要保持不变．

叉乘的分配律

叉乘一个最重要的特性，就是它满足分配律．

$$\mathbf{A}\times(\mathbf{B}+\mathbf{C})=\mathbf{A}\times\mathbf{B}+\mathbf{A}\times\mathbf{C} \tag{6}$$

由式 2 及上式可以推出

$$(\mathbf{A}+\mathbf{B})\times\mathbf{C}=-\mathbf{C}\times(\mathbf{A}+\mathbf{B})=-\mathbf{C}\times\mathbf{A}-\mathbf{C}\times\mathbf{B}=\mathbf{A}\times\mathbf{C}+\mathbf{B}\times\mathbf{C} \tag{7}$$

从几何的角度理解，这个结论并不显然（见矢量叉乘分配律的几何证明[391]）.

叉乘的坐标运算

按照上面的定义，在右手系中，三个坐标轴的单位矢量 $\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}$ 满足

$$\hat{\mathbf{x}}\times\hat{\mathbf{y}}=\hat{\mathbf{z}}\qquad\hat{\mathbf{y}}\times\hat{\mathbf{z}}=\hat{\mathbf{x}}\qquad\hat{\mathbf{z}}\times\hat{\mathbf{x}}=\hat{\mathbf{y}}\tag{8}$$

由式 2 可得

$$\hat{\mathbf{y}}\times\hat{\mathbf{x}}=-\hat{\mathbf{z}}\qquad\hat{\mathbf{z}}\times\hat{\mathbf{y}}=-\hat{\mathbf{x}}\qquad\hat{\mathbf{x}}\times\hat{\mathbf{z}}=-\hat{\mathbf{y}}\tag{9}$$

根据定义，一个矢量叉乘自身，模长为 0. 所以叉乘结果是零矢量 $\mathbf{0}$. 于是又有

$$\hat{\mathbf{x}}\times\hat{\mathbf{x}}=\mathbf{0}\qquad\hat{\mathbf{y}}\times\hat{\mathbf{y}}=\mathbf{0}\qquad\hat{\mathbf{z}}\times\hat{\mathbf{z}}=\mathbf{0}\tag{10}$$

式 8，式 9 和式 10 中共 9 条等式描述了 $\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}$ 中任意两个叉乘的结果.

把矢量 $\mathbf{A}$ 和 $\mathbf{B}$ 分别在直角坐标系的三个单位矢量展开，得到

$$\mathbf{A}=a_x\,\hat{\mathbf{x}}+a_y\,\hat{\mathbf{y}}+a_z\,\hat{\mathbf{z}}\qquad\mathbf{B}=b_x\,\hat{\mathbf{x}}+b_y\,\hat{\mathbf{y}}+b_z\,\hat{\mathbf{z}}\tag{11}$$

(a_x,a_y,a_z) 和 (b_x,b_y,b_z) 分别是 $\mathbf{A}$ 和 $\mathbf{B}$ 的坐标. 根据叉乘的分配律（式 6 式 7），可得到如下 9 项

$$\begin{aligned}\mathbf{A}\times\mathbf{B}&=(a_x\,\hat{\mathbf{x}}+a_y\,\hat{\mathbf{y}}+a_z\,\hat{\mathbf{z}})\times(b_x\,\hat{\mathbf{x}}+b_y\,\hat{\mathbf{y}}+b_z\,\hat{\mathbf{z}})\\&=+a_xb_x(\hat{\mathbf{x}}\times\hat{\mathbf{x}})+a_xb_y(\hat{\mathbf{x}}\times\hat{\mathbf{y}})+a_xb_z(\hat{\mathbf{x}}\times\hat{\mathbf{z}})\\&\quad+a_yb_x(\hat{\mathbf{y}}\times\hat{\mathbf{x}})+a_yb_y(\hat{\mathbf{y}}\times\hat{\mathbf{y}})+a_yb_z(\hat{\mathbf{y}}\times\hat{\mathbf{z}})\\&\quad+a_zb_x(\hat{\mathbf{z}}\times\hat{\mathbf{x}})+a_zb_y(\hat{\mathbf{z}}\times\hat{\mathbf{y}})+a_zb_z(\hat{\mathbf{z}}\times\hat{\mathbf{z}})\end{aligned}\tag{12}$$

注意每一项中的运算在式 8 到式 10 中都能找到答案，于是上式化为

$$\mathbf{A}\times\mathbf{B}=(a_yb_z-a_zb_y)\,\hat{\mathbf{x}}+(a_zb_x-a_xb_z)\,\hat{\mathbf{y}}+(a_xb_y-a_yb_x)\,\hat{\mathbf{z}}\tag{13}$$

令 $\mathbf{C}=\mathbf{A}\times\mathbf{B}$，则 $\mathbf{C}$ 的分量表达式为

$$\begin{cases}c_x=a_yb_z-a_zb_y\\c_y=a_zb_x-a_xb_z\\c_z=a_xb_y-a_yb_x\end{cases}\tag{14}$$

式 13 可以用三阶行列式[432] 表示为

$$\mathbf{A}\times\mathbf{B}=\begin{vmatrix}\hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}}\\ a_x & a_y & a_z\\ b_x & b_y & b_z\end{vmatrix} \tag{15}$$

与普通行列式不同的是，这个行列式中第一行的元素是矢量，所以得出的结果也是矢量.

叉乘在物理中的应用如:圆周运动的线速度（式 5 [883]）,洛伦兹力（式 1 [1180]）,安培力（式 1 [1186]）.

例 1　求三角形面积与法向量

空间直角坐标系中三角形的三点分别为 $O(0,0,0)$，$A(1,1,0)$，$B(-1,1,1)$. 求三角形的面积和一个单位法向量.

令 O 到 A 的矢量和 O 到 B 的矢量分别为

$$\begin{aligned}\mathbf{a}&=(1,1,0)-(0,0,0)=(1,1,0)\\ \mathbf{b}&=(-1,1,1)-(0,0,0)=(-1,1,1)\end{aligned} \tag{16}$$

三角形的面积为

$$S=\frac{1}{2}ab\sin\theta \tag{17}$$

其中 θ 是 $\mathbf{a}$ 与 $\mathbf{b}$ 的夹角. 根据式 1，有[11]

$$S=\frac{1}{2}ab\sin\theta=\frac{1}{2}|\mathbf{a}\times\mathbf{b}| \tag{18}$$

令

$$\mathbf{v}=\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}}\\ 1 & 1 & 0\\ -1 & 1 & 1\end{vmatrix}=\hat{\mathbf{x}}-\hat{\mathbf{y}}+2\,\hat{\mathbf{z}} \tag{19}$$

坐标为 $(1,-1,2)$，模长为 $|\mathbf{v}|=\sqrt{1+1+2^2}=\sqrt{6}$，所以面积为 $S=\sqrt{6}/2$.

根据叉乘的几何定义，$\mathbf{v}=(1,-1,2)$ 就是三角形的法向量，进行归一化[12] 得单位法向量为

$$\hat{\mathbf{v}}=\frac{\mathbf{v}}{|\mathbf{v}|}=\frac{(1,-1,2)}{\sqrt{6}}=\left(\frac{\sqrt{6}}{6},-\frac{\sqrt{6}}{6},\frac{\sqrt{6}}{3}\right) \tag{20}$$

[11]可见 $|\mathbf{a}\times\mathbf{b}|$ 是以 $\mathbf{a}$ 和 $\mathbf{b}$ 为边的平行四边形的面积.

[12]把矢量长度变为 1，方向不变

矢量叉乘分配律的几何证明

预备知识 矢量的叉乘[387]

证明 $\mathbf{A}\times(\mathbf{B}+\mathbf{C})=\mathbf{A}\times\mathbf{B}+\mathbf{A}\times\mathbf{C}$

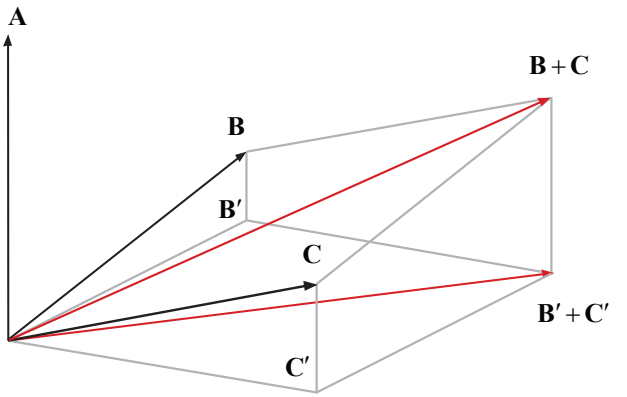


图 1: 把 $\mathbf{B},\mathbf{C},\mathbf{D}$ 投影到与 $\mathbf{A}$ 垂直的平面上

首先令

$$\mathbf{D}=\mathbf{B}+\mathbf{C} \tag{1}$$

把矢量 $\mathbf{B},\mathbf{C},\mathbf{D}$ 在与矢量 $\mathbf{A}$ 垂直的平面上投影，分别得到 $\mathbf{B}',\mathbf{C}',\mathbf{D}'$. 显然，$\mathbf{D}'=\mathbf{B}'+\mathbf{C}'$.

现在先证明

$$\mathbf{A}\times\mathbf{B}=\mathbf{A}\times\mathbf{B}' \tag{2}$$

这是叉乘的一个基本的性质. 首先，根据叉乘的几何定义[387]，$\mathbf{A}\times\mathbf{B}$ 与 $\mathbf{A}\times\mathbf{B}'$ 的方向相同. 另外

$$|\mathbf{A}\times\mathbf{B}|=|\mathbf{A}||\mathbf{B}|\sin\theta_{AB}=|\mathbf{A}||\mathbf{B}'|=|\mathbf{A}\times\mathbf{B}'| \tag{3}$$

所以二者模长也相等，证毕.

同理有

$$\mathbf{A}\times\mathbf{C}=\mathbf{A}\times\mathbf{C}' \tag{4}$$

$$\mathbf{A}\times\mathbf{D}=\mathbf{A}\times\mathbf{D}' \tag{5}$$

所以，要证明

$$\mathbf{A}\times\mathbf{D}=\mathbf{A}\times\mathbf{B}+\mathbf{A}\times\mathbf{C} \tag{6}$$

只需要证明

$$\mathbf{A}\times\mathbf{D}'=\mathbf{A}\times\mathbf{B}'+\mathbf{A}\times\mathbf{C}' \tag{7}$$

即可.

由于 $\mathbf{B}',\mathbf{C}',\mathbf{D}'$ 都与 $\mathbf{A}$ 垂直，所以 $\mathbf{A}$ 与之叉乘的效果相当于 $\mathbf{B}',\mathbf{C}',\mathbf{D}'$ 的模长分别乘以 $|\mathbf{A}|$，且绕 $\mathbf{A}$ 逆时针分别旋转 $90°$. 所以上式就是在说，若已知 $\mathbf{B}'+\mathbf{C}'=\mathbf{D}'$，那么把它们分别乘以常数并旋转 $90°$ 后这个加法仍然成立. 这是显然的. 证毕.

连续叉乘的化简

预备知识 矢量的叉乘[387]

连续两个叉乘的化简也叫 BAC-CAB 定理

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=\mathbf{B}(\mathbf{A}\cdot\mathbf{C})-\mathbf{C}(\mathbf{A}\cdot\mathbf{B}) \tag{1}$$

$$(\mathbf{B}\times\mathbf{C})\times\mathbf{A}=\mathbf{C}(\mathbf{A}\cdot\mathbf{B})-\mathbf{B}(\mathbf{A}\cdot\mathbf{C}) \tag{2}$$

要证明这个定理可以将每个叉乘在各个基底上展开（式 13 [387]）.

习题 1

由叉乘的坐标定义（式 13 [387]）证明式 1.

这里对连续叉乘的几何意义略作说明，可以用于理解该公式的结构. 以式 1 为例，根据叉乘的几何意义我们知道 $\mathbf{B}\times\mathbf{C}$（命名为 $\mathbf{D}$）方向垂直于 $\mathbf{B}$ 和 $\mathbf{C}$ 所在平面. 又因为 $\mathbf{A}\times\mathbf{D}$ 垂直于 $\mathbf{D}$，所以最终的矢量再次落到 $\mathbf{B}$ 和 $\mathbf{C}$ 所在平面上，所以等式右边是 $\mathbf{B}$ 和 $\mathbf{C}$ 的线性组合.

下面来介绍一种简单的记忆方法，括号外的矢量在哪边，括号内靠近那边的矢量所在的项前面就是正号，另一项前面则是负号，如图 1 所示.

三矢量的混合积

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=\mathbf{B}(\mathbf{A}\cdot\mathbf{C})-\mathbf{C}(\mathbf{A}\cdot\mathbf{B})$$

$$(\mathbf{B}\times\mathbf{C})\times\mathbf{A}=\mathbf{C}(\mathbf{A}\cdot\mathbf{B})-\mathbf{B}(\mathbf{A}\cdot C)$$

图 1: 三矢量叉乘的化简

预备知识 矢量的叉乘[387]

我们定义以下运算

$$(\mathbf{A}\times\mathbf{B})\cdot\mathbf{C} \tag{1}$$

为矢量 $\mathbf{A},\mathbf{B},\mathbf{C}$ 的**混合积**，即 $\mathbf{A}$ 和 $\mathbf{B}$ 先做叉乘，然后再把得到的矢量与 $\mathbf{C}$ 做内积（点乘）[471]．注意即使我们把括号去掉这个表达式也没有歧义，因为叉乘只能在矢量之间进行，而点乘的结果是标量．混合积满足

$$\mathbf{A}\times\mathbf{B}\cdot\mathbf{C}=\mathbf{B}\times\mathbf{C}\cdot\mathbf{A}=\mathbf{C}\times\mathbf{A}\cdot\mathbf{B} \tag{2}$$

这个公式可由图 1 记忆．图中箭头的方向由叉乘的方向（顺时针或逆时针）决

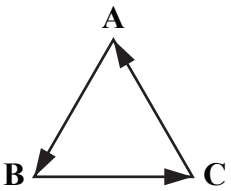


图 1: 式 2 记忆法

定，与内积无关，即 $\mathbf{A}\times\mathbf{B}\cdot\mathbf{C}=\mathbf{C}\cdot(\mathbf{A}\times\mathbf{B})$．如果混合积的顺序取与箭头相反的方向，根据叉乘的性质，需要在前面加上负号（叉乘不满足乘法交换律）．式 2 与式 3 互为相反数

$$\mathbf{C}\times\mathbf{B}\cdot\mathbf{A}=\mathbf{B}\times\mathbf{A}\cdot\mathbf{C}=\mathbf{A}\times\mathbf{C}\cdot\mathbf{B} \tag{3}$$

注意即使将混合积省略括号记为 $\mathbf{A}\times\mathbf{B}\cdot\mathbf{C}$ 或者 $\mathbf{C}\cdot\mathbf{A}\times\mathbf{B}$ 也应该理解为先叉乘后内积．$\mathbf{A}\times(\mathbf{B}\cdot\mathbf{C})$ 没有定义，因为矢量不能叉乘标量．

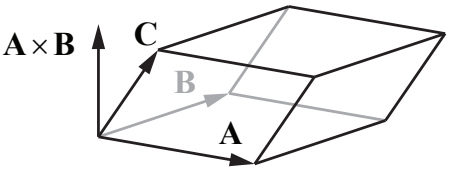


图 2: 矢量混合积的几何意义

几何法证明

如图 2，以三个矢量为棱作平行六面体. 由例 1 [387] 可知 $|\mathbf{A}\times\mathbf{B}|$ 就是 $\mathbf{A},\mathbf{B}$ 所在平行四边形的面积. 令 $\mathbf{A}\times\mathbf{B}=|\mathbf{A}\times\mathbf{B}|\hat{\mathbf{n}}$，则 $\hat{\mathbf{n}}$ 为平面的法向量，平行六面体的高为 $|\hat{\mathbf{n}}\cdot\mathbf{C}|$，所以平行六面体的体积等于底面积乘以高

$$V=|\mathbf{A}\times\mathbf{B}||\hat{\mathbf{n}}\cdot\mathbf{C}|=|\mathbf{A}\times\mathbf{B}\cdot\mathbf{C}| \tag{4}$$

同理可得对于同一平行六面体

$$V=|\mathbf{B}\times\mathbf{C}\cdot\mathbf{A}|=|\mathbf{C}\times\mathbf{A}\cdot\mathbf{B}| \tag{5}$$

这里只证明了式 2 的绝对值，要证明正负号，定义 $\hat{\mathbf{n}}\cdot\mathbf{C}<0$ 时 V 为负值即可.

代数法证明

预备知识 行列式[432]

不难证明三矢矢积若展开成分量的形式，等于三个矢量组成的行列式

$$\mathbf{A}\times\mathbf{B}\cdot\mathbf{C}=\begin{vmatrix}A_x & A_y & A_z\\ B_x & B_y & B_z\\ C_x & C_y & C_z\end{vmatrix} \tag{6}$$

而利用行列式中任意两行置换符号改变，即可证明式 2.

第二章

矩阵和行列式

矩阵

本书中**矩阵**符号用加粗的正体字母来表示，而对应的**矩阵元**一般用斜体加**行标**和**列标**表示．例如矩阵 $\mathbf{A}$ 的第 i 行第 j 列的的**矩阵元**表示为 A_{ij}．特殊地，行数等于列数的矩阵叫做**方阵**．只有一行的矩阵和只有一列的矩阵分别叫做**行矢量**和**列矢量**．

矩阵的对角线

我们先定义矩阵的**对角线（diagonal）**是行标等于列标的所有矩阵元（图 1），称为对角元（diagonal element），左上角的对角元是第一个．对角线的概念常在讨论方阵时使用，但严格来说也适用于长方形矩阵．方阵从右上角到左下角的所有矩阵元则称为**反对角线（anti-diagonal）**．另外，把对角线向上或向下平移一个元素也称为**副对角线（subdiagonal）**，所以对角线也称为**主对角线（main diagonal）**．

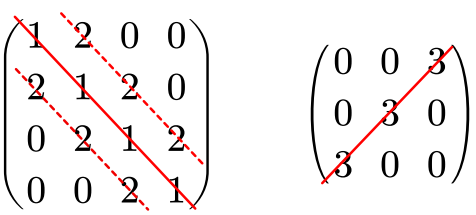


图 1: 左：方阵的对角线（实线）和副对角线（虚线），右：方阵的反对角线

矩阵的转置

则任意矩阵 $\mathbf{A}$ 的**转置（transpose）**记为 $\mathbf{A}^{\mathrm{T}}$．转置操作把 $\mathbf{A}$ 的第 i 行变为 $\mathbf{A}^{\mathrm{T}}$ 的第 i 列，相当于把矩阵沿对角线翻转．即任意矩阵元满足

$$A_{ij}^{\mathrm{T}} = A_{ji} \tag{1}$$

注意转置操作不影响对角线上的矩阵元．另外行矢量转置后变为列矢量，反之亦然．

$$(x_1, x_2 \ldots, x_n)^{\mathrm{T}} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{2}$$

为了排版方便，本书在正文中通常用 $(x_1, x_2 \ldots, x_n)^{\mathrm{T}}$ 表示列矢量．

矩阵的乘法

矩阵最常见的运算是矩阵的乘法．从 $(x_1, \ldots, x_n)$ 到 $(y_1, \ldots, y_m)$ 的线性映射[451]

$$\begin{cases} y_1 = A_{11}x_1 + A_{12}x_2 + \ldots + A_{1n}x_n \\ y_2 = A_{21}x_1 + A_{22}x_2 + \ldots + A_{2n}x_n \\ \quad \vdots \\ y_m = A_{m1}x_1 + A_{m2}x_2 + \ldots + A_{mn}x_n \end{cases} \tag{3}$$

可用矩阵与列矢量的乘法表示为（有时候也会用方括号表示矩阵和矢量）

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1n} \\ A_{21} & A_{22} & \ldots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \ldots & A_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{4}$$

令列矢量 $\mathbf{x} = (x_1, x_2 \ldots, x_n)^{\mathrm{T}}$，$\mathbf{y} = (y_1, y_2 \ldots, y_n)^{\mathrm{T}}$，系数矩阵为 $\mathbf{A}$，上式可记为

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{5}$$

注意 $\mathbf{A}$ 的列数必须和 $\mathbf{x}$ 的行数相等．由此可以定义**矩阵乘以列矢量**的运算规则：$m \times n$ 矩阵乘以 $n \times 1$ 列矢量会得到 $m \times 1$ 的列矢量．要计算 y_i，就用 $m \times n$ 矩阵的第 i 行的 n 个数和 $x_1 \ldots x_n$ 分别相乘再相加（即内积[383]）

$$y_i = \sum_{j=1}^{n} A_{ij}x_j \tag{6}$$

若有 l 个不同的 $\mathbf{x}$ 和 $\mathbf{y}$，第 k 个记为 $\mathbf{x}_k=(x_{1k},\dots,x_{nk})^{\mathrm{T}}$ 和 $\mathbf{y}_k=(y_{1k},\dots,y_{mk})^{\mathrm{T}}$，对应的变换为

$$\begin{pmatrix} y_{1k} \\ y_{2k} \\ \vdots \\ y_{mk} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{pmatrix} \begin{pmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{nk} \end{pmatrix} \tag{7}$$

可以将所有的 $\mathbf{x}_k$ 和 $\mathbf{y}_k$ 分别横向拼成 $n\times l$ 和 $m\times l$ 的矩阵

$$\mathbf{X}=\begin{pmatrix} x_{11} & \cdots & x_{1l} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nl} \end{pmatrix} \qquad \mathbf{Y}=\begin{pmatrix} y_{11} & \cdots & y_{1l} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{ml} \end{pmatrix} \tag{8}$$

现在把 l 组线性变换用一条式子表示为

$$\begin{pmatrix} y_{11} & \cdots & y_{1l} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{ml} \end{pmatrix} = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} x_{11} & \cdots & x_{1l} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nl} \end{pmatrix} \tag{9}$$

由此，可以定义一般的矩阵乘法：$m\times n$ 的矩阵 $\mathbf{A}$ 和 $n\times l$ 的矩阵 $\mathbf{X}$ 相乘得到 $m\times l$ 的矩阵 $\mathbf{Y}$，Y_{ij} 等于 $\mathbf{A}$ 的第 i 行和 $\mathbf{X}$ 的第 j 列内积.

$$\mathbf{Y}=\mathbf{A}\mathbf{X} \tag{10}$$

矩阵元公式为

$$Y_{ij}=\sum_{k=1}^{n} A_{ik}X_{kj} \tag{11}$$

再次注意两个相乘的矩阵，左边矩阵的列数必须等于右边矩阵的行数. 我们可以用图 2 来记忆矩阵乘法.

矩阵乘法一般不满足交换律，举一个反例：

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}=\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}\neq\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}=\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \tag{12}$$

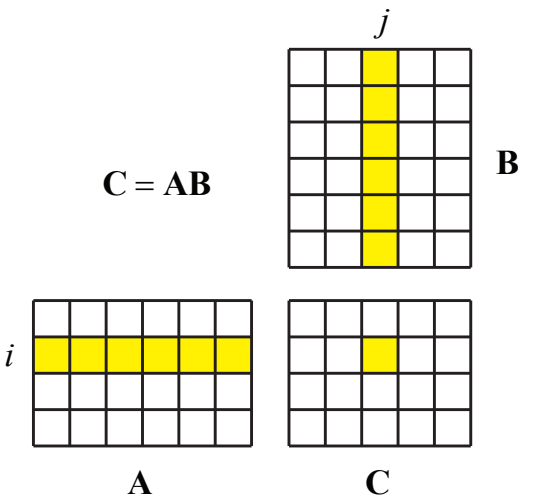


图 2: 矩阵乘法的示意图：$\mathbf{C}$ 的 (i,j) 矩阵元等于 $\mathbf{A}$ 的第 i 行和 $\mathbf{B}$ 的第 j 列逐个元素相乘再相加

矩阵的乘法分配律

矩阵的乘法满足分配律

$$\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC} \tag{13}$$

$$(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC} \tag{14}$$

令式 13 左边等于矩阵 $\mathbf{D}$，则其矩阵元为

$$D_{ij} = \sum_k A_{ik}(B_{kj}+C_{kj}) \tag{15}$$

拆括号得

$$D_{ij} = \sum_k A_{ik}B_{kj} + \sum_k A_{ik}C_{kj} \tag{16}$$

而这恰好是 $\mathbf{AB}+\mathbf{AC}$ 的矩阵元．证毕．式 14 的证明类似．

现在我们可以得出线性变换（式 5）的一个重要性质．对若干列矢量 $\mathbf{v}_1, \mathbf{v}_2, \ldots$

$$\mathbf{A}\left(\sum_i c_i\mathbf{v}_i\right) = \sum_i c_i\mathbf{A}\mathbf{v}_i \tag{17}$$

也就是说若干列矢量的线性组合的线性变换等于每个列矢量分别进行线性变换再进行同样的线性组合．

矩阵乘法的结合律

我们希望证明三个矩阵的乘法满足结合律

$$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} \tag{18}$$

令

$$\mathbf{D} = \mathbf{A}(\mathbf{BC}) \tag{19}$$

$\mathbf{D}$ 的矩阵元为

$$D_{ij} = \sum_l A_{il}(BC)_{lj} = \sum_l A_{il}\left(\sum_k B_{lk}C_{kj}\right) \tag{20}$$

拆括号，得

$$D_{ij} = \sum_k \sum_l (A_{il}B_{lk}C_{kj}) \tag{21}$$

对 C_{kj} 进行合并同类项，得

$$D_{ij} = \sum_k \left(\sum_l A_{il}B_{lk}\right)C_{kj} \tag{22}$$

括号中恰好是 $\mathbf{A}$ 乘以 $\mathbf{B}$ 所得矩阵的矩阵元 $(AB)_{ik}$ 所以

$$D_{ij} = \sum_k (AB)_{ik}C_{kj} \tag{23}$$

即

$$\mathbf{D} = (\mathbf{AB})\mathbf{C} \tag{24}$$

证毕.

单位矩阵

单位矩阵（unit matrix）是对角线上的元素全为 1，非对角线上的元素全为 0 的方阵．通常记为通常记为 $\mathbf{I}$．为了强调矩阵的维数 N，也可记为 $\mathbf{I}_N$．单位矩阵的矩阵元可用克罗内克 δ 函数（式 6 [386]）表示为

$$I_{ij} = \delta_{ij} \tag{25}$$

任何矩阵左乘或右乘单位矩阵，仍然得到矩阵本身．单位矩阵的转置仍为单位矩阵.

对角矩阵

只有对角线不为零的方阵叫做**对角矩阵（diagonal matrix）**，只有反对角线不为零的矩阵叫做**反对角矩阵（anti-diagonal matrix）**．对角矩阵和反对角矩阵通常是指方阵．

根据上述矩阵乘法的定义，对角矩阵 $\mathbf{\Lambda}$ 左乘任意矩阵 $\mathbf{A}$ （即 $\mathbf{\Lambda A}$）相当于把 $\mathbf{A}$ 阵的第 i 行乘以 $\mathbf{\Lambda}$ 的第 i 个对角元；右乘（即 $\mathbf{A\Lambda}$）相当于把第 j 列乘以第 j 个对角元．

逆矩阵

详见逆矩阵[405]．记方阵 $\mathbf{M}$ 的**逆矩阵**为 $\mathbf{M}^{-1}$，且满足

$$\mathbf{M}^{-1}\mathbf{M}=\mathbf{I} \tag{26}$$

其中 $\mathbf{I}$ 是单位矩阵．也就是说，任意一个矩阵（或列矢量）$\mathbf{A}$ 乘以矩阵 $\mathbf{M}$ 再乘以其逆矩阵 $\mathbf{M}^{-1}$ 仍然得到 $\mathbf{A}$ 本身．

虽然矩阵乘法一般不满足交换律，但矩阵和对应的逆矩阵满足[1]，即

$$\mathbf{M}\mathbf{M}^{-1}=\mathbf{M}^{-1}\mathbf{M}=\mathbf{I} \tag{27}$$

逆矩阵 $\mathbf{M}^{-1}$ 所代表的线性变换就是 $\mathbf{M}$ 代表的线性变换的逆变换，令 $\mathbf{x}$ 和 $\mathbf{y}$ 为列矢量，如果有

$$\mathbf{y}=\mathbf{M}\mathbf{x} \tag{28}$$

那么我们在等式两边左乘 $\mathbf{M}^{-1}$ 再把等式左右互换，则上式变为

$$\mathbf{x}=\mathbf{M}^{-1}\mathbf{y} \tag{29}$$

要求逆矩阵，一种简单直接但低效的方法就是先令 $\mathbf{y}=(1\ 0\ \ldots)^{\mathrm{T}}$，代入式 3 [405] 解线性方程组得 $\mathbf{x}$，将 $\mathbf{x},\mathbf{y}$ 代入式 4 可知 $\mathbf{x}$ 就是 $\mathbf{M}$ 的第一列，再令 $\mathbf{y}=(0\ 1\ 0\ \ldots)^{\mathrm{T}}$，解线性方程组可得 $\mathbf{M}$ 的第二列，以此类推就可以得到完整的 $\mathbf{M}$.

[1] 这个定理暂时不证

平面旋转矩阵

预备知识 矩阵[396]

我们之前学过的“平面旋转变换[525]”属于线性变换[451]，以下用矩阵 $\mathbf{R}_2$ 表示．虽然我们可以直接把变换写成矩阵乘以列矢量的形式，但这里我们用另一种方法推导一次，能更好地帮助理解和记忆．

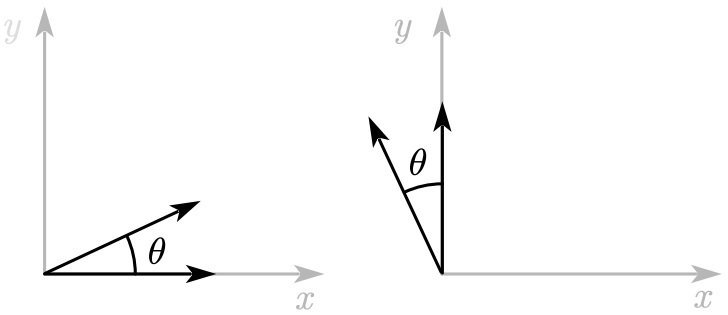


图 1: 把单位矢量 $\hat{\mathbf{x}}$ 和 $\hat{\mathbf{y}}$ 逆时针旋转 θ

已知单位矢量 $\hat{\mathbf{x}}=(1,0)^{\mathrm{T}},\hat{\mathbf{y}}=(0,1)^{\mathrm{T}}$ 逆时针旋转 θ 得（图 1 ）

$$\mathbf{R}_2\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}\cos\theta\\\sin\theta\end{pmatrix}\qquad\mathbf{R}_2\begin{pmatrix}0\\1\end{pmatrix}=\begin{pmatrix}-\sin\theta\\\cos\theta\end{pmatrix}\tag{1}$$

要求任意矢量 $\mathbf{v}=(v_1,v_2)^{\mathrm{T}}$ 的旋转矩阵，可以将 $\mathbf{v}$ 表示成 $\hat{\mathbf{x}}$ 和 $\hat{\mathbf{y}}$ 的线性组合（式 6 [375]）$\mathbf{v}=v_1\hat{\mathbf{x}}+v_2\hat{\mathbf{y}}$．由式 17 [396]，该线性组合的旋转变换等于 $\hat{\mathbf{x}},\hat{\mathbf{y}}$ 分别做旋转变换再做同样的线性组合，即

$$\begin{aligned}\mathbf{R}_2(v_1\hat{\mathbf{x}}+v_2\hat{\mathbf{y}})&=v_1\mathbf{R}_2\hat{\mathbf{x}}+v_2\mathbf{R}_2\hat{\mathbf{y}}\\&=v_1\begin{pmatrix}\cos\theta\\\sin\theta\end{pmatrix}+v_2\begin{pmatrix}-\sin\theta\\\cos\theta\end{pmatrix}\\&=\begin{pmatrix}\cos\theta&-\sin\theta\\\sin\theta&\cos\theta\end{pmatrix}\begin{pmatrix}v_1\\v_2\end{pmatrix}\end{aligned}\tag{2}$$

所以旋转矩阵为

$$\mathbf{R}_2=\begin{pmatrix}\cos\theta&-\sin\theta\\\sin\theta&\cos\theta\end{pmatrix}\tag{3}$$

这与平面旋转变换[525] 得出的结果一致.

把这个推导推广到一般情况，就是如果已知每个基底 $\boldsymbol{\beta}_i$ 的线性变换（记变换矩阵为 $\mathbf{A}$）结果为 $\boldsymbol{\alpha}_i = \mathbf{A}\boldsymbol{\beta}_i$，那么变换矩阵的第 i 列就是第 i 个列矢量 $\boldsymbol{\alpha}_i$.

绕任意点旋转

由式 5 [525] 可得绕任一点 (x_0, y_0) 旋转为

$$\begin{aligned}\begin{pmatrix} x' \\ y' \end{pmatrix} &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1-\cos\theta & -\sin\theta \\ \sin\theta & 1-\cos\theta \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}\end{aligned} \tag{4}$$

可见绕任意点旋转并不能简单表示为单个矩阵和 $(x, y)^{\mathrm{T}}$ 相乘，所以不是线性变换. 所以习惯上讨论旋转变换时都是默认关于原点旋转.

主动和被动理解

我们将任意矢量 $\mathbf{v}$ 的旋转变换记为 $\mathbf{u} = \mathbf{R}_2\mathbf{v}$，我们把 $\mathbf{v}$ 看做是二维空间中某矢量关于基底 $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ 的坐标. 若我们把 $\mathbf{u}$ 看做是另一矢量关于 $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ 的坐标，那么 $\hat{\mathbf{u}}$ 就等于 $\hat{\mathbf{v}}$ 逆时针旋转 θ 角. 旋转矩阵的这种理解被称为**主动**的.

另一种可能的理解是，$\mathbf{u}$ 和 $\mathbf{v}$ 代表二维空间中的同一矢量关于不同基底的展开. 我们把 $\mathbf{u}$ 使用的基底记为 $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2$，$\mathbf{v}$ 使用的基底记为 $\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2$. 我们有

$$u_1\hat{\mathbf{u}}_1 + u_2\hat{\mathbf{u}}_2 = v_1\hat{\mathbf{v}}_1 + v_2\hat{\mathbf{v}}_2 \tag{5}$$

将 $\mathbf{R}_2$ 的矩阵元记为 R_{ij}，不难证明两组基底之间的关系为

$$\begin{cases} \hat{\mathbf{v}}_1 = R_{11}\hat{\mathbf{u}}_1 + R_{21}\hat{\mathbf{u}}_2 \\ \hat{\mathbf{v}}_2 = R_{12}\hat{\mathbf{u}}_1 + R_{22}\hat{\mathbf{u}}_2 \end{cases} \tag{6}$$

将矩阵元代入可知，基底 $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2$ 分别是基底 $\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2$ 顺时针旋转 θ 角所得. 我们把这种理解叫做**被动**的，即旋转矩阵表示同一个矢量的**基底变换**.

逆矩阵

我们既可以使用“平面旋转变换[525]”中求逆变换的方法把 θ 变为 $-\theta$ 再化简求出 $\mathbf{R}_2$ 的逆矩阵，也可以通过解方程求逆矩阵（式 28 [396]）. 但最方便的是，由于 $\mathbf{R}_2$ 是一个单位正交阵，我们只需要把矩阵转置即可得到逆矩阵.

$$\mathbf{R}_2^{-1} = \mathbf{R}_2^{\mathrm{T}} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{7}$$

三维旋转矩阵

预备知识 平面旋转矩阵[402]，自由度[930]

类比平面旋转矩阵[402]，空间旋转矩阵是三维坐标的旋转变换，所以应该是 3×3 的方阵. 不同的是平面旋转变换只有一个自由度 θ，而空间旋转变换除了转过的角度还需要考虑转轴的方向，三维空间中的方向有两个自由度，所有三维旋转矩阵共有 3 个自由度.

若已经知道空间直角坐标系中三个单位正交矢量

$$\hat{\mathbf{x}} = (1,0,0)^{\mathrm{T}} \quad \hat{\mathbf{y}} = (0,1,0)^{\mathrm{T}} \quad \hat{\mathbf{z}} = (0,0,1)^{\mathrm{T}} \tag{1}$$

经过三维旋转矩阵变换以后变为另外三个正交归一矢量，令坐标变为

$$(a_{11},a_{21},a_{31})^{\mathrm{T}} \quad (a_{12},a_{22},a_{32})^{\mathrm{T}} \quad (a_{13},a_{23},a_{33})^{\mathrm{T}} \tag{2}$$

类比平面旋转矩阵[402]，可以得到旋转矩阵为

$$\mathbf{R}_3 = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{3}$$

这 9 个矩阵元只有 3 个是独立的，因为我们有 6 个条件：每个列矢量模长等于 1（3 个等式），且两两间正交（3 个等式）.

除了通过三个单位矢量构建旋转矩阵，我们可以通过由转轴的方向和旋转的角度来计算每个矩阵元，参考“绕轴旋转矩阵[416]”和“四元数[419]”.

逆矩阵

如果我们把式 2 中的三个正交归一基底记为

逆矩阵

- 本词条存在未完成的内容.

预备知识 矩阵[396]

逆矩阵

[2]如果两个方阵的乘积是单位矩阵，

$$\mathbf{BA} = \mathbf{I} \tag{1}$$

那么 $\mathbf{B}$ 就叫做 $\mathbf{A}$ 的**逆矩阵**，或者记为 $\mathbf{A}^{-1}$. 从矢量空间的角度来看，矩阵 $\mathbf{A}$ 把第一个矢量空间映射[98] 到第二个空间，而 $\mathbf{A}^{-1}$ 就代表改映射的逆映射. 这里的一个隐含要求是 $\mathbf{A}$ 必须代表一一映射，如果是多对一的映射，将不存在反映射. 一一映射要求 $\mathbf{A}$ 的每一列都线性无关，所以 $\mathbf{A}$ 的行数必须大于等于列数. 另外显然一个矩阵最多只能有一个逆矩阵.

我们讨论逆矩阵时一般假设 $\mathbf{A}$ 和 $\mathbf{A}^{-1}$ 都是方阵[3]，所以 $\mathbf{A}$ 是一个 N 维方阵，代表两个 N 维矢量空间的一一映射，即每个空间中任何一个矢量在另一个空间中都能找到唯一对应的矢量. 同理，$\mathbf{A}^{-1}$ 的性质也相同.

从以上的分析中易证，如果 $\mathbf{B}$ 是 $\mathbf{A}$ 的逆矩阵，那么 $\mathbf{A}$ 必定也是 $\mathbf{B}$ 的逆矩阵，我们说 $\mathbf{A}$ 和 $\mathbf{B}$ **互逆**，即

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I} \tag{2}$$

[2]参考 Wikipedia 相关页面.

[3]以后如无特殊说明，我们都使用这个规定.

求逆矩阵

令 $\mathbf{x}$ 和 $\mathbf{y}$ 为 N 维列矢量，如果有

$$\mathbf{A}\mathbf{x}=\mathbf{y} \tag{3}$$

那么我们在等式两边左乘 $\mathbf{A}^{-1}$ 得

$$\mathbf{x}=\mathbf{A}^{-1}\mathbf{y} \tag{4}$$

要求逆矩阵，一种直接的方法就是先令 $\mathbf{y}=(1,0,\dots)^{\mathrm{T}}$，代入式 3 解线性方程组得 $\mathbf{x}$，将 $\mathbf{x},\mathbf{y}$ 代入式 4 可知 x 就是 $\mathbf{A}^{-1}$ 的第一列，再令 $\mathbf{y}=(0,1,0,\dots)^{\mathrm{T}}$，解线性方程组可得 $\mathbf{A}$ 的第二列，以此类推.

在一些实际问题中，我们需要大量求解式 3 形式的方程组，每个方程组都有同样的 $\mathbf{A}$ 和不同的 y，一个高效的做法是先求出 $\mathbf{A}^{-1}$，这样解每个方程组就只需要做一次矩阵乘法（式 4）即可.

（未完成）法 2：将 $\mathbf{A}$ 和 $\mathbf{I}$ 做同样的行变换，使 $\mathbf{A}$ 变换后称为单位矩阵. 证明：行变换相当于左乘一个矩阵 $\mathbf{B}$. $\mathbf{BA}=\mathbf{I}$，$\mathbf{BI}=\mathbf{B}$

（未完成）法 3：伴随矩阵（adjugate matrix）除以行列式，参考同济线代.

叉乘的矩阵形式

预备知识 矢量叉乘[387]

对任意矢量 $\mathbf{a}$ 和 $\mathbf{b}$，令

$$\mathbf{c}=\mathbf{a}\times\mathbf{b} \tag{1}$$

该运算可以看作列矢量 $\mathbf{b}$ 到列矢量 $\mathbf{c}$ 的线性变换. 我们知道线性变换可以用矩阵表示，所以必存在矩阵 $\mathbf{A}$，满足

$$\mathbf{c}=\mathbf{A}\mathbf{b} \tag{2}$$

令 $\mathbf{a}$ 的坐标为 (a_x,a_y,a_z)，根据叉乘的分量表达式（式 14 [387]），易得变换矩阵为

$$\mathbf{A}=\begin{pmatrix}0 & -a_z & a_y\\ a_z & 0 & -a_x\\ -a_y & a_x & 0\end{pmatrix} \tag{3}$$

这是一个**反对称矩阵**，即 $A_{ij} = -A_{ji}$.

同理，式 1 也可以看作是 $\mathbf{a}$ 到 $\mathbf{c}$ 的线性变换

$$\mathbf{c} = \mathbf{B}\mathbf{a} \tag{4}$$

其中

$$\mathbf{B} = \begin{pmatrix} 0 & b_z & -b_y \\ -b_z & 0 & b_x \\ b_y & -b_x & 0 \end{pmatrix} \tag{5}$$

这恰好与式 3 符号相反.

矩阵的秩

预备知识 行列式[432]

我们定义矩阵[396] 的**列秩（column rank）**等于其线性无关[378] 的列数，**行秩（row rank）**等于线性无关的行数. 以后我们会证明任意的矩阵的行秩和列秩是相同的[408]，所以都可以直接叫做矩阵的**秩（rank）**.

例 1

矩阵

$$\begin{pmatrix} 1 & 1 & 4 \\ 1 & 0 & 2 \\ 0 & 1 & 2 \end{pmatrix} \tag{1}$$

中，如果我们把矩阵看做是三个列矢量 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 组成，那么 $\mathbf{v}_1, \mathbf{v}_2$ 显然是线性无关的（不共线），而 $\mathbf{v}_3$ 可以表示为 $\mathbf{v}_1, \mathbf{v}_2$ 的线性组合，即

$$\mathbf{v}_3 = 2\mathbf{v}_2 + 2\mathbf{v}_2 \tag{2}$$

所以它们之中只有两个矢量线性无关，该矩阵的秩为 2.

当然，我们也可以认为 $\mathbf{v}_2, \mathbf{v}_3$ 线性无关，排除 $\mathbf{v}_1$，同样得到秩为 2. 一般来说，若有（式 1 [378]）

$$\sum_i c_i \mathbf{v}_i = \mathbf{0} \qquad (c_i \neq 0) \tag{3}$$

我们就可以把任意一个 $\mathbf{v}_i$ 排除，再次求解上式，直到上式无解，那么可以确定剩下的矢量就是线性无关的，他们的数量就是矩阵的秩．这种方法计算量过大，下文我们会介绍更简单的方法．

根据定义，一个矩阵的秩 R 必定小于或等于矩阵的行数 M 以及列数 N（取较小者）．对于方阵，若三者相等，即 $M = N = R$，我们就称其为**满秩矩阵（full rank matrix）**．满秩矩阵意味着矩阵中所有行（列）都线性无关．

判断一个方阵是否为满秩矩阵的一种常见方法是计算方阵的行列式[432]，若结果不为零，则矩阵是满秩的，否则不是（定理 7 [435]）．注意非满秩的情况下行列式并不能判断秩具体是多少．

高斯消元法计算秩

预备知识 高斯消元法[477]

要确定任意矩阵秩的大小，我们可以先用高斯消元法将矩阵变换为梯形矩阵．矩阵的秩数就是梯形矩阵中不为零的行数．这是因为行变换不会改变矩阵的秩．

证明：如果通过行变换可以把矩阵的某行变为零，那么就说明该行必定可以表示为其他行的线性组合；而梯形矩阵中不为零的行都是线性无关的（具体证明留作习题）；

例 2

我们用高斯消元法计算式 1 的秩，该矩阵经过行变换 $\mathbf{r}_2 \leftrightarrow \mathbf{r}_1$，$\mathbf{r}_2 - \mathbf{r}_1$，$\mathbf{r}_3 - \mathbf{r}_2$ 后变为梯形矩阵

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \tag{4}$$

有两个不为零的行，所以矩阵的秩为 2．

厄米共轭算符的映射结构

预备知识 矩阵与矢量空间[456]，正交子空间[527]，线性映射的结构[468]，厄米共轭[422]

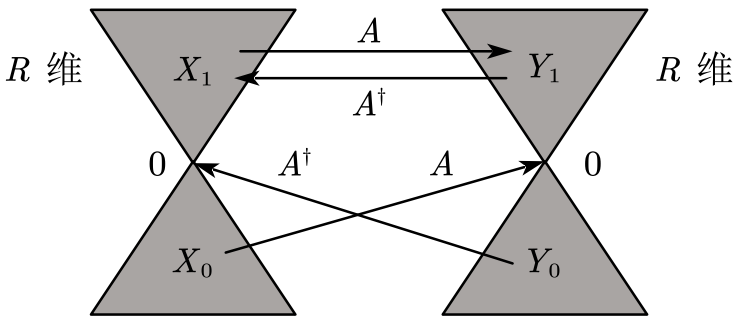


图 1: 定理 1 中的韦恩图，三角形表示线性空间，$X_0 \cap X_1$ 和 $Y_0 \cap Y_1$ 都是零向量.

定理 1

令 X, Y 为有限维线性空间，线性算符 $A : X \to Y$ 的零空间（定理 2 [451]）为 X_0，其厄米共轭[422]（也叫自伴算符）$A^\dagger : Y \to X$ 的零空间为 Y_0；令两算符的值空间分别为 $Y_1 = A(X)$，$X_1 = A^\dagger(Y)$. 那么 X_0, X_1 是 X 中的正交补（定义 2 [527]），Y_0, Y_1 是 Y 中的正交补. 即

$$X = X_0 \oplus X_1 \qquad X_0 \perp X_1 \tag{1}$$

$$Y = Y_0 \oplus Y_1 \qquad Y_0 \perp Y_1 \tag{2}$$

其中 $\oplus$ 表示直和[465]，$\perp$ 表示两空间正交[527].

证明见下文.

推论 1

定理 1 中 X_1, Y_1 维数相同，满足（见图 1）

$$A(X_1) = Y_1 \qquad A^\dagger(Y_1) = X_1 \tag{3}$$

且这两个映射都是双射.

根据定理 1 [468]，由于值空间的维数只能小于或等于定义域空间的维数，所以 X_1, Y_1 维数相同.

推论 2

矩阵的行秩等于列秩[407].

证明：令矩阵 $\mathbf{A}$ 的算符为 A，该矩阵的列秩就是 Y_1 的维数．而 $\mathbf{A}$ 的行秩等于 $\mathbf{A}^\dagger$ 的列秩，也就是 X_1 的维数．而 X_1, Y_1 维数相同．证毕．

证明

以下证明定理 1 中 Y_1 的正交补就是 Y_0（X_0, X_1 的证明同理）．令 Y_1 的正交补为 Y_0'，那么 $y \in Y_0'$ 的充分必要条件是 y 和 Y_1 中所有矢量都正交，即

$$\langle Ax|y\rangle = 0 \qquad (\forall x \in X) \tag{4}$$

而根据零空间的定义（定理 2 [451]）$y \in Y_0$ 的充分必要条件是 $A^\dagger y = 0$，即

$$\langle x|A^\dagger y\rangle = 0 \qquad (\forall x \in X) \tag{5}$$

而根据（链接未完成）式 4 和式 5 等效．所以 $Y_0' = Y_0$．证毕．

矩阵的迹

预备知识 相似变换和相似矩阵[429]

定义 1　矩阵的迹

令域 $\mathbb{F}$ 上的 N 维方阵 $\mathbf{A}$ 的矩阵元为 $a_{ij} \in \mathbb{F}$，它的迹（**trace**）定义为对角线上矩阵元之和

$$\mathrm{tr}(A) = \sum_{i=0}^{N} a_{ii} \tag{1}$$

矩阵的迹是刻画矩阵性质的一个量，它的优点在于满足以下一些性质，从而成为了分析矩阵的变换的利器．

性质

定理 1　线性

矩阵的求迹操作是线性的，即对于域 $\mathbb{F}$ 上的方阵 $\mathbf{A}$，$\mathbf{B}$ 和域中元素 c_1，c_2，

有

$$\mathrm{tr}(c_1\mathbf{A}+c_2\mathbf{B})=c_1\,\mathrm{tr}(\mathbf{A})+c_2\,\mathrm{tr}(\mathbf{B}) \tag{2}$$

证明留作练习.

定理 2　交换性

矩阵乘法的迹满足（$\mathbf{A}$ 和 $\mathbf{B}$ 不必是方阵，但要求乘积是方阵．注意 $\mathbf{AB}$ 的尺寸和 $\mathbf{BA}$ 未必相同）

$$\mathrm{tr}(\mathbf{AB})=\mathrm{tr}(\mathbf{BA}) \tag{3}$$

证明：令式 3 中 $\mathbf{A}$ 为 $M\times N$ 的矩阵，$\mathbf{B}$ 为 $N\times M$ 的矩阵

$$\mathrm{tr}(\mathbf{AB})=\sum_{i=1}^{M}\sum_{k=1}^{N}a_{ik}b_{ki}=\sum_{i=1}^{M}\sum_{k=1}^{N}b_{ki}a_{ik}=\sum_{i=1}^{N}\sum_{k=1}^{M}b_{ik}a_{ki}=\mathrm{tr}(\mathbf{BA}) \tag{4}$$

证毕.

定理 3

相似矩阵的迹相等.

证明：根据交换性式 3

$$\mathrm{tr}(\mathbf{P}^{-1}\mathbf{AP})=\mathrm{tr}(\mathbf{APP}^{-1})=\mathrm{tr}(\mathbf{A})$$

证毕.

定理 4

矩阵的迹等于它的所有 N 个本征值 λ_i 相加，如果某个本征值有 n 重简并，就视为 n 个本征值

$$\mathrm{tr}(\mathbf{A})=\sum_{i=0}^{N}\lambda_i \tag{5}$$

习题 1

对于矩阵 $\mathbf{A}$，若记其转置为 $\mathbf{A}^{\mathrm{T}}$，共轭为 $\overline{\mathbf{A}}$，厄米共轭为 $\mathbf{A}^{\dagger}$，那么有：

$$\mathrm{tr}(A^{\mathrm{T}})=\mathrm{tr}(A),\qquad \mathrm{tr}(\overline{\mathbf{A}})=\overline{\mathrm{tr}(\mathbf{A})} \tag{6}$$

由此可得推论：$\mathrm{tr}(\mathbf{A}^{\dagger})=\mathrm{tr}(\overline{\mathbf{A}^{\mathrm{T}}})=\overline{\mathrm{tr}(\mathbf{A})}$.

例 1

$\mathrm{tr}(AA^{\dagger})=0$ 当且仅当 $\mathrm{tr}(A)=0$.

矩阵的迹的应用实例

习题 2

对于域 $\mathbb{C}$ 上的方阵 $\mathbf{A}$，如果 $\mathbf{A}^2 = \mathbf{A}\mathbf{A}^\dagger$，求证 $\mathbf{A} = \mathbf{A}^\dagger$.

提示：使用例 1 的结论，证明 $\mathbf{A} - \mathbf{A}^\dagger = 0$ 即可.

高斯消元法求逆矩阵

预备知识 高斯消元法解线性方程组[477]，逆矩阵[405]

对可逆矩阵 $\mathbf{M}$ 做行变换相当于左乘一个矩阵 $\mathbf{R}$. 假设某种行变换能使 $\mathbf{M}$ 变为单位矩阵，即

$$\mathbf{R}\mathbf{M} = \mathbf{I} \tag{1}$$

那么根据定义 $\mathbf{R}$ 就是 $\mathbf{M}$ 的逆矩阵即 $\mathbf{R} = \mathbf{M}^{-1}$. 利用这个性质，我们可以同时对 $\mathbf{M}$ 和 $\mathbf{I}$ 做相同的行变换，当 $\mathbf{M}$ 变为单位矩阵后，$\mathbf{I}$ 就变为 $\mathbf{M}^{-1}$：

$$\mathbf{R}\mathbf{M} = \mathbf{I} \qquad \mathbf{R}\mathbf{I} = \mathbf{M}^{-1} \tag{2}$$

注意 $\mathbf{M}$ 可逆当且仅当它是一个满秩矩阵，即每行都线性无关的方阵.

例 1　求逆矩阵

求矩阵 $\mathbf{M} = \begin{pmatrix} 1 & 4 \\ 2 & 9 \end{pmatrix}$ 的逆矩阵.

解：先并列写出 $\mathbf{M}$ 和单位矩阵 $\mathbf{I}$，以下所有行变换都对两个矩阵同时进行

$$\begin{pmatrix} 1 & 4 \\ 2 & 9 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3}$$

第一行乘 -2 加到第二行得

$$\begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \tag{4}$$

第二行乘 -4 加到第一行得

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 9 & -4 \\ -2 & 1 \end{pmatrix} \tag{5}$$

所以 $\mathbf{M}^{-1}=\begin{pmatrix}9 & -4\\ -2 & 1\end{pmatrix}$.

同理，我们也可以把上文中所有行变换改为列变换，列变换相当于把矩阵 $\mathbf{M}$ 右乘一个矩阵 $\mathbf{C}$：

$$\mathbf{MC}=\mathbf{I}\qquad \mathbf{IC}=\mathbf{M}^{-1} \tag{6}$$

习题 1

用列变换的方法求例 1 中的逆矩阵.

克拉默法则

- 本词条存在未完成的内容.

预备知识 行列式[432]

[4]**克拉默法则（Kramer's rule）**是一种直接用行列式解线性方程组的方法. 把线性方程组记为矩阵乘法[396] 的形式

$$\mathbf{Ax}=\mathbf{b} \tag{1}$$

其中 $\mathbf{A}$ 为系数矩阵. 当 $\mathbf{A}$ 为 $N\times N$ 的方阵且行列式 $|\mathbf{A}|\neq 0$ 时（即满秩矩阵[407]），方程有唯一解（见“线性方程组解的结构[482]”）. 该解可以用克拉默法则直接写出：

$$x_i=\frac{|\mathbf{A}_i|}{|\mathbf{A}|}\qquad (i=1,\dots,N) \tag{2}$$

其中 $\mathbf{A}_i$ 是把 $\mathbf{A}$ 的第 i 列替换为 $\mathbf{b}$ 而来.

例 1　解方程组

令式 1 中 $\mathbf{A}=\begin{pmatrix}2 & 1\\ -1 & 3\end{pmatrix}$，$\mathbf{b}=\begin{pmatrix}4\\ 5\end{pmatrix}$，求解方程组.

解：$|\mathbf{A}|=7, |\mathbf{A}_1|=\begin{vmatrix}4 & 1\\ 5 & 3\end{vmatrix}=7, |\mathbf{A}_2|=\begin{vmatrix}2 & 4\\ -1 & 5\end{vmatrix}=14$. 代入式 2 得 $\mathbf{x}=\begin{pmatrix}1\\ 2\end{pmatrix}$.

在数值计算时，克拉默法则解方程组效率较低，直接用高斯消元法求逆矩阵高斯消元法求逆矩阵[412] 会更快.

[4]参考 Wikipedia 相关页面.

证明

证明需要代数余子式 $\mathbf{A}^*$（未完成）

$$\mathbf{A}^{-1}=\frac{\mathbf{A}^*}{|\mathbf{A}|} \tag{3}$$

$$\mathbf{x}=\frac{\mathbf{A}^*\mathbf{b}}{|\mathbf{A}|} \tag{4}$$

过渡矩阵

预备知识 矢量空间的表示[458]

过渡矩阵

给定域 $\mathbb{F}$ 上的 n 维矢量空间 V. 如果 V 有两个基 $\{\mathbf{e}_i\}_{i=1}^n$ 和 $\{\mathbf{e}'_i\}_{i=1}^n$，那么由于各 $\mathbf{e}'_j$ 也是 V 中的元素，故可以表示成 $\mathbf{e}'_j=\sum_{i=1}^{n}a_{ij}\mathbf{e}_i$，因此我们可以把矢量也当作矩阵元素，写出以下矩阵等式：

$$(\mathbf{e}'_1,\cdots\mathbf{e}'_n)=(\mathbf{e}_1,\cdots\mathbf{e}_n)\begin{pmatrix}a_{11},a_{12},\cdots,a_{1n}\\a_{21},a_{22},\cdots,a_{2n}\\\vdots\ddots\vdots\\a_{n1},a_{n2},\cdots,a_{nn}\end{pmatrix} \tag{1}$$

矩阵 $\mathbf{Q}=\begin{pmatrix}a_{11},a_{12},\cdots,a_{1n}\\a_{21},a_{22},\cdots,a_{2n}\\\vdots\ddots\vdots\\a_{n1},a_{n2},\cdots,a_{nn}\end{pmatrix}$ 就被称为基 $\{\mathbf{e}_i\}_{i=1}^n$ 到基 $\{\mathbf{e}'_i\}_{i=1}^n$ 的一个**过渡矩阵（transition matrix）**.

例 1

证明：过渡矩阵必须是可逆矩阵.

过渡矩阵和基的变换

过渡矩阵本身是表示基之间的变换的，但它也可以用来表示两个不同的基下各种表示的变换，比如向量的坐标、线性变换的矩阵等.

不同基下坐标的变换

如果在基 $\{\mathbf{e}_i\}_{i=1}^n$ 中，向量 $\mathbf{v}$ 的坐标是列向量 $\mathbf{c}=\begin{pmatrix}c_1,\cdots,c_n\end{pmatrix}^T$，那么就有

$$\mathbf{v}=(\mathbf{e}_1,\cdots\mathbf{e}_n)\begin{pmatrix}c_1\\ \vdots\\ c_n\end{pmatrix} \tag{2}$$

同样地，如果在基 $\{\mathbf{e}'_i\}_{i=1}^n$ 中，$\mathbf{v}$ 的坐标是列向量 $\mathbf{c}'=\begin{pmatrix}c'_1,\cdots,c'_n\end{pmatrix}^T$，则也有

$$\mathbf{v}=(\mathbf{e}'_1,\cdots\mathbf{e}'_n)\begin{pmatrix}c'_1\\ \vdots\\ c'_n\end{pmatrix} \tag{3}$$

考虑到 $\mathbf{v}=\mathbf{v}$，联立式 1，式 2 和式 3 即可得到：

$$\begin{pmatrix}a_{11},a_{12},\cdots,a_{1n}\\ a_{21},a_{22},\cdots,a_{2n}\\ \vdots\ddots\vdots\\ a_{n1},a_{n2},\cdots,a_{nn}\end{pmatrix}\begin{pmatrix}c'_1\\ \vdots\\ c'_n\end{pmatrix}=\begin{pmatrix}c_1\\ \vdots\\ c_n\end{pmatrix} \tag{4}$$

式 4 也可以简单记为 $\mathbf{Q}\mathbf{c}'=\mathbf{c}$，或者 $\mathbf{c}'=\mathbf{Q}^{-1}\mathbf{c}$.

不同基下线性变换的矩阵的变换

设 T 是 V 上的一个线性变换，基 $\{\mathbf{e}_i\}_{i=1}^n$ 到 $\{\mathbf{e}'_i\}_{i=1}^n$ 的过渡矩阵是 $\mathbf{Q}$. 向量 $\mathbf{v}$ 在基 $\{\mathbf{e}_i\}_{i=1}^n$ 和 $\{\mathbf{e}'_i\}_{i=1}^n$ 下的坐标分别是 $\mathbf{c}$ 和 $\mathbf{c}'$，而 T 的矩阵分别是 $\mathbf{M}$ 和 $\mathbf{M}'$.

$T\mathbf{v}$ 在基 $\{\mathbf{e}_i\}_{i=1}^n$ 和 $\{\mathbf{e}'_i\}_{i=1}^n$ 下的坐标就分别是 $\mathbf{Mc}$ 和 $\mathbf{M}'\mathbf{c}'$．由坐标的变换可知，$\mathbf{QM}'\mathbf{c}'=\mathbf{Mc}$，即 $\mathbf{M}'\mathbf{c}'=\mathbf{Q}^{-1}\mathbf{Mc}$.

又因为 $\mathbf{Qc}'=\mathbf{c}$，故有 $\mathbf{M}'\mathbf{c}'=\mathbf{Q}^{-1}\mathbf{MQc}$.

也就是说，如果在基 $\{\mathbf{e}_i\}_{i=1}^n$ 下，线性变换 T 的矩阵是 $\mathbf{M}$，那么在基 $\{\mathbf{e}'_i\}_{i=1}^n$ 下，T 的矩阵应该是 $\mathbf{Q}^{-1}\mathbf{MQc}$.

由此可见，矩阵的**相似变换**就是同一个线性变换在不同基下的矩阵之间的变换.

绕轴旋转矩阵

预备知识 空间旋转矩阵[404]，圆周运动的速度[883]

直角坐标系中，某点 $\mathbf{r}=(x,y,z)^{\mathrm{T}}$ 以单位矢量 $\hat{\mathbf{A}}=(A_x,A_y,A_z)^{\mathrm{T}}$ 为轴按右手定则转动 θ 角的得到的点 $\mathbf{r}'=(x',y',z')^{\mathrm{T}}$ 可用矩阵乘法计算

$$\mathbf{r}'=\mathbf{R}_\theta\mathbf{r} \tag{1}$$

其中 $\mathbf{R}_\theta$ 为**绕轴旋转矩阵**

$$\mathbf{R}_\theta=\begin{pmatrix} aA_x^2+c & aA_xA_y-sA_z & aA_xA_z+sA_y \\ aA_yA_x+sA_z & aA_y^2+c & aA_yA_z-sA_x \\ aA_zA_x-sA_y & aA_zA_y+sA_x & aA_z^2+c \end{pmatrix} \tag{2}$$

其中

$$c=\cos\theta \qquad s=\sin\theta \qquad a=1-\cos\theta \tag{3}$$

事实上，数学上更规范的做法是用四元数[419]表示该矩阵.

推导

推导的思路是用 $\hat{\mathbf{A}}$, $\mathbf{r}$ 和 θ 三个已知量经过数乘，内积[383]和叉乘[387]三种运算，表示出旋转后的矢量 $\mathbf{r}'$，再拆成三个分量，即可得到线性变换，进而写出矩阵．注意该思路与推导平面旋转矩阵[402]的思路不一样.

如图，$\mathbf{r}$ 绕单位矢量 $\hat{\mathbf{A}}$ 旋转 θ 角后得到 $\mathbf{r}'$．$\mathbf{r}$ 在 $\hat{\mathbf{A}}$ 方向投影得到的矢量为

$$\mathbf{r}_3=(\hat{\mathbf{A}}\cdot\mathbf{r})\hat{\mathbf{A}} \tag{4}$$

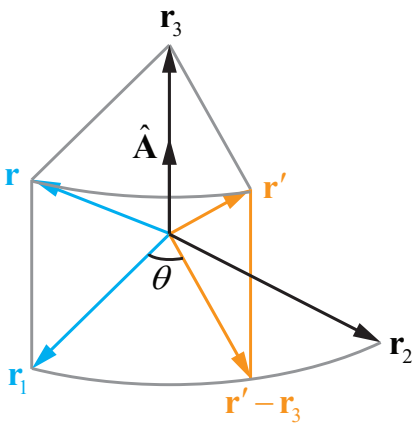


图 1: 绕轴旋转矩阵的推导

在与 $\hat{\mathbf{A}}$ 垂直方向的分量为

$$\mathbf{r}_1 = \mathbf{r} - \mathbf{r}_3 \tag{5}$$

为了构成一组正交基底，令

$$\mathbf{r}_2 = \hat{\mathbf{A}} \times \mathbf{r}_1 \tag{6}$$

则 $\mathbf{r}_2$ 相当于 $\mathbf{r}_1$ 绕 $\hat{\mathbf{A}}$ 旋转 $90°$．现在有了正交的 $\mathbf{r}_1$, $\mathbf{r}_2$ 就可以表示出 $\mathbf{r}_1$ 绕 $\hat{\mathbf{A}}$ 旋转 θ 角后的结果

$$\mathbf{r}' - \mathbf{r}_3 = \mathbf{r}_1 \cos\theta + \mathbf{r}_2 \sin\theta \tag{7}$$

即

$$\mathbf{r}' = \mathbf{r}_1 \cos\theta + \mathbf{r}_2 \sin\theta + \mathbf{r}_3 \tag{8}$$

将式 4 式 5 式 6 代入式 8，即可求出 $\mathbf{r}'$ 关于 $\hat{\mathbf{A}}$, $\mathbf{r}$ 和 θ 的矢量表达式．把结果写成分量的形式，化简可得到 x', y', z' 关于 x, y, z 的线性变换与系数矩阵[396]．

由旋转矩阵推导出匀速圆周运动的线速度

我们可以用旋转矩阵得到 $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$ （式 5 [883]），这也验证了旋转矩阵的正确性．

在无穷小的时间 t 内，点 P 绕轴转过 θ 角，则 $\theta = \omega t \to 0$，此时有 $\sin\theta \to \theta$ 和 $\cos\theta \to 1$．旋转矩阵变为

$$\mathbf{R}_\theta = \begin{pmatrix} 1 & -A_z\theta & A_y\theta \\ A_z\theta & 1 & -A_x\theta \\ -A_y\theta & A_x\theta & 1 \end{pmatrix} \tag{9}$$

下面 $\mathbf{R}_\theta$ 乘以某点的列矢量，得到变换后的坐标，再减掉变换前的坐标，得位移矢量 $\mathbf{s}$

$$\begin{aligned}\mathbf{s} &= \mathbf{v}t \\ &= \begin{pmatrix} 1 & -A_z\theta & A_y\theta \\ A_z\theta & 1 & -A_x\theta \\ -A_y\theta & A_x\theta & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ &= \theta \begin{pmatrix} 0 & -A_z & A_y \\ A_z & 0 & -A_x \\ -A_y & A_x & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ &= \theta \hat{\mathbf{A}} \times \mathbf{r} = (\boldsymbol{\omega} t) \times \mathbf{r}\end{aligned} \tag{10}$$

两边除以 t，得 $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$.

旋转矩阵的导数

预备知识 圆周运动的速度[883]，瞬时转轴[977]，叉乘的矩阵形式[406]，旋转矩阵[404]

我们来证明含参数 t [5]的旋转矩阵 $\mathbf{R}(t)$ 的导数 $\dot{\mathbf{R}}$ 满足

$$\dot{\mathbf{R}} = \boldsymbol{\Omega}\mathbf{R} \tag{1}$$

其中

$$\boldsymbol{\Omega} = \begin{pmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{pmatrix} \tag{2}$$

其中 $(\omega_x, \omega_y, \omega_z)$ 是旋转的瞬时转轴的坐标[6].

[5] t 不一定表示时间，可以是任意参数

[6] 如果使用主动理解，即旋转矩阵将同一个坐标系中的一个矢量旋转后得到另一个矢量．如果使用被动理解，即旋转矩阵将同一个矢量在两个坐标系之间进行变换，则矢量 $(\omega_x, \omega_y, \omega_z)$ 取反方向即可

证明

我们可以把一个任意的不随 t 变化的列矢量 $\mathbf{r}_0$ 旋转得

$$\mathbf{r}(t) = \mathbf{R}(t)\mathbf{r}_0 \tag{3}$$

关于 t 求导得

$$\dot{\mathbf{r}} = \dot{\mathbf{R}}\mathbf{r}_0 \tag{4}$$

令列矢量 $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)^{\mathrm{T}}$，由式 5 [883] 得

$$\dot{\mathbf{r}} = \boldsymbol{\omega} \times \mathbf{r} \tag{5}$$

将叉乘表示为矩阵乘法（式 2 [406]）得

$$\dot{\mathbf{r}} = \boldsymbol{\Omega}\mathbf{r} \tag{6}$$

其中

$$\boldsymbol{\Omega} = \begin{pmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{pmatrix} \tag{7}$$

将式 3 代入式 6 右边，再对比式 4，可得式 1．证毕．

四元数与旋转矩阵

- 本词条需要更多讲解，便于帮助理解．
- 本词条存在未完成的内容．

预备知识 绕轴旋转矩阵[416]，旋转矩阵的导数[418]，四元数[593]

四元数可以用来简洁地表示三维空间中的旋转，极大地减少了计算量．

三维空间中的一个向量表示为标部为 0 的四元数：$v = \left(0, \mathbf{v}\right)$．如果我们绕着一个单位向量 $\hat{\mathbf{n}}$ 把 $\mathbf{v}$ 旋转一个角度 θ，所得的结果应该是哪个向量呢？取四元数 $q = \left(\cos\frac{\theta}{2}, \hat{\mathbf{n}}\sin\frac{\theta}{2}\right)$，那么旋转后的向量就可以表示为

$$qvq^{-1} \tag{1}$$

比如说，取 $v=(0,1,0,0)$，它代表一个 x 轴上的单位向量．如果想要把它绕着 z 轴上的单位向量转 $\pi/2$，那么结果向量的四元数表示应该是 $(0,0,1,0)$．按照我们定义的规则，旋转表示为四元数 $q=(\sqrt{2}/2,0,0,\sqrt{2}/2)$，按照四元数的乘法规则易得 $qvq^{-1}=(0,0,1,0)$．

证明

取四元数 $q=\left(\cos\frac{\theta}{2},\hat{\mathbf{n}}\sin\frac{\theta}{2}\right)$ 和 $v=(0,\mathbf{v})$，直接计算得：

$$\begin{aligned}
qvq^{-1} &= \left(\cos\tfrac{\theta}{2},\hat{\mathbf{n}}\sin\tfrac{\theta}{2}\right)\cdot(0,\mathbf{v})\cdot q^{-1}\\
&= (-\mathbf{v}\cdot\hat{\mathbf{n}}\sin\frac{\theta}{2},\mathbf{v}\cos\frac{\theta}{2}+\hat{\mathbf{n}}\times\mathbf{v}\sin\frac{\theta}{2})\cdot(\cos\frac{\theta}{2},-\hat{\mathbf{n}}\sin\frac{\theta}{2})\\
&= (0,\cos^2\frac{\theta}{2}\mathbf{v}+\cos\frac{\theta}{2}\sin\frac{\theta}{2}\hat{\mathbf{n}}\times\mathbf{v}+(\hat{\mathbf{n}}\cdot\mathbf{v})\sin^2\frac{\theta}{2}\hat{\mathbf{n}}\\
&\quad+\cos\frac{\theta}{2}\sin\frac{\theta}{2}\hat{\mathbf{n}}\times\mathbf{v}-\sin^2\frac{\theta}{2}\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\mathbf{v}))\\
&= (0,\cos^2\frac{\theta}{2}\mathbf{v}-\sin^2\frac{\theta}{2}\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\mathbf{v})+\sin\theta\hat{\mathbf{n}}\times\mathbf{v}\\
&\quad+(\hat{\mathbf{n}}\cdot\mathbf{v})\sin^2\frac{\theta}{2}\hat{\mathbf{n}})
\end{aligned}\tag{2}$$

如果取 $\mathbf{v}\perp\hat{\mathbf{n}}$，考虑到此时 $\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\mathbf{v})=-\mathbf{v}$ 且 $\hat{\mathbf{n}}\cdot\mathbf{v}=0$，那么上式可以化简为

$$qvq^{-1}=(0,\cos\theta\mathbf{v}+\sin\theta\hat{\mathbf{n}}\times\mathbf{v})\tag{3}$$

这正是 $\mathbf{v}$ 绕 $\hat{\mathbf{n}}$ 旋转 θ 的结果．

如果取 $\mathbf{v}\parallel\hat{\mathbf{n}}$，考虑到此时 $\hat{\mathbf{n}}\times\mathbf{v}=0$ 且 $(\hat{\mathbf{n}}\cdot\mathbf{v})\mathbf{v}=\mathbf{v}$，那么上式化为

$$qvq^{-1}=(0,\cos^2\frac{\theta}{2}\mathbf{v}+\sin^2\frac{\theta}{2}\mathbf{v})=(0,\mathbf{v})=v\tag{4}$$

这也是 $\mathbf{v}$ 绕 $\hat{\mathbf{n}}$ 旋转 θ 的结果．

综上所述，取任意向量 $\mathbf{v}$，把它分为垂直和平行于旋转轴的 $\mathbf{v}_\perp$ 和 $\mathbf{v}_\parallel$，分别对应四元数 $v_\perp$ 和 $v_\parallel$，那么 $qv_\perp q^{-1}$ 和 $qv_\parallel q^{-1}$ 分别对应这两个向量旋转后的结果．故取 $v=v_\perp+v_\parallel$ 为 $\mathbf{v}$ 对应的四元数后可由四元数乘法分配律得到，qvq^{-1} 对应 $\mathbf{v}$ 旋转后的向量．

一点补充：考虑到四元数限定向量方向时退化为复数（见**四元数**[593] 的讨论），故 $\mathbf{v}\parallel\hat{\mathbf{n}}$ 的情况其实可以直接套用复数的交换性来得到，即 $qvq^{-1}=qq^{-1}v=v$．

旋转矩阵

我们可以用**四元数（quaternions）** $\mathbf{q}=[s,\mathbf{v}]$ 来表示绕轴旋转矩阵，其中

$$s=\cos\left(\frac{\theta}{2}\right)\qquad v=|\mathbf{v}|=\sin\left(\frac{\theta}{2}\right)\hat{\mathbf{A}} \tag{5}$$

则绕轴旋转矩阵[416] 可以表示为

$$\mathbf{R}(\theta)=\begin{pmatrix}1-2v_y^2-2v_z^2 & 2v_xv_y-2sv_z & 2v_xv_z+2sv_y\\ 2v_xv_y+2sv_z & 1-2v_x^2-2v_z^2 & 2v_yv_z-2sv_x\\ 2v_xv_z-2sv_y & 2v_yv_z+2sv_x & 1-2v_x^2-2v_y^2\end{pmatrix} \tag{6}$$

四元数的乘法运算可以表示两个旋转矩阵相乘，即把两次旋转合并为一次旋转

$$[s_1,\mathbf{v}_1][s_2,\mathbf{v}_2]=[s_1s_2-\mathbf{v}_1\cdot\mathbf{v}_2,s_1\mathbf{v}_2+s_2\mathbf{v}_1+\mathbf{v}_1\times\mathbf{v}_2] \tag{7}$$

导数

若从坐标系 B 到坐标系 A 的基底变换矩阵为 $\mathbf{R}$，当 B 绕原点以角速度 $\boldsymbol{\omega}$ 旋转时有（式 1 [418]）

$$\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t}=\boldsymbol{\Omega}\mathbf{R} \tag{8}$$

其中 $\boldsymbol{\Omega}$ 乘以任意位置矢量 $\mathbf{r}$ 等于 $\boldsymbol{\omega}\times\mathbf{r}$

$$\boldsymbol{\Omega}=\begin{pmatrix}0 & -\omega_z & \omega_y\\ \omega_z & 0 & -\omega_x\\ -\omega_y & \omega_x & 0\end{pmatrix} \tag{9}$$

若旋转矩阵 $\mathbf{R}$ 对应的四元数为 $\mathbf{q}$，则

$$\dot{\mathbf{q}}(t)=\frac{1}{2}[0,\boldsymbol{\omega}(t)]\mathbf{q}(t) \tag{10}$$

对称矩阵

- 本词条处于草稿阶段.
- 直接从“厄米矩阵”里面修改搬过来

预备知识 矩阵[396]

定义 1

若方阵 $\mathbf{A}$ 的任意矩阵元 $a_{i,j}$ 满足 $a_{i,j}=a_{j,i}$，那么 $\mathbf{A}$ 就叫做对称矩阵（**symmetric matrix**）.

根据定义，一个矩阵 $\mathbf{A}$ 是对称矩阵当且仅当它等于自己的转置，即

$$\mathbf{A}^{\mathrm{T}}=\mathbf{A} \tag{1}$$

在线性代数中，对称矩阵一般指实对称矩阵，即矩阵元都是实数. 实对称矩阵也是一个厄米矩阵[422].

N 阶对称矩阵或厄米矩阵存在 N 个正交归一的本征矢[425].

厄米矩阵

预备知识 对称矩阵[421]

厄米共轭

我们把矩阵元可以取复数的矩阵叫做**复数矩阵**. 与矩阵转置（式 1 [396]）类似，要对一个复数矩阵做**厄米共轭（Hermitian conjugate）**，就先将其做转置，再对每个矩阵元取复共轭. 矩阵 $\mathbf{A}$ 的厄米共轭记为 $\mathbf{A}^{\dagger}$，其第 i 行第 j 列的矩阵元为

$$A^{\dagger}_{i,j}=A^{*}_{j,i} \tag{1}$$

注意当矩阵元都是实数时，厄米共轭就是转置. $\mathbf{A}^{\dagger}$ 也称为 $\mathbf{A}$ 的**伴随矩阵（adjoint matrix）**.

厄米共轭的基本性质

任意常数乘以厄米矩阵再取共轭，等于该常数的复共轭乘以矩阵的厄米共轭

$$(c\mathbf{A})^\dagger = c^*\mathbf{A}^\dagger \tag{2}$$

类比转置，矩阵相乘的厄米共轭等于分别做厄米共轭，逆序排列，再相乘

$$(\mathbf{AB}\dots\mathbf{C})^\dagger = \mathbf{C}^\dagger\dots\mathbf{B}^\dagger\mathbf{A}^\dagger \tag{3}$$

习题 1　证明

请根据相关定义证明式 2 和式 3.

厄米矩阵

若 $\mathbf{A}$ 的厄米共轭是其本身，即

$$\mathbf{A}^\dagger = \mathbf{A} \tag{4}$$

那么我们称其为**厄米矩阵**. 厄米矩阵关于对角线对称的任意两个矩阵元互为复共轭

$$A_{i,j} = A^*_{j,i} \tag{5}$$

特殊地，对角线上的矩阵元的复共轭等于本身（$A_{i,i} = A^*_{i,i}$），所以厄米矩阵对角线上的矩阵元都是实数.

事实上，厄米矩阵也可以等效定义为满足

$$\mathbf{v}_i^\dagger(\mathbf{A}\mathbf{v}_j) = (\mathbf{A}\mathbf{v}_i)^\dagger\mathbf{v}_j \tag{6}$$

对任意列矢量 $\mathbf{v}_i$ 和 $\mathbf{v}_j$ 都成立的矩阵. 这可以用式 3 证明.

矩阵的本征方程

预备知识　线性方程组与矢量空间[482]

若已知矩阵 $\mathbf{A}$，我们把线性方程组

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \tag{1}$$

称为矩阵 $\mathbf{A}$ 的**本征方程**．式中 $\mathbf{A}$ 是已知的，而 λ 和 $\mathbf{v}$ 是未知的．显然，当 $\mathbf{v} = \mathbf{0}$ 时方程恒成立，所以我们通常只对非零解感兴趣．也就是说，我们希望找到一些非零矢量 $\mathbf{v}$，使得矩阵 $\mathbf{A}$ 乘以该矢量以后方向不变[7]．对于每个这样的矢量，我们用一个标量 λ 来描述其模长的改变．我们把这些矢量叫做**本征矢（eigen vector）**，把对应的 λ 叫做**本征值（eigen value）**．

若令 $\mathbf{I}$ 为 $N \times N$ 的单位矩阵[8]，则本征方程本质上是一个齐次方程组

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0} \tag{2}$$

括号中的矩阵相当于把矩阵 $\mathbf{A}$ 的对角线上的元都减去 λ 得到的方阵．要确保方程有非零解，只需令系数矩阵 $\mathbf{A} - \lambda\mathbf{I}$ 不是满秩的，即行列式为零

$$|\mathbf{A} - \lambda\mathbf{I}| = 0 \tag{3}$$

这是一个关于 λ 的 N 阶多项式，称为**特征多项式（characteristic polynomial）**．特征多项式必存在 N 个复数根（包括重根），记为 λ_i （$i = 1, 2 \ldots N$）．将它们依次代入式 2，就可以分别解出对应的本征矢．考虑到式 2 是一个齐次方程，所以 $\mathbf{A} - \lambda_i\mathbf{I}$ 的零空间中所有矢量都是本征矢，且零空间至少是一维的．我们把这个空间叫做 λ_i 的**本征矢空间**，是 $\mathbf{v}$ 所在的矢量空间的子空间．

令 λ_i 的本征矢空间的维度是 n_i，若 $n_i = 1$，我们说 λ_i 是**非简并（non-degenerate）**的，若 $n_i > 1$ 就说 λ_i 是 n_i 重**简并（degenerate）**的，把 n_i 叫做**简并数（degeneracy）**．

> 未完成：简并子空间应该写到哪里？

对角化与相似变换

求解矩阵的本征方程有时候也称为矩阵的**对角化**，原因如下：若我们把矩阵 $\mathbf{A}$ 的 i 个本征值和本征列矢量记为 λ_i 和 $\mathbf{v}_i$．如果把本征值按顺序组成对角

[7] “方向”只是从几何矢量[370]中沿用过来的一个习惯说法，注意式 1 中的所有量都可以是复数．两个矢量方向相同意味着一个矢量乘以标量（包括复数）可以得到另一个．

[8] 即对角线上的元为 1，其他元为 0，见“矩阵[396]”

矩阵 $\mathbf{\Lambda}$，把本征矢从按顺序从左到右组成方阵 $\mathbf{P}$，那么根据矩阵乘法[396] 的定义 $\mathbf{AP}$ 相当于分别计算 $\mathbf{A}\mathbf{v}_i$ 再从左到右排列．而 $\mathbf{P\Lambda}$ 相当于把 $\lambda_i\mathbf{v}_i$ 从左到右排成方阵．二者应该相等，所有

$$\mathbf{AP} = \mathbf{P\Lambda} \tag{4}$$

由于方阵 $\mathbf{P}$ 是满秩[407] 的（每列线性无关），必定存在逆矩阵 $\mathbf{P}^{-1}$．两边右乘 $\mathbf{P}^{-1}$ 得

$$\mathbf{A} = \mathbf{P\Lambda P}^{-1} \tag{5}$$

这种从 $\mathbf{A}$ 到 $\mathbf{\Lambda}$ 的变换被称为**相似变换（similarity transform）**．如果能找到使 $\mathbf{\Lambda}$ 为对角矩阵的 $\mathbf{P}$ 就相当于解出了本征方程式 1，这就是“对角化”名字的由来．

对称矩阵的本征问题

- 本词条存在未完成的内容．

预备知识 对称矩阵[421]，矩阵的本征方程[423]

在物理中，我们遇到的本征问题中的矩阵往往是对称矩阵或厄米矩阵．我们下面来证明 N 维对称矩阵 $\mathbf{A}$ 存在 N 个两两正交归一的本征矢 $\mathbf{v}_1, \ldots, \mathbf{v}_N$，且本征值都是实数．另外，本征矢一般用实数表示，但也可以乘以一个任意复数．

未完成：需要把以下结论用定理的形式列出来

证明本征值为实数

本征方程为

$$\mathbf{A}\mathbf{v}_i = \mathbf{v}_i \tag{1}$$

我们先假设本征值和本征矢都可能是复数，将本征方程左边乘以 $\mathbf{v}_i$ 得

$$\mathbf{v}_i^\dagger \mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i^\dagger \mathbf{v}_i \tag{2}$$

将等式两边取厄米共轭（注意矢量也可以看成矩阵），由式 3 [422] 和式 2 [422] 可得

$$\mathbf{v}_i^\dagger \mathbf{A}^{\mathrm{T}} \mathbf{v}_i = \mathbf{v}_i^\dagger \mathbf{A} \mathbf{v}_i = \lambda_i^* \mathbf{v}_i^\dagger \mathbf{v}_i \tag{3}$$

对比两式，得 $\lambda_i = \lambda_i^*$，所以 λ_i 必为实数.

将本征方程记为

$$(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{v}_i = \mathbf{0} \tag{4}$$

就会发现这是一个实系数的齐次方程组，所以必然能把所有线性无关的解都用实数矢量表示. 但注意把任意解 $\mathbf{v}_i$ 乘以一个任意复数，同样也是方程的解.

证明本征矢的正交性

简并空间内，我们可以认为地指定正交归一基底，所以只需要证明不同本征值对应的本征矢正交即可.

$$s = \mathbf{v}_1^{\mathrm{T}}(\mathbf{A}\mathbf{v}_2) = \mathbf{v}_1^{\mathrm{T}}(\lambda_2 \mathbf{v}_2) = \lambda_2 \mathbf{v}_1^{\mathrm{T}} \mathbf{v}_2 \tag{5}$$

使用矩阵乘法结合律式 18 [396] 以及式 3 [422] 得

$$s = (\mathbf{A}\mathbf{v}_1)^{\mathrm{T}} \mathbf{v}_2 = \lambda_1^* \mathbf{v}_1^{\mathrm{T}} \mathbf{v}_2 = \lambda_1 \mathbf{v}_1^{\mathrm{T}} \mathbf{v}_2 \tag{6}$$

以上两矢相等，因为 $\lambda_1 \neq \lambda_2$，所以 $\mathbf{v}_1^{\mathrm{T}} \mathbf{v}_2 = 0$.

事实上，厄米矩阵也可以定义为满足

$$\mathbf{v}_1^{\mathrm{T}}(\mathbf{A}\mathbf{v}_2) = (\mathbf{A}\mathbf{v}_1)^{\mathrm{T}} \mathbf{v}_2 \tag{7}$$

的矩阵.

厄米矩阵的本征问题

- 本词条处于草稿阶段.

预备知识 厄米矩阵[422]，矩阵的本征方程[423]，正交子空间[527]

本词条将“对称矩阵的本征值问题[425]”拓展到厄米矩阵，结论和过程都相似. 我们来证明 N 维厄米矩阵 $\mathbf{H}$ 存在 N 个两两正交归一的本征矢 $\mathbf{v}_1, \ldots, \mathbf{v}_N$.

未完成：需要把以下结论用定理的形式列出来

定理 1

厄米矩阵的本征问题具有以下两个性质

1. 本征值为实数
2. 本征值不同的本征矢正交
3. 存在本征矢构成的正交归一基底

从正交子空间的角度来转述 2,3 条就是：所有本征子空间 $V_1 \dots V_N$ 两两正交且互补，即

$$V_i \perp V_j \qquad (i,j=1,\dots,N, i\neq j) \tag{1}$$

$$V_1 \oplus V_2 \oplus \cdots \oplus V_N = V \tag{2}$$

证明

本征值为实数

本征方程为

$$\mathbf{H}\mathbf{v}_i = \mathbf{v}_i \tag{3}$$

将本征方程左边乘以 $\mathbf{v}_i$ 得

$$\mathbf{v}_i^\dagger \mathbf{H}\mathbf{v}_i = \lambda_i \mathbf{v}_i^\dagger \mathbf{v}_i \tag{4}$$

将等式两边取厄米共轭（注意矢量也可以看成矩阵），由式 3 [422] 和式 2 [422] 可得

$$\mathbf{v}_i^\dagger \mathbf{H}^\dagger \mathbf{v}_i = \mathbf{v}_i^\dagger \mathbf{H}\mathbf{v}_i = \lambda_i^* \mathbf{v}_i^\dagger \mathbf{v}_i \tag{5}$$

对比两式，得 $\lambda_i = \lambda_i^*$，所以 λ_i 必为实数.

本征矢的正交性

下面来证明不同本征值对应的本征矢正交，即

$$\mathbf{v}_i^\dagger \mathbf{v}_j = 0 \qquad (a_i \neq a_j) \tag{6}$$

首先令

$$s = \mathbf{v}_i^\dagger(\mathbf{H}\mathbf{v}_j) = \mathbf{v}_i^\dagger(\lambda_j\mathbf{v}_j) = \lambda_j\mathbf{v}_i^\dagger\mathbf{v}_j \tag{7}$$

使用矩阵乘法结合律式 18 [396] 以及式 3 [422] 得

$$s = (\mathbf{H}\mathbf{v}_i)^\dagger\mathbf{v}_j = \lambda_i^*\mathbf{v}_i^\dagger\mathbf{v}_j = \lambda_i\mathbf{v}_i^\dagger\mathbf{v}_j \tag{8}$$

以上两式相减得

$$(\lambda_j - \lambda_i)\mathbf{v}_i^\dagger\mathbf{v}_j = 0 \tag{9}$$

因为 $\lambda_i \neq \lambda_j$，所以 $\mathbf{v}_i^\dagger\mathbf{v}_j = 0$.

简并

在“矩阵的本征方程[423]”中，我们定义若令 λ_i 的本征矢空间的维数是 n_i，当 $n_i = 1$，我们说 λ_i 是**非简并（non-degenerate）**的，当 $n_i > 1$ 就说 λ_i 是 n_i 重**简并（degenerate）**的，把 n_i 叫做**简并数（degeneracy）**.

根据式 6，对于厄米矩阵，所有不同的 λ_i 对应的本征矢子空间是两两正交的，且 $\sum_i n_i = N$，所以所有这些子空间的直和就是 $\mathbf{A}$ 的定义域空间，即他们互为正交补[527].

未完成：如何证明 $\sum_i n_i = N$ 呢？估计要证明 m 重跟可以使得行稀疏矩阵的秩为 $N - m$.

完备正交基底

对任意的 N 维厄米矩阵 $\mathbf{H}$ 都能得到由本征矢构成的 N 个正交归一基底.

- 如果不存在简并，这组基底是唯一的，且它们的本征值各不相同.
- 如果存在简并，每个子空间中可以选出 n_i 个正交归一基底，将他们放在一起就得到总空间中的 N 个正交归一基底. 注意每个简并子空间中的正交归一基底的选取都是任意的.

未完成：举一个综合的例子，包含简并的

相似变换和相似矩阵

预备知识　酉矩阵[440]

定义 1　相似变换

令 $\mathbf{U}$ 为 N 维酉矩阵．N 维方阵 $\mathbf{A}$ 的相似变换为

$$\mathbf{B}=\mathbf{U}\mathbf{A}\mathbf{U}^{\dagger} \tag{1}$$

注意酉矩阵满足 $\mathbf{U}^{\dagger}=\mathbf{U}^{-1}$．

由于酉矩阵乘以酉矩阵还是酉矩阵，多次相似变换可以看作一次相似变换．

对角化

预备知识　厄米矩阵的本征问题[426]

若相似变换可以使矩阵变为对角矩阵，我们把这个过程称为对角化．

一个 N 维矩阵可以被对角化当且仅当它是（实）对称矩阵或厄米矩阵．

$$\mathbf{U}^{\dagger}\mathbf{A}\mathbf{U}=\mathbf{\Lambda} \tag{2}$$

对角化后，对角矩阵 $\mathbf{\Lambda}$ 的对角元就是矩阵 $\mathbf{A}$ 的本征值 λ_i，$\mathbf{U}$ 的第 i 列矢量就是 λ_i 对应的本征矢．所以我们时常把“对角化”作为“解矩阵的本征方程”的同义词．

（未完成）

例 1　由本征值和本征矢求矩阵

已知本征方程

$$\mathbf{A}\mathbf{v}=\lambda\mathbf{v} \tag{3}$$

的 N 个本征值和本征矢为 λ_i 和 $\mathbf{v}_i$，求矩阵 $\mathbf{A}$．

把式 2 两边分别左乘 $\mathbf{U}$，右乘 $\mathbf{U}^{\dagger}$，得

$$\mathbf{A}=\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\dagger} \tag{4}$$

逆序数

预备知识 排列[106]

[9]我们把集合 $\{1,\dots,N\}$ 的某种排列记为 π，该排列的元素按照顺序分别记为 $\pi_1,\dots,\pi_N$．对于任意 $i<j$，如果满足 $\pi_i>\pi_j$ 我们就把 i,j 或者 π_i,π_j 称为排列 π 的一个**逆序对（inversion）**．一个排列中所有逆序对的个数就叫**逆序数（inversion number）**．

逆序数主要的应用有定义行列式[432] 和列维—奇维塔（Levi-Civita）符号[431]．在这两个应用中，我们只对逆序数的奇偶性感兴趣．我们把逆序数为奇数的排列叫做**奇排列**，逆序数为偶数的排列叫做**偶排列**．把按从小到大的顺序排列叫做**标准排列**．标准排列逆序数为零，是一个偶排列．

例 1

- 标准排列 $1,2,3,4$ 中对任意 $i<j$ 都有 $\pi_i<\pi_j$，不存在逆序对，逆序数为零．
- $4,3,2,1$ 中对任意 $i<j$ 都有 $\pi_i>\pi_j$，任何两个元素都是逆序对，逆序数等于组合 $C_4^2=6$．
- 排列 $1,6,2,5,3,4$ 中，逆序对有 $(6,2)$，$(6,5)$，$(6,3)$，$(6,4)$，$(5,3)$，$(5,4)$．共 6 个．

一般情况下，要计算逆序数，我们可以依次令 $i=1,\dots,N$，对每个 i，将其右边所有小于 π_i 的元素数求和即可．

定理 1

交换排列中任意两个不同元素，逆序数奇偶性改变．

[9]参考 Wikipedia 相关页面．

证明

在排列 p_n 中，考虑第 i,j（$i<j$）两个元素的置换，不妨把他们的值分别记为 m,n．不妨假设 $m<n$（$n<m$ 的讨论同理）．为了方便讨论，我们把任意两个不同元素都用弧线连起来，这些弧线有的代表逆序对，有的不是．我们可以把这些连线分为 4 类来讨论：

1. 对于不连接到 m 或 n 的线，置换不改变它们之中的逆序数．
2. 对于从 m 或 n 连接到 m 左边或 n 右边的线，置换同样不改变他们之中的逆序数．
3. 若 m,n 不相邻，对于从 m 或 n 连接到 m,n 之间某个元素 q 的两条线，只有当 $m<q<n$ 时逆序数会改变：从两个非逆序对变为逆序对．但无论有多少满足条件的 q，逆序数只会增加偶数个．
4. m,n 在置换前不是逆序对，置换后变为逆序对，使逆序数增加 1，改变逆序数的奇偶性．

综上所述，逆序对在置换后必定改变奇数个，所以逆序数奇偶性发生变化．证毕．

推论 1

把奇排列变为标准排列需要置换奇数次，把偶排列变为标准排列需要置换偶数次．

证明显然．

列维—奇维塔符号

预备知识 逆序数[430]

[10]**列维—奇维塔符号（Levi-Civita symbol，简称 LC 符号）**是一个函数，记为 $\epsilon_{i_1,i_2,\ldots,i_N}$．$N$ 叫做它的**维数（dimension）**．它的自变量是 N 个正整数 $i_1,\ldots,i_N$，

[10]本文参考 Wikipedia 相关页面．

可以称为角标．每个角标从 $1,2,\ldots,N$ 中取值．函数值只能取 $0,1,-1$ 中的一个．当 $i_1,\ldots,i_N$ 中有任意两个重复时函数值为 0；若没有重复，则函数值为 $(-1)^{N_p}$，N_p 为排列 $i_1,\ldots,i_N$ 的逆序数[430]

综上，N 维 LC 符号的角标共有 N^N 种不同可能，而使函数值不为零的有 $N!$ 种，其中函数值为 ±1 的各占一半．

三阶 LC 符号

三阶 LC 符号是矢量分析中常用的，可以直接用穷举法来定义．在不至于混淆的情况下我们可以把角标之间的逗号省略．

$$\epsilon_{123}=\epsilon_{231}=\epsilon_{312}=1 \tag{1}$$

$$\epsilon_{321}=\epsilon_{213}=\epsilon_{132}=-1 \tag{2}$$

当 i,j,k （只能取 1,2,3）中任意两个重复时，$\epsilon_{ijk}=0$.

三阶行列式可以记为

$$\det A=\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\epsilon_{i,j,k}a_{1,i}a_{2,j}a_{3,k} \tag{3}$$

所以两个几何矢量的叉乘[387] 可以记为

$$\mathbf{u}\times\mathbf{v}=\sum_{i,j,k}\epsilon_{i,j,k}u_jv_k\hat{\mathbf{e}}_i \tag{4}$$

行列式

预备知识 矩阵[396]，线性相关性[378]

[11]**行列式（determinant）**是线性代数中的一个重要工具，常用于判断线性相关性或解线性方程组（见“克拉默法则[413]”）．一个 $N\times N$ 的方阵[396] $\mathbf{A}$ 的行列式叫做 N 阶行列式，记为 $|\mathbf{A}|$，矩阵元可以是实数或者复数．行列式的运算的结果是一个数，若结果不为零，则所有列矢量（或行矢量）线性无关，否

[11]本文参考 Wikipedia 相关页面 以及 [4].

则线性相关．物理中经常出现二阶和三阶行列式，我们先介绍它们的性质，然后介绍高阶的情况．

令二阶方阵 $\mathbf{A}$ 第 i 行第 j 列的矩阵元为 $a_{i,j}$，其行列式为

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \tag{1}$$

三维方阵 $\mathbf{A}$ 的行列式为[12]

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{pmatrix} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{pmatrix} \tag{2}$$

行列式也可以记为 $\det\{\mathbf{A}\}$．

几何理解

预备知识 右手定则[20]

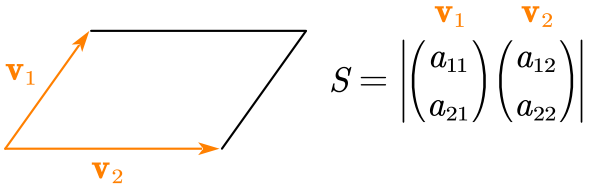


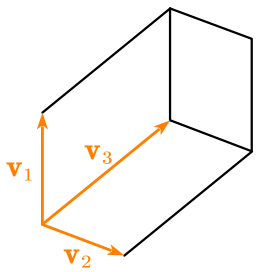
图 1: 二阶行列式的绝对值对应平行四边形的面积

在集合理解中，我们假设矩阵元都是实数．如图 1，二阶行列式的绝对值对应平行四边形的面积（可以认为是二维空间中的“体积”），若把行列式的两列看成两个几何矢量的坐标，他们就是平行四边形的两条边．当 $\mathbf{v}_1$ 逆时针转动[13]得到 $\mathbf{v}_2$ 时，行列式的值为正，反之为负．

如图 2，二阶行列式代表一个平行四面体的体积，若把行列式的 3 列看成 3 个几何矢量的坐标，他们就是平行四面体的 3 条边．若 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 的位置关系与 x, y, z 轴的关系相似（符合右手定则[20]）则行列式的值为正，反之为负．

[12]括号中的内容是一行

[13]假设转动角度小于 180°，下同．



$$V = \left| \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}_{\mathbf{v}_1} \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}_{\mathbf{v}_2} \begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix}_{\mathbf{v}_3} \right|$$

图 2: 三阶行列式对应平行四面体的体积

习题 1

请证明二阶行列式对应平行四边形的面积，三阶行列式对应平行四面体的体积

以上两个结论也可以拓展到更高维的情况，即 N 阶行列式表示 N 维空间中平行体的体积[438].

另外，以后我们会看到行列式转置（即 a_{ij} 变为 a_{ji}）以后不影响它的值（定理 8 [435]），所以我们也可以把每行看成是平行体一条边的坐标.

由行列式的几何意义，我们很容易理解为什么行列式的值为零当且仅当行矢量或列矢量线性相关：例如三维的情况下，$\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 线性相关意味着三矢量共面，体积必为零.

行列式的一般定义

预备知识 逆序数[430]

N 阶行列式（N 为正整数[14]）共有 $N!$ 项，每一项都是 N 个矩阵元的乘积[15]. 每一项中的 N 个矩阵元的行标和列标各不相同，我们既可以在每一项中按照行标来排序，也可以按照列标，我们选用前者. 排序后，行列式展开后的任意一项可记为（先不考虑前面的 $\pm$ 号）

$$\prod_{i=1}^{N} a_{i,\pi_{n,i}} = a_{1,\pi_{n,1}} \cdot a_{2,\pi_{n,2}} \cdots \tag{3}$$

[14] 一阶行列式定义为 $|a_{11}| = a_{11}$，虽然几乎从不被使用

[15] 高阶行列式的计算较为复杂，可通过数学软件计算，详见 Matlab，Mathematica 和 Wolfram Alpha 的计算方法.

其中列标 $\pi_{n,i}$ 是集合 $1,2\dots N$ 排列[106] 后的第 i 个数，显然该数列共有 $N!$ 种不同的排列，这里用 n 表示第 n 种排列，也表示行列式展开的第 n 项.

现在来考虑式 3 前面的 $\pm$ 号. 这由 π_n 的逆序数[430] r 决定. N 阶行列式可以用逆序数定义为

$$\det A=\sum_{n=1}^{N!}(-1)^r\prod_{i=1}^{N}a_{i,\pi_{n,i}} \tag{4}$$

Levi-Civita 符号

$$\det A=\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\cdots\sum_{i_N=1}^{N}\epsilon_{i_1,\dots,i_N}a_{1,i_1}a_{2,i_2}\dots a_{N,i_N} \tag{5}$$

证明与式 4 等效：该式共有 N^N 项求和，根据 LC 符号的定义，$i_1,\dots,i_N$ 中任意两个角标相同时该项为零，所以不为零的项中 $i_1,\dots,i_N$ 必为 π_n 的一种，仍然只有 $N!$ 种不同的可能，这时 $\epsilon_{i_1,\dots,i_N}=(-1)^r$.

行列式的性质

预备知识 行列式[432]

以下是行列式常见的性质，我们从定义出发容易一一证明. 由于行列式的绝对值表示体积[438]，我们也可以从几何上理解这些定理. 这些定理对复数元素的行列式同样适用，只是复数行列式不再具有直观的几何意义.

定理 1

若行列式中某行或某列全为 0，其结果等于 0.

证明：根据定义（式 4 [432]），行列式展开后，相加的每一项都含有每一行（或每一列）的一个元素，所以当某列全为 0，则结果为 0.

几何理解：若平行体的某条边长等于 0，其体积也等于 0.

定理 2

矩阵的任意一列（或任意一行）乘以常数，行列式的值也要乘以该常数.

证明：思路和定理 1 的证明一样，行列式展开后，相加的每一项都含有每一行（或每一列）的一个元素，所以当某列全部元素乘以常数，则相加的每一项都乘以该常数.

几何理解：将平行体任意一条边长乘以一个常数，它的体积也需要乘以该常数.

定理 3

将行列式的两列交换，结果取相反数.

证明：根据行列式的定义，交换两行或两列会给展开后的每一项增加一个逆序数，导致行列式的值取相反数.

定理 4

若将行列式的某行或某行的每一个元素都表示为两个数之和，那么它就可以表示为两个行列式相加：

$$\begin{aligned}&\begin{vmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,N} \\ \dots & \dots & \dots & \dots \\ b_{i,1}+c_{i,1} & b_{i,2}+c_{i,2} & \dots & b_{i,N}+c_{i,N} \\ \dots & \dots & \dots & \dots \\ a_{N,1} & a_{N,2} & \dots & a_{N,N} \end{vmatrix} \\ =&\begin{vmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,N} \\ \dots & \dots & \dots & \dots \\ b_{i,1} & b_{i,2} & \dots & b_{i,N} \\ \dots & \dots & \dots & \dots \\ a_{N,1} & a_{N,2} & \dots & a_{N,N} \end{vmatrix} + \begin{vmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,N} \\ \dots & \dots & \dots & \dots \\ c_{i,1} & c_{i,2} & \dots & c_{i,N} \\ \dots & \dots & \dots & \dots \\ a_{N,1} & a_{N,2} & \dots & a_{N,N} \end{vmatrix}\end{aligned} \tag{1}$$

列的情况也同理.

证明：行列式展开后，相加的每一项都含有每一行（或每一列）的一个元素，所以显然成立.

几何理解：若把平行体的一条边变为两条折线，那么可以作出两个叠加的平行体，体积之和等于原来的体积.

定理 5

把矩阵的第 i 行（列）叠加上“第 j 行（列）乘任意常数 λ”，行列式的值

不变.

证明：根据定理 4 和定理 2，我们可以把这样操作后的行列式拆分为两个行列式，第一个与原来相同；第二个在原来的基础上把第 i 行替换为第 j 行，再把行列式的值乘以 λ. 而第二个行列式由于存在两行相同，结果为 0.

几何理解：以平行四边形为例，由于其体积是底乘以高，令 $\mathbf{v}_1$ 为底，$\mathbf{v}_2$ 在垂直于 $\mathbf{v}_1$ 方向的投影为高，则将 $\mathbf{v}_2$ 变为 $\mathbf{v}_2 + \lambda\mathbf{v}_1$（$\lambda$ 为常数）后高不变，所以面积不变. 高维情况同理.

定理 6

若行列式中的行矢或列线性相关，行列式的值为 0.

线性相关意味着存在某一行 i，可以表示为其他行的线性组合. 那么我们可以通过将这些其他的行依次乘以常数，然后加到第 i 行上，使得第 i 行全为 0. 根据定理 5，这么做不改变行列式的值，而根据式 1，行列式的值为 0.

几何理解：二维情况下两矢量线性相关意味着他们共线，平行四边形面积为 0. 三维情况下线性相关意味着三个矢量共面，平行四面体体积为 0. 高维情况也可类比.

定理 7

行列式的值为 0 当且仅当行列式中存在线性相关的列（行）.

定理 8

矩阵转置（将所有 $a_{i,j}$ 与 $a_{j,i}$ 交换[396]）后行其列式的值不变.

这个定理没有显然的几何理解，可以直接用行列式的定义证明（式 4 [432]）. 根据这个定理，以上凡是涉及到“行”的定理和说明，都可以替换为“列”，请读者自行回顾一次.

拓展

行列式的代数性质可以抽象为**外代数**（见例 6 [590] 以及例子后的解释），进而用于定义**外微分**[604]，描述微分形式的代数性质.

行列式与体积

预备知识 行列式的性质[435]

在“行列式[432]”中我们看到了二阶和三阶行列式从几何上分别对应平行四边形的面积（即二维体积）和平行六面体的体积．我们现在来证明 $N>0$ 维空间的情况．

我们先来看最简单的例子：一个对角线[16]元素都为 1，其他元素为零的行列式

$$\begin{vmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{vmatrix} = 1 \tag{1}$$

这代表 N 维空间中边长都是 1 的立方体的体积．

N 维空间中的任意平行体都可以由该立方体经过两个操作（包括任意次序多次操作）得到．一个是将某条边长乘以常数 λ（可以为负），第二个是将某条边所在的矢量乘以常数 λ 然后加到另一条边上．前者使体积乘以 $|\lambda|$，后者保持体积不变．

如果我们把行列式的每一行（或每一列，下同）对应到平行体的每条边，那么根据定理 2 [435] 和定理 5 [435]，前者同样使行列式的值乘以常数 λ，后者同样保持行列式的值不变．从立方体开始所以经过一系列变换后，行列式的绝对值同样表示对应平行体的体积．

克莱姆法则

预备知识 矩阵[396]

克莱姆法则（也译作克拉默法则（Kramer’s rule）可以直接用行列式表示

[16]行列式或矩阵的对角线特指所有行标和列标相同的元素

出以下线性方程组的各个解．将线性方程组记为

$$\mathbf{A}\mathbf{x}=\mathbf{y} \tag{1}$$

其中 $\mathbf{A}$ 是已知的 N 行 N 列方阵，$\mathbf{y}$ 是已知的 N 维列矢量，$\mathbf{x}$ 是未知的 N 维列矢量．当系数行列式 $|\mathbf{A}|$ 不为零时，方程有唯一解．$\mathbf{x}$ 的各个元为

$$x_i=\frac{|\mathbf{B}_i|}{|\mathbf{A}|} \tag{2}$$

其中矩阵 $\mathbf{B}_i$ 是把矩阵 $\mathbf{A}$ 的第 i 列替换为 $\mathbf{y}$ 得到的矩阵．

海森矩阵

一个二阶可导的多元函数 $f(\mathbf{x})$ 的**海森矩阵（Hessian）** $\mathbf{H}$ 定义为

$$H_{ij}=\frac{\partial^2 f}{\partial x_i\partial x_j} \tag{1}$$

$f(\mathbf{x})$ 的泰勒展开[338] 的前两项可以用梯度矢量和海森矩阵表示为

$$f(\mathbf{x})=f(\mathbf{x}_0)+(\mathbf{x}-\mathbf{x}_0)\boldsymbol{\nabla}f(\mathbf{x}_0)+\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)\mathbf{H}(\mathbf{x}-\mathbf{x}_0)+\mathcal{O}\big((x-x_0)^3\big) \tag{2}$$

海森矩阵可以看做梯度矢量 $\boldsymbol{\nabla}f$ 的雅可比矩阵，即

$$H_{ij}=\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) \tag{3}$$

所以有

$$\mathrm{d}(\boldsymbol{\nabla}f)=\mathbf{H}\,\mathrm{d}\mathbf{x} \tag{4}$$

如果 $f(\mathbf{x})$ 是一个二阶函数，海森矩将不随 $\mathbf{x}$ 变化．所以有

$$\boldsymbol{\nabla}f(\mathbf{x})-\boldsymbol{\nabla}f(\mathbf{x}_0)=\mathbf{H}(\mathbf{x}-\mathbf{x}_0) \tag{5}$$

则函数的极值点为（令 $\boldsymbol{\nabla}f=0$）

$$\mathbf{x}=\mathbf{x}_0-\mathbf{H}^{-1}\boldsymbol{\nabla}f(\mathbf{x}_0) \tag{6}$$

这就是牛顿法寻找函数极小值的主要思路．

正交矩阵酉矩阵

- 本词条处于草稿阶段.
- 需要整合

预备知识 正交矩阵

酉矩阵（unitary matrix）也叫**幺正矩阵**. 当矩阵元为实数时也叫**正交矩阵（orthogonal matrix）**，是正交矩阵的复数拓展，即矩阵元可以是复数. 酉矩阵 $\mathbf{U}$ 的定义同样为

$$\sum_k U_{ki}^* U_{kj} = \delta_{ij} \tag{1}$$

但由于复数的列矢量没有对应的几何矢量[370]，所以这里的正交完全是复线性空间中的广义正交.

单位正交矩阵的回收内容

预备知识 正交归一基底[386]，逆矩阵[405]

若一个实数方阵的每一列的模长都等于 1，且任意两列都正交，那么这个矩阵就是一个**单位正交阵**. 若 $\mathbf{U}$ 为单位正交阵，则其矩阵元满足

$$\sum_k U_{ki} U_{kj} = \delta_{ij} \tag{2}$$

其中 δ_{ij} 是克罗内克 δ 函数. 所以若把 N 阶单位正交阵的每一列看做 N 维空间中的一个单位矢量的直角坐标，那么这些单位矢量就组成该空间的一组正交归一基底.

式 2 也可以用矩阵转置和矩阵乘法表示为

$$\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I} \tag{3}$$

其中 $\mathbf{I}$ 是单位矩阵. 根据逆矩阵的定义，我们得到单位正交矩阵的一个重要性质，即其逆矩阵等于转置矩阵.

$$\mathbf{U}^{-1} = \mathbf{U}^{\mathrm{T}} \tag{4}$$

由逆矩阵的性质（式 2 [405]）得

$$\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{U}\mathbf{U}^{-1} = \mathbf{I} \tag{5}$$

表示为矩阵元的运算就是

$$\sum_k U_{ik}U_{jk} = \delta_{ij} \tag{6}$$

所以单位正交矩阵的所有行同样正交归一．易证式 2 和式 6 互为充分必要条件，都可以作为单位正交阵的定义．

几何理解

为了更形象地理解单位正交阵的上述性质，我们以二维几何矢量和二维实数方阵为例讨论，三维的情况同理可得．

任意的二维实数方阵可以看做是两个单位几何矢量 $\hat{\mathbf{u}},\hat{\mathbf{v}}$ 在正交归一基底 $\hat{\mathbf{x}},\hat{\mathbf{y}}$ 上的坐标（即内积）．

$$\mathbf{U} = \begin{pmatrix} \hat{\mathbf{u}}\cdot\hat{\mathbf{x}} & \hat{\mathbf{v}}\cdot\hat{\mathbf{x}} \\ \hat{\mathbf{u}}\cdot\hat{\mathbf{y}} & \hat{\mathbf{v}}\cdot\hat{\mathbf{y}} \end{pmatrix} \tag{7}$$

显然矩阵的两个列矢量满足正交归一．由内积的交换律，我们同样可以把矩阵的两行分别看做单位矢量 $\hat{\mathbf{x}},\hat{\mathbf{y}}$ 在正交归一基底 $\hat{\mathbf{u}},\hat{\mathbf{v}}$ 上的坐标，所以矩阵的两个行矢量同样正交归一．

现在我们令任意几何矢量以 $\hat{\mathbf{x}},\hat{\mathbf{y}}$ 为基底的坐标为 (x,y)，以 $\hat{\mathbf{u}},\hat{\mathbf{v}}$ 为基底的坐标为 (u,v)，即

$$\hat{\mathbf{v}} = u\hat{\mathbf{u}} + v\hat{\mathbf{v}} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} \tag{8}$$

由正交归一基底的性质，

$$\begin{cases} x = (u\hat{\mathbf{u}} + v\hat{\mathbf{v}})\cdot\hat{\mathbf{x}} \\ y = (u\hat{\mathbf{u}} + v\hat{\mathbf{v}})\cdot\hat{\mathbf{y}} \end{cases} \qquad \begin{cases} u = (x\hat{\mathbf{x}} + y\hat{\mathbf{y}})\cdot\hat{\mathbf{u}} \\ v = (x\hat{\mathbf{x}} + y\hat{\mathbf{y}})\cdot\hat{\mathbf{v}} \end{cases} \tag{9}$$

写成矩阵的形式，就是

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{u}}\cdot\hat{\mathbf{x}} & \hat{\mathbf{v}}\cdot\hat{\mathbf{x}} \\ \hat{\mathbf{u}}\cdot\hat{\mathbf{y}} & \hat{\mathbf{v}}\cdot\hat{\mathbf{y}} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \qquad \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{x}}\cdot\hat{\mathbf{u}} & \hat{\mathbf{y}}\cdot\hat{\mathbf{u}} \\ \hat{\mathbf{x}}\cdot\hat{\mathbf{v}} & \hat{\mathbf{y}}\cdot\hat{\mathbf{v}} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{10}$$

注意上式左边的矩阵是 $\mathbf{U}$，右边的矩阵是 $\mathbf{U}^{\mathrm{T}}$，而右边的变换是左边的逆变换，所以 $\mathbf{U}^{\mathrm{T}} = \mathbf{U}^{-1}$．

幺正变换

- 本词条处于草稿阶段.

预备知识 范数赋范空间[794]，内积内积空间[471]，幺正变换酉矩阵[440]

定义 1　幺正变换

对有限维赋范空间[794] 中的线性变换[451] $U:V\to V$，若对任意 $u,v\in V$ 都满足变换前后内积不变，即 $\langle Uu|Uv\rangle=\langle u|v\rangle$，那么 U 就称为一个（有限维的）幺正变换（**unitary transform**）.

定理 1

有限维赋范空间中的线性变换 $U:V\to V$ 是幺正变换当且仅当对任意 $v\in V$ 都有 $\|Uv\|=\|v\|$.

定理 2

有限维幺正变换 $U:V\to V$ 在 V 中一组正交归一基底 $\{e_i\}$ 下可以表示为酉矩阵[440] $\mathbf{U}$.

我们知道正交归一基底 $\{e_i\}$ 下，线性算符 $A:V\to V$ 可以表示为矩阵，矩阵元为（链接未完成）$A_{i,j}=\langle e_i|A|e_j\rangle$.

第三章

线性空间

矢量空间

预备知识　几何矢量[370]，函数[103]，抽象[102]

矢量空间（vector space）也叫**向量空间**或**线性空间（linear space）**，是一种满足一定条件的集合[92]，有无穷多个元素，每个元素叫做一个**矢量**或**向量（vector）**．它必须满足，在其中选择任意两个矢量，它们的线性组合仍然在这个空间中（以及一些其他条件）．进行归纳后易得，这个条件等价于“任意有限个矢量的线性组合仍然在这个空间中”（封闭性）．这里的“矢量”是一个广义的概念，是几何矢量[370]的抽象[102]；反过来，几何矢量是矢量的具象（特例）．一个广义的矢量，不一定具有长度和方向，例如下面会看到函数也可以看作矢量．

定义

矢量空间的定义必须依赖一个**域（field）**[588] $\mathbb{F}$．简单来说，域就是能进行加减乘除的对象的一个集合，比如实数域 $\mathbb{R}$ 和复数域 $\mathbb{C}$．这个域本身被称为该矢量空间的**标量域（scalar field）**或**标域**，它的元素被称为矢量空间的**标量（scalar）**，它们不是矢量空间的元素，但是可以用来和矢量进行数乘．那么，我们就说矢量空间是“域 $\mathbb{F}$ 上的”．矢量空间选择的域几乎都是 $\mathbb{R}$ 或 $\mathbb{C}$，即讨论的是**实数（或复数）域上的矢量空间**．所以即使我们不了解域的具体定义也没关系，只要知道实数和复数是两种常见的域就足够了．

定义 1　矢量空间

标量域 $\mathbb{F}$ 上的矢量空间定义了矢量的集合 X、两个矢量间的加法运算 $X\times X\to X$ 以及标量和矢量间的数乘运算 $\mathbb{F}\times X\to\mathbb{X}$．令矢量 $u,v,w\in X$，标量 $a,b\in\mathbb{F}$，矢量加法记为 $u+v$，数乘记为 au．两种运算必须满足以下性质：

加法运算

1. 满足加法交换律 $u+v=v+u$．
2. 满足加法结合律 $(u+v)+w=u+(v+w)$．

3. 存在零矢量，使得 $v+0=v$.

4. 空间中任意矢量 v 存在逆矢量 $-v$，使得 $v+(-v)=0$.

数乘运算

1. 乘法分配律 $a(u+v)=au+av$

2. 乘法分配律 $(a+b)v=av+bv$

3. 乘法结合律 $a(bv)=(ab)v$

说明：加法运算 $X\times X\to X$ 是一个二元映射（第 2 节 [98]），注意运算的结果必须仍然落在 X 中．我们把这样的运算叫做**封闭（closed）**的[1]．数乘运算同样也是封闭的，即一个矢量数乘标量后仍然落在 X 中．我们现在还没有涉及 X 以外的元素，所以封闭性看似有些多余，但以后会看到一个矢量空间 X 的子集 X_1 也可以是矢量空间，称为子空间[464]，X_1 上的两种运算封闭意味着运算结果只能落在 X_1 中而不能是 X 的其他元素．

习题 1　几何矢量

证明 1,2,3 维空间中的所有几何矢量[370] 各自构成实数域 $\mathbb{R}$ 上的矢量空间．

作为一个非几何矢量的例子，我们来看由多项式构成矢量空间．

例 1　多项式

所有不大于 n 阶的多项式 $c_nx^n+c_{n-1}x^{n-1}+\cdots+c_1x+c_0$ 可以构成一个实数或复数矢量空间．定义矢量加法为两多项式相加，满足

- 封闭性：任意两个不大于 n 阶的多项式相加仍然为不大于 n 阶的多项式．
- 交换律：多项式相加显然满足交换律．
- 零矢量：常数 0 可以看做一个 0 阶多项式，任何多项式与之相加都不改变．
- 逆矢量：把任意多项式乘以 -1 就得到它的逆矢量，任意多项式与其逆矢量相加等于 0．

[1] 一些文献中也叫“闭合”

定义矢量数乘为多项式乘以常数，显然也满足数乘的各项要求，不再赘述.

另一个重要的矢量空间，是**函数空间**（**function space**）.

例 2　函数空间

实数到实数的全体函数（$f:\mathbb{R}\to\mathbb{R}$）的集合 F 构成一个线性空间，称为 $\mathbb{R}$ 函数空间. 函数空间中两个向量的加法定义为，对于任何实数 x 和函数（即向量）$f,g\in F$，有 $(f+g)(x)=f(x)+g(x)$；数乘定义为，对于任何实数 a,x 和函数 $f\in F$，有 $(af)(x)=af(x)$.

类似地，复数到复数、实数到复数等的函数都可以构成线性空间；把函数限制在连续函数、可导函数等条件下也依然构成线性空间. 特别地，复数域上的归一化可导函数，构成了复数域上的希尔伯特空间，这是一种无穷维的特殊矢量空间，是量子力学的基础概念，我们将会在将来详细讨论.

注意矢量空间的定义并不需要包含内积（点乘）的概念，但我们可以在其基础上额外定义内积，这样的空间叫做**内积空间**[471]，留到以后介绍. 除了内积，我们可以把“几何矢量的运算[375]”和“线性相关性[378]”中介绍的概念都拓展到一般的矢量空间中，这里不再重复.

习题 2　复数列矢量

我们把 N 个复数 $c_1,\dots,c_N$ 按顺序排成一列（或一行，下同），叫做列矢量（或行矢量，下同）. 列矢量可以看成是 $N\times 1$ 的矩阵[396]. 给它们定义通常意义的加法和数乘运算，这样所有列矢量可以构成一个 N 维矢量空间. 注意由于我们使用了复数，即使 $N\leqslant 3$ 时我们也无法将这些矢量与几何矢量对应起来.

如果我们将基底取为[2] $(1,0,\dots,0)^{\mathrm{T}}$，$(0,1,0,\dots,0)^{\mathrm{T}}$，…，$(0,\dots,0,1)^{\mathrm{T}}$，那么显然任意列矢量 $(c_1,\dots,c_N)^{\mathrm{T}}$ 的坐标就是有序实数 $c_1,\dots,c_N$. 但我们也可以取其他基底，这时坐标就会改变. 所以再次强调坐标和矢量本身是不同的. 我们将会在矢量空间的表示[458] 中详细区分矢量本身和矢量的坐标这两个概念.

习题 3

证明例 1 中多项式空间是 $n+1$ 维空间，x^k （$k=0,\dots,n$）是一组基底（提示：证明它们线性无关，可以表示空间中的任意矢量）.

[2] 上标 T 表示转置，这里是为了排版方便

张成空间

预备知识 线性相关性[378]

矢量的线性组合，使得我们有可能用少量矢量来表示更多的矢量．例如 N 维几何矢量空间[370] 空间中的一组基底可以通过线性组合得到整个矢量空间中的任意矢量，我们就说这组基底**张成（span）**了这个矢量空间．

在本节中，我们讨论如何用更少的矢量来表示整个空间里的所有矢量，并在此基础上拓展，引出线性空间的基以及维度等概念．

张成空间

定义 1　张成空间

给定线性空间 V，用 V 的若干矢量构成的集合 $S=\{\mathbf{v}_\alpha\}$，可以是无穷集合，可以线性相关，也可以有限或者线性无关．从 S 中任意地选择有限个矢量进行线性组合所得到的集合 $\{\sum_i^N c_i\mathbf{v_i}|N\in\mathbb{Z},c_i\in\mathbb{R},\mathbf{v}_i\}$ 称为 S 所张成（**span**）的空间，记为 $\operatorname{span}S$ 或者 $\langle\{\mathbf{v}_\alpha\}\rangle$．

容易证明，定义中的 $\operatorname{span}S$ 符合矢量空间[444] 的定义．

张成空间的概念，最直观的例子就是子空间．比如说，三维的几何矢量空间 V 中，任意选定两个向量 $\mathbf{v}$ 和 $\mathbf{u}$，这两个向量的所有线性组合构成的集合，是 V 中的一个过原点的平面．这样，$\operatorname{span}\{\mathbf{v},\mathbf{u}\}$ 构成了一个线性空间，同时还是 V 的子集，因此又被称作 V 的一个子空间．

习题 1

给定一个三维线性空间，把其中的每个向量表示成一个行向量 (x,y,z)，其中 $x,y,z\in\mathbb{R}$．取三个向量 $\mathbf{v}_1=(1,2,0)$，$\mathbf{v}_2=(3,2,4)$ 和 $\mathbf{v}_3=(1,1,1)$，那么这三个向量所张成的空间是一个点、一条线、一个平面还是整个三维空间本身？三个向量是否线性相关？它们的线性相关性和张成空间的样子有什么关系？

实际上，我们完全可以摆脱“用给定空间的向量来张成一个子空间”，而形式化地定义张成的概念．取任何一个集合 S，我们可以用 S 在域 $\mathbb{K}$ 上直接构造一个线性空间 $V=\{\sum a_\alpha s_\alpha|a_\alpha\in\mathbb{K},s_\alpha\in S\}$．这里的 $a_\alpha s_\alpha$ 表示把数字 a_α

和元素 s_α 组合在一起，变成一个不属于 S 的新元素[3]，而 $\sum$ 表示这些新元素组合起来又得到了新的元素，组合符号用加号 $+$ 表示．这时回过头来，我们可以把每个 s_α 看成是 V 的一个向量，它们张成了 V 这个线性空间．

基和基向量

我们用向量来张成空间，就是在用这些向量来表示张成空间里面的所有向量的．从线性相关性[378] 中我们知道，张成空间时所使用的向量，如果它们线性相关，那么就会有冗余向量，也就是说其中一些向量本身就可以被剩下的向量表示出来，因此我们就算把冗余的向量都剔除了，剩余的向量仍然可以张成同一个空间．

进一步，有冗余向量就意味着，张成空间里的每一个向量都可以有多个表示方法．

习题 2

习题 1 中所给出的三个向量是线性相关的．请使用这个例子来讨论以下两个问题：

- 这三个向量中剔除几个向量后就不再有冗余向量？
- 剔除的冗余向量有几种选择？
- 用这三个向量的线性组合来表示向量 $(1,0,0)$，组合方式是唯一的吗？

如果一组向量是线性无关的，那么就没有这讨厌的冗余向量了，我们就可以很方便地用这组向量来讨论张成空间的性质．

定义 2　基和基向量

如果线性空间 V 可以表示为一组向量 $\{\mathbf{v}_i\}$ 的张成空间，并且这组向量是线性无关的，那么我们说 $\{\mathbf{v}_i\}$ 是 V 的一组基底（**basis**）或简称基，各 $\mathbf{v}_i$ 被称作一个基向量（**basis vector**）．

有了基之后，线性空间中的**每一个**向量都可以**唯一**地表示成基向量的线性组合．这里，每一个向量都可以被表示成基向量的线性组合是因为定义中要求基底能张成整个线性空间，而线性组合的唯一性是因为定义中要求基底是线性

[3] 只有一个情况例外，那就是当 $a_\alpha = 1$ 时，将 $1s_\alpha$ 认为就是 s_α 本身，从而这时并没有得到新元素．

无关的．这是极为重要的性质，它使得线性空间上的一切线性性质，比如线性函数、线性变换、张量的表示等等，都只和基底的选择有关．比如说，对于线性函数，一旦选定了基底，那么只需要计算出基向量的函数值，我们就可以得到一切向量的函数值，而不用对每一个向量都作一番计算．

定理 1　基底表示的唯一性

对于线性空间 V，给定一组基 $\{\mathbf{e}_i\}$．对于任何向量 $\mathbf{v} \in V$，它表示为该基底的线性组合的方式是唯一的．

证明：

设 $\mathbf{v} = \sum a_i \mathbf{e}_i = \sum b_i \mathbf{e}_i$，其中 $a_i, b_i \in \mathbb{K}$，而 $\mathbb{K}$ 是定义 V 所用的域．那么我们有：

$$\sum (a_i - b_i)\mathbf{e}_i = \sum a_i \mathbf{e}_i - \sum b_i \mathbf{e}_i = \mathbf{v} - \mathbf{v} = \mathbf{0} \tag{1}$$

由于 $\{\mathbf{e}_i\}$ 线性无关，上式就意味着每一个 $a_i - b_i$ 都是 0，也就是说每一个 a_i 都等于 b_i．因此不可能有两种不同的组合方式能得到 $\mathbf{v}$．

证毕．

定义 3　向量的坐标

对于线性空间 V，给定一组基 $\{\mathbf{e}_i\}$．如果向量 $\mathbf{v} = \sum a_i \mathbf{e}_i$，那么我们称 $\{a_i\}$ 是 $\mathbf{v}$ 的一组坐标（**coordinates**）．

定理 1 意味着选定基底以后，各向量的坐标是唯一的，但基底的选择本身不是唯一的．比如说，二维平面上，任何两个不平行的向量都可以构成这个空间的一个基．基的选择不一样，向量的坐标也不一样．对此的详细讨论请参见**矢量空间的表示**[458]．

线性空间的维度

对于一组有限个向量，如果它们线性相关，就总可以找出一个冗余向量，把它剔除；如果剔除一个冗余向量以后还有冗余向量，就重复这个操作，直到不再有冗余向量为止．这样剔除若干步后所得到的线性无关的向量组，被称为原先向量组的一个**极大线性无关组**．由于每一步剔除的冗余向量可以不同，最终剩下来的极大线性无关组也可以不同．

例 1

依然利用习题 1 的例子．向量组 $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ 的极大线性无关组一共有三个，分别是 $\{\mathbf{v}_1, \mathbf{v}_2\}$，$\{\mathbf{v}_2, \mathbf{v}_3\}$ 和 $\{\mathbf{v}_1, \mathbf{v}_3\}$．

由于每一步剔除的向量都是“冗余”的，因此每一步的剔除都不会改变向量组所张成的空间．尽管剔除的方式不同可能导致剩下来的极大线性无关组不同，但是它们都张成同一个空间，这就意味着这些极大线性无关组一定有某些共性，我们把这个共性称作张成空间的维度．

定理 2

如果线性空间 V 可以表示为一组有限个向量 $\{\mathbf{w}_i\}$ 的张成空间，那么 $\{\mathbf{w}_i\}$ 的任何两个极大线性无关组中的元素数量都是一样的．

定理 2 的证明有很多方式，由于本书的教学思路在这里还没引入线性方程组的概念，因此不会使用线性方程组解的性质来证明，而是用“替换法”．

证明：

假设 S 是若干向量的集合，其中 $\{\mathbf{v}_i\}_{i=1}^n$ 和 $\{\mathbf{u}_j\}_{j=1}^m$ 都是 S 的极大线性无关组．那么由于“极大”，$\{\mathbf{v}_i\}_{i=1}^n$ 中的向量都可以用 $\{\mathbf{u}_j\}_{j=1}^m$ 的线性组合来表示，反之亦然；由于“线性无关”，两者的互相表示是唯一的．

反设 $n \neq m$，不妨就设 $n > m$．

如果从 $\{\mathbf{v}_i\}_{i=1}^n$ 中把 $\mathbf{v}_1$ 拿掉，那么必然存在若干 $\mathbf{u}_j$ 无法被剩下的 $\{\mathbf{v}_i\}_{i=2}^n$ 表示．我们把这些无法被表示的 $\mathbf{u}_j$ 添加进 $\{\mathbf{v}_i\}_{i=2}^n$，取代 $\mathbf{v}_1$，这样又得到一个新的极大线性无关组[4]．对于把 $\mathbf{v}_1$ 替换掉后的新极大无关组，再把 $\mathbf{v}_2$ 拿掉，然后相应地替换为拿掉之后无法表示出的所有 $\mathbf{u}_j$．依次进行．

以上步骤中，由于每一次替换都只拿掉一个 $\mathbf{v}_i$，却要替换为**至少**一个 $\mathbf{u}_j$，而且每次添加的 $\mathbf{u}_j$ 都和之前的没有交集[5]．再加上 $n > m$，因此替换的元素最多到 $\mathbf{v}_m$．这样就得出 $\{\mathbf{u}_j\}_{j=1}^m \cup \{\mathbf{v}_i\}_{j=m+1}^n$ 也是一个极大无关组．由于这个极大无关组比 $\{\mathbf{u}_j\}_{j=1}^m$ 多了一个非空集合 $\{\mathbf{v}_i\}_{j=m+1}^n$，因此和 $\{\mathbf{u}_j\}_{j=1}^m$ 的极大线性无关性**矛盾**．因此假设不成立，即 $n = m$．

令 $S = \{\mathbf{w}_i\}$ 或者 $S = V$，就得到定理的证明．

证毕．

替换法的证明仅仅用到了目前已知的少量向量性质，并不涉及方程组理

[4]新向量组**线性无关**是因为所添加的元素都是 $\{\mathbf{v}_i\}_{i=2}^n$ 所无法表示、本身也线性无关的，而**极大**是因为它可以表出所有 $\mathbf{u}_j$ 了．

[5]因为之前添加的 $\mathbf{u}_j$ 已经在新极大无关组里了，肯定能被表示出来．

论，请仔细体会其逻辑．事实上，本书将会在建立了线性空间的直觉后，用线性空间的思想，即几何的思想，来解释实系数线性方程组的概念．

定理 2 意味着，线性空间的不同基底总包含相同数量的基向量．实际上蕴含的是非常符合直觉的一个概念，维度．

定义 4　线性空间的维度

对于线性空间 V，其任意一个极大线性无关组的元素数量，称为 V 的维度（**dimension**），记为 $\dim V$．

在定理 2 中我们限制了 V 被有限个向量张成，也就是说 $\dim V$ 是有限的．事实上，我们通常讨论的线性空间是有限维的，无穷维的情况会在将来进一步讨论．

线性映射

- 重整中：线性变换和线性映射改为同义词

预备知识　子空间[464]

线性映射

定义 1　线性映射

给定域[588] $\mathbb{F}$（通常取实数域 $\mathbb{R}$ 或复数域 $\mathbb{C}$）上的线性空间 V 和 U．如果有映射[98] $f : V \to U$ 满足：对于任意的向量 $v_1, v_2 \in V$ 和标量 $c_1, c_2 \in \mathbb{F}$，都有

$$f(c_1 v_1 + c_2 v_2) = c_1 f(v_1) + c_2 f(v_2) \tag{1}$$

则称 f 是 V 到 U 的一个线性映射（**linear map**），也叫线性变换（**linear transform**），或者线性算符（**linear operator**）．

一些教材中只把定义域和值域相同的线性映射 $f : V \to V$ 称为线性变换，本书中不做这种区分．

定义 1 的内涵比看上去广一些，不仅仅是对 V 中两个向量的线性组合成立. 容易证明对于任意的一组有限个向量 $\{v_i\} \subseteq V$ 和一组对应的标量 $\{c_i\} \subseteq \mathbb{F}$，都有

$$f\left(\sum_i c_i v_i\right) = \sum_i c_i f(v_i) \tag{2}$$

特殊地，若令所有 $c_i = 0$ 可得

推论 1

线性映射 $f : V \to U$ 总把 $0 \in V$ 映射到 $0 \in U$.

习题 1　几何矢量的线性变换

请证明 N 维几何线性空间 V_N 中，以下变换 $f : V \to V$ 都是线性变换：

- 缩放：把矢量乘以一个固定标量 λ.
- 旋转：把矢量绕任意固定轴逆时针旋转固定角度 θ.
- 镜像：把矢量关于某条过原点的固定直线做镜像对称.
- 投影：把二维矢量垂直投影到过原点的固定直线，或者把三维矢量垂直投影到过原点的平面（也可以拓展到 N 维空间中的矢量投影到 $N-1$ 维平面）.
- 以上几种变换的任意组合.

注意在以上映射中，除了投影外都是双射[98]，且都把零向量映射到零向量. 多对一映射也会把一些非零向量映射到零向量.

习题 2

请证明两个线性映射 f, g 的复合映射 $f \circ g$ 仍然是线性映射. 注意 f 的值域必须是 g 的定义域的子集.

在映射[98] 中，我们把一个映射的像定义为值域中所有被映射到的元素的集合.

定理 1　值空间

线性映射 $f : V \to U$ 的像 $M = f(V) \subseteq U$ 是一个线性空间，即 U 的子空

间[464]．称为 f 的值空间（**range space**）．为了方便我们也可以把该映射记为 $f : V \to M \subseteq U$．

证明留作习题，使用线性空间的定义[444] 即可．

未完成：这两段应该放到后面展开讲解，太浓缩

如果 $\{e_i\}_{i=1}^n$ 是 V 的一组基底，那么任意 $v \in V$ 都可以唯一地表示为 $v = \sum_i c_i e_i$ 的形式，其中 $c_i \in \mathbb{F}$．这样，由于线性性，$f(v) = \sum_i c_i f(e_i)$．也就是说，只需要知道了基向量被 f 映射到哪里，也就可以计算出任意向量映射到哪里．于是，和线性函数一样，确定一个线性映射的时候，我们最多只能自由选择基向量映射到哪里，只不过这里的函数值不再是数字，而是 U 中的向量．

在向量空间的表示[458] 中我们还会看到，选定两个空间的基以后，一个线性映射也可以看成是多个线性函数的排列，因此线性映射和线性函数性质很相似．线性函数是一种特殊的线性映射，V 上的所有线性函数组成的向量空间叫做**对偶空间**[460]．

线性变换

未完成：以下需要改：线性变换和映射是同义词！

要知道一个具体的线性变换的性质，我们只需要知道它把基向量都映射到哪里去了．我们知道，基向量张成线性空间 V 本身，而任意一组基向量变换后所得到的向量组，张成[447] 了线性变换的**像空间**（**image space**），即 V 中全体向量变换后的结果之集合．如果基向量变换后的向量组还是线性无关的，那么这个向量组也是一组基，从而像空间就是 V 本身；但有的时候变换后的向量组线性相关了，这时候的像空间就是 V 的一个真子集了．极端情况下甚至把所有向量都变换为零向量，这样的线性变换的像空间就只有零向量了．

如果一个线性变换的像空间是 V 的一个真子集，那么这个真子集一定是 V 中的一个“过原点的平面”，我们把这种线性变换称为**退化**（**degenerate**）的．反过来，V 中的一个“过原点的平面”可以是某线性变换的像空间．一个线性变换唯一对应一个像空间，但是一个像空间总是对应无穷多个线性变换．

如果 $v \in V$ 在基 $\{e_i\}$ 中表示为 $\sum c_i e_i$，那么它变换后的向量就是 $Tv = T\sum c_i e_i = \sum c_i T e_i$．如果向量组 $\{Te_i\}$ 是线性相关的，那么就意味着存在并非

全为零的 $\{c_i\}$ 使得 $Tv = \sum c_i Te_i = 0$；而 $\{c_i\}$ 并非全为零意味着 v 不是零向量．这就是说，退化的线性变换会把一些非零向量变换成零向量．

习题 3

如果线性变换非退化，那么非零向量有可能被变换为零向量吗？提示：退化和非退化的区别在于基做了线性变换后得到的向量组还是不是线性无关的，或者说还是不是一组基．

退化的线性变换会把非零向量变成零向量这一事实，引申出了零化子的概念：

定理 2　零空间

给定线性映射 $f : V \to U$，V 中被变映射到 U 中零向量的全体向量所构成的集合 $\{v \in V | f(v) = 0\}$ 构成 V 的子空间，称为 f 的零空间（**null space**），也称为零化子空间或零化子．

证明留做习题：只需要根据定义证明 $\{v \in V | f(v) = 0\}$ 是一个线性空间即可．

总结与拓展

到目前为止，我们对于向量的描述依然是抽象的，并未涉及许多具体的性质，如向量垂直、向量长度、向量坐标等概念．

在将来的词条中，我们会看到如何在给定了具体的基时用矩阵来描述向量和线性变换，以及当基的选择变化时，这些矩阵该如何变化．

零化子空间是一种子空间，而我们也会在将来的词条中解释子空间的概念．

线性映射的坐标表示

- 本词条处于草稿阶段．
- 本词条需要重新创作和整合，融入章节逻辑体系．
- 需要强调线性变换是矢量的变换，矢量本身和坐标是有区别的，坐标和基底有关，所以同一个线性映射使用不同基底，线性方程组的系数不一样

预备知识 平面旋转变换[525]

代数理解

从代数的角度来说，对于给出几个数，把它们分别与一些常数相乘再把积相加，得到另外几个数的的过程就叫**线性变换**．例如，在“平面旋转变换[525]”中，直角坐标系中任意一点 P 的坐标 (x,y) 绕远点旋转角 α 以后的坐标为

$$\begin{cases} x' = (\cos\alpha)x + (-\sin\alpha)y \\ y' = (\sin\alpha)x + (\cos\alpha)y \end{cases} \tag{1}$$

这就是一个常见的线性变换，任意给出两个实数 x,y，通过与常数相乘再相加的方法得到两个新的实数 x',y'．

有些线性变换是一一对应的，例如上面的例子中，任何一组 x,y，有且仅有一组 x',y' 与之对应，反之亦然．在这种情况下，这个变换存在**逆变换**．

线性变换的矩阵表示

由 n 个数 $x_1\dots x_n$ 变换到 m 个数 $y_1\dots y_n$ 的线性变换的一般形式为

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \quad\vdots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{cases} \tag{2}$$

这里一共有 $m\times n$ 个系数，每个系数的下标由两个数组成，a_{ij} 是计算 y_i 时 x_j 前面的系数．为了书写方便，把这些系数写成一个 m 行 n 列的数表，用圆括号括起来，就是表示该变换的**矩阵**[396]．

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \tag{3}$$

几何理解

预备知识 几何矢量[370]

我们假设式 2 中的数都是实数，那么我们可以把 $(x_1,\dots,x_n)$ 和 n 维空间中的几何矢量意义对应，即把它们看作是 X 空间中矢量 $\mathbf{x}$ 的 n 个坐标．同理我们把 $(y_1,\dots,y_n)$ 看作是 m 维空间 Y 中的几何矢量 $\mathbf{y}$ 的坐标．

变换是指，给出 X 空间中的任意 $\mathbf{x}$，都能通过某种规则**映射**（即对应）到 Y 空间中的某个矢量 $\mathbf{y}$．注意变换不能是一个 $\mathbf{x}$ 对应到多个 $\mathbf{y}$．变换的概念与函数类似，若将某变换记为 $\hat{T}$，那么可以记

$$\mathbf{y}=\hat{T}\mathbf{x} \tag{4}$$

线性变换是一种满足特定性质的变换．指给出 X 空间中的任意两个矢量 $\mathbf{x}_1$ 和 $\mathbf{x}_2$ 以及两个实数 c_1 和 c_2，这个变换 $\hat{L}$ 满足

$$\hat{L}(c_1\mathbf{x}_1+c_2\mathbf{x}_2)=c_1\hat{L}\mathbf{x}_1+c_2\hat{L}\mathbf{x}_2 \tag{5}$$

矩阵与线性映射

预备知识 线性映射[451]

我们已经知道 $M\times N$ 的矩阵[396] 可以表示一个 N 维列矢量的线性组合，得到一个 M 维列矢量（式 4 [396]）．

$$\mathbf{y}=\mathbf{A}\mathbf{x} \tag{1}$$

我们可以把这个 $\mathbf{x}$ 看做任意一个 N 维矢量空间[444] X 中某矢量关于某组基底 $\{\mathbf{x}_i\}$ 的坐标，而把 $\mathbf{y}$ 看做任意一个 M 维矢量空间（以下称为 Y 空间）中某矢量关于某组基底 $\{\mathbf{y}_i\}$ 的坐标[6]．当基底的选取不同，同一个矩阵表示的映射也不同．

[6]要再次强调列矢量 $\mathbf{x},\mathbf{y}$ 并不是矢量本身，而是矢量关于选定基底的坐标

这样，我们就通过矩阵 $\mathbf{A}$ 建立了从 X 空间到 Y 空间的一个映射[98] $A: X \to Y$. 即 X 空间的任意矢量 $\mathbf{x}$，都可以映射到 Y 空间中唯一矢量 $\mathbf{y}$. 注意该映射既不一定是单射也不一定是满射，而是要取决于 $\mathbf{A}$ 的具体性质.

特殊地，当矩阵 $\mathbf{A}$ 为方阵时，矩阵 $\mathbf{A}$ 可以用于表示 X 空间到自身的自映射 $A: X \to X$，即 $\mathbf{x}$ 和 $\mathbf{y}$ 都是 X 空间中的矢量的坐标，但 $\{\mathbf{x}_i\}$ 和 $\{\mathbf{y}_i\}$ 仍然可以是 X 空间中两组不同的基底. 同样，自映射也未必是单射或满射.

由矩阵与列矢量乘法的分配律（式 17 [396]）可知 X 空间中若干个矢量做任意线性组合然后映射到 Y 空间，等于这些矢量先分别映射到 Y 空间再做同样的线性组合，即

$$\mathbf{A}\left(\sum_i c_i \mathbf{x}_i\right) = \sum_i c_i (\mathbf{A}\mathbf{x}_i) \tag{2}$$

这说明，**任何矩阵 $\mathbf{A}$ 表示的映射 $A: X \to Y$ 必定是线性映射**[451].

作为一个简单的例子，我们来看平面旋转矩阵[402]

$$\mathbf{R}_2 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{3}$$

这是一个方阵，对应二维矢量空间（例如二维几何矢量构成的空间）的自映射. 对于这个矩阵我们有“主动”和“被动”两种理解，前者假设基底不变而矢量旋转，后者假设矢量不变而基底旋转[7]. 这个映射中，映射前后的矢量有一一对应[98] 关系.

我们还可能有**多对一**映射，即多个矢量映射后可能得到同一个矢量（例 1）. 来看一个例子.

例 1　投影矩阵

我们考虑一个投影变换：将平面上任意几何矢量投影到 $\hat{\mathbf{x}}+\hat{\mathbf{y}}$ 方向上得到该方向的矢量. 已知该变换是线性的，写出变换矩阵（变换前后使用同一组正交归一基底 $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2$）

解：与“平面旋转矩阵[402]”中的方法同理，先考虑各基底的投影变换. $\hat{\mathbf{x}}_1 = (1,0)^{\mathrm{T}}$ 投影后变为 $(1,1)^{\mathrm{T}}/\sqrt{2}$，$\hat{\mathbf{x}}_2$ 投影后同样变为 $(1,1)^{\mathrm{T}}/\sqrt{2}$，所以投影

[7] 注意“主动”和“被动”并不是两种唯一的理解，例如我们可以选择让基底顺时针旋转 $\theta/2$，矢量逆时针旋转 $\theta/2$.

变换矩阵即使两个列矢量组成的矩阵

$$\mathbf{P}=\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} \tag{4}$$

注意该变换中虽然每个矢量都映射到同一空间的唯一的矢量，但不同的矢量有可能映射到同一个矢量．所以这是一个多对一映射．

定义域和值域

预备知识　子空间[464]

式 1 表示的线性映射中，**定义域（domain）**是 X 空间中的任意矢量，而**值域（range）**却不一定是整个 Y 空间，也可能是 Y 的一个子空间．任何情况下，**线性映射的值域必定是一个矢量空间**．例如例 1 中投影变换的值域就是沿 $\hat{\mathbf{x}}_1+\hat{\mathbf{x}}_2$ 方向的任意矢量（包括零矢量）构成的一维矢量空间，是二维矢量空间中的一个子空间．

矩阵的列

我们先来看第一空间的基底 $\boldsymbol{\alpha}_1$ 如何映射到第二空间．$\boldsymbol{\alpha}_1$ 对应的列矢量是 $(1,0,\dots)^{\mathrm{T}}$，作为 $\mathbf{x}$ 输入矩阵 $\mathbf{A}$ 得 $\mathbf{y}$ 等于 $\mathbf{A}$ 的第一列．同理，矩阵的第 i 列就是 $\boldsymbol{\alpha}_i$ 映射到第二个空间的矢量在基底 $\{\boldsymbol{\beta}_i\}$ 上的坐标．

未完成：未完成

矩阵元

线性算符对应矩阵，记为 $\hat{A}$，如果有一组正交归一基底 $|u_i\rangle$，那么矩阵元为

$$A_{i,j}=\langle u_i|A|u_j\rangle \tag{5}$$

矢量空间的表示

- 本词条需要重新创作和整合，融入章节逻辑体系.

预备知识 张成空间[447]

由于矢量空间中运算的线性性，可以使用矩阵来表示任何一个矢量空间中的元素和线性变换. 对于一个域 $\mathbb{F}$ 上的 n 维线性空间中的矢量，我们惯例上使用一个 n 行 1 列的矩阵来表示，称为**列向量**. 线性变换被表示成一个 $n \times n$ 的矩阵. 这些矩阵中的元素都必须取自 $\mathbb{F}$.

需要注意的是，这些表示都依赖于该矢量空间的**基**的选取.

用基向量来表示向量和线性变换

给定域 $\mathbb{F}$ 上的 n 维线性空间 V 和它的一个基 $\{e_i\}_{i=1}^n$. 由于 V 中的每一个向量都可以唯一地表示成基向量的线性组合，因此我们可以用线性组合的系数来构成一个列向量，作为这个向量在基 $\{e_i\}_{i=1}^n$ 下的**坐标**. 比如，向量 $a_1e_1 + \cdots + a_ne_n$ 在这个基下的坐标就是

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \tag{1}$$

基的选择不同，同一个向量的坐标也就不一样.

在研究线性变换的时候，我们只需要关注线性变换对基向量的变换，就可以据此计算出任意向量的线性变换. 如果某一个线性变换 T 把基向量 e_i 变换成 $a_{i1}e_1 + \cdots + a_{in}e_n$，那么我们可以在这个基下把 T 表示成一个矩阵：

$$M = \begin{pmatrix} a_{11}, a_{12}, \cdots, a_{1n} \\ a_{21}, a_{22}, \cdots, a_{2n} \\ \vdots \ddots \vdots \\ a_{n1}, a_{n2}, \cdots, a_{nn} \end{pmatrix} \tag{2}$$

这样，如果把 v 的坐标是列向量 $\mathbf{c}$，那么 $\mathbf{Mc}$ 就是 Tv 的坐标.

同样地，线性变换的矩阵表示，也依赖于基的选取.

对偶空间

预备知识 张成空间[447]

线性函数

中学数学中一个贯穿始终的研究对象就是**函数（function）**．函数是**映射（mapping）**的一种，在现代数学的习惯中通常把从任意集合到一个数字集合[8]的映射就被称为一个函数．特别地，我们把定义线性空间时使用的域 $\mathbb{K}$ 都当作数字集合，多数情况下这些域都是实数域、复数域或者它们的某种等价类划分[9]．

如果给定一个域 $\mathbb{K}$ 上的线性空间 V，把 V 中的任意向量 $\mathbf{v}$ 映射到 $\mathbb{K}$ 中的一个元素 $f(\mathbf{v})$ 上，那么映射 $f:V\to\mathbb{K}$ 就是线性空间 V 上的一个函数．你可以任意指定函数的对应规则，但我们只关心所有可能的函数中最容易研究的一类，线性函数．

线性函数由于具有可加性，通俗来说就是“函数图像是平直的”，因此非常容易讨论．在微分几何中，尽管我们研究的不一定是平直的函数，但是总能在局部取平直的线性函数来考察任一点附近的函数性质，这其实就是微分的思想．换句话说，微分实际上是一种“线性近似”，在给定点把可微函数近似为一个线性函数，方便讨论．

定义 1　线性函数

给定一个域 $\mathbb{K}$ 上的线性空间 V，如果映射 $f:V\to\mathbb{K}$ 满足线性性，即对于任意 $\mathbf{v},\mathbf{u}\in V$ 以及任意的 $a,b\in\mathbb{K}$，都有 $f(a\mathbf{v}+b\mathbf{u})=af(\mathbf{v})+bf(\mathbf{u})$，那么我们称 f 是 V 上的一个线性函数（**linear function**）．

线性性质的好处就在于，线性函数的性质只依赖于基向量的函数值．就是说，当我们取定了一组基 $\{\hat{\mathbf{e}}_i\}$，那么只要知道了线性函数 f 对它们的取值 $f(\hat{\mathbf{e}}_i)$，我们就可以利用线性性算出任意向量 $\mathbf{v}\in V$ 的函数值 $f(\mathbf{v})$，因为每一个 $\mathbf{v}$ 都是基向量的某个线性组合．

[8] 数字集合就是我们通常叫做数字的元素的集合，比如实数集 $\mathbb{R}$，有理数集 $\mathbb{R}$，整数集 $\mathbb{R}$，复数集 $\mathbb{C}$ 等，甚至 $\{0,1\}$ 这样的集合都叫数字集合．

[9] 如整数[95] 中提到的模运算，其元素也被称为数字，只不过不是通常意义上的整数．

如果要定义 V 上一个任意的函数，我们的自由度很高，可以对每一个向量都指定一个函数值，各向量的函数值并不互相影响；但是对于线性函数，自由度就只有 $\dim V$ 个，一旦选定了基向量的函数值，所有向量的函数值就都确定了．不过基向量的函数值仍然可以任意指定．

函数之间可以进行加减法和乘法运算，这使得函数自然构成一个环[569]，线性函数自然也不例外．目前我们只关心线性函数之间的加减法．

定义 2　线性函数的加减法

给定一个域 $\mathbb{K}$ 上的线性空间 V，对于两个线性函数 $f,g:V\to\mathbb{K}$，定义函数的和 $(f+g):V\to\mathbb{K}$ 如下：对于任意 $\mathbf{v}\in V$，有 $(f+g)(\mathbf{v})=f(\mathbf{v})+g(\mathbf{v})$.

这样一来，我们就可以对函数也进行加减法，就像对数字所做的那样．同样地，我们可以定义函数的数乘，即用域中的数字和函数相乘来得到新的函数．对于线性函数，数乘定义如下：

定义 3　线性函数的数乘

给定一个域 $\mathbb{K}$ 上的线性空间 V，标量 $a\in\mathbb{R}$ 以及线性函数 $f:V\to\mathbb{K}$，定义函数的数乘 $af:V\to\mathbb{R}$ 如下：对于任何 $\mathbf{v}\in V$，有 $(af)(\mathbf{v})=a\cdot f(\mathbf{v})$.

线性函数的数乘对加法具有分配律：

定理 1　线性函数的数乘对加法的分配律

给定一个域 $\mathbb{K}$ 上的线性空间 V，对于两个线性函数 $f,g:V\to\mathbb{K}$ 和标量 $a\in\mathbb{R}$，有：$a(f+g)=af+ag$.

这一定理的证明是非常直接的，应用域 $\mathbb{K}$ 上的乘法分配律就可以．

对偶空间

从定义 2 和定义 3，以及定理 1 我们可以看出，对于一个域 $\mathbb{K}$ 上的线性空间 V，V 上的全体线性函数也满足向量空间的定义，构成一个向量空间．我们把这个空间称作 V 的**对偶空间**.

定义 4　对偶空间

给定一个域 $\mathbb{K}$ 上的线性空间 V，记 V^* 为集合 $\{V\text{上全体线性函数}\}$．在 V^* 上定义加法和数乘如定义 2 和定义 3，那么 V^* 构成一个线性空间，称作 V 的

对偶空间.

看起来，V^* 称作线性函数空间更直接，为什么要特地取“对偶”一名字呢？初看对偶空间的定义，你可能会觉得 V 和 V^* 的元素之间是不对等的，V^* 中的函数是用来对 V 中的向量进行映射的．可是反过来看的话，V 中向量也可以看成是 V^* 中向量的映射．这就是说，$f \in V^*$ 可以看成是一个 $V \to \mathbb{K}$ 的映射，而 $\mathbf{v} \in V$ 也可以看成是 $V^* \to \mathbb{K}$ 的映射．

定理 2　V 是 V^* 的对偶空间

考虑一个域 $\mathbb{K}$ 上的线性空间 V 及其对偶空间 V^*，那么对于任意 $\mathbf{v} \in V$，$\mathbf{v}$ 是 V^* 上的线性函数.

如果我们形式上进行一下简化，把 f 记为向量形式 $\mathbf{f}$，然后把 $f(\mathbf{v})$ 记为 $\mathbf{fv}$，那么可以更清晰地看出它们互为彼此的线性函数：给定一个域 $\mathbb{K}$ 上的线性空间 V 及其对偶空间 V^*，取 $a_i \in \mathbb{K}$、$\mathbf{f}_i \in V^*$ 和 $\mathbf{v}_i \in V$，那么从 $\mathbf{f}_0(a_1\mathbf{v}_1 + a_2\mathbf{v}_2) = a_1\mathbf{f}_0\mathbf{v}_1 + a_2\mathbf{f}_0\mathbf{v}_2$ 可知，$\mathbf{f}$ 是 V 上的线性函数；从 $(a_1\mathbf{f}_1 + a_2\mathbf{f}_2)\mathbf{a}_0 = a_1\mathbf{f}_1\mathbf{v}_0 + a_2\mathbf{f}_2\mathbf{v}_0$ 可知，$\mathbf{v}$ 是 V^* 上的线性函数.

我们可以这么理解：$\mathbf{fv}$ 实际上就是“取 V 和 V^* 中各一个元素，把它们俩对应到一个标量上”的过程．两个空间形式上平等，互为对偶空间，甚至可以写成 $V^{**} = V^*$．由此，我们可以更对称地定义对偶空间的概念：

定义 5　对偶空间

给定域 $\mathbb{K}$ 上两个 n 维线性空间 V 和 W．如果指定了一个双线性映射 $m : V \times W \to \mathbb{K}$[10]，那么称 V 和 W 是对偶的．如果把 $\mathbf{v} \in V$ 视为线性函数 $f_{\mathbf{v}} : W \to \mathbb{K}$，那么对于任意 $\mathbf{w} \in W$ 有 $f_{\mathbf{v}}(\mathbf{w}) = f(\mathbf{v}, \mathbf{w})$.

对偶空间的表示

给定线性空间 V 作为原空间，再给定 V 的一组基，我们就可以把 V 中的向量表示为列矩阵．对应地，V^* 中的向量表示为行矩阵，这个行矩阵是该向量在对偶基下的坐标．这样一来，向量和对偶向量互相乘起来得到标量的过程，就是这两个矩阵相乘的过程，其中对偶向量的行矩阵放在左边，向量的列矩阵

[10] 即对于 $a_i, b_i \in \mathbb{K}$、$\mathbf{v}_i \in V$ 和 $\mathbf{w}_j \in W$，有 $m(\sum_{i=1}^{n} a_i\mathbf{v}_i, \sum_{j=1}^{n} b_j\mathbf{w}_j) = \sum_{i,j=1}^{n} a_i b_j m(\mathbf{v}_i)m(\mathbf{w}_j)$

放在右边.

对偶基

给定了 V 的基 $\{\mathbf{e}_i\}$ 后，为了方便计算，我们通常规定 V^* 的基 $\{\mathbf{f}_j\}$ 满足关系：$\mathbf{e}_i\mathbf{f}_j = \delta_{ij}$. 这样我们就由 V 的基**唯一**地决定了 V^* 的基，称这样的 V^* 的基为 V 的**对偶基（dual space）**.

事实上，以上定义是对偶基的一个特例. 一般地，如果在给定了 V 的基 $\{\mathbf{e}_i\}$ 后，有一个对称矩阵 g_{ij}，它被定义为 V 的度量在给定基下的表示[11]，那么对偶基的计算应该满足 $\mathbf{e}_i\mathbf{f}_j = g_{ij}$. 注意，这里并没有要求 g_{ij} 是正定的矩阵. 具体的度量矩阵的定义要由问题情景来决定，比如说，狭义相对论中的闵可夫斯基度量就是：

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{1}$$

在闵可夫斯基度量下，如果一个向量在某基下的坐标是 $(a_1, a_2, a_3, a_4)^{\mathrm{T}}$，一个对偶向量在**对偶基**下的坐标是 (b_1, b_2, b_3, b_4)，那么它们相乘之后的结果就是 $a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4$.

狄拉克符号

狄拉克符号是一种用来描述量子态的表示方法. 量子力学中认为每个量子态都是一个向量，定义在复数域上的线性空间中，因此狄拉克符号表示的都是复向量.

狄拉克符号分为左矢和右矢，其中左矢相当于行矩阵，右矢相当于列矩阵；左矢和右矢在两个不同的空间中，称为左矢空间和右矢空间，这两个空间彼此对偶. 每个量子态向量都有双重表示，一个左矢表示，一个右矢表示. 同一个量子态的左矢和右矢是对偶向量，故满足以下条件：在同一个基[12]下，如果同

[11] 如果两个向量的坐标分别是 x_i 和 y_j，那么它们的内积定义为 $x_iy_jg_{ij}$.

[12] 量子力学中所说的基和所谓的“表象”相关，比如在空间表象下的基就是空间的特征态，也就是在空间坐标下表示的波函数为 δ 函数的那些态；动量表象下的基就是动量的特征态，这些基向量都是在全空间概率均匀分布的波函数，不过用动量坐标来表示的话，这些波函数又是 δ 函数了. 同一个波函数在两种坐标之间的表示由傅里叶变换计算出，具体方式详见量子力学.

一个量子态向量 s 的左矢表达是 $\langle s|$，右矢表达是 $|s\rangle$，那么有 $\langle s|\cdot|s\rangle=1$．为了简便，我们也把这个表达写成 $\langle s|\cdot|s\rangle=\langle s|s\rangle$．

一般来说，量子态空间都是无穷维希尔伯特空间，并不能把向量坐标简单表示为矩阵．但是我们依然可以用矩阵来做类比，在这样的类比下，如果一个表象下的一个量子态的右矢中第 i 个坐标是复数 s_i，那么其左矢中第 i 个坐标就是 s_i^*，也就是对应坐标值取其复数共轭，这样就能满足条件 $\langle s|s\rangle=1$．

由于量子态的另一种理解方式是波函数，所以我们其实可以把左矢、右矢的乘积理解为波函数的乘积积分，就像最常见的函数向量一样．具体来说，如果有两个量子态 s 和 t，分别对应波函数 f_s 和 f_t，以及右矢 $|s\rangle$ 和 $|t\rangle$，那么我们有对应关系：

$$\langle s|t\rangle=\int_{\text{全空间}} f_s^* f_t dx \tag{2}$$

注意左矢对应的 f_s 取了复共轭．

子空间

预备知识 线性相关和线性无关[378]

如果一个 N_S 维的矢量空间 $\mathcal{S}$ 中的所有矢量都属于另一个 N 维的矢量空间 $\mathcal{V}$，且量空间中加法和数乘运算的定义相同．那么前者就是后者的子空间．严格来说所有矢量空间都是它本身的子空间（即 $N_S=N$ 的情况），但我们一般是指 $N_S<N$ 的情况．

注意若我们在 $\mathcal{V}$ 中任意找一组基底，不一定能恰好从中选出 N_S 个使其作为 $\mathcal{S}$ 空间的基底．若想让 $\mathcal{V}$ 空间的一组基底包含 $\mathcal{S}$ 空间的基底，我们可以先在 $\mathcal{S}$ 空间中选 N_S 个基底，再在 $\mathcal{S}$ 空间外选取 $N-N_S$ 个基底即可[13]．

例 1　三维空间中的平面

在三维的几何矢量空间 $\mathcal{V}$ 中，若建立直角坐标系，可以选 $\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}$ 作为一组正交归一的基底．现在来看过原点且以矢量 $\hat{\mathbf{x}}+\hat{\mathbf{y}}+\hat{\mathbf{z}}$ 为法向量的平面，平

[13]这样的基底一定是存在的，因为 N 维空间中任意给出 N_S 个线性无关的矢量，就必定能再找到另外 $N-N_S$ 个线性无关的矢量

面方程为[14]

$$x+y+z=0 \tag{1}$$

所有与该平面重合的矢量可以构成这个三维空间中的一个二维子空间 $\mathcal{S}$. 证明：平面上的两个矢量相加仍然落在平面上，数乘也同样落在平面上，详细过程略.

但是，由于 $\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}$ 中任意一个都落在该子空间外面，所以不可能选出两个作为子空间的基底. 如果需要是选一组 $\mathcal{V}$ 的基底且包含 $\mathcal{S}$ 空间的基底，可以现在 $\mathcal{S}$ 空间中选两个基底（坐标满足式 1），例如坐标为 $(1,-2,1)/\sqrt{6}$ 和 $(1,0,-1)/\sqrt{2}$ 的两个矢量. 再在空间外取一个基底，如 $(1,1,1)/\sqrt{3}$. 注意这里给出的三个矢量是一组正交归一基底，但原则上只需要线性无关即可. 在该情况下，线性无关意味着三个几何矢量两两不共线且不共面（例 1 [378]）.

矢量张成的空间

如果在 N 维矢量空间 V 中给定 M 个矢量（不一定要求 $M\leqslant N$，也不一定都线性无关）$|v_1\rangle,\ldots,|v_M\rangle$，那么容易证明

$$|v\rangle=\sum_{i=1}^{M}c_i\,|v_i\rangle \tag{2}$$

的集合就是 V 的一个子空间（包括 V 本身）. 我们把这个空间叫做 $|v_1\rangle,\ldots,|v_M\rangle$ **张成**的空间. 当 $|v_1\rangle,\ldots,|v_M\rangle$ 线性无关时，该空间维 M 维，否则就小于 M 维.

要获得张成空间的一组基底，我们可以在 M 个矢量中，将所有可以表示为其他矢量线性组合的矢量剔除，直到所有剩下的矢量线性无关，就得到了一组基底.

子空间的直和补空间

- 本词条存在未完成的内容.

[14] 参考高中数学教材中的立体几何章节

预备知识　子空间[464]

定义 1　直和

令域 $\mathbb{F}$ 上的向量空间 V_1 和 V_2 为 V 的两个子空间[464]，满足

$$V_1 \cap V_2 = \{0\} \tag{1}$$

且任意 $v \in V$ 都能表示为 V_1 和 V_2 中向量的线性组合，即

$$v = c_1 v_1 + c_2 v_2 \qquad (v_1 \in V_1,\ v_2 \in V_2,\ c_1, c_2 \in \mathbb{F}) \tag{2}$$

那么空间 V 就是 V_1 和 V_2 的直和空间，用直和（**direct sum**）运算记为

$$V = V_1 \oplus V_2 \tag{3}$$

我们把这两个子空间叫做互补的，即 V_2 是 V_1 在 V 中的补空间（**complement space**），反之亦然.

事实上，定义中式 2 也可以改为

$$v = v_1 + v_2 \qquad (v_1 \in V_1,\ v_2 \in V_2) \tag{4}$$

即“任意 $v \in V$ 都能表示为 V_1 和 V_2 中向量之和”. 这是因为 $c_1 v_1 \in V_1$，$c_2 v_2 \in V_2$.

直和空间 $V_1 \oplus V_2$ 中的所有向量可以分为三组，分别是 V_1 中的向量，V_2 的向量，以及只能表示为 V_1 和 V_2 中非零向量之和的向量.

多个子空间的直和

同理，也可以把向量空间 V 表示为多个子空间的直和

$$V = V_1 \oplus \cdots \oplus V_n \tag{5}$$

这可以理解为先把 V_1, V_2 做直和，再把所得空间与 V_3 做直和，等. 事实上，根据向量加法的分配律，容易证明直和也满足分配律，即

$$(V_1 \oplus V_2) \oplus V_3 = V_1 \oplus (V_2 \oplus V_3) \tag{6}$$

所以多个空间的直和也无需添加括号.

以下讨论为了方便只使用两个子空间，但对多个子空间同样适用.

直和空间的基底

从基底的角度来看，若 V_1 和 V_2 中分别有一组基底 α_i $(i=1,\dots,N_1)$ 和 β_i $(i=1,\dots,N_2)$，那么直和空间中 V 的任意向量可以表示为

$$v=\sum_i a_i\alpha_i+\sum_j b_j\beta_j \qquad (v\in V, a_i, b_i\in\mathbb{F}) \tag{7}$$

定理 1

把 V_1 的一组基底和 V_2 的一组基底按任意顺序排列，可以得到 $V_1\oplus V_2$ 的一组基底.

证明：由于已经有式 7，我们只需要证明 $\alpha_1,\dots,\alpha_{N_1},\beta_1,\dots,\beta_{N_2}$ 是线性无关的．使用反证法，若有不全为零的系数使

$$\sum_i a_i\alpha_i+\sum_j b_j\beta_j=0 \tag{8}$$

令

$$u=\sum_i a_i\alpha_i=-\sum_j b_j\beta_j \tag{9}$$

那么 $u\neq 0$ 且 $u\in V_1$ 且 $u\in V_2$．这违反了定义 1．证毕.

推论 1

直和空间 $V=V_1\oplus V_2$ 的维数等于向量空间 V_1,V_2 的维数之和.

证明：定理 1 中的基底有 N_1+N_2 个．证毕.

注意 N_1,N_2 可以等于零，零维线性空间仅由零向量一个元素构成.

例 1

若三维空间中有两个不共线的几何向量 $\mathbf{v_1},\mathbf{v_2}$，它们张成一个平面，或二维子空间．另有一个向量 $\mathbf{v_3}$，独自张成一条直线，即一维空间.

若 $\mathbf{v_3}$ 落在 $\mathbf{v_1},\mathbf{v_2}$ 张成的平面内，则三个向量的所有线性组合仍然在该平面内，所以直和空间仍然是该平面.

若 $\mathbf{v_3}$ 落在平面外，则三个向量将会张成整个三维空间，所以直和就是三维空间．此时两个子空间在该三维空间中互补.

定理 2　唯一分解

直和空间 $V = V_1 \oplus V_2$ 中，对任意 $v \in V$ 都存在唯一的 $v_1 \in V_1$ 和 $v_2 \in V_2$ 使得

$$v = v_1 + v_2 \tag{10}$$

证明：根据定理 1，式 7 的分解中每个系数（坐标）a_i, b_i 都是唯一的，右边的两个求和就分别是式 10 的 v_1 和 v_2，所以 v_1, v_2 也是唯一的.

习题 1

把以上定理和推论拓展到多个子空间的情况并证明.

未完成：补是不唯一的（例 1）

未完成：给出一个空间和子空间，如何求一个补空间？

线性映射的结构

预备知识　补空间[465]

如果线性映射是单射[98] 的（即定义域空间到值域空间[451] 是双射的），那么它的结构简单明了，没有太多值得讨论的. 我们现在来看一个揭示多对一线性映射的结构的重要定理. 考虑线性映射 $A : X \to Y$，线性空间 X 的零空间（定理 2 [451]）X_0 中的每个矢量都被 A 映射到 Y 空间中的零矢量，即 $A(X_0) = \{0\}$.

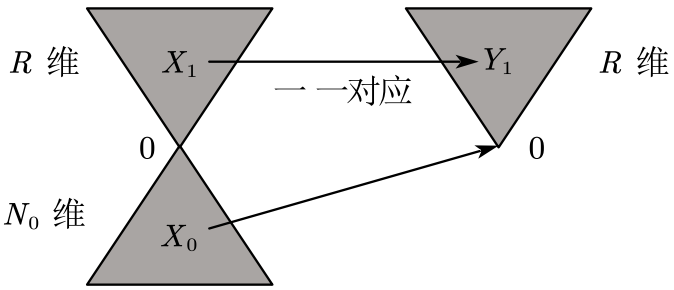


图 1: 两个 X 子空间的线性映射：X_0 是零空间，X_1 是 X_0 在 X 中的补空间

定理 1

令 $A: X \to Y$ 的零空间为 X_0，X_1 为 X_0 在 X 中任意一个补空间（即 $X_0 \oplus X_1 = X$，定义 1 [465]），令值空间 $Y_1 = A(X)$，那么映射 $A: X_1 \to Y_1$ 是一一对应的（即双射）.

证明见下文. 我们可以形象地把定理 1 用图 1 表示，图中每个三角形代表一个矢量空间，由于 X_0, X_1 是互补的，它们只相交于零矢量. 注意对于给定的映射 A，X_0 和 Y_1 是确定的，而 X_0 的补空间 X_1 可以随意选取.

特殊地，若 X_0 是零维空间即 $X_0 = \{0\}$，则补空间 $X_1 = X$. 此时可以把图 1 中表示 X_0 的三角形去掉.

该定理可以更好地帮助我们理解线性方程组解的结构[482]：当 $\mathbf{Ax} = \mathbf{b}$ 的解集等于一个特解加齐次解集，齐次解. $\mathbf{b}$ 是 Y_1 中的某向量，特解是 X_1 中映射到 $\mathbf{b}$ 的向量，而齐次解就是零空间 X_0.

推论 1

令定理 1 中 X, X_0, Y_1 的维数分别为 N_X, N_0, R，那么

$$N_X = N_0 + R \tag{1}$$

证明：见式 1，由于 X_1 与 Y_1 一一对应，X_1 的维数也是 R，又因为 $X_0 \oplus X_1 = X$，X 的维数等于 X_0, X_1 维数之和（推论 1 [465]）. 证毕.

推论 2

线性映射 $A: X \to Y$ 中值空间 $A(X) \subseteq Y$ 的维数小于或等于定义域空间 X 的维数.

证明：式 1 中 $N_X, N_0, R > 0$，所以有 $R \leqslant N_X$.

推论 3

线性映射 $A: X \to Y$ 是单射（即 $A: X \to A(X)$ 是双射）当且仅当零空间 X_0 中只有零向量.

说明：零空间 X_0 只有零矢量一个元素当且仅当式 1 中 $N_X = R$ 或 $N_0 = 0$.

证明：前者证明后者：使用推论 1 [451] 以及单射的定义，可得 $0 \in X$ 是零空间中的唯一向量. 后者证明前者：$N_0 = 0$ 即定理 1 中的 $X = X_1$，与值空间 $Y_1 = A(X)$ 有一一对应关系. 证毕.

证明

要证明定理 1，首先证明 $A(X_1)=A(X)$．任意 $x\in X$ 可以表示为 $x=x_0+x_1$，其中 $x_0\in X_0$，$x_1\in X_1$．$Ax=Ax_0+Ax_1=Ax_1$．这说明任意 $y_1\in Y_1=A(X)$ 都能找到 x_1 使 $Ax_1=y_1$，所以 $A(X_1)=Y_1$．

现在证明 X_1 和 Y_1 一一对应：令 $u,v\in X_1$，我们要证明如果 $Au=Av$ 那么 $u=v$．算符 A 是线性的，所以 $A(u-v)=0$，所以 $u-v\in X_0$．由封闭性，$u-v\in X_1$．由于补空间只交于零矢量 $X_0\cap X_1=\{0\}$，所以 $u-v=0$，即 $u=v$．证毕．

未完成：线性方程组词条[482] 中说明：线性方程组的解空间就是 X_1 中的特解加上齐次解 X_0．

狄拉克符号

预备知识 对偶空间[460]，内积[471]

在物理中，尤其是量子力学中，我们经常会见到**狄拉克符号（Dirac notation）**．矢量空间 X 中，如果使用狄拉克符号，那么其中的元素（矢量）$x\in X$ 可以记为 $|x\rangle$，我们把它叫做**右矢（ket）**．对应地，我们把 X 的对偶空间 X^* 中的矢量叫做**左矢（bra）**[15]，记为 $\langle x|$．一般来说[16]，每个 X^* 中的每个左矢和 X 中的右矢是一一对应[98]．我们说任意 $|x\rangle$ 和对应的 $\langle x|$ 互为**对偶**．

两矢量的加减法记为 $|x\rangle\pm|y\rangle$，$|x\rangle$ 与标量 λ 的数乘可以记为 $\lambda|x\rangle$ 或 $|\lambda x\rangle$．

在对偶空间中，加减法、标量积、算符作用同样记为 $\langle x|\pm\langle y|$，$\lambda\langle x|$ 和 $A\langle x|$．但要注意的是，我们将 $|\lambda x\rangle$ 的对偶矢量记为 $\langle\lambda x|$，这样一来，在复矢量空间中把 λ 写在左矢的里面和外面意义就变得不一样，即

$$\langle\lambda x|=\lambda^*\langle x| \tag{1}$$

如果这个矢量空间中定义了内积，那么 X 中任意两个矢量 $|x\rangle,|y\rangle$ 的内积用狄拉克符号记为 $\langle x|y\rangle$．我们也可以理解为内积是左矢空间中 $\langle x|$ 和右矢空间中 $|y\rangle$ 的二元运算．

[15] bra 和 ket 这两个单词是由英文单词 bracket 拆分而来的．braket 在这里指 $\langle\cdot|\cdot\rangle$．

[16] 严格来说，这要求空间 X 是**完备**的，完备的内积空间又叫**希尔伯特空间**[816]（**Hilbert space**），是量子力学中主要讨论的矢量空间．一般来说我们只在希尔伯特空间中使用狄拉克符号．

算符（映射）$A : X \to X$ 作用在 $|x\rangle$ 上同样可以记为 $A|x\rangle$ 或 $|Ax\rangle$．$|Ax\rangle$ 的对偶记为 $\langle Ax|$ 而不是 $A\langle x|$ 或 $\langle xA|$，详见“伴随算符[??]”.

内积内积空间

预备知识 几何向量的内积[383]，向量空间[444]

[17]在向量空间中，我们可以另外定义任意两个向量的**内积（inner product）**运算，运算的结果是一个实数或复数．内积运算不是向量空间所必须的，但物理中的向量空间几乎都定义了内积运算．我们把定义了内积运算的空间称为**内积空间（inner product space）**，完备的内积空间也叫**希尔伯特空间**[816]**（Hilbert space）**．本词条中我们使用狄拉克符号[470]．

定义 1

对域 $\mathbb{F}$ 上的向量空间[444]，定义内积为满足以下条件的二元运算 $\langle\cdot|\cdot\rangle : V \times V \to \mathbb{F}$．令向量 $u, v \in V$ 以及标量 $a, b \in \mathbb{F}$

- 正定

$$\langle v|v\rangle > 0 \quad (v \neq 0) \tag{1}$$

- 线性

$$\langle w|au + bv\rangle = a\langle w|u\rangle + b\langle w|v\rangle \tag{2}$$

- 交换律

$$\langle u|v\rangle = \langle v|u\rangle^* \tag{3}$$

说明：

- 定义中我们使用了狄拉克符号[470]，也有地方将内积记为 $\langle u, v\rangle$.
- 向量空间依赖于一个域 $\mathbb{F}$ 通常取实数或复数域．对于实数（复数）域的向量空间，内积的结果必须是实数（复数）．对实数的情况，定义中的共轭（*）可以略去．

[17]参考 Wikipedia 相关页面.

- 式 1 意味者无论 $\mathbb{F}$ 是什么，一个向量和自身的内积必须是**实数**.
- 令式 2 中 $a,b=0$ 可得**零向量和任何向量的内积为零**. 如果两个向量内积为零，我们就说他们是**正交（orthogonal）**的，这意味着**零向量和任何向量正交**.
- 根据式 2，零向量和任何向量的内积都必定是 0. 结合式 3 和式 2，可以得到 $\langle au+bv|w\rangle=a^*\langle w|u\rangle+b^*\langle w|v\rangle$.
- 结合式 2 和式 3 可得 $\left\langle \sum_i a_iu_i\middle|\sum_j b_jv_j\right\rangle=\sum_{i,j}a_i^*b_j\langle u_i|v_j\rangle$.

内积一个重要性质就是满足柯西不等式[800]

$$|\langle u|v\rangle|^2\leqslant\langle u|u\rangle\cdot\langle v|v\rangle \tag{4}$$

即两个向量内积绝对值的平方小于它们分别和自身内积再相乘. 由柯西不等式可以证明内积空间必然可以定义范数（证明见下文）

$$\|v\|=\langle v|v\rangle \tag{5}$$

内积的坐标表示

N 维内积空间中，必存在 N 个正交归一基底，任意向量 v 可以在这组基底上找到对应的坐标 $(v_1,\dots,v_N)$，那么任意两个向量 u,v 的内积可以用坐标表示为

$$\langle u|v\rangle=\sum_i u_i^*v_i \tag{6}$$

但若基底不是正交的，令基底为 $\beta_1,\dots,\beta_N$，那么根据内积的线性容易证明两个向量的内积为

$$\langle u|v\rangle=\mathbf{u}^\dagger\mathbf{M}\mathbf{v}=\sum_{i,j}M_{i,j}u_i^*v_j \tag{7}$$

其中矩阵 $\mathbf{M}$ 的矩阵元为 $M_{i,j}=\langle\beta_i|\beta_j\rangle$. 当 $\langle\beta_i|\beta_j\rangle=\delta_{i,j}$ 时，$\mathbf{M}$ 变为单位矩阵，就回到了式 6. 可见式 7 才是最一般坐标内积公式.

勾股定理

内积空间的**勾股定理（Pythagorean theorem）**：对任意两个正交的向量，有

$$\langle u+v|u+v\rangle = \langle u|u\rangle + \langle v|v\rangle \tag{8}$$

证明：

$$\langle u+v|u+v\rangle = \langle u|u\rangle + \langle v|v\rangle + \langle u|v\rangle + \langle v|u\rangle \tag{9}$$

根据正交的定义，$\langle u|v\rangle = 0$．证毕．

证明内积必定是范数

要证 $\langle x|x\rangle$ 满足范数的要求，最关键是证明性质

$$\|x+y\|^2 \leqslant (\|x\| + \|y\|)^2 = \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| \tag{10}$$

即证

$$\langle x+y|x+y\rangle - \langle x|x\rangle - \langle y|y\rangle \leqslant 2\|x\|\|y\| \tag{11}$$

其中

$$\langle x+y|x+y\rangle = \langle x|x\rangle + \langle y|y\rangle + 2\,\mathrm{Re}[\langle x|y\rangle] \tag{12}$$

代入，即证

$$\mathrm{Re}[\langle x|y\rangle]^2 \leqslant \|x\|^2\|y\|^2 = \langle x|x\rangle\,\langle y|y\rangle \tag{13}$$

由柯西不等式

$$\mathrm{Re}[\langle x|y\rangle]^2 + \mathrm{Im}[\langle x|y\rangle]^2 = |\langle x|y\rangle|^2 \leqslant \langle x|x\rangle\,\langle y|y\rangle \tag{14}$$

而 $\mathrm{Im}[\langle x|y\rangle]^2 > 0$．证毕．

第四章

线性方程组

线性方程组

求解解线性方程组是线性代数学习中的一个主要线索. 虽然我们在中学阶段学过一些简单的技巧，但要讨论最一般的情况和背后的数学结构，我们还需要经历一个较为漫长的学习.

低维简单例子

二元一次方程组

在许多实际生活中，我们往往需要解决类似的问题：

小明拿着家长给的 100 元去超市买饮料. 超市里有 3 元钱一瓶的可口可乐和 5 元钱一瓶的奶茶，最后他带着 28 瓶饮料回家过暑假了. 求问小明买了多少瓶可口可乐，多少瓶奶茶？

这个问题实际上是一个二元一次方程组的问题，我们设小明买了 x 瓶可口可乐，y 瓶奶茶，可以列出一个方程组：

$$\begin{cases} 100 = 3x + 5y (a) \\ 28 = x + y (b) \end{cases} \tag{1}$$

一个简单的解决方法是计算 $(a) - 3 * (b)$，得到 $16 = 0x + 2y$ 即 $y = 8$，进一步就知道 $x = 20$.

三元一次方程组

类似的在三维坐标系里面，考察三个平面：$S_1 : x - 3y - 2z = 3$，$S_2 : -2x + y - 4z = -9$ 与 $S_3 : -x + 3y - z = -7$ 的交点. 那么这个问题等价于解决如下三元一次方程组：

$$\begin{cases} x - 3y - 2z = 3 \\ -2x + y - 4z = -9 \\ -x + 3y - z = -7 \end{cases} \tag{2}$$

解得

$$\begin{cases} x = 2 \\ y = -1 \\ z = 1 \end{cases} \tag{3}$$

那么 $(2, -1, 1)$ 就是三维坐标系中平面 S_1，S_2 与 S_3 的交点.

一般情况下的定义

一般的，形如：

$$\begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n &= y_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n &= y_2 \\ \cdots \quad \cdots \quad \cdots & \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n &= y_m \end{cases} \tag{4}$$

的等式组统称为线性方程组，也可以根据其未知数的个数称为 n 元一次方程组.

其中 $x_1 \ldots x_n$ 为 n 个未知量，$y_1 \ldots y_m$ 与 $a_{1,1}, a_{1,2} \ldots a_{1,n}, a_{2,1} \ldots a_{n,m}$ 为给定的系数.（形如 $a_{i,j}$ 的系数表示它是方程组中第 i 个方程的 x_j 对应的系数，也即第 i 行第 j 列系数）

求解线性方程组

高斯消元法[477] 是求解线性方程通解的一个重要方法，适用于所有类型的解，包括无解，唯一解和无穷多个解. 另外高斯消元法还引入了线性代数中一种重要的变换即**行变换**.

对于有唯一解的情况，另一种方法是使用克莱姆法则[438]，但计算量要大许多，一般不对较大的方程组使用.

当全部 y_i 都为零时，我们把方程叫**齐次的（homogeneous）**（见[477]）. 对于 N 元 N 次线性齐次方程组，系数矩阵的行列式[432] 可以用于判断方程的解的个数：若行列式不为零，则方程有无穷多个解；若行列式为零，则方程有唯一解. 对于非齐次 N 元 N 次方程，若行列式不为零，方程有唯一解；若行列式为零，方程有无穷多个解或者无解. 这主要是因为行列式可以用于判断系数方阵中的行或列是否线性无关.

若想彻底了解线性方程组的解的结构，我们必须要从矢量空间的角度来理解[482]：首先我们需要先学习矢量空间[444]，子空间[464]，线性变换[451]，矩阵的秩[407]，然后才能进一步讨论为什么在不同情况下方程组的解会有不同的结构.

高斯消元法解线性方程组

预备知识 矩阵[396]

线性方程组（有 $x_1 \dots x_n$ 这 n 个未知量，所以也叫 **n 元一次方程组**）

$$\begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n &= y_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n &= y_2 \\ \cdots \quad \cdots \quad \cdots & \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n &= y_m \end{cases} \tag{1}$$

可以写成矩阵和列矢量相乘的形式

$$\mathbf{A}\mathbf{x} = \mathbf{y} \tag{2}$$

其中 $\mathbf{A}$ 是维度 $m \times n$ 的矩阵，称为**系数矩阵**.

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} \tag{3}$$

$\mathbf{x}$ 是 n 维列矢量 $(x_1, x_2, \dots, x_m, \dots, x_n)^{\mathrm{T}}$. $\mathbf{y}$ 是 m 维列矢量 $(y_1, y_2, \dots, y_m)^{\mathrm{T}}$，称为**常数列**. $\mathbf{A}$ 和 $\mathbf{y}$ 通常看做已知的，而 $\mathbf{x}$ 看做未知的，即方程组待求的**解**. 下面我们介绍一种解线性方程组的简单的方法，**高斯消元法（Gauss elimination）**. 先来看一个简单的例子.

例 1

我们先回顾一下初中阶段如何解线性方程组，例如

$$\begin{cases} 2x + 3y = 21 \\ 5x - 2y = 5 \end{cases} \tag{4}$$

一种方法是将第一条等式两边除以 2 再移项得到

$$x = -\frac{3}{2}y + \frac{21}{2} \tag{5}$$

再代入第二个条式消去 x 得

$$-\frac{19}{2}y + \frac{105}{2} = 5 \tag{6}$$

解得 $y = 5$，再代入第一条等式得 $x = 3$.

另一种更方便的解法是，将第一条等式（两边）乘以 $-5/2$，加到第二条上消去 x 得

$$-\frac{19}{2}y = -\frac{95}{2} \tag{7}$$

解得 $y = 5$，再代入第一条等式得 $x = 3$. 为什么可以这么做？简单来说是因为如果两条等式都成立，将它们两边相加得到的新的等式同样成立. 下面我们来详细讲解第二种方法.

为了书写简单，我们可以用所谓的增广矩阵（**augmented matrix**）来表示矩阵和常数列，即把式 4 记为

$$\left(\begin{array}{cc|c} 2 & 3 & 21 \\ 5 & -2 & 5 \end{array}\right) \tag{8}$$

同样把第一行乘以 $-5/2$，加到第二行上得

$$\left(\begin{array}{cc|c} 2 & 3 & 21 \\ 0 & -19/2 & -95/2 \end{array}\right) \tag{9}$$

如果这个方程有多个未知数，且方程的数量和未知数相同（系数矩阵为方阵），理想情况下我们可以用第一行消去所有 $i > 1$ 行的第一个系数，再用第二行消去所有 $i > 2$ 行的第 2 个系数，以此类推，最后得到一个三角形系数矩阵. 三角形系数矩阵的最后一行只有最后一个变量的系数不为零，我们求出这个变量后，再代入倒数第二行（只有两个未知量）求出另一个未知量，最后就可以得到所有未知量的值. 这个过程叫做反向迭代（**backward substitution**）.

注意这只是一种理想的情况，如果在处理第 i 行的时候发现 $a_{i,i} = 0$，则该方法无法进行下去. 为此我们需要稍微复杂一些的方法，也就是高斯消元法的一般步骤.

一般步骤

我们将式 1 形式的方程组用增广矩阵表示为

$$\mathbf{B} = (\mathbf{A}, \mathbf{y}) = \left(\begin{array}{cccc|c} a_{1,1} & a_{1,2} & \cdots & a_{1,n} & y_1 \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} & y_m \end{array}\right) \tag{10}$$

定义以下三种矩阵（或增广矩阵）变换为**初等行变换**．初等行变换不改变方程组的解[1]．

1. 对调矩阵中的第 i 行与第 j 行，记作 $\mathbf{r}_i \leftrightarrow \mathbf{r}_j$

2. 将矩阵第 i 行的所有元素乘以一个非零数 k，记作 $\mathbf{r}_i \times k$

3. 把矩阵第 i 行的所有元素乘以数 k 后加到第 j 行上，记作 $\mathbf{r}_i + \mathbf{r}_j \times k$

任何非零矩阵（或增广矩阵）经过有限次初等行变换后总可以使系数矩阵 $\mathbf{A}$ 转化为**梯形矩阵（echelon form）**．我们把梯形矩阵定义为满足该条件的矩阵：第 i 行的第一个非零系数[2] $a_{i,q(i)}$ 的列标 $q(i)$ 总是大于第 $i-1$ 行的第一个非零元 $a_{i-1,q(i-1)}$ 的列标 $q(i-1)$．与例 1 不同的是，当系数矩阵经过行变换化为梯形矩阵后，最后若干行有可能都为零，最后一个非零行也未必只有一个非零元．

高斯消元法的一般步骤如下：

- 先处理第 $i=1$ 行，如果 $a_{1,1}=0$ 但某 $i'>1$ 的行有 $a_{i',1}\neq 0$，就先进行行变换[3] $\mathbf{r}_i \leftrightarrow \mathbf{r}_{i'}$．如果第一列全为 0，我们就无视第 1 列，从第 2 列重新开始，以此类推．记此时第 1 行第一个非零元的列标为 $q(1)$．接下来做若干次行变换 $\mathbf{r}_{i'} + \mathbf{r}_1 \times k$ 使所有第 $i'>1$ 行的 $a_{i',p(1)}$ 都为 0．

[1] 这点我们留到以后证明

[2] 为了书写简单，我们将每次变换后的系数矩阵元仍然用 $a_{i,j}$ 来表示，虽然它的值可能已经发生变化．若需要区分，可以将第 n 次变换后的矩阵元用 $a_{i,j}^{(n)}$ 表示．

[3] 如果 $a_{1,1}\neq 0$ 则不需要

- 依次处理第 $i=2\dots m-1$ 行[4]．要处理第 i 行，先令 $q(i)=q(i-1)+1$，如果此时矩阵元 $a_{i,q(i)}=0$，但某 $i'>i$ 的行有 $a_{i',q(i)}\neq 0$，就先进行行变换 $\mathbf{r}_i\leftrightarrow\mathbf{r}_{i'}$．若不存在这样的 i'，我们就改令 $q(i)=q(i-1)+2$ 并重新开始该步骤，以此类推．接下来做若干次行变换 $\mathbf{r}_{i'}+\mathbf{r}_i\times k$ 使所有第 $i'>i$ 行的 $a_{i',p(i)}$ 都为 0.

完成后，系数矩阵就会变为梯形矩阵.

例 2　解线性方程组

$$\begin{cases} x_1 & +x_2 & -x_3 & +x_4 & = & 3 \\ 2x_1 & +2x_2 & -2x_3 & +x_4 & = & 7 \\ x_1 & +x_2 & & +2x_4 & = & 3 \\ 2x_1 & +2x_2 & -x_3 & +5x_4 & = & 4 \end{cases} \tag{11}$$

解：

将该方程组写成增广矩阵形式

$$\mathbf{B}=\left(\begin{array}{cccc|c} 1 & 1 & -1 & 1 & 3 \\ 2 & 2 & -2 & 1 & 7 \\ 1 & 1 & 0 & 2 & 3 \\ 2 & 2 & -1 & 5 & 4 \end{array}\right) \tag{12}$$

对第一列

$$\begin{array}{l} \mathbf{r}_2-2\mathbf{r}_1 \\ \mathbf{r}_3-1\mathbf{r}_1 \\ \mathbf{r}_4-2\mathbf{r}_1 \end{array} \quad\Longrightarrow\quad \mathbf{B}=\left(\begin{array}{cccc|c} 1 & 1 & -1 & 1 & 3 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 3 & -2 \end{array}\right) \tag{13}$$

发现此时，矩阵第二列自第二行以下全为零，所以需要依次向下一列寻找不为零的元素．继续消元

$$\begin{array}{l} \mathbf{r}_2\leftrightarrow\mathbf{r}_4 \\ \mathbf{r}_3-1\mathbf{r}_2 \\ \mathbf{r}_4-0\mathbf{r}_2 \end{array} \quad\Longrightarrow\quad \mathbf{B}=\left(\begin{array}{cccc|c} 1 & 1 & -1 & 1 & 3 \\ 0 & 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & -2 & 2 \\ 0 & 0 & 0 & -1 & 1 \end{array}\right) \tag{14}$$

[4]如果在处理 $i=m-1$ 之前就得到了梯形矩阵，则可提前终止.

$$\mathbf{r}_4 - 0.5\mathbf{r}_3 \quad \Longrightarrow \quad \mathbf{B} = \left(\begin{array}{cccc|c} 1 & 1 & -1 & 1 & 3 \\ 0 & 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & -2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \tag{15}$$

我们按照上面的方法用第 3 行求 x_4，代入第 2 行求得 x_3. 然而当我们想用第 1 行求 x_1 的时候却出发现我们还没求出 x_2. 解决办法是，我们令 $x_2 = c$ 且假设 c 可以取任意值，再把 x_1 用 c 表示. 方程的解可以表示为

$$\begin{cases} x_1 = 5 - c \\ x_2 = c \\ x_3 = 1 \\ x_4 = -1 \end{cases} \tag{16}$$

或者写成列矢量的形式

$$\mathbf{x} = c \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 5 \\ 0 \\ 1 \\ -1 \end{pmatrix} \tag{17}$$

将该式代入方程组，可以验证 c 取任意值时方程组都成立.

如果多个 x_i 的值都未知，我们就分别假设它们等于不同的任意常数 c_i 即可.

解的数量

根据以上步骤，当系数矩阵变为梯形矩阵后，可以用以下步骤判断解的数量：

1. 若存在系数 $a_{i,j}$ 全为零的行 i，但是对应的常数 y_i 却不为零，则方程组无解.

2. 若存在 d 个系数 $a_{i,j}$ 全为零的行 i，且对应的所有 y_i 也都为零，则方程有无穷个解，且需要 d 个任意常数来表示所有可能的解.

3. 若不存在系数 $a_{i,j}$ 全为零的行，则系数矩阵可以化为三角矩阵，使对角线上的元素全不为零. 此时方程有唯一解.

解的结构

按照高斯消元法的一般步骤，如果方程有解，我们总可以将解表示为一些常矢量的线性组合加上一个常矢量.

$$\mathbf{x} = \sum_{i=1}^{d} c_i \mathbf{x}_i + \mathbf{x}_0 \tag{18}$$

其中 c_i 是 d 个任意常数（当方程有唯一解时 $d = 0$），无论这些常数取什么值，$\mathbf{x}$ 都是方程的解. 另一方面，给出方程的任意一个解，总能找到一些常数 c_i 与之对应. 式 18 叫做方程的**通解（general solution）**，通解中的任意一个就做方程的**特解（special solution）**.

将式 18 代入方程组，使用矩阵乘法分配律（式 13 [396]）得到

$$\mathbf{A} \sum_{i=1}^{d} c_i \mathbf{x}_i + \mathbf{A}\mathbf{x}_0 = \mathbf{y} \tag{19}$$

由于 $\mathbf{x}_0$ 是方程的一个特解，有 $\mathbf{A}\mathbf{x}_0 = \mathbf{y}$. 所以

$$\mathbf{A} \sum_{i=1}^{d} c_i \mathbf{x}_i = \mathbf{0} \tag{20}$$

线性方程组 $\mathbf{A}\mathbf{x} = \mathbf{y}$ 中，如果 $\mathbf{y} = \mathbf{0}$ （粗体的 $\mathbf{0}$ 表示零矢量，即每个元都是 0），则方程组被称为是**齐次（homogeneous）**的. 上式说明 $\sum_i c_i \mathbf{x}_i$ 就是齐次方程组 $\mathbf{A}\mathbf{x} = \mathbf{0}$ 的通解.

综上所述，对于任意有解的**非齐次（inhomogeneous）**方程组 $\mathbf{A}\mathbf{x} = \mathbf{y}$，我们可以将通解（式 18）从形式上理解为齐次方程组 $\mathbf{A}\mathbf{x} = \mathbf{0}$ 的通解与非齐次方程组的任一特解相加.

线性方程组解的结构

- 本词条处于草稿阶段.

预备知识 矩阵与线性变换[456]，高斯消元法[477]，矩阵的秩[407]

线性方程组可以记为

$$\mathbf{A}\mathbf{x}=\mathbf{y} \tag{1}$$

其中 $\mathbf{A}$ 是 $M\times N$ 的矩阵，$\mathbf{x}$ 是 N 维列矢量，$\mathbf{y}$ 是 M 维列矢量，$\mathbf{A}\mathbf{x}$ 表示矩阵与列矢量相乘（式 4 [396]）．$\mathbf{A}$ 和 $\mathbf{y}$ 是已知的，$\mathbf{x}$ 是未知的，被称为方程组的**解（solution）**．

从矢量空间[444] 的角度来看，$\mathbf{x}$ 是一个 N 维矢量空间（以下称为 X 空间）中一个矢量关于某组基底的坐标，$\mathbf{y}$ 是一个 M 维矢量空间（以下称为 Y 空间）中一个矢量关于某组基底的坐标．矩阵 $\mathbf{A}$ 可以将 X 空间中的任意矢量映射到 Y 后的坐标．

我们知道 $\mathbf{A}$ 的第 i 列代表的矢量就是 X 空间中的第 i 个基底映射到 Y 空间的对应矢量．我们把 A 的 N 列对应的 N 个矢量记为 $\{\boldsymbol{\alpha}_i\}$．先来看一个定理

满秩方阵

我们知道矩阵的秩 R 等于线性无关的行数或列数[407]，下面来根据秩来分类讨论方程组的解空间结构．最简单的情况是 $\mathbf{A}$ 为满秩，即 $R=M=N$．这时由于 $\{\boldsymbol{\alpha}_i\}$ 两两线性无关，它们可以作为 Y 空间的一组基底，与 X 空间的基底一一对应．那么这个映射既是单射又是满射．对于 Y 空间的任意矢量 $\mathbf{y}$，X 空间都存在唯一的解 $\mathbf{x}$．特殊地，当 $\mathbf{y}=\mathbf{0}$ 时（即方程是**齐次**的），唯一解就是 $\mathbf{x}=\mathbf{0}$．

$R=M<N$

当 $\mathbf{A}$ 的秩等于 M 且小于 N 时，映射变为从 N 维空间到更小的 M 维空间．即非单射：虽然任意的 $\mathbf{x}$ 仍然映射到唯一的 $\mathbf{y}$，但任意的 $\mathbf{y}$ 却对应无穷多个 $\mathbf{x}$．

当方程是齐次的时候，零空间（定理 2 [451]）X_0 是 $N-M$ 维的（为什么？）．这种情况下，我们希望能解出零空间的 $N-M$ 个基底，使得这组基底的任意线性组合都是齐次方程的解．

对于非齐次方程，我们可以先求对应的齐次方程组的零空间的一组基底，再求出非齐次方程的任意一个解（**特解**），那么非齐次方程组的**解集**（所有解

的集合）就等于零空间中的所有矢量与特解相加．注意非齐次方程的解集并不构成一个矢量空间，因为它不包含零矢量（$\mathbf{x}=\mathbf{0}$ 总是对应 $\mathbf{y}=\mathbf{0}$，所以不可能是非齐次方程组的解），解集中若干矢量的线性组合也不一定仍然属于解集．

未完成：证明

$R<M$

当 $R<M$ 时，$\{\boldsymbol{\alpha}_i\}$ 中只有 R 个线性无关，它们在 Y 空间中张成[447] 一个 R 维子空间 Y_0．如果 $\mathbf{y}$ 在 Y_0 中（可以通过 $\mathbf{y}$ 是否与 $\{\boldsymbol{\alpha}_i\}$ 线性无关来判断），方程组就存在解，如果落在子空间外，方程组就无解．

线性映射的结构 2

预备知识 线性映射的结构[468]，线性方程组[475]

下面我们从线性映射和向量空间的角度理解线性方程组 $\mathbf{Ax}=\mathbf{b}$．

定义 1　线性方程

对给定的线性映射 $A:X\to Y$ 和 $b\in Y$，线性方程为

$$Ax=b \tag{1}$$

所有满足该式的 $x\in X$ 的集合 X_s 叫做方程的解集．

首先注意 A 未必把 Y 中的每个元素都射中，即值空间[451] $Y_1=A(X)\subseteq Y$ 只是 Y 的一个子空间．所以只有 $b\in Y_1$ 时式 1 有解，否则无解（解集为空集）．用映射的语言，解集 X_s 就是集合 $\{b\}$ 的逆像[98] $A^{-1}(\{b\})$．

当式 1 中 $b=0$ 时方程叫做算符 A 的**齐次方程**．根据定义，齐次方程的解就是映射的零空间（定理 2 [451]）．

定理 1

线性方程式 1 的解集可以表示为

$$X_s=X_0+x_1 \tag{2}$$

其中 x_1 为 X_s 中的任意元素，X_0 为映射的零空间．

说明：$X_0 + x_1$ 表示把 X_0 中的每一个向量与 x_1 相加得到的集合．易证当 $x_1 \neq 0$ 时解集 X_s 不是一个向量空间（例如不存在零向量）．

首先证明集合 $X_0 + x_1$ 中的元素满足 $Ax = b$．令任意 $x_0 \in X_0$

$$A(x_0 + x_1) = Ax_0 + Ax_1 = 0 + b = b \tag{3}$$

证毕．再来证明解集中不存在 $X_0 + x_1$ 之外的矢量．令 $x_2 \in X_S$ 且 $x_2 \neq x_1$，那么

$$A(x_2 - x_1) = Ax_2 - Ax_1 = b - b = 0 \tag{4}$$

即 $x_2 - x_1 \in X_0$，即 $x_2 \in x_1 + X_0$．证毕．

线性映射的结构

到此为止我们就可以非常清晰地勾画出多对一线性映射的结构了 $A : X \to Y$．我们先找到定理 1 [468] 中的零空间 X_0 和它的补空间 X_1，其中 X_1 的元素和值域空间 $Y_1 = A(X)$ 一一对应．那么对每个 $x_1 \in X_1$，线性映射就会把集合 $x_1 + X_0$ 所有元素映射到 Y_1 中的同一个元素 $y_1 = Ax_1$ 上．

未完成：需要一个 3 维几何矢量空间的真实例子，零空间 1 维，补空间 2 维

超定线性方程组

预备知识 最小二乘法[267]，线性方程组与矢量空间[482]

令 A 为 $M \times N$ 的复数矩阵，$\mathbf{x}$ 和 $\mathbf{y}$ 为复数列矢量，当 $M > N$ 时，以下线性方程组称为**超定方程组**（只有 $\mathbf{x}$ 是未知）

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{1}$$

我们把 $\mathbf{y}$ 和 $\mathbf{A}$ 拼接成一个 $M \times (N+1)$ 的矩阵，当这个矩阵的 M 个行矢量中只有小于或等于 N 个线性无关时，我们只需取所有线性无关的行即可得到非超定的线性方程组．举一个简单的例子，如果第 2 条方程（第 2 行）是第 1 条方程（第 1 行）乘以常数，那么这两条方程中我们只需保留一条即可．

如果有大于 N 个线性无关的行（由于每行只有 $N+1$ 个元，那么最多只可能有 $N+1$ 个线性无关的行），那么超定方程无解．为什么？

但我们仍然可以寻找一个最优的 $\mathbf{x}$，使以下误差函数取最小值

$$|\mathbf{A}\mathbf{x}-\mathbf{y}|^2=\sum_k\left(\sum_j A_{kj}x_j-y_k\right)\left(\sum_j A_{kj}x_j-y_k\right)^* \tag{2}$$

所以这是一个最小二乘法问题．令误差函数分别对每个 $\mathrm{Re}[x_i]$ 和 $\mathrm{Im}[x_i]$ 求导等于 0，得

$$\sum_j\left(\sum_i A_{ik}^\dagger A_{kj}\right)x_j=\sum_k A_{ik}^\dagger y_k \tag{3}$$

即

$$\mathbf{A}^\dagger\mathbf{A}\mathbf{x}=\mathbf{A}^\dagger\mathbf{y} \tag{4}$$

通常情况下该方程只有一个解，也就是最小二乘法的解．

对比式 1 可以发现式 4 只是在左右两侧同时乘以 $\mathbf{A}$ 的厄米共轭．所以任何能满足式 1 的解也可以通过式 4 解得．

第五章

张量代数

张量

预备知识 矩阵[396]

“张量”一词最早出现在结构力学里，用来描述应力的性质，因此被叫做**张力（tensor）**，这个名字沿用到了今天，只是翻译改成了**张量**．现在我们知道，张力只是一种更一般的数学对象，张量，的特例，而这种对象在物理学的各个角落频繁出现．因此，我们会从现代角度，从根本上定义张量的概念，并展示如何将张量表示成常见的多维矩阵的形式．为了便于理解本条内容，读者应有扎实的线性代数功底．本词条所介绍的是最根本的张量的定义，没有附加任何其它结构或限制，也暂未使用爱因斯坦求和约定，以方便初学者入门．

简介

张量本身，是一种多重线性映射[451]．特别地，二阶张量可以看成是把两个向量映射成一个数字，也可以看成是把一个向量映射成一个向量．这里的数字是在定义以上向量的域[588] 中的，在物理中一般是实数域或复数域．

许多物理量都可以看成是线性映射．比如说，质量可以看成一个张量，它把加速度向量映射成一个力向量；转动惯量[962] 或惯性张量[973] 也可以看成张量，它把角速度向量映射成一个角动量向量．张量的阶数不同，性质看起来也不同，但它们都有一个共性，即都是线性（式 2 [451]）的．在未来的学习中遇到张量时，不妨思考“这个张量是为了把什么映射为什么而定义的”，有利于加深理解．

1-线性函数，2-线性函数以及它们的表示

我们知道，给定两个向量空间 V 和 W，我们可以构造从 V 到 W 的线性映射．在这种情况下，自变量变量都也只有一个元素．多重线性映射，比如 k 重线性映射，就是拿一系列 $V_1, V_2, \cdots, V_k$ 来进行一个**多元映射**，让这个多元映射中的某一个自变量 $v_i \in V$ 单独变化时，映射看起来就像一个普通的线性映射．

最简单的情况，就是考虑 W 是一个一维空间．在这种情况下，W 也可以看作实数域本身，而不是线性空间．将集合中的元素映射到数字集合上的映射，

我们用一个特殊的术语来称呼，**函数**，也就是中学课程中函数一词，毕竟中学所涉及到的映射基本上都是实数到实数的映射，自然被称为函数．那么当线性映射的目标是一维的 W 时，由于此时 W 被看成实数域，我们特别地把这个映射称为**线性函数**．下面，我们就要用线性函数的概念来一步步引入多重线性映射和张量的概念及其表示．

线性函数用向量表示

定义 1　线性函数

给定实数域 $\mathbb{R}$ 上的 n 维线性空间 V，称 $f:V\to\mathbb{R}$ 为 V 到域 $\mathbb{R}$ 上的一个线性函数，如果 f 满足以下线性性：对于任意的 $v_1,v_2\in V$ 和 $a_1,a_2\in\mathbb{R}$，都有 $a_1f(v_1)+a_2f(v_2)=f(a_1v_1+a_2v_2)$．

如果把域 $\mathbb{R}$ 本身看成一个一维的线性空间，那么线性函数就是 V 到这个一维空间上的线性映射．因此，我们只需要任取一个 V 的基，研究基向量被 f 映射到哪里，就可以计算出任意的 $v\in V$ 被映射到哪里了．

设 V 有基 $\{e_i\}_{i=1}^n$．如果基向量 e_i 被映射为 $f(e_i)=m_i\in\mathbb{R}$，那么对于任意的向量 $v=a_1e_1+\cdots+a_ne_n$，根据线性性可以得到：

$$\begin{aligned}f(a_1e_1+\cdots+a_ne_n)&=f(a_1e_1)+\cdots+f(a_ne_n)\\&=m_1a_1+\cdots+m_na_n\end{aligned}\tag{1}$$

其中 $m_1a_1+\cdots+m_na_n$ 可以看成是向量 m 点乘[471] v 所得到的，其中 $m=m_1e_1+m_2e_2+\cdots+m_ne_n$．

也就是说，每个线性函数 f 对应一个向量 m，使得 $f(v)=m\cdot v$．这个向量 m，在以上给定的基 $\{e_i\}_{i=1}^n$ 中的坐标为 $(m_1,m_2,\cdots,m_n)^T$．当然，用别的基表达，坐标也会不一样，但向量还是那个向量．

由于线性函数是一个向量空间到实数上的映射，我们也将其称为一个 **1-线性函数**，以和其它多重线性函数区分，比如接下来讨论的 2-线性函数．

2-线性函数用矩阵表示

继续使用上述给定的 n 维线性空间 V，但是我们现在拿出 2 个 V 来构造映射 $f:V\times V\to\mathbb{R}$．

定义 2　2-线性函数

称 $f: V \times V \to \mathbb{R}$ 为一个 2-线性映射，如果对于任意固定的 v_0，$f(v_0, v)$ 和 $f(v, v_0)$ 都是 v 的线性函数，那么称 f 是一个 V^2 上的 2-线性函数或双线性函数（**bilinear function**）.

双线性函数的另一种表达方法，是

$$\begin{aligned} & f(a_1 v_1 + a_2 v_2, b_1 u_1 + b_2 u_2) \\ = & \, a_1 f(v_1, b_1 u_1 + b_2 u_2) + a_2 f(v_2, b_1 u_1 + b_2 u_2) \\ = & \, a_1 b_1 f(v_1, u_1) + a_2 b_1 f(v_2, u_1) + a_1 b_2 f(v_1, u_2) + a_2 b_2 f(v_2, u_2) \end{aligned} \tag{2}$$

为了表示这个映射，我们需要给两个 V 取定基．不过这里有两个 V，所以可以取两组不同的基，各 V 各自用一个单独的基．我们设第一个 V 的基是 $\{e_i\}_{i=1}^n$，第二个的是 $\{e'_i\}_{i=1}^n$．选定各自的基以后，向量都可以表示成其坐标构成的列向量，而双线性函数就被表示为一个矩阵（以 2 维的 V 为例）：

$$\begin{aligned} v = a_1 e_1 + a_2 e_2 \to \mathbf{c}_v &= \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \\ u = b_1 e'_1 + b_2 e'_2 \to \mathbf{c}_u &= \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \\ f \to \mathbf{M} &= \begin{pmatrix} f(e_1, e'_1) & f(e_1, e'_2) \\ f(e_2, e'_1) & f(e_2, e'_2) \end{pmatrix} \end{aligned} \tag{3}$$

这样，我们就有 $f(v, u) = \mathbf{c}_v^T \mathbf{M} \mathbf{c}_u$．注意这里 $\mathbf{c}_v^T$ 表示 $\mathbf{c}_v$ 作为矩阵的转置[1].

习题 1

把列向量看成矩阵，根据矩阵的运算法则，将式 3 代入式 2，验证 $f(v, u) = \mathbf{c}_v^T \mathbf{M} \mathbf{c}_u$.

上面给出的例子是 2 维的 V 的情况．推广到 n 维的 V 时，$\mathbf{M}$ 要扩展成一个 $n \times n$ 的矩阵，其中第 i 行 j 列的元素是 $m_{ij} = f(e_i, e'_j)$.

我们来看一个简单的 2-线性函数的例子.

[1]如果使用其它的基，那么向量 v、u 的坐标和映射 f 的矩阵会有不同表示，但是计算出来的 $\mathbf{c}_v^T \mathbf{M} \mathbf{c}_u$ 仍然是一致的.

例 1　功率

三维空间里一个物体的受力，可以构成一个三维实线性空间 V，这个物体的速度，也可以构成一个 V[2]. 用 F 表示物体受到的力，v 表示物体的速度，那么这个力对物体做功的功率 P 就是一个双线性函数：$P = F \cdot v$. 如果选定了受力空间和速度空间的基，还可以把向量 F 和 v 分别表示成列向量 $\mathbf{c}_F$ 和 $\mathbf{c}_v$，那么就有 $P = \mathbf{c}_F^T \mathbf{P} \mathbf{c}_v$，这里的 $\mathbf{P}$ 就是一个双线性函数在选定基底下的矩阵，它的作用是表示功率的计算方式. 如果我们按以下方式选择基：受力空间的基是沿着 x、y 和 z 方向的大小为 $1N$ 的力，而速度空间的基是沿着 x、y 和 z 方向的大小为 $1\,\mathrm{m/s}$ 的速度，那么在这两组基下，表示功率的双线性函数的矩阵就是单位矩阵 $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$；如果把受力空间的基改为沿着 y、x 和 z 方向，那么表示功率的双线性函数的矩阵就变成了 $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. 此时，如果再把速度空间的基改为沿着 x、z 和 y 方向，那么功率函数的矩阵就变成了 $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

可以看到，功率函数把受力、速度两个向量映射为一个实数，功率. 选定受力空间和速度空间的不同基，可以得到功率函数的不同矩阵表达. 这些矩阵表达了功率函数的特点，却随着基的变化而变化. 类比线性代数中所学的线性映射，都可以表示为一个矩阵，虽然矩阵随着基的选取的不同可能也会不同，但是这些矩阵都表示同一个线性映射. 同样，我们也可以把功率函数看成本身不变的，只是其矩阵表示随着基的选取不同而可能不同.

矩阵的拓展

为了方便推广 2-线性函数的概念，我们需要用到向量矩阵的概念，因此要简单回顾一下矩阵运算[396]. 矩阵只是一种运算的表达方式. 虽然我们常见的矩阵元素都是实数或者复数，但是只要是可以相加和相乘的元素都可以当作矩

[2]回忆：$\mathbb{R}$ 上的线性空间，只要维度相同，都可以看成同一个空间.

阵元素，例如行向量、列向量都可以作为矩阵的元素.

例 2　矩阵运算的例子

- 向量可以和数字相乘. 如果 v_i 表示向量，a_i 表示数字，那么 $\begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix}$ 是三个向量排成的矩阵，$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ 是三个数字排成的矩阵，按照向量的数乘来进行矩阵乘法，$\begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = (a_1v_1 + a_2v_2 + a_3v_3)$ 就是一个向量，而 $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} = \begin{pmatrix} a_1v_1 & a_1v_2 & a_1v_3 \\ a_2v_1 & a_2v_2 & a_2v_3 \\ a_3v_1 & a_3v_2 & a_3v_3 \end{pmatrix}$ 是 9 个向量排成的矩阵.
- 向量之间可以有点乘. 因此 $\begin{pmatrix} v_1 & v_2 \\ v_3 & v_4 \end{pmatrix} \cdot \begin{pmatrix} v_5 \\ v_6 \end{pmatrix} = \begin{pmatrix} v_1 \cdot v_5 + v_2 \cdot v_6 \\ v_3 \cdot v_5 + v_4 \cdot v_6 \end{pmatrix}$ 是两个数字排成的矩阵.
- 3 维向量之间可以有叉乘. 因此 $\begin{pmatrix} v_1 & v_2 \\ v_3 & v_4 \end{pmatrix} \times \begin{pmatrix} v_5 \\ v_6 \end{pmatrix} = \begin{pmatrix} v_1 \times v_5 + v_2 \times v_6 \\ v_3 \times v_5 + v_4 \times v_6 \end{pmatrix}$ 是两个向量排成的矩阵.

2-线性函数的向量矩阵表示

设一个 2-线性函数 $f : V^2 \to \mathbb{R}$ 在基 $\{e_i\}_{i=1}^n$（为方便计，两个 V 用同一组基）下的矩阵表达是

$$\begin{pmatrix} m_{ij} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{pmatrix} \tag{4}$$

则把每一行当成一个行向量，记为 $\mathbf{m}_i = \begin{pmatrix} m_{i1} & m_{i2} & \cdots & m_{in} \end{pmatrix}$，那么 f 可以表达为行向量的行矩阵：

$$\mathbf{M} = \begin{pmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \cdots & \mathbf{m}_n \end{pmatrix} \tag{5}$$

对于任意向量 $v, u \in V$，如果它们在同一个基下的列向量是 $\mathbf{c}_v$ 和 $\mathbf{c}_u$，那么 $\mathbf{M}\mathbf{c}_v$ 是一个行向量，$\mathbf{M}\mathbf{c}_v\mathbf{c}_u$ 是一个 $\mathbb{R}$ 中的数[3]．

因此，$\mathbf{M}$ 可以表示双线性函数 f，规则是：$f(v,u) = \mathbf{M}\mathbf{c}_v\mathbf{c}_u$．

你可能注意到了，2-线性函数的表示开始变得复杂了．好消息是，在后续的小节中我们会介绍爱因斯坦求和约定[498]，这是一个能大大简化以上表达的规则．

把线性函数称作张量

一个 2-线性函数，又被称为一个**二阶张量**．同样地，一个 1-线性函数被称为一个**一阶张量**．由此还可以推广，一个实数被称为一个**零阶张量**．这就是现代数学中张量的定义之一．所谓的 k-阶张量就是一个 k-线性函数．

不过，张量不仅仅是到数域上的映射．换句话说，张量不仅仅是线性函数，准确来说应该是线性映射．我们接下来就要深入讨论这一点．

重新思考 2-线性函数

预备知识 惯性张量[973]

我们再次观察一下 2-线性函数．虽然我们定义 2-线性函数是把两个向量映射到一个数字，但是在上面“2-线性函数的向量矩阵表示”的例子中我们也发现，一个 2-线性函数也可以看成是把一个向量映射到一个 1-线性函数上．比如说，例子中的函数，在给定的基下，就把 $\mathbf{c}_v$ 映射到了 $\mathbf{M}\mathbf{c}_v$ 上，而当你把 $\mathbf{M}\mathbf{c}_v$ 当成一个整体的时候，会发现它是关于 $\mathbf{c}_u$ 的一个 1-线性函数．由于 1-线性函数可以对应成一个向量，我们也可以把一个 2-线性函数看成是把一个向量映射到一个向量上．

总而言之，一个 2-线性函数可以看成**是把两个向量映射到一个数字**，也可以是**把一个向量映射到一个 1-线性函数上**，还可以是**把一个向量映射到一个**

[3]此处把向量 $\mathbf{m}_1^T$ 看成 $\mathbf{m}$ 的元素，向量 v, u 看成列矩阵，进行矩阵乘法，应用数乘来乘以每一个数和向量 $\mathbf{m}_1^T$．

向量上. 最后一种解读方式是不是很眼熟？没错，这种语境下，它就是线性代数研究了很久的“线性变换”或“两个线性空间之间的线性映射”.

有了这个认识，我们就可以举出更多的 2-线性映射的例子了.

例 3　转动惯量

角速度被定义为一个向量，其大小就是角速度的大小，方向则按照右手定则确定. 给定参考点，一个质点的角动量被定义为 $r \times mv$，其中 r 是从参考点到质点位置的位移向量，m 是质点的质量，v 是质点的速度向量. 一个质点系统的角动量，定义为所有质点的角动量之和.

对于一个质量均匀分布的球体，由于它高度对称，如果它的角速度是 ω，那么它的角动量也沿着 ω 的方向，因此可以表示为 $L = I\omega$，其中 L 是球体的角动量，数字 I 被称为转动惯量[962].

但是，如果一个刚体是 x 轴上一对对称质点构成的，它们的质量都是 m，到原点的距离都是 r，那么当这个刚体的角速度在 y, z 方向上时，角动量非零，且和角速度同向；当角速度在 x 方向上时，角动量为零. 也就是说，转动惯量一般没法简单表示为一个实数 I，而应该用一个 2-线性函数来表示，即惯性张量[973]. 在这个例子中，在给定的直角坐标系上，惯性张量表示为一个矩阵

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 2mr^2 & 0 \\ 0 & 0 & 2mr^2 \end{pmatrix} \tag{6}$$

拿出你的手机，把它近似看成一个长方体. 取一个直角坐标系，使得三个坐标轴分别垂直于手机的正面、侧面和顶面，那么你的手机的转动惯量就可以表示成

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \tag{7}$$

其中通常有 $a > b > c$，因为你的手机的厚度从薄到厚依次是正面、侧面和顶面对应的方向. 如果选用其它坐标系，这个转动惯量通常就是另一个矩阵的样子，但不管选什么坐标系表示手机的转动惯量矩阵，它们都是相似的.

接下来一个例子也是极为常用的二阶张量.

例 4　电极化率

我们简单介绍一下电极化率的概念．如果你已经熟悉电极化率的概念，可以跳过介绍部分（从图 1 之后看起）；如果你想了解电极化率的更多相关知识，请参考电动力学相关词条．

电偶极子是指一对电荷量相同的正负电荷构成的微观系统，它有一个性质叫电偶极矩，是一个向量．如果两个电荷分别带电荷 $\pm q$，而从负电荷到正电荷的位移是 r，那么我们定义此电偶极矩就是 $p = qr$．当然，宏观上我们可以认为这两个电荷位置实际上是重合的，但仍然具有电偶极矩 p，也就是所谓的“理想电偶极子”．

从数学上来说，电荷是点粒子的一个性质，它是一个数字；偶极矩也是点粒子[4]的一个性质，它是一个向量．

一些不对称的分子中，电子云的电荷中心可能和原子核的电荷中心不重合，一个这样的分子就可以看成是一个电偶极子．如果我们把一定区域里的电偶极子的电偶极矩都按照向量加法加起来，得到的总向量再除以这个区域的体积，我们就得到这个区域里电偶极矩的密度；当选定区域趋于一个点时，还能得到电偶极矩在一个点上的密度．当然，这里假设电偶极矩是连续的，因为宏观实验上所说的点，其实并不是严格意义上的点，也包含了大量的电偶极子．

一般情况下，即使有大量电偶极子存在，物质也不会显出电性，因为各个地方的电偶极子都互相抵消了，使得电偶极矩的密度[5]处处为零．但是如果物质处在电场下，就不一定了．正电荷趋于往电场方向移动，负电荷趋于往电场方向的负方向移动，于是电偶极子就趋于让自己的方向和电场方向平行．当然，电偶极子的长度也会发生变化，改变电偶极矩的大小，但是这种改变对电偶极矩密度的改变相比转向带来的改变，微不足道[6]．当原本指向均匀地分布在各个方向上的电偶极子开始指向同一个方向的时候，各个点上就会出现非零的电偶极矩了，物质也开始显出电性．

电偶极矩密度不为零的时候，物质里会出现束缚电荷，但这不是我们目前关心的．我们目前关心的是，电偶极矩密度是怎么被电场影响的．

对于可以自由运动的电偶极子，比如水分子，这个规律其实非常简单：记

[4]当我们把电偶极子的两个电荷视为在同一位置的时候，我们就把它看成在这个位置的一个粒子了．

[5]电动力学中，将电偶极矩的密度称为电极化强度（polarization）．

[6]对于没有自带电偶极矩的分子，或者单独的原子，转向效应是不存在的，这个时候只剩下电偶极矩被拉伸带来的效应了．

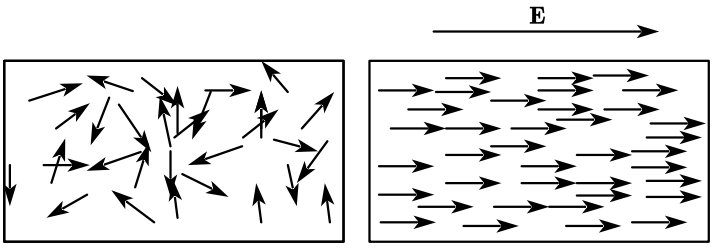


图 1: 电场引起电偶极矩密度变化的示意图．左图中物质里的电偶极子方向随机均匀分布，整体上不显电性；右图中物质里分布着水平向右的电场，在电场作用下电偶极子们都指向水平向右了，出现了水平向右的电偶极矩密度，宏观上开始表现出电性（存在束缚电荷）．

P 是某一点的电偶极矩密度，E 是物质内的电场分布，那么就有 $P = \epsilon E$，其中 ϵ 是一个实数．也就是说，P 总是和 E 平行，并且大小之比恒为 ϵ．这个时候，我们可以说水具有电极化张量 $\mathbf{M}$，使得 $P = ME$．如果按照国际单位制，把 E 和 P 的基都选为同一组直角坐标系上的单位向量，那么水的电极化张量就可以表示为

$$\begin{pmatrix} \epsilon & 0 & 0 \\ 0 & \epsilon & 0 \\ 0 & 0 & \epsilon \end{pmatrix} \tag{8}$$

这是单位矩阵的 ϵ 倍．

但是电极化张量并非总是这么简单的．如果在一个晶体里，固有的电偶极子都是指向水平方向的，但是水平向右和向左的电偶极子互相抵消，总体上还是不显电性；同时，这些电偶极子无论如何都被限制在水平方向上，那么它们就无法转向．当晶体中存在水平向右的电场 E 的时候，水平向右的电偶极子被拉长、向左的被缩短，于是出现了水平向右的电偶极矩密度 $P = \epsilon E$．如果晶体中的电场不在水平方向，那么由于电偶极矩密度还是限制在了水平方向上，因此电场和电偶极矩密度不再平行了．按照国际单位制，把 E 和 P 的基都选为同一组直角坐标系上的单位向量，那么这个晶体里的电极化张量就变成了

$$\begin{pmatrix} \epsilon & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{9}$$

它不再是单位矩阵的 ϵ 倍.

实际情况下，由于晶格对晶体中电荷在不同方向上的运动有不同程度的阻碍，晶体里的电极化张量不一定是单位矩阵的倍数. 不过，由于改变基时矩阵变为相似矩阵，那么如果电极化张量在某个基下的矩阵表示是可对角化的，那么总能找到 E 的一组基，使得电极化张量的矩阵是一个对角矩阵；换句话说，沿着这基向量方向上的电场，引起的电偶极矩密度依然和电场平行，只不过不同的方向上倍数不一样.

k 阶张量

k-线性函数

把 2-线性函数的定义推广，就得到 k-线性函数的定义.

习题 2

请尝试给出 k-线性函数的定义. 答案就在题目下方.

定义 3　k-线性函数

给定域 $\mathbb{R}$ 上的 n 维线性空间 V，称 $f : V^k \to \mathbb{R}$ 为 V 到域 $\mathbb{R}$ 上的一个 k-线性函数或线性泛函（**linear functional**），如果 f 对于每一个自变量都是线性的；或者说，当 k 个自变量中只有一个在变化，其它分量不变，那么 f 对于那个变化的自变量是线性函数. 特别地，线性函数又被称为 1-线性函数.

有了 k-线性函数的概念，我们就可以定义 k-阶张量了：

定义 4　张量

一个 k-线性函数，被称为一个 k-阶张量.

张量的阶数

从前面的讨论我们知道，一个二阶张量还可以看成是把一个向量映射到一个向量；向量可以和一阶张量一一对应，因此二阶张量也可以看成是把一个一阶张量映射到一个一阶张量. 同样地，一个三阶张量可以看成是把两个线性函数映射到一个一阶张量.

事实上，一个 k-阶张量 $f(v_1, v_2, \cdots, v_k)$ 是 k-线性函数，固定住其中几个自变量，比如说 v_1 到 v_m，那么 $f(v_1, v_2, \cdots, v_k)$ 可以看成是剩下 $(k-m)$ 个自

变量的 $k-m$-线性函数 $g(v_{m+1},\cdots,v_k)$. 也就是说，f 把前 m 个向量映射成一个 $k-m$ 线性函数. 因此，我们也可以说，一个 k 阶张量把 m 个向量映射为一个 $(k-m)$-线性函数. 再换一个思路来看，一个 k 阶张量是把 m 个向量映射成 $k-m$ 个向量.

以上观察意味着，我们有一个非常简单的法则来判断一个张量的阶数：阶数就是涉及的向量空间的数目. 比如说，如果一个张量把 3 个向量映射为 2 个向量，那么它的阶数就是 5；如果一个张量把 2 个向量映射为一个数字，那么它的阶数就是 2.

小结与拓展

一个 n 维的 k 阶张量，可以表示为一个 k 维的超立方体矩阵，其中每条边包含了 n 个实数.

我们在之前的定义中都把线性空间看成是定义在实数域 $\mathbb{R}$ 上的，而线性函数是线性空间映射到实数. 实际上，这一点可以推广到，把线性空间定义在任何域 $\mathbb{F}$ 上，线性函数就是线性空间映射到这个域 $\mathbb{F}$ 上. 这样也可以讨论线性空间上的张量.

目前物理学中常用到的域，是实数域 $\mathbb{R}$ 和复数域 $\mathbb{C}$.

像转动惯量这样的物理量，为什么可以看成线性映射呢？这是因为物理量彼此不是独立存在的，多个物理量之间相互作用而影响了其它的物理量，本身就是一种多个物理量映射到其它物理量或它们的性质的过程，就像牛顿第二定律可以看成是把加速度映射到力上，而质量的作用就是作为映射的函数. 因此，物理量本身也可以看成是映射的过程，而在许多物理理论中，我们都可以线性地描述这些映射关系，因此常常可以把物理量看成是线性映射.

爱因斯坦求和约定

预备知识 对偶空间[460]

引入的动机

在线性代数中，我们通常研究的是向量、对偶向量和线性映射等对象，也就是最多涉及二阶及以下的张量．在这种情况下，纸面上可以很方便地写出低阶张量的矩阵形式，毕竟零阶张量的坐标就是一个 1×1 矩阵、一阶张量的坐标就是一个 $1\times n$ 或 $n\times 1$ 矩阵，二阶张量的坐标就是一个 $n\times n$ 矩阵．

在张量代数和一切需要大量应用张量的理论，如微分几何中，我们不可避免地要经常涉及高阶的张量，它们的坐标就没法用矩阵表示．我们当然可以把矩阵拓展为立体阵等概念，但随着阶数上升，这种表示法的复杂程度几何级增加；我们也可以使用**张量**[488] 词条中所提过的向量矩阵的方法，比起立体阵要清楚一些，但套娃式的表达方式也对理解一个张量的性质造成了障碍．

爱因斯坦求和约定正是为了简洁地表达高阶张量的坐标运算而存在的．

定义和例子

在爱因斯坦求和约定中，各维度矩阵的元素用上下标来表示．具体来说，对于二维的矩阵，上标表示行数，下标表示列数，于是 a^i_j 就表示“第 i 行第 j 列的元素”. 进一步，我们也可以直接用 a^i_j 来表示一个二维矩阵. 在“张量[488]”词条的式 5 中我们引入了“行矩阵的行矩阵”这一类**套娃**概念来描述二阶张量——比如说，一个“列矩阵的列矩阵”就可以表示为 a^{ij}，各 a^{ij} 就表示这个矩阵中“第 i **列**的矩阵的第 j **列**的元素”.

由此可见，张量[488] 词条中想方设法进行的复杂表示，都可以简洁地用上下指标来描述．“行矩阵的行矩阵”这样套娃的表达，用指标来表示其实就是 m_{ij}，两个下标．在这里，m_{ij} 中的 i 表示第 i 个行矩阵，而 j 表示这个行矩阵里的第 j 个元素——由此可见，两个下标的地位是不一样的，先后次序很重要．因此在现代数学和物理的文献中，为了减少歧义，也通常明确地写出两个坐标的次序，比如 $m^i{}_j{}^k$ 就表示三个指标的优先顺序分别是 i、j、k．

为了把矩阵乘法的规则推广到上下指标的描述法，我们有了爱因斯坦求和约定．

用一句话来总结爱因斯坦求和约定，就是：**当式子中任何一个指标出现了两次，并且一次是上标、一次是下标时，那么该式表示的实际上是对这个角标一切可能值的求和**．换言之，如果角标 i 作为上标和下标各出现了一次，那么

式子相当于添加了一个关于 i 的求和符号 $\sum_i$.

我们举例来说明：

例 1　线性函数

从张量[488] 中我们知道，一个 1-线性函数可以表示为一个向量，这样的向量常被称为余向量、补向量或者 1-形式. 通常，我们用下标来表示一个余向量的各分量：$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$；而用上标来表示一个通常的几何向量：$\mathbf{v} = (v^1, v^2, v^3)$. 注意，上标不是乘方.

$\boldsymbol{\alpha}$ 和 $\mathbf{v}$ 的内积是

$$\sum_{i=1,2,3} \alpha_i v^i$$

用爱因斯坦求和约定，$\boldsymbol{\alpha}$ 和 $\mathbf{v}$ 的内积就可以写为 $\alpha_i v^i$.

例 2　矩阵运算

对于矩阵 $\mathbf{A}$，我们把其第 i 行第 j 列的元素表示为 A^i_j.

- 矩阵乘法表示为：如果 $\mathbf{A} = \mathbf{BC}$，那么 $A^i_j = B^i_k C^k_j$.
- 矩阵 $\mathbf{A}$ 的迹为 A^i_i.

由于重复出现而实际上应该是求和的指标，被称为**赝指标**或者**哑指标**（**dummy index**），因为它们不是真正的指标，而是可以用任意字母代替的. 没有求和的指标是固定的，是真正的指标. 比如说，$B^i_k C^k_j$ 中 k 可以是任何字母，但是 i 和 j 是不可以替换成别的字母的，因为它们由 A^i_j 决定了. 在这里，哑指标实际上是表示遍历全部可能的真指标.

求和的哑指标不一定是遍历 $1, 2, 3$，也可能是对更多或者更少的指标求和，甚至是无穷集合. 指标的范围要根据具体情境来决定.

爱因斯坦求和约定的表示方法脱胎于矩阵乘法的要求，但是却不依赖于矩阵行和列的形式，转而关注指标间的配合，相比传统的矩阵表达，能更方便地推广到高阶张量的情形中.

若干约定俗成的表示

- 向量的坐标分量用上标表示，对偶向量的坐标分量用下标表示，这对应于对偶空间[460] 词条中说的“向量坐标表示为列矩阵，对偶向量坐标表

示为行矩阵”．用物理学中更常用的术语来说，逆变向量分量用上标表示，协变向量分量用下标表示．

- 在闵可夫斯基时空中，x^0 表示时间坐标 t，而 x^1，x^2 和 x^3 分别表示空间坐标 x，y 和 z．有时候还会使用 x^4 来表示虚时间分量 $\mathrm{i}t$．注意这里取了光速 $c=1$．

抽象指标

在涉及微分几何时我们常使用抽象指标，这是爱因斯坦求和约定的延伸，请参见**抽象指标**[694] 词条．

如有约定俗成的补充，请继续添加．

张量的坐标变换

预备知识 张量[488]，过渡矩阵[414]，爱因斯坦求和约定[498]

在**张量**[488] 词条中我们看到，张量表示为矩阵依赖于相关的各线性空间中基的选择．选定了线性空间的基以后，线性空间中的各向量都可以用其坐标所构成的列矩阵来表示；同样地，选定了两个线性空间各自的基以后，一个二阶张量也可以表示为一个矩阵．一般地，我们把选定了线性空间各自的基以后用来表示张量的矩阵，称作张量的**坐标**，**分量**或**坐标分量**．

正如**爱因斯坦求和约定**[498] 一节中所说，零阶张量的坐标表示为一个 1×1 矩阵，一阶张量的坐标表示为一个 $n\times 1$ 或 $1\times n$ 矩阵，一个二阶张量的坐标表示为一个 $n\times n$ 矩阵．三阶及以上的张量就没法用我们已经熟知的矩阵来表示和运算了，因此我们不可避免地要应用爱因斯坦求和约定．从初学者的角度出发，本节将先用传统线性代数的语言描述一阶、二阶张量的坐标变换规则，然后使用爱因斯坦求和约定来描述一般的张量坐标变换规则．如果你还不熟悉爱因斯坦求和约定，建议详细阅读，比对两种表示方法的异同．

传统记号表示的一阶、二阶张量坐标变换

一阶张量的坐标变换

一阶张量是将一个向量映射为一个数，因此只涉及一个线性空间，最为简单.

给定 k 维线性空间 V，及其上一个张量 $f : V \to \mathbb{R}$. 如果 V 的基是 $\{\mathbf{e}_1, \cdots \mathbf{e}_k\}$，那么坐标为 $\mathbf{c} = (x_1 \cdots x_k)^{\mathrm{T}}$ 的向量 $\mathbf{v}$ 被映射为：

$$f(\mathbf{v}) = \sum_{i=1}^{k} x_i f(\mathbf{e}_i) \tag{1}$$

因此，f 可以表示为 V 中的一个行向量 $\mathbf{F}$，坐标为 $(f(\mathbf{e}_1), \cdots, f(\mathbf{e}_k))$. 对于 V 中任何向量 $\mathbf{v}$，都有 $f(\mathbf{v}) = \mathbf{F}\mathbf{c}$（按矩阵乘法）.

若取另一个基 $\{\mathbf{e}'_1, \cdots,$
$bvece'_k\}$，其中过渡矩阵为 $\mathbf{Q}$. 如果在新的基下 $\mathbf{v}$ 的坐标变为 $\mathbf{c}'$，那么 $\mathbf{Q}\mathbf{c}' = \mathbf{c}$[7].

设在新的基下，f 表示为行向量 $\mathbf{F}'$，那么应有 $f(\mathbf{v}) = \mathbf{F}\mathbf{c} = \mathbf{F}'\mathbf{c}'$. 考虑到 $\mathbf{Q}\mathbf{c}' = \mathbf{c}$，我们可知对于任何坐标 $\mathbf{c}, \mathbf{c}'$ 都有 $\mathbf{F}\mathbf{Q}\mathbf{c}' = \mathbf{F}'\mathbf{c}'$，因此

$$\mathbf{F}\mathbf{Q} = \mathbf{F}' \tag{2}$$

这就是一阶张量的坐标变换.

二阶张量的坐标变换

二阶张量涉及两个同构的线性空间 V，而且是在物理学中最为常见的张量形式，因此我们将详细讨论该情况. 在这里，我们把二阶张量 f 理解为从 V_1 和 V_2 到标量域 $\mathbb{K}$ 的一个线性映射，其中 V_1 和 V_2 同构，$\dim V_1 = \dim V_2 = k$.

给 V_1 指定一组基 $\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\}$，给 V_2 指定一组基 $\{\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_n\}$，在这两组基下，张量 f 被表示为一个矩阵 $\mathbf{F}$，而向量 $\mathbf{v}_1 \in V_1$ 和 $\mathbf{v}_2 \in V_2$ 在这两组基下的坐标列矩阵分别为 $\mathbf{c}(\mathbf{v}_1)$ 和 $\mathbf{c}(\mathbf{v}_2)$. 此时，$f(\mathbf{v_1}, \mathbf{v_2}) = \mathbf{c}(\mathbf{v_2})^{\mathrm{T}}\mathbf{F}\mathbf{c}(\mathbf{v_1})$.

如果给 V_1 和 V_2 进行基变换，过渡矩阵分别为 $\mathbf{P}$ 和 $\mathbf{Q}$，则在新的基下，两向量的坐标分别为 $\mathbf{c}'(\mathbf{v}_1) = \mathbf{P}^{-1}\mathbf{c}(\mathbf{v}_1)$ 和 $\mathbf{c}'(\mathbf{v}_2) = \mathbf{Q}^{-1}\mathbf{c}(\mathbf{v}_2)$，那么 $f(\mathbf{v}_1, \mathbf{v}_2) =$

[7] 见过渡矩阵[414].

$\mathbf{F}\mathbf{v}_1\mathbf{v}_2$. 此时有 $f(\mathbf{v_1},\mathbf{v_2}) = \mathbf{c}(\mathbf{v_2})^{\mathrm{T}}\mathbf{F}\mathbf{c}(\mathbf{v_1}) = \mathbf{c}'(\mathbf{v_2})^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{F}\mathbf{P}\mathbf{c}'(\mathbf{v_1})$. 因此在新基下，$f$ 的矩阵为 $\mathbf{Q}^{\mathrm{T}}\mathbf{F}\mathbf{P}$.

爱因斯坦求和约定表示的一般张量的坐标变换

高阶张量的坐标变换就无法简单地用矩阵的乘法来表示，因此我们采用爱因斯坦求和约定.

一阶张量的坐标变换

给定 k 维线性空间 V，及其上一个张量 $f : V \to \mathbb{R}$. 如果 V 的基是 $\{\mathbf{e}_1, \cdots \mathbf{e}_k\}$，在这组基下，张量 f 的坐标为 T_i，那么我们应有 $T_i = f(\mathbf{e}_i)$.

对 V 进行基的变换，如果过渡矩阵是 a_{ij}，即新的基中基向量 $\mathbf{e}'_i = a_{ji}\mathbf{e}_j$，那么我们易得：

$$T'_i = f(\mathbf{e}'_i) = f(a_{ji}\mathbf{e}_j) = a_{ji}f(\mathbf{e}_j) = a_{ji}T_j \tag{3}$$

注意指标的位置，以及哪些指标是用于求和的赝指标.

二阶张量的坐标变换

给定线性空间 V_1 和 V_2，其中 $\dim V_1 = \dim V_2 = k$. 令 $f : V_1 \times V_2 \to \mathbb{F}$ 是一个二阶张量.

给 V_1 指定一组基 $\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\}$，给 V_2 指定一组基 $\{\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_n\}$，在这两组基下，张量 f 的坐标是 $T_{ij} = f(\mathbf{a}_i, \mathbf{b}_j)$.

如果给 V_1 和 V_2 进行基变换，过渡矩阵分别为 m^{ij} 和 n^{ij}，则在新的基下，$\mathbf{a}'_i = m^{ki}\mathbf{a}_k$，$\mathbf{b}'_j = n^{lj}\mathbf{b}_l$，因此张量 f 在新基下的坐标为：

$$T'_{ij} = f(\mathbf{a}'_i, \mathbf{b}'_j) = f(m^{ki}\mathbf{a}_k, n^{lj}\mathbf{b}_l) = m^{ki}n^{lj}f(\mathbf{a}_k, \mathbf{b}_l) = m^{ki}n^{lj}T_{kl} \tag{4}$$

同样，注意指标的位置，以及哪些指标是用于求和的赝指标.

一般张量的坐标变换

相信从式 3 和式 4 中你已经比较得出了一般张量的坐标变换规律，因此我们不再详细展示计算过程，而是直接将坐标变换的方法表述如下：

定理 1　张量的坐标变换

给定域 $\mathbb{F}$ 上的一组有限个线性空间 $\{V_\alpha\}_{\alpha=1}^n$，设 $f: V_1 \times \cdots \times V_n \to \mathbb{F}$. 给各 V_α 指定一组基，使得 f 在这些基下的坐标为 $T_{i_1 i_2 \cdots i_n}$. 如果对每个 V_α 进行过渡矩阵为 m_{ij}^α 的基变换，那么在变换后的基下，f 的坐标为

$$T'_{i_1 i_2 \cdots i_n} = m^{k_1 i_1} m^{k_2 i_2} \cdots m^{k_n i_n} T_{k_1 k_2 \cdots k_n} \tag{5}$$

注意该式对各 $k_1, k_2, \cdots, k_n$ 求和.

以上变换仅仅是对于张量的坐标而言，并不改变张量本身. 今天的物理学家所说的张量变换，通常是指不同坐标下**同一个**张量的坐标表示之间的转换. 因此有的物理学家会将张量定义为“像张量一样变换的量（a tensor is something that transfers like a tensor）”，比如朗道. 这句定义听上去有些循环论证，但它的意思其实是“像 $T'_{i_1 i_2 \cdots i_n} = m^{k_1 i_1} m^{k_2 i_2} \cdots m^{k_n i_n} T_{k_1 k_2 \cdots k_n}$ 一样变换”. 这个定义使得许多理论物理学家认为**仿射联络**[708] 并不是张量. 事实上，他们所说的这个张量是指流形上的张量场，而仿射联络确实不是一个张量场；仿射联络本身是一个三阶张量，只不过它所涉及的向量并不是流形上每个点的切向量（因此不是张量场），而是向量场本身. 也就是说仿射联络是把一个向量场本身看成一个向量后所进行的线性映射.

协变和逆变

预备知识 过渡矩阵[414]，爱因斯坦求和约定[498]

协变和逆变的概念在物理学中极为普遍，它描述的是物理量随着参考系变化等变换而变换的特点. 在现代物理学语言中，常常使用各种各样的线性空间来描述物理系统，其中一个向量表示一种状态，一组基底表示一种看待此系统的视角，比如参考系、表象等等，而向量的坐标则意味着在给定视角（基底）下的物理量. 因此，协变和逆变实际上描述的是各种向量的坐标随着基底变换的变换特征.

简单来说，协变就是指坐标的变换矩阵和基底变换的过渡矩阵相同，而逆变就是指坐标的变换矩阵和过渡矩阵互逆.

对偶基下的张量坐标变换

在张量[488]词条中我们知道了，定义张量的时候只需要一系列和某一线性空间 V 同构的线性空间就可以，并没有对这些线性空间的基有特殊要求．但是张量的具体表示依赖于各基的选取，所以我们现在制定一下基的选取规则．

首先，我们**不使用任意的** V，那样范围过于广．从现在开始，定义张量时所涉及的一系列 V，不管是作为定义域的一部分还是值域的一部分，都**只能从** V **和** V^* **中选取**，其中 V^* 是 V 的对偶空间．所有的 V 用同一组基，所有的 V^* 则用对应的对偶基．对偶基的定义请参考对偶空间[460]词条．

这样一来，我们只需要确定一组 V 的基，就相当于给出了所有 V 和 V^* 的基，从而确定了任何涉及 V 和 V^* 的张量的坐标．

协变和逆变的概念来自张量坐标变换的方式．我们先观察一个最为简单的例子，来直观感受一下什么是变换的方式．

例 1

考虑一维实线性空间（也就是实数域本身）$\mathbb{R}$，其对偶空间 $\mathbb{R}^*$ 就是全体实系数一次函数的集合 $\{f:\mathbb{R}\to\mathbb{R}|f(x)=ax\}$．为了方便表示，我们把函数 ax 记为 $f_a(x)$．

现在取定 $\mathbb{R}$ 上的一组基．由于 $\mathbb{R}$ 是一维的，基只需要一个基向量，通常取实数 1，但也可以取任何别的实数．如果 $\mathbb{R}$ 的基是 $\{a\}$，那么 $\mathbb{R}^*$ 的基就是 $f_{\frac{1}{a}}$，这样才满足对偶基的要求：$f_{\frac{1}{a}}(a)=\frac{1}{a}a=1$．

现在给出任意一个函数 $f_b\in\mathbb{R}^*$，由于 $f_b(x)=bx=ab(\frac{1}{a}x)=abf_{\frac{1}{a}}(x)$，故在如上给定的基和对偶基下，$f_b$ 的坐标为 (ab)．如果给出任意一个实数 $y\in\mathbb{R}$，那么也可以计算出 y 的坐标是 (y/a)．到这一步，你可能已经大致看出来协变和逆变的关系了．

如果我们换一组 $\mathbb{R}$ 的基，比如说换成 a'，那么过渡矩阵就是 (a'/a)．在新基下，f_b 的坐标为 $a'b$，y 的坐标为 y/a'，也就是说他们的坐标分别是原来的坐标乘以矩阵 (a'/a) 和 (a/a')．这两个变换矩阵，第一个是过渡矩阵本身，第二个是过渡矩阵的逆；所以我们说，f_b 的坐标关于 $\mathbb{R}$ 的坐标变换是一致的，或者说是共变（**covariant**）的，也叫协变的；反过来，y 的坐标变换就是反变（**contravariant**）的，也叫逆变的．

作为简单总结，我们可以说，如果 $\mathbb{R}$ 的基向量变成原来的 2 倍，那么 f_b

的坐标变为原来的 2 倍，y 的坐标变为原来的 $1/2$ 倍.

当基变化时，向量的坐标变换都是过渡矩阵的逆. 对偶基的定义，使得它的过渡矩阵总是原空间基的过渡矩阵的逆. 这就是为什么 f_b 反而协变的.

因此，对于一般的情况，我们把协变和逆变的概念定义如下：

定义 1　协变和逆变向量

给定线性空间 V，则 $\mathbf{v}\in V$ 称为其逆变向量，$\mathbf{w}\in V^*$ 称为其协变向量.

一定要分清楚对偶向量和向量，虽然都叫向量，但它们并不在一个空间里.

上下指标

预备知识　张量的坐标变换[501]

向量和对偶向量都可以看成一阶张量，前者把一个对偶向量映射为一个标量，后者把一个向量映射为一个标量. 我们已经知道，如果原空间 V 的过渡矩阵写为 a^i_j，其中第 i 行 j 列的元素是 a^i_j，那么原空间的向量坐标变换矩阵就是其逆矩阵 $(a^i_j)^{-1}$，而对偶空间的向量坐标变换矩阵是 a^i_j 本身.

爱因斯坦求和约定里规定了上下标之间的关系，但是那只是一种形式，并未说明上下标到底指代什么对象. 我们现在就来定义上下标的含义.

定义 2

给定线性空间 V，将其向量的坐标表示为带上标的数，对偶向量的坐标表示为带下标的数. 这样，对于向量或者说一阶张量，上下标是在告诉读者，这个向量是 V 中的还是 V^* 中的.

在线性代数的范围里，给定线性空间 V 的基后，我们通常把 V 的向量坐标表示为列矩阵，而把 V^* 的向量坐标表示为行矩阵. 因此上标 i 表示“第 i 个列矩阵的元素”，也就是“行数”，类似地下标就表示“列数”. 按照矩阵的乘法规则，即“行乘以列再求和”，我们就得到了爱因斯坦求和约定的上下标规则：求和的两个指标必须一上一下.

爱因斯坦求和约定的好处是，可以把“行数”、“列数”的含义推广到高维矩阵上，也就是推广到高阶张量上，相当于把高维矩阵的各个指标分为两类，一类是“行”，一类是“列”.

升降指标

如果给定了一个向量 $\mathbf{v}$，其在某基下的坐标是 v^a，那么这个向量的对偶向量 $\mathbf{v}^*$ 在对偶基下的坐标就表示为 v_a．这样，一个向量变成其对偶向量的过程中坐标的变化[8]，各元素也发生变化，就叫做**指标的下降**．反过来，把一个对偶向量变成原向量，就叫做**指标的提升**．

由于对偶空间之间是平等的，指标的上下、升降就决定于“选择哪个作为原空间”，也就是说是相对的概念．

向量指标的升降也可以推广到一般的张量指标的升降，规则和对偶向量是一样的：升降后的张量和原张量相乘，结果是 1．

需要补充的是，以上讨论实际上是默认“度量张量为 δ^i_j[9]的情况．如果用度量张量来定义内积，或者用伪度量张量来定义伪内积[10]，那么升降指标的规则应该推广为：升降后的张量和原张量再乘以度量张量后，结果是 1．这就是我们在进行指标升降的时候会乘以一个度量张量的原因．

张量的协变和逆变

给定线性空间 V，则每个逆变向量都可以看成 $V^* \to \mathbb{F}$[11]的一个一阶张量；反过来，每个协变向量也都是 $V \to \mathbb{F}$ 的一阶张量．

一般地，如果一个张量 T 涉及 n 个 V 和 m 个 V^*[12]，那么我们可以把它记为 $T^{b_1b_2\cdots b_m}_{a_1a_2\cdots a_n}$，这样就可以用爱因斯坦求和约定来计算它对各向量的作用．我们用二阶张量来举一个例子，请对照张量[488] 词条教的理解方式和技巧来看．

例 2

给定线性空间 V，设 T^b_a 是涉及一个 V 和一个 V^* 的张量．

对于逆变向量 $v^a \in V$，这个张量把它映射为一个协变向量（也即 1-线性函数）$w_b = v^aT^b_a$．

对于逆变向量 $v^a \in V$ 和协变向量 w_b，这个张量把它们俩映射为一个数字 $v^aw_bT^a_b$．

---

[8] 就是指坐标从列矩阵变成行矩阵

[9] 当且仅当 $i=j$ 时，$\delta^i_j = \delta^{ij} = \delta_{ij} = 1$．

[10] 最常见的就是闵可夫斯基度规，在相对论中随处可见．

[11] $\mathbb{F}$ 是任意的域，一般是实数域 $\mathbb{R}$ 或复数域 $\mathbb{C}$．

[12] 参见张量[488] 的小节“张量的阶数”．

根据张量的坐标变换[501]给的变换规则，对于 V 上的张量 $T^{b_1b_2\cdots b_m}_{a_1a_2\cdots a_n}$，当我们给定 V 的过渡矩阵 a^{ij} 的时候，每个 V^* 的过渡矩阵都是 a_{ij}，其中 a^{ij} 和 a_{ij} 互为逆矩阵，即 $a^{ik}a_{kj}=\delta^i_j$[13]．这样一来，整个张量的坐标变换就成了 m 个 $a_{i_\alpha k\alpha}$ 和 n 个 $a^{k_\beta j_\beta}$ 的乘积，虽然它们并不会互相抵消，但是它们的行列式可以互相抵消，从而使得张量坐标在变换前后行列式不变．

需要补充一个二阶张量的变换例子．

张量积

预备知识 张量[488]，爱因斯坦求和约定[498]

概念

在张量[488]词条中我们知道，一个张量可以看成是将若干个线性空间 V 映射到域 $\mathbb{F}$ 上的映射．

域是可以进行加减乘除的集合，因此我们可以在域上定义映射的乘积．比如如果有实数域上的实函数 $f,g:\mathbb{R}\to\mathbb{R}$，那么我们可以定义两个函数的乘积 $f\cdot g$ 如下：$\forall x\in\mathbb{R}, f\cdot g(x)=f(x)\cdot g(x)$．就是说，把各点的函数值相乘，得到的结果还是一个函数，它就是两个函数的乘积．

同样的方法可以应用到张量上，这样所得到映射乘积，就是所谓的张量积．

定义 1　张量积

给定线性空间 V，在线性空间 V 上有 $n-$ 线性函数 f 和 $m-$ 线性函数 g，则可以定义 $(n+m)-$ 线性函数 $f\cdot g$，方式为：对于任意 $x_i,y_j\in V$，$f\cdot g(x_1,\cdots,x_n,y_1,\cdots,y_m)=f(x_1,\cdots,x_n)\cdot g(y_1,\cdots,y_m)$．

用矩阵可以直观地看出张量积的“升维”性质，但是这个方法很难推广到任意阶的张量．我们举一个一阶张量相乘得到二阶张量的例子：

[13] $\delta^i_j=1$ 当且仅当 $i=j$；否则，$\delta^i_j=0$．按照我们规定的，上标表示行数，下标表示列数，可知它表示单位矩阵．

例 1　一阶张量的张量积

在一个 3 维线性空间 V 里默认取某个基时，有向量 $x^i \in V$ 和一个对偶向量（不一定是 x^i 的对偶）$y_j \in V^*$，那么 x^i 可以看成是把 V^* 里的向量映射为标量的张量，y_j 可以看成是把 V 里的向量映射为标量的张量．如果在给定基下，把 V 中元素的坐标都写为列矩阵，V^* 中的坐标写为行矩阵，比如 $x^i = \left(x^1, x^2, x^3\right)^{\mathrm{T}}$，$y_j = \left(y_1, y_2, y_3\right)$，那么可以通过乘积得到矩阵 $\mathbf{M} = \left(x^1, x^2, x^3\right)^{\mathrm{T}} y_j = \left(y_1, y_2, y_3\right)$，这个矩阵就表示把一个向量和一个对偶向量映射到一个标量上的二阶张量．

为便于理解，我显式地写一遍这个矩阵乘法过程：

$$\mathbf{M} = \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} \times \left(y_1, y_2, y_3\right) = \begin{pmatrix} x^1 y_1, x^1 y_2, x^1 y_3 \\ x^2 y_1, x^2 y_2, x^2 y_3 \\ x^3 y_1, x^3 y_2, x^3 y_3 \end{pmatrix} \tag{1}$$

如果参与进行张量积的张量阶数超过 1，那么就不再能像上边这样简单地用矩阵表达了，因此我们必须使用爱因斯坦求和约定来简化表达．我把表达的规则写成以下定理：

定理 1

如果在给定基下，两个张量分别表示为 $a^{i_1 i_2 \cdots i_m}_{j_1 j_2 \cdots j_n}$ 和 $b^{k_1 \cdots k_s}_{l_1 \cdots l_t}$，那么它们的张量积就是 $a^{i_1 i_2 \cdots i_m}_{j_1 j_2 \cdots j_n} b^{k_1 \cdots k_s}_{l_1 \cdots l_t}$．张量积的阶数为各张量阶数之和．

比如说，张量 a^{ij}_k 和 b^r_s 的张量积是 $a^{ij}_k b^r_s$，其阶数为 $3 + 2 = 5$．

以上张量的表示中要么写出了线性映射的自变量，要么写出了抽象指标．但有的时候为了简便，我们也可能不显式写出张量的自变量等细节，这个时候表示张量的方式就是使用符号 $\otimes$．比如说，考虑到一个向量 $\mathbf{v}$ 本身是一个“将一个对偶向量映射为一个数”的一阶张量，一个对偶向量 $\mathbf{u}^*$ 本身是一个“将一个向量映射为一个数”的一阶张量，那么它们的张量积就表示为 $\mathbf{v} \otimes \mathbf{u}^*$，这样就和它们直接的乘积 $\mathbf{v}\mathbf{u}^*$ 区分开了——$\mathbf{v} \otimes \mathbf{u}^*$ 是一个二阶张量，而 $\mathbf{v}\mathbf{u}^*$ 是一个数．

实例

例 2　投影算子

给定线性空间 V 里的一根坐标轴（比如考虑二维实空间里的 x 轴），沿着这根坐标轴有一个单位向量 $\hat{\mathbf{n}}$. 如果映射 f 把 V 中的任意向量映射为其在这根坐标轴上的投影向量，那么 f 是一个二阶张量. 取 $\hat{\mathbf{n}}$ 的对偶向量 $\hat{\mathbf{n}}^*$，那么 f 可以表示为张量积 $\hat{\mathbf{n}} \otimes \hat{\mathbf{n}}^*$.

具体来说，如果考虑二维实线性空间中的 $\hat{\mathbf{n}} = (\frac{\sqrt{3}}{2}, \frac{1}{2})^{\mathrm{T}}$，其对偶向量为 $\hat{\mathbf{n}}^* = (\frac{\sqrt{3}}{2}, \frac{1}{2})$，那么“投影到 $\hat{\mathbf{n}}$ 所在轴”的投影张量就可以表示为：

$$\hat{\mathbf{n}} \otimes \hat{\mathbf{n}}^* = \begin{pmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{pmatrix} \times \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{3}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{1}{4} \end{pmatrix} \tag{2}$$

容易发现，这个矩阵的行列式是 0，这是符合预期的，因为投影变换是不可逆的.

在狄拉克的表示中，向量表示为右矢 $|\hat{\mathbf{n}}\rangle$，而其对偶向量表示为 $\langle\hat{\mathbf{n}}|$，因此该投影算子表示为 $|\hat{\mathbf{n}}\rangle\langle\hat{\mathbf{n}}|$. 这样一来，把 $|\mathbf{x}\rangle$ 投影到 $|\hat{\mathbf{n}}\rangle$ 的方向上，结果就是 $|\hat{\mathbf{n}}\rangle\langle\hat{\mathbf{n}}|\mathbf{x}\rangle$；把 $\langle\mathbf{x}|$ 投影到 $\langle\hat{\mathbf{n}}|$ 的方向上，结果就是 $\langle\mathbf{x}|\hat{\mathbf{n}}\rangle\langle\hat{\mathbf{n}}|$.

张量的分类

- 本词条处于草稿阶段.

预备知识　协变和逆变[504]

协变和逆变[504] 词条中提到，尽管张量是多个任意的线性空间 V 之间的多重线性映射，我们通常只考虑用 V 和其对偶空间 V^* 定义的张量，这样只需要定义一个空间的基就可以得到所有空间的基了.

回忆张量[488] 词条中所教的判断张量阶数的方法：看一共有多少线性空间参与映射，不过是作为自变量的一部分还是像的一部分. 在协变和逆变[504] 词条中我们介绍了定义一个张量时给线性空间分类的方法，即分为 V 和 V^*，这样就可以进一步把张量分出类型来，比阶数更细致一些.

定义 1

一个 (m,n) 型张量是涉及了 n 个逆变向量和 m 个协变向量的 $(m+n)$ 阶张量.

一个 (m,n) 型张量可以把 n 个逆变向量和 m 个协变向量映射为一个数字，也可以是映射成低阶张量. 比如说，如果 h_c^{ab} 是一个 $(2,1)$ 型张量，那么它乘以一个逆变向量 v^c，就得到一个 $(2,0)$ 型张量 $g^{ab}=v^c h_c^{ab}$；同样，它乘以两个协变向量后就得到一个 $(0,1)$ 型张量，也就是一个协变向量.

第六章

未归类

张量积空间

预备知识 多元函数的傅里叶级数[291]，子空间[464]

定义

我们以二维傅里叶级数来引入张量积的概念. 我们已知一维函数的傅里叶级数展开可以看做是一个矢量在一组正交归一基底上的展开，自然地，我们也希望能从矢量的空间的角度理解二维傅里叶级数展开. 为了讨论方便，我们只讨论可以用有限项傅里叶级数展开的函数.

如果 x 和 y 方向分别取 N_x 和 N_y 个函数基底，那么二维傅里叶级数中就有 N_xN_y 个基底，即被展开的函数是一个 N_xN_y 维空间中的矢量.

观察二维傅里叶级数中的函数基底可以发现，每一个基底都是一个 N_x 维空间中的 x 基底和一个 N_y 维空间中的 y 基底的乘积（两个一元函数相乘变为二元函数）. 从矢量的角度来看，这是一种以前没有见过的矢量乘法，它既不是数乘（得到同维度的矢量）也不是内积（得到标量），而是将两个不同矢量空间中的矢量相乘得到一个更高维空间中的矢量，使新矢量的维度等于前两个矢量各自的维度相乘. 我们把这样的矢量乘法叫做**张量积**，记为

$$|v\rangle = |x\rangle \otimes |y\rangle \tag{1}$$

我们把张量积得到的矢量所在的空间叫做**张量积空间**. 为了书写方便我们时常省略张量积号记为 $|x\rangle\,|y\rangle$，或者将张量积的结果记为一个整体 $|v\rangle = |x, y\rangle$.

我们将两个一维傅里叶变换空间中的基底分别为 $\{|x_i\rangle\}$ 和 $\{|y_j\rangle\}$，那么我们定义张量积空间为基底 $\{|x_i\rangle\,|y_j\rangle\}$ 张成的空间，里面的任意矢量都可以用基底展开.

$$|v\rangle = \sum_{i,j} C_{ij}\,|x_i, y_j\rangle \tag{2}$$

定义张量积满足**结合律**，即两个低维空间中任意各选一个矢量 $|x\rangle$ 和 $|y\rangle$，它们都可以在各自空间的基底中展开，则它们的张量积为

$$|x\rangle\,|y\rangle = \sum_i a_i\,|x_i\rangle \sum_j b_j\,|y_j\rangle = \sum_{i,j} a_ib_j\,|x_i, y_j\rangle \tag{3}$$

注意张量积空间的的任意一个矢量不一定可以表示为一个张量积运算，例如 $|x_1\rangle|y_1\rangle+2|x_2\rangle|y_3\rangle$．也可类比函数的情况，$f_x(x)f_y(y)$ 可以记为 $f(x,y)$，但 $f(x,y)$ 不一定能分解为 $f_x(x)f_y(y)$．

任何矢量空间的基底都需要有固定的顺序，一旦确定在所有的计算中都不能改变．张量积基底通常按照以下两种方式排序[1]

$$\{|x_1,y_1\rangle,|x_1,y_2\rangle\dots|x_2,y_1\rangle,|x_2,y_2\rangle\dots\} \tag{4}$$

$$\{|x_1,y_1\rangle,|x_2,y_1\rangle\dots|x_1,y_2\rangle,|x_2,y_2\rangle\dots\} \tag{5}$$

排序以后，我们就可以用单下标来区分不同的基底，令

$$|x,y\rangle_\alpha=|x_i,y_j\rangle \tag{6}$$

对应两种排序，分别有

$$\alpha=N_y(i-1)+j\quad\text{或}\quad\alpha=i+N_x(j-1)\qquad(1\leqslant\alpha\leqslant N_xN_y) \tag{7}$$

矢量的内积

若我们需要在张量积空间中讨论模长和正交这样的概念就必须先定义**内积**：张量积空间中两矢量（先假设它们可以表示为单个张量积）的内积等于把每个低维空间中对应矢量分别做内积再相乘．即 $|c\rangle|d\rangle$ 和 $|a\rangle|b\rangle$ 的内积为

$$(\langle d|\langle c|)(|a\rangle|b\rangle)=\langle d|\langle c|a\rangle|b\rangle=\langle c|a\rangle\langle d|b\rangle \tag{8}$$

注意 $|c\rangle|d\rangle$ 的厄米共轭记为 $\langle d|\langle c|$ 而不是 $\langle c|\langle d|$．这样就可以很容易看出需要把 $\langle c|a\rangle$ 组合做内积而不是 $\langle d|a\rangle$．但如果把两矢量记为 $|a,b\rangle$ 和 $|c,d\rangle$，那么内积记为[2]

$$\langle c,d|a,b\rangle=\langle c|a\rangle\langle d|b\rangle \tag{9}$$

张量积空间中的内积是否满足交换律取决于两个低维空间中的内积是否满足交换律．按照内积的一般性质 $\langle u|v\rangle=\langle v|u\rangle^*$，所以张量积空间中的也有

$$\langle a,b|c,d\rangle=\langle a|c\rangle\langle b|d\rangle=\langle c|a\rangle^*\langle d|b\rangle^*=\langle c,d|a,b\rangle^* \tag{10}$$

[1] 由于这里每个基底都有两个角标，使初学者有把基底排成长方形的冲动，这是错误的．

[2] 书写习惯上，ket 和对应的 bra 中的记号应保持相同．所以 $|c,d\rangle$ 的厄米共轭记为 $\langle c,d|$ 而不是 $\langle d,c|$．

如果两个低维空间中的基底都是正交归一的，那么张量积空间中的基底也是正交归一的

$$\langle x_{i'}, y_{j'}|x_i, y_j\rangle = \langle x_{i'}|x_i\rangle \langle y_{j'}|y_j\rangle = \delta_{i,i'}\delta_{j,j'} \tag{11}$$

以后我们一般讨论正交归一基底.

知道了张量积空间基底之间的内积后，要计算张量积空间中任意两个矢量的内积，只需先将它们分解到基底上，再按照内积的分配律和正交归一化条件即可得到熟悉的内积公式

$$\begin{aligned}\langle v'|v\rangle &= \left(\sum_{i',j'} C'^{*}_{i',j'} \langle y_{j'}| \langle x_{i'}|\right)\left(\sum_{i,j} C_{i,j} |x_{i'}\rangle |y_{j'}\rangle\right) \\ &= \sum_{i',j'}\sum_{i,j} C'^{*}_{i',j'} C_{i,j}\delta_{i,i'}\delta_{j,j'} = \sum_{i,j} C'^{*}_{i,j} C_{i,j}\end{aligned} \tag{12}$$

正交归一基底的**完备性**仍然可以记为

$$\sum_{i,j} |x_i, y_j\rangle \langle x_i, y_j| = \mathbf{I} \tag{13}$$

其中 $\mathbf{I}$ 是张量积空间的单位算符.

我们现在从张量积空间的角度来看二维傅里叶级数，将被展开的函数记为矢量 $|f\rangle$，则

$$|f\rangle = \sum_{\alpha} C_\alpha |x, y\rangle_\alpha = \sum_{i,j} C_{i,j} |x_i, y_j\rangle \tag{14}$$

其中

$$C_{i,j} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{x,i}(x) f_{y,j}(y) f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \langle x_i, y_j|f\rangle \tag{15}$$

我们可以验证基底的完备性（式 13），将式 15 代入式 14 得

$$|f\rangle = \sum_{i,j} |x_i, y_j\rangle \langle x_i, y_i|f\rangle = \left(\sum_{i,j} |x_i, y_j\rangle \langle x_i, y_i|\right) |f\rangle \tag{16}$$

对所有 $|f\rangle$ 都成立，这就间接证明了

$$\sum_{\alpha} |x_i, y_j\rangle \langle x_i, y_i| = \mathbf{I} \tag{17}$$

若张量积空间中一个矢量可以表示为单个张量积 $|u\rangle |v\rangle$，则其基底展开为

$$|u\rangle |v\rangle = \left(\sum_i a_i |u_i\rangle\right) \otimes \left(\sum_j b_j |u_j\rangle\right) = \sum_{i,j} a_i b_j |u_i, u_j\rangle \tag{18}$$

张量积空间的子空间

令 $|u_i\rangle$ 和 $|v_j\rangle$ 分别为两个低维空间的基底，则张量积空间中存在两种“天然”的子空间．一种将张量积基底根据 i 的值分成许多组，每组张成一个子空间．同理，也可以根据 j 来划分子空间．我们姑且分别称它们为 u 子空间和 v 子空间．

任意一个张量空间中的矢量可以看做是每个子空间中的一个矢量的线性组合

$$\begin{aligned}\sum_{i,j} C_{ij}\,|u_i\rangle\,|v_j\rangle &= \sum_j \left(\sum_i C_{ij}\,|u_i\rangle\right)\otimes|v_j\rangle = \sum_j |a_j\rangle\,|v_j\rangle \\ &= \sum_i |u_i\rangle\otimes\left(\sum_j C_{ij}\,|v_j\rangle\right) = \sum_i |u_i\rangle\,|b_i\rangle\end{aligned} \tag{19}$$

上式中的 $|a_j\rangle$ 可以理解为矢量在每个 $|v_j\rangle$ 子空间的**分量**（u 空间的矢量），$|b_i\rangle$ 理解为矢量在每个 $|u_i\rangle$ 子空间中的**分量**（v 空间的矢量）．

算符

两个低维空间的线性算符 $\hat{A}$ 和 $\hat{B}$ 做张量积可以得到张量积空间中的算符 $\hat{A}\otimes\hat{B}$．我们定义 $\hat{A}\otimes\hat{B}$ 为线性算符，该算符对任意基底的作用结果为

$$\hat{A}\otimes\hat{B}\,|u_i,v_j\rangle = (\hat{A}\,|u_i\rangle)\otimes(\hat{B}\,|v_j\rangle) \tag{20}$$

要对任意矢量作用，只需将矢量记为张量积空间基底的线性组合，然后再分别对张量积基底作用即可．

特殊地，可以用 $\hat{A}\otimes\hat{I}$ 运算将 $\{|u_i\rangle\}$ 空间中的 $\hat{A}$ 拓展到张量积空间中来

$$\hat{A}\otimes\hat{I}\,|u_i,v_j\rangle = (\hat{A}\,|u_i\rangle)\otimes|v_j\rangle \tag{21}$$

这说明算符 $\hat{A}\otimes\hat{I}$ 在每个 v 子空间中都是闭合的．所以，算符 $\hat{A}\otimes\hat{I}$ 作用在张量积空间的任意矢量上，相当于 $\hat{A}\otimes\hat{I}$ 对该矢量在各个子空间中的分量分别作用．$\hat{B}$ 也同理，不再赘述．根据定义，不难证明

$$(\hat{A}\otimes\hat{I})(\hat{I}\otimes\hat{B}) = (\hat{I}\otimes\hat{B})(\hat{A}\otimes\hat{I}) = \hat{A}\otimes\hat{B} \tag{22}$$

和矢量空间一样，线性算符空间中也不是任何算符都能表示为两个算符的张量积．

二维矩阵

若张量积空间中的基底正交归一，则张量积算符矩阵元为

$$\begin{aligned}(\hat{A}\otimes\hat{B})_{\alpha,\alpha'} &= (\hat{A}\otimes\hat{B})_{i,j,i',j'} = \langle u_i v_j|\,\hat{A}\otimes\hat{B}\,|u_{i'}v_{j'}\rangle \\ &= \langle u_i v_j|\,(\hat{A}\,|u_{i'}\rangle\otimes\hat{B}\,|v_{j'}\rangle) = \langle u_i|\,\hat{A}\,|u_{i'}\rangle\,\langle v_j|\,\hat{B}\,|v_{j'}\rangle\end{aligned} \tag{23}$$

注意矩阵元的位置与基底的排序有关，如果按照式 4 排序，则每个 u 子空间的基底都在一起．现在用分块矩阵的概念，若把矢量根据 u 子空间分成一段段，每一段是 u 子空间中分量的系数，矩阵也会分成一些正方形的小块，第 (m,n) 小块中为

$$A_{mn}\mathbf{B} \tag{24}$$

所以，$\hat{A}\otimes\hat{B}$ 的矩阵是把 $\mathbf{A}$ 的每个矩阵元 A_{mn} 拓展成矩阵分块 $A_{mn}\mathbf{B}$.

同理，若基底以式 5 排序，即每个 v 子空间的基底都在一起，则 $\hat{A}\otimes\hat{B}$ 的矩阵就是把 $\mathbf{B}$ 的每个矩阵元 B_{mn} 拓展成矩阵分块 $B_{mn}\mathbf{A}$.

任何线性算符都可以表示为算符张量的线性组合，所以一般的算符仍然可以表示为分块矩阵，但每个分块未必成比例．

四维张量

将直积空间基底排序的做法丧失了两个空间基底的对称性．我们可以将张量积空间中矢量 $|c\rangle$ 的坐标 $C_{i,j}$ 写成矩阵的形式（注意概念上这仍然是一个矢量而不表示线性映射）．矩阵的第 i 行就是 $|u_i\rangle$ 子空间分量的坐标，第 j 列就是 $|v_j\rangle$ 子空间分量的坐标．

$\hat{A}\otimes\hat{B}$ 作用在该矩阵上，就相当于用矩阵 $\mathbf{A}$ 处理每一列（矩阵 $\mathbf{C}$ 的行变换），再用矩阵 $\mathbf{B}$ 处理每一行（矩阵 $\mathbf{C}$ 的列变换），也可以是先列变换再行变换．

$$\sum_{j'} B_{jj'}\left(\sum_{i'} A_{ii'}C_{i'j'}\right) = \sum_{i'} A_{ii'}\left(\sum_{j'} B_{jj'}C_{i'j'}\right) \tag{25}$$

如果 $\mathbf{A}$ 或 $\mathbf{B}$ 是对角的，那么分别只需要把矩阵的每一行或每一列的系数乘以对角元，再分别使用矩阵 $\mathbf{B}$ 或 $\mathbf{A}$ 作用即可．这就相当于对每个子空间中的分量独立作用．

我们可以用四维张量来表示这个算符，令算符 $T = A \otimes B$，则 $T_{i,j,i'j'} = A_{i,i'}B_{j,j'}$，那么算符与矢量的乘法为

$$(T\,|c\rangle)_{i,j} = \sum_{i',j'} T_{i,j,i'j'}C_{i',j'} \tag{26}$$

一般的线性算符也都可以用四维张量表示.

多重张量积

例如三个空间中的矢量 $|a\rangle$，$|b\rangle$，$|c\rangle$ 的张量积为 $|a\rangle\,|b\rangle\,|c\rangle$，可以表示为三维矩阵．算符可以表示为 $2\times 3 = 6$ 维张量．算符作用在矢量就是

$$y_{i,j,k} = \sum_{i',j',k'} Q_{i,j,k,i',j',k'}x_{i',j',k'} \tag{27}$$

其他运算也都可以类比.

部分内积

[3]如果我们定义两矢量的张量积与 u 空间中的矢量的**部分内积**运算为

$$\langle u_1|\,(|u\rangle\,|v\rangle) = \langle u_1|u\rangle\,|v\rangle \tag{28}$$

且该运算是线性的，那么 u 空间中的任意矢量乘以张量积空间中任意矢量的运算为

$$\begin{aligned}\left(\sum_k x_k\,\langle u_k|\right)\left(\sum_{i,j} C_{ij}\,|u_i\rangle\,|v_j\rangle\right) &= \sum_{i,j} C_{ij}\left(\sum_k x_k\,\langle u_k|u_i\rangle\right)|v_j\rangle \\ &= \sum_j \left(\sum_i x_i C_{ij}\right)|v_j\rangle\end{aligned} \tag{29}$$

即对每个 $|v_j\rangle$ 子空间中的分量分别做内积，得到 v 空间的矢量.

同理，也可以定义 v 空间的部分内积

$$\langle v_1|\,(|u\rangle\,|v\rangle) = (\langle v_1|v\rangle)\,|u\rangle \tag{30}$$

那么 v 空间中的任意左矢 $\langle v_1|$ 乘以任意张量积空间的矢量等于将每个 $|u_j\rangle$ 子空间中的分量分别与 $\langle v_1|$ 内积，得到 u 空间的矢量.

[3]这个运算并不是标准运算而是本书的一个定义

部分矩阵元

先定义算符的**部分运算为** $(\hat{A}\otimes\hat{B})\,|b_j\rangle=(\hat{A}\,|\cdot\rangle)\otimes(\hat{B}\,|b_j\rangle)$，这是一个 a 空间的算符作用在某个 a 空间矢量后再与 $|b\rangle$ 做张量积的运算.

定义**部分矩阵元**如 $\langle b_i|\hat{A}\otimes\hat{B}|b_j\rangle=\langle b_i|\hat{B}|b_j\rangle\,\hat{A}$ 是 a 空间的算符. 所以 $\hat{A}\otimes\hat{B}$ 作用在任意矢量上，就相当于先把 $\hat{A}$ 作用在每个 $|b_j\rangle$ 子空间的分量上，再将这些子空间根据矩阵 $\langle b_i|\hat{B}|b_j\rangle$ 进行线性变换. 特殊地，如果 $\langle b_i|\hat{B}|b_j\rangle$ 是对角矩阵，那么只需要把 $\hat{A}$ 作用在每个 $|b_j\rangle$ 子空间的分量上再乘以对角矩阵元 $\langle b_j|\hat{B}|b_j\rangle$ 即可.

$\langle a_i|\hat{A}\otimes\hat{B}|a_j\rangle=\langle a_i|\hat{A}|a_j\rangle\,\hat{B}$ 的讨论同理.

其他笔记

关于本征问题的定理

- 如果考虑张量积空间中的本征问题，$\hat{A}$ 的本征矢 $|eig_i\rangle$ 具有 n 重简并，简并空间的基底分别为 $|eig_i\rangle\,|v_1\rangle\,,|eig_i\rangle\,|v_2\rangle\ldots.$
- $\hat{A}\otimes\hat{B}$ 的本征值共有 $m\times n$ 个，m 和 n 分别是 A 和 B 的维度，若 $a_1,a_2,\ldots,a_m$ 和 $b_1,b_2,\ldots,b_n$ 分别是 A 和 B 的本征值，那么 $A\otimes B$ 的本征值和本征矢分别为 a_ib_j 和 $|u_i,v_j\rangle$ （$0\leqslant i\leqslant m,0\leqslant j\leqslant n$）.
- $A\otimes I+I\otimes B$ 的本征值和本征矢分别为 a_i+b_j 和 $|u_i,v_j\rangle$.
- 两个厄米矩阵张量积仍然是厄米矩阵.
- 两个厄米矩阵相加仍然是厄米矩阵.

算符的展开

另外，量子力学中的算符（我见过的）都可以表示为两个小空间中的算符的张量积的线性组合

$$\Omega=\sum_k A_k\otimes B_k \tag{31}$$

例如中心力场中的单电子哈密顿算符

$$H=K_r+\frac{L^2}{2mr^2}+V(r)-q\boldsymbol{\mathcal{E}}\cdot\mathbf{r} \tag{32}$$

其中 $K_r, V(r)$ 就是 R 空间的算符与 Y 空间的单位算符的张量积，L^2 算符（式 4 [1409]）是 R 空间中的单位算符与 Y 空间中算符的张量积．而最后一项没有微分算符，只是一个函数，这个行数同样也可以展开为张量积空间中的基底的线性组合，即两空间中的算符的张量积的线性组合．

一种简单的证明方法是，例如 2 乘 2 的矩阵 $\mathbf{A}$ 与 2 乘 2 的矩阵 $\mathbf{B}$ 直积，就是把 $\mathbf{A}$ 的每个矩阵元乘以 $\mathbf{B}$ 变为一个 4 乘 4 的矩阵．要得分解任意 4 乘 4 的矩阵，我们只需要取 $\mathbf{A}_i$ 分别为一组矩阵基底 $[1,0;0,0]$，$[0,1;0,0]$，$[0,0;1,0]$，$[0,0;0,1]$，然后再令 $\mathbf{B}_i$ 为每个子矩阵即可．

单空间的基底展开

张量积空间中的任意矢量可以只在一个小空间的基底上展开，例如波函数可以看做径向波函数空间和角向波函数空间的张量积．波函数可以只在角向波函数空间中的基底——球谐函数上展开

$$|\Psi\rangle = \sum_j |R_j\rangle\,|Y_j\rangle \tag{33}$$

其中 j 代表 l, m 组合的某种排序．

如果我们在一个小空间中使用基底，而另一个空间中不用．例如 Y 空间中我们用球谐函数作为基底，可以得到一个“算符矩阵”，即每个矩阵元为 $\mathbf{R}$ 空间的算符．例如薛定谔方程

$$H\,|\Psi\rangle = \mathrm{i}\frac{\partial}{\partial t}\,|\Psi\rangle \tag{34}$$

使用了 Y 空间的基底后可以表示为

$$H\sum_j |R_j\rangle\,|Y_j\rangle = \mathrm{i}\sum_j \frac{\partial}{\partial t}\,|R_j\rangle\,|Y_j\rangle \tag{35}$$

左乘 $\langle Y_i|$ 变为

$$\sum_j \langle Y_i|H|Y_j\rangle\,|R_j\rangle = \mathrm{i}\frac{\partial}{\partial t}\,|R_i\rangle \quad \Rightarrow \quad \sum_j H_{ij}\,|R_j\rangle = \mathrm{i}\frac{\partial}{\partial t}\,|R_i\rangle \tag{36}$$

其中每个 H_{ij} 都是 R 空间的算符．

从这里我们总结出，张量积空间中的矢量可以展开为一个小空间中的 N 个矢量分别与另一个小空间中的 N 个基底的张量积．所谓的“坐标”就是第

一个小空间中的 N 个矢量排成的“列矢量”. 张量积空间中的矢量左乘一个小空间中的矢量可以得到另一个小空间中的矢量. 张量积空间的算符分别左乘右乘一个小空间中的正交归一基底可以得到另一个小空间中的算符.

另一个例子是多通道散射问题（见小时物理笔记）. 就是将总波函数看做是某个变量 R （例如某个电子和质心的距离）的空间与剩下所有自由度的空间的张量积. 每个 channel 就对应剩下所有自由与的空间中的一个基底，系数就是 R 空间中的一个径向波函数 ψ_i. 如果有 N 个通道，就可以得到 N 个 ψ_i 的 coupled equation，其实就是矩阵形式的薛定谔方程.

另外，参考一下 wikipedia 的 tensor product 页面吧.

线性代数导航

几何矢量

高中数学和物理中最熟悉的矢量就是**几何矢量**，这里先回顾几何矢量（以下简称矢量）并引入一些新的概念.

矢量的存在与坐标系无关，可以将其想象成空间中的一些有长度有方向的箭头. 我们对它的位置不感兴趣，所有长度和方向相同的矢量都视为同一矢量. 对于讨论问题的不同，我们有时仅需要处于同一平面（**二维空间**）的所有矢量，有时需要**三维空间**中的所有矢量，最简单的情况下只需要沿某条线（**一维空间**）的所有矢量（这时我们可以规定一个正方向，且仅使用矢量的模长加正负号来表示矢量以简化书写）.

矢量的一些基本运算[370] 同样不需要有任何坐标系的概念，**矢量相加**按照三角形法则或平行四边形法则即可. **矢量数乘**就是把矢量的模长乘以一个实数，若乘以正数，方向不变，若乘以负数，取相反方向. **矢量的线性组合**是把若干矢量分别乘以一个实数再相加得到新的矢量.

矢量的**内积**[383] 等于一个矢量在另一个矢量上的投影长度乘以另一个矢量的模长得到一个实数，矢量的**模长**等于矢量与自身内积再开方，把矢量除以自身模长使模长变为单位长度的过程叫做**归一化**. 若两矢量内积为零，这两个矢量相互**正交**[4].

[4]对于几何矢量，正交就是方向垂直，不加区分.

两矢量**叉乘**[387]得到的矢量垂直于两矢量，模长为一个矢量在另一个矢量垂直方向的投影长度乘以另一个矢量的模长.

为了方便描述矢量之间的关系，我们选取一些**线性无关**的矢量作为所有几何矢量的**基底**，使空间中的任何矢量可以用这些基底的唯一一种线性组合来表示，N 维空间需要 N 个基底. 一般来说，基底不必互相正交. 我们先把这些基底排序，任意矢量表示成它们的线性组合时，把式中的 N 个系数按照顺序排列，就是该矢量的**坐标**，通常用列矢量表示. 由于线性组合的唯一性，每个矢量的坐标是唯一的.

为了方便计算任意矢量的坐标，往往取**正交归一**的基底[386]（所有基底模长为 1，任意两基底互相正交）. 这样，任意矢量的坐标都可以通过与基底的内积得到.

我们可以设计一种规则把某个空间的任意矢量对应（**映射**）到另一个矢量，叫做**变换**[5]. 如果对于某个变换，任意矢量线性组合的变换等于这些矢量分别进行该变换再线性组合，这个变换就是**线性变换**[454]. 在某组基底下，矢量的线性变换可以用其坐标的线性变换表示，并且可以写成矩阵与坐标列矢量相乘的形式.

旋转矩阵[402]可以有两种理解，一是矢量绕某个轴相对于当前的正交归一基底转动，其坐标产生了变换，二是矢量本身没有变，只是其坐标在两个不同的正交归一基底中不同. 这种矩阵的特点是所有列（行）矢量都正交归一，所以叫做**单位正交阵**. 单位正交阵的特点是逆矩阵等于转置矩阵.

矢量微积分

N 维矢量可以作为一个或多个标量的函数（**矢量函数**），可以看成是 N 个普通函数与矢量基底的数乘. 矢量函数同样可以对其自变量求导（或求偏导），也可以积分. 不同的是，矢量函数还可以进行曲线积分和面积分

施密特正交归一化

[5]更广义地，变换可以在不同的空间中进行，例如把一个三维空间中的矢量映射到一个二维空间中的矢量

预备知识 正交归一基底[386]

若在 M 维矢量空间中任意给出 $N \leqslant M$ 个线性无关的矢量，如何得到一组正交归一化的基底呢？我们可以用**施密特正交归一化（Schmidt orthonormalization）**．先看一个二维的例子

例 1　二维空间中的几何矢量

已知两个几何矢量 $\mathbf{v}_1, \mathbf{v}_2$ 坐标分别为 $(2,1)$，$(1,2)$．这两个矢量不共线，说明它们线性无关．但容易看出它们既不归一也不正交，下面来进行施密特正交归一化．

先把 $\mathbf{v}_1$ 归一化，并记为 $\mathbf{u}_1$

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|} = \frac{1}{\sqrt{5}}(2,1) \tag{1}$$

然后，用内积来计算 $\mathbf{v}_2$ 在 $\mathbf{v}_1$ 方向的投影长度

$$\mathbf{v}_2 \cdot \mathbf{u}_1 = \frac{4}{\sqrt{5}} \tag{2}$$

所以 $\mathbf{v}_2$ 在平行于 $\mathbf{v}_1$ 方向的分量为

$$\mathbf{v}_2^{\parallel} = (\mathbf{v}_2 \cdot \mathbf{u}_1)\mathbf{u}_1 = \frac{4}{5}(2,1) \tag{3}$$

将 $\mathbf{v}_2$ 减去平行的分量，就是垂直的分量

$$\mathbf{v}_2^{\perp} = \mathbf{v}_2 - \mathbf{v}_2^{\parallel} = \left(-\frac{3}{5}, \frac{6}{5}\right) \tag{4}$$

归一化并记为

$$\mathbf{u}_2 = \frac{\mathbf{v}_2^{\perp}}{\left|\mathbf{v}_2^{\perp}\right|} = \frac{1}{3\sqrt{5}}(-3,6) \tag{5}$$

现在可以验证，基底 $\mathbf{u}_1$ 和 $\mathbf{u}_2$ 是正交归一的，即 $|\mathbf{u}_1| = |\mathbf{u}_2| = 1$，且 $\mathbf{u}_1 \cdot \mathbf{u}_2 = 0$.

若给出 M 维矢量空间中的 N 个（$N \leqslant M$）线性无关矢量

1. 将第 1 个矢量归一化得到第 1 个基底
2. 将第 2 个矢量分解为与第 1 个矢量平行和垂直的两个分量，并将垂直分量归一化得到第 2 个基底

3. 将第 3 个矢量分解为三个部分，即分别平行于前两个基底的分量和一个垂直分量，并将垂直分量归一化得到第 3 个基底

4. 对第 $n = 4, \ldots, N$ 个基底重复该步骤，得到第 n 个基底

用公式来表示这个过程，就是：

$$\mathbf{v}_1^{\perp} = \mathbf{v}_1 \tag{6}$$

$$\mathbf{u}_i = \frac{\mathbf{v}_i^{\perp}}{\left|\mathbf{v}_i^{\perp}\right|} \qquad (i = 1, \ldots, N) \tag{7}$$

$$\mathbf{v}_i^{||} = \sum_{j=1}^{i-1} (\mathbf{v}_i \cdot \mathbf{u}_j)\mathbf{u}_j \qquad (i = 2, \ldots N) \tag{8}$$

$$\mathbf{v}_i^{\perp} = \mathbf{v}_i - \mathbf{v}_i^{||} \qquad (i = 2, \ldots N) \tag{9}$$

习题 1

对三维空间中的矢量 $(2, 1, 1)$，$(1, 2, 1)$ 和 $(1, 1, 2)$ 进行施密特正交归一化.

推导

这里来解释式 8 和式 9．我们假设已经知道 $i - 1$ 个正交归一的矢量，由于 N 维空间中必然存在 N 个正交归一基底，我们可以设剩下 u_i 的也已经知道（或者可以任意取）．于是 v_i 可以用基底展开为

$$\mathbf{v}_i = \sum_{j=1}^{N} c_j \mathbf{u}_j \tag{10}$$

式 8 得到前 $i - i$ 项之和

$$\mathbf{v}_i^{||} = \sum_{j=1}^{i-1} c_j \mathbf{u}_j \tag{11}$$

所以式 9 就是第 i 项到第 N 项之和

$$\mathbf{v}_i^{\perp} = \sum_{j=i}^{N} c_j \mathbf{u}_j \tag{12}$$

所以对 $j = 1, \ldots, i - 1$，都有 $\mathbf{v}_i^{\perp} \cdot \mathbf{u}_j = 0$．也就是说和已有的 $i - 1$ 个基底都正交.

平面旋转变换

预备知识 三角恒等式[112]，极坐标系[119]

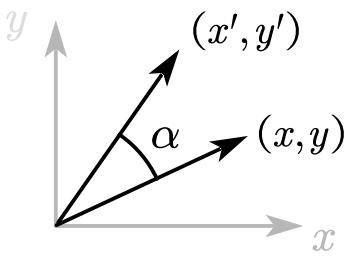


图 1: 把矢量绕原点旋转 α 角

已知直角坐标系中一点 $P(x,y)$，P 绕原点逆时针旋转 α 角（$\alpha \in R$）之后变为 $P'(x',y')$ 则有

$$x' = (\cos\alpha)x + (-\sin\alpha)y \tag{1}$$
$$y' = (\sin\alpha)x + (\cos\alpha)y \tag{2}$$

其逆变换如下，即已知 $P'(x',y')$ 求 $P(x,y)$

$$x = (\cos\alpha)x' + (\sin\alpha)y' \tag{3}$$
$$y = (-\sin\alpha)x' + (\cos\alpha)y' \tag{4}$$

绕任点旋转

要绕任意点 (x_0,y_0) 旋转，只需要先把矢量末端平移 $(-x_0,-y_0)$，绕原点旋转后再平移 (x_0,y_0) 即可

$$\begin{aligned} x' &= (\cos\alpha)(x-x_0) + (\sin\alpha)(y-y_0) + x_0 \\ y' &= (-\sin\alpha)(x-x_0) + (\cos\alpha)(y-y_0) + y_0 \end{aligned} \tag{5}$$

推导

平面上一点 $P(x,y)$ 也可以用极坐标 (r,θ) 表示，一般情况下令极点与原点重合，极径与 x 轴重合，则有

$$x = r\cos\theta \qquad y = r\sin\theta \tag{6}$$

把点 P 绕原点逆时针旋转 α 角变为 P'，则 P' 极坐标为 $(r,\theta+\alpha)$．根据上式计算为 P' 的直角坐标 (x',y') 并用两角和公式（式 3 [112]）化简如下

$$x' = r\cos(\theta+\alpha) = r\cos\theta\cos\alpha - r\sin\theta\sin\alpha = x\cos\alpha - y\sin\alpha \tag{7}$$

$$y' = r\sin(\theta+\alpha) = r\sin\theta\cos\alpha + r\cos\theta\sin\alpha = x\sin\alpha + y\cos\alpha \tag{8}$$

这就证明了式 1 和式 2 两式．

若要证式 3 和式 4 有两种方法．一是将式 1 和式 2 式中的 x,y 看成未知数，解二元一次方程组．另一种方法的思路是，既然 P 逆时针旋转 α 角为 P'，那么把 P' 顺时针旋转 α 角可得到 P．而“顺时针旋转 α 角”就是“逆时针旋转 $-\alpha$ 角”．把变换式 1 和式 2 中的 α 换为 $-\alpha$ 再化简得

$$x = \cos(-\alpha)x' - \sin(-\alpha)y' = \cos\alpha x' + \sin\alpha y' \tag{9}$$

$$y = \sin(-\alpha)x' + \cos(-\alpha)y' = -\sin\alpha x' + \cos\alpha y' \tag{10}$$

证毕．

对易算符

预备知识 映射[98]

定义两个算符（即映射）$A:X\to X$ 和 $B:X\to X$ 的**交换子（commutator）**（也叫**对易算符**）为

$$[A,B] = AB - BA \tag{1}$$

两个算符相等在这里的意义是它们作用在任意 $x\in X$ 上，结果都相等．算符的乘法 AB 在这里表示复合算符（即复合映射 $A\circ B$）．

根据定义，当 $AB=BA$ 就有 $[A,B]=0$，这时我们就说它们**对易（commutes）**或者**交换**，否则就**不对易**．

习题 1

一般来说，矩阵乘法不满足交换律. 试证明任意两个平面旋转矩阵[402]（二维几何矢量空间[370] 的算符）对易（实数 $\alpha \neq \beta$）

$$\mathbf{A} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} \tag{2}$$

习题 2

令 X 为所有连续可导函数的集合，定义算符 A 作用在任意 $f \in X$ 上的结果 Af 为函数 $(Af)(x) = xf(x)$，定义算符 B 为 $(Bf)(x) = \mathrm{d}f/\mathrm{d}x$. 试证明对易算符 $[A, B]$ 是单位算符，即 $[A, B]f = f$.

常见性质

$$[B, A] = -[A, B] \tag{3}$$

分配律

$$[A, B + C] = [A, B] + [A, C] \qquad [A + B, C] = [A, C] + [B, C] \tag{4}$$

$$[A, BC] = [A, B]C + B[A, C] \tag{5}$$

雅可比恒等式（Jacobi identity）

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0 \tag{6}$$

正交子空间

- 本词条存在未完成的内容.

预备知识 内积[471]，直和[465]，张成空间[447]

定义 1　正交子空间

一个内积空间 V 中，如果两个子空间 V_1 和 V_2 任意各选一个矢量 v_1 和 v_2，它们的内积都有 $\langle v_1, v_2 \rangle = 0$，那么我们就说者两个子空间是正交的.

构造正交子空间的一种简单的方法是，在 V 中找到两组矢量 $x_1,\dots,x_m$ 和 $y_1,\dots,y_m$，确保对任意 x_i 和 y_j 正交，那么 $x_1,\dots,x_m$ 张成[447] 的子空间必定和 $y_1,\dots,y_m$ 张成的子空间正交.

定理 1

从基底的角度来看，两个空间正交的充分必要条件是：如果从两空间各选一组基底 $\alpha_i\,(i=1,\dots,N_1)$ 和 $\beta_i\,(i=1,\dots,N_2)$，有对任意 i,j 都有 $\langle\alpha_i,\beta_j\rangle=0$.

习题 1

证明两个正交子空间中，只有零矢量是共同矢量.

正交子空间的直和

若两个正交子空间的维数分别为 N_1 和 N_2，它们之和等于母空间的维数 N，那么就说它们是**互补（complementary）**的. 若分别在这两个空间中取一组基底，那么将他们合并起来就得到了母空间中的一组基底.

特殊地，如果 V_1 和 V_2 的两组基底合并后仍然正交归一，那么合并后就得到了直和空间 $V=V_1\oplus V_2$ 中的一组正交归一基底. 但注意直和空间中的任意一组正交归一基底未必可以划分为 V_1，V_2 空间中的两组基底.

例 1

三维几何矢量空间中，建立直角坐标系，那么 $\hat{\mathbf{x}}$ 和 $\hat{\mathbf{y}}$ 张成的二维矢量空间（平面）与 $\hat{\mathbf{z}}$ 张成的一维矢量空间（直线）正交.

例 2

虽然 xy 平面和 xz 平面是两个垂直的平面，但它们并不是两个正交子空间. 例如矢量 $\hat{\mathbf{x}}$ 是两个平面共同的矢量，但 $\hat{\mathbf{x}}$ 和它本身不正交.

正交补

我们在“直和[465]”中已经定义了补空间的概念，现在来定义一种特殊的补空间.

定义 2　正交补空间

在 V 空间中，若 V_1 和 V_2 正交且 $V=V_1\oplus V_2$，那么 V_1 和 V_2 互为对方的正交补空间（**Orthogonal complement**），简称正交补.

未完成：正交补是唯一的

未完成：如何求正交补？

矩阵指数

定义

实数域上的指数函数 e^x 可以进行 Maclaulin 展开：

$$\mathrm{e}^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \tag{1}$$

展开式使得我们只需要用 x 的幂就可以表示指数 e^x．我们把这一点应用到矩阵中，就可以用方阵的幂来定义出矩阵的指数：

定义 1　矩阵指数

给定方阵 $\mathbf{M}$，定义

$$\mathrm{e}^{\mathbf{M}} = \sum_{n=0}^{\infty} \frac{\mathbf{M}^n}{n!} \tag{2}$$

并称之为矩阵 $\mathbf{M}$ 的指数（**matrix exponential**）．其中对于任意方阵 $\mathbf{M}$，都有 $\mathbf{M}^0 = \mathbf{I}$，$\mathbf{I}$ 是单位矩阵．

矩阵指数在常微分方程中非常常用，是用来解线性齐次方程组的利器．一个矩阵的指数本身还是一个矩阵．

矩阵指数的性质

相似变换的统一

由**过渡矩阵**[414] 可知，如果矩阵 $\mathbf{M}$ 在某基下表示一个线性变换，那么当基按过渡矩阵 $\mathbf{Q}$ 改变时，同一个线性变换的矩阵表示就变为 $\mathbf{Q}^{-1}\mathbf{M}\mathbf{Q}$．在原基下，$\mathrm{e}^{\mathbf{M}}$ 可以表示另一个线性变换，而它在 $\mathbf{Q}$ 下的变换是

$$\mathbf{Q}^{-1}\mathrm{e}^{\mathbf{M}}\mathbf{Q} = \mathrm{e}^{\mathbf{Q}^{-1}\mathbf{M}\mathbf{Q}} \tag{3}$$

也就是说，$\mathrm{e}^{\mathbf{M}}$ 所表示的变换，在基变换的时候，其矩阵表示的变换相当于给 $\mathbf{M}$ 变换后再取矩阵指数．这意味着我们也可以定义线性变换的指数——也可以反过来说，这是因为我们可以定义线性变换的指数，方式也是使用 Maclaulin 级数．

事实上，如果 $\mathcal{T}_i$ 表示若干线性变换，我们可以用映射的复合来定义线性变换的乘法：那么对于任意向量 $\mathbf{v}$，$\mathcal{T}_i^n(\mathbf{v}) = \mathcal{T}_i(\mathcal{T}_i^{n-1}(\mathbf{v}))$，其中 $\mathcal{T}_i^1 = \mathcal{T}_i$．类似地，也可以定义线性变换的加法：$(\mathcal{T}_1 + \mathcal{T}_2)(\mathbf{v}) = \mathcal{T}_1(\mathbf{v}) + \mathcal{T}_2(\mathbf{v})$．这样，有了乘法和加法，就可以计算线性变换的级数了，而 Maclaulin 级数就可以定义为其指数：

$$\mathrm{e}^{\mathcal{T}} = \sum_{n=0}^{\infty} \frac{\mathcal{T}^n}{n!} \tag{4}$$

其中 $\mathcal{T}^0$ 是恒等变换，对应单位矩阵．

式 3 意味着，如果 $\mathbf{M}$ 是 $\mathcal{T}$ 在某基下的矩阵表示，那么 $\mathrm{e}^{\mathcal{T}}$ 在该基下的矩阵表示就是 $\mathrm{e}^{\mathbf{M}}$．

运算性质

设 $\mathbf{M}, \mathbf{N} \in \mathrm{gl}(n, \mathbb{F})$，$a, b \in \mathbb{F}$，则容易得出以下性质：

如果 $\mathbf{MN} = \mathbf{NM}$，那么我们有 $\mathrm{e}^{\mathbf{M}}\mathrm{e}^{\mathbf{N}} = \mathrm{e}^{\mathbf{M}+\mathbf{N}}$．

$\mathrm{e}^{(\mathbf{M}^{\mathrm{T}})} = (\mathrm{e}^{\mathbf{M}})^{\mathrm{T}}$，$\mathrm{e}^{(\mathbf{M}^{\dagger})} = (\mathrm{e}^{\mathbf{M}})^{\dagger}$．

定理 1　矩阵指数的行列式与矩阵的迹

对于 $\mathbf{M} \in \mathrm{gl}(n, \mathbb{F})$，有 $\left|\mathrm{e}^{\mathbf{M}}\right| = \mathrm{e}^{\mathrm{tr}(\mathbf{M})}$．即：矩阵指数的行列式，等于矩阵迹的指数．

证明：

我们只需要考虑上三角矩阵 $\mathbf{M}$ 的情况即可，因为任何矩阵总可以通过相似变换变成上三角矩阵．此时，$\mathbf{M}$ 的迹就是主对角元素之和，而 $\mathbf{M}^k$ 的第 i 个主对角元素都是 $\mathbf{M}$ 的第 i 个主对角元素的 k 次方．

如果只看主对角元素，那么可以记 $\mathbf{M}$ 为 $(m_1, m_2, \cdots, m_n)$，其中各 m_i 是 $\mathbf{M}$ 的第 i 个元素．类似地，$\mathbf{M}^k$ 就可以记为 $(m_1^k, m_2^k, \cdots, m_n^k)$．代入矩阵指数的定义式，可得 $\mathrm{e}^{\mathbf{M}}$ 的对角线元素为 $(\mathrm{e}_1^m, \mathrm{e}_2^m, \cdots, \mathrm{e}_n^m)$．

由于上三角矩阵的乘积还是上三角矩阵，可知 $\mathrm{e}^{\mathbf{M}}$ 是上三角矩阵，因此 $\left|\mathrm{e}^{\mathbf{M}}\right| = \mathrm{e}_1^m \times \mathrm{e}_2^m \times \cdots \times \mathrm{e}_n^m = \mathrm{e}^{\mathrm{e}_1^m + \mathrm{e}_2^m + \cdots + \mathrm{e}_n^m} = \mathrm{e}^{\mathrm{tr}(\mathbf{M})}$．

证毕.

第四部分

代数学进阶

第一章

群论

群

预备知识 逻辑量词，整数[95]，映射[98]

基本概念

群是一种描述对称的数学对象，几乎每一个对称现象的背后都有一个群. 群论起源于伽罗瓦对于五次以上代数方程求根公式的思考，后被应用到了数学和物理的诸多领域当中.

我们在“映射[98]”的式 3 中定义了二元运算为 $A\times B\to C$，如果 $A=B=C$，即一个集合中的任意两个元素运算后的结果仍然在这个集合中，那么我们说这个二元运算是**封闭的（closed）**，有的地方也会称作**“闭合的”**. 二元运算的符号可以任意决定，如果使用 $\cdot$ 作为运算符，那么二元运算就可以简单写为 $a\cdot b=c$. 注意，这里的 $\cdot$ 只是表示某一个运算，不一定是我们通常的乘法或点乘运算.

定义 1　群

一个群 $(G,\cdot)$ 是在集合 G 上赋予了一个二元运算 $\cdot$ 的结构，该运算满足以下要求：

1. 封闭性（**closure**）：$\forall x,y\in G, x\cdot y\in G$，即任意 G 中元素 x,y 满足 $x\cdot y$ 仍是 G 中元素
2. 结合性（**associativity**）：$\forall x,y,z\in G, x\cdot(y\cdot z)=(x\cdot y)\cdot z$
3. 单位元（**identity element**）存在性：$\exists e\in G,\forall x\in G, e\cdot x=x\cdot e=x$
4. 逆元（**inverse element**）存在性：$\forall x\in G,\exists y\in G, x\cdot y=y\cdot x=e$. 通常我们会把这样的 y 称作 x 的逆元，并记为 x^{-1}

严格来说，这样的一个群应该表示为 $(G,\cdot)$，而 G 表示的是没有赋予运算的集合. 但是为了方便讨论，我们通常也会直接用定义群的这个集合来称呼这个群，比如简单地把上述定义的群叫做**群** G.

实际上，我们可以用更为弱化的公理系统来定义群，比如第 4 条只要求存在左逆元，即只要求 $\forall x \in G, \exists y \in G, y \cdot x = e$．在这种情况下我们仍然可以证明左逆元都是右逆元（定理 2）．有很多不同的弱化版本公理系统也能等价地定义出群来，但是为了方便理解，我们用了以上对称的公理系统．

集合的元素数量被称为集合的**基数**或**势**[92]，而群 G 的元素数量也可以称为群的**阶（order）**，记作 $|G|$．阶数有限的群称为有限群．以后如果没有特别说明，默认将群元素 x 的逆元记为 x^{-1}，将群的单位元记为 e．

群的例子

习题 1　二元群

定义一个只含有两个元素的集合，记为 $\{0,1\}$．在这个集合上定义运算“$+$”，由于只有四种运算方式，所以可以通过列举出每一个运算的结果来定义这个运算：

$$0+0=0 \qquad 0+1=1 \qquad 1+0=1 \qquad 1+1=0 \tag{1}$$

- 请用一个 2×2 的表格表示运算规则
- 请根据定义 1 验证这个二元集合配上运算 $+$ 构成一个群

在以上例子中，0 可以理解为“偶数”，1 可以理解为“奇数”，而群的运算可以理解为奇数加奇数等于偶数加偶数是偶数，而奇数加偶数是奇数．另外，注意尽管 $0+1=1+0$，我依然把它们分别单独写了出来，这是因为群的定义不要求**交换律**成立，也就是说，群运算允许 $x\cdot y \neq y\cdot x$．群元素选为 0 和 1 没有特殊原因，只是代表这是群里两个不同的元素而已，任何由两个元素构成的群我们都看作同一个．运算满足交换律的群被称为**阿贝尔群（abelian group）**或**交换群（commutative group）**，否则称为**非阿贝尔群（non-abelian group）**或**非交换群（non-commutative group）**．习惯上，我们把阿贝尔群的运算叫做加法，记为“$+$”，而把非阿贝尔群的运算叫做乘法，记为“$\cdot$”，甚至简化为没有符号，比如 $ab \neq ba$．不过，即使是非阿贝尔群中也可能存在两个元素 a 和 b，使得 $ab=ba$；这时我们说 a 和 b **交换（commute）**．

一般地，由于在朴素集合论[92] 中我们最多只讨论了集合的基数问题，集合中的元素具体如何命名是没有约束的，因此在集合论意义下元素数目相同的

集合都看作同一个. 比如说，我们认为 $\{\Delta, 3, K\}$ 和 $\{1, 2, 3\}$ 是同一个集合. 而现在在集合上定义了一个群运算以后所得到的群，即使构成它们的集合相同，群也可能由群运算的不同而产生不同的结构，从而被看作是不同的群.

例 1　整数加法群

所有整数的集合 $\mathbb{Z}$，配合通常的整数加法运算构成一个群.

例 2　n 元循环群

取一个由 n 个元素组成的集合 G，由于集合元素命名的任意性，不妨把 G 记为 $\{0, 1, \cdots n-1\}$，定义运算为模 n 的加法，即在一个有 n 个整点的钟表上的加法（见“整数[95]”）. 那么这个运算构成 G 上的一个群运算，所构成的群 G 称为 n 元循环群（**n-element cyclic group**），通常记为 C_n 或者 $\mathbb{Z}/n\mathbb{Z}$.

命名为 C_n 是取“cyclic”的含义，而命名为 $\mathbb{Z}/n\mathbb{Z}$ 是为了说明循环群是整数加法群 $\mathbb{Z}$ 的商群（例 1 [547]），而商群是将来会提到的重要概念.

例 3　n 元置换群

首先给定一个 n 元集合，记作 $K = \{1, 2, \cdots, n\}$，并将 K 中的元素按现有的顺序编号. 把 K 看作是 n 个桶中分别装了 1 个写着编号的球，初始状态下球的编号和桶的编号一致. 我们可以把球从桶里面拿出来并进行任意的置换，保持每个桶里还是只有一个球，但是球的编号不一定和桶的编号一致了. 每一个置换可以详细描述为“把 1 号桶的球和 2 号桶的球交换”，“把 1 号桶的球放入 3 号桶，3 号桶的放入 4 号桶，4 号桶的放入 1 号桶”等等.

我们用全体“置换”动作，即所有 K 到自身的一一映射，来作为元素，构成一个集合，称作 n 个元素的置换集合（n 元置换集），记为 S_n. S_n 一共有 $n!$ 个元素[1]. 从原始状态进行任意置换，所得到的结果状态和置换是一一对应的，所以我们也可以用“从原始状态进行置换 f 所得的结果”来表示置换 f 本身.

置换之间可以定义一个运算“$\circ$”，被称为置换间的复合，它是这样定义的：如果 f 和 g 是两个置换，那么 $g \circ f$ 就是先进行 f 置换，再进行 g 置换. 注意先后次序是从右到左进行的.

我们也可以这样来理解一个置换：原始状态下，n 号桶中的小球为 n. 进行一次 f 置换后，n 号桶中的小球就变成了 $f(n)$，再进行一次 g 置换，那么 n

[1] n 个桶各里装了 1 个小球，1 号桶有 n 种装球的可能性，2 号桶因此还剩下 $n-1$ 中可能性，以此类推，这 n 个桶一共有 $n \times (n-1) \times \cdots \times 1$ 种装球的可能性，每种可能性对应一个从初始状态而来的置换方式. 因此，置换的数量一共有 $n \times (n-1) \times \cdots \times 1 = n!$ 种

号桶中现在装的小球就变为 $g(f(n))$. 这个过程也可以看成是进行了一次 $g \circ f$ 运算，让 n 号桶中的小球变成 $g \circ f(n)$.

现在我们有了一个由置换组成的集合以及置换之间的运算，我们来验证在这个运算下，所有的置换构成一个群：

- 显然，任意两个置换的复合还是一个置换，因此该运算是封闭的.
- 映射的符合满足结合律可以通过带入任何一个 K 中的数字来验证.
- 单位元素是恒等映射，即保持所有数字不动的映射.
- 任意一个置换 f 由于是满射，所以对于每一个 K 中的数字 i, 都存在原像 j 满足 $f(j) = i$，又因为 f 是单射，所以原像唯一. 定义 $f^{-1} : K \to K$ 把每一个 i 对应到它的唯一的原像. 可以验证，这样定义了 f 的逆元.

因此，$(S_n, \circ)$ 是一个群. 注意当 $n > 2$ 时 S_n 是不交换（非阿贝尔）的.

在以上两个例子中可以看到，尽管元素数量一样，$\mathbb{Z}_{24}$ 和 S_4 的元素数量都是 24，但是前者是阿贝尔群，后者则不交换，显然两个群的运算结构不可能一样. 这是一个集合论意义上等价但群论意义上不等价的例子，换句话说，同一个集合上有时可以定义多个不同的运算使之成为不同的群.

例 4　二面体群

对称性的意思，是在某种变换（通常是翻转变换或者旋转变换）下保持不变的性质. 比如说，平面上的一个正方形，绕着几何中心旋转角度为 $\pi/4$（即角度制下的 $90°$）的时候和没有旋转是完全重合的，那么我们就说正方形是关于角度为 $\pi/4$ 的旋转对称的. 所有能够使得这个正方形不变的平面变换，比如特定角度的旋转、关于特定轴的翻转，配上变换间的复合运算（即先进行一个变换，再进行另一个），构成了一个 8 阶群（其元素构成为恒等变换，$\pi/4$ 的倍数的旋转和沿着四条对称轴的翻转），称为这个正方形的二面体群（**dihedral group**）记为 D_8. 感兴趣的读者可以尝试验证这些变换的复合是封闭的. 一般的，对于任意一个正多边形，保持其位置不变的旋转和翻转都会构成的一个群，统称为二面体群，记作 D_n.

对称群的概念非常重要且常见. 只要有“不变性”概念存在的地方就可能存在对称群. 一个事物的对称群的结构揭示了这个事物的不变性的特点.

习题 2

证明

- 实数集 $\mathbb{R}$ 以及通常的加法构成一个群.
- 实数集 $\mathbb{R}$ 除去 0 的集合，以及通常的乘法构成一个群.

例 5　n 阶可逆方阵群

给定域 F 上，全体 $n \times n$ 可逆矩阵[405] 构成的集合，配上矩阵乘法就构成了一个非阿贝尔群.

这样的群被简记为 $GL(n, F)$. 当不至于混淆时，这里的 F 一般会指全体实数或者全体复数，这时也会把该群简记为 $GL(n)$. 多数情况下，$GL(n)$ 也是不交换的.

唯一性定理

群中的单位元和每个元素的逆元都是唯一的：

定理 1　群运算满足消去律

给定一个群 G，若对于 $a, b \in G$ 有某个 $x \in G$ 使得 $ax = bx$，那么必然有 $a = b$；类似地，如果 $xa = xb$，也必然有 $a = b$.

该定理的证明留作习题[2].

唯一性是所有群都有的性质，但是等到我们讨论环[569] 的时候，由于环的乘法不要求逆元一定存在，我们没法对环证明这个唯一性. 事实上，很多环都没有唯一性.

习题 3　单位元的唯一性

从定理 1 可知左右单位元分别是唯一的. 请证明左单位 e_1 元等于右单位元 e_2（提示：将它们相乘）.

定理 2　逆元的唯一性

对任意 x，假设它存在一个左逆元 a，那么我们必然有 $ax = e$. 考虑结合性可知，$ae = a = ea = (ax)a = a(xa)$，于是 $xa = e$，即 a 也是 x 的右逆元. 也就是说，只要左逆元存在，那么右逆元也存在，并且等于左逆元. 反之亦然.

[2]提示：用 x^{-1} 去参与运算试试

习题 4

证明 $(ab)^{-1}=b^{-1}a^{-1}$

置换的奇偶性

预备知识 群[534]

在“群[534]”中，我们已经简单的介绍了由一个固定的有限集合 $\{1,2,...,n\}$ 上所有置换组成的集合连同映射的符合所构成的群：对称群. 在这一节我们将更加深入的讨论对称群的一些基本结构和性质.

置换的表示方式

一个置换实际上就是一个函数，要把一个函数表示出来要么给出表达式（一个通用的对应关系），要么把所有的点对应的函数值一一列举出来. 对于置换来说我们通常采用后者.

例 1

在 S_4 中设一个置换 $\sigma=\begin{pmatrix}1&2&3&4\\2&1&4&3\end{pmatrix}$

这个写法的意思是，作为一个置换. σ 把 1 对应到了 2，2 对应到 1，3 对应到 4，4 对应到 3. 值得注意的是，第一行虽然习惯于按照大小来排列从 1 到 n 这些数，但是实际上只要每个数字都出现一次就够了，不按照大小顺序来排列同样也可以表达出一个函数所包含的全部信息. 另外，把两行换位置互换所得到的置换 $\begin{pmatrix}2&1&4&3\\1&2&3&4\end{pmatrix}$ 就是 σ 的逆元.

对于一些比较特殊的置换，我们有更简便的表示方法.

例 2

仍然考虑 S_4，设 $\tau=\begin{pmatrix}1&2&3&4\\2&3&1&4\end{pmatrix}$ 在这个置换下，1，2，3 进行了“轮换"，即每个数字都被对应到它们中的下一个. 4 被保持不动. 这个 τ 简记为 $(1\,2\,3)$，它的意思是，每个数字在这个置换下被替换成后一个数字，最后一个被替换成第一个，也因此 $(1\,2\,3)=(3\,1\,2)=(2\,3\,1)$. 那些不受这个置换影响的数字则不

出现在这个记法当中．一个轮换中出现的数字的个数被称为这个轮换的长度．长度为 2 的轮换一般叫做对换，即两个数字对换位置，其他的保持不动．单位置换记为（1）．

定义 1

两个轮换中出现的数字（或者说被两个轮换改动位置的点）的集合若是不相交，则称这两个轮换不相交．

引理 1

不相交的轮换的乘积是交换的：设 $\tau_1 = (i_1\ i_2...i_k), \tau_2 = (j_1\ j_2...j_s)$，则 $\tau_1\tau_2 = \tau_2\tau_1$.

这个证明思路很简单，因为这两个轮换不相交，可以把它们看成是两个不同的集合重新排列．因此先对哪一个集合重排不影响最终结果．感兴趣的读者可以尝试把证明写出来．

一般来说，不是所有的置换都可以简记成一个轮换，换句话说，能写成轮换的只是一小部分置换．例 1 中的 σ 就不是一个轮换．但是，下面的定理表明，每个置换都可以唯一的表达成轮换的乘积：

定理 1

S_n 中每一个非单位的置换 σ 都可以写成不相交的长度大于等于 2 的轮换的乘积．并且这个分解在不计次序的情况下是唯一的．

这里简单说一下证明思路，具体的证明只需补充细节即可．考虑 σ 的一个动点（被其改变位置的一个数字）：i_1. 记 $i_2 = \sigma(i_1)...i_n = \sigma(i_{n-1})$ 直到出现某个 $i_k = i_1$ 为止（因为一共只有有限多个数字，这种情况必定发生）．我们得到一个轮换 $\sigma_1 = (i_1\ i_2...i_k)$. σ 与 σ_1 在 i_1 到 i_k 这些数上的作用是完全相同的，而 σ_1 保持其他数字不动，可以理解为 σ_1 是 σ 的一部分．显然 σ_1 的长度至少是 2（否则 i_1 将是不动点）．现在考虑 $\sigma\sigma_1^{-1}$. 这个置换就是 σ 相比于 σ_1 多出的部分．它的动点显然少于 σ, 对动点个数做归纳（或者重复上述步骤）可以得到定理中 σ 的轮换分解．

例 3

例 1 中的 $\sigma = \begin{pmatrix} 1\,2\,3\,4 \\ 2\,1\,4\,3 \end{pmatrix}$ 可以分解成 $(1\ 2)(3\ 4)$

置换的奇偶性

引理 2

任意一个轮换 $(i_1\ i_2 ... i_k)$ 都可以写成对换的乘积. 因此根据定理 1，任意置换也可以写成对换的乘积.

证明：$(i_1\ i_2 ... i_k) = (i_1\ i_k)(i_1\ i_{k-1})...(i_1\ i_2)$.

例 4

上述分解并不唯一，例如在 S_3 中，$(1\ 2\ 3) = (1\ 3)(1\ 2) = (2\ 3)(1\ 3)$.

不仅分解的方式不唯一，不同的分解用到的对换的个数甚至也是不同的. 但是尽管如此，不同分解所需对换个数的奇偶性是一致.

定理 2　置换的奇偶性

设 $\sigma \in S_n$, 且有分解 $\sigma = \tau_1\tau_2...\tau_k = \pi_1\pi_2...\pi_s$. 其中 τ_j, π_l 均为对换. 则 k 与 s 有相同的奇偶性.

此证明并不复杂，但比较繁琐，因此略去. 从定理中可以看出，任意的两个分解中用到的对换个数要么同时是奇数，要么同时是偶数. 换句话是，这个奇偶性不依赖于分解，只与对换 σ 本身有关.

定义 2

若一个置换 $\sigma \in S_n$ 可以写成偶数个对换的乘积，则称 σ 为偶置换. 若可以写成奇数个对换的乘积，则称 σ 为奇置换.

习题 1

证明对于任意 $\sigma \in S_n$，σ 与 σ^{-1} 有相同的奇偶性.

习题 2

证明在 S_n 当中，奇置换与偶置换各占一半. (提示，固定一个对换 τ，考虑映射：奇置换 $\to$ 偶置换，$\sigma \mapsto \sigma\tau$

子群

预备知识　群[534]

子群

定义 1　子群

给定一个群 $(G,\cdot)$，如果集合 G 有一个子集 H，使得 $e\in H$ 且 H 中的元素在运算 $\cdot$ 下仍然封闭，那么显然[3] $(H,\cdot)$ 也构成一个群．称 H 是群 G 的子群（**subgroup**）．

虽然群和子群的联系很紧密，但是我们通常还是把它们看作由完全无关的集合所构成的，只不过可以自然地应用已经存在的群运算和子集关系来定义子群的运算．这样，将已有的运算直接用在子集上，有时被称作在子集上**导出**或**诱导**（**induce**）了一个运算，有时也称子集上的运算是**限制在子集上的运算**．比如在定义里，群 H 的运算实际上被认为是和 G 的运算不一样，严格来说应该记为 $\cdot|_H$，意思是“限制在 H 上的 $\cdot$”．但是不至于引起混淆的时候，也可以简单记为“$\cdot$”，并认为是同一个运算．

一般来说，当我们说 H 是一个子群时，强调的是 H 和 G 的关系；但如果我们说 H 是一个群，我们关心的是 H 本身作为群的性质，而没有强调它和其它群的关系．

例 1　整数加群的子群

我们已经知道，全体整数配上通常的加法，能构成一个群（例 1 [534]），可以简称为“整数加群”．整数加群的集合中包括了全体整数．如果我们只取偶数，会发现偶数之间的加法和加法的逆（减法）的结果还是偶数，也就是说，全体偶数对于整数加法是封闭的，可以构成整数加群的一个子群．这个子群，我们记为 $2\mathbb{Z}$．

类似地，如果取某个整数 n 的全体倍数构成集合，这个集合上的加法也是封闭的，构成子群，记为 $n\mathbb{Z}$．比如说，$3\mathbb{Z}=\{0,\pm3,\pm6,\pm9,\cdots\}$．特别地，也可以把 $\mathbb{Z}$ 本身看成是 1 的倍数构成的集合．

定理 1　子群的判别法

[3] 由于已经知道 $(G,\cdot)$ 构成一个群了，群的四条公理中，结合性、单位元存在性以及逆元存在性都被满足了．

设有群 G 及其子集 H，则 H 是 G 的子群当且仅当：对于任意 $x, y \in H$，有 $x^{-1}y \in H$.

充分性证明：

- 单位元存在性：我们需要证明群 G 中的单位元属于 H. 取 $x \in H$，且 $y = x$，那么 $e = y^{-1}x \in H$.
- 逆元存在性：由于 x 是群 G 的元，x^{-1} 必然存在，但我们还要证明 $x^{-1} \in H$. 取 $x \in H$，$y = e \in H$，那么 $x^{-1} = x^{-1}y \in H$.
- 封闭性：取 $x, y \in H$，则 $x^{-1} \in H$，进而 $xy = (x^{-1})^{-1}y \in H$.
- 结合性自然继承自 G 的结合性.

必要性证明：如果 H 是个群，显然对于任意 $x, y \in H$，有 $x^{-1}y \in H$.
证毕.

陪集和同余

预备知识 子群[541]

定义 1　左陪集

给定一个群 G 和它的一个子群 H. 从 G 中任意挑一个元素 x 出来，用它来左乘 H 中的每一个元素，得到一个集合 $\{x, xh_1, xh_2, \cdots\}$，这里的 h_n 要取遍每一个在 H 中的元素. 这个集合，也可以写为 $\{xh | h \in H\}$，被叫做子群 H 关于元素 x 的左陪集（**left coset**），记作 xH.

类似地还可以定义右陪集，不过二者选其一作为代表来研究就可以了，我们习惯上主要研究左陪集.

注意这个表示方法：$xH = \{xh | h \in H\}$. 这是一种非常简洁的表达方法，可以理解成 x 逐个左乘 H 中的元素得到的集合，也可以理解成是 x 和 H 之间的一个运算，结果是另一个集合. 类似地，我们也可以用群里任意两个子集（不一定要求是子群）A 和 B 来生成一个新的子集 $AB = \{ab | a \in A, b \in B\}$，

就是用 A 中的每一个元素去左乘 B 中的每一个元素，得到的所有结果的集合. 当然，这种表示方法也可以推广到一切“元素之间可以进行运算的集合”：集合 A 和 B 之间的运算，就是 A 和 B 所有可能的元素运算的结果构成的集合. 比如说，考虑到整数集合上可以进行小学所学的乘法和加法运算，我们也可以把 $n\mathbb{Z}$ 的表示方法理解为用 n 去乘全体整数所得到的集合，这个集合在加法下构成群[4].

如果 $h \in H$，而 H 是一个**有限群**，那么由于封闭性，$hH = H$. 这是因为，用 h 去左乘 H 中的一切元素（包括 h 自己），那么一方面由于封闭性，运算结果还是在 H 内部；另一方面由于群运算的唯一性，每一个左乘运算都不相同. 这个论断不能简单地用于无限群.

所以，如果我们用 xH 中任意元素 xh_0 来左乘 H，得到的 xh_0H 仍然是同一个左陪集：$xH = xh_0H$. 所以我们可以用左陪集中的任何一个元素 x' 来作为代表，把这个左陪集写成 $x'H$.

特别地，考察这个形式的集合[5]：$\{e, h, hh, hhh, hhhh, \cdots\}$. 那么同样地由于封闭性和运算唯一性可知，这个集合还是群 H 的一个子集. 特别地，H 的群运算限制在这个集合上能构成一个循环群（例 2 [534]）. 只要我们把 n 个 h 相乘的结果记为 h^n，n 个 h^{-1} 相乘的结果记为 h^{-n}，那么如此生成的循环群就可以用指数的加法运算来处理了. 由单个元素可以生成循环群这一概念，我们引入以下定义：

定义 2　阶

给定一个群 G 和一个 $g \in G$，如果存在一个正整数 n 使得 $g^n = e$，那么我们称 g 是一个有限阶的元素. 特别地，所有满足条件的 n 中最小的那一个，被称为 g 的阶（**order**）或者指数，记为 $\operatorname{ord} g$. 特别地，如果任何整数 n 都不能使 $g^n = e$，那么记 $\operatorname{ord} g = \infty$；规定 $\operatorname{ord} e = 0$.

例 1　$n\mathbb{Z}$ 的左陪集

整数加群的群运算是通常的加法，所以我们可以把元素 k 所在的左陪集记为 $k + n\mathbb{Z}$. 比如说，当 n 为 6，k 为 1 时，$1 + 6\mathbb{Z} = \{\cdots, -11, -5, 1, 7, 13, \cdots\}$；当 k 为 8 时，$8 + 6\mathbb{Z} = \{\cdots, -10, -4, 2, 8, 14, \cdots\} = 2 + 6\mathbb{Z}$.

[4]在乘法下一般不构成群，除非 n 是素数，在这种情况下 $n\mathbb{Z}$ 配合乘法和加法就构成了最常见的一种有限域.

[5]其中省略号表示一直列举下去，但排除与前面重复的元素. 所以这个集合可能是有限的.

$6\mathbb{Z}$ 一共有 6 个不同的左陪集；类似地，$n\mathbb{Z}$ 一共有 n 个不同的左陪集.

注意左陪集不一定是子群，例如 $1+6\mathbb{Z}$ 中没有单位元.

例 2　$3\mathbb{Z}$ 的子群和左陪集

$3\mathbb{Z}$ 虽然是 $\mathbb{Z}$ 的子群，但是它也可以从中再分离出子群来. 由于 $3\mathbb{Z}$ 的集合是由全体 3 的倍数构成的，因此全体 6 的倍数构成的集合 $6\mathbb{Z}$ 是 $3\mathbb{Z}$ 的真子集，我们已经知道它构成群了. 也就是说，$6\mathbb{Z}$ 既是 $\mathbb{Z}$ 的子群，又是 $3\mathbb{Z}$ 的子群，甚至还是 $2\mathbb{Z}$ 的子群.

从例 1 中我们知道，$6\mathbb{Z}$ 作为 $\mathbb{Z}$ 的子群，有 6 个左陪集；但作为 $3\mathbb{Z}$ 的左陪集的时候只有 $6/3=2$ 个左陪集.

下图可以形象地表示 $\mathbb{Z}$，$3\mathbb{Z}$ 以及 $6\mathbb{Z}$ 之间的关系，一般的群和子群的关系也可以类似理解.

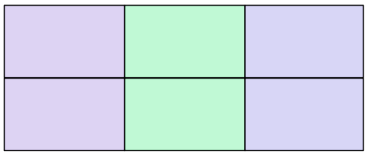


图 1: 子群和陪集的示意图. 整张图表示 $\mathbb{Z}$ 本身；中间的两个绿色区域表示 $3\mathbb{Z}$，左边的两个紫色区域表示 $1+3\mathbb{Z}$ 而右边的两个紫色区域表示 $2+3\mathbb{Z}$；中间上面的绿色区域表示 $6\mathbb{Z}$，而下面的绿色区域表示 $3+6\mathbb{Z}$，剩下的四个紫色区域分别表示 $6\mathbb{Z}$ 的另外四个左陪集.

在这个图中可以清楚地看到，$3\mathbb{Z}=6\mathbb{Z}\cup(3+6\mathbb{Z})$. 这很容易理解，因为 3 的倍数无非两种情况，6 的倍数或者 6 的倍数再加 3.

左陪集的意义是将群划分成互不相交的子集，这是一个等价类划分，也就是说，“x 和 y 属于同一个左陪集”是一个等价关系[858]. 于是我们有了如下定理：

定理 1

左陪集划分是一个等价类划分.

证明：

给定一个群 G，两个元素 $x, y \in G$，再给定它的一个子集 H，那么“y 在左陪集 xH 中”的等价表述，可以是“$y \in xH$”；在这个属于关系两边同时左乘一个 x^{-1}，还能得到更常用的等价表述：$x^{-1}y \in H$[6].

我们用最后这个等价表述来检查，“在同一个左陪集中”这一关系，是否是等价关系.

- 自反性：对于任意的 $x \in G$，由于 $e \in H$，故显然有 $x = xe \in xH$. 因此 x 和自己在同一个左陪集中.
- 对称性：如果 $x^{-1}y \in H$，由于 H 是个群，故 $y^{-1}x = (x^{-1}y)^{-1} \in H$. 因此 x 也在左陪集 yH 中. 对称性还说明，如果 y 在 xH 中，那么 x 也在 yH 中，因此这两个表述可以合二为一为“x，y 在同一个左陪集中”.
- 传递性：如果 $x^{-1}y \in H, y^{-1}z \in H$，那么由于 H 的封闭性，$(x^{-1}y)(y^{-1}x) \in H$，拆开括号后得到 $x^{-1}z \in H$. 因此 z 也在 xH 中.

到此，我们证明了“在同一个左陪集中”是一个等价关系. 由这个等价关系划分的左陪集，是一种等价划分，左陪集彼此互不相交.

证毕

定义 3　同余

对于群 G 和其子群 H，如果 $x, y \in G$ 满足 x, y 在同一个左陪集中，那么我们称 x, y 模 H 同余[7].

知道了左陪集是等价类，我们很容易发现群的一个优美的结构. 在集合论中，子集的基数[92] 可以是任意的，只要它小于等于原集合的基数就可以；但是子群的阶却必须能够整除原群的阶. 这就是群论的**拉格朗日定理（Lagrange's Theorem）**：

定理 2　拉格朗日定理

给定群 G，如果 H 是 G 的子群，那么 $|H|$ 可以整除 $|G|$.

[6]如果你不理解为什么可以像解方程一样两边同时乘以一个元素，请再琢磨琢磨上文中 xH 的定义是什么.

[7]对比整数[95] 中的同余概念. 两个“同余”有什么联系吗？

证明：

假设 $x \in G$. 根据 xH 的定义，$|xH| \leq |H|$. 又由群运算的唯一性，这个不等式应该取等号：$|xH| = |H|$. 所以每个左陪集的基数都一样大.

左陪集彼此不相交，每个元素 x 都属于左陪集 xH，因此 $|G|$ 等于各个左陪集的基数之和，也就是 $|H|$ 的倍数.

证毕

拉格朗日定理还说明，任何有限群 G 的元素，其指数都是 $|G|$ 的因子.

正规子群

预备知识 词条示例[??]

正规子群

在上文中，我们在群上定义了元素和子集间、子集和子集间的运算，还知道了左陪集是一种等价类划分. 假设我们有群 G 和它的子群 H，能不能用 H 的全部左陪集当元素，构成一个集合，并在上面定义一个和 G 的运算相关的群运算呢？

最直接的想法，就是直接进行左陪集之间的运算：对于左陪集 A 和 B，有 $AB = \{ab|a \in A, b \in B\}$. 但是一般来说，如果 G 的运算不是可交换的，我们这样算出来的 AB 甚至不一定是一个左陪集，那么这个左陪集之间的运算甚至不满足封闭性，自然不能是个群运算了.

如何让左陪集之间的运算满足封闭性呢？

假设我们有 $x, y \in G$，它们对应的左陪集分别是 xH 和 yH. 由于 $xH = \{xh_1|h_1 \in H\}$,$yH = \{yh_2|h_2 \in H\}$,那么根据定义有:$xHyH = \{xh_1yh_2|h_1, h_2 \in H\}$. 每一个 xh_1yh_2 都在左陪集 xh_1yH 中，我们希望每一个 h_1 对应的 xh_1yH 都是同一个左陪集.

在定理 1 [543] 的证明中我们提到了一个判断两元素是否在同一个左陪集中的方法. 现在假设 xh_1y 和 xh_2y 在同一个左陪集里，应用判断方法得到：$(xh_1y)^{-1}(xh_2y) \in H$. 由于 $(xh_1y)^{-1} = y^{-1}h_1^{-1}x^{-1}$，我们有：

$$y^{-1}h_1^{-1}x^{-1}xh_2y \in H \tag{1}$$

也就是说

$$y^{-1}h_1^{-1}h_2y \in H \tag{2}$$

由于这里的 y 是 G 中任意元素，h_1 和 h_2 是 H 中任意元素，故我们要求 $y^{-1}Hy \subseteq H$ 对于任何 $y \in G$ 都成立，其中 $y^{-1}Hy = \{y^{-1}h_iy | h_i \in H\}$. 巧的是，让 H 满足了这个要求之后，封闭性就满足了.

习题 1

证明：对于群 G 和它的子群 H，如果 H 满足对于任意的 $g \in G$，都有 $g^{-1}Hg \subseteq H$，那么对于任意的 $x, y \in G$，$xHyH$ 中的元素都在同一个左陪集中.

推论 1

由于群运算的唯一性，从 $g^{-1}Hg \subseteq H$ 我们还可以推知 $|g^{-1}Hg| = |H|$，因此所需要的条件等价于 $g^{-1}Hg = H$，即 $Hg = gH$，也就是左陪集等于右陪集.

定义 1　正规子群

满足推论 1 中条件的子群，被称作正规子群（**normal subgroup**）. 我们把“H 是 G 的正规子群”简记为 $H \triangleleft G$ 或 $G \triangleright H$[8].

现在我们考察一下，如果 H 满足了条件，那么在左陪集的集合中通过 G 的运算而导出的集合间的运算，是不是一个群运算.

- 封闭性：由习题 1 知，该运算有封闭性.
- 结合性：由于 G 的运算具有结合性可知，设 $x, y, z \in G$，那么 $(xh_1yh_2)zh_3 = xh_1(yh_2zh_3)$ 对任意的 $h_n \in H$ 都成立，故而 $(xHyH)zH = xH(yHzH)$ 成立.
- 单位元存在性：H 就是单位元.
- 逆元存在性：$xHx^{-1}H = HH = H$.

这样一来，在左陪集的集合上也诱导了一个群运算，构成了一个群. 我们把这个群叫做群 G 模去正规子群 H 的**商群**（**quotient group**），记作 G/H. 由拉格朗日定理（定理 2）的证明过程可知，$|G/H| = |G|/|H|$.

[8] 三角形总指向子群.

例 1　$\mathbb{Z}$ 的正规子群和商群

$n\mathbb{Z}$ 是 $\mathbb{Z}$ 的正规子群，这对任何 n 都成立．证明很简单，因为 $\mathbb{Z}$ 是交换群，所以元素位置可以自由互换，因此 $g^{-1}Hg = g^{-1}gH = H$ 对任何子集都成立，因此交换群的子群都是正规子群．

考虑 $\mathbb{Z}$ 模去 $3\mathbb{Z}$ 得到的商群，我们记为 $\mathbb{Z}/3\mathbb{Z} = \mathbb{Z}_3$．这个商群一共有 3 个元素，每个元素都是一个左陪集：$0+3\mathbb{Z} = 3\mathbb{Z} = \{\cdots, -6, -3, 0, 3, 6, \cdots\}$，$1+3\mathbb{Z} = \{\cdots, -5, -2, 1, 4, 7, \cdots\}$ 和 $2+3\mathbb{Z} = \{\cdots, -4, -1, 2, 5, 8, \cdots\}$．换句话说，集合 $\mathbb{Z}_3 = \{3\mathbb{Z}, 1+3\mathbb{Z}, 2+3\mathbb{Z}\}$．

注意，虽然 $\mathbb{Z}_3$ 的元素是集合，我们仍然要把它们看成一个个的元素．事实上，$\mathbb{Z}_n$ 的群结构和一个有着 n 个数字的钟表（例 2 [534]）是一模一样的，因此我们也用 $\mathbb{Z}_n$ 来表示循环群．

例 2　置换群和交错群

我们知道，置换群 S_n（例 3）是 n 个元素的变换所构成的群．为了简洁地表达变换，我们可以用以下记号：

$(1,2,3)$ 表示一个变换，它是将 1 号桶中的球放入 2 号桶，2 号桶中的放入 3 号桶，3 号桶中的又放入 1 号桶．显然，$(2,3,1)$ 表示的是同一个变换．$(1,2,3)(4,5)$ 表示的是先进行 $(1,2,3)$ 变换，再进行 $(4,5)$ 变换．显然，这两个变换没有涉及相同数字，所以它们是彼此分离的，可以交换：$(1,2,3)(4,5) = (4,5)(1,2,3)$．

如果两个变换有相同数字，那么我们可以把它们合并为一个括号：$(1,2)(2,3)$ 代表先“把 1 号桶中的球放入 2 号桶，2 号桶中的放入 1 号桶”，再“把 2 号桶中的球放入 3 号桶，3 号桶中的放入 2 号桶”，最终结果是“1 号桶中的球放入 3 号桶，2 号桶中的放入 1 号桶，3 号桶中的球放入 2 号桶”，也就是说，$(1,2)(2,3) = (1,3,2)$．注意 $(1,2,3) \neq (1,3,2)$．

只涉及到两个元素之间的变换显然是最简单的，我们称之为对换．一切变换都是由若干个对换相乘（复合运算）得到的．可以证明（例 1 [565]），每进行一次对换，桶中球的“逆序数”的奇偶性会改变．初始状态下的逆序数是 0，一个偶数；因此如果一个变换可以被奇数个对换组合得到，那么这个变换会使得桶中球的逆序数变成奇数，这个变换就叫做奇变换；同样地，偶数个对换组合得到的变换称为偶变换，它会使得桶中球的逆序数仍为偶数．

偶变换之间的复合仍然能得到偶变换，因此全体偶变换构成置换群 S_n 的一个子群，称为交错群（**alternating group**），记为 A_n．交错群是置换群的一个

正规子群，元素数量是置换群的一半——也就是说，$|S_n/A_n| = 2$ 对任何 n 都成立.

例 3　中心

给定群 G 和它的一个子集 S，记 $C(S)$ 为 G 中所有能和 S 中元素交换的元素之集合，即 $C(S) = \{g \in G | \forall s \in S, gs = sg\}$，称之为 S 在 G 中的中心化子（**centralizer**），有时为了强调也记为 $C_G(S)$. 特别地，称 $C(G)$ 是 G 的中心（**center**）.

考虑到对于任意 $c \in C(G)$ 和任意 $g \in G$ 都有 $cg = gc$，那么任取 $h \in G$，也必有 $hch^{-1}g = hh^{-1}cg = cg = gc = gchh^{-1} = ghch^{-1}$，即 $hch^{-1} \in C(G)$. 因此，$C(G) \triangleright G$.

任意给定群 G 的子群 H，它不一定是正规子群，但我们可以适当“修剪”群 G 来获得一个子群，使得 H 在这个子群里是正规子群. 这样的子群就是以下定义的正规化子：

定义 2　正规化子

给定群 G 及其子群 S. 如果 H 是包含 S 的最小的子群[9]，记 $N(S) = \{n \in G | nH = Hn\}$，那么 $N(S)$ 构成一个子群，而 $H \triangleright N(S)$. 称 $N(S)$ 是 S 的正规化子（**normalizer**）.

注意区分中心化子和正规化子的定义，对于同样的 S，其正规化子比中心化子的定义要弱一些，并且正规化子本身不一定是 G 的正规子群.

小结

对于任何群 $G \triangleright H$，所有左陪集的元素数量都是 $|H|$，包括 H 本身. 这些左陪集彼此没有交集，但群 G 是所有左陪集的并集，我们由此得到了拉格朗日定理，揭示了一个任何群都具有的简洁的结构特征. 商群 G/H 的元素数量，正是 G 和 H 的元素数量的商.

当子群 H 满足，对于任意的 $g \in G$ 都有 $g^{-1}Hg = H$ 时，H 就成了一个正规子群. 只有正规子群才能生成商群；非正规的子群划分出来的等价类，按照二元关系[858] 中提到的方法构成的商集，却并不能用 G 的运算来导出一个群运

[9] 我们通常称 H 是由 S 生成的子群，用 $< S >$ 表示.

算．当然，你也可以把这个商集看成和 G 完全无关的集合，给它独立定义一个群运算，但那样这个群就和 G 没有一点关系了，不能叫做 G 的商群．

G/H 的群运算是 G 的运算导出来的，所以继承了 G 中很多相似的性质；但是 G/H 把同一个左陪集中各个元素的差别都抹去了，相当于一个简化版的 G．所以说，商群继承了原群的部分特征，但是没有完全继承．

两个定义相似的概念：中心化子和正规化子．

接下来当我们讲到群之间的映射的时候，商群与原群“似而不同”的结构非常重要．

群的直积和直和

预备知识 正规子群[547]

直积

群的直积，是在群作为集合的笛卡尔积上，由群运算自然导出的一个群．

定义 1　两个群的直积

给定群 G 和 H，群运算的符号省略．在集合 $G\times H$ 上定义运算：对于任意 $(g_i,h_i)\in G\times H$，有 $(g_1,h_1)(g_2,h_2)=(g_1g_2,h_1h_2)$．集合 $G\times H$ 配合以上定义的运算，构成一个群，称为群 G 和 H 的直积（**direct product**）．

容易看出，两个群直积的单位元是 (e,e)——注意这里的两个 e 分属不同的群，通常是不同的元素．

这个定义分割开了参与直积的不同群的运算，因此可以很方便地直接推广到任意多个群的直积：

定义 2　任意多个群的直积

给定任意多个群 $\{H_i\}$，在这些群作为集合的笛卡尔积上，各分量运算分别进行运算，且遵循各自所属群的运算规则．该笛卡尔积配合该运算规则构成一个群，称为这些群的直积（**direct product**），记为 $\bigotimes_i H_i$ 或 $\prod_i H_i$．

在群论中还有一个和直积很类似的概念，常使人混淆，这就是群的**直和**．

直和实际上是直积的一个特例：任意给定群，都可以使用这些群来构造直积，但是直和指的是已经给定了一个群，使用它的特定子群来生成它.

定义 3　群的直和

当 G 和 H 都是交换群时，我们也称 $G \times H$ 为这两个群的直和（**direct sum**），此时也可以把它表示为 $G + H$. 另外，任意个交换群的直积 $\bigotimes_i H_i$ 也可以表示为 $\bigoplus_i H_i$ 或 $\sum_i H_i$.

直和这一术语的来源不难理解：交换群的群运算通常被称为“加法”.

半直积

群的直积可以推广为以下概念：

定义 4　半直积

给定群 G，如果有 G 的一个正规子群 N 和一个子群 H，使得 $G = NH$[10]，并且 $N \cap H = \{e\}$，那么我们称 G 是 N 和 H 的半直积（**semi-direct product**），记为 $G = N \rtimes H$.

可以注意到，直积是半直积的一种，只要把 $\{(g, e)\}$ 和 $\{g\}$ 等同、把 $\{(e, h)\}$ 和 $\{h\}$ 等同即可. 这样，尽管本节中直积是用“运算的笛卡尔积”来定义的，而半直积是用“已有的群运算”来定义的，这两个在特定情况下是等价的.

半直积不一定是直积，这是因为定义中我们只要求参与运算的两个群中的一个为正规子群，而如果两个群 G 和 H 进行直积，那么容易证明它们俩都是群 $G \times H$ 的正规子群. 从这也可以看出来为什么此处半直积的定义要先给出 G，而不是像直积的定义一样直接用两个群的乘积得到 $G \times H$.

事实上，我们也可以用以上定义半直积的语言来描述直积：给定群 G，如果有 G 的两个正规子群 H 和 N，满足 $H \cap N = \{e\}$，并且 $G = NH$，那么称 G 是 N 和 H 的直积，记为 $G = N \times H$.

半直积的定义也可以不依赖于预先给定的 G，只是这样会稍显复杂一些：

定义 5　半直积

给定群 N 和 H，并且有同态：$f : H \to \operatorname{Aut} N$. 在笛卡尔积集合 $N \times H$ 上定义运算：对于 $n_i \in N, h_i \in H$，有 $(n_1, h_1) \cdot (n_2, h_2) = (n_1 \cdot f_{h_1}(n_2), h_1 \cdot h_2)$.

[10] 就是说，集合 $G = \{nh | n \in N, h \in H\}$.

集合 $N \times H$ 配合这个运算可以得到一个群，称为群 N 和群 H 关于同态 f 的半直积，记为 $N \rtimes_f H$.

我们梳理一下这个定义. 为了定义半直积，我们要用到三个部分，即两个群 N 和 H，再加上一个同态 f. 要注意，f 不是 H 到 N 的同态，而是到 $\mathrm{Aut}\, N$ 的，后者是 N 的**自同构群**. 也就是说，每个 $f(h)$ 表示一个 $N \to N$ 的自同构映射，为了方便，简写为 f_h.

单群

预备知识 群的直积和直和[551]

单群和各种序列

单群可以和素数类比. 我们知道，任何一个非平凡的整数（即除了 0 和 1）都可以分解为素数的乘积，素数是正整数的“砖石”. 类似地，任何一个具有正规子群的群都可以分解为其正规子群和另一个群的半直积，而不存在**非平凡**正规子群的群就可以类比为群的“砖石”. 这样的砖石被称为**单群**.

定义 1　单群

如果群 G 不存在正规子群，那么称其为一个单群（**simple group**）.

群的分解和数字的分解有很多类似之处，但依然有很多不同，因此本节中使用的类比更多地是为了方便建立直观印象，而非对号入座，请注意.

每一个非平凡整数都可以通过因子分解，一步步拆成素因子更少的数，直到得到素数. 比如说，60 可以拆成 30，30 可以拆成 15，15 再拆成 3，而 3 就无法再拆了. 群也类似，可以把群拆成自己的正规子群，直到不再能拆，也就是得到单群.

定义 2　正规子群列

给定 n 个群 G_i，如果每个 G_{i+1} 都是 G_i 的正规子群，而 G_n 是单群，那么可以得到一个序列 $G_0 \triangleright G_1 \triangleright \cdots \triangleright G_n$，称之为一个正规子群序列（**normal subgroup sequence**）.

注意，正规子群列只要求相邻的两个群具有“正规子群”的关系，而 G_{i+2} 不一定是 G_i 的正规子群.

有时候，正规子群列的相邻两项中间可以再插入一个群，使之仍然成为一个正规子群. 这样的行为有时被称为“加细”. 如果一个正规子群列已经无法加细，那么我们称之为一个**合成序列（composition sequence）**.

类比到非负整数的质因数分解，合成序列就好比是每次只拿走一个质因子，序列的每一项都比上一项少一个质因子. 非负整数的“合成序列”不一定唯一，但是所有合成序列的长度一定是一样的. 群的合成序列也类似，合成序列不唯一，但其长度必然是一样的. 这就是以下 Jordan-Holder 定理在群论中的情形：

定义 3　Jordan-Holder 定理

若一个群 G 可以分解为合成列 $G = G_0 \triangleright G_1 \triangleright \cdots \triangleright G_r$ 和 $G = H_0 \triangleright H_1 \triangleright \cdots \triangleright H_s$，那么必有 $r = s$.

有限单群

有限群的研究重点是有限单群，知道了所有有限单群的性质也就能相应推出任意有限群的性质. 但我们是否知道所有有限单群了呢？答案是肯定的. 事实上，目前的研究表明，全体有限单群都可以归入特定的几个类别里，如果这是真的，那么我们确实已经知道了所有有限单群.

定义 4　有限单群分类定理[11]

如果 G 是一个单群，那么 G 必属于以下几类群中的一个：

- $\mathbb{Z}_p$，*其中 p 是素数；*
- *交错群 A_n，其中 $n \geq 5$；*
- *典型李群；*
- *例外、缠绕李群；*
- 26 *个散在单群.*

[11] 可参见 http://www.ams.org/notices/200407/fea-aschbacher.pdf.

有限单群的分类定理的证明历时长且极为复杂，汇聚了历代无数数学家的工作．这庞杂的证明在过去曾被指出有漏洞，尽管这些漏洞都已经被补上了，但过于复杂的证明和毫无美感的分类结果还是让人质疑其正确性．事实上，当代数学家仍然在尝试给出更简洁的证明，截至 2019 年已经有八册证明发表，预计完成后约 5000 页证明．

群同态

预备知识 正规子群[541]

同构

让我们来观察两个群 $(\mathbb{Z},+)$ 和 $(2\mathbb{Z},+)$．如果我们把 $2\mathbb{Z}$ 中的 2 都看成 1，4 都看成 2，以此类推，将 $2k$ 都看成 k，那么两个群的运算规则是一模一样的．比如说，$2\mathbb{Z}$ 中有 $2+4=6$，对应的是 $\mathbb{Z}$ 中 $1+2=3$ 的等式．

我们研究集合和群的时候，元素叫什么名字并不重要，重要的是元素之间是否相同以及运算规则是怎样的．那么，如果我们真的将 $2\mathbb{Z}$ 中的元素 $2k$ 都重命名为 k，它就和 $\mathbb{Z}$ 没什么区别了．所以在群的意义上，如果不考虑子群关系，单独把 $\mathbb{Z}$ 和 $2\mathbb{Z}$ 拿出来的时候，我们就认为它们是不可区分的，完全相同的两个群．

如果我们建立一个映射 $f:\mathbb{Z}\to 2\mathbb{Z}$，定义为 $f(k)=2k$，那么这个 f 就是一个双射，它在两个群的元素之间一一对应地建立了联系．这样，对于任意整数 m,n，有 $f(m)+f(n)=f(m+n)$，也就是说“先运算再映射”和“先映射再运算”结果是相同的．

类似地，对于任意的两个群 G 和 K，如果存在一个**双射** $f:G\to K$，使得对于任意的 $x,y\in G$ 都满足 $f(x)f(y)=f(xy)$，那么这两个群的运算结构就是一模一样的．这时我们说这两个群是**同构（isomorphic）**的，而这个使得它们同构的双射就被称为 G 和 K 之间的**同构映射（isomorphic mapping）**，也可以简称**同构（isomorphism）**这里加粗的两个“同构”，前者是形容词，后者是名词．

由于同构使得两个群各方面表现一模一样，研究同构其实没有太大意义，

我们甚至直接把同构的两个群看成同一个群，不管元素具体怎么命名的．有意思的结构，是以下定义的“同态映射”．

同态

同构映射是一个双射．如果把这个要求拿掉，我们就得到同态的概念：

定义 1　同态映射

对于两个群 G 和 K，如果映射（不一定是双射）$f : G \to K$ 使得 $\forall x, y \in G, f(x)f(y) = f(xy)$，那么称 G 和 K 是同态（**homomorphic**）的，称 f 是同态映射（**homomorphic mapping**）或同态（**homomorphism**）．

定义 2　像和核

沿用定义 1 的设定．K 中被映射到的元素构成的集合，称为 f 的像（**image**），记作 $f(G)$．G 中映射到 K 的单位元 e_K 的元素构成的集合，称为 f 的核（**kernal**），记为 $\ker(f)$．

注意，$f(G) \subset K$，$\ker(f) \subset G$．

同态的两个群，运算结构很相似但又不完全一样．在以上定义的例子中，K 的行为就像是一个弱化版的 G，可能会丢失一些细节，但保留的方面和 G 是一模一样的．这么说可能不够具体，我们用习题来理解同态的“似而不同”．

习题 1　群同态基本定理

设两个群 G 和 K，$f : G \to K$ 是一个同态．求证：

- $\ker(f)$ 是 G 的一个正规子群[12]．
- 对于 $x, y \in G$，如果 x_1 和 y_1 分别和 x, y 同余，或者换句话来说，$x_1^{-1}x \in H$ 和 $y_1^{-1}y \in H$，那么 $f(x) = f(x_1)$，$f(y) = f(y_1)$．
- 由前两条的结论，证明可以用 f 来导出一个映射 $f' : G/\ker(f) \to K$ 是一个同构．

由习题 1，同态的实质就是商群 $G/\ker(f)$ 和 K 之间的同构．$G/\ker(f)$ 继承了 G 的运算，但是由于把同余的元素全都当作同一个了，也就丢失了一部分细节．因此我们说同态的两个群也是“似而不同”的．

[12]这保证了群 $G/\ker(f)$ 存在．

群作用

预备知识　群同态[555]

群在自身上的作用

给定一个群 G，我们任意拿出一个元素 $a \in G$，用 a 去左乘 G 中的所有元素（包括 a 自己），那么我们可以把 a 看成一种 G 到自身上的映射：$f_a : G \to G$，使得对于任意 $x \in G$，$f_a(x) = ax$.

群 G 中的每一个元素都可以像这样生成一个映射，把这些映射全部放在一起，我们也可以整体上看成一个映射 $f : G \times G \to G$，满足：对于任意的 $a, x \in G$，有 $f(a, x) = f_a(x) = ax$. 这样的 $G \times G$ 到 G 上的映射，被称为一个**群作用（group action）**.

按照上述定义得到的作用，通常称为**左平移**作用；相应地，我们也可以让元素 a 对 x 的作用是 $g_a(x) = xa$；这样的作用被称为**右平移**.

群作用

更一般地，对于任何集合 X，群 G 中每个元素都可以代表 $X \to X$ 的一个映射. 我们当然可以任意规定这些映射，但如果这些映射满足一定条件的话，就会构造出一个很有意思的结构：

定义 1　群作用

设群 G 和集合 X，G 中每个元素都是 X 到自身的映射，记 $g \in G$ 将 $x \in X$ 映射为 $g \cdot x \in X$. 如果所有这些映射满足满足下面两条公理：

- 结合律：对于 $g_1, g_2 \in G, x \in X$，$(gh) \cdot x = g \cdot (h \cdot x)$.
- 单位元是恒等映射：G 的单位元 e 将任何 $x \in X$ 映射到自身：$e \cdot x = x$.

那么我们称群 G 作用（**acts**）于集合 X 上.

群作用可以像定义里一样记为 $g \cdot x$，也可以记为 X 到 X 的若干映射 $f_g(x) = g \cdot x$，还可以整体上看成 $G \times X$ 到 X 的一个映射 $f(g, x) = g \cdot x$.

群作用的例子

例 1　平移作用

本条开头定义的左平移和右平移，就是群在自身集合上的作用.

例 2　伴随作用

对于任意 $a \in G$，令 $f_a : G \to G$ 满足 $\forall x \in G, f_a(x) = axa^{-1}$，则这些映射定义了一个群 G 在集合 G 上的作用，称为伴随作用.

例 3　线性变换

参考线性变换[454]. 我们已经知道，n 阶非奇异矩阵配上乘法可以构成一个群；相应地，满秩线性变换（可逆线性变换）配上映射的复合运算构成一个群. 非奇异矩阵乘法是给定了基向量以后，满秩线性变换的复合的表示.

取 n 维实数向量空间 X，那么 X 是向量的集合；$GL(n,\mathbb{R})$ 是 X 上可逆线性变换的群，显然 $GL(n,\mathbb{R})$ 按照通常的线性变换定义，构成了在 X 上的一个作用.

群作用的性质

当我们讨论群 G 在集合 X 上的作用时，一共有两个集合要关心.

如果我们固定 X 中的一个元素 x，那么每个 G 中元素 g 都把 x 映射到某个 $f_g(x) \in X$ 上. 所有能这样被映射到的元素 $f_g(x)$ 构成了 X 的一个子集，称为元素 x 的**轨道（orbit）**. 我们也可以找出所有不移动 x 的群元素 g，即满足 $f_g(x) = x$ 的 $g \in G$，那么所有这样的 g 构成了 G 的一个子群 F_x，称为 x 的**迷向子群（isotropy subgroup）**或者**稳定子群（stablizer subgroup）**.

习题 1　迷向子群

证明 F_x 构成群（例 2 [565]）.

如果对于任何 $x \in X$，x 的轨道都是整个 X，那么我们称这个作用是**可递的**，此时 X 就是 G 的**齐性空间**. 如果对于任何 $x \in X$，x 的轨道只是 $\{x\}$，那么这个作用就是**平凡（trivial）**的.

如果对于任何 $x \in X$，任何 $g \in G - \{e\}$，都使得 $g \cdot x \neq x$，那么我们说这个作用是**有效的**. 有效性等价于说任何 $x \in X$ 的迷向子群都是 $\{e\}$.

在之前的例子中，平移作用既是可递的，又是有效的. 但是伴随作用不能

保证有效性和可递性，具体情况要看群的结构性质．全体可逆线性变换构成的群作用在非零向量空间上，这个作用是可递的，也是有效的——注意一定得是非零向量空间，把零向量排除在外．

定理 1

设群 G 作用在 X 上，在 X 上定义关系 $\sim$ 如下：$\forall x,y\in X, x\sim y \iff \exists g\in G, g\cdot x=y$，或者说，$x\sim y$ 当且仅当 y 在 x 的轨道里．那么，$\sim$ 是一个等价关系．

由群 G 的封闭性和逆元存在性分别可以证明定理 1 中关系 $\sim$ 的传递性和对称性．这个定理说明，轨道划分是一种等价类划分．

定义 2

设群 G 按照伴随作用，作用在自身上．

- 对于 $g\in G$，记 C_g 为 g 在伴随作用下的轨道，称 C_g 为 g 的共轭类（**conjugate class**），每一个 $h\in C_g$ 都称为 g 的共轭元素（**conjugate**）．
- 对于 $g\in G$，记 $C_G(g)$ 是 g 在伴随作用下的迷向子群，称 $C_G(g)$ 为 g 在 G 中的中心化子（**centralizer**）．
- 记 $C(G)=\bigcap_{g\in G} C_G(g)$，称为群 G 的中心（**center**）．

群的中心还有另一个定义：所有可以和 G 中一切元素交换的元素构成的集合，就是 $C(G)$．

定理 2　轨道-迷向子群定理

设群 G 作用在集合 X 上．固定一个 $x\in X$，那么对于 $g,h\in G$，我们有 $g\cdot x=h\cdot x \iff g^{-1}h\cdot x=x \iff g^{-1}h\in F_x \iff h\in gF_x$．也就是说，如果两个群元素 g,h 能够把 x 映射到同一个元素上，那么 g 和 h 在迷向子群 F_x 的同一个左陪集上．

推论 1

由定理 2，$|G|/|F_x|=|O_x|$，其中 O_x 是 x 的轨道．

习题 2　Burnside 引理

设群 G 作用在集合 X 上．对于给定的 $g\in G$，记 $X^g=\{x\in X|g\cdot x=x\}$，

O_x 是 $x \in X$ 的轨道，那么 $|\{O_x|x \in X\}| = \frac{1}{|G|}\sum_{g\in G}|X^g|$. 就是说，$X$ 上轨道的数目，等于每个 $g \in G$ 作用后不产生效果的元素数量之平均值. 证明此引理（见例 3 [565]）.

商空间

预备知识 二元关系[858]，子空间[465]

[13]设 W 是 n 维矢量空间 V 的一个 m 维子空间. 我们想通过 W 来定义 V 中元素的一个等价关系，并由此得到 V 的一个分类. 对于 $v_1, v_2 \in V$，如果 $v_1 - v_2 \in W$，则称它们是关于 $\mathrm{mod}W$ 相合的，记为 $v_1 = v_2$. 不难证明这是一个等价关系. 因此，它就确定了 V 的一个分类. 元素 v 关于 $\mathrm{mod}W$ 的相合类，即 V 中所有与 v 关于 $\mathrm{mod}W$ 的相合的元的全体，用符号 (v) 标记. 这样，有相合类为元素的商集合

$$V/W = \{(v)|v \in V\} \tag{1}$$

现在我们在商集合 V/W 中引入线性运算，使它也成为一个线性空间. 当然这种线性运算应与 V 中原有的线性运算要有联系. 为此，对于相合类的加法和数乘，我们自然采用下列定义

$$\begin{aligned}(v_1) + (v_2) &= (v_1 + v_2)\\ a(v) &= (av), a \in K\end{aligned} \tag{2}$$

这里的定义与相合类的代表的选取无关，故是有意义的，容易证明 V/W 在这些运算下构成体 K 上的矢量空间. V/W 称为 V 关于 $\mathrm{mod}W$ 的**商空间**，(0) 然是它的零元.

下面我们再来分析一下 V, W 和 V/W 三者基之间的关系. 设 $u_1, u_2, \cdots, u_m$ 是 W 的一个基. 我们再补充 $n-m$ 个元素 $u_{m+1}, \cdots u_n$ 使 $u_1, u_2, \cdots, u_n$ 成为 V 的一个基. 于是对 V 中任一元 $v = a^i u_i$ 所确定的相合类，有

$$(v) = \left(a^i u_i\right) = a^i\left(u_i\right) \tag{3}$$

（未完成）

[13]参考《物理学中的几何方法》

群的表示

预备知识 群作用[557]

表示

表示（representation）是数学中的重要部分．为了能够简洁明了地讨论数学对象的性质，我们有时候需要用一个易于理解的表达方法来描述这些对象．比如说，对于学龄前的小朋友，$1+1=2$ 的概念可能过于抽象，不利于理解；这个时候我们往往会使用“一个苹果加一个苹果等于两个苹果”来表示相同的概念，小朋友会更容易理解．

表示论常常是厚重的大部头作品．在本节中，我们仅初步介绍**群**的表示思想．

群的表示

最直观和易于讨论的群，莫过于置换群．事实上，当伽罗华（Galois）第一次提出群的概念的时候，并没有像我们现代理论那样高度抽象和严格；他主要都在讨论置换群的性质．遗憾的是，当年的数学家们都迷惑于伽罗华研究这种东西的意义何在．

我们使用置换群来尝试表示任意的群．

定义 1　群在置换群上的表示

设有群 G 和一个置换群 S_n．如果存在同态 $\phi: G \to S_n$，那么我们称 ϕ 是群 G 在 S_n 上的一个表示．

我们也可以使用线性空间来对群进行表示，利用线性空间的线性变换．

定义 2　群在线性空间上的表示

设有群 G 和一个线性空间 V，记 V 上的全体可逆线性变换为 $GL(V)$[14]．如果存在同态 $\phi: G \to GL(V)$，那么我们称 ϕ 是群 G 在 V 上的一个表示．

[14]注意 V 上全体可逆矩阵例 5 [534] 也是这样的表达方式．矩阵可以看作是线性变换的表示．

自由群

预备知识　群同态[555]

字母和词

定义 1　字母

没有赋予任何运算关系的元素，被称为一个字母（**letter**）.

给定没有赋予任何运算结构的任意集合 S，它的每个元素都可以作为一个字母.

例 1　字母的例子

- 随便取一个元素，取名为 x，那么 x 可以看成一个字母.
- 取 52 个英文字母构成的集合，则每个元素都可以当作字母.

字母就是纯粹的符号，除了“指代一个元素”以外没有更多的结构. 没有更多结构当然没有研究的意义不过，所以我们考虑拓展字母的用途，比如把若干字母按顺序排列起来，得到新的元素. 这种元素不同于字母，我们把它叫做“词”.

定义 2　词

有限个字按照一定顺序排列构成的元素，称作一个词（**word**）或字. 在计算机理论中，也称为字符串.

例 2　词的例子

- 单元素集合的元素 x，可以构成词 x，xx，$xxxxxx$ 等.
- 52 个英文字母构成集合 S，那么 S 中的元素 l, e, t, r 都是字母，它们可以进行有限排列，得到词 $letter$.

单个字母也可以看成是特殊的词，这样一来，就可以把字的排列拓展成词之间的一种**运算**．两个词首尾相连，可以构成一个更大的排列，仍然是有限个字的排列，因此得到的还是词．

定义 3　词的运算：首尾相连

给定集合 S．用 S 中的元素作为字母所排出的词，构成一个“词集合”，记为 $W(S)$．在集合 $W(S)$ 上可以定义一个运算“$\cdot$”如下：设 $a,b\in W(S)$，那么 $a\cdot b=c$，其中 c 是 a 和 b 首尾相连的结果．

例 3　词的运算

还是用英文字母构成的集合 S．lov 是一个词，ely 也是一个词，$lov\cdot ely=lovely$ 是这两个词进行首尾相连运算的结果．

我们也常常省略运算符号“$\cdot$”，这时 lov 和 ely 的连接运算就直接写成 $lovely$ 了，和运算的结果形式上一样，非常简练．

自由生成群

我们希望改进一下词集合和词运算，构造出一个群．为了做到这一点，我们首先需要扩展一下词集合．

定义 4　字母的逆

给定集合 S．给 S 中的每一个元素 x 都赋予一个“逆字母”，记为 x^{-1}．

字母的逆是用来满足群的“逆元存在性”的．如果一个词中出现某个字母和对应的逆字母相连接，那么它们就必须被“消除”．比如说，$lsdd^{-1}l^{-1}m^{-1}m$ 被认为和 lsl^{-1} 是相同的．特别地，对于任何字母 x,y，我们认为 xx^{-1}，$x^{-1}x$，yy^{-1}，$y^{-1}y$ 都是同一个词，称为**空词（empty word）**或**空字**．

有了这个规则，我们就可以构造出一个群了：

定理 1　自由生成群

给定集合 S，用它的元素作为字母．S 中的全体字母和它们的逆字母可构成词，并且构词过程中把相邻的互逆字母消除，这样得到的词所构成的集合记为 $F(S)$．在 $F(S)$ 中定义运算为词的首尾相连，那么 $F(S)$ 配合该运算构成一个群，称为由集合 S 生成的自由群（**free group**），或自由生成群．

自由群的单位元就是**空词**．一个词的字母都取逆以后反序排列，就得到了这个词的逆元．比如说，$le^{-1}tter$ 的逆元素就是 $r^{-1}e^{-1}t^{-1}t^{-1}r^{-1}el^{-1}$．

自由群是结构复杂度最高的群，这是因为所有的群都可以看成某个自由群的商群，而商群是把原来的群中一些细节特征忽略掉的结果．这个结论是由以下定理保证的：

定理 2　自由群的一般性质

给定任意的集合 S 和群 G．如果 $f:S\to G$ 是两个集合间的任意映射，那么总可以把 f 拓展成群同态 $\varphi:F(S)\to G$，并且这种拓展是唯一的．

证明见例 4 [565]．

考虑集合 G 到群 G 上的映射 $f:f(g)=g,\forall g\in G$，那么由 f 拓展而来的同态 φ 就是 $F(G)$ 到 G 的满射．结合**群同态基本定理**（习题 1 [555]）可推知，这意味着 G 是 $f(G)$ 的商群．

群的自由积

预备知识　自由群[562]

将自由群的概念推广，即可得到两个群之间的自由积的概念．

自由积的构造

给定两个群 G 和 H，取集合 $G\cup H$ 上的自由群 $F(G\cup H)$，则 $F(G\cup H)$ 的元素形如 $x_1x_2\cdots x_k$ 的有限长字符串，其中 k 是某个正整数，各 x_i 都是 $G\cup H$ 的元素．

在 $F(G\cup H)$ 上定义一个等价关系：如果字符串 $g_1g_2\cdots g_k$ 中各 $g_i\in G$，那么令 $g_1g_2\cdots g_k\sim g_1\cdot g_2\cdot\cdots\cdot g_k$，即把该字符串等同于各字母在群 G 中运算的结果；同样地把字母都是 H 中元素的字符串等同于这些元素在群 H 中的运算结果；把 G 和 H 的单位元等同于空词．比如说，在整数加法群 $\mathbb{Z}$ 中，把字符串 123 等同于数字 $1+2+3$ 所代表的字符串，即只有一个字母 6 的字符串．

这样一来，商群 $F(G\cup H)/\sim$ 中的字符串就形如 $g_1h_1g_2h_2\cdots g_kh_k$、$g_1h_1g_2h_2\cdots g_k$、$h_1g_1h_2g_2\cdots h_kg_k$ 或 $h_1g_1h_2g_2\cdots h_k$ 的字符串，或者简单来说，有限长的 g 和 h

的交替字符串，其中 $g, g_i \in G$，$h, h_i \in H$.

称商群 $F(G \cup H)/\sim$ 为群 G 和群 H 的**自由积（free product）**，记为 $G * H$.

共合积

定义 1　共合积

设有三个拓扑空间 F, G, H，且有同态 $\phi : F \to G$ 和 $\varphi : F \to H$，那么可以定义 G 和 H 关于 F 的共合积 $G *_F H$ 如下：

$G *_F H$ 是 $G * H$ 的商空间，$G * H/\sim$，其中等价关系为：$\forall f \in F, \phi(f) \sim \varphi(f)$.

简单来说，$G *_F H$ 的元素依然是 H 和 G 中元素交替排列的字符串，但是在给定同态 $\phi : F \to G$ 和 $\varphi : F \to H$ 时，把所有 $\phi(f) \sim \varphi(f)$ 都看成等价元素.

群论中的证明和习题解答

子群和正规子群

例 1　对换和逆序数

问题来源请见例 2 [547].

群作用

例 2　迷向子群

问题来源请见习题 1 [557].

已知群 G 作用在集合 X 上，且对于 $x \in X$ 有 G 的子集 $F_x = \{g \in G | g \cdot x = x\}$.

那么 $\forall g \in F_x$，$(g^{-1}g) \cdot x = e \cdot x = x$. 又因为 $g \cdot x = x$，所以 $g^{-1} \cdot x = g^{-1} \cdot (g \cdot x) = (g^{-1}g) \cdot x = x$.

因此，$\forall g_1, g_2 \in F_x$，有 $(g_1^{-1}g_2) \cdot x = g_1^{-1} \cdot x = x$，也就是说，$(g_1^{-1}g_2) \in F_x$. 根据判别式定理 1 [541]，$F_x$ 构成 G 的子群.

例 3　Burnside 引理

首先转写一个表达：$\sum_{g \in G} |X^g|$ ＝“满足 $g \cdot x = x$ 的 (g, x) 数量”＝

$\sum_{x\in X}|F_x|$.

由推论 1 [557]，$\sum_{x\in X}|F_x|=\sum_{x\in X}|G|/|O_x|=|G|\sum_{x\in X}1/|O_x|$

现在关键是 $\sum_{x\in X}1/|O_x|$ 是多少．显然，$\sum_{x\in O_x}1/|O_x|=1$，也就是说 $1/|O_x|$ 在每一个 O_x 上求和的结果是 1，那么它在整个 X 上求和的结果，刚好就是所有轨道的数量 $|\{O_x|x\in X\}|$.

把以上结果整合，我们有：$\sum_{g\in G}|X^g|=|G|\cdot|\{O_x|x\in X\}|$.

例 4　自由群的一般性质

问题来源请见定理 2 [562]

设集合 S 到群 G 上有一个集合间的映射 f. 我们来构造一个 $\varphi:F(S)\to G$ 的同态.

首先，由于 φ 是 f 的扩张，因此要定义 $\forall s\in S,\varphi(s)=f(s)$.

其次，由于 φ 应为一个群同态，因此要定义 $\forall s_i\in S,\varphi(s_1s_2s_3\cdots)=\varphi(s_1)\varphi(s_2)\varphi(s_3)\cdots$. 同时，还需要有 $\varphi(s^{-1})=\varphi(s)^{-1}$.

这样一来，φ 的映射规则就确定下来了，并且容易验证它是一个群同态. 因此这是唯一一个符合同态条件的扩张.

李群

预备知识　拓扑群[651]，流形[688]

李群的概念

当拓扑群的拓扑部分构成光滑流形时，所得到的拓扑群被称为李群.

定义 1　实李群

给定一个实光滑流形 G，若在 G 上定义了一个群运算“$\cdot$”，且满足流形空间之间的映射 $f:G\times G\to G$ 是一个光滑映射，其中 $f(g_1,g_2)=g_1\cdot g_2^{-1}$，那么称 G 是一个李群（**Lie group**）.

例 1　实李群的例子

1. 拓扑群[651] 词条中例 1 [651] 所举的两个例子，实数轴 $\mathbb{R}$ 和单位圆 S^1 都是实流行，都构成实李群.

2. 拓扑群[651] 词条中例 2 [651] 所举的例子，一般线性群，是一个实李群.

3. 一般线性群 $\mathrm{GL}(k,\mathbb{R})$ 的子群，正交群 $\mathrm{O}(k,\mathbb{R})$，以及特殊正交群 $\mathrm{O}(k,\mathbb{R})$[15]都是实李群.

为了加深对李群的印象，我们再举出一个重要的反例：

定义 2　环面[16]

记 $\mathbb{T}=S^1\times S^1$ 为李群 S^1 的积李群，即其群乘法为 S^1 的群乘法的积，流形为 S^1 的流形的积. 取 $\mathbb{T}^2$ 的子群 $H=\{(\mathrm{e}^{\theta\mathrm{i}},\mathrm{e}^{a_0\theta\mathrm{i}})|\theta\in\mathbb{R}\}$，其中 a_0 是一个无理数.

按以上方式定义的群 H 是一个拓扑群，但不是流形，因而不是李群. 这是因为无理数 a_0 导致 H 的图像是环面上密绕的一条线，在任何一个点的任何一个开邻域里都有无数条不连通的线段，因此任何点处都无法找到同构于欧几里得空间的图.

不过我们依然可以在群 H 上构造出李群来. 把 H 的拓扑进行如下改动：定义映射 $\phi:\mathbb{R}\to H$，其中 $\phi(\theta)=(\mathrm{e}^{\theta\mathrm{i}},\mathrm{e}^{a_0\theta\mathrm{i}})$；再利用 ϕ 来构造 H 的拓扑为“$U\subseteq H$ 为开集当且仅当 $\phi^{-1}(U)$ 为 $\mathbb{R}$ 的开集”. 这样定义的拓扑空间 H 就是一个同胚于 $\mathbb{R}$ 的实李群了.

和李代数的联系

李代数虽然是对李群性质的抽象，但是其本身可以脱离李群来定义，参见**李代数**[611] 词条. 事实上，学有余力的学生完全可以在熟悉了线性代数的概念之后直接进入李代数的学习，无需流形的知识[17].

李代数是李群在单位元 e 处的切空间.

---

[15] 正交群指的是全体保度量线性变换构成的群，等价于全体行列式的绝对值是 1 的矩阵构成的乘法群. 特殊正交群是正交群的子群，是由其中行列式为 +1 的矩阵构成的.

[16] 参考 Wikipedia 相关页面.

[17] 我国李群和李代数专家朱富海教授一直致力于向低年级本科生科普李代数知识，他在 bilibili.com 和知乎都有发布“给大一学生的 Lie 代数”系列，可参见这里.（2021 年 1 月更新）.

第二章

环论

环

预备知识　群[534]

环的定义

群是代数学中研究的最简单的结构，只由一个集合配上一个运算构成，这个运算只有 4 条公理进行限制.

通常，在群之后会介绍的一个更复杂一些的概念，是**环**. 一个环就是由一个集合配上两个运算构成的集合，通常把这两个运算叫做加法和乘法；环上的加法和乘法分别构成群，不过乘法群不包括加法的单位元，而加法群是阿贝尔群.

为了简洁地定义环，先定义两个群论中并没有涉及到的概念.

定义 1　半群和幺半群

给定集合 G 及其上的一个运算，运算符号忽略. 如果运算满足：

- 封闭性
- 结合性

那么称 G 配合该运算构成一个半群（**semi-group**）. 如果半群中含单位元，则构成一个幺半群（**monoid**）. 这里的“幺”是“一”的意思.

由此可见，群就是每个元素都有对应可逆元的幺半群. 有了幺半群的概念，就可以很方便地定义环了.

定义 2　环

一个环（**ring**）是一个集合 R 与两种运算“加”和“乘”，分别记为 $+$ 和 $\cdot$. 其中加法配合 R 中所有元素构成一个阿贝尔群，加法群的单位元通常称为零元，记为 0；乘法配合 R^* 中所有元素构成一个半群，其中 $R^* = R - \{0\}$. 除了加法和乘法分别构成两个群之外，还要求加法和乘法满足分配律

- $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$

- $(b+c)\cdot a=(b\cdot a)+(c\cdot a)$

如果乘法还含有单位元，则称其为幺元，记为 1.

通常，为了方便表示，我们也会省略环中乘法的符号，将 $a\cdot b$ 写为 ab.

由定义，环的加法必须是可交换的，但乘法却不一定．如果 R 的乘法也交换的话，我们就称 R 为一个**交换环（commutative ring）**.

例 1　整数环

整数集合配上通常的加法和减法，构成一个交换环，记为 $\mathbb{Z}$.

例 2　数域

有理数集合配上通常的加法和减法，构成一个交换环，记为 $\mathbb{Q}$．类似地，实数构成交换环 $\mathbb{R}$，复数构成交换环 $\mathbb{C}$．这三个环都是数域的例子，数域是指包含整数的域，而域[588] 是之后基于环而讨论的概念.

例 3　多项式环

设有交换环 R，x 是一个自变量，那么集合 $\{\sum_{i=0}^{n}a_ix^i|n\in\mathbb{Z}^+,a_i\in R\}$ 可以看成是 x 的多项式构成的集合，其系数取遍 R．这些多项式之间的加法和乘法由 R 的加法和乘法诱导（定义），并且构成了一个交换环，称为 R 的多项式环.

记 R 上的多项式环为 $R[x]$.

环的定义在一个细节上有争议，那就是乘法需不需要有幺元．有些书中的定义不要求有幺元，也就是说乘法只构成半群即可，这就使得对于任意正整数 n，n 的全体倍数构成的集合 $n\mathbb{Z}$ 也是一个环；在这种定义里，会把含幺元的环称作**含幺环**或者**幺环**．由于不含幺元的结构一般不研究，所以主流数学界干脆将环定义为有幺元的．本书为了方便表述，认为环都不一定由幺元，将幺环和环区分开，请读者注意.

子环

定义 3　子环

给定一个环 R，如果 S 是 R 的子集，并且在继承 R 的两个运算后也构成环，那么称 S 是 R 的子环（**subring**）.

环的理想和商环

预备知识　环[569]

概念的来源

环 R 的**理想** I，是一种“**正规子环**[1]”，即它使得商集 R/I 能自然成环，就像正规子群的作用一样．之所以不叫正规子环，是因为理想最初来自代数数论，库默尔（Ernst Eduard Kummer）定义了一个他称为“理想数”的概念，证明了费马大定理在 $n<100$ 时大多数情况成立．后来，戴德金（Julius Wilhelm Richard Dedekind）发现，库默尔定义的理想数，正是能诱导出商环的“正规子环”，于是直接借用了“理想数”的名字，将其命名为“理想”．

给定环 R 和它的子环 I，那么 I 要满足什么条件才能使得 R/I 成环呢？

显然，I 关于环的加法，得构成一个正规子群，而这是天然满足的，因为环的加法群是阿贝尔群，而阿贝尔群的一切子群都是正规子群．

R/I 中的每个元素，被定义为 I 作为子群的陪集．元素 $a\in R$ 所在的陪集就是 $a+I$．陪集之间显然可以进行加法运算：

$$(a+I)+(b+I)=a+b+I+I=(a+b)+I \tag{1}$$

这满足诱导运算的要求：a 的陪集加 b 的陪集，等于 a 加 b 的陪集．

为了让 R/I 诱导一个环乘法，我们还需要：a 的陪集乘 b 的陪集，等于 a 乘 b 的陪集．也就是说，

$$(a+I)(b+I)=ab+aI+Ib+I=ab+I \tag{2}$$

这就意味着 $aI+Ib\in I$，对于任意的 $a,b\in R$ 都成立．那就是说，对于任意的 $a,b\in R$，都有 $aI\in I$ 和 $Ib\in I$ 才行（注意 a,b 中有一个是幺元的情况）．

如果 I 是 R 的子环，并且对于任意的 $a\in R$，都有 $aI,Ia\in I$，那么我们就可以利用 R 的环运算，诱导出 R/I 上的环运算．这样的 I 就是我们要的**理想**．

[1]要强调的是，理想不一定有乘法单位元，而本书默认环都有乘法单位元，因此不能说理想是子环．

理想的定义

定义 1　左理想

给定环 R，则 R 的一个子集 I 是 R 的一个左理想（**left ideal**），当且仅当 I 是 R 的加法子群，且对乘法满足左吸收律：$\forall r \in R$，有 $rI \in I$.

类似地，可以定义环的**右理想**，即满足右吸收律：$\forall r \in R$，有 $Ir \in I$.

如果一个子环既是左理想也是右理想，那么称它为一个**双边理想（2-sided ideal）**，简称**理想（ideal）**. 另外，如果一个理想不是环本身，则称其为一个**真理想**.

吸收律的一个直接推论是，如果一个理想包含乘法单位元，那么这个理想就是环本身.

习题 1

证明：给定环 R，如果 $I, J \subset R$ 都是它的理想，那么 $IJ = \{ij | i \in I, j \in J\}$ 等于 $I \cap J$.

这个习题的结论在将来类比素理想和素元素能清晰展示这两个概念之间的相似之处.

定理 1

给定环 R，如果 $I, J \subset R$ 都是它的理想，那么 $I \cap J$ 是它的理想.

定义 2　商环

给定环 R 和它的一个理想 I，R/I 自然构成一个环，称为 R 模去 I 的商环（**quotient ring**）.

例 1　整数环的理想和商环

对于正整数 n，记 $n\mathbb{Z}$ 为集合 $\{nm | m \in \mathbb{Z}\}$，即全体 n 的倍数构成的集合. $n\mathbb{Z}$ 是 $\mathbb{Z}$ 的理想. 记商环 $\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}_n$.

例 1 中定义的 $\mathbb{Z}_n$ 相当于**整数**[95] 词条所描述的“具有 n 个钟点的钟表”.

理想和商环的概念能帮助理解多项式环的结构. 在讨论多项式环以前，我们要先定义一个概念：

定义 3　极小多项式

给定环 R 和某个元素 a（不一定属于 R），那么 $f(a)=\sum_{i=0}^{n} r_i a^i$ 是 a 的一个多项式．有的 f 使得 $f(a)=0$，这样的多项式就叫做 a 的零化多项式（**null polynomial**）．非零的零化多项式中次数最低的，被称为 a 的极小多项式（**minimal polynomial**）．由于多项式环是一个欧几里得环[2]，因此还可得知，一切零化多项式都能被极小多项式整除[3]．如果元素 a 有零化多项式，那么称它是 R 的代数元；反之，称它是 R 的超越元．

显然，超越元的不同多项式的值彼此不相等，也就是说，超越元的多项式可以和多项式本身进行一一对应，或者说超越元的多项式和多项式环是同构的．但是代数元的不同多项式有可能值相同．我们依然用 $R[a]$ 来表示 a 的所有多项式的值的集合，其中多项式的系数取遍 R．

换言之，当 a 是超越元时，$R[a]\cong R[x]$．

例 2　多项式环

定义多项式环 $R[x]$ 的时候，我们说 x 是一个自变量，言下之意就是，$R[x]$ 中的多项式 $f(x)$ 都表示函数；两个多项式 $f(x)$ 和 $g(x)$ 只有每个系数都相等才被认为相等，如果任何系数不相等，那么认为 $f(x)\neq g(x)$．

如果取某个元素 a（不一定属于 R），那么 $f(a)$ 就不是一个函数，而是一个固定的元素了（不一定属于 R）．比如说，取 R 是整数环，那么 $f(x)=2x+x^2$ 就是一个函数，图像表示为一条抛物线；但 $f(\sqrt{2})=2\sqrt{2}+2$ 就是一个元素，尽管它不在整数环中．两个多项式 f 和 g 即使不同，也可能对于同一个元素 a 满足 $f(a)=g(a)$．

给定环 R 的代数元 a，那么 a 的全体零化多项式构成了 $R[x]$ 的一个理想 R_0．

习题 2　多项式环的性质

证明 a 的零化多项式集合 R_0 构成 $R[x]$ 的理想，并验证：$R[a]\cong R[x]/R_0$．

环同态

[2]将在以后讨论．

[3]如果多项式 g 整除 f，意味着存在一个多项式 h，使得 $f=gh$

预备知识 环的理想和商环[571]，

环同态的概念和群同态是类似的，只不过我们需要兼顾两个运算的性质.

定义 1　环同态

给定环 R_1 和 R_2，如果存在映射 $f: R_1 \to R_2$，使得对于任意的 $r, s \in R_1$ 都有 $f(r)+f(s)=f(r+s)$ 和 $f(rs)=f(r)f(s)$，那么我们称 f 是一个从 R_1 到 R_2 的环同态映射（**ring homomorphism**），而环 R_1 和 R_2 是同态（**homomorphic**）的环.

定义 2　环同构

如果一个环同态是双射，那么我们称它为一个环同构（**ring isomorphism**），相应的两个环就是同构（**isomorphic**）的. 记“R_1 和 R_2 同构”为 $R_1 \cong R_2$.

同构的两个环具有完全相同的运算结构，而同态的两个环的运算结构似而不同，和群的同态、同构类似.

环同态基本定理

预备知识 群同态基本定理：习题 1 [555]

定义 3　核

给定环同态 $f: R_1 \to R_2$，记 $\ker f = \{r \in R_1 | f(r)=0\}$，称为同态 f 的核.

和群同态基本定理类似，我们有如下环同态基本定理：

定义 4　环同态基本定理

给定环同态 $f: R_1 \to R_2$，则 $\ker f$ 是 R_1 的理想，并且 $R_1/\ker f \cong R_2$.

整环

预备知识 环同态[573]

我们常见的整数环属于一类性质极为良好的环，我们称这个分类为“无零因子交换幺环”，或者“整环”. 注意区分“整数环”和“整环”这两个相似的术语，前者特指整数构成的环，后者则指一类环.

定义 1　零因子

在环 R 中，如果有两个非零的元素 a,b 使得 $ab=0$，那么我们称 a 是一个左零因子，而 b 是一个右零因子．如果某元素即使左零因子又是右零因子，那么我们称它是一个零因子（**zero divisor**）．

简单来说，零因子就是“相乘得到零的非零元素”，其名称的含义就是“零的因子”．整数环里不存在零因子，但是这个概念也不难理解：考虑环 $\mathbb{Z}_{12}$，在这个环里，$3\neq 0$，$4\neq 0$，但是 $3\times 4=12=0$．

定义 2　整环

对于环 R，如果它的乘法交换并且没有零因子，那么我们称这个环是一个整环（**domain**）．

整环的概念可以记为“无零因子交换幺环”，用以指代它的三个关键特点：“无零因子”、“交换”和“有乘法单位元”．最后一个特点在本书的语境下显得冗余，因为我们限定环都是含有乘法单位元的；强调幺环只是为了避免使用其它术语体系时可能的混淆．

例 1　整环的例子

- 整数环
- 多项式环
- 高斯整数环 $\mathbb{Z}[\mathrm{i}]$

常用概念

接下来是一系列非常有用的概念，也是后续的进阶词条的基础．

定义 3　因子和整除性

在整环 R 中，如果对于 $a,b\in R$，存在 $r\in R$ 使得 $b=ar$，那么我们称 a 整除 b，记为 $a|b$，同时称 a 是 b 的一个因子（**factor**）或除数（**divisor**）．

定义 4　单位

对于整环 R，如果 $u\in R$ 有乘法逆元 $u^{-1}\in R$，那么称 u 是 R 的一个单位（**unit**）．

定义 5　真因子

*在整环 R 中，如果对于 $a,b\in R$，存在 $r\in R$ 使得 $b=ar$，并且 r 不是一个单位，那么称 a 是 b 的一个真因子（**proper factor**）.*

真因子的特点是单向性，如果 a 是 b 的一个真因子，那么绝不可能出现 $b|a$，因为真因子的定义中要求 $b=ar$ 中的 r 是乘法不可逆的．从因子分解的角度，我们可以定义如下等价关系：如果两个元素 $a,b\in R$ 之间只差一个单位因子，即存在单位 u 使得 $a=bu$，那么我们可以把 a 和 b 等价起来．检查一下，如此定义的关系满足等价关系的三个公理，因此是一个等价关系．作为一个例子，整数环中的单位只有两个，1 和 -1，因此这个等价关系应用到整数环中就是把所有 n 和其对应的 $-n$ 等价起来．在讨论因子分解时，我们常常使用这个等价划分．

接下来介绍的两个概念，素元素和不可约元素，都是整数中素数概念的推广，只不过选用了素数的不同特点．

定义 6　素元素

*对于整环中的元素 $p\in R$，如果它满足“如果任何 $a,b\in R$ 使得 $p|ab$，则必有 $p|a$ 或者 $p|a$”，则称它为 R 的一个素元素（**prime element**）.*

定义 7　不可约元素

*对于整环中不是幺元的元素 $p\in R$，如果它在 R 中没有真因子，那么称它为一个不可约元素（**irreducible element**）.*

对于一般的整环，素元素必然是不可约元素：

定理 1　素元素必是不可约元素

给定整环 R 和其素元素 p，则 p 必然是不可约元素.

证明：

反设 p 可约，则存在 p 的两个真因子 a,b 使得 $p=ab$，从而 $p|ab$．由素元素的定义必须有 $p|a$ 或 $p|b$，而这和“真因子”的单向性矛盾．故设定不成立，即 p 不可约.

证毕.

素理想与极大理想

预备知识 整环[574]

本节介绍两类极为重要的理想：素理想和极大理想.

定义 1　素理想

给定交换环 R，如果 P 是它的一个真理想，并且满足“对于任意 $a,b\in R$，如果 $ab\in P$，则必有 $a\in P$ 或 $b\in P$”，则称 P 是 R 的一个素理想（**prime ideal**）.

注意，素理想的定义不一定非得是整环，也可以是含有零因子的交换环. 比较一下它和素元素的定义，形式上是完全一样的.

定义 2　极大理想

给定交换环 R，如果 M 是它的一个真理想，并且满足“对于任意 R 的理想 I，如果 $M\subsetneq I$，那么 $I=R$”，则称 M 是 R 的一个极大理想（**maximal ideal**）.

极大理想的定义很直观，任何比极大理想大的理想都只能是 R 本身.

素理想的一大作用，是把零因子收集了起来，使得商环能成为一个整环. 同时，素理想和素元素之间高度类似. 我们可以把这些话写成如下紧凑的定理：

定理 1

给定交换环 R 和它的一个理想 P，则以下条件等价：

1. P 是 R 的素理想.
2. 如果 R 的两个子集 S_1、S_2 满足“若 $S_1S_2\in P$，则 $S_1\in P$ 或 $S_2\in P$”.
3. R/P 是一个整环.

证明：

（1 到 2）：根据素理想的定义和 S_1S_2 的定义，对于任意的 $s_1\in S_1$ 和 $s_2\in S_2$，那么 s_1 和 s_2 中必有一个是 P 的元素. 如果 $S_1\not\subset P$，那么必存在某一个 $s_1\in S_1$

使得 $s_1 \notin P$. 这样，任意和这个 s_1 配对的 $s_2 \in S_2$ 都必须是 P 的元素了，从而 $S_2 \subseteq P$.

（2 到 3）：任取 P 的左陪集 S_1 和 S_2，如果 $S_1S_2 = P$，那么必有 S_1 和 S_2 的其中之一就是 P. 换句话说，不存在非零（非 P）的元素相乘等于 0（P）. 从而 R/P 无零因子，加上继承自 R 的交换性，它就是一个整环.

（3 到 2）：而左陪集的运算可以看成任意选取代表元素进行运算后取结果所在的左陪集. 对于任意的 $a, b \in R$，如果 $ab \in P$，那么由于 R/P 是整环，必有 $a \in P$ 或者 $b \in P$. 这就是素理想的定义.

证毕.

整环的定义已经非常良好了，再进一步，添上“乘法单位元存在性”就能得到域. 事实上，这正是极大理想的作用：

定理 2

给定交换环 R 和它的一个理想 M，则以下两个条件等价：

1. M 是 R 的极大理想.

2. R/M 是一个域.

证明：

（1 到 2）：我们只需要证明，如果 $\{0\}$ 是 R 的极大理想，那么 R 是域. 这是因为如果 M 是 R 的极大理想，那么 $\{M\}$ 就是 R/M 的极大理想.

任取 $r \in R - \{0\}$，令 $I_r = \{sr | s \in R\}$，则显然 I_r 是一个理想. 由于 $I_r \supsetneq \{0\}$ 以及 $\{0\}$ 是极大理想，可知 $I_r = R$. 也就是说，存在 s 使得 $sr = 1$.

（2 到 1）：同样地，我们只需要证明域 R 的理想只有 R 和 $\{0\}$ 两种.

设 I 是一个 R 的理想，并且 $I \neq \{0\}$. 由于域的逆元存在性以及 I 的吸收律，必有 $1 \in I$，因此 I 必须是 R 本身.

证毕.

例 1　整数环中的理想

整数环 $\mathbb{Z}$ 中的理想都是形如 $n\mathbb{Z} = \{nk | k \in \mathbb{Z}\}$ 的集合. 其中当且仅当 p 是素数时，$p\mathbb{Z}$ 是一个素理想，同时也是极大理想.

由定理 2 可知，对于素数 p，必有 $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$ 是一个域，而且还是只含有限个元素的域，故被称为有限域.

要得出 $\mathbb{Z}_p$，还可以应用另一个定理：

定理 3

有限整环必是域.

证明留给读者．思路提示：由于整环没有零因子，故乘法满足消去律，利用这个消去律和有限性即可证.

爱森斯坦判别式

预备知识 素理想与极大理想[577]

爱森斯坦判别式可以用来判断多项式是否可约．满足爱森斯坦判别式条件的多项式必然是不可约的，即无法表示为两个多项式的乘积.

最简单的多项式是整系数多项式，也被称为**整数环上的**多项式，因此我们先介绍此类多项式上的爱森斯坦判别式．这只是更一般的爱森斯坦判别式的特例.

定理 1　整数环上的爱森斯坦判别式

设有整系数多项式 $f(x)=f_0+f_1x+f_2x^2+\cdots+f_nx^n$，若存在素数 p 使得：

1. $p\nmid f_n$；
2. 对于 $i\in\{0,1,2,\cdots,n-1\}$，都有 $p|f_i$；
3. $p^2|f_0$.

那么 $f(x)$ 就是不可约多项式.

证明：

反设 $f(x)$ 可约，则可以写成 $f(x)=(h_0+h_1x+h_2x^2+\cdots+h_rx^r)(g_0+g_1x+g_2x^2+\cdots+g_sx^s)$.

由于 $f(x)=f_0+f_1x+f_2x^2+\cdots+f_nx^n$，故[4]

$$f_i=\sum_{j+k=i}h_jg_k \tag{1}$$

利用条件 3：注意到 $f_0=h_0g_0$，而 $p^2\nmid f_0=h_0g_0$，故 h_0 和 g_0 中有一个是不能被 p 整除的．不妨设 $p\nmid h_0$．

利用条件 2：由于 $p\mid f_0=h_0g_0$，故 $p\mid g_0$．又由于 $p\mid f_1=h_0g_1+h_1g_0$ 且 $p\mid g_0$，故 $p\mid h_0g_1$；再由于 $p\nmid h_0$，故 $p\mid g_1$．

归纳：类似地，$p\mid f_2=h_0g_2+h_1g_1+h_2g_0$，而 $p\mid g_1,p\mid g_0$ 和 $p\nmid h_0$ 又能推出 $p\mid g_2$．以此类推，可以得到 $p\mid g_s$．

利用条件 1：由于 $p\nmid f_n=h_rg_s$，故应有 $p\nmid g_s$，与上一步矛盾！

由此可知反设不成立．

证毕．

虽然判别式的表达看起来很复杂，但是从证明过程可以看出来，三个条件的设置都自有其考虑．

爱森斯坦判别式不止是可以用于整系数多项式，也可用于任意整环上的多项式．整系数多项式的爱森斯坦判别式只需要整除知识即可理解，而任意整环上的形式则需要素理想的概念．

定理 2　整环上的爱森斯坦判别式

设整环 R 上有多项式 $f(x)=f_0+f_1x+f_2x^2+\cdots+f_nx^n$，若存在 R 的素理想 P，使得：

1. $f_n\notin P$；
2. *对于* $i\in\{0,1,2,\cdots,n-1\}$，*都有* $f_i\in P$；
3. $f_0\notin P^2$，*其中* $P^2=\{p_1p_2|p_i\in P\}$．

那么 $f(x)$ 就是不可约多项式．

其证明可以通过改写定理 1 的过程得到，把每一个形如 $p|a$ 的式子都写为 $a\in P$ 即可，其中 a 只是个符号，可以表示任意 f_i,g_i,h_i 等．读者可自行对比尝试证明．注意素数和整数环 $\mathbb{Z}$ 的素理想之间的关系．

[4]这种运算也被称为卷积．

真因子树

预备知识　整环[574]

真因子树的概念，是笔者优化了“因子链”的概念而得出的一套描述因式分解理论的框架.

概念的描述

定义 1　真因子

给定整环 R，对于 $r,s\in R$，如果 $s|r$ 且 $r\nmid s$，那么称 r 是 s 的真因子.

定义 2　单位

给定整环 R，对于 $u\in R$，如果 u^{-1} 是存在的，那么称 u 是 R 的一个单位（**unit**）. R 中全体单位的集合，记为 U.

显然，如果有单位 u 使得 $r=us$，那么 r 和 s 互相不是真因子. 我们将这样的 r,s 视为等价的：

定义 3

给定整环 R，定义集合 R 上的一个等价关系：对于 $r,s\in R$，r 等价于 s 当且仅当存在单位 u 使得 $r=su$. 等价的元素视为同一个元素，或者说把每个等价类看成一个元素，得到的集合是 R 模去该等价关系的商集[5]，记为 R_u.

举例来说，在**整数环** $\mathbb{Z}$ 上，对于任意正整数 n，我们把它等价于 $-n$，于是 $\mathbb{Z}_u$ 也可以看成是非负整数的集合.

有了 R_u 的概念，就可以定义本节核心的概念了：真因子树.

定义 4　真因子树

给定整环 R，对于 $r\in R$，如果存在非单位的 $a,b\in R$ 使得 $ab=r$，那么可以从 r 画两个箭头分别指向 a 和 b，而 (a,b) 就是 r 的一个因子分解；同样，如果 a 和 b 可以继续分解为其它非单位元素之积，那么也可以继续画出箭头指

[5] 见二元关系[858].

向它们对应的因子分解. 如是反复，直到不能继续进行下去为止，所获得的整个结构称为 r 的一棵真因子树.

每个从 r 出发一路指向末端的路径，称为 r 的一个枝条，枝条中涉及到的元素数量，称为枝条的长度. 一棵真因子树中最长的纸条的元素数量，称为这棵树的长度.

r 的真因子树一般不止一棵.

用真因子树进行描述

以下讨论限制在整环 R 的集合上. 利用真因子树的语言来直接翻译各种概念的方式如下：

定义 5

- 不可约元素：在某一棵树中为末端.
- 素元素：p 是素元素，当且仅当对于任意 r，如果 p 在 r 的一棵树上，那么也必在 r 的每一棵树上[6].
- 有限析因性：对于任意 r，存在一个正整数 N_r，使得 r 的任意枝条长度不超过 N_r.
- 唯一析因性：有限析因，且对于任意 r，其任何两棵树的末端元素构成的集合都是一样的.

由以上改写不难看出，不可约元素在任何树中都是末端. 另外，我们还有以下性质：

定理 1　素元素

整环 R 中素元素 p 的定义是：对于任意 $r \in R$，如果 p 在 r 的某个枝条上，那么 p 必然在 r 的每一棵树中.

应用素元素的定义，不难证明该定理.

[6]这是因为按照素元素的定义，如果 p 在 r 的某个枝条上，而 $r = ab$，那么 p 必在 a 或 b 的某个枝条上.

定理 2　素元素必是不可约元素

整环 R 中的素元素都是不可约元素.

证明：

反设素元素 $p \in R$ 不是末端，那么就可以从 p 开始画出两根枝条，分别以 a 和 b 开始；而由于素元素的定义，它又必须在这两根枝条中某一根的后面，从而在自己的后面，而这是不可能的. 因此反设不成立，素元素必是末端.

证毕.

今后我们也会引用真因子树的概念来方便阐释因式分解相关的问题.

唯一析因环

预备知识　真因子树[581]

唯一析因环，又称唯一分解整环，顾名思义，就是每个元素都有唯一的不可约因式分解. 这个定义用真因子树的语言来描述颇为方便.

定义 1　唯一析因环

对于整环 R，如果它具有有限析因性和唯一析因性[7]，那么称其为一个唯一析因环（**unique factorization domain**），常简称 UFD.

唯一析因环的好处显而易见，每个元素都可以唯一对应一种素元素分解，比如每个非平凡整数都可以表示为素数的乘积，且这种乘积是唯一的.

例 1　正面例子

- 整数环 $\mathbb{Z}$ 是唯一析因环.
- 域 $\mathbb{F}$ 上的多项式环 $\mathbb{F}[x]$ 是唯一析因环.

唯一析因环的概念脱胎于一个经典的错误. 柯西等人曾以为自己证明了**费马大定理**，而事实上他们的证明依赖了一个直觉上成立的假设，用现代语言来说就是“所有的环都是唯一析因环”. 然而很可惜，存在不唯一析因的环，这也成了此类证明中的命门.

[7] 见**真因子树**词条的定义 5 [581].

例 2　反面例子

我们给出一些不是唯一析因环的例子.

- 对于素数 $p>1$，环 $\mathbb{Z}[\sqrt{-p}]=\{a+b\mathrm{i}\sqrt{p}|a,b\in\mathbb{Z}\}$ 存在有多种因式分解方式的元素. 比如说，在环 $\mathbb{Z}[2]$ 上，6 可以分解为 $2\times(1+\mathrm{i}\sqrt{2})\times(1-\mathrm{i}\sqrt{2})$，也可以分解为 $(2+\mathrm{i}\sqrt{2})\times(2-\mathrm{i}\sqrt{2})$.
- 复数域上的全体整函数构成的环不是 UFD，因为存在可以无限析因的元素[8]：$\sin\pi z=\pi z\prod_{i=1}^{\infty}(1-\frac{z^2}{i^2})$.

未完成：关于整函数的概念，缺少“全纯函数”词条作为预备知识.

主理想整环

预备知识 唯一析因环[583]

在寻找唯一析因环的例子时，一个非常容易识别的类别就是本节介绍的的主理想整环.

定义 1　主理想

对于整环 R，任取 $a\in R$，记 $\langle a\rangle=\{ra|r\in R\}$ 为 a 生成的理想.

单个元素生成的理想，称为一个主理想（**principal ideal**）.

定义 2　主理想环

对于环 R，如果它的理想全都是主理想，那么称它为一个主理想环（**principal ideal ring**），常简称 PIR.

对于整环 R，如果它是 PIR，那么称它为一个主理想整环（**principal ideal domain**），常简称 PID.

欧几里得环

[8]本例取自维基百科相关页面.

预备知识 主理想整环[584]

定义 1　欧几里得环

给定整环 R，如果存在映射 $\delta: R \to \mathbb{Z}^+ \cup \{0\}$[9]，使得对于任意 $a, b \in R$ 且 $b \neq 0$，都存在 $q, r \in R$ 使得 $a = qb + r$，且 $\delta(r) < \delta(b)$，则称 R 为一个欧几里得环（**Euclid Ring**），或者欧几里得整环（**Euclid Domain**）.

直观来说，欧几里得环就是“可以做辗转相除法的环”. 这从定义就可以看出来：任取 a，再用任意的 b 去尝试除以它，总能得到 $qb + r$ 的形式，其中 r 相当于除法的余数. 虽然任意环中的元素都可以这么做分解，而欧几里得环就特殊在它还关系到一个非负整数赋值，使得余数的赋值总是小于非零除数 b 的，这样就使得我们可以多次分解，也就是进行辗转相除.

为了严格证明以上说法成立，我们还需要讨论欧几里得环的一个性质：

定理 1

设 R 是一个欧几里得环，那么对于 $r \in R$，有：

$$r = 0 \iff \delta(r) = 0.$$

证明：

任选 $a \in R$ 且 $a \neq 0$，那么 $a = 1a + 0$，故由欧几里得环的定义可推知，对于任意 $a \in R - \{0\}$，必有 $\delta(0) < \delta(a)$，由此得证.

证毕.

有了定理 1 ，我们还可以补上定义中没有说明的一点：当 $\delta(a) \leq \delta(b)$ 的时候，我们可以取 $q = 0$ 而 $r = a$ 来满足 $a = qb + r$ 且符合欧几里得环的性质. 因此，一般只讨论 $\delta(a) > \delta(b)$ 的情形.

我们来看几个典型的欧几里得环.

例 1　整数环

对于 $n \in \mathbb{Z}$，令 $\delta(n) = n$，则容易看出 $\mathbb{Z}$ 是一个欧几里得环.

例 2　多项式环

任给域 $\mathbb{K}$，其多项式环 $\mathbb{K}[x]$ 是一个欧几里得环. 取 $\delta(f(x)) = 2^{\deg(f)}$ 即可.

[9]即给每一个环中元素赋予一个非负整数.

$(deg)(f)$ 是 f 的次数，也就是系数非零的最高项的幂次．这里 δ 要取 2 的指数是因为 $\deg(0)=-\infty$，而我们希望 $\delta(0)=0$．

例 3　高斯整数环

高斯整数环是复平面上的全体坐标为整数的点的集合 $\mathbb{Z}[\mathrm{i}]$ 构成的环，并且是一个欧几里得环．证明见下．

证明：

对于任意 $z\in\mathbb{Z}[\mathrm{i}]$，令 $\delta(z)=|z|^2$．

任给 $a,b\in\mathbb{Z}[\mathrm{i}]$，不妨设 $\delta(a)>\delta(b)$．于是必有 $u,v\in\mathbb{Q}$，使得 $\frac{a}{b}=(u+v\mathrm{i})$．

于是必有 $c,d\in\mathbb{Z}$，使得 $|c-u|\leq\frac{1}{2}$ 和 $|d-v|\leq\frac{1}{2}$．

令 $q=c+d\mathrm{i}$，$r=(c+d\mathrm{i})b-(u+v\mathrm{i})b=(c+d\mathrm{i})b-a$，即可满足 $a=qb+r$ 且 $\delta(r)<\delta(b)$．

证毕．

定理 2

欧几里得环都是主理想整环．

证明：

设 R 是一个欧几里得环，令 I 为 R 的一个理想．不妨设 $I\neq\{0\}$．

由于 $\delta(I)=\{\delta(i)|i\in I\}\subseteq\mathbb{Z}^+\cup\{0\}$，故在**非零**的 $\delta(i)$ 中存在最小值 $\delta(b)$，其中 $b\in I$．

任取 $a\in I$，则存在 $q,r\in R$ 且 $\delta(r)<\delta(b)$，使得 $a-qb=r$．

由于 $a,b\in I$ 且 I 有吸收律[10]，故 $r\in I$．由于 $\delta(b)$ 是 $\delta(I)$ 中的最小非负值，因此必有 $\delta(r)=0$．由定理 1，$r=0$．

因此，$b|a$ 对任意 $a\in I$ 成立．

故 $I=\langle b\rangle$，从而得证．

证毕．

[10] 见**环的理想和商环**[571]，这意味着 $qb\in I$．

第三章

域论

域

预备知识　环[569]

体和域的概念

定义 1　体

给定一个集合 H，如果这个集合中定义了两个运算，加法“+”和乘法“$\times$”，并且 H 对于加法构成一个阿贝尔群，而 $H-\{0\}$ 构成群（0 为 H 加法群的单位元），并且乘法对加法满足分配律，即对于任何 $a,b,c\in H$，满足 $a\times(b+c)=a\times b+a\times c$，那么我们称 $(H,+,\times)$ 构成一个体（**skew field**），或称可除环（**division ring**）、除环.

像在环论中省略乘法的符号一样，我们也常常把体中的“$\times$”符号省略，比如说，将分配律表示为 $a(b+c)=ab+ac$.

简单来说，体就是能进行加减乘除的一个集合，其中加法是可交换的，乘法却不一定. 由于乘法不一定交换，这就使得除法运算相对复杂，但我们在此不过多展开.

例 1　四元数体

四元数[593] 词条中所定义的全体四元数构成的集合，配上所定义的加法和乘法，构成一个体，称为四元数体.

例 2　矩阵体

某个域上的全体 n 阶可逆矩阵，配上矩阵加法和乘法，构成一个体.

类比子群和子环的定义，我们也可以定义子体.

定义 2　子体

体 H 的子集 S 如果满足其对已有的加法和乘法仍然构成体，那么称 S 是 H 的一个子体.

子体的概念引出了以下关键概念.

定义 3　素体与素域

一个体的子体之交显然还是一个子体，因此每个体都存在唯一的非平凡不可约子体，称为这个体的素体（**prime field**）或素域.

素体又被称作素域的原因是，素体必然是 $\mathbb{Z}_p$ 或 $\mathbb{Q}$，其中 p 是素数. 这两种体都是域.

那么我们刚才讨论中的域是什么呢？

定义 4　域

给定一个体 $\mathbb{F}$，如果 $\mathbb{F}$ 的乘法满足交换律，那么称其为一个域（**field**）.

一个域的子体总是乘法可交换的，因此也都称为域的**子域**（**sub-field**）.

数学界主流将域视作乘法可交换的体，因此当谈到域时，总是认为乘法可交换. 少数数学家会把我们以上定义的体称为域，而将我们定义的域称为交换域，但这并不是主流，因此本书使用以上定义.

例 3　数域

包含整数集的域，称为数域. 最重要的数域有三个，有理数域 $\mathbb{Q}$，实数域 $\mathbb{R}$ 和复数域 $\mathbb{C}$，其中 $\mathbb{Q}$ 是最小的数域，也就是说任何数域都包含它；实数域是有理数域的完备化，意味着有理数域中的收敛数列都收敛于某个实数；复数域是最大的数域，也就是说任何数域都是复数域的子域.

注意，$\mathbb{Z}_p$ 并不是 $\mathbb{Z}$ 的子域，因为在 $\mathbb{Z}_p$ 中，$(p-1)+1=0$，而这在 $\mathbb{Z}$ 中是不可能的.

素域的概念对于描述任意的域是关键，以至于我们用素域定义了一个概念，称作域的特征：

定义 5　域的特征

给定域 $\mathbb{F}$，如果它所包含的素域是 $\mathbb{Z}_p$，那么称 $\mathbb{F}$ 的特征（**character**）是 p；如果它的素域是 $\mathbb{Q}$，那么称它的特征是 0.

由定义可见，特征的值取素数或者 0，这个值在很大程度上决定了域的代数性质.

域上的代数

预备知识 矢量空间[444]，环[569]

“代数（algebra）”一词，有两个含义.

第一个含义是指一个数学分支，代数学. 代数学研究集合中各种各样的运算结构，我们从小学就开始涉及了. 一次方程的移项、两边同乘等操作都是代数学研究的性质；理工科到了本科还必须研究线性代数（或称高等代数），作为大量理工学科的数学基础；此外，本部分“抽象代数”则研究了更为基础的一些代数学，但抽象代数本身也只是代数学这门广博学科的起点.

第二个含义，是代数学中研究的一种数学对象，代数. 简单来说，代数就是一种定义了向量乘法的向量空间. 当然，乘法的性质要具体讨论.

定义 1　狭义的代数定义

设 A 是域 $\mathbb{K}$ 上的向量空间，若在 A 中再定义代数乘法 $\times$，使得 $(A,+,\times)$ 成为环，并且 $\forall a\in\mathbb{K},\mathbf{u},\mathbf{v}\in A$ 有

$$a(\mathbf{u}\times\mathbf{v})=(a\mathbf{u})\times\mathbf{v}=\mathbf{u}\times(a\mathbf{v}) \tag{1}$$

则称 A 为域 $\mathbb{K}$ 上的代数，简称（结合）代数.

如果没有特别说明，一个代数一般是指狭义的结合代数，也就是说，是在向量空间里定义向量的乘法，使得它还能构成一个环. 当然，乘法只是一种运算，它只要求对加法满足**分配律**[1]，不一定是结合的. 如果所定义的乘法不是结合的，那么我们称这样的结构是一个**非结合代数**.

代数的例子

下面我们来看几个例子，加深对代数这一概念的理解.

结合代数

例 1　矩阵代数

域 $\mathbb{K}$ 上的 $n\times n$ 的矩阵的全体 $\mathrm{GL}(n,\mathbb{K})$（例 5 [534]）在矩阵的加法，标量

[1]分配律有时也被表示为线性性，在定义 1 中由于我们要求 $(A,+,\times)$ 成环，故已经默认了具有分配律.

与矩阵的相乘运算下构成一个线性空间，在矩阵的加法和矩阵的乘法运算下构成一个环，因此它是一个结合代数.

例 2

对于域 $\mathbb{K}$ 及有限群 $G=\{g_1,g_2,\cdots,g_n\}$，我们可构成群代数 $A(G)=\left\{u|u=\sum a^i g_i, a^i\in\mathbb{K}\right\}$，其中加法为

$$u+v=\sum a^i g_i+\sum b^i g_i=\sum\left(a^i+b^i\right)g_i$$

数乘为

$$au=\sum\left(aa^i\right)g_i, a\in\mathbb{K}$$

而代数乘法为

$$u\times v=\left(\sum a^i g_i\right)\times\left(\sum b^i g_i\right)=\sum_{i,j}a^i b^j\left(g_i g_j\right)$$

这是一个结合代数.

李代数

例 3　三维实线性李代数

在 3 维实向量空间中，以两个向量的叉积 $\mathrm{A}\times\mathrm{B}$ 来定义它们的代数乘法运算，则它们构成一个代数，此时有

$$\begin{aligned}&\mathrm{A}\times\mathrm{A}=0\\&\mathrm{A}\times\mathrm{B}=-\mathrm{B}\times\mathrm{A}\\&\mathrm{A}\times(\mathrm{B}+\mathrm{C})=\mathrm{A}\times\mathrm{B}+\mathrm{A}\times\mathrm{C}\\&(\mathbf{B}+\mathrm{C})\times\mathrm{A}=\mathrm{B}\times\mathrm{A}+\mathrm{C}\times\mathrm{A}\end{aligned}\tag{2}$$

以及

$$(\mathrm{A}\times\mathrm{B})\times\mathrm{C}+(\mathrm{B}\times\mathrm{C})\times\mathrm{A}+(\mathrm{C}\times\mathrm{A})\times\mathrm{B}=0\tag{3}$$

3 维向量空间配上向量外积作为乘法，得到一个非结合代数，三维实线性李代数.

通常把式 3 称为**雅可比恒等式**. 具有这种性质的代数称为**李代数（Lie Algebra）**. 我们会在将来详细讨论李代数.

例 4

$\mathrm{gl}(n,\mathbb{C})$ 除矩阵的加法及数乘外，再定义代数乘法 $\mathrm{A}\times\mathrm{B}=[\mathrm{A},\mathrm{B}]=\mathrm{AB}-\mathrm{BA},\mathrm{A},\mathrm{B}\in\mathrm{gl}(n,\mathbb{C})$．显然 $\mathrm{gl}(n,\mathbb{C})$ 构成一个李代数．

下面一个例子需要一些分析力学中的内容．如果不知道分析力学也没关系，知道即可．

例 5

在分析力学中，正则变量记为 $p_i,q_i,i=1,2,\cdots,s$．函数 $u(p,q)$ 和 $v(p,q)$ 的泊松括号定义为

$$\{u,v\}=\sum_{i=1}^{s}\left(\frac{\partial u}{\partial q_i}\frac{\partial v}{\partial p_i}-\frac{\partial u}{\partial p_i}\frac{\partial v}{\partial q_i}\right) \tag{4}$$

于是我们得到了一个李代数．

未完成：也许可以引用李代数中使用李代数反推出结合代数的方法来描述泊松括号对应的结合代数．

外代数

预备知识 直和[465]

例 6　外代数

给定域 $\mathbb{K}$ 上的 n 维线性空间 V，任取其一组基 $\{\mathbf{e}_i\}_1^n$．我们进行以下构造：

- 对于每两个元素 $\mathbf{v},\mathbf{u}\in V$，我们定义一个新的元素 $\mathbf{v}\wedge\mathbf{u}$，只要求它满足线性性和反对称性[2]．这样，对于两个非零且不相等的 $\mathbf{v},\mathbf{u}\in V$，$\mathbf{v}\wedge\mathbf{u}$ 是一个不在 V 中的新元素；当 $\mathbf{v}=\mathbf{u}$ 或其中之一为 0 时，$\mathbf{v}\wedge\mathbf{u}=0\in V$．由于线性性，我们用集合 $\{\mathbf{e}_i\wedge\mathbf{e}_j\}_{i\neq j}$ 作为基底，在域 $\mathbb{K}$ 上构造一个新的线性空间，记为 $A^2(V)$．由于该运算得到的元素要么是 0，要么是 V 以外的元素，因此我们把它称为外积（**exterior product**）．

[2]线性性即 $\forall a_1\mathbf{v}_1+a_2\mathbf{v}_2,u\in V$，必有 $(a_1\mathbf{v}_1+a_2\mathbf{v}_2)\wedge\mathbf{u}=a_1\mathbf{v}_1\wedge\mathbf{u}+a_2\mathbf{v}_1\wedge\mathbf{u}$．反对称性即 $\forall\mathbf{v},\mathbf{u}\in V$，有 $\mathbf{v}\wedge\mathbf{u}=-\mathbf{u}\wedge\mathbf{v}$．

- 将以上步骤推广，对于每 k 个元素 $\mathbf{v}_i \in V$，我们定义一个新的元素 $\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \cdots \wedge \mathbf{v}_k$，要求它满足线性性、反对称性和结合性[3]. 由于反对称性，k 个 $\{e_i\}$ 的外积中只有 C_n^k 个不为零，用这些不为零的外积作为基底在 $\mathbb{K}$ 上构造一个线性空间，记为 $A^k(V)$. 特别地，$V = A^1(V)$.

- 同样由于反对称性，若干元素 $\mathbf{v}_i$ 中只要有两个相等，它们的外积就为零. 这就意味着我们最多可以取到 $A^n(V)$，而对于 $m > n$，都有 $A^m(V) = \{0\}$. 这样，将各 $A^k(V)$ 视为互相不同的线性空间，其中只有 $n+1$ 个非平凡，它们作为向量空间的直和 $\bigoplus_{k=1}^{n} A^k(V)$ 也是一个向量空间，记为 $A(V)$. 由前面的步骤可知，任取 $\mathbf{v}, \mathbf{u} \in A(V)$，都有 $\mathbf{v} \wedge \mathbf{u} \in A(V)$，因此外积在 $A(V)$ 上是封闭的.

- 将各 $A_k(V)$ 的加法都记为 $+$，它们的直和 $A(V)$ 的加法也记为 $+$，那么用外积作为乘法，$(A(V), +, \wedge)$ 构成一个代数，称为线性空间 V 上的外代数（**exterio algebra**），又称格拉斯曼代数（**Grassmann algebra**）.

n 维线性空间 V 上的 k 次外代数可用于表示 V 上的 k 维“体积”，而当 $k = n$ 时就是我们所熟悉的行列式：取定线性空间 V 的基 $\{\mathbf{e}_i\}$，那么 n 个元素的外积 $\sum_i (a_{1i}\mathbf{e}_i) \wedge \sum_i (a_{2i}\mathbf{e}_i) \wedge \cdots \wedge \sum_i (a_{ni}\mathbf{e}_i) = \left|\{a_{ij}\}\right| \mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n$.

外代数最重要的例子是外微分[604]，它是以**微分形式**的集合构建的线性空间的外代数.

四元数

预备知识 复数[142]，矢量叉乘[387]，域[588]

四元数的历史

复数可以定义为具有特定乘法有序实数对，也可以看成是在实数域上添加一个新的元素 i 所得到的扩域.

[3]结合性即对于任意三个元素 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$，都有 $(\mathbf{v}_1 \wedge \mathbf{v}_2) \wedge \mathbf{v}_3 = \mathbf{v}_1 \wedge (\mathbf{v}_2 \wedge \mathbf{v}_3)$，这样就可以摆脱括号而直接将这三个元素的外积表示为 $\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \mathbf{v}_3$.

复数应用极为广泛，因为它具有非常良好的代数性质和分析性质，同时可以用来表示二维矢量空间，极大地方便了许多几何运算．John Derbyshire 在他的科普作品 Unknown Quantity: Real and Imaginary History of Algebra 中写道，“如果从一维的实数到二维的复数的变更会给我们带来这样巨大的力量和见识，那么为什么要止步呢？难道就没有正等待着被发现的其他种类的数，比如说超复数，它们本身就是三维的吗？还有，难道那些数就不能让我们进一步认识数学吗?”

19 世纪初的爱尔兰数学家 William Hamilton 就曾试图寻找一种具有特定乘法的三元数组，期望能和三维空间的坐标一一对应．然而他努力多年却未能找到可以满足结合律、分配律等的三元数乘法．事实上，这样的三元数乘法是不存在的，但 Hamilton 当时不知道．

1843 年 10 月 16 日，Hamilton 在和妻子一同前往爱尔兰皇家学会议会的路上，突然灵光一现，意识到了问题的本质．他“很不理智”地拿出刀来，在布鲁姆桥的一块石头上刻下了以下公式：

$$\mathrm{i}^2=\mathrm{j}^2=\mathrm{k}^2=\mathrm{ijk}=-1 \tag{1}$$

这一公式包含了一个四元数的良好的乘法定义．

四元数的定义

现代四元数的定义是把每一个四元数看成一个实数和一个三维实矢量的组合：$(a_0,\mathbf{a})$．在这个语境下，i，j，k 被看成是一个标准正交基中的矢量，为了避免将来的混淆，把这三个基矢量重新表示成 $1,\hat{\mathbf{i}},\hat{\mathbf{j}}$ 和 $\hat{\mathbf{k}}$ 后，四元数的运算可以定义如下：

定义 1　四元数的四则运算

记四元数集合为 $\mathbb{H}=\{(a_0,\mathbf{a})|a_0\in\mathbb{R},\mathbf{a}\in\mathbb{R}^3\}$，$\forall(a_0,\mathbf{a}),(b_0,\mathbf{b})\in\mathbb{H}$，加法定义为

$$(a_0,\mathbf{a})+(b_0,\mathbf{b})=(a_0+b_0,\mathbf{a}+\mathbf{b}) \tag{2}$$

乘法定义为（乘号可以省略或者用其他符号）

$$(a_0,\mathbf{a})\times(b_0,\mathbf{b})=(a_0b_0-\mathbf{a}\cdot\mathbf{b},a_0\mathbf{b}+b_0\mathbf{a}+\mathbf{a}\times\mathbf{b}) \tag{3}$$

注意，虽然四元数的加法显然满足交换律，但由于一般来说矢量的叉积不可交换，因此四元数乘法一般不具有交换律．只有 $\mathbf{a}$ 平行于 $\mathbf{b}$（包括其中之一为 $\mathbf{0}$）的时候，才会有 $\mathbf{a}\times\mathbf{b}=\mathbf{0}=\mathbf{b}\times\mathbf{a}$．四元数的乘法满足分配律（习题 1）．

四元数 $A=(a_0,\mathbf{a})$ 的实数部分 a_0 被称为**标部**，矢量部分 $\mathbf{a}$ 称为**矢部**．为了方便，有时候也把这个四元数记为 $a_0+\mathbf{a}$，类比复数的表示法 $(a,b)=a+b\mathrm{i}$．当标部为 0 时，可以直接将四元数记为矢量 $\mathbf{a}$，同理，当矢部为零时也可以记为标量 a_0．这样，$a_0+\mathbf{a}$ 就可以看作两个四元数相加．和复数同理，我们认为矢部为 0 的四元数就是实数．

如果我们把四元数的矢部限制在一个方向上，那么得到的运算结构实际上就是复数．你可以通过比较四元数乘法规则和复数乘法规则来看出这一点——注意在限制条件下任何矢部的向量积都是 $\mathbf{0}$.

除了把四元数看成一个标量加一个三维矢量以外，我们也可以把四元数看成一个四维矢量，就像复数既可以看成一个实数配上一个一维矢量，也可以看成一个二维矢量一样．这种情况下，四元数实际上就是实数域上的一个四维**代数**[590]4．

和复数一样，四元数也有共轭和模的概念．

定义 2　共轭、模方和模

- 令四元数为 $A=a_0+\mathbf{a}$，记 $\widetilde{A}=a_0-\mathbf{a}$，称为 A 的共轭（**conjugate**）．
- 记 $||A||=||\widetilde{A}||=A\widetilde{A}=\widetilde{A}A=a_0^2+|\mathbf{a}|^2$，称为 A 的模方．
- 记 $|A|=\sqrt{||A||}$，称为 A 的模．

习题 1

- 证明四元数满足乘法结合律
- 证明 $||A||=||\widetilde{A}||=\widetilde{A}A=a_0^2+|\mathbf{a}|^2$

[4]从代数角度来说，Hamilton 爵士当年苦苦思索多年的问题就是“是否存在实数域上的三维代数”，今天我们已经知道，实数域上只存在一维、二维、四维和八维的“具有积性的可除代数”，分别对应实数、复数、四元数和八元数，其中四元数已经不具有交换性，而八元数连结合性都丢了．这里的“积性”是指 $|a||b|=|ab|$，而“可除性”就是“乘法单位元存在性”.

- 若将单位矢量 $\hat{\mathbf{e}}$ 看作四元数（实部为零），证明 $\hat{\mathbf{e}}^2=-1,\ \hat{\mathbf{e}}^3=-\hat{\mathbf{e}}$

四元数的指数表示

记 e 是自然对数的底数，那么对于任意实数 x，有 $\mathrm{e}^x=\sum_{n=0}^{\infty}x^n/n!$. 如果把实数 x 替换为其它能够进行乘方的对象，比如说矩阵、复数等，那么还可以把指数的概念拓展到更多的对象上，构造出复数的指数、矩阵的指数等.

四元数也可以乘方，因此可以定义四元数的指数：

定义 3　四元数的指数表示

令四元数为 $A=a_0+\mathbf{a}$，则使用四元数的乘法（式 3），定义（类比复数的指数函数定义[351]）

$$\mathrm{e}^A=\mathrm{e}^{a_0}\mathrm{e}^{\mathbf{a}}=\mathrm{e}^{a_0}\sum_{n=0}^{\infty}\frac{\mathbf{a}^n}{n!}=\mathrm{e}^{a_0}\sum_{n=0}^{\infty}\frac{(a\hat{\mathbf{a}})^n}{n!}=\mathrm{e}^{a_0}[\cos a+\sin(a\hat{\mathbf{a}})] \tag{4}$$

其中 $\hat{\mathbf{a}}$ 是和 $\mathbf{a}$ 同方向的单位矢量，a 是 $\mathbf{a}$ 作为四元数的模或者作为三维实矢量的模.

习题 2　与复数的类比

比较复数和四元数的指数表示，说明为什么可以将复数看成矢部限制在一个方向上的四元数.

由于四元数的加法满足交换律，而乘法一般不满足，因此一般情况下四元数并不满足 $\mathrm{e}^{A+B}=\mathrm{e}^A\mathrm{e}^B$. 因此，四元数的指数表示法在计算中其实用得不多.

四元数的拓展

我们以上定义的四元数，可以看成四个基矢量 1，$\hat{\mathbf{i}}$，$\hat{\mathbf{j}}$ 和 $\hat{\mathbf{k}}$ 的线性组合的集合，而这些线性组合的系数取自实数域 $\mathbb{R}$. 如果系数还可以取自复数域 $\mathbb{C}$，定义出来的数被称为**双四元数**，或**复四元数**. 一个双四元数可以看成一个复数和一个复数域上的三维矢量的组合.

对于一个双四元数 $A=a_0+a_1\hat{\mathbf{i}}+a_2\hat{\mathbf{j}}+a_3\hat{\mathbf{k}}$，共轭的概念就被细分了：

定义 4　双四元数的共轭

- 四元共轭：$\widetilde{A}=a_0-a_1\hat{\mathbf{i}}-a_2\hat{\mathbf{j}}-a_3\hat{\mathbf{k}}$
- 反共轭：$A^C=a_0^*+a_1^*\hat{\mathbf{i}}+a_2^*\hat{\mathbf{j}}+a_3^*\hat{\mathbf{k}}$
- 转置：$A^T=a_0+a_1\hat{\mathbf{i}}-a_2\hat{\mathbf{j}}+a_3\hat{\mathbf{k}}$
- 复共轭：$A^*=a_0^*-a_1^*\hat{\mathbf{i}}+a_2^*\hat{\mathbf{j}}-a_3^*\hat{\mathbf{k}}$
- 厄密（hermitian）共轭：$A^\dagger=a_0^*-a_1^*\hat{\mathbf{i}}-a_2^*\hat{\mathbf{j}}-a_3^*\hat{\mathbf{k}}$

双四元数的模方是用四元共轭定义的：$||A||=A\widetilde{A}=\widetilde{A}A=\sum_{n=0}^{3}a_n^2$．但是和四元数不同的是，双四元数的模方并不总是非负实数，而是有可能取任何复数．

习题 3　双四元数共轭模方的关系

证明：$||A||=||\widetilde{A}||=||A^T||=||A^C||^*=||A^*||^*=||A^\dagger||^*$.

习题 4　双四元数共轭的乘法规律

设 A 和 B 是两个双四元数，证明：

$$\begin{cases}A^CB^C=(AB)^C\\A^*B^*=(AB)^*\\\widetilde{A}\widetilde{B}=\widetilde{BA}\\A^TB^T=(BA)^T\\A^\dagger B^\dagger=(BA)^\dagger\end{cases}\tag{5}$$

和普通四元数不同的是，双四元数中有可能存在非零但模为 0 的数，比如下面例 1 所举例的 u 和 v.

类似地，让四元数的系数取值为四元数，所得到的数被称为**二级四元数**．一般地，让四元数的系数取值为 n **级四元数**，则能得到 $n+1$ **级四元数**．但是由于四元数不再是一个域，也没有交换律，用其做系数的效果远远不如复数系数，应用范围相对有限，在此不作详细介绍．

四元数的矩阵表示

如果将四元数的四个基 $1, \hat{\mathbf{i}}, \hat{\mathbf{j}}$ 和 $\hat{\mathbf{k}}$ 分别对应到以下四个矩阵

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \hat{\mathbf{i}} = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix} \qquad \hat{\mathbf{j}} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad \hat{\mathbf{k}} = \begin{pmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \tag{6}$$

那么一个任意的四元数或双四元数 $A = a_0 + \mathbf{a}$ 可以用矩阵表示为

$$\begin{pmatrix} a_0 + \mathrm{i}a_1 & a_2 + \mathrm{i}a_3 \\ -a_2 + \mathrm{i}a_3 & a_0 - \mathrm{i}a_1 \end{pmatrix} \tag{7}$$

例 1　矩阵表示的例子

$$1 + \begin{pmatrix} 1 \\ 0 \\ \mathrm{i} \end{pmatrix} = 1 + 1\hat{\mathbf{i}} + \mathrm{i}\hat{\mathbf{k}} = \begin{pmatrix} 1+\mathrm{i} & -1 \\ -1 & 1-\mathrm{i} \end{pmatrix} \tag{8}$$

$$u = \frac{1}{2}(1 - \mathrm{i}\hat{\mathbf{i}}) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{9}$$

$$v = -\frac{1}{2}(\hat{\mathbf{j}} + \mathrm{i}\hat{\mathbf{k}}) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{10}$$

容易验证，四元数的加法和乘法运算，与把基分别换成这四个矩阵以后的矩阵加法和乘法运算是同构的.

在矩阵表示中，A 的转置共轭就是矩阵 A 的转置，复共轭就是 A 的各矩阵元取复共轭，厄密共轭就是 A 的厄密共轭矩阵（矩阵转置后再取各矩阵元的复共轭），而四元数的模方就是其矩阵表示的行列式：$||A|| = \det A$. 因此，模方不为零的（双）四元数对应的矩阵是可逆的，其逆矩阵所表示的就是该四元数的逆.

域的扩张

预备知识　域[588]

域的单扩张

如果在一个域中添加不属于域集合的元素，我们可以得到一个更大的集合．要让这个新集合成为域，我们就得定义新元素和原来域中元素相加和相乘的结果；无论怎么定义，这个结果必须满足域的公理．如果在集合中任何元素都无法成为某个运算结果，那么我们就必须再引入新的元素来作为这个结果．以此类推，不停地添加新元素，直到最后不需要添加新元素了，那最后这个集合就是一个新的域，它包含了原来的域．这个域是原来的域的**扩张**，并且包含最初那个新元素的最小的域，因此被称为**单元素扩张**，简称**单扩张**．

域的单扩张具体是怎么进行的呢？我们将从例子开始说明，最后引入域的单扩张的定义．

习题 1　有理数域的 $\sqrt{2}$ 扩张

给定有理数域 $\mathbb{Q}$，则 $\sqrt{2}$ 并不是 $\mathbb{Q}$ 的元素．将 $\sqrt{2}$ 添加进去，那么由于 $\sqrt{2}$ 是实数域 $\mathbb{R}$ 的元素，这提示我们可以把新的运算结果按照 $\mathbb{R}$ 中的运算来定义．这样，$\mathbb{Q}$ 中添加 $\sqrt{2}$ 的单扩域的集合就是 $\{a+b\sqrt{2}|a,b\in\mathbb{Q}\}$．

请验证，任意 $a+b\sqrt{2}$ 都必须在扩域中，而这个集合也满足了加法和乘法的封闭性，所以构成了一个域；换句话说，这个集合是包含 $\sqrt{2}$ 和全体有理数的最小的域．

习题 2　超越扩域

依然给定有理数域 $\mathbb{Q}$，这次添加的元素是 π．那么扩域应当是

$$\{\frac{\sum_{i=0}^{N}a_i\pi^i}{\sum_{j=0}^{M}b_j\pi^j}|a_i,b_j\in\mathbb{Q},b_j\text{不全为零，}N,M\in\mathbb{Z}^+\}\tag{1}$$

也就是 π 的所有有理系数多项式的分式构成的集合，其中分母不是 0．

验证这是一个域，并且是包含全体有理数和 π 的最小域．注意 a_i 全为 0 的情况意味着什么．

这是 $\mathbb{Q}$ 中添加 π 后的单扩张．

观察这两个域扩张的特点，我们容易发现，域添加某个元素后的单扩张，必须包含这个元素的所有多项式，系数取自域中；包含了所有多项式以后，也

就构成了一个域，域的扩张就完成了．因此，域的单扩张就是把新元素的多项式都添加进去的过程．

但是两个场景是不太一样的．$\mathbb{Q}$ 中添加 π 的过程中，包含了所有 π 的有理系数多项式，但是添加 $\sqrt{2}$ 的时候却只需要最多 1 阶的多项式．这是因为 $\sqrt{2}$ 的 2 次方是一个有理数，落入了 $\mathbb{Q}$ 中，因此 $\sqrt{2}$ 的 2 阶及以上的有理系数多项式都可以表示为最多 1 阶的多项式．

换一种更方便的表述，那就是 x^2-2 这个多项式是阶数最小的能将 $\sqrt{2}$ 映射为 0 的多项式．这就引出**最小多项式**的概念．

定义 1　最小多项式

给定一个环 R 和任意元素 x_0，其中 x_0 不一定是 R 中的元素．记

$$R[x_0]=\{\sum_{i=0}^{N} r_i x_0^i | r_i \in R, N \in \mathbb{Z}^+\}$$

称 $R[x_0]$ 是 x_0 在 R 中的多项式环．在 $R[x_0]$ 中，如果存在一个多项式 $f(x_0)=0$，并且任何更低阶的多项式的值都不是 0，那么称 $f(x_0)$ 是 x_0 在 R 上的最小多项式（**minimal polynomial**），或极小多项式．使不存在值为 0 的多项式（这也意味着所有多项式的值甚至都不在 R 中），那么我们说，x_0 在 R 中不具有最小多项式．

在习题 1 和习题 2 中，π 和 $\sqrt{2}$ 的本质区别就是：在 $\mathbb{Q}$ 中，π 没有最小多项式，但是 $\sqrt{2}$ 具有最小多项式 x^2-2．这决定了同一个域 $\mathbb{Q}$ 分别添加 π 和 $\sqrt{2}$ 后的不同单扩张．

可是，我们把域抽象地看成满足一定条件的集合时，和所有集合一样，域中的元素叫什么名字并不重要．对于习题 1 和习题 2 中的情况，我们其实是借用了 $\mathbb{R}$ 中已有的元素名称来引出两个不同的单扩域的．如果不进行这样的类比引出，而只是单纯地说在 $\mathbb{Q}$ 中加入某个新的元素 x_0 来进行单扩张，那么结果可能是 $\{a+bx_0|a,b\in\mathbb{Q}\}$，可能是 $\{\sum_{i=0}^{N} a_i x_0^i | a_i \in \mathbb{Q}, N \in \mathbb{Z}^+\}$，可能是 $\{a+bx_0+cx_0^2|a,b,c\in\mathbb{Q}\}$，甚至还可能形式上也是 $\{a+bx_0|a,b\in\mathbb{Q}\}$，只不过 $x_0^2=2/3$ 了．这时该怎么描述域的单扩张呢？

答案是应用最小多项式对单扩张的决定作用．

定义 2　域的单扩张

给定域 $\mathbb{F}$ 和一个元素 x_0，其中 x_0 不一定在 $\mathbb{F}$ 中．定义 $\mathbb{F}$ 添加 x_0 的单扩

张是集合

$$\{\frac{\sum_{i=0}^{N} a_i x_0^i}{\sum_{j=0}^{M} b_j x_0^j} | a_i, b_j \in \mathbb{F}, \quad \sum_{j=0}^{M} b_j x_0^j \neq 0, \quad N, M \in \mathbb{Z}^+\} \tag{2}$$

记为 $\mathbb{F}(x_0)$；容易验证，这是一个域.

在定义 2 中，如果 x_0 在 $\mathbb{F}$ 中没有最小多项式，那么 $\mathbb{F}(x_0)$ 中的各元素彼此不同，此时我们称 x_0 是 $\mathbb{F}$ 的**超越元（transcendental element）**；

如果 x_0 在 $\mathbb{F}$ 中有最小多项式 $f(x)$，那么称 x_0 是 $\mathbb{F}$ 的**代数元（algebraic element）**，此时在 $\mathbb{F}(x_0)$ 中，若记 g, h 为任意多项式，就有 $g(x_0)$ 和 $g(x_0) + f(x_0)h(x_0)$ 的值相等. 这提示我们，在这种情形下，似乎可以把 $\mathbb{F}(x_0)$ 看成是某个超越元 x 的单扩张 $\mathbb{F}(x)$ 集合中，把 $g(x)$ 和 $g(x) + f(x)h(x)$ 认为是等价的，所得到的**商集**. 实际上，严格来说不是商集，而是商集中再把类似

$$\frac{\sum_{i=0}^{N} a_i x_0^i}{f(x)h(x)}$$

的元素剔除以后剩下的集合. 当然了，0 不要剔除.

域的扩张次数

以上，我们讨论了如何添加一个元素，并根据元素的最小多项式来得到域的单扩张. 对一个域进行单扩张以后，再进行单扩张，如此反复，我们可以获得域的多次扩张.

第四章

外代数

外代数

预备知识 张成空间[447]，四元数[593]，域上的代数[590]

外代数是一种利用已有线性空间构造“代数”这一对象的通用方法，同时蕴含了对三维矢量分析中代数结构的本质解释.

给定线性空间 V，任取 $x,y\in V$，定义 $x\wedge y\notin V$ 是一个新的元素，其中符号 $\wedge$ 称作**外积（exterior product）**，有时也叫做**楔积（wedge product）**，前者是因为这个运算得到的是 V 以外的新元素，后者是由于符号长得像个楔子. 注意，为了方便，我们没有使用线性代数中常见的粗体正体符号来表示向量.

利用各 $x\wedge y$ 构造新的线性空间：定义 $x\wedge y=-y\wedge x$ 对所有 $x,y\in V$ 成立，这同时意味着 $x\wedge x=0$. 集合 $\{x\wedge y|x,y\in V\}$ 张成的线性空间，记为 $\bigwedge^2 V$. 同时，为了统一考虑，记 $V=\bigwedge^1 V$.

$\bigwedge^1 V$ 和 $\bigwedge^2 V$ 之间也可以进行楔积，并且满足**结合律**：$x\wedge(y\wedge z)=(x\wedge y)\wedge z$，由此可以拿掉结合括号，定义 $x\wedge y\wedge z=x\wedge(y\wedge z)=(x\wedge y)\wedge z$. 集合 $\{x\wedge y\wedge z|x,y,z\in V\}$ 张成的线性空间，记为 $\bigwedge^3 V$.

同理，我们可以构造出任意阶的 $\bigwedge^k V$. 要注意的是，如果 $k>\dim V$，那么 $\bigwedge^k V=\{0\}$. 另外，把 V 的基本域 $\mathbb{F}$ 看成一个一维线性空间，记 $\mathbb{F}=\bigwedge^0 V$.

不同线性空间之间可以用直和组合在一起，因此以上这些空间也都可以作直和，得到一个 $\bigwedge V=\bigoplus_k\bigwedge^k V=\mathbb{F}\oplus\bigwedge^1 V\oplus\bigwedge^2 V\cdots$. 这个 $\bigwedge V$，就被称作 V 上的**外积空间（exterior product space）**或**楔积空间（wedge product space）**. $\mathbb{F}$ 是 V 的基域，视为 $\mathbb{F}$ 自身上的一维线性空间.

定理 1　外代数

任给域 $\mathbb{F}$ 上的线性空间 V，则外积 $\wedge$ 是一个 $\bigwedge V$ 上的向量乘法，并且满足结合性，有单位元 $1\in\bigwedge^0 V$，故构成一个 $\mathbb{F}$ 上的代数. 称这个代数为 V 上的外代数（**exterior algebra**）或格拉斯曼代数（**Grassmann algebra**）.

外代数中的元素可以有形象的几何理解. $\bigwedge^1 V$ 中的元素就是 V 中的元素，我们可以想象成箭头. $\bigwedge^2 V$ 中的元素可以看成箭头对，或者是箭头对表示的平行四边形. 同样，$\bigwedge^k V$ 中的元素都可以看成是 k 个箭头张成的一个 k 维对象.

外代数有一个重要的性质，我们用习题 1 和习题 2 来阐述：

习题 1

证明：如果 $k > \dim V$，那么 $\bigwedge^k V = \{0\}$.

习题 2

证明：对于 $\dim V = k$，有 $\dim \bigwedge V = 2^k$. 思路提示：考虑各 $\dim \bigwedge^i V$ 的值，再对比 $(1+1)^k$ 的二项式展开.

三维欧几里得空间 $\mathbb{R}^3$ 中的叉乘实际上就是外积. 这是因为，$\dim \mathbb{R}^3 = \dim \bigwedge^2 \mathbb{R}^3$，这样一来，如果给定 $\mathbb{R}^3$ 的标准正交基 $\{x, y, z\}$，那么我们可以建立同构 $* : \bigwedge^2 \mathbb{R} \to \mathbb{R}^3$，使得 $*(x \wedge y) = z, *(y \wedge z) = x, *(z \wedge x) = y$，这样就可以通过这个同构来把外积变成 $\mathbb{R}^3$ 内部的向量积. 这一映射也是叉乘的“右手定则”的来源，我们也完全可以规定 $*(x \wedge y) = -z, *(y \wedge z) = -x, *(z \wedge x) = -y$，这样定义出来的叉乘就是符合左手定则的了.

三维线性空间是唯一可以构造反交换代数的非平凡空间，就是因为只有三维的 V 才满足 $\dim V = \dim \bigwedge^2 V$，因而可以建立 $\bigwedge^2 V$ 和 V 之间的同构，从而把楔积变成叉积. 相应地，比复数更高维的可除代数只有四元数.

外代数是一个“**分次线性空间（graded vector space）**”，就是说，它作为一个线性空间，每个向量具有一个“次数”，定义如下：每个 $\bigwedge^k V$ 中的向量，其**次数（grade）**就是 k；对于任意向量 $v \in V$，我们总可以把它拆分成各 $\bigwedge^k V$ 中基向量的线性组合，这些基向量中次数最高的就定义为 v 的次数.

外导数

预备知识 微分形式[700]，外代数[603]

外导数的定义

外导数是一种流形上的**微分形式外代数**[1]上的映射，其术语分为两部分，“外”和“导数”. “外”，指的是它把各 $\Omega^k(M)$ 中的元素映射到 $\Omega^k(M)$ **之外**；“导数”，指的是它具有和求导类似的性质. 实际上，矢量分析中的求导就是外导

[1]即流形上的余切向量场集合作为线性空间所生成的外代数.

数的一个特例——你可能会问，求导并不具有“外”的特点，怎么就是特例了呢？我们会在本节中解释这一点.

定义 1　外导数

给定流形 M，其外微分代数是 $\Omega(M)$. 定义映射 $\mathrm{d}:\Omega(M)\to\Omega(M)$，满足：

1. $\forall\omega\in\Omega^k(M)$，有 $\mathrm{d}\omega\in\Omega^{k+1}(M)$.
2. 对于光滑函数 $f\in C^\infty(M)$，$\mathrm{d}f$ 就是 f 的方向导数（1-形式）.
3. 线性性：任取 $a,b\in\mathbb{R}$ 和 $\omega,\mu\in\Omega(M)$，有 $\mathrm{d}(a\omega+b\mu)=a\mathrm{d}\omega+b\mathrm{d}\mu$.
4. **Leibniz** 性：对于 $\omega\in\Omega^k(M),\mu\in\Omega^1(M)$，我们有 $\mathrm{d}(\omega\wedge\mu)=\omega\wedge\mathrm{d}\mu+(-1)^k\mu\wedge\mathrm{d}\omega$[2].
5. 幂零性：任取 $\omega\in\Omega(M)$，都有 $\mathrm{d}(\mathrm{d}\omega)=0$.

称这个映射为 $\Omega(M)$ 或者说 M 上的一个外导数，又称外微分.

定义中的 Leibniz 性要特别注意. 考虑到外积的**反对称性**，我们也完全可以把这一条写成 $\mathrm{d}(\omega\wedge\mu)=\omega\wedge\mathrm{d}\mu+\mu\wedge\mathrm{d}\omega$. 写成定义中减号的形式，是为了更方便计算出诸如 $\mathrm{d}(\mathrm{d}\omega\wedge\mu\wedge\mathrm{d}\nu)=-\mathrm{d}\omega\wedge\mathrm{d}\mu\wedge\mathrm{d}\nu$ 的结果.

三维欧几里得空间

在**外代数**[603] 中我们提到过，$\mathbb{R}^3$ 和 $\bigwedge^2\mathbb{R}^3$ 同构. 在流形 $\mathbb{R}^3$ 上，2-形式构成的线性空间 $\Omega^1(\mathbb{R}^3)\cong\mathbb{R}^3$，因为是由基 $\{\mathrm{d}x,\mathrm{d}y,\mathrm{d}z\}$ 张成的. 这样，我们也可以定义 $\Omega^2(\mathbb{R}^3$ 到 $\Omega^1(\mathbb{R}^3)$ 之间的同构. 这个同构的存在，意味着我们可以把**旋度**和**散度**视为外导数的特例. 我们观察以下例子来说明这一点：

旋度

考虑 $\mathbb{R}^3$ 中任意的 1-形式 $\omega_x\mathrm{d}x+\omega_y\mathrm{d}y+\omega_z\mathrm{d}z$，其中各 ω_i 是 0-形式，即光滑函数. 考虑到外导数对于光滑函数就是方向导数，我们可以得知，对于任意

[2]注意 μ 特指 1-形式.

的 $a \in \{x, y, z\}$，有 $\mathrm{d}\omega_a = \partial_x \omega_a \mathrm{d}x + \partial_y \omega_a \mathrm{d}y + \partial_z \omega_a \mathrm{d}z$．这样，我们就可以计算出：

$$\begin{aligned} &\mathrm{d}(\omega_x \mathrm{d}x + \omega_y \mathrm{d}y + \omega_z \mathrm{d}z) \\ =&(\partial_y \omega_z - \partial_z \omega_y)\mathrm{d}y \wedge \mathrm{d}z + \\ &(\partial_z \omega_x - \partial_x \omega_z)\mathrm{d}z \wedge \mathrm{d}x + \\ &(\partial_x \omega_y - \partial_y \omega_x)\mathrm{d}x \wedge \mathrm{d}y \end{aligned} \tag{1}$$

如果我们把 $\mathrm{d}y \wedge \mathrm{d}z$ 看成 $\mathrm{d}x$、把 $\mathrm{d}z \wedge \mathrm{d}x$ 看成 $\mathrm{d}y$、把 $\mathrm{d}x \wedge \mathrm{d}y$ 看成 $\mathrm{d}z$，那么式 1 右边各项系数刚好对应向量场 $\left(\omega_x, \omega_y, \omega_z\right)^{\mathrm{T}}$ 的旋度．因此，我们把三维欧几里得空间中的旋度，看成是 $\Omega^1(\mathbb{R}^3) \to \Omega^2(\mathbb{R}^3)$ 的外导数．

散度

考虑 $\mathbb{R}^3$ 中任意的 2-形式 $\omega_x \mathrm{d}y \wedge \mathrm{d}z + \omega_y \mathrm{d}z \wedge \mathrm{d}x + \omega_z \mathrm{d}x \wedge \mathrm{d}y$．同样，考虑到 $\mathrm{d}\omega_a = \partial_x \omega_a \mathrm{d}x + \partial_y \omega_a \mathrm{d}y + \partial_z \omega_a \mathrm{d}z$，我们可以计算得：

$$\begin{aligned} &\mathrm{d}(\omega_x \mathrm{d}y \wedge \mathrm{d}z + \omega_y \mathrm{d}z \wedge \mathrm{d}x + \omega_z \mathrm{d}x \wedge \mathrm{d}y) \\ =&(\partial_x \omega_x + \partial_y \omega_y + \partial_z \omega_z)\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z \end{aligned} \tag{2}$$

观察结果的系数可见，2-形式的外导数实际上就是其散度．因此，我们把三维欧几里得空间中的散度，看成是 $\Omega^2(\mathbb{R}^3) \to \Omega^3(\mathbb{R}^3)$ 的外导数．

习题 1　旋度和散度

利用外导数的定义，证明式 1 和式 2 .

梯度

由定义 1 的第二条，三维欧几里得空间中的梯度，是 $\Omega^0(\mathbb{R}^3) \to \Omega^1(\mathbb{R}^3)$ 的外导数．

自此，三维空间中的梯度、旋度和散度都可以统一为一个概念了：外导数．值得注意的是，矢量分析中的定理“梯度的旋度为零”和“旋度的散度为零”，可以统一理解为外导数的**幂零性**：$\mathrm{d}^2 = 0$．

三维空间的特殊性

尝试求解下列代数方程：

$$\frac{n(n-1)}{2} = n \tag{3}$$

答案是 0 和 3，对吧？非零解只有 3，这就是三维空间的特殊之处，也是三维向量分析如此丰富的原因．只有在三维空间中，我们才能用以上外导数的概念来导出散度、旋度等概念．

式 3 究竟是什么呢？n 代表的是一个线性空间 V 的维度，而 $n(n-1)/2$ 代表的是 $\bigwedge^2 V$ 的维度．式 3 想求解的是，什么情况下 V 会和 $\bigwedge^2 V$ 同构，而“非零解只有 3”意味着唯一的情况就是，“V 是一个三维空间”．

向量叉乘的“右手定则”，实际上就是一个同构 $f : \bigwedge^2 V \to V$，其中 V 是 M 上各点的微分形式空间，或者说“余切向量空间”，而 $f(\mathrm{d}y \wedge \mathrm{d}z) = \mathrm{d}x, f(\mathrm{d}z \wedge \mathrm{d}x) = \mathrm{d}y, f(\mathrm{d}x \wedge \mathrm{d}y) = \mathrm{d}z$．我们当然可以定义其它的同构，比如 $f(\mathrm{d}y \wedge \mathrm{d}z) = -\mathrm{d}x, f(\mathrm{d}z \wedge \mathrm{d}x) = -\mathrm{d}y, f(\mathrm{d}x \wedge \mathrm{d}y) = -\mathrm{d}z$ 的同构，此时导出的向量叉乘遵循的就是“左手定则”了．

麦克斯韦方程组的外微分形式：前两个方程

未完成：百科中尚未创建“规范场论”部分．使用引用 [25].

在 $\mathbb{R}^3$ 中考虑电磁场，三个空间轴分别为 x, y, z 轴．

考虑麦克斯韦方程组中的两个方程，$\nabla \cdot \mathbf{B} = 0$ 和 $\nabla \times \mathbf{E} = \partial_t \mathbf{B}$．为了尝试用外代数来表达这两个式子，我们就要把 $\mathbf{B}$ 表示成一个 2-形式 $B = B_z \mathrm{d}x \wedge \mathrm{d}y + B_x \mathrm{d}y \wedge \mathrm{d}z + B_y \mathrm{d}z \wedge \mathrm{d}x$，把 $\mathbf{E}$ 表示成一个 1-形式 $E = E_x \mathrm{d}x + E_y \mathrm{d}y + E_z \mathrm{d}z$，这样以上两个方程的左边就都可以写成外导数的形式，从而有：

$$\mathrm{d}B = 0 \tag{4}$$

和

$$\mathrm{d}E = \partial_t B \tag{5}$$

其中 $\mathbb{R}^4$ 可以写成三维空间和一维时间的乘积：$\mathbb{R}^3 \times \mathbb{R}$．这个四维欧几里得空间中的时间轴记为 x^0 轴，空间轴则记为 x^1, x^2, x^3 轴．要注意在这种表示下，$\partial_t B$ 就成了 $\partial_0 B$．

此时，如果我们用一个统一的 2-形式 $F = B + E \wedge \mathrm{d}x^0$ 来表示电磁场[3]，那么式 4 和式 5 就可以统一用一个式子来表达：

$$\mathrm{d}F = 0 \tag{6}$$

证明过程此处略去，留给规范场论部分中相关章节来处理．麦克斯韦方程组的剩下两个方程的表示，参见**霍奇星算子**[608] 词条．

霍奇星算子

预备知识 外导数[604]

未完成：未完成：需添加更多讨论以及麦克斯韦方程组的表示．

星算子的定义

考虑 n 维线性空间 V 上的外代数 $\bigwedge V$，我们注意到各阶的外积空间具有明显的对称性：$\dim \bigwedge^k V = C_n^k = C_n^{n-k} = \dim \bigwedge^{n-k} V$．这意味着这样的一对空间之间存在线性同构，我们使用星算子 $\star$ 来描述这一同构．

星算子是一个映射，把一个 $\bigwedge^k V$ 中的元素 ω 映射为一个 $\bigwedge^{n-k} V$ 中的元素 $\star\omega$．为了方便描述星算子的定义，我们先引入一些新的表示方法．

选定 V 的基 $\{\mathbf{e}_i\}$，那么任意 $\omega \in \bigwedge^k V$ 都可以表示为各 $\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \wedge \cdots \wedge \mathbf{e}_{i_k}$ 的线性组合，因此我们只需要描述 $\star \mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \wedge \cdots \wedge \mathbf{e}_{i_k}$ 即可定义星算子．

为了方便，我们只考虑 $\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \wedge \cdots \wedge \mathbf{e}_{i_k}$ 中各 $i_{r+1} > i_r$ 的情况，也就是下标顺序排列的情况[4]．规定了只考虑下标顺序排列的情况后，就可以暂时不管顺序的问题，把 $\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \wedge \cdots \wedge \mathbf{e}_{i_k}$ 表示为集合 $\{i_1, i_2, \cdots, i_k\}$．利用这个表达，我们就可以定义 $\star\{i_1, i_2, \cdots i_k\} = \{1, 2, \cdots, n\} - \{i_1, i_2, \cdots, i_k\}$．这样，我们就定义出了 $\bigwedge V$ 中各基向量的星算子了．再加上一条“星算子是线性的”，即 $\star(\sum a_i \omega_i) = \sum a_i \star \omega_i$，就得到星算子的完整定义了：

[3] 电场外积一个 $\mathrm{d}x^0$ 是为了凑成合适的 2-形式．

[4] 乱序排列的情况无非两种，奇排列和偶排列，根据外代数的定义，前者加上负号即可，后者和顺序排列是相等的．

定义 1　霍奇星算子

定义域 $\mathbb{F}$ 上 n 维线性空间 V 上的外代数 $\bigwedge V$ 的自同构映射 $\star : \bigwedge V \to \bigwedge V$，满足：

1. 线性性：对于任意 $a_i \in \mathbb{F}$，$\omega_i \in \bigwedge V$，有 $\star(\sum a_i\omega_i) = \sum a_i \star \omega_i$；
2. 如果把下标顺序排列的 $\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \wedge \cdots \wedge \mathbf{e}_{i_k}$ 表示为集合 $\{i_1, i_2, \cdots, i_k\}$，那么 $\star\{i_1, i_2, \cdots i_k\} = \{1, 2, \cdots, n\} - \{i_1, i_2, \cdots, i_k\}$.

称这一同构为霍奇星算子（**Hodge star operator**）.

麦克斯韦方程组的外微分形式：后两个方程 [25]

星算子描述的是外积空间中一个非常显眼的对偶，即 $\bigwedge^k V$ 和 $\bigwedge^{n-k} V$ 之间的同构.

第五章

复半单李代数

李代数

预备知识 域上的代数[590]，爱因斯坦求和约定[498]

概念

李代数是对域上的代数[590]进行的一种推广．域上的代数是指域上的线性空间配合了一个矢量乘法，使得这个线性空间在矢量乘法下也能构成一个环．李代数也是域上线性空间配合了一个矢量乘法，这个矢量乘法和构成环的乘法几乎一样，但是有一点显著不同：将环乘法的结合律替代为 Jacobi 结合性．

李代数在现代理论物理的时空研究中应用广泛，常用于描述光滑张量场之间的作用——要注意的是，此时我们把一个光滑张量场本身看成一个向量．

定义 1　李代数

给定域 $\mathbb{F}$ 上的一个线性空间 V．在 V 上定义一个“乘法”运算：对于任意 $\mathbf{v}_1,\mathbf{v}_2\in V$，将它们的运算结果记为 $[\mathbf{v}_1,\mathbf{v}_2]$．称这个结构 $(V,[*,*])$ 为一个李代数，如果它满足以下性质：

- 封闭性对于任意 $\mathbf{v}_1,\mathbf{v}_2\in V$,$[\mathbf{v}_1,\mathbf{v}_2]\in V$.
- 双线性性对于任意 $k_i,c_i\in\mathbb{F}$ 和任意 $\mathbf{v}_i,\mathbf{u}_i\in V$,都有 $[k_1\mathbf{v}_1+k_2\mathbf{v}_2,c_1\mathbf{u}_1,c_2\mathbf{u}_2]=k_1c_1[\mathbf{v}_1,\mathbf{u}_1]+k_1c_2[\mathbf{v}_1,\mathbf{u}_2]+k_2c_1[\mathbf{v}_2,\mathbf{u}_1]+k_2c_2[\mathbf{v}_2,\mathbf{u}_2]$.
- 反对称性对于任意 $\mathbf{v},\mathbf{u}\in V$，有 $[\mathbf{v},\mathbf{u}]=-[\mathbf{u},\mathbf{v}]$.
- **Jacobi** 结合性对于任意 $\mathbf{x},\mathbf{y},\mathbf{z}\in V$,有 $[\mathbf{x},([\mathbf{y},\mathbf{z}])]+[\mathbf{z},([\mathbf{x},\mathbf{y}])]+[\mathbf{y},([\mathbf{z},\mathbf{x}])]=0$.

由于这个乘法用括号 $[*,*]$ 表示，因此有时也称之为“括积”.

李代数是对李群的性质进行进一步的抽象，是对李群的一些核心性质的代数刻画．

李代数定义中的双线性性可以通过反对称性和单对称性[1]推出，所以和本书中群的定义一样，是有冗余的定义．如果用爱因斯坦求和约定[498] 来表达，双线性性还可以写为 $[a_i\mathbf{v}_i, b_j\mathbf{u}_j] = a_i b_j[\mathbf{v}_i, \mathbf{u}_j]$.

定义 2　交换李代数

设 $\mathfrak{g}$ 为域 $\mathbb{F}$ 上的 n 维李代数，那么如果对于任意的 $\mathbf{x}, \mathbf{y} \in \mathfrak{g}$，都有 $[\mathbf{x}, \mathbf{y}] = 0$，则称这是一个交换李代数（**commutative Lie algebra**）或阿贝尔李代数（**abelian Lie algebra**）.

交换李代数之所以要求括积为零，是因为反对称性和交换性结合必然得到括积为零.

李代数和结合代数的关系极为紧密．事实上，每个结合代数都可以唯一地导出一个李代数，这由以下定理决定：

定理 1　结合代数导出李代数

设 $\mathfrak{A}$ 是域 $\mathbb{F}$ 上的一个结合代数，即 $\mathfrak{A}$ 是 $\mathbb{F}$ 上的线性空间，且定义了一个乘法 $\times$ 使得它还成为一个环，那么如果定义 $\forall \mathbf{x}, \mathbf{y} \in \mathfrak{A}, [\mathbf{x}, \mathbf{y}] = \mathbf{x} \times \mathbf{y} - \mathbf{y} \times \mathbf{x}$，则这个括号 $[*, *]$ 配上 $\mathfrak{A}$ 构成一个李代数.

证明：

封闭性是显然的，因为 $\times$ 是封闭的.

由这个括积的定义，反对称性也是显然的，因为 $\mathbf{x} \times \mathbf{y} - \mathbf{y} \times \mathbf{x} = -(\mathbf{y} \times \mathbf{x} - \mathbf{x} \times \mathbf{y})$ 恒成立.

由反对称性，我们只需要证明单线性性即可：$\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathfrak{A}$ 和 $a, b \in \mathbb{F}$，有：

最后是 Jacobi 结合性：

$$[\mathbf{x}, a\mathbf{y} + b\mathbf{z}] = a\mathbf{xy} + b\mathbf{xz} - a\mathbf{yx} - b\mathbf{zx} = a\mathbf{xy} + b\mathbf{xz} - a\mathbf{yx} - b\mathbf{zx} \tag{1}$$

这里省略了乘法符号，以显示紧凑.

由式 1，轮换 $\mathbf{x}, \mathbf{y}$ 和 $\mathbf{z}$ 的顺序后把得到的三个式子相加，即可以得到 Jacobi 结合性.

证毕.

[1] 即对于参与运算的双方中某一方线性就可以，比如对后面的线性：$[\mathbf{v}, a\mathbf{x} + b\mathbf{y}] = a[\mathbf{v}, \mathbf{x}] + b[\mathbf{v}, \mathbf{y}]$.

知道了每个结合代数可以对应一个李代数后，我们自然会好奇，每个李代数是不是也能对应一个结合代数？答案是肯定的，见[2].

例子

结合代数对应李代数这一事实，使得我们很容易联想到一个常见的结合代数：矩阵代数．在本文中，将域 $\mathbb{F}$ 上的 n 阶可逆方阵的集合记为 $\mathrm{gl}(n,\mathbb{F})$，那么这个集合自然构成 $\mathbb{F}$ 上的一个线性空间（以矩阵加法为向量加法），而矩阵乘法则使之构成一个环，因此这是一个结合代数．

例 1　一般线性李代数

域 $\mathbb{F}$ 上的 $\mathrm{gl}(n,\mathbb{F})$ 是一个结合代数．由这个结合代数可以导出李代数，其中对于矩阵 $\mathbf{A}$ 和 $\mathbf{B}$，括积的定义为 $[\mathbf{A},\mathbf{B}]=\mathbf{AB}-\mathbf{BA}$．该李代数被称为一般线性李代数，名称和一般线性群 $\mathrm{gl}(n,\mathbb{F})$ 对应．

例 2　特殊线性李代数

对于域 $\mathbb{F}$ 和正整数 n，记 $\mathrm{sl}(n,\mathbb{F})$ 为 $\mathrm{gl}(n,\mathbb{F})$ 中迹为 0 的矩阵构成的结合代数，则它可以导出一个特殊线性李代数，名称也和特殊线性群 $\mathrm{sl}(n,\mathbb{F})$ 对应．

例 3　三维向量叉积

域 $\mathbb{R}$ 上的三维线性空间 $\mathbb{R}^3$ 中，将括积定义为叉积：$\forall \mathbf{v},\mathbf{u}\in\mathbb{R}^3$，有 $[\mathbf{v},\mathbf{u}]=\mathbf{v}\times\mathbf{u}$．那么这个线性空间配上叉积可以得到一个李代数．

结构常数

预备知识 张量的坐标变换[501]

李代数的括积的作用是把两个向量映射为一个向量，而且还要求具有双线性性，因此括积实际上是一个三阶张量．如果设 $\mathfrak{g}$ 为域 $\mathbb{F}$ 上的 n 维李代数，而 $\{\mathbf{x}^i\}$ 为它的一组基，那么对于任意的 $\mathbf{x}^i$ 和 $\mathbf{x}^j$，存在 $C_k^{ij}\in\mathbb{F}$，使得 $[\mathbf{x}^i,\mathbf{x}^j]=C_k^{ij}\mathbf{x}^k$．这样的 C_k^{ij} 就是括积张量的三维矩阵表示．

显然，括积张量的三维矩阵表示依赖于基的选取，和所有其它张量一样．当选定基以后，所得到的矩阵称为 $\mathfrak{g}$ 关于基 $\{\mathbf{x}_i\}$ 的**结构常数**．

[2]尚未创作到相关词条

定理 2

设 $\mathfrak{g}$ 为域 $\mathbb{F}$ 上的 n 维李代数，它在基 $\{\mathbf{x}^i\}$ 下的结构常数为 C_k^{ij}，那么有：

- $C_k^{ij} = -C_{ji}^k$.
- $C_{ij}^l C_{lk}^m + C_{jk}^l C_{li}^m + C_{ki}^l C_{lj}^m = 0$[3].

定理的证明留给读者，这是直接从李代数的反对称性和 Jacobi 结合性得来的.

结构常数唯一地对应李代数，也就是说，我们也可以任取一个线性空间，然后通过定义满足定理 2 的结构常数来定义括积，从而得到一个李代数.

定理 3　结构常数的变换

设 $\mathfrak{g}$ 为域 $\mathbb{F}$ 上的 n 维李代数，它在基 $\{\mathbf{x}^i\}$ 下的结构常数为 C_k^{ij}，在基 $\{\mathbf{y}^i\}$ 下的结构常数为 D_k^{ij}，而且有过渡矩阵[414] a_j^i 使得 $\mathbf{y}^i = a_j^i \mathbf{x}^j$，那么有变换法则：

$$D_k^{ij} a_j^k = a_m^i a_n^j C_j^{mn} \tag{2}$$

证明：

$$[\mathbf{y}^i, \mathbf{y}^j] = D_k^{ij} \mathbf{y}^k = D_k^{ij} a_j^k \mathbf{x}^j \tag{3}$$

$$[\mathbf{y}^i, \mathbf{y}^j] = [a_m^i \mathbf{x}^m, a_n^j \mathbf{x}^n] = a_m^i a_n^j [\mathbf{x}^m, \mathbf{x}^n] = a_m^i a_n^j C_j^{mn} \mathbf{x}^j \tag{4}$$

比较式 3 和式 4，消去 $\mathbf{x}^j$ 即可得 $D_k^{ij} a_j^k = a_m^i a_n^j C_j^{mn}$.

证毕.

回顾张量的坐标变换[501] 词条，我们发现这就是三阶张量的坐标变换式.

李代数的子代数、理想与商代数

预备知识　李代数[611]

[3]注意使用爱因斯坦求和约定.

子代数与理想

和抽象代数中的子群、子环等类比，李代数也可以有次级结构，即子代数.

设 $\mathfrak{m},\mathfrak{n}$ 是李代数 $\mathfrak{g}$ 的非空子集，定义子集间的运算为 $\mathfrak{m}+\mathfrak{n}=\{M+N|M\in\mathfrak{m},N\in\mathfrak{n}\}$，以及 $[\mathfrak{m},\mathfrak{n}]=\{[M,N]|M\in\mathfrak{m},N\in\mathfrak{n}\}$. 那么如果 $\mathfrak{m},\mathfrak{n}$ 和 $\mathfrak{p}$ 都是 $\mathfrak{g}$ 作为线性空间的子空间，我们容易证明以下性质：

- $[\mathfrak{m}+\mathfrak{n},\mathfrak{p}]\subseteq[\mathfrak{m},\mathfrak{p}]+[\mathfrak{n},\mathfrak{p}]$；
- $[\mathfrak{m},\mathfrak{n}]=[\mathfrak{n},\mathfrak{m}]$；
- $[\mathfrak{m},[\mathfrak{n},\mathfrak{p}]]\subseteq[\mathfrak{n},[\mathfrak{m},\mathfrak{p}]]+[\mathfrak{p},[\mathfrak{n},\mathfrak{m}]]$.

定义 1　李代数的子代数

若 $\mathfrak{g}$ 是李代数，$\mathfrak{h}$ 是它作为线性空间的子空间，且有 $[\mathfrak{h},\mathfrak{h}]\subseteq\mathfrak{h}$，那么称 $\mathfrak{h}$ 是 $\mathfrak{g}$ 的子代数（**sub (Lie) algebra**）.

尽管李代数并不成环，但是我们也可以仿照环的理想的定义，用吸收律来定义李代数的理想：

定义 2　李代数的理想

若 $\mathfrak{g}$ 是李代数，$\mathfrak{h}$ 是它作为线性空间的子空间，且有 $[\mathfrak{g},\mathfrak{h}]=\mathfrak{h}$，那么称 $\mathfrak{h}$ 是 $\mathfrak{g}$ 的理想（**ideal**）.

显然，$\mathfrak{g}$ 和 $\{0\}$ 都是 $\mathfrak{g}$ 的理想，称为平凡理想（**trivial ideal**）.

我们通常把定义中 $[\mathfrak{g},\mathfrak{h}]=\mathfrak{h}$ 这一条，称为“吸收性”或者说“吸收律”，方便记忆. 从定义可以看到，李代数的理想必为其子代数. 进一步，理想和子代数还满足以下性质：

- 如果 $\mathfrak{h}_i$ 是子代数，那么 $\mathfrak{h}_1\cap\mathfrak{h}_2$ 也是子代数.
- 如果 $\mathfrak{h}_1$ 是子代数而 $\mathfrak{h}_2$ 是理想，那么 $\mathfrak{h}_1+\mathfrak{h}_2$ 是子代数.
- 如果 $\mathfrak{h}_i$ 是理想，那么 $\mathfrak{h}_1+\mathfrak{h}_2$、$\mathfrak{h}_1\cap\mathfrak{h}_2$ 和 $[\mathfrak{h}_1,\mathfrak{h}_2]$ 都是理想，并且有如下包含关系：

$$[\mathfrak{h}_1,\mathfrak{h}_2]\subseteq\mathfrak{h}_1\cap\mathfrak{h}_2\subseteq\mathfrak{h}_i\subseteq\mathfrak{h}_1+\mathfrak{h}_2 \tag{1}$$

证明是很简单的，留作练习．注意最后的包含关系里的符号区别，其中一共三个子集符号．

例 1　理想的例子

域 $\mathbb{F}$ 上的 n 阶可逆矩阵的集合 $\mathrm{gl}(n,\mathbb{F})$，构成一个李代数．考虑到对于任意矩阵 $A,B\in\mathrm{gl}(n,\mathbb{F})$，我们有 $\mathrm{trace}(AB)=\mathrm{trace}(A)\,\mathrm{trace}(B)=\mathrm{trace}(B)\,\mathrm{trace}(A)=\mathrm{trace}(BA)$，因此必有 $\mathrm{trace}([A,B])=0$．这就提示我们去考虑迹为 0 的全体 n 阶方阵的集合，记为 $\mathfrak{t}$，容易验证它就是 $\mathrm{gl}(n,\mathbb{F})$ 的一个理想．

半单李代数

和**单群**[553] 类似，我们可以定义单李代数：

定义 3　单李代数

给定李代数 $\mathfrak{g}$，如果它没有非平凡理想，即除了 $\{0\}$ 和 $\mathfrak{g}$ 自身外没有别的理想，那么我们称它为一个单李代数（**simple Lie algebra**）．

类似地，如果加一个限制，还有半单李代数的概念：

定义 4　半单李代数

给定李代数 $\mathfrak{g}$，如果它没有非平凡交换理想，那么我们称它为一个半单李代数（**semi-simple Lie algebra**）．

第五部分

拓扑学

第一章

点集拓扑

拓扑空间

预备知识　集合[92]

拓扑空间（topological space）是能够定义连续性，连通性，收敛等性质的最一般化的数学空间．度量空间和流形等都是拓扑空间的例子．

拓扑

从实函数的连续性[156]中我们知道开集是讨论连续性的基础，所以拓扑学首先要定义什么是开集．我们提取了开集最重要的特征，然后用这些特征来定义开集．

定义 1　拓扑

对于任意给定的集合 X，如果我们按照一定规则将它的子集划分为开集（**open set**）和其它子集，那么所有开集的集合就叫做集合 X 的一个拓扑 $\mathcal{T}$．这个规则是：

1. 空集 $\varnothing$ 和 X 本身必须是开子集
2. 有限交封闭：有限个开子集的交集为开子集
3. 任意并封闭：任意个开子集的并集为开子集

X 的全体子集构成一个族，即 X 的幂集 2^X，显然有 $\mathcal{T} \subset 2^X$．同一个集合上的拓扑一般不止一种．

每一个开集 U 的补集 $U^C = X - U$，被称为一个**闭集（closed set）**．2^X 的元素中，有些是开集，有些是闭集，有的既开又闭，也有的既不开也不闭．

例 1　凝聚拓扑和离散拓扑

对给定的集合 X，若只定义 $\varnothing$ 和 X 本身为开集，X 的其他子集为非开，则这个拓扑称为凝聚拓扑或者平凡拓扑（**trivial topology**），这是符合定义 1 的元素最少的拓扑．

相对地，若令 X 的任意子集都为开集，则得到离散拓扑（**discrete topology**），这是元素最多的拓扑．

例 2　Sierpinski 拓扑

令集合为 $X=\{0,1\}$，赋予拓扑 $\mathcal{T}=\{\varnothing,\{0\},X\}$，则我们得到了一个 Sierpinski 空间.

光有正例也并不容易建立对“拓扑”这一概念的直观，我们再看一个简单的反例：

例 3　反例

令集合为 $X=\{0,1,2\}$，则集族 $\{\varnothing,\{0\},\{1\},\{2\},X\}$ 并不是一个拓扑，因为 $\{0\}\cup\{1\}=\{0,1\}$ 并不在这个集族里，从而不满足任意并封闭条件.

定义 2　拓扑基

对于给定的拓扑空间 $(X,\mathcal{T})$，如果 $\mathcal{B}$ 是 $\mathcal{T}$ 的一个子族，而且使得拓扑 $\mathcal{T}$ 中的任何开集都可以表示为 $\mathcal{B}$ 中若干开集（个数是任意的，可以是非负整数也可以是任意无限数）的并，那么称 $\mathcal{B}$ 是 $\mathcal{T}$ 的一个拓扑基（**topological basis**）.

拓扑基的定义中，要求 $\mathcal{B}$ 中的元素都是 X 的开集，而 X 中一切开集都可以由这些元素取并集而得到. 事实上，也可以不用强调 $\mathcal{B}$ 中的元素都是开集，而是简单地说一个集合 U 是开集当且仅当 U 是 $\mathcal{B}$ 中若干元素的并，这其中当然也包括单个元素，所以这种说法蕴含了“$\mathcal{B}$ 中的元素都是开集”.

任意拿一些 X 的子集来构成一个族 $\mathcal{A}$，这个族不一定是某个拓扑的拓扑基；也就是说，$\mathcal{A}$ 中任意多个元素的并集所构成的集合，并不一定是一个拓扑. 所以不是所有的 $\mathcal{A}$ 都可以成为某个拓扑的拓扑基. 如果 $\mathcal{A}$ 是某个拓扑的拓扑基，则简单称其为一个拓扑基[1].

例 4　度量拓扑

参见度量空间[785] 词条.

通常的欧几里得空间 $\mathbb{R}^n=\{(x_1,\cdots x_n)|x_i\in\mathbb{R}\}$，记 $\{(y_1,\cdots y_n)|y_i\in\mathbb{R},\sum_i(y_i-x_i)^2\leq r^2\}$ 为以 $(x_1,\cdots x_n)$ 为球心、$r\geqslant 0$ 为半径的开球（**open sphere**）. 那么全体开球的集合 $\mathcal{B}$ 是某个拓扑 $\mathcal{T}$ 的拓扑基，$\mathcal{T}$ 此时就是一个度量拓扑（**metric topology**）. 这个拓扑就是最常见的实空间上的拓扑，有很直观

[1] 2^X 的子族 $\mathcal{A}$ 是一个拓扑基，当且仅当任意 $A_1,A_2\in\mathcal{A}$ 以及任意的 $x\in A_1\cap A_2$，则存在 $A_x\subseteq A_1\cap A_2$，使得 $x\in A_x$ 且 $A_x\in\mathcal{A}$. 这是拓扑基的判别法，不要求掌握，感兴趣的读者可以自行证明.

的几何意义.

特殊地，当 $N=1$ 时，$\mathbb{R}$ 上所有开区间的集合构成一个拓扑基，也就是说任意个开区间的并和有限个开区间的交都是开区间.

由开球生成的这个拓扑，还有一种定义方法：$A\subset\mathbb{R}^n$ 是一个开集，当且仅当对于任意的 $x\in A$，存在一个半径 r，使得 $\mathbb{R}^n$ 中到 x 距离小于 r 的所有点都在 A 内. 也就是说，开集的点都是内点.

证明所有点都是内点的集合必定可以表示为开球的并：$\mathbb{R}^n$ 中的集合 X 的内点的定义是必然存在以其为圆心的，由 X 的元素构成的开球，所以如果所有的点都是内点，则 X 必然是开球的并. 反证：如果 X 中有一点不属于任何由 X 的元素构成的开球，那么它就不是内点，所以 X 也不可能是一个开集.

定义 3　子拓扑

如果已经给定了一个拓扑 $(X,\mathcal{T})$，那么在 X 的一个子集 A 上可以继承一个拓扑 $\mathcal{T}|_A$，定义为：$\mathcal{T}|_A=\{U\cap A|U\in\mathcal{T}\}$. 那么 $\mathcal{T}|_A$ 称为 $\mathcal{T}$ 的子拓扑（**subtopology**）或者限制拓扑，$(A,\mathcal{T}|_A)$ 构成了 $(X,\mathcal{T})$ 的一个子拓扑空间（**subspace**）.

子拓扑的开集，就是原拓扑空间 X 的开集和子空间集合 A 的交.

例 5　度量空间的子空间

设 $(\mathbb{R}^2,\mathcal{T}_2)$ 是二维实度量空间，即 $x-y$ 平面；$(\mathbb{R},\mathcal{T}_1)$ 是一维实度量空间，即 x 轴. 那么 $\mathcal{T}_1$ 刚好是 $\mathcal{T}_2$ 的子拓扑.

小结和一点拓展

给定集合 X 上的一个拓扑 $\mathcal{T}$，就是 X 上所有开集的集合. 也就是说，一个拓扑是 X 的幂集[2]的子集：$\mathcal{T}\subset 2^X$. 给定拓扑，相当于规定了哪些子集是开集. 一个子集可以是开集，也可以是闭集；可以既开又闭，还可以既不开也不闭.

我们如上定义了拓扑基的概念，一个拓扑基 $\mathcal{B}$ 也是 2^X 的一个子集，但是它不一定成为拓扑，因为它不要求满足任意并封闭和有限交封闭.

事实上，任意给定 2^X 的一个子集 $\mathcal{S}$，即使它不是一个拓扑，我们也可以找到一个包含它的最小拓扑 $T(\mathcal{S})$，这个时候称 $\mathcal{S}$ 是 $T(\mathcal{S})$ 的一个**子基**（**sub-basis**）.

[2]幂集的概念请见集合[92] 中脚注.

可以证明，先把 $\mathcal{S}$ 的有限多个元素[3]拿出来计算出交集，把一切这样的有限交放在一起，构成一个新的集合 $U(\mathcal{S})$；然后再把 $U(\mathcal{S})$ 中的任意多元素拿出来，取并集，再构成一个新的集合 $I(U(\mathcal{S}))$. 这样先取全体有限交、再取全体任意并所得到的集合 $I(U(\mathcal{S}))$，就是 $T(\mathcal{S})$.

拓扑基也是子基，但是它特殊的地方在于，一个拓扑基 $\mathcal{B}$ 只需要进行一次任意并运算就可以得到拓扑了：$T(\mathcal{B}) = I(\mathcal{B})$，省去了考虑交集的麻烦.

对于任意的 $\mathcal{S} \subset 2^X$，判定它是否是（某个拓扑的）拓扑基的一个常用等价条件是：对于任意的 $S_1, S_2 \in \mathcal{S}$ 以及任意的 $x \in S_1 \cap S_2$，总能找到一个 S_3，使得 $x \in S_3 \subset S_1 \cap S_2$.

点集的内部、外部和边界

预备知识 拓扑空间[619]

拓扑空间定义了什么样的集合算开集，而开集的补集被称为闭集. 除此之外，还有一些集合是既不开也不闭的. 比如说，在 $\mathcal{R}$ 上定义的度量拓扑中，开区间都是开集，闭区间和孤立点都是闭集，但是半开半闭区间两者都不属于.

基本概念

定义 1

给定拓扑空间 $(X, \mathcal{T})$ 以及它的一个子集 A.

- 如果对于 $x \in X$，存在一个开集 $V \in \mathcal{T}$ 使得 $x \in V \subset A$，那么称这个 x 是 A 的一个内点（**interior point**）.
- 如果存在一个开集 $V \in \mathcal{T}$ 使得 $x \in V \subset A^C$，那么称这个 x 是 A 的一个外点（**exterior point**）.
- 如果一个 x 既不是 A 的外点也不是内点，那么称 x 是 A 的一个边界点（**boarder point**）.
- 全体内点构成的集合，称为 A 的内部（**interior**），记为 A°.

[3]注意，$\mathcal{S}$ 的元素本身是集合.

- 全体外点构成的集合，称为 A 的外部（**exterior**）.
- 全体边界点构成的集合，称为 A 的边界（**boarder**）.
- 如果包含点 $x \in X$ 的任何开集 U_x，都含有和 x 不相同的、A 的元素：$(U_x - \{x\}) \cap A \neq \varnothing$，那么我们说 x 是 A 的一个聚点或者极限点（**limit point**）.

外点就是“补集的内点”.

聚点的概念在数学分析和高等微积分里就会出现. 在微积分中，一个集合 $A \subset \mathbb{R}$ 的聚点 x，就是不管取多小的 $r > 0$ 作为半径，总有某个 A 中的点 x_r 落在 $(x - r, x + r)$ 里面. A 自己的点 $x \in A$ 当然是 A 的聚点，但是 $x \notin A$ 也可以是聚点. 比如说，区间 $(0, 1)$ 的全体聚点构成的集合，就是 $[0, 1]$.

有了这些定义，我们就可以更直观地理解开集和闭集了：

习题 1

给定 $(X, \mathcal{T})$ 以及它的一个子集 A. 那么，A 是开集当且仅当 A 的点都是内点；A 是闭集当且仅当 A 的聚点都在 A 里. 证明留做习题.

同样地，使用区间为例子来理解开集、闭集和内点、聚点的关系会非常直观.

由定义，对于任何拓扑空间的子集 A，外点都不是聚点，但边界点和内点都是聚点；内点都是 A 的点，但边界点不一定是. 这么一来，计算边界点的方法，就是求出所有聚点，再减去所有内点即可.

对于 A，它的聚点构成的集合被称为 A 的**导出集**（**induced set**）或者**导集**，在本书中记为[4] A'.

例 1

- 令 $A = \{\frac{1}{n} | n \in \mathbb{Z}^+\}$，那么 $A' = \{0\}$.
- 令 $A = (0, 1) \cup (1, 2)$，那么 $A' = [0, 2]$

[4]有的教材中，A' 代表的是补集，但在本书中我们使用 A^C 表示补集了，于是 A' 可以腾出来表示导集.

习题 2

证明：任何集合的导集都是闭集.

在数学分析中，我们会使用包含点 x 的任意长度的开区间，来描述 x 的“附近”. 这一概念在拓扑学中被拓展为“邻域”：

定义 2　邻域

给定拓扑空间 X 和 $x \in X$，则任意开集 $U_x \ni x$，被称为 x 的一个邻域（**neighborhood**）.

注意，这里定义的邻域必须是开集. 在紧致性[628] 一节中，我们会提到一个容易混淆的概念，“紧邻域”，它并不一定是开集，因此不一定是邻域. 紧邻域的名字是有些混淆，但是没有办法，约定俗成这么叫了[5].

闭包

给定任意拓扑空间 X 的任意子集 A，那么 A° 是 A 所包含的最大的开集. 对称地，包含 A 的最小的闭集被称为 A 的**闭包**（**closure**），记为 $\bar{A}$. 闭包很容易得到：

定理 1

给定任意拓扑空间 X 的任意子集 A，则有 $\bar{A} = A \cup A'$.

推论 1

考虑到 $A' - A^\circ$ 是 A 的边界，故 $\bar{A}$ 等于 A 和 A 的边界的并.

以区间作为例子：如果 $A = (0,1)$，那么 $A' = [0,1]$，$\bar{A} = [0,1]$.

以圆盘作为例子：如果 $D = \{(x,y) \in \mathbb{R}^2 | x^2 + y^2 < 1\}$，那么 $D' = \bar{D} = \{(x,y) \in \mathbb{R}^2 | x^2 + y^2 \leq 1\}$. 尝试在 $x^2 + y^2 = 1$ 的圆盘上添加部分点进 D，得到的 D 还是具有相同的闭包. 这是因为这些添加的点都是在最初的 D 的边界上的.

稀疏和稠密

[5]事实上，一种常见的表达方式是在“邻域”两字前加其它修饰，这样表达的往往不是邻域（即不是开集），而是包含某个邻域的集合.“紧邻域”表示包含一个邻域的紧集，“闭邻域”表示包含一个邻域的闭集，它们往往都不是开集，从而不是邻域.

预备知识 集合的极限[783]

给定任意拓扑空间 X 的任意子集 A，如果 $A^\circ = \varnothing$，即 A 没有内点，那么称 A 是**稀疏（sparse）**的；如果 $\bar{A} = B \subset X$，也就是说 B 是 A 和 A 的边界的并，那么称 A 在 B 中是**稠密（dense）**的.

无论取什么开集，它都不可能被稀疏的集合完全包含，总有遗漏的点，所以有了“稀疏”的名称. 而当 A 在 B 中稠密时，不可能找一个开集来圈出 B 中完全和 A 不相交的一块区域，所以有了“稠密”的名称.

全体有理数的集合 $\mathbb{Q}$ 在 $\mathbb{R}$ 中既稠密，又稀疏.

例 2　Cantor 集

令 $A_0 = [0,1]$，也就是说，A_0 是一段长度为 1 的线段. 把这条线段的中间挖去 $1/3$，得到 $A_1 = [0,1/3] \cup [2/3,1]$. 现在 A_1 有了两段长为 $1/3$ 的线段，同样地，分别挖去它们中间的 $1/3$，得到 A_2. 以此类推，A_n 有 2^n 条长度为 $1/3^n$ 的线段，把每条线段的中间挖掉 $1/3$，就可以得到 A_{n+1}. 取这一列集合的极限 $C = \lim_{n\to\infty} A_n$，所得到的 C 被称为 **Cantor** 集，这是一个非常重要的集合，常用来作特例来阐释各种集合的性质.

这是一个稀疏集，同时又是一个完备集（**perfect set**）[6]另外，在测度论中，这个集合的测度是 0.

连续映射和同胚

预备知识 映射[98]，拓扑空间[619]，函数的连续性[156]

给定两个拓扑空间 $(X,\mathcal{T}_X)$ 和 $(Y,\mathcal{T}_Y)$，定义一个映射 $f: X \to Y$. 在集合论意义上，这个映射 f 可以有任何可能的形式，但是在拓扑学意义上，我们只研究满足特定条件的一类映射，称为连续映射. 那么什么是连续映射呢？

[6]导集还是自身的集合，称为完备集或者完美集.

连续映射

定义 1　连续映射[7]

给定两拓扑空间 $(X,\mathcal{T}_X)$ 和 $(Y,\mathcal{T}_Y)$. 映射 $f:X\to Y$ 在某一点 $x_o\in X$ 上连续，当且仅当对于任意 Y 中的开集 $U\ni f(x_0)$，存在 X 中的开集 $V\ni x_0$，使得 $f(V)\subseteq U$.

如果 f 在 X 中的任意一点上都连续，那么称 f 是一个连续映射（**continuous mapping**）.

习题 1　连续映射的等价定义

给定两拓扑空间 X，Y，那么 $f:X\to Y$ 是连续映射，当且仅当对于任意的 $U\in\mathcal{T}_Y$，有 $f^{-1}(U)\in\mathcal{T}_X$. 即开集的逆映射还是开集.

证明这一点.

习题 1 中的等价定义，可以简单记为“开集的逆映射还是开集”，类似地，容易证明它等价于“闭集的逆映射还是闭集”. 这个定义和微积分中实函数的连续性定理 1 本质上是一样的. 在微积分和实变函数中，这个等价定义的用途似乎没那么多，但是在点集拓扑中要更为常用. 接下来讨论的同胚概念就是一个例子.

同胚

在代数学中我们提到了同构和同态的概念[555]. 在任何理论框架中，同构的两个对象都被看成是同一个对象，这就要求同构的两个对象在指定理论框架中有完全一样的行为. 拓扑学的框架和代数学不一样，因此同构的定义也不一样. 由于历史原因，拓扑学意义上的同构，被称为同胚.

定义 2　同胚

对于拓扑空间 X 和 Y，如果存在一个双射 $f:X\to Y$，使得对于任意的开集 $A\in\mathcal{T}_X$ 和 $B\in\mathcal{T}_Y$，都有 $f(A)\in\mathcal{T}_Y$ 和 $f^{-1}(B)\in\mathcal{T}_X$，那么我们就说 X 和 Y 是同胚（**homeomorphic**）的；这个双射 f 被称为一个同胚映射（**homeomorphic mapping**）或者简称为同胚（**homeomorphism**），记为 $X\approx Y$.

[7]这个定义是类比实函数中逐点连续的 $\epsilon-\delta$ 表达.

这个定义其实就是直接要求存在一个双射，使得两个空间的开集一一对应．这个定义很直观，因为拓扑学就是定义开集的理论，开集之间一一对应的话，在拓扑意义下就有完全一样的行为了；但实践中一般不可能一个一个地比较两个空间的开集，看它们是不是都互相配对了，所以我们就引入了以下等价的定义：

定理 1　同胚的等价定义

对于拓扑空间 X 和 Y，如果存在一个双射 $f : X \to Y$，使得 f 和 f^{-1} 都是连续映射，那么我们有 $X \cong Y$．

证明是很简单的，直接应用习题 1 的定义就可以．

例 1

考虑一维实数空间 $\mathbb{R}$，令所有开区间的集合为拓扑基，定义其上的拓扑．取一个开区间 $I = (0, 1)$，将 I 视为 $\mathbb{R}$ 的子空间．那么这两个空间是同胚的．同胚映射取 $f : I \to \mathbb{R}$，其中 $\forall x \in I, f(x) = \tan\left(\pi \cdot \left(x - \frac{1}{2}\right)\right)$ 即可．

例 2

类似例 1，$\mathbb{R}^2 \cong I \times I$．其中 $I \times I$ 是一个正方形区域，去除了边界．

例 3　反例

通常的度量空间 $\mathbb{R}^2$ 和一个球表面 S^2 不同胚．在 $\mathbb{R}^2$ 平面上挖去一个点得到的空间也和 $\mathbb{R}^2$ 不同胚．

我们还可能经常接触到和同胚类似的一个概念，称为“嵌入”．把拓扑空间 X 嵌入到空间 Y，就是说让 X 和 Y 的一个子空间同构，由于同构的空间都被看成是同一个，这就相当于 X 成为了 Y 的一部分．

定义 3　嵌入

设拓扑空间 X 和 Y．若存在连续映射 $f : X \to Y$，使得 $X \cong f(X)$，那么称 f 是一个嵌入（映射）（**embedding**），称 X 由 f 嵌入（**embed**）到 Y 中．

回忆一下映射的四种情况[98]，很容易发现，只要是单连续映射都可以成为嵌入映射．

同胚不变性

数学中至关重要的一种研究思想，就是关注**不变性（invariance）**．我们所研究的对象哪些地方是恒定不变的，标志了这个对象的本质．

比如说，一个 8×8 的国际象棋棋盘，可以用 1×2 的长方形木条铺满；如果把棋盘的左上角和右下角的格子去掉，还能不能铺满呢？答案是否定的，因为如果你考虑棋盘黑白格子的分布，那么每个木条肯定要覆盖一个黑色格子和一个白色格子；去掉的两个格子肯定是同色的，这样一来两种颜色的格子数量就不一样了，肯定不会被铺满．在这个例子中，被覆盖的黑白格子数量无论如何都会是相同的，利用这个不变性就可以很轻松地解答问题．

在物理学中也常常见到不变性的身影．牛顿力学中有伽利略不变性[??]，在任何惯性系下加速度都是一样的；但是由另一个不变性：光速不变原理，却可以导出不同于牛顿力学的狭义相对论．

在点集拓扑中，一个拓扑空间可以有各种各样的性质，比如单位区间 $I=(0,1)$ 具有长度 1，而 $\mathbb{R}$ 具有长度 ∞．但我们从例 1 中知道，这两个拓扑空间应该是同胚的，所以“长度”这一属性不是它们的本质属性．为了得到拓扑空间的本质属性，我们可以探索存在哪些**同胚不变性**，即对于任何同胚的空间来说都一样的性质．

最重要的四个同胚不变性，分别是：**紧致性**[628]，**连通性**[632]，**道路连通性**[634] 以及**分离性**[636]．对每一个性质，本书都单独开辟了一个词条，请点击性质名称后面的链接查看．

紧致性

预备知识 连续映射和同胚[625]

[8]有的时候，研究包含了无穷多个点的子集会涉及到很多并不是那么重要的性质和细节．有些子集的行为，看起来就像是一个点或者几个点，把它揉成一个或几个点也无关紧要，有一点类似于等价类划分把等价元素都揉成了同一个．可以被这样揉成少数几个点的子集，是用以下定义的“覆盖”来描述的．

[8]本文参考 Wikipedia 相关页面．

覆盖

给定任何集合 X 和它的一个子集 A，我们常常可以用一系列 X 的子集 X_i 来把 A 包括在这些子集的并集里，这种行为叫做**用 $\{X_i\}$ 来覆盖 A**．比如说，取集合 $\mathbb{R}$，那么我们可以用一系列子集 $\{[n,n+10)\}$ 来覆盖 $\mathbb{R}$ 本身，其中 n 取遍所有整数．当然，覆盖的方式不止一种．

定义 1　覆盖

给定集合 $A\subset X$．如果 X 的若干子集的集合 $\{X_i\}\subseteq 2^X$ 满足 $A\subseteq\bigcup X_i$，那么我们说集合 $\{X_i\}$ 覆盖了 A．称 $\{X_i\}$ 为 A 的一个覆盖（**cover**）．

当然，用 $\{[n,n+10)\}$ 来覆盖 $\mathbb{R}$ 有些浪费，我们只需要其中的一部分就足够覆盖 $\mathbb{R}$ 了，比如说只取 n 是偶数，甚至只取 n 是 9 的倍数的情况．由此我们可以引出**子覆盖**的概念．

定义 2　子覆盖

给定集合 $A\subseteq X$ 和它的一个覆盖 $\{X_i\}$．如果 $\{X_i\}$ 的某个子集 $\{Y_i\}$ 也能覆盖 A，那么显然 $\{Y_i\}$ 也是 A 的一个覆盖，称它为 A 对于 $\{X_i\}$ 的子覆盖．

在拓扑学中，我们常常关心拓扑空间中开子集的性质．因为根据子拓扑的定义 3 [619] 以及开集的有限交封闭性，开子集构成的拓扑空间中的开集，都是原空间中的开集．因此我们会特别关注用开子集来进行覆盖．

定义 3　开覆盖

给定拓扑空间 X 和它的一个子集 A．如果 A 有一个覆盖 $\{X_i\}$．如果各 X_i 是开集，那么我们称这个覆盖为 A 的开覆盖（**open cover**）．

紧子集

定义 4　紧子集

给定拓扑空间 X 和它的一个子集 A．如果不论取 A 的哪一个开覆盖 $\{X_i\}$，总存在一个有限子覆盖 $\{Y_i\}\subseteq\{X_i\}$[9]，那么称 A 是 X 的一个紧子集（**compact subset**），或者说 A 在 X 中是紧致（**compact**）的，简称 A 是紧的．

紧集就像若干点一样，任意开覆盖都有有限子覆盖．为了直观感受紧致性，

[9]即子覆盖 $\{Y_i\}$ 是有限集合．

我们可以考察通常的度量空间 $\mathbb{R}$．开区间 $(0,1)$ 不是紧致的，因为如果取覆盖是 $\{(1/n,1)\}$，那么这个覆盖的任何有限子集都不足以覆盖 $(0,1)$．但是，闭区间 $[0,1]$ 是紧致的，像 $\{(1/n,1)\}$ 这样的族[10]是没法覆盖闭区间的．事实上，在 $\mathbb{R}$ 中只有闭集是紧的，但直接证明非常麻烦，但使用**分离性**[636] 中的方法就可以方便地证明．

一般来说，讨论拓扑空间 X 中的子集 A 是不是紧致的，直接从 X 中找一切可能的开覆盖很麻烦，但是用拓扑基就可以大大简化开覆盖的种类．这就引出了如下定理：

定理 1

给定拓扑空间 X 和它的一个子集 A，任取 X 的一个拓扑基 $\mathcal{B}$．A 是紧子集，当且仅当从 $\mathcal{B}$ 得到的开覆盖都有有限子覆盖．

这个定理非常直白，因为 X 的开集都是 $\mathcal{B}$ 的元素取并得到的．有了这个定理，我们可以把一个开覆盖中的每个开集再拆分成更小的开集的并．还是用 $\mathbb{R}$ 度量空间举例：

例 1

实数集 $\mathbb{R}$ 有一个拓扑基 $\mathcal{B}=\{(a,b):a,b\in\mathbb{R}\}$，也就是说，$\mathcal{B}$ 是全体开区间的集合．对于集合 $(0,1)$，我们可以取开覆盖 $\{(-1,1/3)\cup(2/3,2),(1/4,3/4)\}$，这个开覆盖中包含了两个开集．但是这个覆盖也可以看成是三个开区间构成的集合：$\{(-1,1/3),(2/3,2),(1/4,3/4)\}$，而这三个开区间都来自 $\mathcal{B}$．

如果拓扑空间 X 是自己的紧子集，那么我们说这个空间是一个**紧空间**．紧空间中的闭集，都是紧的，如同以下定理所说．

定理 2

紧空间的闭子集都是紧的．

假设 X 是一个紧拓扑空间，而 $A\subseteq X$ 是 X 的一个闭子集．假设 $\{X_i\}_{i\in\mathbb{Z}^+}$[11] 是 A 的一个开覆盖，而因为 A 是闭集，故 $X-A$ 是开集．记 $X_0=X-A$，那么 $X_0\cup\{X_i\}_{i\in\mathbb{Z}^+}=\{X_i\}_{i=0,1,2,\cdots}$ 就是 X 的一个开覆盖．因为 X 本身是紧的，所以 $\{X_i\}_{i=0,1,2,\cdots}$ 必然有有限子覆盖 $\{Y_i\}$，其中 Y_i 只有有限多个，每个 Y_i 都是某

[10] 族（family）是指集合的集合，见词条[92]．

[11] 下角标表示 i 的取值范围是正整数

个 X_j. 那么 $\{Y_i\}$ 就是 A 的有限开覆盖. 当然了，我们希望证明的是 $\{X_i\}_{i\in\mathbb{Z}^+}$ 有有限子覆盖，只需要把 X_0 从 $\{Y_i\}$ 中剔除，那么剩下的就是 A 的有限开覆盖，并且还是 $\{X_i\}_{i\in\mathbb{Z}^+}$ 的子覆盖.

局部紧

有的时候，一个空间往往整体上不是紧的，但它总在部分看起来“好像”有紧性. 这就像地球表面是弯曲的，但我们局部看来它很接近一个不弯曲的平面. 为了说清楚什么叫做“部分”看起来是紧的，我们需要扩充一下邻域的概念：

定义 5　紧邻域

给定拓扑空间 X 和 $x\in X$. 如果存在紧子集 $A\subseteq X$ 使得 x 是 A 的内点，那么称 A 是 x 的一个紧邻域.

紧邻域不一定是开集，因此不一定是邻域. 与之类似的概念还有“闭领域”，其含义也是以 x 作为内点的闭集. 注意，不管是邻域、紧邻域还是闭邻域，它们的共性在于，x 必须是它们的内点，也就是说，必须有一个开集 O_x 使得 $x\in O_x$ 且 O_x 在相应“邻域”中.

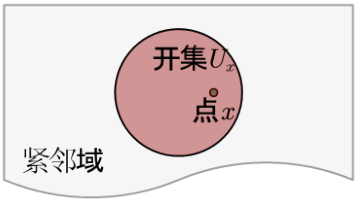


图 1: 点 x 和紧邻域的关系.

紧邻域就好像是包含了 x 的一块紧子集，如果每个 x 附近都有这么一个紧子集，我们就可以在“局部”研究空间的紧性. 因此我们有了以下“局部紧”的定义：

定义 6　局部紧空间

给定拓扑空间 X 和 $x\in X$. 如果对于某个 $x\in X$ 及其任意邻域 U_x，都存在一个 x 的紧邻域 $C\subseteq U_x$，那么称空间 X 在 x 处局部紧（**locally compact**）. 如果 X 在每一个点都局部紧，称这个空间是局部紧空间.

许多紧空间之间的关系，都可以直接加上“局部”二字以后保持成立.

连通性

预备知识 连续映射和同胚[625]

连通性的概念

什么叫连通性？直观来说，在平面 $\mathbb{R}^2$ 上画两个不相交也不相切的圆，那么这两个圆所包含的区域就是不连通的. 不连通的各部分显然是可以被“孤立”出来的，也就是说，如果两圆不相交也不相切，就一定能各自找到一个开集，让这两个圆分别被一个开集包含，而这两个开集还互不相交. 这就是定义连通性的方法.

我们把这两个圆包含的区域单独拿出来，构造一个子空间. 在这个子空间中，两个圆各自有一个特点：它们既是开集也是闭集. 如果将两个圆分别记为 C_1 和 C_2，子空间记为 $A = C_1 \cup C_2$，包含它们的开集分别是 $U_1 \supseteq C_1$ 和 $U_2 \supseteq C_2$，且 $U_1 \cap U_2 = \varnothing$，那么由子拓扑的定义，$C_1 = C_1 \cap U_1$ 是子空间的开集，同理 C_2 也是子空间的开集；然而在子空间中还有 $C_1 = A - C_2$，所以 C_1 还应该是一个闭集，同理 C_2 也是一个闭集.

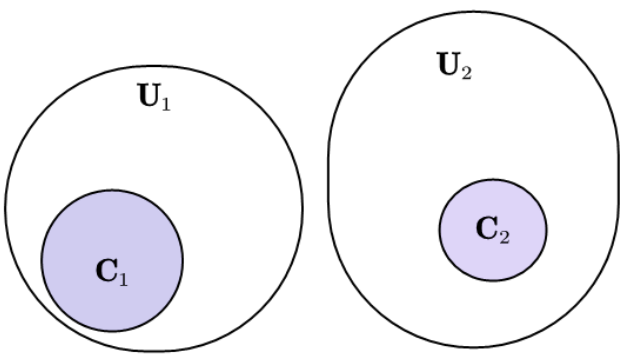


图 1: C_1，C_2，U_1 和 U_2 的示意图

如果你选择的是一个连通的空间，比如说一个圆所包含的区域构造的子空间，那么这种“既开又闭”的情况是不会存在的. 因此，我们可以根据既开又闭的性质来定义连通性.

定义 1　孤立分支

给定拓扑空间 X 和它的一个非空真子集 A. 如果 A 在 X 中既是开集也是闭集，那么称 A 是 X 的一个孤立分支.

注意，定义孤立分支的时候特别强调了 A 是一个非空真子集，这样就把 $\varnothing$ 和 X 本身排除在外了，因为按照定义，它们俩必须是既开又闭的. 在上图中的 C_1 和 C_2 就是两个孤立分支.

有了孤立分支的概念，就能直接引入连通性的概念了：

定义 2　连通性

拓扑空间 X 是连通（**connected**）的，当且仅当 X 中没有孤立分支.

连通性也可以用别的方式定义.

习题 1　连通性的等价定义

给定拓扑空间 X，证明“X 是连通的”和以下命题等价：

- X 不可以表示为两个非空不相交的开集的并；
- X 不可以表示为两个非空不相交的闭集的并；
- 如果 B 是只有两个元素的离散拓扑，那么不存在 X 到 B 的满射.

对于任何拓扑空间，我们可以讨论这个空间本身是不是连通的，也可以研究各个子集构成的子空间是不是连通的. 如果一个子集作为拓扑空间是连通的，我们就说这是一个**连通子集（connected subset）**.

例 1　$\sin\frac{1}{x}$

在 $(0,\infty)$ 上定义函数 $\sin\frac{1}{x}$，取函数图像上的所有点的集合作为 $\mathbb{R}^2$ 的子集：$S=\{(x,y)|x>0, y=\sin\frac{1}{x}\}$. 取 S 的闭包 $\bar{S}$，那么 S 和 $\bar{S}$ 都是连通的.

连通性的性质

引入连通性的概念，一个关键的好处是它刻画了拓扑空间的一个基本性质. 正如我们在同胚[625] 中提过的，连通性是一个同胚不变性，也就是说，同胚的拓扑空间的连通性是完全一致的. 这由以下定理保证：

定理 1　连通性的同胚不变性

设有拓扑空间 X 和 Y，令 $f:X\to Y$ 是一个连续映射，那么，对于 X 的任何连通子集 A，$f(A)$ 也是 Y 的连通子集.

应用“开集的逆映射还是开集”可以很容易地证明这一定理. 考虑到同胚映射在两个方向（$X\to Y$ 和 $Y\to X$）上都是连续映射，那么同胚映射把两个拓扑空间的连通子集彼此对应起来了.

在例 1 中我们知道 S 和 $\bar{S}$ 都是连通的，这不是偶然的.

定理 2　闭包的连通性

给定拓扑空间 X 和它的一个连通子集 A. 如果 $A\subseteq B\subseteq\bar{A}$，那么 B 也是连通的. 特别地，$\bar{A}$ 是连通的.

由“集合的内部、外部和边界[622]”中推论 1 可知，定理 2 中定义的 B 就是 A 添上若干边界点生成的；如果把边界点全都添上了，那就是 $\bar{A}$.

道路连通性

预备知识　连通性[632]

连通性的概念还不够直观，因此我们还有一个更强的性质，道路连通性. 连通性的定义思路是用不相交的开集来分割不连通的各部分，而道路连通性的定义思路是用一条连续的曲线来连接任意两点，如果不连通则这条曲线应该会被截断.

道路连通性的概念

定义 1　道路

在区间 $I=[0,1]$ 上取通常的度量（子）拓扑. 对于任意的拓扑空间 X，如果存在一个连续映射 $f:I\to X$，那么称该映射的像 $f(I)$ 为 X 中的一条道路（**path**）.

道路的定义要更直观一些，而且在别的数学分支也会出现，比如复变函数中会经常讨论复平面上的道路. 需要注意的是，这里定义道路所用的区间是闭

区间，也就是说它有起点和终点．可以把 I 看成是从 0 到 1 流逝的时间，而对应的 $f(I)$ 就是这段时间里，一个点在 X 中连续运动的轨迹．这个点运动的“速度”不会是无穷大，否则就不是连续映射了；这样，在有限时间内，这个运动轨迹必然也是“有限”的．这里打双引号是因为“速度”和“有限”在一般情况下只是一个类比，只有在度量空间里面我们才可以讨论长度，进而有了“速度”的概念．

例 1　道路的例子

- 在度量空间 $\mathbb{R}$ 中，I 本身就是一条道路．任何闭区间都是一条道路．
- 在 $\sin\frac{1}{x}$[12]空间中，对于任何 $a,b>0$，从 $x=a$ 到 $x=b$ 的一段函数图像也构成道路
- 还是在 $\sin\frac{1}{x}=S$ 空间中，不过考虑的是 $\bar{S}=S\cup\{(0,y)|y\in[-1,1]\}$，就是在 S 上再添加 $x=0$ 处的一条长度为 2 的线段．如果选择 $\{(0,y)|y\in[-1,1]\}$ 上的任意一点作为起点/终点，而选择 S 上的一点作为终点/起点，那么中间需要走一条无限长的曲线才能让二者相连，而道路都是有限的，所以这样的两点之间就没有道路．

两条道路 $f(I)$ 和 $g(I)$ 可以首尾相接，得到新的道路 $h(I)$，其中

$$\begin{cases} h(t)=f(2t), t\in[0,\dfrac{1}{2}] \\ h(t)=g(2t-1), t\in[\dfrac{1}{2},1] \end{cases} \tag{1}$$

道路的概念可以用来引出一个更直观的连通性：

定义 2　道路连通性

设拓扑空间 X 和它的子集 A．如果 $\forall a,b\in A$，总能找到一条道路连接 a 和 b，那么称 A 是道路连通（**path-connected**）的．

但是道路连通不等价于连通，连通的子集是可能不道路连通的．我们在例 1 中定义的 $\sin(1/x)$ 空间的闭包空间 $\bar{S}$ 就是一个反例．$\bar{S}$ 是连通的，因为 S 连

[12] 见例 1.

通，根据闭包的连通性定理 2，$\bar{S}$ 也得连通；但是由例 1 给出的反例，我们知道 $\bar{S}$ 不是道路连通的.

但是道路连通的子集一定连通.

定理 1　道路连通性推出连通性

对于拓扑空间 X 及其子集 A，如果 A 是道路连通的，那么 A 必是连通的.

证明

设 A 是道路连通但不是连通的. 那么必然有 $A \subseteq U \cup V$，其中 U, V 和 A 的交集都非空. 取 $a \in U$ 和 $b \in V$，那么由道路连通性，必然存在一条道路 $f : I \to A$，使得 $f(0) = a, f(1) = b$. 由于道路是连续映射，我们有 $f^{-1}(U)$ 和 $f^{-1}(V)$ 都是 I 的非空开子集，而且 $I = f^{-1}(U) \cup f^{-1}(V)$. 由于 I 是连通的，它不可能是两个非空开子集的并. 因此矛盾，所以 A 必须是连通的.

证毕

从这个证明的思路可以看出，道路连通性本质上是在描述每一条道路的连通性，由于道路的端点是任意取的，这就要求 A 本身必须连通了.

道路连通性的性质

道路连通性也是一个同胚不变量.

定理 2　道路连通性的同胚不变性

设有拓扑空间 X 和 Y，令 $f : X \to Y$ 是一个连续映射，那么，对于 X 的任何道路连通子集 A，$f(A)$ 也是 Y 的道路连通子集.

分离性

预备知识　紧致性[628]

分离性的种类一览

分离性是描述一个拓扑空间里，任意的点、子集等彼此之间能被不相交的开集分开的程度. 我会在这里先列出常见的分离性和它们的简单解释，但你不需要掌握所有分离性，只有其中两个是很重要的.

定义 1　分离性的种类

- T_0 分离性，是指取空间中不同的两点 x, y，总存在一个开集 U，使得 U 包含其中一点而不包含另一点.
- T_1 分离性，是指取空间中不同的两点 x, y，总存在两个开集分别含有其中一个点，但各自不含有另一个点.
- T_2 分离性，是指取空间中不同的两点 x, y，总存在两个开集，分别含有其中一个点，并且这两个开集不相交.
- T_3 分离性，是指取空间中不同的两点 x, y 或者将 y 替换为一个不含 x 的闭集，那么总存在两个开集，分别含有其中一个点或闭集，并且这两个开集不相交.
- 正规分离性，是指取空间中不相交的两个闭集 A, B，总存在两个开集，分别含有其中一个闭集，并且这两个开集不相交.
- T_4 分离性，既正规分离又 T_2 分离的性质.

这些分离性之间的区别很细微，看起来很绕，对不对？数学家们将分离性的分类做得比这个要详细得多，除了列表里的，他们还研究了诸如 $T_{2.5}$ 分离性，$T_{3.5}$ 分离性，R_1 分离性，完全正规分离性，正则分离性，正则 Hausdorff 分离性等非常多的分离性. 但常用的重要分离性只有其中两个，T_2 **分离性**和**正规分离性**. 其中 T_2 分离性又被称为 **Hausdorff 分离性**.

Hausdorff 空间和正规空间

为了方便计，我将重新誊写一遍两个重要分离性的定义.

定义 2　Hausdorff 空间和正规空间

给定拓扑空间 X.

- 称 X 是 Hausdorff 空间，如果任意的 $x, y \in X$，都可以被两个不相交的开集 U_x, U_y 分别包含.

- 称 X 是正规空间，如果任意的两个闭集 $A, B \subseteq X$，都可以被两个不相交的开集 U_A, U_B 分别包含.

我们在紧致性[628] 一节的开头提到，紧子集的行为常常和单个点是相似的，而紧子集又常常是闭集；在这里，Hausdorff 空间和正规空间的概念差别，无非就是一个讨论点和点的关系，另一个讨论闭集和闭集的关系. 它们也因此有一些类似的性质.

定理 1　分离性的继承

Hausdorff 空间的任意子空间还是 Hausdorff 的. 正规空间的任意闭子空间还是正规的.

定理证明是很简单的. 注意正规空间的继承性要求必须是闭子集构成的空间，因为只有这样才能保证子空间的闭集仍然是原来空间的闭集，从而直接继承原空间的分离性. 举个反例，线段 $[0, 10]$ 上的度量空间是正规的，如果取子集 $(1, 3) \cup (4, 5)$ 来构成子空间，那么根据子拓扑的定义 3 [619]，$[2, 4] \cap [(1, 3) \cup (4, 5)] = [2, 3)$ 是 $(1, 3) \cup (4, 5)$ 空间的闭集，但它显然不是 $[0, 10]$ 空间的闭集. 当然了，$(1, 3) \cup (4, 5)$ 空间也不是正规空间.

紧的 Hausdorff 空间有一个非常良好的性质：它的紧子集和闭集是等价的.

定理 2　紧 Hausdorff 空间的性质

- 紧 Hausdorff 空间的紧子集一定是闭的.
- 紧 Hausdorff 空间是正规空间.

这个定理的证明是非常巧妙的，我列举如下，感兴趣的读者可以仔细体会：

证明：要证明一个集合是闭集，等价于证明它的补集是开集.

取拓扑空间 X. 设 A 是 X 的紧子集，任取 $x \in X - A$.

取 A 中任意一点 a，由 Hausdorff 分离性，存在两个开集 $U_a \ni a, V_a \ni x$，并且 U_a 和 V_a 不相交. 对每一个 $a \in A$ 都取出这样的 U_a，V_a 对，那么 $\{U_a\}_{a\text{取遍}A}$ 是 A 的一个覆盖. 由于 A 是紧子集，这个覆盖存在有限子覆盖.

也就是说，存在有限个 $a_i \in A, i = 1, 2, \cdots n$，使得 $\{U_{a_i}\}$ 就足以覆盖 A 了. 这里的**有限**非常关键，因为它使得 $\bigcap_{i=1,2,\cdots n} V_{a_i}$ 是开集的有限交，因此仍然是开

集[13].

记 $\bigcup\limits_{i=1,2,\cdots n} V_{a_i}=U$，且 $\bigcap\limits_{i=1,2,\cdots n} V_{a_i}=V$，则 U,V 都是开集，$U\supseteq A$，$V\ni x$，并且 U,V 没有交集 $\to A,V$ 没有交集.

因此，x 是 $X-A$ 的一个内点. 由于 x 是任意取的，故可知每一个 $x\in X-A$ 都是其内点，故 $X-A$ 是开集，故 A 是闭集.

类似地可以对两个不相交的闭集中的点彼此配对，应用 Hausdorff 分离性和 Hausdorff 空间中闭集等价于紧集，可以类似地证明紧 Hausdorff 空间都是正规空间.

证毕.

一点紧化空间

Hausdorff 空间不一定是紧空间，但是总可以添上一个点以后成为紧空间. 最常见的例子就是二维平面添上一个点以后成为一个球面，原本的二维平面因为无穷延伸，所以不是紧空间；但是添上一点再相应定义一些新的开集以后，它就等价于一个有限的球面了，从而变成了紧空间. 这种添上一点使非紧空间变成紧空间的操作，叫做“一点紧化”.

假设有一个 Hausdorff 空间 $(X,\mathcal{T})$，它不紧致. 我给它添上一个点，叫做无穷远点，记为 P，得到一个新的集合 $X\cup\{\mathcal{P}\}=X'$，在这个集合 X' 上定义拓扑：开集一共分两种，含 P 和不含 P 的. 不含 P 的开集都是原先 X 中的开集，而含 P 的开集 O，都是某个 $K\subseteq X$ 的补集：$O=X'-K$. **其中 K 是 X 的一个紧集**. 这样构成了一个新的空间 $(X',\mathcal{T}')$，称为 $(X,\mathcal{T})$ 的**一点紧化空间**.

这样添上一个点并加入和含这个点的开集的定义，就能得到一个紧 Hausdorff 空间. 二维平面变成球面的过程就是这样的.

例 1　黎曼球面

平面几何所研究的空间 $\mathbb{R}^2$ 是一个 Hausdorff 空间，但它不紧致——这是很显然的，如果你取 A_n 是以原点为圆心、n 为半径的圆盘，那么 $\{A_n\}$ 自然是整个平面的一个覆盖，而它不存在有限子覆盖. 当我们给这个平面添加一个无

[13]回顾拓扑的定义，只要求了任意并和有限交的封闭性；以 $\mathbb{R}$ 为例也可看出，对于所有正整数 n，开区间 $(-1,1/n)$ 的交集是 $(-1,0]$，这显然不是一个开集.

穷远点 P 并按照上述方式进行一点紧化以后，所得到的空间实际上等价于一个球面，记为 S^2.

反过来，你也可以把 $\mathbb{R}^2$ 平面看成是球面 S^2 挖去了一个点 P 构成的. 事实上，如果你把一个球放在这个平面的原点处，让球的南极点和原点重合，然后从球的北极点拉出一条条射线穿过球面和平面，球面上和平面上被穿过的点相互联系，就几乎构成了球面和平面的一个双射，几乎每一个球面上的点都唯一对应平面上的一个点，除了北极点本身. 这个北极点就是我们所讨论的 P 点，那个在平面上并不存在的无穷远点.

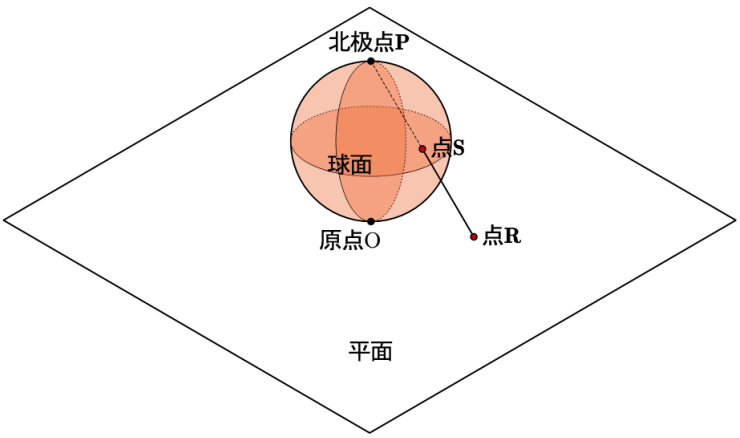


图 1: 黎曼球面示意图. 图中球面的南极点和平面的原点 O 是重合的. 从北极点 P 拉出的一条射线在 S 穿过球面，和平面相交于 R. 球面和平面上的各点 S 和 R 以这种方式一一对应.

这样给 $\mathbb{R}^2$ 平面添上一个无穷远点并定义新的开集以后所得到的球面，被称为黎曼球面. 通常我们会规定这个球面半径为 1.

乘积拓扑

预备知识 拓扑空间[619]

有限维乘积拓扑

给定两个拓扑空间 $(X_1, \mathcal{T}_1)$ 和 $(X_2, \mathcal{T}_2)$，那么我们可以在乘积集合 $X = X_1 \times X_2$ 中定义一个拓扑 $\mathcal{T}$，其拓扑基为 $\{O_1 \times O_2 | O_1 \in \mathcal{T}_\infty, O_2 \in \mathcal{T}_\in\}$. 就是说，乘积集合的拓扑基，是两个空间中开集的笛卡尔积的集合.

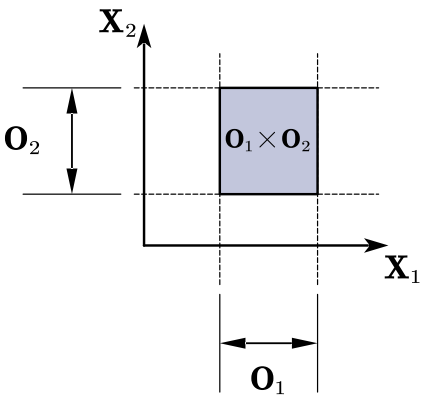


图 1: 二维乘积拓扑的拓扑基中每个元素都是如图所示的一个“矩形”.

由拓扑基生成拓扑的方式（取任意并），容易发现，如果上述 O_1 和 O_2 取的只是 X_1 和 X_2 空间的某两个拓扑基，也能得到一样的定义.

一般地，N 个拓扑空间的集合做笛卡尔积，这个笛卡尔积集合上的**乘积拓扑**定义为：

定义 1　有限维乘积拓扑

设 $(X_n, \mathcal{T}_n)$ 是若干拓扑空间，n 取值范围为 $[1, N] \cap \mathbb{Z}$. 那么乘积空间 $X_1 \times X_2 \times X_3 \times \cdots \times X_N = \prod\limits_{n=1,2,\cdots,N} X_n = X$ 中的拓扑由拓扑基 $\mathcal{B}$ 生成，其中 $\mathcal{B} = \{\prod\limits_{n=1,2,\cdots,N} O_n | \forall n, O_n \in \mathcal{T}_n\}$.

任意维乘积拓扑

如果用于进行笛卡尔积的拓扑空间数量大于等于 $\aleph_0$，那么我们常用的乘积拓扑定义会和有限维情况的说法略有不同. 在这里，我们使用**子基（sub-basis）**来定义乘积拓扑：

定义 2　任意维乘积拓扑

设 $(X_\alpha, \mathcal{T}_\alpha)$ 是若干拓扑空间，α 不再是整数指标，而是用一个集合 Λ 中的元素来表达的指标：$\alpha \in \Lambda$. 这样的集合 Λ 称为一个指标集（**set of indexes**）.

空间 $X = \prod\limits_{\alpha\in\Lambda} X_\alpha$ 的拓扑 $\mathcal{T}$ 由子基 $\mathcal{S}$ 生成，其中 $\mathcal{S} = \{O_{\alpha_0} \times \prod\limits_{\alpha\in\Lambda-\{\alpha_0\}} X_\alpha | \alpha_0 \in \Lambda, O_{\alpha_0} \in \mathcal{T}_{\alpha_0}\}$. 就是说，$\mathcal{S}$ 中的每一个元素，都是某个 X_{α_0} 的开集 O_{α_0} 和其它所有 X_α 乘积的结果.

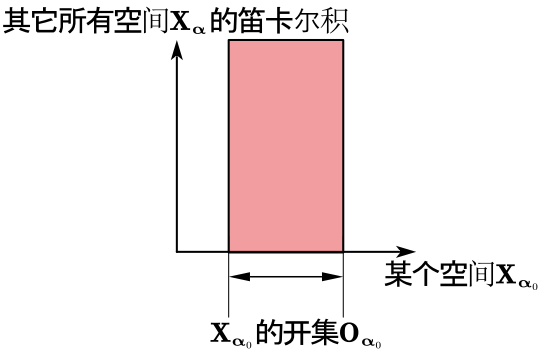


图 2: 子基 $\mathcal{S}$ 中的一个元素，类似一个“柱形”（类比 $\mathbb{R}$ 空间之间的乘积）.

其中，各 X_α 空间称作 X 的一个分量（**component**），类比向量空间的称呼.

任意维乘积拓扑也可以用拓扑基定义，只不过拓扑基中的元素不再是各分量空间的开集之乘积，而是有限个分量空间的开集和其它全体分量空间本身的乘积.

任意维乘积拓扑的定义是包含了有限维的定义的，因为当进行乘积的空间是有限个（即 $|\Lambda| < \aleph_0$）的时候，任意维乘积拓扑的定义就和有限维乘积拓扑的定义一致了.

乘积映射

给定集合间的映射 $f : A \to X$ 和 $g : B \to Y$，可以定义映射 $f \times g : A \times B \to X \times Y$. 其中 $\forall a \in A, b \in B$，有 $f \times g(a, b) = (f(a), g(b))$.

习题 1

证明：如果上述集合是拓扑空间，并且 f 和 g 都是连续映射，那么 $f \times g$

也是乘积空间之间的连续映射.

商拓扑

预备知识 二元关系[858]，映射[98]，拓扑空间[619]

商集

拓扑空间本身作为一个集合，可以在其上定义某种等价关系，由此可以得到等价类所构成的**商集**. 我们可以将同一个等价类中的点看成是同一个点，把商集看成是原集合中部分点相互粘合所得到的集合.

例 1　烟卷

在 $\mathbb{R}^2$ 中取一个矩形的闭子集：$A=\{(x,y)\in\mathbb{R}^2 : x\in[-1,1], y\in[0,1]\}$. 在这个闭子集上定义一个等价关系 $\sim$ 如下：对于 $y\in[0,1]$，有 $(1,y)\sim(-1,y)$；其它点都只和自己等价. 如果把 A 看成一个矩形纸条，那么商集 $A/\sim$ 可以看成是这个纸条的两边对应点粘在一起，卷成一个烟卷的样子.

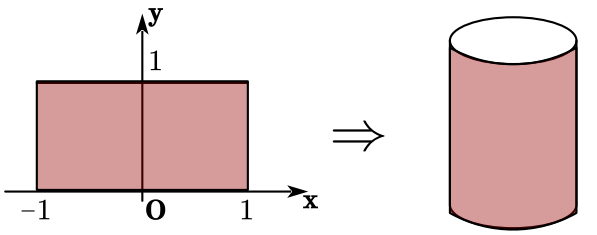


图 1: 卷纸条的过程示意. 左边是矩形纸条 A，右边是模去等价关系 $\sim$（或者说把等价的点粘在一起）以后得到的集合.

我们给出纸条 A 的准确坐标只是为了方便描述上述等价关系. 实际上，在 $\mathbb{R}^2$ 中选择任意的有界连通闭子集，都可以找到这个闭子集和 A 的同胚映射，在拓扑意义下它们就是相同的集合. 取一个任意的四边形、三角形或者不规则图形都可以，你可以直观地想象成揉捏一个橡皮泥，只要不撕裂它，不把本来分开的点粘在一起，那揉出来的任何形状都是同胚的. 当然，用矩形纸条描述以上商集最为方便.

例 2　甜甜圈

还是利用例 1 中的 A, 这次等价关系 $\sim$ 为：对于 $x \in [-1,1], y \in [0,1]$，有 $(x,0) \sim (x,1), (1,y) \sim (-1,y)$，其它点只和自己等价．这一次，商集 $A/\sim$ 可以看成把纸条的左右两边粘合，上下两边也粘合．先粘合哪些点并没有要求，所以可以先粘合成烟卷，再粘合成甜甜圈或者说轮胎的形状．

这种粘合若干边的商集很常见，我们可以定义一种图形化的表示方法，来表示等价关系．以烟卷为例，我们把待粘合的边表示为一个箭头，取名为 x，然后把两个 x 箭头按照“头对头，尾对胃”的方向粘在一起，如图所示：

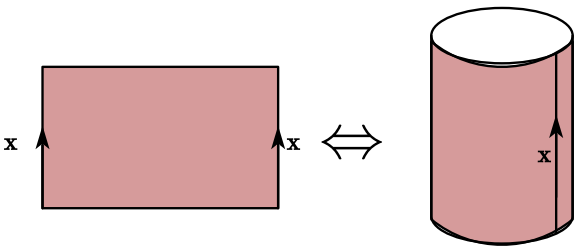


图 2: 名称相同的箭头视为同一个，把它们粘连在一起．图示两个形状是相同的，左边的有两个同为 x 的箭头，要把它们粘连起来，结果是烟卷形状；右边只有一个箭头，不用进行粘连操作了，它本身也就是烟卷形状．

用这种箭头表示法，也可以通过在矩形上标注箭头 x 和 y 来表示一个甜甜圈：

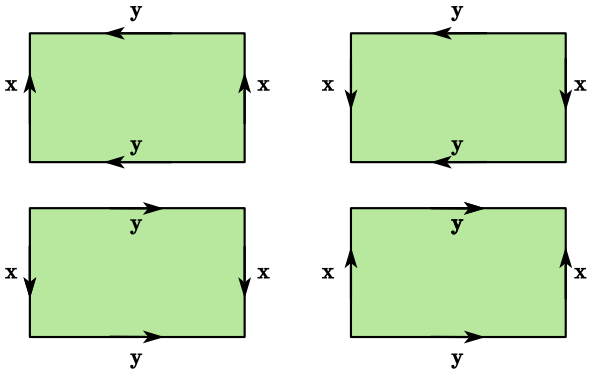


图 3: 甜甜圈的四种箭头表示方法．这四种方法是完全等价的．



[ht]

图 4: 一个莫比乌斯环．将两个未命名的箭头粘在一起

例 3　莫比乌斯环

用箭头表示法可以很轻松地表示一个莫比乌斯环（**Mobius**）．考虑到莫比乌斯环是把纸条的一边翻转以后再和对边粘连而成的，我们可以如图 4 画出莫比乌斯环的箭头表示法．如果只涉及一组边的粘合，我们可以在不引起混淆的情况下省去对箭头的命名．

例 4　克莱因瓶

用箭头表示法来表示一个克莱因瓶（**Klein's Bottle**）：

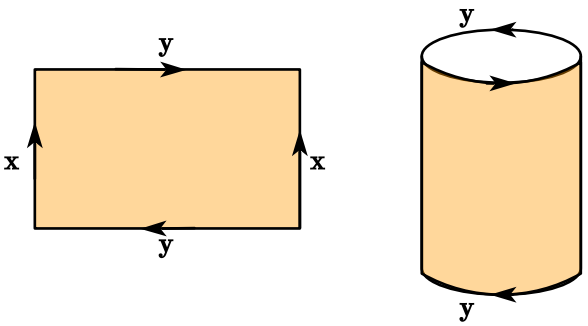


图 5: 克莱因瓶的一种表示．我们可以把克莱因瓶看成先把 x 粘起来，再把 y 粘起来的操作，同样也可以看成先粘连 y 再粘连 x，只是后一种思路复杂得多，更难想象．

注意克莱因瓶的定义中，两个 y 箭头的朝向相反．如果先粘连 y 箭头的话，所得到的就是一个莫比乌斯环，由此可以看出莫比乌斯环和克莱因瓶的相似性．

例 5　射影平面

射影平面和克莱因瓶、莫比乌斯带一样，如果一只蚂蚁在射影平面上爬行

并且不离开平面也不触及边缘，它总能爬到出发点的对面．用箭头表示法来表示射影平面如下：

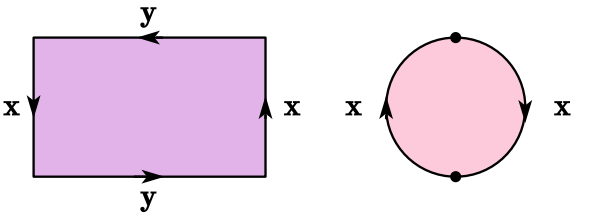


图 6: 射影平面示意图．矩形和圆形的示意图是等价的．射影平面很难直观地画出来，多数情况下只能这样表示．

如果在射影平面的中间挖去一个洞，那么我们就得到了莫比乌斯带．射影平面实际上是莫比乌斯带将边缘粘连起来的结果．

如何看出射影平面挖去一个洞之后就变成莫比乌斯带呢？我们需要反其道而行之，把平面“暂时”剪开，之后再粘连起来．如图 7 所示．

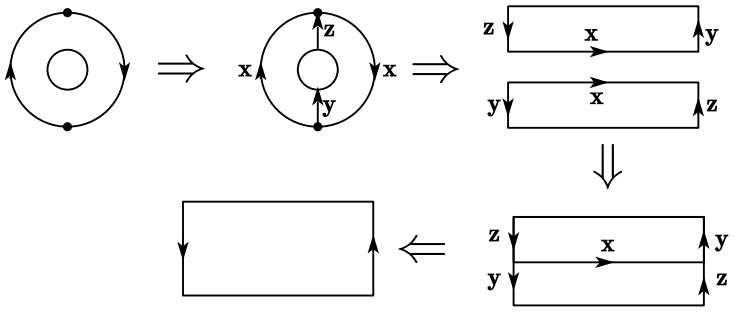


图 7: 从射影平面中间挖去一个洞得到莫比乌斯带示意图．第一步是标注出两个箭头 z 和 y，第二步沿着 z 和 y 箭头将图形剪开，第三步是将 x 箭头粘连起来，第四步是将粘连起来的 x 擦掉，并将 y 和 z 合并，省略名称．可见这就是一个莫比乌斯环．

商拓扑

给定拓扑空间 $(X,\mathcal{T})$，以及集合 X 上任意一个等价关系 $\sim$，定义集合间的映射 $f:X\to X/\sim$，把 X 中的每一个点 P 都映射到 P 所属的等价类上．

那么商集 $X/\sim$ 上可以定义一个新的拓扑 $\mathcal{T}_\sim = \{U \in X/\sim : f^{-1}(U) \in \mathcal{T}\}$. 将这个新的拓扑称为**商拓扑**（**quotient topology**）.

注意，商集中某个点的逆映射，不是 X 中对应的这个点，而是这个点所属的整个等价类.

例 6

还是使用例 1 中的 A，这次把等价关系 $\sim$ 定义为：矩形的底部 $\{(x,0)|x \in [-1,1]\}$ 所有点都是等价的，其它部分只和自己等价. 映射 $f : A \to A/\sim$ 把 A 中的点映射到其等价类上.

假设 A 的底部在 $A/\sim$ 中对应的一个点叫 P，那么 $X/\sim$ 中任何包含 P 的开集，其逆映射都是 X 中包含整个底部的开集. 如果记 $R = \{(x,y) \in \mathbb{R}^2 : x \in (-1/2, 1/2), y \in [0,1]\}$，并记底部为 B，那么 $f^{-1}(f(R)) \neq R$，而是 $f^{-1}(f(R)) = R \cup B$，而这不是一个开集，所以 $f(B)$ 不是 $X/\sim$ 的开集.

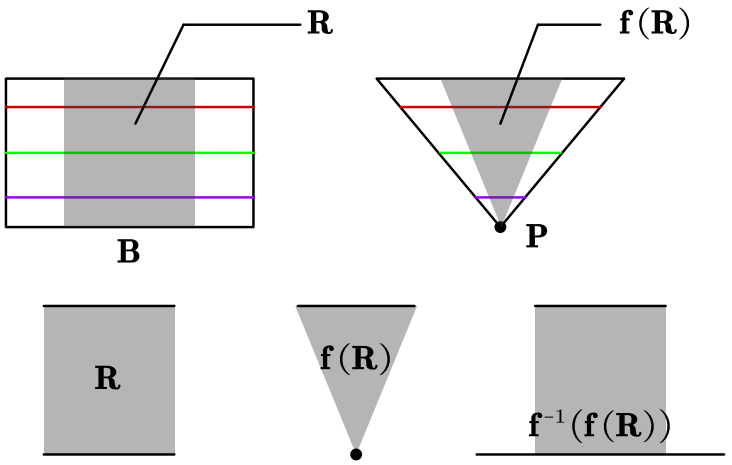


图 8: A，R，B，$A/\sim$，$f(R)$ 和 $f^{-1}(f(R))$ 的示意图. 图中阴影部分表示 R 或 $f(R)$. 可见 $f^{-1}(f(R))$ 和 R 相比，底边两端突出了，不再是 X 的开集，因此 $f(R)$ 不是 $X/\sim$ 的开集.

由例 6 可见，X 和 $X/\sim$ 的开集不一定都会一一对应.

锥空间

给定任意拓扑空间 X，再取 $\mathbb{R}$ 的一个子空间 $I=[0,1]$，那么 $X\times I/\{(x,0)|x\in X\}$ 就是 $X\times I$ 中把全体 $(x,0)$（底部）粘成一个点的商空间，称作 X 的**锥空间**，记为 $\widetilde{C}X$. 如果再把锥空间的顶部也粘连起来，得到的 $\widetilde{C}X/\{(x,1)|x\in X\}$ 称为 X 的**双角锥空间**，记为 $\widetilde{S}X$.

映射空间

预备知识 紧致性[628]

紧开拓扑

给定拓扑空间 X 和 Y，Y^X 表示从 X 到 Y 的所有**连续映射**的集合. 我们在此连续映射集合上定义一个拓扑.

定义 1　紧开拓扑

给定拓扑空间 X 和 Y，在 X 中取一个紧子集 K，在 Y 中取一个开集 U，那么所有使得 K 的像包含在 U 中的连续映射构成的集合 $\{f\in X^Y|f(K)\subseteq U\}$，定义为 X^Y 中的一个开集. 由于这个开集是由紧子集 K 和开集 U 决定的，我们可以记为 $M_{K,U}$. 全体 $M_{K,U}$ 构成的集合，作为子基生成一个拓扑 $\mathcal{T}$[14]，那么 $\mathcal{T}$ 称为一个紧开拓扑. $(X^Y,\mathcal{T})$ 称为 X 到 Y 的（连续）映射空间.

紧开拓扑的定义初看并不直观，实际上它和度量空间有紧密的联系. 在特定条件下，紧开拓扑和度量拓扑是等价的，因此紧开拓扑可以看成度量拓扑更一般的形式.

定理 1　紧开拓扑与度量拓扑的等价性

设拓扑空间 X 为紧空间，Y 为度量空间，其距离函数记为 d，在 Y^X 上定义度量 ρ 如下：对于 $f,g\in Y^X$，令 $\rho(f,g)$ 等于集合 $\{\mathrm{d}(f(x),g(x))\}_{x\in X}$ 中的上

[14]注意，这里是子基，而非拓扑基.

确界．则这个度量导出的拓扑，等价于紧开拓扑．

一般来说，当我们提到映射空间时，总是指连续映射构成的集合，配上紧开拓扑所构成的拓扑空间．

映射空间的性质

定理 2　映射空间的重要性质

- 若 X 是局部紧空间，令映射 $e: Y^X \times X \to Y$ 满足 $\forall f \in Y^X, x \in X$，有 $e(f,x) = f(x)$．则 e 是一个连续映射．
- 若 $f: X \times Z \to Y$ 是连续映射，那么令 $\tilde{f}: Z \to Y^X$ 满足 $\tilde{f}(z)(x) = f(x,z)$[15]，那么 $\tilde{f}$ 也是连续映射．如果 X 还是一个局部紧空间，那么可以反过来说，如果 $\tilde{f}$ 是连续映射，那么 f 是连续映射．
- 如果 Z 是 Hausdorff 空间，那么 $E: Y^{X\times Z} \to (Y^X)^Z$ 就是连续映射，其中 $E(f) = \tilde{f}$．如果 X 也是 Hausdorff 的，那么 E 是一个同胚映射．

以上三条定理没有简略的严格证明，在此从略；感兴趣的读者请参考拓扑学相关教材．这三条定理的性质是代数拓扑的基础．

空间偶和带基点空间

预备知识　映射空间[648]

定义 1　不交并

给定集合 A 和 B．如果在对这两个集合取并时特别区分两个集合中的元素，强行认为它们没有交集，这样取出来的并集被称为 A 和 B 的不交并，记为 $A \amalg B$．

例 1　不交并的例子

[15]注意，这里的 $\tilde{f}(z)$ 是 Y^X 的一个元素，也就是 $X \to Y$ 的一个映射．

- 给定集合 $A=\{1,2,3\}$，$B=\{2,3,4\}$，给两集合中的元素打上各自的标签以进行区分，得到 $A'=\{1_A,2_A,3_A\}$ 和 $B'=\{2_B,3_B,4_B$，那么并集 $A'\cup B'=\{1_A,2_A,3_A,2_B,3_B,4_\}=\{1,2_A,3_A,2_B,3_B,4\}$ 也就是不交并 $A\amalg B$.

- 给定两个一维球(圆)$A=\{(x,y)|x^2+y^2=1\}$ 和 $B=\{(x,y)|(x-1)^2+y^2=3\}$，那么 $A\cup B$ 是两个粘在一起的圆，而 $A\amalg B$ 是两个分开的圆．由于在拓扑意义上，任何两个圆都是同胚的，取不交并时也强行将其分开了，我们就可以把任何圆简单记为 S^1 再作不交并．一般地，S^n 表示任意一个 n 维球面.

定义 2　空间偶

- 把拓扑空间 X 和它的一个子空间 A 绑定，记为 (X,A)，称其为一个空间偶.

- 如果一个连续映射 $f:X\to Y$ 还满足 $f(A)\subseteq B$，即限制在 A 上时可以认为 $f:A\to B$，那么我们也称这是一个空间偶之间的映射 $f:(X,A)\to(Y,B)$.

- 空间偶的映射空间 $(Y,B)^{(X,A)}$ 是所有满足 $f(A)\subseteq B$ 的连续映射 $f:X\to Y$ 的集合．显然这是 Y^X 的子集，因此我们定义其拓扑为 Y^X 的子拓扑.

- 乘积空间偶 $(X,A)\times(Y,B)$ 定义为 $(X\times Y,(X\times B)\cup(A\times Y))$．特别地，$X\cong(X,\varnothing)$，所以由 $X\times(Y,B)=(X\times Y,X\times B)$.

定义 3　带基点空间

取拓扑空间 X 中的一个点 x_0．称空间偶 $(X,\{x_0\})=(X,x_0)$ 是一个带基点空间，称 x_0 为这个带基点空间的基点.

定义 4　带基点空间的一点并和压缩积

- 给定带基点空间 (X,x_0) 和 (Y,y_0). 记 $(X \amalg Y)/(x_0 \sim y_0) = (X,x_0) \vee (Y,y_0) = (X \vee Y, x_0 \vee y_0)$，称为 (X,x_0) 和 (Y,y_0) 的一点并. 换句话说，一点并就是将两个带基点空间进行不交并，然后再把基点粘在一起.

- 记 $(X \wedge Y, x_0 \wedge y_0) = (X,x_0) \wedge (Y,y_0) = X \times Y/\sim$，其中 $\sim$ 定义为 $\forall x \in X, y \in Y$ 有 $(x,y_0) \sim (x_0,y) \sim (x_0,y_0)$.

拓扑群

预备知识 连续映射[625]，自由群[562]

拓扑群，顾名思义，是一种同时具有群和拓扑结构的数学对象. 但是光有两个结构还不够，拓扑群还要求满足拓扑和群运算之间的一个联系.

定义 1　拓扑群

给定一个集合 G，若在 G 上定义了一个拓扑 $\mathcal{T}$ 和一个群运算 "$\cdot$"，且满足拓扑空间之间的映射 $f: G \times G \to G$ 是一个连续映射，其中 $f(g_1,g_2) = g_1 \cdot g_2^{-1}$，那么称 G 是一个拓扑群（**topological group**）.

要求 $f(g_1,g_2) = g_1 \cdot g_2^{-1}$ 是一个连续映射，也就保证了 $f_1(g_1,g_2) = g_1 \cdot g_2$ 和 $f_2(g) = g^{-1}$ 都是连续映射.

通常，我们也会省略运算符号，而简单把 $g_1 \cdot g_2$ 记为 $g_1 g_2$.

定义 2　拓扑群同态

设 G 和 H 是两个拓扑群，它们之间有映射 $f: G \to H$. 如果 f 在拓扑意义上是连续映射，在代数意义上是群同态，那么称 f 是一个拓扑群之间的同态.

下面我们看两个简单的例子.

例 1　拓扑群的例子

- 取实数轴 $\mathbb{R}$，定义通常的度量拓扑以及加法群，则 $\mathbb{R}$ 是一个拓扑群.

- 取复平面上的单位圆 $S^1 = \{e^{\theta i} | \theta \in \mathbb{R}\}$，定义其拓扑为二维欧几里得空间中的子拓扑，群运算为复数乘法，则 S^1 是一个拓扑群.

- 定义映射 $f:\mathbb{R}\to S^1$ 为 $f(t)=\mathrm{e}^{2\pi t\mathrm{i}}$，则 f 是一个拓扑群同态.

以上两个例子非常容易想象出来. 下面，我们稍微深入一些，讨论一个不那么直观的拓扑群，这对将来理解微分几何富有启发意义.

例 2　一般线性群

给定实数域 $\mathbb{R}$ 和一个正整数 k，则全体 $\mathbb{R}$ 上的 k 阶矩阵构成一个 k^2 维欧氏空间，也就是一个拓扑空间. 对于这个空间中的每个点（也就是矩阵），其第 i 行 j 列的矩阵元就是该点在第 $k(i-1)+j$ 个坐标轴上的分量.

如上定义的矩阵空间中，所有非退化矩阵构成一个子集，记为 $\mathrm{GL}(k,\mathbb{R})$. 在 $\mathrm{GL}(k,\mathbb{R})$ 上用子拓扑，则它成为一个拓扑空间.

$\mathrm{GL}(k,\mathbb{R})$ 上，用矩阵乘法作为运算，则它还构成一个群.

通过以上方式，在 $\mathrm{GL}(k,\mathbb{R})$ 上分别定义的拓扑和群结构，合在一起就使得 $\mathrm{GL}(k,\mathbb{R})$ 成为一个拓扑群.

$\mathrm{GL}(k,\mathbb{R})$ 作为拓扑空间可以理解为 $\mathbb{R}^{k^2}$ 空间里挖去了若干孤立点，每个孤立点都代表一个退化矩阵. 定义 $f:\mathrm{GL}(k,\mathbb{R})\times\mathrm{GL}(k,\mathbb{R})\to\mathrm{GL}(k,\mathbb{R})$，其中 $f(g_1,g_2)=g_1\cdot g_2^{-1}$，那么要让它构成拓扑群就必须满足 f 是一个连续映射，使得这一点成立的关键在于行列式是矩阵空间的连续函数，以及行列式的积性[16].

[16]即矩阵乘积的行列式等于矩阵行列式的乘积.

第二章

同伦论

映射的同伦和空间的同伦

预备知识 映射空间[648]

同胚的两个拓扑空间是完全相同的，在拓扑学意义下无法区分．证明两个拓扑空间同胚通常是非常困难的，但证明两个拓扑空间不同胚却可以有很多容易计算的方法，其中一个方法就是证明两个拓扑空间不**同伦**．

同伦是一种比同胚弱一些的关系，是拓扑学中的一种代数性质．为了介绍拓扑空间中的同伦，我们首先要讨论映射的同伦．我们对于映射之间的关系和拓扑空间之间的关系都使用了“同伦”这一术语，但不会造成混淆，因为两个“同伦”之间是相互导出的概念，并且所针对的对象不同．

映射的同伦

定义 1　映射的同伦

设有两个拓扑空间 X 和 Y，以及它们之间的两个连续映射 $f_0, f_1 : X \to Y$；记 $I = [0,1]$，取度量空间 $\mathbb{R}$ 的子空间拓扑．如果存在一个连续映射 $H : X \times I \to Y$，使得对于任何 $x \in X$，都有 $H(x,0) = f_0(x)$ 和 $H(x,1) = f_1(x)$．那么我们称 f_0 和 f_1 是同伦的映射，记为 $f_0 \overset{H}{\cong} f_1$，或者简单记为 $f_0 \cong f_1$；H 称为从 f_0 到 f_1 的一个同伦或者伦移．

同伦的本质是道路．这句话有两个含义：第一，固定 X 中的某一点 x_0，那么 $H(x_0, I) = p(I)$ 就是一个 I 到 Y 的映射，即一条道路；第二，在映射空间 Y^X 中来看，f_0 和 f_1 分别是两个点，而 H 就可以看成是一个从闭区间 I 到 Y^X 上的映射，即道路．因此，我们有以下两条重要定理，一个写为习题，一个写为定理：

习题 1　映射的同伦是道路（对给定点）

各空间和映射的定义同定义 1．给定 $x_0 \in X$，对于任何的 $t \in I$，构造映射 $p : I \to Y$，其中 $p(t) = H(x_0, t)$．证明 p 是一个连续映射，从而说明 p 是一条道路．

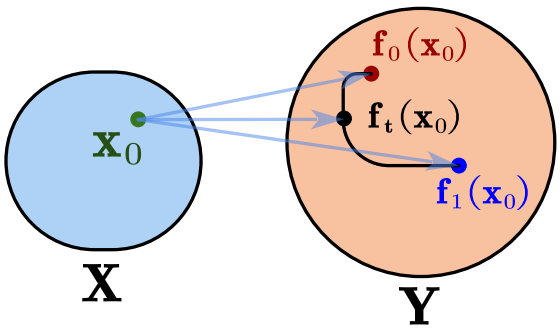


图 1: 习题 1 的示意图．固定 x_0 之后，$p(t)=H(x_0,t)$ 在 Y 中画出一条轨迹．

定理 1　映射的同伦是道路（对同伦本身）

各空间和映射的定义同定义 1．如果 $p:I\to Y^X$ 满足 $p(t)=f_t$，那么 p 是一个连续映射．由此可推论出，H 本身可以看成是映射空间里的一条道路．

证明：

直接引用定理 2 [648]，令 $Z=I$，$f=H$，则 $p=\widetilde{f}$，因此 p 是连续映射．

证毕．

以上两条定理分别对于 X 中的给定点以及 X 整体的角度描述了同伦和道路的关系．第二个定理说明，两映射同伦，等价于说两映射是同一条道路的两端，还等价于说两映射是同一条道路上的两点．由于道路可以首尾相接从而得到新的道路，因此同一个映射空间中，如果 $f\cong g$ 而 $g\cong h$，那么一定有 $f\cong h$．这就是说，同伦关系是具有传递性的．考虑到同伦关系显然还有自反性和对称性，我们直接可得如下定理：

定理 2

映射的同伦关系是一个等价关系．

映射之间还有复合运算，同伦关系也会继承到复合运算上．

定理 3　映射同伦关系的继承

设 X，Y 和 Z 是三个拓扑空间，$f_0,f_1:X\to Y$，$g_0,g_1:Y\to Z$ 分别是连续映射，且有 $f_0\overset{F}{\cong}f_1$ 和 $g_0\overset{G}{\cong}g_1$．那么 $g_0\circ f_0\cong g_1\circ f_1$．

证明：

由于 $F : X \times I \to Y$ 和 $g_0 : Y \to Z$ 都是连续映射，故 $g_0 \circ F : X \times I \to Z$ 是一个连续映射，故 $g_0 \circ F$ 是一个连接了 $g_0 \circ f_0$ 和 $g_0 \circ f_1$ 的道路. 故 $g_0 \circ f_0 \cong g_0 \circ f_1$.

类似地，$G \circ (f_1 \times 1_I) : X \times I \to Z$[1]是一个连接了 $g_0 \circ f_1$ 和 $g_1 \circ f_1$ 的道路，故 $g_0 \circ f_1 \cong g_1 \circ f_1$.

综上，由映射同伦的传递性，$g_0 f_0 \cong g_1 f_1$.

证毕.

拓扑空间的同伦

定义 2　空间同伦

设有两个拓扑空间 X 和 Y，如果存在连续映射 $f : X \to Y$ 和 $g : Y \to X$，使得 $gf \cong 1_X$ 且 $fg \cong 1_Y$，那么我们称 X 和 Y 是同伦的，或者说具有相同的伦型. f 称为从 X 到 Y 的同伦等价，反过来 g 是从 Y 到 X 的同伦等价，二者互为同伦逆.

同伦和同胚一样，也是一个等价关系，只是要更弱一些，使得不同胚的空间也有可能同伦. 当然，同胚的空间必然同伦.

习题 2

证明空间的同伦具有传递性，从而推论同伦是一个等价关系.

思考

在映射空间[648] 中所定义的紧开拓扑看起来很不接地气. 本小节习题 1 直观地阐释了同伦的意义，即对于每个固定点 $x_0 \in X$，同伦 H 都会在 Y 中画出一条道路，这意味着每个点的映射都是**连续地变化**的. 定理 1 说明，整体上来看，H 本身是映射空间 Y^X 中的一条道路. 从这个联系来看，思考一下为什么要用紧开拓扑来定义映射空间.

---

[1] $1_I : I \to I$ 是恒等映射，$f_1 \times 1_I$ 的定义见乘积拓扑[640] 中的习题 1.

可缩空间

预备知识　同伦[654]

收缩核映射

首先我们要引入三个依次增强的概念．第一个是**保核收缩映射**，就是将拓扑空间映射到自身的子集（称为核）上，同时还保证核的元素都是映射的不动点；第二个是**形变收缩映射**，它是一种保核收缩映射，但是多了一个条件，要求拓扑空间中各点都连续地移动到核中；第三个是**强形变收缩映射**，它是一种形变收缩映射，同样多了一个条件，要求移动过程中时时刻刻保持核中的元素是不动点．

下面，我们严格描述这三个概念．

定义 1　保核收缩映射

设拓扑空间 X 及其子集 A．如果存在一个连续映射 $f : X \to X$，使得 $f(X) = A$，并且 $\forall a \in A, f(a) = a$，那么称 f 是 X 到 A 的一个保核收缩映射，同时称 A 是 f 的收缩核．保核收缩映射也可以简称保核收缩．

定义 2　形变收缩映射

设拓扑空间 X 及其子集 A．如果 f 是 X 到 A 的一个保核收缩，且 $f \cong i_X$，其中 i_X 是 $X \to X$ 的恒等映射，那么称 f 是 X 到 A 的一个形变收缩映射，也可简称形变收缩．称 A 是 X 的形变收缩核．

从**映射的同伦和空间的同伦**[654] 词条的习题 1 可知，同伦意味着映射的每个点都画出一条道路，或者说连续移动．因此，如果 X 有一个到 A 的形变收缩，那么 X 中的所有点都可以同时、连续地移动到 A 当中．

形变收缩是比保核收缩更强的概念，这就意味着存在某个映射，它是保核收缩，但不是形变收缩．

例 1

设 X 是由平面上的一个开圆盘 A 和一个与 A 不相交的开圆环 B 取并集而得的．简单粗暴地设映射 $f : X \to X$ 满足 $\forall a \in A, b \in B$，存在一个 $a_0 \in A$，

使得 $f(a)=a, f(b)=a_0$. 那么 f 是一个保核收缩，但不是形变收缩——很明显，B 中的点无法通过一条道路连续地移动到 $a_0\in A$ 上.

定义 3　强形变收缩映射

设拓扑空间 X 及其子集 A. 如果 f 是 X 到 A 的一个形变收缩，$H: X\times I\to X$ 是从 i_X 到 f 的同伦，且满足 $\forall t\in I, a\in A$，有 $H(a,t)=a$，那么称 f 是 X 到 A 的一个强形变收缩映射，简称强形变收缩. 称 A 是 X 的强形变收缩核.

强形变收缩要求在 X 的各点连续移动到 A 中时，A 中的点本身保持不动.

从定义可以看出来，形变收缩还意味着 X 和 A 是同伦的空间. 这启发我们对同伦的认识：同伦的空间是能够通过连续变化而变成彼此的. 和同胚不同的是，同伦只要求了连续变化，并不要求双射，这就使得我们可以将多个点映射到同一个点，甚至将拓扑空间降维. 弦论中，形变收缩的性质对弦的世界环性质有重要意义.

可缩空间

可缩空间这一名称很直白，就是指可以被连续压缩为一个点的空间.

定义 4　零伦和可缩空间

设拓扑空间 X 及其上一点 x_0. 如果存在一个从 X 到 x_0 的形变收缩 f，那么称 X 是一个可缩空间，称 f 是一个零伦的映射.

例 2　锥空间

锥空间 $\widetilde{C}X$ 的定义见商拓扑[643] 的“锥空间”小节. 对于任意的拓扑空间 X，其锥空间 $\widetilde{C}X$ 都是可缩的. 事实上，我们可以直接写出这个收缩 $H: CX\times I\to\overline{(x,0)}$ 如下：

$\forall x\in X, t_n\in I$, 有 $H(\overline{(x,t_1)},t_2)=\overline{(x,t_1t_2)}$.

也可以想象成，锥空间逐渐向顶点压缩.

基本群

预备知识　同伦[654]，群论

本节我们将介绍拓扑空间上的一个重要的代数结构，这将会是我们研究同伦（或者说拓扑空间的连续变化）的有力工具．这个结构是一个群，被称为**基本群**．回忆一下，在点集拓扑中我们讨论了四类同胚不变量；基本群的地位是类似的，它是同伦不变量．

我们首先将简单讨论基本群的想法是怎么来的，从直观上定义一个道路运算开始，逐渐调整定义直到满足群的四个公理．

基本群的构造过程

道路类的积

一个自然的想法是，研究拓扑空间中的道路之间的运算，称这个运算为道路之间的积．这种运算首先要**满足封闭性**，即两个道路的积还是一条道路．最简单的情况，就是两条道路首尾相连而成一条新的道路．因此，我们首先定义道路之间的积为首尾相连：

定义 1　道路的积

给定拓扑空间 X 及其上的两条道路：$f,g:I\to X$，并设 $f(1)=g(0)$，即 f 的终点是 g 的起点．定义 f 和 g 的积为 $f*g=h$，其中

$$h=\begin{cases} f(2t), & t\in[0,1/2] \\ g(2t-1), & t\in[1/2,1] \end{cases} \tag{1}$$

显然，h 也是一条道路，并且其起点是 f 的起点、终点是 g 的终点．

如上定义的道路 f 和 g 的积，虽然满足封闭性，但是并**不满足结合性**，这用一个简单的例子就可以说明：

习题 1　道路的积不满足结合性

取通常的一维欧几里得空间 $\mathbb{R}$．设 $\mathbb{R}$ 上有三条道路，$f=t$，$g=t+1$ 和 $h=t+2$．请验证，$(f*g)*h\neq f*(g*h)$．

简单来说，道路的积不满足结合律，是因为虽然 $(f*g)*h$ 和 $f*(g*h)$ 的轨迹是重合的，但是道路上的点走过这条轨迹的“速度”不一样，因此不被视作一条道路．如果我们能把 $(f*g)*h$ 和 $f*(g*h)$ 归入同一个等价类里，转而研究这种道路等价类上的运算，那么结合性就可以满足了．

为了定义这样的等价类，我们不使用一般的同伦，那样的范围太广了，以至于结构太简单；我们使用如下限制的同伦：

定义 2　保起点和终点同伦

如果两条道路 f,g 有相同的起点和终点，$f \overset{H}{\cong} g$，且对于任何 $t \in [0,1]$，都有 $H(0,t) = f(0) = g(0)$ 和 $H(1,t) = f(1) = g(1)$，那么我们称 f 和 g 保起点和终点同伦，记为 $f \dot{\cong} g$.

定义 3　道路类

给定拓扑空间 X.如果两条道路: $f,g : I \to X$ 满足 $f(0) = g(0), f(1) = g(1)$ (即有相同的起点和终点)，并且 $f \dot{\cong} g$，那么将它们归入同一个等价类．这样划分出的等价类，称为 X 上的道路类（**path class**)，记为 $[f]$，其中 f 是 $[f]$ 中任意一条道路.

显然，如果两条道路轨迹相同，那么它们一定是在同一个道路类里的，这就**解决了结合性的问题**.

定义 4　道路类的积

给定拓扑空间 X 和其上的两个道路类 $[f]$ 和 $[g]$．记 $[f]$ 和 $[g]$ 之间的积为 $[f] * [g] = [f * g]$．为了方便，通常也记 $[f] * [g] = [f][g]$.

要注意的是，道路类的积可以省略运算符号 $*$，但是道路的积不可以．因为道路本身还是一种映射，映射之间除了道路积，还有复合等运算；尽管道路之间通常没法复合，但是为了尽可能避免歧义，就不省略道路的积中的 $*$ 了.

基点和回路类

道路类的运算仍然有缺陷，那就是并非任意道路类之间都可以作积，必须是首尾相连的道路才行．解决这个问题很简单，就是选定一个基点，只研究同时以这个基点为起点和终点的道路类就可以了．起点和终点重合的道路，称作**回路**或者**闭路（closed path)**．如果一条回路或者回路类以 x_0 为起点和终点，那么也称 x_0 是该回路或者回路类的基点.

两条回路如果保起点和终点同伦，那么也可以称它们是**保基点同伦**.

由道路连通分支的性质可知，选定基点以后，经过这个基点的回路都逃不出基点所在的道路连通分支．**因此同伦论中研究的通常是道路连通的空间**.

定义 5　回路类的集合

给定道路连通空间 X 及一个基点 $x_0 \in X$．记 $\Omega(X, x_0)$ 为所有以 x_0 为基点的回路类的集合．

基本群的定义

习题 2　基本群

给定道路连通空间 X 及一个基点 $x_0 \in X$．给 $\Omega(X, x_0)$ 赋予道路类的积作为运算．证明这个运算满足群的四条公理．

因此，$\Omega(X, x_0)$ 配合该运算构成一个群，记为 $\pi_1(X, x_0)$，称为 X 上关于基点 x_0 的基本群（**fundamental group**）．

习题 2 的提示：封闭性和结合性，在构造道路类的积时已经满足了——道路类的积这个概念的引入就是为了满足这两条性质的．基本群的单位元是回路类 $[e_{x_0}]$，其中 $e_{x_0} : I \to \{x_0\}$，就是随着 t 变化却恒映射为 x_0 的回路；回路类 $[f(t)]$ 的逆元为 $[f(t)]^{-1} = [f(1-t)]$．

注意，严格的证明需要构造出 e_{x_0} 到 $f * f^{-1}$ 的同伦 H．

基本群的例子

例 1　单连通空间

考虑二维欧几里得空间 $\mathbb{R}^2$．这个单连通空间上的所有道路之间都保起点和终点同伦，因此关于任何基点的回路都保基点同伦，换句话说，关于任何基点的回路类只有一个．因此，$\mathbb{R}^2$ 上关于任何基点的基本群都是平凡群（只含一个元素）．

基本群是平凡群的空间，被称为单连通空间（**simply connected space**）．

任意维的欧几里得空间 $\mathbb{R}^N$ 都是单连通的；这些空间中对应的球面 S^{N-1} 也都是单连通的．

例 2　带一个孔洞的空间

考虑 $\mathbb{R}^2$ 上挖去一个点的空间 $A_1 = \mathbb{R}^2 - \{(0,0)\}$，那么对于任何基点 $x_0 \in A_1$，基本群 $\pi_1(A_1, x_0) = \mathbb{Z}$．其中正整数 n 对应的回路类，是顺时针绕 $(0,0)$ 圈数为 n 的回路集合，$-n$ 对应的就是逆时针绕了 n 圈的回路类，0 对应的就是没有绕 $(0,0)$ 的回路类．当然，也可以反过来定义 n 对应逆时针绕转、

$-n$ 对应顺时针绕转.

如果挖去的不是一个点，而是一个圆，那么基本群也是 $\mathbb{Z}$. 不管挖去的是一个点还是一个单连通的子集，我们都看成是挖去了一个孔洞.

例 3　带多个孔洞的空间

$\mathbb{R}^2$ 上挖去 2 个孔洞，那么考虑回路类时就要区分是绕转了哪个孔洞. 这样一来，挖去两个孔洞后的空间的基本群就是自由群[562]$<\{x_1, x_2\}>$. 一般地，挖去 k 个孔洞以后的空间的基本群是自由群 $<\{x_i | i = 1, 2, \cdots, k\}>$.

基点的选择

在前面的论述中，我们总是说“任意基点”，而论述的内容和具体选择哪个基点无关. 事实上，在道路连通空间中讨论基本群时，确实和基点无关.

定理 1　基点的选取不影响基本群的结构

给定道路连通空间 X 和两个基点 $x_0, y_0 \in X$，则基本群 $\pi_1(X, x_0)$ 和 $\pi_1(X, y_0)$ 是同构的.

证明：

取 x_0 到 y_0 的任意道路 h，则以 y_0 为基点的回路类 $[g]$，可以对应到以 x_0 为基点的回路类 $[h][g][h]^{-1}$. 记这样的对应为 T_h : {以y_0为基点的回路类} $\to$ {以x_0为基点的回路类}，其中 $T_h([g]) = [h][g][h]^{-1}$. 由同伦定义可知 T_h 是一个**双射**.

对于 y_0 为基点的回路类 $[f]$ 和 $[g]$，

$$T_h([f][g]) = [h][f][g][h^{-1}] = [h][f][h]^{-1}[h][g][h]^{-1} = T_h([f])T_h([g])$$

也就是说，回路积和映射可互换顺序，意味着 T_h 是一个**群同态**.

双射的群同态是**群同构**.

证毕.

既然道路连通空间中，基点的不同并不会导致基本群结构的不同，因此我们可以忽视基点的存在，而简单地称任何基点上构造的基本群都是同一个. 在这种语境下，我们可以把道路连通空间的基本群记为 $\pi_1(X)$.

高阶同伦群

预备知识 基本群[658]

基本群的符号被定义为 π_1．你可能好奇为什么要有一个下标 1——这是因为基本群只是同伦群中的一种，而仅仅用 π_1 群来描述一个拓扑空间往往是不够的．

考虑三个拓扑空间：二维球面（如地球表面）S^2，挖去中心的三维球体（如挖去地心的地球）D^3-x_0 以及 n 维几何空间 $\mathbb{R}^n$．容易看出，这三个空间中的任意道路都彼此同伦，因此对任何基点来说都只存在一个回路类，意味着它们的基本群都是平凡群（只有一个元素）．这样一来，对于空间中有几个洞、空间的维度是什么等信息就丢失了，我们就需要推广基本群的概念来描述这些性质．这就是高阶同伦群．

球和球面

在 n 维欧几里得空间中，集合 $\{(x_1,x_2,\cdots,x_n)|\sum_{i=1}^n x_i^2\leq 1\}$ 也被称为半径为 1 的 n 维球体，记为 B^n；集合 $\{x_1,x_2,\cdots,x_n|\sum_{i=1}^n x_i^2=1\}$ 是 B^n 的表面，记为 ∂B^n，或 S^{n-1}．记号 ∂B^n 表达的是“B^n 的边界[622]”，而 S^{n-1} 表达的是“n 维球面”．

更一般地，所有和球同胚的拓扑空间都被看成球，所有和球面同胚的都被看成球面．比如说，n 维空间里的“立方体”，$\{(x_1,\cdots,x_n|x_i\in[0,1],\forall i=1,\cdots,n)\}$，都可以被看成是球．

定理 1　球面粘接

对于任意的正整数 n，把 B^n 表示为一个立方体 $\{(x_1,\cdots,x_n|x_i\in[0,1],\forall i=1,\cdots,n)\}$．$B^n$ 显然是 $\mathbb{R}^n$ 的子空间．在 B^n 上取一个等价关系 $\sim$，使得 B^n 边界上的点都等价，其它点则只和自己等价．则商空间 $B^n/\sim$ 同胚于 S^n．

举个例子，一张平面桌布就是一个 B^2 空间，桌布的边缘是 S^1；如果把桌布边缘粘合成一个点，那么所得的商空间就是 S^2．

n 阶回路及其回路积

基本群是用保基点回路类定义的. 由于道路都是区间 I 到拓扑空间的嵌入映射，而回路是首尾相连的，因此一条回路可以看成是 S^1 到拓扑空间的嵌入映射. 这时，我们也可以把基本群称作 1 **阶同伦群**.

类似地，如果把 S^n 嵌入到拓扑空间中，并将道路的首尾相连运算推广到 n 维情况，那么我们还可以得到 n **阶同伦群**的定义.

1 阶道路的首尾相连运算

由定义 1 [634] 可知，拓扑空间 X 中的一条道路 p 是从区间 I 到 X 内的一个连续映射. 两条道路可以进行首尾相连运算，由定义 1 [658] 定义. 在这里用图像表示首尾相连运算：

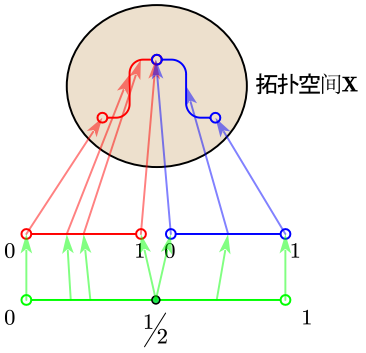


图 1: 道路积的示意图. 图中红色和蓝色分别表示两条道路，红色道路的终点是蓝色道路的起点. 绿色线段也是一个区间 I，它映射到红蓝区间上后再嵌入到 X 上，其中中点同时映射到红色区间的终点和蓝色区间的终点，这意味着把这两个点粘在一起.

如图所示，红色区间到 X 的映射是一条道路，蓝色区间到 X 的映射也是一条道路，它们的首位相连运算所得的道路则是绿色区间映射到红蓝区间后再到 X 的映射. 绿色区间也可以看成红蓝区间粘在一起后，进行适当放缩的结果. 因此，我们也可以把道路的首尾相连运算看成两个区间粘成一个区间后，再作映射.

1 阶回路的回路积

继续参考图 1 ，如果红色道路的起点和终点相同，那么它就是一条回路. 回路也可以看成是区间两端粘在一起再嵌入到 X 中，这样起点和终点肯定相同了. 因此，回路可以看成是 S^1 到 X 的嵌入映射，它的起点和终点统称**基点**.

红蓝回路保基点粘接的过程，就是把红色区间的终点和蓝色区间的起点分别解开后再粘到一起，得到绿色区间，然后再把绿色区间的两端粘起来，得到一个绿色的 S^1，再嵌入 X.

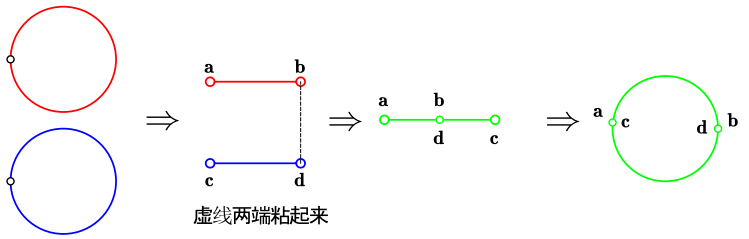


图 2: 一维球 S^1 的保基点粘接过程. 红蓝 S^1 可以看成是红蓝区间把起点 a, d 和终点 b, c 粘在一起的结果. 如果红蓝 S^1 的基点相同（图中分开画了，但是此处认为基点相同），那么可以定义保基点粘接，方法是把基点拆开成起点和终点，然后把红色区间的终点 b 和蓝色区间的起点 d 粘起来，再把红色区间的起点 a 和蓝色区间的终点 c 都粘回基点上.

这样粘接的好处是方便想象. 在进行保基点同伦变换的时候，你可以想象成基点不动，而道路其它点连续运动，期间保持完整道路的形状以及一直在空间 X 中. 把红蓝回路粘接后相当于允许 b 和 d 离开基点运动.

当然，由于道路连通空间中基点的选择不影响基本群的结构，因此我们也可以抽象地把回路的首尾相连看成各自的起始点拆开后再把彼此的首尾相连的过程，图示依然参考图 2 .

2 阶回路的回路积

2 阶回路定义为 S^2 到拓扑空间 X 的嵌入映射. 类似地，为了方便想象，我们把 S^2 看成正方形 I^2 把四条边都粘成一个点的结果. 两条 2 阶回路的粘接过程如图所示：

同样地，这样粘接之后，2 阶回路的保基点同伦变换就可以想象成红蓝边

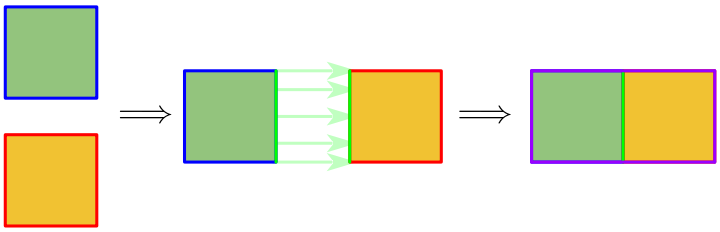


图 3: 绿色和黄色正方形分别是两个 D^2，它们的边界 ∂D^2 分别是蓝色和红色的 S^1．把蓝色和红色的 S^1 都粘到基点上，那么就得到绿黄两个球面 S^2．绿黄球面的粘接，就是把绿黄正方形的某一边（亮绿色表示）从基点处解开，然后把亮绿色边粘在一起，但不是粘成一点，而是两条边粘成一边．

都限制在基点上，但是亮绿色那一边自由了．我们会在接下来专门用一个小节来讨论想象的方法．

n 阶回路的回路积

将 1、2 阶回路的粘接方法推广开，就得到 n 阶回路的定义和它们的“首尾相连”运算：

定义 1　n 阶回路

给定拓扑空间 X，则 X 中一个 n 阶回路是 S^n 到 X 的嵌入映射．

特别地，1 阶回路又称为回路．

定义 2　n 阶回路的粘接

将 S^n 看成超立方体 D^n 将表面 ∂D^n 粘接到一个点的商空间，则两个 S^n 的粘接就是把两个超立方体的某一个超平面解除粘连，然后将这两个平面点对点粘成一个平面．

为了方便想象，回路的粘接还有一个更抽象的等价定义：

习题 1

两个 S^n 的粘接，就是各取一个道路连通子集 D^n，将这两个子集的边界 $\partial D^n = S^{n-1}$ 粘成一个 S^{n-1}，粘连规则是任取一个同胚映射 $f : S^{n-1} \to S^{n-1}$，将 $x \in S^{n-1}$ 和 $f(x)$ 粘在一起．这样粘连以后，再把子集 D^n 挖去，剩下的 S^n 就是球面粘接的结果了．思考这是为什么，图解在下方．

以 S^2 的粘接为例，如图所示.

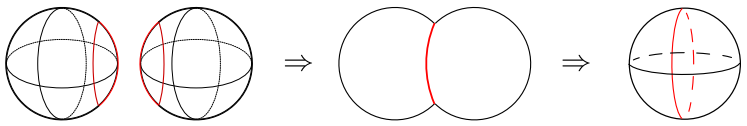


图 4: S^2 的粘接过程. 分别在两个球面上取一小块弧面，弧面的边界用红色表示；将弧面除边界外的部分挖去不用，然后把弧面粘起来，得到一个葫芦表面一样的东西，经过适当的同伦变形得到一个球面.

高阶同伦群

定义 3　n 阶回路类

设 $p_1, p_2 : S^n \to X$ 是拓扑空间 X 中的两条 n 阶回路，它们都经过基点 $x_0 \in X$，那么如果 $H : S^1 \times I \to X$ 是 p_1, phi_2 的同伦，并且对于任何 $t \in I$，同伦 H 都经过基点 x_0，那么称 p_1, p_2 是彼此的保基点同伦. 按照保基点同伦关系划分的等价类，称为回路类.

定义 4　n 阶同伦群

给定拓扑空间 X，记其中 n 阶回路的“首尾相连”运算为 $*$，则对于回路类 $[p_1]$ 和 $[p_2]$，定义其积为 $[p_1][p_2] = [p_1 * p_2]$. n 阶回路类的集合配上回路类的积构成一个群，称为 X 上的 n 阶同伦群，记为 $\pi_n(X)$.

从 S^n 的粘接过程中我们可以看到，当 $n > 1$ 的时候，粘接的方式不唯一，随便取一种同胚都可以进行粘接. 这导致了如下深刻的结果[2]：

定理 2　同伦群交换性定理

当 $n > 1$ 时，$\pi_n(X)$ 是阿贝尔群.

2 阶同伦群的想象方法

本小节介绍 2 阶同伦群的想象方法，展示其和基本群的不同和相似之处，希望能帮助读者体会更高阶同伦群的性质. 遗憾的是，直观的图示只能应用于

[2]文献来源：V.A. Vassiliev，《拓扑学导论（中译本）》2.2 节，高等教育出版社.

π_1 和 π_2 群，而 π_3 群就已经需要四维空间才能描述，更高阶的同伦群也需要更多的维度.

球面的同伦群

预备知识 高阶同伦群[663]

以各球面 S^n 为拓扑空间，则它们的各阶同伦群 π_i 列举如下[3]：（第一行数字表示同伦群的阶数 i，第一列数字表示球面的维度 n）

表 1: S^n 球面的同伦群 π_i（0 表示只有一个元素的平凡群）

	1	2	3	4	5	6	7	8	9	10	11	12
1	$\mathbb{Z}$	0	0	0	0	0	0	0	0	0	0	0
2	0	$\mathbb{Z}$	$\mathbb{Z}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_{12}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_3$	$\mathbb{Z}_{15}$	$\mathbb{Z}_2$	$\mathbb{Z}_2 \times \mathbb{Z}_2$
3	0	0	$\mathbb{Z}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_{12}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_3$	$\mathbb{Z}_{15}$	$\mathbb{Z}_2$	$\mathbb{Z}_2 \times \mathbb{Z}_2$
4	0	0	0	$\mathbb{Z}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z} \times \mathbb{Z}_{12}$	$\mathbb{Z}_2 \times \mathbb{Z}_2$	$\mathbb{Z}_2 \times \mathbb{Z}_2$	$\mathbb{Z}_{24} \times \mathbb{Z}_3$	$\mathbb{Z}_{15}$	$\mathbb{Z}_2$
5	0	0	0	0	$\mathbb{Z}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_{24}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_{30}$
6	0	0	0	0	0	$\mathbb{Z}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_{24}$	0	$\mathbb{Z}$	$\mathbb{Z}_2$
7	0	0	0	0	0	0	$\mathbb{Z}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_{24}$	0	0
8	0	0	0	0	0	0	0	$\mathbb{Z}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_{24}$	0

该表格有两个很明显的规律：第一，当同伦群的阶数 i 小于球面的维度 n 的时候，同伦群一定是平凡群；第二，当同伦群的阶数等于球面的维度的时候，同伦群一定是正整数群 $\mathbb{Z}$.

覆叠空间

预备知识 连续映射和同胚[625]

覆叠空间是一种常见的简化空间描述的方式，可以更轻松地描述许多复杂

[3]文献来源：Allen Hatcher, Algebraic Topology, 2002 by Cambridge University Press, pp.339(Section 4.1).

空间的性质．覆叠的思想在微分几何中极为重要，而在理论物理中也偶尔会使用该方法．

覆叠的概念

定义 1　覆叠映射和覆叠空间

设有拓扑空间之间的满连续映射 $p : C \to X$．如果对于任意的 $x \in X$，存在开集 $U_x \in \mathcal{T}_X$ 且 $x \in U_x$，使得 $p^{-1}(U_x)$ 是 C 中若干不相交开子集的并，并且每个这样的开子集都通过 p 和 U_x 同胚，那么称 C 是 X 的覆叠空间（**covering space**），p 是其覆叠映射（**covering map**）或覆叠投影（**covering projection**）开集 U_x 是点 x 的典范邻域．

覆叠映射 p 的逆把典范邻域 U_x 映射为若干个和 U_x 同胚的开集，这也可以表示为存在一个**离散空间** F_x 使得 $p^{-1}(U_x) \approx U_x \times F_x$．

需要注意的是，覆叠空间并不保证任何邻域都是典范邻域，即不是任何开集 U 的逆映射 $p^{-1}(U)$ 都同构于 $U \times F$．

如果 X 是一个连通空间的话，我们还可以得知对于任何典范邻域 U_x 对应的离散空间 F_x，其基数 $|F_x|$ 都是一样的．这由以下定理严格描述：

定理 1　连通空间的覆叠

设 $p : C \to X$ 是一个覆叠映射，X 是连通的．对于任意的 $x, y \in X$，如果记它们的典范邻域分别是 U_x 和 U_y，且 $p^{-1}(U_x) \approx U_x \times F_x$ 和 $p^{-1}(U_y) \approx U_y \times F_y$，则有 $|F_x| = |F_y|$．

定理 1 意味着如果 X 连通，那么其覆叠映射把它同构到 $|F_x|$ 个自身的拷贝上去了，其中 F_x 是任意点 $x \in X$ 的典范邻域对应的离散空间．

覆叠映射的例子

例 1

取实度量空间 $\mathbb{R}$ 和复平面上的单位圆 $S^1 = \{e^{2\pi i t} \in \mathbb{C} | t \in \mathbb{R}\}$．建立映射 $p : \mathbb{R} \to S^1$，其中 $p(t) = e^{2\pi i t}$，则 p 是一个覆叠映射，$\mathbb{R}$ 是 S^1 的覆叠空间．

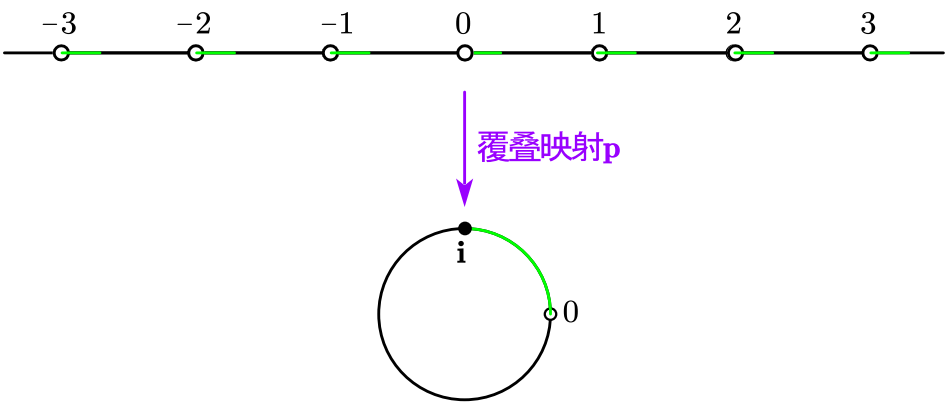


图 1: $\mathbb{R}$ 到 S^1 的覆叠映射示意图．绿色段分别表示 S^1 上的一个典范邻域和它关于覆叠映射 p 的原像．

基本群的计算

预备知识 基本群[658]，可缩空间[657]，群的自由积[564]

虽然把各阶同伦群都考虑进去以后可以很详细地刻画空间的伦型，但是高阶同伦群大多极其难计算．本节简单介绍一阶同伦群，即基本群的计算方法．

基本群计算定理

我们列举一些方便用于计算基本群的定理如下．

定理 1　积空间的基本群

给定拓扑空间 X 和 Y，则 $\pi_1(X\times Y)=\pi_1(X)\times\pi_1(Y)$.

定理 2　Seifert-van Kampen 定理

设拓扑空间 X 可以被它的两个开集 U_1 和 U_2 覆盖，即 $X=U_1\cup U_2$；若 U_1,U_2 和 $U_1\cap U_2$ 都是道路连通空间，取 $U_1\cap U_2$ 中一个点作为基点来构造各空间的基本群．设 $f_i:U_1\cap U_2\to U_i$ 为恒等嵌入，即 $\forall x\in U_1\cap U_2, f_i(x)=x\in U_i$．如果用 f_i 来定义 $\pi_1(U_1\cap U_2)$ 到 $\pi_(U_i)$ 上的同态，那么有：$\pi_1(X)=\pi_1(U_1)*_{\pi_1(U_1\cap U_2)}\pi_1(U_2)$.

Seifert-van Kampen 定理难以简洁表达，不过如果能充分理解**群的自由积**[564]，应该容易理解该定理．不过，我们可以考虑该定理的弱化版本，此版

本用处也很广泛：

定理 3　弱化版 Seifert-van Kampen 定理

设拓扑空间 X 可以被它的两个开集 U_1 和 U_2 覆盖，并且 $U_1 \cap U_2$ 是单连通（例 1 [658]）的，那么有 $\pi_1(X) = \pi_1(U_1) * \pi_1(U_2)$.

除此之外，我们在**可缩空间**[657] 中提到的形变收缩也能用于大大简化基本群的计算.

定理 4　形变收缩核的同伦

给定拓扑空间 X. 如果 $f : X \to A \in X$ 是一个形变收缩，且 A 是其收缩核，那么 $X \cong A$.

证明很简单，由形变收缩的定义可知，$f : X \to A$ 和 $1_X : A \to X$ 是彼此的同伦逆；1_X 是 X 到自身的恒等映射，由于此处将其限制在 A 上，也可以记为 $1_A = 1_X|_A$.

定理 4 很好地展示了同胚和同伦的区别：如果存在一个形变收缩，把拓扑空间压成更低维度的情况，比如将烟卷压成圆环，那么收缩前后的空间是同伦的，但它们由于维度不同，就不会同胚.

定理 5　锥空间的基本群

设 X 是任意非空的拓扑空间，则 $\pi_1(\widetilde{C}X) = \{e\}$，即为只有一个元素的平凡群.

锥空间的基本群总是平凡群，也就是说所有回路都是保基点同伦的. 这是因为，首先锥空间一定是道路连通空间，因此基点可以任意选择，不妨选为锥顶点；其次，任何一条回路都可以通过各点沿着锥空间的 I 分量连续地收缩到锥顶点上，从而和恒等于锥顶点的回路同伦.

基本群计算实例

预备知识　复数[142]，覆叠空间[668]

S^1 是用于构建许多拓扑空间的伦型的原料，因此严格证明 S^1 的基本群非常重要. 例 1 提供了严格证明的大体思路，限于篇幅，只能由有兴趣的读者自行补充细节了.

例 1　S^1 的基本群

将 S^1 看成复平面上的单位圆 $\{e^{2\pi it} \in \mathbb{C} | t \in \mathbb{R}\} = \{x + yi \in \mathbb{C} | x = \cos 2\pi t, y = \sin 2\pi t\}$. 取通常的实度量空间 $\mathbb{R}$，建立映射 $p : \mathbb{R} \to S^1$，其中 $p(t) = e^{2\pi it}$，则 p 是一个覆叠映射.

把 S^1 看成 $\mathbb{R}$ 的商拓扑空间，其中等价关系 $\sim$ 为：$x \sim y \iff |x - y| \in \mathbb{Z}$，即把实数轴绕到圆上. 对于圆周上任意一个点 $e^{2\pi it_0}$，可以取典范邻域 $U_{t_0} = \{e^{2\pi it} | t \in (t_0 - 1/4, t_0 + 1/4)\}$. 这个典范邻域中 $1/4$ 的选择是任意的，换成任何小于 $1/2$ 的正数也可以，我们只需要用该典范邻域来说明接下来定义的提升映射 $\tilde{f}$ 是唯一的.

取 S^1 的基点为 $p(0) = 1$，设 S^1 中有一条道路 $f : I \to S^1$. 我们可以把 f 提升为 $\mathbb{R}$ 中的道路 $\tilde{f} : I \to \mathbb{R}$，使得 $f = p \cdot \tilde{f}$. 如果取定 $\tilde{f}(0) = 0$，那么这种提升是唯一的（为什么？结合 U_{t_0} 的性质想一想）. 这样，我们就可以通过唯一的提升，把道路 f 都表示为 $\tilde{f}$. 这样做的好处是，$\tilde{f}$ 的图像容易画出来.

任意回路 $f : I \to S^1$ 表示为 $\tilde{f} : I \to \mathbb{R}$ 后，必然是 $I \times \mathbb{R}$ 平面上，从点 $(0, 0)$ 出发，结束于 $(1, f(1))$ 的一段连续函数，其中 $f(1)$ 是一个整数[4]. 在这个例子的情况下，函数 g 和 f 保基点同伦当且仅当 $\tilde{g}$ 的图像也是从点 $(0, 0)$ 出发，结束于 $(1, f(1))$ 的连续函数.

有了以上约定，我们就可以把 S^1 中的道路积表示如图.

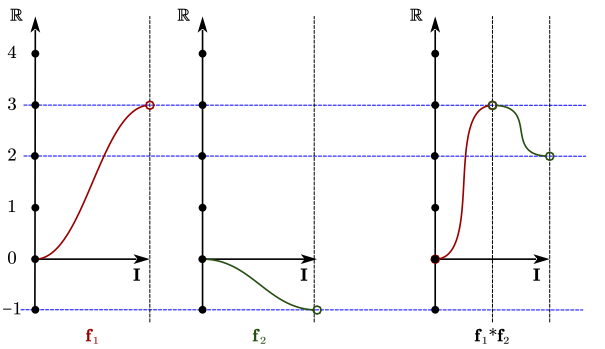


图 1: S^1 上的回路 f_1 和 f_2，以及它们的回路积 $f_1 * f_2$.

记按照以上约定所得到的 $\tilde{f}(1)$ 为回路 f 的度数，记为 $\deg f$. 由图 1 易知，保基点同伦的回路都有相同的度数，并且 $\deg f_1 * f_2 = \deg f_1 + \deg f_2$.

因此，容易得出一维球面的基本群：$\pi_1(S^1) = \mathbb{Z}$.

[4]因为 f 要构成回路，$f = p \cdot \tilde{f}$，而只有整数能被 p 映射到 S^1 的基点上.

例子中的讨论是较为严谨的思路，实际上可以直观地把 $\pi_1(S^1)$ 中的各元素（回路类）看成是顺时针或逆时针绕过整个圆周 n 周后回到基点的回路所构成的类，其中 n 是一个整数，因此基本群同构于整数加群．比如说，如果选定逆时针为正方向，那么顺时针旋转 2 圈后回到基点的回路类就对应于 -2.

例 2　和 S^1 有关的空间中的基本群

由于 $\pi_1(S^1)=\mathbb{Z}$，结合本节所述定理，我们可以轻松计算如下空间的基本群：

- 甜甜圈空间可以表示为 $S^1\times S^1$，因此其基本群为 $\pi_1(S^1\times S^1)=\mathbb{Z}\times\mathbb{Z}$.
- 高维甜甜圈 $S^1\times\cdots\times S^1$ 的基本群是 $\mathbb{Z}\times\cdots\times\mathbb{Z}$.
- 二阶圈图 $S^1\vee S^1$（定义 4 [649]）的基本群是 $\mathbb{Z}*\mathbb{Z}$.
- 高阶圈图 $S^1\vee\cdots\vee S^1$ 的基本群是 $\mathbb{Z}*\cdots*\mathbb{Z}$.

例 3　莫比乌斯带的基本群

莫比乌斯带 M 可以通过强形变收缩（定义 3 [657]）来同伦于其中轴线 S^1，因此莫比乌斯带的基本群和 S^1 一样，都是 $\mathbb{Z}$.

理解莫比乌斯带基本群的难点是看出中轴线是其强形变收缩核，为了方便描述如何构造对应的强形变收缩，我们把莫比乌斯带看成射影平面挖去中心的结果，如所示.

由于射影平面本身可以看成一个圆盘 B^2 的商拓扑空间，我们同样可以把 M 看成圆盘 B^2 挖去中心后再取商拓扑，也就是看成圆环的商拓扑空间．不失一般性地，把这个圆环看成半径为 2 的圆和半径为 1 的圆之间的部分．取映射 $f:M\times I\to M$，其中对于任意 $(x,y)\in M, t\in I$，都有 $f((x,y),t)=(1+t/(1-\sqrt{x^2+y^2}))(x,y)$，则 f 是一个强形变收缩，其收缩核就是圆环的外边缘；取商拓扑可见，这个外边缘正是莫比乌斯带的中轴线.

习题 1　莫比乌斯带的强形变收缩

验证例 3 中的 f 是同伦，进而证明它确实是强形变收缩．提示：先考虑未将圆环外边缘的对径点粘合时的情况，再对比考虑粘合后的效果.

习题 2

将莫比乌斯带看成一个射影平面挖去中心后的空间．如果将沿着外边缘走

了半圈的回路类[5]记为 $a \in \pi_1(M)$，那么绕着中心缺口顺时针旋转一周的回路类是 $\pi_1(M)$ 中的哪个元素？

答案是 a^2.

[5]走了半圈就回到了起点，因为对径点粘在一起了.

第六部分

微分几何

第一章

欧几里得空间

三维欧几里得空间中的曲线

预备知识 矢量的导数[274]，张成空间[447]

三维欧几里得空间是我们最为熟悉的空间，其性质也非常好，易于研究. 本节讨论的是三维空间中的曲线的性质，引入曲率、挠率等概念，介绍了 Frenet 向量组以及该组中向量之间的关系.

一般欧几里得空间中曲线的概念

定义 1　曲线

令 I 是实数轴 $\mathbb{R}$ 上的一个区间，则称连续函数 $f: I \to \mathbb{R}^n$ 为 $\mathbb{R}^n$ 中的一条曲线（**curve**），按拓扑学的习惯也称道路（**path**）. 此处连续是指函数的 n 个分量都是 $\mathbb{R} \to \mathbb{R}$ 的连续函数.

我们可以任意取定一个坐标系，把向量值函数 f 分为 n 个标量值函数，简称为 f 的分量，由此来理解连续的含义. 你可能自然会想确认，f 的分量的连续性，和取定坐标系的方式是否有关？答案是无关的. 这是因为我们可以用另一种方式来理解此处的“连续”，那就是取集合 I 与 $\mathbb{R}^3$，配上通常的拓扑——即 I 取 $\mathbb{R}$ 的子拓扑，$\mathbb{R}^3$ 取 $\mathbb{R}$ 的乘积拓扑——所得到的拓扑空间，那么 f 就是拓扑空间之间的映射，其连续性取决于拓扑意义上的连续性，和具体坐标系的选择就无关了.

要强调的一点是，即便两条曲线的轨迹一样，它们也不一定是同一条曲线. 比如说，取 $f, g: \mathbb{R} \to \mathbb{R}^2$ 这两个函数，定义为 $f(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$ 和 $g(t) = \begin{pmatrix} \cos 2t \\ \sin 2t \end{pmatrix}$，那么尽管它们的轨迹都是平面上的单位圆，但由于两条曲线的“速度”不一样，我们依然把它们认为是不同的曲线.

定义 2　可微曲线

如果曲线 $f: I \to \mathbb{R}^n$ 对于任意分量都是可微的，那么称 f 是一个可微曲线（**differentiable curve**）或可微道路（**differentiable path**），有时也记为 C^1 的曲线.

同样地，f 本身的可微性，和具体坐标系的选择有关系吗？由于“微分”这一概念是实数空间特有的，我们没法像前面一样直接用拓扑的概念绕过坐标系的选择；但答案是一样的，**无关**．证明这一点的思路也可以应用到之前对一般曲线的讨论上去：考虑变换坐标系后各点的变换，会发现变换后的分量就是过渡矩阵乘以原先的函数列矩阵，也就是说，变换坐标后新的函数分量，是旧分量的某种线性组合．这样一来，变换前可微当且仅当变换后也可微，从而可知和变换无关．

定义 3　光滑曲线

如果曲线 $f : I \to \mathbb{R}^n$ 对于任意分量都是光滑的，那么称 f 是一个光滑曲线（**smooth curve**）或光滑道路（**smooth path**），有时也记为 C^∞ 的曲线．

和上述坐标变换的讨论类似，可证明光滑曲线的光滑性不依赖于坐标系的选择．

定义 4　正则曲线

Frenet 向量组

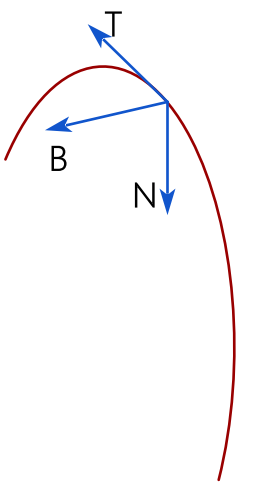


图 1: Frenet 向量组示意图．途中三个蓝色向量就是红色曲线在某一点处的 Frenet 向量组，分别是单位切向量 T、单位主法向量 N 和单位副法向量 B．

Frenet 向量组是**三维**欧几里得空间 $\mathbb{R}^3$ 中特有的概念，是刻画**可微曲线**性质的优良工具.

Frenet 向量组中的第一个向量，是**单位切向量**. 关于 t 对曲线 $f(t)$ 求导就能得到切向量，如果把 t 看作时间，那么切向量 $\frac{\mathrm{d}}{\mathrm{d}t}f(t)$ 也可以看成是曲线上一点的切速度. 单位切向量就是方向与切速度相同，但长度为 1 的向量.

为了得到单位切向量,我们可以用切速度除以切速度的大小来获得:$\frac{\mathrm{d}}{\mathrm{d}t}f(t)/\left|\frac{\mathrm{d}}{\mathrm{d}t}f(t)\right|$. 但更简单的方法是，令切速度的大小恒为 1，或者说只讨论切速度大小恒为 1 的**可微曲线**. 当切速度恒为 1 的时候，曲线走过的弧长就恒等于经过的时间，于是我们说这时是用**弧长**来作曲线的参数. 弧长参数通常用字母 s 表示.

定义 5　单位切向量

使用弧长参数来描述可微曲线 $f(s)$，记 $T(s)=\frac{\mathrm{d}}{\mathrm{d}s}f(s)$，称为曲线在参数为 s 处的单位切向量（**unit tangent vector**）.

Frenet 向量组中的第二个向量，是**单位主法向量**. 对于处处可微的曲线，切向量总是存在的，但是主法向量不一定存在. 主法向量的定义如下：

定义 6　主法向量

使用弧长参数来描述可微曲线 $f(s)$，其单位切向量为 $T(s)$. 如果 $\frac{\mathrm{d}}{\mathrm{d}s}T(s)$ 不为零，则称 $\kappa(s)N(s)=\frac{\mathrm{d}}{\mathrm{d}s}T(s)$ 为曲线在 s 处的主法向量（**normal vector**）. 其中 $N(s)$ 是单位向量，κ 是一个正数，称作曲线在这一点处的曲率（**curvature**）.

由于使用弧长参数以后，单位切向量的大小不变，因此 $N(s)$ 恒与 $T(s)$ 垂直. $N(s)$ 就是那第二个成员，单位主法向量.

Frenet 向量组中的最后一个成员，是**单位副法向量**，是用前两个向量叉乘而来的.

定义 7　副法向量

条件设定如定义 5 和定义 6，称 $B(s)=T(s)\times N(s)$ 为曲线在 s 处的单位副法向量（**unit binormal vector**）.

以上定义的副法向量必然是单位向量，因为 T 和 N 都是单位向量且正交.

Frenet 向量组可以用于构成所谓的 **Frenet 坐标架**，也称**活动坐标架**. 相比预先规定的自然坐标系，活动坐标架忽略了曲线在空间中的具体位置，更简洁和方便地表示了曲线本身的局部性质.

定义 8　密切平面

条件设定如定义 5 和定义 6，由 $T(s)$ 和 $N(s)$ 两向量在点 $f(s)$ 处所张成的二维空间，称为曲线在这一点处的密切平面（**osculating plane**）.

对于**平面曲线**，即整体都在一个平面上的曲线，其所在的平面就是每个点的密切平面.

曲率和扭率

曲率已经在定义 6 中定义清楚了，是表征单位切向量旋转速率的数字. 要说明的是，如果单位切向量在某点处不变化，那么主法向量的概念就不复存在，也就没有 Frenet 向量组的后两位成员了. 在这种情况下，我们说曲线在这一点处的曲率为. 更一致的曲率定义方法如下：

定义 9　曲率

使用弧长参数来描述可微曲线 $f(s)$，其单位切向量为 $T(s)$. 称 $\kappa(s) = \left|\frac{\mathrm{d}}{\mathrm{d}s}T(s)\right|$ 为曲线在 s 处的曲率（**curvature**）.

当 Frenet 向量组存在时，还有一个重要的概念叫**扭率**或**挠率**，在有的文献中也会被称为**第二曲率**.

在定义扭率之前，我们要先讨论副法向量的一个性质：$\frac{\mathrm{d}}{\mathrm{d}s}B(s)$ 和 $N(s)$ 平行. 首先，因为单位副法向量的长度不变，故有 $\frac{\mathrm{d}}{\mathrm{d}s}B(s)$ 和 $B(s)$ 垂直. 由于 $B = T \times N$，故 $\frac{\mathrm{d}}{\mathrm{d}s}B(s) = [\frac{\mathrm{d}}{\mathrm{d}s}T(s)] \times N(s) + T(s) \times \frac{\mathrm{d}}{\mathrm{d}s}B(s) = T(s) \times \frac{\mathrm{d}}{\mathrm{d}s}B(s)$，进而可知 $\frac{\mathrm{d}}{\mathrm{d}s}B(s)$ 垂直于 $T(s)$. 因此，$\frac{\mathrm{d}}{\mathrm{d}s}B(s)$ 垂直于 T 和 B，必然和 N 在一条线上了.

定义 10　扭率

使用弧长参数来描述可微曲线 $f(s)$，令其 Frenet 向量组为 $T(s), N(s), B(S)$. 令实数 τ 满足 $\frac{\mathrm{d}}{\mathrm{d}s}B(s) = -\tau N(s)$，那么称实数 τ 为曲线在 s 处的扭率（**torsion**）.

曲率描述了曲线单位切向量方向变化的速率，越快则曲线弯曲程度越强. 扭率则描述了曲线偏离平面曲线的程度，也就是其扭曲程度；也可以理解为，扭率描述了密切平面旋转的速率.

Frenet-Serret 公式

三维欧几里得空间中的可微曲线 $f(s)$，如果处处有非零曲率 $\kappa(s)$，且其扭率为 $\tau(s)$（可以为零），则它一定有 Frenet 向量组．T，N 和 B 具有如下关系：

定理 1　Frenet-Serret 公式

对于曲线的 Frenet 向量组 $\{T(s), N(s), B(s)\}$，我们有如下关系：

$$\frac{\mathrm{d}}{\mathrm{d}s}T(s) = \kappa N(s) \tag{1}$$

$$\frac{\mathrm{d}}{\mathrm{d}s}N(s) = -\kappa(s)T(s) + \tau(s)B(s) \tag{2}$$

$$\frac{\mathrm{d}}{\mathrm{d}s}B(s) = -\tau(s)N(s) \tag{3}$$

用矩阵的形式可以更简洁地表示为

$$\begin{pmatrix} T' \\ N' \\ B' \end{pmatrix} = \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix} \begin{pmatrix} T \\ N \\ B \end{pmatrix} \tag{4}$$

矩阵表示也更好记忆．

证明：

1. 根据 N 的定义，直接可得式 1．
2. 根据扭率的定义，直接可得式 3．
3. 考虑到三个向量的方向关系，以及它们都是单位向量的事实，我们可以得到 $N = B \times T$．因此，$\frac{\mathrm{d}}{\mathrm{d}s}N = \frac{\mathrm{d}B}{\mathrm{d}s} \times T + B \times \frac{\mathrm{d}T}{\mathrm{d}s} = -\tau N \times T + B \times \kappa N = \tau B - \kappa T$．

证毕．

光滑映射

预备知识 线性映射[451]，矢量的导数[274]，欧几里得空间

欧几里得空间作为线性空间，彼此之间可以进行线性映射；同时，欧几里得空间中可以进行微分操作，因此我们还可以研究这些线性映射的微积分性质．本节讨论一类可以无穷次求（偏）导的线性映射，称为光滑映射．

为了讨论光滑映射，我们首先引入光滑函数的概念．

光滑函数

定义 1　实数集上的光滑函数

给定函数 $f:\mathbb{R}\to\mathbb{R}$，如果在某一点 $x_0\in\mathbb{R}$，对于任意的 $k\in\mathbb{Z}^+$，存在 f 的 k 阶导数 $f^{(k)}(x_0)$，则称 f 是在点 x_0 处任意阶可导的，或 C^∞ 的，或光滑的（**smooth**）．今后，我们将术语 C^∞ 和光滑看成完全等价的．如果 f 处处光滑，则称其本身是一个光滑函数．

光滑函数的概念可以通过偏导数来推广到多自变量的情况．

定义 2　多元光滑函数

给定函数 $f:\mathbb{R}^n\to\mathbb{R}$，如果在某一点 $(x_0^1,x_0^2,\cdots x_0^n)\in\mathbb{R}^n$，对于任意的 $k\in\mathbb{Z}^+$，存在 f 的 k 阶偏导数

$$\frac{\partial^k}{\partial x^{i_1}\partial x^{i_2}\cdots\partial x^{i_k}}f(x_0^1,x_0^2,\cdots x_0^n)$$

其中各 i_j 都是 1 到 n 的任意整数，可以有重复；则称 f 在 $(x_0^1,x_0^2,\cdots x_0^n)$ 处光滑．如果 f 处处光滑，则简称 f 光滑．

光滑映射

两个欧几里得空间之间的线性映射 $f:\mathbb{R}^m\to\mathbb{R}^n$，可以看成是 m 个自变量的 n 维向量值函数；进一步，也可以把向量看成 n 个标量，也就是说把这个线性映射看成 n 个线性函数组合在一起．

定义 3　光滑映射

设 $f:\mathbb{R}^m\to\mathbb{R}^n$ 是欧几里得空间之间的线性映射．对于 $\forall\mathbf{x_0}\in\mathbb{R}^m$，记 $f(\mathbf{x_0})=(f_1(\mathbf{x_0}),f_2(\mathbf{x_0}),\cdots,f_n(\mathbf{x_0}))$．如果每一个 f_i 都是光滑函数，则称 f 是一个光滑映射（**smooth map**）．

光滑函数和解析函数的比较

解析函数是一种性质良好的函数，它可以表示成幂级数的形式．如果一个函数解析，那么我们总可以在任何点处对它进行泰勒展开，展开结果就是其幂级数表达．能进行泰勒展开固然需要可以无穷求导，也就是说解析函数必然是光滑的；但反过来却不成立，有的光滑函数就不解析．

例 1

考虑函数 $f:\mathbb{R}\to\mathbb{R}$，其中当 $x>0$ 时 $f(x)=\exp\{(-1/x)\}$，当 $x\leq 0$ 时 $f(x)=0$．可以验证，f 处处光滑，但是它在 $x=0$ 处的泰勒展开是常数 0，而它显然不是一个常函数，因此 f 在 $x=0$ 处不解析．

切空间（欧几里得空间）

预备知识 偏导数[266]，光滑映射[681]

切空间是切向量的集合．从几何直观来说，切向量就是和某个平面或者超平面相切于一点的向量．

在线性代数中，我们将原点默认为所有向量的起点，这样只需要终点就可以表示向量了，因此向量被一一对应到点上．这种方式获得的向量，实际上是更广义的切向量中的一类，即从原点“发射”出去的向量．我们也可以空间中的其它点作为起点，来得到发射出去的向量．当然，如果把向量理解成“具有长度和方向的量”以及此概念的推广，那么向量的起点是无关紧要的；而切向量是区分了起点的．

为什么叫做切向量呢？我们将在讨论光滑流形在欧几里得空间中的嵌入时看到原因．

从一个点 P 发射出去的向量，被称为点 P 处的**切向量**．因此，线性代数中所研究的空间可以看成是原点处的切向量构成的空间，称作**原点处的切空间**．

如果一个欧氏空间被嵌入到更高维的一个空间中，比如说将 $\mathbb{R}^2$ 以抛物面的形状嵌入到 $\mathbb{R}^3$ 中，那么切空间的意义就非常直观了：点 P 处的切空间就是 $\mathbb{R}^3$ 中的一个与该抛物面切于点 P 的平面，取点 P 作为该平面的原点，那么这个平面也可以看成一个二维实线性空间 $\mathbb{R}^2$．

切空间的定义

几何向量定义

对于一个欧几里得空间 $\mathbb{R}^n$，我们可以简单地把某点 $x \in \mathbb{R}^n$ 处的切向量定义为以该点为起点的向量，而所有该点处的切向量构成一个 x 上的切空间．显然，切空间就是一个线性空间．

几何向量的定义在欧几里得空间里非常直观，但是当我们讨论曲面，或者更一般地说，流形上的切向量时，就不是那么方便了，因此我们会采用以下定义方式．

道路定义

如果在欧几里得空间 $\mathbb{R}^n$ 中有一个曲面 S，假设它是一个光滑的曲面，也就是说可以把它看成某个光滑函数的等值面[1]，那么我们可以使用道路来求出曲面某一点处的**切向量（tangent vector）**．

取点 $p \in S$，令 $r : I \to S$ 是一条光滑[2]的道路，且满足 $r(0) = p$，那么 $r(t)$ 是从 x 处出发的一条向量．从几何的角度来说，这个向量不一定是曲面 S 的切向量，它有可能穿过 S；不过从微积分中我们知道，极限 $\lim\limits_{t\to 0} \frac{r(t)}{|r(t)|}$ 是 p 处的一个切向量．用这个极限计算出来的切向量涵盖了所有可能的方向，但都是**单位向量**；为了讨论 p 处的整个切空间，可以改用 $\lim\limits_{t\to 0} \frac{r(t)}{t}$ 来定义，这样所有长度的切向量都存在了，甚至包括无穷长度的切向量．

如果一条道路 r 按以上定义导出的向量存在[3]且有限长，那么我们称 r 是**收敛**的．如果导出的向量不存在或者长度为无穷，那么称 r 是**发散**的．如果道路 r 导出切向量 $\mathbf{v}$，我们也说 r **收敛于** $\mathbf{v}$．

不同的道路可能收敛同一个向量．我们把收敛于同一个切向量的道路称为等价的．这样一来，我们就可以对道路进行等价划分，同一个等价类的道路收敛于同一个切向量，不同等价类的道路收敛于不同的切向量．于是，我们就可以把这样定义的道路等价类和切向量一一对应；或者说，**称道路等价类为向量**．

---

[1] 对于函数 f，任意数值 $a \in \mathcal{F}$，称 $f^{-1}(a)$ 是 f 的一个等值面．

[2] 由于预备知识中光滑的定义依赖于欧氏空间，这里的光滑是指 r 作为 $I \to \mathbb{R}^n$ 的映射而言的．

[3] 光滑道路必然存在导出的向量；虽然我们只考虑光滑道路，但并不妨碍将道路导出切向量的思想应用到一切道路上，这样的话有可能有的道路并不存在导出的向量．

对于道路 r，把 $[r]$ 记为所有和 r 等价的道路之集合，也就是 r 所在的等价类.

一个点处的切向量全体构成一个**切空间（tangent space）**，这是一个线性空间. 两个道路等价类的和该怎么计算呢？在两个道路等价类 $[r_1]$ 和 $[r_2]$ 中任取代表道路 r_1 和 r_2，那么令 $r(t)=r_1(r)+r_2(t)$，得到和道路 r，那么可定义道路等价类的向量加法：$[r_1]+[r_2]=[r]$. 类似地，可以定义数乘：$a[r]=[ar]$.

不仅是在 S 上，在欧几里得空间 $\mathbb{R}^n$ 本身中也可以用道路定义切向量，这个定义和几何向量定义的切向量是一致的.

导子定义

同样设在欧几里得空间 $\mathbb{R}^n$ 中有一个光滑曲面 S，取点 $p\in S\subseteq\mathbb{R}^n$. 令 $\partial/\partial x_i$ 表示对 $\mathbb{R}^n$ 的第 i 个坐标求偏导数，$\partial/\partial x_i|_p$ 表示在点 p 处求这个偏微分. 那么如果在 p 的某邻域内 S 的部分可以看成是一个函数 $f:\mathbb{R}^{n-1}\to\mathbb{R}$，则 $\partial/\partial x_i|_p f$ 是 S 在 p 处的一个切向量. 全体形如 $\sum a_i\cdot\partial/\partial x_i|_p f$ 的向量构成了 p 处的切空间.

进一步地抽象，不考虑具体的 f 形式，则 $\partial/\partial x_i|_p$ 是 S 在 p 处的一个切向量. 全体形如 $\sum a_i\cdot\partial/\partial x_i|_p$ 的向量构成了 p 处的切空间.

$\partial/\partial x_i|_p f$ 表示的是一个偏导数的值，而 $\partial/\partial x_i|_p$ 是表示求偏导数这一操作，因而被称作算子. 具体地，在这个例子里，求偏导数的操作被称为一个**导子（derivation）**，任何导子相加或乘以某个数字的结果还是导子. 导子之间自然引出的向量加法和数乘使它们构成了一个线性空间. 导子和偏导数的关系，就像电场和静电力的关系一样. 在电场中放入一个测试电荷，测试电荷就会受到电场施加的静电力，但我们通常研究的是电场本身，测试电荷只是方便描述其性质，并不是关键；同样地，取一个测试函数，用导子作用于测试函数就可以得到一个数值，但我们研究的是导子本身，要摆脱具体的测试函数.

不仅是在 S 上，在欧几里得空间 $\mathbb{R}^n$ 本身中也可以用导子定义切向量，此时导子定义、道路定义和几何向量定义的切向量是一致的.

三种定义的联系

对于流形上的图 (U,φ) 或流形上的某个嵌入 i，沿着相应道路对时间 $t\in I$ 求导，可以得到 $\varphi(U)$ 中的向量，即几何向量定义；也可以得到 $i(U)$ 上的切向量，这是几何向量定义难以清楚描述的. 道路定义就是沿着道路求方向导数，

但将直线道路取代成任意光滑道路，所求的方向导数是光滑道路在起点处切线的方向．这是因为我们对图的选择是任意的，一个图中的直线道路到另一个图中不一定还是直线，但一定是光滑道路．而导子本身就是在求方向导数．

今后，当我们提到切向量时，总是指道路定义和导子定义，有时会交替使用，读者应熟悉两种定义之间的等价性．

切丛

详见**切丛**[698]．

欧氏空间中任意一点处都有切空间，所有这些切空间构成的集合，被称为一个**切丛（tangent bundle）**．$\mathbb{R}^n$ 上每个点的切空间都同构于 $\mathbb{R}^n$，因此切丛可以看成集合 $\mathbb{R}^n \times \mathbb{R}^n = \mathbb{R}^{2n}$．

物理学中有一个很常见的切丛的实例，那就是切向速度空间．给定一个时空，一个质点在时空中每一个点处都可以取不同的瞬时速度，在每个点上可能的瞬时速度的集合，就构成了这个点上的一个切空间．所有点上所有可能的瞬时速度的集合，就构成了该时空中的一个切丛．

第二章

流形

流形

预备知识 拓扑空间[619]

实流形（real manifold）是一种拓扑空间，其每个点都有一邻域与与欧几里得空间中的开集同胚（homeomorphic）. 如果这些欧几里得空间是 n 维的，那么就叫做 n 维流形. 因此，一个实流形可以看成是我们熟知的欧几里得空间“拼接”而成的.

拓扑流形

定义 1　图和图册

设 N 是一个拓扑空间，满足 Hausdorff 分离性以及第二可数性[1]. 如果存在开集 $U \in \mathcal{T}_N$ 和一个正整数 n，使得 U 同胚于 $\mathbb{R}^n$，同胚映射为 $\varphi : U \to \mathbb{R}^n$，那么称 (U, φ) 是 N 上的一张图（**chart**）. 如果图的一个集合 $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$ 覆盖了 N，即 $\bigcup\{U_\alpha\} = N$，那么称这个集合 $\mathcal{A}$ 是一个图册（**atlas**）.

图 (U, φ) 可以看成是给 U 中各点 x 赋予了一个坐标值 $\varphi(x)$.

定义 2　实拓扑流形

设 N 是一个道路连通的拓扑空间，满足 Hausdorff 分离性以及第二可数性，且有一个图册 $\mathcal{A}$，则称 $(N, \mathcal{A})$ 是一个实拓扑流形（**real topological manifold**）.

从这些定义可看到，流形是“局部地”和欧几里得空间同胚的数学对象. 低维欧几里得空间可以很方便地用我们的几何直觉来理解，大大方便了建立对于流形的直觉.

要注意的是，以上定义只要求了流形的每个点附近都有领域，使之局部地同胚于欧几里得空间，这样的流形被称为拓扑流形，但并不是我们将在微分几何中讨论的重点；我们将来讨论的重点概念是“光滑流形”.

例 1　S^n 流形

[1] 拓扑空间如果有一个可数拓扑基（即一个拓扑基，包含最多 $\aleph_0$ 个基本开集），则称之为第二可数的.

n 维球面 S^n 都是实流形.

- 记 $S^1 = \{(\cos 2\pi t, \sin 2\pi t) \in \mathbb{R}^2 | t \in [0,1]\}$，那么我们可以通过挖去一个点后使用正切函数来构造图：$U = S^1 - \{(1,0)\} = \{(\cos 2\pi t, \sin 2\pi t) \in \mathbb{R}^2 | t \in (0,1)\}$ 可以同胚于 $\mathbb{R}$，同胚映射为 $\varphi(\cos 2\pi t, \sin 2\pi t) = \tan \pi t - \pi/2$；类似地，$V = S^1 - \{(-1,0)\} = \{(\cos 2\pi t, \sin 2\pi t) \in \mathbb{R}^2 | t \in (-1/2, 1/2)\}$ 也可以同胚于 $\mathbb{R}$，同胚映射为 $\phi(\cos 2\pi t, \sin 2\pi t) = \tan \pi t$. 这样一来，$\{(U,\varphi),(V,\phi)\}$ 就是 S^1 的两个图的集合，并且覆盖 S^1，因此这个集合是一个图册，从而得出 S^1 是一个实拓扑流形.

- 类似地，从 S^n 中分别挖去两个不同的点所得到的 U 和 V 都可以同胚于 $\mathbb{R}^n$，从而得到图册.

例 2　莫比乌斯带

将莫比乌斯带表示为矩形纸带两边扭转后粘合，我们可以将其用如图所示的 5 个开集 U_i 覆盖，每个开集都同胚于 $\mathbb{R}^2$，因此我们可以利用它们来构建一个含有 5 个图的图册.

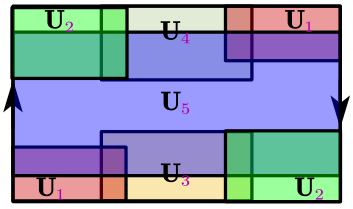


图 1: 莫比乌斯带和覆盖它的五个开集. 注意 $U1$（红色）和 $U2$（绿色）在图示中有“两块”区域，当然它们实际上是相连的.

光滑流形

考虑一个拓扑流形 N 的两个图 (U,φ) 和 (V,ϕ). 如果 $U \cap V \neq \varnothing$，那么在 $U \cap V$ 中的每个点就有两套坐标；当然，也可以把 $\varphi \circ \phi^{-1} : \phi(V) \to \varphi(U)$ 和 $\phi \circ \varphi^{-1} : \varphi(U) \to \phi(V)$ 看成 $\mathbb{R}^n$ 的开子集上的多元向量值函数. 只要这个函数是可以任意进行微分的，就能方便我们研究.

定义 3　光滑函数

设 $f : \mathbb{R}^n \to \mathbb{R}$ 是一个 n 元函数，如果 $\frac{\partial^{k_1+k_2+\cdots+k_n}}{\partial^{k_1}x^1 \partial^{k_2}x^2 \cdots \partial^{k_n}x^n} f$ 在一点 $x \in \mathbb{R}^n$ 处

对于任意非负整数 k_i 成立，即 f 的任意阶偏导数存在，那么称 f 在 x 处光滑；如果 f 处处光滑，也称它是一个光滑函数（**smooth function**），记为 C^∞ 的函数．如果向量值函数的各分量函数都是光滑函数，那么也称这个向量值函数是一个光滑映射（**smooth map**）．

当一个函数的 m 阶偏导数存在并且连续时，我们说它是 C^m 的，因此任意阶偏导数存在时自然被记为 C^∞．今后我们将“光滑”和 C^∞ 视为同义词，不加区分．

需要注意的是，光滑和解析不是等价的概念．解析函数一定光滑，但光滑函数不一定解析，一个典型例子是例 1 [681]．可参考幂级数[206].

定义 4　相容

考虑一个拓扑流形 N 的两个图 (U,φ) 和 (V,ϕ)．如果 $U\cap V\neq\varnothing$，且 $\varphi\circ\phi^{-1}:\phi(V)\to\varphi(U)$ 和 $\phi\circ\varphi^{-1}:\varphi(U)\to\phi(V)$ 都是光滑映射，那么我们称这两个图是相容的（**compatible**）．

从拓扑流形例子可以看出，同一个拓扑空间 N 可以有不同的图册，对应的虽然是同一空间，但按照定义却是不同流形．这意味着流形不仅仅是指空间 N 本身，还指它的局部坐标系结构．这样把拓扑流形分类通常是无意义的，实际上我们需要一个更为统一的研究对象，那就是光滑流形．我们接连使用以下两个定义来定义光滑流形．

定义 5　极大图册

设 $(N,\mathcal{A})$ 是一个拓扑流形．如果 N 的任何一个图，满足“只要它和 $\mathcal{A}$ 中所有图都相容，它就一定在 $\mathcal{A}$ 中”，那么就称 $\mathcal{A}$ 是一个极大图册．

定义 6　光滑流形

拥有极大图册的拓扑流形，被称为一个光滑流形（**smooth manifold**）．光滑流形的图册，被称为该流形上的一个微分结构（**differential structure**）．

将以上定义中的 $\mathbb{R}^n$ 替换为 $\mathbb{C}^n$，图册中的" 光滑映射" 替换为" 复解析映射", 我们就得到复流形的定义：

定义 7　复流形

将定义 1 中图的定义限定为到 $\mathbb{C}^n$ 中开集的同胚，图册的定义限定为 $\mathbb{C}^n$

开集之间的复解析映射，那么定义 2 中所定义的实光滑流形就变成了复流形（**complex manifold**）.

这就是我们将来讨论的主要对象了．今后如无特别说明，本书中“流形”一词都特指“光滑流形”.

通常，为了方便，我们也会将流形 $(N,\mathcal{A})$ 简单记为 N.

子流形

- 本词条处于草稿阶段.

预备知识 流形[688]

子流形

定义 1　子流形

设 N 是一个 n 维流形，K 是它的一个子集．如果在 K 上任意点 $x_0\in K\subseteq N$，都存在一个 N 的图 (U,φ)，使得 $\varphi(U\cap K)$ 是 $\mathbb{R}^n$ 中“令后 $n-k$ 个坐标为 0”所得的平面，其中为方便，记 $\varphi|_K:U\cap K\to\mathbb{R}^k$，且 $\forall x\in U\cap K,\varphi(x)=\varphi|_K(x)$，那么用各 $(U,\varphi|_K)$ 的图册可以构成集合 K 上的拓扑流形，取该图册的极大图册来赋予 K 所得的光滑流形，即为 N 的子流形.

简单来说，定义子流形的需要两步．第一步是利用原流形的 N 图册来构建一个子流形 K 的图册，具体方式见定义 1，其中使用的定义方式是为了方便讨论，并非唯一的方式；第二步是把这个图册扩充为极大图册．这样，配备了此极大图册的 K 就是我们需要的子流形.

等值面/水平集

积流形

预备知识　流形[688]，乘积拓扑[640]

定义 1　积流形

给定两个（微分）流形 $\mathcal{M}$ 和 $\mathcal{N}$，考虑它们的笛卡尔积 $\mathcal{P}=\mathcal{M}\times\mathcal{N}$．$\mathcal{P}$ 赋予乘积拓扑；$\mathcal{M}$ 的任意图 (U,ϕ) 与 $\mathcal{N}$ 的任意图 (V,φ) 的乘积都是 $\mathcal{P}$ 的图，而 $\mathcal{P}$ 的图册由这些图生成．这样，$\mathcal{P}$ 就能成为一个流形，称为 $\mathcal{M}$ 和 $\mathcal{N}$ 的积流形（**Product Manifold**）．

流形上的切空间

- 本词条存在未完成的内容．

预备知识　流形[688]，切空间（欧几里得空间）[683]

对于流形 N，如果能将它嵌入到某个 $\mathbb{R}^k$ 中，嵌入映射为 $i:N\to\mathbb{R}^k$，那么根据**切空间（欧几里得空间）**[683] 中关于曲面 S 的讨论，我们可以使用道路或者导子来算出特定嵌入 i 下流形 N 的切空间和切丛．但是和测试电荷、测试函数类似，特定嵌入也只是一个测试函数，我们讨论流形本身时不依赖特定的嵌入，这就体现出道路和导子定义的好处了．

和大多数教材不同的是，本书中在许多地方使用道路的等价类来定义流形上的切向量，这样比起导子要更加容易可视化．需要注意的是，我们仍然会将导子定义和道路定义看成完全相同的东西，所以会在合适的时候选择使用相应的定义来描述切向量，因此读者应牢牢掌握两种定义为什么是等价的．

流形上切空间的定义

定义 1　切向量

给定流形 N，则在其上一点 $p\in N$ 处的一个切向量就是从 p 出发的一条道路 r 所在的等价类 $[r]$．其中，两道路 r_1 和 r_2 等价当且仅当存在 p 处的一个图 (U,φ)，使得道路 $\varphi\circ r_1$ 和 $\varphi\circ r_2$ 都收敛于 $\varphi(U)$ 中的同一个切向量．

切向量的定义只要求两条道路在某一个图中对应的欧几里得空间里的切向量等价．这种定义方法是合理的，这由以下定理保证：

定理 1

给定流形 N，其上一点 $p \in N$ 处有两个图 (U, φ) 和 (V, ϕ)．p 出发的两条道路 r_1 和 r_2，它们 $\varphi \circ r_1$ 和 $\varphi \circ r_2$ 收敛于同一个切向量，当且仅当 $\phi \circ r_1$ 和 $\phi \circ r_2$ 也收敛于同一个切向量．

证明：为方便计，将 $\phi \circ \varphi^{-1}$ 记为 f．

由于 $f : \mathbb{R}^n \to \mathbb{R}^n$ 是一个双向光滑双射，即 f 和 f^{-1} 都是双射且光滑，于是它的 Jacobi 矩阵 $\partial f / \partial \mathbf{v}$ 是非奇异的；换句话说，如果把向量值函数 f 的第 i 个分量函数记为 $f_i : \mathbb{R}^n \to \mathbb{R}$，那么 f_i 的梯度 $\boldsymbol{\nabla} f_i$ 处处存在且不为零．类似地，f^{-1} 的分量函数的梯度也处处存在且不为零．在以下证明中，为了方便，我们将直接使用 Jacobi 矩阵的表示方法．

$\varphi \circ r_1$ 对应的向量是 $\mathrm{d}(\varphi \circ r_1) / \mathrm{d}t = (\partial \varphi / \partial \mathbf{v}) \cdot (\mathrm{d}r_1 / \mathrm{d}t)$，其中 $\mathbf{v}$ 表示 $\mathbb{R}^n$ 中的向量，$\partial \varphi / \partial \mathbf{v}$ 取在点 $\varphi(p)$ 处的值．相应地，$\varphi \circ r)_2$ 表示的向量是 $(\partial \varphi / \partial \mathbf{v}) \cdot (\mathrm{d}r_2 / \mathrm{d}t)$. 由于这两个向量是同一个，故 $(\partial \varphi / \partial \mathbf{v}) \cdot (\mathrm{d}r_1 / \mathrm{d}t) = (\partial \varphi / \partial \mathbf{v}) \cdot (\mathrm{d}r_2 / \mathrm{d}t)$.

现在直接计算 $\phi \circ r_1$ 和 $\phi \circ r_2$ 对应的向量：$\mathrm{d}(\phi \circ r_1) / \mathrm{d}t = \mathrm{d}(\phi \circ \varphi^{-1} \circ \varphi \circ r_1) / \mathrm{d}t = \mathrm{d}(f \circ \varphi \circ r_1) / \mathrm{d}t = (\partial f / \partial \mathbf{v})(\partial \varphi / \partial \mathbf{v})(\mathrm{d}r_1 / \mathrm{d}t)$．类似地，$\phi \circ r_2$ 对应的向量是 $(\partial f / \partial \mathbf{v})(\partial \varphi / \partial \mathbf{v})(\mathrm{d}r_2 / \mathrm{d}t)$．由于 $(\partial \varphi / \partial \mathbf{v}) \cdot (\mathrm{d}r_1 / \mathrm{d}t) = (\partial \varphi / \partial \mathbf{v}) \cdot (\mathrm{d}r_2 / \mathrm{d}t)$，故 $(\partial f / \partial \mathbf{v})(\partial \varphi / \partial \mathbf{v})(\mathrm{d}r_1 / \mathrm{d}t)(\partial f / \partial \mathbf{v})(\partial \varphi / \partial \mathbf{v})(\mathrm{d}r_2 / \mathrm{d}t)$，这就证明了必要性；而充分性则是由于 $\partial f / \partial \mathbf{v}$ 非奇异，故非零的 $(\partial \varphi / \partial \mathbf{v})(\mathrm{d}r_1 / \mathrm{d}t)$ 不会映射到零向量上，进而两个向量之差如果不为零则其在 f 下的映射也不为零．

证毕.

定理 1 意味着流形 N 上的道路，在一个图里的等价划分，和在任何其它图里的等价划分是一致的，这就使得我们可以摆脱对特定图的依赖，直接定义两条道路等价当且仅当它们在任意一个图里等价．类似地，我们也可以摆脱对特定图的依赖来定义道路等价类的加法．

定义 2　切向量的加法

给定流形 N，其上一点 $p \in N$．p 出发的两个道路等价类 $[r_1]$ 和 $[r_2]$ 的和定义为，任取 p 附近一图 (U, φ)，令 $[r_1] + [r_2] = [\varphi^{-1}(\varphi(r_1) + \varphi(r_2))]$.

定理 2

给定流形 N，其上一点 $p \in N$ 处有两个图 (U, φ) 和 (V, ϕ)．对 p 出发的两

条道路 r_1 和 r_2，按以上定义所得的和（注意，这个和是 N 上的一个道路等价类），在两个图中的计算结果一致.

证明：

为方便计，记 $f = \phi \circ \varphi^{-1}$.

根据定理 1 的证明，我们知道 f 的各分量 f_i 都处处有非零梯度. 设 r_1 和 r_2 的起点都是 $p \in N$，那么用 (U, φ) 计算可得：$r_1 + r_2 = \varphi^{-1}(\varphi \circ r_1 + \varphi \circ r_2)$；而用 (V, ϕ) 计算可得：$r_1 + r_2 = \phi^{-1}(\phi \circ r_1 + \phi \circ r_2) = \phi^{-1}(f \circ \varphi \circ r_1 + f \circ \varphi \circ r_2)$. 我们希望这样计算出来的两个 $r_1 + r_2$ 是等价的（一般不相等）.

将以上计算出来的两个 $r_1 + r_2$ 都放到 (U, φ) 中进行比较，它们等价当且仅当对 t 求导的结果一样. 用 (U, φ) 本身算出来的 $r_1 + r_2$ 当然是 $\varphi r_1 + \varphi r_2$；而在 (V, ϕ) 中算出来的 $r_1 + r_2$ 则是 $\varphi\phi^{-1}(f\varphi r_1 + f\varphi r_2) = f^{-1}(f\varphi r_1 + f\varphi r_2)$. 直接计算得，$\mathrm{d}/\mathrm{d}t(\varphi r_1 + \varphi r_2) = \mathrm{d}/\mathrm{d}t(\varphi r_1) + \mathrm{d}/\mathrm{d}t(\varphi r_2)$，而 $\mathrm{d}/\mathrm{d}t(f^{-1}(f\varphi r_1 + f\varphi r_2) = \partial f^{-1}/\partial \mathbf{v}(\partial f/\partial \mathbf{v}(\mathrm{d}/\mathrm{d}t(\varphi r_1) + \mathrm{d}/\mathrm{d}t(\varphi r_2)))$. 考虑到 $\partial f/\partial \mathbf{v}$ 和 $\partial f^{-1}/\partial \mathbf{v}$ 都是非奇异的，故它们的乘积是单位矩阵，于是以上两个求导的结果是一致的，因此两种方式计算出来的 $r_1 + r_2$ 是等价的.

证毕.

这个定理保证了定义 2 的合理性，即无论用哪个图去计算切向量（道路等价类）的和，结果仍然是同一个等价类. 定理成立的关键在于，两个图相容使得 $f = \phi \circ \varphi^{-1}$ 是一个双向光滑的双射，因此 $\partial f/\partial \mathbf{v}$ 和 $\partial f^{-1}/\partial \mathbf{v}$ 都存在且非奇异，这就使得计算过程中可以把 f 变换的影响消除.

微分

未完成：内容

抽象指标

预备知识 爱因斯坦求和约定[498]

- 本词条缺少预备知识，初学者可能会遇到困难.

引入的动机

在欧几里得空间中，我们常常可以把一个向量看成若干个标量的组合，类似地也可以把一个向量场看成若干个标量场的组合. 这是因为在欧几里得空间中我们已经默认了一组标准正交基，使得向量都可以表示成坐标形式，每个坐标都可以单独看成一个标量.

欧几里得空间中可以唯一地把向量看成其分量的组合，这使得问题的讨论具体而直观. 比如说，当我们讨论向量和对偶向量的乘积时，就可以说是各对应分量的乘积之和；按照爱因斯坦求和约定，如果有向量 $\mathbf{v}=\left(v^x,v^y,v^z\right)$ 和对偶向量 $\mathbf{u}=\left(u_x,u_y,u_z\right)$，那么这两个向量的内积就是 v^iu_i.

在微分几何中，我们研究的主要对象是流形. 流形是局部同胚于欧几里得空间的对象，这样就可以按照这个同胚在流形上的每个点附近定义一个局部坐标系. 但是每个点上，可以选择不同的欧几里得空间来进行同胚，也就有了不同的坐标系. 当我们讨论流形上一点的切空间时，选择不同的局部坐标系会导致切空间的基的选择有所不同. 这样一来，流形上一个切向量的坐标分量就是不确定的，必须给定了局部坐标系才可以讨论其分量.

在流形上讨论向量、对偶向量乃至一般的张量时，只有给定了具体的局部坐标系才可以像欧几里得空间里那样拆成坐标形式来直观讨论. 总是要选定局部坐标系会让讨论变得繁琐，因此，为了在保留坐标表示的优势的同时避免选定坐标系的烦恼，我们引入了抽象指标的表示方法.

定义及例子

对于流形上一点的切向量 $\mathbf{v}$，我们用抽象指标记号将其记为 $\mathbf{v}^a$. a 是这个切向量的一个抽象指标，取什么符号都不重要；单一的抽象指标意味着当我们选定任意一个局部坐标系以确定切空间的基后，这个 $\mathbf{v}$ 在这组基下的坐标表示是由一个指标来确定的，也就是说，它的分量是 v^1，v^2，v^3 等等. 类似地，一个对偶向量 $\mathbf{u}_b$ 意味着确定切空间的基以及其对偶基后，这个 $\mathbf{u}_b$ 在这组基下的坐标分量是 u_1，u_2，u_3 等.

由于无论在什么局部坐标系下，$\mathbf{v}$ 和 $\mathbf{u}$ 的乘积永远是 v^iu_i 的形式，因此我们把这一规律用抽象指标表示为 $\mathbf{vu}=\mathbf{v}^a\mathbf{u}_a$，其中抽象指标 a 按照爱因斯坦

求和约定进行了求和．虽然没有给定具体坐标系的时候没法进行“求和”的运算，但是抽象指标求和式表达的意义是“如果我们选定了**任意**局部坐标系，那么可以用 v^i 和 u_i 分别代替 $\mathbf{v}^a$ 和 $\mathbf{u}_a$ 来表示具体求和”．也就是说，抽象指标 a 是具体指标 i 的推广，是忽略了具体坐标系后的抽象．

类似地，流形上一点处的一个二阶协变张量可以表示成 $\mathbf{T}_{ab}$，它将向量 $\mathbf{v}^a$ 和 $\mathbf{u}^b$ 映射为一个数字 $\mathbf{v}^a\mathbf{u}^b\mathbf{T}_{ab}$．一个三阶的混合张量可以是 $\mathbf{T}^a_{bc}$，它将向量 $\mathbf{v}^b$，$\mathbf{u}^c$ 和对偶向量 $\mathbf{w}_a$ 映射为一个数字 $\mathbf{w}_a\mathbf{v}^b\mathbf{u}^c\mathbf{T}^a_{bc}$．

纤维丛

- 本词条处于草稿阶段．

预备知识 乘积拓扑[640]，矢量空间[444]

定义

直观来说，纤维丛是指在一个拓扑空间 B 的每一个点都长出来另一个拓扑空间 F 所得到的一个空间．每一个点 $x \in B$ 上的 F 被称为一根**纤维（fibre）**，这些纤维所在的 B 称为**底空间（base space）**，而整个结构 (B, F) 就是一个**纤维丛（fibre bundle）**．

准确的定义如下所述，其中 E 就是“B 上每个点都长出一个 F 的丛空间”：

定义 1　纤维丛

给定拓扑空间 B 和 F，如果存在一个拓扑空间 E 和一个连续满射 $\pi : E \to B$，使得对于任意的 $x \in B$，都有 $\pi^{-1}(x) \cong F$，那么称这个结构 (E, F, B, π) 为一个纤维丛（**fibre bundle**），称 E 是这个纤维丛的全空间（**total space**），F 是其纤维（**fibre**），B 是其底空间（**base space**），有时也译作基空间．

如果把 B 想象成一块土地，F 想象成一棵草，那么 E 就是“土地上长了一片草”这一概念，E 的每个元素就是某棵草上的一个点．定义中的连续满射 f 的作用是把这样的一个点映射到相应的草所在的地点．

要注意的是，E 不完全等同于 $B \times F$．对于 $B \times F$ 来说，任意给定两个 $x_1, x_2 \in B$，我们自然可以找到 $x_1 \times F$ 和 $x_2 \times F$ 上的一一对应关系，这是由集

合笛卡尔积的定义决定的．但是纤维丛 E 上，如果上述 $x_1 \neq x_2$，那么两个地方长出来的纤维是没有天然的双射对应的的[2]．这就是“纤维丛”这一名称的深意，而乘积空间应该被想象纤维被粘在一起的情况，只是纤维丛的一个定义了额外联系的特例．

两个纤维丛之间可以有映射偶：

定义 2　纤维丛的态射

设 (E_1, F_1, B_1, f_1) 和 (E_2, F_2, B_2, f_2) 是两个纤维丛，

纤维丛的例子

向量丛

向量丛是纤维丛的特例，即纤维都是向量空间的情况．

定义 3　向量丛

给定拓扑空间 B 和线性空间 V，如果存在一个拓扑空间 E 和一个连续满射 $\pi : E \to B$，使得对于任意的 $x \in B$，都有 $\pi^{-1}(x) \cong V$，那么称这个结构 (E, V, B, π) 为一个向量丛（**vector bundle**）．

向量丛之间也有丛映射：

定义 4　丛映射

给定向量丛 (E, V_E, M, π_E) 和 $(F, V_F, N, \pi_F])$，其中 M 和 N 是实流形．我们定义一个“光滑丛映射（C^∞ **bundle map**）”为 $E \to F$ 的映射偶 $\varphi : E \to F$ 和 $\overline{\varphi} : M \to N$，使得：

$$\overline{\varphi} \circ \pi_E = \varphi \circ \pi_F \tag{1}$$

且在任意 $p \in M$ 处，$\varphi|_p$[3]是从 $p \times V_E$ 到 $\overline{\varphi}(p) \times V_F$ 的映射，并且是一个线性映射．

虽然，一个向量丛 (E, V, B, ϕ) 不能简单等同于 $B \times V$，不过 $B \times V$ 本身也是一个纤维丛，称之为**平凡（trivial）**的纤维丛．

---

[2]在微分几何中，我们研究的切丛是纤维丛的一种，而所谓的“联络”实际上就是指定了不同纤维间的双射．

[3]即只考虑 p 处纤维的映射 φ．

向量丛和切丛

- 本词条处于草稿阶段.

未完成：创建“环上的模”词条后，应将其引用为本词条的预备知识

预备知识 纤维丛[696]

“切丛”是“切向量丛”的简称. 顾名思义，它是切向量构成的向量丛，而这里的切向量是在流形上定义的，因此切丛的底空间是流形.

定义 1　切丛

*给定流形 M，以 M 为底空间，把各 $p \in M$ 上的 T_pM 视为该点处的一根纤维，得到的纤维丛就称为流形 M 上的切丛（**tangent bundle**）.*

一个切向量场可以视为切丛的一种特殊的子集，称为“**截面（section）**”. 使用这个术语是为了强调这种子集的特殊性，它在每一根纤维上都取且仅取一个点，看起来就像是纤维丛的一个截面. 同样地，一个光滑切向量场有时也被称作切丛上的一个**光滑截面**.

局部来看，流形 M 上切丛的每根纤维是一个线性空间；整体来看，每个切向量场本身都可以看成一个向量，构成一个线性空间，记为 $\mathfrak{X}(M)$. $\mathfrak{X}(M)$ 作为线性空间的**基域**和 M 是一致的，具体来说，实流形 M 上的 $\mathfrak{X}(M)$，其数乘时使用的“数字”就是实数. 但是我们常研究另一种情况，即把对 $\mathfrak{X}(M)$ 进行数乘时的“数字”取为 M 上的一个光滑函数. 这个时候，$\mathfrak{X}(M)$ 不再能被看成一个线性空间，而应该是一个环 $C^\infty(M)$ 上的模. 这里 $C^\infty(M)$ 是 M 上全体光滑函数的集合，它只是一个环，不满足乘法逆元存在性.

从流形整体来看，一个 1-形式，或者说余切向量场，本身也可以看成是模 $\mathfrak{X}(M)$ 上的一个线性映射.

流形 M 上也可以定义切丛之外的向量丛，同样有截面的概念，只不过此时的截面不再是切向量场了. 向量丛 E 上全体光滑截面的集合，记为 $\Gamma(E)$. 由上所述，光滑向量场的集合 $\mathfrak{X}(M)$ 是 $\Gamma(E)$ 的特例，正如切丛是向量丛的特例.

向量丛之间有两类比较重要的映射：

定义 2　点算子

设 E 和 F 是同一个底空间 M 上的两个向量丛，称丛映射 $\varphi : E \to F$ 是一个点算子（**point operator**），如果对于任意的光滑截面 $s \in \Gamma(E)$ 和点 $p \in M$，当 $s(p) = 0$ 时必有 $\varphi(s)(p) = 0$.

换句话说，如果称 $Z_s\{p \in M | s(p) = 0\}$ 为 s 的零化子（**null set**），那么 φ 是一个点算子当且仅当 $Z_s \subseteq Z_{\varphi s}$.

定义 3　局部算子

设 E 和 F 是同一个底空间 M 上的两个向量丛，称丛映射 $\varphi : E \to F$ 是一个局部算子（**local operator**），如果对于任意的光滑截面 $s \in \Gamma(E)$，开集 $U \subseteq E$ 和点 $p \in M$，当 $s(U) = \{0\}$ 时必有 $\varphi(s)(U) = \{0\}$.

换句话说，如果记 Z_s 的内部[4]为 Z_s°，那么 φ 是一个局部算子当且仅当 $Z_s^\circ \subseteq Z_{\varphi s}^\circ$.

局部算子等价于模 $\mathfrak{X}(M)$ 上的一个线性映射，我把这一点表述为如下定理：

定理 1

给定流形 M 上的两个向量丛 E 和 F. 如果 $\varphi : \Gamma(E) \to \Gamma(F)$ 是 $C^\infty(M)$-线性的[5]，那么 φ 必为一个局部算子.

习题 1

证明定理 1 . 思路之一可以是利用 bump function 来构造一个使得 $fs = s$ 的函数 f，然后讨论 s, fs 和 φs 的零化子的内部之间的关系.

习题 2　定理 1 的反思

在把 $\mathfrak{X}(M)$ 整体视为一个环 $C^\infty(M)$ 上的模的时候，局部算子的性质有什么几何意义吗？

最后，我们再介绍一个概念，**参考系**. 时空都是流形，而在时空理论中我们常讨论不同参考系下的物理定律，因此有必要定义一下流形上的参考系是什么.

定义 4　参考系

[4] 见**点集的内部、外部和边界**[622] 词条.

[5] 即 $\forall s \in \Gamma(E)$ 和 $\forall f \in C^\infty(M)$，都有 $\varphi(fs) = f\varphi(s)$.

给定流形 M 上的一个 r 维向量丛 E. 设 E 上一个光滑截面的集合 $Fr = \{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_r\}$ 满足：对于任意 $p \in M$，$\{\mathbf{e}_1(p), \mathbf{e}_2(p), \cdots, \mathbf{e}_r(p)\}$ 构成纤维 E_p 的一组基，那么称 Fr 是 E 上的一个参考系（**frame**）.

微分形式

预备知识 对偶空间[460]，流形上的切空间[692]

1-形式

在流形 M 上给定一个光滑函数 f 和点 $p \in M$ 处的一个切向量 $\mathbf{v}$，则由切向量的道路定义容易计算出 $\mathbf{v}f(p)$，这个值不依赖于图的选择. 不过为了方便理解，既然这是一点处的性质，我们不妨取 p 所在的一个图来考虑，这样就把问题范围从任意流形 M 化为任意欧几里得空间了.

欧几里得空间一点处的微分形式

在欧几里得空间 $\mathbb{R}^n$ 中，$\mathbf{v}f(p) = \mathbf{v} \cdot \nabla f$. 我们可以看成是给定 f，任意的向量 $\mathbf{v}$ 都会被映射为 $\mathbf{v} \cdot \nabla f$，这是一个线性映射. 不同的 f 有可能对应相同的 ∇f，只要它们全微分的形式是相同的就可以——由此诞生了“微分形式”的概念.

我们用 $\mathbb{R}$ 上的光滑函数举一个例子. 考虑函数 $f(x) = x + 1$ 和 $g(x) = \mathrm{e}^x$. 在 $x = 0$ 处，它们的函数值是相同的，斜率也相同，如图图 1 所示. 这样一来，在这个点附近，忽略高阶无穷小的情况下，这两个函数是无法进行区分的. 如果两个函数表示两个斜面，那么当你站在 $x = 0$ 这一点上的时候，无论是你手表上的高度计读数还是倾斜的感觉，都无法告诉你自己到底站在函数 f 上还是 g 上.

我们把**一个点处**函数值和斜率完全相同的全体函数，归为一个等价类，称为这一点处的一个**微分形式（differential form）**. 取这个名字是因为，一个微分形式等价类里包含所有函数，其“微分的形式”都是一样的. 比如说，图 1 中的两个函数，在 $x = 0$ 处的微分都是 $\mathrm{d}f = \mathrm{d}g = \mathrm{d}x$. 如果在 $x = 1$ 处观察，两函数不仅函数值不同，而且“微分的形式”也不一样：$\mathrm{d}f = \mathrm{d}x \neq \mathrm{e}\mathrm{d}x = \mathrm{d}g$.

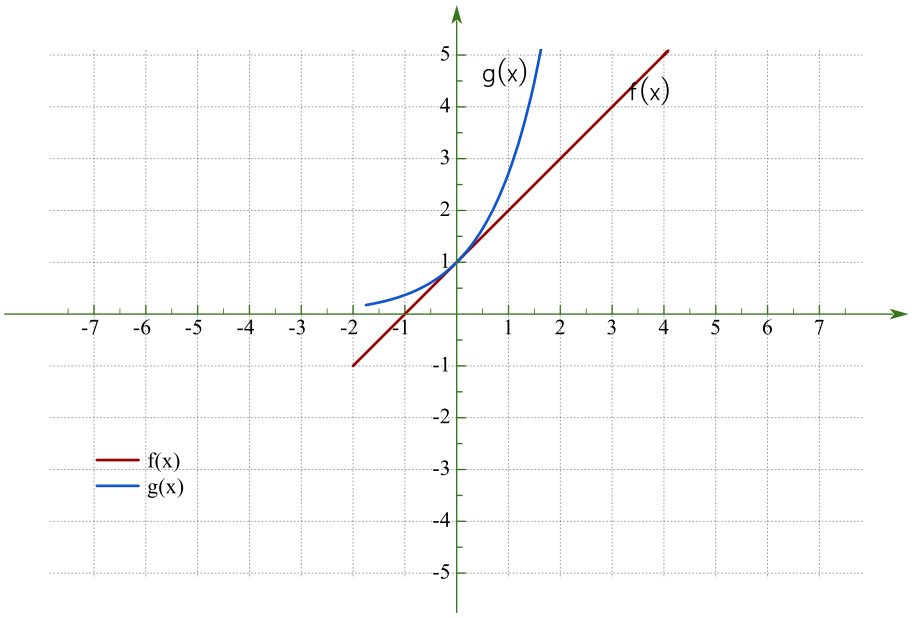


图 1: 函数 $f = x + 1$ 和 $g = \mathrm{e}^x$ 的例子.

上述微分形式的定义，可以拓展到任何 $\mathbb{R}^n$ 上. 一般地，如果 $\mathbb{R}^n$ 上的两个光滑函数 f, g 在一点 $p \in \mathbb{R}^n$ 处的函数值相同，并且全微分也相同，那么我们说这两个函数是 p 处同一个微分形式等价类里的元素. 直观来说，一个微分形式就是 $\mathbb{R}^n \times \mathbb{R}$ 经过某点 $(p, h(p))$ 的一个平面，所有图像与该平面相切于 $(p, h(p))$ 的函数都属于这个微分形式. 在图 1 的例子中，$p = 0, h(p) = 1$.

最后，在流形上，由于我们只关心微分形式的“微分的形式”，而不关心具体的函数值，我们也会忽略函数值的差别，而把一点处全微分相同的函数们都归入同一个微分形式中. 这是因为在流形上我们把微分形式看成余切向量场，用来和切向量相互作用，函数值是不影响这种作用的.

流形上的微分形式

在流形 M 上给定一个光滑函数 $f : M \to \mathbb{R}$，考虑点 $p \in M$ 处的一个切向量 $\mathbf{v} : I \to M$，我们可以计算出 f 在 p 处沿着 v 方向的方向导数：

$$\mathbf{v}f = \lim_{t \to 0^+} \frac{f(\mathbf{v}(t)) - f(\mathbf{v}(0))}{t} \tag{1}$$

由于是求方向导数，因此如果函数 g 和 f 属于 p 处的同一个等价类，那么就有 $\mathbf{v}f = \mathbf{v}g$. 用向量分析的语言来说，$\mathbf{v}f$ 就是 $\mathbf{v} \cdot \nabla f$，其中 ∇f 是 f 的梯

度向量．p 点处任意微分形式里的全体函数，其在该点的梯度向量都是一样的，因此我们也可以定义这就是微分形式在该点的梯度向量．

我们知道，向量内积实际上是向量和对偶向量的相互作用．这样一来，我们也可以把一点处的一个微分形式看成该点处切向量的一个对偶向量，它的作用是把该点处的切向量映射成一个数字，并且这个映射是线性的．也就是说，一点处的一个微分形式，实际上就是该点处切空间上的一个对偶向量或者叫余切向量．

在欧几里得空间中我们可以很方便地写出一点处任意微分形式的表达：$a_i \mathrm{d}x^i$．注意这里使用了爱因斯坦求和约定，其中 $a_i \in \mathbb{R}$，而 x^i 是欧几里得空间的坐标．你也可以这么理解，$\mathrm{d}x^i$ 就是函数 $x^i + c$ 的微分形式，其中 c 是任意常数．这样，$a_i \mathrm{d}x^i$ 就被视为函数 $f = a_i x^i$ 所在的微分形式了．我们把这个微分形式表示为 $\mathrm{d}f$．

定义 1　流形上的微分形式

*给定实流形 M．称 M 上的一个光滑余切向量场为一个微分形式（**differential form**），或者更具体地，一个 **1-**形式（**1-form**）．*

在 $p \in M$ 处的一个余切向量，可以表示为 $\mathrm{d}f$，其中 f 是定义在 p 的一个邻域中的光滑函数，且对于任意切向量 $\mathbf{v}$，都有 $\mathbf{v}$ 和 $\mathrm{d}f$ 的乘积恰为切向量对函数的作用：$\mathbf{v}f$．

k-形式

预备知识　外代数[603]

我们已经把微分形式定义为光滑余切向量场了．这样一来，流形 M 上一点 p 处的全体微分形式，构成的余切向量集合，就是一个线性空间 T_p^*M．我们自然可以在这个线性空间上定义一个外代数 $\bigwedge T_p^*M$．

定义 2　k-形式

*给定实流形 M 上一点 p 处的余切空间 T_p^*M，其外积空间 $\bigwedge^k T_p^*M$ 中的任一元素为 p 处的一个 k-形式（k-**form**）．*

M 上的一个光滑[6]k-形式场[7]，称为 M 上的一个 k-形式．M 上 k-形式的集

[6]此处“光滑”依然指“在任意坐标图中场系数是光滑函数”．

[7]这是一个 k 阶张量场．

合，记为 $\Omega^k(M)$.

流形上的光滑张量场的集合，本身也可以看成一个线性空间．在本节的例子里，流形上 k-形式的集合，本身也是一个线性空间，每个 k-形式都是该空间里的一个向量．从这个意义上来说，$\Omega^k(M)$ 中的每一个 k-形式场也可以看成是 M 上把余切向量场看成一个向量时的 k 次外积．

在度量张量的帮助下，余切向量也可以和切向量进行一一对应，此时我们就可以把余切向量也表示成一个切向量，其和其它切向量的相互作用表现为内积．这样一来，某点上的一个 1-形式可以理解为一个箭头；2-形式可以理解为两个箭头张成的面积，注意这个面积是有方向的；类似地，一个 k-形式就可以理解为 k 维的有向体积．

费罗贝尼乌斯定理

预备知识 切空间[692]

在微分几何中，**费罗贝尼乌斯定理（Frobenius' theorem）**给出了一阶拟线性偏微分方程组可积的充分必要条件．它有许多重要的几何推论．

对合分布与可积性

设 M 是 n 维微分流形．一个**光滑 k 维分布**（smooth k-dimensional distribution，注意这不是分析学意义下代表广义函数的分布）$\mathfrak{D}$ 是指切丛 TM 的 k 维光滑子丛．等价地说，这表示在任何一点 $p\in M$ 都给出 T_pM 的 k 维子空间 $\mathfrak{D}_p$，而且 $\mathfrak{D}_p$ 光滑地依赖于 p.

维数为 k 的光滑分布 $\mathfrak{D}$ 称为**对合的（involutive）**，如果对于 $\mathfrak{D}$ 的任意两个光滑截面 X,Y，$[X,Y]$ 也还是 $\mathfrak{D}$ 的截面．

定理 1　费罗贝尼乌斯定理

设 $\mathfrak{D}$ 是 M 上的 k 维对合分布．则在任意一点 $p\in M$，都有局部坐标系 $\{x^i\}$ 使得 $\mathfrak{D}$ 在该点处由坐标向量 $\partial_1,\cdots,\partial_k$ 张成．等价地说，在任意一点 p 都有一过点 p 的 k 维子流形 N_p 使得 $\mathfrak{D}$ 恰为 N_p 的切丛．

这样的分布称为**可积的（integrable）**．由此在小邻域内得到的子流形称为

分布的**积分子流形（integral submanifold）**，积分子流形的族称为 M 局部的一个**正则叶理（regular foliation）**．它将流形 M 在局部上划分成了许多相互不交叠的叶片．

定理的证明大意如下．不妨设 M 是 $\mathbb{R}^n$ 中坐标原点的小邻域．取这小邻域里 $\mathfrak{D}$ 的局部标架 $X_1,\cdots,X_k$，使得 $X_i(0)=\partial_i|_0$ (这总可以通过适当的旋转和缩放做到)，并定义映射 $\Pi: M\to\mathbb{R}^k$ 为

$$\Pi(x^1,\cdots,x^n)=(x^1,\cdots,x^k).$$

则对于 $p\in M$ 和 p 处的切向量 $v=\sum_{i=1}^n v^i\partial_i|_p$, 有

$$d\Pi(p)(v)=v^1\partial_1|_{\Pi(p)}+\ldots+v^k\partial_k|_{\Pi(p)}.$$

于是切丛的子丛 $\ker d\Pi$ 的秩是 $n-k$，且在原点的小邻域内与 $\mathfrak{D}$ 互补．故 $(d\Pi|_{\mathfrak{D}})$ 实际上是从子丛 $\mathfrak{D}$ 到切丛 $T\Pi(M)$ 的同构，因此 $(d\Pi|_{\mathfrak{D}})^{-1}$ 在丛 $T\Pi(M)=\operatorname{span}(\partial_1,\cdots,\partial_k)$ 上是良好定义的，而且其像正是 $\mathfrak{D}$．命 $V_i=(d\Pi|_{\mathfrak{D}})^{-1}\partial_i$，则诸 V_i 都是 $\mathfrak{D}$ 的截面，根据对合性质，$[V_i,V_j]$ 也是 $\mathfrak{D}$ 的截面，而且

$$d\Pi[V_i,V_j]=[\partial_i,\partial_j]=0.$$

于是 $[V_i,V_j]=0$，因此它们实际上是坐标向量场．因此 $\mathfrak{D}$ 是由坐标向量场张成的．

等价表述

设 $\mathfrak{D}$ 是 k 维光滑分布. 如果微分 p-形式 ω 作用在 $\mathfrak{D}^{\otimes p}$ 的任何截面上都得到零，则称 ω 消没 $\mathfrak{D}$．显然两个消没 $\mathfrak{D}$ 的微分形式的外积仍然消没 $\mathfrak{D}$．所有消没 $\mathfrak{D}$ 的微分形式组成 M 的外微分代数的子代数，称为 $\mathfrak{D}$ 的消没子（annihilator）．根据外微分的计算法则，可见 $\mathfrak{D}$ 为对合分布当且仅当对于任意消没 $\mathfrak{D}$ 的 ω，$d\omega$ 仍旧消没 $\mathfrak{D}$．

在流形 M 上给定 $n-k$ 个线性无关的微分形式 $\omega_{k+1},\ldots,\omega_n$ 后，方程组

$$\omega_l=0,\ l=+1,\ldots,n-k$$

称为一个**普法夫系（Pfaffian system）**，它确定了一个 k 维分布，即所有被 $\omega_{k+1},\ldots,\omega_n$ 消没的子空间给出的分布．因此，费罗贝尼乌斯定理有下列等价形式：

定理 2

设 $\omega_k, ..., \omega_n$ 是 M 上线性无关的微分形式，则普法夫系

$$\omega_l = 0,\ l = k+1, ..., n-k$$

给出可积的 k 维分布当且仅当每个 $d\omega_l$ 皆由 $\omega_{k+1}, ..., \omega_n$ 生成，或等价地，

$$d\omega_l \wedge \omega_{k+1} \wedge ... \wedge \omega_n = 0,\ \forall l.$$

这称为费罗贝尼乌斯条件.

应用举例

费罗贝尼乌斯定理在微分几何中有广泛的应用．大致来说，它是" 偏导数可交换" 这一事实的推广：偏微分方程组受制于偏导数可交换这一条件，因此可积的偏微分方程组必须满足由此导出的必要条件，而费罗贝尼乌斯定理断言这些条件也是可积的充分条件.

考虑 n 维自变量 $x = (x^1, ..., x^n)$ 的 m 个未知函数 $u = (u^1, ..., u^m)$ 的拟线性偏微分方程组：

$$\frac{\partial u^\alpha}{\partial x^i} = f_i^\alpha(x, u),\ 1 \le \alpha \le m; 1 \le i \le n. \tag{1}$$

这里 $f_i^\alpha(x, z)$ 是已知的函数．为了确定这个方程组的解，需要的是一个 “初始条件”，即指定 u 在某 x_0 处的值．方程的可解性问题如下：

设有开集 $U \subset \mathbb{R}^n$，$V \subset \mathbb{R}^m$，函数组 $f_i^\alpha(x, z)$ 定义在 $U \times V$ 上．则对于 $x_0 \in U$ 和 $z_0 \in V$，是否存在定义在 U 上，取值于 V 中，且满足式 1 和初始条件 $u(x_0) = z_0$ 的函数组 $u = (u^1, ..., u^m)$?

如果 $u = u(x), u(x_0) = z_0$ 是这方程组的解，那么根据偏导数的交换性，显然可得

$$\frac{\partial f_i^\alpha}{\partial x^j} + \sum_{\beta=1}^m f_j^\beta \frac{\partial f_i^\alpha}{\partial z^\beta} = \frac{\partial f_j^\alpha}{\partial x^i} + \sum_{\beta=1}^m f_i^\beta \frac{\partial f_j^\alpha}{\partial z^\beta}, \forall i, j, \alpha \tag{2}$$

在点 (x_0, z_0) 处成立．因此，欲使得式 1 对于任何初值都可积，则式 2 必须在 $U \times V$ 中处处成立．显然式 2 处处成立当且仅当向量场

$$X_i = \frac{\partial}{\partial x^i} + \sum_{\alpha=1}^m f_i^\alpha(x, z) \frac{\partial}{\partial z^\alpha},\ i = 1, ..., n$$

两两对易．然而另一方面，$u = u(x)$ 是式 1 的解当且仅当子流形 $\{(x, u(x)) : x \in U\}$ 是 n 维分布 $\mathrm{span}(X_1, ..., X_n)$ 的积分子流形．因此，根据费罗贝尼乌斯定理，诸 X_i 两两对易实为式 1 对于任何初值均可积的充要条件．故式 2 在 $U \times V$ 上处处成立是式 1 对于任何初值均可积的充要条件．

如果用普法夫系来表示则更简单．命

$$\omega^\alpha := dz^\alpha - \sum_{i=1}^{n} f_i^\alpha dx^i, 1 \leq \alpha \leq m. \tag{3}$$

则式 1 等价于 $U \times V$ 上的普法夫系 $\omega^\alpha = 0, 1 \leq \alpha \leq m$，或者直接计算可见微分 1-形式 $\{\omega^\alpha\}_{\alpha=1}^m$ 消没的那些切向量恰是 m 维分布 $\mathrm{span}\{X_i\}_{i=1}^n$．式 2 是这个普法夫系给出 $U \times V$ 上可积分布的充要条件，而通过简单的外微分运算可看出它进一步等价于

$$d\omega^\alpha = 0, 1 \leq \alpha \leq m$$

在整个 $U \times V$ 上恒成立．用这个办法来判定可积性非常方便．

第三章

联络

联络 (向量丛)

预备知识 向量丛[??]

本节采用爱因斯坦求和约定.

设 M 是 n 维微分流形，E 是其上秩为 k 的光滑向量丛.

定义与例子

向量丛 E 上的一个**联络（connection）**是指一个映射 $D:\Gamma(E)\otimes TM\to\Gamma(E)$，或者等价地 $D:\Gamma(E)\to T^*M\otimes\Gamma(E)$，满足如下条件：

1. 对于截面 $\xi\in\Gamma(E)$ 和任何切向量场 X，$X\to D_X\xi$ 是 $C^\infty(M)$-线性的，即
$$D_{fX_1+fX_2}\xi=fD_{X_1}\xi+gD_{X_2}\xi,\ f,g\in C^\infty(M).$$
2. 对于截面 $\xi\in\Gamma(E)$ 和任何切向量场 X，$\xi\to D_X\xi$ 是实线性的，即对于实数 c_1,c_2 有
$$D_X(c_1\xi_1+c_2\xi_2)=c_1D_X\xi_1+c_2D_X\xi_2.$$
3. 对于任何 $f\in C^\infty(M)$，都有莱布尼兹律：
$$D_X(f\xi)=X(f)\xi+fD_X\xi.$$

直观上说，在向量丛上给定联络，就是给定一个" 符合张量规律的导数运算".

如果取 $M=\mathbb{R}^n$，E 为平凡向量丛 $\mathbb{R}^n\times\mathbb{R}^k$，则截面 ξ 就是通常的 k-维向量值函数，通常的对笛卡尔坐标的微分运算
$$D\xi=(\partial_i\xi^\alpha)_{1\le i\le n}^{1\le\alpha\le k}$$
就是一个联络.

不平凡的例子在黎曼几何中多有出现.

联络形式

设 D 是向量丛 E 上的联络. 设 $\{e_i\}_{i=1}^n, \{\theta^i\}_{i=1}^n, \{s_\alpha\}_{\alpha=1}^k$ 分别是 $TM, T^*M, \Gamma(E)$ 的局部光滑标架，其中 $\{e_i\}$ 和 $\{\theta^i\}$ 是对偶标架. 在这些局部光滑标架之下，按照联络的定义，有

$$D_X\xi = X^i \cdot \left[e_i(\xi^\alpha)s_\alpha + \xi^\alpha D_{e_i}s_\alpha\right].$$

若命 $D_{e_i}s_\alpha = \Gamma_{\alpha i}^\beta s_\beta$，则有

$$D_X\xi = X^i \cdot \left[e_i(\xi^\alpha) + \Gamma_{\beta i}^\alpha \xi^\beta\right] s_\alpha.$$

称系数 $\Gamma_{\beta i}^\alpha$ 为联络系数（coefficients of connection）或克氏符（Christoffel symbol）. 于是就得到 1-形式的 $k \times k$ 矩阵

$$\omega = (\omega_\beta^\alpha) = (\Gamma_{\beta i}^\alpha \theta^i).$$

它们称为联络 D 在局部的**联络 1-形式矩阵（matrix of connection 1-forms）**. 从而可写

$$D\xi = (d\xi^\beta + \omega_\alpha^\beta \xi^\alpha) \otimes s_\beta.$$

注意，克氏符和联络 1-形式都只能局部定义. 若 $\{s'_\beta\}$ 是 E 的另一个局部标架，同 $\{s_\alpha\}$ 之间的转换公式为 $s_\beta = b_\beta^\alpha s_\alpha$，则

$$b_\alpha^\gamma \omega'^\alpha_\beta \otimes s_\gamma = \omega'^\alpha_\beta s'_\alpha = Ds'_\beta = (db_\beta^\gamma + b_\beta^\alpha \omega_\alpha^\gamma) \otimes s_\gamma.$$

从而新标架下的联络 1-形式矩阵 ω' 同原标架下的 ω 之间的转换关系是

$$\omega' = db \cdot b^{-1} + b \cdot \omega \cdot b^{-1}. \tag{1}$$

这是联络 1-形式在向量丛的局部标架变换下的转换公式，由此可见联络 1-形式矩阵的定义依赖标架的选取，从而并非全局定义的张量场.

在 $\mathbb{R}^n \times \mathbb{R}^k$ 上，笛卡尔坐标系下的微分

$$D\xi = (\partial_i \xi^\alpha)_{1\le i\le n}^{1\le \alpha\le k}$$

在经过坐标变换之后也要改变形式. 这就是联络形式定义的最初来源.

存在定理

由转换公式式 1 此可得联络的存在定理，或者说，只要给出了满足转换公式式 1 的 1-形式矩阵，就等于给出了联络：

定理 1

设 $\mathfrak{U}$ 是流形 M 的开覆盖，使得在每一个开集 $U \in \mathfrak{U}$ 上都给定了切丛的局部标架 $\{e_i^U\}$，E 的局部标架 $\{s_\alpha^U\}$ 以及 1-形式的 $k \times k$ 方阵 ω^U．若对于任意两个相交的 $U, V \in \mathfrak{U}$，在 $U \cap V$ 总成立

$$\omega^V = db_U^V \cdot \{b_V^U\}^{-1} + b_U^V \cdot \omega^U \cdot \{b_U^V\}^{-1},$$

其中 b_U^V 是从 $\{s_\alpha^U\}$ 到 $\{s_\alpha^V\}$ 的转换矩阵，则存在唯一一个 E 上的联络，使得其在局部标架 $\{s_\alpha^U\}$ 之下的联络 1-形式矩阵是 $\{\omega^U\}$.

还可以适当选取 E 的标架，使得联络 1-形式矩阵在指定的点处有简单的形式：

定理 2

给定任何一点 $p \in M$ 以及一个 E 上的联络 D，都存在 E 在 p 附近的局部标架，使得这标架下的联络 1-形式矩阵 $\omega(p) = 0$.

证明大意先取一个 E 的标架 $\{s_\alpha\}$，并设

$$\omega_\alpha^\beta = \Gamma_{\alpha i}^\beta dx^i.$$

命 $\{x^i\}$ 为 p 附近的坐标系，使得 $x^i(p) = 0$．则命

$$b_\alpha^\beta = \delta_\alpha^\beta - \Gamma_{\alpha i}^\beta x^i,$$

并定义 E 的新的局部标架 $s'_\alpha = b_\alpha^\beta s_\beta$．则从 $db(p) = -\omega(p)$ 和转换公式式 1 就得到新标架下 $\omega'(p) = 0$.

注意，这不表示可以取到局部标架使得联络 1-形式矩阵在 p 的邻域内都等于零．某个局部标架下的联络系数局部为零是非常特殊的性质．详见平行性（向量丛）[713].

对偶联络; 联络的和与积

E 上的联络 D 自然诱导出对偶丛 E^* 上的联络 D^*：如果 $\xi \in \Gamma(E), \eta^* \in \Gamma(E^*)$，则

$$d\langle \eta^*, \xi \rangle = \langle D^*\eta^*, \xi \rangle + \langle \eta^*, D\xi \rangle.$$

如果 $\{s_\alpha\}$ 是 E 的局部标架而 $\{s^{*\gamma}\}$ 是对偶标架，那么 D^* 的联络 1-形式矩阵在此标架下为

$$\omega_\alpha^{*\beta} = \langle D^* s^{*\beta}, s_\alpha \rangle = -\langle s^{*\beta}, Ds_\alpha \rangle = -\omega_\alpha^\beta,$$

或者可简写为 $D^* s^{*\beta} = -\omega_\alpha^\beta s^{*\alpha}$. 若截面 $\eta^* \in \Gamma(E^*)$ 有局部表达式 $\eta^* = \eta_\beta s^{*\beta}$，则

$$D^*\eta^* = (d\eta_\beta - \omega_\beta^\alpha \eta_\alpha) \otimes s^{*\beta}.$$

如果 D_1, D_2 分别是向量丛 E_1, E_2 上的联络，则在直和 $E_1 \oplus E_2$ 和张量积 $E_1 \otimes E_2$ 上的和与积分别为

$$D(\xi_1 \oplus \xi_2) = D\xi_1 \oplus D\xi_2,\ D(\xi_1 \otimes \xi_2) = D\xi_1 \otimes \xi_2 + \xi_1 \otimes D\xi_2.$$

曲率 (向量丛)

预备知识 联络（向量丛）[708]

本节采用爱因斯坦求和约定.

设 M 是 n 维微分流形，E 是其上秩为 k 的光滑向量丛. 设给定了 E 上的联络 D.

定义

给定截面 $\xi \in \Gamma(E)$ 和切向量场 X, Y，联络 D 的沿着 X, Y 作用在 ξ 上的**曲率算子（curvature operator）**定义为

$$R(X,Y)\xi = -D_X D_Y \xi + D_Y D_X \xi + D_{[X,Y]}\xi.$$

直观来说，曲率算子是量度导数算子的可对易性的：如果取 X, Y 为坐标向量 ∂_i, ∂_j，那么 $[X,Y] = 0$，从而 $R(\partial_i, \partial_j)\xi$ 就是沿着坐标方向的二阶导数之差.

曲率方阵

为计算 $R(X,Y)\xi$，设 $\{s_\alpha\}_{\alpha=1}^k$ 是 E 的局部标架，$\xi=\xi^\alpha s_\alpha$ 为 ξ 在此标价下的局部表达式，ω 是此标架下的联络 1-形式矩阵．从而

$$D_X\xi = X(\xi^\alpha)\xi_\alpha + \xi^\beta\omega_\beta^\alpha(X)s_\alpha,$$

$$\begin{aligned} D_YD_X\xi &= Y[X(\xi^\alpha)]s_\alpha + Y[\xi^\beta\omega_\beta^\alpha(X)]s_\alpha + [X(\xi^\beta)+\xi^\gamma\omega_\gamma^\beta(X)]\omega_\beta^\alpha(Y)s_\alpha \\ &= Y[X(\xi^\alpha)]s_\alpha + Y(\xi^\beta)\omega_\beta^\alpha(X)s_\alpha + \xi^\beta Y[\omega_\beta^\alpha(X)]s_\alpha \\ &\quad + [X(\xi^\beta)+\xi^\gamma\omega_\gamma^\beta(X)]\omega_\beta^\alpha(Y)s_\alpha, \end{aligned}$$

于是

$$\begin{aligned} &(-D_XD_Y+D_YD_X)\xi \\ &= [Y,X](\xi^\alpha)s_\alpha + [Y(\xi^\beta)\omega_\beta^\alpha(X) - X(\xi^\beta)\omega_\beta^\alpha(Y)]s_\alpha + \left(\xi^\beta Y[\omega_\beta^\alpha(X)] - \xi^\beta X[\omega_\beta^\alpha(Y)]\right)s_\alpha \\ &\quad + \left(X(\xi^\beta)\omega_\beta^\alpha(Y) - Y(\xi^\beta)\omega_\beta^\alpha(X) + \xi^\gamma\omega_\gamma^\beta(X)\omega_\beta^\alpha(Y) - \xi^\gamma\omega_\gamma^\beta(Y)\omega_\beta^\alpha(X)\right)s_\alpha \\ &= [Y,X](\xi^\alpha)s_\alpha + \left(\xi^\beta Y[\omega_\beta^\alpha(X)] - \xi^\beta X[\omega_\beta^\alpha(Y)]\right)s_\alpha \\ &\quad + \left(\xi^\gamma\omega_\gamma^\beta(X)\omega_\beta^\alpha(Y) - \xi^\gamma\omega_\gamma^\beta(Y)\omega_\beta^\alpha(X)\right)s_\alpha \\ &= \left([Y,X](\xi^\alpha) + \xi^\beta\omega_\beta^\alpha[Y,X]\right)s_\alpha + \left(\xi^\beta Y[\omega_\beta^\alpha(X)] - \xi^\beta X[\omega_\beta^\alpha(Y)] - \xi^\beta\omega_\beta^\alpha[Y,X]\right)s_\alpha \\ &\quad + \left(\xi^\gamma\omega_\gamma^\beta(X)\omega_\beta^\alpha(Y) - \xi^\gamma\omega_\gamma^\beta(Y)\omega_\beta^\alpha(X)\right)s_\alpha \\ &= D_{[Y,X]}\xi - (d\omega_\beta^\alpha - \omega_\beta^\gamma\wedge\omega_\gamma^\alpha)(X,Y)\xi^\beta s_\alpha. \end{aligned}$$

故

$$R(X,Y)\xi = -(d\omega_\beta^\alpha - \omega_\beta^\gamma\wedge\omega_\gamma^\alpha)(X,Y)\xi.$$

由

$$\Omega_\beta^\alpha = d\omega_\beta^\alpha - \omega_\beta^\gamma\wedge\omega_\gamma^\alpha$$

给出的 1-形式的矩阵 $\Omega=(\Omega_\beta^\alpha)$ 称为联络 D 的**曲率方阵（curvature matrix）**.

张量运算

对于任何截面，$\xi\in\Gamma(E)$，都有

$$R(X,Y)\xi = -\Omega(X,Y)\xi.$$

容易验证在新的局部标架 $\{s'_\alpha\}$ 之下，如果它与原来的标架之间的转换公式为 $s_\alpha = b_\alpha^\beta s_\beta$，则有

$$\Omega' = b \cdot \Omega \cdot b^{-1},$$

因此 $(X, Y, \xi) \to R(X,Y)\xi$ 是张量运算：它对于 X, Y, ξ 都是线性的，而且只依赖于它们在某一点处的值.

给定切丛的局部标架 $\{e_i\}_{i=1}^n$ 后，可写 $R(X,Y)\xi = R_{ij\beta}^\alpha X^i Y^j s_\alpha$. 这个张量运算对于 X, Y 是反对称的，而且可写（注意负号!）

$$\Omega_\beta^\alpha = -\frac{1}{2} R_{ij\beta}^\alpha \theta^i \wedge \theta^j.$$

这里 $\{\theta^i\}_{i=1}^n$ 是同 $\{e_i\}_{i=1}^n$ 对偶的局部标架.

有些文献中 $R(X,Y)\xi$ 的定义比上文定义多一个负号，这样在曲率方阵的定义中就不必出现负号.

第二毕安基恒等式

将等式 $\Omega_\beta^\alpha = d\omega_\beta^\alpha - \omega_\beta^\gamma \wedge \omega_\gamma^\alpha$ 微分，得到

$$\begin{aligned} d\Omega_\beta^\alpha &= -d\omega_\beta^\gamma \wedge \omega_\gamma^\alpha + \omega_\beta^\gamma \wedge d\omega_\gamma^\alpha \\ &= -(\Omega_\beta^\gamma + \omega_\beta^\lambda \wedge \omega_\lambda^\gamma) \wedge \omega_\gamma^\alpha + \omega_\beta^\gamma \wedge (\Omega_\gamma^\alpha + \omega_\gamma^\lambda \wedge \omega_\lambda^\alpha) \\ &= \omega_\beta^\gamma \wedge \Omega_\gamma^\alpha - \Omega_\beta^\gamma \wedge \omega_\gamma^\alpha. \end{aligned}$$

这称为**第二毕安基恒等式（second Bianchi identity）**.

平行性 (向量丛)

预备知识 曲率（向量丛）[711]，费罗贝尼乌斯定理[703]

本节采用爱因斯坦求和约定.

设 M 是 n 维微分流形，E 是其上秩为 k 的光滑向量丛. 设给定了 E 上的联络 D.

平行截面

向量丛 E 的截面 $\xi \in \Gamma(E)$ 称为在联络 D 之下**平行的（parallel）**，如果

$$D\xi = 0.$$

在局部标架 $\{s_\alpha\}_{\alpha=1}^k$ 之下，如果 $\xi = \xi^\alpha s_\alpha$，$\omega$ 是此标架下的联络 1-形式矩阵，则 $D\xi = 0$ 等价于

$$d\xi^\alpha + \xi^\beta \omega_\beta^\alpha = 0,\ 1 \le \alpha \le k.$$

进一步给定切丛和余切丛的局部标架 $\{e_i\}, \{\theta_j\}$ (二者为对偶) 之后，就有了克氏符 $\Gamma_{\beta i}^\alpha$，因此上式进一步等价于方程组

$$e_i(\xi^\alpha) + \xi^\beta \Gamma_{\beta i}^\alpha = 0,\ 1 \le \alpha \le k,\ 1 \le i \le n.$$

这是普法夫系．在局部上，根据费罗贝尼乌斯定理[703]，这方程组可解（局部上相当于有 k 个函数 $\{\xi^\alpha\}_{\alpha=1}^k$ 满足上面的偏微分方程组）当且仅当

$$0 = d(\omega_\beta^\alpha \xi^\beta) = d\xi^\beta \wedge \omega_\beta^\alpha + \xi^\beta d\omega_\beta^\alpha = \xi^\beta \wedge \Omega_\beta^\alpha.$$

这里 Ω 是 D 的曲率方阵．这也就表示这个方程组可解当且仅当

$$\xi^\beta \wedge \Omega_\beta^\alpha = 0.$$

因此如果 D 的曲率在某个开集上等于零，则上述普法夫系在此开集上是局部可积的，于是此开集上存在 E 的局部平行截面．

这样看来，曲率是平行截面存在的障碍．

$\mathbb{R}^n$ 上平凡丛 $\mathbb{R}\times^n\mathbb{R}^k$ 上的平凡联络就是通常的微分运算，因此当然是可交换的，曲率算子为零．设 ∂_α 是 $\mathbb{R}^k$ 上的一个仿射坐标向量，则它在此联络下就是平行的．这很符合" 平行" 的直观意义．

平行移动

向量丛在区域上的平行截面很可能不存在，但却可以定义沿着某条道路平行的截面．

设 $\gamma : [0, a] \to M$ 是 Lipschitz 连续的道路，$\xi_0 \in E_{\gamma(0)}$ 是沿着道路的截面．它的严格定义是一个映射：$\xi : [0, a] \to E$，使得 $\xi(t) \in E_{\gamma(t)}$．称此截面是**沿着**

道路 γ 平行的（parallel along γ），如果 $D_{\gamma'}\xi(t)=0$．此时称 $\xi(a)$ 是 $\xi(0)$ 沿着 γ 的**平行移动（parallel transport）**.

在 M 的局部坐标系 $\{x^i\}_{i=1}^n$ 和 E 的局部标架 $\{s_\alpha\}_{\alpha=1}^k$ 之下，可写 $\gamma(t)=(\gamma^i(t))$，$\xi=\xi^\alpha s_\alpha$，则 $D_{\gamma'}\xi=0$ 等价于

$$\frac{d\xi^\alpha}{dt}+\Gamma^\alpha_{\beta i}(\gamma(t))\frac{d\gamma^i}{dt}(t)\xi^\beta=0.$$

这是 k 个函数 ξ^α 的齐次线性常微分方程组，系数是有界的，所以给定初值之后它有唯一的 Lipschitz 连续的解.

这就给出了一个线性变换 $P_\gamma: E_{\gamma(0)}\to E_{\gamma(a)}$，将任何一个 $\xi(0)\in E_{\gamma(0)}$ 通过平行移动变为 $\xi(a)\in E_{\gamma(a)}$．从微分方程本身可得出终值与 γ 重参数化的方法无关．简单计算给出 P_γ^{-1} 实为沿着反向道路 γ^{-1} 的平行移动.

若命 $P_\gamma^{\tau,t}$ 表示从 $E_{\gamma(\tau)}$ 到 $E_{\gamma(t)}$ 的沿着 γ 的平行移动，则简单地计算给出：对于任何截面 $\xi\in\Gamma(E)$ 和任何使得 $\gamma'(t)$ 存在的 t，皆有

$$D_{\gamma'(t)}\xi(\gamma(t))=\lim_{\tau\to t}\frac{P_\gamma^{\tau,t}\xi(\gamma(\tau))-\xi(\gamma(t))}{\tau-t}.$$

这样看来，沿着道路的平行移动实则决定了整个联络本身的取值.

对于 $\mathbb{R}^n$ 上平凡丛 $\mathbb{R}^n\times\mathbb{R}^k$ 上的平凡联络，沿着两条有同样起终点的道路的平行移动给出同样的结果：它就是通常意义下实数空间中的平行移动，移动的结果显然跟移动的路径无关．然而对于一般的联络来说这是不对的．这种现象是因为一般的联络曲率不为零．这是和乐定理[715] 要讨论的内容.

和乐群（向量丛）

预备知识 平行性（向量丛）[713]

本节采用爱因斯坦求和约定.

设 M 是 n 维微分流形，E 是其上秩为 k 的光滑向量丛．设给定了 E 上的联络 D.

和乐群

固定一点 $p\in M$．考虑纤维 E_p 上所有形如 $P_\gamma: E_p\to E_p$ 的平行移动，其中 γ 是起终点都在 p 处的 C^1 道路．容易验证：$P_\gamma\cdot P_\eta=P_{\gamma\cdot\eta}$，其中 $\gamma\cdot\eta$ 表示道

路的复合. 也容易验证沿着常值道路的平行移动是恒等变换，以及 $P_\gamma^{-1}=P_{\gamma^{-1}}$. 因此所有形如这样的变换构成了 E_p 上的一般线性群 $GL(E_p)$ 的子群，叫做 D 的基点在 p 处的**和乐群（holonomy group）**，常记为 $\mathrm{Hol}_p(D)$.

如果 $q\in M$ 是另外一点，那么 $\mathrm{Hol}_q(D)$ 与 $\mathrm{Hol}_p(D)$ 是同构的：只需要构造任何连接 p 和 q 的光滑道路就可以了. 如果限于考虑沿着零伦的闭道路的平行移动构成的群，则得到**限制和乐群（restricted holonomy group）**，记为 $\mathrm{Hol}_p^0(D)$. 容易看出商群 $\mathrm{Hol}_p(D)/\mathrm{Hol}_p^0(D)$ 可作为基本群 $\pi_1(M)$ 的商群.

群 $\mathrm{Hol}_p(D)$ 和 $\mathrm{Hol}_p^0(D)$ 都作为一般线性群 $GL(E_p)$ 的子群赋予拓扑. 如果 $\gamma:S^1\to M$ 是基点为 p 的零伦闭道路，那么存在同伦 $\gamma_u(t):[0,1]\times S^1\to M$ 使得 $\gamma_0(t)=\gamma(t)$，$\gamma_1(t)\equiv p$. 则每一个 P_{γ_u} 都是 $\mathrm{Hol}_p^0(D)$ 的元素，而且根据常微分方程的基本理论，$u\to P_{\gamma_u}$ 是连接 P_γ 和 id 的道路. 这样一来 $\mathrm{Hol}_p^0(D)$ 是 $GL(E_p)$ 的道路连通子群. 山边英彦（Yamabe）定理说：李群的道路连通子群是李子群，因此 $\mathbf{Hol}_p^0(D)$ **是** $GL(E_p)$ **的李子群**. 同理 $\mathrm{Hol}_p(D)$ 也是李子群. $\mathrm{Hol}_p^0(D)$ 实际上是 $\mathrm{Hol}_p^0(D)$ 的单位分支.

和乐定理

$\mathrm{Hol}_p^0(D)$ 和 $\mathrm{Hol}_p(D)$ 的李代数是一样的，因为二者的商群是可数的离散群. $\mathrm{Hol}_p^0(D)$ 的李代数叫做联络 D 的基点为 p 的**和乐代数（holonomy algebra）**，记为 $\mathfrak{h}_p(D)$. 它是 $\mathrm{End}(E_p)$ 的李子代数.

嘉当（E. Cartan）给出了曲率算子的一个直观解释，揭示了曲率算子与和乐群之间的关系. 固定两个切向量 $X,Y\in T_pM$. 考虑从二维三角形 Δ_2 (二维方块 $[0,1]^2$ 的左下角) 到 M 的光滑映射 $f:\Delta_2\to M$，使得 $f(0,0)=p$，$\partial_x f(0,0)=X$，$\partial_y f(0,0)=Y$. 令 $f_u(x,y):=f(ux,uy)$，这给出了从周线 $f_{\partial\Delta_2}:\partial\Delta_2\to M$ 到点 p 的一个收缩同伦. 设 t 是周线 $f_{\partial\Delta_2}$ 的弧长参数，并以 γ_u 表示参数为 u 的同伦闭道路. 则 $\gamma_u(t)=f(ux(t),uy(t))$，而且容易算出

$$\left.\frac{d}{du}P_u\right|_{u=0}=0,\ \left.\frac{d^2}{du^2}P_u\right|_{u=0}=-2R(X,Y).$$

粗略来说，这表示" 绕小周线一周" 得到的平行移动算子实际上是由曲率算子给出的. 将这个命题精细化，就可看出和乐群 $\mathrm{Hol}_p^0(D)$ 的无穷小生成元（也就是 $\mathfrak{h}_p(D)$ 的元素）是曲率算子. 由此可得到和乐定理（holonomy theorem）：

定理 1　Ambrose-Singer 和乐定理，向量丛版本

设 R 是联络 D 的曲率算子. 则对于任何一点 $p \in M$, 和乐代数 $\mathfrak{h}_p(D)$ 是由变换 $\{R(X,Y) : X, Y \in T_pM\}$ 的集合生成的.

第四章

黎曼几何

黎曼度量与伪黎曼度量

预备知识 内积[471]，切丛[698]，余切丛（未完成），张量场[??]

本节采用爱因斯坦求和约定.

黎曼度量（Riemannian metric）或伪黎曼度量（pseudo-Riemannian metric）是黎曼几何或伪黎曼几何所要求的基本结构. 赋予黎曼度量/伪黎曼度量的微分流形被称为黎曼流形/伪黎曼流形，它们既是几何学研究的对象，也是广义相对论得以展开的舞台.

这里将一直设 M 是一个 n 维实微分流形.

黎曼度量

定义 1　黎曼度量

M 上的一个黎曼度量 g 是指丛 $T^*(M)\otimes T^*(M)$ 的一个对称的正定截面. 等价地，给出黎曼度量 g，就相当于在每一点 p 的切空间 T_pM 上指定一个内积 g_p. 指定了黎曼度量的微分流形称为黎曼流形. 有时也会把黎曼度量记为 $\langle\cdot,\cdot\rangle_p$.

给定局部坐标系 $\{x^i\}$ 后，黎曼度量 g 的局部表达式是

$$g_{ij}(x)dx^i\otimes dx^j,$$

这里 $g_{ij}(x)=g_{ji}(x)$，而且对于任何向量 $(X^i)_{i=1}^n\neq 0$ 都有

$$g_{ij}X^iX^j>0.$$

有了黎曼度量，就可以谈论诸如长度/面积/体积之类的度量性质了. 例如，设 $\gamma:[a,b]\to M$ 是黎曼流形 (M,g) 上的一条道路，则定义其长度为

$$L[\gamma]:=\int_a^b\sqrt{g_{\gamma(t)}(\gamma'(t),\gamma'(t))}dt.$$

如果用局部坐标系给出曲线的局部参数方程 $x(t)=(x^i(t))_{i=1}^n$，则

$$L[\gamma]:=\int_a^b\sqrt{g_{ij}(x(t))(\dot{x}^i(t),\dot{x}^j(t))}dt.$$

容易验证：曲线的长度不依赖于其参数化的方式.

洛伦兹度量

Levi-Civita 联络

预备知识　黎曼度量[719]，联络（向量丛）[708]

第七部分

概率与统计

第一章

基础

随机变量 概率分布函数

预备知识 定积分[198]

生活中有许多现象可以看做是随机的，例如掷骰子的点数．事实上骰子作为一个宏观物体，其运动可以用一个复杂的动力学方程来精确描述．但经过诸如“摇匀”这种混沌过程后，方程的最终结果对初始条件极为敏感，使结果难以预测．这时我们就有充分的理由将该结果看作是随机的，并用一个变量来表示可能的结果（就像把方程中的未知数用 x 表示）．我们把这样的变量称为**随机变量**．

随机变量可以是**离散**的也可以是**连续**的，例如掷骰子的点数只能取 1 到 6 的离散值，而打靶时子弹离靶心的距离就可以用一个连续的随机变量表示．一些更复杂事件的结果可能需要用到不止一个随机变量来描述，本文只讨论单个随机变量，但结论容易拓展到多个变量．

对于一些离散的随机变量，可能发现每个离散值得到的概率也都是恒定的．对一个公平的骰子，所有的点数得到的概率都是 $1/6$；对一个公平的硬币，掷到正反两面的概率都是 $1/2$．如果骰子或硬币是不公平的，不同结果会对应不同的概率，但这些概率也是固定的．对于连续的随机变量，得到不同值的概率可能也是固定的，然而这些值有无穷多个，应该如何描述他们对应的概率呢？

连续随机变量的分布函数

我们可以用**概率分布函数（probability distribution function，PDF）**来描述一个变量取各个值的概率．假设一个连续随机变量 x 可以在某个区间内取值，我们就把该区间分为 n 份，第 i 个子区间的长度为 Δx_i 然后我们做大量的实验（记为 N 次），把随机变量得到的每个值分类归入这 n 个子区间中，并把第 i 个区间中值的个数记为 N_i．现在我们可以画出一种表示概率的**直方图（histogram）**，令第 i 个区间的长方形高度为 $y_i = N_i/(N\Delta x_i)$，则每个长方形的面积 $y_i\Delta x_i = N_i/N$ 表示随机变量的值落在第 i 个区间的概率，注意所有长方形的面积之和为 1．

现在，我们令区间数 $n \to \infty$ 且每个区间长度 $\Delta x_i \to 0$，则离散的 y_i 值就可以表示为函数 $y = f(x)$．我们可以用定积分来表示“所有长方形的面积之和为 1”，即[1]

$$\int_{-\infty}^{+\infty} f(x) = 1 \tag{1}$$

该式叫做概率分布函数的**归一化**．满足归一化意味着，所有情况发生的概率总和为 1．

若我们要求随机变量落在区间 $[a, b]$ 内的概率，就求 $[a, b]$ 区间内分布函数下方的面积即可．更常见地，我们可以用微分式

$$\mathrm{d}P = f(x)\,\mathrm{d}x \tag{2}$$

表示 x 处长度为 $\mathrm{d}x$ 的区间微元对应的概率 $\mathrm{d}P$．所以 $f(x)$ 又被称为**概率密度（probability density）**．

平均值

大学物理中，随机变量 x 的平均值通常被表示为 $\bar{x}$ 或者 $\langle x \rangle$，我们以后都会使用．

对于离散的情况，某个量的平均值等于每个可能的值出现的概率乘以该值再求和，即

$$\langle x \rangle = \sum_i x_i P_i \tag{3}$$

要求某个分布的平均值，我们同样可以将整个区间划分为 n 个子区间，每个区间的概率近似为 $P_i = f(x_i)\Delta x_i$，则平均值为

$$\langle x \rangle \approx \sum_{i=0}^{n} x_i P_i = \sum_{i=1}^{n} x_i f(x_i) \Delta x_i \tag{4}$$

用定积分[198] 的思想，当子区间无限多且取无限小时，上式变为

$$\langle x \rangle = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x \tag{5}$$

[1]注意积分上下限是 x 取值的区间，以下为了方便表示，我们取整个实数域，可以理解为超出区间的部分概率分部函数为 0．

方差

离散情况下，若已知平均值 $\langle x\rangle$，**方差**（每个数据点离平均值距离的平方的平均值）可定义为

$$\sigma_x^2 \approx \sum_{i=0}^{n}(x_i-\bar{x})^2 P_i \tag{6}$$

与计算平均值的思路类似，将方差拓展到连续变量的情况得

$$\sigma_x^2 = \int_{-\infty}^{+\infty}(x-\bar{x})^2 f(x)\,\mathrm{d}x \tag{7}$$

习题 1

某直流电源存在微小误差，其电压随时间的函数为

$$U(t) = U_0 + \varepsilon\sin(\omega t) \tag{8}$$

为衡量误差大小，请计算电压的方差（用 ε 表示）. 提示：由于电压变化是周期性的，可以只在一个周期内积分.

任意函数的平均

更一般地，我们可以对离散的随机变量 x_i 定义任意函数 $g(x)$ 的平均值为

$$\langle g\rangle = \sum_{i=0}^{n} g(x_i)P_i \tag{9}$$

例如在计算平均值和方差时，$g(x)$ 分别取 x 和 $(x-\bar{x})^2$.

拓展到连续的随机变量，有

$$\langle g\rangle = \int_{-\infty}^{+\infty} g(x)f(x)\,\mathrm{d}x \tag{10}$$

例 1　分子的平均动能

某气体中含有大量分子（阿伏伽德罗常数数量级：10^{23}），若假设某时刻它们的速度大小 v 的分布函数为

$$f(v) = A\sin^2\left(\frac{\pi v}{v_{max}}\right) \qquad (v\in[0,v_{max}]) \tag{11}$$

其中 A 为常数. 请分别计算：

1. 常数 A，使 $f(v)$ 满足归一化（式 1）
2. 分子速度大小的平均值
3. 分子速度大小方差
4. 分子动能 $E_k = mv^2/2$ 的平均值
5. 分子动能的方差

随机变量的变换

预备知识 概率分布函数[723]

求新变量的分布函数

我们先来讨论这样一个问题：令两个随机变量 x_1, x_2 间有函数关系 $x_1 = g(x_2)$，若已知 x_1 的分布函数为 $f_1(x_1)$，求 x_2 的分布函数 $f_2(x_2)$.

将两个概率分布写成微分形式，有

$$\mathrm{d}P = f_1(x_1)\,\mathrm{d}x_1 \tag{1}$$

$$\mathrm{d}P = f_2(x_2)\,\mathrm{d}x_2 \tag{2}$$

若将式 1 中的 x_1 替换成 $g(x_2)$，$\mathrm{d}x_1$ 替换成 $g'(x_2)\,\mathrm{d}x_2$，有

$$\mathrm{d}P = f_1[g(x_2)]g'(x_2)\,\mathrm{d}x_2 \tag{3}$$

对比式 2，得

$$f_2(x) = f_1[g(x)]g'(x) \tag{4}$$

这样，就求出了 x_2 的分布函数 $f_2(x)$.

例 1

已知 $f_1(x_1) = 3x_1^2$，$x_1 \in [0,1]$，$x_2 = x_1^2$，求 x_2 的分布函数.

用 x_2 表示 x_1 得 $x_1 = \sqrt{x_2}$，代入 $\mathrm{d}P = f_1(x_1)\,\mathrm{d}x_1$，得

$$\mathrm{d}P = 3\sqrt{x_2}^2\,\mathrm{d}(\sqrt{x_2}) = \frac{3}{2}\sqrt{x_2}\,\mathrm{d}x_2 \tag{5}$$

所以 x_2 的分布函数为 $f_2(x_2) = 3\sqrt{x_2}/2$.

求两变量的关系

预备知识 可分离变量的微分方程

另一个常见的问题是已知 x_1 和 x_2 的分布函数 $f_1(x_1), f_2(x_2)$，求两个随机变量需要满足的函数关系.

对比式 1 和式 2 可得一个已分离变量的微分方程

$$f_1(x_1)\,\mathrm{d}x_1 = f_2(x_2)\,\mathrm{d}x_2 \tag{6}$$

将方程两边积分即可得到两变量所满足的函数关系

$$F_1(x_1) = F_2(x_2) + C \tag{7}$$

其中函数 F_1, F_2 分别是函数 f_1, f_2 的一个原函数，待定常数 C 通常可以由 x_1 和 x_2 的取值范围确定.

这个问题的一个常见应用是在程序中生成指定分布函数的随机变量. 在许多编程语言中，随机数生成器只能生成一个从 0 到 1 均匀分布的随机变量（即 $f(x) = 1$），若我们需要一个其他分布的随机变量，就可以使用以上方法.

例 2

已知随机变量 x_1 $(x_1 \in [0,1])$ 的分布函数为 $f_1(x_1) = 1$，求函数关系 $x_2 = g(x_1)$ 使得 x_2 的分布函数为 $f_2(x_2) = 2x_2$ $(x_2 \in [0,1])$.

将 f_1, f_2 代入式 6 并两边积分得

$$x_1 = x_2^2 + C \tag{8}$$

由于 x_1 和 x_2 的区间关系得 $C = 0$，所以有 $x_2 = \sqrt{x_1}$.

例 3

给出两个随机变量 ξ_1, ξ_2 $(\xi_1, \xi_2 \in [0,1])$，分布函数均为 $f(\xi_i) = 1$，用 ξ_1, ξ_2 表示某随机点的极坐标 (r, θ) 使得该点在单位圆内均匀随机分布.

要使随机点在单位圆内随机分布，θ 显然应该在 $[0, 2\pi]$ 间均匀随机分布，所以令 $\theta = 2\pi\xi_2$ 即可. 要决定 r 的分布函数，我们把单位圆划分为许多小圆环，随机点出现在某圆环内的概率等于该圆环的面积比单位圆的面积，即

$$\mathrm{d}P = \frac{2\pi r\,\mathrm{d}r}{\pi} = 2r\,\mathrm{d}r \tag{9}$$

所以 r 的分布函数为 $2r$. 令 r 与 ξ_1 间存在函数关系，由式 6 得

$$1\,\mathrm{d}\xi_1 = 2r\,\mathrm{d}r \tag{10}$$

两边积分得 $\xi_1 = r^2$ （$\xi_1 = 0$ 时 $r = 0$，所以积分常数为零），即 $r = \sqrt{\xi_1}$. 所以，随机点的极坐标取 $(\sqrt{\xi_1}, 2\pi\xi_2)$ 即可.

多变量分布函数

预备知识 概率分布函数[723]，重积分[283]

二维分布函数

我们以打靶为例来引入二元概率分布函数，以及在直角坐标系和极坐标系中如何分析各种平均值. 我们在靶上建立 x-y 直角坐标系，类比一维概率分布函数的定义，二维概率分布可以用一个二元函数表示为 $f(x,y)$，也叫**概率密度**（**probability density**）. 子弹落在一个长方形区域内（令 $x \in [x_1, x_2]$，$y \in [y_1, y_2]$）的概率可以用二重积分表示

$$P = \int_{y_1}^{y_2}\int_{x_1}^{x_2} f(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{1}$$

从几何上理解，$f(x,y)$ 可以看作一张三维空间中的曲面，而这个二重积分则表示曲面和 x-y 平面之间的一个体积（图 1）.

类比一维情况，概率归一化条件为

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1 \tag{2}$$

由于概率的量纲是 1，$f(x,y)$ 的国际单位量纲就是 m^{-2}，即面积的倒数.

平均值

类比式 9 [723]，离散情况下，令 P_i 是 (x_i, y_i) 出现的概率，函数 $g(x,y)$ 的平均值为

$$\langle g\rangle = \sum_{i=1}^{N} g(x_i, y_i)P_i \tag{3}$$

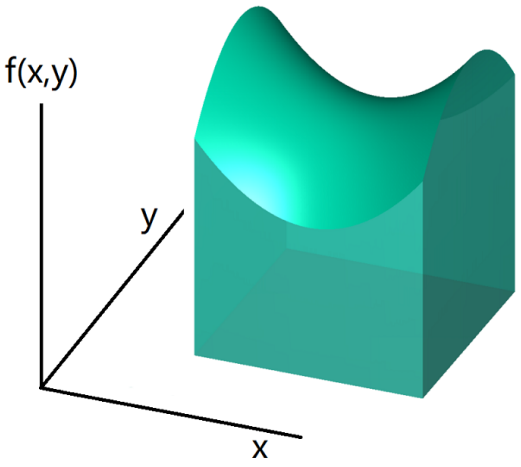


图 1: $f(x,y)$ 曲面下的体积

拓展到连续的情况，类比式 10 [723]，$g(x,y)$ 的平均值可以定义为

$$\langle g\rangle = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{4}$$

例如位置矢量的 x 和 y 分量平均值分别为

$$\langle x\rangle = \iint xf(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{5}$$

$$\langle y\rangle = \iint yf(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{6}$$

以上两式也可以用位置矢量 $\mathbf{r}$ 表示为（这时 $\mathbf{g}$ 的函数值是矢量）

$$\langle \mathbf{r}\rangle = \iint \mathbf{r}f(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{7}$$

点 (x,y) 到原点距离的距离 r 的平均值

$$\langle r\rangle = \iint \sqrt{x^2+y^2}f(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{8}$$

注意区分该式和式 7 式的意义，平均位置和平均距离是两个不同的概念：如果 $f(x,y)$ 是关于原点中心对称的函数，则平均位置为 $(0,0)$，而平均距离不为零.

距离平方 r^2 的平均值

$$\langle r^2\rangle = \iint (x^2+y^2)f(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{9}$$

概率分布的极坐标表示

预备知识 极坐标系[119]

除了直角坐标系外，我们也可以使用极坐标[119]表示分布函数，记为[2] $f(r,\theta)$，那么概率归一化条件变为

$$\int_0^\infty \int_0^{2\pi} f(r,\theta)\ r\,\mathrm{d}\theta\,\mathrm{d}r = 1 \tag{10}$$

函数 $g(r,\theta)$ 的平均值变为

$$\langle g\rangle = \int_0^\infty \int_0^{2\pi} g(r,\theta) f(r,\theta)\ r\,\mathrm{d}\theta\,\mathrm{d}r \tag{11}$$

例如式 9 变为

$$\left\langle r^2\right\rangle = \int_0^\infty \int_0^{2\pi} r^2 f(r,\theta)\ r\,\mathrm{d}\theta\,\mathrm{d}r = \int_0^\infty \int_0^{2\pi} r^3 f(r,\theta)\,\mathrm{d}\theta\,\mathrm{d}r \tag{12}$$

圆对称分布

继续打靶的例子，我们接下来讨论圆对称的概率分布（在三维情况下叫做球对称），即 $f(r,\theta)$ 的值只和 r 有关而与 θ 无关的情况．为了书写方便可记 $f(r)\equiv f(r,\theta)$．归一化条件式 10 化简为（先完成关于 θ 积分得 2π）

$$2\pi\int_0^\infty r f(r)\,\mathrm{d}r = 1 \tag{13}$$

一个容易混淆的概念是，$f(r)$ 并不是变量 r 的概率分布函数．也就是说 $\int_a^b f(r)\,\mathrm{d}r$ 并不是子弹落在圆环 $r\in[a,b]$ 内的概率．若将 r 的概率分布函数记为 $F(r)$，子弹落在圆环中的概率应该是

$$P_{ab} = \int_a^b F(r)\,\mathrm{d}r = \int_a^b 2\pi r f(r)\,\mathrm{d}r \tag{14}$$

由于两个定积分在任何区间都相等，两个被积函数必须相等，即

$$F(r) = 2\pi r f(r) \tag{15}$$

这时以上的平均值也可以得到进一步化简，例如式 12 变为

$$\left\langle r^2\right\rangle = \int_0^\infty r^2 F(r)\,\mathrm{d}r = 2\pi\int_0^\infty r^3 f(r)\,\mathrm{d}r \tag{16}$$

[2]在数学上，该式两边是两个不同的二元函数，应该使用不同的函数名如 $g(r,\theta)=f(x,y)$，但在一些物理教材中，我们有时为了方便不区分函数名，直接把 $g(r,\theta)$ 写成 $f(r,\theta)$.

高斯分布（正态分布）

- 本词条存在未完成的内容.

预备知识 分布函数[723]

高斯分布函数为

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{1}$$

其中 μ 是分布的的平均值，σ 是标准差．满足归一化条件

$$\int_{-\infty}^{\infty}f(x)\,\mathrm{d}x=1 \tag{2}$$

未完成：图，例题，平均值，方差

推导

高斯分布（Gaussian Distribution）又叫**正态分布（Normal Distribution）**，具有如下形式

$$f(x)=A\exp\left[-\lambda(x-x_0)^2\right] \tag{3}$$

可见其主要特征就是指数函数中含有 Δx^2 项．由对称性，分布函数是关于 $x=x_0$ 的偶函数，所以平均值显然为 $\mu=x_0$．首先我们补充两个积分，由换元积分法[188]（$x=\sqrt{t}$）以及 Γ 函数[254] 的性质得

$$\int_{-\infty}^{+\infty}\exp\left(-x^2\right)\mathrm{d}x=\int_{0}^{+\infty}t^{-1/2}\mathrm{e}^{-t}\,\mathrm{d}t=\left(-\frac{1}{2}\right)!=\sqrt{\pi} \tag{4}$$

$$\int_{-\infty}^{+\infty}x^2\exp\left(-x^2\right)\mathrm{d}x=\int_{0}^{+\infty}t^{1/2}\mathrm{e}^{-t}\,\mathrm{d}t=\frac{1}{2}!=\frac{1}{2}\left(-\frac{1}{2}\right)!=\frac{\sqrt{\pi}}{2} \tag{5}$$

根据分布函数的归一化条件，结合式 4 得

$$1=\int_{-\infty}^{+\infty}f(x)\,\mathrm{d}x=A\int_{-\infty}^{+\infty}\exp\left[-\lambda(x-x_0)^2\right]\mathrm{d}x=A\sqrt{\frac{\pi}{\lambda}} \tag{6}$$

即 $A=\sqrt{\lambda/\pi}$. 再来计算高斯分布的方差，结合式 5 得

$$\sigma^2=\int_{-\infty}^{+\infty}(x-x_0)^2A\exp\left[-\lambda(x-x_0)^2\right]\mathrm{d}x=\frac{1}{2\lambda} \tag{7}$$

用式 6 和式 7 解得 $\lambda=1/(2\sigma^2)$ 和 $A=1/(\sigma\sqrt{2\pi})$，代入式 3 可得高斯分布式 1.

二项分布

预备知识 二项式定理[107]

二项分布

如果记 X 为 n 重伯努利试验中成功（记为事件 A）的次数，则 X 的可能取值为 $0,1,\ldots,n$. 记 p 为每次试验中 A 发生的概率，即 $P(A)=p$，则 $P(\overline{A})=1-p$.

而我们知道，n 重伯努利试验的基本结果可以记作

$$\omega=(\omega_1,\omega_2,\cdots,\omega_n) \tag{1}$$

其中 ω_i 为 A，或者为 $\overline{A}$. 这样的 ω 共有 2^n 个，这 2^n 个样本点 ω 组成了样本空间 Ω.

下面求事件 X 的分布列，即 $\{X=k\}$ 的概率. 若某个样本点

$$\omega=(\omega_1,\omega_2,\cdots,\omega_n)\in\{X=k\} \tag{2}$$

意味着 $\omega_1,\omega_2,\cdots,\omega_n$ 中有 k 个 A，$n-k$ 个 $\overline{A}$. 由事件的独立性知：

$$P(\omega)=p^k(1-p)^{n-k} \tag{3}$$

而事件 $\{X=k\}$ 中这样的 ω 共有 $\binom{n}{k}$ 个，所以 X 的分布列为：

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k},k=0,1,\cdots,n \tag{4}$$

这个分布成为**二项分布**，记为 $X\sim b(n,p)$.

那么它的和是不是为 1 呢？这很容易验证．我们有

$$\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}=[p+(1-p)]^n=1 \tag{5}$$

并且从上式可以看出，二项概率 $\binom{n}{k}p^k(1-p)^{n-k}$ 恰好是二项式 $[p+(1-p)]^n$ 的展开式中的第 $k+1$ 项，这正是其名称的由来．

显然，二项概率是一种离散分布．它很常见，举例来说：

例 1　二项分布的例子

- 检查 10 件产品，10 件产品中不合格品的个数 X 服从二项分布 $b(10,p)$，其中 p 为不合格品率．
- 调查 50 个人，50 个人中患色盲的人数 Y 服从二项分布 $b(50,p)$，其中 p 为色盲率．
- 射击 5 次，5 次中命中次数 Z 服从二项分布 $b(5,p)$，其中 p 为射手的命中率．

下面来计算一些具体的例题．

例 2　治愈的概率

某特效药的临床有效率为 0.95，今有 10 人服用，问至少有 8 人治愈的概率是多少？

直接计算，我们有

$$\begin{aligned}P(X\geqslant 8)&=P(X=8)+P(X=9)+P(X=10)\\&=\binom{10}{8}0.95^8 0.05^2+\binom{10}{9}0.95^9 0.05+\binom{10}{10}0.95^{10}\\&=0.0746+0.3151+0.5988=0.9885\end{aligned} \tag{6}$$

即 10 人中有 8 人以上被治愈的概率为 0.9885．

例 3

设随机变量 $X\sim b(2,p), Y\sim b(3,p)$．若 $P(X\geqslant 1)=5/9$, 试求 $P(Y\geqslant 1)$．

由 $P(X \geqslant 1) = 5/9$，知 $P(X = 0) = 4/9$，所以 $(1-p)^2 = 4/9$，由此得 $p = 1/3$．再由 $Y \sim b(3,p)$ 可得

$$P(Y \geqslant 1) = 1 - P(Y = 0) = 1 - \left(1 - \frac{1}{3}\right)^3 = \frac{19}{27} \tag{7}$$

二点分布

$n = 1$ 时的二项分布 $b(1,p)$ 称为**二点分布**，或称 **0-1 分布**，或称**伯努利分布**．其分布列为

$$P(X = x) = p^x(1-p)^{1-x}, x = 0, 1 \tag{8}$$

或者，用表格可以列为：

x	0	1
P	$1-p$	p

(9)

二点分布 $b(1,p)$ 主要用来描述一次伯努利试验中成功（记为 A）的次数（0 或 1）．很多随机现象的样本空间 Ω 常可一分为二，记为 A 与 $\overline{A}$，由此形成伯努利试验．n 重伯努利试验是由 n 个相同的、独立进行的伯努利试验组成，若将第 i 个伯努利试验中 A 出现的次数记为 $X_i (i = 1, 2, \cdots, n)$, 则 X_i 相互独立，且服从相同的二点分布 $b(1,p)$．此时其和 $X = X_1 + X_2 + \cdots + X_n$ 就是 n 重伯努利试验中 A 出现的总次数，它服从二项分布 $b(n,p)$．这就是二项分布 $b(n,p)$ 与二点分布 $b(1,p)$ 之间的联系，即服从二项分布的随机变量是 n 个独立同为二点分布的随机变量之和．

二项分布的数学期望和方差

设随机变量 $X \sim b(n,p)$，则

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}(1-p)^{(n-1)} \\ &= np[p + (1-p)]^{n-1} = np \end{aligned} \tag{10}$$

又因为

$$\begin{aligned}
E\left(X^2\right) &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} (k-1+1) k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} + \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} + np \\
&= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} (1-p)^{(n-2)-(k-2)} + np \\
&= n(n-1)p^2 + np
\end{aligned} \tag{11}$$

由此得 X 的方差为

$$\operatorname{Var}(X) = E\left(X^2\right) - (E(X))^2 = n(n-1)p^2 + np - (np)^2 = np(1-p) \tag{12}$$

因为二点分布是 $n = 1$ 时的二项分布 $b(1, p)$，所以二点分布的数学期望为 p，方差为 $p(1-p)$.

为了看出不同的 p 的值，其二项分布 $b(n, p)$ 的变化情况，我们给出 $n = 10$ 时，不同 p 值的二项分布的概率值：我们再来看一下按照上面的数据描绘出的

表 1: 二项分布 $b(n, p)$

k	0	1	2	3	4	5	6	7	8	9	
b(10,0.2)	0.107	0.268	0.302	0.201	0.088	0.027	0.006	0.001			
b(10,0.5)	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.
b(10,0.8)				0.001	0.006	0.027	0.088	0.201	0.302	0.268	0.

线条图：

从上图可以看出：

- 位于均值 np 附近概率较大.
- 随着 p 的增加，分布的峰逐渐右移.

下面来看一个例题.

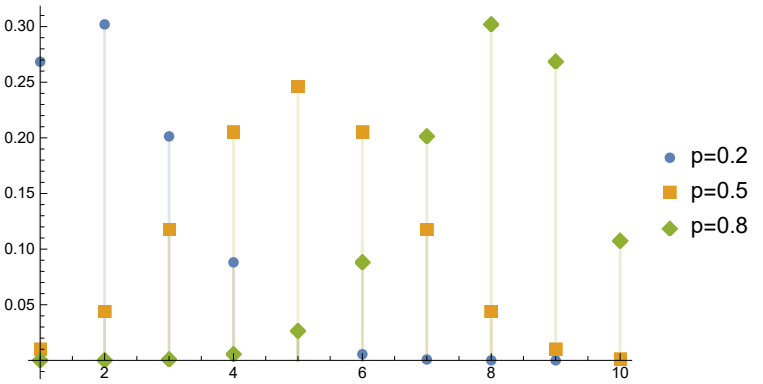


图 1: 二项分布 $b(n,p)$ 的线条图

例 4

甲、乙两棋手约定进行 10 局比赛，以赢的局数多者为胜．设在每局中甲赢的概率为 0.6，乙赢的概率为 0.4．如果各局比赛是独立进行的，试问甲胜、乙胜、不分胜负的概率各为多少？

以 X 表示 10 局比赛中甲赢的局数，则 $X\sim b(10,0.6)$．所以

$$\begin{aligned} P(\text{甲胜})&=P(X\geqslant 6)=\sum_{k=6}^{10}\binom{10}{k}0.6^k0.4^{10-k}=0.6330\\ P(\text{乙胜})&=P(X\leqslant 4)=\sum_{k=0}^{4}\binom{10}{k}0.6^k0.4^{10-k}=0.1663\\ P(\text{不分胜负})&=P(X=5)=\binom{10}{5}0.6^50.4^5=0.2007 \end{aligned}\tag{13}$$

可见甲胜的可能性达 63.3%，而乙胜的可能性只有 16.63%，它比不分胜负的可能性还要小．最后两个概率之和 0.3670 表示乙不输的概率．

中心极限定理

预备知识 高斯分布

结论

若一个随机变量 x 的平均值为零，方差为 $\langle x^2\rangle$，令 $X=\sum_i^N x_i$，当 $N\gg 1$ 时，X 满足高斯分布，且方差为

$$\left\langle X^2\right\rangle = N\left\langle x^2\right\rangle \tag{1}$$

证明略.

二维随机走动

预备知识 中心极限定理[736]

结论

若平面上某点从坐标原点出发，每一步沿随机方向走动一个随机步长，步长的分布函数为 $f(r)$，N $(N\gg 1)$ 步之后，该点的位置分布可以用**圆高斯分布**表示

$$P(X,Y)=\frac{a}{\pi}\mathrm{e}^{-a(X^2+Y^2)} \quad 或 \quad P(R)=2aR\mathrm{e}^{-aR^2} \tag{1}$$

其中

$$a=\frac{1}{N\left\langle r^2\right\rangle} \qquad \left\langle r^2\right\rangle=\int_0^\infty r^2 f(r)\,\mathrm{d}r \tag{2}$$

由分布函数可得，随机点最终离原点的距离的平均值和方均根为

$$\langle R\rangle=\frac{\sqrt{\pi}}{2}\sqrt{N\left\langle r^2\right\rangle} \qquad \sqrt{\langle R^2\rangle}=\sqrt{N\left\langle r^2\right\rangle} \tag{3}$$

推导

我们先来分析随机点的 x 坐标．假设每一步在 x 方向投影的长度为 x_i，N 步以后，该点的 x 坐标为 X，则

$$\left\langle x^2\right\rangle=\int_0^\infty\int_0^{2\pi}(r\cos\theta)^2\cdot f(r)\,\mathrm{d}r\cdot\frac{1}{2\pi}\,\mathrm{d}\theta=\frac{1}{2}\left\langle r^2\right\rangle \tag{4}$$

根据中心极限定理[736]，X 满足高斯分布，且

$$\left\langle X^2\right\rangle = N\left\langle x^2\right\rangle = \frac{1}{2}N\left\langle r^2\right\rangle \tag{5}$$

对 y 轴分析也有类似的结果，将 $P(X,Y)$ 分布归一化后，可以得到式 1．我们由式 1 求 $\langle X^2\rangle$，得

$$\left\langle X^2\right\rangle = \int_0^\infty\int_0^\infty X^2P(X,Y)\,\mathrm{d}X\,\mathrm{d}Y = \int_0^\infty\sqrt{\frac{a}{\pi}}X^2\mathrm{e}^{-aX^2}\,\mathrm{d}X = \frac{1}{2a} \tag{6}$$

对比式 5 和式 6 即可得到式 2．证毕．

平均值的不确定度

- 本词条处于草稿阶段．
- 本词条缺少预备知识，初学者可能会遇到困难．

例题：比如说有一个概率分布 $y=f(x)$．它的

平均值（数学期望）为 $\bar{x}=\int_{-\infty}^{+\infty}xf(x)\,\mathrm{d}x$，

方差为 $\sigma_x^2=\int_{-\infty}^{+\infty}(x-\bar{x})^2f(x)\,\mathrm{d}x=\left(\int_{-\infty}^{+\infty}x^2f(x)\,\mathrm{d}x\right)-\bar{x}^2=\overline{x^2}-\bar{x}^2$

标准差为 $\sigma_x=\sqrt{\int_{-\infty}^{+\infty}(x-\bar{x})^2f(x)\,\mathrm{d}x}$．

如果测量一个数据，这三个值就可以用来衡量这个数据的特征．

但如果测量 n 次平均值，那显然平均值显然要比一次测量更可靠，$\sigma_{\bar{x}}<\sigma_x$．各种教科书上都会给出 $\sigma_{\bar{x}}=\frac{1}{\sqrt{n}}\sigma_x$ 或者 $\sigma_{\bar{x}}^2=\frac{1}{n}\sigma_x^2$．那么这个公式到底怎么来的呢？

其实在上式中，$\sigma_{\bar{x}}^2$ 的定义是

$$\sigma_{\bar{x}}^2=\int_{-\infty}^{+\infty}\left(\frac{1}{n}\sum_{i=1}^{n}x_i-\bar{x}\right)^2f(x_1)f(x_2)\ldots f(x_n)\,\mathrm{d}x_1\ldots\mathrm{d}x_n \tag{1}$$

下面就从这个定义证明 $\sigma_{\bar{x}}^2=\frac{1}{n}\sigma_x^2$．

先考虑两次测量，即 $n=2$ 的情况．先后得到 x_1,x_2 的概率密度是 $f_2(x_1,x_2)=f(x_1)f(x_2)$．不难证明归一化：

$$\iint f(x_1)f(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2=\int f(x_1)\,\mathrm{d}x_1\int f(x_2)\,\mathrm{d}x_2=1\times1=1 \tag{2}$$

先看 $(x_1+x_2)/2$ 的平均值，令 $y=(x_1+x_2)/2$.

$$\begin{aligned}\bar{y}&=\iint\frac{x_1+x_2}{2}f(x_1)f(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2\\&=\frac{1}{2}\int x_1f(x_1)\,\mathrm{d}x_1\int f(x_2)\,\mathrm{d}x_2+\frac{1}{2}\int f(x_1)\,\mathrm{d}x_1\int x_2f(x_2)\,\mathrm{d}x_2\\&=\frac{\bar{x}}{2}+\frac{\bar{x}}{2}=\bar{x}\end{aligned}\tag{3}$$

结论是，进行两次测量取平均值，数学期望就是测量一次的数学期望．这个结论是符合常识的．

根据同样的方法，可以测量方差．

$$\begin{aligned}\sigma_{\bar{x}}^2&=\iint\left(\frac{x_1+x_2}{2}-\bar{x}\right)^2f(x_1)f(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2\\&=\frac{1}{2}\left(\overline{x^2}-\bar{x}^2\right)=\frac{1}{2}\sigma_x^2\end{aligned}\tag{4}$$

所以 $\sigma_{\bar{x}}^2=\frac{1}{2}\sigma_x^2$，且 $\sigma_{\bar{x}}=\frac{1}{\sqrt{2}}\sigma_x$

对于 $n>2$ 的情况，利用求和符号和积分运算法则，也很容易证明 $\sigma_{\bar{x}}=\frac{1}{\sqrt{n}}\sigma_x$

第八部分

偏微分方程和特殊函数

第一章

基础

分离变量法

- 本词条处于草稿阶段.

预备知识 多元函数的傅里叶级数[291]

在求解各种物理问题中的偏微分方程时，我们经常使用**分离变量法**. 在分离变量法中，我们假设微分方程的解可以表示为

$$f(x_1,\dots,x_N)=\sum_{i_1,\dots,i_N} c_{i_1,\dots,i_N} f_{i_1}(x_1)f_{i_2}(x_2)\dots f_{i_N}(x_N) \tag{1}$$

即每个变量都具有一组一元函数，这些一元函数的乘积的线性组合可以表示方程的解. 若将该式代入偏微分方程，可以分别得到关于每个变量 x_i 的常微分方程，我们就说这个偏微分方程式**可分离变量**的.

拉普拉斯方程

预备知识 梯度[307]，散度[311]

若令一个函数的拉普拉斯等于零，那么我们就得到了一个**偏微分方程**，即**拉普拉斯方程**

$$\nabla^2 u=0 \tag{1}$$

从物理上，二元函数的拉普拉斯方程可以理解为一个静止的，不受外力的薄膜[1020].

球坐标系中的矢量算符

预备知识 旋度[315]

球坐标系中标量函数 $u(r,\theta,\phi)$ 和矢量函数 $\mathbf{v}(r,\theta,\phi)$ 的梯度，散度，旋度和拉普拉斯算符的公式如下．其中 r 是极径，θ 是极角，ϕ 是方位角．

梯度算符

$$\boldsymbol{\nabla} u=\frac{\partial u}{\partial r}\hat{\mathbf{r}}+\frac{1}{r}\frac{\partial u}{\partial \theta}\hat{\boldsymbol{\theta}}+\frac{1}{r\sin\theta}\frac{\partial u}{\partial \phi}\hat{\boldsymbol{\phi}} \tag{1}$$

散度算符

$$\boldsymbol{\nabla}\cdot\mathbf{v}=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 v_r\right)+\frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}(\sin\theta v_\theta)+\frac{1}{r\sin\theta}\frac{\partial v_\phi}{\partial \phi} \tag{2}$$

旋度算符

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{v}=&\frac{1}{r\sin\theta}\left[\frac{\partial}{\partial\theta}(\sin\theta v_\phi)-\frac{\partial}{\partial\phi}\right]\hat{\mathbf{r}}+\frac{1}{r}\left[\frac{1}{\sin\theta}\frac{\partial v_r}{\partial\phi}-\frac{\partial}{\partial r}(rv_\phi)\right]\hat{\boldsymbol{\theta}}\\&+\frac{1}{r}\left[\frac{\partial}{\partial r}(rv_\theta)-\frac{\partial v_r}{\partial\theta}\right]\hat{\boldsymbol{\phi}}\end{aligned} \tag{3}$$

拉普拉斯算符

$$\nabla^2 u=\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}u)=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial u}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2 u}{\partial\phi^2} \tag{4}$$

拉普拉斯算符的分解

为了书写方便本书中定义两个算符 ∇_r^2 和 ∇_Ω^2 满足

$$\nabla^2=\nabla_r^2+\frac{\nabla_\Omega^2}{r^2} \tag{5}$$

其中 ∇_r^2 只对 r 求偏导，∇_Ω^2 只对 θ,ϕ 求偏导．∇_Ω^2 还可以进一步分解为

$$\nabla_\Omega^2=\boldsymbol{\nabla}_\Omega\cdot\boldsymbol{\nabla}_\Omega \tag{6}$$

其中（详见“球坐标系中的角动量算符[1393]”）

$$\boldsymbol{\nabla}_\Omega=-\left(\sin\phi\frac{\partial}{\partial\theta}+\cot\theta\cos\phi\frac{\partial}{\partial\phi}\right)\hat{\mathbf{x}}+\left(\cos\phi\frac{\partial}{\partial\theta}-\cot\theta\sin\phi\frac{\partial}{\partial\phi}\right)\hat{\mathbf{y}}+\frac{\partial}{\partial\phi}\hat{\mathbf{z}} \tag{7}$$

推导

预备知识 正交函数系[221]

球坐标系中，位置矢量可以表示为

$$\mathbf{r} = r\hat{\mathbf{r}} = r\sin\theta\cos\phi\,\hat{\mathbf{x}} + r\sin\theta\sin\phi\,\hat{\mathbf{y}} + r\cos\theta\hat{\mathbf{z}} \tag{8}$$

同样，球坐标系的三个单位矢量由三个坐标增加的方向确定

$$\begin{cases} \dfrac{\partial\mathbf{r}}{\partial r} = \sin\theta\cos\phi\,\hat{\mathbf{x}} + \sin\theta\sin\phi\,\hat{\mathbf{y}} + \cos\theta\,\hat{\mathbf{z}} \\ \dfrac{\partial\mathbf{r}}{\partial\theta} = r\cos\theta\cos\phi\,\hat{\mathbf{x}} + r\cos\theta\sin\phi\,\hat{\mathbf{y}} - r\sin\theta\,\hat{\mathbf{z}} \\ \dfrac{\partial\mathbf{r}}{\partial\phi} = -r\sin\theta\sin\phi\,\hat{\mathbf{x}} + r\sin\theta\cos\phi\,\hat{\mathbf{y}} \end{cases} \tag{9}$$

归一化得三个单位矢量为

$$\begin{cases} \hat{\mathbf{r}} = \sin\theta\cos\phi\,\hat{\mathbf{x}} + \sin\theta\sin\phi\,\hat{\mathbf{y}} + \cos\theta\,\hat{\mathbf{z}} \\ \hat{\boldsymbol{\theta}} = \cos\theta\cos\phi\,\hat{\mathbf{x}} + \cos\theta\sin\phi\,\hat{\mathbf{y}} - \sin\theta\,\hat{\mathbf{z}} \\ \hat{\boldsymbol{\phi}} = -\sin\phi\,\hat{\mathbf{x}} + \cos\phi\,\hat{\mathbf{y}} \end{cases} \tag{10}$$

不难验证这三个单位矢量两两间内积为零，即两两垂直，所以球坐标系也属于正交曲线坐标系．对比式 9 与式 10，有

$$\frac{\partial\mathbf{r}}{\partial r} = \hat{\mathbf{r}} \qquad \frac{\partial\mathbf{r}}{\partial\theta} = r\hat{\boldsymbol{\theta}} \qquad \frac{\partial\mathbf{r}}{\partial\phi} = r\sin\theta\,\hat{\boldsymbol{\phi}} \tag{11}$$

所以位置矢量的微分可以表示为

$$\mathrm{d}\mathbf{r} = \hat{\mathbf{r}}\,\mathrm{d}r + r\hat{\boldsymbol{\theta}}\,\mathrm{d}\theta + r\sin\theta\hat{\boldsymbol{\phi}}\,\mathrm{d}\phi \tag{12}$$

结合式 3 [288] 到式 8 [288] 即可完成推导．

球坐标系中的拉普拉斯方程

- 本词条存在未完成的内容．

预备知识 球坐标系中的拉普拉斯算符[742]，分离变量法解偏微分方程（链接未完成）

三维空间的拉普拉斯方程为

$$\nabla^2 f(\mathbf{r}) = 0 \tag{1}$$

球坐标系中，可以将拉普拉斯算子分解为径向和角向两部分（式 5 [742]）

$$\nabla^2 = \nabla_r^2 + \frac{\nabla_\Omega^2}{r^2} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial u}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 u}{\partial\phi^2} \tag{2}$$

则式 1 两边乘以 r^2 得

$$\left(r^2\nabla_r^2 - \nabla_\Omega^2\right) f = 0 \tag{3}$$

注意第一项只含有 r 的偏导，第二项只含有 θ,ϕ 的偏导．用分离变量法，令 $f(\mathbf{r}) = R(r)Y(\hat{\mathbf{r}})$（$\hat{\mathbf{r}}$ 是 θ,ϕ 的简写），则分离后的**径向方程**和**角向方程**分别为

$$r^2\nabla_r^2 R(r) = l(l+1)R(r) \tag{4}$$

$$\nabla_\Omega^2 Y(\hat{\mathbf{r}}) = -l(l+1)Y(\hat{\mathbf{r}}) \tag{5}$$

角向方程还可以进一步分离变量，令 $Y(\theta,\phi) = \Theta(\theta)\Phi(\phi)$，有

$$l(l+1)\sin^2\theta + \sin\theta\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Theta}{\partial\theta}\right)\Big/\Theta + \frac{\partial^2\Phi}{\partial\phi^2}\Big/\Phi = 0 \tag{6}$$

注意前二项只含 θ，第三项只含 ϕ．一般令前两项之和为常数 m^2，则第三项为 $-m^2$．这样我们就成功分离出了三个常微分方程，下面分别介绍．

径向方程是**欧拉型方程**（链接未完成）

$$r^2\frac{\mathrm{d}^2 R}{\mathrm{d}r^2} + 2r\frac{\mathrm{d}R}{\mathrm{d}r} - l(l+1)R = 0 \tag{7}$$

使用变量代换 $t = \ln r$ 解得

$$R(r) = C_1 r^l + \frac{C_2}{r^{l+1}} \tag{8}$$

关于 θ 的方程是

$$\sin^2\theta\frac{\mathrm{d}^2\Theta}{\mathrm{d}\theta^2} + \sin\theta\cos\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta} + [l(l+1)\sin^2\theta - m^2]\Theta = 0 \tag{9}$$

使用变量代换 $x=\cos\theta$ 得

$$(1-x^2)\frac{\mathrm{d}^2\Theta}{\mathrm{d}x^2}-2x\frac{\mathrm{d}\Theta}{\mathrm{d}x}+\left[l(l+1)-\frac{m^2}{1-x^2}\right]\Theta=0 \tag{10}$$

或

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[(1-x^2)\frac{\mathrm{d}\Theta}{\mathrm{d}x}\right]+\left[l(l+1)-\frac{m^2}{1-x^2}\right]\Theta=0 \tag{11}$$

该式被称为**连带勒让德方程**[755]，解为连带勒让德多项式 $P_l^m(x)=P_l^m(\cos\theta)$.

关于 ϕ 的方程是

$$\frac{\mathrm{d}^2\Phi}{\mathrm{d}\phi^2}=-m^2\Phi \tag{12}$$

该方程的解为 $\mathrm{e}^{\mathrm{i}m\phi}$. 原则上 m 可以取任意实数，但由于球坐标中的**循环边界条件**要求 $\Phi(\phi+2\pi)=\Phi(\phi)$，$m$ 只能取任意整数.

综上，球坐标中拉普拉斯方程的通解为

$$f(r,\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(C_{l,m}r^l+\frac{C'_{l,m}}{r^{l+1}}\right)P_l^m(\cos\theta)\mathrm{e}^{\mathrm{i}m\phi} \tag{13}$$

我们一般把在单位球面上归一化后的 $\Theta(\theta)\Phi(\phi)$ 称为**球谐函数**[761]，记为 $Y_{l,m}(\theta,\phi)$（满足角向方程式 5 ），则通解也可记为[1]

$$f(r,\theta,\phi)=\sum_{l,m}\left(A_{l,m}r^l+\frac{B_{l,m}}{r^{l+1}}\right)Y_{l,m}(\theta,\phi) \tag{14}$$

柱坐标系中的矢量算符

- 本词条处于草稿阶段.

梯度

$$\boldsymbol{\nabla}u=\frac{\partial u}{\partial r}\hat{\mathbf{r}}+\frac{1}{r}\frac{\partial u}{\partial\theta}\hat{\boldsymbol{\theta}}+\frac{\partial u}{\partial z}\hat{\mathbf{z}} \tag{1}$$

散度

$$\boldsymbol{\nabla}\cdot\mathbf{v}=\frac{1}{r}\frac{\partial}{\partial r}(rv_r)+\frac{1}{r}\frac{\partial v_\theta}{\partial\theta}+\frac{\partial v_z}{\partial z} \tag{2}$$

[1]以后为了方便，我们常在求和符号中省略 l,m 的范围.

旋度

$$\nabla \times \mathbf{v} = \left(\frac{1}{r}\frac{\partial v_z}{\partial \theta} - \frac{\partial v_\theta}{\partial z}\right)\hat{\mathbf{r}} + \left(\frac{\partial v_r}{\partial z} - \frac{\partial v_z}{\partial r}\right)\hat{\boldsymbol{\theta}} + \frac{1}{r}\left[\frac{\partial}{\partial r}(rv_\theta) - \frac{\partial v_r}{\partial \theta}\right]\hat{\mathbf{z}} \tag{3}$$

拉普拉斯算子

$$\nabla^2 u = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{\partial^2 u}{\partial z^2} \tag{4}$$

柱坐标系中的拉普拉斯方程

预备知识 分离变量法简介，拉普拉斯方程（直角坐标），柱坐标的拉普拉斯算符[746]

拉普拉斯方程为

$$\nabla^2 u = 0 \tag{1}$$

拉普拉斯在柱坐标中分离变量后，径向方程为贝塞尔方程（式 1 [757]）

$$x\frac{\mathrm{d}}{\mathrm{d}x}\left(x\frac{\mathrm{d}y}{\mathrm{d}x}\right) + (x^2 - m^2)y = 0 \tag{2}$$

其中 $x = lr$．下面我们来推导．

分离变量法

柱坐标系中的拉普拉斯方程为

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{\partial^2 u}{\partial z^2} = 0 \tag{3}$$

令 $u = R(r)\Phi(\theta)Z(z)$，代入方程得

$$\frac{1}{rR}\frac{\partial}{\partial r}\left(r\frac{\partial R}{\partial r}\right) + \frac{1}{r^2\Phi}\frac{\partial^2 \Phi}{\partial \theta^2} + \frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} = 0 \tag{4}$$

前两项只是 r 和 θ 的函数，第三项只是 z 的函数，所以它们分别为常数．令

$$\frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} = l^2 \tag{5}$$

则前两项为

$$\frac{1}{rR}\frac{\partial}{\partial r}\left(r\frac{\partial R}{\partial r}\right)+\frac{1}{r^2\Phi}\frac{\partial^2\Phi}{\partial\theta^2}=-l^2 \tag{6}$$

为了继续分离 r 和 θ，两边乘以 r^2，则左边第二项只是关于 θ 的函数，剩下的部分只是关于 r 的函数．令

$$\frac{1}{\Phi}\frac{\mathrm{d}^2\Phi}{\mathrm{d}\theta^2}=-m^2 \tag{7}$$

则剩下的部分为 m^2，即

$$r\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}R}{\mathrm{d}r}\right)+(l^2r^2-m^2)R=0 \tag{8}$$

令 $x=lr$，$y(x)=R(r)$ 则

$$x\frac{\mathrm{d}}{\mathrm{d}x}\left(x\frac{\mathrm{d}y}{\mathrm{d}x}\right)+(x^2-m^2)y=0 \tag{9}$$

到此为止，三个变量已经完全分离，各自的微分方程为式 5，式 7，式 8．

$Z(z)$ 的通解为 $C_1\mathrm{e}^{lz}+C_2\mathrm{e}^{-lz}$，$\Phi(\theta)$ 的通解为 $\mathrm{e}^{\mathrm{i}m\theta}$．式 8 的解不能用有限的初等函数表示，式 9 为贝塞尔方程的标准形式．

需要注意的是，贝塞尔函数的阶数 m 是角向方程 $(\mathrm{d}^2\Phi/\mathrm{d}\theta^2)/\Phi=-m^2$ 的参数，而不是径向方程的参数 l．参数 l 被包含在自变量 x 中．

泊松方程

预备知识 拉普拉斯方程[742]

[2]三维空间的**泊松方程（Poisson's equation）**可以记为

$$\Delta u=f(x,y,z) \tag{1}$$

泊松方程可以看做是非齐次的拉普拉斯方程．

泊松方程与时间无关，显然不适用冲量定理法．我们可以采用特解法．先不管边界条件，任取这泊松方程的一个特解 v，然后令 $u=v+w$．这就把问题转化为求解 w，而 $\Delta w=\Delta u-\Delta v=\Delta u-f=0$，这不再是泊松方程而是拉普拉斯方程．在一定边界条件下求解拉普拉斯方程是我们之前研究过的问题．

[2]参考 Wikipedia 相关页面．

例 1

在圆域 $\rho<\rho_0$ 上求解泊松方程的边值问题：

$$\begin{cases}\Delta u=a+b\left(x^2-y^2\right)\\ u|_{\rho=\rho_0}=c\end{cases}\tag{2}$$

先设法找泊松方程的一个特解．显然 $\Delta(ax^2/2)=a,\Delta(ay^2/2)=a$，为了对称起见，取 $a(x^2+y^2)/4$．又 $\Delta(bx^4/12)=bx^2,\Delta(by^4/12)=by^2$．这样，找到一个特解

$$\begin{aligned}&=\tfrac{a}{4}\left(x^2+y^2\right)+\tfrac{b}{12}\left(x^4-y^4\right)=\tfrac{a}{4}\rho^2+\tfrac{b}{12}\left(x^2+y^2\right)\left(x^2-y^2\right)\\&=\tfrac{a}{4}\rho^2+\tfrac{b}{12}\rho^4\cos 2\varphi\end{aligned}\tag{3}$$

令

$$u=v+w=\frac{a}{4}\rho^2+\frac{b}{12}\rho^4\cos 2\varphi+w\tag{4}$$

就把问题转化为 w 的定解问题：

$$\begin{cases}\Delta w=0\\ w|_{\rho=\rho_0}=c-\tfrac{a}{4}\rho_0^2-\tfrac{b}{12}\rho_0^4\cos 2\varphi\end{cases}\tag{5}$$

在极坐标系中用分离变数法求解拉普拉斯方程，即

$$\begin{aligned}w(\rho,\varphi)=&C_0+D_0\ln\rho+\sum_{m=1}^{\infty}\rho^m\left(A_m\cos m\varphi+B_m\sin m\varphi\right)\\&+\sum_{i}^{\infty}\rho^{-m}\left(C_m\cos m\varphi+D_m\sin m\varphi\right)\end{aligned}\tag{6}$$

w 在圆内应当处处有限．但上式的 $\ln\rho$ 和 ρ^{-m} 在圆心为无限大，所以应当排除，就是说 $D_0=0,C_m=0,D_m=0$．于是，

$$w(\rho,\varphi)=\sum_{m=0}^{\infty}\rho^m\left(A_m\cos m\varphi+B_m\sin^{+}m\varphi\right)\tag{7}$$

把上式代入边界条件，

$$\sum_{m=0}^{\infty}\rho_0^m\left(A_m\cos m\varphi+B_m\sin m\varphi\right)=c-\frac{a}{4}\rho_0^2-\frac{b}{12}\rho_0^4\cos 2\varphi\tag{8}$$

比较两边系数，得

$$A_0 = c - \frac{a}{4}\rho_0^2, A_2 = -\frac{b}{12}\rho_0^2, A_m = 0(m \neq 0, 2); B_m = 0 \tag{9}$$

这样，所求解是

$$u = v + w = c + \frac{a}{4}\left(\rho^2 - \rho_0^2\right) + \frac{b}{12}\rho^2\left(\rho^2 - \rho_0^2\right)\cos 2\varphi \tag{10}$$

例 2

在矩形域 $0 \leq x \leq a, 0 \leq y \leq b$ 上求解泊松方程的边值问题

$$\Delta_2 u = -2 \tag{11}$$

$$u|_{x=0} = 0, u|_{x=a} = 0 \tag{12}$$

$$u|_{y=0} = 0, u|_{y=b} = 0 \tag{13}$$

先找泊松方程的一个特解 v，显然，$v = -x^2$ 满足 $\Delta v = -2$. 其实，$v = -x^2 + c_1 x + c_2$（c_1 和 c_2 是两个积分常数）也满足 $\Delta v = -2$. 我们打算选择适当的 c_1 和 c_2，使 v 满足齐次边界条件式 12. 容易看出，$c_1 = a, c_2 = 0$. 这样，

$$v(x, y) = x(a - x) \tag{14}$$

令

$$u(x, y) = v + w = x(a - x) + w(x, y) \tag{15}$$

把上式代入 u 的定解问题，就把它转化为 w 的定解问题

$$\Delta w = 0 \tag{16}$$

$$w|_{x=0} = 0, w|_{x=a} = 0 \tag{17}$$

$$w|_{y=0} = x(x - a), w|_{y=b} = x(x - a) \tag{18}$$

显然，满足式 16 和式 17 的解可表为

$$w(x, y) = \sum_{n=1}^{\infty}\left(A_n e^{\frac{n\pi y}{a}} + B_n e^{-\frac{n\pi y}{a}}\right)\sin\frac{n\pi x}{a} \tag{19}$$

为确定系数 A_n 和 B_n，以式 19 代入边界条件式 18，有

$$\begin{array}{l} \sum_{n=1}^{\infty}\left(A_n + B_n\right)\sin\frac{n\pi x}{a} = x(x - a) \\ \sum_{n=1}^{\infty}\left(A_n \mathrm{e}^{\frac{n\pi b}{a}} + B_n \mathrm{e}^{-\frac{n\pi b}{a}}\right)\sin\frac{n\pi x}{a} = x(x - a) \end{array} \tag{20}$$

把式 20 的右边也展为傅里叶正弦级数：

$$x(x-a)=\sum_{n=1}^{\infty}C_n\sin\frac{n\pi x}{a} \tag{21}$$

其中

$$C_n=\frac{2}{a}\int_0^a\left(x^2-ax\right)\sin\frac{n\pi x}{a}\mathrm{d}x=\frac{4a^2}{n^3\pi^3}\left[(-1)^n-1\right] \tag{22}$$

以式 21 代入式 20 的两边，比较两边的傅里叶系数，

$$\begin{aligned}A_n+B_n&=C_n\\ A_n\mathrm{e}^{\frac{n\pi b}{a}}+B_n\mathrm{e}^{-\frac{n\pi b}{a}}&=C_n\end{aligned} \tag{23}$$

由此可得

$$\begin{aligned}A_n&=\frac{1-\mathrm{e}^{-n\pi b/a}}{\mathrm{e}^{n\pi b/a}-\mathrm{e}^{-n\pi b/a}}C_n=\frac{\mathrm{e}^{-n\pi b/2a}\left(\mathrm{e}^{n\pi b/2a}-\mathrm{e}^{-n\pi b/2a}\right)}{\mathrm{e}^{n\pi b/a}-\mathrm{e}^{-n\pi b/a}}\\ &=\frac{\mathrm{e}^{-n\pi b/2a}}{\mathrm{e}^{n\pi b/2a}+\mathrm{e}^{-n\pi b/2a}}C_n=\frac{\mathrm{e}^{-n\pi b/2a}}{\cosh\left(n\pi b/2a\right)}C_n\\ B_n&=\frac{\mathrm{e}^{n\pi b/a}-1}{\mathrm{e}^{n\pi b/a}-\mathrm{e}^{-n\pi b/a}}C_n=\frac{\mathrm{e}^{n\pi b/2a}\left(\mathrm{e}^{n\pi b/2a}-\mathrm{e}^{-n\pi b/2a}\right)}{\mathrm{e}^{n\pi b/a}-\mathrm{e}^{-n\pi b/a}}\\ &=\frac{\mathrm{e}^{n\pi b/2a}}{\mathrm{e}^{n\pi b/2a}+\mathrm{e}^{-n\pi b/2a}}C_n=\frac{\mathrm{e}^{n\pi b/2a}}{\cosh\left(n\pi b/2a\right)}C_n\end{aligned} \tag{24}$$

于是代回式 19 成为

$$w(x,y)=\sum_{n=1}^{\infty}\frac{\cosh[n\pi(y-b/2)/a]}{\cosh(n\pi b/2a)}C_n\sin\frac{n\pi x}{a} \tag{25}$$

我们又知道，对于 $n=2k(k=1,2,\cdots)$，$C_n=0$．对于 $n=2k-1(k=1,2,\cdots)$，$C_n=-8a^2/(2k-1)^3\pi^3$．这样，

$$w(x,y)=-\frac{8a^2}{\pi^3}\sum_{k=1}^{\infty}\frac{\cosh[(2k-1)\pi(y-b/2)/a]}{(2k-1)^3\cosh[(2k-1)\pi b/2a]}\sin\frac{(2k-1)\pi x}{a} \tag{26}$$

把 $w(x,y)$ 加上 $x(x-a)$ 就是所求的 $u(x,y)$.

球坐标系中的亥姆霍兹方程

预备知识 球坐标系中的拉普拉斯方程[744]

亥姆霍兹方程为[3]

$$(\nabla^2 + k^2)f = 0 \tag{1}$$

球坐标系中的求解过程与拉普拉斯方程[744]中的过程类似．将拉普拉斯算子分解为径向和角向两部分（式 5 [742]）

$$\nabla^2 = \nabla_r^2 + \frac{\nabla_\Omega^2}{r^2} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial u}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 u}{\partial\phi^2} \tag{2}$$

则式 1 乘 r^2 得

$$\left(r^2\nabla_r^2 + k^2r^2 - \nabla_\Omega^2\right)f = 0 \tag{3}$$

注意前两项只含有 r 的偏导，第三项只含有 θ,ϕ 的偏导．用分离变量法，令 $f(\mathbf{r}) = R(r)Y(\hat{\mathbf{r}})$，则分离后的角向方程和径向方程分别为

$$\nabla_\Omega^2 Y(\hat{\mathbf{r}}) = -l(l+1)Y(\hat{\mathbf{r}}) \tag{4}$$

$$\left[r^2\nabla_r^2 + k^2r^2 - l(l+1)\right]R(r) = 0 \tag{5}$$

我们已知角向的解为球谐函数[761] $Y_{l,m}(\hat{\mathbf{r}})$.而径向方程比普拉斯方程中的（式 4 [744]）多出了一项 k^2r^2，使用变量代换 $\rho = kr$ 得

$$\rho^2\frac{\partial^2 R}{\partial\rho^2} + 2\rho\frac{\partial R}{\partial\rho} + [\rho^2 - l(l+1)]R = 0 \tag{6}$$

该方程被称为**球贝塞尔方程**[759]，两个线性无关解分别是第一和第二类**球贝塞尔函数** $j_l(\rho)$ 和 $y_l(\rho)$.

综上，方程的通解为

$$f(\mathbf{r}) = \sum_{l,m}[A_l j_l(kr) + B_l y_l(kr)]Y_{l,m}(\hat{\mathbf{r}}) \tag{7}$$

柱坐标中的亥姆霍兹方程

[3] 一般地，k 可以是复数，所以 k^2 也可以是负实数.

- 本词条处于草稿阶段.

预备知识 亥姆霍兹方程，柱坐标系中的拉普拉斯方程[747]

亥姆霍兹方程为

$$\nabla^2 u = -k^2 u \tag{1}$$

将柱坐标系中的拉普拉斯方程右边加上一项得到亥姆霍兹方程.

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}+\frac{\partial^2 u}{\partial z^2}=-k^2 u \tag{2}$$

使用分离变量法，令 $u=R(r)\Phi(\theta)Z(z)$，代入方程得

$$\frac{1}{rR}\frac{\partial}{\partial r}\left(r\frac{\partial R}{\partial r}\right)+\frac{1}{r^2\Phi}\frac{\partial^2 \Phi}{\partial \theta^2}+\frac{1}{Z}\frac{\partial^2 Z}{\partial z^2}=-k^2 \tag{3}$$

$\Theta(\theta)$ 和 $Z(z)$ 的常微分方程和解与拉普拉斯方程中的相同

$$\frac{1}{Z}\frac{\partial^2 Z}{\partial z^2}=l^2 \qquad \frac{1}{\Phi}\frac{\mathrm{d}^2\Phi}{\mathrm{d}\theta^2}=-m^2 \tag{4}$$

径向方程变为

$$r\frac{\partial}{\partial r}\left(r\frac{\partial R}{\partial r}\right)+[(l^2+k^2)r^2-m^2]R=0 \tag{5}$$

令 $x=\sqrt{l^2+k^2}r$ 以及 $u(x)=R(r)$，做变量替换得

$$x\frac{\mathrm{d}}{\mathrm{d}x}\left(x\frac{\mathrm{d}y}{\mathrm{d}x}\right)+(x^2-m^2)y=0 \tag{6}$$

与拉普拉斯方程的情况（式 9 [747]）一样，我们得到了贝塞尔方程（式 1 [757]）. 但注意 x,r 之间的缩放比例取也取决于 k^2.

勒让德多项式

勒让德方程为

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[(1-x^2)\frac{\mathrm{d}}{\mathrm{d}x}P_l(x)\right]+l(l-1)P_l(x)=0 \tag{1}$$

我们仅在区间 $x\in[-1,1]$ 上考虑 l 为非负整数的情况．方程的解 $P_l(x)$ 是关于 x 的 l 阶多项式

$$P_l(x)=\frac{1}{2^l}\sum_{s=0}^{[l/2]}\frac{(-1)^s(2l-2s)!}{s!(l-s)!(l-2s)!}x^{l-2s} \tag{2}$$

其中 $[x]$ 是向下取整函数．当 x 是整数，$[x]=x$，当 x 是非整数，$[x]$ 是小于 x 的最大整数．

这里列出前几个多项式

$$\begin{aligned} P_0(x)&=1 & P_3(x)&=\frac{1}{2}(5x^3-3x)\\ P_1(x)&=x & P_4(x)&=\frac{1}{8}(35x^4-30x^2+3)\\ P_2(x)&=\frac{1}{2}(3x^2-1) & P_5(x)&=\frac{1}{8}(63x^5-70x^3+15x) \end{aligned} \tag{3}$$

勒让德多项式也可以用**罗德里格斯**公式表示

$$P_l(x)=\frac{1}{2^l l!}\frac{\mathrm{d}^l}{\mathrm{d}x^l}\left(x^2-1\right)^l \tag{4}$$

由于求导会改变函数的奇偶性，由上式可以证明当 l 为偶（奇）数时 $P_l(x)$ 是偶（奇）函数，所以只有 x 的偶（奇）次方项．

正交归一性质

归一化系数为

$$A_l=\sqrt{\frac{2l+1}{2}} \tag{5}$$

满足正交归一化条件

$$\int_{-1}^{1}A_{l'}P_{l'}(x)\cdot A_lP_l(x)\,\mathrm{d}x=\delta_{l,l'} \tag{6}$$

其他性质

$$P_l(1)=1 \qquad P_l(-1)=(-1)^l \tag{7}$$

勒让德方程的级数解

令

$$P_l(x)=\sum_{n=0}^{\infty}c_nx^n \tag{8}$$

代入方程，对比系数得到递推公式

$$c_{n+2}=\frac{n(n+1)-l(l+1)}{(n+2)(n+1)}c_n=\frac{(n-l)(n+1+l)}{(n+2)(n+1)}c_n \tag{9}$$

可见偶数项系数可用 c_0 表示，奇数项系数可用 c_1 表示. 所以 c_0 和 c_1 可以看做是二阶微分方程的两个任意常数.

当 l 为整数时，可以证明 $n>l$ 以上的系数都为 0，令最高次项系数为

$$c_l=\frac{(2l)!}{2^l(l!)^2} \tag{10}$$

就得到了式 2.

生成函数

勒让德多项式可以表示为以下函数对 r 的泰勒展开的系数

$$\frac{1}{\sqrt{1+r^2-2rx}}=\sum_{l=0}^{\infty}P_l(x)r^l \tag{11}$$

其中 $1/\sqrt{1+r^2-2rx}$ 叫做勒让德多项式的生成函数或母函数

连带勒让德多项式

预备知识 勒让德多项式[753]

[4]连带勒让德方程来源于在球坐标系中使用分离变量法解拉普拉斯方程（球坐标系中的拉普拉斯方程）中，连带勒让德方程为

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[(1-x^2)\frac{\mathrm{d}}{\mathrm{d}x}P_l^m(x)\right]+\left[l(l-1)-\frac{m^2}{1-x^2}\right]P_l^m(x)=0 \tag{1}$$

[4]参考 Wikipedia 相关页面.

当 $m=0$ 时，该方式变为勒让德方程（式 1 [753]）.

我们只讨 l 为非负整数，m 为绝对值不大于 l 的整数的情况，且只讨论区间 $x\in[-1,1]$. 此时连带勒让德函数的解 $P_l^m(x)$ 可用勒让德多项式 $P_l(x)$ 生成

$$P_l^m(x)=(-1)^m(1-x^2)^{m/2}\frac{\mathrm{d}^m}{\mathrm{d}x^m}P_l(x) \tag{2}$$

注意当 $m=0$ 时有 $P_l^m(x)=P_l(x)$. 当 m 为奇数时，$P_l^m(x)$ 并不是一个多项式，但习惯上我们称之为勒让德多项式. 式中 $(-1)^m$ 叫做 **Shortley-Condon 相位**，量子力学中通常会使用该相位，但在一些文献中不包含这个相位（同样满足微分方程）.

归一化

勒让德多项式的归一化系数为（证明略）

$$A_{l,m}=\sqrt{\frac{2l+1}{2}\frac{(l-m)!}{(l+m)!}} \tag{3}$$

满足正交归一条件

$$\int_{-1}^{1}[A_{l',m}P_{l'}^m(x)][A_{l,m}P_l^m(x)]\,\mathrm{d}x=\delta_{l,l'} \tag{4}$$

注意只有同一个 m 的多项式之间正交.

其他性质

- 当 $m=0$ 时，见勒让德多项式[753] 的性质.
- **端点值**：对任意 l,m 有

$$P_l^m(1)=\delta_{m,0}\qquad P_l^m(-1)=(-1)^l\delta_{m,0} \tag{5}$$

- m 取相反数时，对称或反对称关系为

$$A_{l,-m}P_l^{-m}(x)=(-1)^mA_{l,m}P_l^m(x) \tag{6}$$

或者

$$P_l^{-m}(x)=(-1)^m\frac{(l-m)!}{(l+m)!}P_l^m(x) \tag{7}$$

贝塞尔函数

预备知识 常微分方程[211]

贝塞尔微分方程（Bessel's differential equation）通常在使用分离变量法求解柱坐标中的拉普拉斯方程[747]时得到：

$$x\frac{\mathrm{d}}{\mathrm{d}x}\left(x\frac{\mathrm{d}y}{\mathrm{d}x}\right)+(x^2-\alpha^2)y=0 \tag{1}$$

其中 α 叫做**阶数（order）**. 两个线性无关的解分别是**第一类贝塞尔函数（Bessel function of the first kind）** $J_\alpha(x)$ 和**第二类贝塞尔函数（Bessel function of the second kind）** $Y_\alpha(x)$. 这里只讨论 $x>0$ 且 α 为整数或半整数的情况.

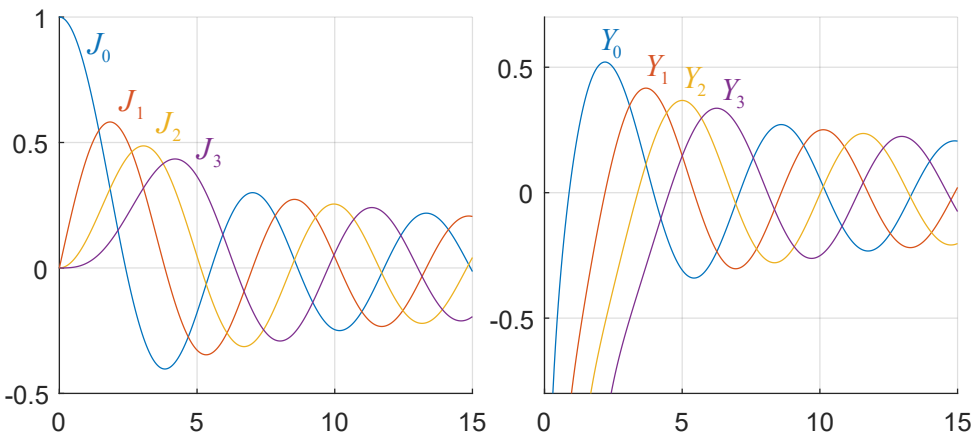


图 1: 第一类和第二类贝塞尔函数（非负整数阶）

$J_\alpha(x)$ 的级数形式为

$$J_\alpha(x)=\sum_{m=0}^{\infty}\frac{(-1)^m}{m!\Gamma(m+\alpha+1)}\left(\frac{x}{2}\right)^{2m+\alpha} \tag{2}$$

其中使用了 Γ 函数[254]. $Y_\alpha(x)$ 可以通过 $J_\alpha(x)$ 来定义

$$Y_\alpha(x)=\frac{J_\alpha(x)\cos\alpha\pi-J_{-\alpha}(x)}{\sin\alpha\pi} \tag{3}$$

容易证明

$$J_{1/2}(x)=\sqrt{\frac{2}{\pi x}}\sin x=-Y_{-1/2}(x) \tag{4}$$

$$J_{-1/2}(x)=\sqrt{\frac{2}{\pi x}}\cos x=-Y_{1/2}(x) \tag{5}$$

变量替换

令 $u(x)=\sqrt{x}y(x)$，对式 1 替换变量可得更简洁的方程，使一阶导数项消失

$$x^2\frac{\mathrm{d}^2u}{\mathrm{d}x^2}+\left(x^2-\alpha^2+\frac{1}{4}\right)u=0 \tag{6}$$

显然 $u(x)$ 的两组解就是 $\sqrt{x}J_\alpha(x)$ 和 $\sqrt{x}Y_\alpha(x)$.

汉克尔函数

贝塞尔方程的两个线性无关解也可以用第一类和第二类汉克尔函数来表示[5]

$$H_\alpha^{(1)}(x)=J_\alpha(x)+\mathrm{i}Y_\alpha(x)\qquad H_\alpha^{(2)}(x)=J_\alpha(x)-\mathrm{i}Y_\alpha(x) \tag{7}$$

常用性质

令 Z 为 $J,Y,H^{(1)},H^{(2)}$ 的任意一种，则

$$Z_{-\alpha}(x)=(-1)^\alpha Z_\alpha(x) \tag{8}$$

递推关系

$$\frac{2\alpha}{x}Z_\alpha(x)=Z_{\alpha-1}(x)+Z_{\alpha+1}(x) \tag{9}$$

一阶导数

$$\frac{\mathrm{d}Z_\alpha}{\mathrm{d}x}=\frac{1}{2}[Z_{\alpha-1}(x)-Z_{\alpha+1}(x)] \tag{10}$$

渐进形式（$x\gg1$）

$$J_\alpha(x)\to\sqrt{\frac{2}{\pi x}}\cos\left(x-\frac{\alpha\pi}{2}-\frac{\pi}{4}\right) \tag{11}$$

$$Y_\alpha(x)\to\sqrt{\frac{2}{\pi x}}\sin\left(x-\frac{\alpha\pi}{2}-\frac{\pi}{4}\right) \tag{12}$$

正交关系

$$\int_0^1 J_\alpha(u_{\alpha,m}x)J_\alpha(u_{\alpha,n}x)x\,\mathrm{d}x=\frac{\delta_{m,n}}{2}[J_{\alpha+1}(u_{\alpha,m})]^2 \tag{13}$$

[5]可类比欧拉公式 $\exp(\pm\mathrm{i}x)=\cos x\pm\mathrm{i}\sin x$ （式 3 [351]）.

其中 $u_{\alpha,m}$ 是 $J_\alpha(x)$ 的第 m 个根.

连续正交关系（容易从渐进形式得到）

$$\int_0^\infty \sqrt{k_1}J_\alpha(k_1x)\sqrt{k_2}J_\alpha(k_2x)x\,\mathrm{d}x=\delta(k_2-k_1) \tag{14}$$

修正贝塞尔函数

修正贝塞尔方程为

$$x^2\frac{\mathrm{d}^2y}{\mathrm{d}x^2}+x\frac{\mathrm{d}y}{\mathrm{d}x}-(x^2+\alpha^2)y=0 \tag{15}$$

其解为两个**修正贝塞尔函数（Modified Bessel Function）**，第一类为 $I_\alpha(x)$，第二类为 $K_\alpha(x)$，与贝塞尔函数的关系为

$$I_\alpha(x)=\mathrm{i}^{-\alpha}J_\alpha(\mathrm{i}x)\qquad K_\alpha(x)=\frac{\pi}{2}\mathrm{i}^{\alpha+1}H_\alpha^{(1)}(\mathrm{i}x) \tag{16}$$

球贝塞尔函数

预备知识 贝塞尔函数[757]

[6]**球贝塞尔方程（spherical Bessel's equation）**为

$$x^2\frac{\mathrm{d}^2y}{\mathrm{d}x^2}+2x\frac{\mathrm{d}y}{\mathrm{d}x}+[x^2-n(n+1)]y=0 \tag{1}$$

两个线性无关的解分别为第一类球贝塞尔函数 $j_n(x)$ 和第二类球贝塞尔函数 $y_n(x)$，见图 1. 它们可以通过贝塞尔函数[757] J，Y 来定义

$$j_n(x)=\sqrt{\frac{\pi}{2x}}J_{n+1/2}(x)\qquad y_n(x)=\sqrt{\frac{\pi}{2x}}Y_{n+1/2}(x) \tag{2}$$

同样也可以定义两类**球汉克尔函数（spherical Hankel's function）**

$$h_n^{(1)}(x)=\sqrt{\frac{\pi}{2x}}H_{n+1/2}^{(1)}(x)=j_n(x)+\mathrm{i}y_n(x) \tag{3}$$

[6]参考 Wikipedia 相关页面.

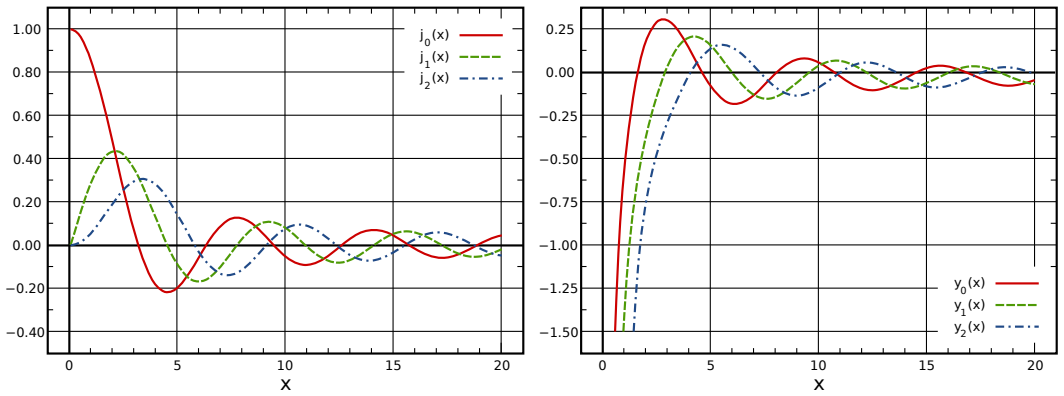


图 1: 球贝塞尔函数和球汉克尔函数（来自 Wikipedia）

$$h_n^{(2)}(x) = \sqrt{\frac{\pi}{2x}} H_{n+1/2}^{(2)}(x) = j_n(x) - \mathrm{i}y_n(x) \tag{4}$$

另一种等效的定义方式使用 Rayleigh's 方程

$$j_n(x) = (-x)^n \left(\frac{1}{x}\frac{\mathrm{d}}{\mathrm{d}x}\right)^n \frac{\sin x}{x} \tag{5}$$

$$y_n(x) = -(-x)^n \left(\frac{1}{x}\frac{\mathrm{d}}{\mathrm{d}x}\right)^n \frac{\cos x}{x} \tag{6}$$

性质

一阶导数（f 是 $j, y, h^{(1)}, h^{(2)}$ 中的任意一种）

$$f_n'(z) = f_{n-1}(z) - \frac{n+1}{z} f_n(z) \tag{7}$$

渐进形式当 $x \gg 1$ 时，球贝塞尔函数的渐进表达式为

$$j_l(x) \to \sin(x - l\pi/2)/x \qquad y_l(x) \to -\cos(x - l\pi/2)/x \tag{8}$$

$$h_l^{(1)}(x) \to (-\mathrm{i})^{l+1} \mathrm{e}^{\mathrm{i}x}/x \qquad h_l^{(2)}(x) \to \mathrm{i}^{l+1} \mathrm{e}^{-\mathrm{i}x}/x \tag{9}$$

由渐进形式可得径向归一化积分为（以 j_l 为例，推导见式 18 [230]）

$$\begin{aligned} \int_0^\infty k' j_l(k'r) \cdot k j_l(kr) r^2 \,\mathrm{d}r &= \int_0^\infty \sin(k'r - l\pi/2) \sin(kr - l\pi/2) \,\mathrm{d}r \\ &= \frac{\pi}{2}[\delta(k' - k) - \delta(k' + k)] \end{aligned} \tag{10}$$

修正球贝塞尔函数

修正球贝塞尔方程为

$$x^2\frac{\mathrm{d}^2y}{\mathrm{d}x^2}+2x\frac{\mathrm{d}y}{\mathrm{d}x}-[x^2+n(n+1)]y=0 \tag{11}$$

两个线性无关解称为**第一类修正球贝塞尔函数（modified spherical Bessel function of the first kind）**和**第二类修正球贝塞尔函数**

$$i_n(x)=\sqrt{\frac{\pi}{2x}}I_{n+1/2}(x)=\mathrm{i}^{-n}j_n(\mathrm{i}x) \tag{12}$$

$$k_n(x)=\sqrt{\frac{2}{\pi x}}K_{n+1/2}(x)=\frac{\pi}{2}\mathrm{i}^{n+2}h_n^{(1)}(\mathrm{i}x) \tag{13}$$

渐进形式为

$$i_n(x)\to\frac{\mathrm{e}^x}{2x}\qquad k_n(x)\to\frac{\pi}{2}\frac{\mathrm{e}^{-x}}{x} \tag{14}$$

球谐函数

预备知识 球坐标的拉普拉斯方程[744]，连带勒让德多项式[755]

当球坐标中的拉普拉斯方程（球坐标系中的拉普拉斯方程[744]）分离变量后，关于极角 θ 的函数为连带勒让德多项式 $P_l^m(\cos\theta)$，方向角函数为 $\mathrm{e}^{\mathrm{i}m\phi}$．我们定义**球谐函数**为[7][8]

$$Y_{l,m}(\hat{\mathbf{r}})=Y_{l,m}(\theta,\phi)=A_{l,m}P_l^m(\cos\theta)\mathrm{e}^{\mathrm{i}m\phi} \tag{1}$$

其中 l,m 为整数，$l\geq 0$，$-l\leq m\leq -l$．$A_{l,m}$ 是归一化系数，使得 $|Y_{l,m}(\hat{\mathbf{r}})|^2$ 在单位球面上的面积分等于 1．

$$A_{l,m}=\sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} \tag{2}$$

球谐函数可以看作是将单位球面上的每一点（或者三维空间中的每个方向）映射到一个复数函数值．常见的球谐函数见“球谐函数列表[765]”，函数图常用图 1 的方式绘制．

[7]有些教材也将球谐函数记为 Y_l^m

[8]特殊地，$Y_{l,0}(\hat{\mathbf{r}})=\sqrt{(2l+1)/(4\pi)}P_l(\cos\theta)$，$P_l$ 是勒让德多项式[753]．

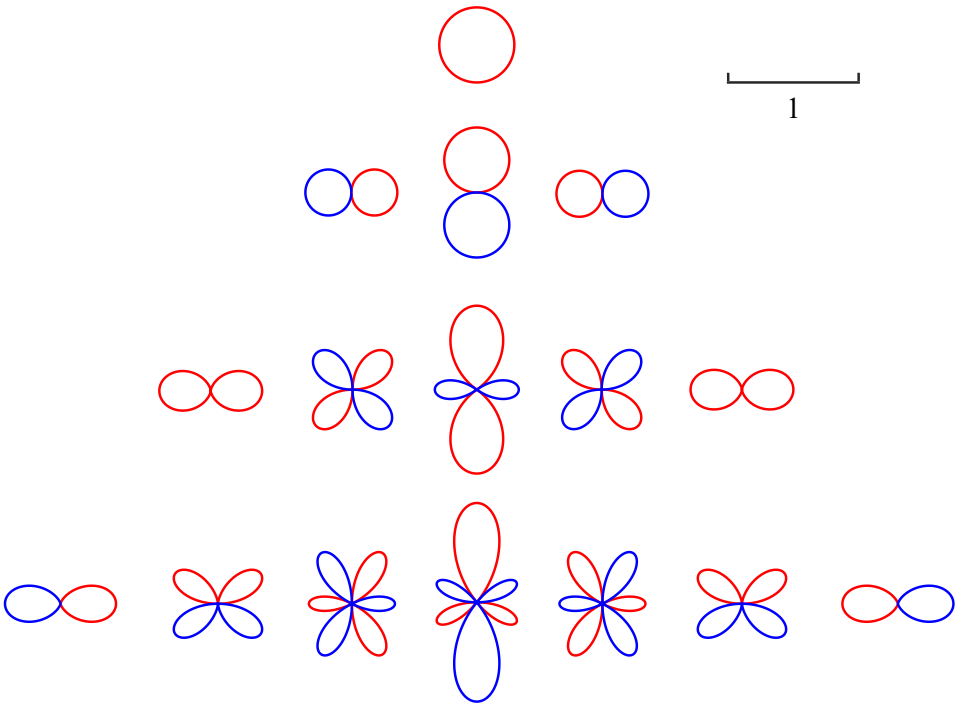


图 1: $r(\theta) = |Y_{l,m}(\theta, 0)|$ 的极坐标曲线，注意 $Y_{l,m}(\theta, 0)$ 是实数．红色代表 $Y_{l,m}(\theta, 0) > 0$，蓝色代表 $Y_{l,m}(\theta, 0) < 0$．第 1 行到第 4 行分别为 $l = 0$ 到 3，每行从左到右分别为 $m = -l$ 到 l．图中的右上角标明了 r 的单位长度．这里的球谐函数使用了 Condon–Shortley 相位（见下文）.

偏微分方程

球谐函数同时满足偏微分方程

$$\nabla^2_\Omega Y_{lm} = -l(l+1)Y_{lm} \tag{3}$$

和

$$\frac{\partial}{\partial \phi} Y_{lm} = \mathrm{i} m Y_{lm} \tag{4}$$

其中

$$\nabla^2_\Omega Y_{lm} = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\left(\frac{\partial^2}{\partial\phi^2}\right) \tag{5}$$

中括号中的算符是球坐标系拉普拉斯算子（式 4 [742]）∇^2 中与 r 无关的部分.

归一化系数

由球谐函数的归一化条件，

$$\begin{aligned} 1 &= \int |Y_{l,m}(\hat{\mathbf{r}})|^2 \,\mathrm{d}\Omega = \int_0^{\pi}\int_0^{2\pi} |Y_{l,m}(\theta,\phi)|^2 \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\phi \\ &= |A_{l,m}|^2 \int_{-1}^{1} |P_l^m(\cos\theta)|^2 \,\mathrm{d}(\cos\theta) \int_0^{2\pi} \left|\mathrm{e}^{\mathrm{i}m\phi}\right|^2 \mathrm{d}\phi \\ &= \frac{2\pi}{|A'_{l,m}|^2} |A_{l,m}|^2 \end{aligned} \tag{6}$$

其中 $A'_{l,m}$ 是 $P_l^m(x)$ 的归一化系数（见式 3 [755]），代入后可得 $A_{l,m}$.

Condon-Shortley 相位

与连带勒让德多项式相同，在定义球谐函数时我们也可以选择是否包含 **Condon-Shortley 相位** $(-1)^m$（物理中一般选择包含）. 如果包含，我们可以选择将其包含在连带勒让德多项式中（如式 2 [755]），或者包含在球谐函数的定义中. 如果不包含，该相位在两个定义中都不出现.

正交归一性

由勒让德函数的正交归一性（式 4 [755]）以及 $\mathrm{e}^{\mathrm{i}m\phi}$ 的正交归一性（式 5 [228]），不难证明球谐函数的正交归一性

$$\int Y^*_{l',m'}(\hat{\mathbf{r}}) Y_{l,m}(\hat{\mathbf{r}}) \,\mathrm{d}\Omega = \delta_{ll'}\delta_{mm'} \tag{7}$$

其他性质

$$Y_{l,m}(\hat{\mathbf{r}}) = (-1)^m Y^*_{l,-m}(\hat{\mathbf{r}}) \tag{8}$$

这里的 $(-1)^m$ 是由连带勒让德多项式的性质而来，而共轭由 $\exp(\mathrm{i}m\phi)$ 因子而来.

根据“球坐标与直角坐标的转换[128]”中定义的直角坐标系可得

$$x = \sqrt{\frac{2\pi}{3}} r(Y_{1,-1} - Y_{1,1}) \qquad y = \mathrm{i}\sqrt{\frac{2\pi}{3}} r(Y_{1,-1} + Y_{1,1}) \qquad z = 2\sqrt{\frac{\pi}{3}} r Y_{1,0} \tag{9}$$

旋转变换

$$Y_{l,m}(\hat{\mathbf{r}}') = \sum_{m'=-l}^{l} D^{(l)}_{m,m'}(\mathcal{R})^* Y_{l,m'}(\hat{\mathbf{r}}) \tag{10}$$

其中 $\mathbf{D}^{(l)}$ 是 Wigner D 矩阵[766]. 从量子力学的角度来说，总角动量是与方向无关的，只有角动量在某方向的分量有关.

中心对称

l 为偶数时，球谐函数是中心对称的（偶宇称[343]），否则是反对称的（奇宇称）.

$$Y_{l,m}(-\hat{\mathbf{r}}) = (-1)^l Y_{l,m}(\hat{\mathbf{r}}) \tag{11}$$

可以用图 1 验证.

积分

三个球谐函数之积的积分可以表示成两个 CG 系数[852] 或 3j 符号[855] 相乘[9]

$$\begin{aligned}
&\int Y_{l_1 m_1}(\hat{\mathbf{r}}) Y_{l_2 m_2}(\hat{\mathbf{r}}) Y_{l_3 m_3}(\hat{\mathbf{r}})\,\mathrm{d}\Omega \\
&= (-1)^{m_3}\sqrt{\frac{(2l_1+1)(2l_2+1)}{4\pi(2l_3+1)}}\begin{bmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & -m_3 \end{bmatrix} \\
&= \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}}\begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}
\end{aligned} \tag{12}$$

使用式 8 把第一个球谐函数变为复共轭得

$$\begin{aligned}
&\int Y^*_{l_1 m_1}(\hat{\mathbf{r}}) Y_{l_2 m_2}(\hat{\mathbf{r}}) Y_{l_3 m_3}(\hat{\mathbf{r}})\,\mathrm{d}\Omega \\
&= (-1)^{m_1}\sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}}\begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_1 & l_2 & l_3 \\ -m_1 & m_2 & m_3 \end{pmatrix}
\end{aligned} \tag{13}$$

该式在量子力学中有一些应用，如“氢原子的选择定则[1449]”.

[9] 见 Bransden 附录 A4，以及 Wikipedia 的 3j/CG coefficients 页面

应用

平面波的球谐展开[767]，电多极子展开[1212]，球坐标系中的拉普拉斯方程[744]，类氢原子的定态波函数[1536] 等.

球谐函数表

预备知识 球谐函数[761]

下面列出一些常见的球谐函数

$$Y_{l,m}(\theta,\phi)=\sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)\mathrm{e}^{\mathrm{i}m\phi} \tag{1}$$

$$l=0 \qquad Y_{0,0}=\sqrt{\frac{1}{4\pi}} \tag{2}$$

$$l=1 \qquad \begin{cases} Y_{1,0}=\sqrt{\dfrac{3}{4\pi}}\cos\theta \\ Y_{1,\pm1}=\mp\sqrt{\dfrac{3}{8\pi}}\sin\theta\,\mathrm{e}^{\pm\mathrm{i}\phi} \end{cases} \tag{3}$$

$$l=2 \qquad \begin{cases} Y_{2,0}=\sqrt{\dfrac{5}{16\pi}}(3\cos^2\theta-1) \\ Y_{2,\pm1}=\mp\sqrt{\dfrac{15}{8\pi}}\sin\theta\cos\theta\,\mathrm{e}^{\pm\mathrm{i}\phi} \\ Y_{2,\pm2}=\sqrt{\dfrac{15}{32\pi}}\sin^2\theta\,\mathrm{e}^{\pm2\mathrm{i}\phi} \end{cases} \tag{4}$$

$$l=3 \qquad \begin{cases} Y_{3,0}=\sqrt{\dfrac{7}{16\pi}}(5\cos^3\theta-3\cos\theta) \\ Y_{3,\pm1}=\mp\sqrt{\dfrac{21}{64\pi}}\sin\theta(5\cos^2\theta-1)\,\mathrm{e}^{\pm\mathrm{i}\phi} \\ Y_{3,\pm2}=\sqrt{\dfrac{105}{32\pi}}\sin^2\theta\cos\theta\,\mathrm{e}^{\pm2\mathrm{i}\phi} \\ Y_{3,\pm3}=\mp\sqrt{\dfrac{35}{64\pi}}\sin^3\theta\,\mathrm{e}^{\pm3\mathrm{i}\phi} \end{cases} \tag{5}$$

Wigner D 矩阵

预备知识 球谐函数[761]

若我们把球谐函数 $Y_{l,m}(\hat{\mathbf{r}})$ 绕原点进行某种旋转，得到的函数可以表示成球谐函数的线性组合，且只需要同一个 l 子空间中的球谐函数，若将旋转算符（主动）记为 $\mathcal{R}$，则

$$\mathcal{R}\,|l,m\rangle = \sum_{m'} |l,m'\rangle\ \langle l,m'|\mathcal{R}|l,m\rangle \tag{1}$$

我们把系数矩阵称为 Wigner D 矩阵

$$D^l_{m',m} = \langle l,m'|\mathcal{R}|l,m\rangle \tag{2}$$

表达式

一种定义旋转的方法就是利用三个欧拉角表示旋转，即先将 $|l,m\rangle$ 绕 z 轴逆时针旋转 γ 角，再绕 y 轴逆时针旋转 β 角，最后绕 z 轴旋转 α 角．我们将旋转算符记为 $\mathcal{R}(\alpha,\beta,\gamma)$．由球谐函数的定义式 1 [761]，第一个旋转得到 $\exp(-\mathrm{i}m\gamma)\,|l,m\rangle$．

第二个旋转为

$$\mathcal{R}(0,\beta,0)\,|l,m\rangle = \sum_{m'} \langle l,m'|\mathcal{R}(0,\beta,0)|l,m\rangle\,|l,m'\rangle \tag{3}$$

我们把系数矩阵称为 Wigner d 矩阵，是 Wigner D 矩阵的一个特例．

$$d^l_{m',m}(\beta) = D^l_{m',m}(0,\beta,0) = \langle l,m'|\mathcal{R}(0,\beta,0)|l,m\rangle \tag{4}$$

其表达式为（推导略）

$$\begin{aligned} d^l_{m',m}(\beta) =& \sqrt{(l+m')!(l-m')!(l+m)!(l-m)!}\times \\ & \sum_s \frac{(-1)^{m'-m+s}\left(\cos\frac{\beta}{2}\right)^{2l+m-m'-2s}\left(\sin\frac{\beta}{2}\right)^{m'-m+2s}}{(l+m-s)!s!(m'-m+s)!(l-m'-s)!} \end{aligned} \tag{5}$$

其中 s 的取值范围需要保证被阶乘的数为非负．

第二次旋转完以后我们得到若干个 $|l,m'\rangle$ 的线性组合，要再次绕 z 轴旋转 α 角，只需要把它们分别乘以 $\exp(-\mathrm{i}m'\alpha)$ 即可．所以 Wigner D 矩阵的完整表达式为

$$D^l_{m',m}(\alpha,\beta,\gamma)=\mathrm{e}^{-\mathrm{i}m'\alpha}d^l_{m',m}(\beta)\mathrm{e}^{-\mathrm{i}m\gamma} \tag{6}$$

平面波的球谐展开

- 本词条需要更多讲解，便于帮助理解．

预备知识 平面波[1008]，球谐函数[761]，球贝塞尔函数[759]

复数形式的平面波可以展开为[1011]

$$|\mathbf{k}\rangle=\frac{1}{(2\pi)^{3/2}}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\mathrm{i}^l Y^*_{lm}(\hat{\mathbf{k}})\sqrt{\frac{2}{\pi}}j_l(kr)Y_{lm}(\hat{\mathbf{r}}) \tag{1}$$

可以证明，一组正交归一的基底为（见“径向函数的归一化[769]”）

$$|s_{l,m}(k)\rangle=s_{l,m}(k,\mathbf{r})=\frac{1}{r}\sqrt{\frac{2}{\pi}}krj_l(kr)Y_{lm}(\hat{\mathbf{r}}) \tag{2}$$

平面波式 1 可以表示为相同能量球谐波的线性组合[12]

$$|\mathbf{k}\rangle=\sum_{l,m}\frac{\mathrm{i}^l}{k}Y^*_{lm}(\hat{\mathbf{k}})\,|s_{l,m}(k)\rangle \tag{3}$$

$|s_{l,m}(k)\rangle$ 的完备性可以表示为

$$\int_0^\infty \mathrm{d}k\sum_{l,m}|s_{l,m}(k)\rangle\langle s_{l,m}(k)|=I \tag{4}$$

就像一维傅里叶变换可以表示为

$$|f\rangle=\int_{-\infty}^{\infty}\mathrm{d}k\,|k\rangle\langle k|f\rangle \tag{5}$$

[10] 由 $Y^*_{l,m}=(-1)^mY_{l,-m}$ 易证这里的复共轭可以加在任意一个球谐函数上．

[11] 由 $Y_{l,m}(-\hat{\mathbf{k}})=(-1)^lY_{l,m}(\hat{\mathbf{k}})$ 易得 $\langle\mathbf{k}|$.

[12] 这里不作证明

其中 $|k\rangle = \exp(\mathrm{i}kx)/\sqrt{2\pi}$.

特殊地，如果平面波的方向指向极轴（$\theta = 0$），由对称性，我们只需要 $m = 0$ 的球谐波即可组成平面波．将 $\theta = 0$ 代入式 1 得

$$\mathrm{e}^{\mathrm{i}kz} = \sum_{l=0}^{\infty}(2l+1)\mathrm{i}^l j_l(kr)P_l(\cos\theta) \tag{6}$$

球谐展开函数的傅里叶变换

若三维函数具有球谐展开的形式

$$f(\mathbf{r}) = \frac{1}{r}\sum_{l,m} u_{lm}(r)Y_l^m(\hat{\mathbf{r}}) \tag{7}$$

要做傅里叶变换

$$g(\mathbf{k}) = \langle \mathbf{k}|f\rangle = \frac{1}{(2\pi)^{3/2}}\int f(\mathbf{r})\mathrm{e}^{-\mathrm{i}\mathbf{k}\mathbf{r}}\,\mathrm{d}^3r \tag{8}$$

将式 3 代入上式得

$$g(\mathbf{k}) = \frac{1}{k}\sum_{l,m} g_{l,m}(k)Y_l^m(\hat{\mathbf{k}}) \tag{9}$$

其中

$$g_{l,m}(k) = \sqrt{\frac{2}{\pi}}\mathrm{i}^{-l}\int_0^{+\infty} u_{lm}(r)kr j_l(kr)\,\mathrm{d}r \tag{10}$$

例 1　类氢原子基态的动量谱

类氢原子基态的波函数为（见式 2 [1536]，使用原子单位）

$$\psi(\mathbf{r}) = \frac{Z^{3/2}}{\sqrt{\pi}}\mathrm{e}^{-Zr} \tag{11}$$

显然只有 $l = 0, m = 0$ 球谐项．而 $Y_{0,0} = 1/\sqrt{4\pi}$，所以径向波函数为

$$R_{00}(r) = 2Z^{3/2}\mathrm{e}^{-Zr} \tag{12}$$

所以傅里叶变换为（注意 $j_0(x) = \sin x/x$）

$$g(\mathbf{k}) = \frac{\sqrt{2}}{k\pi}\int_0^{\infty}\mathrm{e}^{-r}\sin(kr)r\,\mathrm{d}r = \frac{2\sqrt{2}}{\pi(k^2+1)^2} \tag{13}$$

$$g(\mathbf{k}) = \left(\frac{2}{Z}\right)^{3/2}\frac{1}{\pi(k^2/Z^2+1)^2} \tag{14}$$

我们也可以将沿 z 轴正方向的三维平面波用球坐标表示（不使用球谐函数），再在球坐标中与波函数积分，结果相同．

球谐波的归一化

预备知识 球贝塞尔函数[759]，平面波的的正交归一化[1348]

若一组径向函数基底 $u_{l,m}(k,r)$ 满足归一化条件

$$\int_0^\infty u_{l,m}^*(k',r)u_{l,m}(k,r)\,\mathrm{d}r = \delta(k-k') \tag{1}$$

则球坐标中一组完备正交归一的函数基底为

$$|s_{l,m}(k)\rangle = \frac{1}{r}u_{l,m}(k,r)Y_{l,m}(\hat{\mathbf{r}}) \tag{2}$$

满足正交归一条件

$$\begin{aligned}&\langle s_{l,m}(k)|s_{l',m'}(k')\rangle\\ =&\int_0^\infty \frac{1}{r}u_{l,m}^*(k,r)\frac{1}{r}u_{l',m'}(k',r)r^2\,\mathrm{d}r\int Y_{l,m}^*(\hat{\mathbf{r}})Y_{l',m'}(\hat{\mathbf{r}})\,\mathrm{d}\Omega\\ =&\delta_{l,l'}\delta_{m,m'}\delta(k-k')\end{aligned} \tag{3}$$

空间中任意复函数可以用式 2 中的基底展开为

$$\begin{aligned}f(\mathbf{r}) &= \sum_{l,m}\int_0^\infty c_{l,m}(k)\,|s_{l,m}(k)\rangle\,\mathrm{d}k\\ &= \frac{1}{r}\sum_{l,m}\left(\int_0^\infty c_{l,m}(k)u_{l,m}(k,r)\,\mathrm{d}k\right)Y_{l,m}(\hat{\mathbf{r}})\end{aligned} \tag{4}$$

函数在任意基底上的投影为

$$\begin{aligned}\langle s_{l,m}(k)|f(\mathbf{r})\rangle &= \sum_{l',m'}\int_0^\infty \mathrm{d}k'\,c_{l',m'}(k')\,\langle s_{l,m}(k)|s_{l',m'}(k')\rangle\\ &= \sum_{l',m'}\int_0^\infty \mathrm{d}k'\,c_{l',m'}(k')\delta_{l,l'}\delta_{m,m'}\delta(k-k')\\ &= c_{l,m}(k)\end{aligned} \tag{5}$$

令 $|f\rangle$ 的分波展开形式为

$$f(\mathbf{r}) = \frac{1}{r}\sum_{l,m}g_{l,m}(r)Y_{l,m}(\hat{\mathbf{r}}) \tag{6}$$

则 $g_{l,m}(r)$ 和 $c_{l,m}(k)$ 有类似傅里叶变换的关系

$$g_{l,m}(r)=\int_0^\infty c_{l,m}(k)u_{l,m}(k,r)\,\mathrm{d}k \tag{7}$$

$$c_{l,m}(k)=\int_0^\infty u^*_{l,m}(k,r)g_{l,m}(r)\,\mathrm{d}r \tag{8}$$

归一化的球贝塞尔函数

由于 $\mathrm{e}^{\mathrm{i}kx}$ 从负无穷到正无穷的归一化积分为 $2\pi\delta(k-k')$，易得 $\sin(kx)$ 的积分为 $\pi\delta(k-k')$（先表示成指数形式），从 0 到正无穷的积分为 $\pi\delta(k-k')/2$.

注意归一化只需要渐进表达式即可（因为局部的不同相对于无穷积分来说可以忽略）. 球贝塞尔函数的渐进形式为 $j_l(kr)\to\sin(kr-l\pi/2)/(kr)$，所以 $krj_l(kr)$ 从 0 到正无穷的归一化积分同样是 $\pi\delta(k-k')/2$. 所以归一化的球贝塞尔函数为

$$u_{l,m}(k,r)=\sqrt{\frac{2}{\pi}}krj_l(kr) \tag{9}$$

渐进形式为

$$u_{l,m}(k,r)\to\sqrt{\frac{2}{\pi}}\sin(kr-l\pi/2) \tag{10}$$

分离变量法与张量积空间

预备知识 拉普拉斯方程[742]

我们可以从张量积空间和本征问题的角度来理解分离变量法.

第一我们假设考虑的方程是线性的. 即解可以表示为齐次解的线性组合加一个非齐次解.

以球坐标的拉普拉斯方程为例，这是一个齐次方程

$$\nabla^2u=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial u}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2u}{\partial\phi^2}=0 \tag{1}$$

我们可以将 u 所在的三维函数空间看做是三个变量各自的函数空间的张量积空间，张量积空间中的基底为三个单变量函数空间中的基底的张量积

$$u(r,\theta,\phi)=f(r)g(\theta)h(\phi) \tag{2}$$

要分离变量，可以从最简单的变量开始，分离出该变量的算符，然后求该算符的本征方程的一组本征函数（本征矢），作为该变量的函数空间的基底. 这样，该算符就可以被替换为常数. 替换后，尝试分离下一个变量的算符，以此类推.

拉普拉斯方程中，首先最容易分离的是关于 ϕ 的算符

$$\frac{\partial^2}{\partial\phi^2} \tag{3}$$

令该算符的本征函数 $h_m(\phi)$ 为 ϕ 函数空间的基底，令本征值为常数 $-m^2$

$$\frac{\partial^2 h}{\partial\phi^2} = -m^2 h \tag{4}$$

将算符替换为本征值，然后容易发现后两项中都有公共的 $1/r^2$ 因子，提取出来就可以分离出只含有 θ 的算符

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) - \frac{m^2}{\sin^2\theta} \tag{5}$$

令该算符的本征函数 $g_l(\theta)$ 为 θ 函数空间的基底，令本征值为常数 $-l(l+1)$.

最后，我们剩下仅关于 r 的方程，同样可以表示为本征方程

$$\frac{\partial}{\partial r}\left(r^2\frac{\partial f}{\partial r}\right) = l(l+1)f \tag{6}$$

令 $f_l(r)$ 为 r 函数空间的基底.

根据施图姆—刘维尔定理，三组本征基底都是正交且完备的，所以张量积空间中的基底也是正交完备的

$$u_{l,m}(r,\theta,\phi) = f_l(r)g_l(\theta)h_m(\phi) \tag{7}$$

广义球谐函数

预备知识 CG 系数[852]

有了 CG 系数的相位约定和球谐函数的相位约定，就可以定义**广义球谐函数（Generalized Spherical Harmonics）**

$$\mathcal{Y}^{L,M}_{l_1,l_2}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)=\sum_{m_1,m_2}\begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M\end{bmatrix} Y_{l_1m_1}(\hat{\mathbf{r}}_1)Y_{l_2m_2}(\hat{\mathbf{r}}_2) \tag{1}$$

由于只有 $m_1+m_2=M$ 时 CG 系数才不为零，求和就由双重求和变为了单个求和．上下限

宇称

宇称（parity）算符 Π 的作用是把所有自变量乘以 -1 得到的函数．本征函数是所有中心对称或反对车的函数，本征值为分别为 ±1．

球谐函数是宇称算符的本征矢，本征值为 $(-1)^l$ （式 8 [761]），易得广义球谐函数也是宇称算符的本征矢，本征值为 $(-1)^{l_1+l_2}$

$$\Pi\mathcal{Y}^{L,M}_{l_1,l_2}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)=\mathcal{Y}^{L,M}_{l_1,l_2}(-\hat{\mathbf{r}}_1,-\hat{\mathbf{r}}_2)=(-1)^{l_1+l_2}\mathcal{Y}^{L,M}_{l_1,l_2}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2) \tag{2}$$

交换对称性

广义球谐函数不具有交换对称性[13]（除非 $l_1=l_2$）

$$P_{12}\mathcal{Y}^{L,M}_{l_1,l_2}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)=\mathcal{Y}^{L,M}_{l_1,l_2}(\hat{\mathbf{r}}_2,\hat{\mathbf{r}}_1)=(-1)^{l_1+l_2-L}\mathcal{Y}^{L,M}_{l_2,l_1}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2) \tag{3}$$

也可以记为

$$\mathcal{Y}^{L,M}_{l_1,l_2}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)=(-1)^{l_1+l_2-L}\mathcal{Y}^{L,M}_{l_2,l_1}(\hat{\mathbf{r}}_2,\hat{\mathbf{r}}_1) \tag{4}$$

但是我们可以轻易构造对称（+）或反对称（-）的函数

$$\mathcal{Y}^{L,M}_{l_1,l_2}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)\pm(-1)^{l_1+l_2-L}\mathcal{Y}^{L,M}_{l_2,l_1}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2) \tag{5}$$

$$\Psi(\mathbf{r}_1,\mathbf{r}_2)=R(r_1,r_2)\mathcal{Y}^{L,M}_{l_1,l_2}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)-R(r_1,r_2)\mathcal{Y}^{L,M}_{l_2,l_1}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)\qquad(l_1+l_2-L=odd) \tag{6}$$

[13]式中 l_1+l_2-L 前面的负号可以变为加号，但习惯上仍然写成负号，因为虽然球谐函数要求 L 为整数，但在涉及到电子自旋时 L 有可能是半整数，这时就不能变为加号了．

证明可以用 CG 系数的对称性（交换左边两列，式 7 [852]）

$$\begin{aligned}\mathcal{Y}_{l_1,l_2}^{L,M}(\hat{\mathbf{r}}_2,\hat{\mathbf{r}}_1) &= \sum_{m_1,m_2}\begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M\end{bmatrix} Y_{l_1,m_1}(\hat{\mathbf{r}}_2)Y_{l_2,m_2}(\hat{\mathbf{r}}_1) \\ &= (-1)^{l_1+l_2-L}\sum_{m_1,m_2}\begin{bmatrix} l_2 & l_1 & L \\ m_2 & m_1 & M\end{bmatrix} Y_{l_2,m_2}(\hat{\mathbf{r}}_1)Y_{l_1,m_1}(\hat{\mathbf{r}}_2) \\ &= (-1)^{l_1+l_2-L}\mathcal{Y}_{l_2,l_1}^{L,M}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)\end{aligned} \tag{7}$$

从式 3 也可以得到对易关系

$$\left[L^2, P_{12}\right] = [L_z, P_{12}] = 0 \tag{8}$$

即交换两个粒子不改变总角动量.

操作类算符（宇称，平移，交换）如果与某物理量算符对易，就说明波函数经过该操作，改物理量守恒. 如果哈密顿中的某一项算符（例如 z 方向的电场的 dipole）如果与某物理量算符（例如 L_z）对易，就说明波函数经过该传播子传播，该物理量守恒.

正交性

将式 10 [852] 用这里的符号表示，就是

$$\int \mathcal{Y}_{l_1',l_2'}^{L',M'}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)\mathcal{Y}_{l_1,l_2}^{L,M}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)\,\mathrm{d}\Omega_1\,\mathrm{d}\Omega_2 = \delta_{l_1,l_1'}\delta_{l_2,l_2'}\delta_{L,L'}\delta_{M,M'} \tag{9}$$

其他性质

$$\mathcal{Y}_{l_1,l_2}^{L,-M}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2) = (-1)^{l_1+l_2+L+M}\mathcal{Y}_{l_1,l_2}^{L,M}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)^* \tag{10}$$

其中 $*$ 表示复共轭．推导如下，

$$\begin{aligned}\mathcal{Y}_{l_1,l_2}^{L,-M}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2) &= \sum_{m_1+m_2=M}\begin{bmatrix} l_1 & l_2 & L \\ -m_1 & -m_2 & -M\end{bmatrix} Y_{l_1,-m_1}(\hat{\mathbf{r}}_1)Y_{l_2,-m_2}(\hat{\mathbf{r}}_2) \\ &= (-1)^{l_1+l_2+L}\sum_{m_1+m_2=M}\begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M\end{bmatrix} Y_{l_1,-m_1}(\hat{\mathbf{r}}_1)Y_{l_2,-m_2}(\hat{\mathbf{r}}_2) \\ &= (-1)^{l_1+l_2+L+M}\sum_{m_1+m_2=M}\begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M\end{bmatrix} Y_{l_1,m_1}^{*}(\hat{\mathbf{r}}_1)Y_{l_2,m_2}^{*}(\hat{\mathbf{r}}_2) \\ &= (-1)^{l_1+l_2+L+M}\mathcal{Y}_{l_1,l_2}^{L,M}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)^{*}\end{aligned} \tag{11}$$

误差函数

预备知识 高斯积分[339]

[14]**误差函数（error function）**的定义为

$$\operatorname{erf}(x)=\frac{1}{\sqrt{\pi}}\int_{-x}^{x}\mathrm{e}^{-t^2}\,\mathrm{d}t=\frac{2}{\sqrt{\pi}}\int_{0}^{x}\mathrm{e}^{-t^2}\,\mathrm{d}t \tag{1}$$

这是一个奇函数．根据式 5 [339] 有

$$\operatorname{erf}(\pm\infty)=\pm 1 \tag{2}$$

由式 13 [202]，误差函数的导数为

$$\frac{\mathrm{d}}{\mathrm{d}x}\operatorname{erf}(x)=\frac{2}{\sqrt{\pi}}\mathrm{e}^{-x^2} \tag{3}$$

即 e^{-x^2} 的不定积分（原函数）为

$$\int \mathrm{e}^{-x^2}\,\mathrm{d}x=\frac{\sqrt{\pi}}{2}\operatorname{erf}(x)+C \tag{4}$$

级数展开

[14]参考 Wikipedia 相关页面．

预备知识 泰勒级数[181]

由指数函数的级数展开得

$$\mathrm{e}^{-x^2}=\sum_{n=0}^{\infty}\frac{1}{n!}(-x^2)^n=\sum_{n}\frac{(-1)^n}{n!}x^{2n} \tag{5}$$

对各项做不定积分代入式 4 可得误差函数的级数展开为

$$\operatorname{erf}(x)=\frac{2}{\sqrt{\pi}}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)n!}x^{2n+1}=\frac{2}{\sqrt{\pi}}\left(x-\frac{x^3}{3}+\frac{x^5}{10}-\frac{x^7}{42}+\frac{x^9}{216}\cdots\right) \tag{6}$$

例 1

令 a 为实数且 $a>0$，计算不定积分

$$\int\exp\left(-ax^2+bx\right)\mathrm{d}x \tag{7}$$

解：我们可以先将指数部分凑平方得

$$-ax^2+bx=-t^2+\frac{b^2}{4a} \tag{8}$$

其中

$$t=\sqrt{a}x-\frac{b}{2\sqrt{a}} \tag{9}$$

使用上式进行换元积分得

$$\int\exp\left(-ax^2+bx\right)\mathrm{d}x=\frac{1}{\sqrt{a}}\mathrm{e}^{b^2/(4a)}\int\mathrm{e}^{-t^2}\,\mathrm{d}t=\frac{1}{2}\sqrt{\frac{\pi}{a}}\mathrm{e}^{b^2/(4a)}\operatorname{erf}\left(\sqrt{a}x-\frac{b}{2\sqrt{a}}\right) \tag{10}$$

例 2

计算无穷积分（即 e^{-ax^2} 的傅里叶变换[237]）

$$g(k)=\int_{-\infty}^{+\infty}\mathrm{e}^{-ax^2}\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x \tag{11}$$

令 $b=-\mathrm{i}k$，使用例 1 的结论有

$$g(k)=\frac{1}{2}\sqrt{\frac{\pi}{a}}\mathrm{e}^{-k^2/(4a)}\operatorname{erf}\left(\sqrt{a}x+\frac{\mathrm{i}k}{2\sqrt{a}}\right)\Bigg|_{-\infty}^{+\infty}=\sqrt{\frac{\pi}{a}}\mathrm{e}^{-k^2/(4a)} \tag{12}$$

虚误差函数

预备知识 误差函数[774]

虚误差函数的定义为

$$\operatorname{erfi}(x)=-\mathrm{i}\operatorname{erf}(\mathrm{i}x)=\frac{1}{\sqrt{\pi}}\int_{-x}^{x}\mathrm{e}^{t^2}\,\mathrm{d}t=\frac{2}{\sqrt{\pi}}\int_{0}^{x}\mathrm{e}^{t^2}\,\mathrm{d}t \tag{1}$$

推导如下，使用换元法 $t=\mathrm{i}\tau$ 有

$$\operatorname{erfi}(x)=\frac{-\mathrm{i}}{\sqrt{\pi}}\int_{-\mathrm{i}x}^{\mathrm{i}x}\mathrm{e}^{t^2}\,\mathrm{d}t=\frac{-\mathrm{i}}{\sqrt{\pi}}\int_{-x}^{x}\mathrm{e}^{-(\mathrm{i}\tau)^2}\,\mathrm{d}(\mathrm{i}\tau)=\frac{1}{\sqrt{\pi}}\int_{-x}^{x}\mathrm{e}^{\tau^2}\,\mathrm{d}\tau \tag{2}$$

同理可得

$$\operatorname{erfi}(\mathrm{i}x)=\mathrm{i}\operatorname{erf}(x) \tag{3}$$

其导函数为

$$\frac{\mathrm{d}}{\mathrm{d}x}\operatorname{erfi}(x)=\frac{2}{\sqrt{\pi}}\mathrm{e}^{x^2} \tag{4}$$

所以

$$\int\mathrm{e}^{x^2}\,\mathrm{d}x=\frac{\sqrt{\pi}}{2}\operatorname{erfi}(x)+C \tag{5}$$

与误差函数的级数展开同理，$\operatorname{erfi}(x)$ 的级数展开为

$$\operatorname{erfi}(x)=\frac{2}{\sqrt{\pi}}\sum_{n=0}^{\infty}\frac{x^{2n+1}}{(2n+1)n!}=\frac{2}{\sqrt{\pi}}\left(x+\frac{x^3}{3}+\frac{x^5}{10}+\frac{x^7}{42}+\frac{x^9}{216}\cdots\right) \tag{6}$$

超几何函数

广义超几何函数表示为[15]

$${}_pF_q(a_1,\ldots,a_p;b_1,\ldots,b_q;z)=\sum_{n=0}^{\infty}\frac{(a_1)_n\ldots(a_p)_n}{(b_1)_n\ldots(b_q)_n}\frac{z^n}{n!} \tag{1}$$

其中 $(a)_n=a(a+1)\ldots(a+n-1)$，叫做 **Pochhammer 符号**.

[15]但据说当 $p>q+1$ 时级数不收敛

合流超几何函数 ${}_1F_1(a;b;z)$

求平面库仑波函数需要使用 ${}_1F_1(a;b;z)$，其级数展开为

$$ {}_1F_1(a;b;z)=\sum_{n=0}^{\infty}\frac{(a_1)_n}{(b_1)_n}\frac{z^n}{n!} \tag{2} $$

渐进展开为

$$ \begin{aligned} {}_1F_1(a;b;z)=&\frac{(-1)^a\Gamma(b)}{\Gamma(b-a)}\sum_{n=0}^{\infty}(-1)^n\frac{(a)_n(a-b+1)_n}{n!}z^{-n-a}\\ &+\frac{\Gamma(b)}{\Gamma(a)}\sum_{n=0}^{\infty}\frac{(b-a)_n(1-a)_n}{n!}z^{-n+a-b}\mathrm{e}^z \end{aligned} \tag{3} $$

连续分数展开为

$$ {}_1F_1(a;b;z)=1+\frac{az/b}{1+\ldots}\,\frac{-c_1z}{1+c_1z+\ldots}\,\frac{-c_2z}{1+c_2z+\ldots} \tag{4} $$

$$ c_n=\frac{a+n}{(n+1)(b+n)} \tag{5} $$

椭圆积分

预备知识 换元积分法[188]

第一类不完全椭圆积分（incomplete elliptic integral of the first kind）

$$ F(\phi,k)=\int_0^{\phi}\frac{\mathrm{d}x}{\sqrt{1-k^2\sin^2x}} \tag{1} $$

椭圆积分有时候也记为 $F(\phi|k^2)=F(\sin\phi;k)$.

使用积分换元法，令 $t=\sin x$，有

$$ F(\phi,k)=\int_0^{\phi}\frac{\mathrm{d}t}{\sqrt{(1-t^2)(1-k^2t^2)}} \tag{2} $$

另一种换元法是令 $\theta=2x$，$k=\csc(\theta_0/2)$，$F(\phi,k)$ 也可以表示为

$$ F\left(\phi,\csc\frac{\theta_0}{2}\right)=\frac{1}{\sqrt{2}}\sin\frac{\theta_0}{2}\int_0^{2\phi}\frac{\mathrm{d}\theta}{\sqrt{\cos\theta-\cos\theta_0}} \tag{3} $$

数值计算

$F(\phi|m)$ 的 Matlab 函数为 `ellipticF(phi, m)`，Mathematica 函数为 `EllipticF[phi, m`

（图未完成）

库仑函数

- 本词条缺少预备知识，初学者可能会遇到困难.
- 本词条需要更多讲解，便于帮助理解.

预备知识 常微分方程[211]，Gamma 函数[254]

[16]在球坐标系中解库仑势场中的定态薛定谔方程[1328]，会得到径向方程（式 7 [1431]）

$$\frac{\mathrm{d}^2u}{\mathrm{d}\rho^2}+\left[1-\frac{2\eta}{\rho}-\frac{l(l+1)}{\rho^2}\right]u=0 \tag{1}$$

其解为**第一类库仑函数** $F_l(\eta,\rho)$ 和**第二类库仑函数** $G_l(\eta,\rho)$. 第一类库仑函数的解析式为[17]

$$F_l(\eta,\rho)=A_l(\eta)\rho^{l+1}\mathrm{e}^{\mp\mathrm{i}\rho}M(l+1\mp\mathrm{i}\eta,2l+2,\pm2\mathrm{i}\rho) \tag{2}$$

其中

$$A_l(\eta)=\frac{2^l\mathrm{e}^{-\pi\eta/2}|\Gamma(l+1+\mathrm{i}\eta)|}{(2l+1)!} \tag{3}$$

$M(a,b,z)$ 是**库默尔（Kummer）M 函数**，也叫第一类合流超几何函数[776]，记为 ${}_1F_1(a;b;z)$. 库仑函数也可以用**惠特克（Whittaker）M 函数**来表示得更紧凑

$$F_l(\eta,\rho)=\left(\frac{\mp\mathrm{i}}{2}\right)^{l+1}A_l(\eta)M_{\pm\mathrm{i}\eta,l+\frac{1}{2}}(\pm2\mathrm{i}\rho) \tag{4}$$

其中惠特克 M 函数与库默尔 M 函数的关系为

$$M_{\kappa,\mu}(z)=\mathrm{e}^{-z/2}z^{\mu+1/2}M(\mu-\kappa+1/2,1+2\mu,z) \tag{5}$$

[16]参考 NIST 相关页面

[17]这里的两套正负号是等效的.

$F_l(\eta,\rho)$ 是一个实函数，类似 ρ 乘以第一类球贝塞尔函数[18] $j_l(\rho)$，有

$$F_l(\eta,0)=0 \qquad \left.\frac{\mathrm{d}F_l(\eta,\rho)}{\mathrm{d}\rho}\right|_{\rho=0}=\begin{cases}A_0(\eta) & (l=0)\\ 0 & (l>0)\end{cases} \tag{6}$$

且渐进形式为

$$F_l(\eta,\rho)\overset{\rho\to\infty}{\longrightarrow}\sin\left(\rho-\frac{\pi l}{2}-\eta\ln(2\rho)+\sigma_l\right) \tag{7}$$

其中 $\sigma_l(\eta)$ 是**库仑相移（Coulomb phase shift）**

$$\sigma_l(\eta)=\arg[\Gamma(l+1+\mathrm{i}\eta)] \tag{8}$$

其中 arg 函数是复数的幅角（式 5 [142]）.

第二类库仑函数 $G_l(\eta,\rho)$ 可以由 $H_l^{\pm}(\eta,\rho)$ 得到，后者是两类库仑函数的线性组合[19]

$$H_l^{\pm}(\eta,\rho)=G_l(\eta,\rho)\pm\mathrm{i}F_l(\eta,\rho) \tag{9}$$

定义为

$$H_l^{\pm}(\eta,\rho)=(\mp\mathrm{i})^l\mathrm{e}^{\pi\eta/2\pm\mathrm{i}\sigma_l(\eta)}W_{\mp\mathrm{i}\eta,l+1/2}(\mp2\mathrm{i}\rho) \tag{10}$$

其中 $W_{\kappa,\mu}(z)$ 是**惠特克（Whittaker）W 函数**

$$W_{\kappa,\mu}(z)=\mathrm{e}^{-z/2}z^{\mu+1/2}U(\mu-\kappa+1/2,1+2\mu,z) \tag{11}$$

其中 $U(a,b,z)$ 是**库默尔（Kummer）U 函数**，定义为

$$U(a,b,z)=\frac{\Gamma(1-b)}{\Gamma(a+1-b)}M(a,b,z)+\frac{\Gamma(b-1)}{\Gamma(a)}z^{1-b}M(a+1-b,2-b,z) \tag{12}$$

式 11 代入式 10 得

$$H_l^{\pm}(\eta,\rho)=\mathrm{e}^{\pm\mathrm{i}\theta_l(\eta,\rho)}(\mp2\mathrm{i}\rho)^{l+1\pm\mathrm{i}\eta}U(l+1\pm\mathrm{i}\eta,2l+2,\mp2\mathrm{i}\rho) \tag{13}$$

其中

$$\theta_l(\eta,\rho)=\rho-\eta\ln(2\rho)-\frac{l\pi}{2}+\sigma_l(\eta) \tag{14}$$

是式 7 中的渐进相位.

[18] 当 $\eta=0$ 时二者相等.

[19] 类比 $\exp(\mathrm{i}x)=\cos x+\mathrm{i}\sin x$

$G_l(\eta,\rho)$ 的渐进形式为

$$\lim_{\rho\to\infty} G_l(\eta,\rho)=\cos\theta_l(\eta,\rho) \tag{15}$$

库仑函数的导数可以由惠特克函数的导数[20]得到.

$$\frac{\partial}{\partial z}[M_{a,b}(z)]=\left(\frac{1}{2}-\frac{a}{z}\right)M_{a,b}(z)+\frac{1}{x}\left(a+b+\frac{1}{2}\right)M_{a+1,b}(z) \tag{16}$$

$$\frac{\partial}{\partial z}[W_{a,b}(z)]=\frac{1}{2z}[(z-2a)W_{a,b}(z)-2W_{a+1,b}(z)] \tag{17}$$

可见程序中求导数大概要比求函数值多用一倍时间（因为要求两次惠特克函数），但同时也顺便求出了函数值. 如果同时需要两类库仑函数及它们的导数，只需求 H^+ 函数的导数，这就顺便求出了 H^+，再分别取实部虚部即可. 另外只需要把 H^+ 做复共轭即可得到 H^-.

一种验证函数值是否正确的方法是使用以下性质

$$\mathcal{W}\{G,F\}=\mathcal{W}\{H^{\pm},F\}=1 \tag{18}$$

其中 $\mathcal{W}\{f_1(x),f_2(x)\}$ 是二阶**朗斯基行列式（Wronskian determinant）**

$$\mathcal{W}\{f_1(x),f_2(x)\}=\begin{vmatrix} f_1(x) & f_2(x) \\ f_1'(x) & f_2'(x) \end{vmatrix}=f_1(x)f_2'(x)-f_2(x)f_1'(x) \tag{19}$$

由渐进形式可得径向归一化积分[21]与球贝塞尔函数的相同

$$\int_0^\infty F_l(Z/k',k'r)F_l(Z/k,kr)\,\mathrm{d}r=\int_0^\infty \sin(k'r)\sin(kr)\,\mathrm{d}r=\frac{\pi}{2}[\delta(k-k')-\delta(k+k')] \tag{20}$$

[20]可以使用 Wolfram Alpha 或 Mathematica

[21]积分时可忽略 sin 中的额外相位，但笔者不会证.

第九部分

数学分析

第一章

基础

集合的极限

预备知识 集合[92]，极限[149]

特征函数

如果我们有可数个集合构成的族 $\{A_n\}$，每个 A_n 都是编了号的集合，那么对于给定的元素，我们可以观察它是否存在于各 A_n 中．为了表达方便，我们用一个**特征函数**，$\mathcal{X}$，来表达属于关系：对于元素 x 和集合 A_n，当 $x \in A_n$ 时 $\mathcal{X}_{A_n}(x) = 1$；当 $x \notin A_n$ 时 $\mathcal{X}_{A_n}(x) = 0$.

有了特征函数，我们就可以套用数列的极限来讨论集合的极限．

上极限和下极限

给定一列编好了号的集合：$\{A_n\}$．为了方便之后的讨论，先定义两列集合：$U_k = \bigcup\limits_{i \geq k} A_i$，$J_k = \bigcap\limits_{i \geq k} A_i$.

这么取的两列集合有很棒的性质：对于任何正整数 k，$U_{k+1} \subseteq U_k$，而 $J_{k+1} \supseteq J_k$．如果说真子集时比原来的集合要小的话，我们可以认为 U_k 是一个**单调不增**的集列，而 J_k **单调不减**．这么一来，对于任何元素 x，不管 x 是一个数字，一种动物，还是一片凋落在安田讲堂[1]前的银杏叶，它的特征函数 $\mathcal{X}_{U_n}(x)$ **单调不增**，也就是说一旦某个 U_n 不包含 x，那么任何 $m > n$ 的 U_m 也不包含 x；同样地，$\mathcal{X}_{J_n}(x)$ **单调不减**，也就是说也就是说一旦某个 J_n 包含 x，那么任何 $m > n$ 的 J_m 也包含 x.

单调函数都是必然有极限的，这么一来，任给元素 x，$\mathcal{X}_{U_n}(x)$ 和 $\mathcal{X}_{J_n}(x)$ 都有极限．那么我们可以定义 U_n 和 J_n 的极限如下：

定义 1　上极限

给定一列编好了号的集合：$\{A_n\}$，令 $U_k = \bigcup\limits_{i \geq k} A_i$．定义 $\lim\limits_{k \to \infty} U_k = U$，其中 $U = \{x | \text{存在无穷多个正整数} n, \text{使} \mathcal{X}_{U_n}(x) = 1\}$．称这个 U 是集列 $\{A_n\}$ 的上极限集合（**upper limit set**）.

[1] 安田讲堂是东京大学的大厅.

简单来说，某元素 x 在上极限 U 中，当且仅当 x 在**无穷多个** A_n 中.

类似地有：

定义 2　下极限

给定一列编好了号的集合：$\{A_n\}$，令 $J_k = \bigcap_{i\geq k} A_i$. 定义 $\lim_{k\to\infty} J_k = J$，其中 $J = \{x|$存在正整数N使得$\forall n > N, \mathcal{X}_{J_n}(x) = 1\}$. 称这个 J 是集列 $\{A_n\}$ 的下极限集合（**lower limit set**）.

简单来说，某元素 x 在下极限 J 中，当且仅当从集合序列的某一个（比如 A_N）开始，x 在之后的**每一个** A_n 中. 也可以简单地说，只有有限个 A_n 不包含 x.

如果从 A_n 的视角来看的话，上极限 U 中的元素是包含于无穷多个 A_n 的. 我们也可以记上极限为 $U = \overline{\lim_{n\to\infty}} A_n$.

下极限 J 中的元素，从某个编号 m 开始，包含于**每一个** $n > m$ 的 A_n. 我们也可以记下极限为 $J = \underline{\lim_{n\to\infty}} A_n$.

上下极限的计算

给定一列编好了号的集合：$\{A_n\}$，则

$$\overline{\lim_{n\to\infty}} A_n = \bigcap_{k=1}^{\infty}\bigcup_{i=k}^{\infty} A_i = \bigcap_{k=1}^{\infty} U_k \tag{1}$$

类似地

$$\underline{\lim_{n\to\infty}} A_n = \bigcup_{k=1}^{\infty}\bigcap_{i=k}^{\infty} A_i = \bigcup_{k=1}^{\infty} J_k \tag{2}$$

这两个计算公式的依据是单调性. 由于 U_k 单调不增，J_k 单调不减，于是 $\bigcap_{k=1}^{N} U_k = U_N$，因此

$$\lim_{k\to\infty} U_k = \lim_{N\to\infty}\bigcap_{k=1}^{N} U_k = \bigcap_{k=1}^{\infty} U_k \qquad \lim_{k\to\infty} J_k = \lim_{N\to\infty}\bigcup_{k=1}^{N} J_k = \bigcup_{k=1}^{\infty} J_k \tag{3}$$

极限

给定一列编好了号的集合：$\{A_n\}$，那么它的上极限和下极限总是存在的.

但是，如果有一个元素 x 很不听话，随着 n 向无穷增大，$\mathcal{X}_{A_n}(x)$ 一会儿是 1，一会儿是 0，那就没法定义 $\{A_n\}$ 的极限，因为你没法说清楚这个不听话的 x 究竟是不是极限集合中的元素．所以，为了良好地定义 $\{A_n\}$ 的极限，就不允许不听话的 x 出现．

单调集列自然不存在不听话的 x，但是单调性是一个过分强的要求．为了让集列 $\{A_n\}$ 有极限，我们只需要它的上极限等于下极限：$\overline{\lim\limits_{n\to\infty}} A_n = \underline{\lim\limits_{n\to\infty}} A_n$．由于每一个 $J_k \subseteq U_k$，故已经必有 $\underline{\lim\limits_{n\to\infty}} A_n \subseteq \overline{\lim\limits_{n\to\infty}} A_n$，所以我们只要求 $\overline{\lim\limits_{n\to\infty}} A_n \subseteq \underline{\lim\limits_{n\to\infty}} A_n$．当 $\overline{\lim\limits_{n\to\infty}} A_n \subseteq \underline{\lim\limits_{n\to\infty}} A_n$ 时，上下极限相等，我们就统一把它们叫做 $\{A_n\}$ 的**极限（limit set）**.

度量空间

预备知识 集合[92]

度量空间是除拓扑空间[619] 外最广义的空间．它在集合的基础上增加了距离或长度的概念．

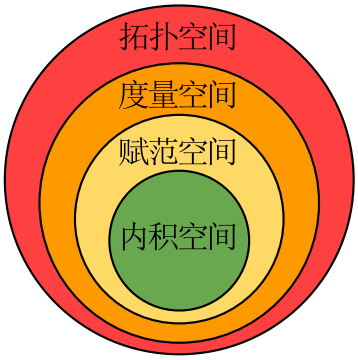


图 1: 用维恩图表示几种不同空间之间的关系，从内到外分别是内积空间[471]，赋范空间[794]，度量空间，拓扑空间[619]（修改自维基百科）

定义 1　度量空间

一个集合中任意两个元素 u, v 间若定义了满足以下条件的距离函数（**distance function**）$d(u, v)$（函数值为实数），那它就是一个度量空间（**metric space**）．集合中的每个元素就叫空间中的一个点．

- 正定性：$d(u,v)\geq 0$，且 $d(u,v)=0$ 当且仅当 $u=v$
- 对称性：$d(u,v)=d(v,u)$
- 三角不等式：$d(u,v)\leqslant d(u,w)+d(w,v)$

其中“三角不等式”就是通常所说的“三角形两边之和不小于第三边”，移向后就变为“两边之差不大于第三边”.

未完成：修改批注：将原先的四个条件整合成三个条件，并冠以数学界习惯的名称，方便学生记忆.

例 1　欧几里得空间

N 维欧几里得空间 $\mathbb{R}^N$ 中通常定义距离函数为

$$d(x,y)=\sqrt{\sum_{i=1}^{N}(x_i-y_i)^2} \tag{1}$$

那么它是一个度量空间（证明留做习题）.

特殊地，实数域 $\mathbb{R}$ 通常的距离函数为 $d(x,y)=|x-y|$.

日常生活中，我们关于距离的直观概念都是建立在例 1 的基础上的，但度量空间是非常广义和抽象[102] 的. 例如上例中 $d(x,y)=|x^3-y^3|$ 也可以是 $\mathbb{R}$ 的距离函数；又例如我们可以把一些函数的集合看成一个度量空间：

例 2

所有 $f:\mathbb{R}\to\mathbb{R}$ 函数[103] 的集合是一个度量空间，如果定义距离函数为

$$d(f,g)=\max|f(x)-g(x)| \tag{2}$$

证明留做习题.

推论 1

度量空间 X 的子集 A 若继承 X 的距离函数，那么 A 也是一个度量空间，称为 X 的子空间（**subspace**）.

证明显然.

对比线性空间

虽然例 1 和例 2 的集合也可以用于定义线性空间，但二者却有较大区别：比起线性空间[444]，度量空间有距离或长度的概念而线性空间却不一定（见范数[794]）．线性空间必须定义“加法”和“标量积”两种运算而度量空间不必．线性空间的“零元”在度量空间也不必存在．

度量空间中的概念

预备知识 度量空间[785]

度量空间与一般集合的最大区别就是元素之间有了距离的概念．利用距离函数我们可以定义许多度量空间特有的基本概念．

定义 1　邻域

给定一个半径 $r>0$，度量空间 X 中的一点 x 周围所有满足 $d(x,y)<r$ 的点 $y\in X$（包括 x 自己）就是 x 在 X 中的一个邻域（**neighborhood**）N_r．如果将邻域去掉 x 本身，就叫做去心邻域（**deleted/punctured neighbourhood**）．

注意邻域取决于所讨论的度量空间，例如即使当 x 和 r 不变，当 X 分别取有理数集和实数集时，邻域 N_r 也是不同的．如果 x 同时属于两个不同空间 A,B，那么我们应该区分 x 在 A 中的邻域和 x 在 B 中的邻域：前者是 A 的子集，而后者是 B 的子集．以下许多概念也与讨论的空间有关，所以在讨论时我们需要明确使用哪个空间．

邻域至少包含一个点，而去心邻域可以是空集．

定义 2　内点

给定度量空间 X 的一个子集 A．如果某点 $x\in A$ 在 X 中的某邻域是 A 的子集，那么 x 就是集合 A 的内点（**interior point**）．

换言之：给定度量空间 X 的一个子集 A 以及 $x\in A$，如果存在 $r>0$ 使所有 X 中所有与 x 距离小于 r 的点都属于 A，那么 x 就是 A 的内点．

例 1

一个点是否是集合 A 的内点取决于 X 的定义．例如令 A 为 $(-1,1)$ 中的有

理数，X 为有理数集，则 A 中的任意点都是内点．但如果令 X 为实数集，那么 A 中的任意点都不是内点，因为可以证明任意两个有理数之间都存在实数．若令 $X = A$，那么所有点都是内点．

定义 3　极限点，离散点

给定度量空间 X 中的一点 x，如果 x（在 X 中）[2]的任意去心邻域都不为空，那么 x 就是集合 X 的一个极限点（**limit point**）．如果一个点不是极限点，它就是离散点（**discrete point**）．

例 2

有理数集或实数集（$\mathbb{R}$）构成的度量空间中任意一点都是极限点（证明显然）．

定义 4

如果讨论度量空间 X 和它的子空间 A，那么如果存在一点 $x \in X$（但未必 $x \in A$），使得 x 在 A 中的所有去心邻域都不为空，那么就说 x 是 A 在 X 中的极限点．

根据定义 3，A 在 X 的极限点都是 X 的极限点．当 $x \in A$ 时可以直接说 x 是 A 的极限点，但 $x \notin A$ 时则不能．

例如有理数（作为 $\mathbb{R}$ 的一个子集）的（在 $\mathbb{R}$ 中的）极限点却不一定是有理数，例如 $\sqrt{2}$ 的任意邻域中都有无穷个有理数，所以是有理数集的一个极限点，但 $\sqrt{2}$ 却是无理数．

推论 1

度量空间 X 中的点 x 是极限点的充分必要条件是，x 的任意去心邻域都有无穷多个点．

证明：使用反证法．如果 x 的某个去心领域只有有限个点，那么必定能找到离 x 最近的一点 y，那么对于任意的 $r < d(x, y)$，x 的去心邻域为空，与定义矛盾．证毕．

推论 2

有限元素的集合都不存在极限点．

[2]在定理中，我们一般假设只存在提到的集合，所以这里的“（在 X 中）”可以省略．

定义 5　序列的极限

给定度量空间 X 中的无穷个点组成的序列（**sequence**）$x_1, x_2, \ldots$ 以及一点 x，若对任意给定的 $\epsilon > 0$，总存在 N 使得当 $n > N$ 时就有 $d(x_n, x) < \epsilon$，那么 x 就是该序列的极限（**limit**）.

需要注意：

- 度量空间中序列的极限未必是极限点，例如整数集 $\mathbb{Z}$ 中的序列 $1, 2, 3, 3, 3, \ldots$ 的极限是 3，但 $\mathbb{Z}$ 中任意一点都不是（$\mathbb{Z}$ 的）极限点.
- 序列的极限是否存在可能与 X 有关. 例如令 X 为正实数集，则序列 $1, 1/2, 1/3, \ldots$ 不存在极限，但若 X 改为实数集，那么该序列就存在极限. 若问题中存在两个集合 $A \subsetneq X$ 时，我们需要指定哪个集合中的极限.

推论 3

给定度量空间 X 的任意极限点 x，都可以构建一个 X 中的序列使其极限为 x.

给定 $A \subsetneq X$ 的在 X 中的任意极限点 $y \in X, y \notin A$，都可以构建一个 A 中的序列使其在 X 中的极限点为 y.

证明显然.

开集和闭集

我们初高中所学的开区间就是实数集 $\mathbb{R}$ 的开集. 任意开区间的并和有限开区间的交也是 $\mathbb{R}$ 的开集. 下面我们对任意度量空间通过距离的概念给出开集的定义.

定义 6　度量空间的开集

若度量空间 X 的子集 A 中的任意一点都是内点（定义 2），那么 A 就是一个开集（**open set**）.

换言之：给定度量空间 X 的一个子集 A，若对于任意 $x \in A$ 都存在 $r > 0$ 使得 X 中与 x 距离小于 r 的所有点都属于 A，那么 A 就是一个开集.

如果 A 不是开集，就说它是非开的（**not open**）.

由于 A 中的点是否为内点取决于 X，所以 A 是否为开集也是相对于 X 而言的. 例 1 中的三种情况下 A 分别是开集、非开集、开集.

定义 7　度量空间的闭集

若度量空间 X 的子集 A 是开集，那么 A 关于 X 的补集就是闭集（**closed set**）.

根据这个定义，显然闭集的补集都是开集. 显然 A 是否是闭集同样取决于 X 的选取.

定理 1

度量空间 X 的子集 A 是闭集的充分必要条件是：A 在 X 中的所有极限点（定义 4）都属于 A.

注意“开集”和“闭集”并不是反义词，一个集合如果看作它本身的子集（$A = X$），则它既开又闭. 例如将例 1 中的开区间改为闭区间，则三种情况下 A 分别是“非开但闭”，“非开非闭”，“又开又闭”.

拓展

即使讨论的不是度量空间（没有距离的概念），我们也可以通过更广义的方式定义开集，见“拓扑空间[619]”. 度量空间是拓扑空间的一种，可以证明上述定义的开集同样满足拓扑空间中对开集的要求，即“有限交任意并”也都是开集.

柯西序列 完备度量空间

预备知识 度量空间中的概念[787]

定义 1　柯西序列

给定某个度量空间 X 中的序列 $\{x_i\}$，当满足以下条件时，他就叫做柯西序列（**Cauchy sequence**）.

对任意 $\varepsilon > 0$，存在 N，当 $m, n \geqslant N$ 时就满足 $d(u_m, u_n) < \varepsilon$.

也就是说，柯西序列要求随着项数趋于无穷，该项之后的所有项之间的距离趋于零.

定义 2　收敛和发散序列

本书把度量空间中的序列称为收敛的（**convergent**）当且仅当它是柯西序列．如果序列不收敛，那它就是发散的（**divergent**）．

定理 1

如果度量空间 X 中的序列存在极限（定义 5 [787]），则它是一个柯西序列．

证明留做习题．

定理 2

度量空间 X 中的柯西序列 $\{x_n\}$ 若在 X 中不存在极限．则可以定义一个新的元素 $x \notin X$，并定义它和任意 $y \in X$ 的距离为

$$d(x,y) = \lim_{n\to\infty} d(x_n, y) \tag{1}$$

（可以证明该极限存在）把 x 添加进 X 后，构成新的度量空间 X'．在 X' 中 x 是序列 $\{x_i\}$ 的极限．

未完成：证明

注意一些教材中对“收敛”的定义更为严格，即只把上述定义中 $x \in X$ 的情况称为收敛．在这种定义下，一些柯西数列既不收敛也不发散（定理 2 中 $x \notin X$ 的情况）．

实数域 $\mathbb{R}$ 上的柯西序列必定是收敛的，这是我们可以通过判断数列是否为柯西序列从而判断该序列是否收敛．但如果我们从 $\mathbb{R}$ 中把收敛的那点挖走，那么这个柯西序列在这个集合中就不收敛．所以柯西序列是否收敛取决于它所属于的集合．

度量空间的完备性

定义 3　完备空间

如果度量空间 X 中任意柯西序列都收敛到 $x \in X$（即存在极限），那么该赋范空间就是完备（**complete**）的．

在定理 2 中我们介绍了一种方法，可以给不完备度量空间 X 不断加入新的元素．如果我们不断重复这一过程，就可以在 X 的基础上得到一个最小的完备空间，叫做 A 的**完备化空间**（**complete metric space**）．

完备的赋范空间常称为**巴拿赫空间（Banach space）**. 其中，完备的内积空间特别称作**希尔伯特空间（Hilbert space）**.

“完备”可以形象理解为空间中没有“漏洞”. 有限维空间都是完备的. 可数维空间都是不完备的. 例如有理数集和多项式组成的空间就是不完备的（柯西序列的极限可以是 e^x，但是 e^x 并不属于该空间）.

完备空间

- 和 cauchy.tex 重复，需要整合

预备知识 度量空间[785]，极限[149]

柯西列

度量空间是定义了“距离”的空间，因此可以讨论点与点之间的远近. 微积分的基础是**极限**的概念；直观来说，如果一个数列 $\{a_n\}$ 在 n 趋于正无穷时有极限 a_0，意思就是说 a_n 随着 n 的增大，越来越靠近 a_0. 因此，表达式 $\lim\limits_{n\to\infty} a_n = a_0$ 可以表述为“不管我们要求的接近程度 ϵ 有多小（同时大于零），总能找到足够大的正整数 N，使得只要 $n > N$，那么 a_n 到 a_0 的距离就小于 ϵ”；或者用数学符号，更紧凑一点，写为：$\forall \epsilon > 0$，$\exists N \in \mathbb{Z}^+$，使得 $\forall n > N$，有 $|a_n - a_0| < \epsilon$.

什么情况下，一个数列 $\{a_n\}$ 存在这样的极限 a_0 呢？这样的数列被称作柯西数列，定义如下.

定义 1　柯西数列

在 $\mathbb{R}$ 中，一个数列 $\{a_n\}$ 被称为柯西数列（**Cauchy Series**），当且仅当对于任意的正距离 ϵ，存在 $N \in \mathbb{Z}$，使得只要 $n, m > N$，就有：$|a_n - a_m| < \epsilon$.

这样的数列就是可以有极限的数列.

习题 1　柯西数列的收敛性

证明：柯西数列必有极限；有极限的数列必是柯西数列.

以上定义中，我们只讨论了“距离”这一性质，因此这些定义都可以推广到任意有距离概念的集合中，也就是度量空间中.

定义 2　柯西列

给定度量空间 $\mathcal{M}$，记两点 x，$y \in \mathcal{M}$ 之间的度量为 $d(x,y)$. 称点列 $\{a_n\}_{n=1}^{\infty} \subseteq \mathcal{M}$ 为一个柯西点列，当且仅当 $\forall \epsilon > 0$，$\exists N \in \mathbb{Z}^+$，使得 $\forall n, m > N$，有 $d(a_n, a_m) < \epsilon$.

柯西点列、柯西数列也可以简称/统称为柯西列.

一个柯西列只能收敛到一个点上.

定理 1　收敛点的唯一性

给定度量空间 $\mathcal{M}$，如果 $\{a_n\}_{n=1}^{\infty}$ 是一个柯西列，那么这个柯西列只收敛到一个点上.

证明是很简单的. 反设柯西列 $\{a_n\}$ 有两个收敛点 x_0 和 y_0，那么找出这两个点的距离 $L = d(x_0, y_0)$，由收敛点的定义，存在正整数 N 使得 $\forall n > N$，使得 $d(a_n, x_0) < L/2$ 和 $d(a_n, y_0) < L/2$. 这样的 a_n 不可能存在，因为它到 x_0 和 y_0 两点的距离之和小于 x_0 和 y_0 两点之间的距离了. 这就产生了矛盾，说明设定不成立，也就是说柯西列不可能有两个收敛点.

完备空间

在度量空间[785] 词条中我们发现，有理数集 $\mathbb{Q}$ 上的柯西列不一定存在极限；而实数集 $\mathbb{R}$ 上的柯西列一定存在极限. 也就是说，在**这个**度量下，$\mathbb{R}$ 可以看成 $\mathbb{Q}$ 中添加了所有收敛点构成的集合.

像 $\mathbb{R}$ 这样包含了自身所有柯西列的收敛点的集合，被称为**完备集（Complete Set）**. 包含了自身所有柯西列的收敛点的度量空间，被称为**完备（度量）空间（Complete Metric Space）**.

考虑到一个柯西列唯一对应一个收敛点，我们也可以把柯西列本身看成一个点. 那么，一个不完备的度量空间 $\mathcal{M}$，也可以通过把所有柯西列作为一个点添加到这个空间里来得到一个完备的空间. 如果一个柯西列在 $\mathcal{M}$ 中已经有收敛点了，那么添加这个柯西列相当于添加了已经存在的收敛点，也就是什么

都没添加；如果一个柯西列在 $\mathcal{M}$ 中没有收敛点，那就把这个柯西列本身当作一个点添加进去.

注意，柯西列本身是 $\mathcal{M}$ 的一个子集，它不是 $\mathcal{M}$ 中的一个点. 我们把柯西列等价于其收敛点后，就可以把它看成点来添加进 $\mathcal{M}$ 中了. 添加缺失的收敛点以得到完备空间的过程，叫做**度量空间的完备化**.

对于这些新添加的点，它的度量应该是什么呢？

定义 3　完备化度量

设 $\{a_n\}$ 是度量空间 $\mathcal{M}$ 中的柯西列，其在空间中没有收敛点. 那么添加其收敛点并命名为 a_0，对于 $\mathcal{M}$ 中任意一点 x，定义度量 $d(x,a_0)$ 为 $\lim\limits_{x,a_n}$.

可以看到，新的点和其它点的度量是用数列极限定义的. 感兴趣的读者可以自行验证，对于柯西点列，这样的极限一定存在.

范数赋范空间

预备知识 矢量空间[444]，度量空间[785]

范数（norm）可以看作几何矢量[370] 的模长在一般矢量空间上的拓展.

定义 1

设 X 是实数或复数域上的矢量空间. X 上的范数是满足如下条件的非负函数 $\|\cdot\|$：

1. $\|x\| \geqslant 0$ （正定）
2. $\|x\| = 0$ 当且仅当 $x = 0$
3. $\|\lambda x\| = |\lambda|\|x\|$
4. $\|x_1 + x_2\| \leqslant \|x_1\| + \|x_2\|$ （三角不等式）

如果一个矢量空间中定义了范数，我们就把它称为赋范空间（**normed space**）.

赋范线性空间都具有度量

$$d(x,y) := \|x-y\| \tag{1}$$

所以赋范空间都是度量空间[785]．作为度量空间时完备的赋范空间称为**巴拿赫空间（Banach space）**[801]．

一个线性空间上可能可以定义许多个范数．线性空间 X 上的两个范数 $\|\cdot\|_1$ 和 $\|\cdot\|_2$ 称为**等价（equivalent）**的，如果有正实数 $C>1$ 使得如下不等式对于任何 $x\in X$ 都成立：

$$C^{-1}\|x\|_1 \leq \|x\|_2 \leq C\|x\|_1.$$

内积空间（inner product space）[471] 是非常重要的特殊的赋范空间．如果 H 是内积空间，$\langle\cdot,\cdot\rangle \to \mathbb{C}$ 是其上的内积，则若命 $\|x\| = \sqrt{\langle x,x\rangle}$，$H$ 便成为一个赋范空间．

有限维空间上的范数

设 $p\geq 1$．定义 $\mathbb{R}^N$ 或 $\mathbb{C}^N$ 空间（即 N 维实数或复数列矢量空间）的 **p-范数**为

$$\|x\|_p = \left(\sum_{i=1}^{N}|x_i|^p\right)^{1/p} \tag{2}$$

物理中常见的是 **2-范数**，也叫**欧几里得范数（Euclidean norm）**即

$$\|x\|_2 = \sqrt{|x_1|^2+|x_2|^2+\cdots+|x_N|^2} \tag{3}$$

它是由内积

$$\langle x,y\rangle = \sum_{i=1}^{N} x_i\bar{y}_i$$

诱导的．

在极限 $p\to\infty$ 之下，绝对值最大的 x_i 对求和的贡献将远大于其他分量，所以可定义**无穷范数（infinity norm）**为

$$\|\mathbf{x}\|_\infty = \max\{|x_i|\}. \tag{4}$$

除此之外，有限维实或复线性空间上还可以定义许多种不同的范数．不过，有限维实或复线性空间上的任意两个范数必然彼此等价．它们都给出空间上唯一的一个自然拓扑（即使得所有线性泛函均连续的拓扑）．在任何范数之下，有限维实或复线性空间都是巴拿赫空间[801]．

赋范空间中的极限

既然赋范空间属于度量空间，那么我们可以延用度量空间中序列极限的定义．只需要令距离函数 $d(x,y)=\|x-y\|$ 即可．

在赋范空间中，极限

$$\lim_{n\to\infty} x_n = x \tag{5}$$

的定义是

$$\lim_{n\to\infty} \|x_n - x\| = 0 \tag{6}$$

函数的范数

欧氏空间上的函数可以定义多种范数．对于函数 f，最常见的范数有极大范数

$$\|f\|_{L^\infty} := \sup_{x\in\mathbb{R}^N} |f(x)|;$$

还有 L^p 范数（这里 $p\geq 1$）：

$$\|f\|_{L^p} := \left(\int_{\mathbb{R}^N} |f(x)|^p dx\right)^{1/p}.$$

特别地，量子力学中经常要考虑的范数是 L^2 范数，它是由内积

$$\langle f,g\rangle = \int_{\mathbb{R}^N} f(x)\bar{g}(x)dx$$

诱导的．这些定义都可以推广到一般的拓扑空间或测度空间上．

赋范空间上的线性算子

两个赋范线性空间 $(X,\|\cdot\|_X),(Y,\|\cdot\|_Y)$ 之间的线性算子 T 若满足

$$\|Tx\|_Y \leq C\|x\|_X$$

则称为是**有界的（bounded）**；使得上面不等式成立的 C 的下确界称作算子 T 的**算子范数（operator norm）**，常记为 $\|T\|_{X\to Y}$，不至于混淆时也可简略记为 $\|T\|$．赋范线性空间之间的有界线性算子与赋范线性空间之间的连续线性算子是同一个数学对象．由线性性质，

$$\|T\|_{X\to Y} = \sup_{\|x\|_X\leq 1} \|Tx\|_Y.$$

从 X 到 Y 的有界线性算子的集合常记为 $\mathfrak{B}(X,Y)$. 若赋予算子范数，则它也是一个赋范线性空间.

如果存在一有界线性算子 $T:X\to Y$，使得 T 是双射，且其逆映射也是有界的，则称赋范线性空间 X,Y 是**同构的（isomorphic）**，算子 T 称为两者之间的同构映射（isomorphism）. 此时 T 既是线性空间范畴下的可逆算子，也是两个度量空间之间的同胚. 因此，任何拓扑性质，例如完备性，可分性等等，在同构映射下都不变.

如果有 $\|Tx\|_Y=\|x\|_X$，则称 T 是赋范线性空间之间的**等距（isometry）**. 如果既是同构映射也是等距映射，则称为等距同构（isometric isomorphism），此时两赋范线性空间则称为等距同构的. 注意，对于无穷维线性空间来说，一个空间到自己的单线性映射不一定是满的，这与有限维线性空间截然不同. 例如在平方可和序列空间 l^2 中，平移映射

$$(x_1,x_2,x_3,...)\to(0,x_1,x_2,x_3,...)$$

是单的，而且甚至是等距映射，但它不是满的.

赋范空间上的结构

给定两个赋范线性空间 $(X,\|\cdot\|_X),(Y,\|\cdot\|_Y)$ 之后，其直积 $X\times Y$ 上可定义一些相互等价的范数而使之成为赋范线性空间，例如

$$\|(x,y)\|_1=\|x\|_X+\|y\|_Y,\qquad \|(x,y)\|_\infty=\max(\|x\|_X,\|y\|_Y).$$

更一般地，可定义

$$\|(x,y)\|_p=\left(\|x\|_X^p+\|y\|_Y^p\right)^{1/p},\ 1\le p\le\infty$$

它们上述两个范数都等价.

如果 X 是赋范线性空间，则其子空间也是赋范线性空间. 如果 $M\subset X$ 是的闭子空间，则商空间 X/M 上可定义范数为

$$\|x+M\|:=\inf_{m\in M}\|x+m\|.$$

此时 X/M 也成为赋范线性空间. 赋范线性空间之间的有界线性算子 $T:X\to Y$ 可以被分解为

$$T=T_1\circ\pi,\quad T:X\xrightarrow{\pi}X/\operatorname{Ker}(T)\xrightarrow{T_1}Y,$$

其中 $\pi : x \to x + \mathrm{Ker}(T)$ 是商投影，$T_1 : x + \mathrm{Ker}(T) \to T(x)$ 是商映射，它是到像空间 $\mathrm{Ran}(T)$ 的一一映射，但逆映射却不一定有界．关于有界算子的像空间的性质，详见巴拿赫定理[803]．

对于赋范线性空间 X 的真闭子空间 M，成立**里斯引理（Riesz's lemma）**：任给 $\alpha \in (0,1)$，都存在 x 使得 $\|x\| = 1$，且 $\mathrm{dist}(x, M) \geq \alpha$．详见[798]．

里斯引理（赋范空间）

- 本词条需要更多参考文献．

预备知识 范数[794]

里斯引理（Riesz's lemma）是赋范空间理论中的一个基本引理．它的表述如下：

引理 1　里斯引理 (Riesz's lemma)

设 X 是赋范空间，$\|\cdot\|$ 是它的范数，M 是它的真闭子空间．则任给 $\alpha \in (0,1)$，都存在 x 使得 $\|x\| = 1$，且 $\mathrm{dist}(x, M) \geq \alpha$．

这个引理有鲜明的直观意义：任给赋范空间中的“平面”，总有一个单位向量与这平面“几乎垂直”．

证明

证明是直接的构造．任取 $y \in X \setminus M$，命 $\delta = \mathrm{dist}(y, M)$．由于 M 是闭的，故 $\delta > 0$．根据距离的定义，任给 $\varepsilon > 0$，皆存在 $z_0 \in M$ 使得

$$\delta \leq \|y - z_0\| < \delta + \varepsilon.$$

命 $x = (y - z_0)/\|y - z_0\|$，则 $\|x\| = 1$，而且对于任何 $z \in M$ 皆有

$$\|x - z\| = \left\| \frac{y - z_0}{\|y - z_0\|} - z \right\| = \left\| \frac{y - (z_0 + z\|y - z_0\|)}{\|y - z_0\|} \right\|.$$

由于 $z_0 + z\|y - z_0\| \in M$，所以上式右边大于

$$\frac{\delta}{\delta + \varepsilon}.$$

于是对于给定的 $0<\alpha<1$，只要取 ε 足够小即可使得

$$\operatorname{dist}(x,M)=\inf_{z\in M}\|x-z\|>\frac{\delta}{\delta+\varepsilon}>\alpha.$$

应用举例

如果 X 是无穷维的赋范空间，那么存在递增的子空间族

$$M_1\subset M_2\subset M_3\subset\ldots,$$

使得 $\dim(M_k)=k$．有限维的子空间都是闭子空间，所以根据里斯引理，存在 $x_k\in M_k\setminus M_{k-1}$ 使得

$$\|x_k\|=1,\ \operatorname{dist}(x_k,M_k)>\frac{1}{2}.$$

这样一来，恒有 $\|x_m-x_n\|>1/2$ 成立，于是 X 中闭单位球面上的序列 $\{x_n\}$ 没有收敛子序列．这表示**无穷维赋范空间中的闭单位球不可能是紧的.**

投影算符

预备知识 内积[471]

本词条只介绍有限维或可数维矢量空间的投影算符.

定理 1　正交分解

令 M 为 N 维矢量空间 X 的子空间．那么任意 $u\in X$ 有唯一的分解

$$u=v+w\qquad(v\in M,w\in M^{\perp})\tag{1}$$

其中 $M^{\perp}$ 是 M 在 X 中的正交补.

对每个子空间 $M\subseteq X$，我们定义对应的**投影算符** P，根据式 1 将每个 $u\in X$ 映射到 v.

定理 2

令投影算符将 N 维空间 X 中的矢量投影到其子空间 A 中，$\{a_i\}$ 是 A 的一组正交归一基底．那么投影算符可以表示为

$$P=\sum_i|a_i\rangle\langle a_i|\tag{2}$$

由该定理易证定理 1（将 u 分别分解到 M 和 $M^\perp$ 的正交归一基底上，这个投影存在且唯一）.

投影算符是厄米算符

投影算符 P 是厄米算符（也叫自伴算符），即对任意 $u,v \in X$ 满足 $\langle u|Pv\rangle = \langle Pu|v\rangle$.

投影算符有 0 和 1 两个本征值. 对应的本征矢分别是 M 和 $M^\perp$ 空间中的所有矢量.

证明

对任意 $u,v \in X$，有

$$\langle u|Pv\rangle = \sum_i braketua_i \langle a_i|v\rangle \tag{3}$$

$$\langle Pu|v\rangle = \overline{\langle v|Pu\rangle} = \sum_i \overline{braketva_i}\overline{\langle a_i|u\rangle} = \sum_i braketua_i \langle a_i|v\rangle = \langle u|Pv\rangle \tag{4}$$

证毕.

柯西—施瓦茨不等式

$\mathbb{C}^N$ 空间

对 $\mathbb{C}^N$ 空间，有

$$\left|\sum_i^N u_i^* v_i\right|^2 \leqslant \sum_j |u_j|^2 \sum_k |v_k|^2 \tag{1}$$

内积空间

预备知识 内积[471]

内积空间中任意两个矢量满足柯西不等式

$$|\langle u|v\rangle|^2 \leqslant \langle u|u\rangle \cdot \langle v|v\rangle \tag{2}$$

证明

我们使用勾股定理来证明．令 u 在 v 方向的投影为 x，在 v 垂直方向的投影为 y，即

$$x=\frac{\langle v|u\rangle}{\langle v|v\rangle}v\qquad y=u-x \tag{3}$$

容易证明 $u=x+y$ 且 $\langle x|y\rangle=0$．这样就可以使用勾股定理式 8 [471]

$$\langle u|u\rangle=\langle x|x\rangle+\langle y|y\rangle\geqslant\langle x|x\rangle=\frac{|\langle u|v\rangle|^2}{\langle v|v\rangle} \tag{4}$$

两边乘以 $\langle v|v\rangle$，得式 2．特殊地，不难证明当且仅当 u,v 共线时 $y=0$，即 $\langle y|y\rangle=0$，不等式取等号．证毕．

巴拿赫空间

预备知识 柯西序列完备度量空间[790]

定义 1　巴拿赫空间

完备的赋范空间叫做巴拿赫空间（**Banach space**）．

巴拿赫空间的例子

例 1　$\mathbb{C}^N$ 空间

N 维实空间 $\mathbb{R}^N$ 或者复空间 $\mathbb{C}^N$ 在任何范数之下都是完备的，因此在任何范数之下都是巴拿赫空间．

例 2　连续函数空间

有限闭区间上的所有连续函数构成的空间记为 $X:=C[a,b]$，这是一个不可数维的空间．若令范数为

$$\|f\|:=\max_{a\leqslant x\leqslant b}|f(x)|,$$

则它成为一个巴拿赫空间．这范数下收敛的连续函数序列恰为一致收敛的连续函数序列．而若赋予 X 以 L^p 范数

$$\|f\|_{L^p}:=\left(\int_{\mathbb{R}^N}|f(x)|^p dx\right)^{1/p},$$

则在此范数下它不是完备的.

更一般地，对于任何紧 Hausforff 空间 K，连续函数空间 $C(K)$ 在范数

$$\|f\| := \sup_{x\in K} |f(x)|$$

之下也是巴拿赫空间.

例 3　L^p 空间

对于测度空间 $(\Omega, \mathcal{A}, \mu)$，定义可测函数的 L^p 范数为

$$\|f\|_{L^p(\mu)} = \left(\int_\Omega |f(x)|^p d\mu(x)\right)^{1/p},$$

$$\|f\|_{L^\infty(\mu)} = \text{ess sup}_\Omega |f|.$$

如果将几乎处处相等的函数视为相同，则当 $1 \le p \le \infty$ 时 $\|\cdot\|_{L^p(\mu)}$ 便是一个范数，而满足 $\|f\|_{L^p(\mu)}$ 的可测函数的线性空间就是 $L^p(\mu)$. 它是完备的. 当 $\Omega = \mathbb{N}$，而测度 μ 为普通的计数测度时，可测函数就是通常的序列，此时将空间记为 l^p，而序列的范数是

$$\|x\|_p = \left(\sum_{n=1}^{\infty} |x(n)|^p\right)^{1/p}.$$

本例内容可参考 Bogachev, V. I. (2007). Measure theory (Vol. 1). Springer Science & Business Media., 第三章.

巴拿赫空间上的线性算子

巴拿赫空间上有界算子的定义与赋范空间中一样. 不过，在分析数学中，也常常需要考虑不一定有界的线性算子. 一般来说，对于两个巴拿赫空间 X, Y，分析数学中最经常考虑的是**稠定算子（densely defined operator）**，即定义在 X 的某个稠密子空间上的线性算子 T. 这个稠密子空间称为算子的定义域，常记为 $\text{Dom}(T)$. 稠定算子不一定可以延拓为全空间上的有界算子. 如果算子 T_1 是算子 T 的延拓，即 $\text{Dom}(T) \subset \text{Dom}(T_1)$ 且在 $\text{Dom}(T)$ 上有 $T_1 = T$，则写 $T \subset T_1$.

对于巴拿赫空间 X, Y 之间的（不一定有界的）线性算子 T，其图像（graph）定义为 $X \times Y$ 的子空间

$$\text{G}(T) := \{(x, Tx) : x \in \text{Dom}(T)\}.$$

在 Dom(T) 上定义的范数 $[\|x\|]_T := \|x\|_X + \|Tx\|_Y$ 称为图范数（graph norm）. 如果 G(T) 是闭子空间，则称算子 T 为**闭算子（closed operator）**. 如果存在延拓 $T_1 \supset T$ 使得 T_1 为闭算子，则 T 称为**可闭化（closable）**的. 闭图像定理（Closed graph theorem）[803] 说明：在巴拿赫空间上，定义于全空间的闭算子是有界的. 谱理论就是针对闭算子展开的.

巴拿赫空间的基本性质

赋范线性空间在度量空间的意义下经过完备化之后成为巴拿赫空间. 完备的赋范空间在赋以等价范数后还是完备的. 可分的赋范线性空间的完备化空间还是可分的.

一个赋范空间 $(X, \|\cdot\|)$ 是完备的，当且仅当由 $\sum_{n=1}^{\infty} \|x_n\| < \infty$ 总可以推出 $\sum_{n=1}^{\infty} x_n$ 按照范数收敛.

如果 $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ 是巴拿赫空间，那么其直积 $X \times Y$ 在范数 $\|(x,y)\|_p$ 之下也是巴拿赫空间. 有界线性算子空间 $\mathfrak{B}(X,Y)$ 在算子范数下也是巴拿赫空间. 如果 X 巴拿赫空间，$M \subset X$ 是其闭子空间，则商空间 X/M 在商范数之下也是巴拿赫空间. 这里用到的定义见[794].

无穷维巴拿赫空间的代数维数（即其中极大线性无关向量组的势）一定是不可数的. 这是因为，如果 $\{x_k\} \subset X$ 是任何序列，则若命 X_n 是 $\{x_1, ..., x_n\}$ 张成的子空间，那么每个 X_n 都是真闭子空间. 按照贝尔纲定理，它们的并集是第一纲集，但巴拿赫空间作为完备度量空间是第二纲集，所以不可能有 $X = \bigcup_n X_n$.

无穷维巴拿赫空间中的闭单位球 $\{\|x\| \leq 1\}$ 在其范数拓扑下一定不是紧的. 详见里斯引理[798].

巴拿赫定理

预备知识 巴拿赫空间[801]

在巴拿赫空间的理论中，**巴拿赫定理（Banach theorems）**是一组关于巴拿赫空间上的有界线性算子的定理. 它们都依赖于巴拿赫空间的完备性，并且在分析数学中有广泛的应用. 可参考 Kôsaku Yosida: Functional analysis. Grundlehren

der mathematischen Wissenschaften 123, Springer-Verlag, 1980 (6th ed.).

一致有界原理（共鸣定理）

定理 1　一致有界原理（共鸣定理，Uniform Boundedness Principle）

设 X 是巴拿赫空间，Y 是赋范线性空间，$\mathfrak{F} \subset \mathfrak{B}(X,Y)$ 是有界线性算子的族．如果对于任何 $x \in X$ 皆有 $\sup_{T\in\mathfrak{F}} \|Tx\|_Y < \infty$，那么实际上必有

$$\sup_{T\in\mathfrak{F}} \|T\|_{\mathfrak{B}(X,Y)} = \sup_{T\in\mathfrak{F}} \left(\sup_{\|x\|_X \le 1} \|Tx\|_Y \right) < \infty.$$

也就是说，从巴拿赫空间出发的线性算子的族，如果是逐点有界的，则一定是一致有界的．

这个定理的证明用到了贝尔纲定理（Baire category theorem），而这依赖于 X 的完备性．如果 X 不完备，则定理不成立．一个反例如下：取 X 为仅有有限项非零的序列组成的空间，并赋以范数 $\|x\| = \max_k |x_k|$．这个空间不是完备的．定义 $T_n : X \to X$ 为

$$T_n x = (x_1, 2x_2, ..., nx_n, 0, ...),$$

则 $\|T_n\| = n$，因此族 $\{T_n\}$ 不是一致有界的，但对于任何固定的 $x \in X$，当 n 充分大时，$T_n x$ 将恒等于常值，因此对于任何固定的 $x \in X$，都有 $\sup_n \|T_n x\| < \infty$．

该定理的一个应用是说明：连续函数的傅里叶级数不必逐点收敛．

开映像原理

定理 2　开映像原理（Open Mapping Theorem）

设 X, Y 是巴拿赫空间，$T : X \to Y$ 是连续线性算子，而且是满映射．则 T 是开映射，即 X 中的开集经 T 变为 Y 中的开集．

这个定理的证明也用到了贝尔纲定理（Baire category theorem），而这依赖于 X, Y 的完备性．如果 X 不完备，则定理不成立．实际上有很简单的反例可以作为说明：如果 $Z \subset X$ 是稠密的真子空间，那么自然的嵌入映射 $Z \to X$ 就显然不是开映射．

开映像原理有如下的直接推论：如果 T 是巴拿赫空间 X, Y 之间的连续线性算子，且值域 $\mathrm{Ran}(T)$ 是 Y 的闭子空间，则商映射 $X/\mathrm{Ker}(T) \to \mathrm{Ran}(T)$ 是同

构．另外，开映像原理还说明：一个线性空间上的两个范数若都是完备的，则这两个范数必然等价．

闭图像定理

定理 3　闭图像定理（Closed Graph Theorem）

设 $T : X \to Y$ 是巴拿赫空间之间的闭算子．则 T 是连续的．

这个定理实际上可作为开映像原理的推论：如果 $\mathrm{G}(T) \subset X \times Y$ 是闭子空间，则 X 在图范数 $[|x|]_T = \|x\|_X + \|Tx\|_Y$ 下成为巴拿赫空间．但巴拿赫空间之间的映射 $I : (X, [|\cdot|]_T) \to (X, \|\cdot\|_X)$, $I(x) = x$ 是连续的单满射，从而根据开映像原理它也是开映射，因此其逆映射是连续的．这表示 T 是有界算子．

一致收敛

- 本词条处于草稿阶段．

一个函数列 $f_n(x)$ 一致收敛到 $f(x)$ 的定义是：当对于任意 $\epsilon > 0$，存在 N，当 $n \geqslant N$ 时对任意 x 都有

$$|f_n(x) - f(x)| < \epsilon \tag{1}$$

或者记为

$$\lim_{n\to\infty} (\max |f_n(x) - f(x)|) = 0 \tag{2}$$

黎曼积分与勒贝格积分

- 本词条处于草稿阶段．
- 本词条缺少预备知识，初学者可能会遇到困难．

[3]我们来定义区间 $[a, b]$ 的实函数 $f(x)$ 的黎曼积分．令有序实数列

$$\mathcal{P} = \{a = x_0 < x_1 < \cdots < x_n = b\} \tag{1}$$

[3]参考这个讲义．

令

$$m_k = \inf_{x_{k-1}\le x\le x_k} f(x) \qquad M_k = \sup_{x_{k-1}\le x\le x_k} f(x) \quad (1\le k\le n) \tag{2}$$

对应的下和上黎曼求和为

$$\underline{I}(f,\mathcal{P}) = \sum_{k=1}^{n} m_k \Delta k \qquad \bar{I}(f,\mathcal{P}) = \sum_{k=1}^{n} M_k \Delta_k \tag{3}$$

其中 $\Delta_k = x_k - x_{k-1}$，$1\le k\le n$. 下和上黎曼积分的定义为

$$\underline{I}(f) = \sup_{\mathcal{P}} \underline{I}(f,\mathcal{P}) \qquad \bar{I}(f) = \inf_{\mathcal{P}} \bar{I}(f,\mathcal{P}) \tag{4}$$

显然，$\underline{I}(f)\le \bar{I}(f)$. f 叫做**黎曼可积（Riemann integrable）**当且仅当上下黎曼积分相等

$$I(f) = \underline{I}(f) = \bar{I}(f) \tag{5}$$

我们以下用 $I(f)$ 表示黎曼积分，$\int$ 符号表示勒贝格积分.

定理 1

任何连续函数 $f\in C[a,b]$ 都是黎曼可积的.

定理 2

任何连续函数 $f\in C[a,b]$, 勒贝格积分和黎曼积分相等.

定理 3

令 f 为 $[a,b]$ 上的有界实函数，那么 f 是黎曼可积的当且仅当 f 的所有不连续点的勒贝格测度为零.

这样的函数可以在勒贝格测度为零的集合上修改，使其 Borel-measurable，若这样做，勒贝格积分和黎曼积分结果相等.

数学分析笔记

- 本词条缺少预备知识，初学者可能会遇到困难.
- 这是一个总结，应该放到所有相应内容之后，以及给出详细词条的链接

本文参考 [8].

Chap 1. 实数系和复数系

- **有理数（rational number）**记为 Q，实数记为 R
- 虽然任意两个不同的有理数间还有一个有理数，但是有理数集中还是会有“间隙”，而实数集填补了这些间隙.
- **集合（set）：属于（in）** $x \in A$，**不属于（not in）** $x \notin A$
- **空集（empty set）**，**非空（none empty）**，**子集（subset）** $A \subseteq B$，**超集（superset）** $B \supseteq A$，**真子集（proper subset）**
- **有序集（ordered set）**，任意不相等的两个元可以比较大小
- **有上界（bounded above）**：任意元小于等于超集中的某个元. **上界（upper bound）**
- **最小上界（least upper bound, supremem）**：$\alpha = \sup E$；**最大下界（greatest lower bound，infimum）**：$\alpha = \inf E$
- 如果对于任意非空有上界的 $E \subset S$，都有 $\sup E \in S$，那 S 就具有 **upper bound property**
- **域（field）**集合 F 定义了**加法**和**乘法**. 加法满足：闭合性，交换律，结合律，存在 0 元，存在逆元. 乘法满足：闭合性，交换律，结合律，存在单位元，存在倒数. 加法和乘法满足分配律.
- 有理数集是一个域
- **有序域（ordered field）**
- 存在一个有序域 R 具有 upper bound property，且有理数集 Q 是其子集. R 就是实数.
- 实数的**阿基米德性质**：存在整数 n 使 $nx > y$ （$x > 0$）
- $x \in R$，$x > 0$，n 为整数，存在实数 y 使 $y^n = x$
- **稠密（dense）**：两个不同的实数间必有一个有理数

- **extended real number system** 是在实数集基础上加入 $\pm\infty$ 两个符号．对任何实数有 $-\infty < x < +\infty$．所有非空子集都有最小上界和最大下界．相比于无穷，实数集中的元被称为 **finite**．
- 复数是一对有序实数 (a, b)，定义了加法和乘法后，就变成了一个域．定义 $\mathrm{i} = (0, 1)$．
- 对正整数 k，R^k 定义为所有 k 个有序实数的集合 $\mathbf{x} = (x_1, \ldots, x_k)$，其中 x_i 叫做坐标．
- 定义 R^k 中的内积为 $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{k} x_i y_i$
- 定义模长为 $|x| = (\mathbf{x} \cdot \mathbf{x})^{1/2}$
- 定义了内积和模长的 R^k 被称为**欧几里得 k 空间（euclidean k-space）**．这也是一个**度量空间（metric space）**（见下文）．通常 R^1 叫做线，R^2 叫做面

Chap 2. 基本拓扑

- **函数**是两个集合 A，B 之间的**映射**；定义域，值域，值．$f(A)$ 就是集合 A 的 **image**；$f(A) \subset B$．如果 $f(A) = B$，那么 f 把 A 映射到（onto）B．$f^{-1}(E)$ **inverse image**．**1-1 映射**
- 如果 A 到 B 存在 1-1 映射，记为 $A \sim B$：**reflexive** $A \sim A$，**symmetric** $A \sim B \to B \sim A$，**transitive** $A \sim B, B \sim C \to A \sim C$
- 定义 J_n 为集合 $1, 2, \ldots, n$，定义 J 为 $1, 2, \ldots$
- $A \sim J_n$ 为**有限（finite）**，不是有限就是**无限（infinite）**，$A \sim J$ 为**可数（countable）**[4]．**至多可数（at most countable）**就是有限或者可数
- 对无限集来说，“含有有限个元”很模糊，但是 1-1 映射的定义仍然有效（只要写出一个表达式）
- 有限集不可能与真子集等效，而无限集可以

[4] 也叫 enumerable 或者 denumerable

- 数列是 J 的映射 $f(n)=x_n$，记为 $\{x_n\}$．x_n 叫做一项．如果 $x_n\in A$，那该序列就叫 A 中（元素）的序列．
- 可数集的无穷子集仍然是可数的
- **Sequence of set** $\{E_\alpha\}$，每个 $\alpha\in A$ 都对应一个 E_α（其实就是 set of sets，但不这么叫）
- **并集（Union）**：$\bigcup\limits_{m=1}^{n}E_m$，**交集（intersection）**：$\bigcap\limits_{m=1}^{n}E_m$
- 并集和交集的混合运算法则和加法和乘法差不多
- 如果 E_n $(n=1,2\dots)$（无穷多个）是可数的，那么它们的 union 仍然是可数的
- 如果 A 是 at most countable，且对 $\alpha\in A$，B 也是 at most countable，那么 $T=\bigcup_{\alpha\in A}B_\alpha$ 也是 at most countable
- 如果 A 可数，$a_i\in A$，而 B_n（n 为固定的正整数）是所有 $(a_1,\dots,a_n)$ 的集合，那么 B_n 也是可数的
- 自然数和有理数（可以看作两个有序整数）都可数，无理数不可数
- 如果集合 X 中的元素可以叫做**点（point）**，如果一个值为实数的函数 $d(p,q)$, $p\in X$, $q\in X$ 满足：当 $p=q$，$d(p,q)=0$，当 $p\neq q$，$d(p,q)>0$，$d(p,q)=d(q,p)$，$d(p,q)\leqslant d(p,r)+d(r,q)$，$r\in X$，我们就说这是一个**度量空间（metrix space）**，函数 d 叫做**距离函数（distance function）**，或者**度规（metric）**
- **segment** (a,b) 是所有 $a<x<b$ 的实数，**interval** $[a,b]$ 是所有 $a\leqslant x\leqslant b$ 的实数．
- interval 也叫 1-方格，类似地，R^k 中可以定义 **k-方格（k-cell）**，2-方格是长方形
- 类似地，R^k 空间也可以定义 **开/闭球（open/closed ball）**

- **convex**：$E \subset R^k$ 对任意 $0 < \lambda < 1$ 满足 $\lambda\mathbf{x} + (1-\lambda)\mathbf{x} \in E$．例如，ball 和 k-方格都是 convex 的.
- 度量空间中，**邻域（neighborhood）**N_r：到某点距离小于 r 的集合（$r > 0$）
- **极限点（limit point）** p：所有邻域存在一个与 p 不同的点（无论半径有多小）
- 如果不是极限点，那就是 **孤立点（isolated point）**
- 如果所有极限点都属于集合，这个集合就是**闭（closed）**的.
- 如果关于点 p 的某个邻域是集合 E 的子集，p 就是 E 的**内点（interior point）**
- 如果 E 中的任意一点都是内点，E 就是 **开（open）**的
- **补集（complement）**
- 如果一个闭集合中每一点都是它的极限点,那么该集合就是**完全（perfect）**的
- 如果集合中任意一点都在某个 r 为实数的邻域内，这个集合就是 **有界的（bounded）**
- 集合 E 在集合 X 上**稠密（dense）**：X 中任意一点都是 E 的一个极限点或者 E 中的一点.（例如有理数在实数上稠密）
- 任何邻域都是开的
- 如果 p 是一个极限点，那么它的任何邻域都有无限多个点
- 有限个点的集合没有极限点
- $(\bigcup_\alpha E_\alpha)^c = \bigcap_\alpha (E_\alpha^c)$ 其中 c 代表补集
- 集合 E 是开的当且仅当它的补集是闭的．E 是闭的当且仅当它的补集是开的.
- 任意多开集合的并集仍然是开的，任意多闭集合的交集仍然是闭的

- 有限个开集合的交集仍然是开的，有限个闭集合的并集仍然是闭的
- 设 X 是度量空间，如果 $E \subset X$，E' 表示 E 在 X 中所有极限点组成的集．那么，把 $\bar{E} = E \cup E'$ 叫做 E 的**闭包（closure）**
- 设 X 是度量空间，而 $E \subset X$，那么 (a) $\bar{E}$ 是闭的，(b) $E = \bar{E}$ 当且仅当 E 闭，(c) 如果闭集 $F \subset X$ 且 $E \subset F$，那么 $\bar{E} \subset F$．由 (a) 和 (c)，E 是 X 中包含 E 的最小闭子集
- 设 E 是一个不空实数集，上有界，令 $y = \sup E$，那么 $y \in \bar{E}$．因此，如果 E 闭，那么 $y \in E$.
- 令 $Y \subset X$，$E \subset Y$，E 是开的当且仅当 $E = Y \bigcap G$，对某个 $G \subset X$
- 若 X 的一组开子集 $\{G_\alpha\}$ 使 $E \subset \bigcup_\alpha G_\alpha$，那么 $\{G_\alpha\}$ 就是 E 的**开覆盖（open cover）**
- **紧集（compact set）:** 如果 $\{G_\alpha\}$ 是 K 的开覆盖，那么存在有限个 $\alpha_1, \ldots, \alpha_n$ 使得 $K \subset G_{\alpha_1} \bigcup \cdots \bigcup G_{\alpha_n}$
- 有限集都是紧集
- 如果 $E \subset Y \subset X$，那么 E 可能是 Y 中的开区间而不是 X 的开区间.
- 假设 $K \subset Y \subset X$. 那么 K 在 X 中是紧的当且仅当它在 Y 中也是紧的.
- 度量空间的紧子集是闭的
- 紧集的闭子集也是紧的
- 如果 F 是闭的且 K 是紧的，那么 $F \bigcap K$ 是紧的
- 如果 $\{K_\alpha\}$ 是度量空间 X 的一组紧子集且任意有限个 $\{K_\alpha\}$ 的交集为非空，那么 $\bigcap K_\alpha$ 也是非空的
- 如果 E 是紧集 K 的无穷子集，那么 E 在 K 中存在极限点
- 如果 $\{I_n\}$ 是 R^1 中的一组区间，使得 $I_n \supset I_{n+1} (n = 1, 2, 3, \ldots)$，那么 $\bigcap_1^\infty I_n$ 非空

- k-方格是紧的
- 对 R^k 中的集合 E，这三个条件等价：(a) E 闭且有界．(b) E 是紧的．(c) E 中的任意无限集在 E 中存在极限点
- R^k 中任何有界的无限集在 R^k 中有（至少）一个极限点
- 令 P 为 R^k 内的非空完全集．那么 P 是不可数的
- The **Cantor set** shows that there exist perfect sets in R^1 which contain no segment.
- Two subsets A and B of a metric space X are said to be **separated** if both $A \cap \bar{B}$ and $\bar{A} \cap B$ are empty, i.e., if no point of A lies in the closure of B and no point of B lies in the closure of A.
- A subset E of the real line R^1 is connected if and only if it has the following property: If $x \in E$, $y \in E$, and $x < z < y$, then $z \in E$.

Chap 3. 数列与级数

- A sequence $\{p_n\}$ in a metric space X is said to converge if there is a point $p \in X$ with the following property: For every $\epsilon > 0$ there is an integer N such that $n \geqslant N$ implies that $d(p_n, p) < \epsilon$. (Here d denotes the distance in X.) In this case we also say that $\{p_n\}$ converges to p, or that p is the limit of $\{p_n\}$ and we write $p_n \to p$, or $\lim_{n\to\infty} p_n = p$.
- the set of all points p_n is the range of $\{p_n\}$. The range of a sequence may be a finite set, or it may be infinite. The sequence $\{p_n\}$ is said to be bounded if its range is bounded.

未完成：以下所有内容

Chap 4. 连续性

Chap 5. 微分法

Chap 6. Riemann-Stieltjes 积分

Chap 7. 函数序列与函数项级数

Chap 8. 一些特殊函数

Chap 9. 多元函数

Chap 10. 微分形式的积分

Chap 11. Lebesgue 理论

第十部分

泛函分析

第一章

笔记

希尔伯特空间

- 本词条处于草稿阶段.

预备知识 巴拿赫空间[801]

定义完备的内积空间为**希尔伯特空间（Hilbert space）**. 其中“完备”的概念见巴拿赫空间定义.

由于内积可以用于定义范数，希尔伯特空间必定是巴拿赫空间. 希尔伯特空间在巴拿赫空间的基础上增加了内积运算，从而产生了“正交”的概念.

泛函分析笔记 1

Banach 空间

1.1 Linear Spaces and Dimension

- [1]$\mathbb{R}$ 和 $\mathbb{C}$ 分别表示实数域和复数域，$\mathbb{K}$ 表示二者中的一个
- $\mathbb{K}^N$ 表示 N 元 **tuple** $(\xi_1, \xi_2, \ldots, \xi_N)$
- 某区间上的连续函数 $u : [a, b] \to \mathbb{R}$ 可以表示为 $C[a, b]$
- $\mathbb{K}$ **上（over** $\mathbb{K}$**）**的矢量空间（linear space 就是 vector space) 表示只能以 $\mathbb{K}$ 的元乘以某个矢量
- $C[a, b]$ 是无穷维矢量空间

1.2 Normed Spaces and Convergence

- 用 $\|u\| > 0$ 表示**范数（norm）**

[1]本文参考：[9]

- 定义了范数的空间就叫**赋范空间（normed space）**，满足 (1) $\|u\| \geqslant 0$，(2) $\|u\| = 0$ iff $u = 0$，(3) $\|\alpha u\| = |\alpha|\|u\|$，(4) $\|u + v\| \leqslant \|u\| + \|v\|$
- 定义两个矢量之间的**距离（distance）**为 $\|u - v\|$
- 模可以用于定义极限 $\lim_{n\to\infty} u_n = u$ 为 $\lim_{n\to\infty} \|u_n - u\| = 0$，即 u_n 收敛到 u
- 上一条中，(1) u 是唯一的，(2) u_n 是有界的（bounded），(3) $\|u_n\| \to \|u\|$，(4) $u_n + v_n \to u + v$，(5) $\alpha_n u_n \to \alpha u$
- 柯西序列（Cauchy sequence）[790]：对任意 $\varepsilon > 0$，存在 N，当 $n, m \geqslant N$ 就有 $\|u_n - u_m\| < \varepsilon$
- 在赋范空间中，每个收敛序列都是柯西序列

1.3 Banach Spaces and the Cauchy Convergence Criterion

- 赋范空间 X 是 Banach 空间[801] 当且仅当每个柯西数列都收敛
- 在 Banach 空间中，收敛序列都是柯西序列
- 空间 $X := C[a, b]$ 是实数 Banach 空间，定义范数为 $\|u\| := \max_{a\leqslant x\leqslant b} |u(x)|$. $u_n \to u$ 意味着 $\|u_n - u\| = \max_{a\leqslant x\leqslant b} |u_n(x) - u(x)| \to 0$. 也就是 $u_n(x)$ 一致收敛[805] 到 u
- 如果柯西序列 u_n 的子序列 $u_{n'} \to u$，那么 $u_n \to u$
- 若 $\sum_{j=1}^{\infty} \|u_{j+1} - u_j\| < \infty$，那么 u_n 是柯西序列
- 集合 $U_\varepsilon(u_0) := \{u \in X : \|u - u_0\| < \varepsilon\}$ 叫做 u_0 的 ε-**邻域（neighborhood）**

1.4 Open and Closed Sets

- X 的子集 M 是**开集**当且仅当对任意 $u \in M$ 都存邻域是 M 的子集
- X 的子集 M 是**闭集**当且仅当每个序列的极限都属于 M
- X 的子集 M 是 closed 当且仅当 $X - M$ 是开的

1.5 Operators

- M 和 Y 是集合，$u \in M, v \in Y$ 算符 $A : M \to Y$ 代表映射 $v = Au$，其中 M 是**定义域（domain of definition）**，也记为 $D(A)$．**值域（range）**是 $A(M) := \{v \in Y : v = Au, u \in M\}$，也记为 $R(A)$
- A 叫**满射（surjective）**当且仅当 $A(M) = Y$，叫做**单射（injective）**当且仅当 $Au = Av$ 意味着 $u = v$，**双射（bijective）**如果前两者都符合
- 如果 A 是 bijective，存在逆算符 $A^{-1} : Y \to M$，定义为 $A^{-1}v = u$ 当且仅当 $Au = v$
- 算符也叫函数
- $A : M \subseteq X \to Y$ 表示 $A : M \to Y$ 且 $M \subseteq X$，当 $Y = \mathbb{K}$，就把 A 叫做**泛函（functional）**

1.9 Continuity

- **Sequentially continuous**：$\lim_{n\to\infty} u_n = u$（$u \in M$）意味着 $\lim_{n\to\infty} Au_n = Au$
- **continuous operator**：对任意 $u \in M$，以及 $\varepsilon > 0$，存在 $\delta > 0$ 使得 $\|v - u\| < \delta$（$v \in M$）意味着 $\|Av - Au\| < \varepsilon$
- **uniformly continuous operator**：上面的 δ 不取决于 u
- **Lipschitz continuous**：存在 $L > 0$ 使得 $\|Av - Au\| \leq L\|v - u\|$ 对所有 $u, v \in M$ 都成立
- continuous 和 sequentially continuous 是充要条件
- 两个连续算符的复合算符 $C = A \circ B$ 也是连续的
- 两个集合叫 **homeomorphic**（注意不是 homomorphic）当且仅当存在一个算符，满足连续，双射，且逆算符也连续

1.10 Convexity

- **凸集（convex set）**：线性空间中的子集 M 是凸的当且仅当 $u, v \in M$ 且 $0 \leq \alpha \leq 1$ 意味着 $\alpha u + (1-\alpha)v \in M$ （两点之间的连线也属于集合）
- **凸函数**：凸集合上的函数 $f : M \to \mathbb{R}$ 满足 $f(\alpha u + (1-\alpha)v) \leq \alpha f(u) + (1-\alpha)f(v)$ 对所有 $u, v \in M$ 和 $0 \leq \alpha \leq 1$ 都成立
- 赋范空间中的函数 $f(u) := \|u\|$ 是凸函数
- **linear subspace，closed linear subspace**
- 显然所有线性子空间都是凸的
- **linear hull**：$\operatorname{span} M$，**closure**：$\bar{M}$，**convex hull**：$\operatorname{co} M$
- $u \in \operatorname{co} M$：对固定的 $n = 1, 2, \ldots$，$u = \alpha_1 u_1 + \cdots + \alpha_n u_n$，其中 $u_1, \ldots, u_n \in M$，$0 \leq \alpha_1, \ldots, \alpha_n \leq 1$，$\alpha_1 + \cdots + \alpha_n = 1$

1.11 Compactness

- 如果赋范空间的集合 M 满足每个序列都有收敛的子序列，那么 M 就是**相对紧的（relatively compact）**
- 如果赋范空间的集合 M 满足每个序列都有收敛的子序列且收敛到 M 中，那么 M 就是**紧的（compact）**
- 如果存在 $r \geqslant 0$ 使任意 $u \in M$ 都有 $\|u\| \leqslant r$，那么 M 就是**有界的（bounded）**
- M 是紧的当且仅当它是相对紧的且闭合
- 每个紧集都是有界的
- $\mathbb{K}^N$ 上的子集若使用 $\|u\| := |u|_\infty$，那么它是相对紧的当且仅当它是有界的
- **Arzela-Ascoli theorem**：令 $X := C[a, b]$，且 $\|u\| := \max_{a \leqslant x \leqslant b} |u(x)|$．那么若 $M \subseteq X$ 有界且一致连续，那么 M 就是相对紧的

- **Weierstrass 定理：**令 $f : M \to \mathbb{R}$ 为赋范空间中非空紧子集 M 上的连续函数，那么 f 在 M 上存在一个最大值和最小值
- 令 X, Y 为 $\mathbb{K}$ 上的赋范空间，令 $A : M \subseteq X \to Y$ 中 M 为非空紧集，A 为连续算符．那么 A 是一致连续的
- **finite ε net** 对任意 $\varepsilon > 0$,存在有限多个点 $v_1, \ldots, v_J \in M$ 使得 $\min_{1 \le j \le J} \|u - v_j\| \le \varepsilon$ （$\forall u \in M$）
- $A : M \subseteq X \to Y$ 称为**紧的（compact）**当且仅当 A 是连续的且 A 将有界集变换到相对紧集
- 算符 $(Au)(x) := \int_a^b F(x, y, u(y))\,\mathrm{d}y$ 是紧的，其中 $-\infty < a < b < \infty$，$u(x)$ 有界

1.13 The Minkowski Functional and Homeomorphisms

- 赋范空间 X 中的两个范数叫做 **equivalent** 当且仅当存在正数 α 和 β 使得 $\alpha\|u\| \leqslant \|u\|_1 \leqslant \beta\|u\|$ 对所有的 $u \in X$ 成立
- 有限维空间的任意两个范数都是 equivalent 的
- 每个有限维赋范空间都是 Banach space.
- 如果 $u_0, \ldots, u_N$ 满足 $u_1 - u_0, u_2 - u_0, \ldots, u_N - u_0$ 线性无关，那么就说它们在**一般位置（general position）**．这个性质与 $u_0, \ldots, u_N$ 的顺序无关
- N-单纯形（N-simplex）是 $\mathcal{S} := \mathrm{co}\,\{u_0, \ldots, u_N\}$ （co is convex hull），其中 $u_0, \ldots, u_N$ 在一般位置．0-单纯形是一个点
- 单纯形的 **barycenter** 是 $b := \sum_{j=0}^{N} u_j/(N+1)$
- $\mathcal{S}$ 的 k-face 是单纯形中 $k+1$ 个不同 vertices 的 convex hull
- $\mathrm{diam}\, M := \sup_{u,v \in M} \|u - v\|$ is the **diameter** of M, and $\mathrm{dist}(u, M) := \inf_{w \in M} \|u - w\|$ is the **distance** of the point u from the point M

1.14 The Brouwer Fixed-Point Theorem

- 当 M 是有限维赋范空间中紧的，凸的，非空的子集，连续算符 $A : M \to N$ 拥有 **fixed point**

1.15 The Schauder Fixed-Point Theorem

- 紧算符 $A : M \to M$ 有一个 fixed point 如果 M 是 Banach 空间的一个有界的，闭的，突的，非空的子集

1.20 Linear Operators

- **线性算符** $A : L \subseteq X \to Y$ 是线性的当且仅当 $A(\alpha u + \beta v) = \alpha Au + \beta Av$
- 用 $L(X, Y)$ 表示线性连续算符 $A : X \to Y$，X 是 $\mathbb{K}$ 上的赋范空间，Y 是 $\mathbb{K}$ 上的 Banach 空间．$L(X, Y)$ 是 $\mathbb{K}$ 上的 Banach 空间，范数就是算符的范数
- 有限维矢量空间中的线性算符可以表示为矩阵，所有这些算符都是连续的
- 算符的**零空间（null space）**为 $N(A) := \{u \in X : Au = 0\}$
- 线性算符 A 是连续的，当且仅当存在 $c > 0$ 使 $\|Au\| \leqslant c\|u\|$ 对所有 u 都成立．即当且仅当 A 是有界的
- 线性算符是单射的当且仅当 $N(A) = 0$
- 定义线性连续算符 $A : X \to Y$ 的**算符范数（operator norm）**为 $\|A\| := \sup_{\|v\| \leqslant 1} \|Av\|$
- 当 $X \neq 0$，有 $\|A\| := \sup_{\|v\|=1} \|Av\|$

1.21 The Dual Space

- 令 X 为 $\mathbb{K}$ 上的一个赋范空间，一个线性的连续算符 $f : X \to \mathbb{K}$ 称为**线性连续泛函（linear continuous functional）**

- 所有 X 上的线性连续泛函叫做 X 的**对偶空间（dual space）** X^*，$X^* = L(X, \mathbb{K})$
- $f \in X^*$ 作用在 $u \in X$ 上可以记为 $\langle f, u\rangle := f(u)$
- $f \in X^*$ 的范数为 $\|f\| := \sup_{\|v\|\leqslant 1} |f(v)|$，所以 $|f(u)| \leqslant \|f\|\|u\|$
- 令 X 为 $\mathbb{K}$ 上的赋范空间，那么对偶空间 X^* 使用上述范数就是 $\mathbb{K}$ 上的 Banach 空间

1.23 Banach Algebras and Operator Functions

- By a **Banach algebra** $\mathcal{B}$ over $\mathbb{K}$ we understand a Banach space over $\mathbb{K}$, where an additional multiplication AB is defined such that $AB \in \mathcal{B}$ for all $A, B \in \mathcal{B}$. More over, for $A, B, C \in \mathcal{B}$ and $\alpha \in \mathbb{K}$, $(AB)C = A(BC)$, $A(B+C) = AB+AC$, $(B+C)A = BA+CA$, $\alpha(AB) = (\alpha A)B = A(\alpha B)$, $\|AB\| \leqslant \|A\|\|B\|$. Exist $E \in \mathcal{B}$ such that $AE = EA$ for all $A \in \mathcal{B}$ and $\|E\| = 1$
- define **operator function** through $F(A) := \sum_{j=0}^{\infty} a_j A^j$, and $F(z) = \sum_{j=0}^{\infty} a_j z^j, z \in \mathbb{K}$
- Let X be a Banach space over $\mathbb{K}$. For each $A \in L(X, X)$ with $\|A\| < r$, $F(A) \in L(X, X)$

1.25 Application to the Spectrum

- 考虑 $Au = \lambda u, u \in X, \lambda \in \mathbb{C}$
- 令 $A \in L(X, X)$，X 是非空的复 Banach 空间．λ 是**本征值（eigen value）**当 $u \neq 0$
- **预解集（resolvent set）** $\rho(A)$**:** 使得 $(A-\lambda I)^{-1} : X \to X$ 存在且 $\in L(X, X)$. $(A - \lambda I)$ 叫做 resolvent
- 预解集中的 λ 代入本征方程只能解得 0 矢量
- $\sigma(A) := \mathbb{C} - \rho(A)$ 叫做**谱（spectrum）**

- 谱是 $\mathbb{C}$ 中的紧子集且 $|\lambda| \leqslant \|A\|$ 对所有 $\lambda \in \sigma(A)$ 成立
- 每个本征值都属于谱
- 预解集 $\rho(A)$ 是一个开集
- Banach 空间 X 上的一个算符 $B : X \to X$ 如果值域 $R(B)$ 是闭的且零空间是有限维的，那么他就是 **semi-Fredholm** 的
- **本质谱（essential spectrum）** $\sigma_e(A)$ 包括所有使得 $(A - \lambda I)$ 不是 semi-Fredholm 的 λ
- $\sigma_e(A) \subseteq \sigma(A)$
- $\sigma_e(A)$ 就是所有具有无穷**简并（degeneracy）**的本征值的集合
- 如果 X 是有限维的，那么 A 的本质谱是空的

1.26 Density and Approximation

- $M \subseteq X$ 被称为在 X 中**稠密的（dense）**当且仅当 $\bar{M} = X$，其中 $\bar{M}$ 是 M 的闭包
- **可数的（countable）**, **至多可数（at most countable）**
- X 叫**可分的（separable）**的当且仅当它存在至多可数的稠密子集 $M \subseteq X$
- **Weierstrass 近似理论**: $X := C[a, b]$，$-\infty < a < b < \infty$．所有实系数多项式的集合在 X 内稠密
- $C[a, b]$ 是可分的
- 任何有限维赋范空间都是可分的
- 令 X 为可分的赋范空间．存在一个序列 $\{X_n\}$（X_n 是 X 的有限维线性子空间），使得 $X_1 \subseteq X_2 \subseteq \cdots \subseteq X$ 以及 $\bigcup_{n=1}^{\infty} X_n = X$

Problems

- 令 $C^k[a,b]$ （$k=1,2,\dots$）为连续函数 $u:[a,b]\to\mathbb{R}$，满足 k 阶导数连续．令 $\|u\|:=\sum_{j=0}^{k}\max_{a\leq x\leq b}\left|u^{(j)}(x)\right|$

泛函分析笔记 2

2.1 Hilbert Spaces

- 希尔伯特空间必须使用勒贝格积分（见附录）
- 内积[471] 记为 $(u|v),(u|v)\in K$
- **pre-Hilbert 空间**就是定义了内积的线性空间
- **柯西—施瓦兹不等式**[800] 是（pre-）Hilbert 空间中最重要的性质．
- pre-Hilbert 空间都是赋范空间，范数为 $\|u\|:=\sqrt{(u|u)}$
- 希尔伯特空间定义：1. 是一个内积空间，2. 是一个 Banach 空间（或者任意柯西序列的极限都属于它本身）

2.2 Standard Examples

- 空间 $X:=\mathbb{K}^N$ 是一个希尔伯特空间，内积为 $(x|y):=\sum_j\bar{\xi}_j\eta_j$，范数为 $\|x\|=(x|x)^{1/2}$
- 令 $-\infty<a<b<\infty$．$\forall u,v\in C[a,b]$，定义内积为 $(u|v)=\int_a^b uv\,\mathrm{d}x$，那么这是一个 pre-Hilbert 空间而不是 Hilbert 空间，记为 $C_*[a,b]$．该空间是以下 $L_2(a,b)$ 的稠密子集
- 令 $-\infty\leqslant a<b\leqslant\infty$，令 $L_2(a,b)$ 为所有可测函数 $u:]a,b[\to\mathbb{R}$（其中 $]a,b[$ 表示开区间）$\{x\in R:a<x<b\}$，满足 $\int_a^b|u|^2\,\mathrm{d}x<\infty$．定义内积为 $(u|v):=\int_a^b uv\,\mathrm{d}x$，那么 $L_2(a,b)$ 是无穷维的实希尔伯特空间

- $L_2(a,b)$ 的 identification principle: 两个函数 u 和 v 是同一个元素当且仅当 $u(x)=v(x)$ 对几乎所有 $x\in]a,b[$ 成立

- 令 G 表示 $\mathbb{R}^N$ （$N\geqslant 1$）中的可测非空子集，$L_2^{\mathbb{K}}(G)$ 表示可测函数 $u:G\to\mathbb{K}$ 的集合，满足 $\int_G |u|^2\,\mathrm{d}x<\infty$．那么 $L_2^{\mathbb{K}}(G)$ 是一个希尔伯特空间，内积定义为 $(u|v):=\int_G \bar{u}v\,\mathrm{d}x$，称为**勒贝格空间（Legesgue space）**

- 请写出 $L_2^{\mathbb{K}}(G)$ 中的**施瓦兹（Schwarz）不等式**

- 令 G 为 $\mathbb{R}^N$ 中的非空开子集（$N>1$）．那么 $C^k(G)$ 表示 k 阶偏导连续的函数 $u:G\to R$ 的集合

- $C^k(\bar{G})$ 包含 C^k 中所有满足各阶偏导数能拓展到 G 的闭包 $\bar{G}$ 上的函数

- 如果 $u\in C^k(G)$ 对所有的 $k=0,1,\ldots$ 都成立，那么我们记 $u\in C^\infty(G)$．同理可以定义 $C^\infty(\bar{G})$

- $C_0^\infty(G)$ 是所有 $C^\infty(G)$ 中的函数，满足在 G 的紧子集 C 外恒为零

- 令 G 为 $\mathbb{R}^N$ 中的一个非空开集，$N\geqslant 1$．那么 (i) $C_0^\infty(G)$ 和 $C(\bar{G})$ 在 $L_2(G)$ 中稠密

- $C_0^\infty(G)_{\mathbb{C}}$ 和 $C(\bar{G})_{\mathbb{C}}$ 在 $L_2^{\mathbb{C}}(G)$ 中稠密

- 分部积分的高阶拓展：$\int_G(\partial_j u)v\,\mathrm{d}x=\int_{\partial G}uvn_j\,\mathrm{d}O-\int_G u\partial_j v\,\mathrm{d}x$．其中 $x=(\xi_1,\ldots,\xi_N)$，$\partial_j u:=\partial u/\partial\xi_j$，$n_j$ 是边界 ∂G 的法向量 $n=(n_1,\ldots,n_N)$．$\int \mathrm{d}O$ 代表边界上的积分，二维情况下 $\int \mathrm{d}O$ 是逆时针的环积分．（推导见式 1 [306]）．当 u 或 v 在 ∂G 上为零时，边界积分为零

- 上一条的分部积分公式对所有 $u,v\in C^1(\bar{G})$ 成立，G 是 $\mathbb{R}^N$ 中的有界非空开集，边界足够光滑．没有边界积分的分部积分对所有 $u\in C^1(G)$ 和 $v\in C_0^\infty(G)$ 成立．G 是 $\mathbb{R}^N$ 中的一个非空开集

- 微积分基本定理 $\int_a^b w'\,\mathrm{d}x=w\big|_a^b$ 在高维中拓展为**高斯定理（Gauss theorem）** $\int_G\partial_j w\,\mathrm{d}x=\int_{\partial G}wn_j\,\mathrm{d}O$．令 $w=uv$，可得分部积分

2.3 Bilinear Forms

- 赋范空间 X 上的 **bounded bilinear form** 是一个函数 $a : X \times X \to \mathbb{K}$ 且具有性质 (1) 双线性（bilinear）：对所有 $u, v, w \in X$ 以及 $\alpha, \beta \in \mathbb{K}$，$a(\alpha u + \beta v, w) = \alpha a(u, w) + \beta a(v, w)$ 以及 $a(w, \alpha u + \beta v) = \alpha a(w, u) + \beta a(w, v)$，(2) **有界性（Boundedness）**存在常数 $d > 0$ 使得 $|a(u, v)| \leq d\|u\|\|v\|$ 对所有 $u, v \in X$ 成立

- 双线性函数 $a(\cdot, \cdot)$ 叫做对称的当且仅当 $a(u, v) = a(v, u)$ 对任意 $u, v \in X$ 成立，叫做 **positive** 当且仅当 $0 \leq a(u, u)$ 对所有 $u \in X$ 成立. 叫做 **strongly positive** 当且仅当存在常数 $c > 0$ 使 $c\|u\|^2 \leq a(u, u)$ 对所有 $u \in X$ 成立

2.4 The Main Theorem on Quadratic Variational Problems

- 令 $a : X \times X \to \mathbb{R}$ 是实希尔伯特空间 X 上的 symmetric, bounded, strongly positive, bilinear form，$b : X \to \mathbb{R}$ 是一个 X 上的线性连续函数. 那么 variational problem $a(u, u)/2 - b(u) = \min!$ （$u \in X$）有唯一解，且该式等效于 variational equation $a(u, v) = b(v)$ 对所有 $v \in X$ 成立.（想一想哈密顿原理[1065] 如何推出欧拉—拉格朗日方程[1052]）

2.5 The Functional Analytic Justification of the Dirichlet Principle

- **广义导数**：分部积分 $\int_G u\partial_j v \, \mathrm{d}x = -\int_G (\partial_j u)v \, \mathrm{d}x$ 对所有 $v \in C_0^\infty(G)$ 和 $u \in C^1(G)$ 成立. 若 $w, u \in L_2(G)$ 能使所有 $\int_G u\partial_j v \, \mathrm{d}x = -\int_G wv \, \mathrm{d}x$ 对所有 $v \in C_0^\infty(G)$ 成立，那么 w 就是 u 的广义导数（G 是 $\mathbb{R}^N$ 中的非空开集）. 同样记 $w = \partial_j u$

- 广义导数 $w = \partial_j u$ 能在一个 N 维零测度集外被唯一确定

- **Sobolev 空间** $W_2^1(G)$：G 是 $\mathbb{R}^N$ 中的非空开集，$u, \partial_j u \in L_2(G)$. 定义内积为 $(u|v)_{1,2} := \int_G (uv + \sum_j \partial_j u \partial_j v) \, \mathrm{d}x$

- $W_2^1(G)$ 是一个希尔伯特空间，如果我们认为两个在大多数地方相等的函数是同一个函数

- 定义 $\mathring{W}_2^1(G)$ 为 $C_0^\infty(G)$ 在 $W_2^1(G)$ 上的闭包

- $\mathring{W}_2^1(G)$ 是 $W_2^1(G)$ 的（实的）子希尔伯特空间
- 从广义的角度理解，$\mathring{W}_2^1(G)$ 中函数的边界值为 0

2.8 Generalized Functions and Linear Functionals

- **多重指标（multiindex）**：$\alpha = (\alpha_1, \ldots, \alpha_N)$，令 $|\alpha| := \alpha_1 + \cdots + \alpha_N$, 以及 $\partial^\alpha u := \partial_1^{\alpha_1} \ldots \partial_N^{\alpha N} u = \partial^{|\alpha|} u / (\partial \xi_1^{\alpha_1} \ldots \partial \xi_N^{\alpha N})$
- 分部积分：G 为 $\mathbb{R}^N \geq 1$ 上的所有非空开区间．对所有 $u, v \in C_0^\infty(G)$ 以及所有多重指标 α，$\int_G u \partial^\alpha v \,\mathrm{d}x = (-1)^{|\alpha|} \int_G (\partial^\alpha u) v \,\mathrm{d}x$ （重复使用分部积分即可证明）
- 令 $\mathcal{D}(G) := C_0^\infty(G)$．令 $\phi_n, \phi \in \mathcal{D}(G)$．$\phi_n \to \phi$ 的定义是：对所有的多重指标 α，K 上有一致收敛[805] $\partial^\alpha \phi_n \to \partial^\alpha \phi(x)$
- 如果 $G =]\alpha, \beta[$，那么 $\mathcal{D}(\alpha, \beta) := \mathcal{D}(G)$
- **广义函数（generalized function）** $U \in \mathcal{D}'(G)$ 定义：线性连续泛函 $U : \mathcal{D}(G) \to \mathbb{R}$．广义函数也叫**分布（distribution）**
- $L_2(G) \subseteq \mathcal{D}'(G)$ 的意思是每个 $u \in L_2(G)$ 都对应（identified with）一个广义函数 $U(\phi) := \int_G u(x)\phi(x) \,\mathrm{d}x$ 对所有 $\phi \in \mathcal{D}(G)$ 成立．$U \in \mathcal{D}'(G)$．如果 $u = v$，那么 $U = V$
- **狄拉克 δ_y 分部**：$\delta_y(\phi) := \phi(y)$ 对所有 $\phi \in \mathcal{D}(G)$ 成立
- 定义广义函数 $U \in \mathcal{D}'(G)$ 的导数 $\partial^\alpha U$ 为 $(\partial^\alpha U)(\phi) := (-1)^{|\alpha|} U(\partial^\alpha \phi)$ 对所有 $\phi \in \mathcal{D}(G)$ 成立
- 如果 $u \in \mathcal{D}(G)$ 对应的广义函数为 U，$\partial^\alpha u$ 对应的广义函数为 V，那么 $V = \partial^\alpha U$
- 如果 $U \in \mathcal{D}'(G)$，那么 $\partial^\alpha U \in \mathcal{D}'(G)$ 对所有的 α 成立
- 广义函数存在任意阶导数

- 广义函数的极限 $U_n \to U$ 的定义：$U_n(\phi) \to U(\phi)$ 对所有 $\phi \in \mathcal{D}(G)$ 都成立
- 在 $L_2(G)$ 中 $u_n \to u$ 意味着对应的 $U_n \to U$
- 在 $\mathcal{D}(\alpha, \beta)$ 中，$f_{y,\epsilon}$ 对应的泛函 $F_{y,\epsilon} \to \delta_y$ 当 $\epsilon \to +0$. $f_{y,\epsilon}$ 是区间 $[y-\epsilon, y+\epsilon]$ 外为零，积分为 1 的函数
- 在 $\mathcal{D}(\alpha, \beta)$ 中，方程 $-U'' = \delta_y$ 的解 U 对应的就是格林函数，$U(\phi) = \int \mathcal{G}(x, y)\phi(x)\,\mathrm{d}x$
- 令 $u, w \in L_2(G)$，广义导数 $w = \partial^\alpha u$ 的定义为：对应的广义函数满足 $W = \partial^\alpha U$. 这比之前定义的广义导数更一般化
- 广义导数（除了零测度集）是唯一确定的

2.9 Orthogonal Projection

- **orthogonal complement**：$M^\perp := w \in X : (w|v) = 0 \ \forall \ v \in M$
- **perpendicular principle**：令 M 为希尔伯特空间 X 上闭合的线性子空间. 对于给定的 $u \in X$，$\|u - v\| = \min!$ 存在唯一的解 v，且 $u - v \in M^\perp$
- **orthogonality decomposition**：如果给定 $u \in M$，要求 $w \in M^\perp$，那么 $u = v + w$ 是唯一的

2.10 Linear Functionals and the Riesz Theorem

- **Riesz teorem**：令 X 为 $\mathbb{K}$ 上的希尔伯特空间，令 X^* 为 X 的对偶空间. 那么 $f \in X^*$ 当且仅当存在 $v \in X$ 使得 $f(u) = (v|u)$ 对所有 $u \in X$. v 可以由 f 唯一确定，且 $\|f\| = \|v\|$
- 如果 f 是 Hilbert 空间中的非零线性连续函数，那么它的零空间 $N(f)$ 是一个闭合平面且 orthogonal complement $N(f)^\perp$ 是一维的

2.11 The Duality Map

- **duality map** $J : X \to X^*$ 把 $v \in X$ 映射到 $f(u) = (v|u)$（对所有 $u \in X$）
- 定义 $\langle f, u\rangle = f(u)$（$f \in X^*, u \in X$），那么 $\langle J(v), u\rangle := (v|u)$（对所有 $u, v \in X$）
- 对偶映射 J 是双射的，连续的以及 norm preserving．即 $\|J(u)\| = \|u\|$ 对所有 $u \in X$
- 如果 X 是实 Hilbert 空间，那么 J 是线性的．如果 X 是复 Hilbert 空间，那么 J 是反线性的，即 $J(\alpha v + \beta w) = \bar{\alpha} Ju + \bar{\beta} Jw$（对所有 $\alpha, \beta \in \mathbb{C}, u, w \in X$）

2.13 The Linear Orthogonality Princple

- 以下三个条件互相等价：(1) 二次最小值问题的 existence principle (2) 垂直定理 (3) Riesz 定理

2.14 Nonlinear Monotone Operators

- 实 Hilbert 空间 X 上的**强单调（strongly monotone）**算符（注意不一定是线性的！）$A : X \to X$ 定义为：存在常数 $c > 0$ 使得 $(u - v|Au - Av) \geq c\|u - v\|^2$ 对所有 $u, v \in X$ 成立
- 对任意 $z \in X$ 以及强单调，Lipschitz 连续的算符 A，$Au = z$ 存在唯一解 $u \in X$

Problems

- **special tensor product**：经典分析中，两个函数的张量积定义为 $(\phi \otimes \psi)(x, y) := \phi(x)\psi(y)$．令广义函数 $U, \delta \in \mathcal{D}'(\mathbb{R})$．定义 $(U \otimes V)(\chi) = U(V(\chi))$

泛函分析笔记 3

3.1 Orthonormal Series

- **正交归一系（orthonormal system）**：$(u_k|u_m) = \delta_{k,m}$ 对所有 k, m 成立
- 令 X 为实 Hilbert 空间，$\{u_0, u_1, \dots\}$ 为 X 中至多可数的正交归一系．对有限正交归一系，如果 $u = \sum_{n=0}^{N}(u_n|u)u_n$ 对所有 $u \in X$ 成立，它就叫做**完备的（complete）**．对可数的正交归一系，令 $s_m := \sum_{n=0}^{m}(u_n|u)u_n$，如果 $u = \lim_{m\to\infty} s_m$ 对所有 $u \in X$ 成立，它就叫做**完备的（complete）**
- 有限的正交归一系在 Hilbert 空间 X 中是完备的当且仅当它是 X 的一组基底
- 令 $\{u_n\}$ 为Hilbert空间中可数的正交归一系.如果无穷级数 $u = \sum_{n=0}^{\infty} c_n u_n$ 对某个 $u \in X$ 收敛，那么 $c_n = (u_n|u)$ 对所有 n 成立
- **Bessel inequality**：$\sum_{n=0}^{m} |(u_n|u)|^2 \le |u|^2$ 对所有 $u \in X$ 和 m 成立
- **Convergence criterion**：令 $\{u_n\}$ 为 Hilbert 空间 X 中的可数正交归一系．那么 $\sum_{n=0}^{\infty} c_n u_n$ 收敛当且仅当 $\sum_{n=0}^{\infty} |c_n|^2$ 收敛
- 令 $\{u_n\}$ 为可数正交归一系，那么以下两个条件等效：(1) 它在 X 中是完备的．(2) 它的 linear hull （span）在 X 中是稠密的
- 对 Hilbert 空间 X 中可数完备的正交归一系 $\{u_n\}$：(1) **Parseval equation**：$(u|v) = \sum_{n=0}^{\infty} \bar{c}_n(u)c_n(v)$，(2) $\|u\|^2 = \sum_{n=0}^{\infty} |(u_n|u)|^2$，(3) 如果 $(u_n|u) = 0$ 对所有 n 和固定的 $u \in X$ 成立，那么 $u = 0$
- 对每个 $u \in L_2(-\pi, \pi)$，傅里叶级数收敛．即 $\lim_{m\to\infty} \int_{-\pi}^{\pi} [u(x) - a_0/2 - \sum_k a_k \cos kx + b_k \sin kx]^2 \, \mathrm{d}x = 0$

3.5 Unitary Operators

- 令 X 和 Y 为 $\mathbb{K}$ 上的希尔伯特空间．算符 $U : X \to Y$ 叫做 **unitary** 当且仅当 U 是线性的，满射的，且 $(Uv|Uw) = (v|w)$ 对所有 $v, w \in X$ 成立

- 如果算符 U 是 unitary 的，那么它就是双射的，连续的，且 $\|Uv\| = \|v\|$ 对所有 $v \in X$ 成立．而且，存在逆算符 $U^{-1} : Y \to X$，同样是 unitary 的

3.6 The Extension Principle

- 对 Banach 空间 X 和 Y，令线性算符 $A : D \subseteq X \to Y$ 的范数为有限值，$D \subseteq X$ 是稠密线性的．那么 (1) A 可以唯一地拓展（extended）到 $A : X \to Y$ 上，范数仍然为有限．（2）如果 A 在 D 上的紧算符，那么拓展后也是紧算符

3.7 Applications to the Fourier Transformation

- $\mathcal{S}$ 空间：包含所有 C^∞ 函数 $u : \mathbb{R} \to \mathbb{C}$，满足 $\|u\|_{p,q} < \infty$ 对所有 $p, q = 0, 1, \ldots$ 成立．其中 $\|u\|_{p,q} := \sup_{x\in\mathbb{R}}(1 + |x|^p) \sum_{n=0}^{q} \left|u^{(n)}(x)\right|$．极限 $u_n \to u$ 以 $\|u\|_{p,q}$ 为准

- $\mathcal{S}$ 空间中的函数叫做**在无穷远处 rapidly decreasing**．且满足 $\left|\int_{-\infty}^{\infty} u^{(n)}(x)\,\mathrm{d}x\right| < \infty$ 对所有 $n = 0, 1, \ldots$ 成立．且分部积分的边界项为零

- 傅里叶变换 $F : \mathcal{S} \to \mathcal{S}$ 是线性的连续双射的．反傅里叶变换同样是连续的

- 傅里叶变换 $F : \mathcal{S} \to \mathcal{S}$ 可以唯一地拓展到酉算符 $F : L_2^{\mathbb{C}}(\mathbb{R}) \to L_2^{\mathbb{C}}(\mathbb{R})$．注意 $C_0^\infty(\mathbb{R})_{\mathbb{C}} \subseteq \mathcal{S} \subseteq L_2^{\mathbb{C}}(\mathbb{R})$

3.8 The Fourier Transform of Tempered General Functions

- $\mathcal{S}'$ 定义为所有线性连续映射 $T : \mathcal{S} \to \mathbb{C}$．$T$ 叫做 **tempered generalized functions** 或者 **tempered distributions**

- 定义 $(FT)(u) := T(Fu)$ （$u \in \mathcal{S}$）

- 算符 $F : \mathcal{S}' \to \mathcal{S}'$ 是线性且双射的

- 令 $v : \mathbb{R} \to \mathbb{C}$ 为可测的有界函数．算符 $T(u) = \int_{-\infty}^{\infty} v(x)u(x)\,\mathrm{d}x$（$\forall u \in \mathcal{S}$）．那么 $T \in \mathcal{S}'$

- 令 $y \in \mathbb{R}$，定义 **tempered delta distribution** δ_y：$\delta_y(u) := u(y)$（$\forall u \in \mathcal{S}$）. 那么 (1) $\delta_y \in \mathcal{S}'$，(2) $F\delta_y = (2\pi)^{-1/2}\text{“}e^{-iky}\text{”}$，(3) $(2\pi)^{-1/2}F^{-1}(\text{“}e^{-iky}\text{”}) = \delta_y$. 其中定义 $\text{“}e^{-iky}\text{”}(u) := \int_{-\infty}^{\infty} e^{-iky}u(k)\,\mathrm{d}k$ （$\forall u \in \mathcal{S}$）

Problems

- **Tensor Product** $X \otimes Y$

泛函分析笔记 4

4.1 Symmetric Operators

- Hilbert 空间 X 上的线性算符 $A : D(A) \subseteq X \to X$ 是对称的当且仅当 $D(A)$ 在 X 中稠密且 $(Au|u) = (u|Av)$ 对所有 $u, v \in D(A)$ 成立

- Hilbert 空间 X 上的线性对称算符 $A : D(A) \subseteq X \to X$ 满足：(1) $(Au|u)$ 是实数，(2) 所有本征值为实数，(3) 不同本征值对应的本征矢正交，(4) 本征矢构成的至多可数的完备正交归一系对应的本征值包含 A 的所有本征值

4.2 The Hilbert-Schmidt Theory

- 令 $A : X \to X$ 为可分 Hilbert 空间中的一个非空线性对称紧算符，那么 (1) A 存在本征矢构成的完备正交归一系

- 所有本征值 λ 为实数，且每个 $\lambda \neq 0$ 存在有限简并

- 不同本征值对应的本征矢正交

- 如果 A 有可数个本征值（注意 $\lambda = 0$ 不是本征值），那么本征值序列 $\lambda_n \to 0$

泛函分析笔记 5

5.1 Extensions and Embeddings

- 线性空间 X 到 Y 的算符 A 和 B 记为 $B \subseteq A$ 当且仅当定义域 $D(B) \subseteq D(A)$ 且它们在 $D(B)$ 上是同一算符．这时 A 是 B 的 **extension**
- $A = B$ 当且仅当 $A \subseteq B$ 且 $B \subseteq A$
- **embedding** $X \subseteq Y$ 是**连续的**当且仅当存他们之间存在线性，单射，连续的算符
- **embedding** $X \subseteq Y$ 是**紧的**当且仅当存他们之间存在线性，单射，紧的算符

5.2 Self-Adjoint Operators

- 令线性算符 $A : D(A) \subseteq X \to X$ 的定义域 $D(A)$ 在希尔伯特空间 X 上稠密．定义 $v \in D(A^*)$ 当且仅当存在 $w \in X$ 使 $(v|Au) = (w|u)$ 对任意 $u \in D(A)$ 都成立．令 $A^*v := w$，就得到了**伴随（adjoint）**算符 $A^* : D(A^*) \subseteq X \to X$．所以 $D(A^*)$ 是满足定义的最大集合
- 伴随算符：(1) 是线性的 (2) $(\alpha A)^* = \bar{\alpha} A^*$ (3) $A \subseteq B$ 意味着 $B^* \subseteq A^*$
- 如果 $D(A^*)$ 在 X 上稠密，那么存在 $(A^*)^*$，记为 A^{**}
- 希尔伯特空间 X 上的线性算符 A 是**对称的（symmetric）**当且仅当 $A \subseteq A^*$，即 $(Au|v) = (u|Av)$ 对所有 $u, v \in D(A)$ 成立
- 希尔伯特空间 X 上的线性算符 A 是**自伴的（self-adjoint）**当且仅当 $A = A^*$，注意 $D(A) = D(A^*)$．自伴算符都是对称算符
- 希尔伯特空间 X 上的线性算符 A 是 **skew-symmetric** 当且仅当 $A \subseteq -A^*$
- 希尔伯特空间 X 上的线性算符 A 是 **斜自伴（skew-adjoint）**当且仅当 $A = -A^*$

- 希尔伯特空间 X 上的线性连续算符 A 的伴随算符也是线性连续的，且 $\|A\| = \|A^*\|$ 以及 $A^{**} = A$

- 令 $f : [a,b] \times [a,b] \to \mathbb{R}$ 为连续函数且 $-\infty < a < b < \infty$．定义算符 A 为 $(Au)(x) := \int_a^b f(x,y)u(y)\,\mathrm{d}y$，令 $X := L_2(a,b)$．那么 (1) $A : X \to X$ 是线性紧算符，(2) $(A^*u)(x) = \int_a^b f(y,x)u(y)\,\mathrm{d}y$，$A^*$ 也是线性紧算符．(3) 如果 $f(x,y) = f(y,x)$ 对任意 $x, y \in [a,b]$ 成立，那么算符 A 是自伴算符

- 希尔伯特空间 X 上的任意的线性自伴算符 A 都是 **maximally symmetric**．也就是说，如果 S 是对称算符且 $A \subseteq S$，那么 $A = S$

- 令 $X := L_2^{\mathbb{C}}(\mathbb{R})$，且 $(Au)(x) := u'(x)$ （$x \in \mathbb{R}$），$D(A) := u \in X : u' \in X$, u' 为广义导数，那么 (1) 算符 A 是斜自伴算符，(2) iA 是自伴算符

- 令 B 为上一条中的 A 改成“非广义”的求导算符，且定义域为 $X \bigcap C^1(\mathbb{R})$，那么 B 是 skew-symmetric 算符且 iB 是对称算符

- 令 $X := L_2^{\mathbb{C}}(\mathbb{R})$，定义 $(Mu)(x) := xu(x)$（$\forall x \in R$），$D(M) := \{u \in X : Mu \in X\}$．那么算符 M 是自伴算符

- 回忆：令 M 为 Hilbert 空间 X 的闭线性子空间，任何 $u \in X$ 存在唯一的分解 $u = v + w$（$v \in M$，$w \in M^{\perp}$）．**正交投影算符** $P : X \to M$ 定义为 $Pu := v$

- 正交投影算符 P 是线性的，连续的，自伴的．且 $P^2 = P$．如果 M 为非空，那么 $\|P\| = 1$

- 如果 $P : X \to X$ 是一个线性连续的自伴算符且 $P^2 = P$，那么 P 就是正交投影算符

- 令 $A : D(A) \subseteq X \to X$ 为 Hilbert 空间 X 上的线性对称算符，且 $R(A)$ 在 X 上稠密，那么 $(A^{-1})^* = (A^*)^{-1}$．如果这个 A 是自伴算符，那么 A^{-1} 也是

- 对 Hilbert 空间 X 上的线性算符 $U : X \to X$，以下条件等价．(1) U 是 unitary 算符，(2) $UU^* = U^*U = I$，(3) U 是双射且 $U^{-1} = U^*$，(4) U 是满射且 $\|Uv\| = \|v\|$ 对所有 $v \in X$ 成立

5.9 Semigroups, One-Parameter Groups, and Teir Physical Relevance

- 令 X 为 Banach 空间．X 上的 **semigroup** $\{S(t)\}_{t\geq 0}$ 包含一系列算符 $S(t) : X \to X$ （$\forall t \geq 0$）使得：$S(t+s) = S(t)S(s)$ （$\forall t, s \geq 0$）以及 $S(0) = I$

- Semigroup $\{S(t)\}_{t\geq 0}$ 的 **generator** $A : D(A) \subseteq X \to X$ 定义为 $Au := \lim_{t\to+0}[S(t) - I]t^{-1}u$，$u \in D(A)$ 当且仅当这个极限存在

- 令 $\mathcal{S}_+$ 为 semigroup，$u(t) := S(t)u_0$ （$\forall t \geq 0$）．那么 $\mathcal{S}_+$ 是**强连续的 strongly continuous** 当且仅当对任意 $u_0 \in X$，函数 $u : [0, \infty[\to \mathbb{R}$ 是连续的

- $\mathcal{S}_+$ 被称为 **nonexpansive** 当且仅当 $\mathcal{S}_+$ 是强连续的且 $\|S(t)u_0 - S(t)v_0\| \leq \|u_0 - v_0\|$ （$\forall u_0, v_0 \in X$，对每个 $t \geq 0$）

- $\mathcal{S}_+$ 被称为**线性的（linear）**当且仅当 $S(t) : X \to X$ （$\forall u_0, v_0 \in X$）是线性连续的

- Banach 空间 X 上的 **one-parameter group** $\{S(t)\}_{t\in\mathbb{R}}$ 包含一系列算符 $S(t) : X \to X$ （$\forall t \in \mathbb{R}$），使得：$S(t+s) = S(t)S(s)$ （$\forall t, s \in \mathbb{R}$）以及 $S(0) = I$

- One-parameter group 的 **generator** $A : D(A) \subseteq X \to X$ 定义为 $Au := \lim_{t\to 0}[S(t) - I]t^{-1}u$，$u \in D(A)$ 当且仅当这个极限存在

- one-parameter group 的强连续和线性类比 semigroup

- $\mathcal{S}$ 叫做**一致连续（uniformly continuous）**当且仅当 $\mathcal{S}$ 是线性的且 $\|S(t+h) - S(t)\| \to 0$ 当 $h \to 0$ （$\forall t \in \mathbb{R}$）

- 一致连续的 $\mathcal{S}$ 也是强连续的

- 令 $A : X \to X$ 为 Banach 空间 X 上的线性连续算符，令 $S(t) := e^{tA}$ （$\forall t \in \mathbb{R}$）．那么 (1) $\mathcal{S} = \{S(t)\}$ 是 X 上的线性 one-parameter group，generator 为 A (2) $\mathcal{S}$ 是一致连续（强连续）的．(3) 给出 $u_0 \in X$，令

$u(t) := e^{tA}u_0$ （$\forall t \in \mathbb{R}$），那么 $u = u(t)$ 是微分方程 $u'(t) = Au(t)$（$-\infty < t < \infty$），$u(0) = u_0$ 的唯一解

- 如果上面的 A 是线性连续的，那么上面的微分方程无法用于描述自然中的不可逆过程

- **one-parameter unitary group**：强连续，one-parameter group，每个 $S(t)$ 都是 unitary 的

- 令 A 为复 Hilbert 空间 X 中的线性连续的自伴算符，令 $S(t) := e^{iAt}$（$\forall t \in \mathbb{R}$），那么 $\{S(t)\}$ 是 one-parameter unitary group，generator 是 iA

5.13 Applications to the Schrodinger Equation

- $u'(t) = -iAu(t)$ （$-\infty < t < \infty$），$u(0) = u_0$

- 令 $A : D(A) \subseteq X \to X$ 为复 Hilbert 空间 X 的自伴算符，

5.14 Applications to Quantum Mechanics

- Hilbert 空间中单位矢量 $\psi \in X$ （$(\psi|\psi) = 1$）叫做**态（state）**

- Hilbert 空间中的自伴算符 A 叫做可观测量

- 测量

- 不确定原理

- **动量算符** $A : D(A) \subseteq X \to X$ 定义为 $(A\phi)(x) := -i\hbar\, \mathrm{d}\phi(x)/\mathrm{d}x$ （$\forall x \in \mathbb{R}$）．其中 $D(A) := \{\phi \in X : \phi' \in X\}$，$\phi'$ 是广义导数

- $(B\phi)(x) := x\phi(x)$（$x \in \mathbb{R}$）叫做**位置算符**，其中 $D(B) := \phi \in X : B\phi \in X$

- 位置算符和动量算符都是自伴算符（见上文）

5.15 Generalized Eigenfunctions

- 令 $X := L_2^{\mathbb{C}}(\mathbb{R})$，$A : D(A) \subseteq X \to X$ 为对称算符，$\mathcal{S} \subseteq D(A)$ 那么非零的 tempered distribution $T \in \mathcal{S}'$ 叫做 A 的 **广义本征函数（generalized eigenfunction）**（具有实数本征值）当且仅当 $T(A\phi) = \lambda T(\phi)$ 对所有 $\phi \in \mathcal{S}$ 成立
- 广义本征函数系 $\{T_\alpha\}_{\alpha\in\mathcal{A}}$ 称为**完备的（complete）**当且仅当 “$T_\alpha(\phi) = 0$ 对所有 $\alpha \in \mathcal{A}$ 和固定的 $\phi \in \mathcal{S}$ 成立” 意味着 $\phi = 0$
- 令 $T(\phi) := (\psi|\phi) = \int_{-\infty}^{\infty} \overline{\psi(x)}\phi(x)\,\mathrm{d}x$ （$\forall\phi \in \mathcal{S}$），那么 (1) 对每个 $\psi \in X$，$T \in \mathcal{S}'$，(2) 映射 $\psi \mapsto T$ 是从 X 到 $\mathcal{S}'$ 的线性双射算符
- 上面 A 中的每个本征函数 $\psi \in D(A)$ 可以映射到一个广义本征函数
- 令 $p \in \mathbb{R}$, $T_p(\phi) := \int_{-\infty}^{\infty} \overline{\phi_p(x)}\phi(x)\,\mathrm{d}x$（$\forall\phi \in \mathcal{S}$）其中 $\phi_p(x) := \exp(ipx/\hbar)$. $T_p \in \mathcal{S}'$
- $\{\phi_p\}_{p\in\mathbb{R}}$ （对应的广义函数 $T_p(\phi)$）是动量算符的完备广义本征函数
- $\{\delta_y\}_{y\in\mathbb{R}}$ 是位置算符 B 的完备广义本征函数.即 $\delta_y(B\phi) = y\delta_y(\phi)$（$\forall\phi \in \mathcal{S}$）

5.20 A Look at Scattering Theory

- 实际运动:$\psi(t) := \exp(-itH/\hbar)\psi(0)$,自由运动:$\psi_0(t) := \exp(-itH_0/\hbar)\psi_0(0)$
- 运动 $\psi = \psi(t)$ 被称为 **asymptotically free** 当 $t \to +\infty$
- $\psi(0)$ 是 asymptotically free motion 的初态当且仅当 $\psi(0)$ 与 H 的所有本征矢（bound states）正交
- 散射态 ϕ_p 是（对应）H_0 的广义本征函数 $T_p(\phi)$. $T_p(H_0\phi) = p^2/(2m)T_p(\phi)$（$\forall\phi \in \mathcal{S}, p \in \mathbb{R}$）

5.21 The Language of Physicists in Quantum Physics and the Justification of the Dirac Calculus

- 令 $\phi_k(x) := (2\pi)^{-1/2}e^{ikx}$ $x, k \in \mathbb{R}$

Appendix

The Lebesgue Measure

- N-cuboid：$C := \{(\xi_1, \ldots, \xi_2) \in \mathbb{R}^N : a_j < \xi_j < b_j\}$ （$j = 1, \ldots, N$）
- N-cuboid 的体积为 $\mathrm{vol}(C) := \prod_{j=1}^{N}(b_j - a_j)$
- 勒贝格测度将 $\mathbb{R}^N$ 中的经典的体积从足够有规律的集合拓展到一些“无规律”的集合．令 $\mathcal{A}$ 为 $\mathbb{R}^N$ 中可测子集的集合
- (1) $\mathbb{R}^N$ 中任意开集和闭集属于 $\mathcal{A}$．(2) 如果 $A, B \in \mathcal{A}$，那么 $A \cup B \in \mathcal{A}$，$A \cap B \in \mathcal{A}$，$A - B \in \mathcal{A}$．(3) $A_n \in \mathcal{A}$ 的无穷交集和无穷并集仍然属于 $\mathcal{A}$．(4) 每个 $A \in \mathcal{A}$ 可以赋予一个数 $0 \le \mu(A) \le \infty$, 叫做 N 维**测度（measure）**，$A \in \mathcal{A}$ 叫做**可测的（measurable）**．(5) 如果 $A \cap B = \emptyset$，那么 $\mu(A \cup B) = \mu(A) + \mu(B)$．这对无穷个 A_n 仍然适用．(6) 如果 C 是一个 N-cuboid，那么 $C \in \mathcal{A}$ 以及 $\mu(C) = \mathrm{vol}(C)$．(7) A 的测度为零当且仅当对任意 $\varepsilon > 0$，存在可数个 N-cuboid $C_1, C_2, \ldots$ 使得 $A \subseteq \bigcup_{j=1}^{\infty} C_j$ 且 $\sum_{j=1}^{\infty} \mu(C_j) < \varepsilon$．(8) 如果 A 的 N 维测度为零且 $B \subseteq A$，那么 B 的 N 维测度也是零．(9) $\mathcal{A}$ 叫做**最小的（minimal）**，如果 $\mathcal{A}'$ 满足以上条件，就有 $\mathcal{A} \subseteq \mathcal{A}'$
- 测度 μ 在 $\mathcal{A}$ 上唯一．μ 叫做勒贝格测度，也记为 $\mathrm{meas}(A) := \mu(A)$（$\forall A \in \mathcal{A}$）
- $\mathbb{R}^N$ 中有限个或可数个点的 N 维测度为零
- 一个性质 P **几乎到处（almost everywhere）**都成立，当且仅当 P 对 $\mathbb{R}^N$ 中除了测度为零的集合的点都成立
- **几乎所有（almost all）**也类似．例如几乎所有的实数都是无理数
- 例如令 $M \subseteq \mathbb{R}^N$，$u(x) = \lim_{n\to\infty} u_n(x)$ （对几乎所有 $x \in M$）
- $\mu(M) \le \mu(G)$ 当 $M \subseteq G$，$\mu(\mathbb{R}^N) = +\infty$ 且 $\mu(\emptyset) = 0$

Step Functions

- **阶梯函数（step function）**：$u : M \subseteq \mathbb{R}^N \to \mathbb{K}$，piecewise constant．$M$ 是可测的，存在有限个两两不相交的 $M_j \subseteq M$，使得他们的测度为有限且 $u(x) = a_j$ （$\forall x \in M_j$ 和 $\forall j$），$u(x) = 0$（otherwise）

- 定义阶梯函数的积分为 $\int_M u \, \mathrm{d}x := \sum_j \mathrm{meas}(M_j) a_j$

Measurable Functions

- **可测函数（Measurable Function）**：$u : M \subseteq \mathbb{R}^N \to \mathbb{K}$ (1) M 是可测的 (2) 存在阶梯函数的序列 $u(x) = \lim_{n\to\infty} u_n(x)$ （对几乎所有 $x \in M$）

- **Luzin 定理**：令 M 为 $\mathbb{R}^N$ 的可测子集．函数 $u : M \to \mathbb{K}$ 是可测的当且仅当对任意 δ，它在除测度为 δ 的集合外是连续的

- 函数 $f : M \subseteq \mathbb{R}^N \to \mathbb{K}$ 是可测的如果他在可测集 M 上几乎到处连续

- 可测函数的线性组合和极限也是可测的

- 可测函数在定义域中测度为 0 的子集上改变了，那么新的函数仍然是可测的

Lebesgue Integral

- **勒贝格积分（Lebesgue Integral）**：函数 $u : M \subseteq \mathbb{R}^N \to \mathbb{K}$ 是**可积（integrable）**的当且仅当两个条件成立：(1) 存在由阶梯函数构成的函数列 u_n 使得 $u(x) = \lim_{n\to\infty} u_n(x)$ 对几乎所有 $x \in M$ 成立 (2) 对任意 $\varepsilon > 0$，存在 $n_0(\varepsilon)$ 使得 $\int_M |u_n(x) - u_m(x)| \, \mathrm{d}x < \varepsilon$ 对所有 $n, m \geq n_0(\varepsilon)$ 成立

- 如果 u 是可积的，那么定义 $\int_M u \, \mathrm{d}x := \lim_{n\to\infty} \int_M u_n \, \mathrm{d}x$

- 显然，可积函数必定可测

装备希尔伯特空间

本节中都用 $\mathcal{H}$ 来代表所要考虑的 Hilbert 空间，用 $\langle\cdot,\cdot\rangle$ 来代表其上的内积.

按照通常的说法，量子力学的基本舞台是 **Hilbert 空间**，即完备的内积空间．然而在这里，我们要考虑的数学对象比 Hilbert 空间更精细一些：我们的基本舞台不再是一般的 Hilbert 空间，而是所谓的**装备 Hilbert 空间（rigged Hilbert space, RHS）**．为了定义装备 Hilbert 空间，首先要定义**核空间（nuclear space）**．它是 Alexander Grothendieck 的博士论文研究的内容．我们这里采用如下较狭窄但已经足够用的定义：

定义 1　核空间（Nuclear Space）

一个核空间 Φ 是一族递降的 Hilbert 空间 $\Phi_1 \supset \Phi_2 \supset \ldots$ 的交，赋予自然的拓扑，这族 Hilbert 空间满足如下条件：

1. Φ_{n+1} 的范数比 Φ_n 的范数更强，而且 Φ_n 是 Φ_{n+1} 在 Φ_n 的范数之下的完备化；

2. 对于任何 n，总有 $m > n$ 使得嵌入映射 $T_{m,n} : \Phi_m \hookrightarrow \Phi_n$ 是核算子（nuclear operator），即它的 Hilbert-Schmidt 迹范数是良好定义的.

从此以后，沿用 Dirac 的左右矢记号：以 $|\phi\rangle$ 代表 Φ 的元素，$\langle u|$ 代表其对偶空间 Φ^* 的元素，二者之间的配对就以 $\langle u|\phi\rangle$ 代表．今后还要考虑包含映射 $\Phi \hookrightarrow \mathcal{H}$，它自然诱导出包含映射 $\mathcal{H} \hookrightarrow \Phi^*$．为方便计，对于 $f \in \Phi \subset \mathcal{H}$，以 $|f\rangle$ 表示之; 对于 $g \in \mathcal{H}$，以 $\langle g|$ 表示 g 在包含映射 $\mathcal{H} \hookrightarrow \Phi^*$ 下的像.

定义 2　装备 Hilbert 空间（Rigged Hilbert Space, RHS）

装备 Hilbert 空间包含一可分 Hilbert 空间 $\mathcal{H}$ 与一稠密子空间 $\Phi \subset \mathcal{H}$，满足如下条件：Φ 上有一更强的拓扑 τ_Φ，使得 (Φ, τ_Φ) 成为一核空间，而且有 $\Phi_0 = \mathcal{H}$，且自然的包含映射 $\Phi \hookrightarrow \mathcal{H}$ 在此拓扑下连续.

注意到自然的三重偶 $\Phi \hookrightarrow \mathcal{H} \hookrightarrow \Phi^*$ 与 $\mathcal{H}$ 的内积相容，即如果 $f \in \Phi \subset \mathcal{H}$，$g \in \mathcal{H}$，则 $\langle g, f\rangle_{\mathcal{H}} = \langle g|f\rangle$．此三重偶称为 **Gelfand 三重偶（Gelfand triple）**.

为什么要如此大费周章地考虑 Gelfand 三重偶？原来，它的原型乃是 Hilbert 空间语言中的位置表象空间 $L^2(\mathbb{R}^n)$ 以及其上定义的算子 $X, i\nabla$，即位置算符

和动量算符; 这两个算子的定义域都不是整个 $L^2(\mathbb{R}^n)$，它们最自然的定义域乃是 Schwartz 函数空间 $\mathcal{S}(\mathbb{R}^n)$，即所有光滑且各阶导数都迅速衰减的函数的空间，赋予 Fréchet 空间拓扑．位置算子和动量算子在 Schwartz 空间上都是连续算子．而 Schwartz 空间 $\mathcal{S}(\mathbb{R}^n)$ 本身是加权的 Sobolev 空间系

$$\Phi_k := \left\{ f \in L^2(\mathbb{R}^n) : (-\Delta + |x|^2)^k f \in L^2(\mathbb{R}^n) \right\}$$

的交，而包含映射 $T_{m+n,m} : \Phi_{m+n} \to \Phi_m$ 是核算子．这时就有 Gelfand 三重偶

$$\mathcal{S}(\mathbb{R}^n) \hookrightarrow L^2(\mathbb{R}^n) \hookrightarrow \mathcal{S}'(\mathbb{R}^n) .$$

也可以注意到，上述定义的 Sobolev 范数中的算子 $-\Delta + |x|^2$ 正是谐振子的薛定谔算符．

这个三重偶应该包含了所有物理上感兴趣的对象．实际上，它显然囊括了平方可积的束缚态，即可归一化的波函数; 也囊括了不平方可积的散射态，例如自由单色波 $e^{ik\cdot x}$．前者属于 $L^2(\mathbb{R}^n)$，而后者应该用 Schwartz 分布即 $\mathcal{S}'(\mathbb{R}^n)$ 的元素来代表．另外，物理上的" 位置本征态" 即 $\delta(x - x_0)$ 显然并不是数学意义下的函数，但它属于对偶空间 $\mathcal{S}'(\mathbb{R}^n)$，而且当然满足

$$X^j \delta(x - x_0) = x_0^j,$$

因为对于任何 $\phi \in \mathcal{S}(\mathbb{R}^n)$，按照定义皆有

$$\begin{aligned} \left[\langle \delta(x - x_0)| \, X^j \right] |\phi\rangle &:= \langle \delta(x - x_0)| \left[X^j \, |\phi\rangle \right] \\ &= x_0^j \phi(x_0) = \left[\langle \delta(x - x_0)| \, x_0^j \right] |\phi\rangle . \end{aligned}$$

由此可见，$\delta(x - x_0)$ 是坐标分量算符 X^j 的本征态，特征值为 x_0^j．同理可得 $e^{ik\cdot x}$ 是动量分量算符 $i\partial_{x^j}$ 的本征态，特征值为 $-k_j$; 也是自由粒子 Hamilton 算符 $-\Delta$ 的本征态，特征值为 $|k|^2$．动量本征态和位置本征态正是" 不可归一化的态矢量" 的例子．

这个例子正说明了为什么要引入装备 Hilbert 空间作为量子力学的舞台：物理上感兴趣的许多算符在数学意义下的 Hilbert 空间（即所有可归一化的态的空间）中可能没有本征态，或者本征态集合不完备; 但引入 RHS 之后，算符在对偶空间 Φ^* 就可以有足够多的本征态．在上面的例子中，任何可归一化的

态 $f \in L^2(\mathbb{R}^n)$ 都可以写成位置本征态的叠加（这里的积分在数学上的严格定义应该是 Schwartz 分布之间的卷积）：

$$f(x) = \int_{\mathbb{R}^n} f(y)\delta(x-y)dy;$$

或者动量本征态/自由 Hamilton 算符本征态的叠加（这里的积分取主值）：

$$f(x) = \int_{\mathbb{R}^n} \hat{f}(k)e^{ik\cdot x}dk.$$

这些都是物理意义十分直观的结论．这个结果并不是偶然的．它是 RHS 中自伴算子谱定理的特例．

宇称算符

预备知识 多元函数积分和宇称[343]

对多元函数 $f(\mathbf{x}) = f(x_1, \ldots, x_N)$．定义宇称算符 Π 如下

$$\Pi f(\mathbf{x}) = f(-\mathbf{x}) \tag{1}$$

若函数内积[471] 定义为（星号表示复共轭）

$$\langle f|g\rangle = \int f^*(\mathbf{x})g(\mathbf{x})\,\mathrm{d}^N x \tag{2}$$

容易证明宇称算符是一个厄米算符．对本征方程

$$\Pi f(\mathbf{x}) = \lambda f(\mathbf{x}) \tag{3}$$

容易证明[2]宇称算符的本征值 λ 只可能等于 1 或 -1．我们说 $\lambda = 1$ 的函数具有**偶宇称（even parity）**，$\lambda = -1$ 的函数具有**奇宇称（odd parity）**．它们分别满足

$$f(-\mathbf{x}) = \pm f(\mathbf{x}) \tag{4}$$

对于一元函数，具有奇宇称的就是**奇函数（odd function）**，具有偶宇称的就是**偶函数（even function）**．

[2]分别令 $\mathbf{x}$ 为某对称的两点 $\mathbf{x}_1$ 和 $-\mathbf{x}_1$，使函数值不为零．本征方程要求 $f(-\mathbf{x}_1) = \lambda f(\mathbf{x}_1)$ 且 $f(\mathbf{x}_1) = \lambda f(-\mathbf{x}_1)$，所以必有 $\lambda^2 = 1$，$\lambda = \pm 1$.

第二章

其他

堆放排列组合

比热力学课本上简单得多的方法推导出这种排列组合来.

题目是这样的，

如果有 n 个不加区分的小球，有 N 个有编号的盒子（$N \geqslant n$ ），那么把所有小球都放到盒子里有几种方法（每个盒子能装的个数没有限制）?

现在把所有的情况根据非空盒子的个数分类. 非空盒子个数可能为 1 个（n 个小球都在里面），2 个，一直到 n 个（每个盒子只装 1 个）. 如果用 i 个盒子装小球，那么首先从 N 个盒子里面选择 i 个会有 C_n^i 种情况. 然后要考虑的是，如果用已选的 i 个有编号盒子装 n 个小球，又有几种情况. 用所谓的插空法得到共有 C_{n-1}^{i-1} 种情况. 所以一个 i 对应 $C_N^i C_{n-1}^{i-1}$ 种情况.

最后把所有不同 i 的情况数加在一起，得出所有情况的总数为

$$\sum_{i=1}^{n} C_N^i C_{n-1}^{i-1} \tag{1}$$

又由于 $C_a^b = a!/[(a-b)!b!] = C_a^{a-b}$，上式可变为

$$\sum_{i=1}^{n} C_N^i C_{n-1}^{n-i} \tag{2}$$

又由于 $\sum_i C_a^i C_b^{n-i} = C_{a+b}^n$ （i 取所有可能的整数使得 $i \leqslant a$，$n-i \leqslant b$ ）（见），上式变为

$$\sum_{i=1}^{n} C_N^i C_{n-1}^{n-i} = C_{N+n-1}^n = \frac{(N+n-1)!}{n!(N-1)!} \tag{3}$$

这就是最后的答案.

范德蒙恒等式

预备知识 二项式定理[107]

[1]**范德蒙恒等式（Vandermonde's identity）**是指

$$\binom{a+b}{n}=\sum_i\binom{a}{i}\binom{b}{n-i}=\sum_i\binom{a}{n-i}\binom{b}{i} \tag{1}$$

其中求和是对所有使得表达式有意义的非负整数 i 进行的.

证明 1

假设有编了号的 $a+b$ 个小球. 不分顺序抓取 n 个，求总共有几种情况（用 N 表示）.

方法 1：根据定义，有 $N=\binom{a+b}{n}$ 种情况.

方法 2：先把球分成 A，B 两组，分别有 a 个和 b 个. 如果在 A 组中抽取 i 个球（有 $\binom{a}{i}$ 种情况），在 B 组中只能抽取 $n-i$ 个（有 $\binom{b}{n-i}$ 种情况），所以一个 i 对应 $\binom{a}{i}\binom{b}{n-i}$ 种情况. 所有可能的 i 一共有 $N=\sum_i\binom{a}{i}\binom{b}{n-i}$ 种情况.

由于这个问题只有一个答案，所以有 $\binom{a+b}{n}=\sum_i\binom{a}{i}\binom{b}{n-i}$.

但 i 的范围具体从多少取到多少，由 a，b 是否大于 n 来决定. 当 a,b 都大于 n 时，i 可以从 0 取到 n，如果其中至少有一个小于 n，那么 i 的取值不能使 C 的上标大于下标.

证毕.

[1]参考 Wikipedia 相关页面

证明 2

我们也可以通过二项式定理来证明

$$\begin{aligned}(1+x)^{m+n} &= \sum_{r=0}^{m+n} \binom{m+n}{r} x^r \\ &= (1+x)^m (1+x)^n \\ &= \sum_{0}^{m} \binom{m}{i} x^i \sum_{j=0}^{n} \binom{n}{j} x^j \\ &= \sum_{r=0}^{m+n} \sum_{k=0}^{r} \binom{m}{k} \binom{n}{r-k} x^r\end{aligned} \tag{2}$$

解三棱锥顶角

预备知识 高中立体几何

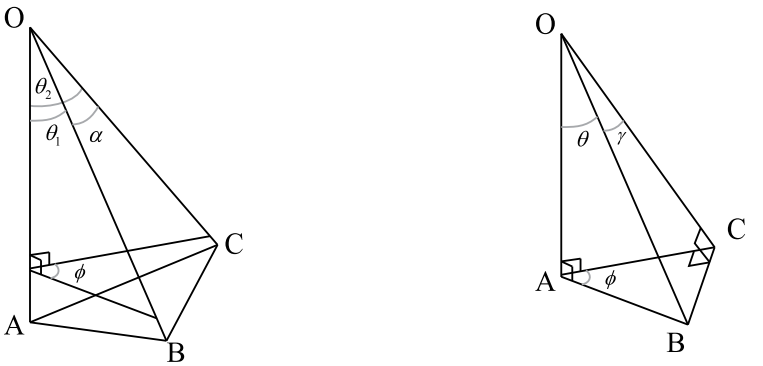


图 1:（左图）已知 θ_1，θ_2 和 ϕ 求 α；（右图）已知 θ 和 ϕ 求 γ

我们考虑如图 1 中三棱锥顶点 O 处的 3 条棱和 3 个面之间的角度关系. 这里涉及了三种角：两条棱之间的夹角（线线角）θ，棱和面的夹角（线面角）γ，以及面和面的夹角（面面角）ϕ. 对给定的顶点，每种角都有各有 9 个，共 9 个. 我们先不讨论底面 ABC 的位置如何，如果顶点的三条棱的方向都确定，我们就说顶点的形状确定.

显然，我们无需知道所有 9 个角的大小才能确定顶点的形状．这里给出一个类似于三角形余弦定理的公式，将线线角和面面角联系起来．线线角可以类比余弦定理中的边长，面面角类比余弦定理中的夹角．

$$\cos\alpha = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos\phi \tag{1}$$

如果已知该式中的 4 个角中的 3 个，就可以求出另一个角．与余弦定理类似，在求解 θ_1 或 θ_2 时可能存在两个解，也可能无解．注意如果我们交换 θ_1 和 θ_2 的值，上式仍然满足（这相当于创造一个镜像三棱锥）．

一个关于线面角的常用公式是

$$\sin\gamma = \sin\phi\sin\theta_1 \tag{2}$$

其中 γ 是线段 OB 和三角形 OAC 的线面角．

例 1　已知三个线线角求线面角

若已知 θ_1，θ_2 和 α，则

$$\cos\phi = \frac{\cos\alpha - \cos\theta_1\cos\theta_2}{\sin\theta_1\sin\theta_2} \tag{3}$$

证明

我们可以用球坐标系来证明式 1．令极坐标 (θ_1,ϕ_1) 和 (θ_2,ϕ_2) 代表的两个单位矢量的坐标分别为

$$(\sin\theta_1\cos\phi_1, \sin\theta_1\sin\phi_1, \cos\theta_1) \qquad (\sin\theta_2\cos\phi_2, \sin\theta_2\sin\phi_2, \cos\theta_2) \tag{4}$$

$$\cos\alpha = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos(\phi_2 - \phi_1) \tag{5}$$

证毕．

再来证明式 2，我们仍然使用球坐标，将射线 OA 作为极轴，面 OAC 为 xz 平面，线面角中“线”的极坐标为 $(1,\theta,\phi)$，该点的 y 坐标为 $\sin\theta\sin\phi$ 而 $y = \sin\gamma$，所以 $\sin\gamma = \sin\theta\sin\phi$．证毕．

足球顶点坐标的计算

预备知识 解三棱锥顶角[846]，空间旋转矩阵[404]



图 1: 足球透视图

足球的各种常用数据都可以由一个条件推出，那就是它有 60 个顶点．考虑到每个顶点都有 3 条同样长度的棱，棱的个数为

$$(60\text{ 个顶点}) \times (3\text{ 条棱})/(\text{每条棱有 }2\text{ 个顶点}) = 90 \tag{1}$$

另外每个顶点都由一个正五边形和两个正六边形共用，所以五边形的个数为

$$(60\text{ 个顶点}) \times (\text{每个顶点 }1\text{ 个五边形})/(\text{每个五边形有 }5\text{ 个顶点}) = 12 \tag{2}$$

同理，六边形的个数为

$$(60\text{ 个顶点}) \times (\text{每个顶点 }2\text{ 个六边形})/(\text{每个六边形有 }6\text{ 个顶点}) = 20 \tag{3}$$

坐标计算

以下所用到的方法不仅适用于足球，还适用于其他多面体如正十二面体，正二十面体等．以足球的一个五边形作为底面，得到足球的侧视图和俯视图如图 2．为了方便讨论，我们根据这两个视图把足球的顶点按高度分为 A 到 H 共 8 层．并且逆时针给出相应的数字编号．

由于足球由正六边形和正五边形拼成，足球的每一条棱都相等．在以下计算中，不妨设棱长为 1．通过以下几个步骤，就可以求解足球的各点的直角坐标．

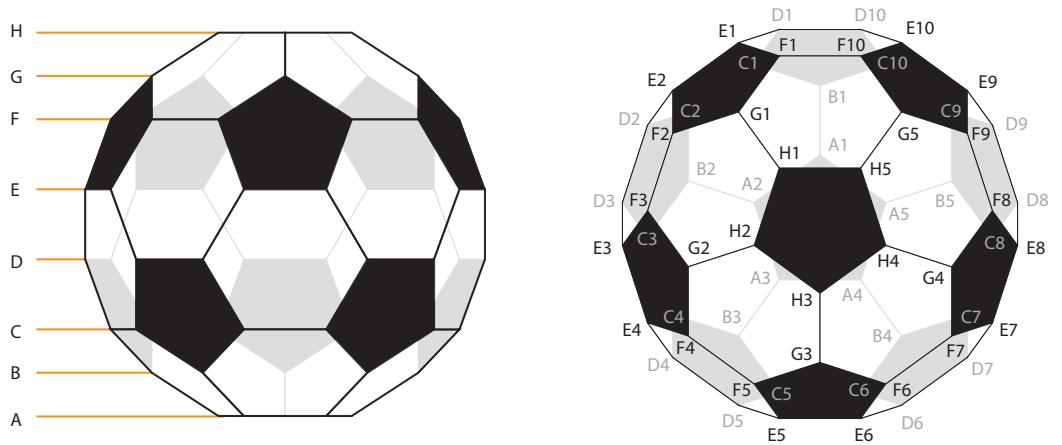


图 2: 各层和各顶点的命名，右图中黑色字体对应五边形的顶点，灰色字体对应六边形的顶点

求出 A_1 到 A_5 的坐标

直角坐标系中，令 A 层的五边形与 xy 平面重合，如下图. 为了表示方便，令

$$S=\sin\frac{2\pi}{5}\qquad C=\cos\frac{2\pi}{5}\qquad s=\sin\frac{\pi}{5}\qquad c=\cos\frac{\pi}{5}\tag{4}$$

则点 A_1 到 A_5 的坐标分别为

$$\left(0,\frac{1}{2s},0\right)\quad\left(-\frac{S}{2s},\frac{C}{2s},0\right)\quad\left(-\frac{1}{2},-\frac{c}{2s},0\right)\quad\left(\frac{1}{2},-\frac{c}{2s},0\right)\quad\left(\frac{S}{2s},\frac{C}{2s},0\right)\tag{5}$$

三个重要的角

由式 1 [846] 可得以下角的大小（以下每个角都是锐角）.

两个六边形之间的面面角满足

$$\sin\phi_{66}=\frac{2}{3}\tag{6}$$

五边形与六边形之间的面面角满足

$$\sin\phi_{56}=\frac{4s}{\sqrt{15}}\tag{7}$$

由式 2 [846]，五边形与相邻棱的线面角满足

$$\sin\gamma=\frac{2s}{\sqrt{5}}\tag{8}$$

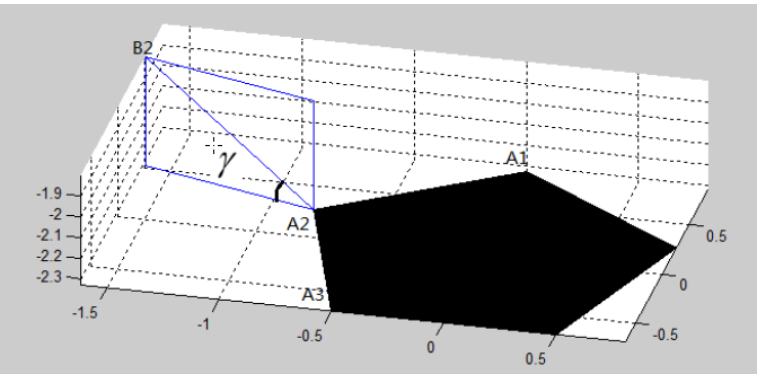


图 3: 求线段 A_1B_1 与底面的线面角

求出 B_1 的坐标

若把从 A_5 指向 A_1 的矢量记为 $\mathbf{v}_1$，A_2 指向 A_1 的矢量记为 $\mathbf{v}_2$，A_1 指向 B_1 的矢量记为 $\mathbf{u}$，可得

$$\mathbf{u}=\frac{\mathbf{v}_1+\mathbf{v}_2}{|\mathbf{v}_1+\mathbf{v}_2|}\cos\gamma+\frac{\mathbf{v}_2\times\mathbf{v}_1}{|\mathbf{v}_2\times\mathbf{v}_1|}\sin\gamma=\frac{\sqrt{1-3S^2/4}}{2s}(\mathbf{v}_1+\mathbf{v}_2)+\frac{\sqrt{3}}{2}(\mathbf{v}_2\times\mathbf{v}_1)\quad(9)$$

由式 5 得

$$\mathbf{v}_1=\frac{-S\hat{\mathbf{x}}+(C-1)\hat{\mathbf{y}}}{2s}\qquad\mathbf{v}_2=\frac{(-S+s)\hat{\mathbf{x}}+(C+c)\hat{\mathbf{y}}}{2s}\qquad(10)$$

代入可得 $\mathbf{u}$（略），加上 A_1 的坐标可得 B_1 的坐标．接下来我们并不需要用同样的方法求出 B 层的其他坐标，因为我们可以通过一个简单的矢量公式依次求出剩下所有顶点的坐标．

求剩下的坐标

已知空间中正 N 边形任意连续 3 点的坐标，求第 4 点的坐标，可以用以下矢量公式来计算

$$\mathbf{r}_4=\mathbf{r}_1-\mathbf{r}_2+\mathbf{r}_3+2\sin\left(\frac{N-4}{2N}\pi\right)\cdot(\mathbf{r}_3-\mathbf{r}_2)\qquad(11)$$

其中 $\mathbf{r}_i$ 是第 i 点的位置矢量[878]．

根据已知的 A_2, A_1, B_1 三个点，该式可以算出 C_1 的坐标，再把 A_1, B_1, C_1 的坐标代入得到 C_2，再由 B_1, C_1, C_2 得到 B_2，以此类推．当 C 层完成后，又可以同理算出其他层的所有顶点．下面我们来推导这个公式．

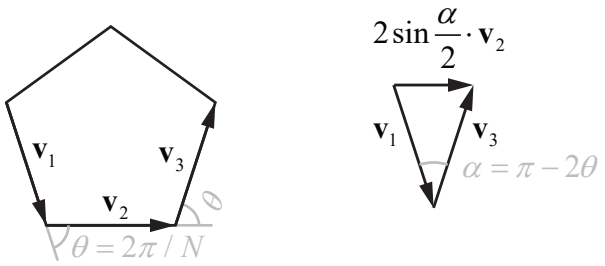


图 4: 由正 N 边形的 3 个顶点坐标求第 4 个顶点坐标．图中以五边形为例，但图中的公式适用于任意的 $N > 2$

令 $\mathbf{v}_i = \mathbf{r}_{i+1} - \mathbf{r}_i$，从图 4 中不难发现

$$\mathbf{v}_1 + \mathbf{v}_3 = 2\sin\left(\frac{N-4}{2N}\pi\right)\cdot\mathbf{v}_2 \tag{12}$$

将 $\mathbf{r}_i$ 代入即可得到式 11．

足球的高度和中点坐标

我们从底部的五边形出发，沿着任意一条外接棱的方向前进，依次会经过棱，五边形，六边形，六边形，顶部的五边形．如果从反方向出发，顺序恰好相反．这两条路线恰好可以绕足球一整圈．

按照第一种路线到达顶部，转过的角度依次是 $\gamma, \gamma, \phi_{56}, \phi_{66}, \phi_{56}$．可以验证，它们相加等于 π．

考虑到经过的五边形（边长为 1）的中轴线长度是 $(1+c)/(2s)$，经过的六边形的中轴线长度是 $\sqrt{3}$，经过的棱长为 1，底部五边形到顶部五边形的高度为[2]

$$\begin{aligned} h &= \sin\gamma + \frac{1+c}{2s}\sin(2\gamma) + \sqrt{3}\sin(2\gamma+\phi_{56}) + \sqrt{3}\sin(2\gamma+\phi_{56}+\phi_{66}) \\ &= \sqrt{\frac{41}{10}\sqrt{5}+\frac{25}{2}} \end{aligned} \tag{13}$$

球心到底面和顶面的距离为 $h/2$．

由勾股定理可以得到中心点到任意顶点的距离为

$$R = \sqrt{\left(\frac{h}{2}\right)^2 + \left(\frac{1}{2s}\right)^2} = \frac{1}{4}\sqrt{18\sqrt{5}+58} \tag{14}$$

[2]事实上这个式子很难化简，结果来自 Wikipedia 查找并通过数值验证．

Clebsch-Gordan 系数

- 本词条需要更多讲解，便于帮助理解.

预备知识 角动量加法[1403]

相位约定

任何算符的一组归一化本征基底各自乘以一个任意相位因子 $e^{i\phi_I}$ 仍然是一组归一化的本征基底（本征值不变）. 所以基底变换矩阵（如 CG 矩阵），每一行或每一列分别乘以一个相位因子，仍然表示相同的基底变换. 这些相位怎么取被称为相位约定（Phase Convention）. 一般的约定是使所有 CG 系数为实数，且第一行和第一列（对应最大的 m_1 和最大的 L）的矩阵元大于零（Wikipedia，Griffiths）.

解析表达式为

按照这种相位约定，CG 系数的解析表达式为

$$\begin{aligned}
&\langle l_1, m_1, l_2, m_2 | l_1, l_2, L, M\rangle = \begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{bmatrix} \\
&= \sqrt{\frac{(2L+1)(L+l_1-l_2)!(L-l_1+l_2)!(l_1+l_2-L)!}{(l_1+l_2+L+1)!}} \\
&\times \sqrt{(L+M)!(L-M)!(l_1-m_1)!(l_1+m_1)!(l_2-m_2)!(l_2+m_2)!} \\
&\times \sum_{k=k_{min}}^{k_{max}} \frac{(-1)^k}{k!(l_1+l_2-L-k)!(l_1-m_1-k)!(l_2+m_2-k)!} \\
&\times \frac{1}{(L-l_2+m_1+k)!(L-l_1-m_2+k)!}
\end{aligned} \tag{1}$$

其中求和的上下限需要保证每个含有 k 的阶乘自变量都大于等于 0，所以有

$$\begin{aligned}
k_{min} &= \max\{0,\ \ l_2-m_1-L,\ \ l_1+m_2-L\} \\
k_{max} &= \min\{l_1+l_2-L,\ \ l_1-m_1,\ \ l_2+m_2\}
\end{aligned} \tag{2}$$

基本选择定则

从“角动量加法[1403]”中可以得到

$$|l_1 - l_2| \leqslant L \leqslant l_1 + l_2 \tag{3}$$

从物理上理解，两个矢量相加得到第三个矢量 $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}$，则他们的模长 v_1, v_2, v 必须满足三角不等式 $|v_1 - v_2| \leqslant v \leqslant v_1 + v_2$.

我们还可以得到

$$m_1 + m_2 = M \tag{4}$$

$$-L \leqslant M \leqslant L \qquad -l_1 \leqslant m_1 \leqslant l_1 \qquad -l_2 \leqslant m_2 \leqslant l_2 \tag{5}$$

当有可能出现半整数时，还要求[3]

$$M + L \in \mathbb{N} \qquad m_1 + l_1 \in \mathbb{N} \qquad m_2 + l_2 \in \mathbb{N} \tag{6}$$

当且仅当一个 CG 系数满足以上选择定则时，它才会在 CG 表中出现（可能是 0）.

对称性

由于 3j 符号[855] 具有更简洁的对称性，我们可以用其推导 CG 系数的对称性[4]

$$\begin{aligned}
\begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{bmatrix} &= (-1)^{l_1+l_2-L} \begin{bmatrix} l_1 & l_2 & L \\ -m_1 & -m_2 & -M \end{bmatrix} \\
&= (-1)^{l_1+l_2-L} \begin{bmatrix} l_2 & l_1 & L \\ m_2 & m_1 & M \end{bmatrix} \\
&= (-1)^{l_1-m_1} \sqrt{\frac{2L+1}{2l_2+1}} \begin{bmatrix} l_1 & L & l_2 \\ m_1 & -M & -m_2 \end{bmatrix} \\
&= (-1)^{l_2+m_2} \sqrt{\frac{2L+1}{2l_1+1}} \begin{bmatrix} L & l_2 & l_1 \\ -M & m_2 & -m_1 \end{bmatrix}
\end{aligned} \tag{7}$$

[3]推导过程中注意所有数都可能取半整数. 若两个数都是半整数或都是整数，它们相加等于整数，否则相加等于半整数. 半整数乘以 2 等于奇数，整数乘以 2 等于偶数. 类比：若两个数都是奇数或偶数，它们相加等于偶数，否则相加等于奇数.

[4]下式中如果 L 是整数，那么前面取正负号都行，但如果是半整数，则只能取负号，推导时要注意.

对称性选择定则

符合基本选择定则的 CG 系数也可能为 0．我们可以用更多的选择定则来找到这些系数．由 3j 符号[855] 的选择定则（式 7，式 8），当第一行三个数之和为奇数时[5]

$$\begin{bmatrix} l_1 & l_2 & L \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} l & l & L \\ m & m & M \end{bmatrix} = \begin{bmatrix} l_1 & l & l \\ m_1 & m & -m \end{bmatrix} = \begin{bmatrix} l & l_2 & l \\ m & m_2 & -m \end{bmatrix} = 0 \tag{8}$$

注意即使加入了这些选择定则，CG 系数仍然可能为零，事实上要找到所有的零系数非常困难[6]．

特殊情况

一些情况下式 1 有更简单的表达式

$$\begin{bmatrix} l_1 & l_2 & 0 \\ m_1 & m_2 & 0 \end{bmatrix} = \frac{(-1)^{l_1-m_1}}{\sqrt{2l_1+1}} \tag{9}$$

正交归一性

由于 $|l_1, l_2, L, M\rangle$ 都是正交归一的，所以有

$$\begin{aligned} &\sum_{m_1,m_2} \langle l_1', l_2', L', M'|l_1, m_1, l_2, m_2\rangle \langle l_1, m_1, l_2, m_2|l_1, l_2, L, M\rangle \\ &= \delta_{l_1,l_1'}\delta_{l_2,l_2'}\delta_{L,L'}\delta_{M,M'} \end{aligned} \tag{10}$$

即

$$\sum_{m_1,m_2} \begin{bmatrix} l_1' & l_2' & L' \\ m_1 & m_2 & M' \end{bmatrix} \begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{bmatrix} = \delta_{l_1,l_1'}\delta_{l_2,l_2'}\delta_{L,L'}\delta_{M,M'} \tag{11}$$

根据选择定则，只需要对满足式 4 和式 5 的 m_1 和 m_2 即可（双重求和变为一次求和）．

[5] 由 CG 系数的对称性推导得到的条件会复杂很多，但与该条件是等价的．

[6] 见 T A Heim, J Hinze and A R P Rau, Some classes of ''nontrivial zeroes' of angular momentum addition coefficients，一个例子是 $\begin{bmatrix} 3 & 3 & 2 \\ -2 & 2 & 0 \end{bmatrix}$，注意第一行之和是偶数．

与球谐函数的关系

三个球谐函数之积的积分可以表示成两个 CG 系数或 3j 符号[855] 相乘[7]

$$\begin{aligned}&\int Y_{l_1m_1}(\hat{\mathbf{r}})Y_{l_2m_2}(\hat{\mathbf{r}})Y_{l_3m_3}(\hat{\mathbf{r}})\,\mathrm{d}\Omega\\&=(-1)^{m_3}\sqrt{\frac{(2l_1+1)(2l_2+1)}{4\pi(2l_3+1)}}\begin{bmatrix}l_1&l_2&l_3\\0&0&0\end{bmatrix}\begin{bmatrix}l_1&l_2&l_3\\m_1&m_2&-m_3\end{bmatrix}\\&=\sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}}\begin{pmatrix}l_1&l_2&l_3\\0&0&0\end{pmatrix}\begin{pmatrix}l_1&l_2&l_3\\m_1&m_2&m_3\end{pmatrix}\end{aligned}\tag{12}$$

Wigner 3j 符号

预备知识 CG 系数[852]

每个 3j 符号和 CG 系数都一一对应且相差一个常数，用圆括号表示 3j 符号，方括号表示 CG 系数，有

$$\begin{pmatrix}j_1&j_2&j_3\\m_1&m_2&m_3\end{pmatrix}=\frac{(-1)^{j_1-j_2-m_3}}{\sqrt{2j_3+1}}\begin{bmatrix}j_1&j_2&j_3\\m_1&m_2&-m_3\end{bmatrix}\tag{1}$$

对称性

3j 符号具有很好的对称性．首先，任意交换两列等于在前面加 $(-1)^{j_1+j_2+j_3}$

$$\begin{aligned}&\begin{pmatrix}j_3&j_2&j_1\\m_3&m_2&m_1\end{pmatrix}=\begin{pmatrix}j_1&j_3&j_2\\m_1&m_3&m_2\end{pmatrix}=\begin{pmatrix}j_2&j_1&j_3\\m_2&m_1&m_3\end{pmatrix}\\&=(-1)^{j_1+j_2+j_3}\begin{pmatrix}j_1&j_2&j_3\\m_1&m_2&m_3\end{pmatrix}\end{aligned}\tag{2}$$

如果交换两次，3j 符号不变

$$\begin{pmatrix}j_1&j_2&j_3\\m_1&m_2&m_3\end{pmatrix}=\begin{pmatrix}j_2&j_3&j_1\\m_2&m_3&m_1\end{pmatrix}=\begin{pmatrix}j_3&j_1&j_2\\m_3&m_1&m_2\end{pmatrix}\tag{3}$$

[7] 见 Bransden 附录 A4，以及 Wikipedia 的 3j/CG coefficients 页面

将第二行取相反数也等于在前面加 $(-1)^{j_1+j_2+j_3}$

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ -m_1 & -m_2 & -m_3 \end{pmatrix} = (-1)^{j_1+j_2+j_3} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \tag{4}$$

选择定则

3j 符号的选择定则直接告诉我们哪些 3j 符号等于 0．有了选择定则，我们就无需计算不符合定则的 3j 符号．

从 CG 系数的选择定则可得**三角不等式**（三个不等式等效）

$$|j_1 - j_3| \leqslant j_2 \leqslant j_1+j_3 \qquad |j_2 - j_3| \leqslant j_1 \leqslant j_2+j_3 \qquad |j_3 - j_1| \leqslant j_2 \leqslant j_3+j_1 \tag{5}$$

以及

$$m_1 + m_2 + m_3 = 0 \tag{6}$$

除此之外，以上每个对称性也可以得到一个选择定则：当 $j_1+j_2+j_3$ 为奇数时，如果任意两列相同，结果为 0

$$\begin{pmatrix} j & j & j_3 \\ m & m & m_3 \end{pmatrix} = \begin{pmatrix} j_1 & j & j \\ m_1 & m & m \end{pmatrix} = \begin{pmatrix} j & j_2 & j \\ m & m_2 & m \end{pmatrix} = 0 \tag{7}$$

当 $j_1+j_2+j_3$ 为奇数时，若 $m_1 = m_2 = m_3 = 0$，结果也为 0

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ 0 & 0 & 0 \end{pmatrix} = 0 \tag{8}$$

与球谐函数的关系

三个球谐函数之积的积分可以表示成两个 CG 系数[852] 或 3j 符号相乘[8]

$$\begin{aligned} &\int Y_{l_1m_1}(\hat{\mathbf{r}}) Y_{l_2m_2}(\hat{\mathbf{r}}) Y_{l_3m_3}(\hat{\mathbf{r}}) \,\mathrm{d}\Omega \\ &= (-1)^{m_3} \sqrt{\frac{(2l_1+1)(2l_2+1)}{4\pi(2l_3+1)}} \begin{bmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & -m_3 \end{bmatrix} \\ &= \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \end{aligned} \tag{9}$$

[8] 见 Bransden 附录 A4，以及 Wikipedia 的 3j/CG coefficients 页面

Wigner 6j 符号

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix} = \sum_{m_1,\ldots,m_6} (-1)^{\sum_{k=1}^{6}(j_k-m_k)} \begin{pmatrix} j_1 & j_2 & j_3 \\ -m_1 & -m_2 & -m_3 \end{pmatrix} \times \begin{pmatrix} j_1 & j_5 & j_6 \\ m_1 & -m_5 & m_6 \end{pmatrix} \begin{pmatrix} j_4 & j_2 & j_6 \\ m_4 & m_2 & -m_6 \end{pmatrix} \begin{pmatrix} j_4 & j_5 & j_3 \\ -m_4 & m_5 & m_3 \end{pmatrix} \tag{1}$$

Wigner 9j 符号

预备知识　6j 符号[857]

9j 符号可以用 6j 符号定义为

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \\ j_7 & j_8 & j_9 \end{Bmatrix} = \sum_i (-1)^{2i}(2i+1) \begin{Bmatrix} j_1 & j_4 & j_7 \\ j_8 & j_9 & i \end{Bmatrix} \begin{Bmatrix} j_2 & j_5 & j_8 \\ j_4 & i & j_6 \end{Bmatrix} \begin{Bmatrix} j_3 & j_6 & j_9 \\ i & j_1 & j_2 \end{Bmatrix} \tag{1}$$

对称性

任意两行行或两列交换，在前面添加 $(-1)^S$，其中 S 是 9 个 j 之和.

特殊情况

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \\ j_7 & j_8 & 0 \end{Bmatrix} = \frac{\delta_{j_3,j_6}\delta_{j_7,j_8}}{\sqrt{(2j_3+1)(2j_7+1)}}(-1)^{j_2+j_3+j_4+j_7} \begin{Bmatrix} j_1 & j_2 & j_3 \\ j_5 & j_4 & j_7 \end{Bmatrix} \tag{2}$$

二元关系

预备知识　集合[92]

关系

在“集合[92]”中我们只关心了集合的基数，即集合中元素的数目．在这种语境下，任何两个元素数量相同的集合都可以看作是同一个集合．但是仅仅讨论集合的基数未免太过单调，缺少了很多有意思的理论，于是我们希望在集合的元素之间建立一些结构，来进行更细致的划分和研究．

关系（relation）是集合上最基础的一种结构．给定一个关系，我们就可以讨论一些元素之间**是否满足**这个关系．比如说，如果取一家三口构成一个集合，$\sim$ 代表的是“年龄大于”，那么我们可以说“爸爸对于孩子具有这个关系”，但是反过来“孩子对于爸爸不具有这个关系”．从这个例子可以看出，关系的表达方式很灵活，而且可以是有方向性的．讨论关系时，我们唯一关心的是给定元素之间是否具有这样的关系．

关系可以用在两个元素之间，也可以用在三个元素之间，甚至可以用在不特定的元素之间．

二元关系

在绝大多数数学和物理领域，我们只关心集合上的**二元关系（binary relation）**．如果 $\sim$ 是在集合 A 上定义的一个二元关系，那么任意给定两个元素，我们都可以讨论它们之间是否具有这种关系，但如果给定三个元素，讨论就没有意义了．比如，如果 $\sim$ 的定义是“年龄大于”，那么把三个人的年龄都拿过来比较就没有意义；不过，如果 $\sim$ 的定义是“比后面两个人的年龄都大”，那么 $\sim$ 就可以用在三个人身上．

对于集合 A 上的二元关系 $\sim$，如果 $x, y \in A$ 满足这个关系，我们可以把这句话表述为 $x \sim y$．如果不满足，则可以表述为 $x \not\sim y$．用这样简洁的表示方法，我们可以把以上“年龄大于”的关系表述如下：

$$\text{爸爸} \sim \text{孩子} \qquad \text{妈妈} \sim \text{孩子} \qquad \text{孩子} \not\sim \text{爸爸} \qquad \text{孩子} \not\sim \text{妈妈} \tag{1}$$

爸爸和妈妈之间，在没有进一步信息的情况下，无法判断是否满足这个关系，但这个关系是存在的．也就是说，元素间的关系一共有两种状态：满足和不满足．

等价关系

最基础的一类关系，是**等价关系（equivalence）**．在集合 A 上定义的关系 $\sim$ 是一个等价关系，要求满足条件：

1. **自反性**：$\forall x \in A, x \sim x$；
2. **对称性**：$\forall x, y \in A, x \sim y \Rightarrow y \sim x$；
3. **传递性**：$\forall x, y, z \in A, x \sim y, y \sim z \Rightarrow x \sim z$．

如果把“具有关系 $\sim$”看成是两个元素间相互连接（没有方向性，因为对称性），传递性保证了当多个元素相连时，这些元素也两两互联；自反性甚至保证了每个元素必然和自身相连．由此一来，我们可以根据等价关系把集合 A 划分成**等价类（equivalence class）**，每个等价类都是 A 的一个子集，所包含的元素彼此相连．

等价类的划分是数学中非常重要的思维方法，它可以将每一个等价类中的元素都看成是无差别的，大大简化一个集合的复杂程度．有了等价关系后，我们可以用相应的等价类来组成一个新的集合，这样的集合被称作**商集（quotient set）**．和原来集合中的元素不一样，商集中的元素是原集合中的子集．

作为例子，我们取全体非负整数的集合 $\mathbb{Z}$，并在上面定义关系 $\sim$ 为“两数的差是 3 的倍数”，那么容易验证，$1 \sim 4, 7 \sim 304, 0 \not\sim 77, 77 \not\sim 1$．利用这个等价关系，我们可以把非负整数集合划分成三个等价类，分别是 $\{0, 3, 6, 9, 12 \cdots\}$，$\{1, 4, 7, 10, 13 \cdots\}$ 和 $\{2, 5, 8, 11, 14 \cdots\}$．这样，我们可以得到一个含有三个元素的商集，用数学专业术语来说，叫做集合 A 模去关系 $\sim$ 所得的商集．这里的**模（mod）**的意思来自于“除法”，在群论[541] 中会看到为什么用除法来命名商集．

群论笔记

参考文献：Mathematical Methods for Physicists (Arfken et. al.)

17.1 Introduction to Group Theory

- **群**的定义：定义了乘法的集合，(1) 闭合，(2) 结合律，(3) 单位元，(4) 逆元
- **阶数**：离散群的阶数等于元素数，连续群的阶数等于参数的个数
- 满足交换律的群就叫**阿贝尔（abelian）**群
- **循环（cyclic）**群：群内的所有元素都可以用某个元素表示为 $I, a, a^2, \ldots$
- 循环群是阿贝尔群
- 两个群中的元素如果一一对应（包括乘法运算），则他们是 **isomorphic** 的，如果是多对一，就是 **homomorphic** 的
- 如果一个群的子集也是一个群，就称为**子群（subgroup）**

17.2 Representation of Groups

17.3 Symmetry and Physics

17.4 Distrete Groups

17.5 Distrete Products

- **直积**：两个群 G 和 H 通过直积运算得到一个新群 $G \otimes H$，元素为 (g, h)，其中 $g \in G$，$h \in H$．$G \otimes H$ 中的乘法运算为对应元素分别运算 $(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1 h_2)$，注意两个元素的乘法可能不一样．

17.6 Symmetric Group

- **对称群（symmetric group）** S_n 是 n 个不同对象的 permutation 操作组成的群，所以共有 $n!$ 个元素

17.7 Continuous Groups

- $SO(n)$（**special orthogonal**）表示 n 维空间旋转群（$n(n-1)/2$ 阶），如果加上 reflection，就是 $O(n)$.
- $SU(n)$（**special unitary**），阶数是 n^2-1
- **李群（Lie group）：**

范畴论

- 本词条缺少预备知识，初学者可能会遇到困难.
- 本词条需要更多参考文献.
- 本词条需要更多讲解，便于帮助理解.

基本概念

概括地说，范畴是由对象（Object）及其间的态射（Morphism）组成的数学结构.

定义 1　范畴

范畴 $\mathcal{C}$ 包含以下结构.

1. 集合 $\mathrm{Ob}(\mathcal{C})$，其元素称作 $\mathcal{C}$ 的对象；
2. 集合 $\mathrm{Mor}_{\mathcal{C}}$，其元素称作 $\mathcal{C}$ 的态射，配上一对映射 $\mathrm{Mor}(\mathcal{C}) \rightrightarrows_a$；
3. 对于每三个元素 $x,y,z \in \mathrm{Ob}(\mathcal{C})$，存在态射复合映射 $\circ : \mathrm{Mor}_{\mathcal{C}}(Y,Z) \times \mathrm{Mor}_{\mathcal{C}}(X,Y) \to \mathrm{Mor}_{\mathcal{C}}(X,Z)$.

这些结构满足以下规则.

1. 对于每个元素 $X \in \mathrm{Ob}(\mathcal{C})$，存在单位态射 $\mathrm{id}_X \in \mathrm{Mor}_{\mathcal{C}}(X,X)$，使得 $\mathrm{id}_X \circ \phi = \phi$ 和 $\psi \circ \mathrm{id}_X = \psi$ 成立，其中 $\phi \in \mathrm{Mor}_{\mathcal{C}}(Y,X)$ 对于任意 $Y \in \mathrm{Ob}(\mathcal{C})$，$\psi \in \mathrm{Mor}_{\mathcal{C}}(X,Z)$ 对于任意 $Z \in \mathrm{Ob}(\mathcal{C})$；

2. 复合具有结合性,即对于 $\phi \in \mathrm{Mor}_{\mathcal{C}}(Y,Z), \psi \in \mathrm{Mor}_{\mathcal{C}}(X,Y), \chi \in \mathrm{Mor}_{\mathcal{C}}(W,X)$，有 $(\phi \circ \psi) \circ \chi = \phi \circ (\psi \circ \chi)$；

3. 任意态射集合 $\mathrm{Mor}_{\mathcal{C}}(X,Y)$ 和 $\mathrm{Mor}_{\mathcal{C}}(X',Y')$ 不交，除非 $X = X', Y = Y'$，此时它们相等.

由于以下练习，我们讨论单位态射的时候可以省去选择，只讨论唯一的单位映射.

习题 1

证明同个元素的任意两个单位态射是相等的.

有时我们也会使用 dom （**域**）和 cod （**陪域**）来讨论态射. 值得注意的是，我们一般情况下要求元素为一个集合而不是一个类，仅在如下范畴中我们放宽要求允许元素为一个真类：

- 集合范畴 (Set)，态射为映射；
- 交换群范畴 (Ab)，态射为交换群同态；
- 群范畴 (Grp)，态射为群同态；
- 左 G-群作用范畴 (G−Set)，态射为 G-映射；
- 对于含幺交换环 R，模范畴 (R−Mod)，态射为模同态；
- 对于域 F，向量空间范畴 (F−Vect)，态射为线性映射；
- 拓扑空间范畴 (Top)，态射为连续映射；
- 含幺交换环范畴 (Ring)，态射为环同态.

我们称 Ob 为一集合的范畴**集合范畴**，称其为**真类范畴**若其 Ob 不构成集合. 另外，若一真类范畴的每个态射类都构成集合，我们称它是一个**局部集合范畴**. 为简化我们所使用的符号，若所讨论的范畴在上下文中没有歧义，可使用符号 Ob 和 $\mathrm{Mor}(X,Y)$；另对于态射 $\phi \in \mathrm{Mor}(X,Y)$，我们通常使用映射符号将其写为 $\phi : X \to Y$.

定义 2　范畴论同构

对于一个态射 $\phi : X \to Y$，若存在另一个态射 $\psi : Y \to X$ 使得 $\phi \circ \psi = \mathrm{id}_Y$ 和 $\psi \circ \phi = \mathrm{id}_X$ 成立，则我们说它是一个范畴论同构．此时 ϕ 也被称为一个可逆态射，而 $\phi^{-1} = \psi$ 则被称为它的逆．

若一态射的域和陪域相同，则它可被称为一个**自态射**，元素 $X \in \mathrm{Ob}$ 的所有自态射形成一个幺半群，被称为 $\mathrm{Endo}_{\mathcal{C}}(X)$；若它还是一个同态，则被称为**自同态**，所有元素 X 的自同态形成一个群，被称为 $\mathrm{Auto}_{\mathcal{C}}(X)$．

定义 3　群胚

若范畴 $\mathcal{C}$ 的所有态射都是范畴论同构，则称范畴 $\mathcal{C}$ 为一群胚．

群胚是群的一般化结构．一个群 G 可以生成一个群胚，使得其结构被保留：这个群胚由一个元素 X 构成，其仅有的态射为 $\mathrm{Mor}(X, X) = G$，态射复合被定义为 $a \circ b = ab$，请自行验证如此定义的范畴为一个群胚．集合 A 也可生成群胚，定义

$$\mathrm{Ob} = A \quad \mathrm{Mor}(X, Y) = \begin{cases} \emptyset & X \neq Y \\ \{\mathrm{id}\} & X = Y \end{cases} \tag{1}$$

请自行验证它是一个群胚．

定义 4　函子

范畴 $\mathcal{A}, \mathcal{B}$ 间的协变函子 $F : \mathcal{A} \to \mathcal{B}$ 包含以下结构．

1. 若 $\mathcal{A}$ 和 $\mathcal{B}$ 均为集合范畴，则函子包含一个映射 $F_{\mathrm{Ob}}^{\mathrm{Set}} : \mathrm{Ob}(\mathcal{A}) \to \mathrm{Ob}(\mathcal{B})$，否则函子包含一个从 $\mathrm{Ob}(\mathcal{A})$ 到 $\mathrm{Ob}(\mathcal{B})$ 的关联 $F_{\mathrm{Ob}}^{\mathrm{PrCl}}$；

2. 函子包含一个映射 $F_{\mathrm{Mor}} : \mathrm{Mor}_{\mathcal{C}}(X, Y) \to \mathrm{Mor}_{\mathcal{B}}(F_{\mathrm{Ob}}^{\mathrm{Set/PrCl}}(X), F_{\mathrm{Ob}}^{\mathrm{Set/PrCl}}(Y))$，对于每个 $x, y \in \mathrm{Ob}(A)$．

函子满足如下公理．

1. 对于任意元素 $X \in \mathrm{Ob}(A)$，$F(\mathrm{id}_X) = \mathrm{id}_{F(x)}$；

2. 对于任意可复合的元素对 (ϕ, ψ)，$F(\phi \circ_{\mathcal{A}} \psi) = F(\phi) \circ_{\mathcal{B}} F(\psi)$．

我们仅考虑域和陪域为上方列出的真类范畴的具有 $F_{\mathrm{Ob}}^{\mathrm{PrCl}}$ 的函子．我们考虑的大部分范畴和函子都是**数学对象**（可用 ZFC 公理化集合论描述的对象）．注意若函子对象关系的类型（映射或关联）明确，我们可以将其省略．另外我们也会在语义明确时省略下标 Ob 和 Mor．任何范畴 $\mathcal{A}$ 都有一个显然的单位函子 $\mathrm{id}_{\mathcal{A}}$，同时对于两个函子 $F:\mathcal{A}\to\mathcal{B}$ 和 $G:\mathcal{B}\to\mathcal{C}$，我们可以定义它们的复合．

TODO opposite category and contravariant functor

TODO natural transformation

Hom 函子和 Yoneda 引理

定义 5　Hom 函子

若 $\mathcal{C}$ 是一个局部集合范畴，对于每个 $A\in\mathcal{C}$，记 A-协变 Hom 函子为 $h^A=\mathrm{Hom}(A,-):\mathcal{C}\to\mathrm{Set}$，其定义如下：

- *若 $X\in\mathrm{Ob}(\mathcal{C})$，则 $h^A(X)=\mathrm{Mor}_{\mathcal{C}}(A,X)\in\mathrm{Ob}(\mathrm{Set})$；*
- *若 $f\in\mathrm{Mor}_{\mathcal{C}}(X,Y)$，则对于 $g:A\to X$，有 $h^A(f)=g\mapsto f\circ g:A\to Y\in h^A(Y)$．*

记 B-逆变 Hom 函子为 $h_B=\mathrm{Hom}(-,B):\mathcal{C}\to\mathrm{Set}$，其定义如下：

- *若 $X\in\mathrm{Ob}(\mathcal{C})$，则 $h_A(X)=\mathrm{Mor}_{\mathcal{C}}(X,B)\in\mathrm{Ob}(\mathrm{Set})$；*
- *若 $f\in\mathrm{Mor}_{\mathcal{C}}(X,Y)$，则对于 $g:X\to B$，有 $h_A(f)=g\mapsto g\circ f:Y\to B\in h_A(Y)$．*

若我们使用现代代数几何的语言，如上定义的逆变 Hom 函子还可被称为一个**预层**．

定义 6　Hom 函子范畴

我们使用 $\mathrm{Set}^{\mathcal{C}}$ 来记协变 Hom 函子所构成的范畴，对于每个 $\mathcal{C}$ 中的态射 $\phi:A'\to A$，我们可以获得一个自然变换 $h_{\mathrm{co}}(\phi):(h^A:\mathcal{C}\to\mathrm{Set})\Rightarrow(h^{A'}:\mathcal{C}\to\mathrm{Set})$．我们可以简单地看出，下图是交换的对于任意 $\psi\in\mathrm{Mor}_{\mathcal{C}}(X,Y)$，于是 $h_{\mathrm{co}}(\phi)$ 的确是一自然变换．类似地，我们可以定义逆变 Hom 函子所构成的范畴 $\mathrm{Set}^{\mathcal{C}^{\mathrm{op}}}$，唯一不同的一点是我们需要给出一个相反的映射 $\phi:B\to B'$ 来使得

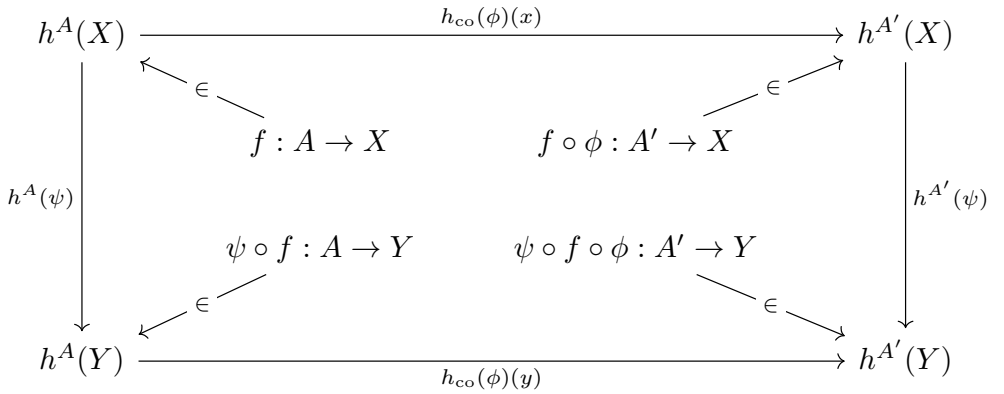


图 1: $h_{\mathrm{co}}(\phi)$ 的自然性

下图交换. 注意下图中 $h_B(-)$ 是一逆变函子，所以关于图交换的结论是对于 $\psi : Y \to X$ 而言的.

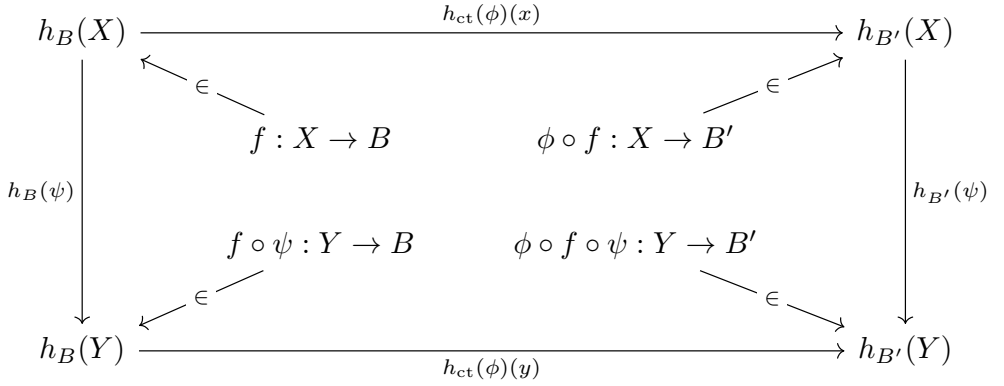


图 2: $h_{\mathrm{ct}}(\phi)$ 的自然性

如上定义的 $\mathrm{Set}^{\mathcal{C}^{\mathrm{op}}}$ 在代数几何中也被记为 $\mathrm{PSh}(\mathcal{C})$.

习题 2

请验证 $h^- : \mathcal{C} \to \mathrm{Set}^{\mathcal{C}}$ 和 $h_- : \mathcal{C} \to \mathrm{Set}^{\mathcal{C}^{\mathrm{op}}}$ 为函子. 提示：观察上述定义，可以注意到前者为逆变函子，后者为协变函子.

定理 1　Yoneda 引理

对于任意局部集合范畴 $\mathcal{C}$ 和 $A, B \in \mathrm{Ob}(\mathcal{C})$，

1. 给定协变函子 $F : \mathcal{C} \xrightarrow{\mathrm{co}} \mathrm{Set}$，存在双射 $\mathrm{Nat}(h^A, F) \simeq F(A)$；
2. 给定逆变函子 $G : \mathcal{C} \xrightarrow{\mathrm{ct}} \mathrm{Set}$，存在双射 $\mathrm{Nat}(h_B, G) \simeq G(B)$

证明. 对于自然变换 $\Phi : h^A \Rightarrow F$，我们将它映射到 $u = m(\Psi) = \Phi_A(\mathrm{id}_A)$（注意此处 Φ_A 是 Set 中 $h^A(A) \to F(A)$ 的态射，即一个映射）；对于 $v \in F(A)$，将它映射到如 $\Psi_X(f : \mathrm{Mor}_{\mathcal{C}}(A, X)) = F(f)(v)$ 定义的自然变换 $n(v) = \Psi$. 以下练习证明了 $m \circ n = \mathrm{id}$ 和 $n \circ m = \mathrm{id}$，即得所需双射. 另一部分的证明留作练习.

习题 3

请尝试证明在上述证明中，$m \circ n = \mathrm{id}$ 和 $n \circ m = \mathrm{id}$.

解答.

- $m \circ n$：若 $v \in F(A)$，则 $n(v) = \Psi$，其中 $\Psi_X(f) = F(f)(v)$，我们有 $m(\Psi) = \Psi_A(\mathrm{id}_A) = F(id_A)(v) = id_{F(A)}(v) = v$；
- $n \circ m$：若 $\Phi \in \mathrm{Nat}(h^A, F)$，则下图对任意 $\phi : X' \to Y'$ 和 $\psi : A \to X'$ 交换，即 $F(\phi)(\Phi_{X'}(\psi)) = \Phi_{Y'}(h^A(\phi)(\psi))$ 成立. 取 $X' = A$, $Y' = X$, $\phi = f : A \to X$, $\psi = \mathrm{id}_A : A \to A$，则有 $F(f)(\Phi_A(\mathrm{id}_A)) = \Phi_X(h^A(f)(\mathrm{id}_A)) = \Phi_X(f \circ \mathrm{id}_A) = \Phi_X(f)$. 此时计算 $(n \circ m)(\Phi), m(\Phi) = \Phi_A(\mathrm{id}_A) = u, n(u) = \Psi$，其中 $\Psi_X(f) = F(f)(\Phi_A(\mathrm{id}_A))$. 据前文，有 $\Phi_X(f) = F(f)(\Phi_A(\mathrm{id}_A)) = \Psi_X(f)$，即得 $\Psi = \Phi$.

$$\begin{array}{ccc} h^A(X') & \xrightarrow{\Phi_X} & F(X') \\ {\scriptstyle h^A(\phi)}\downarrow & & \downarrow{\scriptstyle F(\phi)} \\ h^A(Y') & \xrightarrow[\Phi_Y]{} & F(Y') \end{array}$$

图 3: 此图交换

习题 4

模仿上述证明，证明 Yoneda 引理的第二部分.

余元公式

余元公式为

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin \pi x} \tag{1}$$

引理 1　伽马函数的三种等价定义：

欧拉定义

$$\Gamma(z) := \int_0^\infty e^{-t}t^{z-1}\,\mathrm{d}t \tag{2}$$

高斯定义

$$\Gamma(z) := \lim_{n\to\infty}\frac{n!n^z}{z(z+1)\dots(z+n)} = \prod_n^\infty \frac{\left(1+\frac{1}{n}\right)^z}{\left(1+\frac{z}{n}\right)} \tag{3}$$

魏尔斯特拉斯定义

$$\frac{1}{\Gamma(x)} := ze^{\gamma z}\prod_n^\infty\left[\left(1+\frac{z}{n}\right)\mathrm{e}^{-\frac{z}{n}}\right] \tag{4}$$

首先是[9]

$$(\cos z + i\sin z)^n = \cos\ nz + i\sin nz \tag{5}$$

所以有

$$\sin nz = n(\cos^{n-1} z)\sin z - \frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}(\cos^{n-3} z\cdot\sin z) + \dots \tag{6}$$

仅仅考虑 $n = 2k+1$ 的情况，然后考虑到 $\cos^2 z = 1-\sin^2 z$，则必有

$$\sin(2k+1)z = \sin\ z\cdot P(\sin^2 z) \tag{7}$$

其中 $P(\sin^2 z)$ 代表以 $\sin^2 z$ 为变量的 k 次多项式.

考虑到在 $z_m = \frac{m\pi}{2k+1}, m = 1,2,\dots,k$ 时,对于每一个 z_m ,都有 $\sin(2k+1)z = 0, \sin\ z \neq 0$ 则可知每一个 z_m 都是 $P(\sin^2 z)$ 的根. 即得

$$P(\sin^2 z) = A(1-\frac{\sin^2 z}{\sin^2 z_1})(1-\frac{\sin^2 z}{\sin^2 z_2})\dots(1-\frac{\sin^2 z}{\sin^2 z_k}) \tag{8}$$

考虑到

$$A = P(0) = \lim_{z\to 0}\frac{\sin(2k+1)z}{\sin z} = 2k+1 \tag{9}$$

则有

$$P(\sin^2 z) = (2k+1)(1-\frac{\sin^2 z}{\sin^2 z_1})(1-\frac{\sin^2 z}{\sin^2 z_2})\dots(1-\frac{\sin^2 z}{\sin^2 z_k}) \tag{10}$$

也即

$$\sin(2k+1)z = \sin z\cdot\ (2k+1)(1-\frac{\sin^2 z}{\sin^2 z_1})\dots(1-\frac{\sin^2 z}{\sin^2 z_k}) \tag{11}$$

[9]这个证法来自于《微积分学教程》第二卷第 408 目

令 $z=\frac{x}{2k+1}$，则为

$$\sin\ x=\sin\left(\frac{x}{2k+1}\right)\cdot(2k+1)(1-\frac{\sin^2\frac{x}{2k+1}}{\sin^2 z_1})\ldots(1-\frac{\sin^2\frac{x}{2k+1}}{\sin^2 z_k}) \tag{12}$$

现在把 $\sin x$ 乘积分成两部分，即截取

$$V_k^a=(1-\frac{\sin^2\frac{x}{2k+1}}{\sin^2 z_{a+1}})\ldots(1-\frac{\sin^2\frac{x}{2k+1}}{\sin^2 z_k}) \tag{13}$$

而将前面的部分记为

$$U_a^k=(2k+1)\sin\left(\frac{x}{2k+1}\right)(1-\frac{\sin^2\frac{x}{2k+1}}{\sin^2 z_1})\ldots(1-\frac{\sin^2\frac{x}{2k+1}}{\sin^2 z_a}) \tag{14}$$

其中 $0<a<k$.

由 $\lim_{k\to\infty}(2k+1)\sin\frac{x}{2k+1}=x$

$$\lim_{k\to\infty}\left(\frac{\sin\frac{x}{2k+1}}{\sin\frac{m\pi}{2k+1}}\right)^2=\left(\frac{x}{m\pi}\right)^2,m=1,2,\ldots,k \tag{15}$$

可得

$$U_a=\lim_{k\to\infty}U_a^k=x(1-\frac{x^2}{\pi^2})\ldots(1-\frac{x^2}{a^2\pi^2}) \tag{16}$$

现在考虑 V_a^k. 考虑到在 $0<\varphi<\frac{\pi}{2}$ 时，$\frac{2}{\pi}\varphi<sin\varphi$，则有

$$\sin^2\frac{h\pi}{2k+1}>\frac{4}{\pi}\frac{(h\pi)^2}{(2k+1)^2},\ h=a+1,\ldots k \tag{17}$$

而 $\sin^2\frac{x}{2k+1}<\frac{x^2}{(2k+1)^2}$ 是易得的. 于是

$$1>V_a^k>(1-\frac{x^2}{4(a+1)^2})\ldots(1-\frac{x^2}{4k^2}) \tag{18}$$

因为最开始并没有限定 a 的具体数值，所以总可以找到 $a=a_0$，使得 $4(a_0+1)>x^2$

引理 2

对于充分大的 n 而言，$\prod_{n=1}^{\infty}(1+a_n)$ 收敛的充要条件是 $\sum_{n=1}^{\infty}a_n$ 收敛.

当然，不需要一定从 n=1 开始，去掉有限项都不改变收敛或发散的结果.

引理 3

若无穷乘积 $\prod_{n=1}^{\infty} p_n$ 收敛，那么总有充分大的 m 使得 $\lim_{m\to\infty}\prod_{n=m+1}^{\infty} p_n = 1$.

当然，这要求 $p_n \neq 0$，所有这时应该把 $x = 0, \pm\pi, \pm 2\pi \ldots$ 分出来单独考虑.

由于

$$\sum_{h=a_0+1}^{\infty} \frac{x^2}{4h^2}\text{收敛，所以} \lim_{a_0\to\infty}\lim_{k\to\infty}(1-\frac{x^2}{4(a_0+1)^2})\ldots(1-\frac{x^2}{4k^2}) = 1 \tag{19}$$

即为

$$\sin x = \lim_{a\to\infty} U_a = x\cdot\prod_{n=1}^{\infty}(1-\frac{x^2}{n^2\pi^2}) \tag{20}$$

这里可以直接看出，这对于 $x = 0, \pm\pi, \pm 2\pi \ldots$ 也是成立的.

另外，有

$$\sin \pi x = \pi x\cdot\prod_{n=1}^{\infty}(1-\frac{x^2}{n^2}) \tag{21}$$

考虑到（威尔斯特斯拉定义）

$$\Gamma(x) = \frac{1}{x}\prod_{n=1}^{\infty}\frac{(1+\frac{1}{n})^x}{1+\frac{x}{n}} \tag{22}$$

以及 $\Gamma(1+x) = x\Gamma(x)$，则有

$$\Gamma(1-x) = -x\Gamma(-x) = \prod_{n=1}^{\infty}\frac{(1+\frac{1}{n})^{-x}}{1-\frac{x}{n}} \tag{23}$$

于是

$$\Gamma(x)\Gamma(1-x) = \frac{1}{x}\cdot\prod_{n=1}^{\infty}\frac{1}{(1-\frac{x^2}{n^2})} = \frac{\pi}{\sin\pi x} \tag{24}$$

代数学基本定理

- 本词条存在未完成的内容.

预备知识 复变函数[346]，多项式环

代数学基本定理的表述

定义 1　代数学基本定理

设 $f(z)=\sum_{i=0}^{n} a_i z^i$ 为复数域上的多项式，那么必存在 $z_0 \in \mathbb{C}$，使得 $f(z_0)=0$.

由多项式的性质知，如果 $f(z_0)=0$，那么必然存在一个 $n-1$ 阶多项式 $g(z)$，使得 $f(z)=g(z)\cdot(z-z_0)$. 对 $g(z)$ 重复此步骤，再对接下来得到的每一个次数大于 1 的多项式都重复此步骤，最终结果是存在一系列 $z_i \in \mathbb{C}$，使得 $f(z)=\prod_{i=0}^{n}(z-z_i)$.

代数学基本定理的证明

本小节将用不同思路证明代数学基本定理.

代数拓扑方法

预备知识 基本群[658]

考虑 $\mathbb{C}$ 的度量拓扑，其中度量 $d(z_1,z_2)=|z_1-z_2|$；换句话说，把 $\mathbb{C}$ 看成二维欧氏空间. 假设存在多项式 $f(z)$ 使得任意复数 z 都满足 $f(z)\neq 0$，那么 $|f(z)|\neq 0$，于是可以引入映射 $p:\mathbb{C}\to S^1$，其中 $p(z)=f(z)/|f(z)|$；换句话说，p 将每个 $z\in\mathbb{C}$ 先映射到 $f(z)$，再保持辐角不变并移动到单位圆上.

由于 $f(z)$ 是连续映射，且 $f(z)/|f(z)|$ 没有奇点，于是 $p(z)=f(z)/|f(z)|$ 也是连续映射.

在 $\mathbb{C}$ 上选择任意一点 z_0 作为基点来构造基本群. 设任意回路 $r:S^1\to\mathbb{C}$，以及零回路 $n:S^1\to\mathbb{C}$，其中 $n(t)=z_0$ 对所有 t 成立. 由于 $\mathbb{C}$ 是可缩空间，因此必有 $r\cong n$. 由定理 3 [654]，$f\circ r\cong f\circ n$，因此单位圆上的回路也彼此同伦. 然而我们知道 S^1 的基本群是 $\mathbb{Z}$，意味着并非所有回路都彼此同伦，因此出现了矛盾.

矛盾意味着假设错误，也就是说，$p(z)$ 不是连续映射；换句话说，$p(z)$ 必然有零点.

第十一部分

经典力学

第一章

质点运动学

物理量和单位转换

量纲和单位制

[1]我们知道物理公式中有不同的物理量，例如时间，长度，电荷，等．我们把这些概念称为**量纲（physical dimension 或 dimension）**[2]．对公式中不同物理量之间量纲关系进行分析叫做**量纲分析（dimensional analysis）**．

同一个量纲又可以有不同的**单位（unit of measurement 或 unit）**．例如长度量纲的单位可以有“英尺”，“米”，“公里”，等．

不同量纲可以分别取指数再相乘，得到新的量纲．例如通过公式 $v = x/t$，就可以知道速度的量纲由长度的 1 次方和时间的 -1 次方相乘而来．如果一个量没有任何单位，那么我们就说它具有**单位量纲**或者**无量纲**．

有时候为了表明一些无量纲的量的用途也会在后面加一些记号，例如弧度是由弧长除以半径定义的，无量纲，但我们有时候仍然在后面加 Rad 强调它是一个弧度．又例如单位“摩尔”（符号 mol）只是一个无量纲的数，约等于 6.02×10^{23}．我们也可以在后面写上“个”，但和 rad 一样都不是一种独立的量纲．又例如原子单位制[1315] 中所有物理量都进行了无量纲化，为了强调原子单位可以在后面写上 au．

单位转换

预备知识 几何矢量[370]

在物理中我们首先要区分两类变量，一种是只有数值没有单位，例如“甲的身高是乙的 1.27 倍”．另一种既有数值也有单位，例如“甲的身高是 1.8 m”．我们把前者叫做一个**数**，把后者叫做**物理量**．

为了方便理解，我们这里把物理量和一维几何矢量[370] 的部分性质做一个类比：一个一维几何矢量本身也不能用一个数描述，而是需要先选取一个矢量基底，然后用一个数（即坐标）乘以这个基底才能表示这个矢量．例如对于同一个矢量，若甲选择的基底是乙的基底的两倍，那么甲的坐标将会是乙的一半．

[1]参考 Wikipedia 相关页面．

[2]注意 dimension 也可以表示空间的维度，二者是完全不同的概念．

用公式表示，假设某个一维矢量是 $\mathbf{v}$，两种不同的基底分别是 $\mathbf{u}_1$ 和 $\mathbf{u}_2$，则

$$\mathbf{v} = x_1\mathbf{u}_1 = x_2\mathbf{u}_2 \tag{1}$$

其中 x_1 和 x_2 是数学量．现在，如果我们已知 $\mathbf{u}_1$ 和 $\mathbf{u}_2$ 的关系，例如

$$\mathbf{u}_1 = c\mathbf{u}_2 \tag{2}$$

其中 c 也是一个数学量．我们就可以直接将这个关系代入式 1 中的 $\mathbf{u}_1$，得

$$x_1(c\mathbf{u}_2) = x_2\mathbf{u}_2 \tag{3}$$

现在等式两边都使用同样的基底，于是他们的坐标也必定相同，即 $x_2 = cx_1$．

同样，若基底间的关系是 $\mathbf{u}_2 = b\mathbf{u}_1$，显然有 $b = 1/c$，代入式 1 中的 $\mathbf{u}_2$ 得 $x_1 = bx_2$．

例 1

$$3\ \mathrm{km} = 3 \times 1000\ \mathrm{m} = 3000\ \mathrm{m} \tag{4}$$

也可以拓展到复合单位

$$2\ \frac{\mathrm{km}}{\mathrm{h}} = 2 \cdot \frac{1000\ \mathrm{m}}{3600\ \mathrm{s}} = \frac{5}{9}\ \mathrm{m/s} \tag{5}$$

与一维矢量不同的是，两个相同的物理量可以相除并得到一个无量纲的数，且这个数不依赖于单位制的选取．这么做的时候，可以类比为将两个矢量的模长相除．例如

$$\frac{200\ \mathrm{cm}}{1\ \mathrm{m}} = \frac{2\ \mathrm{m}}{1000\ \mathrm{mm}} = \frac{2\ \mathrm{m}}{1\ \mathrm{m}} = 2 \tag{6}$$

所以，另一种单位转换的办法，是对每种物理量先凑出一个等于 1 的无量纲分数，称为转换常数．例如 $\beta_x = 1\ \mathrm{m}/(1000\ \mathrm{mm}) = 1$，使得分母是原单位，分子是新单位，然后将它乘以需要转换的物理量．例如

$$2000\ \mathrm{mm} = 2000\ \mathrm{mm} \times \frac{1\ \mathrm{m}}{1000\ \mathrm{mm}} = 2\ \mathrm{m} \tag{7}$$

这与上面的方法是完全等效的．对于不同量纲下的任意物理量 Ω 和 Ω'，使用转换常数 β_Ω 可以把该过程记为

$$\Omega' = \beta_\Omega\Omega \tag{8}$$

例：厘米—克—秒单位制[1975]．

物理公式的单位制转换

使用不同的单位制时，物理公式可能保持不变也可能会增减一些常数.

例 2

无论用国际单位制还是厘米—克—秒（CGS）单位制[1975]，都有 $x = vt$. 但如果使用一个新的单位制，令 $x' = \beta_x x$，$v' = \beta_v v$，$t' = \beta_t t$，代入得

$$x' = \frac{\beta_v \beta_t}{\beta_x} v't' \tag{9}$$

若从国际单位转换成 CGS 单位，则 $\beta_x = 0.01\ \mathrm{m/cm}$，$\beta_t = 1$，$\beta_v = 0.01\ \mathrm{m/cm}$，恰好能使公式中的转换常数都消去使其保持不变. 但若时间单位改用小时，则 $\beta_t = 3600\ \mathrm{s/h}$. 代入上式变为

$$x' = \beta_v v't' \tag{10}$$

这样，公式中就多出了一个常数.

在这个简单的例子中，所有的转换常数虽然具有不同的数值，但本质上都还是无量纲的 1. 但在一些单位制中，同一个物理量的量纲却有可能不同，所以转换常数也不等于 1.

例 3

例如在高斯单位制[1976] 中，为了让国际单位制的库伦定律

$$F = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2} \tag{11}$$

变为

$$F' = \frac{q'_1 q'_2}{r'^2} \tag{12}$$

已知高斯单位制中力和长度的单位为 $\mathrm{g \cdot cm/s^2}$ 和 cm. 易得 q' 的量纲为 $\sqrt{\mathrm{g \cdot cm^3}}/\mathrm{s}$.

把 $F = \beta_F F'$，$q = \beta_q q'$，$r = \beta_x r'$ 代入式 11 得

$$F' = \frac{\beta_q^2}{4\pi\epsilon_0 \beta_x^2 \beta_F} \frac{q'_1 q'_2}{r'^2} \tag{13}$$

对比式 12 得 $\beta_q^2/(4\pi\epsilon_0 \beta_x^2 \beta_F) = 1$，即

$$\beta_q = \sqrt{4\pi\epsilon_0 \beta_F}\, \beta_x \approx 3.336 \times 10^{-10}\ \mathrm{C \cdot s/\sqrt{g \cdot cm^3}} \tag{14}$$

并不是单位量纲.

另见无量纲的物理公式[876].

不同物理量的运算

首先要注意，把两个具有不同量纲的物理量相加减没有任何意义，把一个无量纲的数与一个物理量相加减也没有意义. 例如把长度和时间相加，把速度与数字 2 相加都是没有意义的.

然而，几个相同或不同量纲的物理量相乘或相除却是很常见的. 例如长度乘以长度等于体积，长度除以时间等于速度. 相乘或相除后所得物理量的单位也可以写成若干个基本单位相乘或者相除，最终表示成若干个单位的幂相乘的形式（可以用“·”表示相乘，也可以省略）. 例如

$$2\ \mathrm{m}\times 3\ \mathrm{m}=6\ \mathrm{m}\cdot\mathrm{m}=6\ \mathrm{m}^2 \tag{15}$$

$$6\ \mathrm{m}/3\ \mathrm{s}=2\ \mathrm{m}/\mathrm{s}=2\ \mathrm{m}\cdot\mathrm{s}^{-1} \tag{16}$$

有时候我们会给这些复合单位一个新的记号,例如根据牛顿第二定律（式 2[899]），力的国际单位应该是 $\mathrm{kg}\cdot\mathrm{m}\cdot\mathrm{s}^{-2}$，但我们也可以定义一个等效单位 N （牛）.

无量纲的物理公式

预备知识 万有引力[1027]，物理量和单位转换[873]

与数学公式不同，物理公式中的变量（即**物理量**）通常既包含数值又包含单位，如 $x=10\ \mathrm{cm}$. 但有时候我们需要将含有单位的公式转换为不含单位的物理公式. 例如数值计算时，或者为了书写简单（常见于量子力学，见原子单位[1315]）. 我们通过几个例子来说明转换过程.

例 1　牛顿第二定律

国际单位制下的牛顿第二定律公式为

$$F=ma \tag{1}$$

我们先定义几个含单位的常量例如 $\beta_F = 1\ \mathrm{N}$，$\beta_m = 1\ \mathrm{kg}$，$\beta_a = 1\ \mathrm{m/s^2}$．把不含单位的的力，质量和加速度分别为 F_0, m_0, a_0，则有 $F = F_0\beta_F$，$m = m_0\beta_m$，$a = a_0\beta_a$．代入上式得

$$F_0 = \frac{\beta_m\beta_a}{\beta_F} m_0 a_0 \tag{2}$$

由于国际单位定义 $1\ \mathrm{N} = 1\ \mathrm{kg}\cdot\ \mathrm{m/s^2}$，上式中 $\beta_m\beta_a/\beta_F = 1$，所以不含单位的牛顿第二定律为

$$F_0 = m_0 a_0 \tag{3}$$

以上的做法看起来似乎并没有什么意义，这是因为我们把每个常量 β 都定义为一个相应的国际单位．事实上，只要保证 $\beta_m\beta_a/\beta_F = 1$，这三个常量是可以任取的．例如令 $\beta_m = 4\ \mathrm{g}$，$\beta_a = 25\ \mathrm{cm/s^2}$，$\beta_F = 10^{-3}\ \mathrm{N}$，上式仍然成立．有了这个变换，有时候我们只需要做一次计算就能得到不同参数下的结果．

例 2　万有引力公式

国际单位下的万有引力公式为

$$F = G\frac{Mm}{r^2} \tag{4}$$

其中 $G \approx 6.674 \times 10^{-11}\ \mathrm{m^3kg^{-1}s^{-2}}$．在例 1 的基础上定义 $r = \beta_x r_0$，则有

$$F_0 = \frac{G\beta_m^2}{\beta_F\beta_x^2}\frac{M_0 m_0}{r_0^2} \tag{5}$$

我们若想让新的不含单位的万有引力公式也不含 G 以减少计算量，即

$$F_0 = \frac{M_0 m_0}{r_0^2} \tag{6}$$

只需令所有的 β 满足 $G\beta_m^2/(\beta_F\beta_x^2) = 1$，例如 $\beta_x = 1\ \mathrm{m}$，$\beta_m = 1.224 \times 10^5\ \mathrm{kg}$，$\beta_F = 1\ \mathrm{N}$．

注意无单位的物理公式和有单位的物理公式（无论用什么单位）存在本质的不同，例如在式 3 中，如果受力物体的质量恰好等于 β_m，那么公式可以直接写为

$$F_0 = a_0 \tag{7}$$

而这种写法对含单位的公式来说是错误的，因为单位不同的两个物理量不可以相等（或相加）．我们也可以直接把数字和无单位的物理量相加，如 $1 + x_0$．

在不至于混淆的情况下（例如通篇都使用无量纲公式），我们就可以把以上的角标 0 去掉．

无单位公式转换为含单位公式

对某个物理量 Q，有

$$Q = \beta_Q Q_0 \tag{8}$$

要把无单位公式变为含单位公式，就把式中所有 Q_0 用 Q/β_Q 替换即可．以式 6 为例，替换后得

$$F = \frac{\beta_F \beta_x^2}{\beta_m^2} \frac{Mm}{r^2} = G\frac{Mm}{r^2} \tag{9}$$

位置矢量 位移

预备知识 几何矢量[370]

位置矢量（位矢）就是从坐标原点指向某一点的矢量，通常记为 $\mathbf{r}$．当定义了一个坐标系，那么坐标系中一点的位置就可以用位矢表示．

有时候表示一个关于位置的函数，通常将位矢 $\mathbf{r}$ 作为自变量．例如一个物体内密度关于位置的分布可以表示为 $\rho(\mathbf{r})$．在直角坐标系中，就相当于 $\rho(x,y,z)$，在球坐标系中就相当于 $\rho(r,\theta,\phi)$ 这么做的好处是书写简洁，而且不需要指定坐标系的种类．

在物体运动过程中，可以把物体的位矢看做时间的矢量函数 $\mathbf{r}(t)$，则**位移** $\Delta\mathbf{r}$ 是一段时间 $[t_1,t_2]$ 内物体初末位矢的矢量差

$$\Delta\mathbf{r} = \mathbf{r}(t_2) - \mathbf{r}(t_1) \tag{1}$$

注意位移只与一段时间内物体的初末位置有关，与路径无关．

预备知识 全微分[269]，矢量的微分，矢量内积[383]

例 1　**证明** $\mathrm{d}r = \hat{\mathbf{r}} \cdot \mathrm{d}\mathbf{r}$

这个证明的几何意义是，位矢模长的微小变化等于位矢的微小变化在位矢方向的投影．

这里以平面直角坐标系中的位矢为例证明．令位矢 $\mathbf{r}$ 的坐标为 (x,y)，模长为 $r = \sqrt{x^2+y^2}$，模长的全微分为

$$\mathrm{d}r = \frac{\partial r}{\partial x}\mathrm{d}x + \frac{\partial r}{\partial y}\mathrm{d}y = \frac{x}{\sqrt{x^2+y^2}}\mathrm{d}x + \frac{y}{\sqrt{x^2+y^2}}\mathrm{d}y \tag{2}$$

考虑到 $x/\sqrt{x^2+y^2}$ 和 $y/\sqrt{x^2+y^2}$ 分别为 $\hat{\mathbf{r}}=\mathbf{r}/r$ 的两个分量，$\mathrm{d}x$ 和 $\mathrm{d}y$ 分别为 $\mathrm{d}\mathbf{r}$ 的两个分量，根据内积的定义[383] 上式变为

$$\mathrm{d}r=\hat{\mathbf{r}}\cdot\mathrm{d}\mathbf{r} \tag{3}$$

速度 加速度（一维）

预备知识 位移[878]，复合函数求导[174]

速度和加速度都是矢量，但如果我们考虑质点的一维运动（沿直线运动），那么我们可以指定一个正方向并沿运动方向建立坐标轴．这样一来，我们就可以把一维情况下的位移、速度、加速度这些矢量用一个标量来表示，可以用标量的正负号区分矢量的方向，正号代表指向正方向，负号代表指向负方向，标量的绝对值就等于矢量的模长．所以以下我们用坐标 x 来表示一维位移，实数 v 和 a 来表示一维速度和加速度．事实上，这些标量可以看作是对应矢量的坐标[370]．

物理学中，**速度（velocity）**和**加速度（acceleration）**通常指瞬时值．在一维运动中，**瞬时速度（instantaneous velocity）**被定义为一段极短时间 Δt 内质点的位移[878] Δx 除以这段时间，**瞬时加速度（instantaneous acceleration）**被定义为一段极短时间 Δt 内质点的速度变化 Δv 除以这段时间，而这些恰好是导数[158] 的定义．用极限[149] 和导数[158] 来表示，就是

$$v(t)=\lim_{\Delta t\to 0}\frac{x(t+\Delta t)-x(t)}{\Delta t}=\frac{\mathrm{d}x(t)}{\mathrm{d}t} \tag{1}$$

$$a(t)=\lim_{\Delta t\to 0}\frac{v(t+\Delta t)-v(t)}{\Delta t}=\frac{\mathrm{d}v(t)}{\mathrm{d}t} \tag{2}$$

根据高阶导数[169] 的定义，加速度就是位矢的二阶导数，即导数的导数

$$a(t)=\frac{\mathrm{d}^2x(t)}{\mathrm{d}t^2} \tag{3}$$

为什么经典力学中，我们通常之关心位移的一阶和二阶导数而不关心更高阶呢？因为牛顿第二定律中出现了加速度，把它和质点的受力紧紧关联起来，另一方面，和速度成正比的动量是重要的守恒量．

例 1　匀加速运动

若已知某直线运动的位移—时间函数为

$$x(t) = x_0 + v_0 t + \frac{1}{2} a_0 t^2 \tag{4}$$

试证明这是一个匀加速运动.

对位移求导得到速度为

$$v(t) = v_0 + a_0 t \tag{5}$$

再次求导（二阶导数）得到加速度为 $a(t) = a_0$ 是常数. 可见这是一个匀加速运动.

事实上，任何匀加速直线运动都可以表示为式 4 的形式，详见“匀加速直线运动[887]”.

例 2　简谐振动

已知简谐振动（详见“简谐振子[913]”）的位移函数为 $x(t) = A\cos(\omega t)$，运用复合函数求导[174] 得速度为 $v(t) = -A\omega \sin(\omega t)$，加速度为 $a(t) = -A\omega^2 \cos(\omega t)$.

由速度或加速度求位移

预备知识　牛顿—莱布尼兹公式[202]

既然一维速度 $v(t)$ 是位置 $x(t)$ 的导数，那么 $x(t)$ 是 $v(t)$ 的原函数[187]. 由牛顿—莱布尼兹公式得速度在一段时间的定积分等于初末位置之差，即

$$x(t_2) - x(t_1) = \int_{t_1}^{t_2} v(t)\,\mathrm{d}t \tag{6}$$

所以若已知某时刻质点的位置 $x(t_0) = x_0$，和速度函数 $v(t)$，就可以求得任意时刻的位置.

$$x(t) = x_0 + \int_{t_0}^{t} v(t')\,\mathrm{d}t' \tag{7}$$

注意为了区分积分变量和积分上限，我们把积分变量改成 t'. 这是一个常见的做法.

例 3　匀速直线运动

若一维运动的质点速度始终为 v_0，由式 7 得

$$\begin{aligned} x(t) &= x_0 + \int_{t_0}^{t} v_0 \,\mathrm{d}t \\ &= x_0 + v_0(t - t_0) \end{aligned} \tag{8}$$

与式 6 和式 7 同理，一维速度和加速度之间也有类似关系

$$v(t_2) - v(t_1) = \int_{t_1}^{t_2} a(t) \,\mathrm{d}t \tag{9}$$

$$v(t) = v_0 + \int_{t_0}^{t} a(t') \,\mathrm{d}t' \tag{10}$$

把式 10 带入式 7 再次定积分得（注意积分变量再次重命名）

$$\begin{aligned} x(t) &= x_0 + \int_{t_0}^{t} \left[v_0 + \int_{t_0}^{t'} a(t'') \,\mathrm{d}t'' \right] \mathrm{d}t' \\ &= x_0 + v_0(t - t_0) + \int_{t_0}^{t} \left[\int_{t_0}^{t'} a(t'') \,\mathrm{d}t'' \right] \mathrm{d}t' \end{aligned} \tag{11}$$

这里做了两次积分，即二重积分[283].

简单的例子见“匀加速直线运动[887]”.

习题 1

你是否能看出，以 t' 和 t'' 建立直角坐标系，式 4 的二重积分的区域是一个三角形？

速度 加速度

预备知识 速度 加速度 （一维）[879]，矢量的导数[274]，矢量积分

在大学物理中，“位移”，“速度”和“加速度”都是矢量，既包括了大小，也包括方向. 如果没有特殊说明，它们一般是指“瞬时速度”和“瞬时加速度”.

速度的定义

考察一个质点在运动过程中在某时刻经过某一点的速度，就取质点在这一点附近的一小段位移 $\Delta\mathbf{r}$，以及物体完成这段位移需要的时间 Δt．那么当 Δt 无穷小时，若 $\Delta\mathbf{r}/\Delta t$ 存在极限，则这个极限就是速度矢量 $\mathbf{v}$．写成极限的形式，就是

$$\mathbf{v}=\lim_{\Delta t\to 0}\frac{\Delta\mathbf{r}}{\Delta t} \tag{1}$$

速度与位矢的关系

质点在运动时，其位矢 $\mathbf{r}$ 是时间 t 的函数，质点在 t_1 时刻的位矢为 $\mathbf{r}(t_1)$，经过时间 Δt，位矢为 $\mathbf{r}(t_1+\Delta t)$，所以物体在 Δt 时间内的位移为

$$\Delta\mathbf{r}=\mathbf{r}(t_1+\Delta t)-\mathbf{r}(t_1) \tag{2}$$

式 2 代入式 1，得

$$\mathbf{v}(t_1)=\lim_{\Delta t\to 0}\frac{\mathbf{r}(t_1+\Delta t)-\mathbf{r}(t_1)}{\Delta t} \tag{3}$$

根据矢量求导[274] 的定义，这就是位矢对时间的导数，即

$$\mathbf{v}=\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} \tag{4}$$

简单的例子：匀速圆周运动的速度（求导法）[883]．

加速度的定义

通常情况下，质点运动轨迹上的每一点都会对应一个确定的速度矢量[3]，类比速度的定义，加速的定义为

$$\mathbf{a}(t_1)=\lim_{\Delta t\to 0}\frac{\Delta\mathbf{v}}{\Delta t}=\lim_{\Delta t\to 0}\frac{\mathbf{v}(t_1+\Delta t)-\mathbf{v}(t_1)}{\Delta t}=\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} \tag{5}$$

结合速度的定义，加速度为

$$\mathbf{a}=\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}=\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)=\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} \tag{6}$$

所以，加速度是速度对时间的导数，或者位矢对时间的二阶导数．

简单的例子：匀速圆周运动的速加速度（求导法[158]）．

[3]注意上面的速度在定义时虽然取了两点，但是取极限以后，速度和位置是一一对应的，也就和时间一一对应，而不是两个位置和时间对应一个速度．

由速度或加速度计算位矢

如果已知速度关于时间的函数 $\mathbf{v}(t)$，以及初始时间 t_0 和位置 $\mathbf{r}_0$，该如何得到位移—时间函数 $\mathbf{r}(t)$ 呢？类比一维的情况[879]，我们也可以通过矢量函数的定积分[277]（见例 1）来求出速度—时间函数进而求出位移—时间函数

$$\mathbf{v}(t)=\mathbf{v}_0+\int_{t_0}^{t}\mathbf{a}(t)\,\mathrm{d}t \tag{7}$$

$$\mathbf{r}(t)=\mathbf{r}_0+\int_{t_0}^{t}\mathbf{v}(t)\,\mathrm{d}t \tag{8}$$

简单的例子：匀加速运动[889].

圆周运动的速度

几何法

预备知识 小角正弦值极限[152]，速度的定义[881]

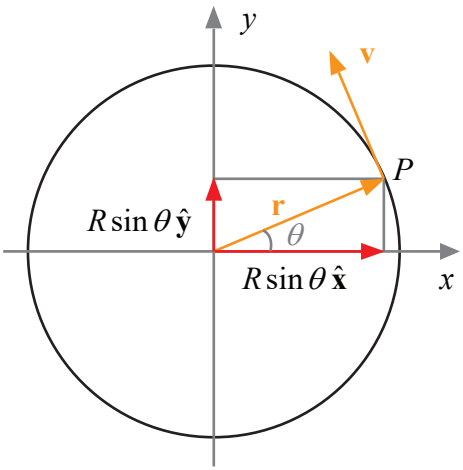


图 1: 匀速圆周运动的速度

如图设一个点 P 做半径为 R 的圆周运动，角速度为 ω（可以是时间的函数），那么在一段微小时间 Δt 内，可以认为 ω 是常量，点 P 转过的角度为

$\Delta\theta=\omega\Delta t$. 这样，根据小角正弦值极限[152]，当 $\Delta t\to 0$ 时，点 P 在 Δt 内走过的位移长度（线段的长度）趋近于弧的长度，即 $|\Delta\mathbf{s}|\to R\omega\Delta t$.

根据速度的定义

$$\mathbf{v}=\lim_{\Delta t\to 0}\frac{\Delta\mathbf{s}}{\Delta t} \tag{1}$$

速度的大小为

$$v=\lim_{\Delta t\to 0}\frac{|\Delta\mathbf{s}|}{\Delta t}=\lim_{\Delta t\to 0}\frac{\omega R\Delta t}{\Delta t}=\omega R \tag{2}$$

速度的方向显然与过 A 点的圆的切线重合.

求导法

预备知识 矢量的导数 求导法则[274]

如图，在平面直角坐标系（单位矢量分别为 $\hat{\mathbf{x}}$，$\hat{\mathbf{y}}$ ）中，令一个绕原点做逆时针匀速圆周运动的质点的位矢为 $\mathbf{r}$，与 $\hat{\mathbf{x}}$ 的夹角是时间的函数 $\theta(t)$，圆周运动的半径为 R. 那么任意时刻 t 将位矢 $\mathbf{r}$ 沿着 x 与 y 轴方向分解，有

$$\mathbf{r}(t)=R\cos\theta(t)\,\hat{\mathbf{x}}+R\sin\theta(t)\,\hat{\mathbf{y}} \tag{3}$$

其中 $\hat{\mathbf{x}}$ 是 x 轴正方向的单位矢量，$\hat{\mathbf{y}}$ 是 y 轴正方向的单位矢量. 由速度的定义 $\mathbf{v}=\mathrm{d}\mathbf{r}/\mathrm{d}t$，即

$$\begin{aligned}\mathbf{v}&=\frac{\mathrm{d}}{\mathrm{d}t}(R\cos\theta\,\hat{\mathbf{x}}+R\sin\theta\,\hat{\mathbf{y}})=-R\dot{\theta}\sin\theta\,\hat{\mathbf{x}}+R\dot{\theta}\cos\theta\,\hat{\mathbf{y}}\\&=\dot{\theta}R[\cos(\theta+\pi/2)\hat{\mathbf{x}}+\sin(\theta+\pi/2)\hat{\mathbf{y}}]\end{aligned} \tag{4}$$

定义**瞬时角速度**（简称**角速度**）等于 θ 关于时间的导数 $\omega=\dot{\theta}$，则速度大小为 $v=\omega R$，方向为 $\hat{\mathbf{r}}$ 逆时针旋转 $\pi/2$，即圆的切线方向.

三维空间的情况

预备知识 矢量叉乘[387]

如图 2，在三维空间中，圆周运动所在的平面可以任意选取，我们可以将角速度拓展成一个矢量 $\boldsymbol{\omega}$，其方向垂直于该平面并由右手定则[20] 确定. 令坐标系的原点在圆周运动的轴上，用位矢 $\mathbf{r}$ 表示点 P 的位置，则圆周运动的半径为

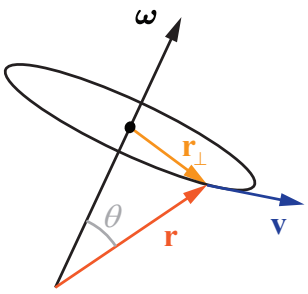


图 2: 角速度与线速度

$r_\perp = r\sin\theta$，其中 θ 是 $\mathbf{r}$ 与 $\boldsymbol{\omega}$ 的夹角．所以圆周运动速度的大小为 $v = \omega r\sin\theta$．根据矢量叉乘的几何定义[387]，有

$$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r} \tag{5}$$

圆周运动的加速度

预备知识 圆周运动的速度[883]

匀速圆周运动的加速度（几何法）

在圆周运动中，位矢 $\mathbf{r}$ 是时间的函数．对时间求导后，我们得到速度矢量关于时间的函数．对速度也进行同样的操作，就不难得到圆周运动的加速度[881]．

$$\mathbf{a} = \lim_{\Delta t\to 0}\frac{\Delta\mathbf{v}}{\Delta t} \tag{1}$$

现在我们用几何的方法来求该极限．根据矢量减法的定义，计算 $\Delta\mathbf{v} = \mathbf{v}_2 - \mathbf{v}_1$ 要先把 $\mathbf{v}_1$ 和 $\mathbf{v}_2$ 的起点放在一起（例如都放在原点），再从 $\mathbf{v}_1$ 的终点指向 $\mathbf{v}_2$ 的终点得到 $\Delta\mathbf{v}$ （图 1）．

我们已知匀速圆周运动的速度大小为 $|\mathbf{v}| = R\omega$，根据“微小正弦极限[152]”中的结论，把 $\Delta\mathbf{v}$ 的长度用弧长近似，得

$$|\Delta\mathbf{v}| = |\mathbf{v}|\Delta\theta = (R\omega)\omega\Delta t \tag{2}$$

所以质点的加速度大小为

$$a = \lim_{\Delta t\to 0}\frac{|\Delta\mathbf{v}|}{\Delta t} = \lim_{\Delta t\to 0}\frac{\omega^2 R\Delta t}{\Delta t} = \omega^2 R \tag{3}$$

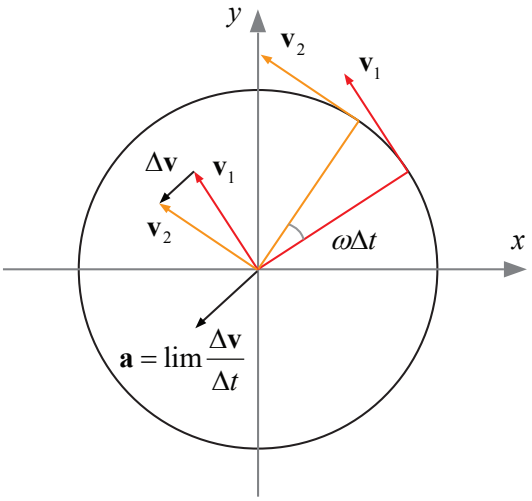


图 1: 把速度矢量移到原点再相减

由图可得加速度的方向是速度方向逆时针偏转 $\pi/2$. 又由于速度方向是位移方向逆时针偏转 $\pi/2$，所以匀速圆周运动的加速度的方向与位矢的方向相反.

结合模长和方向，令 $\mathbf{r}$ 为位矢（取圆心为坐标原点），就得到加速度的矢量形式

$$\mathbf{a} = -\omega^2\mathbf{r} \tag{4}$$

由速度与角速度的关系，它的模长也可以表示为 v^2/r 或者 ωv.

圆周运动的加速度（求导法）

现在我们来推导一般圆周运动的加速度（不要求匀速），将圆周运动的速度（式 4 [883]）再次对时间求导得加速度

$$\mathbf{a} = \dot{\mathbf{v}} = -R\dot{\theta}^2(\cos\theta\,\hat{\mathbf{x}} + \sin\theta\,\hat{\mathbf{y}}) + \ddot{\theta}R[\cos(\theta+\pi/2)\hat{\mathbf{x}} + \sin(\theta+\pi/2)\hat{\mathbf{y}}] \tag{5}$$

当角速度 $\omega = \dot{\theta}$ 为常量时（匀速圆周运动），上式第二项为零，第一项与式 4 相同，当角速度随时间变化时，由于 $\ddot{\theta}R = \dot{\omega}R = \dot{v}$，上式可以记为

$$\mathbf{a} = -\omega^2\mathbf{r} + \dot{v}\hat{\mathbf{v}} \tag{6}$$

其中 $\hat{\mathbf{v}}$ 是速度方向的单位矢量. 所以变速圆周运动除了向心加速度外，还有一个沿速度方向的加速度.

三维空间的情况

预备知识 连续叉乘的化简[392]

若要把式 4 拓展到三维空间中围绕过圆心的轴转动的任意匀速圆周运动，可以对式 5 [883]（$\mathbf{v}=\boldsymbol{\omega}\times\mathbf{r}$）求时间导数（令 $\boldsymbol{\omega}$ 为常矢量）得

$$\mathbf{a}=\dot{\mathbf{v}}=\boldsymbol{\omega}\times\mathbf{v} \tag{7}$$

将式 5 [883] 再次代入，得

$$\mathbf{a}=\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r}) \tag{8}$$

要验证该式与式 4 吻合，把连续叉乘化为内积得

$$\mathbf{a}=\omega^2 r\cos\theta\,\hat{\boldsymbol{\omega}}-\omega^2\mathbf{r}=-\omega^2(\mathbf{r}-r\cos\theta\,\hat{\boldsymbol{\omega}})=-\omega^2\mathbf{r}_\perp \tag{9}$$

其中 θ 是 $\hat{\boldsymbol{\omega}}$ 和 $\hat{\mathbf{r}}$ 之间的夹角，$\mathbf{r}_\perp$ 是从圆周运动的圆心指向点 P 的矢量，相当于式 4 中的 $\mathbf{r}$．证毕．

我们再来考虑变速圆周运动的情况，当 $\boldsymbol{\omega}$ 的模长随时间变化时，式 7 的求导变为

$$\mathbf{a}=\boldsymbol{\omega}\times\mathbf{v}+\dot{\boldsymbol{\omega}}\times\mathbf{r}=-\omega^2\mathbf{r}_\perp+\dot{\boldsymbol{\omega}}\times\mathbf{r} \tag{10}$$

定义**角加速度（angular acceleration）** $\boldsymbol{\alpha}=\dot{\boldsymbol{\omega}}$，并将式 5 [883] 代入，得

$$\mathbf{a}=\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})+\boldsymbol{\alpha}\times\mathbf{r} \tag{11}$$

由定义易证右边第二项等于式 6 中的 $\dot{v}\hat{\mathbf{v}}$．

匀加速直线运动

预备知识 速度加速度（一维）[879]

质点做匀速直线运动时，我们延运动的直线建立坐标轴 x，则最一般的运动方程

$$x(t)=x_0+v_0(t-t_0)+\frac{1}{2}a_0(t-t_0)^2 \tag{1}$$

其中 x_0, v_0, a_0 分别是 t_0 时刻的位置，速度和加速度．注意沿 x 轴正方向的速度和加速度取正号，沿反方向取负号．

匀加速直线运动速度变化为

$$v(t) = v_0 + a_0(t - t_0) \tag{2}$$

另外有一条不含时间的公式

$$v_2^2 - v_1^2 = 2a_0(x_2 - x_1) \tag{3}$$

其中 v_1, v_2 分别是质点经过点 x_1 和 x_2 时的速度．

注意只有初速度矢量和加速度矢量方向相同（包括初速度为零）时匀加速运动的轨迹才是直线，否则轨迹就是抛物线，见“匀加速运动[889]”.

未完成：把“匀加速运动”的自由落体移动过来

未完成：插入 ep1 的飞机丢导弹例题

推导

做匀速直线运动的质点在 t_0 时的位置为 x_0，速度为 v_0，且加速度始终等于常数 a_0，求任意时刻的速度和加速度 $x(t)$.

我们首先把 $a(t) = a_0$ 代入式 10 [879]

$$v(t) = v_0 + \int_{t_0}^{t} a_0 \,\mathrm{d}t' \tag{4}$$

马上得到式 2．再次积分（式 7 [879]）

$$x(t) = x_0 + \int_{t_0}^{t} [v_0 + a_0(t' - t_0)] \,\mathrm{d}t' \tag{5}$$

得式 1．这个过程相当于直接使用式 11 [879].

要得到式 3，由式 1 和式 2 分别得

$$x_2 - x_1 = v_1(t_2 - t_1) + \frac{1}{2}a_0(t_2 - t_1)^2 \tag{6}$$

$$t_2 - t_1 = \frac{v_2 - v_1}{a_0} \tag{7}$$

代入消去 t 得式 3．直观来看，如果画速度—时间图，$x_2 - x_1$ 可以表示成梯形的面积，利用梯形公式可以得到该式．

式 3 的另一种推导是，在学习牛顿第二定律[899] 和动能定理[904] 后，我们可以直接写出

$$\frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2 = F(x_2 - x_1) = ma(x_2 - x_1) \tag{8}$$

两边除以 $m/2$ 得到式 3 .

匀加速运动

预备知识 速度 加速度[881]

若在一段时间内，质点的加速度矢量不随时间变化（常矢量）$\mathbf{a}$，那么我们说质点做**匀加速运动（constant acceleration motion）**．由“速度 加速度[881]”中的式 7 和式 8，速度和位移函数分别为

$$\mathbf{v}(t) = \mathbf{v}_0 + \int_{t_0}^{t} \mathbf{a}\,\mathrm{d}t = \mathbf{v}_0 + \mathbf{a}\cdot(t - t_0) \tag{1}$$

$$\mathbf{r}(t) = \mathbf{r}_0 + \int_{t_0}^{t} \mathbf{v}(t)\,\mathrm{d}t = \mathbf{r}_0 + \mathbf{v}_0\cdot(t - t_0) + \frac{1}{2}\mathbf{a}\cdot(t - t_0)^2 \tag{2}$$

匀加速直线运动

一个最简单的直线匀加速运动是**自由落体运动**．自由落体运动是初速度 $\mathbf{v}_0 = 0$，竖直向下加速度为重力加速度恒为 g 的匀加速直线运动．其中 $g \approx 9.8\ \mathrm{m/s^2}$ 是**重力加速度**，也可以用常矢量 $\mathbf{g}$ 表示．代入式 1 和式 2 得

$$\mathbf{v}(t) = \mathbf{g}\cdot(t - t_0) \tag{3}$$

$$\mathbf{r}(t) = \mathbf{r}_0 + \frac{1}{2}\mathbf{g}\cdot(t - t_0)^2 \tag{4}$$

为了方便，在一维运动时我们可以直接用标量表示位移，速度和加速度，这样以上两式中的矢量都可以用标量表示．

未完成：自由落体移动到“匀加速直线运动”

抛体运动

作为一个稍复杂的情况，抛体运动是加速度为 $\mathbf{g}$，初速度为 $\mathbf{v}_0$ 的匀加速运动．将 $\mathbf{a}=\mathbf{g}$ 代入式 1 和式 2 得

$$\mathbf{v}(t)=\mathbf{v}_0+\mathbf{g}\cdot(t-t_0) \tag{5}$$

$$\mathbf{r}(t)=\mathbf{r}_0+\mathbf{v}_0\cdot(t-t_0)+\frac{1}{2}\mathbf{g}\cdot(t-t_0)^2 \tag{6}$$

对比式 4 和式 6 可以发现抛体运动就是自由落体运动与匀速直线运动的矢量叠加．所以如果我们在一个相对于当前参考系以 $\mathbf{v}_0$ 运动的参考系中观察抛体运动，就会是自由落体运动．

匀速曲线运动

预备知识 曲率半径[179]，圆周运动的加速度[885]

要研究质点的任意曲线运动，如果我们已经知道曲线的形状和运动的速度，最重要的是算出它的加速度，以便用牛顿第二定律进行受力分析．质点沿曲线运动时，即使它的速度不变，也存在加速度，与速度的方向垂直．最典型的例子就是匀速圆周运动．我们现在把对圆周运动加速度的分析拓展到沿任意曲线的匀速运动，以后还会拓展到变速曲线运动[??]．

在以下的推导中我们会发现可以直接使用匀速圆周运动的向心加速度[885]公式来计算任意匀速曲线运动，只要把半径替换成曲线某点处的曲率半径．

加速度推导

质点做匀速曲线运动时，由于速度矢量模长 $|\mathbf{v}|$ 恒定，加速度[881] 矢量 $\mathbf{a}$ 完全由速度矢量 $\mathbf{v}$ 的方向改变而产生，就像匀速圆周运动那样．回顾加速度的定义（式 5 [881]）

$$\mathbf{a}=\lim_{\Delta t\to 0}\frac{\Delta\mathbf{v}}{\Delta t} \tag{1}$$

类似用几何法推导匀速圆周运动的速度[883] 那样，我们可以近似认为[152] 当速度矢量 $\mathbf{v}$ 转过一个小角度 $\Delta\theta$ 时，它的增量 $\Delta\mathbf{v}$ 垂直于 $\mathbf{v}$，且大小为

$$|\Delta\mathbf{v}| = v\Delta\theta \tag{2}$$

代入式 2 得加速度大小为

$$|\mathbf{a}| = \lim_{\Delta t\to 0}\frac{|\Delta\mathbf{v}|}{\Delta t} = v\lim_{\Delta t\to 0}\frac{\Delta\theta}{\Delta t} = v\omega \tag{3}$$

其中 ω 是速度矢量 $\mathbf{v}$ 在某个时刻旋转的角速度．方向与 $\mathbf{v}$ 垂直，即曲线切线变化的方向．

如果使用角速度矢量（图 2 [883]）$\boldsymbol{\omega}$，那么加速度矢量可以用矢量叉乘[387] 表示为

$$\mathbf{a} = \boldsymbol{\omega}\times\mathbf{v} \tag{4}$$

那么当质点经过曲线某一点时，如何求 ω 呢？我们可以使用曲率[179] 的概念．令质点所在位置的曲率半径为 R，根据曲率半径的定义（式 1 [179]），Δt 内质点在曲线上走过的长度为 $\Delta l = v\Delta t$，所以切线的角度变化率为

$$\omega = \lim_{\Delta t\to 0}\frac{\Delta\theta}{\Delta t} = \lim_{\Delta t\to 0}\frac{\Delta\theta}{\Delta l}\frac{\Delta l}{\Delta t} = \frac{v}{R} \tag{5}$$

再带入式 3 得

$$|\mathbf{a}| = v\omega = \frac{v^2}{R} = \omega^2 R \tag{6}$$

这和匀速圆周运动的向心加速度（式 4 [885]）的形式一样，只是把半径换为曲率半径．这是意料之中的，因为圆就是曲率半径恒为 R 的特殊曲线．

极坐标中的速度和加速度

预备知识 速度 加速度[881]，极坐标中单位矢量的偏导[287]

若已知某点的极坐标关于时间的函数 $r(t)$ 和 $\theta(t)$，求该点的速度和加速度．

极坐标中的位置矢量可以用 $\mathbf{r} = r\hat{\mathbf{r}}$ 表示，注意其中径向单位矢量可以看做复合函数 $\hat{\mathbf{r}}[\theta(t)]$．根据定义，速度是位矢的一阶导数，在力学中经常在变量上面加一点表示对时间的一阶导数，两点表示二阶导数，根据矢量的求导法则

$$\mathbf{v} = \dot{r}\hat{\mathbf{r}} + r\frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}t} \tag{1}$$

由链式法则和式 1 [287]，上式中

$$\frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}t}=\frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}\theta}\dot{\theta}=\dot{\theta}\hat{\boldsymbol{\theta}} \tag{2}$$

所以极坐标中的速度为

$$\mathbf{v}=\dot{r}\hat{\mathbf{r}}+r\dot{\theta}\hat{\boldsymbol{\theta}} \tag{3}$$

这是符合直觉的，**径向**速度等于位矢模长的导数，而**角向**速度等于位矢模长乘以角速度.

我们再来计算加速度，用同样的方法对速度求一阶导数得

$$\begin{aligned}\mathbf{a}&=\ddot{r}\hat{\mathbf{r}}+\dot{r}\frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}t}+\dot{r}\dot{\theta}\hat{\boldsymbol{\theta}}+r\ddot{\theta}\hat{\boldsymbol{\theta}}+r\dot{\theta}\frac{\mathrm{d}\hat{\boldsymbol{\theta}}}{\mathrm{d}t}\\&=\ddot{r}\hat{\mathbf{r}}+\dot{r}\dot{\theta}\hat{\boldsymbol{\theta}}+\dot{r}\dot{\theta}\hat{\boldsymbol{\theta}}+r\ddot{\theta}\hat{\boldsymbol{\theta}}-r\dot{\theta}^2\hat{\mathbf{r}}\\&=(\ddot{r}-r\dot{\theta}^2)\hat{\mathbf{r}}+(2\dot{r}\dot{\theta}+r\ddot{\theta})\hat{\boldsymbol{\theta}}\end{aligned} \tag{4}$$

这个结论并不是那么显而易见. 我们将加速度的径向和角向分量分别记为 a_r 和 a_θ，其中 a_θ 还可以记为另一种更紧凑形式即

$$a_\theta=\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}t}\left(r^2\frac{\mathrm{d}\theta}{\mathrm{d}t}\right) \tag{5}$$

速度的坐标系变换

预备知识 几何矢量[370]

无相对转动

若两个坐标系 S 和 S' 之间无相对转动（注意我们不要求坐标轴 $\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}$ 和 $\hat{\mathbf{x}}',\hat{\mathbf{y}}',\hat{\mathbf{z}}'$ 同方向），那么某时刻两坐标系之间的相对速度是唯一确定的，即 S' 系中任意一个固定点相对于 S 系任意一个固定点的速度. 我们把 S' 系相对于 S 系的速度记为 $\mathbf{v}_r$ 根据高中所学的速度叠加原理，若某点 P 相对于 S 系的速度瞬时为 $\mathbf{v}_S$，相对于 S' 的瞬时速度为 $\mathbf{v}_{S'}$ 我们有

$$\mathbf{v}_S=\mathbf{v}_{S'}+\mathbf{v}_r \tag{1}$$

其中三个矢量都可以是时间的函数．注意该式与点 P 的位置无关只和速度有关．

注意式 1 中的矢量都是不依赖于坐标系的几何矢量，不能将 $\mathbf{v}_S$ 和 $\mathbf{v}'_S$ 等同于点 P 在 S 系和 S' 系中的三个位置坐标的求导．如果要将式 1 写成分量的形式，三个矢量必须使用同一坐标系．我们来举例说明．

例 1

令 $\hat{\mathbf{y}}' = \hat{\mathbf{x}}$，$\hat{\mathbf{z}}' = \hat{\mathbf{y}}$，$\hat{\mathbf{x}}' = \hat{\mathbf{z}}$，$\mathbf{v}_r = 2\hat{\mathbf{x}} = 2\hat{\mathbf{y}}'$，点 P 在 S 系中坐标关于时间的导数为 $(1, 2, 3)$．请将式 1 表示为三个分量的形式．

容易得出，点 P 在 S' 系中的坐标关于时间的导数为 $(3, -1, 2)$．

我们先来看错误的理解：将 $\mathbf{v}_S$ 等同于 $(1, 2, 3)$，$\mathbf{v}_{S'}$ 等同于 $(3, -1, 2)$，这时会发现，无论 $\mathbf{v}_r$ 取 $(2, 0, 0)$ 还是 $(0, 2, 0)$ 都不能让等式对三个分量同时成立．

正确的做法是将三个矢量都放到同一坐标系中．先使用 S 系，$\mathbf{v}_S$ 的分量仍然取 $(1, 2, 3)$，$\mathbf{v}_{S'}$ 在 S 系中为 $(-1, 2, 3)$，$\mathbf{v}_r$ 在 S 系中取 $(2, 0, 0)$．这样式 1 就成立了．

也可以三个矢量都放到 S' 系中，有 $(3, 1, 2) = (3, -1, 2) + (0, 2, 0)$，同样成立．

对于一般的情况，两参考系中矢量的坐标变换需要使用空间旋转矩阵[404]．

一般情况

对于任意两个坐标系，他们之间的相对运动除了平移可能还有转动，即 $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ 和 $\hat{\mathbf{x}}', \hat{\mathbf{y}}', \hat{\mathbf{z}}'$ 之间的关系可能随时间变化．这时式 1 是否仍然成立呢？

要回答这个问题我们首先要修改 $\mathbf{v}_r$ 的定义．按照上一节的定义，如果坐标系间存在相对转动，$\mathbf{v}_r$ 将与两个固定点的选取有关．若定义某时刻点 P 在两坐标系中的坐标分别为 (x_p, y_p, z_p) 和 (x'_p, y'_p, z'_p)，则 $\mathbf{v}_r$ 可以定义为 S' 系中的固定点 (x'_p, y'_p, z'_p) 相对于 S 系中的固定点 (x_p, y_p, z_p) 的瞬时速度．这时仍有

$$\mathbf{v}_S = \mathbf{v}_{S'} + \mathbf{v}_r \tag{2}$$

再次强调，这三个矢量也是不依赖坐标系的几何矢量．若要记为分量的形式需要使用同一坐标系．

例 2

令 S' 系 $t=0$ 时与 S 系重合并绕 z 轴逆时针以角速度 ω 转动（不一定为匀速），又令点 P 的运动方程为 $\mathbf{r}(t)=\alpha t\hat{\mathbf{x}}'$，验证式 2.

首先将 $\mathbf{r}(t)$ 用 $\hat{\mathbf{x}},\hat{\mathbf{y}}$ 基底表示为

$$\mathbf{r}(t)=\alpha t(\cos\omega t\,\hat{\mathbf{x}}+\sin\omega t\,\hat{\mathbf{y}}) \tag{3}$$

将 $\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}$ 视为常矢量，$\mathbf{r}(t)$ 关于时间求导得点 P 相对于 S 系的速度

$$\mathbf{v}_S=\alpha(\cos\omega t\,\hat{\mathbf{x}}+\sin\omega t\,\hat{\mathbf{y}})+\alpha\omega t(-\sin\omega t\,\hat{\mathbf{x}}+\cos\omega t\,\hat{\mathbf{y}}) \tag{4}$$

将 $\hat{\mathbf{x}}',\hat{\mathbf{y}}',\hat{\mathbf{z}}'$ 视为常矢量，$\mathbf{r}(t)$ 关于时间求导得点 P 相对于 S' 系的速度

$$\mathbf{v}_{S'}=\frac{\mathrm{d}}{\mathrm{d}t}(\alpha t\hat{\mathbf{x}}')=\alpha\hat{\mathbf{a}}=\alpha(\cos\omega t\,\hat{\mathbf{x}}+\sin\omega t\,\hat{\mathbf{y}}) \tag{5}$$

最后，t 时刻两坐标系在点 P 处的相对速度（见式 5 [883]）为

$$\begin{aligned}\mathbf{v}_r&=\boldsymbol{\omega}\times\mathbf{r}=(\omega\hat{\mathbf{z}}')\times(\alpha t\hat{\mathbf{x}}')\\&=\alpha\omega t\hat{\mathbf{z}}'\times\hat{\mathbf{x}}'=\alpha\omega t\hat{\mathbf{y}}'=\alpha\omega t(-\sin\omega t\,\hat{\mathbf{x}}+\cos\omega t\,\hat{\mathbf{y}})\end{aligned} \tag{6}$$

将以上三式代入式 2 可验证其成立. 注意以上我们将所有的矢量用 $\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}$ 基底表示三个矢量，同理我们也可以将所有矢量用 $\hat{\mathbf{x}}',\hat{\mathbf{y}}',\hat{\mathbf{z}}'$.

证明

（未完成）

加速度的坐标系变换

预备知识 速度的参考系变换[892]

无相对转动

类比式 1 [892]，若两个参考系之间只有平移没有转动，令某时刻点 P 相对于 S 系和 S' 系的加速度分别为 $\mathbf{a}_S$ 和 $\mathbf{a}_{S'}$，再令两坐标系中任意两个固定点（例如各自的原点）之间的加速度为 $\mathbf{a}_r$，那么有

$$\mathbf{a}_S=\mathbf{a}_{S'}+\mathbf{a}_r \tag{1}$$

同样地，如果要将该式写成分量的形式，三个矢量必须使用同一坐标系（见例 1 [892]）.

一般情况

类比式 2 [892]，若两参考系之间有可能存在转动，牵连加速度 $\mathbf{a}_r$ 的定义会变得比牵连速度 $\mathbf{v}_r$ 更微妙，因为牵连速度与坐标系选取无关，而牵连加速度却有关！我们举例解释.

例 1　牵连加速度

令 S' 系原点沿着 S 系的单位矢量 $\mathbf{x}$ 匀速运动，且相对 S 系以恒定角速度矢量 $\boldsymbol{\omega} = 2\mathbf{y}$ 转动. 令 $t = 0$ 时两系完全重合，且点 P 恰好经过原点. 我们来讨论此刻原点处的牵连速度.

显然，两参考系 $t = 0$ 时刻的固定点就是各自的原点 O_S 和 $O_{S'}$. $O_{S'}$ 延 S 系的 x 轴匀速运动，所以 S 系的观察者会认为牵连加速度为零. 然而在 S' 系的观察者看来，O_S 始终在做速度不为零的曲线运动，所以 $t = 0$ 时牵连加速度不为零.

我们在 S 系中讨论问题. 定义 t 时刻点 P 在 S' 系中的固定点相对于 S 系的加速度为 $\mathbf{a}_r$. 那么可以证明（比式 1 多出了一项）

$$\mathbf{a}_S = \mathbf{a}_{S'} + \mathbf{a}_r + 2\boldsymbol{\omega} \times \mathbf{v}_{S'} \tag{2}$$

其中 $\boldsymbol{\omega}$ 是 S' 系相对于 S 系的瞬时角速度. 最后一项被称为**科里奥利加速度（Coriolis Acceleration）**

$$\mathbf{a}_c = 2\boldsymbol{\omega} \times \mathbf{v}_{S'} \tag{3}$$

若我们把 S' 相对于 S 的运动分解为原点的平移加绕原点的转动，那么牵连加速度 $\mathbf{a}_r$ 也可以分解为平移加速度和旋转加速度的矢量和，而旋转加速度又可以分为向心加速度和角加速度引起的切向加速度.

$$\mathbf{a}_r = \mathbf{a}_T + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) + \dot{\boldsymbol{\omega}} \times \mathbf{r} \tag{4}$$

其中平移加速度 $\mathbf{a}_T$ 的定义为 S' 系原点相对于 S 系原点（关于 S 系）的加速度.

习题 1

请使用例 2 [892] 的情景验证式 2.

习题 2

试证明式 4.

证明（旋转矩阵）

预备知识 旋转矩阵的导数[418]

我们在 S 系中以坐标的形式证明式 2，即式中的矢量都看作是 S 系中的三个坐标. 令点 P 在两系中的坐标分别为 $\mathbf{r}_S(t)=(x,y,z)^{\mathrm{T}}$ 和 $\mathbf{r}_{S'}(t)=(x',y',z')^{\mathrm{T}}$，且坐标变换可以用一个旋转矩阵 $\mathbf{R}(t)$ 和一个平移矢量 $\mathbf{d}(t)$ 表示为

$$\mathbf{r}_S=\mathbf{R}\mathbf{r}_{S'}+\mathbf{d} \tag{5}$$

两边关于时间求导得[4]

$$\dot{\mathbf{r}}_S=\dot{\mathbf{R}}\mathbf{r}_{S'}+\mathbf{R}\dot{\mathbf{r}}_{S'}+\dot{\mathbf{d}} \tag{6}$$

再求导并整理得

$$\ddot{\mathbf{r}}_S=\mathbf{R}\ddot{\mathbf{r}}_{S'}+(\ddot{\mathbf{R}}\mathbf{r}_{S'}+\ddot{\mathbf{d}})+2\dot{\mathbf{R}}\dot{\mathbf{r}}_{S'} \tag{7}$$

下面我们只需证明这三项分别对应式 2 的各项即可.

在 S 系中，显然有 $\mathbf{a}_S=\ddot{\mathbf{r}}_S$. P 在 S' 系中的加速度为 $\ddot{\mathbf{r}}_{S'}$，乘以旋转矩阵就变换到 S 系中，所以 $\mathbf{a}_{S'}=\mathbf{R}\ddot{\mathbf{r}}_{S'}$.

若 S' 系中的固定点 $\mathbf{r}_{S'}$ 不随时间变化，则求二阶导数得 S' 系中固定点相对于 S 系中固定点的加速度（在 S 系中的坐标）

$$\mathbf{a}_r=\ddot{\mathbf{R}}\mathbf{r}_{S'}+\ddot{\mathbf{d}} \tag{8}$$

再来看式 7 最后一项，将式 1 [418] 代入，得

$$2\dot{\mathbf{R}}\dot{\mathbf{r}}_{S'}=2\boldsymbol{\Omega}(\mathbf{R}\dot{\mathbf{r}}_{S'})=2\boldsymbol{\omega}\times\mathbf{v}_{S'} \tag{9}$$

这就是式 2 的最后一项. 证毕.

[4]用一点表示时间导数，两点表示时间二阶导数

第二章

质点动力学

力的分解与合成

预备知识 几何矢量[370]

在经典力学中，力可以用几何矢量[370]表示．力的分解与合成可以看作一个**基本假设**．这个假设是牛顿运动定律[899]的基础，因为牛顿三定律中的“力”都是指质点所受的合力．

当若干个力 $\mathbf{F}_i$（$i=1,2,\ldots,N$）作用在同一个质点上时，等效于一个力 $\mathbf{F}$ 作用在同一个质点上．

$$\mathbf{F}=\sum_{i=1}^{N}\mathbf{F}_i=\mathbf{F}_1+\mathbf{F}_2+\cdots+\mathbf{F}_N \tag{1}$$

注意这里的加号表示几何矢量[370]的加法而不是数的加法．我们把 $\mathbf{F}$ 叫做 N 个 $\mathbf{F}_i$ 的**合力**，每个 $\mathbf{F}_i$ 叫做一个**分力**．式 1 从左到右的过程叫做**力的分解**，从右到左的过程叫做**力的合成**．

这里所说的“等效”可以指这个质点受力后的运动情况，也可以指物体发生的形变，例如该质点固定在弹簧上，弹簧发生的形变．

回顾两个几何矢量的加法，我们就得到了所谓的**平行四边形法则**或者**三角形法则**．若 $N>2$，用“首尾相接”的方法即可．注意这个过程不需要坐标系的概念．若建立了直角坐标系，我们也可以先计算这些矢量的坐标，然后使用坐标计算矢量加法（式 7 [379]）．

多次分解

注意在式 1 中我们甚至可以进行多次分解，即继续令某个（或每个）力等于若干力相加

$$\mathbf{F}_i=\sum_{j}\mathbf{F}_{i,j} \tag{2}$$

那么 $\mathbf{F}$ 就可以最终分解为

$$\mathbf{F}=\sum_{i,j}\mathbf{F}_{i,j} \tag{3}$$

这仍然符合分解的定义，即一个力矢量表示为多个力矢量相加，本质上并无不同．

牛顿运动定律 惯性系

预备知识 加速度[881]，力的合成与分解[898]

牛顿的三定律可表述如下，为了避免讨论物体的质心及转动，这里我们只讨论质点.

- **第一定律** 不受力或受合力为零的质点做匀速直线运动或静止.
- **第二定律** 质点所受合外力等于质点的质量乘以加速度.
- **第三定律** 两质点的相互作用力等大反向.

第一定律

牛顿第一定律的作用是定义**惯性坐标系（Inertial frame of reference，简称惯性系）**：惯性系存在，且满足牛顿第一定律的参考系就是惯性系.

- **推论** 相对某惯性系静止或匀速运动且没有相对转动的参考系也是惯性系，否则不是惯性系.

推论证明：若已知 A 系为惯性系，B 系相对 A 系的平移速度为 $\mathbf{v}_{AB}$，质点在两系中的瞬时速度分别记为 $\mathbf{v}_A, \mathbf{v}_B$，则由“绝对速度 = 牵连速度 + 相对速度”得

$$\mathbf{v}_B = \mathbf{v}_A + \mathbf{v}_{AB} \tag{1}$$

若 $\mathbf{v}_A$ 与 $\mathbf{v}_{AB}$ 都是常矢量，那么显然 $\mathbf{v}_B$ 也是常矢量，即 B 系为惯性系. 若两系之间有任何相对的加速度（包括加速平移和转动），那么 $\mathbf{v}_{AB}$ 将随时间或位置变化，也就不能保证 $\mathbf{v}_B$ 一定是常矢量，所以 B 系就不是惯性系.

第二定律

牛顿第二定律只能在惯性系中使用，在非惯性系中需要用惯性力[920] 进行修正. 用矢量 $\mathbf{F}$ 表示合力，牛顿第二定律记为

$$\mathbf{F} = m\mathbf{a} \tag{2}$$

高中物理只强调圆周运动和直线运动，所以一般 F 和 a 都记为标量，请读者不要局限在高中思维．这是一条矢量表达式，加速度矢量 $\mathbf{a}$ 是位置矢量 $\mathbf{r}$ 关于时间的二阶导数，或者速度矢量 $\mathbf{v}$ 关于时间的导数．加速度和速度不必共线也不必垂直．

动量定理

牛顿本人对第二定律的表述使用了动量定理[910]，记质点的动量为 $\mathbf{p}$，则

$$\mathbf{F}=\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} \tag{3}$$

在经典力学中，由于质量不发生变化，式 2 和式 3 是等效的，但令人惊讶的是，牛顿所用的形式在狭义相对论中仍然成立[1]，而式 2 却不成立．

第三定律

广义来说，牛顿第三定律就是动量守恒定律[944]．牛顿第三定律在任何参考系中都适用，但是要注意两点．第一，在非惯性系中，由于惯性力作为一个数学上的修正，并不是真正的力，所以不存在反作用力．第二，在考虑电磁力时，由于电磁场可能具有动量，所以动量守恒定律要求所有物体与电磁场的动量之和守恒，而不仅仅是质点的总动量守恒．在考虑两带电粒子的相互作用力时，若假设粒子的运动速度较慢，则磁场可以忽略，电磁场动量始终为零，此时两粒子的总动量守恒，相互作用力等大反向．

圆周运动的向心力

预备知识 圆周运动的加速度[885]，牛顿运动定律[899]

在惯性参考系中，质点的运动符合牛顿第二定律 $\mathbf{F}=m\mathbf{a}$．所以要使一个质点做半径为 R 的圆周运动，那么我们就可以把任意时刻圆周运动的加速度矢量（式 4 [885]）乘以质量 m 得到它该时刻受到的合力．我们把这个力叫做**向心力（centripetal force）**．

$$\mathbf{F}_c=-m\omega^2\mathbf{r}=-\frac{mv^2}{r}\hat{\mathbf{r}}=m\boldsymbol{\omega}\times\mathbf{v} \tag{1}$$

[1]在狭义相对论中，动量的定义有所不同．

我们需要注意区分向心力和离心力[924]，向心力不是圆周运动给物体施加的力，而是维持圆周运动所需要的力．如果我们使用惯性参考系，就不存在离心力的概念．只有当我们使用旋转的参考系（非惯性系），才能讨论离心力，因为离心力是一种惯性力[920]，是在非惯性系中讨论物体运动的一个数学工具．

那么为什么例如在车转弯的时候我们会感觉有离心力把我们向外推呢？事实上人的惯性使得人有继续保持直线运动的倾向，而车突然改变方向偏离原来的直线，所以在地面惯性系看来是车在向内推人，但人因身处车中，本能地觉得车是静止的，以为牛顿第二定律仍然成立，所以感觉自己受到了一种无形的离心力．这是一种错觉，就像我们本能地认为光是沿直线传播，所以会觉得伸进水中的筷子被折弯了，或者觉得镜子里面还有一个空间．

有趣的推导

为了更好地理解为何向心力会产生圆周运动，我们可以假设一个小球在光滑水平面运动的过程中被反弹若干次，使得它的轨道称为闭合的正 N 边形，每个顶点距离中心为 R．每次撞击后，速度不变，方向改变 $2\theta = 2\pi/N$．

未完成：图

所以在每次碰撞时，我们可以把速度垂直分解为径向（指向圆心）和角向速度．碰撞后角向速度不变，径向速度增量为 $2v\sin\theta$，即动量改变为

$$\Delta p = 2mv\sin\theta \tag{2}$$

正 N 边形每条边的长度为 $L = 2R\sin(\theta)$，所以相邻两次碰撞的时间为

$$\Delta t = \frac{2R\sin\theta}{v} \tag{3}$$

我们可以令 $N \to \infty$，这样，多边形轨迹就趋近于圆周运动，每次撞击时，冲量趋近于 0，时间间隔也趋近于 0，我们就可以把这些离散的撞击看成时一个连续的力．平均来说，力的大小等于动量改变除以时间

$$F_c = \frac{\Delta p}{\Delta t} = \frac{mv^2}{R} = m\omega^2 R \tag{4}$$

注意该式甚至不需要取极限 $\Delta t \to 0$ 就可以成立．

未完成：例题和习题

重力 重量

预备知识 万有引力[1027]，圆周运动的向心力[900]

一般来说，重力的定义并不明确．有的地方直接把一个质点在某点的**万有引力（gravity）**定义为重力．但另一些情况下把弹簧秤的度数定义为重力（例如在地面参考系）．为了区分，我们建议不要使用“重力”，而是直接将前者称为（万有）引力，后者称为**重量（weight）**或者**视重**．英语中，“重力”并没有单独对应的词汇．

什么情况下引力会和重量不同？答案是当质点（在惯性系中）具有加速度时．最常见的例子就是地球表面虽然可以近似为惯性系，但严格来说却不是，所以相对地面静止的一点在惯性系中具有加速度．如果只考虑地球的自转产生的加速度[2]，那么地表任意一个相对静止的点[3]在（相对地轴静止的）惯性系中都会做圆周运动，它的加速度等于向心加速度．

惯性系中的分析

为了方便，我们先选取与地轴相对静止的惯性系分析．当一个相对地表静止的质点挂在弹簧秤上达到平衡时，它所受的引力 $\mathbf{G}$ 和弹簧对它的拉力 $\mathbf{T}$（或者台秤的支持力）的合力提供圆周运动的向心力[900]．

$$\mathbf{F}_c = \mathbf{G} + \mathbf{T} \tag{1}$$

所以不妨定义重量矢量为弹簧拉力的逆矢量

$$\mathbf{F}_w = -\mathbf{T} = \mathbf{G} - \mathbf{F}_c \tag{2}$$

那么重量，也就是弹簧秤的示数大小就是模长 $F_w = |\mathbf{F}_w|$．

可以看出，当质点无加速度（$\mathbf{F}_c = \mathbf{0}$）时，重量矢量与引力相等．

（未完成：计算一下重力加速度）

[2]地球绕太阳系公转的加速度远小于自转，一般可忽略不计，有兴趣的读者可以自行计算．

[3]除了两个极点

地表参考系中的分析

预备知识 离心力[924]

非惯性系中，保持平衡的条件同样是合力为零，但合力要包括惯性力．在地表参考系，惯性力就是地球自转的离心力 $\mathbf{F}'_c$.

$$\mathbf{G}+\mathbf{T}+\mathbf{F}'_c=\mathbf{0} \tag{3}$$

由于离心力刚好与向心力相反，即 $\mathbf{F}'_c=-\mathbf{F}_c$，代入后仍然可以得到式 1．注意离心力并不真的存在，只是一个数学上的“把戏”.

功 功率

预备知识 矢量的内积[383]，定积分[198]

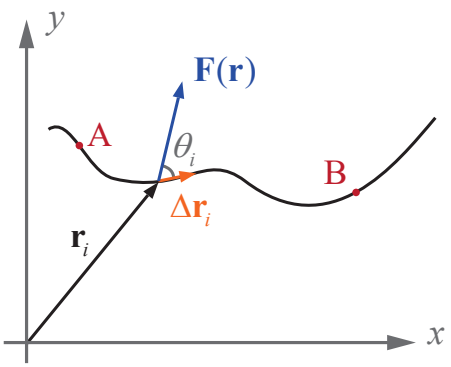


图 1: 在一小段位移中，把变力看做恒力

如图 1，当质点沿着曲线运动时，有一个力作用在其上，当质点的位置为 $\mathbf{r}$ 时，力为 $\mathbf{F}(\mathbf{r})$．下面求质点从点 A 运动到点 B 的过程中，力对质点的做功.

把从 A 到 B 这段曲线看成由许多小位移 $\Delta\mathbf{r}_1,\Delta\mathbf{r}_2\ldots\Delta\mathbf{r}_n$ 组成，对其中第 i 个进行分析．由于 $\Delta\mathbf{r}_i$ 很短，质点经过 $\Delta\mathbf{r}_i$ 的过程中位矢 $\mathbf{r}$ 几乎不变，记为常矢量 $\mathbf{r}_i$．在这小段中，$\mathbf{F}(\mathbf{r})$ 也可以近似看成是恒力 $\mathbf{F}(\mathbf{r}_i)$.

现在把 $\mathbf{F}(\mathbf{r}_i)$ 分解成垂直于 $\Delta\mathbf{r}_i$ 和平行于 $\Delta\mathbf{r}_i$ 的两个正交分量，其中垂直分量不做功，平行分量的大小为 $|\mathbf{F}(\mathbf{r}_i)|\cos\theta_i$，该分量做功大小为

$$\Delta W_i=|\mathbf{F}(\mathbf{r}_i)||\Delta\mathbf{r}_i|\cos\theta_i \tag{1}$$

上式可以表示成矢量内积[383]的形式

$$\Delta W_i = \mathbf{F}(\mathbf{r}_i) \cdot \Delta \mathbf{r}_i \tag{2}$$

把上式对所有的 i 求和，就得到了做功的近似表达式

$$W_{ab} = \sum_{i=1}^{n} \Delta W_i \approx \sum_{i=1}^{n} \mathbf{F}(\mathbf{r}_i) \cdot \Delta \mathbf{r}_i \tag{3}$$

事实上，当曲线分割的越细，即 n 越大时，上式就越精确地成立．类比定积分[198]中的介绍，令 $n \to \infty$，把求和符号换成积分符号，把表示增量的 Δ 换成微分符号 d，则不等号可以变为等号．

$$W_{ab} = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{F}(\mathbf{r}_i) \cdot \Delta \mathbf{r}_i = \int_{C_{ab}} \mathbf{F}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} \tag{4}$$

不同于一元函数的积分，这一类特殊的积分叫做**线积分**，详见“线积分[295]”.

力的功率

功率（瞬时）的定义为做功的变化率，即

$$P = \lim_{\Delta t\to 0} \frac{\Delta W}{\Delta t} = \frac{\mathrm{d}W}{\mathrm{d}t} \tag{5}$$

根据式 2，力的功率为

$$P = \lim_{\Delta t\to 0} \frac{\mathbf{F}(\mathbf{r}_i) \cdot \Delta \mathbf{r}_i}{\Delta t_i} = \mathbf{F} \cdot \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \mathbf{F} \cdot \mathbf{v} \tag{6}$$

动能 动能定理（单个质点）

预备知识 功 功率[903]，牛顿第二定律[899]

令质点的质量为 m，速度[881]为 $\mathbf{v}$，则质点的**动能**定义为

$$E_k = \frac{1}{2} m \mathbf{v}^2 = \frac{1}{2} m v^2 \tag{1}$$

注意这里的 $\mathbf{v}^2 = \mathbf{v} \cdot \mathbf{v}$ 表示速度矢量和自身的内积[471]，结果等于其模长的平方．

定理 1　质点动能定理

一段时间内质点动能的变化等于合外力对质点做的功．从变化率（即时间导数）的角度来看，动能定理也可以表述为质点的动能变化率等于合外力对质点的功率．

推导

力对质点做功的功率[903] 为

$$P=\frac{\mathrm{d}W}{\mathrm{d}t}=\mathbf{F}\cdot\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}=\mathbf{F}\cdot\mathbf{v} \tag{2}$$

再来看动能的变化率

$$\frac{\mathrm{d}}{\mathrm{d}t}E_k=\frac{1}{2}m\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{v}\cdot\mathbf{v}) \tag{3}$$

由“矢量内积的求导[274]”式 6，$\mathrm{d}(\mathbf{v}\cdot\mathbf{v})/\mathrm{d}t=2\mathbf{v}\cdot\mathrm{d}\mathbf{v}/\mathrm{d}t=2\mathbf{v}\cdot\mathbf{a}$，上式变为

$$\frac{\mathrm{d}}{\mathrm{d}t}E_k=m\mathbf{a}\cdot\mathbf{v}=\mathbf{F}\cdot\mathbf{v} \tag{4}$$

最后一步使用了牛顿第二定律（式 2 [899]）．注意式 2 与式 4 相等，所以动能变化率等于合外力的功率．

力场 势能

预备知识 矢量场[294]，功[903]，牛顿—莱布尼兹公式[202]

力场

高中物理中我们已经学过一些场的概念，即质点受场的力取决于质点在场中的位置．例如地球表面局部的引力场可以近似看做一个恒力场（称为为**重力场**），即在一定区域内，质点总受向下的，大小恒为 mg 的重力（矢量式 $\mathbf{F}=m\mathbf{g}$）．又例如水平面上一根原长忽略不计的弹簧，一端固定在原点，另一端连接质点，那么质点受力总指向原点，大小等于劲度系数和位矢模长的之积 kr．用矢量的方法表示，就是 $\mathbf{F}=-k\mathbf{r}$．

总结到一般情况，**力场**可以用场对质点施加的力（矢量）关于质点位置（即位矢[878]）的矢量函数表示，所以力场是一种矢量场．

例 1　引力场

球坐标原点处质量为 M 的质点在周围造成的引力场为

$$\mathbf{F}(\mathbf{r}) = -G\frac{M}{r^2}\hat{\mathbf{r}} \tag{1}$$

若位矢用 $\mathbf{r}$ 来表示（$\mathbf{r} = r\hat{\mathbf{r}}$），则

$$\mathbf{F}(\mathbf{r}) = -G\frac{M}{r^3}\mathbf{r} \tag{2}$$

现在变换到直角坐标系中，有

$$\begin{cases} \mathbf{r} = x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}} \\ r = \sqrt{x^2+y^2+z^2} \end{cases} \tag{3}$$

代入上式，展开得

$$\mathbf{F}(\mathbf{r}) = -\frac{GMx}{(x^2+y^2+z^2)^{3/2}}\hat{\mathbf{x}} - \frac{GMy}{(x^2+y^2+z^2)^{3/2}}\hat{\mathbf{y}} - \frac{GMz}{(x^2+y^2+z^2)^{3/2}}\hat{\mathbf{z}} \tag{4}$$

显然球坐标系中的引力场表达式比直角坐标系中的要简洁得多．由此可见，对不同的矢量场选择适当的坐标系往往可以简化问题．

若质点从场的一点移动到另一点的过程中，力场对质点做的功[903] 只与初末位置有关，而与质点移动的路径无关，那么这个力场就是一个**保守场**．这时我们可以给该质点定义一个**势能函数**，势能函数是一个关于位矢的标量函数，一般记为 $V(\mathbf{r})$，具有能量量纲．当质点从一点以任意路径移动到另一点时，场对质点做的功等于质点初位置的势能减末位置的势能，即

$$\int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{F}(\mathbf{r})\cdot \mathrm{d}\mathbf{r} = V(\mathbf{r}_1) - V(\mathbf{r}_2) \tag{5}$$

一维势能函数

现在先假设质点只能沿一条直线运动，且力也始终与直线平行．显然质点从一点到另一点的路径只可能有一条，所以任何一维力场都是保守场．若给直线定义一个正方向，单位矢量为 $\hat{\mathbf{x}}$，任何一维力场可以记为

$$\mathbf{F}(x) = F(x)\hat{\mathbf{x}} \tag{6}$$

质点的位置矢量可记为 $\mathbf{r}=x\hat{\mathbf{x}}$．由于 $\hat{\mathbf{x}}\cdot\hat{\mathbf{x}}=1$，质点从 $x=a$ 移动到 $x=b$ 过程中场做的功为

$$\int_{\mathbf{r}_1}^{\mathbf{r}_2}\mathbf{F}(\mathbf{r})\cdot\mathrm{d}\mathbf{r}=\int_a^b F(x)\,\mathrm{d}x \tag{7}$$

根据势能的定义，对任意的 a 和 b，上式应该等于 $V(a)-V(b)$．根据牛顿—莱布尼兹公式[202]，势能函数恰好就是 $F(x)$ 的负原函数，所以 $F(x)$ 是 $V(x)$ 负导函数．

$$V(x)=-\int F(x)\,\mathrm{d}x\qquad F(x)=-\frac{\mathrm{d}V(x)}{\mathrm{d}x} \tag{8}$$

需要注意的是，由于原函数有无穷多个（由不定积分中任意常数的取值决定），所以势能函数也存在无穷多个，且都相差一个常数．为了确定势能函数，我们需要指定场中某一点的势能值，如果令某点势能为零，那么这点就叫做**零势点**．

例 2　弹簧的势能

一个原长可忽略的轻弹簧劲度系数为 k，一端固定在原点，另一端连接质点．质点只能沿 $\hat{\mathbf{x}}$ 方向运动，规定质点在原点时势能为 0，求弹簧的势能关于质点位置坐标 x 的函数．

由题意，式 6 中 $F(x)=-kx$，不定积分并取负值得到含有待定常数的势能函数

$$V(x)=-\int(-kx)\,\mathrm{d}x=\frac{1}{2}kx^2-C \tag{9}$$

为了确定待定常数，代入 $V(0)=0$，解得 $C=0$．所以所求势能为

$$V(x)=\frac{1}{2}kx^2 \tag{10}$$

多维势能函数

预备知识　梯度定理[307]

假设力场 $\mathbf{F}(\mathbf{r})$ 是平面或三维空间中的保守场，对应势能为 $V(\mathbf{r})$，初始点为 $\mathbf{r}_i$，终点为 $\mathbf{r}_f$．对 $-V(\mathbf{r})$ 使用梯度定理[307] 得

$$\int_{\mathbf{r}_i}^{\mathbf{r}_f}\boldsymbol{\nabla}[-V(\mathbf{r})]\cdot\mathrm{d}\mathbf{l}=V(\mathbf{r}_i)-V(\mathbf{r}_f) \tag{11}$$

我们把该式与式 5 比较，不难发现力场是势能函数的负梯度

$$\mathbf{F}(\mathbf{r}) = -\boldsymbol{\nabla} V(\mathbf{r}) \tag{12}$$

即在保守场的某点中，力的方向是势能下降最快的方向，大小是该方向的负方向导数．由梯度的定义，力场的各个分量分别为对应方向的负偏导数

$$F_x(\mathbf{r}) = -\frac{\partial V(\mathbf{r})}{\partial x} \quad F_y(\mathbf{r}) = -\frac{\partial V(\mathbf{r})}{\partial y} \quad \dots \tag{13}$$

例 3　二维简谐振子

若已知二维的势能函数为 $V(x,y) = k_1(x+y)^2/2 + k_2(x-y)^2/2$，求力场．若已知场函数求势能函数，又该如何求？

把势能函数代入式 13 中，求偏导，得场为

$$\begin{aligned} \mathbf{F}(\mathbf{r}) &= -\frac{\partial V}{\partial x}\hat{\mathbf{x}} - \frac{\partial V}{\partial y}\hat{\mathbf{y}} \\ &= -[(k_1+k_2)x + (k_1-k_2)y]\hat{\mathbf{x}} - [(k_1-k_2)x + (k_1+k_2)y]\hat{\mathbf{y}} \end{aligned} \tag{14}$$

现在我们根据“梯度定理[307]”中的式 26 从场逆推势能．首先对力场的 x 分量和 y 分量分别关于 x 和 y 做不定积分得到任意两个原函数并记为 G_x 和 G_y 得

$$G_x(x,y) = -\frac{1}{2}(k_1+k_2)x^2 - (k_1-k_2)xy \tag{15}$$

$$G_y(x,y) = -(k_1-k_2)xy - \frac{1}{2}(k_1+k_2)y^2 \tag{16}$$

代入得（注意这里的场是势能函数的的负梯度而不是梯度，另外注意下式中的常数项都并入 C 中）

$$\begin{aligned} V(x,y) &= -G_y(x,y) + G_y(x,y_0) - G_x(x,y_0) + C \\ &= \frac{1}{2}(k_1+k_2)x^2 + (k_1-k_2)xy + \frac{1}{2}(k_1+k_2)y^2 + C \\ &= \frac{1}{2}k_1(x+y)^2 + \frac{1}{2}k_2(x-y)^2 + C \end{aligned} \tag{17}$$

若规定零势点 $V(0,0) = 0$，代入上式得 $C = 0$.

两质点间的势能

如果两质点 A 和 B 的位矢分别为 $\mathbf{r}_A$ 和 $\mathbf{r}_B$，相对位移为 $\mathbf{R}=\mathbf{r}_B-\mathbf{r}_A$，两质点距离为 $R=|\mathbf{R}|$．且 A 对 B 的作用力为 $\mathbf{F}=F(R)\hat{\mathbf{R}}$，$B$ 对 A 的反作用力为 $-\mathbf{F}$．现在考虑一个过程中力对两质点做的总功．

在一段微小时间 $\mathrm{d}t$ 内，两质点分别移动了 $\mathrm{d}\mathbf{r}_A$，和 $\mathrm{d}\mathbf{r}_B$，则相互作用力对二者做功为

$$\mathrm{d}W=\mathbf{F}\cdot\mathrm{d}\mathbf{r}_B+(-\mathbf{F})\cdot\mathrm{d}\mathbf{r}_A=\mathbf{F}\cdot\mathrm{d}\mathbf{R}=F(R)\hat{\mathbf{R}}\cdot\mathrm{d}\mathbf{R}=F(R)\,\mathrm{d}R \tag{18}$$

（最后一步的证明见“位置矢量[878]”中的例 1）定积分得

$$W=\int_{R_1}^{R_2}F(R)\,\mathrm{d}R \tag{19}$$

现在我们借用一维势能的定义式 8 来定义势能函数为 $F(R)$ 的负原函数，则力在一段时间内对两质点做的总功就等于末势能减初势能

$$W=V(R_2)-V(R_1) \tag{20}$$

含时势能

以上的讨论中，我们默认力场的分布不随时间变化，所得势能显然也不随时间变化．但在一些情况下，我们也可以定义随时间变化的势能．

机械能守恒（单个质点）

预备知识 动能定理[904]，势能[905]

若质点只受不随时间变化的保守力作用[4]，那么物体在某段时间内从 A 点移动到 B 点，力场对物体做功能等于初末势能函数之差

$$W_{AB}=V(\mathbf{r}_A)-V(\mathbf{r}_B) \tag{1}$$

[4]即势能函数不随时间变化

而根据动能定理，力场对质点做功等于质点的末动能减初动能

$$W_{AB} = E_{kB} - E_{kA} = \frac{1}{2}mv_B^2 - \frac{1}{2}mv_A^2 \tag{2}$$

结合以上两式，得

$$E_{kA} + V(\mathbf{r}_A) = E_{kB} + V(\mathbf{r}_B) \tag{3}$$

我们现在定义质点在某个时刻的动能加势能为**机械能**. 上式就是单个质点机械能守恒的表达式.

动量 动量定理（单个质点）

预备知识 牛顿第二定律[899]

令质点质量为 m，速度为 $\mathbf{v}$，定义其**动量**为

$$\mathbf{p} = m\mathbf{v} \tag{1}$$

注意动量是矢量，与速度（矢量）的方向相同，且取决于坐标系.

现在把动量和速度都看做时间的函数. 等式两边求导，速度对时间的导数等于加速度 $\mathbf{a}$

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = m\mathbf{a} \tag{2}$$

根据牛顿第二定律，$m\mathbf{a}$ 等于质点所受合外力 $\mathbf{F}$（注意力和加速度也都是时间的函数），所以

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{F} \tag{3}$$

这就是**动量定理**，即动量的变化率等于合外力. 在牛顿力学中，动量定理和牛顿第二定律是完全等效的.

动量定理也可以写成微分形式

$$\mathrm{d}\mathbf{p} = \mathbf{F}\,\mathrm{d}t \tag{4}$$

也就是在极微小时间内的动量变化等于力乘以这段时间.

现在用定积分[198]中的微元思想考虑动量从时刻 t_1 到 t_2 的总变化，我们可以把这段时间划分为 N 段微小时间，第 i 段所在的时刻记为 t_i，每小段时间内 $\mathbf{F}$ 可认为是恒力 $\mathbf{F}(t_i)$

$$\mathbf{p}(t_2)-\mathbf{p}(t_1)=\sum_{i=1}^{N}\Delta\mathbf{p}_i=\sum_{i=1}^{N}\mathbf{F}(t_i)\Delta t_i \tag{5}$$

当 $N\to\infty,\Delta t\to 0$ 时该式可以用定积分（矢量函数）表示[5]

$$\mathbf{p}(t_2)-\mathbf{p}(t_1)=\int_{t_1}^{t_2}\mathbf{F}(t)\,\mathrm{d}t \tag{6}$$

这是**动量定理**的积分形式．特殊地，对于恒力 $\mathbf{F}$，右边的积分等于 $(t_2-t_1)\mathbf{F}$，上式记为

$$\Delta\mathbf{p}=\mathbf{F}\Delta t \tag{7}$$

角动量 角动量定理 角动量守恒（单个质点）

预备知识 牛顿第二定律[899]，力矩[947]

质点的角动量

在我们讨论物体的转动时通常会引入**角动量（angular momentum）**的概念，这与直线运动中的动量有许多相似之处．

定义 1　角动量

一个质点的质量为 m，某时刻速度为 $\mathbf{v}$，则其动量为 $\mathbf{p}=m\mathbf{v}$．在三维空间中建立坐标系，原点为 O，O 点到质点的位置矢量[878]为 $\mathbf{r}$．定义该质点关于 O 点的角动量为

$$\mathbf{L}=\mathbf{r}\times\mathbf{p}=m\,\mathbf{r}\times\mathbf{v} \tag{1}$$

由叉乘的几何定义[387]可知，当速度与位矢平行时角动量为 $\mathbf{0}$，垂直时角动量模长为距离和动量模长的积 $L=rp$.

[5]通常省略以上的推导而直接表达为“式 4 两边定积分得到式 6”

注意角动量的值取决于参考系的选取，如果我们选用不同的参考系，那么式 1 中的 $\mathbf{r}$ 就会发生改变，进而改变角动量．事实上我们也可以取空间中的任意一点 $\mathbf{r}_0$ 作为角动量的参考点，定义角动量为

$$\mathbf{L} = (\mathbf{r} - \mathbf{r}_0) \times \mathbf{p} \tag{2}$$

这相当于以 $\mathbf{r}_0$ 为原点建立一个新的参考系，然后在其中计算角动量．

例 1　质点圆周运动的角动量

考虑一个质量为 m 的质点绕坐标原点做半径为 r 的逆时针匀速圆周运动

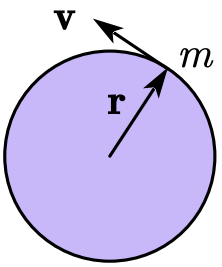


图 1: 圆周运动的角动量

（图 1），角速度为 ω，速度为 $v = \omega r$．对于圆周运动，位置矢量和速度矢量始终垂直，令 $\hat{\mathbf{z}}$ 为垂直纸面向上的单位矢量，则角动量为

$$\mathbf{L} = mvr\hat{\mathbf{z}} = mr^2\omega\hat{\mathbf{z}} \tag{3}$$

所以质点做圆周运动时，角动量矢量不随时间变化，即下文将要介绍的角动量守恒．

例 2　质点直线运动的角动量

虽然角动量通常用于描述转动，但也可以描述质点任意运动．例如当质点

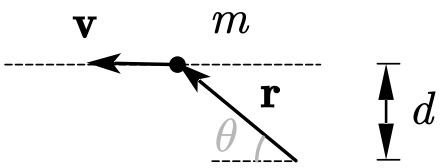


图 2: 直线运动的角动量

做匀速直线运动，运动轨迹与原点的距离为 d，则角动量始终为

$$\mathbf{L} = mvr\sin\theta\hat{\mathbf{z}} = mvd\hat{\mathbf{z}} \tag{4}$$

所以做直线运动的质点同样有角动量守恒.

单质点角动量定理

令质点在某时刻受到的力矩[947]为 $\boldsymbol{\tau}$，可以证明

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t}=\boldsymbol{\tau} \tag{5}$$

这就是（单个质点的）**角动量定理**.

特殊地，若质点受到的力矩为零，则 $\mathrm{d}\mathbf{L}/\mathrm{d}t=\mathbf{0}$，即角动量不随时间变化. 这个现象叫做**角动量守恒（conservation of angular momentum）**. 由力矩的定义（式 3 [947]）$\boldsymbol{\tau}=\mathbf{r}\times\mathbf{F}$，可见以下两种情况下力矩为零，角动量守恒.

1. 质点受合力 $\mathbf{F}=\mathbf{0}$，即质点静止或做匀速直线运动（例 2 ）.
2. $\mathbf{F}$ 与 $\mathbf{r}$ 同向，即质点只受关于 O 点的**有心力**. 如例 1 的圆周运动中，向心力就属于有心力.

证明

我们来证明单个质点的角动量定理. 令质点的速度为 $\mathbf{v}=\mathrm{d}\mathbf{r}/\mathrm{d}t$，加速度为 $\mathbf{a}=\mathrm{d}\mathbf{v}/\mathrm{d}t$，叉乘的求导法则（式 7 [274]）与标量乘法求导类似，牛顿第二定律[899]为 $\mathbf{F}=m\mathbf{a}$，两个同方向矢量叉乘为零，

$$\begin{aligned}\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t}&=\frac{\mathrm{d}(\mathbf{r}\times\mathbf{p})}{\mathrm{d}t}=m\frac{\mathrm{d}(\mathbf{r}\times\mathbf{v})}{\mathrm{d}t}=m\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\times\mathbf{v}+\mathbf{r}\times\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}\right)\\&=m(\mathbf{v}\times\mathbf{v}+\mathbf{r}\times\mathbf{a})=\mathbf{r}\times(m\mathbf{a})\\&=\mathbf{r}\times\mathbf{F}=\boldsymbol{\tau}\end{aligned} \tag{6}$$

证毕.

简谐振子

预备知识 胡克定律，牛顿第二定律[899]

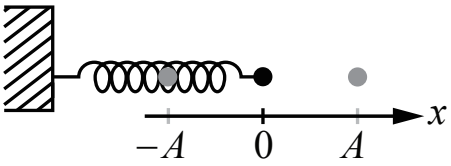


图 1: 简谐振子模型

如图 1，质量为 m 的质点固定在弹性系数为 k 的弹簧的一端，弹簧另一端固定．在 $t=0$ 时，若质点不在平衡位置，或者有一个初速度，则接下来会发生振动（忽略弹簧的质量，任何摩擦以及重力）．以质点拉伸弹簧的方向为 x 轴正方向，质点的平衡位置为 $x=0$．当质点在位置 x 时，根据胡克定律，受力为 $F=-kx$．根据牛顿第二定律[899] $F=ma=m\ddot{x}$ （$\ddot{x}$ 代表对时间的二阶导数）．两式消去 F，得

$$m\ddot{x}=-kx \tag{1}$$

这是一个单变量函数 $x(t)$ 与其二阶导数的关系式．我们把这样含有单变量函数及其导数或高阶导数的等式叫做**常微分方程**[211]．由于上式中最高阶导数是二阶，所以叫做**二阶微分方程**．要解该方程，就是要寻找一个函数 $x(t)$，使它的二阶导数与 $-x(t)$ 成正比，比例系数为 k/m．注意到 $\cos''t=-\cos t$ 具有类似的性质[6]，不妨继续猜测 $x=\cos(\omega t)$，则 $\ddot{x}=-\omega^2\cos\omega t$．所以只要令 $\omega=\sqrt{k/m}$ 即可满足方程．这说明，弹簧的振动可以用余弦函数来描述．但是这只是方程的一个解．任意情况的振动可以表示为以下函数（令 A 和 ϕ_0 为两个任意实数）

$$x=A\cos(\omega t+\phi_0)\qquad\left(\omega=\sqrt{k/m}\right) \tag{2}$$

这叫做微分方程式 1 的**通解**（系统的方法参考二阶常系数齐次微分方程的通解[215] ），即无论常数 A,ϕ_0 取任意值，微分方程总能得到满足．

满足这种形式的运动叫做**简谐运动（或简谐振动）**．其中 A 为**振幅**，$\omega t+\phi_0$ 为**相位**，ϕ_0 为**初相位**（即 $t=0$ 时刻的相位）．但是如何决定 A 和 ϕ_0 呢？根据上面给出的条件还不能判断．由于有两个待定常数，我们需要两个额外条件才能解出．常见的情况是给出初始时刻 $t=0$ 时质点的位置 $x(0)$ 和速度 $\dot{x}(0)$，这就叫做**初值条件**．

例如给出 $x(0)=0$，$\dot{x}(0)=v_0$，把方程的通解代入，得 $A\cos\phi_0=0$，

[6] $\sin t$ 也有同样的性质，所以以下讨论对 $\sin t$ 也成立

$-A\omega\sin\phi_0 = v_0$，解得 $\phi_0 = \pi/2$，$A = -v_0\omega$．所以

$$x = -v_0\omega\cos\left(\omega t + \frac{\pi}{2}\right) = v_0\omega\sin\omega t \qquad \left(\omega = \sqrt{k/m}\right) \tag{3}$$

能量

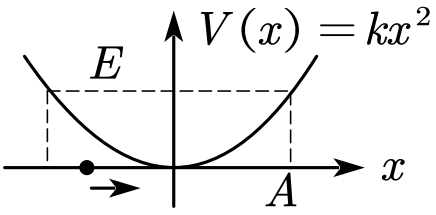


图 2: 简谐振子势能曲线

简谐振子的总能量等于质点动能加弹簧的弹性势能．当位移最大时，势能为 0，总能量等于势能，位移为 0 时势能为 0，总能零等于动能

$$E = \frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \frac{1}{2}kA^2 = \frac{1}{2}mv_0^2 \tag{4}$$

其中 v_0 是质点经过原点处的速度．

受阻落体

预备知识 匀加速运动[889]

在自由落体的基础上，若假设质点受到的空气阻力的大小与其速度成正比，比例系数为 α，那么根据牛顿第二定律[899] 可以列出动力学方程（假设向下为正方向）

$$ma = F = mg - \alpha v \tag{1}$$

考虑到加速度是速度的导数，上式变为

$$\frac{\mathrm{d}v}{\mathrm{d}t} = g - \frac{\alpha}{m}v \tag{2}$$

这是速度关于时间的函数 $v(t)$ 与其一阶导数 $\dot{v}(t)$ 的关系式，即微分方程[211]．与自由落体问题不同的是，这个方程的右边含有未知函数 $v(t)$，所以不可能直

接将等式两边积分解得 $v(t)$．我们可以根据微分与导数的关系，将上式两边同乘 $\mathrm{d}t$ 并整理得

$$\frac{1}{g-\alpha v/m}\,\mathrm{d}v=\mathrm{d}t \tag{3}$$

这样我们就得到了 v 和 t 的微分[173] 关系，即每当 t 增加一个微小量时，如何求 v 对应增加的微小量．注意等式左边仅含 v，右边仅含 t，所以这一步叫做**分离变量**，我们称式 2 为**可分离变量**的微分方程．假设 v 和 t 之间的关系可以表示为

$$F(v)=G(t) \tag{4}$$

那么对等式两边微分即可得到式 3 的形式．令 $f(v)$ 和 $g(t)$ 分别为 $F(v)$ 和 $G(t)$ 的导函数，有

$$f(v)\,\mathrm{d}v=g(t)\,\mathrm{d}t \tag{5}$$

对比式 3 可得 $f(v)=1/(g-\alpha v/m)$ 和 $g(t)=1$，把二者做不定积分[187] 得原函数．首先显然 $G(t)=t+C_1$．对 $f(v)$ 积分可用“积分表[192]”中的式 1 和 式 3 得

$$F(v)=-\frac{m}{\alpha}\ln\left|g-\frac{\alpha}{m}v\right|+C_2=-\frac{m}{\alpha}\ln\left(g-\frac{\alpha}{m}v\right)+C_2 \tag{6}$$

上式中绝对值符号可去掉是因为在式 2 中根据物理情景可知 $\mathrm{d}v/\mathrm{d}t$ 始终大于零．把两原函数代回式 4 （这时可以把 C_1 和 C_2 合并为一个待定常数 C），整理可得

$$v=\frac{m}{\alpha}\left(g-\mathrm{e}^{-\alpha C/m}\mathrm{e}^{-\alpha t/m}\right) \tag{7}$$

这就是微分方程式 2 的通解，可代入原微分方程以验证是否成立．以后我们把以上这种由式 5 形式求式 4 形式的步骤简称为“对方程两边积分”．由于方程阶数为 1，通解仅含有一个待定常数．为了确定这个待定常数，我们用题目给出的**初值条件**，即 $t=0$ 时 $v=0$，代入通解可解得 C，再把 C 代回通解得满足初始条件的**特解**

$$v(t)=\frac{mg}{\alpha}\left(1-\mathrm{e}^{-\alpha t/m}\right) \tag{8}$$

从该式可以看出，当 $t=0$ 时，质点速度为 0，符合初始条件，而当 $t\to+\infty$ 时，$v(t)\to mg/\alpha$．可见质点的速度会无限趋近一个最大值，而这个最大值恰好可以使阻力 αv 等于重力 mg．利用这一条件，即使不解微分方程，也可以很快算出质点的末速度．

单摆

预备知识 简谐振子[913]，匀速圆周运动[885]

理想的单摆由一个质点和一个质量不计的细绳（或细杆）组成．绳的一头连接质点，另一头固定不动．我们来对单摆做受力分析．如图 1，令质点质量为 m，受重力大小为 mg，受绳的拉力大小为 T．将重力沿与绳平行的方向和垂直的方向正交分解，分力大小分别为 $mg\cos\theta$ 和 $mg\sin\theta$．由于绳的限制，质点只允许做圆周运动，所以绳的拉力与重力平行绳的分量必然提供质点的向心力．

$$T - mg\cos\theta = ma_c \tag{1}$$

对于变速圆周运动，向心加速度仍然可以用 $a_c = v^2/L$ 求解，其中 L 是绳长即圆的半径（证明见“变速圆周运动”）．在求单摆运动时，拉力 T 的大小并不重要，我们更关心的是摆角 θ 随时间的变化．

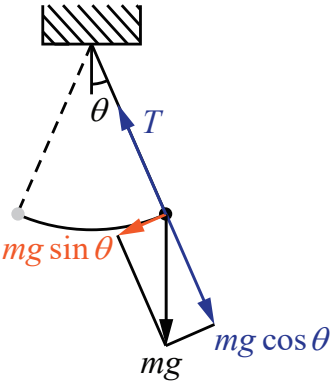


图 1: 单摆

令质点向右运动时速度为正，角速度和速度的关系为 $\dot\theta = v/L$，对其两边求导得角加速度和加速度的关系

$$\ddot\theta = a_\theta/L \tag{2}$$

其中 a_θ 是质点延垂直绳方向的加速度．现在沿垂直绳方向运用牛顿第二定律，并代入上式中的 a_θ 得

$$-mg\sin\theta = ma_\theta = -m\ddot\theta L \tag{3}$$

两边消去质量可得摆角 θ 关于时间的二阶微分方程.

$$L\ddot{\theta} = -g\sin\theta \tag{4}$$

解出该方程即可得到单摆做任意幅度摆动的规律. 虽然我们还不知道方程的解，但观察方程可知单摆的运动规律只与摆长 L 和重力加速度 g 有关，而与质点的质量无关. 所以改变同一单摆的质量不会改变它的运动规律.

小幅度摆动

预备知识 小角正弦值极限[152]

遗憾的是，式 4 的解并不能用有限个基本初等函数表示，而是需要使用椭圆积分[1061]. 我们先来考虑一种简单的情况，即单摆进行小的幅度摆动. 当 $\theta \to 0$ 时，可以把式 4 中的 $\sin\theta$ 近似为 θ. 令质点从最低点到当前位置之间的弧长为 s，则有 $s = \theta L$ 和 $\ddot{s} = a_\theta = \ddot{\theta}L$，式 4 变为 s 的微分方程

$$\ddot{s} = -\frac{g}{L}s \tag{5}$$

观察该式可以发现其结构与简谐振子的微分方程（式 1 [913]）非常相似. 用同样的方法，可得通解为

$$s = A\cos(\omega t + \varphi_0) \qquad \left(\omega = \sqrt{g/L}\right) \tag{6}$$

考虑到 $\theta \to 0$ 的情况下，圆弧可以近似为线段，所以可以认为此时质点在做一维简谐运动，用坐标 x 代替弧长 s.

圆锥摆

- 本词条处于草稿阶段.

预备知识 圆周运动的向心力[900]

质量不计的细线一端固定，另一端连接一个质点质量为 m. 细线与竖直方向夹角为 θ，质点做圆周运动.

未完成：图

我们令 $\alpha = \pi/2 - \theta$

$$T\sin\alpha = mg \tag{1}$$

$$T\cos\alpha = m\omega^2 R\cos\alpha \tag{2}$$

解得圆周运动的角速度为

$$\omega = \sqrt{\frac{g}{R\cos\theta}} = \sqrt{\frac{T}{mR}} \tag{3}$$

可见张角 θ，张力 T 和角速度 ω 之间只有一个自由度[930].

傅科摆

预备知识 单摆[917]

傅科摆是首个直接证明地球自转的实验. 试想如果把一个不受任何阻力的单摆放在地球的北极，那么地球每自转一定角度，单摆的摆平面不变，所以以地球为参考系观察，摆平面将反方向转动，这样就能证明地球在自转. 现实中，为了能克服阻力和微扰长时间摆动，通常使用质量较大，摆臂较长的摆作为傅科摆.

但若傅科摆被放在北纬 α 角处，摆平面的将会以怎样的角速度转动呢？事实证明，若令地球自转的角速度为 ω_0，则单摆相对地面转动的角速度 ω 将等于

$$\omega = \omega_0\sin\alpha \tag{1}$$

傅科摆角速度的一种几何推导

预备知识 矢量叉乘[387]

设 $\mathbf{R}$ 为地心指向傅科摆的矢量，当地纬度为 α，地轴指向北的单位矢量为 $\hat{\mathbf{k}}$，显然 $\hat{\mathbf{k}}\cdot\hat{\mathbf{R}} = \sin\alpha$.

若把任意矢量 $\mathbf{P}$ 围绕某单位矢量 $\hat{\mathbf{M}}$ 以右手定则旋转角微元 $\mathrm{d}\theta$，有

$$\mathrm{d}\mathbf{P} = \hat{\mathbf{M}} \times \mathbf{P}\,\mathrm{d}\theta \tag{2}$$

开始时，令傅科摆在最低点的速度方向的单位向量为 $\hat{\mathbf{A}}$（$\hat{\mathbf{A}}\cdot\mathbf{R}=0$），在傅科摆下方的水平地面上标记单位向量 $\hat{\mathbf{B}}$，使开始时 $\hat{\mathbf{B}}=\hat{\mathbf{A}}$．当傅科摆随地球在准静止状态下移动位移 $\mathrm{d}\mathbf{s}$（$\mathrm{d}\mathbf{s}\cdot\mathbf{R}=0$）后，由式 2 可得

$$\mathrm{d}\hat{\mathbf{A}} = \hat{\mathbf{M}} \times \hat{\mathbf{A}}\cdot\mathrm{d}\theta = \frac{\mathbf{R}\times\mathrm{d}\mathbf{s}}{|\mathbf{R}\times\mathrm{d}\mathbf{s}|}\times\hat{\mathbf{A}}\frac{\mathrm{d}s}{R} \tag{3}$$

注意这只是一个比较符合物理直觉的假设，这里并不给出证明．当地球转动 $\mathrm{d}\theta$ 时，上式中 $\mathrm{d}\mathbf{s}=\hat{\mathbf{k}}\times\mathbf{R}\,\mathrm{d}\theta$，而地面上的标记 $\hat{\mathbf{B}}$ 也围绕地轴转动，所以 $\mathrm{d}\hat{\mathbf{B}}=\hat{\mathbf{k}}\times\hat{\mathbf{A}}\,\mathrm{d}\theta$．

下面计算 $\mathrm{d}\hat{\mathbf{A}}-\mathrm{d}\hat{\mathbf{B}}$．因为 $\mathbf{R}\cdot\mathrm{d}\mathbf{s}=0$，所以 $|\mathbf{R}\times\mathrm{d}\mathbf{s}|=R\,\mathrm{d}s$，所以

$$\begin{aligned}\mathrm{d}\hat{\mathbf{A}} &= \frac{\mathbf{R}\times\mathrm{d}\mathbf{s}}{R^2}\times\hat{\mathbf{A}} = \frac{1}{R^2}\mathbf{R}\times(\hat{\mathbf{k}}\times\mathbf{R}\,\mathrm{d}\theta)\times\hat{\mathbf{A}}\\ &= \hat{\mathbf{R}}\times(\hat{\mathbf{k}}\times\hat{\mathbf{R}})\times\hat{\mathbf{A}}\,\mathrm{d}\theta = [(\hat{\mathbf{R}}\cdot\hat{\mathbf{R}})\hat{\mathbf{k}}-(\hat{\mathbf{R}}\cdot\hat{\mathbf{k}})\hat{\mathbf{R}}]\times\hat{\mathbf{A}}\,\mathrm{d}\theta\\ &= (\hat{\mathbf{k}}-\hat{\mathbf{R}}\sin\alpha)\times\hat{\mathbf{A}}\,\mathrm{d}\theta\end{aligned} \tag{4}$$

$$\mathrm{d}\hat{\mathbf{A}}-\mathrm{d}\hat{\mathbf{B}} = (\hat{\mathbf{k}}-\hat{\mathbf{R}}\sin\alpha)\times\hat{\mathbf{A}}\,\mathrm{d}\theta-\hat{\mathbf{k}}\times\hat{\mathbf{A}}\,\mathrm{d}\theta = -\sin\alpha\hat{\mathbf{R}}\times\hat{\mathbf{A}}\,\mathrm{d}\theta \tag{5}$$

所以地球转过 $\mathrm{d}\theta$ 角以后，$\hat{\mathbf{A}}$ 与 $\hat{\mathbf{B}}$ 之间的夹角为

$$\mathrm{d}\gamma = \left|\mathrm{d}\hat{\mathbf{A}}-\mathrm{d}\hat{\mathbf{B}}\right| = \sin\alpha\,\mathrm{d}\theta \tag{6}$$

两边除以 $\mathrm{d}t$ 得角速度

$$\omega = \omega_0\sin\alpha \tag{7}$$

惯性力

预备知识 牛顿第二定律[899]，加速度的坐标系变换[894]

若一个质点 m 某时刻在惯性系 S 中的加速度为 $\mathbf{a}_S$，在另一个非惯性系 S' 中的加速度为 $\mathbf{a}_{S'}$，则可假设质点受到一个**惯性力**

$$\mathbf{f} = m(\mathbf{a}_{S'}-\mathbf{a}_S) \tag{1}$$

使得牛顿运动定律在非惯性系 S' 中仍然成立．惯性力作为一个数学工具，既没有施力物体，也不是真实的力．另外，惯性力取决于参考系的选取甚至质点的运动，真实的受力不依赖于参考系．

例 1　加速的电梯

在向上加速的电梯中，电梯给人的支持力大于人的重力．这在地面参考系（惯性系）中的解释是，电梯给人的支持力除了要抵消人的重力，还要提供额外的向上的力使人产生向上的加速度．但是人从直觉上认为自己所处的是惯性系，符合牛顿运动定律．那么唯一合理的解释就是自己被施加了额外的向下的力．为了使自己保持“静止”或“受力平衡”，电梯会给人一个额外的支持力．

例 2　转弯的车

车向左转时人感觉到向右的“离心力”．同样，这一现象在地面参考系中解释，是车为了使人具有向左的向心加速度，给人一个向左的力，同时人对车施加一个向右的反作用力．然而，车中的人直觉上认为自己所处的是惯性系，符合牛顿运动定律，那么唯一合理的解释就是自己受到了向右的“离心力”．为了使人保持“静止”或“受力平衡”，车必须给人一个向左的反作用力．

从这两个例子可以看出，如果要使牛顿运动定律在非惯性系中也成立，则需要假设一些力的存在，即惯性力．人的直觉总会假设自己的参考系是惯性系，这就解释了为什么日常生活中我们常说离心力却不说向心力．用离心力来描述现象并没有错，这只是从更符合直觉的非惯性系的角度来分析而已．

平动非惯性系

假设某个非惯性系 S' 相对于惯性系 S 没有旋转只有平移，且 t 时刻的相对加速度为 $\mathbf{a}(t)$．这样，S' 中的任何一个静止点相对于 S 系的加速度都是 $\mathbf{a}(t)$．设 S' 系中有一质点 m，相对于 S' 系的加速度为 $\mathbf{a}_{S'}(t)$，那么在惯性系中质点的加速度为

$$\mathbf{a}_S = \mathbf{a}_{S'} + \mathbf{a} \tag{2}$$

即**绝对加速度**[7]等于相对加速度加牵连加速度．运用牛顿第二定律得质点真实的受力为（注意真实受力不随参考系变化）

$$\mathbf{F} = m\mathbf{a}_S = m(\mathbf{a} + \mathbf{a}_{S'}) \tag{3}$$

非惯性系中的观察者假设自己的参考系中牛顿定律仍然成立，并假设存在惯性力 $\mathbf{f}$，对于某个质点有

$$\mathbf{F} + \mathbf{f} = m\mathbf{a}_{S'} \tag{4}$$

所以

$$\mathbf{f} = m\mathbf{a}_{S'} - \mathbf{F} = -m\mathbf{a} \tag{5}$$

这说明，平动非惯性系中任何一个物体受到的惯性力大小与质量和非惯性系的加速度的乘积成正比，方向与相对加速度方向相反．

在例 1 中，电梯的参考系就是一个平动的非惯性系．如果人站在一个弹簧秤上，秤的示数将等于人的体重加上惯性力 ma（电梯向上加速时 a 取正值）．这用惯性力来解释，就是静止的人受到与电梯加速度方向相反的惯性力，大小等于 ma（注意在电梯参考系中人没有加速度，所以是“受力平衡”的）．

非平动参考系

非平动参考系相对于惯性系除了平移运动还可能做旋转运动．对于非平动参考系，式 2 并不满足（即使把 $\mathbf{a}$ 看做位置和时间的函数），以后在“科里奥利力[925]”中我们会推导 $\mathbf{a}_S$ 和 $\mathbf{a}_{S'}$ 间的具体关系．所以我们不能得到式 5．由式 3 和式 4，对任何非惯性系都成立的结论只有

$$\mathbf{f} = m\mathbf{a}_{S'} - \mathbf{F} = m(\mathbf{a}_{S'} - \mathbf{a}_S) \tag{6}$$

所以在最一般的情况下，质点所受惯性力等于质量乘以两参考系中加速度之差．

滑块和运动斜面问题

[7]这里把绝对加速度定义为任意惯性系中的测得质点的加速度．由于不同惯性系之间的相对加速度为零，绝对加速度与惯性系的选择无关．

- 本词条处于草稿阶段.

预备知识 人船模型

在滑块斜面问题的基础上，如果我们假设斜面质量为 M, 与地面无摩擦，那么滑块相对斜面的加速是多少呢？

令 x, y 为滑块水平方向和竖直方向移动的距离．X 为斜面水平方向移动的距离.

$$x = \frac{M}{M+m} l\cos\theta \qquad y = -l\sin\theta \qquad X = -\frac{m}{M+m} l\cos\theta \tag{1}$$

以下介绍三种方法，都可以解得滑块相对斜面的加速度为

$$a = \ddot{l} = \frac{g\sin\theta(M+m)}{M + m\sin^2\theta} \tag{2}$$

未完成：图

受力分析法

未完成：未完成

非惯性系法

在斜面的参考系，滑块会受到向右的惯性力 $-m\ddot{X}$，所以沿斜面向下使用牛顿第二定律得

$$-m\ddot{X}\cos\theta + mg\sin\theta = m\ddot{l} \tag{3}$$

解得式 2.

拉格朗日方程法

考虑动量守恒，这个系统只有一个自由度．我们把滑块相对斜面向下移动的距离记为 l．为了描述系统动能和势能，我们另外引入几个可以用 l 表示的

变量：则拉格朗日量为

$$\begin{aligned} L = T - V &= \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) + \frac{1}{2}M\dot{X}^2 - mgy \\ &= \frac{1}{2}m\left(1 - \frac{m\cos^2\theta}{M+m}\right)\dot{l}^2 + mg\sin\theta \cdot l \end{aligned} \tag{4}$$

代入拉格朗日方程

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{l}} = \frac{\partial L}{\partial l} \tag{5}$$

解得式 2.

离心力

预备知识 匀速圆周运动的加速度[885]，惯性力[920]，连续叉乘的化简[392]

令参考系 abc 和 xyz 的 c 轴和 z 轴始终重合．其中 xyz 是惯性系，abc 以恒定的角速度 ω 绕 z 轴逆时针转动．求 abc 系中一个质量为 m 的静止质点所受的惯性力，即**离心力（centrifugal force）**.

令质点的坐标 (a,b,c) 离 c 轴的距离为 $r_\perp = \sqrt{a^2+b^2}$，对应的径向矢量为 $\mathbf{r}_\perp = (a,b,0)^{\mathrm{T}}$．在 xyz 系中，质点做匀速圆周运动，相对于 xyz 系的加速度（用 abc 系的坐标表示）为

$$\mathbf{a}_{xyz} = -\omega^2\mathbf{r}_\perp = -\omega^2 \begin{pmatrix} a \\ b \\ 0 \end{pmatrix}_{abc} \tag{1}$$

质点相对于 abc 系静止，相对加速度为零

$$\mathbf{a}_{abc} = \mathbf{0} \tag{2}$$

所以由惯性力[920] 中的结论，惯性力为

$$\mathbf{f} = m(\mathbf{a}_{abc} - \mathbf{a}_{xyz}) = \omega^2\mathbf{r}_\perp = m\omega^2 \begin{pmatrix} a \\ b \\ 0 \end{pmatrix}_{abc} \tag{3}$$

注意离心力向外，与直觉相符．注意这个结论只适用于质点相对于 abc 系静止的情况，若有相对运动，则惯性力除了离心力，还会有一项科里奥利力[925]．

若转轴取任意方向 $\hat{\boldsymbol{\omega}}$，由式 8 [885] 得

$$\mathbf{a}_{xyz} = \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) \tag{4}$$

$$\mathbf{f} = -m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) \tag{5}$$

科里奥利力

预备知识 离心力[924]，平面旋转矩阵[402]

科里奥利力（Coriolis Force）是匀速旋转的参考系中由质点运动产生的惯性力．

$$\mathbf{F}_c = 2m\mathbf{v}_{S'} \times \boldsymbol{\omega} \tag{1}$$

其中 $\mathbf{v}_{S'}$ 是质点相对于旋转参考系 S' 的瞬时速度，$\boldsymbol{\omega}$ 是旋转系相对于某惯性系 S 转动的角速度矢量[883]．式中的乘法是叉乘[387]．在匀速转动参考系（属于非惯性系）中，若质点保持相对静止，则惯性力只有离心力．然而当质点与转动参考系有相对速度时，惯性力中还会增加一个与速度垂直的力，这就是科里奥利力．地理中的地转偏向力就是科里奥利力，可用上式计算（见“地球表面的科里奥利力[927]”）．

由叉乘的定义可得，科里奥利力与速度矢量始终保持垂直，所以科里奥利力不会对质点做功．

推导

预备知识 连续叉乘的化简[392]，圆周运动的速度[883]，加速度的坐标变换[894]

我们可以直接根据惯性力的定义（式 1 [920]）和加速度的坐标变换（式 3 [894]）得到任意非惯性系 S' 中质点的总惯性力（S 为任意惯性系）为

$$\mathbf{F}_c = m(\mathbf{a}_{S'} - \mathbf{a}_S) = -m\mathbf{a}_r + 2m\mathbf{v}_{S'} \times \boldsymbol{\omega} \tag{2}$$

其中第一项包含平移惯性力和转动惯性力，转动惯性力又可划分为离心力以及角加速度产生的惯性力（见式 8），但与质点相对 S' 的速度无关，所以只将科里奥利力定义为第二项．

另一种推导

类比式 1 [892]，若 S' 系与 S 系原点始终重合，且相对 S 系以角速度 $\boldsymbol{\omega}$ 旋转，对任意一个随时间变化的矢量（假设一阶导数存在），我们把它在 S 和 S' 系中的时间导数分别记为 $(\dot{\mathbf{A}})_S$ 和 $(\dot{\mathbf{A}})_{S'}$，则有

$$(\dot{\mathbf{A}})_S = (\dot{\mathbf{A}})_{S'} + \boldsymbol{\omega} \times \mathbf{A} \tag{3}$$

最有一项参考式 5 [883]. 注意该式中的矢量为几何矢量[370] 而不是列矢量，若要将该式记为坐标形式，应该使用同一坐标系[892].

我们先将 $\mathbf{A}$ 替换为质点的位矢 $\mathbf{r}$，得参考系中质点的速度关系为（即式 1 [892]）

$$\mathbf{v}_S = \mathbf{v}_{S'} + \boldsymbol{\omega} \times \mathbf{r} \tag{4}$$

两边在 S 系中对时间求导得

$$\mathbf{a}_S = (\dot{\mathbf{v}}_{S'})_S + \boldsymbol{\omega} \times \mathbf{v}_S + \dot{\boldsymbol{\omega}} \times \mathbf{r} \tag{5}$$

注意 S' 系中的加速度 $\mathbf{a}_{S'}$ 并不是 $(\dot{\mathbf{v}}_{S'})_S$，而是 $(\dot{\mathbf{v}}_{S'})_{S'}$. 令式 3 中的 $\mathbf{A} = \mathbf{v}_{S'}$，得

$$(\dot{\mathbf{v}_{S'}})_S = \mathbf{a}_{S'} + \boldsymbol{\omega} \times \mathbf{v}_{S'} \tag{6}$$

将式 4 和式 6 代入式 5，得

$$\mathbf{a}_S = \mathbf{a}_{S'} + 2\boldsymbol{\omega} \times \mathbf{v}_{S'} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) + \dot{\boldsymbol{\omega}} \times \mathbf{r} \tag{7}$$

所以旋转参考系中的总惯性力（式 1 [920]）为

$$\mathbf{f} = m(\mathbf{a}_{S'} - \mathbf{a}_S) = 2m\mathbf{v}_{S'} \times \boldsymbol{\omega} - m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) - m\dot{\boldsymbol{\omega}} \times \mathbf{r} \tag{8}$$

其中第一项被称为科里奥利力（唯一一项与 $\mathbf{v}_{S'}$ 有关的），第二项为离心力（式 5 [924]），第三项为角加速度产生的惯性力.

旋转参考系的“机械能守恒”

预备知识 科里奥利力[925]，机械能守恒（单个质点）[909]

在匀速旋转参考系中，如果加入惯性力的修正，则牛顿定律同样适用．那么我们是否同样有某种“修正版”的单质点机械能守恒呢？答案是肯定的，在旋转参考系中，我们仍然定义每个质点的动能为 $E_k = mv^2/2$（v 是相对旋转参考系的速度），假设质点所受的所有非惯性力在旋转参考系中都是保守力[905]，对应的势能函数 $V(\mathbf{r})$．那么机械能修正后的守恒量为

$$\frac{1}{2}mv^2 + V(\mathbf{r}) - \frac{1}{2}m\omega^2 r^2 = \text{const} \tag{1}$$

其中 ω 是旋转参考系相对于惯性系的角速度．

对于质点系，如果质点两两之间的力也是保守力，例如库仑力或者万有引力，那么我们只需要对每个质点计算上式并求和即可得到系统的守恒量．

例子：雅可比常量[1093]．

推导

证明的思路很简单，在“机械能守恒（单个质点）[909]”中证明的基础上，我们只需要额外考虑惯性力的做功即可．

由于科里奥利力始终与每个质点的运动方向垂直，所以对质点不做功，而离心力却有可能会做功，所以式 1 中的前两项不是守恒量．每个质点所受的离心力只与位置有关，于是我们可以得到一个离心力场

$$F_c(\mathbf{r}) = m\omega^2\mathbf{r} \tag{2}$$

这是一个中心力场，所以必定是一个保守场[1034]，沿径向积分容易得到对应的等效势能函数为

$$V_c(\mathbf{r}) = -\frac{1}{2}m\omega^2 r^2 \tag{3}$$

现在，离心力做功等于 $V_c(\mathbf{r})$ 的减小，所以把这个等效势能加入原来的机械能，就又得到了一个守恒量．

地球表面的科里奥利力

预备知识 科里奥利力[925]

令延地轴向北的单位矢量为 $\hat{\mathbf{z}}$，质点所在经线与赤道交点的单位矢量为 $\hat{\mathbf{x}}$，则 $\hat{\mathbf{y}}=\hat{\mathbf{z}}\times\hat{\mathbf{x}}$. 若质点运动的方向与所在经线的夹角为 ϕ（顺时针为正），运动平面的法向量与赤道平面的夹角为 θ（若把地球近似看做球形，则 θ 是质点所在纬度[8]）. 这样，运动平面内正北方向的单位矢量为 $\hat{\mathbf{z}}'=-\sin\theta\,\hat{\mathbf{x}}+\cos\theta\,\hat{\mathbf{z}}$，正东方向的单位矢量为 $\hat{\mathbf{y}}$，正上方为 $\hat{\mathbf{x}}'=\hat{\mathbf{y}}\times\hat{\mathbf{z}}'=\sin\theta\,\hat{\mathbf{z}}+\cos\theta\,\hat{\mathbf{x}}$，速度方向的单位矢量为

$$\hat{\mathbf{v}}=\sin\phi\,\hat{\mathbf{y}}+\cos\phi\,\hat{\mathbf{z}}'=-\cos\phi\sin\theta\,\hat{\mathbf{x}}+\sin\phi\,\hat{\mathbf{y}}+\cos\phi\cos\theta\,\hat{\mathbf{z}}\tag{1}$$

现在可以计算科里奥利力

$$\mathbf{F}_{col}=2mv\omega\,\hat{\mathbf{v}}\times\hat{\mathbf{z}}=2mv\omega(\cos\phi\sin\theta\,\hat{\mathbf{y}}+\sin\phi\,\hat{\mathbf{x}})\tag{2}$$

其向北，向东，向上的分量分别为

$$\mathbf{F}_{col}\cdot\hat{\mathbf{z}}'=-2mv\omega\sin\theta\sin\phi\qquad\mathbf{F}_{col}\cdot\hat{\mathbf{y}}=2mv\omega\sin\theta\cos\phi\tag{3}$$

$$\mathbf{F}_{col}\cdot\hat{\mathbf{x}}'=2mv\omega\sin\phi\cos\theta\tag{4}$$

可以证明水平分力可以表示为

$$\mathbf{F}_{col}^{\parallel}=(\mathbf{F}_{col}\cdot\hat{\mathbf{y}})\hat{\mathbf{y}}+(\mathbf{F}_{col}\cdot\hat{\mathbf{z}}')\hat{\mathbf{z}}'=2m\mathbf{v}\times\boldsymbol{\omega}'\tag{5}$$

其中 $\boldsymbol{\omega}'=\omega\sin\theta\,\hat{\mathbf{x}}'$. 可见科氏力的水平分量始终与速度垂直，且在地球的两极（$\theta=\pi/2$）处取最大值 $2mv\omega$，在赤道处为 0.

要特别注意的是，地球表面的非惯性力除了科里奥利力外还有离心力，但离心力一般被地球的椭球形弥补，可以不计.

例 1

假设 30 吨重的高铁车厢在北纬 30 度以 300 km/h 的速度行驶，其水平方向的科氏力大小为

$$\begin{aligned}F_{col}&=2\times30,000\text{ kg}\times\frac{300,000\text{ m}}{3600\text{ s}}\times\frac{2\pi}{24\text{ h}\times3600\text{ s}}\times\sin\frac{\pi}{6}\\&=60.32\text{ N}\end{aligned}\tag{6}$$

[8]但严格来说地球由于受离心力，赤道宽，两极窄.

第三章

质点系与刚体

自由度

预备知识 柱坐标系[121]，球坐标系[123]

我们已知描述一个质点在三维空间中的位置需要至少三个变量，无论使用直角坐标，柱坐标或球坐标．我们说这个质点有三个**自由度（degrees of freedom）**．同理，若空间中有 N 个质点可以自由移动，我们就说这些质点组成的系统有 $3N$ 个自由度．

需要注意的是，以上讨论的粒子都是自由的，即没有**约束**．若给系统施加约束，自由度就会相应减少．例如，若用长度为 R 的细棒将一个质点和坐标原点相连，则质点只能够在半径为 R 的球面上运动，这时我们在直角坐标系中施加了约束条件

$$x^2 + y^2 + z^2 = R^2 \tag{1}$$

当我们确定了 x, y, z 中的任意两个变量后，第三个变量就可以根据上式确定，所以这时质点就只有两个自由度．我们把选取的两个自由坐标叫做**独立（independent）变量**，剩下的一个坐标就是**非独立**的[1]．另一种更直观的方法是把这个质点的位置用球坐标 (r, θ, ϕ) 表示，约束条件变为

$$r = R \tag{2}$$

我们马上就得到两个独立的变量 (θ, ϕ) 用于完全确定质点的位置．我们把式 1 或式 2 这样的等式称为一个约束条件，若给自由度为 N 的系统加上 M 个约束条件，则系统的自由度减少为 $N - M$．

另一个例子是考虑两个质点，它们之间由长度为 R 细棒相连，那么原来的 6 个自由度的系统在一个约束条件 $(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 = R^2$ 下变为 5 个自由度．要具体选取五个独立变量，我们可以用三个独立变量先确定第一个质点的位置，然后再用剩下两个独立变量加上约束条件确定第二个质点的位置．我们还可以再继续加约束条件，例如限制第一个质点只能在 z 轴上运动，这事实上是两个约束条件 $x_1 = 0$ 和 $y_1 = 0$，使两个质点的自由度最终减少为 3 个．我们可以取一个独立变量描述第一个质点在 z 轴上的坐标，另外两个独立变量描述第二个质点相对于第一个质点的方向．

[1]注意这个例子中第三个变量可能会有正负两个解，但我们仍然认为它是非独立的．

质点系

预备知识 牛顿第三定律[899]

在考虑多个物体构成的系统时，我们有时候可以把每个物体都近似为一个质点，这样我们就得到了由有限个质点构成的系统，简称为**质点系**或者**质点组**.

令质点系中有 N 个质点，每个质点的受力都可以分为两类，一是系统外界物体给该质点的力，称为**外力**，二是来自系统内其他质点的力，称为**内力**. 对第 i 个质点，以下将它受到的所有外力和内力之和分别记为 $\mathbf{F}_i^{out}$ 和 $\mathbf{F}_i^{in}$，即单个质点的**合内力**以及**合外力**，所以单个质点所受的合力为

$$\mathbf{F}_i = \mathbf{F}_i^{in} + \mathbf{F}_i^{out} \tag{1}$$

系统中所有质点所受的**合力**等于**合内力**加**合外力**[2]

$$\mathbf{F}_{tot} = \sum_i \mathbf{F}_i = \sum_i^N \mathbf{F}_i^{in} + \sum_i^N \mathbf{F}_i^{out} = \mathbf{F}_{tot}^{in} + \mathbf{F}_{tot}^{out} \tag{2}$$

若将第 j 个质点对第 i 个质点的内力记为 $\mathbf{F}_{j\to i}$ 则上式中

$$\sum_i \mathbf{F}_i^{in} = \sum_{i,j}^{i\neq j} \mathbf{F}_{j\to i} \tag{3}$$

任意两个质点 k 和 l 对该求和的贡献是一对相互作用力 $\mathbf{F}_{k\to l} + \mathbf{F}_{l\to k}$，而根据牛顿第三定律[899]，相互作用力之和为零. 所以上式求和为零. 所以，质点系中**合内力为零**，系统所受合力等于合外力

$$\mathbf{F}_{tot} = \mathbf{F}_{tot}^{out} = \sum_i^N \mathbf{F}_i^{out} \tag{4}$$

质心 质心系

[2]角标“tot”表示 total.

预备知识　体积分[283]，质点系[931]

质心的定义

质心（center of mass）通俗来讲可以理解为质量的中心，是系统中各个质点的位置矢量[878]关于质量的加权平均值．我们先看几个例子．

两个等质量质点的质心

对于两个质量相等的质点，它们的质心显然在它们连线的中点处，无论它们的质量是多少．如果它们都在 x 轴上，则质心的位置就是两质点 x 坐标的中点

$$x_c = (x_1 + x_2)/2 \tag{1}$$

其中角标 c 表示 center of mass，有时候也会写做 CM．

在二维和三维空间的情况下，质点的位置用位置矢量 $\mathbf{r}$ 描述，将它们的位置矢量分别记为 $\mathbf{r}_1$ 和 $\mathbf{r}_2$，则质心的位置为

$$\mathbf{r}_c = (\mathbf{r}_1 + \mathbf{r}_2)/2 \tag{2}$$

即两个位置矢量的平均值．

两个不同质量质点的质心

当两个质点质量不一样时（分别记为 m_1 和 m_2），质心会更靠近更重的质点．如果它们都在 x 轴上，我们就用加权平均值

$$x_c = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2} \tag{3}$$

注意 $m_1 = m_2$ 就得到了式 1．另一个极端是当一个质量远大于另一个，如 $m_1 \gg m_2$，这时质心就趋近于 x_1 了．

二维和三维空间的情况下也类似有

$$\mathbf{r}_c = \frac{m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2}{m_1 + m_2} \tag{4}$$

当 $m_1 = m_2$ 就得到式 2．

习题 1

证明两质点的质心必定在其连线上，即 $\mathbf{r}_1-\mathbf{r}_c$ 和 $\mathbf{r}_2-\mathbf{r}_c$ 共线.

习题 2

试证明式 4 中质心到两质点的距离与它们的质量成反比，即

$$\frac{|\mathbf{r}_1-\mathbf{r}_c|}{|\mathbf{r}_2-\mathbf{r}_c|}=\frac{m_2}{m_1} \tag{5}$$

质点系的质心

若质点系中有 N 个质点，令第 i 个质点质量为 m_i，位置为 $\mathbf{r}_i$，总质量为 $M=\sum_i m_i$，则该质点系的**质心**定义为

$$\mathbf{r}_c=\frac{1}{M}\sum_i m_i\mathbf{r}_i \tag{6}$$

在直角坐标系中，我们可以将上式的矢量求和分解为对 x,y,z 方向的分量分别求和（因为矢量相加等于两个分量分别相加）. 令 $\mathbf{r}_i=x_i\hat{\mathbf{x}}+y_i\hat{\mathbf{y}}+z_i\hat{\mathbf{z}}$，即矢量 $\mathbf{r}_i$ 的坐标为 (x_i,y_i,z_i)，有

$$x_c=\frac{1}{M}\sum_i m_ix_i \qquad y_c=\frac{1}{M}\sum_i m_iy_i \qquad z_c=\frac{1}{M}\sum_i m_iz_i \tag{7}$$

例 1

空间直角坐标系中四个质点质量分别为 1 kg，2 kg，3 kg，4 kg，坐标分别为 $(0,0,0)$，$(1,0,0)$，$(0,2,0)$，$(0,0,3)$（单位：米）. 求该系统质心的位置.

解：系统总质量为 10 kg，直接使用式 7 得

$$x_c=\frac{1}{10\text{ kg}}(0\text{ m}\times1\text{ kg}+1\text{ m}\times2\text{ kg}+0\text{ m}\times3\text{ kg}+0\text{ m}\times4\text{ kg})=\frac{1}{5}\text{ m} \tag{8}$$

$$y_c=\frac{1}{10\text{ kg}}(0\text{ m}\times1\text{ kg}+0\text{ m}\times2\text{ kg}+2\text{ m}\times3\text{ kg}+0\text{ m}\times4\text{ kg})=\frac{3}{5}\text{ m} \tag{9}$$

$$z_c=\frac{1}{10\text{ kg}}(0\text{ m}\times1\text{ kg}+0\text{ m}\times2\text{ kg}+0\text{ m}\times3\text{ kg}+3\text{ m}\times4\text{ kg})=\frac{6}{5}\text{ m} \tag{10}$$

所以质心的坐标为 $(1/5,3/5,6/5)$（单位：米）.

质心的分解

若我们把质点系划分为若干组，可以先计算每组的质心，再计算“质心的质心”就可以得到系统的总质心．我们举例说明

例 2

令四个质点中的前两个为 a 组，后两个为 b 组，则它们的质心分别为

$$\mathbf{r}_a = (m_1\mathbf{r}_1 + m_2\mathbf{r}_2)/M_a \qquad \mathbf{r}_b = (m_3\mathbf{r}_3 + m_4\mathbf{r}_4)/M_b \tag{11}$$

其中 $M_a = m_1 + m_2$，$M_b = m_3 + m_4$．再计算“质心的质心”得整个系统的质心为

$$\mathbf{r}_c = \frac{M_a\mathbf{r}_a + M_b\mathbf{r}_b}{M_a + M_b} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + m_3\mathbf{r}_3 + m_4\mathbf{r}_4}{m_1 + m_2 + m_3 + m_4} \tag{12}$$

这个结果符合式 6.

习题 3

边长为 $2L$ 的大正方形和边长为 L 的小正方形拼接在一起，假设密度均匀，求拼接后物体的质心（提示：可以分别求出两正方形的质心在求“质心的质心”）.

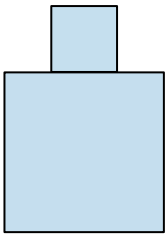


图 1: 求质心

习题 4

如图 2 所示，一个半径为 R 的圆盘中间挖去一个半径为 r 的圆盘，两圆心距离为 d，假设密度均匀，求该物体质心（提示：我们可以把该形状看成一个完整的圆盘和一个具有“负质量”的小圆盘叠加而成的，分别计算二者的质心再计算“质心的质心”）.

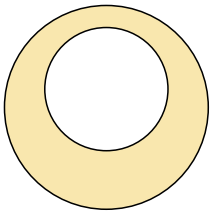


图 2: 求质心

质心与重心

质心在物理中有什么用呢？一个基本的应用就是质心就是物体的**重心**．我们只讨论恒定重力场中物体的重心．其定义是：若重力场对物体关于某点的合力矩[947] 恒为 0，这个点就是它的**重心**．“合力矩为零”意味着，如果物体初始时以任意姿态静止，那么它将一直保持静止．虽然我们还没系统学习力矩，但可以用初中学过的“力乘力臂”进行计算（式 1 [947]）．

例 3

轻杆[3]两端有质量分别为 m_1 和 m_2 的小球，轻杆可以绕系统质心在竖直平面上自由转动．试证明重力对系统的力矩恒为 0．

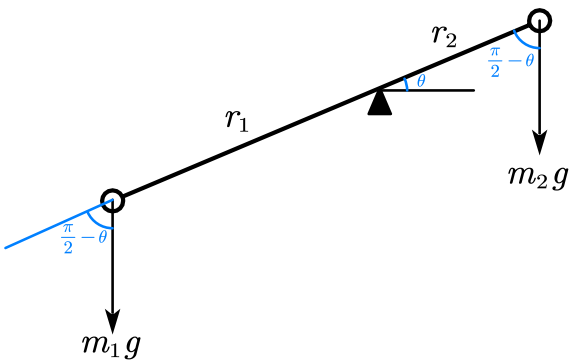


图 3: 轻杆与两小球

解：以逆时针为正，合力矩为

$$M = r_1 m_1 g \cos\theta - r_2 m_2 g \cos\theta = (r_1 m_1 - r_2 m_2) g \cos\theta \tag{13}$$

由式 5，括号中两项相等，所以无论 θ 去何值，合力矩都为 0．

[3]轻杆是指质量可忽略不计的杆

我们把一般性的证明留到“重心[??]”中.

连续质量分布

对连续质量分布，令密度关于位置的函数为 $\rho(\mathbf{r})$，总质量为密度的体积分

$$M = \int \rho(\mathbf{r})\,\mathrm{d}V \tag{14}$$

要计算质心，我们可以把整个物体划分为许多小块（微元），如果每一块都很小，我们可以假设第 i 块的位置为 $\mathbf{r}_i$，密度为常数 $\rho(\mathbf{r}_i)$，体积为 ΔV_i，所以质量为 $\Delta m_i = \rho(\mathbf{r}_i)\Delta V_i$. 我们把每个小块都用 $\mathbf{r}_i$ 处的一个质量为 Δm_i 的质点来代替，那么质心为

$$\mathbf{r}_c = \frac{1}{M}\sum_i \Delta m_i \mathbf{r}_i = \frac{1}{M}\sum_i \mathbf{r}_i \rho(\mathbf{r}_i)\Delta V_i \tag{15}$$

当所有的微元的体积都趋近于零时，我们就可以将该式用体积分[283] 表示为

$$\mathbf{r}_c = \frac{1}{M}\int \mathbf{r}\rho(\mathbf{r})\,\mathrm{d}V \tag{16}$$

这个积分中的被积函数是矢量，结果也是矢量，该如何计算呢？答案就是像式 7 那样分别对矢量的每个分量积分，得到结果的每个分量（可见求和具有的性质，积分通常也有）.

$$\begin{cases} x_c = \dfrac{1}{M}\iiint x\sigma(\mathbf{r})\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \\ y_c = \dfrac{1}{M}\iiint y\sigma(\mathbf{r})\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \\ z_c = \dfrac{1}{M}\iiint z\sigma(\mathbf{r})\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \end{cases} \tag{17}$$

如果要计算的物体是一个厚度可以忽略不计的薄片，令 $\sigma(\mathbf{r})$ 为面密度（单位面积的质量），我们就可以用面积分代替体积分.

$$\mathbf{r}_c = \frac{1}{M}\int \mathbf{r}\sigma(\mathbf{r})\,\mathrm{d}S \tag{18}$$

即

$$\begin{cases} x_c = \dfrac{1}{M}\iint x\sigma(\mathbf{r})\,\mathrm{d}x\,\mathrm{d}y \\ y_c = \dfrac{1}{M}\iint y\sigma(\mathbf{r})\,\mathrm{d}x\,\mathrm{d}y \end{cases} \tag{19}$$

例 4　长方形的质心

在平面直角坐标系中，长方形均匀薄片的 4 个点分别为 $(0,0)$，$(a,0)$，(a,b)，$(b,0)$，面密度 σ 为常数，试计算其质心.

解：长方形的总质量为 $M=ab\sigma$. 使用式 18（分别对矢量的两个分量积分）得

$$x_c=\frac{1}{ab\sigma}\int_0^b\int_0^a x\sigma\,\mathrm{d}x\,\mathrm{d}y=\frac{1}{a}\int_0^a x\,\mathrm{d}x=\frac{a}{2} \tag{20}$$

$$y_c=\frac{1}{ab\sigma}\int_0^b\int_0^a y\sigma\,\mathrm{d}x\,\mathrm{d}y=\frac{1}{b}\int_0^b y\,\mathrm{d}y=\frac{b}{2} \tag{21}$$

可见质心的坐标为 $(a/2,b/2)$，恰好在长方形的中心.

我们再补充两个例子用于练习积分的运算

例 5　三角形的质心

在平面直角坐标系中，三角形均匀薄片的 3 个点分别为 $(-a,0)$，$(b,0)$，$(0,c)$. 试计算其质心.

解：令面密度 σ 为常数，则总质量为 $M=(a+b)c\sigma/2$. 两条斜边的直线方程分别为

$$x=f_1(y)=a(y-c)/c\qquad x=f_2(y)=b(c-y)/c \tag{22}$$

做面积分得（先积 x 再积 y）

$$\begin{aligned}x_c&=\frac{2}{(a+b)c\sigma}\int_0^c\int_{f_1(y)}^{f_2(y)}x\sigma\,\mathrm{d}x\,\mathrm{d}y\\&=\frac{1}{(a+b)c}\int_0^c[f_2^2(y)-f_1^2(y)]\,\mathrm{d}y=\frac{b-a}{3}\end{aligned} \tag{23}$$

$$\begin{aligned}y_c&=\frac{2}{(a+b)c\sigma}\int_0^c\int_{f_1(y)}^{f_2(y)}y\sigma\,\mathrm{d}x\,\mathrm{d}y\\&=\frac{2}{(a+b)c}\int_0^c[f_2(y)-f_1(y)]y\,\mathrm{d}y=\frac{c}{3}\end{aligned} \tag{24}$$

不难发现，这就是初中所学的三角形的重心，即底边中线的三等分点，或三条中线的交点.

由于质点系的积分和求和具有同样的性质，在以下的证明中，我们只需对质点系加以证明，结论对于连续质量分布的物体也同样适用.

质心的唯一性

既然质心的定义取决于参考系（因为 $\mathbf{r}_i$ 取决于参考系），那么不同参考系中计算出的质心是否是空间中的同一点呢？例如将例 4 中的长方形平移 Δs，质心是否也会平移 Δs？我们只需要证明，在 A 坐标系中得到的质心 $\mathbf{r}_{Ac}$ 与 B 坐标系中得到的质心 $\mathbf{r}_{Bc}$ 满足关系

$$\mathbf{r}_{Ac} = \mathbf{r}_{AB} + \mathbf{r}_{Bc} \tag{25}$$

其中 $\mathbf{r}_{AB}$ 是 A 系原点指向 B 系原点的矢量．首先根据定义

$$\mathbf{r}_{Ac} = \frac{1}{M}\sum_i m_i \mathbf{r}_{Ai} \qquad \mathbf{r}_{Bc} = \frac{1}{M}\sum_i m_i \mathbf{r}_{Bi} \tag{26}$$

由位矢的坐标系变换，$\mathbf{r}_{Ai} = \mathbf{r}_{AB} + \mathbf{r}_{Bi}$，所以

$$\mathbf{r}_{Ac} = \frac{1}{M}\sum_i m_i(\mathbf{r}_{AB} + \mathbf{r}_{Bi}) = \mathbf{r}_{AB} + \frac{1}{M}\sum_i m_i \mathbf{r}_{Bi} = \mathbf{r}_{AB} + \mathbf{r}_{Bc} \tag{27}$$

质心参考系

定义质点系的**质心参考系**（或**质心系**）为原点固定在质心上且没有转动的参考系（平动参考系）．根据质心的唯一性（式 25），在质心系中计算质心（式 6）仍然落在原点，即

$$\sum_i m_i \mathbf{r}_{ci} = \mathbf{0} \tag{28}$$

其中 $\mathbf{r}_{ci}$ 是质心系中质点 i 的位矢．

注意质心系并不一定是惯性系，只有当合外力为零质心做匀速直线运动时，质心系才是惯性系．在非惯性系中，每个质点受惯性力．

未完成：来源？链接？

质心系中总动量

把式 28 两边对时间求导，得

$$\sum_i m_i \mathbf{v}_{ci} = \mathbf{0} \tag{29}$$

注意到等式左边是质心系中质点系的总动量，所以我们得到质心系的一个重要特点，**质心系中总动量为零**．

质点系的动量

预备知识 动量 动量定理（单个质点）[910]，质心质心系[931]

质点系的总动量为

$$\mathbf{p}=\sum_i m_i\mathbf{v}_i=\sum_i m_i\dot{\mathbf{r}}_i=\frac{\mathrm{d}}{\mathrm{d}t}\sum_i m_i\mathbf{r}_i \tag{1}$$

由质心的定义（式 6 [931]）

$$\sum_i m_i\mathbf{r}_i=M\mathbf{r}_c \tag{2}$$

其中 $\mathbf{r}_c$ 为质心的位置，$M=\sum_i m_i$ 为质点系的总质量．两边对时间求导并代入式 1 得

$$\mathbf{p}=M\mathbf{v}_c \tag{3}$$

其中 $\mathbf{v}_c=\dot{\mathbf{r}}_c$ 是质心的速度．

式 3 告诉我们一个重要的结论：**在求一个系统的总动量时，我们可以把它等效为其质心处具有相同质量的质点．**

例 1　滚动的圆盘

一个质量为 M 圆盘在地面延直线滚动，圆心的速度为 $\mathbf{v}$．若将其分割为许多小份，使用 $\sum_i m_i\mathbf{v}_i$（或者用积分形式）求总动量会比较麻烦．但如果直接用式 3，我们可以马上写出它的总动量为 $\mathbf{p}=M\mathbf{v}$，甚至不需要知道它的半径和角速度，也不需要知道它和地面是否存在打滑．

通过式 3 也可以直观地得出：质心参考系中系统的总动量为零．

刚体

预备知识 质心[931]，自由度[930]

当我们要考虑一个物体的质量分布带来的力学效应时，就不能再将其简化为一个质点．许多情况下我们考虑的物体在某过程中形变较小可忽略不计，这

时我们就可以忽略它运动过程中的的任何形变，从而大大简化问题．我们把这种模型叫做**刚体**．在分析刚体时，我们通常把刚体看做是质点系．要这么做，我们可以把刚体划分为无限多个体积无限小的微元，再把每个微元近似为一个质量相同的质点即可．

刚体的自由度

在没有任何约束的情况下，三维空间中每个质点有 3 个**自由度**，即用三个完全独立的变量才能完全确定位置，所以 N 个质点组成的质点系共有 $3N$ 个自由度．然而完全确定一个刚体的位置只需要 6 个变量，这是因为刚体模型通过假设“任意两个质点之间距离不变”，给质点系的位置施加了 $3N-6$ 个约束条件．如何得出 6 个自由度呢？我们可以假设第一个质点有 3 个自由度，第二个质点由于要与第一个质点保持距离不变，只有 $3-1=2$ 个自由度，而第三个质点要与前两个质点保持距离不变，只有 $3-2=1$ 个自由度．有了前三个质点后（假设它们不共线），剩下所有质点的位置都可以由与这三个质点的距离确定，所以任何刚体都有 $3+2+1=6$ 个自由度．

我们也可以这么划分 6 个自由度：令其中 3 个决定刚体上某点的位置，2 个决定该点到刚体上另一点的矢量的方向（球坐标[123] 中的两个角度），最后一个决定刚体绕该矢量旋转的角度．

若刚体只能绕固定点旋转，那么刚体就只有 3 个转动自由度．若在此基础上，刚体只能绕固定轴旋转，那它就只剩一个自由度了（转过的角度）．

刚体的运动

如果已知刚体的受力（可以关于时间变化），如何计算刚体的运动呢？我们可以分别计算刚体所受的所有外力的矢量和，称为**合外力**，以及所有外力关于刚体质心质心[931] 的矢量和，称为**合外力矩**．然后通过质点系的动量定理[944]，角动量定理角动量定理[954] 以及刚体的刚体的转动惯量[962] 或者惯性张量[973] 就可以得到刚体在每个时刻的姿态．

如果刚体在旋转过程中转轴始终保持在同一个方向，那么计算过程将十分简单，因为不需要使用惯性张量，具体过程见“刚体的平面运动方程[970]”．但如果刚体转动时转轴可能发生改变，那么就必须使用惯性张量和参考系变换，使计算更为复杂，详见“刚体的运动方程[979]”．

轻杆模型

预备知识 速度 加速度[881]

轻杆在这里指是质量可以忽略不计，且不能伸缩的刚性杆．轻杆和质点一样都是一种理想化的模型．我们先来讨论轻杆的运动学特性，而把动力学放到后面．

不可伸长条件

速度约束

我们来思考“不可伸长”这一条件会对杆两端的速度和加速度带来什么约束．令杆的长度为 L，两端点的位置矢量分别为 $\mathbf{r}_1$ 和 $\mathbf{r}_2$，则杆的长度平方为 $L^2=(\mathbf{r}_2-\mathbf{r}_1)^2$．另外记端点的速度为 $\mathbf{v}_1,\mathbf{v}_2$．

如果我们将 $\mathbf{r}_1$ 指向 $\mathbf{r}_2$ 的单位矢量记为 $\hat{\mathbf{r}}$

$$\hat{\mathbf{r}}=\frac{|\mathbf{r}_2-\mathbf{r}_2|}{L} \tag{1}$$

长度平方随时间的变化的率为[4]

$$\frac{\mathrm{d}}{\mathrm{d}t}L^2=\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{r}_2-\mathbf{r}_1)^2 \tag{2}$$

由矢量求导法则

$$\frac{\mathrm{d}}{\mathrm{d}t}L^2=2L\frac{\mathrm{d}L}{\mathrm{d}t}=2(\mathbf{r}_2-\mathbf{r}_1)\cdot(\mathbf{v}_2-\mathbf{v}_1)=2L(\mathbf{v}_2-\mathbf{v}_1)\cdot\hat{\mathbf{r}} \tag{3}$$

即

$$\frac{\mathrm{d}L}{\mathrm{d}t}=\mathbf{v}_2\cdot\hat{\mathbf{r}}-\mathbf{v}_1\cdot\hat{\mathbf{r}} \tag{4}$$

所以要满足长度恒定不变，只需要满足（充分必要条件）

$$\mathbf{v}_2\cdot\hat{\mathbf{r}}=\mathbf{v}_1\cdot\hat{\mathbf{r}} \tag{5}$$

该式的意思是，两个端点的速度延杆的分量在任意时刻都相等．

[4]为什么我们要用长度的平方而不是直接对 $|\mathbf{r}_2-\mathbf{r}_1|$ 求导？我们的确可以这么做并得到同样的结果，但是使用平方会使计算更方便．

例 1

如图 3，杆的两端被固定在两个垂直的轨道上运动，两端的运动速度分别为 v_1, v_2，已知杆与水平轨道的夹角为 θ，求 v_1 和 v_2 的关系．解：由式 5，两

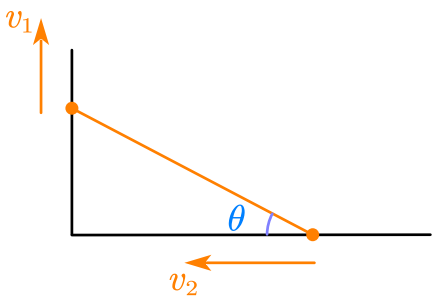


图 1: 杆的运动

速度在杆方向的分量应该相等，即

$$v_2 \cos\theta = v_1 \sin\theta \tag{6}$$

即

$$v_2 = v_1 \tan\theta \tag{7}$$

例 2　人拉船模型

如图 2，人在岸上通过一小滑轮用绳子以速度 v 拉船，绳子与水平面的夹角为 θ，求船的速度 u.

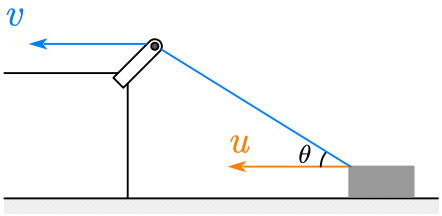


图 2: 人拉船模型

解：我们可以把从滑轮到船的这段绳子的长度缩短的速度用式 4 来计算，这个速度也就是人的速度

$$v = -\frac{\mathrm{d}L}{\mathrm{d}t} = u\cos\theta \tag{8}$$

所以

$$u = \frac{v}{\cos\theta} \tag{9}$$

习题 1

三个小朋友 ACB 两两相距 30 m，接下来的任意时刻 A 向着 B 跑，B 向着 C 跑，C 向着 A 跑，速度都相等．求他们相遇时，各跑了多少路程．

角速度

预备知识 圆周运动的加速度[885]

若已知两端点的速度，如何求出角速度呢？我们可以在以 $\mathbf{r}_1$ 为原点的无转动的参考系中观察，此时轻杆另一端的速度为 $\mathbf{v}_2' = \mathbf{v}_2 - \mathbf{v}_1$ 且与 $\hat{\mathbf{r}}$ 垂直，所以角速度大小为

$$\omega = \frac{|\mathbf{v}_2 - \mathbf{v}_1|}{r} \tag{10}$$

矢量表达式为

$$\boldsymbol{\omega} = \hat{\mathbf{r}} \times \frac{\mathbf{v}_2 - \mathbf{v}_1}{r} \tag{11}$$

加速度约束

若我们想知道对加速度的约束，只需对式 3 两边再求一次时间导数，即式 2 的二阶时间导数．

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}(\mathbf{r}_2 - \mathbf{r}_1)^2 = (\mathbf{v}_2 - \mathbf{v}_1)^2 + (\mathbf{r}_2 - \mathbf{r}_1)\cdot(\mathbf{a}_2 - \mathbf{a}_1) = 0 \tag{12}$$

用式 10 把 $(\mathbf{v}_2 - \mathbf{v}_1)^2$ 用 ω 表示，再两边除以 r 得

$$\mathbf{a}_1\cdot\hat{\mathbf{r}} - \mathbf{a}_2\cdot\hat{\mathbf{r}} = \omega^2 r \tag{13}$$

ω 是杆的瞬时角速度的大小．

例 3

在例 1 的模型中，若已知端点的速度 v_1, v_2 和 θ，求两点加速度 a_1, a_2 的关系．

解：由式 11 和式 13 得，

$$\omega = (v_2\cos\theta + v_1\sin\theta)/r \tag{14}$$

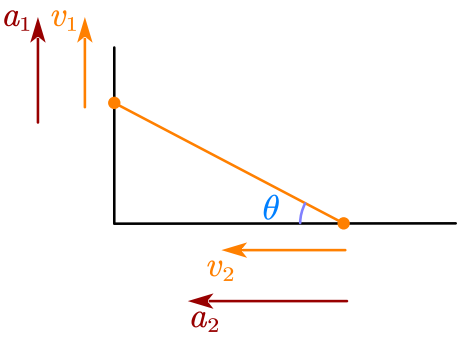


图 3: 杆的运动

$$a_2\cos\theta - a_1\sin\theta = \omega^2 r = (v_2\cos\theta + v_1\sin\theta)^2/r \tag{15}$$

动力学

预备知识 转动惯量[962]

轻杆没有质量，也没有转动惯量．假设我们只能在轻杆的两端对其施加两个力 $\mathbf{F}_1, \mathbf{F}_2$，这两个力会满足什么条件呢？首先，轻杆受到的合力必须为 $\mathbf{0}$，否则他就会马上被加速到无限快．这意味着这两个力大小相等，方向相反（$\mathbf{F}_1 + \mathbf{F}_2 = \mathbf{0}$）．其次，它受到的和力矩必须也为 $\mathbf{0}$，否则就会瞬间拥有无限大的角速度．这意味着两个力必须共线，即都延杆的方向．于是我们可以令 $\mathbf{F}_1 = F\hat{\mathbf{r}}$，$\mathbf{F}_2 = -F\hat{\mathbf{r}}$.

现在，无论轻杆如何运动，这两个力对轻杆做功的功率为

$$P = \mathbf{F}_1\cdot\mathbf{v}_1 + \mathbf{F}_2\cdot\mathbf{v}_2 = F(\hat{\mathbf{r}}\cdot\mathbf{v}_1 - \hat{\mathbf{r}}\cdot\mathbf{v}_2) \tag{16}$$

由式 5 可知功率恒为 0.

动量定理 动量守恒

预备知识 质点系的动量[939]

以下我们将根据牛顿定律推导出系统的**动量定理**，即系统总动量的变化率等于合外力

$$\dot{\mathbf{p}} = \mathbf{F}^{out} \tag{1}$$

由此可以得出系统所受合外力为零时系统总动量不随时间变化，即**动量守恒**（**conservation of momentum**）.

由式 3 [939]，还可以把式 1 改写为类似牛顿第二定律的形式，一些教材中被称为**质点系的牛顿第二定律**

$$M\mathbf{a}_c = \mathbf{F}^{out} \tag{2}$$

其中 $\mathbf{a}_c$ 是质点系质心的加速度. 系统合外力为零时，质心加速度为零，即静止或者做匀速运动.

推导

在牛顿力学中任何系统都可以看做质点系，质点系中第 i 个质点可能受到系统内力 $\mathbf{F}_i^{in}$ 或系统外力 $\mathbf{F}_i^{out}$. 由单个质点的动量定理[910]，

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}_i = \mathbf{F}_i^{in} + \mathbf{F}_i^{out} \tag{3}$$

总动量的变化率为

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} = \sum_i \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}_i = \sum_i \mathbf{F}_i^{in} + \sum_i \mathbf{F}_i^{out} \tag{4}$$

由“质点系[931]”中的结论，上式右边第一项求和是系统合内力，恒为零. 于是我们得到系统的动量定理

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} = \sum_i \mathbf{F}_i^{out} \tag{5}$$

可见当和外力（即等式右边）为零时，动量 $\mathbf{P}$ 不随时间变化，也就是**动量守恒**.

未完成：导弹爆炸，人船模型

例 1　静止原子核的转变

一个原来静止的原子核,经放射性衰变,放出一个动量为 $9.22\times10^{-16}\mathrm{g}\cdot\mathrm{cm/s}$ 的电子，同时该核在垂直方向上又放出一个动量为 $5.33\times10^{-16}\mathrm{g}\cdot\mathrm{cm/s}$ 的中微子. 问蜕变后原子核的动量的大小和方向.

解：由于这个静止的原子核在蜕变的全过程中没有受到其他外力，所以对该原子核构成的系统，总动量守恒．即有

$$\mathbf{p}_{\mathrm{B}}+\mathbf{p}_{\mathrm{e}}+\mathbf{p}_{\nu}=0 \tag{6}$$

即有

$$p_{\mathrm{B}}=|\mathbf{p}_{\mathrm{B}}|=|-\mathbf{p}_{\mathrm{e}}-\mathbf{p}_{\nu}|=\sqrt{p_{\mathrm{e}}^{2}+p_{\nu}^{2}}=10.65\times10^{-16}\mathrm{g}\cdot\mathrm{cm/s} \tag{7}$$

$$\theta=\arctan\frac{5.33}{9.22}=30^{\circ} \tag{8}$$

质点系的动能 柯尼希定理

预备知识 质点系的动量[939]

柯尼希定理

某参考系中质点系的动能等于该参考系中其质心的动能加上质心系中质点系的动能，即

$$E_k=\frac{1}{2}Mv_c^2+\frac{1}{2}\sum_i m_i v_{ci}^2 \tag{1}$$

其中 M 是所有质点的质量和，v_c 是质心系相对于当前参考系的运动速度，m_i 是第 i 个质点的质量，v_{ci} 是第 i 个质点在质心系中的速度．

例 1　圆环滚动的动能

一个圆环在水平地面上延直线无摩擦地滚动，其半径为 R，质量为 m，角速度为 ω，求地面参考系中圆环的动能．

圆环质心的速度大小为 $v_c=\omega R$，圆环相对于圆心旋转的线速度大小处处为 $v_{ci}=\omega R$，代入式 1 得动能为

$$E_k=\frac{1}{2}mv_c^2+\frac{1}{2}\sum_i m_i v_{ci}^2=\frac{1}{2}m\omega^2R^2+\frac{1}{2}\sum_i m_i\omega^2R^2=m\omega^2R^2 \tag{2}$$

证明

在当前参考系中，第 i 个质点的运动速度为

$$\mathbf{v}_i = \mathbf{v}_c + \mathbf{v}_{ci} \tag{3}$$

于是有

$$\begin{aligned} E_k &= \frac{1}{2}\sum_i m_i \mathbf{v}_i^2 = \frac{1}{2}\sum_i m_i (\mathbf{v}_c + \mathbf{v}_{ci})^2 \\ &= \frac{1}{2}\sum_i m_i \mathbf{v}_c^2 + \frac{1}{2}\sum_i m_i \mathbf{v}_{ci}^2 + \sum_i m_i \mathbf{v}_c \cdot \mathbf{v}_{ci} \end{aligned} \tag{4}$$

现在只需证明 $\sum_i m_i \mathbf{v}_c \cdot \mathbf{v}_{ci} = 0$ 即可．考虑到

$$\sum_i m_i \mathbf{v}_c \cdot \mathbf{v}_{ci} = \mathbf{v}_c \cdot \sum_i m_i \mathbf{v}_{ci} \tag{5}$$

而质心系中的质点系动量为零（式 29 [931]），所以

$$\sum_i m_i \mathbf{v}_{ci} = \mathbf{0} \tag{6}$$

证毕．

力矩

预备知识 矢量的叉乘[387]

平面力矩

如果只考虑一个厚度不计的片状物体在平面上的运动和受力，受力点位矢为 $\mathbf{r}$，力为 $\mathbf{F}$，那么对于一个给定的参考点（除非明确指出，一般取坐标原点），就可以计算物体的受到的力矩．

根据初中所学的方法，应该先作出“力臂” $\mathbf{r}_\perp$ 与力的方向垂直（图 1 左）．力矩的大小（用 τ 表示）为

$$\tau = |\mathbf{r}_\perp||\mathbf{F}| = |\mathbf{r}||\mathbf{F}|\sin\theta \tag{1}$$

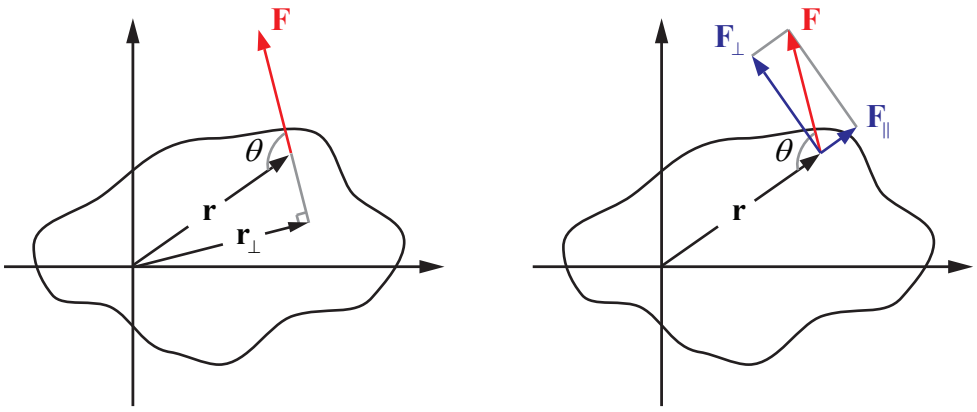


图 1: 力矩的两种几何理解

其中 θ 是 $\mathbf{r}$ 与 $\mathbf{F}$ 的夹角或其补角[5]．从另一种角度来看，也可以把力 $\mathbf{F}$ 正交分解为平行于 $\mathbf{r}$ 的分量和垂直于 $\mathbf{r}$ 的分量（图 1 右）．其中平行分量不产生力矩，垂直分量产生的力矩为

$$\tau = |\mathbf{r}||\mathbf{F}_\perp| = |\mathbf{r}||\mathbf{F}|\sin\theta \tag{2}$$

为了区分力矩的两个不同的方向（逆时针和顺时针），通常有两种做法：一是用正负号加以区分，例如规定逆时针的力矩为正，顺时针为负．这种定义把力矩看做一种标量（就像我们讨论一维运动时，将速度表示成标量，用正负号区分方向）．

一般定义

若物体受到若干个力，且受力点不在一个平面内，或者力方向不在同一平面内，则应该在三维空间内考虑力矩，这时力矩只能是矢量，使用位置矢量和力矢量的叉乘[387] 定义为

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F} \tag{3}$$

根据这种定义，单个力的力矩大小还是 $\tau = |\mathbf{r}||\mathbf{F}|\sin\theta$，但是得到的力矩是矢量．在平面问题中，逆时针的力矩垂直纸面指向读者，而顺时针的力矩方向相反．

力矩的叠加

[5]因为 $\sin(\pi - \theta) = \sin\theta$

预备知识 质心[931]，重积分[283]

若一个物体在多个位置 $\mathbf{r}_i$ 分别受力为 $\mathbf{F}_i$ （$i=1,\dots,N$）[6]，那么定义总力矩为每个点的力矩之和

$$\boldsymbol{\tau}=\sum_{i=1}^{N}\boldsymbol{\tau}_i=\sum_{i=1}^{N}\mathbf{r}_i\times\mathbf{F}_i \tag{4}$$

如果受力是连续分部的，假设受力的体密度为 $\mathbf{f}(\mathbf{r})$，那么可以由体积分计算总力矩

$$\boldsymbol{\tau}=\int\mathbf{r}\times\mathbf{f}(\mathbf{r})\,\mathrm{d}V \tag{5}$$

这个积分可以展开为三个分量的体积分.

习题 1　重力的力矩

若一个物体处于匀强重力场中，若它的密度分布为 $\rho(\mathbf{r})$ 求它受重力的力矩.

事实上这个问题与其用式 5，还有一种更简单的办法：假设这个物体由许多质点组成，每个质点位置为 $\mathbf{r}_i$，质量为 m_i，令重力加速度矢量为 $\mathbf{g}$（这是一个常矢量，指向下），则总力矩为

$$\boldsymbol{\tau}=\sum_i\mathbf{r}_i\times(m_i\mathbf{g})=\left(\sum_i m_i\mathbf{r}_i\right)\times\mathbf{g} \tag{6}$$

这里使用了叉乘的分配律. 再使用质心的定义，令总质量为 $M=\sum_i m_i$，得

$$\boldsymbol{\tau}=M\mathbf{r}_c\times\mathbf{g} \tag{7}$$

可见在计算匀强重力场对物体得力矩时，我们可以假设所有的重力都集中与质心一点.

力矩的坐标系变换

一般来说，由于受力点的位置矢量 $\mathbf{r}$ 与坐标系的选取有关，现在来看力矩在不同坐标系之间的变换.

[6]如果某个点处受到多个力，我们可以计算该点受到的合力记为 F_i

在坐标系 A 中，第 i 个受力点的位置矢量为 $\mathbf{r}_{Ai}$，物体的合力矩为

$$\boldsymbol{\tau}_A = \sum_i \mathbf{r}_{Ai} \times \mathbf{F}_i = \mathbf{0} \tag{8}$$

在另一坐标系 B 中，B 原点指向 A 原点的矢量为 $\mathbf{r}_{BA}$，合力矩为

$$\begin{aligned} \boldsymbol{\tau}_B &= \sum_i (\mathbf{r}_{Ai} + \mathbf{r}_{BA}) \times \mathbf{F}_i = \sum_i \mathbf{r}_{Ai} \times \mathbf{F}_i + \sum_i \mathbf{r}_{BA} \times \mathbf{F}_i \\ &= \boldsymbol{\tau}_A + \mathbf{r}_{BA} \times \sum_i \mathbf{F}_i \end{aligned} \tag{9}$$

其中最后两步使用了叉乘的分配律[387]．由结论可以看出，变换坐标系，力矩需要加上原坐标系相对新坐标系的位移叉乘物体的合力．由此也可以得出，**若物体的合力为零，则力矩与参考系无关．**

刚体的静力平衡

预备知识 动量定理[944]，角动量定理[954]

[7]在惯性系中，如果刚体所受的所有合外力与合外力矩[947] 都为零，则我们说它处于**静力平衡（static equilibrium）**．其中合外力（矩）是指所有施加在刚体上的力（矩）的矢量和．通常来说，静力平衡意味着刚体保持静止不动，但严格来说刚体质心可以做匀速运动，刚体也可以绕质心做匀速转动．在非惯性系中，若加入惯性力的修正，该结论仍然成立．

注意当合外力为零时，合外力矩与参考点（参考系）的选取无关（式 9 [947]）．

证明

把刚体看做由许多质点组成，合外力为零时刚体动量守恒[944]，而动量等于“质心的动量”[939] $\mathbf{p}_c = M_c \mathbf{v}_c$，所以质心做匀速运动或不动．

刚体合外力矩为零时，质点系角动量守恒[954]，而角动量等于质心的角动量 $\mathbf{L}_c = \mathbf{r}_c \times \mathbf{p}_c$ 加上质心系中的角动量（式 8 [952]）．当质心匀速直线运动或不动时 $\mathbf{L}_c$ 不变，所以质心系中刚体的角动量也不变，所以刚体绕质心做匀速转动或不转动．

[7]参考 Wikipedia 相关页面和 [11].

例 1　轻杆三力平衡

如图，一个长度为 L 质量不计的细杆，中间和两端受力分别为 $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3$.

未完成：图，选取不同受力点

未完成：吊桥例题，见 EP1 20201021，缆绳受力与重物位置的关系.

例 2

如图 1，一个质量为 m 的线轴被斜挂在墙上，线轴与墙面的摩擦系数为 μ，线轴的大圆半径为 R，小圆半径为 r，求当 α 满足什么条件时，线轴才能不滑落.

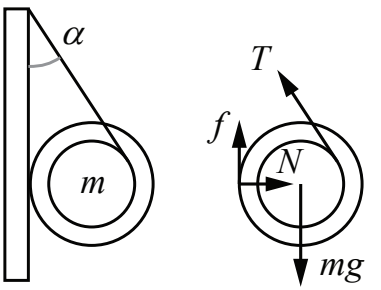


图 1: 线轴的平衡

我们先来看线轴受哪几个力：重力 mg，绳的拉力 T，墙的支持力 N 和摩擦力 f. 由摩擦系数的定义和刚体平衡条件可得

$$\begin{cases} f \leqslant \mu N & \text{（摩擦系数）} \\ N - T\sin\alpha = 0 & \text{（水平方向受力平衡）} \\ T\cos\alpha + f - mg = 0 & \text{（竖直方向受力平衡）} \\ Tr - fR = 0 & \text{（力矩平衡）} \end{cases} \tag{1}$$

其中最后一条力矩平衡是以圆心为原点计算力矩，虽然原则上我们可以取任意点计算力矩，但取在圆心计算最为简单. 除了 α 我们有三个未知数 T, f, N，用以上三条等式恰好可以把这三个未知数消去，可得关于 α 的不等式

$$\sin\alpha \geqslant \frac{r}{\mu R} \tag{2}$$

一个有趣的地方在于，不等式中没有出现质量 m. 事实上，我们不使用那条含有 mg 的等式也可以顺利得到答案.

角动量

预备知识 质点的角动量[911]，质点系[931]

系统的角动量

角动量是矢量，若把系统看做质点系，则系统的角动量等于所有质点的角动量矢量相加[375].

$$\mathbf{L}=\sum_i \mathbf{L}_i=\sum_i \mathbf{r}_i\times\mathbf{p}_i=\sum_i m_i\mathbf{r}_i\times\mathbf{v}_i \tag{1}$$

例 1　旋转圆环

在例 1 [911] 中，我们知道单个质点做圆周运动的角动量为 $\mathbf{L}=mr^2\omega\mathbf{z}$. 现在考虑一个半径为 r 质量为 M 的细圆环绕处于原点的圆心做逆时针圆周运动，每个质点的轨迹就是圆环本身. 如果我把圆环看成由许多质点 m_i 组成，每个质点都做上述圆周运动，则总角动量为

$$\mathbf{L}=\sum_i m_i\omega r^2\hat{\mathbf{z}}=M\omega r^2\hat{\mathbf{z}} \tag{2}$$

角动量的坐标系变换

可类比力矩的坐标系变换（式 9 [947]），坐标系 A 中总角动量为

$$\mathbf{L}_A=\sum_i \mathbf{r}_{Ai}\times\mathbf{p}_i \tag{3}$$

变换到与 A **相对静止的**坐标系 B 中，从 B 原点指向 A 原点的矢量为 $\mathbf{r}_{BA}$，那么总角动量为

$$\mathbf{L}_B=\sum_i(\mathbf{r}_{BA}+\mathbf{r}_{Ai})\times\mathbf{p}_i=\mathbf{r}_{BA}\times\sum_i\mathbf{p}_i+\mathbf{L}_A \tag{4}$$

可见当系统总动量为零时，两参考系中计算系统角动量的结果相同. 注意由于两参考系相对静止，以上的 $\mathbf{p}_i$ 在两参考系中都是相同的.

例 2　旋转圆环 2

如果我们把例 1 中的旋转圆环平移一下再计算（关于原点）的角动量，根据上述讨论，结果仍然是相同的．这是因为圆环中的每个质点都存在一个反方向运动的质点，圆环的总动量为零．

角动量的质心系分解

若当前参考系不是质心系，令质心的位置矢量为 $\mathbf{r}_c$，速度为 $\mathbf{v}_c$．每个质点在质心系中的位置为 $\mathbf{r}_{ci}$，速度为 $\mathbf{v}_{ci}$，令系统总质量为 $M=\sum_i m_i$，则系统角动量可以表示为

$$\begin{aligned}\mathbf{L}&=\sum_i m_i\mathbf{r}_i\mathbf{v}_i=\sum_i m_i(\mathbf{r}_c+\mathbf{r}_{ci})\times(\mathbf{v}_c+\mathbf{v}_{ci})\\&=\mathbf{r}_c\times(M\mathbf{v}_c)+\sum_i m_i\mathbf{r}_{ci}\times\mathbf{v}_{ci}+\mathbf{r}_c\times\sum_i m_i\mathbf{v}_{ci}+\left(\sum_i m_i\mathbf{r}_{ci}\right)\times\mathbf{v}_c\end{aligned}\tag{5}$$

根据质心系的性质，最后两项中的求和为零．另外根据式 3 [939]，$M\mathbf{v}_c$ 就是系统总动量，也相当于所有质量集中于质心获得的动量．所以我们可以把第一项定义所谓的“质心的角动量”

$$\mathbf{L}_0=\mathbf{r}_c\times(M\mathbf{v}_c)\tag{6}$$

而第二项就是质心系中系统的角动量

$$\mathbf{L}_c=\sum_i m_i\mathbf{r}_{ci}\times\mathbf{v}_{ci}\tag{7}$$

所以式 5 可以记为

$$\mathbf{L}=\mathbf{L}_0+\mathbf{L}_c\tag{8}$$

所以**任何坐标系中，系统的总角动量等于其质心的角动量加上相其相对质心的角动量**．注意与上一小节不同的是，我们不要求两个参考系相对静止．

习题 1　旋转圆环 3

在例 2 的基础上，如果圆环一边旋转，圆心一边以速度 $\mathbf{v}$ 做匀速直线运动，直线与当前原点距离为 d，求系统的总角动量．角动量是否守恒？

角动量定理 角动量守恒

预备知识 角动量[952]

角动量定理可以表示为

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t}=\boldsymbol{\tau} \tag{1}$$

即系统总角动量（矢量）$\mathbf{L}$ 对时间的变化率等于所有外力矩的矢量和（**合外力矩**）$\boldsymbol{\tau}$. 角动量定理可以类比动量定理[944]，其中角动量与系统动量 $\mathbf{p}$ 对应，合外力矩与合外力 $\mathbf{F}$ 对应. 注意一般情况下，$\mathbf{L}$ 和 $\boldsymbol{\tau}$ 的方向不一定相同，只有在例如刚体绕固定轴转动[962] 的特殊情况时，二者才相同. 式 1 的证明见下文.

作为一个简单的情况，我们来看刚体的绕轴转动[962]. 此时刚体的转动惯量以及合外力矩的方向都固定的转轴平行[8]，于是我们可以规定一个正方向，把 $\mathbf{L},\boldsymbol{\tau}$ 用实数表示：当实数为正，则矢量指向正方向，反之则指向反方向.

然而一般情况下，$\mathbf{L}$ 和 $\boldsymbol{\tau}$ 的方向都可以随时间改变，且二者方向未必相同. 这可以类比对质点的曲线运动（如圆周运动）使用动量定理：加速度不仅取决于速度大小的变化，还取决于速度方向的变化；动量和力的方向也未必相同.

例 1　陀螺的进动

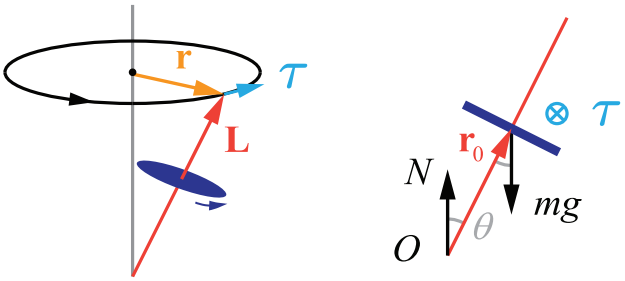


图 1: 陀螺的进动

如图 1（左），陀螺旋转时，若它的轴与竖直方向有一定倾角，轴会绕一个竖直轴缓慢旋转，这种现象被称为进动（**precession**）. 为了便于分析，我们

[8]这可以类比质点做直线运动的动量定理[944]，动量和力都沿同一方向.

先假设陀螺进动的角速度比陀螺自转的角速度要慢得多．这样，我们就可以认为陀螺的角动量 $\mathbf{L}$ 与陀螺的轴平行．

陀螺在进动过程中，角动量大小不变，但方向不断变化，所以角动量变化率 $\mathrm{d}\mathbf{L}/\mathrm{dt}$ 不为零．令 $\mathbf{L}$ 的起点为原点，$\mathbf{L}$ 末端在圆形轨迹上运动．$\mathrm{d}\mathbf{L}/\mathrm{dt}$ 的方向始终沿着该圆形轨迹的切线方向．根据角动量定理，陀螺所受的力矩 $\boldsymbol{\tau}$ 也具有同样的大小和方向．

那么这个力矩是如何产生的呢？我们对陀螺进行受力分析如图 1（右），要计算陀螺所受力矩，我们取轴的底端为原点，假设陀螺的轴没有质量，则地面对陀螺的支持力 N 产生的力矩为零，而重力产生的力矩为 $\boldsymbol{\tau} = \mathbf{r}_0 \times (m\mathbf{g})$，其大小为 $mgr_0 \sin\theta$，方向垂直纸面向里，恰好符合陀螺进动的要求．

比较违反直觉的地方在于，陀螺受到的重力是延使陀螺倾倒的方向施加的，然而陀螺不但丝毫不会倾倒（如果不计摩擦），反而其重心会向着与重力垂直的方向移动．要具体计算陀螺进动的快慢，我们还需要知道角动量和陀螺自转的角速度的关系，见习题 1 [962]．

角动量守恒

当式 1 中系统受合外力矩为零时，角动量变化率为零，即系统总角动量不随时间变化．这时我们说该系统**角动量守恒（conservation of angular momentum）**．

例 2　面团碰撞

在没有引力的太空中，两个面团在各自质心系中的角动量分别为 $\mathbf{L}_1$ 和 $\mathbf{L}_2$．他们的质心在某惯性系 S 的 x 轴上相向运动然后相撞并融为一体，求碰撞后的角动量．

解：根据角动量的质心系分解（式 8 [952]），在惯性系 S 中他们的角动量仍然为 $\mathbf{L}_1$ 和 $\mathbf{L}_2$．把两个面团看成一个系统，总角动量为 $\mathbf{L} = \mathbf{L}_1 + \mathbf{L}_2$．由于不受任何外力（矩），根据角动量守恒，相撞后总角动量仍然为 $\mathbf{L}$．

习题 1　面团碰撞 2

在没有引力的太空中，质量为 m_1, m_2 的两个面团在各自质心系中的角动量分别为 $\mathbf{L}_1$ 和 $\mathbf{L}_2$．他们的质心分别在某惯性系 S 的 x 轴和 $y = 1$ 直线上运动，x 方向速度分别为 v_1, v_2．若它们相撞并融为一体，求碰撞后的角动量．

角动量分量守恒

在一些情况下，我们不能完全保证合外力矩为零，而只能得出合外力矩在某个方向 $\hat{\mathbf{e}}$ 的投影（即分量）为零

$$\boldsymbol{\tau}\cdot\hat{\mathbf{e}}\equiv 0 \tag{2}$$

把式 1 两边同时点成 $\hat{\mathbf{e}}$，得

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{L}\cdot\mathbf{e})=0 \tag{3}$$

这就说明 $\mathbf{L}$ 在 $\hat{\mathbf{e}}$ 方向得投影不随时间变化，即角动量分量守恒.

角动量分量守恒最常见的例子是刚体绕固定轴的无摩擦转动[962]：若不施加额外的力矩，轴只可能给刚体施加垂直轴方向的力矩，所以刚体在平行轴方向的角动量分量守恒.

例 3　单车轮与转椅实验

小明开始时坐在静止的无摩擦转椅上，两手握住一个单车轮的轴的两端，单车轮在水平面上转动. 这时小明将单车轮上下翻转（仍保持转动），问小明与转椅会如何转动？

假设开始时车轮的角动量向上，那么翻转后车轮的角动量向下，即角动量增量向下. 由于角动量守恒，小明的身体和转椅的角动量必须有一个向上的增量，所以转椅最后的旋转方向与轮子开始时的旋转方向相同.

证明角动量定理

证明可类比系统的动量定理[944]. 我们已经知道单个质点的角动量，而任何物体都可以划分成若干足够小的微元，每个微元可以看成一个质点. 令第 i 个质点的位矢为 $\mathbf{r}_i$，角动量为 $\mathbf{L}_i$，力矩为 $\boldsymbol{\tau}_i$，单个质点的角动量定理[911] 为

$$\frac{\mathrm{d}\mathbf{L}_i}{\mathrm{d}t}=\boldsymbol{\tau}_i=\boldsymbol{\tau}_i^{in}+\boldsymbol{\tau}_i^{out} \tag{4}$$

其中 $\boldsymbol{\tau}_i^{in}$ 和 $\boldsymbol{\tau}_i^{out}$ 为质点 i 受到的系统内其他质点的力矩和来自系统外的力矩. 将该式对所有 i 求和，得到总角动量 $\mathbf{L}$ 变化率

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t}=\sum_i\frac{\mathrm{d}\mathbf{L}_i}{\mathrm{d}t}=\sum_i\boldsymbol{\tau}_i^{in}+\sum_i\boldsymbol{\tau}_i^{out} \tag{5}$$

现在我们只需证明质点系的合内力矩为零即可

$$\sum_i \boldsymbol{\tau}_i^{in} = \sum_i \left(\mathbf{r}_i \times \sum_j^{j\neq i} \mathbf{F}_{j\to i} \right) = \sum_{i,j}^{i\neq j} \mathbf{r}_i \times \mathbf{F}_{j\to i} \tag{6}$$

其中 $\mathbf{F}_{j\to i}$ 是质点 j 对质点 i 的力．现在只考虑任意两个质点 k 和 l，在求和中的贡献为

$$\mathbf{r}_k \times \mathbf{F}_{l\to k} + \mathbf{r}_l \times \mathbf{F}_{k\to l} \equiv \boldsymbol{\tau}_{l\to k} + \boldsymbol{\tau}_{k\to l} \tag{7}$$

即 k 对 l 的力矩加 l 对 k 的力矩（两质点的和内力矩）．所以若能证明任意两质点的和内力矩为零，则质点系的合内力矩为零．

我们先来看几何证明．如图 2，根据定义，力矩的大小等于力的模长乘以力臂的长度[947]，而一对相互作用力的大小相同，又由于二者共线，力臂也重合，所以两个力矩大小相等．但是两个力矩的方向一个是顺时针（指向纸内），一个是逆时针（指向纸外），所以两力矩互相抵消，相加为零．

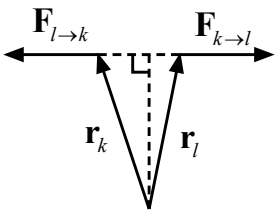


图 2: 两质点的相互作用力对总力矩贡献为零

再在看代数的方法：我们先沿着两质点的连线写出相互作用力 $\mathbf{F}_{l\to k} = \alpha(\mathbf{r}_k - \mathbf{r}_l)$，$\mathbf{F}_{k\to l} = \alpha(\mathbf{r}_l - \mathbf{r}_k)$，其中 α 是一个常数．直接计算两力矩和得

$$\mathbf{r}_k \times (\mathbf{r}_k - \mathbf{r}_l)\alpha + \mathbf{r}_l \times (\mathbf{r}_l - \mathbf{r}_k)\alpha = 0 \tag{8}$$

证毕．

二体系统

预备知识 自由度[930]，动量定理[944]

我们现在考虑两个仅受相互作用的质点 A 和 B，它们的质量分别为 m_A 和 m_B．由于不受系统外力，在任何惯性系中它们的质心[931] 都会做匀速直线运动（式 2 [944]）．所以系统的质心参考系是一个惯性系，以下我们选取质心参考系，系统的质心将一直处于原点．

现在定义它们的**相对位矢**（也叫**相对坐标**）为点 A 指向点 B 的矢量

$$\mathbf{R}=\mathbf{r}_B-\mathbf{r}_A \tag{1}$$

且定义**相对速度**和**相对加速度**分别为 $\mathbf{R}$ 的导数 $\dot{\mathbf{R}}$ 和二阶导数 $\ddot{\mathbf{R}}$．在质心系中观察，由于质心始终处于原点，两质点的位矢 $\mathbf{r}_A$ 和 $\mathbf{r}_B$ 满足

$$m_A\mathbf{r}_A+m_B\mathbf{r}_B=\mathbf{0} \tag{2}$$

联立式 1 和式 2 可以发现在质心系中 $\mathbf{R},\mathbf{r}_A,\mathbf{r}_B$ 间始终存在一一对应的关系（共线且模长呈固定比例），所以质心系中不受外力的二体系统只有三个自由度[930]

$$\mathbf{r}_A=\frac{-m_B}{m_A+m_B}\mathbf{R}\qquad \mathbf{r}_B=\frac{m_A}{m_A+m_B}\mathbf{R} \tag{3}$$

运动方程

现在令质点 A 对 B 的作用力为 $\mathbf{F}$ （与 $\mathbf{R}$ 同向），则由牛顿第三定律，B 对 A 有反作用力 $-\mathbf{F}$．两质点加速度分别为（牛顿第二定律）$\mathbf{a}_A=-\mathbf{F}/m_A$，$\mathbf{a}_B=\mathbf{F}/m_B$．所以相对加速度为

$$\ddot{\mathbf{R}}=\ddot{\mathbf{r}}_B-\ddot{\mathbf{r}}_A=\frac{m_A+m_B}{m_Am_B}\mathbf{F} \tag{4}$$

若定义两质点的**约化质量**为

$$\mu=\frac{m_Am_B}{m_A+m_B} \tag{5}$$

且将上式两边同乘约化质量，我们得到相对位矢的牛顿第二定律

$$\mathbf{F}=\mu\ddot{\mathbf{R}} \tag{6}$$

也就是说，在质心系中使用相对位矢，二体系统的运动规律就相当于单个质量为 μ，位矢为 $\mathbf{R}$ 的质点的运动规律，我们姑且将其称为**等效质点**．而 A 对 B 的作用力可以看成等效质点的受力．

例 1　两天体圆周运动

令质量分别为 m_1 和 m_2 的天体距离为 R，在万有引力作用下绕质心做圆周运动，求角速度 ω.

解：两天体之间的引力大小为（式 1 [1027]）

$$F = \frac{Gm_1m_2}{R^2} \tag{7}$$

如果不使用等效质点的概念，我们可以先用式 3 得到两个天体做圆周运动的半径，然后再令引力等于其中一个天体的向心力[900]

$$F = m_1\omega^2 r_1 \tag{8}$$

也可以不求 r_1, r_2，直接使用等效天体的圆周运动向心力

$$F = \mu\omega^2 R \tag{9}$$

易得以上两式是等效的，解得

$$\omega = \sqrt{\frac{G(m_1+m_2)}{R^3}} \tag{10}$$

机械能

再来看系统的动能．令 $\mathbf{v} = \dot{\mathbf{R}}$，使用式 3 把系统在质心系中的总动能用相对位矢表示得

$$E_k = \frac{1}{2}(m_A\dot{\mathbf{r}}_A^2 + m_B\dot{\mathbf{r}}_B^2) = \frac{1}{2}\frac{m_Am_B}{m_A+m_B}\mathbf{v}^2 = \frac{1}{2}\mu\mathbf{v}^2 \tag{11}$$

这恰好是等效质点动能.

若两质点间的相互作用力的大小只是二者距离 $R = |\mathbf{R}|$ 的函数，我们可以用一个标量函数 $F(R)$ 来表示力与距离的关系，即

$$\mathbf{F}(\mathbf{R}) = F(R)\hat{\mathbf{R}} \tag{12}$$

注意 $F(R) > 0$ 时两质点存在斥力，$F(R) < 0$ 时存在引力.

根据“势能[905]”中的式 20，我们可以定义势能函数 $V(R)$ 为 $F(R)$ 的一个负原函数．现在写出二体系统在质心系中的机械能为

$$E = \frac{1}{2}\mu\dot{\mathbf{R}}^2 + V(R) \tag{13}$$

由于系统不受外力，机械能守恒.

动量守恒

在质心系中，两质点的总动量恒为零，动量守恒．令它们的动量分别为 $\mathbf{p}_A$ 和 $\mathbf{p}_B$，有

$$\mathbf{p}_B = m_B\dot{\mathbf{r}}_B = \mu\mathbf{v} \tag{14}$$

令等效质点的动量为 $\mathbf{p} = \mu\mathbf{v}$，则

$$\mathbf{p} = \mathbf{p}_B = -\mathbf{p}_A \tag{15}$$

即等效质点的动量等于质点 B 的动量或质点 A 动量的逆矢量．

角动量守恒

由式 15 可得二体系统总角动量为

$$\mathbf{L} = \mathbf{r}_A \times \mathbf{p}_A + \mathbf{r}_B \times \mathbf{p}_B = \mathbf{R} \times \mathbf{p} \tag{16}$$

即二体系统的总角动量等于等效质点的角动量．

如果两个质点之间的相互作用力沿它们连线的方向，那么所有两个力都经过质心，即都不受任何力矩，所以系统角动量守恒．

二体碰撞

预备知识 二体系统[957]

注意以下讨论的碰撞不必要求在一瞬间发生，可以拓展到有限距离的作用力甚至无穷远但不断衰减的作用力．例如考虑两个带电荷的质点的碰撞．在无穷远处时，二者之间的作用可忽略，此时的速度可定义为初速度．当发生相互作用后，把两质点互相远离到相距无穷远时的速度定义为末速度．

一维情况

高中物理中，若两质点的运动限制在同一直线上且碰撞为完全弹性碰撞，我们可以联立能量守恒和动量守恒两条式子来解出碰撞后的速度．但这里介绍

另一种更简单的方法，即利用质心系求解．为了区别于质心系，我们把原参考系叫做**实验室参考系**（简称为实验系）．令两质点质量分别为 m_1 和 m_2，实验系中初速度分别为 v_{10} 和 v_{20}，需要求实验系中的末速度 v_1 和 v_2．根据定义，系统质心的位置为 $x_c=(m_1x_1+m_2x_2)/(m_1+m_2)$，等式两边对时间 t 求导，得质心的速度为

$$v_c=(m_1v_1+m_2v_2)/(m_1+m_2) \tag{1}$$

现在我们在质心系中考虑该问题．在三维情况下，质心系中的二体系统只有三个自由度（式 3 [957]），不难类推在一维情况下二体系统只有一个自由度．所以无论维度多少，在质心系中考虑二体碰撞问题将会简单得多．

由速度叠加原理，初始时两质点在质心系中的速度分别为

$$v_{c10}=v_{10}-v_c \qquad v_{c20}=v_{20}-v_c \tag{2}$$

先考虑质心系中的完全弹性碰撞，由于两质点的速度大小始终成正比（质心系[931]式 29），为了使能量守恒，碰撞只能有一种结果，即两质点的速度方向都取反方向而速度大小保持不变．现在我们重新回到实验系中，两质点的末速度分别为

$$v_1=v_c+(-v_{c10})=2v_c-v_{10} \qquad v_2=v_c+(-v_{c20})=2v_c-v_{20} \tag{3}$$

代入式 1 即可得到最后结果．

若问题为非完全弹性碰撞，可设质心系中碰撞后与碰撞前的能量比值为 $\alpha^2<1$，即速度的比值为 α．碰撞后两质点的质心系速度分别变为 $-\alpha v_{c10}$ 和 $-\alpha v_{c20}$，变换到实验系中速度为

$$\begin{aligned} v_1&=v_c+(-\alpha v_{c10})=(1+\alpha)v_c-\alpha v_{10}\\ v_2&=v_c+(-\alpha v_{c20})=(1+\alpha)v_c-\alpha v_{20} \end{aligned} \tag{4}$$

三维的情况

由于在多维情况下，碰撞损失的能量可能与碰撞的角度有关，这里仅讨论最常见的完全弹性碰撞．碰撞的轨迹如图 1 所示．由于质心系中系统的总动量始终为 0（质心系[931]式 29），初状态和末状态中两质点的速度方向相反，但延长线一般不重合（否则就变为上面的一维情况）．质点末状态速度与初状态速

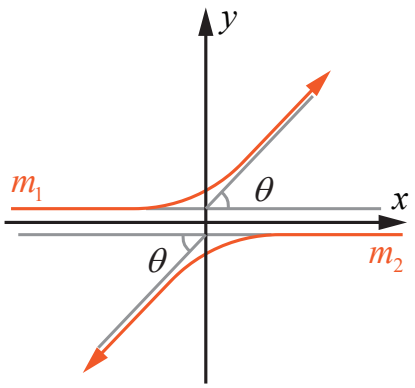


图 1: 质心系中的二体碰撞

度的夹角叫做**散射角（scattering angle）**. 注意质心系中两质点散射角相同，而实验系中两散射角不必相同. 求散射角需要知道具体的作用力形式，以下讨论假设我们已知质心系中的散射角. 若两质点间的相互作用力与两点的连线共线，那么质心系中两质点的运动轨迹将始终在同一平面上，而这在实验系中一般不成立（当两个质点的入射延长线为两条不平行且不相交的直线时）.

令质量分别为 m_1 和 m_2 的两质点初始速度为 $\mathbf{v}_{10}$ 和 $\mathbf{v}_{20}$，为了方便，我们规定质心系的 x 轴与第一个质点在质心系中的入射方向相同，则质心系中的初始速度分别为

$$\mathbf{v}_{c10} = (v_{10} - v_c)\hat{\mathbf{x}} \qquad \mathbf{v}_{c20} = (-v_{20} - v_c)\hat{\mathbf{x}} \tag{5}$$

其中质心速度为

$$\mathbf{v}_c = (m_1 v_1 + m_2 v_2)\hat{\mathbf{x}}/(m_1 + m_2) \tag{6}$$

完全弹性碰撞说明能量守恒，类比一维的情况可得质心系中两质点末速度的大小分别等于初速度大小. 所以末速度分别为

$$\mathbf{v}_{c1} = (v_{10} - v_c)(\hat{\mathbf{x}}\cos\theta + \hat{\mathbf{y}}\sin\theta) \qquad \mathbf{v}_{c2} = -(v_{20} + v_c)(\hat{\mathbf{x}}\cos\theta + \hat{\mathbf{y}}\sin\theta) \tag{7}$$

最后再变换到实验系，得末速度分别为

$$\mathbf{v}_1 = \mathbf{v}_c + \mathbf{v}_{c1} \qquad \mathbf{v}_2 = \mathbf{v}_c + \mathbf{v}_{c2} \tag{8}$$

刚体定轴转动转动惯量

预备知识 刚体[939]，角动量定理[954]

刚体的绕轴转动

若刚体绕固定轴转动，那么刚体的位置只需一个变量即可完全确定（一个自由度），我们令该变量为转角 θ．θ 关于时间 t 的导数就是刚体绕轴旋转的角速度 ω．我们还可以定义角速度 ω 关于时间的导数（即 θ 关于时间的二阶导数）为**角加速度（angular acceleration）**，记为 α．

我们可以把刚体的绕轴转动类比质点的直线运动，把 θ，ω 和 α 分别类比为直线运动中的位置 x，速度 v 和加速度 a，因为后三个变量之间的数学关系是完全相同的．于是我们可以立即得到匀变速转动（即 α 为常数）的一些公式，如

$$\theta = \theta_0 + \omega t + \frac{1}{2}\alpha t^2 \tag{1}$$

$$\omega_1^2 - \omega_0^2 = 2\alpha\theta \tag{2}$$

在以上三个标量的基础上，我们可以定义它们的矢量形式 $\boldsymbol{\theta}$，$\boldsymbol{\omega}$ 和 $\boldsymbol{\alpha}$，令它们的方向为转轴的方向，用右手定则[20] 来判断．

要判断刚体上任意一点的速度，使用式 5 [883] 即可（见图 1）

$$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r} \tag{3}$$

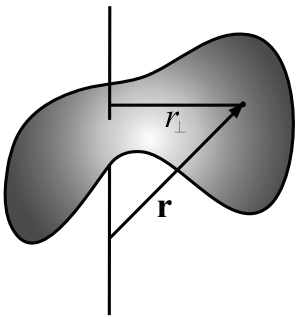


图 1: 刚体绕轴旋转时任意一点的线速度

角动量与转动惯量

设刚体绕固定轴转动，令轴的方向为 $\hat{\mathbf{z}}$．假设轴光滑，则轴对刚体可施加 x,y 两个方向的力矩，却不能施加 z 方向的力矩．所以根据角动量定理（式 1 [954]），角动量 $\mathbf{L}$ 的 z 分量 L_z 守恒．我们下面来推导 L_z 与角速度 ω 的关系．矢量 $\mathbf{L}$ 与矢量 $\boldsymbol{\omega}$ 的关系见惯性张量[973]．

对于刚体上的单个质点，$L_z=(\mathbf{r}\times\mathbf{p})\cdot\hat{\mathbf{z}}$．首先把质点的位矢在水平方向和竖直方向分解，$\mathbf{r}=\mathbf{r}_z+\mathbf{r}_\perp$．由于 $\mathbf{p}$ 一直沿水平方向，根据叉乘的几何定义，$\mathbf{r}_z\times\mathbf{p}$ 也是沿水平方向，只有 $\mathbf{r}_\perp\times\mathbf{p}$ 沿 z 方向．另外，在圆周运动中，半径始终与速度垂直，所以 $\mathbf{r}_\perp$ 始终与 $\mathbf{p}$ 垂直．得出结论

$$L_z=|\mathbf{r}_\perp||\mathbf{p}|=mr_\perp v=mr_\perp^2\omega \tag{4}$$

若把刚体分成无数小块，每小块的质量分别为 m_i，离轴的距离 $r_{\perp i}=\sqrt{x_i^2+y_i^2}$，则刚体的角动量 z 分量为

$$L_z=\omega\sum_i m_i r_{\perp i}^2 \tag{5}$$

用积分写成

$$L_z=\omega\int r_\perp^2\,\mathrm{d}m=\omega\int r_\perp^2\rho(\mathbf{r})\,\mathrm{d}V \tag{6}$$

定义刚体绕固定轴旋转的**转动惯量（moment of inertia）**为

$$I=\int r_\perp^2\,\mathrm{d}m=\omega\int r_\perp^2\rho(\mathbf{r})\,\mathrm{d}V \tag{7}$$

（注意角动量的大小不仅取决于刚体的质量分布，还取决于转轴的位置和方向）则刚体沿轴方向的角动量分量为

$$L_z=I\omega \tag{8}$$

现在来看“角动量定理[954]”的式 1，注意等号两边是矢量，所以各个分量必须相等，我们有

$$\frac{\mathrm{d}L_z}{\mathrm{d}t}=\tau_z \tag{9}$$

将式 8 代入式 9，并利用角加速度的定义得

$$I\alpha=\tau_z \tag{10}$$

这就是刚体绕轴转动的动力学方程，其形式可类比质点做直线运动时的牛顿第二定律[899]．

例 1　物理摆

如图 2，已知质量为 M 的薄片绕某点的转动惯量为 I，转轴到刚体质心的长度为 r_c，转轴和质心的连线与竖直方向夹角为 θ，求刚体的运动方程.

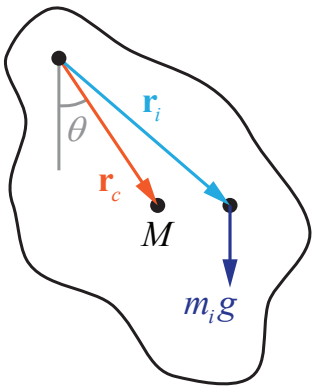


图 2: 物理摆

首先我们把刚体看做质点系，以转轴为原点计算刚体的合力矩为（由于这是一个平面问题，力矩必然垂直于该平面）

$$\begin{aligned}\boldsymbol{\tau} &= \sum_i \mathbf{r}_i \times (m_i \mathbf{g}) = \left(\sum_i m_i \mathbf{r}_i\right) \times \mathbf{g} = M\mathbf{r}_c \times \mathbf{g} \\ &= Mgr_c \sin\theta\end{aligned} \tag{11}$$

这就说明，刚体所受力矩相当于质量为 M，长度为 r_c 的单摆所受的力矩. 代入式 10 得刚体摆的运动方程为

$$I\ddot{\theta} = Mgr_c \sin\theta \tag{12}$$

可以验证当刚体的质量全部集中在质心时（$I = Mr_c^2$）我们就得到了单摆的运动方程式 4 [917].

习题 1　陀螺进动的角速度

在“角动量定理[954]”的例 1 中，如果除 r_0, m, g 外，还知道陀螺的转动惯量为 I 和陀螺的角速度 ω，试证明陀螺进动的角速度为

$$\Omega = \frac{mgr_0}{I\omega} \tag{13}$$

注意进动角速度与陀螺倾角 θ 无关.

垂直轴的角动量

以上的讨论中，我们有意避免讨论垂直轴方向的角动量分量 L_x, L_y．一般情况下，我们不能保证他们是守恒的．在一些特殊情况下，例如刚体的形状和质量分布关于转轴呈轴对称，那么容易证明刚体（关于任意固定点）的总角动量 $\mathbf{L}$ 只可能延转轴方向，即 $L_x = L_y = 0$ 守恒．

在一些不对称的情况下，例如一个倾斜的细杆绕转轴旋转，转轴就需要对细杆施加一个不停旋转的力矩，细杆也会对轴施加一个反力矩，这类似于作用力和反作用力，详见“刚体定轴转动 2[978]”.

平行轴定理与垂直轴定理

预备知识 转动惯量[962]

平行轴定理

若我们已知刚体关于一个通过其质心的轴（称为**质心轴**）的转动惯量为 I_0，那么我们可以通过平行轴定理简单地求出刚体关于另一个与质心轴平行的轴的转动惯量 I，而无需重新算一次定积分．令两个轴之间的距离为 R，刚体质量为 M，则计算公式为

$$I = I_0 + MR^2 \tag{1}$$

要证明该式，我们把刚体看做质点系，令质心轴到质点 m_i 的垂直矢量为 $\mathbf{r}_i$，平行轴到质心轴的垂直矢量为 $\mathbf{R}$，则刚体关于平行轴的转动惯量为

$$I = \sum_i m_i(\mathbf{R} + \mathbf{r}_i)^2 = R^2 \sum_i m_i + \sum_i m_i r_i^2 + 2\mathbf{R} \cdot \sum_i m_i \mathbf{r}_i \tag{2}$$

由于质心轴经过刚体的质心，上式最后一项中的求和为零（式 28 [931]），而右边第二项恰好是 I_0，右边第一项中 $\sum_i m_i = M$，立即可得式 1．证毕．

垂直轴定理

若我们要求一个刚体薄片关于一条与其垂直的轴（称为**垂直轴**）的转动惯量 I，我们可以在薄片上取两个互相垂直且与垂直轴相交的轴并分别计算薄片

关于这两条轴的转动惯量 I_x 和 I_y．这样就有

$$I = I_x + I_y \tag{3}$$

要证明该式，我们建立空间直角坐标系，令垂直轴与 z 轴重合，另外两条轴分别与 x 轴和 y 轴重合．把刚体看做质点系，令质点 m_i 的坐标为 $(x_i, y_i, 0)$

$$I = \sum_i m_i(x_i^2 + y_i^2) = \sum_i m_i x_i^2 + \sum_i m_i y_i^2 = I_x + I_y \tag{4}$$

证毕．

常见几何体的转动惯量

预备知识 转动惯量[962]

细圆环 薄圆柱环

细圆环和薄圆柱环的所有质量与转轴的距离都为 R，可以看成许多质点的叠加，每个质点转惯量为 $m_i R^2$，所以

$$I = \sum_i m_i R^2 = MR^2 \tag{1}$$

细棒（端点轴）

细棒的线密度为 $\lambda = M/L$，如果划分成长度为 Δr 的小段，第 i 段距离转轴 r_i，有

$$I = \lim_{\Delta r \to 0} \sum_i \lambda \Delta r \cdot r_i^2 = \int_0^L \lambda r^2 \,\mathrm{d}r = \frac{1}{3}\lambda L^3 = \frac{1}{3}ML^2 \tag{2}$$

细棒（中心轴）

细棒（中心轴）可以看做两个等质量的细棒（端点轴），质量都为 M_1，每个具有转动惯量（式 2 ）$M_1R^2/3$，乘以二得总转动惯量为

$$I = \frac{1}{3}MR^2 = \frac{1}{12}ML^2 \tag{3}$$

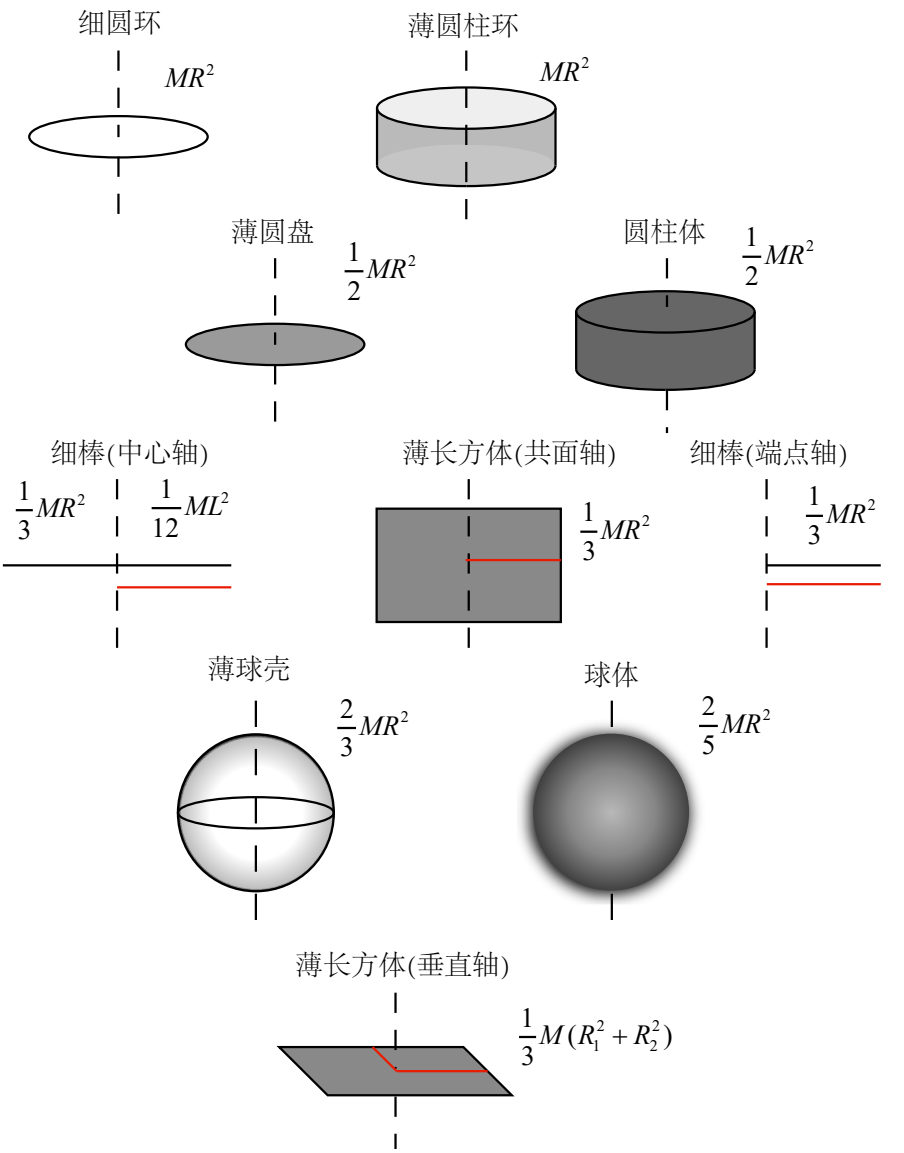


图 1: 常见几何体的转动惯量，虚线为转轴，物体质量 M 均匀分布，R 为几何体的半径或红线标注的长度.

其中 $L=2R$. 由此可以看出，**若一个物体可以拆分成转动惯量相同的若干部分，那么转动惯量公式不变.**

薄长方体（共面轴）

薄长方体（共面轴）可以看成许多细棒（中心轴）组成，所以转动惯量的系数仍然为

$$I = \frac{1}{3}MR^2 = \frac{1}{12}ML^2 \tag{4}$$

薄圆盘 圆柱

薄圆盘可以看做许多宽度为 Δr 的细圆环组成[9]，质量面密度为 $\sigma = M/(\pi R^2)$，第 i 个圆环的半径为 r_i，面积为 $2\pi r_i \Delta r$，总转动惯量为

$$I = \int_0^R r^2 \,\mathrm{d}m = \int_0^R \sigma \cdot 2\pi r \,\mathrm{d}r \,\mathrm{d}r^2 = 2\pi\sigma \int_0^R r^3 \,\mathrm{d}r \tag{5}$$

也可以在极坐标极坐标[119] 中直接根据定义写出积分

$$I = \int r^2 \sigma \,\mathrm{d}s = \int_0^{2\pi}\int_0^R \sigma r^2 \cdot r \,\mathrm{d}r \,\mathrm{d}\theta = 2\pi\sigma \int_0^R r^3 \,\mathrm{d}r = \frac{1}{2}\sigma\pi R^2 R^2 = \frac{1}{2}MR^2 \tag{6}$$

圆柱可看做由许多相同的薄圆盘组成，转动惯量系数相同.

薄球壳

球壳可以看做由许多细圆环组成，质量面密度为 $\sigma = M/(4\pi R^2)$，球坐标中，令第 i 个圆环对应的极角为 θ，宽度为 $R\,\mathrm{d}\theta$，面积为 $\mathrm{d}s_i = 2\pi R\sin\theta \cdot R\,\mathrm{d}\theta$，半径为 $r_\perp = R\sin\theta$，总转动惯量为

$$\begin{aligned} I &= \int_0^\pi r_\perp^2 \,\mathrm{d}m = \int R^2 \sin^2\theta \cdot \sigma \cdot 2\pi R \sin\theta \cdot R \,\mathrm{d}\theta \\ &= 2\pi\sigma R^4 \int \sin^3\theta_i \,\mathrm{d}\theta = 2\pi\sigma R^4 \int_0^\pi \sin^3\theta \,\mathrm{d}\theta \end{aligned} \tag{7}$$

也可以在球坐标中直接写出球面积分

$$\begin{aligned} I &= \int r^2\sigma \,\mathrm{d}s = \int_0^{2\pi}\int_0^\pi (R\sin\theta)^2 \sigma R^2 \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\phi = 2\pi\sigma R^4 \int_0^\pi \sin^3\theta \,\mathrm{d}\theta \\ &= 2\pi\sigma R^4 \int_{-1}^1 (1-\cos^2\theta) \,\mathrm{d}\cos\theta = \frac{2}{3}(\sigma 4\pi R^2)R^2 = \frac{2}{3}MR^2 \end{aligned} \tag{8}$$

其中对 θ 的积分使用了换元积分法.

[9] 然而薄圆盘不能看做由许多过圆心的细棒组成，因为这样面密度就是不均匀的. 另外注意每个细环的转动惯量并不相同（因为半径各不相同），所以不能直接用圆环的转动惯量公式.

球体

球体可以看做由许多薄球壳组成，体密度为 $\rho = M/(4\pi R^3/3)$，令第 i 个球壳半径为 r，厚度为 $\mathrm{d}r$，体积为 $4\pi r^2\,\mathrm{d}r$，总转动惯量为

$$I = \int_0^R \frac{2}{3}r^2\,\mathrm{d}m = \frac{2}{3}\rho\int_0^R r^2\,\mathrm{d}V = \frac{2M}{R^3}\int_0^R r^4\,\mathrm{d}r = \frac{2}{5}MR^2 \tag{9}$$

也可以在球坐标中直接体积分

$$\begin{aligned} I &= \int (r\sin\theta)^2\,\mathrm{d}m = \int_0^{2\pi}\int_0^{\pi}\int_0^R (r\sin\theta)^2\sigma r^2\sin\theta\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi \\ &= \frac{3M}{2R^3}\int_0^{\pi}\sin^3\theta\,\mathrm{d}\theta\int_0^R r^4\,\mathrm{d}r = \frac{2M}{R^3}\int_0^R r^4\,\mathrm{d}r = \frac{2}{5}MR^2 \end{aligned} \tag{10}$$

其中对 θ 的积分使用了换元积分法.

薄长方体（垂直轴）

由“薄长方体（共面轴）”可知两个共面方向的转动惯量分别为（式 4 ）$MR_1^2/3$ 和 $MR_2^2/3$，使用垂直轴定理式 3 [966] 可得关于垂直轴的转动惯量为二者之和

$$I = \frac{1}{3}M(R_1^2 + R_2^2) = \frac{1}{12}M(L_1^2 + L_2^2) \tag{11}$$

其中 $L_1 = 2R_1$，$L_2 = 2R_2$ 分别是两条边长.

长方体

另外由于长方体可以看作许多薄片延轴方向叠加，其转动惯量公式也相同

$$I = \frac{1}{3}M(R_1^2 + R_2^2) = \frac{1}{12}M(L_1^2 + L_2^2) \tag{12}$$

其中 $L_1 = 2R_1$，$L_2 = 2R_2$ 分别是长方体垂直于转轴的两条边长.

刚体的平面运动方程

预备知识 动量定理[944]，转动惯量[962]

任意惯性系中，若刚体质量为 M，质心为 $\mathbf{r}_c$，刚体受若干个力 $\mathbf{F}_i$，作用点分别为 $\mathbf{r}_i$，若刚体**只延一个固定的方向转动**（如刚体的二维运动），且该方向关于质心的转动惯量为 I，则质心运动方程和绕质心转动的方程分别为

$$M\mathbf{a}_c = \sum_i \mathbf{F}_i \tag{1}$$

$$I\boldsymbol{\alpha} = \sum_i \boldsymbol{\tau}_i = \sum_i (\mathbf{r}_i - \mathbf{r}_c) \times \mathbf{F}_i \tag{2}$$

其中 $\mathbf{a}_c$ 是质心的加速度，$\boldsymbol{\alpha}$ 是绕质心转动的角加速度．这是说，我们可以把刚体的运动分解成质心的移动和相对质心的转动，并用合力计算前者，用关于质心的合力矩计算后者．

推导

我们把刚体看做质点系来证明，已知刚体总动量等于质心动量[939] $\mathbf{p}_c$．在任意惯性系中，由动量定理[944]，动量变化率为

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}_c = M\mathbf{a}_c = \sum_i \mathbf{F}_i \tag{3}$$

现在我们用角动量定理证明式 2．由于质心与刚体的相对位置不变，质心系中刚体必须绕质心转动，且角动量为（式 8 [962]）$\mathbf{L}_c = I\boldsymbol{\omega}$，角动量变化率为[10]

$$\frac{\mathrm{d}\mathbf{L}_c}{\mathrm{d}t} = I\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} = I\boldsymbol{\alpha} \tag{4}$$

要特别注意的是，除非合力为零，质心系并不是惯性系，所以使用角动量定理要考虑刚体的惯性力．但幸运的是质心系中惯性力（式 5 [920]）$-m_i\mathbf{a}_c$ 产生的合力矩为零[931]

$$\sum_i \mathbf{r}_{ci} \times (-m_i\mathbf{a}_c) = \mathbf{a}_c \times \sum_i m_i\mathbf{r}_{ci} = \mathbf{0} \tag{5}$$

[10]注意第一步成立的条件是 I 不变，而一般情况下 I 与刚体的转轴有关，所以只能假设刚体延同一方向转动．唯一的例外是物体的转动惯量与方向无关的情况，例如球体．刚体的变向转动较为复杂，不做讨论．

现在我们可以继续角动量定理[954] 得

$$I\boldsymbol{\alpha} = \sum_i \mathbf{r}_{ci} \times \mathbf{F}_i \tag{6}$$

由于质心系相对于任何惯性系没有相对转动，所以在任意惯性系中刚体的角加速度仍然为 $\boldsymbol{\alpha}$．但受力点的位矢变为 $\mathbf{r}_i = \mathbf{r}_c + \mathbf{r}_{ci}$，即

$$I\boldsymbol{\alpha} = \sum_i (\mathbf{r}_i - \mathbf{r}_c) \times \mathbf{F}_i \tag{7}$$

例 1　球体或圆柱延斜面无摩擦滚动

如图 1，在一个倾角为 θ 的斜面上，一个半径为 R 质量为 M 的均匀的球体（或圆柱）从静止开始向下滚动，其转动惯量为 I，求质心的加速度和滚动的角加速度．

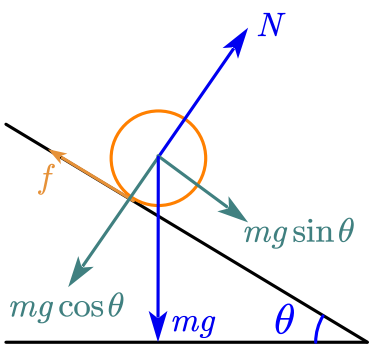


图 1: 圆柱延斜面无摩擦滚动

解：首先，根据无摩擦的条件，系统只有一个自由度[930]，即圆柱的位移或者转角，二者的关系为

$$s = R\theta \tag{8}$$

两边求二阶导数，得加速度与角加速度的关系为

$$a = R\alpha \tag{9}$$

受力分析如图，圆柱受三个力：重力，支持力和静摩擦力．将物体受到的重力分解为垂直斜面方向和沿斜面方向的两个分力．其中垂直方向的分力与斜面提供的支持力抵消，而平行方向的分量和摩擦力的合力决定质心的加速度式 1

$$mg\sin\theta - f = Ma \tag{10}$$

再来分析关于质心的转动，由于重力和摩擦力都在质心，所以对合力矩贡献为 0．所以只有摩擦力的贡献为 $\tau = fR$，由式 2 得

$$I\alpha = fR \tag{11}$$

由式 9 到式 11 三式解得加速度为

$$a = \frac{g\sin\theta}{1 + I/(MR^2)} \tag{12}$$

可见转动惯量为 0 时，结果与无摩擦滚动一致．而转动惯量越大，加速度就越小．

例 2

一根质量为 M 长为 L 的均匀细棒延 y 方向静止放置在水平面 xy，从 $t = 0$ 时起在其上端施加一个 x 方向的恒力，描述细棒如何运动．如果木棒与地面的摩擦系数为 μ，答案又如何？

首先考虑质心的运动，细棒所受外力只有一个恒力，所以由式 1 质心沿 x 方向做匀加速运动．再来看质心系中细棒的转动由“转动惯量[962]”中例 1 可知细棒绕其质心做单摆运动．

惯性张量

- 本词条存在未完成的内容．

预备知识 常见几何体的转动惯量[967]，张量[488]，瞬时转轴[977]

[11]我们已经初步学习了刚体绕固定轴的转动，现在我们做一个更一般的讨论，也就是刚体绕某个固定点的转动．刚体绕固定轴转动只有一个自由度 θ，而绕固定点转动却可以有 3 个自由度（见“刚体[939]”）．

惯性张量

刚体绕固定点旋转的角动量等于**惯性张量 I** 乘以瞬时角速度矢量 $\boldsymbol{\omega}$

$$\mathbf{L} = \mathbf{I}\boldsymbol{\omega} \tag{1}$$

[11]参考 Wikipedia 相关页面．

惯性张量是一个 3 维方阵，其阵元一般记为

$$\mathbf{I}=\begin{pmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{pmatrix} \tag{2}$$

我们现在假设刚体绕坐标原点旋转，如果将 x,y,z 分别记为 x_1,x_2,x_3，则 $\mathbf{I}$ 的矩阵元可以通过体积分计算

$$I_{ij}=\delta_{i,j}\int r^2\rho(\mathbf{r})\,\mathrm{d}V-\int x_ix_j\rho(\mathbf{r})\,\mathrm{d}V \qquad (i,j=1,2,3) \tag{3}$$

其中 $\mathbf{r}$ 是刚体上某点的位置矢量[878]，$r=|\mathbf{r}|$，$\rho(\mathbf{r})$ 是位置 $\mathbf{r}$ 处的密度．也可以表示成矩阵积分的形式，即对每个矩阵元分别积分

$$\mathbf{I}=\int\rho(\mathbf{r})\begin{pmatrix} y^2+z^2 & -xy & -xz \\ -yx & x^2+z^2 & -yz \\ -zx & -zy & x^2+y^2 \end{pmatrix}\mathrm{d}V \tag{4}$$

可以看出，惯性张量与坐标系的选取有关，而这又可以划分为原点的位置和坐标轴的取向．

例 1　长方体的惯性张量

令重心在原点，密度为常数，质量为 M 的长方体延 x,y,z 方向的长度分别为 a,b,c，那么有

$$\begin{aligned} &\iiint xy\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z=\iiint yz\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z=\iiint zx\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z=0 \\ &\iiint x^2\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z=\frac{1}{12}a^3bc \\ &\iiint y^2\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z=\frac{1}{12}ab^3c \\ &\iiint z^2\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z=\frac{1}{12}abc^3 \end{aligned} \tag{5}$$

使用 $M=\rho abc$ 可得

$$\mathbf{I}=\frac{1}{12}M\begin{pmatrix} b^2+c^2 & 0 & 0 \\ 0 & a^2+c^2 & 0 \\ 0 & 0 & a^2+b^2 \end{pmatrix} \tag{6}$$

可以发现这个是一个对角矩阵. 这就意味着, 如果长方体分别绕 x, y, z 轴转动, 它的角动量也分别延 x, y, z 轴. 由式 11 [967] 可以验证三个方向的转动惯量是对的. 当转轴与坐标轴不重合时, 可见转动惯量的方向与转动方向一般不同.

特殊地, 当 $a = b = c$ （正方体）时, 惯性张量变为单位矩阵乘以常数

$$\mathbf{I} = \frac{1}{6} M a^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{7}$$

由于单位矩阵乘以任何矢量都得到相同的矢量, 正方体延任意转轴转动时, 角动量都和角速度矢量成正比

$$\mathbf{L} = \mathbf{I}\boldsymbol{\omega} = \frac{1}{6} M a^2 \boldsymbol{\omega} \tag{8}$$

这可以验证式 12 [967] 的结果.

例 2　细杆的角动量

假设细杆沿 x 轴放置, 长度为 L. 要计算它的惯性张量, 我们可以从头积分（习题 1）, 也可以假设细杆是一个细长的长方体, 直接用例 1 的结论（令 $a = L, b \to 0, c \to 0$）得

$$\mathbf{I} = \frac{1}{12} M \begin{pmatrix} 0 & 0 & 0 \\ 0 & L^2 & 0 \\ 0 & 0 & L^2 \end{pmatrix} \tag{9}$$

可以看到, 如果转轴延 x 方向, 即 $\boldsymbol{\omega} = (\omega, 0, 0)^{\mathrm{T}}$, 角动量为零, 如果延 y, z 方向, 则角动量大小都等于

$$L = \frac{1}{12} M L^2 \omega \tag{10}$$

对应的转动惯量为 $I = ML^2/12$, 这与式 3 [967] 一致.

习题 1

直接使用积分计算细杆的角动量（式 9）.

推导

我们把刚体看作质点系刚体[939]，刚体绕固定点转动时，刚体上的每个点都在做瞬时圆周运动. 根据角动量的定义式 1 [952] 以及圆周运动的速度式 5 [883]

$$\mathbf{L}=\sum_i m_i\mathbf{r}_i\mathbf{v}_i=\sum_i m_i\mathbf{r}_i\times(\boldsymbol{\omega}\times\mathbf{r}_i)=\sum_i m_i r_i^2\boldsymbol{\omega}-\sum_i m_i(\boldsymbol{\omega}\cdot\mathbf{r}_i)\mathbf{r}_i \tag{11}$$

其中使用了“连续叉乘的化简[392]”中的式 1. 将上式右边写成分量的形式，并将求和表示为刚体密度 $\rho(\mathbf{r})$ 的积分得

$$\begin{aligned}\begin{pmatrix}L_x\\L_y\\L_z\end{pmatrix}&=\int\rho r^2\begin{pmatrix}\omega_x\\\omega_y\\\omega_z\end{pmatrix}\mathrm{d}V-\int\rho\begin{pmatrix}xx&xy&xz\\yx&yy&yz\\zx&zy&zz\end{pmatrix}\begin{pmatrix}\omega_x\\\omega_y\\\omega_z\end{pmatrix}\mathrm{d}V\\&=\left[\int\rho\begin{pmatrix}r^2-xx&-xy&-xz\\-yx&r^2-yy&-yz\\-zx&-zy&r^2-zz\end{pmatrix}\mathrm{d}V\right]\begin{pmatrix}\omega_x\\\omega_y\\\omega_z\end{pmatrix}\end{aligned} \tag{12}$$

对比式 1 可知，方括号中的积分就是惯性张量 $\mathbf{I}$（式 4）.

未完成：以下内容开新词条

坐标旋转变换

预备知识 空间旋转矩阵[404]

上文中提到惯性张量矩阵与坐标系的选取有关，一般情况下当刚体相对于选取的参考系发生平移或任意方向的转动，惯性张量都会发生改变. 我们先来讨论如果惯性张量在两个原点相同朝向不同的坐标系之间如何变换，这叫做惯性张量的**旋转变换**. 先建立一个与刚体始终相对静止的参考系叫做**体坐标系（body frame）**，一般我们选择体坐标系的原点在刚体质心处，但刚体受力运动时，体坐标系并不是一个惯性系，所以我们还要选一个惯性系，即**实验室坐标系（lab frame）**. 令实验室系矢量到体坐标系矢量的变换矩阵[404] 为 $\mathbf{R}$，记体坐标系和实验室系中的惯性张量分别为 $\mathbf{I}_0$ 和 $\mathbf{I}$，则二者的关系为

$$\mathbf{I}=\mathbf{R}\mathbf{I}_0\mathbf{R}^{-1}=\mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}} \tag{13}$$

其中使用了酉矩阵的性质 $\mathbf{R}^{-1}=\mathbf{R}^{\mathrm{T}}$（式 4 [440]）. 注意该式是一个相似变换[429]，所以 $\mathbf{I}$ 和 $\mathbf{I}_0$ 互为相似矩阵.

证明：要计算 $\mathbf{L}=\mathbf{I}\boldsymbol{\omega}$，把式 13 右边三个矩阵从右到左依次作用在 $\boldsymbol{\omega}$ 上. 先计算 $\mathbf{R}^{\mathrm{T}}\boldsymbol{\omega}$ 把实验室系中的角速度矢量 $\boldsymbol{\omega}$ 变换到体坐标系中的 $\boldsymbol{\omega}_0$. 再计算 $\mathbf{I}_0\boldsymbol{\omega}_0$ 得到体坐标系中的角动量 $\mathbf{L}_0$，最后计算 $\mathbf{R}\mathbf{L}_0$ 得到实验室系中的角动量 $\mathbf{L}$. 证毕.

例 3

如果转轴和轻杆法线的夹角为 θ，（未完成）

坐标平移变换

（未完成）

刚体的瞬时转轴

预备知识 刚体[939]

当刚体绕某个固定点做任意转动时，我们在每个时刻仍然能找到一个经过该点的**瞬时转轴**以及延转轴的瞬时角速度（矢量）$\boldsymbol{\omega}$.

某时刻瞬时转轴的定义是：该时刻刚体在内任意落在转轴上的点速度为零. 事实上，我们只需要在某时刻找到刚体上瞬时速度为 0 的任意两个不同点，就可以过这两点作出瞬时转轴，并确保该直线上的所有点瞬时速度都为 0.

例 1　进动陀螺的瞬时转轴

在陀螺进动的例子（例 1 [954]）中，我们可能会认为陀螺的瞬时转轴就是陀螺的轴. 但陀螺的轴时时刻刻都在运动，除了与地面接触的点外，任意一点的瞬时速度都不为 0. 所以此时陀螺的轴并不是瞬时转轴.

注意陀螺的自转和进动的方向相同（都是顺时针或逆时针），我们可以知道真正的瞬时转轴同样经过与地面的接触点，但会在陀螺轴的上方. 我们只需要找到陀螺圆盘表面上瞬时速度为 0 的一点即可求出瞬时转轴的倾角.

刚体定轴转动 2

- 本词条处于草稿阶段.

预备知识 惯性张量[973]

转轴提供的力矩

之前我们在讨论刚体绕固定轴转动时只涉及了角动量和力矩延轴方向的分量 L_z，τ_z. 在学习了惯性张量以后我们就可以开始讨论垂直轴方向的角动量和力矩了.

我们假设刚体与转轴之间无摩擦，那么转轴就不可能给杆提供延轴方向的力矩，但是却可以提供垂直轴方向的力矩，这里记为 $\boldsymbol{\tau}_a$. 我们把所有外力对刚体的力矩记为 $\boldsymbol{\tau}$，于是根据角动量定理有

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = \boldsymbol{\tau} + \boldsymbol{\tau}_a \tag{1}$$

之前我们看到，刚体的运动完全是由外力矩的 z 分量（延轴方向分量）τ_z 决定的，所以我们可以断定 $\boldsymbol{\tau}$ 的任何垂直分量都被 $\boldsymbol{\tau}_a$ 中的“反作用力矩”抵消了[12]. 所以不失一般性，我们接下来假设 $\boldsymbol{\tau}$ 与转轴共线.

再来看刚体的角动量（注意不仅是 z 分量）. 在学惯性张量时我们看到，一般情况下角动量 $\mathbf{L}$ 和角速度 $\boldsymbol{\omega}$ 是不共线的. 我们来举一个例子.

========== 未完成 ==========

例如把一个质量均匀分布的长杆倾斜固定在转轴上匀速旋转，它绕轴旋转时的角动量始终垂直于长杆且随之一起转动. 这时运用角动量定理就会发现转轴必须给杆提供一个垂直于转轴的力矩.

刚体的动能定理

[12] 做一个类比，假设质点只能延光滑的轨道运动（轨道可以是弯曲的），那么只有外力延轨道的方向的分量能影响质点的运动，垂直轨道方向的分量被轨道对质点的反作用力抵消了. 注意轨道给质点的力除了这个反作用力外，还有一部分用于提供质点在垂直轨道方向的加速度分量.

预备知识 刚体的平面运动方程[970]，柯尼希定理[946]，轻杆模型

如果我们把刚体[939]看作质点系，那么系统中任意两个质点间距离保持不变，我们可以假设这些质点之间以不可伸长的轻杆相连，使它们不能相对运动.

平动动能与转动动能

预备知识 柯尼希定理[946]

由柯尼希定理，刚体的动能可以分为**平动动能**和**转动动能**两部分．下面来证明一段时间内，平动动能的增加等于合外力关于质心位移的做功，而转动动能的增加等于质心系中合外力矩的做功.（未完成!）

刚体的运动方程

预备知识 刚体的平面运动方程[970]，惯性张量[973]

一般情况下下刚体的运动方程要比平面运动复杂许多，但我们仍然可以将运动分解为质心的运动以及刚体绕质心的旋转，前者由合力决定，所以仍然有（式 1 [970]）

$$M\mathbf{a}_c = \sum_i \mathbf{F}_i \tag{1}$$

所以相对于平面运动，该问题的困难在于绕质心转动的计算．虽然角动量定理仍然满足，但转动惯量将有可能随时间变化．这是因为刚体的瞬时转轴相对刚体的位置可能会随时间变化，而不同的转轴对应的角动量一般不同.

下面我们会看到，刚体绕固定点转动的角动量定理可以记为

$$\boldsymbol{\tau} = \mathbf{I}\boldsymbol{\alpha} + \boldsymbol{\omega} \times \mathbf{L} \tag{2}$$

对比平面运动的式 2 [970]，转动惯量变为了惯性张量，且多了一项角速度叉乘角动量．当二者共线时，叉乘为零，就回到了平面运动的式子.

转动方程

我们仍然可以用角动量定理来推导刚体的转动方程，但这里的角动量要用惯性张量来表示（式 1 [973] 和式 6 [973]）

$$\mathbf{L}=\mathbf{I}\boldsymbol{\omega}=\mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}\boldsymbol{\omega} \tag{3}$$

其中 $\mathbf{I}_0$ 不随时间变化，$\mathbf{L}$，$\boldsymbol{\omega}$ 和 $\mathbf{R}$ 都是时间的函数. 代入角动量定理（式 1 [954]）得

$$\boldsymbol{\tau}=\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t}=\frac{\mathrm{d}\mathbf{I}}{\mathrm{d}t}\boldsymbol{\omega}+\mathbf{I}\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t}=\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}\right)\boldsymbol{\omega}+\mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \tag{4}$$

其中 $\mathrm{d}\boldsymbol{\omega}/\mathrm{d}t$ 是角加速度，记为 $\boldsymbol{\alpha}$，角速度和角加速度的关系可以类比速度和加速度[881]. 另外注意这里对矩阵求导就是对每个元分别求导.

我们把力矩 $\boldsymbol{\tau}$ 看作是一个关于时间的已知函数，把旋转矩阵 $\mathbf{R}$ 和角速度 $\boldsymbol{\omega}$ 看做关于时间的未知函数（即微分方程的解）. $\mathbf{R}$ 和 $\boldsymbol{\omega}$ 完整描述了刚体绕固定点转动的**状态**，就像位置和动量可以完整描述了一个质点运动的状态.

另外，$\boldsymbol{\omega}$ 和 $\mathbf{R}$ 之间的关系就速度和位移的关系，假设体坐标系中固定在刚体上的任意一点坐标为 $\mathbf{r}$ （不随时间变化），变换到实验室坐标系中为 $\mathbf{R}\mathbf{r}$. 对时间求导得该点在实验室坐标系的速度为

$$\mathbf{v}=\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t}\mathbf{r} \tag{5}$$

而角速度和速度之间有 $\mathbf{v}=\boldsymbol{\omega}\times(\mathbf{R}\mathbf{r})$（式 5 [883]）. 我们可以把叉乘用矩阵乘法表示为

$$\mathbf{v}=\boldsymbol{\Omega}\mathbf{R}\mathbf{r} \tag{6}$$

其中

$$\boldsymbol{\Omega}=\begin{pmatrix}0 & -\omega_z & \omega_y\\ \omega_z & 0 & -\omega_x\\ -\omega_y & \omega_x & 0\end{pmatrix} \tag{7}$$

是一个反对称矩阵，即

$$\boldsymbol{\Omega}^{\mathrm{T}}=-\boldsymbol{\Omega} \tag{8}$$

由于 $\mathbf{r}$ 是任意的，对比式 5 和式 6 得

$$\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t}=\boldsymbol{\Omega}\mathbf{R} \tag{9}$$

现在我们可以化简式 4 右边第一项，根据链式法则[174] 和式 9，式 8

$$\begin{aligned}\frac{\mathrm{d}\mathbf{I}}{\mathrm{d}t}\boldsymbol{\omega} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}\right)\boldsymbol{\omega} \\ &= \frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}\boldsymbol{\omega} + \mathbf{R}\mathbf{I}_0\left(\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t}\right)^{\mathrm{T}}\boldsymbol{\omega} \\ &= \boldsymbol{\Omega}\mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}\boldsymbol{\omega} + \mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}\boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{\omega} \\ &= \boldsymbol{\Omega}\mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}\boldsymbol{\omega} - \mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\omega} \\ &= \boldsymbol{\Omega}\mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}\boldsymbol{\omega} \\ &= \boldsymbol{\omega}\times\mathbf{L}\end{aligned} \tag{10}$$

其中使用了 $\boldsymbol{\Omega}\boldsymbol{\omega} = \boldsymbol{\omega}\times\boldsymbol{\omega} = \mathbf{0}$．所以式 4 变为

$$\begin{aligned}\boldsymbol{\tau} &= \boldsymbol{\omega}\times\mathbf{L} + \mathbf{I}\boldsymbol{\alpha} \\ &= \boldsymbol{\Omega}\mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}\boldsymbol{\omega} + \mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t}\end{aligned} \tag{11}$$

整理得（注意 $\mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}$ 得逆矩阵是 $\mathbf{R}\mathbf{I}_0^{-1}\mathbf{R}^{\mathrm{T}}$，$\mathbf{I}_0^{-1}$ 是 $\mathbf{I}_0$ 的逆矩阵）

$$\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} = \mathbf{R}\mathbf{I}_0^{-1}\mathbf{R}^{\mathrm{T}}\left(\boldsymbol{\tau} - \boldsymbol{\Omega}\mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}\boldsymbol{\omega}\right) \tag{12}$$

式 9 和式 12 就是完整的绕质心转动方程．这是一个一阶常微分方程组[333]，写成标量的形式共有 12 条，未知数分别为 $\omega_x, \omega_y, \omega_z$，$R_{i,j}$ 共 12 个．

事实上旋转矩阵 $\mathbf{R}$ 其实只有三个独立的自由度，如果我们能用三个变量表示 $\mathbf{R}$，就可以得到只含 6 个未知数的 6 个方程．一种方法是使用欧拉角，但列出来后式子会比较复杂．另一种方法是用 4 元数，即用 4 个变量表示 $\mathbf{R}$，可以得到形式相对简单的方程，见“刚体运动方程（四元数）[981]”.

刚体运动方程（四元数）

- 本词条需要更多讲解，便于帮助理解.

预备知识 刚体的运动方程[979]，四元数与旋转矩阵[419]

我们用四元数 $\mathbf{q}$ 和角速度 $\boldsymbol{\omega}$ （共 7 个标量）来表示刚体绕固定点旋转的运动状态．下面来列运动方程（7 元一阶微分方程组）．

式 10 [419] 中已经给出了 4 条（$\mathbf{q}$ 关于时间的导数）

$$\frac{\mathrm{d}\mathbf{q}}{\mathrm{d}t} = \frac{1}{2}[0, \boldsymbol{\omega}]\mathbf{q} \tag{1}$$

而之前的式 12 [979] 给出了另外 3 条（$\boldsymbol{\omega}$ 关于时间的导数）

$$\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} = \mathbf{R}\mathbf{I}_0^{-1}\mathbf{R}^{\mathrm{T}}\left(\boldsymbol{\tau} - \boldsymbol{\Omega}\mathbf{R}\mathbf{I}_0\mathbf{R}^{\mathrm{T}}\boldsymbol{\omega}\right) \tag{2}$$

其中 $\mathbf{I}_0$ 是体坐标系中的惯性张量，$\mathbf{R}$ 是体坐标系到实验室坐标系的旋转矩阵，$\boldsymbol{\tau}$ 是力矩（已知）．注意现在我们可以用四元数 $\mathbf{q}$ 表示 $\mathbf{R}$ （式 6 [419]）．这样就得到了所有的运动方程．

用数值计算来解这个方程见“刚体转动数值模拟[1894]”．

第四章

软体和流体力学

绳的法向压力

预备知识 定积分[198]

我们知道当一根质量忽略不计的绳子紧绷时，如果它不受任何法相的力（即垂直于绳的力），它将保持一条直线．如果绳的一部分存在弯曲，则说明该部分受法向力．如图 1，水平放置的绳子的张力为 T，经过半径为 R 的滑轮后弯曲了 θ 角．由于我们假设滑轮与绳之间无摩擦[1]，经过滑轮后绳的张力仍然为 T．

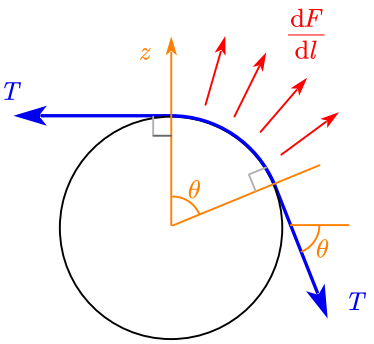


图 1: 滑轮对绳的线压强

显然，绳弯曲是由于在于滑轮接触的部分处处都受到法向的力．为了描述这种力的大小，我们取长度为 Δl 的一小段绳（弯曲忽略不计），若它受法向力大小为 ΔF，就用 $\Delta F/\Delta l$ 来描述法向力的大小．当 $\Delta l \to 0$，就记为 $\mathrm{d}F/\mathrm{d}l$，称为**线压强**，是一个标量（压强是单位面积的压力，线压强是单位长度的压力）．注意力是相互的，滑轮对绳的压力与绳对滑轮的压力处处等大反向，所以总压力也等大反向．

绳子每点处的线压强与什么有关呢？由直觉我们可以猜到它和绳的张力成正比，和滑轮的半径有关，但与 θ 无关．我们先假设接触的部分先压强处处相等，对绳做受力分析．

绳能保持静止，说明它受到的合力为 0，即两端的拉力以及滑轮对它的力相加为 0．若我们只考所有力的 z 分量，可知滑轮对绳向上的分离等于绳右端

[1]等效地，也可以假设滑轮可以无摩擦滚动

拉力的 z 分量，大小为 $T\sin\theta$.

滑轮对绳合的 z 分量可以用定积分计算

$$F_z=\int_0^\theta \cos\theta'\frac{\mathrm{d}F}{\mathrm{d}l}R\,\mathrm{d}\theta'=\frac{\mathrm{d}F}{\mathrm{d}l}R\int_0^\theta\cos\theta'\,\mathrm{d}\theta'=\frac{\mathrm{d}F}{\mathrm{d}l}R\sin\theta \tag{1}$$

所以有

$$\frac{\mathrm{d}F}{\mathrm{d}l}R\sin\theta=T\sin\theta \tag{2}$$

即

$$\frac{\mathrm{d}F}{\mathrm{d}l}=\frac{T}{R} \tag{3}$$

所以线压强与张力成正比，和曲率半径成反比．这符合我们上面的猜测．

我们可以验证水平方向也有受力平衡，滑轮对绳水平方向的合力为（向右为正）

$$\int_0^\theta \sin\theta'\frac{\mathrm{d}F}{\mathrm{d}l}R\,\mathrm{d}\theta'=\frac{\mathrm{d}F}{\mathrm{d}l}R\int_0^\theta\sin\theta'\,\mathrm{d}\theta'=\frac{\mathrm{d}F}{\mathrm{d}l}R(1-\cos\theta)=T(1-\cos\theta) \tag{4}$$

绳子两端的拉力在水平方向的合力为 $-T+T\cos\theta$，所以水平合力同样为 0.

对于任意弯曲且光滑的绳子，要计算它任意一点受到的线压强，我们只需要知道该点处的张力与曲率半径 ρ 即可，即

$$\frac{\mathrm{d}F}{\mathrm{d}l}=\frac{T}{\rho} \tag{5}$$

例 1　磁场中线圈的张力

一个垂直纸面向外的匀强磁场中，强度为 B．有一个平行纸面放置的柔软单股线圈．当线圈中通有逆时针电流 I 时，线圈就会受到向外的安培力并紧绷成一个圆形．求线圈的张力．

解：磁场对线圈的线压强在这里由安培力充当，我们可以首先把线圈划分为许多小段，每段长度为 Δl，且近似为直线．则每段所受安培力为

$$\Delta F=IB\Delta l \tag{6}$$

所以线压强为 IB．由式 3 得张力为 $T=RIB$.

杨氏模量

- 本词条处于草稿阶段.

[2]**杨氏模量（Young's modulus）**也叫**杨氏模数**或者**弹性模量**，是衡量材料弹性的力学性质．其定义为

$$E = \frac{FL_0}{A\Delta L} \tag{1}$$

其中 F 是受力，L_0 是原长，A 是横截面积，ΔL 是伸长或压缩的长度．杨氏模量具有压强的量纲（国际单位：帕斯卡[1970]）通常我们假设 $\Delta L \ll L_0$，因为和弹簧一样，过大的形变会导致非线性效应.

例 1

一根横截面直径为 $2r = 2\ \mathrm{mm}$ 的钢丝，松弛长度为 $L_0 = 1\ \mathrm{m}$，已知该钢丝的杨氏模量为 $E = 200\ \mathrm{GPa}$，要将其拉长 $\Delta L = 1\ \mathrm{mm}$ 需要在两端施加多大的力？

解：横截面为 $A = \pi r^2$，将各个量代入式 1 解的张力为 $F = 628.3\ \mathrm{N}$.

流 流密度

预备知识 矢量场[294]

流，流量

单位时间流经某个截面的某种物理量叫**流（current）**．最常见的例子是电流[1123]，即单位时间流经某截面的电荷量，用极限[149] 定义为

$$I = \lim_{\Delta t \to 0} \frac{\Delta q}{\Delta t} \tag{1}$$

注意流的大小一般与所选择截面的位置，方向，面积都有关系．我们也可以把电荷 q 替换成其他物理量如质量、能量、粒子数，分别得到质量流、能流、粒子流.

[2]本文参考维基百科相关页面.

一段时间内流经截面的某种物理量的总量就叫做**流量**，流量是时间 t 的函数，所以由导数[158] 的定义，流是流量关于时间的导数．反之，流量是流在某段时间的定积分[198]．

$$\Delta q = \int_{t_1}^{t_2} I(t)\,\mathrm{d}t \tag{2}$$

流密度

流密度（current density）可以用于描述某时刻流体在的空间流动的速率．我们以水流为例，在一条河流或管道中，某时刻 t 在空间中任意一点 $\mathbf{r}$ 处，都对应一个水流速度 $\mathbf{v}$，如果我们在该点放置一个与速度垂直的微小截面（通常叫**面元**），令其面积为 ΔS，在一段微小时间 Δt 内流经截面的质量为 Δm，那么**质量流密度（mass current density）**可以用极限定义为

$$\mathbf{j}(\mathbf{r},t) = \hat{\mathbf{n}} \lim_{\Delta S,\Delta t\to 0} \frac{\Delta m}{\Delta S \Delta t} \tag{3}$$

其中 $\hat{\mathbf{n}}$ 表示面元正方向法向量或者面元处流体的速度方向．流密度是一个关于位置的矢量函数，即矢量场[294]．式 3 中的质量可以替换为不同的物理量，若替换为能量则称为**能流密度**，若是粒子数则称为**粒子流密度**，若是电荷量则称为**电流密度**[1123]，等等．

我们也可以根据密度和速度来定义流密度

$$\mathbf{j}(\mathbf{r},t) = \rho(\mathbf{r},t)\mathbf{v}(\mathbf{r},t) \tag{4}$$

其中密度定义为（质量同样可以替换成其他物理量）

$$\rho(\mathbf{r},t) = \lim_{\Delta V\to 0} \frac{\Delta m}{\Delta V} \tag{5}$$

式 4 的定义和式 3 是等效的，因为时间 Δt 内流经 ΔS 的体积为 $\Delta V = \Delta S \cdot v\Delta t$，代入式 5 再代入式 4 就得到了式 3．

流密度与流

预备知识 曲面积分通量[297]

我们来讨论如何通过流密度来计算流．通过例 1 [297] 和面积分（通量）的定义（式 1 [297]），易得流密度作为矢量场在某截面上的通量就是该截面的流：

$$I(t) = \int \mathbf{j}(\mathbf{r},t)\,\mathrm{d}\mathbf{s} \tag{6}$$

浮力 阿基米德原理

等效法

[3]我们先用一个简单易懂的方式解释浮力. 假设在重力加速度为 g 的环境中，容器中密度为 ρ_0 的液体完全静止. 这时令液体内部有一任意形状的闭合曲面，体积为 V_0. 把曲面内部的液体作为一个整体做受力分析，其质量为 $m=\rho_0 V_0$，所受重力为 $mg=\rho_0 V_0 g$. 由于曲面中液体保持静止，说明曲面外的液体对曲面内的液体施加了相同大小的浮力. 现在我们如果把曲面内的液体替换为一块密度为 ρ 的物体，由于曲面形状不改变，外界液体对该物体的浮力仍然为

$$F=\rho_0 V_0 g \tag{1}$$

注意 V_0 为物体在水中部分的体积，如果物体只有部分在水中，V_0 将小于物体的体积.

例 1　柱体的浮力

柱体的浮力容易通过简单的压强计算获得. 例如一个边长为 L 的立方体完全浸泡在某种液体中，其四个侧面受到的压强都是沿水平方向的，且互相抵消，对浮力贡献为零. 下表面受到向上的压强为 $P_1=\rho g h_1$ 对浮力的贡献为 $F_1=L^2 P_1$. 同理，上表面受到向下的压强为 $P_2=\rho g h_2$，对浮力的贡献为 $F_2=-L^2 P_2$. 所以总浮力为

$$F=F_1+F_2=\rho g L^2(h_1-h_2)=\rho g L^3=\rho g V \tag{2}$$

其中 ρ 是液体的体积. 对于其他柱体（如圆柱，三棱柱等），若竖直放置，同样可以通过以上方法得到式 1.

散度法

预备知识　牛顿—莱布尼兹公式的高维拓展[314]

[3]本文参考 Wikipedia 相关页面.

现在我们用面积分的方法表示浮力. 令 z 轴竖直向上，且水面处 $z=0$，则水面下压强为

$$P=-\rho_0 g z \tag{3}$$

现在把上述的闭合曲面划分为许多个微面元，第 i 个面元用矢量 $\Delta\mathbf{s}_i$，表示，其中模长为面元的面积，方向为从内向外的法向. 这个面元受到外界液体的压力为

$$\Delta\mathbf{F}_i=-P\Delta\mathbf{s}_i=\rho_0 g z\Delta\mathbf{s}_i \tag{4}$$

现在把所有面元所受的压力求和，并用曲面积分[297] 表示为

$$\mathbf{F}=\oint \rho_0 g z\,\mathrm{d}\mathbf{s} \tag{5}$$

这就是物体所受的浮力. 使用式 1 [314] 得

$$\mathbf{F}=\int \boldsymbol{\nabla}(\rho_0 g z)\,\mathrm{d}V=\rho_0 g V_0\hat{\mathbf{z}} \tag{6}$$

可见该结论与“等效法”中得出的一致.

伯努利方程

- 本词条处于草稿阶段.

[4]不可压缩流体的方程.

$$\frac{v^2}{2}+gz+\frac{p}{\rho}=\text{常数} \tag{1}$$

其中 g 是重力加速度，z 是高度，p 是压强，ρ 是液体的密度（液体不可压缩，是常数）.

推导

如图，一根管子的粗细不同两部分的横截面面积分别为 A_1, A_2，压强分别为 p_1, p_2，速度分别为 v_1, v_2.

[4]参考 Wikipedia 相关页面.

穿过横截面 $i=1,2$ 的功率分别为

$$P_i = p_i A_i v_i + \frac{1}{2}(\rho v_i A_i)v_i^2 + \rho v_i A_i g z_i \tag{2}$$

而且我们要求

$$P_1 = P_2 \tag{3}$$

另外由于液体不可压缩，有

$$A_1 v_1 = A_2 v_2 \tag{4}$$

所以式 3 两边同时除以 $\rho v_i A_i$ 得

$$\frac{v_1^2}{2} + g z_1 + \frac{p_1}{\rho} = \frac{v_2^2}{2} + g z_2 + \frac{p_2}{\rho} \tag{5}$$

未完成：这样的推导如何拓展到开放空间的情况呢？举例：水龙头下的乒乓球，香蕉球，机翼，两张纸中间吹气

黏度

- 本词条处于草稿阶段.

[5]如果无穷大静止水平面上，有一层厚度为 h 的液体，液体的表面覆盖一层无穷大平板，平板匀速运动，使水的上表面以速度 v 移动（未必和平板相对静止），下表面速度为零. 实验证明，平板单位面积受到水的阻力大小 f （具有压强的量纲）可以通过一个简单的公式计算

$$f = \mu \frac{v}{h} \tag{1}$$

其中 μ 是一个和液体的种类有关的常数，被称为**粘度（viscosity）**. 粘度可以理解为液体的摩擦力.

流体运动的描述方法

[5]参考 Wikipedia 相关页面.

- 本词条处于草稿阶段.

预备知识 速度加速度[881]

流体力学中，流体运动的描述方法包括拉格朗日法与欧拉法，两种方法在物理意义和数学表达方面各有特点，下面我们具体解释.

流体与固体的区别

我们从流体与固体的区别引入两种描述方法，固体无论处于静止还是运动都可以通过有限的静变形承受剪切力，其形状不易变化，在运动学中，可以只从几何角度来描述物体的位置随时间的变化.

流体在静止时无法通过有限的静变形承受剪切力，在运动状态下虽能产生剪切力，剪切力却不能维持流体内部各点位置的有序，反而使其产生连续不断的变形，流体内部各质点位置变化很大，单从几何角度描述其运动将复杂而困难，需要专门的处理方法.

固体力学中，只着眼于需要描述的物体，物体之外的都叫做环境，某一物体在环境中运动，与环境之间发生力的作用从而改变运动状态，这种运动描述方法称为拉格朗日法.

流体也适用拉格朗日法，但流体不断变形，从几何角度需要进行复杂的坐标变换，因此，直接描述一小部分流体在整个流场空间（环境）的运动不大方便. 我们考虑是不是可以只研究一个特定不变的空间呢，着眼于流体经过这个空间时发生的变化以及与这个空间的相互作用，是否也能全面描述流体的运动呢，实践证明这也是可行的，这种方法称为欧拉法.

拉格朗日法

我们先尝试使用拉格朗日法描述流体的运动，考虑一个无穷小的不规则流体微团 A，为研究其运动，我们跟随此微团在时间和空间中运动，从而描述微团的空间位置随时间变化的轨迹. 可以想象我们是坐在船上，船跟随河水向前运动.

运动过程中，微团形状不断变化，但仍将其看作一个整体，独立的变量为时间 t 与变动的空间坐标 $s=(x,y,z)$，拉格朗日法这样描述流体微团的运动：

在 t 时刻，微团 A 的空间位置为 $\mathbf{r}=r(t,s)$，某一瞬时的位置与时间和空间坐标都有关.

在 t 时刻，微团 A 的速度为 $\mathbf{V}=\mathrm{d}\mathbf{r}/\mathrm{d}x$，某一瞬时的速度是空间坐标对时间的导数.

在 t 时刻，微团 A 的加速度为 $\mathbf{a}=\mathrm{d}\mathbf{V}/\mathrm{d}x=\mathrm{d}^2\mathbf{r}/\mathrm{d}x^2$.

因为微团在空间不断运动，拉格朗日法的坐标本身也在空间内不断变化，由于一般的流体微团内部都包含大量的质点，且由于微团的变形，不同质点的运动情况复杂多样，拉格朗日法不再方便.

欧拉法

欧拉法描述流体运动的方式则是将注意力集中到流体随时间流过的固定的空间位置 A 上，研究一个固定的发生流体运动的空间，空间形状可以任意而空间坐标是固定的. 可以想象我们站在桥上，观察桥下河水的流动.

欧拉法的独立变量是时间 t 和所研究空间的固定坐标 $s=(x,y,z)$，欧拉法这样描述通过固定空间的流体的运动：

在 t 时刻，A 处流体质点的速度为 $\mathbf{V}=u(t.x,y,z)\hat{\mathbf{i}}+v(t,x,y,z)\hat{\mathbf{j}}+w(t,x,y,z)\hat{\mathbf{k}}$，速度是某一时刻位于空间某一点的流体质点的速度，其与时间和空间坐标都有关.

至于 t 时刻的加速度，需要使用物质导数（Substantial Derivative）的概念，参见物质导数[992]，实际上，此时的 A 处流体质点的加速度为 $\mathbf{a}=\partial\mathbf{V}/\partial t+(\mathbf{V}\cdot\boldsymbol{\nabla})\mathbf{V}$，这个式子由两部分组成，前一部分只与时间相关，后一部分与空间相关，使用这种表达，欧拉法更容易描述流体运动的具体情况.

物质导数（实质导数）

- 本词条处于草稿阶段.

物质导数定义

在流体力学中，物质导数实际上是**拉格朗日法下某变量对时间的导数**，但它可以表示成**对欧拉变量的全导数**形式，因其特殊性，常用大写的微分符号来

表示，称为物质导数（Substantial Derivative），或实质导数，又或者随体导数. 设 $\boldsymbol{\Phi}$ 是流体的某种性质，物质导数的一般形式为：

$$\frac{\mathrm{D}\boldsymbol{\Phi}}{\mathrm{D}t} = \frac{\partial \boldsymbol{\Phi}}{\partial t} + (\mathbf{V}\cdot\boldsymbol{\nabla})\boldsymbol{\Phi} \tag{1}$$

公式中，等号左边的项为物质导数，等号右边的第一项称作当地导数，第二项称作对流导数或者牵连导数，这两项是欧拉法下的描述.

推导过程

第五章

振动与波动

振动的指数形式

预备知识 二阶常系数齐次微分方程[215]

简谐振子的微分方程

$$m\ddot{x} = -kx \tag{1}$$

是一个二阶常系数齐次微分方程．其复数域的通解可以表示为

$$x(t) = C_1 \mathrm{e}^{\mathrm{i}\omega t} + C_2 \mathrm{e}^{-\mathrm{i}\omega t} \tag{2}$$

其中 C_1, C_2 是任意复常数．由于指数函数的运算往往比三角函数方便，物理或工程中常常用指数函数表示振动，即把式 1 的通解记为[1]

$$\tilde{x}(t) = \tilde{A}\mathrm{e}^{-\mathrm{i}\omega t} \tag{3}$$

其中 $\tilde{A}$ 是一个复数[2]，称为**复振幅**，$\tilde{A}$ 的模长 $A = |\tilde{A}|$ 就是振幅，$\tilde{A}$ 幅角的相反数 $\varphi_0 = -\arg(\tilde{A})$ 就是初相位[3]．当我们用式 3 表示振动时，其实部表示质点的坐标，**虚部没有物理意义**．

为了验证式 3 的确包含了实数域的通解，我们可以先把复振幅表示为 $\tilde{A} = A\mathrm{e}^{-\mathrm{i}\varphi_0}$，代入式 3，再取实部得

$$x(t) = \mathrm{Re}[\tilde{x}(t)] = A\,\mathrm{Re}\left[\mathrm{e}^{-\mathrm{i}(\omega t+\varphi_0)}\right] = A\cos(\omega t + \varphi_0) \tag{4}$$

速度和加速度

由于复数的求导等效于对实部和虚部分别求导，当我们求速度或加速度时，也可以直接对式 3 求导．

$$\tilde{v}(t) = \frac{\mathrm{d}\tilde{x}}{\mathrm{d}t} = -\mathrm{i}\omega\tilde{A}\mathrm{e}^{-\mathrm{i}\omega t} \tag{5}$$

[1] 式 3 中 $\mathrm{e}^{-\mathrm{i}\omega t}$ 里的负号是一种习惯，有些教材中也会使用正号．无论使用哪一种，必须在计算中保持一致．

[2] 在变量上方加波浪线通常为了强调该变量是一个复数，但为了书写方便有时候也会省略，需要从语境中判断．

[3] 如果式 3 中没有负号，则初相位定义为 $\varphi_0 = \arg(\tilde{A})$

它的实部就是质点的速度

$$v(t) = \mathrm{Re}[\tilde{v}(t)] = A\,\mathrm{Re}\left[-\mathrm{i}e^{-\mathrm{i}\omega(\omega t+\varphi_0)}\right] = -\omega A\sin(\omega t+\varphi_0) \tag{6}$$

可以验证这和直接对式 4 求导相同．然而式 5 的复数表示要简洁得多．同理，加速度的复数形式为

$$\tilde{a}(t) = \frac{\mathrm{d}^2\tilde{x}}{\mathrm{d}t^2} = -\omega^2\tilde{A}e^{-\mathrm{i}\omega t} \tag{7}$$

$$\begin{aligned} a(t) &= \mathrm{Re}[\tilde{a}(t)] = -\omega^2\,\mathrm{Re}[\tilde{x}(t)] = -\omega^2 x(t) \\ &= -\omega^2 A\cos(\omega t+\varphi_0) \end{aligned} \tag{8}$$

相比于对式 4 求二阶导数，式 7 同样简洁得多．

振动的叠加

这里举另一个例子说明使用指数函数表示振动比三角函数方便．假设有若干个频率相同但振幅和初相位各不相同的振动 $x_i(t) = A_i\cos(\omega t+\varphi_{0i})$，现在我们来计算它们叠加的结果，即 $\sum_i x_i(t)$．若用两角和公式（式 4 [112]）直接计算，得

$$\begin{aligned} \sum_i x_i(t) &= \sum_i [A_i\cos\varphi_{0i}\cos(\omega t) - A_i\sin\varphi_{0i}\sin(\omega t)] \\ &= \sum_i (A_i\cos\varphi_{0i})\cos(\omega t) - \sum_i (A_i\sin\varphi_{0i})\sin(\omega t) \end{aligned} \tag{9}$$

分别令 $C = \sum_i A_i\cos\varphi_{0i}$，$D = \sum_i A_i\sin\varphi_{0i}$，且令 $A = \sqrt{C^2+D^2}$，以及令 φ_0 满足 $\cos\varphi_0 = C/A$，$\sin\varphi_0 = D/A$，则上式变为

$$\sum_i x_i(t) = A[\cos\varphi_0\cos(\omega t) - \sin\varphi_0\sin(\omega t)] = A\cos(\omega t+\varphi_0) \tag{10}$$

可见任意多个相同频率的简谐波叠加仍然是该频率的一个简谐波．

若我们用指数形式的振动来进行同样的计算，第 i 个振动可表示为 $\tilde{x}_i(t) = \tilde{A}_i e^{-\mathrm{i}\omega t}$，其中 $\tilde{A}_i = A_i e^{-\mathrm{i}\omega\varphi_{0i}}$．求和得

$$\sum_i \tilde{x}_i(t) = \left(\sum_i \tilde{A}_i\right) e^{-\mathrm{i}\omega t} \tag{11}$$

令 $\tilde{A}=\sum_i \tilde{A}_i$，$A=|\tilde{A}|$，$\varphi_0=-\arg(\tilde{A})$，则最后结果为

$$\sum_i \tilde{x}_i(t)=\tilde{A}\mathrm{e}^{-\mathrm{i}\omega t} \tag{12}$$

$$\sum_i x_i(t)=\mathrm{Re}\left[\tilde{A}\mathrm{e}^{-\mathrm{i}\omega t}\right]=A\cos(\omega t+\varphi_0) \tag{13}$$

不难证明上式的 A 和 φ_0 与式 10 得到的 A 和 φ_0 相同，但是这里的推导的过程却更为简洁.

能量法解谐振动问题

预备知识 简谐振子[913]，机械能守恒[909]

由于谐振动系统的总能量是常量，反过来我们可以用机械能守恒定律来导出谐振动的运动方程. 仍以弹簧振子为例，在任一时刻，振子作谐振动的总机械能等于势能加上动能，即

$$E=E_\mathrm{k}+E_\mathrm{p}=\frac{1}{2}mv^2+\frac{1}{2}kx^2 \tag{1}$$

由于总机械能 E 是常量，不随时间而改变，上式对时间 t 求导得

$$mv\frac{\mathrm{d}v}{\mathrm{d}t}+kx\frac{\mathrm{d}x}{\mathrm{d}t}=0 \tag{2}$$

由速度和坐标间的关系 $v=\mathrm{d}x/\mathrm{d}t$，代入整理后，即得谐振动的运动方程

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2}+\frac{k}{m}x=0 \tag{3}$$

能量法在解简谐运动（广义来说，机械能守恒的）问题中非常有帮助，下面来看一个例题.

例 1　考虑弹簧质量的弹簧振子

当考虑弹簧的质量时，用能量法求解弹簧振子的周期. 设弹簧的质量为 m_0，且小于振动物体的质量 m，弹簧的劲度系数为 k，弹簧的原长为 L.

在振动过程中任一时刻，设物体的位移为 x，速度为 v. 如图 1 所示. 弹簧振动时，可以认为弹簧在振动的任一时刻内，各等长小段的变形相同，弹簧

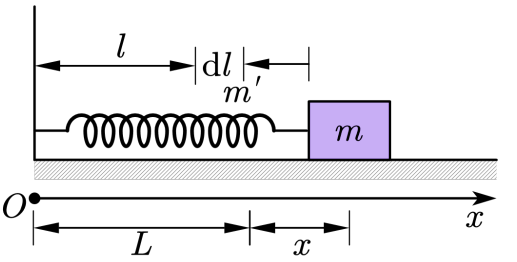


图 1: 考虑弹簧质量的弹簧振子

各截面处的位移是线性变化的. 在离弹簧固定端距离 l 的地方取一小段微元长度 $\mathrm{d}l$, 质量即为 $\dfrac{m'}{L}\mathrm{d}l$, 位移为 $\dfrac{x}{L}l$, 因而该段微元长度的速度就等于 $\dfrac{v}{L}l$, 写出其微元动能的表达式:

$$\mathrm{d}E'_{\mathrm{k}} = \frac{1}{2}\left(\frac{m'}{L}\mathrm{d}l\right)\left(\frac{l}{L}v\right)^2 = \frac{m'v^2}{2L^3}l^2\mathrm{d}l \tag{4}$$

积分, 可得整个弹簧的动能:

$$E'_{\mathrm{k}} = \int_0^L \frac{m'v^2}{2L^3}l^2\mathrm{d}l = \frac{1}{2}\frac{m'}{3}v^2 \tag{5}$$

于是, 整个弹簧振子系统的总机械能

$$E = \frac{1}{2}kx^2 + \frac{1}{2}mv^2 + \frac{1}{2}\frac{m'}{3}v^2 \tag{6}$$

由于整个系统没有能量损耗, 所以机械能守恒, 即

$$\frac{1}{2}kx^2 + \frac{1}{2}mv^2 + \frac{1}{2}\frac{m'}{3}v^2 = E = \text{constant} \tag{7}$$

将此方程对时间 t 求导, 得

$$\left(m + \frac{m'}{3}\right)\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + kx = 0 \tag{8}$$

将其改写成简谐振动的微分方程形式:

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + \omega^2 x = 0 \qquad \left(\omega^2 = \frac{k}{m + m'/3}\right) \tag{9}$$

于是可得运动周期:

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{m + m'/3}{k}} \tag{10}$$

拍频

预备知识 简谐振子[913]

同一直线上两个不同频率的谐振动的合成

设一质点在一直线上同时参与两个不同频率的谐振动，其振动表达式为

$$\begin{aligned} x_1 &= A_1 \cos\left(\omega_1 t + \phi_{01}\right) \\ x_2 &= A_2 \cos\left(\omega_2 t + \phi_{02}\right) \end{aligned} \tag{1}$$

根据叠加原理，合运动的位移为

$$x = x_1 + x_2 = A_1 \cos\left(\omega_1 t + \phi_{01}\right) + A_2 \cos\left(\omega_2 t + \phi_{02}\right) \tag{2}$$

为方便计算，设 $A_1 = A_2 = A, \phi_{01} = \phi_{02} = \phi_0$，则上式可化成

$$x = 2A \cos\left(\frac{\omega_2 - \omega_1}{2} t\right) \cos\left(\frac{\omega_2 + \omega_1}{2} t + \phi_0\right) \tag{3}$$

对于通常实际遇到的情况而言，两个频率比较接近，且 $|\omega_2 - \omega_1| \ll \omega_1$，式 3 中第一项因子随时间作缓慢地变化，第二项因子是角频率近于 ω（即接近于 ω_1, ω_2）的简谐函数，因此合成的振动可近似看成是角频率为 $(\omega_1 + \omega_2)/2 \approx \omega_1 \approx \omega_2$，振幅为 $|2A\cos(\omega_2 - \omega_1)t/2|$ 的谐振动．这种两个频率较大且差值较小的谐振动合成时，合振幅出现时强时弱周期性缓慢变化的现象，叫做**拍（beat）**．

拍出现的频率叫做**拍频**，等于两分振动频率之差．

拍现象在技术上有重要应用．例如，管乐器中的双簧管就是利用两个簧片振动频率的微小差别产生颤动的拍音；调整乐器时，使它和标准音叉出现的拍音消失来校准乐器；拍现象常用于汽车速度监视器、地面卫星跟踪等．此外，在各种电子学测量仪器中，也常常用到拍现象．

受阻简谐振子

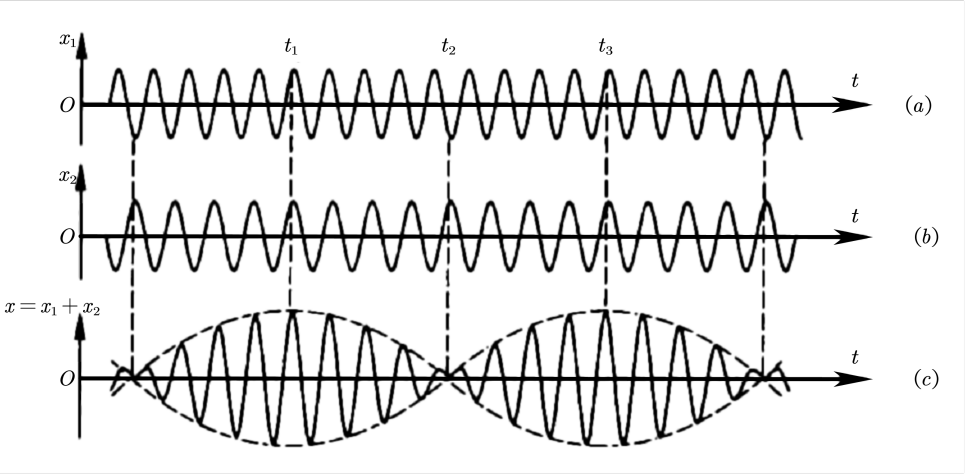


图 1: 拍

预备知识 二阶常系数齐次微分方程的[218]

结论

两个复数根（$\alpha^2-4km<0$，最常讨论和应用的情况）

$$y=C_1\mathrm{e}^{rx}\cos(\omega x+C_2) \tag{1}$$

其中

$$r=-\frac{\alpha}{2m}\qquad \omega=\sqrt{\omega_0^2-\frac{\alpha^2}{4m^2}} \tag{2}$$

其中 $\omega_0=\sqrt{k/m}$ 是**固有振动频率**（$\alpha=0$ 时的振动频率），r 和 ω 满足

$$r^2+\omega^2=\omega_0^2 \tag{3}$$

推导

在弹簧振子的振动方程基础上，若振子还受到一个与速度成正比的阻力 $f=-\alpha v$，则振动方程如下（$\alpha\neq 0$）.

$$my''=-\alpha y'-ky \tag{4}$$

之所以设为正比，是因为所得方程是线性方程，便于求解. 根据“二阶常系数齐次微分方程[218]”，解特征方程 $mr^2+\alpha r+k=0$，可得通解分为三种情况.

1. 有两个不同的实根 r_1, r_2 （$\alpha^2 - 4km > 0$），方程的通解为

$$y = C_1 \mathrm{e}^{r_1 x} + C_2 \mathrm{e}^{r_2 x} \tag{5}$$

这种情况下，阻力系数太大以至于质点直接减速回到平衡位置而无法发生任何振动.

2. 有一个重根 r （$\alpha^2 - 4km = 0$），方程的通解为

$$y = C_1 \mathrm{e}^{rx} + C_2 x \mathrm{e}^{rx} \tag{6}$$

这是质点振动与不振动的临界点.

3. 两个复数根 r_1, r_2 （$\alpha^2 - 4km < 0$，最常讨论和应用的情况）

$$y = C_1 \mathrm{e}^{rx} \cos(\omega x + C_2) \tag{7}$$

其中

$$r = -\frac{\alpha}{2m} \qquad \omega = \frac{1}{2m}\sqrt{4mk - \alpha^2} = \sqrt{\omega_0^2 - \frac{\alpha^2}{4m^2}} \tag{8}$$

这种情况下，质点做振幅不断衰减的振动，衰减系数 r 与阻力系数成正比. 若令（见简谐振子[913] 的振动频率）

$$\omega_0 = \sqrt{\frac{k}{m}}, \quad \gamma = \frac{\alpha}{2\sqrt{mk}} \tag{9}$$

则

$$r = -\omega_0 \gamma, \quad \omega = \omega_0 \sqrt{1 - \gamma^2} \tag{10}$$

满足 $r^2 + \omega^2 = \omega_0^2$.

简谐振子的品质因数

预备知识 简谐振子受迫运动[1004]

简谐振子的品质因数为

$$Q = \frac{k}{\alpha \omega_m} \tag{1}$$

其中 k 为弹性系数，α 为线性阻尼，ω_m 为共振角频率.

弹簧振子的品质因数有几种定义方法，它们是等效的

从能量定义

$$Q=\frac{2\pi E}{W} \tag{2}$$

其中 E 是当 $\omega=\omega_m$ 时弹簧振子做简谐运动的总能量，W 是 $\omega=\omega_m$ 时 $f(t)$ 在一个周期内给弹簧振子做的功，等于阻力在一个周期内消耗的能量．下面根据该定义推导．

当达到共振频率时（$\omega=\omega_m$），外力一周期做功

$$W=\int_0^{2\pi/\omega}\mathrm{Re}[y'(t)]\,\mathrm{Re}[f(t)]\,\mathrm{d}t \tag{3}$$

令 $A=A_x+\mathrm{i}A_y$，则

$$\mathrm{Re}[y'(x)]=\mathrm{Re}[\mathrm{i}\omega(A_x+\mathrm{i}A_y)(\cos\omega t+\mathrm{i}\sin\omega t)]=-\omega A_y\cos\omega t-\omega A_x\sin\omega t \tag{4}$$

$$\mathrm{Re}\,f(t)=B\cos\omega t \tag{5}$$

代入方程，得

$$W=-\omega B\int_0^{2\pi/\omega}(A_y\cos\omega t+A_x\sin\omega t)\cos\omega t\,\mathrm{d}t=-BA_y\pi \tag{6}$$

另外弹簧振子的总能量为 $E=k|A|^2/2$，所以品质因数为

$$Q=\frac{2\pi E}{W}=-\frac{k|A|^2}{BA_y} \tag{7}$$

此时

$$A=\frac{B}{m(\omega_0^2-\omega^2)+\mathrm{i}\alpha\omega}=\frac{B}{\frac{\alpha^2}{2m}+\mathrm{i}\omega\alpha} \tag{8}$$

所以

$$|A|=\frac{|B|}{\sqrt{\alpha^2\omega^2+\frac{\alpha^4}{4m^2}}} \tag{9}$$

$$A_y=\mathrm{Im}[A]=\frac{-B\omega\alpha}{\alpha^2\omega^2+\frac{\alpha^2}{4m^2}} \tag{10}$$

代入式 7，得品质因数为

$$Q=\frac{k}{\alpha\omega_m} \tag{11}$$

从幅频曲线定义

当线性阻尼 α 很小时，品质因数为

$$Q = \frac{f_m}{\Delta f} = \frac{\omega_m}{\Delta\omega} \tag{12}$$

其中 $\Delta\omega$ 是系统的能量—频率曲线的半高宽（FWHM）. 由此可以看出品质因数越大，幅频曲线就越尖锐越细.

由于弹簧振子能量为 $E = k|A|^2_{max}/2$，$\Delta\omega$ 也是幅频曲线的“$1/\sqrt{2}$ 高宽”. 即满足

$$|A| \geqslant \frac{1}{\sqrt{2}}|A|_{max} \tag{13}$$

的 ω 区间的宽度.

由于

$$\omega_m = \sqrt{\frac{k}{m} - \frac{\alpha^2}{2m^2}} = \sqrt{\omega_0^2 - \frac{\alpha^2}{2m^2}} \tag{14}$$

当阻力系数 α 很小时，弹簧振子的共振频率接近于无阻力时的频率. 根据（未完成），

$$\omega_m = \sqrt{\omega_0^2 - \frac{\alpha^2}{2m^2}} = \omega_0 - \frac{\alpha^2}{4\omega_0 m^2} \tag{15}$$

所以可以认为 $\omega_m \approx \omega_0$.

把幅频关系

$$|A| = \frac{|B|}{m\sqrt{(\omega^2 - \omega_m^2)^2 + (\omega_0^4 - \omega_m^4)}} \tag{16}$$

代入式 13 得

$$(\omega^2 - \omega_m^2)^2 + (\omega_0^4 - \omega_m^4) = 2(\omega_0^4 - \omega_m^4) \tag{17}$$

解得

$$\omega^2 = \omega_m^2 \pm \sqrt{\omega_0^4 - \omega_m^4} \tag{18}$$

所以区间以 ω_0 为中心且

$$\Delta(\omega^2) = 2\sqrt{\omega_0^4 - \omega_m^4} \tag{19}$$

而 $\Delta(\omega^2) \approx 2\omega_m\Delta\omega$，所以

$$\Delta\omega = \frac{\Delta(\omega^2)}{2\omega_m} = \frac{\sqrt{\omega_0^4 - \omega_m^4}}{\omega_m} \tag{20}$$

其中

$$\sqrt{\omega_0^4-\omega_m^4}=\sqrt{(\omega_0^2-\omega_m^2)(\omega_0^2+\omega_m^2)}\approx\sqrt{\frac{\alpha^2}{2m^2}\cdot 2\omega_m^2}=\frac{\alpha\omega_m}{m} \tag{21}$$

所以 $\Delta\omega=\alpha/m$．所以

$$Q=\frac{\omega_m}{\Delta\omega}\approx\frac{\omega_0}{\Delta\omega}=\frac{m\omega_0}{\alpha}=\frac{m}{\alpha}\sqrt{\frac{k}{m}}=\frac{k}{\alpha\omega_0}\approx\frac{k}{\alpha\omega_m} \tag{22}$$

与第一种定义结论一致．

简谐振子受迫运动

预备知识 受阻简谐振子[999]，振动的指数形式[995]

在受阻简谐振子的基础上，若给振子额外施加一个周期变化的力（驱动力），得到微分方程如下．

$$m\ddot{y}=-\alpha\dot{y}-ky+f(t) \tag{1}$$

以下只讨论 $f(t)$ 为简谐函数的情况，令其振幅为 B，频率为 ω．我们姑且假设经过足够长的时间后，该弹簧振子也会做简谐振动，振动频率等于 $f(t)$ 的频率．若能找到一个这样的解，就说明该假设是对的．

为了方便计算，我们用指数形式表示振动，设

$$\tilde{y}(t)=\tilde{A}\mathrm{e}^{-\mathrm{i}\omega t}\qquad\tilde{f}(t)=\tilde{B}\mathrm{e}^{-\mathrm{i}\omega t} \tag{2}$$

由于我们的微分方程是线性的，如果复数形式的 $y(t)$ 和 $f(t)$ 能满足微分方程，那么它们的实部也能满足该微分方程．将它们代入式 1，得

$$\tilde{A}=\frac{\tilde{B}}{k-\omega^2m-\mathrm{i}\alpha\omega} \tag{3}$$

由于弹簧振子的固有频率为 $\omega_0=\sqrt{k/m}$，上式可用 ω_0 表示为

$$\tilde{A}=\frac{\tilde{B}}{m(\omega_0^2-\omega^2)-\mathrm{i}\alpha\omega} \tag{4}$$

上式两边求模长，得到简谐振子的振幅 $A=|\tilde{A}|$ 与驱动力频率 ω 的关系，称为**幅频关系**

$$A=\frac{B}{\sqrt{(\omega^2-\omega_0^2)^2m^2+\alpha^2\omega^2}} \tag{5}$$

假设 $\tilde{B}$ 的幅角为零，对 $\tilde{A}$ 求负幅角，得简谐振子初相位 φ_0 与驱动频率 ω 的关系，称为**相频关系**（$\arctan 2$ 的定义见）

$$\varphi_0=-\arg\tilde{A}=\arctan 2\left[m(\omega_0^2-\omega^2),-\alpha\omega\right] \tag{6}$$

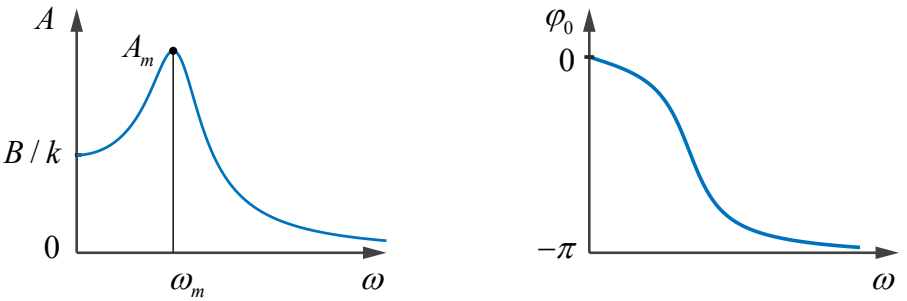


图 1: 幅频曲线和相频曲线

式 5 的根号内是关于 ω^2 的二次函数，求得二次函数最小值的位置为

$$\omega_m=\sqrt{\omega_0^2-\frac{\alpha^2}{2m^2}} \tag{7}$$

幅频曲线可改写为

$$A=\frac{B}{m\sqrt{(\omega^2-\omega_m^2)^2+(\omega_0^4-\omega_m^4)}} \tag{8}$$

所以 A 的最大值为

$$A_m=\frac{B}{m\sqrt{\omega_0^4-\omega_m^4}} \tag{9}$$

由式 8 可得，当 $\omega=0$ 和 $\omega\to+\infty$ 时，振幅分别为 B/k 和 0．前者代表施加的是一个恒力，结论符合胡克定律．

再来观察相频曲线，注意初相位始终为负，说明简谐运动的相位始终落后于驱动力的相位，且频率越快，落后越多．由式 6 可知当 $\omega\to+\infty$ 时相位恰好落后 π．

共振

- 本词条缺少预备知识，初学者可能会遇到困难.

[4]如同讨论类氢原子[1536] 中的那样，原子系统是存在能级的，从而可以用特定频率的光去激发，而这种行为也可以一般的描述为共振．Wikipedia 说

In mechanical systems, resonance is a phenomenon that only occurs when the frequency at which a force is periodically applied is equal or nearly equal to one of the natural frequencies of the system on which it acts. This causes the system to oscillate with larger amplitude than when the force is applied at other frequencies.

而对应到原子里面，这个 force 就是周期的电磁场，而 oscillation 则是原子在基态和激发态之间振荡．对应的，这个振荡不会仅仅发生在共振频率，其它频率也会存在振荡只不过振幅不如共振点大．那么，当频率偏离共振后，振幅的变化其实反映了这个共振的“好坏”：如果稍微偏离一点，振幅就大幅衰减，那么说明这个共振非常锐利而精确，几乎只发生在共振点，这样的共振我们说它有更高的“品质”．具体说来，我们用品质因数（quality factor）来定义：

$$Q = \frac{f}{\Delta f} \tag{1}$$

其中，而 f 是共振频率，Δf 是共振的带宽（bandwidth），对于不同的共振线性有不同的定义[5].

举一个例子：对于一般的室温原子气体，这个参数基本由室温的热运动造成的多普勒展宽决定. 原子的能级一般在可见光波段，即 $f \sim 500\text{THz}$[6]. 对应的，假如原子序数为 100，室温 $T = 300\text{K}$，对应的最可几速率为 $\bar{v} = \sqrt{2k_BT/m} \sim 300\text{m/s}$，从而对应了一个 $\Delta f = f\bar{v}/c \sim 500 \times 10^{-6}\text{THz} = 500\text{MHz}$ 的展宽．这个对应了品质因子 $Q \sim 10^6$．这个展宽一般大于常规原子能级的展宽（在 10MHz 量级），从而会把原子本身能级的性质掩盖掉.

[4]参考 Wikipedia 相关页面.

[5]比如，对于 Lorentzian 这种有无限大的方差的线型来说就是半高全宽（Full Width at Half Maximum, FWHM），而对于 Gaussian 线型来说就是它的标准差.

[6]在此后的讨论中，**频率**统一由 $f = \cdots\text{Hz}$ 这种形式表示，而对应的**角频率** $\omega = 2\pi \times \cdots\text{Hz}$ 来表示，两个 $\cdots$ 是一样的.

常见的品质好（$Q > 100$）的共振有石英石的共振 $Q \sim 10^4 - 10^6$，也就是说现在的石英表在一年里一般也就误差 $5 - 10$ 分钟量级，而机械表就要差一个档次了，很可能到一个小时．此外，人工制造的微观结构，比如回音壁腔，能达到 10^9 的品质因子，是目前人类构造出的最高的品质因子层级．而地球自转的品质因子在 10^7 水平，脉冲星自转的 Q 在 10^{10} 水平．这种精度实际上让人们第一次的间接的认识到了引力波，因为脉冲星释放了引力波，释放了能量，其自转速度开始下降．由于它超高的品质因数，只要频率相对下降 10^{-10} 这么多就能够从频谱上看到明显变化．当然，这种随着时间变化的频谱不具有可重复性，也就没办法用来作为频率标准．

共振有不同的线型．刚才说的热展宽，由于原子速度是玻尔兹曼分布，所以是 Gaussian 线型；而原子自己的能级则是 Lorentzian 线型，它对应于谐振子的阻尼振荡[1004]，其中幅频关系式 5 为

$$A = \frac{B}{\sqrt{(\omega^2 - \omega_0^2)^2 m^2 + \alpha^2 \omega^2}} \tag{2}$$

而对应的能量，或者说是系统的响应，$E \propto A^2$，在 $Q \sim \omega_0/\gamma \gg 1$ 的情况下可以近似满足

$$E = \frac{C}{\Delta\omega^2 + \gamma^2/4}. \tag{3}$$

接下来，我们讨论对一个这种高 Q 阻尼振荡里面，频率的测量到底能有多精准．这个问题之所以重要是因为当今最准确的测量就是对于频率的测量，所以这个问题的讨论实际上对应于我们对于物理量的测量到底能精确到什么程度．当然，这个问题最最直接的回答是由 Fourier 定理保证的，$\Delta t \Delta\omega \geq 1/2$．这个形式与 Heisenberg 不确定性原理非常像，其原因之一就是量子力学的量子态是有其波动内涵的．然而这个回答是非常微妙的，因为对于 $\Delta\omega$ 的定义其实并不明确．

考虑一个经典系统，给定初始条件，它的振荡是有阻尼且有噪音的，具体波形是 $x = x_0(1+\epsilon)\cos(\omega t)e^{-\gamma t}$，其中 $\epsilon \in [-r, r]$ 是随机参数，$r < 1$ 是随机强度，代表信噪比的倒数．考虑我们有采样率 $2000 s^{-1}$ 的信号，$\omega = 2\pi \times 447\text{Hz}$, $\gamma = 2\pi \times 1\text{Hz}$，接下来的几幅图则展示了不同的信噪比，$\Delta t$，在相同 γ, ω 下的频谱．

平面波

预备知识 矢量内积[383]，简谐振子[913]

通常我们说“波”时，它几乎可以是任意形状的，例如一个波包[1011]．但我们先来讨论一种最简单的波动，叫做**简谐波（sinusoidal wave）**，在任意时刻，它在传播方向的波形是一个正弦函数，范围是无穷大．简谐波也叫**平面波（plane wave）**，本书对这两个词不加区分．叫平面波是因为三维平面波的等相位面（见下文）都是平面，但其他维度习惯上也可以使用这个词，这在量子力学中较常见[7]．

一维情况

我们先来看一个一维的平面波，一个常用的例子是一根无限长的弦，静止的时候弦与 x 轴重合，在任何时刻 t，弦的波函数（即形状）可以用 $y(x,t)$ 来描述．若

$$y(x,t) = A\cos(kx - \omega t + \phi_0) \tag{1}$$

则我们把这个波函数称为**平面波**，如图 1 所示[8]．

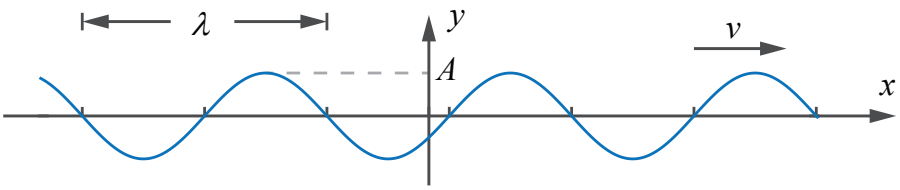


图 1: 平面波

我们定义图中的 A 为振幅，定义一个空间周期为**波长（wave length）**，记为 λ．与波长一一对应的一个量是式 1 中的 k，称为**波数（wave number）**．波长与波数的关系可以类比简谐振子[913] 的角频率 ω 与时间周期 T 的关系，即

$$k = \frac{2\pi}{\lambda} \tag{2}$$

[7] 见几乎任意一本常见量子力学入门教材

[8] 需要注意的是，图中的横轴是位置 x 而不是时间 t，要避免将质点振动的位移—时间图与该图混淆．

我们再来看波函数随时间的变化，如果在弦的某个位置做一个标记并观察其运动，则式 1 中 x 可视为常数，我们立即得到一个简谐振动，角频率为 ω，初相位为[9] $-kx-\phi_0$.

我们在观察平面波的时候，通常会想象它在移动（虽然弦上每个点的 x 坐标并不改变），我们把这种移动的速度叫做**波速（wave velocity）** v. 把式 1 稍作整理得

$$y(x,t)=A\cos\left[k\left(x-\frac{\omega}{k}t\right)+\phi_0\right] \tag{3}$$

由于函数 $f(x-x_0)$ 可以看做 $f(x)$ 向 x 轴正方向平移 x_0 得到的函数，上式也可以看做 $t=0$ 时刻的波函数向 x 轴正方向平移 $\omega t/k$ 得到的波函数. 将平移距离除以 t 就得到了单位时间移动的距离，即波速

$$v=\frac{\omega}{k} \tag{4}$$

如果将 $\omega=2\pi/T$ 和 $k=2\pi/\lambda$ 代入上式，得到波速的另一个表达式

$$v=\frac{\lambda}{T} \tag{5}$$

这里的 T 是振动周期. 也可以令振动频率 $f=1/T$，则上式又变为

$$v=\lambda f \tag{6}$$

横波与纵波

以上我们看到的波函数表示**横波**，即质点振动的方向与波的传播方向垂直. 与横波相对的另一类波叫做**纵波**，即质点振动方向与波的传播方向相同. 纵波的波函数与横波相同，只是函数值的意义由垂直方向的位移改为了平行方向的位移（不妨记为 ξ）

$$\xi=A\cos(kx-\omega t+\phi_0) \tag{7}$$

广义来说，振幅可以是一个任意方向的矢量 $\mathbf{A}$，它可以既不平行也不垂直于传播方向.

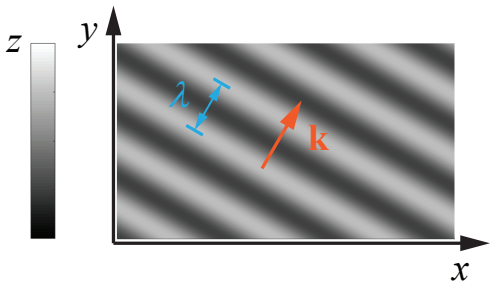


图 2: 二维平面波

二维和三维的平面波

如图 2，我们可以用函数 $z(x,y,t)$ 表示一个二维的平面波（横波）．波长的定义与一维情况相同，在 $k=2\pi/\lambda$ 的基础上，我们定义**波矢 k** 的方向为波速的方向．观察图中的波可以发现，沿波矢方向移动 l，相位变化为 kl，沿垂直波矢方向移动 l，相位不改变，沿任意其他方向移动 l，相位变化为 $kl\cos\theta$，其中 θ 是移动方向与 $\mathbf{k}$ 方向的夹角．垂直于波矢的任意直线就是二维平面波的**等相位面**（线）．我们可以用内积来表示相位随空间的变化

$$\Delta\phi=\mathbf{k}\cdot\Delta\mathbf{r}=k_x\Delta x+k_y\Delta y \tag{8}$$

于是我们可以写出波函数为

$$z=A\cos(\mathbf{k}\cdot\mathbf{r}-\omega t+\phi_0) \tag{9}$$

要表示纵波，同样把 z 换位 ξ 即可．

类似地，三维空间中的平面波可表示为

$$\mathbf{s}=\mathbf{A}\cos(\mathbf{k}\cdot\mathbf{r}-\omega t+\phi_0) \tag{10}$$

其中 $\mathbf{k}$ 和 $\mathbf{r}$ 是三维矢量．注意这里的 $\mathbf{r}$ 表示介质静止时某质点的位矢．如果波函数表示横波，矢量振幅 $\mathbf{A}$ 必须垂直于波矢 $\mathbf{k}$，其方向叫做**极化方向**．如果波函数表示纵波，$\mathbf{A}$ 必须与 $\mathbf{k}$ 同向．

波函数的复数表示

[9]由于余弦函数是偶函数，我们不妨将 cos 的自变量取相反数使 ωt 的符号为正．

预备知识 振动的指数形式[995]

用复数表示波函数，往往可以化简书写和计算．类比式 3 [995]，我们可以把平面波表示为指数形式[10]

$$\tilde{\mathbf{s}} = \mathbf{A}\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}-\omega t+\phi_0)} \tag{11}$$

注意只有实部表示质点的位移，虚部无物理意义．

未完成：平面波的叠加？波包分解成平面波？相速度？群速度？

波包

预备知识 平面波[1008]

[11]**波包（wave packet）**是一种局域的波动．也就是说在某个时刻，波包的振幅关于位置变化，超出一定范围后，振幅以某种方式衰减（但不一定严格为零，例如高斯波包[1012]）．

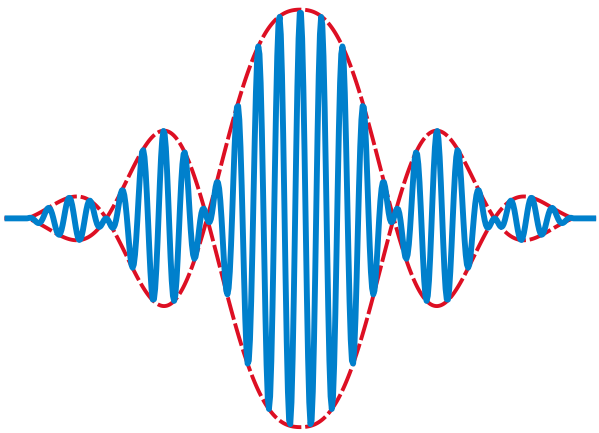


图 1: 波包，红色虚线是包络线（来自维基百科）

[10]现在我们知道为什么振动的指数形式中 ωt 要带一个负号了，这样就可以让波动的指数形式中 $\mathbf{k}\cdot\mathbf{r}$ 项为正．

[11]参考 Wikipedia 相关页面．

波函数的复数表示

简单的波包通常可以表示为以下形式

$$f(\mathbf{r},t)=A(\mathbf{r},t)\exp(\mathbf{k}\cdot\mathbf{r}-\omega t) \tag{1}$$

其中 A 是复振幅．经典物理中使用复数表示振动或者波动时．

这种形式的常见的波包是高斯波包[1012]．

高斯波包

- 本词条处于草稿阶段．

预备知识 波包[1011]，高斯分布[731]

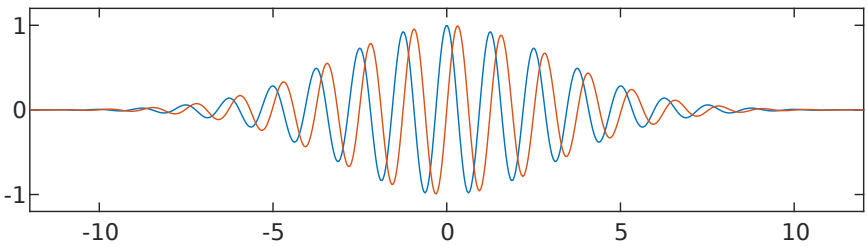


图 1: 高斯波包（式 1 ），蓝色为实部，红色为虚部，$x_0=0$，$A_0=1$，$a=1/20$，$k_0=5$．

[12]**高斯波包（Gaussian wave packet）**是指轮廓为高斯分布的波包，在光学和量子力学中有重要应用．高斯波包用复函数表示为（A_0 为复数）

$$f(x)=A_0\mathrm{e}^{-a(x-x_0)^2}\mathrm{e}^{\mathrm{i}k_0x} \tag{1}$$

FWHMI （光强半高宽）是多少？即 $f(x)^2$ 大于峰值一半时的宽度．令

$$\mathrm{e}^{-2a\Delta x^2}=1/2 \tag{2}$$

得

$$\mathrm{FWHMI}=2\Delta x=\sqrt{\frac{2\ln 2}{a}} \tag{3}$$

[12]参考 Wikipedia 相关页面．

频谱

预备知识 傅里叶变换（指数）[237]

（未完成：哪里有介绍频谱的概念？）

要求式 1 的傅里叶变换 $g(k)$，由例 1 [237] 以及傅里叶变换性质式 13 [237] 和式 14 [237] 得

$$g(k)=A_0\sqrt{\frac{\pi}{a}}\exp\left[-\frac{(k-k_0)^2}{4a}\right] \tag{4}$$

群速度

- 本词条处于草稿阶段.

当波速不随频率变化时，波形整体移动.

波速是频率的函数

$$v_g=\frac{\mathrm{d}\omega}{\mathrm{d}k} \tag{1}$$

（未完成：推导一下量子力学中中心动量为 k 的波包的相速度为什么是 $2k$）

多普勒效应（一维匀速）

预备知识 平面波[1008]

多普勒效应（Doppler effect）讨论的是，当机械波的波源和（或）接收者相对于波的介质运动时，发射的频率和接收到的频率之间的变化关系．本文只讨论波源和接收者沿同一直线匀速运动的情况．另外本文不讨论相对论效应，即假设波速远小于真空中的光速（另见“光的多普勒效应[1283]”）.

例 1

生活中一种常见的多普勒效应是，一辆疾驰的车一边鸣笛一边驶过行人，人听到的音调就会先高后低．这是因为，车经过人之前不断靠近人，经过人后再不断远离人．可见多普勒效应和运动的速度有关.

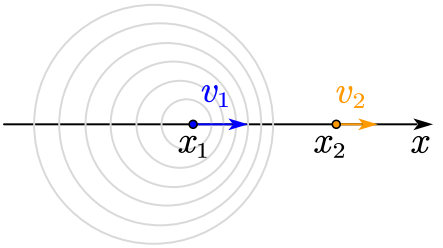


图 1: 多普勒效应

在分析多普勒效应时，一种方便的做法是选取介质为参考系，例如有均匀的风时，参考系随风运动．假设介质处处均匀且静止，波在介质中传播的速度（**波速**）v 处处相等，且与方向无关（**各向同性**）．

令图 1 中波源甲的运动方程，即位置关于时间的函数为 $x_1(t)$，对时间求导得到速度 $v_1(t)=\dot{x}_1(t)$．同理，接收者乙位置和速度分别为 $x_2(t)$ 和 $v_2(t)$．注意 v_1, v_2 都是向右运动为正，向左运动为负．令波源的频率为 f_1，接收者收到的频率为 f_2．当两点分别以小于波速的速度（$|v_1|, |v_2| < v$）匀速运动时，有

$$\frac{f_2}{f_1}=\frac{v\mp v_2}{v\mp v_1} \tag{1}$$

当 $x_1<x_2$ 时该式取负号，当 $x_2<x_1$ 时取正号，推导见下文．

习题 1

在一条长直马路上，风速为 10 m/s，一个单车以 5 m/s 向顺风而行，迎面驶来一辆摩托车，速度为 20 m/s．若摩托车一直以 600 Hz 的频率鸣喇叭，声速为 340 m/s，求摩托车经过单车前后骑单车的人听到的频率．

令风马路为 x 轴，风向为正方向，在风的参考系中，摩托车速度为

$$v_1=-20-10=-30(\text{ m/s}) \tag{2}$$

单车速度为

$$v_2=5-10=-5\text{ m/s} \tag{3}$$

另外 $v=340$ m/s，相遇以前 $x_2<x_1$，所以

$$f_2=\frac{v+v_2}{v+v_1}f_1=648.4\text{ Hz} \tag{4}$$

相遇以后 $x_1<x_2$，所以

$$f_2=\frac{v-v_2}{v-v_1}f_1=559.5\text{ Hz} \tag{5}$$

注意我们必须在风的参考系中计算，否则得不到正确的结果（除非 $v_1 = v_2$）.

反射的多普勒效应

当我们考虑波从一个运动的点上反射．每当这个点接收一个周期的波，就同时反射一个周期的波．所以我们可以把这个点作为一个新的波源，发射的频率与接收的频率相同．这样我们就可以重复使用式 1 得到最终接收者的频率.

例 2

假设 $x_3 < x_1 < x_2 < x_4$，当波从 x_1 发出，经过 x_2, x_3 分别反射，被 x_4 接收，有

$$\frac{f_4}{f_1} = \frac{f_4}{f_3}\frac{f_3}{f_2}\frac{f_2}{f_1} = \frac{v - v_4}{v - v_3}\frac{v + v_3}{v + v_2}\frac{v - v_2}{v - v_1} \tag{6}$$

可见如果这四个点都向中间运动，波每反射一次都会使频率变快.

推导

本质上，多普勒效应可以等效为追及问题，可以想象甲以一定的频率 f_1 向乙发射速度为 v 的子弹，子弹的位置对应波峰的位置，两个相邻子弹之间的间距对应波长．若甲乙相对介质静止不动或者以相同的速度运动，则乙接收到子弹的频率和甲发射的频率是一样的，但若甲乙不断靠近，则接收子弹的频率就会更高，若不断远离，则接受的频率就更低.

以 $x_1 < x_2$ 为例，假设波源在某时刻在 x_{10} 位置向右发射一枚子弹，经过周期 $T_1 = 1/f_1$ 后又发射第二枚时，第一枚的位置为 $x_{10} + vT_1$，而此时波源的位置（也就是第二枚子弹的位置）为 $x_{10} + v_1T_1$．所以两枚子弹相距（也就是波长）为

$$\lambda = (v - v_1)T_1 \tag{7}$$

同理，接收者收到这两枚子弹的时间间隔为

$$T_2 = \lambda/(v - v_2) \tag{8}$$

以上两式消去 λ，有

$$\frac{f_2}{f_1} = \frac{T_1}{T_2} = \frac{v - v_2}{v - v_1} \tag{9}$$

这样就得到了式 1．$x_2 < x_1$ 的推导同理可得.

多普勒效应

预备知识 多普勒效应（一维变速）

我们相对波介质静止建立参考系，令移动的波源位置为 $\mathbf{r}_0(t)$，振动相位为 $\phi_0(t)$，t 时刻波源到位矢 $\mathbf{r}$ 的距离为 $|\mathbf{r}-\mathbf{r}_0(t)|$. 令点观测点接收到的相位为 $\phi(\mathbf{r},t)$，而这个相位是波源在 t_0 时刻发出得，则有

$$|\mathbf{r}-\mathbf{r}_0(t-\Delta t)|=v\Delta t \tag{1}$$

其中 v 为波在介质中传播的速度，Δt 是波从波源 $\mathbf{r}_0(t-\Delta t)$ 处传到 $\mathbf{r}(t)$ 所需要的时间．$t-\Delta t$ 时刻波源发出的波阵面（相位不变）在 t 时刻到达 $\mathbf{r}$，所以

$$\phi(\mathbf{r},t)=\phi_0(t-\Delta t) \tag{2}$$

根据式 1 和式 2 可以求出空间中的相位场 $\phi(\mathbf{r},t)$．如果观察者在 $\mathbf{r}$ 处不动，对时间求导即可得到振动的角频率．

若观察者移动，令位置为 $\mathbf{r}_1(t)$，相位为 $\phi_1(t)$，则

$$\phi_1(t)=\phi(\mathbf{r}_1(t),t) \tag{3}$$

对时间求导，可得任意时刻点观察者接收到的角频率．

频率公式

对式 3 求时间 t 得全导数[273]，得

$$\frac{\mathrm{d}\phi_1}{\mathrm{d}t}=\boldsymbol{\nabla}\phi\cdot\mathbf{v}_1+\frac{\partial\phi}{\partial t} \tag{4}$$

为了书写方便，我们令 $t-\Delta t=t_0$，即波源在 t_0 时刻发出的波阵面在 t 到达 $\mathbf{r}$．t_0 可以看成 t 和 $\mathbf{r}$ 的函数 $t_0(t,\mathbf{r})$

$$\phi(\mathbf{r},t)=\phi_0(t_0)=\omega t_0(\mathbf{r},t) \tag{5}$$

$$\frac{\partial\phi}{\partial t}=\omega\frac{\partial t_0}{\partial t}\qquad\boldsymbol{\nabla}\phi=\omega\boldsymbol{\nabla}t_0 \tag{6}$$

由几何关系

$$|\mathbf{r}-\mathbf{r}_0(t_0)|=v(t-t_0) \tag{7}$$

上式约束了 $t_0,t,\mathbf{r}$ 三者的关系．两边的增量可以写为（令 $\mathbf{r}-\mathbf{r}_0$ 的单位矢量为 $\hat{\mathbf{u}}$）

$$\mathrm{d}|\mathbf{r}-\mathbf{r}_0(t_0)|=\hat{\mathbf{u}}\cdot\mathrm{d}\mathbf{r}-\mathbf{u}\cdot\mathrm{d}\mathbf{r}_0=\hat{\mathbf{u}}\cdot\mathrm{d}\mathbf{r}-\hat{\mathbf{u}}\cdot\mathbf{v}_0\,\mathrm{d}t_0 \tag{8}$$

若要求 $\partial t_0/\partial t$ （隐含 $\mathbf{r}$ 不变），令 $\mathrm{d}\mathbf{r}=0$

$$-\hat{\mathbf{u}}\cdot\mathbf{v}_0\,\mathrm{d}t_0=v(\mathrm{d}t-\mathrm{d}t_0) \tag{9}$$

两边除以 $\mathrm{d}t$，得

$$\frac{\partial t_0}{\partial t}=\frac{1}{1+\hat{\mathbf{u}}\cdot\mathbf{v}_0/v} \tag{10}$$

要求 $\boldsymbol{\nabla}t_0$，先令 $\mathrm{d}t=0$

$$\hat{\mathbf{u}}\cdot\mathrm{d}\mathbf{r}=-(v-\hat{\mathbf{u}}\cdot\mathbf{v}_0)\,\mathrm{d}t_0 \tag{11}$$

若假设波源速度不超过波速，可以看出 $\mathrm{d}\mathbf{r}$ 沿 $-\hat{\mathbf{u}}$ 方向时 $\mathrm{d}t_0$ 最大．所以根据梯度的几何意义，$\boldsymbol{\nabla}t_0$ 的方向是 $-\hat{\mathbf{u}}$．为了求梯度大小，令 $\mathrm{d}\mathbf{r}=-\hat{\mathbf{u}}\,\mathrm{d}r$，代入得

$$\boldsymbol{\nabla}t_0=-\hat{\mathbf{u}}\frac{\partial t_0}{\partial r}=\frac{-\hat{\mathbf{u}}}{v-\hat{\mathbf{u}}\cdot\mathbf{v}_0} \tag{12}$$

现在带回，得

$$\omega_1=\frac{\mathrm{d}\phi_1}{\mathrm{d}t}=\omega\left(\boldsymbol{\nabla}t_0\cdot\mathbf{v}_1+\frac{\partial t_0}{\partial t}\right)=\omega\frac{v-\hat{\mathbf{u}}\cdot\mathbf{v}_1}{v-\hat{\mathbf{u}}\cdot\mathbf{v}_0} \tag{13}$$

这就是一般的普勒效应公式．

一维波动方程

预备知识 导数与差分[185]，平面波[1008]，偏导数[266]

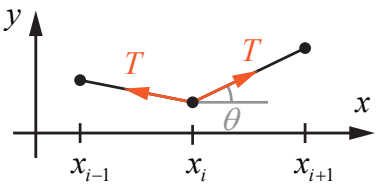


图 1: “微元法”分析弦的波动

横波

我们假设有一根无限长的弦，质量线密度为 λ，弦的**张力**（即拉力）为 T，弦静止时与 x 轴重合．假设 t 时刻的**波函数**（即弦的形状）为 $y(x,t)$，且弦的振幅较小，下面我们来求波函数所满足的微分方程，称为**一维波动方程**.

我们用“微元法”的思想，把弦划分为许多小线段，每段长度为 h，质量为 $m_i=\lambda h$，且质量都集中在左端的端点 x_i 处．下面我们来考察质点 m_i 的受力情况（图 1）．令 m_{i+1} 对 m_i 的拉力方向与 x 轴夹角为 θ，由于弦的波动较小，可以认为 θ 很小，这样来自 m_{i+1} 的拉力的两个分量为

$$\begin{cases} F_x = T\cos\theta \approx T \\ F_y = T\sin\theta \approx T\tan\theta = T\dfrac{y(x_i+h)-y(x_i)}{h} \end{cases} \tag{1}$$

同理，来自 m_{i-1} 的拉力的两个分量为

$$\begin{cases} F'_x \approx -T \\ F'_y \approx T\dfrac{y(x_i-h)-y(x_i)}{h} \end{cases} \tag{2}$$

把式 1 和式 2 相加，得 m_i 受 x 方向的合力为零，y 方向的合力为

$$F_y = T\frac{y(x_i+h)-2y(x_i)+y(x_i-h)}{h} \tag{3}$$

结合牛顿定律，并代入 $m_i=\lambda h$，有

$$\lambda\frac{\partial^2 y}{\partial t^2} = T\frac{y(x_i+h)-2y(x_i)+y(x_i-h)}{h^2} \tag{4}$$

注意到 h 很小，由式 5 [185] 可得等式右边为 $T\,\partial^2 y/\partial x^2$．于是我们得到波动方程为

$$\frac{\partial^2 y}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 y}{\partial t^2} = 0 \tag{5}$$

其中

$$v=\sqrt{\frac{T}{\lambda}} \tag{6}$$

我们稍后会看到 v 就是波的速度[13].

由于这个方程中未知函数 $y(x,t)$ 是一个多元函数，且出现了偏微分，我们把它叫做**偏微分方程**.

纵波



图 2: 一个常见的纵波模型

以上弦模型中的波动显然是一个横波，我们也可以建立一个纵波的模型. 如图 2，我们假设一条无限长的弹簧上的质量都集中于等间距的 x_i，间距为 h. 若弹簧的平均线密度为 λ，则每个质点的质量为 $m_i=\lambda h$. 为了描述弹簧的弹性，我们令单位长度弹簧的弹性系数为 k，则长为 h 的一小段弹簧弹性系数为 k/h. 若用 ξ_i 来描述 m_i 在 x 方向的位移，我们可以列出 m_i 的运动方程为

$$\lambda h\frac{\partial^2\xi}{\partial t^2}=F_i=\frac{k}{h}[\xi(x_i+h)-2\xi(x_i)+\xi(x_i-h)] \tag{7}$$

同样利用式 5 [185] 得波动方程为

$$\frac{\partial^2\xi}{\partial x^2}-\frac{\lambda}{k}\frac{\partial^2\xi}{\partial t^2}=0 \tag{8}$$

方程的解

预备知识 偏微分算符[277]

[13]有些教材也会把式 5 乘以 -1 或者乘以 c^2，导致读者分不清系数到底是 $1/c^2$ 还是 c^2. 一个简单的办法是看量纲，$\partial/\partial x$ 的量纲是长度分之一，$\partial/\partial t$ 是时间分之一，而 v 的量纲是速度（长度除以时间）.

算符分解法

我们可以先把式 5 记为算符的形式，即

$$\left(\frac{\partial^2}{\partial x^2}-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)y(x,t)=0 \tag{9}$$

也可以记为

$$\left(\frac{\partial}{\partial x}+\frac{1}{c^2}\frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial x}-\frac{1}{c^2}\frac{\partial}{\partial t}\right)y(x,t)=0 \tag{10}$$

注意两个括号的顺序可以互换．所以我们可以先寻找一阶偏微分方程

$$\left(\frac{\partial}{\partial x}\frac{1}{c^2}\frac{\partial}{\partial t}\right)y(x,t)=0 \tag{11}$$

的解得通解为

$$y(x,t)=c_1f_1(x-ct)+c_2f_2(x+ct) \tag{12}$$

其中 $f_i(x)$ 是任意函数．这个通解得物理意义很明确．即一个向右平移的波和一个向左平移的波，速度大小均为 c.

傅里叶变换法

（未完成）通解可以写成平面波的叠加（积分）.

二维波动方程

预备知识 一维波动方程[1017]，梯度[307]，散度[311]

假设我们有一个紧绷的橡皮薄膜，单位长度的张力为 λ．面密度为 σ，均为常数．取一小块微元进行受力分析，其质量为 $m=\sigma S$．要分析其受力，可以将面元的边界划分成许多小段，每小段给面元施加的竖直方向的力等于 λ 乘以法相斜率 $\boldsymbol{\nabla}u$ 再乘以该小段长度 $\mathrm{d}l$．注意只有再假设其振动幅度很小时这才成立，因为斜率是 $\tan\theta$，而严格来说我们需要 $\sin\theta$，类比同一维波动方程的推导[1017].

$$\frac{\partial^2u}{\partial t^2}=\frac{\lambda}{\sigma S}\oint[\boldsymbol{\nabla}u(\mathbf{r})]\cdot\hat{\mathbf{n}}(\mathbf{r})\,\mathrm{d}l \tag{1}$$

$\hat{\mathbf{n}}(\mathbf{r})$ 为边界在 $\mathbf{r}$ 处向外的法向量.

当 $S \to 0$，根据散度的定义

$$\frac{1}{S}\oint [\boldsymbol{\nabla} u(\mathbf{r})] \cdot \hat{\mathbf{n}}(\mathbf{r})\,\mathrm{d}l \to \boldsymbol{\nabla}\cdot(\boldsymbol{\nabla} u) = \nabla^2 u \tag{2}$$

所以波动方程为

$$\nabla^2 u - \frac{\sigma}{\lambda}\frac{\partial^2 u}{\partial t^2} = 0 \tag{3}$$

特殊地，当薄膜保持静止不动，我们就得到了拉普拉斯方程[301].

波的能量

预备知识 一维波动方程[1017]，二项式定理（非整数幂）[108]

波的能量

当机械波传播到介质中的某处时，该处原来不动的质点开始振动，因而具有**动能**，同时该处的介质也将产生形变，因而也具有**势能**.

波动传播时，介质由近及远地振动着，由此可见，能量是向外传播的. 这是波动的重要特征.

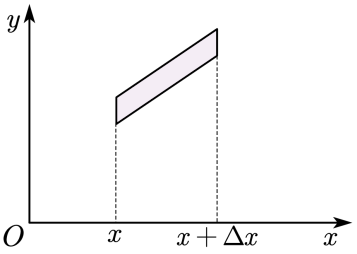


图 1: 一小段线元

如图 1 所示，在弦线上在 x 处取线元 Δx，设弦线的线密度（单位长度的质量）为 ρ_l，其质量为 $\rho_l \Delta x$. 当弦线中有平面简谐波传播时，设波函数为

$$y = A\cos\left[\omega\left(t - \frac{x}{u}\right) + \phi_0\right] \tag{1}$$

线元的动能为

$$\Delta E_{\mathrm{k}}=\frac{1}{2}\rho_l\Delta x\left(\frac{\partial y}{\partial t}\right)^2 \tag{2}$$

弦线上有张力作用，线元由原长 Δx 变为 Δl，也就是说伸长了 $\Delta l-\Delta x$．这是在张力的作用下伸长的，即它的弹性势能应等于弦张力 F 在线元伸长过程中做的功，也就是

$$\Delta E_{\mathrm{p}}=F(\Delta l-\Delta x) \tag{3}$$

在 Δx 很小时

$$\begin{aligned}\Delta l&=\sqrt{(\Delta x)^2+(\Delta y)^2}=\Delta x\left[1+\left(\frac{\Delta y}{\Delta x}\right)^2\right]^{1/2}\\&\approx\Delta x\left[1+\left(\frac{\partial y}{\partial x}\right)^2\right]^{1/2}\approx\Delta x\left[1+\frac{1}{2}\left(\frac{\partial y}{\partial x}\right)^2\right]\end{aligned} \tag{4}$$

此处的最后一个约等号运用了广义二项式定理（式 1 [108]）．因此

$$\Delta E_{\mathrm{p}}=\frac{1}{2}F\left(\frac{\partial y}{\partial x}\right)^2\Delta x \tag{5}$$

所以，线元的总机械能

$$\Delta E=\Delta E_{\mathrm{k}}+\Delta E_{\mathrm{p}}=\frac{1}{2}\rho_l\left(\frac{\partial y}{\partial t}\right)^2\Delta x+\frac{1}{2}F\left(\frac{\partial y}{\partial x}\right)^2\Delta x \tag{6}$$

对于平面简谐波有

$$\Delta E_{\mathrm{k}}=\frac{1}{2}\rho_l\Delta x\left(\frac{\partial y}{\partial t}\right)^2=\frac{1}{2}\rho_l\Delta x\omega^2A^2\sin^2\left[\omega\left(t-\frac{x}{u}\right)+\phi_0\right] \tag{7}$$

$$\Delta E_{\mathrm{p}}=\frac{1}{2}F\Delta x\left(\frac{\partial y}{\partial x}\right)^2=\frac{1}{2}F\Delta x\frac{1}{u^2}\omega^2A^2\sin^2\left[\omega\left(t-\frac{x}{u}\right)+\phi_0\right] \tag{8}$$

由于 $u=\sqrt{\dfrac{F}{\rho_l}}$，所以有

$$\Delta E_{\mathrm{p}}=\frac{1}{2}\rho_l\Delta x\omega^2A^2\sin^2\left[\omega\left(t-\frac{x}{u}\right)+\phi_0\right] \tag{9}$$

即

$$\Delta E_{\mathrm{p}}=\Delta E_{\mathrm{k}} \tag{10}$$

而总机械能为

$$\Delta E=\Delta x\rho_l\omega^2A^2\sin^2\left[\omega\left(t-\frac{x}{u}\right)+\phi_0\right] \tag{11}$$

能量密度

为了更精确地描述波在介质中的分布情况，引入能量密度的概念.

介质中单位体积的波动能量，称为**波的能量密度**，用 w 表示. 设弦线的横截面积为 S，其体密度为 ρ，它与线密度 ρ_l 的关系为 $\rho_l = \rho S$. 则能量密度

$$w = \frac{\Delta E}{S\Delta x} = \rho\omega^2 A^2 \sin^2\left[\omega\left(t - \frac{x}{u}\right) + \phi_0\right] \tag{12}$$

波的能量密度是随时间而变化的，通常取其在一个周期内的平均值，用 $\overline{w}$ 表示，称为**平均能量密度**. 因为正弦函数的平方在一个周期内的平均值为 $\frac{1}{2}$，即

$$\frac{1}{T}\int_0^T \sin^2\omega t\mathrm{d}t = \frac{1}{2} \tag{13}$$

所以能量密度在一个周期内的平均值为

$$\overline{w} = \frac{1}{2}\rho A^2\omega^2 \tag{14}$$

由上式可见，机械波的能量与振幅的平方、频率的平方都成正比.

波的强度

- 没有推导，可以以机械波为例；阻抗需要链接

预备知识 波的能量[1021]，流密度[986]

为了描述能量随着波动的进行而在介质中传播的情况，引入能流的概念.

单位时间内通过介质中某面积的能量称为通过该面积的**能流**（energy flow）. 设在某一介质中，垂直于波速 u 的方向，取面积 S，则在单位时间内通过 S 面的能量等于体积 uS 中的能量，如图 1 所示

此能量是周期性变化的，通常取其一个周期的时间平均值，得到平均能流为

$$\overline{P} = \overline{w}uS \tag{1}$$

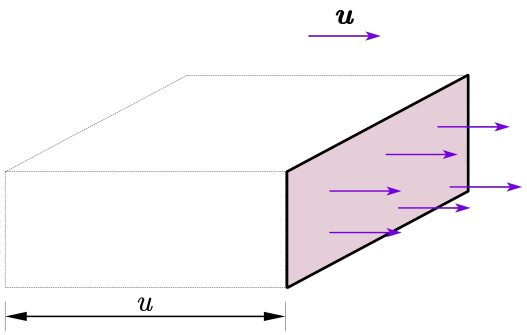


图 1: 体积 uS 内的能量在单位时间内通过 S 面

式中，$\overline{w}$ 为平均能量密度．通过与波动传播方向垂直的单位面积的平均能流，称为**平均能流密度**，又称**波的强度（intensity of wave）**，用 I 来表示，即

$$I = \overline{w}u = \frac{1}{2}\rho u\omega^2 A^2 = \frac{1}{2}Z\omega^2 A^2 \tag{2}$$

其中

$$Z = \rho u \tag{3}$$

是表征介质特性的一个常量，称为介质的**特性阻抗**．式 2 表明，弹性介质中简谐波的强度正比于振幅的平方，正比角频率（或频率）的平方，还正比于介质的特性阻抗．在国际单位制中，波的强度单位为 $\mathrm{W/m^2}$．

冲击波

当波源运动的速度 v_S 超过波的速度 u 时，这时波源将位于波前的前方．如图 1 所示．当波源在 A 位置时发出的波，在其后 t 时刻的波阵面为半径等于 ut 的球面，但此时刻波源已前进了 $v_S t$ 的距离到达 B 位置，在整个 t 时间内，波源发出的波的各波前的切面形成一个圆锥面，这锥形的顶角满足

$$\sin\alpha = \frac{ut}{v_{\mathrm{s}}t} = \frac{u}{v_{\mathrm{s}}} \tag{1}$$

随着时间的推移，各波前不断扩展，锥面也不断扩展，这种以点波源为顶点的圆锥形的波称为**冲击波（shock wave）**．v_S/u 通常称为**马赫数（Mach number）**，α 称为**马赫角（Mach angle）**．锥面就是受扰动的介质与未受扰动的介质的分界面，在两侧有着压强、密度和温度的突变．

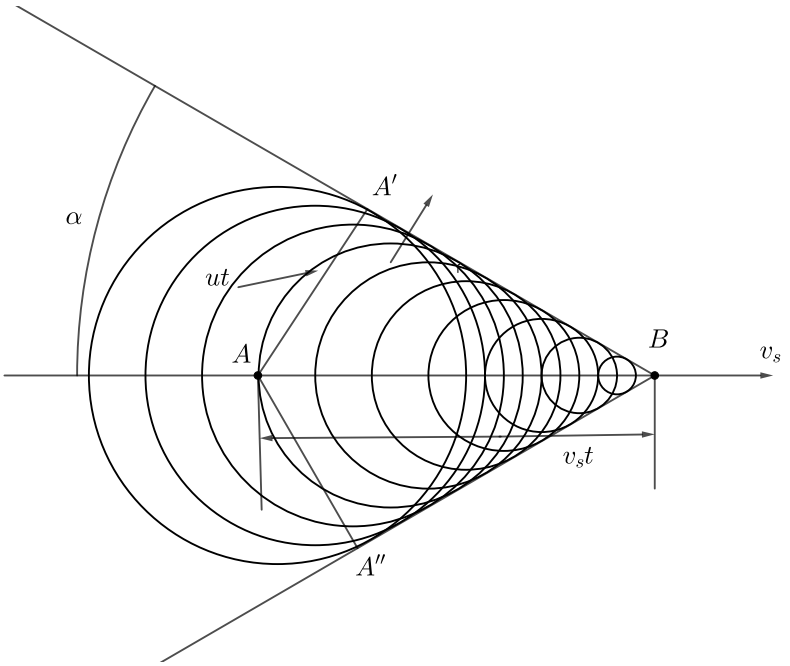


图 1: 冲击波

飞机、炮弹等以超音速飞行时，都会在空气中激起冲击波．过强的冲击波能使掠过地区的物体遭到损坏（如使玻璃窗震碎等），这种现象称为“声暴”.

第六章

中心力场问题

万有引力 引力势能

预备知识 牛顿运动定律[899]

万有引力和引力场

若两个质点质量分别为 m_1 和 m_2，位置矢量分别为 $\mathbf{r}_1$ 和 $\mathbf{r}_2$，则质点 1 对质点 2 的**万有引力（gravitational force）**为

$$\mathbf{F}_{12}=-\frac{Gm_1m_2}{r_{12}^2}\hat{\mathbf{r}}_{12}=-\frac{Gm_1m_2}{|\mathbf{r}_2-\mathbf{r}_1|^3}(\mathbf{r}_2-\mathbf{r}_1) \tag{1}$$

其中 $r_{12}=|\mathbf{r}_2-\mathbf{r}_1|$ 是两点间的距离，$\hat{\mathbf{r}}_{12}=(\mathbf{r}_2-\mathbf{r}_1)/|\mathbf{r}_2-\mathbf{r}_1|$ 是从 1 指向 2 的单位矢量．由该式，质点 2 对质点 1 的万有引力为 $\mathbf{F}_{21}=-\mathbf{F}_{12}$，符合牛顿第三定律．

我们类比高中所学电场的概念，把以上 m_1 对 m_2 的作用力看做是 m_1 在空间中产生的引力场对 m_2 的作用力．定义 m_1 产生的**引力场**为

$$\mathbf{g}(\mathbf{r})=-\frac{Gm_1}{|\mathbf{r}-\mathbf{r}_1|^3}(\mathbf{r}-\mathbf{r}_1) \tag{2}$$

若 m_1 是某天体的质量，这里的 $\mathbf{g}$ 就是它的重力加速度．可见重力加速度会随位置的不同而变化．

万有引力势能

预备知识 力场 势能[905]，球坐标系中的梯度算符[742]

在寻找万有引力的势能以前，我们先来证明所有具有 $\mathbf{F}(\mathbf{r})=F(r)\hat{\mathbf{r}}$ 形式的力场都是保守场[905]．质点延一段轨迹 $\mathcal{L}$ 从 $\mathbf{r}_1$ 移动到 $\mathbf{r}_2$ 时，力场对质点做功[903] 可以用线积分[295] 表示（以 r 作为参数）

$$W=\int_{\mathcal{L}}\mathbf{F}(\mathbf{r})\cdot\mathrm{d}\mathbf{r}=\int_{r_1}^{r_2}F(r)\,\mathrm{d}r \tag{3}$$

其中第二步是因为

$$\mathbf{F}(\mathbf{r})\cdot\mathrm{d}\mathbf{r}=F(r)\hat{\mathbf{r}}\cdot(\mathrm{d}r\,\hat{\mathbf{r}}+r\,\mathrm{d}\theta\,\hat{\boldsymbol{\theta}})=F(r)\,\mathrm{d}r \tag{4}$$

显然线积分的结果只与初末位置（与原点的距离）有关，而与路径 $\mathcal{L}$ 的选择无关，$\mathbf{F}(\mathbf{r})$ 是保守力场.

现在我们来寻找引力对应的势能. 假设质量为 M 的质点固定在坐标原点，考察质量为 m 的质点位置矢量为 $\mathbf{r}$. 由于场对物体做功等于初势能减末势能[905] 令质点沿着引力场从 $\mathbf{r}_1$ 延任意曲线移动到 $\mathbf{r}_2$，我们有

$$V(\mathbf{r}_1)-V(\mathbf{r}_2)=\int_{r_1}^{r_2}F(r)\,\mathrm{d}r=-GMm\int_{r_1}^{r_2}\frac{1}{r^2}\,\mathrm{d}r=-GMm\left(\frac{1}{r_1}-\frac{1}{r_2}\right) \tag{5}$$

可见任意位置的势能函数可以取

$$V(\mathbf{r})=V(r)=-\frac{GMm}{r} \tag{6}$$

根据势能的定义也可以给 $V(\mathbf{r})$ 加上任意常数，但习惯上我们令无穷远处势能为 0，而上式恰好满足这点. 拓展到任意具有 $\mathbf{F}(\mathbf{r})=F(r)\hat{\mathbf{r}}$ 形式的力场，其势能可以用不定积分得到

$$V=-\int F(r)\,\mathrm{d}r \tag{7}$$

这与一维的情况相同.

我们也可以反过来通过引力势能求出引力场.使用球坐标的梯度算符式 1 [742] 得

$$\mathbf{F}=-\boldsymbol{\nabla}V=GMm\left(\hat{\mathbf{r}}\frac{\partial}{\partial r}+\hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial\theta}+\hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial}{\partial\phi}\right)\frac{1}{r}=-\frac{GMm}{r^2}\hat{\mathbf{r}} \tag{8}$$

也可以使用直角坐标的梯度

$$\mathbf{F}=-\boldsymbol{\nabla}V=GMm\left(\hat{\mathbf{x}}\frac{\partial}{\partial x}+\hat{\mathbf{y}}\frac{\partial}{\partial y}+\hat{\mathbf{z}}\frac{\partial}{\partial z}\right)\frac{1}{\sqrt{x^2+y^2+z^2}} \tag{9}$$

以 x 分量为例

$$\frac{\partial}{\partial x}\left(x^2+y^2+z^2\right)^{-1/2}=-x(x^2+y^2+z^2)^{-3/2}=-\frac{x}{r^3} \tag{10}$$

另外两个分量类似可得 $-y/r^3$ 和 $-z/r^3$，代入式 9 得

$$\mathbf{F}=-\boldsymbol{\nabla}V=-\frac{GMm}{r^3}(x\hat{\mathbf{x}}+y\hat{\mathbf{y}}+z\hat{\mathbf{z}})=-\frac{GMm}{r^2}\frac{\mathbf{r}}{r}=-\frac{GMm}{r^2}\hat{\mathbf{r}} \tag{11}$$

壳层定理

预备知识 万有引力[1027]

[1]在经典力学中，**壳层定理（shell theorem）**指的是在计算一个球对称质量分布的球壳形物体产生的引力场时，我可以将其做以下等效．令该物体总质量为 M，半径为 R，要计算 $\mathbf{r}$ 处的引力场．

- 在计算球外一点（$r \geqslant R$）处的引力场时，可以将球体等效为球心处质量为 M 的质点产生的引力场．
- 在计算球内一点（$r < R$）时，球体半径大于 r 的球壳部分可以忽略．

该定理由牛顿在《自然哲学的数学原理》中首次证明．事实上，由于库仑定律和牛顿定律在形式上是一样的，这个定理可以很自然地使用电场的高斯定律[1161] 证明．以下我们讨论一些更直接的方法．

均匀圆环的引力场

为了证明该定理，我们先来看一个更简单的问题．如图 1，求一个质量为 M 半径为 R 的圆环对圆环轴线上任意一点 P 产生的引力场，已知点 P 到圆环上任意一点的连线长度为 r，与 z 轴的夹角为 θ．

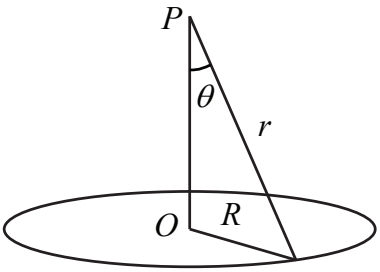


图 1: 圆环的引力场

由问题的对称性，所求引力场由 P 指向 O．若我们将圆环划分为许多小段，每小段质量为 m_i，则 m_i 在点 P 产生的引力场在 PO 方向的分量为

$$g_i = \frac{Gm_i}{r^2}\cos\theta \tag{1}$$

所以总引力场大小等于

$$g = \sum_i g_i = \frac{GM}{r^2}\cos\theta \tag{2}$$

[1]参考 Wikipedia 相关词条．

均匀球壳的引力场（几何法）

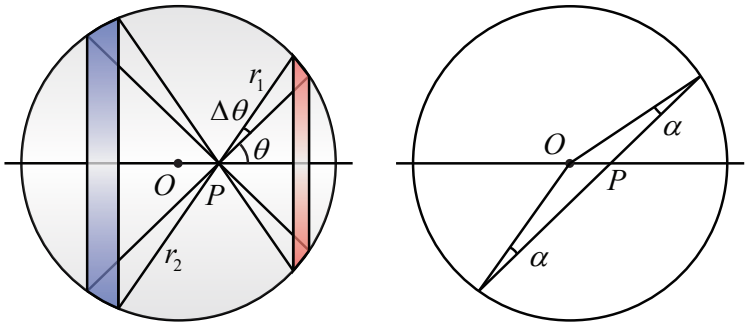


图 2: 几何法

我们现在来证明一个质量面密度为 σ 的均匀球壳在其内部一点 P 产生的引力场为零．如图 2，令球心为 O，并过 OP 作一个轴，这样球面上任意一点都对应一个张角 θ．根据不同的 θ（$0 < \theta < \pi/2$）可将球壳划分为许多对细圆环，每个圆环对应一个 $\Delta\theta$．当 $\Delta\theta \to 0$ 时，如果能证明任意一对细圆环在 P 点产生的引力场都能互相抵消，那么球壳对 P 点的总引力场就为零．

我们先要求出两个圆环的面积，以左图中右边的圆环为例，圆环的周长为 $2\pi r_1 \sin\theta$，当 $\Delta\theta \to 0$ 时，圆环的宽度为 $r_1\Delta\theta/\cos\alpha$（$\alpha$ 的定义见右图），所以圆环的面积等于周长乘以宽度，再乘以面密度 σ 得到右圆环的质量

$$M_1 = 2\pi r_1^2 \sigma \Delta\theta \sin\theta/\cos\alpha \tag{3}$$

同理，左圆环的质量为

$$M_2 = 2\pi r_2^2 \sigma \Delta\theta \sin\theta/\cos\alpha \tag{4}$$

将 M_1, r_1 和 M_2, r_2 分别代入式 2 可得 $g_1 = g_2$ 即两圆环在 P 点产生的引力场大小相等，方向相反，总引力场为零．证毕．

均匀球壳内的引力场（积分法）

下面我们通过直接积分的方法求解球壳内外的引力场分布．注意这个积分较为繁琐，不感兴趣的读者可以直接看结论．

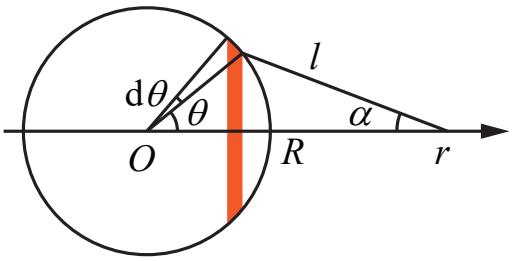


图 3: 积分法

如图 3，令场点离原点 O 的距离为 r （虽然图中 $r > R$，但 $r < R$ 时以下推导同样成立）. 和以上推导类似，图中圆环的面积为周长乘以宽度，质量为

$$\mathrm{d}M = 2\pi R\sin\theta \cdot R\,\mathrm{d}\theta \cdot \sigma = 2\pi R^2\sigma\sin\theta\,\mathrm{d}\theta \tag{5}$$

圆环在场点产生的引力场为

$$\mathrm{d}g = \frac{G\,\mathrm{d}M}{l^2}\cos\alpha \tag{6}$$

其中 $\cos\alpha = (r - R\cos\theta)/l$. 由余弦定理，$l = R^2 + r^2 - 2Rr\cos\theta$. 将 $\mathrm{d}M$，$\cos\alpha$ 和 l 代入式 6 再对 θ 作定积分得

$$g = 2\pi R^2 G\sigma\int_0^\pi \frac{(r - R\cos\theta)\sin\theta}{(R^2 + r^2 - 2Rr\cos\theta)^{3/2}}\,\mathrm{d}\theta \tag{7}$$

将上式中的定积分记为 I，使用第一类换元积分法，令 $x = \cos\theta$，得

$$\begin{aligned} I &= \int_{-1}^{1} \frac{r - Rx}{(R^2 + r^2 - 2Rrx)^{3/2}}\,\mathrm{d}x \\ &= \frac{1}{2r}\int_{-1}^{1} \frac{(r^2 - R^2) + (R^2 + r^2 - 2Rrx)}{(R^2 + r^2 - 2Rrx)^{3/2}}\,\mathrm{d}x \\ &= \frac{r^2 - R^2}{2r}\int_{-1}^{1} \frac{1}{(R^2 + r^2 - 2Rrx)^{3/2}}\,\mathrm{d}x + \frac{R^2}{2r}\int_{-1}^{1} \frac{1}{\sqrt{R^2 + r^2 - 2Rrx}}\,\mathrm{d}x \end{aligned} \tag{8}$$

这两个积分可以由“积分表[192]”中的式 2 结合式 1 得到.

$$\begin{aligned} I &= \frac{r^2 - R^2}{2r^2 R}\left(\frac{1}{|r - R|} - \frac{1}{|r + R|}\right) - \frac{1}{2r^2 R}(|r - R| - |r + R|) \\ &= \begin{cases} 2/r^2 & (r > R) \\ 0 & (r < R) \end{cases} \end{aligned} \tag{9}$$

令球壳的质量为 $M=4\pi R^2\sigma$，则将式 9 代入式 7 得引力场为

$$g=\begin{cases}GM/r^2 & (r>R)\\ 0 & (r<R)\end{cases} \tag{10}$$

当 $r<R$ 时，引力场为零，与上面的“几何法”结论一致，当 $r>R$ 时，我们发现引力场等同于球心处质量相同的质点产生的引力场.

球体的引力场

我们来考虑一个总质量为 M，半径为 R，体密度随着到球心的距离 s 而改变的球体在球内外产生的引力场. 我们可以把球体划分成任意多个质量均匀的薄球壳，若 $r>R$，每个球壳在场点的引力场都等效于球心处等质量的质点产生的引力场，所以球体的引力场仍然等于球心处等质量质点产生的引力场.

当 $r<R$ 时，所有半径大于 r 的球壳在场点的场强为零，而所有半径小于 r 的球壳在场点的引力场等于球心处等质量的质点产生的引力场. 记密度函数为 $\rho(s)$. 综上，有

$$g=\begin{cases}GM_0(r)/r^2 & (r<R)\\ GM/r^2 & (r>R)\end{cases} \tag{11}$$

其中

$$M_0(r)=\int_0^r 4\pi s^2\rho(s)\,\mathrm{d}s \tag{12}$$

均匀球体的势函数分布

最后，我们来讨论均匀球体的势函数在空间中的分布情况. 先写出其引力场表达式，只需令式 11 中的密度函数为常数 ρ 即可

$$g=\begin{cases}4G\pi\rho r/3 & (r<R)\\ GM/r^2 & (r>R)\end{cases} \tag{13}$$

记在距离球心距离为 r 的场点处，单位质量所具有的引力势能为 $V(r)$, 称之为势函数. 规定距离球心无穷远处的势能为零，则势函数的定义式为：

$$V(r)=\int_\infty^r g(r)\,\mathrm{d}r \tag{14}$$

将式 13 代入，得均匀球体的势函数分布

$$V(r)=\begin{cases}2G\pi\rho(r^2-R^2)/3-GM/r & (r<R)\\ -GM/r & (r>R)\end{cases} \tag{15}$$

开普勒三定律

预备知识 开普勒问题[1037]

开普勒定律描述了行星围绕固定中心天体的运动的规律. 第一定律给出了行星运动轨道的几何形状，第二定律描述了行星在轨道不同位置的相对速度，第三定律给出周期与轨道长轴的关系. 开普勒三定律最早是开普勒根据对太阳系中行星的观测数据总结而来的，但牛顿运用他的第二定律[899] 和万有引力定律[1027]，将开普勒三定律精确地推导了出来.

第一定律：每一个行星都沿各自的椭圆轨道环绕中心天体，中心天体则处在椭圆的一个焦点上.

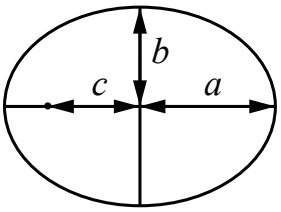


图 1: 开普勒第一定律

第二定律：相等时间内，中心天体与行星的连线所扫过的面积是相等的.

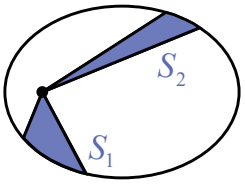


图 2: 开普勒第二定律

第三定律：椭圆轨道半长轴（a）的立方与周期（公转一圈所用的时间 T）的平方成正比.

数值模拟及证明

以下的推导中，需要忽略行星与恒星的大小，行星之间的相互引力以及中心天体的运动．在进行解析推导之前，不妨先看看更为直观的数值模拟．

- 天体运动的简单数值计算[1883]
- 开普勒第一定律的证明[1043]
- 开普勒第二和第三定律的证明[1045]

中心力场问题

预备知识 极坐标系[891]，二体系统[957]，角动量守恒（单个质点）[911]，机械能守恒（单个质点）[909]

[2]**中心力场问题**可以表述为：在惯性系中，若一个质点只受来自某固定点的力

$$\mathbf{F}(\mathbf{r}) = F(r)\hat{\mathbf{r}} \tag{1}$$

求质点的运动规律．

首先注意力场 $\mathbf{F}(\mathbf{r})$ 是一个保守场（见式 3 [1027]），所以中心力场问题也可以用势能函数 $V(r)$ 来描述（式 7 [1027]），且有

$$F(r) = -\frac{\mathrm{d}V(r)}{\mathrm{d}r} \tag{2}$$

我们已知二体系统[957] 的运动可以等效为单个质点的中心力场问题，所以在质点的中心力场问题的讨论中，只需把质点质量 m 和位矢 $\mathbf{r}$ 分别替换成约化质量 μ 和相对矢量 $\mathbf{R}$ 即可拓展到二体系统．

[2]本文参考 [10].

平面性

质点的角动量为

$$\mathbf{L}=\mathbf{r}\times m\dot{\mathbf{r}} \tag{3}$$

对角动量 $\mathbf{L}$ 关于时间 t 求导

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t}=m\dot{\mathbf{r}}\times\dot{\mathbf{r}}+m\mathbf{r}\times\ddot{\mathbf{r}}=m\mathbf{r}\times\ddot{\mathbf{r}} \tag{4}$$

由式 1 可得

$$\ddot{\mathbf{r}}=\frac{F(r)}{m}\hat{\mathbf{r}} \tag{5}$$

于是

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t}=m\mathbf{r}\times\ddot{\mathbf{r}}=r\hat{\mathbf{r}}\times F(r)\hat{\mathbf{r}}=0 \tag{6}$$

即**角动量守恒**（角动量的方向和模长均守恒）.

因为角动量垂直于位置矢量和速度矢量所在的平面，故每一时刻位置矢量和速度矢量都在同一平面内，并且加速度矢量也在此平面内. 因此，质点在中心力场中的运动是一个**平面问题**.

极坐标中的运动方程

由于式 1 中的 $F(r)$ 与位置矢量 $\mathbf{r}$ 的方向无关，在极坐标系[119] 中处理中心力场问题通常比较简单. 极坐标中质点的速度和加速度[891] 分别为

$$\dot{\mathbf{r}}=\dot{r}\hat{\mathbf{r}}+r\dot{\theta}\hat{\boldsymbol{\theta}} \tag{7}$$

$$\ddot{\mathbf{r}}=(\ddot{r}-r\dot{\theta}^2)\hat{\mathbf{r}}+\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}t}\left(r^2\dot{\theta}\right)\hat{\boldsymbol{\theta}} \tag{8}$$

由式 7 得质点的角动量在极坐标中的表示为

$$\mathbf{L}=\mathbf{r}\times(m\dot{\mathbf{r}})=mr\hat{\mathbf{r}}\times(\dot{r}\hat{\mathbf{r}}+r\dot{\theta}\hat{\boldsymbol{\theta}})=mr^2\dot{\theta}\hat{\mathbf{z}} \tag{9}$$

其中 $\hat{\mathbf{z}}$ 是垂直于极坐标平面的单位矢量（这个符号来自柱坐标系[121]）.

另外在 $\hat{\mathbf{r}}$ 方向可得

$$m(\ddot{r}-r\dot{\theta}^2)=F(r) \tag{10}$$

使用式 9 消去式 10 中的 $\dot{\theta}$，得

$$m\ddot{r}=F(r)+\frac{L^2}{mr^3} \tag{11}$$

该式被称为中心力场问题的**径向方程**.

一维等效势能与稳定轨道

由于式 11 中不含 θ，我们可以将其等效为一个一维问题，等号右侧看做等效力 $F'(r)$．求等效力的反原函数可得一维**等效势能**

$$V'(r) = V(r) + \frac{L^2}{2mr^2} \tag{12}$$

自然地，我们可以利用等效一维问题中的能量守恒列出 $r(t)$ 的一阶微分方程[3]

$$E = \frac{1}{2}m\dot{r}^2 + V'(r) = \frac{1}{2}m\dot{r}^2 + \frac{L^2}{2mr^2} + V(r) \tag{13}$$

即

$$\dot{r} = \sqrt{\frac{2}{m}[E - V'(r)]} \tag{14}$$

这是一个可分离变量的一阶常微分方程，分离变量然后两边积分得

$$t = \int_{r_0}^{r} \frac{\mathrm{d}r}{\sqrt{\frac{2}{m}[E - V'(r)]}} \tag{15}$$

积分后即可逆向得到 $r(t)$.

从一维等效势能还可以判断轨道的稳定性，我们来看一个例子

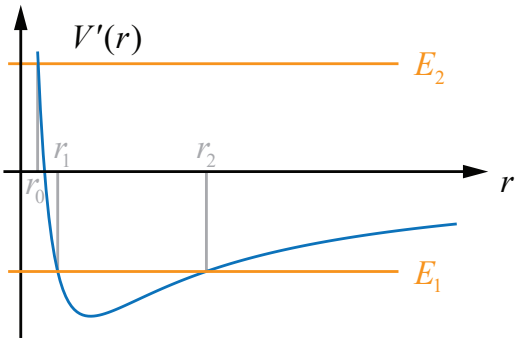


图 1: 万有引力的一维等效势能

例 1　万有引力

对万有引力，$V(r) = -GMm/r$，等效势能的大致图像如图 1．注意 $V'(r)$

[3]式 13 可以在极坐标系中直接推出，先列出 $E = m\mathbf{v}^2/2 + V(r)$，再将式 7 代入，并用式 9 消去 $\dot{\theta}$ 即可．

的形状还取决于常数 L，但根据“用极值点确定函数图像[172]”，曲线总存在一个最小值，且 $\lim\limits_{r\to\infty} V'(r)=0$，$\lim\limits_{r\to 0} V'(r)=+\infty$.

若质点具有能量 $E_2>0$，由图可得这个质点不可能一直绕力心旋转，而是从无穷远处入射，在距离 r_0 时开始远离力心，最终回到无穷远. 若质点具有能量 $E_1<0$，由图可知 r 始终在 $[r_1,r_2]$ 区间内往返变动（在“开普勒问题[1037]”中，我们将会知道 E_1 和 E_2 分别对应椭圆轨道和双曲线轨道）. 特殊地，当质点能量等于 $V'(r)$ 的最小值时，它与力心的距离将保持不变，即轨道为圆形. 若给处于圆形轨道的质点一个扰动，r 将在曲线最低点附近振动，且振动频率由最低点处曲线的二阶导数决定，我们将这种不会因为扰动而彻底改变的轨道叫做稳定轨道.

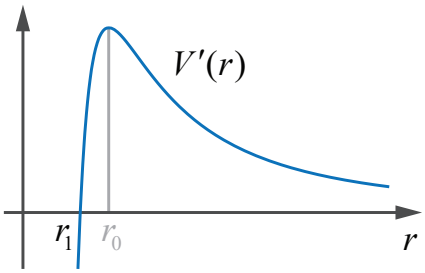


图 2: 四次方反比力的一维等效势能

例 2　四次方反比力

作为一个不稳定轨道的例子，我们来考察 $V(r)=-k/r^3$，其中 k 是一个大于零的常数. 等效势能的大致图像如图 2. 若质点的能量大于 $V'(r_0)$，则质点会从无穷远入射，穿过力心然后回到无穷远，若质点的能量小于零，它将被困在 $r<r_1$ 的圆形势阱内并不断穿过力心. 若质点的能量恰好为 $V'(r_0)$，那么它将以 r_0 为半径做圆周运动，然而任何微小的扰动都会使其从势能曲线顶端向两侧滑落，从而彻底改变轨道的性质. 我们说这样的轨道是不稳定的.

开普勒问题

预备知识 万有引力[1027]，椭圆的三种定义[133]，双曲线的三种定义[134]，抛物线的三种定义[136]

在中心力场问题[1034] 中，若 $F(r)$ 是平方反比的力（斥力为正引力为负），即

$$F(r)=\frac{k}{r^2}\qquad V(r)=\frac{k}{r} \tag{1}$$

则该问题被称为**开普勒问题**. 其中 k 为非零实数. 例如对于万有引力 $k=-GMm$，对于异种电荷间的库仑力[1123]，有[4] $k=Qq/(4\pi\epsilon_0)$.

在开普勒问题中，可以证明[1043] 质点的运动轨道是圆锥曲线的一种，力心处于焦点. 质点的机械能（动能加势能）E 和角动量 L 可以唯一地确定轨道的形状和大小. 轨道的形状一般由离心率 e 描述，大小由半通径 p 描述（式 1 [131]）. $E<0$ 对应椭圆轨道，$E=0$ 对应抛物线轨道[5]，$E>0$ 对应双曲线轨道. 注意双曲线轨道有两支，当 $k<0$（引力）时取离中心天体较近的一支，$k>0$（斥力）时取较远的一支.

$$e=\sqrt{1+\frac{2EL^2}{mk^2}} \tag{2}$$

$$p=\frac{L^2}{m|k|} \tag{3}$$

椭圆或双曲线的大小和形状也可以由参数 a,b 描述. a,b 与 e,p 的对应关系见“椭圆的三种定义[133]”和“双曲线的三种定义[134]”.

$$a=\frac{|k|}{2|E|} \tag{4}$$

$$b=\frac{L}{\sqrt{2m|E|}} \tag{5}$$

要求位置和时间的关系，见“开普勒问题的运动方程[1040]”.

证明

如果我们已知质点轨道为圆锥曲线，只需要简单的代数方法就可以得到上述关系. 而证明轨道是圆锥曲线则要复杂得多，见“普勒第一定律的证明[1043]”.

[4] 高中所学的库仑定律的系数 k 在大学物理中通常记为 $1/(4\pi\epsilon_0)$，其中 ϵ_0 为真空中的电介质常数.

[5] 显然只有引力（$k<0$）可以产生非正机械能，即椭圆轨或抛物线轨道.

椭圆轨道

令椭圆轨道（$k<0$）距离焦点的最近和最远距离分别为 r_1 和 r_2，列出总能量（动能加势能）守恒

$$\frac{1}{2}mv_1^2+\frac{k}{r_1}=\frac{1}{2}mv_2^2+\frac{k}{r_2} \tag{6}$$

以及角动量守恒

$$mv_1r_1=mv_2r_2 \tag{7}$$

把式 7 中的 v_2 代入式 6，可得

$$v_1^2=\frac{-2k/m}{r_1+r_2}\frac{r_2}{r_1} \tag{8}$$

代入式 6 的左边，并使用 $r_1+r_2=2a$ （式 10 [133]）得到总能量

$$E=\frac{k}{2a} \tag{9}$$

把式 8 代入式 7 的左边，并使用 $r_1r_2=(a+c)(a-c)=b^2$ 得角动量

$$L=b\sqrt{\frac{-mk}{a}} \tag{10}$$

将式 9 和式 10 逆转即可得到式 4 和式 5．要得到式 2 式 3，只需使用式 7 [133] 和式 8 [133] 即可．

抛物线轨道

已知抛物线轨道（$k<0$）的总能量为零，抛物线轨道离焦点的最近距离为焦距 $p/2$，该点处，动量和能量为

$$L=mv_0\frac{p}{2} \tag{11}$$

$$0=E=\frac{1}{2}mv_0^2+\frac{k}{p/2} \tag{12}$$

两式消去 v_0 得角动量为 $L=\sqrt{mkp}$．证毕．

双曲线轨道

无论 k 的正负如何，令双曲线轨道离焦点最近的距离为 r_1，可列出总能量守恒

$$\frac{1}{2}mv_0^2 = \frac{1}{2}mv_1^2 + \frac{k}{r_1} \tag{13}$$

该式左边表示质点在无穷远处的总能量，此时势能为 0，总能量等于动能．再来看角动量守恒

$$mv_0 b = mv_1 r_1 \tag{14}$$

该式左边为无穷远处的角动量．由式 11 [134] 可知，在无穷远处，双曲线的渐近线与焦点的距离为 b.

用以上两式消去 v_1，再利用 $r_1 = a - c$，得

$$E = \frac{1}{2}mv_0^2 = \frac{|k|}{2a} \tag{15}$$

再将该式的 v_0 代入式 14 左边得

$$L = b\sqrt{\frac{m|k|}{a}} \tag{16}$$

开普勒问题的运动方程

预备知识 开普勒问题[1037]，中心力场问题[1034]

若已知轨道形状，我们来计算质点在轨道上的位置如何关于时间变化．由式 15 [1034] 得

$$t = \sqrt{\frac{m}{2}} \int_{r_0}^{r} \frac{\mathrm{d}r'}{\sqrt{E - k/r' - L^2/(2mr'^2)}} \tag{1}$$

该式对任何圆锥曲线轨道都适用，其中 r_0 是轨道离中心最近的一点，令质点经过该点时 $t = 0$．把这个积分的结果 $t(r)$ 取反函数，就可以得到 $r(t)$．同理，有

$$\mathrm{d}t = \frac{mr^2}{L}\,\mathrm{d}\theta \tag{2}$$

将式 1 [131] 代入，积分得

$$t = \frac{L^3}{mk^2} \int_{\theta_0}^{\theta} \frac{\mathrm{d}\theta'}{(1 - e\cos\theta')^2} \tag{3}$$

对**抛物线**（$e=1$），有

$$t=\frac{L^3}{2mk^2}\left(\tan\frac{\theta}{2}+\frac{1}{3}\tan^3\frac{\theta}{2}\right) \tag{4}$$

对于**椭圆**（$e<1$），可以用一个参数**偏近点角（eccentric anomaly）** ψ 来代替 θ 会更方便．ψ 的定义为

$$r=a(1-e\cos\psi) \tag{5}$$

其中 a 是半长轴．当 θ 从 0 变化到 2π 时，ψ 也从 0 变化到 2π，只是速度不一样．

$$t=\sqrt{\frac{ma^3}{-k}}(\psi-e\sin\psi) \tag{6}$$

该式被称为**开普勒方程（Kepler's equation）**，开普勒第二定律也可以由该式验证．

对于**双曲线**（$k<0$），偏近点角使用式 7 定义，可取任意实数，使得

$$r=a(e\cosh\xi-1) \tag{7}$$

$$t=\sqrt{\frac{ma^3}{-k}}(e\sinh\xi-\xi) \tag{8}$$

未完成：双曲线 $k>0$ 情况是否也相同？

推导

未完成：……

拉普拉斯—龙格—楞次矢量

预备知识 开普勒问题[1037]

在开普勒问题中，我们定义**拉普拉斯—龙格—楞次矢量（Laplace-Runge-Lenz Vector）**（缩写为 **LRL 矢量**）为

$$\mathbf{A} = \mathbf{p} \times \mathbf{L} - mk\hat{\mathbf{r}} \tag{1}$$

其中 $\mathbf{p}$ 为质点动量，$\mathbf{L}$ 为轨道角动量，k 是一个常数（对万有引力 $k = GMm$，对库仑力 $k = Qq/(4\pi\epsilon_0)$．$\hat{\mathbf{r}}$ 为质点位矢 $\mathbf{r}$ 的单位矢量．在开普勒问题中，可以证明 $\mathbf{A}$ 是一个守恒量．

守恒证明

我们下面证明 $\dot{\mathbf{A}} = 0$．对式 1 求时间导数，考虑到中心力场中质点角动量 $\mathbf{L}$ 守恒，有

$$\dot{\mathbf{A}} = \dot{\mathbf{p}} \times \mathbf{L} - mk\dot{\hat{\mathbf{r}}} \tag{2}$$

其中由牛顿第二定律[899] 和万有引力定律[1027]，有

$$\dot{\mathbf{p}} = \mathbf{F} = -\frac{k}{r^2}\hat{\mathbf{r}} \tag{3}$$

又由“极坐标中单位矢量的偏导[287]”得

$$\dot{\hat{\mathbf{r}}} = \frac{\partial \hat{\mathbf{r}}}{\partial \theta}\frac{\mathrm{d}\theta}{\mathrm{d}t} = \dot{\theta}\hat{\boldsymbol{\theta}} \tag{4}$$

最后由式 9 [1034]，极坐标系中的角动量等于

$$\mathbf{L} = mr^2\dot{\theta}\hat{\mathbf{z}} \tag{5}$$

将式 3 至式 5 代入式 2 得

$$\dot{\mathbf{A}} = -\frac{k}{r^2}\hat{\mathbf{r}} \times (mr^2\dot{\theta}\hat{\mathbf{z}}) - mk\dot{\theta}\hat{\boldsymbol{\theta}} = -mk\dot{\theta}(\hat{\mathbf{r}} \times \hat{\mathbf{z}} + \hat{\boldsymbol{\theta}}) = \mathbf{0} \tag{6}$$

最后一个等号成立是因为 $\hat{\mathbf{r}} \times \hat{\mathbf{z}} = -\hat{\boldsymbol{\theta}}$，可以类比直角坐标系中的 $\hat{\mathbf{x}} \times \hat{\mathbf{z}} = -\hat{\mathbf{y}}$．证毕．

轨道方程 比耐公式

预备知识 中心力场问题[1034]

我们来看“中心力场问题[1034]”中得到的两条运动方程（式 10 和式 9）

$$\ddot{r} - r\dot{\theta}^2 = F(r)/m \tag{1}$$

$$mr^2\dot{\theta} = L \tag{2}$$

为了得到极坐标中 $r(\theta)$ 的微分方程（**轨道方程**），我们以下用式 2 消去式 1 中的 t. 首先可以把 r 看做复合函数 $r[\theta(t)]$，再用链式法则[174] 处理式 1 的第一项

$$\begin{aligned}\ddot{r} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}r}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t}\right) = \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\mathrm{d}r}{\mathrm{d}\theta}\right)\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2 + \frac{\mathrm{d}r}{\mathrm{d}\theta}\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} \\ &= \frac{\mathrm{d}^2r}{\mathrm{d}\theta^2}\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2 + \frac{\mathrm{d}r}{\mathrm{d}\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)\frac{\mathrm{d}\theta}{\mathrm{d}t}\end{aligned} \tag{3}$$

然后把式 2 代入式 1 消去所有 $\dot{\theta} = \mathrm{d}\theta/\mathrm{d}t$，得到 r 关于 θ 的微分方程

$$\frac{\mathrm{d}^2r}{\mathrm{d}\theta^2}\left(\frac{L}{r^2}\right)^2 + \frac{\mathrm{d}r}{\mathrm{d}\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{L}{r^2}\right)\frac{L}{r^2} - r\left(\frac{L}{r^2}\right)^2 = mF(r) \tag{4}$$

即

$$\frac{\mathrm{d}^2r}{\mathrm{d}\theta^2} + r^2\frac{\mathrm{d}r}{\mathrm{d}\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{1}{r^2}\right) - r = \frac{mr^4}{L^2}F(r) \tag{5}$$

这就是轨道方程. 这个方程比较复杂，但可以通过换元法化为十分简洁的形式. 令

$$u \equiv \frac{1}{r} \tag{6}$$

代入式 5，得到 u 关于 θ 的微分方程

$$\frac{\mathrm{d}^2u}{\mathrm{d}\theta^2} + u = -\frac{m}{L^2u^2}F\left(\frac{1}{u}\right) \tag{7}$$

这个二阶微分方程被称为**比耐公式**.

开普勒第一定律的证明

预备知识 开普勒三定律[1033]

行星轨道是以中心天体为焦点的任意圆锥曲线[6]．极坐标中，圆锥曲线的方程[131] 为

$$r = \frac{p}{1 - e\cos\theta} \tag{1}$$

令太阳中心天体在坐标原点，则行星沿该轨道运行．

用 LRL 矢量证明

预备知识 拉普拉斯—龙格—楞次矢量[1041]

我们先来看一种无需微分方程的推导，将开普勒问题中的 LRL 矢量 $\mathbf{A}$ 内积位矢 $\mathbf{r}$ 得

$$\mathbf{A}\cdot\mathbf{r} = (\mathbf{p}\times\mathbf{L})\cdot\mathbf{r} - mkr \tag{2}$$

由矢量混合积公式式 2 [392]，右边第一项为

$$(\mathbf{p}\times\mathbf{L})\cdot\mathbf{r} = (\mathbf{r}\times\mathbf{p})\cdot\mathbf{L} = L^2 \tag{3}$$

令 θ 为从 $\hat{\mathbf{A}}$ 转向 $\hat{\mathbf{r}}$ 的夹角（令极轴与矢量 $\mathbf{A}$ 平行，如图 1），则式 2 变为

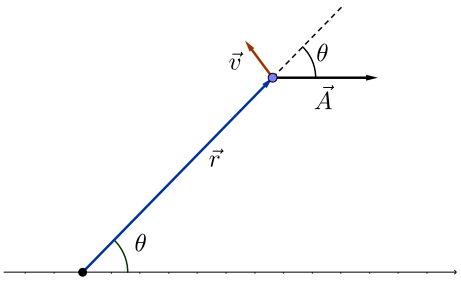


图 1: LRL 矢量与位置矢量夹角

$$Ar\cos\theta = L^2 - mkr \tag{4}$$

[6]行星轨道不一定是椭圆，也可以是抛物线或者双曲线，但是抛物线或双曲线轨道是从无穷远来到无穷远去的轨道，不会绕中心天体旋转．所以开普勒定律作为行星运动的经验公式，只描述了椭圆．

可得极坐标中的轨道为圆锥曲线的极坐标方程[7]

$$r(\theta)=\frac{p}{1+e\cos\theta} \tag{5}$$

其中通径为 $p=L^2/(mk)$，离心率为 $e=A/(mk)$.

用比耐公式证明

预备知识 比耐公式[1042]，二阶常系数非齐次微分方程的通解[218]

将平方反比力 $F(r)=-k/r^2$ 即 $F(1/u)=-ku^2$ 代入比耐公式

$$\frac{\mathrm{d}^2u}{\mathrm{d}\theta^2}+u=-\frac{m}{L^2u^2}F\left(\frac{1}{u}\right) \tag{6}$$

通解[218] 为

$$u(\theta)=\frac{1}{p}[1-e\cos(\theta+\phi_0)] \tag{7}$$

其中

$$p=\frac{L^2}{mk} \tag{8}$$

将式 7 代入 $r=1/u$，得到圆锥曲线式 1. 证毕.

平方反比斥力

当有心力从引力变为斥力时，令 $F(r)=k/r^2$ 即 $F(1/u)=ku^2$ 代入比耐公式，解得

$$u(\theta)=-\frac{1}{p}[1+e\cos(\theta+\phi_0)] \tag{9}$$

与式 7 相比，中的常数项由正号变为负号，这使得极坐标的双曲线方程表达双曲线离焦点较远的一支（见式 6 [131]）.

开普勒第二和第三定律的证明

[7]对比式 1 会发现分母的正负号反了，这相当于把圆锥曲线旋转了 180°，并不影响形状.

预备知识 开普勒三定律[1033]

第二定律

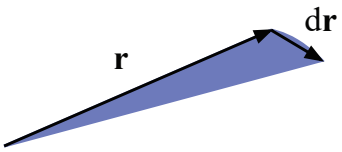


图 1: 微小时间 $\mathrm{d}t$ 内位矢扫过的面积

令质点的位矢为 $\mathbf{r}$，在很小一段时间 $\mathrm{d}t$ 内移动了 $\mathrm{d}\mathbf{r}$，于是位矢扫过的面积就是以 $\mathbf{r}$ 和 $\mathrm{d}\mathbf{r}$ 为两条边的三角形的面积

$$\mathrm{d}S = |\mathbf{r}||\mathrm{d}\mathbf{r}|(\sin\theta)/2 \tag{1}$$

其中 θ 为两条矢量的夹角. 若把面积看成矢量，方向垂直于三角形所在的平面，则根据叉乘的定义[387] 有 $\mathrm{d}\mathbf{S} = \mathbf{r} \times \mathrm{d}\mathbf{r}/2$. 两边除以 $\mathrm{d}t$，得扫过面积的速率为

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \frac{1}{2}\mathbf{r} \times \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \frac{1}{2}\mathbf{r} \times \mathbf{v} = \frac{\mathbf{L}}{2m} \tag{2}$$

其中 $\mathbf{L}$ 是质点的轨道角动量. 我们已知在中心力场问题[1034] 中，角动量守恒. 把上式记为标量形式，即

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{L}{2m} \tag{3}$$

这说明面积 S 随时间增加的速率为常数. 证毕.

第三定律

我们要试图找到开普勒问题中椭圆轨道的周期与其形状大小之间的关系. 一个直接的思路就是用椭圆的面积除以开普勒第二定律中的面积变化率得到周期. 用式 10 [1037] 将式 3 中的角动量用轨道参数 a, b 表示，有

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{L}{2m} = \frac{b}{2}\sqrt{\frac{k}{ma}} \tag{4}$$

另外已知椭圆的面积为 $S = \pi ab$，除以上式可得周期

$$T = \frac{S}{\mathrm{d}S/\mathrm{d}t} = 2\pi a^{3/2}\sqrt{\frac{m}{k}} \tag{5}$$

这就证明了轨道周期的平方和半长轴的三次方成正比.

散射

- 本词条处于草稿阶段.

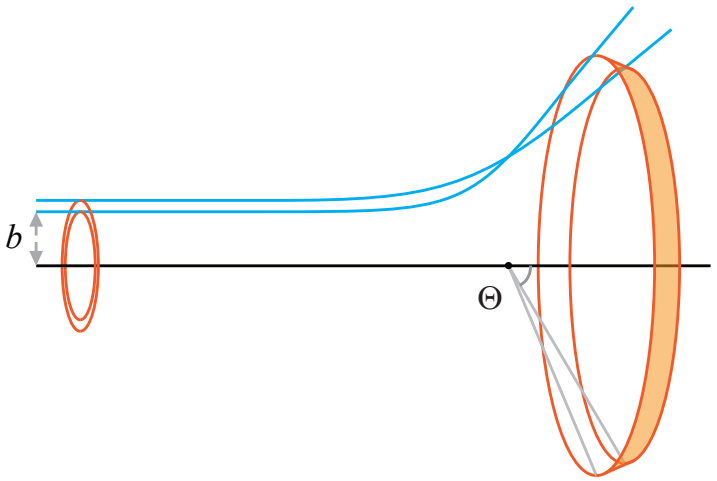


图 1: 微分截面的定义

卢瑟福散射

- 本词条处于草稿阶段.

预备知识 开普勒问题开普勒问题[1037]，散射[1047]

定义**碰撞参量**为双曲线的渐近线到焦点的距离，若轻质点一直做匀速直线运动，则碰撞参量就是两质点的最近距离．由双曲线的性质，碰撞参量等于双曲线的参数 b.

看做经典散射，微分截面等于

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}=\frac{b\,\mathrm{d}b\,\mathrm{d}\phi}{\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi}=\frac{b}{\sin\theta}\frac{\mathrm{d}b}{\mathrm{d}\theta} \tag{1}$$

由双曲线性质，偏射角满足

$$\cot\frac{\theta}{2}=\frac{b}{a} \tag{2}$$

由反开普勒问题 $E=kQq/(2a)$，消去 a 得

$$b=\frac{kQq}{2E}\cot\frac{\theta}{2} \tag{3}$$

求导代入微分截面得

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}=\left[\frac{kQq}{4E\sin^2(\theta/2)}\right]^2 \tag{4}$$

闭合轨道的条件

预备知识 比耐公式[1042]

比耐公式可记为

$$\frac{\mathrm{d}^2u}{\mathrm{d}\theta^2}+u=-\frac{m}{l^2u^2}f\left(\frac{1}{u}\right) \tag{1}$$

其中 $u=1/r$，有心力 f 向外为正，l 为轨道角动量．首先易得圆形轨道满足

$$l=-mr^3f \tag{2}$$

任何 $f<0$ 的力场都支持圆形轨道，然而根据等效一维势能，$V'=V+l^2/(2mr^2)$ 必须在最低点的二阶导数大于零轨道才能稳定．在稳定情况下，令半径为 r_0，如果给天体一个微扰，r 会呈周期性波动，我们不妨假设这个波动很小，使

$$f(r)=f(r_0)+f'(r_0)(r-r_0) \tag{3}$$

则可以证明式 1 的解为

$$u=u_0+a\cos\beta\theta \tag{4}$$

其中

$$\beta^2=3+\frac{r}{f}\frac{\mathrm{d}f}{\mathrm{d}r}\bigg|_{r=r_0} \tag{5}$$

这里 β 的意义是天体每转过一周关于 r_0 振动的次数，对于平方反比力的椭圆轨道，显然有 $\beta=1$．所以，当 β 为有理数（即 $\beta=n_1/n_2$）时，轨道是闭合的．

为了要求在任意距离的半径上轨道都闭合，β 必须不能随 r 变化，所以可以把式 5 看成 $f(r)$ 的微分方程，通解为

$$f(r)=-\frac{k}{r^{3-\beta^2}} \tag{6}$$

Bertrand 定理

我们以上只考虑了一阶微扰的情况，如果轨道与圆形轨道偏离较大，如何找到闭合条件呢？我们可以将 f 由上面的微分近似变为高阶泰勒展开，再来寻找比耐公式的解．J. Bertrand 在 1873 年证明，只有当 $\beta^2=1$ 或 $\beta^2=4$ 的时候才能满足所有可能的轨道都闭合，而这两种情况分别对应平方反比力，以及胡克定律．这个定理被称为 **Bertrand 定理**.

Bohr-Sommerfeld 氢原子模型

- 本词条处于草稿阶段.

[8]使用原子单位，氢原子 n,l,m 束缚态中，总角动量为 $L=\sqrt{l(l+1)}$，能量为 $E=-Z^2/(2n^2)$．但不知为何 Sommerfeld 模型中取 $L=l$．也就是说当 $m=l$ 时角动量可以刚延 z 方向.

Sommerfeld 模型中量子化条件为（T 表示轨道的一个周期）

$$\int_T m\dot{r}\,\mathrm{d}r=nh \tag{1}$$

即

$$2\int_{a-c}^{a+c}\sqrt{2m\left(E-\frac{k}{r}-\frac{L^2}{2mr^2}\right)}\,\mathrm{d}r=kh \tag{2}$$

若取 $L=l$ 可证明

$$n=l+k \tag{3}$$

开普勒问题

$$e=\sqrt{1+\frac{2EL^2}{mk^2}} \tag{4}$$

$$a=\frac{k}{2|E|} \tag{5}$$

[8]参考 Wikipedia 相关页面 .

Matlab 数值积分

以下程序用于数值验证式 3. 注意原子单位中 $h = 2\pi$.

```
n = 2; l = 1;
Z = 1; k = -Z;
E = -Z^2/(2*n^2);
L = l;
me = 1;
a = abs(k)/(2*abs(E)); b = L/sqrt(2*me*abs(E));
c = sqrt(a^2 - b^2);
f = @(r)real(sqrt(2*me*(E - k./r - L^2./(2*me*r.^2))));
r = linspace(a-c, a+c, 500); figure; plot(r, f(r));
I = 2*integral(f, a - c, a + c);
disp(I/(2*pi))
```

第七章

分析力学

拉格朗日方程

预备知识 牛顿第二定律[899]，自由度[930]，偏导数[266]

若一个系统的状态可以由几个独立的**广义坐标（generalized coordinates）**$q_i(t)$ $(i=1,2\ldots,N)$ 和时间 t 的函数描述，则广义坐标关于时间的函数可通过**拉格朗日方程（Lagrange equation）**[1][2]解出.

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i}=\frac{\partial L}{\partial q_i}\quad (i=1,\ldots,N) \tag{1}$$

其中 L 是系统的**拉格朗日量（lagrangian）**，等于系统动能减势能（我们先假设势能不是 $\dot{q}_i$ 的函数）. 为了方便，我们也会把 $q_1,\ldots,q_N$ 和 $\dot{q}_1,\ldots,\dot{q}_N$ 分别记为 q 和 $\dot{q}$. 那么拉格朗日量可以记为

$$L(q,\dot{q},t)=T(q,\dot{q},t)-V(q,t) \tag{2}$$

其中 T 是系统的动能，V 是势能.

拉格朗日力学与牛顿力学完全等效，其优势在于，第一，只需写出拉格朗日量，就可由简单求导得到完整的动力学微分方程而无需进行受力分析. 第二，方程的形式不随广义坐标的选取而改变. 第三，无需解方程即可以得到一些守恒量.

我们以下先通过一些例子来学习如何使用拉格朗日方程，而把方程的证明或推导留到“达朗贝尔定理[1063]”以及“哈密顿原理[1065]”中.

例 1　单个质点

若把一个质点作为系统，其直角坐标 x,y,z 就可以看做一组广义坐标，把它们看做时间的函数 $x(t),y(t),z(t)$，则三个函数完整地描述了每个时刻质点的位置. 三个函数对时间的导数（即速度的三个分量）记为 $\dot{x},\dot{y},\dot{z}$，则质点的动能为 $T=m(\dot{x}^2+\dot{y}^2+\dot{z}^2)/2$. 令质点的势能为 $V(x,y,z,t)$，拉格朗日量为

$$L(x,y,z,\dot{x},\dot{y},\dot{z},t)=T-V=\frac{1}{2}m(\dot{x}^2+\dot{y}^2+\dot{z}^2)-V(x,y,z,t) \tag{3}$$

[1] 也叫**欧拉-拉格朗日方程（Euler-Lagrange Equation）**

[2] 方程中对 $\dot{q}$ 求偏导可能会使初学者看起来有些奇怪不知道该怎么求，这里明确一下该偏导在数学上是如何操作的：可将式 2 中 L 的所有 $\dot{q}_i$ 记为 v_i，然后将 q_i 和 v_i 看作是完全没有关联的变量，求偏导时对 $q_1,\ldots,q_N,v_1,v_2\ldots,t$ 中的一个求，而保持其他量不变

现在列出 x 坐标的拉格朗日方程，把上式对 $\dot{x}$ 求偏导（$x, y, z, \dot{y}, \dot{z}, t$ 视为常数），再对时间求导得

$$\frac{\partial L}{\partial \dot{x}} = m\dot{x} \qquad \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}} = m\ddot{x} \tag{4}$$

再把拉格朗日量对 x 求偏导得

$$\frac{\partial L}{\partial x} = -\frac{\partial V}{\partial x} \tag{5}$$

以上两式代入拉格朗日方程，使用 x 方向分力与势能的关系 $F_x = -\partial V/\partial x$，得到 x 分量上的牛顿第二定律

$$m\ddot{x} = F_x \tag{6}$$

y 和 z 坐标的拉格朗日方程也可以得到对应的牛顿第二定律．三个分量的牛顿第二定律结合初始时刻的位置和速度就可以解出函数 $x(t), y(t), z(t)$．

例 2　中心力场问题

仍然以一个质点作为系统，假设质点在一个平面内运动，使用极坐标 r, θ 为广义坐标，假设势能只是 r 的函数 $V(r)$，极坐标系中质点的速度（式 3 [891]）为

$$\mathbf{v} = \dot{r}\hat{\mathbf{r}} + r\dot{\theta}\hat{\boldsymbol{\theta}} \tag{7}$$

动能等于

$$T = \frac{1}{2}m\mathbf{v}^2 = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) \tag{8}$$

类似于例 1，两个广义坐标对应的拉格朗日方程分别为

$$m\ddot{r} = mr\dot{\theta}^2 - \frac{\partial V}{\partial r} \tag{9}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(mr^2\dot{\theta}\right) = 0 \tag{10}$$

其中括号中就是 θ 的广义动量．这是一个常微分方程组，直接求解较为复杂．但注意第二条方程等号右边为零，可以先对两边不定积分得

$$mr^2\dot{\theta} = L \tag{11}$$

其中 L 为任意常数．根据角动量的定义[952]，结合式 7 不难发现 L 就是质点的角动量，所以式 10 的物理意义就是角动量守恒．

守恒量通常可以简化问题，若已知初始条件 $r(0),\theta(0),\dot{r}(0),\dot{\theta}(0)$，可先算出角动量 L，再把式 11 代入式 9 消去 $\dot{\theta}$ 得到关于 r 的单变量微分方程，进而解出 $r(t),\theta(t)$.

在该例中出现了一个**守恒量（conserved quantity）** $\partial L\Big/\partial\dot{\theta}$，原因是 θ 对应的拉格朗日方程右边为零

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial\dot{\theta}}=\frac{\partial L}{\partial\theta}=0 \tag{12}$$

即拉格朗日量与 θ 无关（或者表述成“不含 θ”）.

一般地，**若拉格朗日量与某个广义坐标 q_i 无关，则该坐标对应的广义动量 p_i 是一个守恒量**，我们把这样的坐标叫做**循环坐标（cyclic coordinate）**. **广义动量**（也叫**共轭动量，conjugate momentum**）被定义为

$$p_i=\frac{\partial L}{\partial\dot{q}_i} \tag{13}$$

例 3　耦合弹簧振子

如图，三个弹性系数为 k 的弹簧（忽略质量）将两个质量为 m 的质点串联在两个固定点之间，令一个广义坐标为中间弹簧的伸长距离 x，另一个广义坐标为两质点的质心与其平衡位置的距离 X，则两质点相对平衡位置的位移分别为

$$x_1=X-x/2\qquad x_2=X+x/2 \tag{14}$$

动能为

$$T=\frac{1}{2}m\dot{x}_1^2+\frac{1}{2}m\dot{x}_2^2=m\left(\dot{X}^2+\frac{1}{4}\dot{x}^2\right) \tag{15}$$

势能为

$$V=\frac{1}{2}kx^2+\frac{1}{2}kx_1^2+\frac{1}{2}kx_2^2=k\left(X^2+\frac{3}{4}x^2\right) \tag{16}$$

可求得拉格朗日方程组为

$$m\ddot{x}=-3kx\qquad m\ddot{X}=-kX \tag{17}$$

可见该问题中使用这种广义坐标的好处就是两个微分方程都只含有一个变量，易于求解. 由“简谐振子[913]”中的结论可得 $x(t)$ 和 $X(t)$ 的通解为

$$x(t)=x_0\cos\left(\sqrt{\frac{3k}{m}}t+\phi_x\right)\qquad X(t)=X_0\cos\left(\sqrt{\frac{k}{m}}t+\phi_X\right) \tag{18}$$

若用 x_1, x_2 坐标结合受力分析列出微分方程，则会得到两条各含两个变量的微分方程．在振动系统中，这种能使方程分离变量的广义坐标叫做简正坐标．

以上几个例子中，每个质点的坐标都是独立的，在没有约束时，D 维空间中的 N 质点有 $D \times N$ 个独立的广义坐标．在一些问题中，我们往往会遇到不同形式的**约束**，例如当两个质点之间由轻棍连接时，约束方程为 $(x_1 - x_2)^2 + (y_1 - x_2)^2 + (z_1 - z_2)^2 = L^2$ 这时两个质点的自由度[930] 就由 6 个变为 5 个．约束可以是时间的函数，例如细棍的长度可以是关于时间的函数 $L(t)$．这里我们只讨论**完整约束**，即可以表示成以下形式

$$f_i(q_1, q_2, \ldots, q_N, t) = 0 \qquad (i = 1, 2, \ldots, M) \tag{19}$$

若给 N 个自由度的系统施加 $M < N$ 个完整约束，系统的自由度就变为 $N - M$ 个自由度，并可以用 $N - M$ 个广义坐标描述．

例 4　不含时约束

两个质量为 m 的质点分别被限制在 x 轴和 y 轴上移动，且他们之间由长度为 l 的轻杆相连．x 轴的质点另与一平行于 x 轴的弹簧相连，平衡位置为 x_0，恢复系数为 k，求该系统的运动方程．

该系统只有一个自由度，我们可以把 x 轴质点的 x 坐标作为广义坐标，则由约束方程 $x^2 + y^2 = l^2$ 可得 y 轴质点的速度为

$$\dot{y} = -\frac{x\dot{x}}{\sqrt{l^2 - x^2}} \tag{20}$$

所以系统的拉格朗日量为

$$L = \frac{1}{2}m\dot{x}^2\left(1 + \frac{x^2}{l^2 - x^2}\right) + \frac{1}{2}k(x - x_0)^2 \tag{21}$$

代入拉格朗日方程即可．

例 5　含时约束

一个无穷长的细杆被固定在原点的转轴上，并在水平面上以角速度 ω 转动．一个质点可以沿细杆无摩擦运动，求质点的运动方程．

首先给细杆定义一个正方向，并取质点的离原点的距离为 r．质点速度（式 3 [891]）为 $\mathbf{v} = \dot{r}\hat{\mathbf{r}} + r\omega\hat{\boldsymbol{\theta}}$，势能为 0．系统的拉格朗日量为

$$L = T - V = \frac{1}{2}m(\dot{r}^2 + r^2\omega^2) \tag{22}$$

代入拉格朗日方程，得到运动方程

$$\ddot{r} = \omega^2 r \tag{23}$$

该函数的通解为

$$r(t) = C_1 \mathrm{e}^{\omega t} + C_2 \mathrm{e}^{-\omega t} \tag{24}$$

例 6　综合例题（包括解方程）

一根无限长杆在竖直平面绕杆上某点以角速度 ω 匀速转动，一质点可以延杆做无摩擦运动，质点与转轴由弹簧相连，弹性系数为 k，原长为 r_0，求质点的运动轨迹.

令质点离转轴的距离为 r（正方向任意选取），拉格朗日量为

$$L = T - V = \frac{1}{2}m(\dot{r}^2 + r^2\omega^2) - mgr\sin(\omega t) - \frac{1}{2}k(r - r_0)^2 \tag{25}$$

令 $\omega_0^2 = k/m - \omega^2$，由拉格朗日方程得质点的轨道方程

$$\ddot{r} + \omega_0^2 r = -g\sin(\omega t) + \frac{kr_0}{m} \tag{26}$$

这是一个二阶常系数非齐次微分方程[218]，直接套公式即可[3].

要得到齐次解，我们会发现 ω_0^2 的正负号很重要，先来讨论角速度 ω 较小的情况即 $\omega_0^2 > 0$，齐次解为

$$r(t) = C_1 r_1(t) + C_2 r_2(t) = C_1\cos(\omega_0 t) + C_2\sin(\omega_0 t) \tag{27}$$

要求非齐次解，先求朗斯基行列式

$$W(t) = \begin{vmatrix} r_1 & r_2 \\ r_1' & r_2' \end{vmatrix} = \omega_0 \tag{28}$$

$$\begin{aligned} \int \frac{r_1 f(t)}{W}\,\mathrm{d}t =& \frac{g}{\omega_0(\omega^2 - \omega_0^2)}[\omega_0\sin(\omega_0 t)\sin(\omega t) + \omega\cos(\omega_0 t)\cos(\omega t)] \\ &+ \frac{kr_0}{m\omega_0^2}\sin\omega_0 t \end{aligned} \tag{29}$$

[3] 也可以用Wolfram Alpha，例如输入 y'' + a^2*y = −g*sin(w*x) + b

$$\int \frac{r_2 f(t)}{W}\,\mathrm{d}t = \frac{-g}{\omega_0(\omega^2-\omega_0^2)}[\omega_0\cos(\omega_0 t)\sin(\omega t)-\omega\sin(\omega_0 t)\cos(\omega t)] - \frac{kr_0}{m\omega_0^2}\cos\omega_0 t \tag{30}$$

通解为

$$\begin{aligned} r(t) &= C_1 r_1 + C_2 r_2 - r_1\int\frac{r_2 f(t)}{W}\,\mathrm{d}t + r_2\int\frac{r_1 f(t)}{W}\,\mathrm{d}t \\ &= C_1\cos(\omega_0 t) + C_2\sin(\omega_0 t) + \frac{g}{\omega^2-\omega_0^2}\sin(\omega t) + \frac{kr_0}{m\omega_0^2} \end{aligned} \tag{31}$$

可见这种情况下，质点的径向运动是两个角频率分别为 ω 和 ω_0 的简谐振动的叠加.

同理，当角速度 ω 较大即 $\omega_0^2<0$ 时，通解为

$$r(t) = C_1 \mathrm{e}^{\omega_0 t} + C_2\mathrm{e}^{-\omega_0 t} + \frac{g}{\omega^2-\omega_0^2}\sin(\omega t) + \frac{kr_0}{m\omega_0^2} \tag{32}$$

这种情况下，质点一边以角频率 ω 延径向做简谐振动，一边以指数形式远离或接近原点（取决于初始条件）.

拉氏方程和极值问题

预备知识 哈密顿原理最小作用量[1065]

我们之前通过变分法得出，当位形空间中轨迹 $\{q_i(t)\}$ 的两端固定而中间变化时，要使作用量 $S[\{q_i(t)\}]$ 取极值，我们只需要解拉格朗日方程式 7 [1065]即可.

事实上，拉氏方程不仅能够解出动力学系统随时间的演化，还能解决泛函分析中一类更广的极值问题，例如著名的速降线问题. 因为 L 函数除了拉格朗日量以外，还可以是其他函数.

（未完成：举一些 Goldstein 中的例题）

贝尔特拉米等式

- 本词条处于草稿阶段.

预备知识 拉格朗日方程[1052]

贝尔特拉米等式（Beltrami identity）是指，当拉格朗日量 $L[\{q_i\},\{\dot{q}_i\},t]$ 不显含时间时，可以与一个更简单的方程等效

$$L-\dot{q}\frac{\partial L}{\partial \dot{q}}=C \tag{1}$$

其中 C 为待定常数.

最速降线问题

预备知识 拉氏方程和极值问题[1057]

[4]在一个竖直平面上任意取两个不同的点 P_1,P_2，P_1 的高度大于等于 P_2. 一个静止的质点在 P_1 处释放，在恒定的重力的作用下沿某曲线 $\mathcal{L}$ 的轨道无摩擦地运动到 P_2，耗时为 t. 若一条曲线 $\mathcal{L}$ 使得 t 最短，那么这个曲线就是 P_1,P_2 两点间的**最速降线（brachistochrone）**.

通过下文的推导可以发现，最速降线必定是摆线. 然而连接两点的摆线有无穷多条，其中只有一条是最速降线. 最速降线对应的摆线是起点参数为 $\theta=0$ 的那条，即初始斜率为无穷大.

推导

[5]令 x 轴向右，y 轴向下. 质点由原点出发，在重力加速度 g 的作用下沿函数 $y(x)$ 的曲线运动. 为书写方便令质量为 $m=1$. 由动能定理得

$$v=\sqrt{2gy} \tag{1}$$

[4]参考 Wikipedia 相关页面.

[5]参考 Wolfram MathWorld 相关页面，以及 [10] 的相关章节.

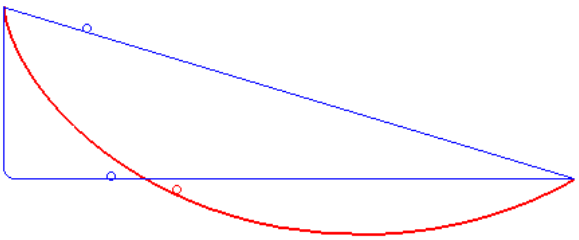


图 1: 红色：摆线；蓝色：直线和圆角折线．计算证明摆线最快．

曲线任意一点得斜率为（y 轴向下，x 轴向右）

$$\cos\theta = \frac{1}{\sqrt{1+\dot{y}^2}} \tag{2}$$

注意 $\dot{y}$ 代表 $y(x)$ 对 x 求导而不是对时间求导．x 方向速度分量为

$$\frac{\mathrm{d}x}{\mathrm{d}t} = v\cos\theta \tag{3}$$

即从 x_1 滑落到点 x_2 的时间为

$$t_{12} = \int_{x_1}^{x_2} \frac{\mathrm{d}x}{v\cos\theta} = \int_{x_1}^{x_2} \sqrt{\frac{1+\dot{y}^2}{2gy}}\,\mathrm{d}x \tag{4}$$

要求 t_{12} 关于 $y(x)$ 的极值，可以使用拉氏方程，令“拉格朗日量”为被积函数

$$L(y,\dot{y},x) = \sqrt{\frac{1+\dot{y}^2}{2gy}} \tag{5}$$

注意不显含 x，即 $\partial L/\partial x = 0$．下面来一步步代入拉格朗日方程．先把 L 对 $\dot{y}$ 求偏导，注意要把 y 看作常数

$$\frac{\partial}{\partial\dot{y}}L(y,\dot{y},x) = \frac{1}{\sqrt{2gy}}\frac{\dot{y}}{\sqrt{1+\dot{y}^2}} \tag{6}$$

然后对 x 求全导数（把 $y,\dot{y}$ 看成 x 的函数）

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial L}{\partial\dot{y}} = \frac{1}{\sqrt{2gy}\sqrt{1+\dot{y}^2}}\left(-\frac{\dot{y}^2}{2y} + \ddot{y} - \frac{\dot{y}^2\ddot{y}}{1+\dot{y}^2}\right) \tag{7}$$

同理，将 $\dot{y}$ 看作常数，对 y 求偏导得

$$\frac{\partial L}{\partial y} = -\frac{1}{2}\sqrt{\frac{1+\dot{y}^2}{2gy}}\frac{1}{y} \tag{8}$$

以上两式代入拉氏方程（式 1 [1052]），即令两式相等，并化简得

$$2y\ddot{y}+\dot{y}^2+1=0 \tag{9}$$

该方程的解为

$$\frac{y}{a}=1-\cos\left[\frac{x+\sqrt{y(2a-y)}}{a}\right] \tag{10}$$

其中 a 是大于零的参数．令方括号部分为 θ，该方程能写成 θ 的参数方程，该曲线被称为**摆线（cycloid）**或**滚轮线**．

$$\begin{cases} x=a(\theta-\sin\theta) \\ y=a(1-\cos\theta) \end{cases} \tag{11}$$

并且有

$$\dot{y}=\frac{\sin\theta}{1-\cos\theta}\qquad \ddot{y}=-\frac{1}{a(1-\cos\theta)^2} \tag{12}$$

代入式 9 可验证成立．

由于我们一开始假设了起点处 $y_1=0$，在式 11 中对应的是滚轮线的任意尖端（$\theta_1=2\pi n$）．不妨假设起点处 $\theta_1=0$，即 $x_1=0$．接下来，由于 (x_2,y_2) 必须落在曲线上，a 的值可以由此确定．

代入式 4 得质点沿最速降线滑动的时间为

$$t_{12}=\sqrt{\frac{a}{g}}(\theta_2-\theta_1)=\sqrt{\frac{a}{g}}\theta_2 \tag{13}$$

其中 θ_2 为 x_2,y_2 对应的参数．

证明摆线是等时线

摆线有另一个奇妙的性质：对于固定的 a，无论从曲线哪个位置释放质点，质点来回振动的周期都是相同的．这个结论可以用于制造精确的钟摆，由惠更斯于 1638 年在《摆钟论》中提出，比牛顿的《原理》（1686）早了约 50 年．

令质点的起始位置对应的参数为 $\theta=\theta_0$，滚动到底部时参数为 $\theta=\pi$．若

不改变坐标系，则所需时间变为[6]

$$\begin{aligned} t_{12} &= \int_{x_1}^{x_2} \sqrt{\frac{1+\dot{y}^2}{2g(y-y_0)}}\,\mathrm{d}x = \int_{x_1}^{x_2} \sqrt{\frac{2/(1-\cos\theta)}{2ga(\cos\theta_0-\cos\theta)}}\,\mathrm{d}x \\ &= \sqrt{\frac{a}{g}}\int_{\theta_0}^{\pi} \sqrt{\frac{1-\cos\theta}{\cos\theta_0-\cos\theta}}\,\mathrm{d}\theta = \pi\sqrt{\frac{a}{g}} \end{aligned} \tag{14}$$

可以发现这与 θ_0 无关，且 $\theta_0 = 0$ 时与式 13 吻合．所以从摆线的任何位置释放周期都相同．摆动周期为

$$T = 4\pi\sqrt{\frac{a}{g}} \tag{15}$$

这相当于摆长为 $2a$ 的单摆[917] 的小幅摆动周期，因为摆线底部的曲率半径[179] 是 $2a$.

单摆（大摆角）

预备知识 单摆[917]，椭圆积分[777]

令最大摆角为 θ_0，能量守恒，机械能为

$$E = \frac{1}{2}ml^2\dot{\theta}^2 - mgl\cos\theta = -mgl\cos\theta_0 \tag{1}$$

所以 θ 处的角速度为

$$\dot{\theta} = \sqrt{\frac{2g}{l}(\cos\theta - \cos\theta_0)} \tag{2}$$

令 $t = 0$ 时 $\theta = 0$ 且 $\dot{\theta} > 0$，该微分方程的解可以用椭圆积分式 3 [777] 表示

$$t(\alpha) = \sqrt{\frac{l}{2g}}\int_0^{\alpha} \frac{\mathrm{d}\theta}{\sqrt{\cos\theta-\cos\theta_0}} = \sqrt{\frac{l}{g}}\csc\frac{\theta_0}{2}F\left(\frac{\alpha}{2},\csc\frac{\theta_0}{2}\right) \qquad (0 \leqslant \alpha \leqslant \theta_0) \tag{3}$$

周期可以表示为从最低点第一次摆到最高点所需时间的 4 倍

$$T = 4t(\theta_0) = 4\sqrt{\frac{l}{g}}\csc\frac{\theta_0}{2}F\left(\frac{\theta_0}{2},\csc\frac{\theta_0}{2}\right) \tag{4}$$

[6] 最后一步积分笔者使用了 Mathematica.

Wikipedia 给出的公式为

$$T = 4\sqrt{\frac{l}{g}}F\left(\frac{\pi}{2}, \sin\frac{\theta_0}{2}\right) \tag{5}$$

已使用数值验证二者相等.

（图未完成）（未完成：周期的级数展开，$\theta(t)$ 级数解）

双摆和三摆

预备知识 拉格朗日方程[1052]

双摆

（未完成）

三摆

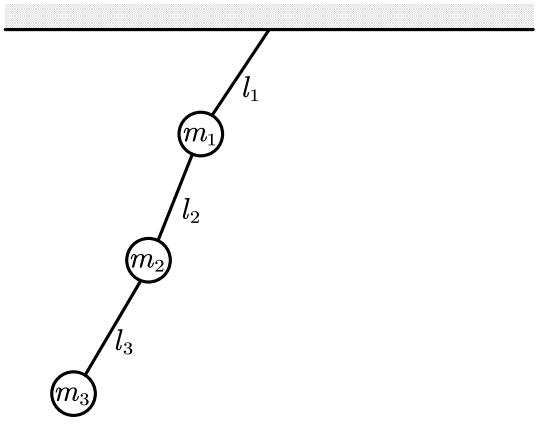


图 1: 三摆

如图 1 所示，三根质量不计的杆长度分别为 l_1, l_2, l_3，三个质点的质量分别为 m_1, m_2, m_3．我们把这个模型叫做**三摆**．三摆常用于演示物理学中的混沌现象.

运动方程

$$\begin{aligned} T = {} & \frac{1}{2}m_1(l_1\omega_1)^2 + \frac{1}{2}m_2[(l_1\omega_1\cos\theta_1 + l_2\omega_2\cos\theta_2)^2 + \\ & (l_1\omega_1\sin\theta_1 + l_2\omega_2\sin\theta_2)^2] \\ & + \frac{1}{2}m_3[(l_1\omega_1\cos\theta_1 + l_2\omega_2\cos\theta_2 + l_3\omega_3\cos\theta_3)^2 + \\ & (l_1\omega_1\sin\theta_1 + l_2\omega_2\sin\theta_2 + l_3\omega_3\sin\theta_3)^2] \end{aligned} \tag{1}$$

$$\begin{aligned} V = {} & -m_1gl_1\cos\theta_1 - m_2g(l_1\cos\theta_1 + l_2\cos\theta_2) \\ & - m_3g(l_1\cos\theta_1 + l_2\cos\theta_2 + l_3\cos\theta_3) \end{aligned} \tag{2}$$

$$L(\theta_1,\theta_2,\theta_3,\omega_1,\omega_2,\omega_3) = T(\omega_1,\omega_2,\omega_3) - V(\theta_1,\theta_2,\theta_3) \tag{3}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \omega_i} = \frac{\partial L}{\partial \theta_i} \quad (i = 1,2,3) \tag{4}$$

这是一个二阶微分方程组.

达朗贝尔定理

以下使用牛顿第二定律证明拉格朗日方程，其中会使用到一个关于约束的定理叫做达朗贝尔定理

由牛顿第二定律证明拉格朗日方程

注意以下所有的函数偏导都是把 $q_1, q_2 \ldots q_N, \dot{q}_1, \dot{q}_2 \ldots \dot{q}_N, t$ 作为变量，即对一个变量求导数而把其他变量看做常数. 另外，势能 V 和位矢 $\mathbf{r}_j$ 与 $\dot{q}_i$ 无关，偏导为 0.

系统动能为

$$T = \frac{1}{2}\sum_j m_j\mathbf{v}_j^2 = \frac{1}{2}\sum_j m_j\dot{\mathbf{r}}_j\cdot\dot{\mathbf{r}}_j \tag{1}$$

由矢量内积的求导法则

$$\frac{\partial T}{\partial \dot{q}_i} = \sum_j m_j\dot{\mathbf{r}}_j\cdot\frac{\partial \dot{\mathbf{r}}_j}{\partial \dot{q}_i} \tag{2}$$

其中质点 j 的速度可以用全导数[273] 公式

$$\dot{\mathbf{r}}_j = \sum_k \frac{\partial \mathbf{r}_j}{\partial q_k}\dot{q}_k + \frac{\partial \mathbf{r}_j}{\partial t} \tag{3}$$

对 $\dot{q}_i$ 求偏导，注意位矢与 $\dot{q}_i$ 无关，所以求偏导时 $\partial r_j/\partial q_k$ 与 $\partial r_j/\partial t$ 可看做常数.

$$\frac{\partial \dot{\mathbf{r}}_j}{\partial \dot{q}_i} = \frac{\partial \mathbf{r}_j}{\partial q_i} \tag{4}$$

代入式 2 并对时间求导得到拉格朗日方程的左边（我们暂时只讨论 $\partial V/\partial \dot{q}_i = 0$ 的情况）

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}_i} = \sum_j m_j \ddot{\mathbf{r}}_j \cdot \frac{\partial \mathbf{r}_j}{\partial q_i} + \sum_j m_j \dot{\mathbf{r}}_j \cdot \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathbf{r}_j}{\partial q_i} \tag{5}$$

拉格朗日方程的右边为

$$\frac{\partial L}{\partial q_i} = \frac{\partial T}{\partial q_i} - \frac{\partial V}{\partial q_i} \tag{6}$$

其中第一项为

$$\frac{\partial T}{\partial q_i} = \sum_j m_j \dot{\mathbf{r}}_j \cdot \frac{\partial \dot{\mathbf{r}}_j}{\partial q_i} = \sum_j m_j \dot{\mathbf{r}}_j \cdot \frac{\partial}{\partial q_i}\frac{\mathrm{d}\mathbf{r}_j}{\mathrm{d}t} \tag{7}$$

第二项被定义为**广义力**

$$Q_i = -\frac{\partial V}{\partial q_i} = \sum_j \left(-\frac{\partial V}{\partial x_j}\frac{\partial x_j}{\partial q_i} - \frac{\partial V}{\partial y_j}\frac{\partial y_j}{\partial q_i} - \frac{\partial V}{\partial z_j}\frac{\partial z_j}{\partial q_i} \right) = \sum_j \mathbf{F}_j^{(a)} \cdot \frac{\partial \mathbf{r}_j}{\partial q_i} \tag{8}$$

其中 $\mathbf{F}_j^{(a)} = -\boldsymbol{\nabla}_j V$ 被称为**非约束力**. 所以要证明拉格朗日方程，即证明式 5 等于式 7 加式 8，首先需要证明

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathbf{r}_j}{\partial q_i} = \frac{\partial}{\partial q_i}\frac{\mathrm{d}\mathbf{r}_j}{\mathrm{d}t} \tag{9}$$

也就是证明全导数和偏导数运算可对易. 使用全导数[273] 的定义，以及混合偏导[266] 的性质，有

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathbf{r}_j}{\partial q_i} = \sum_k \frac{\partial}{\partial q_k}\frac{\partial \mathbf{r}_i}{\partial q_i}\dot{q}_k + \frac{\partial}{\partial t}\frac{\partial \mathbf{r}_j}{\partial q_i} = \sum_k \frac{\partial}{\partial q_i}\frac{\partial \mathbf{r}_i}{\partial q_k}\dot{q}_k + \frac{\partial}{\partial q_i}\frac{\partial \mathbf{r}_j}{\partial t} = \frac{\partial}{\partial q_i}\frac{\mathrm{d}\mathbf{r}_j}{\mathrm{d}t} \tag{10}$$

然后我们需要证明

$$\sum_j \left(\mathbf{F}_j^{(a)} - m\ddot{\mathbf{r}}_j \right) \cdot \frac{\partial \mathbf{r}_j}{\partial q_i} = 0 \qquad (i = 1 \ldots N) \tag{11}$$

即可证明拉格朗日方程. 该式被称为**达朗贝尔定理**. 注意由于这里的 $\mathbf{F}_j^{(a)}$ 为质点 j 所受的非约束力而不是合力，所以求和项的小括号一般不为 0.

达朗贝尔定理证明

令第 j 个质点所受和力为 $\mathbf{F}_j = \mathbf{F}_j^{(a)} + \mathbf{F}_j^{(c)}$，两项分别为非约束力和约束力．由牛顿第二定律 $\mathbf{F}_j - m\ddot{\mathbf{r}}_j = 0$，所以

$$\sum_j \left(\mathbf{F}_j^{(a)} + \mathbf{F}_j^{(c)} - m\ddot{\mathbf{r}}_j\right) \cdot \frac{\partial \mathbf{r}_j}{\partial q_i} = 0 \qquad (i = 1 \dots N) \tag{12}$$

现在我们只需证明

$$\sum_j \mathbf{F}_j^{(c)} \cdot \frac{\partial \mathbf{r}_j}{\partial q_i} = 0 \qquad (i = 1 \dots N) \tag{13}$$

由于以上偏微分中时间保持不变，约束力不做功，该求和为零，证毕．

最小作用量哈密顿原理

预备知识 拉格朗日方程[1052]

若一个经典力学系统某时刻的状态完全由 N 个独立的广义坐标 $q_1 \dots q_N$ 描述（以下把 $q_1 \dots q_N$ 记为 $\{q_i\}$），那么可以把 $\{q_i\}$ 看做是 N 维空间中的一点，这个空间叫做**位形空间（configuration space）**．系统变化的过程可以看做位形空间中的一点随时间变化而走出的**轨迹（trajectory）**．若该轨迹 $\{q_i(t)\}$ 已知，可定义 t_1 与 t_2 之间系统的**作用量（action）**为

$$S[\{q_i(t)\}] = \int_{t_1}^{t_2} L[\{q_i(t)\}, \{\dot{q}_i(t)\}, t]\,\mathrm{d}t \tag{1}$$

注意 $\dot{q}_i(t)$ 代表 $q_i(t)$ 对时间求导，所以以下讨论中我们要求所有函数 $q_i(t)$ 必须在 $[t_1, t_2]$ 区间处处可导．其中 L 是拉氏量（式 2 [1052]）．

我们想要探究这么一个问题：当我们固定轨迹的起点和终点，即预先规定 t_1, t_2 时刻每个 $q_i(t)$ 的值，这时候取什么样的轨迹才能使作用量 S 取得最小值或者最大值呢？

这个问题和求一元函数的最大最小值[170]有些相似，事实上我们并不能一步到位找到“最值”，而是需要先求出若干个“极值”或“驻点”，然后在这些

解中进一步选取最大值或最小值．类比过来，泛函的极值是指上述轨迹发生微小改变（但两个端点保持不变）时泛函的值几乎不变（即小于一阶无穷小）．

我们可以把 $S[\{q_i(t)\}]$ 看成某种广义的函数，显然这个函数的函数值是一个实数，但自变量却不直接是 N 个实数，而是 N 个实函数 $q_i(t)$．我们把这种函数的函数叫做**泛函（functional）**．数学专业中，讨论泛函的课程叫做**泛函分析（functional analysis）**．

爱尔兰数学家**威廉·哈密顿（William Hamilton）**意识到，要求经典力学系统的演化（即位形空间中的轨迹），除了通过牛顿三定律，也可以通过求作用量 S 的极值得出[7]．我们把这叫做**哈密顿原理（Hamilton's principle）**．哈密顿原理也被称为**最小作用量（least action）**原理[8]．

定理 1　哈密顿原理

若一个经典力学系统在 t_1 和 t_2 时刻的坐标分别为 $\{q_i(t_1)\}$ 和 $\{q_i(t_2)\}$，那么在这段时间内所有连接这两点的轨迹中，只有一个能使作用量 S 取最值，而这个轨迹就是该系统真实演变的轨迹．

在下面的推导中我们将看到，要求这个轨迹，只需要求解拉格朗日方程式 1 [1052] 即可

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right)=\frac{\partial L}{\partial q_i}\qquad (i=1\ldots N) \tag{2}$$

由哈密顿原理导拉格朗日方程

假设满足哈密顿原理的轨迹为 $\{q_i'(t)\}$，为了让轨迹发生微小改变，现取一个变量 α 及任意 N 个一阶可导实函数 $\{\eta_i(t)\}$，令 $q_i(t,\alpha)=q_i'(t)+\alpha\eta_i(t)$．由于 α 变化的过程中仍然要保持初末时刻的 q_i 不变，$\eta_i(t)$ 必须满足 $\eta_i(t_1)=\eta_i(t_2)=0$．

现在拉格朗日量最终是 t 和 α 的函数，而作用量则完全是 α 的函数．

根据哈密顿原理，在 $\alpha=0$ 处有 $\mathrm{d}S/\mathrm{d}\alpha=0$．为书写方便，以下所有对 α 的（偏）导数都默认在 $\alpha=0$ 时求得．

$$\frac{\mathrm{d}S}{\mathrm{d}\alpha}=\int_{t_1}^{t_2}\frac{\partial}{\partial\alpha}L[\{q_i(t,\alpha)\},\{\dot{q}_i(t,\alpha)\},t]\,\mathrm{d}t \tag{3}$$

[7] 这有些类似于几何光学中的费马定理，可以代替反射和折射定律

[8] “最小”只是习惯的叫法，极值可以是极小值，极大值或鞍点．

注意 α 在时间积分中只是参数，所以可以置换偏导和积分的顺序．这里使用偏导是为了强调求导时保持 t 不变．使用偏导的链式法则[270] 有

$$\frac{\mathrm{d}S}{\mathrm{d}\alpha}=\sum_i\int_{t_1}^{t_2}\left[\frac{\partial L}{\partial q_i}\frac{\partial q_i}{\partial\alpha}+\frac{\partial L}{\partial\dot{q}_i}\frac{\partial^2 q_i}{\partial\alpha\partial t}\right]\mathrm{d}t \tag{4}$$

对第二项使用分部积分[190] 得

$$\int_{t_1}^{t_2}\frac{\partial L}{\partial\dot{q}_i}\frac{\partial^2 q_i}{\partial\alpha\partial t}\,\mathrm{d}t=\left.\frac{\partial L}{\partial\dot{q}_i}\frac{\partial q_i}{\partial\alpha}\right|_{t=t_1}^{t=t_2}-\int_{t_1}^{t_2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial\dot{q}_i}\right)\frac{\partial q_i}{\partial\alpha}\,\mathrm{d}t \tag{5}$$

其中 $\partial q_i/\partial\alpha=\eta_i$，在 t_1,t_2 时刻都为 0，第一项消失．代入式 4 得

$$\sum_i\int_{t_1}^{t_2}\left[\frac{\partial L}{\partial q_i}-\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial\dot{q}_i}\right)\right]\eta_i(t)\,\mathrm{d}t=0 \tag{6}$$

由于 $\eta_i(t)$ 可以任取，方括号内为零．要证明这点只需取 $\eta_i(t)=\delta_{ij}\delta(t-t')$ 代入即可．于是我们得到拉格朗日方程组

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial\dot{q}_i}\right)=\frac{\partial L}{\partial q_i}\qquad(i=1\ldots N) \tag{7}$$

运动积分

预备知识 拉格朗日方程[1052]，齐次函数的欧拉定理[336]

我们知道拉格朗日方程是关于广义坐标 $q_i(i=1,2,\cdots,s)$ 的二阶微分方程．在一些情况下，在系统运动过程中，存在 q_i 和 $\dot{q}_i$ 的某些函数，它们不随时间而变，这些函数称为系统的**运动积分**（integral of motion）．运动积分是通常的守恒律，如动量守恒定律、角动量守恒定律、机械能守恒定律的概念的推广．可以看出，运动积分相对于拉格朗日方程而言降了一阶，是一阶的微分方程，所以运动积分有时也称为**第一次积分**．力学中的对称性是十分重要的，而运动积分的存在与否，与系统的对称性有密切的关系．一个系统如果有尽可能多的运动积分，将对问题的求解带来极大的方便．

可遗坐标与广义动量积分

若拉格朗日函数 L 不含某个广义坐标 q_β，即 $\dfrac{\partial L}{\partial q_\beta}=0$，则这种广义坐标叫做**可遗坐标（ignorable coordinates）**，也称为**循环坐标（cyclic coordinates）**. 于是，拉格朗日方程写为

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_\beta}\right)=0 \tag{1}$$

这是说，广义动量 $p_\beta=\dfrac{\partial L}{\partial \dot{q}_\beta}$ 是守恒的，

$$p_\beta=\text{常数（如 } L \text{ 不含有 } q_\beta\text{）} \tag{2}$$

这叫做**广义动量积分**.

我们知道，若循环坐标 q_β 是系统的整体平移坐标，即拉格朗日函数 L 对于整体平移是不变的，可知广义动量积分归结为动量守恒定律. 若拉格朗日函数不包含整体转动坐标，即拉格朗日函数 L 对于整体转动是不变的（各向同性），则广义动量积分归结为角动量守恒定律.

在矢量力学中，动量守恒定律和角动量守恒定律是以牛顿第三定律为先决条件，即内力的矢量和为零、内力的力矩和为零，而**广义动量积分则并不以牛顿第三定律为先决条件**. 这点在后续讨论电磁场时十分明显，很难根据电磁场与粒子的相互作用来谈牛顿第三定律.

例 1　两个楔子的加速度

未完成：整合到“滑块和运动斜面问题（blkSlp.tex）”中

质量为 M 的光滑大楔子置于光滑的水平桌面上，质量为 m 的光滑小楔子沿着大楔子的光滑斜边滑下，如图 1 所示. 求这两个楔子的加速度.

解：大楔子可在水平方向运动，小楔子在大楔子斜边上运动系统有两个自由度.

取桌面上的固定点 O，把大楔子的质心相对于 O 点的水平坐标记作 X. 把小楔子的质心相对于大楔子斜边底端、沿着斜边计算的坐标记作 q（其实取的时候不需要一定是质心，物体上的任一点都可以）. 这个系统的广义坐标就是 X 和 q.

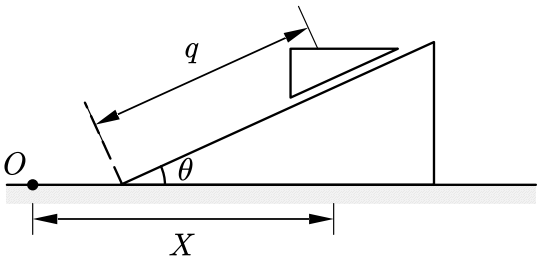


图 1: 两个楔子

主动力是两个楔子所受的重力，它们都是势力．大楔子的势能在运动过程中不变，不需要考虑，只需要讨论小楔子的势能．计算动能需注意，小楔子的速度不仅仅是沿斜边的 $\dot{q}$，而且还有随着大楔子在水平方向运动的速度 $\dot{X}$．可得：

$$\begin{aligned} T &= \frac{1}{2}M\dot{X}^2 + \frac{1}{2}m\left[v_{\text{水平}}^2 + v_{\text{竖直}}^2\right] \\ &= \frac{1}{2}M\dot{X}^2 + \frac{1}{2}m\left[(\dot{X}+\dot{q}\cos\theta)^2 + \dot{q}^2\sin^2\theta\right] \\ V &= mgq\sin\theta \\ L &= T - V = \frac{1}{2}(M+m)\dot{X}^2 + \frac{1}{2}m\dot{q}^2 + m\dot{X}\dot{q}\cos\theta - mgq\sin\theta \end{aligned} \tag{3}$$

由拉格朗日方程

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}[M\dot{X} + m(\dot{X}+\dot{q}\cos\theta)] = 0 \\ \dfrac{\mathrm{d}}{\mathrm{d}t}[m(\dot{X}\cos\theta + \dot{q})] + mg\sin\theta = 0 \end{cases} \tag{4}$$

第一个方程指出，这个系统在水平方向的动量守恒．事实上，X 是可遗坐标，所以相应的广义动量守恒．

由运动方程解得大楔子的加速度

$$\ddot{X} = \frac{mg\sin\theta\cos\theta}{M + m\sin^2\theta} \tag{5}$$

以及小楔子相对于大楔子的加速度

$$\ddot{q} = -\frac{(M+m)g\sin\theta}{M + m\sin^2\theta} \tag{6}$$

从此例题可以再次看出用拉格朗日方法解题的优越性．

广义能量积分

把拉格朗日量 $L(q,\dot{q},t)$ 对 t 求全导数得

$$\frac{\mathrm{d}L}{\mathrm{d}t}=\sum_i\frac{\partial L}{\partial q_i}\dot{q}_i+\sum_i\frac{\partial L}{\partial\dot{q}_i}\frac{\mathrm{d}\dot{q}_i}{\mathrm{d}t}+\frac{\partial L}{\partial t} \tag{7}$$

对右边第一项的偏微分使用拉格朗日方程（式 1 [1052]），得

$$\begin{aligned}\frac{\mathrm{d}L}{\mathrm{d}t}&=\sum_i\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial\dot{q}_i}\right)\dot{q}_i+\sum_i\frac{\partial L}{\partial\dot{q}_i}\frac{\mathrm{d}\dot{q}_i}{\mathrm{d}t}+\frac{\partial L}{\partial t}\\&=\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_i\frac{\partial L}{\partial\dot{q}_i}\dot{q}_i\right)+\frac{\partial L}{\partial t}\end{aligned} \tag{8}$$

即

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_i\frac{\partial L}{\partial\dot{q}_i}\dot{q}_i-L\right)=-\frac{\partial L}{\partial t} \tag{9}$$

当拉格朗日量不显含时间时，即 $\partial L/\partial t=0$，我们发现括号中的量是一个守恒量，叫做**能量函数（energy function）**[9]

$$h=\sum_i\frac{\partial L}{\partial\dot{q}_i}\dot{q}_i-L \tag{10}$$

该过程也叫做**雅可比积分（Jacobian integral）**．在一定条件下这个量是系统的总能量，详见“哈密顿正则方程[1075]”．

未完成：下面的部分需要整（删）合（掉）

我们需要清楚广义能量函数的意义．首先，势能 V 是与广义速度无关的，因此 H 的定义式 10 中的 $\dfrac{\partial L}{\partial\dot{q}_i}$．可代之以 $\dfrac{\partial T}{\partial\dot{q}_i}$．

设变换式 $\mathbf{r}_i=\mathbf{r}_i(q)$ 不显含时间，即 $\partial\mathbf{r}_i/\partial t=0$，则

$$\dot{\mathbf{r}}_i=\sum_{i=1}^{s}\frac{\partial\mathbf{r}_i}{\partial q_i}\dot{q}_i \tag{11}$$

[9] h 在数值上和以后要学的哈密顿量[1075] 相等，但哈密顿量看作是 q、广义动量 p 和 t 的函数，而能量函数看作是 $q,\dot{q},t$ 的函数．

于是

$$\begin{aligned} T &= \sum_{i=1}^{n} \frac{1}{2} m_i \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i = \sum_{i=1}^{n} \frac{1}{2} m_i \sum_{i=1}^{s} \dot{q}_i \frac{\partial \mathbf{r}_i}{\partial q_i} \cdot \sum_{\beta=1}^{s} \frac{\partial \mathbf{r}_i}{\partial q_\beta} \dot{q}_\beta \\ &= \sum_{i=1}^{n} \sum_{i=1}^{s} \sum_{\beta=1}^{s} \frac{1}{2} m_i \frac{\partial \mathbf{r}_i}{\partial q_i} \cdot \frac{\partial \mathbf{r}_i}{\partial q_\beta} \dot{q}_i \dot{q}_\beta \end{aligned} \tag{12}$$

这是广义速度的二次齐次多项式．根据齐次函数的欧拉定理，有

$$\sum_{i=1}^{s} \frac{\partial T}{\partial \dot{q}_i} \dot{q}_i = 2T \tag{13}$$

其实这也可以直接验证：

$$\begin{aligned} \frac{\partial T}{\partial \dot{q}_\gamma} &= \sum_{i=1}^{n} \sum_{i=1}^{s} \frac{1}{2} m_i \frac{\partial \mathbf{r}_i}{\partial q_i} \cdot \frac{\partial \mathbf{r}_i}{\partial q_\gamma} \dot{q}_i + \sum_{i=1}^{n} \sum_{\beta=1}^{s} \frac{1}{2} m_i \frac{\partial \mathbf{r}_i}{\partial q_\gamma} \cdot \frac{\partial \mathbf{r}_i}{\partial q_\beta} \dot{q}_\beta \\ &= \sum_{i=1}^{n} \sum_{i=1}^{s} m_i \frac{\partial \mathbf{r}_i}{\partial q_i} \cdot \frac{\partial \mathbf{r}_i}{\partial q_\gamma} \dot{q}_i \end{aligned} \tag{14}$$

于是

$$\sum_{\gamma=1}^{s} \frac{\partial T}{\partial \dot{q}_\gamma} \dot{q}_\gamma = \sum_{i=1}^{n} \sum_{i=1}^{s} \sum_{\gamma=1}^{s} m_i \frac{\partial \mathbf{r}_i}{\partial q_i} \cdot \frac{\partial \mathbf{r}_i}{\partial q_\gamma} \dot{q}_i \dot{q}_\gamma = 2T \tag{15}$$

由此，广义能量函数

$$H = \sum_{i=1}^{s} p_i \dot{q}_i - L = 2T - (T - V) = T + V \tag{16}$$

这样，在变换式 $\mathbf{r}_i = \mathbf{r}_i(q)$ 不显含时间的条件下，动能是广义速度的二次齐次式，广义能量函数 H 就是机械能．如果约束是非定常的，则变换式 $\mathbf{r}_i = \mathbf{r}_i(q, t)$ 难免显含时间．即使约束是稳定的，也可能由于选择了某些广义坐标（例如平移坐标系），变换式 $\mathbf{r}_i = \mathbf{r}_i(q, t)$ 显含时间 t．在变换式显含时间的情况下，

$$\dot{\mathbf{r}}_i = \frac{\partial \mathbf{r}_i}{\partial t} + \sum_{i=1}^{s} \frac{\partial \mathbf{r}_i}{\partial q_i} \dot{q}_i \tag{17}$$

于是

$$\begin{aligned} T &= \sum_{i=1}^{n} \frac{1}{2} m_i \left(\frac{\partial \mathbf{r}_i}{\partial t} + \sum_{i=1}^{s} \frac{\partial \mathbf{r}_i}{\partial q_i} \dot{q}_i \right) \cdot \left(\frac{\partial \mathbf{r}_i}{\partial t} + \sum_{\beta=1}^{s} \frac{\partial \mathbf{r}_i}{\partial q_\beta} \dot{q}_\beta \right) \\ &= \sum_{i=1}^{n} \left\{ \frac{1}{2} m_i \left(\frac{\partial \mathbf{r}_i}{\partial t} \right)^2 + \sum_{i=1}^{s} m_i \frac{\partial \mathbf{r}_i}{\partial t} \cdot \frac{\partial \mathbf{r}_i}{\partial q_i} \dot{q}_i \right. \\ &\quad \left. + \sum_{i=1}^{s} \sum_{\beta=1}^{s} \frac{1}{2} m_i \frac{\partial \mathbf{r}_i}{\partial q_i} \cdot \frac{\partial \mathbf{r}_i}{\partial q_\beta} \dot{q}_i \dot{q}_\beta \right\} \end{aligned} \tag{18}$$

这包含三个部分，它们分别是广义速度的零次、一次、二次的齐次多项式，分别记作 T_0, T_1, T_2，即 $T = T_0 + T_1 + T_2$．根据齐次函数的欧拉定理，

$$\sum_{i=1}^{s} \frac{\partial T}{\partial \dot{q}_i} \dot{q}_i = 0T_0 + 1T_1 + 2T_2 = T_1 + 2T_2 \tag{19}$$

由此，广义能量函数

$$H = \sum_{i=1}^{s} p_i \dot{q}_i - L = (T_1 + 2T_2) - (T_0 + T_1 + T_2 - V) = T_2 - T_0 + V \tag{20}$$

这样，在变换式显含时间的条件下，广义能量函数 H 并非机械能，但是具有机械能的量纲，故名之为**广义能量**.

例 2　直线运动汽车里的谐振子

在匀速直线运动的汽车中，有一谐振子在光滑水平槽中往返振动．取沿振动方向的坐标为 q，原点在谐振子的平衡点．选汽车为参考系，于是它是惯性系．谐振子的拉格朗日函数 $L = T - V = m\dot{q}^2/2 - kq^2/2$．易见 $\partial L/\partial t = 0$，所以 H 守恒．另一方面，由于动能 $T = m\dot{q}^2/2$ 是广义速度 $\dot{q}$ 的二次单项式，所以 H 就是机械能．诚然，$H = p\dot{q} - L = (\partial L/\partial \dot{q})\dot{q} - L = m\dot{q}^2/2 + kq^2/2 = T + V$.

改取地面为参考系，这也可认为是惯性系．如果谐振子的振动方向平行于汽车运动方向，则谐振子的 $L = T - V = m\left(\dot{q} + v_0\right)^2/2 - kq^2/2 = m\dot{q}^2/2 + mv_0\dot{q} + mv_0^2/2 - kq^2/2$，其中 v_0 是汽车的速度．因 $\partial L/\partial t = 0$，所以 H 守恒．但动能 T 不是 $\dot{q}$ 的二次齐次式，所以 H 并非机械能．实际上我们可以通过写出 H 的表达式来说明这一点：$H = p\dot{q} - L = m\left(\dot{q} + v_0\right)\dot{q} - L = m\dot{q}^2/2 - mv_0^2/2 + kq^2/2 \neq T + V$.

如果汽车并非匀速，我们考察它的匀加速运动，即速度为 at．仍以地面为参考系，则谐振子的 $m(\dot{q} + at)^2/2 - kq^2/2$，这时 $\partial L/\partial t \neq 0$，所以 H 不

守恒．另一方面，T 不是 $\dot{q}$ 的二次齐次式，所以 H 也不是机械能．实际上，$H = m\dot{q}^2/2 - ma^2t^2/2 + kq^2/2 \neq T + V$．

勒让德变换

预备知识 分部积分[190]

勒让德变换的定义很多，有的定义虽然严谨却不易理解．本节采用其最方便理解的方式来定义．

定义

单自变量勒让德变换

定义 1

给定一个实单自变量函数 $f(t)$，假设它是一个凸函数，即其导函数 $f'(t)$ 严格单调．将 $f'(t)$ 重命名为一个新的实变量 s，如果存在一个实函数 $g(s)$，使得 $g'(s) = t$，那么称 $g(s)$ 是 $f(t)$ 的勒让德变换（**Legendre transformation**）．

简单来说，勒让德变换就是有两个函数，$f(t)$ 和 $g(s)$，它们使用不同的自变量，但是各自求导以后却能得到对方的自变量：$f'(t) = s$，$g'(s) = t$．这个关系是**对称**的，因此 $f(t)$ 也是 $g(s)$ 的勒让德变换．使用勒让德变换，可以把凸函数 $f(t)$ 所表示的某些性质改用 $g(s)$ 来表示，相当于用新自变量描述同样的信息．

如何求 $f(t)$ 勒让德变换呢？如果已知 $f(t)$ 的表达式，那么就可以计算出 $f'(t) = s$，由定义，$f'(t)$ 是一个单调函数，因此存在反函数 $F(s) = t$．由于 $g'(s) = t = F(s)$，我们就可以写出：

$$g(s) = \int t\,\mathrm{d}s = \int F(s)\,\mathrm{d}s \tag{1}$$

许多情形下，我们并不知道 $f(t)$ 的显式表达，这个方法就可能并不适用．多亏了分部积分的结果，我们可以有以下简单的方式来得到一个函数的勒让德变换．

与分部积分的关系

假设 $f(t)$ 和 $g(s)$ 互为彼此的勒让德变换．由分部积分可得，

$$\begin{aligned} \mathrm{d}ts &= s\,\mathrm{d}t + t\,\mathrm{d}s \\ &= f'(t)\,\mathrm{d}t + g'(s)\,\mathrm{d}s \quad \text{（这一步由勒让德变换得）} \\ &= \mathrm{d}f + \mathrm{d}g \end{aligned} \tag{2}$$

这个结果说明，$\mathrm{d}g = \mathrm{d}ts - \mathrm{d}f$，由此得

$$g = ts - f \tag{3}$$

简单总结一下，把 $f(t)$ 变换成 $g(s)$ 的过程，就是用相互交换的两个元素 t 和 s 作乘积，再减去 f．注意 t 和 s 的关系是：$f'(t) = s$，$g'(s) = t$.

多自变量勒让德变换

把微分拓展到偏微分，我们可以自然地推广出多自变量的勒让德变换.

定义 2

给定一个实函数 $f(t_1, t_2, \cdots, t_n)$，假设对于任意 $i = 1, 2, \cdots, n$，都有 $\frac{\partial}{\partial t_i} f$ 是关于 t_i 的凸函数．将 $\frac{\partial}{\partial t_i} f$ 重命名为 s_i，如果存在 $g(s_1, s_2, \cdots, s_k, t_{k+1}, \cdots, t_n)$，使得对于任意 $i = 1, 2, \cdots k$，都有 $\frac{\partial}{\partial s_i} g = t_i$，那么称 g 是 f 替换了自变量 $t_1, \cdots, t_k$ 后的勒让德变换.

同样地，多自变量勒让德变换也可以用分部积分来求得．如果 g 是定义 2 中所说的替换了前 k 个 t_i 的 f 的勒让德变换，那么有：

$$g = \sum_{i=1,2,\cdots,k} t_i s_i - f \tag{4}$$

勒让德变换的性质

以下描述中，假设 $f_i(t)$ 的勒让德变换是 $g_i(t)$.

习题 1　标度性质

证明：

- 如果 $f_2(t)=af_1(t)$，那么 $g_2(s)=ag_1(s/a)$，
- 如果 $f_2(t)=f_1(at)$，那么 $g_2(s)=g_1(s/a)$.

推论 1　标度性质的推论

如果 $f_1(t)$ 是一个 m 次齐次函数[336]，那么 $g_1(s)$ 就是一个 n 齐次函数，其中 $1/m+1/n=1$.

习题 2　平移性质

证明：

- $f_2(t)=f_1(t)+c\Rightarrow g_2(s)=g_1(s)-c$，其中 c 是任意常数，下同；
- $f_2(t)=f_1(t+c)\Rightarrow g_2(s)=g_1(s)-cs$

习题 3　反函数性质

证明：

$f_2(t)=f_1^{-1}(t)\Rightarrow g_2(s)=-sg_1(1/p)$

哈密顿正则方程

预备知识　拉格朗日方程[1052]，勒让德变换[1073]

哈密顿量

拉格朗日力学中使用 N 个广义坐标 $q_1,\dots,q_N$ 和 N 个广义速度 $\dot{q}_1,\dots,\dot{q}_N$ 来描述力学系统的状态，我们分别简记为 q 和 $\dot{q}$. 而在哈密顿力学中，系统的状态由广义坐标和对应的 N **广义动量（generalized momentum）** $p_1,\dots,p_N$，来描述，简记为 p. 我们定义广义动量为（式 13 [1052]）

$$p_i=\frac{\partial L}{\partial \dot{q}_i} \tag{1}$$

p 也可以叫做**正则动量（canonical momentum）**或者**共轭动量（conjugate momentum）**.

定义系统的**哈密顿量（Hamiltonian）**为

$$H(q,p,t)=\sum_i \dot{q}_i p_i - L \tag{2}$$

其中 L 为拉格朗日量（式 1 [1052]）．我们在能量积分（式 10 [1067]）中见到过，但在哈密顿力学中，我们必须要把等式右边的所有 $\dot{q}$ 用 p 来表示，使哈密顿量是 (q,p,t) 的函数．

从拉格朗日函数 L 变为哈密顿量 H 的这种变换，叫做**勒让德变换（Legendre transformation）**[1073]．

哈密顿量与能量

当拉格朗日量 L 等于系统动能减势能时，哈密顿量等于系统能量．所以当我们在构建一个系统的拉格朗日量时，往往不需要做勒让德变换，而是直接用广义坐标和广义动量写出系统能量．另外，

证明

将系统看做质点系，由于 $L=T-V$ 且 V 与 $\dot{q}$ 无关，有

$$p_i=\frac{\partial L}{\partial \dot{q}_i}=\frac{\partial}{\partial \dot{q}_i}\sum_j \frac{1}{2}m_j\dot{\mathbf{r}}_j^2=\sum_j m_j\dot{\mathbf{r}}_j\frac{\partial \dot{\mathbf{r}}_j}{\partial \dot{q}_i}=\sum_j m_j\dot{\mathbf{r}}_j\frac{\partial \mathbf{r}_j}{\partial q_i} \tag{3}$$

其中最后一步利用了式 4 [1063]．所以

$$\sum_i \dot{q}_i p_i=\sum_j m_j\dot{\mathbf{r}}_j\sum_i\frac{\partial \mathbf{r}_j}{\partial q_i}\frac{\mathrm{d}q_i}{\mathrm{d}t}=\sum_j m_j\dot{\mathbf{r}}_j^2=2T \tag{4}$$

其中第二步用到了 $\mathbf{r}_j(q_1,\ldots,q_N)$ 的全微分[269]，注意该函数不能显含[270] t．上式代回式 2，有 $H=2T-(T-V)=T+V$ 即系统总能量．证毕．

哈密顿方程

哈密顿方程（Hamilton's equations）也叫**哈密顿正则方程（Hamilton's canonical equations）**是一组关于广义坐标和广义动量的微分方程组，共有 $2N$ 条方程：

$$\dot{q}_i=\frac{\partial H}{\partial p_i}\qquad \dot{p}_i=-\frac{\partial H}{\partial q_i}\qquad (i=1\ldots N) \tag{5}$$

与拉格朗日方程相比，虽然方程的个数多了一倍，但是方程却由二阶变为了一阶．可见在方程中，q 和 p 是对称的，具有同样的地位．

例 1　直角坐标系中的质点运动

直角坐标系中一个质点的拉格朗日量为 $L=m(\dot{x}^2+\dot{y}^2+\dot{z}^2)/2-V$，易得共轭动量就是通常定义的动量（用 x_1,x_2,x_3 表示 x,y,z）

$$p_{x_i}=\frac{\partial L}{\partial \dot{x}_i}=m\dot{x}_i \qquad \text{(i = 1,2,3)} \tag{6}$$

哈密顿量等于总能量（注意要写成 x_i 和 p_{x_i} 的函数，不能含有 $\dot{x}_i$）为

$$H=\sum_i \frac{p_{x_i}^2}{2m}+V \tag{7}$$

代入哈密顿方程得

$$\dot{x}_i=\frac{p_{x_i}}{m} \qquad \dot{p}_{x_i}=-\frac{\partial V}{\partial x_i} \qquad (i=1,2,3) \tag{8}$$

显然上式的第一条是（普通）动量与速度的关系，第二条则是牛顿第二定律．

例 2　中心力场问题

在式 2 [1052] 的基础上，广义坐标 $p_r=r,\theta$ 的广义动量分别为 $m\dot{r}$ 和 $L=mr^2\dot{\theta}$．可以写出质点的哈密顿量（即能量）为

$$H=T+V=\frac{1}{2}m(\dot{r}^2+r^2\dot{\theta}^2)+V(r)=\frac{p_r^2}{2m}+\frac{L^2}{2mr^2}+V(r) \tag{9}$$

代入哈密顿方程得

$$m\ddot{r}=mr\dot{\theta}^2-\frac{\partial V}{\partial r} \tag{10}$$

以及角动量守恒 $\dot{L}=0$.

哈密顿正则方程的推导

对式 2 全微分，有

$$\mathrm{d}H=\sum_i \dot{q}_i\,\mathrm{d}p_i+\sum_i p_i\,\mathrm{d}\dot{q}_i-\mathrm{d}L \tag{11}$$

对拉格朗日量全微分，有

$$\mathrm{d}L=\sum_i \frac{\partial L}{\partial q_i}\,\mathrm{d}q_i+\sum_i \frac{\partial L}{\partial \dot{q}_i}\,\mathrm{d}\dot{q}_i+\frac{\partial L}{\partial t}\,\mathrm{d}t \tag{12}$$

将拉格朗日方程（式 1 [1052]）的右边和广义动量的定义（式 13 [1052]）代入该式，得

$$\mathrm{d}L = \sum_i \dot{p}_i \,\mathrm{d}q_i + \sum_i p_i \,\mathrm{d}\dot{q}_i + \frac{\partial L}{\partial t}\,\mathrm{d}t \tag{13}$$

代入式 11，得

$$\mathrm{d}H = -\sum_i \dot{p}_i \,\mathrm{d}q_i + \sum_i \dot{q}_i \,\mathrm{d}p_i - \frac{\partial L}{\partial t}\,\mathrm{d}t \tag{14}$$

由于 H 是 p_i, q_i, t 的函数，其全微分为

$$\mathrm{d}H = \sum_i \frac{\partial H}{\partial q_i}\,\mathrm{d}q_i + \sum_i \frac{\partial H}{\partial p_i}\,\mathrm{d}p_i + \frac{\partial H}{\partial t}\,\mathrm{d}t \tag{15}$$

最后，对比以上两式可得哈密顿正则方程（式 5），以及

$$\frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t} \tag{16}$$

泊松括号

- 本词条处于草稿阶段.

预备知识 哈密顿正则方程[1075]

给定函数 $u(q,p,t)$ 和 $v(q,p,t)$，定义泊松括号为

$$\{u,v\} = \sum_i \frac{\partial u}{\partial q_i}\frac{\partial v}{\partial p_i} - \frac{\partial v}{\partial q_i}\frac{\partial u}{\partial p_i} \tag{1}$$

容易证明泊松括号满足

$$\{v,u\} = -\{u,v\} \tag{2}$$

泊松括号与守恒量

对任意不显含时的物理量 $\omega(q,p)$ 都有

$$\dot{\omega} = \sum_i \left(\frac{\partial \omega}{\partial q_i}\dot{q}_i + \frac{\partial \omega}{\partial p_i}\dot{p}_i\right) = \sum_i \left(\frac{\partial \omega}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial q_i}\frac{\partial \omega}{\partial p_i}\right) = \{\omega, H\} \tag{3}$$

所以若泊松括号恒等于零，则该物理量守恒.

同理，当 $\omega(q,p,t)$ 显含时间时有

$$\dot{\omega}=\{\omega,H\}+\frac{\partial\omega}{\partial t} \tag{4}$$

量子力学中的对易算符对应泊松括号. 该式对应量子力学中的算符平均值演化方程.

正则变换 2

- 本词条处于草稿阶段.

预备知识 泊松括号[1078]

首先回顾拉格朗日方程[1052] 为

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial\dot{q}_i}=\frac{\partial L}{\partial q_i}\qquad L(q,\dot{q},t)=T-V \tag{1}$$

正则动量为 $p_i=\partial L/\partial\dot{q}_i$，广义力为 $\partial L/\partial q_i$，拉氏方程就是广义力与正则动量的牛顿第二定律. 对于任何广义坐标，拉格朗日方程的形式不变.

勒让德变换后，得到哈密顿正则方程为

$$\dot{q}_i=\frac{\partial H}{\partial p_i}\qquad \dot{p}_i=-\frac{\partial H}{\partial q_i} \tag{2}$$

其中 $H(p,q)=T+V$. 我们通常简记 $q_1,\dots,q_N$ 为 q，简记 $\dot{q}_1,\dots,\dot{q}_N$ 为 $\dot{q}$.

正则变换

若广义坐标 q 变换到另一套广义坐标 q'，假设变换不显含时间，每个 q_k 都是新坐标的函数 $q_k(q')$，有

$$\dot{q}_i=\sum_j\frac{\partial q_i}{\partial q'_j}\dot{q}'_j \tag{3}$$

所以

$$\frac{\partial\dot{q}_k}{\partial\dot{q}'_i}=\frac{\partial q_k}{\partial q'_i} \tag{4}$$

另外可见广义速度的变换也和广义坐标有关：$\dot{q}_i = \dot{q}_i(q', \dot{q}')$.

拉格朗日量是系统的状态量，所以 $L(q', \dot{q}', t) = L[q(q'), \dot{q}(q', \dot{q}'), t]$，所以

$$p'_i = \frac{\partial L}{\partial \dot{q}'_i} = \sum_k \frac{\partial L}{\partial \dot{q}_k} \frac{\partial \dot{q}_k}{\partial \dot{q}'_i} = \sum_k \frac{\partial q_k}{\partial q'_i} p_k \tag{5}$$

这就从坐标变换推出了动量变换. 对于任何广义坐标以及对应的正则动量，哈密顿方程的形式不变（因为拉格朗日方程的形式不变，哈密顿方程是由拉格朗日方程推出来的），也有其他情况也不变. 所有使正则方程成立的坐标叫做**正则坐标（canonical coordinates）**. 下面推导判断正则坐标的一般条件.

对于不显含时的物理量 $\omega(q, p)$，有（式 3 [1078]）

$$\dot{\omega} = \{\omega, H\} = \sum_i \left(\frac{\partial \omega}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial q_i} \frac{\partial \omega}{\partial p_i} \right) \tag{6}$$

现在若把 H 看成是 $H[q'(q,p), p'(q,p)]$，

$$\frac{\partial H}{\partial p_i} = \sum_k \frac{\partial H}{\partial q'_k} \frac{\partial q'_k}{\partial p_i} + \frac{\partial H}{\partial p'_k} \frac{\partial p'_k}{\partial p_i} \tag{7}$$

$$\frac{\partial H}{\partial q_i} = \sum_k \frac{\partial H}{\partial q'_k} \frac{\partial q'_k}{\partial q_i} + \frac{\partial H}{\partial p'_k} \frac{\partial p'_k}{\partial q_i} \tag{8}$$

代入并对 H 的偏微分合并同类项得

$$\dot{\omega} = \{\omega, H\} = \sum_k \left[\frac{\partial H}{\partial q'_k} \{\omega, q'_k\} + \frac{\partial H}{\partial p'_k} \{\omega, p'_k\} \right] \tag{9}$$

注意泊松括号是对 q, p 进行偏微分，记为 $\{\}_{q,p}$. 分别代入 $\omega = q'_i, p'_i$，得到转换坐标后的哈密顿方程的一般形式. 为了保持正则方程的形式，必须要求

$$\{q'_i, q'_k\}_{q,p} = \{p'_i, p'_k\}_{q,p} = 0 \tag{10}$$

$$\{q'_i, p'_k\}_{q,p} = \delta_{ik} \tag{11}$$

这就是**判断正则变换的一般条件**.

可以证明，用任何正则坐标作为泊松括号的角标，其值都不变：

$$\{u, v\}_{q,p} = \sum_i \left(\frac{\partial u}{\partial q_i} \frac{\partial v}{\partial p_i} - \frac{\partial v}{\partial q_i} \frac{\partial u}{\partial p_i} \right) \tag{12}$$

其中

$$\frac{\partial u}{\partial q_i}\frac{\partial v}{\partial p_i}=\sum_j\left(\frac{\partial u}{\partial q'_j}\frac{\partial q'_j}{\partial q_i}+\frac{\partial u}{\partial p'_j}\frac{\partial p'_j}{\partial q_i}\right)\sum_k\left(\frac{\partial v}{\partial q'_k}\frac{\partial q'_k}{\partial p_i}+\frac{\partial v}{\partial p'_k}\frac{\partial p'_k}{\partial p_i}\right) \tag{13}$$

$$\frac{\partial v}{\partial q_i}\frac{\partial u}{\partial p_i}=\sum_k\left(\frac{\partial v}{\partial q'_k}\frac{\partial q'_k}{\partial q_i}+\frac{\partial v}{\partial p'_k}\frac{\partial p'_k}{\partial q_i}\right)\sum_j\left(\frac{\partial u}{\partial q'_j}\frac{\partial q'_j}{\partial p_i}+\frac{\partial u}{\partial p'_j}\frac{\partial p'_j}{\partial p_i}\right) \tag{14}$$

现在我们要得到 $\{u,v\}_{q',p'}=\sum_i\left(\frac{\partial u}{\partial q'_i}\frac{\partial v}{\partial p'_i}-\frac{\partial v}{\partial q'_i}\frac{\partial u}{\partial p'_i}\right)$，可以把上两式代入式 12 后对 $\frac{\partial u}{\partial q'}\frac{\partial v}{\partial p'}$ 和 $\frac{\partial v}{\partial q'_i}\frac{\partial u}{\partial p'_i}$ 合并同类项，得

$$\begin{aligned}\{u,v\}_{q,p}&=\sum_{jk}\frac{\partial u}{\partial q'_j}\frac{\partial v}{\partial p'_k}\sum_i\left(\frac{\partial q'_j}{\partial q_i}\frac{\partial p'_k}{\partial p_i}-\frac{\partial p'_k}{\partial q_i}\frac{\partial q'_j}{\partial p_i}\right)\\&\quad-\sum_{jk}\frac{\partial v}{\partial q'_k}\frac{\partial u}{\partial p'_j}\sum_i\left(\frac{\partial q'_k}{\partial q_i}\frac{\partial p'_j}{\partial p_i}-\frac{\partial p'_j}{\partial q_i}\frac{\partial q'_k}{\partial p_i}\right)\\&=\sum_{jk}\frac{\partial u}{\partial q'_j}\frac{\partial v}{\partial p'_k}\{q'_j,p'_k\}_{q,p}-\sum_{jk}\frac{\partial v}{\partial q'_k}\frac{\partial u}{\partial p'_j}\{q'_k,p'_j\}_{q,p}\end{aligned} \tag{15}$$

代入正则坐标条件（式 11 ），得

$$\begin{aligned}\{u,v\}_{q,p}&=\sum_{jk}\frac{\partial u}{\partial q'_j}\frac{\partial v}{\partial p'_k}\delta_{jk}-\sum_{jk}\frac{\partial v}{\partial q'_k}\frac{\partial u}{\partial p'_j}\delta_{jk}=\sum_j\left(\frac{\partial u}{\partial q'_j}\frac{\partial v}{\partial p'_j}-\frac{\partial v}{\partial q'_j}\frac{\partial u}{\partial p'_j}\right)\\&=\{u,v\}_{q',p'}\end{aligned} \tag{16}$$

证毕.

第八章

轨道力学

二体问题综述

预备知识 中心力场问题[1034]

由两个质量相近的普通天体组成的系统，称为“二体系统”.

一般情况下，两个天体的体积相比距离可以忽略不计，故可将天体视为质点，研究它们的相对位置的变化关系.

研究两个天体的相对运动，首先需要选取参考点，有两种常用的选择：

- 以其中一个天体为参考点，研究另一个天体的相对距离和相对速度变化；
- 以系统质心为参考点，分别研究两个天体相对质心的运动.

无外力作用下，系统质心的加速度为零，因此**由质心为原点建立的参考系是惯性参考系**，这也是最常用的参考系.

研究质点系的运动模式的方法，也有两种常用选择：

- 将每个质点的位置矢量 $\mathbf{r}$ 视为时间 t（或其他参数）的未知函数，建立运动微分方程 $\ddot{\mathbf{r}}=\mathbf{f}(t)$，求解该微分方程并代入初值条件，解出运动模式；
- 将每个质点每个时刻的位置矢量 $\mathbf{r}$ 和动量矢量 $\mathbf{v}$ 视为未知数，由动量定理、角动量定理、能量守恒等建立关于位置矢量和速度矢量的方程组，进而求解.

第一种方法最为常见，但是在天体力学和其他力学问题中，很多时候运动微分方程难以求解，因此需综合运用以上两种方法.

相似性

以质心为原点建立坐标系. 记天体 M 的质量为 m_1、位置矢量为 $\mathbf{r}_1$，天体 N 的质量为 m_2、位置矢量为 $\mathbf{r}_2$.

在此坐标系中，无外力作用下，由**动量定理**可得

$$\begin{cases} m_1\mathbf{r}_1+m_2\mathbf{r}_2=0 \\ m_1\dot{\mathbf{r}}_1+m_2\dot{\mathbf{r}}_2=0 \end{cases} \tag{1}$$

可见，两天体的位置矢量和速度矢量分别成固定的比例关系，即两天体相对质心的运动模式是相似的，只需研究其中一个天体，就能得出系统的运动规律.

平面性

记 $\mu = m_1/m_2$，系统对质心的角动量为

$$\mathbf{L} = \mathbf{r}_1 \times m_1\dot{\mathbf{r}}_1 + \mathbf{r}_2 \times \dot{\mathbf{r}}_2 = m_1(1+\mu)\mathbf{r}_1 \times \dot{\mathbf{r}}_1 \tag{2}$$

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = \dot{\mathbf{r}}_1 \times \dot{\mathbf{r}}_1 + \mathbf{r}_1 \times \ddot{\mathbf{r}}_1 = \mathbf{r}_1 \times \ddot{\mathbf{r}}_1 \tag{3}$$

由万有引力定律可得

$$\ddot{\mathbf{r}}_1 = \frac{Gm_2}{|\mathbf{r}_2 - \mathbf{r}_1|^3}(\mathbf{r}_2 - \mathbf{r}_1) = -\frac{Gm_2}{(1+\mu)^2|\mathbf{r}_1|^3}\mathbf{r}_1 \tag{4}$$

于是

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = \mathbf{r}_1 \times \ddot{\mathbf{r}}_1 = -\mathbf{r}_1 \times \frac{Gm_2}{(1+\mu)^2|\mathbf{r}_1|^3}\mathbf{r}_1 = 0 \tag{5}$$

即**角动量守恒**（角动量的方向和模长均守恒）. 因为角动量垂直于位置矢量和速度矢量所在的平面，故每一时刻位置矢量和速度矢量都在同一平面内，并且加速度矢量也在此平面内. 因此，二体系统的运动问题是一个**平面问题**.

轨道方程通解

记 $k = Gm_1m_2$，则天体 M 的拉普拉斯-龙格-楞次矢量（L-R-L 矢量）可以表示为：

$$\mathbf{A}_1 = m_1\dot{\mathbf{r}}_1 \times \mathbf{L} - \frac{m_1 k}{(1+\mu)|\mathbf{r}_1|}\mathbf{r}_1 \tag{6}$$

二体系统中，L-R-L 矢量 $\mathbf{A}_1$ 守恒（此处证明略，可参考词条“拉普拉斯-龙格-楞次矢量[1041]”）.

矢量 $\mathbf{A}_1$ 与位置矢量 r_1 作内积，并取坐标形式为极坐标

$$\begin{aligned}\mathbf{A}_1 \cdot \mathbf{r}_1 &= m_1\dot{\mathbf{r}}_1 \times \mathbf{L} \cdot \mathbf{r}_1 - \frac{m_1 k}{(1+\mu)} r_1 \\ &= m_1(\mathbf{r}_1 \times \dot{\mathbf{r}}_1) \cdot \mathbf{L} - \frac{m_1 k}{(1+\mu)} r_1 \\ &= \frac{L^2 - m_1 k}{1+\mu} r_1\end{aligned} \tag{7}$$

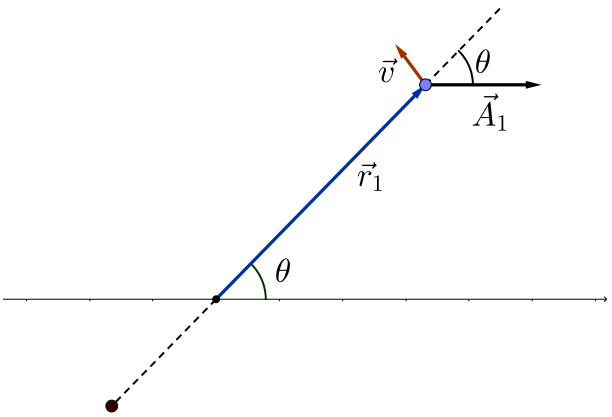


图 1: L-R-L 矢量与位置矢量

令 θ 为从矢量 $\mathbf{A}_1$ 转向位置矢量 $\mathbf{r}_1$ 的夹角（极轴与从矢量 $\mathbf{A}_1$ 平行，如图 1）. 则式 7 可化为

$$A_1 r_1 \cos\theta = \frac{L^2 - m_1 k}{1+\mu} r_1 \tag{8}$$

移项整理可得轨道的极坐标方程

$$r_1 = \frac{p}{1 + e\cos\theta} \tag{9}$$

其中，$p = L^2/(m_1 k)$，$e = A_1(1+\mu)/(m_1 k)$.

可见天体 M 的轨迹是圆锥曲线，其偏心率由 L-R-L 矢量 $\mathbf{A}_1$ 决定. 天体 N 的轨迹与之相似. 当两个天体的质量相差悬殊，则大质量的天体运动范围极小，且系统质心非常靠近大质量天体，两天体的相对运动模式则近似为单质点在引力场中的运动（即开普勒问题，恒星—行星运动模式）.

通过求解天体 M 的运动微分方程式 4 ，也可以得到相同的结果（推导过程参考“开普勒问题[1037]”和“轨道方程—比耐公式[1042]”）.

在图 2 中展示了二体系统的四种运动轨迹，分别为圆形轨迹、椭圆形轨迹、抛物线形轨迹和双曲线形轨迹.

若已知的轨迹上的一点，容易求出天体在该点处的速度矢量. 对于天体 M，可将速度矢量分解为切向速度 v_τ 与法向速度 v_n. 在极坐标系中，切向速

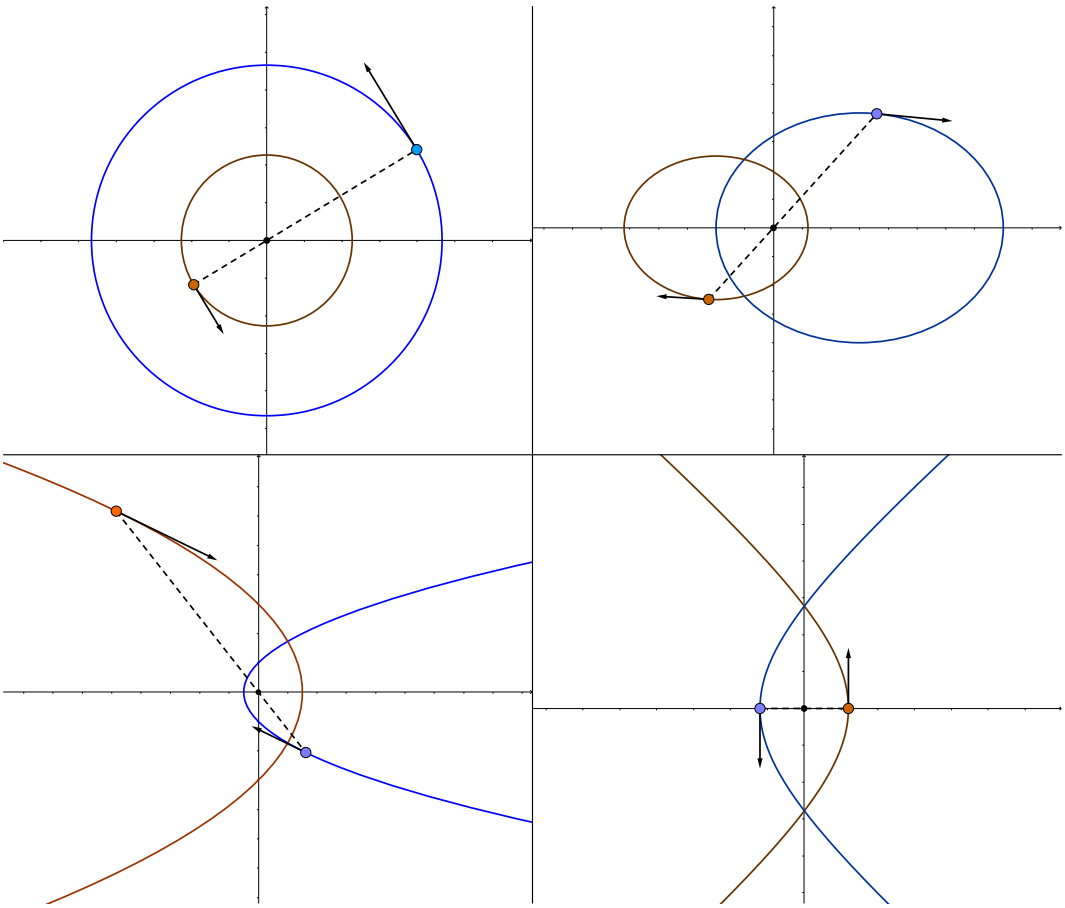


图 2: 双体系统运动轨迹

度可表示为 $v_\tau = r_1 \cdot \theta$，而上文已证明守恒的系统角动量为

$$\mathbf{L} = m_1(1+\mu)\mathbf{r}_1 \times \dot{\mathbf{r}}_1 = m_1(1+\mu)r_1^2\dot{\theta}\hat{\mathbf{z}} \tag{10}$$

因此，切向速度为

$$v_\tau = \frac{L}{m_1(1+\mu)r_1} \tag{11}$$

又因为法向速度 $v_n = \dot{\mathbf{r}}_1$，对式 9 求导可得

$$v_n = \frac{\mathrm{d}r_1}{\mathrm{d}t} = -\frac{ep\dot{\theta}\cos\theta}{(1+e\cos\theta)^2} = -\frac{epL\cos\theta}{m_1(1+\mu)r_1^2(1+e\cos\theta)^2} = -\frac{A_1\cos\theta}{m_1 L} \tag{12}$$

时间参数

通过对二体系统守恒量的分析，我们发现角动量和 L-R-L 矢量直接决定了轨道的形状和大小，再令极轴正方向与 L-R-L 矢量平行，则轨道方程在坐标系

中就有唯一确定的表达式．以上的讨论中，我们避免了求解二阶运动微分方程的繁琐过程，然而美中不足的是，我们还没有建立天体运动的位置—时间关系式．下面就对时间参数展开讨论．

式 10 通过角动量的定义式建立了轨道的角度参数和时间的关系，可以将该式改写为

$$\frac{\mathrm{d}\theta}{\mathrm{d}t}=\frac{L}{m_1(1+\mu)r_1^2} \tag{13}$$

将轨道方程式 9 代入，并分离变量

$$\mathrm{d}t=\frac{(1+\mu)L^3}{m_1k^2}\frac{\mathrm{d}\theta}{(1+e\cos\theta)} \tag{14}$$

方程两边积分，可得

$$t-t_0=\frac{(1+\mu)L^3}{m_1k^2}\int_0^{\theta}\frac{\mathrm{d}\theta}{(1+e\cos\theta)^2} \tag{15}$$

其中 t_0 为初始时间，习惯上取 $\theta=0$ 的时刻为时间的零点，即 $t_0=0$．等号右边的积分根据轨道形状有不同的表达式，以圆形和抛物线为例

$$\int_0^{\theta}\frac{\mathrm{d}\theta}{(1+e\cos\theta)^2}=\begin{cases}\theta & (e=0)\\ \dfrac{1}{2}\tan\left(\dfrac{\theta}{2}\right)+\dfrac{1}{6}\tan^3\left(\dfrac{\theta}{2}\right) & (e=1)\end{cases} \tag{16}$$

对应椭圆和双曲线的积分表达式非常复杂，有需要的读者可查阅标准数学手册积分表．

由于式 15 中的被积函数在其有效定义域上为正值，因此所得的时间 t 单调递增，即 θ 与 t 存在一一对应的关系．

综上所述，可以得出结论：**二体系统的运动问题在给定的初始条件下具有确定解**．

习题 1　关于系统机械能

试证明二体系统机械能守恒．

提示：可以选用以下两种方法

- 写出天体 M 和 N 的二阶运动微分方程，分别与其速度矢量内积；
- 系统机械能等于动能与引力势能的和，写出机械能表达式并适当代换，最终可用角动量和 L-R-L 矢量表达系统机械能．

轨道参数 时间变量

预备知识 开普勒第一定律的证明[1043]

在“开普勒第一定律的证明”中，我们得到了极坐标下的轨道方程

$$r = \frac{p}{1 + e\cos\theta} \tag{1}$$

其中，$p = L^2/(mk)$，$e = A/(mk)$.

对于轨迹上任意一点，容易求出质点在该处的速度矢量. 可将速度矢量分解为切向速度 v_τ 与法向速度 v_n. 在极坐标系中，切向速度可表示为 $v_\tau = r\dot{\theta}$，角动量为 $\mathbf{L} = mr^2\dot{\theta}\hat{\mathbf{z}}$，因此，切向速度为

$$v_\tau = \frac{L}{mr} \tag{2}$$

另有法向速度 $v_n = \dot{r}$，对式 1 求导可得

$$v_n = \frac{\mathrm{d}r}{\mathrm{d}t} = -\frac{ep\dot{\theta}\sin\theta}{(1 + e\cos\theta)^2} = -\frac{epL\sin\theta}{mr^2(1 + e\cos\theta)^2} = -\frac{A\sin\theta}{mL} \tag{3}$$

初值条件

由轨道方程式 1 可知，轨道上 $\theta = 0$ 的点距离坐标原点最近，并且由式 2 和 式 3 可以发现，该点处的切向速度最大，法向速度为零. 在恒星-行星系统中，称这个最近点为“**近日点**”，通常取该点的运动状态为初值条件. 记这个最近距离为 r_0，该点处速度为 v_0. 由此初值条件可得

$$L = mv_0r_0 \tag{4}$$

$$A = m^2v_0^2r_0 - mk = m^2\left(v_0^2r_0 - GM\right) \tag{5}$$

$$e = \frac{A}{mk} = \frac{v_0^2r_0}{GM} - 1 \tag{6}$$

轨道上任意一点处，质点所具有的机械能为

$$E = \frac{1}{2}m(v_\tau + v_n)^2 - \frac{k}{r} \tag{7}$$

由于系统不受外力或外力矩作用，故机械能守恒，由初值条件亦可求出系统机械能

$$\begin{aligned} E &= \frac{1}{2}mv_0^2 - \frac{k}{r_0} \\ &= \frac{L^2}{2mr_0^2} - \frac{k}{r_0} \\ &= \frac{L^2}{2m}\frac{(1+e)^2)}{p^2} - \frac{k(1+e)}{p} \\ &= \frac{mk^2}{2L^2}\left(e^2-1\right) \end{aligned} \tag{8}$$

式 8 与“开普勒问题[1037]”中式 2 所给出的轨道离心率与能量、角动量的关系式一致.

由此可见，初值条件可以唯一确定轨道的形状、大小等几何参数.

时间变量

在许多实际应用中，往往需要确定天体位置与时间的关系. 比如常见的椭圆轨道，我们不仅需要知道周期，还需要计算天体运行至任意位置的时间. 下面就对时间参数展开讨论.

可将极坐标下的角动量表达式 $L = mr^2\dot{\theta}$ 改写为

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{L}{mr^2} \tag{9}$$

将轨道方程式 1 代入，并分离变量

$$\mathrm{d}t = \frac{L^3}{mk^2}\frac{\mathrm{d}\theta}{(1+e\cos\theta)} \tag{10}$$

方程两边积分，可得

$$t - t_0 = \frac{L^3}{mk^2}\int_0^\theta \frac{\mathrm{d}\theta}{(1+e\cos\theta)^2} \tag{11}$$

其中 t_0 为初始时间，习惯上取 $\theta = 0$ 的时刻为时间的零点，即 $t_0 = 0$. 等号右边的积分根据轨道形状有不同的形式.

圆轨道

圆轨道是最简单的情况，离心率 $e = 0$，可得

$$t = \frac{L^3}{mk^2}\theta \tag{12}$$

$$v^2 r = GM \tag{13}$$

沿圆轨道运动的天体，角速度恒定，即**匀速圆周运动**．令 $\theta = 2\pi$ 可得运转周期为

$$T = \frac{2\pi L^3}{mk^2} = \frac{2\pi v^3 r^3}{GM} = 2\pi\sqrt{\frac{r^3}{GM}} \tag{14}$$

椭圆轨道

椭圆轨道是最常见的情形，离心率 $0 < e < 1$

$$t = \frac{L^3}{mk^2(1-e^2)^{\frac{3}{2}}}\left[2\arctan\left(\sqrt{\frac{1-e}{1+e}}\tan\frac{\theta}{2}\right) - \frac{e\sqrt{1-e^2}\sin\theta}{1+e\cos\theta}\right] \tag{15}$$

规定当 $(n-1)\pi < \theta \leqslant n\pi$ 时，$\arctan\left(\sqrt{\frac{1-e}{1+e}}\tan\frac{\theta}{2}\right)$ 在 $(\frac{n-1}{2}\pi, \frac{n}{2}\pi]$ 范围内取值．

同样的，令 $\theta = 2\pi$ 可得运转周期为

$$T = \frac{2\pi L^3}{mk^2(1-e^2)^{\frac{3}{2}}} \tag{16}$$

根据式 15 绘制的时间—角度变化规律，如图 1 所示

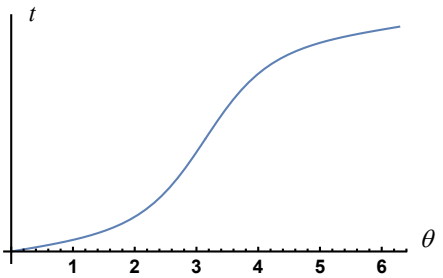


图 1: 椭圆轨道时间变量

抛物线轨道

抛物线轨道离心率 $e = 1$

$$t = \frac{L^3}{2mk^2}\tan\left(\frac{\theta}{2}\right) + \frac{1}{6}\tan^3\left(\frac{\theta}{2}\right) \tag{17}$$

在抛物线轨道上运动的天体，其动能与势能之和为**零**．

双曲线线轨道

双曲线轨道离心率 $e>1$

$$t=\frac{L^3}{mk^2(e^2-1)}\left[\frac{e\sin\theta}{1+e\cos\theta}-\frac{1}{\sqrt{e^2-1}}\ln\left(\frac{\sqrt{e+1}+\sqrt{e-1}\tan\frac{\theta}{2}}{\sqrt{e+1}-\sqrt{e-1}\tan\frac{\theta}{2}}\right)\right] \tag{18}$$

双曲线轨道的时间—角度变化规律，如图 2 所示

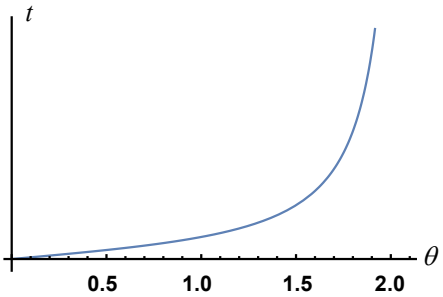


图 2: 双曲线轨道时间变量

由于式 11 中的被积函数在其有效定义域上为正值，因此所得的时间 t 单调递增，即 θ 与 t 存在一一对应的关系．然而位置—时间的关系式往往涉及超越方程的求解，不能用初等函数式表达，只能借助牛顿迭代法等数值方法计算．

限制性三体问题

- 符号 mu 的定义和“二体问题”中的不同，是否要改过来？
- 提了旋转参考系，但是没有用到
- 参考 Wikipedia 添加内容

预备知识 开普勒问题[1037]

[1]在中心力场问题中，只需把质点质量 m 和位矢 $\mathbf{r}$ 分别替换成约化质量 μM 和相对矢量 $\mathbf{R}$ 即可将单质点中心力场问题的规律拓展到二体系统[957]．

在图 1 所示的惯性坐标系 XOY 中，天体 A 和天体 B 距离为 R，质量分别为 M_1 和 M_2，总质量为 M，原点 0 为该二体系统的质心．记 $M_1=\mu M$，则

[1]参考 Wikipedia 相关页面．

$M_2=(1-\mu)M$．由“开普勒问题[1037]”中的结论可知，天体 A 和天体 B 的运动轨迹是坐标平面中两条相似的圆锥曲线．

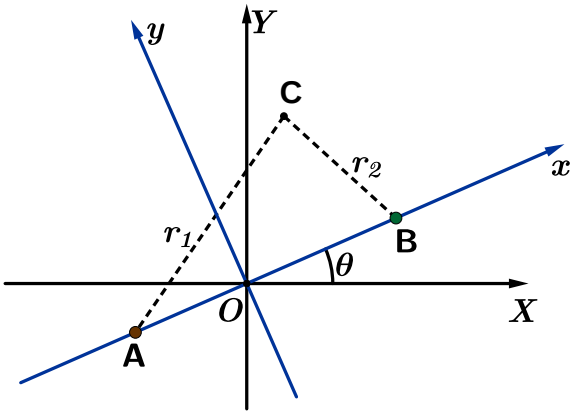


图 1: 限制性三体问题

在图 1 的二体系统中再添加一个质点，其质量 m 足够小，以至于该质点对原二体系统的影响可以忽略不计，即我们忽略 m 对其他天体的引力，但仍然受其他天体的引力．在此条件下该系统的运动问题，称为**限制性三体问题**（**restricted three-body problem**）．

假设当前时刻，A、B 连线相对惯性坐标系的角速度为 ω，与 X 轴正方向夹角为 θ．以 A、B 连线为 x 轴建立旋转坐标系 xOy．记质点 C 在旋转系中的坐标为 (x,y,z)，在惯性系中的位置矢量为 $\mathbf{r}$，两者有如下关系

$$\mathbf{r}=\begin{pmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix} \tag{1}$$

将上式关于时间求导，可得

$$\dot{\mathbf{r}}=\begin{pmatrix}(\dot{x}-\omega y)\cos\theta-(\omega x+\dot{y})\sin\theta\\ (\dot{x}-\omega y)\sin\theta-(\omega x+\dot{y})\cos\theta\\ \dot{z}\end{pmatrix} \tag{2}$$

于是，可求出质点的动能

$$T=\frac{1}{2}m|\dot{\mathbf{r}}|^2=\frac{1}{2}m\left[(\dot{x}-\omega y)^2+(\omega x+\dot{y})^2+\dot{z}^2\right] \tag{3}$$

记质点 C 与天体 A、B 的距离分别为 r_1 和 r_2，则质点 C 的引力势能可表达为

$$\begin{aligned} U &= -\frac{\mu GMm}{r_1} - \frac{(1-\mu)GMm}{r_2} \\ &= -\frac{\mu GMm}{\sqrt{[x+(1-\mu)R]^2+y^2+z^2}} - \frac{(1-\mu)GMm}{\sqrt{(x-\mu R)^2+y^2+z^2}} \end{aligned} \tag{4}$$

在惯性系中，质点 C 对坐标原点的角动量为

$$\mathbf{L} = m\mathbf{r} \times \dot{\mathbf{r}} \tag{5}$$

这里只考虑角动量在 Z 轴方向上的分量

$$\begin{aligned} L_z &= m \begin{vmatrix} x\cos\theta - y\sin\theta & x\sin\theta + y\cos\theta \\ (\dot{x}-\omega y)\cos\theta - (\omega x + \dot{y})\sin\theta & (\dot{x}-\omega y)\sin\theta - (\omega x+\dot{y})\cos\theta \end{vmatrix} \\ &= m\left[\omega\left(x^2+y^2\right) + x\dot{y} - \dot{x}y\right] \end{aligned} \tag{6}$$

在没有外力或外力矩的情况下，三体系统的总角动量和总机械能守恒．由于先前假设质点 C 的引入不影响原二体系统的运动，因此在限制性三体问题中，天体 A 和天体 B 的机械能及角动量依旧守恒，于是质点 C 的机械能和角动量也守恒．

未完成：这是错的，例如月球的角动量就不守恒，无论在惯性系还是旋转系

以上的讨论中，并不限制天体 A 和天体 B 的距离 R 和角速度 ω 在运动过程中为定值，换句话说，无论原二体系统的运动模式是何种圆锥曲线，以上的讨论都成立．

雅可比常量

- 本词条处于草稿阶段．

预备知识 限制性三体问题[1091]，旋转参考系的机械能守恒[926]

$$\frac{1}{2}mv^2 - \frac{Gmm_1}{r_1} - \frac{Gmm_2}{r_2} - \frac{1}{2}m\omega^2 r^2 = \text{const} \tag{1}$$

两边同除 $-m/2$ 即可得到守恒的**雅可比常量（Jacobi constant）**．

$$C_j = 2G\left(\frac{m_1}{r_1} + \frac{m_2}{r_2}\right) + \omega^2 r^2 - v^2 \tag{2}$$

雅可比常量 2

- 本词条需要更多参考文献.

预备知识 限制性三体问题[1091]，拉格朗日方程[1052]

[2]在“限制性三体问题[1091]”中，我们推导了第三天体 C 的动能、势能、角动量表达式. 这里对限制性三体问题进一步特殊化，假设原二体系统围绕质心作圆周运动，即大质量天体 A 和 B 的距离 R 在运动过程中不变，则在旋转坐标系中，A、B 是相对静止的. 并且天体 A、B 将作匀速圆周运动（参考“轨道参数-时间变量[1088]”中关于圆轨道的讨论），即角速度 ω 也恒定不变.

动力学方程

三体系统在空间中自由度[930] 为 9，但由于天体 A 和 B 的运动模式已知，故限制性三体系统的自由度为 3，可用质点 C 在旋转坐标系中的广义坐标 x,y,z 描述. 将质点 C 的动能 T（式 3 [1091]）和势能 U（式 4 [1091]）代入拉格朗日方程[1052]）

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial(T-U)}{\partial\dot{q}_i}=\frac{\partial(T-U)}{\partial q_i}\quad(q_i\ \to\ x,y,z) \tag{1}$$

可得质点 C 的动力学微分方程组

$$-\omega^2x-2\omega\dot{y}+\ddot{x}=-\frac{\partial U}{\partial x}=-\frac{\mu GM}{r_1^3}[x+(1-\mu)R]-\frac{(1-\mu)GM}{r_2^3}(x-\mu R) \tag{2}$$

$$-\omega^2y+2\omega\dot{y}+\ddot{y}=-\frac{\partial U}{\partial y}=-\frac{\mu GM}{r_1^3}y-\frac{(1-\mu)GM}{r_2^3}y \tag{3}$$

$$\ddot{z}=-\frac{\partial U}{\partial z}=-\frac{\mu GM}{r_1^3}z-\frac{(1-\mu)GM}{r_2^3}z \tag{4}$$

雅可比常量

由于质点 C 的角动量在惯性坐标轴上的分量和机械能分别守恒，故这些物理量的线性组合也是守恒量. 为统一量纲，可以令角动量分量乘以角速度.

[2]本文参考 Wikipedia 相关页面.

在所有这些线性组合中，有一种组合具有实际的物理意义，记这个组合量为 E

$$\begin{aligned} E &= (T+U) - \omega L_z \\ &= \frac{1}{2}m\left(\dot{x}^2 - \dot{y}^2 + \dot{z}^2\right) - \frac{1}{2}m\omega^2\left(x^2+y^2\right) - \frac{\mu GMm}{r_1} - \frac{(1-\mu)GMm}{r_2} \end{aligned} \tag{5}$$

物理量 E 具有能量的量纲．其中 $\frac{1}{2}m(\dot{x}^2 - \dot{y}^2 + \dot{z}^2)$ 是质点 C 相对旋转坐标系的动能，E 的其余部分只与位置相关，故可解释为势能．在势能部分，引力势能 $-\mu GMm/r_1 - (1-\mu)GMm/r_2$ 与坐标系的选取无关，而 $-\frac{1}{2}m\omega^2(x^2+y^2)$ 是由于坐标系 xOy 的旋转引起的，可将其解释为质点 C 所受的向心力由于坐标轴自身的旋转而产生的势能．守恒量 E 称为**雅可比常量**，可看作是第三天体相对于旋转坐标系具有的能量．

推导

将式 2 乘以 $\dot{x}$，式 3 乘以 $\dot{y}$，式 4 乘以 $\dot{z}$，再将三个式子相加，可得

$$\begin{aligned} &\dot{x}\ddot{x} + \dot{y}\ddot{y} + \dot{z}\ddot{z} - \omega^2(x\dot{x} + y\dot{y}) \\ &= -\frac{\mu GM}{r_1^3}[x\dot{x} + (1-\mu)R\dot{x} + y\dot{y} + z\dot{z}] - \frac{(1-\mu)GM}{r_2^3}(x\dot{x} - \mu R\dot{x} + y\dot{y} + z\dot{z}) \end{aligned} \tag{6}$$

其中，注意到

$$\dot{x}\ddot{x} + \dot{y}\ddot{y} + \dot{z}\ddot{z} = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right) \tag{7}$$

$$\omega^2(x\dot{x} + y\dot{y}) = \frac{1}{2}\omega^2\frac{\mathrm{d}}{\mathrm{d}t}\left(x^2+y^2\right) \tag{8}$$

$$-\frac{\mu GM}{r_1^3}[x\dot{x} + (1-\mu)R\dot{x} + y\dot{y} + z\dot{z}] = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mu GM}{r_1}\right) \tag{9}$$

$$-\frac{(1-\mu)GM}{r_2^3}(x\dot{x} - \mu R\dot{x} + y\dot{y} + z\dot{z}) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{(1-\mu)GM}{r_2}\right) \tag{10}$$

将以上四式代入式 6 可得

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right) - \frac{\omega^2}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(x^2+y^2\right) - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\mu GM}{r_1} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{(1-\mu)GM}{r_2} = 0 \tag{11}$$

于是

$$\frac{1}{2}\left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right) - \frac{1}{2}\omega^2\left(x^2+y^2\right) - \frac{\mu GM}{r_1} - \frac{(1-\mu)GM}{r_2} = Const \tag{12}$$

式 12 就是单位质量的**雅可比常量（Jacobi constant）**.

拉格朗日点

- 本词条存在未完成的内容.

预备知识 雅可比常量[1094]

[3]对于大质量天体 A 和 B 围绕质心作匀速圆周运动的限制性三体问题，是否存在若干个特殊位置，使得处于该位置并满足一定的初始条件的第三质点将在运动中保持与天体 A 和 B 的相对位置不变，即第三质点同样围绕系统质心作匀速圆周运动，且角速度与天体 A 和 B 相同？这样的位置称为“**平动点**”或“**拉格朗日点（Lagrangian point）**”.

提取式 1 [1093] 中的势能部分并除以质量，记为等效势 $\widetilde{U}$

$$\widetilde{U}=-\frac{Gm_1}{r_1}-\frac{Gm_2}{r_2}-\frac{1}{2}\omega^2r^2 \tag{1}$$

其中 $r_i=\sqrt{(x-x_i)^2-(y-y_i)^2}$.

等效势能在 xoy 平面内部分区域的分布如图 1 所示，图中 A、B 两点为等效势能的理论奇点，在其附近形成“势能陷阱”.

要计算平衡点，即 $\widetilde{U}$ 的极值点，只需令

$$\boldsymbol{\nabla}\widetilde{U}=\mathbf{0} \tag{2}$$

得到

$$\frac{-Gm_1(x-x_1)}{r_1^3}-\frac{Gm_2(x-x_2)}{r_2^3}+\omega^2x=0 \tag{3}$$

$$-\frac{Gm_1y}{r_1^3}-\frac{Gm_2y}{r_2^3}+\omega^2y=0 \tag{4}$$

即

$$-\frac{Gm_1(\mathbf{r}-\mathbf{r}_1)}{r_1^3}-\frac{Gm_2(\mathbf{r}-\mathbf{r}_2)}{r_2^3}+\omega^2\mathbf{r}=\mathbf{0} \tag{5}$$

显然这相当于寻找两个万有引力和离心力合力为零的点. 式 4 有显然的解 $y=0$，带入式 3 得

[3]参考 Wikipedia 相关页面.

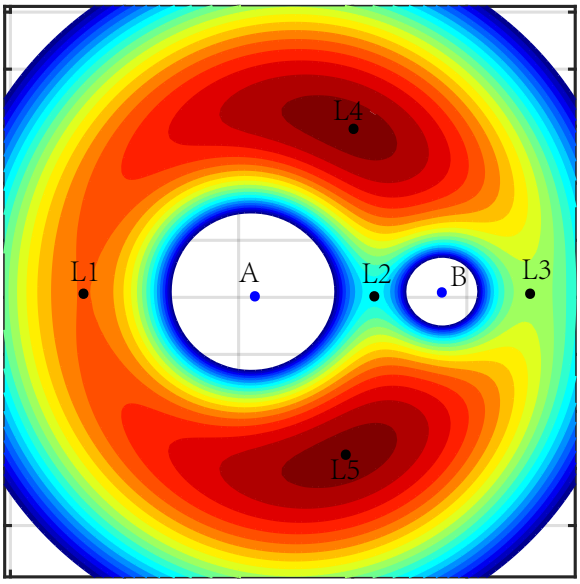


图 1: 等效势能的分布（红色代表高势能，蓝色代表低势能）

未完成：见注释

使用拉格朗日方程推导

未完成：该小节有待整理

保持与天体 A 和 B 的相对位置不变，即要求第三质点在旋转坐标系中满足 $\dot{x}=\dot{y}=\dot{z}=\ddot{x}=\ddot{y}=\ddot{z}=0$，代入第三质点的动力学微分方程组（式 2 [1094]，式 3 [1094]，式 4 [1094]）可得方程组

$$-\omega^2 x=-\frac{\partial U}{\partial x}=-\frac{\mu GM}{r_1^3}[x+(1-\mu)R]-\frac{(1-\mu)GM}{r_2^3}(x-\mu R) \tag{6}$$

$$-\omega^2 y=-\frac{\partial U}{\partial y}=-\left[\frac{\mu GM}{r_1^3}+\frac{(1-\mu)GM}{r_2^3}\right]y \tag{7}$$

$$0=-\frac{\partial U}{\partial z}=-\left[\frac{\mu GM}{r_1^3}+\frac{(1-\mu)GM}{r_2^3}\right]z \tag{8}$$

观察式 8 可得 $z=0$，即待求点在天体 A、B 的运动平面内.

若 $y=0$，则式 7 恒成立. 并且注意到 A、B 作匀速圆周运动时，根据外有引力等于向心力的关系，可得

$$\omega^2=\frac{GM}{R^3} \tag{9}$$

此时式 6 便可化简为

$$\frac{x}{R^3}=\frac{\mu}{|x+(1-\mu)R|^3}[x+(1-\mu)R]+\frac{1-\mu}{|x-\mu R|^3}(x-\mu R) \tag{10}$$

此方程没有解析解，只能借助数值方法求得其满足一定精度要求的数值解．可以验证在天体 A、B 之间及两边延长线方向上各有一个解，分别记为 L_1, L_2, L_3．

若 $y\neq 0$，则有

$$\begin{cases}\dfrac{x}{R^3}=\dfrac{\mu}{r_1^3}[x+(1-\mu)R]+\dfrac{1-\mu}{r_2^3}(x-\mu R)\\ \dfrac{1}{R^3}=\dfrac{\mu}{r_1^3}+\dfrac{1-\mu}{r_2^3}\end{cases} \tag{11}$$

解得 $r_1=r_2=R$，即平动点与 A、B 形成正三角形，这样的平动点有两个，分别记为 L_4, L_5．

拉格朗日点的力学本质是，第三质点在该位置所受的引力合力恰好等于维持匀速旋转所需的的向心力．五个拉格朗日点与 A、B（在旋转坐标系内）的相对位置大致如图 2 所示．

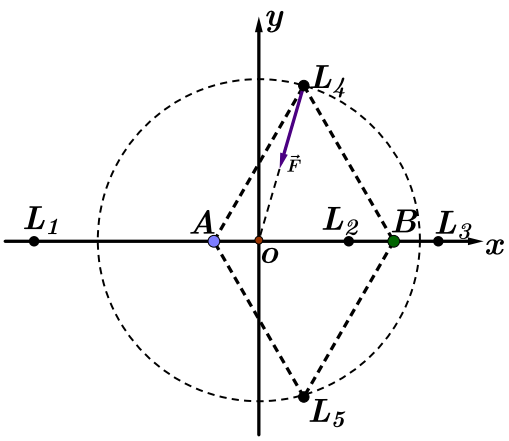


图 2: 拉格朗日点位置

光子火箭

- 本词条处于草稿阶段.

一个光子的能量

$$\hbar\omega = mc^2 \tag{1}$$

一个光子的动量

$$p = mc = \frac{\hbar\omega}{c} \tag{2}$$

功率和推力的关系

$$F = p\frac{\mathrm{d}N}{\mathrm{d}t} = \frac{\hbar\omega}{c}\frac{\mathrm{d}N}{\mathrm{d}t} = \frac{P}{c} \tag{3}$$

所以 $1N$ 的推力需要 3×10^8 W 的功率！这与光子的能量无关.

如果使用光子火箭，用反物质湮灭产生的光子推动，有

$$F = \frac{P}{c} = c\frac{\mathrm{d}m}{\mathrm{d}t} \tag{4}$$

产生 100 吨的推力只需要每秒消耗 3.3 克反物质，而功率却是惊人的 3×10^{14} W，可以每秒钟汽化约 10^5 立方米的水（从常温常压到水蒸气）.

一般情况

若使用经典力学，用于推进的介质喷出速度越大，同样的推力功率就越大.

功率

$$P = \frac{v^2}{2}\frac{\mathrm{d}m}{\mathrm{d}t} \tag{5}$$

或者

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \frac{2P}{v^2} \tag{6}$$

推力

$$F = v\frac{\mathrm{d}m}{\mathrm{d}t} = \frac{2P}{v} \tag{7}$$

所以在设计推进器时，式 6 式 7 会互相制约，在同样的功率下，要想推力翻倍，工质需要消耗原来的 4 倍.

火箭加速的完整计算

在没有引力的环境．火箭以以相对速度 u 和工质消耗 $\alpha = \mathrm{d}m/\mathrm{d}t$ 喷出燃料．

$$F = u\alpha \tag{8}$$

火箭质量

$$m(t) = M - \alpha t \tag{9}$$

加速度

$$a(t) = \frac{F}{m(t)} = \frac{u}{M/\alpha - t} \tag{10}$$

速度

$$v(t) = \int_0^t a(t')\,\mathrm{d}t' = -u\ln(1 - \alpha t/M) \tag{11}$$

若火箭质量减去燃料为 M_0，那么时间最大取 $t = (M - M_0)/\alpha$，最大速度为

$$v_f = u\ln(M/M_0) \tag{12}$$

或者

$$M = M_0 \mathrm{e}^{v_f/u} \tag{13}$$

可见燃料随莫速度呈指数增长．注意这与 α 无关．

甚至 α 随时间变化时我们也可以得到式 13 的结论

$$v = \int_0^t \frac{u\alpha(t')}{m(t')}\,\mathrm{d}t' = \int_{M_0}^{M} \frac{u\,\mathrm{d}m}{m} = u\ln\frac{M}{m} \tag{14}$$

可见 v 是火箭总质量 m 的函数，无论 α 如何随时间变化．

太空电梯

预备知识 万有引力[1027]，离心力[924]，积分方程

[4]太空电梯通常是从地球同步轨道外的一个空间站悬挂一根缆绳到地球表面，使得我们可以通过该缆绳将一定质量的物体直接运送到太空中．缆绳的张

[4]关于太空电梯的概念参考 Wikipedia 相关页面．

力由空间站的离心力提供. 碳纳米管材料的发现使得太空电梯从的可行性大大提高，本文中我们来计算理想情况下太空电梯所需缆绳的质量. 根据已有的碳纳米管材料数据，下面的计算发现所需缆绳的质量仅仅是最大载重（包括动力系统）的 2.7 到 13.5 倍. 可见碳纳米管材料的大规模生产将有可能开启全新的航天时代.

以下的计算是风速为零的理想情况，为了避免极端天气，可在空间站安装伸缩机，只有在适合的天气将缆绳下放到地面. 太空电梯的另一个问题是需要避免低轨道卫星和太空碎片的碰撞，这些不在本文讨论范围内.

理想情况计算

在地球所在的旋转参考系中，令缆绳上某点的张力为 $F(r)$，r 为该点到地心的距离. 令缆绳底部悬挂重物的重力为 F_0，则缆绳底部所需张力为 $F(r_0)=F_0$，r_0 是地球赤道半径. 显然在近地处缆绳张力 $F(r)$ 随着高度 r 的增加，因为 $F(r)$ 等于底部重物的重力加上 r 下方所有缆绳的重力减离心力. 当 r 大于同步轨道半径时，$F(r)$ 慢慢减小，这是因为同步轨道以上缆绳受到的离心力已经大于万有引力. 为了节约材料，缆绳的线密度（单位长度的质量）也可以设计为高度的函数 $\lambda(r)$. 于是缆绳的张力可以通过以下积分计算，其中括号的第一项是引力[1027]，第二项是离心力[924].

$$F(r)=F_0+\int_{r_0}^{r}\left(\frac{GM}{r'^2}-\omega^2 r'\right)\lambda(r')\,\mathrm{d}r' \tag{1}$$

其中 ω 是地球自转角速度.

令缆绳材料单位截面可承受的最大张力为 p，那么截面面积至少为

$$A(r)=F(r)/p \tag{2}$$

绳的密度为常数 ρ，那么线密度为

$$\lambda(r)=\rho A(r)=\rho F(r)/p \tag{3}$$

带回式 1 有积分方程

$$F(r)=\frac{GM\rho}{p}\int_{r_0}^{r}\frac{F(r')}{r'^2}\,\mathrm{d}r'-\frac{\omega^2\rho}{p}\int_{r_0}^{r}F(r')r'\,\mathrm{d}r' \tag{4}$$

令两个积分前面的常数为 α, β，两边求导（右边使用式 13 [202]）得微分方程

$$\frac{\mathrm{d}F}{\mathrm{d}r} = \alpha\frac{F(r)}{r^2} - \beta r F(r) \tag{5}$$

分离变量，解得

$$F = C\exp\left(-\frac{\alpha}{r} + \frac{\beta}{2}r^2\right) \tag{6}$$

令初始条件为 $F(r_0) = F_0$，F_0 是载重．代入得到方程的解为

$$F(r) = F_0\exp\left[\frac{GM\rho}{p}\left(\frac{1}{r_0} - \frac{1}{r}\right) - \frac{\omega^2\rho}{2p}(r^2 - r_0^2)\right] \tag{7}$$

目前最强的材料碳纳米管保守估计[5] $p = 2.5\text{-}6.6 \times 10^{10}$ Pa，密度使用 $\rho = 1.34\times10^3\,\mathrm{kg/m^3}$．地球半径取 6.370×10^6 m，同步轨道半径为 4.2164×10^7 m．地球质量 $M = 5.972 \times 10^{24}$ kg．

代入上式得 $F = 2.7\text{-}13.5F_0$．

附：Matlab 代码

```
G = 6.6743e-11; % 引力常数
M = 5.972e24; % 地球质量
r0 = 6.370e6; % 地球半径
r = 4.2164e7; % 同步轨道半径
rho = 1.34e3; % 碳纳米管密度 0.037-1.34
p = 2.5e10; % 碳纳米管截面压强 2.5-6.6
w = 2*pi/(24*3600); % 地球角速度

ratio = exp(G*M*rho/p*(1/r0 - 1/r) - w^2*rho/(2*p)*(r^2 - r0^2))
    ;
```

[5] 来源：Nature 2019 相关文章．

第十二部分

光学

第一章

几何光学

惠更斯原理

水面上有一个波传播时，如果没有遇到障碍物，波将保持原来的波面形状前进，若在前进中遇到一个有小孔的障碍物 AB，如图 1 所示．只要小孔的孔径 a 比波长 λ 小，我们就可看到，穿过小孔的波总是半圆形的波，与原来波的形状无关，这说明小孔可以看作是一个新波源，所发射的波称为子波．

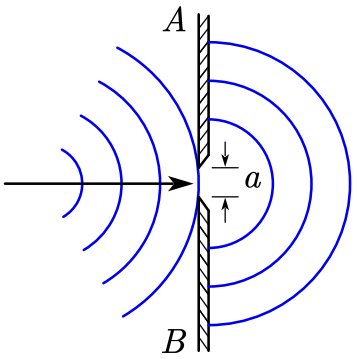


图 1: 障碍物的小孔成为新的波源

惠更斯（C. Huygens）于 1678 年提出了关于波传播的几何法则：在波的传播过程中，波阵面（波前）上的每一点都可行做是发射子波的波源，在其后的任一时刻，这些子波的包迹就成为新的波阵面．这就是**惠更斯原理（Huygens principle）**.

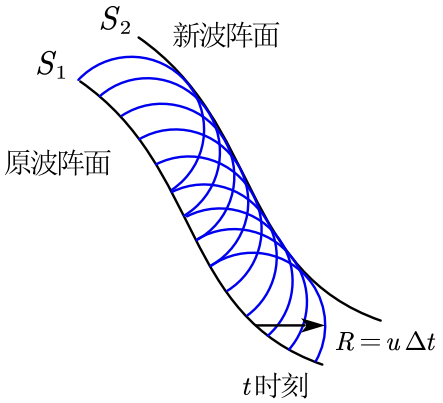


图 2: 惠更斯原理

如图 2 所示，设 S_1 为某一时刻 t 的波阵面，根据惠更斯原理，S_1 上的每一点发出的球面子波，经 Δt 时间后形成半径为 t 的球面，在波的前进方向上，这些子波的包迹 S_2 就成为 $t+\Delta t$ 时刻的新波阵面．惠更斯原理对任何波动过程都是适用的，不论是机械波或电磁波，只要知道某一时刻的波阵面，就可根据这一原理用几何方法来决定任一时刻的波阵面，因而在很广泛的范围内解决了波的传播问题．图 3 中用惠更斯原理描绘出球面波和平面波的传播．根据惠更斯原理，还可以简捷地用作图方法说明波在传播中发生的衍射、散射、反射和折射等现象． 应该指出，惠更斯原理并没有说明各个子波在传播中对某一

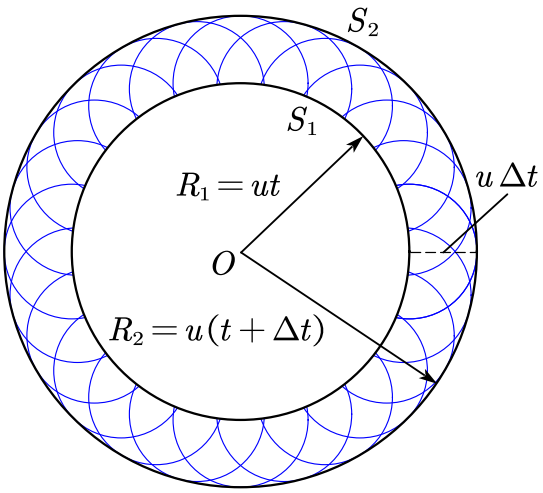


图 3: 用惠更斯原理求作球面波的新的波阵面

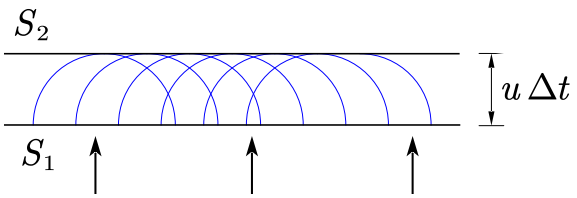


图 4: 用惠更斯原理求作平面波的新的波阵面

点振动的相位和振幅究竟有多少贡献，不能给出沿不同方向传播的波的强度分布，后来菲涅耳对惠更斯原理作了补充，这将在惠更斯——菲涅尔原理一节中加以介绍．

光的折射 斯涅尔定律

预备知识 惠更斯原理[1105]

折射率

介质的折射率可以由光在介质中的速度 v 定义，令 c 为真空中的光速，则

$$n=\frac{c}{v} \tag{1}$$

这说明折射率和速度成反比．由于真空中的光速是光的最大速度，所以折射率必然大于等于 1．

折射定律

折射定律也叫**斯涅尔定律**，在几何光学中描述光从一种介质由光滑的界面入射到另一种介质时角度的变化．令两种介质的折射率分别为 n_1 和 n_2，光线在两种介质中关于法线的夹角分别为 θ_1 和 θ_2，则两介质中的光线共面，且

$$n_1\sin\theta_1=n_2\sin\theta_2 \tag{2}$$

推导

要推导斯涅尔定律，就不得不考虑光的波动性．我们可以用惠更斯原理来解释这个公式．如图 1 所示，光波从介质 1 传播进入介质 2，其入射角、折射角分别为 θ_1、θ_2，传播速度分别为 v_1、v_2，假设 $v_1>v_2$．在时间 t_j 时，光波的波前会包含点 A_j 和点 B_j 的位置，标记这时的波前为 $\overline{A_jB_j}$．假设时间 t_j 与 t_{j+1} 之间的间隔为常数 Δt，则以下几个直线段之间的长度相等关系成立：

$$\begin{aligned}A_0A_1=B_0B_1=B_1B_2=B_2B_3=v_1\Delta t\\ A_1A_2=A_2A_3=A_3A_4=B_3B_4=v_2\Delta t\end{aligned} \tag{3}$$

从波前 $\overline{A_1B_1}$ 的每一个点波源发射出的球面次波，分别在介质 1、介质 2 的传播速度为 v_1、v_2，$\overline{A_2B_2}$ 必须正切这些球面次波．特别而言，在时间间隔 Δt 之

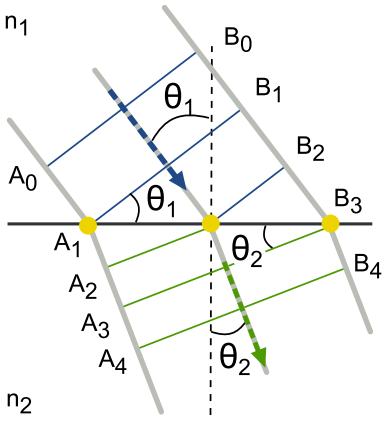


图 1: 波的折射

后，波前 $\overline{A_2B_2}$ 在介质 1 的部分必须平行于相距 $v_1\Delta t$ 的波前 $\overline{A_1B_1}$，而波前 $\overline{A_2B_2}$ 在介质 2 的部分必须正切从点波源 A_1 发射出的半径为 $v_2\Delta t$ 的球面次波. 所以，在通过界面时，会出现弯曲的波前 $\overline{A_2B_2}$.

由于光波传播的方向垂直于波前，所以在介质 1、介质 2 里，波前与界面之间的夹角分别等于入射角 θ_1、折射角 θ_2. 直线段长度 B_1B_3 与 A_1A_3 之间的关系为

$$B_1B_3/\sin\theta_1 = A_1B_3 = A_1A_3/\sin\theta_2 \tag{4}$$

即

$$\frac{v_1}{\sin\theta_1} = \frac{v_2}{\sin\theta_2}\,. \tag{5}$$

应用折射率 n 的定义式，即 $n=\dfrac{c}{v}$（c 为光速），我们得出**斯涅尔定律**：

$$n_1\sin\theta_1 = n_2\sin\theta_2 \tag{6}$$

其中，n_1、n_2 分别为介质 1、介质 2 的折射率.

薄透镜

预备知识 折射定律[1107]

单个球面的成像公式

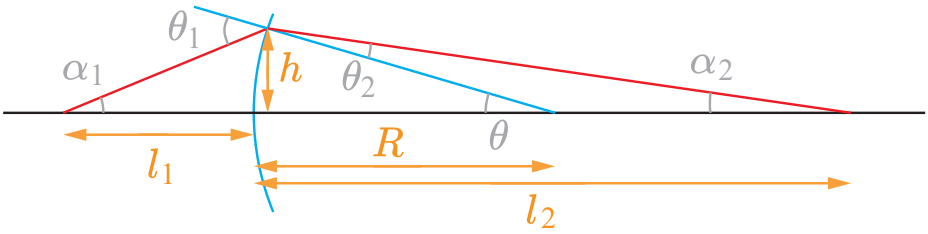


图 1: 单球面成像

如图 1，我们考虑两种折射率分别为 n_1 和 n_2 得介质被一个球面划分为左右两部分．当光线从左边入射时，经过界面反射．

傍轴条件：图中所有角度都很小[1]．球面近似是平面，球面上任意一点横坐标相同．

$$l_1\alpha_1 = l_2\alpha_2 = R\theta \tag{1}$$

由三角形性质得

$$\alpha_1 = \theta_1 - \theta \qquad \alpha_2 = \theta - \theta_2 \tag{2}$$

$$n_1\theta_1 = n_2\theta_2 \tag{3}$$

消去 θ_1 和 θ_2 得

$$\frac{n_1}{l_1} + \frac{n_2}{l_2} = \frac{n_2 - n_1}{R} \tag{4}$$

这已经比较接近凸透镜成像公式了．注意当 l_1 或者 l_2 取负数时同样成立，这意味着物或者像在透镜的另一侧（即虚物或者虚像）．若透镜的圆心在左侧，将式中 R 也改为负数即可（请读者自行证明这两个结论）．最后注意我们并不要求式中 n_1, n_2 哪个更大．

薄透镜

如果一个透镜的两个面可以近似为球面，且它们之间的距离比起物距和像距来可以忽略不记，那么我们就称它为**薄透镜**．我们假设透镜外介质的折射率

[1] 即 $\sin\beta \approx \tan\beta \approx \beta$，见“小角正弦极限[152]”

为 n_1，透镜内折射率为 n_2．我们另透镜一侧的半径为 R_1，物距为 u，像距为 v_1，另一侧半径为 R_2（另焦点在另一边为正），物距为 u_2，像距为 v．于是有

$$\frac{n_1}{u}+\frac{n_2}{v_1}=\frac{n_2-n_1}{R_1} \tag{5}$$

$$\frac{n_2}{u_2}+\frac{n_1}{v}=\frac{n_2-n_1}{R_2} \tag{6}$$

由于透镜厚度可以忽略，任何情况下都有

$$v_1=-u_2 \tag{7}$$

将式 5 与式 6 相加再用式 7 消去含有 v_1 和 u_2 的两项，就得到了熟悉的**薄透镜成像公式**

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f} \tag{8}$$

其中

$$\frac{1}{f}=\left(\frac{n_2}{n_1}-1\right)\left(\frac{1}{R_1}+\frac{1}{R_2}\right) \tag{9}$$

再次声明这里使用的正负号规范：实物距离为正，实像距离为正，凸透镜半径为正；反之为负．

薄透镜叠加

与上一节同理，若两个焦距分别为 f_1 和 f_2 的薄透镜叠加，当其间距远小于物距和像距时，合成后的透镜焦距 f 满足

$$\frac{1}{f}=\frac{1}{f_1}+\frac{1}{f_2} \tag{10}$$

推导：另第一个透镜的物距为 u，像距为 v_1，第二个透镜物距为 u_2，像距为 v，由于两透镜距离可忽略，第一个透镜的实像像距等于第二个透镜的虚物物距，所以分别可以列出

$$\frac{1}{u}+\frac{1}{v_1}=\frac{1}{f_1} \tag{11}$$

$$\frac{1}{u_2}+\frac{1}{v}=\frac{1}{f_2} \tag{12}$$

$$v_1=-u_2 \tag{13}$$

联立得

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f_1}+\frac{1}{f_2} \tag{14}$$

令组合透镜的等效焦距为 f，即等式右边为 $1/f$，得到式 10.

第二章

波动光学

可见光谱

可见光的波长范围在大约 400nm 到 700nm 之间．各个波长对应的颜色如图 1，图中我们进一步将可见光根据波长划分为红，橙，黄，绿，蓝，紫几个区间．



图 1: 电脑生成的标准可见光谱（图片来自维基百科）

人眼和显示器

我们知道人眼可以识别不同的波长是因为我们视网膜上有三种不同的感光细胞，它们分别主要感受红，绿，蓝三种波长的光，再根据这三种信号的大小在大脑中还原出不同的颜色．显示器也根据人眼的这种特性对每个像素设置了红，绿，蓝三种颜色[1]．

当电脑储存一张彩色图片时，每个像素会被储存为 red, green, blue （或记为 RGB）三个数，每个数一般取 0 到 255 的整数，有时候也取 0 到 1 的实数．图 1 中，RGB 三个分量的大小如图 2 所示，这可以反映出人眼中三种受体对不同波长的敏感程度．

双缝干涉中一个重要极限

预备知识 泰勒级数[181]

该近似最常用于波动理论，如光的远场干涉衍射．

$$|\mathbf{r}-\mathbf{r}_0| = r - \hat{\mathbf{r}}\cdot\mathbf{r}_0 = r - r_0\cos\theta \tag{1}$$

[1] 一些动物的感光细胞不一样，所以当它们看屏幕是，看到的颜色可能与我们不同．

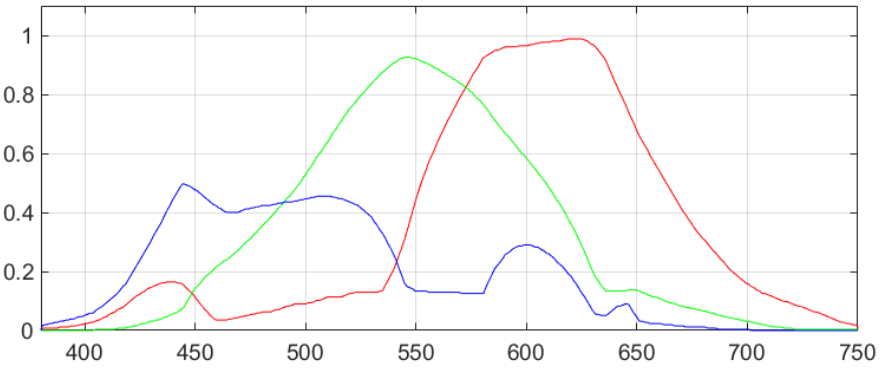


图 2: RGB 三个分量

θ 为 $\mathbf{r}$ 与 $\mathbf{r}_0$ 的夹角．这个近似的误差远小于 r_0．

该近似的几何意义是，若三角形的两条边远长于第三条，那么这两条边的长度差等于第三条在任意一条临边上的投影．

推导

令 $\varepsilon = r_0/r$，$\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}_0 = \cos\theta$，该命题变为证明

$$\lim_{\varepsilon\to 0} \frac{1}{r_0}[|\mathbf{r} - \mathbf{r}_0| - (r - r_0\cos\theta)] = 0 \tag{2}$$

或

$$\lim_{\varepsilon\to 0} \frac{1}{r_0}\left[\sqrt{r^2 + r_0^2 - 2rr_0\cos\theta} - (r - r_0\cos\theta)\right] = 0 \tag{3}$$

把一个 r 提出中括号外，得

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[\sqrt{1 + \varepsilon^2 - 2\varepsilon\cos\theta} - (1 - \varepsilon\cos\theta)\right] = 0 \tag{4}$$

现在把根号对 ε 进行泰勒展开，保留其一阶无穷小，得

$$\sqrt{1 + \varepsilon^2 - 2\varepsilon\cos\theta} = 1 - \varepsilon\cos\theta + \mathcal{O}\left(\varepsilon^2\right) \tag{5}$$

代入得

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\mathcal{O}\left(\varepsilon^2\right) = \lim_{\varepsilon\to 0} \mathcal{O}(\varepsilon) = 0 \tag{6}$$

证毕．

杨氏双缝干涉实验

预备知识 双缝干涉中一个重要极限[1112]

用两个点波源作光的干涉实验的典型代表，是杨氏实验.

实验装置

杨氏实验的装置如下图所示，在普通单色光源前面放一个开有小孔 S 的

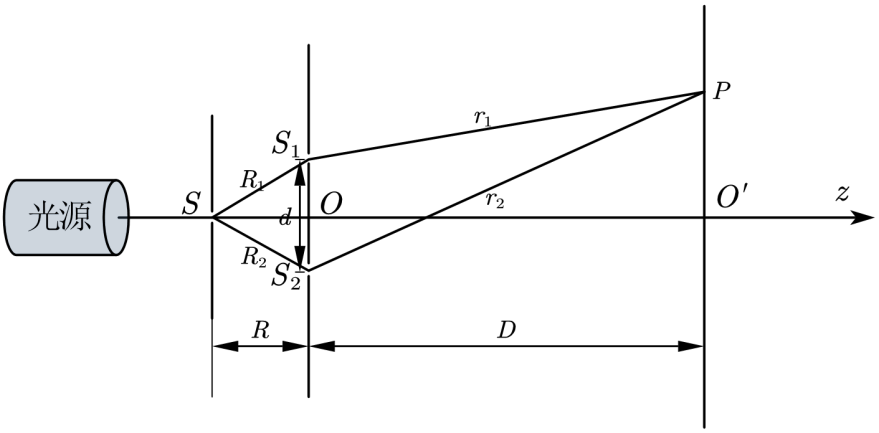


图 1: 杨氏双缝干涉

屏，作为单色点光源. 在 S 的照明范围内再放一个开有两个小孔 S_1, S_2 的屏. 按惠更斯原理，S 将作为两个次波源向前发射次波（球面波），形成交叠的波场. 在较远的地方放置一接收屏，屏上可以观测到一组几乎是平行的直线条纹，如图 1 (b) .

但这样的亮度太低了. 为了提高干涉条纹的亮度，实际中 S, S_1, S_2 用三个互相平行的狭缝（S_1 和 S_2 在同一竖直平面上并都与地面平行，S 在一竖直平面上且与地面平行，且两竖直平面互相平行）并且可不用屏幕接收，而代之以目镜直接观测. 激光出现以后，人们可以用氦氖激光束直接照明双孔，在屏幕上即可获得一套相当明显的干涉条纹，供许多人同时观看.

干涉明暗条纹的位置

现在来分析，用普通光源做杨氏实验时，由双孔出射的两束光波之间的相位差．设 $SS_1 = R_1, SS_2 = R_2$，用 φ_0 代表点波源 S 的初相位，则次波源 S_1, S_2 的初相位分别为

$$\varphi_{10} = \varphi_0 + \frac{2\pi}{\lambda} R_1, \quad \varphi_{20} = \varphi_0 + \frac{2\pi}{\lambda} R_2 \tag{1}$$

从而

$$\varphi_{10} - \varphi_{20} = \frac{2\pi}{\lambda} (R_1 - R_2) \tag{2}$$

由此可见，两次波之间的相位差与 φ_0 无关，即使中 φ_0 变了，相位差中 $\varphi_{10} - \varphi_{20}$ 也不变．下面来作一些计算．

令双孔间距为 d, 屏幕与双孔屏间的距离为 D，屏幕上横向观测范围为 X，我们设 $d^2 \ll D^2$（这被称为**远场条件**），$X^2 \ll D^2$（这被称为**傍轴条件**）.

设 S_1、S_2 离 S 的距离相等，即 $R_1 = R_2$, 从而 $\varphi_{10} = \varphi_{20}$，为了简化运算，可以取二者皆为 0，不影响结论．如图 1，从线段 S_1S_2 的中点 O 作 z 轴垂直于双孔屏和接收屏．设接收屏上点 P 的横向距离为 X，OP 与 z 轴的夹角为 θ．在远场条件下，可认为 $S_1P \parallel S_2P$，在傍轴条件下可认为 OP 也与它们平行．从 S_1 作 OP 和 S_2P 的垂线交 S_2P 于 N．则 $\overline{S_2N}$ 近似等于光程差：

$$r_1 - r_2 \approx \overline{S_2N} = d \sin\theta \approx \frac{dx}{D} \tag{3}$$

容易得到相位差

$$\delta(P) = \frac{2\pi (r_1 - r_2)}{\lambda} = \frac{2\pi d}{\lambda D} x \tag{4}$$

干涉条纹的形状，即等强度线是一组纵向（即与图 1 垂直）的平行直线．强度随 $\delta(P)$ 作周期性变化．干涉条纹的间距定义为两条相邻亮纹（强度极大）或两条暗纹（强度极小）之间的距离．因为两条相邻条纹之间的光程差相差 λ，令式 3 中 $r_1 - r_2 = \Delta L = \lambda$，$x$ 写成条纹间隔 Δx，则有

$$\Delta x = \frac{\lambda D}{d} \tag{5}$$

由上述分析可知，当 P 点处为明纹时，有

$$x = \pm k \frac{D\lambda}{d} \quad k = 0, 1, 2, \cdots \tag{6}$$

相应于 $k=0$ 的明纹称为**零级明纹**或**中央明纹**．相应于 $k=1, k=2, \cdots$ 的明纹，称为第一级、第二级、……明纹．

当 P 点处为暗纹时，有

$$x=\pm(2k+1)\frac{D\lambda}{2d} \quad k=0,1,2,\cdots \tag{7}$$

若光源中包含 λ_1 和 λ_2 两条谱线，则屏上有两套间距不等的条纹同时存在，它们非相干地叠加在一起，如图 2 所示．若光源发出的是白光，则在中央

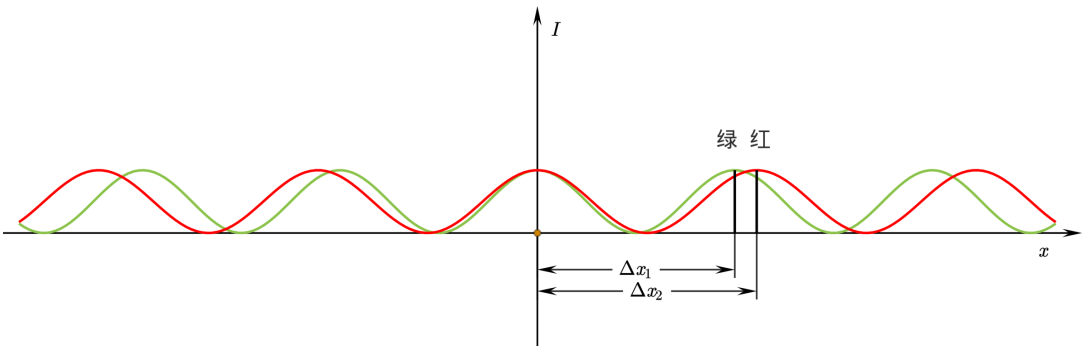


图 2: 两套不相同颜色干涉条纹不相干叠加

零级的白色亮纹两侧，对称地排列着若干条彩色条纹．

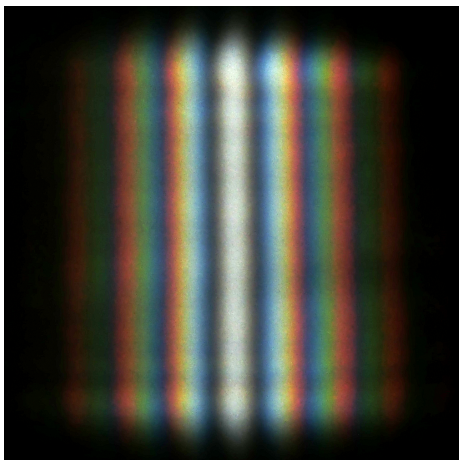


图 3: 白光的杨氏双缝干涉图样（图片来自维基百科）

劳埃德镜实验

预备知识 杨氏双缝干涉实验[1114]，半波损失[??]

劳埃德（H.Lloyd）于 1834 年提出了一种更简单的观察干涉的装置. 如图 1 所示，MN 为一块平玻璃板，用作反射镜，S_1 是一狭缝光源，从光源发出的光波，一部分掠射（即入射角接近 90°）到玻璃平板上，经玻璃表面反射到达屏上；另一部分直接射到屏上. 这两部分光也是相干光，它们同样是用分波阵面得到的. 反射光可看成是由虚光源 S_2 发出的. S_1 和 S_2 构成一对相干光源，对干涉条纹的分析与杨氏实验也相同. 中画有阴影的区域表示相干光在空间叠加的区域. 这时在屏上可以观察到明暗相间的干涉条纹. 应该指出，在劳埃德

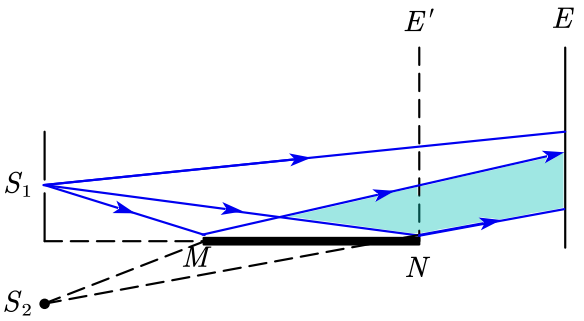


图 1: 劳埃德镜实验

镜实验中，如果把屏幕移近到和镜面边缘 N 相接触，即图 1 中 E' 的位置，这时从 S_1 和 S_2 发出的光到达接触处的路程相等，应该出现明纹，但实验结果却是暗纹，其他的条纹也有相应变化. 这一实验事实说明了由镜面反射出来的光和直接射到屏上的光在 N 处的相位相反，即相位差为 π. 由于直射光的相位不会变化，所以只能认为光从空气射向玻璃平板发生反射时，反射光的相位跃变了 π.

进一步的实验表明：光从光疏介质射到光密介质界面反射时，在掠射（入射角 $i = 90°$ 或正入射，即 $i = 0$）的情况下，反射光的相位较之入射光的相位有 π 的突变，这一变化导致了反射光的波程在反射过程中附加了半个波长，故常称为**半波损失**. 今后在讨论光波叠加时，若有半波损失，在计算波程差时必须计及，否则会得出与实际情况不符的结果.

普通光源的发光机理

一般普通光源（指非激光光源）发光的机理是处于激发态的原子（或分子）的自发辐射，即光源中的原子（或分子）吸收了外界能量而处于激发态，这些激发态是极不稳定的，它会自发地回到低能量的激发态或基态，在这过程中，原子向外发射电磁波（光波）．每个原子的发光是间歇的．一个原子经一次发光后，只有在重新获得足够能量后才会再次发光每次发光的持续时间极短，约为 10^{-8}s．可见原子发射的光波是一段频率一定、振动方向一定、有限长的光波，通常称为光波列图 1.

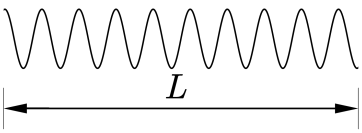


图 1: 光波波列

同一原子在不同时刻所发出的波列之间振动方向和相位也各不相同．在普通光源中，大量原子在发光，各个原子的激发和辐射参差不齐，而且彼此之间没有联系，是一种随机过程，因而不同原子在同一时刻所发出的波列在频率、振动方向和相位上各自独立，可见，普通光源中原子发光，可谓此起彼伏、瞬息万变．

单色光

预备知识 平面波[1008]

可见光是波长在 $400-760$nm，亦即频率在 $4.3\times10^{14}\sim7.5\times10^{14}$Hz 之间的电磁波．具有单一频率的光波称为单色光（monochromatic light）．然而，严格的单色光是不存在的．任何光源所发出的光波都有一定的频率（或波长）范围，在此范围内，各种频率（或波长）所对应的强度是不同的，以波长为横坐标，强度为纵坐标，可以直观地表示出这种强度与波长间的关系，称为**光谱曲**

线，或称**谱线（spectrum）**，如图 1 所示，谱线所对应的波长范围越窄，则称光的单色性越好.

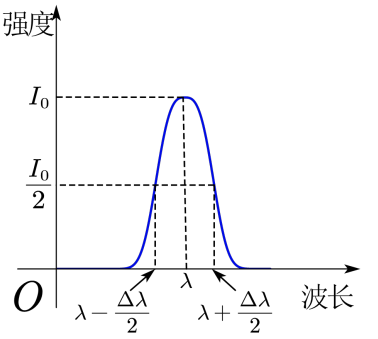


图 1: 谱线及其宽度

设谱线中心处的波长为 λ，强度为 I_0，通常用最大强度的一半所包围的波长范围 $\Delta\lambda$ 当作**谱线宽度**（line width），它是标志谱线单色性好坏的物理量. 普通单色光源，如钠光灯、镉灯、汞灯等，谱线宽度的数量级为 $0.1 \sim 10^{-3}\mathrm{nm}$，而激光的谱线宽度只有 $10^{-9}\mathrm{nm}$，甚至更小.

高斯光束

- 本词条处于草稿阶段.
- 本词条缺少预备知识，初学者可能会遇到困难.
- 翻译成中文

Wave Eq.

$$\nabla^2\mathbf{E}-\frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2}=0 \tag{1}$$

assume propagation within small angle of z axis

$$\mathbf{E}=\hat{\boldsymbol{\epsilon}}E(\mathbf{r},t)=2\hat{\boldsymbol{\epsilon}}U(x,y,z)\mathrm{e}^{\mathrm{i}(kz-\omega t)} \tag{2}$$

$U(x,y,z)$ is the envelope. Plug in, use ‘slowly varying envelope approximation’

$$2\mathrm{i}k\frac{\partial U}{\partial z}=\frac{\partial^2 U}{\partial x^2}+\frac{\partial^2 U}{\partial y^2} \tag{3}$$

The general solution is a linear combination of the following basis

$$U_{mn}(x,y,z)=\frac{C}{w(z)}\exp\left[-\frac{r^2}{w^2(z)}\right]\exp\left[ik\frac{r^2}{2R(z)}\right]H_m\left[\frac{\sqrt{2}x}{w(z)}\right]H_n\left[\frac{\sqrt{2}y}{w(z)}\right]\mathrm{e}^{-\mathrm{i}\phi_{mn}(z)} \tag{4}$$

$$\phi_{mn}(z)=(m+n+1)\tan^{-1}(z/z_R) \tag{5}$$

$$w(z)=w_0\sqrt{1+z^2/z_R^2}\qquad R(z)=z+z_R^2/z\qquad z_R=\pi w_0^2/\lambda \tag{6}$$

$$H_0=1\qquad H_1=2x\qquad H_2=4x^2-1 \tag{7}$$

This is called the **Hermite-Gauss mode**, denoted TEM_{mn}. H are Hermite polynomials and ϕ is the **Gouy phase-shift**, z_R is the **Rayleigh length**. The second exp factor makes the wave front a spherical wave with curvature $R(z)$, because

$$l-R=\sqrt{R^2+z^2}-R\approx\frac{r^2}{2R} \tag{8}$$

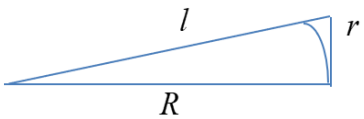


图 1: triangle

TEM_{00} is the **fundamental Gaussian mode**.

In cylindrical coordinates, the basis change to **Laguerre-Gauss modes** TEM^*_{lm}

$$U_{lm}(r,\theta,z)=\frac{C'}{w(z)}\left[\frac{\sqrt{2}r}{w(z)}\right]^{|m|}\exp\left[-\frac{r^2}{w^2(z)}\right]\exp\left[\mathrm{i}\frac{r^2}{2R(z)}\right]L_l^{|m|}\left[\frac{2r^2}{w^2(z)}\right]\mathrm{e}^{\mathrm{i}m\theta}\mathrm{e}^{-\mathrm{i}\phi_{lm}(z)} \tag{9}$$

$$\phi_{lm}(z)=(2l+|m|+1)\tan^{-1}(z/z_R) \tag{10}$$

This is analogous to solving SHO in polar coordiantes while Hermite-Gauss modes are in Cartesian coordinates.

第十三部分

电动力学

第一章

基础

电流

预备知识 导数[158]

若在导线中取一个任意横截面，并定义一个正方向，通过该截面的电流定义为

$$I = \frac{\mathrm{d}q}{\mathrm{d}t} \tag{1}$$

其中 $\mathrm{d}q$ 为 $\mathrm{d}t$ 时间内由正方向通过截面的净电荷量[1].

考虑一根电荷线密度为 λ 的导线中电荷都以速度 v 沿正方向运动，时间 $\mathrm{d}t$ 内通过的长度为 $v\,\mathrm{d}t$，通过的电荷为 $\lambda v\,\mathrm{d}t$，所以电流为

$$I = \lambda v \tag{2}$$

电流密度

预备知识 电流[1123]，流密度[986]

电流某时刻在空间中的分布情况可以用**电流密度**（矢量）$\mathbf{j}(\mathbf{r})$ 来描述，其方向与 $\mathbf{r}$ 处的电荷运动方向相同，详见“流密度[986]”.

$$I = \int \mathbf{j} \cdot \mathrm{d}\mathbf{S} \tag{1}$$

库仑定律

预备知识 万有引力[1027]

学习万有引力以后，库仑力就可以轻而易举地类比过来，所以我们这里不再做重复的推导.

[1]具体来说，正电荷向正方向流动和负电荷向负方向流动，电流都去正号，反之取负号.

先来定义**点电荷**，点电荷就是在质点的基础上（忽略物体的大小与形状），增加了一个总电荷量的属性．与万有引力的式 1 [1027] 类似，两个点电荷间的库仑力的矢量表达式为

$$\mathbf{F}_{12}=k\frac{q_1q_2}{r_{12}^2}\hat{\mathbf{r}}_{12}=\frac{1}{4\pi\epsilon_0}\frac{q_1q_2}{|\mathbf{r}_2-\mathbf{r}_1|^3}(\mathbf{r}_2-\mathbf{r}_1)\tag{1}$$

其中 q_1,q_2 分别是两个点电荷的电荷量．k 是库仑常数，但在大学物理中我们几乎不会见到这个符号，它通常被替换为

$$k=\frac{1}{4\pi\epsilon_0}\approx 8.9876\times 10^9\ \mathrm{Nm^2/C^2}\tag{2}$$

ϵ_0 是**真空中的电介质常数（vacuum permittivity）**，我们以后会见到[1161]．

注意比起万有引力的式 1 [1027]，式 1 没有负号，q_1,q_2 可以是负数（代表负电荷）．我们容易看出两电荷同号相吸，异号相斥．

电场

预备知识 库仑定律[1123]

在经典电磁理论中，电荷与电荷之间的作用力是通过场的作用产生的．所以点电荷 q_1 对另一个点电荷 q_2 的库仑力可以理解为 q_1 在其周围产生了电场对 q_2 的作用（反之亦然）．

注意在点电荷模型中，我们假设一个点电荷产生的电场对其自身合力为 0．另外，我们一般不讨论点电荷所在位置处的电场强度，我们可以说电场在该点处没有定义．

电场是可以叠加的，当空间中有 N 个点电荷，空间中某点（除了这些电荷所在的点）处的电场等于每个点电荷在该点产生的电场之和．注意这个求和是矢量相加．

在某个时刻，空间中的电场是位置 $\mathbf{r}$ 的矢量函数，即任意一个 $\mathbf{r}$ 对应一个唯一矢量 $\mathbf{E}$．我们把这个函数记为 $\mathbf{E}(\mathbf{r})$．当另一个点电荷处于这个电场中，它就会受到电场力

$$\mathbf{F}(\mathbf{r})=q\mathbf{E}(\mathbf{r})\tag{1}$$

这个力也是位置的函数，也就是一个力场[905]．

点电荷的电场

现在我们可以用式 1 推出点电荷电场的表达式．q_1 对 q_2 的库仑力为（式 1 [1123]）

$$\mathbf{F}_{12} = q_2\mathbf{E}_1(\mathbf{r}_2) = \frac{1}{4\pi\epsilon_0}\frac{q_1q_2}{|\mathbf{r}_2-\mathbf{r}_1|^2}\hat{\mathbf{r}}_{12} \tag{2}$$

其中 $\mathbf{E}_1(\mathbf{r})$ 是 q_1 单独产生的电场分布．两边除以 q_2 得

$$\mathbf{E}_1(\mathbf{r}_2) = \frac{1}{4\pi\epsilon_0}\frac{q_1}{|\mathbf{r}_2-\mathbf{r}_1|^2}\hat{\mathbf{r}}_{12} \tag{3}$$

所以，任意点位于 $\mathbf{r}_i$ 处的点电荷 q_i 产生的电场为

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{q_i}{|\mathbf{r}-\mathbf{r}_i|^2}\hat{\mathbf{R}}_i \tag{4}$$

其中 $\hat{\mathbf{R}}_i = (\mathbf{r}-\mathbf{r}_i)/|\mathbf{r}-\mathbf{r}_i|$ 是由 $\mathbf{r}_i$ 指向 $\mathbf{r}$ 的单位矢量．由于电场可叠加，空间中的 N 个电荷产生的电场为

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\sum_{i=1}^{N}\frac{q_i}{|\mathbf{r}-\mathbf{r}_i|^2}\hat{\mathbf{R}}_i \tag{5}$$

连续分布电荷的电场

预备知识 重积分[283]

当电荷时连续分布时，我们可以用电荷密度 $\rho(\mathbf{r})$ 表示其分布情况．式 5 的求和变为体积分．

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^2}\hat{\mathbf{R}}\,\mathrm{d}^3r' \tag{6}$$

其中 $\hat{\mathbf{R}} = (\mathbf{r}-\mathbf{r}')/|\mathbf{r}-\mathbf{r}'|$ 是由 $\mathbf{r}'$ 指向 $\mathbf{r}$ 的单位矢量．类比矢量求和，要对一个矢量函数进行积分，分别对每个分量积分即可．

例 1　均匀带电直棒的电场

设有一均匀带电直棒，长度为 L，总电荷量为 q，线外一点 P 离开直棒的垂直距离为 a，P 点和直棒两端的连线与直棒之间的夹角分别为 θ_1 和 θ_2．如图 1 所示．求 P 点的电场强度．

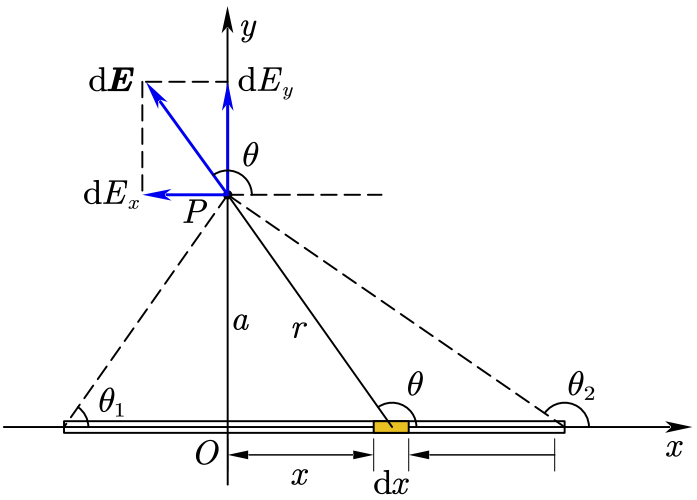


图 1: 均匀带电的细直棒

如图，取 P 点到直棒的垂足 O 为原点，坐标轴 x 沿带电直棒，y 通过 P 点．设直线上每单位长度所带的电荷量为 λ（λ 就是电荷线密度），即 $\lambda = q/L$. 首先在棒上离原点为 x 处取一长度为 $\mathrm{d}x$ 的电荷元，其电荷量为 $\mathrm{d}q = \lambda\,\mathrm{d}x$，在 P 点激发的电场强度 dE 为

$$\mathrm{dE} = \frac{1}{4\pi\varepsilon_0}\frac{\lambda \mathrm{d}x}{r^2}\mathrm{e}_r \tag{7}$$

式中 e_r 是从 $\mathrm{d}x$ 指向 P 点的单位矢量．设 dE 与 x 轴之间的夹角为 θ，则 dE 沿 x 轴和 y 轴的两个分量分别为 $\mathrm{d}E_x = \mathrm{d}E\cos\theta, \mathrm{d}E_y = \mathrm{d}E\sin\theta$．从图可知

$$x = a\tan\left(\theta - \frac{\pi}{2}\right) = -a\cot\theta, \quad \mathrm{d}x = a\csc^2\theta\mathrm{d}\theta, \quad r^2 = x^2 + a^2 = a^2\csc^2\theta \tag{8}$$

所以

$$\mathrm{d}E_x = \frac{\lambda}{4\pi\varepsilon_0 a}\cos\theta\mathrm{d}\theta, \quad \mathrm{d}E_y = \frac{\lambda}{4\pi\varepsilon_0 a}\sin\theta\mathrm{d}\theta \tag{9}$$

将上列两式积分，得

$$\begin{aligned} E_x &= \int \mathrm{d}E_x = \int_{\theta_1}^{\theta_2}\frac{\lambda}{4\pi\varepsilon_0 a}\cos\theta\mathrm{d}\theta = \frac{\lambda}{4\pi\varepsilon_0 a}(\sin\theta_2 - \sin\theta_1) \\ E_y &= \int \mathrm{d}E_y = \int_{\theta_1}^{\theta_2}\frac{\lambda}{4\pi\varepsilon_0 a}\sin\theta\mathrm{d}\theta = \frac{\lambda}{4\pi\varepsilon_0 a}(\cos\theta_1 - \cos\theta_2) \end{aligned} \tag{10}$$

电场强度的大小为

$$E = \sqrt{E_x^2 + E_y^2} = \frac{\lambda}{4\pi\varepsilon_0 a}\sqrt{2 - 2\cos(\theta_1 - \theta_2)} \tag{11}$$

其方向可用 E 与 x 轴的夹角 α 表示：

$$\alpha = \arctan\frac{E_y}{E_x} = \arctan\frac{\sin\theta_2 - \sin\theta_1}{\cos\theta_1 - \cos\theta_2} \tag{12}$$

如果这一均匀带电直棒是无限长的，亦即 $\theta_1 = 0, \theta_2 = \pi$，那么

$$E = \frac{\lambda}{2\pi\varepsilon_0 a} \tag{13}$$

式 13 表明，无限长带电直棒附近某点的电场强度 E 与该点离带电直棒的距离 a 成反比，E 的方向垂直于直棒．若 λ 为正，E 沿 y 轴的正方向；若 λ 为负，E 沿 y 轴的负方向．以上结果对有限长的细直棒来说，在靠近直棒中部附近的区域（$a \ll L$）也近似成立．

例 2　带电圆环的电场

[2]在 yz 平面有一半径为 R 的圆环，均匀带有电荷量 q. 试计算圆环轴线（x 轴）上任意一点 P 处的电场强度．

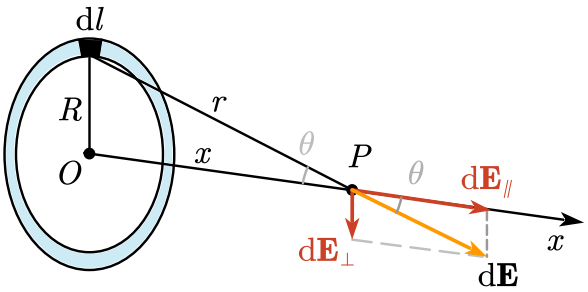


图 2: 均匀带电圆环轴线上任一点处的电场强度

根据本例的轴对称性特征，将 dE 分解为平行于 x 轴线的分量 $\mathrm{dE}_{\parallel}$ 和垂直于轴线的分量 $\mathrm{dE}_{\perp}$，如图 2 所示．

根据对称性，不难发现各电荷元的电场强度在垂直于 x 轴方向上的分矢量 dE 上相互抵消，所以 P 点的合电场强度是平行于 x 轴的那些分矢量 $\mathrm{dE}_{\parallel}$ 的总和，即

$$E = \int \mathrm{d}E_{\parallel} = \int \mathrm{d}E\cos\theta = \frac{1}{4\pi\varepsilon_0}\frac{q}{2\pi R}\frac{\cos\theta}{r^2}\oint \mathrm{d}l \tag{14}$$

也就是

$$E = \frac{qx}{4\pi\varepsilon_0\left(x^2 + R^2\right)^{3/2}} \tag{15}$$

[2]本例可类比“球体的引力场[1028]”中式 2 的推导．

若 q 为正电荷，则 E 的方向沿 x 轴正方向；若 q 为负，则 E 的方向沿 x 轴负方向．当 $x=0$ 时，即在圆环中心，$E=0$；当 $x \gg R$，即 P 点远离圆环时，$(x^2+R^2)^{3/2} \approx x^3$，则上式可近似地写作

$$E=\frac{q}{4\pi\varepsilon_0 x^2} \tag{16}$$

也就是说，与环上电荷全部集中在环心处的一个点电荷所激发的电场强度相同．这点与我们的物理直觉也是相符合的．

例 3　带电圆盘的电场

试计算均匀带电圆盘轴线上与盘心 O 相距为 x 的任一给定点 P 处的电场强度．设盘的半径为 R，电荷面密度（即单位面积上的电荷量）为 σ．如图 3 所示．

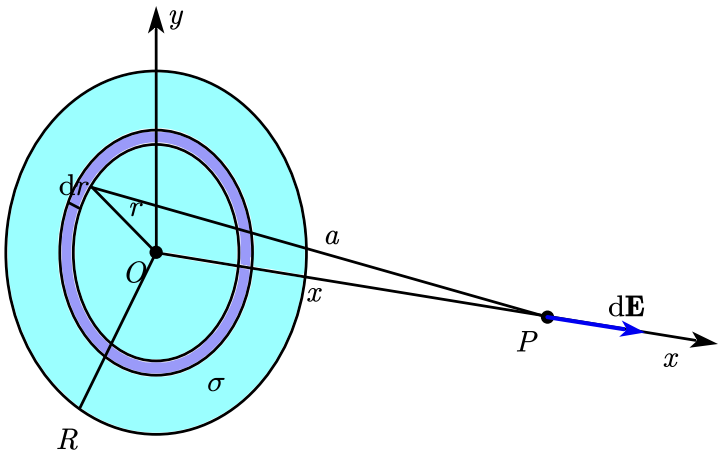


图 3: 带电圆盘的电场

根据上面一个例子，很容易想到，可以把圆盘分成一系列同心的细圆环，每个细圆环可看作为电荷元，圆盘轴上各点处的电场强度就是这些半径不同的细圆环产生的电场强度的叠加．半径为 r，宽度为 $\mathrm{d}r$ 的细圆环所带的电荷量为

$$\mathrm{d}q=\sigma 2\pi r\mathrm{d}r \tag{17}$$

利用例题例 2 里的结果，可得到此带电细圆环在 P 点激发的电场强度为

$$\mathrm{d}E=\frac{x\mathrm{d}q}{4\pi\varepsilon_0\left(x^2+r^2\right)^{3/2}}=\frac{1}{4\pi\varepsilon_0}\frac{x}{\left(x^2+r^2\right)^{3/2}}\sigma 2\pi r\mathrm{d}r=\frac{\sigma x}{2\varepsilon_0}\frac{r\mathrm{d}r}{\left(x^2+r^2\right)^{3/2}} \tag{18}$$

由于各带电细圆环在 P 点激发的电场强度的方向都相同，而带电圆盘的电场强度就是这些带电细圆环所激发的电场强度的矢量和，所以

$$E=\int \mathrm{d}E=\frac{\sigma x}{2\varepsilon_0}\int_0^R\frac{r\mathrm{d}r}{(x^2+r^2)^{3/2}}=\frac{\sigma}{2\varepsilon_0}\left(1-\frac{x}{\sqrt{R^2+x^2}}\right) \tag{19}$$

电场强度 E 的方向与圆盘相垂直，其指向则视 σ 的正负而定．$\sigma>0$，则 E 沿 x 轴正方向；$\sigma<0$，则 E 沿 x 轴负方向．

我们来看两个特殊情况．

若 $R\gg x$，即在 P 点看来可认为均匀带电圆盘为无限大，则 P 点的电场强度可由对上式取极限求得：

$$E=\frac{\sigma}{2\varepsilon_0} \tag{20}$$

只要 P 点与任意带电平面间的距离远小于该点到带电平面边缘各点的距离，即对均匀带电平面中部附近各点来说，这平面都可看作是无限大，其电场强度都可以由式 20 近似表示．这表明无限大均匀带电平面所激发的电场与距离 x 无关，在平面两侧各点有大小相等、方向都与平面相垂直的匀强电场或称作均匀电场（**uniform electric field**）．

若 $x\gg R$，利用二项式定理展开的近似式，略去 R/x 的高次项后可以把式 19 的括号中后一项处理为

$$\left(1+\frac{R^2}{x^2}\right)^{-1/2}=1-\frac{1}{2}\frac{R^2}{x^2}+\frac{3}{8}\left(\frac{R^2}{x^2}\right)^2-\cdots\approx 1-\frac{1}{2}\frac{R^2}{x^2} \tag{21}$$

于是 P 点的电场强度大小为

$$E=\frac{\sigma R^2}{4\varepsilon_0 x^2}=\frac{q}{4\pi\varepsilon_0 x^2} \tag{22}$$

式中 $q=\sigma\pi r^2$ 是圆盘所带电荷量．由此可见，当 P 点离开圆盘的距离比圆盘本身的大小大得多时，P 点的电场强度与电荷量 q 集中在圆盘的中心的一个点电荷在该点所激发的电场强度相同．

磁场

预备知识　电场[1124]

本书中**磁场（magnetic field）**指的是**磁感应强度（magnetic inductance）**，一般记为 $\mathbf{B}$．磁场是三维空间中的矢量场[294]，即空间中的每一点 $\mathbf{r}$ 都对应一个磁场矢量 $\mathbf{B}(\mathbf{r})$．磁场一般用广义洛伦兹力式 4 [1180] 来定义：空间中某点的磁场使得运动时经过该点的点电荷所受的电磁力为

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{1}$$

其中 q 是电荷量，$\mathbf{E}$ 是该点的电场，$\mathbf{v}$ 是速度．

由于历史原因，“磁场强度”这个而名字已经被占用，所以 $\mathbf{B}$ 只好叫做磁感应强度．在比较新的教材中，磁场一般指磁感应强度．磁场也可以使用安培力[1186] 来定义，但安培力在微观本质上也是洛伦兹力．

磁感应强度单位为**特斯拉（Tesla）**，即 $\mathrm{kg \cdot C^{-1} s^{-1}}$．

与电场线一样，我们可以在空间中画出许多有方向的磁感线，使得磁感线上任意一点的方向等于该点处磁场的方向，磁感线在单位横截面的条数与截面处的磁场大小近似成正比．磁感线通常用于磁场的可视化和帮助理解，并不是实际存在的，也并不精确．

电势 电势能

预备知识　力场势能[905]

电场做功

如果空间中存在某种分布的静电场（关于位置的矢量函数）$\mathbf{E}(\mathbf{r})$，一个电荷为 q 的点电荷在任意位置 $\mathbf{r}$ 都会受到电场力

$$\mathbf{F}(\mathbf{r}) = q\mathbf{E}(\mathbf{r}) \tag{1}$$

这样我们就得到了一个由电场力构成的力场，与电场处处成正比．

从某点 $\mathbf{r}_1$ 沿着某个路径移动到另一点 $\mathbf{r}_2$，电场对它做功为（式 4 [903]）

$$W = \int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{F}(\mathbf{r})\,\mathrm{d}\mathbf{r} = q\int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{E}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} \tag{2}$$

势能与保守场

要讨论势能或者电势的概念，我们首先需要确保力场 $\mathbf{F}(\mathbf{r})$ 或电场 $\mathbf{E}(\mathbf{r})$ 是一个保守场，即只有上式的结果只与初末位置有关而与路径无关时才可能存在一个势能函数．以后我们会看到电场可以分为两类：由电荷直接产生的电场是无旋的，即保守场，而由变化磁场产生的电场是有旋的，即非保守场（见“力场势能[905]”，“旋度斯托克斯定理[315]”）．

这个问题有两种解决方法，第一是要求净电荷分布不随时间变化，且电流保持恒定[3]，这类问题称为**静电学（electrostatics）**．在静电学问题中，空间中没有变化的磁场，所以所有的电场都是由净电荷产生的无旋场，总可以定义电势．

第二种方法是，在非静电学的情况下，我们将电场中的保守场部分，即电荷直接产生的电场分离出来定义电势（能），即在式 2 的积分中，$\mathbf{E}(\mathbf{r})$ 并不取总电场，而是取其中的无旋部分．例如电路中的电压（电势差）[1143] 就是通过这种方法定义的．因为一般来说电路不满足静电学的条件，例如电路中的线圈可能存在电磁感应．

电势能

记**电势能（electric potential energy）**函数（关于位置的实函数）为 $E_p(\mathbf{r})$，就有“力场对物体做功等于势能减少”，或者“势能增加等于力场做的负功”即

$$E_p(\mathbf{r}_2) - E_p(\mathbf{r}_1) = -W = -q\int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{E}(\mathbf{r})\cdot \mathrm{d}\mathbf{r} \tag{3}$$

根据上式，我们只能判断两点之间的势能差，若要完全确定势能函数，还需要指定**零势点**的位置 $\mathbf{r}_0$（可以取无穷远）使得 $E_p(\mathbf{r}_0)=0$（类比万有引力势能[1027]）．于是任意一点的势能为

$$E_p(\mathbf{r}) = -q\int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{E}(\mathbf{r}')\cdot \mathrm{d}\mathbf{r}' \tag{4}$$

微分关系为

$$\mathbf{F} = q\mathbf{E}(\mathbf{r}) = -\boldsymbol{\nabla} E_p \tag{5}$$

[3]注意有电流就意味着有电荷在移动，所以我们强调的是“净”电荷保持静止

例 1　匀强电场

假设空间中存在匀强电场 $\mathbf{E} = E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}} + E_z\hat{\mathbf{z}}$，即电场矢量与位置无关. 以原点为零势点 $\mathbf{r}_0$，求点电荷 q 在空间中任意一点 $\mathbf{r}$ 的电势能.

根据式 4，取积分路径为原点到 $\mathbf{r}$ 的线段，则

$$E_p(\mathbf{r}) = -q\mathbf{E}\int_{\mathbf{0}}^{\mathbf{r}} \mathrm{d}\mathbf{r}' = -q\mathbf{E}\cdot\mathbf{r} \tag{6}$$

当 $\mathbf{r}$ 和 $\mathbf{E}$ 同方向时，就得到了高中所学的 $-qEr$.

电势

注意以上公式中，无论是做功还是势能都与电荷量 q 有关（成正比）. 为了更直接地描述电场本身的性质，我们可以把这些公式两边都除以 q. 若定义**电势** $V(\mathbf{r})$ 满足

$$E_p(\mathbf{r}) = qV(\mathbf{r}) \tag{7}$$

则式 3 和式 4 就变为

$$V(\mathbf{r}_2) - V(\mathbf{r}_1) = -\int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{E}(\mathbf{r})\cdot\mathrm{d}\mathbf{r} \tag{8}$$

$$V(\mathbf{r}) = -\int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{E}(\mathbf{r})\cdot\mathrm{d}\mathbf{r} \tag{9}$$

同样，该积分只与初末位置有关而与路径无关. 这样一来，电势就只是电场的属性而与电荷量无关了. 微分关系为

$$\mathbf{E}(\mathbf{r}) = -\boldsymbol{\nabla}V(\mathbf{r}) \tag{10}$$

这相当于式 5 两边除以 q.

两个点电荷的电势

假设空间中有两个点电荷，q_1 产生的电场可以根据式 2 [1124] 计算，现在考虑 q_2 在该场中的势能，取无穷远处为零势点，使用式 4 可以计算 q_2 的势能为（这个过程可以类比万有引力势能[1027] 的推导过程）

$$E_{p2} = -q_2\int_{\infty}^{\mathbf{r}_1} \frac{1}{4\pi\epsilon_0}\frac{q_1}{r_{12}^3}\mathbf{r}_{12}\cdot\mathrm{d}\mathbf{r} = \frac{1}{4\pi\epsilon_0}\frac{q_1q_2}{r_{12}} \tag{11}$$

有趣的是，计算 q_1 在 q_2 产生的电场中的势能结果也是一样的．事实上，电势能不储存于任何一个电荷中而是储存于电场中．以后我们会学习如何直接从总电场分布来计算势能，结果还是和上式一样．所以以后我们对这两种情况不加区分，都记为

$$E_p(r) = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r_{12}} \tag{12}$$

注意当两电荷异号时，q_1q_2 为负数，此时与万有引力一样，距离越近，势能越小．两电荷同号时 q_1q_2 为正，距离越远势能越小．

例 2

两个电荷量为 Q 的电荷分别被固定在 $(-c,0)$ 和 $(c,0)$ 两点处，另一质量为 m 电荷量为 q 的点电荷从 $(0,a)$ 延直线移动到 $(0,b)$，试问它的动能增加了多少？

解：由式 12 可知，点电荷 q 在 $(0,y)$ 处的电势能为

$$E_p = \frac{1}{4\pi\epsilon_0}\frac{2Qq}{\sqrt{y^2+c^2}} \tag{13}$$

所以动能增加等于势能减少，即

$$\Delta E_k = V_a - V_b = \frac{Qq}{2\pi\epsilon_0}\left(\frac{1}{\sqrt{a^2+c^2}} - \frac{1}{\sqrt{b^2+c^2}}\right) \tag{14}$$

点电荷的电势能

由式 12 两边除以 q_2（或直接用式 9 ）得 $\mathbf{r}_1$ 处单个点电荷在 $\mathbf{r}$ 处产生的电势为

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{q}{|\mathbf{r}-\mathbf{r}_1|} \tag{15}$$

由于电场可以矢量叠加，易证电势和电势能也可以叠加

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\sum_i \frac{q_i}{|\mathbf{r}-\mathbf{r}_i|} \tag{16}$$

多个点电荷的电势能

当空间中存在 N 个点电荷，取无穷远处电势为零，那么第 i 个点电荷在剩下 $N-1$ 个点电荷产生的电场中的电势就是

$$V_i(\mathbf{r}_i) = \frac{1}{4\pi\epsilon_0}\sum_j^{(j\neq i)} \frac{q_j}{|\mathbf{r}_j-\mathbf{r}_i|} \tag{17}$$

电势能为 $E_{pi}(\mathbf{r}_i) = q_i V_i(\mathbf{r}_i)$. 这个能量的意义是：保持其他电荷位置不变，把点电荷 q_i 从无穷远处移动到 $\mathbf{r}_i$ 处需要克服电场力所做的功.

如果要计算 N 个电荷的总电势能，我们需要假设开始时这些点电荷两两间距离都是无穷远，然后把他们依次移动到 $\mathbf{r}_i$ 处所需要的总功：移动第 q_1 个不需要做功，移动 q_2 需要客服 q_1 的电场力做功，移动第 j 个电荷需要克服前 $j-1$ 个电荷的电场做功. 所以

$$E_p = \frac{1}{4\pi\epsilon_0} \sum_{i,j}^{(i<j)} \frac{q_i q_j}{|\mathbf{r}_j - \mathbf{r}_i|} \tag{18}$$

其中小于号是为了在求和过程中不重复计算同一对电荷两次. 如果不要求小于，则需要把结果除以 2

$$E_p = \frac{1}{8\pi\epsilon_0} \sum_{i,j}^{(j\neq i)} \frac{q_i q_j}{|\mathbf{r}_j - \mathbf{r}_i|} \tag{19}$$

将式 17 代入该式得

$$E_p = \frac{1}{2} \sum_i q_i V_i(\mathbf{r}_i) \tag{20}$$

显然以上三种对电势能的定义是等效的.

连续电荷分布的情况

将式 17 拓展到连续电荷分布的情况，令电荷密度分布为 $\rho(\mathbf{r})$，则某点的电势可以通过对整个空间体积分得到

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}^3 r' \tag{21}$$

例 3　求有限长细导体棒的电势分布

设带电直导线长为 L，线电荷密度为 λ. 以导体棒中点为坐标原点，建立 xOy 坐标系. 现考虑导体棒的中垂线上与其距离为 d 的 P 点的电势. 如下图所示.

对导线上的某一电荷微元，有 $\mathrm{d}q = \lambda \mathrm{d}x$，该电荷微元到导线点 P 的距离为 $r = \sqrt{x^2 + d^2}$，该电荷微元对点 P 贡献的电势为

$$\mathrm{d}V = \frac{1}{4\pi\epsilon_0} \frac{\mathrm{d}q}{r} = \frac{1}{4\pi\epsilon_0} \frac{\lambda \mathrm{d}x}{\sqrt{x^2 + d^2}} \tag{22}$$

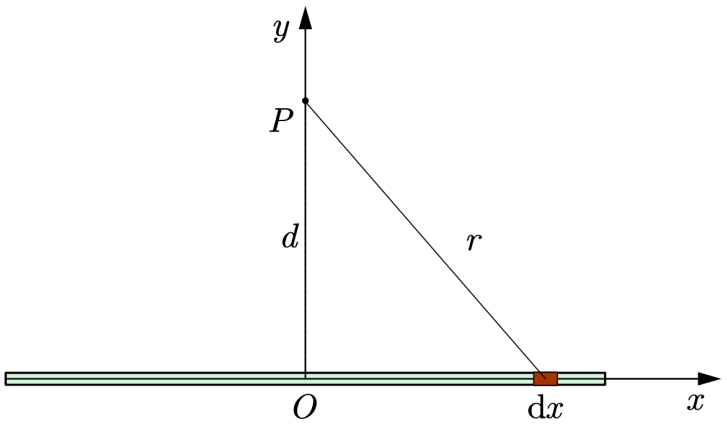


图 1: 有限长导体棒中轴线上的电势

那么整段细导体棒对 P 点的电势的贡献为

$$\begin{aligned} V &= \frac{\lambda}{4\pi\epsilon_0}\int_{-\frac{L}{2}}^{\frac{L}{2}}\frac{\mathrm{d}x}{\sqrt{x^2+d^2}} \\ &= \frac{\lambda}{4\pi\epsilon_0}\ln\left(x+\sqrt{x^2+d^2}\right)\Big|_{-\frac{L}{2}}^{\frac{L}{2}} \\ &= \frac{\lambda}{4\pi\epsilon_0}\ln\frac{\sqrt{L^2+4d^2}+L}{\sqrt{L^2+4d^2}-L} \\ &= \frac{\lambda}{4\pi\epsilon_0}\ln\frac{(\sqrt{L^2+4d^2}+L)^2}{2d} \\ &= \frac{\lambda}{2\pi\epsilon_0}\ln\frac{\sqrt{L^2+4d^2}+L}{2d} \end{aligned} \tag{23}$$

要注意的是，上述推导是以无穷远处为势能零点．若令棒长 $L\to\infty$，将得出任意位置的电势为无穷大的错误结论，此时以离无限长导体棒有限长度的位置为电势零点较为合适．

（未完成，使用式 21，对比例 1 [1124] 的结论）

例 4　求均匀带电圆盘轴线上的电势

设圆盘半径为 R，圆盘的电荷面密度为 σ．以圆盘中心为坐标原点，并在过原点且与圆盘垂直的方向上建立 x 轴．令无穷远处为零势点．参考图如下

用极坐标把圆盘平面分成许多面元，则坐标为 r、φ 的面元的面积为 $\mathrm{d}S = r\mathrm{d}\varphi\mathrm{d}r$，相应的电荷量为 $\mathrm{d}q = \sigma\mathrm{d}S = \sigma r\mathrm{d}\varphi\mathrm{d}r$．那么，该电荷微元对 P 点的电势贡献为

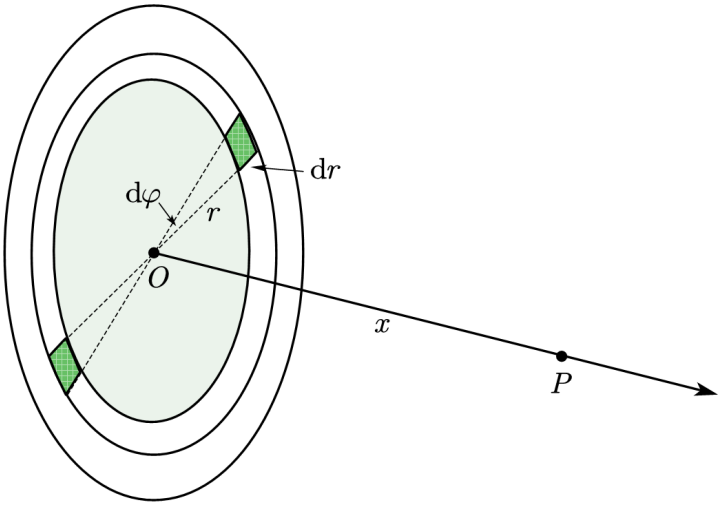


图 2: 均匀带电圆盘轴线上的电势

$$\mathrm{d}V = \frac{1}{4\pi\varepsilon_0}\frac{\mathrm{d}q}{r} = \frac{\sigma}{4\pi\varepsilon_0}\frac{r\mathrm{d}\varphi\mathrm{d}r}{\sqrt{r^2+x^2}} \tag{24}$$

对 $\mathrm{d}V$ 积分，有

$$\begin{aligned} V &= \frac{1}{4\pi\varepsilon_0}\iint\frac{\mathrm{d}q}{r} \\ &= \frac{\sigma}{4\pi\varepsilon_0}\iint\frac{r\mathrm{d}\varphi\mathrm{d}r}{\sqrt{r^2+x^2}} \\ &= \frac{\sigma}{4\pi\varepsilon_0}\int_0^{2\pi}\mathrm{d}\varphi\int_0^R\frac{r}{\sqrt{r^2+x^2}}\mathrm{d}r \\ &= \frac{\sigma}{2\varepsilon_0}\left(\sqrt{R^2+x^2}-x\right) \end{aligned} \tag{25}$$

（未完成，使用式 21，对比例 2 [1124] 的结论）

例 5　求均匀带电球内外的电势分布

设电荷面密度为 σ，令无穷远处为零势点

（未完成，使用式 21，对比例 2 [1124] 的结论）

我们也可以类比式 20，写出连续电荷分布的电势能为

$$E_p = \frac{1}{2}\int V(\mathbf{r})\rho(\mathbf{r})\,\mathrm{d}^3r \tag{26}$$

电偶极子

预备知识　电场[1124]

电偶极子

在电磁学和基础的电动力学框架下，“电荷”可以被认为是一种粒子，它没有大小，同时具有一个特征，该特征用标量（数字）来刻画，也就是电荷．电荷能产生电场，同时由于电荷是一个标量，没有方向之分，这种电场也就是球对称的．

现在我们介绍的模型可以被认为是一种新的粒子，也没有大小，同时具有一个特征，称作电偶极矩，但是该特征用一个向量来刻画．电偶极子也能产生电场，但是由于电偶极矩是向量，有方向之分，这种电场就不再是球对称的了．我们从实际模型出发，来导入电偶极子的概念．

定义 1　物理电偶极子

在电磁学框架下，给定两个等量异号的电荷，假设它们的电荷量分别是 $\pm q$，其中 $q \geq 0$．如果负电荷到正电荷的位移向量是 $\mathbf{r}$，那么我们把这两个粒子构成的系统看成一个叫物理电偶极子（**physical electric dipole**）的对象．物理电偶极子电荷量为 0，但是它有另一个性质叫电偶极矩（**electric dipole moment**）来取代电荷的概念．电偶极矩定义为 $q\mathbf{r}$，是一个向量．

电磁学框架下的“粒子”，是没有大小但可以具有确定空间位置的一个存在．我们希望用同样的视角来看待电偶极子，把它也当成一个没有大小但是具有确定空间位置的一个存在．物理电偶极子显然不能达到这个效果，因为它必须有长度，正负电荷的位置是不一样的——如果一样，电偶极矩就为 0 了．不过，如果我们考虑问题的空间尺度远大于电偶极子的长度，那么我们可以近似地看成正负电荷的位置是一样的，这样就可以确定电偶极子的位置了．更进一步，为了形式上简洁，我们抽象出了一个原本不存在的新概念，理想电偶极子．

定义 2　理想电偶极子

在电磁学框架下引入新的粒子，称作理想电偶极子（**ideal electric dipole**）．

理想电偶极子是一个点粒子，其标量电荷量为 0，但是有一个向量电偶极矩（**electric dipole moment**）．理想电偶极子能产生电场，该电场的分布取决于电偶极矩的大小和方向．

理想电偶极子是物理电偶极子的抽象，抹去了“大小”的概念．从引入理想电偶极子的角度考虑，我们希望的是，在可以忽略物理电偶极子长度的情形下，我们可以直接把它们视为理想电偶极子．这就使得理想电偶极子的性质需要从物理电偶极子来导出了．

这些性质包括理想电偶极子是如何产生电场的．导出的方法是，我们首先计算物理电偶极子的电场分布，这是电磁学框架下可以做到的；然后在**保持电偶极矩不变**的情况下**缩短两电荷间的距离**．正负电荷的间距趋于 0 时的电场分布，就是理想电偶极子的电场分布了．

我们接下来就应用这个思想来推导电偶极子产生的电场和电势．

物理电偶极子的电场

令空间中两个位置不同的点电荷具有等量的异号电荷，则他们构成一对**物理电偶极子**，令他们的电荷量分别为 $-q$ 和 q，位置矢量分别为 $\mathbf{r}_1$ 和 $\mathbf{r}_2$，则它们的总电场为两个电荷各自电场的矢量和（见式 5 [1124]）

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\left(\frac{-q(\mathbf{r}-\mathbf{r}_1)}{|\mathbf{r}-\mathbf{r}_1|^3} + \frac{q(\mathbf{r}-\mathbf{r}_2)}{|\mathbf{r}-\mathbf{r}_2|^3}\right) \tag{1}$$

总电势同样是两个点电荷的电势之和

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\left(\frac{-q}{|\mathbf{r}-\mathbf{r}_1|} + \frac{q}{|\mathbf{r}-\mathbf{r}_2|}\right) \tag{2}$$

这个电偶极子的**电偶极矩（electric dipole moment）**为

$$\mathbf{p} = -q\mathbf{r}_1 + q\mathbf{r}_2 = q(\mathbf{r}_2 - \mathbf{r}_1) \tag{3}$$

如果改用电偶极矩来描述 $|\mathbf{r}| \gg |\mathbf{r}_1| + |\mathbf{r}_2|$ 时的情况，即所谓的“远处”，则电势分布为

$$V_d(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{\mathbf{p}\cdot\mathbf{r}}{r^3} = \frac{1}{4\pi\epsilon_0}\frac{\mathbf{p}\cdot\hat{\mathbf{r}}}{r^2} \tag{4}$$

对电势求梯度就能得到远处的电场分布

$$\mathbf{E}_d(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{1}{r^3}[3(\mathbf{p}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{p}] \tag{5}$$

注意这两个量分别随 r 的平方反比和三次方反比下降.

从远处，也就是 $|\mathbf{r}| \gg |\mathbf{r}_1| + |\mathbf{r}_2|$ 的地方看来，电偶极子的长度是可以忽略的，所以我们可以把式 4 和式 5 看成是**理想电偶极子**的电势和电场分布.

换句话说，物理电偶极子在远处时的性质，就是理想电偶极子在任意位置的性质. 这就是我们需要理想电偶极子所具有的性质.

偶极子远处性质的推导

电势

我们从式 2 出发，并且考虑到 $|\mathbf{r}| \gg |\mathbf{r}_1|+|\mathbf{r}_2|$, $|\mathbf{a}-\mathbf{b}| = \sqrt{\mathbf{a}^2+\mathbf{b}^2-2\mathbf{a}\cdot\mathbf{b}}$, 以及 $\lim_{\mathrm{d}x\to 0}\sqrt{x+\mathrm{d}x} = \sqrt{x}+\mathrm{d}x/2\sqrt{x}$, 从而得到[4]:

$$\begin{aligned}
V(\mathbf{r}) &= \frac{1}{4\pi\epsilon_0}\left(\frac{-q}{|\mathbf{r}-\mathbf{r}_1|}+\frac{q}{|\mathbf{r}-\mathbf{r}_2|}\right)\\
&= \frac{q}{4\pi\epsilon_0}\left(\frac{|\mathbf{r}-\mathbf{r}_1|-|\mathbf{r}-\mathbf{r}_1|}{|\mathbf{r}-\mathbf{r}_2||\mathbf{r}-\mathbf{r}_2|}\right)\\
&= \frac{q}{4\pi\epsilon_0}\left(\frac{\sqrt{\mathbf{r}^2+\mathbf{r}_1^2-2\mathbf{r}\cdot\mathbf{r}_2}-\sqrt{\mathbf{r}^2+\mathbf{r}_2^2-2\mathbf{r}\cdot\mathbf{r}_1}}{|\mathbf{r}-\mathbf{r}_1||\mathbf{r}-\mathbf{r}_2|}\right)\\
\doteq V_d(\mathbf{r}) &= \frac{q}{4\pi\epsilon_0}\left(\frac{\sqrt{\mathbf{r}^2}-\frac{2\mathbf{r}\cdot\mathbf{r}_1}{2\sqrt{\mathbf{r}^2}}-\sqrt{\mathbf{r}^2}+\frac{2\mathbf{r}\cdot\mathbf{r}_2}{2\sqrt{\mathbf{r}^2}}}{\mathbf{r}^2}\right)\\
&= \frac{1}{4\pi\epsilon_0}\left(\frac{q(\mathbf{r}\cdot(\mathbf{r}_2-\mathbf{r}_1))}{\mathbf{r}^2|\mathbf{r}|}\right)\\
&= \frac{1}{4\pi\epsilon_0}\left(\frac{\mathbf{r}\cdot\mathbf{p}}{r^3}\right)
\end{aligned} \tag{6}$$

电场

理想电偶极子的电场，或者物理电偶极子在远处的电场，可以通过求电势的梯度来得到. 引用式 9 [319]，令 $\mathbf{a}=\mathbf{p}$ 为常量，$\mathbf{b}=\mathbf{r}/r^3$，并考虑球坐标系下

[4]符号 $\doteq$ 此处指约等于，或者取极限后的结果.

的 nabla 算子：$\nabla = \hat{r}\partial/\partial r + \hat{\theta} r \partial/\partial\theta + \hat{\varphi} r\sin\theta\partial/\partial\varphi$，则有：

$$
\begin{aligned}
\mathbf{E}_d(\mathbf{r}) &= -\nabla V_d(\mathbf{r}) = -\frac{1}{4\pi\epsilon_0}\nabla\frac{\mathbf{p}\cdot\mathbf{r}}{r^3} \\
&= -\frac{1}{4\pi\epsilon_0}\Big((\mathbf{p}\cdot\nabla)\frac{\mathbf{r}}{r^3} + \mathbf{p}\times(\nabla\times\frac{\mathbf{r}}{r^3})\Big) \\
&= -\frac{1}{4\pi\epsilon_0}(\mathbf{p}\cdot\nabla)\frac{\mathbf{r}}{r^3} \\
&= -\frac{1}{4\pi\epsilon_0}(p_x\frac{\partial}{\partial x} + p_y\frac{\partial}{\partial y} + p_z\frac{\partial}{\partial z})\begin{pmatrix} \frac{x}{(x^2+y^2+z^2)^{3/2}} \\ \frac{y}{(x^2+y^2+z^2)^{3/2}} \\ \frac{z}{(x^2+y^2+z^2)^{3/2}} \end{pmatrix} \\
&= -\frac{1}{4\pi\epsilon_0}\begin{pmatrix} p_x\frac{x^2+y^2+z^2-3x^2}{(x^2+y^2+z^2)^{5/2}} + p_y\frac{-3xy}{(x^2+y^2+z^2)^{5/2}} + p_z\frac{-3xz}{(x^2+y^2+z^2)^{5/2}} \\ +p_x\frac{-3xy}{(x^2+y^2+z^2)^{5/2}} + p_y\frac{x^2+y^2+z^2-3y^2}{(x^2+y^2+z^2)^{5/2}} + p_z\frac{-3yz}{(x^2+y^2+z^2)^{5/2}} \\ +p_y\frac{-3xz}{(x^2+y^2+z^2)^{5/2}} + p_y\frac{-3yz}{(x^2+y^2+z^2)^{5/2}} + p_z\frac{x^2+y^2+z^2-3z^2}{(x^2+y^2+z^2)^{5/2}}) \end{pmatrix} \\
&= \frac{1}{4\pi\epsilon_0}\frac{1}{r^3}\begin{pmatrix} p_x\frac{3x^2-r^2}{r^2} + p_y\frac{3xy}{r^2} + p_z\frac{3xz}{r^2} \\ p_x\frac{3xy}{r^2} + p_y\frac{3y^2-r^2}{r^2} + p_z\frac{3yz}{r^2} \\ p_y\frac{3xz}{r^2} + p_y\frac{3yz}{r^2} + p_z\frac{3z^2-r^2}{r^2}) \end{pmatrix} \\
&= \frac{1}{4\pi\epsilon_0}\frac{1}{r^3}\left[\begin{pmatrix} p_x\frac{3x^2}{r^2} + p_y\frac{3xy}{r^2} + p_z\frac{3xz}{r^2} \\ p_x\frac{3xy}{r^2} + p_y\frac{3y^2}{r^2} + p_z\frac{3yz}{r^2} \\ p_y\frac{3xz}{r^2} + p_y\frac{3yz}{r^2} + p_z\frac{3z^2}{r^2}) \end{pmatrix} - \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix}\right] \\
&= \frac{1}{4\pi\epsilon_0}\frac{1}{r^3}[3(\mathbf{p}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{p}]
\end{aligned}
\tag{7}
$$

总结

理想电偶极子是我们往电动力学框架中添加的模型，其来源是为了方便描述物理电偶极子在远处的性质．理想电偶极子和电荷一样产生电场，区别在于电荷是标量，其产生的电场是球对称的，而理想电偶极子是一个向量，其产生的电场只是关于该向量旋转对称．实际上存不存在理想电偶极子呢？这是电动力学框架下无法回答的．物理理论并不完美刻画现实，它们只是能很好地解释现实的数学模型．对于电的本质是什么，后续的理论，从量子力学，量子场论到规范场论，弦理论等，都会给出各自的解释，但是要记住它们都只是数学模型而已．

除了作为标量的电荷、作为向量的理想电偶极子以外，我们还可以引入更多模型，如电四极子等，用来简化更多实际场景．具体方式请参见**电多极矩**相关词条．

未完成：有了电多极矩后进行引用．

电偶极子 2

预备知识 电偶极子[1137]

可以拓展到多个电荷的情况或者连续分布的情况

$$\mathbf{p}=\sum_i \mathbf{r}_i q_i \tag{1}$$

$$\mathbf{p}=\int \mathbf{r}\rho(\mathbf{r})\,\mathrm{d}V \tag{2}$$

注意只有被求和或者积分的所有电荷之和为零，偶极子 $\mathbf{p}$ 才不随参考系改变

$$\sum_i(\mathbf{r}_i+\mathbf{d})q_i=\sum_i \mathbf{r}_i q_i+\mathbf{d}\sum_i q_i \tag{3}$$

若电荷之和不为零，我们可以定义一个和质心性质类似的中心

$$\mathbf{r}_0=\frac{\sum_i \mathbf{r}_i q_i}{\sum_i q_i} \tag{4}$$

可以证明这个位置和参考系无关．

$$\frac{\sum_i(\mathbf{r}_i+\mathbf{d})q_i}{\sum_i q_i}=\mathbf{r}_0+\mathbf{d} \tag{5}$$

如果以 $\mathbf{r}_0$ 为原点，偶极子为零．

那多级展开到底应该关于哪一点进行呢？笔者认为最好的选择是（想像一个巨大的正电荷左右分别有两个等大反号的小电荷，中心当然应该是在大电荷上）

$$\mathbf{r}_0=\frac{\sum_i \mathbf{r}_i|q_i|}{\sum_i |q_i|} \tag{6}$$

这个位置同样与坐标系无关．

导体

预备知识 电场的高斯定律[1161]

导体的电荷分布

带净电荷的导体有一个显著的性质就是净电荷只分布于其表面，且表面的电荷分布使得导体内部电场强度为零．我们可以用高斯定律来证明这个性质．假设导体达到了静电平衡，则其内部的电荷分布稳定且没有任何（宏观上的）的电流．如果导体内部存在任何净电荷，在净电荷周围做一个高斯面，则高斯面上必然会有电场，进而产生电流，这就违背了静电平衡的假设．

由于内部电场强度为零，导体在静电平衡时处处电势相等，是一个**等势体**．

（未完成）

例 1　匀强电场中的球形导体

匀强电场 $\mathbf{E}_0$ 中有一个半径为 R 的导体球处于静电平衡，求导体球表面的电荷分布．

我们取电场的方向为球坐标的极轴，由问题的对称性，电荷分布关于极轴对称，所以电荷面密度可以表示为 $\sigma(\theta)$．我们现在需要寻找一个 $\sigma(\theta)$ 使得表面电荷在导体球内部可以产生与外电场 $\mathbf{E}_0$ 等大反向的匀强电场．

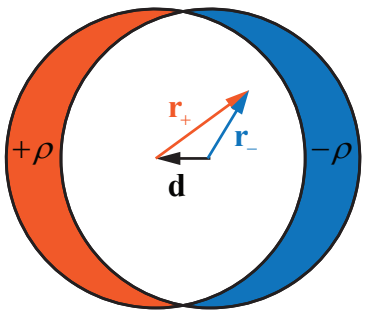


图 1: 两个重叠的均匀带电导体球

这里介绍一种巧妙的办法，我们先来看以下一个模型．先假设不存在外电场，令两个电荷密度分别为 $\pm\rho$ 的均匀带电球在极轴方向错开距离 d （$d < R$）

两球重合的部分正负电荷叠加，净电荷为零．令 $\pm\rho$ 的球心指向空间中某点的矢量分别为 $\mathbf{r}_\pm$，根据例 3 [1161] 中的结论，我们知道两球产生的电场分别为

$$\mathbf{E}_\pm = \frac{\rho}{3\epsilon_0}\mathbf{r}_\pm \tag{1}$$

令 $-\rho$ 的球心指向 $+\rho$ 球心的矢量为 $\mathbf{d}$，有 $\mathbf{r}_+ - \mathbf{r}_- = -\mathbf{d}$，所以两球重合区域中任意一点的总电场为

$$\mathbf{E} = \mathbf{E}_+ + \mathbf{E}_- = \frac{\rho}{3\epsilon_0}(\mathbf{r}_+ - \mathbf{r}_-) = -\frac{\rho}{3\epsilon_0}\mathbf{d} \tag{2}$$

这是一个匀强电场，恰好符合我们的要求．现在就可以令 $\mathbf{d}\to\mathbf{0}$，然后计算电荷面密度，面密度与厚度 h 成正比．由图 2 可得 $h = d\cos\theta$．

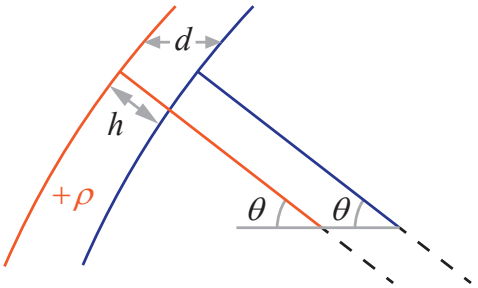


图 2: 两个重叠的均匀带电导体球 θ

所以面密度就是 $\sigma(\theta) = \rho h = \rho d\cos\theta = 3\epsilon_0 E\cos\theta$．若 $\mathbf{E}$ 要与外电场 $\mathbf{E}_0$ 抵消，令 $E = E_0$ 可得

$$\sigma(\theta) = 3\epsilon_0 E_0\cos\theta \tag{3}$$

电压

预备知识 电势电势能[1130]

电压（voltage）就是电势差的同义词，通常在讨论电路时使用，本词条只讨论电路中的电压．电势差的定义为（式 8 [1130]）

$$U_{21} = V(\mathbf{r}_2) - V(\mathbf{r}_1) = -\int_{\mathbf{r}_1}^{\mathbf{r}_2}\mathbf{E}_0(\mathbf{r})\cdot\mathrm{d}\mathbf{r} \tag{1}$$

在“电势电势能[1130]”中，我们强调了要定义电荷的电势能或电势，我们必须要使用无旋的电场（保守场），而电路中一般来说既存在无旋场也存在有旋场（例如线圈的感生电场）. 这时我们规定式 1 中的 $\mathbf{E}_0(\mathbf{r})$ **只包含电路中的净电荷产生的电场**.

以下我们来举若干例子说明. 首先讨论电阻两端的电压，见电阻[1157]. 再来看另一些例子

例 1　电容的电压

令平行板电容器[1145]两端分别有正负净电荷，我们知道在平行板之间会产生近似匀强的电场 $\mathbf{E}$，这个电场是电路中的净电荷产生的，所以需要计入式 1. 所以从带正电荷的平行板到带负电荷的平行板进行积分，结果为 $-Ed$（d 是平行板的距离）. 所以带电容器带正电荷的一端的电势比另一端要高.

例 2　化学电源

对于化学电源（即电池）两端产生的电压，我们应该如何通过上述的定义理解呢？这里介绍一个简单的模型. 假设电池的两端构成一个平行板电容器（例 2 [1145]），带正电荷的一端是电源的正极，带负电的一端是电源的负极，两个平行板之间存在导电材料，使电子可以在它们之间移动.

两极板之间存在从正极到负极的匀强电场，使得它们之间的电子受到从负极到正极的电场力. 所以如果电池中没有化学反应，电子将不断地从负极移动到正极，使得两极板的净电荷消失，电源两端电势差变为零. 但若有化学反应，在这个模型中化学反应的作用相当于给每个电子施加一个与电场力等大反向的恒力，使得两极板间的电子合力为零，以保持两平行板上的净电荷不被中和.

当我们将电源接入电路开始工作时，负极的电子不断通过外电路移动到电源的正极，这时平行板之间的电场稍微变弱，电场力变得比化学反应提供的恒力略弱，电子就在这个恒力的作用下不断从平行板的正极移向负极，从而保持电源两端的电势差，产生源源不断的电流.

例 3　感生电压

假设我们有一个 N 匝的不闭合线圈，两端接在理想电压表上. 若线圈中存在变化的磁场，磁通量 $\Phi(t)=\alpha t$ 随时间线性增加，那么根据电磁感应我们知道电压表会显示稳定的读数 α.

注意这并不是一个静电学问题，变化的磁场会沿着线圈产生涡旋电场，这是一个有旋场，所以在计算电路电压时我们不能将这部分电场算入（详见“电势[1130]”）.

那为什么线圈两端还会产生电压呢？因为整个导线（忽略电阻）作为一个导体，其内部的总电场必须为零，所以当线圈放在涡旋电场中，线圈就会自动调整其净电荷分布，使线圈中处处存在与涡旋电场等大反向的，由净电荷产生的电场．而根据定义，这部分电场需要算入式 1 的积分中．由于线圈有 N 匝，从导线的一端线积分到另一端，需要延导线一起环绕 N 圈，所以线圈两端的电压（也叫感生电动势）与匝数成正比.

由此我们可以看到，“导体内部电场为零/导体是等势体”的说法和“导线两端存在电压/电势差”并不矛盾．前者指的是总电场，后者指的是总电场中的无旋部分的电势.

例 4　动生电压

假设一根导体在间距为 L 的两个导体轨道上以速度 v 匀速运动，导体电阻不计．空间中存在垂直于轨道平面向下的匀强磁场，大小为 B．求轨道之间的电压.

根据高中所学的洛伦兹力计算方法，导体中的电子将受到向右的洛伦兹力 $F = evB$，导体为了保持内部平衡（即电子受力为零），会在其表面重新分布净电荷，产生一个向左的静电场使得 $eE = evB$．所以轨道之间的电势差为 $U = EL = vBL$，且右端的电势高于左端.

注意该例不符合“导体内部电场为零/导体是等势体”，因为存在磁场时这只对静止的导体适用．在学习“电磁场的参考系变换[1246]”后我们会发现如果我们将参考系与导体一起移动，在这个参考系中导体棒内部电场的确等于零.

电容

预备知识　电势电势能[1130]

单个导体的电容量

一个导体的**电容量（capacitance）**等于电荷量除以导体电势（导体平衡时为等势体）

$$C = \frac{Q}{V} \tag{1}$$

要证明任意导体的势能和电荷量成正比，根据式 21 [1130]，当导体上的电荷密度分布 $\rho(\mathbf{r})$ 乘以常数 λ，空间中任意一点的电势同样会乘以 λ. 所以对于某个形状的导体，其电容是固定的.

例 1　导体球的电容

若规定无穷远处为零势点，由高斯定律，半径为 R，带电为 Q 的导体球的电势为 $V = Q/(4\pi\epsilon_0 R)$，所以其电容为

$$C = \frac{Q}{V} = 4\pi\epsilon_0 R \tag{2}$$

两导体之间的电容量

若两导体带等量异种电荷[5] $\pm Q$，电势差为 V，则两导体间的电容量同样被定义为

$$C = \frac{Q}{V} \tag{3}$$

现在来证明两导体间的电容只与他们的形状，相对位置以及空间中的电介质分布有关，而与电荷量无关. 这就要求证明电势差始终与电荷 Q 成正比，如果我们假设量导体表面的电荷面密度 σ 始终与 Q 成正比（证明见，那么由库仑定律可知空间中任意一点的场强也与 Q 成正比. 而电势差

$$V = \int_{+}^{-} \mathbf{E}(r) \cdot \mathrm{d}\mathbf{l} \tag{4}$$

（式中的“±”分别代表带电荷量为 $\pm Q$ 的导体表面上的一点，积分路径任意选取）和场强成正比也就是和 Q 成正比. 证毕.

例 2　平行板电容器

电容器中最常见也是最基本的模型就是平行板电容器，我们假设空间中存

[5]这里的 Q 不一定要求大于零，± 号只是一个用于区分两个导体的记号而已，不代表电荷的正负.

在均匀的电介质，介电常数为 ϵ 空间中两块面积为 S 的方形导体板相距为 d 平行放置，带电量分别是 $\pm Q$，如果忽略边缘效应（即假设只有两板之间的长方体空间中存在匀强电场），由高斯定律，两板之间的电场为 $E = \sigma/\epsilon = Q/(S\epsilon)$，电势差为 $V = Ed = Qd/(S\epsilon)$，所以电容等于

$$C = \frac{Q}{V} = \epsilon\frac{s}{d} \tag{5}$$

可见在相同介质中，平行板电容器的电容量与板的面积成正比，与距离成反比，比例系数为 ϵ. 这就是为什么 ϵ 也被称为电容率. 当不存在电介质时，$\epsilon = \epsilon_0$，所以 ϵ_0 被称为真空中的电容率.

例 3　同心球壳电容器

若有两个半径分别为 r 和 R 的同心球壳导体（$r < R$），电荷量分别为 $+Q$ 和 $-Q$，求两球壳间的电容量.

由高斯定律两球壳间的电场只与小球壳的电荷量有关，所以两球壳的电势差为

$$V = \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{r} - \frac{1}{R}\right) \tag{6}$$

所以电容量为

$$C = \frac{Q}{V} = \frac{4\pi\epsilon_0 rR}{R - r} \tag{7}$$

当 $r \to R$ 时，球壳的面积为 $S = 4\pi R^2$，两板间距为 $d = R - r$，则上式的电容趋于平行板电容器的电容.

对比式 7 和式 2 可以发现两个足够近的导体获得的电容量要比单个导体大得多，这也是为什么电路中的电容都有两个极.

电压与电流的关系

若电容器两端的电压随时间变化为 $V(t)$，那么如何计算流过电容器的电流[6]呢？如果我们定义从 $+Q$ 极流入从 $-Q$ 极流出的电流方向为正方向，那么 $I = \mathrm{d}Q/\mathrm{d}t$，对 $Q = CV$ 两边求时间导数得

$$I = C\frac{\mathrm{d}V}{\mathrm{d}t} \tag{8}$$

[6]我们在讨论电路的时候习惯性地说电流“流过电容器”，而事实上并没有电荷在电容器内部从一极移动到另一极.

习题 1

试证明在交流电路中，流经电容的电流的相位比电压的相位提前 $\pi/2$.

电容器的能量

为了改变电容器的电荷量 Q，我们需要将电荷从一个导体移动到另一个导体，为了不影响电荷及电场的分布，从 $Q=0$ 开始，我们每次只从 $-Q$ 移动极少量的电荷 ΔQ 到 $+Q$，移动的过程中外力需要克服电场力做功 $V\Delta Q$，而 V 又是 Q 的函数，第 i 次移动前，电荷量为 $\pm Q_i$，由定积分[198] 的思想

$$W=\int V(Q)\,\mathrm{d}Q=\int \frac{Q}{C}\,\mathrm{d}Q=\frac{1}{2}\frac{Q^2}{C}=\frac{1}{2}CV^2 \tag{9}$$

可见电容器的能量与电荷量或电势的平方成正比. 注意电容器的能量本质上是导体表面上连续电荷分布的电势能，所以另一种计算上式的方法就是直接使用式 26 [1130] 得

$$W=\frac{1}{2}[QV_+ + (-Q)V_-]=\frac{1}{2}QV=\frac{1}{2}\frac{Q^2}{C}=\frac{1}{2}CV^2 \tag{10}$$

电介质的简单模型

预备知识 电偶极子[1137]

先来介绍一个简单的模型：假设**电介质（dielectric）**是一种完全不导电的物质，其中正电荷密度（假设非常大）分布为 $\rho(\mathbf{r})$. 若电介质中处处无电场，那么负电荷密度为 $-\rho(\mathbf{r})$，电解质呈电中性；若某点附近有电场，该点处负电荷不动，正电荷向电场方向移动一个非常小的距离 $\mathbf{d}$. 那么我们就把介质中某一点的**极化强度（polarization density 或 polarization）**定义为

$$\mathbf{P}=\rho\mathbf{d} \tag{1}$$

若取一个体积为 V 的体积元，把连续电荷分布看成是由许多点电荷构成的，令其中的正点电荷为 q_i，每个正电荷对应一个负电荷 $-q_i$，相对位移为 $\mathbf{d}$，则极化强度可用电偶极矩式 3 [1137] 表示为

$$\mathbf{P}=\lim_{V\to 0}\sum_i \frac{q_i}{V}\mathbf{d}=\lim_{V\to 0}\frac{1}{V}\sum_i \mathbf{p}_i \tag{2}$$

所以极化强度也可以理解为电偶极矩的体密度.

极化强度与电场

若电解质中某点处电场为 $\mathbf{E}$，如果该介质是（a) 各向同性 (b) 均匀 (c) 线性的，该点处的极化强度与电场的关系就满足

$$\mathbf{P} = \chi\epsilon_0\mathbf{E} \tag{3}$$

其中常数 χ 叫做**电极化率（electric susceptibility）**，ϵ_0 是真空中的电容率．特殊地，我们可以认为真空也是一种极化率为零的电介质．

若电介质不满足上述三个条件，在最一般的情况下，电极化率可以用一个 3×3 的矩阵 $\boldsymbol{\chi}$ 表示，称为**电极化张量**．

电介质的微观结构

- 需要与“电介质的简单模型”整合

预备知识 电偶极子[1137]

电介质（dielectric）是电阻率很大、导电能力很差的物质，其主要特征在于它的原子或分子中的电子与原子核的结合力很强，电子处于束缚状态．当电介质处在电场中时，在电介质中，不论是原子中的电子、还是分子中的离子或是晶体点阵上的带电粒子，在电场的作用下都会在原子大小的范刚内移动，当达到静电平衡时，在电介质表面层或在体内会出现极化电荷，这个现象称作电介质的**极化（polarization）**．

下面就研究电场与电介质间的相互作用，从而说明电介质的一些性质．

电介质的电结构

电介质中每个分子都由正、负电荷组成，是一个复杂的带电系统一般来说，正、负电荷在分子中都不集中在一点，但在考虑这些电荷在离分子较远处所产生的电场时，或是考虑一个分子受外电场作用时，可以认为分子中的全部正电荷可用一等效正电荷来代替，全部负电荷用一等效负电荷来代替等效正、负电荷在分子中所处的位置，分别称为该分子的**正负电荷“中心”**．

按照电介质分子内部的电结构不同，可以把电介质分为两大类：有极分子和无极分子．有一类电介质，如水蒸气（H_2O）、氯化氢（HCl）、一氧化碳（CO）、氨（NH_3）等，分子内正、负电荷中心不相重合，这类分子称为**有极分子（polar molecules）**．设有极分子的正电荷中心和负电荷中心之间的距离为 l，分子中全部正电荷或负电荷的总电荷量为 q，则有极分子的等效电偶极矩 $\mathrm{p} = q\mathrm{l}$．整块电介质可以看成是无数个电偶极子的聚集体，虽然每一个分子的等效电偶极矩不为零，但由于分子的无规则热运动，各个分子的电偶极矩的方向是杂乱无章地排列的，所以不论从电介质的整体来看，还是从电介质中的某一小体积（其中包含有大量的分子）来看，其中各个分子电偶极矩的矢量和 $\sum \mathrm{p}$ 平均说来等于零，电介质是呈电中性的．另有一类电介质，如氢（H_2）、氦（He）、氮

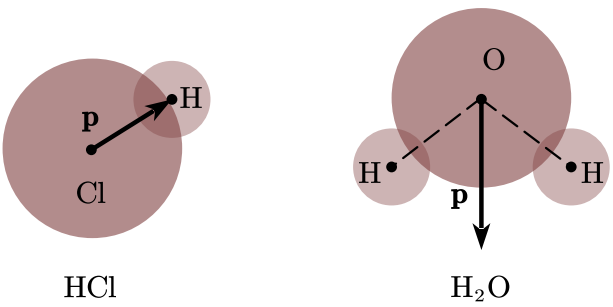


图 1: 有极分子的电偶极矩示例

（N_2）、甲烷（CH_4）等，它们的正、负电荷的中心重合在一起，它的等效电偶极矩等于零，凡属于这种类型的分子叫做**无极分子（nonpolar molecules）**．由于每个分子的等效电偶极矩 $\mathrm{p} = 0$，电介质整体也是呈电中性的．

基于有极分子和无极分子的电结构不同，它们在外电场中所受到的作用也不相同，下面将分别讨论．

电介质的极化

无极分子电介质的位移极化

当无极分子电介质处在外电场中时，在电场力作用下分子中的正、负电荷中心将发生相对位移，形成一个电偶极子，它们的等效电偶极矩 p 的方向都沿着电场的方向图 2 (b) 所示．在电介质内部，由于相邻电偶极子的正负电荷相互靠近，如果电介质是均匀的，则在它内部仍然保持电中性，但是在电介质的

两个和外电场强度 E_0 相垂直的表面层里（厚度为分子等效电偶极矩的轴长 l）将分别出现正电荷和负电荷．这些电荷不能离开电介质，也不能在电介质中自由移动，我们称之为**极化电荷（polarized charge）**．把在电场作用下能移动一宏观距离的电荷统称为**自由电荷（free charge）**．这种在外电场作用下，在电介质中出现极化电荷的现象叫做**电介质的极化**．分子的电偶极矩 p 的大小与电场强度成正比，外电场越强，每个分子的正、负电荷中心之间的相对位移越大，分子的电偶极矩也越大，电介质两表面上出现的极化电荷也越多，被极化的程度越高．当外电场撤去后，正、负电荷的中心又重合在一起，见 $p = 0$，图 2 (a)．电介质表面上的极化电荷也随之消失．由于无极分子的极化来自正、负电荷中心的相对位移，所以常叫做**位移极化（displacement polarization）**．

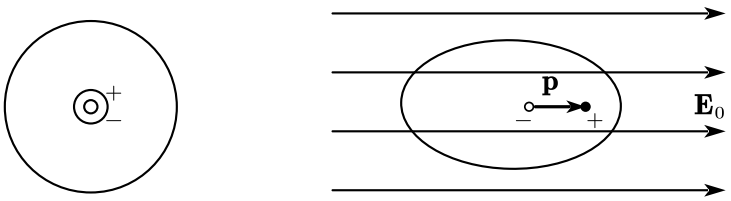


图 2: 无极分子极化示意图

有极分子电介质的取向极化

对于有极分子电介质来说，每个分子本来就等效为一个电偶极子，它在外电场的作用下，将受到力矩的作用，使分子的电偶极矩 p 转向电场的方向，如图 3 (b) 所示．这样，宏观上看，在电介质与外电场垂直的两表面上也会出现极化电荷，如图 3 (c) 所示．当外电场撤去后，由于分子无规热运动和分子间的相互碰撞都会破坏分子偶极矩沿电场方向的取向排列，使之回到沿各个方向的均匀分布（图 3 (a)），表面的极化电荷也随之消失．可见，有极分子电介质的极化程度取决于外电场的强弱和电介质的温度，外电场越强且温度越低，分子电偶极矩沿电场取向排列的概率也越大．有极分子的极化就是等效电偶极子转向外电场的方向，所以叫做**取向极化（orientation polarization）**．一般说来，分子在取向极化的同时还会产生位移极化，但是，对有极分子电介质来说，在静电场作用下，取向极化的效应比位移极化的效应强得多，因而其主要的极化

机理是取向极化.

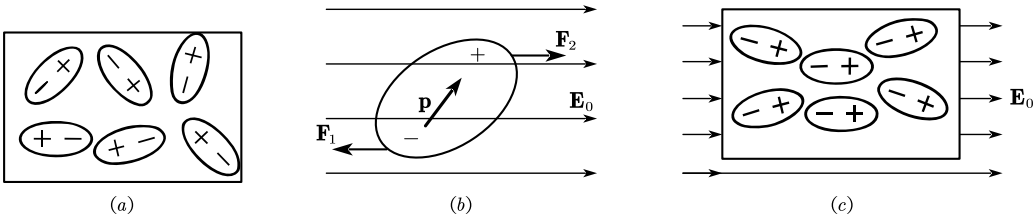


图 3: 有极分子的极化示意图

当前广为应用的家用微波炉，就是介质分子（水分子）在高频电场中反复极化的一个实际应用！水分子作为一个有极分子，其电偶极子在电场力矩的作用下，要转向与外电场方向一致的排列. 如果电场方向交替变化，水分子的电偶极矩 p 也要跟随电场方向反复转动，在这个过程中水分子作高频振动，引起快速摩擦而产生热量. 当微波频率为 2450MHz 时，水分子能极大地吸收微波的电磁能量，达到加热、煮熟食物的目的. 微波在金属面上反射，却很容易穿透空气、玻璃、塑料等物质，且极大地被食物中的水、油、糖所吸收.

电极化强度

- 需要与“电介质的简单模型”整合

预备知识 电介质[1149]

在电介质[1149] 一节中，我们从分子的电结构出发，说明了两类电介质极化的微观过程虽然不同，但宏观的效果却是相同的，都是在电介质的两个相对表面上出现了异号的极化电荷，在电介质内部有沿电场方向的电偶极矩. 因此下面从宏观上描述电介质的极化现象时，就不分两类电介质来讨论了.

在电介质内任取一物理无限小的体积元 ΔV（但其中仍有大量的分子），当没有外电场时，这体积元中所有分子的电偶极矩的矢量和 $\sum \mathrm{p}$ 等于零. 但是，在外电场的影响下，由于电介质的极化，$\sum \mathrm{p}$ 将不等于零. 外电场愈强，被极化的程度愈大，$\sum \mathrm{p}$ 的值也愈大. 因此我们取单位体积内分子电偶极矩的矢量和，即

$$\mathrm{P} = \frac{\sum \mathrm{p}}{\Delta V} \tag{1}$$

作为量度电介质极化程度的基本物理量，称为该点（ΔV 所包围的一点）的电极化强度（electric polarization）或 P 矢量．在国际单位制中，电极化强度的单位是 $\mathrm{C/m^2}$．

电极化强度与极化电荷的关系

- 这只是一个很特殊的例子，一般规律是 $\nabla \times \mathbf{P} = \rho$ 未完成

预备知识 电极化强度[1152]

极化电荷是由于电介质极化产生的，因此电极化强度与极化电荷之间必定存在一定的关系．对于均匀电介质，其极化电荷只集中在表面层里或在两种不同的界面层里．电介质极化后产生的一切宏观效应就是通过这些电荷来体现的．下面我们就来研究均匀电介质极化电荷面密度与电极化强度之间的关系．

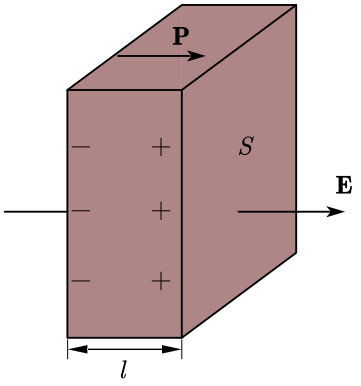


图 1: 极化电荷面密度与电极化强度

设有一厚为 l、表面积为 S 的电介质薄片图 1 放置在一均匀电场 E 中，那么在薄片两表面产生了极化电荷，薄片的电极化强度 P 平行于电场强度 E．薄片总的电偶极矩 $\sum \mathrm{p}$ 是电极化强度的大小与薄片体积的乘积 PSl，这相当于薄片表面的极化电荷 q' 与薄片两表面正负电荷分开的距离 l 的乘积，即

$$\left|\sum \mathrm{p}\right| = PSl = q'l \tag{1}$$

因此，薄片表面的极化电荷面密度就等于电极化强度的大小，即 $\sigma' = P$. 这个结果假定了薄片表面与 P 垂直. 在一般情况下，设 $\mathrm{e_n}$ 为薄片表面的单位法向矢量，那么

$$\sigma' = \mathrm{P} \cdot \mathrm{e_n} = P_\mathrm{n} \tag{2}$$

即介质极化所产生的极化电荷面密度等于电极化强度沿介质表面外的分量. 在薄片侧面，由于 P 的方向与侧面法线垂直，所以侧面上的极化电荷面密度为零.

极化电流

预备知识 电极化强度与极化电荷的关系[1153]

这里推导介质中麦克斯韦方程组的 $\boldsymbol{\nabla} \times \mathbf{H}$ 的那条.

首先证明磁介质产生的电流 $\boldsymbol{\nabla} \times \mathbf{M} = \mathbf{j}_M$，即证 $\oint \mathbf{j}_M \cdot \mathrm{d}\mathbf{s} = \oint \mathbf{M} \cdot \mathrm{d}\mathbf{l}$. 假设磁偶极子都是小线圈组成，曲面内部净电流为零，曲面边界只有穿过小线圈才能在曲面上产生净电流.

$$I = I_1(\pi R^2 \hat{\mathbf{n}} \cdot \mathrm{d}\mathbf{l})n = \mathbf{M} \cdot \mathrm{d}\mathbf{l} \tag{1}$$

证毕.

再来看电介质的极化电流

$$\mathbf{j}_p = \frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} \tag{2}$$

所以根据

$$\boldsymbol{\nabla} \times \mathbf{B} = \mu_0(\mathbf{j}_f + \mathbf{j}_M + \mathbf{j}_p + \mathbf{j}_E) = \mu_0 \mathbf{j}_f + \mu_0 \boldsymbol{\nabla} \times \mathbf{M} + \mu_0 \frac{\partial \mathbf{P}}{\partial t} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \tag{3}$$

定义**磁场强度**为

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M} \tag{4}$$

则

$$\boldsymbol{\nabla} \times \mathbf{H} = \mathbf{j}_f + \frac{\partial \mathbf{D}}{\partial t} \tag{5}$$

另外，假设磁介质为线性

$$\mathbf{M} = \chi_M \mathbf{H} \qquad \mu_r = 1 + \chi_M \qquad \mathbf{H} = \mathbf{B}/\mu \qquad \mathbf{j}_M = \chi_M \mathbf{j}_f \tag{6}$$

这就是说，$\chi > 0$ 的磁介质会使电流加强.

介质中的静电场

预备知识 电极化强度与极化电荷的关系[1153]

如果把激发外电场的原有电荷系称为自由电荷，并用 E_0 表示它们所激发的电场强度，而用 E' 表示极化过程完成之后极化电荷所激发的电场强度．那么，空间任一点最终的合电场强度 E 应是上述两类电荷所激发电场强度的矢量和，即

$$\mathbf{E} = \mathbf{E}_0 + \mathbf{E}' \tag{1}$$

由于在电介质中，自由电荷的电场与极化电荷的电场的方向总是相反，所以在电介质中的合电场强度 E 与外电场强度 E_0 相比显著地削弱了．

对于各向同性线性电介质，电极化强度 P 和介质内部的合电场强度 E 的关系为

$$\mathrm{P} = \chi_\mathrm{e}\varepsilon_0\mathrm{E} \tag{2}$$

式中的比例因数 χ_e 和电介质的性质有关，叫做电介质的**电极化率（electric susceptibility）**，是量纲为 1 的量．

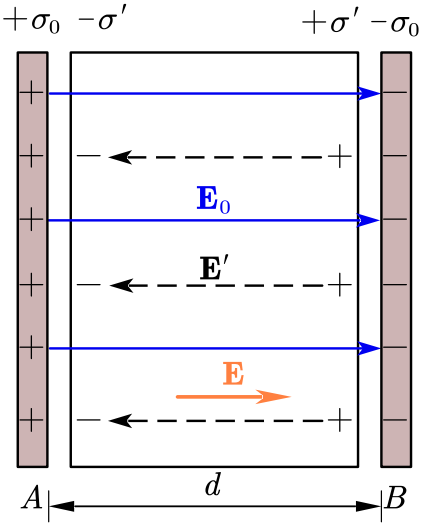


图 1: 电介质中的电场强度

为了定量地了解电介质内部电场强度被削弱的情况，我们讨论如下特例：图 1 表示在两块“无限大”极板间充有电极化率为 χ_{e} 的均匀电介质，设两极板上的自由电荷面密度为 $\pm\sigma_0$，电介质表面上的极化电荷面密度为 $\pm\sigma'$．自由电荷的电场强度大小 $E_0=\sigma_0/\varepsilon_0$，在图中用实线表示；极化电荷的电场强度大小为 $E'=\sigma'/\varepsilon_0$，在图中用虚线表示．E' 的方向和 E_0 的方向相反，因此极板间电介质中的合电场强度 E 的大小为

$$E=E_0-E'=\frac{\sigma_0}{\varepsilon_0}-\frac{\sigma'}{\varepsilon_0} \tag{3}$$

考虑到极化电荷面密度为 σ，极板间电介质中的合电场强度 E 的大小又可写为

$$E=E_0-\frac{P}{\varepsilon_0}=E_0-\chi_{\mathrm{e}}E \tag{4}$$

即

$$E=\frac{E_0}{1+\chi_{\mathrm{e}}} \tag{5}$$

说明电介质内部的电场强度 E 被削弱为外电场强度 E_0 的 $1+\chi_{\mathrm{e}}$．下面我们将看到 $(1+\chi_{\mathrm{e}})$ 正是电介质的相对电容率 ε_{r}．我们知道，两极板间的电势差为

$$U=Ed=\frac{\sigma_0 d}{\varepsilon_0\left(1+\chi_{\mathrm{e}}\right)} \tag{6}$$

设极板的面积为 S，则极板上总的电荷量为 $q=\sigma_0 S$，按电容器电容的定义，极板间充满均匀电介质后的电容为

$$C=\frac{q}{U}=\frac{\varepsilon_0\left(1+\chi_{\mathrm{e}}\right)S}{d}=\left(1+\chi_{\mathrm{e}}\right)C_0 \tag{7}$$

而我们知道 $C=\varepsilon_{\mathrm{r}}C_0$，所以

$$\varepsilon_{\mathrm{r}}=1+\chi_{\mathrm{e}} \tag{8}$$

这就解释了电容器中充满电介质后其电容增大的实验事实．又令

$$\varepsilon=\varepsilon_r\varepsilon_0=\left(1+\chi_{\mathrm{e}}\right)\varepsilon_0 \tag{9}$$

称作电介质的**电容率**或**介电常量**．与真空中的电容率 ε_0 有相同的单位．电极化率 χ_{e}、相对电容率 ε_{r}，和电容率 ε 都是表征电介质性质的物理量，三者中知道任何一个即可求得其他两个．

这里要特别说明一点，式 8 虽然是从平行板电容器中均匀电介质的特例引出的，但它却是普遍适用的．而式 5 表明，在均匀电介质充满整个电场的情况下，电介质内部的电场强度 E 为电场强度 E_0 的 $1/\varepsilon_{\mathrm{r}}$ 倍，这一结论并不是普遍成立的，但电介质内部的电场强度通常要减弱，这个现象却是普遍成立的．

此外做一个小扩展，上面研究的是各向同性电介质，电极化率和电容率都是常量，但自然界也存在一些电介质，在一定的温度范围内电容率随电场强度而变化，它们的极化规律有着复杂的非线性关系．例如钛酸钡等，在外电场撤除后仍保留有剩余的极化，这样的材料称作**铁电体**，另一类电介质在外力的作用下发生机械变形（拉伸或压缩）时，也能产生电极化现象，称作**压电效应**．如石英晶体等就具有压电效应．压电效应的反效应叫做电致伸缩，即晶体在电场中会产生伸长或收缩的效应．还有一类材料在外电场撤销后，会长期保留其极化状态，就像永磁体保留有磁性一样．这样的电介质称作**永驻体**．上述这些具有特殊性质的材料有着重要的应用，可以制成各种换能器和传感器，满足人们不同的需求．它们是材料科学的基础研究内容之一．

电阻 欧姆定律

预备知识 电流[1123]，电压[1143]

欧姆定律（Ohm's law）为

$$U = IR \tag{1}$$

其中 U 是**电阻器**[7]两端的电压，I 是流经电阻器的电流，R 是电阻器的电阻．注意电压产生的电场是电阻器中电流的原因．

作为一个理想模型，柱形（例如长方体，圆柱体）电阻器的电阻为

$$R = \frac{\varrho L}{S} \tag{2}$$

其中 S 为电阻的横截面积，L 为电阻的长度，ϱ 为材料的**电阻率**．电阻率体现了材料电阻能力，和材料性质有关，也可能会随温度，压强，光照（例如光敏电阻），等环境因素变化．我们也通常把电阻率的倒数 $1/\varrho$ 叫做**电导率**．

[7]在可能混淆的情况下，为了区分“电阻”这个词作为电路元件和作为物理量，我们把前者叫做“电阻器”，后者叫做“电阻”．

微观形式的欧姆定律为

$$\mathbf{E} = \varrho \mathbf{j} \tag{3}$$

这个公式告诉我们，电阻材料中某点的电流密度与电场成正比.

电阻的简单模型

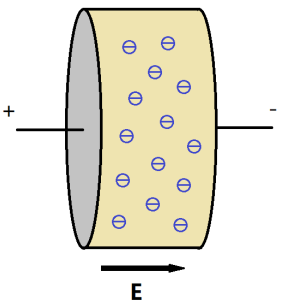


图 1: 简单的电阻模型：一个平行板电容器中均匀填充导电介质

我们这里用一个简单的经典力学模型推导上文中的概念和公式，但严格来说，这个推导需要使用量子力学和半导体理论. 假设一个平行板电容器中间有某种均匀的导电材料，该材料中自由电子的电荷密度为 $-\rho$ （$\rho > 0$）为定值，每个电子受到的阻力与电子速度成正比，比例常数 $\alpha > 0$. 即

$$\mathbf{f} = -\alpha \mathbf{v} \tag{4}$$

当我们在电容器两板上施加电压时，内部会产生匀强电场，使电子受到电场力

$$\mathbf{F} = -e\mathbf{E} \tag{5}$$

电子在该电场力下加速（由于电子质量很小，加速过程很快，可以假设是一瞬间完成的），直到阻力等于电场力时加速停止，进行匀速运动. 于是有

$$-e\mathbf{E} - \alpha \mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v} = -\frac{e}{\alpha}\mathbf{E} \tag{6}$$

所以电阻内电流密度大小为

$$\mathbf{j} = -\rho \mathbf{v} = \frac{e\rho}{\alpha}\mathbf{E} \tag{7}$$

电流大小为

$$I = jS = \frac{\rho E e S}{\alpha} \tag{8}$$

电阻中的电场大小可以由两端电压表示为

$$E = \frac{U}{L} \tag{9}$$

代入上式得

$$U = I\frac{\alpha L}{\rho e S} \tag{10}$$

我们定义**电阻率**为

$$\varrho = \frac{\alpha}{\rho e} \tag{11}$$

然后再根据电阻率定义**电阻**为

$$R = \frac{\varrho L}{S} \tag{12}$$

可见它与长度 L 成正比，与横截面成反比．将式 11 代入式 12 再代入式 10，可得**欧姆定律**

$$U = IR \tag{13}$$

式 7 也可以使用电阻率记为

$$\mathbf{E} = \varrho \mathbf{j} \tag{14}$$

这相当于欧姆定律的微观形式．

电感

预备知识　安培环路定理[1179]，法拉第电磁感应定律[1191]

电感器（inductor）是电路中的一个元件，当通过它的电流变化时，它的两端会产生反向电动势从而抵抗电流的改变．我们用物理量**电感**来描述电感器的这种性质的强弱，记为 L，定义如下

$$U = -L\frac{\mathrm{d}I}{\mathrm{d}t} \tag{1}$$

也就是说反向电动势与电流的时间导数成正比，比例系数就是电感．





图 1: 左：各种电感器，右：在电路图中的符号

简单的电感模型

电感器的基本原理是利用线圈的自感，即当线圈中有电流通过时，会在它的内部及周围产生与电流大小成正比的磁场．当电流发生改变时，这个磁场也随之改变，而根据法拉第电磁感应定律[1191]，变化的磁场又会反过来作用在线圈上产生电动势．

由例 2 [1179]，单位长度线圈数为 n，长度为 l，忽略边缘效应的螺线管，内部磁场强度为

$$Bl = \mu_0 nlI \Rightarrow B = \mu_0 nI \tag{2}$$

磁通量为

$$\Phi = BS = \mu_0 nSI \tag{3}$$

感生电动势为

$$U = -nL\frac{\mathrm{d}\Phi}{\mathrm{d}t} = -\mu_0 n^2 lS\frac{\mathrm{d}I}{\mathrm{d}t} \tag{4}$$

令电感为

$$L = \mu_0 n^2 lS \tag{5}$$

有

$$U = -L\frac{\mathrm{d}I}{\mathrm{d}t} \tag{6}$$

电感储存的能量

与电容类似，电感器中也可以储存能量．

$$E_B = \frac{1}{2}LI^2 \tag{7}$$

推导

$$E_B = -\int_0^{I_m} UI\,\mathrm{d}t = \int_0^{I_m} L\frac{\mathrm{d}I}{\mathrm{d}t}I\,\mathrm{d}t = L\int_0^{I_m} I\,\mathrm{d}I = \frac{1}{2}LI^2 \tag{8}$$

电场的高斯定律

预备知识 电场[1124]，散度散度定理[311]

[8]**电场的高斯定律（Gauss's law for electric field）**是麦克斯韦方程组[1223]中四条方程中的一条，是经典电磁学的基本假设之一，无法推导．用电场线来形象地描述该定理，就是正电荷会产生电场线，负电荷会消除电场线．产生和消除的电场线和电荷量成正比．

在静电学条件下，即空间中净电荷不随时间变化，且电流恒定不变时，高斯定律和库伦定律[1123]是等价的．然而在非静电学条件下，高斯定律依然成立而库伦定律却不成立[9]．所以高斯定律是更普适的定理，库仑定律可以看作是高斯定律在静电学条件下的特殊形式．

积分形式

在空间中任意选取一个闭合曲面 $\mathcal{S}$，电场在这个曲面上从内向外的通量等于被曲面包围的总电荷量除以真空中的介电常数．

$$\oint_{\mathcal{S}} \mathbf{E}(\mathbf{r})\cdot\mathrm{d}\mathbf{s} = \frac{1}{\epsilon_0}\int_{\mathcal{S}} \rho(\mathbf{r})\,\mathrm{d}V \tag{1}$$

作为另一种形象的理解，我们可以想象正电荷会以固定速率向周围释放或吸收一种不可压缩的流体，负电荷则吸收这种流体．释放和吸收的速率和电荷绝对值成正比，电场可以看作该流体的速度场．所以无论取什么形状的闭合曲面，单位时间流经曲面的总流量都只取决于曲面内部的电荷总量．

[8]本文参考 Wikipedia 相关页面．

[9]作为一个反例，我们在电动力学简介[15]中已经讨论过：若一个静止的电荷在原地快速抖动一下，电磁场会以光速把这个扰动向周围传播，所以远处的电荷并不能立即感受到电场的变化．这显然超出了库伦定律描述的范围．

微分形式

空间任意一点的电场散度等于电荷密度除以真空介电常数.

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \tag{2}$$

根据散度定理[311]，该式和式 1 是完全等价的.

电通量

如果把电场想像成是某种不可压缩液体的速度场（液体流动的速度矢量在空间上的分布），那么对于电场中假想的一个不闭合的空间曲面，电通量就相当于单位时间流过该曲面的体积. 类比流量和通量中得到的公式

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \oint \mathbf{v} \cdot \mathrm{d}\mathbf{s} \tag{3}$$

电通量为（符号为 Φ，国际单位 $\mathrm{Nm^2/C}$）

$$\Phi = \oint \mathbf{E} \cdot \mathrm{d}\mathbf{s} \tag{4}$$

即电通量就是电场在所选曲面上的通量.

高斯定律的积分形式

高斯定律的积分形式说的是，选取任意闭合曲面为高斯面（由内向外为正方向），高斯面上的电通量等于高斯面内的总电荷除以常数. 即

$$\oint \mathbf{E} \cdot \mathrm{d}\mathbf{s} = \frac{1}{\epsilon_0} \int \rho \, \mathrm{d}V \tag{5}$$

其中常数 ϵ_0 为**真空中的电介质常数（vacuum permitivity）**，又叫**真空中的电容率**（原因见“电容[1145]”）. 为了便于理解和记忆，可以把电场想像成某种流体的场，这种流体从正电荷流出，流入负电荷，流出和流入的速率（单位时间的体积）和电荷的大小成正比，比例系数为 $1/\epsilon_0$. 从数学上来讲，有电荷的地方电场就有**源**，ρ/ϵ_0 就叫做电场的**源密度**.

例 1　点电荷和球壳的电场

令点电荷的电荷量为 Q，以点电荷为圆心选取半径为 r 的球形高斯面. 由

系统的对称性可知，高斯面上的电场大小必处处相等．所以高斯面的电场通量（式 5 左边）等于

$$\oint \mathbf{E}\cdot \mathrm{d}\mathbf{s} = E\cdot 4\pi r^2 \tag{6}$$

式 5 的右边等于 Q/ϵ_0，无需积分．两边除以 $4\pi r^2$ 得高斯面上电场的大小为

$$E = \frac{Q}{4\pi\epsilon_0 r^2} \tag{7}$$

与库仑定律的结论一致．这说明高斯定律可以推导出库伦定律．

再来看均匀带电球壳产生的电场，若在其外部作球形的高斯面，得到的结论与点电荷一致，但若在球内部作球形高斯面，由于高斯面内没有电荷，根据对称性可知球壳内部电场为零．

例 2　带电平面和平行板电容器的电场

空间中有一块无限大的厚度不计的均匀带电平面，面电荷密度为 σ，求空间中的电场．

如图，我们令平面位于取一个柱形的高斯面，上下底面离平面的距离都为 z．（未完成）

例 3　均匀带电球的电场

令球的半径为 R，电荷密度为 ρ，令无穷远处为零势点，求均匀带电球内外的电场分布．

带点球的总电荷量为 $Q = 4\pi R^3\rho/3$，用与例 1 类似的方法可知球外电场与点电荷的电场相同．再来看球内的电场，以半径为 r（$r < R$）作高斯面，高斯面内的电荷为 $q = 4\pi r^3\rho/3$，高斯面上的电场为

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0 r^2}\hat{\mathbf{r}} = \frac{\rho}{3\epsilon_0}\mathbf{r} \tag{8}$$

高斯定律的微分形式

根据矢量分析中的散度定理[311]，若对任意闭合曲面，一个标量场在曲面内的体积分等于一个矢量场在曲面上的面积分，该标量场就是该矢量场的散度．把该结论用于上式，可得电场的散度为

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \rho/\epsilon_0 \tag{9}$$

[ht]

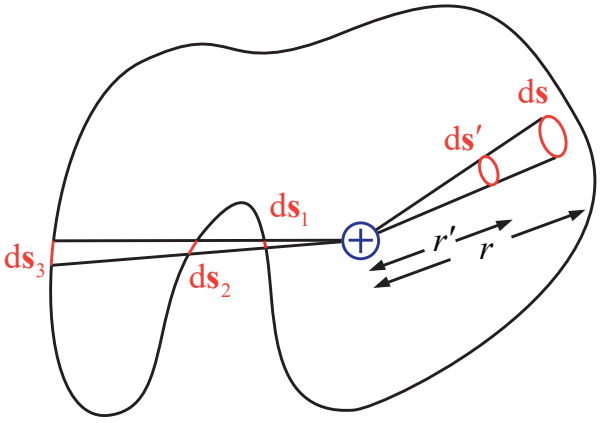


图 1: 图 1

[ht]



图 2: 图 2

未完成：直接证明微分形式

由库仑定律推导高斯定律（静电学）

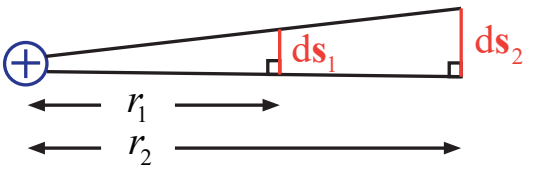
由例 1 可知静电学情况下，由高斯定律可以轻易得出库伦定律．现在我们从库仑定律证明高斯定律．这里先给出一个符合几何直觉的幼稚证明，便于理解．严谨的证明见“电场的高斯定律证明[1166]”.

首先只考虑高斯面内一个点电荷 Q 对一个面元 ds 的通量．若元 ds 与相对于点电荷的位矢 $\mathbf{r}$ 的夹角为 θ，通过该面元的电通量为

$$\mathrm{d}\Phi = \mathbf{E}\cdot\mathrm{d}\mathbf{s} = \frac{kQ}{r^2}\,\mathrm{d}s\cdot\cos\theta \tag{10}$$

下面证明保持以 ds 为底面，Q 为顶点的圆锥的任意一个截面的电通量都是一样的（即式 4 节 中 ds 和 ds′ 的电通量一样）．要证明这点，首先考虑两个与 Q 距离相同，但角度不同的截面（见图 4 节 ），其中 ds 垂直于 $\mathbf{r}$．它们的电通量分别为

$$\mathrm{d}\Phi_a = E\cdot\mathrm{d}s_a\cdot\cos\theta \tag{11}$$



[ht]

图 3: 图 3

$$\mathrm{d}\Phi_b = E \cdot \mathrm{d}s_b \tag{12}$$

然而显然有 $\mathrm{d}s_b = \cos\theta\ \mathrm{d}s_a$（注意圆锥非常细长的时候，母线近似都平行），所以有

$$\mathrm{d}\Phi_a = \mathrm{d}\Phi_b \tag{13}$$

即，原地改变截面的角度，电通量不变．下面再考虑角度垂直但与电荷距离不同的情况（式 4 节 中的 $\mathrm{d}s_1$ 与 $\mathrm{d}s_2$ ）它们的电通量分别为

$$\mathrm{d}\Phi_1 = \frac{kQ}{r_1^2}\,\mathrm{d}s_1 \tag{14}$$

$$\mathrm{d}\Phi_2 = \frac{kQ}{r_2^2}\,\mathrm{d}s_2 \tag{15}$$

然而根据几何关系，有

$$\frac{\mathrm{d}s_1}{r_1^2} = \frac{\mathrm{d}s_2}{r_2^2} \tag{16}$$

所以仍然有．

$$\mathrm{d}\Phi_1 = \mathrm{d}\Phi_2 \tag{17}$$

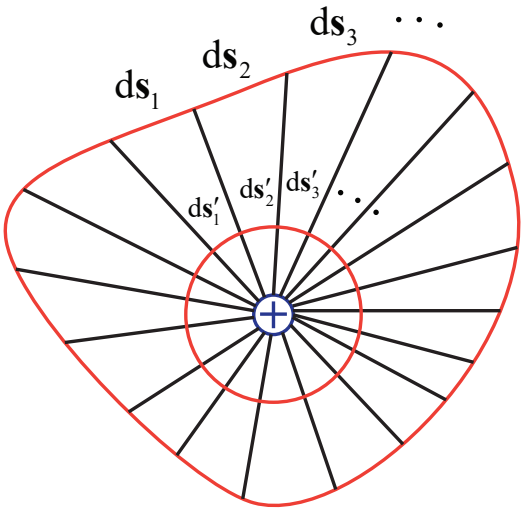
若把整个闭合曲面划分成无数个小面元（如式 4 节 ），每个小面元 $\mathrm{d}\mathbf{s}_i$ 根据上面推理，都等效为半径 R 上的一块垂直面元 $\mathrm{d}\mathbf{s}_i$．这样，这些等效面元可以重新组成一个半径为 R 的球体．而球体的通量为

$$\Phi = \frac{kQ}{R^2} \cdot 4\pi R^2 = 4\pi kQ \tag{18}$$

令真空中的介电常数 $\epsilon_0 = 1/(4\pi k)$，就得到高斯定律

$$\Phi = Q/\epsilon_0 = \frac{1}{\epsilon_0}\int \rho\,\mathrm{d}V \tag{19}$$

若高斯面内有许多点电荷或电荷分布，则根据电场叠加原理，对每个点电荷应用上述结论再相加即可．另外，若出现高斯面重叠的情况（如图 1 中的 $\mathrm{d}\mathbf{s}_1$，



[ht]

图 4: 图 4

$\mathbf{ds}_2$，$\mathbf{ds}_3$ ），高斯定律仍然成立．这是因为 $\mathbf{ds}_1$ 和 $\mathbf{ds}_2$ 的电通量大小相等，然而方向相反，对总电通量的贡献抵消，只有剩下的 $\mathbf{ds}_3$ 才对总电通量有贡献，这和不重叠的情况一样．

最后来证明高斯面外部的点电荷对高斯面的总电通量没有贡献．因为从高斯面外的点电荷（电场源）“流入”高斯面的电通量全部“流出”，这是由于从电荷出发的所有圆锥要么不被高斯面截断，要么被高斯面截断两次，一次进一次出，且两个界面上的电通量大小相同，总电通量贡献为零．

电场的高斯定律证明

预备知识 电场的高斯定律[1161]，球坐标系中的梯度散度[742]

以下我们用使用库仑定律和散度定理[311] 严谨地证明电场的高斯定律[1161]．

我们先看一个位于原点，电荷为 q 的点电荷产生的电场（式 4 [1124]）

$$\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\epsilon_0}\frac{\hat{\mathbf{r}}}{r^2} \tag{1}$$

要计算某点的散度，最方便的做法是使用球坐标公式（式 2 [742]）得

$$\nabla \cdot \frac{\hat{\mathbf{r}}}{r^2} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{1}{r^2}\right) = 0 \tag{2}$$

注意由于式 1 在原点处无定义，也不存在偏导数，该结论不适用于 $\mathbf{r}=\mathbf{0}$.

由于散度算符是线性的，即使空间中有许多点电荷，第 i 个产生的电场为 $\mathbf{E}_i(\mathbf{r})$，空间中任何点（除了点电荷的位置）的电场散度都为零.

$$\boldsymbol{\nabla}\cdot\left(\sum_i \mathbf{E}_i\right)=\sum_i \boldsymbol{\nabla}\cdot\mathbf{E}_i=0 \tag{3}$$

我们还是假设只有一个点电荷位 q 于坐标原点，现在我们以原点为球心做一个半径为任意 $R>0$ 的球面 $\mathcal{S}$，并计算电场在球面上的通量为

$$\Phi=\oint_{\mathcal{S}}\mathbf{E}\cdot\mathrm{d}\mathbf{s}=\frac{q}{4\pi\epsilon_0 R^2}\cdot 4\pi R^2=\frac{q}{\epsilon_0} \tag{4}$$

然而根据散度定理（式 1 [311]），如果 $\boldsymbol{\nabla}\cdot\mathbf{E}$ 在球内处处为零，应该有 $\Phi=0$ 才对. 所以问题应该出在原点，由于散度定理要求矢量场在闭合曲面内部处处可偏导，所以式 4 的结果并不严格适用于散度定理[10].

绕开奇点

我们可以以另一种方式避开 $\mathbf{r}=0$ 处的奇点使用散度定理. 结合式 3 和式 4，可以用散度定理证明：如果在球面 $\mathcal{S}$ 外部再任取一个闭合曲面 $\mathcal{S}'$（正方向也向外），那么电场在 $\mathcal{S}'$ 上的通量应该也是相同的：

我们把这两个曲面共同看成是它们之间的球壳形体积 $\bar{\mathcal{V}}$ 的内外表面（但 $\mathcal{S}$ 的正方向需要改变）. 而 $\bar{\mathcal{V}}$ 中散度处处有定义且等于零，所以由散度定理得内外表面的电场总通量也为零

$$\Phi=-\oint_{\mathcal{S}}\mathbf{E}(\mathbf{r})\cdot\mathrm{d}\mathbf{s}+\oint_{\mathcal{S}'}\mathbf{E}(\mathbf{r})\cdot\mathrm{d}\mathbf{s}=\int_{\bar{\mathcal{V}}}\boldsymbol{\nabla}\cdot\mathbf{E}(\mathbf{r})\,\mathrm{d}V=0 \tag{5}$$

移项可得两个面积分相等. 证毕.

所以，对于任意两个包含同一个点电荷 q 的曲面（正方向都向外），电场在它们上的通量都等于 q/ϵ_0，而 q 也不一定需要在原点. 再根据电场的叠加原理和点乘的分配律（式 3 [383]），如果曲面内有多个点电荷 $q_1,\ldots,q_N$，那么它们在曲面上的通量等于每个点电荷产生的通量之和，即

$$\oint_{\mathcal{S}}\mathbf{E}(\mathbf{r})\cdot\mathrm{d}\mathbf{s}=\frac{1}{\epsilon_0}\sum_i q_i \tag{6}$$

[10]但有办法可以弥补，见“点电荷电场的散度”

这就证明了高斯定律的积分形式. 若曲面内的电荷是连续分布的，我们只需要把求和变为对 $\mathcal{S}$ 内部的电荷密度分布 $\rho(\mathbf{r})$ 的体积分即可

$$\oint_{\mathcal{S}} \mathbf{E}(\mathbf{r}) \cdot \mathrm{d}\mathbf{s} = \frac{1}{\epsilon_0} \int \rho(\mathbf{r}) \, \mathrm{d}V \tag{7}$$

注意如果电荷密度 $\rho(\mathbf{r})$ 处处为有限值，那么电场的散度也将处处有定义. 对比散度定理可得空间中任意一点 $\mathbf{r}$ 处电场的散度与电荷密度成正比.

$$\boldsymbol{\nabla} \cdot E = \frac{\rho}{\epsilon_0} \tag{8}$$

这就证明了电场高斯定律的微分形式.

导体的静电平衡

- 本词条处于草稿阶段.

在一类问题中，我们知道各个导体中带有的净电荷或者电势，却不知道具体的分布情况，应该如何求解电荷分布呢？

泊松方程

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \rho/\epsilon_0 \tag{1}$$

$$\mathbf{E} = -\boldsymbol{\nabla} V \tag{2}$$

代入得泊松方程[748]

$$\nabla^2 V = -\rho/\epsilon_0 \tag{3}$$

由于导体是等势体，而导体外部的空间没有净电荷，所以这相当于对导体外部空间求解边界值为常数的拉普拉斯方程

$$\nabla^2 V = 0 \tag{4}$$

若已知导体的电势，那么边界条件已知. 若已知导体中的净电荷，由于净电荷只集中于物体表面，所以在表面处有

$$-\frac{\partial V}{\partial n} = E = \sigma/\epsilon_0 \tag{5}$$

再进行积分即可得出金属的净电荷．如果只有一块金属，可以先假设金属的 V 为任意常数，解出以后再乘以一个常数使净电荷符合条件即可．

未完成：例题：金属契子

未完成：数值解泊松方程：relaxation method

磁场的高斯定律

- 本词条处于草稿阶段．

预备知识 电场的高斯定律[1161]

[11]**磁场的高斯定律（Gauss's law for magnetism）**是麦克斯韦方程组[1223]中的一条方程．

磁场的高斯定律：对于任意磁场 $\mathbf{B}(\mathbf{r})$ 和任意闭合曲面，曲面上的磁通量[1185]为零．

$$\oint \mathbf{B}(\mathbf{r}) \cdot \mathrm{d}\mathbf{s} = 0 \tag{1}$$

电场的能量

预备知识 电容[1145]

[12]首先我们来回顾一下真空中平行板电容器（例 2 [1145]）的能量

$$E_p = \frac{1}{2}CV^2 = \frac{1}{2}\epsilon_0 \frac{S}{d}(Ed)^2 = \frac{1}{2}\epsilon_0 \tau E^2 \tag{1}$$

其中 $\tau = Sd$ 为平行板间长方体的体积．这条公式容易让我们联想到电势能储存在电场之中．在电动力学中，这种理解是正确的，我们不妨把电场含有的能

[11]本文参考 Wikipedia 相关页面．

[12]本文参考 [12]．

量称为**电场能**．通过上式，我们假设空间中任意一点的电场能密度为 $\epsilon_0 E^2/2$，则总电场能可以通过体积分得到

$$E_p = \frac{1}{2}\epsilon_0 \int \mathbf{E}(\mathbf{r})^2 \,\mathrm{d}^3 r \tag{2}$$

可以证明该式定义的电场能与式 26 [1130] 定义的电势能是完全等效的

$$E_p = \frac{1}{2} \int V(\mathbf{r})\rho(\mathbf{r}) \,\mathrm{d}^3 r \tag{3}$$

证明

预备知识 分部积分的高维拓展[306]

电场的高斯定律[1161]

$$\rho = \epsilon_0 \boldsymbol{\nabla} \cdot \mathbf{E} \tag{4}$$

代入式 3 得

$$E_p = \frac{\epsilon_0}{2} \int V(\boldsymbol{\nabla} \cdot \mathbf{E}) \,\mathrm{d}^3 r \tag{5}$$

由 3 维分部积分式 4 [306] 得

$$E_p = \frac{\epsilon_0}{2} \oint \mathbf{E} V \,\mathrm{d}\mathbf{s} - \frac{\epsilon_0}{2} \int \mathbf{E} \cdot \boldsymbol{\nabla} V \,\mathrm{d}V \tag{6}$$

当我们把积分区域拓展到无限大时，右边第一个面积分趋于 0．右边第二个积分可以利用电场与电势的关系 $\boldsymbol{\nabla} V = -\mathbf{E}$（式 10 [1130]）．带入可得式 2．证毕．

LC 振荡电路

预备知识 简谐振子[913]，电容[1145]，电感[1159]

电路中电压和电流的周期性变化称为**电磁振荡（electromagnetic oscillation）**，电磁振荡与机械振动有类似的运动形式．产生电磁振荡的电路称为**振荡电路（resonant circuit）**．最简单的振荡电路是由一个电容器与一个自感线圈串联而成的，称为 **LC 电路（LC circuit）**．

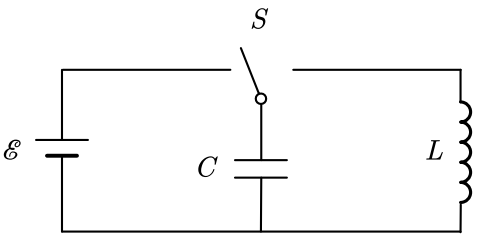


图 1: LC 振荡电路

如图 1 所示的电路，先使电源给电容器充电，然后将电键接通 LC 回路，在振荡电路刚被接通的瞬间，电容器两极板上的电荷最多，板间的电场也最强，电场的能量全部集中在电容器的两极板间.

当电容器放电时，因自感的存在，电路中的电流将逐渐增大到最大值，两极板上的电荷也相应地逐渐减小到零. 在此过程中，电流在自感线圈中激起磁场，到放电结束时，电容器两极板间的电场能量全部转化成线圈中的磁场能量. 在电容器放电完毕时，电路中的电流达到最大值. 这时，就要对电容器作反方向的充电. 由于线圈的自感作用，随着电流的逐渐减弱到零，电容器两极板上的电荷又相应地逐渐增加到最大值. 同时，磁场能量又全部转化成电场能量.

然后，电容器又通过线圈放电，电路中的电流逐渐增大，不过这时电流的方向与前放电时相反，电场能量又转化成磁场能量. 此后，电容器又被充电，回复到原状态，完成了一个完全的振荡过程.

由上述可知，在 LC 电路中，电荷和电流都随时间做周期性的变化，相应地电容器中的电场强度和线圈中的磁感应强度以及电场能量和磁场能量也都随时间作周期性变化，而且不断地相互转化着. 如果电路中没有任何能量损耗（如电阻的焦耳热、电磁辐射等），那么这种变化将在电路中一直持续下去，这种电磁振荡称为**无阻尼自由振荡**.

下面我们定量地研究无阻尼自由振荡，找出电容器极板上的电荷和电路中的电流随时间变化的规律.

设在某一时刻，电容器极板上的电荷量为 q, 电路中的电流为 i，并取 LC 回路的顺时针方向为电流的正方向. 线圈两端的电势差应和电容器两极板之间的电势差相等，即

$$-L\frac{\mathrm{d}i}{\mathrm{d}t}=\frac{q}{C} \tag{1}$$

由电流的定义式 $i=\dfrac{\mathrm{d}q}{\mathrm{d}t}$，我们有：

$$\frac{\mathrm{d}^2q}{\mathrm{d}t^2}=-\frac{1}{LC}q \tag{2}$$

令 $\omega^2=\dfrac{1}{LC}$，得

$$\frac{\mathrm{d}^2q}{\mathrm{d}t^2}=-\omega^2q \tag{3}$$

根据二阶线性常系数齐次微分方程[215] 的第 3 种情况，其通解为

$$q=Q_0\cos\left(\omega t+\phi_0\right) \tag{4}$$

式中 Q_0 为极板上电荷量的最大值，称为电荷量振幅．ϕ_0 是振荡的初相位，Q_0 和 ϕ_0 的数值由初始条件决定．ω 是振荡的角频率．无阻尼自由振荡的频率和周期分别为

$$\nu=\frac{\omega}{2\pi}=\frac{1}{2\pi\sqrt{LC}},\quad T=2\pi\sqrt{LC} \tag{5}$$

相比电荷量的变化情况，我们更关注的是电流的变化情况．由电流的定义式，对式 4 两边求导数，得到

$$i=\frac{\mathrm{d}q}{\mathrm{d}t}=-\omega Q_0\sin\left(\omega t+\phi_0\right) \tag{6}$$

即为电路中任一时刻的电流．

令 $\omega Q_0=I_0$ 表示电流的最大值，称为**电流振幅**，则上式为

$$i=-I_0\sin\left(\omega t+\phi_0\right)=I_0\cos\left(\omega t+\phi_0+\frac{\pi}{2}\right) \tag{7}$$

式 4 和式 7 表明，在 LC 振荡电路中，电荷和电流都作谐振动，且振荡频率相同，电流的相位比电荷的相位超前 $\pi/2$．如图 2 所示．

现在让我们来考虑 LC 振荡电路中的能量．在任一时刻 t，电容器极板上的电荷量为 q，相应的电场能量为

$$W_\mathrm{e}=\frac{1}{2}\frac{q^2}{C}=\frac{Q_0^2}{2C}\cos^2\left(\omega t+\phi_0\right) \tag{8}$$

设此时的电流为 i，那么线圈内的磁场能量为

$$W_\mathrm{m}=\frac{1}{2}Li^2=\frac{L\omega^2Q_0^2}{2}\sin^2\left(\omega t+\phi_0\right) \tag{9}$$

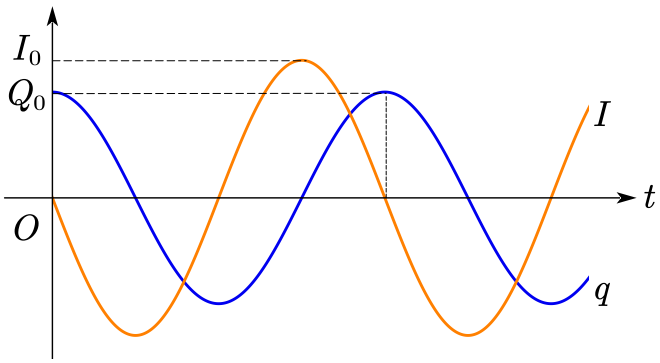


图 2: 电荷与电流的振荡

把上两式相加，并由 $\omega^2 = \dfrac{1}{LC}$，可得总能量：

$$W = W_{\mathrm{e}} + W_{\mathrm{m}} = \frac{Q_0^2}{2C}\cos^2(\omega t + \phi_0) + \frac{Q_0^2}{2C}\sin^2(\omega t + \phi_0) = \frac{Q_0^2}{2C} \tag{10}$$

式 10 说明，在无阻尼自由振荡电路中，尽管电能和磁能都随时间而变化，但总的电磁能量却保持不变.

从上面的分析可以知道，电磁振荡中的电荷量及电流对应机械振动中的位移和速度，自感对应于惯性，起着电流惯性作用磁场能藉对应于动能，电场能量对应于势能. 在力电类比中，我们将对机械振动和电磁振荡之间的等效关系做一个完整的总结.

力电振动类比

电磁振荡和机械振动的规律非常相似，所以运用力电类比就可以把电磁振荡和机械振动对应起来，只要知道一种振动的解，就可以用类比方法得到另一种振动的解. 虽然机械振动比较直观，但由于电学的迅速发展，人们对交变电路规律的熟悉程度已经超过机械振动. 因此，在工程上，常常把复杂的机械振动问题用力电类比方法化成交变电路问题，然后通过计算或实验测定，找出它们的解机械振动和电磁振荡对应的物理量列在下表中.

LC 受迫振荡电路

表 1: 机械振动和电磁振荡对应的物理量

机械振动		电磁振荡 (串联电路)	
位移	x	电荷	q
速度	v	电流	i
质量	m	电感	L
劲度系数	k	电容的倒数	$\dfrac{1}{C}$
阻力系数	γ	电阻	R
驱动力	F	电动势	$\mathscr{E}$
弹性势能	$\dfrac{1}{2}kx^2$	电场能量	$\dfrac{1}{2}\dfrac{q^2}{C}$
动能	$\dfrac{1}{2}mv^2$	磁场能量	$\dfrac{1}{2}Li^2$

预备知识 LC 振荡电路[1170]

当电路中有电阻存在时，由于能量的损耗，电荷和电流的振幅将逐渐减小．如果在电路中加入一个电动势作周期性变化的电源，如图 1 所示，可以

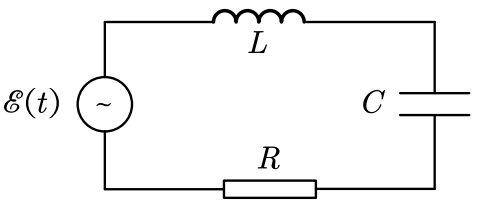


图 1: 受迫振荡电路

连续不断地供给能量，即可使电流振幅保持不变，这种在外加周期性电动势持续作用下产生的振荡，称为**受迫振荡（forced oscillation）**．设电源的电动势为 $\mathscr{E}=\mathscr{E}_0\cos\omega_{\mathrm{d}}t$，则受迫振荡的微分方程可写成

$$L\frac{\mathrm{d}^2q}{\mathrm{d}t^2}+R\frac{\mathrm{d}q}{\mathrm{d}t}+\frac{q}{C}=\mathscr{E}_0\cos\omega_{\mathrm{d}}t \tag{1}$$

在稳定状态下其解为

$$q=Q_0\cos\left(\omega_d t+\phi\right) \tag{2}$$

通常我们感兴趣的不是电荷，而是电流的振荡，由上式得

$$\begin{aligned} i = \frac{\mathrm{d}q}{\mathrm{d}t} &= -\omega Q_0 \sin\left(\omega_\mathrm{d} t + \phi\right) \\ &= \omega_\mathrm{d} Q_0 \cos\left(\omega_\mathrm{d} t + \phi + \frac{\pi}{2}\right) \\ &= I_0 \cos\left(\omega_\mathrm{d} t + \phi'\right) \end{aligned} \tag{3}$$

式中

$$I_0 = \frac{\mathscr{E}_0}{\sqrt{R^2 + \left(\omega_\mathrm{d} L - \frac{1}{\omega_\mathrm{d} C}\right)^2}} \tag{4}$$

$$\tan\phi' = \frac{\frac{1}{\omega_\mathrm{d} C} - \omega_\mathrm{d} L}{R} \tag{5}$$

可以看到，电流 i 的振荡角频率与电动势的角频率相同，但两者的相位并不相同.

由式 4 不难看出，当电路满足条件 $\omega_\mathrm{d} L = \dfrac{1}{\omega_\mathrm{d} C}$ 时，电流将有最大的振幅. 从这个条件将 w_d 解出来，得

$$\omega_\mathrm{d} = \sqrt{\frac{1}{LC}} \tag{6}$$

这就是说，当外加电动势的频率和自由振荡的的频率相等时，电流的振幅为最大，其值等于 $\dfrac{\mathscr{E}_0}{R}$，这时，电流和外加电动势之间的相位差 $\phi = 0$. 这种在周期性电动势作用下，电流振幅达到最大值的现象称为**电共振（electrical resonance）**. 收音机中的调谐，就是调节电容器的电容使电路与其某一种频率的无线电信号发生共振，以选取电台.

比奥萨伐尔定律

预备知识 电流产生磁场[??]，线积分[295]

闭合回路的磁场

若已知空间中的电流分布，**比奥萨伐尔定律（Biot-Savart law）**给出了任意一点的磁场分布. 我们先讨论一种角为简单的情况，即一根粗细可忽略的闭

合导线中的恒定电流产生的磁场[13]．如果有多个这样的导线，可以分别计算它们的磁场并叠加即可．假设导线中电流为 I，空间中任意位置 $\mathbf{r}$ 的磁场可以用线积分表示为

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \oint \frac{I \, \mathrm{d}\mathbf{r}' \times \hat{\mathbf{R}}}{R^2} \tag{1}$$

其中积分路径是导线所在的环路，延某个指定的正方向进行．如果电流也是沿正方向，那么 I 大于零，否则小于零．$\mathbf{r}'$ 是导线上某点的位置，矢量 $\mathrm{d}\mathbf{r}'$ 代表导线上的一小段的长度和方向．$\mathbf{R} = \mathbf{r} - \mathbf{r}'$，$\hat{\mathbf{R}}$ 表示 $\mathbf{R}$ 的单位矢量，$R = |\mathbf{R}|$ 是模长．

使用这个公式需要注意两点．第一，电流及其分布不随时间变化．这点可以从“不存在瞬时作用”理解，假设某时刻电流突然从 0 变为某个值，由于电磁场传播需要一定时间，环路上不可能瞬间出现磁场．第二，空间中不能有变化的电场，因为变化的电场也会产生磁场．

电流密度形式

更一般地，若电流分布以电流密度 $\mathbf{j}(\mathbf{r})$ 来描述，那么

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \oint \frac{\mathbf{j}(\mathbf{r}') \times \hat{\mathbf{R}}}{R^2} \, \mathrm{d}V \tag{2}$$

详细说明

比奥萨伐尔定律是电磁学的基本假设之一，所以以下并不是推导，而是解释公式的意义．

一个粗细忽略不计的电流回路中有电流 I，如何确定该回路在空间中任意一点所产生的磁场呢？由于磁场与电场一样可以叠加，我们可以把回路划分成极小的线段，分别计算每个小线段在某点 $\mathbf{r}$ 产生的磁场，然后求和．当这些小线段的长度趋近于零，求和就变成了线积分[295]．

那如何计算一小段长度为 $\mathrm{d}l$ 的电流（电流元）产生的磁场呢？如图（$\mathrm{d}l$ 存在正方向，与 $\mathbf{R}$ 的夹角为 θ），为了表示电流的方向，我们先把 $\mathrm{d}l$ 变为矢量 $\mathrm{d}\mathbf{l}$，方向为电流的正方向（当 $I > 0$ 时，电流方向与 $\mathrm{d}\mathbf{l}$ 相同，$I < 0$ 时相反）．

[13] 在电动力学[15] 中，我们提到不存在瞬间作用，电磁场的传播需要时间．而且变化的磁场也会产生变化的电场，这将使计算变得十分复杂，我们将在“推迟势[1227]”中介绍．

现在我们要求空间中任意一点 $\mathbf{r}$（把这点叫做场点）的磁场，设电流元的位置为 $\mathbf{r}$（把这点叫做原点），且设

$$\mathbf{R}=\mathbf{r}-\mathbf{r}' \tag{3}$$

首先，磁场大小正比于 I，这是合理的，因为如果把两小段电流元重叠放在一起，那么根据叠加原理，任何地方的磁场都会增加一倍．其次，与点电荷的电场（库仑定律）类似，场强与距离的平方成反比．最后，由于电流有特定的方向，磁场不再具有球对称性，而是以导线为轴的中心对称．磁场大小正比于 $|\mathrm{d}\mathbf{l}\times\mathbf{R}|=R\sin\theta\,\mathrm{d}l$（磁场方向垂直于 $\mathrm{d}\mathbf{l}$ 和 $\mathbf{R}$ 所在平面，符合右手定则，见矢量的叉乘[387]）．这说明，在 I 和 R 不变时，垂直于电流元的点（$\theta=\pi/2$）具有最大场强，与其共线的点（$\theta=0$）场强为零．

要注意的是，虽然我们是对单独一个 $I\,\mathrm{d}\mathbf{l}$ 分析，但稳定的电流元在现实中并不存在．因为线电流必须组成环路，否则在两个端点处就会分别积累大量的异号电荷（见电流的连续性），从而产生变化的电场及磁场，而在静态电磁场问题中，我们要求净电荷和电流的分布不随时间改变．所以在利用比奥萨伐尔公式时，必须要以对整个闭合回路积分（无穷长直导线或者螺线管等理想化问题除外）若给所有的小电流源编号为 $I\,\mathrm{d}\mathbf{l}_i$，令第 i 个电流源的起点为 $\mathbf{r}'_i$，$\mathbf{R}_i=\mathbf{r}-\mathbf{r}'_i$．把所有电场矢量相加，变为

$$\mathbf{B}(\mathbf{r})=\frac{\mu_0}{4\pi}\sum_i\frac{I\,\mathrm{d}\mathbf{l}_i\times\mathbf{R}_i}{R_i^2} \tag{4}$$

当电流元无穷短，数量无穷多的时候，上式写为积分的形式，且由于第 i 个电流元的终点就是第 $i+1$ 个电流元的起点，$\mathrm{d}\mathbf{l}_i=\mathbf{r}'_{i+1}-\mathbf{r}'_i=\mathrm{d}\mathbf{r}'$，矢量积分写为

$$\mathbf{B}(\mathbf{r})=\frac{\mu_0}{4\pi}\oint\frac{I\,\mathrm{d}\mathbf{r}'\times\hat{\mathbf{R}}}{R^2} \tag{5}$$

注意 $\mathbf{R}$ 是场点 $\mathbf{r}$ 和源点 $\mathbf{r}'$ 的函数，积分时把 $\mathbf{r}$ 视为常数而对 $\mathbf{r}'$ 积分．为了使公式更明确，在难度较大的电磁学教材中把上式直接记为

$$\mathbf{B}(\mathbf{r})=\frac{\mu_0}{4\pi}\oint\frac{I\,\mathrm{d}\mathbf{r}'\times(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3} \tag{6}$$

例 1　无限长直导线的电场

如图 1，令导线与 x 轴重合，并使原点到场点的距离最近，有 $x=r\tan\theta$，微

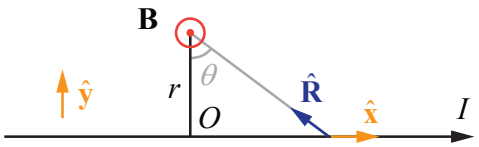


图 1: 无限长直导线的电场

分得 $|\mathrm{d}\mathbf{r}'| = \mathrm{d}x = r\sec^2\theta\,\mathrm{d}\theta$，另有 $\hat{\mathbf{x}}\times\hat{\mathbf{R}} = \hat{\mathbf{z}}\sin(\theta+\pi/2) = \hat{\mathbf{z}}\cos\theta$，$R = r/\cos\theta$. 代入式 5 得

$$\mathbf{B} = \frac{\mu_0}{4\pi} I\hat{\mathbf{z}} \int_{-\pi/2}^{\pi/2} \frac{(r\sec^2\theta\,\mathrm{d}\theta)\cos\theta}{r^2/\cos\theta^2} = \frac{\mu_0}{4\pi}\frac{I}{r}\hat{\mathbf{z}}\int_{-\pi/2}^{\pi/2}\cos\theta\,\mathrm{d}\theta = \frac{\mu_0}{2\pi}\frac{I}{r}\hat{\mathbf{z}} \tag{7}$$

可见磁场大小与距离 r 成反比，总是垂直于导线，且方向符合右手定则.

矢量积分的计算方法

比奥萨伐尔定律的积分中含有矢量微元的叉乘，看起来和普通的矢量积分不同，但是在常见的简单问题中，可以从几何理解上直接转换为标量的积分（见无限长直导线的磁场和环形电流轴线的磁场）. 如果是更一般的问题，则可以把叉乘分解成 3 个分量，然后变为 6 个标量积分

$$\begin{aligned}\mathrm{d}\mathbf{r}'\times\mathbf{R} &= \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \mathrm{d}x' & \mathrm{d}y' & \mathrm{d}z' \\ x-x' & y-y' & z-z' \end{vmatrix} \\ &= \hat{\mathbf{x}}[(z-z')\,\mathrm{d}y' - (y-y')\,\mathrm{d}z'] + \hat{\mathbf{y}}[\dots] + \hat{\mathbf{z}}[\dots]\end{aligned} \tag{8}$$

$$\begin{aligned}\mathbf{B}(\mathbf{r}) &= \frac{\mu_0}{4\pi}\oint\frac{I\,\mathrm{d}\mathbf{r}'\times\hat{\mathbf{R}}}{R^2} = \frac{\mu_0 I}{4\pi}\oint\frac{\mathrm{d}\mathbf{r}'\times\mathbf{R}}{R^3} \\ &= \hat{\mathbf{x}}\frac{\mu_0 I}{4\pi}\left[\oint\frac{1}{R^3}(z-z')\,\mathrm{d}y' - \oint\frac{1}{R^3}(y-y')\,\mathrm{d}z'\right] + \hat{\mathbf{y}}[\dots] + \hat{\mathbf{z}}[\dots]\end{aligned} \tag{9}$$

电流密度的形式

假设电流的空间分布是连续变化的而不能看成一条截面不计曲线，我们需要用电流密度 $\mathbf{J}$ 来表示空间的电流分布. 现在考虑一个粗细不能忽略的环路，$\mathbf{r}'$ 处的截面积为 A（取截面时应垂直于电流），通过截面的电流为 $I = A\mathbf{J}$，所

以电流元变为 $I\,\mathrm{d}\mathbf{l} = \mathbf{J}A\,\mathrm{d}l = \mathbf{J}\cdot\mathrm{d}V$（根据定义，$\mathrm{d}\mathbf{l}$ 与 $\mathbf{J}$ 的正方向相同），$\mathrm{d}V$ 是电流元的体积．于是比奥萨伐尔定律的环路积分变为体积分

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{J}(\mathbf{r}')\times\hat{\mathbf{R}}}{R^2}\,\mathrm{d}V' \tag{10}$$

注意积分内的电流密度是关于源点的函数而不是场点的函数．理论上，体积分应该在导线内部进行，然而导线外部电流密度为零，故积分可以对全空间进行．类比式 6，更明确的写法是

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{J}(\mathbf{r}')\times(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}\,\mathrm{d}V' \tag{11}$$

积分时 $\mathbf{r}$ 视为常数．

安培环路定理

- 最后的证明未完成

预备知识 比奥萨伐尔定律[1175]，斯托克斯定理[318]

在空间中选取一环路（称为**安培环路**）并定义一个正方向，那么磁感应强度在该环路上的线积分[295] 等于穿过环路的总电流（电流的正方向由右手定则[20] 判断）乘以真空中的磁导率 μ_0．

$$\oint \mathbf{B}\cdot\mathrm{d}\mathbf{r} = \mu_0 I \tag{1}$$

这就是**安培环路定理（Ampere's circuital Law）**．

上式中有几点需要注意的地方．第一，电流及其分布不随时间变化．这点可以从“不存在瞬时作用”理解，假设某时刻电流突然从 0 变为某个值，由于电磁场传播需要一定时间，环路上不可能瞬间出现磁场．第二，空间中不能有变化的电场，因为变化的电场也会产生磁场，改变环路积分的结果．

例 1　无限长直导线的磁场

若导线的电流为 I，在其周围作一个半径为 r 的安培环路，由对称性，环路上任意一点的磁感应强度大小相同且沿正方向．所以式 1 等于

$$2\pi r B = \mu_0 I \tag{2}$$

所以磁感应强度大小的分布为

$$B(r)=\frac{\mu_0}{2\pi}\frac{I}{r} \tag{3}$$

这与使用比奥萨伐尔定律（例 1 [1175]）得出的结论一致.

例 2　无限长螺线管的磁场

（图未完成）圆柱形均匀缠绕的螺线管单位长度匝数为 n 沿螺线管的轴线方向取一个长方形回路，根据对称性，垂直于轴线的方向不会有任何磁场. 所以现在螺线管外面的部分可以任意伸缩，如果伸到无穷远，则磁场为零. 所以环路积分完全由内部的平行边贡献

$$BL=\mu_0 I_{tot}=\mu_0 nLI \tag{4}$$

所以外部磁场为零，内部磁场为匀强.

$$B=\mu_0 nI \tag{5}$$

在实际情况中，如果螺线管比较细长，那我们仍然可以近似认为它的内部为匀强磁场.

微分形式

根据斯托克斯公式（式 1 [318]），可以把式 1 记为微分形式

$$\boldsymbol{\nabla}\times\mathbf{B}=\mu_0\mathbf{j} \tag{6}$$

安培定理与奥萨法尔定律

既然比奥萨法尔定律可以让我们通过空间中的电流直接计算处任意位置的磁场，那么我们就可以通过比奥萨法尔定律直接推导出安培环路定理. 事实上，这两个定律是等价的，如果要求空间中任意环路都满足安培定理，那么比奥萨法尔定律也必须成立（可以类比库仑定律和电场高斯定律之间的关系）.

（未完成，见 [12] 233 页）

洛伦兹力

预备知识 矢量的叉乘[387]

磁场 $\mathbf{B}$ 中，电荷为 q，以速度 $\mathbf{v}$ 运动的点电荷受到的洛伦兹力通过叉乘[387]定义

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B} \tag{1}$$

即洛伦兹力与速度和磁场的方向垂直，大小等于 qvB 乘以速度与磁场夹角的正弦值．可见当速度与磁场垂直时洛伦兹力最大，平行时没有洛伦兹力．

磁场对电荷不做功

由于任意时刻，磁场力的方向垂直于运动方向，所以静磁场不对电荷做功（类比向心力不对圆周运动做功），证明如下．洛伦兹力的瞬时功率为

$$P = \mathbf{F} \cdot \mathbf{v} = q\,\mathbf{v} \times \mathbf{B} \cdot \mathbf{v} \tag{2}$$

由矢量混合积的运算

$$\mathbf{v} \times \mathbf{B} \cdot \mathbf{v} = \mathbf{v} \times \mathbf{v} \cdot \mathbf{B} = 0 \tag{3}$$

因为矢量叉乘本身等于 0.

广义洛伦兹力

预备知识 极限[149]

广义上的洛伦兹力是指电磁场给电荷施加的所有作用力，即电场力加洛伦兹力．麦克斯韦方程组只描述了由电荷的分布及运动情况如何计算电磁场．而广义洛伦兹力则解释了已知电磁场分布如何计算电荷的受力．对于点电荷

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{4}$$

对于连续的电荷分布

$$\mathbf{f} = \rho(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{5}$$

其中 $\mathbf{f}$ 是受力密度，用极限的方法定义为无穷小体积的受力除以该体积

$$\mathbf{f} = \lim_{V \to 0} \frac{\mathbf{F}}{V} \tag{6}$$

霍尔效应

- 本词条处于草稿阶段.

[14]1879 年霍尔（E. C. Hall）首先观察到，把一载流导体薄片放在磁场中时，如果磁场方向垂直于薄片平面，则在薄片的上、下两侧面会出现微弱的电势差．这一现象称为**霍尔效应（Hall effect）**．此电势差称为**霍尔电势差**．实验测定，霍尔电势差的大小与电流 I 及磁感应强度 B 成正比，而与薄片沿 B 方向的厚度 d 成反比．它们的关系可写成：

$$V = R_{\mathrm{H}} \frac{IB}{d} \tag{1}$$

其中 R_H 是**霍尔系数（Hall coefficient）**，等于自由电子体密度 ρ 的倒数（式 4 ）.

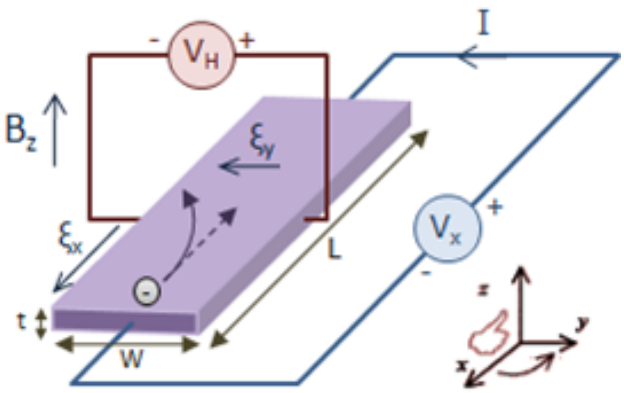


图 1: 霍尔效应示意图（来自维基百科）

推导

平衡时电场力与洛伦兹力相等

$$Eq = vBq \tag{2}$$

令 ρ 为自由电子的电荷密度

$$E = jB/\rho \tag{3}$$

[14]参考 Wikipedia 相关页面.

两端电压为

$$V = IB/(d\rho) \tag{4}$$

霍尔电阻（Hall resistance）

$$R = V/I = B/(d\rho) \tag{5}$$

磁场的能量

- 本词条存在未完成的内容.
- 本词条缺少预备知识，初学者可能会遇到困难.

预备知识 磁场矢势

磁场 $\mathbf{B}(\mathbf{r})$ 的能量密度为

$$W = \frac{1}{2\mu_0}\int \mathbf{B}^2\,\mathrm{d}V \qquad 或 \qquad W = \frac{1}{2}\int \mathbf{A}\cdot\mathbf{J}\,\mathrm{d}V \tag{1}$$

其中 μ_0 是真空中的磁导率，$\mathbf{A}$ 是磁场矢势，$\mathbf{J}$ 是电流密度，积分是对全空间积分（或者对被积函数不为零的空间积分）.

幼稚的推导

首先我们根据能量守恒的思想，假设给一个电感 l 充电的能量都以“磁场能”的形式储存起来，且每个点的能量密度只是磁场强度绝对值的函数.

在例 2 [1179] 中，我们知道螺线管中的磁场强度为

$$B = \mu_0 nI \tag{2}$$

该螺线管的电感为

$$L = \mu_0 n^2 lS \tag{3}$$

所以

未完成：...

$$W=\frac{1}{2\mu_0}\int \mathbf{B}^2\,\mathrm{d}V \tag{4}$$

简单的推导

我们首先考虑一个单匝线圈的磁场能量

未完成：图，画单股线圈，画出大概的磁场分布，画出两根相邻的电缆外接电源

假设线圈的电阻为零

假设 $t=0$ 时 $I=0$，此时没有磁场，磁场能量为零．接下来 I 随着 t 慢慢增加．反向电动势为（定义与电流相同的方向为正）

$$\varepsilon=-\frac{\mathrm{d}\Phi}{\mathrm{d}t}=-\frac{\mathrm{d}}{\mathrm{d}t}\int \mathbf{B}\cdot\mathrm{d}\mathbf{s}=-\frac{\mathrm{d}}{\mathrm{d}t}\int(\boldsymbol{\nabla}\times\mathbf{A})\cdot\mathrm{d}\mathbf{s}=-\frac{\mathrm{d}}{\mathrm{d}t}\oint \mathbf{A}\cdot\mathrm{d}\mathbf{l} \tag{5}$$

电源克服反电动势的功率为

$$\frac{\mathrm{d}W}{\mathrm{d}t}=-\varepsilon I=I\frac{\mathrm{d}}{\mathrm{d}t}\int \mathbf{B}\cdot\mathrm{d}\mathbf{s}=\int I\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}t}\cdot\mathrm{d}\mathbf{s} \tag{6}$$

由于磁场与电流成正比（见比奥萨伐尔定律[1175]），不妨设 $\mathbf{B}=\mathbf{b}I$　．则

$$I\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}t}=\mathbf{b}I\frac{\mathrm{d}I}{\mathrm{d}t}=\frac{\mathbf{b}}{2}\frac{\mathrm{d}I^2}{\mathrm{d}t}=\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}I\mathbf{B}\right) \tag{7}$$

所以

$$\frac{\mathrm{d}W}{\mathrm{d}t}=\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\int I\mathbf{B}\cdot\mathrm{d}\mathbf{s}\right) \tag{8}$$

注意两边都是对时间的导数．两边对时间积分，得

$$W=\frac{1}{2}\int I\mathbf{B}\cdot\mathrm{d}\mathbf{s} \tag{9}$$

注意当 $I=0$ 时 $W=0$，所以积分常数为零．注意在上述过程中，并没有假设电流以什么样的函数随时间变化（只要是缓慢变化即可）．

$$W=\frac{I}{2}\int \mathbf{B}\cdot\mathrm{d}\mathbf{s}=\frac{I}{2}\int \boldsymbol{\nabla}\times\mathbf{A}\cdot\mathrm{d}\mathbf{s}=\frac{1}{2}\oint I\mathbf{A}\cdot\mathrm{d}\mathbf{l} \tag{10}$$

未完成：推导

磁通量

- 本词条处于草稿阶段.
- 应该由浅入深，矢势放到后面．添加例题.

预备知识　曲面积分通量[297]

[15]令空间中磁感应强度为 $\mathbf{B}(\mathbf{r})$ 我们可以通过曲面积分来定义一个通过某曲面的**磁通量（magnetic flux）**为

$$\Phi = \int \mathbf{B}(\mathbf{r}) \cdot \mathrm{d}\mathbf{s} \tag{1}$$

形象来说，磁通量也可以看作是磁感线通过曲面的条数．反方向的磁感线共线为负，和正方向磁感线抵消.

磁通量只与曲面的边界有关．这是因为 $\mathbf{B}$ 的散度恒为零（磁场的高斯定律）

未完成：证明

利用磁场矢势及旋度定理，磁通量变为

$$\Phi = \int \boldsymbol{\nabla} \times \mathbf{A} \cdot \mathrm{d}\mathbf{s} = \oint \mathbf{A} \cdot \mathrm{d}\mathbf{r} \tag{2}$$

另外，由于磁场的散度为零，根据高斯定律，任何闭合曲面的磁通量都是 0. 用另一种方式来理解：如果选定一个闭合回路，以该闭合回路为边界的任何曲面的磁通量都相等.

闭合线圈的磁通量

如何计算一个闭合线圈对自己产生的磁通量呢？利用磁场矢势公式

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0 I}{4\pi} \oint \frac{\mathrm{d}\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} \tag{3}$$

[15]参考 [12] 和 Wikipedia 相关页面.

注意在该积分中，$\mathbf{r}$ 视为常量，积份完后，积分变量 $\mathbf{r}$ 消失．现在根据式 2 再次将上式对 $\mathbf{r}$ 进行同一环路积分得到磁通量

$$\Phi = \frac{\mu_0 I}{4\pi} \oint \oint \frac{\mathrm{d}\mathbf{r}' \, \mathrm{d}\mathbf{r}}{|\mathbf{r} - \mathbf{r}'|} \tag{4}$$

安培力

预备知识 洛伦兹力[1180]

在匀强磁场中，一段笔直的细导线中通有电流 I，导线的长度和正方向用矢量 $\mathbf{L}$ 表示[16]，若电流与 $\mathbf{L}$ 的方向相同则取正值，若相反则取负值．导线 $\mathbf{L}$ 受到的安培力为

$$\mathbf{F} = I\mathbf{L} \times \mathbf{B} \tag{1}$$

该式可由洛伦兹力推导，见下文．

当磁场分布不均匀，或导线是弯曲的，可用“微元法”的思想，把该导线分为许多小段然后对每小段的安培力 $\mathrm{d}\mathbf{F} = I\mathrm{d}L \times \mathbf{B}$ 进行矢量求和，即曲线积分）．

$$\mathbf{F} = I \int_L \mathrm{d}\mathbf{r} \times \mathbf{B} \tag{2}$$

L 为导线所在的曲线，积分方向沿电流方向．

推导（匀强磁场中的直导线）

假设导线中正电荷运动而负电荷不动[17]，运动的正电荷线密度（单位长度的电荷量）为 λ，速度为 $\mathbf{v}$．那么电流为

$$I = \lambda v \tag{3}$$

所有运动的正电荷受到的洛伦兹力为

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B} \tag{4}$$

[16]根据电流连续性定理，电流不可能在端点凭空出现或消失，所以我们可以认为 $\mathbf{L}$ 是一个回路中的一段．

[17]事实上是负电荷 $-\lambda$ 以 $-\mathbf{v}$ 运动而正电荷不动，但这样假设得到的安培力相同，且方便记忆和推导．

当电流方向 $\mathbf{v}$ 与 $\mathbf{L}$ 相同时有 $L\mathbf{v}=v\mathbf{L}$，此时定义电流为正，$I=\lambda v$，相反时有 $L\mathbf{v}=-v\mathbf{L}$，定义电流为负，$I=-\lambda v$．所以

$$q\mathbf{v}=\lambda L\mathbf{v}=\pm\lambda v\mathbf{L}=I\mathbf{L} \tag{5}$$

代入式 4 得

$$\mathbf{F}=I\mathbf{L}\times\mathbf{B} \tag{6}$$

磁矩

预备知识 力矩[947]，安培力[1186]

当我们把一个通有电流的线圈放置在磁场中时，这个线圈往往回受到安培力产生的力矩．本词条讨论如何计算一些简单的情况，先看一道例题．

例 1　匀强磁场中的长方形线圈

假设一个粗细可以忽略不计的长方形电流环路被放置在匀强磁场 $\mathbf{B}$ 中（图未完成），两条边的边长分别为 a,b，电流为 I．线圈平面的法向量（由右手定则[20] 定义）和磁场夹角为 θ，边长为 a 边始终垂直于磁场．求线圈所受力矩．

解：（未完成）力矩的方向使得线圈的法向磁场的方向转动，大小正比于 $\sin\theta$．

为了更好地表示这个结果，定义**磁矩（magnetic moment）**为

$$\boldsymbol{\mu}=I\mathbf{A} \tag{1}$$

其中 $\mathbf{A}$ 的方向是线圈的法向量，模长等于线圈的面积，我们不妨把它叫做面积矢量．根据矢量叉乘[387] 的几何定义，线圈所受力矩为

$$\boldsymbol{\tau}=\boldsymbol{\mu}\times\mathbf{B} \tag{2}$$

事实上，可以证明这个结论与线圈的形状无关，线圈甚至可以不共面（此时需要重新定义 $\mathbf{A}$），详见“磁场中闭合电流的力矩[1189]”．另外，当我们由多匝线圈时，只需将结果乘以匝数即可．

旋转的电荷

对于绕轴做圆周运动的点电荷，令角速度为 ω，等效电流为 q/T，$T=2\pi/\omega$ 是转动周期，圆周面积 $A=\pi r^2$，代入式 1 得

$$\boldsymbol{\mu}=\frac{1}{2}q\omega r^2 \tag{3}$$

磁场中闭合电流的合力

预备知识 安培力[1186]，斯托克斯定理，静磁场的高斯定律，

假设空间中有任意磁场 $\mathbf{B}(\mathbf{r})$，闭合电流回路 L 中有电流 I．则其受到的安培力可以表示为线积分 $\mathbf{F}=\oint I\,\mathrm{d}\mathbf{l}\times\mathbf{B}$．积分方向为电流方向．若磁场是匀强磁场，则立即得到 $\mathbf{F}=I(\oint\mathrm{d}\mathbf{l})\times\mathbf{B}=\mathbf{0}$ 若磁场是任意的，那么

$$\begin{aligned}\mathbf{F}&=\oint_L I\,\mathrm{d}\mathbf{l}\times\mathbf{B}=\hat{\mathbf{x}}I\oint_L\mathrm{d}\mathbf{l}\times\mathbf{B}\cdot\hat{\mathbf{x}}+\hat{\mathbf{y}}I\oint_L\mathrm{d}\mathbf{l}\times\mathbf{B}\cdot\hat{\mathbf{y}}+\hat{\mathbf{z}}I\oint_L\mathrm{d}\mathbf{l}\times\mathbf{B}\cdot\hat{\mathbf{z}}\\&=\hat{\mathbf{x}}I\oint_L(\mathbf{B}\times\hat{\mathbf{x}})\cdot\mathrm{d}\mathbf{l}+\hat{\mathbf{y}}I\oint_L(\mathbf{B}\times\hat{\mathbf{y}})\cdot\mathrm{d}\mathbf{l}+\hat{\mathbf{z}}I\oint_L(\mathbf{B}\times\hat{\mathbf{z}})\cdot\mathrm{d}\mathbf{l}\\&=\hat{\mathbf{x}}I\int_\Sigma\boldsymbol{\nabla}\times(\mathbf{B}\times\hat{\mathbf{x}})\cdot\mathrm{d}\mathbf{s}+\hat{\mathbf{y}}I\int_\Sigma\boldsymbol{\nabla}\times(\mathbf{B}\times\hat{\mathbf{y}})\cdot\mathrm{d}\mathbf{s}+\hat{\mathbf{z}}I\int_\Sigma\boldsymbol{\nabla}\times(\mathbf{B}\times\hat{\mathbf{z}})\cdot\mathrm{d}\mathbf{s}\end{aligned} \tag{1}$$

其中用到了斯托克斯定理，Σ 是以闭合曲线 L 为边界的曲面．上式中

$$\boldsymbol{\nabla}\times(\mathbf{B}\times\hat{\mathbf{x}})=\mathbf{B}(\boldsymbol{\nabla}\cdot\hat{\mathbf{x}})+(\hat{\mathbf{x}}\cdot\boldsymbol{\nabla})\mathbf{B}-\hat{\mathbf{x}}(\boldsymbol{\nabla}\cdot\mathbf{B})-(\mathbf{B}\cdot\boldsymbol{\nabla})\hat{\mathbf{x}}=(\hat{\mathbf{x}}\cdot\boldsymbol{\nabla})\mathbf{B}=\frac{\partial\mathbf{B}}{\partial x} \tag{2}$$

这里用到了 $\hat{\mathbf{x}}$ 的任意微分为 0 以及 $\boldsymbol{\nabla}\cdot\mathbf{B}=0$ 的性质．对称地，将上式中的 $\hat{\mathbf{x}}$ 替换成 $\hat{\mathbf{y}}$ 和 $\hat{\mathbf{z}}$，等式也成立．所以

$$\mathbf{F}=\hat{\mathbf{x}}I\int_\Sigma\frac{\partial\mathbf{B}}{\partial x}\cdot\mathrm{d}\mathbf{s}+\hat{\mathbf{y}}I\int_\Sigma\frac{\partial\mathbf{B}}{\partial y}\cdot\mathrm{d}\mathbf{s}+\hat{\mathbf{z}}I\int_\Sigma\frac{\partial\mathbf{B}}{\partial z}\cdot\mathrm{d}\mathbf{s} \tag{3}$$

写成分量的形式，就是

$$F_x=I\int_\Sigma\frac{\partial\mathbf{B}}{\partial x}\cdot\mathrm{d}\mathbf{s}\qquad F_y=I\int_\Sigma\frac{\partial\mathbf{B}}{\partial y}\cdot\mathrm{d}\mathbf{s}\qquad F_z=I\int_\Sigma\frac{\partial\mathbf{B}}{\partial z}\cdot\mathrm{d}\mathbf{s} \tag{4}$$

磁场中闭合电流的力矩

预备知识 磁矩[1187]，梯度[307]

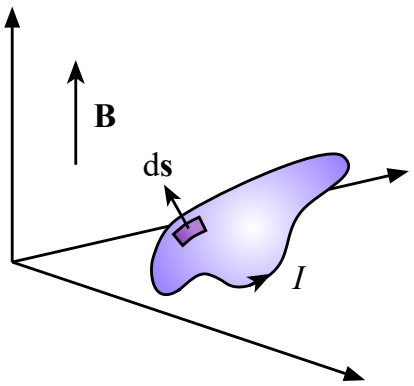


图 1: 闭合电流在磁场中所受的力矩

如图 1，在匀强磁场 $\mathbf{B}$ 中，一段闭合环路电流（忽略粗细）所受的力矩为

$$\boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B} \tag{1}$$

其中 $\boldsymbol{\mu}$ 是**磁矩（megnetic moment）**，等于电流 I 乘以以这个环路为边界的任意曲面的面积矢量，曲面法向量的方向通过右手定则[20] 判断.

$$\boldsymbol{\mu} = I\mathbf{A} \tag{2}$$

其中面积矢量 $\mathbf{A}$ 的定义为（简单来说就是把图 1 中曲面上的所有小面积元求矢量和）

$$\mathbf{A} = \oint \mathrm{d}\mathbf{s} \tag{3}$$

一种简单的情况是，若闭合环路在同一个平面上，那么 $\mathbf{A}$ 的模长就等于环路围成的面积.

更一般地，若磁场为非匀强，则力矩为

$$\boldsymbol{\tau} = I \int \mathrm{d}\mathbf{s} \times \boldsymbol{\nabla}(\mathbf{B} \cdot \mathbf{r}) \tag{4}$$

推导

预备知识 连续叉乘进行化简[392]，斯托克斯定理[315]

如图 1，线圈中有闭合电流 I，以及任意磁场分布 $\mathbf{B}(\mathbf{r})$，现在求线圈所受力矩．我们可以把线圈划分为许多小段 $\mathrm{d}\mathbf{r}$，每小段的安培力产生的力矩为

$$\mathrm{d}\boldsymbol{\tau} = \mathbf{r} \times \mathrm{d}\mathbf{F} = \mathbf{r} \times (I\,\mathrm{d}\mathbf{r} \times \mathbf{B}) \tag{5}$$

对连续叉乘进行化简[392] 得

$$\mathrm{d}\boldsymbol{\tau} = \mathbf{r} \times (I\,\mathrm{d}\mathbf{r} \times \mathbf{B}) = I(\mathbf{B}\cdot\mathbf{r})\,\mathrm{d}\mathbf{r} - I\mathbf{B}(\mathbf{r}\cdot\mathrm{d}\mathbf{r}) \tag{6}$$

对 $\mathbf{r}$ 沿闭合回路进行环积分得总力矩为

$$\boldsymbol{\tau} = \int \mathrm{d}\boldsymbol{\tau} = I\oint(\mathbf{B}\cdot\mathbf{r})\,\mathrm{d}\mathbf{r} - I\mathbf{B}\oint \mathbf{r}\cdot\mathrm{d}\mathbf{r} \tag{7}$$

其中

$$\oint \mathbf{r}\cdot\mathrm{d}\mathbf{r} = \oint r\hat{\mathbf{r}}\cdot\mathrm{d}\mathbf{r} = \oint r\,\mathrm{d}r = \frac{r^2}{2}\bigg|_{r_0}^{r_0} = 0 \tag{8}$$

这是因为环积分的起点和终点到原点的距离都相同．所以

$$\begin{aligned}\boldsymbol{\tau} &= I\oint(\mathbf{B}\cdot\mathbf{r})\,\mathrm{d}\mathbf{r}\\ &= \hat{\mathbf{x}}I\oint(\mathbf{B}\cdot\mathbf{r})\hat{\mathbf{x}}\cdot\mathrm{d}\mathbf{r} + \hat{\mathbf{y}}I\oint(\mathbf{B}\cdot\mathbf{r})\hat{\mathbf{y}}\cdot\mathrm{d}\mathbf{r} + \hat{\mathbf{z}}I\oint(\mathbf{B}\cdot\mathbf{r})\hat{\mathbf{z}}\cdot\mathrm{d}\mathbf{r}\end{aligned} \tag{9}$$

对第一项进行分析，剩下两项类推即可．由斯托克斯定理得

$$\begin{aligned}\hat{\mathbf{x}}I\oint(\mathbf{B}\cdot\mathbf{r})\hat{\mathbf{x}}\cdot\mathrm{d}\mathbf{r} &= \hat{\mathbf{x}}I\int \boldsymbol{\nabla}\times[(\mathbf{B}\cdot\mathbf{r})\hat{\mathbf{x}}]\cdot\mathrm{d}\mathbf{s}\\ &= \hat{\mathbf{x}}I\int \boldsymbol{\nabla}(\mathbf{B}\cdot\mathbf{r})\times\hat{\mathbf{x}}\cdot\mathrm{d}\mathbf{s}\\ &= \hat{\mathbf{x}}I\int \mathrm{d}\mathbf{s}\times\boldsymbol{\nabla}(\mathbf{B}\cdot\mathbf{r})\cdot\hat{\mathbf{x}}\end{aligned} \tag{10}$$

其中面积分在以环路为边界的任意曲面进行．对 $\hat{\mathbf{y}}$ 和 $\hat{\mathbf{z}}$ 项也同样处理，得式 4．在匀强电场的情况下，$\mathbf{B}$ 是常矢量，$\boldsymbol{\nabla}(\mathbf{B}\cdot\mathbf{r}) = \mathbf{B}$，得式 1．

法拉第电磁感应定律

- 本词条需要更多讲解，便于帮助理解.
- 本词条存在未完成的内容.

预备知识 磁通量[1185]

电磁感应定律的积分形式

我们在高中物理学过，闭合线圈产生的感生电动势等于线圈内磁通量随时间的变化率．方向由楞次定律[??] 决定．即

$$\varepsilon = -\frac{\mathrm{d}\Phi}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}t}\int \mathbf{B}\cdot\mathrm{d}\mathbf{a} = -\int \frac{\mathrm{d}\mathbf{B}}{\mathrm{d}t}\cdot\mathrm{d}\mathbf{a} \tag{1}$$

这里的 $\mathbf{a}$ 表示面积，积分的曲面是以线圈为边界的任意曲面．另一方面，感生电动势是由感生电场产生的．

$$\varepsilon = \oint \mathbf{E}\cdot\mathrm{d}\mathbf{r} \tag{2}$$

这里的路径积分是沿着线圈进行的．规定线圈正方向以后，式 1 中曲面的正方向由右手定则[20] 决定．

根据麦克斯韦方程组，电场产生的原因有两种，一种是电荷产生电场（电场的高斯定律[1161]），另一种是变化的磁场产生电场（法拉第电磁感应）．前者在式 2 中的环路积分为零，对电动势没有贡献．所以式 2 中的 $\mathbf{E}$ 既可以只包含感生电场，也可以是总电场．我们一般理解为总电场．

对比上面两式，得

$$\oint \mathbf{E}\cdot\mathrm{d}\mathbf{r} = -\int \frac{\partial\mathbf{B}}{\partial t}\cdot\mathrm{d}\mathbf{a} \tag{3}$$

如果我们假设感生电场只与电场的分布和变化率有关，则这个公式对空间中任何假想中的回路都成立，而不需要有真正的线圈存在．注意上式中的磁场是空间中的所有磁场．

电磁感应定律的微分形式

预备知识 斯托克斯定理[318]

斯托克斯定理告诉我们，若对任意闭合回路，一个矢量场对曲面正方向的面积分等于另一个场在曲面边界线正方向的线积分，那么前者是后者的旋度.

未完成：“斯托克斯定理”并没有提到这个逆向的结论

应用到上式，可得电场的旋度为

$$\boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{4}$$

位移电流

- 本词条处于草稿阶段.

位移电流（displacement current）

定义为

$$\mathbf{j}_d = \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \tag{1}$$

分子电流和分子磁矩

我们知道任何实物都是由分子、原子组成的，而分子或原子中任何一个电子都不停地同时参与两种运动：一种是环绕原子核的轨道运动，另一种是电子本身的自旋运动. 这两种运动都等效于一个电流分布，因而能产生磁效应把分子或原子看作一个整体，分子或原子中各个电子对外界所产生磁效应的总和，可用一个等效的圆电流表示，统称**分子电流**. 这种分子电流具有一定的磁矩，称为**分子磁矩（molecular magnetic moment）**，用符号 m_{mole} 表示.

在外磁场 B_0 作用下，分子或原子中和每个电子相联系的磁矩都受到磁力矩的作用，由于分子或原子中的电子以一定的角动量作高速转动，这时，每个电子除了保持上述两种运动以外，还要附加电子磁矩以外磁场方向为轴线的

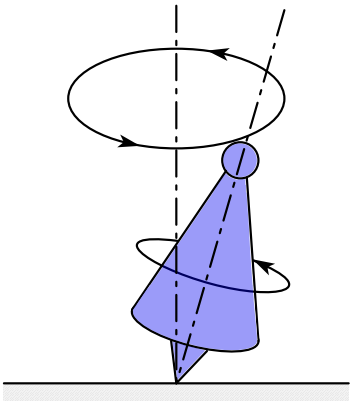


图 1: 陀螺的进动示意图

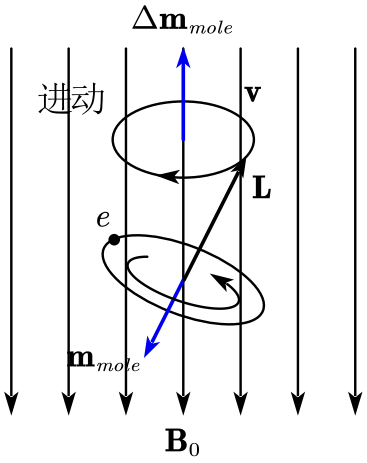


图 2: 在外磁场中电子的进动和附加磁矩

转动，称为电子的进动这与力学中所讲的高速旋转着的陀螺，在重力矩的作用下，以重力方向力轴线所作的进动十分相似，如图所示.

我们可以得知，无论电于原来的磁矩与磁场方向之间的夹角是何值，在外磁场 B_0 中，电子角动量 L 进动的转向总是和 B_0 的方向构成右手螺旋关系，如图 2 和图 3 所示. 电子的进动也相当于一个圆电流. 因为电子带负电，这种等效的圆电流的此举的方向永远与 B_0 的方向相反，原子或分子中各个电子因进动而产生的磁效应的总和也可用一个等效的分子电流的磁矩来表示，因进动而产生的等效电流的磁矩称为**附加磁矩**，用 Δm_{mole} 表示.

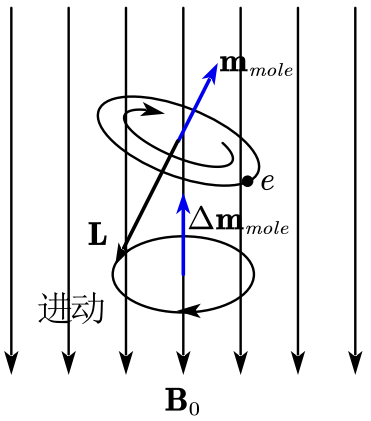


图 3: 在外磁场中电子的进动和附加磁矩

磁介质

处于磁场中的实物物质，都会呈现不同程度的磁性，我们把物质具有磁性的过程称为**磁化（magnetization）**.一切能够被磁化的实物称为**磁介质（magnetic material）**．设某一电流分布在真空中激发的磁感应强度为 B_0，磁场中放进了某种磁介质后，磁化了的磁介质激发附加磁感应强度 B'，这时磁场中任一点的磁感应强度 B 等于 B_0 和 B' 的矢量和，即

$$\mathrm{B} = \mathrm{B}_0 + \mathrm{B}' \tag{1}$$

由于磁介质的磁化特性不同，有一些磁介质磁化后使磁介质中的磁感应强度 B 稍大于 B_0，**这类磁介质称为顺磁质（paramagnetic material）**，例如锰、铬、铂、氮等都属于顺磁性物质；另一些磁介质磁化后使磁介质中的磁感应强度 B 稍小于 B_0，即这类磁介质称为**抗磁质（diamagnetic material）**，例如水银、铜、铋、硫、氯、氢、银、金、锌、铅等都属于抗磁性物质．一切抗磁质以及大多数顺磁质都有一个共同点，那就是它们所激发的附加磁场极其微弱，B 和 B_0 相差很小．此外还有另一类磁介质，它们磁化后所激发的附加磁感应强度 B' 远大于 B_0，这类能显著地增强磁场的物质，称为**铁磁质（ferromagnetic material）**，例如铁、镍、钴、钆以及这些金属的合金，还有铁氧体等物质都是铁磁质．

磁化强度

预备知识 磁介质[1194]，线积分[295]

为了表征磁介质磁化的程度，与讨论电介质时定义极化强度一样引进一个物理量，叫做**磁化强度（magnetization intensity）**. 在被磁化后的磁介质内，任取一体积元 ΔV. 在这体积元中所有分子固有磁矩的矢量和 $\sum \mathrm{m}_{mole}$ 加上附加磁矩的矢量和 $\sum \Delta \mathrm{m}_{mole}$ 与该体积元的比值，即单位体积内分子磁矩的矢量和，称为**磁化强度**，用 M 表示. 即

$$\mathrm{M}=\frac{\sum \mathrm{m}_{mole}+\sum \Delta \mathrm{m}_{mole}}{\Delta V} \tag{1}$$

对于顺磁质，$\sum \Delta \mathrm{m}_{mole}$ 可以忽略；对于抗磁质，$\sum \mathrm{m}_{mole}=0$；对于真空，$\mathrm{M}=0$. 如果在介质中各点的 M 相同，就称磁介质被均匀磁化. 在国际单位制中，M 的单位是 A/m.

顺磁质经磁化后，M 的方向与该处的磁场 B 一致，它在磁介质内所激发的附加磁场 B′ 的方向也与 B_0 的方向相同. 抗磁质经磁化后，M 的方向与该处磁场 B 相反，它在磁介质内所激发的附加磁场 B′ 的方向也与 B_0 的方向相反. 磁介质的磁化情况，可以用磁化强度 M 来描述，也可以用磁化电流来反映. 磁化电流实质上是分子电流的宏观表现，它与磁化强度 M 之间必然存在一定的联系. 下面，我们将用直观的方法找出能测定的宏观的磁化强度与磁化电流之间的关系. 为简单起见，我们选一特例来讨论. 设有一无限长的载流直螺线管，管内充满均匀磁介质，电流在螺线管内激发均匀磁场，如图 1 所示. 在此磁场中磁介质被均匀磁化，这时磁介质中各个分子电流平面将转到与磁场的方向相垂直.

我们来看磁介质内任一截面上分子电流排列的情况. 从图 2 和图 3 中可以看出，在磁介质内部任意一点处，总是有两个方向相反的分子电流通过，结果相且抵消；只有在截面边缘处，分子电流未被抵消，形成与截面边缘重合的圆电流. 对磁介质的整体来说，未被抵消的分子电流是沿着柱面流动的，称为**安培表面电流**（或叫**磁化面电流**）. 对顺磁性物质，安培表面电流和螺线管上导线中的电流 I 方向相同；对抗磁性物质，则两者方向相反. 我们上面的图中都是顺磁质的情况. 设 α_s 为圆柱形磁介质表面上单位长度的磁化电流，即磁化

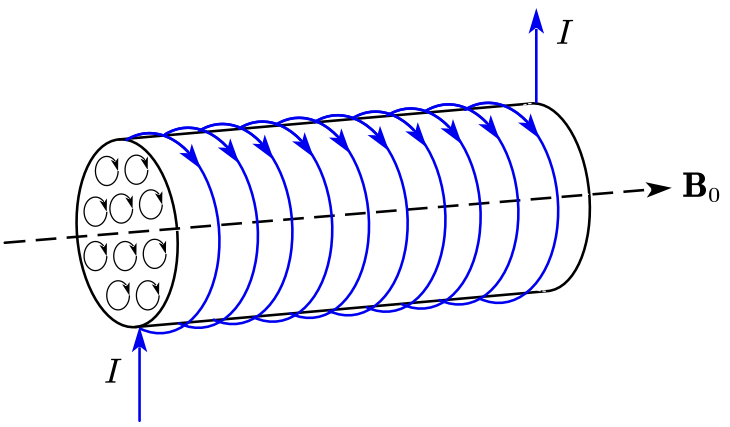


图 1: 无限长的载流直螺线管

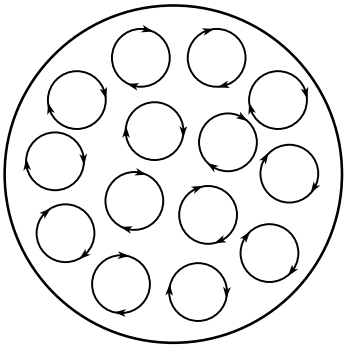


图 2: 任一截面上分子电流排列

面电流的线密度，S 为磁介质的截面积，l 为所选取的一段磁介质的长度．在 l 长度上，表面电流的总量值为 $I_s = \alpha_s l$，因此在这段磁介质总体积 Sl 中的总磁矩为

$$\sum \mathrm{m}_{mole} = I_s \mathrm{S} = \alpha_s l \mathrm{S} \tag{2}$$

按定义，M 为单位体积内的磁矩，所以

$$M = \frac{|\sum \mathrm{m}_{mole}|}{V} = \frac{\alpha_s Sl}{Sl} = \alpha_s \tag{3}$$

即磁介质表面某处的磁化面电流线密度的大小等于该处磁化强度的量值．上述结果是从均匀磁介质被均匀磁化的特例导出的，在一般情况中，应该是磁介质表面上磁化面电流线密度应等于该处磁化强度的切线分量．而且在不均匀磁介质内部，由于排列着的分子电流未能相互抵消，此时磁体内各点都有磁化电流．

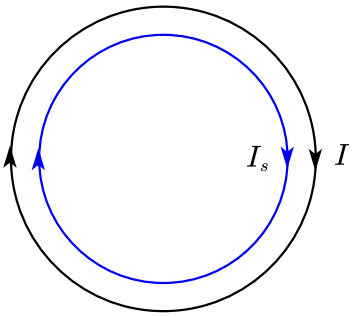


图 3: 任一截面上分子电流排列

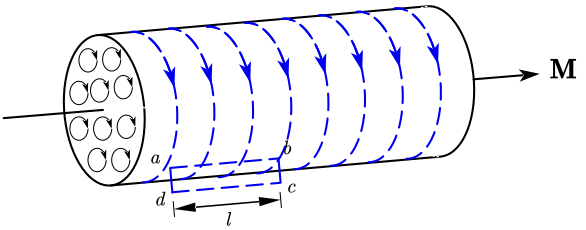


图 4: 无限长的载流直螺线管

为了求得磁化电流与磁化强度的联系，我们来计算磁化强度对闭合回路的线积分 $\oint \mathrm{M}\cdot\mathrm{dl}$ 与磁化电流的关系．仍用前述特例，在图 4 所示的圆柱形磁介质的边界附近，取一长方形闭合回路 $abcd$，ab 边在磁介质内部，它平行于柱体轴线，长度为 l．而 bc、ad 两边则垂直于柱面在磁介质内部各点处，M 都沿 ab 方向，大小相等，在柱外各点处 $\mathrm{M}=0$，所以 M 沿 bc、cd、da 三边的积分为零，因而 M 对闭合回路 $abcd$ 的积分等于 M 沿 ab 边的积分，即

$$\oint \mathrm{M}\cdot\mathrm{dl}=\int_a^b \mathrm{M}\cdot\mathrm{dl}=M|ab|=Ml \tag{4}$$

将 $M=\alpha_s$, 代入后得

$$\oint \mathrm{M}\cdot\mathrm{d}l=\alpha_s l=I_s \tag{5}$$

这里，$\alpha_s l=I_s$ 就是通过以闭合回路 $abcd$ 为边界的任意曲面的总磁化电流，所以式 5 表明**磁化强度对闭合回路的线积分等于通过回路所包围的面积内的总磁化电流**．式 5 虽是从均匀磁化介质及长方形闭合回路的简单特例导出的，但却是在任何情况下都普遍适用的关系式．

顺磁质的磁化

对顺磁质而言，虽然每个原子或分子有一定的磁矩，但由于分子的无规则热运动，各个分子磁矩排列的方向是十分纷乱的，对顺磁质内任何一个体积元来说，其中各分子的分子磁矩的矢量和 $\sum \mathrm{m}_{mole}=0$，因而对外界不显示磁效应．在外磁场 B_0 的作用下，分子磁矩 m_{mole} 的大小不改变，但是外磁场 B_0 要促使，m_{mole} 绕磁场方向进动，并具有一定的能量．这里的能量，是说固有磁矩 m_{mole} 在外磁场中的能量为 $W_m=-\mathrm{m}_{mole}\cdot\mathrm{B}$．由此可见，当 m_{mole} 与 B_0 方向相同时，能量最低．

我们知道，介质中存在着大量原子或分子，由于这些原子或分子之间的相互作用和碰撞，促使分子磁矩 m_{mole} 改变方向，从而改变 m_{mole} 在外磁场中的能量状态．一方面，从能量的角度来看，分子磁矩尽可能要处于低的能量状态，即 m_{mole} 与外磁场方向一致的状态；另一方面，分子热运动又破坏了分子磁矩沿磁场方向有序排列．当达到热平衡时，原子或分子的能量遵守玻耳兹曼分布律，处在较低能量状态的原子数或分子数比高能量状态的要多，亦即其分子磁矩 m_{mole} 靠近外磁场方向的分子数较多．显然，磁场越强，温度越低，分子磁矩 m_{mole} 排列也越整齐，这时，在顺磁体内任取一体积元 ΔV，其中各分子磁矩的矢量 $\sum \mathrm{m}_{mole}$ 将有一定的量值，因而在宏观上呈现出一个与外磁场同方向的附加磁场．这就是顺磁性的来源．

应当指出，顺磁质受到外磁场的作用后，其中的原子或分子也会产生抗磁性，但在通常情况下，多数顺磁质分子的附加磁矩 $\sum \Delta\mathrm{m}_{mole}$ 比 $\sum \mathrm{m}_{mole}$ 小很多，所以这些磁介质主要显示出顺磁性．

抗磁质的磁化

- 本词条处于草稿阶段．
- 待审核

在抗磁质中，每个原子或分子中所有电子的**轨道磁矩（orbital magnetic moment）**和**自旋磁矩（spin magnetic moment）**的矢量和等于零，在外磁场 B_0

中电子轨道运动的平面在磁场中会发生进动，而且其轨道角动量进动的方向在任何情况下都是沿着磁场的方向，和电子轨道运动的速度方向无关，并在同一外磁场 B_0 中都以相同的角速度进动．因此，这时抗磁质中每个分子或原子中所有的电子形成一个整体绕外磁场的进动，从而产生一个附加磁矩 $\Delta \mathrm{m}_{mole}$，$\Delta \mathrm{m}_{mole}$ 的方向与 B_0 的方向相反，大小与 B_0 的大小成正比．这样，抗磁材料在外磁场的作用下，磁体内任一体积元中大量分子或原子的附加磁矩的矢量和 $\sum \Delta \mathrm{m}_{mole}$ 有一定的量值，结果在磁体内激发一个和外磁场方向相反的附加磁场，这就是抗磁性的起源．抗磁性既然起源于外磁场对电子轨道运动作用的结果，应该在任何原子或分子的结构中都会产生，因此它是一切磁介质所共有的性质．

有磁介质时的安培环路定理

预备知识 安培环路定理[1179]，磁化强度[1195]

当电流的磁场中有磁介质存在时，空间任一点的磁感应强度 B 等于导线中传导电流所激发的磁场与磁介质磁化后磁化电流所激发的附加磁场的矢量和，这时安培环路定理应写成

$$\oint \mathrm{B} \cdot \mathrm{dl} = \mu_0 \left(\sum I + I_s \right) \tag{1}$$

等式右边的两项电流是穿过以回路为边界的任一曲面的总电流，即传导电流 $\sum I$ 和磁化电流 I_s 的代数和．一般说来，I 是可以测量的，可认为它是已知的；而 I_s 不能事先给定，也无法直接测量，它依赖于介质磁化的情况，而介质的磁化情况又依赖于磁介质中的磁感应强度 B，因此直接求解方程式 1 很复杂．为了解决这一困难，我们利用关系式 5 [1195] 将式 1 改写成

$$\oint \mathrm{B} \cdot \mathrm{dl} = \mu_0 \left(\sum I + \oint \mathrm{M} \cdot \mathrm{dl} \right) \tag{2}$$

或

$$\oint \left(\frac{\mathrm{B}}{\mu_0} - \mathrm{M} \right) \cdot \mathrm{dl} = \sum I \tag{3}$$

然后引进一个新的物理量，称为**磁场强度（magnetic intensity）**，用符号 H 表示，定义为

$$\mathrm{H} = \frac{\mathrm{B}}{\mu_0} - \mathrm{M} \tag{4}$$

这样，便有下列简单的形式：

$$\oint \mathrm{H} \cdot \mathrm{dl} = \sum I \tag{5}$$

式 5 称为有磁介质时的安培环路定理．它表明 H 矢量的环流只和传导电流 I 有关，而在形式上与磁介质的磁性无关．因此引入磁场强度 H 这个物理量以后，在磁场分布具有高度对称性时，能够使我们比较方便地处理有磁介质时的磁场问题，安培环路定理和静磁场的另一普遍规律一磁场中的高斯定律一起，是处理静磁场问题的基本定理．

在国际单位制中，H 的单位是 A/m．

式 4 表示了磁介质中任一点处磁感应强度 B、磁场强度 H 和磁化强度 M 之间的普遍关系，不论磁介质是否均匀，甚至对铁磁性物质都能适用．但是，磁化强度 M 不仅和磁介质的性质有关，也和磁介质所在处的磁场有关．我们定义

$$\chi_m = \frac{M}{H} \tag{6}$$

为磁介质的**磁化率（magnetic susceptibility）**，是一与磁介质的性质有关的物理量．因为 M 和 H 所用的单位相同，所以磁化率 χ_m，是单位为 1 的量．如果磁介质是均匀的，则 χ_m 是常量；如果磁介质是不均匀的，则 χ_m 是空间位置的函数．对于顺磁质，$\chi_m > 0$，磁化强度 M 和磁场强度 H 的方向相同；对于抗磁质，$\chi_m < 0$，磁化强度 M 和磁场强度 H 的方向相反．式 6 又可写为：

$$\mathrm{M} = \chi_m \mathrm{H} \tag{7}$$

将其代入式 4 中可解得

$$\mathrm{B} = \mu_0 \mathrm{H} + \mu_0 \mathrm{M} = \mu_0 \left(1 + \chi_m\right) \mathrm{H} \tag{8}$$

通常称

$$\mu_r = 1 + \chi_m \tag{9}$$

为该介质的**相对磁导率（relative magnetic permeability）**. 于是式 7 就可以写为：

$$\mathrm{B} = \mu_0\mu_\mathrm{r}\mathrm{H} \triangleq \mu\mathrm{H} \tag{10}$$

其中 $\mu = \mu_0\mu_\mathrm{r}$ 成为磁介质的**磁导率（magnetic permeability）**. 对于真空中的磁场来说，由于 $\mathrm{M} = 0$，所以“真空”的 $\mu_r = 1, \chi_m = 0$. 真空中各点处的磁场强度 H 等于该点磁感应强度 B 的 $1/\mu_0$ 倍，即 $\mathrm{H} = \mathrm{B}/\mu_0$.

电路

预备知识 欧姆定律[1157]

零势能点

虽然我们可以规定电路中某点电势为零，但大多数情况下这么做没有什么意义，因为在这些电路中我们只讨论两点间的电压（电势差）. 例如电阻器两端的电压与电流和电阻值有关而与零势点的选取无关.（未完成）

不存在净电荷

宏观上来说，我们一般假设电路中的任意一点不能存在净电荷（电容器除外）. 这是因为电路中通常讨论的电压所能产生的净电荷往往可以忽略不计. 这里通过一个例子说明.

例 1

一个半径为 1 cm 的导体小球与无穷远之间的电容约为 1.1×10^{-12} F（见例 1 [1145]）. 如果一个 2 V 电源的正负极分别通过长导线连接两个这样的小球，每个小球上最终只能积累 $\pm 1.1 \times 10^{-12}$ C 的电荷，仅相当于 1 A 的电流流动 1.1×10^{-12} 秒.

这就是为什么在电路中，我们一般认为**开路中电流为零，只有回路才能产生电流**.

有了这个假设，我们还可以得到基尔霍夫电流定律[1202]. 这个定理简单来说就是若一条电流为 I 的导线分叉为两条电流为 I_1 和 I_2 导线，那么必有

$I_1 + I_2 = I$．如果该式不成立，就说明节点处在不断积累净电荷，从而违反了以上假设．

瞬间平衡

我们通常假设电路可以瞬间达到平衡状态．例如给电源两端接上电阻后，我们假设回路中的电流瞬间就能达到最终的稳定值而无需经历由小变大的过程．又例如给电源两端接上电容后，我们假设电容瞬间被充满．在实际中达到平衡所需的时间也是极短的，这时因为电路中自由电子的密度很高，每克材料中都约有阿伏伽德罗常数数量级的自由电子，电路中常见的电流对应的电子运动速度是极慢的，所以它们加速的时间可以忽略．

接地

在电路中如果我们要规定零势点，可以用接地符号．这时我们假设大地的电势为零，且电流可以流入和流出．

在实际运用中，接地符号往往并不需要真的接地，只是一种用于简化电路图的手段，相当于把所有接地的点都用导线连接起来．例如（图未完成：电源负极接地，正极并联一些电阻电容，它们的另一端也接地）

基尔霍夫电路定律

预备知识 电路[1201]

定律阐述

定理 1　基尔霍夫电流定律

基尔霍夫电流定律又称为基尔霍夫第一定律，规定在电路中所有进入某节点的电流的总和等于所有离开这节点的电流的总和．或者说，假设进入某节点的电流为正值，离开这节点的电流为负值，则所有涉及这节点的电流的代数和等于零．以方程式表达，对于电路的任意节点，有

$$\sum_{k=1}^{n} I_k = 0 \tag{1}$$

其中，I_k 是第 k 个进入或离开这节点的电流，是流过与这节点相连接的第 k 个支路的电流，可以是实数或复数.

证明

考虑电路的某节点，跟这节点相连接有 n 个支路. 假设进入这节点的电流为正值，离开这节点的电流为负值，则这节点的总电流 I 等于流过支路 k 的电流 I_k 的代数和：

$$I = \sum_{k=1}^{n} I_k \tag{2}$$

将这方程式对某段时间 $[t_1, t_2]$ 内积分，可以得到这段时间该节点电荷的增加

$$q = \sum_{k=1}^{n} q_k \tag{3}$$

其中 $q = \int_{t_1}^{t_2} I(t)\,\mathrm{d}t$，$q_k = \int_{t_1}^{t_2} I_k(t)\,\mathrm{d}t$ 是流过支路 k 的电荷.

若 $q > 0$，则说明有正电荷会累积于该节点，$q < 0$ 表示负电荷会累积于节点. 在讨论电路时，我们一般假设任意一点不存在净电荷[1201]，所以 q 和 I 都恒为零.

例 1

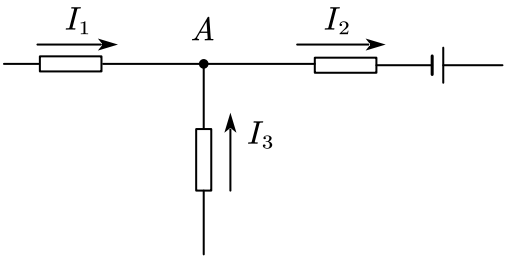


图 1: 列节点方程

各支路电流往往是未知量，它们的方向事先并不知道. 这时，可以先给每个支路电流假设一个方向，并按照这一方向列出方程. 求解方程后，如果求得某支路电流的数值为正，则该电流的实际方向与假设方向相同，否则相反. 这个假设的电流方向叫做电流的正方向. 给每一支路电流假设一个正方向之后，

就可用代数量描写每条支路的电流，代数量的绝对值反映电流的大小，代数量的正负则反映电流的实际方向．正方向一经选定，节点方程节点方程的形式（等号左右两边应写哪些电流）就完全确定．例如，为列出图 1 中节点 A 的方程，可任意地选定与 A 有关的三个支路电流的正方向如图箭头所示，从而写出如下的节点方程：

$$I_1 + I_3 - I_2 = 0 \tag{4}$$

电路中有 n 个节点的时候，一共有 $n-1$ 个节点方程是独立的．这 $n-1$ 个独立方程构成**基尔霍夫第一方程组**．将它们与基尔霍夫第二方程组联立后，可以根据已知电动势及电阻求得每一支路的电流．

定理 2　基尔霍夫电压定律

基尔霍夫电压定律又称为基尔霍夫第二定律，表明沿着闭合回路所有元件两端的电势差（电压）的代数和等于零．或者，换句话说，沿着闭合回路的所有电动势的代数和等于所有电压降的代数和．以方程式表达，对于电路的任意闭合回路，

$$\sum_{k=1}^{m} U_k = 0 \tag{5}$$

其中，m 是此闭合回路的元件数目，U_k 是元件两端的电压，可以是实数或复数．

证明

根据电势差的定义（式 1 [1143]）

$$U_{21} = V(\mathbf{r}_2) - V(\mathbf{r}_1) = -\int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{E}_0(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} \tag{6}$$

如果令路径起点为 $\mathbf{r}_1$，终点为 $\mathbf{r}_N$，中途有若干点 $\mathbf{r}_2, \ldots, \mathbf{r}_{N-1}$．那么有可以将路径积分划分为若干段，总电势差等与每段电势差之和

$$U_{N1} = U_{21} + U_{32} + \cdots + U_{N,N-1} \tag{7}$$

其中 $U_{j,I} = -\int_{\mathbf{r}_i}^{\mathbf{r}_j} \mathbf{E}_0(\mathbf{r}) \cdot \mathrm{d}\mathbf{r}$．如果取一个环路作为积分路径，起点终点相接，电势差为零．即 $\mathbf{r}_1 = \mathbf{r}_N$，$U_{N1} = 0$．立即可得式 5．证毕．

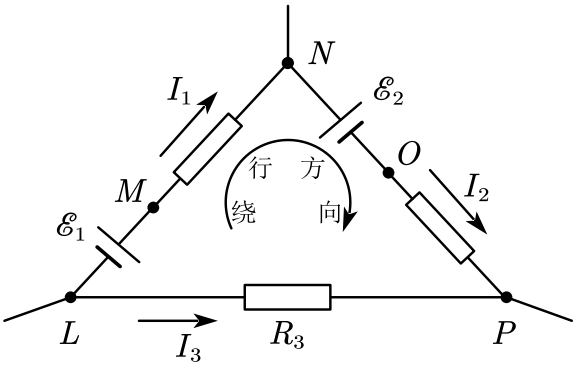


图 2: 复杂电路中的一个回路

例 2

据图 2，由基尔霍夫电压定律可知，

$$U_{ML} + U_{NM} + U_{ON} + U_{PO} + U_{LP} = 0 \tag{8}$$

设想有一个观察者从 L 点出发沿图中圆形箭头所示的方向绕行回路一周回到 L 点，他沿途看到电势有时升高有时降低，而且升、降的总量相等．由图看出，电势从 L 到 M 升高了数值 $\mathscr{E}_1$，从 M 到 N 降低了数值 I_1R_1，依次类推，于是式 8 可以表示为

$$\mathscr{E}_1 - I_1R_1 - \mathscr{E}_2 - I_2R_2 + I_3R_3 = 0 \tag{9}$$

由以上讨论不难看式 9 中每项前面的正、负号应由以下规则确定：任意选定一个绕行回路的方向（叫做绕行方向），当绕行方向从负极进入电源时（如 $\mathscr{E}_1$），其电动势前写 $+$ 号，否则写 $-$ 号（如 $\mathscr{E}_2$）；当绕行方向与电阻的电流正方向相同时（如 R_1 和 R_3），该电阻的 IR 项前写 $-$ 号，否则写 $+$ 号（如 R_2）．再次强调电流可正可负，正电流表示与正方向形同，负电流则相反，所以 IR 也可正可负．

一个电路可以包含许多回路，但它们的方程并非都是独立的．如下图电路所示：AR_2BR_1A、AR_3BR_2A 及 AR_3BR_1A．

前两个回路的方程显然独立，因为每个回路都包含一条另一回路所不包含的支路．但第三个回路的方程就不独立，它可由前两个方程推出．电路中所有

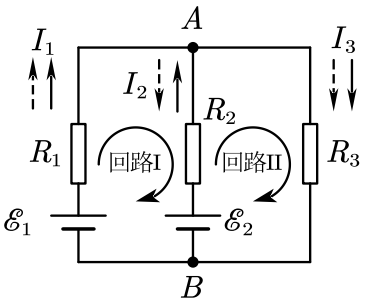


图 3: 独立回路

独立的回路方程构成**基尔霍夫第二方程组**. 为了列出独立的回路方程，可以选择这样的回路，其中每个至少包含一条其他回路所不包含的支路. 一个完整电路的支路数 b、节点数 n 和独立回路数 m 之间的关系为

$$b = m + n - 1 \tag{10}$$

如果全部电动势及电阻皆已知，则电路共有 b 个未知的支路电流. 另一方面，由前述可知，这个电路必有 $n-1$ 个独立的节点方程及 m 个独立的回路方程，即共有 $m+n-1$ 个独立方程，恰与未知量个数 b 相等，因此可唯一地解出各支路电流. 当然，除支路电流外，电动势或电阻也可作为未知量，只要未知量个数为 b，同样可以求解. 可见基尔霍夫方程组原则上可以解决一切线性直流电路的计算问题.

当电动势是待求量而且连电源的极性也未知时，可以任意地给电动势选定一个正方向（即假设一对正、负极，电动势的正方向是指从假设的负极到正极的方向），并把电动势作为代数量列出基氏第二方程，方程中 $\mathscr{E}$ 前的 $+$、$-$ 号应根据绕行方向是否进入假设的负极来决定. 求解后，如果 $\mathscr{E} > 0$，则实际极性与假设极性相同，否则相反.

定律应用

首先，根据上节所述，可以总结出用基尔霍夫方程组解题的步骤如下：

1. 任意选定各支路电流的正方向；

2. 数出节点数 n，任取其中 $n-1$ 个写出 $n-1$ 个节点方程；

3. 数出支路数 b，选定 $m=b-n+1$ 个独立回路，任意指定每个回路的绕行方向，列出 m 个回路方程；

4. 对所列的 $(n-1)+(b-n+1)=b$ 个方程联立求解；

5. 根据所得电流值的正负判断各电流的实际方向.

下面来看几道具体的电路例题.

例 3

图 3 中，已知 $\mathscr{E}_1=32\text{V},\mathscr{E}_2=24\text{V},R_1=5\,,R_2=6\,,R_3=54$，求各支路的电流.

选定 I_1、I_2、I_3 的正方向如图实箭头所示，虚箭头则代表实际方向，待求解后方可确定. 观察知，节点数 $n=2$，所以只有一个节点方程：

$$I_3-I_1-I_2=0 \tag{11}$$

又因支路数 $b=3$，故独立回路数 $m=b-n+1=2$. 选图中 I、II 两个独立回路，约定其绕行方向如图圆形箭头所示，列出回路方程：

$$\begin{aligned}\text{回路I}:&\quad \mathscr{E}_1-I_1R_1+I_2R_2-\mathscr{E}_2=0\\ \text{回路II}:&\quad \mathscr{E}_2-I_2R_2-I_3R_3=0\end{aligned} \tag{12}$$

联立式 11、式 12，解得：

$$\begin{cases}I_1=1\text{A}\\ I_2=-0.5\text{A}\\ I_3=0.5\text{A}\end{cases} \tag{13}$$

可见 I_1、I_3 的实际方向与假设的正方向相同，I_2 的实际方向与正方向相反. 三个电流的实际方向在图中用虚箭头标出.

例 4　桥式电路

已知图 4 中 $R_1=50\Omega$，$R_2=40\Omega$，$R_3=15\Omega$，$R_4=26\Omega$，$R_5=10\Omega$，求 A、B 之间的总电阻.

设想在 A、B 之间接入端电压为 U 的无内阻电源使成完整的电路，用基尔霍夫方程组求出从 A 流进的电流 I，则比值 U/I 便是 A、B 之间的总电阻. 把惠斯通电桥[1209] 中的 I_G、R_G 分别改为 I_5,R_5，即可得出答案. 求解可得 $I=U/32\Omega$，即 A、B 之间的总电阻为 32Ω.

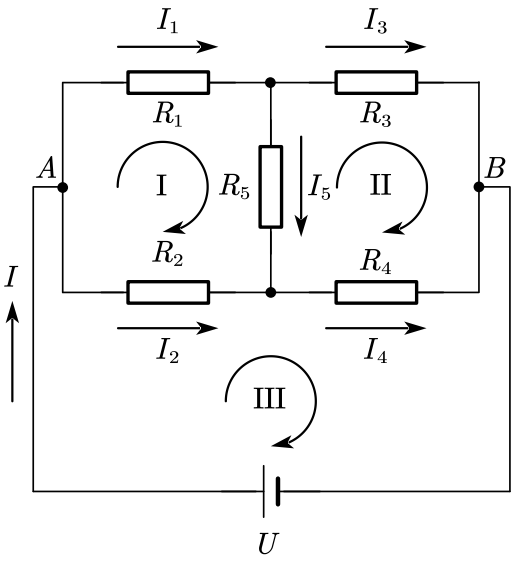


图 4: 桥式电路

电容—电阻电路充放电曲线

- 本词条处于草稿阶段.

预备知识 电容[1145]

回路中有电源，电阻和电容．开关闭合以后，求电容两端电压随时间的变化．

未完成：图

环路电压需要满足

$$U - IR - U_c = 0 \tag{1}$$

根据电路中电流处处相同，可以列出

$$I = C\frac{\mathrm{d}U_c}{\mathrm{d}t} \tag{2}$$

代入得

$$RC\frac{\mathrm{d}U_c}{\mathrm{d}t} + U_c + U = 0 \tag{3}$$

这是一个一阶线性常微分方程[212]．初始时 $U_c=0$，解得

$$U_c(t)=U\left(1-\mathrm{e}^{-t/(RC)}\right) \tag{4}$$

可以看到当 $t\to\infty$ 时 $U_c=U$．

未完成：放电情况

惠斯通电桥

预备知识 基尔霍夫电路定律[1202]

惠斯通电桥（Wheatstone bridge）也叫**惠斯登**电桥．（介绍未完成）

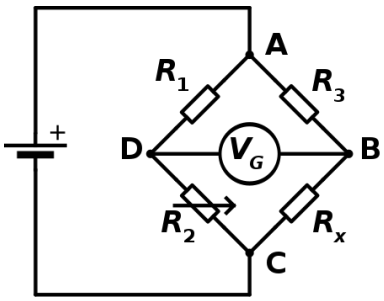


图 1: 惠斯通电桥

求电桥图 2 中电流计的电流 I_G 与电源电动势及各臂电阻的关系（电源内阻可忽略）．

选定各支路电流 $I_1,I_2,I_3,I_4,I_{\mathrm{G}}$ 及 I 的正方向如图 2 中实箭头所示．因节点数 $n=4$，故可列出三个节点方程：

$$\begin{aligned}
&\text{节点}A:\quad I=I_1+I_2\\
&\text{节点}B:\quad I_1=I_3+I_G\\
&\text{节点}C:\quad I_3+I_4=I
\end{aligned} \tag{1}$$

又因支路数 $b=6$，故独立回路数 $m=b-n+1=3$ 选图中 I、II、III 三个独立回路，约定其绕行方向如图圆形箭头所示，列出回路方程：

$$\begin{aligned}
&\text{回路I}:\quad I_1R_1+I_{\mathrm{G}}R_{\mathrm{G}}-I_2R_2=0\\
&\text{回路II}:\quad I_3R_3-I_4R_4-I_{\mathrm{G}}R_{\mathrm{G}}=0\\
&\text{回路III}:\quad I_2R_2+I_4R_4=\mathscr{E}
\end{aligned} \tag{2}$$

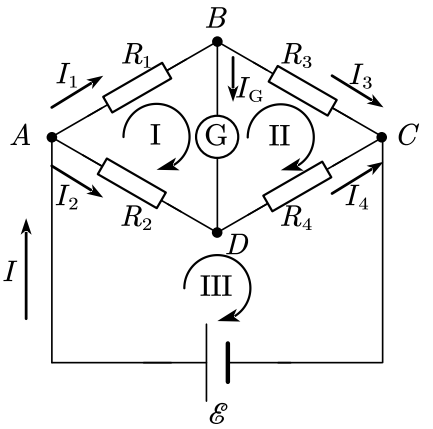


图 2: 惠斯通电桥

六个方程联立解得：

$$I_{\mathrm{G}}=\frac{\left(R_2R_3-R_1R_4\right)\mathscr{E}}{R_1R_3\left(R_2+R_4\right)+R_2R_4\left(R_1+R_3\right)+R_{\mathrm{G}}\left(R_1+R_3\right)\left(R_2+R_4\right)} \tag{3}$$

由上式可知电桥平衡（即 $I_{\mathrm{G}}=0$）的充要条件为

$$R_1R_4=R_2R_3 \tag{4}$$

式 3 说明，当 $R_2R_3-R_1R_4>0$ 时 $I_{\mathrm{G}}>0$，电流 I_{G} 的实际方向与正方向一致（向下）；反之，当 $R_2R_3-R_1R_4<0$ 时 $I_{\mathrm{G}}<0$，I_{G} 的实际方向与正方向相反（向上）.

第二章

电动力学 2

电荷守恒 电流连续性方程

预备知识 散度定理[311]，电流密度[1123]

总电荷守恒：一个封闭系统的总电荷保持不变（正反粒子带有不同的电荷，可以成对创造或湮灭，但总电荷仍然保持不变）.

然而这并没有阻止一些电荷从一个地点凭空消失同时在另一个地点出现，所以我们还可以描述得更精确一些. 在空间中任意选取一个闭合曲面，曲面内的总电荷的减少的速率等于由内向外通过曲面的速率.

$$\oint \mathbf{j}\cdot\mathrm{d}\mathbf{s} = -\frac{\mathrm{d}}{\mathrm{d}t}\int \rho\,\mathrm{d}V \tag{1}$$

由于不同变量的积分和求导可以交换

$$\frac{\mathrm{d}}{\mathrm{d}t}\oint \rho\,\mathrm{d}V = \oint \frac{\partial\rho}{\partial t}\,\mathrm{d}V \tag{2}$$

且由散度定理[311]，上式可写为

$$\int \boldsymbol{\nabla}\cdot\mathbf{j}\,\mathrm{d}V + \int \frac{\partial\rho}{\partial t}\,\mathrm{d}V = 0 \tag{3}$$

由于该体积分对任何闭合曲面都成立，所以

$$\boldsymbol{\nabla}\cdot\mathbf{j} + \frac{\partial\rho}{\partial t} = 0 \tag{4}$$

电多极展开

预备知识 球谐函数[761]

若空间中的一个球内（$r<a$）存在静止的电荷分布 $\rho(\mathbf{r})$，那么球外的电势 $V(\mathbf{r})$（$r>a$）可以展开为径向函数和球谐函数之积的形式

$$V(\mathbf{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{C_{l,m}}{r^{l+1}}Y_{l,m}(\hat{\mathbf{r}}) \tag{1}$$

其中常数 $C_{l,m}$ 为

$$C_{l,m}=\frac{1}{(2l+1)\epsilon_0}\int\rho(\mathbf{r})r^lY^*_{l,m}(\hat{\mathbf{r}})\,\mathrm{d}^3r \tag{2}$$

当 $m=0$ 时，$l=1$ 的项就是电单极子（各向同性），$l=2$ 的项就是电偶极子[1141]（$\propto\cos\theta$），$l=N$ 的项叫做电 N 极子.

推导

首先我们给出单个点电荷势能的展开公式

$$\frac{1}{|\mathbf{r}_2-\mathbf{r}_1|}=4\pi\sum_{l=0}^{\infty}\frac{1}{2l+1}\frac{r_<^l}{r_>^{l+1}}\sum_{m=-l}^{l}Y^*_{l,m}(\hat{\mathbf{r}}_1)Y_{l,m}(\hat{\mathbf{r}}_2) \tag{3}$$

其中 $r_<:=\min\{r_1,r_2\}$，$r_>:=\max\{r_1,r_2\}$. 这就是为什么我们要求电荷必须在 $r<a$ 的球内而 $V(\mathbf{r})$ 必须在球外计算.

然后根据库仑势能有

$$V(\mathbf{r})=\frac{1}{4\pi\epsilon_0}\int\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}^3r' \tag{4}$$

令式 3 中 $r_<=r'$，$r_>=r$，可得式 1.

电磁场标势和矢势

- 本词条处于草稿阶段.

预备知识 法拉第电磁感应[1191]，磁矢势[1215]

用标势和矢势表示电磁场

$$\mathbf{E}=-\boldsymbol{\nabla}\varphi-\frac{\partial\mathbf{A}}{\partial t} \tag{1}$$

$$\mathbf{B}=\boldsymbol{\nabla}\times\mathbf{A} \tag{2}$$

推导

首先定义 $\mathbf{A}$，则由法拉第电磁感应定律（式 2 [1223]）

$$\boldsymbol{\nabla} \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = \boldsymbol{\nabla} \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = \mathbf{0} \tag{3}$$

这说明括号中的矢量可以表示为一个标量函数的梯度，即标势 φ，负号是为了在静电场的情况下使得标势等于电势.

标势和矢势的麦克斯韦方程组

将式 1 和式 2 代入麦克斯韦方程组可以得到两条与麦克斯韦方程组等效的方程

$$\nabla^2 \varphi + \frac{\partial}{\partial t}(\boldsymbol{\nabla} \cdot \mathbf{A}) = -\frac{\rho}{\epsilon_0} \tag{4}$$

$$\left(\nabla^2 \mathbf{A} - \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{A}}{\partial t^2} \right) - \boldsymbol{\nabla} \left(\boldsymbol{\nabla} \cdot \mathbf{A} + \mu_0 \epsilon_0 \frac{\partial \varphi}{\partial t} \right) = -\mu_0 \mathbf{J} \tag{5}$$

磁标势

磁标势推导

$$\oint_L \mathbf{H} \cdot \mathrm{d}l = \int_S \mathbf{J} \cdot \mathrm{d}S \tag{1}$$

当 S 内无电流亦环路 L 内无电流时

$$\oint_L \mathbf{H} \cdot \mathrm{d}\mathbf{l} = 0 \tag{2}$$

在该区域，有 $\mathbf{J} = 0$，则在该区域内，磁场方程为

$$\boldsymbol{\nabla} \times \mathbf{H} = 0 \tag{3}$$

$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0 \tag{4}$$

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) = f(\mathbf{H}) \tag{5}$$

式 5 的写法为函数形式因为在如铁磁性物质中，线性关系 $\mathbf{B}=\mu_0\mathbf{H}$ 不成立. 而 $\mathbf{H}$ 与 $\mathbf{B}$ 的关系可以由磁滞回线确定. 把式 5 代入式 4 得

$$\boldsymbol{\nabla}\cdot\mathbf{H}=-\boldsymbol{\nabla}\cdot\mathbf{M} \tag{6}$$

把分子电流看作由一对假想的磁荷组成的磁偶极子，则和电场中的 $\boldsymbol{\nabla}\cdot\mathbf{P}=-\rho_p$ 对应.

$$\rho_m=-\mu_0\boldsymbol{\nabla}\cdot\mathbf{M} \tag{7}$$

因而在 $\mathbf{J}=0$ 区域内，由式 3 式 6 式 7 开始微分方程可以写为

$$\boldsymbol{\nabla}\times\mathbf{H}=0 \tag{8}$$

$$\boldsymbol{\nabla}\cdot\mathbf{H}=\frac{\rho_m}{\mu_0} \tag{9}$$

对比与静电场微分方程

$$\boldsymbol{\nabla}\cdot\mathbf{E}=\frac{\rho_f+\rho_p}{\epsilon} \tag{10}$$

$$\boldsymbol{\nabla}\times\mathbf{E}=0 \tag{11}$$

故而可以引入磁场标势

$$\mathbf{H}=-\boldsymbol{\nabla}\cdot\varphi_m \tag{12}$$

磁矢势

- 本词条处于草稿阶段.

预备知识 旋度的逆运算[325]

由于磁场 $\mathbf{B}(\mathbf{r})$ 任何情况都是一个无旋场[1169]，所以根据定理 1 [325] 必定存在一个矢量场 $\mathbf{A}(\mathbf{r})$ 使得

$$\boldsymbol{\nabla}\times\mathbf{A}=\mathbf{B} \tag{1}$$

且 $\mathbf{A}$ 可以通过下式计算

$$\mathbf{A}(\mathbf{r})=\frac{1}{4\pi}\int\mathbf{B}(\mathbf{r}')\times\frac{\mathbf{R}}{R^3}\,\mathrm{d}V'+\mathbf{H}(\mathbf{r}) \tag{2}$$

其中 $\mathbf{r},\mathbf{r}'$ 分别是坐标原点指向三维直角坐标 (x,y,z) 和 (x',y',z') 的位置矢量，$\mathbf{R}=\mathbf{r}'-\mathbf{r}$，$R=|\mathbf{R}|$，体积分 $\int\mathrm{d}V'=\int\mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z'$ 的区域是空间中 $\mathbf{B}$ 不为零的区域，$\times$ 表示矢量叉乘[387]，$\mathbf{H}(\mathbf{r})$ 是一个任意无旋场.

规范

（详见“规范变换[1217]”）由于 $\mathbf{A}(\mathbf{r})$ 不止一种，我们有时候需要某种**规范（gauge）**来将其唯一确定下来．例如在**库伦规范（Coulomb Gauge）**中，我们要求

$$\boldsymbol{\nabla}\cdot\mathbf{A}=\mathbf{0} \tag{3}$$

根据式 2 [325]，我们只需要令 $\mathbf{H}(\mathbf{r})$ 是一个调和场即可，事实上库伦规范直接规定 $\mathbf{H}(\mathbf{r})=0$.

磁多极矩

预备知识 电多极展开[1212]

引理 1　分子电流观点

对于物质磁性的解释，把每个宏观体积元内的分子看成完全一样的电流环，及具有同样的面积 a 和取向（由面元矢量 $\mathbf{a}$ 表示），环内具有同样的电流 I，而磁性由分子电流引发．故而后面式 4 可以把电流密度的体积分转变为电流的环路积分．

矢势的多级展开

磁场矢势在空间中定义为

$$\mathbf{A}(\mathbf{x})=\frac{\mu_0}{4\pi}\int_v\frac{\mathbf{J}'(\mathbf{x})\mathrm{d}V'}{R} \tag{1}$$

势能零点在无穷远点，而电流分布在小区域 V 内，则，可以对体积 V 内的 $\mathbf{A}(\mathbf{x})$ 进行多级泰勒展开（R 为 $\mathbf{x}$ 到体积 V 内的势能零点的距离）

$$\mathbf{A}(\mathbf{x})=\frac{\mu_0}{4\pi}\int_V\mathbf{J}'(\mathbf{x})\left[\frac{1}{R}-\mathbf{x}'\cdot\boldsymbol{\nabla}\cdot\frac{1}{R}+\frac{1}{2!}\sum_{i,j}x_i'x_j'\frac{\partial^2}{\partial x_i\partial x_j}\frac{1}{R}+\cdots\right]\mathrm{d}\mathbf{V}' \tag{2}$$

式 2 中，第一项

$$\mathbf{A}^{(0)}(\mathbf{x})=\frac{\mu_0}{4\pi R}\int_V\mathbf{R}J'(\mathbf{x})\,\mathrm{d}V' \tag{3}$$

由于电流的连续性，把电流分为许多闭合的流管，对于每一个流管来说，有

$$\int_V \mathbf{J}(\mathbf{x}')\,\mathrm{d}V' = \oint_L I\,\mathrm{d}l = I\oint_L \mathrm{d}l = 0 \tag{4}$$

得

$$\mathbf{A}^{(0)} = 0 \tag{5}$$

对展开式第二项

$$\mathbf{A}^{(1)} = -\frac{\mu_0}{4\pi}\int_V \mathbf{J}(\mathbf{x}')\mathbf{x}'\cdot\boldsymbol{\nabla}\cdot\frac{1}{R}\,\mathrm{d}V' \tag{6}$$

由式 4 得

$$\mathbf{A}^{(1)} = -\frac{\mu_0 I}{4\pi}\int_l \mathbf{x}'\cdot\boldsymbol{\nabla}\cdot\frac{1}{R}\,\mathrm{d}V' = \frac{\mu_0 I}{4\pi}\int_l \mathbf{x}'\cdot\frac{\mathbf{R}}{R^3}\,\mathrm{d}l' \tag{7}$$

其中有 $\mathrm{d}\mathbf{x}' = \mathrm{d}l'$，因为 $\mathbf{x}'$ 为闭合流管上的坐标．全微分绕闭合回路积分为 0

$$0 = \oint_L \mathrm{d}(\mathbf{x}'\cdot\mathbf{R})\mathbf{x}' = \oint_L (\mathbf{x}'\cdot\mathbf{R})\mathrm{d}l' + \oint_l (\mathrm{d}l'\cdot\mathbf{R})\mathbf{x}' \tag{8}$$

而

$$\oint_L (\mathbf{x}'\cdot\mathbf{R})\,\mathrm{d}l' = \frac{1}{2}\oint_L (\mathbf{x}'\times\mathrm{d}l')\times\mathbf{R} \tag{9}$$

$$\mathbf{A}^{(1)} = \frac{\mu_0}{4\pi R^3}\cdot\frac{I}{2}\oint_L (\mathbf{x}'\times\mathrm{d}l')\times\mathbf{R} = \frac{\mu_0}{4\pi}\frac{\mathbf{m}\times\mathbf{R}}{R^3} \tag{10}$$

其中 $\mathbf{m} = \frac{1}{2}\oint_L(\mathbf{x}'\times\mathrm{d}l')$ 被称为磁矩，而对于一个小线圈（分子电流），他所围的面元 $\Delta\mathbf{S}$ 可以表示为

$$\Delta\mathbf{S} = \frac{1}{2}\oint_L \mathbf{x}'\times\mathrm{d}\mathbf{l}' \tag{11}$$

故而，有等式

$$\mathbf{m} = I\Delta\mathbf{S} \tag{12}$$

规范变换

预备知识 电磁场标势和矢势[1213]

虽然标势和矢势可以唯一确定电磁场，但是同一个电磁场却可以由不同的标势和矢势得到.

首先给矢势（式 2）加上某个标量函数的梯度，磁场不变.

$$\mathbf{B}' = \boldsymbol{\nabla} \times \mathbf{A}' = \boldsymbol{\nabla} \times (\mathbf{A} + \boldsymbol{\nabla}\lambda) = \boldsymbol{\nabla} \times \mathbf{A} = \mathbf{B} \tag{1}$$

但如果 λ 随时间变化，式 2 [1213] 中的电场会改变. 我们只需把标势减去 λ 关于时间的导数即可使电场不变

$$\mathbf{E}' = -\boldsymbol{\nabla}\varphi' - \frac{\partial \mathbf{A}'}{\partial t} = -\boldsymbol{\nabla}\left(\varphi - \frac{\partial \lambda}{\partial t}\right) - \frac{\partial}{\partial t}(\mathbf{A} + \boldsymbol{\nabla}\lambda) = \mathbf{E} \tag{2}$$

所以标势和矢势的**规范变换**为

$$\begin{cases} \mathbf{A}' = \mathbf{A} + \boldsymbol{\nabla}\lambda \\ \varphi' = \varphi - \dfrac{\partial \lambda}{\partial t} \end{cases} \tag{3}$$

任何产生相同电磁场的标势矢势都可以通过规范变换联系起来.

常见的两种规范是**库仑规范**[1219] 和**洛伦兹规范**[1218].

洛伦兹规范

- 本词条处于草稿阶段.

预备知识 规范变换[1217]

如果令

$$\boldsymbol{\nabla} \cdot \mathbf{A} = -\mu_0 \epsilon_0 \frac{\partial \varphi}{\partial t} \tag{1}$$

那么标势和矢势就符合**洛伦兹规范**.

麦克斯韦方程组（式 4 [1213] 式 5 [1213]）将变为十分对称的形式

$$\nabla^2 \varphi - \mu_0 \epsilon_0 \frac{\partial^2 \varphi}{\partial t^2} = -\frac{\rho}{\epsilon_0} \tag{2}$$

$$\nabla^2 \mathbf{A} - \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \mathbf{j} \tag{3}$$

库仑规范

- 本词条处于草稿阶段.

预备知识 规范变换[1217]

当我们选择 λ 使得 $\boldsymbol{\nabla}\cdot\mathbf{A}=0$ 时，就得到了**库仑规范**. 库仑规范下，标势和矢势的麦克斯韦方程组（式 4 [1213] 式 5 [1213]）化简为（真空中的光速为 $c=1/\sqrt{\mu_0\epsilon_0}$）

$$\nabla^2\varphi=-\frac{\rho}{\epsilon_0} \tag{1}$$

$$\nabla^2\mathbf{A}-\frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2}=-\mu_0\mathbf{j}+\frac{1}{c^2}\boldsymbol{\nabla}\left(\frac{\partial\varphi}{\partial t}\right) \tag{2}$$

其中式 1 的形式与静止电荷分布的泊松方程形式一样，但同样适用于变化的电荷. 这看起来似乎是瞬时作用，但由于标势和矢势都只是数学上的量而不是物理上存在的量，所以是完全正确的.

当空间中没有电荷只有电磁场时，以上两式进一步化简为

$$\nabla^2\varphi=0 \tag{3}$$

$$\nabla^2\mathbf{A}-\frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2}=\frac{1}{c^2}\boldsymbol{\nabla}\left(\frac{\partial\varphi}{\partial t}\right) \tag{4}$$

若假设无穷远处没有电荷电流也没有电场磁场，那么无穷远处的标势矢势需满足

$$\nabla^2\varphi=0\qquad\boldsymbol{\nabla}\varphi+\frac{\partial\mathbf{A}}{\partial t}=0 \tag{5}$$

以及 $\mathbf{A}$ 为调和场. 任何满足 $\nabla^2\lambda=0$ 的规范变换都能满足这些条件. 可见只有 $\boldsymbol{\nabla}\cdot\mathbf{A}=0$ 并不能唯一确定标势矢势，还需要一定的边界条件. 库仑规范的另一个条件是：令标势的边界条件为无穷远处 $\varphi=0$，于是标量势可以唯一确定为

$$\varphi(\mathbf{r},t)=\frac{1}{4\pi\epsilon_0}\int\frac{\rho(\mathbf{r}',t)}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}^3r' \tag{6}$$

无源的情况

如果空间中没有电荷和电流（称为**无源**），那么 φ 处处为零．由此得到一个常见的结论是

$$\boldsymbol{\mathcal{E}}(t)=-\frac{\partial \mathbf{A}}{\partial t} \tag{7}$$

矢势的波动方程也变得非常简单

$$\nabla^2\mathbf{A}-\frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2}=0 \tag{8}$$

该方程的通解（先不要求 $\boldsymbol{\nabla}\cdot\mathbf{A}=0$）就是由任意极化方向[1] $\hat{\boldsymbol{\varepsilon}}$ 和传播方向的光速平面波

$$\hat{\boldsymbol{\varepsilon}}\cos(\mathbf{k}\cdot\mathbf{r}-\omega t+\phi)\qquad(\omega=ck) \tag{9}$$

的线性组合．再加上 $\boldsymbol{\nabla}\cdot\mathbf{A}=0$ 条件，得 $\hat{\boldsymbol{\varepsilon}}\cdot\mathbf{k}=0$，即极化方向与传播方向垂直．

电磁场的能量守恒 坡印廷矢量

预备知识 麦克斯韦方程组[1223]，电场的能量[1169]，磁场的能量[1183]

结论

坡印廷矢量

真空中电磁场的能流密度（即单位面积单位时间通过某个垂直的微小截面的能量）为

$$\mathbf{s}=\frac{1}{\mu_0}\mathbf{E}\times\mathbf{B} \tag{1}$$

$\mathbf{s}$ 就是坡印廷矢量．

例 1　平面电磁波的能流密度

我们知道真空中的平面电磁波[1230] 的电场和磁场方向垂直，所以二者叉乘

[1] 这里指 $\mathbf{A}$ 的方向

的模长等于他们各自的模长相乘．叉乘的方向就是电磁波传播的方向．考虑到每个位置的磁场 $|B(\mathbf{r})| = |E(\mathbf{r})|/c$，再利用式 4 [1230] 得

$$s = \frac{1}{\mu_0 c}|\mathbf{E}|^2 = \epsilon_0 c|\mathbf{E}|^2 \tag{2}$$

但注意在任意一点处，这个值是随时间以 $\sin^2$ 波动的，在波节处，能流密度为零，而在波峰处为最大，平均值等于 $1/2$．所以任意一点的平均能流密度为

$$\langle s\rangle = \frac{1}{2}\epsilon_0 c|\mathbf{E}_{max}|^2 \tag{3}$$

电磁场能量守恒积分形式

$$\int_V \frac{\mathrm{d}w}{\mathrm{d}t}\,\mathrm{d}V + \frac{\partial}{\partial t}\int_V \rho_E\,\mathrm{d}V + \oint_\Omega \mathbf{s}\cdot\mathrm{d}\mathbf{a} = 0 \tag{4}$$

选取任意的一个闭合曲面 Ω，内部空间记为 V，以下三者之和为零．

1. 电磁场对 V 中所有电荷做功的功率
2. V 中电磁场能量增加的速率
3. 以及通过曲面 Ω 流出的能量的速率

电磁场能量守恒微分形式

$$\frac{\mathrm{d}w}{\mathrm{d}t} + \frac{\partial\rho}{\partial t} + \boldsymbol{\nabla}\cdot\mathbf{s} = 0 \tag{5}$$

空间中选取任意一点，以下三者之和为零．

1. 电磁场对电荷的功率密度
2. 电磁场能量密度增量
3. 能流密度散度

推导

类比电流的连续性方程式 4 [1212]（即电荷守恒），若电磁场不对电荷做功，能量守恒可以写成

未完成：ρ_E 的表达式引用一下

$$\frac{\partial \rho_E}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{s} = 0 \tag{6}$$

的形式．其中 $\mathbf{s}$ 是电磁场的能流密度（也叫**坡印廷矢量**）（参考流密度）．但若再考虑上电磁场对电荷做功，则还需要加上一项做功做功功率密度 $\partial w/\partial t$，即单位时间单位体积电磁场对电荷做的功）．

$$\frac{\partial w}{\partial t} + \frac{\partial \rho_E}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{s} = 0 \tag{7}$$

第一项中电磁场对电荷做功即广义洛伦兹力做功（功率密度）

$$\frac{\partial w}{\partial t} = \mathbf{f} \cdot \mathbf{v} = \rho(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \mathbf{v} = \mathbf{j} \cdot \mathbf{E} \tag{8}$$

假设电磁场的能量守恒式 7 成立，那么 $\mathbf{j} \cdot \mathbf{E} = -\partial \rho_E/\partial t - \boldsymbol{\nabla} \cdot \mathbf{s}$．等式右边只与场有关，所以应该把电流密度 $\mathbf{j}$ 用麦克斯韦方程组替换成场的表达式，即

$$\mathbf{j} = \frac{1}{\mu_0} \boldsymbol{\nabla} \times \mathbf{B} - \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \tag{9}$$

代入得

$$\begin{aligned} \frac{\partial w}{\partial t} &= \left(\frac{1}{\mu_0} \boldsymbol{\nabla} \times \mathbf{B} - \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right) \cdot \mathbf{E} \\ &= \frac{1}{\mu_0} (\boldsymbol{\nabla} \times \mathbf{B}) \mathbf{E} - \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \mathbf{E} \end{aligned} \tag{10}$$

式 7 第二项中，ρ_E 是电场能量密度和磁场能量密度之和，即

$$\begin{aligned} \boldsymbol{\nabla} \times (\mathbf{B} \times \hat{\mathbf{x}}) &= \mathbf{B}(\boldsymbol{\nabla} \cdot \hat{\mathbf{x}}) + (x \boldsymbol{\nabla}) \cdot \mathbf{B} - \hat{\mathbf{x}} (\boldsymbol{\nabla} \cdot \mathbf{B}) - (\mathbf{B} \cdot \boldsymbol{\nabla}) \hat{\mathbf{x}} \\ &= (x \boldsymbol{\nabla}) \cdot \mathbf{B} = \frac{\partial \mathbf{B}}{\partial x} \end{aligned} \tag{11}$$

现在我们可以把式 10，式 11 代入式 7 中，求出 $\boldsymbol{\nabla} \cdot \mathbf{s}$.

$$\begin{aligned} \boldsymbol{\nabla} \cdot \mathbf{s} &= -\frac{\partial w}{\partial t} - \frac{\partial \rho_E}{\partial t} \\ &= -\frac{1}{\mu_0} (\boldsymbol{\nabla} \times \mathbf{B}) \mathbf{E} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \mathbf{E} - \frac{\partial}{\partial t} \left(\frac{1}{2} \epsilon_0 \mathbf{E}^2 + \frac{1}{2} \frac{\mathbf{B}^2}{\mu_0} \right) \\ &= -\frac{1}{\mu_0} (\boldsymbol{\nabla} \times \mathbf{B}) \mathbf{E} - \frac{1}{\mu_0} \frac{\partial \mathbf{B}}{\partial t} \mathbf{B} \end{aligned} \tag{12}$$

其中 $(\nabla \times \mathbf{B}) \cdot \mathbf{E} = (\nabla \times \mathbf{E}) \cdot \mathbf{B} - \nabla \cdot (\mathbf{E} \times \mathbf{B})$，因为 $\nabla \cdot (\mathbf{E} \times \mathbf{B}) = (\nabla \times \mathbf{E}) \cdot \mathbf{B} - (\nabla \times \mathbf{B}) \cdot \mathbf{E}$（Gibbs 算子相关公式）．代入得

$$\begin{aligned} \nabla \cdot \mathbf{s} &= -\frac{\partial w}{\partial t} - \frac{\partial \rho_E}{\partial t} \\ &= -\frac{1}{\mu_0}(\nabla \times \mathbf{B})\mathbf{E} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}\mathbf{E} - \frac{\partial}{\partial t}\left(\frac{1}{2}\epsilon_0 \mathbf{E}^2 + \frac{1}{2}\frac{\mathbf{B}^2}{\mu_0}\right) \\ &= -\frac{1}{\mu_0}(\nabla \times \mathbf{E}) \cdot \mathbf{B} - \frac{1}{\mu_0}\frac{\partial \mathbf{B}}{\partial t}\mathbf{B} + \frac{1}{\mu_0}\nabla \cdot (\mathbf{E} \times \mathbf{B}) \end{aligned} \tag{13}$$

其中 $\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t$，代入得

$$\nabla \cdot \mathbf{s} = \nabla \cdot \left(\frac{1}{\mu_0}\mathbf{E} \times \mathbf{B}\right) \tag{14}$$

即

$$\mathbf{s} = \frac{1}{\mu_0}\mathbf{E} \times \mathbf{B} \tag{15}$$

这就是电磁场的能流密度．

事实上，给 $\mathbf{s}$ 再加上任意一个散度为零的场，式 14 都能满足，但为了简洁起见，一般写成式 15．

麦克斯韦方程组

- 本词条处于草稿阶段．

[2]麦克斯韦方程组描述了经典电磁理论中电荷如何影响电磁场，以及电磁场变化的规律．它有多种不同的表示方法，本词条中会分别讨论．

微分形式

麦克斯韦方程组共有四条方程

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \tag{1}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{2}$$

[2]参考 Wikipedia 相关页面．

$$\nabla \cdot \mathbf{B} = 0 \tag{3}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \tag{4}$$

其中式 1 到式 4 分别是电场的高斯定律证明[1166]，法拉第电磁感应定律[1191]，磁场的高斯定律[1169]，安培环路定理[1179]（加位移电流）. 注意电场和磁场不是完全对称的，可以通过引入磁单极子[1226] 的概念使它们完全对称.

积分形式

$$\oint \mathbf{E} \cdot \mathrm{d}\mathbf{s} = \frac{1}{\epsilon_0} \int \rho \, \mathrm{d}V \tag{5}$$

$$\oint \mathbf{E} \cdot \mathrm{d}\mathbf{l} = -\int \frac{\partial \mathbf{B}}{\partial t} \cdot \mathrm{d}\mathbf{s} \tag{6}$$

$$\oint \mathbf{B} \cdot \mathrm{d}\mathbf{s} = 0 \tag{7}$$

$$\oint \mathbf{B} \cdot \mathrm{d}\mathbf{l} = \mu_0 \int \mathbf{j} \cdot \mathrm{d}\mathbf{s} + \mu_0 \epsilon_0 \int \frac{\partial \mathbf{E}}{\partial t} \cdot \mathrm{d}\mathbf{s} \tag{8}$$

高斯单位制

高斯单位制[1976] 中的麦克斯韦方程组具有更为对称的形式

$$\nabla \cdot \mathbf{E} = 4\pi\rho \tag{9}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \tag{10}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{11}$$

$$\nabla \times \mathbf{B} = \frac{4\pi}{c}\mathbf{j} + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} \tag{12}$$

麦克斯韦方程组（介质）

预备知识 麦克斯韦方程组[1223]

$$\begin{aligned}
\boldsymbol{\nabla}\cdot\mathbf{D} &= \rho_f \\
\boldsymbol{\nabla}\times\mathbf{E} &= -\frac{\partial\mathbf{B}}{\partial t} \\
\boldsymbol{\nabla}\cdot\mathbf{B} &= 0 \\
\boldsymbol{\nabla}\times\mathbf{H} &= J_f + \frac{\partial\mathbf{D}}{\partial t}
\end{aligned} \tag{1}$$

其中电位移矢量 $\mathbf{D}=\epsilon_0\mathbf{E}+\mathbf{P}$，磁场强度 $\mathbf{H}=\frac{\mathbf{B}}{\mu_0}-\mathbf{M}$.

在各向同性线性介质中，有 $\mathbf{P}=\chi_E\epsilon_0\mathbf{E}$，$\mathbf{M}=\chi_B\mathbf{H}$. 代入上式得 $\mathbf{D}=(1+\chi_E)\epsilon_0\mathbf{E}$ 和 $\mathbf{H}=\frac{\mathbf{B}}{(1+\chi_B)\mu_0}$.

定义相对介电常数为 $\epsilon_r=1+\chi_E$，相对磁导率为 $\mu_r=1+\chi_B$，则 $\mathbf{D}=\epsilon_r\epsilon_0\mathbf{E}$，$\mathbf{H}=\frac{\mathbf{B}}{\mu_r\mu_0}$，

电磁场推迟势

预备知识 洛伦兹规范[1218]

[3]在静电学情况下（$\rho(\mathbf{r})$ 和 $\mathbf{j}(\mathbf{r})$ 不随时间变化），标势和矢势也与时间无关，这时它们满足（式 2 [1218] 式 3 [1218] 中时间导数项为零）

$$\nabla^2\varphi=-\frac{\rho}{\epsilon_0} \tag{1}$$

$$\nabla^2\mathbf{A}=-\mu_0\mathbf{j} \tag{2}$$

解的

$$V(\mathbf{r})=\frac{1}{4\pi\epsilon_0}\int\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}V' \tag{3}$$

$$\mathbf{A}(\mathbf{r})=\frac{\mu_0}{4\pi}\int\frac{\mathbf{j}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}V' \tag{4}$$

在非静电学情况下，电荷密度 $\rho(\mathbf{r},t)$ 和电流 $\mathbf{j}(\mathbf{r},t)$ 既是空间的函数也是时间的函数. 定义**推迟时间（retarded time）**为

$$t_r\equiv t-|\mathbf{r}-\mathbf{r}'|/c \tag{5}$$

[3]参考 [12].

那么可以证明（式 2 [1218] 式 3 [1218] 的解）

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\rho(\mathbf{r}', t_r)}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}V' \tag{6}$$

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{j}(\mathbf{r}', t_r)}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}V' \tag{7}$$

磁单极子

- 本词条处于草稿阶段.

预备知识 麦克斯韦方程组[1223]

[4]若定义磁荷 q_m 的单位为 Am（安培 • 米），ρ_m 为磁荷的体密度，则

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \frac{\rho}{\epsilon_0} \tag{1}$$

$$\boldsymbol{\nabla}\times\mathbf{E} = -\mu_0\mathbf{j}_m - \frac{\partial\mathbf{B}}{\partial t} \tag{2}$$

$$\boldsymbol{\nabla}\cdot\mathbf{B} = \mu_0\rho_m \tag{3}$$

$$\boldsymbol{\nabla}\times\mathbf{B} = \mu_0\mathbf{j} + \mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t} \tag{4}$$

洛伦兹力

$$\mathbf{F} = q\left(\mathbf{E} + \frac{\mathbf{v}}{c}\times\mathbf{B}\right) \tag{5}$$

高斯单位

高斯单位制[1976] 下，麦克斯韦方程组和洛伦兹力具有更对称的形式

$$\boldsymbol{\nabla}\cdot\mathbf{E} = 4\pi\rho \tag{6}$$

$$\boldsymbol{\nabla}\times\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t} - \frac{4\pi}{c}\mathbf{j}_m \tag{7}$$

$$\boldsymbol{\nabla}\cdot\mathbf{B} = 4\pi\rho_m \tag{8}$$

[4]参考 Wikipedia 相关页面.

$$\boldsymbol{\nabla}\times\mathbf{B}=\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}+\frac{4\pi}{c}\mathbf{j} \tag{9}$$

洛伦兹力

$$\mathbf{F}=q\left(\mathbf{E}+\frac{\mathbf{v}}{c}\times\mathbf{B}\right)+q_m\left(\mathbf{B}-\frac{\mathbf{v}}{c}\times\mathbf{E}\right) \tag{10}$$

非齐次亥姆霍兹方程 推迟势

- 本词条处于草稿阶段.

预备知识 格林函数

对波动方程

$$\nabla^2u-\frac{1}{c^2}\frac{\partial^2u}{\partial t^2}=f(\mathbf{r},t) \tag{1}$$

用格林函数法求解，令格林函数满足

$$\nabla^2G-\frac{1}{c^2}\frac{\partial^2G}{\partial t^2}=-\delta(\mathbf{r}-\mathbf{r}')\delta(t-t') \tag{2}$$

这个方程的意义是，如果在时刻 t，$\mathbf{r}$ 处出现一个极短的脉冲，会生成怎样的波函数 G. 求出 G 以后，我们可以把非齐次项 $f(\mathbf{r},t)$ 看成由许许多多这样的脉冲组成，对空间和时间进行积分. 即可得到式 1 的解.

$$u(\mathbf{r},t)=\iint f(\mathbf{r}',t')G(\mathbf{r},t,\mathbf{r}',t')\,\mathrm{d}V'\,\mathrm{d}t' \tag{3}$$

其中 $\int\mathrm{d}V'$ 表示对 $\mathbf{r}'$ 进行体积分. 下文用傅里叶变换法解式 2 解出格林函数为

$$G(\mathbf{r},t,\mathbf{r}',t')=\frac{1}{4\pi}\frac{\delta[(t-|\mathbf{r}-\mathbf{r}'|/c)-t']}{|\mathbf{r}-\mathbf{r}'|} \tag{4}$$

式 4，式 3 得式 1 的解为

$$u(\mathbf{r},t)=\frac{1}{4\pi}\int\frac{f(\mathbf{r}',t-|\mathbf{r}-\mathbf{r}'|/c)}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}V' \tag{5}$$

式 5 与静电场的势能公式很像，但是场源的时间做了修正. 其中 $t-|\mathbf{r}-\mathbf{r}'|/c\equiv t_{ret}$ 定义为**推迟时间（retarded time）**，$|\mathbf{r}-\mathbf{r}'|/c$ 是波从 $\mathbf{r}'$ 到 $\mathbf{r}$ 所需的时间. 也就是说，$f(\mathbf{r},t)$ 并不能马上影响 $u(\mathbf{r},t)$，而是需要一个“信号传播时间”才行. 在静电场中，由于场源不随时间变化，所以不需要考虑时间延迟.

具体过程

解格林函数

从物理意义上，要求格林函数在以 $\mathbf{r}'$ 为中心的任意方向都相同，即只是 $|\mathbf{r}-\mathbf{r}'|$ 的函数．以下为了方便，令 $R=|\mathbf{r}-\mathbf{r}'|$．由于等式右边除了原点外都是零，方程为齐次方程．齐次解为平面波

$$G(\mathbf{r},t)=\sum A(\mathbf{k},\omega)\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\mathrm{e}^{\mathrm{i}\omega t} \tag{6}$$

求和是对所有满足 $\omega/k=c$ 和边界条件的 ω 和 $\mathbf{k}$ 求和（或积分）．但显然该解在原点不满足要求．为了排除原点，且满足对称性，以 $\mathbf{r}'$ 为原点建立球坐标，改用球坐标中的拉普拉斯方程，方程变为

$$\frac{1}{R^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(R^2\frac{\mathrm{d}G}{\mathrm{d}R}\right)-\frac{1}{c^2}\frac{\mathrm{d}^2G}{\mathrm{d}t^2}=-\delta(\mathbf{r}-\mathbf{r}')\delta(t-t') \tag{7}$$

在 $R\neq 0$ 的条件下解齐次方程，首先分离变量，得到分离变量解（提示：$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}G}{\mathrm{d}r}\right)=\frac{1}{r}\frac{\mathrm{d}^2(rG)}{\mathrm{d}r^2}$ ）

$$\frac{1}{R}\left[C_1(\omega)\mathrm{e}^{\mathrm{i}\omega R/c}+C_2(\omega)\mathrm{e}^{-\mathrm{i}\omega R/c}\right]\mathrm{e}^{\mathrm{i}\omega t} \tag{8}$$

所以通解为式 8 对不同的 ω 求和．然而 ω 是连续的，所以改用傅里叶变换法解方程式 7

$$\begin{cases} A(R,\omega)=\displaystyle\int_{-\infty}^{+\infty}G(R,t,t')\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}t \\ G(R,t,t')=\dfrac{1}{2\pi}\displaystyle\int_{-\infty}^{+\infty}A(R,\omega)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega \end{cases} \tag{9}$$

式 9 就是式 8 对连续 ω 的求和（积分）得到的通解，待定系数包含在 A 里面，下面的式 12 验证了这点．另外，式 7 右边含时 δ 函数的傅里叶变换为

$$\delta(t-t')=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{\mathrm{i}\omega t'}\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega \tag{10}$$

式 9，式 10 代入方程式 7 得到经过时间傅里叶变换的偏微分方程，与式 7 等效．

$$\nabla^2A(R,\omega)+\frac{\omega^2}{c^2}A(R,\omega)=-\mathrm{e}^{\mathrm{i}\omega t'}\delta(\mathbf{r}-\mathbf{r}') \tag{11}$$

注意这是关于位置的偏微分方程，ω 视为常数．解这条方程，就相当于解出了固定振动频率 ω 的波源所产生的同频率的波动方程．齐次解为（解方程提示：$\frac{1}{R^2}\frac{\mathrm{d}}{\mathrm{d}R}\left(R^2\frac{\mathrm{d}A}{\mathrm{d}R}\right)=\frac{1}{R}\frac{\mathrm{d}^2}{\mathrm{d}R^2}(RA)$，令 $\frac{1}{R^2}\frac{\mathrm{d}}{\mathrm{d}R}\left(R^2\frac{\mathrm{d}A}{\mathrm{d}R}\right)=\frac{1}{R}\frac{\mathrm{d}^2}{\mathrm{d}R^2}\,\mathrm{d}(RA)$ ）

$$A(R,\omega)=\frac{1}{R}\left[C_1(\omega)\mathrm{e}^{\mathrm{i}\omega R/c}+C_2(\omega)\mathrm{e}^{-\mathrm{i}\omega R/c}\right] \tag{12}$$

由于这是方程 $\mathbf{r}\neq\mathbf{r}'$ 的通解，而式 11 的右边可以看做 $\mathbf{r}=\mathbf{r}'$ 时的边界条件，接下来利用边界条件找到适合的待定系数 $C_1(\omega)$ 和 $C_2(\omega)$．首先当 $\mathbf{r}\to\mathbf{r}'$ 时，$R\to 0$，$A\to(C_1+C_2)/R$．所以

$$\nabla^2 A(R,\omega)=(C_1+C_2)\nabla^2\frac{1}{R}=-4\pi(C_1+C_2)\delta(\mathbf{r}-\mathbf{r}') \tag{13}$$

（关于 $\nabla^2\frac{1}{R}$ 见空间狄拉克 *delta* 函数，）．代入式 11 左边第一项，得

$$-4\pi(C_1+C_2)\delta(\mathbf{r}-\mathbf{r}')+\frac{\omega^2}{c^2}A(R,\omega)=-\mathrm{e}^{\mathrm{i}\omega t'}\delta(\mathbf{r}-\mathbf{r}') \tag{14}$$

由于空间 delta 函数 $\delta(\mathbf{r}-\mathbf{r}')\sim\frac{1}{R^3}$，所以相比之下 $\omega^2A(R,\omega)/c^2\sim 1/R$ 在 $R\to 0$ 时可以忽略不计．等式两边对比系数得

$$C_1+C_2=\frac{1}{4\pi}\mathrm{e}^{\mathrm{i}\omega t'} \tag{15}$$

再考虑式 12 所代表的波函数分量，第一项代表波源向外传播的球形波，第二项代表向波源传播的，所以 $C_2=0$，式 12 变为

$$A(R,\omega)=\frac{1}{R}C_1(\omega)\mathrm{e}^{\mathrm{i}\omega R/c}=\frac{1}{4\pi R}\mathrm{e}^{\mathrm{i}\omega t'}\mathrm{e}^{\mathrm{i}\omega R/c} \tag{16}$$

现在可以把上式进行反傅里叶变换式 9 得到格林函数

$$\begin{aligned}G(R,t,t')&=\frac{1}{2\pi}\int_{-\infty}^{+\infty}A(R,\omega)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{1}{4\pi R}\mathrm{e}^{\mathrm{i}\omega t'}\mathrm{e}^{\mathrm{i}\omega R/c}\delta(\mathbf{r}-\mathbf{r}')\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega\\&=\frac{1}{4\pi R}\cdot\frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{\mathrm{i}\omega(t'+R/c-t)}\,\mathrm{d}\omega=\frac{1}{4\pi R}\delta(t-R/c-t')\\&=\frac{1}{4\pi}\frac{\delta[(t-|\mathbf{r}-\mathbf{r}'|/c)-t']}{|\mathbf{r}-\mathbf{r}'|}\end{aligned} \tag{17}$$

这就是式 2 的解式 4.

电场波动方程

预备知识 麦克斯韦方程组（介质）[1224]，矢量算符公式

介质中

非线性光学中一般认为介质具有 $\mu = \mu_0$，且假设 $\nabla \cdot \mathbf{E} = 0$ 仍然成立

介质中没有自由电荷或自由电流.

类似真空情况的推导过程，有

$$\nabla \times (\nabla \times \mathbf{E}) = -\frac{\partial}{\partial t}(\nabla \times \mathbf{B}) = -\mu_0 \frac{\partial}{\partial t}(\nabla \times \mathbf{H}) = -\mu_0 \frac{\partial^2}{\partial t^2}\mathbf{D} \tag{1}$$

把电位移矢量的定义 $\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}$ 代入上式，化简为

$$\nabla \times (\nabla \times \mathbf{E}) = -\frac{\partial}{\partial t}(\nabla \times \mathbf{B}) = -\mu_0 \frac{\partial}{\partial t}(\nabla \times \mathbf{H}) = -\mu_0 \frac{\partial^2}{\partial t^2}\mathbf{D} \tag{2}$$

真空中的平面电磁波

- 本词条处于草稿阶段.

预备知识 麦克斯韦方程组[1223]

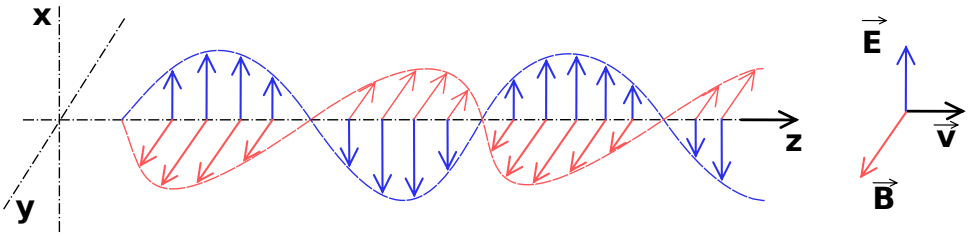


图 1: 平面电磁波的电磁场分布. 注意于电磁场矢量与 x, y 坐标无关，并占据整个空间（图片来自维基百科）

[5]平面电磁波如图 1 所示．同一位置处电场与磁感互相垂直，且模长长比例

$$|E(\mathbf{r})| = c|B(\mathbf{r})| \tag{1}$$

任意一点的能量密度为

$$\rho_E = \frac{1}{2}\left(\epsilon_0 E^2 + \frac{B^2}{\mu_0}\right) = \epsilon_0 E^2 \tag{2}$$

其中电场和磁场各贡献一般．平均能流密度（光强）为

$$I = \frac{1}{2} c\epsilon_0 E_0^2 \tag{3}$$

推导见例 1 [1220]，也可以认为瞬时能流密度等于能量密度乘以波速 c，对于简谐波，需要除以二得平均值．

波速等于真空中的光速 c，且

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} = 299,792,458\ \mathrm{m/s} \tag{4}$$

推导

（描述未完成）

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{E}) = -\boldsymbol{\nabla} \times \frac{\partial \mathbf{B}}{\partial t} = -\epsilon_0 \mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{5}$$

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{E}) = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\nabla^2 \mathbf{E} \tag{6}$$

$$\nabla^2 E = \epsilon_0 \mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{7}$$

$$E_y = E_0 \cos(\omega t - kx) \tag{8}$$

平面电磁波的能量叠加

[5]参考 [12] 相关章节．

预备知识 真空中的平面电磁波[1230]

已知单个平面电磁波的平均能流密度为 $c\epsilon_0 E_0^2/2$，那么两个同方向不同频率电磁波叠加后能流密度是多少呢？

考虑两个电磁波 $E_1\sin(k_1x-\omega_1t)$ 和 $E_2\sin(k_2x-\omega_2t)$，它们叠加后的能流密度是多少呢？先看电磁场能量密度

$$\begin{aligned}\rho_E &= \epsilon_0[E_1\sin(k_1x-\omega_1t)+E_2\sin(k_2x-\omega_2t)]^2\\ &= \epsilon_0E_1^2\sin^2(k_1x-\omega_1t)+\epsilon_0E_2^2\sin^2(k_2x-\omega_2t)\\ &\quad +2\epsilon_0E_1E_2\sin(k_1x-\omega_1t)\sin(k_2x-\omega_2t)\end{aligned}\tag{1}$$

其中前两项是每个平面波各自的能量密度，第三项是干涉项，根据式 12 [112] 可知他是两个简谐函数之和，平均值为零.

所以 N 个同方向不同频率平面波叠加后的平均能流密度等于它们各自能流密度之和

$$|\mathbf{j}| = \frac{1}{2}c\epsilon_0\sum_i E_i^2\tag{2}$$

电磁波包的能谱

预备知识 傅里叶变换（指数）[237]，平面电磁波的能量叠加[1231]

对于真空中的平面波电磁波，波速恒定为 c，如果知道某点 x_0 处的电场—时间关系 $g(t)$，如何求波函数 $f(x-ct)$ 呢？代入 $x=x_0$ 可知 $g(t)=f(x_0-ct)$，所以

$$f(x) = g\left(\frac{x_0-x}{c}\right)\tag{1}$$

当这个波包完整经过一个平面后，穿过平面的能量面密度等于能量体密度（式 2 [1230]）在传播方向的积分（积分上下限为 $\pm\infty$）

$$\sigma_E = \epsilon_0\int f(x)^2\,\mathrm{d}x = \epsilon_0\int g^2\left(\frac{x_0-x}{c}\right)\mathrm{d}x = c\epsilon_0\int g^2(u)\,\mathrm{d}u\tag{2}$$

另一种方法是把坡印廷矢量[1220] 对时间积分，同样能得到该式.

能量的频率分布

根据傅里叶变换的归一化不变性（式 15 [237]），所以若令 g 的傅里叶变换为 $\tilde{g}$ 则[6]

$$\sigma_E = c\epsilon_0 \int_{-\infty}^{+\infty} |\tilde{g}(\omega)|^2 \,\mathrm{d}\omega = 2c\epsilon_0 \int_0^{+\infty} |\tilde{g}(\omega)|^2 \,\mathrm{d}\omega \tag{3}$$

这相当于把波包看作是许多不同频率简谐波的叠加，总能量面密度是每个简谐波的能量面密度叠加．所以能量面密度的频率分布为[7]

$$s(\omega) = 2c\epsilon_0 |\tilde{g}(\omega)|^2 \tag{4}$$

考虑到光子能量为 $E = \omega\hbar$，光子能量分布为

$$s(E) = \frac{2c\epsilon_0}{\hbar} |\tilde{g}(\omega)|^2 \left(\frac{E}{\hbar}\right) \tag{5}$$

用矢势表示

在库仑规范[1219] 下，矢势为 $A(t)$ 对于波包有

$$g(t) = -\frac{\mathrm{d}A(t)}{\mathrm{d}t} \tag{6}$$

由傅里叶变换的求导公式（式 16 [237]）得

$$\tilde{g}(\omega) = -\mathrm{i}\omega \tilde{A}(\omega) \tag{7}$$

代入式 4 得

$$s(\omega) = 2c\epsilon_0 \omega^2 \left|\tilde{A}(\omega)\right|^2 \tag{8}$$

若使用高斯单位制[1976]，$s(\omega) = \omega^2 \left|\tilde{A}\right|^2 / (2\pi c)$.

电偶极子辐射

[6] 第二个等号中，由式 5 [237] 得 $|\tilde{g}(-\omega)|^2 = |\tilde{g}(\omega)|^2$，所以负半轴的积分与正半轴相等.

[7] 原子单位：$s(\omega) = c|\tilde{g}(\omega)|^2/(2\pi)$

- 本词条处于草稿阶段.

预备知识 电磁场推迟势[1225]

[8]令原点处的电偶极子为

$$\mathbf{p}(t)=p_0\cos(\omega t)\hat{\mathbf{z}} \tag{1}$$

使用洛伦兹规范，在 $r\gg\lambda$ 的近似下

$$\varphi(r,\theta,t)=-\frac{p_0\omega}{4\pi\epsilon_0 c}\left(\frac{\cos\theta}{r}\right)\sin[\omega(t-r/c)] \tag{2}$$

$$\mathbf{A}(r,\theta,t)=-\frac{\mu_0 p_0\omega}{4\pi r}\sin[\omega(t-r/c)]\hat{\mathbf{z}} \tag{3}$$

进而得

$$\mathbf{E}=-\frac{\mu_0 p_0\omega^2}{4\pi}\left(\frac{\sin\theta}{r}\right)\cos[\omega(t-r/c)]\hat{\boldsymbol{\theta}} \tag{4}$$

$$\mathbf{B}=-\frac{\mu_0 p_0\omega^2}{4\pi c}\left(\frac{\sin\theta}{r}\right)\cos[\omega(t-r/c)]\hat{\boldsymbol{\phi}} \tag{5}$$

其中用到了近似（d 是偶极子的长度，$k=2\pi/\lambda$）

$$d\ll\frac{1}{k}\ll r \tag{6}$$

辐射功率

$$\mathbf{s}(\mathbf{r},t)=\frac{1}{\mu_0}\mathbf{E}\times\mathbf{B}=\frac{\mu_0 p_0^2\omega^4}{16\pi^2 c}\frac{\sin^2\theta}{r^2}\cos^2[\omega(t-r/c)]\hat{\mathbf{r}} \tag{7}$$

时间平均值为

$$\langle\mathbf{s}\rangle=\frac{1}{\mu_0}\mathbf{E}\times\mathbf{B}=\frac{\mu_0 p_0^2\omega^4}{32\pi^2 c}\frac{\sin^2\theta}{r^2}\hat{\mathbf{r}} \tag{8}$$

总辐射功率（时间平均）为

$$\langle P\rangle=\oint\langle\mathbf{s}\rangle\cdot\mathrm{d}\mathbf{a}=\frac{\mu_0 p_0^2\omega^4}{12\pi c} \tag{9}$$

[8]参考 [12].

介质中的波动方程

对非磁介质

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{E})=-\mu_0\frac{\partial^2\mathbf{D}}{\partial t^2} \tag{1}$$

平面波时

$$\nabla^2\mathbf{E}-\mu_0\frac{\partial^2\mathbf{D}}{\partial t^2}=0 \tag{2}$$

$$\mathbf{D}=\epsilon_0\mathbf{E}+\mathbf{P}=\epsilon_0(1+\tilde{\chi})\mathbf{E}+\mathbf{P}^{NL}=\tilde{\epsilon}\mathbf{E}+\mathbf{P}^{NL} \tag{3}$$

$$\nabla^2\mathbf{E}-\mu_0\tilde{\epsilon}\frac{\partial^2\mathbf{E}}{\partial t^2}=\mu_0\frac{\partial^2\mathbf{P}^{NL}}{\partial t^2} \tag{4}$$

波浪号表示复数，$\tilde{\epsilon}$ 和 $\tilde{\chi}$ 都只是 ω 而不是场强的函数．先来看线性的情况（$\mathbf{P}^{NL}=\mathbf{0}$），例如洛伦兹模型．

$$\nabla^2\mathbf{E}-\mu_0\tilde{\epsilon}\frac{\partial^2\mathbf{E}}{\partial t^2}=\mathbf{0} \tag{5}$$

（见齐次波动方程，先做时间傅里叶变换，再解齐次亥姆霍兹方程，通解是所有平面波）然而这里的 $\tilde{k}^2=\mu_0\tilde{\epsilon}\omega$ 是复数，平面波变为指数衰减的单频单向波．以 z 方向传播 x 方向极化为例，令 $\tilde{k}=k+\kappa$

$$E_x(z,t)=\tilde{E}_{0x}\mathrm{e}^{\mathrm{i}(\tilde{k}z-\omega t)}=\tilde{E}_{0x}\mathrm{e}^{-\kappa z}\mathrm{e}^{\mathrm{i}(kz-\omega t)} \tag{6}$$

$$\tilde{k}=\omega\sqrt{\mu_0\tilde{\epsilon}}=\frac{\omega}{c}\sqrt{1+\chi^{(1)}} \tag{7}$$

通解仍然为所有可能的单频单向波的线性组合．实折射率和吸收系数定义为

$$n=\frac{ck}{\omega}=\mathrm{Re}\left[\sqrt{1+\chi^{(1)}}\right]\approx 1+\frac{1}{2}\mathrm{Re}\left[\chi^{(1)}\right] \tag{8}$$

$$\alpha=2\kappa=\frac{2\omega}{c}\mathrm{Im}\left[\sqrt{1+\chi^{(1)}}\right]\approx\frac{\omega}{c}\mathrm{Im}\left[\chi^{(1)}\right] \tag{9}$$

$$\tilde{\epsilon}=\epsilon_0(n+\mathrm{i}\alpha c/2\omega)^2 \tag{10}$$

注意区分 ϵ（permittivity），ϵ_r（dielectric constant）和 χ（susceptivility）．有时候 ϵ_r．不同的书符号可能不一样，以名称和语境为准．

导体中的电磁波

预备知识 电场波动方程[1230]

[9] 良导体的情况下，设任何净电荷消散的时间都非常快，可以认为 $\rho_f = 0$. 另外自由电流仅由自由电子在电场中运动产生，$\mathbf{j}_f = \sigma\mathbf{E}$. 代入介质中的麦克斯韦方程组得波动方程

$$\nabla^2\mathbf{E} = \epsilon\mu\frac{\partial^2\mathbf{E}}{\partial t^2} + \mu\sigma\frac{\partial\mathbf{E}}{\partial t} \tag{1}$$

设电磁场的形式为

$$\mathbf{E} = \hat{\mathbf{x}}\tilde{E}_0\mathrm{e}^{\mathrm{i}(\tilde{k}z-\omega t)} \qquad \mathbf{B} = \hat{\mathbf{y}}\tilde{B}_0\mathrm{e}^{\mathrm{i}(\tilde{k}z-\omega t)} \tag{2}$$

代入波动方程得

$$\tilde{k}^2 = \epsilon\mu\omega^2 + \mathrm{i}\sigma\mu\omega \tag{3}$$

令 $\tilde{k} = k + \mathrm{i}\kappa$，得

$$k = \omega\sqrt{\frac{\epsilon\mu}{2}}\sqrt{\sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2}+1} \qquad \kappa = \omega\sqrt{\frac{\epsilon\mu}{2}}\sqrt{\sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2}-1} \tag{4}$$

对于良导体，$\sigma \gg \epsilon\omega$，有

$$k = \kappa = \omega\sqrt{\frac{\mu\sigma}{2\omega}} \tag{5}$$

电磁场变为

$$\mathbf{E} = \hat{\mathbf{x}}\tilde{E}_0\mathrm{e}^{-\kappa z}\mathrm{e}^{\mathrm{i}(kz-\omega t)} \qquad \mathbf{B} = \hat{\mathbf{y}}\tilde{B}_0\mathrm{e}^{-\kappa z}\mathrm{e}^{\mathrm{i}(kz-\omega t)} \tag{6}$$

定义**趋肤深度（skin depth）**为

$$d = 1/\kappa \tag{7}$$

使用 $\boldsymbol{\nabla}\times\mathbf{E} = -\partial\mathbf{B}/\partial t$ （这里 $\boldsymbol{\nabla} \equiv \mathrm{i}\tilde{k}\hat{\mathbf{z}}$）

$$\tilde{B}_0 = \tilde{k}\tilde{E}_0/\omega \tag{8}$$

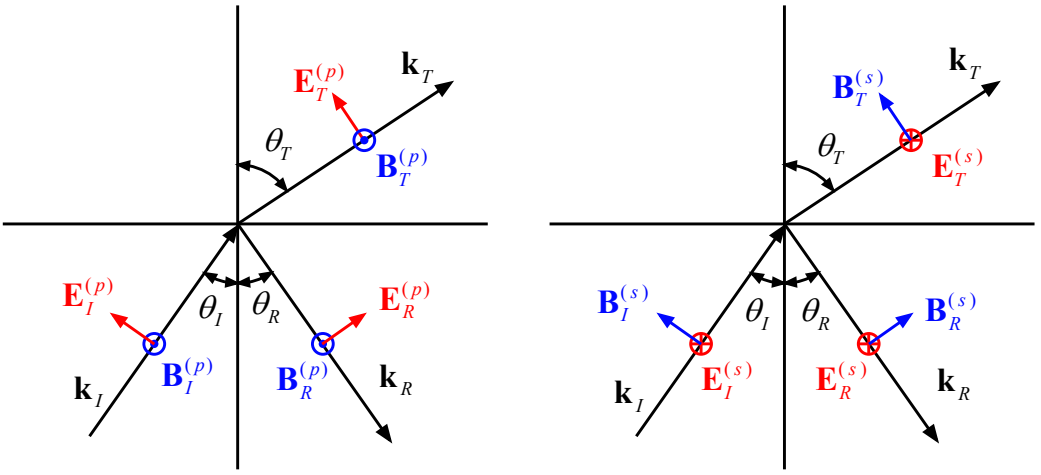
所以电场磁场存在相位差

$$\phi = \arg(\tilde{k}) \tag{9}$$

[9] 参考 David Griffiths 的电动力学导论

菲涅尔公式 布儒斯特角

预备知识 麦克斯韦方程组（介质）[1224]



$$\frac{E_R^{(p)}}{E_I^{(p)}}=\frac{B_R^{(p)}}{B_I^{(p)}}=\frac{m_2\cos\theta_i-m_1\cos\theta_T}{m_2\cos\theta_i+m_1\cos\theta_T}$$
$$\frac{E_T^{(p)}}{E_I^{(p)}}=\frac{B_T^{(p)}}{B_I^{(p)}}=\frac{2m_1\cos\theta_I}{m_2\cos\theta_i+m_1\cos\theta_T}$$
$$\frac{E_R^{(s)}}{E_I^{(s)}}=\frac{B_R^{(s)}}{B_I^{(s)}}=\frac{m_1\cos\theta_i-m_2\cos\theta_T}{m_1\cos\theta_i+m_2\cos\theta_T}$$
$$\frac{E_T^{(s)}}{E_I^{(s)}}=\frac{B_T^{(s)}}{B_I^{(s)}}=\frac{2m_1\cos\theta_I}{m_1\cos\theta_i+m_2\cos\theta_T}$$

图 1: 菲涅尔公式

利用具体的电磁场的边界条件

- $\boldsymbol{\nabla}\cdot\mathbf{D}=0$ 和 $\boldsymbol{\nabla}\cdot\mathbf{B}=0$ 分别对应 $\epsilon\mathbf{E}_\perp=\epsilon'\mathbf{E}'_\perp$ 和 $\epsilon\mathbf{B}_\perp=\epsilon'\mathbf{B}'_\perp$.
- $\boldsymbol{\nabla}\cdot\mathbf{E}=0$ 和 $\boldsymbol{\nabla}\cdot\mathbf{H}=0$ 分别对应 $\mathbf{E}_{//}=\mathbf{E}'_{//}$ 和 $\mathbf{B}_{//}/\mu=\mathbf{B}'_{//}/\mu'$.

现在分两种情况讨论

1. 极化方向垂直于入射面（图 1 右）

$$\frac{E_R^{(s)}}{E_I^{(s)}}=\frac{m_1\cos\theta_I-m_2\cos\theta_T}{m_1\cos\theta_I+m_2\cos\theta_T}\qquad\frac{E_T^{(s)}}{E_I^{(s)}}=\frac{2m_1\cos\theta_I}{m_1\cos\theta_I+m_2\cos\theta_T}\tag{1}$$

2. 极化方向平行于入射面（图 1 左）

$$\frac{E_R^{(p)}}{E_I^{(p)}}=\frac{m_2\cos\theta_I-m_1\cos\theta_T}{m_2\cos\theta_I+m_1\cos\theta_T}\qquad\frac{E_T^{(p)}}{E_I^{(p)}}=\frac{2m_1\cos\theta_I}{m_2\cos\theta_I+m_1\cos\theta_T}\tag{2}$$

其中 $m_I=n_I/\mu_I=c\sqrt{\epsilon_I/\mu_I}$，一般情况下介质的磁导率于真空区磁导率的区别可忽略，即可以把 m_I 替换为折射率 n_I．另外注意菲涅尔公式包含相位信息，即以上的 E 可以是复振幅．

布儒斯特角

我们这里考虑常见的 $n_2>n_1$ 且 $\mu_1=\mu_2$ 情况．由式 2 容易证明当入射角为**布儒斯特角（Brewster's angle）**时反射光的平行（p）分量消失．布儒斯特角等于

$$\theta_B=\arctan(n_2/n_1)\tag{3}$$

盒中的电磁波

预备知识 电场波动方程[1230]，分离变量法[742]

空间中一个电阻不计的金属盒中有电磁波．金属盒的大小为 $(0\leqslant x\leqslant a,0\leqslant y\leqslant b,0\leqslant z\leqslant c)$，电场的波动方程为

$$\nabla^2\cdot\mathbf{E}=\frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2}\tag{1}$$

矢量相等的充要条件是三个分量分别相等

$$\nabla^2E_x=\frac{1}{c^2}\frac{\partial^2E_x}{\partial t^2}\qquad\nabla^2E_y=\frac{1}{c^2}\frac{\partial^2E_x}{\partial t^2}\qquad\nabla^2E_z=\frac{1}{c^2}\frac{\partial^2E_x}{\partial t^2}\tag{2}$$

下面以 E_x 为例，用分离变量法得出通解．先令 $E_x=X(x)Y(y)Z(z)T(t)$．代入上式，两边同除 $X(x)Y(y)Z(z)T(t)$ 得

$$\frac{\mathrm{d}^2X}{\mathrm{d}x^2}\Big/X+\frac{\mathrm{d}^2Y}{\mathrm{d}y^2}\Big/Y+\frac{\mathrm{d}^2Z}{\mathrm{d}z^2}\Big/Z=\frac{1}{c^2}\frac{\mathrm{d}^2T}{\mathrm{d}t^2}\Big/T\tag{3}$$

由于上式每一项都是一个独立变量的函数，所以每一项都等于一个常数．令这些常数为

$$\begin{aligned}&\frac{\mathrm{d}^2X}{\mathrm{d}x^2}\Big/X=-k_x^2 \qquad \frac{\mathrm{d}^2Y}{\mathrm{d}y^2}\Big/Y=-k_y^2\\&\frac{\mathrm{d}^2Z}{\mathrm{d}z^2}\Big/Z=-k_z^2 \qquad \frac{1}{c^2}\frac{\mathrm{d}^2T}{\mathrm{d}t^2}\Big/T=-\omega^2\end{aligned}\tag{4}$$

（取负号是因为我们只对三角函数解感兴趣，指数函数解在这里无关）代入上式，这些常数满足

$$k_x^2+k_y^2+k_z^2=\omega^2\tag{5}$$

上面三式的通解是

$$\begin{cases}X=C_1\cos(k_xx)+C_2\sin(k_xx)\\Y=C_3\cos(k_yy)+C_4\sin(k_yy)\\Z=C_5\cos(k_zz)+C_6\sin(k_zz)\end{cases}\tag{6}$$

时间函数的解取 $T=C\cos(\omega t)$（时间函数的相位不重要）由理想导体的电磁场边界条件

$$E_{//}=0 \qquad \frac{\partial E_\perp}{\partial n}=0\tag{7}$$

$\partial E_x/\partial x=0$（$x\to a$ 时）；$E_x=0$（$y\to b$ 或 $z\to c$ 时）．把上面的通解代入条件，得

$$X=C_1\cos\Big(\frac{n_x\pi}{a}x\Big) \qquad Y=C_4\sin\Big(\frac{n_y\pi}{b}y\Big) \qquad Z=C_6\sin\Big(\frac{n_z\pi}{c}z\Big)\tag{8}$$

三个变量相乘，令 $C_1C_4C_6=E_{x0}$，得

$$E_x=E_{x0}\cos\Big(\frac{n_x\pi}{a}x\Big)\sin\Big(\frac{n_y\pi}{b}y\Big)\sin\Big(\frac{n_z\pi}{c}z\Big)\tag{9}$$

同理对 E_y，E_z 分析，得到电场的三个分量在盒内的分布

$$\begin{cases}E_x=E_{x0}\cos\Big(\frac{n_x\pi}{a}x\Big)\sin\Big(\frac{n_y\pi}{b}y\Big)\sin\Big(\frac{n_z\pi}{c}z\Big)\\E_y=E_{y0}\sin\Big(\frac{n_x\pi}{a}x\Big)\cos\Big(\frac{n_y\pi}{b}y\Big)\sin\Big(\frac{n_z\pi}{c}z\Big)\\E_z=E_{z0}\sin\Big(\frac{n_x\pi}{a}x\Big)\sin\Big(\frac{n_y\pi}{b}y\Big)\cos\Big(\frac{n_z\pi}{c}z\Big)\end{cases}\tag{10}$$

$$T=C\cos(\omega t)\tag{11}$$

且满足 $\omega=\pi\sqrt{n_x^2/a^2+n_y^2/b^2+n_z^2/c^2}$ 特殊地，当盒子是立方体的时候，$a=b=c=L$ 时，$\omega=\pi\sqrt{n_x^2+n_y^2+n_z^2}/L$.

电磁场的动量守恒 动量流密度张量

预备知识 能流密度，张量，张量的散度

结论

假设电磁场动量守恒，则动量流密度张量为

$$T_{ij}=\epsilon_0\left(\frac{1}{2}\mathbf{E}^2\delta_{ij}-E_iE_j\right)+\frac{1}{\mu_0}\left(\frac{1}{2}\mathbf{B}^2\delta_{ij}-B_iB_j\right) \tag{1}$$

该张量也成为**麦克斯韦应力张量（Maxwell Stress Tensor）**

推导

能量是标量，所以能流密度就是矢量．但动量本身就是矢量，要如何表示动量流密度呢？我们可以分析动量在某方向分量的流密度．根据张量的散度

$$(\boldsymbol{\nabla}\cdot\mathbf{T})_j=\sum_i\frac{\partial}{\partial x_i}T_{ji} \tag{2}$$

假设电磁场满足动量守恒，在闭合空间中，有“转换速率 + 流出速率 + 增加速率 = 0”（类比电磁场的能量守恒公式）．则电磁场的动量守恒会有

$$\mathbf{f}+\boldsymbol{\nabla}\cdot\mathbf{T}+\mu_0\epsilon_0\frac{\partial\mathbf{s}}{\partial t}=\mathbf{0} \tag{3}$$

由广义洛伦兹力计算电荷的受力密度 $\mathbf{f}$

$$\mathbf{f}=\rho(\mathbf{E}+\mathbf{v}\times\mathbf{B})=\rho\mathbf{E}+\mathbf{j}\times\mathbf{B}\qquad(\mathbf{j}=\rho\mathbf{v}) \tag{4}$$

由于式 3 的后两项是电磁场的量，不能含有关于电荷的量，所以接下来要通过麦克斯韦方程组把电荷密度 ρ 和电流密度 $\mathbf{j}$ 替换成电磁场．

$$\rho=\epsilon_0\boldsymbol{\nabla}\cdot\mathbf{E}\qquad\left(\boldsymbol{\nabla}\cdot\mathbf{E}=\frac{\rho}{\epsilon_0}\right) \tag{5}$$

$$\mathbf{j}=\frac{1}{\mu_0}\boldsymbol{\nabla}\times\mathbf{B}-\epsilon_0\frac{\partial\mathbf{E}}{\partial t}\qquad\left(\boldsymbol{\nabla}\times\mathbf{B}=\mu_0\mathbf{j}+\epsilon_0\mu_0\frac{\partial\mathbf{E}}{\partial t}\right) \tag{6}$$

代入上式，得

$$\mathbf{f}=\epsilon_0(\boldsymbol{\nabla}\cdot\mathbf{E})\mathbf{E}+\frac{1}{\mu_0}(\boldsymbol{\nabla}\times\mathbf{B})\times\mathbf{B}-\epsilon_0\frac{\partial\mathbf{E}}{\partial t}\times\mathbf{B} \tag{7}$$

其中

$$\begin{aligned}\frac{\partial\mathbf{E}}{\partial t}\times\mathbf{B}&=\frac{\partial}{\partial t}(\mathbf{E}\times\mathbf{B})-\mathbf{E}\times\frac{\partial\mathbf{B}}{\partial t}\\&=\frac{\partial}{\partial t}(\mathbf{E}\times\mathbf{B})-(\boldsymbol{\nabla}\times\mathbf{E})\times\mathbf{E}\qquad\left(\boldsymbol{\nabla}\times\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial t}\right)\end{aligned} \tag{8}$$

代入上式得

$$\begin{aligned}\mathbf{f}&=\epsilon_0(\boldsymbol{\nabla}\cdot\mathbf{E})\mathbf{E}+\frac{1}{\mu_0}(\boldsymbol{\nabla}\times\mathbf{B})\times\mathbf{B}+\epsilon_0(\boldsymbol{\nabla}\times\mathbf{E})\times\mathbf{E}-\epsilon_0\frac{\partial}{\partial t}(\mathbf{E}\times\mathbf{B})\\&=\epsilon_0[(\boldsymbol{\nabla}\cdot\mathbf{E})\mathbf{E}+(\boldsymbol{\nabla}\times\mathbf{E})\times\mathbf{E}]+\frac{1}{\mu_0}(\boldsymbol{\nabla}\times\mathbf{B})\times\mathbf{B}-\epsilon_0\mu_0\frac{\partial\mathbf{s}}{\partial t}\end{aligned} \tag{9}$$

为了使式中电磁场的公式更加对称，不妨加上一项

$$\frac{1}{\mu_0}(\boldsymbol{\nabla}\cdot\mathbf{B})\mathbf{B}=\mathbf{0}\qquad(\boldsymbol{\nabla}\cdot\mathbf{B}=0) \tag{10}$$

得

$$\mathbf{f}=\epsilon_0[(\boldsymbol{\nabla}\cdot\mathbf{E})\mathbf{E}+(\boldsymbol{\nabla}\times\mathbf{E})\times\mathbf{E}]+\frac{1}{\mu_0}[(\boldsymbol{\nabla}\cdot\mathbf{B})\mathbf{B}+(\boldsymbol{\nabla}\times\mathbf{B})\times\mathbf{B}]-\epsilon_0\mu_0\frac{\partial\mathbf{s}}{\partial t} \tag{11}$$

一般来说，凡是出现两个连续的叉乘要尽量化成内积，下面计算 $(\boldsymbol{\nabla}\times\mathbf{E})\times\mathbf{E}$. 由吉布斯算子（劈形算符）的相关公式

$$\boldsymbol{\nabla}(\mathbf{A}\cdot\mathbf{B})=\mathbf{A}\times(\boldsymbol{\nabla}\times\mathbf{B})+\mathbf{B}\times(\boldsymbol{\nabla}\times\mathbf{A})+(\mathbf{A}\cdot\boldsymbol{\nabla})\mathbf{B}+(\mathbf{B}\cdot\boldsymbol{\nabla})\mathbf{A} \tag{12}$$

令 $\mathbf{A}=\mathbf{B}=\mathbf{E}$，得

$$\boldsymbol{\nabla}\left(\mathbf{E}^2\right)=2\mathbf{E}\times(\boldsymbol{\nabla}\times\mathbf{E})+2(\mathbf{E}\cdot\boldsymbol{\nabla})\mathbf{E} \tag{13}$$

即 $(\boldsymbol{\nabla}\times\mathbf{E})\times\mathbf{E}=(\mathbf{E}\cdot\boldsymbol{\nabla})\mathbf{E}-\boldsymbol{\nabla}\left(\mathbf{E}^2\right)/2$ 同理得

$$(\boldsymbol{\nabla}\times\mathbf{B})\times\mathbf{B}=(\mathbf{B}\cdot\boldsymbol{\nabla})\mathbf{B}-\frac{1}{2}\boldsymbol{\nabla}\left(\mathbf{B}^2\right) \tag{14}$$

代入得

$$\begin{aligned}\mathbf{f}=&\epsilon_0[(\boldsymbol{\nabla}\times E)\mathbf{E}+(\mathbf{E}\cdot\boldsymbol{\nabla})\mathbf{E}-\frac{1}{2}\boldsymbol{\nabla}\left(\mathbf{E}^2\right)]\\&+\frac{1}{\mu_0}[(\boldsymbol{\nabla}\cdot\mathbf{B})\mathbf{B}+(\mathbf{B}\cdot\boldsymbol{\nabla})\mathbf{B}-\frac{1}{2}\boldsymbol{\nabla}\left(\mathbf{B}^2\right)]-\epsilon_0\mu_0\frac{\partial\mathbf{s}}{\partial t}\end{aligned} \tag{15}$$

与式 3 对比，可以看出动量流密度张量的散度为

$$\begin{aligned}\boldsymbol{\nabla}\cdot\mathbf{T} =& -\epsilon_0[(\boldsymbol{\nabla}\cdot\mathbf{E})\mathbf{E}+(\mathbf{E}\cdot\boldsymbol{\nabla})\mathbf{E}-\frac{1}{2}\boldsymbol{\nabla}\left(\mathbf{E}^2\right)]\\ &-\frac{1}{\mu_0}[(\boldsymbol{\nabla}\cdot\mathbf{B})\mathbf{B}+(\mathbf{B}\cdot\boldsymbol{\nabla})\mathbf{B}-\frac{1}{2}\boldsymbol{\nabla}\left(\mathbf{B}^2\right)]\end{aligned} \tag{16}$$

接下来由二阶张量的散度计算公式，通过对比系数，就可以求出动量流密度张量 $\mathbf{T}$ （三阶矩阵）.

下面把等式右边的部分用求和符号表示（求和符号是张量分析中最常见的符号，只有熟练运用才能学好张量分析）. 下面推导用到了克罗内克 δ 函数，且定义任意矢量加上下标表示第个分量，例如

$$\mathbf{A}_j=\begin{cases}A_x & (j=1)\\ A_y & (j=2)\\ A_z & (j=3)\end{cases}\qquad x_j=\begin{cases}x & (j=1)\\ y & (j=2)\\ z & (j=3)\end{cases} \tag{17}$$

$(\boldsymbol{\nabla}\cdot\mathbf{E})\mathbf{E}+(\mathbf{E}\cdot\boldsymbol{\nabla})\mathbf{E}-\frac{1}{2}\boldsymbol{\nabla}\left(\mathbf{E}^2\right)$ 是一个矢量，它的第 j 个分量为

$$\begin{aligned}&[(\boldsymbol{\nabla}\cdot\mathbf{E})\mathbf{E}+(\mathbf{E}\cdot\boldsymbol{\nabla})\mathbf{E}-\frac{1}{2}\boldsymbol{\nabla}\left(\mathbf{E}^2\right)]_j\\ =&\sum_i\frac{\partial E_i}{\partial x_i}E_j+\sum_i E_i\frac{\partial E_j}{\partial x_i}-\frac{1}{2}\sum_i\frac{\partial\mathbf{E}^2}{\partial x_i}\delta_{ij}\\ =&\sum_i\left(\frac{\partial E_i}{\partial x_i}E_j+E_i\frac{\partial E_j}{\partial x_i}-\frac{1}{2}\frac{\partial\mathbf{E}^2}{\partial x_i}\delta_{ij}\right)\\ =&\sum_i\left(\frac{\partial}{\partial x_i}(E_iE_j)-\frac{1}{2}\frac{\partial\mathbf{E}^2}{\partial x_i}\delta_{ij}\right)\\ =&\sum_i\frac{\partial}{\partial x_i}\left(E_iE_j-\frac{1}{2}\mathbf{E}^2\delta_{ij}\right)\end{aligned} \tag{18}$$

同理

$$\left[(\boldsymbol{\nabla}\cdot\mathbf{B})\mathbf{B}+(\mathbf{B}\cdot\boldsymbol{\nabla})\mathbf{B}-\frac{1}{2}\boldsymbol{\nabla}\left(\mathbf{B}^2\right)\right]_j=\sum_i\frac{\partial}{\partial x_i}\left(B_iB_j-\frac{1}{2}\mathbf{B}^2\delta_{ij}\right) \tag{19}$$

所以

$$(\boldsymbol{\nabla}\cdot\mathbf{T})_j=\sum_i\frac{\partial}{\partial x_i}\left[\epsilon_0\left(\frac{1}{2}\mathbf{E}^2\delta_{ij}-E_iE_j\right)+\frac{1}{\mu_0}\left(\frac{1}{2}\mathbf{B}^2\delta_{ij}-B_iB_j\right)\right] \tag{20}$$

而由张量散度的定义

$$(\boldsymbol{\nabla}\cdot\mathbf{T})_j=\sum_i\frac{\partial}{\partial x_i}T_{ij}\tag{21}$$

得到动量流密度张量为

$$T_{ij}=\epsilon_0\left(\frac{1}{2}\mathbf{E}^2\delta_{ij}-E_iE_j\right)+\frac{1}{\mu_0}\left(\frac{1}{2}\mathbf{B}^2\delta_{ij}-B_iB_j\right)\tag{22}$$

理论上，在 $\mathbf{T}$ 上面加上任意一个满足 $\boldsymbol{\nabla}\cdot\mathbf{T}'=\mathbf{0}$ 的张量场，均可以使电磁场动量守恒，但是若规定无穷远处动量流密度为零，则可以证明 $T'_{ij}=0$.

磁旋比 玻尔磁子

预备知识 转动惯量[962]，磁矩[1187]

如果一个带电刚体，质量密度和电荷密度成正比，当它绕轴转动时，角动量[??] 为

$$\mathbf{L}=I\boldsymbol{\omega}=\boldsymbol{\omega}\int r^2\rho_m\,\mathrm{d}V\tag{1}$$

磁矩定义为

$$\boldsymbol{\mu}=Ia\hat{\boldsymbol{\omega}}=\int\frac{\mathrm{d}Q}{2\pi/\omega}\pi r^2\hat{\boldsymbol{\omega}}=\frac{1}{2}\boldsymbol{\omega}\int r^2\rho_e\,\mathrm{d}V\tag{2}$$

其中 r 为质量元到转轴的距离．两式比较，得

$$\boldsymbol{\mu}=\frac{q}{2m}\mathbf{L}\tag{3}$$

但对基本粒子（例如电子）的实验中，发现上式还需要一个修正因子，称为 **g 因子（g-factor）**

$$\boldsymbol{\mu}=g\frac{q}{2m}\mathbf{L}\tag{4}$$

定义**磁旋比（gyromagnetic ratio）**[10]为

$$\gamma=\frac{gq}{2m}\tag{5}$$

对于粒子的自旋，$L=\hbar\sqrt{l(l+1)}$．所以 $\mu=\sqrt{l(l+1)}\hbar gq/(2m)$．对于电子，实验测得

$$g_e=2.0023193043617(15)\approx 2\tag{6}$$

[10]有时也叫 **magnetogyric ratio**

$$\boldsymbol{\mu} = g_e \frac{e\hbar}{2m_e} \sqrt{\frac{1}{2}\left(1+\frac{1}{2}\right)} = \frac{\sqrt{3}}{2} g_e \mu_B \approx \sqrt{3}\mu_B \tag{7}$$

其中 μ_B 为**玻尔磁子（Bohr magneton）**，定义为

$$\mu_B = \frac{e\hbar}{2m_e} \tag{8}$$

第三章

电动力学 3

电磁场的参考系变换

- 本词条处于草稿阶段.

在 S 参考系中有一个电流为 I 的无限长直导线，令电流方向为 $\hat{\mathbf{z}}$. 离导线距离为 r_0 处有一个电荷为 q 的粒子沿正方向以速度 v 运动. 如果观察者的参考系 S' 相对 S 沿 $\hat{\mathbf{z}}$ 方向相对导线运动，速度为 u，在该参考系中求粒子所受的力. 这个力和 u 有关吗?

错误的分析

当 $u=0$ 时，电流在导线周围产生的磁场使粒子受到垂直于导线洛伦兹力. 而当 $u=v$ 时，导线同样产生磁场，而粒子却是静止的所以不受力. 这两个结论互相矛盾. 当 u 取其他不同值时，得到的受力也各不相同.

正确的分析

上面 $u=0$ 时的结论是正确的，空间中只有磁场没有电场. 然而当 $u\neq 0$ 时，由于导线中的异号电荷运动快慢不同，相对论尺缩效应使两种电荷的线密度产生区别，从而产生垂直导线的电场，与洛伦兹力共同作用在粒子上，使粒子受力与 u 无关.

具体计算

作为一个简单的导线模型，我们假设导线中的所有正电荷保持相对静止，其线电荷密度为 λ，以速度 v_0 相对导线延正方向运动，导线中的所有负电荷同样保持相对静止，线电荷密度为 $-\lambda$，以速度 $-v_0$ 相对导线延负方向运动. 这样，导线的电流为

$$I=2\lambda v_0 \tag{1}$$

点电荷感受到的磁场大小为

$$B=\frac{\mu_0}{2\pi}\frac{I}{r_0} \tag{2}$$

方向右右手螺旋定则决定. 那么它的受力大小为 $F=qvB$，方向指向导线.

（未完成）

拉格朗日电磁势

- 本词条处于草稿阶段.

预备知识 电磁场标势和矢势[1213]，拉格朗日方程[1052]

由于

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}_i}-\frac{\partial T}{\partial q_i}=Q_i \tag{1}$$

若非常规势能 $U(q_1,q_2...\dot{q}_1,\dot{q}_2...,t)$ 与广义力满足

$$Q_i=\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial U}{\partial \dot{q}_i}-\frac{\partial U}{\partial q_i} \tag{2}$$

且定义拉格朗日量为 $L=T-U$，则代入可得拉格朗日方程

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i}=\frac{\partial L}{\partial q_i} \tag{3}$$

现在证明在任意电磁场中运动的点电荷的非常规势为

$$U=q(\Phi-\mathbf{A}\cdot\mathbf{v}) \tag{4}$$

其中 Φ 和 $\mathbf{A}$ 分别为点电荷所在位置的电磁标势和矢势，$\mathbf{v}$ 为点电荷的速度.

直角坐标中的证明

以下用直角坐标证明（令 $x_1,x_2,x_3=x,y,z$）

$$F_i=Q_i=\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial U}{\partial \dot{x}_i}-\frac{\partial U}{\partial x_i} \tag{5}$$

以 x 分量为例

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial U}{\partial \dot{x}}=-q\frac{\mathrm{d}}{\mathrm{d}t}A_x=-q\left[(\boldsymbol{\nabla}A_x)\cdot\mathbf{v}+\frac{\partial A_x}{\partial t}\right] \tag{6}$$

$$\frac{\partial U}{\partial x}=q\left(\frac{\partial \Phi}{\partial x}-\frac{\partial \mathbf{A}}{\partial x}\cdot\mathbf{v}\right) \tag{7}$$

代入式 5 右边得

$$q\left[-\frac{\partial\Phi}{\partial x}-\frac{\partial A_x}{\partial t}+v_y\left(\frac{\partial A_y}{\partial x}-\frac{\partial A_x}{\partial y}\right)-v_z\left(\frac{\partial A_x}{\partial z}-\frac{\partial A_z}{\partial x}\right)\right] \tag{8}$$

根据广义洛伦兹力及电磁势的定义

$$\mathbf{F}=q(\mathbf{E}+\mathbf{v}\times\mathbf{B})\qquad \mathbf{E}=-\boldsymbol{\nabla}\Phi-\frac{\partial\mathbf{A}}{\partial t}\qquad \mathbf{B}=\boldsymbol{\nabla}\times\mathbf{A} \tag{9}$$

$$\begin{aligned}F_x&=q(E_x+v_yB_z-v_zB_y)\\&=q\left[-\frac{\partial\Phi}{\partial x}-\frac{\partial A_x}{\partial t}+v_y\left(\frac{\partial A_y}{\partial x}-\frac{\partial A_x}{\partial y}\right)-v_z\left(\frac{\partial A_x}{\partial z}-\frac{\partial A_z}{\partial x}\right)\right]\end{aligned} \tag{10}$$

对比可得式 5 在 x 方向成立，y,z 分量的证明类似．另外容易证明多个点电荷组成的系统的非常规势等于每个点电荷的非常规势之和．

广义坐标的证明

下面来证明对于任意广义坐标 $\{q_i\}$，式 4 都满足式 2．首先令变换关系为

$$x_i=x_i(q_1,q_2,\ldots,t) \tag{11}$$

其中 x_i 包括每个质点的 x,y,z 坐标，则

$$\frac{\partial U}{\partial\dot{q}_i}=\sum_j\frac{\partial U}{\partial\dot{x}_j}\frac{\partial\dot{x}_j}{\partial\dot{q}_i}=\sum_j\frac{\partial U}{\partial\dot{x}_j}\frac{\partial x_j}{\partial q_i} \tag{12}$$

注意第二步用到了式 4 [1063]．对时间求全导数得

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial U}{\partial\dot{q}_i}=\sum_j\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial U}{\partial\dot{x}_j}\frac{\partial x_j}{\partial q_i}+\sum_j\frac{\partial U}{\partial\dot{x}_j}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial x_j}{\partial q_i} \tag{13}$$

另外有

$$\frac{\partial U}{\partial q_i}=\sum_j\frac{\partial U}{\partial x_j}\frac{\partial x_j}{\partial q_i}+\sum_j\frac{\partial U}{\partial\dot{x}_j}\frac{\partial\dot{x}_j}{\partial q_i} \tag{14}$$

易证式 13 和式 14 的右边第二项相等．将以上两式代入式 2 右边得（第二步利用了式 5）

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial U}{\partial\dot{x}_i}-\frac{\partial U}{\partial x_i}=\sum_j\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial U}{\partial\dot{x}_j}\frac{\partial x_j}{\partial q_i}-\sum_j\frac{\partial U}{\partial x_j}\frac{\partial x_j}{\partial q_i}=\sum_j F_j\frac{\partial x_j}{\partial q_i} \tag{15}$$

而这恰好是 Q_i 的定义．证毕．

哈密顿量

首先求广义动量

$$\mathbf{p}=\frac{\partial L}{\partial \dot{\mathbf{q}}}=m\mathbf{v}+q\mathbf{A} \tag{16}$$

则哈密顿量为

$$H=\mathbf{p}\dot{\mathbf{q}}-L=\frac{1}{2m}(\mathbf{p}-q\mathbf{A})^2+q\Phi \tag{17}$$

电磁场角动量分解

电磁场的动量为

$$\mathbf{p}=\epsilon_0\int \mathrm{d}V\,\mathbf{E}\times\mathbf{B} \tag{1}$$

角动量为

$$\mathbf{J}=\mathbf{r}\times\mathbf{p}=\epsilon_0\int \mathrm{d}V\,\mathbf{r}\times\mathbf{E}\times(\boldsymbol{\nabla}\times\mathbf{A}) \tag{2}$$

现在假设电磁场只在一定范围内不为零，且体积分的边界处场强为零．假设该范围内没有净电荷与电流，则

$$\begin{aligned}\mathbf{r}\times\mathbf{E}\times(\boldsymbol{\nabla}\times\mathbf{A})&=\mathbf{r}\times[\boldsymbol{\nabla}(\mathbf{E}\cdot\mathbf{A})-\mathbf{A}(\mathbf{E}\cdot\boldsymbol{\nabla})]_{\partial A}\\&=[(\mathbf{r}\times\boldsymbol{\nabla})(\mathbf{E}\cdot\mathbf{A})-(\mathbf{E}\cdot\boldsymbol{\nabla})(\mathbf{r}\times\mathbf{A})]_{\partial A}\end{aligned} \tag{3}$$

其中转微分算符 $[]_{\partial A}$ 的作用是先把方括号内的 $\boldsymbol{\nabla}$ 作为普通矢量进行计算，再把展开结果中每一项的偏微分作用在 A 的分量上．上式第一项为 $\sum\limits_i E_i(\mathbf{r}\times\boldsymbol{\nabla})A_i$，第二项为

$$-[(\mathbf{E}\cdot\boldsymbol{\nabla})(\mathbf{r}\times\mathbf{A})]_{\partial A}=-(\mathbf{E}\cdot\boldsymbol{\nabla})(\mathbf{r}\times\mathbf{A})+[(\mathbf{E}\cdot\boldsymbol{\nabla})(\mathbf{r}\times\mathbf{A})]_{\partial r} \tag{4}$$

其中第二项为 $[(\mathbf{E}\cdot\boldsymbol{\nabla})(\mathbf{r}\times\mathbf{A})]_{\partial r}=[(\mathbf{E}\cdot\boldsymbol{\nabla})\mathbf{r}]\times\mathbf{A}=\mathbf{E}\times\mathbf{A}$，第一项中

$$\begin{aligned}(\mathbf{E}\cdot\boldsymbol{\nabla})(\mathbf{r}\times\mathbf{A})&=[(\mathbf{E}\cdot\boldsymbol{\nabla})(\mathbf{r}\times\mathbf{A})]_{\partial ErA}-[(\mathbf{E}\cdot\boldsymbol{\nabla})(\mathbf{r}\times\mathbf{A})]_{\partial E}\\&=[(\mathbf{E}\cdot\boldsymbol{\nabla})(\mathbf{r}\times\mathbf{A})]_{\partial ErA}\end{aligned} \tag{5}$$

这是因为 $[(\mathbf{E}\cdot\boldsymbol{\nabla})(\mathbf{r}\times\mathbf{A})]_{\partial E}=(\boldsymbol{\nabla}\cdot\mathbf{E})(\mathbf{r}\times\mathbf{A})=0$．综上，

$$\mathbf{J}=\epsilon_0\int \mathrm{d}V\sum_i E_i(\mathbf{r}\times\boldsymbol{\nabla})A_i+\epsilon_0\int \mathrm{d}V\,\mathbf{E}\times\mathbf{A}+\epsilon_0\int \mathrm{d}V\,[(\mathbf{E}\cdot\boldsymbol{\nabla})(\mathbf{r}\times\mathbf{A})]_{\partial ErA} \tag{6}$$

现在证明最后一项为 0．以 x 分量为例，

$$\begin{aligned}\hat{\mathbf{x}}\int \mathrm{d}V\,[(\mathbf{E}\cdot\boldsymbol{\nabla})(\mathbf{r}\times\mathbf{A})]_{\partial E r A} &= \int \mathrm{d}V\,\boldsymbol{\nabla}\cdot[\mathbf{E}(\hat{\mathbf{x}}\cdot\mathbf{r}\times\mathbf{A})] \\ &= \oint \mathrm{d}\mathbf{s}\,\mathbf{E}(\hat{\mathbf{x}}\cdot\mathbf{r}\times\mathbf{A}) = 0\end{aligned} \tag{7}$$

最后一步是因为边界处场强为零．现在我们可以看出角动量由两部分组成

$$\mathbf{J}=\mathbf{L}+\mathbf{S}\qquad \mathbf{L}=\epsilon_0\int \mathrm{d}V\sum_i E_i(\mathbf{r}\times\boldsymbol{\nabla})A_i\qquad \mathbf{S}=\epsilon_0\int \mathrm{d}V\,\mathbf{E}\times\mathbf{A} \tag{8}$$

其中 $\mathbf{L}$ 是轨道角动量，$\mathbf{S}$ 是自旋角动量．

第十四部分

相对论

第一章

狭义相对论

狭义相对论的基本假设

预备知识 公理系统[91]，麦克斯韦方程组[1223]

“光速在任意参考系下都不变”这一理念，和广为流传的误解不同，并不是实验中得来的．部分书籍会简单粗暴地告诉你，Michelson-Morley（迈克尔逊-莫雷）实验（下称 **MM 实验**）是为了寻找以太存在的证据而进行的，实验结果表明在误差范围内光速在任何参考系都是一样的，因此提出了“光速不变原理”，也就是狭义相对论的两个公理之一．然而，这是对历史的错误描述．一些物理课本使用这样的描述是有其教育意义的，因为用这样的误解很容易引入狭义相对论，而不需要学生有扎实的电动力学基础．

本书秉承准确、翔实的原则，将当代物理史学界对“光速不变原理”的由来阐释如下．

关于 Michelson-Morley 实验的常见误解

误解 1：MM 实验的目的是寻找以太

[1]不是的．

MM 实验是在 1887 年进行的．科学家们曾经就“光到底是粒子还是波”争论不休，早年由于牛顿的绝对权威，几乎所有人都顺从他而坚持光的粒子说．在 1800 年至 1850 年之间，波动说才逐渐占据了上风．但是在当时的常识看来，波动是介质运动的结果，也就是说，有波动就必须有介质．光既然是波，那么它就一定有介质，物理学家们将其称为**以太（ether）**．

很明显，根据定义，以太会和光作用．但是以太会不会和其它物质作用呢？这个问题没有直接的答案．光学界的泰斗杨（Young）和菲涅尔（Fresnel）认为以太是不会和普通物质相互作用的；而斯托克斯（Stokes）认为以太应该和普通物质作用，从而会被普通物质拖拽而产生运动．

杨的观点来自于一个天文学现象：恒星的**像差（aberration）**[2]．而斯托克斯的观点是，由于光有偏振，意味着光是横波；即使是可见光也有极高的频

[1]本小节的内容主要来自 University of Minnesota 在 2019 秋季的 History of 20th Century Physics 课程

[2]“Upon considering the phenomena of the aberration of the stars I am disposed to believe, that the luminiferous ether pervades the substance of all material bodies with little or no resistance, as freely perhaps as the wind passes

率[3]，因此以太必须坚如磐石，怎么可能“像风吹过树林一般”？所以以太必须和普通物质相互拖拽[4].

MM 实验由此产生，其目的是检验以太究竟会不会和普通物质作用．**也就是说，MM 实验默认了以太的存在**．实验的基本思想是，如果以太不和普通物质作用，那么以太可以看成是一种绝对静止的参考系，地球在公转、自转等运动中一定会相对以太而运动，否则就过于凑巧了．但是如果以太和普通物质作用，那么地球就会拖拽以太，在地面上以太应该近乎静止．如果我们在地面检测光在两个相互垂直的方向上的传播速度，那么杨的理论预言光速会有不同，而斯通克斯的理论则语言光速不会变化．

当然了，实验结果是光速在两个方向上的差别非常小，因此当时人们认为 MM 实验证明了以太和普通物质有相互作用．

误解 2：爱因斯坦是受了 MM 实验的启发而提出“光速不变”的

不是的．

尽管 MM 实验不是为了寻找以太，但实验结果也可以解读为我们今天所说的“光速不变原理”．但事实上，爱因斯坦并没有拿 MM 实验当回事，在他 1905 年的论文里并没有提到 MM 实验——至少没有明确提到．在很长一段时间里，物理史学者甚至都不确定爱因斯坦那时知不知道 MM 实验（现在的结论是，他知道的）．1905 年之后，爱因斯坦倒是频繁提到 MM 实验和狭义相对论的关系，但是并没有说过实验支撑了光速不变原理．

1931 年，爱因斯坦和迈克尔逊在加州帕萨迪纳会面．爱因斯坦很不解地问迈克尔逊为什么要做那么一个实验，后者回答说，"Because it’s fun!（因为很有意思啊！）"

误解 3：洛伦兹只不过提出了尺缩钟慢原理

洛伦兹的贡献实际上更多．

---

through a grove of trees.”Thomas Young 于 1804 年所说．笔者翻译其原话为：“考虑恒星的像差现象，我自然而然地相信，传播光的以太在任何物质中弥散，只有很少甚至没有阻力，就好像风吹过树林一般．”

[3]比如说，589 纳米波长的光在我们眼中是黄色的，其频率高达 $\nu = c/\lambda = 5 \times 10^{17}\,\mathrm{Hz}$．其中 c 为光速，$\lambda = 5.89 \times 10^{-10}\,\mathrm{m}$ 为波长

[4]笔者对此论述有疑义：如果以太和普通物质不相互作用，那它坚如磐石也和普通物质无关啊．那个时代的物理学家普遍有很多来自日常经验的先入为主思想，在今天看来逻辑不严谨其实并不奇怪．

斯托克斯显然认为，两个方向的光速不变意味着以太相对地面静止，从而说明地球很可能和以太相互作用．洛伦兹（Lorentz）和斐兹杰惹（FitzGerald）提出，这有可能是因为干涉仪的臂长在和以太运动方向平行的时候更短而造成的．也就是说，他们依然认为以太是存在的，只不过用尺缩假设来支持杨的想法．实验物理学家们此时有了三个研究方向：寻找以太风、探测以太和普通物质的拖拽以及测量尺缩效应．

但是洛伦兹想得更多．由于那时已经有了麦克斯韦理论，人们知道光是一种电磁波．如果以太是光的介质，那就意味着以太是电场和磁场的介质，电磁场分布在以太里而不是凭空存在的．如果普通物质不和以太作用，那么物体在以太中的运动完全不会造成扰动；但是带电物质会被电磁场所作用，也就是说以太反过来会对物质有作用．那么以太究竟算不算可以和普通物质作用呢？

另外，由于麦克斯韦理论中并没有指明是在哪个惯性系成立，因此如果我们假设它在和以太相对静止的参考系成立而地球和以太有相对运动的话，那么麦克斯韦理论应该在地球上不成立．但是大量广泛的应用表明，麦克斯韦理论在地球上是适用的．

为了解释上述矛盾，洛伦兹猜想，同一个电磁场在不同的惯性系中应该会有所不同．他的猜想解释了为什么任何至多有一阶近似的光学实验都不可能探测出以太的运动．他进一步完善了这个理论，给出了电磁场在不同惯性系中的变化．巧合的是，这种变化和物质的尺缩效应是完全一样的．洛伦兹认为这只是一个巧合，但这个结果极大地启发了闵可夫斯基时空．相对论思想其实是在洛伦兹工作的基础上更进一步，把巧合看成是必然的，解释为时空本身的变换．所以洛伦兹不只是提出了尺缩钟慢原理和洛伦兹变换，他对电磁理论的修正也启发了相对论的时空观．

光速不变原理和相对性原理

在洛伦兹的基础上更进一步，以太的存在已经没什么逻辑上的必要了．假设光速不会变化是由时空本身的属性造成的，而不是任何类似以太的物质的作用，那么我们就得到了光速不变原理：光速在任何参考系中都不变．今天的公制单位中，使用铯-133 原子中两个超精细分裂能级的跃迁频率来定义秒，而使用光速来定义米：一米就是光在 $1/299792458$ 秒内行进的距离．也就是说，光速被**定义**为 $c = 299792458\,\mathrm{m/s}$．

注意，由于单位是可以任意取的，如果你强行定义 $\tau = 299792458\,\mathrm{s}$，那么在这个单位下也有 $c = 1\,\mathrm{m}/\tau$．因此理论物理学家通常不会使用冗长的 c 的表达，而是简单地令 $c = 1$．这样计算结果非常简洁，并且如果想要回归日常单位，只需要做量纲分析，在合适的地方重新把 c 加上就可以了．比如说，当你算出来某个东西的速度是 0.2，考虑到速度的量纲，你自然能得出这个东西的速度应该为 $0.2c$．

抛弃以太而选择光速不变原理，相当于赋予了任何惯性参考系以相同的地位，没有哪个更特殊．秉承相同的理念，爱因斯坦进而认为任何惯性参考系都应该遵循完全相同的物理规律，不应该有地位差别．

洛伦兹对电磁理论的修正说明我们从伽利略时代以来就坚守的时空观并不能保证物理规律的一致性．这意味着我们需要新的时空观．

事件与尺缩效应

预备知识 狭义相对论的基本假设[1253]

事件

在和时空相关的理论中，当我们描述一件事的时候，我们并不关心这件事具体是什么，只关心它发生在何时何地．因此为了将来的讨论，我们首先需要定义“事件”的概念．

一个**事件（event）**是指在时空坐标系中的一个点．事件所发生的时间、地点，就是事件作为一个点的坐标．

对事件的观测

狭义相对论的核心是光．在任何参考系中，光速不变．光的其它性质并不能保证一定不变，如光强的分布，偏振的角度等．

除了光速以外，事件本身也是不随惯性系变化的．这就是说，在任何惯性系 K_1 中同时同地发生的事情，在任何惯性系 K_2 中也是同时同地发生的．鉴于我们已经为了简便，将事件简单表达为它所发生的时间和地点，那么同时同地发生的事件都应看成同一个事件．

更进一步，由于讨论事件时我们只关心其发生的时间和位置，即只关心其时空坐标，因此也可以直接用**事件的时空坐标**来指代**事件本身**.

光速和事件的不变性，是目前我们观测事件的最基本工具．我们将在实践中体会利用光速不变和事件不变来观测事件的方法.

同时性的相对性

考虑一根的铁轨，向左向右都无限延伸．在这铁轨上取一个点作为原点，向右作为正方向，可以画一个 x_1 轴，用来测量和铁轨静止的参考系中的事件位置，这个参考系称作**铁轨系**，记为 K_1.

现在，铁轨上从左到右开过一辆火车．和火车静止的参考系也可以沿着铁轨画一个 x_2 轴，只不过它是用来描述火车参考系中事件位置的，称作**火车系**，记为 K_2.

在铁轨系中，如果某时刻看到火车的两个不重叠的轮子同时发光，那么这两道光会在铁轨上的两个发光点的中点相遇，而“相遇”也是一个事件．从发光到相遇，两束光通过了相同的路程，由于光速不变，它们经过了相同的时间，由此反推可知发光的时间是一样的．但是同样的三个事件在火车系看来是不一样的：在火车系中，“相遇”发生在更靠近后轮的位置，也就是说，在火车看来前轮所发的光走过了更长的路程，花了更长的时间，从“相遇”的时间反推回去，可知在火车眼里是前轮先发光.

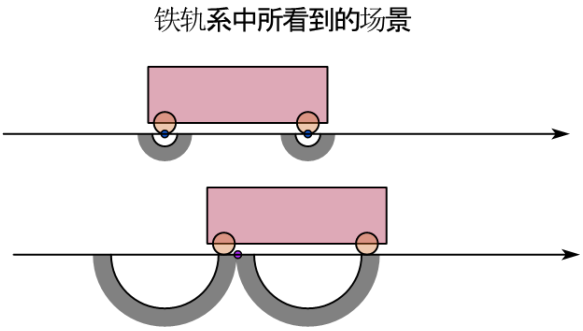


图 1: 在铁路系中所看到的三个事件，分别用三个点表示．上图是两个轮子同时发光的两个事件；下图是一段时间以后，火车运动了一段距离，而两束光相遇的事件.

事实上，在 K_1 中同时但不同地发生的事情，在 K_2 中必然不同时发生．“同

时”这一概念并非绝对，两个事件是否同时，取决于从什么参考系来观察它们.

习题 1　火车系中的事件

图 1 中是以铁轨的视角，选取了两个时刻，描述了“前轮发光”、“后轮发光”和“光束相遇”这三个事件. 请你尝试画出火车的视角下三个事件的先后关系. 提示：你需要三个关键时刻，依次是“前轮发光”时，“后轮发光”时和“光束相遇”时.

尺缩效应

我们还是使用上一节定义的火车系 K_2 和铁轨系 K_1. 如果说，在铁轨上标记了两个点 A 和 B，使得前轮通过 B 时发光，后轮通过 A 时发光. 在 K_1 中，前轮和后轮分别同时通过这两个点，也就是说，在 K_1 中，火车两轮的间距和 A、B 的间距一样；但是在 K_2 中来看，前轮先发光，后轮后发光，这就意味着火车两轮的间距比 A、B 的间距要长.

这说明，运动的物体应该比静止时看起来要短. 由于没有任何点是特殊的，所以这种运动造成的收缩在每一个地方都是一样的，或者说，运动造成的尺缩效应是均匀的. 那么运动造成的收缩的比例应该怎么计算呢？

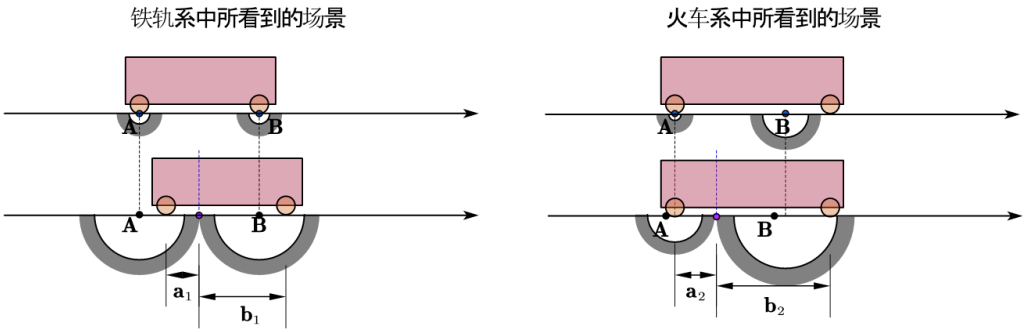


图 2: 尺缩效应配图

把 A、B 的间距看成 AB 的长度，车轮间距看成火车的长度. 设 AB 的静止长度（在 K_1 中的长度）为 $2S$，而火车的静止长度（在 K_2 中的长度）为 $2L$.

记火车相对铁轨的运动速度为（沿着 x 正方向）v，考虑到两个参考系中

没有哪个更特殊，则铁轨相对火车的运动速度为 $-v$；同样，火车在铁轨系中的收缩比例，也应该和铁轨在火车系中的收缩比例相等．记这个收缩比例是 $m \in (0,1)$，那么“在**铁轨系**中火车和 AB 长度一样”意味着：

$$2mL = 2S \tag{1}$$

在**火车系**中，火车的长度是 $2L$，大于 AB 的长度 $2mS$，所以前轮先碾过 B 点发光，然后才轮到后轮碾过 A 点发光．前后轮发光各自是一个独立的事件，所以它们是否同时取决于参考系的选择；但是有一个东西在两个参考系中都是一样的，那就是两束光相遇的位置，因为两束光的相遇是一个单独的事件．

我们来考察一下，在两个参考系中，光相遇的位置对应于车上的哪个地方．如下图所示，在 K_1 中，设相遇点到火车后轮的距离是 a_1，到火车前轮的距离是 b_1；在 K_2 中，设相遇点到火车后轮的距离是 a_2，到火车前轮的距离是 b_2．由于收缩是均匀的，相遇点在两个参考系中都是同一个点，因此

$$\frac{a_1}{b_1} = \frac{a_2}{b_2} \tag{2}$$

根据两个场景的不同，具体计算一下 a_1，b_1，a_2，b_2，得到[5]：

$$a_1 = S - v \cdot \frac{S}{c}, b_1 = S + v \cdot \frac{S}{c}, a_2 + b_2 = 2L, \frac{b_2 - a_2}{c} = \frac{2L - 2mS}{v} \tag{3}$$

整理式 3 并代入式 1 得：

$$\frac{a_1}{b_1} = \frac{c - v}{c + v} \tag{4}$$

$$\frac{a_2}{b_2} = \frac{v - c(1 - m^2)}{v + c(1 - m^2)} \tag{5}$$

把式 2 代入式 4 和式 5 中得：

$$\frac{c - v}{c + v} = \frac{v - c(1 - m^2)}{v + c(1 - m^2)} \tag{6}$$

在等式右边上下同乘以 c/v 得：

[5]第四个等式两端都是 K_2 中两个轮子发光的时间间隔

$$\frac{c-v}{c+v}=\frac{c-\frac{c^2}{v}(1-m^2)}{c+\frac{c^2}{v}(1-m^2)} \tag{7}$$

这样就可以直接得到：

$$v=\frac{c^2}{v}(1-m^2) \tag{8}$$

解得

$$m=\sqrt{1-\frac{v^2}{c^2}} \tag{9}$$

结论就是，如果一个物体静止时的长度为 L，那么在某一惯性系中若它沿着自身长度的方向运动速度为 v，则在此参考系中它的长度是 $\sqrt{1-\frac{v^2}{c^2}}\cdot L<L$.

时间的变换与钟慢效应

预备知识 尺缩效应[1256]

伽利略变换中时间是绝对而独立的，和空间无关．但我们在事件与尺缩效应[1256] 词条中已经知道，引入光速不变原理之后会发现，两个独立事件[6]是否同时，取决于观察的惯性系是哪个．

测量事件发生时间

在一个一维空间中，发生了一个事件，我们自然能知道这个事件发生的位置．实际上，因为我们要求事件发生的时候都同时发射一道球面光，在空间各个地方布满光感探测器以后，就可以通过探测的结果来反推球面光的形状，进而计算出球心位置．当然，我们没必要真的这么麻烦地计算，因为毕竟我们是在做思想实验，只需要明确“球面波的形状可测量”，进而得出“事件发生的位置是可知的”就行了．

但是事件发生的时间需要一点技巧去求得．很容易想到，利用光速不变原理就可以把对位置的知识转化成对时间的知识：首先让时间和空间的原点处

[6]注意，我们已经声明过将同时同地的事件都看作同一个，因此此处的“独立事件”不可能是同时同地的．

（$t=0, x=0$）发生一个事件，不妨称之为**参考事件**．在某时某地发生了另一个需要我们观测的事件，那么参考事件的球面光和待观测事件的球面光相遇也是一个事件．相遇事件发生的位置到原点的距离，除以光速，就可以得到相遇事件发生的时间；同理也可以得到相遇事件和待观测事件所发生的时间差，进而计算出待观测事件发生的时间．

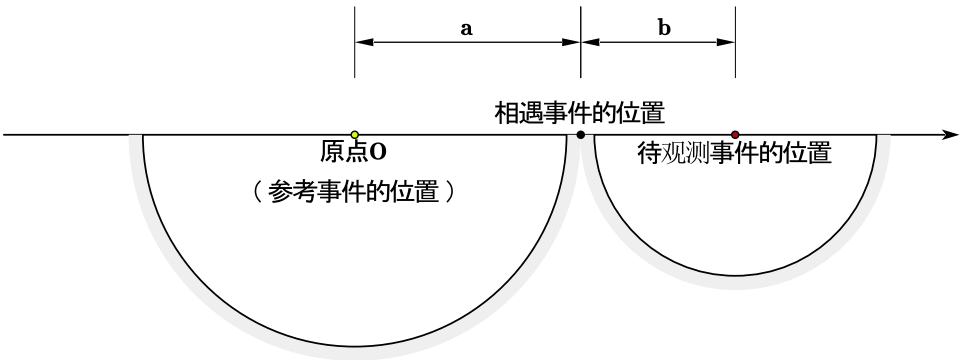


图 1: 测量待观测事件发生的时间．相遇事件发生的时间是 a/c，它和待观测事件的时间间隔是 $-b/c$．也就是说，在如图所示的情形下，待观测事件发生的时间是 $(a-b)/c=[2a-(a+b)]/c$．注意待观测事件的位置不是 b，而是 $a+b$.

这个方法在我们推导两个惯性系中时间的变换时非常好用．注意一点：我们要求相遇事件是参考事件的光球和待观测事件的光球**相切**的时候，不论是内切还是外切．外切可能出现在待观测事件是在负的时间上发生的的情况下——这时我们需要拓展事件光球的概念，在事件发生以前，光球向事件发生地收缩，在发生的那一刻收缩为点，重新发散出去．

时间的变换

沿用铁轨系 K_1 和火车系 K_2 的设定，令 K_2 相对 K_1 以速度 v 运动．我们现在想计算一下，在 K_1 中任意点 (x_1,t_1) 发生的事件，在 K_2 中应该对应哪个坐标 (x_2,t_2)．由于这里的 x_1 是任意的，我们就需要火车无限长，不妨直接把火车和铁轨都抽象为一根无限长的坐标轴．

令两个坐标系的原点在各自的 $t_1=0$，$t_2=0$ 时刻重合．我们可以任意指定一个事件，定义它在两个坐标系中的时空坐标都是 $(0,0)$，从而实现“重合”的要求．这样一来，在两个坐标系的时空原点处发生的参考事件就成了同一个

事件.

待观测事件在 K_1 中的坐标是 (x_1, t_1)，意味着相遇事件在 K_1 中的坐标是 $(t_1c + \frac{x_1-t_1c}{2}, t_1 + \frac{x_1-t_1c}{2c}) = (\frac{x_1+t_1c}{2}, \frac{x_1+t_1c}{2c})$. 注意相遇事件的时间和空间坐标中间只差了一个因子 c，这是因为相遇事件必然在参考事件的光球上. 由上一个小节知，利用相遇事件和待观测事件的空间坐标，就可以算出待观测事件的时间坐标；虽然我们已经知道了 t_1，但你仍然可以验算一下，当作练习.

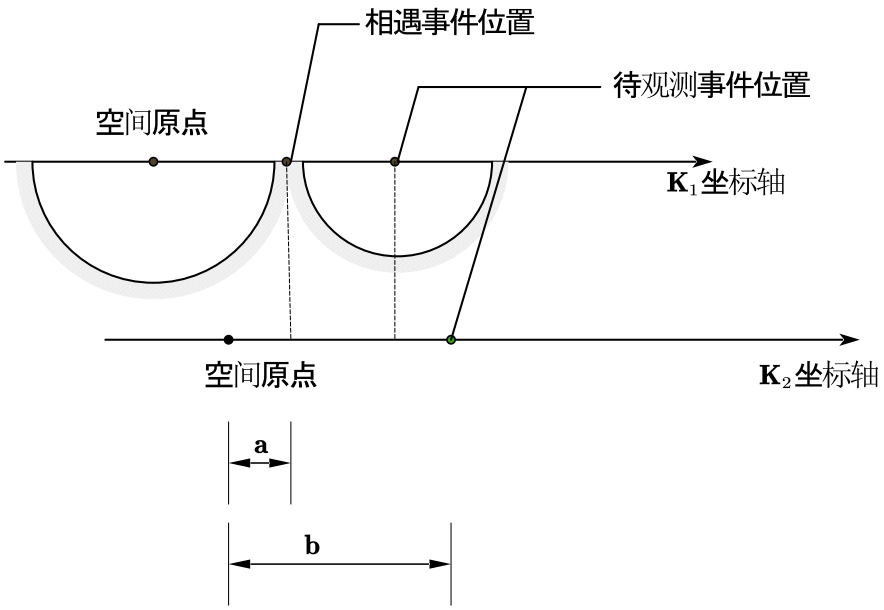


图 2: 相遇事件在两个坐标轴中的表示. a 是 K_2 中相遇事件的位置，b 是 K_2 中待观测事件的位置. 这个图是在 K_1 的视角下绘制的，请注意 K_2 坐标轴的尺缩效应使得 K_2 的尺子看起来更密集了一些.

同样地,我们知道相遇事件在 K_2 中待观测事件的位置是 $(x_1-vt_1)/\sqrt{1-\frac{v^2}{c^2}}$[7]; 类似地,相遇事件的位置[8] 是 $(\frac{x_1+t_1c}{2}-v\frac{x_1+t_1c}{2c})/\sqrt{1-\frac{v^2}{c^2}} = \frac{x_1+t_1c}{2}\cdot(1-\frac{v}{c})/\sqrt{1-\frac{v^2}{c^2}}$

代入图 1 的结果,可知 K_2 中待观测事件的时间是 $\Big[(x_1+t_1c)\cdot(1-\frac{v}{c})/\sqrt{1-\frac{v^2}{c^2}} - (x_1 -$

$\frac{1}{c} = \frac{t_1-\frac{v}{c^2}x_1}{\sqrt{1-\frac{v^2}{c^2}}}$.

由此，我们得出了时间的变换：K_1 中的事件 (x_1, t_1)，在 K_2 中的坐标是

[7]直接由尺缩效应得. $x_1 - vt_1$ 部分是在 K_1 中看到的 K_2 的原点和待观测事件的距离.

[8]可以用 $\frac{x_1+t_1c}{2}$ 代替 x_1，$\frac{x_1+t_1c}{2c}$ 代替 t_1，代入 $(x_1 - vt_1)/\sqrt{1-\frac{v^2}{c^2}}$ 得到.

$((x_1-vt_1)/\sqrt{1-\frac{v^2}{c^2}},(t_1-\frac{v}{c^2}x_1)/\sqrt{1-\frac{v^2}{c^2}})$．这就是**一维空间中的洛伦兹变换**.

钟慢效应和双生子悖论

我们依然考虑火车系 K_2 和铁轨系 K_1．在两个参考系的空间原点处放一只钟表，精确地记录着流逝的时间．火车上的钟表由于跟着火车一起运动，从**铁轨系**的视角来看，它的坐标随着时间 t_1 变化：(vt_1,t_1)．应用上一节的结论，可以计算出该钟表在**火车系**中的坐标：$(0,t_1\sqrt{1-\frac{v^2}{c^2}})$．空间坐标恒为零是肯定的，因为我们把它挂在了火车的原点上；时间坐标正比于 t_1，但是乘以了一个小于 1 的因子 $\sqrt{1-\frac{v^2}{c^2}}$．也就是说，在 K_1 看来，火车上的时间始终要过得比 K_1 中慢.

两个坐标系并没有哪个更特殊，所以可以简单地把钟慢效应理解为：以速度 v 运动的物体，其时间的流逝速度只有静止观察者手上那块表的时间流逝速度的 $\sqrt{1-\frac{v^2}{c^2}}$．这里的时间流逝是最基本的物理性质，意味着如果一艘飞船相对你运动，在你看来飞船上的一切物理活动都变慢了，化学反应速率也以相同的比例变慢，质子衰变的速率也如此.

基于这个效应，人们自然提出了一个问题：如果有一对同年同月同日生的双生子，一个人留在地球上，另一个人乘飞船高速远航．若干年后宇航双生子回到地球，两人拥抱在一起．这时候，谁的年龄更大？

对于此问题的解答，请参见**双生子佯谬**[1304] 词条.

洛伦兹变换

预备知识 时间的变换[1260]

洛伦兹变换

在时间的变换与钟慢效应[1260] 一节中，我们自然而然地推导出了一维空间中的洛伦兹变换：

$$\begin{cases} x_2 = \dfrac{x_1 - vt_1}{\sqrt{1 - v^2/c^2}} \\ t_2 = \dfrac{t_1 - vx_1/c^2}{\sqrt{1 - v^2/c^2}} \end{cases} \tag{1}$$

当然，由于没有哪个惯性系更特殊，考虑到 K_1 相对 K_2 的速度是 $-v$，我们有：

$$\begin{cases} x_1 = \dfrac{x_2 + vt_2}{\sqrt{1 - v^2/c^2}} \\ t_1 = \dfrac{t_2 + vx_2/c^2}{\sqrt{1 - v^2/c^2}} \end{cases} \tag{2}$$

当然了，这一逆变换也是可以手动从式 1 中解出来的.

例 1　垂直方向上的洛伦兹变换

依然考虑火车模型. 现在在火车和铁轨的原点都树一根杆子，从铁轨系看来两根杆子高度相同. 请说明为什么从火车系看来两根杆子高度依然相同.（提示：竖直的杆子和水平的杆子本质区别是什么？利用同时性的相对性说明一下. 你可能需要考虑“两个事件的同时性不是绝对的”以及“同一个事件在任何参考系都是同一个事件”.）

例 1 的结果说明，尺缩效应和同时性的相对性只发生在火车运动的方向上. 这就是说，垂直于火车的坐标轴不会因为火车的运动而发生变化. 这就引出了完整版的三维空间中——或者说四维时空中的洛伦兹变换：

$$\begin{cases} x' = \dfrac{x - vt}{\sqrt{1 - v^2/c^2}} \\ y' = y \\ z' = z \\ t' = \dfrac{t - vx/c^2}{\sqrt{1 - v^2/c^2}} \end{cases} \qquad \begin{cases} x = \dfrac{x' + vt'}{\sqrt{1 - v^2/c^2}} \\ y = y' \\ z = z' \\ t = \dfrac{t' + vx'/c^2}{\sqrt{1 - v^2/c^2}} \end{cases} \tag{3}$$

这里我们应用了多数资料中使用的字母习惯，避免读者造成混淆. 在上式中，$x_1 = x$，$t_1 = t$，$x_2 = x'$，$t_2 = t'$.

矩阵表示

洛伦兹变换也可以用矩阵表示为：

$$L=\begin{bmatrix} \frac{1}{\sqrt{1-v^2/c^2}} & -\frac{v/c^2}{\sqrt{1-v^2/c^2}} & 0 & 0 \\ -\frac{v}{\sqrt{1-v^2/c^2}} & \frac{1}{\sqrt{1-v^2/c^2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

将 L 称为**洛伦兹矩阵**或**洛伦兹变换矩阵**.

如果一个事件在 K_1 和 K_2 中的坐标分别是 $\begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix}$ 和 $\begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix}$，那么有

$$\begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix}=L\begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} \tag{5}$$

斜坐标系

预备知识 直角坐标系，几何矢量[370]

斜坐标系是通常的直角坐标系的延伸. 以二维空间中为例，任取两根相互不平行的坐标轴 x 和 y，都能对平面上任意一点 P 进行定位，方法是认为坐标轴上已经给定了标尺，从 P 处画一条平行于 y 轴的直线，它与 x 轴相交处的标尺数值即为 P 的 x 坐标；类似地，画一条平行于 x 轴的直线，它与 y 轴相交处的标尺数值即为 P 的 y 坐标.

斜坐标系的标尺不一定是均匀的. 事实上，只要标尺数值沿着坐标轴的方向一直递增就可以.

与直角坐标的转换

为方便计，我们只考虑标尺均匀的斜坐标系.

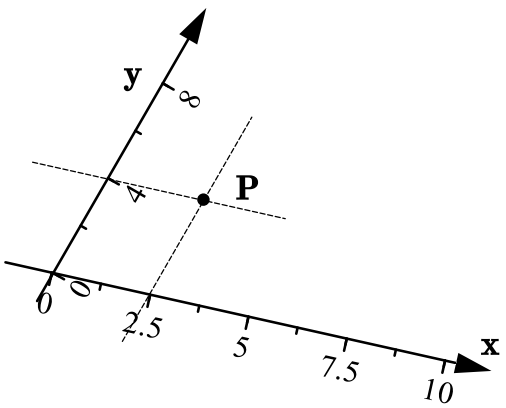


图 1: 一个斜坐标系的示意图．图中 x、y 两轴的夹角是 72.9°，两轴的标尺已经给出．在这个坐标系中，P 的坐标是 $< 2.5, 4 >$.

给定直角坐标系 xOy，由相互垂直的 x 轴和 y 轴构成，两轴交于原点 O．另给定斜坐标系 $x'Oy'$，由 x' 轴和 y' 轴构成，两轴在直角坐标系中的斜率分别为 $T_x = \tan\theta_x$ 和 $T_y = \cot\theta_y$，也相交于原点 O.

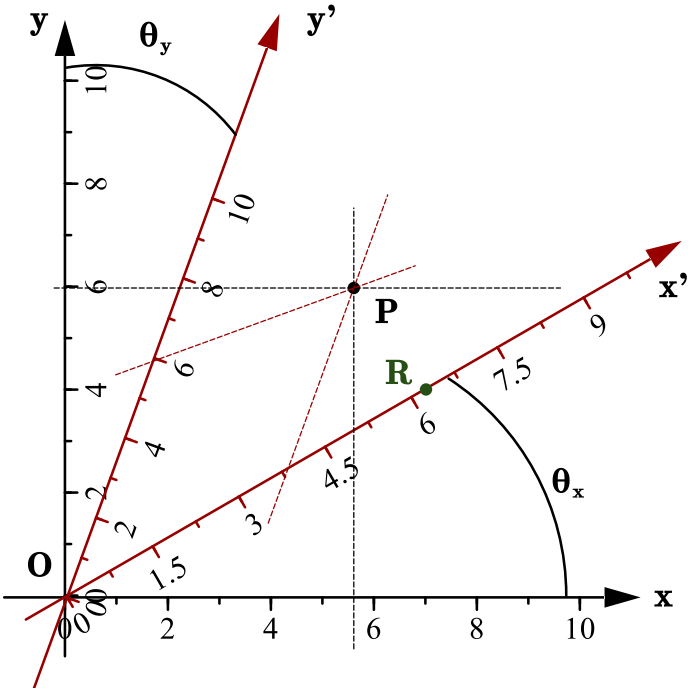


图 2: 本小节中所使用的坐标系和点示意图.

如果把两个斜坐标轴看成直角坐标轴中的直线，那么我们可以研究斜坐标

轴上两点之间的距离．假设取定 x' 轴上一点 R，在直角坐标系中，R 到 O 的距离是 $|RO|$．如果 x' 轴和 x 轴用相同的标尺，那么 R 在斜坐标系中的坐标应该是 $(|RO|,0)$．但是，如果 x' 轴和 x 轴所用的标尺不一样，那么 R 在斜坐标系中的坐标 $(R_{x'},0)$ 就不再等于 $(|RO|,0)$．这个时候，为了方便描述斜坐标系的标尺，我们引入一个新的概念：**拉伸比例**.

如果把 x' 轴看成可压缩和伸长的轴线，那么当它被拉伸时，其标尺“密度”会下降，也就是说，x' 轴上标尺数值为 1 的点到 O 的距离会越来越大．x' 轴伸长的比例，被称为 x' 轴的**拉伸比例**，大小为 $|RO|/R_{x'}$．这样一来，如果 x' 轴的**拉伸比例**是 r_x，那么在直角坐标系中 x' 轴上长度为 L 的线段，在斜坐标系中长度就是 L/r_x.

设有点 P，其直角坐标为 (x_0,y_0)，斜坐标为 (x_0',y_0')．在直角坐标系中，把两条斜坐标轴和两条平行斜坐标轴而穿过 P 的直线方程写出来，分别联立，可以解出 P 的**斜坐标投影**的直角坐标，进而得到投影点在直角坐标系中到 O 的长度，将其除以拉伸比例后即可得到相应的 x' 和 y' 坐标.

例 1　斜坐标系的转化

对于在直角坐标系中的斜率分别为 $T_x=\tan\theta_x$ 和 $T_y=\cot\theta_y$，也相交于原点 O 的斜坐标轴，若 x' 和 y' 的拉伸比例分别为 r_x 和 r_y，且点 P 的直角坐标为 (x_0,y_0)，求证 P 的斜坐标为 $(x_0',y_0')=(\frac{y_0-T_yx_0}{T_x-T_y}\cdot\frac{\sqrt{1+T_x^2}}{r_x},\frac{x_0-T_xy_0}{T_y-T_x}\cdot\frac{\sqrt{1+T_y^2}}{r_y})$

证明：

x' 轴在直角坐标系中的表达式为 $y=T_xx$，将它和过 P 且平行于 y' 轴的直线 $y=T_yx+y_0-T_yx_0$ 联立，解得 P 在 x' 轴上投影的位置 $(\frac{y_0-T_yx_0}{T_x-T_y},\frac{y_0-T_yx_0}{T_x-T_y}\cdot T_x)$，其到 O 的距离为 $\frac{y_0-T_yx_0}{T_x-T_y}\cdot\sqrt{1+T_x^2}$．考虑到拉伸比例 r_x，可知 P 在斜坐标系中的 x' 坐标为：$\frac{y_0-T_yx_0}{T_x-T_y}\cdot\sqrt{1+T_x^2}/r_x$.

同理可解得[9] P 在斜坐标系中的 y' 坐标为：$\frac{x_0-T_xy_0}{T_y-T_x}\cdot\sqrt{1+T_y^2}/r_y$.

证毕.

与几何向量的联系

把直角坐标系中的点 $P=(x_0,y_0)$ 可以看成是从原点 O 指向 P 的向量 $\overrightarrow{OP}$．令沿着 x 轴正方向的单位向量为 $\hat{\imath}$，沿着 y 轴正方向的单位向量为 $\hat{\jmath}$，那

[9] 也可以利用对称性，将上式中一切 x 和 y 对换得到.

么有 $\overrightarrow{OP}=x_{0\curvearrowright}+y_{o\curvearrowright}$.

从原点出发给定 x' 轴，y' 轴，以及它们的拉伸比例 r_x，r_y. 令沿着 x' 轴正方向的单位向量为 $_{\curvearrowright'}$，沿着 y' 轴正方向的单位向量为 $_{\curvearrowright'}$，点 P 在此斜坐标系中的坐标为 (x'_0,y'_0)，那么有 $\overrightarrow{OP}=x'_0r_{x\curvearrowright'}+y'_or_{y\curvearrowright'}$.

也就是说，斜坐标可以看成是以 $r_{x\curvearrowright'}$ 和 $r_{y\curvearrowright'}$ 作为基底来表达的 $\overrightarrow{OP}$ 的分量.

斜坐标系表示洛伦兹变换

预备知识 洛伦兹变换[1263]，斜坐标系[1265]

洛伦兹变换的斜坐标表示

一维空间的洛伦兹变换表述为

$$\begin{cases} x'=\dfrac{x-vt}{\sqrt{1-v^2/c^2}} \\ t'=\dfrac{t-vx/c^2}{\sqrt{1-v^2/c^2}} \end{cases} \tag{1}$$

如果把 (x,t) 作为直角坐标系，尝试用斜坐标系来表示 x' 和 t'，那么我们需要确定四个关键参数：斜坐标轴的斜率 T_x 和 T_t，以及其拉伸比例 r_x 和 r_t.

习题 1

如果用直角坐标表示铁轨系中的事件坐标，那么这些事件在火车系中的坐标可以一个斜坐标系来表示. 将一维空间洛伦兹变换的表达式和斜坐标系的坐标转换进行比较，推导出这个斜坐标系的坐标轴斜率和拉伸比例.

答案如图所示.

约化光速

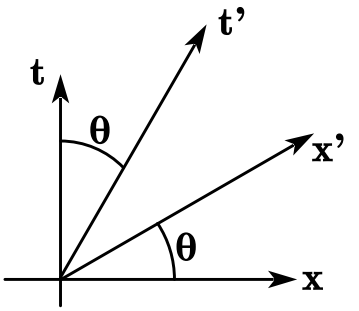


图 1: 洛伦兹变换的斜坐标表达，其中 $\tan\theta = v/c$，两坐标轴的拉伸比例均为 $\sqrt{(1+v^2/c^2)}/\sqrt{(1-v^2/c^2)}$.

预备知识 斜坐标系表示洛伦兹变换[1268]，无单位的物理公式[876]

约去量纲

物理量运算的规则有一个约束，合法的带单位多项式中，相加减的必然是相同的物理量．在无单位的物理公式[876] 中我们知道，当单位统一时，可以将物理量的单位和数值分开计算．

例 1　约去量纲的例子

不考虑相对论效应．一辆火车以速度 3.6 km/h 向东行驶，车上有一个小朋友以速度 4 m/s 向东奔跑，那么小朋友相对铁轨的速度就是 $3.6\times\frac{\mathrm{km}}{\mathrm{h}}+4\times\frac{\mathrm{m}}{\mathrm{s}}=3.6\times\frac{1000\cdot\mathrm{m}}{3600\,\mathrm{s}}+4\times\frac{\mathrm{m}}{\mathrm{s}}=(3.6\times\frac{1000}{3600}+4)\frac{\mathrm{m}}{\mathrm{s}}=5\frac{\mathrm{m}}{\mathrm{s}}=5\,\mathrm{m/s}$.

约去光速

如果取长度单位为 m，时间单位为 $\tau = 299792458\,\mathrm{s}$，那么光速就可以写为 $1\,\mathrm{m}/\tau$．用 τ 取代 s 作为时间单位，那么一切涉及光速的等式中，我们都可以把光速的**数值**写为 1，大大简化计算，而光速的量纲 $[\mathrm{m/s}] = [\mathrm{m}/\tau]$ 是独立于数值进行计算的．在这种写法中，0.1 倍光速就可以写为 $0.1\,\mathrm{m}/\tau$.

所以，在具体数值的计算中，使用上述单位制，则任何地方的光速的**数值**都可以视为 1．当然，任何物理量都可以通过选择适当的单位制，来让其数值上等于 1，但在狭义相对论中只有令光速的数值为 1 是最方便的，因为任何惯性系中光速都不变．

如果**任何**数值计算中都可以把光速设为 1，那么我们可以干脆把这个规律一般化，把光速本身就当成 1，没有量纲；或者说，这种情况下，我们把长度和时间看成同一个物理量，其换算关系是 $1\,\mathrm{m} = 299792458\,\mathrm{s}$，正如长度的换算关系 $1\,\mathrm{km} = 1000\,\mathrm{m}$ 一样．使用这种技巧，那么狭义相对论中的一切公式都会简洁得多．

当然，实际应用中我们还是要把长度和时间分开来看的，这就要求我们在适当的地方要能够把光速的具体数值代回去，使得**量纲和物理量相符，并且相加减的项具有相同的量纲**，方式是进行量纲分析．我将用例子解释如下：

例 2

- 如果约去光速后，某长度的表达式为 xv，那么在国际单位制下，这个长度应该是 xv/c.
- 如果 $tv^2 + x$ 表示某个长度，那么在国际单位制下，这个长度应该是 $tv^2/c + x$.
- 如果 $tv^2 + x$ 表示某个加速度，那么在国际单位制下，无论如何都没法把 c 添加回去以得到合适的量纲，因此这个表达式不可能表示加速度.

约去光速后洛伦兹变换的表达

约去光速后，洛伦兹变换变得高度对称．

$$\begin{cases} x' = \dfrac{x - vt}{\sqrt{1 - v^2}} \\ y' = y \\ z' = z \\ t' = \dfrac{t - vx}{\sqrt{1 - v^2}} \end{cases} \qquad \begin{cases} x = \dfrac{x' + vt'}{\sqrt{1 - v^2}} \\ y = y' \\ z = z' \\ t = \dfrac{t' + vx'}{\sqrt{1 - v^2}} \end{cases} \tag{1}$$

同时洛伦兹矩阵也变为：

$$L=\begin{bmatrix} \frac{1}{\sqrt{1-v^2}} & -\frac{v}{\sqrt{1-v^2}} & 0 & 0 \\ -\frac{v}{\sqrt{1-v^2}} & \frac{1}{\sqrt{1-v^2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2}$$

回归国际单位制也很简单．考虑到 x' 是长度，结合“相加减项必然有相同的量纲”，可以在适当位置添加光速以满足这两个条件，得到 $x'=\frac{x-vt}{\sqrt{1-v^2/c^2}}$．同样，考虑到 t' 是时间，可以得到 $t'=\frac{t-vx/c^2}{\sqrt{1-v^2/c^2}}$．

由此可见，约去光速的表达可以很自然地转化为国际单位制的表达，但是前者更加简洁优美，计算也不那么繁琐了．事实上，理论物理学家几乎不会真的用上 $c=299792458\,\mathrm{m/s}$，而是都默认 $c=1$．今后，我们将沿用 $c=1$ 的表达方法，约去 c；如果你不习惯的话，不妨在见到每一个表达式的时候，练习一下该如何适当添加 c 以回归国际单位制．

注意，在约去光速后，速度 v 的取值范围是 $[0,1]$，应理解为光速的倍数．

习题 1

将事件与尺缩效应[1256]、时间的变换与钟慢效应[1260]、洛伦兹变换[1263]和斜坐标系表示洛伦兹变换[1268]中的所有推导，在约去光速的条件下重新进行一遍，体会约去光速的便捷性[10]．

洛伦兹变换的代数推导

预备知识 约化光速[1268]，伽利略变换[??]

火车模型对于初学者建立狭义相对论的直观理解有很大帮助，但它在逻辑上其实较为繁琐，需要推导出尺缩效应，再以此推导出时间的变换，才能得到洛伦兹变换．本节采用更紧凑的初等代数方法重新推导一遍洛伦兹变换，更为直接，同时也可以当作约化光速表示法的一个实例．

[10]事实上，笔者在推导这些词条中的公式时，就是使用了约去光速的技巧，再在结果中把光速添加回去而得到的．

洛伦兹变换

为方便讨论，设 $t=0$ 时两参考系 S 与 S' 的原点重合，x,y,z 轴方向相同．S' 系相对 S 系沿 x 轴方向以速度 v 匀速运动．S 系中的坐标 (x,y,z,t) 一一对应到 S' 系中的坐标 (x',y',z',t')．每一个坐标等价于在这个坐标上发生的一个事件．

洛伦兹变换的另一个假设是，事件的四个坐标在不同惯性系中的变换一定是线性变换，这对应着我们在火车模型中所提到的，尺缩效应是均匀的，以及由此推导而得的时间变换的线性性．这也称作物理定律的平移对称及时间对称．

可以证明，满足上述要求的线性变换为

$$\begin{cases} x'=\dfrac{x-vt}{\sqrt{1-v^2/c^2}} \\ y'=y \\ z'=z \\ t'=\dfrac{t-vx/c^2}{\sqrt{1-v^2/c^2}} \end{cases} \qquad \begin{cases} x=\dfrac{x'+vt'}{\sqrt{1-v^2/c^2}} \\ y=y' \\ z=z' \\ t=\dfrac{t'+vx'/c^2}{\sqrt{1-v^2/c^2}} \end{cases} \tag{1}$$

该变换被称为**洛伦兹变换**．下面，我们就来具体推导这个变换．

推导

由于两个参考系是完全对称的，根据相对性原理，它们之间的坐标变换应该具有完全相同的形式．式 1 中的变换和逆变换有个别符号上的差异，因为习惯上的坐标系的定义使得 S' 相对 S 的速度是正的，而 S 相对于 S' 速度却是负的．为了保持绝对的镜像对称我们只需把 S' 的 x 轴取反方向．

先考虑 (x,t) 两个坐标，令线性变换为

$$\begin{cases} x'=ax+bt \\ t'=mx+nt \end{cases} \tag{2}$$

解得逆变换为

$$\begin{cases} x=\dfrac{nx'-bt'}{an-bm} \\ t=\dfrac{mx'-at'}{bm-an} \end{cases} \tag{3}$$

由相对性原理，正变换和逆变换的系数必须完全相同，对比系数并化简得

$$\begin{cases} a = -n \\ bm - an = 1 \end{cases} \tag{4}$$

为了书写方便，我们使用**约化光速**的表示[1268]，令 $c = 1$（无单位）.

由光速不变原理，当某点延 x 轴以光速运动，即 $\mathrm{d}x/\mathrm{d}t = 1$ 时，它在另一个参考系中的速度必须也是光速，即 $\mathrm{d}x'/\mathrm{d}t' = -1$（注意两个参考系中 x 轴方向相反）. 所以

$$\frac{\mathrm{d}x'}{\mathrm{d}t'} = \frac{a\,\mathrm{d}x + b\,\mathrm{d}t}{m\,\mathrm{d}x + n\,\mathrm{d}t} = \frac{a+b}{m+n} = -1 \tag{5}$$

根据定义，S 系中的任意不动点（$\mathrm{d}x/\mathrm{d}t = 0$）在 S' 系中速度为 v，所以

$$\frac{\mathrm{d}x'}{\mathrm{d}t'} = \frac{a\,\mathrm{d}x + b\,\mathrm{d}t}{m\,\mathrm{d}x + n\,\mathrm{d}t} = \frac{b}{n} = v \tag{6}$$

根据两坐标系定义，我们要求 t 不变时，x 增加 x' 减少，所以

$$a < 0 \tag{7}$$

联立式 4 到式 7 可解得所有系数

$$a = -\frac{1}{\sqrt{1-v^2}} \qquad b = \frac{v}{\sqrt{1-v^2}} \qquad m = -\frac{v}{\sqrt{1-v^2}} \qquad n = \frac{1}{\sqrt{1-v^2}} \tag{8}$$

所以变换和逆变换为

$$\begin{cases} x' = -\dfrac{x+vt}{\sqrt{1-v^2}} \\ t' = \dfrac{t+vx}{\sqrt{1-v^2}} \end{cases} \qquad \begin{cases} x = -\dfrac{x'+vt'}{\sqrt{1-v^2}} \\ t = \dfrac{t'+vx'}{\sqrt{1-v^2}} \end{cases} \tag{9}$$

若按照一般的坐标系习惯，把 x' 改成 $-x'$，就与式 1 一致了[11].

下面再来考虑其他维度，设

$$\begin{cases} x' = \dfrac{x-vt}{\sqrt{1-v^2}} + ay + bz \\ t' = \dfrac{t-vx}{\sqrt{1-v^2}} + cy + ez \end{cases} \tag{10}$$

[11] 要从该式得到国际单位的公式，只需把 x, x', v, v' 都替换为国际单位的物理量除以转换常数 c 即可

求逆变换，对比系数，得出 a,b,c,e 为零．同理可得

$$\begin{cases} y' = y \\ z' = z \end{cases} \tag{11}$$

至此我们就完全推导出了洛伦兹变换．

相对论速度变换

预备知识 斜坐标表示洛伦兹变换[1268]

一维空间情况

沿用铁轨系 K_1 和火车系 K_2 的设定，令 K_2 相对 K_1 以速度 v 运动，以直角坐标系表示 K_1，相应地用斜坐标系表示 K_2．如果一个点在 K_1 中以速度 u 运动，那么它在直角坐标系中扫过一条直线，称为它的**世界线**；这在斜坐标系中也算一条直线．因此，这个点在 K_2 中也是以匀速运动的，记为 u'．我们希望计算出 u' 相对于 u 的关系．

几何解法

如图所示，K_1 系的坐标为 x 和 t，K_2 系的坐标为 x' 和 t'，由题设有 $\tan\theta = v$．图中蓝线表示所讨论点的运动轨迹，因此 $\tan\varphi = u$．我们的目标是计算出蓝线在 K_2 中的斜率，也就是 $\frac{a}{b}$．

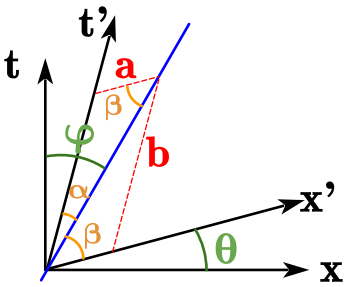


图 1: 几何解相对论速度变换的示意图．

由正弦定理，$\frac{a}{b}=\frac{\sin\alpha}{\sin\beta}$．其中 $\alpha=\varphi-\theta$，$\beta=\frac{\pi}{2}-\theta-\varphi$．则

$$\begin{aligned}\frac{\sin\alpha}{\sin\beta}&=\frac{\sin(\varphi-\theta)}{\sin(\frac{\pi}{2}-\theta-\varphi)},\\&=\frac{\sin(\varphi-\theta)}{\cos(\varphi+\theta)},\\&=\tan(\varphi-\theta)\cdot\frac{\cos(\varphi-\theta)}{\cos(\varphi+\theta)},\\&=\tan(\varphi-\theta)\cdot\frac{\cos\varphi\cos\theta+\sin\varphi\sin\theta}{\cos\varphi\cos\theta-\sin\varphi\sin\theta},\\&=\tan(\varphi-\theta)\cdot\frac{1+\tan\varphi\tan\theta}{1-\tan\varphi\tan\theta},\\&=\frac{\tan\varphi-\tan\theta}{1-\tan\varphi\tan\theta},\\&=\frac{u-v}{1-uv}\end{aligned}\tag{1}$$

也就是说，当物体在 K_1 里以速度 u 运动时，在相对 K_1 以速度 v 运动的火车 K_2 看来，其速度是 $u'=(u-v)/(1-uv)$．

这就是一维空间里，物体在两个参考系之间的速度变换公式．

代数解法

设物体在 K_1 中的运动轨迹是 $x=ut+c$，其中 c 是一个常数．那么利用洛伦兹变换将 x 和 t 用 x' 和 t' 表示，我们有

$$\frac{x'+vt'}{\sqrt{1-v^2}}=u\cdot\frac{t'+vx'}{\sqrt{1-v^2}}+c\tag{2}$$

整理得

$$x'=\frac{u-v}{1-uv}t'+c\sqrt{1-v^2}/(1-uv)\tag{3}$$

因此 $u'=(u-v)/(1-uv)$．

分析解法

以上几何解法和代数解法都是对于在两个参考系中都匀速运动的点而言的．事实上，这一速度变换公式也可以用在任意运动状态的点上．

考虑到洛伦兹变换和一阶微分的形式不变性[12]，可以有：

$$u' = \frac{dx'}{dt'} = \frac{dx - vdt}{dt - vdx} = \frac{\frac{dx}{dt} - v}{1 - v\frac{dx}{dt}} = \frac{u - v}{1 - uv} \tag{4}$$

和前面的结果一致.

三维空间情况

现在我们尝试推导所有方向上的速度变化，而不仅仅是一维空间里，物体速度和火车速度在同一方向上的时候.

假设 K_2 相对 K_1 的运动速度是 $\begin{pmatrix} v, & 0, & 0 \end{pmatrix}^{\mathrm{T}}$. 设在 K_1 中，有一质点以速度 $\begin{pmatrix} u_x, & u_y, & u_z \end{pmatrix}^{\mathrm{T}}$ 运动，其在 K_2 中的速度是 $(u'_x, \quad u'_y, \quad u'_z)^{\mathrm{T}}$.

选取 K_1 和 K_2 的原点位置，使得两原点重合且该质点的轨迹穿过原点. 那么该质点的时空轨迹，在 K_1 中的表达就是：

$$\begin{pmatrix} t, & u_x t, & u_y t, & u_z t \end{pmatrix}^{\mathrm{T}} \tag{5}$$

对式 5 进行洛伦兹变换，就得到该质点在 K_2 中的时空轨迹：

$$\begin{pmatrix} t', & u'_x t', & u'_y t', & u'_z t' \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} \frac{t - vu_x t}{\sqrt{1-v^2}}, & \frac{u_x t - vt}{\sqrt{1-v^2}}, & u_y t, & u_z t \end{pmatrix}^{\mathrm{T}} \tag{6}$$

如果只对 t' 进行洛伦兹变换并将它代入 $u'_\alpha t'$（$\alpha \in \{x, y, z\}$），那么有

$$\begin{pmatrix} t', & u'_x t', & u'_y t', & u'_z t' \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} \frac{t - vu_x t}{\sqrt{1-v^2}}, & u'_x \frac{t - vu_x t}{\sqrt{1-v^2}}, & u'_y \frac{t - vu_x t}{\sqrt{1-v^2}}, & u'_z \frac{t - vu_x t}{\sqrt{1-v^2}} \end{pmatrix}^{\mathrm{T}} \tag{7}$$

令式 6 和式 7 的右边相等，可以解出：

$$\begin{pmatrix} u'_x, & u'_y, & u'_z \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} \frac{u_x - v}{1 - u_x v}, & u_y \cdot \frac{\sqrt{1-v^2}}{1 - u_x v}, & u_z \cdot \frac{\sqrt{1-v^2}}{1 - u_x v} \end{pmatrix}^{\mathrm{T}} \tag{8}$$

这就是完整的速度变换公式.

[12] 即如果 $y = f(x)$，那么总有 $dy = f'(x)dx$. 更高阶的微分是不能这样写的，比如 $d^2y = f''(x)dx^2$ 就不可以.

速度变换的向量表达

上面所得出的完整速度变换公式，显式地写出了各分量的值．我们也可以用更紧凑的表达方式，将 $\left(u_x,\quad u_y,\quad u_z\right)^{\mathrm{T}}$ 和 $\left(u'_x,\quad u'_y,\quad u'_z\right)^{\mathrm{T}}$ 分别看成向量 $\mathbf{u}$ 和 $\mathbf{u}'$，那么可以把速度变换表示成

$$\mathbf{u}'=\frac{\sqrt{1-v^2}\mathbf{u}-\frac{1-\sqrt{1-v^2}}{v^2}(\mathbf{u}\cdot\mathbf{v})\mathbf{v}-\mathbf{v}}{1-\mathbf{u}\cdot\mathbf{v}} \tag{9}$$

如果按照惯例，记 $\gamma=\frac{1}{\sqrt{1-v^2}}$，则有

$$\mathbf{u}'=\frac{\sqrt{1-v^2}\mathbf{u}-\frac{\gamma}{\gamma+1}(\mathbf{u}\cdot\mathbf{v})\mathbf{v}-\mathbf{v}}{1-\mathbf{u}\cdot\mathbf{v}} \tag{10}$$

相对论加速度变换

预备知识 速度变换[1274]

一般情况下的加速度变换

问题的限制条件

为了简化讨论，不失一般性，我们可以把情景按以下方式设定：

假设 K_2 相对 K_1 的运动速度是 $\mathbf{v}=\begin{pmatrix}v\\0\\0\end{pmatrix}$．设在 K_1 中，有一质点以速度 $\mathbf{u}=\begin{pmatrix}u_x\\u_y\\u_z\end{pmatrix}$ 运动，其在 K_2 中的速度是 $\mathbf{u}'=\begin{pmatrix}u'_x\\u'_y\\u'_z\end{pmatrix}$．同时，质点在 K_1 中有加速度 $\mathbf{a}=\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}=\begin{pmatrix}a_x\\a_y\\a_z\end{pmatrix}$，在 K_2 中加速度则为 $\mathbf{a}'=\frac{\mathrm{d}}{\mathrm{d}t'}\mathbf{u}'=\begin{pmatrix}a'_x\\a'_y\\a'_z\end{pmatrix}$．选取两坐标系的原点位置使它们重合、在质点轨迹上，并且设质点一直以匀速运动，直到在 K_1 中测量的时间 t 时才有了非零加速度．

计算的整体思路是，将 $\frac{\mathrm{d}}{\mathrm{d}t'}\mathbf{u}'$ 拆成 $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}'\cdot\frac{\mathrm{d}t}{\mathrm{d}t'}$，分别计算两个微商，然后再乘起来．一阶微分的形式不变性保证了这个思路的合法性．

注意，这里并不能简单地令 $t'=\sqrt{1-v^2}t$，因为质点的位置并不总是在 K_2 的原点处，我们需要选所讨论位置的 t'．

于是有

$$\mathbf{u}'=\frac{\sqrt{1-v^2}-\frac{1-\sqrt{1-v^2}}{v^2}(\mathbf{u}\cdot\mathbf{v})\mathbf{v}-\mathbf{v}}{1-\mathbf{u}\cdot\mathbf{v}} \tag{1}$$

$$t'=\frac{t-vu_xt}{\sqrt{1-v^2}} \tag{2}$$

从式 2 可得，

$$\frac{\mathrm{d}t}{\mathrm{d}t'}=\frac{1}{\mathrm{d}t'/\mathrm{d}t}=\frac{\sqrt{1-v^2}}{1-u_xv} \tag{3}$$

习题 1

已知对于向量 $\mathbf{A}$，$\mathbf{B}$ 和 $\mathbf{C}$ 有恒等式：$\mathbf{C}\times(\mathbf{AB})=(\mathbf{C}\cdot\mathbf{B})\mathbf{A}-(\mathbf{C}\cdot\mathbf{A})\mathbf{B}$，结合式 1，请证明：

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}'=\frac{1-v^2}{(1-\mathbf{u}\cdot\mathbf{v})^2}[\mathbf{a}-\frac{1}{\sqrt{1-v^2}+1-v^2}(\mathbf{a}\times\mathbf{v})\times\mathbf{v}+\sqrt{1-v^2}(\mathbf{a}\times\mathbf{u})\times\mathbf{v}] \tag{4}$$

将式 3 和式 4 相乘，即得到

$$\mathbf{a}'=\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}'\cdot\frac{\mathrm{d}t}{\mathrm{d}t'}=\frac{(1-v^2)^{3/2}}{(1-\mathbf{uv})^3}[\mathbf{a}-\frac{1}{\sqrt{1-v^2}+1-v^2}(\mathbf{a}\times\mathbf{v})\times\mathbf{v}+\sqrt{1-v^2}(\mathbf{a}\times\mathbf{u})\times\mathbf{v}] \tag{5}$$

同样地，由于在 K_2 眼中，质点以 $\mathbf{u}'$ 运动且 K_1 以 $-\mathbf{v}$ 运动，因此可以得到逆变换：

$$\mathbf{a}=\frac{(1-v^2)^{3/2}}{(1+\mathbf{u}'\mathbf{v})^3}[\mathbf{a}'-\frac{1}{\sqrt{1-v^2}+1-v^2}(\mathbf{a}'\times\mathbf{v})\times\mathbf{v}-\sqrt{1-v^2}(\mathbf{a}'\times\mathbf{u}')\times\mathbf{v}] \tag{6}$$

瞬时自身系中的变换

对于运动的质点，和质点一直相对静止的参考系称为质点的**自身系**．如果质点在惯性系中有非零加速度，那么自身系就不是惯性系，不在狭义相对论的

讨论范围内. 但是，对于质点轨迹上的给定点，我们可以取质点在这里的速度，然后选择一个在这一瞬间和质点相对静止的惯性系. 这样在某一点处和质点相对静止的参考系，称为质点的**瞬时自身系**.

在本节讨论中，如果设 $\mathbf{u}=\mathbf{v}$，即 $\mathbf{u}'=\mathbf{0}$，那么在加速度出现时，K_2 就是质点在这一事件位置的瞬时自身系. 质点在瞬时自身系中的加速度是 $\mathbf{a}'$，那么由式 6，其在 K_1 中的加速度是

$$\mathbf{a}=(1-v^2)^{3/2}[\mathbf{a}'-\frac{1}{\sqrt{1-v^2}+1-v^2}(\mathbf{a}'\times\mathbf{v})\times\mathbf{v}] \tag{7}$$

如果 $\mathbf{a}'\parallel\mathbf{v}$，则有

$$\mathbf{a}=(1-v^2)^{3/2}\mathbf{a}' \tag{8}$$

因此 $\mathbf{a}\parallel\mathbf{v}$.

如果 $\mathbf{a}'\perp\mathbf{v}$，则有

$$\mathbf{a}=(1-v^2)\mathbf{a}' \tag{9}$$

因此 $\mathbf{a}\perp\mathbf{v}$.

时空的四维表示

预备知识 斜坐标表示洛伦兹变换[1268]，坐标变换与过渡矩阵

概念

四位置和四位移

把时间坐标和空间坐标都看成时空的坐标，那么一个事件在时空中的位置就被称为其**四位置（4-position）**. 四位置本身是一个向量，但其坐标表示取决于所选取的惯性参考系. 同一个四位置在不同惯性系中的坐标，可以利用洛伦兹矩阵作为过渡矩阵来相互转化.

两个事件之间的四位置之差，称为这两个事件之间的**四位移（4-displacement）**.

四速度

假设某质点在三维空间中运动. 经典物理中认为，质点轨迹上某一点的速度是一个向量，其方向与该点处轨迹相切. 但速度的大小具体是多少，三维轨迹完全没有提供足够的信息. 同样的轨迹完全可以是用不同的瞬时速度来走过的.

为了完全从几何角度描述该质点的速度，我们可以把视角提高到四维时空，将三维的轨迹拉升到四维，这样就有充足的信息来描述质点的速度大小了，而三维轨迹就是四维轨迹在三维空间中的投影. 取四维轨迹上某一点的切向量 $\mathbf{v}$，使其在时间轴上的投影为 1（单位时间长度），那么 $\mathbf{v}$ 在三维空间中的投影就是速度.

这个例子启发我们研究四维速度比研究三维速度更加全面，由此有了以下概念：

定义 1　四速度

一个质点的四速度（**4-velocity**），定义为质点的四维运动轨迹的切向量 $\mathbf{v}$，满足 $\mathbf{v}$ 在瞬时自身系中的时间轴投影长度是 1.

四速度的概念可以应用在经典力学中，也可以应用于相对论.

相对论框架下，在某惯性参考系里的观察者看来，一个质点的四速度的时间分量，就是这个质点系所处的参考系的“时间流逝速度”，即质点的手表和观察者的手表的转动速度之比[13]. 如果质点相对观察者静止，那么质点的四速度就是 $(1,0,0,0)$，即非零分量只有时间分量，且时间流逝速度为 1.

如果在经典力学框架下讨论四速度，那么其时间分量就永远是 1，这其实意味着经典情况下不存在钟慢效应.

四加速度

定义 2　四加速度

一个质点的四加速度（**4-acceleration**），定义为质点的四速度对固有时间（瞬时自身系中的时间）求导的结果.

[13] 由于我们采用 $c=1$ 的约定，速度是一个无量纲量.

四位移、四速度与四加速度的性质

四位移的不变性

习题 1

设有惯性参考系 K_1，另一惯性系 K_2 以速度 $\mathbf{v}=\left(v,0,0\right)^T$ 相对 K_1 运动. 请通过计算证明：对于两个事件 A 和 B 而言，它们在两个参考系中的四位移都是一样的.

四速度的表达

设有惯性参考系 K_1，一个质点在 K_1 中某点的速度是 $\mathbf{u}=\left(u,0,0\right)^T$，那么它在 K_1 的四速度 $\mathbf{U}$ 的**方向**和 $\left(1,u,0,0\right)^T$ 一致[14]，设 $\mathbf{U}=k\cdot\left(1,u,0,0\right)^T$. 则由于 $\mathbf{U}$ 对应瞬时自身系 K_2 中单位时间的长度，再考虑到 K_2 的坐标轴在 K_1 中的表示具有**拉伸比例** $\sqrt{(1+u^2)}/\sqrt{(1-u^2)}$，知 $\mathbf{U}$ 在 K_1 中的长度为 $\sqrt{(1+u^2)}/\sqrt{(1-u^2)}$，其在时间轴上的投影就是 $1/\sqrt{(1-u^2)}=\gamma$.

因此，速度为 $\mathbf{u}=\left(u,0,0\right)^T$ 的质点，其四速度是 $\mathbf{U}=\gamma\left(1,u,0,0\right)^T=\left(\gamma,\gamma u,0,0\right)^T=\left(1/\sqrt{1-u^2},u/\sqrt{1-u^2},0,0\right)^T$.

四加速度

题设同“四加速度的表达”，设质点在某点的四加速度是 $\mathbf{A}=\mathrm{d}\mathbf{U}\,/\mathrm{d}\tau$，其中 τ 是质点的瞬时自身系 K_2 中的固有时. 记 $\mathbf{U}=(t,\mathbf{u})$，$\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t}=\mathbf{a}$. 由于在 K_1 中 $\mathrm{d}t\,/\,\mathrm{d}\tau=1/\sqrt{1-u^2}=\gamma$，再考虑到分部积分和 $\mathrm{d}\gamma\,/\,\mathrm{d}t=\mathbf{a}\cdot\mathbf{u}/\gamma^3$，可以得：

$$
\begin{aligned}
\mathbf{A}&=\frac{\mathrm{d}\mathbf{U}}{\mathrm{d}\tau}\\
&=\frac{\mathrm{d}\mathbf{U}}{\mathrm{d}t}\cdot\frac{\mathrm{d}t}{\mathrm{d}\tau}\\
&=(\gamma^4\mathbf{a}\cdot\mathbf{u},\gamma^2\mathbf{a}+\gamma^4(\mathbf{a}\cdot\mathbf{u})\mathbf{u})\\
&=(\gamma^4\mathbf{a}\cdot\mathbf{u},\gamma^4(\mathbf{a}+(\mathbf{a}\times\mathbf{u})\times\mathbf{u})
\end{aligned}
\tag{1}
$$

[14]也就是和瞬时自身系 K_2 的时间轴重合.

世界线

我们知道，一个事件被看作是四维时空中的一个点，它有给定的时间和空间坐标．如果我们考虑一个粒子的运动，其在时空中的轨迹就是一条曲线．这条曲线就被称为粒子的**世界线（world line）**.

闵可夫斯基空间

预备知识 时空的四维表示[1279]，度量空间[785]

时空间隔

经典力学讨论的时空是一个欧几里得空间，不同惯性系之间的伽利略变换是一个保度量变换，即时空点之间的度量在变换前后不变．但是相对论下的欧几里得度量不再是不变的，两个时空点之间的度量，在洛伦兹变换后一般会变化．不过相对论里有另一个类似欧几里得度量的东西，它在洛伦兹变换下保持不变.

习题 1　洛伦兹不变量

给定参考系 K_1 和 K_2，其中 K_2 相对 K_1 以速度 $\mathbf{v}=(v,0,0)^T$ 运动．假设有两个事件，它们在 K_1 中的四位置分别为 (t_1,x_1,y_1,z_1) 和 (t_2,x_2,y_2,z_2)，在 K_2 中的四位置分别为 (t_1',x_1',y_1',z_1') 和 (t_2',x_2',y_2',z_2')．请利用洛伦兹变换，通过计算证明

$$\begin{aligned}&(t_1-t_2)^2-(x_1-x_2)^2-(y_1-y_2)^2-(z_1-z_2)^2\\=&(t_1'-t_2')^2-(x_1'-x_2')^2-(y_1'-y_2')^2-(z_1'-z_2')^2\end{aligned}\tag{1}$$

由习题 1 可知，在任何惯性参考系中，$\sqrt{-(t_1-t_2)^2+(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2}$ 都是不变的．我们把这个量记为 ΔS，称为两事件之间的**时空间隔（interval）**.

闵可夫斯基度规

我们知道，n 维欧几里得空间是 n 个实数集合的笛卡尔积，配上用勾股定理定义的度量所得到的．这样的度量，在物理学上称为欧几里得度规．闵可夫

斯基给四维时空定义了一个新的度规，即时空间隔．以时空间隔作为度规的四维空间，被称为闵可夫斯基空间．

时空间隔并不是数学意义上的**度量**（定义 1 [785]），因为它无法满足正定性，三角不等式也普遍不成立或不存在[15]．

另一种看待闵可夫斯基度规的方法是，把四维时空看成复数集合的笛卡尔积，依然使用勾股定理所定义的度量，但是把时间看成必须是纯虚数．也就是说，事件的时间为 t 的时候，其在时间轴上的坐标是 $\mathrm{i}t$．这样一来，就可以把事件 $(\mathrm{i}t, x, y, z)$ 的范数写成 $\sqrt{(\mathrm{i}t_1 - \mathrm{i}t_2)^2 + (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$，形式上和实数欧几里得空间的欧几里得范数是一样的，只不过用的是复数；进而得到形式上的“复”欧几里得度量．需要特别注意的是，闵可夫斯基度规并不能看成酉空间中的四维复数度量．

光的多普勒效应

预备知识 多普勒效应[1016]，洛伦兹变换[1263]

不论是牛顿力学框架下还是相对论意义下，光的频率都可能在不同参考系之间有所不同．由于观察者和光源之间相对运动，造成观察者所测得的光的频率与光源的频率不一致，这个现象被称为光的多普勒效应．由于光速不变原理，在相对论框架下讨论光的多普勒效应反而更为简单．

光的多普勒效应的推导

由于在任何参考系中，光的速度都一样，因此只需要知道光的波长就可以根据以下公式得到光的频率：

$$f = \frac{1}{\lambda} \tag{1}$$

其中，f 是光的频率，λ 是光的波长，1 是光速．

假设光源所在的参考系为 K_1，观察者所在的参考系为 K_2，而 K_2 相对 K_1 的运动速度为 $\vec{v} = (v, 0, 0)^T$．设光的频率在 K_1 为 f，在 K_2 中为 f'，那么光的波长在 K_1 中为 $\lambda = 1/f$，在 K_2 中为 $\lambda' = 1/f'$．

[15]时空间隔不是实数的时候，无法比较大小．

一维情况

设在 K_1 看来，观察者和光的运动方向是一致的.

取 K_1 中光的两个相邻的波峰[16]，作为两个以光速运动的点. 为简化讨论，不妨设其中一个点的世界线[17]表示为

$$\begin{pmatrix} t_1 \\ t_1 \\ 0 \\ 0 \end{pmatrix} \tag{2}$$

而另一个点为

$$\begin{pmatrix} t_2 \\ t_2+\lambda \\ 0 \\ 0 \end{pmatrix} \tag{3}$$

代入洛伦兹变换可得，在 K_2 中两波峰的世界线分别为

$$\begin{pmatrix} \frac{t_1-vt_1}{\sqrt{1-v^2}} \\ \frac{t_1-vt_1}{\sqrt{1-v^2}} \\ 0 \\ 0 \end{pmatrix} \tag{4}$$

和

$$\begin{pmatrix} \frac{t_2-v(t_2+\lambda)}{\sqrt{1-v^2}} \\ \frac{(t_2+\lambda)-vt_2}{\sqrt{1-v^2}} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{t_2-vt_2-v\lambda}{\sqrt{1-v^2}} \\ \frac{t_2-vt_2+\lambda}{\sqrt{1-v^2}} \\ 0 \\ 0 \end{pmatrix} \tag{5}$$

在各参考系中计算波长，就要计算对应参考系中同一时间下两波峰的空间坐标的差. 对于 K_1，“同一时间”意味着 $t_1=t_2$；对于 K_2，“同一时间”意味着 $t_1-vt_1=t_2-v(t_2+\lambda)$，

[16] 也可以是相邻波谷，或者任何相邻的两个同相位的点.

[17] 见**时空的四维表示**[1279].

将“同一时间”条件分别代入两个参考系中两个点的世界线，消去 t_1 计算后发现，两波峰在 K_1 中的距离始终是 λ，而在 K_2 中的距离始终是 $\frac{(1+v)\lambda}{\sqrt{1-v^2}}$. 由于已经设定这束光在 K_2 中的波长是 λ'，因此有

$$\lambda' = \frac{(1+v)\lambda}{\sqrt{1-v^2}} = \frac{\sqrt{1+v}}{\sqrt{1-v}}\lambda \tag{6}$$

因此在 K_2 中，光的频率变为

$$f' = \frac{1}{\lambda'} = \frac{\sqrt{1-v}}{\sqrt{1+v}\lambda} = \frac{\sqrt{1-v}}{\sqrt{1+v}}f \tag{7}$$

这就是一维情况下光的多普勒效应.

三维情况

三维情况下，我们要考虑在 K_1 看来光的传播方向和 K_2 的运动方向不一致的情况.

只要在垂直 x 轴的方向上，y 轴和 z 轴的方向可以任意选取，因此不妨设在 K_1 看来，光的传播方向在 $x-y$ 平面的第一象限，与 x 轴夹角为 θ.

类似一维情况中，取相邻两个波峰，考虑它们在 K_1 中的世界线：

$$\begin{pmatrix} t_1 \\ t_1\cos\theta \\ t_1\sin\theta \\ 0 \end{pmatrix} \tag{8}$$

和

$$\begin{pmatrix} t_2 \\ (t_2+\lambda)\cos\theta \\ (t_2+\lambda)\sin\theta \\ 0 \end{pmatrix} \tag{9}$$

代入洛伦兹变换，得到它们在 K_2 中的世界线：

$$\begin{pmatrix} \frac{t_1 - vt_1\cos\theta}{\sqrt{1-v^2}} \\ \frac{t_1\cos\theta - vt_1}{\sqrt{1-v^2}} \\ t_1\sin\theta \\ 0 \end{pmatrix} \tag{10}$$

和

$$\begin{pmatrix} \frac{t_2-v(t_2+\lambda)\cos\theta}{\sqrt{1-v^2}} \\ \frac{(t_2+\lambda)\cos\theta-vt_2}{\sqrt{1-v^2}} \\ (t_2+\lambda)\sin\theta \\ 0 \end{pmatrix} \tag{11}$$

在 K_2 中，“同一时间”条件为 $t_1 - vt_1\cos\theta = t_2 - v(t_2+\lambda)\cos\theta$. 类似一维情况，代入该条件并求两波峰的空间坐标之差，所得到的差是一个向量

$$\lambda \begin{pmatrix} \frac{\cos\theta\sqrt{1-v^2}}{1-v\cos\theta} \\ \frac{\sin\theta}{1-v\cos\theta} \\ 0 \end{pmatrix} \tag{12}$$

该向量的模长就是

$$\lambda' = \lambda \cdot \frac{\cos\theta\sqrt{1-v^2}}{1-v\cos\theta} \tag{13}$$

因此在 K_2 中，光的频率变为

$$f' = \frac{1}{\lambda'} = \frac{1-v\cos\theta}{\cos\theta\sqrt{1-v^2}} f \tag{14}$$

这就是三维空间中一般情况下光的多普勒效应.

如果 $\theta = 0$，那么情况退化为一维空间的问题. 作为验证将 $\theta = 0$ 代入式 14 后结果和式 7 一致；事实上，容易验证，三维情况下的所有式子在 $\theta = 0$ 时都退化为对应的一维情况下的式子.

洛伦兹群

预备知识　群[534]，李群，算符对易与共同本征函数[1381]，闵可夫斯基空间[1282]，协变和逆变[504]

洛伦兹变换

物质运动可以看成一连串事件的发展过程. 在四维时空里，每个事件都有相应的坐标. 对于同一事件，在惯性系 A 里用 $(t, \mathbf{x})$ 表示，在惯性系 A' 里用

$(t', \mathbf{x}')$ 表示．这就是四维时空的四矢量，洛伦兹变换就是作用在一个惯性系的四维矢量上，使之变为另一个惯性系的四矢量．洛伦兹变换必须是线性的，譬如一个惯性系中的匀速运动，换了参考系也依然是匀速运动．从 $(t, \mathbf{x})$ 到 $(t', \mathbf{x}')$ 的线性变换是保证一切物理定律在不同惯性系下形式不变的基本要求．

接下来我们考虑两个特殊事件．设惯性系 A 和 A' 在 $t = t' = 0$ 时在原点重合，此时原点发出一股光信号，该事件为 $(0, 0, 0, 0)$．在此后的某个时刻，P 点接收一光信号．在惯性系 A 里该事件为为 (t, x, y, z), 在惯性系 A' 里该事件为 (t', x', y', z')．采取自然单位制 $c = 1$．那么我们有

$$\begin{aligned} t^2 - (x^2 + y^2 + z^2) &= 0 \\ t'^2 - (x'^2 + y'^2 + z'^2) &= 0 \end{aligned} \tag{1}$$

对于一般的两个事件，它们之间可能并不是用光信号联系，甚至可以没有联系．所以以上两个二次式的结果取值是任意的．通过线性变换，可以把上述方程里参考系 A' 中的二次式变为关于 (t, x, y, z) 的二次式．然而由于初始事件在原点重合，这个线性变换只能是

$$t^2 - (x^2 + y^2 + z^2) = C\left[(t'^2 - (x'^2 + y'^2 + z'^2)\right] \tag{2}$$

其中 C 是常数因子．由于空间中不存在特殊方向，所以 C 只与两惯性系的相对速度大小有关．反过来，我们也有

$$t'^2 - (x'^2 + y'^2 + z'^2) = C\left[(t'^2 - (x'^2 + y'^2 + z'^2)\right] \tag{3}$$

解得 $C^2 = 1$，由变换的连续性我们取 $C = 1$．称式 1 的二次式为**时空间隔**，则洛伦兹变换是不改变时空间隔的线性变换，即

$$s^2 = t'^2 - (x'^2 + y'^2 + z'^2) = t^2 - (x^2 + y^2 + z^2) \tag{4}$$

洛伦兹群

洛伦兹变换的集合形成**洛伦兹群（Lorentz Group）**．采取闵可夫斯基空间的度规张量 $g_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$，引入时空坐标四矢量标记

$$\begin{aligned} x^\mu &= \left(x^0, x^i\right) \quad (i = 1, 2, 3) \\ &= (t, \mathbf{x}) \end{aligned} \tag{5}$$

时空间隔可以表示为

$$\begin{aligned} S^2 &= x^0x^0 - x^ix^i = x^\mu x^\nu g_{\mu\nu} \\ &= x'^0x'^0 - x'^ix'^i = x'^\mu x'^\nu g_{\mu\nu} \end{aligned} \tag{6}$$

设联系两个惯性系的洛伦兹矩阵为 Λ, 矩阵元为 $\Lambda^\mu_\nu \quad (\mu,\nu = 0,1,2,3)$. 洛伦兹变换可以表示为

$$x'^\mu = \Lambda^\mu_\nu x^\nu = \Lambda^\mu_0 x^0 + \Lambda^\mu_i x^i \tag{7}$$

代入式 6 我们有

$$x^\mu x^\nu g_{\mu\nu} = x^\rho x^\sigma g_{\rho\sigma} \tag{8}$$

保持时空距离不变的要求导致

$$g_{\rho\sigma} = g_{\mu\nu}\Lambda^\mu_\rho\Lambda^\nu_\sigma \tag{9}$$

考虑一个无穷小变换

$$\Lambda^\mu_\nu = \delta^\mu_\nu + \omega^\mu_\nu \tag{10}$$

代入式 9 得到 $\omega_{\mu\nu} = -\omega_{\nu\mu}$ 所以洛伦兹群有六个独立参元素，分别是三个转动自由度和三个平动自由度.

下面我们来探讨洛伦兹群的性质. 四维矢量的矢量积与时空坐标四矢量一样，在洛伦兹变换下不变. 设任意四维矢量 X，这意味着

$$X^TX = (\Lambda X)^T(\Lambda X) = X^T(\Lambda^T\Lambda)X \tag{11}$$

可见，洛伦兹群是正交群. 将洛伦兹群记为 $\mathcal{O}(3,1)$, 其中 $\mathcal{O}$ 为 “$Orthogonal$” 正交的意思.

未完成：正交群引用

分类

设洛伦兹矩阵用 Λ 表示，可以证明洛伦兹矩阵的行列式 $\det\Lambda = \pm 1$.

取洛伦兹矩阵的 00 分量，根据式 9 有

$$1 = g_{\mu\nu}\Lambda^0_\rho\Lambda^0_\sigma = (\Lambda^0_0)^2 - (\Lambda^0_i)^2 \tag{12}$$

所以 $|\Lambda^0_0| \geqslant 1$. 根据 Λ^0_0 的范围（$\Lambda^0_0 \geqslant 1$ 是 orthochronous 的洛伦兹变换）和 Λ（$\det\Lambda = 1$ 是 proper 的变换）的行列式可以将洛伦兹变换分为四类

1. proper orthochronous($\Lambda_+^\uparrow$)，对应 $\det\Lambda=1\quad \Lambda_0^0\geqslant 1$.

2. proper non-orthochronous($\Lambda_+^\downarrow$)，对应 $\det\Lambda=1\quad \Lambda_0^0\leqslant -1$

3. improper orthochronous($\Lambda_-^\uparrow$)，对应 $\det\Lambda=-1\quad \Lambda_0^0\geqslant 1$

4. improper non-orthochronous($\Lambda_+^\downarrow$)，对应 $\det\Lambda=-1\quad \Lambda_0^0\leqslant -1$

推导

采取矩阵记法，设两个惯性系的坐标矢量分别为 X 和 X'，度规矩阵为 g，洛伦兹矩阵为 Λ. 则式 6，式 7，式 9 可以分别写为

$$S^2=X^{\mathrm{T}}gX \tag{13}$$

$$X'=\Lambda X \tag{14}$$

$$g=\Lambda^{\mathrm{T}}g\Lambda \tag{15}$$

将式 15 两边取行列式，有

$$\det g=\det\Lambda^{\mathrm{T}}\cdot\det g\cdot\det\Lambda \tag{16}$$

命题得证.

举例

1. **转动：**纯粹的空间转动下，$x'^0=x^0,x'^i=a^{ij}x^j$，这时洛伦兹矩阵写为

$$\Lambda=\begin{pmatrix}1&0&0&0\\0&a^{11}&a^{12}&a^{13}\\0&a^{21}&a^{22}&a^{23}\\0&a^{31}&a^{32}&a^{33}\end{pmatrix} \tag{17}$$

由于 a^{ij} 构成的子矩阵为正交矩阵，则此时 $\det\Lambda=\pm1$，所以 Λ 可能是 $\Lambda_+^\uparrow$ 或 $\Lambda_-^\uparrow$.

2. **平动（boost）**：对于沿着 x 轴的平动，变换为

$$\begin{aligned} x'^0 &= x^0\cosh\eta - x^1\sinh\eta \\ x'^1 &= -x^0\sinh\eta + x^1\cosh\eta \\ x'^2 &= x^2, \quad x'^3 = x^3 \end{aligned} \tag{18}$$

参数定义为

$$\frac{v}{c} = \tanh\eta \tag{19}$$

$$\left(1-(\frac{v}{c})^2\right)^{-\frac{1}{2}} = \cosh\eta \tag{20}$$

$$\left(1-(\frac{v}{c})^2\right)^{-\frac{1}{2}}\cdot\frac{v}{c} = \sinh\eta \tag{21}$$

那么对应的洛伦兹矩阵为

$$\Lambda = \begin{pmatrix} \cosh\eta & -\sinh\eta & 0 & 0 \\ -\sinh\eta & \cosh\eta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{22}$$

并有

$$\begin{aligned} \det\Lambda &= \cosh^2\eta - \sinh^2\eta = 1 \\ \Lambda_0^0 &= \cosh\eta \geqslant 1 \end{aligned} \tag{23}$$

所以 Λ 属于 $\Lambda_+^\uparrow$

3. **时间反演**此时 $\Lambda = diag(-1,1,1,1)$，明显的，Λ 属于 $\Lambda_-^\downarrow$.

4. **空间反演**此时洛伦兹矩阵为闵可夫斯基度规，即 $\Lambda = diag(1,-1,-1,-1)$. 明显的，此时 Λ 属于 $\Lambda_+^\downarrow$

可以根据群性质证明，所有的变换都可以写成 $\Lambda_+^\uparrow$ 与时间反演变换 Λ_T 以及宇称变换 Λ_P 的乘积．在一些书上，会把洛伦兹群 $\Lambda_+^\uparrow$ 写成 $\mathcal{SO}(3,1)^\uparrow$，表明这是 **Special Orthogonal groups**，**Special** 指的就是行列式为 1 这个性质．所以

$$\mathcal{O}(3,1) = \left\{\mathcal{SO}(3,1)^\uparrow, \Lambda_P\mathcal{SO}(3,1)^\uparrow, \Lambda_T\mathcal{SO}(3,1)^\uparrow, \Lambda_P\Lambda_T\mathcal{SO}(3,1)^\uparrow\right\}$$

所以只需要对 $\mathcal{SO}(3,1)^\uparrow$，即 $\Lambda_+^\uparrow$ 进行研究即可．下述的洛伦兹变换一律指正规正时洛伦兹变换．

洛伦兹变换的群元形式

上述小节已证明，洛伦兹群有 6 个参数，是反对称矩阵 $\omega_{\mu\nu}$ 的六个独立元素，对应着六个独立生成元．把生成元表示为 $J^{\mu\nu}$，因为有反对称指标 (μ,ν) 则有 $J^{\mu\nu}=-J^{\nu\mu}$．根据李群理论，一个广义的洛伦兹群元 Λ 可以写为

$$\Lambda = e^{-\frac{\mathrm{i}}{2}\omega_{\mu\nu}J^{\mu\nu}} \tag{24}$$

四矢量表象下的生成元

在四矢量表象下，生成元由 4Œ4 矩阵 $(J^{\mu\nu})^{\rho}_{\sigma}$ 表示．可以证明，

$$(J^{\mu\nu})^{\rho}_{\sigma} = \mathrm{i}\left(g^{\mu\rho}\delta^{\nu}_{\sigma} - g^{\nu\rho}\delta^{\mu}_{\sigma}\right) \tag{25}$$

洛伦兹群的李代数

可以证明，洛伦兹群的李代数为：

$$\left[J^{\mu\nu}, J^{\rho\sigma}\right] = \mathrm{i}\left(g^{\nu\rho}J^{\mu\sigma} - g^{\mu\rho}J^{\nu\sigma} - g^{\nu\sigma}J^{\mu\rho} + g^{\mu\sigma}J^{\nu\rho}\right) \tag{26}$$

通过重组 $J^{\mu\nu}$，可以得到两类空间矢量

$$J^{i} = \frac{1}{2}\epsilon^{ijk}J^{jk}, \quad K^{i} = J^{i0} \tag{27}$$

在这个角度上，可以写为

$$\begin{aligned}
\left[J^{i}, J^{j}\right] &= \mathrm{i}\epsilon^{ijk}J^{k} \\
\left[J^{i}, K^{j}\right] &= \mathrm{i}\epsilon^{ijk}K^{k} \\
\left[K^{i}, K^{j}\right] &= -\mathrm{i}\epsilon^{ijk}J^{k}
\end{aligned} \tag{28}$$

第一个对易子是 $SU(2)$ 的李代数，这意味着式 27 中 J^{i} 是角动量．第二个对易子则表明了 $\mathbf{K}$ 是平动矢量．

托马斯进动

预备知识 洛伦兹群[1286]

当两个洛伦兹平动复合的时候，其结果通常不是一个洛伦兹平动，而是一个平动加上一个转动.

从物理直觉上来说，考虑三个参考系 K_1，K_2 和 K_3．每个参考系都架设三把标尺，彼此垂直，各标尺分别测量相应参考系里的 x，y 和 z 坐标．令 K_2 以速度 $\mathbf{v}$ 相对 K_1 运动，同时保证在 K_1 和 K_2 看来，双方对应的标尺都相互平行；再令 K_3 以速度 $\mathbf{w}$ 相对 K_2 运动，同时保证在 K_2 和 K_3 看来，双方对应的标尺都相互平行，那么这个时候 K_3 的标尺还和 K_1 的标尺对应平行吗？

答案是否定的．这是因为尺缩效应只发生在沿着参考系相对速度的方向，而垂直于相对速度的方向不会发生尺缩效应．粗略的解释如图 1 所示，K_3 的标尺相对于 K_1 的标尺不再平行，而是转动了一个角度.

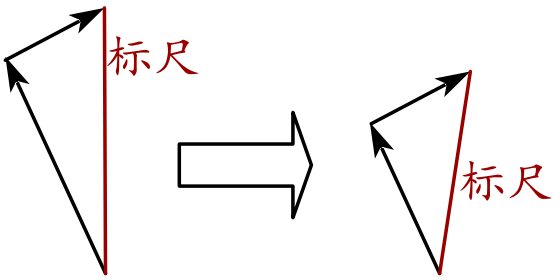


图 1: 托马斯进动的示意图．左图表示 K_3 眼中自己的一根标尺，它是物理存在；垂直的线段表示这个标尺，下面的长箭头表示 K_3 相对 K_1 的运动方向，而上面的短箭头表示垂直于该运动的方向．在 K_1 看来，沿着运动方向发生了尺缩效应，下面的箭头变短，而垂直运动方向的箭头没有变短，结果就是标尺的长度缩短，并且其方向转动了一个角度.

这种纯粹由于相对运动而产生的标尺旋转，被称为**托马斯进动（Thomas Precession）**，由 Llewellyn Thomas 于 1914 年提出．电子的磁矩变化、傅科摆的运动修正，都需要考虑到托马斯进动的影响.

推导

出于方便和实用性的考量，我们研究的是一个系统加速运动中的指向变化．这里的系统可以是一个电子，也可以是一套傅科摆，指向则可以是任意指

定的．以下为了方便表述，我们用一个电子的运动来说明．

设电子在实验室中以速度 $\mathbf{v}=(v,0,0)^{\mathrm{T}}$ 运动，同时具有一个加速度 $\mathbf{a}=(a_x,a_y,0)$．记实验室参考系为 K_1，以速度 $\mathbf{v}$ 运动的参考系为 K_2，而以速度 $\mathbf{v}+\mathbf{a}\mathrm{d}t$ 运动的参考系为 K_3．这里 $\mathrm{d}t$ 是一段极短的时间，而 $\mathbf{v}+\mathbf{a}\mathrm{d}t$ 是电子在 $\mathrm{d}t$ 时刻后的速度．记 $a=|\mathbf{a}|=\sqrt{a_x^2+a_y^2}$.

以上设定中尽管限制了 $\mathbf{v}$ 在 K_1 参考系的 x 轴上、$\mathbf{a}$ 在 $x-y$ 平面上，但是并不失一般性，而且方便计算．这样，K_1 到 K_2 的洛伦兹过渡矩阵就是

$$\mathbf{L}_{1-2}=\begin{pmatrix}\frac{1}{\sqrt{1-v^2}} & \frac{-v}{\sqrt{1-v^2}} & 0 & 0\\ \frac{-v}{\sqrt{1-v^2}} & \frac{1}{\sqrt{1-v^2}} & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix} \tag{1}$$

而 K_2 到 K_3 的过渡矩阵是

$$\mathbf{L}_{2-3}=\begin{pmatrix}\frac{1}{\sqrt{1-(a\mathrm{d}t)^2}} & -\frac{a_x\mathrm{d}t}{\sqrt{1-(a\mathrm{d}t)^2}} & -\frac{a_y\mathrm{d}t}{\sqrt{1-(a\mathrm{d}t)^2}} & 0\\ -\frac{a_x\mathrm{d}t}{\sqrt{1-(a\mathrm{d}t)^2}} & 1+(\frac{1}{\sqrt{1-(a\mathrm{d}t)^2}}-1)\frac{(a_x\mathrm{d}t)^2}{(a\mathrm{d}t)^2} & (\frac{1}{\sqrt{1-(a\mathrm{d}t)^2}}-1)\frac{a_x\mathrm{d}ta_y\mathrm{d}t}{(a\mathrm{d}t)^2} & 0\\ -\frac{a_y\mathrm{d}t}{\sqrt{1-(a\mathrm{d}t)^2}} & (\frac{1}{\sqrt{1-(a\mathrm{d}t)^2}}-1)\frac{a_x\mathrm{d}ta_y\mathrm{d}t}{(a\mathrm{d}t)^2} & 1+(\frac{1}{\sqrt{1-(a\mathrm{d}t)^2}}-1)\frac{(a_y\mathrm{d}t)^2}{(a\mathrm{d}t)^2} & \\ 0 & 0 & 0 & 1\end{pmatrix} \tag{2}$$

那么 K_1 到 K_3 的过渡矩阵就是这两个矩阵的复合（注意它们相乘的方向，并且忽略掉 $\mathrm{d}t$ 大于一阶的项）：

$$\begin{aligned}\mathbf{L}_{1-3}&=\mathbf{L}_{2-3}\mathbf{L}_{1-2}\\ &=\begin{pmatrix}\frac{1}{\sqrt{1-v^2}} & -\frac{v}{\sqrt{1-v^2}} & a_y\mathrm{d}t & 0\\ -\frac{v}{\sqrt{1-v^2}} & \frac{1}{\sqrt{1-v^2}} & 0 & 0\\ -a_y\mathrm{d}t & va_y\mathrm{d}t & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\end{aligned} \tag{3}$$

这个变换就不是一个平动，而是一个平动和一个转动的复合．

如果记 $\mathbf{L}_{3-1}$ 是从 K_3 直接**平动**回到 K_1 的洛伦兹过渡矩阵[18]，那么我们有

[18]就是说，直接从 K_3 变换回 K_1，保持双方的标尺平行．该矩阵的具体形式此处省略了．

$$\mathbf{L}_{1-3}\mathbf{L}_{3-1}=\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & (\frac{1}{\sqrt{1-v^2}}-1)\frac{a_y\mathrm{d}t}{v} & 0\\ 0 & (1-\frac{1}{\sqrt{1-v^2}})\frac{a_y\mathrm{d}t}{v} & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}=\mathbf{R} \tag{4}$$

这是一个 $x-y$ 平面内的转动．也就是说，当粒子有了一个 y 方向的加速度的时候，其参考系指向会变化．

$\mathbf{R}$ 的含义是,在 $\mathrm{d}t$ 内,这个指向在 $x-y$ 平面上**顺时针**转动了 $\arcsin(\frac{1}{\sqrt{1-v^2}}-1)\frac{a_y\mathrm{d}t}{v}$ 的角度，由于该角度很小，因此由 $\lim_{\theta\to 0}\frac{\sin\theta}{\theta}=1$ 可知，这个转动角度也可以写为 $\mathrm{d}\Omega=(\frac{1}{\sqrt{1-v^2}}-1)\frac{a_y\mathrm{d}t}{v}$.

整理以上结果，我们可以得到，当电子以速度 $\mathbf{v}$ 运动且有加速度 $\mathbf{a}$ 时，其坐标系指向按角速度 $\boldsymbol{\omega}$ 偏转，其中 $\boldsymbol{\omega}=\frac{\mathrm{d}\Omega}{\mathrm{d}t}=(\frac{1}{\sqrt{1-v^2}}-1)\frac{\mathbf{a}\times\mathbf{v}}{v^2}$.

当 v 也很小，以至于 $\frac{1}{\sqrt{1-v^2}}\approx 1+\frac{1}{2}v^2$ 时，这个角速度还可以近似为 $\frac{\mathbf{a}\times\mathbf{v}}{2}$.

第二章

相对论动力学

相对论动力学假设

预备知识 相对论加速度变换[1277]，时空的四维表示[1279]

什么是动力学

初学牛顿力学时，你也许注意到了，课本将它分为两大部分：运动学和动力学.

运动学的研究范畴，基本上是纯数学的，只讨论了什么是位移、速度、加速度等概念，以及这些概念之间的数学联系. 在运动学中有一个看起来很自然的假设，在了解相对论之前常常被忽略，那就是在不同参考系中这些概念和它们的联系是如何变化的.

动力学的研究范畴，则加上了“力”的概念，讨论力是如何影响物体的运动状态、力有什么性质等的. 在牛顿力学中，力的性质由牛顿三定律决定，由此可以引申出更深刻的动量守恒和能量守恒等定律.

在狭义相对论中，参考系的变换导致的运动学变换复杂了很多，力的作用也不再像牛顿力学中那样简单.

常见误解的辟谣

一个常见的**误导**是，许多科普读物和视频会使用光子钟等模型来“推导”出爱因斯坦的明星方程，$E=mc^2$，或者用我们约定的单位制，$E=m$. 这些推导的思路是用牛顿第二定律 $F=m\mathrm{d}x/\mathrm{d}t$ 来定义“力”，然后利用洛伦兹变换计算在不同惯性系中，物体加速度的变化. 当然，在使用伽利略变换的牛顿力学中，任何惯性系下物体的加速度都一样，因此质量是不会变化的；但是在狭义相对论中，由于不同惯性系下加速度一般是不同的，加之相对性原理要求力是不随参考系而变化的[1]，因此套入牛顿第二定律以后，必须认为质量是可以变化的. 这种思路所计算出的质点质量变化，就是 $m/\sqrt{1-v^2}$，其中 v 是质点在观察者眼中的运动速度. 这个质量，就是所谓的“动质量”. 以 v^2 为自变

[1]比如说，一个正在施加张力的弹簧，应该看成是在任何参考系下都产生一样的张力，而不受弹簧尺缩效应的影响.

量，对 $m/\sqrt{1-v^2}$ 进行 Maclaulin 展开（即 $v_0^2=0$ 处的泰勒展开），我们可以得到

$$\frac{m}{\sqrt{1-v^2}}=m+\frac{1}{2}mv^2+\frac{3}{4}mv^4+\cdots \tag{1}$$

其中第二项就是牛顿力学中的动能．在 v 远远小于光速 1 的时候，我们可以只看前两项，也就是 $m+mv^2/2$，或者为了更明显地表现量纲，我们把光速添加回去：$mc^2+mv^2/2$．这看上去像是物体的动能加上某种能量．把 mc^2 看成是物体固有的一种能量，那么式 1 就变成了

$$\frac{m}{\sqrt{1-v^2/c^2}}c^2=mc^2+\frac{1}{2}mv^2+\frac{3}{4c^2}mv^4+\cdots=\text{固有能量}+\text{动能} \tag{2}$$

把质点的固有能量 + 动能理解为质点含有的总能量，记为 E，那么就有

$$\frac{m}{\sqrt{1-v^2}}c^2=E \tag{3}$$

这就是著名的质能方程．

但是，这种推导思路是误导的．

从相对论加速度变换[1277] 词条中可以看到，在质点加速度方向垂直和平行于速度方向时，加速度的变化系数是不一样的，这就导致在沿用牛顿第二定律来计算质量的时候，会出现质量随着方向变化的情况．旧版教科书会认为，这意味着不应该把质量简单地看成一个标量，而应该看成一个二阶张量[488]．但现代早已抛弃了“质量随方向变化”的概念，甚至也不再强调“动质量”的概念．

本节将会解释现代的相对论动力学概念．

四动量

设质点静止时的质量是 m．在参考系 K_1 中，质点的四速度是 $\mathbf{U}$，定义 $\mathbf{P}=m\mathbf{U}$ 为质点的**四动量（4-momentum）**．

具体地，如果在 K_1 中，质点的速度是 $(v_x,v_y,v_z)^T$，那么其四速度就是 $\gamma(1,v_x,v_y,v_z)^T$，其中 $\gamma=1/\sqrt{1-v^2}$．这样，质点的四动量就表示为

$$\frac{m}{\sqrt{1-v^2}}\begin{pmatrix}1\\ v_x\\ v_y\\ v_z\end{pmatrix} \tag{4}$$

特别地，四动量的时间分量又被称为**能量**.

动力学假设

在牛顿力学中，动量守恒和能量守恒是三定律的推论. 在实验中，这两个守恒定律在很高的精度上一直成立. 你可以认为推论的成立是对牛顿定律的证明，也可以换个角度，认为守恒定律才是更基本的规律.

狭义相对论中，加速度的变换复杂多了，因此更难以猜想应该怎么拓展牛顿定律. 由于牛顿定律是有效的，我们相信它在限定范围内，也就是宏观低速的情况下是正确的，这也意味着，狭义相对论成立的必要条件之一，是当光速取为无穷大的时候，整个狭义相对论理论就变成了牛顿理论. 现在的问题是，该从什么方向拓展牛顿定律呢？

最简洁的可能性，是**四动量守恒**. 也就是说，在任何惯性系看来，一个孤立系统中的若干质点相互作用前后，系统的总四动量守恒. 这就是相对论的动力学假设，描述了相互作用的规律，作为牛顿定律的拓展.

需要强调的是，这仅仅是一个假设，无法由别的定律推导出来. 也就是说，这个动力学假设是一个公理[91]，和牛顿第二第三定律的地位一样. 如果这个公理所得出的理论和实际相差太大，那么理论物理学家会选择抛弃它，改用别的假设. 幸运的是，迄今为止，四动量守恒和实验很好地相符，因此沿用至今.

尘埃云的能动张量

预备知识 动力学假设[1296]，张量[488]

均匀等速尘埃云的动量通量

均匀等速尘埃云

在任意参考系中，空间中分布若干质点．这些质点的集合，被称为一片**尘埃云（dust）**，各质点被称为**尘埃粒子**．如果在某个参考系中，一片尘埃云的各质点都保持静止，那么我们称这个参考系是尘埃云的**自身系**，称尘埃云为**等速尘埃云**，因为这意味着在其它参考系中，尘埃粒子的速度都会是相同的．如果在某个参考系中，尘埃粒子的质量相同、在空间中均匀分布，那么我们称这片尘埃云是**均匀的**．

任何一片**均匀尘埃云**，都可以看成是许多**均匀等速尘埃云**的叠加，只要把属于各个速度的尘埃粒子分别拿出来构成尘埃云即可．而任何一片尘埃云，也可以看成是局部均匀的．因此，研究均匀等速尘埃云的性质最为容易，也可以方便地拓展到任意尘埃云的性质中．

尘埃云数量通量密度

假设空间中有一片均匀等速尘埃云，在其自身系中各点的粒子数量密度都是 n，其中 n 是一个实数．也就是说，在尘埃云的自身系中，在任何体积 V 中，尘埃粒子的数量都是 nV．

取 K_1 参考系作为观察者，设观察者认为尘埃云的速度是 $\mathbf{v}=(v_x,v_y,v_z)^T$．由于尺缩效应，在 K_1 中，空间各点尘埃粒子的数量密度变为 $n/\sqrt{1-v^2}$．

固定 x 坐标，取此处 $y-z$ 平面上的一个单位面积 A_{yz}．如所示，在单位时间内，通过 A_{yz} 的是体积 $A_{yz}\cdot v_x=v_x$ 内的粒子，数量是

$$\frac{n}{\sqrt{1-v^2}}\cdot v_x=\frac{nv_x}{\sqrt{1-v^2}} \tag{1}$$

类似地，用 i,j,k 代表任意的字母 x,y,z，那么对于固定的 i 坐标，单位时间内通过 $j-k$ 平面上的单位面积的粒子数量是 $nv_i/\sqrt{1-v^2}$．通过一个平面的尘埃粒子数量，称为尘埃云通过这个平面的数量通量；而当所通过平面的面积是单位面积时，数量通量也可以称为数量密度的通量，或者数量通量的密度．

总结下来，如果取所考察平面的面积向量为 $\mathbf{S}$，那么对于上述均匀等速尘埃云，单位时间内通过这个平面的粒子数量是 $\frac{n}{\sqrt{1-v^2}}|\mathbf{v}\cdot\mathbf{S}|$．

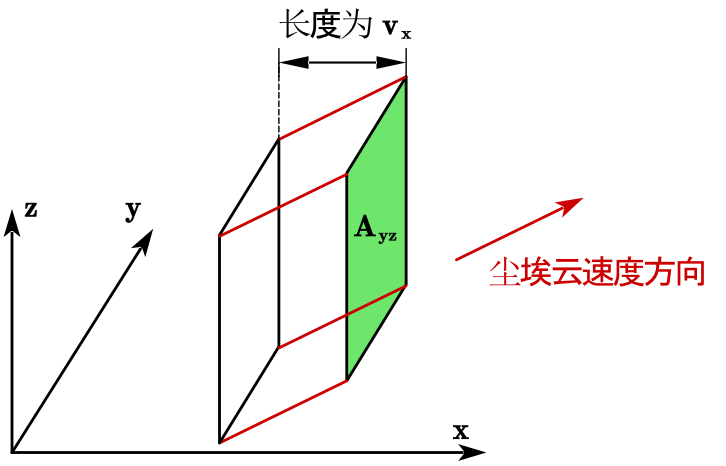


图 1: 粒子数量通量示意图. 绿色面表示计算粒子数量通量的参考平面 A_{yz}；一个粒子在单位时间后通过这个平面，当且仅当这个粒子现在正在图示体积内. 该体积的四条红色边和尘埃云的速度方向平行，而左右两个平面之间的距离是 v_x，故体积为 $A_{yz} \cdot v_x$.

尘埃云动量通量密度

尘埃云设定同上，并假设每个尘埃云粒子的质量是 m，那么它在 K_1 中的动量是 $m/\sqrt{1-v^2}$. 结合数量通量密度可知，在单位时间内通过面积 $\mathbf{S}$ 的尘埃云动量是

$$\frac{n}{\sqrt{1-v^2}}|\mathbf{v}\cdot\mathbf{S}|\cdot\frac{m}{\sqrt{1-v^2}}\mathbf{v}=\frac{nm\mathbf{v}}{1-v^2}|\mathbf{v}\cdot\mathbf{S}|=\frac{nm}{1-v^2}|\mathbf{v}\cdot\mathbf{S}|\cdot\begin{pmatrix}v_x\\v_y\\v_z\end{pmatrix} \tag{2}$$

固定 x 坐标时，单位时间内通过单位面积的尘埃云动量的 y 分量就是

$$\frac{nm}{1-v^2}v_xv_y \tag{3}$$

一般地，固定 i 坐标时，单位时间内通过单位面积的尘埃云动量的 j 分量就是

$$\frac{nm}{1-v^2}v_iv_j \tag{4}$$

四动量通量密度和能量-动量张量

上面所讨论的动量通量密度，在四维闵可夫斯基空间中可以表示为穿过单位时间、单位 A_{yz} 面积的**三维“平面”**[2]的世界线数量乘以各世界线代表的动量.

图 2 是四维闵可夫斯基空间中的动量通量示意图，绿色的垂直线段表示单位时间长度里的单位面积 A_{yz}，相当于用一维线段表示了一个三维的体积；若干红色线段表示各尘埃粒子的世界线. 动量通量密度，就是通过单位时间单位面积的世界线数量乘以各粒子的动量[3].

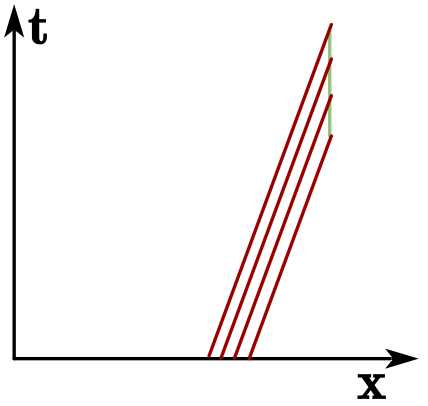


图 2: 四维闵可夫斯基空间中的动量通量

既然都用到四维表示了，我们也可以把结论拓展到四动量，以及固定时间的情况.

如果设各尘埃粒子的四速度是 $\mathbf{U}=1/\sqrt{1-v^2}(1,v_x,v_y,v_z)^T$，可以定义一个矩阵 $\mathbf{M}$，其第 i 行 j 列的元素代表“固定 i 坐标时，单位时间内通过单位面积的尘埃云动量的 j 分量”，那么就有

$$\mathbf{M}=\frac{nm}{1-v^2}\begin{pmatrix}1 & v_x & v_y & v_z\\ v_x & v_x^2 & v_xv_y & v_xv_z\\ v_y & v_xv_y & v_y^2 & v_yv_z\\ v_z & v_xv_z & v_yv_z & v_z^2\end{pmatrix}=nm\cdot\mathbf{U}\mathbf{U}^T \tag{5}$$

[2]多维几何学中，将 n 维空间中的一个 $n-1$ 维子空间称为一个**超平面（hypersurface）**，因此这里的三维平面，指的是四维空间中的三维超平面.

[3]这里有一个视错觉现象，即绿色线段看起来不是垂直的，而是微微倒向左边.

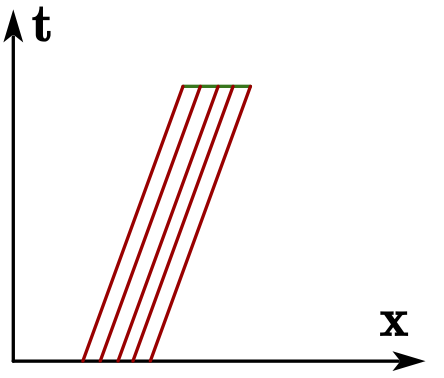


图 3: 固定时间时，穿过单位三维空间超平面的世界线示意图.

这里的 $\mathbf{U}$ 是一个列矩阵，也就是说，$\mathbf{M}$ 是向量 $n\mathbf{U}$ 和 $m\mathbf{U}$ 的张量积．$\mathbf{M}$ 因而是一个二阶张量，被称作**能量-动量张量**.

第三章

广义相对论

双生子佯谬

预备知识 斜坐标系表示洛伦兹变换[1268]

双生子佯谬，又称孪生子佯谬，是一个著名的相对论问题．“佯谬”一词的意思是“看起来像是错误但实际上不是”，其中“谬”指“错误”，而“佯”指“假的”．它曾经被认为是一个悖论，但在今天已经被完美解决了，所以成了一个佯谬．和**时间的变换与钟慢效应**[1260]词条的结尾所指出的一样，我们将从双生子佯谬入手，尝试讨论非惯性系眼中的时空．

问题描述

假设地球在某个惯性系 K_1 中静止，忽略一切引力等作用，把地球考虑成一个质点．为方便理解，也可以说 K_1 是“地球系”．在地球上有一对完全同龄的双胞胎，其中弟弟始终留在地球上，而哥哥则乘坐飞船离开地球．称飞船的参考系 K_2 是“飞船系”，同样看成一个质点．一段时间以哥哥返回并降落在地球上．称飞船的参考系 K_2 是“飞船系”．当飞船降落后，兄弟俩的年龄是否有差异？差异是什么？

双生子佯谬的解答

我们现有的工具只有狭义相对论，而弟弟所在的地球系是一个惯性系，因此我们可以先从地球系开始讨论．假设兄弟俩各戴着一只完美的手表，走时绝无误差，并且双方都把飞船出发的一刻设为时间零点．这样，我们只需要比较哥哥返回地球时两人手表的读数即可．

问题的最简形式，就是哥哥以匀速直线运动离开地球，某一时刻瞬间反向，沿着原道路以相反速度回到地球．这样，我们甚至不需要关心飞船降落的过程，因为当飞船和地球重合的时候，它们的时空坐标完全相同，被认为是同一个事件，无论在哪个参考系看来都是同时发生的，不会出现同时性的相对性．这样，无论飞船有没有和地球保持静止，我们都可以比较哥哥和弟弟各自手上的手表读数．

事实上并不会存在这样瞬间反向的运动，那么这个模型有物理意义吗？答案是有的，我们只需要耍一个小小的花招．

考虑两艘飞船，都在哥哥的航道上，一艘自地球出发做匀速直线运动，另一艘飞船自远方飞向地球，速度和第一艘相反．这样，两艘飞船都是惯性系，都可以用我们已有的狭义相对论方法去讨论它们眼中的时空．这两艘飞船的运动可以看成是，飞船 1 始终离开地球，飞船 2 先靠近地球，然后穿过地球，即两艘飞船一直匀速，只是中途重合了一瞬间；但是也可以看成，飞船 1 离开地球后，到达重合点突然变成飞船 2，飞向地球；飞船 2 也在重合点处突然变成飞船 1，飞离地球．为了把这个双飞船的模型等价转化为问题中单飞船的情况，我们只需要让两艘飞船对表，使得它们重合时手表读数相同，然后只观察“飞船 1 飞到重合点后变成飞船 2 再回到地球”这一过程就可以，将“飞船 2 飞到重合点后变成飞船 1 再飞向无穷远”这一过程视为**不存在**．

坐标网格与等时线划分

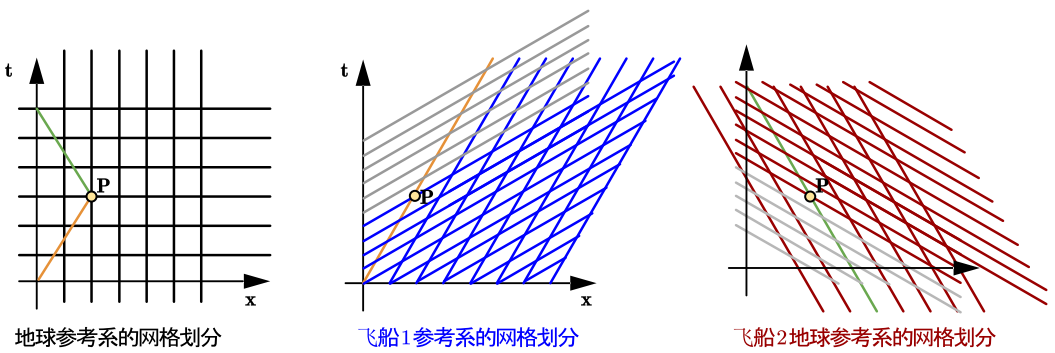


图 1: 在**地球系**中所看到的三个惯性系的网格划分

我们将地球、飞船 1 和飞船 2 的坐标网格分别表示如图 1，三个网格都是在**地球系**中划分的．图中的点 P 表示“飞船 1 和飞船 2 相互转换”这一事件．显然，地球系自身的等时线都和 x 轴平行，等距线都和 t 轴平行，因此它的网格看起来是一个矩形划分，如图中左边的黑色网格所示；飞船 1 的坐标网格如中间的蓝色划分，其中橙色线代表空间坐标始终为 0 的点，即飞船 1 本身的轨迹，而灰色线代表忽略掉的等时线，因为它们出现在 P 点之后，是不存在的；飞船 2 的坐标网格如右边的红色划分，其中绿色线代表空间坐标始终为 0 的点，即飞船 2 本身的轨迹，而灰色线同样代表忽略掉的等时线，因为出现在 P 之前，不存在．

图 1 没有明显画出来的是，各等距线被灰色等时线所覆盖的部分也要忽略掉.

现在，把两个飞船的坐标网格放到一起，擦掉各自被忽略掉的等时线和等距线部分，就得到了飞船系在整个运动过程中的等时线划分，如图 2 所示：

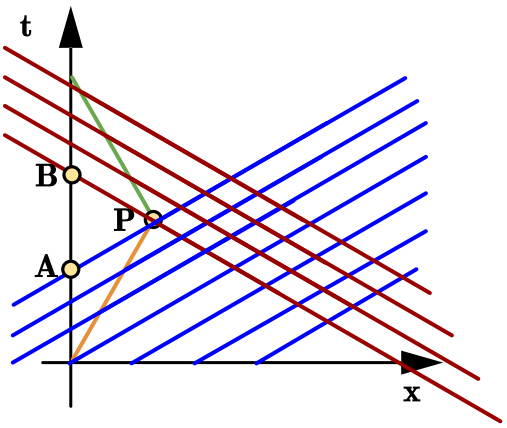


图 2: 飞船系的等时线划分

地球系眼中的飞船时间流逝

在地球上的弟弟看来，尽管飞船的速度在 P 点处突然反向，但是大小是不变的，因此飞船上的时间流速一直都是地球上时间流速的 $\sqrt{1-v^2}$ 倍，其中 v 是指飞船的速度大小. 如果飞船回到地球的一瞬间，弟弟的手表显示时间是 t_0 的话，那么哥哥的手表显示的时间应该是 $t_0\sqrt{1-v^2}$.

飞船系眼中的地球时间流逝

在飞船上的哥哥看来,地球上的时间流速应该是飞船上时间流速的 $\sqrt{1-v^2}$ 倍，其中 v 是指地球的速度大小，和地球眼中飞船的速度大小是相等的. 但是和地球不同的是，在飞船反向的那一刻，即事件 P 处，有两条等时线，分别和 t 轴相交于 A 点和 B 点. 也就是说，在飞船看来，A 和 B 是同时发生的事件，甚至由于它们到飞船的距离也一样，因此在飞船看来是同一个事件.

A 事件和 B 事件分别代表弟弟的手表在地球上显示 t_1 和 t_2，其中 $t_1 < t_2 < t_0$. 也就是说，在飞船上的哥哥看来，当飞船突然变向的时候，地球上的弟弟的手表显示的时间瞬间从 t_1 跳到了 t_2. 整体上，哥哥眼中弟弟的时间流

逝速度还是比飞船慢的，但是在飞船变向时弟弟的时间突然跃变了一下，所以导致最终二人相遇时，弟弟手表的读数还是大于哥哥的.

习题 1

假设二人再次相遇时，弟弟的手表显示时间为 t_0. 计算在地球系中 A 和 B 的时间差 Δt.

习题 1 的计算结果应为 $\Delta t = v^2 t_0$. 如果代入地球系计算的结果，即在哥哥看来整个飞行过程持续时间为 $t_0\sqrt{1-v^2}$，那么飞船系中弟弟的时间流速只有哥哥的 $\sqrt{1-v^2}$ 倍，因此弟弟的时间应该连续流逝了 $t_0\sqrt{1-v^2}\sqrt{1-v^2} = t_0(1-v^2)$；加上弟弟的时间有 $\Delta t = v^2 t_0$ 的跃变，因此在哥哥看来，相遇时弟弟的手表读数是 $\Delta t_0 + t_0(1-v^2) = t_0$，与地球系中计算的结果相同.

解答总结

当哥哥不在地球上时，读表的同时性是相对的，一方眼中的同时读表，在另一方看来就是一先一后的，不是同时. 因此兄弟俩如果要比较彼此的手表读数，就得见一次面，因为见面属于同一事件（时空坐标完全相同），无论在哪个参考系看来都是同时的，见面时的两人手表的读数是绝对的，这时候的比较才有意义. 但是双方要再次见面，哥哥就不可避免地要经历变速过程. 如果是瞬间变速，那么在哥哥看来弟弟的手表读数会在变速的一瞬间跃变，导致最终还是具有更大的读数. 因此当两人见面时，无论在谁看来都是弟弟的时间流逝更多，因此弟弟的年龄应该比哥哥大.

实际情况中我们不可能观测到对方的手表读数跃变，因为我们不可能瞬间变速，而是要经历一个加速过程. 加速过程可以看成无数匀速过程相加[1]，在加速过程中每一个时空点的瞬时自身系中观察地球的时空坐标，那么我们会发现在加速过程中，哥哥眼中弟弟的时间流逝速度会大大加快，导致最终相遇时总的时间流逝还是弟弟更多.

[1]严格来说，可以把加速过程分成 n 段匀速过程，研究其中各事件的时空坐标，再取 $\lim\limits_{n\to\infty}$ 的极限就得到加速过程中各事件的时空坐标.

思考

在图 2 中 P 的右侧有一块锥形区域，其中每个点都经过了一红一蓝两个等时线，这意味着什么？

如果在某个比 P 点距离地球更远的地方是火星，其中火星、地球和飞船始终保持在一条直线上（理想地），火星上有一个人也拿着手表在计时．在整个飞行过程中，从飞船系看，火星上这个人的手表读数是如何变化的？

因果结构

预备知识 连通性[632]，黎曼度量与伪黎曼度量[719]

在研究时空的理论中，因果结构是一个基本问题．根据 David B. Malament 于 1977 年发表的论文 *The class of continuous timelike curves determines the topology of spacetime*[2]，在一个给定的时空中，确定了弱因果性就可以完全确定该时空的拓扑结构．本词条对因果结构相关的部分概念进行了收集和解释，方便阅读相关论文时进行对照．

概念

时空

一个**时空（spacetime）**是指一个无界的、连通的四维伪黎曼流形 (M, g)，其中 g 是一个洛伦兹度规．该度规在任何观察者的参考系中可以表示为对角矩阵 $\mathrm{diag}(1, -1, -1, -1)$．

时间流逝方向分为**未来导向**和**过去导向**的．这个分类是指时空中**类时**切向量的等价划分．如果两个类时切向量的内积为正，那么我们把它们归入同一类，由此把所有类时向量分成两类，称其中一类为未来导向的，另一类为过去导向的．未来和过去的选择是随机的，因为在目前的研究中两个等价类是完全等同的，尚未找到区分它们的细致结构——然而实际的宇宙中似乎有明显的时间导向，因此寻找描述两个等价类根本差异的结构也是数学物理学中一个令人感兴趣的话题．

[2]Malament, Journal of Mathematical Physics 18:7, 1399-1404.

曲线

一条**曲线（curve）**是指一个从实数轴上的任意开区间 I 到拓扑空间的连续映射，在时空理论中这个拓扑空间自然是指时空. 如果曲线是一个光滑映射，那么也被称为一条**光滑曲线（smooth curve）**. 另外，此术语等同于一些拓扑学家口中的**道路（path）**.

一条**类时曲线（timelike curve）**是指一条切向量处处类时的曲线.

一条**类空曲线（spacelike curve）**是指一条切向量处处类空的曲线.

一条**类光曲线（lightlike curve）**或**零曲线（null curve）**是指一条切向量处处类光的曲线.

一条**因果曲线（causal curve）**或**非类空曲线（non-spacelike curve）**是指一条类时或类光的曲线，它以“因果”命名是因为它完全可以成为一个粒子的世界线，因而体现了粒子随着时间运动的因果顺序.

因果结构

对于时空中的两个点 $x, y \in M$，我们定义它们之间的不同关系如下：

- y 是 x 的**类时未来（timelike/chronological curve）**，如果存在一条**未来导向**的**类时**曲线自 x 出发，并经过 y. 这个关系常记为 $x \ll y$.
- y 是 x 的**严格因果未来**，如果存在一条**未来导向**的**类时或类光**曲线 $\gamma : I \to M$，使得存在 $t_1 < t_2$，$\gamma(t_1) = x$ 且 $\gamma(t_2) = y$. 这个关系常记为 $x < y$.
- y 是 x 的**因果未来**，如果 y 是 x 的严格因果未来或者 $x = y$[3]. 这个关系常记为 $x \leq y$ 或者 $x \prec y$.

对于时空中的两个集合 A 与 O，

- 记集合 $I^+(A, O)$ 为“能成为某个 $p \in A$ 的**类时未来**的 $q \in O$”的集合，称之为 A **关于** O **的类时未来（choronological future of** A **relative to** O**）**.
- 记集合 $J^+(A, O)$ 为“能成为某个 $p \in A$ 的**严格因果未来**的 $q \in O$”的集合，称之为 A **关于** O **的因果未来（causal future of** A **relative to** O**）**.

[3]或者说，存在一条**未来导向**的**类时或类光**曲线 $\gamma : I \to M$，使得存在 $t_1 \leq t_2$，$\gamma(t_1) = x$ 且 $\gamma(t_2) = y$.

- 记集合 $E^+(A,O)=J^+(A,O)-I^+(A,O)$，称之为 **horismos future of** A **relative to** O[4].

将以上定义中的“类时未来”换成“类时过去”，即将类时未来中的“未来导向”替换为“过去导向”，可以相应得到 $I^-(A,O)$、$J^-(A,O)$ 和 $E^-(A,O)$.

记:

- $I^+(A,O)\cup I^-(A,O)=I(A,O)$
- $J^+(A,O)\cup J^-(A,O)=J(A,O)$
- $E^+(A,O)\cup E^-(A,O)=E(A,O)$

以上所有表示中，如果 $A=\{q\}$，那么也可以用 q 代替 A，比如 $I(A,O)$ 可以表示为 $I(q,O)$. 如果 O 是整个时空 M，那么可以省略不写，如 $I(A,M)$ 可以表示为 $I(A)$.

如果 $I(A,O)=\varnothing$，那么称 A 在 O 中是**非时序的（achronal）**.

因果结构对时空结构的决定意义

给定两个时空 (M,g) 和 (M',g'). 如果 $f:M\to M'$ 是一个双射，并且对于 M 中的两个点 $p\ll q$ 必有 $f(p)\ll f(q)$，那么称 f 是**因果同构（causal isomorphism）**，也可以说这两个时空是**因果同构（causally isomorphic）**的.

在 Malament 1977 年的论文（词条开始部分所提的那篇）中，证明了如果两个时空间存在一个因果同构，并且两个时空都进行了未来定向和过去定向，那么这两个时空就是同胚的.

[4]尚未找到正式翻译，笔者暂定翻译为“地平未来”（根据 horismos 的构词）或者“测地未来”（根据它的物理意义，沿着测地线）.

第十五部分

量子力学

第一章

量子力学 1

玻尔原子模型

预备知识 圆周运动[885]，库仑力[1123]

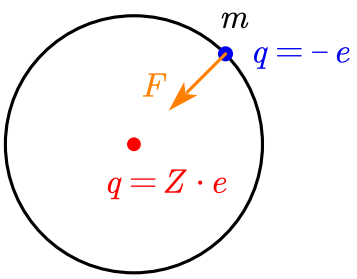


图 1: 玻尔原子模型

玻尔原子模型（Bohr Model）（图 1）是量子力学发展的早期被提出的一种解释**类氢原子**光谱的模型．该模型中，我们假设原子核具有 Z 个正电荷．对于氢原子 H 有 $Z=1$，氦离子 He^{+} 有 $Z=2$，锂离子 Li^{++} 有 $Z=3$ 等等．

由于原子核质量远大于电子，我们先假设它固定不动，唯一的核外电子按照牛顿力学和库伦定律运动，再人为地加上一个条件（**量子化条件**）使电子的轨道角动量只能取一些特定的（**离散**的）值．这样，轨道的半径也只能取离散的值，每个半径 r_n 对应一个机械能（动能加势能）E_n，我们把这些能量叫做**能级**．如果电子从一条能量较高的轨道跃迁到另一条能量较低的轨道，那么一个光子将被产生，带走两个轨道的能量差．反之，如果恰好有一个入射光子的能量是两条轨道机械能之差，那这个光子就可以被低能量轨道的电子吸收，使其跃迁到高能量轨道．

虽然这个模型成功地解释了氢原子各个能级的能量以及氢原子光谱，但它却并不是完全按照量子力学的的方法来计算的．按照（现代的）量子力学，电子需要用波函数描述，波函数由薛定谔方程计算得到，所以不具有经典力学中“轨道”的概念．

根据玻尔模型，氢原子的各个能级的能量为

$$E_n=-\frac{me^4}{32\pi^2\epsilon_0^2\hbar^2}\frac{Z^2}{n^2}\approx -13.6\ \mathrm{eV}\frac{Z^2}{n^2}\qquad (n=1,2,\dots) \tag{1}$$

可见 n 越大，能级越高．我们把 $n=1$ 的状态叫做**基态**，其他状态叫做**激发态**．

各能级的轨道半径如下．特殊地，我们把氢原子基态（$Z=1$，$n=1$）的电子轨道的半径叫做**玻尔半径（Bohr radius）**，记为 a_0．

$$r_n = a_0\frac{n^2}{Z} \qquad (n=1,2,\dots) \tag{2}$$

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{m_e e^2} \approx 5.292\times10^{-11}\ \mathrm{m} \tag{3}$$

能级公式推导

所有原子或离子中最简单的一类叫**类氢原子**，类氢原子只有一个核外电子，以及一个带 Z 个元电荷的原子核．以下的计算假设二者为质点和点电荷，原子核不动，电子绕原子核做圆周运动．运用经典力学和库仑力公式，可求出电子在不同半径下做圆周运动的能量．库仑定律与牛顿定律（圆周运动）分别为

$$F = \frac{1}{4\pi\epsilon_0}\frac{(Ze)e}{r^2} \qquad F = ma = m\frac{v^2}{r} \tag{4}$$

解得电子速度平方为

$$v^2 = \frac{e^2 Z}{4\pi\epsilon_0 m r} \tag{5}$$

动能与势能分别为

$$E_K = \frac{1}{2}mv^2 = \frac{1}{8\pi\epsilon_0}\frac{Ze^2}{r} \qquad E_P = -\frac{1}{4\pi\epsilon_0}\frac{Ze^2}{r} = -2E_k \tag{6}$$

总能量为

$$E = E_K + E_P = -\frac{Ze^2}{8\pi\epsilon_0 r} \tag{7}$$

到此为止，我们还没有用到量子力学．然而这样的模型与真实的类氢原子相比有两个致命的缺陷：第一，根据电动力学，圆周运动的电子会向外辐射电磁波，能量减少，最终坠入原子核; 第二，该模型允许氢原子的能量具有连续值（因为 r 可连续变化），而实验中氢原子只能放出特定能量的光子，说明只能取特定的能量，即存在离散的**能级**，我们把能级由低到高记为 $E\ (n=1,2,3\dots)$．

以上矛盾说明微观世界的粒子不遵守经典力学和电磁学．玻尔为了解释实验，在经典力学和电磁学上加入了一个条件：角动量量子化．

以原子核为原点，电子轨道平面的法向量为 z 轴，由于电子的位矢 $\mathbf{r}$ 与动量 $\mathbf{p}$ 始终垂直，电子的角动量为

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} = mvr\hat{\mathbf{z}} \tag{8}$$

玻尔引入的角动量量子化条件为

$$L = mvr = n\hbar \tag{9}$$

其中 n 可以取任意正整数，$\hbar$ 为**约化普朗克常量**

$$\hbar = \frac{h}{2\pi} \tag{10}$$

该条件也可以等效理解为驻波条件，即允许的圆形轨道长度是德布罗意波长的整数倍.

$$2\pi r = \frac{h}{mv}n \tag{11}$$

注意式 9 与式 11 等效. 把式 5 代入了该条件，解得可能的轨道半径为式 2. 注意轨道与 n^2 成正比，和 Z 成反比.

将 r_n（式 2）代入式 7，得到能级表达式为

$$E_n = -\frac{mZ^2e^4}{32\pi^2\epsilon_0^2\hbar^2}\frac{1}{n^2} \approx -13.6\ \text{eV}\frac{Z^2}{n^2} \tag{12}$$

对氢原子，有 $Z = 1$，最低的能级为 $n = 1$，所以氢原子**基态**的能级 E_0 约为 -13.6 eV. 这是一个著名的常数（若使用原子单位[1315]，这个值恰好是 $-1/2$）.

将 r_n 代入式 5 还可以得到电子速度为

$$v_n = \frac{Ze^2}{4\pi\epsilon_0\hbar}\frac{1}{n} \tag{13}$$

原子单位制

预备知识 无单位的物理公式[876]，玻尔原子模型[1313]

[1]在量子力学的许多理论或数值计算中，选用**原子单位（Hartree atomic unit）**会更方便．需要特别注意的是，原子单位下的物理公式在使用时的习惯和其他单位制有所不同，例如会出现 $E=\omega$（能量等于角频率，见式 8）这种看似不符合量纲分析的公式．要理解这种记号，这里推荐的方法是把使用原子单位的公式都理解成无单位的物理公式[876]．

若声明使用原子单位，物理量的数值后面就不需要标记单位，例如“氢原子基态能量为 $E\approx -0.5$”．有时候为了强调我们使用原子单位，我们会在数值后面加上“a.u.”，如“$E\approx 0.5\,\mathrm{a\,u}$”．这里的“a.u.”可等效为“单位 1”，类似弧度单位“Rad”[2]．

无单位的薛定谔方程

预备知识 薛定谔方程[1331]

国际单位下的一维含时薛定谔方程为

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2}+V\Psi=\mathrm{i}\hbar\frac{\partial\Psi}{\partial t} \tag{1}$$

与“无单位的物理公式[876]”中的方法类似，我们需要给公式中出现的每个单位定义一个常数，分别记为 β_x，β_m，β_t，β_E，β_Ψ，且令 $x=x_a\beta_x$，$m=m_a\beta_m$，$t=t_a\beta_t$，$V=V_a\beta_E$，$\Psi=\Psi_a\beta_\Psi$（注意 x_a, m_a 等都是不带单位的）．代入式 1，各项同除 $\beta_E\beta_\Psi$，得[3]

$$-\left(\frac{\hbar^2}{\beta_m\beta_x^2\beta_E}\right)\frac{1}{2m_a}\frac{\partial^2\Psi_a}{\partial x_a^2}+V_a\Psi_a=\mathrm{i}\left(\frac{\hbar}{\beta_E\beta_t}\right)\frac{\partial\Psi_a}{\partial t_a} \tag{2}$$

为了让公式尽可能简洁，我们令两个括号都为 1，得

$$\beta_E=\frac{\hbar^2}{\beta_m\beta_x^2}\qquad \beta_t=\frac{\beta_m\beta_x^2}{\hbar} \tag{3}$$

于是无单位的薛定谔方程为

$$-\frac{1}{2m_a}\frac{\partial^2\Psi_a}{\partial x_a^2}+V_a\Psi_a=\mathrm{i}\frac{\partial\Psi_a}{\partial t_a} \tag{4}$$

[1]参考 Wikipedia 相关条目．以及 [19] 的附录．

[2]例如在扇形面积公式 $S=\theta R^2/2$ 中的 θ 可以看作具有单位“Rad”，但面积的单位却只是“m^2”而无需记为“$\mathrm{Rad}\cdot\mathrm{m}^2$”．

[3]根据偏微分的定义，常数可以移到偏微分算符外，如 $\partial^2/\partial(\beta_x x_a)^2=(1/\beta_x^2)\,\partial^2/\partial x_a^2$

我们再看波函数的归一化公式

$$1 = \int |\Psi|^2 \,\mathrm{d}x = \beta_\Psi^2 \beta_x \int |\Psi_a|^2 \,\mathrm{d}x_a \tag{5}$$

为了使归一化公式的形式不变，必须令

$$\beta_\Psi = \beta_x^{-1/2} \tag{6}$$

同理，对 N 维波函数有 $\beta_\Psi = \beta_x^{-N/2}$．由式 6 和式 3 可知我们只剩下两个自由的转换常数，例如只要确定 β_x 和 β_m，剩下的 β 也就确定了．

我们来看光子的能量公式

$$E = \hbar\omega \Longrightarrow E_a = \frac{\hbar\beta_\omega}{\beta_E}\omega_a \tag{7}$$

为满足周期的定义（$T_a = 2\pi/\omega_a$），必须令 $\beta_\omega = 1/\beta_t$．将上文得 β_E 和 β_t 代入可得原子单位下光子能量等于角频率

$$E_a = \omega_a \tag{8}$$

我们以后会发现，使用原子单位后所有公式中的普朗克常数 $\hbar$ 都恰好被消去了，所以一些文献中直接说“原子单位中 $\hbar = 1$”．这句话的一种理解方式是把 $\hbar$ 看成一个角动量常量，而原子单位制中角动量的转换常数恰好为 $\beta_L = \hbar$（见式 1），所以转换后该角动量常数等于 1a.u.

原子单位

最常见的情况下，原子单位定义 β_m 等于电子的质量，β_x 等于玻尔半径，再由式 3 和式 6 确定 $\beta_E, \beta_t, \beta_\Psi$，如表 1 所示[4]．注意许多常数都与氢原子的玻尔模型[1313]（原子核不动）的基态（表中简称基态）有关．

表中还定义了一些其他的物理量的转换常数，它们的定义可以使以下无单位公式成立（以后我们在不至于混淆的情况下省略角标 a）

$$\omega = \frac{2\pi}{T} \tag{9}$$

$$x = vt \tag{10}$$

[4]为了区别能量与电场，以下用 E 表示能量，用 $\mathcal{E}$ 表示电场．

表 1: 原子单位转换常数表

物理量	β	描述	数值（国际单位）
质量 m	m_e	电子质量	$9.10938215 \times 10^{-31}$
长度 x	$a_0 = \dfrac{4\pi\epsilon_0\hbar^2}{m_e e^2}$	玻尔半径	$5.2917721067 \times 10^{-11}$
速度 v	$\dfrac{\hbar}{m_e a_0}$	基态电子速度	$2.1876912633 \times 10^{6}$
时间 t	$m_e a_0^2/\hbar$	长度除以速度	$2.418884326 \times 10^{-17}$
角频率 ω	$\dfrac{\hbar}{m_e a_0^2}$	时间的倒数	$6.579683921 \times 10^{15}$
能量 E	$\dfrac{\hbar^2}{m_e a_0^2} = \dfrac{e^2}{4\pi\epsilon_0 a_0}$	基态电子势能大小	$4.3597446499 \times 10^{-18}$
角动量 L	$m_e v_0 a_0 = \hbar$	长度乘以动量	$1.054571800 \times 10^{-34}$
电荷 q	e 或 q_e	电子电荷	$1.6021766208 \times 10^{-19}$
电场强度 $\mathcal{E}$	$\dfrac{e}{4\pi\epsilon_0 a_0^2}$	基态轨道电场强度	$5.1422067070 \times 10^{11}$
磁感应强度 B	$\dfrac{\hbar}{e a_0^2}$		$2.350517567 \times 10^{5}$
电势 V	$\dfrac{e}{4\pi\epsilon_0 a_0}$	基态轨道电势	27.211386019

$$\mathbf{L} = m\mathbf{r} \times \mathbf{v} \quad \text{（角动量）} \tag{11}$$

$$\mathcal{E} = \frac{q}{r^2} \quad \text{（点电荷电场）} \tag{12}$$

$$\rho_{eng} = \frac{1}{8\pi}|\mathbf{E}|^2 \quad \text{（电磁场能量密度）} \tag{13}$$

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B} \quad \text{（洛伦兹力）} \tag{14}$$

$$U = \frac{q}{r} = \mathcal{E}x \quad \text{（点电荷电势）} \tag{15}$$

$$V = qU = -q\mathcal{E}x \quad \text{（匀强电场电势能）} \tag{16}$$

$$\boldsymbol{\mathcal{E}} = -\boldsymbol{\nabla}\varphi - \frac{\partial}{\partial t}\mathbf{A} \quad \text{（标势和矢势）} \tag{17}$$

$$\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A} \quad \text{（标势和矢势）} \tag{18}$$

薛定谔方程为

$$-\frac{1}{2m}\frac{\partial^2\Psi}{\partial x^2}+V\Psi=\mathrm{i}\frac{\partial\Psi}{\partial t} \tag{19}$$

注意当考察对象为电子时，式中 $m=1$，可省略.

例 1　匀强电场中电子的薛定谔方程

令式 19 中 $m=1$，$q=-1$，再将式 16 代入，得

$$-\frac{1}{2}\frac{\partial^2\Psi}{\partial x^2}+\mathcal{E}x\Psi=\mathrm{i}\frac{\partial\Psi}{\partial t} \tag{20}$$

习题 1　氢原子的基态能量

计算玻尔模型中氢原子基态的能量（答案：$-1/2$）.

电磁学相关

把精细结构常数[1321] 记为 α. 国际单位中有 $a_0=\hbar/(\alpha c m_e)$. 原子单位中有 $c=1/\alpha$，$\epsilon_0=1/(4\pi)$，$\mu_0=4\pi\alpha^2$，$\mu_0\epsilon_0=\alpha^2$，电磁场能量密度 $\rho_E=(|\mathbf{E}|^2+|c\mathbf{B}|^2)/(8\pi)$. 坡印廷矢量为 $\mathbf{s}=\mathbf{E}\times\mathbf{B}/(4\pi\alpha^2)$. 对平面电磁波 $E=cB$.

另一种原子单位

当问题涉及一主要角频率 ω 的时候（例如研究原子在单频激光中的变化），可选择 $\beta_E=\hbar\omega$ 做能量单位. 同样令 β_m 等于电子质量，β_q 等于元电荷，由式 3 得

$$\beta_x=\sqrt{\frac{\hbar}{m_e\omega}}\qquad\beta_t=\frac{1}{\omega} \tag{21}$$

为了使式 16 成立，得

$$\beta_{\mathcal{E}}=\frac{\hbar\omega}{e\beta_x} \tag{22}$$

一种常见的情况是平面电磁波中的电场用国际单位表示为 $\mathcal{E}(t)=\mathcal{E}_0\cos(\omega t)$，而原子单位下该式为

$$\mathcal{E}(t)=\mathcal{E}_0\cos t \tag{23}$$

注意右边不含 ω，形式更简洁.

另一个常见的例子是简谐振子[1344]，若使用其振动的固有频率来定义 β_t，其薛定谔方程为

$$-\frac{1}{2}\frac{\partial^2\Psi}{\partial x^2}+\frac{1}{2}x^2\Psi=\mathrm{i}\frac{\partial\Psi}{\partial t} \tag{24}$$

能级为

$$E_n = \frac{1}{2} + n \qquad (n = 0, 1, 2 \dots) \tag{25}$$

归一化的基态波函数为

$$\psi_0(x) = \pi^{-1/4} \mathrm{e}^{-x^2/2} \tag{26}$$

例 2　转换为含单位的公式

现在我们按照“无单位的物理公式[876]”中介绍的方法将式 24 转换为含单位的公式. 即先把所有无单位的物理量替换成有单位的物理量除以对应的 β 常数，得（两边已同乘 β_Ψ）

$$-\beta_x^2 \frac{1}{2} \frac{\partial^2 \Psi}{\partial x^2} + \frac{1}{\beta_x^2} \frac{1}{2} x^2 \Psi = \beta_t \mathrm{i} \frac{\partial \Psi}{\partial t} \tag{27}$$

将式 21 代入，两边乘以 $\omega\hbar$ 得国际单位下的简谐振子薛定谔方程

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \Psi}{\partial x^2} + \frac{1}{2} m\omega^2 x^2 \Psi = \mathrm{i}\hbar \frac{\partial \Psi}{\partial t} \tag{28}$$

类似地，也可以将式 25 变为

$$E = \left(\frac{1}{2} + n\right) \beta_E = \left(\frac{1}{2} + n\right) \omega\hbar \qquad (n = 0, 1, 2 \dots) \tag{29}$$

自然单位制

- 本词条处于草稿阶段.

预备知识　原子单位制[1315]

在高能物理和场论中，我们往往使用一套单位制，使得 $\hbar = 1$ 以及 $c = 1$，下面我们对这种单位制进行说明.

在“原子单位制[1315]”的第一节中，我们已经确定了一系列常数 β，使得薛定谔方程变成更简洁的形式，但还留有两个自由度 β_x 和 β_m. 我们现在使用 $c = 1$ 这个条件，即规定速度的转换常数为 $\beta_v = c$. 如果我们希望满足 $x = vt$ 这个简单的公式，那么必须有 $\beta_x = \beta_v \beta_t$. 将式 3 [1315] 的 β_t 代入得

$$\beta_x \beta_m = \frac{\hbar}{c} \tag{1}$$

所以此时转换常数只剩一个自由度，但注意 Hartree 原子单位制（表 1 [1315]）不满足这个关系．例如令 β_x 为玻尔半径，那么 β_m 将不会等于电子质量，而是电子质量乘以精细结构常数[1321]．反之如果令 β_m 为电子质量，那么 β_x 将等于玻尔半径乘以精细结构常数．

指数衰减

- 本词条处于草稿阶段．

预备知识 一阶线性微分方程[212]

[5]衰变的速率和总量成正比，或者说每个粒子单位时间衰变的概率一样．

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -\lambda N \tag{1}$$

方程的解

$$N(t) = N_0 \mathrm{e}^{-\lambda t} \tag{2}$$

半衰期 T_h 定义为，一半粒子发生衰变所需要的时间．满足

$$N(T_h) = \frac{N_0}{2} \iff \mathrm{e}^{-\lambda T_h} = \frac{1}{2} \tag{3}$$

解得

$$T_h = \frac{\ln 2}{\lambda} \tag{4}$$

精细结构常数

预备知识 国际单位制[1970]

[6]**精细结构常数（fine structure constant）**是量子力学中一个没有量纲的常数，定义为[7]

$$\alpha = \frac{e^2}{4\pi\epsilon_0 \hbar c} \approx 0.0072973525693(11) \approx \frac{1}{137.036} \tag{1}$$

[5]参考 Wikipedia 相关页面．

[6]参考 Wikipedia 相关页面．

[7]采用 2018 CODATA 推荐值．

其中 e 是元电荷，π 是圆周率，ϵ_0 是真空电介质常数，$\hbar$ 是约化普朗克常数，c 是真空中的光速．根据 2019 年采用的国际单位制最新标准，这些常数除了 ϵ_0 都是精确的，所以对精细常数的精确测量同时也相当于对 ϵ_0 或 μ_0 的精确测量．

量子力学与矩阵

预备知识 量子力学简介[60]，重积分[283]，厄米矩阵本征值问题[426]

在量子力学简介[60] 中，我们知道粒子的状态由波函数描述．许多量子力学教材会从波函数和薛定谔方程讲起（如无限深势阱[1335]，简谐振子[1340]），而另一些教材会从矩阵的形式讲起．前者由于涉及到求解（偏）微分方程，数学要求会相对较高，而后者只需涉及一些有限维的线性代数．二者乍看起来不同，但实际上几乎是等效的[8]．为了让读者先对量子力学的理论结构由一个总体的了解，避免迷失在数学计算中，我们先介绍矩阵的形式．

以下将例 2 [60] 展开讨论，使用线性代数的语言，继续完善量子力学的基本假设并应用到例子中．该例中的角动量本征态 $|x+\rangle$，$|x-\rangle$ 和 $|z+\rangle$，$|z-\rangle$ 之间的关系是线性的．这很容易让我们联想到矢量空间[444] 的基底．根据矢量空间的定义，这四个本征态符合作为矢量的要求：可以给它们乘以常数，可以把它们相加（见式 2 [60]），等等（参考例 1 [444]）．正是由于这个原因，我们把量子力学中的态（即波函数）称为**状态矢量（state vector）**，简称**态矢**．

当明确了它们是矢量后，我们就可以对这些态（波函数）定义内积[471]．以（某个时刻的）一维（波）函数 $f(x)$，$g(x)$ 为例（可以用狄拉克符号记为 $|f\rangle$，$|g\rangle$），我们将其内积定义为（注意这里的内积是有顺序的，星号代表对函数值取复共轭）

$$\langle f|g\rangle = \int_{-\infty}^{+\infty} f(x)^* g(x)\,\mathrm{d}x \tag{1}$$

如果态矢 $|f\rangle$ 和 $|g\rangle$ 是二维或三维的波函数，我们只需要用重积分即可（积分的范围是全空间，即三个方向的上下限都为无穷大）[9]

$$\langle f|g\rangle = \int f(\mathbf{r})^* g(\mathbf{r})\,\mathrm{d}^3 r = \iiint f(x,y,z)^* g(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \tag{2}$$

[8] 最大的区别和难点是前者是无穷维而后者是有限维

[9] 再次强调，例 2 [60] 讨论的态矢是抽象的，并不能用波函数表示，但为了方便理解，我们姑且假设它们可以．

容易证明该定义满足 $\langle f|g\rangle = \langle g|f\rangle^*$，即交换内积的顺序，相当于把结果取复共轭.

角动量算符

在“量子力学简介[60]”中，由于数学工具上的不足，我们并没有提及某个物理量的本征态是怎么得到的，这里来简单介绍.

量子力学中，每个物理量都对应一个**算符（operator）**，算符可以想象为对（波）函数的一种操作，算符作用在（波）函数上可以得到一个新的函数. 例如某时刻函数为 $\sin x$，求导算符 $\mathrm{d}/\mathrm{d}x$ 作用在 $\sin x$ 上就得到一个新的函数 $\cos x$. 又例如坐标 x 也可以作为一个算符，我们定义将其作用在任意函数 $\Psi(x,t)$ 上，就是将其相乘，即 $x\Psi(x,t)$. 又例如任意函数 $f(x)$ 也可以是一个算符，我们定义将其作用在 $\Psi(x,t)$ 上得 $f(x)\Psi(x,t)$.

不难发现这些算符都是**线性**的，如果它们在讨论的有限维[10]空间[11]中是封闭的，我们就可以用矩阵来表示它们[12]（见“矩阵与线性变换[456]”）.

我们这里考察的空间的确是有限维（二维）的，习惯上我们用 $|z+\rangle$ 和 $|z-\rangle$ 作为基底（也可以用 $|x+\rangle$ 和 $|x-\rangle$ 或 $|y+\rangle$ 和 $|y-\rangle$，进行基底变换即可）. 在这个二维空间中，我们直接给出（以后再推导）x,y,z 三个方向的角动量算符的矩阵（如果去掉 $\hbar/2$，这三个矩阵就是著名的**泡利矩阵**）为

$$\mathbf{L}_x = \frac{\hbar}{2}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \qquad \mathbf{L}_y = \frac{\hbar}{2}\begin{pmatrix}0 & -\mathrm{i}\\ \mathrm{i} & 0\end{pmatrix} \qquad \mathbf{L}_z = \frac{\hbar}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \tag{3}$$

注意它们都是厄米矩阵[422]，即满足 $\mathbf{L}_i^\dagger = \mathbf{L}_i$.

那么我们为什么需要算符或者矩阵呢？简单来说，**一个物理量的本征矢就是就是对应算符（矩阵）的本征矢，对应的测量值就是本征值**. 这是量子力学的又一基本假设.

以 $\mathbf{L}_z$ 为例，其本征方程为

$$\frac{\hbar}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}c_1\\ c_2\end{pmatrix} = \lambda\begin{pmatrix}c_1\\ c_2\end{pmatrix} \tag{4}$$

[10]无限维空间在数学上会产生许多麻烦.

[11]这里要提醒一下态矢所在的矢量空间的维度和物理空间的维度不是一回事，例如将一维（物理上的空间维度，即粒子做直线运动）波函数 $\sin(x)$ 和 $\cos(x)$ 作为基底，可以组成一个二维态矢空间.

[12]以后我们还会详细讲算符和矩阵之间的联系，如果你暂时不理解，可以先接受这个结论.

由于这个矩阵已经是对角矩阵，两个本征矢和本征值分别为

$$\begin{aligned}\begin{pmatrix}c_1\\c_2\end{pmatrix}&=\begin{pmatrix}1\\0\end{pmatrix}&\lambda&=+\frac{\hbar}{2}\\\begin{pmatrix}c_1\\c_2\end{pmatrix}&=\begin{pmatrix}0\\1\end{pmatrix}&\lambda&=-\frac{\hbar}{2}\end{aligned}\tag{5}$$

如我们所料，两个态矢就是我们使用的两个基底 $|z+\rangle$ 和 $|z-\rangle$（因为它们就是这么定义的），本征值也同样吻合.

我们再来解 x 方向角动量的本征方程

$$\frac{\hbar}{2}\begin{pmatrix}0&1\\1&0\end{pmatrix}\begin{pmatrix}c_1\\c_2\end{pmatrix}=\lambda\begin{pmatrix}c_1\\c_2\end{pmatrix}\tag{6}$$

解得两个态矢和本征值分别为（量子力学中我们需要把所有的本征矢归一化，意义见下文）

$$\begin{aligned}\begin{pmatrix}c_1\\c_2\end{pmatrix}&=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}&\lambda&=+\frac{\hbar}{2}\\\begin{pmatrix}c_1\\c_2\end{pmatrix}&=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}&\lambda&=-\frac{\hbar}{2}\end{aligned}\tag{7}$$

根据定义，这两个矢量就是 $|x+\rangle$ 和 $|x-\rangle$，考虑到我们现在用的基底是 $|z+\rangle$ 和 $|z-\rangle$，马上得到式 2 [60]

$$\begin{cases}|x+\rangle=\dfrac{1}{\sqrt{2}}|z+\rangle+\dfrac{1}{\sqrt{2}}|z-\rangle\\|x-\rangle=\dfrac{1}{\sqrt{2}}|z+\rangle-\dfrac{1}{\sqrt{2}}|z-\rangle\end{cases}\tag{8}$$

最后，y 方向的本征方程和两个解为

$$\frac{\hbar}{2}\begin{pmatrix}0&-\mathrm{i}\\\mathrm{i}&0\end{pmatrix}\begin{pmatrix}c_1\\c_2\end{pmatrix}=\lambda\begin{pmatrix}c_1\\c_2\end{pmatrix}\tag{9}$$

$$\begin{aligned}\begin{pmatrix}c_1\\c_2\end{pmatrix}&=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\\mathrm{i}\end{pmatrix}&\lambda&=+\frac{\hbar}{2}\\\begin{pmatrix}c_1\\c_2\end{pmatrix}&=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-\mathrm{i}\end{pmatrix}&\lambda&=-\frac{\hbar}{2}\end{aligned}\tag{10}$$

即

$$\begin{cases} |y+\rangle = \dfrac{1}{\sqrt{2}}|z+\rangle + \dfrac{\mathrm{i}}{\sqrt{2}}|z-\rangle \\ |y-\rangle = \dfrac{1}{\sqrt{2}}|z+\rangle - \dfrac{\mathrm{i}}{\sqrt{2}}|z-\rangle \end{cases} \tag{11}$$

以上的过程中有几点需要注意．由于测量到的物理量（本征值）必须是实数，算符对应的矩阵必须是厄米矩阵．N 维厄米矩阵会有 N 个正交归一的本征矢．不难验证，$\langle x+|x-\rangle$，$\langle y+|y-\rangle$ 和 $\langle z+|z-\rangle$ 都为 0．例如

$$\langle y+|y-\rangle = \begin{pmatrix} 1 \\ \mathrm{i} \end{pmatrix}^{\dagger} \begin{pmatrix} 1 \\ -\mathrm{i} \end{pmatrix} = \begin{pmatrix} 1, -\mathrm{i} \end{pmatrix} \begin{pmatrix} 1 \\ -\mathrm{i} \end{pmatrix} = 0 \tag{12}$$

算符和本征问题

预备知识 量子力学[60]，平面波[1008]，厄米矩阵的本征值问题[426]

“量子力学[60]”中我们已经简单介绍了量子力学的基本假设．这里我们来进行更详细的说明，注意我们仍然只讨论做一维直线运动的单个微观粒子．首先来做一个总结

- 粒子的状态由波函数描述．
- 波函数按照薛定谔方程随时间变化．
- 某时刻若对粒子测量一个物理量，则先找到该物理量的本征函数 $\Psi_1, \Psi_2, \ldots$ 以及对应的本征值，把粒子在该时刻的波函数 $\Psi(x,t)$ 表示成这些本征函数的线性叠加 $\Psi = C_1\Psi_1 + C_2\Psi_2 \ldots$，测得第 i 个本征值的概率就是系数 C_i 的模长平方．本征值有可能是连续的或是离散的．
- 测量完后，粒子的波函数坍缩为第 i 个本征函数．

算符和本征问题

之前提到，位置的本征函数是一些无穷窄的函数，叫做 δ 函数，对应的位置本征值则是这些 δ 函数所在的位置．动量的本征函数是一些平面波，对应的

动量本征值就是平面波的空间频率乘以一个常数．然而我们并没有说明某个物理量的本征波函数是怎么得到的，以下将进一步介绍．

在量子力学中，每个物理量都可以对应一个**算符**，算符可以想象为对波函数的一种操作，算符作用在波函数上可以得到一个新的波函数．例如某时刻波函数为 $\sin x$，求导算符 $\mathrm{d}/\mathrm{d}x$ 作用在 $\sin x$ 上就得到一个新的波函数 $\cos x$．又例如坐标 x 也可以作为一个算符，我们定义将其作用在任意波函数 $\Psi(x,t)$ 上，就是将其相乘，即 $x\Psi(x,t)$．又例如任意函数 $f(x)$ 也可以是一个算符，我们定义将其作用在 $\Psi(x,t)$ 上得 $f(x)\Psi(x,t)$．

在书写习惯上，我们将某物理量 Q 的算符用 $\hat{Q}$ 表示，如位置的算符用 $\hat{x}$ 表示，动量的算符用 $\hat{p}$ 表示．当我们熟练以后，为了书写简洁往往将“ˆ”符号省略．

要得到某个物理量的本征函数，我们需要解本征方程

$$\hat{Q}\psi(x)=\lambda\psi(x) \tag{1}$$

其中 λ 是本征值，$\psi(x)$ 是本征函数，二者都是未知的．如果本征值是离散的（如束缚态的能量），我们就可以用整数角标来 i (来区分不同的本征值和本征函数，将它们分别记为 λ_i 和 $\psi_i(x)$，如果本征值是连续的（如位置，动量），我们就将角标 i 替换为一个实数，比方说记为 α．用 $\lambda(\alpha)$ 区分不同本征值，将对应的波函数记为 $\psi_\alpha(x)$．注意本征函数不含时间变量 t．另外，用于由于物理量的算符，其本征值必定是实数．我们把这类算符叫做**厄米算符（Hermitian operator）**[13]，以后会具体介绍．

我们姑且认为，量子力学的基本假设规定[14]**位置的算符** $\hat{x}$ 是 x，**动量的算符** $\hat{p}$ 是微分算符 $-\mathrm{i}\hbar\,\partial/\partial x$．其中 i 是虚数单位，$\hbar$ 是一个常数，叫做**约化普朗克常数**，即普朗克常数 h 除以 2π．

可以定性地验证位于坐标 x_0 处的 δ 函数是 $\hat{x}$ 的本征矢，且本征值为 x_0．我们把位于原点处的 δ 函数记为 $\delta(x)$，那么 x_0 处的 δ 函数就是 $\delta(x-x_0)$．将本征函数和本征值代入本征方程，得

$$x\delta(x-x_0)=x_0\delta(x-x_0) \tag{2}$$

[13] 数学上也叫**自伴算符（self-adjoint operator）**

[14] 严格来说这并不是基本假设的一部分，但现阶段这么认为并没有大碍．说实话，本书并不打算深究这一点．另外，算符在不同的波函数**表象**下具有不同的形式，这里使用的是**位置表象**．表象的概念以后会学习，现在先不用担心．

我们可以从函数图像上对该式做一个定性证明：由于 $\delta(x-x_0)$ 只有在 x_0 处一个无穷窄的区间不为零，所以将 δ 函数乘以 x，就相当于在这个不为零的无穷窄区间乘以 x_0.

动量的本征方程为

$$\hat{p}\psi(x) = -\mathrm{i}\hbar\frac{\partial}{\partial x}\psi(x) = p\psi(x) \tag{3}$$

代入即可证明本征函数为[15] $\psi(x) = \exp(\mathrm{i}kx)$，对应的动量本征值为 $p = \hbar k$. $\exp(\mathrm{i}kx)$ 是平面波的复数形式，量子力学中的平面波指的是这个而不是 $\sin(kx)$ 或 $\cos(kx)$.

利用波长和波数的关系 $\lambda = 2\pi/k$，以及 $\hbar = h/2\pi$，我们就可以得到著名的**德布罗意公式**

$$\lambda = \frac{h}{p} \tag{4}$$

所以德布罗意公式描述的是动量本征值和本征函数（平面波）的波长之间的关系，即动量和波长成反比. 所以平面波的波长 λ 也叫**德布罗意波长**.

量子力学的基本假设规定，其他所有算符都可以通过位置和动量算符拼凑而成，其形式与经典力学中对应物理量的形式相同. 例如，经典力学中的动能为 $p^2/(2m)$，那么量子力学中的动能算符（习惯上用大写字母 T 表示）就是

$$\hat{T} = \frac{\hat{p}^2}{2m} \tag{5}$$

要理解算符运算并不难，这里的 $\hat{p}^2$ 也可以记为 $\hat{p}\hat{p}$，即两个动量算符**相乘**. 两个算符相乘的定义是，将右边的算符先作用在波函数上，再将左边的算符作用在波函数上. 所以 $\hat{p}^2$ 作用在波函数 $\psi(x)$ 上，就相当与先关于 x 求偏导，乘以常数 $-\mathrm{i}\hbar$，再关于 x 求偏导，再乘以常数 $-\mathrm{i}\hbar$.

$$\hat{p}^2\psi = -\mathrm{i}\hbar\frac{\partial}{\partial x}\left(-\mathrm{i}\hbar\frac{\partial}{\partial x}\psi\right) \tag{6}$$

由于常数可以提到求导算符的外面，这就相当于关于 x 求二阶偏导，然后在乘以 $(-\mathrm{i}\hbar)^2 = -\hbar^2$.

$$\hat{p}^2\psi(x) = -\hbar^2\frac{\partial^2}{\partial x^2}\psi(x) \tag{7}$$

[15] 我们也可以通过解微分方程得到

所以

$$\hat{p}^2 = -\hbar^2 \frac{\partial^2}{\partial x^2} \tag{8}$$

$$\hat{T} = \frac{\hat{p}^2}{2m} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} \tag{9}$$

有了动能的算符，我们就可以列出动能的本征方程并解出其本征函数和本征值 E_k（注意角标 k 代表 kinetic，不是波数）

$$\hat{T}\psi(x) = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x) = E_k\psi(x) \tag{10}$$

巧的是，解出动能的本征函数与动量的本征函数相同，也是 $\exp(-\mathrm{i}kx)$，但对应的本征值不同，是 $E_k = \hbar^2k^2/(2m)$. 这事实上并不是巧合，因为 $\hat{T}$ 正比于 $\hat{p}^2$，当第一个 $\hat{p}$ 作用在动量的本征函数上，得到的是常数乘以该本征函数，再经第二个 $\hat{p}$ 作用，同样相当于乘以一个常数

$$\begin{aligned} -\mathrm{i}\hbar\frac{\partial}{\partial x}\left(-\mathrm{i}\hbar\frac{\partial}{\partial x}\mathrm{e}^{-\mathrm{i}kx}\right) &= -\mathrm{i}\hbar\frac{\partial}{\partial x}\left(\hbar k\mathrm{e}^{-\mathrm{i}kx}\right) = \hbar k\left(-\mathrm{i}\hbar\frac{\partial}{\partial x}\mathrm{e}^{-\mathrm{i}kx}\right) \\ &= \hbar^2k^2\mathrm{e}^{-\mathrm{i}kx} \end{aligned} \tag{11}$$

两边除以 $2m$，就验证了动能的本征方程，并得到本征值. 需要注意的是，同一个动能本征值 E_k 对应两个互为相反数的波数 k，即存在两个线性无关的本征态（分别是向左和向右的平面波），且这两个本征态的任意线性组合都是同一个能量本征值的本征函数. 我们把这种一个本征值对应多个本征函数的情况叫做**简并**，如果最多由 N 个本征函数，就有 N **重简并**. 所以一维情况下，能量具有二重简并.

定态薛定谔方程

- 本词条存在未完成的内容.
- 本词条需要更多讲解，便于帮助理解.

预备知识 量子力学基本假设[1435]，矢量算符[300]

定态薛定谔方程

定态薛定谔方程就是哈密顿算符的本征方程，本征值就是能量 E.

$$H\psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{1}$$

单个粒子问题中，哈密顿算符对应粒子的总能量，总能量算符可以表示为动能算符和势能算符之和

$$H = T + V \tag{2}$$

一维定态薛定谔方程

一维运动的单个质点，波函数是坐标 x 的函数 $\Psi(x)$

$$T = -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} \qquad V = V(x) \tag{3}$$

所以定态薛定谔方程为

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\Psi}{\mathrm{d}x^2} + V(x)\Psi = E\Psi \tag{4}$$

束缚态

这是一个二阶线性常微分方程. 数学上来看无论 E 是多少，必有两个线性无关的解，它们的线性组合也是解（二维解空间）. 然而物理上可能存在的波函数必须要可归一化：即随着 $|x| \to \infty$，$|\psi|^2$ 在极限 $|x| \to \infty$ 的过程中下降得比 $1/x$ 要快. 可以证明只有对于某些离散的 E 我们才能解出这些波函数. 我们把每个 E_i 叫做**能级（energy level）**，对应的波函数叫做**束缚态（bound state）**.

所有的能级都处于能量区间

$$\min V(x) < E_i < \min V(\pm\infty) \tag{5}$$

其中 $\min V(x)$ 是函数 $V(x)$ 的最小值（可以是负无穷），$\min V(\pm\infty)$ 表示 $V(x)$ 的正无穷极限和负无穷极限中较小的一个. 因为当 E 太小时，波函数必定会在无穷远处爆炸（无穷大），而当 E 太大时，波函数虽然不会爆炸但也不会趋于零. 式 5，要求势能函数 $V(x)$ 中存在某种形状的凹陷，又称为**势阱（potential well）**.

例子：无限深势阱[1335]，有限深势阱[1362]，简谐振子（升降算符）[1340].

散射态

当 $E > \min V(\pm\infty)$ 时，虽然波函数不满足归一化，但它们仍然有重要的应用．我们把它们叫做**散射态（scattering states）**．计算散射态时，通常我们要求两个极限 $V(\pm\infty)$ 都存在．这样在满足无穷远处 $E > V$ 的方向，波函数在无穷远处将会是简谐波．最直接的例子是平面波函数．

虽然散射态本身不能归一化，但是它们的线性组合却可以．例如 $V(x) \equiv 0$ 势能中的高斯波包[1012] 就可以看作由平面波线性叠加而来，即反傅里叶变换．

某种意义上，给定一个势能函数，所有的束缚态和散射态可以构成一组正交归一的函数基底，用于展开所有可归一化的波函数．

简并

简并是指一个能量本征值 E 对应多个线性无关的归一化波函数的情况．可以证明一维束缚态不存在简并，散射态存在双重简并．

多维定态薛定谔方程

二维或三维的情况下，波函数是位置矢量[878] 的函数 $\Psi(\mathbf{r})$

$$T = -\frac{\hbar^2}{2m}\nabla^2 \qquad V = V(\mathbf{r}) \tag{6}$$

定态薛定谔方程为

$$-\frac{\hbar^2}{2m}\nabla^2\Psi + V(\mathbf{r})\Psi = E\Psi \tag{7}$$

例子：三维简谐振子（球坐标）[1428]．

未完成：为什么束缚态必须是实函数（或者乘以一个相位因子）？为什么实函数的平均动量必须等于 0（实函数的傅里叶变换对对称函数）？物理上意味着什么？如果势能是偶函数，那么波函数为什么一定是基函数或者偶函数？这可以通过哈密顿算符和宇称算符的对易性来解释．

波函数的对称性

预备知识 宇称算符[842]

若哈密顿算符和宇称算符 Π 对易，则它们具有一组共同的本征波函数，其中每个都具有奇宇称或者偶宇称．例如对于一维定态薛定谔方程，若势能函数是偶函数，即 $V(x)=V(-x)$，则波函数必定是奇函数或者偶函数，即

$$\psi(-x)=\pm\psi(x) \tag{8}$$

薛定谔方程

- 本词条处于草稿阶段．
- 本词条需要更多讲解，便于帮助理解．

预备知识 定态薛定谔方程[1328]，分离变量法[742]

$$H\left|\Psi\right\rangle=\mathrm{i}\frac{\partial}{\partial t}\left|\Psi\right\rangle \tag{1}$$

波函数是位置矢量[878] 的函数 $\Psi(\mathbf{r},t)$

$$T=-\frac{\hbar^2}{2m}\nabla^2 \qquad V=V(\mathbf{r}) \tag{2}$$

定态薛定谔方程为

$$-\frac{\hbar^2}{2m}\nabla^2\Psi+V(\mathbf{r})\Psi=\mathrm{i}\frac{\partial}{\partial t}\Psi \tag{3}$$

能量守恒系统

当哈密顿算符 H 不随时间变化时，我们说这个系统能量守恒[1377]．这时我们可以用分离变量法，令

$$\Psi(\mathbf{r},t)=\psi(\mathbf{r})T(t) \tag{4}$$

$$H\psi=E\psi \tag{5}$$

以及

$$\mathrm{i}\hbar\frac{\partial}{\partial t}T=ET \tag{6}$$

其中式 5 就是定态薛定谔方程[1328]，即哈密顿算符的本征方程．式 6 有简单的解

$$T(t) = \mathrm{e}^{-\mathrm{i}Et/\hbar} \tag{7}$$

根据 H 的不同，本征值 E 可以取离散或连续的值．先看只取离散值的简单情况（如无限深势阱[1335]），令能级为 E_n （$n = 1, 2, \dots$），那么含时薛定谔方程的通解为

$$\Psi(\mathbf{r}, t) = \sum_n C_n \psi_n(\mathbf{r}) \mathrm{e}^{-\mathrm{i}E_n t/\hbar} \tag{8}$$

其中 C_n 为待定系数，由初始条件决定．一个简单的例子见“无限深势阱中的高斯波包[1337]”.

如果 E 只在某个区间内取连续值，我们同样可以使用分离变量法，只是求和变为积分，系数变为能量的函数

$$\Psi(\mathbf{r}, t) = \int C(E) \psi_E(\mathbf{r}) \mathrm{e}^{-\mathrm{i}Et/\hbar} \,\mathrm{d}E \tag{9}$$

若能量是简并的，即定态薛定谔方程中一个能量 E 由多个线性无关的解 $\psi_{E,i}(x)$（$i = 1, 2 \dots$），那么上式变为

$$\Psi(\mathbf{r}, t) = \int \sum_i C_i(E) \psi_{E,i}(\mathbf{r}) \mathrm{e}^{-\mathrm{i}Et/\hbar} \,\mathrm{d}E \tag{10}$$

一个经典的例子见“一维自由粒子（量子）[1345]”.

不确定原理

预备知识 平均值[1376]，柯西不等式[800]

位置—动量不确定原理

单个粒子一维运动的波函数 $\psi(x)$ 的位置和动量的标准差为 σ_x 和 σ_p

$$\sigma_x \sigma_p \geqslant \frac{\hbar}{2} \tag{1}$$

例 1　无限深势阱的束缚态

（未完成）证明束缚态满足式 1.

例 2　高斯波包

（未完成）证明高斯波包可以使式 1 取等号.

不确定原理的拓展

任意两个物理量 A 和 B 都满足

$$\sigma_a\sigma_b \leqslant \frac{1}{2\mathrm{i}}\langle v|[A,B]|v\rangle \tag{2}$$

式 1 可以看成是该式的特例：令 $A=x, B=p$，根据正则对易关系

$$[A,B]=[x,p]=\mathrm{i}\hbar \tag{3}$$

代入式 2 得式 1.

证明

令 $f=(A-a)v$，$g=(B-b)v$，使用柯西不等式[800] 有

$$\sigma_a^2\sigma_b^2=\langle f|f\rangle\langle g|g\rangle \leqslant |\langle f|g\rangle|^2=(\mathrm{Re}\langle f|g\rangle)^2+(|\mathrm{Im}\langle f|g\rangle|)^2 \tag{4}$$

其中

$$\begin{aligned}\mathrm{Im}\langle f|g\rangle &= \frac{1}{2\mathrm{i}}(\langle f|g\rangle-\langle g|f\rangle)=\frac{1}{2\mathrm{i}}\langle v|[(A-a),(B-b)]|v\rangle \\ &= \frac{1}{2\mathrm{i}}\langle v|[A,B]|v\rangle\end{aligned} \tag{5}$$

忽略式 4 中的 Re 项（大于等于 0），得到式 2．容易证明，$\langle v|[A,B]|v\rangle$ 必然是一个纯虚数，所以式 2 右边必为实数.

多维空间中的量子力学

预备知识　张量积空间[513]，梯度[307]

单个粒子在多维空间中的波函数

本词条中的“多维”指的是二维和三维．与牛顿力学一样，在学习完粒子的直线运动（一维）后，我们希望能了解粒子在平面上的运动（二维），或者

空间中的运动（三维）．在多于一维的情况下，波函数变为位置矢量 $\mathbf{r}$ 以及时间 t 的函数

$$\Psi(\mathbf{r},t) \tag{1}$$

例如在二维直角坐标系中，$\Psi(\mathbf{r},t)=\Psi(x,y,t)$，又例如在三维的球坐标系中，$\Psi(\mathbf{r},t)=\Psi(r,\theta,\phi,t)$

矢量算符

在多维空间中，位置和动量分别从标量拓展为矢量，所以对应地，位置算符和动量算符也分别拓展为**矢量算符**（我们暂时把这两个算符的定义看作是量子力学的基本假设）

$$\hat{\mathbf{r}}=\mathbf{r}=x\hat{\mathbf{x}}+y\hat{\mathbf{y}}+z\hat{\mathbf{z}} \tag{2}$$

$$\begin{aligned}\hat{\mathbf{p}}&=-\mathrm{i}\hbar\boldsymbol{\nabla}=\hat{p}_x\hat{\mathbf{x}}+\hat{p}_y\hat{\mathbf{y}}+\hat{p}_z\hat{\mathbf{z}}\\&=\left(-\mathrm{i}\hbar\frac{\partial}{\partial x}\right)\hat{\mathbf{x}}+\left(-\mathrm{i}\hbar\frac{\partial}{\partial y}\right)\hat{\mathbf{y}}+\left(-\mathrm{i}\hbar\frac{\partial}{\partial z}\right)\hat{\mathbf{z}}\end{aligned} \tag{3}$$

当矢量算符作用在波函数上后，得到的函数的自变量仍然是 $(\mathbf{r},t)$，而函数值却变为一个复数矢量（矢量的三个分量都是复数）．把位置算符作用在波函数上，就是把波函数分别乘以 x,y,z 的坐标，并作为函数值的三个分量．而把动量算符作用在波函数上，就先把各个方向的动量算符分别作用，并作为函数值的三个分量．所以在矢量算符的本征方程

$$\hat{\mathbf{Q}}\Psi(\mathbf{r})=\boldsymbol{\lambda}\Psi(\mathbf{r}) \tag{4}$$

中，我们只有使用矢量本征值才能保证等号两边都是矢量．矢量算符的本征方程也可以写成三个分量的形式

$$\hat{Q}_x\Psi(\mathbf{r})=\lambda_x\Psi(\mathbf{r})\qquad\hat{Q}_y\Psi(\mathbf{r})=\lambda_y\Psi(\mathbf{r})\qquad\hat{Q}_z\Psi(\mathbf{r})=\lambda_z\Psi(\mathbf{r}) \tag{5}$$

不难验证，位置的本征函数就是三维 δ 函数

$$\delta(\mathbf{r}-\mathbf{r}_0)=\delta(x-x_0)\delta(y-y_0)\delta(z-z_0) \tag{6}$$

对应的本征值为 $\mathbf{r}_0$．动量的本征函数就是三维平面波

$$\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r})=\exp(\mathrm{i}k_xx)\exp(\mathrm{i}k_yy)\exp(\mathrm{i}k_zz) \tag{7}$$

对应的本征值为 $\mathbf{p}=\hbar\mathbf{k}$.

无限深势阱

预备知识 定态薛定谔方程[1328]，二阶常系数齐次微分方程[215]

只考察质量为 m 的粒子沿 x 方向的运动情况．势能函数为

$$V(x)=\begin{cases}0 & (0\leqslant x\leqslant a)\\ +\infty & (x<0 \text{ 或 } x>a)\end{cases} \tag{1}$$

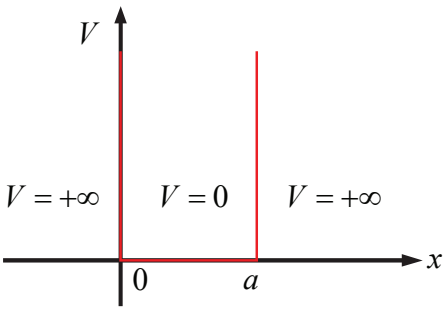


图 1: 无限深势阱

求解定态薛定谔方程

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi(x)+V(x)\psi(x)=E\psi(x) \tag{2}$$

结论

第 n 个能级为（$n=1,2,3\ldots$）

$$E_n=\frac{\pi^2\hbar^2}{2ma^2}n^2 \tag{3}$$

对应的能量本征波函数为

$$\psi_n(x)=\sqrt{\frac{2}{a}}\sin\left(\frac{n\pi}{a}x\right) \tag{4}$$

一些教材中也会把势阱中间作为坐标原点，这相当于把这里的势能 $V(x)$ 和所有 $\psi_n(x)$ 都向左移动 $a/2$．能级仍然保持不变，波函数变为（为了简洁适

当乘以 -1，物理意义不变）

$$\psi_n(x)=\sqrt{\frac{2}{a}}\begin{cases}\sin(n\pi x/a) & (\text{偶数}n)\\ \cos(n\pi x/a) & (\text{奇数}n)\end{cases} \tag{5}$$

事实上我们可以通过平移的方法把势阱移动到任何位置.

推导

先考虑势阱内部（$0\leqslant x\leqslant a$，$V=0$），方程变为

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi(x)=E\psi(x) \tag{6}$$

这是二阶常系数齐次微分方程．通解为

$$\psi(x)=C_1\cos(kx)+C_2\sin(kx)\qquad k=\frac{\sqrt{2mE}}{\hbar} \tag{7}$$

其中

（通解也可以写成指数函数 $C\mathrm{e}^{\mathrm{i}kx}$，加上边界条件后的结论一样）现在讨论边界条件：在有限深势阱束缚态中将会看到，如果势阱外部势能是有限值，波函数将会按照指数函数衰减，势能越高衰减得越快．而现在势阱外部势能为无穷大，就可以直接认为波函数在势阱外部始终为零．所以边界条件为

$$\psi(0)=0,\quad \psi(a)=0 \tag{8}$$

这两个条件代入以上通解中，解得

$$C_2=0,\quad k=\frac{n\pi}{a}\ \ (n=1,2,3\dots) \tag{9}$$

C_1 的取值暂时不能确定，但先将通解写为 $\psi(x)=C\sin\left(\frac{n\pi}{a}x\right)$，常数 C 就可以通过波函数的归一化来确定：

$$1=\int_{-\infty}^{+\infty}|\psi(x)|^2\,\mathrm{d}x=\int_{-a}^{+a}\left|C\sin\left(\frac{n\pi}{a}x\right)\right|^2\mathrm{d}x=|C|^2\frac{a}{2} \tag{10}$$

严格来说，C 可以是复数，解为 $C=\sqrt{2/a}\mathrm{e}^{\mathrm{i}\theta}$．但是为了方便通常把归一化常数中的相位因子 $\mathrm{e}^{\mathrm{i}\theta}$ 默认为 1．所以归一化的波函数为

$$\psi(x)=\sqrt{\frac{2}{a}}\sin\left(\frac{n\pi}{a}x\right) \tag{11}$$

另外，由

$$\frac{\sqrt{2mE}}{\hbar} = k = \frac{n\pi}{a} \tag{12}$$

可以得出能级是离散的结论．即

$$E_n = \frac{\pi^2\hbar^2}{2ma^2}n^2 \quad (n = 1, 2, 3\ldots) \tag{13}$$

无限深势阱中的高斯波包

预备知识 无限深势阱[1335]，高斯波包[1012]

我们下面来计算无限深势阱中一个高斯波包的运动．定性来说，波包会一边移动一边扩散（变宽变矮），且在两个势阱壁之间来回反弹．反弹的过程中会发生干涉．

结果如图 1 和图 2，动画见百科互动演示．

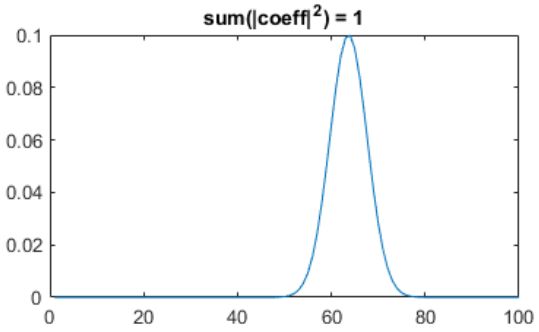


图 1: 束缚态概率分布，x 轴为束缚态的 n，y 轴为概率，求和约等于 1

计算

本文使用原子单位制[1315]，并假设粒子质量为 1．假设无限深势阱的区间为 $[0, L]$，能量的本征波函数（本征态）为（式 4 [1335]）

$$\psi_n(x) = \sqrt{\frac{2}{L}}\sin(k_n x) \tag{1}$$

$$k_n = \frac{n\pi}{L} \tag{2}$$

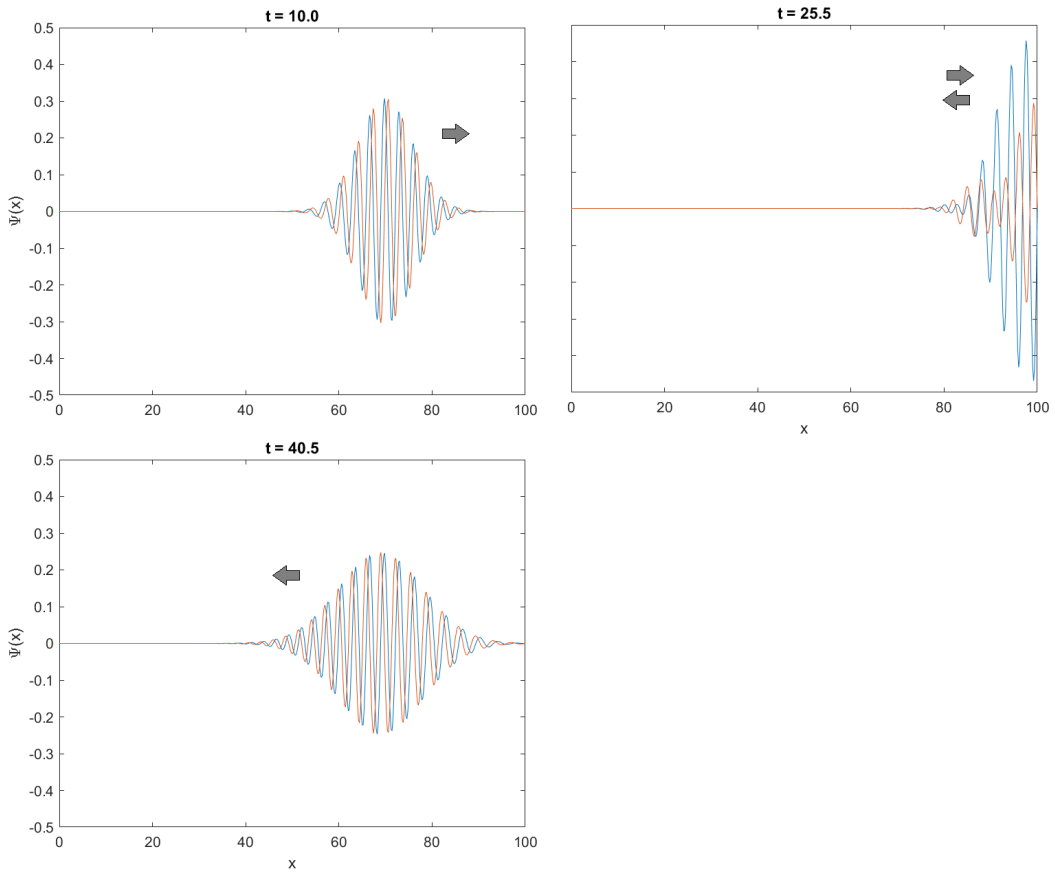


图 2: 波包遇到势阱壁后发生反弹，过程中发生干涉

能量的本征值为

$$E_n = \frac{k_n^2}{2} \tag{3}$$

初始时波函数为高斯波包（式 1 [1346]）

$$\psi(x,0) = \frac{1}{(2\pi\sigma_x^2)^{1/4}} \mathrm{e}^{-(x-x_0)^2/(2\sigma_x)^2} \mathrm{e}^{\mathrm{i}k_0 x} \tag{4}$$

第一步是把初始波函数投影到能量本征态上

$$C_n = \int_0^L \psi_n^*(x)\psi(x,0)\,\mathrm{d}x = \int_0^L \sqrt{\frac{2}{L}} \sin(k_n x)\psi(x,0)\,\mathrm{d}x \tag{5}$$

那么接下来，波函数的演化可以表示为

$$\psi(x,t) = \sum_i C_i \mathrm{e}^{-\mathrm{i}E_n t}\psi_n(x) \tag{6}$$

若我们假设初始波包宽度足够小，使得波函数在势阱外的函数值可以忽略不记，则式 5 的定积分可以拓展到无穷区间，即傅里叶变换．我们已经知道初始高斯波包（指数）傅里叶变换的结果为式 2 [1346]

$$g(k)=\frac{1}{(2\pi\sigma_p^2)^{1/4}}\mathrm{e}^{-(k-k_0)^2/(2\sigma_k)^2}\mathrm{e}^{-\mathrm{i}x_0(k-k_0)} \tag{7}$$

将正弦函数记为指数的形式为

$$\sqrt{\frac{2}{L}}\sin(k_nx)=\mathrm{i}\sqrt{\frac{\pi}{L}}\left(\frac{\mathrm{e}^{-\mathrm{i}k_nx}}{\sqrt{2\pi}}-\frac{\mathrm{e}^{\mathrm{i}k_nx}}{\sqrt{2\pi}}\right) \tag{8}$$

所以式 5 变为

$$C_n=\mathrm{i}\sqrt{\frac{\pi}{L}}[g(k_n)-g(-k_n)] \tag{9}$$

代入式 6 即可．

升降算符

- 本词条处于草稿阶段．

预备知识 本征方程

结论

已知某个算符 $\hat{Q}$，若能找到另一个算符 $\hat{Q}_+$，使得 $[\hat{Q},\hat{Q}_+]=h\hat{Q}_+$ 成立（h 是大于零的实数），这个算符就是 $\hat{Q}$ 对应的升算符．同理，若有 $\hat{Q}_-$ 使得 $[\hat{Q},\hat{Q}_-]=-h\hat{Q}_-$ 成立，这个算符就是对应的降算符．升降算符的作用是把一个本征函数变为本征值更大或者更小的本征函数．即

$$\hat{Q}(\hat{Q}_\pm\psi)=(q\pm h)(\hat{Q}_\pm\psi) \tag{1}$$

意义

有时候如果算符过于复杂求解本征方程比较困难，就可以尝试寻找升降算符．升降算符可以让我们不用求解本征方程就可以快速地找到本征值．具体见简谐振子和轨道角动量．

证明

如果 ψ 是 $\hat{Q}$ 的一个本征函数，且本征值为 λ，那么根据对易关系 $[\hat{Q},\hat{Q}_+]=h\hat{Q}_+$ 有

$$\hat{Q}(\hat{Q}_+\psi)=\hat{Q}_+(\hat{Q}\psi)+h\hat{Q}_+\psi=\hat{Q}_+(\lambda\psi)+h\hat{Q}_+\psi=(\lambda+h)(\hat{Q}_+\psi) \tag{2}$$

降算符的证明同理.

本征函数的归一化

注意升降算符并不一定能保持函数的归一化. 若假设 $\hat{a}_\pm|\psi_n\rangle=A_n|\psi_{n+1}\rangle$，其中 $|\psi_n\rangle$ 和 $|\psi_{n+1}\rangle$ 都是归一化的本征态，那么由归一化条件要求

$$\langle\psi_n|\hat{a}_\pm^\dagger\hat{a}_\pm|\psi_n\rangle=|A_n|^2\langle\psi_{n+1}|\psi_{n+1}\rangle=|A_n|^2 \tag{3}$$

习惯上令 A_n 为实数，即上式开方.

简谐振子（升降算符）

预备知识 简谐振子[913]，升降算符[1339]

结论

$$a_\pm=\frac{1}{\sqrt{2m\omega\hbar}}(m\omega x\mp\mathrm{i}p)\quad a_+\psi_n=\sqrt{n+1}\psi_{n+1}\quad a_-\psi_n=\sqrt{n}\psi_{n-1} \tag{1}$$

$$E_n=\left(n+\frac{1}{2}\right)\omega\hbar\qquad\hat{n}=a_+a_-\qquad H=\omega\hbar\left(\hat{n}+\frac{1}{2}\right) \tag{2}$$

$$\psi_0(x)=\frac{1}{\pi^{1/4}\beta^{1/2}}\mathrm{e}^{-(x/\beta)^2/2}\qquad\beta=\sqrt{\frac{\hbar}{m\omega}} \tag{3}$$

推导

在经典的简谐振子模型中，若质点沿 x 轴方向振动，且在 $x=0$ 处平衡，则势能函数 $V(x)=kx^2/2$．由于自由振动的频率为 $\omega=\sqrt{k/m}$，所以势能可记为

$$V(x)=\frac{1}{2}m\omega^2x^2 \tag{4}$$

在量子力学中，这个模型要用薛定谔方程来求解．该模型的哈密顿算符为

$$H=\frac{p^2}{2m}+V=\frac{p^2}{2m}+\frac{1}{2}m\omega^2x^2=\frac{1}{2m}[p^2+(m\omega x)^2] \tag{5}$$

定态薛定谔方程（能量的本征方程）为

$$H\psi=E\psi \tag{6}$$

由于这个方程需要使用幂级数，但作为一种巧妙的方法，先利用升降算符[1339]来得到能量的本征值，再求本征函数．这里直接给出 H 的升降算符，它们分别可以把本征值升降 $\omega\hbar$ （证明见下文）

$$a_\pm=\frac{1}{\sqrt{2m\omega\hbar}}(m\omega x\mp \mathrm{i}p)\qquad \Delta E=\omega\hbar \tag{7}$$

根据式 1 [1339]，对任意一个 H 的本征函数 ψ_n，有

$$H(a_\pm\psi_n)=(E_n\pm\hbar\omega)(a_\pm\psi_n) \tag{8}$$

这也就是说，简谐振子的定态薛定谔方程的解中，本征值 E_n 取离散值，且相邻两个能级相差 $\Delta E=\hbar\omega$.

类似于无限深势阱[1335]，谐振子也应该有一个最低能级 E_0 和对应的 $\psi_0(x)$．所以 a_- 必然对 ψ_0 无效，即得到的波函数没有物理意义，所以不妨猜测 $a_-\psi_0=0$ 即

$$(m\omega x+\mathrm{i}p)\psi_0=0\quad\Rightarrow\quad\frac{\mathrm{d}}{\mathrm{d}x}\psi_0=-\frac{m\omega x}{\hbar}\psi_0 \tag{9}$$

这是一阶齐次线性微分方程，通解为

$$\psi_0(x)=\frac{1}{\pi^{1/4}\sqrt{\beta}}\mathrm{e}^{-(x/\beta)^2/2} \tag{10}$$

其中 $\beta=\sqrt{\hbar/m\omega}$ 具有长度量纲．不难验证上式是定态薛定谔方程式 6 的解，本征值为 $E_0=\omega\hbar/2$.

归一化条件

证明 $\psi_{n+1}=a_+\psi_n/\sqrt{n+1}$，$\psi_{n-1}=a_-\psi_n/\sqrt{n}$

$$\begin{aligned}A^2&=\int(a_+\psi_n)^*(a_+\psi_n)\,\mathrm{d}x=\int(a_+^*a_+\psi_n)^*(\psi_n)\,\mathrm{d}x\\&=\int(a_-a_+\psi_n)^*(\psi_n)\,\mathrm{d}x\end{aligned}\tag{11}$$

而 $a_-a_+\psi_n=H\psi_n/(\omega\hbar)+\psi_n/2=(n+1)\psi_n$．所以

$$A^2=(n+1)\int\psi_n^*\psi_n\,\mathrm{d}x=n+1\qquad A=\sqrt{n+1}\tag{12}$$

所以 $\psi_{n+1}=a_+\psi_n/\sqrt{n+1}$

波函数

现在只需要对基态波函数不断使用升算符和归一化系数就可以得到任意激发态波函数（已归一化）

$$\psi_n=\frac{a_+}{\sqrt{n!}}\psi_0\tag{13}$$

升降算符的证明

根据升降算符的定义，要证明 $a_\pm$ 是升降算符只需证明对易关系

$$[H,a_\pm]=\pm\omega\hbar a_\pm\tag{14}$$

根据升降算符定义

$$[H,a_\pm]=\frac{1}{2m\sqrt{2m\hbar\omega}}\left[(m\omega x)^2+p^2,m\omega x\mp\mathrm{i}p\right]\tag{15}$$

不难证明

$$\left[\hat{A}+\hat{B},\hat{C}+\hat{D}\right]=\left[\hat{A},\hat{C}\right]+\left[\hat{A},\hat{D}\right]+\left[\hat{B},\hat{C}\right]+\left[\hat{B},\hat{D}\right]\tag{16}$$

和

$$\left[x^2,x\right]=\left[p^2,p\right]=0\tag{17}$$

将两式代入后，只剩下两个交叉项

$$[H,a_\pm]=\frac{\omega}{2\sqrt{2m\hbar\omega}}\left(\mp\mathrm{i}\left[x^2,p\right]+\left[p^2,x\right]\right) \tag{18}$$

同样可以证明

$$\left[\hat{A}\hat{B},\hat{C}\right]=\hat{A}\left[\hat{B},\hat{C}\right]+\left[\hat{A},\hat{C}\right]\hat{B} \tag{19}$$

所以

$$\left[x^2,p\right]=x[x,p]+[x,p]x=2\mathrm{i}\hbar x \tag{20}$$

$$\left[p^2,x\right]=-p[x,p]-[x,p]p=-2\mathrm{i}\hbar p \tag{21}$$

代入得

$$[H,a_\pm]=\frac{\pm\omega\hbar}{\sqrt{2m\hbar\omega}}(x\mp\mathrm{i}p)=\pm\omega\hbar a_\pm \tag{22}$$

证毕.

简谐振子升降算符归一化

预备知识 简谐振子（升降算符）[1340]

首先要提醒的是，一般算符满足的一个条件是 $\langle g|\hat{Q}f\rangle=\langle\hat{Q}^\dagger g|f\rangle$. 但是对于厄米算符，$\hat{Q}^\dagger=\hat{Q}$，所以有 $\langle g|\hat{Q}f\rangle=\langle\hat{Q}g|f\rangle$.

对于谐振子的升降算符 $a_\pm=(m\omega x\mp\mathrm{i}p)/\sqrt{2m\omega\hbar}$，有

$$\begin{aligned}a_-a_+&=\frac{1}{2m\omega\hbar}(m^2\omega^2x^2+p^2-\mathrm{i}m\omega[x,p])\\&=\frac{1}{\omega\hbar}\left[\frac{1}{2m}(m^2\omega^2x^2+p^2)+\frac{\omega\hbar}{2}\right]\\&=\frac{1}{\omega\hbar}H+\frac{1}{2}\end{aligned} \tag{1}$$

$$\begin{aligned}|a_+\psi_n|^2=\langle a_+\psi_n|a_+\psi_n\rangle=\langle\psi_n|a_-a_+\psi_n\rangle&=\left\langle\psi_n\middle|\left(\frac{1}{\omega\hbar}H+\frac{1}{2}\right)\psi_n\right\rangle\\&=\left(n+\frac{1}{2}\right)+\frac{1}{2}=n+1\end{aligned} \tag{2}$$

所以有 $a_+\psi_n=\sqrt{n+1}\psi_{n+1}$ （同理 $a_-\psi_n=\sqrt{n}\,\psi_{n-1}$ ）. 再次提醒，归一化系数后面可以加上任意相位因子 $\mathrm{e}^{\mathrm{i}\theta}$，同样能满足归一化条件，但一般省略.

简谐振子（级数）

预备知识 简谐振子（升降算符）[1340]

结论

量子力学中，简谐振子的能级为

$$E_n=\left(\frac{1}{2}+n\right)\hbar\omega \tag{1}$$

波函数为

$$\psi_n(x)=\frac{1}{\sqrt{2^n n!}}\left(\frac{\alpha^2}{\pi}\right)^{1/4}H_n(u)\mathrm{e}^{-u^2/2} \tag{2}$$

其中

$$\alpha\equiv\sqrt{\frac{m\omega}{\hbar}}\qquad u\equiv\alpha x \tag{3}$$

$H_n(u)$ 叫做 Hermite 多项式

$$H_n(u)\equiv(-1)^n\mathrm{e}^{u^2}\frac{\mathrm{d}^n}{\mathrm{d}u^n}\left(\mathrm{e}^{-u^2}\right) \tag{4}$$

前 6 阶 Hermite 多项式分别为

$$\begin{aligned}&H_0(u)=1 && H_3(u)=8u^3-12u\\&H_1(u)=2u && H_4(u)=16u^4-48u^2+12\\&H_2(u)=4u^2-2 && H_5(u)=32u^5-160u^3+120u\end{aligned} \tag{5}$$

前 4 个波函数分别为（注意函数的奇偶性与角标的奇偶性相同）

$$\begin{aligned}&\psi_0(x)=\left(\frac{\alpha^2}{\pi}\right)^{1/4}\mathrm{e}^{-u^2/2} && \psi_2(x)=\left(\frac{\alpha^2}{\pi}\right)^{1/4}\frac{1}{\sqrt{2}}(2u^2-1)\mathrm{e}^{-u^2/2}\\&\psi_1(x)=\left(\frac{\alpha^2}{\pi}\right)^{1/4}\sqrt{2}u\mathrm{e}^{-u^2/2} && \psi_3(x)=\left(\frac{\alpha^2}{\pi}\right)^{1/4}\frac{1}{\sqrt{3}}u(2u^2-3)\mathrm{e}^{-u^2/2}\end{aligned} \tag{6}$$

推导

薛定谔方程为

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}+\frac{1}{2}m\omega^2x^2\psi=E\psi \tag{7}$$

一维自由粒子（量子）

- 本词条存在未完成的内容.

预备知识 傅里叶变换与连续正交归一基底[240]，薛定谔方程[1331]，原子单位制[1315]

本文使用原子单位制[1315]．当含时薛定谔方程[1331] 中势能函数 $V(x)=0$ 时，有

$$-\frac{1}{2m}\frac{\partial^2}{\partial x^2}\Psi(x,t)=\mathrm{i}\frac{\partial}{\partial t}\Psi(x,t) \tag{1}$$

一般用分离变量法解该方程，通解为式 10 [1331]．但首先要解出对应的定态薛定谔方程

$$-\frac{1}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi_E(x)=E\psi_E(x) \tag{2}$$

只有 $E>0$ 时有可归一化的解，也就是熟悉的平面波（链接未完成）

$$\psi_E(x)=\mathrm{e}^{\mathrm{i}kx} \tag{3}$$

其中 $k=\pm\sqrt{2mE}$，注意定态薛定谔方程具有二重简并，即一个 E 对应两个线性无关解．令 $\omega=E$，则通解为

$$\Psi(x,t)=\int_0^{+\infty}\left[A_+(\omega)\mathrm{e}^{\mathrm{i}(|k|x-\omega t)}+A_-(\omega)\mathrm{e}^{\mathrm{i}(-|k|x-\omega t)}\right]\mathrm{d}\omega \tag{4}$$

一种更简洁的表示方法是做变量替换 $\omega(k)=k^2/(2m)$，把通解记为

$$\Psi(x,t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}C(k)\mathrm{e}^{\mathrm{i}(kx-\omega t)}\,\mathrm{d}k \tag{5}$$

直观上来理解，式 4 和式 5 都表示许多不同频率的平面波的叠加，所以是等效的，系数 $A_\pm(\omega)$ 和 $C(k)$ 具有对应关系，以后我们会看到他们之间如何转换[1348]．注意式 5 恰好是 $C(k)$ 的反傅里叶变换（式 2 [237]）．当 $t=0$ 时有

$$C(k)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\Psi(x,0)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x \tag{6}$$

这样我们就从初始波函数 $\Psi(x,0)$ 得到了系数.

式 5 相当于把波函数在正交归一的基底 $\psi_k(x)=\mathrm{e}^{\mathrm{i}kx}/\sqrt{2\pi}$ 展开，基底满足（式 11 [240]）

$$\langle\psi_{k'}|\psi_k\rangle=\delta(k'-k) \tag{7}$$

我们把这种归一化叫做**动量归一化**（原子单位中 k 就是动量）；而式 4 则把波函数在另一组正交归一基底上展开，这组基底一般使用**能量归一化**，其基底 $\psi_{E,1},\psi_{E,2}$（方括号中的两项）满足

$$\langle\psi_{E',i'}|\psi_{E,i}\rangle=\delta_{i',i}\delta(E'-E) \tag{8}$$

详见“动量归一化和能量归一化[1348]”.

一个简单的例子见“一维高斯波包（量子）[1346]”.

事实上，无论我们使用什么样的初始波包，它接下来的演化都是平淡无奇的，可以定性地划分为匀速移动和扩散. 在熟悉了本文以后，我们可以开始考虑一个稍微复杂一些的问题：当势能函数 $V(x)$ 在某个区间不为零时，波包经过这个区间会如何变化？这就是一维散射问题[1351].

一维自由高斯波包（量子）

- 本词条需要更多讲解，便于帮助理解.
- 本词条存在未完成的内容.

预备知识 无限深势阱中的高斯波包[1337]，一维自由粒子（量子）[1345]

我们这里要讨论的是量子力学中单个粒子的一个特定的一维波函数 $\psi(x,t)$. 它描述的粒子的空间分布是一个高斯分布（正态分布）[731].

设 $t=0$ 时的波函数（已归一化）

$$\psi(x,0)=\frac{1}{(2\pi\sigma_x^2)^{1/4}}\mathrm{e}^{-(x-x_0)^2/(2\sigma_x)^2}\mathrm{e}^{\mathrm{i}\frac{p_0}{\hbar}x} \tag{1}$$

那么动量表象波函数具有对称的形式[16]

$$\psi(p,0)=\frac{1}{(2\pi\sigma_p^2)^{1/4}}\mathrm{e}^{-(p-p_0)^2/(2\sigma_p)^2}\mathrm{e}^{-\mathrm{i}\frac{x_0}{\hbar}(p-p_0)} \tag{2}$$

[16]也可以把式 1 和式 2 同时除以常数 $\mathrm{e}^{\mathrm{i}p_0x_0}$ 使式 1 最后的 x 变为 $x-x_0$，式 2 最后的 $p-p_0$ 变为 p.

其中 σ_x 为位置的标准差，σ_p 为动量的标准差，满足不确定原理

$$\sigma_x\sigma_p=\frac{\hbar}{2} \tag{3}$$

含时波函数为

$$\psi(x,t)=\frac{1}{(2\pi\sigma_x^2)^{1/4}}\frac{1}{\sqrt{1+\frac{\mathrm{i}\hbar t}{2m\sigma_x^2}}}\exp\left[\frac{-(x-p_0t/m)^2}{(2\sigma_x)^2\left(1+\frac{\mathrm{i}\hbar t}{2m\sigma_x^2}\right)}\right]\exp\left[\frac{\mathrm{i}p_0}{\hbar}\left(x-\frac{p_0t}{2m}\right)\right] \tag{4}$$

推导

如果我们想要一维波函数的概率分布为高斯分布[731]，即

$$|\psi(x)|^2=\frac{1}{\sigma_x\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2\sigma_x^2}} \tag{5}$$

先假设波函数为实数，有

$$\psi(x)=\frac{1}{(2\pi\sigma_x^2)^{1/4}}\mathrm{e}^{-\left(\frac{x}{2\sigma_x}\right)^2} \tag{6}$$

变换到动量表象，得[17]

$$\psi(p)=\frac{1}{(2\pi\sigma_p^2)^{1/4}}\mathrm{e}^{-\left(\frac{p}{2\sigma_p}\right)^2} \tag{7}$$

其中 $\sigma_p=\hbar/(2\sigma_x)$，可见高斯波包一个独特的性质就是在位置和动量表象下都是高斯分布.

由于波函数为实数，动量平均值为零. 为了让波函数有一个动量 p_0，而维持 $|\psi(x)|^2$ 和 $|\psi(p)|^2$ 的形状不变，我们可以直接将动量表象中的波函数平移 p_0，得

$$\psi(p)=\frac{1}{(2\pi\sigma_p^2)^{1/4}}\mathrm{e}^{-\left(\frac{p-p_0}{2\sigma_p}\right)^2} \tag{8}$$

由傅里叶变换的性质，对应的位置表象波函数需要乘以因子 $\exp(\mathrm{i}p_0x/\hbar)$ 变为

$$\psi(x)=\frac{1}{(2\pi\sigma_x^2)^{1/4}}\mathrm{e}^{-\left(\frac{x}{2\sigma_x}\right)^2}\mathrm{e}^{\mathrm{i}p_0x/\hbar} \tag{9}$$

[17]可以用 Wolfram Alpha 或 Mathematica 计算积分.

类似地，也可以将 $\psi(x)$ 平移 x_0

$$\psi(x)=\frac{1}{(2\pi\sigma_x^2)^{1/4}}\mathrm{e}^{-\left(\frac{x-x_0}{2\sigma_x}\right)^2}\mathrm{e}^{\mathrm{i}p_0(x-x_0)/\hbar} \tag{10}$$

而 $\psi(p)$ 则需要乘以因子 $\exp(-\mathrm{i}x_0p/\hbar)$

$$\psi(p)=\frac{1}{(2\pi\sigma_p^2)^{1/4}}\mathrm{e}^{-\left(\frac{p-p_0}{2\sigma_p}\right)^2}\mathrm{e}^{-\mathrm{i}x_0p/\hbar} \tag{11}$$

将以上两式同乘一个常数 $\mathrm{e}^{\mathrm{i}p_0x_0/\hbar}$ 就得到式 1 和式 2.

平面波的的正交归一化

预备知识 一维自由粒子（量子）[1345]

本文使用原子单位制[1315]. 回顾一维自由粒子[1345] 中的分析，我们需要把能量本征态即平面波进行**动量归一化**或者**能量归一化**

$$\int_{-\infty}^{+\infty}\psi_{k'}(x)^*\psi_k(x)\,\mathrm{d}x=\delta(k'-k)\qquad(k\in\mathbb{R}) \tag{1}$$

$$\int_{0}^{+\infty}\psi_{E',i'}(x)^*\psi_{E,i}(x)\,\mathrm{d}x=\delta_{i',i}\delta(E'-E)\qquad(E>0) \tag{2}$$

其中 $*$ 表示复共轭，满足该式的 ψ_k 就是动量归一化的平面波：

$$\psi_k(x)=\frac{\mathrm{e}^{\mathrm{i}kx}}{\sqrt{2\pi}} \tag{3}$$

但什么样的归一化系数能满足能量归一化条件呢？

我们可以令定态薛定谔方程的两个线性无关解 $\psi_{E,1}(x),\psi_{E,2}(x)$ 分别为 $-k$ 和 k 的平面波（暂时规定 $k=\sqrt{2mE}>0$），只是归一化系数和式 3 不同. 只需要使用 δ 函数的性质式 13 [230]，把 $E'-E$ 看成 k 的函数（k' 视为常数）[18]，唯一的正根为 k'

$$f(k)=\frac{k'^2}{2m}-\frac{k^2}{2m}\qquad|f'(k')|=\frac{k'}{m} \tag{4}$$

[18] 反之亦然：k' 看成变量，k 看成常数

所以

$$\delta(E'-E)=\delta[f(k)]=\frac{m}{k'}\delta(k'-k)=\frac{m}{k}\delta(k'-k) \tag{5}$$

代入式 2 得**能量归一化**后的平面波

$$\psi_{E,1}(x)=\sqrt{\frac{m}{2\pi k}}\mathrm{e}^{\mathrm{i}kx}\qquad \psi_{E,2}(x)=\sqrt{\frac{m}{2\pi k}}\mathrm{e}^{-\mathrm{i}kx} \tag{6}$$

现在，任意波函数 $\psi(x)$ 就可以用能量归一化的基底展开为

$$\begin{aligned}\psi(x)&=\int_0^{+\infty}[A(k)\psi_{E,1}(x)+A(-k)\psi_{E,2}(x)]\,\mathrm{d}E\\&=\int_0^{+\infty}\sqrt{\frac{k}{m}}\left[A(k)\frac{\mathrm{e}^{\mathrm{i}kx}}{\sqrt{2\pi}}+A(-k)\frac{\mathrm{e}^{-\mathrm{i}kx}}{\sqrt{2\pi}}\right]\mathrm{d}k\\&=\int_{-\infty}^{+\infty}\sqrt{\frac{|k|}{m}}A(k)\psi_k(x)\,\mathrm{d}k\end{aligned} \tag{7}$$

对比用动量归一化展开

$$\psi(x)=\int_{-\infty}^{+\infty}C(k)\psi_k(x)\,\mathrm{d}k \tag{8}$$

两种系数间满足关系

$$C(k)=\sqrt{\frac{|k|}{m}}A(k) \tag{9}$$

其他正交归一本征态

事实上归一化其实有无数种，因为每个能量 E 的本征波函数空间是一个二维简并空间，我们可以把以上正交归一的 $\psi_{E,1},\psi_{E,2}$ 做一个任意幺正变换后仍然得到两个正交归一的解．最常见的方法例如令（角标 o 代表 odd，e 代表 even）

$$\begin{aligned}\psi_{E,e}(x)&=\frac{1}{\sqrt{2}}[\psi_{E,1}(x)+\psi_{E,2}(x)]=\sqrt{\frac{m}{\pi k}}\cos(kx)\\\psi_{E,o}(x)&=\frac{1}{\sqrt{2}\,\mathrm{i}}[\psi_{E,1}(x)-\psi_{E,2}(x)]=\sqrt{\frac{m}{\pi k}}\sin(kx)\end{aligned} \tag{10}$$

他们满足式 2 （习题）．这个变换使用了矩阵 $\begin{pmatrix}1 & 1\\ -\mathrm{i} & \mathrm{i}\end{pmatrix}$，可以证明使用任何 2 乘 2 的酉矩阵[440] 做变换都可以．

令 $k>0$，把 $\psi_{\pm k}$ 做类似的变换，我们也可以写出动量归一化的一种变体[19]

$$\begin{aligned}\psi_{k,e}(x)&=\frac{1}{\sqrt{2}}[\psi_k(x)+\psi_{-k}(x)]=\frac{1}{\sqrt{\pi}}\cos(kx)\\ \psi_{k,o}(x)&=\frac{1}{\sqrt{2}\,\mathrm{i}}[\psi_k(x)-\psi_{-k}(x)]=\frac{1}{\sqrt{\pi}}\sin(kx)\end{aligned}\qquad(k>0)\tag{11}$$

这就是三角傅里叶变换[236] 中使用的基底．他们满足归一化条件（习题 1 [230]）

$$\int_{-\infty}^{+\infty}\psi_{k',i}(x)^*\psi_{k,i}(x)\,\mathrm{d}x=\delta(k'-k)\qquad(k>0,i=e,o)\tag{12}$$

无限深圆形势阱

- 本词条处于草稿阶段．

预备知识 定态薛定谔方程[1328]，柱坐标中的亥姆霍兹方程[752]

$$-\frac{1}{2m}\nabla^2\psi(\mathbf{r})=E\psi(\mathbf{r})\tag{1}$$

令 $k=\sqrt{2mE}$，得到标准形式的亥姆霍兹方程

$$\nabla^2\psi(\mathbf{r})=-k^2\psi(\mathbf{r})\tag{2}$$

使用分离变量法，角向波函数为

$$\Theta_{m_z}(\theta)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{im_z\theta}\tag{3}$$

径向方程为式 6 [752]（$l=0$，$x=kr$）

$$x\frac{\mathrm{d}}{\mathrm{d}x}\left(x\frac{\mathrm{d}y}{\mathrm{d}x}\right)+(x^2-m_z^2)y=0\tag{4}$$

所以径向波函数为

$$R_{m_z}(r)=J_{m_z}(kr)\tag{5}$$

[19]注意这样得到的并不是动量算符的本征函数，所以严格来说动量归一化只有一种可能．

从物理上来看，能量分为角速度产生的能量以及径向速度产生的能量，角动量已经由 m_z 固定，所以 $r \to 0$ 时角速度产生的能量占主导，$J_{m_z}(kr)$ 的局部波长变长，而 $r \to \infty$ 时径向速度几乎占据所有能量，这时 $J_{m_z}(kr)$ 的局部波数趋近于 k.

再考虑到边界条件要求 $R_{m_z}(a) = 0$，可以得到每个能级的 k_n，$E_n = k_n^2/2$. 波函数的正交性也可以从贝塞尔函数对应的正交归一性质得到.

有限深球势阱

- 本词条处于草稿阶段.

预备知识 球贝塞尔函数[759]

束缚态

角向方程为球谐函数，势阱内部的径向波函数（unscaled）为球贝塞尔函数 $j_l(kr)$，外部为第二类修正球贝塞尔函数（式 13 [759]）$k_l(\kappa r)$，$k = \sqrt{2mE}$ 和 $\kappa = \sqrt{2m(V-E)}$. 在势阱边界处，函数和一阶导数需要相等.

一维散射（量子）

- 本词条存在未完成的内容.

预备知识 平面波的的动量和能量归一化[1348]

本文使用原子单位制[1315]. 在量子力学中，散射是一个十分重要的过程. 例如为了探索例如原子的结构，一种重要的手段就是用其他粒子轰击原子，并探测出射粒子在不同方向上的动量或能量分布等. 著名的卢瑟福散射就是用电子轰击原子的方法解开了原子结构之谜. 虽然卢瑟福最初使用经典力学来分析这类散射过程（当时量子力学还没有出现），但原子尺度下量子力学比经典力学要精确得多.

在经典力学中，散射[1047] 过程考虑一个粒子从无穷远处入射，经过一个势能 $V(\mathbf{r})$ 后发生偏折．量子力学中，这一过程可以用波包来描述：一个波包 $\psi(x,t)$ 在初始时刻从远处入射，我们想知道经过 $V(\mathbf{r})$ 散射后，$\psi(x,t)$ 会如何变化，例如不同出射方向的概率流密度[1386] 以及动量如何分布？

由于三维空间的量子散射所需的数学较为复杂，我们先学习一维散射．三维散射的许多性质都可以从一维情况类比，但也有更丰富的特性．

我们已经学习了一维自由粒子[1345] 如何随时间演化，例如一维高斯波包[1346] 在自由演化过程中，他的中心会像经典粒子一样以恒定速度移动，但同时波包还会慢慢变宽．在此基础上，如果在波包前进方向上添加一个不为零的 $V(x)$ （例如一个势垒或势阱图 3 [60]），那么波包将会在各个方向上发生散射（一维运动只有左右两个方向）．我们希望能计算该过程中波函数如何变化．我们要求势能函数 $V(x)$ 只在局部不为零，即满足 $V(\pm\infty)\to 0$．

原则上，我们可以直接把初始波函数和势能代入含时薛定谔方程[1331] 进行求解

$$-\frac{1}{2m}\frac{\partial^2}{\partial t^2}\psi(x,t)+V(x)\psi(x,t)=\mathrm{i}\frac{\partial}{\partial t}\psi(x,t) \tag{1}$$

即使没有解析解，也可以通过数值方法，依次求出每个时间步长 t_n （$n=1,2,\dots$）的波函数 $\psi(x,t_n)$．

解析分析

即使能轻易得到数值解，解析的分析也能帮我们更好地理解散射问题（或者任何问题）．和“一维自由粒子（量子）[1345]”中的过程类似，分离变量法可以得到式 1 的通解（式 10 [1331]）．首先看定态薛定谔方程定态薛定谔方程

$$-\frac{1}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi_E+V(x)\psi_E=E\psi_E\qquad(E>0) \tag{2}$$

其中 E 可以连续取值，每个 E 都对应不同的解．当 $E<0$ 时，有可能存在束缚态，但我们在散射问题中不考虑．这是一个二阶常微分方程，每个 E 有两个线性无关的解，记为 $\psi_{E,1}(x)$ 和 $\psi_{E,2}(x)$，于是通解可以表示为

$$\Psi(x,t)=\int_0^{+\infty}[C_1(E)\psi_{E,1}(x)+C_2(E)\psi_{E,2}(x)]\mathrm{e}^{-\mathrm{i}E(t-t_0)}\,\mathrm{d}E \tag{3}$$

系数 $C_i(E)$ 由初始波函数决定，满足

$$\Psi(x,t_0)=\int_0^{+\infty}[C_1(E)\psi_{E,1}(x)+C_2(E)\psi_{E,2}(x)]\,\mathrm{d}E \tag{4}$$

注意由于我们规定 $V(\pm\infty)\to 0$，在无穷远处 $\psi_{E,i}(x)$ 显然都是平面波，波数都是 $k=\sqrt{2mE}$. 另外由于式 2 中所有的系数都是实数，$\psi_{E,1},\psi_{E,2}$ 也可以都是实函数.

具体如何从 $\Psi(x,t_0)$ 得到系数 $C_i(E)$ 呢？如果我们确保所有不同的基底 $\psi_{E,i}$ 都是正交归一[20]的（见“傅里叶变换与连续正交归一基底[240]”）

$$\langle\psi_{E',i'}|\psi_{E,i}\rangle=\delta_{i',i}\delta(E'-E) \tag{5}$$

那么就有

$$C_i(E)=\langle\psi_{E,i}|\Psi(t_0)\rangle\qquad(i=1,2) \tag{6}$$

这就是**能量归一化**. 当然我们同样有不同的归一化方式（事实上有无数种），例如类似于式 5 [1345] 的**动量归一化**，令式 2 的本征函数为 $\psi_k(x)$（$k\in\mathbb{R}$，$k^2/(2m)=E$），满足归一化条件

$$\langle\psi_{k'}|\psi_k\rangle=\delta(k'-k) \tag{7}$$

那么波函数表示为

$$\Psi(x,t)=\int_{-\infty}^{+\infty}C(k)\psi_k(x)\mathrm{e}^{-\mathrm{i}E(t-t_0)}\,\mathrm{d}k \tag{8}$$

系数通过初始波函数的投影计算

$$C(k)=\langle\psi_k|\Psi(x,t_0)\rangle \tag{9}$$

归一化的具体过程要比平面波更复杂，详见“动量归一化和能量归一化[1348]”.

一个基本的例子见“方势垒散射数值计算[1369]”.

未完成：未完成，需要包含波包系数计算[1365] 方势垒散射数值计算[1369]

散射幅

未完成：未完成

[20]通常解二阶常微分方程只要求找到两个线性无关的解，但正交归一的条件要更苛刻.

一维散射态的正交归一化

预备知识 一维散射（量子）[1351]

本文使用原子单位制[1315]．[21]类似于平面波的归一化[1348]，一维散射态也有不同的归一化方式，但情况要更为复杂．

为了方便先假设 $V(x)$ 关于原点对称，且 $V(x)$ 只在区间 $[-L,L]$ 内不为零．由于 $V(x)$ 的对称性，我们必定能找到实值的奇函数和偶函数两种解．令 $k=\sqrt{2mE}>0$，在区间 $[-L,L]$ 外，波函数就是正弦函数加上一个相移

$$\psi_k(x)=A\sin(kx+\phi)\qquad(x>L)\tag{1}$$

其中 ϕ 是 k 的函数，称为**相移（phase shift）**．为方便书写下文把 $\phi(k),\phi(k')$ 分别记为 ϕ,ϕ'．

令奇函数和偶函数散射态分别为实函数 $\psi_{k,1}(x)$ 和 $\psi_{k,2}(x)$ 我们希望通过添加适当的归一化系数后，波函数能满足归一化条件（式 12 [1348]）

$$\int_{-\infty}^{+\infty}\psi_{k',i}(x)^*\psi_{k,i}(x)\,\mathrm{d}x=\delta(k'-k)\qquad(k>0,i=1,2)\tag{2}$$

定理 1

式 2 对所有性质良好的 $V(x)$ 都成立，且式 1 中归一化系数和简谐波相同，即 $A=1/\sqrt{\pi}$（式 11 [1348]）．

部分证明

对于奇偶性不同的两个函数，他们显然式正交的．首先已知

$$\int_0^{+\infty}\sin(k'x)\sin(kx)\,\mathrm{d}x=\frac{\pi}{2}\delta(k'-k)\tag{3}$$

现在添加相位 $\phi(k)$ 后，有不定积分

$$\int\sin(k'x+\phi')\sin(kx+\phi)\,\mathrm{d}x=\frac{\sin[(k'-k)x+(\phi'-\phi)]}{2(k'-k)}-\frac{\sin[(k'+k)x+(\phi'+\phi)]}{2(k'+k)}\tag{4}$$

[21] 写作参考这篇帖子.

在 $(0,n)$ 做定积分取极限 $n\to\infty$ 后发现比 $\delta(x)$ 多了两项

$$\int_0^{+\infty}\sin(k'x+\phi')\sin(kx+\phi)\,\mathrm{d}x=\frac{\pi}{2}\delta(k'-k)+\frac{\sin(\phi'+\phi)}{2(k'+k)}-\frac{\sin(\phi'-\phi)}{2(k'-k)}\tag{5}$$

所以在区间 $[0,+\infty)$ 上 $\sin(kx+\phi)$ 并不正交[22].

使用归一化系数 $1/\sqrt{2}$，式 2 的积分为（利用波函数的奇偶性）

$$\begin{aligned}\langle\psi_{k'}|\psi_k\rangle&=2\int_0^{+\infty}\frac{1}{\sqrt{\pi}}\sin(k'x+\phi')\frac{1}{\sqrt{\pi}}\sin(kx+\phi)\,\mathrm{d}x+2I(k,k')\\&=\delta(k'-k)+\frac{\sin(\phi'+\phi)}{\pi(k'+k)}-\frac{\sin(\phi'-\phi)}{\pi(k'-k)}+2I(k,k')\end{aligned}\tag{6}$$

其中 $2I(k,k')$ 修正了 $[-L,L]$ 区间实际波函数和 $\sin(kx+\phi)$ 的不同

$$I(k,k')=\int_0^L\psi_{k'}(x)^*\psi_k(x)\,\mathrm{d}x-\int_0^L\frac{1}{\sqrt{\pi}}\sin(k'x+\phi')\frac{1}{\sqrt{\pi}}\sin(kx+\phi)\,\mathrm{d}x\tag{7}$$

如果能证明对于任意 $V(x)$，式 6 的最后的三项之和都为零，那么我们就证明了式 2 的正交归一关系．读者可以尝试用一些具体的例子证明，如方势垒[1365].

不对称势能

归一化

由于势能 $V(x)$ 不对称，我们无法保证上文的 $\psi_{k,1},\psi_{k,2}$ 的奇偶性．但通过适当的操作仍然能使式 2 成立．先给出结论：令区间 $[-L,L]$ 外的波函数为

$$\psi_{k,i}(x)=\begin{cases}A_{+,i}\sin(kx+\phi_{+,i}) & (x>L)\\ A_{-,i}\sin(kx-\phi_{-,i}) & (x<-L)\end{cases}\qquad(i=1,2)\tag{8}$$

那么，归一化系数需要满足

$$\frac{1}{2}\left(|A_-|^2+|A_+|^2\right)=\frac{1}{\pi}\tag{9}$$

当 $A_+=A_-$ 时就有上文的归一化系数 $1/\sqrt{\pi}$.

[22]但在 $(-\infty,\infty)$ 上却正交

部分证明：把正交化积分划分为正负半轴两部分进行（ψ_k 取 $\psi_{k,1},\psi_{k,2}$ 中一个）

$$\begin{aligned}\langle\psi_{k'}|\psi_k\rangle &= |A_-|^2\int_{-\infty}^{0}\sin(kx+\phi_+)\sin(kx+\phi_+)\,\mathrm{d}x + I_-(k,k')\\ &\quad + |A_+|^2\int_{0}^{+\infty}\sin(kx+\phi_+)\sin(kx+\phi_+)\,\mathrm{d}x + I_+(k,k')\\ &= |A_-|^2\frac{\pi}{2}\delta(k'-k) + |A_+|^2\frac{\pi}{2}\delta(k'-k)\end{aligned}\tag{10}$$

把式 9 代入有 $\langle\psi_{k'}|\psi_k\rangle=\delta(k'-k)$.

正交化

和上文类似，我们可以保证 $i=i'$ 时式 2 成立．但由于我们缺失了函数的奇偶性，一般无法保证 $i\neq i'$ 时也成立．可以证明定态薛定谔方程必定存在一对满足以下正交条件的 $\psi_{k,1},\psi_{k,2}$

$$\lim_{n\to\infty}\frac{1}{n}\int_{-n}^{+n}\psi_{k,1}(x)^*\psi_{k,2}(x)\,\mathrm{d}x=0\tag{11}$$

显然上文的奇函数和偶函数根据该定义是正交的．注意把 $\psi_{k,1},\psi_{k,2}$ 分别乘以不同的相位因子同样满足上式以及式 2.

若我们得到两个不正交的线性无关解 $\psi_{k,1},\psi'_{k,2}$，这时我们可以使用一个类似于施密特正交化[522] 的操作．先把 $\psi_{k,1}$ 按照式 9 归一化，那么

$$\psi_{k,2}(x)=\psi'_{k,2}(x)-\alpha\psi_{k,1}(x)\tag{12}$$

$$\begin{aligned}\alpha &= \lim_{n\to\infty}\frac{\pi}{n}\int_{-n}^{n}\psi_{k,1}(x)^*\psi'_{k,2}(x)\,\mathrm{d}x\\ &= \frac{\pi}{2}[A_{-,1}A_{-,2}\cos(\phi_{-,1}-\phi_{-,2})+A_{+,1}A_{+,2}\cos(\phi_{+,1}-\phi_{+,2})]\end{aligned}\tag{13}$$

这相当于把 $\psi'_{k,2}$ 中与 $\psi_{k,1}$ 不正交的部分减去．最后再对 $\psi_{k,2}$ 归一化使其满足式 9 即可使 $\psi_{k,1},\psi_{k,2}$ 正交归一．

部分证明：假设对于任意给定 k 必定存在满足式 2 和式 11 的一对正交归一解 $\psi_{k,1},\psi_{k,2}$．它们是 2 维解空间的一组基底，那么 $\psi'_{k,2}$ 必定可以表示为

$$\psi'_{k,2}=\alpha\psi_{k,1}+\beta\psi_{k,2}\tag{14}$$

所以

$$\lim_{n\to\infty}\frac{1}{n}\int_{-n}^{+n}\psi_{k,1}(x)^*\psi'_{k,2}(x)\,\mathrm{d}x = \alpha\lim_{n\to\infty}\frac{1}{n}\int_{-n}^{+n}\psi_{k,1}(x)^*\psi_{k,1}(x)\,\mathrm{d}x+\beta\lim_{n\to\infty}\frac{1}{n}\int_{-n}^{+n}\psi_{k,1}(x)^*\psi_{k,2}(x)\,\mathrm{d}x \tag{15}$$

其中由式 9 得（$[-L,L]$ 的积分在极限中消失）

$$\lim_{n\to\infty}\frac{1}{n}\int_{-n}^{+n}\psi_{k,1}(x)^*\psi_{k,1}(x)\,\mathrm{d}x=\frac{1}{2}\left(|A_-|^2+|A_+|^2\right)=\frac{1}{\pi} \tag{16}$$

把式 16 和式 11 代入式 15 得式 13．证毕．

行波的归一化

$$\psi_{k,i}=\begin{cases}A_i\exp(\mathrm{i}kx)+B_i\exp(-\mathrm{i}kx) & (x<-L)\\ C_i\exp(\mathrm{i}kx)+D_i\exp(-\mathrm{i}kx) & (x>L)\end{cases}\quad(i=a,b) \tag{17}$$

和上文类似的方法得归一化条件为

$$|A_i|^2+|B_i|^2+|C_i|^2+|D_i|^2=\frac{1}{\pi} \tag{18}$$

对于平面波，显然有 $A=C=1/\sqrt{2\pi}$，$B=D=0$．和“平面波的的正交归一化[1348]”中结论相同．

一种 $\psi_{k,1}$ 常用的行波边界条件为 $D_a=0$，它的物理意义是粒子从左边入射，发生反射和透射．满足归一化条件（链接未完成）

$$|A_a|^2+|B_a|^2+|C_a|^2=\frac{1}{\pi} \tag{19}$$

又根据概率流[1386] 守恒得 $|A_a|^2=|B_a|^2+|C_a|^2$，得 $|A_a|=1/\sqrt{2\pi}$．为了方便我们一般直接规定

$$A_a=\frac{1}{\sqrt{2\pi}} \tag{20}$$

一般情况下我们通过施密特正交化得到与之正交归一的 ψ_b．我们先令不一定正交的 ψ'_b 满足边界条件 $A_b=0$，即粒子从右边入射，投影得

$$\alpha=\pi(B_a^*B_b+C_a^*C_b) \tag{21}$$

于是（未归一化的）ψ_b 为

$$\psi_b = \psi_b' - \alpha\psi_a \tag{22}$$

但我们下面会看到当 $V(x)$ 对称（偶函数）时，$\psi_{k,2}(x) = \psi_{k,1}(-x)$.

变换矩阵

$$\psi_{k,i}(x) = \begin{cases} A_{+,i}\sin(kx + \phi_{+,i}) & (x > L) \\ A_{-,i}\sin(kx - \phi_{-,i}) & (x < -L) \end{cases} \quad (i = 1, 2) \tag{23}$$

令变换矩阵为 $\begin{pmatrix} C_1 & C_2 \\ C_2^* & -C_1^* \end{pmatrix}$，满足 $|C_1|^2 + |C_2|^2 = 1$，有

$$\psi_{k,a}(x) = C_1\psi_{k,1}(x) + C_2\psi_{k,2}(x) \tag{24}$$

$$\psi_{k,b}(x) = C_2^*\psi_{k,1}(x) - C_1^*\psi_{k,2}(x) \tag{25}$$

令 $\psi_{k,a}$ 满足上述边界条件 $D_a = 0, A_a = 1/\sqrt{2\pi}$ 分别得到

$$\begin{aligned} C_1 A_{+,1}\mathrm{e}^{-\mathrm{i}\phi_{+,1}} + C_2 A_{+,2}\mathrm{e}^{-\mathrm{i}\phi_{+,2}} &= 0 \\ C_1 A_{-,1}\frac{\mathrm{e}^{-\mathrm{i}\phi_{-,1}}}{2\mathrm{i}} + C_2 A_{-,2}\frac{\mathrm{e}^{-\mathrm{i}\phi_{-,2}}}{2\mathrm{i}} &= \frac{1}{\sqrt{2\pi}} \end{aligned} \tag{26}$$

这就可以求出 C_1, C_2，以及 $\psi_{k,a}, \psi_{k,b}$.

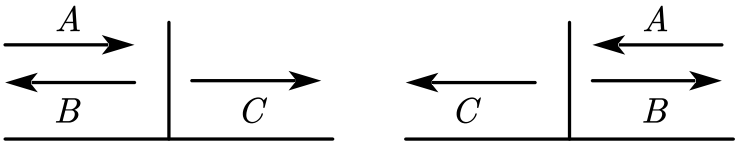


图 1: 对称且正交的行波解．左：$\psi_{k,a}$，右：$\psi_{k,b}$

可以证明[23]如果 $V(x)$ 是对称的（$A_{\pm,i} = 1/\sqrt{2}$，$\phi_{+,i} = \phi_{-,i}$），那么 ψ_b 就和 ψ_a 镜像对称（图 1 ）

$$\psi_{k,b}(x) = \psi_{k,a}(-x) \tag{27}$$

[23]证明过程：对 $\psi_{k,b}$ 写出式 26 的对应条件，C_1, C_2 替换为 $C_2^*, -C_1^*$，这时会发现该条件和式 26 是相同的.

由正交条件式 11 得

$$\mathrm{Re}[B^*C]=0 \Longleftrightarrow \cos(\arg B-\arg C)=0 \tag{28}$$

即 B,C 的幅角总是相差 $\pi/2$.

一维 delta 势能散射

- 本词条处于草稿阶段.

预备知识 一维散射（量子）[1351]

假设原点处有一 δ 势能 $V(x)=a\delta(x)$. 当 $a<0$ 时它是一个势阱，存在离散的束缚态，当 $a>0$ 时它是势垒，值存在连续的散射态.

原点的边界条件为

$$\Delta\psi'=2a\psi \tag{1}$$

非束缚态的本征函数具有二重简并. 如果取对称和反对称作为基底，分别为（$k>0$）

$$f_\pm(k,x)=\frac{1}{\sqrt{\pi}}\begin{cases}\sin(kx+\phi) & (x>0)\\ \pm f_\pm(k,-x) & (x<0)\end{cases}\qquad \phi=\arctan(k/a) \tag{2}$$

$$f_1(k,x)=\frac{1}{\sqrt{\pi}}\sin(kx) \tag{3}$$

满足归一化条件

$$\langle f_i(k')|f_j(k)\rangle=\delta_{i,j}\delta(k-k') \tag{4}$$

我们把这样的解叫做**驻波（standing wave）**.

现在我们试图变换每个二维简并空间中的基底，令

$$\begin{aligned}h_1(k,x)&=C_{11}f_1(k,x)+C_{12}f_0(k,x)\\ h_2(k,x)&=C_{21}f_1(k,x)+C_{22}f_0(k,x)\end{aligned} \tag{5}$$

若要满足同样的归一化条件

$$\langle h_{i'}(k')|h_i(k)\rangle=\delta_{i,i'}\delta(k-k') \tag{6}$$

其中

$$\langle h_{i'}(k')|h_i(k)\rangle = \sum_{j'} C^*_{i',j'} \langle f_{j'}(k')| \sum_j C_{i,j} |f_j(k)\rangle = \sum_j C^*_{i',j} C_{i,j} \delta(k-k') \tag{7}$$

系数需要满足

$$\sum_j C^*_{i',j} C_{i,j} = \delta_{i,i'} \tag{8}$$

即变换矩阵 $\mathbf{C}$ 必须是一个酉矩阵.

我们先来寻找代表从左边入射并向左边反射且向右边出射的基底 $h_1^{(+)}$，令边界条件为

$$h_1^{(+)}(x) \propto \exp(\mathrm{i}kx) \quad (x>0) \tag{9}$$

解得系数分别为[24]

$$C_1 = \frac{-\mathrm{i}\mathrm{e}^{\mathrm{i}\phi}}{\sqrt{2}} \qquad C_2 = \frac{\mathrm{i}}{\sqrt{2}} \tag{10}$$

代入得（仍然要求 $k>0$）

$$h_1^{(+)}(k,x) = \frac{1}{\sqrt{2\pi}} \begin{cases} \mathrm{e}^{\mathrm{i}kx} - \cos\phi \mathrm{e}^{\mathrm{i}\phi}\mathrm{e}^{-\mathrm{i}kx} & (x<0) \\ -\mathrm{i}\sin\phi \mathrm{e}^{\mathrm{i}\phi}\mathrm{e}^{\mathrm{i}kx} & (x>0) \end{cases} \tag{11}$$

要得到 h_2，我们只需要补全矩阵 $\mathbf{C}$ 的第二行，使其与第一行正交，得到[25]

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{pmatrix} -\mathrm{i}\mathrm{e}^{\mathrm{i}\phi} & \mathrm{i} \\ -\mathrm{i}\mathrm{e}^{\mathrm{i}\phi} & -\mathrm{i} \end{pmatrix} \tag{12}$$

代入得（$k>0$）

$$h_2^{(+)}(k,x) = \frac{1}{\sqrt{2\pi}} \begin{cases} -\mathrm{i}\sin\phi \mathrm{e}^{\mathrm{i}\phi}\mathrm{e}^{-\mathrm{i}kx} & (x<0) \\ \mathrm{e}^{-\mathrm{i}kx} - \cos\phi \mathrm{e}^{\mathrm{i}\phi}\mathrm{e}^{\mathrm{i}kx} & (x>0) \end{cases} \tag{13}$$

可以直接用归一化积分验证正交归一化条件式 6 和原点处的边界条件式 4（如果边界条件是线性的，符合边界条件的函数线性叠加仍然符合边界条件）.

可以发现，$h_1(k,x)$ 和 $h_2(k,x)$ 分别是两个**入射波（incoming wave）**，即从所有方向入射，一个方向出射.

[24]注意求解的时候列出的是齐次方程组，所以该式中的系数同乘一个任意相位因子仍然是解，这里选择相位因子使得式 11 中的入射波具有最简洁的形式.

[25]矩阵的第二行同样也可以乘以任意相位因子，我们选择该因子是的式 13 的入射波最简洁.

我们只需把 $h_1(k,x)$ 和 $h_2(k,x)$ 取复共轭，即可得到两个方向的**出射波（outgoing wave）**，即从一个方向入射，各个方向出射．相对于驻波，可以称入射波和出射波为**行波（travelling wave）**．

从入射波到出射波的变换矩阵为

$$\mathbf{T}=\begin{pmatrix}\mathrm{e}^{\mathrm{i}\phi}\cos\phi & -\mathrm{i}\sin\phi\\ -\mathrm{i}\sin\phi & \mathrm{e}^{-\mathrm{i}\phi}\cos\phi\end{pmatrix} \tag{14}$$

在三维问题中，我们同样可以得到驻波，入射波和出射波三组不同的本征波函数．

波包的散射

无论是驻波还是行波都不具有物理意义，因为一个粒子不可能有无限细的能量带宽和无限大的位置分布．我们现在来考虑如果初始时刻我们有一个波包，这个波包会如何移动．

我们可以选取一组基底将波包进行分解，写出各个基底随时间变化的形式，再进行线性组合即可得到波包随时间的变化．

本词条中的例子有一个特殊的地方在于 δ 函数在原点产生的边界条件要求波函数空间中的所有波函数都必须满足（因为所有的基底都必须满足）．所以如果初始波包在原点处不为零，那么初始波包也需要满足原点的边界条件．若是其他有限函数如方势垒和方势阱，则没有这个限制．但一般来说这个限制并不重要，因为我们大部分情况下考虑波包从无穷远处入射，初始时刻原点的波函数为零．

一维阶梯势能

- 本词条处于草稿阶段．

预备知识 定态薛定谔方程[1328]

使用原子单位

$$V(x)=\begin{cases}+\infty & (x\leqslant -l)\\ 0 & (-l\leqslant x<0)\\ V_0 & (0\leqslant x\leqslant l)\\ +\infty & (l\leqslant x)\end{cases} \tag{1}$$

阶梯两侧得波数满足

$$k^2=2V_0+k_1^2 \tag{2}$$

令波函数的一个解为

$$\psi_1(x)=\begin{cases}A_1\exp(\mathrm{i}kx) & (-l\leqslant x<0)\\ C_1\exp(\mathrm{i}kx)+D_1\exp(-\mathrm{i}kx) & (0\leqslant x\leqslant l)\end{cases} \tag{3}$$

在原点匹配函数值和导数，得

$$\begin{cases}A_1=C_1+D_1\\ kA_1=k_1C_1-k_1D_1\end{cases} \tag{4}$$

解得

$$\begin{cases}C_1=\dfrac{k_1+k}{2k_1}A_1\\ D_1=\dfrac{k_1-k}{2k_1}A_1\end{cases} \tag{5}$$

令波函数的一个解为

$$\psi_2(x)=\begin{cases}B_2\exp(-\mathrm{i}kx) & (-l\leqslant x<0)\\ C_2\exp(\mathrm{i}kx)+D_2\exp(-\mathrm{i}kx) & (0\leqslant x\leqslant l)\end{cases} \tag{6}$$

在原点匹配函数值和导数，得

$$\begin{cases}A_1=C_1+D_1\\ -kA_1=k_1C_1-k_1D_1\end{cases} \tag{7}$$

有限深势阱

- 本词条存在未完成的内容.

预备知识 定态薛定谔方程[1328]

[26]本文使用原子单位制[1315]. 一维定态薛定谔方程式 4 [1328] 为

$$-\frac{1}{2m}\frac{\partial^2}{\partial x^2}\psi(x)+V(x)\psi(x)=E\psi(x) \tag{1}$$

令势能函数为

$$V(x)=\begin{cases}-V_0 & (-L/2\leqslant x\leqslant L/2)\\ 0 & (\text{其他})\end{cases} \tag{2}$$

该势能叫做**有限深势阱（finite square well）**.

有限深势阱既包含连续的本征态（散射态），又可能包含有限个离散的束缚态. 所以有限深势阱是研究一维散射问题的一个简单模型.

束缚态

由于 $V(x)$ 是对称的，波函数必定是奇函数或者偶函数（式 8 [1328]）. 令

$$k=\sqrt{2mE}\qquad \kappa=\sqrt{2m(E+V_0)} \tag{3}$$

第 1,3 区间的通解为 $E<V$

$$\psi(x)=C_1\mathrm{e}^{kx}+C_2\mathrm{e}^{-kx} \tag{4}$$

为了让无穷远处波函可归一化，所以

$$\psi_1=A\mathrm{e}^{kx}\qquad \psi_3=D\mathrm{e}^{-kx} \tag{5}$$

第 2 区间的通解为 $E>V$

$$\psi_2(x)=B\cos(\kappa x)+C\sin(\kappa x) \tag{6}$$

[26]参考 Wikipedia 相关页面.

奇波函数

当波函数为奇函数时，易得 $\psi(0)=0$ 即 $B=0$，且 $A=-D$．再考虑 $x=L/2$ 处波函数及一阶导数连续有

$$\begin{aligned} C\sin(\kappa L/2) &= D\exp(-kL/2) \\ \kappa C\cos(\kappa L/2) &= -kD\exp(-kL/2) \end{aligned} \tag{7}$$

其中可以把 E 看成未知量，决定 k,κ，C,D 也是未知量．两式相除得

$$\tan(\kappa L/2) = -\kappa/k \tag{8}$$

未完成：图未完成，式 8 和式 10 的图解

这是一个超定方程，可能存在解．解出后再次代入式 7 可以求得比值 D/C．归一化即可确定 C,D．

偶波函数

当波函数为奇函数时，易得 $\psi'(0)=0$ 即 $C=0$，且 $A=D$．与奇函数的情况同理得

$$\begin{aligned} B\cos(\kappa L/2) &= D\exp(-kL/2) \\ \kappa B\sin(\kappa L/2) &= kD\exp(-kL/2) \end{aligned} \tag{9}$$

相除得

$$\tan(\kappa L/2) = k/\kappa \tag{10}$$

未完成：什么时候有束缚态什么时候没有？

散射态

散射态的计算只需要把“方势垒[1365]”中 $E>V_0$ 的情况下取 $V_0<0$ 即可，这里不再重复．

方势垒的定态波函数

- 未完成：透射率反射率画图

预备知识 无限深势阱[1335]，原子单位制[1315]

本文采用原子单位制[1315]．我们要解一维定态薛定谔方程（参考式 4 [1315]）为

$$-\frac{1}{2m}\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}+V\psi=E\psi \tag{1}$$

令方势垒长度为 $2l$，关于原点对称．势能函数为

$$V(x)=\begin{cases}0 & (x<-l)\\ V_0 & (-l\leqslant x<l)\\ 0 & (l\leqslant x)\end{cases} \tag{2}$$

一些教材把势垒放在区间 $[0,a]$，这时只需令 $a=2l$，把解出的波函数做一个平移即可．以下讨论两种常见的解，一种类似于 $\sin(kx),\cos(kx)$ 另一种类似于 $\exp(\mathrm{i}kx),\exp(-\mathrm{i}kx)$．

奇偶函数解

对称势能的好处是存在奇和偶的实函数解，且它们自动正交．

$E\geq V_0$ 的情况

令

$$k=\sqrt{2mE}\qquad b=\sqrt{2m(E-V_0)} \tag{3}$$

令对称解和反对称的波函数分别为

$$\psi^e_{\mathbf{k}}(x)=\begin{cases}A_1\cos(bx) & (0\leqslant x\leqslant l)\\ C_1\cos(kx)+D_1\sin(kx) & (l<x)\\ \psi^e_{\mathbf{k}}(-x) & (x<0)\end{cases} \tag{4}$$

$$\psi^o_{\mathbf{k}}(x)=\begin{cases}B_2\sin(bx) & (0\leqslant x\leqslant l)\\ C_2\cos(kx)+D_2\sin(kx) & (l<x)\\ -\psi^o_{\mathbf{k}}(-x) & (x<0)\end{cases}\tag{5}$$

其中 $l<x$ 的部分也可以分别记为

$$G_i\sin(kx+\phi_i)\qquad(i=1,2)\tag{6}$$

$$G_i=\sqrt{C_i^2+D_i^2}\qquad\phi_i=\operatorname{Arctan}(C_i,D_i)\qquad(i=1,2)\tag{7}$$

在 $x=l$ 处匹配波函数和一阶导数，解得

$$\begin{cases}\dfrac{C_1}{A_1}=\cos(kl)\cos(bl)+\dfrac{b}{k}\sin(kl)\sin(bl)\\ \dfrac{D_1}{A_1}=-\dfrac{b}{k}\cos(kl)\sin(bl)+\sin(kl)\cos(bl)\end{cases}\tag{8}$$

$$\begin{cases}\dfrac{C_2}{B_2}=\cos(kl)\sin(bl)-\dfrac{b}{k}\sin(kl)\cos(bl)\\ \dfrac{D_2}{B_2}=\dfrac{b}{k}\cos(kl)\cos(bl)+\sin(kl)\sin(bl)\end{cases}\tag{9}$$

对波函数归一化，使 $\int\psi'_k(x)\psi_k(x)=\delta(k'-k)$，使无穷远处振幅为 $1/\sqrt{\pi}$，得

$$\begin{aligned}A_1&=\frac{1}{\sqrt{\pi}\sqrt{\cos^2(bl)+(b^2/k^2)\sin^2(bl)}}\\ B_2&=\frac{1}{\sqrt{\pi}\sqrt{\sin^2(bl)+(b^2/k^2)\cos^2(bl)}}\end{aligned}\tag{10}$$

$0<E<V_0$ 的情况

令 $\kappa=\sqrt{2m(V_0-E)}$

$$\psi^e_{\mathbf{k}}(x)=\begin{cases}A_1\cosh(\kappa x) & (0\leqslant x\leqslant l)\\ C_1\cos(kx)+D_1\sin(kx) & (l<x)\end{cases}\tag{11}$$

$$\psi^o_{\mathbf{k}}(x)=\begin{cases}A_2\sinh(\kappa x) & (0\leqslant x\leqslant l)\\ C_2\cos(kx)+D_2\sin(kx) & (l<x)\end{cases}\tag{12}$$

在 $x=l$ 处匹配波函数和一阶导数，分别得

$$\begin{cases} A_1\cosh(\kappa l)=C_1\cos(kl)+D_1\sin(kl) \\ \kappa A_1\sinh(\kappa l)=-kC_1\sin(kl)+kD_1\cos(kl) \end{cases} \tag{13}$$

$$\begin{cases} A_2\sinh(\kappa l)=C_2\cos(kl)+D_2\sin(kl) \\ \kappa A_2\cosh(\kappa l)=-kC_2\sin(kl)+kD_2\cos(kl) \end{cases} \tag{14}$$

解得

$$\begin{cases} \dfrac{C_1}{A_1}=\cosh(\kappa l)\cos(kl)-\dfrac{\kappa}{k}\sinh(\kappa l)\sin(kl) \\ \dfrac{D_1}{A_1}=\cosh(\kappa l)\sin(kl)+\dfrac{\kappa}{k}\sinh(\kappa l)\cos(kl) \end{cases} \tag{15}$$

$$\begin{cases} \dfrac{C_2}{A_2}=\sinh(\kappa l)\cos(kl)-\dfrac{\kappa}{k}\cosh(\kappa l)\sin(kl) \\ \dfrac{D_2}{A_2}=\sinh(\kappa l)\sin(kl)+\dfrac{\kappa}{k}\cosh(\kappa l)\cos(kl) \end{cases} \tag{16}$$

对波函数归一化，使 $\int\psi_k'(x)\psi_k(x)=\delta(k'-k)$，得

$$A_1=\frac{1}{\sqrt{\pi}\sqrt{\cosh^2(\kappa l)+(\kappa/k)^2\sinh^2(kl)}} \tag{17}$$

$$A_2=\frac{1}{\sqrt{\pi}\sqrt{\sinh^2(\kappa l)+(\kappa/k)^2\cosh^2(kl)}} \tag{18}$$

行波解

令左、中、右三个区间 exp 项的系数分别为 A,B、C,D、E,F．左边右入射和出射，右边只有出射没有反射

$$\psi_{k,1}(x)=\begin{cases} A\exp(\mathrm{i}kx)+B\exp(-\mathrm{i}kx) & (x<-L) \\ C\exp(\mathrm{i}bx)+D\exp(-\mathrm{i}bx) & (-L\le x\le L) \\ F\exp(-\mathrm{i}kx) & (x<-L) \end{cases} \tag{19}$$

$$\psi_{k,2}(x)=\psi_{k,1}(-x) \tag{20}$$

要归一化只需满足 $A=1/\sqrt{2\pi}$ （式 20 [1354]）.

$E > V_0$ 的情况

当 $E > V_0$ 时，系数解为

$$\begin{cases} A = 1/\sqrt{2\pi} \\ B = \mathcal{N} \cdot (k^2 - b^2)\sin(2bl)\exp(-\mathrm{i}kl) \\ C = \mathcal{N} \cdot \mathrm{i}k(k+b)\exp(-\mathrm{i}bl) \\ D = \mathcal{N} \cdot \mathrm{i}k(b-k)\exp(\mathrm{i}bl) \\ E = \mathcal{N} \cdot 2\mathrm{i}kb\exp(-\mathrm{i}kl) \end{cases} \tag{21}$$

其中

$$\mathcal{N} = \frac{1}{\sqrt{2\pi}} \frac{(k^2+b^2)\sin(2bl) - 2\mathrm{i}kb\cos(2bl)}{(k^2-b^2)^2\sin^2(2bl) + 4k^2b^2} \exp(-\mathrm{i}kl) \tag{22}$$

可以验证 $\psi_{k,1}, \psi_{k,2}$ 正交的条件（式 28 [1354]）$\mathrm{Re}[B^*E] = 0$.

透射率，反射率分别为

$$T = \left|\frac{E}{A}\right|^2 = \frac{4k^2b^2}{(k^2-b^2)^2\sin^2(2bl) + 4k^2b^2} \tag{23}$$

$$R = \left|\frac{B}{A}\right|^2 = \frac{(k^2-b^2)\sin^2(2bl)}{(k^2-b^2)^2\sin^2(2bl) + 4k^2b^2} \tag{24}$$

$0 < E < V_0$ 的情况

当 $0 < E < V_0$ 时，系数解为

$$\begin{cases} A = 1/\sqrt{2\pi} \\ B = \mathcal{N} \cdot (\kappa^2 + k^2)\sinh(2\kappa l)\exp(-\mathrm{i}kl) \\ C = \mathcal{N} \cdot k(\mathrm{i}\kappa - k)\exp(-\kappa l) \\ D = \mathcal{N} \cdot k(\mathrm{i}\kappa + k)\exp(\kappa l) \\ E = \mathcal{N} \cdot 2\mathrm{i}k\kappa\exp(-\mathrm{i}kl) \end{cases} \tag{25}$$

$$\mathcal{N} = \frac{1}{\sqrt{2\pi}} \frac{(k^2-\kappa^2)\sinh(2\kappa l) - 2\mathrm{i}k\kappa\cosh(2\kappa l)}{(k^2+\kappa^2)^2\sinh^2(2\kappa l) + 4k^2\kappa^2} \exp(-\mathrm{i}kl) \tag{26}$$

$$T = \left|\frac{E}{A}\right|^2 = \frac{4k^2\kappa^2}{(k^2+\kappa^2)^2\sinh^2(2\kappa l) + 4k^2\kappa^2} \tag{27}$$

$$R = \left|\frac{B}{A}\right|^2 = \frac{(\kappa^2 + k^2)\sinh^2(2\kappa l)}{(k^2+\kappa^2)^2\sinh^2(2\kappa l) + 4k^2\kappa^2} \tag{28}$$

投影系数

任意平方可积的波函数可以表示为

$$\psi(x) = \int_0^{+\infty} C_1(k)\psi_{k,1}(x) + C_2(k)\psi_{k,2}(x)\,\mathrm{d}k \tag{29}$$

若 $\psi(x)$ 只在 $x < -L$ 不为零，则有

$$\begin{aligned} C_1(k) &= \langle\psi_{k,1}|\psi\rangle = \tilde{\psi}(k) + \sqrt{2\pi}B^*\tilde{\psi}(-k) \\ C_2(k) &= \langle\psi_{k,2}|\psi\rangle = \sqrt{2\pi}E^*\tilde{\psi}(-k) \end{aligned} \tag{30}$$

其中 $\tilde{\psi}(k)$ 是 $\psi(x)$ 的傅里叶变换

$$\tilde{\psi}(k) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\psi(x)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x \tag{31}$$

含时的具体例子见“方势垒散射数值计算[1369]”.

方势垒散射数值计算

- 本词条处于草稿阶段.
- 未完成：需要改进程序，计算波函数随时间的变化.（给系数加上 $\exp(-\mathrm{i}Et)$ 即可）

预备知识 方势垒的定态波函数[1365]

代码 1: FSBsct.m

```
% 使用类 sin, cos 基底
clear; close all;

% === params ====
k0 = 4; a = 1/6;
```

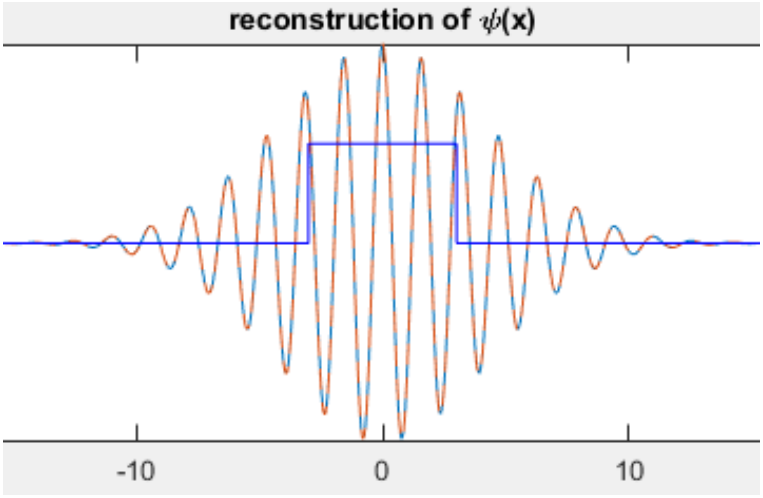


图 1: 运行结果

```
Nk = 300; kmax = k0 + 12*a; kmin = max(kmax/Nk, k0 - 12*a);
x0 = 0;
Nx = 2000; xmin = x0-6/a; xmax = x0+6/a;
psi = @(x)exp(1i*k0*x).*exp(-(a*(x-x0)).^2);
L = 3; V0 = 1; m = 1;
% ==============

% plot FFT
k = linspace(kmin, kmax, Nk); dk = k(2)-k(1);
[g1, k1] = FFT(fftresize(psi(x),Nx*2), dx);
g = interp1(k1, g1, k);
figure; plot(k, abs(g), '.');
axis([kmin, kmax, 0, max(abs(g))*1.1]);
xlabel k; title 'FFT of \psi(x)';

A = zeros(1,Nk); B = A;
E = k.^2/(2*m);
psi1 = zeros(size(x));
for i = 1:Nk
    A(i) = integral(@(x)psi(x).*FSB(x,E(i),L,m,V0,0), ...
            xmin, xmax, 'RelTol', 1e-16);
    B(i) = integral(@(x)psi(x).*FSB(x,E(i),L,m,V0,1), ...
            xmin, xmax, 'RelTol', 1e-16);
```

```
    psi1 = psi1 + dk*A(i)*FSB(x,E(i),L,m,V0,0) ...
                + dk*B(i)*FSB(x,E(i),L,m,V0,1);
end
figure; plot(x, real(psi(x)));
hold on; plot(x, real(psi1), '--');
plot([xmin, -L,-L, L, L, xmax], [0, 0, 0.5, 0.5, 0, 0], 'b');
axis([xmin,xmax,-1,1]);
title 'reconstruction of \psi(x)'
```

```
% 使用类 exp(ikx), exp(-ikx) 基底
clear; close all;

% === params ====
k0 = 4; a = 1/6;
Nk = 300; kmax = k0 + 12*a; kmin = max(kmax/Nk, k0 - 12*a);
x0 = 0;
Nx = 2000; xmin = x0-6/a; xmax = x0+6/a;
psi = @(x)exp(1i*k0*x).*exp(-(a*(x-x0)).^2);
L = 3; V0 = 1; m = 1;
% ==============

% plot FFT
k = linspace(kmin, kmax, Nk); dk = k(2)-k(1);
[g1, k1] = FFT(fftresize(psi(x),Nx*2), dx);
g = interp1(k1, g1, k);
figure; plot(k, abs(g), '.');
axis([kmin, kmax, 0, max(abs(g))*1.1]);
xlabel k; title 'FFT of \psi(x)';

C1 = zeros(1,Nk); C2 = C1;
E = k.^2/(2*m);
psi1 = zeros(size(x));
for i = 1:Nk
    C1(i) = integral(@(x)psi(x).*conj(FSB2(x,E(i),L,m,V0,1)),
        ...
            xmin, xmax, 'RelTol', 1e-16);
```

```
    C2(i) = integral(@(x)psi(x).*conj(FSB2(x,E(i),L,m,V0,-1)),
        ...
             xmin, xmax, 'RelTol', 1e-16);
    psi1 = psi1 + dk*C1(i)*FSB2(x,E(i),L,m,V0,1) ...
                + dk*C2(i)*FSB2(x,E(i),L,m,V0,-1);
end
figure; plot(x, real(psi(x)));
hold on; plot(x, real(psi1), '--');
plot([xmin, -L,-L, L, L, xmax], [0, 0, 0.5, 0.5, 0, 0], 'b');
axis([xmin,xmax,-1,1]);
title 'reconstruction of \psi(x)'
```

动量表象的薛定谔方程

预备知识 薛定谔方程[1331]

本文使用原子单位制[1315]．一维薛定谔方程为

$$-\frac{1}{2m}\frac{\partial^2}{\partial x^2}\psi(x,t)+V(x,t)\psi(x,t)=\mathrm{i}\frac{\partial}{\partial t}\psi(x,t) \tag{1}$$

先来看一维的情况：可以证明，如果 $V(x,t)$ 关于 x 可以被泰勒展开，那么动量表象的薛定谔方程为

$$-\frac{k^2}{2m}\varphi(k,t)+V\left(\mathrm{i}\frac{\partial}{\partial k}\right)\varphi(k,t)=\mathrm{i}\frac{\partial}{\partial t}\varphi(k,t) \tag{2}$$

证明：把

$$\psi(x,t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\varphi(k,t)\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x \tag{3}$$

代入式 1，再利用式 19 [237] 即可．证毕．

从另一种方式来理解，表象无关的薛定谔方程为

$$\hat{H}\left|\psi\right\rangle=\mathrm{i}\frac{\partial}{\partial t}\left|\psi\right\rangle \tag{4}$$

$$\hat{H}=\frac{\hat{p}^2}{2m}+V(\hat{x}) \tag{5}$$

而动量表象中 $\hat{p}=p$，$\hat{x}=\mathrm{i}\,\partial/\partial k$，代入即可．

未完成：例子

类似地，动量表象得三维薛定谔方程为

$$-\frac{\mathbf{k}^2}{2m}\varphi(\mathbf{k},t)+V(\mathrm{i}\boldsymbol{\nabla}_k)\varphi(\mathbf{k},t)=\mathrm{i}\frac{\partial}{\partial t}\varphi(\mathbf{k},t) \tag{6}$$

拉比频率

- 本词条处于草稿阶段.

[27]双态系统中

式 13 [1414]

$$A_{mn}=\langle\phi_m(t)|\,H'(t)\,|\phi_n(t)\rangle=\langle\psi_m|\,H'(t)\,|\psi_n\rangle\,\mathrm{e}^{\mathrm{i}(E_m-E_n)t/\hbar} \tag{1}$$

$$\mathbf{Ac}=\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{c} \tag{2}$$

若微扰哈密顿矩阵的对角元消失（例如束缚态有 $\langle\psi_n|\mathbf{r}|\psi_n\rangle=0$），有 $\omega_0=E_2-E_1$

$$\begin{cases}\dot{c}_1=\dfrac{1}{\mathrm{i}\hbar}H'_{12}\mathrm{e}^{-\mathrm{i}\omega_0 t}c_2\\ \dot{c}_2=\dfrac{1}{\mathrm{i}\hbar}H'_{21}\mathrm{e}^{\mathrm{i}\omega_0 t}c_1\end{cases} \tag{3}$$

拉比频率（Rabi Frequency）

若矩阵 $\mathbf{H}'$ 是一个常数矩阵乘以一个 $\cos(\omega t)$，当 $\omega\to\omega_0$ 时有拉比频率为

$$\omega_r=\frac{1}{2}\sqrt{(\omega-\omega_0)^2+(|V_0|/\hbar)^2} \tag{4}$$

解

$$c_2(t)=\frac{2A_{21}^{\dagger}}{\mathrm{i}\hbar\omega_r}\exp[-\mathrm{i}(\omega-\omega_0)t/2]\sin(\omega_r t/2) \tag{5}$$

$$|c_1(t)|^2+|c_2(t)|^2=1 \tag{6}$$

$$P_{21}(t)=|c_2(t)|^2=\frac{4|A_{12}|^2}{\hbar^2\omega_r^2}\sin^2(\omega_r t/2) \tag{7}$$

[27]本文参考 Griffiths 和 Bransden

二维无限深势阱

预备知识 无限深势阱[1335]，分离变量法

[28]现在我们来看一个粒子的二维运动．定态薛定谔方程为

$$H\Psi(x,y)=E\Psi(x,y) \tag{1}$$

其中哈密顿算符为

$$H=-\frac{1}{2}\nabla^2+V(x,y)=-\frac{1}{2}\frac{\partial^2}{\partial x^2}-\frac{1}{2}\frac{\partial^2}{\partial y^2}+V(x,y) \tag{2}$$

前两项分别为 x 和 y 方向的动能算符，$V(x,y)$ 为势能（算符）

$$V(x,y)=\begin{cases}0 & (0\leqslant x\leqslant a,\ 0\leqslant y\leqslant b)\\ +\infty & (\text{其他情况})\end{cases} \tag{3}$$

$V(x,y)$ 可以看成 $V_x(x)$ 和 $V_y(y)$ 两个函数相加，他们的定义为

$$V_x(x)=\begin{cases}0 & (0\leqslant x\leqslant a)\\ +\infty & (\text{其他情况})\end{cases} \tag{4}$$

$$V_y(y)=\begin{cases}0 & (0\leqslant y\leqslant b)\\ +\infty & (\text{其他情况})\end{cases} \tag{5}$$

于是总哈密顿算符可以记为两部分，每部分是一个一维简谐振子的哈密顿算符

$$H=H_x+H_y \tag{6}$$

$$\begin{cases}H_x=-\dfrac{1}{2}\dfrac{\partial^2}{\partial x^2}+V_x(x)\\ H_y=-\dfrac{1}{2}\dfrac{\partial^2}{\partial y^2}+V_y(y)\end{cases} \tag{7}$$

[28]本词条使用原子单位

为什么这样做呢？这使我们可以使用分离变量法．令

$$\Psi(x,y)=\Psi_x(x)\Psi_y(y) \tag{8}$$

代入薛定谔方程两边同除以 $\Psi_x(x)\Psi_y(y)$ 得

$$\frac{H_x\Psi_x(x)}{\Psi_x(x)}+\frac{H_y\Psi_y(y)}{\Psi_y(y)}=E \tag{9}$$

由于第一项只和 x 有关，第二项只和 y 有关，所以它们都是常数，令

$$\begin{cases}H_x\Psi_x(x)=E_x\Psi_x(x)\\ H_y\Psi_y(y)=E_y\Psi_y(y)\end{cases} \tag{10}$$

他们分别是一维简谐振子的定态薛定谔方程.他们的各个束缚态波函数（式 4 [1335]）分别为

$$\Psi_{x,i}(x)=\sqrt{\frac{2}{a}}\sin\left(\frac{n\pi}{a}x\right)\quad(i=1,2,...) \tag{11}$$

$$\Psi_{y,j}(y)=\sqrt{\frac{2}{b}}\sin\left(\frac{n\pi}{b}y\right)\quad(j=1,2,...) \tag{12}$$

对应的能量分别为

$$E_{x,i}=\frac{\pi^2}{2a^2}i^2\qquad E_{y,j}=\frac{\pi^2}{2b^2}j^2 \tag{13}$$

则总哈密顿算符的束缚态可以记为它们的乘积

$$\Psi_{i,j}(x,y)=\Psi_{x,i}(x)\Psi_{y,j}(y)=\frac{2}{\sqrt{ab}}\sin\left(\frac{i\pi}{a}x\right)\sin\left(\frac{j\pi}{b}x\right) \tag{14}$$

对应的能量为

$$E_{i,j}=E_{x,i}+E_{y,j}=\frac{\pi^2}{2}\left(\frac{i^2}{a^2}+\frac{j^2}{b^2}\right) \tag{15}$$

另外，由于 $\Psi_{x,i}(x)$ 和 $\Psi_{y,j}(y)$ 都是完备的，它们的乘积也是完备的，即二维势阱中任意波函数可以表示为

$$\Psi(x,y)=\sum_{i,j}C_{i,j}\Psi_{x,i}(x)\Psi_{y,j}(y) \tag{16}$$

事实上，这就是二维傅里叶级数.

平均值（量子力学）

预备知识 测量理论[1435]

我们先来回顾测量理论．假设某个物理量 Q 有离散的本征态 $|\phi_i\rangle$（$i = 1, 2\ldots$），对应的本征值为 q_i，满足

$$Q\,|\phi_i\rangle = q_i\,|\phi_i\rangle \tag{1}$$

且满足正交归一条件

$$\langle\phi_i|\phi_j\rangle = \delta_{i,j} \tag{2}$$

记粒子处于 $|\psi\rangle$ 状态，可表示为本征态的线性组合

$$|\psi\rangle = \sum_i c_i\,|\phi_i\rangle \tag{3}$$

对其测量 Q，得到第 q_i 的概率为

$$P_i = |c_i|^2 = |\langle\phi_i|\psi\rangle|^2 \tag{4}$$

现在我们可以定义 Q 的平均值为

$$\langle Q\rangle = \sum_i P_i q_i = |c_i|^2 q_i \tag{5}$$

这意味着，如果我们取大量处于 $|\psi\rangle$ 状态的系统，分别测量 Q 再取平均，结果就是该式．

平均值还有一个更常见的公式，与式 5 等效．

$$\langle Q\rangle = \langle\psi|Q|\psi\rangle \tag{6}$$

要验证，可以将式 3 代入该式，得

$$\begin{aligned}\langle Q\rangle &= \left(\sum_i c_i^*\,\langle\phi_i|\right) Q \left(\sum_j c_j\,|\phi_j\rangle\right) \\ &= \sum_{i,j} c_i^* c_j\,\langle\phi_i|\,Q\,|\phi_j\rangle\end{aligned} \tag{7}$$

再代入式 1 和式 2 得

$$\langle Q\rangle=\sum_{i,j}c_i^*c_jq_i\left\langle\phi_i|\phi_j\right\rangle=\sum_{i,j}c_i^*c_jq_i\delta_{i,j}=\sum_i|c_i|^2q_i \tag{8}$$

证毕.

守恒量（量子力学）

预备知识 平均值[1376]，无限深势阱[1335]

在经典力学中，某个时刻测量系统的某个量 Q 可以得到唯一确定的值. 如果这个值不随时间变化，就说它**守恒**. 而在量子力学中，这个值（算符的本征值）是不确定的，那守恒量由该如何理解呢？比如两个无限深势阱中的粒子具有相同的波函数，测得的能量却可能各不相同，是否意味着能量不守恒？

答案是否定的，在量子力学中，我们定义若对系统的任意状态测量某个物理量得到任意本征值的概率都不随时间改变，这个量就是守恒的. 例如在无限深势阱中任意随时间变化的波包，测得每个能量本征值的概率都不随时间变化，这就说明能量守恒. 由这个定义可知，守恒量的平均值也不随时间变化.

能量守恒

定理 1　能量守恒

若哈密顿算符 H 不随时间变化，就有能量守恒.

证明：令能量的所有本征态为（为了方便起见，假设本征态是离散的）$\{|u_i\rangle\}$，系统状态随时间的演化可以记为（C_i 为常数，由初始状态决定）

$$|v(t)\rangle=\sum_iC_i\mathrm{e}^{-\mathrm{i}E_it}|u_i\rangle \tag{1}$$

令能量 E_j 对应的所有本征态的角标的集合为 D_j，那么测得 E_j 的概率为

$$P(E_j)=\sum_{i\in D_j}|C_i|^2 \tag{2}$$

不随时间变化，所以有能量守恒. 证毕.

例 1

能量守恒的简单系统如：无限深势阱[1335]，简谐振子[1340]，自由粒子，有限深势阱[1351]，阶梯势能[1361] 等．这些系统的哈密顿算符都不随时间变化．将任意时刻的波函数投影到能量的本征态上，得到的系数的模方都是相同的．

其他守恒量

预备知识 算符对易与共同本征矢[1384]

定理 2　对易与守恒量

令 H 为哈密顿算符，Q 是另一个物理量的算符．以下两个命题互为充分必要条件[116]

1. Q 对应一个守恒量
2. $[H,Q]=0$

该定理在经典力学中的对应定理是：若物理量 $\omega(q,p)$ 和哈密顿量 H 的泊松括号[1078] $\{\omega,H\}$ 为零，则 ω 是一个守恒量．

证明：我们还是从离散本征态的简单情况来证明这个定理，且假设哈密顿算符不含时[29]．首先由 2 证明 1，若两算符对称，则它们必有一组完备的共同本征矢 $\{|u_i\rangle\}$，满足

$$\begin{cases} H|u_i\rangle = E_i|u_i\rangle \\ Q|u_i\rangle = q_i|u_i\rangle \end{cases} \tag{3}$$

系统状态随时间的变化可以记为

$$|v(t)\rangle = \sum_i C_i \mathrm{e}^{-\mathrm{i}E_i t}|u_i\rangle \tag{4}$$

其中 C_i 为常数．令集合 D_j 包含 q_j 对应的所有基底的角标．于是，任意时刻测到 q_j 的概率等于相关系数的模方和

$$P(q_j) = \sum_{i\in D_j} \left|C_i \mathrm{e}^{-\mathrm{i}E_i t}\right|^2 = \sum_{i\in D_j} |C_i|^2 \tag{5}$$

也是常数．证毕．

由 1 证明 2 要更麻烦一些．我们新开一节，不感兴趣的读者可以跳过．

[29]如果含时，可以将时间分割成许多小区间，假设每个小区间中不随时间变化

证明守恒量算符与哈密顿算符对易

先假设能量的本征态基底为 $\{|u_i\rangle\}$，系统状态随时间变化为

$$|v(t)\rangle = \sum_i C_i \mathrm{e}^{-\mathrm{i}E_i t} |u_i\rangle \tag{6}$$

令算符 Q 任意本征值 q_m 对应的本征矢子空间的一组基底为 $\{|v_k\rangle\}$，那么任意时刻测得 q_m 的概率为

$$\begin{aligned} P(q_m) &= \sum_k \left| \langle v_k | \sum_i C_i \mathrm{e}^{-\mathrm{i}E_i t} |u_i\rangle \right|^2 \\ &= \sum_k \left(\sum_i C_i \mathrm{e}^{\mathrm{i}E_i t} \langle v_k | u_i \rangle \right)^\dagger \left(\sum_j C_j \mathrm{e}^{-\mathrm{i}E_j t} \langle v_k | u_j \rangle \right) \\ &= \sum_{i,j} \left(C_i^* C_j \sum_k \langle u_i | v_k \rangle \langle v_k | u_j \rangle \right) \mathrm{e}^{\mathrm{i}(E_i - E_j)t} \end{aligned} \tag{7}$$

若 Q 守恒，则要求对于任何 C_i 可能的取值，该式的结果都与时间无关．所以对任意 $i \neq j$ 都有

$$\sum_k \langle u_i | v_k \rangle \langle v_k | u_j \rangle = 0 \tag{8}$$

稍加思考可以得出要么 $|u_j\rangle$ 在 q_m 子空间中，要么 $|u_j\rangle$ 在 q_m 子空间上的投影为 0．所以任何 q_m 子空间（假设是 N_m 维）同样可以由 N_m 个 $|u_i\rangle$ 张成．所以可以找到 H 和 Q 的一套共同本征矢，所以两算符对易．证毕．

好量子数

- 举例子

预备知识 量子力学的基本假设[1435]

我们知道量子力学中，可测量的物理量对应厄米算符 Q，本征方程为

$$Q |\psi_i\rangle = q_i |\psi_i\rangle \tag{1}$$

定义 1

如果厄米算符 Q 和哈密顿算符 H 对易，即 $[H,Q]=0$，那么 Q 的本征值就被称为好量子数（**good quantum number**）.

$[H,Q]=0$ 说明 Q 是一个守恒量[1377]，也说明两个算符具有一组共同的正交归一的完备本征矢[1381].

薛定谔方程的分离变量法

预备知识 张量积空间[513]

如果定态薛定谔方程

$$H\left|\Psi\right\rangle = E\left|\Psi\right\rangle \tag{1}$$

中的状态 $\left|\Psi\right\rangle$ 可以看作是两个小空间的张量积空间[30]，且总哈密顿 H 可以分解为

$$H = H_1\otimes I + I\otimes H_2 \tag{2}$$

的形式. 那么我们先分别解出两个子空间的定态薛定谔方程

$$\begin{cases} H_1\left|\Psi_{1,i}\right\rangle = E_{1,i}\left|\Psi_{1,i}\right\rangle \\ H_2\left|\Psi_{2,j}\right\rangle = E_{2,j}\left|\Psi_{2,j}\right\rangle \end{cases} \tag{3}$$

这两组解分别构成了两个小空间的一组完备的正交归一基底. 我们也可以证明 $\left|\Psi_{1,i}\right\rangle\otimes\left|\Psi_{2,j}\right\rangle$ 满足张量积空间中的薛定谔方程（式 1）. 代入得

$$\begin{aligned} (H_1\otimes I + I\otimes H_2)\left|\Psi_{1,i}\right\rangle\otimes\left|\Psi_{2,j}\right\rangle &= (H_1\left|\Psi_{1,i}\right\rangle)\otimes\left|\Psi_{2,j}\right\rangle + \left|\Psi_{1,i}\right\rangle\otimes(H_2\left|\Psi_{2,j}\right\rangle) \\ &= (E_{1,i}+E_{2,j})\left|\Psi_{1,i}\right\rangle\otimes\left|\Psi_{2,j}\right\rangle \end{aligned} \tag{4}$$

证毕.

事实上，这就是偏微分方程的分离变量法. 两个小空间既可以是同一个粒子在两个不同方向上的状态空间，例如二维无限深势阱[1374]），也可以是两个不同粒子的状态空间例如双粒子无限深势阱.

[30] 也可以是 $N>2$ 个小空间的张量积空间，以下结论类比可得

对易厄米算符与共同本征矢

预备知识 厄米矩阵[422]，本征矢的简并

对易与交换子

本文讨论 $N=2,\dots$ 维矢量空间 X 上的任意两个厄米算符 $A:X\to X$ 和 $A:X\to X$．当我们在 X 中确定一组正交归一基底后，它们可以分别表示为 $N\times N$ 的厄米矩阵 $\mathbf{A}$ 和 $\mathbf{B}$.

我们知道一般来说两个矩阵（线性算符）的乘法不满足交换律 $AB=BA$，即**不对易（non-commutative）**，但也存在满足交换律的特殊算符对，即**对易的（commutative）**.

定义算符 A,B 的**对易子（commutator）**为

$$[A,B]=AB-BA \tag{1}$$

那么对易可以表示为 $[A,B]=0$，反之 $[A,B]\neq 0$.

定理 1

有限维矢量空间 V 中两个厄米算符 A 和 B，以下两个命题互为充分必要条件

1. A,B 对易（$[A,B]=0$）.
2. V 中存在一组正交归一基底，同时是 A,B 的本征矢.

我们可以分为简并和非简并的情况讨论．当 A 和 B 的本征问题都不存在简并，那么这组共同正交归一基底是唯一的．若简并，A 的每个本征子空间内，都能找到 B 本征矢构成的正交归一基底.

未完成：本征子空间需要链接

证明条件 $2 \to 1$

设算符 A 和 B 有一组共同的本征矢 v_i，则它们同时满足 A 和 B 的本征方程

$$\begin{cases} Av_i = a_i v_i \\ Bv_i = b_i v_i \end{cases} \tag{2}$$

对任何 v_i，都有

$$A(Bv_i) = A(b_i v_i) = b_i A v_i = a_i b_i v_i \tag{3}$$

$$B(Av_i) = B(a_i v_i) = a_i B v_i = a_i b_i v_i \tag{4}$$

所以 $ABv_i = BAv_i$ 即

$$[A, B] = AB - BA = 0 \tag{5}$$

即两算符对易．证毕．

证明条件 $1 \to 2$

要证明 $1 \to 2$，只需证明 A 的一套本征矢都满足 B 的本征方程即可．

算符 A 非简并情况（B 是否简并没关系）

先解出算符 A 的本征方程 $Av_i = a_i v_i$，如果 A 算符不发生简并（见本征矢的简并）那么本征值各不相同，且给定一个本征值 a_i 其解只可能是 v_i 或者 v_i 乘以一个任意复常数（注释：其实也可以再相乘一个算符 A 不涉及的物理量的函数，例如总能量算符 H 的本征矢还可以再成一个时间因子 $\mathrm{e}^{\mathrm{i}\omega t}$ ）．

因为算符对易，有

$$A(Bv_i) = B(Av_i) = a_i(Bv_i) \tag{6}$$

把式中的 Bv_i 看成一个新的波函数，上式说明 Bv_i 是算符 A 和本征值 a_i 的另一个本征矢．根据以上分析，Bv_i 必定是 v_i 乘以某个复常数（命名为 b_i ），即

$$Bv_i = b_i v_i \tag{7}$$

而这正是 B 的本征方程（而 B 也是厄米矩阵，所以作为本征值 b_i 的数域从复数缩小到实数）．证毕．

算符 A 简并情况

假设算符 A 的所有本征值为 a_i （各不相同），任意一个 a_i 有 n_i 重简并. 若 $n_i = 1$，对应唯一一个 v_i，那么根据上文对非简并情况的推理，v_i 就已经是 B 的本征矢了. 若 $n_i > 1$，存在一个 n_i 维希尔伯特子空间，里面任何一个函数都是 a_i 对应的本征矢，所以要在子空间中寻找共同本征矢，只需在子空间中寻找 B 的本征矢即可. 令 ϕ_i 为本征值为 a_i 的子空间中的任意函数，利用对易关系

$$A(B\phi_i) = B(A\phi_i) = a_i(B\phi_i) \tag{8}$$

这条式子说明 $B\phi_i$ 是 A 和 a_i 的一个本征矢，即 $B\phi_i$ 仍然在 a_i 的简并子空间中. 所以 B 对子空间来说是一个闭合的厄米算符，所以必有 N 个线性无关的本征矢. 证毕.

以下的内容应该归到厄米算符里面讲（厄米算符在希尔伯特空间中是无穷维的矩阵，但是如果一个厄米算符在一个子空间中闭合，那么就可以通过以下方法找到 N 个线性无关的本征矢. 先在空间中任意选取 n_i 个线性无关的正交本征矢 $v_{i1}, v_{i2} \dots v_{in_i}$ 作为子空间的基底（本征矢的简并）），并可以用基底 $v_{i1}, v_{i2} \dots v_{in_i}$ 展开.

令 $Bv_{ij} = \sum_{k=1}^{n_i} W_{jk} v_{ik}$ （$W_{jk} = \langle v_{ij}| B |v_{ik}\rangle$，可以是复数），则 B 在该子空间可以表示成一个 n_i 维的方形矩阵（记为 W ）.

以 $v_{i1}, v_{i2} \dots v_{in_i}$ 为子空间的基底，子空间内任意函数 $\phi = x_1 v_{i1} + x_2 v_{i2} \dots$ 可以记为 $|\phi\rangle = (x_1, x_2, \dots, x_{n_i})^{\mathrm{T}}$. 根据算符的矩阵表示[1388]，$B$ 在子空间的矩阵元就是系数 W_{jk}，

$$W = \begin{pmatrix} W_{11} & \dots & W_{1n_i} \\ \vdots & \ddots & \vdots \\ W_{n_i 1} & \dots & W_{n_i n_i} \end{pmatrix} \tag{9}$$

所以 B 在子空间范围内的本征方程的矩阵形式就是

$$W |\phi_k\rangle = b_{ik} |\phi_k\rangle \tag{10}$$

所以 B 在子空间的本征值就是 W 的本征值，本征矢就是 W 的本征矢对应的波函数.

最后要证明的就是 W 矩阵必然存在 n_i 个本征矢. 由于 B 是厄米算符，W 必然是厄米矩阵，而 n_i 维的厄米矩阵必然存在 n_i 个两两正交的复数本征矢和

实数本征值（厄米接矩阵）.

综上所述，对每一个 n_i 重简并的 a_i，都存在 n_i 个两两正交的本征矢作为 A，B 算符的共同本征矢．证毕.

矩阵对易与共同本征矢

- 重复词条，待整合（删除）

由于矩阵乘法满足结合律（式 18 [396]），算符对易意味着，对于任意一个列矢量 $\mathbf{v}$（看成单列矩阵），先将 $\mathbf{A}$ 对其作用再将 $\mathbf{B}$ 对其作用，等于以相反的顺序作用．即

$$\mathbf{B}(\mathbf{A}\mathbf{v}) = \mathbf{A}(\mathbf{B}\mathbf{v}) \tag{1}$$

为什么要讨论两个矩阵是否对易？因为对易与矩阵的本征问题紧密关联

证明

首先用条件 2 证明条件 1．记这组共同本征矢为 $\mathbf{v}_i$ $(i = 1, \ldots, N)$，且令本征值为 a_i, b_i，即

$$\begin{cases} \mathbf{A}\mathbf{v}_i = a_i \mathbf{v}_i \\ \mathbf{B}\mathbf{v}_i = b_i \mathbf{v}_i \end{cases} \tag{2}$$

那么对所有 $i = 1, \ldots, N$ 都有

$$\mathbf{B}(\mathbf{A}\mathbf{v}_i) = \mathbf{B}(a_i \mathbf{v}_i) = a_i \mathbf{B}\mathbf{v}_i = a_i b_i \mathbf{v}_i \tag{3}$$

$$\mathbf{A}(\mathbf{B}\mathbf{v}_i) = \mathbf{A}(b_i \mathbf{v}_i) = b_i \mathbf{A}\mathbf{v}_i = a_i b_i \mathbf{v}_i \tag{4}$$

可见式 1 对每个基底都成立，所以对空间中的任意矢量也成立（因为任意矢量可以表示成基底的线性组合，而算符都是线性的），所以 $\mathbf{AB} = \mathbf{BA}$.

再来用条件 1 证明条件 2．这要更复杂一些．我们解 $\mathbf{A}$ 的本征方程

$$\mathbf{A}\mathbf{v}_i = a_i \mathbf{v}_i \tag{5}$$

由于 $\mathbf{A}$ 是厄米矩阵，我们必定可以得到一组（N 个）两两正交的本征矢量 $\mathbf{v}_i$ 和本征值 a_i．由于两个算符对易，有

$$\mathbf{A}(\mathbf{B}\mathbf{v}_i) = \mathbf{B}(\mathbf{A}\mathbf{v}_i) = \mathbf{B}(a_i \mathbf{v}_i) = a_i(\mathbf{B}\mathbf{v}_i) \tag{6}$$

观察等式两端，这说明 $\mathbf{Bv}_i$ 矢量同样也是本征值 a_i 对应的本征矢.

现在分两种情况讨论. 第一种 $\mathbf{A}$ 的本征矢不存在简并（注意我们并不在乎 $\mathbf{B}$ 是否有简并）. 这样每个 a_i 对应一个一维子空间，空间中的所有矢量都共线，所以 $\mathbf{Bv}_i$ 和 $\mathbf{v}_i$ 共线，即后者乘以常数等于前者. 令该常数为 b_i，得

$$\mathbf{Bv}_i = b_i\mathbf{v}_i \tag{7}$$

这就证明了 $\mathbf{v}_i$ 同样是 $\mathbf{B}$ 的本征矢.

更复杂的是如果 $\mathbf{A}$ 的某个本征值 a_i 存在简并，那它的所有本征矢可以组成 $n_i > 1$ 维的子空间. 如果 $\mathbf{v}_i$ 属于这个子空间，那么式 5 和式 6 未必说明 $\mathbf{Bv}_i$ 与 $\mathbf{v}_i$ 共线，而是只能说明 $\mathbf{Bv}_i$ 也在这个子空间中. 或者说 $\mathbf{B}$ 在该子空间中具有**封闭性**. 可以证明在该子空间中，矩阵 $\mathbf{B}$ 必定有 n_i 个正交归一的本征矢，它们既是 $\mathbf{A}$ 的本征矢（因为该子空间中的任何矢量都是），也是 $\mathbf{B}$ 的本征矢.

类氢原子的约化质量

预备知识 玻尔原子模型

玻尔氢原子模型假设原子核的质量远大于电子质量. 得出的结论是

$$E_n = -\frac{mZ^2e^4}{8\epsilon_0^2h^2}\frac{1}{n^2} \approx -13.6\ \mathrm{eV}\frac{Z^2}{n^2} \tag{1}$$

当考虑原子核运动的时候（但仍然忽略万有引力），上式变为

$$E_n = -\frac{\mu Z^2e^4}{8\epsilon_0^2h^2}\frac{1}{n^2} \tag{2}$$

其中 $\mu = m_1m_2/(m_1+m_2)$ 叫做**约化质量**. m_1 和 m_2 分别是原子核质量和核外电子质量.　　要注意这时的 E_n 包括了原子核的动能.

推导

1. 经典力学的条件

$$m_1\frac{v_1^2}{r_1} = m_2\frac{v_2^2}{r_2} = \frac{1}{4\pi\epsilon_0}\frac{Ze^2}{(r_1+r_2)^2} \tag{3}$$

2. 角动量量子化条件（注意角动量是总角动量!）

$$m_1v_1r_1 + m_2v_2r_2 = n\hbar \tag{4}$$

3. 质心系条件

$$m_1r_1 = m_2r_2 \qquad m_1v_1 = m_2v_2 \tag{5}$$

要求的能量为

$$E_n = \frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2 - \frac{1}{4\pi\epsilon_0}\frac{Ze^2}{r_1 + r_2} \tag{6}$$

联立这几条式子，就可以解出考虑原子核运动的能级公式. 推荐的步骤是，把所有的 r_2 和 v_2 写成 m_1r_1/m_2 和 m_1v_1/m_2 . 然后用和玻尔原子模型中一样的方法解出来即可.

概率流密度

预备知识 薛定谔方程[1331]

一维情况下，对于某个波函数 $\psi(x,t)$，定义概率流为

$$j(x,t) = \frac{\mathrm{i}\hbar}{2m}\left(\psi\frac{\partial\psi^*}{\partial x} - \psi^*\frac{\partial\psi}{\partial x}\right) = \frac{\hbar}{m}\operatorname{Im}\left[\psi^*\frac{\partial}{\partial x}\psi\right] \tag{1}$$

某个区间中的概率增加率等于流入该区间的概率流

$$\frac{\mathrm{d}}{\mathrm{d}t}P_{ab}(t) = j(a,t) - j(b,t) \tag{2}$$

三维情况下，概率流的定义变为

$$\mathbf{j}(\mathbf{r},t) = \frac{\mathrm{i}\hbar}{2m}(\psi\boldsymbol{\nabla}\psi^* - \psi^*\boldsymbol{\nabla}\psi) = \frac{\hbar}{m}\operatorname{Im}[\psi^*\boldsymbol{\nabla}\psi] \tag{3}$$

且有

$$\frac{\mathrm{d}}{\mathrm{d}t}P_{\mathcal{V}}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{V}}|\psi(\mathbf{r},t)|^2\,\mathrm{d}V = \int_{\mathcal{S}}\mathbf{j}(\mathbf{r},t)\cdot\mathrm{d}\mathbf{s} \tag{4}$$

或写为概率守恒公式（类比电荷守恒[1212]）

$$\frac{\mathrm{d}}{\mathrm{d}t}(\psi^*\psi) + \boldsymbol{\nabla}\cdot\mathbf{j} = \mathbf{0} \tag{5}$$

习题 1　平面波

求三维平面波 $A\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r})$ 的概率流密度. 答案: $|A|^2\hbar\mathbf{k}/m$, 注意这恰好等于概率密度乘以粒子速度（原因见下文）.

习题 2　球面波

求球面波 $A\exp(\mathrm{i}kr)/r$ 的概率流密度（$r=|\mathbf{r}|$）. 答案: $|A|^2\hbar\mathbf{k}/(mr^2)$, 同样等于概率密度乘以粒子速度, 注意单位时间通过任意球面的概率都是一样的.

推导

对一维情况有

$$\frac{\mathrm{d}}{\mathrm{d}t}P_{ab}=\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b\psi^*\psi\,\mathrm{d}x=\int_a^b\left(\psi\frac{\partial}{\partial t}\psi^*+\psi^*\frac{\partial}{\partial t}\psi\right)\mathrm{d}x \tag{6}$$

一维薛定谔方程以及复共轭为

$$\mathrm{i}\hbar\frac{\partial\psi}{\partial t}=-\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2}+V\psi \tag{7}$$

$$-\mathrm{i}\hbar\frac{\partial\psi^*}{\partial t}=-\frac{\hbar^2}{2m}\frac{\partial^2\psi^*}{\partial x^2}+V\psi^* \tag{8}$$

代入上式的时间微分，得

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}P_{ab}&=\frac{\mathrm{i}\hbar}{2m}\int_a^b\left(\psi^*\frac{\partial^2\psi}{\partial x^2}-\psi\frac{\partial^2\psi^*}{\partial x^2}\right)\mathrm{d}x=\frac{\mathrm{i}\hbar}{2m}\int_a^b\frac{\partial}{\partial x}\left(\psi^*\frac{\partial\psi}{\partial x}-\psi\frac{\partial\psi^*}{\partial x}\right)\mathrm{d}x\\&=\frac{\mathrm{i}\hbar}{2m}\left(\psi^*\frac{\partial\psi}{\partial x}-\psi\frac{\partial\psi^*}{\partial x}\right)\Bigg|_{x=a}^{x=b}=j(a)-j(b)\end{aligned} \tag{9}$$

三维情况的证明可类比.

概率的速度

类比经典力学或电磁学中的 $\mathbf{j}=\rho\mathbf{v}$，若定义概率流速度为概率流除以概率密度，则平面波 $\psi(x)=A\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}$ 的概率流速为

$$\mathbf{v}=\mathbf{j}/|\psi|^2=\frac{\mathrm{i}\hbar}{2m}\big(-|A|^2\mathrm{i}\mathbf{k}-|A|^2\mathrm{i}\mathbf{k}\big)/|A|^2=\frac{\hbar\mathbf{k}}{m}=\frac{\mathbf{p}}{m}=\mathbf{v}_{CM} \tag{10}$$

所以平面波的概率流速度等于具有相同动量的经典粒子的速度.

算符的矩阵表示

- 本词条缺少预备知识，初学者可能会遇到困难.
- 这是事实上是线性代数的内容，并没有涉及到量子力学

考虑一个比较基本的问题，算符的“功能”是什么呢？算符就是对函数的一种操作方法．给出一个波函数，经过算符作用，可以得到一个新的波函数．以下给出量子力学中算符的两个重要性质

1. 算符都是线性的，即对任意 n 个波函数 $\psi_1, \psi_2 \dots \psi_n$，算符 $\hat{Q}$ 满足

$$\hat{Q}(c_1\psi_1 + c_2\psi_2 \dots c_n\psi_n) = c_1\hat{Q}\psi_1 + c_2\hat{Q}\psi_2 \dots c_n\hat{Q}\psi_n \tag{1}$$

2. 算符的本征方程的本征值都是实数．因为根据测量理论，本征值就是可能出现的测量结果，所以本征值一定是实数.

我们已经知道，波函数可以用列矢量表示．既然算符都是线性的，而矩阵可以表示列矢量的线性变换，是否可以用矩阵代替算符，从而作用于列向量呢？根据性质 1，若 ψ 是算符是算符 $\hat{Q}$ 的本征函数，$\lambda_1 \dots \lambda_n$ 是对应的本征值（实数），则

$$\begin{cases} \hat{Q}\psi &= \hat{Q}(c_1\psi_1 + \cdots + c_n\psi_n) \\ &= c_1\hat{Q}\psi_1 + \cdots + c_n\hat{Q}\psi_n \\ &= \lambda_1 c_1\psi_1 + \cdots + \lambda_n c_n\psi_n \end{cases} \tag{2}$$

若把上面的波函数表示成列矢量，就相当于在算符 $\hat{Q}$ 的作用下任意一个列矢量 $|\psi\rangle = (c_1, \dots, c_n)^{\mathrm{T}}$ 总是会变成 $(\lambda_1 c_1, \dots, \lambda_n c_n)^{\mathrm{T}}$．这个变换可以用矩阵

$$\mathbf{Q} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \tag{3}$$

来表示，即

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} \lambda_1 c_1 \\ \vdots \\ \lambda_n c_n \end{pmatrix} \tag{4}$$

所以矩阵 $\mathbf{Q}$ 就是算符 $\hat{Q}$ 的矩阵形式，把算符作用在波函数上得到新的波函数，等效于把算符对应的矩阵作用在波函数对应的列矢量上，得到新的波函数对应的列矢量.

用矩阵和列向量表示的本征方程如下

$$\mathbf{Q}\,|\psi\rangle = \lambda\,|\psi\rangle \tag{5}$$

解得 $\lambda = \lambda_i$ 时，$|\psi\rangle = |\psi_1\rangle = (0,\ldots,1,\ldots,0)^{\mathrm{T}}$（只有第 i 个分量等于 1，其余分量等于 0），而 $|\psi_i\rangle$ 正是波函数 ψ_i 对应的列向量.

在任意基底中的矩阵

上面的讨论中用矩阵 $\mathbf{Q}$ 表示算符 $\hat{Q}$，其局限性在于，只能使用 $\hat{Q}$ 的本征函数 $\psi_1 \ldots \psi_n$ 作为基底. 现在若用其他基底（正交归一的）$\phi_1 \ldots \phi_n$，能否求出算符 $\hat{Q}$ 对应的矩阵 $\mathbf{Q}_1$ 呢？

下面讨论中，为了避免混淆，用 $|f\rangle_\phi$ 表示波函数 f 以 $\phi_1 \ldots \phi_n$ 为基底的列矢量，$|f\rangle_\psi$ 表示波函数 f 以 $\psi_1 \ldots \psi_n$ 为基底的列矢量.

现取任意一波函数 f，$|f\rangle_\psi = (c_1 \ldots c_n)^{\mathrm{T}}$，$|f\rangle_\phi = (d_1 \ldots d_n)^{\mathrm{T}}$. 虽然它们表示同一个波函数 f，但是由于选取的基底不同，列向量也不同. 下面讨论它们之间的变换关系.

若把 f 按 ϕ_i 展开，有

$$\begin{aligned}\int \phi_i^* f\,\mathrm{d}x &= \int \phi_i^*(d_1\phi_1 + \ldots + d_n\phi_n)\,\mathrm{d}x = \sum_{j=1}^{n} d_j \int \phi_i^*\phi_j\,\mathrm{d}x \\ &= \sum_{j=1}^{n} d_j\delta_{ij} = d_i\end{aligned} \tag{6}$$

若把 f 按 ψ_i 展开，有

$$d_i = \int \phi_i^* f\,\mathrm{d}x = \int \phi_i^*(c_1\psi_1 + \cdots + c_n\psi_n)\,\mathrm{d}x = \sum_{j=1}^{n} c_j \int \phi_i^*\psi_j\,\mathrm{d}x \tag{7}$$

上式用矩阵和列矢量表示，即 $(d_1 \ldots d_n)^{\mathrm{T}} = \mathbf{P}(c_1 \ldots c_n)^{\mathrm{T}}$，即 $|f\rangle_\phi = P|f\rangle_\psi$. 其中 P 矩阵的矩阵元 $P_{ij} = \int \phi_i^*\psi_j dx$. $\mathbf{P}$ 叫做基底变换矩阵（或表象变换矩阵）.

若令 $\hat{Q}f = g$，根据前面的内容，$Q\,|f\rangle_\psi = |g\rangle_\psi$ 其中 $Q = \mathrm{diag}(\lambda_1 \ldots \lambda_n)$.

下面应用基底变换矩阵，有 $|f\rangle_\psi = P^{-1}|f\rangle_\phi$；$|g\rangle_\psi = P^{-1}|g\rangle_\phi$. 代入上式得

$$QP^{-1}|f\rangle_\phi = P^{-1}|g\rangle_\phi \tag{8}$$

两边左乘 P 得

$$PQP^{-1}|f\rangle_\phi = |g\rangle_\phi \tag{9}$$

令 $\mathbf{Q}_1 = \mathbf{PQP}^{-1}$，得

$$Q_1|f\rangle_\phi = |g\rangle_\phi \tag{10}$$

所以 $\mathbf{Q}_1$ 就是要求的矩阵.

下面证明 $\mathbf{Q}_1$ 是厄米矩阵.

我们先学习所谓幺正矩阵. 这里给出幺正矩阵的一种定义：若把矩阵 $\mathbf{P}$ 的每一列划分成一个列向量，从左到右分别为 $|p_1\rangle \ldots |p_n\rangle$，若满足 $\langle p_i|p_j\rangle = \delta_{ij}$ 则矩阵 $\mathbf{P}$ 叫做幺正矩阵.

容易证明式 $*$ 中的 P 就是幺正矩阵（证明略）.

性质 1：幺正矩阵一个很重要的性质就是其厄米共轭等于其逆矩阵，$\mathbf{P}^* = \mathbf{P}^{-1}$

证明：要证明 $\mathbf{P}^* = \mathbf{P}^{-1}$，只需证明 $\mathbf{P}^*\mathbf{P}$ 是单位矩阵即可. 根据矩阵乘法的定义，

$$(P^*P)_{ij} = \sum_{k=1}^{n}(P^*)_{ik}P_{kj} \tag{11}$$

根据厄米共轭的定义，

$$\sum_{k=1}^{n}(P^*)_{ik}P_{kj} = \sum_{k=1}^{n}(P_{ki})^*P_{kj} = \langle p_i|p_j\rangle\,\delta_{ij} \tag{12}$$

所以 $\mathbf{P}^*\mathbf{P}$ 是 n 阶的单位矩阵. 证毕.

在上文中，$\mathbf{Q}$ 是所谓的实数元的对角矩阵，所以 $\mathbf{Q}^* = \mathbf{Q}$

另外容易证明，$(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*$. 所以

$$\mathbf{Q}_1^* = (\mathbf{PQP}^{-1})^* = (\mathbf{P}^{-1})^*(\mathbf{PQ})^* = \mathbf{P}(\mathbf{Q}^*\mathbf{P}^*) = \mathbf{PQP}^{-1} = \mathbf{Q}_1 \tag{13}$$

所以 $\mathbf{Q}_1$ 是厄米矩阵.

角动量（量子）

- 本词条处于草稿阶段.
- 本词条缺少预备知识，初学者可能会遇到困难.

预备知识 角动量[952]，对易厄米算符与共同本征矢[1381]，升降算符[1339]

思路：根据力学量（测量量）的经典表达式，可以写出对应的算符（这其实是量子力学的一个重要假设，课本往往将其忽略，而直接告诉你可以这样做）

从经典公式到算符

经典力学中单个粒子的角动量公式是

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} \tag{1}$$

其中 $\mathbf{r}$ 是参考点到物体的位矢，$\mathbf{p}$ 是粒子的动量. 或者写成直角坐标系中的分量形式（令原点为参考点）

$$L_x = yp_z - zp_y \qquad L_y = zp_x - xp_z \qquad L_z = xp_y - yp_x \tag{2}$$

现在我们用式 2 定义三个方向的角动量算符，此时 x, y, z, p_x, p_y, p_z 也应理解为算符. 同样地，若用 $\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}$ 表示位置矢量算符，用 $\mathbf{p} = p_x\hat{\mathbf{x}} + p_y\hat{\mathbf{y}} + p_z\hat{\mathbf{z}}$ 表示动量矢量算符，那角动量矢量算符可以用式 1 定义. 我们还可以定义角动量平方（标量）算符（也可以记为非粗体的 L^2）

$$\mathbf{L}^2 = \mathbf{L} \cdot \mathbf{L} = L_x^2 + L_y^2 + L_z^2 \tag{3}$$

除了 x, y, z 三个方向的角动量分量，我们可以将任意方向的角动量分量 $\hat{\mathbf{n}} \cdot \mathbf{L}$ 表示为算符且都与 $\mathbf{L}^2$ 算符对易.

$$L_n = n_x L_x + n_y L_y + n_z L_z \qquad (n_x^2 + n_y^2 + n_z^2 = 1) \tag{4}$$

角动量算符 L, L_z 通常在球坐标中表示，详见“球坐标系中的角动量算符[1393]”.

对易关系

对于角动量分量，理想的状况是，如果能解出本征方程

$$\mathbf{L}\psi = \mathbf{l}\psi \tag{5}$$

我们就能得到矢量本征值 $\mathbf{l}$，然后测量 $\mathbf{L}$ 本征态的结果就一定是 $\mathbf{l}$. 但事实上，$\mathbf{L}$ 几乎从来不单独使用，因为上式无解. 为什么？要解上式，充分必要条件就是要存在 ψ，使三个分量同时有解

$$L_x\psi = l_x\psi \qquad L_y\psi = l_y\psi \qquad L_z\psi = l_z\psi \tag{6}$$

不幸的是，L_x，L_y，L_z 中任意两个都不对易，所以没有共同的本征函数（见“算符对易和共同本征矢函数[1381]”）. 可以证明三个算符之间的对易关系为

$$[L_x, L_y] = \mathrm{i}\hbar L_z \qquad [L_y, L_z] = \mathrm{i}\hbar L_x \qquad [L_z, L_x] = \mathrm{i}\hbar L_y \tag{7}$$

事实上，三个分量中我们只能同时知道一个（海森堡不确定原理）. 通常情况下，我们选择解 L_z 的本征方程 $L_z\psi = l_z\psi$.

比较幸运的是，L^2 和 L_x, L_y, L_z 都对易（或者任意 L_n），所以必然存在一套本征函数，同时是 L_x, L_y, L_z 其中一个和 L^2 的本征函数. 我们习惯上约定计算 L^2 和 L_z 的共同本征矢.

升降算符和本征值

如果要解 L^2 和 L_z 的共同本征函数，通常的方法是先把算符的表达式转换到球坐标中再解本征方程[742]. 但是我们现在先用一种更简单的（但非常重要的）方法，升降算符（在简谐振子问题中已经见过），来绕过本征函数直接求出共同波函数的简并情况以及对两个算符的本征值.

可以证明 L_z 的升降算符分别为

$$L_\pm = L_x \pm \mathrm{i}L_y \tag{8}$$

根据升降算符[1339] 中的一种定义，要证明它们是升降算符，只要证明 $[L_z, L_\pm] \propto L_\pm$ 即可. 结论是（证明见常见算符对易表）

$$[L_z, L_\pm] = \pm\hbar L_\pm \tag{9}$$

类似简谐振子的升降算符，我们还需要一个归一化系数使 $L_\pm |l,m\rangle = A_\pm |l,m\pm1\rangle$ 成立（见“轨道角动量升降算符归一化[1394]”）．结论是

$$L_\pm |l,m\rangle = \hbar\sqrt{l(l+1)-m(m\pm1)}\,|l,m\pm1\rangle \tag{10}$$

由于 $|l,m\rangle$ 也是

（未完成）

表 1: 轨道角动量符号（剩下的按字母表顺序排序）

符号	S	P	D	F	G	H	I	K
L	0	1	2	3	4	5	6	7
符号	L	M	N	O	Q	R	T	U
L	8	9	10	11	12	13	14	15

球坐标系中的角动量算符

预备知识 角动量（量子）[1391]

本文使用原子单位制[1315]．在量子力学中，我们一般把角动量算符放在球坐标中表示．把轨道角动量算符在直角坐标系中的定义（式 2 [1391]）通过链式法则[270] 用球坐标表示（留作习题）．

$$L_x = \mathrm{i}\left(\sin\phi\frac{\partial}{\partial\theta} + \cot\theta\cos\phi\frac{\partial}{\partial\phi}\right) \tag{1}$$

$$L_y = \mathrm{i}\left(-\cos\phi\frac{\partial}{\partial\theta} + \cot\theta\sin\phi\frac{\partial}{\partial\phi}\right) \tag{2}$$

$$L_z = -\mathrm{i}\frac{\partial}{\partial\phi} \tag{3}$$

$$L^2 = L_x^2 + L_y^2 + L_z^2 = -\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right) - \frac{1}{\sin^2\theta}\frac{\partial^2 u}{\partial\phi^2} \tag{4}$$

注意 L^2 恰好是球坐标系中拉普拉斯算符的角向部分（式 5 [742]）∇_Ω^2 的负．

$$L^2 = -\nabla_\Omega^2 \tag{5}$$

这并不奇怪，经典力学中球坐的哈密顿量可以记为（式 9 [1075]）

$$H=\frac{p_r^2}{2m}+\frac{L^2}{2mr^2}+V \tag{6}$$

其中 $p_r=m\dot{r}$，$L=mr^2\dot{\theta}$. 而量子力学的哈密顿算符在球坐标中可以用式 5 [742] 分解为

$$H=-\frac{1}{2m}\nabla^2+V=\frac{-\nabla_r^2}{2m}+\frac{-\nabla_\Omega^2}{2mr^2}+V \tag{7}$$

这让我们很容易猜出 $p_r^2=-\nabla_r^2$ 和式 5.

类比动量算符 $\mathbf{p}=-\mathrm{i}\boldsymbol{\nabla}$，我们可以定义 $\boldsymbol{\nabla}_\Omega$ 满足

$$\mathbf{L}=-\mathrm{i}\boldsymbol{\nabla}_\Omega \tag{8}$$

于是 ∇_Ω^2 可以看作是两个 $\boldsymbol{\nabla}_\Omega$ 相乘而得.

角动量算符的本征函数

预备知识 球谐函数[761]

我们已经知道 L^2, L_z 对易且具有共同本征矢 $|l,m\rangle$，现在我们在球坐标中求解它的波函数. 来看本征方程

$$L_z|l,m\rangle=m|l,m\rangle \tag{9}$$

$$L^2|l,m\rangle=l(l+1)|l,m\rangle \tag{10}$$

它的解就是球谐函数 $Y_{l,m}(\theta,\phi)$. 但本征波函数应该是三维的，所以任意波函数 $R(r)Y_{l,m}(\theta,\phi)$ 都是 L^2 和 L_z 的共同本征波函数.

轨道角动量升降算符归一化

预备知识 轨道角动量[1391]

首先要提醒，一般来说，算符满足的一个条件是 $\langle g|\hat{Q}f\rangle=\langle\hat{Q}^*g|f\rangle$. 但是对于厄米算符，$\hat{Q}^*=\hat{Q}$，所以有 $\langle g|\hat{Q}f\rangle=\langle\hat{Q}g|f\rangle$.

对于角动量升算符

$$L_+L_- = (L_x + \mathrm{i}L_y)(L_x - \mathrm{i}L_y) = L_x^2 + L_y^2 - \mathrm{i}[L_x, L_y] = L^2 - L_z^2 + \hbar L_z \tag{1}$$

所以

$$\begin{aligned} L_+L_-\psi_{l,m} &= \hbar^2 l(l+1)\psi_{l,m} - \hbar^2 m^2\psi_{l,m} + m\hbar^2\psi_{l,m} \\ &= \hbar^2[l(l+1) - m(m-1)]\psi_{l,m} \end{aligned} \tag{2}$$

所以

$$\langle L_-\psi_{l,m}|L_-\psi_{l,m}\rangle = \langle\psi_{l,m}|L_+L_-\psi_{l,m}\rangle = \hbar^2[l(l+1) - m(m-1)] \tag{3}$$

所以

$$L_-\psi_{l,m} = \hbar\sqrt{l(l+1) - m(m-1)}\psi_{l,m-1} \tag{4}$$

同理可证

$$L_+\psi_{l,m} = \hbar\sqrt{l(l+1) - m(m+1)}\psi_{l,m+1} \tag{5}$$

严格来说，归一化系数后面加上任意相位因子 $\mathrm{e}^{\mathrm{i}\theta}$ 后仍能满足式 3，但一般省略.

自旋角动量

- 本词条处于草稿阶段.

预备知识 轨道角动量[1391]，张量积空间[513]

未完成：简单介绍斯特恩-格拉赫实验

自旋是量子力学中的基本粒子特有的性质，描述粒子的波函数不包含自旋的信息，自旋处于单独的有限维希尔伯特空间中，和波函数的空间做张量积以后用于描述粒子的状态.

1. 自旋角动量三个分量算符 S_x, S_y, S_z 的互相对易关系以及自旋模长平方算符 S^2 的对易关系

2. 与轨道角动量同理，存在一组本征态 $|s,m\rangle$

   （$s=0,1/2,1,3/2\ldots$，$m=-s,-s+1\ldots,s-1,s$ 但是每种粒子都有固有的 s ）满足

$$S^2|s,m\rangle=\hbar^2 s(s+1)|s,m\rangle \quad \text{和} \quad S_z|s,m\rangle=\hbar m|s,m\rangle \tag{1}$$

3. 存在升降算符 $S_\pm=S_x\pm \mathrm{i}S_y$，且（根号项是归一化系数）

$$S_\pm|s,m\rangle=\hbar\sqrt{s(s+1)-m(m\pm1)}\,|s,m+1\rangle \tag{2}$$

4. 对于 $s=1/2$ 的粒子，一共有 2 个本征态，分别是 $|1/2,1/2\rangle$，$|1/2,-1/2\rangle$. 它们的角动量模长平方都是 $3\hbar^2/4$，角动量 z 分量都是 $\hbar/2$. 以这两个本征态为基底，令第一个为 $\chi_+=(1,0)^{\mathrm{T}}$，第二个为 $\chi_-=(0,1)^{\mathrm{T}}$. 可以得出角动量平方算符的矩阵为

$$\mathbf{S}^2=\frac{3\hbar^2}{4}\begin{pmatrix}1&0\\0&1\end{pmatrix} \qquad \mathbf{S}_z=\frac{\hbar}{2}\begin{pmatrix}1&0\\0&-1\end{pmatrix} \tag{3}$$

   根据 $S_+\chi_-=\hbar\chi_+$ 和 $S_-\chi_+=\hbar\chi_-$，得到

$$S_x=\frac{\hbar}{2}\begin{pmatrix}0&1\\1&0\end{pmatrix} \qquad S_y=\frac{\hbar}{2}\begin{pmatrix}0&-\mathrm{i}\\\mathrm{i}&0\end{pmatrix} \tag{4}$$

   然后，定义泡利矩阵.

$$\sigma_x=\begin{pmatrix}0&1\\1&0\end{pmatrix} \qquad \sigma_y=\begin{pmatrix}0&-\mathrm{i}\\\mathrm{i}&0\end{pmatrix} \qquad \sigma_z=\begin{pmatrix}1&0\\0&-1\end{pmatrix} \tag{5}$$

   其实，根据对易关系直接就可以得到泡利矩阵.

自旋角动量矩阵

预备知识 自旋，简谐振子升降算符归一化[1343]

我们已知 $1/2$ 自旋的矩阵可以用 $\hbar/2$ 乘以泡利矩阵得到．以下我们试图计算任意 $n/2$ 自旋粒子的三个自旋分量算符 S_x，S_y 和 S_z 对应的矩阵．这些矩阵一般使用 S_z 的本征态 $|s,m\rangle$ 作为基底（另外两组基底同理可得）．注意当 s 为整数时，本文的结论同样适用于轨道角动量（把算符 S 和量子数 s 分别替换为 L 和 l 即可）．

基本思路是先求出升降算符 $S_\pm = S_x \pm \mathrm{i}S_y$ 的矩阵．再把它们分别相加和相减得到 S_x 和 S_y 的矩阵．S_z 算符在其本征态基底下是本征值组成的对角矩阵．

已知归一化系数为

$$S_\pm |s,m\rangle = \hbar\sqrt{s(s+1)-m(m\pm1)}\,|s,m\pm1\rangle \tag{1}$$

所以

$$\langle s,m|S_\pm|s,m'\rangle = \delta_{m,m'\pm1}\hbar\sqrt{s(s+1)-mm'} \tag{2}$$

可见 $\mathbf{S}_+$ 只有下方的子对角线不为零，$\mathbf{S}_-$ 只有上方的子对角线不为零．

最后得三个矩阵为

$$\langle s,m|S_x|s,m'\rangle = \frac{1}{2}(\delta_{m,m'+1}+\delta_{m,m'-1})\hbar\sqrt{s(s+1)-mm'} \tag{3}$$

$$\langle s,m|S_y|s,m'\rangle = \frac{1}{2\mathrm{i}}(\delta_{m,m'+1}-\delta_{m,m'-1})\hbar\sqrt{s(s+1)-mm'} \tag{4}$$

$$\langle s,m|S_z|s,m'\rangle = \delta_{m,m'}m\hbar \tag{5}$$

可以验证当 $s=1/2, m=\pm1/2$ 时，我们就得到了三个泡利矩阵．

第二章

量子力学 2

算符的指数函数 波函数传播子

预备知识 一阶线性常微分方程组[333]

不含时，有限维的情况

我们可以证明对任意厄米矩阵 $\mathbf{H}$，$\mathbf{U}(t)=\exp(-\mathrm{i}\mathbf{H}t)$ 都是酉矩阵，成为**传播子（propagator）**．要证明一个矩阵是酉矩阵，只需要证明

$$\mathbf{U}^\dagger\mathbf{U}=\mathbf{I} \tag{1}$$

由 $\exp(-\mathrm{i}\mathbf{H}t)$ 的级数定义以及厄米算符的性质可得 $\exp(-\mathrm{i}\mathbf{H}t)^\dagger=\exp(\mathrm{i}\mathbf{H}t)$，所以

$$\mathbf{U}(t)^\dagger\mathbf{U}(t)=\exp(\mathrm{i}\mathbf{H}t)\exp(-\mathrm{i}\mathbf{H}t)=\exp(\mathbf{0})=\mathbf{I} \tag{2}$$

注意只有 $[\mathbf{A},\mathbf{B}]=0$ 时才有

$$\exp(\mathbf{A})\exp(\mathbf{B})=\exp(\mathbf{A}+\mathbf{B}) \tag{3}$$

不含时，无穷维的情况

算符和有限维矩阵的性质往往有相同之处，然而当拓展到无穷维的情况时往往就需要高级得多的数学（泛函分析），我们暂不详细介绍这些数学，而是直接通过类比给出结论．

将“一阶线性常微分方程组[333]”中式 1 拓展成偏微分方程[1]，令 A 为算符．

$$\frac{\partial}{\partial t}f(x,y,\ldots,t)=Af(x,y,\ldots,t) \tag{4}$$

那么当 A 不含 t 时，有

$$f(x,y,\ldots,t)=\exp(At)f(x,y,\ldots,0) \tag{5}$$

若 A 含有 t，形式解式 7 [333] 变为

$$f(x,y,\ldots,t)=\hat{\mathcal{T}}\exp\left[\int_0^t A(t')\mathrm{d}t'\right]f(x,y,\ldots,t) \tag{6}$$

[1]形象理解：将矢量 $\mathbf{v}$ 看作有无穷多个元，且取值连续，就成了 $f(x,t)$

含时薛定谔方程的解

我们把哈密顿算符 H 看作是无穷维矩阵，薛定谔方程可记为与一阶线性常微分方程组（式 1 [333]）相同的形式

$$\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = -\mathrm{i}H|\psi(t)\rangle \tag{7}$$

当哈密顿算符 H 不含时，解为（根据式 2 [333]）

$$|\psi(t)\rangle = \exp(-\mathrm{i}Ht)|\psi(0)\rangle \tag{8}$$

当哈密顿算符含时，形式上可以把解记为

$$|\psi(t)\rangle = \hat{\mathcal{T}}\exp\left(\int_0^t H(t')\mathrm{d}t'\right)|\psi(0)\rangle \tag{9}$$

我们把以上的 exp 就是波函数的传播子，其定义依然是使用指数函数的级数展开.

平移算符

预备知识 算符的指数函数[1399]

一维情况

在一维的情况下，**平移算符（translation operator）** $T(a)$ 可以把函数[2] $f(x)$ 整体向右平移 a 得到 $f(x-a)$. 假设 $f(x)$ 是无穷阶可导函数，a 是常数，那么 $f(x-a)$ 关于某点 x 的泰勒级数[181] 可以用表示为

$$f(x-a) = \left[1 + (-a)\frac{\partial}{\partial x} + \frac{1}{2!}(-a)^2\frac{\partial^2}{\partial x^2} + \dots\right]f(x) \tag{1}$$

其中方括号中的部分可以表示为一个算符的指数函数，即

$$T(a) = \exp\left(-a\frac{\partial}{\partial x}\right) \tag{2}$$

我们把这个算符叫做平移算符.

[2]这里的讨论是一般性的，所以这里的函数不一定是量子力学中的波函数

例 1

令 $f(x)=x^2$，现在我们使用平移算符将其向右平移 a.

$$\begin{aligned}\exp\left(-a\frac{\partial}{\partial x}\right)x^2&=\left(1-a\frac{\partial}{\partial x}+\frac{1}{2!}a^2\frac{\partial^2}{\partial x^2}\cdots\right)x^2\\&=x^2-2ax+a^2=(x-a)^2\end{aligned}\tag{3}$$

我们还可以再次使用平移算符，

$$\begin{aligned}\exp\left(-b\frac{\partial}{\partial x}\right)(x-a)^2&=\left(1-b\frac{\partial}{\partial x}+\frac{1}{2!}b^2\frac{\partial^2}{\partial x^2}\cdots\right)(x-a)^2\\&=(x-a)^2-2b(x-a)+b^2=(x-a-b)^2\end{aligned}\tag{4}$$

这就验证了 $T(b)T(a)=T(a+b)$，即

$$\exp\left(-a\frac{\partial}{\partial x}\right)\exp\left(-b\frac{\partial}{\partial x}\right)=\exp\left[-(a+b)\frac{\partial}{\partial x}\right]\tag{5}$$

在量子力学中，由于（位置表象下的）动量算符为 $p=-\mathrm{i}\hbar\,\partial/\partial x$，平移算符可记为

$$T(a)=\exp\left(-\frac{\mathrm{i}}{\hbar}ap\right)\tag{6}$$

三维情况

将以上的一维动量算符 p 换成三维的 $\mathbf{p}=-\mathrm{i}\hbar\boldsymbol{\nabla}$ 即可

$$T(\mathbf{a})=\exp(-\mathbf{a}\cdot\boldsymbol{\nabla})=\exp\left(-\frac{\mathrm{i}}{\hbar}\mathbf{a}\cdot\mathbf{p}\right)\tag{7}$$

推导同理.

旋转算符

预备知识 平移算符[1400]，梯度[307]

如果要将一个三维函数 $f(\mathbf{r})$ 绕 z 轴旋转角 α，我们可以使用极坐标 $f(r,\theta,\phi)$，并用平移算符 $\exp(-\alpha\,\partial/\partial\phi)$ 对坐标 ϕ 进行“平移”. 那么球坐

标中的算符 $\partial/\partial\phi$ 在直角坐标中如何表示呢？令 $s=\sqrt{x^2+y^2}$，该算符的意义是求函数在 $\hat{\boldsymbol{\phi}}$ 方向的方向导数乘以 s．$\hat{\boldsymbol{\phi}}=(-y/s,x/s,0)$，所以

$$\frac{\partial f}{\partial\phi}=s\boldsymbol{\nabla}f\cdot\hat{\boldsymbol{\phi}}=x\frac{\partial f}{\partial y}-y\frac{\partial f}{\partial x}=\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right)f \tag{1}$$

于是我们就得到直角坐标系中绕 z 轴逆时针旋转的算符为

$$R_z(\alpha)=\exp\left[-\alpha\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right)\right] \tag{2}$$

同理，我们可以将极坐标的极轴指向 x 或 y 轴正方向，从而得出绕 x 或 y 轴逆时针旋转角 α 的算符分别为

$$\begin{aligned}R_x(\alpha)&=\exp\left[-\alpha\left(y\frac{\partial}{\partial z}-z\frac{\partial}{\partial y}\right)\right]\\R_y(\alpha)&=\exp\left[-\alpha\left(z\frac{\partial}{\partial x}-x\frac{\partial}{\partial z}\right)\right]\end{aligned} \tag{3}$$

例 1

要将 $f(x,y,z)=x$ 绕 z 轴旋转 α 角，就计算

$$\begin{aligned}&\exp\left[-\alpha\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right)\right]x\\=&\,x-\alpha\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right)x+\frac{1}{2!}\alpha^2\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right)^2x-\ldots\end{aligned} \tag{4}$$

其中

$$\begin{aligned}&\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right)x=-y\\&\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right)^2x=\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right)(-y)=-x\\&\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right)^3x=\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right)(-x)=y\\&\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right)^4x=\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right)y=x\\&\ldots\end{aligned} \tag{5}$$

所以

$$\begin{aligned}&\exp\left[-\alpha\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right)\right]x\\&=\left(1-\frac{1}{2!}\alpha^2+\frac{1}{4!}\alpha^4\dots\right)x+\left(\alpha-\frac{1}{3!}\alpha^3+\frac{1}{5!}\alpha^5\dots\right)y\\&=x\cos\alpha+y\sin\alpha\end{aligned}\tag{6}$$

显然这就是旋转后所得的函数.

在量子力学中，由直角坐标系中角动量算符 L_x, L_y, L_z 的定义，三个旋转算符分别可以记为

$$\exp\left(-\mathrm{i}\frac{\alpha L_x}{\hbar}\right)\qquad\exp\left(-\mathrm{i}\frac{\alpha L_y}{\hbar}\right)\qquad\exp\left(-\mathrm{i}\frac{\alpha L_z}{\hbar}\right)\tag{7}$$

角动量加法（量子力学）

预备知识 轨道角动量[1391]，张量积空间[513]

考虑两个系统的角动量空间，基底分别为 $|l_1,m_1\rangle$ 和 $|l_2,m_2\rangle$，其中 l_1,l_2 固定空间的维度分别为 $2l_1+1$ 和 $2l_2+1$．两空间的角动量算符分别为

$$L_1^2\qquad L_{1x}\qquad L_{1y}\qquad L_{1z}\tag{1}$$

$$L_2^2\qquad L_{2x}\qquad L_{2y}\qquad L_{2z}\tag{2}$$

我们用这两个空间生成 $(2l_1+1)(2l_2+1)$ 维的张量积空间，基底为 $|l_1,m_1\rangle\otimes|l_2,m_2\rangle$（或记为 $|l_1,m_1,l_2,m2\rangle$）．张量积空间中一组显然的 **Complete Set of Commutable Operators（CSCO）**为

$$\{L_{1z},L_{2z}\}\tag{3}$$

也就是说只需要 m_1,m_2 两个量子数就可以唯一确定一个基底.

在张量积空间上定义两个系统总角动量算符为[3]

$$\begin{aligned}L_x&=L_{1x}\otimes I+I\otimes L_{2x}\\L_y&=L_{1y}\otimes I+I\otimes L_{2y}\\L_z&=L_{1z}\otimes I+I\otimes L_{2z}\end{aligned}\tag{4}$$

[3]注意有时候为了方便我们会直接记为如 $L_x=L_{1x}+L_{2x}$ 的形式，严格来说这是不对的，因为等号左边的算符所在的空间是等号右边两个算符所在空间的张量积，而并非同一空间中的两个算符相加.

$$L^2 = L_x^2 + L_y^2 + L_z^2 \tag{5}$$

可以证明新增的对易关系为

$$[L^2, L_z] = [L^2, L_1^2] = [L^2, L_2^2] = [L_z, L_{1z}] = [L_z, L_{2z}] = 0 \tag{6}$$

由此可以得到另一组 CSCO 为

$$\{L^2, L_z\} \tag{7}$$

即我们可以想得到张量积空间的另一组基底．令两个算符的本征值（量子数）分别为 L 和 M，这组基底可记为 $|L, M\rangle$．首先由对易关系，$|l_1, m_1, l_2, m_2\rangle$ 已经是 L_z 的本征矢，每个本征值 M 对应一个本征子空间（因为有简并），以下称为“M 子空间”．子空间的维数 N_M 是满足 $m_1 + m_2 = M$ 的不同 m_1, m_2 组合数（图 1 左）．例如图中当 $M = -1/2$ 时 $N_M = 4$．

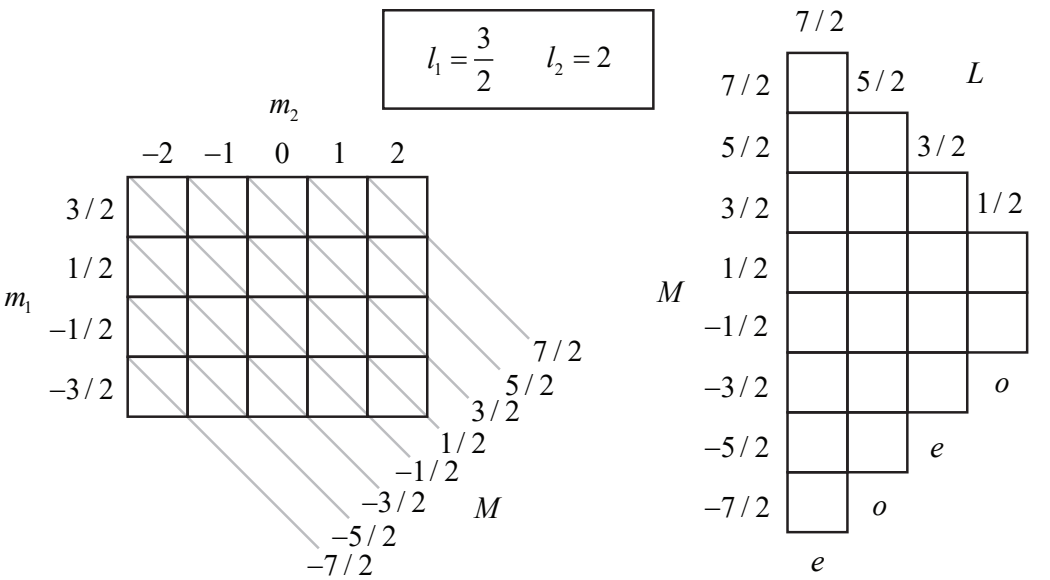


图 1: 角量子数分别为 $l_1 = 3/2$ 和 $l_2 = 2$ 的两个粒子进行角动量加法．图中每个方格代表张量积空间中一个基底．左图中每个基底是 $|l_1, m_1, l_2, m_2\rangle$ 的形式，每个条斜线穿过的基底张成一个 M 子空间．而右图每个基底是 $|L, M\rangle$ 的形式．基底变换在每个 M 子空间中独立进行，即左图的一条斜线变成右图的一行．

$$N_M = \begin{cases} l_1 + l_2 - |M| + 1 & (|M| > |l_1 - l_2|) \\ 2\min\{l_1, l_2\} + 1 & (|M| \leqslant |l_1 - l_2|) \end{cases} \tag{8}$$

M 子空间中每组基底中 m_1 的最大值为

$$\max\{m_1\} = \begin{cases} l_1 & (M \geqslant l_1 - l_2) \\ l_2 + M & (M < l_1 - l_2) \end{cases} \tag{9}$$

例如图中当 $M = 1/2$ 时 m_1 的最大值是 $3/2$.

线性代数中基底的排列顺序十分重要，顺序会影响坐标和矩阵的值．我们通常令 $|l_1, m_1, l_2, m_2\rangle$（$m_2 = M - m_1$）按 m_1 从大到小的顺序排列，$|L, M\rangle$ 按 L 从大到小排序.

可以证明在每个 M 子空间中，$|L, M\rangle$ 取 $L = l_1 + l_2, l_1 + l_2 - 1, \ldots$（共 N_M 个）．L 在所有 M 子空间的最小值是 $|l_1 - l_2|$，当 $|M| \leqslant |l_1 - l_2|$ 时取得（图 1 右边最后一列），所以 L 在整个张量积空间中的范围是

$$|l_1 - l_2| \leqslant L \leqslant l_1 + l_2 \tag{10}$$

若已知两个粒子的 L，那么它们各自的 l_1, l_2 可能是多少呢？根据式 10 可以画出图 2．例如当 $L = 0$ 时只能允许 $l_1 = l_2$.

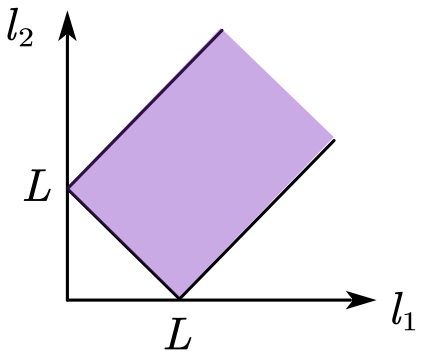


图 2: 对于给定的 L，阴影部分是允许的 l_1, l_2，阴影向右上方无限延申.

CG 系数

既然每个 M 子空间可以通过 N_M 个正交归一的 $|l_1, m_1, l_2, m_2\rangle$ 基底或者 $|L, M\rangle$ 基底展开，那么就有必要讨论它们之间的酉变换矩阵 $\mathbf{U}_M$．其矩阵元 $\langle l_1, m_1, l_2, m_2 | L, M\rangle$ 被称为 **Clebsch－Gordan 系数**或者简称 **CG 系数**．详见

“Clebsch–Gordan 系数[852]”.

$$|L,M\rangle=\sum_{m_1,m_2}\langle l_1,m_1,l_2,m_2|L,M\rangle\cdot|l_1,m_1,l_2,m_2\rangle \tag{11}$$

$$|l_1,m_1,l_2,m_2\rangle=\sum_{L}\langle L,M|l_1,m_1,l_2,m_2\rangle\cdot|L,M\rangle \tag{12}$$

注意式 11 对 m_1,m_2 求和需要满足 $M=m_1+m_2$，所以以上两式都是 N_M 项求和．CG 系数也可以用另外两种符号记为

$$\begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{bmatrix}=C^{L,M}_{l_1,m_1,l_2,m_2}=\langle l_1,m_1,l_2,m_2|L,M\rangle \tag{13}$$

要求出 CG 系数，我们只需要在每个 M 子空间中把 $|l_1,m_1,l_2,m_2\rangle$ 基底下的 L^2 矩阵对角化

$$L^2=L_1^2+L_2^2+2(L_{1x}L_{2x}+L_{1y}L_{2y}+L_{1z}L_{2z}) \tag{14}$$

其中只有 $L_{1x}L_{2x}+L_{1y}L_{2y}$ 不是对角矩阵．可以利用升降算符表示为

$$2(L_{1x}L_{2x}+L_{1y}L_{2y})=L_{1+}L_{2-}+L_{1-}L_{2+} \tag{15}$$

于是以 $|l_1,m_1,l_2,m_2\rangle$ 为基底，L^2 矩阵的矩阵元为

$$\begin{aligned}&\langle l_1,m_1',l_2,m_2'|\,L^2\,|l_1,m_1,l_2,m_2\rangle=\hbar^2\times\\ &\begin{bmatrix} \delta_{m_1',m_1}\delta_{m_2',m_2}[l_1(l_1+1)+l_2(l_2+1)+2m_1m_2] \\ +\delta_{m_1',m_1+1}\delta_{m_2',m_2-1}\sqrt{l_1(l_1+1)-m_1(m_1+1)}\sqrt{l_2(l_2+1)-m_2(m_2-1)} \\ +\delta_{m_1',m_1-1}\delta_{m_2',m_2+1}\sqrt{l_1(l_1+1)-m_1(m_1-1)}\sqrt{l_2(l_2+1)-m_2(m_2+1)}\end{bmatrix}\end{aligned} \tag{16}$$

可以看出这是一个三对角矩阵，其 N_M 个正交归一的本征列矢量就是 $|L,M\rangle$（$L=l_1+l_2,l_1+l_2-1,\dots$）在 $|l_1,m_1,l_2,m_2\rangle$ 基底下的坐标．把这些列矢量按基底顺序从左到右排列就是 $\mathbf{U}_M$.

升降算符法

要如何从 L^2 矩阵的一般形式式 16 得到其本征值以及本征矢呢？一种巧妙的办法是先从唯一非简并的两个 M 空间入手（参考 [14]），由式 14 和式 15 得

$$\begin{aligned}L^2\,|l_1,\pm l_1,l_2,\pm l_2\rangle&=\hbar^2l_1(l_1+1)+\hbar^2l_2(l_2+1)+2\hbar^2l_1l_2\\&=\hbar^2(l_1+l_2)(l_1+l_2+1)\end{aligned} \tag{17}$$

所以令 L^2 的量子数为 L，本征值为 $\hbar^2 L(L+1)$，那么 $|l_1, m_1 = \pm l_1, l_2, m_2 = \pm l_2\rangle$ 对应 $L = l_1 + l_2$．所以

$$|l_1, m_1 = \pm l_1, l_2, m_2 = \pm l_2\rangle = |L = l_1 + l_2, M = \pm(l_1 + l_2)\rangle \tag{18}$$

接下来，对 $|l_1, l_1, l_2, l_2\rangle$ 使用 L_z 的降算符 $L_- = L_-^{(1)} + L_-^{(2)}$ 就得到了图 1 的第二行左边第一个基底（先不考虑归一化）

$$|L = l_1 + l_2, M = l_1 + l_2 - 1\rangle = L_- |l_1, l_1, l_2, l_2\rangle = |l_1, l_1 - 1, l_2, l_2\rangle + |l_1, l_1, l_2, l_2 - 1\rangle \tag{19}$$

但 $M = l_1 + l_2 - 1$ 子空间是一个 2 维子空间，所以我们容易找到另一个正交归一基底 $|l_1, l_1 - 1, l_2, l_2\rangle + |l_1, l_1, l_2, l_2 - 1\rangle$，并且这就是 $|L = l_1 + l_2 - 1, M = l_1 + l_2 - 1\rangle$．

我们可以再次对 $M = l_1 + l_2 - 1$ 子空间中的两个基底使用将算符，得到 $M = l_1 + l_2 - 2$ 中前两个基底 $|l_1 + l_2, M\rangle$，和 $|l_1 + l_2 - 1, M\rangle$．但这是一个三维子空间，所以可以找到第三个正交基底 $|l_1 + l_2 - 2, M\rangle$．以此类推，就可以找到所有 $(2l_1 + 1)(2l_2 + 1)$ 个 $|L, M\rangle$ 基底．

氢原子基态的波函数

预备知识 量子力学[60]，多变量分布函数[728]

由于波函数的统计诠释，统计在量子力学中经常碰到，所以这里我们通过一个例子来熟悉一下统计的一些常见计算．

氢原子是唯一有解析解的原子，因为它结构简单，只有一个核外电子．由于核外电子质量又远小于原子核的质量，忽略核的运动，且不计万有引力．

氢原子基态的波函数为

$$\psi(\mathbf{r}) = A\mathrm{e}^{-r/a} \tag{1}$$

其中 $a = 4\pi\varepsilon_0\hbar^2/(m_e e^2)$ 是量子理论中一个重要的常数，**玻尔半径**．由于这是个球对称函数，所以氢原子的波函数通常在球坐标中表示，即表示成三个球坐标的函数 $\psi(\mathbf{r}) = \psi(r, \theta, \phi)$．其模长平方同样表示粒子在某点出现的概率密度．由于氢原子基态的波函数是球对称的，波函数只是 r 的函数．

归一化

概率密度必须归一化，也就是说，在所有地方找到电子的概率之和为必为 1．所以可以用归一化来确定波函数前面的系数 A．把概率密度对整个空间体积分

$$1=\int|\psi(\mathbf{r})|^2\,\mathrm{d}V=\int_0^{+\infty}A^2\mathrm{e}^{-2r/a}\cdot 4\pi r^2\,\mathrm{d}r=A^2\pi a^3 \tag{2}$$

所以 $A=1/\sqrt{\pi a^3}$，即归一化的波函数为

$$\psi(\mathbf{r})=\frac{1}{\sqrt{\pi a^3}}\mathrm{e}^{-r/a} \tag{3}$$

位置的平均值

根据连续概率分布中平均值（或数学期望）的定义（式 7 [728]）

$$\langle\mathbf{r}\rangle=\int\mathbf{r}|\psi(\mathbf{r})|^2\,\mathrm{d}V=\mathbf{0} \tag{4}$$

积分显然为零，因为波函数关于中心呈球对称分布，各个方向的 $\mathbf{r}$ 互相抵消了．所以如果对足够多个处于基态的氢原子测量电子的位置，并求平均位置（矢量），一定会在原子核处．

电子离原子核距离的平均值

如果在上题中，求平均值的不是位置矢量，而是位置的大小，那么结果显然是大于零的．

$$\langle r\rangle=\int r|\psi(\mathbf{r})|^2\,\mathrm{d}V=\int rA^2\mathrm{e}^{-2r/a}\cdot 4\pi r^2\,\mathrm{d}r=\frac{3}{8}a^4A=\frac{3}{2}a \tag{5}$$

注意这比玻尔半径要大．

电子最可能出现的位置

一个位置 $\mathbf{r}$ 的波函数模长平方 $|\psi(\mathbf{r})|^2$ 越大，电子越有可能出现在这个位置．所以现在要求的是概率密度出现最大值的位置．

根据指数函数的性质，最大值 $|\psi(\mathbf{r})|^2_{\max}=\left(\mathrm{e}^{-0/a}/\sqrt{\pi a^3}\right)^2=1/(\pi a^3)$，最大值位置为 $\mathbf{r}=\mathbf{0}$．所以电子最可能出现在原子核处．

电子与原子核最有可能的距离

若定义径向概率密度为

$$f(r)=\lim_{\Delta r\to 0}\frac{\int_r^{r+\Delta r}|\psi(\mathbf{r})|^2\cdot 4\pi r^2\,\mathrm{d}r}{\Delta r}=4\pi r^2|\psi(\mathbf{r})|^2 \tag{6}$$

那么电子与原子核的距离落在 $[r_1,r_2]$ 区间中的概率就是

$$P(r_1\le r\le r_2)=\int_{r_1}^{r_2}f(r)\,\mathrm{d}r \tag{7}$$

若要求最可能出现的半径，即 $f(r)$ 的最大值对应的 r，可以对其求导（见“导数与函数极值[170]”）. 令 $\mathrm{d}f(r)/\mathrm{d}r=0$ 即 $8r\mathrm{e}^{-2r/a}/a^3-(4/a^3)(2/a)r^2\mathrm{e}^{-2r/a}=0$，解得 $r=a$.

这个重要结论说明，**玻尔半径就是氢原子基态中电子与原子核最有可能的距离**. 注意不要把“最有可能的距离”和“最有可能的位置”的距离混淆：电子出现在 $r=a$ 球面上任意**给定一点**的概率（概率密度，下同）都没有 $r=0$ 的概率大，但是出现在 $r=a$ 球面上**任意一点**的概率则要比出现在 $r=0$ 这点的概率大得多.

球坐标系中的定态薛定谔方程

预备知识 定态薛定谔方程[1328]，球坐标系中的亥姆霍兹方程[751]，球坐标系中的角动量算符[1393]

本文使用原子单位制[1315]. 我们希望在球坐标中求解定态薛定谔方程

$$-\frac{1}{2m}\nabla^2\psi(\mathbf{r})+V(\mathbf{r})\psi(\mathbf{r})=E\psi(\mathbf{r}) \tag{1}$$

当势能 $V(\mathbf{r})$ 处处为零时，这就变成了球坐标系中的亥姆霍兹方程[751]. 以下考虑一般情况，但过程也和求解亥姆霍兹方程类似.

使用球坐标的拉普拉斯算子（式 4 [742]）可以将哈密顿算符表示为

$$H=-\frac{1}{2m}\nabla^2+V(\mathbf{r})=K_r+\frac{L^2}{2mr^2}+V(\mathbf{r}) \tag{2}$$

其中径向动量算符和角动量平方算符为（式 5 [1393]）分别为

$$K_r = -\frac{1}{2m}\nabla_r^2 = -\frac{1}{2m}\left(\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r}\right) = -\frac{1}{2mr^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\partial}{\partial r}\right) \tag{3}$$

$$L^2 = -\nabla_\Omega^2 = -\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 u}{\partial\phi^2}\right] \tag{4}$$

注意角动量算符不含 r.

定态薛定谔方程式 1 变为

$$\left(K_r + \frac{L^2}{2mr^2} + V - E\right)\Psi(\mathbf{r}) = 0 \tag{5}$$

我们假设势能函数只与粒子到原点的距离有关，即 $V = V(r)$. 两边乘以 r^2 可以将 r 与角向变量 θ, ϕ（简写为 $\hat{\mathbf{r}}$）分离，令 $\Psi = R(r)Y(\hat{\mathbf{r}})$.

解得 $Y(\hat{\mathbf{r}})$ 为球谐函数 $Y_{l,m}(\hat{\mathbf{r}})$ 满足

$$L^2 Y_{l,m}(\hat{\mathbf{r}}) = l(l+1)Y_{l,m}(\hat{\mathbf{r}}) \tag{6}$$

分离变量后 $R(r)$ 满足的方程一般被称为**径向薛定谔方程**

$$K_r R_l(r) + \left[V(r) + \frac{l(l+1)}{2mr^2}\right]R_l(r) = ER(r) \tag{7}$$

我们可以通过变量替换将其化为更简洁的形式. 实际上是不同 l 的一系列方程.

定义（**约化径向波函数（scaled radial wave function）**）为

$$\psi_l(r) = rR_l(r) \tag{8}$$

代入式 7，第一项变为

$$K_r R_l = -\frac{1}{2m}\left(\frac{\mathrm{d}^2 R_l}{\mathrm{d}r^2} + \frac{2}{r}\frac{\mathrm{d}R_l}{\mathrm{d}r}\right) = -\frac{1}{2mr^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}R_l}{\mathrm{d}r}\right) = -\frac{1}{2mr}\frac{\mathrm{d}^2\psi_l}{\mathrm{d}r^2} \tag{9}$$

所以式 7 两边乘 r 后化简为

$$-\frac{1}{2m}\frac{\mathrm{d}^2\psi_l}{\mathrm{d}r^2} + \left[V(r) + \frac{l(l+1)}{2mr^2}\right]\psi_l = E\psi_l \tag{10}$$

这是径向薛定谔方程更常见的形式，方括号中可以看作**一维等效势能**（类比经典力学的情况式 12 [1034]）. 该方程的解取决于 $V(r)$ 的具体形式. 一般来说，对

于每个较小的 l，可能找到有限个束缚态，我们把它用一个整数 n 编号，叫做**主量子数（principal quantum number）**，而 l 叫做**角量子数（angular quantum number）**，束缚态能量 $E_{n,l}$ 由这两个量子数共同决定．随着 l 不断增加，方括号中的等效势能越来越浅，最终有可能导致径向方程不存在束缚态解．氢原子是一个显著的例外，由于 $1/r$ 势能的特殊性，它不但有无穷个束缚态，而且束缚态能量和 l 无关，这在一般情况下是不成立的．显然，对于有限深的势能函数 $V(r)$ 只存在有限个束缚态．

总波函数体积分要求

$$\int |R|^2|Y|^2 r^2 \,\mathrm{d}\Omega \,\mathrm{d}r = 1 \tag{11}$$

球谐函数已经满足 $\int |Y|^2 \,\mathrm{d}\Omega = 1$，所以，要求

$$\int |R|^2 r^2 \,\mathrm{d}r = 1 \tag{12}$$

即

$$\int |\psi_l(r)|^2 \,\mathrm{d}r = 1 \tag{13}$$

柱坐标系中的薛定谔方程

- 本词条处于草稿阶段．

预备知识 球坐标系中的定态薛定谔方程[1409]

$$u(r) = \sqrt{r}R(r) \tag{1}$$

$$H = K_r + \frac{L_z^2}{2mr^2} \tag{2}$$

$$K_r R = -\frac{1}{2m}\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}R}{\mathrm{d}r}\right) = -\frac{1}{2m}\frac{1}{\sqrt{r}}\left(\frac{\mathrm{d}^2u}{\mathrm{d}r^2} + \frac{u}{4r^2}\right) \tag{3}$$

$$\frac{L_z^2}{2mr^2}\psi = \frac{1}{2m}\frac{m_z^2}{r^2}\psi \tag{4}$$

所以径向方程为

$$-\frac{1}{2m}\frac{\mathrm{d}^2u}{\mathrm{d}r^2} + \left[V(r) + \frac{1}{2m}\left(\frac{m_z^2}{r^2} - \frac{1}{4r^2}\right)\right]u = Eu \tag{5}$$

球坐标中的薛定谔方程

- 本词条需要更多讲解，便于帮助理解.
- 这篇大概是用来数值解 TDSE 的.

预备知识 球坐标系中的定态薛定谔方程[1409]，张量积空间[513]

本文使用原子单位制[1315]. 无论 TDSE 是否可分离变量，在球坐标系中用球谐函数都是常用的做法

$$\Psi(\mathbf{r},t)=\sum_{l,m}R_l(r,t)Y_{l,m}(\hat{\mathbf{r}})=\sum_{l,m}\frac{1}{r}\psi_{l,m}(r,t)Y_{l,m}(\hat{\mathbf{r}}) \tag{1}$$

我们可以将 $\psi_l(r,t)Y_{l,m}(\hat{\mathbf{r}})$ 看做径向空间和角向空间中态矢的张量积. 我们将 l,m 的组合进行排序并给每个组合一个全局下标 i 或 j.

$$\Psi(\mathbf{r},t)=\sum_{j}R_j(r,t)Y_j(\hat{\mathbf{r}})=\sum_{j}\frac{1}{r}\psi_j(r,t)Y_j(\hat{\mathbf{r}}) \tag{2}$$

将波函数代入含时薛定谔方程

$$H\sum_j|R_j\rangle|Y_j\rangle=\mathrm{i}\frac{\partial}{\partial t}\sum_j|R_j\rangle|Y_j\rangle \tag{3}$$

左乘 $\langle Y_i|$，可以将角向坐标积去，得到一组径向函数的 coupled equation. 这不完全是 TDSE 的矩阵形式，因为我们没有在径向选取基底[4].

$$\sum_j\langle Y_i|H|Y_j\rangle|R_j\rangle=\mathrm{i}\frac{\partial}{\partial t}|R_j\rangle \tag{4}$$

如果 TDSE 可以分离变量，$|Y_j\rangle$ 是 H 的本征矢，那各个径向波函数将会是独立的（uncoupled）.

例如无外场时的中心力场哈密顿算符有三项

$$H=K_r+\frac{L^2}{2mr^2}+V(r) \tag{5}$$

[4]另一种理解是在径向选取 $\delta(r-r_0)$ 作为基底，但本征值连续的基底比较复杂，就不这么想吧.

它们分别都是对角矩阵，K_r 和 $V(r)$ 的每个对角元都是一样的，我们已经见到过，第二项与 l 有关．含时薛定谔方程变为

$$-\frac{1}{2m}\frac{\mathrm{d}^2\psi_{l,m}}{\mathrm{d}r^2}+\left[V(r)+\frac{l(l+1)}{2mr^2}\right]\psi_{l,m}=\mathrm{i}\frac{\partial}{\partial t}\psi_{l,m} \tag{6}$$

但是如果我们有一个任意方向的外场，哈密顿算符就会新增一项 $V_f=\mathbf{E}\cdot\mathbf{r}$．显然这个哈密顿算符不是对角矩阵

$$\langle Y_i|\mathbf{E}\cdot\mathbf{r}|Y_j\rangle \tag{7}$$

那这一项就叫做 coupling term．由此可以计算出选择定则（selection rule）．

量子力学中的变分法

当平均能量是波函数的鞍点时，波函数就是能量的本征态．对一维单粒子

$$E=\langle\psi|H|\psi\rangle \tag{1}$$

但注意这里的波函数必须已经归一化．由于变分法需要假设任意的增量函数 $\delta\psi$，我们只好用一个不要求归一化的能量平均值公式

$$E=\langle\psi|H|\psi\rangle/\langle\psi|\psi\rangle \tag{2}$$

现在假设波函数增加 $\delta\psi$

$$E\langle\delta\psi|\psi\rangle+E\langle\psi|\delta\psi\rangle=\langle\delta\psi|H|\psi\rangle+\langle\psi|H|\delta\psi\rangle \tag{3}$$

由于 $\delta\psi$ 是任意的，我们也可以使用 $\mathrm{i}\delta\psi$

$$-E\langle\delta\psi|\psi\rangle+E\langle\psi|\delta\psi\rangle=-\langle\delta\psi|H|\psi\rangle+\langle\psi|H|\delta\psi\rangle \tag{4}$$

以上两式等效，两式相减，得（相当于 ψ^* 与 ψ 是两个独立的变量函数）

$$E\langle\delta\psi|\psi\rangle=\langle\delta\psi|H|\psi\rangle \tag{5}$$

该式对任意微小函数增量 $\delta\psi$ 都要求成立．现在如果令 $\delta\psi=\delta(x)$，我们得到薛定谔方程

$$H|\psi\rangle=E|\psi\rangle \tag{6}$$

归一化条件下的变分法也可以由拉格朗日乘数法完成，令

$$L = \langle\psi|H|\psi\rangle - \lambda[\langle\psi|\psi\rangle - 1] \tag{7}$$

类似以上过程，同样有

$$\langle\delta\psi|H|\psi\rangle - \lambda\langle\delta\psi|\psi\rangle = 0 \tag{8}$$

即

$$H|\psi\rangle = \lambda|\psi\rangle \tag{9}$$

显然，乘数 λ 就是本征态能量.

含时微扰理论

预备知识 薛定谔方程[1331]

薛定谔方程的矩阵形式

[5]在讲微扰理论之前，我们先来看如何把含时薛定谔方程写为矩阵的形式. 含时薛定谔方程的一般形式为

$$H|\psi(t)\rangle = \mathrm{i}\hbar\frac{\partial}{\partial t}|\psi(t)\rangle \tag{1}$$

我们把哈密顿算符分为不含时部分 H_0 和含时部分 $H'(t)$

$$H = H_0 + H'(t) \tag{2}$$

我们已经知道 $H'(t) = 0$ 的情况下含时薛定谔方程的通解：先解出对应的定态薛定谔方程的波函数 $|\psi_n\rangle$ 和能级 E_n，通解可表示为（为简单起见我们暂时假设 H_0 只有离散的束缚，例如简谐振子[1340]）

$$|\psi_0(t)\rangle = \sum_n c_n|\psi_n\rangle\,\mathrm{e}^{-\mathrm{i}E_nt/\hbar} \tag{3}$$

[5]本文参考 [14].

注意其中 c_n 为常数，由初始波函数决定．我们可以定义一组**含时**基底

$$|\phi_n(t)\rangle = |\psi_n\rangle \, \mathrm{e}^{-\mathrm{i}E_n t/\hbar} \qquad (n = 1, 2, \dots) \tag{4}$$

用于展开式 1 中的含时波函数．注意任何时刻这组基底都正交归一．

$$|\psi(t)\rangle = \sum_n c_n(t) \, |\phi_n(t)\rangle = \sum_n c_n(t) \, |\psi_n\rangle \, \mathrm{e}^{-\mathrm{i}E_n t/\hbar} \tag{5}$$

由于基底并不是总哈密顿算符 H 的本征矢，系数需由常数拓展为时间的函数 $c_n(t)$.

选择了基底后，就可以把薛定谔方程表示为矩阵的形式．把上式代入薛定谔方程（式 1）得

$$\begin{aligned} &\sum_n c_n(t) H_0 \, |\phi_n(t)\rangle + \sum_n c_n(t) H'(t) \, |\phi_n(t)\rangle \\ &= \mathrm{i}\hbar \sum_n \frac{\mathrm{d}}{\mathrm{d}t} c_n(t) \, |\phi_n(t)\rangle + \mathrm{i}\hbar \sum_n c_n(t) \frac{\mathrm{d}}{\mathrm{d}t} \, |\phi_n(t)\rangle \end{aligned} \tag{6}$$

考虑到

$$H_0 \, |\phi_n(t)\rangle = \mathrm{i}\hbar \frac{\mathrm{d}}{\mathrm{d}t} \, |\phi_n(t)\rangle \tag{7}$$

可化简为

$$\sum_n c_n(t) H'(t) \, |\phi_n(t)\rangle = \mathrm{i}\hbar \sum_n \frac{\mathrm{d}}{\mathrm{d}t} c_n(t) \, |\phi_n(t)\rangle \tag{8}$$

两边左乘 $\langle\phi_m(t)| = \mathrm{e}^{\mathrm{i}E_m t/\hbar} \langle\psi_m|$ （即要求每个分量相等），且令

$$\omega_{mn} = \frac{E_m - E_n}{\hbar} \tag{9}$$

$$H'_{mn}(t) = \langle\psi_m| \, H'(t) \, |\psi_n\rangle \tag{10}$$

$$\tilde{H}_{mn}(t) = \langle\phi_m(t)| \, H'(t) \, |\phi_n(t)\rangle = H'_{mn} \mathrm{e}^{\mathrm{i}\omega_{mn} t/\hbar} \tag{11}$$

得

$$\sum_n \tilde{H}_{mn}(t) c_n(t) = \mathrm{i}\hbar \frac{\mathrm{d}}{\mathrm{d}t} c_m(t) \tag{12}$$

写成矩阵形式为（对矢量求导即对每个分量分别求导[274]）

$$\tilde{\mathbf{H}}\mathbf{c} = \mathrm{i}\hbar \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{c} \tag{13}$$

到此为止我们还没有做任何近似，该式和式 1 完全等效．

含时微扰理论

若哈密顿算符中的势能包含时间，只有极少数情况下存在解析解．这时我们可以用**含时微扰理论**来近似求解．类比不含时微扰理论，我们引入一个常数 λ 来分离不同阶数的近似，最后只需令 $\lambda = 1$ 即可．理论上当阶数足够高时，近似解将会收敛到精确解．

令哈密顿算符，系数矢量分别为

$$H = H_0 + \lambda H'(t) \tag{14}$$

$$\mathbf{c}(t) = \mathbf{c}^{(0)}(t) + \lambda \mathbf{c}^{(1)}(t) + \lambda^2 \mathbf{c}^{(2)}(t) \dots \tag{15}$$

代入式 13，根据 $\lambda = 1$ 的阶数分离方程，得

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{c}^{(0)}(t) = \mathbf{0} \qquad （0\ 阶近似） \tag{16}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{c}^{(1)}(t) = \frac{1}{\mathrm{i}\hbar}\tilde{\mathbf{H}}(t)\mathbf{c}^{(0)}(t) \qquad （1\ 阶近似） \tag{17}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{c}^{(n)}(t) = \frac{1}{\mathrm{i}\hbar}\tilde{\mathbf{H}}(t)\mathbf{c}^{(n-1)}(t) \qquad （n\ 阶近似） \tag{18}$$

为了求解各阶近似，我们假设 $t = 0$ 时只有 0 阶系数 $\mathbf{c}^{(0)}(0)$ 不为零．若给出初始波函数 $|\psi(0)\rangle$，可用 $|\phi_n(0)\rangle = |\psi_n\rangle$ 展开得到 $\mathbf{c}^{(0)}(0)$．式 16 说明零阶系数矢量为常数，所以零阶近似解就是 $\mathbf{c}^{(0)}(t) = \mathbf{c}^{(0)}(0)$．继续把 $\mathbf{c}^{(0)}(t)$ 代入式 17，两边对时间从 0 到 t 定积分（矢量的积分即对每个分量分别积分）得

$$\mathbf{c}^{(1)}(t) - \mathbf{c}^{(1)}(0) = \frac{1}{\mathrm{i}\hbar}\int_0^t \tilde{\mathbf{H}}(t)\mathbf{c}^{(0)}(t)\,\mathrm{d}t \tag{19}$$

代入 $\mathbf{c}^{(1)}(0) = 0$，$\mathbf{c}^{(0)}(t) = \mathbf{c}^{(0)}(0)$，得一阶近似解为

$$\mathbf{c}^{(1)}(t) = \frac{1}{\mathrm{i}\hbar}\int_0^t \tilde{\mathbf{H}}(t)\mathbf{c}^{(0)}(0)\,\mathrm{d}t \tag{20}$$

类似地，对式 18 积分，若已知 $\mathbf{c}^{(n-1)}(t)$，有

$$\mathbf{c}^{(n)}(t) = \frac{1}{\mathrm{i}\hbar}\int_0^t \tilde{\mathbf{H}}(t)\mathbf{c}^{(n-1)}(t)\,\mathrm{d}t \tag{21}$$

所以要想得到 n 阶近似解，积分 n 次即可．为了明确起见，式 21 的分量表达式为

$$c_i^{(n)}(t) = \frac{1}{\mathrm{i}\hbar}\int_0^t \sum_j H'_{ij}(t)\mathrm{e}^{\mathrm{i}\omega_{ij}t}c_j^{(n-1)}(t)\,\mathrm{d}t \tag{22}$$

简单的一阶微扰

大多数情况下我们只使用一阶近似，一种简单且常见的情况是，若初态为 H_0 的某个本征态 $|\psi_j\rangle$

$$c_i(t)=\frac{1}{\mathrm{i}\hbar}\int_0^t\langle i|H'(t)|j\rangle\,\mathrm{e}^{\mathrm{i}\omega_{ij}t}\,\mathrm{d}t \tag{23}$$

我们把 $c_i(t)$ 叫做**跃迁幅（transition amplitude）**．所以在一阶近似中，波函数在 t 时刻出现在 $|\psi_i\rangle$ 的概率约为

$$P_{ij}(t)=|c_i(t)|^2 \tag{24}$$

注意一阶微扰仅当 $P_{ij}\ll 1$ 时有效．

在此基础上，一种更简单的情况是：如果 $H'(t)$ 中的时间函数可以分离出来

$$H'(t)=Wf(t) \tag{25}$$

其中 W 是一个不含时的算符．令 $W_{ij}=\langle\psi_i|W|\psi_j\rangle$，有

$$H'_{ij}(t)=\langle\psi_i|H'|\psi_j\rangle=W_{ij}f(t) \tag{26}$$

此时一阶微扰公式（式 20）变为 $f(t)$ 的反傅里叶变换[237]．末态 $|\psi_i\rangle$ （$i\neq j$）的系数为

$$c_i(t)=\frac{W_{ij}}{\mathrm{i}\hbar}\int_0^t f(t)\mathrm{e}^{\mathrm{i}\omega_{ij}t}\,\mathrm{d}t \tag{27}$$

几种含时微扰

预备知识 含时微扰理论[1414]

令初态到末态能量差对应的光子频率为

$$\omega_{fi}=\frac{E_f-E_i}{\hbar} \tag{1}$$

$$S=\frac{1}{\mathrm{i}\hbar}\int_{-\infty}^{+\infty}\langle f|\hat{H}'(t)|i\rangle\,\mathrm{e}^{\mathrm{i}\omega_{fi}}\,\mathrm{d}t \tag{2}$$

当 $\langle f|H'(t)|i\rangle = W_{fi}g(t)$ 时

$$P_{fi} = |S|^2 = \frac{|W_{fi}|^2}{\hbar^2}\left|\int_{-\infty}^{+\infty} g(t)\mathrm{e}^{\mathrm{i}\omega_{fi}t}\,\mathrm{d}t\right|^2 \tag{3}$$

瞬时脉冲 $g(t) = \delta(t - t_0)$

$$\left|\int_{-\infty}^{+\infty} g(t)\mathrm{e}^{\mathrm{i}\omega_{fi}t}\,\mathrm{d}t\right|^2 = \left|\int_{t_0-\epsilon}^{t_0+\epsilon} \delta(t-t_0)\mathrm{e}^{\mathrm{i}\omega_{fi}t}\,\mathrm{d}t\right|^2 = 1 \tag{4}$$

代入得

$$P_{fi} = \frac{|W_{fi}|^2}{\hbar^2} \tag{5}$$

方形脉冲 $g(t)$（从 $t = t_1$ 到 $t = t_2$）

$$\begin{aligned}\left|\int_{-\infty}^{+\infty} g(t)\mathrm{e}^{\mathrm{i}\omega_{fi}t}\,\mathrm{d}t\right|^2 &= \left|\int_{t_1}^{t_2} \mathrm{e}^{\mathrm{i}\omega_{fi}t}\,\mathrm{d}t\right|^2 = \left|\frac{\mathrm{e}^{\mathrm{i}\omega_{fi}t_2} - \mathrm{e}^{\mathrm{i}\omega_{fi}t_1}}{\mathrm{i}\omega_{fi}}\right|^2 \\ &= \frac{\sin^2[\omega_{fi}(t_2-t_1)/2]}{[\omega_{fi}(t_2-t_1)/2]^2}(t_2-t_1)^2 \\ &= \Delta t^2\,\mathrm{sinc}^2[\omega_{fi}\Delta t/2]\end{aligned} \tag{6}$$

概率为

$$P_{fi} = \frac{|W_{fi}|^2}{\hbar^2}\Delta t^2\,\mathrm{sinc}^2[\omega_{fi}\Delta t/2] \tag{7}$$

于瞬时脉冲相比，主要跃迁到附近的 E_2 能级．且时间越长能量变化越小．

简谐振动 $g(t) = \mathrm{e}^{-\mathrm{i}\omega t}$ 乘以方形脉冲（从 $t = t_1$ 到 $t = t_2$）

与上面的推导类似，结果为

$$c_i(t) = -\frac{W_{fi}}{\hbar}\frac{\mathrm{e}^{\mathrm{i}(\omega_{ij}-\omega)t} - 1}{\omega_{ij} - \omega} \tag{8}$$

$$P_{fi} = |c_i(t)|^2 = \frac{|W_{fi}|^2}{\hbar^2}\Delta t^2\,\mathrm{sinc}^2[(\omega_{fi}-\omega)\Delta t/2] \tag{9}$$

这说明，跃迁倾向于增加能量 $\hbar\omega$，时间越长，就越靠近 $\hbar\omega$．要注意真实的简谐微扰往往是 $\cos(\omega t)$，分解为两项积分后，会有干涉效应，结果较为复杂．但若 $\omega \gg \omega_{fi}$ 时可以忽略干涉项．

注意当 $\Delta t \to \infty$ 时，$\mathrm{sinc}^2(x)/\sqrt{\pi} \to \delta(x)$.

简谐振动 $g(t)=\cos(\omega t)$ 乘以方形脉冲（从 $t=t_1$ 到 $t=t_2$）

由于积分关于 $g(t)$ 是线性的，可以将 $\cos(\omega t)$ 拆分为两个指数函数

$$c_i(t) = -\frac{W_{fi}}{2\hbar}\left[\frac{\mathrm{e}^{\mathrm{i}(\omega_{ij}+\omega)t}-1}{\omega_{ij}+\omega} + \frac{\mathrm{e}^{\mathrm{i}(\omega_{ij}-\omega)t}-1}{\omega_{ij}-\omega}\right] \tag{10}$$

作为一个近似，如果 $|\omega-\omega_{ij}| \ll \omega+\omega_{ij}$ 那么第一项可以忽略不计，所以

$$P_{fi} = |c_i(t)|^2 = \frac{|W_{fi}|^2}{4\hbar^2}\Delta t^2\,\mathrm{sinc}^2[(\omega_{fi}-\omega)\Delta t/2] \tag{11}$$

当 $\omega=\omega_{fi}$ 时，跃迁概率和时间平方成正比.

含连续态的微扰理论

- 本词条处于草稿阶段.
- 本词条需要更多参考文献.

预备知识 不含时微扰理论

一般的束缚 + 连续微扰理论. 假设我们有两个束缚态和连续态，总的波函数可以写成

$$|\psi\rangle = C_1|1\rangle + C_2|2\rangle + \int \phi(\mathbf{k})|\mathbf{k}\rangle\,\mathrm{d}^3k \tag{1}$$

令归一化条件为 $\langle 1|1\rangle = \langle 2|2\rangle = 1$，$\langle \mathbf{k}'|\mathbf{k}\rangle = \delta^3(\mathbf{k}'-\mathbf{k})$，otherwise $=0$.

$\mathbf{H}'$ 矩阵可以想象成是这个样子的

方格子代表 $C_{ij} = \langle i|H'|j\rangle$，横条代表 $H_{\mathrm{i}\mathbf{k}'} = \langle i|H'|\mathbf{k}'\rangle$，纵条代表 $H_{\mathbf{k}j} = \langle \mathbf{k}|H'|j\rangle$.

与离散的情况相似，微扰理论的推导方法是先把 $|\psi\rangle$ 代入含时薛定谔方程，然后两边分别左乘基底 $\langle i|$ 和 $\langle \mathbf{k}|$，注意后者这里要使用动量归一化条件把对 $\mathbf{k}$ 的积分消去. 微扰递推公式为

$$C_i^{(n+1)}(t) = \frac{1}{\mathrm{i}\hbar}\int \mathrm{d}t'\left(\sum_{j\neq i} H'_{ij}C_j^{(n)} + \int H_{\mathrm{i}\mathbf{k}'}\phi^{(n)}(\mathbf{k}')\,\mathrm{d}^3k'\right) \tag{2}$$

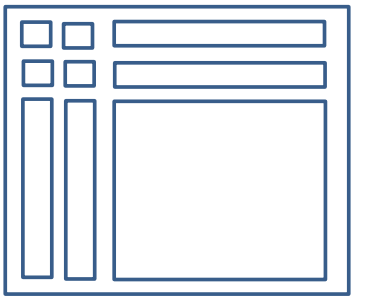


图 1: $\mathbf{H}'$ 矩阵的结构

$$\phi^{(n+1)}(\mathbf{k}) = \frac{1}{\mathrm{i}\hbar}\int \mathrm{d}t' \left(\sum_j H'_{\mathbf{k}j} C_j^{(n)} + \int H_{\mathbf{k}\mathbf{k}'} \phi^{(n)}(\mathbf{k}') \, \mathrm{d}^3 k' \right) \tag{3}$$

关于 $|\mathbf{k}\rangle$ 的定义，若势能函数是局部的，那么在无穷远处波函数是平面波，由此来定义 $\mathbf{k}$．这样，在计算 $\langle \mathbf{k}'|\mathbf{k}\rangle$ 时，由于积分范围是无穷，可以忽略局部势能对波函数的影响，所以归一化系数就是 $\mathrm{e}^{\mathrm{i}\mathbf{k}\mathbf{r}} = \mathrm{e}^{\mathrm{i}k_x x}\mathrm{e}^{\mathrm{i}k_y y}\mathrm{e}^{\mathrm{i}k_z z}$ 的归一化系数即 $1/(2\pi)^{3/2}$．

量子散射

预备知识 散射[1047]，概率流密度[1386]

散射截面

[6]本文使用原子单位制[1315]．散射截面 σ 等于一定时间内被散射的粒子数除以单位截面入射的粒子数．那么从经典力学的角度，如果想象入射粒子流密度是均匀的，σ 可以看做是一个障碍物（无远程作用）的最大横截面面积，微分截面 $\mathrm{d}\sigma/\mathrm{d}\Omega$ 可以理解为单位立体角的散射截面．量子力学中，如果考虑单粒子以平面波入射，那么 σ 等于被散射的概率流（概率/时间）除以入射的概率流密度（概率/时间/面积）．概率流定义为

$$\mathbf{j} = \frac{\mathrm{i}}{2m}(\psi \boldsymbol{\nabla}\psi^* - \psi^* \boldsymbol{\nabla}\psi) \tag{1}$$

[6]本文参考 [19] 相关章节．

微分截面（differential cross section）是能量的函数，可以用概率流定义为

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}=\lim_{r\to\infty}\frac{(\mathbf{j}_{out}\cdot\hat{\mathbf{r}})r^2}{|\mathbf{j}_{in}|} \tag{2}$$

对所有方向积分得总**散射截面（scattering cross section）**

$$\sigma=\int\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\,\mathrm{d}\Omega \tag{3}$$

短程势的边界条件

在三维情况下，每个能量 E 的本征函数都是无穷维简并的．且根据不同的边界条件我们可以获得不同的正交归一基底．常见的边界条件如平面波入射．这是一种物理意义较强的选择，因为如果一个无穷远处的入射波包具有很窄的能量带宽，我们就可以把它近似看作是平面波[7]．平面波经过散射后，会向各个方向发射球面波．于是规定波函数的边界条件为（以下用箭头表示 $r\to\infty$ 时函数的渐进形式）

$$\psi_{\mathbf{k}}^{(+)}(\mathbf{r})\to(2\pi)^{-3/2}\left[\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r})+f(k,\hat{\mathbf{r}})\frac{\exp(\mathrm{i}kr)}{r}\right] \tag{4}$$

$f(k,\hat{\mathbf{r}})$ 是**散射幅**．注意该边界条件只能对短程势能使用，即满足

$$\lim_{r\to\infty}rV(r)=0 \tag{5}$$

的势能，原因见下文．显然库仑势能不属于短程势，我们将在“库仑散射[1431]”讨论．满足边界条件式 4 的波函数也会满足正交归一条件

$$\int\psi_{\mathbf{k}'}^{(+)}(\mathbf{r})^*\psi_{\mathbf{k}}^{(+)}(\mathbf{r})\,\mathrm{d}^3r=\delta(\mathbf{k}-\mathbf{k}') \tag{6}$$

将式 4 代入式 2 可得微分截面就是散射幅的模方

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}=|f(k,\hat{\mathbf{r}})|^2 \tag{7}$$

[7]如果不是，那么入射波包可以由这些散射态叠加而成，之后的时间演化就是这些散射态分别乘以 $\exp(-\mathrm{i}Et)$ 再叠加．但由于这时能量没有精确的定义，我们不能讨论微分截面关于能量的函数．

光学定理

由概率守恒对散射幅的约束，可以得出**光学定理（Optical Theorem）**

$$\sigma = \frac{4\pi}{k}\,\mathrm{Im}[f(k,\hat{\mathbf{k}})] \tag{8}$$

光学定理的物理意义是：球面波往其他方向发射出去的总概率流等于在 $\hat{\mathbf{k}}$ 方向抵消的概率流.

未完成：证明见 [19] 习题 12.2

分波展开

预备知识 球坐标中的薛定谔方程[1412]

除了式 4 的边界条件外，我们也可以找到符合另一种常用边界条件的正交归一散射态，即要求每个散射态同时是能量和角动量 L^2, L_z 的本征态. 这与“氢原子的定态波函数[1536]”类似，只是我们要求能量 $E > 0$.

这时每个能量同样有无穷维简并. 我们在球坐标中解方程. 先讨论简单的情况，即势能为中心势能 $V(r)$，薛定谔方程可以在球坐标中分离变量，使波函数表示为

$$\psi(\mathbf{r}) = \frac{1}{r}\sum_l c_l \psi_{l,m}(r) Y_{l,m}(\hat{\mathbf{r}}) \tag{9}$$

其中每一项被称为一个**分波（partial wave）**，满足上述边界条件. 因为在中心势能 $V(r)$ 下，使解偏微分方程变为解常微分方程，即径向方程

$$-\frac{1}{2m}\frac{\mathrm{d}^2\psi_{l,m}}{\mathrm{d}r^2} + \left[V(r) + \frac{l(l+1)}{2mr^2}\right]\psi_{l,m} = \mathrm{i}\frac{\partial}{\partial t}\psi_{l,m} \tag{10}$$

未完成：这一段讲解有点问题，要说明我们只需要 $m = 0$

相移

对于短程势能 $V(r)$，即

$$\lim_{r\to\infty} rV(r) = 0 \tag{11}$$

可以证明[8]短程势能在无穷远处对相位的影响可以忽略．$\psi_l(r)$ 的渐进表达式就完全可以由**相移（phase shift）** δ_l 来描述．由于微分截面也是在无穷远处定义的，我们可以直接由相移计算微分截面．

我们定义相移 δ_l 满足

$$\psi_l(k,r) \to \sin\left(kr - \frac{l\pi}{2} + \delta_l\right) \tag{12}$$

为什么这么定义呢？因为相移是相对的，我们把 $V(r)\equiv 0$ 时的相位作为基准

$$\psi_l(k,r) = krj_r(kr) \to \sin\left(kr - \frac{l\pi}{2}\right) \tag{13}$$

可以证明（使用式 1 [767] 把式 4 中的平面波展开，再对比系数），将球面波叠加得到式 4 形式后散射幅为

$$f(k,\theta) = \sum_{l=0}^{\infty} f_l(k) P_l(\cos\theta) \tag{14}$$

$$f_l(k) = \frac{2l+1}{k}\exp(\mathrm{i}\delta_l)\sin\delta_l \tag{15}$$

注意这与 ϕ 无关，$f(\hat{\mathbf{r}})$ 是一个轴对称的分布．总截面也可以表示为每个分波的截面之和

$$\sigma(k) = \int |f(k,\theta)|^2\,\mathrm{d}\Omega = \sum_{l=0}^{\infty}\sigma_l(k) \tag{16}$$

$$\sigma_l(k) = \frac{4\pi}{2l+1}|f_l(k)|^2 = \frac{4\pi}{k^2}(2l+1)\sin^2\delta_l(k) \tag{17}$$

这说明，从散射截面只和相移 $\delta_l(k)$ 有关．而数学上可以证明对应同一列 $\delta_l(k)$ 的势能并不是唯一的．

[8]不严谨的证明：无穷远处相位的变化约等于 $\int_0^\infty [V(r) - l(l+1)/(2mr^2)]\,\mathrm{d}r$，只要满足 $rV(r)\to 0$ 这个积分就收敛．

散射矩阵

定义 S 矩阵（虽然叫矩阵，其实是一系列函数）为

$$S_l(k) = \exp(2\mathrm{i}\delta_l) \tag{18}$$

定义 K 矩阵为

$$K_l(k) = \tan\delta_l \tag{19}$$

二者关系为

$$S_l(k) = \frac{1+\mathrm{i}K_l(k)}{1-\mathrm{i}K_l(k)} \tag{20}$$

称为矩阵的原因是，在多通道散射（未完成）中，需要考虑从第不同通道入射和出射的情况，这些量就成为了矩阵，每个矩阵元都是一个函数.

推导散射截面和相位的关系

预备知识 平面波的球谐展开[767]

存在一组变换系数 c_l 使

$$\psi(\mathbf{r}) \to \frac{1}{kr}\sum_l c_l \sin\left(kr - \frac{\pi l}{2} + \delta_l\right) Y_{l,0}(\mathbf{r}) \tag{21}$$

满足边界条件式 4. 令式 4 中散射幅为

$$f(k,\theta) = \sum_l (2l+1)a_l(k)P_l(\cos\theta) \tag{22}$$

假设我们已经在球坐标中解出了 $\psi_{k,l}$，即径向波函数 $R_{k,l}(r)$ 与相移，如何获得 $f(k,\theta)$，即系数 $a_l(k)$ 呢？把 ψ_k 用 $\psi_{k,l}$ 基底展开，即对 P_l 展开，再逐项对比系数即可. 首先展开平面波

$$\mathrm{e}^{\mathrm{i}kz} = \sum_{l=0}^{\infty} \mathrm{i}^l(2l+1)j_l(kr)P_l(\cos\theta) \to \sum_{l=0}^{\infty}(2l+1)\frac{\mathrm{e}^{\mathrm{i}kr} - \mathrm{e}^{-\mathrm{i}(kr-l\pi)}}{2\mathrm{i}kr}P_l(\cos\theta) \tag{23}$$

将式 22 和式 23 代入式 4，再逐项与式 21 对比，得

$$a_l(k) = \frac{1}{k}\sin\delta_l \exp(\mathrm{i}\delta_l) \tag{24}$$

再次代入式 22 可得散射幅与相移的关系.

量子散射的延迟

- 本词条处于草稿阶段.

预备知识 一维散射（量子）[1351]

一个一维波包用傅里叶变换[236] 表示为

$$\psi(x)=\int_{-\infty}^{+\infty}A(\omega)\exp(\mathrm{i}kx-\omega t)\,\mathrm{d}\omega \tag{1}$$

自由粒子情况下 $\omega=k^2/(2m)$. 如果经过一个势阱，不同平面波透射后发生相移 $\phi(\omega)$，经过后，波包为

$$\psi'(x,t)=\int_{-\infty}^{+\infty}A(k)\exp[\mathrm{i}kx-\omega t+\phi(\omega)]\,\mathrm{d}k \tag{2}$$

想象一个特殊情况：经过势阱后 $A(k)$ 不变，$\phi(\omega)=\Delta t\omega$，那么波函数变为

$$\psi'(x,t)=\int_{-\infty}^{+\infty}A(k)\exp[\mathrm{i}kx-\omega(t-\Delta t)]\,\mathrm{d}k=\psi(x,t-\Delta t) \tag{3}$$

这样波包在时间轴上向右平移了 Δt_{EWS}，即延迟. 近似来说，如果波包带宽较窄，频率中心为 ω_0，那么在带宽以内可以把 $\phi(\omega)$ 近似看成是线性的，那么延迟近似为

$$\Delta t_{EWS}(\omega)=\frac{\mathrm{d}\phi}{\mathrm{d}\omega} \tag{4}$$

如果要取一个与波包形状无关的延迟的定义，那么这个定义是最佳选择. 注意这样定义的延迟与频率有关.这个延迟被称为 **Wigner 延迟**或者 **EWS (Eisenbud-Wigner-Smith) 延迟**.

驻相法

使用**驻相法（stationary phase method）**可以分析出波包的近似位置，速度以及时间延迟. 把式 2 中的 x,t 固定，令相位 $kx-\omega t+\phi$ 对 ω 求导为零，有

$$x=v_g\left(t-\frac{\mathrm{d}\phi}{\mathrm{d}\omega}\right) \tag{5}$$

其中 $v_g=\mathrm{d}\omega/\mathrm{d}k$ 是群速度. 同样可以得到式 4.

波恩近似（散射）

预备知识 量子散射[1420]，格林函数[??]

我们还是要解出连续态的不含时波函数，且无穷远处的动量为 $\mathbf{k}_i$（入射平面波的动量）.

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = E\psi \tag{1}$$

$$(\nabla^2 + k^2)\psi = \frac{2m}{\hbar^2}V\psi \equiv U(\mathbf{r})\psi \tag{2}$$

这是非齐次亥姆霍兹方程，其格林函数为（球面波）

$$G(R) = -\frac{\mathrm{e}^{\mathrm{i}kR}}{4\pi R} \tag{3}$$

满足

$$(\nabla^2 + k^2)G(\mathbf{r}) = \delta^3(\mathbf{r}) \tag{4}$$

薛定谔方程的积分形式为

$$\psi(\mathbf{r}) = \psi_0(\mathbf{r}) + \int G(|\mathbf{r}-\mathbf{r}'|)U(\mathbf{r}')\psi(\mathbf{r}')\,\mathrm{d}^3r' \tag{5}$$

ψ_0 是自由粒子波函数，由于无穷远处积分项消失（$1/r$），$\psi_0(\mathbf{r}\to\infty)$ 要求具有动量 $\mathbf{k}_i$，唯一的选择是平面波

$$\psi_0(\mathbf{r}) = A\mathrm{e}^{\mathrm{i}\mathbf{k}_i\cdot\mathbf{r}} \tag{6}$$

由于微分截面定义在无穷远处，我们把格林函数取无穷远处的极限（远场），注意这个极限在定义中，所以并不算是一个近似．这是关于 $\mathbf{r}'$ 的平面波

$$|\mathbf{r}-\mathbf{r}'| \approx r - \hat{r}\cdot\mathbf{r}' \approx r \tag{7}$$

$$G(\mathbf{r},\mathbf{r}') = -\frac{\mathrm{e}^{\mathrm{i}k|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \to -\frac{\mathrm{e}^{\mathrm{i}kr}}{4\pi r}\mathrm{e}^{-\mathrm{i}\mathbf{k}_f\cdot\mathbf{r}'} \tag{8}$$

其中 $\mathbf{k}_f = k\hat{\mathbf{r}}$ 是出射的方向，注意 $|\mathbf{k}_i| = |\mathbf{k}_f|$ 意味着弹性散射.

积分方程求近似解的一般方法是先把一个近似解代入积分内，积分得到一阶修正后的解，再次代入，得到二阶修正后的解，以此类推迭代．波恩近似中，

假设势能相对于入射动能较弱，积分项相当于微扰，所以令初始（零阶）波函数为 $\psi_0(\mathbf{r})$．代入式 5 得一阶修正的波函数，叫做**第一波恩近似**

$$\psi^{(1)}(\mathbf{r}) = A\mathrm{e}^{\mathrm{i}\mathbf{k}_i\cdot\mathbf{r}} - A\frac{m}{2\pi\hbar^2}\frac{\mathrm{e}^{\mathrm{i}kr}}{r}\int \mathrm{e}^{\mathrm{i}(\mathbf{k}_i-\mathbf{k}_f)\cdot\mathbf{r}'}V(\mathbf{r}')\,\mathrm{d}^3r' \tag{9}$$

根据定义，散射幅为

$$f(k,\hat{r}) = -\frac{m}{2\pi\hbar^2}\int \mathrm{e}^{\mathrm{i}(\mathbf{k}_i-\mathbf{k}_f)\cdot\mathbf{r}'}V(\mathbf{r}')\,\mathrm{d}^3r' \tag{10}$$

这相当于势能函数的空间傅里叶变换．

高阶波恩近似

把式 5 多次代入式 5 的积分中，得到精确解的“积分级数”形式

$$\begin{aligned}\psi(\mathbf{r}) = \psi_0(\mathbf{r}) &+ \int \mathrm{d}^3r'\,G(k,\mathbf{r},\mathbf{r}')U(\mathbf{r}')\psi_0(\mathbf{r}') \\ &+ \int \mathrm{d}^3r'\,G(k,\mathbf{r},\mathbf{r}')U(\mathbf{r}')\int \mathrm{d}^3r''\,G(k,\mathbf{r}',\mathbf{r}'')U(\mathbf{r}'')\psi_0(\mathbf{r}'') \\ &+ \int \mathrm{d}^3r'\,G(k,\mathbf{r},\mathbf{r}')U(\mathbf{r}')\times \\ &\int \mathrm{d}^3r''\,G(k,\mathbf{r}',\mathbf{r}'')U(\mathbf{r}'')\int \mathrm{d}^3r'''\,G(k,\mathbf{r}'',\mathbf{r}''')U(\mathbf{r}''')\psi_0(\mathbf{r}''')\ldots\end{aligned} \tag{11}$$

若只计算是指包含前 n 行，就叫第 n 波恩近似．具体计算时，偶尔会用到二阶，基本不会用到三阶或以上．

非常有趣的是，即使我们不假设零阶波函数是平面波，波函数展开成上式时取前 n 行的结果仍然是相同的．

质心系中的多粒子问题

- 本词条处于草稿阶段．

以一维的 H_2^+ 为例，原子核质量为 M，电子质量为 m，哈密顿量为

$$H = T_{N1} + T_{N2} + T_e + V(X_1, X_2, x) \tag{1}$$

若势能函数只与位置有关，那么一般我们会把方程化为质心系方程．令总波函数为

$$\Psi(X_1,X_2,x)=\psi_{cm}(X_{cm})\psi(X_1,X_2,x) \tag{2}$$

构建不同坐标的哈密顿量．第一可以直接用数学的偏微分变量替换，第二可以通过经典力学的哈密顿量来计算．

用两个质量不同的粒子为例会不会更好一点，rigid rotator！或者氢原子！先解释一下约化质量是怎么来的！为什么玻尔模型用约化质量会有那么高的精度，对应的量子力学原理就在这里！

三维简谐振子（球坐标）

- 本词条处于草稿阶段．

预备知识 球坐标系中的径向方程[1409]

希望求解定态薛定谔方程（式 7 [1328]）

$$-\frac{\hbar^2}{2m}\nabla^2\Psi+V(\mathbf{r})\Psi=E\Psi \tag{1}$$

其中势能函数为

$$V(\mathbf{r})=\frac{1}{2}m\omega^2r^2 \tag{2}$$

我已知角向波函数是球谐函数[1409] $Y_{l,m}(\hat{\mathbf{r}})$．只需解出方程（式 10 [1409]）即可

$$-\frac{1}{2m}\frac{\mathrm{d}^2\psi_l}{\mathrm{d}r^2}+\left[V(r)+\frac{l(l+1)}{2mr^2}\right]\psi_l=E\psi_l \tag{3}$$

总波函数和能级为

$$\psi_{n,l,m}=R_{n,l}(r)Y_{l,m}(\Omega)\qquad E_{n,l}=\left(2n+l+\frac{3}{2}\right)\hbar\omega \tag{4}$$

由于角动量量子数 l 只决定离心势能 $\frac{\hbar^2}{2m}\frac{l(l+1)}{r^2}$ 的大小，所以 l 可以取任意非负整数．径向波函数为．令 $x=r/\beta$

$$R_{n,l}(r)=\frac{1}{\beta^{3/2}\pi^{1/4}}\sqrt{\frac{2^{n+l+2}n!}{(2n+2l+1)!!}}x^lL_n^{l+1/2}(x^2)\mathrm{e}^{-x^2/2} \tag{5}$$

其中 $L_n^{l+1/2}$ 是广义拉盖尔多项式[??].

前几个束缚态为（简并 $\deg = \sum(2l+1)$ ）

- $E = 3\hbar\omega/2$ （$\deg = 1$ ）

$$R_{0,0}(r) = \frac{1}{\beta^{3/2}\pi^{1/4}} 2\mathrm{e}^{-\frac{1}{2}x^2} \tag{6}$$

- $E = 5\hbar\omega/2$ （$\deg = 3$ ）

$$R_{0,1}(r) = \frac{1}{\beta^{3/2}\pi^{1/4}} \frac{2\sqrt{6}}{3} x\mathrm{e}^{-\frac{1}{2}x^2} \tag{7}$$

- $E = 7\hbar\omega/2$ （$\deg = 6$ ）

$$R_{0,2}(r) = \frac{1}{\beta^{3/2}\pi^{1/4}} \frac{4}{\sqrt{15}} x^2\mathrm{e}^{-x^2/2} \tag{8}$$

$$R_{1,0}(r) = \frac{1}{\beta^{3/2}\pi^{1/4}} \frac{2\sqrt{6}}{3} \left(\frac{3}{2} - x^2\right)\mathrm{e}^{-x^2/2} \tag{9}$$

- $E = 9\hbar\omega/2$ （$\deg = 8$ ）

$$R_{0,3}(r) = \frac{1}{\beta^{3/2}\pi^{1/4}} 4\sqrt{\frac{2}{105}} x^3\mathrm{e}^{-x^2/2} \tag{10}$$

$$R_{1,1}(r) = \frac{1}{\beta^{3/2}\pi^{1/4}} \frac{4}{\sqrt{15}} \left(\frac{5}{2} - x^2\right) x\mathrm{e}^{-x^2/2} \tag{11}$$

球面散射态与平面散射态的转换

预备知识 平面波的球谐展开[767]

我们已经知道没有势能时，每个正能量（即每个 k 值）的本征态存在无穷维的简并，我们既可以取无穷多离散的球面波作为简并空间的基底，也可以取无穷多不同方向的平面波作为简并空间的基底，两组基底展开同一个本征值为 $E = k^2/2$ 的子空间.

现在我们来讨论存在短程有心力的情况，令短程力的势能函数为 $V(r)$ 且满足短程条件

$$\lim_{r\to\infty} r^2 V(r) = 0 \tag{1}$$

之所以这么定义短程力，是因为当 r 很大时，$V(r)$ 相比与径向方程中的离心势能项可以忽略不计.

这时如果要求解球面波的散射态，可以取一个较大的 r_0 把 r 分成两部分，$r < r_0$ 的部分一般没有解析解，我们可以用数值方法求解. $r > r_0$ 部分的通解是两类球贝塞尔函数的线性组合，即 $Akrn_l(kr) + Bkrj_l(kr)$. 解出第一部分以后，可以在 r_0 将两部分波函数匹配，使函数值和一阶导数连续，解得 A, B 系数. 然后求出相移 $\delta_l(k)$，使[9]

$$Akrj_l(kr) - Bkry_l(kr) \to Ckrj_l[kr + \delta_l(k)] \tag{2}$$

即

$$\delta_l(k) = \arctan\frac{B}{A} \tag{3}$$

要归一化，我们只需令 $C = \sqrt{2/\pi}$. 要证明带有相移的径向波函数满足归一化条件，我们只需证明微小相移不影响归一化积分的结果

$$\begin{aligned}
&\int_0^\infty \sin(kr)\sin[k'r + \delta'(k)(k'-k)]\,\mathrm{d}r \\
=&\int_0^\infty \sin(kr)\sin(k'r)\cos[\delta'(k)\,\mathrm{d}k]\,\mathrm{d}r \\
=&\int_0^\infty \sin(kr)\sin(k'r)\,\mathrm{d}r = \frac{\pi}{2}\delta(k-k')
\end{aligned} \tag{4}$$

归一化以后，我们就得到了球面波形式的散射态.

我们希望能找到平面波对应的非束缚本征态，即当 $r \to \infty$ 的时候我们仍然希望看到平面波（想象平面水波遇到一个石头，只会产生局部的扰动）.

（说明未完成，没时间了，直接上公式吧）

边界条件为[10]

$$\psi_{\mathbf{k}}(\mathbf{r}) \to \frac{1}{(2\pi)^{3/2}}\left[\mathrm{e}^{\mathbf{k}\cdot\mathbf{r}} + f(\hat{\mathbf{r}})\frac{\mathrm{e}^{\mathrm{i}kr}}{r}\right] \tag{5}$$

[9]我们在第二类贝塞尔函数 $y_l(kr)$ 前面加上负号使其渐进式具有 $\cos(kr - l\pi/2)$ 的形式

[10]满足该边界条件的波函数貌似也满足归一化条件 $\delta(\mathbf{k}-\mathbf{k}')$.

其中 $f(\hat{\mathbf{r}})$ 是一个只与方向有关的函数，叫做**散射幅**．散射幅的模方就是微分截面

$$\frac{\partial\sigma}{\partial\Omega}=\frac{|\mathbf{j}_{out}(\mathbf{r})|r^2}{|\mathbf{j}_{in}(\mathbf{r})|}=|f(\hat{\mathbf{r}})|^2 \tag{6}$$

将各项球谐展开，令

$$f(\hat{\mathbf{r}})=\sum_{l,m}B_{l,m}Y_{l,m}(\hat{\mathbf{r}}) \tag{7}$$

对比每个球谐项的径向波函数得

$$C_{l,m}\sqrt{\frac{2}{\pi}}\sin[kr-l\pi/2+\delta_l(k)]=A_{l,m}\sqrt{\frac{2}{\pi}}\sin(kr-l\pi/2)+B_{lm}\frac{1}{(2\pi)^{3/2}}\mathrm{e}^{\mathrm{i}kr} \tag{8}$$

其中 $A_{l,m}$ 是已知的平面波的球谐展开系数（式 3 [767]），解得另外两个系数为（分别对比 $\exp(\mathrm{i}kr)$ 和 $\exp(-\mathrm{i}kr)$ 分量的系数）

$$C_{l,m}=\mathrm{e}^{\mathrm{i}\delta_l(k)}A_{l,m}=\frac{\mathrm{i}^l}{k}\mathrm{e}^{\mathrm{i}\delta_l(k)}Y^*_{lm}(\hat{\mathbf{k}}) \tag{9}$$

$$B_{l,m}=4\pi\sin\delta_l(k)\mathrm{e}^{-\mathrm{i}l\pi/2}C_{l,m} \tag{10}$$

库仑散射（量子）

预备知识 量子散射[1420]

[11]本文使用原子单位制[1315]．**库仑波函数（Coulomb wave function）**是类氢原子的散射态波函数．是库仑势能 $V(r)=Z/r$ （$Z<0$）下定态薛定谔方程[1328] 对能量 $E>0$ 的解．

库仑波函数常用的边界条件有两种，以波数 $\mathbf{k}$ 的平面波入射（球面波出射）的库仑波函数记为 $\psi^{(+)}_{\mathbf{k}}(\mathbf{r})$，平面波出射（球面波入射）的散射态记为 $\psi^{(-)}_{\mathbf{k}}(\mathbf{r})$．第一种边界条用于计算散射

$$\psi^{(\pm)}_{\mathbf{k}}(\mathbf{r})\to\frac{1}{(2\pi)^{3/2}}\left\{\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}+\mathrm{i}\eta\ln(kr-\mathbf{k}\cdot\mathbf{r})}+f(\hat{\mathbf{k}})\frac{\exp[\mathrm{i}kr-\mathrm{i}\eta\ln(2kr)]}{r}\right\}\qquad(\mathbf{r}\cdot\mathbf{k}\to\mp\infty) \tag{1}$$

[11]参考资料一个是 Wikipedia，另一个是“F Morales et al 2016 J. Phys. B: At. Mol. Opt. Phys. 49 245001”的附录．据说 [17] 也有．

在抛物线坐标系[124] 中可以解得[12]满足该边界的库仑波函数的解析式为

$$\psi_{\mathbf{k}}^{(\pm)}(\mathbf{r}) = \frac{1}{(2\pi)^{3/2}}\Gamma(1 \pm \mathrm{i}\eta)\mathrm{e}^{-\pi\eta/2}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}{}_1F_1(\mp\mathrm{i}\eta; 1; \pm\mathrm{i}kr - \mathrm{i}\mathbf{k}\cdot\mathbf{r}) \tag{2}$$

其中 $\eta = mZ/k$ 叫做 **Sommerfeld 参数**，$\rho = kr$，k 是能量为 $E = k^2/(2m)$ 的平面波的波数，Z 是原子核和电子的电荷之积（特别注意 $Z < 0$）. 上式满足

$$\psi_{\mathbf{k}}^{(+)}(\mathbf{r}) = \psi_{-\mathbf{k}}^{(-)}(r)^* \tag{3}$$

$$\langle\psi_{\mathbf{k}}^{(\pm)}|\psi_{\mathbf{k}'}^{(\pm)}\rangle = \delta(\mathbf{k} - \mathbf{k}') \tag{4}$$

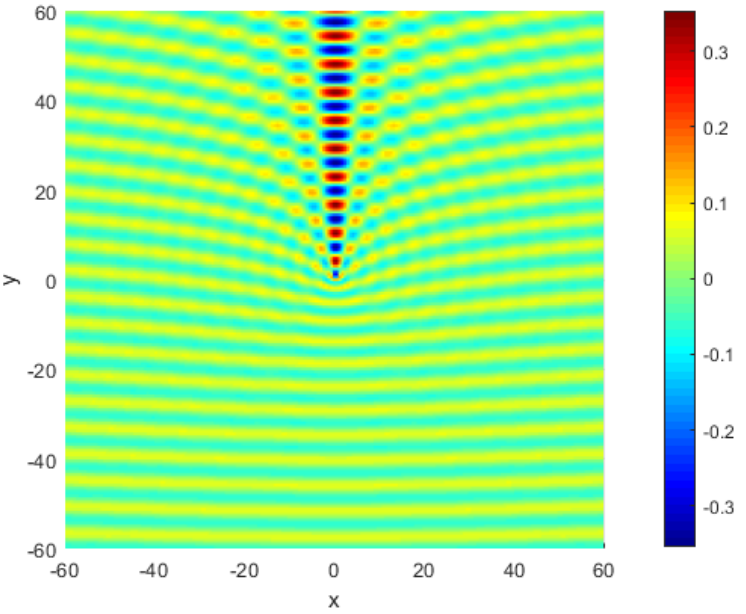


图 1: $\psi_{\mathbf{k}}^{(+)}(\mathbf{r})$ 的实部，$Z = -5, \mathbf{k} = \mathbf{y}$，$\eta = -5$. 当 $y \to -\infty$ 时趋近于平面波. 从经典力学考虑，粒子越靠近原子核入射偏折越多. 注意左右两边的平面波偏折后会相交并产生干涉.

另一种边界条件是在球坐标中用分离变量解薛定谔方程，本征态为

$$\psi_{l,m}(\mathbf{r}) = \frac{1}{r}\sqrt{\frac{2}{\pi}}F_l(\eta, kr)Y_{l,m}(\hat{\mathbf{r}}) \tag{5}$$

正交归一条件为

$$\langle\psi_{k',l',m'}|\psi_{k,l,m}\rangle = \delta_{l,l'}\delta_{m,m'}\delta(k' - k) \tag{6}$$

[12]推导见 [19].

球坐标中的库仑波函数

预备知识 薛定谔径向方程[1409]

球坐标中的库仑波函数与上面“库仑平面波”之间的关系，可以类比平面波与球面波的关系[767]．即 k 相同的所有方向的平面波基底张成的空间与球面波基底张成的空间相同，即能量为 $E=k^2/(2m)$ 的无穷维子空间．

库仑势的薛定谔径向方程（式 10 [1409]）为

$$-\frac{1}{2m}\frac{\mathrm{d}^2u}{\mathrm{d}r^2}+\left[\frac{Z}{r}+\frac{l(l+1)}{2mr^2}\right]u=\frac{k^2}{2m}u \tag{7}$$

其中 $u(r)$ 是 Scaled 波函数，l 是角量子数，Z 是原子核电荷与电子电荷的乘积（注意小于零）．方括号中的等效势能如图 2．

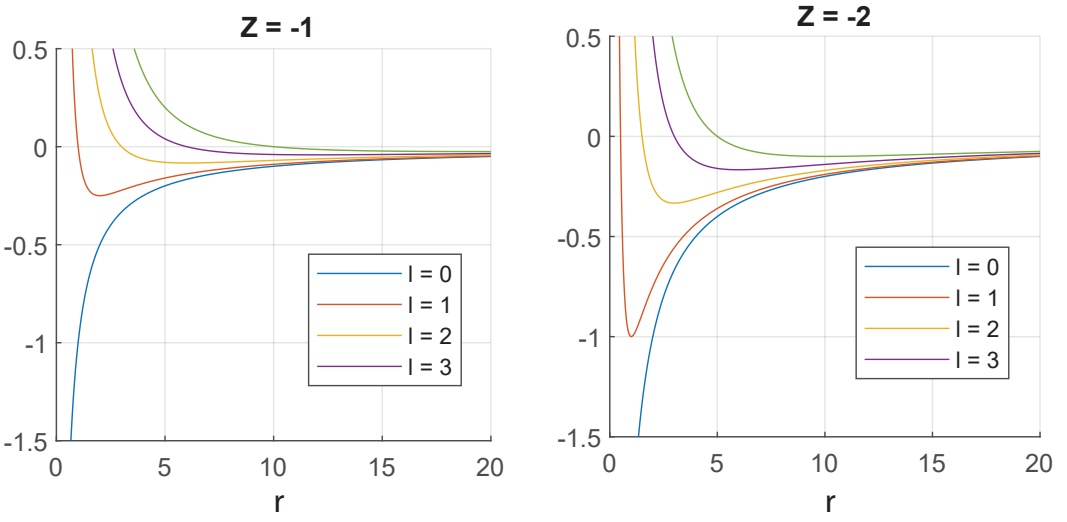


图 2: 球坐标中的径向等效势能

使用 $\eta=mZ/k$ 以及 $\rho=kr$ 换元，得

$$\frac{\mathrm{d}^2u}{\mathrm{d}\rho^2}+\left[1-\frac{2\eta}{\rho}-\frac{l(l+1)}{\rho^2}\right]u=0 \tag{8}$$

两个线性无关解为**第一类库仑函数** $F_l(\eta,\rho)$ 和**第二类库仑函数** $G_l(\eta,\rho)$[778]．散射态中，我们假设 $k>0$（$E>0$）．但理论上我们也可以取 k 为正虚数（$E<0$），这样当 E 为束缚态能量时就可以得到束缚态的径向波函数（式 3 [1536]）．

可以证明完备正交归一的**库仑球面波**为

$$|C_{l,m}(k)\rangle = \frac{1}{r}\sqrt{\frac{2}{\pi}}F_l(\eta, kr)Y_{l,m}(\hat{\mathbf{r}}) \tag{9}$$

满足

$$\langle C_{l',m'}(k')|C_{l,m}(k)\rangle = \delta_{l,l'}\delta_{m,m'}\delta(k-k') \tag{10}$$

现在可以将第一种边界条件的库仑波函数（式 1）展开为

$$\psi_{\mathbf{k}}^{(\pm)}(\mathbf{r}) = \sum_{l,m} a_{l,m}^{(\pm)}(\mathbf{k})\,|C_{l,m}(k)\rangle \tag{11}$$

其中

$$a_{l,m}^{(\pm)}(\mathbf{k}) = \frac{\mathrm{i}^l}{k}\exp[\pm\mathrm{i}\sigma_l(\eta)]Y_{l,m}^*(\hat{\mathbf{k}}) \tag{12}$$

σ_l 是库仑相移（式 8 [778]）. 注意 $a_{l,m}$ 并不是 $\langle C_{l,m}(k)|\psi_{\mathbf{k}}^{(\pm)}\rangle$，后者模长为无穷大，也可记为 $a_{l,m}\delta(0)$.

与平面波的情况（傅里叶变换）类似，要将一个球谐展开的波函数 $|f\rangle$ 投影到库仑波函数上，就先投影到库仑球面波上，然后进行幺正变换

$$\langle\psi_{\mathbf{k}}^{(\pm)}|f\rangle = \sum_{l,m} a_{l,m}^{(\pm)}(\mathbf{k})^*\,\langle C_{l,m}(k)|f\rangle \tag{13}$$

若 $|f\rangle$ 的 scaled 径向波函数为 $u_{l,m}(r)$，则

$$\langle\psi_{\mathbf{k}}^{(\pm)}|f\rangle = \frac{1}{k}\sum_{l,m} g_{l,m}(k)Y_{l,m}(\hat{\mathbf{k}}) \tag{14}$$

其中

$$g_{l,m}(k) = \sqrt{\frac{2}{\pi}}\mathrm{i}^{-l}\mathrm{e}^{\mp\mathrm{i}\sigma_l}\int_0^\infty F_l(\eta, kr)u_{l,m}(r)\,\mathrm{d}r \tag{15}$$

将波函数做傅里叶变换和投影到库仑波函数有什么区别呢？例如做氢原子电离的 TDSE，若想求 $t=+\infty$ 时的动量分布，理论上只要在 t 足够大时做傅里叶变换即可，但如果 t 不够大（电场已消失），电离波包所受的库仑力还不可忽略，那么虽然得到了瞬时的动量分布，但却与 $t=+\infty$ 的不同. 这时因为动量算符与哈密顿算符不对易，动量不守恒. 但若在电场消失的时候，投影到 $\psi_{\mathbf{k}}^{(-)}$ 上，由于它是哈密顿的本征函数，投影的模方（概率）不会随时间改变.

量子力学的基本假设

- 本词条处于草稿阶段.

预备知识 哈密顿正则方程[1075]

一维的情况

- 粒子的状态由 Hilbert 空间中的波函数 $|\psi(t)\rangle$ 表示.
- 要得到一个物理量 $\omega(x,p)$ 对应的算符，就把 x 和 p 换成对应的算符 $\hat{x}$ 和 $\hat{p}$. 这两个算符的定义为

$$\langle x|\,\hat{x}\,|x'\rangle = x\delta(x-x') \tag{1}$$

$$\langle x|\,\hat{p}\,|x'\rangle = -\mathrm{i}\hbar\delta'(x-x') \tag{2}$$

- 找到哈密顿量对应的哈密顿算符，波函数的演化由薛定谔方程决定

$$\hat{H}\,|\psi(t)\rangle = \mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}\,|\psi(t)\rangle \tag{3}$$

- 如果要测量一个物理量 $\hat{\Omega}(x,p)$，那么先求出所有的归一化本征函数 $|\omega_i\rangle$ 和对应的本征值 ω_i，测量到 ω_i 的概率为 $|\langle\omega|\psi\rangle|^2$. 测量完之后波函数由 $|\psi\rangle$ 变为 $|\omega_i\rangle$.

注意这里的 x 和 p 分别是哈密顿方程中的广义坐标和广义动量，而不必是位置和动量.

希尔伯特空间既包括可以正常归一化的波函数，也包括能用狄拉克 δ 函数归一化的波函数.

如果经典哈密顿量中出现了 xp 项，那么算符要写成 $(\hat{x}\hat{p}+\hat{p}\hat{x})/2$ 以保证物理量的算符是 Hermitian 矩阵. 如果出现了 x 和 p 的更高次项，就只能靠实验判断.

本词条参考 [16].

电磁场中的薛定谔方程及规范变换

预备知识 点电荷的拉格朗日和哈密顿量[1247]，量子化[1435]，原子单位制[1315]

本文使用原子单位制[1315]．电动力学中，电磁场中单个粒子的哈密顿量为

$$H=\frac{1}{2m}(\mathbf{p}-q\mathbf{A})^2+q\varphi \tag{1}$$

其中 φ 和 $\mathbf{A}$ 分别是电磁场的标势和矢势[1213]，都是位置 $\mathbf{r}$ 和时间的函数．$\mathbf{p}$ 是 $\mathbf{r}$ 的广义动量[1075]，

$$\mathbf{p}=m\mathbf{v}+q\mathbf{A} \tag{2}$$

将 $\mathbf{p}=-\mathrm{i}\hbar\boldsymbol{\nabla}$，代入得量子化后的哈密顿算符为

$$\begin{aligned}H&=\frac{\mathbf{p}^2}{2m}-\frac{q}{2m}(\mathbf{A}\cdot\mathbf{p}+\mathbf{p}\cdot\mathbf{A})+\frac{q^2}{2m}\mathbf{A}^2+q\varphi\\&=-\frac{1}{2m}\nabla^2+\mathrm{i}\frac{q}{2m}(\mathbf{A}\cdot\boldsymbol{\nabla}+\boldsymbol{\nabla}\cdot\mathbf{A})+\frac{q^2}{2m}\mathbf{A}^2+q\varphi\end{aligned} \tag{3}$$

注意算符 $\boldsymbol{\nabla}\cdot\mathbf{A}$ 是指先把波函数乘以矢势再取散度而不是直接对 $\mathbf{A}$ 取散度（想想量子力学中算符相乘的定义）．

另外要注意 $\mathbf{p}=-\mathrm{i}\boldsymbol{\nabla}$ 代表的是式 2 的**广义动量**而不是 $m\mathbf{v}$．所以一般规范下的平面波 $\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r})$ 的能量是

$$E=\frac{(\mathbf{k}-q\mathbf{A})^2}{2m} \tag{4}$$

在长度规范[1437] 下，$\mathbf{A}\equiv\mathbf{0}$，这时才有常见的 $E=\mathbf{k}^2/(2m)$．

如果对电磁场进行规范变换（式 3 [1217]）

$$\mathbf{A}=\mathbf{A}'+\boldsymbol{\nabla}\chi\qquad\varphi=\varphi'-\frac{\partial\chi}{\partial t} \tag{5}$$

为了仍然满足含时薛定谔方程，波函数需要乘以一个相位因子，但物理观测结果却不会改变

$$\Psi(\mathbf{r},t)=\exp(\mathrm{i}q\chi)\Psi'(\mathbf{r},t) \tag{6}$$

其中 χ 是式 3 [1217] 中的任意标量函数 λ．将以上三式代入含时薛定谔方程并化简后，等于把 $\mathbf{A},\varphi,\Psi$ 分别直接加上一瞥撇．所以对于任意规范，式 3 都保持相同的形式．

常见的规范如库仑规范[1437]，以及偶极子近似[??] 下的长度规范[1437] 和速度规范[1438].

库仑规范（量子）

预备知识 电磁场中的单粒子薛定谔方程[1436]

库仑规范[1219] 推导起来较为简单，用角标 C 表示库仑规范，则矢势满足 $\boldsymbol{\nabla}\cdot\mathbf{A}_C=0$，标势 φ_C 满足（式 6 [1219]）

$$\varphi(\mathbf{r},t)=\frac{1}{4\pi\epsilon_0}\int\frac{\rho(\mathbf{r}',t)}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}^3r' \tag{1}$$

结合矢量算符法则式 3 [304] 得

$$\boldsymbol{\nabla}\cdot(\mathbf{A}_C\Psi)=(\boldsymbol{\nabla}\cdot\mathbf{A}_C)\Psi+\mathbf{A}_C\cdot(\boldsymbol{\nabla}\Psi)=\mathbf{A}_C\cdot(\boldsymbol{\nabla}\Psi) \tag{2}$$

所以式 3 [1436] 简化为

$$H=-\frac{1}{2m}\nabla^2+\mathrm{i}\frac{q}{m}\mathbf{A}_C\cdot\boldsymbol{\nabla}+\frac{q^2}{2m}\mathbf{A}_C^2+q\varphi_C \tag{3}$$

长度规范

- 本词条存在未完成的内容.

预备知识 电磁场中的薛定谔方程及规范变换[1436]，偶极子近似（量子）[??]

本文使用原子单位制[1315].我们只在使用偶极子近似时讨论**长度规范(length gauge)**，因为我们接下来需要 $\mathbf{A}$ 与位置无关．当空间中存在静止的电荷分布时，我们可以把标量势能分为 $V+\varphi$ 两部分．前者由静止电荷根据库伦定律计算，不参与规范变换，在这里我们甚至可以不把它看成电磁力而只是某种其他势能．令不含时哈密顿算符为

$$H_0=\frac{\mathbf{p}^2}{2m}+qV \tag{1}$$

其中 $\mathbf{p}$ 是所选规范下的广义动量算符（式 2 [1436]）

$$\mathbf{p} = m\mathbf{v} + q\mathbf{A} = -\mathrm{i}\boldsymbol{\nabla} \tag{2}$$

未完成：以上这段论述应该放在偶极子近似里面

我们令 $\mathbf{A}_C, \varphi_C, \Psi_C$ 代表库仑规范，$\mathbf{A}_L, \varphi_L, \Psi_L$ 代表长度规范．将后者代入与规范无关的哈密顿量（式 3 [1436]）得

$$H_L = H_0 - \frac{q}{2m}(\mathbf{A}_L \cdot \mathbf{p} + \mathbf{p} \cdot \mathbf{A}_L) + \frac{q^2}{2m}\mathbf{A}_L^2 + q\varphi_L \tag{3}$$

长度规范的思路是：如果使用某种 $\mathbf{A}' = 0$ 的规范，就可以简化该式．用不带撇的变量表示库仑规范，我们令

$$\Psi_C(\mathbf{r}, t) = \exp(\mathrm{i}q\chi_L)\Psi_L(\mathbf{r}, t) \tag{4}$$

$$\chi_L(\mathbf{r}, t) = \mathbf{A}_C(t) \cdot \mathbf{r} \tag{5}$$

利用规范变换为（式 5 [1436]），以及 $-\partial\mathbf{A}_C/\partial t = \boldsymbol{\mathcal{E}}$，（$\boldsymbol{\mathcal{E}}(t)$ 是除静电场以外的含时电场）（库伦规范 + 偶极子近似，引用未完成）以及 $\varphi_C = 0$

$$\mathbf{A}_L = \mathbf{A}_C - \boldsymbol{\nabla}\chi_L = \mathbf{0} \tag{6}$$

可见**长度规范下矢势为零**，广义动量（式 2 [1436]）变为普通动量

$$\mathbf{p}_L = m\mathbf{v} = -\mathrm{i}\boldsymbol{\nabla} \tag{7}$$

再看标势的变换：

$$\varphi_L(t) = \varphi_C + \frac{\partial\chi_L}{\partial t} = -\boldsymbol{\mathcal{E}}(t) \cdot \mathbf{r} \tag{8}$$

由于形式不变，把式 6 和式 8 代入式 3 得长度规范下的哈密顿算符为

$$H_L = H_0 - q\boldsymbol{\mathcal{E}}(t) \cdot \mathbf{r} \tag{9}$$

薛定谔方程为

$$H_L\Psi^L = \mathrm{i}\frac{\partial}{\partial t}\Psi^L \tag{10}$$

速度规范

预备知识 长度规范[1437]，库仑规范（量子）[1437]

本文使用原子单位制[1315]．和长度规范中的思路一样，我们只在使用偶极子近似时讨论**速度规范（velocity gauge）**．用角标 V 表示速度规范，先从规范不变的哈密顿算符（式 3 [1437]）出发

$$H_V = H_0 - \frac{q}{2m}(\mathbf{A}_V \cdot \mathbf{p} + \mathbf{p} \cdot \mathbf{A}_V) + \frac{q^2}{2m}\mathbf{A}_V^2 + q\varphi_V \tag{1}$$

长度规范的思路是把上式中的 $\mathbf{A}_V^2$ 消去．对库仑规范使用规范变换（式 5 [1436]）

$$\Psi_C(\mathbf{r}, t) = \exp(\mathrm{i}q\chi_V)\Psi_V(\mathbf{r}, t) \tag{2}$$

$$\chi_V(t) = -\frac{q}{2m}\int_{-\infty}^{t} \mathbf{A}_C^2(t')\,\mathrm{d}t' \tag{3}$$

得

$$\mathbf{A}_V = \mathbf{A}_C - \boldsymbol{\nabla}\chi_V = \mathbf{A}_C \tag{4}$$

可见**速度规范下的矢势和库仑规范的相同**，以下统一记为 $\mathbf{A}_C$．这使得广义动量（式 2 [1436]）也和库伦规范的相同，

$$\mathbf{p}_V = \mathbf{p}_C = m\mathbf{v} + q\mathbf{A}_C = -\mathrm{i}\boldsymbol{\nabla} \tag{5}$$

再看标势的变换：

$$\varphi_V = \varphi_C + \frac{\partial \chi_V}{\partial t} = -\frac{q}{2m}\mathbf{A}_C^2 \tag{6}$$

式 4 和式 6 带入式 1 可以消去 $\mathbf{A}_V^2$ 项得

$$H_V = H_0 - \frac{q}{m}\mathbf{A} \cdot \mathbf{p} \tag{7}$$

薛定谔方程为

$$H_V\Psi_V = \mathrm{i}\frac{\partial}{\partial t}\Psi_V \tag{8}$$

对比式 2 和式 4 [1437] 得长度规范与速度规范中的波函数转换关系为

$$\Psi_V = \exp[\mathrm{i}q(\chi_L - \chi_V)]\Psi_L = \exp\left[\mathrm{i}q\mathbf{A} \cdot \mathbf{r} + \mathrm{i}\frac{q^2}{2m}\int_{-\infty}^{t} \mathbf{A}^2(t')\,\mathrm{d}t'\right]\Psi_L \tag{9}$$

加速度规范

预备知识 速度规范[1438]，平移算符[1400]

本文使用原子单位制[1315]．首先注意**加速度规范（acceleration gauge）**并不是一种规范而只是薛定谔方程的一种变换，说它是规范只是习惯上的叫法．该变换也叫做 **Kramers-Henneberger 变换**或 **K-H 变换**．

在速度规范[1438] 下，电场和矢势满足和库仑规范同样的关系（式 7 [1219]）

$$\boldsymbol{\mathcal{E}}(t) = -\frac{\partial \mathbf{A}}{\partial t} \tag{1}$$

注意速度规范默认偶极子近似[??]，即空间中电场和矢势都与位置无关．一个电荷为 q 的粒子在电磁波到来之前处于静止，那么接下来它在电磁波作用下的位移为

$$\boldsymbol{\alpha}(t) = -\frac{q}{m}\int_{-\infty}^{t} \mathbf{A}(t')\,\mathrm{d}t' \tag{2}$$

波函数变换是一个位移为 $\boldsymbol{\alpha}$ 的平移

$$\Psi_V(\mathbf{r},t) = \Psi_A(\mathbf{r}-\boldsymbol{\alpha},t) \tag{3}$$

也可以用平移算符[1400] 记为

$$\Psi_V(\mathbf{r},t) = \mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{p}}\Psi_A(\mathbf{r},t) \tag{4}$$

这相当于一个参考系变换，我们把 Ψ_A 所在的参考系叫做 **K-H 参考系（K-H frame）**，是一个非惯性系．注意严格来说式 4 要求波函数在整个空间无穷阶可导，而式 3 却不用，但习惯上我们只是把平移算符看成平移的一种记号，并不要求波函数无穷阶可导．

为什么说 K-H 变换不是一个规范变换？因为如果我们如果试图找到式 6 [1436] 中的 χ，会发现 $\chi = -\boldsymbol{\alpha}\cdot\mathbf{p}/q$，而这是一个微分算符，不是位置和时间的函数．

可以证明 K-H 系中，哈密顿量变为

$$H_A = \frac{\mathbf{p}^2}{2m} + V[\mathbf{r}+\boldsymbol{\alpha}(t)] \tag{5}$$

其中 $V(\mathbf{r})$ 是不含时的势能函数（例如原子核对电子的库仑势能）．我们可以把加速度规范想象成在 K-H 非惯性系中使用速度规范，仍有

$$\mathbf{p} = m\mathbf{v} + q\mathbf{A} = -\mathrm{i}\boldsymbol{\nabla} \tag{6}$$

但这里的 $\mathbf{v}$ 是相对于 K-H 系的.

注意这里不存在表示电场力的项，我们可以理解为 K-H 参考系中的惯性力[920] 与电场力抵消了.

对应的薛定谔方程仍然是

$$H_A\Psi_A = \mathrm{i}\frac{\partial}{\partial t}\Psi_A \tag{7}$$

加速度规范的一个简单应用见 Volkov 波函数[1441].

Volkov 波函数

- 本词条存在未完成的内容.

预备知识 加速度规范[1439]

本文使用原子单位制[1315]. 若初始时刻 t_0 一个自由粒子的波函数是平面波，波数为 $\mathbf{k}$. 接下来空间中出现了随时间变化的电场 $\boldsymbol{\mathcal{E}}(t)$（我们采用偶极子近似[??]，即假设电场不随位置变化，感生磁场可以忽略），那么我们把该电场中的含时波函数称为 **Volkov 波函数**.

以下我们分别在长度规范[1437]、速度规范[1438]、加速度规范[1439] 中求解 Volkov 波函数. 在 $t \leq t_0$ 时，令 $\boldsymbol{\mathcal{E}}(t) = \mathbf{A}(t) = \mathbf{0}$. 令三种规范下 $t \leq t_0$ 的波函数都是平面波（不同规范仅当非零电场出现后才有区别）

$$\Psi_{\mathbf{k}}(\mathbf{r}, t \leq t_0) = (2\pi)^{-3/2} \exp\left(\mathrm{i}\mathbf{k}\cdot\mathbf{r} - \mathrm{i}Et\right) \tag{1}$$

其中 $E = \mathbf{k}^2/(2m)$ 是初始动能（下同）. 注意以下所有规范中，算符 $\mathbf{p} = -\mathrm{i}\boldsymbol{\nabla}$ 都是广义动量[1436].

加速度规范

求解 Volkov 波函数最容易的方法就是使用加速度规范，由于空间中没有净电荷，薛定谔方程（式 7 [1439]）中 $V(\mathbf{r}) = 0$

$$\frac{\mathbf{p}^2}{2m}\Psi^A_{\mathbf{k}} = \mathrm{i}\frac{\partial}{\partial t}\Psi^A_{\mathbf{k}} \tag{2}$$

显然式 1 就是该方程的解

$$\Psi_{\mathbf{k}}^{A}(\mathbf{r},t)=(2\pi)^{-3/2}\exp\left(\mathrm{i}\mathbf{k}\cdot\mathbf{r}-\mathrm{i}Et\right) \tag{3}$$

可见在 K-H 参考系中，波函数始终保持平面波的形式.

例 1　高斯波包与电磁场

在 K-H 参考系中，若使用偶极子近似且令势能函数 $V(\mathbf{r})=0$，我们会发现高斯始终是高斯波包. 电磁波消失以后，K-H 系变为原来的惯性系，这是因为电场不能含有直流分量.（未完成：讲详细点?）

速度规范

使用式 3 [1439] 对式 3 做规范变换，得速度规范下的 Volkov 波函数为

$$\Psi_{\mathbf{k}}^{V}(\mathbf{r},t)=(2\pi)^{-3/2}\exp\left\{\mathrm{i}\mathbf{k}\cdot[\mathbf{r}-\boldsymbol{\alpha}(t)]-\mathrm{i}Et\right\} \tag{4}$$

其中 $\boldsymbol{\alpha}(t)$ 对应的是一个经典点电荷在电场中的位移（式 2 [1439]）

$$\boldsymbol{\alpha}(t)=-\frac{q}{m}\int_{t_0}^{t}\mathbf{A}(t')\,\mathrm{d}t' \tag{5}$$

容易验证式 4 是以下薛定谔方程（见式 8 [1438]）的解

$$\mathrm{i}\frac{\partial}{\partial t}\Psi^{V}=\left[\frac{\mathbf{p}^2}{2m}-\frac{q}{m}\mathbf{A}(t)\cdot\mathbf{p}\right]\Psi^{V} \tag{6}$$

注意 E 只是初始时的动能，电场中的粒子能量不守恒. 任意时刻，波函数都是 $m\mathbf{v}$ 和动能 E_k 的本征矢，本征值和经典粒子的动量动能相同.

$$m\mathbf{v}(t)=\mathbf{k}-q\mathbf{A}(t) \tag{7}$$

$$E_k(t)=\frac{[\mathbf{k}-q\mathbf{A}(t)]^2}{2m} \tag{8}$$

初始时刻有 $m\mathbf{v}(t_0)=\mathbf{k}$.

长度规范

要求长度规范长度规范下的 Volkov 波函数只需要对式 4 再次做规范变换即可（式 9 [1438]），得[13]

$$\Psi_{\mathbf{k}}^{L}(\mathbf{r},t)=(2\pi)^{-3/2}\exp\left[\mathrm{i}\mathbf{p}(t)\cdot\mathbf{r}-\frac{\mathrm{i}}{2m}\int_{t_0}^{t}\mathbf{p}(t')^2\,\mathrm{d}t'-\mathrm{i}Et_0\right] \tag{9}$$

[13] 如果不在乎波函数的全局相位，可以把式 9 方括号最后的常数项去掉.

其中 $\mathbf{p}(t)$ 就是的经典粒子的动量（式 7）. 注意本文中的 $\mathbf{A}(t)$ 都是指库仑规范或速度规范中的矢势[14]，长度规范下广义动量就是 $m\mathbf{v}$（式 7 [1437]）.

容易验证式 9 是以下薛定谔方程的解（式 10 [1437]）

$$\left[\frac{\mathbf{p}^2}{2m} - q\boldsymbol{\mathcal{E}}(t)\cdot\mathbf{r}\right]\Psi^L = \mathrm{i}\frac{\partial}{\partial t}\Psi^L \tag{10}$$

电磁场中的类氢原子

- 本词条处于草稿阶段.

预备知识 长度规范和速度规范[1438]

[15]假设原子核不动（无限核质量近似），使用库仑规范得

$$\varphi = \frac{Z}{r} \tag{1}$$

其中 $Z<0$ 是核电荷数.

式？变为

$$H = H_0 + H_I = -\frac{1}{2m}\nabla^2 + \frac{Z}{r} - \frac{\mathrm{i}}{m}\mathbf{A}\cdot\boldsymbol{\nabla} + \frac{1}{2m}\mathbf{A}^2 \tag{2}$$

其中含时哈密顿算符 H_I 是后两项.

以下我们使用偶极近似，则 $\mathbf{A}$ 只是时间的函数.

跃迁概率（一阶微扰）

- 本词条处于草稿阶段.

[14]长度规范下的矢势恒为零（式 6 [1437]）.

[15]本词条使用原子单位

预备知识 电磁波包的频谱[1232]，含时微扰理论[1414]

本文使用原子单位制[1315]．一阶微扰理论就是单光子电离，即两能级之间的能量等于单个光子的能量．

长度规范下的微扰跃迁理论

含时微扰理论（式 22 [1414]）为

$$c_i(+\infty) = -\mathrm{i}\int_{-\infty}^{+\infty} \langle i|H'(t)|j\rangle\, \mathrm{e}^{\mathrm{i}\omega_{ij}t}\,\mathrm{d}t \tag{1}$$

长度规范[1437] 中，

$$H'(t) = -q\boldsymbol{\mathcal{E}}(t)\cdot\mathbf{r} \tag{2}$$

$\boldsymbol{\mathcal{E}}$ 只是 t 的函数，可以分离

$$\langle i|H'(t)|j\rangle = -q\boldsymbol{\mathcal{E}}(t)\cdot\langle i|\mathbf{r}|j\rangle \tag{3}$$

令电场的傅里叶变换（式 1 [237]）为

$$\tilde{\boldsymbol{\mathcal{E}}}(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \boldsymbol{\mathcal{E}}(t)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t \tag{4}$$

则式 3 代入式 1 得

$$c_i(+\infty) = \mathrm{i}q\,\langle i|\mathbf{r}|j\rangle\cdot\int_{-\infty}^{+\infty}\boldsymbol{\mathcal{E}}(t)\mathrm{e}^{\mathrm{i}\omega_{ij}t}\,\mathrm{d}t = \mathrm{i}\sqrt{2\pi}q\,\langle i|\mathbf{r}|j\rangle\cdot\tilde{\boldsymbol{\mathcal{E}}}(-\omega_{ij}) \tag{5}$$

令 $\tilde{\boldsymbol{\mathcal{E}}}(\omega) = \tilde{\mathcal{E}}(\omega)\hat{\mathbf{e}}$，跃迁概率为

$$P_{j\to i} = |c_i(+\infty)|^2 = 2\pi q^2|\hat{\mathbf{e}}\cdot\langle i|\mathbf{r}|j\rangle|^2\left|\tilde{\mathcal{E}}(\omega_{ij})\right|^2 \tag{6}$$

结合（式 4 [1232]）

$$s(\omega) = 2c\epsilon_0\left|\tilde{\mathcal{E}}(\omega)\right|^2 \tag{7}$$

得

$$P_{j\to i} = \frac{\pi q^2}{c\epsilon_0}|\hat{\mathbf{e}}\cdot\langle i|\mathbf{r}|j\rangle|^2 s(\omega_{ij}) \tag{8}$$

对于束缚态 $|i\rangle$，$P_{j\to i}$ 是从 $|j\rangle$ 跃迁到 $|i\rangle$ 的概率；而对于连续态的 $|i\rangle$（如原子电离），若 $|i\rangle$ 对应出射方向的渐进动量 $\mathbf{k}$，那么 $P_{j\to i}$ 是 $\mathbf{k}$ 空间的三维概率密度分布函数[1419]．

另外注意只有电磁波包中频率为 ω_{ij} 的平面波分量对跃迁有贡献，所以我们也可以直接将能量差 ω_{ij} 替换为 ω．

对氢原子的具体计算见式 3 [1447]．

速度规范下的微扰跃迁理论

[16]速度规范[1438] 中，

$$H'(t) = -\frac{q}{m}\mathbf{A}\cdot\mathbf{p} = \frac{\mathrm{i}q}{m}\mathbf{A}\cdot\boldsymbol{\nabla} \tag{9}$$

$\mathbf{A}$ 只是 t 的函数，可以分离

$$\langle i|H'(t)|j\rangle = -\frac{q}{m}\mathbf{A}(t)\cdot\langle i|\mathbf{p}|j\rangle \tag{10}$$

令矢势的傅里叶变换为

$$\tilde{\mathbf{A}}(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\mathbf{A}(t)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t \tag{11}$$

代入式 1 得

$$c_i(t) = \frac{\mathrm{i}q}{m}\langle i|\mathbf{p}|j\rangle\cdot\int_{-\infty}^{+\infty}\mathbf{A}(t)\mathrm{e}^{\mathrm{i}\omega_{ij}t}\,\mathrm{d}t = \mathrm{i}\sqrt{2\pi}\frac{q}{m}\langle i|\mathbf{p}|j\rangle\cdot\tilde{\mathbf{A}}(-\omega_{ij}) \tag{12}$$

令 $\tilde{\mathbf{A}}(\omega) = \tilde{A}(\omega)\hat{\mathbf{e}}$，则跃迁概率为

$$P_{j\to i} = |c_i(t)|^2 = \frac{2\pi q^2}{m^2}|\hat{\mathbf{e}}\cdot\langle i|\mathbf{p}|j\rangle|^2\left|\tilde{A}(\omega_{ij})\right|^2 \tag{13}$$

结合波包的频谱公式（式 8 [1232] 变为原子单位）

$$s(\omega) = \frac{c}{2\pi}\omega^2\left|\tilde{A}(\omega_{ij})\right|^2 \tag{14}$$

$$P_{j\to i} = \frac{4\pi^2q^2}{cm^2\omega_{ij}^2}|\hat{\mathbf{e}}\cdot\langle i|\mathbf{p}|j\rangle|^2 s(\omega_{ij}) \tag{15}$$

对于束缚态 $|i\rangle$，$P_{j\to i}$ 是从 $|j\rangle$ 跃迁到 $|i\rangle$ 的概率；而对于连续态的 $|i\rangle$（如原子电离），若 $|i\rangle$ 对应出射方向的渐进动量 $\mathbf{k}$，那么 $P_{j\to i}$ 是 $\mathbf{k}$ 空间的三维概率密度分布函数[1419].

另外注意只有电磁波包中频率为 ω_{ij} 的平面波分量对跃迁有贡献，所以我们也可以直接将能量差 ω_{ij} 替换为 ω.

[16]参考 [17] 含时微扰相关章节.

两种规范比较

注意 $|i\rangle, |j\rangle$ 是没有电磁场时的能量本征态，波函数与规范无关. 把式 2 [1451] 和式 7 [1232] 带入式 12 可以证明两种规范等效（式 5 等于式 12）. 但是如果例如 $|i\rangle$ 是平面波，则不同规范结果不同.

单电子跃迁截面（一阶微扰）

- 本词条处于草稿阶段.

预备知识 微分截面[1420]，跃迁概率（一阶微扰）[1443]

束缚态到束缚态

从式 15 [1443] 来看，若电子从束缚态 $|j\rangle$ 跃迁到另一个束缚态 $|i\rangle$，跃迁概率（即吸收能量）和光谱中的 ω_{ij} 的强度成正比. 所以自然地，可以把**束缚态之间的跃迁截面**定义为

$$E = \omega_{ij} P_{j\to i} = \sigma_{j\to i} s(\omega_{ij}) \tag{1}$$

得

$$\sigma_{j\to i} = \frac{4\pi^2 q^2}{cm^2\omega_{ij}} |\hat{\mathbf{e}} \cdot \langle i|\mathbf{p}|j\rangle|^2 \tag{2}$$

束缚态到连续态

对于频率为 ω 的平面电磁波[1230]，截面可以想象成与电磁波传播方向垂直放置的一块面积为 $\sigma(\omega)$ 的面元，使得原子从电磁波中吸收的平均功率恰好等于电磁波经过该面元的功率. 对于波包，单位频率下原子吸收的能量为

$$\frac{\mathrm{d}E}{\mathrm{d}\omega} = \sigma(\omega) s(\omega) \tag{3}$$

如果再对立体角微分得某频率的微分截面

$$\frac{\partial^2 E}{\partial\omega\partial\Omega} = \frac{\partial\sigma}{\partial\Omega} s(\omega) \tag{4}$$

当波包经过以后，$\mathbf{A}=0$．把 ω 看作 k 的函数 $\omega(\mathbf{k})=k^2/(2m)+I_0$，$-I_0$ 是束缚态 $|j\rangle$ 的能量，上式积分得总能量

$$E=\iint \frac{\partial\sigma}{\partial\Omega}s(\omega)\,\mathrm{d}\omega\,\mathrm{d}\Omega=\iint \frac{\partial\sigma}{\partial\Omega}s(\omega)\frac{k}{m}\,\mathrm{d}k\,\mathrm{d}\Omega \tag{5}$$

另一方面，又有

$$E=\iint \omega P_{j\to\mathbf{k}}k^2\,\mathrm{d}k\,\mathrm{d}\Omega \tag{6}$$

未完成：$P_{j\to i}$ 是什么？引用和说明未完成

其中 ω 是光子能量．对比以上两式可以把微分截面表示为 $\mathbf{k}$ 的函数，也可以看成是单位矢量 $\hat{\mathbf{k}}$ 和 ω 的函数（$\mathbf{k}$ 的模长由 ω 决定）

$$\frac{\partial\sigma}{\partial\Omega}=\frac{km\omega}{s(\omega)}P_{j\to\mathbf{k}} \tag{7}$$

其中 $\omega(\mathbf{k})=k^2/(2m)+I_0$，下同．

长度规范下有（式 8 [1443]）

$$\frac{\partial\sigma}{\partial\Omega}=\frac{4\pi^2m\omega kq^2}{c}|\langle\mathbf{k}|\hat{\mathbf{e}}\cdot\mathbf{r}|j\rangle|^2 \tag{8}$$

速度规范下有（式 15 [1443]）

$$\frac{\partial\sigma}{\partial\Omega}=\frac{4\pi^2kq^2}{cm\omega}|\langle\mathbf{k}|\hat{\mathbf{e}}\cdot\mathbf{p}|j\rangle|^2 \tag{9}$$

氢原子电离计算（一阶微扰）

- 本词条处于草稿阶段．

预备知识 跃迁概率（含时微扰）[1443]

本文使用原子单位制[1315]．

长度规范

归一化的平面波和归一化的氢原子基态为

$$|\mathbf{k}\rangle = (2\pi)^{-3/2}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \qquad |0\rangle = \frac{1}{\sqrt{\pi}}\mathrm{e}^{-r} \tag{1}$$

长度规范下的跃迁偶极子，可以在极坐标系中积分

$$\langle\mathbf{k}|\mathbf{r}|0\rangle = \frac{\hat{\mathbf{k}}}{(2\pi)^{3/2}\sqrt{\pi}}\int_0^{+\infty}\int_0^{\pi}\mathrm{e}^{-r}\mathrm{e}^{-\mathrm{i}kr\cos\theta}r\cos\theta\cdot 2\pi r^2\sin\theta\,\mathrm{d}\theta\,\mathrm{d}r \tag{2}$$

换元，令 $u=\cos\theta$，得[17]

$$\begin{aligned}\langle\mathbf{k}|\mathbf{r}|0\rangle &= \frac{1}{\sqrt{2}\pi}\hat{\mathbf{k}}\int_0^{+\infty}r^3\mathrm{e}^{-r}\int_{-1}^{1}\mathrm{e}^{-\mathrm{i}kru}u\,\mathrm{d}u\cdot\mathrm{d}r\\ &= \mathrm{i}\frac{\sqrt{2}\hat{\mathbf{k}}}{k\pi}\int_0^{+\infty}r^2\mathrm{e}^{-r}\left[\cos(kr)-\frac{1}{kr}\sin(kr)\right]\mathrm{d}r\\ &= -\mathrm{i}\frac{8\sqrt{2}}{\pi}\frac{\mathbf{k}}{(k^2+1)^3}\end{aligned} \tag{3}$$

未完成：代入未完成

速度规范

注意一阶微扰理论中的初态和末态波函数都是无微扰（无外场）情况下的，与规范无关．要计算 $\langle\mathbf{k}|\boldsymbol{\nabla}|0\rangle$，先看积分

$$\int\exp(-\mathrm{i}\mathbf{k}\cdot\mathbf{r})\exp(-r)\,\mathrm{d}^3r = \frac{8\pi}{(k^2+1)^2} \tag{4}$$

使用算符 $\boldsymbol{\nabla}$ 的反厄米性得

$$\begin{aligned}&\int\exp(-\mathrm{i}\mathbf{k}\cdot\mathbf{r})\boldsymbol{\nabla}\exp(-r)\,\mathrm{d}^3r = -\int[\boldsymbol{\nabla}\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r})]^*\exp(-r)\,\mathrm{d}^3r\\ &= \mathrm{i}\mathbf{k}\int\exp(-\mathrm{i}\mathbf{k}\cdot\mathbf{r})\exp(-r)\,\mathrm{d}^3r = \mathrm{i}\frac{8\pi\mathbf{k}}{(k^2+1)^2}\end{aligned} \tag{5}$$

[17]最后一步可通过 Wolfram Alpha 获得

乘以归一化系数得

$$\langle \mathbf{k}|\boldsymbol{\nabla}|0\rangle = \mathrm{i}\frac{2\sqrt{2}}{\pi}\frac{\mathbf{k}}{(k^2+1)^2} \tag{6}$$

该式代入式 9 [1446]（$q=-1$）得微分截面为

$$\frac{\partial\sigma}{\partial\Omega} = \frac{32}{mc\omega}\frac{k(\mathbf{k}\cdot\hat{\mathbf{e}})^2}{(k^2+1)^4} = \frac{64}{mc}\frac{k(\mathbf{k}\cdot\hat{\mathbf{e}})^2}{(k^2+1)^5} \tag{7}$$

对于质子数为 Z 类氢原子有

$$\frac{\partial\sigma}{\partial\Omega} = \frac{32Z^5}{mc\omega}\frac{k(\mathbf{k}\cdot\hat{\mathbf{e}})^2}{(k^2+Z^2)^4} \tag{8}$$

两种规范对比

如果 $|\mathbf{k}\rangle$ 是库仑函数（能量本征态）应该有（式 2 [1451]）

$$\langle \mathbf{k}|\mathbf{r}|0\rangle = -\frac{\langle \mathbf{k}|\boldsymbol{\nabla}|0\rangle}{m\omega_{k0}} \tag{9}$$

其中 $\omega_{k0} = k^2/2 + 1/2$，但实际上式 3 和式 6 满足

$$\langle \mathbf{k}|\mathbf{r}|0\rangle = -2\frac{\langle \mathbf{k}|\boldsymbol{\nabla}|0\rangle}{m\omega_{k0}} \tag{10}$$

这说明在使用平面波近似库伦函数时，长度规范的 transition amplitude 恰好是速度规范的 2 倍，截面是四倍（待求证）.

教材中推导微分截面一般使用速度规范，因为速度规范的结果与实验吻合更好.

氢原子的选择定则

- 本词条处于草稿阶段.

预备知识 3j 符号[855]，球谐函数[761]，含时微扰理论[1414]

氢原子的**选择定则（selection rule）**是指在哪些情况下**跃迁偶极子矩阵元（transition dipole matrix element）** $\langle\psi_{n,l,m}|\mathbf{r}|\psi_{n',l',m'}\rangle$ 为零．该矩阵在含时微扰理论中出现，如果矩阵元为零，说明在一阶微扰近似下 $|\psi_{n',l',m'}\rangle$ 不能在电场哈密顿量为 $\boldsymbol{\mathcal{E}}\cdot\mathbf{r}$（长度规范[1437]）的作用下跃迁到 $|\psi_{n,l,m}\rangle$．但即使矩阵元为零，仍有可能通过多阶微扰仍然存在耦合（即矩阵的 N 次方中该矩阵元不为零）．从物理上来看就是先从初态跃迁到中间态，再从中间态跃迁到末态．如果高阶微扰的跃迁也被禁止（即矩阵的任意次方中该矩阵元都为零），那么就是绝对禁止的．

利用 3j 符号的对称性推导

预备知识 3j 符号[855]

相比于算符对易的方法[18]，3j 符号的好处是不仅能得到选择定则，还可以直接算出偶极子矩阵元角向积分的具体值而无需手动积分[19]．长度规范[1437] 中电场哈密顿量为 $\boldsymbol{\mathcal{E}}\cdot\mathbf{r}$．其中 $\mathbf{r}$ 的三个分量可以用球谐函数表示为（式 9 [761]）

$$x=\sqrt{\frac{2\pi}{3}}r(Y_{1,-1}-Y_{1,1})\qquad y=\mathrm{i}\sqrt{\frac{2\pi}{3}}r(Y_{1,-1}+Y_{1,1})\qquad z=\sqrt{\frac{4\pi}{3}}rY_{1,0}\tag{1}$$

所以

$$\begin{aligned}&\langle\psi_{n,l,m}|\boldsymbol{\mathcal{E}}\cdot\mathbf{r}|\psi_{n',l',m'}\rangle=\boldsymbol{\mathcal{E}}\cdot\langle\psi_{n,l,m}|\mathbf{r}|\psi_{n',l',m'}\rangle\\&=\int R_{n,l}(r)rR_{n',l'}(r)r^2\,\mathrm{d}r\times\\&\Big[\sqrt{\frac{2\pi}{3}}E_x(\langle Y_{l,m}|Y_{1,-1}|Y_{l',m'}\rangle-\langle Y_{l,m}|Y_{1,1}|Y_{l',m'}\rangle)\\&+\mathrm{i}\sqrt{\frac{2\pi}{3}}E_y(\langle Y_{l,m}|Y_{1,-1}|Y_{l',m'}\rangle+\langle Y_{l,m}|Y_{1,1}|Y_{l',m'}\rangle)\\&+\sqrt{\frac{4\pi}{3}}E_z\langle Y_{l,m}|Y_{1,0}|Y_{l',m'}\rangle\Big]\end{aligned}\tag{2}$$

其中三个球谐函数之积的积分为（式 13 [761]）

$$\langle Y_{l,m}|Y_{1,m_1}|Y_{l',m'}\rangle=(-1)^m\sqrt{\frac{3(2l+1)(2l'+1)}{4\pi}}\begin{pmatrix}l&1&l'\\0&0&0\end{pmatrix}\begin{pmatrix}l&1&l'\\-m&m_1&m'\end{pmatrix}\tag{3}$$

[18] 参考 [14].

[19] 当然，手动 3j 符号也比较繁琐，可以借助 Wolfram Alpha 或 Mathematica，Matlab 中我也写了一个程序（同样可以符号计算），还没放进百科．

令 $\Delta m = m - m'$，$\Delta l = l - l'$．使用 3j 的选择定则（式 6 [855]）得 $\Delta m = -m_1$．所以

$$\Delta m = \begin{cases} 0 & (\text{电场只延 } z \text{ 方向}) \\ 0, \pm 1 & (\text{其他方向电场}) \end{cases} \tag{4}$$

由三角约束（$|l - l'| \leqslant 1 \leqslant l + l'$）得 $\Delta l = 0, \pm 1$．但由式 8 [855] 得 $l + l' + 1$ 为奇数时式 3 为零，所以只能有

$$\Delta l = \pm 1 \tag{5}$$

这是两个常见的选择定则．

另外 3j 符号还有其他选择定则．即使符合所有常见的选择定则，偶极子矩阵元仍然有可能为零．找到所有 3j 符号（或 CG 系数）为零的情况是十分困难的．

未完成：这段可能有误，需核实；径向积分是否有选择定则？

物理意义

z 方向的电场不会改变电子 z 方向的角动量，所以 L_z 守恒[1377]，Δm 为 0．另外由于光子的角动量量子数为 $l = 1$ （链接未完成），所以 $\Delta l = 0, \pm 1$，但 $\Delta l = 0$ 仍然需要对称性来排除．

跃迁偶极子矩阵的三种形式

预备知识 长度规范[1437]

[20]本文使用原子单位制[1315]．在长度规范[1437] 下，对于某种势能 V 的束缚态 $|\psi_a\rangle$ 和 $|\psi_b\rangle$，可以证明以下三种形式的**跃迁偶极子矩阵（transition dipole matrix）**相等．

$$D^L_{ba} = q \langle \psi_b | \mathbf{r} | \psi_a \rangle \tag{1}$$

$$D^V_{ba} = -\frac{q}{m(E_b - E_a)} \langle \psi_b | \boldsymbol{\nabla} | \psi_a \rangle \tag{2}$$

[20]本文参考 [19]

$$D_{ba}^{A}=\frac{q}{m(E_b-E_a)^2}\langle\psi_b|\boldsymbol{\nabla}V|\psi_a\rangle \tag{3}$$

这三种形式分别称为偶极子矩阵的**长度形式（length form）**、**速度形式（velocity form）**和**加速度形式（acceleration form）**．注意他们都使用长度规范，不要和速度规范[1438] 和加速度规范混淆．

证明速度形式

$$H_0=-\frac{\nabla^2}{2m}+V(\mathbf{r}) \tag{4}$$

$$\langle\psi_b|\mathbf{r}|\psi_a\rangle=\frac{\langle H_0\psi_b|\mathbf{r}|\psi_a\rangle-\langle\psi_b|\mathbf{r}|H_0\psi_a\rangle}{E_b-E_a}=\frac{\langle\psi_b|H_0\mathbf{r}-\mathbf{r}H_0|\psi_a\rangle}{E_b-E_a} \tag{5}$$

其中

$$H_0\mathbf{r}-\mathbf{r}H_0=-\frac{1}{2m}(\nabla^2\mathbf{r}-\mathbf{r}\nabla^2) \tag{6}$$

注意这里的算符 $\nabla^2\mathbf{r}$ 作用在波函数上是指 $\nabla^2(\mathbf{r}\psi)$ 而不是 $(\nabla^2\mathbf{r})\psi$．由式 8 [304] 不难证明

$$\nabla^2(\mathbf{r}\psi)=2\boldsymbol{\nabla}\psi+\mathbf{r}\nabla^2\psi \tag{7}$$

所以

$$H_0\mathbf{r}-\mathbf{r}H_0=-\frac{1}{m}\boldsymbol{\nabla}\psi \tag{8}$$

代入式 5 得

$$\langle\psi_b|\mathbf{r}|\psi_a\rangle=-\frac{\langle\psi_b|\boldsymbol{\nabla}|\psi_a\rangle}{m(E_b-E_a)} \tag{9}$$

代入式 1 可得速度形式．证毕．

证明加速度形式

从速度形式出发，再次进行式 5 类似的操作

$$\langle\psi_b|\boldsymbol{\nabla}|\psi_a\rangle=\frac{\langle\psi_b|H_0\boldsymbol{\nabla}-\boldsymbol{\nabla}H_0|\psi_a\rangle}{E_b-E_a} \tag{10}$$

其中

$$H_0\boldsymbol{\nabla}-\boldsymbol{\nabla}H_0=V\boldsymbol{\nabla}-\boldsymbol{\nabla}V=-(\boldsymbol{\nabla}V) \tag{11}$$

代入式 10 再带入式 2 可得加速度形式，证毕．

康普顿散射

- 本词条缺少预备知识，初学者可能会遇到困难.

预备知识 光子，相对论动量，相对论能量

康普顿效应是射线与物质相互作用的三种效应之一. 康普顿效应是指入射光子与物质原子中的核外电子产生非弹性碰撞而被散射的过程. 碰撞时，入射光子把部分能量转移给电子，使它脱离原子成反冲电子，而散射光子的能量和运动方向发生变化. 如图 1 所示，其中 $h\nu$ 是入射光子的能量，$h\nu'$ 是散射光子的能量，θ 是散射光子的散射角，e 是反冲电子，ϕ 是反冲电子的反冲角. 由于

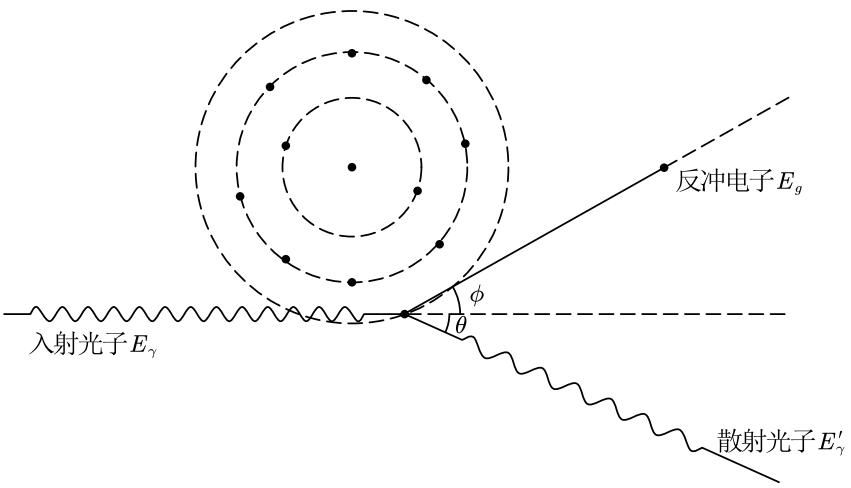


图 1: 康普顿散射

发生康普顿散射的 γ 光子能量比电子的束缚能要大的多，所以 γ 光子与原子中的电子相互作用时，可以把电子的束缚能忽略，看成是自由电子，并视为散射发生以前电子是静止的，动能为 0，只有静止能 m_0c^2. 散射后，电子获得速度 v，此时电子的能量

$$E = \frac{m_0c^2}{\sqrt{1-\beta^2}} \tag{1}$$

动量为

$$mv = \frac{m_0v}{\sqrt{1-\beta^2}} \tag{2}$$

其中 $\beta = v/c$，c 为光速.

由相对论的能量和动量守恒定律可以得到：

$$\begin{cases} m_0c^2 + h\nu & = \dfrac{m_0c^2}{\sqrt{1-\beta^2}} + h\nu' \\ \dfrac{h\nu}{c} & = \dfrac{h\nu'}{c}\cos\theta + \dfrac{m_0v}{\sqrt{1-\beta^2}}\cos\phi \\ \dfrac{h\nu'}{c}\sin\theta & = \dfrac{m_0v}{\sqrt{1-\beta^2}}\sin\theta \end{cases} \tag{3}$$

解方程式 3，可得

$$h\nu' = \frac{h\nu}{1 + \frac{h\nu}{m_0c^2}(1-\cos\theta)} \tag{4}$$

其中 $h\nu/c$ 是入射光子的动量，$h\nu'/c$ 是散射光子的动量，此式表示散射光子能量与入射光子能量及散射角的关系.

全同粒子

预备知识　多体量子力学简介，角动量加法（量子力学）[1403]，自旋角动量[1395]

在量子力学中，我们假设许多基本粒子是不可区分的，例如电子，质子，中子等，我们把不可区分的粒子称为是**全同粒子（identical particles）**. 例如我们在两个位置同时测量到两个电子，我们只能知道两处各有一个电子，而不知道哪个电子在哪里. 对于宏观物体，我们可以通过做标记或者追踪轨迹的方式轻易把不同它们区分开，但量子力学中的粒子既不能做标记也没有轨迹的概念，所以我们唯一可以区分的粒子的方法就是利用它们的物理属性，如质量，电荷量，自旋等物理性质.

波函数的对称性

先不考虑自旋[1395]，我们如何用它们的波函数体现“不可区分”呢？对于位置表象下的双粒子波函数 $\psi(\mathbf{r}_1, \mathbf{r}_2)$，在任意两点 $\mathbf{r}_1, \mathbf{r}_2$ 同时发现两个电子的概率密度函数为

$$f(\mathbf{r}_1, \mathbf{r}_2) = |\psi(\mathbf{r}_1, \mathbf{r}_2)|^2 \tag{1}$$

为了体现两粒子不可区分，我们要求概率分布满足 $f(\mathbf{r}_2,\mathbf{r}_1)=f(\mathbf{r}_1,\mathbf{r}_2)$. 于是波函数必须满足以下两个条件之一

$$\psi(\mathbf{r}_2,\mathbf{r}_1)=\pm\psi(\mathbf{r}_1,\mathbf{r}_2) \tag{2}$$

如果取正号，我们把这样的波函数称为**对称的（symmetric）**，取负号时则称为**反对称的（antisymmetric）**. 两个全同粒子的波函数只可能取这两种情况之一. 如果波函数既不是对称也不是反对称，那么就称为**不对称的（not symmetric）**. 注意对于全同粒子的波函数，$\mathbf{r}_1,\mathbf{r}_2$ 的角标不再代表粒子的编号，而仅仅用于区分两个不同位置.

以后会看到，具有交换对称或反对称的波函数在根据薛定谔方程演化的过程中一直会保持它的对称性.

自旋态的对称性

如果只考虑双粒子的自旋态矢量空间[444]，同样由于概率原因我们要求全同粒子同样在该空间的态矢满足交换对称.

例 1

以常见的双电子自旋为例. 单电子自旋维 $1/2$，自旋态在 2 维空间，两个基底通常记为 $\uparrow$ 和 $\downarrow$，分别代表自旋 z 分量 $m_s=\pm1/2$. 而双电子的自旋态是该空间和自己的张量积，是 $2\times2=4$ 维的，正交归一基底自然地可以取 $\uparrow\uparrow,\uparrow\downarrow,\downarrow\uparrow,\downarrow\downarrow$. 但中间两个既不不是对称也不是反对称的，所以我们可以将他们稍作变换，得到 4 个具有交换对称性的正交归一基底

$$\uparrow\uparrow \qquad \frac{1}{\sqrt{2}}(\uparrow\downarrow+\downarrow\uparrow) \qquad \downarrow\downarrow \tag{3}$$

$$\frac{1}{\sqrt{2}}(\uparrow\downarrow-\downarrow\uparrow) \tag{4}$$

其中式 3 的三个基底是对称的，即三重态（**triplet**）；式 4 的一个基底是反对称的，即单态（**singlet**）. 显然，三重态的任意线性组合都是对称的自旋态[21]，他们构成了 3 维的子空间[464]，而反对称的只有单态为基底的 1 维子空间.

但这里的对称性具体如何定义呢？我们可以用 $\chi_{1,2}$ 表示这个 4 维空间中的任意矢量，定义 $\chi_{2,1}$ 就是把 $\chi_{1,2}$ 的每项中两个箭头（单粒子态）互换，例如

[21] 当我们这么说时，默认还没有做归一化

↓↑ 变为 ↑↓ 又如 ↓↑ − ↑↓ 变为 ↑↓ − ↓↑. 于是（反）对称条件就是

$$\chi_{2,1} = \pm\chi_{1,2} \tag{5}$$

我们可以把 ↑↑ 和 ↓↓ 视为两电子自旋方向相同的态，$(\uparrow\downarrow \pm \downarrow\uparrow)/\sqrt{2}$ 视为两电子自旋方向相反的态. 对于后者，我们只能说“一个电子自旋为 ↑ 另一个为 ↓”却不能说哪个是 ↑ 哪个是 ↓. 这正体现了全同粒子的不可区分性. 相比之下，↑↓ 就表明第一个电子是 ↑，第二个是 ↓，所以不能对两个电子使用.

一般地，对于自旋量子数为 s 的两个全同粒子，单个粒子处于 $N = 2s+1$ 维矢量空间中，双粒子态空间就是 N^2 维的. 其中对称子空间是 $N(N+1)/2$ 维的，反对称子空间是 $N(N-1)/2$ 维的，相加恰好等于 N^2. 所以他们是两个互补的正交子空间[527]. 我们把推导留到“粒子交换算符[1458]”中.

玻色子和费米子

当我们考虑两个粒子的总状态时，我们可以把总空间视为双粒子波函数空间和双粒子自旋态空间的张量积空间. 双粒子态矢总可以分解为波函数和自旋的张量积的线性组合

$$|\Psi_{1,2}\rangle = \sum_i \psi_i(\mathbf{r}_1, \mathbf{r}_2)\chi_{1,2}^{(i)} \tag{6}$$

在我们往往讨论的只是可以记为一项的简单态

$$|\Psi_{1,2}\rangle = \psi(\mathbf{r}_1, \mathbf{r}_2)\chi_{1,2} \tag{7}$$

与式 2 同理，全同粒子需要满足

$$|\Psi_{1,2}\rangle = \pm|\Psi_{2,1}\rangle \tag{8}$$

对于式 7 来说，$|\Psi_{1,2}\rangle$ 是对称的当且仅当波函数和自旋具有相同对称性（都是对称或者都是反对称），$|\Psi_{1,2}\rangle$ 是反对称的当且仅当波函数和自旋具有相反对称性（一个对称一个反对称）.

波函数或自旋的两种交换对称性会带来实验上可观测的结果. 实验表明，我们可以把所有的基本粒子划分为两类，自旋为整数的称为**玻色子（boson）**，具有对称态矢；自旋为半整数的称为**费米子（fermion）**，具有反对称态矢. 这

是量子力学关于全同粒子的基本假设．至于为什么自旋可以决定波函数对称性，需要使用量子场论才能解释．

例如电子是自旋为 1/2 的费米子，为了保证总状态的反对称，如果两电子的自旋处于单态（反对称），那么波函数就必须是对称的；反之如果两电子处于三重态（对称），那么波函数必须是反对称的．

多个粒子的情况

以上我们只讨论了两个全同粒子，若有更多，那么费米子和玻色子分别要求系统的态矢在交换任意两个粒子时都保持对称和反对称．

泡利不相容原理

预备知识 全同粒子[1454]

[22]在量子力学中，**泡利不相容原理（Pauli exclusion principle）**指的是，在含有若干个全同费米子（即自旋为 1/2 的粒子）的系统中，任意两个费米子不能处于同一个态．

事实上费米子对态矢的反对称性要求已经包含了泡利不相容原理，所以该原理只是费米子性质的一个推论．若假设系统只有两个全同费米子，用矢量空间的语言来解释泡利不相容原理就是：双粒子态表示为张量积 $|\psi\rangle|\psi\rangle$ 时是一个交换对称的态矢，而费米子的态矢必须是反对称的，所以两个费米子不可能处于 $|\psi\rangle|\psi\rangle$ 形式的态．

泡利不相容原理常用于解释原子中的电子排布．

例 1　原子壳层

在原子壳层理论中，如果忽略电子之间的相互作用，那么单个电子具有一系列正交归一的能量本征态，其中能量最低的态叫做基态，具有唯一的波函数 $\psi_g(\mathbf{r})$．如果两个电子都处于基态，就意味着双电子波函数为

$$\psi(\mathbf{r}_1,\mathbf{r}_2)=\psi_g(\mathbf{r}_1)\psi_g(\mathbf{r}_2) \tag{1}$$

[22]参考 Wikipedia 相关词条 和 [14].

这个波函数显然是交换对称的.

但电子是费米子，即总态矢必须是反对称的，所以总自旋态必须是反对称的（见式 8 [1454] 的说明），所以如果双电子都处于基态，那么唯一可能的状态就是

$$\Psi(\mathbf{r}_1,\mathbf{r}_2)\chi_{1,2}=\frac{1}{\sqrt{2}}\psi_g(\mathbf{r}_1)\psi_g(\mathbf{r}_2)(\uparrow\downarrow-\downarrow\uparrow) \tag{2}$$

根据泡利不相容原理，$\psi_g(\mathbf{r}_1)\psi_g(\mathbf{r}_2)\uparrow\uparrow$ 和 $\psi_g(\mathbf{r}_1)\psi_g(\mathbf{r}_2)\downarrow\downarrow$ 态不可能存在是因为它们表示“两个电子处于同一个状态”.

粒子交换算符

- 本词条存在未完成的内容.

预备知识 全同粒子[1454]，厄米矩阵的本征问题[426]

对双粒子波函数，定义**粒子交换算符（particle exchange operator）**为

$$P_{1,2}\psi(\mathbf{r}_1,\mathbf{r}_2)=\psi(\mathbf{r}_2,\mathbf{r}_1) \tag{1}$$

可以证明这是一个厄米算符，即

$$\langle\phi|P_{1,2}\psi\rangle=\langle P_{1,2}\phi|\psi\rangle \tag{2}$$

不难证明该算符有且只有 1 和 -1 两个本征值，根据厄米算符本征方程的性质[426]，对应的本征矢空间是两个**互补的正交子空间**[527]，量空间分别由对称波函数和反对称波函数（以下分别记为 ψ_+ 和 ψ_-）的集合构成，即满足

$$\psi_\pm(\mathbf{r}_2,\mathbf{r}_1)=\pm\psi_\pm(\mathbf{r}_1,\mathbf{r}_2) \tag{3}$$

这两个子空间外的波函数既非对称也非反对称.

若交换算符作用在双粒子自旋态上，使用式 5 [1454] 的记号表示就是

$$P_{1,2}\chi_{1,2}=\chi_{2,1} \tag{4}$$

一般希尔伯特空间中的交换算符

那么一般来说我们如何定义双粒子态空间中的交换算符呢？首先我们要确定单粒子空间，并找到一组单粒子的基底 $\{|i\rangle\}$. 然后把两个单粒子空间做张量积得到双粒子的态空间，基底取 $\{|i\rangle\,|j\rangle\}$. 交换算符作用在某个双粒子态矢上，就先将其在基底 $\{|i\rangle\,|j\rangle\}$ 上展开，再将每个基底 $|i\rangle\,|j\rangle$ 替换为 $|j\rangle\,|i\rangle$ 即可：

$$P_{1,2}\sum_{i,j}C_{i,j}\,|i\rangle\,|j\rangle=\sum_{i,j}C_{i,j}\,|j\rangle\,|i\rangle=\sum_{i,j}C_{j,i}\,|i\rangle\,|j\rangle \tag{5}$$

连续本征矢

交换算符的定义也可以拓展到连续本征矢的情况. 例如在式 1 中，单粒子的位置本征态基底记为 $|\mathbf{r}\rangle$，张量积空间的基底记为 $|\mathbf{r}_1\rangle\,|\mathbf{r}_2\rangle$，那么式 5 中求和变为积分

$$P_{1,2}\int\psi(\mathbf{r}_1,\mathbf{r}_2)\,|\mathbf{r}_1\rangle\,|\mathbf{r}_2\rangle\,\mathrm{d}^3r_1\,\mathrm{d}^3r_2=\int\psi(\mathbf{r}_2,\mathbf{r}_1)\,|\mathbf{r}_1\rangle\,|\mathbf{r}_2\rangle\,\mathrm{d}^3r_1\,\mathrm{d}^3r_2 \tag{6}$$

多个粒子的情况

$N>2$ 个全同粒子波函数的交换算符也可以通过上述方法借助基底来定义. 注意这时不止一个，我们可以把交换粒子 i,j（$i\neq j$）的算符记为 $P_{i,j}$，共有 $C_N^2=n(n-1)/2$ 个. 例如三粒子空间中，

$$P_{2,3}\,|i\rangle\,|j\rangle\,|k\rangle=|i\rangle\,|k\rangle\,|j\rangle \tag{7}$$

这些算符中有一些是对易的（如 $P_{1,2}$ 和 $P_{3,4}$），一些是不对易的（如 $P_{1,2}$ 和 $P_{2,3}$）.

在多个全同粒子的态空间中，玻色子的态矢只能处于所有交换算符的对称子空间的交集中，这个交集仍然是矢量空间；费米子的态矢只能处于所有交换算符的反对称子空间的交空间中.

显然，同一个交换算符的乘积是单位算符：

$$P_{i,j}^2=1 \tag{8}$$

即同一个交换算符作用两次后，任意态矢都保持不变.

与哈密顿算符对易

预备知识 守恒量（量子力学）[1377]

对于全同粒子，交换算符与哈密顿算符对易[23]. 这保证了 P 是一个守恒量. 也就是全同粒子的波函数在按照薛定谔方程演化过程中将一直保持对称性或反对称性.

本征态与测量

两个全同粒子的本征态（或者其他任何态）也必须是必须是对称或反对称的，例如双粒子的位置本征态

$$\frac{1}{\sqrt{2}}[\delta_{\mathbf{r}_a}(\mathbf{r}_1)\delta_{\mathbf{r}_b}(\mathbf{r}_2) \pm \delta_{\mathbf{r}_b}(\mathbf{r}_1)\delta_{\mathbf{r}_a}(\mathbf{r}_2)]$$

只能告诉我们一个粒子在 $\mathbf{r}_a$ 处另一个粒子在 $\mathbf{r}_b$ 处，仍然不能区分它们.

于是根据测量理论，我们仍然要将总波函数（已满足对称或反对称）投影到本征态上去. 例如位置的概率分布为

$$\begin{aligned} f(\mathbf{r}_1,\mathbf{r}_2) &= \left|\frac{1}{\sqrt{2}}\int [\delta_{\mathbf{r}_1}(\mathbf{r}_1')\delta_{\mathbf{r}_2}(\mathbf{r}_2') \pm \delta_{\mathbf{r}_2}(\mathbf{r}_1')\delta_{\mathbf{r}_1}(\mathbf{r}_2')]\psi_{\pm}(\mathbf{r}_1',\mathbf{r}_2')\,\mathrm{d}^3r_1'\,\mathrm{d}^3r_2'\right|^2 \\ &= \left|\frac{1}{\sqrt{2}}\psi_{\pm}(\mathbf{r}_1,\mathbf{r}_2) \pm \frac{1}{\sqrt{2}}\psi_{\pm}(\mathbf{r}_2,\mathbf{r}_1)\right|^2 \\ &= \left|\frac{1}{\sqrt{2}}\psi_{\pm}(\mathbf{r}_1,\mathbf{r}_2) + \frac{1}{\sqrt{2}}\psi_{\pm}(\mathbf{r}_1,\mathbf{r}_2)\right|^2 \\ &= 2|\psi_{\pm}(\mathbf{r}_1,\mathbf{r}_2)|^2 \end{aligned} \tag{9}$$

注意该式中 $f(\mathbf{r}_1,\mathbf{r}_2)$ 不区分 $\mathbf{r}_1,\mathbf{r}_2$ 的顺序. 也就是说 $f(\mathbf{r}_1,\mathbf{r}_2)$ 告诉我们一个粒子在 $\mathbf{r}_1$，另一个在 $\mathbf{r}_2$ 的概率密度.

对应地，在做归一化时如果对全空间积分，在每一个 $f(\mathbf{r}_1,\mathbf{r}_2)$ 都重复计算了 $f(\mathbf{r}_2,\mathbf{r}_1)$. 所以最直接的解决的方法是对全空间积分后除以 2，使归一化条件为

$$\frac{1}{2}\int f(\mathbf{r}_1,\mathbf{r}_2)\,\mathrm{d}^3r_1\,\mathrm{d}^3r_2 = \int |\psi_{\pm}(\mathbf{r}_1,\mathbf{r}_2)|^2\,\mathrm{d}^3r_1\,\mathrm{d}^3r_2 = 1 \tag{10}$$

[23]对非全同粒子则不成立，例如两个质量不同的粒子的动能算符与交换算符不对易

另一种方法是不除 2，而是选取上式积分范围的一半使得对任何 $f(\mathbf{r}_1,\mathbf{r}_2)$，$f(\mathbf{r}_2,\mathbf{r}_1)$ 不会被重复计算，在一维运动情况下，我们可以只对 r_1,r_2 平面满足 $r_1<r_2$ 的一半积分．在多维运动时，我们只需要选取任意一个坐标，例如令 $y_1<y_2$ 即可将整个积分范围划分为满足要求的两半[24]．

例 1　单粒子位置分布

若两个全同粒子的波函数为 $\psi(\mathbf{r}_1,\mathbf{r}_2)$，求单个粒子的位置分布 $P(\mathbf{r})$．注意我们并不能求 $P(\mathbf{r}_1)$ 或者 $P(\mathbf{r}_2)$．

我们只需要将式 9 对其中一个 $\mathbf{r}$ 全空间积分即可

$$f(\mathbf{r})=\int f(\mathbf{r},\mathbf{r}')\,\mathrm{d}^3r'=2\int|\psi_\pm(\mathbf{r},\mathbf{r}')|^2\,\mathrm{d}^3r' \tag{11}$$

根据式 10，发现一个粒子的总概率是 $\int f(\mathbf{r})\,\mathrm{d}^3r=2$ 而不是 1．这并没有问题，因为空间中共有两个粒子．

一般投影

类比式 9 可以发现，将任何对称或反对称的双粒子波函数 $\psi(\mathbf{r}_1,\mathbf{r}_2)$ 投影到形式为

$$[\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\pm\phi_2(\mathbf{r}_1)\phi_1(\mathbf{r}_2)]/\sqrt{2} \tag{12}$$

且具有相同交换对称性的波函数上，都会得到相同的两项相加[25]．所以为了方便计算，我们可以只投影到第一项 $\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)$ 上再乘以 $\sqrt{2}$ 即可．

> 未完成：以下内容来自个人笔记，有待整理和审评

算符

现在来看上述的对称或反对称空间允许什么样的算符，即什么算符在这两个子空间是闭合的．闭合是指算符作用完后仍然落在同一空间，即

$$P_{1,2}\Omega\,|\Psi\rangle=\pm\Omega\,|\Psi\rangle \tag{13}$$

[24] 也可以使用任何可以区分粒子交换的条件，例如 $x_1^2+y_1^2+z_1^2<x_2^2+y_2^2+z_2^2$，将角标 1 和 2 互换后，两条不等式只能满足一条．

[25] 如果投影到对称性相反的波函数上，结果为零

即所有满足 $P_{1,2}\Omega = \pm\Omega$ 的算符.

首先,同一个单粒子算符对所有粒子作用之和满足该条件,例如总动能 $T = T_1 + T_2$. 满足 $P_{1,2}T = T$. 又例如外势能（非相互作用势能）$V = V(\mathbf{r}_1) + V(\mathbf{r}_2)$. 不允许单独出现单粒子算符，例如 $V(\mathbf{r}_1)$.

同理，不允许出现不同的算符分别作用于两粒子 $U_1 + V_2$.

从物理意义上来说，就是不允许出现任何可以区分两个全同粒子的算符. 因为全同粒子绝对不可区分，所以任何作用（算符）不可能对两个全同粒子具有不同的形式.

另一类允许的算符是多粒子算符，例如两粒子间的相互作用 $V(\mathbf{r}_1, \mathbf{r}_2) = V(|\mathbf{r}_2 - \mathbf{r}_1|)$.

可分离哈密顿

在对称或反对称子空间中，如果总哈密顿量可以写成各个粒子的哈密顿量之和 $H = \sum H_i$，(注意 H_i 对每个粒子的形式必须一样)，则可先计算单个粒子的本征方程 $H_i |\psi_{in}\rangle = E_{in} |\psi_{in}\rangle$ (对任何 i 形式都一样)，总能量为 $E_{n_1,n_2,\ldots} = \sum_i E_{n_i}$，总波函数为 Slater 行列式（费米子）以及对称化算符（玻色子），双粒子系统中即 $(|\psi_i\rangle |\psi_j\rangle \pm |\psi_j\rangle |\psi_i\rangle)/\sqrt{2}$.

双粒子系统的证明：$|\psi_i\rangle$ 是单粒子薛定谔方程的波函数，可区分双粒子的完备基底为 $|\psi_j\rangle |\psi_i\rangle$，但这些基底中只有 $|\psi_k\rangle |\psi_k\rangle$ 落在玻色子子空间，剩下的基底都不是 $P_{1,2}$ 的本征值. 为了构建对称和反对称子空间，先把所有 $i \neq j$ 的基底分为成对的 $|\psi_i\rangle |\psi_j\rangle$ $|\psi_j\rangle |\psi_i\rangle$，把每一对分别相加和相减再归一化得到许多对新的基底 $(|\psi_i\rangle |\psi_j\rangle \pm |\psi_j\rangle |\psi_i\rangle)/\sqrt{2}$ (显然这么做不失完备性)，不难看出，所有含加号的基底张成完备的双波色子空间，所有含减号的基底张成完备的双费米子空间. 这说明在双粒子系统中，全空间等于对称子空间和反对称子空间的直和（即两个子空间的基底一起张成全空间）. 然而对多粒子系统，这并不成立，但用 Slater 行列式和“正 Slater 行列式”得到的两组基底仍然是完备的. 继续双粒子的情况，现在证明上面得到的两个子空间的基底已经是总哈密顿量的本征矢了

$$\begin{aligned}(H_1 + H_2)(|\psi_i\rangle |\psi_j\rangle \pm |\psi_j\rangle |\psi_i\rangle) &= E_i |\psi_i\rangle |\psi_j\rangle \pm E_j |\psi_j\rangle |\psi_i\rangle + E_j |\psi_i\rangle |\psi_j\rangle \pm E_i |\psi_j\rangle |\psi_i\rangle \\ &= (E_i + E_j)(|\psi_i\rangle |\psi_j\rangle \pm |\psi_j\rangle |\psi_i\rangle)\end{aligned} \tag{14}$$

证毕. 多维的情况可以用对称和反对称算符的运算证明,（详见 Bransden）.

可分离态的时间演变

从物理意义上来看,可分离变量意味着粒子间没有相互作用,例如 $V(|\mathbf{r}_i,\mathbf{r}_j|)$ 项. 那么没有相互作用的两个粒子是否需要对称化或反对称化呢？如果需要，那么原则上我们将要把全宇宙所有的全同粒子（反）对称化！这是一定要的，因为量子力学对全同粒子的基本假设就是态矢只能落在对称空间或反对称空间，这两个空间之外的所有态矢不可能存在，它们违反了全同粒子假设！另外，$\mathbf{r}_1,\mathbf{r}_2$ ，这些角标只是符号工具，二次量子化中将彻底抛弃角标.

根据所有全同粒子对称化的原理,看两个粒子的波函数 $(|\psi_1\rangle\,|\psi_2\rangle\pm|\psi_2\rangle\,|\psi_1\rangle)/\sqrt{2}$，如果 $(|\psi_i\rangle$ 与 $|\psi_j\rangle$ 相距很远，波函数完全没有重合，时刻 t_1 在 $|\psi_1\rangle$ 的附近测到一个粒子，然后时刻 t_2 在 $|\psi_2\rangle$ 范围测到一个粒子，能不能说这是刚才的粒子呢？不能！只能说全宇宙有 N 个这种粒子，t_1 的时候其中一个在 $|\psi_1\rangle$ 附近，t_2 的时候其中一个在 ψ_2 附近.

对称化与反对称化

预备知识 粒子交换算符[1458]

如果给定一个多粒子波函数处于对称或反对称子空间之外，特定情况下我们可以将其**对称化（symmetrize）**或**反对称化（antisymmetrize）**. 先用粒子交换算符表示例 1 [1454] 的过程.

例 1

假设单个粒子态空间的一组单位正交基底是 $\{|i\rangle\}$，两个粒子某时刻的态矢可以用单个张量积表示 $|i\rangle\,|j\rangle$ 且. 若 $i=j$，则显然这个态已经是对称的. 若 $i\neq j$，则态矢是不对称的，我们可以通过乘以 $(1\pm P_{12})/\sqrt{2}$ 来对称或反对称化（分别取正号和负号）

$$\frac{1}{\sqrt{2}}(1\pm P_{12})\,|i\rangle\,|j\rangle=\frac{1}{\sqrt{2}}(|i\rangle\,|j\rangle\pm|j\rangle\,|i\rangle) \tag{1}$$

由于等式中的两项也是正交归一的，我们需要另乘归一化系数 $1/\sqrt{2}$. 可以验

证，该式就是 $P_{1,2}$ 的本征矢，本征值分别为 ± 1．这种方法利用了交换算符的性质式 8 [1458]．

更一般地，我们也可以不要求 $|i\rangle|j\rangle$ 正交，但这样式 1 中的两项也变得不正交，我们就要重新计算归一化系数了．

在该例中，我们把 $(1 \pm P_{12})/\sqrt{2}$ 称为**对称化算符（symmetrizer）**或**反对称化算符（antisymmetrizer）**．

多粒子对称化

若将例 1 中的两个粒子改为 N 个粒子，如何将状态 $|i_1\rangle|i_2\rangle\ldots|i_N\rangle$ （反）对称化呢？稍加思考会发现，若要对称化，我们只需要将 $|i_1\rangle, |i_2\rangle, \ldots, |i_N\rangle$ 的所有不同的排列相加再归一化即可．如果这 N 个单粒子态都是不同的，那么一共有 $N!$ 种排列．我们用 $p_n(i)$ 来表示，例如 则正交化的结果为

表 1: $N = 3$ 的 6 种排列

	$p_1(i)$	$p_2(i)$	$p_3(i)$	$p_4(i)$	$p_5(i)$	$p_6(i)$
$i = 1$	1	1	2	2	3	3
$i = 2$	2	3	1	3	1	2
$i = 3$	3	2	3	1	2	1

$$\frac{1}{\sqrt{N!}}\sum_{n=0}^{N!}|p_n(1)\rangle|p_n(2)\rangle\ldots|p_n(N)\rangle \tag{2}$$

为了验证这是一个对称态，我们可以用例如 $P_{i,j}$ 作用在上面，这相当于把表 1 中的第 i, j 行调换，容易得出这不会改变式 2．

如果 $|i_1\rangle, \ldots, |i_N\rangle$ 中出现重复，情况就要更复杂一些．令其中只有 $M < N$ 种不同的单粒子态，重复的次数分别是 $n_1, \ldots, n_M$，有 $\sum_i n_i = N$．这样一来不同的排列减少至 $N!/(n_1!n_2!\ldots n_M!)$ 种，我们仍然可以写出类似式 2 的表达式，但求和只有 $N!/(n_1!n_2!\ldots n_M!)$ 个正交归一的项，所以归一化系数也变为 $\sqrt{n_1!n_2!\ldots n_M!/N!}$．

对称化算符

要把对称化用算符表示出来也不难，我们可以对每一种不同的排列 $p_n(1), p_n(2), \ldots$ 都定义一个对称化算符 P_n. 定义反对称化算符为（求和的项数同样取决于是否出现重复）

$$\mathcal{S} = 1 + \sum_n P_n \tag{3}$$

这样对称化就可以优雅地表示为 $\mathcal{S}\,|i_1\rangle \ldots |i_N\rangle$.

多粒子反对称化

我们同样也可以使用排列算符对 N 粒子态 $|i_1\rangle\,|i_2\rangle \ldots |i_N\rangle$ 进行反对称化，但前提是我们必须要求其中 N 个单粒子态都是不同的. 这时因为如果 $|j\rangle = |k\rangle$，那么无论如何排列，得到的态关于交换算符 $P_{j,k}$ 都是对称的. 这就是泡利不相容原理[1457].

由式 1 的启发，我们可以尝试改变式 2 或式 3 中一些项的正负号来达到反对称化. 事实上行列式的定义[432] 中已经给出了我们需要规则，即使用逆序数的奇偶性来决定正负号. 对称化的结果可以用行列式记为

$$\begin{vmatrix} |i_1\rangle & |i_2\rangle & \ldots & |i_N\rangle \\ |i_1\rangle & |i_2\rangle & \ldots & |i_N\rangle \\ \ldots & \ldots & \ldots & \ldots \\ |i_1\rangle & |i_2\rangle & \ldots & |i_N\rangle \end{vmatrix} \tag{4}$$

这个行列式被称为**斯莱特行列式（Slater determinant）**.

要证明反对称性很简单，任意交换算符 $P_{i,j}$ 作用在式 4 , 相当于把行列式的两行置换，而根据行列式的性质（定理 3 [435]），这会使结果取相反数. 证毕.

通过在式 3 中根据 P_n 的逆序数在其前面适当添加正负号 S_n（见式 4 [432]），我们也可以定义反对称化算符

$$\mathcal{A} = 1 + \sum_n S_n P_n \tag{5}$$

子空间的维度

先来看有限维的情况，假设单个粒子态所在空间是 M 维的（例如只考虑自旋空间，$M = 2s+1$），那么只有当粒子数 $N \leq M$ 时，才能允许全同费米子

的态矢．通过基底的不同组合[106] 以及对称化，我们一共可以找到 C_M^N 种不同的反对称基底，所以[26]反对称子空间是 C_M^N 维的．要得到这 C_M^N 个基底，我们只需要从 M 个单粒子基底中无顺序不重复地选出 N 个（$|i_1\rangle, \ldots, |i_N\rangle$）．每选一次就可以得到一个对称基底 $\mathcal{A}\,|i_1\rangle \ldots |i_N\rangle$，它们张成整个反对称子空间．物理上，这意味着 N 个全同费米子占据 M 个状态，每种可能就是一个基底．

对于玻色子，我们不要求 $N \leq M$，因为多个玻色子可以处于同一个单粒子态．可以证明对称子空间的维数等于“从 M 个态中无序地选 N 个，允许重复”的个数，但遗憾的是我们不能把这个数写成一个简洁的公式．对称子空间基底的构建也同理，每次选取出 N 个态后都可以构建一个基底 $\mathcal{S}\,|i_1\rangle \ldots |i_N\rangle$．

密度矩阵

预备知识 矩阵的迹[410]，投影算符[799]，量子力学的基本假设[1435]

[27]若一个系综中的 N 个系统中，有 n_i（$i = 1, 2, \ldots, k$）个在状态 $|i\rangle$（这里假设 $|i\rangle$ 是正交归一的）．那么这个系综可以用**密度矩阵（density matrix）**（或算符）描述

$$\rho = \sum_i p_i \,|i\rangle\langle i| \tag{1}$$

其中 $p_i = n_i/N$ 是随机选一个系统，处于状态 $|i\rangle$ 的概率．若所有系统都处于同一个 $|i\rangle$，那么这个系综就是**纯的（pure）**，否则就是**混合的（mixed）**．

非纯态的等效

纯态 $|\psi\rangle$ 可以唯一地表示为密度矩阵 $\rho = |\psi\rangle\langle\psi|$，它的意义也相当明确．

然而对于非纯态，我们可以使用不同的正交归一的（纯态）基底的概率组合来表示，测量结果却是一样的．因为测量并不能区分量子概率和经典概率．

例如有一束电子．若这束电子是纯态的，我们就可以通过多次测量得到这个纯态（除了一个总体相位因子）．例如通过测量 $|y+\rangle$ 和 $|y-\rangle$ 的概率，我们可以确定 $|x+\rangle + c\,|x-\rangle$ 中的复数 c．

[26]这不是证明只是对结论的一个简单描述．

[27]参考 Shankar, Principles of Quantum Mechanics 2ed，以及 Wikipedia

但若这束电子的自旋方向是随机的，我们既可将其等效为随机的一半 $|x+\rangle$ 和 $|x-\rangle$ 构成的，也可以等效为随机的一半 $|y+\rangle$ 和 $|y-\rangle$ 构成的．虽然“实际”上它们是不一样的，但任何测量都无法区分这两种情况．所以密度矩阵可以是 $\rho = (|x+\rangle\langle x+| + |x-\rangle\langle x-|)/2$ 也可以是 $\rho = (|y+\rangle\langle y+| + |y-\rangle\langle y-|)/2$．事实上，它们都是单位矩阵的一半．注意如果概率不是均分的，如 $\rho = (|x+\rangle\langle x+|/3 + 2|x-\rangle\langle x-|)/3$ 就无法用 $|y\pm\rangle$ 基底表示（一般的对角矩阵经过相似变换后不是对角矩阵）．

纯态和非纯态都可以对应一个唯一密度矩阵．

测量

对于某个物理量对应的算符 Ω，它的**系综平均值（ensemble average）**为

$$\langle\bar{\Omega}\rangle = \sum_i p_i \langle i|\Omega|i\rangle \tag{2}$$

这个平均值既包含了每个 $|i\rangle$ 的平均，又包含了对每个系统的平均．

系综平均也可以用迹表示为 $\mathrm{tr}(\Omega\rho)$．根据迹的定义，

$$\mathrm{tr}(\Omega\rho) = \sum_j \langle j|\Omega\rho|j\rangle = \sum_{i,j} p_i \langle j|\Omega|i\rangle\langle i|j\rangle = \sum_i p_i \langle i|\Omega|i\rangle = \langle\bar{\Omega}\rangle \tag{3}$$

证毕．

对于纯态，获得测量值 ω 的概率可以看作投影算符 $\mathbb{P}_\omega$ 的平均值（满足 $\mathbb{P}_\omega^2 = \mathbb{P}_\omega$）

$$P(\omega) = |\langle\omega|\psi\rangle|^2 = \langle\mathbb{P}_\omega\psi|\mathbb{P}_\omega\psi\rangle = \langle\psi|\mathbb{P}_\omega|\psi\rangle \tag{4}$$

所以对于混合态，测量值 ω 的概率为

$$\overline{P(\omega)} = \mathrm{tr}(\mathbb{P}_\omega\rho) \tag{5}$$

密度矩阵的性质

- 密度矩阵算符是厄米算符（自伴算符）
- 密度矩阵的迹为 1

Hartree-Fork 方法

Hartree 方法的精髓是假设多粒子波函数 $\Psi(\mathbf{r}_1, \mathbf{r}_2 \dots \mathbf{r}_N)$ 等于每个单粒子波函数（包括自旋）的乘积 $u_1(q_1)u_2(q_2)\dots u_N(q_N)$（Hartree 函数），其中不同的单粒子波函数要求正交归一．然后找到最优的单粒子波函数 $u_1, u_2, \dots u_N$ 使总哈密顿的平均值最小．所以该方法属于变分法．得到的能级大于精确能级．

注意 Hartree 函数并不满足全同粒子的对易关系，既不是对称也不是反对称．对于全同费米子，方法是令多粒子函数为单粒子函数（包括自旋）的 Slater 行列式．

$$\Psi(q_1, q_2 \dots) = \frac{1}{\sqrt{N!}} \begin{vmatrix} u_1(q_1) & u_2(q_1) & \cdots \\ u_1(q_2) & u_2(q_2) & \cdots \\ \vdots & \vdots & \ddots \end{vmatrix} \tag{1}$$

变分法的拉格朗日乘数函数为

$$L = \langle \Psi | \hat{H} | \Psi \rangle - \sum_{ij} \varepsilon_{ij} \langle u_i | u_j \rangle \tag{2}$$

为了化简第二项，令原基底为另一组基底的幺正变换

$$\begin{pmatrix} u_1 \\ u_2 \\ \vdots \end{pmatrix} = \hat{U} \begin{pmatrix} u'_1 \\ u'_2 \\ \vdots \end{pmatrix} \tag{3}$$

可以证明第一项不变．因为

$$\begin{aligned} \Psi(q_1, q_2 \dots) &= \frac{1}{\sqrt{N!}} \begin{vmatrix} u_1(q_1) & u_2(q_1) & \cdots \\ u_1(q_2) & u_2(q_2) & \cdots \\ \vdots & \vdots & \ddots \end{vmatrix} = \frac{1}{\sqrt{N!}} \left| \mathbf{U} \begin{pmatrix} u'_1(q_1) & u'_1(q_2) & \cdots \\ u'_2(q_1) & u'_2(q_2) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \right| \\ &= \frac{|\mathbf{U}|}{\sqrt{N!}} \begin{vmatrix} u'_1(q_1) & u'_2(q_1) & \cdots \\ u'_1(q_2) & u'_2(q_2) & \cdots \\ \vdots & \vdots & \ddots \end{vmatrix} = \frac{\mathrm{e}^{\mathrm{i}\theta}}{\sqrt{N!}} \begin{vmatrix} u'_1(q_1) & u'_2(q_1) & \cdots \\ u'_1(q_2) & u'_2(q_2) & \cdots \\ \vdots & \vdots & \ddots \end{vmatrix} \end{aligned} \tag{4}$$

注意这里用到了 $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$ 和 $|\mathbf{U}| = \mathrm{e}^{\mathrm{i}\theta}$．这是说，平均能量关于 u 的公式不受幺正变换的影响．对第二项，

$$\sum_{ij} \varepsilon_{ij} \langle u_i | u_j \rangle = \int (u_1'^*, u_2'^* \dots) \mathbf{U}^\dagger \varepsilon \mathbf{U} \begin{pmatrix} u_1' \\ u_2' \\ \vdots \end{pmatrix} \mathrm{d}q \tag{5}$$

若我们选择幺正变换，使得 $\mathbf{U}^\dagger \boldsymbol{\varepsilon} \mathbf{U} = E_i \delta_{ij}$，即把矩阵 $\boldsymbol{\varepsilon}$ 对角化（下文可知 $\boldsymbol{\varepsilon}$ 为对称矩阵），得

$$\sum_{ij} \varepsilon_{ij} \langle u_i | u_j \rangle = \sum_i E_i \langle u_i | u_j \rangle \tag{6}$$

现在把所有的撇号省略，拉格朗日函数为

$$L = \langle \Psi | \hat{H} | \Psi \rangle - \sum_i E_i \langle u_i | u_i \rangle \tag{7}$$

即约束条件只需要归一化，正交会自动完成．现在来化简第一项．首先把总波函数中的行列式记为求和的形式

$$\Psi = \frac{1}{\sqrt{N!}} \sum_{N!} (-1)^p \hat{P} \Psi_H \equiv \sqrt{N!} \hat{A} \Psi_H \tag{8}$$

其中 $\Psi_H = u_1(q_1) u_2(q_2) \dots$ 是 Hartree 函数，$\hat{P}$ 是置换算符，相当于做 p 次双粒子置换（p 是逆序数），行列式展开后共有 $N!$ 项．$\hat{A}$ 为反对称化算符，由于 $\hat{H}$ 和 $\hat{A}$ 存在一组共同本征矢，$[\hat{H}, \hat{A}] = 0$，另外可以证明，$\hat{A}^2 = \hat{A}$（意义是反对称化只需要一次）（可先证明 $N = 2, 3$，高阶行列式同理）．

$$\begin{aligned} \langle \Psi | \hat{H} | \Psi \rangle &= N! \langle \Psi_H | \hat{A} \hat{H} \hat{A} | \Psi_H \rangle = N! \langle \Psi_H | \hat{H} \hat{A}^2 | \Psi_H \rangle \\ &= N! \langle \Psi_H | \hat{H} \hat{A} | \Psi_H \rangle = \sum_{N!} (-1)^p \langle \Psi_H | \hat{H} \hat{P} | \Psi_H \rangle \end{aligned} \tag{9}$$

我们现在考虑多电子原子（离子）问题

$$\hat{H} = \sum_i \hat{h}_i + \sum_{i<j} \frac{1}{r_{ij}} \qquad \hat{h}_i = -\frac{1}{2} \nabla^2 + \frac{Z}{r_i} \tag{10}$$

$$\begin{aligned} \langle \Psi | \sum_i \hat{h}_i | \Psi \rangle &= \sum_i \sum_{N!} (-1)^p \langle \Psi_H | \hat{h}_i \hat{P} | \Psi_H \rangle = \sum_i \langle \Psi_H | \hat{h}_i | \Psi_H \rangle \\ &= \sum_i \langle u_i | \hat{h}_i | u_i \rangle \end{aligned} \tag{11}$$

这是因为只有当 $\hat{P}$ 为 1 时（行列式的对角项，逆序数 $p=0$ ）积分才不为零．同理，$\hat{H}$ 剩下的部分为

$$\langle\Psi|\sum_{i<j}\frac{1}{r_{ij}}|\Psi\rangle=\sum_{i<j}\sum_{N!}(-1)^p\langle\Psi_H|\frac{1}{r_{ij}}\hat{P}|\Psi_H\rangle \tag{12}$$

现在 $P=1$ 或者 $P=P_{ij}$ （$p=1$ ）时积分都可能不为零．所以

$$\begin{aligned}\langle\Psi|\sum_{i<j}\frac{1}{r_{ij}}|\Psi\rangle&=\sum_{i<j}\langle\Psi_H|\frac{1}{r_{ij}}(1-P_{ij})|\Psi_H\rangle\\&=\sum_{i<j}\langle u_iu_j|\frac{1}{r_{ij}}|u_iu_j\rangle-\sum_{i<j}\langle u_iu_j|\frac{1}{r_{ij}}|u_ju_i\rangle\\&\equiv\sum_{i<j}J_{ij}-\sum_{i<j}K_{ij}\end{aligned} \tag{13}$$

注意这里 $|u_iu_j\rangle^\dagger$ 记为 $\langle u_iu_j|$ 而不是 $\langle u_ju_i|$．另外易证 $J_{ij}=J_{ji}$，$K_{ij}=K_{ji}$（交换积分变量即可）所以

$$L=\sum_i\langle u_i|\hat{h}_i|u_i\rangle+\sum_{i<j}\langle u_iu_j|\frac{1}{r_{ij}}\langle u_iu_j|-\sum_{i<j}\langle u_iu_j|\frac{1}{r_{ij}}|u_ju_i\rangle-\sum_iE_i\langle u_i|u_j\rangle \tag{14}$$

现在，类似于变分法中的过程，把任意一个 $\langle u_k|$ 变为 $\langle u_k+\delta u_k|$，减去上式，令为 0，得

$$\langle\delta u_k|h_k|u_k\rangle+\sum_j^{(j\neq k)}\langle\delta u_ku_j|\frac{1}{r_{jk}}|u_ku_j\rangle-\sum_j^{(j\neq k)}\langle\delta u_ku_j|\frac{1}{r_{ij}}|u_ju_k\rangle-E_k\langle\delta u_k|u_k\rangle=0 \tag{15}$$

即

$$\langle\delta u_k|\left[h_k|u_k\rangle+\sum_j^{(j\neq k)}\langle u_j|\frac{1}{r_{jk}}|u_j\rangle|u_k\rangle-\sum_j^{(j\neq k)}\langle u_j|\frac{1}{r_{ij}}|u_k\rangle|u_j\rangle-E_k|u_k\rangle\right]=0 \tag{16}$$

最后，由于 $\langle\delta u_k|$ 可以取任意微小函数，与之相乘的 ket 必须为零

$$h_k|u_k\rangle+\left[\sum_j^{(j\neq k)}\langle u_j|\frac{1}{r_{jk}}|u_j\rangle\right]|u_k\rangle-\sum_j^{(j\neq k)}\left[\langle u_j|\frac{1}{r_{ij}}|u_k\rangle\right]|u_j\rangle=E_k|u_k\rangle \tag{17}$$

这是所谓的，非线性耦合微分积分本征方程组．

注意虽然现在 trial 波函数满足全同费米子的反对称，一般却不是总自旋角动量的本征函数（其实也不是总轨道角动量的本征函数，除非把无穷个不同的 Ψ_H 求和）. 为了实现这点，可以先指定总自旋角动量 $|S,M\rangle$，然后通过对行列式线性组合构建自旋部分为 $|S,M\rangle$ 的 trial 波函数.

例 1　氦原子

对于 He 原子的 singlet 自旋态 $(\uparrow\downarrow-\downarrow\uparrow)/\sqrt{2}$，可以用一个行列式构建 trial 波函数. 由于自旋为反对称，轨道波函数必须为对称，对于基态，这意味着两个轨道波函数相同

$$\Psi=\frac{1}{\sqrt{2}}\begin{vmatrix}\phi(\mathbf{r}_1)\uparrow & \phi(\mathbf{r}_1)\downarrow\\ \phi(\mathbf{r}_2)\uparrow & \phi(\mathbf{r}_2)\downarrow\end{vmatrix}=\phi(\mathbf{r}_1)\phi(\mathbf{r}_2)(\uparrow\downarrow-\downarrow\uparrow)/\sqrt{2} \tag{18}$$

把 $u_1=\phi\uparrow$，$u_2=\phi\downarrow$ 代入本征方程组得单个轨道波函数的本征方程

$$\hat{h}\,|\phi\rangle+\langle\phi|\,\frac{1}{r_{12}}\,|\phi\rangle\,|\phi\rangle=E\,|\phi\rangle \tag{19}$$

当然也可以直接把 $\Psi=\phi(\mathbf{r}_1)\phi(\mathbf{r}_2)(\uparrow\downarrow-\downarrow\uparrow)/\sqrt{2}$ 代入能量平均值公式，用分母 $\langle\phi|\phi\rangle$ 归一化或者用拉格朗日乘数法做，得到的方程与上式一样

$$\begin{aligned}L&=2\,\langle\phi\phi|\,\hat{h}\,|\phi\phi\rangle+\langle\phi\phi|\,\frac{1}{r_{12}}\,|\phi\phi\rangle-\lambda[\langle\phi|\phi\rangle-1]\\&=2\,\langle\phi|\,\hat{h}\,|\phi\rangle+\langle\phi\phi|\,\frac{1}{r_{12}}\,|\phi\phi\rangle-\lambda[\langle\phi|\phi\rangle-1]\end{aligned} \tag{20}$$

注意拉格朗日乘数法中的优化函数可以使用限制条件化简. 令增量为 0

$$2\,\langle\delta\phi|\,h\,|\phi\rangle+2\,\langle\delta\phi\phi|\,\frac{1}{r_{12}}\,|\phi\phi\rangle-\lambda\,\langle\delta\phi|\phi\rangle=0 \tag{21}$$

其中使用了 $\delta\,\langle\phi\phi|=\langle\delta\phi\phi|+\langle\phi\delta\phi|$.

$$\hat{h}\,|\phi\rangle+\langle\phi|\,\frac{1}{r_{12}}\,|\phi\rangle\,|\phi\rangle=\frac{\lambda}{2}\,|\phi\rangle \tag{22}$$

对于其他的自旋态，往往需要行列式的线性组合才能构造总自旋本征态，这个比较复杂，先来看另一种方法. 我们可以直接指定总自旋态，如果要求轨道波函数反对称，可把 $u_i(q_i)=\phi_i(\mathbf{r}_i)$ 直接代入本征方程组即可，而当要求轨道波函数对称时，可以把以上的反对称化算符 $\hat{A}$ 中的 $(-1)^p$ 去掉改成对称化算符 $\hat{B}$

$$\hat{B}\,|\Psi_H\rangle=\frac{1}{N!}\sum_{N!}\hat{P}\,|\Psi_H\rangle \tag{23}$$

$\hat{B}$ 同样满足 $[\hat{H},\hat{B}]=0$，$\hat{B}^2=\hat{B}$．以上推导全部有效（所有负号改成正号即可），新的本征方程组变为

$$h_k|u_k\rangle+\left[\sum_j^{(j\neq k)}\langle u_j|\frac{1}{r_{jk}}|u_j\rangle\right]|u_k\rangle+\sum_j^{(j\neq k)}\left[\langle u_j|\frac{1}{r_{ij}}|u_k\rangle\right]|u_j\rangle=E_k|u_k\rangle \tag{24}$$

例 2　氦原子

考虑 He 的 $1s2s$，^1S 态，即自旋为 singlet $|0,0\rangle=(\uparrow\downarrow-\uparrow\downarrow)/\sqrt{2}$．两个空间波函数不相同．直接把 $u_1=\phi_{1s}$，$u_2=\phi_{2s}$ 代入上面的对称本征方程组，得

$$\hat{h}_1|\phi_{1s}\rangle+\langle\phi_{2s}|\frac{1}{r_{12}}|\phi_{2s}\rangle|\phi_{1s}\rangle+\langle\phi_{2s}|\frac{1}{r_{12}}|\phi_{1s}\rangle|\phi_{2s}\rangle=E_1|\phi_{1s}\rangle \tag{25}$$

$$\hat{h}_2|\phi_{2s}\rangle+\langle\phi_{1s}|\frac{1}{r_{12}}|\phi_{1s}\rangle|\phi_{2s}\rangle+\langle\phi_{1s}|\frac{1}{r_{12}}|\phi_{2s}\rangle|\phi_{1s}\rangle=E_2|\phi_{2s}\rangle \tag{26}$$

如果直接用 $|\Psi\rangle=(\phi_1\phi_2+\phi_2\phi_1)/\sqrt{2}$，结果相同．

多通道散射

- 本词条处于草稿阶段．

预备知识　量子散射[1420]

[28]多通道散射（multi-channel scattering）研究的是非弹性散射（inelastic scattering）．例如我们用一个电子或入射到一个原子或者粒子上，将总波函数表示为

$$\Psi=\sum_i\Phi_i(l,m,\mathbf{r}_2,\mathbf{r}_3,\dots)\psi_i(r) \tag{1}$$

其中 $\psi_i(r)$ 表示入射电子的径向波函数，Φ_i 表示其他电子的波函数以及入射电子的量子数 l,m．

关于费米子波函数的反对称，我们假设 Φ 满足 $\mathbf{r}_2,\mathbf{r}_3\dots$ 等电子的反对称条件．先忽略总波函数反对称的要求，如果需要反对称，只需用反对称化算符处理即可．

[28]参考 AMO2 课的笔记，以及 Theoretical Atomic Physics (Friedrich) Chap 3.3

对于弹性散射，虽然我们我们解微分方程得到的两种线性无关的球面波通常是实函数，如 $j_l(kr), y_l(kr)$（代表驻波），但这里我们需要的是行波，就是类似 $h^{(\pm)}(kr) = j_l(kr) \pm y_l(kr)$ 这种，同样是两个．而 $h_l^{(+)}(kr)$ 代表出射波，$h_l^{(-)}$ 代表入射波．$j_l = (h_l^{(+)} + h_l^{(-)})/2$，这样就是一进一出．

对于弹性散射，我们可以把 $h_l^{(+)} + h_l^{(-)}$ 叠加得到球面波．对非弹性散射的本征态，我们也首先希望得到从某个通道（channel）i 入射的球面波 $f_i^{(-)}$ 加上所有通道的出射波 $f_f^{(+)}$．

Adiabatic 笔记

- 本词条处于草稿阶段．

预备知识 量子散射[1420]

Multi-channel 散射可以取 diabatic 基底（相当于 $R \to \infty$ 时的状态），在这种情况下，$R \to \infty$ 时势能矩阵就没有任何 coupling，而当 R 较小时就会有 coupling．

但如果我们将势能矩阵对角化，得到的基底就叫做 adiabatic 基底．

$$H\Psi = E\Psi \tag{1}$$

$$H = T_s + H_{ad}(\mathbf{R}_s) \tag{2}$$

s 角标代表 slow，ad 下标代表 adiabatic．

$$H_{ad}\Phi_\nu = u_\nu(\mathbf{R}_s)\Phi_\nu(\mathbf{R}_s, \Omega_f) \tag{3}$$

f 角标代表 fast．解的时候 $\mathbf{R}_s$ 是常数．u_ν 包括离散和连续．

波函数表示为

$$\Psi(\mathbf{R}_s, \Omega_f) = \sum\!\!\!\!\!\!\int_\nu F_\nu(\mathbf{R}_s)\Phi_\nu(\mathbf{R}_s, \Omega_f) \tag{4}$$

代入薛定谔方程得

$$-\frac{1}{2\mu}\{[\boldsymbol{\nabla}_{\mathbf{R}_s} + \mathbf{P}(\mathbf{R}_s)]^2 + \mathbf{u}\}\mathbf{F} = E\mathbf{F} \tag{5}$$

其中

$$\mathbf{P}_{i,j} = \langle \Phi_i | \boldsymbol{\nabla}_{\mathbf{R}_s} | \Phi_j \rangle \tag{6}$$

这里的积分是对 Ω_f 进行．得到的是一个矩阵，矩阵元为矢量．

为了明确起见，我们把式 5 记为分量的形式为

$$-\frac{1}{2\mu}\left[\nabla^2_{\mathbf{R}_s} F_i + 2\sum_j (P_{ij}\boldsymbol{\nabla}_{\mathbf{R}_s} F_j) + P^{(2)}_{i,j} F_j\right] = EF_j \tag{7}$$

其中 $\mathbf{P}^{(2)}$ 矩阵定义为 $\langle \Phi_i | \nabla^2_{\mathbf{R}_s} | \Phi_j \rangle$，其实就是 $\mathbf{P}$ 矩阵的平方．

到此为止所有公式都是精确的．

晶体衍射

- 本词条处于草稿阶段．

米勒指数

这里只讨论长方体晶格，假设三条边分别为 a, b, c．晶格面是通过许多格点的平面．取一个格点为原点，平面过三条坐标轴的截距分别定义为 $a/h, b/k, c/l$．其中 h, k, l 必须是整数．这样，平面方程为

$$\frac{x}{a/h} + \frac{y}{b/k} + \frac{z}{c/l} = 1 \tag{1}$$

法向量为 $(h/a,\ k/b,\ l/c)/\sqrt{(h/a)^2 + (k/b)^2 + (l/c)^2}$．现在来看相邻的两平面相距多少．截点有仍然需要落在格点上，所以只能是所有截距变为两倍．两平面的距离为法向量与任何一个截距的增量矢量内积

$$d = \frac{1}{\sqrt{(h/a)^2 + (k/b)^2 + (l/c)^2}} \qquad \frac{1}{d^2} = \frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \tag{2}$$

布拉格衍射公式

两个平面，不管格点在上面如何分布，若入射光和出射光和平面夹角都为 θ，那么光程差为 $\delta = 2d\sin\theta$，干涉条件为

$$2d\sin\theta = n\lambda \tag{3}$$

Crystalline Scattering Factor

晶格中各个原子的位置用 (x,y,z) 表示，坐标为 (ax,by,cz)．在进行布拉格衍射时，同一个 cell 里面的不同格点会产生不同平面组，即不同相位．例如两点 (x_1,y_1,z_1) 和 (x_2,y_2,z_2) 所在的两个平面

$$\begin{aligned} d_{12} &= \frac{(h/a,\ k/b,\ l/c)}{\sqrt{(h/a)^2+(k/b)^2+(l/c)^2}} \cdot [a(x_2-x_1), b(y_2-y_1), c(z_2-z_1)] \\ &= \frac{(h\Delta x + k\Delta y + l\Delta z)}{\sqrt{(h/a)^2+(k/b)^2+(l/c)^2}} = d(h\Delta x + k\Delta y + l\Delta z) \end{aligned} \tag{4}$$

然而发生衍射时，d 对应的相位差为 2π，所以 d_{12} 对应的相位差为

$$\delta = 2\pi(h\Delta x + k\Delta y + l\Delta z) \tag{5}$$

如果一个晶格有多个原子，每个原子的散射振幅为 f_i，那么总振幅为

$$F = \sum_i f_i \mathrm{e}^{2\pi \mathrm{i}(hx_i+ky_i+lz_i)} \tag{6}$$

超导唯象解释——伦敦方程

预备知识 麦克斯韦方程[1223]，库伦规范[1219]

本文为唯象论，不涉及超导本质．考虑超导本质需用量子的角度去解释即 BCS 理论．

伦敦第一方程

假定超导体中有两种载流电子——正常传到电子与超导电子，对于超导电子有

$$\frac{\partial}{\partial t}\mathbf{J}_s = \alpha \mathbf{E} \tag{1}$$

其中 $\mathbf{J}_s$ 代表超导体中的超导电流密度，$\alpha = n\dfrac{n_s e^2}{m}$，该方程理解，在超导电子运动速度远小于光速 c 的情况下，磁力对超导电子的影响忽略，此时由 $\mathbf{F}=m\mathbf{a}$ 写在电场中的形式 $m\frac{\partial v}{\partial t} = -e\mathbf{E}$ 推出．实验证明与超导体的相关性质吻合．为第二方程引出给出先决条件 $E=0$, 否则电子速度会不断上升．

伦敦第二方程

$$\nabla \times \mathbf{J}_s = -\alpha \mathbf{B} \tag{2}$$

由式 1 取旋度加之 $\nabla \times E = \frac{\partial B}{\partial t}$ 推得.

伦敦规范

在库伦规范的前提下，矢势 $\mathbf{A}$ 并不唯一，为了使矢势 $\mathbf{A}$ 唯一确定，而在超导体表面 S 上引入限定 $\mathbf{A}$ 的法向分量为 0, 即

$$\nabla \cdot \mathbf{A} = 0 \tag{3}$$

$$\mathbf{e}_n \cdot \mathbf{A}|_s = 0 \tag{4}$$

伦敦方程解释超导现象

定理 1　迈斯纳效应

当材料处于超导态时，随着进入导体内部深度的增加，磁场迅速衰减，磁场主要存在于导体表面一定厚度的薄层内.

恒定情形时（正常电子所导致的电流为零，超导电流与深度有关.）对于超导体内的磁场和电流满足的麦克斯韦—伦敦方程为

$$\nabla \cdot \mathbf{B} = 0 \tag{5}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}_s \tag{6}$$

$$\nabla \cdot \mathbf{J}_s = 0 \tag{7}$$

$$\nabla \times \mathbf{J}_s = -\alpha \mathbf{B} \tag{8}$$

将式 6 取旋度，按照矢量乘积展开，同时代入式 5 与式 2 得

$$\nabla^2 \mathbf{B} = \frac{1}{\lambda_L^2} \mathbf{B} \tag{9}$$

其中 $\lambda_L = 1/\sqrt{\mu_0 \alpha} = \sqrt{\dfrac{m}{\mu_0 n_s e^2}}$，由此可以看出磁场 $\mathbf{B}$ 随着位置变化迅速变化. 方程解亦可能形式为磁场随着位置变化迅速上升或下降，但当考虑现实情况迈斯纳效应时，可以确定真解.

求解超导电流，通过对式 2 取旋度，及式 6 和式 5 的代入得

$$\nabla^2 \mathbf{J}_s = \frac{1}{\lambda_L^2} \mathbf{J}_s \tag{10}$$

同式 9 相同形式，故而可以得到相同的结果，即超导电流分布在超导体表面.

第三章

量子力学与量子场论

前言

本章选自 laserdog 编写的量子力学的讲义，并由小时进行了重新排版和一些格式调整．以下是讲义的前言．

前言

这是一本量子力学的教材，整体结构利用了矩阵力学的形式，主要目的为弥补国内教材和量子力学课的匮乏．

原稿由英文写成，但是在制作讲义的时候放弃了英文：我写的讲义肯定是不如 Sakurai 的大作，或者其他类似的书的．既然这样的话，为什么要放弃我本就不多的目的——方便中国的读者们阅读呢？

因此，本讲义为中文的，当然如果有时间的话我会把我的英文原稿也弄出来，不过意义并不很大．

本讲义主要参考文献为 Sakurai[1]，也适当的参考了我自己学习量子力学的时候各个老师写的讲义．

基本概念

描述一个态

我们用如下记号：$|\alpha\rangle$ 来描述一个态，这种描述也称为“右矢”．从某种程度上，它可以理解为我们线性代数里面学过的列矢量：

$$|\alpha\rangle = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{1}$$

当然，这么写出来的话，肯定要指明我们实在怎样的一组坐标下面来进行这个描述的．一旦清楚的说明坐标了之后，这个态的意义也就明确了起来．在

[1] J. N. J. J. Sakurai, Modern Quantum Mechanics, 2nd ed., edited by L. Kenny (Addison-Wesley, 2011).

量子力学的课程中，一个很容易造成困扰的概念就是“叠加态”．请大家一定记住，叠加态并不是什么奇怪的事情．比如，你在三维直角坐标系中，有一个点 $(1,1,1)/\sqrt{3}$，你会觉得“不可思议！竟然可以同时处于 x 轴、y 轴、z 轴的混合态”吗？不会吧．所以，在量子力学里面，你也不必对此大惊小怪——大家其实是同样的数学结构．

有右矢自然就有左矢．代数的讲，一个右矢对应的左矢是它的 Hermitian 共轭[2]，通常会用 $h.c$ 来表示（Hermitian Conjugation）．共轭大家肯定知道，就是复数里面把幅角取反即可了，但是如果在此基础上再做一个转置的话，就构成了 Hermitian 共轭，形成了左矢，也就是一个行向量．

$$\langle\alpha| = (x_1^* \; x_2^* \; \cdots \; x_n^*) \tag{2}$$

举一个简单的例子，

例 1

如果规定

$$|\alpha\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathrm{i} \\ 1 \end{pmatrix} \tag{3}$$

那么

$$\langle\alpha| = \frac{1}{\sqrt{2}}(-\mathrm{i}, 1) \tag{4}$$

左矢和右矢之间可以做内积，就像一个行矢量和列矢量做内积（乘法）一样，得到一个数 $\langle\alpha|\beta\rangle = x \in \mathbb{C}$．还是刚才那个例子，

例 2

$$\langle\alpha|\alpha\rangle = 1 \tag{5}$$

这样一组对应的左矢和右矢内积得 1 的称为单位向量，也叫归一的向量，归一的态，等等很多叫法．如果一个态没有归一，对它进行归一化的过程为

$$|\beta\rangle \to \frac{|\beta\rangle}{\sqrt{\langle\beta|\beta\rangle}} \tag{6}$$

这里面其实用到了内积的一个小性质：$\langle\alpha|\beta\rangle = (\langle\beta|\alpha\rangle)^*$．读者可以验证一下．

[2]其实这有约定俗成的翻译：厄米；然而，这个翻译实在是不能达意，我并不打算使用它．

我们说了这么久的态，但是我们知道在波动力学里面，我们描述一个态是利用它的波函数，一般是坐标的函数 $\psi(x)$．这个和我们的态有什么关系呢？

其实很有关系的．我先不着急透露，我们先想想内积的几何意义．$\langle\alpha|\beta\rangle$ 是不是相当于求出 $|\beta\rangle$ 是有多少成分在 $|\alpha\rangle$ 上么？也就是说，

$$|\beta\rangle = \langle\alpha|\beta\rangle|\alpha\rangle + \text{与}|\alpha\rangle\text{垂直的成分} \tag{7}$$

那我们考虑这么一组态[3]，$|x\rangle$，它们代表一个态被完全的限制在了 x 处无法移动．那么，一个归一的态 $|\psi\rangle$ 如果能写成坐标空间的波函数形式的话，其最有可能的样子就是

$$\psi(x) = \langle x|\psi\rangle \tag{8}$$

由于无论是波函数还是态矢量（右矢），差一个复数相位都是不会有什么区别的，所以我们的粒子在 $x \to x' + \Delta x'$ 找到的概率，就像一般的量子力学问题里面说的那样，是

$$\int_{x'}^{x'+\Delta x'} |\psi(x)|^2 \,\mathrm{d}x = \int_{x'}^{x'+\Delta x'} \langle\psi|x\rangle\langle x|\psi\rangle \tag{9}$$

如果对全空间积分，这个概率是 1，也就是说

$$\int_{-\infty}^{\infty} \langle\psi|x\rangle\langle x|\psi\rangle = 1 \tag{10}$$

这是一个平凡的结论，但是实际上这很有一点深意的，本章中间的部分可以看到．

捎带说一句，像 $\langle x|\alpha\rangle$ 这样，可以叫 $|\alpha\rangle$ 的 x 表象．类似的还有动量表象等．不同表象之间的变换等问题随后会讲．

算符

首先介绍一个很重要的算符：单位算符．别急，这个是很无聊，但是却有很重要的作用．而且，我们介绍的并不是那么无聊的形式．

考虑一组完备正交归一的态 $|\lambda_i\rangle$．单位算符 I 就可以写成

$$I = \sum_i |\lambda_i\rangle\langle\lambda_i| \tag{11}$$

这就被称为 **resolution of identity**．为了便于理解，我们来举一个例子．

[3]实际上有不可数无穷个态，而且归一的办法和我们之前说的不完全一样，但是这里就先不管数学上这样会不会带来问题了

例 3

三维空间一个矢量

$$\mathbf{v} = v_x\mathbf{i} + v_y\mathbf{j} + v_z\mathbf{k} \tag{12}$$

其中，v_x, v_y, v_z 可以通过 $\langle i\ (or\ j,k)|v\rangle$ 得到.

也就是说，

$$\begin{aligned} |v\rangle &= v_x|i\rangle + v_y|j\rangle + v_z|k\rangle = |i\rangle\langle i|v\rangle + |j\rangle\langle j|v\rangle + |k\rangle\langle k|v\rangle \\ &= (|i\rangle\langle i| + |j\rangle\langle j| + |k\rangle\langle k|)|v\rangle \end{aligned} \tag{13}$$

正如式 11，我们有

$$|\alpha\rangle = \sum_i |\lambda_i\rangle\langle\lambda_i|\alpha\rangle \tag{14}$$

还有很多一般的关于算符的理论. 相信通过前面的理解，我们知道算符在某一组特定的正交完备归一基 $|\lambda_i\rangle$ 下可以表示为矩阵，而矩阵元

$$A_{ij} = \langle\lambda_i|A|\lambda_j\rangle \tag{15}$$

一个物理量显然是一个实数，这实际上要求了它对应的算符是 Hermitian 算符，即

$$\hat{A} = \hat{A}^\dagger, \quad i.e., \quad A_{ij} = A^*_{ji} \tag{16}$$

（实际上，在满足 $\mathcal{PT}$ 对称性下这两件事情并不完全一致，不是 Hermitian 算符也可以得到全部实数的本征值，在本书中不考虑这种事情）而不是物理量的算符并不非要是 Hermitian 的，如

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \& \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{17}$$

介绍几个重要的二维 Hermitian：

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{18}$$

上面四个矩阵被称为 **Pauli 矩阵**，有的时候也会写成四维（矩阵）矢量的形式，$\boldsymbol{\sigma}$.

这四个矩阵十分重要，尤其是在研究自旋 1/2 体系的时候. 这部分内容在后面讲角动量的时候你会时常见到.

关于算符还有很多内容，考虑到这是一个启发性的讲义，这里就不赘述了. 我们可以从 Exercise 里面逐步学习：

习题 1

定义

$$\hat{A}=\begin{pmatrix}0 & 0 & 0 & \cdots\\ 1 & 0 & 0 & \cdots\\ 0 & \sqrt{2} & 0 & \cdots\\ 0 & 0 & \sqrt{3} & \cdots\\ & & & \ddots\end{pmatrix} \tag{19}$$

求出下面算符的本征态和对应的本征值：

$$\hat{H}=\hat{A}^{\dagger}\hat{A}+\frac{1}{2} \tag{20}$$

接下来我们看一下算符的对易关系，其实与其相关的矩阵的对易关系想必你们都学过：

$$\begin{aligned}[\hat{A},\hat{B}] &\overset{def}{\equiv} \hat{A}\hat{B}-\hat{B}\hat{A}\\ \{\hat{A},\hat{B}\} &\overset{def}{\equiv} \hat{A}\hat{B}+\hat{B}\hat{A}\end{aligned} \tag{21}$$

这里不仅介绍了对易，也介绍了反对易．这两者同样常见，不同的是，前者经常出现于玻色体系，而后者经常出现于费米体系．

习题 2

在习题 1 中，我们定义了算符 $\hat{A}$．那么，请计算

$$[\hat{A},\hat{A}^{\dagger}]=? \tag{22}$$

习题 3

试证明，对于 $\forall\hat{A},\hat{B},\hat{C}$

$$[\hat{A},[\hat{B},\hat{C}]]+[\hat{B},[\hat{C},\hat{A}]]+[\hat{C},[\hat{A},\hat{B}]]=0 \tag{23}$$

接下来简要介绍一下算符的完备性，证明暂时略去．可以证明的是，存在逆的 Hermitian 算符，即存在 $\hat{B}\hat{A}=I$ 的时候，如果这个 Hermitian 算符的本征态维数有限，那么这个本征态是完备的．如果无限维，这件事情不一定成立．至于有人问为什么无限维会带来 bug 呢？举个简单的例子证明无限维的时候好多概念都用不了．

例 4

定义[4]一个无限维的完备正交归一解为 $|i\rangle, i = 0, 1, \cdots \infty$，那么考虑算符

$$A = \sum_{i=0}^{\infty} |i\rangle\langle i+1|, \quad B = \sum_{i=1}^{\infty} |i\rangle\langle i-1| \tag{24}$$

那么，显然有

$$AB = I, \quad BA = I - |0\rangle\langle 0| \tag{25}$$

也就是说，逆矩阵这件事情实际上在无限维的时候定义有一定困难.

以上就是关于算符的一些一般性质.

特殊的算符

我们知道，大多数情况下，一个粒子的哈密顿量都会包含动能部分，而动能的表示又是依靠动量的. 很常见的，我们就会考虑动量算符 $\hat{p}$. 这个算符的本征态和对应的本征值为

$$\hat{p}|p_0\rangle = p_0|p_0\rangle \tag{26}$$

而由于是连续的谱，归一化条件将是利用 delta 函数的

$$\langle p_1|p_2\rangle = \delta(p_1 - p_2) \tag{27}$$

这种连续的谱也会发生在坐标算符 $\hat{x}$ 上. 类似的有

$$\hat{x}|x_0\rangle = x_0|x_0\rangle \tag{28}$$

$$\langle x_1|x_2\rangle = \delta(x_1 - x_2) \tag{29}$$

我们前面介绍过的 resolution of identity，这里的形式也需要改变：

$$\int \mathrm{d}p\, |p\rangle\langle p| = I \tag{30}$$

$$\int \mathrm{d}x\, |x\rangle\langle x| = I \tag{31}$$

不过这些都是小变化，仔细想想的话，其实很容易理解.

[4]从这里开始，后面很可能有的时候算符不写它的那个“帽子”ˆ了.

我们接下来要考虑的实际上是一个非常奇怪的问题：$\langle p|x\rangle =?$．答案很简单，形式为

$$\langle p|x\rangle = \frac{1}{\sqrt{2\pi\hbar}}\mathrm{e}^{-\mathrm{i}px/\hbar} \tag{32}$$

但是具体怎么做的呢？我们先看一个貌似与这个问题并不是很相关的问题：平移问题．

让我们考虑沿着 x 方向的平移算符 $\hat{\mathcal{J}}$:

$$\hat{\mathcal{J}}(\mathrm{d}x')|x'\rangle = |x'+\mathrm{d}x'\rangle \tag{33}$$

有了这组性质之后我们实际上可以得到任何态 $|\alpha\rangle$ 是如何在这样一个平移下变换的，因为我们可以把它展开成坐标本征态的线性组合

$$\hat{\mathcal{J}}(\mathrm{d}x')|\alpha\rangle = \int \hat{\mathcal{J}}(\mathrm{d}x')|x'\rangle\langle x'|\alpha\rangle\,\mathrm{d}x' = \int |x'+\mathrm{d}x'\rangle\langle x'|\alpha\rangle\,\mathrm{d}x' \tag{34}$$

有几点值得注意：这个算符满足很多性质，比如它是 Unitary 的，再比如说它是非常常见的可以线性组合的

$$\begin{cases} \hat{\mathcal{J}}(\mathrm{d}x')\hat{\mathcal{J}}^{\dagger}(\mathrm{d}x') = 1 \\ \hat{\mathcal{J}}(\mathrm{d}x')\hat{\mathcal{J}}(\mathrm{d}x'') = \hat{\mathcal{J}}(\mathrm{d}x'+\mathrm{d}x'') \\ \hat{\mathcal{J}}(-\mathrm{d}x') = \hat{\mathcal{J}}^{-1}(\mathrm{d}x') \\ \hat{\mathcal{J}}(0) = I \end{cases} \tag{35}$$

考虑到量子力学的基本假设[5]，$[x,p]=\mathrm{i}\hbar$，我们注意到

$$\hat{x}\hat{\mathcal{J}}(\mathrm{d}x')|x'\rangle = (x'+\mathrm{d}x')|x'+\mathrm{d}x'\rangle \tag{36}$$

$$\hat{\mathcal{J}}(\mathrm{d}x')\hat{x}|x'\rangle = x'|x'+\mathrm{d}x'\rangle \tag{37}$$

也就是说

$$[\hat{x},\hat{\mathcal{J}}(\mathrm{d}x')] \sim \mathrm{d}x' \tag{38}$$

这里面的约等于号是考虑到 $|x'+\mathrm{d}x'\rangle \sim |x'\rangle$．利用在 $\mathrm{d}x'\to 0$ 的时候 $\hat{\mathcal{J}}(\mathrm{d}x')\to 1$，我们得到

$$\hat{\mathcal{J}}(\mathrm{d}x') \sim 1 - \frac{\mathrm{i}}{\hbar}\hat{p}\,\mathrm{d}x' \tag{39}$$

[5]我后面**可能**会总结一下基本假设，但也可能不总结了 orz

例 5

可以证明 $\hat{P}$ 是 Hermitian 的．一般这种问题的做法是看矩阵元，比如我有一套本征态 $|i\rangle$，定义 $P_{ij} \equiv \langle i|P|j\rangle$．如果他是 Hermitian 的，那么 $P_{ij} = P_{ji}^*$．但是这里的话，我们没有办法选取离散的本征态，不过即使只是连续本征态，也是可以的．最直觉的做法就是选取坐标本征态，我们规定 $P(x,x') = \langle x|P|x'\rangle$．我们只需要证明 $P(x,x') = P(x',x)^*$．这里简要地给出一个“令人信服”但不那么严格的说法．

我们有一个最方便的办法来计算，就是利用 $\mathrm{d}x'$ 很小的时候的 $\hat{\mathcal{J}}(\mathrm{d}x')$．

$$\langle x|\hat{\mathcal{J}}(a)|x'\rangle \sim \langle x|x'\rangle - \frac{\mathrm{i}}{\hbar}a\langle x|P|x'\rangle \tag{40}$$

也就是说

$$P(x,x') \sim \mathrm{i}\frac{\hbar}{a}[\langle x|\hat{\mathcal{J}}(a)|x'\rangle - \langle x|x'\rangle] = \mathrm{i}\frac{\hbar}{a}[\langle x|x'+a\rangle - \langle x|x'\rangle] \tag{41}$$

而很类似的

$$P(x',x) \sim \mathrm{i}\frac{\hbar}{a}[\langle x'|x+a\rangle - \langle x'|x\rangle] \tag{42}$$

$$P(x,x')^* \sim -\mathrm{i}\frac{\hbar}{a}[\langle x'+a|x\rangle - \langle x'|x\rangle] = \mathrm{i}\frac{\hbar}{a}[\langle x'|x\rangle - \langle x'+a|x\rangle] \tag{43}$$

取 $a \to 0$，上面的式子严格等于，而且两者的关系更一目了然．

投影算符

这是一个概念，看起来并不是很有趣，但是我们在后面学习微扰理论的时候会发现它很重要．

它一般指这样一个算符：我们现在有一组正交完备归一的基 $\{|\alpha_i\rangle\}$．我们很关心一个初始在 $|\alpha_j\rangle$ 的基的行为，并且可以预料到，在其刚刚开始演化的时候，其大部分成分都还这个态上．这就有点像你把一根直线转了一个小角度，它沿原来方向的分量还是很大，如图 1．

很显然的，这个投影算符就可以用 $|\alpha_j\rangle\langle\alpha_j|$ 来表示．我们还是利用我们之前介绍过的一个例子来说明．

例 6

三维空间一个矢量

$$\mathbf{v} = v_x\mathbf{i} + v_y\mathbf{j} + v_z\mathbf{k} \tag{44}$$

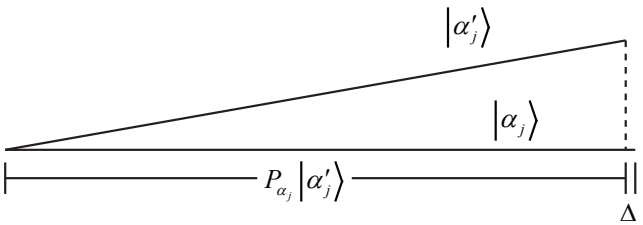


图 1: 偏移一点的矢量与投影

如果我要获得它的 x 方向的投影，我首先要保证它在 x 方向上，其次要保证它的大小．于是，最后得到

$$\mathbf{v}_x = v_x\mathbf{i} = (\mathbf{v}\cdot\mathbf{i})\mathbf{i} = \langle i|v\rangle|i\rangle = |i\rangle\langle i|\mathbf{v} \tag{45}$$

投影算符不仅指投影到一个方向的，还指投影到一个平面上的投影算符，或者投影到 Hilbert 空间上更“高维度”的空间．

这个投影算符虽然现在看起来没有什么用，但是后面你会看到它的作用的．几个有趣的性质读者可以作为练习

习题 4

请读者证明：如果 $\hat{P}$ 是投影算符，则

(1) $\hat{P}\hat{P} = \hat{P}$

(2) $1 - \hat{P}$ 也是投影算符．

算符的函数

我们很多时候可以写一些算符的函数来简化算符．大部分时候这种函数都是一目了然的，比如 $\hat{A}^2 = \hat{A}\cdot\hat{A}$，但有的时候也存在一些很微妙的东西．值得注意的是，无论什么时候，都要记得算符不一定是可对易的．

一个简单的函数，$f(\hat{A}) = \hat{A}^2$ 就容易造成初学者的失误．比如

$$f(\hat{A}+\hat{B}) = \hat{A}^2 + \hat{A}\hat{B} + \hat{B}\hat{A} + \hat{B}^2 \tag{46}$$

稍微复杂点的函数，比如

$$\exp\left(\hat{A}\right) = I + \hat{A} + \frac{1}{2!}\hat{A}^2 + \cdots \tag{47}$$

也是非常常见的一种算符函数．

习题 5

考虑一个体系，它的 Hilbert 空间是二维的，并且记 $\begin{pmatrix}1\\0\end{pmatrix}=|\uparrow\rangle, \begin{pmatrix}0\\1\end{pmatrix}=|\downarrow\rangle$. 现在有一个算符

$$\hat{U}=\exp(\mathrm{i}\alpha(t)\sigma_x) \tag{48}$$

其中，σ_x 是式 18 定义的 Pauli 矩阵，那么求

$$\hat{U}\begin{pmatrix}1\\0\end{pmatrix}=? \tag{49}$$

显然，我们有

$$\hat{U}=I_{2\times2}+\alpha(t)\sigma_x+\frac{1}{2!}\left(\alpha(t)\sigma_x\right)^2+\dots \tag{50}$$

由于 Pauli 矩阵有很好的性质：

$$\sigma_x^2=I_{2\times2} \tag{51}$$

我们有

$$\begin{aligned}\hat{U}&=\left(1-\frac{1}{2!}\alpha(t)^2+\frac{1}{4!}\alpha(t)^4\cdots\right)I_{2\times2}+\left(\alpha(t)+\frac{1}{3!}\alpha(t)^3+\cdots\right)\mathrm{i}\sigma_x\\&=\cos\alpha(t)I_{2\times2}+\sin\alpha(t)\mathrm{i}\sigma_x\end{aligned} \tag{52}$$

从而，我们有

$$\hat{U}\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}\cos\alpha(t)\\\mathrm{i}\sin\alpha(t)\end{pmatrix} \tag{53}$$

这个例子实际上在我们后面看到动力学分析中是一个很常见的“量子演化问题”的基本处理．以后我们会再次看到这个问题的．

以上是一个带答案的练习，现在我们出一个证明性的练习．

习题 6

求证：

$$\mathrm{e}^{\hat{A}}\hat{B}\mathrm{e}^{-\hat{A}}=\hat{B}+[\hat{A},\hat{B}]+\frac{1}{2!}[\hat{A},[\hat{A},\hat{B}]]+\cdots \tag{54}$$

这个等式叫做 Baker-Hausdorff formula，我们可能会经常地遇到这个式子．

⋆ 密度矩阵，与量子信息基本概念

量子信息可以说是只要具备完善的量子力学知识就能够入门，但是入的好不好得看初等数学功力的一个方向.

量子信息里面一个重要的概念就是**密度矩阵**. 对于**纯态（Pure State）**，即一个可以被用态矢量描述的态，归一态矢量为 $|\psi\rangle$，其密度矩阵可以用

$$\hat{\rho}_\psi = |\psi\rangle\langle\psi| \tag{55}$$

来描述. 这样的描述可以略去态矢量的相位，是自由度更小的更“好”的描述.

纯态密度矩阵有很多性质，比如

$$\hat{\rho} = \hat{\rho}^\dagger,\ \mathrm{Tr}(\hat{\rho}) = 1,\ \mathrm{Tr}(\hat{\rho}^2) = 1 \tag{56}$$

这是由于

$$\rho^2 = |\psi\rangle\langle\psi|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi| = \rho \tag{57}$$

一个算符在这个态下的平均值

$$\langle\hat{O}\rangle = \langle\psi|\hat{O}|\psi\rangle = \mathrm{Tr}(\hat{O}\hat{\rho}) \tag{58}$$

与纯态相对的是**混态（Mixed State）**. 我们来想这么一个问题：纯态就是一个态，按照统计的观点，它的熵就是 $\ln 1 = 0$. 而如果我们这个体系描述的东西内部很复杂，各个态之间自发的跳动，我们不能简单的用一个态来描述，怎么办呢？我们仍然能用密度矩阵来描述. 一个最基本的要求就是这个密度矩阵的性质得和纯态一致，即

$$\hat{\rho} = \hat{\rho}^\dagger,\ \mathrm{Tr}(\hat{\rho}) = 1 \tag{59}$$

一个算符在这个态下的平均值

$$\langle\hat{O}\rangle = \langle\psi|\hat{O}|\psi\rangle = \mathrm{Tr}(\hat{O}\hat{\rho}) \tag{60}$$

我们再来看看密度矩阵的一些行为. 由于它是 Hermitian，我们通过某种对角化，显然可以把密度矩阵写成

$$\rho = \sum_i \lambda_i |i\rangle\langle i|, \quad \sum_i \lambda_i = 1, \quad \forall i \lambda_i > 0 \tag{61}$$

这部分有点数学，而且有的时候并不有趣，我们来看一个练习.

习题 7

假设存在一系列的密度矩阵 ρ_j，和一些复数 $c_j, \sum c_j = 1 \& \forall j, c_j > 0$．证明：

$$\sum_j c_j \rho_j \tag{62}$$

也是一个密度矩阵．

一个系统处于某个态的时候可以求它的熵，其中，目前广泛接受并应用的熵是 von Neumann 熵，其定义为

$$S = -\mathrm{Tr}(\hat{\rho} \ln \hat{\rho}) = -\sum_i \lambda_i \ln \lambda_i \tag{63}$$

当然，对于纯态，$S = 0$，但是对于混态则不然．会想热力学，我们知道熵有一个很有趣的结论，我们这里留作练习．

习题 8

von Neumann 熵满足两个系统的混杂会增大熵，数学上讲也就是

$$S[\lambda\hat{\rho}_1 + (1-\lambda)\hat{\rho}_2] \geq \lambda S[\hat{\rho}_1] + (1-\lambda) S[\hat{\rho}_2], \quad \forall \hat{\rho}_1, \rho_2, \quad 0 < \lambda < 1 \tag{64}$$

试证明这点（答案：Nielsen[6], section 11.3.5）．

有的时候，我们要考虑多粒子的问题．当然真正的量子多粒子问题我们在“全同粒子的统计[1493]”中再仔细考虑，我们这里先看一些简单的情况，比如定域的（从而就不需要考虑全同性了）．有两个粒子 a, b，其各自 Hilbert 空间为 $\mathcal{H}_a, \mathcal{H}_b$．那么，整体 Hilbert 空间，代数的讲，就变成了两个粒子各自 Hilbert 空间的**张量积**[7]：$\mathcal{H} = \mathcal{H}_a \otimes \mathcal{H}_b$．这时候我们定义**约化密度矩阵”（reduced density matrix）**，来得到单粒子的信息．比如，我们如果想了解 a 粒子，那么就可以写它的对 a 的约化密度矩阵 $\hat{\rho}_a = \mathrm{Tr}_b(\hat{\rho})$，其是通过对 $\mathcal{H}_b$ 做部分取迹做到的．技术上讲，这是通过用一组完备的基 $|\phi_i\rangle \in \mathcal{H}_b$ 来做到的：

$$\langle \psi_1 \hat{\rho}_a | \psi_2 \rangle = \sum_i \langle \psi_1 \otimes \phi_i | \hat{\rho} | \psi_2 \otimes \phi_i \rangle \tag{65}$$

我们用一个练习来结束这一部分．

[6] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2010).

[7] 张量积的矩阵表达：$A \otimes B$ 就是把 A 中的每个元素 A_{ij} 变成一个矩阵 $A_{ij} \cdot B$．

习题 9

求 trace 自然会发生很多事情，一个显然的事情就是两体关联的失去：

$$\begin{aligned} S_{a\otimes b}[\hat{\rho}] &= \mathrm{Tr}_{a\otimes b}(-\hat{\rho}\ln\hat{\rho}) \leq S_a[\hat{\rho}_a] + S_b[\hat{\rho}_b] \\ &= \mathrm{Tr}_a(-\hat{\rho}_a\ln\hat{\rho}_a) + \mathrm{Tr}_b(-\hat{\rho}_b\ln\hat{\rho}_b) = S_{a\otimes b}[\hat{\rho}_a\otimes\hat{\rho}_b] \end{aligned} \tag{66}$$

试证明这一点（答案：Trace[8]），并验证对于 Bell 基，比如

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle|1\rangle - |1\rangle|0\rangle) \tag{67}$$

这个不等式是成立的.

这是量子信息论的一个经常需要考虑的事情，很多量子信息相关的书，比如 XiaoGang Wen 的书[9] 都介绍了这点，在此不再赘述.

我们这里再讲有关密度矩阵的一个知识：**在有限温度的时候，密度矩阵如何表示**（一般我们处理的时候都是处理零温问题，不考虑统计分布）. 其实很简单：

$$\rho = \exp\left(-\hat{H}/k_BT\right)/Z \tag{68}$$

其中，$Z = \mathrm{Tr}\exp\left(-\hat{H}/k_BT\right)$. 更多的关于密度矩阵，量子信息的内容就不在此赘述了.

测量，不确定性原理

测量发生于一个 Hermitian 算符 $\hat{A}$ 在纯态或者混态 $\hat{\rho}$ 上. 最终得到的结果是算符 $\hat{A}$ 的本征值 λ，而测量得到这个本征值 λ 的概率是 $P_\lambda = \mathrm{Tr}(\hat{P}_\lambda\hat{\rho})$，而这里 $\hat{P}_\lambda = \sum|\hat{A}=\lambda\rangle\langle\hat{A}=\lambda|$ 是投影算符. 可以想象，统计上来说，平均的测量值是 $\hat{A}$ 在态 $\hat{\rho}$ 上的期待值，也就是 $\mathrm{Tr}(\hat{A}\hat{\rho}) = \sum_\lambda \lambda P_\lambda$. 其中，对于纯态，就是简单的 $\langle\psi|\hat{A}|\psi\rangle, \hat{\rho} = |\psi\rangle\langle\psi|$.

量子力学有几个假设，其中比较有名的（你们或许在此前就听说过）就是‘‘塌缩’’假设:如果测量的结果是 λ,那么这个量子态就会塌缩到 $\hat{P}_\lambda\hat{\rho}\hat{P}_\lambda/\mathrm{Tr}(\hat{P}_\lambda\hat{\rho})$. 特例就是当本征值为 λ 的态只有一个的话，系统就会塌缩到 $|\hat{A}=\lambda\rangle$ 上. 下面一个练习是对这一点的直接应用.

---

[8] H. Araki and E. H. Lieb, *Communications in Mathematical Physics* 18, 160 (1970).

[9] B. Zeng, X. Chen, D. Zhou, and X. Wen, ArXiv e-prints (2015), arXiv:1508.02595 [cond-mat.str-el].

习题 10

考虑一个系统，它的态可以用角动量 l、角动量 z 分量 m 的共同本征态展开为

$$\begin{aligned}|\psi\rangle =&\frac{1}{\sqrt{2}}\left(\frac{1}{2}|l=1,m=1\rangle+\frac{1}{\sqrt{2}}|l=1,m=0\rangle+\frac{1}{2}|l=1,m=-1\rangle\right)\\&+\frac{1}{\sqrt{2}}|l=0,m=0\rangle\end{aligned}\tag{69}$$

现在测它的角动量 z 分量，测到 $m=0$. 求测量完的体系处于什么态（用密度矩阵和态矢分别写出）.

量子信息学的一个很重要的（也看起来很平庸的）结论是：通过测量，我们得到了信息. 而得到了信息也就是说，熵见笑了. 这换成数学语言就是说

$$S[\hat{\rho}]\geq\sum_{\lambda}P_{\lambda}S[\hat{\rho}_{\lambda}]\tag{70}$$

其中，$\hat{\rho}_{\lambda}$ 表示测量到塌缩到的 λ 态. 我们看一个例子.

习题 11

考虑一个密度矩阵

$$\hat{\rho}=\frac{1}{4}[\mathbb{I}-\sigma_1\otimes\sigma_2]\tag{71}$$

1. 求验算：它是否是一个纯态？

2. 定义算符 $\hat{A}=\sigma_1\otimes\sigma_0$，求它的本征值，并定义投影到这些本征值的投影算符，计算进行该算符的测量后得到各个本征值的概率，并验证我们之前对于测量得到信息的结论.

接下来，也就是本章的最后，我们来看**不确定性原理**.

对于 Hermitian 算符 $\hat{A}\&\hat{B}$，不确定原理告诉我们

$$(\langle\hat{A}^2\rangle-\langle\hat{A}\rangle^2)(\langle\hat{B}^2\rangle-\langle\hat{B}\rangle^2)\geq\frac{1}{4}|\langle[\hat{A},\hat{B}]\rangle|^2\tag{72}$$

其中，$\langle\cdot\rangle$ 是其在某个量子态 $\hat{\rho}$ 下的期待值.

证明如下：首先，定义所谓两个算符（不一定是 Hermitian）的内积：

$$(\hat{A},\hat{B})=\langle\hat{A}^{\dagger}\hat{B}\rangle=\mathrm{Tr}(\hat{A}^{\dagger}\hat{B}\hat{\rho})\tag{73}$$

不难看出，它是满足内积的性质的（线性、共轭）．现在考虑两个新的算符 $\hat{A}' = \hat{A} - \langle\hat{A}\rangle, \hat{B}' = \hat{B} - \langle\hat{B}\rangle$，此时

$$\begin{aligned}\frac{1}{4}|\langle[\hat{A},\hat{B}]\rangle|^2 &= \frac{1}{4}|\langle[\hat{A}',\hat{B}']\rangle|^2 \\ &= \frac{1}{2}(\hat{A}',\hat{B}')(\hat{B}',\hat{A}') - \frac{1}{4}(\hat{A}',\hat{B}')^2 - \frac{1}{4}(\hat{B}',\hat{A}')^2 \\ &= [\mathrm{Im}(\hat{A}',\hat{B}')]^2 \\ &\leq |(\hat{A}',\hat{B}')|^2 \leq (\hat{A}',\hat{A}')(\hat{B}',\hat{B}') \\ &= (\langle\hat{A}^2\rangle - |\langle\hat{A}\rangle|^2)(\langle\hat{B}^2\rangle - |\langle\hat{B}\rangle|^2)\end{aligned} \tag{74}$$

补充几个练习题

习题 12

如果 $[\hat{A},\hat{B}] = \hat{B}^2$，用 $\hat{B}$ 表示 $\mathrm{e}^{\hat{A}}\hat{B}\mathrm{e}^{\hat{A}}$.

习题 13

矢量 $\mathbf{A},\mathbf{B}$ 为三维矢量，而 $\boldsymbol{\sigma}$ 则是 Pauli 矩阵的矢量，计算 $[\mathbf{A}\cdot\boldsymbol{\sigma}, B\cdot\boldsymbol{\sigma}]$

习题 14

矢量 $\mathbf{n}$ 为三维实数矢量，而 $\boldsymbol{\sigma}$ 则是 Pauli 矩阵的矢量，计算 $\exp(\mathrm{i}\theta\mathbf{n}\cdot\boldsymbol{\sigma})$，并计算 $\exp(\mathrm{i}\theta\mathbf{n}\cdot\boldsymbol{\sigma})(\mathbf{A}\cdot\boldsymbol{\sigma})\exp(-\mathrm{i}\theta\mathbf{n}\cdot\boldsymbol{\sigma})$

全同粒子的统计

为什么要讲全同粒子呢？因为，我们的世界有两种粒子，玻色子和费米子，他们的性质截然不同，但又都很重要．上一章讲了基本的数学结构，这一章就是对这两种粒子建立这种数学结构．要注意，只有玻色子和费米子这两种粒子在低维的时候很可能被打破，形成一种“任意子（anyon）”，甚为可怕，其在量子信息中也有很重要的作用，是处理非 Abelian 交换的，可做逻辑门．更多信息可以参考[10].

[10] C. Nayak *et al.*, *Rev. Mod. Phys.* 80, 1083 (2008).

Fock 空间

一个简单的说法，Fock 空间就是不同粒子数的态所在的空间．例如，一个玻色体系，有两个能级，那么第一个能级 n 个粒子，第二个能级 m 个粒子的态，我们并不关心其具体的波函数形式 $\psi(x)$ 这样的，而是抽象的写成 $|n,m\rangle$．当然，这里我们还没有明确的定义玻色和费米体系都是什么样子，这个只是简单的给一个大致想法．接下来，我们仔细的构建这套理论．

有关排列群

首先，我们看一些有关于排列群的知识．这部分不是必要的，对于理解某些特殊的问题时有奇效[11]．

全同粒子的传统处理方法

如果我有 n 个粒子，由于是全同的，我说这 n 个粒子在 $\mathrm{x}_1,\mathrm{x}_2\ldots\mathrm{x}_n$ 这 n 个位置组成的**无序数组**的情况可以用类似我们 single particle 的办法描述，也就是使用波函数

$$\psi(\mathrm{x})=\psi(\mathrm{x}_1,\mathrm{x}_2\ldots\mathrm{x}_n) \tag{1}$$

然而这个波函数的 n 个位置参数显然是有序的．我们可以断言，在这 n 个坐标进行重新排列之后，波函数应该不变——或者至少说，密度矩阵不变，波函数有相位差别．

在波函数的形式仅仅是一个函数——并不是什么旋量啊之类的奇怪的东西，而仅仅是一个复数的时候（这也被叫做排列群的一维表示），粒子自然而然地被分为了两类：波色子和费米子．

我们取一个排列算符 $\sigma\in S_n$，并将这个算符作用在波函数的参数上，得到的应该是一个对应的一维表示 $R(\sigma)$ 作用在波函数上

$$\psi(\sigma(\mathrm{x}_1,\mathrm{x}_2\ldots\mathrm{x}_n))=R(\sigma)\psi(\mathrm{x}_1,\mathrm{x}_2\ldots\mathrm{x}_n) \tag{2}$$

显然满足的是 $|R(\sigma)|=1$．现在考虑一对位置的交换，比如把 i 位置和 j 位置进行交换，这样的操作记为 $\sigma_{i,j}$．显然的有 $R(\sigma_{i,j})^2=1$，也就是说

$$R(\sigma_{i,j})=\pm1 \tag{3}$$

[11]但是大多数时候这是用来装逼的，不想看的话跳过就可以了

如果更具体的分析这两种群表述的区别，在量子场论里面我们能一目了然的看到这与自旋之间有着密切的关系．在后面角动量的部分和相对论性量子力学部分我们也会稍微的提及一点．不过在这里我只给出结论：得到 $+1$ 结果的是玻色子（Boson），两个粒子交换波函数不变；得到 -1 结果的是费米子（Fermion），两个粒子交换波函数变成原来的负数．

然而，如果波函数不是一个数的话，这有可能就是一个排列群的二维或者更复杂的表示了，我们在此不赘述．

多体 Hilbert 空间的数学结构

波函数所在的 Hilbert 空间，在这套框架下是没有确定粒子数的空间，也就是说最终的 Hilbert 空间——Fock 空间，是各个粒子数的 Hilbert 空间的**直和**．

$$\mathcal{F} = \mathcal{H}_0 \oplus \mathcal{H}_1 \dots \tag{4}$$

而我们可以一目了然的说明的就是，0 粒子状态就是一种我们称为“真空”的状态，可以用 $|0\rangle$ 来指代．更准确一点地说，真空态相当于在 1 粒子波函数的 Hilbert 空间中没有粒子（波函数等于 0），而 1 粒子波函数的 Hilbert 空间就是单体波函数所在的线性空间，也是如果我们有一些简单的波动力学基础的话最熟悉的那种．

而更高粒子数的子空间 $\mathcal{H}_n$，则是指通过对称化或者反对称化之后的 n 个单粒子空间的**张量积的子空间**，$(\mathcal{H}_1)^{\otimes n} \equiv \mathcal{H}_1 \otimes \mathcal{H}_1 \cdots \otimes \mathcal{H}_1$，如图 1 就是一个张量积的示意．

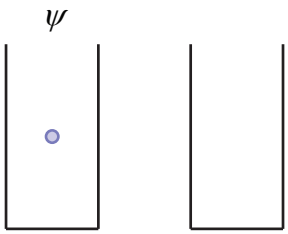


图 1: 体系的波函数为 $|\psi\rangle_{\text{left}} \otimes |0\rangle_{\text{right}}$

而我们注意到，$\mathcal{H}_n$ 仅仅是张量积的子空间，要建立完整的数学描述还需要告诉我们如何从这个张量积空间获得这个需要的子空间．也就是说，形式上

的，我们需要找到这么一个线性映射：

$$\mathcal{S}: \qquad (\mathcal{H}_1)^{\otimes n} \mapsto \mathcal{H}_n \tag{5}$$

$$|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle \mapsto |\psi_1, \psi_2, \ldots, \psi_n\rangle \tag{6}$$

很显然的，我们就得到了两种情况下的这个线性映射 $\mathcal{S}$：

- Boson Case

$$|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle \mapsto \sum_{\sigma\in S_n} |\psi_{\sigma(1)}\rangle \otimes |\psi_{\sigma(2)}\rangle \otimes \cdots \otimes |\psi_{\sigma(n)}\rangle \tag{7}$$

- Fermion Case

$$|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle \mapsto \sum_{\sigma\in S_n} (-1)^\sigma |\psi_{\sigma(1)}\rangle \otimes |\psi_{\sigma(2)}\rangle \otimes \cdots \otimes |\psi_{\sigma(n)}\rangle \tag{8}$$

其中，熟悉线性代数的同学肯定会猜到，这个 $(-1)^\sigma$ 就是一个重排操作 σ 中交换的次数，或者说逆序数.

接下来我们考虑多体问题 Hilbert 空间的基．最基础的版本是我们的波函数是空间坐标的函数，当然是多个空间坐标：$\psi(x_1, \ldots, x_n)$．但是，由于统计性质的要求，我们要求（对 Boson 取 $+$，Femrion 取 $-$）

$$\psi(x_1, \ldots, x_n) = \frac{1}{\sqrt{n!}} \sum_{\sigma\in S_n} (\pm 1)^\sigma \langle x_1|\psi_{\sigma(1)}\rangle \ldots \langle x_n|\psi_{\sigma(n)}\rangle \tag{9}$$

习题 1　练习（Majorana 算符）

考虑两个格点，这两个格点上有相同的费米子模式．记 $|1\rangle_{1,2} = c^\dagger_{1,2}|0\rangle$，

(1) 在

$$\begin{pmatrix} |0\rangle_1|0\rangle_2 \\ |0\rangle_1|1\rangle_2 \\ |1\rangle_1|0\rangle_2 \\ |1\rangle_1|1\rangle_2 \end{pmatrix} \tag{10}$$

的基下，写出 $c_{1,2}, c^\dagger_{1,2}$ 的矩阵形式．此处我们定义 $|1\rangle_1|1\rangle_2 \equiv c^\dagger_1 c^\dagger_2 |0\rangle_1|0\rangle_2$.

近似理论：微扰

- 本词条缺少预备知识，初学者可能会遇到困难.

无论在一般的量子力学框架下，还是其延伸出的各种问题，比如量子多体问题、量子统计问题、量子场论问题等等，可以精确求解的本身就是少之又少. 很多时候，一些必要的近似手段才真正地发挥了大作用.

一个最成功的近似理论就是微扰理论. 假如我们对某一个体系了解的很充分，那么对这个体系在小的已知扰动（比如，由杂质或者外界场影响等等）的行为可以非常好的用微扰进行处理，得到很多十分有效的结论.

不含时 Time-independent

很多微扰是不含时的微扰——比如说，我的杂质就一直扔在那里了，不管你时间怎么变化我的杂质又不会消失. 这时候的微扰论实际上是比较基础的，我们不用引入很多奇葩观念就可以轻松地解决问题. 在这种时候，有一个很有趣的现象我们后面会看到.

注意到，如果微扰很小的话，微扰之后本征态的能量、态都很接近微扰之前的能量，比如说，$|n^{(0)}\rangle, E_n^{(0)}$ 分别表示原来的本征态和能量，微扰之后在它附近的能量本征态为 $|n,\lambda\rangle, E_n$，λ 为一个描述微扰大小的参数. 所以，我们的一般想法就是：假定满足如下关系：

$$E_n = E_n^{(0)} + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + \cdots |\psi_n\rangle = |\psi_n^{(0)}\rangle + \lambda|\psi_n^{(1)}\rangle + \lambda^2|\psi_n^{(2)}\rangle \tag{1}$$

注意，一目了然，这并没有进行**归一化**. 另外，很直观的，假定 $|\psi_n^{(k>0)}\rangle$ 与原本征态 $|\psi_n\rangle$ 正交.

非简并问题

为什么上来先强调非简并问题，这实际上在后面可以看出来. 一个比较直观的说法是这样的：如果简并了，那么我实际上在这两个简并态之间随意组合都可以得到处于该能量的本征态，尽管其中只有一些是有一定的物理意义（也就是其他算符本征态）的. 在这么多种本征态组合中，究竟哪个和微扰之后的本征态最为接近呢？这就是一个技术上的问题. 并不是不能考虑，但是在这里我们先处理点最简单的问题，在后面我们再考虑简并情况.

在进一步说明问题前，我们先考虑一个更更简单的情况：一个可以严格对角化的 2×2 问题.

习题 1

一个体系的非微扰 Hamiltonian 是：

$$H_0 = E_1|e^0\rangle\langle e^0| + E_2|g^0\rangle\langle g^0| = \begin{pmatrix} E_1 & 0 \\ 0 & E_2 \end{pmatrix} \tag{2}$$

其中 $E_1 < E_2$．而微扰部分的 Hamiltonian 则是

$$H_1 = \lambda(|e^0\rangle\langle g^0| + |g^0\rangle\langle e^0|) \tag{3}$$

显然，这是一个实对称矩阵（Hermitian 矩阵），我们可以把它用一个正交矩阵（Unitary 矩阵）进行对角化：

$$H = H_0 + H_1 = U^\dagger \begin{pmatrix} \lambda_- & 0 \\ 0 & \lambda_+ \end{pmatrix} U \tag{4}$$

其中，我们知道

$$\lambda_\pm = \frac{E_1 + E_2}{2} \pm \sqrt{\left(\frac{E_1 - E_2}{2}\right)^2 + \lambda^2} \tag{5}$$

假如这个微扰后的 Hamiltonian 的本征态是 $(a_\pm, b_\pm)^T$，那么我们知道

$$U \begin{pmatrix} a_- \\ b_- \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad U \begin{pmatrix} a_+ \\ b_+ \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{6}$$

从而

$$U^T(\equiv U^\dagger) = \begin{pmatrix} a_- & a_+ \\ b_- & b_+ \end{pmatrix}, \quad U = \begin{pmatrix} a_- & b_- \\ a_+ & b_+ \end{pmatrix} \tag{7}$$

而我们知道：未归一化的本征态可以写成很好的形式，即

$$\begin{pmatrix} a_- \\ b_- \end{pmatrix} \propto \begin{pmatrix} 1 \\ \dfrac{\lambda_- - E_1}{\lambda} \end{pmatrix}, \quad \begin{pmatrix} a_+ \\ b_+ \end{pmatrix} \propto \begin{pmatrix} 1 \\ \dfrac{\lambda_+ - E_2}{\lambda} \end{pmatrix} \tag{8}$$

1. 验证上式

2. 计算：如果在 $t = 0$ 的时候体系处于 $|\psi(t=0)\rangle = |g^0\rangle$，那么其时间演化是怎样的

(Hint and solution of 2. :

设

$$\begin{pmatrix} a_- \\ b_- \end{pmatrix} = \frac{1}{A}\begin{pmatrix} 1 \\ \dfrac{\lambda_- - E_1}{\lambda} \end{pmatrix}, \quad \begin{pmatrix} a_+ \\ b_+ \end{pmatrix} = \frac{1}{B}\begin{pmatrix} 1 \\ \dfrac{\lambda_+ - E_1}{\lambda} \end{pmatrix} \tag{9}$$

那么，在经过 U 变换的坐标系中（$|g\rangle, |e\rangle$ 是完整的 Hamiltonian 的本征态），$|\psi(t=0)\rangle = |g\rangle/A + |e\rangle/B$，一切都变得简单了起来. $|\psi(t)\rangle = \mathrm{e}^{-\lambda_- t/\hbar}|g\rangle/A + \mathrm{e}^{-\lambda_+ t/\hbar}|e\rangle/B$，其中必然有 $1/A^2 + 1/B^2 = 1$（这个验算可能会比较繁琐，Mathematica 结果见图 1）

```
In[1]:= H = {{h1, λ}, {λ, h2}};
        v = Eigenvalues[H];
        λ1 = v[[1]];
        λ2 = v[[2]];
        A = Sqrt[1 + ((λ1 - h1) / λ) ^ 2];
        B = Sqrt[1 + ((λ2 - h1) / λ) ^ 2];
        Simplify[1/A^2 + 1/B^2]

Out[7]= 1
```

图 1: Tedious Calculation Check via Mathematica

其 $|g^0\rangle, |e^0\rangle$ 的分量为

$$\langle g^0|\psi(t)\rangle = \mathrm{e}^{-\lambda_- t/\hbar}\frac{1}{A^2} + \mathrm{e}^{-\lambda_+ t/\hbar}\frac{1}{B^2} \tag{10}$$

其中，

$$A^2 = 1 + \left(\frac{\lambda_- - E_1}{\lambda}\right)^2, \quad B^2 = 1 + \left(\frac{\lambda_+ - E_1}{\lambda}\right)^2 \tag{11}$$

其模长的为

$$\begin{aligned} |\langle g^0|\psi(t)\rangle|^2 &= \frac{1}{A^4} + \frac{1}{B^4} + \frac{2}{A^2B^2}\cos((\lambda_+ - \lambda_-)t/\hbar) \\ &= 1 - \frac{2}{A^2B^2}(1 - \cos((\lambda_+ - \lambda_-)t/\hbar)) \end{aligned} \tag{12}$$

而容易知道（via Mathematica），

$$\lambda_+ - \lambda_- = 2\sqrt{\left(\frac{E_1 - E_2}{2}\right)^2 + \lambda^2}, \quad \frac{1}{A^2B^2} = \frac{\lambda^2}{(E_1 - E_2)^2 + 4\lambda^2} \tag{13}$$

于是

$$
\begin{aligned}
|\langle g^0|\psi(t)\rangle|^2 &= 1-\frac{2\lambda^2}{(E_1-E_2)^2+4\lambda^2}\left(1-\cos\left(2\sqrt{\left(\frac{E_1-E_2}{2}\right)^2+\lambda^2}\frac{t}{\hbar}\right)\right)\\
&= 1-\frac{4\lambda^2}{(E_1-E_2)^2+4\lambda^2}\sin^2\left(\sqrt{\left(\frac{E_1-E_2}{2}\right)^2+\lambda^2}\frac{t}{\hbar}\right)
\end{aligned}
\tag{14}
$$

)

可以做一定的物理分析：当 $E_2\to E_1$ 的时候（亦即简并情况），我们知道任何一个 Hermitian 的微扰之后的本征态都是 $(|g^0\rangle+|e^0\rangle)/\sqrt{2}$，$(|g^0\rangle-|e^0\rangle)/\sqrt{2}$. 在那种情况下，我们的原始态在新的态上的概率是一样的，换句话说，很有可能存在某些时候，两个振幅相同的新的态的 $|g^0\rangle$ 分量相干相消——而我们计算的解析结果也确实是这样，在那个时候，$\sin^2$ 前面的因子为 1，即确实存在某些时刻体系完全处于 $|e^0\rangle$ 上.

那么，一般的非简并问题是怎么处理的呢？首先，我们要定义一些算符. 这一段可能比较枯燥，但是当你看到它的效果之后就会震惊于我们之前进行的这些步骤是多么的重要.

我们首先（重新）定义我们在“基本概念[1479]”中就见到过的投影算符；不同的是，我们要对每一个态都定义如此一个投影算符：

$$
\hat{\mathcal{P}}_n=|\psi_n^0\rangle\langle\psi_n^0|,\quad \hat{\mathcal{Q}}_n=1-\hat{\mathcal{P}}_n \tag{15}
$$

显然有

$$
\hat{\mathcal{P}}_n\hat{\mathcal{Q}}_n=\hat{\mathcal{Q}}_n\hat{\mathcal{P}}_n=0,\quad [\hat{\mathcal{P}}_n,H_0]=[\hat{\mathcal{Q}}_n,H_0]=0 \tag{16}
$$

考虑定态的薛定谔方程 $(H_0+\lambda\hat{V})|\psi_n\rangle=E_n|\psi_n\rangle$，我们首先把 $|\psi_n\rangle$ 拆成与原本正交的部分和与原本重合的部分，也就是说

$$
|\psi_n\rangle=\hat{\mathcal{P}}_n|\psi_n\rangle+\hat{\mathcal{Q}}_n|\psi_n\rangle \tag{17}
$$

其中呢，第二项是很小的微扰项，需要去求解. 我们把薛定谔方程按照此思路重新展开. 先看左边，

$$
\begin{aligned}
\text{left} &= \hat{H}_0\hat{\mathcal{P}}_n|\psi_n\rangle+\lambda\hat{V}\hat{\mathcal{P}}_n|\psi_n\rangle+\hat{H}_0\hat{\mathcal{Q}}_n|\psi_n\rangle+\lambda\hat{V}\hat{\mathcal{Q}}_n|\psi_n\rangle\\
&= E_n^{(0)}\hat{\mathcal{P}}_n|\psi_n\rangle+\lambda\hat{V}\hat{\mathcal{P}}_n|\psi_n\rangle+\hat{H}_0\hat{\mathcal{Q}}_n|\psi_n\rangle+\lambda\hat{V}\hat{\mathcal{Q}}_n|\psi_n\rangle
\end{aligned}
\tag{18}
$$

而右边可以展开成

$$\text{right} = E_n\hat{\mathcal{P}}_n|\psi_n\rangle + E_n\hat{\mathcal{Q}}_n|\psi_n\rangle \tag{19}$$

这两个相等，也就是说

$$\begin{aligned} &E_n^{(0)}\hat{\mathcal{P}}_n|\psi_n\rangle + \lambda\hat{V}\hat{\mathcal{P}}_n|\psi_n\rangle + \hat{H}_0\hat{\mathcal{Q}}_n|\psi_n\rangle + \lambda\hat{V}\hat{\mathcal{Q}}_n|\psi_n\rangle \\ &= E_n\hat{\mathcal{P}}_n|\psi_n\rangle + E_n\hat{\mathcal{Q}}_n|\psi_n\rangle \end{aligned} \tag{20}$$

从而得到

$$(E_n - \hat{H}_0)\hat{\mathcal{Q}}_n|\psi_n\rangle = (E_n^{(0)} - E_n)\hat{\mathcal{P}}_n|\psi_n\rangle + \lambda\hat{V}|\psi_n\rangle \tag{21}$$

将 $\hat{\mathcal{Q}}$ 左乘式 21，我们得到

$$\begin{aligned} \hat{\mathcal{Q}}(E_n - \hat{H}_0)\hat{\mathcal{Q}}_n|\psi_n\rangle &= \hat{\mathcal{Q}}(E_n^{(0)} - E_n)\hat{\mathcal{P}}_n|\psi_n\rangle + \hat{\mathcal{Q}}\lambda\hat{V}|\psi_n\rangle \\ &= 0 + \hat{\mathcal{Q}}\lambda\hat{V}|\psi_n\rangle \end{aligned} \tag{22}$$

而显然 $\hat{\mathcal{Q}} = \hat{\mathcal{Q}}\hat{\mathcal{Q}}$，从而得到

$$\hat{\mathcal{Q}}(E_n - \hat{H}_0)\hat{\mathcal{Q}}_n\hat{\mathcal{Q}}|\psi_n\rangle = \lambda\hat{\mathcal{Q}}\hat{V}|\psi_n\rangle \tag{23}$$

含时 Time-dependent

问题分析：原子自发辐射

我们接下来分析原子自发辐射问题来回顾我们这一章学过的东西．问题背景：我们考虑部分量子化了的原子—光相互作用，其 Hamiltonian 可以用如下方程描述（已经经过部分简化，如偶极近似，旋波近似）：

$$\begin{aligned} H &= H_0 + H_{int}, \\ H_0 &= \sum_{\mathbf{k}} \hbar\omega_{\mathbf{k}} a_{\mathbf{k}}^\dagger a_{\mathbf{k}} + \frac{1}{2}\hbar\omega_{eg}\sigma_z, \\ H_{int} &= \hbar\sum_{\mathbf{k}} \Omega_{\mathbf{k}}\sigma_+ a_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot r} + \Omega_{\mathbf{k}}^*\sigma_- a_{\mathbf{k}}^\dagger \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot r}. \end{aligned} \tag{24}$$

其中，ω_{eg} 表示从激发态 $|e\rangle$ 跃迁到基态 $|g\rangle$ 的光子的角频率，$\hbar\Omega_{\mathbf{k}} \equiv -e\langle e|\mathbf{r}|g\rangle\cdot\hat{\epsilon}_{\mathbf{k}}\varepsilon_{\mathbf{k}}$ 总之就是一个频率量纲的东西并不重要，$a_{\mathbf{k}} = a_{\mathbf{k}}(t) = a_{\mathbf{k},0}\mathrm{e}^{-\mathrm{i}\omega t}$.

习题 2

我们试图引入相互作用汇景，

$$H^I(t) = \mathrm{e}^{\mathrm{i}H_0 t/\hbar} H_{int} e^{-\mathrm{i}H_0 t/\hbar} \tag{25}$$

明显，会带来 nontrivial 的项的是如下两项

$$\begin{aligned} \mathrm{e}^{\mathrm{i}\omega_{\mathbf{k}} a_{\mathbf{k}}^\dagger a_{\mathbf{k}} t} a_{\mathbf{k}} \mathrm{e}^{-\mathrm{i}\omega_{\mathbf{k}} a_{\mathbf{k}}^\dagger a_{\mathbf{k}} t} &= a_{\mathbf{k}} \mathrm{e}^{-\mathrm{i}\omega_{\mathbf{k}} t} \\ \mathrm{e}^{i\omega_{eg}\sigma_z t/2} \sigma_+ \mathrm{e}^{\mathrm{i}\omega_{eg}\sigma_z t/2} &= \sigma_+ \mathrm{e}^{-\mathrm{i}\omega_{eg} t} \end{aligned} \tag{26}$$

试推导这两个式子，并进而写出 Hamiltonian $H^I(t)$.

我们有

$$H^I(t) = \hbar \sum_{\mathbf{k}} \left[\Omega_{\mathbf{k}} \sigma_+ a_{\mathbf{k}} \mathrm{e}^{\mathrm{i}(\omega_{eg} - \omega_{\mathbf{k}})t} + h.c. \right] \tag{27}$$

角动量 2

角动量问题包含一般的轨道角动量，和更有趣点的自旋，两者能产生磁矩，却有着不一样的产生系数（g 因子）. 当然，如果不考虑磁效应的话，角动量有自己的一般理论；但是如果考虑磁效应，利用角动量来解决很多磁学问题是很常见的，一种极端简化的模型就是 Ising Model 和 Heisenberg Model，会在讨论磁学问题的时候详细分析.

一般理论 General Ideas

角动量的问题其实相对来说更独立一点. 它有自己的一套体系. 本部分将从基本的数学结构开始建立角动量理论，其中会逐步的引入高级一点的数学描述（群描述），虽然这部分知识并不难，但是却很多时候并不容易理解.

经典角动量的直观推广

很容易证明，经典的角动量 $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ 对应的算符是一个很个性的算符：它在直角坐标中的三个分量都不对易，而且它们三个都是平权的. 我们知道，能够同时确定的物理量必须满足：它们有共同的本征态，换句话说，它们彼此对易. 因此，我们在这三个分量里面只能有一个能确定. 既然平权，我们不妨

取为 z 方向分量．计算仍然给出这三个方向的角动量分量算符，和总角动量的平方这个标量算符也是对易的[12]．

写出来就是

$$[\hat{L}_z, \hat{\mathbf{L}}^2] = 0 \tag{1}$$

$$[\hat{L}_a, \hat{L}_b] = \mathrm{i}\hbar\epsilon_{abc}L_c \tag{2}$$

其中，后者是个标量算符．本章在后面很可能习惯性的把 $\hbar$ 略去，请注意．

$$\hat{\mathbf{L}}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2 \tag{3}$$

于是，如果仅在角动量框架内的话，我们可以取可以同时确定的本征态 $|l, m\rangle$，其中

$$\hat{\mathbf{L}}^2|l, m\rangle = l(l+1)|\hbar^2 l, m\rangle \tag{4}$$

$$\hat{L}_z|l, m\rangle = m\hbar|l, m\rangle \tag{5}$$

其中呢，$l(l+1)$ 只是写起来比较带感而已，至此我们没给出任何更具体的关于本征值的结论．

习题 1

考虑到角动量算符的定义，

$$\hat{\mathbf{L}} = \frac{1}{2}\left(\hat{\mathbf{p}} \times \hat{\mathbf{r}} - \hat{\mathbf{r}} \times \hat{\mathbf{p}}\right) \tag{6}$$

(1) 请验证式 1 和式 2，其中需要计算

$$[\hat{L}_a, \hat{r}_b] =? \quad [\hat{L}_a, \hat{p}_z] =? \tag{7}$$

(2) 计算矢量算符

$$\hat{\mathbf{L}} \times \hat{\mathbf{L}} =? \tag{8}$$

可以像前面处理一维谐振子那样处理体系，我们定义**阶梯算符**为

$$\hat{L}_\pm = \hat{L}_x \pm i\hat{L}_y \tag{9}$$

[12]这一段里的结论的验算可以参考后面的练习．

这个算符有什么好处呢？我们来做一些简单的计算．一个重要的关系就是

$$[\hat{L}_z, \hat{L}_\pm] = \pm \hat{L}_\pm \tag{10}$$

这个的推导很简单，可以自己完成．我们要讲的，是 $L_\pm|l,m\rangle$ 这个东西，它也是 $L_z, \mathbf{L}^2$ 的本征态．前者很简单，我们只需要计算下面这个式子：

$$L_z L_\pm |l,m\rangle = (L_\pm L_z \pm L_\pm)|l,m\rangle = (m \pm 1) L_\pm |l,m\rangle \tag{11}$$

由此我们可以看出，很有可能

$$L_\pm |l,m\rangle \stackrel{?}{=} c|l, m\pm 1\rangle \tag{12}$$

其中 c 是一个待确定的数，可以明确的知道它和 l,m 有关．只需要确定它是不是 $\mathbf{L}^2$ 的本征值为 l 的本征态就行了．我们可以去计算

$$\mathbf{L}^2 L_\pm |l,m\rangle \tag{13}$$

这次，我们得先做点工作：我们直接重新写 $\mathbf{L}^2$ 的表达形式：

$$\mathbf{L}^2 = L_z(L_z+1) + L_- L_+ \tag{14}$$

由式 10，很容易证明这个写法满足

$$\begin{aligned} \mathbf{L}^2 L_\pm |l,m\rangle &= L_z(L_z+1)L_\pm|l,m\rangle + L_-L_+L_\pm|l,m\rangle \\ &= (m\pm 1)(m+1\pm 1)L_\pm|l,m\rangle + L_-L_+L_\pm|l,m\rangle \end{aligned} \tag{15}$$

验算得到

$$[\hat{L}_+, \hat{L}_-] = 2\hat{L}_z \tag{16}$$

由此，我们分别验算 L_+, L_-．先看 L_+

$$L_-L_+L_+ = (L_+L_- + 2L_z)L_+ = L_+(L_-L_+) - 2L_zL_+ \tag{17}$$

从而

$$\begin{aligned} &\mathbf{L}^2 L_+|l,m\rangle \\ &= (m+1)(m+2)L_+|l,m\rangle + (L_+(L_-L_+) - 2L_zL_+)|l,m\rangle \\ &= (m+1)(m+2)L_+|l,m\rangle + \left(L_+(\mathbf{L}^2 - L_z(L_z+1)) - 2(m+1)L_+\right)|l,m\rangle \\ &= \left[(m+1)(m+2) + l(l+1) - m(m+1) - 2(m+1)\right] L_+|l,m\rangle \\ &= l(l+1)L_+|l,m\rangle \end{aligned} \tag{18}$$

另一方面，L_- 满足

$$L_-L_+L_- = L_-(L_-L_+ - 2L_z) = L_-(L_-L_+) - 2L_-L_z \tag{19}$$

从而

$$\begin{aligned}
&\mathbf{L}^2 L_-|l,m\rangle \\
&= m(m-1)L_-|l,m\rangle + (L_-(L_-L_+) + 2L_-L_z)|l,m\rangle \\
&= m(m-1)L_-|l,m\rangle + \left(L_-(\mathbf{L}^2 - L_z(L_z+1)) + 2mL_-\right)|l,m\rangle \\
&= \left[m(m-1) + l(l+1) - m(m+1) + 2m\right] L_-|l,m\rangle \\
&= l(l+1)L_-|l,m\rangle
\end{aligned} \tag{20}$$

啊，总之，我们得到了我们喜欢的式子

$$L_\pm|l,m\rangle = c|l,m\pm1\rangle \tag{21}$$

至于求 c 如何求，这个还是很有技巧的，我们作为练习引导大家一步步求出来．请参考习题 2，结论为

$$L_\pm|l,m\rangle = \sqrt{l(l+1) - m(m\pm1)}|l,m\pm1\rangle \tag{22}$$

接下来我们看一个实际的例子，理解一下我们干了什么．

例 1

有一个体系，哈密顿量为

$$\hat{H} = \hat{L}_x^2 + \hat{L}_y^2 - \hat{L}_z = \sum_{l,m} E_0 \hat{L}_-\hat{L}_+ \tag{23}$$

E_0 是常数，它的本征态显然为各个角动量态 $|l,m\rangle$，假设每个态的能量为 $E_{l,m}$，可以写成

$$\hat{H} = \sum_{l,m} E_{l,m}|l,m\rangle\langle l,m| \tag{24}$$

而通过插入 resolution of identity，哈密顿量可以写成

$$\begin{aligned}
\hat{H} &= \sum_{l,m} E_0 \hat{L}_-\hat{L}_+ = \sum_{l,m} E_0 \hat{L}_-\hat{L}_+|l,m\rangle\langle l,m| \\
&= \sum_{l,m} E_0 \left[l(l+1) - m(m+1)\right]|l,m\rangle\langle l,m|
\end{aligned} \tag{25}$$

从而轻易得出

$$E_{l,m} = E_0\left[l(l+1) - m(m+1)\right] \tag{26}$$

角动量叠加原理

这一部分的内容实际上可以说是不涉及高深数学知识的角动量的最复杂的知识了.

习题 2

如何求式 21 中的系数 c? 换句话说，如何求

$$\langle l, m \pm 1 | L_{\pm} | l, m \rangle \tag{27}$$

这实际上还包含一个被称为 Condon-Shortley convention 的事情在影响这个 c. 它叫 convention 是说明它确实有一些待确定的自由度，就是这个 c 的相位. 我们后面会看到.

注意到，由算符的 Hermitianity，

$$\langle l, m | L_{\pm}^{\dagger} | l, m \pm 1 \rangle = \langle l, m | L_{\mp} | l, m \pm 1 \rangle = c^* \tag{28}$$

即

$$L_{\mp} | l, m \pm 1 \rangle = c^* | l, m \rangle \tag{29}$$

或者说

$$L_{\mp} L_{\pm} | l, m \rangle = c^* c | l, m \rangle \tag{30}$$

这一下就简单了，因为我们注意到，在 $\mathbf{L}^2$ 里面也有类似的项. 分开来看:

$$L_- L_+ = \mathbf{L}^2 - L_z(L_z + 1) \tag{31}$$

于是

$$L_+ | l, m \rangle = c | l, m+ \rangle, |c|^2 = l(l+1) - m(m+1) \tag{32}$$

类似的，

$$L_+ L_- = L_- L_+ + 2L_z = \mathbf{L}^2 - L_z(L_z - 1) \tag{33}$$

于是

$$L_- | l, m \rangle = c | l, m-1 \rangle, |c|^2 = l(l+1) - m(m-1) \tag{34}$$

著名的 Condon-Shortley convention 规定，

$$c = \sqrt{l(l+1) - m(m \pm 1)} \tag{35}$$

从而

$$L_{\pm} | l, m \rangle = \sqrt{l(l+1) - m(m \pm 1)} | l, m \pm 1 \rangle \tag{36}$$

磁学问题 Magnetic Problems

这部分的问题，一定程度上参考了 Auerbach 的书[13].

冷原子基本知识

研究冷原子主要是两方面（其中后者更重要）：原子的内部结构，和原子原子之间的相互作用.

首先研究原子内部结构，我们可以先写出单原子的 Hamiltonian,

$$\begin{aligned}\hat{H}_{at} = &-\sum_i \left(\frac{\hbar^2\nabla_i^2}{2\mu_i} + V_{ei}(\mathbf{r})\right) + \sum_{i<j} V_{ee}(\mathbf{r}_i - \mathbf{r}_j) \\ &+ \sum_i [\alpha_f \mathbf{S}_i \cdot \mathbf{L}_i + \alpha_{hf}(\mathbf{S}_i + \mathbf{L}_i) \cdot \mathbf{I}_i]\end{aligned} \tag{1}$$

如果只考虑第一项，也就是说每个电子之间是没有相互作用的，得到的是很经典的能级，我们中学就学过. 第二项的作用是给了一个修正，当然一般的表达式并不容易写出来，但是我们知道能量的修正可以写成如下形式

$$E = -\frac{\mu\alpha^2}{2\hbar^2(n+\Delta)^2}, \quad \Delta = \Delta(n,l), \quad \alpha = \frac{e^2}{4\pi\varepsilon_0} \tag{2}$$

除此之外，第二项还能提供一个我们所知的 Hund 规则，单重态 $\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2) - \psi_1(\mathbf{r}_2)\psi_2(\mathbf{r}_1)$ 在离的很近的时候波函数是基本没有的，也就是提供了一个 repulsive 的能量；这也就是说三重态相比单重态能量更低.

第三项，正如他的名字所暗示的，包含着精细结构（fine）和超精细结构（hyperfine）

原子物理里面，一个很基本的概念就是 Zeeman 效应，这是实验上观测到，并在后来基于量子理论建立了正确的描述；很通俗而且愚蠢的 Zeeman 项（系数）的解释是：相对的，原子核（带电）绕电子转圈，引发磁场，与电子的自旋相互作用.

$$E \propto -\boldsymbol{\mu}_s \cdot \mathbf{B}_{\text{eff}} = \frac{Ze^2\mu_0}{8\pi m_0^2 r^3}(\mathbf{S}\cdot L) \tag{3}$$

[13] A. Auerbach, *Interacting Electrons and Quantum Magnetism*, Springer Science & Business Media, 2012.

这一项相比与之前大概是 10^{-3} 的量级；而电子整体的角动量与核自旋的相互作用是超精细结构，相比精细结构是 10^{-3} 的量级.

物理学对原子进行分类，主要有三类（实际上有很多更精细的分化，但在冷原子范畴内我们并不关心原子形成的晶体的结构：我们希望在极低温度的情况下，原子仍然能保证不会结晶而是气态；这也是冷原子的密度极低，十分稀薄从而少体散射就适用的一个原因）：

碱金属：锂钠钾铷铯钫. $^{2S+1}L_J$ 表示的话，这些原子的基态都是 $^2S_{1/2}$，价电子都是 ns^1，区别主要体现在核自旋上（不同的同位素）. ^{87}Rb,^{23}Na 是最常见最爱研究的玻色气体，而 ^{40}K,^{6}Li 则是最常见的费米气体. Fermion or Boson 主要看核自旋是整数还是半整数.

作为例子看一个问题：

^{87}Rb,^{6}Li 在磁场 $\mathbf{B}=B\hat{\mathbf{z}}$ 中，Hamiltonian 可以写成

$$\hat{H}=B(\mu_B J_z+\mu_N I_z)+\alpha(\mathbf{J}\cdot I) \tag{4}$$

这个的处理方法实际上和传统的 LS 耦合基本一样：引入一个新的参数 $\mathbf{F}=J+I$，从而保证

$$\mathbf{J}\cdot I=\frac{1}{2}(\mathbf{F}^2-\mathbf{J}^2-\mathbf{I}^2)=\frac{1}{2}(F(F+1)-J(J+1)-I(I+1)) \tag{5}$$

从而能谱就是

$$E=B(\mu_B J_z+\mu_N I_z)+\frac{\alpha}{2}(F(F+1)-J(J+1)-I(I+1)) \tag{6}$$

除此之外，我们还可以看到另外的东西. 这个 Hamiltonian 不好解，但是我们可以用 $|F,m_F\rangle$ 来展开，并且把超精细结构的项作为 perturbation 来处理. $^{87}\text{Rb}: I=\frac{3}{2};\ ^6\text{Li}: I=1; J=\frac{1}{2}$

对于 ^{87}Rb 来说，能够得到的 $F=2,1$. 其本征态有两系列 $F=2:|2,0\rangle,|2,\pm1\rangle,|2,\pm2\rangle$ $F=1:|1,0\rangle,|1,\pm1\rangle$

但是我们知道，$|F,m_F\rangle$ 必须利用 CG 系数知道它实际上是如何被组合成的，才能用我们的 Hamiltonian 计算.

$$|2,2\rangle=\left|\frac{3}{2},\frac{3}{2}\right\rangle\left|\frac{1}{2},\frac{1}{2}\right\rangle \tag{7}$$

$$|2,1\rangle = \frac{1}{2}\left|\frac{3}{2},\frac{3}{2}\right\rangle\left|\frac{1}{2},-\frac{1}{2}\right\rangle + \frac{\sqrt{3}}{2}\left|\frac{3}{2},\frac{1}{2}\right\rangle\left|\frac{1}{2},\frac{1}{2}\right\rangle \tag{8}$$

$$|2,0\rangle = \frac{\sqrt{2}}{2}\left|\frac{3}{2},\frac{1}{2}\right\rangle\left|\frac{1}{2},-\frac{1}{2}\right\rangle + \frac{\sqrt{2}}{2}\left|\frac{3}{2},-\frac{1}{2}\right\rangle\left|\frac{1}{2},\frac{1}{2}\right\rangle \tag{9}$$

$$|2,-1\rangle = \frac{1}{2}\left|\frac{3}{2},-\frac{3}{2}\right\rangle\left|\frac{1}{2},\frac{1}{2}\right\rangle + \frac{\sqrt{3}}{2}\left|\frac{3}{2},-\frac{1}{2}\right\rangle\left|\frac{1}{2},-\frac{1}{2}\right\rangle \tag{10}$$

$$|2,-2\rangle = \left|\frac{3}{2},-\frac{3}{2}\right\rangle\left|\frac{1}{2},-\frac{1}{2}\right\rangle \tag{11}$$

$$|1,1\rangle = \frac{\sqrt{3}}{2}\left|\frac{3}{2},\frac{3}{2}\right\rangle\left|\frac{1}{2},-\frac{1}{2}\right\rangle - \frac{1}{2}\left|\frac{3}{2},\frac{1}{2}\right\rangle\left|\frac{1}{2},\frac{1}{2}\right\rangle \tag{12}$$

$$|1,0\rangle = \frac{\sqrt{2}}{2}\left|\frac{3}{2},\frac{1}{2}\right\rangle\left|\frac{1}{2},-\frac{1}{2}\right\rangle - \frac{\sqrt{2}}{2}\left|\frac{3}{2},-\frac{1}{2}\right\rangle\left|\frac{1}{2},\frac{1}{2}\right\rangle \tag{13}$$

$$|1,-1\rangle = -\frac{\sqrt{3}}{2}\left|\frac{3}{2},-\frac{3}{2}\right\rangle\left|\frac{1}{2},\frac{1}{2}\right\rangle + \frac{1}{2}\left|\frac{3}{2},-\frac{1}{2}\right\rangle\left|\frac{1}{2},-\frac{1}{2}\right\rangle \tag{14}$$

这样写出来的 Hamiltonian 是一个 8×8 的矩阵，当然大部分矩阵元都是 0，但是对角项和一些部分也不是 0．利用一些技巧，我们知道只有拥有同样的 m_F 的两个态之间才有矩阵元，凭借惊人的毅力，我们把这个矩阵写出来，如图 1 （本人的计算功底很差…如果算错了不要打我[14]）：

$$\begin{pmatrix}
(\frac{3}{2}\mu_B+\frac{1}{2}\mu_N)B+A_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & (\frac{3}{4}\mu_B+\frac{1}{4}\mu_N)B+A_1 & 0 & 0 & 0 & \frac{\sqrt{3}}{4}(\mu_B-\mu_N)B & 0 & 0 \\
0 & 0 & A_1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -(\frac{3}{4}\mu_B+\frac{1}{4}\mu_N)B+A_1 & 0 & 0 & 0 & \frac{\sqrt{3}}{4}(\mu_B-\mu_N)B \\
0 & 0 & 0 & 0 & -(\frac{3}{2}\mu_B+\frac{1}{2}\mu_N)B+A_1 & 0 & 0 & 0 \\
0 & \frac{\sqrt{3}}{4}(\mu_B-\mu_N)B & 0 & 0 & 0 & (\frac{3}{4}\mu_B+\frac{1}{4}\mu_N)B+A_2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & A_2 & 0 \\
0 & 0 & 0 & \frac{\sqrt{3}}{4}(\mu_B-\mu_N)B & 0 & 0 & 0 & -(\frac{3}{4}\mu_B+\frac{1}{4}\mu_N)B+A_2
\end{pmatrix}$$

图 1: 算错的矩阵

$$\begin{aligned}
A_1 &\equiv \frac{3}{4}\alpha = \frac{\alpha}{2}\left(2\times 3 - \frac{3}{2}\times\frac{5}{2} - \frac{1}{2}\times\frac{3}{2}\right) \\
A_2 &\equiv -\frac{5}{4}\alpha = \frac{\alpha}{2}\left(1\times 2 - \frac{3}{2}\times\frac{5}{2} - \frac{1}{2}\times\frac{3}{2}\right)
\end{aligned} \tag{15}$$

发现，只有在 $|2,1\rangle, |1,1\rangle$ 和 $|2,-1\rangle, |1,-1\rangle$ 之间存在 off-diagonal 成分．相

[14]注意：确实发现算错了，但是懒得改，领会精神 orz

当于我们需要对角化如下两个 $2\times$ 矩阵：

$$\begin{pmatrix} \left(\frac{3}{4}\mu_B+\frac{1}{4}\mu_N\right)B+A_1 & \frac{\sqrt{3}}{4}(\mu_B-\mu_N)B \\ \frac{\sqrt{3}}{4}(\mu_B-\mu_N)B & \left(\frac{3}{4}\mu_B+\frac{1}{4}\mu_N\right)B+A_2 \end{pmatrix}, \\ \begin{pmatrix} -\left(\frac{3}{4}\mu_B+\frac{1}{4}\mu_N\right)B+A_1 & \frac{\sqrt{3}}{4}(\mu_B-\mu_N)B \\ \frac{\sqrt{3}}{4}(\mu_B-\mu_N)B & -\left(\frac{3}{4}\mu_B+\frac{1}{4}\mu_N\right)B+A_2 \end{pmatrix} \tag{16}$$

得到这两部分的本征值分别应该为

$$\lambda_{1,2/1}=-\left(\frac{3}{4}\mu_B-\frac{1}{4}\mu_N\right)B+\frac{A_1+A_2}{2}\pm\sqrt{\frac{1}{4}(A_1-A_2)^2+\frac{3}{4}(\mu_B-\mu_N)^2B^2} \tag{17}$$

$$\lambda_{-1,2/1}=\left(\frac{3}{4}\mu_B+\frac{1}{4}\mu_N\right)B+\frac{A_1+A_2}{2}\pm\sqrt{\frac{1}{4}(A_1-A_2)^2+\frac{3}{4}(\mu_B-\mu_N)^2B^2} \tag{18}$$

这两个本征值的行为在 B 非常小和相对大一点的时候有着截然不同的表现．在很小的时候，也就是说 $B\ll|A_1-A_2|/\sqrt{3}(\mu_B-\mu_N)$，我们可以把根号内进行展开，得到

$$\sqrt{\frac{1}{4}(A_1-A_2)^2+\frac{3}{4}(\mu_B-\mu_N)^2B^2}\approx\frac{1}{2}|A_1-A_2|\left(1+\frac{3}{2}\frac{(\mu_B-\mu_N)^2}{(A_1-A_2)^2}B^2\right) \tag{19}$$

得到了一个 Zeeman effect 的二次 shift．而大的磁场下超精细结构 A_1,A_2 很小，几乎忽略，主要修正项都是 $\frac{\sqrt{3}}{2}\Delta\mu B$，是线性的 Zeeman shift．

当然这只是一个关于二次 Zeeman shift 的例子，^{6}Li 同理，有兴趣的同学可以进行练习．

碱金属原子的另一个特点就是明显的第一激发态的劈裂．第一激发态本身应该是 $L=1$，也就是 ^2P．但是如果有了 LS 耦合，不同的 J 会有不同的精细结构劈裂．最简单的就是大家都实验过的 Na 黄光双线．当然如果考虑超精细结构还会更进一步分裂，但是那个效应太小不容易观察到．

要注意，在基态和这两个激发态之间的光跃迁是实验上对（冷）原子进行捕捉和操作的基本手段．冷的目的是他们基本没有很大的动能，这样就少去了很多的复杂因素（反冲能量损失、多普勒效应等）．

最后一类有趣的原子就是高自旋原子．不过我并不对这个感兴趣，就先不写简介了．

接下来讨论一下上面提到的光跃迁过程——说白了就是一个原子在偶极作用下的行为．

光引入的附加 Hamiltonian 是：

$$\hat{H}_d = \mathbf{d} \cdot E = d_j E_j^0 \cos(\omega t - \phi_j) \tag{20}$$

其中 $\mathbf{d} = \sum_\alpha (-e\mathbf{r}_\alpha)$，代表原子的偶极矩.

比如，我们研究碱金属的问题.相对的，我们研究的是从 $^2S_{1/2}$ 到 $^2P_{3/2}$, $^2P_{1/2}$ 的跃迁．两个激发态可以用 D_1, D_2 表示，能级差是有精细机构带来的；一个数量级上的能量估计是：$\Delta_{sp} \gg \Delta_f \gg \Delta_{hf}$．显然，这里我们可以忽略掉超精细结构.

我们定义投影算符，把态投影到基态或者激发态上．显然这里可以使用

$$\mathcal{P}_g = \frac{\mathbf{L}^2}{2\hbar}, \quad \mathcal{P}_e = 1 - \mathcal{P}_g \tag{21}$$

而体系的 Hamiltonian 在无外场的时候则是

$$\hat{H} = \hat{H}_{at} + \hat{H}_d \tag{22}$$

我们前面分析了，无外场

$$\hat{H}_{at} = E_e \mathcal{P}_e + \frac{A_f}{\hbar^2} \mathbf{L} \cdot \mathbf{S} \tag{23}$$

这个解释就是：我们把基态能量扔了，而且不去管那些关于主量子数 n 的细节；在我们需要考虑的范围内，Hamiltonian 就可以写成你在激发态上的成分乘以激发态的能量 E_e．显然，这个 Hamiltonian 并不包含在基态和激发态之间的任何 coupling.

我们定义一个算符

$$\mathcal{U}_{rot}(t) = \mathrm{e}^{-\mathrm{i}\omega t \mathcal{P}_e} = \mathcal{P}_g + \mathcal{P}_e \mathrm{e}^{-\mathrm{i}\omega t} \tag{24}$$

这算符相当于做一个 unitary 的变换[15]，我们现在看外场怎样变换

$$\hat{H}'_d = \mathcal{U}^\dagger_{rot}(t) \hat{H}_d \mathcal{U}_{rot} \tag{26}$$

[15]实际上这就类似相互作用绘景里面常用的样子给反过来而已

$$U_I = \mathrm{e}^{\mathrm{i}\hat{H}_0 t/\hbar} = \mathrm{e}^{\mathrm{i}\omega_e t \mathcal{P}_e} \tag{25}$$

显然的，我们知道无论基态还是激发态都没有特殊的极化方向，$\langle d_j\rangle = 0$，即

$$\mathcal{P}_g d_j \mathcal{P}_g = \mathcal{P}_e d_j \mathcal{P}_e = 0 \tag{27}$$

而另一方面，E_j 又是外场，不与投影算符 $\mathcal{P}$ 有对易关系．因此，我们很容易的可以进行一步化简：

$$\hat{H}'_d = E_j^0 \cos(\phi_j - \omega t)(\mathcal{P}_g d_j \mathcal{P}_e \mathrm{e}^{-\mathrm{i}\omega t} + \mathcal{P}_e \mathrm{e}^{\mathrm{i}\omega t} d_j \mathcal{P}_g) \tag{28}$$

而

$$\cos(\phi_j - \omega t) = \frac{\mathrm{e}^{\mathrm{i}(\phi_j - \omega t)} + \mathrm{e}^{-\mathrm{i}(\phi_j - \omega t)}}{2} \tag{29}$$

定义 $\tilde{E}_j = E_j^0 \mathrm{e}^{\mathrm{i}\phi_j}$，我们得到

$$\hat{H}'_d = \frac{1}{2}\left(\tilde{E}_j \mathrm{e}^{-\mathrm{i}\omega t} + \tilde{E}_j^* \mathrm{e}^{\mathrm{i}\omega t}\right)(\mathcal{P}_g d_j \mathcal{P}_e \mathrm{e}^{-\mathrm{i}\omega t} + \mathcal{P}_e \mathrm{e}^{\mathrm{i}\omega t} d_j \mathcal{P}_g) \tag{30}$$

在近似共振的时候 $\omega \sim \omega_e = E_e/\hbar$，某种程度上可以忽略到以 $\pm 2\mathrm{i}\omega t$ 做震荡的成分（被称为 Rotating Wave Approximation RWA[16]），于是我们得到

$$\hat{H}'_d \approx \frac{1}{2}\left(\tilde{E}_j \mathcal{P}_e d_j \mathcal{P}_g + \tilde{E}_j^* \mathcal{P}_g d_j \mathcal{P}_e\right) \tag{31}$$

因为包含精细结构 $A_f \mathbf{L}\cdot\mathbf{S}/\hbar^2$ 的项已经很小了，其在这个 Unitary 变换下的变化可以忽略．而另一方面，一个 unitary 的变换，如果是含时的，则会带来一个变化，是由于要保证 Hamiltonian 的定义正确[17]

$$|\psi\rangle \to \mathcal{U}(t)|\psi\rangle = |\psi'\rangle \tag{32}$$

$$H'|\psi'\rangle = \mathrm{i}\hbar\frac{d}{dt}|\psi'\rangle = \mathrm{i}\hbar\frac{d\mathcal{U}(t)}{dt}|\psi\rangle + \mathrm{i}\hbar\mathcal{U}(t)\frac{\mathrm{d}}{\mathrm{d}t}|\psi\rangle \tag{33}$$

而 $|\psi\rangle = \mathcal{U}^\dagger(t)|\psi'\rangle$，故

$$H'|\psi\rangle = \left(\mathrm{i}\hbar\frac{d\mathcal{U}(t)}{dt}\mathcal{U}^\dagger(t) + \mathcal{U}(t)H\mathcal{U}^\dagger(t)\right)|\psi'\rangle \tag{34}$$

[16]适用条件：

- 场几乎共振，能级差 $E_e \sim \hbar\omega$
- 场很弱，具体的话要满足 $E_j d \ll \hbar\omega$

[17]注意：这里我的这个 note 写的比较乱，但是我觉得可以阅读；我也懒得改了．

从而

$$H' = \mathrm{i}\hbar\frac{d\mathcal{U}(t)}{dt}\mathcal{U}^\dagger(t) + \mathcal{U}(t)H\mathcal{U}^\dagger(t) \tag{35}$$

$$\hat{H}'_{at} = E_e\mathcal{P}_e - \hbar\omega\mathcal{P}_e + \frac{A_f}{\hbar^2}\mathbf{L}\cdot\mathbf{S} \tag{36}$$

我们研究对基态的二阶微扰．微扰理论的基本知识可以参考 Prof Fa, Wang 的讲义，见 Fa Wang[18].

能量的修正应该是

$$\Delta E = \langle g|H'_d|g\rangle + \frac{\langle g|H'_d|e\rangle\langle e|H'_d|g\rangle}{0 - E'_e} \tag{37}$$

可以一目了然的看出来，H'_d 这种左边右边有不同的投影算符 $\mathcal{P}$ 的，第一项为 0；而第二项可以简化成

$$\Delta E = -\langle g|H'_d|e\rangle H'^{-1}_{at}\langle e|H'_d|g\rangle \tag{38}$$

因为这里我们实际上扔了一个算符进去，并不应该写成这样子，而应该写成一个仅对基态有用的有效 Hamiltonian：

$$H_{\text{eff}} = -\mathcal{P}_g H'_d\mathcal{P}_e H'^{-1}_{at}\mathcal{P}_e H'_d\mathcal{P}_g \sim -\mathcal{P}_g H'_d H'^{-1}_{at} H'_d\mathcal{P}_g \tag{39}$$

其中近似是由于：我们只希望得到二阶的修正，所以 H'_{at} 基本只作用在激发态上，两个对激发态的投影算符可以忽略了．

来咱们继续，虽然发生了很多令人伤心的事情但是冷原子的学习还是要继续的．

我们定义一个二阶矩阵（算符）：

$$\mathcal{D}_{ij} = \mathcal{P}_g d_i \hat{H}^{-1}_{at} d_j \mathcal{P}_g \tag{40}$$

这样的话，考虑到

$$\begin{aligned}\hat{H}_{\text{eff}} &= -\mathcal{P}_g\hat{H}'_d\hat{H}'^{-1}_{at}\hat{H}'_d\mathcal{P}_g \\ &= -\frac{1}{4}\mathcal{P}_g\left(\tilde{E}_j\mathcal{P}_e d_j\mathcal{P}_g + \tilde{E}^*_j\mathcal{P}_g d_j\mathcal{P}_e\right)\hat{H}'^{-1}_{at}\left(\tilde{E}_i\mathcal{P}_e d_i\mathcal{P}_g + \tilde{E}^*_i\mathcal{P}_g d_i\mathcal{P}_e\right)\mathcal{P}_g\end{aligned} \tag{41}$$

[18]Fa Wang, Summary of Lecture 6: perturbation theory, laserroger: Site Part C, 2nd-order perturbation

显然，红色的两项包含诸如 $\mathcal{P}_e\mathcal{P}_g$, $\mathcal{P}_g\mathcal{P}_e$ 这种（$\tilde{E}$ 这种外场在这时候和他们都没有对易关系）的，显然 $=0$，所以最后我们有对基态的等效 Hamiltonian

$$\hat{H}_{\text{eff}} = -\frac{1}{4}\mathcal{P}_g \sum_j \tilde{E}_j^* \mathcal{P}_g d_j \mathcal{P}_e \hat{H}'^{-1}_{at} \sum_i \tilde{E}_i \mathcal{P}_e d_i \mathcal{P}_g \mathcal{P}_g \tag{42}$$

或者，利用我们刚刚定义的二阶矩阵，我们有

$$\hat{H}_{\text{eff}} = -\frac{1}{4}\sum_{i,j} \tilde{E}_i^* \mathcal{D}_{ij} \tilde{E}_j \tag{43}$$

可以把这个 $\mathcal{D}_{ij}$ 拆成三个部分：表示 trace 的各项同性矩阵 $\mathcal{D}^0_{ij} = \frac{1}{3}Tr(\mathcal{D})\delta_{ij}$，全反对称的 $\mathcal{D}^1_{ij} = \frac{1}{2}(\mathcal{D}_{ij} - \mathcal{D}_{ji})$，和全对称无 trace 的 $\mathcal{D}^2_{ij} = \frac{1}{2}(\mathcal{D}_{ij} + \mathcal{D}_{ji}) - \mathcal{D}^0_{ij}$，于是

$$\mathcal{D}_{ij} = \mathcal{D}^0_{ij} + \mathcal{D}^1_{ij} + \mathcal{D}^2_{ij} \tag{44}$$

我们接下来分析一下这种效应带来的 splitting，作为我们对于原子—光相互作用的结束.

1. 极端情况下，可以忽略精细结构，也就是说 $A_f = 0$. 这时候 $\mathcal{D}_{ij}$ 就不再是一个很复杂的算符，而是简单地

$$\mathcal{D}_{ij} = \frac{1}{\Delta_e}\mathcal{P}_g d_i d_j \mathcal{P}_g \tag{45}$$

主要到这里两个到基态的投影算符使得中间的位置算符 d_j 并没有明显的极化效应，所以 $\mathcal{D}_{ij} \propto \delta_{ij}$，$\mathcal{D}^1_{ij} = \mathcal{D}^2_{ij} = 0$.

而

$$\mathcal{D}^0_{ij} = \frac{1}{3\Delta_e}\mathcal{P}_g d_j \mathcal{P}_e \mathcal{P}_e d_j \mathcal{P}_g \delta_{ij} = \frac{1}{3\Delta_e}|\mathcal{P}_g d_j \mathcal{P}_e|^2 \delta_{ij} \overset{def}{\equiv} -4u_s\delta_{ij} \tag{46}$$

其中，很显然的可以得到

$$u_s = -\frac{1}{12\Delta_e}|\langle l=0|\mathbf{d}|l=1\rangle|^2 \tag{47}$$

因此，有效 Hamiltonian 为

$$\hat{H}_{\text{eff}} = u_s \mathbf{E}^2 \tag{48}$$

这说明什么呢？这说明没有精细结构的话，**光是没有办法改变电子的自旋的（Hamiltonian 里面没有包含自旋与其他参数耦合的项）**；而且，这种作用并

不依赖光的偏振，而仅仅和光的强度有关（$\mathbf{E}^2$）；利用这点可以很好的进行实验上的**光陷阱，从而捕捉原子**；而 Δ_e 实际上与原子的速度有关（Doppler 效应），于是说给激光制冷提供了一个实验的理论指导.

2. 考虑精细结构（当然很小的精细结构）的时候的能级分裂

注意到，

$$\hat{H}^{-1} \equiv (\hat{H}_0 + \hat{V})^{-1} = \hat{H}_0^{-1} - \hat{H}_0^{-1}\hat{V}\hat{H}^{-1} \tag{49}$$

[19]我们可以得到

$$\hat{H}_{at} = \frac{1}{\Delta_e}\mathcal{P}_e - \frac{1}{\Delta_e}\frac{A_f}{\hbar^2}\mathbf{L}\cdot\mathbf{S}\hat{H}_{at}^{-1} \tag{51}$$

从而，我们可以分出来和之前一样的 $\mathcal{D}_{ij}$ 部分，和精细结构带来的效应

$$\mathcal{D}_{ij} = \mathcal{P}_g d_i \frac{1}{\Delta_e} d_j \mathcal{P}_g - \alpha \mathcal{P}_g d_i \mathbf{L}\cdot\mathbf{S} d_j \mathcal{P}_g, \quad \alpha \equiv A_f/(\hbar^2\Delta_e) \tag{52}$$

角动量算符和位置算符有着明显的对易关系：

$$[L_i, d_j] = \mathrm{i}\hbar\epsilon_{ijk}d_k \tag{53}$$

因此

$$d_i\mathbf{L}\cdot\mathbf{S}d_j = \mathbf{L}\cdot\mathbf{S}d_i d_j + [\mathbf{L}\cdot\mathbf{S}, d_i]d_j \tag{54}$$

第一项很显然等于 0，因为这是作用在基态上的，而基态的话 $L=0, j=s, \mathbf{L}\cdot\mathbf{S} = j(j+1) - s(s+1) = 0$，因此我们有

$$\mathcal{D}_{ij} = -4u_s\delta_{ij} - \mathrm{i}\alpha\hbar\epsilon_{ilm}S_l\mathcal{D}_{mj} \tag{55}$$

做一级近似，也就是把上面式子的第二项中 $\mathcal{D}_{mj} \sim -4u_s\delta_{ij}$，于是

$$\mathcal{D}_{ij} = -4u_s\delta_{ij} - 4\mathrm{i}u_s\alpha\hbar\epsilon_{ijl}S_l \tag{56}$$

注意到，这里面，第一项对能量的贡献是量级 $1/\Delta_e$ 的，而第二项是 $1/\Delta_e^2$. 热效应同样也是 $1/\Delta_e^2$，于是没办法在保持这种相互作用的程度而减小热效应.

[19]这里可以通过这种办法检验：

$$(\hat{H}_0^{-1} - \hat{H}_0^{-1}\hat{V}\hat{H}^{-1})\hat{H} = 1 + \hat{H}_0^{-1}\hat{V} - \hat{H}_0^{-1}\hat{V} = 1 \tag{50}$$

化简可以得到

$$\hat{H}_{\text{eff}} = u_s|\mathbf{E}|^2 + \mathrm{i}\alpha\hbar u_s(\mathbf{E}^* \times \mathbf{E}) \cdot S \tag{57}$$

可以注意到，这样的光会使得自旋的极化. 例如，圆偏振，$\mathbf{E} = E_0(\mathbf{e}_x + \mathrm{i}\mathbf{e}_y)$，我们会发现

$$\mathbf{E}^* \times E = 2\mathrm{i}E_0^2\mathbf{e}_z \tag{58}$$

不难想象，这在我们将来操作冷原子的时候有很好的应用潜力.

两个原子间的相互作用

冷原子有他很独到的地方：

- 1. 很冷，所以粒子动能很低，我们只需要一个低能理论
- 2. 相互作用很弱，很多时候都是 short-range （$\propto \mathrm{e}^{-kr}$）的，有一个截止距离 r_0
- 3. 原子非常稀薄，可以暂时不使用多体理论，而是使用少体散射解决一定问题

在更深入的研究之前，我们先回顾一下两体散射过程的量子力学（场论下的散射问题我们这里不详细谈）. 我们很显然可以通过并非二次量子化而是基于波函数的量子力学来处理.

散射问题的一般解法是分波法，也就是

$$\Psi(r) = \sum_{l=0}^{+\infty} \frac{\chi_{kl}(r)}{kr} P_l(\cos\theta) \tag{1}$$

其中呢，χ_{kl} 是分波之后的成分，需要进一步求解，而 $P_l(\cos\theta)$ 则是勒让德函数. k 是受能量调控的. $E = \hbar^2k^2/(2m)$. 很容易的写出径向薛定谔方程：

$$\frac{\mathrm{d}^2\chi_{kl}(r)}{\mathrm{d}r^2} - \frac{l(l+1)}{r^2}\chi_{kl}(r) + \frac{m}{\hbar^2}(E - V(r))\chi_{kl}(r) = 0 \tag{2}$$

我们感兴趣的，由于所研究的是低能范围，更偏向于 $l=0$ 的 S 波成分，也就是

$$\frac{\mathrm{d}^2\chi_{k0}(r)}{\mathrm{d}r^2}+\frac{m}{\hbar^2}(E-V(r))\chi_{k0}(r)=0 \tag{3}$$

这个的解在大于我们的截止半径 r_0 的时候一目了然，就是简单地 $\sin(kr+\delta_k)$ 嘛；然而这个 δ 究竟是多少呢？这就要取决于在小于截止半径内的具体势能形式了．不过要注意，这里的这个截止半径确实是很关键的一个要求；但是，有的时候即使并不是 e^{-kr} 而是 $r^{-\alpha}$ 的形式的势能，在某些情况下我们的分析也是正确的，在本文的后半部分会涉及这一点．

关于这个 δ_k 的研究，我们实际上可以通过一个与归一化系数无关的办法，亦即

$$\frac{\chi'(r_0^+)}{\chi(r_0^+)}=\frac{k}{\tan(kr_0+\delta_k)}\sim\frac{k}{\tan(\delta_k)}\overset{def}{\equiv}-\frac{1}{a_s} \tag{4}$$

其中，这个 a_s 是人为定义的一个参数，叫做散射长度；其一个很直观的物理意义就是：在远处波函数的行为可以认为与一个半径为 a_s 的硬球模型一样（其势能为 $V(r)=(1-\theta(r-a_s))\times+\infty$）．

显然，对于低能散射（小的 k），$-ka_s=\tan\delta_k\sim\delta_k$，也就是说，phase shift $\delta_k\propto k$．当然随着 k 变得不那么小，这一现象也逐渐偏离；这里大家需要有一个数量级的感觉，也就是说，在 δ_k 接近饱和（趋近于 $\pi/2$）的时候，k 仍然是相对来说很低能的一个效应．

当然，这里面最重要的参数还是 a_s，尤其是我们这里并没有限制 a_s 的任何取值，它可正可负，反正就是一个参数．

当然，我们这里只是描绘了 $l=0$ 的情况；而事实上，对于任意的 l，其 phase shift 的计算会得到 $\propto k^{2l+1}$ 的样子，就像这里的线性关系．

既然我们这里已经得到了 Ψ 的形式，我们把它做 $r>r_0$ 的展开，并考虑到这时候 r 实际上已经很大了可以只取某一些项：

$$\Psi(r)=\frac{\sin(kr+\delta_k)}{kr}\to\cos\delta_k+\frac{1}{kr}\sin\delta_k=\frac{\sin\delta_k}{k}\left[\frac{1}{r}-\frac{1}{a_s}\right] \tag{5}$$

可以看到，$\Psi(a_s)=0$．这对不同的 k 都是一样的．这是一个散射长度的物理意义．

之前我们也提到过，只要有相同的 δ_k，那么对于不同的势能形式，其原子远处的性质也是一样的．于是我们来研究一个很有趣的 toy model：

$$V(r)=\begin{cases}-V_0 & (r<r_0)\\ 0 & (r\geqslant r_0)\end{cases} \tag{6}$$

于是 Schrödinger 方程就是

$$\chi''+\frac{m}{\hbar^2}(E+V_0)\chi=0 \quad (r<r_0) \tag{7}$$

$$\chi''+\frac{m}{\hbar^2}(E)\chi=0 \quad (r\geqslant r_0) \tag{8}$$

其中，E 表示由于是低能散射，所以能量很小可以忽略．如此我们就可以写出连接条件

$$\sqrt{\frac{V_0m}{\hbar^2}}\cot\left(\sqrt{\frac{V_0m}{\hbar^2}}r_0\right)=-\frac{1}{a_s} \tag{9}$$

如此，我们可以得到一个 a_s 和 V_0 的关系：

$$a_s=-\sqrt{\frac{\hbar^2}{V_0m}}\tan\left(\sqrt{\frac{V_0m}{\hbar^2}}r_0\right) \tag{10}$$

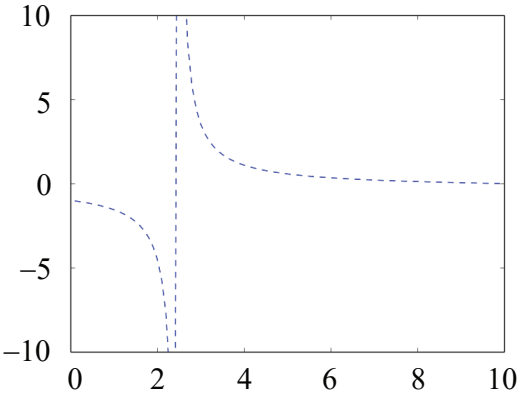


图 1: a_s 和 V_0 的关系

大致如图 1，横坐标 V_0，纵坐标 a_s．可以看到，a_s 会迅速的变成负无穷，然后再变成正无穷．

当然这是一个针对我们已经知道势能形式的验证；如果想要得到一些实验上可以测量的东西的话，就是

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}=\frac{1}{k^2}\left|\sum_{l=0}^{+\infty}(2l+1)\frac{\mathrm{e}^{\mathrm{i}\delta_{lk}}\sin\delta_{lk}}{k}P_l(\cos\theta)\right|^2 \tag{11}$$

之前说的时候我们注意到这个 r_0 是很重要的一个假设；但是如果我们有一个长程力会发生什么呢？实际上，有人证明，如果力的作用是 $r^{-\alpha}$ 的形式，那么对于 l 波，

$$\begin{aligned}\tan\delta_k&\propto k^{2l+1}(2l+1\leq\alpha-2)\\ \tan\delta_k&\propto k^{\alpha-2}(2l+1\geq\alpha-2)\end{aligned} \tag{12}$$

提一句，l 越高，δ_k 与 k 的关系幂次越高，所以相移就 $\to 0$，表明几乎没有相互作用．而我们处理的 s 波问题，要求 $\alpha>3$．van der waals 相互作用是 $ar^{-5}+br^{-11}$ 的，显然满足要求．

如果仔细的考虑分波问题的话，把入射的平面波分解展开成球面波，下面式子的左边应该等于右边

$$\begin{aligned}&f(\theta)\frac{\mathrm{e}^{\mathrm{i}kr}}{r}+\frac{1}{2\mathrm{i}kr}\left[\sum_{l=0}^{+\infty}(2l+1)P_l(\cos\theta)\left(\mathrm{e}^{\mathrm{i}kr}-\mathrm{e}^{-\mathrm{i}(kr-\pi l)}\right)\right]\\ =&\sum_{l=0}^{+\infty}\frac{\chi_{kl}(r)}{kr}P_l(\cos\theta)\end{aligned} \tag{13}$$

而可以显然的看到

$$\begin{aligned}\mathrm{e}^{\mathrm{i}kr}-\mathrm{e}^{-\mathrm{i}(kr-\pi l)}&=\mathrm{e}^{\mathrm{i}\pi l/2}\left(\mathrm{e}^{\mathrm{i}(kr-\pi l/2)}-\mathrm{e}^{-\mathrm{i}(kr-\pi l/2)}\right)\\ &=\mathrm{i}^l 2\mathrm{i}\sin\left(kr-\frac{\pi l}{2}\right)\end{aligned} \tag{14}$$

其中定义得到的散射成分为

$$f(\theta)\frac{\mathrm{e}^{\mathrm{i}kr}}{r} \tag{15}$$

而第二项是平面波 $\mathrm{e}^{\mathrm{i}kz}$ 的拆分．不妨假设

$$f(\theta)=\sum_{l=0}^{+\infty}D_lP_l(\cos\theta) \tag{16}$$

从而，在 $r\to\infty$ 的时候，考虑 l 成分：

$$A_l\sin\Big(kr-\frac{\pi}{2}l+\delta_l\Big)=\mathrm{i}^l(2l+1)\sin\Big(kr-\frac{\pi}{2}l\Big)+k\mathrm{e}^{\mathrm{i}kr}D_l \tag{17}$$

而 D_l 是个数，与 r 无关，那么显然我们就有

$$\frac{\partial}{\partial r}\left[\mathrm{e}^{-\mathrm{i}kr}\left(A_l\sin\Big(kr-\frac{\pi}{2}l+\delta_l\Big)-\mathrm{i}^l(2l+1)\sin\Big(kr-\frac{\pi}{2}l\Big)\right)\right]=0 \tag{18}$$

从而计算得到

$$f(\theta)=\frac{1}{2\mathrm{i}k}\left[\sum_{l=0}^{+\infty}(2l+1)P_l(\cos\theta)\left(\mathrm{e}^{2\mathrm{i}\delta_l}-1\right)\right]=0 \tag{19}$$

和

$$A_l=\mathrm{i}^l(2l+1)\mathrm{e}^{\mathrm{i}\delta_l} \tag{20}$$

对于 s 成分，有

$$\begin{aligned}f_s(\theta)&=\frac{\mathrm{e}^{2\mathrm{i}\delta}-1}{2\mathrm{i}k}=\frac{\sin\delta\mathrm{e}^{\mathrm{i}\delta+\pi/2}}{\mathrm{i}k}=\frac{\sin\delta}{k\mathrm{e}^{-\mathrm{i}\delta}}=\frac{\sin\delta}{k(\cos\delta-\mathrm{i}\sin\delta)}\\&=\frac{1}{-\mathrm{i}k+k/\tan\delta}=\frac{1}{-\mathrm{i}k-1/a_s}\end{aligned} \tag{21}$$

可以看出来，当 $ka_s\ll 1$ 时，$f_s\sim -a_s$，此时 s 波散射的散射截面 $\sigma=8\pi|f(\theta)|^2=4\pi a_s^2$；而另一个极限，$ka_s\gg 1$ 时，$f_s\sim -1/(\mathrm{i}k)$，$\sigma=8\pi/k^2$.

注意到，在第二个极限的时候，散射强度只与动能大小相关，我们称之为 unitary 区域.

就像固体里面一样，我们这里可以定义一个**赝势**从而得到我们想要的 a_s. 具体如何得到呢？我们通过如下这种特殊的势能：δ 势

$$V(r)=\delta(r)\hat{\mathcal{O}}(r) \tag{22}$$

这个势需要满足对于任意的 $\forall r$ 都有统一的薛定谔方程，而且解是我们期望的（低能的）

$$\Psi(r)=\frac{\sin(kr+\delta_k)}{kr}\to\frac{1}{kr}-\frac{1}{ka_s} \tag{23}$$

显然，动能部分得到非常简单的结果

$$-\frac{\hbar^2}{2m}\nabla^2\Psi(r)=-\frac{\hbar^2}{2mk}\nabla^2\frac{1}{r}+-\frac{\hbar^2}{2m}\nabla^2\frac{1}{ka_s}=-\frac{\hbar^2}{2mk}4\pi\delta(r) \tag{24}$$

势能为了满足薛定谔方程，应该，

$$\delta(r)\hat{\mathcal{O}}(r)\Psi(r)=\frac{\hbar^2}{2mk}4\pi\delta(r) \tag{25}$$

从积分角度理解，有

$$\int\delta(r)\hat{\mathcal{O}}(r)\Psi(r)=\hat{\mathcal{O}}(r)\Psi(r)|_{r=0}=\hat{\mathcal{O}}(r)\left(\frac{1}{kr}-\frac{1}{ka_s}\right)|_{r=0}\overset{\text{should}}{=}\frac{\hbar^2}{2mk}4\pi \tag{26}$$

那么，可以看出，如果我们令

$$\hat{\mathcal{O}}(r)=\frac{\hbar^2}{2m}4\pi a_s\partial_r\cdot r \tag{27}$$

其显然有

$$\hat{\mathcal{O}}(r)\frac{1}{kr}=\frac{\hbar^2}{2m}4\pi a_s\partial_r\frac{1}{k}=0 \tag{28}$$

$$\hat{\mathcal{O}}(r)\frac{1}{ka_s}=\frac{\hbar^2}{2m}4\pi a_s\partial_r\frac{r}{ka_s}=\frac{\hbar^2}{2mk}4\pi \tag{29}$$

也就是说满足我们的要求.

接下来，我们考察另外一种情况：我们考虑点接触势，也就是说 $V(r)=g\delta(r)$. 注意，这里的 g 不应该理解为一个参数. 如果强行理解为一个参数，由于波函数的发散，我们通常这里需要进行**重整化操作**.

在这种考虑下，相互作用 Hamiltonian 就可以写为

$$\mathcal{H}=\int\mathrm{d}^3\mathbf{r}\left(\psi^\dagger\left(-\frac{\hbar^2}{2m}\nabla^2\right)\psi+g\psi^\dagger\psi^\dagger\psi\psi\right) \tag{30}$$

我们可以建立两体散射的 T 矩阵，并且只考虑 ladder 图. 如图：

也就是说我们可以写出形如 $T=g+fT$ 的 self-consistant 形式.

通过分析知道，最后 T 矩阵实际上是散射的入射能量的函数，也就是说，可以写成 $T(E)$ 的形式，有

$$T(E)=g+\frac{g}{V}\sum_{\mathbf{k}}\frac{1}{E-\hbar^2\mathbf{k}^2/m}T(E) \tag{31}$$

其中第二项包含一个顶角、可供交换的动量 $\mathbf{k}$ 以及相应带来的传播子. 形式上我们解得

$$\begin{aligned}T(E)&=\frac{g}{1-\frac{g}{V}\sum_{\mathbf{k}}\frac{1}{E-\hbar^2\mathbf{k}^2/m}}\\&=\frac{1}{\frac{1}{g}+\frac{1}{V}\sum_{\mathbf{k}}\frac{1}{\hbar^2\mathbf{k}^2/m}-\frac{1}{V}\sum_{\mathbf{k}}\left(\frac{1}{E-\hbar^2\mathbf{k}^2/m}+\frac{1}{\hbar^2\mathbf{k}^2/m}\right)}\end{aligned} \tag{32}$$

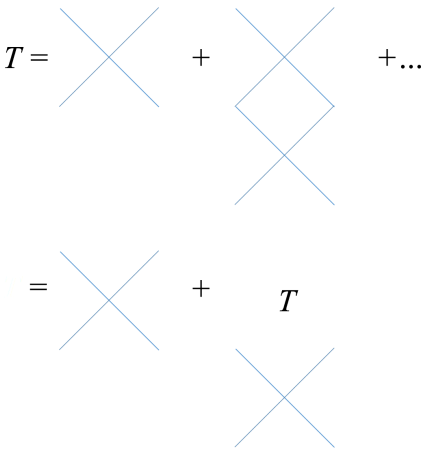


图 2: ladder 图

分母的第三项可以计算：

$$\begin{aligned}&\frac{1}{V}\sum_{\mathbf{k}}\left(\frac{1}{E-\hbar^2\mathbf{k}^2/m}+\frac{1}{\hbar^2\mathbf{k}^2/m}\right)\\=&\frac{mE}{V\hbar^2}\sum_{\mathbf{k}}\frac{1}{\mathbf{k}^2(E-\hbar^2\mathbf{k}^2/m)}\end{aligned}\tag{33}$$

求和可以变成积分，补上差的系数并且根据对称性，我们得到

$$\begin{aligned}=&\frac{V}{(2\pi)^3}\frac{1}{V}\int \mathrm{d}^3\mathbf{k}\left(\frac{1}{E-\hbar^2\mathbf{k}^2/m}+\frac{1}{\hbar^2\mathbf{k}^2/m}\right)\\=&\frac{m}{2\pi\hbar^2}\int \mathrm{d}k\frac{1}{1-\hbar^2k^2/mE}\end{aligned}\tag{34}$$

令 $k_0=\sqrt{mE/\hbar^2}$，结合留数定理（注意这里是奇点在实轴上，我们只有主值）我们有

$$\begin{aligned}=&\frac{m}{2\pi\hbar^2}\int \mathrm{d}k\frac{1}{1-k^2/k_0^2}\\=&\frac{m}{2\pi\hbar^2}\cdot 2\cdot\pi\mathrm{i}\cdot\left(-\frac{k_0}{2}\right)\\=&\frac{-\mathrm{i}mk_0}{2\pi\hbar^2}\end{aligned}\tag{35}$$

于是

$$T(E)=\frac{1}{\frac{1}{g}+\frac{1}{V}\sum_{\mathbf{k}}\frac{1}{\hbar^2\mathbf{k}^2/m}+\frac{\mathrm{i}mk_0}{2\pi\hbar^2}}\tag{36}$$

利用散射振幅与 T 矩阵的关系：

$$T=\frac{4\pi\hbar^2}{m}\frac{1}{1/a_s+\mathrm{i}k}=\frac{1}{\frac{1}{g}+\frac{1}{V}\sum_{\mathbf{k}}\frac{1}{\hbar^2\mathbf{k}^2/m}+\frac{\mathrm{i}mk_0}{2\pi\hbar^2}} \tag{37}$$

实部相等于是我们得到了最重要的一个关系式：

$$\frac{ma_s}{4\pi\hbar^2}=\frac{1}{g}+\frac{1}{V}\sum_{\mathbf{k}}\frac{1}{\hbar^2\mathbf{k}^2/m} \tag{38}$$

对散射相移的一般定义是：散射波函数 $\Psi_{sc}(r)=f(\theta)\mathrm{e}^{\mathrm{i}(kr-\omega t)}/r$ 中的散射振幅 $f(\theta)$ 进行分波展开，

$$f(\theta)=\sum_{l=0}^{+\infty}(2l+1)f_lP_l(\cos\theta) \tag{39}$$

得到的 f_l 满足

$$f_l=\frac{\mathrm{e}^{\mathrm{i}\delta_{lk}}\sin\delta_{lk}}{k} \tag{40}$$

而在能量很低的时候，f_l 对 k 有一个大概的形式依赖关系：（其中利用了 $j_l(kr)\to a(kr/2)^l$ 以及径向部分 $A_l(r)=\mathrm{e}^{\mathrm{i}\delta_l}[\cos\delta_l j_l(kr)-\sin\delta_l n_l(kr)]$，在 $\delta\to 0$ 时 $A_l(r)\to j_l(kr)$）：

$$f_l=-\frac{2m}{\hbar^2}\int_0^{\infty}\mathrm{d}r\, j_l(kr)V(r)A(r)r^2\sim k^{2l} \tag{41}$$

从而得到在 $k\to 0$ 的时候

$$\delta_{lk}\propto k^{2l+1} \tag{42}$$

Feshbach 共振

相互作用 Hamiltonian，我们考虑一个坐标表象张量积一个 open-closed 的 channel 表象．这种情况下，可以写为：

$$\hat{H}=-\frac{\hbar^2\nabla^2}{m}\otimes I_2+V_{2\times 2} \tag{1}$$

其中因为约化所以 $2m \to m$．此外，

$$V_{2\times 2} = \begin{cases} \begin{pmatrix} -V_{\text{open}} & W \\ W & -V_{\text{close}} \end{pmatrix} & (r < r_0) \\ \begin{pmatrix} 0 & 0 \\ 0 & +\infty \end{pmatrix} & (r > r_0) \end{cases} \tag{2}$$

这其实就是对我们之前的物理要求：在远处 close 通道有一个远高于能标的渐进值，近处有一个相互作用（W）但是很小．它是由超精细结构提供的，比 $|V_{\text{open}} - V_{\text{close}}|$ 小几个量级．

在很前面的时候，我们知道，$E \to 0$ 的情况下，波函数可以写为 $\Psi = \frac{r-a_s}{r}|\text{open}\rangle$．

在近处，我们考虑相互作用部分的本征态 $|\pm\rangle$：

$$|+\rangle = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad |-\rangle = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} \tag{3}$$

其中，$\tan 2\theta = 2W/|V_{\text{open}} - V_{\text{close}}|$，本征值可以近似认为是 V_{open} 和 V_{close}．可以用未归一的波函数来描述这两个本征态各自独立的情况：

$$|\Psi\rangle \propto \frac{1}{r}\left[\sin(q_+ r)|+\rangle + A\sin(q_- r)|-\rangle\right] \tag{4}$$

其中，q_+, q_- 为该能量下的波矢，A 是一个系数．由于整体 $E \to 0$ 为零能散射，所以 $\hbar^2 q^2/m$ 就应该等于负的相互作用本征值，也就是有

$$\frac{\hbar^2 q_{+/-}^2}{m} = V_{\text{open}}/V_{\text{close}} \tag{5}$$

考虑在 $r \geqslant r_0$ 的位置，体系波函数就不再会包含 close 成分，也就是说，$\langle\text{close}|\Psi\rangle(r_0) = 0$，从而告诉我们这个系数 A 满足

$$A = -\tan\theta \frac{\sin(q_+ r_0)}{\sin(q_- r_0)} \tag{6}$$

此外，open 态在 r_0 处要连续，不仅是波函数还有其导数．利用之前得到的关系，可以有

$$\left.\frac{L'}{L}\right|_{r=r_0} = \left.\frac{(r-a_s)'}{r-a_s}\right|_{r=r_0} \tag{7}$$

其中

$$L=\sin(q_+r)\cos\theta-\frac{\sin(q_+r_0)}{\sin(q_-r_0)}\sin(q_-r)\tan\theta \tag{8}$$

经过化简我们得到

$$\frac{q_+\cos^2\theta}{\tan(q_+r_0)}+\frac{q_-\sin^2\theta}{\tan(q_-r_0)}=\frac{1}{r_0-a_s} \tag{9}$$

可以看到，因为 $\theta\to 0$，主要成分有 q_+ 贡献．在没有相互作用的时候，我们有

$$\frac{1}{r_0-a_{\text{bg}}}=\frac{q_+}{\tan(q_+r_0)} \tag{10}$$

只有在 $\tan(q_-r_0)\to 0$ 的时候，第二项才有贡献．考虑一个体系的束缚态，能量 E_c，其满足

$$\sin(qr)|_{r=0,r_0}=0,\quad q=\sqrt{\frac{m(V_c+E_c)}{\hbar^2}} \tag{11}$$

在 $E_c\to 0$ 的时候，$q_-\sim q$，从而

$$\tan(q_-r_0)\sim(q_--q)r_0=\frac{1}{\hbar}\sqrt{m}(\sqrt{V_c}-\sqrt{V_c+E_c})=-\frac{mE_c}{2\hbar^2q_-} \tag{12}$$

也就是说，第二项约为

$$-\frac{2\hbar^2q_-^2\theta^2}{mr_0}\frac{1}{E_c}=-\gamma\frac{1}{E_c},\gamma\equiv\frac{2\hbar^2q_-^2\theta^2}{mr_0} \tag{13}$$

此时，

$$\frac{a_s-r_0}{a_{\text{bg}}-r_0}=1-\frac{\Delta B}{B-B_{\text{res}}} \tag{14}$$

其中，B 为施加的磁场，而我们的 E_c 则由磁场调控（因为磁矩不一样）：$E_c\to E_c+\delta\mu B$，而 $\Delta B=\frac{\hbar^2}{\delta\mu m}\gamma(a_{\text{bg}}-r_0)$，$B_{\text{res}}=-E_c/\delta\mu+\Delta B$，都是一些以 B 参数的东西．

BCS-BEC Crossover 的平均场描述

仍然沿用 BCS 的 Hamiltonian，在动量空间里面，

$$H-\mu N=\sum_{\mathbf{k},\sigma}(\epsilon_{\mathbf{k}}-\mu)a^\dagger_{\mathbf{k},\sigma}a_{\mathbf{k},\boldsymbol{\sigma}}+\frac{1}{V}\sum_{\mathbf{k},\mathbf{k}'}U_{\mathbf{k},\mathbf{k}'}a^\dagger_{\mathbf{k}\uparrow}a^\dagger_{-\mathbf{k}\downarrow}a_{-\mathbf{k}\downarrow}a_{\mathbf{k}\uparrow} \tag{1}$$

假设不是一个微扰的情况下，也可以用 $|\Psi\rangle = \prod_{\mathbf{k}}(u_{\mathbf{k}} + v_{\mathbf{k}} a^\dagger_{\mathbf{k}\uparrow} a^\dagger_{-\mathbf{k}\downarrow})|0\rangle$ 来描述基态．仍然定义平均场参数 $\Delta = \langle a_{-\mathbf{k}\downarrow}, a_{\mathbf{k}\uparrow}\rangle$．不同的是，研究冷原子问题的时候，并不是只有在费米面附近才有 U，而是一个队任意动量都可以的情况．重复同样的 Bogoliubov 变换，我们能够解出来 $u^2_{\mathbf{k}}, v^2_{\mathbf{k}} = 1/2(1 \pm \epsilon_{\mathbf{k}}/E_{\mathbf{k}})$，其中元激发的能级关系 $E_{\mathbf{k}} = \sqrt{(\epsilon_{\mathbf{k}} - \mu)^2 + |\Delta|^2}$．利用自洽方程，我们得到

$$-\frac{1}{U} = \frac{1}{V}\sum_{\mathbf{k}} \frac{1}{2E_{\mathbf{k}}} \tag{2}$$

而我们知道，式 38 [1516] 给出了

$$\frac{m a_s}{4\pi\hbar^2} = \frac{1}{g} + \frac{1}{V}\sum_{\mathbf{k}} \frac{1}{\hbar^2\mathbf{k}^2/m} = \frac{1}{V}\sum_{\mathbf{k}} \left(\frac{1}{2\epsilon_{\mathbf{k}}} - \frac{1}{2E_{\mathbf{k}}}\right) \tag{3}$$

定义无量纲参数来简化问题：

$$x^2 = \frac{\hbar^2\mathbf{k}^2}{2m}\frac{1}{\Delta}, \quad x_0 = \frac{\mu}{\Delta}, \quad \epsilon_x = x^2 - x_0 = \frac{\epsilon_{\mathbf{k}} - \mu}{\Delta}, \quad E_x = \sqrt{1 + \epsilon_x^2} = \frac{E_{\mathbf{k}}}{\Delta} \tag{4}$$

把求和化为积分，也就是 $\frac{1}{V}\sum_{\mathbf{k}} \to \int_0^\infty \mathrm{d}^3\mathbf{k}/(2\pi)^3$，利用求对称性，我们很容易转写一些式子：

$$\begin{aligned}
-\frac{m\Delta}{4\pi\hbar^2 a_s} &= \frac{1}{V}\sum_{\mathbf{k}} \left(\frac{\Delta}{2E_{\mathbf{k}}} - \frac{\Delta}{2\epsilon_{\mathbf{k}}}\right) \\
-\frac{m\Delta}{4\pi\hbar^2 a_s} &= \int_0^\infty \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3}\left(\frac{1}{2E_x} - \frac{1}{2x^2}\right) \\
&= \int_0^\infty \frac{\mathrm{d}k \cdot 4\pi k^2}{(2\pi)^3}\left(\frac{1}{2E_x} - \frac{1}{2x^2}\right) \\
-\frac{1}{a_s} &= \frac{2}{\pi\Delta}\frac{\hbar^2}{m}\int_0^\infty \mathrm{d}k \cdot k^2 \left(\frac{1}{2E_x} - \frac{1}{2x^2}\right) \\
k^2 &= \frac{m x^2}{\hbar^2} \\
-\frac{1}{a_s} &= \frac{2}{\pi}\sqrt{\frac{2m\Delta}{\hbar^2}}\int_0^\infty \mathrm{d}x \cdot x^2 \left(\frac{1}{E_x} - \frac{1}{x^2}\right)
\end{aligned} \tag{5}$$

可以看出来，右边的积分纯粹只是 x_0 的函数，可以定义

$$I_1(x_0) = \int_0^\infty \mathrm{d}x \cdot x^2 \left(\frac{1}{E_x} - \frac{1}{x^2}\right) \tag{6}$$

而另一方面，我们可以根据热力学定律知道平均粒子数 $N=-\partial\Omega/\partial\mu$ 得到

$$N=\sum_{\mathbf{k}}\left(1-\frac{\epsilon_{\mathbf{k}}-\mu}{E_{\mathbf{k}}}\right) \tag{7}$$

而这个 N 是要满足我们的要求的，也就是说 $n=N/V=k_F^3/(3\pi^2)$．我们当然也可以用类似的无量纲参数来处理：

$$\begin{aligned}
n&=\frac{1}{V}\sum_{\mathbf{k}}\left(1-\frac{\epsilon_{\mathbf{k}}-\mu}{E_{\mathbf{k}}}\right)\\
&=\frac{1}{V}\sum_{\mathbf{k}}\left(1-\frac{\epsilon_x\Delta}{E_x\Delta}\right)\\
&=\int_0^\infty\frac{\mathrm{d}k\cdot 4\pi k^2}{(2\pi)^3}\left(1-\frac{\epsilon_x}{E_x}\right)\\
&=\frac{1}{2\pi^2}\frac{2m\Delta}{\hbar^2}\sqrt{\frac{2m\Delta}{\hbar^2}}\int_0^\infty\mathrm{d}x\cdot x^2\left(1-\frac{\epsilon_x}{E_x}\right)\\
&=\frac{1}{2\pi^2}\left(\frac{2m\Delta}{\hbar^2}\right)^{3/2}I_2(x_0)
\end{aligned} \tag{8}$$

其中

$$I_2(x_0)=\int_0^\infty\mathrm{d}x\cdot x^2\left(1-\frac{\epsilon_x}{E_x}\right) \tag{9}$$

于是我们得到了两个方程（代替原本的自洽方程来给我们关于系统的信息）

$$\begin{cases}
\dfrac{\Delta}{E_F}=\left(\dfrac{2}{3I_2(x_0)}\right)^{2/3}\\[2ex]
\dfrac{1}{k_Fa_s}=-\dfrac{2}{\pi}\left(\dfrac{2}{3I_2(x_0)}\right)^{1/3}I_1(x_0)
\end{cases} \tag{10}$$

但是，对于 I_1,I_2 的渐进行为并不好分析，我们定义更多的积分

$$I_3(x_0)=\int_0^\infty\mathrm{d}x\frac{x^4\epsilon_x}{E_x^6} \tag{11}$$

$$I_4(x_0)=\int_0^\infty\mathrm{d}x\frac{x^2}{E_x^2} \tag{12}$$

$$I_5(x_0)=\int_0^\infty\mathrm{d}x\frac{x^2}{E_x^3} \tag{13}$$

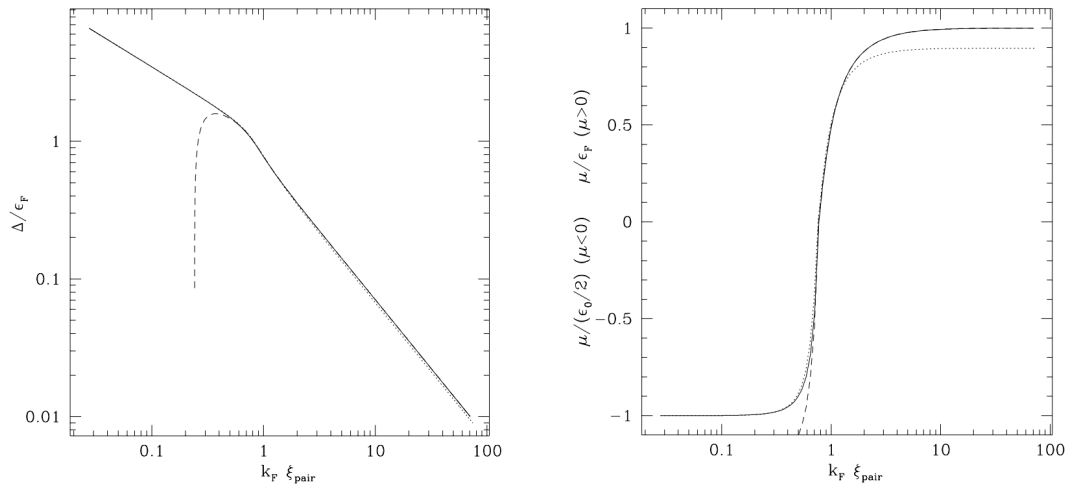


图 1: 左：Δ/E_F vs $k_F\xi_{\text{pair}}$，右：$\mu/E_F(\mu>0), \mu/(\epsilon_0/2)(\mu<0)$，区分两种情况是因为在 Bose 情况下没有费米能的定义，$\epsilon_0 \equiv 1/(ma_s^2)$

$$I_6(x_0) = \int_0^\infty \mathrm{d}x \frac{x^2\epsilon_x}{E_x^3} \tag{14}$$

可以知道，$I_1 = (2x_0I_6 - I_5), I_2 = \frac{2}{3}(x_0I_5 + I_6)$. 而 I_3, I_4 是很好计算的. 如果定义

$$x_1^2 = \frac{x_0 + \sqrt{1+x_0^2}}{2} \tag{15}$$

就可以算出

$$I_3 = \frac{\pi}{16}\frac{x_1(1+x_1^4)}{(1+x_0^2)^{1/2}} \tag{16}$$

$$I_4 = \frac{\pi}{2}x_1 \tag{17}$$

平均场下可以计算费米子对的特征距离

$$\xi_{\text{pair}}^2 = \frac{1}{m^2}\frac{\int_0^\infty \mathrm{d}k(k^4(\epsilon_{\mathbf{k}}-\mu)^2/E_{\mathbf{k}}^6)}{\int_0^\infty \mathrm{d}k(k^2/E_{\mathbf{k}}^2)} = \frac{2}{m\Delta}\frac{I_3}{I_4} \tag{18}$$

还有一些其他的参数信息

$$\frac{\Delta}{E_F} = \left(\frac{1}{x_0I_5 + I_6}\right)^{2/3} \tag{19}$$

$$\frac{\mu}{E_F} = x_0\frac{\Delta}{E_F} \tag{20}$$

$$\frac{1}{k_F a_s} = -\frac{4}{\pi}\frac{x_0 I_6 - I_5}{(x_0 I_5 + I_6)^{1/3}} \tag{21}$$

$$k_F \xi_{\text{pair}} = \sqrt{\frac{1+x_1^4}{2}}\frac{(x_0 I_5 + I_6)^{1/3}}{(1+x_0^2)^{1/4}} \tag{22}$$

现在仔细分析一下渐进行为：

- $x_0 \to \infty$，此时 $\mu \to E_F$，$I_5 \sim \sqrt{x_0}$，$I_6 \sim \ln x_0/(2\sqrt{x_0})$，我们的 $1/k_F a_s \to (2/\pi)\ln x_0$，$a_s \to 0^+$，表现为微弱的吸引力，就是 BCS 态的情况．此时 $k_F \xi_{\text{pair}} \to x_0/\sqrt{2}$.

- $x_0 \to -\infty$，$I_5 \sim \pi/(16|x_0|^{3/2}), I_6 \sim \pi/(4|x_0|^{1/2})$．此时 $\Delta \to \infty, \mu \to -\infty, a_s \to 0^+, k_F \xi_{\text{pair}} \to 0^+$．表现为强吸引力（$\Delta$），就是 BEC 极限．

可以用 $k_f \xi_{\text{pair}}$ 作为参数反解很多信息，如图 1.

BEC 超流

首先，我们回顾一下理想玻色气体的 BEC．假设这种玻色子的色散是 $\epsilon_{\mathbf{k}} = \hbar^2 \mathbf{k}^2/2m$ （实际上长什么样在我们这里一般的分析中并没有什么关系）．我们可以看到，$n_{\mathbf{k}} = 1/(\mathrm{e}^{\beta(\epsilon_{\mathbf{k}}-\mu)} - 1)$．那么，在热力学极限下可以发现如果 $\mu < 0$，换句话说温度处于 T_c，$N \to \infty$，$n_{\mathbf{k}}/N \to 0$．而相变的时候 $T_c : \mu = 0$，这时候物理图像是 de Brogilie 波长 $\approx$ 粒子间距．在 $< T_c$ 时，

$$N = N_0 + \frac{V}{2\pi^3}\int \mathrm{d}^3\mathbf{k}\frac{1}{\mathrm{e}^{\beta\epsilon_{\mathbf{k}}} - 1}, \quad N_0/N \neq 0 \tag{1}$$

呈现宏观占据．看得出来我们虽然很熟悉什么是理想玻色气体的 BEC，但是对于一般的情况，如何规定它是否是处于凝聚的态呢？

首先，我们可以写出某个多体（N 体）波函数 $\Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$ 密度矩阵 ρ．它的矩阵元有 $2N$ 个指标，$\langle x_1, \cdots, x_N|\rho|y_1, \cdots, y_N\rangle$．把它 trace 掉 $N-1$ 个指标，得到

$$\rho(\mathbf{r}, \mathbf{r}') = N\int \sum_s p_s \psi^*(\mathbf{r}, \mathbf{r}_2, \cdots, \mathbf{r}_N) p_s \psi(\mathbf{r}', \mathbf{r}_2, \cdots, \mathbf{r}_N) \mathrm{d}^3\mathbf{r}_2 \cdots \mathrm{d}^3\mathbf{r}_N \tag{2}$$

二参数函数可以按照某一个基展开，

$$\rho(\mathbf{r},\mathbf{r}')=\sum_i N_i\varphi_i^*(\mathbf{r})\varphi_i(\mathbf{r}') \tag{3}$$

对于 N_i，如果所有 $N_i/N=0$，那么就称这个态是 normal phase．如果有一个不等于 0 的，那就是一般的 BEC，如果有多个 $N_i\neq 0$ 叫 fragment BEC．后两者称之存在非对角长程序（off-diagonal long-range order，ODLRO）．

对于正常的 BEC 的情况，$\exists N_0\gg N_{i\neq 0}$，如果我们忽略两体相互作用的细节，我们就可以写出

$$\rho(\mathbf{r},\mathbf{r}')\approx N_0\varphi^*(\mathbf{r})\varphi(\mathbf{r}') \tag{4}$$

其中，满足要求的系统的态是

$$\Psi(\mathbf{r}_1,\mathbf{r}_2,\cdots,\mathbf{r}_N)=\prod_{\otimes}\varphi(\mathbf{r}_i) \tag{5}$$

我们考虑系统的 Hamiltonian．利用赝势，我们写出

$$\hat{H}=\sum_i^N\left(\frac{\nabla_i^2}{2m}+V(\mathbf{r}_i)\right)\sum_{\langle i,j\rangle}\frac{4\pi\hbar^2a_s}{m}\delta^3(\mathbf{r}_{ij})\frac{\partial}{\partial r_{ij}}r_{ij} \tag{6}$$

显然，我们的波函数式 5 在这里面并不存在我们上一章讨论的 $r_i\to r_j$ 的那种发散行为，因为我们实际上忽略了两体相互作用的细节．当然，这种情况下我们很容易发现，没有奇异性的时候，赝势仅仅就是一个 δ 而已，因为其形如 $r_{ij}\partial_{r_{ij}}$ 的成分在 $\delta^3(\mathbf{r}_{ij})$ 下不作任何贡献：

$$\delta^3(\mathbf{r}_{ij})\frac{\partial}{\partial r_{ij}}r_{ij}=\delta^3(\mathbf{r}_{ij}) \tag{7}$$

我们可以通过对 $\langle\varphi|H+\mu N|\varphi\rangle$ 取极值得到本征值方程．

$$\varepsilon=N\int\mathrm{d}^3\mathbf{r}\varphi^*(\mathbf{r})\left(-\frac{\hbar^2\nabla^2}{2m}+V(\mathbf{r})\right)\varphi(\mathbf{r})+\frac{N(N-1)}{2}\int\mathrm{d}^3\mathbf{r}\frac{4\pi\hbar^2a_s}{m}|\varphi(\mathbf{r})|^4 \tag{8}$$

以 φ^* 作为变量，$N\to\infty$ 而且 $\phi=\sqrt{N}\varphi, U\equiv\dfrac{4\pi\hbar^2a_s}{m}$，我们有

$$\Rightarrow\left(-\frac{\hbar^2\nabla^2}{2m}+V(\mathbf{r})\right)\phi+U|\phi|^2\phi=\mu\phi \tag{9}$$

这就是著名的**不含时 Gross-Pitaevskii 方程**.

如果我们考虑的不是一个定态问题，而是 $\phi=\phi(\mathbf{r},t)$，那么我们的这个方程实际上要稍作改动．符合直觉的我们可以得到 **含时 Gross-Pitaevskii 方程**

$$\mathrm{i}\hbar\frac{\partial\phi}{\partial t}=\left(-\frac{\hbar^2\nabla^2}{2m}+V(\mathbf{r})\right)\phi+U|\phi|^2\phi \tag{10}$$

接下来我们做两个魔法操作．首先将 $\phi^*\times$ 式 10 减去它的共轭，显然其左手边为

$$\mathrm{i}\hbar\phi^*\frac{\partial\phi}{\partial t}+\mathrm{i}\hbar\phi\frac{\partial\phi^*}{\partial t}=\mathrm{i}\hbar\frac{\partial\phi^*\phi}{\partial t}=\mathrm{i}\hbar\frac{\partial|\phi|^2}{\partial t} \tag{11}$$

而右手边则为

$$\begin{aligned}&\left(\phi^*\left(-\frac{\hbar^2\nabla^2}{2m}+V(\mathbf{r})\right)\phi+U|\phi|^4\right)-\left(\phi\left(-\frac{\hbar^2\nabla^2}{2m}+V(\mathbf{r})\right)\phi^*+U|\phi|^4\right)\\=&\frac{\hbar^2}{2m}\nabla\cdot(\phi\nabla\phi^*-\phi^*\nabla\phi)\end{aligned} \tag{12}$$

于是，就得到

$$\mathrm{i}\hbar\frac{\partial|\phi|^2}{\partial t}=\frac{\hbar^2}{2m}\nabla\cdot(\phi\nabla\phi^*-\phi^*\nabla\phi)$$

定义 $\phi(\mathbf{r})=\sqrt{\rho}\mathrm{e}^{\mathrm{i}\theta}$，其中 $\rho=\rho(\mathbf{r}),\theta=\theta(\mathbf{r})$．我们从而有

$$\frac{\partial\rho}{\partial t}+\nabla\cdot(\rho\mathbf{v}),\quad \mathbf{v}\overset{\text{def}}{\equiv}\frac{\hbar}{m}\frac{\partial\theta}{\partial t} \tag{13}$$

这就得到了连续性方程．当然，还得补充一个无旋方程

$$\nabla\times\mathbf{v}=0 \tag{14}$$

另一个魔法是将 $\phi^*\times$ 式 10 它的共轭．左手边为

$$\mathrm{i}\hbar\phi^*\frac{\partial\phi}{\partial t}-\mathrm{i}\hbar\phi\frac{\partial\phi^*}{\partial t}=\mathrm{i}\hbar\phi^{*2}\frac{\partial\phi/\phi^*}{\partial t}=-2\hbar\rho\frac{\partial\theta}{\partial t} \tag{15}$$

而右边就有点复杂了，下面的最后一项会很复杂：

$$\begin{aligned}&\left(\phi^*\left(-\frac{\hbar^2\nabla^2}{2m}+V(\mathbf{r})\right)\phi+U|\phi|^4\right)+\left(\phi\left(-\frac{\hbar^2\nabla^2}{2m}+V(\mathbf{r})\right)\phi^*+U|\phi|^4\right)\\=&2V\rho+2U\rho^2-\frac{\hbar^2}{2m}\left(\phi^*\nabla^2\phi+\phi\nabla^2\phi^*\right)\end{aligned} \tag{16}$$

我们显然希望能够提取出来一个 ρ，所以我们给出详细的计算过程：

$$\begin{aligned}\nabla^2\phi &= \nabla^2\left(\sqrt{\rho}\mathrm{e}^{\mathrm{i}\theta}\right)\\ &= \nabla\cdot\left(\frac{\mathrm{e}^{\mathrm{i}\theta}}{2\sqrt{\rho}}\nabla\rho + i\sqrt{\rho}\mathrm{e}^{\mathrm{i}\theta}\nabla\theta\right)\\ &= \frac{\mathrm{e}^{\mathrm{i}\theta}}{2\sqrt{\rho}}\nabla^2\rho + \frac{\mathrm{i}\mathrm{e}^{\mathrm{i}\theta}}{\sqrt{\rho}}\nabla\theta\nabla\rho - \frac{\mathrm{e}^{\mathrm{i}\theta}}{4\rho^{3/2}}(\nabla\rho)^2 - \sqrt{\rho}\mathrm{e}^{\mathrm{i}\theta}(\nabla\theta)^2 + i\sqrt{\rho}\mathrm{e}^{\mathrm{i}\theta}\nabla^2\theta\end{aligned} \tag{17}$$

类似有

$$\begin{aligned}\nabla^2\phi^* &= \nabla^2\left(\sqrt{\rho}\mathrm{e}^{-\mathrm{i}\theta}\right)\\ &= \nabla\cdot\left(\frac{\mathrm{e}^{-\mathrm{i}\theta}}{2\sqrt{\rho}}\nabla\rho - i\sqrt{\rho}\mathrm{e}^{-\mathrm{i}\theta}\nabla\theta\right)\\ &= \frac{\mathrm{e}^{-\mathrm{i}\theta}}{2\sqrt{\rho}}\nabla^2\rho - \frac{\mathrm{i}\mathrm{e}^{-\mathrm{i}\theta}}{\sqrt{\rho}}\nabla\theta\nabla\rho - \frac{\mathrm{e}^{-\mathrm{i}\theta}}{4\rho^{3/2}}(\nabla\rho)^2 - \sqrt{\rho}\mathrm{e}^{-\mathrm{i}\theta}(\nabla\theta)^2 - i\sqrt{\rho}\mathrm{e}^{-\mathrm{i}\theta}\nabla^2\theta\end{aligned} \tag{18}$$

从而有

$$\phi^*\nabla^2\phi + \phi\nabla^2\phi^* = \nabla^2\rho - \frac{1}{2\rho}(\nabla\rho)^2 - 2\rho(\nabla\theta)^2 = 2\sqrt{\rho}\nabla^2\sqrt{\rho} - 2\rho(\nabla\theta)^2 \tag{19}$$

从而我们得到原本的式子右边等于

$$-\frac{\hbar^2}{2m}\frac{1}{\sqrt{\rho}} + \frac{\hbar^2}{2m}(\nabla\theta)^2 + V + U\rho \tag{20}$$

将 ∇ 作用到等式两边，得到

$$m\frac{\partial\mathbf{v}}{\partial t} = -\nabla\cdot\left[-\frac{\hbar^2}{2m}\frac{1}{\sqrt{\rho}}\nabla^2\sqrt{\rho} + \frac{1}{2}m\mathbf{v}^2 + V + U\rho\right] \tag{21}$$

以上的推导都是零温情况，只包含超流成分；有限温的问题可以加入正常流体的成分，从而变成二流体模型．考虑一个均匀系统 $V(\mathbf{r}) = 0$，基态有均匀的密度 ρ_0，且 $\mathbf{v} = 0$．我们可以对非基态的态的密度函数作展开：$\rho = \rho_0 + \delta\rho$．只取 $\delta\rho, \mathbf{v}$ 和 k 的领头阶，我们可以简化刚刚得到的方程：

$$m\frac{\partial\mathbf{v}}{\partial t} = -U\nabla\cdot\delta\rho, \quad \frac{\partial\rho}{\partial t} = -\nabla(\rho\mathbf{v}) \approx -\rho_0\nabla\mathbf{v} \tag{22}$$

联合起来就得到

$$\frac{\partial^2\delta\rho}{\partial t^2} = \frac{U\rho_0}{m}\nabla^2\delta\rho \tag{23}$$

即得到了声子色散

$$\omega = \sqrt{\frac{U\rho_0}{m}}k \tag{24}$$

声速得到为 $c = \sqrt{\dfrac{U\rho_0}{m}}$，是一个低能的线性 gapless 激发（从而无色散）. 这与 U(1) 对称性是联系起来的.

最后我们来说一下超流与临界速度. 这个临界速度就和我们 BCS 里面的临界磁场有点像：超过这个速度，超流就不再是超流了. 考虑一个质量为 m_0，速度 $\mathbf{v}_i$ 的杂质，在与某处于 BEC 态的物体相互作用. 如果有摩擦，也就是说，这个杂质被散射到某一速度 $\mathbf{v}_f$ 了，我们可以在初始 BEC 质心系写能动量守恒

$$m_0\mathbf{v}_i + \mathbf{q} = m_0\mathbf{v}_f, \quad \frac{m_0\mathbf{v}_i^2}{2} = \frac{m_0\mathbf{v}_i^2}{2} + c|\mathbf{q}| \tag{25}$$

得到

$$\mathbf{v}_i \cdot \mathbf{q} - c|\mathbf{q}| = \frac{\mathbf{q}^2}{2m} \tag{26}$$

可见右边等式大于 0，如果 $\mathbf{v}_i < c$，则左边无法满足也为正，即无法发生散射，也就是真正意义的“超流”（不会产生摩擦作用），这个临界速度也就是声速 c.

第十六部分

原子分子物理

第一章

原子

类氢原子的定态波函数

预备知识 球坐标系中的定态薛定谔方程[1409]，原子单位制[1315]

本文使用原子单位制[1315]．**类氢原子（hydrogen-like atom）**被定义为原子核有 Z 个质子（核电荷为 $+Ze$）有一个核外电子的原子/离子，例如氢原子和失去一个电子的氦原子 He^{+}，失去两个电子的锂离子 Li^{++}．

类氢原子的定态薛定谔方程为

$$-\frac{1}{2m}\nabla^2\psi(\mathbf{r})-\frac{Z}{r}\psi(\mathbf{r})=E\psi(\mathbf{r}) \tag{1}$$

类氢原子是唯一存在解析解的原子（离子）．

我们这里只讨论束缚态，即 $E<0$ 的解．从数学上，E 取小于零的任意值时我们都能找到解，但只有当 E 取特定离散值的时候这些波函数才能归一化（否则没有物理意义）．由于类氢原子具有球对称性，球坐标下的波函数具有最简单的形式．波函数的表达式为

$$\psi_{nlm}(r,\theta,\phi)=R_{nl}(r)Y_{l,m}(\theta,\phi) \tag{2}$$

其中 n 是**主量子数**（$n=1,2,\dots$），l 是**角量子数**（$l=0,1,\dots,n-1$），m 是**磁量子数**（$m=-l,-l+1,\dots,l$）．$R_{nl}(r)$ 是归一化的**径向波函数**，$Y_{l,m}(\theta,\phi)$ 是归一化的**球谐函数**（见“球谐函数列表[765]”）．

径向波函数 $R_{nl}(r)$

如果忽略原子核的运动，以下的 a 是玻尔半径（原子单位中 $a=1$），如果不忽略，a 就是约化玻尔半径．注意 Z 和 a 的作用是把径向波函数关于原点收缩 Z/a 倍（并保持波函数归一化）．

$$R_{nl}(r)=\sqrt{\left(\frac{2Z}{na}\right)^3\frac{(n-l-1)!}{2n(n+l)!}}\left(\frac{2Zr}{na}\right)^l L_{n-l-1}^{2l+1}\left(\frac{2Zr}{na}\right)\mathrm{e}^{-Zr/(na)} \tag{3}$$

其中 $L_n^l(x)$ 是**连带拉盖尔多项式（associated Laguerre polynomial）**．

$Z=1$ 时 $rR_{n,l}(r)$ 的函数图[1]如图 2

[1]我们以后会看到 $rR_{l,m}(r)$ 比 $R_{l,m}(r)$ 更常用

以下给出前几个径向波函数，注意所有径向波函数的值都是实数.

$$n=1 \qquad R_{10}(r)=2\left(\frac{Z}{a}\right)^{3/2}\exp(-Zr/a) \tag{4}$$

$$n=2 \qquad \begin{cases} R_{20}(r)=\dfrac{1}{\sqrt{2}}\left(\dfrac{Z}{a}\right)^{3/2}\left(1-\dfrac{1}{2}\dfrac{Zr}{a}\right)\exp\left(-\dfrac{Zr}{2a}\right) \\ R_{21}(r)=\dfrac{1}{2\sqrt{6}}\left(\dfrac{Z}{a}\right)^{3/2}\dfrac{Zr}{a}\exp\left(-\dfrac{Zr}{2a}\right) \end{cases} \tag{5}$$

$$n=3 \qquad \begin{cases} R_{30}(r)=\dfrac{2}{3\sqrt{3}}\left(\dfrac{Z}{a}\right)^{3/2}\left(1-\dfrac{2}{3}\dfrac{Zr}{a}+\dfrac{2}{27}\dfrac{Z^2r^2}{a^2}\right)\exp\left(-\dfrac{Zr}{3a}\right) \\ R_{31}(r)=\dfrac{8}{27\sqrt{6}}\left(\dfrac{Z}{a}\right)^{3/2}\left(1-\dfrac{1}{6}\dfrac{Zr}{a}\right)\dfrac{Zr}{a}\exp\left(-\dfrac{Zr}{3a}\right) \\ R_{32}(r)=\dfrac{4}{81\sqrt{30}}\left(\dfrac{Z}{a}\right)^{3/2}\dfrac{Z^2r^2}{a^2}\exp\left(-\dfrac{Zr}{3a}\right) \end{cases} \tag{6}$$

更多 $R_{n,l}$ 可以用 Mathematica 或者 Wolfram Alpha 生成，命令如

```
n = 4; l = 1;
Sqrt[(2*Z/(n*a))^3 * Factorial[n-l-1]/(2*n*Factorial[n+l])]
*(2*Z*r/(n*a))^l*LaguerreL[n-l-1, 2l+1, 2Z*r/(n*a)] * Exp[-Z*r/(n
```

或者用 Matlab 进行数值计算

```
Rnl = @(Z,n,l,r)sqrt((2*Z/n)^3*factorial(n-l-1)/(2*n*factorial(n
   +l))) *...
2        (2*Z*r/n).^l .* laguerreL(n-l-1, 2*l+1, 2*Z*r/n) .* exp
            (-Z*r/n);
```

性质

我们要求氢原子每个束缚态满足归一化条件

$$\int |\Psi_{n,l,m}(\mathbf{r})|\,\mathrm{d}^3r=\int\int_0^\infty |R_{n,l}(r)Y_{l,m}(\hat{\mathbf{r}})|^2 r^2\,\mathrm{d}r\,\mathrm{d}\Omega=1 \tag{7}$$

先做角向积分，球谐函数已经满足归一化条件

$$\int |Y_{l,m}(\hat{\mathbf{r}})|^2\,\mathrm{d}\Omega=1 \tag{8}$$

得到径向波函数的归一化条件为

$$\int [rR_{n,l}(r)]^2 \,\mathrm{d}r = 1 \tag{9}$$

再来看正交性，我们知道哈密顿算符的本征态之间两两正交（式中 n', l', m' 至少有一个与 n, l, m 不同）

$$\int \Psi^*_{n,l,m}(\mathbf{r})\Psi_{n',l',m'}(\mathbf{r})\,\mathrm{d}^3r = \int\int_0^\infty R_{n,l}(r)R_{n',l'}(r)Y^*_{l,m}(\hat{\mathbf{r}})Y_{l',m'}(\hat{\mathbf{r}})r^2\,\mathrm{d}r\,\mathrm{d}\Omega = 0 \tag{10}$$

同样可以先对角向做积分，若 l', m' 至少有一个与 l, m 不同，那么积分直接为零，径向波函数不需要任何正交条件（也的确不满足）．但若两个球谐函数相同，即 $l' = l$，$m' = m$，$n' \neq n$，那么角向积分等于 1，径向波函数满足

$$\int\int_0^\infty R_{n,l}(r)R_{n',l}(r)r^2\,\mathrm{d}r = 0 \tag{11}$$

径向概率分布

我们来求径向概率分布 $P(r)$．$P(r)$ 的定义为：发现粒子在 $r \in [a, b]$（厚球壳）内的概率为 $\int_a^b P(r)\,\mathrm{d}r$．由于波函数的模长平方就是三维的概率密度，有

$$\int_a^b P(r)\,\mathrm{d}r = \int_a^b \int \left|\frac{1}{r}\psi_{l,m}(r)Y_{l,m}(\hat{\mathbf{r}})\right|^2 \mathrm{d}\Omega\, r^2\mathrm{d}r = \int_a^b |\psi_{l,m}(r)|^2\,\mathrm{d}r \tag{12}$$

对任意 $a, b > 0$ 都成立，所以有

$$P(r) = |\psi_{l,m}(r)|^2 \tag{13}$$

任意波函数可以表示为所有本征波函数的叠加

$$\Psi(\mathbf{r}) = \frac{1}{r}\sum_{l,m}\psi_{l,m}(r)Y_{l,m}(\hat{\mathbf{r}}) \tag{14}$$

其径向概率分布为

$$\int_a^b P(r)\,\mathrm{d}r = \int_a^b \int \left|\frac{1}{r}\sum_{l,m}\psi_{l,m}(r)Y_{l,m}(\hat{\mathbf{r}})\right|^2 \mathrm{d}\Omega\, r^2\mathrm{d}r = \sum_{l,m}\int_a^b |\psi_{l,m}(r)|^2\,\mathrm{d}r \tag{15}$$

对任意 $a, b > 0$ 都成立，所以有

$$P(r) = \sum_{l,m}|\psi_{l,m}(r)|^2 \tag{16}$$

动量表象 动量分布

要求动量表象下的波函数，我们需要将位置表象的波函数投影到归一化的动量的本征矢上，即三维傅里叶变换

$$\psi_{nlm}(\mathbf{p}) = \langle \mathbf{p}|\psi\rangle = \frac{1}{\sqrt{2\pi}} \int \exp(-\mathrm{i}\mathbf{p}\cdot\mathbf{r})\psi(\mathbf{r})\,\mathrm{d}^3 r \tag{17}$$

这个积分在球坐标中完成才是最方便的，具体方法我们将举例子说明（见例 1 [767]）.

正如位置表象下位置的分布函数是 $|\psi(\mathbf{r})|^2$，动量表象下动量的分布函数是 $|\psi(\mathbf{p})|^2$ （也符合测量理论）.

氢原子波函数分析

预备知识 库仑波函数[1431]

光电子动量谱

在计算类氢原子光电离时，当外场消失后，每个能量本征态（散射态）的概率就固定不变了. 然而动量的本征态系数还是会变（除非时间无穷大）. 要得到时间无穷大时电子的三维动量分布，我们可以直接将波函数投影到库仑波函数（渐进平面波）上. 事实上这样的动量谱通常被称为 angular resolved energy spectrum （毕竟是能量的本征态），为了方便我们还是直接叫做动量谱.

$$P(\mathbf{k}) = |f(\mathbf{k})|^2 = \left|\left\langle \Psi_{\mathbf{k}}^{(-)}\middle|\Psi(\mathbf{r})\right\rangle\right|^2 \tag{1}$$

归一化为

$$\int P(\mathbf{k})\,\mathrm{d}^3 k = \iint P(\mathbf{k})\,\mathrm{d}\Omega\cdot k^2\,\mathrm{d}k = 1 \tag{2}$$

那么被电离的光电子（PE 或 photo-electron）的能量谱如何计算呢？先看动量大小的分布，归一化为

$$\int_0^\infty P(k)\,\mathrm{d}k = 1 \tag{3}$$

能量分布归一化为

$$\int P(E)\,\mathrm{d}E = \int P(E)k\,\mathrm{d}k = 1 \tag{4}$$

二者的关系为 $P(k) = kP(E)$.

我们只需要把总波函数投影到动量绝对值为 k 的子空间上即可. 对比以上几式得

$$P(k) = k^2 \int P(\mathbf{k})\,\mathrm{d}\Omega \tag{5}$$

但是类氢原子的波函数一般在球坐标中进行，我们试图直接用（径向）库仑函数来计算. 将波函数投影到归一化得库仑球面波式 9 [1431] 得

$$P(l,m,k) = |f_{l,m}(k)|^2 \tag{6}$$

其中

$$f_{l,m}(k) = \langle C_{l,m}(k)|\Psi(\mathbf{r})\rangle = \sqrt{\frac{2}{\pi}} \int F_l(k,r)\psi_{l,m}(r)\,\mathrm{d}r \tag{7}$$

其中 $\psi_{l,m}(r)$ 是 scaled 的径向波函数式 1 [1412].

归一化为[2]

$$\sum_{l,m} \int P(l,m,k)\,\mathrm{d}k = 0 \tag{8}$$

对比式 3 和式 4 得

$$P(k) = \sum_{l,m} P(l,m,k) = \frac{2}{\pi} \sum_{l,m} \left| \int F_l(k,r)\psi_{l,m}(r)\,\mathrm{d}r \right|^2 \tag{9}$$

虽然这个公式看起来只包括了径向动能的分布，但实际上也有角向的动能，体现在 l 量子数里面[3].

电子轨道与元素周期表

- 本词条存在未完成的内容.

在这里，用最简单的方式介绍原子的壳层结构，并解释元素周期表如何根据壳层结构分出每个周期. 在玻尔原子模型[1313] 中，原子轨道如图 1.

现在要把电子放到这些轨道上面来，使电子的总能量最小. 这种状态叫做原子的基态. 理想的状态是，所有电子都在最小的一圈轨道上，但是由于每条轨道只能容纳一定数目的电子，另一些电子不得不进入其他轨道.

[2] 为什么对 k 得积分没有 k^2 项？这取决有库伦球面波得归一化式 10 [1431].

[3] 想一下库仑函数的微分方程中 l 是如何决定角向动能的？注意与 m 无关.

为了解释每条轨道能容纳多少电子，把每个轨道的“空位”用一行格子描述．当一行格子被电子填满时，该轨道就不能容纳更多电子了．给不同的格子命名：从半径最小的轨道开始，用数字 1，2，3 等依次命名每条轨道，这些数字也叫主量子数，用 n 来表示．在上图中，每条轨道从第一列开始依次用 $S, P, D, F...$ 命名，每一列有相同的角量子数，用 l 来表示．$S, P, D, F\ldots$ 的角量子数依次是 $l = 0, 1, 2, 3\ldots$．把行标和列标组合起来，就能得到任意一个格子的名称，例如第三行第二列的格子叫做 $3P$．对于主量子数为 n 的行，角量子数从 0 增加到 $n-1$，也就是说第 n 行有 n 个格子．

接下来，每个格子又能装不同数目的电子，任意一格能装的电子数等于 $2(2l+1)$．这是因为，对于角量子数为 l 的格子，还存在另一个参数 m，叫磁量子数．对于特定的角量子数 l，m 可以取 $-l, -l+1...0, ...l-1, l$ 等 $2l+1$ 个不同的值．所以每个格子又可以根据不同的 m 细分成 $2l+1$ 个小格子．最后，每一个小格子里面能装两个电子．总共算下来，第 n 行刚好可以装 $2n^2$ 个电子（见图 XVI.1.3）．

到此为止，每条轨道承载电子的数目已经解释清楚了，但是应该如何把电子往格子里面放呢？为了使电子总能量最小，对于氢原子（1 个核外电子），显然电子应该放在 $1S$ 格子里，氦原子（2 个核外电子）可以把两个电子都放在 $1S$ 格子里，从而把 $n=1$ 的轨道填满，这就是第一周期的两个原子的电子分布．对于锂原子（3 个核外电子）可以在氦原子的基础上往 $2S$ 格子里放一个电子．但奇怪的是，填电子的顺序并不是从下到上从左到右，而是如下图中的箭头所示，即按照 $1S, 2S, 2P, 3S, 3P, 4S, 3D, 4P, 5S\ldots$ 的顺序来填上图的格子．格子内的数字表示每格能装下的电子数，即 $2(2l+1)$．

未完成：这张图在第四周期以上不成立，见 [14]

元素周期表的排序

要判断某个原子所在的周期，就先根据原子序号找出上图中所有装有电子的格子，其中 n 最大的格子就是该元素所在的周期．例如 30 号元素，可以按照上图绿色线条的顺序占满 $1S, 2S, 2P, 3S, 3P, 4S, 3D$（这些格子能容纳的总电子数刚好是 30）．其中 $4S$ 的主量子数最大，$n=4$，所以 30 号元素在第四周期．按照这个规律，把上图按照周期分类如下．

未完成：解释 1s22P2 之类的符号吧

能项符号

- 本词条处于草稿阶段.

预备知识 电子轨道与元素周期表[1540]，角动量加法（量子力学）[1403]

[4]**能项符号（term symbol）**描述了原子的总角动量状态

$$^{2S+1}L_J \tag{1}$$

其中 L 表示总轨道角动量（量子数），可以用 $S, P, D, F \ldots$ 等符号分别表示 $L = 0, 1, 2, 3, \ldots$；S 是所有电子的总自旋角动量，J 是所有电子的总角动量，包括轨道角动量和自旋角动量. $2S+1$ 叫做 multiplicity，是指确定了 S 和 L 以后 J 可能的取值个数（见 式 10 [1403]）.

能项符号假设自旋-轨道耦合可以忽略不计，所以 L 和 S 仍然是好量子数[1379].

Hund's Rule

Hund's Rule 可以决定原子处于基态时的能项符号

1.

2.

3.

电子排布

另外有一种符号例如碳原子（$Z = 6$）是 $(1s)^2(2s)^2(2p)^2$ 不知道叫什么.

氦原子的基态是 $(1s)^2$，总轨道角动量 $L = 0$. 由泡利不相容原理，两个电子自旋方向相反（singlet，反对称，$S = 0$），所以波函数是对称的，符号为 $^1S_0$.

[4]参考 [14]，section 2.2: The Periodic Table

氢线（21 厘米线）

- 本词条处于草稿阶段.
- 本词条缺少预备知识，初学者可能会遇到困难.

[5]宇宙中著名的 21cm 谱线是由氢原子基态的超精细常数产生. 频率为 $1,420,405,751.7667 \pm 0.0009$ Hz，对应波长 21.1061140542 cm 该波长的电磁波属于微波.

兰姆位移

- 本词条处于草稿阶段.
- 本词条缺少预备知识，初学者可能会遇到困难.

预备知识 玻尔原子模型[1313]

[6]根据狄拉克的量子理论，氢原子的 $2S_{1/2}$ 和 $2P_{1/2}$ 能级应该具有相同的能量，但兰姆的实验发现它们有微弱的偏差，前者高出约

$$\Delta E = 4.372 \times 10^{-6}\ \text{eV} \tag{1}$$

这被称为**兰姆位移（Lamb shift）**.

图 1 中显示，$n = 2$ 能级的精细结构由 $2P_{3/2}$ 和 $2P_{1/2}$ 构成，而兰姆位移由 $2S_{1/2}$ 和 $2P_{1/2}$ 构成. 兰姆位移和氢原子基态的超精细结构[1543] 在同一个数量级.

补充

（未完成）Hans Bethe 首先用量子电动力学解释了兰姆位移.

（未完成）兰姆位移用于精确测量精细结构常数.

[5]参考 Wikipedia 相关条目.

[6]参考 HyperPhysics 相关条目.

能级修正

事实上，兰姆位移对每个能级都有微小的修正[7]

$$\Delta E_{lamb} = \alpha^5 m_e c^2 \frac{k(n,0)}{4n^3} \qquad (l=0) \tag{2}$$

$$\Delta E_{lamb} = \alpha^5 m_e c^2 \frac{1}{4n^3}\left[k(n,l) \pm \frac{1}{\pi(j+1/2)(l+1/2)}\right] \qquad (l\neq 0,\ j=l\pm 1/2) \tag{3}$$

其中 $k(n,0)$ 约等于 13，$k(n,l)<0.05$. 将 $2S_{1/2}$（$l=0$）和 $2P_{1/2}$（$n=2,\ l=1,\ s=1/2$），得 $\Delta E = 4.365\times 10^{-6}$.

Keldysh 参数

- 本词条处于草稿阶段.

预备知识 ponderomotive 能量[??]

Keldysh parameter

Keldysh parameter 是一个无量纲常数，用于判断什么时候 tunnelling ionization 占主导

$$\gamma = \sqrt{\frac{I_p}{2U_p}} = \frac{\omega}{e\mathcal{E}_0}\sqrt{2m_e I_p} \tag{1}$$

I_p 为电离能量，U_p 为 ponderomotive 能量（**??** [??]）

$$U_p = \frac{q^2\mathcal{E}_0^2}{4m\omega^2} \propto I/\omega^2 \tag{2}$$

γ 越小，Tunnelling ionization 越容易发生．$\gamma<1$ 时 tunnelling ionization 会占主要部分.

[7] 详见 Bethe, H.A.; Salpeter, E.E. (1957). Quantum Mechanics of One- and Two-Electron Atoms. Springer. p. 103.

ADK 电离率

- 本词条处于草稿阶段.

[8] M. Ammosov, N. Delone, V. Krainov. 三个人给出了一般原子的瞬时的 tunnelling ionization rate，叫做 **ADK rate**.

$$\frac{\mathrm{d}P}{\mathrm{d}t} = W_{ADK} = \omega_p |C_{n^*l^*}|^2 G_{lm} \left(\frac{4\omega_p}{\omega_T}\right)^{2n^*-m-1} \exp\left(-\frac{4\omega_p}{3\omega_T}\right) \tag{1}$$

其中

$$\omega_p = I_p \qquad \omega_T = \frac{e\mathcal{E}_0}{\sqrt{2mI_p}} \qquad n^* = \sqrt{\frac{I_p^H}{I_p}} \tag{2}$$

$$|C_{n^*l^*}|^2 = \frac{2^{2n^*}}{n^*\Gamma(n^*+l^*+1)\Gamma(n^*-l^*)} \tag{3}$$

$$G_{lm} = \frac{(2l+1)(l+|m|)!}{2^{|m|}|m|!(l-|m|)!} \tag{4}$$

其中 I_p^H 是氢原子的电离能 I_p，l, m 是角动量量子数，effective quantum number $l^* = 0$ 当 $l \ll n$，否则 $l^* = n^* - 1$.

抛物线坐标系中的类氢原子定态波函数

- 本词条处于草稿阶段.
- 画图密度图应该会很漂亮

预备知识 定态薛定谔方程[1328]，抛物线坐标系[124]

[9]本文使用原子单位制[1315]. 令原子核不动.

$$-\frac{1}{2m}\nabla^2\psi - \frac{Z}{r}\psi = E\psi \tag{1}$$

[8]参考 [19] Chap.15.3

[9]参考 [19] Chap 3.5.

在抛物线坐标系中变为

$$-\frac{1}{2m}\left\{\frac{4}{\xi+\eta}\left[\frac{\partial u}{\partial \xi}\left(\xi\frac{\partial}{\partial \xi}\right)+\frac{\partial u}{\partial \eta}\left(\eta\frac{\partial}{\partial \eta}\right)\right]+\frac{1}{\xi\eta}\frac{\partial^2 u}{\partial \phi^2}\right\}\psi-\frac{2Z}{\xi+\eta}\psi=E\psi \tag{2}$$

分离变量，令

$$\psi(\xi,\eta,\phi)=f(\xi)g(\eta)\Phi(\phi) \tag{3}$$

和球坐标同理，$\Phi(\phi)=\exp(\mathrm{i}m\phi)$．令另外两个分离变量常数满足

$$\nu_1+\nu_2=Z \tag{4}$$

有

$$\frac{\mathrm{d}}{\mathrm{d}\xi}\left(\xi\frac{\mathrm{d}f}{\mathrm{d}\xi}\right)+\left(\frac{\mu E\xi}{2}-\frac{m^2}{4\xi}+\nu_1\right)f=0 \tag{5}$$

$$\frac{\mathrm{d}}{\mathrm{d}\eta}\left(\eta\frac{\mathrm{d}g}{\mathrm{d}\eta}\right)+\left(\frac{mE\eta}{2}-\frac{m^2}{4\eta}+\nu_2\right)g=0 \tag{6}$$

可以化简为 Kummer-Laplace 微分方程. 解出后，ν_1,ν_2 分别对应两个整数 n_1,n_2，称为抛物线量子数，和主量子数 n 的关系为

$$n=n_1+n_2+|m|+1 \tag{7}$$

令 $\rho_1=Z\xi/n$，$\rho_2=Z\eta/n$，定态波函数为

$$\begin{aligned}\psi_{n_1,n_2,m}(\xi,\eta,\phi)=&\frac{Z^{3/2}}{\sqrt{\pi}n^2}\sqrt{\frac{n_1!n_2!}{[(n_1+|m|)!(n_2+|m|)!]^3}}\\&\times \mathrm{e}^{-(\rho_1+\rho_2)/2}(\rho_1\rho_2)^{|m|/2}L^{|m|}_{n_1+|m|}(\rho_1)L^{|m|}_{n_2+|m|}(\rho_2)\mathrm{e}^{\mathrm{i}m\phi}\end{aligned} \tag{8}$$

能量本征值为

$$E_n=-\frac{Z^2}{2n^2} \tag{9}$$

FROG

- 本词条缺少预备知识，初学者可能会遇到困难.

如果我们要测量线偏振的超快激光脉冲的波形（即电场关于时间的函数），我们不能直接用仪器测量，因为电子元件的时间分辨率远远不够．但我们可以使用 Frequency Resolved Optical Gating （FROG）．FROG 有许多不同的实现方法，这里只讨论比较常见的 SHG FROG（Second Harmonic Generation FROG）和 **PCGPA**．

Frog 的大概原理就是把被测量的脉冲（prob）$f(t)$ 和另一个脉冲（gate）$g(t)$ 叠加（注意 $f(t)$ 和 $g(t)$ 都是实函数），然后通过一个二阶非线性介质产生一个 $f(t)g(t)$ 脉冲并分离出来，再测量光谱（即傅里叶变换）．接下来我们可以控制两个脉冲的相对延迟（time delay）τ，就得到不同延迟下 $f(t-\tau)g(t)$ 的光谱，也就是一个二维函数，叫做 **FROG trace**（图 2）．很多情况下，gate 就是 prob 通过一个 beam splitter 分出来的，实验光路如图 1 所示．

$$a(\omega,\tau)=\left|\int_{-\infty}^{\infty}f(t-\tau)g(t)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t\right|^2 \tag{1}$$

只要用特定的算法，就可以从式 1 中解出 $f(t)$ 和 $g(t)$．

PCGPA 算法

PCGPA（Principal Component General Projection Algorithm）就是从式 1 中解出 $f(t)$ 和 $g(t)$ 的一种常见算法．在得到我们假设 $f(t)$ 和 $g(t)$ 都是等时间间隔的离散点 $f_1,f_2\ldots,f_N$ 和 $g_2,g_2\ldots,g_N$，可以看成两个列矢量．我们先来做两矢量的外积，即

$$\mathbf{f}\mathbf{g}^{\mathrm{T}}=\begin{pmatrix}f_1g_1 & \cdots & f_1g_N\\ \vdots & \ddots & \vdots\\ f_Ng_1 & \cdots & f_Ng_N\end{pmatrix} \tag{2}$$

然后我们把第 i 行向左移动 $i-1$ 个元（左边多出的矩阵元补到右边），得到

$$\begin{pmatrix}f_1g_1 & f_1g_2 & \cdots & f_1g_N\\ f_2g_2 & f_2g_3 & \cdots & f_2g_1\\ \vdots & \vdots & \vdots & \vdots\\ f_Ng_N & f_Ng_1 & \cdots & f_Ng_{N-1}\end{pmatrix} \tag{3}$$

现在可以发现每一列恰好是 $f(t-\tau)g(t)$ 的离散形式[10]，第 i 列 $\tau=(i-1)\Delta t$. 然而当 $i>N/2$ 的时候，更自然的理解是 $\tau=(i-N)\Delta t$. 例如与其认为第 N 列是 g 向上移动了 $N-1$ 个元，倒不如认为是向下移动了 1 个元. 根据这种思想，我们可以把矩阵左半和右半调换得到

$$\begin{pmatrix} f_1g_{N/2+1} & f_1g_{N/2+2} & \cdots & f_1g_{N/2} \\ f_2g_{N/2+2} & f_2g_{N/2+3} & \cdots & f_2g_{N/2+1} \\ \vdots & \vdots & \vdots & \vdots \\ f_Ng_{N/2} & f_Ng_{N/2+1} & \cdots & f_Ng_{N/2-1} \end{pmatrix} \tag{4}$$

这样，从左到右的每列分别对应从 $\tau=-(N/2)\Delta t,\ldots,(N/2-1)\Delta t$.

现在我们对每一列做离散傅里叶变换（注意前后都要 fftshift），就得到了含相位的 Frog trace 矩阵，其中的每一列都是式 1 中的

$$\int_{-\infty}^{\infty} f(t-\tau)g(t)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t \tag{5}$$

[10]唯一的区别是从最上面移出的 g_i 跑到了最下面，而在实验中最下面应该由 0 来填补. 但如果 **f** 和 **g** 矢量的首尾都有足够多的 0，这个问题就自动解决了.

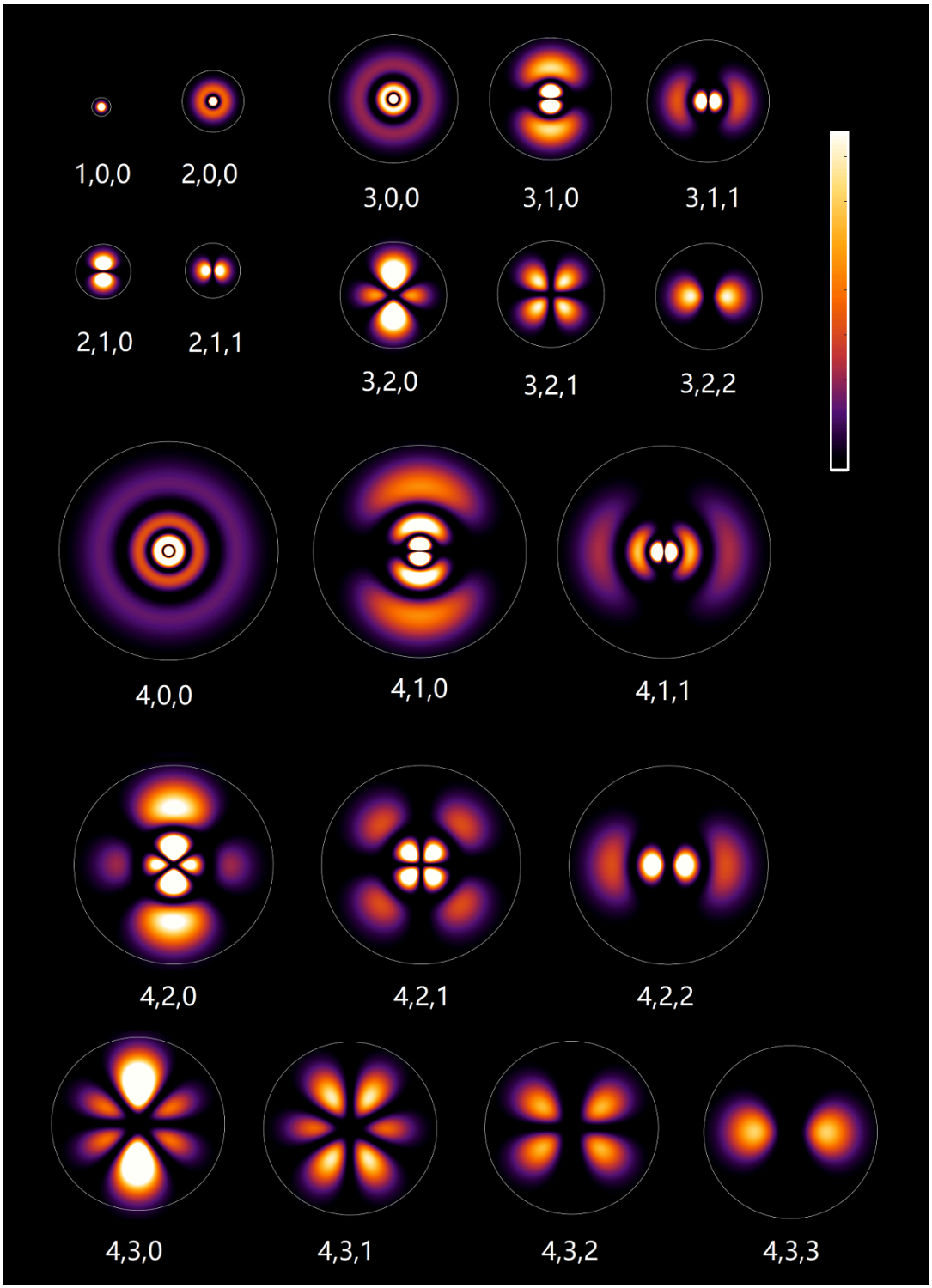


图 1: 氢原子波函数的概率密度 $|\psi_{nlm}(\mathbf{r})|^2$ 的 x-z 截面，大小成比例．每个图中的三个数字分别是量子数 n, l, m，电子出现在白色圆圈内部的概率为 0.95．色条对应的数值是线性的，每个子图中的色条取值范围不相同．另见 Matlab 画图程序[1923]．

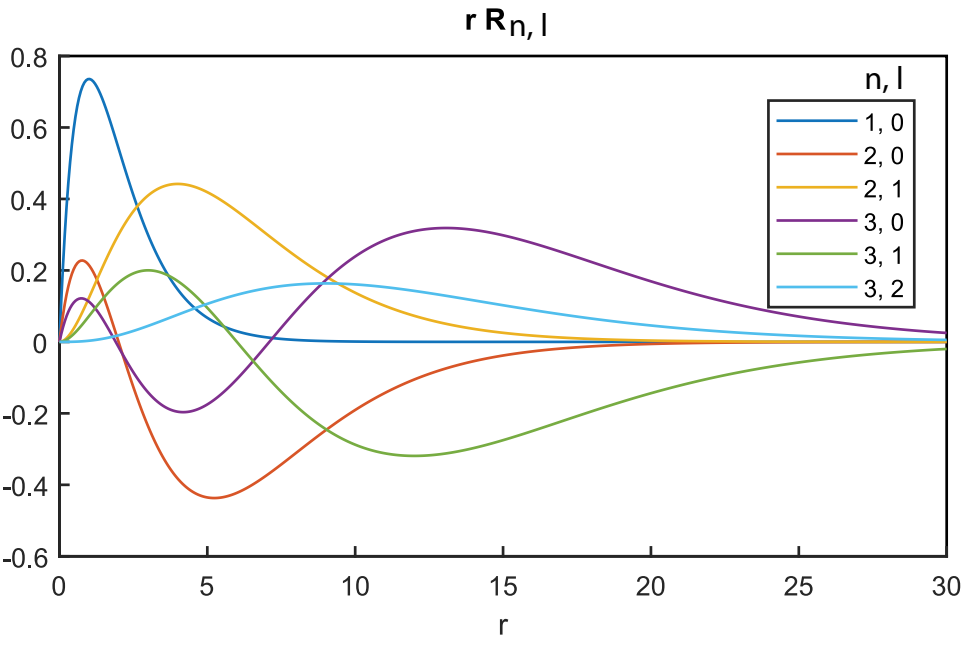


图 2: 径向波函数函数图（使用原子单位[1315]，$a = 1$）

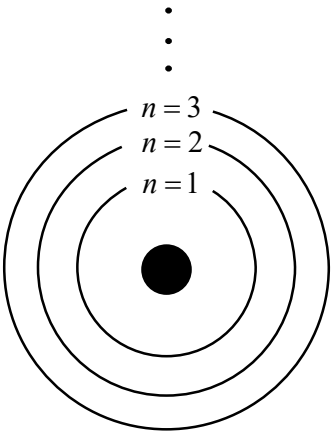


图 1: 电子轨道

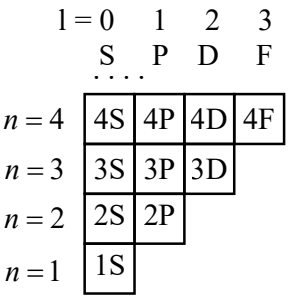
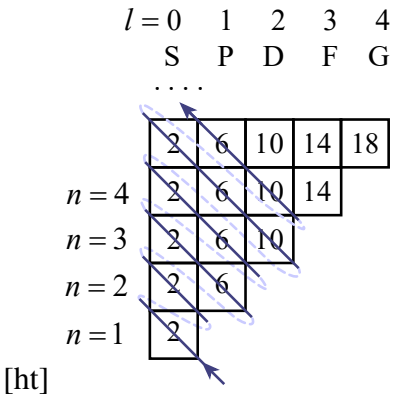


图 2: 用格子描述电子轨道



[ht]

图 3: 轨道的填充顺序

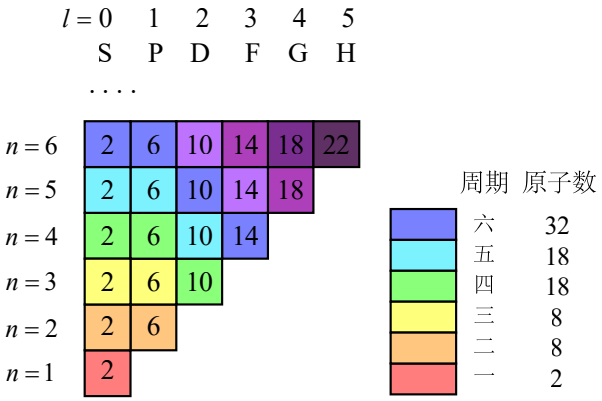


图 4: 划分周期

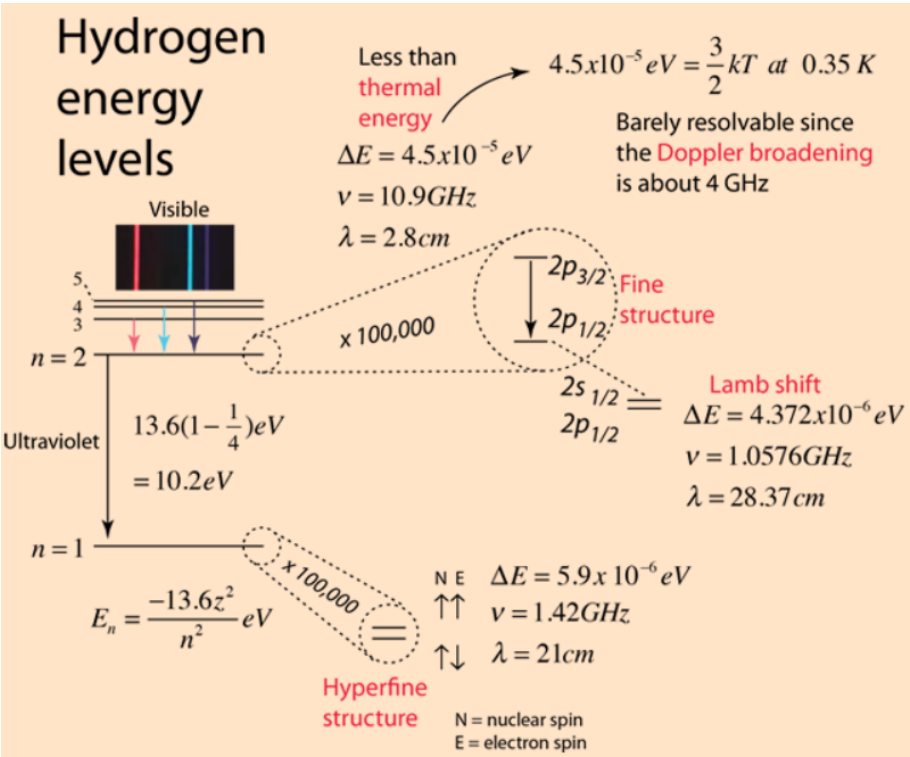


图 1: 氢原子的精细结构，超精细结构，以及兰姆位移（来自 HyperPhysics）

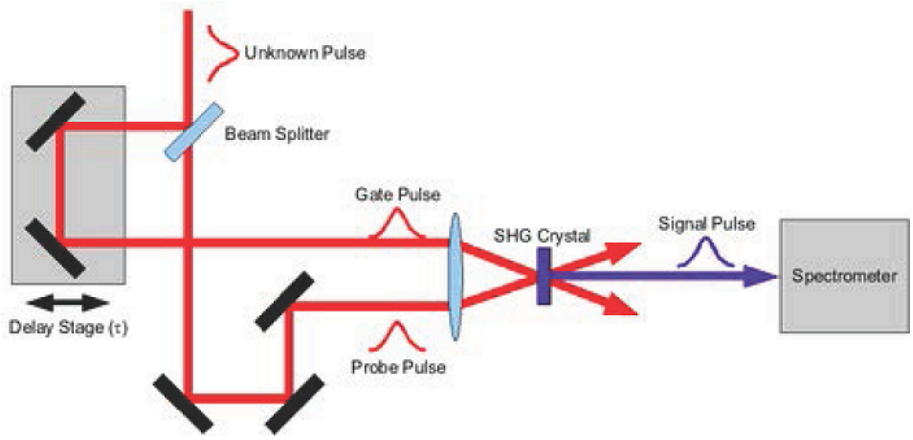


图 1: SHG FROG 的光路（图片来自 WikiPedia）

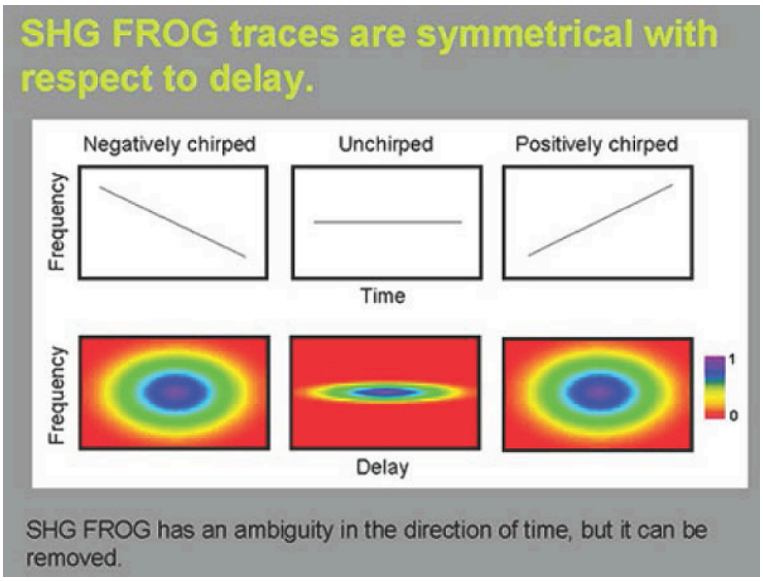


图 2: SHG FROG trace（图片来自网络）

第二章

分子

双原子分子势能曲线

- 本词条处于草稿阶段.

双原子分子势能曲线

双原子分子势能曲线（Potential Energy Curves）是认识和描述双原子分子、离子动力学最方便的工具之一，在研究双原子分子、离子的过程中被广泛地使用．它的横坐标是原子间距，纵坐标是势能．一个典型的双原子分子势能曲线如图 1[1]

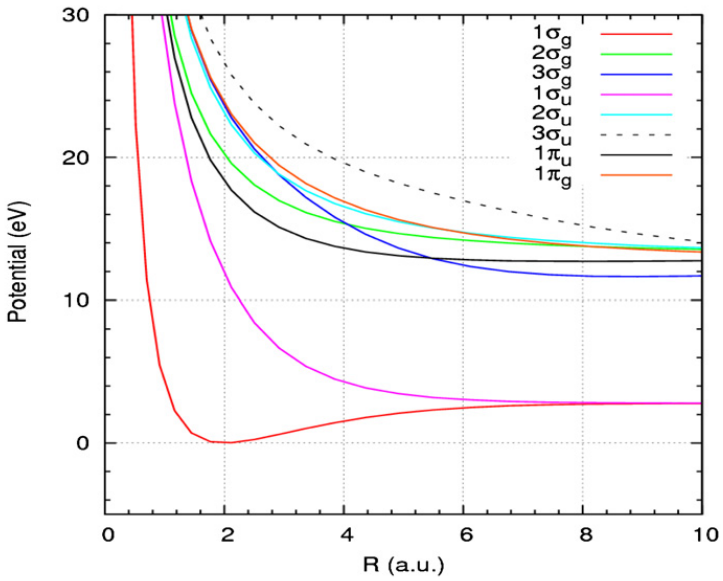


图 1: D_2^+ 势能曲线图

通常在双原子分子势能曲线图中，不同电子态的曲线会在势能曲线旁边标注．通过使用势能曲线，我们可以更好地认识和描述双原子分子的电离、解离等过程．

[1]图片来自：Li,H,et al."Intensity dependence of the attosecond control of the dissociative ionization of D2." Journal of Physics B Atomic Molecular & Optical Physics 47.12(2014):124020

第十七部分

热力学和统计力学

第一章

热力学

气体分子对容器壁的压强

预备知识 动量定理[944]

我们来从微观的角度考察气体是如何对一个光滑平面产生压强的. 一种错误的解释是：每个气体分子像一个有弹性的小球，他们互紧挨着，对彼此和容器壁产生压力，当体积越小，压力也就越大. 然而真实情况[1]是，气体分子之间的距离远远大于他们的体积，且都在不断运动. 是大量分子撞击容器壁产生的“冲击力”对容器产生了等效的压力. 这种撞击可以看作是在一瞬间完成的，就像高中学的完全弹性碰撞. 所以我们需要使用动量和冲量.

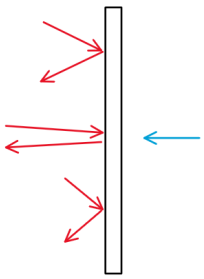


图 1: 分子对容器壁的压强

如图 1，我们假设空间中没有重力，向右为 x 方向，一块面积为 S 的光滑平板初始时静止，令其质量为 M，远大于分子的质量. 平板的左侧不断受到大量粒子从各个随机方向，以各种随机速度的撞击. 虽然每个分子的动量很小，但撞击的频率却很大（数量级与阿伏伽德罗常数相当，即 10^{24} 次每秒）. 由于这些撞击的位置完全是随机的，他们可以被等效为一个均匀的压强. 如何定义等效压强呢？我们可以在平板的右侧对平板施加一个均匀恒定的“真正的”压强 P （或者等效地，直接施加恒力 $F = PS$），如果左边分子的撞击和右边的压强能使平板保持宏观的静止，那么我们就说分子撞击对平板的（等效）压强为 P.

我们给这些分子编号，假设第 i 个分子延 x 方向的速度分量为 $v_{x,i}$，质量都为 m，则碰撞前延 x 方向的动量为 $p_i = mv_{x,i}$. 由于平板的质量远大于单个

[1]除了极端温度或压强的情况

分子的质量，撞击以后可以认为分子 x 方向的速度取相反数，而平行于平板方向的速度不变（类似于光的镜面反射）. 这样，碰撞后单个分子的动量变化（即冲量）为 $\Delta p_i = -2mv_{x,i}$，由动量守恒，平板的动量瞬间增加了 $2mv_{x,i}$. 如果一段时间 Δt 内，有 N 个分子撞击平板，则平板受到向右的总冲量为

$$\Delta p = \sum_{i=1}^{N} 2mv_{x,i} = 2m\sum_{i=1}^{N} v_{x,i} \tag{1}$$

再来看右边的压力对平板的作用. Δt 时间内该作用力对平板向左的冲量为

$$\Delta p = PS\Delta t \tag{2}$$

若要使平板在宏观上不动，总冲量必须为零，以上两式相等，即

$$P = \frac{2m}{S\Delta t}\sum_{i=1}^{N} v_{x,i} = 2m\bar{v}_{x,i}\frac{N}{S\Delta t} \tag{3}$$

其中我们定义 x 方向速度平均值为

$$\bar{v}_{x,i} = \frac{1}{N}\sum_{i=1}^{N} v_{x,i} \tag{4}$$

也就是说，等效压强和分子在垂直容器壁方向的平均动量成正比，与单位面积单位时间撞击容器壁的粒子数成正比.

两块平板

我们再来讨论一个稍微复杂一些的情况. 假设空间中有两块平行平板，他们之间距离为 a，每个粒子会在这两块平板之间来回反弹（假设不飞出边界），第 i 个粒子来回运动一次的周期是 $T_i = 2a/v_{x,i}$，Δt 时间（假设远大于 T_i）内可以与右边的平板发生碰撞的次数为

$$N_i = \frac{\Delta t}{T_i} = \frac{v_{x,i}\Delta t}{2a} \tag{5}$$

所有的分子在 Δt 时间内碰撞右壁的次数为 $N = \sum N_i$.

所以 Δt 时间内，右侧平板获得的总冲量为

$$\Delta p = \sum_{i=1}^{N} 2mv_{x,i}N_i = \frac{m\Delta t}{a}\sum_{i=1}^{N} v_{x,i}^2 = m\overline{v_{x,i}^2}\frac{N\Delta t}{a} \tag{6}$$

其中我们定义了速度平方得平均值（注意这与式 4 的平方 $\bar{v}_{x,i}^2$ 不同）

$$\overline{v_{x,i}^2} = \frac{1}{N}\sum_{i=1}^{N} v_{x,i}^2 \tag{7}$$

与上面同理，为了保持总冲量相等，式 6 和式 2 必须相等，得

$$P = m\overline{v_{x,i}^2}\frac{N}{Sa} \tag{8}$$

从这个公式出发，我们很容易可以得到理想气体状态方程[1560]．注意这里的 Sa 就是两平板之间的体积，而 $mv_{x,i}^2$ 就是单个分子动能 $E_{k,i}$ 的两倍．

理想气体状态方程

预备知识 准静态过程[1580]，分子撞击对容器壁的压强[1558]

理想气体是热力学中一个理想化的模型（可以类比自由落体[889]，简谐振子等[913]），用于探索密闭容器中一定量气体的压强，体积和温度之间的关系．该模型假设密闭容器中有一定量的气体（由原子或分子构成，以下统称为分子），假设气体非常稀薄，分子之间几乎不发生相互作用（例如碰撞），且与容器壁之间的碰撞是完全弹性的且在瞬间完成的，假设每个分子所受重力忽略不计．

理想气体状态方程最开始由实验得到，这是因为在日常环境下，大部分气体与理想气体符合较好．其最常见的形式为

$$PV = nRT \tag{1}$$

其中 P 是气体的压强（处处相等），V 是容器的体积，n 是气体分子的摩尔数，R 是**气体常数（gas constant）**，T 是热力学温度（单位为开尔文“K”见下文）．

一摩尔包含的粒子个数由**阿伏伽德罗常数（Avogadro constant）**精确定义为

$$N_A = 6.02214076 \times 10^{23} \tag{2}$$

为了定义气体常数，我们先来介绍热力学中一个重要的常数——**玻尔兹曼常数（Boltzmann constant）**，它被精确定义为

$$k_B = 1.380649 \times 10^{-23}\ \mathrm{J/K} \tag{3}$$

理想气体常数被定义为

$$R = k_B N_A = 8.31446261815324\ \mathrm{J/K} \tag{4}$$

这个数值同样是精确的.

令容器中总分子数为 $N = nN_A$，所以状态方程（式 1）也可以记为

$$PV = Nk_BT \tag{5}$$

从微观角度，**热力学温度**[1562] 可以由理想气体分子平均动能定义[2]

$$\bar{E}_k = \frac{3}{2}k_BT \tag{6}$$

所以气体分子总动能为

$$E_k = \frac{3}{2}Nk_BT = \frac{3}{2}nRT \tag{7}$$

例 1　标准状况下气体分子密度

我们可以由理想气体状态方程得出化学中一个常用的常数：标准状况下（273.15 K、101 kPa）每摩尔气体的体积约为 22.4L. 将温度和压强代入式 1 得

$$\frac{V}{n} = \frac{RT}{P} = \frac{8.314 \times 273.15}{1.01 \times 10^5}\ \mathrm{m^3/mol} = 22.48\ \mathrm{L/mol} \tag{8}$$

注意理想气体的假设使结果有微小误差.

由经典力学推导

我们在“分子撞击对容器壁的压强[1558]”中已经详细推导了压强和分子速度的关系，以下重复的部分将简略带过. 注意分子密度趋近于 0 的假设导致分子之间没有作用力，每个分子的运动都是可以看作是独立的.

假设长方体容器的 x, y, z 三个方向的边长分别为 a, b, c，则体积为 $V = abc$. 考虑一个初始延任意方向运动的分子，与容器壁发生完全弹性碰撞，它在 x 方向的周期（即运动一个来回所需的时间）为 $2a/v_x$，每个周期带给 $x = a$ 容器壁的冲量为 $2mv_x$，该容器壁面积为 bc，所以受到该粒子的平均压强 P 为冲量除以周期除以面积 mv_x^2/V，即

$$P_iV = mv_{x,i}^2 \tag{9}$$

[2]注意这里的动能是平动动能，我们暂时不讨论气体分子的转动.

注意这里我们用角标 i 来表示第 i 个分子．如果有 N 个分子，质量都为 m，那么

$$PV = \sum P_i V = 2N\left(\frac{1}{2}m\overline{v_x^2}\right) = 2N\bar{E}_{kx} \tag{10}$$

注意 $\overline{v_x^2}$ 和 $\bar{v}_x^2$ 是不同的．前者是先对每个分子的速度取平方再平均．而后者是先计算速度的平均值在平方．等式右边等于 $2N$ 乘以所有分子在 x 方向的平均动能．由于我们假设分子运动各向同性（即不会出现某些方向的分子运动较快），所以平均的总动能等于单方向平均动能的 3 倍（注意这里我们假设是 3 维空间，如果是 N_d 维空间，就是 N_d 倍）

$$\bar{E}_k = \frac{1}{2}m(\overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2}) = \frac{3}{2}m\overline{v_i^2} \tag{11}$$

所以有（参考式 8 [1558]）

$$PV = \frac{2}{3}N\bar{E}_k \tag{12}$$

这说明压强和体积的乘积等于分子总平动动能的 2/3．用式 6 消去 E_k，就得到了式 5 形式的理想气体状态方程．

温度 温标

预备知识 理想气体状态方程[1560]

三种温标

对于气体而言，温度越高意味着气体分子速度越大（式 6 [1560]），而对于固体，温度越高说分振动越剧烈．在理想气体模型中，我们看到温度与气体分子的平均动能成正比．

$$\bar{E}_k = \frac{3}{2}k_B T_K \tag{1}$$

这样我们就在微观上定义了**热力学温标**（单位是开尔文，K，国际单位的一种）．当分子动能为 0 时，热力学温度就是 0 K，即**绝对零度**．根据这个定义，最低的可能温度就是绝对零度[3]．

[3]统计力学中的确有负温度这种说法，但根据定义，它的温度反而比任何温度要高．

在生活中，我们一般使用**摄氏温标**或**华氏温标**来表示温度，它们的单位分别记 ° C，° F，我们以下用 T_C 和 T_F 表示．三种温标的转换关系如下

$$T_K = T_C + 273.15^\circ\ \mathrm{C} \tag{2}$$

$$T_C = \frac{5}{9}(T_F - 32^\circ\ \mathrm{F}) \tag{3}$$

注意一开尔文和一摄氏度的大小一样，只是相差了一个常数．零下 273.15 摄氏度就是绝对零度．

理想气体的内能

- 本词条处于草稿阶段．

预备知识 理想气体状态方程[1560]，能均分定理[??]

理想气体的动能为（式 7 [1560]）

$$E_k = \frac{3}{2}Nk_BT = \frac{3}{2}nRT \tag{1}$$

分子运动中，三个方向的动能占三个自由，而对于多原子分子，还可能出现转动和振动等自由度．令自由度为 i，则内能（即总能量）为

$$E = \frac{i}{2}Nk_BT = \frac{i}{2}nRT \tag{2}$$

压强体积图

- 本词条处于草稿阶段．

预备知识 功[903]

如图 1 所示，做功

$$W = \int_{V_1}^{V_2} P(V)\,\mathrm{d}V \tag{1}$$

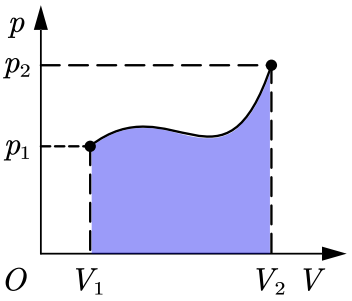


图 1: $p-V$ 图中气体做的功

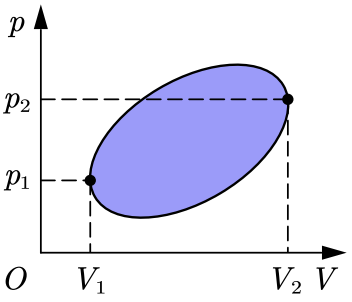


图 2: $p-V$ 图中闭合路径气体做的功

做功就是 $p-V$ 曲线下面的面积.

如图 2 所示，如果是一个闭合路径，那么对外做功就是闭合路径围出的面积（顺时针为正）.

如果是容器中的理想气体，可以画出等温曲线. 这样理想气体状态方程中的三个量都可以从图中看出.

热平衡 热力学第零定律

- 本词条处于草稿阶段.

热平衡

一个系统（例如一个绝热容器中的气体）若放置得足够久，它的内部就会达到热平衡.

热力学第零定律

若物体 A 与物体 C 达到热平衡，且物体 B 与物体 C 也达到热平衡，那么 A 和 B 之间同样有热平衡.

简单来说，热平衡意味着温度处处相等. 那么这个定律也就不证自明了.

热传导定律

预备知识 梯度梯度定理[307]

如果气体内各部分的温度不同，从温度较高处向温度较低处，将有热量的传递，这一现象叫做**热传导（heat conduction）**现象.

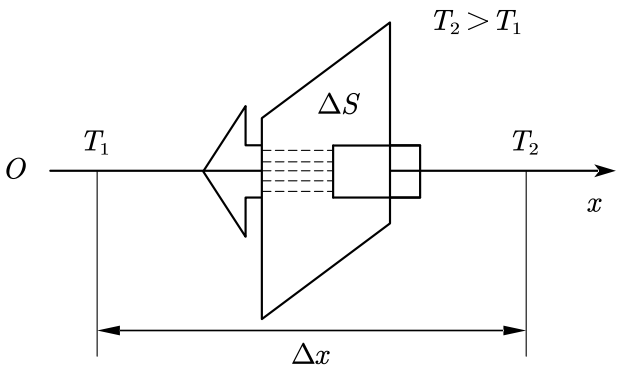


图 1: 热传导现象

如图 1 所示，Ox 轴是气体温度变化最大的方向，在这个方向上气体温度的空间变化率 $\mathrm{d}T/\mathrm{d}x$，叫做**温度梯度**. 设 ΔS 为垂直于 Ox 轴的某指定平面的面积. 实验证明，在单位时间内，从温度较高的一侧，通过这一平面，向温度较低的一侧所传递的热量，与这一平面所在处的温度梯度成正比，同时也与面积 ΔS 成正比，即得**热传导定律**：

$$\frac{\Delta Q}{\Delta t} = -\kappa \frac{\mathrm{d}T}{\mathrm{d}x} \Delta S \tag{1}$$

比例系数 κ 叫做**热导率（thermal conductivity）**. 式中负号表示热量传递的方向是从高温处传到低温处，和温度梯度的方向是相反的热导率的单位是 $\mathrm{W/(m \cdot K)}$.

实验测得，在0°C时，氢的热导率为0.168W/(m·K)，氧力2.42×10^{-1}W/(m·K)，空气为 2.23×10^{-1}W/(m·K)．在 100°C 时，水汽的热导率为 $2.18\times10-1$W/(m·K)．显然，气体的热导率都很小，所以，当气体中不存在对流时，气体可用作很好的绝热材料．

在气体动理论中，对气体热传导现象给出这样的解释：在温度较高的热层中，分子平均动能较大；而在温度较低的冷层中，分子平均动能较小．由于冷热两层分子的互相掺和与相互碰撞，从热层到冷层出现热运动能量的净输运．输运的热运动能量，对单原子气体来说，只是分子的平动动能；而对多原子气体来说，还包含转动和振动的能量在内．

热力学第一定律

- 本词条处于草稿阶段．

预备知识 压力体积图[1563]，理想气体内能[1563]

外部对系统传递的热量等于系统对外做功加上系统的内能增加．

$$\Delta Q = W + \Delta E \tag{1}$$

对于容器中的气体，有

$$W = \int P\,\mathrm{d}V \tag{2}$$

对理想气体，令分子自由度为 i，有

$$E = \frac{i}{2}nRT \tag{3}$$

等压过程

预备知识 理想气体状态方程[1560]

等压过程的特征是系统的压强保持不变，即 p 为常量，$\mathrm{d}p=0$．设想气缸连续地与一系列有微小温度差的恒温热源相接触，同时活塞上所加的外力保持不变．那么接触产生什么效果呢？就是将有微小的热量传给气体，使气体温度稍微升高，气体对活塞的压强也随之较外界所施的压强增加一微量，于是稍微推动活塞对外做功．由于体积的膨胀，压强降低，从而保证气体在内、外压强的量值保持不变的情况下进行膨胀．所以这一准静态过程是一个**等压过程**（**isobaric process**），如图 1 所示．现在我们来计算气体的体积增加 $\mathrm{d}V$ 时所做

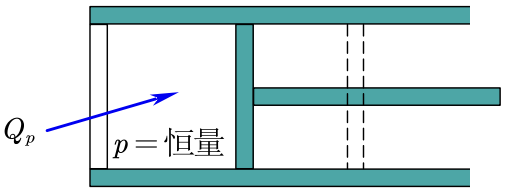


图 1: 气体的等压过程

的功 δW．根据理想气体状态方程，如果气体的体积从 V 增加到 $V+\mathrm{d}V$，温度从 T 增加到 $T+\mathrm{d}T$，那么气体所做的功

$$\delta W=p\mathrm{d}V=\frac{m}{M}R\mathrm{d}T \tag{1}$$

根据热力学第一定律，系统吸收的热量为

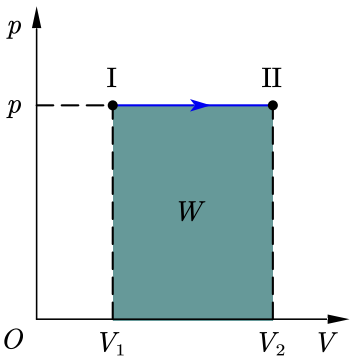


图 2: 等压过程中功的计算

$$\delta Q_p=\mathrm{d}E+\frac{m}{M}R\mathrm{d}T \tag{2}$$

式中，下角标 p 表示压强不变．当气体从状态 $\mathrm{I}(p, V_1, T_1)$ 等压地变为状态 $\mathrm{II}(p, V_2, T_2)$ 时，气体对外做功为

$$W = \int_{V_1}^{V_2} p\mathrm{d}V = p\left(V_2 - V_1\right) \tag{3}$$

或写成

$$W = \int_{T_1}^{T_2} \frac{m}{M} R\mathrm{d}T = \frac{m}{M} R\left(T_2 - T_1\right) \tag{4}$$

所以，整个过程中传递的热量为

$$Q_p = E_2 - E_1 + \frac{m}{M} R\left(T_2 - T_1\right) \tag{5}$$

气体在等压膨胀过程中，所吸收的热量的一部分用来增加内能，另一部分用于气体对外做功；气体在等压压缩过程中，外界对气体做功，同时内能减小，其和等于放出的热量．

我们把 1mol 气体在压强不变的条件下，温度改变 1K 所需要的热量叫做气体的**摩尔定压热容（molar heat capacity at constant pressure）**，用 $C_{p,m}$ 表示，即

$$C_{p,m} = \frac{\delta Q_p}{\frac{m}{M}\mathrm{d}T} \tag{6}$$

根据这个定义可得

$$\delta Q_p = \frac{m}{M} C_{p,m}\mathrm{d}T \tag{7}$$

又因

$$E_2 - E_1 = \frac{m}{M} C_{V,\mathrm{m}}\left(T_2 - T_1\right) \tag{8}$$

我们得到

$$C_{p,m} = C_{V,m} + R \tag{9}$$

式 9 叫做**迈耶（J. R. Meyer）公式**．它的意义是，1 mol 理想气体温度升高 1K 时，在等压过程中比在等体过程中要多吸收 8.31J 的热量．这部分热量去哪了呢？当然是转化为对外所做的膨胀功．由此可见，普适气体常量 R 等于 1 mol 理想气体在等压过程中温度升高 1K 对外所做的功．因 $C_{v,m} = iR/2$，从式 9 可知

$$C_{p,m} = \frac{i}{2} R + R = \frac{i+2}{2} R \tag{10}$$

摩尔定压热容 $C_{p,m}$ 与摩尔定容热容 $C_{V,m}$ 之比，用 γ 表示，叫做 **[摩尔] 热容比（ratio of [molar] heat capacity）**或**绝热指数**，于是

$$\gamma = \frac{C_{p,\mathrm{m}}}{C_{V,\mathrm{m}}} = \frac{i+2}{i} \tag{11}$$

根据式 11 不难算出：对于单原子分子气体，$\gamma = 5/3 \approx 1.67$；双原子刚性分子气体 $\gamma = 1.40$；多原子刚性分子气体 $\gamma \approx 1.33$．它们也都只与气体分子的自由度有关，而与气体温度无关．

无论是定压热容，还是定容热容，它们的共同特点是体现了使物体温度发生变化的难易程度，热容大的物体同样升高 1K，所需要的热量也多，这说明温度不易变化，所以物体的热容是其**热惯性**的量度．

我们来看几个例题加深一下对内容的理解．

例 1　温度计

用作测温的温度计，为了能和被测物体迅速达到热平衡，它的热容必须很小．

例 2　氮气加热

一气缸中贮有氮气，质量为 1.25kg，在标准大气压下缓慢地加热，使温度升高 1K．试求气体膨胀时所做的功 W、气体内能的增量 ΔE 以及气体所吸收的热量 Q_p．（活塞的质量以及它与气缸壁的摩擦均可略去）．

因过程是等压的，所以

$$W = \frac{m}{M} R\Delta T = \frac{1.25}{0.028} \times 8.31 \times 1\mathrm{J} = 371\mathrm{J} \tag{12}$$

而因为氮气的 $i = 5$，所以

$$C_{V,m} = \frac{i}{2} R = 20.8\mathrm{J}/(\mathrm{mol}\cdot\mathrm{K}) \tag{13}$$

于是

$$\Delta E = \frac{m}{M} C_{V,m}\Delta T = \frac{1.25}{0.028} \times 20.8 \times 1\mathrm{J} = 929\mathrm{J} \tag{14}$$

所以，气体在这一过程中所吸收的热量为

$$Q_p = E_2 - E_1 + W = 1300\mathrm{J} \tag{15}$$

等体过程

预备知识 理想气体状态方程[1560]

等体过程的特征是气体的体积保持不变，即 V 为常量，$\mathrm{d}V=0$．设有一气缸，活塞保持固定不动，把气缸连续地与一系列有微小温度差的恒温热源相接触，使气体的温度逐渐上升，压强增大，但是气体的体积保待不变．这样的准静态过程是一个**等体过程（isochoric process）**，如图 1 所示．如图 2 所示，在

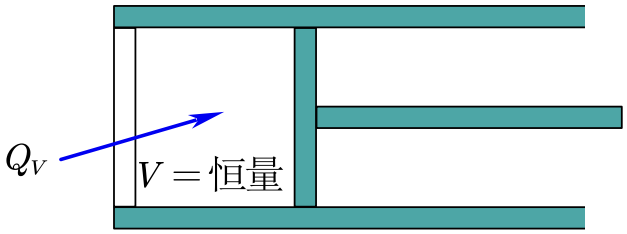


图 1: 气体的等体过程

等体过程中，$\mathrm{d}V=0$，所以 $\delta W=0$．根据热力学第一定律，得

$$\delta Q_V=\mathrm{d}E \tag{1}$$

为了计算向气体传递的热量，我们要用到摩尔热容的概念：同一种气体在不

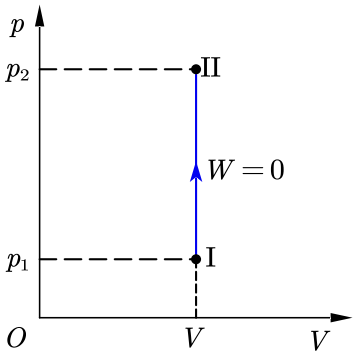


图 2: 等体过程中功的计算

同过程中，有不同的热容．最常用的是等体过程与等压过程中的两种热容．气

体的**摩尔定容热容（molar heal capacity at constant volume）**，是指 1mol 气体在体积不变的条件下，温度改变 1K（或 1°C）所吸收或放出的热量，用 $C_{V,m}$ 表示，其值可由实验测定．这样，质量为 m 的气体在等体过程中，温度改变 $\mathrm{d}T$ 时所需要的热量就是

$$\delta Q_V = \frac{m}{M} C_{V,m} \mathrm{d}T \tag{2}$$

作为 $C_{V,m}$ 的定义式，可将上式改成写成：

$$C_{V,m} = \frac{\delta Q_V}{\frac{m}{M}\mathrm{d}T} \tag{3}$$

把式 2 代入式 1，即得：

$$\mathrm{d}E = \frac{m}{M} C_{V,m} \mathrm{d}T \tag{4}$$

我们这里需要指出的一点是，式 4 是计算过程中理想气体内能变化的通用式子，不仅仅适用于等体过程．我们知道，理想气体的内能只与温度有关，所以一定质量的理想气体在不同的状态变化过程中，如果温度的增量 $\mathrm{d}T$ 相同，那么气体所吸取的热量和所做的功虽然随过程的不同而异，但是气体内能的增量却相同，与所经历的过程无关．那么我们看，现在从等体过程中，我们得到了理想气体温度升高 $\mathrm{d}T$ 时，内能的增量由式 4 给出，那么在任何过程中，都是可以用这个式子来计算理想气体的内能增量的．

我们知道，理想气体的内能为

$$E = \frac{m}{M}\frac{i}{2} RT \tag{5}$$

由此得

$$\mathrm{d}E = \frac{m}{M}\frac{i}{2} R\mathrm{d}T \tag{6}$$

把式 6 与式 4 相比较，我们可以看出：

$$C_{V,m} = \frac{i}{2} R \tag{7}$$

上式说明，理想气体的摩尔定容热容是**一个只与分子的自由度有关的量，它与气体的温度无关**．对于单原子分子气体 $i = 3, C_{V,m} = 12.5\mathrm{J/(mol \cdot K)}$；对于双原子分子气体，不考虑分子的振动，则 $i = 5, C_{V,m} = 20.8\mathrm{J/(mol \cdot K)}$；对于多原子分子气体，不考虑分子的振动，则 $i = 6, C_{V,m} = 24.9\mathrm{J/(mol \cdot K)}$．

等温过程

预备知识 理想气体状态方程[1560]

等温过程的特征是系统的温度保持不变，即 $\mathrm{d}T=0$．由于理想气体的内能只取决于温度，所以在等温过程中，理想气体的内能也保持不变，也就是说 $\mathrm{d}E=0$．

设想一汽缸壁是绝对不导热的，而底部则是绝对导热的（图 1）．现在将气缸的底部和一恒温热源相接触，当活塞上的外界压强无限缓慢地降低时，缸内气体也将随之逐渐膨胀，对外做功气体内能就随之缓慢减少，温度也将随之略微降低．然而，由于气体与恒温热源相接触，当气体温度比热源温度略低时，就有微小的热量传给气体，使气体温度维持原值不变．这一准静态过程就是一个**等温过程（isothermal process）**．

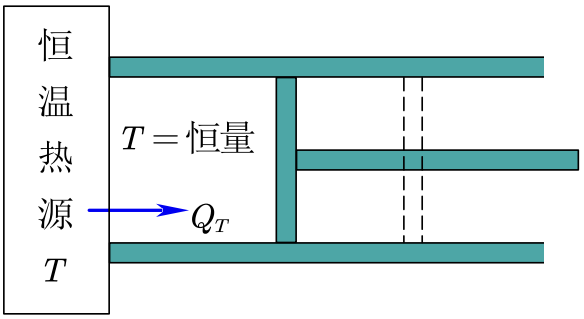


图 1: 气体的等温膨胀

在等温过程中，$p_1V_1=p_2V_2$，系统对外做的功为

$$W=\int_{V_1}^{V_2}p\mathrm{d}V=\int_{V_1}^{V_2}\frac{p_1V_1}{V}\mathrm{d}V=p_1V_1\ln\frac{V_2}{V_1}=p_1V_1\ln\frac{p_1}{p_2}\tag{1}$$

根据理想气体状态方程可得

$$W=\frac{m}{M}RT\ln\frac{V_2}{V_1}=\frac{m}{M}RT\ln\frac{p_1}{p_2}\tag{2}$$

又根据热力学第一定律，系统在等温过程中所吸收的热量应和它所做的功相等，即

$$Q_T=W=\frac{m}{M}RT\ln\frac{V_2}{V_1}=\frac{m}{M}RT\ln\frac{p_1}{p_2}\tag{3}$$

那么等温过程在 $p-V$ 图上长什么样呢？当然是一条双曲线上的一段．这种双曲线就叫做**等温线（isotherm）**．如图 2 所示，$\mathrm{I}\to\mathrm{II}$ 就是一个等温膨胀过程．在等温膨胀过程中，理想气体所吸取的热量全部转化为对外所做的功；反之，在等温压缩时．外界对理想气体所做的功，将全部转化为传给恒温热源的热量．

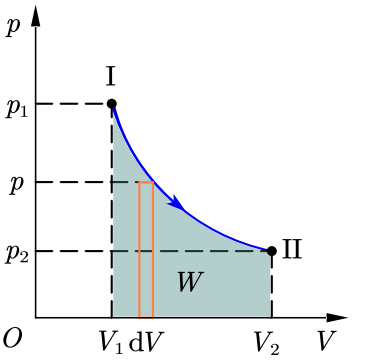


图 2: 等温过程中功的计算

热容量

- 本词条存在未完成的内容．

热容

一个系统在一定条件下的**热容量（heat capacity）**定义为[4]

$$C=\frac{\mathrm{d}Q}{\mathrm{d}T} \tag{1}$$

热容可能跟温度压强等有关．

定义**比热容（specific heat capacity）**为热容除以质量

$$c=\frac{C}{m} \tag{2}$$

[4]这个定义可以类比电容量[1145]

理想气体的等压热热容与等体热容

预备知识 热力学第一定律[1566]

理想气体等体过程[1570] 中的热容叫做**等体热容**

$$C_V = \frac{i}{2}nR \tag{3}$$

而等压过程[1566] 中的热容叫做**等压热容**

$$C_P = \frac{i+2}{2}nR \tag{4}$$

等体热容的推导

（未完成）

$$C_V = \frac{\mathrm{d}Q}{\mathrm{d}T} = \frac{\mathrm{d}E}{\mathrm{d}T} = \frac{i}{2}nR \tag{5}$$

等压热容的推导

（未完成）

$$C_V = \frac{\mathrm{d}Q}{\mathrm{d}T} = \frac{\mathrm{d}E}{\mathrm{d}T} + \frac{\mathrm{d}W}{\mathrm{d}T} = \frac{i}{2}nR + nR = \frac{i+2}{2}nR \tag{6}$$

绝热过程

预备知识 热容量[1573]

在系统状态变化过程中，如果和外界没有热量和粒子交换，这个过程就叫做**绝热过程（adiabatic process）**

$$\Delta Q = 0 \tag{1}$$

根据熵增（式 2 [1576]）的定义，绝热过程是一个等熵过程.

由热力学第一定律式 1 [1566]，

$$W + \Delta E = 0 \tag{2}$$

即系统对外做功和内能增加之和为零.

理想气体的绝热过程中压强体积曲线为

$$PV^\gamma = C \tag{3}$$

其中 C 为常数，γ 为**绝热指数**（i 是气体分子）

$$\gamma = \frac{i+2}{i} \tag{4}$$

绝热指数也可以用等体热容（式 3 [1573]）和等压热容（式 4 [1573]）表示为

$$\gamma = \frac{C_P}{C_V} \tag{5}$$

推导

考虑一个极短的过程，式 2 变为微分形式

$$\mathrm{d}W + \mathrm{d}E = 0 \tag{6}$$

其中（式 1 [1563]）

$$\mathrm{d}W = P\,\mathrm{d}V \tag{7}$$

将理想气体状态方程（式 1 [1560]）两边微分得

$$\mathrm{d}P\,V + P\,\mathrm{d}V = nR\mathrm{d}T \tag{8}$$

将气体的内能公式（式 2 [1563]）两边微分得

$$\mathrm{d}E = \frac{i}{2}nR\,\mathrm{d}T = \frac{i}{2}(V\,\mathrm{d}P + P\,\mathrm{d}V) \tag{9}$$

i 是气体分子自由度. 把式 7 和式 9 代入式 6 得 P 和 V 之间得微分方程

$$\gamma P\,\mathrm{d}V + V\,\mathrm{d}P = 0 \tag{10}$$

其中 γ 为**绝热指数**

$$\gamma = \frac{i+2}{i} \tag{11}$$

绝热

$$PV^\gamma = C \tag{12}$$

熵

预备知识 理想气体状态方程[1560]

在热力学和统计力学中，**熵（entropy）**用于描述系统的无序程度，是一个状态量，通常记为 S．例如若已知理想气体的 P,V,n,T 等状态量，就可以确定它的熵．

宏观定义

熵是一个系统的状态参量，它的增量为

$$\mathrm{d}S = \left.\frac{\Delta Q}{T}\right|_{\text{可逆}} \tag{1}$$

其中，ΔQ 是可逆吸热，也就是说系统应无限地接近平衡态．

根据这个定义，经过任一卡诺循环，熵均回到初始值：

$$\sum \mathrm{d}S_i = 0 \tag{2}$$

然而，要成为真正意义上的「状态参量」的桂冠，式 2 应该对任何循环都成立，而不是只对卡诺循环成立．换言之，我们想要

$$\oint \mathrm{d}S = 0 \tag{3}$$

对于所有循环均成立．这恰恰是正确的．这个证明基于这样的事实：热不能全部变为功而不产生其他效果．因为它意味着 $\eta = 1$，违背了卡诺的结论．

系统在 i 段吸热 ΔQ_i，该热量来自工作于系统现在的温度与热库 T_0 之间的卡诺制冷机．制冷机所需要的功是 ΔW_i．若 $\Delta Q_i < 0$，例如 j 点处，热量由卡诺热机出系统并向热库 T_0 放出该热量．

现在，我们证明，对于任意循环都有 $\oint \mathrm{d}S = 0$．设系统经过如图所示的循环，图中有个温度为 T_0 的辅助热库．在循环上取一小段过程 i．若此过程中的吸热为 ΔQ_i，则设想它是由一个卡诺制冷机供给的，这个制冷机工作于热库 T_0 和系统此时的温度 T_i 之间．工作过程中，制冷机从热库 T_0 吸收的热量为 ΔQ_{0i}，所需的功为 ΔW_i．卡诺制冷机满足

$$\frac{\Delta Q_{0i}}{T_0} = \frac{\Delta Q_i}{T_i} \tag{4}$$

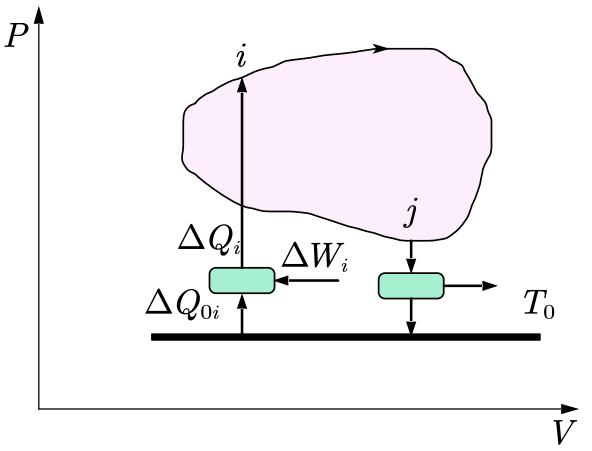


图 1: 系统进行准静态循环

将之对闭合回路求和（在有些段，例如图中的 j 段，ΔQ_i 和 ΔQ_{0i}，可以是负值），得

$$\frac{Q_0}{T_0} = \oint \frac{\mathrm{d}Q}{T} \tag{5}$$

式中，Q_0 是从那个热库吸收热量的总值．如果 Q_0 为零，我们就求得了所要的结果

$$\oint \frac{\mathrm{d}Q}{T} = \oint \mathrm{d}S = 0 \tag{6}$$

我们将看到，的确如此（表示 $\mathrm{d}S$ 时，我们用的是 $\mathrm{d}Q/T$，而不是 $\Delta Q/T$，因为该比值代表微分．）

如果 $Q_0 > 0$，则意味着热库失去了热量．因为系统和所有辅助的卡诺机都已经复原，根据能量守恒，这部分热量一定由卡诺机和我们的系统全部转化为了等量的功，即 $\eta = 1$，而这是不可能的．如果 $Q_0 < 0$，我们可以使整个过程反向进行（因为对于卡诺机和我们的循环，所有步骤都是可逆的），这样的 Q_0 符号就为正，出现如同前面一样的矛盾．

即使没有跟我们一起进行上面的证明，你也一定注意到了，如果要用 $\Delta Q/T = \mathrm{d}S$ 将 ΔQ 和 $\mathrm{d}S$ 联系在一起，那么交换的热量 ΔQ 必须是可逆的．

现在我们还没有完全了解 S 的意义，我们只知道它是个状态参量和它的定义式．下面我们来看一下它到底是啥：先改写一下热力学第一定律，采用 S 进行描述：

$$\mathrm{d}E = \mathrm{d}Q - P\mathrm{d}V \tag{7}$$

因为 E 是状态参量，因此 $\mathrm{d}E$ 与热量如何进入系统无关．我们设热最的交换是可逆的，这样有

$$\mathrm{d}E = T\mathrm{d}S - P\mathrm{d}V \tag{8}$$

从数学上讲，可以认为 E 是 S 和 V 的函数，且

$$T = \frac{\partial E}{\partial S} \tag{9}$$

$$P = -\frac{\partial E}{\partial V} \tag{10}$$

因为我们现在关注的是 S，所以将第一定律写为

$$\mathrm{d}S = \frac{1}{T}\mathrm{d}E + \frac{P}{T}\mathrm{d}V \tag{11}$$

此方程告诉我们 S（对于一定量的气体，如 197mol）是宏观量的函数，如体积 V、能量 U，且

$$\left.\frac{\partial S}{\partial U}\right|_V = \frac{1}{T} \quad \left.\frac{\partial S}{\partial V}\right|_U = \frac{P}{T} \tag{12}$$

要记住这两个方程．

在我们了解熵 S 的含义前，先利用下面这个公式，实际计算几个过程的熵变：

$$\mathrm{d}S = \left.\frac{\mathrm{d}Q}{T}\right|_{\text{可逆}} \tag{13}$$

例 1　质量为 m 的冰在 0°C 时的熔化问题

先看质量为 m 的冰在 0°C 时的熔化问题．我们必须将潜热 $L = 80\mathrm{cal/g}$ 可逆地给系统，也就是用温度为 $(0+\varepsilon)^\circ$C $(\varepsilon \to 0)$ 热库为系统提供热量，使冰水混合物的系统充分吸热，达到水增多一点儿的新平衡态．一直这样做，直到所有的冰在此温度下全部熔化，

$$S_2 - S_1 = \int_1^2 \frac{\mathrm{d}Q}{T} = \frac{mL}{T_{\text{冰}}}\,(\mathrm{cal/K}) \tag{14}$$

式中，$T_{\text{冰}} = 237.16\mathrm{K}$．

在这个简单的例子中，T 固定，等于 $T_{\text{冰}}$，所以它可以提出到积分号的外面，积分的结果为 ml．

接下来我们考虑稍微复杂一些的情况．

例 2　将质量为 m 的水加热（$T_1 \to T_2$）

将质量为 m 的水加热，使其温度由 T_1 升高到 T_2．熵变为

$$S_2 - S_1 = \int_1^2 mc_{水}\frac{\mathrm{d}T}{T} = mc_{水}\ln\frac{T_2}{T_1} \tag{15}$$

再次强调，记住热量应该可逆地进入系统：你不能将温度为 T_1 的水猛地倒入温度为 T_2 的锅中．相反，你将它与一连串的热库接触，每个热库的温度都比前一个热库的温度高一个无限小量，水有足够的时间与每个热库达到平衡，最终使水的温度从 T_1 升高到 T_2．

如果你将水冷却降温，也可以用这个方程，但结果是负值，因为 $T_2 < T_1$．

最后我们计算气体的熵变，其结果非常有意义．温度为 T 的气体等温膨胀，体积从 V_1 增大到 V_2，如图所示．

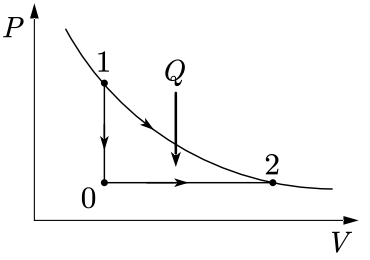


图 2: 1、2 点温度相同

与冰熔化时类似，温度 T 为常量，可以放到积分号的外面：

$$S_2 - S_1 = \int_1^2 \frac{\mathrm{d}Q}{T} = \frac{1}{T}\int_1^2 \mathrm{d}Q = \frac{Q}{T} \tag{16}$$

式中，Q 是气体吸收的总热量．因为等温过程中 E 为定值，Q 等于气体所做的功，因此

$$S_2 - S_1 = \int_1^2 \frac{\mathrm{d}Q}{T} = \int_1^2 \frac{P\mathrm{d}V}{T} = \int_1^2 \frac{nRT\mathrm{d}V}{VT} = nR\ln\frac{V_2}{V_1} \tag{17}$$

如果每个点都有唯一的熵，那么 1、2 两点的熵差与我们怎样从 1 到达 2 无关．因此，如果不沿等温路径的话，我们可以顺着 P 轴下降到达 0 点，该点与 1 点的体积相同，即 $V_0 = V_1$ 且 $P_0 = P_2$．熵变为

$$S_0 - S_1 = nc_v\int_{T_1}^{T_0}\frac{\mathrm{d}T}{T} = n\frac{3R}{2|}\ln\frac{T_0}{T_1} = n\frac{3R}{2}\ln\frac{T_0}{T} \tag{18}$$

$$S_2 - S_0 = nc_p \ln \frac{T_2}{T_0} = n\frac{5R}{2} \ln \frac{T}{T_0} \tag{19}$$

$$\begin{aligned} S_2 - S_1 &= n\frac{3R}{2} \ln \frac{T_0}{T} + n\frac{5R}{2} \ln \frac{T}{T_0} \\ &= n\frac{3R}{2} \ln \frac{T_0}{T_0} + n\frac{2R}{2} \ln \frac{T}{T_0} = nR \ln \frac{T}{T_0} \\ &= nR \ln \frac{V_2}{V_1} \quad \left(\text{用到了} \frac{T}{T_0} = \frac{T_1}{T_0} = \frac{P_1V_1}{P_0V_0} = \frac{P_1}{P_2} = \frac{V_2}{V_1} \right) \end{aligned} \tag{20}$$

这里用到了由图 2 得到的许多等式：$T = T_1 = T_2$，　$V_1 = V_0$ 和 $P_0 = P_2$．因此，求熵变时，我们可以随意选择路径．通常，我们选择最容易的路径．

微观定义

$$S = k_B \ln \Omega \tag{21}$$

准静态过程

平衡态

在不受外界影响的条件下，系统的宏观性质不随时间变化的状态，称为**平衡态（equilibrium state）**，否则就是**非平衡态（nonequilibrium state）**.

具体来说，一定质量的气体在一容器内，如果它与外界没有交换能量，没有外场作用，内部也没有任何形式的能量转化，经过一段时间后，气体各部分终将达到相同的密度、相同的温度、压强等，所有的宏观性质都不随时间而变化，这种状态就是平衡态.

在实际情况中，并不存在完全不受外界影响、而且宏观性质绝对保持不变的系统，所以平衡态只是一个理想的模型，它是在一定条件下对实际情况的概括和抽象.

应当指出，平衡态是指系统的宏观性质不随时间变化；从微观方面看，组成系统的分子的热运动是永不停息的，通过分子的热运动的相互碰撞，其总效果在宏观上表现为不随时间变化的，所以平衡态实际上是**热动平衡状态（thermodynamical equilibrium state）**.

来思考这样一个例子：如果将一根金属棒的两端分别放在沸水和冰水混合物中，经过一段时间后，虽然棒上各处的温度不随时间变化，但这种状态仍不是平衡态，而是**定常态（steady state）**，因为金属棒与外界有能量交换.

准静态过程

当气体的外界条件发生改变时，它的状态就会发生变化. 气体从一个状态不断地变化到另一状态，所经历的是一个状态变化的过程. 过程进展的速度可以很快，也可以很慢，实际过程通常比较复杂. 如果过程进展得十分缓慢，使所经历的一系列中间状态都无限接近平衡状态，这个过程就叫做**准静态过程（quasi-static process）**或**平衡过程（equilibrium process）**. 准静态过程就是实际过程无限缓慢进行时的极限情况.

显然，准静态过程是个理想的过程，它和实际过程毕竟是有差别的，但在许多情况下，可近似地把实际过程当作准静态过程处理，所以准静态过程是个很有用的理想模型. 具体来说，一个系统如果最初处于非平衡态，经过一段时间过渡到一个平衡态，这一过渡时间叫**弛豫时间（relaxation time）**. 在一个实际过程中，如果系统的状态发生一个可以被实验查知的微小变化所需的时间比弛豫时间长得多，那么在任何时刻进行观察时，系统都有充分的时间达到平衡态. 这样的过程就可以当成准静态过程处理.

例如，原来气缸内处于平衡态的气体受到压缩后再达到平衡态所需要的时间，即弛豫时间，大约是 10^{-3}s 或更小，如果在实验中压缩一次所用的时间是 1s，这时间是上述弛豫时间的 10^3 倍，气体这一压缩过程就可以认为是准静态过程. 实际内燃机汽缸气体经历一次压缩时间大约是 10^{-2}s，这个时间也已是上述弛豫时间的 10 倍以上. 从理论上对这种压缩过程作初步研究时，也把它当成准静态过程处理.

卡诺热机

预备知识 等温过程[1572]，绝热过程[1574]，熵[1576]

卡诺热机的工作过程

卡诺循环（Carnot cycle）是一个特别的热力学循环，使用在一个假想的**卡诺热机**上，由法国人尼古拉·卡诺于 1824 年提出，埃米尔·克拉佩龙于 1830 年代至 1840 年代扩充，是为了找出热机的最大的工作效率而分析热机的工作过程.

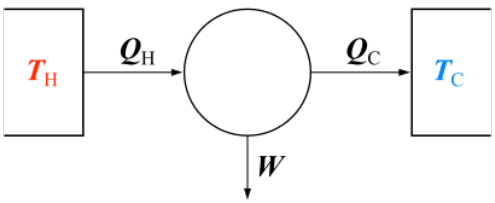


图 1: 热机示意图（来自维基百科）

卡诺循环由两个等温过程，两个绝热过程，下面先通过两个图来直观感受一下卡诺热机的工作过程.

压强-体积图，即 P-V 图，是大家十分熟悉的：温度-熵图，即 T-S 图，则

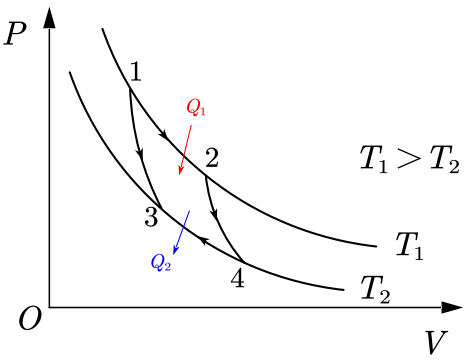


图 2: 卡诺循环的压强-体积图

如下所示：

图 2 $1 \to 2$、图 3 $A \to B$，可逆等温膨胀：此等温的过程中系统从高温热库吸收了热量且全部拿去做功.

图 2 $2 \to 3$、图 3 $B \to C$，等熵（可逆绝热）膨胀：移开热库，系统对环境做功，其能量来自于本身的内能.

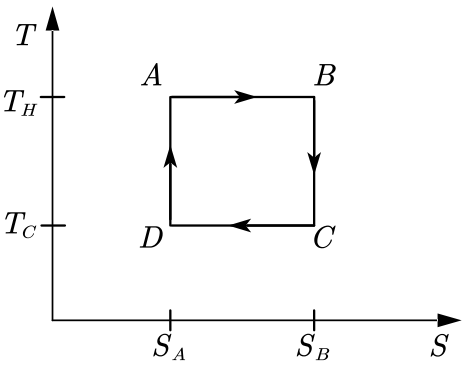


图 3: 卡诺循环的温度-熵图

图 2 $3 \to 4$、图 3 $C \to D$，可逆等温压缩：此等温的过程中系统向低温热库放出了热量．同时环境对系统做正功．

图 2 $4 \to 1$、图 3 $D \to A$，等熵（可逆绝热）压缩：移开低温热库，此绝热的过程系统对环境作负功，系统在此过程后回到原来的状态．

卡诺循环的效率

下面我们来分析一下卡诺循环的效率．气体在等温膨胀过程中，从高温热源吸取热量，

$$Q_1 = \frac{m}{M} RT_1 \ln \frac{V_2}{V_1} \tag{1}$$

气体在等温压缩过程中向低温热源放出热量 Q_2，为便于研究，取绝对值，有

$$Q_2 = \frac{m}{M} RT_2 \ln \frac{V_3}{V_4} \tag{2}$$

应用绝热过程方程 $T_1 V_2^{\gamma-1} = T_2 V_3^{\gamma-1}$ 和 $T_1 V_1^{\gamma-1} = T_2 V_4^{\gamma-1}$ 可得

$$\left(\frac{V_2}{V_1}\right)^{\gamma-1} = \left(\frac{V_3}{V_4}\right)^{\gamma-1} \tag{3}$$

也就是说

$$\frac{V_2}{V_1} = \frac{V_3}{V_4} \tag{4}$$

所以有

$$Q_2 = \frac{m}{M} RT_2 \ln \frac{V_3}{V_4} = \frac{m}{M} RT_2 \ln \frac{V_2}{V_1} \tag{5}$$

取 Q_1 与 Q_2 的比值，可得

$$\frac{Q_1}{T_1}=\frac{Q_2}{T_2} \tag{6}$$

因此卡诺热机的效率为

$$\eta_{\mathrm{c}}=1-\frac{Q_2}{Q_1}=1-\frac{T_2}{T_1} \tag{7}$$

卡诺热机有几条重要性质，我们做个总结.

1. 要完成一次卡诺循环必须有高温和低温两个热源；
2. 卡诺循环的效率只与两个热源的温度有关，高温热源的温度越高，低温热源的温度越低，卡诺循环的效率越大，也就是说节两热源的温度差越大，从高温热源所吸取的热量 Q_1 的“利用价值”越大；
3. 卡诺循环的效率总足小于 1 的.

实际上，卡诺循环还有个逆循环过程，功能应该都可以猜到，也就是制冷机. 通过类似对卡诺循环的分析，可以得到制冷系数

$$w_{\mathrm{C}}=\frac{Q_2}{A}=\frac{Q_2}{Q_1-Q_2}=\frac{T_2}{T_1-T_2} \tag{8}$$

例 1　卡诺制冷机

有一卡诺制冷机，从温度为 $-10^{\circ}\mathrm{C}$ 的冷藏室吸取热量，而向温度为 $20^{\circ}\mathrm{C}$ 的物体放出热量. 设该制冷机所耗功率为 $15\mathrm{kW}$, 问每分钟从冷藏室吸取的热量为多少？

令 $T_1=293\mathrm{K}, T_2=263\mathrm{K}$，则

$$w=\frac{T_2}{T_1-T_2}=\frac{263}{30} \tag{9}$$

每分钟做功为

$$A=15\times10^3\times60\mathrm{J}=9\times10^5\mathrm{J} \tag{10}$$

所以每分钟从冷藏室中吸取的热量为

$$Q_2=wA=7.89\times10^6\mathrm{J} \tag{11}$$

此时，每分钟向温度为 $20^{\circ}\mathrm{C}$ 的物体放出的热量为

$$Q_1=Q_2+A=8.79\times10^6\mathrm{J} \tag{12}$$

分子平均碰壁数

- 本词条存在未完成的内容.

预备知识 理想气体状态方程[1560]

我们来考虑理想气体，即气体分子之间不发生相互作用．当容器中的气体分子平均速度为 $\bar{v}$，分子数密度（单位体积内的分子个数）为 n 时，单位容器面积单位时间受到分子碰撞的平均次数为

$$\frac{1}{4}n\bar{v} \tag{1}$$

这个结论与容器的形状无关.

简单的推导

假设所有分子的速度都是 v，分子数密度为 n（单位体积内的分子数）．假设分子之间不发生碰撞．如果所有的分子都向同一个方向运动，那么单位时间通过面积为 a 的垂直截面的分子数为 nva．如果容器是一个球壳，那么球壳的一半会受到粒子的撞击，单位时间的撞击次数（碰撞率）等于单位时间粒子通过容器最大截面的个数（如图 1），即 $nv(\pi R^2)$．如果有一半的分子向右移动，

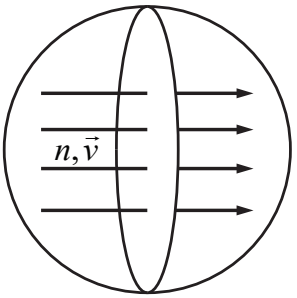


图 1: 分子同向运动的情况

一半向上移动（如图 2），那么每个方向的分子数密度变为原来的一半，总的碰撞率仍为

$$\frac{1}{2}nv\left(\pi R^2\right)+\frac{1}{2}nv\left(\pi R^2\right)=nv\left(\pi R^2\right) \tag{2}$$

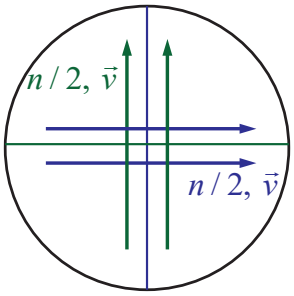


图 2: 分子向两个方向运动的情况

依此类推，如果分子运动的方向被均匀分布在空间的各个方向上，单位时间碰撞数仍然是 $nv(\pi R^2)$. 由于球形容器的表面积为 $4\pi R^2$，所以单位容器壁面积单位时间的碰撞数就是 $nv/4$.

接下来如果把球形容器改成任意形状的容器，由于分子运动在各个方向都是一样，所以结论不变. 另外，一般情况下并不是每个分子都具有相同的速度，所以速度取平均值 v 即可.

积分推导

未完成：建立球坐标系，以及平面

气体分子的速度分布

- 本词条存在未完成的内容.

预备知识 多变量分布函数[728]

当我们描述大量气体分子的运动时，在某时刻可以用一个速度分布函数 $f(v_x, v_y, v_z)$ 描述气体分子的速度分布，或用矢量符号记为 $f(\mathbf{v})$，也可以用球坐标表示为 $f(v, v_\theta, v_\phi)$.

我们以下考虑**各向同性**的速度分部，即 $f(v, v_\theta, v_\phi)$ 与方向 v_θ, v_ϕ 无关，可以简写为 $f(v)$. 另外以下将分子视为质点，不考虑其转动和振动等内部运动.

速度的能均分定理

我们来考虑分子的平均动能以及各个方向的平均动能之间的关系. 由于经典力学中动能为 $E_k = mv^2/2$，我们只需要计算分子速度（以及各个分量速度）平方的平均值.

令 v_i 为速度分量 v_x, v_y, v_z 中的一个，有

$$\left\langle v_i^2 \right\rangle = \int v_i^2 f(v_x, v_y, v_z)\,\mathrm{d}v_x\,\mathrm{d}v_y\,\mathrm{d}v_z \tag{1}$$

由于速度分布各向同性，有

$$\left\langle v_x^2 \right\rangle = \left\langle v_y^2 \right\rangle = \left\langle v_z^2 \right\rangle \tag{2}$$

而速度模长平方的分布为

$$\begin{aligned}\left\langle v^2 \right\rangle &= \int (v_x^2 + v_y^2 + v_z^2) f(v_x, v_y, v_z)\,\mathrm{d}v_x\,\mathrm{d}v_y\,\mathrm{d}v_z \\ &= \sum_{i=x,y,z} \int v_i^2 f(v_x, v_y, v_z)\,\mathrm{d}v_x\,\mathrm{d}v_y\,\mathrm{d}v_z \\ &= \sum_{i=x,y,z} \left\langle v_i^2 \right\rangle = \frac{1}{3}\left\langle v_i^2 \right\rangle\end{aligned} \tag{3}$$

这就说明，分子的平均动能和各个分量之间的平均动能满足

$$\bar{E}_{kx} = \bar{E}_{ky} = \bar{E}_{kz} = \frac{1}{3}\bar{E}_k \tag{4}$$

这相当于把总平均动能**均分**到了三个方向上，所以称为**能均分定理**.

速率分布

若把速度（矢量）的模长 $v = |\mathbf{v}|$ 叫做速率，则类比式 15 [728]，速率的分布函数为

$$F(v) = 4\pi v^2 f(v) \tag{5}$$

令气体分子的数量为 N，则随机一个分子速度绝对值落在 $[v_a, v_b]$ 范围的概率为（类比式 14 [728]）

$$P_{ab} = \int_a^b F(v)\,\mathrm{d}v = \int_a^b 4\pi v^2 f(v)\,\mathrm{d}v \tag{6}$$

所以近似地，该区间中的分子数量为

$$N_{ab} = P_{ab} N \tag{7}$$

（未完成）

麦克斯韦—玻尔兹曼分布

- 本词条处于草稿阶段.

预备知识 随机变量的变换[726]，气体分子的速度分布[1586]

[5]理想气体分子的速率分布由**麦克斯韦—波尔兹曼**分布来描述

$$f(v) = 4\pi \left(\frac{m}{2\pi kT} \right)^{3/2} v^2 \exp\left(-\frac{mv^2}{2kT} \right) \tag{1}$$

这是一个**概率分布函数**[723]，即速度模长在某个区间 $v \in [v_a, v_b]$ 的概率为

$$P_{ab} = \int_{v_a}^{v_b} f(v)\,\mathrm{d}v \tag{2}$$

平均速度为

$$\bar{v} = \sqrt{\frac{8kT}{\pi m}} \tag{3}$$

速度平方平均值为

$$\overline{v^2} = \frac{3kT}{m} \tag{4}$$

概率最大的位置为

$$v_p = \sqrt{\frac{2kT}{m}} \tag{5}$$

动能分布为

$$f(E) = \frac{2}{kT} \sqrt{\frac{E}{\pi kT}} \exp\left(-\frac{E}{kT} \right) \tag{6}$$

[5]参考 [21] 以及维基百科相关页面.

未完成：推导，简单方法参考新概念热学

理想气体分压定律

预备知识 理想气体[1560]

如果一个容器中有多种理想气体，由于我们忽略分子间的相互作用，每种气体分别满足理想气体状态方程式 1 [1560].

$$P_iV = n_iRT \tag{1}$$

其中 n_i 是第 i 种气体的摩尔数，P_i 是该气体对容器壁撞击产生的压强，叫做该气体的**分压**. 由于平衡时所有气体的体积和温度相同，所以每种气体的分压与它的分子个数成正比.

根据“分子撞击对容器壁的压强[1558]”的压强产生原理，当存在多种不同气体时，总压强等于每种气体压强之和.

$$P = \sum_i P_i \tag{2}$$

另外，总分子数等于每种气体的分子数之和

$$n = \sum_i n_i \tag{3}$$

所以将所有气体的式 1 相加，再使用式 2 和式 3，就重新得到了理想气体状态方程

$$PV = nRT \tag{4}$$

可见该方程同样适用于混合气体.

未完成：需要例题

饱和蒸汽压

预备知识 理想气体状态方程[1560]

在一定的温度下，若容器中某种液体和它的蒸汽达到平衡，这时气体的压强就是**饱和蒸汽压**．注意饱和蒸汽压只和液体的种类和温度有关．

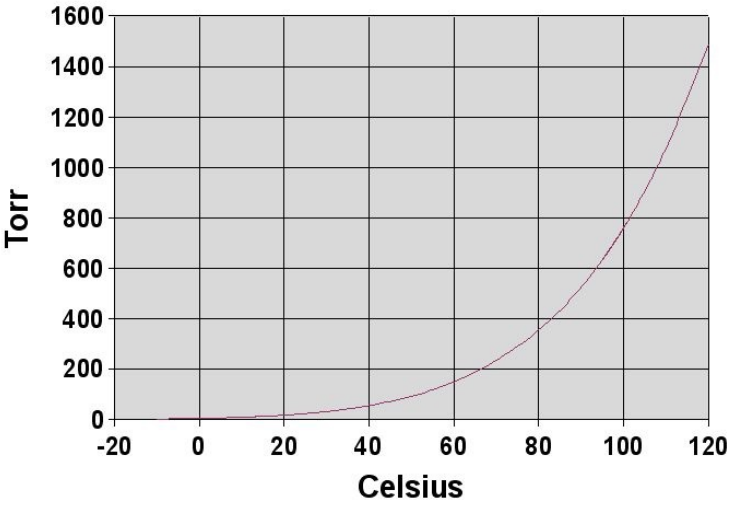


图 1: 水的饱和蒸汽压，横坐标是摄氏温度，纵坐标是毫米汞柱（图片来自维基百科）

我们用一个函数 $P_\alpha(T)$ 描述饱和蒸汽压和温度的关系．例如水蒸气的饱和蒸汽压记为 $P_{H2O}(T)$ （图 1）．若容器的体积可以变化，当容器变大而温度不变时，压强首先会变小，但液体会继续汽化，直到压强恢复到饱和蒸汽压时重新达到平衡．反之，当容器体积变小（温度不变），则压强首先升高，然后气体会逐渐液化直到恢复饱和蒸汽压时重新达到平衡．一般来说，温度升高会导致饱和蒸汽压升高．

为了方便讨论，我们以下假设蒸汽为理想气体[1560]．在常温常压（或更低温低压）下这一般是近似成立的，但在高温高压下假设会失效．将饱和蒸汽压代入理想气体状态方程会发现饱和蒸汽的分子密度只于温度有关

$$\frac{n_\alpha}{V} = \frac{P_\alpha(T)}{RT} \tag{1}$$

习题 1　水蒸气的质量

求 25° C 时一立方米饱和水蒸气的质量．

混合气体

预备知识 理想气体分压定律[1589]

我们接下来讨论容器中除了某种液体及其蒸汽外，还有另一种气体的情况．以水，水蒸气和干燥的空气[6]为例．如果将两种气体都视为理想气体，那么两种气体的总压强可以根据分压定律来计算，即水蒸气单独存在时的压强加上空气单独存在时的压强．另外，假设空气分子不溶于水且不与水分子发生任何反应，则水蒸气的分压仍然是 $P_{H2O}(T)$．各种气体的分子密度仍然可以用式 1 计算．

令干燥空气的分压为 P_{air} 则混合气体总压强 P 为

$$P(T) = P_{H2O}(T) + P_{air}(T) \tag{2}$$

所以若已知总压强和温度，根据式 1 可以计算平衡时两种气体的比例为

$$\frac{n_{H20}}{n_{air}} = \frac{P_{H2O}}{P_{air}} = \frac{P_{H2O}(T)}{P - P_{H2O}(T)} \tag{3}$$

绝对湿度和相对湿度

我们下面来介绍日常生活中经常使用的相对湿度．我们先定义**绝对湿度**为 n'_{H20}/n'_{air} 或者 P'_{H20}/P'_{air}，这里使用一瞥表示实际上的比例而不是平衡时的比例（式 3）．若 P 表示大气压强，那么式 3 就是某温度和汽压下可能出现的最大绝对湿度．因为如果实际湿度稍高于这个值，空气中的水蒸气就会迅速凝结到空气中的尘埃上，形成云或雾，有或者直接凝结到物体表面形成水滴或霜．

另一方面，空气中绝对湿度常常低于式 3 的值，即水和水蒸气处于不平衡的状态．这是因为地表的水（江湖海，土壤等）蒸发的速度有限，且大气并不是长期处于恒温恒压状态．

我们把**相对湿度**定义为水蒸气的分子实际密度（或分压）除以当前温度下饱和水蒸气的分子密度（或分压）．即

$$\phi = \frac{n'_{H2O}}{n_{H2O}} = \frac{P'_{H20}}{P_{H2O}} \tag{4}$$

[6]虽然空气由不同种类的分子构成，但容易证明在以下讨论中，是否把空气看成一种分子结论都一样．

沸腾

在式 3 中，如果我们实时调节容器的体积，使容器中的气体始终保持某个固定的压强 P，那么当温度不断升高时，绝对湿度也会不断升高．当 $P_{H2O}(T)$ 趋近 P 时，容器里面就几乎只有水蒸气了．此时如果温度继续升高使 $P_{H2O}(T)$ 等于或略大于 P，系统就无法保持平衡，水会不断汽化，使容器不断膨胀，直到所有的水都变为水蒸气为止．我们把这个过程叫做**沸腾（boiling）**，并定义满足

$$P_{H2O}(T_b) = P \tag{5}$$

的温度 T_b 为**沸点（boiling point）**．沸点与液体的种类以及压强有关．

大气密度和压强

预备知识 一阶线性微分方程[212]，理想气体分压定律[1589]，积分方程

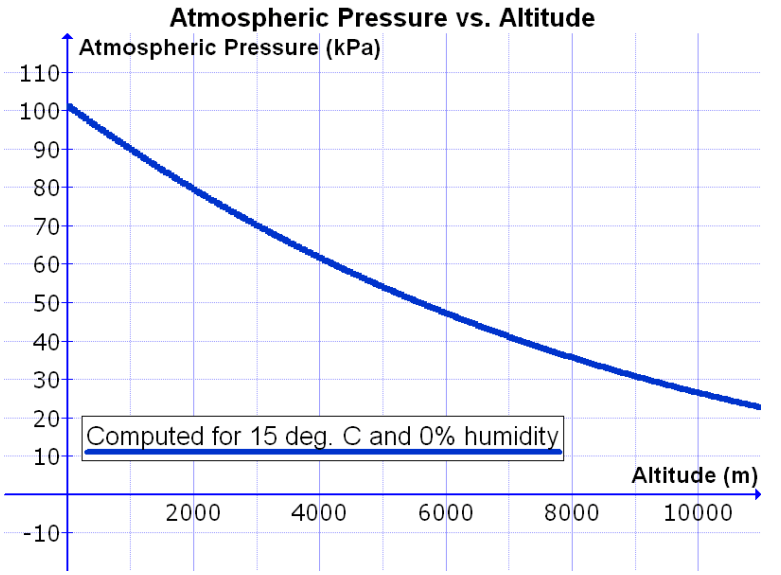


图 1: 根据式 8 以及分压定律得到的大气压强随高度变化图，假设大气中没有水蒸气且温度恒定（来自维基百科）

[7]以下介绍一个理想模型．假设大气是理想气体，密度随高度变化为 $\rho(z)$．

[7]参考 Wikipedia 相关页面以及另一个页面．

所以高度 z 处压强为

$$P(z)=\int_z^\infty \rho(z')g\,\mathrm{d}z' \tag{1}$$

其中由于大气厚度远小于地球半径，我们取 g 为常数．根据理想气体状态方程[1560]，

$$PV=nRT \tag{2}$$

先假设大气只是由一种分子构成，摩尔质量为 μ，即 $m=n\mu$，代入有

$$P=\frac{m}{\mu V}RT=\frac{R}{\mu}\rho T \tag{3}$$

其中 P,T,ρ 都是高度的函数．代入式 1 得关于 $\rho(z)$ 的积分方程

$$\frac{R}{\mu}\rho(z)T(z)=\int_z^\infty \rho(z')g\,\mathrm{d}z' \tag{4}$$

通常来说海拔越高的地方气温越低，如果 $T(z)$ 是已知的，就可以解出 $\rho(z)$．方程两边对 z 求导，整理得

$$\rho'(z)+\frac{1}{T(z)}\left[T'(z)+\frac{\mu g}{R}\right]\rho(z)=0 \tag{5}$$

这是一个一阶线性微分方程，可以直接用公式求解[212]．把解出的 $\rho(z)$ 代入式 3 即可求出对应的压强 $P(z)$．

作为一种简单情况，假设温度不随高度变化，那么方程变为常系数的

$$\rho'(z)+\frac{\mu g}{RT}\rho(z)=0 \tag{6}$$

容易解得

$$\rho(z)=\rho_0\exp\left(-\frac{\mu g}{RT}z\right) \tag{7}$$

或者

$$P(z)=P_0\exp\left(-\frac{\mu g}{RT}z\right) \tag{8}$$

其中 ρ_0,P_0 是某个高度 z_0 处的大气密度和气压．这说明恒温条件下气压随海拔升高呈指数下降，且温度越低下降越快．

当大气中有多种气体时，可以对每种气体分别求解，把 P_0 替换为改气体在 z_0 处的分压[1589]．总密度就是每种气体的密度之和．大气中的水蒸气同样也可能随着高度变化．

黑体辐射定律

- 本词条处于草稿阶段.
- 本词条需要更多参考文献.

单位体积单位频率的能量

$$S_\nu(\nu)=\frac{8\pi h}{c^3}\frac{\nu^3}{\mathrm{e}^{h\nu/(k_BT)}-1} \tag{1}$$

如果要计算波长的分布，根据随机变量的变换[726]，由 $|S_\lambda(\lambda)\,\mathrm{d}\lambda|=|S_\nu(\nu)\,\mathrm{d}\nu|$ 得

$$S_\lambda(\lambda)=\frac{c}{\lambda^2}S_\nu\left(\frac{c}{\lambda}\right)=\frac{8\pi ch}{\lambda^5}\frac{1}{\mathrm{e}^{hc/(k_BT\lambda)}-1} \tag{2}$$

单位面积单位频率单位立体角[286] 的功率

$$B(\nu)=\frac{c}{4\pi}S_\nu(\nu)=\frac{2h}{c^2}\frac{\nu^3}{\mathrm{e}^{h\nu/(k_BT)}-1} \tag{3}$$

在黑体内部，辐射是各向同性的，但在黑体表面，对于给定的一个平面微元，$B(\nu)$ 是垂直于平面的值，与法向量夹角为 θ 的方向的值为 $B(\nu)\cos\theta$.

每个能级（$n\omega\hbar$）的平均能量

$$E(\nu)=\frac{h\nu}{\mathrm{e}^{h\nu/(k_BT)}-1} \tag{4}$$

态密度

$$\rho(\nu)=\frac{8\pi}{c^3}\nu^2 \tag{5}$$

推导

预备知识 盒中的电磁波盒中的电磁波[1238]

维恩位移定律

预备知识　黑体辐射定律[1594]

维恩位移定律（Wien's displacement law）描述了黑体辐射光谱的峰值波长与温度之间的关系

$$\lambda_{max} = \frac{b}{T} \tag{1}$$

其中 b 是**维恩位移常数**

$$b = \frac{hc}{xk_BT} = 2.897771955 \times 10^{-3}\ \mathrm{mK} \tag{2}$$

其中 $x = 4.9651142\ldots$ 是以下超越方程的解

$$(x-5)\mathrm{e}^x + 5 = 0 \tag{3}$$

推导

预备知识　导数与函数极值[170]

我们可以从黑体辐射定律（式 2 [1594]）推导出来

$$I \propto \frac{1}{\lambda^5}\frac{1}{\mathrm{e}^{hc/(\lambda k_BT)}-1} \tag{4}$$

由于该分布处处可导且最大值就是唯一的极大值，根据“导数与函数极值[170]”解方程，解出下式即可.

$$\frac{\mathrm{d}I}{\mathrm{d}\lambda} = 0 \tag{5}$$

斯特藩—玻尔兹曼定律

斯特藩—玻尔兹曼定律（Stefan-Boltzmann law）指的是，黑体单位面积的辐射功率与温度的 4 次方成正比

$$\frac{\mathrm{d}P}{\mathrm{d}S} = \sigma T^4 \tag{1}$$

其中 σ 是**斯特藩—玻尔兹曼常数（Stefan-Boltzmann constant）**

$$\sigma = \frac{2\pi^5 k_B^4}{15c^2h^3} = 5.670374419\cdots \times 10^{-8}\ \mathrm{Wm^{-2}K^{-4}} \tag{2}$$

其中 k_B 是玻尔兹曼常数，c 是真空中的光速，h 是普朗克常数，详见“物理学常数[1973]”.

从黑体辐射定律推导

预备知识 黑体辐射定律[1594]

我们考虑黑体表面一个小平面，以它的法向量（指向真空）为极轴建立球坐标系．单位面积单位频率单位立体角的功率见（式 3 [1594]）

$$I(\nu)=\frac{2h}{c^2}\frac{\nu^3}{\mathrm{e}^{h\nu/(k_BT)}-1} \tag{3}$$

我们只需要对频率和上半球面的立体角做积分即可．但是注意在不垂直黑体表面的方向，功率需要乘以 $\cos\theta$（表面积在该方向的投影是 $\cos\theta$）．

$$\frac{\mathrm{d}P}{\mathrm{d}S}=\int_0^\infty I(\nu)\,\mathrm{d}\nu\int_0^{2\pi}\mathrm{d}\phi\int_0^{\pi/2}\cos\theta\sin\theta\,\mathrm{d}\theta=\pi\int_0^\infty I(\nu)\,\mathrm{d}\nu \tag{4}$$

使用换元积分，令 $u=h\nu/(k_BT)$ 得

$$\frac{\mathrm{d}P}{\mathrm{d}S}=\frac{2\pi h}{c^2}\left(\frac{k_BT}{h}\right)^4\int_0^\infty\frac{u^3}{\mathrm{e}^u-1}\,\mathrm{d}u \tag{5}$$

其中积分的结果可以用黎曼 ζ 函数表示为 $6\zeta(4)=\pi^4/15$．代入即可得式 1 ．

第二章

统计力学

相空间

- 本词条处于草稿阶段.

单个粒子的相空间

单个粒子（看做质点）的状态可以由 3 个位置坐标 (x,y,z) 和三个动量坐标 (p_x,p_y,p_z) 来描述，为了便于拓展到一般情况，我们用 $q_1\equiv x$，$q_2\equiv y$，$q_3\equiv z$ 和 $p_1\equiv p_x$，$p_2\equiv p_y$，$p_3\equiv p_z$ 表示．想象一个由 3 个 q_i 坐标和 3 个 p_i 坐标组成的 6 维空间，体积元定义为

$$\mathrm{d}\Omega_1=\frac{1}{h^3}\,\mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}q_3\cdot\mathrm{d}p_1\,\mathrm{d}p_2\,\mathrm{d}p_3 \tag{1}$$

为了方便表示，简写为 $d\Omega_1=\mathrm{d}^3q\cdot\mathrm{d}^3p$.

$$\Omega_1=\frac{1}{h^3}\int_{\Omega_1}\mathrm{d}^3q\,\mathrm{d}^3p \tag{2}$$

积分对所有可能的 q_i 和 p_i 进行．例如粒子若被限制在一个长宽高分别为 L_x,L_y,L_z 的盒子里，而动量没有限制，那么上面积分变为

$$\Omega_1=\frac{1}{h^3}\int_0^{L_x}\int_0^{L_y}\int_0^{L_z}\int\mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}q_3\,\mathrm{d}^3p=\frac{L_xL_yL_z}{h^3}\int\mathrm{d}^3p \tag{3}$$

经典力学中，对于 N 个粒子的系统，可以用 $3N$ 个位置坐标和 $3N$ 个动量坐标来完全描述系统的状态．则相空间为 $6N$ 维空间.

理想气体的状态密度（相空间）

- 本词条处于草稿阶段.

预备知识 N 维球体的体积[340]

相空间中能量小于 E 的状态数

$$\Omega_0=\frac{V^N}{N!h^{3N}}\frac{(2\pi mE)^{3N/2}}{\Gamma(3N/2+1)} \tag{1}$$

关于能量的状态密度函数为

$$g(E)=\frac{V^N}{N!h^{3N}}\frac{(2\pi m)^{3N/2}}{\Gamma(3N/2)}E^{3N/2-1} \tag{2}$$

关于动量绝对值的状态密度函数为

$$g(p)=\frac{V^N}{N!h^{3N}}\frac{2\pi^{3N/2}}{\Gamma(3N/2)}p^{3N-1} \tag{3}$$

推导

在 N 个独立的，可区分粒子的相空间中，能量小于 E 的状态数（体积除以 h^{3N}，无量纲）为

$$\Omega_0=\frac{1}{N!h^{3N}}\int\limits_{\sum p^2\leqslant 2mE}\mathrm{d}^{3N}q\,\mathrm{d}^{3N}p=\frac{V^N}{N!h^{3N}}\int\limits_{\sum p^2\leqslant 2mE}\mathrm{d}^{3N}p \tag{4}$$

其中积分 $\int_{\cdots}\mathrm{d}^{3N}p$ 可以看做 $n=3N$ 维球体的体积，半径为 $R=\sqrt{2mE}$.

n 维空间中球体的体积[340] 为

$$V_n=\frac{\pi^{n/2}}{\Gamma(n/2+1)}R^n \tag{5}$$

代入 $n=3N$ 和 $R=\sqrt{2mE}$，得

$$\int\limits_{\sum p^2\leqslant 2mE}\mathrm{d}^{3N}p=\frac{(2\pi mE)^{3N/2}}{\Gamma(3N/2+1)} \tag{6}$$

代入式 4 得式 1．再对 E 求导得式 2.

理想气体单粒子能级密度

- 本词条处于草稿阶段.

相空间法

$$\Omega_0=\frac{1}{h^3}\int\limits_{\sum p^2\leqslant 2mE}\mathrm{d}^3q\,\mathrm{d}^3p=\frac{V}{h^3}\frac{4}{3}\pi p^3=\frac{V}{h^3}\frac{4}{3}\pi(2m\varepsilon)^{3/2} \tag{1}$$

$$a(\varepsilon)=\frac{\mathrm{d}\Omega_0}{\mathrm{d}\varepsilon}=\frac{2\pi V(2m)^{3/2}}{h^3}\varepsilon^{1/2} \tag{2}$$

量子力学法

由盒中粒子得，单粒子的能级为

$$\varepsilon=\frac{\hbar^2}{2m}\left[\left(\frac{\pi n_x}{L_x}\right)^2+\left(\frac{\pi n_y}{L_y}\right)^2+\left(\frac{\pi n_z}{L_z}\right)^2\right]=\frac{\hbar^2}{2m}(k_x^2+k_y^2+k_z^2) \tag{3}$$

在 k 空间中，每个能级所占的体积为

$$V_1=\frac{\pi^3}{L_xL_yL_z}=\frac{\pi^3}{V} \tag{4}$$

K 空间中，能量小于 E 的量子态数为（注意 n 为正值，所以只求一个卦限的体积，要乘 $1/8$）

$$\Omega_0=\frac{1}{8}\cdot\frac{\hbar^2}{2m}\frac{4}{3}\pi k^3\Big/\frac{\pi^2}{V}=\frac{V}{h^3}\frac{4}{3}\pi(2m\varepsilon)^{3/2} \tag{5}$$

这就是式 1，对能量求导得式 2 .

理想气体（微正则系综法）

- 本词条处于草稿阶段.

预备知识 理想气体的状态密度（相空间）[1598]

当理想气体处于某个能量 E 时，相空间只是位置相空间（$3N$ 维）以及动量相空间中的一个球表面（$3N-1$ 维）. 这个 $6N-1$ 维相空间中的体积除以 h^{3N} （式 3 [1598]）就可以作为熵公式中的 Ω，得

$$S(E,V,N)=k\ln\Omega=Nk\left(\ln\frac{V}{N\lambda^3}+\frac{5}{2}\right) \tag{1}$$

该式被称为 **Sackur-Tetrode 公式**. 其中

$$\lambda=\frac{h}{\sqrt{4\pi mE/(3N)}}=\frac{h}{\sqrt{2\pi mkT}} \tag{2}$$

式 1 也用到了 Stirling 近似 $\ln N!=N\ln N-N$.

根据熵的微分关系

$$\mathrm{d}S = \frac{1}{T}\,\mathrm{d}E + \frac{P}{T}\,\mathrm{d}V - \frac{\mu}{T}\,\mathrm{d}N \tag{3}$$

所以对 S 求三个偏导得

$$T = \frac{2E}{3Nk} \tag{4}$$

$$P = NkT/V \tag{5}$$

$$\mu = kT\ln\frac{N\lambda^3}{V} \tag{6}$$

正则系宗法

- 本词条处于草稿阶段.

亥姆霍兹自由能

由于最可能的能级的可能性为“1”

$$1 = \Omega \mathrm{e}^{-E/\tau}/Q = \mathrm{e}^{S/k}\mathrm{e}^{-E/\tau}/Q = \mathrm{e}^{-F/\tau}/Q \tag{1}$$

所以

$$F = -kT\ln Q \tag{2}$$

系统总能量

$$E = \frac{\sum_i E_i \mathrm{e}^{-E_i/kT}}{Q} \tag{3}$$

考虑到 $E_i\mathrm{e}^{-E_i/kT} = \tau^2$

$$E = \tau^2\frac{\partial}{\partial\tau}\ln Q \tag{4}$$

理想气体（正则系宗法）

- 本词条处于草稿阶段.

可区分粒子和不可区分粒子

对于可区分粒子，从粒子的角度求和，配分函数为（dis = distinguishable）

$$\begin{aligned} Q_{dis} &= \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\cdots\sum_{i_N=0}^{\infty} \mathrm{e}^{-\beta(\varepsilon_{i_1}+\varepsilon_{i_2}\cdots)} = \sum_{i_1=0}^{\infty}\mathrm{e}^{-\beta\varepsilon_{i_1}}\sum_{i_2=0}^{\infty}\mathrm{e}^{-\beta\varepsilon_{i_2}}\cdots\sum_{i_N=0}^{\infty}\mathrm{e}^{-\beta\varepsilon_{i_N}} \\ &= \left(\sum_{i=0}^{\infty}\mathrm{e}^{-\beta\varepsilon_i}\right)^N = Q_1^N \end{aligned} \tag{1}$$

从能级的角度求和

$$Q_{dis} = \sum_{\{n_i\}}\frac{N!}{n_0!n_1!\ldots}\exp\left(-\beta\sum_{i=0}^{\infty}n_i\varepsilon_i\right) = \sum_{\{n_i\}}\frac{N!}{n_0!n_1!\ldots}\mathrm{e}^{-n_1\varepsilon_1\beta}\mathrm{e}^{-n_2\varepsilon_2\beta}\ldots \tag{2}$$

由式 1 和式 2 物理意义可知，二者相等．再从能级的角度考虑，若粒子不可区分（由于这个配分函数是最常用的，所以不写角标）

$$Q = \sum_{\{n_i\}}\mathrm{e}^{-n_1\varepsilon_1\beta}\mathrm{e}^{-n_2\varepsilon_2\beta}\ldots \tag{3}$$

比较式 2，求和的每项少了一个因子 $N!/(n_0!n_1!\ldots)$

理想气体条件：能级占有率极低，几乎没有两个粒子在同一个能级上，所以大部分 $n_i = 0$，$0! = 1$．个别 $n_i = 1$，$1! = 1$．可以近似认为

$$\frac{N!}{n_0!n_1!\ldots} \approx N! \tag{4}$$

所以

$$Q = \frac{1}{N!}Q_{dis} = \frac{1}{N!}Q_1^N \tag{5}$$

那如何求 Q_1 呢？

对单粒子相空间积分

注意每个量子态对应的相空间体积为 h 的空间维数次方．

$$Q_1 = \frac{1}{h^3}\int \mathrm{e}^{-\frac{p^2}{2m\cdot kT}}\,\mathrm{d}^3p\,\mathrm{d}^3x = \frac{V}{h^3}\int \mathrm{e}^{-\frac{p^2}{2m\cdot kT}}\,\mathrm{d}^3p = \frac{V}{\lambda^3} \tag{6}$$

其中 λ 叫做热力学波长，正比与粒子热运动的德布罗意波

$$\lambda = \frac{h}{\sqrt{2\pi mkT}} \tag{7}$$

对单粒子能级密度积分

用单粒子能级密度 $a(\varepsilon)$ 对玻尔兹曼因子积分.

$$a(\varepsilon)=\frac{2\pi V(2m)^{3/2}}{h^3}\varepsilon^{1/2} \tag{8}$$

$$Q_1=\sum_{i=0}^{\infty}\mathrm{e}^{-\beta\varepsilon_i}=\int_0^{\infty}a(\varepsilon)\mathrm{e}^{-\beta\varepsilon}\,\mathrm{d}\varepsilon=\frac{2\pi V(2m)^{3/2}}{h^3}\int_0^{\infty}\varepsilon^{1/2}\mathrm{e}^{-\beta\varepsilon}\,\mathrm{d}\varepsilon \tag{9}$$

对积分换元，令 $x=\beta\varepsilon$，

$$\int_0^{\infty}\varepsilon^{1/2}\mathrm{e}^{-\beta\varepsilon}\,\mathrm{d}\varepsilon=(kT)^{3/2}\int_0^{\infty}x^{1/2}\mathrm{e}^{-x}\,\mathrm{d}x=\Gamma(3/2)(kT)^{3/2}=\frac{\sqrt{\pi}}{2}(kT)^{3/2} \tag{10}$$

$$Q_1=\sum_{i=0}^{\infty}\mathrm{e}^{-\beta\varepsilon_i}=\int_0^{\infty}a(\varepsilon)\mathrm{e}^{-\beta\varepsilon}\,\mathrm{d}\varepsilon=\frac{2\pi V(2m)^{3/2}}{h^3}\frac{\sqrt{\pi}}{2}(kT)^{3/2}=\frac{V}{\lambda^3} \tag{11}$$

对系统的能级密度积分

现在我们试图直接求 Q，系统的总能级密度为

$$g(E)=\frac{\mathrm{d}\Omega_0}{\mathrm{d}E}=\frac{V^N}{N!h^3}\frac{(2\pi m)^{3N/2}}{(3N/2-1)!}\mathrm{e}^{3N/2-1} \tag{12}$$

$$Q=\int_0^{\infty}g(E)\mathrm{e}^{-E\beta}\,\mathrm{d}E=\frac{V^N(2\pi m)^{3N/2}}{N!h^3(3N/2-1)!}\int_0^{\infty}\mathrm{e}^{3N/2-1}\mathrm{e}^{-\beta E}\,\mathrm{d}E \tag{13}$$

令 $x=\beta E$ 对积分换元，

$$\begin{aligned}\int_0^{\infty}\mathrm{e}^{3N/2-1}\mathrm{e}^{-\beta E}\,\mathrm{d}E&=(kT)^{3N/2}\int_0^{\infty}x^{3N/2-1}\mathrm{e}^{-x}\,\mathrm{d}x\\&=(kT)^{3N/2}\Gamma(3N/2)\\&=(kT)^{3N/2}(3N/2-1)!\end{aligned} \tag{14}$$

代入上式得

$$\begin{aligned}Q&=\frac{V^N(2\pi m)^{3N/2}}{N!h^3(3N/2-1)!}(kT)^{3N/2}(3N/2-1)!\\&=\frac{V^N(2\pi mkT)^{3N/2}}{N!h^3}=\frac{1}{N!}\left(\frac{V}{\lambda^3}\right)^N\\&=\frac{1}{N!}Q_1^N\end{aligned} \tag{15}$$

与之前的结果都一样.

热力学性质

得到系统的配分函数 Q 以后，可由用亥姆霍兹自由能得到热力学的性质

$$F = -kT\ln Q = -kT(N\ln Q_1 - N\ln N + N) \tag{16}$$

$$S = Nk\left(\ln\frac{V}{N\lambda^3} + \frac{5}{2}\right) \tag{17}$$

$$P = -\left(\frac{\partial F}{\partial V}\right)_{T,N} = \frac{NkT}{V} \qquad \text{（理想气体状态方程）} \tag{18}$$

$$\mu = kT\ln\frac{N\lambda^3}{V} = kT\ln\frac{N}{Q_1} \tag{19}$$

在巨正则系综里，定义逸度为 $z = \mathrm{e}^{\mu/(kT)}$，则 $N = zV/\lambda^3 = zQ_1$.

分布函数

若有 N 个粒子组成理想气体，每个能级平均有多少粒子（由于理想气体的条件是能级占有率 $\langle n_i\rangle \ll 1$，但仍然会有分布曲线）

对任何一个粒子来说，出现在 ε_i 能级（非简并）的概率是 $\mathrm{e}^{-\beta\varepsilon_i}/Q_1 = \lambda^3\mathrm{e}^{-\beta\varepsilon_i}/V$. 那么 N 个没有相互作用的粒子在该能级的平均粒子数就为

$$\langle n_i\rangle = \frac{N\lambda^3}{V}\mathrm{e}^{-\beta\varepsilon_i} \tag{20}$$

理想气体的化学能 $\mu = kT\ln N\lambda^3/V$，即 $\mathrm{e}^{\mu/kT} = N\lambda^3/V$. 代入上式，得

$$\langle n_i\rangle = \mathrm{e}^{\beta\mu}\mathrm{e}^{-\beta\varepsilon_i} = \mathrm{e}^{(\mu-\varepsilon_i)/(kT)} \tag{21}$$

这就是麦克斯韦—玻尔兹曼分布.

理想气体（巨正则系综法）

- 本词条处于草稿阶段.

理想气体的巨配分函数

$$\Xi = \exp(zQ_1) = \exp\left(\frac{zV}{\lambda^3}\right) \tag{1}$$

推导

$$\begin{aligned}\Xi &= \sum_{N=0}^{\infty}\sum_{i=1}^{\infty} \mathrm{e}^{(N\mu - E_i)\beta} = \sum_{N=0}^{\infty} z^N Q = \sum_{N=0}^{\infty} z^N \frac{1}{N!} Q_1^N \\ &= \sum_{N=0}^{\infty} \frac{1}{N!}(zQ_1)^N = \exp(zQ_1) = \exp\left(\frac{zV}{\lambda^3}\right)\end{aligned} \tag{2}$$

其中用到了指数函数的泰勒展开（式 6 [181]）.

状态方程推导

首先求出理想气体的巨势

$$\Phi = -kT \ln \Xi = -kT \frac{zV}{\lambda^3} \tag{3}$$

由巨正则系综法[1608]

$$\mathrm{d}\Phi = -P\,\mathrm{d}V - S\,\mathrm{d}T - N\,\mathrm{d}\mu \tag{4}$$

$$P = -\left(\frac{\partial \Phi}{\partial V}\right)_{T,\mu} = kT\frac{z}{\lambda^3} \tag{5}$$

注意 z 是 μ 和 T 的函数（$z = \mathrm{e}^{\mu/(kT)}$），$\lambda$ 是 T 的函数，所以 z 和 λ 在该偏微分中都看做常数.

$$N = -\left(\frac{\partial \Phi}{\partial \mu}\right)_{V,T} = kT\frac{V}{\lambda^3}\left(\frac{\partial z}{\partial \mu}\right)_{V,T} = \frac{Vz}{\lambda^3} \qquad (= zQ_1) \tag{6}$$

若用上面两式消去 z/λ^3 因子，得到理想气体状态方程 $PV = NkT$.

另外，想象在巨正则系综的物理情景中，变化 T 和 μ，从而使式 6 中的粒子数保持不变，则 N 不变时可以看成 T 的函数（而这个函数应该与正则系宗所得到的一样）. 由式 6 得

$$\mu = kT \ln \frac{N\lambda^3}{V} \tag{7}$$

再测试一下状态方程 $PV=-\Phi$，得到 $PV=kTzV/\lambda^3$，这与上面的压强公式（编号）重复，没有新的信息．若把粒子数公式 $N=Vz/\lambda^3$（编号）代入理想气体的巨配分函数 $\Xi=\exp(zV/\lambda^3)$（编号）以及巨势 $\Phi=-kTzV/\lambda^3$（编号），得到两个个相当简洁的表达式，可以方便记忆

$$\Xi=\exp(N) \tag{8}$$

$$\Phi=-NkT \tag{9}$$

理想气体的熵为

$$\begin{aligned}S&=-\left(\frac{\partial\Phi}{\partial T}\right)_{V,\mu}=Vk\frac{T}{\lambda^3}\frac{\partial z}{\partial T}+kTz\frac{\partial}{\partial T}\left(\frac{T}{\lambda^3}\right)\\&=Vk\frac{T}{\lambda^3}\left(-\frac{\mu z}{kT^2}\right)+kTz\frac{\partial}{\partial T}\left(\frac{(2\pi mk)^{3/2}T^{5/2}}{h^3}\right)\\&=-\frac{\mu zV}{T\lambda^3}+kTz\frac{5}{2}\frac{(2\pi mkT)^{3/2}}{h^3}=-\frac{\mu zV}{T\lambda^3}+\frac{5}{2}\frac{kTz}{\lambda^3}\\&=Nk\left(\frac{5}{2}-\frac{\mu}{kT}\right)\end{aligned} \tag{10}$$

这里得出的熵是 μ 和 T 的函数（从巨正则系综的物理情景来看，得出的所有结果都应该是预先设定的参数 μ 和 T 的函数）．

为了和巨正则系综比较，把式 7 代入式 10，即把粒子数人为保持不变，一切看成温度的函数．果然得到了理想气体的熵（Sackur-Tetrode 公式）

$$S=Nk\left(\ln\frac{V}{N\lambda^3}+\frac{5}{2}\right) \tag{11}$$

理解

巨正则系综法的物理情景是：让系统（体积 V ）与粒子源（化学势 μ ）和热源（温度 T ）保持平热平衡，由 μ 和 T 决定粒子数 N，压强 P，能量 E 等等．这与微正则系综或正则系宗的物理情景不一样．但是得到的结论却是一样的．

等间隔能级系统（正则系宗）

- 本词条处于草稿阶段.

预备知识 玻尔兹曼因子配分函数

结论

若一个系统的能量只能取一系列离散的值（能级），但相邻能级间距恰好为 ε，那么该系统在温度 τ 达到热平衡时，平均能量为

$$\langle\varepsilon\rangle = \frac{\varepsilon}{\mathrm{e}^{\varepsilon/\tau}-1} \tag{1}$$

推导 1

令第 n 个能级的能量为 ε_n，能量的平均值为

$$\langle\varepsilon\rangle = \frac{\sum\limits_{n=0}^{\infty}\varepsilon_n\exp(-\varepsilon_n/\tau)}{\sum\limits_{n=0}^{\infty}\exp(-\varepsilon_n/\tau)} \tag{2}$$

其中 $\sum\limits_{n=0}^{\infty}\exp(-\varepsilon_n/\tau)$ 就是配分函数 Q.

对于等间距能级，假设等间距能级 $\varepsilon_n = n\varepsilon$（也可以假设 $\varepsilon_n = \varepsilon_0 + n\varepsilon$，上式的分子分母都多出一个因子 $\exp(-\varepsilon_0/\tau)$，最后的结果相同）. 首先化简配分函数

$$Q = \sum_{n=0}^{\infty}\mathrm{e}^{-n\varepsilon/\tau} = \sum_{n=0}^{\infty}(\mathrm{e}^{-\varepsilon/\tau})^n \tag{3}$$

由于 $\mathrm{e}^{-\varepsilon/\tau} < 1$，根据等比数列求和公式

$$Q = \frac{1}{1-\mathrm{e}^{-\varepsilon/\tau}} \tag{4}$$

再化简分子

$$\sum_{n=0}^{\infty}n\varepsilon\exp(-n\varepsilon/\tau) = \varepsilon\sum_{n=0}^{\infty}n(\mathrm{e}^{-\varepsilon/\tau})^n \tag{5}$$

令 $x = \mathrm{e}^{-\varepsilon/\tau}$，有 $x < 1$. 同样根据等比数列求和公式

$$\sum_{n=0}^{\infty}nx^n = x\frac{\mathrm{d}}{\mathrm{d}x}\sum_{n=0}^{\infty}x^n = x\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{1-x}\right) = \frac{x}{(1-x)^2} \tag{6}$$

把 x 换成 $\mathrm{e}^{-\varepsilon/\tau}$，式 5 变为

$$\frac{\varepsilon \mathrm{e}^{-\varepsilon/\tau}}{(1-\mathrm{e}^{-\varepsilon/\tau})^2} \tag{7}$$

把分子分母代入平均值公式式 2 得到最后结论

$$\langle \varepsilon \rangle = \frac{\varepsilon \mathrm{e}^{-\varepsilon/\tau}}{(1-\mathrm{e}^{-\varepsilon/\tau})^2} \Big/ \frac{1}{1-\mathrm{e}^{-\varepsilon/\tau}} = \frac{\varepsilon \mathrm{e}^{-\varepsilon/\tau}}{1-\mathrm{e}^{-\varepsilon/\tau}} = \frac{\varepsilon}{\mathrm{e}^{\varepsilon/\tau}-1} \tag{8}$$

推导 2

由式 4 已知配分函数

$$Q = \frac{1}{1-\mathrm{e}^{-\varepsilon/\tau}} = \frac{1}{1-\mathrm{e}^{-\varepsilon\beta}} \tag{9}$$

我们也可以直接用能量均值公式

$$\langle \varepsilon \rangle = -\frac{\partial}{\partial \beta} \ln Q = \frac{\partial}{\partial \beta} \ln\left(1-\mathrm{e}^{-\varepsilon\beta}\right) = \frac{\varepsilon \mathrm{e}^{-\varepsilon\beta}}{1-\mathrm{e}^{-\varepsilon\beta}} = \frac{\varepsilon}{\mathrm{e}^{\varepsilon/\tau}-1} \tag{10}$$

结果相同.

巨正则系综法

- 本词条处于草稿阶段.

中心思想

一系统与热源达到平衡，热源温度为 T，化学势为 μ，那么系统在任意一个包含 N 个粒子，能量 E 的非简并状态的概率为

$$\frac{\mathrm{e}^{(\mu N-E)/(kT)}}{\Xi} \tag{1}$$

其中，巨配分函数使所有状态的概率之和为一，起到归一化的作用.

$$\Xi = \sum_N \sum_i \mathrm{e}^{(\mu N-E_i)/(kT)} \tag{2}$$

推导

“能级导向”

$$\begin{aligned}\Xi &= \sum_{N=1}^{\infty}\sum_{\{n_i\}}^{*}\exp\left(N\mu-\sum_{i=1}^{\infty}n_i\varepsilon_i\right)\beta = \sum_{N=1}^{\infty}\sum_{\{n_i\}}^{*}z^N\prod_{i=0}^{\infty}\left(\mathrm{e}^{-\varepsilon_i\beta}\right)^{n_i}\\ &= \sum_{N=1}^{\infty}\sum_{\{n_i\}}^{*}\prod_{i=0}^{\infty}\left(z\mathrm{e}^{-\varepsilon_i\beta}\right)^{n_i} = \sum_{n_1}^{*}\sum_{n_2}^{*}\cdots\prod_{i=0}^{\infty}\left(z\mathrm{e}^{-\varepsilon_i\beta}\right)^{n_i}\\ &= \sum_{n_1}^{*}\left(z\mathrm{e}^{-\varepsilon_i\beta}\right)^{n_1}\sum_{n_2}^{*}\left(z\mathrm{e}^{-\varepsilon_i\beta}\right)^{n_2}\cdots = \prod_{i}^{\infty}\sum_{n_i}^{*}\left(z\mathrm{e}^{-\varepsilon_i\beta}\right)^{n_i}\end{aligned} \tag{3}$$

系统的热力学性质

由最大概率项假设，

$$1 = \frac{\Omega\mathrm{e}^{(\mu N-E)/(kT)}}{\Xi} = \frac{\mathrm{e}^{S/k}\mathrm{e}^{(\mu N-E)/(kT)}}{\Xi} \tag{4}$$

$$\mathrm{e}^{S/k}\mathrm{e}^{(\mu N-E)/(kT)} = \Xi \tag{5}$$

$$E - ST - \mu N = -kT\ln\Xi \tag{6}$$

令 $\Phi = -kT\ln\Xi$ 叫做**巨势**

$$\Phi = E - ST - \mu N \tag{7}$$

$$\Phi = E - ST - \mu N = F - G = E - ST - (E - ST + PV) = -PV \tag{8}$$

考虑到 $\mathrm{d}E = T\,\mathrm{d}S - P\,\mathrm{d}V + \mu\,\mathrm{d}N$

$$\mathrm{d}\Phi = -P\,\mathrm{d}V - S\,\mathrm{d}T - N\,\mathrm{d}\mu \tag{9}$$

所以

$$S = -\left(\frac{\partial\Phi}{\partial T}\right)_{V,\mu} \qquad N = -\left(\frac{\partial\Phi}{\partial\mu}\right)_{V,T} \qquad P = -\left(\frac{\partial\Phi}{\partial V}\right)_{T,\mu} \tag{10}$$

另外有一个求能级分布的公式

$$\langle n_i\rangle = \frac{1}{\Xi}\sum_{N=1}^{\infty}\sum_{\{n_i\}}^{*}n_i\exp\left(N\mu-\sum_{i=1}^{\infty}n_i\varepsilon_i\right)\beta = -\frac{1}{\beta\Xi}\frac{\partial\Xi}{\partial\varepsilon_i} = -kT\frac{\partial}{\partial\varepsilon_i}\ln\Xi = \frac{\partial\Phi}{\partial\varepsilon_i} \tag{11}$$

量子气体（单能级巨正则系综法）

- 本词条处于草稿阶段.

我们可以把一个包含许多粒子的系统看做热池，把每个能级 ε_i 看做一个系统. 为了便于理解，可以把能级想象成一个盒子，所有处于该能级的粒子都在盒内，都具有能量 ε_i. 当 ε_i 系统中粒子数为 n_i 时，系统的总能量为 $E=n_i\varepsilon_i$. 注意对于一个 n_i，由于同种粒子不可区分，系统只有一种状态，所以在当前系统的巨配分函数中，对能量的求和只有一项.

$$\begin{aligned}\Xi&=\sum_{n_i}\sum_{E_j}\mathrm{e}^{(n_i\mu-E_j)\beta}=\sum_{n_i}\mathrm{e}^{(n_i\mu-E)\beta}\\&=\sum_{n_i}\mathrm{e}^{(n_i\mu-n_i\varepsilon_i)\beta}=\sum_{n_i}\left[\mathrm{e}^{(\mu-\varepsilon_i)\beta}\right]^{n_i}\end{aligned}\tag{1}$$

系统（ε_i 能级）中的平均粒子数为

$$\begin{aligned}\langle n_i\rangle&=\frac{1}{\Xi}\sum_{n_i}\sum_{E_j}n_i\mathrm{e}^{(n_i\mu-E_j)\beta}\\&=\frac{1}{\Xi}\sum_{n_i}n_i\left[\mathrm{e}^{(\mu-\varepsilon_i)\beta}\right]^{n_i}\end{aligned}\tag{2}$$

费米子

由于泡利不相容原理，一个能级只能存在 0 或 1 个费米子（这里忽略自旋）.

$$\Xi=\sum_{n_i=0}^{1}\left[\mathrm{e}^{(\mu-\varepsilon_i)\beta}\right]^{n_i}=1+\mathrm{e}^{(\mu-\varepsilon_i)\beta}\tag{3}$$

ε_i 能级的平均粒子数为

$$\langle n_i\rangle=\frac{1}{\Xi}\sum_{n_i=0}^{1}n_i\left[\mathrm{e}^{(\mu-\varepsilon_i)\beta}\right]^{n_i}=\frac{0+\mathrm{e}^{(\mu-\varepsilon_i)\beta}}{1+\mathrm{e}^{(\mu-\varepsilon_i)\beta}}=\frac{1}{\mathrm{e}^{(\varepsilon_i-\mu)\beta}+1}\tag{4}$$

这就是著名的**费米—狄拉克分布**.

玻色子

任何能级都允许同时存在任意数量的玻色子，所以上面两式中对 n_i 的求和上限变为正无穷即可（见等比数列求和以及类等比数列求和）. 但为了使求和收敛，必须要求 $\mathrm{e}^{(\varepsilon_i-\mu)\beta}-1>0$，或者 $\mu<\varepsilon_i$.

$$\Xi=\sum_{n_i=0}^{\infty}\left[\mathrm{e}^{(\mu-\varepsilon_i)\beta}\right]^{n_i}=\frac{1}{1-\mathrm{e}^{(\mu-\varepsilon_i)\beta}} \tag{5}$$

$$\langle n_i\rangle=\frac{1}{\Xi}\sum_{n_i=0}^{1}n_i\left[\mathrm{e}^{(\mu-\varepsilon_i)\beta}\right]^{n_i}=\left[1-\mathrm{e}^{(\mu-\varepsilon_i)\beta}\right]\frac{\mathrm{e}^{(\mu-\varepsilon_i)\beta}}{\left[1-\mathrm{e}^{(\mu-\varepsilon_i)\beta}\right]^2}=\frac{1}{\mathrm{e}^{(\varepsilon_i-\mu)\beta}-1} \tag{6}$$

这就是著名的**玻色—爱因斯坦分布**.

当每个能级的平均粒子数 $\langle n_i\rangle$ 都很小时，即 $\langle n_i\rangle\ll 1$ 时，$\langle n_i\rangle=1/(\mathrm{e}^{(\varepsilon_i-\mu)\beta}\pm 1)$ 的分母 $\gg 1$，分布可以近似为

$$\langle n_i\rangle=\frac{1}{\mathrm{e}^{(\varepsilon_i-\mu)\beta}}=\mathrm{e}^{(\mu-\varepsilon_i)\beta} \tag{7}$$

这就是**麦克斯韦—玻尔兹曼分布**，对应理想气体. 由此可见，当

该分布的总粒子数为

$$N=\sum_{i}^{\infty}\langle n_i\rangle=\mathrm{e}^{\mu\beta}\sum_{i}^{\infty}\mathrm{e}^{-\varepsilon_i\beta}=zQ_1 \tag{8}$$

为了验证该式的正确性，代入理想气体的化学势和单粒子配分函数，上式成立.

$$\mu=kT\ln\frac{N\lambda^3}{V}\qquad Q_1=\frac{V}{\lambda^3} \tag{9}$$

这种方法虽然可以简单地求出分布函数，但却不能求出其他物理量，例如量子气体的压强，熵，等. 因为我们的系统只包含一个能级，而不是大量粒子. 要使用标准的巨正则系综，必须把包含大量粒子的量子气体作为系统，并考虑每个粒子数对应的所有可能的能级分布.

量子气体（巨正则系宗）

- 本词条处于草稿阶段.

假设 N 个玻色子（不可区分）之间没有互相作用，每个粒子具有能级 ε_i（这里先假设每个粒子都只有 m 个能级而不是无穷多个，最后再说明 $m\to\infty$ 时成立）. 化学势为 μ，则巨分配函数为

$$\begin{aligned}\Xi&=\sum_{N=0}^{\infty}\sum_{\{n_i\}}^{*}\mathrm{e}^{(N\mu-\varepsilon_i)\beta}=\sum_{N=0}^{\infty}\sum_{\{n_i\}}^{*}z^N\exp\left(-\beta\sum_{i}^{m}n_i\varepsilon_i\right)\\&=\sum_{N=0}^{\infty}\sum_{\{n_i\}}^{*}\left(\prod_{i=1}^{\infty}z^{n_i}\right)\left(\prod_{i=1}^{\infty}(\mathrm{e}^{-\beta\varepsilon_i})^{n_i}\right)=\sum_{N=0}^{\infty}\sum_{\{n_i\}}^{*}\prod_{i=1}^{\infty}(z\mathrm{e}^{-\beta\varepsilon_i})^{n_i}\end{aligned}\tag{1}$$

其中 $\sum_{\{n_i\}}^{*}$ 求和时，由于同一个能级上的玻色子数量不限，限制条件仅为 $\sum_i^m n_i=N$（对于费米子，由于不相容原理，要求 $\sum_i^m n_i=N$ 以及 $n_i\leqslant 1$）.

巨正则系综的最大优势就是可以利用关系

$$\sum_{N=0}^{\infty}\sum_{\{n_i\}}^{*}(\dots)=\sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty}\cdots\sum_{m=0}^{\infty}(\dots)\tag{2}$$

于是

$$\begin{aligned}\Xi&=\sum_{N=0}^{\infty}\sum_{\{n_i\}}^{*}\prod_{i=1}^{\infty}(z\mathrm{e}^{-\beta\varepsilon_i})^{n_i}=\sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty}\cdots\sum_{n_m=0}^{\infty}\prod_{i=1}^{\infty}(z\mathrm{e}^{-\beta\varepsilon_i})^{n_i}\\&=\sum_{n_1=0}^{\infty}(z\mathrm{e}^{-\beta\varepsilon_1})^{n_1}\sum_{n_2=0}^{\infty}(z\mathrm{e}^{-\beta\varepsilon_2})^{n_2}\cdots\sum_{n_m=0}^{\infty}(z\mathrm{e}^{-\beta\varepsilon_m})^{n_m}\\&=\prod_{i=1}^{m}\sum_{n_i}^{\infty}(z\mathrm{e}^{-\beta\varepsilon_i})^{n_i}\end{aligned}\tag{3}$$

光子气体

- 本词条处于草稿阶段.

预备知识 等间隔能级系统（正则系宗）[1606]

我们仍用正则系宗下推导光子气体的总能能量和压强．已知单模式（单频率）的光子气体配分函数为 $Q(\omega)=1/[1-\exp(-\omega\hbar\beta)]$，系统的配分函数为

$$Q=\prod_i Q(\omega_i)=\prod_i \frac{1}{1-\exp(-\omega_i\hbar\beta)} \tag{1}$$

系统总能量为

$$U=-\frac{\partial}{\partial\beta}\ln Q=\sum_i \frac{\partial}{\partial\beta}\ln\left(1-\mathrm{e}^{-\omega_i\hbar\beta}\right)=\sum_i \frac{\omega_i\hbar\mathrm{e}^{-\omega_i\hbar\beta}}{1-\mathrm{e}^{-\omega_i\hbar\beta}}=\sum_i \frac{\omega_i\hbar}{\mathrm{e}^{\omega_i\hbar\beta}-1} \tag{2}$$

用积分来求和，模式密度为 $\rho(\omega)=V\omega^2/(\pi^2c^3)$

$$U=\frac{V}{\pi^2c^3}\int_0^\infty \frac{\omega^3\hbar}{\mathrm{e}^{\omega\hbar\beta}-1}\,\mathrm{d}\omega \tag{3}$$

用换元积分法，令 $x=\omega\hbar\beta$

$$U=\frac{V}{\pi^2c^3\beta^4\hbar^3}\int_0^\infty \frac{x}{\mathrm{e}^x-1}\,\mathrm{d}x=\frac{\pi^2V}{15c^3\hbar^3}(kT)^4 \tag{4}$$

最后的积分可由 Mathematica 完成．

压强

由正则系宗法的步骤，先求出亥姆霍兹自由能

$$F=-kT\ln Q=kT\sum_i \ln\left(1-\mathrm{e}^{-\omega_i\hbar\beta}\right) \tag{5}$$

理想气体的熵：纯微观分析

预备知识 熵[1576]

设某理想气体由 N 个原子组成，体积为 V，能量为 U，现在，我们利用玻尔兹曼熵公式计算它的熵（可以加上一个常量）．假设 N 为定值，但是 U 和 V 可以变化．我们要求的量是 $S(U,V)$．

理想气体的能量全部为动能，与粒子的位置无关．我们必须求与 U 和 V 相对应的状态数 $\Omega(U,V)$ 的对数．

我们已经知道，

$$\Omega\left(U,V\right)=\left(\frac{V}{a^3}\right)^N\times\Omega_p\left(U\right)\tag{1}$$

式中 V/a^3 是每个原子可占据的位置的数量；$\Omega_p(U)$ 是内能为 U 的气体中，动量微观分布的数目．（计算自由膨胀的**熵变**时，膨胀前后内能不变，因此，我们忽略了 $\Omega_p(U)$．现在 U 可以变化，所以需要计算 $\Omega_p(U)$，这使得我们的工作会更艰苦一些）．内能为（对容器内每个原子的所有可能组态）

$$U=\sum_{i=1}^{N}\frac{1}{2}m\left|\boldsymbol{v}_i\right|^2=\sum_{i=1}^{N}\frac{\left|\boldsymbol{p}_i\right|^2}{2m}=\sum_{i=1}^{N}\frac{p_{ix}^2+p_{iy}^2+p_{iz}^2}{2m}\tag{2}$$

式中 $\boldsymbol{p}=m\boldsymbol{v}$ 是动量．

现在我们构造一个 $3N$ 维的矢量 P：

$$\mathrm{P}=\left(p_{1x},p_{1y},p_{1z},p_{2x},\cdots,p_{Nz}\right)\tag{3}$$

它不过是 N 个动量矢量 $\boldsymbol{p}_i$ 的 3 个分量的集合．如果我们将 P 的分量重新编号为 $j=1,\cdots,3N$，则

$$\mathrm{P}=\left(p_{1x},p_{1y},p_{1z},p_{2x},\cdots,p_{Nz}\right)\tag{4}$$

这就是说

$$P_1=p_{1x},P_2=p_{1y},P_3=p_{1z},P_4=p_{2x},\cdots,P_{3N}=p_{Nz}\tag{5}$$

内能可以写成

$$U=\sum_{j=1}^{3N}\frac{P_j^2}{2m}\tag{6}$$

不考虑位置的话，原子的动量可以取满足(6)的任意值，所以我们必须看看满足条件的动量值有多少个．我们将该条件改写为

$$\sum_{j=1}^{3N}P_j^2=2mU\tag{7}$$

这是 $3N$ 维空间中半径为 $R=\sqrt{2mU}$ 的超球方程，就像

$$x^2+y^2=R^2\quad\text{圆或者 1 维球}\tag{8}$$

或

$$x^2+y^2+z^2=R^2 \quad \text{普通的球或者 2 维球} \tag{9}$$

在数学文献中，$d=2$ 的圆和 $d=3$ 的普通球都被叫作球，周长和表面积都统称为面积．对千我们熟悉的这两个例子，面积是 $2\pi R$ 和 $4\pi R^2$．对于半径为 R 的 d 维球，借助量纲分析可得，其面积按 R^{d-1} 关系变化．在我们的问题中，$R=\sqrt{2mU}$ 且 $d-1=3N-1\approx 3N$．如果将各个动址分为大小为 b^3 的单元，就像 a^3，很小且可以任意选取，那么气体可能具有的总状态数为

$$\Omega(V,U)=V^N U^{3N/2} F(m,N,a,b) \tag{10}$$

式中，我们专注于它随 U 和 V 的变化，将其他因素 m,a,b 和 N 合并，写入了函数 $F(m,N.a,b)$．我们不需要详细地了解 F，因为我们仅要将

$$S=k\ln\Omega=k\left[N\ln V+\frac{3}{2}N\ln U\right]+k\ln F(m,N,a,b) \tag{11}$$

对 U 和 V 求偏导，而 F 对此没有贡献．偏导数为

$$\left.\frac{\partial S}{\partial V}\right|_U=\frac{kN}{V} \tag{12}$$

$$\left.\frac{\partial S}{\partial U}\right|_V=\frac{3kN}{2U} \tag{13}$$

而我们又知道上面的两个导数分别等于 P/T 和 $1/T$，这是由热力学得到的相应结果．于是

$$\frac{kN}{V}=\frac{P}{T} \tag{14}$$

$$\frac{3kN}{2U}=\frac{1}{T} \tag{15}$$

这样，从玻尔兹曼熵公式我们可以得到理想气体状态方程．

伊辛模型

- 本词条处于草稿阶段．
- 本词条缺少预备知识，初学者可能会遇到困难．

[1]平均磁化率

$$\langle M\rangle = \frac{\sum_i M_i \mathrm{e}^{-M_i H\beta}}{\sum_i M_i \mathrm{e}^{-M_i H\beta}} = -kT\frac{\partial}{\partial H}\ln Q \tag{1}$$

平均能量

$$U = -\langle M\rangle H \tag{2}$$

统计力学公式

- 本词条处于草稿阶段.
- 需要增加相应内容的链接.

微分关系

$$H = E + PV \tag{1}$$

$$G = E + PV - ST \tag{2}$$

$$\mu\,\mathrm{d}N + N\,\mathrm{d}\mu = \mathrm{d}G = V\,\mathrm{d}P - S\,\mathrm{d}T + \mu\,\mathrm{d}N \tag{3}$$

微正则系综

$$\mathrm{d}S = \frac{1}{T}\mathrm{d}E + \frac{P}{T}\mathrm{d}V - \frac{\mu}{T}\mathrm{d}N \tag{4}$$

正则系宗

$$-kT\ln Q = F = E - ST \tag{5}$$

$$\mathrm{d}F = -S\,\mathrm{d}T - P\,\mathrm{d}V + \mu\,\mathrm{d}N \tag{6}$$

$$E = -\frac{\partial}{\partial\beta}\ln Q \tag{7}$$

[1]参考 Wikipedia 相关页面.

巨正则系综

$$-PV=\Phi=E-ST-\mu N \tag{8}$$

$$\Phi=-kT\ln\Xi \tag{9}$$

$$\mathrm{d}\Phi=-P\,\mathrm{d}V-S\,\mathrm{d}T-N\,\mathrm{d}\mu \tag{10}$$

$$\langle n_i\rangle=\frac{\partial\Phi}{\partial\varepsilon_i} \tag{11}$$

理想气体

$$V_n=\frac{\pi^{n/2}R^n}{\Gamma(1+n/2)}\qquad（N\text{ 维球体}） \tag{12}$$

$$\Omega_0=\frac{V^N}{N!h^3}\frac{(2\pi mE)^{3N/2}}{(3N/2)!}\qquad（N\text{ 粒子能级密度}） \tag{13}$$

$$a(\varepsilon)=\frac{2\pi V(2m)^{3/2}}{h^3}\varepsilon^{1/2}\qquad（\text{单粒子能及密度}） \tag{14}$$

$$S=Nk\left(\ln\frac{V}{N\lambda^3}+\frac{5}{2}\right)\qquad（\text{熵}） \tag{15}$$

$$N=zQ_1\Rightarrow\mu=kT\ln\frac{N\lambda^3}{V}\qquad（\text{化学势}） \tag{16}$$

$$\Xi=\ln N\qquad（\text{巨势}） \tag{17}$$

量子气体

$$N=Q_1g_{3/2}(z)=\frac{V}{\lambda^3}g_{3/2}(z)\qquad（BE）\tag{18}$$

$$N=Q_1f_{3/2}(z)\qquad（FD）\tag{19}$$

$$\frac{PV}{kT}=Q_1g_{5/2}(z)=\frac{V}{\lambda^3}g_{5/2}(z)\qquad（BE）\tag{20}$$

$$\frac{PV}{kT}=Q_1f_{5/2}(z)\qquad（FD）\tag{21}$$

$$PV=NkT\frac{g_{5/2}(z)}{g_{3/2}(z)}\qquad（BE）\tag{22}$$

$$PV = NkT\frac{f_{5/2}(z)}{f_{3/2}(z)} \qquad （FD）\tag{23}$$

$$E = \frac{3}{2}PV \qquad （BE\text{ 和 }FD）\tag{24}$$

理论上可以通过三式中的任意两式消去 z，但是不能写成解析形式.

$$\begin{aligned} g_n(z) &= z + \frac{z^2}{2^n} + \frac{z^3}{3^n}\cdots \\ f_n(z) &= z - \frac{z^2}{2^n} + \frac{z^3}{3^n}\cdots \end{aligned}\tag{25}$$

BE 凝聚态

$$N = \frac{V}{\lambda_c^3}g_{3/2}(1) \Rightarrow T_c = \frac{h^2}{2\pi mk}\left(\frac{N}{2.612V}\right)^{2/3}\tag{26}$$

$$\frac{N_e}{N} = \frac{\lambda^3}{\lambda_c^3} \Rightarrow N_e = N\left(\frac{T}{T_c}\right)^{3/2} \Rightarrow N_0 = N\left[1 - \left(\frac{T}{T_c}\right)^{3/2}\right]\tag{27}$$

$$N_0 = \frac{1}{\mathrm{e}^{(\varepsilon_0-\mu)/kT} - 1} = \frac{kT}{\varepsilon_0 - \mu}\tag{28}$$

$$\varepsilon_0 - \mu \ll \varepsilon_1 - \varepsilon_0 \Rightarrow \varepsilon_0 - \mu \ll \varepsilon_1 - \mu\tag{29}$$

$$N_1 = \frac{1}{\mathrm{e}^{(\varepsilon_1-\mu)/kT} - 1} < \frac{kT}{\varepsilon_1 - \mu} \ll \frac{kT}{\varepsilon_0 - \mu} = N_0\tag{30}$$

范德瓦尔斯方程

$$\left(P + \frac{aN^2}{V^2}\right)(V - bN) = NkT\tag{31}$$

量子转子能级

角量子数 l 决定能级

$$E_l = l(l+1)\frac{\hbar^2}{2IkT}\tag{32}$$

$2l+1$ 重简并，其中 $I = m_1m_2r_{12}^2/(m_1+m_2)$ 为质心转动惯量. 当 l 为偶数时，两粒子的波函数具有交换对称，奇数时反对称. 两原子核的自旋共有 $s^2 = (2I+1)^2$ 种状态，其中对称态占 $s(s+1)/2$ 种，反对称太占 $s(s-1)/2$ 种.

若两粒子都是费米子（I 为半整数），则总波函数反对称，即 l 为单数核自旋对称，或 l 为偶数核自旋反对称.

弹簧振子能级

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right) \tag{33}$$

非简并.

为什么书上说 $m = 0$ （能级密度与 ε^m 成正比）不能产生凝聚态，然而我在模拟中做到了？

第十八部分

近代物理

第一章

量子场论

标量场

预备知识 洛伦兹群[1286]，欧拉—拉格朗日方程[1052]

定义 1

标量场即洛伦兹变换下不变的场．设在两个惯性系中，标量场分别为 $\phi(x)$ 和 $\phi'(x')$，那么有

$$\phi'(x') = \phi(x) \tag{1}$$

图景是这样的，对于一个固定点 P 而言，洛伦兹变换前后对应的坐标不同，场函数形式不同，结果是 P 的数值不变．

推论 1

做一个无穷小变换

$$x^\rho \to x'^\rho = x^\rho + \delta x^\rho \tag{2}$$

并且

$$\delta x^\rho = \omega^\rho_\sigma x^\sigma = -\frac{i}{2}\omega_{\mu\nu}\left(J^{\mu\nu}\right)^\rho_\sigma x^\sigma \tag{3}$$

其中，$\left(J^{\mu\nu}\right)^\rho_\sigma = i\left(\eta^{\mu\rho}\delta^\nu_\sigma - \eta^{\nu\rho}\delta^\mu_\sigma\right)$ 显然，令 $J^{\mu\nu} = 0$，该标量表示 $(0,0,)$ 能满足标量场的定义．这是个平庸的结果，在场表示里，我们能做到更多．对于固定点 x，我们做场的无穷小变换

$$\delta_0\phi \equiv \phi'(x) - \phi(x) \tag{4}$$

为了找到该表示下的生成元，我们做一阶泰勒展开

$$\delta_0\phi = \phi'(x' - \delta x) - \phi(x) = -\delta x^\rho \partial_\rho \phi(x) \tag{5}$$

代入式 3，我们有

$$\delta_0\phi = \frac{i}{2}\omega_{\mu\nu}\left(J^{\mu\nu}\right)^\rho_\sigma x^\sigma \partial_\rho \phi \equiv -\frac{i}{2}\omega_{\mu\nu}L^{\mu\nu}\phi \tag{6}$$

在这里，我们定义

$$L^{\mu\nu} = -\left(J^{\mu\nu}\right)^\rho_\sigma x^\sigma \partial_\rho = i\left(x^\mu\partial^\nu - x^\nu\partial^\mu\right) \tag{7}$$

我们可以验证，$L^{\mu\nu}$ 满足洛伦兹群的李代数，从而确实是洛伦兹变换的生成元．所以在这里我们给出了无限维的洛伦兹群的标量场表示．（由于场函数随着坐标不同而不同，所以是无限维的）

实际上，如果定义 $p^\mu = +i\partial^\mu$，会有 $L^{\mu\nu} = x^\mu p^\nu - x^\nu p^\mu$，若上标取空间坐标 i,j，我们会发现这就是轨道角动量．由于 $J^{\mu\nu} = L^{\mu\nu} + S^{\mu\nu}$，所以标量场表示是自旋为 0 的．

综上所述，我们得到了洛伦兹群的标量场表示，标量场在该表示下进行洛伦兹变换，由于该表示自旋为 0，所以标量场代表的是自旋为 0 的场．

举例

克莱因—戈登场即为标量场．克莱因—戈登方程为

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\psi - \nabla^2\psi + \frac{m^2c^2}{\hbar^2}\psi = 0 \tag{8}$$

取自然单位，$c = \bar{h} = 1$，得到协变形式

$$\left(\Box^2 + m^2\right)\psi = 0 \tag{9}$$

达朗贝尔算符 $\Box^2 = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2$ 在经典场论，它的拉格朗日量为 $L = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2$，由欧拉-拉格朗日方程可解得克莱因-戈尔登方程方程．根据能量本征值解，我们知道这是一个相对论性的自旋为 0 的标量场．

推论 2

以上我们得到了经典场论下的标量场，接着我们将标量场进行正则量子化．所谓量子化，实际上是把场看做无数个简谐振子，这由产生和湮灭算符体现．

对于量子力学，我们有坐标-动量基本对易关系 $\left[\hat{q^i}, \hat{p^j}\right] = i\delta^{ij}$ 若在海森堡绘景，该对易子则是等时性的．同样的，在量子场论里，我们也可以有类似的操作．场 $\hat{\phi}(t, \mathrm{x})$ 对应于坐标算符 $\hat{q^i}$，场的共轭动量定义为

$$\hat{\Pi}_{\hat{\phi}} = \frac{\partial\hat{\mathcal{L}}}{\partial\left(\partial_0\phi\right)} = \partial_0\hat{\phi} \tag{10}$$

，其中 $\hat{\mathcal{L}}$ 为拉氏量密度，**Action** $S = \int dtL = \int d^4x\mathcal{L}\left(\phi, \partial_\mu\phi\right)$

$\hat{\Pi}(t, \mathrm{x})$ 对应于动量算符 $\hat{q^i}$，则有

$$[\hat{\phi}(t, \mathrm{x}), \hat{\Pi}(t, \mathrm{y})] = i\delta^{(3)}(\mathrm{x} - \mathrm{y}) \tag{11}$$

等时下

$$[\hat{\phi}(t,\mathrm{x}),\hat{\phi}(t,\mathrm{y})]=[\hat{\Pi}(t,\mathrm{x}),\hat{\Pi}(t,\mathrm{y})]=0 \tag{12}$$

实标量场是厄米算符．由此我们可得到克莱因-戈登方程的算符解，即：

$$\hat{\phi}(x)=\int\frac{d^3p}{(2\pi)^3\sqrt{2E_{\mathrm{p}}}}\left(\hat{a}_{\mathrm{p}}e^{-ipx}+\hat{a}_{\mathrm{p}}^{\dagger}e^{ipx}\right) \tag{13}$$

可以验证算符的厄米性．

其中 $ipx=ip_\mu x^\mu=iEt-i\vec{p}\vec{x}$, 可以证明，场算符的对易关系等价于

$$\left[\hat{a}_{\mathrm{p}},\hat{a}_{\mathrm{q}}^{\dagger}\right]=(2\pi)^3\delta^{(3)}(\mathrm{p}-\mathrm{q}) \tag{14}$$

$$[\hat{a}_{\mathrm{p}},\hat{a}_{\mathrm{q}}]=0,\quad\left[\hat{a}_{\mathrm{p}}^{\dagger},\hat{a}_{\mathrm{q}}^{\dagger}\right]=0 \tag{15}$$

自然而然的，从对易关系我们可以看出，场算符的两项由产生算符和湮灭算符线性叠加而成．

能量

我们需要知道量子化标量场后的哈密顿密度形式．由于

$$\hat{\mathcal{H}}=\hat{\Pi}_{\hat{\phi}}\partial_0\hat{\phi}-\hat{\mathcal{L}}=\frac{1}{2}\left[\hat{\Pi}_{\hat{\phi}}^2+(\nabla\hat{\phi})^2+m^2\hat{\phi}^2\right] \tag{16}$$

$$\hat{H}=\int d^3x\hat{\mathcal{H}}=\int\frac{d^3p}{(2\pi)^3}E_p\frac{1}{2}\left(\hat{a}_{\mathrm{p}}^{\dagger}\hat{a}_{\mathrm{p}}+\hat{a}_{\mathrm{p}}\hat{a}_{\mathrm{p}}^{\dagger}\right)=\int\frac{d^3p}{(2\pi)^3}E_p\left(\hat{a}_{\mathrm{p}}^{\dagger}\hat{a}_{\mathrm{p}}+\frac{1}{2}\left[\hat{a}_{\mathrm{p}},\hat{a}_{\mathrm{p}}^{\dagger}\right]\right) \tag{17}$$

第二项是所有谐振子的真空零点能，由于我们实际上的测值是能量差，所以这里出现的发散情况并无大碍．

对于这种操作，我们引入 **normal ordering** 这一概念，对于算符 $\hat{\mathcal{O}}$, 其 normal ordered form 表示为 $:\mathcal{O}:$，把产生算符都放左边，湮灭算符都放右边，从而得到 $:\mathcal{O}:$.

所以

$$\hat{H}=\frac{1}{2}\int d^3x:\hat{\Pi}^2+(\nabla\hat{\phi})^2+m^2\hat{\phi}^2: \tag{18}$$

因果性

我们需要验证这个理论是否符合因果性，即类空距离的算符是否都对易

$$\left[\hat{\mathcal{O}}_1(x), \hat{\mathcal{O}}_2(y)\right] = 0 \quad \forall \quad (x-y)^2 < 0 \tag{19}$$

我们定义

$$\Delta(x-y) = [\hat{\phi}(x), \hat{\phi}(y)] \tag{20}$$

可求得

$$\Delta(x-y) = \int \frac{d^3p}{(2\pi)^3} \frac{1}{2E_{\vec{p}}} \left(e^{-ip\cdot(x-y)} - e^{ip\cdot(x-y)}\right) \tag{21}$$

从上式中我们可以获知：

1. 这是洛伦兹不变的．积分测度的不变性可以由 delta 函数的性质证明，指数函数由于是数，所以也是洛伦兹不变的．

2. 类时下，该函数不为 0，类空反之．我们可以通过举特例证明这一点．（由洛伦兹不变性知特例的结果取值不变）

因而我们证明了标量场理论是符合因果律的．

传播子

在量子力学里，传播子即为跃迁振幅，$\langle \mathrm{x}'', t \mid \mathrm{x}', t_0 \rangle$ 为 t_0 时制备的具有本征值 **x'** 的粒子在 t 时刻，**x"** 处被发现的概率振幅．也就是粒子从一个时空点到另一个时空点的概率振幅．（当然啦，这只是其中一种解释）场论中亦然，由于场算符由产生算符和湮灭算符线性叠加而成，其作用于真空态则意味着在某处产生粒子．所以传播子的表达可为

$$\begin{aligned}\langle 0|\hat{\phi}(x)\hat{\phi}(y)|0\rangle &= \int \frac{d^3p d^3p'}{(2\pi)^6} \frac{1}{\sqrt{4E_{\vec{p}}E_{\vec{p}'}}} \left\langle 0 \left| a_{\vec{p}} a^{\dagger}_{\vec{p}'} \right| 0 \right\rangle e^{-ip\cdot x + ip'\cdot y} \\ &= \int \frac{d^3p}{(2\pi)^3} \frac{1}{2E_{\vec{p}}} e^{-ip\cdot(x-y)} \equiv D(x-y)\end{aligned}$$

可以验证

$$[\hat{\phi}(x), \hat{\phi}(y)] = D(x-y) - D(y-x) = 0 \quad \text{if} \quad (x-y)^2 < 0 \tag{22}$$

费曼传播子

在相互作用场论中，费曼传播子相当重要．它是这么定义的

$$\Delta_F(x-y)=\langle 0|T\hat{\phi}(x)\hat{\phi}(y)|0\rangle=\begin{cases}D(x-y) & x^0>y^0\\ D(y-x) & y^0>x^0\end{cases} \tag{23}$$

其中 T 是 **time ordering** 编时乘积，作用是把时间靠后的算符放在前面

$$T\hat{\phi}(x)\hat{\phi}(y)=\begin{cases}\hat{\phi}(x)\hat{\phi}(y) & x^0>y^0\\ \hat{\phi}(y)\hat{\phi}(x) & y^0>x^0\end{cases} \tag{24}$$

可以用留数定理证明，标量场的费曼传播子是

$$\Delta_F(x-y)=\int\frac{d^4p}{(2\pi)^4}\frac{i}{p^2-m^2}e^{-ip\cdot(x-y)} \tag{25}$$

格林函数

费曼传播子即 Klein-Gordon 方程的格林函数解．

$$\begin{aligned}\left(\partial_t^2-\nabla^2+m^2\right)\Delta_F(x-y)&=\int\frac{d^4p}{(2\pi)^4}\frac{i}{p^2-m^2}\left(-p^2+m^2\right)e^{-ip\cdot(x-y)}\\&=-i\int\frac{d^4p}{(2\pi)^4}e^{-ip\cdot(x-y)}\\&=-i\delta^{(4)}(x-y)\end{aligned} \tag{26}$$

洛伦兹群覆盖群 SL(2,C) 的不可约表示

预备知识 洛伦兹群[1286] 不可约表示，李群

李代数的重新推导

洛伦兹群的李代数是

$$\begin{aligned}[J_i,J_j]&=i\epsilon_{ijk}J_k\\ [J_i,K_j]&=i\epsilon_{ijk}K_k\\ [K_i,K_j]&=-i\epsilon_{ijk}J_k\end{aligned} \tag{1}$$

洛伦兹群元表达为

$$\Lambda = e^{-i\mathrm{J}\cdot\theta + i\mathrm{K}\cdot\phi} \tag{2}$$

引入新算符

$$\mathrm{J}^{\pm} = \frac{\mathrm{J} \pm i\mathrm{K}}{2} \tag{3}$$

经过计算，新的生成元的对易关系如下所示

$$\begin{aligned} \left[J^{+,i}, J^{+,j}\right] &= i\epsilon^{ijk} J^{+,k} \\ \left[J^{-,i}, J^{-,j}\right] &= i\epsilon^{ijk} J^{-,k} \\ \left[J^{+,i}, J^{-,j}\right] &= 0 \end{aligned} \tag{4}$$

显然，这与 $SU(2)$ 群的李代数一致，也就是说洛伦兹群包含了两份 SU(2) 的李代数．然而洛伦兹群不是单连通群．在数学上，给定李代数，它可对应许多李群，而其中唯一的单连通李群可以视为其他群的“母体”，因为总存在从这个单连通群到其他群的映射，反之则不一定，这个母群被数学家们称为“覆盖群”，并且称该群覆盖了同一李代数对应的其他群．通过推导洛伦兹群李代数的不可约表示，可以导出洛伦兹群的覆盖群的表示．（对于非单连通群，不存在 Lie 代数的不可约表示和群的表示之间的一一映射）可以证明，其覆盖群为 $SL(2,C)$．$SU(2)$ 的每一不可约表示都可以用 $SU(2)$ 的 Casimir 元对应的标量 **j** 来标记．（回忆一下，Casimir 算符是群元中与群的所有生成元都对易的算符．本征值在群元所作变换下不变所以可以拿来标记群表示．）设这两份 $SU(2)$ 的 j 为 j_-, j_+, 则 SL(2,C) 可以以 (j_-, j_+) 作为群表示的标记，维数为 $j = (2j_- + 1)(2j_+ + 1)$．

SL(2,C) 与 SO(3,1) 关系的简单证明

令坐标矢量为 (x_0, x_1, x_2, x_3)，$SO(3,1)$ 的群元为 $\Lambda.x'^{\nu} = \Lambda^{\mu}_{v} x^{\nu}$, 洛伦兹变换使得该时空距离不变，即 $x^{\nu}x^{\nu} = x'^{\nu}x'^{\nu}$.

一般的厄米矩阵可以表示为

$$\chi = \begin{bmatrix} x_0 + x_3 & x_1 - \mathrm{i}x_2 \\ x + \mathrm{i}x_2 & x_0 - x_3 \end{bmatrix} \tag{5}$$

显然，该矩阵行列式即为时空距离．对于 $\chi' = M\chi M^*$，该线性变换同样不会改变时空距离，即矩阵的行列式．由于 $\pm M$ 都对应同一个 Λ，所以 $SL(2,C)$ 是 $SO(3,1)$ 的双覆盖．

SL(2,C) 的几个表示

我们通常省略洛伦兹群双覆盖的说法，而把以下几个表示称之为洛伦兹群表示．但我们依然需要明晰，它们是洛伦兹群双覆盖的表示．

$(0,0)$ 表示

该表示维数为 1．$\mathrm{J}^{\pm}=0$，则 J,K 都是 0．所以是作用在标量上的简单表示．

$(\frac{1}{2},0)$ 表示与 $(0,\frac{1}{2})$ 表示

这两个表示都是二维表示，且自旋为 $\frac{1}{2}$，所以它们都是作用在旋量上的表示．对于 $(\frac{1}{2},0)$ 作用的旋量，我们标记为 $(\psi_L)_\alpha,\alpha=1,2$，这就是左手的 Wely 旋量.$(0,\frac{1}{2})$ 作用的旋量，我们标记为 $(\psi_R)_\alpha,\alpha=1,2$．这就是右手的 Wely 旋量．

$$\text{Weyl spinors:}\quad \psi_L\in\left(\frac{1}{2},0\right),\quad \psi_R\in\left(0,\frac{1}{2}\right) \tag{6}$$

接下来我们来推导这两个表示的具体形式．通过定义，我们知晓，$(\frac{1}{2},0)$ 表示意味着 J^- 为 2Œ2 矩阵．同时，$\mathrm{J}^+=0.$ 符合这一点，也符合李代数的解是 $\mathrm{J}^-=\boldsymbol{\sigma}/2$，所以

$$\begin{aligned}\mathrm{J}&=\mathrm{J}^+ +\mathrm{J}^-=\frac{\sigma}{2}\\ \mathrm{K}&=-i\left(\mathrm{J}^+-\mathrm{J}^-\right)=i\frac{\sigma}{2}\end{aligned} \tag{7}$$

代入式 2，左手旋量的洛伦兹变换为

$$\psi_L\to\Lambda_L\psi_L=\exp\left\{(-i\boldsymbol{\theta}-\boldsymbol{\phi})\cdot\frac{\boldsymbol{\sigma}}{2}\right\}\psi_L \tag{8}$$

对于旋转变换我们有

$$\begin{aligned} R_x\left(\theta_x\right) &= e^{-i\theta_x J_x} = e^{-i\theta_x \frac{1}{2}\sigma_x} \\ &= 1 - \frac{i}{2}\theta_x\sigma_x + \frac{1}{2}\left(-\frac{i}{2}\theta_x\sigma_x\right)^2 + \cdots \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{i}{2}\theta_x \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - \frac{1}{2}\left(\frac{-\theta_x}{2}\right)^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \cdots \\ &= \begin{pmatrix} 1 - \frac{1}{2}\left(-\frac{\theta_x}{2}\right)^2 + \cdots & i\frac{-\theta_x}{2} + \cdots \\ i\frac{-\theta_x}{2} + \cdots & 1 - \frac{1}{2}\left(\frac{-\theta_x}{2}\right)^2 + \cdots \end{pmatrix} + \cdots \\ &= \begin{pmatrix} \cos\frac{\theta_x}{2} & -i\sin\frac{\theta_x}{2} \\ -i\sin\frac{\theta_x}{2} & \cos\frac{\theta_x}{2} \end{pmatrix} \end{aligned} \tag{9}$$

同样可得，

$$R_y\left(\theta_y\right) = \begin{pmatrix} \cos\frac{\theta_y}{2} & -\sin\frac{\theta_y}{2} \\ \sin\frac{\theta_y}{2} & \cos\frac{\theta_y}{2} \end{pmatrix} \tag{10}$$

以及，

$$R_z\left(\theta_z\right) = \begin{pmatrix} \cos\frac{\theta_z}{2} - i\sin\frac{\theta_z}{2} & 0 \\ 0 & \cos\frac{\theta_z}{2} + i\sin\frac{\theta_z}{2} \end{pmatrix} = \begin{pmatrix} e^{-i\frac{\theta_z}{2}} & 0 \\ 0 & e^{i\frac{\theta_z}{2}} \end{pmatrix} \tag{11}$$

对于 boost（推动），我们有

$$\begin{aligned} B_x\left(\phi_x\right) &= \begin{pmatrix} \cosh\frac{\phi_x}{2} & \sinh\frac{\phi_x}{2} \\ \sinh\frac{\phi_x}{2} & \cosh\frac{\phi_x}{2} \end{pmatrix} \\ B_y\left(\phi_y\right) &= \begin{pmatrix} \cosh\frac{\phi_y}{2} & -i\sinh\frac{\phi_y}{2} \\ i\sinh\frac{\phi_y}{2} & \cosh\frac{\phi_y}{2} \end{pmatrix} \\ B_z\left(\phi_z\right) &= \begin{pmatrix} \cosh\frac{\phi_z}{2} + \sinh\frac{\phi_z}{2} & 0 \\ 0 & \cosh\frac{\phi_z}{2} - \sinh\frac{\phi_z}{2} \end{pmatrix} = \begin{pmatrix} e^{\frac{\phi_z}{2}} & 0 \\ 0 & e^{-\frac{\phi_z}{2}} \end{pmatrix}. \end{aligned} \tag{12}$$

显然这与矢量的群表示相差甚远．对于 $(0, \frac{1}{2})$，我们可以重复以上过程．从定义中解得 $\mathrm{J} = \sigma/2, \mathrm{K} = -i\sigma/2$．所以右手旋量的变换为

$$\psi_R \to \Lambda_R \psi_R = \exp\left\{(-i\boldsymbol{\theta} + \boldsymbol{\phi}) \cdot \frac{\sigma}{2}\right\} \psi_R \tag{13}$$

（$\frac{1}{2},\frac{1}{2}$）表示

根据定义，我们知道此时 $SL(2,C)$ 包含了两份独立变换的 $SU(2)$ 表示（因为 $[J^{+,i},J^{-,j}]=0$，所以是互相独立的），维数为 4 维．那么它的作用对象必定也具有两个独立变量，代表在不同的 SU(2) 群下进行变换．

$$\left(\frac{1}{2},\frac{1}{2}\right)=\left(\frac{1}{2},0\right)\otimes\left(0,\frac{1}{2}\right) \tag{14}$$

从维数上看，作用对象可能是一个 2Œ2 矩阵．我们希望定义一个表示变换，使得该变换等价于常见的 $SO(3,1)$ 群作用在四维矢量的效果．这种效果上的等价性类似于二维旋转操作可以用单位复数作用到向量对应的复数上来表示，在定义了复数与实数矩阵间的映射后，旋转也可以用实矩阵作用于列向量来表示．那么，这个 2Œ2 矩阵是什么类型的呢？首先考虑复数矩阵，一般的复数矩阵有 8 个自由参数，由前文知，我们的（$\frac{1}{2},\frac{1}{2}$）表示作用对象是四维的，具有 4 个独立参数．可以验证，任意的复数矩阵 M 可以由厄米矩阵 $H\left(H^{\dagger}=H\right)$ 和反厄米矩阵 $A\left(A^{\dagger}=-A\right)$ 组成:$M=H+A$，而厄米矩阵和反厄米矩阵都具有四个独立参数．那么我们可以将变换限制在厄米矩阵上．即寻求（$\frac{1}{2},\frac{1}{2}$）的不可约表示，使得厄米矩阵在变换后依然是厄米矩阵（即厄米矩阵构成四维复数矩阵的不变子空间，该表示是厄米矩阵上的不可约表示），且等价于四矢量变换．（即时空距离不变）一般的厄米矩阵以泡利矩阵为基

$$\sigma^1=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \sigma^2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma^3=\begin{pmatrix}1&0\\0&-1\end{pmatrix} \tag{15}$$

$$\sigma^{\mu}=\left(1,\sigma^i\right),\quad \bar{\sigma}^{\mu}=\left(1,-\sigma^i\right) \tag{16}$$

$$v=\begin{pmatrix}v_0+v_3&v_1-iv_2\\v_1+iv_2&v_0-v_3\end{pmatrix}=v_0\begin{pmatrix}1&0\\0&1\end{pmatrix}+v_1\begin{pmatrix}0&1\\1&0\end{pmatrix}+v_2\begin{pmatrix}0&-i\\i&0\end{pmatrix}+v_3\begin{pmatrix}1&0\\0&-1\end{pmatrix} \tag{17}$$

在该表示下

$$v\to\Lambda_L v\bar{\Lambda}_R \tag{18}$$

$$\bar{\Lambda}_R\in\Lambda_R=e^{\frac{1}{2}(i\theta-\phi)\cdot\sigma} \tag{19}$$

我们不妨展开表示到一阶那么

$$\begin{aligned} v' = \Lambda_L v \Lambda_R &= \left(1 - \frac{i}{2}\theta_i\sigma^i - \frac{1}{2}\phi_i\sigma^i\right)\left(v^0\sigma^0 - v^i\sigma^i\right)\left(1 + \frac{i}{2}\theta_i\sigma^i - \frac{1}{2}\phi_i\sigma^i\right) \\ &= v^0 - v^i\sigma^i - \phi_i\sigma^i v^0 + \frac{i}{2}\theta_i v^j\left[\sigma^i, \sigma^j\right] + \frac{1}{2}\phi_i v^j\left\{\sigma^i, \sigma^j\right\} \\ &= v^0 - v^i\sigma^i - \phi_i\sigma^i v^0 - \epsilon^{ijk}\theta_i v^j\sigma^k + \phi_i v^i \\ &= \left(v^0 + \phi_i v^i\right) - \left(v^i + \phi_i v^0 + \epsilon^{ijk}\theta_j v^k\right)\sigma^i \end{aligned} \tag{20}$$

这正是洛伦兹变换的无穷小形式. $\begin{pmatrix} 1 & \phi_1 & \phi_2 & \phi_3 \\ \phi_1 & 1 & -\theta_3 & \theta_2 \\ \phi_2 & \theta_3 & 1 & \theta_1 \\ \phi_3 & -\theta_2 & \theta_1 & 1 \end{pmatrix}\begin{pmatrix} v^0 \\ v^1 \\ v^2 \\ v^3 \end{pmatrix} = \begin{pmatrix} v^0 + \phi_1 v^1 + \phi_2 v^2 + \\ v^1 + \phi_1 v^0 + \theta_2 v^3 - \\ v^2 + \phi_2 v^0 + \theta_3 v^1 - \\ v^3 + \phi_3 v^0 + \theta_1 v^2 - \end{pmatrix}$

所以厄米矩阵是广义的四维协变矢量，洛伦兹变换最基本的作用对象是旋量.

可以证明 $\xi_R^\dagger\sigma^\mu\psi_R$ 和 $\xi_L^\dagger\bar{\sigma}^\mu\psi_L$ 也是协变的四维矢量.

第二章

高能物理

基本粒子

- 本词条处于草稿阶段.

未完成：图翻译成中文

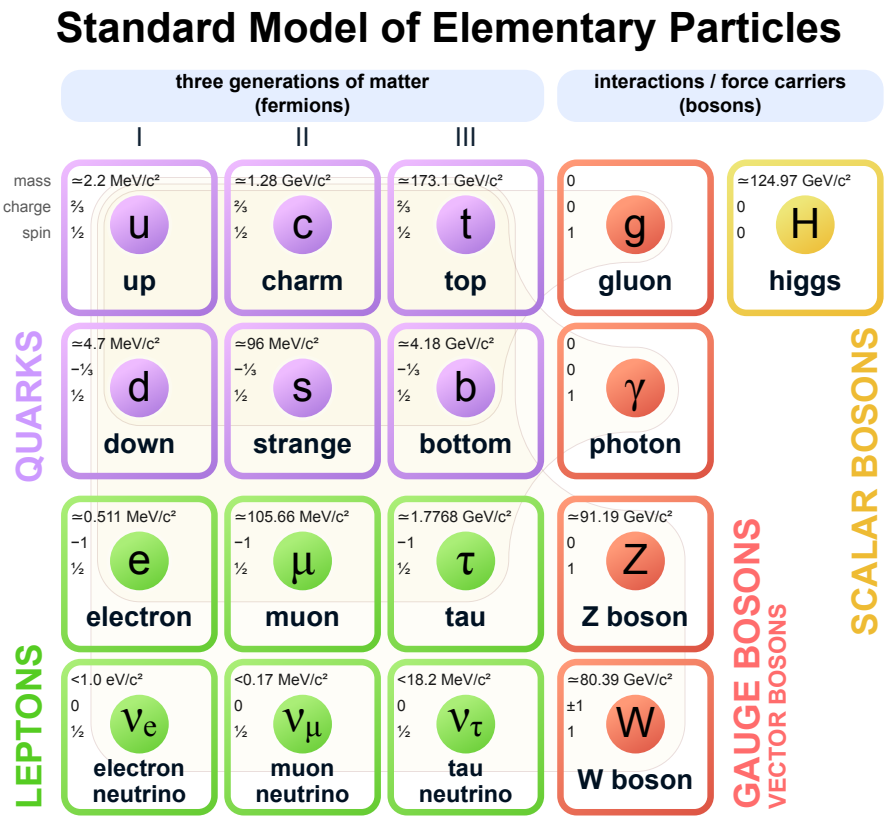


图 1: 标准模型的基本粒子（来自维基百科）

费米子包括 6 种**夸克**（**quark**）和 6 种**轻子**（**lepton**）. 玻色子包括胶子，光子，**Z 玻色子**和 **W 玻色子**，以及**希格斯子**（**higgs**）.

第三章

弦理论

引力量子化

- 本词条需要更多参考文献.
- 本词条缺少预备知识，初学者可能会遇到困难.

广义相对论包含引力波．他们带有 $J=2$ 的轨道角动量．对引力波做量子化能得到自旋 2 的引力子．弦理论自然包含了引力子，因此也是量子引力的有力候选者之一．但是引力子的圈图会有发散．经过计算，我们得到引力子圈图

$$I \sim \int d^4p \to \infty \tag{1}$$

这意味着引力是不可重整的．也就是说，把引力纳入量子场论会产生巨大的问题．弦理论处理这个问题的办法是，用一个一维的弦来代替零维的点粒子．这样的化相互作用就不会在一个点发生了．我们看如下的不确定性关系

$$\Delta x \Delta p \sim \hbar \tag{2}$$

我们可以看出发散的真正原因是因为 $\Delta x \to 0$ 导致的 $\Delta p \to \infty$．为了解决发散问题，在弦理论中，不确定性关系有如下修正

$$\Delta x = \frac{\hbar}{\Delta p} + \alpha' \frac{\Delta p}{\hbar} \tag{3}$$

其中 α' 跟弦的张力之间的关系是

$$\alpha' = \frac{1}{2\pi T_s} \tag{4}$$

弦理论的最小距离单位是

$$x_{\min} \sim 2\sqrt{\alpha'} \tag{5}$$

弦论概述

- 本词条需要更多参考文献.
- 本词条缺少预备知识，初学者可能会遇到困难.

弦论的基本思想是：组成物质的基本单元不是粒子，而是弦．弦有开弦和闭弦两种，形状如下图所示： 弦的激发给出了各种不同的粒子．粒子在时空



图 1: 开弦.

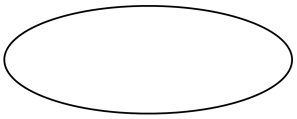


图 2: 闭弦.

中运动的轨迹叫做**世界线（world line）**．弦在时空中运动的轨迹叫做**世界面（world sheet）**．世界面由两个参数 σ 和 τ 来参数化．函数 $x^\mu(\tau,\sigma)$ 把世界面上的坐标 σ 和 τ 映射到时空坐标 x 上．

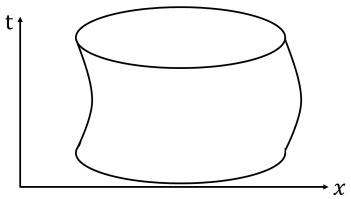


图 3: 世界面 $x^\mu(\tau,\sigma)$.

在弦论世界里面,组成物质的基本单元是长度大约是普朗克长度（$10^{-33}\,\mathrm{cm}$）的弦．像所有其他弦一样，这些基本弦能够振动．不同的振动频率形成了不同性质的粒子．对于一个自旋为 J 质量为 m_J 的粒子，质量和自旋通过如下式子

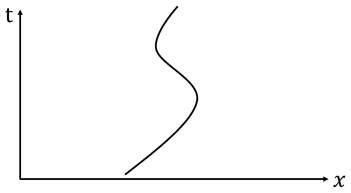


图 4: 世界线 $x^\mu(\tau)$.

联系起来

$$J = \alpha' m_J^2 \,. \tag{1}$$

弦也能分开和合并．比如弦 A 可以分开变成弦 B 和弦 C．这个过程对应于粒子衰变

$$A \to B + C \,. \tag{2}$$

弦论的种类

- 本词条需要更多参考文献.
- 本词条缺少预备知识，初学者可能会遇到困难.

五种弦论

弦论有五种．这五种弦论通过**对偶（duality）**联系在一起．这五种弦论分别是

1. 玻色弦论这个弦论里面只有玻色子．没有超对称．因为这种弦论里面没有费米子，所以这种弦论不能描述物质．这只是一个玩具理论．它包含了开弦，闭弦．需要 26 维的时空．

2. 第一型弦论这个版本的弦论包含了玻色子和费米子．粒子相互作用包含了超对称和规范群 SO（32）．这个理论有十维．

3. 第二 A 型弦论这个版本的弦论同样包含了超对称，开弦和闭弦．第二 A 型弦论的端点能够 attach 到高维物体（D-膜）上面．这个理论里面费米子不是 chiral 的．

4. 第二 B 型弦论跟二 A 型弦论差不多，只不过这种弦论里面的费米子是 chiral 的.

5. Heterotic 弦论包含了超对称，只允许闭弦. 规范群是 $E_8 \times E_8$. 左行和右行模式需要不同的时空维度. Heterotic 弦论有两种.

M-理论

不同的 duality 把这几种不同的弦论联系在了一起. 这些 duality 的名字分别是 S duality 和 T duality. 所以一个大胆的猜测是，在这五种弦论的背后，还有一个更加基本的理论：M 理论.

p-膜

p-膜的意思是，具有 p 个空间维度的物体. D-膜在弦论里面非常重要，因为开弦的端点能够 attach 到它上面. 比如 Yang-Mills 理论，就涉及到 attach 到 D-膜的开弦. 引力并不 attach 到 D-膜上. 这也就解释了为什么引力这样弱.

高维

这个 idea 首先是 Kaluza 和 Klein 首先提出来的. 在五维的空间，引力和电磁力可以很好地统一到同一个框架下. 他们提出了 compactification 这个 idea. 意思就是，当额外的维度足够小的时候，多余的维度就探测不到了.

BRST 量子化

弦理论的量子化方案有三种：协变量子化，光锥量子化和 BRST 量子化. 三种方案的优劣比较如下：

- 协变量子化：洛伦兹不变能明显表现出来. 有鬼. 时空维数是 26 维难以证明.
- 光锥量子化：洛伦兹不变不再明显. 没有鬼. 时空维数是 26 维容易证明.
- BRST 量子化：洛伦兹不变能明显表现出来. 有鬼. 时空维数是 26 维容易证明.

BRST 算符

首先我们来考虑李代数．考虑算符 K_i, 这些算符满足如下的李代数

$$[K_i, K_j] = f_{ij}{}^k K_k \,. \tag{1}$$

其中 $f_{ij}{}^k$ 被称作理论的结构常数．这些结构常数满足

$$f_{ij}{}^m f_{mk}{}^i + f_{jk}{}^m f_{mi}{}^l + f_{ki}{}^m f_{mj}{}^l = 0 \,. \tag{2}$$

现在我们引入两个鬼场，记作 b_i 和 c_j, 它们满足反对易关系

$$\{c^i, b_j\} = \delta^i_j \,. \tag{3}$$

我们现在回忆一下，一个场 $\phi(z, \bar{z})$ 在共形变换 $z \to w(z)$ 下具有如下变换的时候

$$\phi(z, \bar{z}) = \left(\frac{\partial w}{\partial z}\right)^h \left(\frac{\partial w}{\partial \bar{z}}\right)^{\bar{h}} \phi(w, \bar{w}) \,. \tag{4}$$

我们就说，这个场具有共形维度 $(h, \bar{h})$. b 和 c 场的共形维度分别是 2 和 -1. 我们现在来用 b 和 c 这两个鬼场以及 K_i 来构造两个算符．第一个是

$$Q = c^i K_i - \frac{1}{2} f_{ij}{}^k c^i c^j b_k \,. \tag{5}$$

我们假设 $Q = Q^\dagger$．Q 满足如下性质

$$Q^2 = 0 \,. \tag{6}$$

这个关系也可以写作 $\{Q, Q\} = 0$．我们把 BRST 的算符记作 Q 是为了暗示它是这个系统的守恒荷．我们把这叫做 BRST 荷．

第二个完全由鬼场组成的算符叫做鬼场算符 U．它的表达式如下

$$U = c^i b_i \,. \tag{7}$$

这个算符具有整数的特征值．如果李代数的维度是 n, 那么 U 的特征值就是 $0, \ldots, n$．如果一个态 $|\Psi\rangle$ 满足 $U|\Psi\rangle = m|\Psi\rangle$，我们就说这个态具有鬼数 m．Q 能够把鬼数提升 1.

$$U(Q|\Psi\rangle) = (1 + m)(Q|\Psi\rangle) \,. \tag{8}$$

BRST 不变态

一个态 $|\Psi\rangle$ 如果能被 BRST 算符湮灭掉，那这个态就被称为 BRST 不变态.

$$Q|\Psi\rangle = 0 \rightarrow |\Psi\rangle\text{是 BRST 不变的 .} \tag{9}$$

BRST 不变态是一个理论的物理态. 因为 $Q^2 = 0$, 我们可以得知任何一个满足如下条件的态 $|\Psi\rangle$ 都是 BRST 不变的

$$|\Psi\rangle = Q|\chi\rangle \neq 0 . \tag{10}$$

因为

$$Q|\Psi\rangle = Q^2|\chi\rangle = 0 . \tag{11}$$

我们把满足 $|\Psi\rangle = Q|\chi\rangle$ 的态叫做零态. 假设 $|\phi\rangle$ 是物理态，那么 $Q|\phi\rangle = 0$. 注意到

$$\langle\phi|\Psi\rangle = \langle\phi|(Q|\chi\rangle) = (\langle|Q)|\chi\rangle = 0 . \tag{12}$$

因此，一个物理态和一个零态之间的振幅总是为 0. 现在我们知道了，对态 $|\phi\rangle$ 加上一个零态 $|\Psi\rangle = Q|\chi\rangle$ 之后，仍然和 $|\phi\rangle$ 等价.

$$|\phi\rangle\text{等价于}|\phi\rangle + Q|\chi\rangle . \tag{13}$$

如果一个态 $|\Psi\rangle$ 具有鬼数 0，我们有

$$U|\Psi\rangle = 0 . \tag{14}$$

从式 7 以及式 14，我们可以得知 $b_k|\Psi\rangle = 0$. 一个态 $|\Psi\rangle$ 如果满足 $b_k|\Psi\rangle = 0$, 也就是说这个态被 b 湮灭，这样一个态不能被 c 湮灭，因为

$$U = \sum_i c^i b_i = \sum_i \delta^i_i - b_i c^i = n - \sum_i b_i c^i . \tag{15}$$

我们可以得知

$$U|\Psi\rangle = \left(n - \sum_i b_i c^i\right)|\Psi\rangle = n|\Psi\rangle - \sum_i b_i c^i|\Psi\rangle \tag{16}$$

对于鬼数为 0 的态我们有如下结论

$$Q|\Psi\rangle = \sum_i c^i K_i|\Psi\rangle . \tag{17}$$

这说明对于一个鬼数为零的态，我们有

$$K_i|\Psi\rangle = 0\,. \tag{18}$$

我们可以说：一个 BRST 不变而且具有鬼数 0 的态也会在生成元是 K_i 的对称性下保持不变．如果一个态具有鬼数 0，这告诉我们这个态不是一个鬼态，因此我们避免了负的概率．

弦理论里面的 BRST

我们现在选取共形规范 $h_{\alpha\beta} = \eta_{\alpha\beta}$. 在这个情形下，能量动量张量具有一个全纯的部分 $T_{zz}(z)$ 和一个反全纯的部分 $T_{\bar{z}\bar{z}}(\bar{z})$. $T_{zz}(z)$ 有如下展开式

$$T_{zz}(z) = \sum_{m=-\infty}^{\infty} \frac{L_m}{z^{m+2}}\,. \tag{19}$$

$T_{zz}(z)T_{ww}(w)$ 的算子积展开如下

$$T_{zz}(z)T_{ww}(w) = \frac{D/2}{(z-w)^4} - \frac{2T_{ww}(w)}{(z-w)^2} - \frac{\partial_w T_{ww}(w)}{z-w}\,. \tag{20}$$

为了简便起见，我们省略了因子 l_s^2．我们对鬼场进行如下定义

$$b(z)c(w) = \frac{1}{z-w}, \quad \bar{b}(\bar{z})\bar{c}(\bar{w}) = \frac{1}{\bar{z}-\bar{w}}\,. \tag{21}$$

接下来我们写下鬼场的能量动量张量 $T_{\rm gh}(z)$. 它的表达式如下

$$T_{\rm gh} = -2b(z)\partial_z c(z) - \partial_z b(z)c(z)\,. \tag{22}$$

利用鬼场的能量动量张量以及弦的能量动量张量，我们推出了如下的 BRST 流

$$j(z) = c(z)\left(T_{zz}(z) + \frac{1}{2}T_{\rm gh}(z)\right) = c(z)T_{zz}(z) + c(z)\partial_z c(z)b(z)\,. \tag{23}$$

BRST 荷由下式给定

$$Q = \int \frac{dz}{2\pi i} j(z)\,. \tag{24}$$

中心荷来自算子积展开式 20 的第一项，也就是

$$\frac{D/2}{(z-w)^4}\,. \tag{25}$$

这一项也叫做共形反常，因为它让代数不闭合．所以我们想要去掉这一项．我们考虑总的能量动量张量，它是弦的能量动量张量和鬼的能量动量张量之和．

$$T = T_{zz}(z) + T_{\rm gh}(z)\,. \tag{26}$$

鬼的能量动量张量如下

$$T_{\rm gh}(z)T_{\rm gh}(w) = \frac{-13}{(z-w)^4} - \frac{2T_{\rm gh}(w)}{(z-w)^2} - \frac{\partial_w T_{\rm gh}(w)}{z-w} \tag{27}$$

如果时空维度是 26 维，那么鬼场的能量动量张量的算子积展开式 27 的第一项能够把共形反常给消掉．这一结论是直接由 BRST 荷的 $Q^2 = 0$ 推导出来的．

BRST 变换

我们在光锥坐标系下，定义鬼场 c 和反鬼场 b. 其中 c 由 c^+,c^- 两部分组成．b 由 b_{++},b_{--} 两部分组成．我们引入鬼场的能量动量张量 $T^{\rm gh}_{\pm\pm}$

$$\begin{aligned} T^{\rm gh}_{++} &= i(2b_{++}\partial_+ c^+ + \partial_+ b_{++} c^+)\,. \\ T^{\rm gh}_{--} &= i(2b_{--}\partial_- c^- + \partial_- b_{--} c^-)\,. \end{aligned} \tag{28}$$

BRST 变换归纳如下

$$\begin{aligned} \delta X^\mu &= i\epsilon(c^+\partial_+ + c^-\partial_-)X^\mu\,. \\ \delta c^\pm &= \pm i\epsilon(c^+\partial_+ + c^-\partial_-)c^\pm\,. \\ \delta b_{\pm\pm} &= \pm i\epsilon(T_{\pm\pm} + T^{\rm gh}_{\pm\pm})\,. \end{aligned} \tag{29}$$

鬼场的作用量是

$$S_{\rm gh} = \int d^2\sigma(b_{++}\partial_- c^+ + b_{--}\partial_+ c^-)\,. \tag{30}$$

从作用量出发我们可以推出如下运动方程

$$\partial_- b_{++} = \partial_+ b_{--} = \partial_- c^+ = \partial_+ c^- = 0\,. \tag{31}$$

接下来我们写下鬼场的模式展开

$$\begin{aligned} c^+ &= \sum_n \bar{c}_n e^{-in(\tau+\sigma)}\,, \quad c^- = \sum_n c_n e^{-in(\tau-\sigma)}\,, \\ b_{++} &= \sum_n \bar{b}_n e^{-in(\tau+\sigma)}\,, \quad b_{--} = \sum_n b_n e^{-in(\tau-\sigma)} \end{aligned} \tag{32}$$

这些模式满足如下的反对易关系

$$\{b_m, c_n\} = \delta_{m+n,0} \, . \tag{33}$$

并且 $\{b_m, b_n\} = \{c_m, c_n\} = 0$. 鬼场的 Virasoro 算符是用这些模式定义出来的. 使用正规序展开，它们是

$$L_m^{\rm gh} = \sum_n (m-n) : b_{m+n} c_{-n} : \, , \quad \bar{L}_m^{\rm gh} = \sum_n (m-n) : \bar{b}_{m+n} \bar{c}_{-n} : \, . \tag{34}$$

总的 Virasoro 算符定义如下

$$L^{\rm tot} = L_m + L_m^{\rm gh} - a\delta_{m,0} \, . \tag{35}$$

我们可以证明，总的 Virasoro 算符满足下面的对易关系

$$[L_m^{\rm tot}, L_n^{\rm tot}] = (m-n) L_{m+n}^{\rm tot} + A(m) \delta_{m+n,0} \, . \tag{36}$$

右边的 $A(m)$ 项是 Anormaly．它的表达式如下

$$A(m) = \frac{D}{12} m(m^2-1) + \frac{1}{6}(m - 13m^3) + 2am \, . \tag{37}$$

为了让反常消失，我们取 $D = 26, a = 1$. BRST 流由下式给出

$$j = cT + \frac{1}{2} : cT^{\rm gh} : + \frac{3}{2} \partial^2 c \, . \tag{38}$$

BRST 荷由下面的模式展开给出

$$Q = \sum_n c_n L_{-n} + \frac{1}{2} \sum_{m,n} (m-n) : c_m c_n b_{-m-n} : - c_0 \, . \tag{39}$$

我们可以计算 Q^2, 结果如下

$$\begin{aligned} Q^2 &= \frac{1}{2} \sum ([L_m^{\rm tot}, L_n^{\rm tot}] - (m-n) L_{m+n}^{\rm tot}) c_{-m} c_{-n} \\ &\simeq \frac{1}{12}(D - 26) \, . \end{aligned} \tag{40}$$

$Q^2 = 0$ 要求时空维数必须是 26 维（$D = 26$）．回到原始的 BRST 方法，使用经典的 Virasoro 代数

$$[L_m, L_n] = (m-n) L_{m-n} \, . \tag{41}$$

我们可以得到结构常数的表达式如下

$$f_{mn}^{k} = (m-n)\delta_{m+n,k} \, . \tag{42}$$

现在我们来看弦理论里面物理的谱是如何构造出来的．我们先看开弦的例子．这些态是用鬼真空态构造出来的．我们把鬼真空态称作 $|\chi\rangle$. 这个态被所有的正的鬼模湮灭．如果 $n > 0$ 我们有

$$b_n|\chi\rangle = c_n|\chi\rangle = 0 \, . \tag{43}$$

鬼场的零模是一个特殊情况．它们可以用来构造理论的物理态．使用对易关系，我们可以推出零模满足如下方程

$$\{b_0, c_0\} = 1 \, . \tag{44}$$

由式 3 我们可知 $b_0^2 = c_0^2 = 0$. 我们也要求对于物理态 $|\Psi\rangle$，我们有 $b_0|\Psi\rangle = 0$. 现在，我们可以从鬼态的零模出发，构造一个两态的系统．这两个基矢被记作 $|\uparrow\rangle$ 和 $|\downarrow\rangle$．它们满足

$$\begin{aligned} &b_0|\downarrow\rangle = c_0|\uparrow\rangle = 0 \\ &b_0|\uparrow\rangle = |\downarrow\rangle \, , \quad c_0|\downarrow\rangle = |\uparrow\rangle \, . \end{aligned} \tag{45}$$

我们选取 $|\downarrow\rangle$ 作为鬼真空态．我们可以把这个态与动量态 $|k\rangle$ 做张量积，可以得到 $|\downarrow, k\rangle$. 为了得到物理的态，我们把 BRST 荷 Q 作用在上面．我们可以证明

$$Q|\downarrow, k\rangle = (L_0 - 1)c_0|\downarrow, k\rangle \, . \tag{46}$$

$Q|\downarrow, k\rangle = 0$ 的要求给出了在壳条件 $L_0 - 1 = 0$. 这个是快子态．第一激发态如下

$$|\Psi\rangle = (\varsigma \cdot \alpha_{-1} + \xi_1 c_{-1} + \xi_2 b_{-1})|\downarrow, k\rangle \, . \tag{47}$$

ξ_1 和 ξ_2 是常数，ς_μ 是一个有 26 个分量的矢量．因为第一激发态是无质量的．所以它有 24 个独立的分量．为了去掉多余的自由度，我们创造一个满足 $Q|\psi\rangle = 0$ 的物理态．我们可以证明

$$Q|\psi\rangle = 2(p^2 c_0 + (p \cdot \varsigma)c_{-1} + \xi_2 p \cdot \alpha_{-1})|\downarrow, k\rangle \, . \tag{48}$$

$Q|\psi\rangle = 0$ 对参数进行了一些限制．有 26 个正模的态和 2 个负模的态．我们可以引入一些约束来去掉负模的态．第一个约束是取 $p \cdot \varsigma = 0$ 以及 $\xi_1 = 0$. 这可以把负模的态去掉．同时 $p^2 = 0$, 这告诉我们这是一个无质量的态．我们有量纲零模的态

$$k_\mu \alpha_{-1}^\mu |\downarrow, k\rangle \ , \quad c_{-1}|\downarrow, k\rangle \ . \tag{49}$$

这些态跟物理的态是正交的．去掉它们给了我们一个 24 自由度的态．

无鬼定理

无鬼定理的叙述如下：如果时空维数是 $D = 26$，那么负模的态会被自动消去．

RNS 超弦

我们使用 Ramond-Neveu-Schwarz(RNS) 形式来修改玻色弦理论以引入费米子．这个方法在世界面上具有超对称．随后我们会使用具有时空超对称的 Green-Schwarz 形式．当时空维度是 10 的时候，这两个方案是等价的．

首先我们考虑共形规范下的 Polyakov 作用量．

$$S = -\frac{T}{2} \int d^2\sigma \partial_\alpha X^\mu \partial^\alpha X_\mu \ . \tag{1}$$

加入自由费米子 ψ 之后，作用量如下

$$S = -\frac{T}{2} \int d^2\sigma (\partial_\alpha X^\mu \partial^\alpha X_\mu - i\bar{\psi}^\mu \rho^\alpha \partial_\alpha \psi_\mu) \ . \tag{2}$$

ρ^α 是世界面上的狄拉克矩阵．因为世界面是 1+1 维的，所以 ρ^α 也是 1+1 维的狄拉克矩阵．有两个这样的矩阵

$$\rho^0 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \ , \quad \rho^1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \ . \tag{3}$$

Majorana 旋量

$\psi^\mu = \psi^\mu(\sigma, \tau)$ 是两分量 Majorana 旋量．我们把它记作

$$\psi = \begin{pmatrix} \psi_- \\ \psi_+ \end{pmatrix} \ . \tag{4}$$

在洛伦兹变换下，这些场按照矢量的规则变换．我们可以定义 $\bar{\psi}^\mu$

$$\bar{\psi}^\mu = (\psi^\dagger)^\mu \rho^0 \ . \tag{5}$$

我们再定义一个 ρ^3 矩阵如下

$$\rho^3 = \rho^0 \rho^1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ . \tag{6}$$

我们现在进行如下的定义

$$\begin{aligned} \sigma^\pm &= \tau \pm \sigma \ , \\ \partial_\pm &= \frac{1}{2}(\partial_\tau \pm \partial_\sigma) \ , \\ \partial_\tau &= \partial_+ + \partial_- \ , \quad \partial_\sigma = \partial_+ - \partial_- \ . \end{aligned} \tag{7}$$

经过计算，我们可以得出

$$\begin{aligned} S_{\rm F} &= -\frac{T}{2} \int d^2\sigma (-i\bar{\psi}^\mu \rho^\alpha \partial_\alpha \psi_\mu) \\ &= -\frac{T}{2} \int d^2\sigma (-2i)(\psi_- \cdot \partial_+ \psi_- + \psi_+ \cdot \partial_- \psi_+) \\ &= iT \int d^2\sigma (\psi_- \cdot \partial_+ \psi_- + \psi_+ \cdot \partial_- \psi_+) \ . \end{aligned} \tag{8}$$

从上面的作用量我们可以看出如下的运动方程

$$\partial_+ \psi_-^\mu = \partial_- \psi_+^\mu = 0 \ . \tag{9}$$

世界面上的超对称变换

我们现在引入超对称变换的参数 ϵ．这个参数也是一个 Majorana 旋量

$$\epsilon = \begin{pmatrix} \epsilon_- \\ \epsilon_+ \end{pmatrix} \ . \tag{10}$$

因为 ϵ 的组成部分被取成常数，这代表了世界面上的全局坐标．超对称变换有如下形式

$$\begin{aligned} \delta X^\mu &= \bar{\epsilon} \psi^\mu \ , \\ \delta \psi^\mu &= -i\rho^\alpha \partial_\alpha X^\mu \epsilon \ . \end{aligned} \tag{11}$$

我们的作用量在上面的超对称变换下保持不变. 这个变换让自由的玻色子变成费米子，也让自由的费米子变成玻色子. 按照分量来写的话，式 11 的第一个式子可以化简成如下形式

$$\delta X^\mu = \epsilon_- \psi_-^\mu + \epsilon_+ \psi_+^\mu \, . \tag{12}$$

式 11 的第二个式子可以化简成如下形式

$$\begin{aligned} \delta \psi_-^\mu &= -2\partial_- X^\mu \epsilon_+ \, , \\ \delta \psi_+^\mu &= 2\partial_+ X^\mu \epsilon_- \, . \end{aligned} \tag{13}$$

在超对称变换下，我们的拉式量式 2 按照下面进行变换

$$\delta L = -T[\partial_\alpha(\bar{\epsilon}\psi^\mu \partial^\alpha X_\mu) - \partial_\alpha \bar{\epsilon}(\rho^\beta \rho^\alpha \psi^\mu \partial_\beta X_\mu)] \, . \tag{14}$$

因为第一项是全导数项，我们忽略不计，我们可得守恒流是

$$J_\alpha^\mu = \frac{1}{2}\rho^\beta \rho_\alpha \psi^\mu \partial_\beta X_\mu \, . \tag{15}$$

能量动量张量

考虑世界面坐标的平移变换

$$\sigma^\alpha \to \sigma^\alpha + \epsilon^\alpha \, . \tag{16}$$

玻色场 X^μ 的变换如下

$$X^\mu \to X^\mu + \epsilon^\alpha \partial_\alpha X^\mu \, . \tag{17}$$

费米子场的变换如下

$$\psi^\mu \to \psi^\mu + \epsilon^\alpha \partial_\alpha \psi^\mu \, . \tag{18}$$

首先我们看拉式量里面的费米子部分，我们有

$$L_{\mathrm{F}} = -\frac{i}{2}\bar{\psi}^\mu \rho^\alpha \partial_\alpha \psi_\mu \, . \tag{19}$$

我们对这一项进行变分

$$\begin{aligned} \delta L_{\mathrm{F}} &= -\frac{i}{2}(\delta\bar{\psi}^\mu)\rho^\alpha \partial_\alpha \psi_\mu - \frac{i}{2}\bar{\psi}^\mu \rho^\alpha \partial_\alpha(\delta\psi_\mu) \\ &= -\frac{i}{2}(\epsilon^\beta \partial_\beta \bar{\psi}^\mu)\rho^\alpha \partial_\alpha \psi_\mu - \frac{i}{2}\bar{\psi}^\mu \rho^\alpha \partial_\alpha(\epsilon^\beta \partial_\beta \psi_\mu) \, . \end{aligned} \tag{20}$$

经过一番计算，我们得到

$$\delta L_{\mathrm{F}} = \partial_\alpha \epsilon^\beta \left(-\frac{i}{2} \bar{\psi}^\mu \rho^\alpha \partial_\beta \psi_\mu \right) . \tag{21}$$

我们需要进行对称化

$$\delta L_{\mathrm{F}} = \partial_\alpha \epsilon^\beta \left(-\frac{i}{4} \bar{\psi}^\mu \rho^\alpha \partial_\beta \psi_\mu - \frac{i}{4} \bar{\psi}^\mu \rho^\beta \partial_\alpha \psi_\mu \right) . \tag{22}$$

加上玻色部分，我们可以推出如下的能量动量张量

$$T_{\alpha\beta} = \partial_\alpha X^\mu \partial_\beta X_\mu + \frac{i}{4} \bar{\psi}^\mu \rho_\alpha \partial_\beta \psi_\mu + \frac{i}{4} \bar{\psi}^\mu \rho_\beta \partial_\alpha \psi_\mu - (\mathrm{Trace}) \tag{23}$$

减去 Trace 那部分是为了保证能量动量张量保持无迹．这是标度不变性的要求．使用光锥坐标，我们可以把能量动量张量的非零部分计算如下

$$T_{++} = \partial_+ X_\mu \partial_+ X^\mu + \frac{i}{2} \psi_+^\mu \partial_+ \psi_{+\mu} . \tag{24}$$

超对称的流是

$$J_+ = \psi_+^\mu \partial_+ X_\mu , \quad J_- = \psi_-^\mu \partial_- X_\mu . \tag{25}$$

费米子的运动方程是

$$\partial_+ \psi_-^\mu = \partial_- \psi_+^\mu = 0 . \tag{26}$$

玻色子的运动方程是

$$\partial_+ \partial_- X^\mu = 0 . \tag{27}$$

我们可以得到能量动量张量的守恒定律

$$\partial_- T_{++} = \partial_+ T_{--} = 0 . \tag{28}$$

模式展开和边界条件

我们回到费米子的作用量

$$S_{\mathrm{F}} = \int d^2\sigma (\psi_- \partial_+ \psi_- + \psi_+ \partial_- \psi_+) . \tag{29}$$

以第二项为例，我们对其进行变分，我们得到

$$\delta \int d^2\sigma \psi_+ \partial_- \psi_+ = \int d^2\sigma [\delta\psi_+ \partial_- \psi_+ + \psi_+ \partial_- (\delta\psi_+)] . \tag{30}$$

使用分部积分，我们有

$$\int d^2\sigma\psi_+\partial_-(\delta\psi_+) = \int_{-\infty}^{\infty} d\tau\psi_+\delta\psi_+\Big|_{\sigma=0}^{\sigma=\pi} - \int d^2\sigma\partial_-\psi_+\delta\psi_+ \,. \tag{31}$$

对另外一些项也进行类似处理，我们可以得到如下的边界项

$$\delta S_{\mathrm{F}} = \int_{-\infty}^{\infty} d\tau\{(\psi_+\delta\psi_+ - \psi_-\delta\psi_-)\Big|_{\sigma=\pi} - (\psi_+\delta\psi_+ - \psi_-\delta\psi_-)\Big|_{\sigma=0}\} \,. \tag{32}$$

开弦的边界条件

洛伦兹不变要求边界项必须为零．于是我们可以得到

$$\psi_+\delta\psi_+ - \psi_-\delta\psi_- = 0 \,. \tag{33}$$

在 $\sigma = 0$ 处我们可以取

$$\psi_\mu^+(0,\tau) = \psi_\mu^-(0,\tau) \,. \tag{34}$$

更一般地来说 $\psi_+ = \pm\psi_-$ 都可以让边界条件消失．我们一般选择式 34 作为 $\sigma = 0$ 处的边界条件．这让 $\sigma = \pi$ 处的边界条件 ambiguous．我们可以选取两种不同的边界条件

- $\psi_\mu^+(\pi,\tau) = \psi_\mu^-(\pi,\tau)$ (Ramond) 给出的弦的态是时空费米子
- $\psi_\mu^+(\pi,\tau) = -\psi_\mu^-(\pi,\tau)$ (Neveau-Schwarz) 给出的弦的态是时空玻色子

开弦的模展开

首先我们来考虑 R 部分．模式展开如下

$$\begin{aligned} \psi_-^\mu(\sigma,\tau) &= \frac{1}{\sqrt{2}}\sum_n d_n^\mu e^{-in(\tau-\sigma)} \,, \\ \psi_+^\mu(\sigma,\tau) &= \frac{1}{\sqrt{2}}\sum_n d_n^\mu e^{-in(\tau+\sigma)} \,. \end{aligned} \tag{35}$$

Majorana 条件是费米子是实的．这令我们取

$$d_{-n}^\mu = (d_n^\mu)^\dagger \,. \tag{36}$$

n 的取值是 $n = 0, \pm1, \pm2, \ldots$.

NS 部分具有不同的模式展开．展开如下

$$\begin{aligned}\psi_-^\mu(\sigma,\tau) &= \frac{1}{\sqrt{2}}\sum_r b_r^\mu e^{-ir(\tau-\sigma)}\ ,\\ \psi_+^\mu(\sigma,\tau) &= \frac{1}{\sqrt{2}}\sum_e b_r^\mu e^{-ir(\tau+\sigma)}\ .\end{aligned} \tag{37}$$

这里的 r 取值是

$$r = \pm\frac{1}{2}, \pm\frac{3}{2}, \pm\frac{5}{2}\ , \ldots . \tag{38}$$

闭弦的边界条件

闭弦有两种边界条件

- $\psi_\pm(\sigma,\tau) = \psi_\pm(\sigma+\pi,\tau)$ (周期性边界条件)
- $\psi_\pm(\sigma,\tau) = -\psi_\pm(\sigma+\pi,\tau)$ (反周期性边界条件)

闭弦的模式展开

上面的两种边界条件可以分别应用到左行波和右行波上面．模式展开是

$$\begin{aligned}\psi_+^\mu(\sigma,\tau) &= \sum_r \tilde{d}_r^\mu e^{-2ir(\tau+\sigma)}\ ,\\ \psi_-^\mu(\sigma,\tau) &= \sum_r \tilde{d}_r^\mu e^{-2ir(\tau-\sigma)}\ .\end{aligned} \tag{39}$$

如果我们选择 R 部分，那么

$$r = 0, \pm1, \pm2, \ldots . \tag{40}$$

从另一方面来讲，如果我们选择 NS 部分，那么

$$r = \pm\frac{1}{2}, \pm\frac{3}{2}, \ldots . \tag{41}$$

左行的部分和右行的部分是独立的．如果左行和右行的部分能够匹配上，我们就得到了时空玻色子．如果左行和右行的部分不能匹配上，我们就得到了时空费米子．

- 选取 NS 作为左行的部分，选取 NS 作为右行的部分给出时空玻色子．

- 选取 R 作为左行的部分，选取 R 作为右行的部分给出时空玻色子.
- 选取 NS 作为左行的部分，选取 R 作为右行的部分给出时空费米子.
- 选取 R 作为左行的部分，选取 NS 作为右行的部分给出时空费米子.

Super-Virasoro 生成元

我们把 Virasoro 算符进行扩展

$$L_m \to L_m^{(B)} + L_m^{(F)} \,. \tag{42}$$

我们使用下面的定义

$$L_m = \frac{1}{2}\sum_{-\infty}^{\infty} \alpha_{m-n}\cdot\alpha_n + \frac{1}{4}\sum_r (2r-m) b_{-r} b_{m+r} \,. \tag{43}$$

此外，我们还有从超对称流出来的第二个生成元．对于 NS，我们有

$$G_r = \frac{\sqrt{2}}{\pi}\int_{-\pi}^{\pi} d\sigma e^{ir\sigma} J_+ = \sum_{m=-\infty}^{\infty} \alpha_m \cdot b_{r+m} \,. \tag{44}$$

对于 R，我们有

$$F_m = \sum_n \alpha_{-n}\cdot d_{m+n} \,. \tag{45}$$

这里，m 和 n 都是整数．$r = \pm 1/2, \pm 3/2$.

正则量子化

首先我们来回顾玻色弦的正则量子化关系

$$[\alpha_m^\mu, \alpha_n^\nu] = m\delta_{m+n,0}\eta^{\mu\nu} \,. \tag{46}$$

对于闭弦来说，还有一个与之类似的与 $\bar{\alpha}$ 有关的关系．对于超弦理论，我们还需要补充费米子的对易关系

$$\{\psi_A^\mu(\sigma,\tau), \psi_B^\mu(\sigma',\tau)\} = \pi\eta^{\mu\nu}\delta_{AB}\delta(\sigma-\sigma') \,. \tag{47}$$

按照模式来写，我们有如下对易关系

$$\begin{aligned} \{b_r^\mu, b_r^\nu\} &= \eta^{\mu\nu}\delta_{r+s,0} \,, \\ \{d_m^\mu, d_n^\nu\} &= \eta^{\mu\nu}\delta_{m+n,0} \,. \end{aligned} \tag{48}$$

Super-Virasoro 代数

NS Sector

对于 NS sector，我们有如下代数

$$
\begin{aligned}
[L_n, L_m] &= (n-m)L_{n+m} + \frac{c}{12}(n^3 - n)\delta_{n+m,0} \,, \\
[L_n, G_r] &= \frac{1}{2}(n-2r)G_{n+r} \,, \\
\{G_r, G_s\} &= 2L_{r+s} + \frac{c}{12}(4r^2 - 1)\delta_{r+s,0} \,.
\end{aligned}
\tag{49}
$$

第四章

宇宙学

宇宙的演化

- 本词条需要更多参考文献.
- 本词条缺少预备知识，初学者可能会遇到困难.

宇宙演化可以大约分为两个时期，（详见表 1 [1]），从创世之初（$t=0$）到创世后 3 分钟我们可以称为前物质时期. 这个时期可以大致分为两个时期：从宇宙诞生后到 10^{-43} 秒，此时宇宙沐浴于 10^{18} GeV 的高温中，四大相互作用被统一在一起. 随后到 10^{-34} 秒宇宙在高温中经历**暴涨（Inflation）**，宇宙尺度因子迅速扩大（详见图 1 [2]）. 在暴涨快结束时超对称破缺，宇宙中的重元素开始产生.

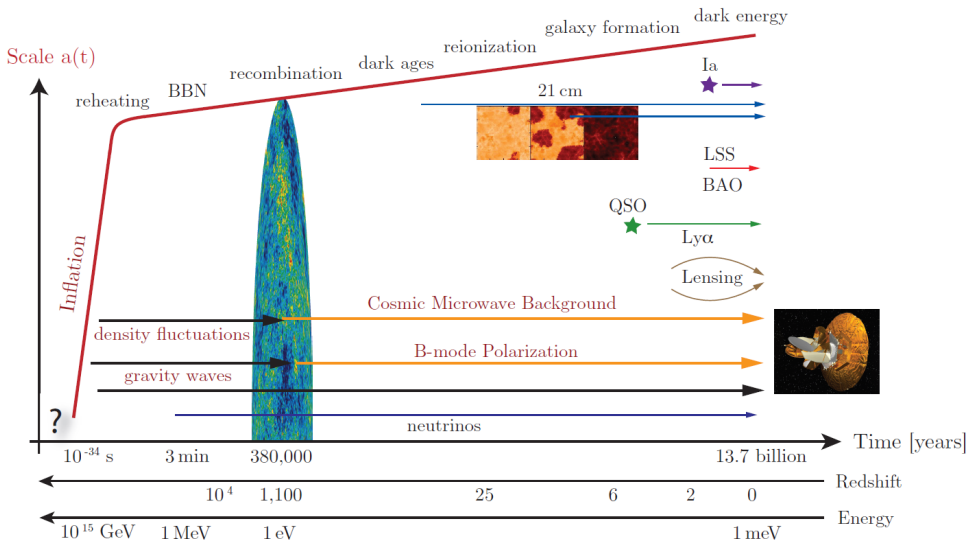


图 1: 宇宙尺度及成分表

从 10^{-10} 秒到宇宙诞生后 3 分钟，宇宙暴涨结束，宇宙从 1 TeV 迅速逐渐冷却到 0.1 MeV. 在此期间暴涨所释放的能量把宇宙重新加热，宇宙中的电弱相互作用开始分离，夸克结合成强子，中子冷却下来，宇宙中的核子开始在相互作用下结合并形成，称为**原核初合成** (BBN)；中微子在创世后 1 秒与其他核

[1] 翻译自 Daniel Baumann，TASI Lectures on Inflation, arXiv:0907.5424，1.2 节，Table1

[2] 引自 Daniel Baumann，TASI Lectures on Inflation, arXiv:0907.5424，1.2 节，Figure2

表 1: 宇宙大事记（问号表示理论尚未给出合理解释）

	时间	能量
普朗克时间奇点?	$< 10^{-43}$ s	10^{18} GeV
弦论尺度?	$> 10^{-43}$ s	$< 10^{18}$ GeV
超统一?	$\sim 10^{-36}$ s	10^{15} GeV
暴涨开始?	$> 10^{-34}$ s	$< 10^{15}$ GeV
超对称破缺?	$< 10^{-10}$ s	> 1 TeV
重元素产生?	$< 10^{-10}$ s	> 1 TeV
电弱统一时期	10^{-10} s	1 TeV
夸克-强子转换时期	10^{-4} s	10^{2} MeV
核子冷却	0.01 s	10 MeV
中微子退偶	1 s	1 MeV
大爆炸原核初合成	3 min	0.1 MeV
物质-辐射密度相等	10^{4} yrs	1 eV
再融合	10^{5} yrs	0.1 eV
宇宙黑暗时期	$10^{5} - 10^{8}$ yrs	
再电离	10^{8} yrs	
星系形成	$\sim 6 \times 10^{8}$ yrs	
暗能量	$\sim 10^{9}$ yrs	
太阳系形成	$\sim 8 \times 10^{9}$ yrs	

子退耦并此后不再与其他物质产生相互作用，一直传播到现在．此时，宇宙开始以辐射为主导，我们也称为辐射主导时期的开始，此外，宇宙中的原初扰动和引力波开始形成．

从创世后 3 分钟到现在，我们可以称为后物质时期．我们也可以把这个时期大致地分为两个时期：从宇宙诞生后 3 分钟到 380.000 年，宇宙一直以辐射为主导（详见图 2 [3]）．随着宇宙的膨胀光子和物质从密度对等逐渐分离，且光子、物质和原初引力波进行充分的相互作用称为**再融合** (recombination)，光子逐渐冷却下来形成**宇宙辐射背景** (CMB)，期间暗物质开始形成．

从宇宙诞生后 380.000 年到现在，宇宙进入物质主导时期，另一部分光子形成了宇宙辐射背景并随着时间演化到现在，一部分光子和原初引力波相互作

[3] 引自 Daniel Baumann，Cosmology，Perface，Figure 1

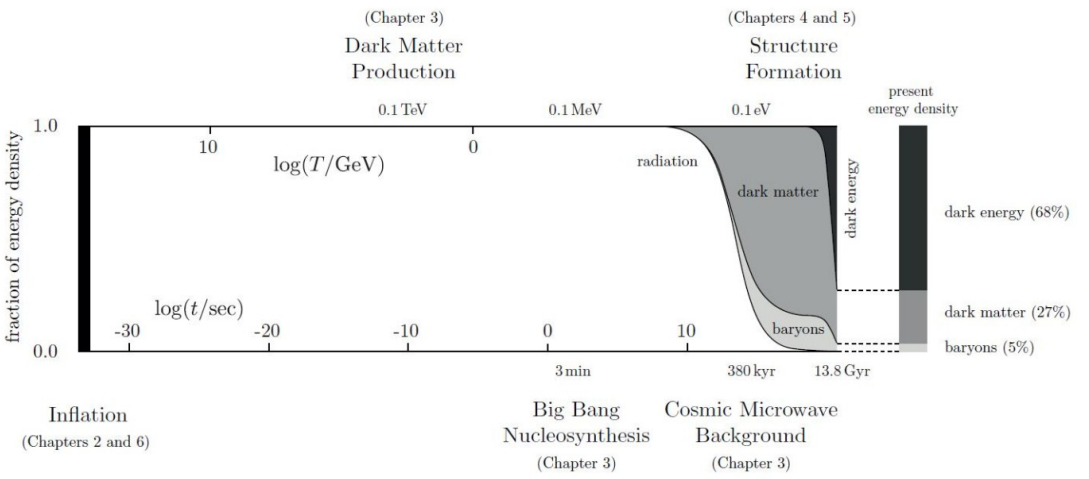


图 2: 宇宙成分变化

用后成为 B-模极化光子; 一部分原初引力波和中微子并不参与到相互作用中，并随着时间演化直到现在．在此期间，各大星系形成于创世后 10^8 年之后，太阳系形成于 8×10^9 年．

宇宙学红移

- 本词条缺少预备知识，初学者可能会遇到困难．

预备知识 FRW 度规 [1659]

由于物体在宇宙中传播的过程中，宇宙也在加速膨胀，所以为了准确的测量物体在宇宙传播过程中的物理量，我们需要引进宇宙学红移的概念．

光子的红移

从量子力学的描述中，光子的波长可以定义为 $\lambda = \frac{h}{p}$．当光在 t_1 时刻发射以波长 λ_1 发射，于 t_0 时刻被接收所观测到的波长为

$$\lambda_0 = \frac{a(t_0)}{a(t_1)}\lambda_1, \tag{1}$$

其中 $a(t)$ 为 t 时刻的宇宙尺度因子．因为宇宙在加速膨胀 $a(t_0) > a(t_1)$, 所以容易得 $\lambda_0 > \lambda_1$．

红移因子

为了计算方便，我们可以通过定义从星系发出的，经过一段时间到达地球后被观测的光的红移为**红移因子（redshift parametre）**

未完成：补充及引用“光的多普勒效应”词条

$$z = \frac{\lambda_0 - \lambda_1}{\lambda_1}, \tag{2}$$

显然从式 1 我们可以推出

$$1 + z = \frac{a(t_0)}{a(t)} = \frac{1}{a(t_1)}. \tag{3}$$

一般地我们设现在的宇宙尺度因子 $a(t_0) = 1$.

哈勃常数

我们把 t_1 时刻的宇宙尺度因子 $a(t_1)$ 以现在的时刻 t_0 为原点作泰勒展开，可得

$$a(t_1) = a(t_0)(1 + (t - t_0)H_0 + \cdots) \tag{4}$$

从式 3 我们可以看出 $z = H_0(t_0 - t_1)$，这里我们定义了**哈勃常数（Hubble Constant）** H_0. 显然我们可以发现红移参量与光走过的距离 $d = c(t_0 - t_1)$ 成正比

$$z \simeq \frac{H_0 d}{c}. \tag{5}$$

哈勃常数常常被定义为以下数值

$$H_0 \equiv 100h\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}, \tag{6}$$

其中 h 测得的数值为[4]

$$h \sim 0.67 \pm 0.01. \tag{7}$$

[4] 12Planck 2013 Results –Cosmological Parameters [arXiv:1303.5076]

宇宙中的距离

预备知识 FRW 度规[1659]，宇宙学红移

非可观测距离

为了有利于定义可观测距离，我们首先定义非观测距离．也称为度规距离．FRW 度规可以写成以下形式

$$ds^2 = -dt^2 + a(t)^2(d\chi^2 + S_k^2(\chi)d\Omega^2), \tag{1}$$

其中

未完成：[xxx]

度规距离可以被定义为空间度规中立体角的因子

$$d_m = S_k(\chi). \tag{2}$$

当取平直宇宙的情况，即 $k = 0$ 时，度规距离可退化为**共动距离（comoving distance）** χ. 共动距离可以用宇宙学红移因子 z 和哈勃常数 $H(z)$ 来表达，具体可写成

$$\chi(z) = \int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_0^z \frac{dz}{H(z)}. \tag{3}$$

这里需要强调，度规距离和共动距离都是非可观测距离．

光度距离

由于 Type IA 超新星被认为是一个拥有绝对光度 L 的星体（每秒发出恒定的能量）, 所以此类星体被称作**标准烛光（standard candle）**．Type IA 超新星爆发的光通过宇宙传到地球后，地球上观测到的能流密度 F(单位面积每秒所观测到的能量) 可以用于测量距离．我们先假设某一静止 Type IA 超新星，能流 F 与此超新性的共动距离 χ 的关系为

$$F = \frac{L}{4\pi\chi^2}. \tag{4}$$

然而，由于宇宙的膨胀，在 FRW 度规中，上述关系可以被以下三个原因所修正：

1、在时间 t_0 时刻超新星的光到达地球，超新星的光穿过地球的固有面积为 $4\pi d_m^2$;

2、在膨胀的宇宙中，以地球作为参照系，超新星在往后退，所以要除以一个宇宙学红移因子 $1+z$;

3、相对于超新星，地球也在往后退，所以地球接收到的光子也会产生效应，因而需要再次除以一个宇宙学红移因子 $1+z$

所以，能流 F 与距离的正确关系应为

$$F=\frac{L}{4\pi d_m(1+z)^2}. \tag{5}$$

角距离

未完成：内容

Friedmann-Robertson-Walker (FRW) 度规

预备知识 度规，张量分析

在宇宙学中，我们常用 **Friedmann-Robertson-Walker (FRW)** 度规，其一般形式为

$$ds^2=g_{\mu\nu}dx^\mu dx^\nu=-dt^2+a(t)^2\Big(\frac{1}{1-kr^2}+r^2d\Omega^2\Big), \tag{1}$$

其中 t 为时间坐标，r 为空间某一点到原点的距离，k 对应空间的弯曲性质：当 $k=1$ 时，空间为三维球；$k=0$ 时空间为平直空间；当 $k=-1$ 时，空间为双曲空间．Ω 为 S^2 球面．$a(t)$ 称为**尺度因子（scalar factor）**.

第五章

宇宙学扰动

标量扰动

- 本词条需要更多参考文献.
- 本词条缺少预备知识，初学者可能会遇到困难.

把爱因斯坦方程进行扰动，我们可以得到

$$\delta G^{\mu}_{\nu} = 8\pi G \delta T^{\mu}_{\nu} \,. \tag{1}$$

我们主要关心标量的扰动，可以得到

$$\begin{aligned}
\delta G^0_0 &= 2a^{-2}[\nabla^2\Phi - 3\mathcal{H}(\Phi' - \mathcal{H}\Psi)] \,, \\
\delta G^i_0 &= -2a^{-2}\partial^i(\Phi' - \mathcal{H}\Psi) \,, \\
\delta G^i_j &= 2a^{-2}\left[(\mathcal{H}^2 + 2\mathcal{H}')\Psi + \mathcal{H}\Psi' - \Phi'' - 2\mathcal{H}\Phi' + \frac{1}{3}\nabla^2(\Phi+\Psi)\right]\delta^i_j \\
&\quad - a^{-2}\left(\partial^i\partial_j - \frac{1}{3}\delta^i_j\nabla^2\right)(\Phi+\Psi) \,.
\end{aligned} \tag{2}$$

对能量动量张量同样进行扰动，我们有

$$\begin{aligned}
\delta T^0_0 &= -\delta\rho \,, \\
\delta T^i_0 &= -(\bar{\rho} + \bar{p})\partial^i v \,, \\
\delta T^i_j &= \delta p \delta^i_j + \left(\partial^i\partial_j - \frac{1}{3}\nabla^2\right)\sigma \,,
\end{aligned} \tag{3}$$

使用上述方程，式 1 的 00 分量的方程可以写成

$$\nabla^2\Phi - 3\mathcal{H}(\Phi' - \mathcal{H}\Psi) = -4\pi G a^2 \delta\rho \,. \tag{4}$$

上述方程可以视作平坦空间的泊松方程在弯曲空间中的推广．定义物理距离 $\mathrm{r} = a\mathrm{x}$ 和物理时间 t, 我们可以把上面方程右边的 a^2 吸收掉．最终我们可以得到如下方程

$$\nabla^2_{\mathrm{r}}\Phi - 3H(\dot{\Phi} - H\Psi) = -4\pi G\delta\rho \tag{5}$$

现在我们来看 $(i,0)$ 方程

$$\Phi' - \mathcal{H}\Psi = 4\pi G a^2(\bar{\rho} + \bar{p})v \,. \tag{6}$$

综合式 4 和式 6 我们可得

$$\nabla^2\phi = -4\pi G a^2[\delta\rho - 3\mathcal{H}(\bar{\rho}+\bar{p})v] \,. \tag{7}$$

等式的右边实际上是规范不变的密度扰动 $(\delta\rho)_*$. 我们可以把上式写成

$$\nabla^2\Phi = -4\pi G a^2\bar{\rho}\delta_* \,, \tag{8}$$

其中 $\delta_* = (\delta\rho)_*/\bar{\rho}$ 是 Gauge-invariant density contrast.

张量扰动

- 本词条需要更多参考文献.
- 本词条缺少预备知识，初学者可能会遇到困难.

对于在视界内部的模式来说，张量扰动对应了在 FRW 背景下传播的引力波.

宇宙演化

在共形牛顿规范下，我们只保留 h_{ij}^{TT}, 可得

$$ds^2 = a^2[-d\eta^2 + (\delta_{ij} + h_{ij}^{TT})dx^i dx^j] \,. \tag{1}$$

对于爱因斯坦张量，我们有 $\delta G_0^0 = 0, \delta G_0^i = 0$ 以及

$$\delta G_j^i = \frac{1}{2a^2}[(h_{ij}^{TT})'' + 2\mathcal{H}(h_{ij}^{TT})' - \nabla^2 h_{ij}^{TT}] \,. \tag{2}$$

于是，扰动的爱因斯坦方程可以写成如下形式

$$(h_{ij}^{TT})'' + 2\mathcal{H}(h_{ij}^{TT})' - \nabla^2 h_{ij}^{TT} = 16\pi G a^2 \sigma_{ij}^{TT} \,. \tag{3}$$

我们可以换到动量空间然后以极化张量为基进行如下展开

$$\tilde{h}_{ij}^{TT}(\eta, \mathrm{k}) = \sum_{A=+,\times} e_{ij}^A(\hat{\mathrm{k}})\tilde{h}_A(\eta, \mathrm{k}) \,, \tag{4}$$

类似地，我们有

$$\tilde{\sigma}_{ij}^{TT}(\eta,\mathrm{k}) = \sum_{A=+,\times} e_{ij}^{A}(\hat{\mathrm{k}})\tilde{\sigma}_{A}(\eta,\mathrm{k})\,. \tag{5}$$

极化张量的定义如下

$$\begin{aligned} e_{ij}^{+}(\hat{\mathrm{k}}) &= \hat{\mathrm{u}}_i\hat{\mathrm{u}}_j - \hat{\mathrm{v}}_i\hat{\mathrm{v}}_j\,, \\ e_{ij}^{\times}(\hat{\mathrm{k}}) &= \hat{\mathrm{u}}_i\hat{\mathrm{v}}_j + \hat{\mathrm{v}}_i\hat{\mathrm{u}}_j\,, \end{aligned} \tag{6}$$

极化张量按照下式归一化

$$e_{ij}^{A}(\hat{\mathrm{k}})e_{ij}^{A'}(\hat{\mathrm{k}}) = 2\delta^{AA'}\,. \tag{7}$$

如果 $\hat{\mathrm{k}}$ 是沿着 $\hat{z}$ 方向的，我们可以选取 $\hat{\mathrm{u}}=\hat{\mathrm{x}}$ 以及 $\hat{\mathrm{v}}=\hat{\mathrm{y}}$ 于是我们有

$$e_{ab}^{+} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad e_{ab}^{\times} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\,. \tag{8}$$

于是式 3 变成了两个关于 $h_A(\eta,k)$ 的独立方程

$$\tilde{h}_A'' + 2\mathcal{H}\tilde{h}_A' + k^2\tilde{h}_A = 16\pi G a^2\tilde{\sigma}_A\,. \tag{9}$$

因此

$$\tilde{h}_A(\eta,\mathrm{k}) \propto \frac{1}{a(\eta)}\sin(k\eta+\alpha) \quad (k\eta \ll 1)\,. \tag{10}$$

假如

$$h_A(\eta,\mathrm{k}) \propto \frac{1}{a(\eta)}\sin(k\eta+\alpha)\,. \tag{11}$$

我们就有

$$h_A'(\eta,\mathrm{k}) \propto \frac{k\cos(k\eta+\alpha)}{a(\eta)} + O\left(\frac{1}{a^2}\right)\,, \tag{12}$$

于是

$$\dot{h}_A(\eta,\mathrm{k}) \propto \frac{k\cos(k\eta+\alpha)}{a^2(\eta)} + O\left(\frac{1}{a^3}\right)\,. \tag{13}$$

因为引力波的强度正比于 $\sum_A\langle\dot{h}_A^2\rangle$. 所以我们得出了 $\rho_{\mathrm{gw}} \propto a^{-4}$. 需要注意的是，只有张量在视界内的时候，它们才能够被引力子描述，也就是一些能量和动量都是能被很好定义的，随着宇宙的膨胀，能量密度按照 $1/a^4$ 变化.

对于超出视界的模式，我们有

$$\tilde{h}'' + \frac{2}{\eta}\tilde{h}' \simeq 0\,, \tag{14}$$

我们把初始条件 $\eta_{\rm in}$ 设在辐射为主导的时期，初始条件设成 $\tilde{h}(\eta_{\rm in},k)=\tilde{h}_{\rm in}(k)$ 以及 $\tilde{h}'(\eta_{\rm in},k)=0$. 由于式 9 是一个线性方程，不同动量的模式不会混合. 所以研究一个特定的模式如何演化我们可以简单地设置 $\tilde{h}_{\rm in}(k)=1$，也就是说，我们研究初始值为 1 下 $\tilde{h}(\eta,k)$ 的演化.

为了方便数值积分，我们定义无量纲量 $x=\log a$ 以及 $\hat{k}=k/\mathcal{H}$ 于是运动方程变为了如下形式

$$\partial_x^2\tilde{h}+(3+\xi)\partial_x\tilde{h}+\hat{k}^2\tilde{h}=0\,, \tag{15}$$

我们在模拟中使用了这样的参数 $\Omega_{\rm M}=0.3$ 以及 $\Omega_\Lambda=1-\Omega_{\rm M}-\Omega_{\rm R}\simeq 0.7$. 数

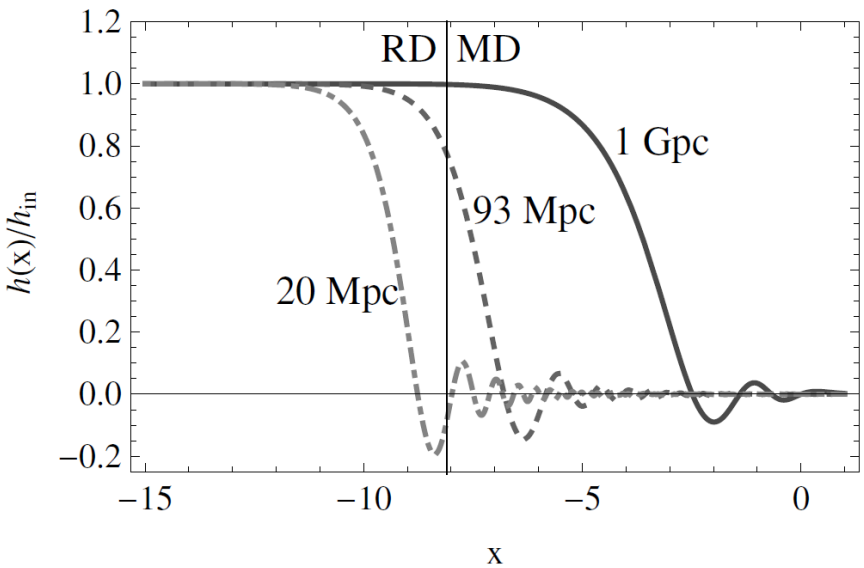


图 1: $\tilde{h}(\eta,k)$ 的演化. 我们把初始条件设为 1. 横坐标是 $x=\log a(\eta)$.

值模拟结果如图 1 所示.

$\tilde{h}(\eta,k)$ 在视界外（$k\eta\ll 1$）总是守恒的. 进入世界之后，不管是在物质为主导时期还是辐射为主导的时期，它都会迅速衰变成 0. 到了退耦和时期，也就是红移 $z_{\rm dec}\simeq 1090$, $x_{\rm dec}\simeq -7.0$, 所有的短波长引力波几乎都消失了，只有长波长的引力波仍然存在. 因此，引力波能影响大尺度上的微波背景辐射而影响不到小尺度上的微波背景辐射.

辐射为主的时期的解析解

辐射为主的时期的引力波运动方程为

$$\tilde{h}'' + \frac{2}{\eta}\tilde{h}' + k^2\tilde{h} = 0\,, \tag{16}$$

这个方程有两个解 $\sin(k\eta)/(k\eta)$ 和 $\cos(k\eta)/(k\eta)$. 加上边界条件 $\tilde{h}(\eta_{\rm in}, k) = \tilde{h}_{\rm in}(k)$ 以及 $\tilde{h}'(\eta_{\rm in}, k) = 0$. 边界条件确定了只有第一个解是可行的，所以我们有

$$\tilde{h}(\eta, k) \simeq \tilde{h}_{\rm in}\frac{\sin(k\eta)}{k\eta}\,. \tag{17}$$

物质为主的时期的解析解

在物质为主的时期，我们有 $\mathcal{H} = 2/\eta$, 运动方程为

$$\tilde{h}'' + \frac{4}{\eta}\tilde{h}' + k^2\tilde{h} = 0\,. \tag{18}$$

这个方程可以严格解，两个独立的解是

$$\begin{aligned} h_1(\eta, k) &= \frac{1}{(k\eta)^2}\left[\frac{\sin(k\eta)}{k\eta} - \cos(k\eta)\right]\,, \\ h_2(\eta, k) &= \frac{1}{(k\eta)^2}\left[\frac{\cos(k\eta)}{k\eta} + \sin(k\eta)\right]\,. \end{aligned} \tag{19}$$

我们考虑一个在物质-辐射平衡时期在视界外的模式，即 $k\eta_{\rm eq} \ll 1$，在辐射为主的时期应为常数解. 因为 $h_1(\eta)$ 满足 $h_1(0) = 1/3$ 以及 $h'(0) = 0$, 我们得到，在物质为主时期的长波长模式大约是下式

$$\tilde{h}(\eta, k) \simeq \frac{3\tilde{h}_{\rm in}(k)}{(k\tau)^2}\left[\frac{\sin(k\eta)}{k\eta} - \cos(k\eta)\right], \quad k\eta_{\rm eq} \ll 1\,. \tag{20}$$

宇宙学常数为主的时期的解析解

宇宙学常数为主的时期，$\tilde{h}$ 满足的运动方程为

$$\tilde{h}'' - \frac{2}{\eta}\tilde{h}' + k^2\tilde{h} = 0\,. \tag{21}$$

最一般的解为

$$\begin{aligned} \tilde{h}(\eta, k) = \tilde{h}_{\rm in}\Big\{&\alpha_k[\cos(k\eta) + k\eta\sin(k\eta)] \\ &+\beta_k[\sin(k\eta) - k\eta\cos(k\eta)]\Big\}\,, \end{aligned} \tag{22}$$

第六章

引力波

引力波的几何描述

- 本词条需要更多参考文献.
- 本词条缺少预备知识，初学者可能会遇到困难.

引力的作用量分为爱因斯坦作用量和物质的作用量 $S = S_E + S_M$ 其中

$$S_E = \frac{c^3}{16\pi G}\int d^4x\sqrt{-g}R\,. \tag{1}$$

能量动量张量的定义如下

$$\delta S_M = \frac{1}{2c}\int d^4x\sqrt{-g}T^{\mu\nu}\delta g_{\mu\nu}\,. \tag{2}$$

对作用量求变分，我们得到了如下的爱因斯坦方程

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu}\,. \tag{3}$$

广义相对论在很大的对称群下都是不变的，比如下面的坐标变换

$$x^\mu \to x'^\mu(x)\,. \tag{4}$$

在上面的坐标变换下，度规按照如下规则变换

$$g_{\mu\nu}(x) \to g'_{\mu\nu} = \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\nu}g_{\rho\sigma}(x)\,. \tag{5}$$

我们把这种对称性叫做广义相对论的规范对称性. 现在我们把度规按照如下的方式展开

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}\,,\quad |h_{\mu\nu}| \ll 1\,. \tag{6}$$

考虑如下的平移坐标变换

$$x^\mu \to x'^\mu = x^\mu + \xi^\mu(x)\,. \tag{7}$$

在这样的坐标变换下，$h_{\mu\nu}$ 按照如下规则变换

$$h_{\mu\nu}(x) \to h'_{\mu\nu}(x) - (\partial_\mu\xi_\nu + \partial_\nu\xi_\mu)\,. \tag{8}$$

现在考虑如下的洛伦兹变换

$$x^\mu \to \Lambda^\mu_\nu x^\nu \ . \tag{9}$$

矩阵 Λ^μ_ν 满足

$$\Lambda^\rho_\mu \Lambda^\sigma_\nu \eta_{\rho\sigma} = \eta_{\mu\nu} \ . \tag{10}$$

在洛伦兹变换下，矩阵 $g_{\mu\nu}$ 的变换如下

$$\begin{aligned} g_{\mu\nu} \to g'_{\mu\nu}(x') &= \Lambda^\rho_\mu \Lambda^\sigma_\mu g_{\rho\sigma}(x) \\ &= \Lambda^\rho_\mu \Lambda^\sigma_\nu [\eta_{\rho\sigma} + h_{\rho\sigma}(x)] \\ &= \eta_{\mu\nu} + \Lambda^\rho_\mu \Lambda^\sigma_\nu h_{\rho\sigma}(x) \ . \end{aligned} \tag{11}$$

因此，在洛伦兹变换下，$h_{\mu\nu}$ 的变换规则如下

$$h'_{\mu\nu}(x) = \Lambda^\rho_\mu \Lambda^\sigma_\nu h_{\rho\sigma}(x) \ . \tag{12}$$

这说明 $h_{\mu\nu}$ 在洛伦兹变换下是张量．转动变换无论如何不会破坏 $|h_{\mu\nu}| \ll 1$ 这个条件．但是对于 boost 我们必须只做总是使得 $h_{\mu\nu} \ll 1$ 能够满足的 boost 的变换．这样的理论叫做线性理论．另外，值得注意的有以下几点

- $h_{\mu\nu}$ 在有限的平移变换 $x^\mu \to x'^\mu = x^\mu + a^\mu$ 和洛伦兹变换 $x^\mu \to \Lambda^\mu_\nu x^\nu$ 下不变．
- $h_{\mu\nu}$ 在无穷小的局域变换 $x^\mu \to x'^\mu = x^\mu + \xi^\mu(x)$ 下不变．

与此形成鲜明对比的是，广义相对论没有庞加莱对称性．这是因为在广义相对论的框架下，平坦空间并没有什么特殊性，而庞加莱对称群是平坦空间里面的对称群．但是，广义相对论拥有全部的坐标变换的不变性，而不仅仅是无穷小的坐标变换的不变性．在线性阶，黎曼张量是

$$R_{\mu\nu\rho\sigma} = \frac{1}{2}(\partial_\nu \partial_\rho h_{\mu\sigma} + \partial_\mu \partial_\sigma h_{\nu\rho} - \partial_\mu \partial_\rho h_{\nu\sigma} - \partial_\nu \partial_\sigma h_{\mu\rho}) \ . \tag{13}$$

在我们感兴趣的物理情形下，总是存在一个参考系使得式 6 成立．我们定义如下的两个量

$$h = \eta^{\mu\nu} h_{\mu\nu} \ , \quad \bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h \ . \tag{14}$$

同样的，我们可以定义另外一个量

$$\bar{h} \equiv \eta^{\mu\nu} \bar{h}_{\mu\nu} = h - 2h = -h \ . \tag{15}$$

我们可以反解出 $h_{\mu\nu}$ 关于 $\bar{h}_{\mu\nu}$ 的表达式.

$$h_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\bar{h} \,. \tag{16}$$

线性化的爱因斯坦方程如下

$$\Box\bar{h}_{\mu\nu} + \eta_{\mu\nu}\partial^\rho\partial^\sigma\bar{h}_{\rho\sigma} - \partial^\rho\partial_\nu\bar{h}_{\mu\rho} - \partial^\rho\partial_\mu\bar{h}_{\nu\rho} = -\frac{16\pi G}{c^4}T_{\mu\nu} \,. \tag{17}$$

现在我们选取如下的洛伦兹规范（也叫做希尔伯特规范，谐振子规范，或者是 De Donder 规范）

$$\partial^\nu\bar{h}_{\mu\nu} = 0 \,. \tag{18}$$

在 $x^\mu \to x'^\mu + \xi^\mu(x)$ 变换下，$\bar{h}_{\mu\nu}$ 的变换如下

$$\bar{h}_{\mu\nu} \to \bar{h}'_{\mu\nu} = \bar{h}_{\mu\nu} - (\partial_\mu\xi_\nu + \partial_\nu\xi_\mu - \eta_{\mu\nu}\partial_\rho\xi^\rho) \,. \tag{19}$$

$\partial^\nu h_{\mu\nu}$ 的变换如下

$$\partial^\nu\bar{h}_{\mu\nu} \to (\partial^\nu\bar{h}_{\mu\nu})' = \partial^\nu\bar{h}_{\mu\nu} - \Box\xi_\mu \,, \tag{20}$$

如果我们需要让 $\partial^\nu\bar{h}_{\mu\nu} = 0$, 假设初始的场满足 $\partial^\nu\bar{h}_{\mu\nu} = f_\mu(x)$. 那我们必须选择满足下面条件的 ξ_μ

$$\Box\xi_\mu = f_\mu(x) \,. \tag{21}$$

这个方程总是有解的．如果我们用 $G(x)$ 来表示达朗贝尔算符的格林函数，那么它将满足

$$\Box G(x-y) = \delta^4(x-y) \,. \tag{22}$$

$\xi_\mu(x)$ 的解为

$$\xi_\mu(x) = \int d^4x G(x-y) f_\mu(y) \,. \tag{23}$$

在这样的规范条件下，式 17 的左边最后三项都为零．我们得到了一个简单的波动方程

$$\Box\bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4}T_{\mu\nu} \,. \tag{24}$$

综合式 18 和式 24 我们得到如下的方程

$$\partial^\nu T_{\mu\nu} = 0 \,. \tag{25}$$

这个方程就是线性理论的能动量守恒方程. 完整的爱因斯坦理论里面的能动量守恒方程是

$$D^\nu T_{\mu\nu} = 0\,. \tag{26}$$

线性理论里面的假设有：作为引力波源的物体一般都被近似为在平坦空间里面运动. 比如互相绕转的双星系统，他们的背景被看作 $\eta_{\mu\nu}$. 这说明我们在用牛顿引力来描述这个系统，而不是完整的爱因斯坦理论.

TT 规范

- 本词条需要更多参考文献.
- 本词条缺少预备知识，初学者可能会遇到困难.

在源的外面，我们有 $T_{\mu\nu} = 0$，所以我们有如下方程

$$\Box \bar{h}_{\mu\nu} = 0\,. \tag{1}$$

其中 $\Box = -(1/c^2)\partial_0^2 + \nabla^2$. 这个方程说明了引力波以光速传播. 因为式 18 [1667] 没有完全确定规范，在没有源的地方我们能够极大地固定规范.

如果 $\Box \xi_\mu = 0$，那么变换 $x^\mu \to x^\mu + \xi^\mu$ 保持 $\partial^\nu \bar{h}_{\mu\nu}$ 不变. 从 $\Box \xi_\mu = 0$ 可以推出

$$\Box \xi_{\mu\nu} = 0\,, \quad \xi_{\mu\nu} \equiv \partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \eta_{\mu\nu}\partial_\rho \xi^\rho\,. \tag{2}$$

我们可以使用 TT 规范

$$h^{0\mu} = 0\,, \quad h^i_i = 0\,, \quad \partial^j j_{ij} = 0\,. \tag{3}$$

加上洛伦兹条件把十个自由度的对称矩阵 $h_{\mu\nu}$ 变为六个自由度. 剩余的四个满足 $\Box \xi_\mu = 0$ 的 ξ^μ 把它降到了两个自由度. 我们把 TT 规范下的度规记作 h^{TT}_{ij}.

注意，TT 规范在含有源的时候是不能取的，这是因为 $\Box \bar{h}_{\mu\nu} \neq 0$. 式 1 有如下的平面波解

$$h^{TT}_{ij}(x) = e_{ij}(\mathrm{k})\mathrm{e}^{\mathrm{i}kx}\,. \tag{4}$$

其中 $k^\mu=(\omega/c,\mathrm{k}),\omega/c=|\mathrm{k}|$．$e_{ij}(\mathrm{k})$ 被称作极化张量．沿着 z 方向前进的度规张量是

$$h_{ij}^{TT}(t,z)=\begin{pmatrix}h_+ & h_\times & 0\\ h_\times & -h_+ & 0\\ 0 & 0 & 0\end{pmatrix}\cos[\omega(t-z/c)]\,. \tag{5}$$

更简要地，我们可以写成

$$h_{ab}^{TT}(t,z)=\begin{pmatrix}h_+ & h_\times\\ h_\times & -h_+\end{pmatrix}\cos[\omega(t-z/c)]\,. \tag{6}$$

其中 $a,b=1,2$. 代入到度规中，我们可以写成

$$\begin{aligned}ds^2=&-c^2dt^2+dz^2+\{1+h_+\cos[\omega(t-z/c)]\}dx^2\\&+\{1-h_+\cos[\omega(t-z/c)]\}dy^2+2h_\times\cos[\omega(t-z/c)]dxdy\,.\end{aligned} \tag{7}$$

给定一个在源外面朝着 $\hat{\mathrm{n}}$ 方向传播的平面波解 $h_{\mu\nu}(x)$. 假设我们已经加入了洛伦兹规范但还没有加上 TT 规范，我们可以通过如下步骤找到 TT 规范下的波．

首先，我们引入如下的张量

$$P_{ij}(\hat{\mathrm{n}})=\delta_{ij}-n_in_j\,. \tag{8}$$

这个张量是对称，横向的（$n^iP_{ij}(\hat{\mathrm{n}})=0$），是一个投影（$P_{ik}P_{kj}=P_{ij}$），它的迹是 $P_{ii}=2$．使用上面定义的 P_{ij}, 我们可以构建如下的投影算符

$$\Lambda_{ij,kl}(\hat{\mathrm{n}})=P_{ik}P_{jl}-\frac{1}{2}P_{ij}P_{kl}\,. \tag{9}$$

这个投影算符仍然是一个投影，因为

$$\Lambda_{ij,kl}\Lambda_{kl,mn}=\Lambda_{ij,mn}\,. \tag{10}$$

并且它对于所有的指标都是横向的 $n^i\Lambda_{ij,kl}=0$, $n^j\Lambda_{ij,kl}=0$, 它对于 (i,j) 和 (k,l) 指标是无迹的

$$\Lambda_{ii,kl}=\Lambda_{ij,kk}=0\,. \tag{11}$$

给定引力波的传播方向，我们可以写出 $\Lambda_{ij,kl}(\hat{\mathrm{n}})$ 的显性表达式

$$\begin{aligned}\Lambda_{ij,kl}(\hat{\mathrm{n}})=&\delta_{ik}\delta_{jl}-\frac{1}{2}\delta_{ij}\delta_{kl}-n_jn_l\delta_{ik}-n_in_k\delta_{jl}\\&+\frac{1}{2}n_kn_l\delta_{ij}+\frac{1}{2}n_in_j\delta_{kl}+\frac{1}{2}n_in_jn_kn_l\,.\end{aligned} \tag{12}$$

我们把这个张量叫做 Lambda 张量. 给定一个在洛伦兹规范下，但是还不在 TT 规范下的平面波 $h_{\mu\nu}$，我们可以写出 TT 规范下的平面波如下.

$$h_{ij}^{TT} = \Lambda_{ij,kl} h_{kl} \,. \tag{13}$$

从等式右边可以看出，它对于 (i,j) 指标是横向，无迹的. 从 $h_{\mu\nu}$ 是真空中波动方程的一个解和它在洛伦兹规范下可以看出 $\Box h_{ij}^{TT} = 0$.

给定任意的对称张量 S_{ij}, 我们定义它的横向无迹部分为

$$S_{ij}^{TT} = \Lambda_{ij,kl} S_{kl} \,. \tag{14}$$

注意到 S_{ij}^{TT} 仍然是对称的.

在 TT 规范下，h_{ij}^{TT} 可以做如下展开

$$h_{ij}^{TT}(x) = \int \frac{d^3k}{(2\pi)^3} (\mathcal{A}_{ij}(\mathrm{k}) \mathrm{e}^{\mathrm{i}kx} + \mathcal{A}_{ij}^*(\mathrm{k}) \mathrm{e}^{-\mathrm{i}kx}) \,. \tag{15}$$

其中 $k^\mu = (\omega/c, \mathrm{k})$. $|\mathrm{k}| = \omega/c = (2\pi f)/c$ 因此 $d^3k = |\mathrm{k}|^2 d|\mathrm{k}| d\Omega = (2\pi/c)^3 f^2 df d\Omega$. 另外 $d^2\hat{\mathrm{n}} = d\cos\theta d\phi$, 所以上述方程读作.

$$h_{ij}^{TT}(x) = \frac{1}{c^3} \int_0^\infty df f^2 \int d^2\hat{\mathrm{n}} (\mathcal{A}_{ij}(f, \hat{\mathrm{n}}) \mathrm{e}^{-2\pi \mathrm{i} f(t - \hat{\mathrm{n}}\cdot\mathrm{x}/c)} + c.c) \tag{16}$$

注意到在这个表达式里面，只有物理的频率 $f > 0$ 进入了展开式. TT 规范条件给出了 $\mathcal{A}_i^i(\mathrm{k}) = 0$ 以及 $k^i \mathcal{A}_{ij}(\mathrm{k}) = 0$. 需要注意的是，如果波是由往不同方向传播的波叠加组成的，$h_{ij}(x)$ 就不能约化到 2×2 的矩阵. 这一点在我们考虑引力波的随机背景的时候会非常重要. 然而，当我们在地球上观测到从单个天体辐射出来的引力波的时候，波的传播方向还是很确定的，我们可以写出如下表达式

$$\mathcal{A}_{ij}(\mathrm{k}) = A_{ij}(f) \delta^{(2)}(\hat{\mathrm{n}} - \hat{\mathrm{n}}_0) \,. \tag{17}$$

现在我们只关注垂直于引力波运动方向的那个平面并且把平面的坐标记作 $a, b = 1, 2$ 并且省略掉 TT 这两个字，于是我们有

$$h_{ab}(t, \mathrm{x}) = \int_0^\infty df (\tilde{f}_{ab}(f, \mathrm{x}) \mathrm{e}^{-2\pi \mathrm{i} f t} + \tilde{h}_{ab}^*(f, \mathrm{x}) \mathrm{e}^{2\pi \mathrm{i} f t}) \,. \tag{18}$$

其中

$$\begin{aligned} \tilde{h}_{ab}(f, \mathrm{x}) &= \frac{f^2}{c^3} \int d^2\hat{\mathrm{n}} \mathcal{A}_{ab}(f, \hat{\mathrm{n}}) \mathrm{e}^{2\pi \mathrm{i} f \hat{\mathrm{n}}\cdot\mathrm{x}/c} \\ &= \frac{f^2}{c^3} A_{ab}(f) \mathrm{e}^{2\pi \mathrm{i} f \hat{\mathrm{n}}_0\cdot\mathrm{x}/c} \,. \end{aligned} \tag{19}$$

对于地面上的实验仪器来说，探测器的长度一般是小于 $\lambda = \lambda/(2\pi)$ 的．如果我们把坐标系的原点取在探测器的中央，我们就有 $\exp\{2\pi i f \hat{\mathrm{n}}\cdot \mathrm{x}/c\} = \exp\{i\hat{\mathrm{n}}\cdot \mathrm{x}/\lambda\} \simeq 1$．如果我们对探测器那个位置的引力波感兴趣，我们就可以忽略所有的 x 的依赖

$$h_{ab}(t) = \int_0^\infty df(\tilde{h}_{ab}(f)\mathrm{e}^{-2\pi \mathrm{i} f t} + \tilde{h}^*_{ab}(f)\mathrm{e}^{2\pi \mathrm{i} f t}) \,, \tag{20}$$

其中 $\tilde{h}_{ab}(f) = \tilde{h}_{ab}(f, \mathrm{x} = 0)$．当然，当我们考虑两个不同的探测器的引力波信号的时候，或者说当我们需要 $h_{ab}(t, \mathrm{x})$ 的空间导数的时候，我们还是需要保留 x 的．从式 6 出发，我们可以得到

$$\tilde{h}_{ab}(f) = \begin{pmatrix} \tilde{h}_+(f) & \tilde{h}_\times(f) \\ \tilde{h}_\times(f) & -\tilde{h}_+(f) \end{pmatrix} . \tag{21}$$

$+$ 和 $\times$ 都是依据选取的波传播的方向所确定的垂直平面确定的．如果我们把这个系统旋转 ψ 角度，那么 h_+ 和 $h_\times$ 将按照如下规则变换

$$\begin{aligned} h_+ &\to h_+ \cos 2\psi - h_\times \sin 2\psi \,, \\ h_\times &\to h_+ \sin 2\psi + h_\times \cos 2\psi \,. \end{aligned} \tag{22}$$

现在我们可以换一种形式写出式 20．首先定义

$$\tilde{h}_{ab}(-f, \mathrm{x}) = \tilde{h}^*_{ab}(f, \mathrm{x}) \,, \tag{23}$$

现在式 20 变成了

$$h_{ab}(t) = \int_{-\infty}^\infty df \tilde{h}_{ab}(f)\mathrm{e}^{-2\pi \mathrm{i} f t} \,. \tag{24}$$

平面波的另外一种展开式是通过引入极化张量 $\mathrm{e}^A_{ij}(\hat{\mathrm{n}})(A = +, \times)$ 来表示的．

$$\mathrm{e}^+_{ij}(\hat{\mathrm{n}}) = \hat{\mathrm{u}}_i\hat{\mathrm{u}}_j - \hat{\mathrm{v}}_i\hat{\mathrm{v}}_j \,, \quad \mathrm{e}^\times_{ij}(\hat{\mathrm{n}}) = \hat{\mathrm{u}}_i\hat{\mathrm{v}}_j - \hat{\mathrm{v}}_i\hat{\mathrm{u}}_j \,. \tag{25}$$

$\hat{\mathrm{u}}$ 和 $\hat{\mathrm{v}}$ 是垂直于传播方向 $\hat{\mathrm{n}}$ 的单位矢量．在这个定义下，极化张量按照如下定义归一化

$$\mathrm{e}^A_{ij}(\hat{\mathrm{n}})\mathrm{e}^{A',ij}(\hat{\mathrm{n}}) = 2\delta^{AA'} \,. \tag{26}$$

选取 z 为 $\hat{\mathrm{n}}$ 方向，另外 $\hat{\mathrm{u}} = \mathrm{x}, \hat{\mathrm{v}} = \hat{\mathrm{y}}$．所以

$$\mathrm{e}^+_{ab} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_{ab} \,, \quad \mathrm{e}^\times_{ab} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}_{ab} \,. \tag{27}$$

其中 $a,b=1,2$ 是 (x,y) 平面．我们按照如下的式子定义 $\tilde{h}_A(f,\hat{\mathrm{n}})$ 的振幅

$$\frac{f^2}{c^3}\mathcal{A}_{ij}(f,\hat{\mathrm{n}})=\sum_{A=+,\times}\tilde{h}_A(f,\hat{\mathrm{n}})\mathrm{e}^A_{ij}(\hat{\mathrm{n}})\,. \tag{28}$$

式 16 可以化简为

$$h_{ab}(t,\mathrm{x})=\sum_{A=+,\times}\int_{-\infty}^{\infty}df\int d^2\hat{\mathrm{n}}\tilde{h}_A(f,\hat{n})\mathrm{e}^A_{ab}(\hat{\mathrm{n}})\mathrm{e}^{-2\pi if(t-\hat{\mathrm{n}}\cdot\mathrm{x}/c)}\,, \tag{29}$$

在上述式子中我们再一次使用了 $\tilde{h}_A(-f,\hat{\mathrm{n}})=\tilde{h}^*_A(f,\hat{\mathrm{n}})$ 这个关系式．

引力波和测试质量的相互作用

- 本词条需要更多参考文献．
- 本词条缺少预备知识，初学者可能会遇到困难．

这一节我们讲引力波和探测器的相互作用．在理想情况下，我们把探测器看作一系列测试质量．在广义相对论里面，数学上选择一个规范对应于物理上选择一个特定的观测者．在前面的学习中我们已经知道了在 TT 规范下，引力波具有非常简单的形式，所以现在我们想要懂得 TT 规范对应的是什么参考系．我们也将学习到，在探测器共动参考系下，描述探测器上的引力波将更加直观．因此，在我们讨论引力波和探测器的相互作用的时候，知道我们使用的是哪个参考系是非常重要的事情．

理解一个特定的规范下的物理意义，我们需要使用两个工具：测地线方程和测地线偏移方程．

测地线方程和测地线偏移

考虑一个含参数 λ 的曲线 $x^\mu(\lambda)$. 两个相距 $d\lambda$ 的点的时空间隔是

$$\begin{aligned}ds^2&=g_{\mu\nu}dx^\mu dx^\nu\\&=g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}d\lambda^2\,.\end{aligned} \tag{1}$$

沿着类空曲线，我们有 $ds^2>0$, 所以我们可以直接开平方

$$ds=(g_{\mu\nu}dx^\mu dx^\nu)^{1/2}\,. \tag{2}$$

这是曲线固有距离的一个量度. 一个类时的曲线有 $ds^2 < 0$. 在这种情况下，我们可以定义固有时间 τ

$$c^2 d\tau^2 = -ds^2 = -g_{\mu\nu} dx^\mu dx^\nu \ . \tag{3}$$

我们可以直接使用 τ 来作为参数 λ. 因此 $x^\mu = x^\mu(\tau)$. 从式 3 我们可以导出

$$g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = -c^2 \ . \tag{4}$$

我们可以定义四动量如下

$$u^\mu = \frac{dx^\mu}{d\tau} \ , \tag{5}$$

式 4 可以写作

$$g_{\mu\nu} u^\mu u^\nu = -c^2 \ . \tag{6}$$

考虑所有可能的满足 $x^\mu(\tau_A) = x_A^\mu$ 和 $x^\mu(\tau_B) = x_B^\mu$ 的类时曲线. 对如下的作用量变分可以得到经典路径

$$S = -m \int_{\tau_A}^{\tau_B} d\tau \ . \tag{7}$$

这给出了测地线方程

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma^\mu_{\nu\rho}(x) \frac{dx^\nu}{d\tau} \frac{dx^\rho}{d\tau} = 0 \ . \tag{8}$$

这是一个测试质量在不受外力的情况下在弯曲的背景下的经典运动方程. 用四速度的话，测地线方程变成如下的形式

$$\frac{du^\mu}{d\tau} + \Gamma^\mu_{\nu\rho} u^\nu u^\rho = 0 \ . \tag{9}$$

现在考虑隔得很近的两个测地线，一个是 $x^\mu(\tau)$，另一个是 $x^\mu(\tau) + \xi^\mu(\tau)$. $x^\mu(\tau)$ 满足式 8 ，$x^\mu(\tau) + \xi^\mu(\tau)$ 满足

$$\frac{d^2(x^\mu + \xi^\mu)}{d\tau^2} + \Gamma^\mu_{\nu\rho}(x + \xi) \frac{d(x^\nu + \xi^\nu)}{d\tau} \frac{dx^\rho + \xi^\rho}{d\tau} = 0 \ . \tag{10}$$

式 10 - 式 8 并且展开到 ξ 的一阶，我们有

$$\frac{d^2 \xi^\mu}{d\tau^2} + 2\Gamma^\mu_{\nu\rho}(x) \frac{dx^\nu}{d\tau} \frac{d\xi^\rho}{d\tau} + \xi^\sigma \partial_\sigma \Gamma^\mu_{\nu\rho}(x) \frac{dx^\nu}{d\tau} \frac{dx^\rho}{d\tau} = 0 \ . \tag{11}$$

引入协变导数

$$\frac{DV^\mu}{d\tau} \equiv \frac{dV^\mu}{d\tau} + \Gamma^\mu_{\nu\rho} V^\nu \frac{dx^\rho}{d\tau} \ . \tag{12}$$

于是式 11 可以被写作

$$\frac{D^2\xi^\mu}{D\tau^2} = -R^\mu_{\nu\rho\sigma}\xi^\rho \frac{dx^\nu}{d\tau}\frac{dx^\sigma}{d\tau}\,, \tag{13}$$

用四速度表示出来，我们有

$$\frac{D^2\xi^\mu}{D\tau^2} = -R^\mu_{\nu\rho\sigma}\xi^\rho u^\nu u^\sigma\,. \tag{14}$$

这个方程告诉我们两个相邻的类时测地线会受到潮汐引力. 潮汐引力是由黎曼张量决定的.

局域惯性系和自由落体参考系

在讨论 TT 参考系和探测器参考系之前，我们可以回顾一下局域惯性系和自由落体参考系. 在广义相对论里，总可以存在一个坐标变换，令给定的坐标点 P 上的所有的 Christoffel 符号都为零.

$$\Gamma^\mu_{\nu\rho}(P) = 0\,. \tag{15}$$

在这个系下，P 点的测地线方程为

$$\left.\frac{d^2x^\mu}{d\tau^2}\right|_P = 0\,. \tag{16}$$

在这个参考系下，测试粒子是自由下落的，虽然只是在时空上的一个点而已. 这样的参考系被称为“局域惯性系”.

这样的参考系可以按照如下的步骤来构造. 在点 P 我们可以选择一组四矢量 e_α. $\alpha = 0,\ldots,3$. 让他们满足 $\eta_{\mu\nu}e^\mu_\alpha e^\nu_\beta = \eta_{\alpha\beta}$. 现在考虑从 P 点出发的测地线. (n^0,n^1,n^2,n^3) 是基矢 e_α 的四个分量. 用这个方法，我们可以走到时空中的所有点.

在一个足够小的空间里面，测地线不会相交. 所以每个点只会碰到有且仅有一条测地线. 因此，这个方法可以让我们给 P 点附近足够小的区域里面的所有点给定坐标. 这样的坐标系统被成为 Riemann normal coordinates. 我们现在来验证 Riemann normal coordiante 确实是局域惯性系的一种具体实现.

首先，从 $\{e_\alpha\}$ 的定义我们可以自然推导出 $g_{\mu\nu}(P) = \eta_{\mu\nu}$ 这样一个事实. 为了证明 $\Gamma^\mu_{\nu\rho}(P)$ 也成立，考虑测地线方程式 8 . 依照定义，坐标对于固有时是线性的，这说明 $d^2x^\mu/d\tau^2 = 0, dx^\mu/d\tau = n^\nu$. 因此测地线方程变成了如下形式

$$\Gamma^\mu_{\nu\rho}(P)n^\nu n^\rho = 0\,. \tag{17}$$

上式对所有 n^μ 都成立，我们得出如下结论

$$\Gamma^\mu_{\nu\rho}(P) = 0\,. \tag{18}$$

Riemann normal coordinates 因此为局域惯性系提供了一个具体的例子.

在局域惯性系里面，一个测试粒子只能在时空中的一点自由运动. 但我们可以做得比这个好得多. 我们可以建造一个测试质量在整个测地线都是自由落体运动的参考系. 一个自由旋转的物体沿着 $x^\mu(\tau)$ 运动会满足如下方程

$$\frac{ds^\mu}{d\tau} + \Gamma^\mu_{\nu\rho} s^\nu \frac{dx^\rho}{d\tau} = 0\,. \tag{19}$$

其中 s^μ 是自旋四矢量. 在静止系下，这个四矢量可以约化到 $s^\mu = (0, \mathrm{s})$. 这个方程是 $ds^\mu/d\tau = 0$ 的协变形式. $ds^\mu/d\tau = 0$ 说明了平坦时空下的角动量守恒. 我们从 P 点开始构造局域惯性系. 我们使用三个陀螺仪来标记空间坐标. 然后我们把这个参考系沿着测地线移动. 从定义可以看出，在这个参考系下，陀螺仪不会随着坐标轴旋转，因为它们定义了轴的方向. 于是，在整个类时测地线上，我们都有 $ds^\mu/d\tau = 0$ 并且 $\Gamma^\mu_{\nu\rho}$ 在整个类时测地线上都为 0. 这样的参考系被称作自由落体参考系（Riemann normal coordinates 的坐标轴用陀螺仪来标记）. 它的坐标被成为 Fermi normal coordinates. 一个自由下落参考系因此是整个测地线上的一个局域的惯性参考系. 这样一个参考系是自由在空间中移动的卫星所使用的参考系.

TT 参考系和 proper detector 参考系

TT 参考系

在 TT 参考系下，引力波可以呈现出特别简单的形式. 从式 8 可以看出

$$\left.\frac{d^2x^i}{d\tau^2}\right|_{\tau=0} = -\left[\Gamma^i_{\nu\rho}(x)\frac{dx^\nu}{d\tau}\frac{dx^\rho}{d\tau}\right]_{\tau=0} = -\left[\Gamma^i_{00}\left(\frac{dx^0}{d\tau}\right)^2\right]_{\tau=0}. \tag{20}$$

推导第二个等号的时候我们用到了在 $\tau = 0$ 的时候 $dx^i/d\tau = 0$. 这是因为我们令测试质量一开始处于静止状态. 现在我们把度规进行展开 $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ 然后按照 $h_{\mu\nu}$ 进行小量展开，克里斯托夫联络变成了

$$\Gamma^\mu_{\nu\rho} = \frac{1}{2}\eta^{\mu\sigma}(\partial_\nu h_{\rho\sigma} + \partial_\rho h_{\nu\sigma} - \partial_\sigma h_{\nu\rho})\,. \tag{21}$$

因此

$$\Gamma^i_{00} = \frac{1}{2}(2\partial_0 h_{0i} - \partial_i h_{00}) \,. \tag{22}$$

在 TT 规范下，这个量为零，因为规范条件会令 $h_{00} = h_{0i} = 0$. 因此，如果在 $\tau = 0$ 时刻 $dx^i/d\tau$ 为 0，在 TT 规范下它的导数 $d^2x^i/d\tau^2$ 同样为 0，因此，$dx^i/d\tau$ 总是为 0. 这表明，在 TT 规范下，如果一个粒子在波传来之前是静止的，那么波传过去之后也会同样保持静止.

在测试粒子之间的距离很小的时候，测地线偏离方程是适用的. 在 TT 规范下，两个本来静止的测试粒子之间的距离 ξ^i 同样也不变. 我们使用测地线偏离方程的第 i 分量. 因为在 $\tau = 0$,$dx^i/d\tau = 0$, $dx^0/d\tau = c$, 所以我们有

$$\left.\frac{d^2\xi^i}{d\tau^2}\right|_{\tau=0} = -\left[2c\Gamma^i_{0\rho}\frac{d\xi^\rho}{d\tau} + c^2\xi^\sigma\partial_\sigma\Gamma^i_{00}\right]_{\tau=0} \,. \tag{23}$$

在 TT 规范下，这个方程可以化简为

$$\left.\frac{d^2\xi^i}{d\tau^2}\right|_{\tau=0} = -\left[\dot{h}_{ij}\frac{d\xi^i}{d\tau}\right]_{\tau=0} \,. \tag{24}$$

因此，在 $\tau = 0$, 我们有 $d\xi^i/d\tau = 0$, 于是我们同样有 $d^2\xi^i/d\tau^2 = 0$. 因此 ξ^i 会一直保持不变. 在 TT 规范下，原本静止的测试质量测到的原时 τ 和坐标时 t 是相等的.

TT 规范表明了，在广义相对论里面，物理的效应并不依赖于坐标系，所以这个理论是对于坐标变换不变的. 在 TT 规范下测试质量的位置保持不变并不说明引力波对测试质量不起作用，只不过是我们通过测试质量来定义这样一个坐标系，使得引力波经过的时候，测试质量保持静止. 如果我们需要看出物理的变化，我们可以看 proper distance 或者 proper time.

比方说考虑这两个事件 $(t, x_1, 0, 0)$ 和 $(t, x_2, 0, 0)$. 在 TT 规范下，坐标距离 $x_2 - x_1 = L$ 保持不变，即使是引力波沿着 z 轴传播也是如此. 但是两个点之间的 proper distance 是

$$s = (x_2 - x_1)[1 + h_+\cos(\omega t)]^{1/2} \simeq L[1 + \frac{1}{2}h_+\cos(\omega t)] \,. \tag{25}$$

第十九部分

科学计算

第一章

Matlab 简介

计算物理导航

在当今时代，数值计算已经成为理科生不可或缺的技能．利用计算机程序计算理论模型，处理实验数据，绘制各种图表等都在学习和研究中都扮演非常重要的角色．虽然一些简单的计算用计算器就可以完成，但许多时候我们需要数以万计的计算量，用计算器就显得不实际了．

计算机语言

计算机的指令等信息在其数字电路中是以二进制 0 和 1（对应低电压和高电压）传播的，然而这样的语言显然不适合人类阅读和编写．为了让计算机按照指定的规则进行运算，我们可以用某种计算机语言（由英文单词，符号和数字组成）写下代码，计算机会把这种代码转换为二进制指令进行相应的运算．

科学计算中，目前比较流行的高级语言有 Matlab，Python，Mathematica 等[1]．这些语言除了可以进行传统的数值计算外，还有符号计算的功能，例如可以对解析表达式进行不定积分和求导等运算．一般来说，Mathematica 主要用于符号计算，Matlab 和 Python 主要用于数值计算．

本章的数值计算以讲解算法（Algorithm）为主，使用 Matlab 编程，且尽量使用简单的语法，这样即使读者熟悉的编程语言不是 Matlab，也能轻易地改写出相应的代码．本章使用的所有 Matlab 语法都会在前几个词条中介绍，不要求读者有任何编程基．

本章的符号计算使用 Mathematica 软件和 Wolfram Alpha 网站[2]．见“Mathematica 基础”．符号计算主要用于计算解析问题（如复杂的不定积分）或者获得高精度的数值结果（例如计算 π 的一万位）．其使用频率远没有数值计算多．

---

[1]高级语言可以理解为自动化程度较高的语言，优点是编程和调试等较简单，易学易用，缺点是效率较低．比上述语言更低级的语言有 Fortran，C++，C 等，更适合做高性能数值计算．一些大计算量的物理问题可能运行一次需要数小时乃至数星期的时间，这时运行效率是首要考虑因素，而编程和调试的难易反而相对次要．

[2]Wolfram Alpha 相当于 Matheematica 的简单网页版，优势在于无需安装软件，用有网络浏览器的设备（包括手机和平板）可随时访问．网址为 wolframalpha.com．

常用算法

为了让读者熟悉一些科学计算中常用的算法，顺便熟悉 Matlab 编程，本章还介绍了用于求函数根值的二分法[1841] 和多区间二分法[1843]，用于排序的冒泡法[1845]，用于搜索多元函数局部最小值的 Nelder-Mead 算法[1851]，以及用于曲线拟合的最小二乘法.

微分方程数值解

微分方程在物理中无处不在，力学中最基本的牛顿第二定律就是一个常微分方程. 虽然我们都希望能得到方程的精确解析解，然而往往只有最简单的一些问题才存在. 对于没有精确解析解的问题，我们要么加入一些近似条件得到近似的解析解，要么求助于数值计算，也可以是二者结合[3]. 本章中我们会学习如何计算常微分方程（组）的数值解[1886].

我们先用一阶微分近似的方法[1882][1883] 举例介绍一般的思路，再介绍更精确的中点法[1888] 和四阶龙格—库塔法[1891].

Matlab 简介

Matlab（Matrix Laboratory）的中文名叫矩阵实验室，是一款著名的科学计算软件，也指这个软件中使用的编程语言. 这里仅介绍最基本的 Matlab 功能和语法，且仅介绍本书使用到的功能.

界面介绍

Matlab 的编程界面（图 1）属于**集成开发环境（IDE/Integrated Develop Environment）**，简而言之就是一切与 Matlab 编程有关的工作都可以在该界面完成[4]. 以下介绍界面中常用的窗口. 要选择显示的窗口，可在 Home 菜单中点击 Layout 按钮，并在 Show 下面勾选需要的窗口.

[3]对于一些复杂的问题（尤其在量子力学中）有时即使用数值计算也不能在合理的时间内解出微分方程，这时就要先使用近似化简问题.

[4]界面语言默认与操作系统语言相同，本书使用英文界面.

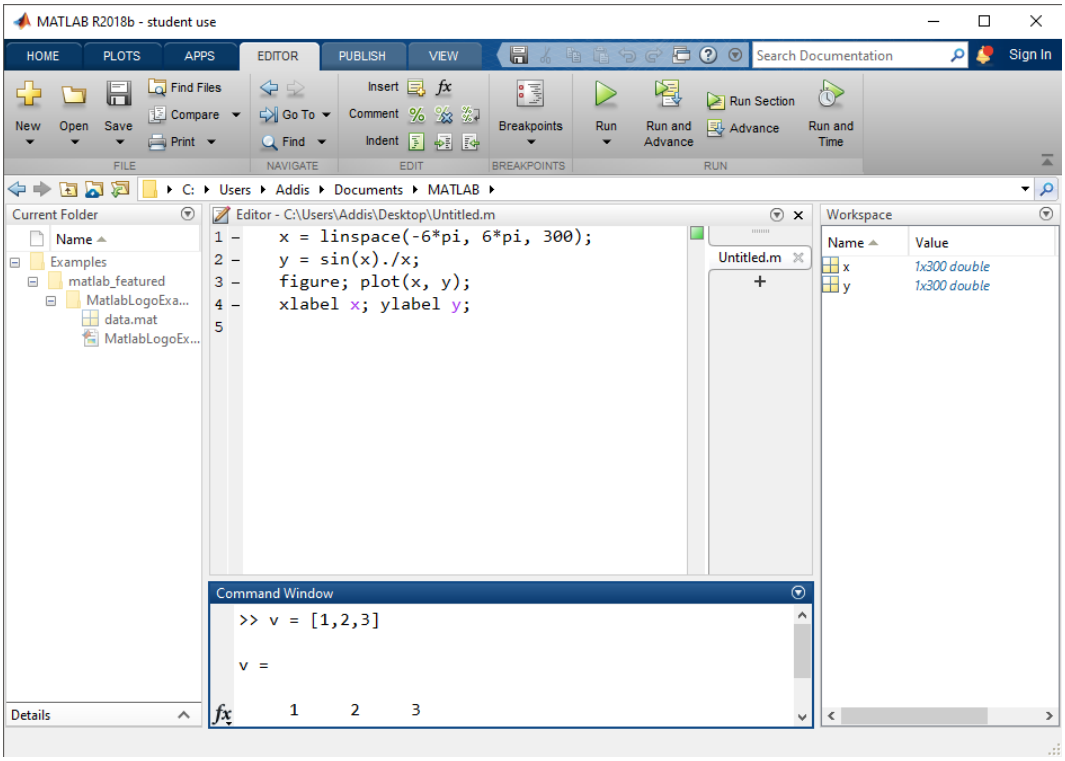


图 1: Matlab 的 IDE 界面

Editor 用于编辑代码[5]，同时具有自动检测语法错误，代码调试等功能．Matlab 的代码文件分为脚本文件和函数文件两种形式，后缀名都为“.m”，用图 1 中 Editor 菜单栏的 Save 按钮可保存代码文件．Matlab 作为一种**解释语言**（**Interpreted Language**）可以直接在 Editor 中运行源代码，无需传统的编译过程．为了让 Matlab 能运行代码文件，需要把文件所在的目录[6]添加到 Matlab 的搜索路径下．

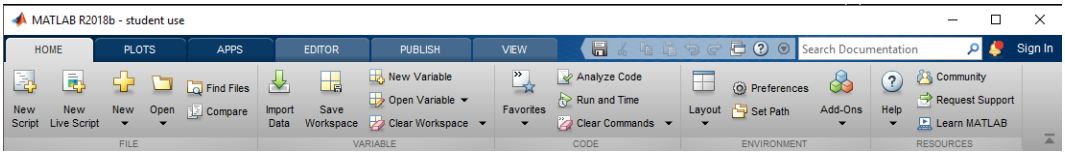


图 2: Home 菜单

如图 2，Home 菜单中的 Set Path 按钮可以设置 Matlab 的搜索路径．点开

[5]从 R2020a 版本起，Matlab 默认支持 UTF-8 编码，笔者这里强烈建议使用较新版的 Matlab，本书的代码下载统一使用 UTF-8 编码

[6]R2020a 版本前，若系统语言是英语，Matlab 不能识别中文目录．

后用 Add Folder 按钮可以添加单个文件夹（不包含子文件夹），用 Add with Subfolders 添加文件夹（包含子文件夹）. 用 Remove 删除已添加的路径，用 Default 还原初始设置，用 Save 保存修改，用 Close 关闭窗口. 若要运行程序，回到 Editor 菜单点击 Run 按钮即可.

Command Window 主要用于输入临时指令或者调试程序，可输入除了函数定义外的任意指令. Command Window 只能按输入顺序执行，不方便修改和编辑，如果指令较长或有多个指令，应该使用 Editor. 在 Command Window 中按回车执行输入的指令，按上箭头可重复已输入的指令.

Workspace 用于查看 Matlab 当前的所有变量的列表. Matlab 的所有变量都可以理解为矩阵，单个值可理解为 `1@1` 的矩阵. 列表中 Name 是变量名，Size 是矩阵维度，Value 是变量值，右上角的下拉菜单中的 Choose Columns 中还可设置显示更多属性，例如 Bytes 是占用字节数，Class 是变量类型，Min 是最小值，Max 是最大值，Mean 是平均值，Median 是中位数，Std 是标准差等. 双击 Workspace 中的变量可显示变量值.

Matlab Online

Matlab Online 是 Matlab 网页版，具有 Matlab 的基本功能，和类似于软件的界面，需要购买了正版 Matlab 的 Matlab 账号登录（学生账号也可以）. 若账号购买了工具箱（Toolbox），也可以使用对应的工具箱. 本书官网 wuli.wiki 提供免费的 Matlab 账号供读者试用和体验 Matlab Online.

计算器

下面我们仅用 Command Window 来熟悉 Matlab 的基本语法. 我们先看如何把 Matlab 当做普通的科学计算器使用. 在 Command Window 中，“>>”提示符表示用户在该位置输入命令，`ans` 是一个特殊的变量[1685] 用于储存计算结果.

```
>> 1.2/3.4 + (5.6+7.8)*9 -1
ans = 119.9529
>> ans + 1
ans = 120.9529
>> 1/exp(1)
ans = 0.3679
```

```
>> exp(-1i*pi)+1
ans = 0
```

常用的运算符号有

```
+, -, *, /, ^（指数）
```

常用的数学函数有

```
sqrt（开方）, exp, sin, cos, tan, cot, asin, acos, atan, acot,
   real（实部）, imag（虚部）, conj（共轭）
```

等（三角函数前面加 `a` 代表其反函数）. 运算的优先顺序与数学上的习惯一样. 注意这些函数的自变量都可以是复数. 为了区分虚数单位 `i` 和变量 `i`，好的习惯是在 `i` 前面加数字（上面的第三条命令）. `pi` 是圆周率.

`mod(N,n)` 是求余运算，计算 `N` 被 `n` 整除后的余数. 注意这个函数有两个变量，用逗号隔开. 要注意在 Matlab 中，这种有输入和输出的命令都是广义的**函数（function）**，不仅是数学函数.

`sign(num)` 函数用于求实数 `num` 的符号. 如果 `num > 0`，则返回 1，若 `num < 0` 则返回 -1，若 `num = 0` 则返回 0.

用大写 `E` 或小写 `e` 表示科学计数法（不允许有空格），如 2.997×10^8 表示为 `2.997e8` 或 `2.997e+8`. 用小写 `pi` 表示圆周率，用 `exp(1)` 表示自然对数底，用 `1+2i` 或 `1+2j` 等表示复数，注意 `i` 和 `j` 前面不能有空格.

如果需要在输出中显示多位小数，可使用 `format long` 命令使结果显示为**双精度**（约 16 位有效数字），用 `format short` 命令恢复默认格式.

```
>> format long; pi
ans = 3.141592653589793
>> exp(1)
ans = 2.718281828459046
```

Matlab 的变量与矩阵

预备知识 Matlab 简介[1682]，自然对数底[153]，矩阵[396]

变量与矩阵

变量（variable）可用于储存数据并通过变量名获取变量值．变量名可以由多个字母，数字和下划线组成．注意变量名区分大小写，且首字符只能是字母．合法的变量名如 `ii`，`a3`，`odd_number`，`evenNumber`．用等号可以对变量**赋值**，被赋值的变量放在等号左边，等号右边的表达式的运算结果会储存在被赋值的变量中，直到再次被赋值．**表达式**由若干变量，常数和算符组成．注意同一个变量可以出现在等号左边以及右边的表达式中．

```
>> a = 1.2/3.4 + (5.6+7.8)*9 -1
a = 119.9529
>> a = atan(a + 1)
a = 1.5625
```

如果新的变量第一次被赋值，它会自动出现在 Workspace 窗口中．注意 Workspace 中的一个特殊的变量 `ans`，如果命令的输出结果没有赋值给变量，就会自动赋值给 `ans`．注意一般不要对 `ans` 赋值．另外一些特殊的变量如 `pi`（圆周率，最好不要赋值），和 `i`（虚数单位）．由于 `i` 时常被用于循环变量，一种好习惯是在虚数单位前加数字，例如 `1i`，`1+3i` 等．自然对数底没有对应的变量，若要使用自然对数底，用 `exp(1)` 即可．

另外，如果在命令后面加分号（semicolon）“`;`”，则命令执行后不输出结果．也可以用分号把多个命令写到一行．

```
>> 1 + 1; a = ans^2
a = 4
```

用 Editor 编写程序时，每个命令后面都需要加分号，需要在 Command Window 输出时，用 `disp` 函数．

```
>> disp('something'); disp(10);
something
10
```

`clear` 命令可以清空 Workspace 中的所有变量，用 `clear <var1>,<var2> ...` 清除指定的变量（`<var1>,<var2>` 是变量名）．用 `clc` 命令可以清空 Command Window （按上箭头仍然可以查看历史命令）．

本书只涉及到 3 种**变量类型（class）：双精度（double），字符（char）**和**逻辑（logical）**.

双精度变量

双精度变量用于储存数值，有效数字约为 16 位（如果是复数，实部和虚部各 16 位），取值范围约为 10^{-308} 到 10^{308}. 如无变量类型声明，所有命令中出现的常数及储存数值的变量都为 `double`.

Matlab 中的所有变量都可以理解为**矩阵**，单值变量（标量，scalar）可以理解为 `1@1` 的矩阵，只有一行或一列的矩阵叫做**行矢量（row vector）**或**列矢量（column vector）**. 一些简单的矩阵操作如下

```
>> a = [1,2,3]
a = 1   2   3
```

用方括号创建矩阵，用逗号分隔每行的矩阵元，行矢量中逗号可省略

```
>> a = [1 2 3]
a = 1   2   3
```

用分号分隔行

```
>> b = [1;2;3]
b =
    1
    2
    3
>> c = [1 2 3; 2 3 4; 3 4 5]
c =
    1   2   3
    2   3   4
    3   4   5
```

方括号还可以用来合并矩阵（注意矩阵尺寸必须合适）

```
>> d = [a; a]
d =
    1   2   3
    1   2   3
>> e = [a a]
```

```
e =
    1   2   3   1   2   3
```

用 size 函数获取矩阵尺寸，如果使用两个输入变量，则第二个用于指定维度（行是第一个维度，列是第二个）. 用 numel 函数获取矩阵元个数.

```
>> size(d)
ans = 2   3
>> size(d,1)
ans = 2
>> numel(d)
ans = 6
```

用 zeros 函数生成全零矩阵

```
>> zeros(2,3)
ans =
    0   0   0
    0   0   0
```

用 zeros([2,3]) 和 zeros(size(d)) 结果也相同. 用 ones 可以生成全 1 矩阵，也可以乘以任意常数

```
>> ones(2,3)*5
ans =
    5   5   5
    5   5   5
```

用 eye(N) 生成 N∈N 的单位矩阵. 用 rand(M,N) 生成随机矩阵，矩阵元从 0 到 1 均匀分布，如果不提供输入变量，rand 相当于 rand(1,1). 用 M:step:N 生成等差数列（行矢量），例如

```
>> 1:2:10
ans = 1   3   5   7   9
>> 0:pi/3:pi*2
ans = 0   1.0472   2.0944   3.1416   4.1888   5.2360   6.2832
>> 10:-2:1
ans = 10   8   6   4   2
```

如果只用一个冒号，那么间隔默认为 1

```
>> 1:3
ans = 1  2  3
```

用 `linspace(x1,x2,Nx)` 生成指定首项尾项和项数的等差数列（行矢量）

```
>> linspace(0,pi,4)
ans = 0  1.0472  1.2566  2.0944  3.1416
```

下面介绍矩阵运算. 形状相同的矩阵可以进行 `+` 和 `-` 运算，矩阵和标量也可以，结果是把每个矩阵元加（减）标量；矩阵乘法“`*`”既可以把常数与矩阵相乘，也可以进行数学上的矩阵乘法；矩阵的幂“`^`”相当于矩阵与自己多次相乘；“`/`”可以把矩阵除以一个常数.

```
>> a = [1 2; 3 4]; b = [1 -1; 2 -2];
>> a + b
ans =
    2  1
    5  2
>> a * b
ans =
    5  -5
    11  -11
```

若两个矩阵尺寸相同，可进行**逐个元素运算**，如

```
>> a .* b
ans =
    1  -2
    6  -8
```

这里把 `a` 的每个矩阵元和 `b` 的对应矩阵元相乘，得到新矩阵的对应矩阵元. 为了区分矩阵乘法，需要在乘号前面加一个英文句号. 类似地，我们也可以进行逐个元素的除法运算（“`./`”）和幂运算（“`.^`”）.

```
>> a ./ b
ans =
    1.0000  -2.0000
    1.5000  -2.0000
>> a.^2
ans =
```

```
    1   4
    9   16
```

逐个元素运算也可以用于一些尺寸不同的情况，例如对 2×3 的矩阵与 2×1 的矢量进行逐个元素运算，相当于将后者复制成 2×3 的矩阵再进行运算．又例如 1×3 的矢量与 2×1 的矢量进行逐个元素运算，相当于把两个矢量都复制成 2×3 的矩阵再进行运算．再例如一个标量和一个矩阵做逐个元素运算相当于把标量复制成同尺寸的矩阵再进行运算．

```
>> m1 = [1 3 5; 2 4 6]; m2 = [1; 2];
>>  m1.*m2
ans =
    1   3   5
    4   8   12
>> m1 = [1, 2, 3]; m2 = [2; 3];
>> m1 + m2
ans =
    3   4   5
    4   5   6
>> a = [1 2; 3 4];
>> 24./a
ans =
    24   12
    8    6
```

单引号“'”可以使实数矩阵转置，或使复矩阵取厄米共轭（即转置并取复共轭）．若只需要对复矩阵转置，用“.'”即可．

```
>> c = a + 1i*b
c =
    1.0000 + 1.0000i   2.0000 - 1.0000i
    3.0000 + 2.0000i   4.0000 - 2.0000i
>> c'
ans =
    1.0000 - 1.0000i   3.0000 - 2.0000i
    2.0000 + 1.0000i   4.0000 + 2.0000i
>> c.'
ans =
```

```
    1.0000 + 1.0000i   3.0000 + 2.0000i
    2.0000 - 1.0000i   4.0000 - 2.0000i
```

使用 fliplr 和 flipud 函数可以分别把矩阵左右翻转和上下翻转，例如

```
>> fliplr(a)
ans =
    2   1
    4   3
>> flipud(a)
ans =
    3   4
    1   2
```

用 sum 函数可以分别求矩阵每列或每行的和．当矩阵为行矢量或列矢量时，sum 对所有矩阵元求和．其他情况下，sum 默认对每列求和，若想对每行求和，可以在第二个输入变量中输入 2 （行是矩阵的第二个维度）．例如

```
>> a = [1,2; 3,4];
>> sum(a)
ans = 4   6
>> sum(a,2)
ans =
    3
    7
```

用 mean 函数可以求平均值，使用格式与 sum 相同．

给 norm 函数输入一个行矢量或列矢量时，得到该矢量的模长（即各矢量元的平方和再开根号）．

用 det 函数可以求方阵的行列式[432]，用 “\” 算符可以解线性方程组 $\mathbf{Ax}=\mathbf{b}$．如[7]

```
>> A = [1,2; 3,4]; det(A)
ans = -2
>> x = [-1; 1]; b = A*x
b =
    1
```

[7]若 $|\mathbf{A}|=0$，即方程组存在无数个解的时候，“\” 算符只能得到一个特解．若 $\mathbf{A}$ 是一个超定方程组，即行数大于列数，算法见 “超定线性方程组[485]”．

```
     1
>> A \ b
ans =
    -1
     1
```

Matlab 自带的数学函数一般支持矩阵自变量，结果是该函数对每个矩阵元分别运算．这样的函数叫做**矢量化（vectorized）**的函数．

```
>> cos(0:pi/4:pi)
ans =
1.0000   0.7071   0.0000   -0.7071   -1.0000
```

用矢量化的运算可以使代码简短易懂，且提高计算效率（因为只需要进行一次函数调用）．

矩阵索引

矩阵索引用于表示矩阵部分矩阵元，例如

```
>> a = [1 2 3; 4 5 6; 7 8 9];
ans =
    1   2   3
    4   5   6
    7   8   9
>> a(1,2)
ans = 2
```

索引括号中的两个数或变量分别代表行标和列标，也可以是矩阵，例如

```
>> a(2:3,1)
ans =
    4
    7
>> a(2:3,1:2)
ans =
    4   5
    7   8
>> a([3,2],[1,2])
ans =
```

```
    7   8
    4   5
>> a(:,2)
ans =
    2
    5
    8
>> a(1:end-1,2:3)
ans =
    2   3
    5   6
```

其中单独一个冒号表示某个维度的全部索引，end 表示某维度的最大索引．注意索引不仅可以用来取值，还可以放在等号左边赋值．

```
>> b = a; b(1:3) = a(2:4)
b =
    4   2   3
    7   5   6
    2   8   9
```

要求左边的矩阵元个数等于右边．唯一的例外是当右边为标量

```
b(1:3) = 0
b =
    0   2   3
    0   5   6
    0   8   9
```

我们还可以用单个索引

```
>> a(2:5)
ans = 4   7   2   5
>> a(7:end)
ans = 3   6   9
```

注意单个索引的顺序是先增加第一个维度（行标），再增加第二个维度（列标）．虽然上面都是以双精度矩阵为例，但这些索引方法适用于任何数据类型的矩阵．

字符变量

字符型变量一般用于控制行输出结果或对生成的图片进行标注. 把 N 个字符放在一对单引号内，可生成 1ŒN 的字符类型数组.

```
>> str1 = '这是一个字符串'; str2 = 'this is a string';
>> [str1, ',' , str2]
ans = '这是一个字符串, this is a string'
>> numel(ans)
ans = 24
```

把双精度类型变为字符串可以用 num2str 函数（注意 2 的英文读音与“to”相同，num 代表“number”，str 代表字符串“string”），通常用于与其他字符矩阵合并，如

```
>> number = 3; str = ['The number is', num2str(number), '.']
str = 'The number is 3.'
```

若要在字符串中加入英文单引号，可用两个英文单引号表示.

逻辑变量

逻辑变量只能具有 0 或 1 两个值，分别代表**假（false）**和**真（true）**. 以下算符

```
>, >=（大于等于）, <, <=, ==（等于）,  ~=  （不等于）
```

可用于比较双精度数组，返回逻辑型数组

```
>> L = 1 + 1 > 3
L = logical 0
```

逻辑“与”，“或”，“非”算符分别为（仅用于逻辑标量）

```
&&    ||    ~
```

当算符两边都为真时，与运算才能为真，若至少有一边为真，或运算就为真. 非运算用于把真假互换. 例如

```
>> 1 > 0 && 2 > 1
ans = logical 1
>> 1 > 0 || 2 < 1
```

```
ans = logical 1
>> ~ (1 > 0)
ans = logical 0
```

注意由于非运算的优先级比 > 运算要高，所以最后一条命令必须要加括号.

在需要的时候，双精度变量可以自动转换为逻辑变量，规则是只有双精度的 0 转换为逻辑 0，其他双精度值一律转换为逻辑 1. 如

```
>> 1.3 && -0.8
ans = logical 1
```

其中 1.3 和 0.8 分别转换为逻辑 1，然后进行与运算得到逻辑 1.

除了上文中介绍的索引方法外，我们还可以用相同大小逻辑数组索引任意矩阵. 注意逻辑索引的输出结果是一个列矢量，以下为了节约空间我们将输出结果改为行矢量.

```
>> a = [1 2 3; 4 5 6; 7 8 9];
>> mark = logical([1 0 0; 0 0 1; 1 0 0]); a(mark)
ans = 1  7  6 % 列矢量
```

逻辑索引常见的例子如

```
>> a(a <= 4)
ans = 1  4  2  3 % 列矢量
```

注意逻辑索引中不能使用双精度类型代替逻辑类型.

find 函数可以用于寻找逻辑矩阵中所有值为 1 的矩阵元的位置. 如果只提供一个输出变量（或不提供输出变量），函数返回单索引，如果提供两个输出变量，函数返回行标和列标. 例如

```
>> a = [1, 0; 0, 1]
a =
    1  0
    0  1
>> find(a)
ans = 1  4
>> [r, c] = find(a)
r = 1  2 % 列矢量
c = 1  2 % 列矢量
```

注意以上的双精度矩阵 a 被自动转换为逻辑矩阵.

any 函数用于判断逻辑矩阵中的每一列是否存在任何值为“真”的矩阵元，若有，则返回真，否则返回假. all 函数用于判断逻辑矩阵中的每一列是否所有矩阵元都为“真”，若是，返回真，否则返回假. 这两个函数的使用格式与 sum 和 mean 类似，当第二个输入变量为 2 时，对每行进行操作.

```
>> a = [1 6 7; 2 7 1; 3 8 9];
>> all(a > 5)
ans = 1×3 logical array
    0   1   0
>> any(a > 7, 2)
ans = 3×1 logical array
    0
    0
    1
>> all(a(:) < 10)
ans = logical 1
```

Matlab 的判断与循环

预备知识 Matlab 的变量与矩阵[1685]

脚本文件

在讲解更复杂的程序结构前，我们先来看脚本文件. **脚本（script）文件**是包含若干个指令的文件，文件后缀名为“.m”. 脚本文件可以单独执行，也在其他文件或 Command Window 中被调用（注意需要将所在文件夹添加到搜索路径）. 后者相当于把被调用脚本的代码直接插入到调用指令处，调用指令就是脚本文件的文件名. 脚本中的每条命令后面应该加分号以隐藏输出结果，若需要输出，用 disp 函数.

```
>> disp('good'); a = 3; disp(['a = ', num2str(a)])
good
a = 3
```

在脚本文件中，可以在行首或命令后用百分号 % 进行**注释（comment）**[8]．注释是程序的说明，使程序更易读，但在执行程序时会被忽略（图 1 [1682]）．

判断结构

现在来看一段代码（脚本文件）．要生成新的脚本文件，可以在 Editor 菜单（图 1 [1682]）的左边单击 New，然后选 Script，或者用快捷键 Ctrl+N．在生成的 Editor 中输入以下代码

```
a = rand(1,1); b = 0.5;
if a > b % a 较大
    disp('a is larger');
else % b 较大
    disp('b is larger');
end
```

这段程序用 rand 函数随机生成一个从 0 到 1 的数，如果随机数大于 0.5 则输出第一段文字，否则输出第二段文字．不难猜测出这里的 if 用于判断，如果条件满足，则只执行 if 和 else 之间的指令．如果条件不满足，则只执行 else 到 end 的指令．注意 else 语句可以不在判断结构中出现，若不出现，当判断条件不满足时程序将直接执行 end 后面的代码．

要执行该代码，在 Editor 菜单中单击 Run 图标（绿色三角形），如果代码没有保存，Matlab 会先弹出保存对话框．再次强调文件必须保存在 Matlab 的搜索路径下．

elseif 语句可用于在判断结构中产生多个分支，如

```
a = rand(1,1);
if a > 0.9
    disp('a in (0.9, 1]');
elseif a > 0.6
    disp('a in (0.6, 0.9]');
elseif a > 0.3
    disp('a in (0.3, 0.6]');
else
```

[8]截止到 Matlab 2017b，在英文版 Matlab IDE 中，任何中文注释都会在 Matlab 重启后变为乱码．若要使用中文注释，建议使用中文操作系统和中文 Matlab．

```
    disp('a in [0, 0.3]');
end
```

这个程序用于判断随机数 a 的区间. 若 if 的条件判断成功, 判断结构就只执行 if 到第一个 elseif 之间的命令. 若 if 判断失败, 程序就继续判断第一个 elseif 中的条件, 若判断成功, 就只执行第一个 elseif 到第二个 elseif 之间的命令, 以此类推. 如果 if 和所有的 elseif 条件都判断失败, 则执行 else 后面的命令.

循环结构

我们先来看 for 循环

```
for ii = 1:3
    disp(['ii^2 = ' num2str(ii^2)]);
end
```

运行结果为

```
ii^2 = 1
ii^2 = 4
ii^2 = 9
```

容易看出这段代码被执行了 3 次, **循环变量** ii 按顺序取 1:3 中的一个矩阵元. 注意选取 ii 作为变量名是为了与虚数单位区分, 当然也可以选择其他变量名. 再来看一个稍复杂的循环

```
Nx = 5;
x = zeros(1,Nx);  % 预赋值
x(1) = 2;
for ii = 2:numel(x)
    x(ii) = x(ii-1)^2;
end
disp(['x = ' num2str(x)])
```

在循环开始前 x(1) 被赋值为 2, 在循环中, 第 ii 个矩阵元依次被赋值为第 ii-1 个矩阵元的平方. 运行结果为

```
x = 2   4   16   256   65536
```

注意在循环前用 zero 对矩阵进行了**预赋值（preallocation）**. 预赋值不是必须的，但如果不进行预赋值，每次循环矩阵的尺寸都要改变，会导致程序运行变慢. 另外注意循环中不允许给循环变量赋值.

再来看另一种循环叫做 while 循环. 下面来看一个例程，输出 100 以内的斐波那契数列（$a_1 = 1, a_2 = 1, a_{n+1} = a_n + a_{n-1}$）.

代码 1: fibonacci.m

```
a1 = 1; disp(a1);
a2 = 1; disp(a2);
a3 = a1 + a2;
while a3 <= 100
    disp(a3);
    a1 = a2;
    a2 = a3;
    a3 = a1 + a2;
end
```

while 结构在每个循环开始会判断 while 后面的条件，如果条件成立，则进行一次循环，否则退出循环. 以上的程序中由于我们事先并不知道我们要进行几次循环，所以选用 while，当最后一项大于 100 时，循环终止. 运行结果为（每个数占一行）：

```
1   1   2   3   5   8   13   21   34   55   89
```

在 for 循环或 while 循环的内部，使用 continue 命令可以直接进入下一个循环（while 的仍然要先判断条件），使用 break 命令可以跳出循环. 以下例程计算 100 以内的斐波那契数列的所有奇数项

代码 2: fibonacciOdd.m

```
a1 = 1; a2 = 1;
disp(a1); disp(a2);
a3 = a1 + a2;
while 1
    a1 = a2;
    a2 = a3;
    a3 = a1 + a2;
    if a3 > 100
```

```
9           break;
10      elseif mod(a3, 2) == 0
11          continue;
12      end
13      disp(a3);
14  end
```

先来看第 4 行，`double` 类型的非零数在这里会自动转换为 `logical` 类型的 1（`true`），只有 `double` 类型的 0 才会转换为 `logical` 类型的 0（`false`）. 乍看之下，`while` 循环将永远执行下去（称为**死循环**），然而第 9 行的 `break` 在 `a3 > 100` 时就会使程序跳出循环. 如果 `a3 <= 100` 且为偶数，则第 10 行的 `elseif` 判断为真，`continue` 命令被执行，程序将直接跳过之后的 `disp` 函数直接进入下一个循环，所以数列的偶数项都不会被输出. 程序的运行结果为（每个数占一行）

```
1   1   3   5   13   21   55   89
```

return 命令

在一个脚本文件的任何地方，如果 `return` 命令被执行，则程序将结束该脚本文件的执行. 如果该脚本文件是被单独执行的，程序将终止. 如果该脚本文件是被其他脚本文件或函数文件调用的，程序将继续执行调用命令的下一个命令.

Matlab 的函数

预备知识 Matlab 的判断与循环[1696]

函数文件

我们已经学了一些函数，现在来看如何自定义函数. Matlab 中定义了函数的文件叫做**函数文件**. 函数文件同样以“.m”作为后缀名，文件中的第一个命令必须是 `function`，用于定义主函数. 文件名必须与主函数同名. 文件中其他函数都是子函数. 主函数可以调用子函数，子函数可以调用同文件中的其他

子函数，但不能调用主函数，主函数和子函数都可以调用 Matlab 的内部函数或搜索路径下其他函数文件中的主函数. 若函数文件在搜索路径下，其他 m 文件或 Command Window 中可以直接调用它的主函数. 注意函数文件中的子函数不能从文件外被调用.

函数的 workspace 是独立的，即函数在执行的过程中，只能读写输入变量，函数内部定义的定量，以及全局变量（暂不介绍，不建议使用），而不能读取调用该函数的代码中的变量. 相比之下，调用脚本相当于把脚本的代码直接插入到调用命令处，所以脚本中可以获取调用脚本的代码中的变量. 注意函数只能通过函数文件定义，不能在脚本文件或控制行中定义.

函数句柄

函数句柄（function handle）是一种特殊的变量类型，可用于定义一个临时的函数，也可传递到其他函数中. 首先，对于已经存在的函数（包括函数文件定义的），可直接在函数名前面加 `@` 生成函数句柄

```
>> f = @sin; f(pi/2)
ans = 1;
```

若要将一个含有变量的表达式变为函数句柄，要在 `@` 后面用小括号指定函数句柄的变量

```
>> A = 3; w = 5; phi = pi/2;
>> f1 = @(x) A*sin(w*x+phi) + (2*x/pi).^2;
>> f1([0,pi/2])
ans =
    3.0000   1.0000;
>> f2 = @(x,phi) A*sin(w*x+phi) + (2*x/pi).^2;
>> f2([0,pi/2],pi/2)
ans =
    3.0000   1.0000;
```

以上的函数句柄在定义时用了逐个元素的幂运算“`.^`”，使函数句柄支持矩阵输入. 其中 `f1` 的变量仅为 `x`，`f2` 的变量为 `x` 和 `phi`. 注意如果在函数句柄定义后改变定义表达式中的变量，函数句柄不变.

```
>> A = 5; f1(0)
ans = 3.0000
```

有时候一些函数并不支持矢量运算，这时可以通过 arrayfun 函数来进行矢量运算，例如

```
>> f = @(x) x^2 + 1/x;
>> arrayfun(f, [1,2;3,4])
ans =
    2.0000   4.5000
    9.3333   16.2500
```

这样做相当于以下代码

```
f = @(x) x^2 + 1/x;
x = [1,2; 3,4];
ans = zeros(size(x));
for ii = 1:numel(x)
    ans(ii) = f(x(ii));
end
```

自定义函数（function）

自定义函数的格式为

```
[输出1, 输出2, ...] = function 函数名(变量1, 变量2, ...)
函数体
end
```

其中 函数名 是字母，数字和下划线的组合，例如 MyFun_123，第一个字符不能是数字或下划线．若函数无变量，则小括号可省略．若函数无输出，则等号及方括号可省略，若只有一个输出，方括号也可省略．在一些情况下，如果 函数体 中没有使用某些输入变量，就可以把这些变量用“~”符号代替．

函数的调用格式为

```
[输出1, 输出2, ...] = 函数名(变量1, 变量2,...)
```

调用函数时，如果输出变量个数少于函数定义中的输出变量个数，则函数仅输出前几个变量．若调用函数时不需要前面的某几个变量，也可用“~”符号代替．

调用函数时，输入变量的个数也可以少于函数定义中的输入变量，但是函数体内部必须要做出相应的措施以防止函数体使用未生成的变量．我们来看下

面一个函数

```
function y =  myfun(x, A, phi, y0)
if nargin < 4
    y0 = 0;
    if nargin < 3
        phi = 0;
        if nargin < 2
            A = 1;
        end
    end
end
y = A*cos(x + phi) + y0;
end
```

注意函数体中使用了一个特殊的变量 nargin，每当函数被调用时，这个变量的值将会等于输入变量的个数（同理，nargout 将等于输出变量的个数）. 以上定义的函数允许 1-4 个输入变量，函数体中的 2-10 行根据 nargin 的值对没有输入的几个变量依次赋值. 例如在控制行中调用该函数

```
>> myfun([0, pi/2])
ans = 1.0000    0.0000
>> myfun(0, 1, pi/2)
ans = 0.0000
```

调用函数时的一种特殊格式是，如果被调用的函数的输入变量是若干个没有空格的字符串，则可以省略括号，逗号和双引号. 例如我们之前见过的 format 函数

```
>> format long;
```

和将在“Matlab 画图[1704]”中见到的

```
>> hold on; axis equal; xlabel x;
```

与脚本文件相同，若函数文件中执行了 return 命令，该函数的执行将立即终止，程序将执行调用该函数的命令的下一个命令. 若该函数是单独被调用的（例如按开始按钮或在控制行被直接调用），则程序结束.

Matlab 画图

预备知识 Matlab 的函数[1700]

Matlab 具有强大的画图功能，这里仅介绍一些基础知识．最常用的画图函数是 `plot`，例如

```
>> x = linspace(0, 2*pi, 100); y = sin(x);
>> plot(x, y);
```

结果如图 1（左）所示．如果要在该坐标系继续画图，要用 `hold on` 命令（`on` 是 `hold` 的输入变量），否则每用一次 `plot`，之前画过的图都会被清除．用 `hold off` 可以重新恢复自动清除．

```
>> y1 = cos(x);
>> hold on; plot(x, y1);
```

结果如图 1（右）所示，注意新增曲线的颜色变化．

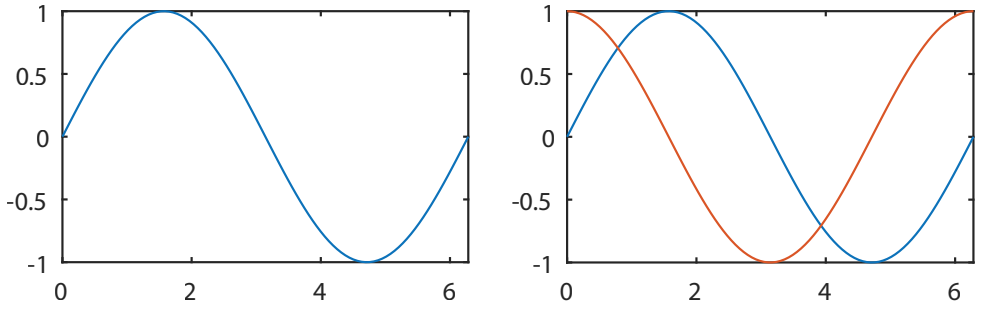


图 1: plot 函数

如果我们要新建一个画图窗口，用 `figure` 函数．若要指定画图的颜色，可以添加 `figure` 的第三个变量，用一个字符表示颜色（red：`'r'`，green：`'g'`，blue：`'b'`，yellow：`'y'`，magenta：`'m'`，cyan：`'c'`，black：`'k'`，white：`'w'`）．例如

```
>> x2 = cos(x); y2 = sin(x);
>> figure; plot(x2, y2, 'r');
```

在新增的窗口中，结果如图 2 （左）所示．注意根据窗口尺寸的不同，x 轴和 y 轴的单位长度一般不同，若要使其相同，可以在 `plot` 后面用 `axis equal` 命令（其中字符串 `equal` 是 `axis` 函数的输入变量），得到图 2 （右）．若要调整

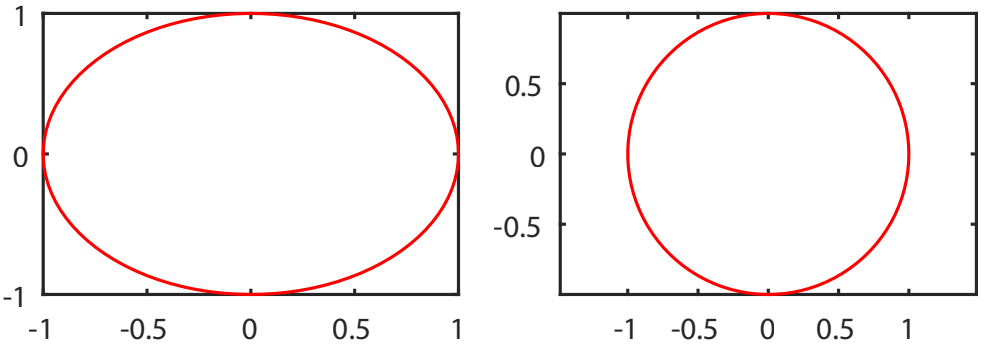


图 2: 红色的单位圆

坐标轴的范围，也可用 `axis` 函数．另外可以在 `plot` 的第三个变量的字符串中设定曲线的形状，用 `xlabel` 和 `ylabel` 函数分别设置 x 轴和 y 轴的文字，用 `title` 函数设置图片标题

```
>> plot(x2, y2, '.-r');
>> axis([-1.2, 1.2, -1.2, 1.2]);
>> xlabel('x'); ylabel('y'); title('unit circle');
```

其中 `'.-'` 表示带点的连线，点的坐标由 `x2` 和 `y2` 决定（另外 `'+-'` 表示带加号的连线，`'o-'` 表示带圆圈的连线）．`axis` 中行矢量中的四个数分别是 x 轴的最小最大值和 y 轴的最小最大值．结果如图 3 （左）所示．

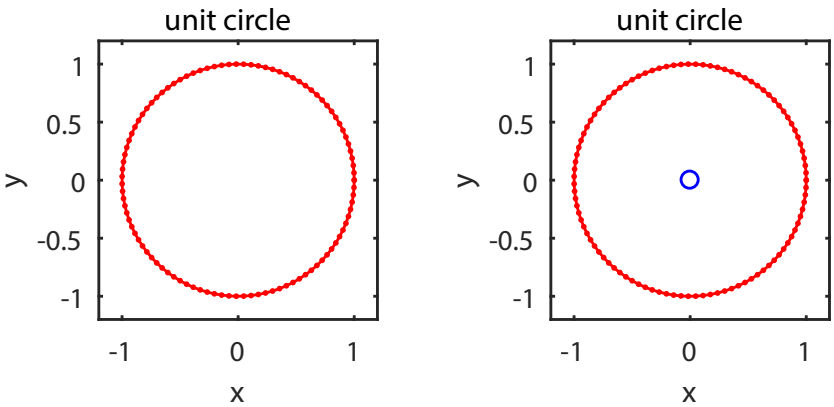


图 3: 红色的单位圆

除了 plot 以外，常用的还有 scatter 函数，用于画散点图．格式与 plot 相似．默认的散点形状是圆圈，但也可以在第三个变量中设置颜色和 '+'，'x'，'.' 等形状．例如

```
>> hold on; scatter(0, 0, 'b');
```

结果如图 3 （右）所示．

最后，如果要关闭当前画图窗口，用 close 函数（无输入变量），如果要关闭所有窗口，用 close all 即可．

Matlab 画箭头矢量场图

未完成：这里只有单个箭头，需要画出矢量场．

这里提供一个在平面上画箭头的 Matlab 函数 arrow，该函数的特点是可以自动检测当前 plot box 的长宽比，使箭头的形状不因此失真．这里是一个演示程序

```
1 figure; axis([0, 1000, 0, 1]);
2 arrow(0, 0, 800, 0.7, 50);
```

运行结果如下，可以看到虽然该图横轴单位距离比纵轴小很多，但画出的箭头仍不失真．

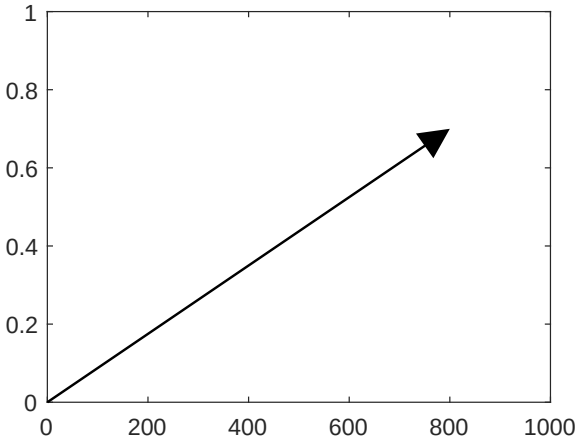


图 1: 运行结果

代码 1: arrow.m

```
% plot arrow without distortion by getting current aspect ratio
% will 'hold on' and 'axis manual'
function arrow(x, y, vx, vy, headWidth, color, lineWidth)
hold on; axis manual;
A = norm([vx,vy]);
if ~exist('headWidth', 'var') || isempty(headWidth)
    headWidth = 0.2 * A;
end
if ~exist('color', 'var') || isempty(color)
    color = 'k';
end
if ~exist('lineWidth', 'var') || isempty(lineWidth)
    lineWidth = 1;
end
[x1, y1] = xy2equi(x, y);
[vx1, vy1] = xy2equi(vx, vy);
A1 = norm([vx1, vy1]);
headWidth1 = headWidth / A * A1;
Ptri1 = arrowTri(x1, y1, vx1, vy1, headWidth1);
tri_x1 = Ptri1(:,1); tri_y1 = Ptri1(:,2);
[tri_x2, tri_y2] = equi2xy(tri_x1, tri_y1);
vx = vx * (1-headWidth/A*0.3); vy = vy * (1-headWidth/A*0.3);
% plot line
plot([x, x + vx], [y, y + vy], 'Color', color, 'LineWidth',
   lineWidth);
hold on;
% plot arrow head (triangle)
fill(tri_x2, tri_y2, 'w', 'FaceColor', color, 'EdgeAlpha', 0);
end

% calculate the arrow triangle
function Ptri = arrowTri(x, y, vx, vy, headWidth)
A = norm([vx,vy]);
u = [vx, vy, 0] / A;
Ptri = [0, -1; 0.5, -1; 0, 0; -0.5, -1; 0, -1]; % triangle
```

```
Ptri = Ptri * headWidth;
Ptri = [Ptri, zeros(5, 1)];
Ptri = turn(Ptri, [0, 1, 0], u);
Ptri = Ptri(:, 1:2) + [x + vx, y + vy];
end
```

Matlab 球坐标中的分布图

- 本词条缺少预备知识，初学者可能会遇到困难.
- 本词条处于草稿阶段.

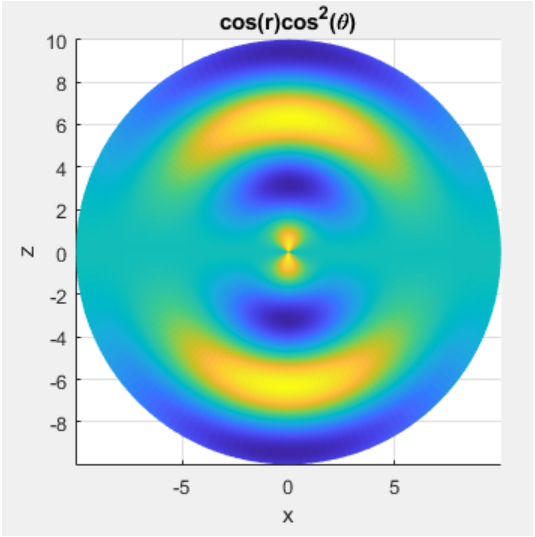


图 1: 示例 1

我们可以用 Matlab 内建的 `surf(X,Y,Z,Val)` 函数在空间中画出任意曲面. 其中 `X,Y,Z,Val` 是尺寸相同的 2 维矩阵，分别代表曲面上每个网格的各点.

```
% surf() in spherical coordinate
% usage 1. R,Th,Ph,Val are 2D matrices in same shape
% usage 2. R,Th,Ph are vector,vector,scalar
% usage 3. R,Th,Ph are vector,scalar,vector
function h = surfSph(R,Th,Ph,Val,varargin)
% resize
```

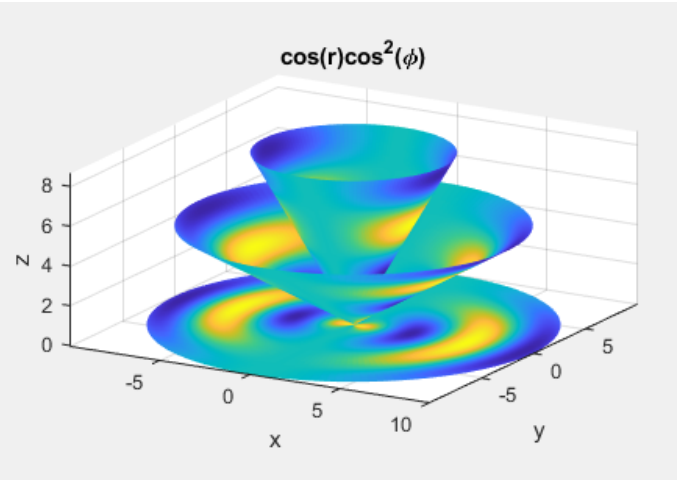


图 2: 示例 2

```
if isvector(R) && isvector(Th)
    [R,Th] = ndgrid(R, Th);
end
if isvector(R) && isvector(Ph)
    [R,Ph] = ndgrid(R,Ph);
end
if isscalar(Th)
    Th = Th*ones(size(R));
end
if isscalar(Ph)
    Ph = Ph*ones(size(R));
end
if (size(R) ~= size(Th) || size(Th) ~= size(Ph)...
        || size(Ph) ~= size(Val))
    error('wrong shape for Th, Ph, Val');
end
% plot
[X,Y,Z] = sph2cart(Ph,pi/2-Th,R);
h = surf(X,Y,Z,varargin{:});
shading flat; axis equal;
xlabel x; ylabel y; zlabel z;
cameratoolbar;
end
```

示例 1

```
Nr = 200; Nth = 150;
R = linspace(0, 10, Nr);
Th = linspace(0, 2*pi, Nth);
[R, Th] = ndgrid(R, Th);
Ph = 0;
Val = cos(R) .* cos(Th).^2;
figure; surfSph(R,Th,Ph,Val);
title('cos(r)cos^2(\theta)');
view(0, 0);
```

运行结果如图 1 .

示例 2

```
Nr = 200; Nth = 150;
R = linspace(0, 10, Nr);
Ph = linspace(0, 2*pi, Nth);
[R, Ph] = ndgrid(R, Ph);
Val = cos(R) .* cos(Ph).^2;
figure; surfSph(R,pi/6,Ph,Val);
hold on; surfSph(R,pi/3,Ph,Val);
surfSph(R,pi/2,Ph,Val);
xlabel x; ylabel y; zlabel z;
title('cos(r)cos^2(\phi)');
view(30, 16);
```

运行结果如图 2 .

Matlab 的程序调试及其他功能

预备知识 Matlab 的函数[1700]

程序调试

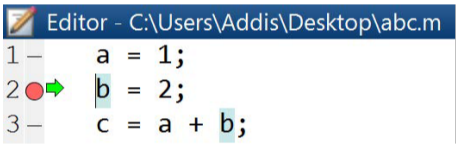


图 1: 在行首设置 Breakpoint

若要调试程序，可选择一行代码并单击该行前面的横线，这时会出现红色圆点 Breakpoint （图 1），程序运行到 Breakpoint 会暂停.

此时要查看变量情况，可通过 Workspace 查看各个变量的情况，也可用光标悬停在某个变量上. 还可以用 Command Window 改变某些变量的值，或画图等. 在这种调试状态下，也可以通过 Edit 菜单中的一些按钮控制接下来程序如何运行（图 2）. 其中“Continue”（快捷键 F5）是继续运行直到下一个

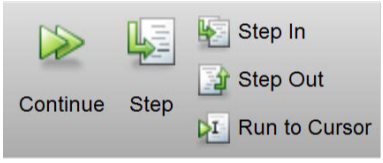


图 2: Step 菜单

Breakpoint 或结束. “Step”（F10）是运行到下一行，“Step In”（F11）是进入子程序并暂停，“Step Out”是运行完当前子程序并回到子程序被调用的地方.“Run to Cursor”是运行到光标所在处.

warning 和 error 函数

有时候当我们的程序中出现了某个错误时，程序会终止并在控制行中返回一个错误信息，例如我们给 `sin` 函数输入两个变量时，控制行的 `error` 信息将提示变量个数太多.

```
>> sin(1,2)
Error using sin
Too many input arguments.
```

另一些情况下当错误不是那么严重时，我们会得到一个 `warning` 提示，程序在输出提示后将继续运行.

我们可以在我们自己的脚本或函数中用 error 或 warning 函数达到同样的效果

```
function myfun(char)
if char == 'w'
    warning('this is a warning');
elseif char == 'e'
    error('this is an error');
else
    disp('Hello World!');
end
end
```

当 error 或 warning 函数被执行时，控制行不但会输出对应的字符串，还会输出它们所在的文件和行号，如

```
>> myfun w
Warning: this is a warning
> In myfun (line 3)
```

单击 myfun 或 line 3，可在 Editor 中显示对应位置．若 myfun 函数被其他函数调用，那么调用的位置也会被逐级列出．使用 warning 的效果类似，但程序不会被终止．

分节

在行首用两个百分号 %% 可以对代码进行分节（图 3）．这样做一是可以使代码结构更清晰，二是可以单独选择某一节运行（Edit 菜单中的“Run Section”按钮）．

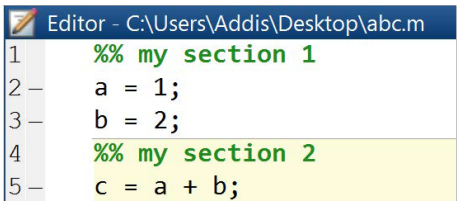


图 3: 代码分节

工具箱（Toolbox）

在购买和安装 Matlab 软件时，可以选择各种各样的工具箱，常用的工具箱有曲线拟合（Curve Fitting，从离散的数据点得到一条曲线），图像处理（Image Processing，图像变换，增强，降噪，二值化等），图像获取（Image Acquisition，从相机获取图像），Matlab 编译器（MATLAB Compiler，编译代码，提高运行速度）．注意使用了工具箱功能的代码在没有工具箱的 Matlab 软件上将无法运行．

第二章

Python 简介

Python 简介

到 2019 年，在一些排名中，Python 已经成为最流行的语言，无论是在科学计算还是在计算机领域都有大量的使用者．本书中我们主要介绍 python 在科学计算中的应用．

如果不想安装软件可以直接用浏览器访问 Jupyter Notebook 运行 Python 程序，要在本地使用 Python 推荐安装 Anoconda．以下我们用前者进行讲解．Jupyter Notebook 的优点是交互式编程，即每输入一个命令都可以立即执行（快捷键 Shift + Enter），利于学习和实验．

作为计算器

请在 Jupyter Notebook 中尝试输入以下命令并执行（运行结果略）．Python 程序使用 # 注释一行，或者用两个 """ 或 ''' 注释多行[1]．注释是为了辅助人阅读代码，不会被执行．

四则运算

```
2 + 2
```

```
123 / 456 # 注意得到的是浮点数
123 // 456 # 两数相除后向下取整
```

幂运算

```
3 ** 2
```

整数除法，即相除再向下取整

```
4 // 3
```

求余

```
4 % 3 # 使得 a == a // b + a % b 恒成立
```

使用括号

[1]事实上 """...""" 或 '''...''' 是一个字符串[1726] 而不是注释，但单独出现的字符串并不会对程序运行造成影响，所以可以用作注释

```
(123 - 234*2)**2 / (34 + 54**4)
```

各种常见的数学函数都在 math 模块中，需要加载.

```
import math
```

使用模块中的函数，要在前面加上模块名和一点. 例如开方（square root）

```
math.sqrt(284)
```

自然指数函数

```
math.exp(5.1)
```

这样做虽然略显麻烦，但可以区分不同模块中同名函数. 在确保没有冲突的情况下我们也可以用以下方法加载模块中的指定函数，如

```
from math import sqrt, exp, sin, cos
```

现在使用这些函数就不需要 `math.` 的前缀了

```
sin(1)
```

我们甚至可以用这种方式引入一个模块中所有函数（和其他内容）而无需前缀. 这样做引起名称冲突的可能性更大，不建议使用.

```
from math import *
```

从模块中不仅可以引入函数，还有常数，例如圆周率和自然对数底（注意 `e` 这种单字母名称很可能会产生冲突，所以不建议取消 `math.` 前缀）

```
sin(pi/2)
```

```
log(e)
```

`math` 模块中的其他常用函数如：绝对值（absolute value）

```
fabs(-32)
```

自然对数

```
log(0.5)
```

以 10 为底的对数

```
log10(1000)
```

弧度转为角度

```
degrees(pi/2)
```

变量

Python 中的**变量**（**variable**）和数学中的不同，Python 的变量可以理解为一个储存数值的容器，我们可以用等号把一个数值储存在一个变量中．例如要计算一个长方体的面积，我们既可以直接把三个数字相乘，也可以先把这三个数字赋值给三个**变量**然后相乘

```
a = 1
b = 2
c = 3
volumn = a*b*c
```

在 Jupyter Notebook 执行这四行发现并没有输出，无论是一次性执行还是分开执行．这是因为赋值命令默认不输出结果．要强制输出结果可以用

```
print(volumn)
```

或者一次输出多个变量

```
print(a, b, c, volumn)
```

注意：在 python2.x 版本中，`print` 是一个关键字，输出内容不需要加括号，python 3.x 中它是一个函数，可以介绍任意多个用逗号隔开的变量，他们甚至可以是不同类型的．例如

```
print volumn  #2.x
print(volumn) # 3.x
```

要强调的是，这里的等号并不是数学上的等于，而是**赋值**，即把等号右边得到的数值储存在左边的变量中．在第 4 行执行时，计算机会先计算等号右边表达式的结果 6，然后将 6 储存在变量 `volumn` 中．`volumn` 这个变量中并不会包含

`a*b*c` 这个信息，只储存 6 这个数值．所以改变 `a, b, c` 后 `volumn` 的值并不会自动改变．

如果要让长方形的某个边长增加 1，我们可以执行

```
a = a + 1
```

如果将等号理解为数学上的等于，这个式子显然是错的．但正确的理解是，先把 `a` 当前的值 1 加上 1 得到 2，然后把 2 **赋值**给 `a`．由于我们没有给 `volumn` 重新赋值，它仍然是 6，要更新 `volumn`，只需要重新执行

```
volumn = a*b*c
```

并用 `print(volumn)` 显示新的值．

等效地，我们也可以用**自加运算** `+=`，将 `a = a + 1` 记为

```
a += 1
```

注意 `+=` 是一个整体的算符，中间不能有空格．类似的运算还有 `-=`，`*=`，`/=` 等．

函数

Python 中的函数与数学中的函数不完全一样，函数可以有若干个输入变量和输出变量（也可以没有）．下面我们定义一个简单的函数来计算长方形的面积

```
def f(a, b, c):
    volumn = a*b*c
    return volumn
```

这段代码用到了两个 Python 的**关键字（keyword）** `def` 和 `return`．关键字是指在程序中有特殊含义的单词，不能作为变量名和函数名的名称．其中 `def` 用于定义函数，`f` 是函数名，`a`，`b` 和 `c` 分别是函数的**输入变量（argument）**．冒号以后是**函数体**，可以有若干行命令．注意这些命令前面必须有**缩进（indentation）**．在以上代码中，函数体的第一行计算面积，第二行将用关键字 `return` 将计算的结果作为输出并退出函数．

现在我们可以使用这个函数，使用方法和 `sin`，`sqrt` 等数学函数一样，只是不同输入变量要用逗号隔开．

```
V = f(1.2, 3.4, 6)
print(V)
```

Python 基本变量类型

Python 基本变量类型一共有，我们可以用 `type()` 函数查看某个变量的类型．例如执行

```
n = 123; x = 3.14; print(type(n)); print(type(x))
```

结果为

```
<class 'int'>
<class 'float'>;
```

整数

与一些编译语言不同，Python 的整数类型（integer）没有长度限制（除超出了内存大小）．例如

```
print(123456789012345678901234567890 + 1)
```

的结果为 `123456789012345678901234567891`．

默认情况下整数用十进制表示，如果需要输入 **2 进制（binary）**，可以在前面加 `0b` 或 `0B`，例如 `0b1001` 表示 10 进制的 `9`．类似地，`0o` 或 `0O` 开头表示 **8 进制（octal）**；`0x` 或 `0X` 开头的表示 **16 进制（hexadecimal）**，16 进制中的 10 到 15 分别用大写或小写字母 `a` 到 `f` 表示．例如 `0xff` 表示十进制的 `255`．不同进制的整数同样没有长度限制．

类型转换

转换格式为 `类型 (变量)`．例如 `int('123')` 会把字符串 `'123'` 变为整数 `123`．

Python 数据类型

下面介绍 Python 中一些常用数据类型.

列表（List）

List（列表）是 Python 中使用最常用的的数据类型之一. 列表中元素的类型可以不相同，它支持数字，字符串甚至可以包含列表. 列表是写在方括号 [] 之间、用逗号分隔开的元素列表.

```
list1 = [1,2,3,4]
list2 = ['a','b',1,3]
print (list1)              # 输出完整列表
print (list1[0])           # 输出列表第一个元素
print (list2[1:3])         # 从第二个开始输出到第三个元素
print (list2[2:])          # 输出从第三个元素开始的所有元素
print (list2 * 2)          # 输出两次列表
```

输出为:

```
[1, 2, 3, 4]
1
['b', 1]
[1, 3]
['a', 'b', 1, 3, 'a', 'b', 1, 3]
```

可以用 list1.append(...) 在列表最后插入一个或多个元素.

元组（Tuple）

元组（tuple）与列表类似，不同之处在于元组的元素**不能修改**. 元组写在小括号 () 里，元素之间用逗号隔开. 元组中的元素类型也可以不相同:

```
tup1 = (1,2,3,4)
tup2 = ('a','b',1,3)
print(tup1)              # 输出完整元组
print (tup1[0])          # 输出元组第一个元素
```

```
print (tup2[1:3])        # 从第二个开始输出到第三个元素
print (tup2[2:])         # 输出从第三个元素开始的所有元素
print (tup2 * 2)
```

输出

```
(1, 2, 3, 4)
1
('b', 1)
(1, 3)
('a', 'b', 1, 3, 'a', 'b', 1, 3)
```

字典（Dictionary）

字典是无序的对象集合．字典当中的元素是通过键来存取的，用 {} 标识，它是一个无序的 key:value 的集合，key 和 value 分别译成**键**和**值**．key 必须使用不可变类型．在同一个字典中，key 必须是**唯一**的．例如统计一个班学生的成绩可以使用字典表示．

```
scores = {'语文': 89, '数学': 92, '英语': 93}
print(scores['语文'])
```

输出：

```
89
```

Numpy 库

数组

我们这里使用 numpy 模块中的数组．numpy 是数值计算中一个常用的模块，主要用于处理各种数组（或矩阵）．numpy 提供的是 ndarray 对象，即 n 维数组对象．这里介绍另一种 import 的方式

```
import numpy as np
```

这相当于给 numpy. 指定一个更简单的名字 np.，方便了使用．你也可以把 np 替换成其他名字．上文使用 math 模块时也可以使用

```
import numpy as m
```

要创建一个一维数组（或者叫做行矢量），如

```
a = np.array([1, 2, 3])
```

这个数组有三个元素，每个元素相当于一个独立的变量．要获取第 i 个元素的值，用 a[i]，如

```
a[0]
```

注意索引 i 从 0 开始而不是 1．所以第一个元素索引为 0，第二个索引为 1 等等．

显示整个数组

```
print(a)
```

要获取数组的一部分

```
a[1:2]
```

注意：python 对数组部分索引获取的时候，不会获取索引范围的最后一个索引值，即上面这个命令只会得到 a[1] 的值为 2，而没有 a[2] 的值．

要间隔一个步长获取数组的一部分，可以使用

```
a = np.array([1,2,3,4,5,6,7,8])
a[1:6:2]
```

括号中的第三个参数表示步长，2 表示隔一个元素取一个元素，上述代码取出 a 中第 1 个（注意不是第 0 个），第 3 个，第 5 个值，返回结果为 2,4,6.

创建一个二维数组，如

```
b = np.array([[1, 2, 3, 4],[4, 5, 6, 7], [7, 8, 9, 10]])
```

数组的大小可以通过其 shape 属性获得．例如获取 a 和 b 的数组大小，

```
a.shape
b.shape
```

最终会得到返回值为

```
(8,)
(3, 4)
```

注意到返回值都有一个小括号，说明返回值类型为一个**元组类型**. 数组 a 的 shape 只有一个元素，因此它是一维数组. 而数组 c 的 shape 有两个元素，因此它是二维数组，其中第 0 轴的长度为 3，第 1 轴的长度为 4.

修改数组的 shape 属性，在保持数组元素个数不变的情况下，改变数组每个轴的长度. 下面的例子将数组 c 的 shape 改为 (4,3)，注意从 (3,4) 改为 (4,3) 并**不是对数组进行转置**[396]，而只是改变每个轴的大小，数组元素在内存中的位置并没有改变：

```
b.shape = 4,3
print(b)
```

返回

```
array([[ 1, 2, 3],
[ 4, 4, 5],
[ 6, 7, 7],
[ 8, 9, 10]])
```

这种操作会永久的改变原数组的 shape. 当某个轴的元素为-1 时，将根据数组元素的个数自动计算此轴的长度，因此下面的程序将数组 b 的 shape 改为 (4,3)：

```
c.shape = 4, -1
```

另外，通过 reshape 方法，可以创建一个改变了尺寸的新数组，原数组的 shape 保持不变. 例如

```
d = b.reshape((4,3))
print(b)
print(d)
```

返回

```
array([[1, 2, 3, 4],
[4, 5, 6, 7],
[7, 8, 9, 10]])
```

```
4
5 array([[ 1, 2, 3],
6 [ 4, 4, 5],
7 [ 6, 7, 7],
8 [ 8, 9, 10]])
```

其实数组 b 和 d 其实共享数据存储内存区域，因此修改其中任意一个数组的元素也会同时修改另外一个数组的内容．例如，

```
1 b[0]=999
2 print(b)
3 print(d)
```

输出

```
1 array([[999, 2, 3, 4],
2 [4, 5, 6, 7],
3 [7, 8, 9, 10]])
4
5 array([[ 999, 2, 3],
6 [ 4, 4, 5],
7 [ 6, 7, 7],
8 [ 8, 9, 10]])
```

一些常用的创建数组的 numpy 函数

arange 函数通过指定开始值、终值和步长来创建一维数组，注意数组**不包括**终值：

```
1 np.arange(0,10,1)
2 array([ 0,1, 2, 3, 4, 5, 6, 7, 8, 9])
```

linspace 函数通过指定开始值、终值和元素个数来创建一维数组，可以通过 endpoint 关键字指定是否包括终值，默认是包括终值：

```
1 np.linspace(0,1,10)
2 array([0.         , 0.11111111, 0.22222222, 0.33333333,
   0.44444444,
```

```
       0.55555556, 0.66666667, 0.77777778, 0.88888889, 1.
                  ])
```

zeros 创建全 0 数组

```
np.zeros(3)
array([0., 0., 0.])
np.zeros((3,5))
array([[0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0.]])
```

另外，ones 和 eye创建全 1 数组，对角元为 1 的数组，用法与zeros类似.

矩阵的操作

下面分别介绍对矩阵[396] 进行一些常见的操作处理，包括转置，拼接，翻转，内积[383]，外积，计算本征值[423]，行列式[432]，逆矩阵[405]，等等在 python 中的实现.

```
A = [1, 2, 3]
B = [[2, 1, 4],[2, 2, 3],[1, 1, 3]]
C = [4, 5, 6]
```

转置

```
np.transpose(B)
```

拼接

```
D = np.concatenate((A,C)) # Concatenate vectors
print(D)
[1 2 3 4 5 6]
```

左右或者上下翻转矩阵，字母 u, d, l, r 分别是 up，down，left，和 right 的缩写.

```
D = np.flipud(A)
print(D)
E = np.fliplr(B)
print(E)
```

输出

```
[3 2 1]
array([[4, 1, 2],
       [3, 2, 2],
       [3, 1, 1]])
```

内积（也叫点乘）[383]，外积（也叫张量积）

```
D = np.dot(A,C)
print(D)
D = np.outer(A,C) # Outer product
print(D)
```

输出

```
32
[[ 4  5  6]
 [ 8 10 12]
 [12 15 18]]
```

计算本征值[423]，逆矩阵[405]，行列式[432]

```
D = np.linalg.eigvals(B)
print(D)
D = np.linalg.inv(B)
print(D)
D = np.linalg.det(B)
print(D)
```

随机数，例如产生 0-1 之间的随机数

```
D = np.random.rand(3,5)
print(D)
```

创建其他随机数方法类似，例如正态分布随机数用 np.random.randn 函数.

Python 字符串处理

预备知识 Python 简介[1715]

字符串处理

字符串（string）可以想 i 想由字符构成的特殊数组，一般用于储存一段文字．一个字符串变量如

```
s = 'python String 字符串'
```

注意也可以使用双引号 "..."．用 type(s) 查看类型得到 str，即 string 的缩写．

input() 函数可以在程序运行的过程中获取用户输入的字符串，例如

```
s = input('请输入：')
print('您输入的内容是：', s)
```

如果希望用户输入其他类型，如整数，那么输入后再进行变量类型转换即可，如 int(s)．

len(s) 用于获取字符串长度

```
print('s = ', s, '; length = ', len(s))
```

也可以用三个单引号或双引号表示多行字符串，如

```
s = '''这是第一行
这是第二行
这是第三行'''
```

注意如果第一个 ''' 后面必须紧接第一个字符，第二个 ''' 必须紧跟最后一个字符．

转义符

如果在字符串中要表示一些特殊的字符，我们可以使用 \ 开头的**转义符（escape character）**，例如 \n 表示换行，\' 表示单引号，\" 表示双引号，\\ 表示 \．如果字符串两边用了双引号，内部的单引号不需要转义，反之亦然．

```
print('it's good') # 错误， python 会把 'it' 理解为一个字符串
print("it's good") # 正确
```

```
print('it\'s good') # 正确
```

如果不想使用转义符，可以在第一个单引号或双引号前面加 r 或 R. 例如 r'a\nb' 中的 \n 不代表换行而是 \ 和 n 两个字符.

截取字符串

字符串的截取与数组的截取相同

```
s = '0123456789'
print(s[0:3]) # 截取第一位到第三位的字符
print(s[:]) # 截取字符串的全部字符
print(s[6:]) # 截取第七个字符到结尾
print(s[:-3]) # 截取从头开始到倒数第三个字符之前
print(s[2]) # 截取第三个字符
print(s[-1]) # 截取倒数第一个字符
print(s[::-1]) # 创造一个与原字符串顺序相反的字符串
print(s[-3:-1]) # 截取倒数第三位与倒数第一位之前的字符
print(s[-3:]) # 截取倒数第三位到结尾
```

输出:

```
012
0123456789
6789
0123456
2
9
9876543210
78
789
```

连接字符串

```
delimiter = ','
mylist = ['Brazil', 'Russia', 'India', 'China']
```

```
print (delimiter.join(mylist))
print('abc'+'def')
输出
Brazil,Russia,India,China
abcdef
```

字符串搜索

- 搜索指定字符串：s.find('t')，搜不到返回 -1
- 指定起始位置搜索：s.find('t', start)
- 指定起始及结束位置搜索：s.find('t', start, end)
- 从右边开始查找：s.find('t')
- 搜索到多少个指定字符串：s.count('t')

```
s='python String function'
print('%s find nono=%d' % (s,s.find('nono')))
print('%s find t=%d' % (s,s.find('t')))
print('%s find t from %d=%d' % (s,1,s.find('t',1)))
print('%s find t from %d to %d=%d' % (s,1,2,s.find('t',1,2)))
print('%s rfind t=%d' % (s,s.rfind('t')))
print('%s count t=%d' % (s,s.count('t')))
```

输出：

```
python String function find nono=-1
python String function find t=2
python String function find t from 1=2
python String function find t from 1 to 2=-1
python String function rfind t=18
python String function count t=3
```

字符串替换

替换 old 为 new：`s.replace('old','new')`
替换指定次数的 old 为 new：`s.replace('old','new',maxReplaceTimes)`

字母处理

- 全部变为大写：`s.upper()`
- 全部变为小写：`s.lower()`
- 大小写互换：`s.swapcase()`
- 首字母大写，其余小写：`s.capitalize()`
- 首字母大写：`s.title()`

```
1 s = 'python String 字符串'
2 print('s =          ', s);
3 print('lower =      ', s.lower()); # 全部变为小写:
4 print('upper =      ', s.upper()); # 全部变为大写:
5 print('swapcase =   ', s.swapcase()); # 大小写互换
6 print('capitalize = ', s.capitalize()); # 首字母大写, 其余小写
7 print('title =      ', s.title()); # 首字母大写
```

输出：

```
s =          python String 字符串
lower =      python string 字符串
upper =      PYTHON STRING 字符串
swapcase =   PYTHON sTRING 字符串
capitalize = Python string 字符串
title =      Python String 字符串
```

格式化相关

获取固定长度，右对齐，左边不够用空格补齐：s.ljust(width)
获取固定长度，左对齐，右边不够用空格补齐：s.ljust(width)
获取固定长度，中间对齐，两边不够用空格补齐：s.ljust(width)
获取固定长度，右对齐，左边不足用 0 补齐

```
s='python String'
print('%s ljust = %s' % (s,s.ljust(20)))
print('%s rjust = %s' % (s,s.rjust(20)))
print('%s center = %s' % (s,s.center(20)))
print('%s zfill = %s' % (s,s.zfill(20)))
输出:
python String ljust = python String
python String rjust =         python String
python String center =     python String
python String zfill = 0000000python String
```

字符串去空格及去指定字符

去两边空格：s.strip()
去左空格：s.lstrip()
去右空格：s.rstrip()
去两边字符串：s.strip('d')，相应的也有lstrip, rstrip
按指定字符分割字符串为数组：s.split(' ')

```
s = ' python String function '
print('%s strip=%s' % (s,s.strip()))
s = 'python String function'
print('%s strip=%s' % (s,s.strip('d')))
```

翻转字符串

```
sStr1 = 'abcdefg'
sStr1 = sStr1[::-1]
print(sStr1)
```

```
输出
gfedcba
```

关于更多高效的字符串处理方法，感兴趣的读者可以查阅正则表达式的使用方法.

Python 文件读写

预备知识 Python 入门[1715]

学习处理文件和保存数据可让你的程序使用起来更容易：用户将能够选择输入什么样的数据，以及在什么时候输入; 用户使用你的程序做一些工作后，可将程序关闭，以后再接着往下做.

Python 文件读写的方法

open() 方法

Python `open()` 方法用于打开一个文件，并返回**文件对象（file object）**，在对文件进行处理过程都需要使用到这个函数，如果该文件无法被打开，会抛出错误.

`open()` 函数常用形式是接收两个参数：**文件名（file）**和**模式（mode）**.

```
open(file, mode='r')
```

此方法有很多其他参数，一般情况下用不到. 其中`file` 必需，是文件路径（相对或者绝对路径）；`mode` 可选，代表文件打开模式. 打开模式有多种方式.

表 1: 打开模式

r	以只读方式打开文件
w	打开一个文件只用于写入. 如果该文件已存在则打开文件，并从开头开始编辑，即原
b	二进制模式
+	打开一个文件进行更新 (可读可写
a	打开一个文件用于追加. 如果该文件已存在，新的内容将会被写入到已有内容
	上述方式组合使用

使用 open() 方法一定要保证关闭文件对象，即调用 close() 方法．例如计算机上有一个文件名为file.txt的文件，内容如表 1 节 所示，需要读取并输出．代码如下



图 1: file 文件内容

```
fo=open(r'C:\Users\shizy\Desktop\file.txt')
print("文件名: ", fo.name)
print("是否已关闭 : ", fo.closed)
print("访问模式 : ", fo.mode)
str = fo.read()
print(str)
fo.close()
```

输出如下

```
文件名:  C:\Users\shizy\Desktop\file.txt
是否已关闭 :  False
访问模式 :  r
3.1415926
1234567
abcdefg
```

上述代码第一个open方法前面我们添加了一个r，表示原生字符串，否则\会对字符串进行转义处理．当然也可以不加r，而使用两个\，效果一样．我们在程序 2-4 行分别使用了文件的三种其他方法：

file.closed 返回 true 如果文件已被关闭，否则返回 false.

file.mode 返回被打开文件的访问模式.

file.name 返回文件的名称.

write() 方法

write()方法可将任何字符串写入一个打开的文件. 需要重点注意的是, Python 字符串可以是二进制数据, 而不是仅仅是文字. write()方法不会在字符串的结尾添加换行符 \n. 语法如下:

```
fo = open("file2.txt", "wb")
fo.write( "Hello Python\n");
# 关闭打开的文件
fo.close()
```

readline() 方法

readline() 会从文件中读取单独的**一行**. 换行符为'\n'. readline() 如果返回一个空字符串, 说明已经已经读取到最后一行.

readlines() 方法

readlines() 将以列表的形式返回该文件中包含的**所有行**, 列表中的一项表示文件的一行. 如果设置可选参数 sizehint, 则读取指定长度的字节, 并且将这些字节按行分割.

上面介绍的文件读写方法虽然没问题, 但是一般情况下我们不使用, 而是使用with 方法. 从而避免忘记关闭文件导致的不可预料的结果.

从文件中读取数据

读取整个文件

下面的程序打开并读取这个文件, 再将其内容显示到屏幕上:

```
with open('file.txt') as f:
contents = f.read()
print(contents)
```

在这个程序中,关键字with 在不再需要访问文件后将其关闭. 在这个程序中,注意到我们调用了open(),但没有调用close(); 你也可以调用open() 和close() 来打开和关闭文件, 但这样做时, 如果程序存在错误, 导致close() 语句未执行, 文件将不会关闭. 这看似微不足道, 但未妥善地关闭文件可能会导致数据丢失或受损. 如果在程序中过早地调用close(), 你会发现需要使用文件时它已关闭, 无法访问, 这会导致更多的错误. 并非在任何情况下都能轻松确定关

闭文件的恰当时机，但通过使用前面所示的结构，可让 Python 去确定：你只管打开文件，并在需要时使用它，Python 自会在合适的时候自动将其关闭.

逐行读取

读取文件时，常常需要检查其中的每一行：你可能要在文件中查找特定的信息，或者要以某种方式修改文件中的文本.

```
with open('file.txt') as f:
    for line in f:
        print(line)
```

利用with来对文件的操作与上一节类似，这里不再赘述.

数据存储

很多程序都要求用户输入某种信息. 不管专注的是什么，程序都把用户提供的信息存储在列表和字典等数据结构中. 用户关闭程序时，你几乎总是要保存他们提供的信息; 一种简单的方式是使用模块json 来存储数据.

模块json 让你能够将简单的 Python 数据结构转储到文件中，并在程序再次运行时加载该文件中的数据. 你还可以使用json 在 Python 程序之间分享数据. 更重要的是，json数据格式并非 Python 专用的，这让你能够将以json格式存储的数据与使用其他编程语言的人分享. 这是一种轻便格式，很有用，也易于学习.

使用 json.dump() 和 json.load()

我们来编写一个存储一组数字的简短程序，再编写一个将这些数字读取到内存中的程序. 第一个程序将使用json.dump() 来存储这组数字，而第二个程序将使用json.load().

函数json.dump() 接受两个实参：要存储的数据以及可用于存储数据的文件对象. 下面演示了如何使用json.dump() 来存储数字列表.

```
import json
numbers = [2, 3, 5, 7, 11, 13]
filename = 'numbers.json'
```

```
with open(filename, 'w') as f_obj:
    json.dump(numbers, f_obj)
```

我们先导入模块json，再创建一个数字列表．在第三行处，我们指定了要将该数字列表存储到其中的文件的名称．通常使用文件扩展名.json来指出文件存储的数据为json格式．接下来，我们以写入模式打开这个文件，让json 能够将数据写入其中．在第五行处，我们使用函数json.dump() 将数字列表存储到文件numbers.json中．

下面再编写一个程序，使用json.load() 将这个列表读取到内存中：

```
import json
filename = 'numbers.json'
with open(filename) as f_obj:
   numbers = json.load(f_obj)
print(numbers)
```

Python 判断与循环

循环语句

我们经常需要对某个操作重复执行多次，可以用 for 循环，例如输出 1 到 3 的平方：

```
for i in range(1,4):
    print(i,i**2)
```

输出

```
1 1
2 4
3 9
```

注意，python 中用缩进来表示代码的范围，通常为一个 tab 制表符（即键盘上的 TAB 键），或者四个空格缩进；for 循环后面冒号必不可少．for 循环还可以遍历一个列表，对每个元素执行相同的操作．例如

```
A = np.array([3, 2, 4, 5, 1, 76])
for aa in A:
    print(3*aa)
```

这里对列表每一个元素扩大 3 倍. 如果用不得循环指标的时候可以用下划线代替:

```
for _ in range(3):
    print('Hello Python')
```

此时连续输出 'Hello Python' 三次.

有时候我们并不知道循环多少次，此时可以用 while 循环. 例如找到最小的 n，使得

$$\sum_{i=1}^{n} i^2 < 50 \tag{1}$$

成立. 它可以通过 while 循环实现. 代码如下

```
a = 0
sum0 = 0.0
while sum0 < 50:
    a = a + 1
    sum0 = sum0+a**2
    print(a, sum0)
```

输出

```
1 1.0
2 5.0
3 14.0
4 30.0
5 55.0
```

有时我们在程序中做什么将取决于前面步骤的结果. 在这种情况下，您可能会发现 if 语句很有帮助. 假设我想知道一个二次方程 $ax^2 + bx + c = 0$ 有多少实根. 下面的代码根据 a、b 和 c 的值告诉我根的数量.

```
a = 5
b = 7
```

```
c = -1
D = b**2 - 4*a*c
if D > 0:
    print("two roots")
if D < 0:
    print("no roots")
if D == 0:
    print("one root")
```

通常 `if` 可以嵌套使用；循环语句可以与条件判断语句结合使用. 例如，一个简单的打分系统：

$$f = \begin{cases} A, & x > 90; \\ B, & x > 80; \\ C, & x > 60. \end{cases} \tag{2}$$

```
x =99
if x>90:
    print('A')
elif x>80:
    print('B')
else:
    print('C')
```

Python 函数

预备知识 Python 入门[1715]

函数是组织好的，可重复使用的，用来实现相关功能的代码段. 它能提高代码的重复利用率. Python 提供了许多**内建函数**，比如 `print()`. 我们也可以自己创建函数，这被叫做**自定义函数**. 需要在程序中多次执行同一项任务时，你无需反复编写完成该任务的代码，而只需调用执行该任务的函数，让 Python 运行其中的代码. 你将发现，通过使用函数，程序的编写、阅读、测试和修复都将更容易.

函数的定义与调用

我们通过一个简单的例子开始介绍：

```
def  func1():
    # 定义一个函数，函数名为func1
     print('hello python')
func1()
```

这个示例演示了最简单的函数结构．第一行的代码行使用关键字def 来告诉 Python 我们要定义一个函数．然后向 Python 指出了函数名，还可能在括号内指出函数为完成其任务需要什么样的信息．在这里，函数名为func1，它不需要任何信息就能完成其工作，因此括号是空的（括号必不可少）；最后，定义以**冒号**结尾，后面的所有**缩进行**构成了函数体．第二行处的文本是**函数注释**，以#开始，通常用来描述了函数是做什么的．第三行print('hello python') 是函数体内的唯一一行代码．**函数调用**让 Python 执行函数的代码．要调用函数，可依次指定函数名以及用括号括起的必要信息，由于这个函数不需要任何信息，因此调用它时只需输入func1() 即可．

向函数传递信息

只需对上面函数稍作修改,就可以让函数func1() 不仅向用户显示hello python，还将可以打印其他信息．在函数定义def func1() 的括号内添加信息．通过在这里添加信息，就可让函数接受你给定的任何值．

```
def  func2(str):
    '''
     改进第一个函数，使得它可以输出更多信息
     输入：字符串
    '''
     print('hello'+str)
func2('C++')
func2('Tom')
```

输出

```
hello C++
```

```
2 hello Tom
```

上述代码在第一行函数定义处多了一个信息，函数功能变得更加丰富．同时我们注意到第二个用三个**单引号**开头，不再是#．三个单引号是用来进行多行注释的．另外注意到 Python 中字符串可以通过+来连接．

实参和形参

前面定义函数func2时，要求函数定义括号里面指定一个信息．调用这个函数并提供这种信息．实际上这个所谓的信息是一个**形参**，而在函数调用过程中，如func2('C++')，值 C++ 是一个**实参**．实参是调用函数时传递给函数的信息．我们调用函数时，将要让函数使用的信息放在括号内．通俗的说，将函数定义中的变量称为形参，将函数调用中的变量称为实参．

传递参数

函数定义中可能包含多个形参，因此函数调用中也可能包含多个实参．向函数传递实参的方式很多，可使用**位置实参**，这要求实参的顺序与形参的顺序相同; 也可使用**关键字实参**，其中每个实参都由变量名和值组成．

位置实参

在调用函数时，Python 必须将函数调用中的每个实参都关联到函数定义中的一个形参．为此，最简单的关联方式是**基于实参的顺序**，这种关联方式被称为位置实参．如一个用来展示个人基本信息的函数定义如下：

```
1 def ShowInfo(name,age,ID):
2     '''
3     定义一个函数，来展示个人信息
4     name: 姓名
5     age: 年龄
6     ID:  学号
7     '''
8     print("my name is "+name+",my age is "+str(age)," and my ID
          is"+ID)
```

```
ShowInfo('XiaoMin', 16,'123456')
```

输出my name is XiaoMin,my age is 16 and my ID is123456.这个函数有三个参数，name，age，ID，调用函数时候也需要指定三个参数，否则会报错，这三个参数的顺序与定义函数顺序一致，这就是位置参数.

关键字实参

关键字实参是传递给函数的**名称-值**对. 在实参中将名称和值关联起来，因此向函数传递实参时不会混淆. 关键字实参无需考虑函数调用中的实参顺序.

函数 ShowInfo()还是原来那样，但调用这个函数时，我们向 Python 明确地指出了各个实参对应的形参. 如：

```
ShowInfo(name='XiaoMin',ID='123456',age=16)
```

输出结果与之前一样，这时候不需要考虑参数的顺序.

参数默认值

编写函数时，可给每个形参指定默认值. 在调用函数中给形参提供了实参时，Python 将使用指定的实参值; 否则，将使用形参的默认值. 因此，给形参指定默认值后，可在函数调用中省略相应的实参. 使用默认值可简化函数调用，还可清楚地指出函数的典型用法. 我们再次以展示个人信息为例：

```
def ShowInfo2(name,age=16,ID):
    '''
    定义一个函数，来展示个人信息
    name: 姓名
    age: 年龄
    ID:  学号
    '''
    print("my name is "+name+",my age is "+str(age)," and my ID
        is"+ID)
```

调用函数

```
ShowInfo(name='XiaoMin',ID='123456')
ShowInfo(name='XiaoLi',ID='10086')
```

这里我们只给定了两个实参，而age使用了默认值，结果就会使用默认值.

函数返回值

函数并非总是直接显示输出，相反，它可以处理一些数据，并返回一个或一组值．函数返回的值被称为**返回值**．在 Python 函数中，可使用return语句将值返回到调用函数的代码行．返回值让你能够将程序的大部分繁重工作移到函数中去完成，从而简化主程序.

下面我们定义一个函数，通过给定分数来返回对应的成绩登记.

```
def score(num):
    '''
    打分
    :param num:
    :return: 等级
    '''
    num = float(num)
    if num>100 or num<0:
        print('输入有误！')
    if num>90:
        return 'A'
    elif num>80:
        return 'B'
    elif num>60:
        return 'C'
    else:
        return 'D'
state =score(89)
print("state",state)
```

通函数score，返回成绩等级.

传递列表

向函数传递列表很有用，这种列表包含的可能是名字、数字或更复杂的对象．将列表传递给函数后，函数就能直接访问其内容．下面使用函数来提高处理列表的效率．假设有一个用户列表，我们要问候其中的每位用户．下面的示

例将一个名字列表传递给一个名为greet_users() 的函数，这个函数问候列表中的每个人.

```
def greet_users(names):
"""向列表中的每位用户都发出简单的问候"""
for name in names:
msg = "Hello, " + name.title() + "!"
print(msg)
usernames = ['xiaomin', 'toy', 'mar']
greet_users(usernames)
```

将列表传递给函数后，函数就可对其进行修改. 在函数中对这个列表所做的任何修改都是**永久性**的，即他可以改变原始列表.

传递任意数量的实参

有时候，我们预先不知道函数需要接受多少个实参，好在 Python 允许函数从调用语句中收集任意数量的实参. 为了达到这个目的，在使用形参前加*para 但不管调用语句提供了多少实参，这个形参都将它们包括进去. 例如求和

```
def sum(*num):
    s=0
    for i in num:
        s = s+i
    return s
print(sum(1,2,3))
print(sum(1,2,3,4))
```

两次调用函数分别传递了三个，四个参数.

Python 画图

python 绘图库 matplotlib 的使用

python 绘图有很多库，其中 matplotlib 库是最受欢迎的绘图库. 如果熟悉 Matlab[1682] 软件的用户，对 matplotlib 库就很容易上手，因为 matplotlib 中很多

命令命名与用法与 matlab 十分类似．首先在使用之前需要导入模块，并取别名 plt．

```
import matplotlib.pyplot as plt
```

如果使用 jupter 等在线编辑器，一般需要加入

```
% matplotlib inline
```

来告诉解释器在浏览器中显示图像．现在我们来看看如何使用这个库．例如在 [0,5] 区间均匀取 30 个点，分别计算 $\sin(x)$ 与 $\exp(x)$, 并作图

```
A = np.linspace(0,5,30)
B = np.sin(A)
C = np.exp(-A)
plt.plot(A,B)
plt.hold(True)
plt.plot(A,C)
```

上述代码通过 hold 把两个图像放在一张画布上面，如图 1 所示．

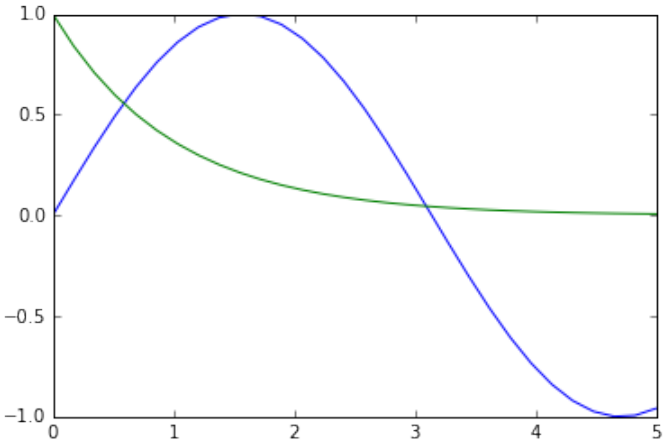


图 1: $\sin(x)$ 与 $\exp\{x\}$

下面给一个综合案例来详细说明这个库的具体使用．结果如图 2 所示．

```
import matplotlib.pyplot as plt
import numpy as np
import matplotlib as mpl

```

```
mpl.rcParams["font.sans-serif"]=['kaiti']
mpl.rcParams['axes.unicode_minus']=False
plt.subplot(211)
x = np.linspace(0,10,1000)
y = np.cos(x)
x1 = np.linspace(0,10,20)
y1= np.cos(x1)
plt.plot(x,y, ls='-',lw='2',label='曲线图')
plt.scatter(x1,y1,c='b',label='散点图')

plt.xlim(-1,11)
plt.ylim(-1.1,1.1)
plt.xlabel('x轴',fontsize=20)
plt.ylabel('y轴',fontsize=20)
plt.grid(ls='-.',c='r')
#绘制平行坐标轴直线
plt.axhline(y=0,c='g',ls='-',lw=3)
plt.axvline(x=0,c='g',ls='-',lw=3)
#绘制垂直于坐标轴的区域
plt.axvspan(xmin=0,xmax=1,facecolor='y',alpha=0.5)#facecolor或者
    fcc
plt.axhspan(ymin=0,ymax=0.2,fc='r',alpha=0.5)
#添加箭头注释
plt.annotate('注释内容',
xy=(2,0),
xytext=(3,0.2),
weight='bold',
color='b',
fontsize=20,
arrowprops=dict(arrowstyle='->',
connectionstyle='arc3',
color='b'))
#添加文本注释
plt.text(2,0.8,r'普通文本 $\sin(\pi x)$',color='b',weight='bold'
    ,fontsize=20)
plt.title('图像标题',fontsize=20)
```

```
39 plt.legend(loc='upper right',title='图例标题',fontsize=20)#里面
       可有参数
40 plt.tick_params(labelsize=20)
41
42 plt.subplot(212)
43 plt.plot(x,y,'r-',lw=2)
44
45 plt.xticks([0,5,10],[r'$\pi$',r'$2\pi$','C'],rotation=20)
46 plt.ylim(1,-1)
47 plt.xlabel('自定义坐标刻度，倒序并旋转',fontsize=20)
48 plt.text(5,0,'MATPLOTLIB',size=20,rotation=30,
49 bbox=dict(boxstyle='round',edgecolor='r',facecolor='gray'))
50 plt.tick_params(labelsize=20)
51 plt.show()
```

先看一下效果图

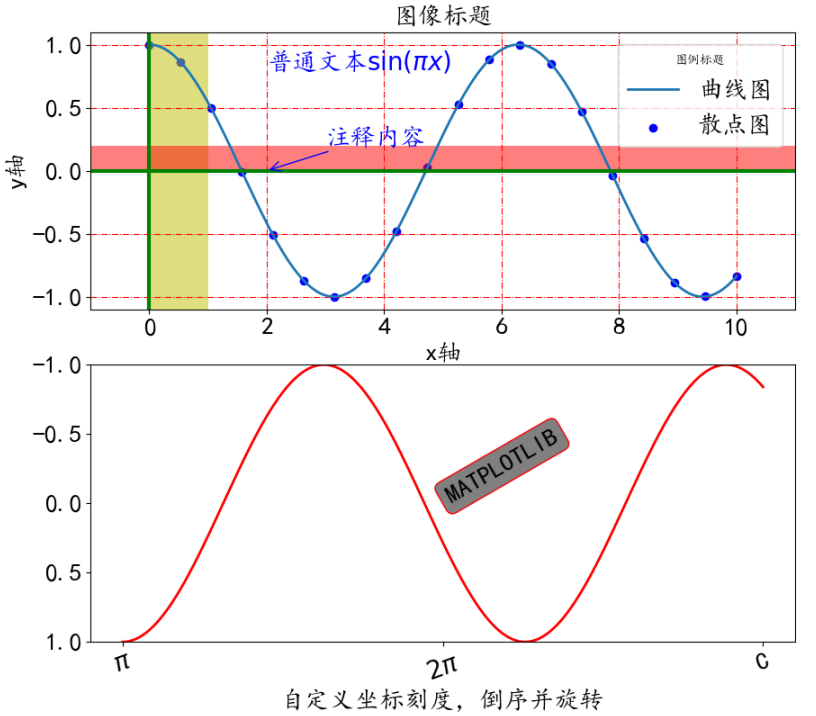


图 2: matplotlib 库的基本使用

1. 代码 1-3 行先导入需要使用的 python 库
2. 5-6 行对字体进行设置，默认情况下中文不显示
3. 7 行在一个画布上面划分两行一列的区域，并将后续作图显示在第一个区域
4. 8-11 行产生 xy 数据
5. 12,13 行分别绘制曲线图与散点图,`ls` 是 `linestyle` 缩写,`lw`是 `linewidth` 缩写，`c` 是 `color`，`label` 是图例名称
6. 15-19 行分别限制坐标轴范围，设置坐标轴名称，已经添加网格线操作.
7. 后续代码，相关说明可以看源码注释部分

Python 的类

```
# 定义 person 类
class person():
    def __init__(self, name, age):
        self.name = name
        self.age = age
    def show(self):
        print("name is", self.name )
        print("age is", self.age )

# 生成对象
p1 = person("jason", "30")
p2 = person("justin", "28")
# 调用成员函数
p1.show();
p2.show();
```

运行结果

```
name is jason
age is 30
name is justin
age is 28
```

Python 模块

预备知识 Python 入门[1715]

在“Python 入门[1715]”中，我们示范了如何从 math 模块中引入一些变量和函数（如 pi 和 sin）. 现在我们来学习 Python 模块的原理.

事实上，Python 的模块可以是任意一个普通的 .py 拓展名的文件. 这个文件中可以定义全局变量，函数，类，甚至是可执行的命令. 当我们要在其他 .py 文件中使用这个文件中的函数或全局变量，只需要声明把前者作为模块使用即可.

例如创建模块文件 mymodule.py 如下

```
# 函数
def plus1(num):
    return num + 1
# 全局变量
num = 3.5
num2 = plus1(num)
```

在该模块中定义了一个函数和两个全局变量. 注意在定义第二个变量时执行了函数 plus1，这个命令会在模块被调用时执行.

使用模块

现在，我们我们在同一文件夹中新建如下 test.py，就可以用 import 语句来使用我们刚刚创建的模块.

```
import mymodule
print('num = ', mymodule.num)
print('num2 = ', mymodule.num2)
```

```
print('num2 + 1 = ', mymodule.plus1(mymodule.num2))
```

在“Python 入门[1715]”中我们还介绍了模块的不同导入方法，这里不再赘述.

除了内建模块外，Python 默认只会在当前文件夹（即路径）下搜索模块文件. 若要添加其他搜索路径，可以在 import 命令之前使用

```
import sys
sys.path.insert(1, 路径)
```

其中 路径 既可以是绝对路径也可以是绝对路径. 例如要指定当前路径的子文件夹 modules，则使用 'modules/' 即可；如果是上层文件夹中的子文件夹，可用 '../modules/'.

模块嵌套

在上面的例子中，如果在 mymodule.py 文件中使用了 import 语句导入其他模块中的函数和变量等，那么在其 test.py 中也可以使用这些函数和变量. 例如在 mymodule.py 第一行插入

```
from math import pi
```

那么在 test.py 中就可以使用 mymodule.pi 获取这个变量. 但如果把这行改为

```
import math
```

那么在 test.py 中要使用 mymodule.math.pi 才能获取这个变量.

Python 内建模块

Python 中有几个**内建模块（built-in module）**，可以随时导入.

math 模块

（未完成）

time 模块

在 Python 中，通常有这几种方式来表示时间：

时间戳（timestamp）：通常来说，时间戳表示的是从 1970 年 1 月 1 日 00:00:00 开始按秒计算的偏移量. 我们运行 `type(time.time())`，返回的是 `float` 类型.

格式化的时间字符串

元组（`struct_time`）：`struct_time` 元组共有 9 个元素：（年，月，日，时，分，秒，一年中第几周，一年中第几天，夏令时）

random 模块

随机生成内容.

```
import random
# 0,1之间时间生成的浮点数  float
print(random.random())
# 随机生成传入参数范围内的数字 即 1, 2, 3
print(random.randint(1, 3))
# 随机生成传入参数范围内的数字, range 顾头不顾尾
print(random.randrange(1, 3))
# 随机选择任意一个数字
print(random.choice([1, '23', [4, 5]]))
```

sys 模块

```
import sys
# 命令行参数List, 第一个元素是程序本身路径
print(sys.argv)
# 结果['D:/Pycharm Community/python内置函数/sysTest.py']
# 获取Python解释程序的版本信息
print(sys.version)
```

os 模块

os 模块是与操作系统交互的一个接口.

```
os.getcwd() # 获取当前工作目录, 即当前python脚本工作的目录路径
os.chdir("dirname")  # 改变当前脚本工作目录; 相当于shell下cd
```

```
os.curdir  # 返回当前目录: ('.')
os.makedirs('dirname1/dirname2')    # 可生成多层递归目录
# 若目录为空，则删除，并递归到上一级目录，如若也为空，则删除，依此类推
os.removedirs('dirname1')
os.mkdir('dirname')    # 生成单级目录
os.rmdir('dirname')    # 删除单级空目录，若目录不为空则无法删除，报错;
# 列出指定目录下的所有文件和子目录，包括隐藏文件，并以列表方式打印
os.listdir('dirname')
os.remove()  # 删除一个文件
os.rename("oldname","newname")  # 重命名文件/目录
os.stat('path/filename')  # 获取文件/目录信息
os.path.abspath(path)  # 返回path规范化的绝对路径
os.path.split(path)  # 将path分割成目录和文件名二元组返回
os.path.exists(path) #  如果path存在，返回True；如果path不存在，返回False
os.path.isabs(path)  # 如果path是绝对路径，返回True
os.path.isfile(path)  # 如果path是一个存在的文件，返回True. 否则返回False
os.path.isdir(path) # 如果path是一个存在的目录，则返回True. 否则返回False
os.path.join(path1[, path2[, ...]])  将多个路径组合后返回
```

Python 异常处理

- 本词条处于草稿阶段.

```
try:
    print(var)
except Exception as exc: # as exc 可以省略，  如果不需要使用
    print(exc) # 打印错误信息
else:
```

```
    print('done') # 运行正常时执行
finally:
    print('----') # 总会执行
```

Exception 代表所有异常，常见的异常类型有：SyntaxError 语法错误，NameError 名字错误（如变量名未定义），TypeError 类型错误，IOError 文件读写错误，等.

如果需要检测多个错误类型，可以把 Exception 替换成多个类型的元组如 (SyntaxError,NameError).

else 和 finally语句可以省略.

如何产生异常呢？每个异常类型可以把字符串转换为改异常类型的一个变量，然后用关键词 raise 即可抛出异常.

```
e = TypeError('这是一个错误')
raise e
```

通常把 raise 放在一个判断从句中，否则程序运行到此处必然会出现异常.

SciPy 数值微分与积分

预备知识 Python 入门[1715]

SciPy函数库在NumPy库[1721] 的基础上增加了众多的数学、科学以及工程计算中常用的库函数. 例如线性代数[521]、常微分方程数值求解[1886]、信号处理、图像处理、稀疏矩阵[1858] 等等. 使用SciPy库之前需要先导入库，

```
import  scipy#导入scipy中所有模块
from scipy import integrate#导入其中积分模块
```

数值积分

数值积分是对定积分的数值求解. 例如可以利用数值积分计算某个形状的面积. 下面让我们来考虑一下如何计算半径为 1 的半圆的面积，根据圆的面积公式，其面积应该等于 π. 单位半圆曲线可以用函数half_circle()表示，对应计算积分的代码如下：

```
def half_circle(x):
    return (1-x**2)**0.5

from scipy import integrate
res, err = integrate.quad(half_circle, -1, 1)
print(res,err)
输出:
1.5707963267948983 1.0002354500215915e-09
```

此处调用了积分模块下的quad函数，这里不仅返回了积分结果，还给出来计算误差. 多重定积分的求值可以通过多次调用quad函数实现，为了调用方便，integrate库提供了dblquad函数进行二重定积分，tplquad函数进行三重定积分.

dblquad函数的调用方式为:

```
dblquad(func2d, a, b, gfun, hfun)
```

对于func2d(x,y)函数进行二重积分,其中a,b为变量x的积分区间,而gfun(x)到hfun(x)为变量y的积分区间.

tblquad函数的调用方式与上面一样.

数值解微分方程组

scipy.integrate库提供了数值积分和常微分方程组求解算法odeint.odeint()函数需要至少三个变量，第一个是微分方程函数，第二个是微分方程初值，第三个是微分的自变量.

下面让我们来看看如何用odeint计算十分经典的洛仑兹吸引子的轨迹. 洛仑兹吸引子由下面的三个微分方程定义:

$$\dot{x} = \sigma \tag{1}$$

$$\dot{y} = x(\rho - y) \tag{2}$$

$$\dot{z} = xy - \beta \tag{3}$$

```
from scipy.integrate import odeint
import numpy as np
```

```
def lorenz(w, t, p, r, b):
    # 给出位置矢量w, 和三个参数p, r, b计算出
    # dx/dt, dy/dt, dz/dt的值
    x, y, z = w
    # 直接与lorenz的计算公式对应
    return np.array([p*(y-x), x*(r-z)-y, x*y-b*z])

t = np.arange(0, 30, 0.01) # 创建时间点
# 调用ode对lorenz进行求解, 用两个不同的初始值
track1 = odeint(lorenz, (0.0, 1.00, 0.0), t, args=(10.0, 28.0,
    3.0))
track2 = odeint(lorenz, (0.0, 1.01, 0.0), t, args=(10.0, 28.0,
    3.0))
from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt
fig = plt.figure()
ax = Axes3D(fig)
ax.plot(track1[:,0], track1[:,1], track1[:,2])
ax.plot(track2[:,0], track2[:,1], track2[:,2])
plt.show()
```

结果如图 1 所示. 这里我们还使用到了第四个参数args，它是一个tuple类型，

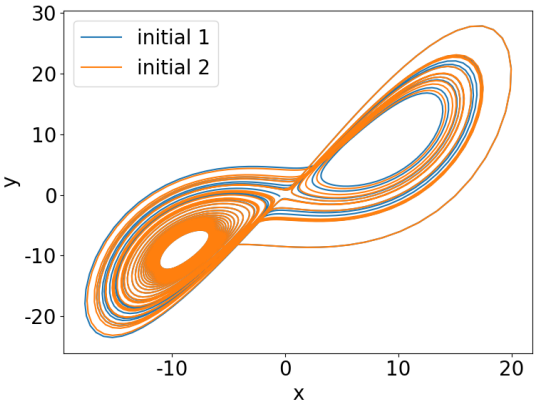


图 1: 洛伦兹微分方程数值解

给函数传递额外的参数. 我们看到即使初始值只相差 0.01，两条运动轨迹也

是完全不同的．在程序中先定义一个lorenz函数，它的任务是计算出某个位置的各个方向的微分值，这个计算直接根据洛仑兹吸引子的公式得出．然后调用odeint，对微分方程求解，odeint有许多参数，这里用到的四个参数分别为：

- lorenz，它是计算某个位移上的各个方向的速度（位移的微分）
- (0.0, 1.0, 0.0)，位移初始值．计算常微分方程所需的各个变量的初始值
- t，表示时间的数组，odeint对于此数组中的每个时间点进行求解，得出所有时间点的位置
- args，这些参数直接传递给lorenz函数，因此它们都是常量

SciPy 最小二乘法

预备知识 Python 入门[1715]

最小二乘拟合

最小二乘法是一种数学优化技术．它通过最小化误差的平方和寻找数据的最佳函数匹配．利用最小二乘法可以简便地求得未知的数据，并使得这些求得的数据与实际数据之间误差的平方和为最小．

通常情况下最小二乘法是这样的:给定一组数,假设是二维数据.$(x_1, y_1), (x_2, y_2), \cdots, (x$, 并且 x, y 是相关的．假设满足方程

$$y = f(x) \tag{1}$$

但是这个方程我们是不知道具体什么样的．我们找到一个近似函数

$$y = g(x) \tag{2}$$

y=g(x)

对于每一个点 (x_i, y_i) 有 $y_i = g(x_i) + \epsilon_i$，$\epsilon_i$ 是残差．那么总的残差平方和为

$$err = \sum_{i=1}^{n}(y_i - g(x_i)]^2 \tag{3}$$

那么我们就是要找到这样的 $g(x)$ 使得如下公式中的 err 函数最小．这种算法被称之为**最小二乘拟合（Least-square fitting）**.

scipy中的子函数库optimize已经提供了实现最小二乘拟合算法的函数leastsq. 下面是用leastsq进行数据拟合的一个例子：$y = A\sin(2\pi x + \theta)$.

```
import numpy as np
from scipy.optimize import leastsq
import pylab as plt
import matplotlib as mpl

mpl.rcParams["font.sans-serif"]=['kaiti']
mpl.rcParams['axes.unicode_minus']=False
def func(x, p):
"""
数据拟合所用的函数: A*sin(2*pi*k*x + theta)
"""
 A, k, theta = p
 return A*np.sin(2*np.pi*k*x+theta)

def residuals(p, y, x):
"""
实验数据x, y和拟合函数之间的差, p为拟合需要找到的系数
"""
  return y - func(x, p)

x = np.linspace(0, -2*np.pi, 100)
A, k, theta = 10, 0.34, np.pi/6 # 真实数据的函数参数
y0 = func(x, [A, k, theta]) # 真实数据
y1 = y0 + 2 * np.random.randn(len(x)) # 加入噪声之后的实验数据
p0 = [7, 0.2, 0] # 第一次猜测的函数拟合参数

# 调用leastsq进行数据拟合
```

```
# residuals为计算误差的函数
# p0为拟合参数的初始值
# args为需要拟合的实验数据
plsq = leastsq(residuals, p0, args=(y1, x))

print(u"真实参数:", [A, k, theta])
print(u"拟合参数", plsq[0]) # 实验数据拟合后的参数

plt.plot(x, y0, label=u"真实数据")
plt.plot(x, y1, label=u"\autoref{PyFit_fig1}带噪声的实验数据")
plt.plot(x, func(x, plsq[0]), label=u"拟合数据")
plt.xticks(fontsize=20)
plt.yticks(fontsize=20)
plt.legend()
plt.show()
```

结果图像如图 1 所示.

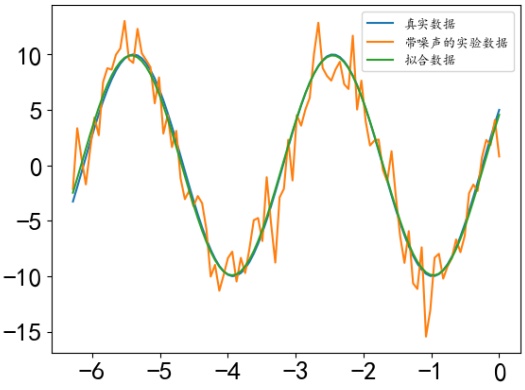


图 1: 最小二乘法 1

我们看到拟合参数虽然和真实参数完全不同，但是由于正弦函数具有周期性，实际上拟合参数得到的函数和真实参数对应的函数是一致的.

第二个例子：拟合 $y = kx + b$.

```
# -*- coding: utf-8 -*-
import numpy as np
from scipy.optimize import leastsq
```

```
import pylab as plt
import matplotlib as mpl

mpl.rcParams["font.sans-serif"]=['kaiti']
mpl.rcParams['axes.unicode_minus']=False
def func(x, p):
 k,b = p
 return k*x+b

def residuals(p, y, x):
  """
 实验数据x, y和拟合函数之间的差, p为拟合需要找到的系数
 """
  return y - func(x, p)

x = np.linspace(0, -2*np.pi, 100)
k,b = 10, 5.0 # 真实数据的函数参数
y0 = func(x, [k,b]) # 真实数据
y1 = y0 + 2 * np.random.randn(len(x)) # 加入噪声之后的实验数据
p0 = [7, 0.2] # 第一次猜测的函数拟合参数

 # 调用leastsq进行数据拟合
 # residuals为计算误差的函数
 # p0为拟合参数的初始值
 # args为需要拟合的实验数据
plsq = leastsq(residuals, p0, args=(y1, x))

print(u"真实参数:", [k,b])
print(u"拟合参数", plsq[0]) # 实验数据拟合后的参数

plt.plot(x, y0, label=u"真实数据")
plt.plot(x, y1, label=u"带噪声的实验数据")
plt.plot(x, func(x, plsq[0]), label=u"拟合数据")
plt.xticks(fontsize=20)
plt.yticks(fontsize=20)
plt.legend()
```

```
40 plt.show()
```

拟合结果：真实参数为 [10, 5.0]，拟合参数[9.96688563 5.0180872]．图像如图 2 所示．

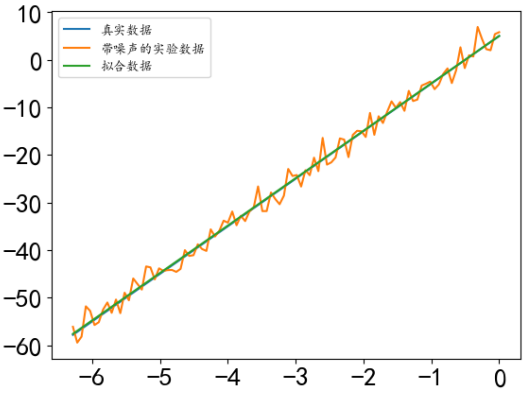


图 2: 最小二乘法 2

第三章

Linux 简介

Linux 基础

- 本词条存在未完成的内容.

虽然大多数人更熟悉 Windows 系统，但由于 Linux 系统的免费开源稳定，许多科学计算软件都是主要支持 Linux，而许多研究组的服务器和计算机集群也已 Linux 系统为主．所以本书使用 Linux 作为编译和运行环境．注意我们只介绍 Linux 的命令行（**terminal**）操作，不涉及任何图形界面．另外，Linux 有不同的版本，本书选用 Ubuntu．

同时使用 Windows 和 Linux

如果你的电脑已经有了 Windows 系统，要使用 Linux 命令行可以有几种方式：

1. 远程链接 Linux 服务器
2. Windows Subsystem for Linux （WSL）[1] 或者 WSL2[2]
3. Cygwin 或者 MinGW
4. Docker
5. 安装虚拟机，如 VirtualBox
6. 安装双系统

其中前 4 种方案可以在使用 Windows 的同时使用 Linux 命令行（注意使用虚拟机需要占用较多资源，可能导致电脑卡顿），而双系统方案一次启动只能进入一个系统．

[1]注意只有 Windows 10 支持 WSL

[2]WSL2 预计 2020 年发布．

Windows 远程软件

在 Windows 下使用 Linux 的另一种方案就是远程连接到另一台装有 Linux 系统的电脑

- putty, kitty，MobaXterm 远程命令行
- WinSCP 文件传输

基本命令

暂时先列出最基础的命令行命令，请自行搜索学习

未完成：有待完善

- `pwd` 当前目录
- `ls` 查看某目录的文件（`-v` 序号排序）
- `cd` 改变当前目录
- `mkdir` 创建目录（`-p` 创建多层）
- `rmdir` 删除目录
- `cp` 复制文件（夹）
- `mv` 移动文件（夹）
- `touch` 创建空文件或更改文件日期
- `rm` 删除文件或文件（夹）
- `man`，`--help`
- `echo` 重复文字
- `>`，`<` 和“|”
- `cat` 显示文本文件内容

- `vim` 编辑文本文件
- `sudo` 超级管理员权限
- `du`，`df` 文件（夹）大小，硬盘容量
- `top` 进程管理，资源占用
- `tar`，`zip` 压缩与解压
- `ln -s` 符号链接（类似 windows 的快捷方式）
- `uname` 系统信息
- `apt-get` 安装软件
- `chmod` 修改文件权限
- `hostname`，`hostname -I` 本机名称，IP 地址
- `ping` 检查是否可以连接到网络地址
- `reboot` 重启
- `ssh` 远程命令行
- `find` 搜索文件或目录 `-exec` 对搜索结果执行命令
- `grep` 搜索字符串或文件内容
- `locate` 搜索计算机的文件

进阶命令

未完成：有待完善

- `sha1sum`，`md5sum` 等指纹
- `rsync` （若文件中含有转义字符需要三个反斜杠 `\\\`）

在 Linux 上编译 C/C++ 程序

本文介绍如何使用 Linux 控制行用 g++ 编译器编译一个 C++ 的 hello world 程序．阅读本文不需要任何基础（包括 C++ 和 Linux）．

如何安装 Linux 控制行？你可以直接在电脑上安装一个 Linux 系统，也可以在 Win10 下使用 WSL（已经安装 win10 系统的话建议使用后者），如果是 win7 系统，可以安装一个 MinGW．安装方法请自行搜索．本文以 Ubuntu 系统为例（Linux 有不同的版本，Ubuntu 是广泛使用的一个）．

在控制行安装 g++

打开控制行后，要检查是否安装了 g++，输入以下命令并按回车．

```
g++ --version
```

这个命令由两部分组成，第一部分是 `g++`，即在控制行中调用 `g++` 这个程序．第二部分是 g++ 程序的 `--version` 选项，即查看已安装 g++ 的版本．如果系统上没有安装 g++，控制行（Terminal）会返回类似这样的提示

```
Command 'g++' not found, but can be installed with:
sudo apt install g++
```

要安装 g++，只需要在 terminal 输入两个命令

```
sudo apt update
sudo apt install g++
```

`sudo` 是 super user do，即获取操作系统的 root 权限，root 权限是 Linux 系统的最高权限．控制行中的一些命令（如安装卸载软件）一般需要 root 权限，所以需要在前面加 sudo （如果你已经以 root 的身份登录 Linux 则不需要）．

`apt` （也可以用 `apt-get`）是 Ubuntu 系统中用来管理程序的命令（安装，卸载，更新等）．注意系统一般要链接网络才能使用 apt 安装程序．在安装任何程序之前，建议新更新一下 apt 的数据库（第一行）．

`update` 是 `apt` 的一个参数（argument），就跟 g++ 的 `--version` 参数一样．Linux 中程序的许多参数以 `--` 或者 `-` 开头．

`install` 是 `apt` 程序的另一个参数. `g++` 是 `install` 的参数，即要安装的软件名.

运行后可以按提示安装 g++，过程中可能需要输入 Linux 的用户密码，以及确认是否安装（输入 `Y` 然后按回车）

```
After this operation, 155 MB of additional disk space will be
   used.
Do you want to continue? [Y/n] Y
```

安装完以后，再次查看 g++ 的 version，得到

```
g++ (Ubuntu 7.4.0-1ubuntu1~18.04.1) 7.4.0
Copyright (C) 2017 Free Software Foundation, Inc.
This is free software; see the source for copying conditions.
   There is NO
warranty; not even for MERCHANTABILITY or FITNESS FOR A
   PARTICULAR PURPOSE.
```

可以看到我们安装的版本是 7.4.0.

创建第一个 C++ 程序

打开一个文本编辑器，创建一个新文本文档，命名为 hello.cpp，输入如下 C++ 代码并保存.

```cpp
// print hello world
#include <iostream>

int main()
{
    std::cout << "hello world!" << std::endl;
}
```

在编译之前，我们首先要介绍 working directory. 控制行一般会在输入的每一行命令之前显示当前的路径（directory）如图 1.

这个目录一般是打开控制行的目录（例如 Ubuntu 中在某个文件夹中右键打开控制行，或者 win10 中按 shift + 右键）. 我们假设刚才创建的 cpp 文件就保存在这个目录下.

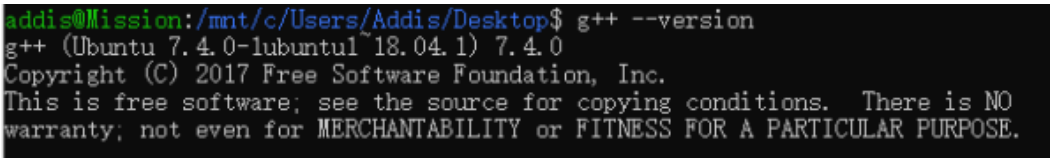


图 1: 蓝色的路径就是当前路径

现在我们来用 g++ 编译这个程序，控制行输入

```
g++ hello.cpp
```

如果编译成功，将会在同一文件夹中生成 a.out 可执行文件，相当于 windows 中的 exe 文件．要运行该文件，使用（`./` 表示当前的 working directory）

```
./a.out
```

运行后可以看到程序在控制行输出了 `hello world!`.

这样，我们就成功在 Linux 中运行了第一个 C++ 程序.

Makefile 简介

预备知识 Linux 基础[1761]

Make 通常被用于自动编译程序．当程序由许多文件构成或使用许多其他的程序库时，编译的命令往往会变得非常复杂，如果每次编译都手动输入这些命令会非常麻烦且容易出错．如果每次输入的编译命令都相同，一个简单的解决方法是直接把这些命令写进一个 shell script，每次编译调用即可.

但这样做会遇到另一个问题，当程序文件（如.cpp 文件）较多时，如果在上次编译以后只修改了某些文件，我们会希望只将这些文件重新编译（生成对应的.o 文件）并再次 link，这样可以大大节约编译时间.

Linux 系统中，`make` 就是为了解决这个问题产生的．`make` 并不是为某个编程语言而设计的，也不会去分析任何文件的内容．用户需要写一个名为 `Makefile` 的文件告诉 `make` 该怎么做．`Makefile` 中主要描述了两件事情，第一是每个文件需要依赖哪些文件才能开始生成，第二就是每个文件需要用哪些命令生成．下面举例说明，为了方便，这个例子只使用 Linux 常用命令而不涉及任何编程语言.

简单的例子

我们先在某个目录创建一个简单的 Makefile 文件，内容如下

代码 1: Makefile

```
1 a :
2         echo "file a" > a
3
4 b : a
5         echo "text in a {" > b
6         cat a >> b
7         echo "}" >> b
```

这个 Makefile 指定了两个要生成的文本文件 a 和 b．先来看含有冒号的两行，冒号表示依赖关系，左边是需要生成的文件名，右边是依赖的文件名．可见 a 没有任何依赖，可以直接生成，而文件 b 依赖于 a，需要等 a 生成后才能开始生成．

在每个含有冒号的行后，可以添加若干行命令，当 make 决定生成某个文件时，就执行这些命令．注意每个命令前必须有一个制表符（用键盘的 tab 键插入）．所以生成 a 文件的命令就是 `echo "file a" > a`，即新建文件 a，内容为 `file a`．同理，生成 b 的命令就是新建文件，内容为 `text in a {...}` 其中 `...` 为 a 中的文本．如果 a 文件不存在，第 6 行命令就无法执行，所以必须指定 a 为 b 的依赖．

每个含有冒号的行以及接下来的所有命令构成一个 **规则（rule）**．冒号左边的文件叫做 **目标（target）**，右边的文件叫做 **依赖（dependency）**．如果有多个依赖文件，用空格隔开即可．

现在可以在命令行先 cd 到 Makefile 所在的目录，然后执行 make 命令．如果一切正常，程序将会在命令行依次输出执行的命令．我们看到 Makefile 中第 2 行被执行以生成 a

```
1 echo "file a" > a
```

生成文件 a 的内容为 `file a`．

但为什么不生成 b 呢？因为 make 默认情况下只生成第一个出现的目标．如果想指定生成 b，执行 `make b` 即可．执行过程中命令行显示 Makefile 的

5-7 行被执行了．但注意第 2 行没有被执行．这是因为 a 文件已经存在了且上一次 make 后没有被修改过，所以就不需要再重复生成一次了．b 中的内容如下

```
text in a {
file a
}
```

现在两个文件都生成了，且没有被修改，所以若再次执行 make b，则会显示

```
make: 'b' is up to date.
```

执行 make 或 make a 结果类似．

接下来我们把 a, b 两个文件都删除，执行一次 make b，会发现第 2 行和 5-7 行都被执行了，因为当 make 要生成 b 时发现它的依赖文件 a 并不存在，所以就会先生成 a．两文件中内容与之前相同．

我们来手动把 a 中的内容修改为 modified file a，再次运行 make b，发现 Makefile 的 5-7 行被执行，且 b 中的内容变为

```
text in a {
modified file a
}
```

这是因为，make 在生成 b 前，不但要检查它的所有依赖都存在还要检查他们在上次 make 后是否有更新．如果任何依赖有更新，b 将会被重新生成．

伪目标

注意 rule 的目标除了是具体的文件，也可以是**伪目标（phony tarket）**，伪目标可以有任意名字．伪目标的 rule 中可以有命令也可以没有．

一个特殊的伪目标是 goal．在上面的 Makefile 中如果我们想让 make 默认生成 b，可以在文件开始插入 goal 作为第一个目标

```
1 goal : b
2
3 a :
4         echo "file a" > a
```

```
5
6  b : a
7          echo "text in a {" > b
8          cat a >> b
9          echo "}" >> b
```

现在执行 make 或者 make goal 就相当于 make b 了.

另一个常用的伪目标是 clean，通常用于清除所有被生成的文件. 修改 Makefile 如下

```
1  goal : b
2
3  a :
4          echo "file a" > a
5
6  b : a
7          echo "text in a {" > b
8          cat a >> b
9          echo "}" >> b
10
11 clean :
12         rm -f a b
```

现在执行 make clean 就可以清除两个生成的文件. 使用 -f 选项的好处是，即使某些被删除的文件不存在，rm 也不会警告.

Makefile 笔记

预备知识 Makefile 简介[1766]

- 要系统地学习 make，参考资料是 GNU Make 文档，提供网页版和 pdf 下载.
- make 程序广义上是声明如何更新一系列文件，以及他们的 dependencies. 当某个文件改变后，任何依赖于它的文件都需要通过指定的规则（rule）来更新

- 突然发现，rule 的“:”前面不可以有空格？
- rule 的第一行用于声明什么文件取决于什么文件，剩下的行声明用什么命令去更新.
- “:”后面的文件会按照顺序更新，而且只会更新一次，若以如果更新第二个文件的时候第一个文件被删掉也不会出错
- “:”后面可以有 phony target，比如说 `make clean` 中的 `clean`
- `rm` 命令后面记得加 `-f` 选项，否则如果文件不存在就会出错导致 make 就会失败
- 可以根据所有 rule 中的 target 和 dependency 画一个树状图，如果某个点的文件不存在，那么就会先运行生成它的 rule，如果图中的任何一点更新了，从这点到顶点的所有文件都要更新. 具体的规则是，如果一个 target 的任何一个 dependency 比它要新，那么 target 就要重新编译.
- “goal :”可以声明 make 的终极目标，如

```
goal: file1 file2 ...
        command1
        command2
```

- 如果没有 goal 的话，default goal 就是第一个开头不为“.”的 target. 剩下的 rule 的顺序应该可以随意任何一个被依赖的文件改变了，或者它们依赖的文件改变了，就会执行 command. make 并不知道 command 的含义. 只是把它传给 shell 来执行.
- implicit rules 大概就是可以仅声明“A: B”，由自定义的命令或者 make 默认的命令来生成“A”. implicit rule 中的 dependency 是至少有这些 dependency，而不是只能有这些 dependency.
- “MAKEFLAGS = -r”大概是用来取消默认 implicit rules，包括所有后缀名识别
- 老的 suffix rule 和 implicit rules 的功能差不多，现在已经过时了，应该用 implicit rule. 一个例子如

```
.f90.o:
        gfortran -c $<
```

其中 `$<` 是 auto variable 中的一个（见 10.5.3 Automatic Variables），在执行的时候被替换成 `:` 右边的第一个 dependency. 现在如果有

```
file1.o: file1.f90 file2.o file3.o
```

那么应该会执行 `gfortran -c file1.f90`.另外,如果“file2.o”或“file3.o”被更新了，这条命令应该也会再执行一次.

- `$^` 列出所有的 prerequisites (“:”右边的内容)
- `$(shell ...)` 可以执行 shell 命令，如 `$(shell echo *.f90)` 可以在当前位置列出所有“.f90”文件.
- `$@` 大概就是 target file （如果“:”左边只有一个文件的话）

第四章

C++ 简介

C++ 基础

Matlab 和 Python 等动态语言虽然用起来方便，但缺点是运行较慢，对于一些计算量大的项目不适合．目前在高性能计算中广泛使用的只有两种语言即 C++ 和 Fortran．虽然 Fortran 普遍被认为是一个过时的语言，但在计算物理中，许多人仍然在使用，一是因为以前遗留下的 Fortran 代码比较多，二是一些年纪较大的学者只会 Fortran．

一本在数值算法中很有名的书是 Numerical Recipes，这本书第三版以前都使用 Fortran 或 C，而第三版却只有 C++，这也是本书选择介绍 C++ 而不是 Fortran 的原因之一．本书将从 Numerical Recipes 中借鉴许多代码上的风格和算法．

C++ 的特征实在多不胜数，事实上无论是什么语言，做计算物理的研究者大多会倾向于只选择一些最简单的语法来使用．我们在这里列出本书使用的 C++ 特性．

基础语法

- 基本类型（`bool`, `char`, `int`, `long`, `long long`, `float`, `double`, `long double`）
- 基本算符（`=`, `+`, `-`, `*`, `/`, `%`, `++`, `--`, `+=`, `-=`, `*=`, `/=`, `?:`) 以及优先级
- scope （scope 内的名字在 scope 外没有定义，scope 内可以定义与 scope 外相同的名字并覆盖，类的 destructor 会在 scope 结束时自动调用）
- 判断（`if`, `else if`, `else`）
- 循环（`for`, `while`，`do while`）
- 函数（函数名重载，变量默认值，算符函数，inline 函数）
- `const`
- 指针

- 引用
- 头文件机制（相当于原地插入头文件中的代码）
- 多文件编译
- one definition rule （ODR)
- 栈（stack）和堆（heap）
- `typedef`
- 数组
- 动态内存管理 `new`, `delete`.
- 异常处理 `throw`, `try`, `catch`

标准库

- `cmath`
- `complex`
- `vector`
- `string`, `string32`
- `iostream` （`cin`，`cout`，`<<` 算符，`>>` 算符）
- `fstream`

较高级的功能

- 类（constructor, destructor, `public`, `private`，数据成员，函数成员，`operator+, -, *,`
- 宏（`#include`, `#define`, `#if`, `#ifdef`, `#ifndef`, `#else`, `#endif`）
- `namespace`
- `constexpr`
- 类的继承

高级功能

- 类模板
- 函数模板

C++ 的整数

本文的**模（modulo）**运算都是指 M 加或减整数个 N 后使结果范围在 0 到 $N-1$ 之间，N 只能是正整数．以下我们假设整数用 n 位二进制表示．

无符号整数

- 取值范围为 0 到 2^n-1
- 如果溢出就把二进制的高位截去，也就是模 2^n

有符号整数

- 取值范围为 -2^{n-1} 到 $2^{n-1}-1$
- 运算 overflow 的结果无定义

转换规则

整数之间的转换规则：

- 超出范围的值转换为无符号整型，将其模 2^n （n 是目标类型的比特数），例如 $n=8$ 时 -1 转换为 255
- 超出范围的值转换为有符号整型，结果无定义

其他基本类型转换规则：

- 其他类型转换为 `bool`：0 变为 `false`，否则变为 `true`
- `bool` 转换为其他类型：`true` 变为 1，`false` 变为 0

- 浮点类型转换为整数类型：向 0 取整
- 整型转为浮点：如果位数太多会不精确

2 的补

有符号整数类型的负数在内存中的二进制表示常采用 **2 的补（2's complement）**. 若采用 2 的补，以下几点成立

- 一个整数和它的相反数相加等于 2^n
- 范围内最大的整数加上 1 等于范围内最小的整数，即加 1 后取相反数（如 $n = 8$ 时 $127 + 1 = -128$）
- 若保持 bit 不变，无符号类型变为有符号类型只需减去 2^n，反之则加上 2^n. 例如 $n = 8$ 时无符号 156 和有符号 -100 的各个 bit 相同.

也有其他表示负数的方法，例如最左边的 bit 表示符号，剩余表示大小.

数学上,可以将无符号的整数类型及其加法构成一个 2^n 元循环群（见例 2 [534]）. 若采用 2 的补，由于群元素可以用任何符号，有符号整数类型可以看作将无符号整数类型的后一半元素减去 2^n，而群运算保持不变.

Integer Literal

[1]这里讨论 c++11 标准. literal 中的字母不区分大小写，`0` 开头表示 8 进制，`0x` 开头表示 16 进制，数字后面可以加 `u`，以及 `l` 或 `ll` 的一个. literal 的类型会根据数值的大小而变化.

- 最小是 `int`，最大是 `long long int` 每一种前面都可以加 unsigned
- 有后缀 `u` 必定是 unsigned
- 没有 `u` 的十进制必定是 signed，其他进制可以是 signed 或 unsigned （首选 signed）
- 后缀 `l` 和 `ll` 分别规定类型至少为 `long` 和 `long long`，如果数值超出就不断升级

[1]参考cppreference

C/C++ 多文件编译笔记

以下规则对 c 和 c++ 语言都适用.

单个 translation unit

- `#include` 头文件（.h）相当于直接把头文件的内容插入到当前位置.
- 原则上头文件可以没有或者有任意后缀名（例如标准库的头文件就没有后缀名）.
- 每个.cpp （也可以用.cxx）文件叫做一个 translation unit . 可以单独编译成一个 object 文件（.o）
- 某个 translation unit 中定义的宏只在这一个 translation unit 中有效，即从定义开始到文件底部（或者到 `#undef`）.
- 头文件开始用 `#pragma once` 可以防止同一个头文件被多次插入同一个 translation unit. 例如 `b.h` 和 `c.h` 中都 include 了 `a.h`，如果 cpp 文件中同时 include 了 `b.h` 和 `c.h`，`a.h` 就不会被插入两次.

多个 translation unit

- 若有多个 translation unit，可以分别编译，然后将所有 object 文件 link 起来就成了可执行程序.
- 一个 translation unit 中定义的函数要在另一个 translation unit 调用前要先声明（declare）.
- 每个 translation unit 编译的时候也不需要有其他 translation unit 存在，即使调用了其他 translation unit 中定义的函数.
- 只有在 link 阶段编译器才会检查一个 o 文件中调用的函数是否能在另一个 o 文件中有定义，如果没有，link 就会失败.

- ODR （one definition rule）要求一个函数不能在同一个或不同 translation unit 中有重复的定义（definition）. 但重复的声明（declaration）是允许的.
- inline 函数或变量需要在使用到它的每个 translation unit 中都有一样的定义（所以一般放到头文件中）. 否则即使预先 declare 也会出错.
- class 的声明也需要在使用到它的每个 translation unit 中都有一样的定义（所以一般放到头文件中）.
- 任何函数或类的 template 的定义同样也需要在使用到它的每个 translation unit 中都有一样的定义（所以一般放到头文件中）.
- `inline` 的含义已经与是否真的 `inline` 几乎无关了，只是为了避免 ODR 错误的一种手段.
- 即使使用了 `inline`，同一个 translation unit 中也不可以出现两个同样的定义.
- 如果一个函数调用的函数在它之后定义，则需要将 declaration 放在它前面.

C++ 的 namespace

- 本词条处于草稿阶段.

- 使用 `namespace` 函数的一个重要例外就是，argument 的 `namespace` 会从 `function` 取消. 这样 namespace 中的函数就有可能和 namespace 外的函数发生冲突！解决的办法就是定义自己的 namespace，或者直接更改函数名.
- STL 中 include 的 c header 并没有任何 namespace，例如 `#include <iostream>` 以后，`getchar()\verb` 没有 namespace. 这是历史遗留问题.
- 定义 namespace 的大括号内部如果 using 其他 namespace 中的 name，那么这个 name 也会在这个 namespace 中出现. 所以尽量不要！

- 定义 namespace 的大括号中如果定义了一个 name，那么大括号外定义的相同的名字就会被隐藏，即使是不同的函数 overloading 也会被隐藏．可以在函数名前面加 global namespace `::`.
- 如果有一个没有 namespace 的函数和一个有 namespace 的同 signature 函数，使用 using declaration 不会发生冲突（没有 namespace 的函数被隐藏），但 using directive 就会冲突！
- 在 global namespace 中不要使用开头有下划线的名字．
- 如果在 `namespace A {}` 内部使用 `using namespace B;\verb` 在 `namespace A {}` 外部使用 `using A::xx;` 是安全的，即不会引入 `using namespace B`. 但如果使用 `using namespace A`，则会引入 `using namespace B`.
- 即使在 `namespace A {}` 内部使用 `using namespace B;` 新定义的函数也不会与 B 中的同名函数 overload，而是会直接覆盖 B 中的函数．

C++ 异常处理

- 本词条缺少预备知识，初学者可能会遇到困难．

在初学编程时，遇到错误我们往往就直接用 `exit()` 终止程序．但有时候我们不希望程序终止，而是希望程序自行对错误进行一定的处理．C 语言的常见办法是把函数的返回值（通常是整型）作为错误代码，`0` 代表成功，其他值对应不同类型的错误．然而函数调用是重重嵌套的，将错误代码层层传递是一件很麻烦的事情，另外，把错误处理和函数调用的语法分离开也可以使代码的结构更明显．

在 C++ 中有专门的异常处理机制，一般使用 `throw`（抛出某种类型的错误）, `try`（检测某段代码的运行）和 `catch`（处理某种类型的错误）三个关键词完成．来看一个简单的例子．

```
1 #include <iostream>
2 #include <string>
3 using namespace std;
```

```
4
5  struct err_info { string where, what;};
6
7  void fun2()
8  {
9          err_info e;
10         e.where = "fun2()"; e.what = "something wrong!";
11         throw e;
12 }
13
14 void fun1() { fun2(); }
15
16 int main () {
17         try { fun1(); }
18         catch (err_info e) {
19                 cout << "where: " << e.where << endl;
20                 cout << "what: " << e.what << endl;
21         }
22 }
```

程序中 main() 调用 fun1()，fun1() 接着调用 fun2()，而 fun2() 必然会出现一个异常，抛出了一个类型为 err_info 的对象 e. 这时无论 fun2() 是否运行完成都会终止，并把 e 传给 fun1()，而 fun1() 并没有处理这个异常的代码，所以 fun1() 同样也终止运行，并把 e 回传给 main(). 由于 main() 中存在处理 err_info 类型错误的代码（catch (err_info e)），所以就会执行相应的错误处理.

调试 C++ 程序

程序调试

对于 Matlab 或 Python 这样的动态语言，由于它们的解释器（interpreter）执行代码的方式就是从代码中读一行执行一行，所以对它们的来说，**调试（debug）**是一件很自然的事情，当某个地方出错了，解释器会告诉你出错的原因，代码

的位置，以及函数调用的顺序（也就是说如果错误出现在某个函数中，解释器就会告诉你例如程序的第几行调用了 A 函数，A 函数中第几行调用了 B 函数，B 函数的第几行哪里出错了）．然而对于编译语言，由于在编译过程中代码被转换成二进制命令，程序在执行二进制命令时如果没有额外的调试信息，则没有办法对应到代码的相应位置．

而使调试更难的是，表面上代码出错的位置并不是最开始出错的位置，或者程序根本不会在运行时检测到错误，但输出了错误的结果．这就需要从发现错误的位置逆向顺藤摸瓜找到错误的根源．而这么做就需要确定并检查代码中一连串的位置．例如一个函数中的某个错误可能是被调用的时候被传递了错误的参数，这时就要去检查调用它的函数，以此类推．

无论使用什么编程语言，调试程序通常有两种方法：一种叫做 **print debug**，就是手动在代码中插入许多额外的输出命令，运行的过程中在屏幕上（或文件中）不断输出信息，如果哪里出错了，我们就可以根据最后输出的信息对应到出错代码的位置．第二种方法是使用某种**调试器（debugger）**，例如下面要介绍的 **GNU Debugger**，简称 gdb．

下面我们来简单讨论 C++ 这样的静态语言中两种调试方法的优势和劣势．print debug 的劣势显而易见，如果你的代码本来没有任何输出，那么程序运行过程中崩溃或者结果错误的时候你将完全不知道发生了什么，只能先在代码的一些主要部分插入一些输出命令，重新编译，复现错误，把出错的位置缩小一些，然后再在这个小范围进一步插入更详细的输出命令，重新编译，复现错误，直到找到最开始的原因．更糟糕的是，如果一个函数被层层调用，即使你知道这个函数中出错了，也可能不知道是谁调用它的时候出错的．如果这个函数在循环中被调用，那输出的调试信息可能会多得难以阅读．除了输出命令，也可以插入一些额外的检查命令，例如检查数组的索引是否超出了数组的长度，某些值是否不在某个范围内，如果不符合就输出错误信息并终止程序．最后，当 bug 被修复以后，我们往往还需要手动删除用于 debug 的代码．

如果使用调试器，你不需要修改程序就能调试程序．调试器通常是和编译器配套的，例如 gcc （C 语言）或者 g++ （C++）的编译器对应的调试器都是 gdb．编译器在编译程序的时候如果使用调试选项（`-g`），那么它们在编译时就会在生成的可执行文件中插入一些调试信息，例如每个命令对应到代码的哪一行，每个变量的名称是什么，等．编译完以后，我们并不直接运行程序，而是将其放到调试器中运行，这样如果出错了调试器就可以根据编译时插入的额

外信息显示出错的行号，以及函数调用顺序．调试器甚至还可以像动态语言一样，逐行运行程序，或者给某行添加**断点（breakpoint）**，当程序运行到断点处或者触发特定的条件时就会暂停（就像 Matlab 的程序调试[1710] 中介绍的那样）．当程序在某行被暂停时，我们还能直接检查每个变量当前的值（同样无须修改程序）．

但调试器也有一些弊端．在编译器正常编译程序时往往会进行一些优化使程序运行得更快，而这些优化使得二进制文件中的命令并不完全和程序代码一一对应．例如编译器可能会把某个中间变量“优化掉”，这时如果进行调试，就无法查看这个变量的值．所以当我们调试程序时，往往不希望编译器进行任何优化（例如 `-O1` 到 `-O3` 等选项）．不优化会导致程序运行变慢，但如果程序本身运行的时间不长，你可能感觉不到太大区别．另一个劣势是，学习和掌握调试器的各种命令（或者 GUI 界面）需要一定的时间．这两个缺点都可以使用 print debug 避免．

在实践中我们往往会根据情况选择其中一种调试方法或者两种混用．通常要写好一个程序，我们花在调试和修改上的时间往往比写程序本身要多许多，要更高效地调试，通常只能从实践中慢慢积累经验，用更少的步骤更多的推理得出问题根源所在．

gdb 使用示例

（未完成）

数据结构：密矩阵

一般来说，矩阵的每个元在计算机内存中逐个储存，这种数据结构通常叫做**密矩阵（dense matrix）**．由于在计算机内存中所有数据都是按顺序排成一行，所以在储存矩阵时我们就有两种选择，一是把矩阵所有行首尾相接，叫做**行主序（row-major）**，二是把矩阵所有列首尾相接，叫**列主序（column-major）**．例如对于 2×2 的矩阵 $\mathbf{A}$，列主序下，矩阵元在内存中的顺序依次 $A_{11}, A_{21}, A_{12}, A_{22}$，而在行主序下顺序为 $A_{11}, A_{12}, A_{21}, A_{22}$．

在 fortran 和 Matlab 中[2]，语言自带的矩阵都是列主序，而在 C/C++ 中如果

[2]Matlab 最初就是由 fortran 编写的．

用指针的指针（或数组的数组）来表示矩阵，得到的将会是行主序．当然，由于 C++ 的灵活性，我们完全可以创造行主序和列主序两种不同的矩阵类．由于 C++ 的数组的指标从 0 开始，那么矩阵的行标和列表也习惯从 0 开始．

在我们用双索引寻找矩阵元时，我们需要先将其转换为单索引．假设矩阵尺寸为 $N_1 \times N_2$，那么

$$\begin{cases} n = i + N_1 j & \text{（列主序）} \\ n = N_2 i + j & \text{（行主序）} \end{cases} \tag{1}$$
$$(i = 0, 1, \ldots, N_1 - 1, \quad j = 0, 1, \ldots, N_2 - 1)$$

行主序和列主序也可以延申至高维矩阵，如果使用列主序，那么当我们在内存中按顺序读取数据的时候，第 1 个索引（index）将变化得最快，第 2 个索引变化得第二快，最后得索引变化得最慢．行主序则相反，最后的索引变化得最快，而第一个最慢．例如 4 维数组的多索引变为单索引的公式为

$$\begin{cases} n = i_1 + N_1 i_2 + N_1 N_2 i_3 + N_1 N_2 N_3 i_4 & \text{（列主序）} \\ n = N_2 N_3 N_4 i_1 + N_3 N_4 i_2 + N_4 i_3 + i_4 & \text{（行主序）} \end{cases} \tag{2}$$

从性能角度来看，单索引要比多个索引要快．

若要由单索引计算多索引，我们可以用整数除法（向下取整）`/` 和求余运算 `%`，例如对列主序的矩阵有

$$\begin{cases} i = n \% N \\ j = n / N \end{cases} \tag{3}$$

数据结构：带对角矩阵

预备知识 密矩阵[1782]

在 CBLAS 中，我们可以用**带对角矩阵（band diagonal matrix）**的数据结构来储存矩阵，减少不必要的零元素的储存和运算．和密矩阵一样，带对角矩阵同样分为行主序和列主序两种．

要把一个密矩阵转换成列主序带对角矩阵，就把每个对角线都排成一行，并保持每个元素的列标不变即可．同理，要转换成行主序带对角矩阵，就把每个对角线都排成一列，保持每个元素的行标不变．

举一个例子，若列主序密矩阵（表 1）中只有几个对角线不为零，当这样的矩阵很大时，就会在左下角和右上角出现大量的矩阵元为零，不仅占用内存，矩阵运算时也效率不高．为了方便我们不妨把不为零的矩阵元依次编号[3]．

列主序带对角矩阵如表 2，列主序带对角矩阵如表 3．

表 1: 密矩阵

01	04				
02	05	08			
03	06	09	12		
	07	10	13	15	
		11	14	16	17

表 2: 列主序带对角矩阵

	04	08	12	15	17
01	05	09	13	16	
02	06	10	14		
03	07	11			

表 3: 行主序带对角矩阵

		01	04
	02	05	08
03	06	09	12
07	10	13	15
11	14	16	17

[3]这并不是标准的编号方式，只是这里为了讲解临时使用的．

索引转换

令表 1 到表 3 中矩阵的行数和列数分别为 N_1, N_2，行标分别为 i, i_c, i_r，列标分别为 j, j_c, j_r．单索引分别为 k, k_c, k_r．再令 i_{diag} 为表 2 中对角线所在的行标，j_{diag} 为表 3 中对角线所在的列标，那么有以下转换公式

$$\begin{cases} i_c = i - j + i_{diag} \\ j_c = j \\ k_c = i_c + N_1 j_c = i + i_{diag} + (N_1 - 1)j \end{cases} \tag{1}$$

$$\begin{cases} i_r = i \\ j_r = j - i + j_{diag} \\ k_r = N_2 i + j = (N_2 - 1)i + j + j_{diag} \end{cases} \tag{2}$$

C++ 矩阵类的实现

预备知识 C++ 中类的定义和继承

未完成：本文代码还没验证过

容器指的就是像 `std::vector` 这样的可以储存多个元素的对象，下面介绍的矢量，矩阵，高维数组等都是容器．

C/C++ 的数组是储存于内存的 stack （栈）中的，而 stack 一般只有几个 Mb，一旦数组太大就会产生 stack overflow 错误．第二是 stack 中的数据的大小都只能在编译时确定（即 constant expression，例如 literal 或者宏定义），所以我们不能在运行是确定数组的长度（例如从文件中读取，或通过运行时的计算得到）．

```
// 用 C/C++ array 的嵌套定义矩阵
#define N1 5
#define N2 5
double a[N1][N2]; // 定义并在栈中分配内存
// 全部元素赋值为 0
```

```
for (long i = 0; i < N1; ++i)
    for (long j = 0; j < N2; ++j)
        a[i][j] = 0;
```

所以在一般来说容器都会 allocate （分配）到内存的 heap （堆）中．在 C++ 中，在 heap 中分配内存用 `new`，释放则用 `delete`．

`std::vector` 可以说是使用最广的矢量容器，理论上我们可以用 `vector` 的 `vector` 定义一个 heap 中的矩阵．

```
// 用 std::vector 的嵌套定义矩阵
vector<vector<double>> a;
// 分配内存
long N1 = 10, N2 = 10;
a.resize(N1);
for (long i = 0; i < N1; ++i)
    a[i].resize(N2);
// 全部元素赋值为 0
for (long i = 0; i < N1; ++i)
    for (long j = 0; j < N2; ++j)
        a[i][j] = 0;
```

注意由于用了多次 `new`，数据在内存中不一定是连续的，甚至不一定按 `i` 的顺序排列．这样不是很好，为了性能和语法上的考虑，一般需要给每个容器分配一块连续的内存（例如我们会想要用单个连续索引来获得矩阵元）．

最后一种方法就是本书 SLISC 库[1790] 中的底层实现方法，即在堆中分配一段连续的内存，并把双索引转换乘单索引来获取矩阵元．

```
Long N1 = 10, N2 = 10;
double *a;
new double a[N1*N2];
// 全部元素赋值为 0
for (long i = 0; i < N1; ++i)
    for (long j = 0; j < N2; ++j)
        a[N2*i + j] = 0; // 行主序
        // a[i + N1*j] = 0; // 列主序
// 或者直接使用单索引赋值更快
for (long i = 0; i < N1*N2; ++i)
```

```
    a[i] = 1;
delete a[];
```

前面两个例子中双索引都是行主序[1782] 的，即第二个索引增加 1，内存地址也增加 1．而这个例子中我们既可以使用行主序也可以使用列主序（第一个索引增加 1，内存地址增加 1），还可以使用单索引．可见这种方式要灵活得多．

注意这只是一个底层的原理，我们需要把这个功能封装到一个矩阵类 Mat 中，在使用的时候可以达到例如以下效果

```
MatDoub a(10, 10); // constructor 会执行 new
// 全部元素赋值为 0
for (long i = 0; i < a.n1(); ++i) // n1() n2() 用于获取行数列数
    for (long j = 0; j < a.n2(); ++j)
        a(i, j) = 0;
// 使用单索引
for (long i = 0; i < a.size(); ++i) // size() 用于获取总长度
    a[i] = 1;

// 改变尺寸，  先用 delete 再用 new 重新分配，  数据丢失
a.resize(20, 20);
```

为了区分行主序和列主序，我们可以设计两个类似的矩阵类 CmatDoub 是列主序矩阵，MatDoub 是行主序矩阵．

我们来看一个简化版的 CmatDoub 定义，不考虑使用 const CmatDoub．

```
class CmatDoub
{
protected:
    int m_N1, m_N2;
    double *p;
public:
    CmatDoub(): m_N1(0), m_N2(0) {}
    CmatDoub(int N1, int N2): m_N1(N1), m_N2(N2) { new double
       m_p[N1*N2]; }
    Doub& operator()(int i, int j) { return p[m_N2*i + j]}
    int n1() { return m_N1; }
    int n2() { return m_N2; }
```

```
    int size() { return m_N1 * m_N2; }
    void resize(int N1, int N2)
    {
        m_N1 = N1; m_N2 = N2;
        if (N1 * N2 != m_N1 * m_N2) {
            delete p[]; new p [N1*N2];
        }
    }
    ~CmatDoub() { delete p[]; }
};
```

用这个矩阵类就可以实现上例中的使用方式.

SLISC 库简介

- 本词条处于草稿阶段.

相对于 fortran 或者 Matlab 等，用 C++ 做数值计算的一个缺陷就是语言本身（或标准库）没有矩阵类型（以及高维矩阵类型）. 但我们可以用第三方库或者自己写一个. 本书选择后者，原因是为了教学需要我们需要保持代码的简单易读，避免复杂的 C++ 语法.

本书中我们将大量使用 **SLISC** 库，即 **Scientific Library in Simple C++**. 代码可以从我们的 GitHub 仓库下载. 该库的特点是尽量不使用 C++ 的复杂语法（如模板）和复杂的类结构，使代码便于阅读学习和修改，同时又保持相对较高的性能. SLISC 使用兼容性高得 C++11 标准.

我们希望把矩阵进行一定程度的封装，但又几乎不损失性能.

流行的矩阵库，如 Eigen，但是代码极其复杂，错误信息不容易懂，修改起来非常困难. 例如 `MatrixXd` 矩阵类就有 6 次继承，一个 2x2 的矩阵在 gdb 里面调试的时候显示出来的变量信息是这样的：

```
<Eigen::PlainObjectBase<Eigen::Matrix<double,-1,-1,0,-1,-1>>> =
{<Eigen::MatrixBase<Eigen::Matrix<double,-1,-1,0,-1,-1>>> =
{<Eigen::DenseBase<Eigen::Matrix<double,-1,-1,0,-1,-1>>> =
{<Eigen::DenseCoeffsBase<Eigen::Matrix<double,-1,-1,0,-1,-1>,
    3>> =
```

```
{<Eigen::DenseCoeffsBase<Eigen::Matrix<double,-1,-1,0,-1,-1>,
    1>> =
{<Eigen::DenseCoeffsBase<Eigen::Matrix<double,-1,-1,0,-1,-1>,
    0>> =
{<Eigen::EigenBase<Eigen::Matrix<double,-1,-1,0,-1,-1>>> =
{<No data fields>}, <No data fields>},
<No data fields>}, <No data fields>}, <No data fields>}, <No
    data fields>},
m_storage = {m_data = 0x855ceb0, m_rows = 2, m_cols = 2}}, <No
    data fields>
```

实际上这里面真正有用的只有最后一行，显示了矩阵在内存中的地址 `m_data`，行数 `m_rows` 以及列数 `m_cols`．这样的库只适合直接拿来用，不适合读和改，尤其是对非计算机专业的人来说．

底层用 MKL 来实现，速度非常快．

所有的函数都尽量使用指针 interface（因为是最兼容的！最笨的办法往往是最灵活的）然后可以再封装一层更友好的 interface

slicing 会有少量的 overhead．由用户自己权衡是使用 slicing 还是指针 interface．

重新定义类型名称

在 C++ 标准中，一些基本类型在内存中的长度在不同计算机上可能会不同．例如 `long` 有时候和 `int` 一样是 4 字节，而另一些时候则和 `long long` 一样是 8 字节．因此，SLISC 仿照 Numerical Recipe 的方式，定义自己的类型名称．以后本书统一使用这些定义．

- `Char` 是 `char`，即 1 字节字符或整数
- `Uchar` 是 `unsigned char`，即 1 字节无符号整数
- `Short` 是 2 字节整数
- `Int` 是 4 字节整数
- `Long` 是 `Int` 或者 `Llong`，取决于宏 `SLS_USE_INT_AS_LONG` 是否定义．SLISC 库种，`Long` 被用来作为矩阵的索引类型．

- `Llong` 是 8 字节整数
- `Float` 是 `float`（4 字节）
- `Doub` 是 `double`（8 字节）
- `Ldoub` 是 `long double`（16 字节）
- `Fcomp` 是 `std::complex<float>`（8 字节）
- `Comp` 是 `std::complex<double>`（16 字节）
- `Lcomp` 是 `std::complex<long double>`（32 字节）

另外我们还定义一些附加类型，用于声明函数的参数. 例如 `Doub_I` 是 `const Doub`，表示函数的输入参数. 又如 `Doub_O` 是 `Doub &`，表示函数的输出参数. `Doub_IO` 也是 `Doub &`，但表示既作为输入也作为输出. 以上的标量类型都同理，在后面加 `_I`，`_O`，`_IO` 分别表示 in，out，in-out.

以后还会看到矢量类和矩阵类，但和标量不同，为了避免不必要的复制，它们都必须 pass by reference 而不是 pass by value. 例如矩阵类 `CmatDoub` 的 `CmatDoub_I` 是 `const CmatDoub &`，`CmatDoub_O` 和 `CmatDoub_IO` 都是 `CmatDoub &`.

SLISC 库编程习惯

- 如无必要尽量不使用 `unsigned` 类型.
- 为了速度，SLISC 中的函数尽量不会改变矩阵尺寸，需要实现手动设置

SLISC 的密矩阵类

- 本词条处于草稿阶段.

预备知识 SLISC 库简介[1788]，数据结构：密矩阵[1782]，C++ 类的定义和继承

我们先以双精度类型 `Doub`（即 `double`）来举例介绍 SLISC 中的密矩阵. 我们习惯上也把 1 维的矩阵称为矢量.

Vbase 类

所有密矩阵都继承于 VbaseDoub，VbaseDoub 不直接使用，仅用于继承. 这里使用继承主要是为了避免在每个密矩阵中都重复定义这些基础功能.

```
class VbaseDoub
{
protected:
    Doub *m_p;
    Long m_N;
public:
    typedef Doub value_type;
    VbaseDoub();
    explicit VbaseDoub(Long_I N);
    VbaseDoub(const VbaseDoub &rhs);

    Doub* p();
    const Doub* p() const;
    Long size() const;
    void resize(Long_I N);
    Doub &operator[](Long_I i);
    const Doub &operator[](Long_I i) const;
    Doub& end();
    const Doub& end() const;
    Doub& end(Long_I i);
    const Doub& end(Long_I i) const;
    void operator<<(VbaseDoub &rhs);
    ~VbaseDoub();
};
```

简单来说，VbaseDoub 就是一个简单的 std::vector，负责内存管理，用方括号 [i] 获取第 i 个矩阵元，用 end() 获取最后一个元，end(i) 获取第 N-i 个元. 用 .size() 获取元素长度，.resize(N) 释放内存并重新分配 N 个元素的新内存. << 用于把一个对象所分配的内存转移到另一个对象中. 所有继承 VbaseDoub 的米矩阵都具有这些功能，除了resize() 会根据矩阵的维度重新定义.

Vec 矢量

为了避免歧义，任何两个矩阵之间禁止使用等号. 如果要复制使用 copy() 函数.

```
class VecDoub : public VbaseDoub
{
public:
    typedef VbaseDoub Base;
    VecDoub() = default;
    explicit VecDoub(Long_I N);
    VecDoub(const VecDoub &rhs);
    VecDoub &operator=(const VecDoub &rhs) = delete;
    void operator<<(VecDoub &rhs);
};
```

Cmat 矩阵

```
class CmatDoub : public VbaseDoub
{
protected:
    typedef VbaseDoub Base;
    Long m_N0, m_N1;
public:
    CmatDoub(): m_N0(0), m_N1(0) {};
    CmatDoub(Long_I N1, Long_I N2);
    CmatDoub(const CmatDoub &rhs);
    CmatDoub &operator=(const CmatDoub &rhs) = delete;
    void operator<<(CmatDoub &rhs);
    Doub& operator()(Long_I i, Long_I j);
    const Doub& operator()(Long_I i, Long_I j) const;
    Long n0() const;
    Long n1() const;
    void resize(Long_I N1, Long_I N2);
    void reshape(Long_I N1, Long_I N2);
};
```

Cmat3 三维数组

```
class Cmat3Doub : public VbaseDoub
{
protected:
    typedef VbaseDoub Base;
    Long m_N0, m_N1, m_N2;
public:
    typedef Doub value_type;
    Cmat3Doub(): m_N0(0), m_N1(0), m_N2(0) {};
    Cmat3Doub(Long_I N1, Long_I N2, Long_I N3);
    Cmat3Doub(const Cmat3Doub &rhs);
    Cmat3Doub &operator=(const Cmat3Doub &rhs) = delete;
    void operator<<(Cmat3Doub &rhs);
    void resize(Long_I N1, Long_I N2, Long_I N3);
    void reshape(Long_I N1, Long_I N2, Long_I N3);
    Doub &operator()(Long_I i, Long_I j, Long_I k);
    const Doub &operator()(Long_I i, Long_I j, Long_I k) const;
    Long n0() const;
    Long n1() const;
    Long n2() const;
};
```

未完成：列出其他矩阵

SLISC 矩阵的基本运算

- 本词条处于草稿阶段.

预备知识 SLISC 的矢量和矩阵[1790]

这里简单列出 `arithmetic.h` 中的函数，用于对矩阵进行基本运算.

`Long size(T_I v)` 相当于 `std::vector` 的 `.size()`，但返回 `Long` 类型，这是为了避免使用 `unsigned` 类型.

`Tr sum_abs(T_I v)` 绝对值求和

`Ts max(T_I v)`，`Ts min(T_I v)` 最大元素和最小元素

`Ts max(Long_O ind, T_I v)`，`Ts min(Long_O ind, T_I v)` 最大元素和最小元素（同时求出其位置）

`Tr max_abs(T_I v)` 最大绝对值

`void mod(T_O v, T1_I v1, Ts1_I s)` 求余

`void mod(T1_O v, Ts1_I s)` 求余

`void real(T_O v, T1_I v1)` 提取实部

`void imag(T_O v, T1_I v1)` 提取虚部

`void abs(T_IO v)` 求绝对值

`void abs(T_O v, T1_I v1)` 求绝对值

`Tret sum(T_I v)` 求和

`Tret prod(T_I v)` 求积

`Tr norm2(T_I v)` 绝对值平方和

`Tr norm(T_I a)` 绝对值平方和再开根号

`void resize_cpy(Tv_IO v, Long_I N, T_I val = 0)` 矢量改变尺寸保留数据（多出数据初始化为 0）

`void resize_cpy(Tv_IO v, Long_I N0, Long_I N1, T_I val = 0)` 矩阵改变尺寸保留数据（多出数据初始化为 0）

`void resize_cpy(Tv_IO v, Long_I N0, Long_I N1, Long_I N2, T_I val = 0)` 3D 改变尺寸保留数据（多出数据初始化为 0）

`void linspace(Tv_O v, Ts_I first, Ts_I last)` 等间距赋值

`void flip(Tv_IO v)` 矢量翻转

`void flip(T_O v, T1_I v1)` 矢量翻转

`void reorder(Tv_O v, To_I order)` 矢量重新排序

`void trans(T_IO v)` 方矩阵转置

`void trans(T_O v, T1_I v1)` 矩阵转置

`void conj(T_IO v)` 共轭

`void conj(T_O v, T1_I v1)` 共轭

`void her(T_IO v)` 厄米共轭

`void her(T_O v, T1_I v1)` 厄米共轭

`void operator+=(T_IO v, Ts_I s)`,`void operator-=(T_IO v, Ts_I s)`,`void operator*=(T_IO v, Ts_I s)`,`void operator/=(T_IO v, Ts_I s)` 矩阵和标量的四则运算

`void operator+=(T_O &v, T1_I v1)`,`void operator+=(T_O &v, T1_I v1)`,`void operator-=(T_O &v, T1_I v1)`,`void operator*=(T_O &v, T1_I v1)`,`void operator/=(T_O &v, T1_I v1)` 矩阵和矩阵的四则运算

`void plus(T_O v, T1_I v1, Ts2_I s)`,`void minus(T_O v, T1_I v1, Ts2_I s)` `void times(T_O v, T1_I v1, Ts2_I s)`,`void divide(T_O v, T1_I v1, Ts2_I s)` 矩阵和标量的四则运算

`void pow(T_IO v, Ts_I s)` 幂运算

`void pow(T_O v, T1_I v1, Ts_I s)` 幂运算

`T dot(T1_I v1, T2_I v2)` 点乘

`void cumsum(T_O v, T1_I v1)` 累积求和

`void mul(T_O y, T1_I a, T2_I x)` 矩阵乘矢量

`void mul(T_IO y, T1_I a, T2_I x, Ts1_I alpha, Ts_I beta)` 矩阵乘矢量（慢）

`void mul(T_O &y, T1_I x, T2_I a)` 行矢量乘矩阵（慢）

`void mul_gen(Ty_O y, Ta_I a, Tx_I &x)` 一般矩阵乘矢量（使用 BLAS）

`void mul_gen(Ty_O &y, Ta_I a, Tx_I x, Tsa_I alpha = 1, Tsa_I beta = 0)` 一般矩阵乘矢量（使用 BLAS）

`void mul_sym(Ty_IO &y, Ta_I a, Tx_I x, Tsa_I alpha = 1, Tsa_I beta = 0` 对称矩阵乘矢量（使用 BLAS）

`void uniq_elm(T_IO v)` 找出不相同的元素（自动 resize）

`void uniq_rows(T_O a, T1_I a1)` 找出不相同的行

`void exp(T_IO v)` 指数函数

`void exp(T_O v, T1_I v1)` 指数函数

SLISC 密矩阵的切割

- 本词条处于草稿阶段.

预备知识 SLISC 的矢量和矩阵[1790]

如果我们想把矩阵的一部分（例如一行）输入到某个函数中怎么办呢？如果我们不使用任何容器而是用指针，函数的输入输出也使用指针，那么可以通过 `step` 参数实现：

```
Doub sum(Doub_I *x, Long_I N, Long_I step)
{
    Doub s = 0;
    for (Long i = 0; i < N; ++i)
        s += x[step*i];
    return s;
}
```

```
Long N0 = 10, N1 = 10; row = 3;
Doub *a; // 列主序矩阵
new Doub a[N0*N1];
for (Long i = 0; i < N0*N1; ++i)
    a[i] = i;
// 对第 3 行求和
cout << "sum a(3,all) = " << sum(a+3, N1, N0) << endl;
// 对第 3 列求和
cout << "sum a(all,3) = " << sum(a+3*N0, N0, 1) << endl;
// 对第 3 列的前 3 个元求和
cout << "sum(0-2,3) = " << sum(a+3*N0, 3, 1) << endl;
delete a[];
```

再看一个例子，如何对子矩阵求和呢？我们需要一个参数叫 leading dimension，该参数表示列标（行主序时：行标）增加 1 时单索引增加的数量.

```
// N0, N1 时子矩阵的尺寸， lda 是原矩阵的行数
Doub sum(Doub_I *a, Long_I N0, Long_I N1, Long_I lda)
{
    Doub s = 0;
    for (Long j = 0; i < N1; ++i) {
        for (Long i = 0; i < N0; ++i)
            s += a[i];
```

```
        a += lda;
    }
    return s;
}
```

使用方法：

```
Long N0 = 10, N1 = 10;
Doub *a;
new Doub a[N0*N1]; // 列主序
for (Long i = 0; i < N0*N1; ++i)
    a[i] = i;
// 对 3-5 行， 4-6 列的子矩阵求和
cout << "sum(3-5,4-6) = " << sum(a+4*N0+3, 3, 3, N0) << endl;
```

可见直接使用裸指针给函数传递数组是非常灵活的，在 C 语言中这也是很常见的做法，见 BLAS[1813] 和 LAPACK．然而这么做的缺点是容易出错，例如索引超出分配的内存，或者忘记用 `delete` 释放内存．这样的错误调试起来会非常困难．

view 类型

在 SLISC 库中，函数的输入或输出除了各种容器外，也可以是 view 类型的一种．例如上面的 `sum` 函数也可以定义为

```
Doub sum(DvecDoub_I x)
{
    Doub s = x[0];
    for (Long i = 1; i < x.size(); ++i)
        s += x[x.step()*i];
    return s;
}
```

其中 `DvecDoub` 是 view 类型的一种，它包含一个指针 `Doub *m_p`，一个长度 `Long *m_N` 以及一个步长 `Long *m_step`．它生成和毁灭时不会分配或释放内存，它只用于表示另一个容器的一部分数据，所以叫做 view，也叫 slicing．

SLISC 提供一系列开头为 `cut` 的函数，可以从容器中截取一部分然后返回 view 类型的一种．例如以上 `sum` 函数可以这样使用

```
CmatDoub a(10, 10);
// 对第 3 行求和
cout << "sum row 3 = " << sum(cut1(a, 3)) << endl;
```

其中 `cut1(a, 3)` 从矩阵 `a` 中切下第 3 行（`1` 表示第 1 个维度，即“行”），然后返回一个临时的 `const DvecDoub &` 类型传给 `sum` 函数.

类似地，`cut0(a, 3)` 从矩阵 `a` 中切下第 3 列（`0` 表示第 0 个维度，即“列”），然后返回一个临时的 `const SvecDoub &` 类型. `SvecDoub` 是一个密矢量的 view，没有 `m_step` 成员，所有元素在内存中都是相邻的. 在 view 类型命名中，`S` 代表 smooth，`D` 代表 dash.

同理，如果要切割一个子矩阵，用 `cut(a, i, N0, j, N1)` 可以切出一个 `DcmatDoub` 类型，它的成员包括一个起始指针 `Doub *m_p`，子矩阵尺寸 `Long *m_N0, *m_N1`，总长度 `Long m_N`，以及 leading dimension `m_lda`.

view 类型和指针一样，也分为 low level const 和 high level const，`const SvecDoub` 代表 high level const，也就是它一旦初始化后数据成员就都是 const，其中 `m_p` 是 high level，即`const Doub *`. view 的 low level const 则有专门的类型如 `SvecDoub_c`，它的成员可以随意改变，但所代表的数据不能被修改.

在函数参数中，`SvecDoub_I` 的定义是 `const SvecDoub_c &` 用于输入；`SvecDoub_O, SvecDoub_IO` 的定义是 `const SvecDoub &`，用于输出.

未完成：具体介绍一下所有类型

SLISC 的 Mcoo 矩阵类

- 本词条处于草稿阶段.

预备知识 SLISC 的密矩阵类[1790]

当一个矩阵对象中有大量 0 元的时候，用密矩阵数据结构会浪费内存和计算量. COO 是一种常用的稀疏矩阵，它的数据结构是把所有的非零元存放在一个矢量中，并用另外两个矢量分别记录这些非零元的行标和列标.

在本书的SLISC库中,我们根据这种数据结构定义了 McooDoub, McooComp 等矩阵类,它和密矩阵一样继承一个 Vbase 类来储存非零矩阵元,另外 m_N0, m_N1 用于储存矩阵尺寸，m_Nnz 储存非零元的长度，该长度未必是 Vbase 的长度，后者通过 reserve(Long_I) 成员函数来调整尺寸. 声明如下：

```
class McooDoub : public VbaseDoub
{
private:
    typedef VbaseDoub Base;
    using Base::m_p;
    using Base::m_N;
    Long m_N0, m_N1, m_Nnz;
    VecLong m_row, m_col;
public:
    using Base::p;
    McooDoub(): m_N0(0), m_N1(0), m_Nnz(0) {};
    McooDoub(Long_I N0, Long_I N1);
    McooDoub(Long_I N0, Long_I N1, Long_I Ncap);
    Long *row_p();
    const Long *row_p() const;
    Long *col_p();
    const Long *col_p() const;
    void push(Doub_I s, Long_I i, Long_I j);
    void set(Doub_I s, Long_I i, Long_I j);
    Long n0() const;
    Long n1() const;
    Long size() const;
    Long nnz() const;
    Long capacity() const;
    Long find(Long_I i, Long_I j) const;
    Doub &ref(Long_I i, Long_I j);
    Doub operator()(Long_I i, Long_I j) const;
    Doub &operator[](Long_I ind);
    Doub operator[](Long_I ind) const;
    Long row(Long_I ind) const;
    Long col(Long_I ind) const;
```

```
    void trim(Long_I Nnz);
    void resize(Long_I N);
    void reserve(Long_I N);
    void reshape(Long_I N0, Long_I N1);
};
```

SLISC 的 band 类

- 本词条处于草稿阶段.

预备知识 数据结构：带对角矩阵[1783]，SLISC 的密矩阵类[1790]

SLISC 把 BLAS[1813] 中的带对角矩阵封装成了 Cband 类（列主序．行主序暂时没有实现）．其中矩阵元储存于 m_a 中，m_N0 是行数，m_a.n1() 是列数．m_Nup, m_Nlow 分别是对角线上方和下方非零副对角线的条数，m_a 的第 m_idiag 行则是对角线所在的行．

```
class CbandDoub
{
public:
    Long m_N0;
    Long m_Nup;
    Long m_Nlow;
    Long m_idiag;
    CmatDoub m_a;

    CbandDoub();
    CbandDoub(Long_I N0, Long_I N1, Long_I Nup, Long_I Nlow,
              Long_I lda = -1, Long_I idiag = -1);
    Doub * p();
    const Doub * p() const;
    Doub operator()(Long_I i, Long_I j) const;
    Doub &ref(Long_I i, Long_I j);
    Long n0() const;
```

```
18      Long n1() const;
19      Long size() const;
20      Long nup() const;
21      Long nlow() const;
22      Long lda() const;
23      Long idiag() const;
24      CmatDoub &cmat();
25      const CmatDoub &cmat() const;
26      DcmatDoub band();
27      DcmatDoub_c band() const;
28      DvecDoub diag();
29      DvecDoub_c diag() const;
30      void resize(Long_I lda, Long_I N1);
31      void resize(Long_I N0, Long_I N1, Long_I Nup,
32                  Long_I Nlow, Long_I lda = -1, Long_I idiag = -1)
                       ;
33      void reshape(Long_I N0, Long_I Nup, Long_I Nlow, Long_I
           idiag = -1);
34      void shift(Long_I idiag);
35  };
```

SLISC 的文件读写

- 本词条处于草稿阶段.

`Bool file_exist(Str_I fname, Bool_I case_sens = true)` 判断指定文件是否存在

`Bool dir_exist(Str_I path)` 判断路径是否存在

`Long file_size(Str_I fname)` 获取文件大小（字节数），`fname` 是文件名.

`Str path2dir(Str_I fname)` 从“路径 + 文件”字符串中提取“路径”部分

`void mkdir(Str_I path)` 创建路径

`void rmdir(Str_I path)` 删除路径

`void ensure_dir(Str_I dir_or_file)` 如果路径不存在，创建该路径（可以任意多层）

`void file_list(vecStr_O fnames, Str_I path, Bool_I append = false)` 列出指定路径中所有文件，`append = true` 则 `fnames` 原有的元素不被清空.

`void file_list_full(vecStr_O fnames, Str_I path, Bool_I append = false` 和 `file_list()` 相同，只是文件名包含路径.

`void folder_list(vecStr_O folders, Str_I path, Bool_I append = false)` 列出指定路径中所有文件夹

`void folder_list_full(vecStr_O folders, Str_I path, Bool_I append = fa` 和 `folder_list()` 相同，但包含路径.

`void file_list_r(vecStr_O fnames, Str_I path, Bool_I append = false)` 列出指定路径中所有文件，包括所有子文件夹中的.

`void file_ext(vecStr_O fnames_ext, vecStr_I fnames, Str_I ext, Bool_I` 在文件列表中选出所有具有某个拓展名的文件

`void file_list_ext(vecStr_O fnames, Str_I path, Str_I ext, Bool_I keep` 在某个路径中寻找所有具有某个拓展名的文件

`void file_copy(Str_I fname_out, Str_I fname_in, Bool_I replace = false` 复制文件 `replace` 用于替换目标的同名文件，否则会一直提示，需要手动删除程序才能继续.

`void file_copy(Str_I fname_out, Str_I fname_in, Str_IO buffer, Bool_I` 用户提供缓存 `buffer` 的 `file_copy`，提高速度.

`void file_move(Str_I fname_out, Str_I fname_in, Bool_I replace = false` 移动文件.

`void file_move(Str_I fname_out, Str_I fname_in, Str_IO buffer, Bool_I` 带用户缓存的版本.

`void open_bin(ofstream &fout, Str_I fname)` 打开二进制文件写入

`void open_bin(ifstream &fin, Str_I fname)` 打开二进制文件读取

`void write(const Char *data, Long_I Nbyte, Str_I fname)` 把字符串写入二进制文件

`Long read(Char *data, Long_I Nbyte, Str_I fname)` 读取二进制文件到字符串，返回实际上读取到的字节数.

`void write(Str_I str, Str_I fname)` 把字符串写入二进制文件

`void write_vec_str(vecStr_I vec_str, Str_I fname)` 把每个字符串写入文件中的一列

`void read_vec_str(vecStr_O vec_str, Str_I fname)`

`void read(Str_O str, Str_I fname)` 读取二进制文件到字符串

`void write(ofstream &fout, Char_I s)` 把一个 `Char` 写入二进制文件，也可以是 `Int`, `Llong`, `Doub`, `Comp`, `Str`.

`void read(ifstream &fin, Char_I s)` 和 `write` 同理.

`void read(CmatInt_O mat, Str_I file, Long_I skip_lines = 0)` 从一个文件中读取密矩阵，矩阵会被 `.resize` 成需要的形状. `CmatInt` 也可以是 `CmatLlong` 或 `CmatDoub`

`void read(VecDoub_O v, Str_I file, Long_I skip_lines = 0)` 从文件中读取单列矢量.

`void last_modified(Str_O yymmddhhmmss, Str_I fname)` 获取文件的最后修改时间.

`void set_buff(ofstream &fout, Str_IO buffer)` 缓存越大，速度越快.

`Bool little_endian()` 操作系统是否是 `little_endian` 的.

`void change_endian(Char *data, Long_I elm_size, Long_I Nelm)` 把一个字符串中每个长度为 `elm_size` 的单元前后翻折.

SLISC 的计时工具

预备知识 SLISC 库简介[1788]

SLISC 库的 `timer.h` 头文件提供了时间相关功能：

- `Str hhmmss()` 返回当前系统时间字符串 `hh:mm:ss`
- `Str hhmmss(Long sec)` 把秒转换为字符串 `hh:mm:ss`
- `Str hhhmmss(Long sec)` 把秒转换为字符串 `h....h:mm:ss` 其中小时可以有任意多位

- `Timer` 类用于计算物理时间（区别于 cpu 时间）`Timer::tic()` 开始计时，`Doub Timer::toc()` 返回流逝的时间（单位：s）
- `CPUTimer` 和 `Timer` 类似，只是用于衡量 CPU 时间.
- `void pause()` 暂停程序直到输入任意字符并按回车，也可以直接按回车.
- `void pause(Doub t)` 暂停程序 `t` 秒.

SLISC 的 matt/matb 文件格式

- 本词条存在未完成的内容.

预备知识 SLISC 简介

SLISC 库提供了两种特殊的文件格式用于储存各种变量和矩阵，保存时可以保存变量名，读取时可以指定要读取的变量名. 两种文件分别使用拓展名 `.matt` 和 `.matb`. 其中 mat 表示 matrix，t 表示 text，t 表示 binary. 顾名思义，第一种是文本文件，第二种是二进制文件. 二进制文件具有占用硬盘空间少，读写速度快的优点. 文本文件可以方便使用其他语言读取.

以下给一个 matb 文件的读写例子，也可以把例子中的 matb 换成 matt.

未完成：matt 的模式目前使用字符串，应该统一一下代码

```
1 #include "SLISC/matb.h"
2 using namespace slisc;
3
4 int main()
5 {
6     // 随便初始化一些变量， 目前支持 SLISC 库中的绝大部分标量和
        密矩阵
7     Int i = 1; Doub d = 3.1; Comp c(1, 2);
8     VecInt v(3); linspace(3, 1, 3); CmatInt a(2,2); linspace(a);
9
```

```
    Matb matb("test.matb", 'w'); // 打开 matb 文件，使用写入模式
    // 相当于 Matb matb; matb.open("test.matb", 'w');
    save(i, "i", matb); // 保存变量，指定变量名，可以是任意字符串
    save(d, "d", matb); // 注意同个文件中保存的变量名不能重复
    save(c, "c", matb);
    save(v, "c", matb);
    save(a, "a", matb);
    matb.close(); // 关闭文件.
    // 在 destructor 中会自动调用，open() 时也会自动调用

    Int i1; Doub d1; Comp c1; VecInt v1(3); CmatInt a1;
    matb.open("test.matb", 'w'); // 重新用读取模式打开文件
    // 读取变量，可以按照任何顺序，不需要全部读取
    // 矩阵会被自动 resize()
    load(i1, "i", matb);
    load(d1, "d", matb);
    load(c1, "c", matb);
    load(v1, "c", matb);
    load(a1, "a", matb);
    matb.close();

    // 此时可以对比读取的变量是否和保存的变量一致
}
```

文件格式

我们先介绍二进制格式 matb．matb 文件使用 little endian 储存（链接未完成），每个变量的前 8 个字节是变量名的长度 `Nname`，所有整数都用 `long long` 储存（下同），然后是 `Nname` 字节的变量名，紧接着是 8 个字节的矩阵元类型编号（对照表未完成），8 个字节的矩阵维度 `Ndim`（标量的维数是 0，向量维度是 1）接着 `8*Ndim` 字节记录每个维度的长度（标量不需要）．最后，再记录 `N * sizeof(元素类型)`，其中 `N` 是元素个数，等于各个维度的长度相乘．

在文件的最后，若一共有 `Nvar` 个变量，那么就用 `Nvar*8` 字节记录下每个

变量的第一个字节在文件中的字节数（第一个变量总是从第 0 个字节开始）（未完成：目前是倒序的，这似乎没什么必要），然后再用 8 字节记录下 Nvar，这是为了读取时查找方便．最后为了验证文件的完整性，用 16 字节写入字符串 "Matb_End_of_File"．该字符串用于读取时验证文件的完整性，若不存在则说明文件被损坏．

对于文本格式 matt，每个整数或浮点数直接转换成字符串写入文本，每两个数之间使用空格隔开，复数则写成两个相邻的数（未完成：现在代码并不是这么做的，而是类似 1+2i）．

Matb 类

```
// all non-data intergers in file are Llong
class Matb {
public:
    Matb();
    Matb(Str_I fname, Char_I rw);
    Char m_rw; // 模式
    ifstream m_in;
    ofstream m_out;
    Llong m_filesize; // 文件大小
    Str m_fname; // 文件名
    vecStr m_name; // 变量名
    vecLlong m_type; // 变量类型
    vector<vecLlong> m_size; // 矩阵尺寸
    vecStr m_data; // 矩阵元二进制数据（仅 'm' 模式使用）
    vecLlong m_ind; // 变量位置

    void open(Str_I fname, Char_I rw); // 打开文件
    Bool isopen(); // 文件是否打开

    // 内部函数
    void read_data(); // 读取到 m_data
    void write_data(Str_I fname); // 写入到 m_data
    Long size() const; // 变量个数
    Long data_size(Long_I i) const; // 第 i 个变量的 m_data 长度
```

```
25      void get_profile(); // 读取变量信息
26      Long data_pos(Long_I var_ind) const; // 第 i 个变量在文件中
            的位置
27      Long search(Str_I name); // 搜索变量编号, 如果不存在返回 -1
28      ~Matb();
29  };
```

修改模式

打开 matb 的模式除了 'r', 'w' 外还可以是 'm'，表示 modify. 在该模式下，打开文件后可以对变量进行增加，删除，修改等操作. 该功能尚未完善.

其他工具

可执行文件 SLISC/util/matbinfo 可以显示 matb 文件中的所有变量信息，用法是 matbinfo 文件名.matb.

可执行文件 SLISC/util/matt2matb 和 matb2matt 可以把 matt 文件和 matb 文件互相转换,例如 matb2matt -d -r 文件名.matb 会生成 文件名.matt. 其中 -d 选项在转换后把原文件删除，如果 文件名.matt 已经存在，-r 选项会重新转换并将其替换.

C++ 中的 SFINAE 技巧

- 本词条缺少预备知识，初学者可能会遇到困难.

限制模板范围

例如我们有一个函数模板将两个对象 v1，v2 相加得到 v

```
1  template <class T, class T1, class T2>
2  void Plus(T &v, const T1 &v1, const T2 &v2)
3  {...}
```

如果我们想要对不同类型的 T，T1，T2 执行不同的代码，我强烈推荐下面要介绍的 SFINAE 技巧来实现.

我们先来定义一个宏函数 MY_IF()（至于为什么要这么定义先不介绍），其中使用了 type_traits 头文件中的 enable_if.这个宏用于输入一个 constexpr 的 bool 表达式，由于表达式中有可能出现一个或多个逗号，所以形式上我们用了任意变量的宏函数.

```
#define MY_IF0(...) typename std::enable_if<(bool)(__VA_ARGS__),
    Int>::type
#define MY_IF(...) MY_IF0(__VA_ARGS__) = 0
```

这里先定义了 MY_IF0()，看起来多此一举，但实际上在一些情况下我们也需要单独使用 MY_IF0().

另外，我们假设存在一些函数模板用于判断 T, T1, T2 的类型（具体怎么定义先不介绍），例如如果 T 是一个矩阵，is_matrix<T>() 就返回 true，而 is_scalar<T>() 和 is_vector<T>() 都返回 false.

```
template <class T> bool is_scalar();
template <class T> bool is_vector();
template <class T> bool is_matrix();
```

现在我们用 MY_IF 来区分不同版本的 Plus 函数. 标量相加的函数如下

```
template <class T, class T1, class T2,
MY_IF(is_scalar<T>() && is_scalar<T1>() && is_scalar<T2>())>
void Plus(T &v, const T1 &v1, const T2 &v2)
{ v = v1 + v2; }
```

注意在我们通过 MY_IF 声明了这个模板什么时候有定义（只有 T, T1, T2 为标量时有定义，例如 int, double, complex 等）.

同样，我们可以再写一个版本的 Plus 定义矢量相加

```
template <class T, class T1, class T2,
MY_IF(is_vector<T>() && is_vector<T1>() && is_vector<T2>())>
void Plus(T &v, const T1 &v1, const T2 &v2)
{
    for (int i = 0; i < v.size(); ++i) {
        v[i] = v1[i] + v2[i];
    }
}
```

我们还可以再定义矩阵相加，矩阵与标量相加，矢量与标量相加等等.

如果我们的函数需要先 declare 再 define（例如 class 的成员函数在 class 定义中 declare，然后在别的地方 define），那就需要在 declaration 中使用 `MY_IF()`，而在 definition 中使用 `MY_IF0()`，其他内容都一样.

使用 SFINAE 的好处是，我们可以限制函数模板 instantiate 的条件，使得一些不合法的使用变得没有定义（比如说我想要用 Plus 把矩阵和矢量相加，又比如复数矩阵相加得到实数矩阵）. 另一个好处是无论我们定义多少个版本的 Plus，只要 `MY_IF` 中的条件总在唯一一个版本中为 `true`，编译器就不会抱怨无法判断使用哪个版本的 Plus 函数.

类型判断

我们下面来介绍这些函数如何实现.

首先，它们必须是 constexpr 函数，也就是说它们必须要能在编译阶段（而不是运行阶段）被调用并返回结果. 其次，它们是模板函数，因为类型 T 不可能作为函数参数，而只能作为模板参数.

我们首先定义 `is_same<T1, T2>()` 函数模板来判断 `T1, T2` 两个类型是否相同（这里用到了标准库的 `type_traits` 头文件）

```
template <class T1, class T2>
constexpr bool is_same()
{ return std::is_same<T1, T2>::value; }
```

注意在 `type_traits` 中，`is_` 开头的函数都是类模板而不是函数模板，理论上我们可以直接拿 `std::is_same` 来用，但为了概念和使用上更简单，我们重新定义了同名的函数模板.

有了 `is_same`，我们就可以很容易地实现 `is_int<T>()`, `is_double<T>()`, 等.

```
template <class T>
constexpr bool is_int()
{ return is_same<T, int>(); }

template <class T>
constexpr bool is_double()
{ return is_same<T, double>(); }
```

`is_complex<T>()` 有所不同，因为 `std::complex` 本身也是一个类模板，可以有不同的类.这里为了简单，姑且就假设我们只使用 `std::complex<double>`. 如果需要支持任意类型的 `std::complex<>`，需要使用下文中定义 `is_vector` 的方法.

```
template <class T>
constexpr bool is_complex()
{ return is_same<T, std::complex<double>>(); }
```

为了简单起见，我们假设“标量”只包括 `int, double, std::complex<double>` 三种，于是可以定义用于判断标量的函数

```
template <class T>
constexpr bool is_scalar()
{ return is_int<T>() || is_double<T>() || is_complex<T>(); }
```

现在我们再来看如何定义 `is_vector<T>()`. 对于矢量，我们既可以使用 `std::vector<>`，也可以自己定义一个矢量类型. 我们通过 template specialization 来实现 `is_vector<>`

```
template <class T> struct is_vector_imp : std::false_type {};
template <class T> struct is_vector_imp<vector<T>> : std::
   true_type {};
template<class T>
constexpr bool is_vector()
{
    return is_vector_imp<T>();
}
```

首先我们对一般的类型 `T` 定义的 `is_vector_imp`（imp 这里表示 implementation），继承 `std::false_type`（你只需要知道 `falst_type` 是一个类，它的对象可以自动转换为 false，`true_type` 类的对象可以自动转换为 `true`）. 然后对凡是符合 `is_vector_imp<vector<T>>` 格式的类进行不同的定义：即继承 `true_type`. 最后，我们再把 `is_vector_imp` 已经实现的功能封装成函数模板 `is_vector<T>()`，就大功告成了.

用于判断矩阵类型的 `is_matrix<T>()` 也可以如法炮制

```
template <class T> struct is_matrix_imp : std::false_type {};
template <class T> struct is_matrix_imp<Matrix<T>> : std::
   true_type {};
template<class T>
constexpr bool is_matrix()
{
    return is_vector_imp<T>();
}
```

然而注意标准库中没有矩阵类型（据说其实有一个，后来烂尾了，几乎没人用），所以我们一般自己定义矩阵类型 `Matrix<T>`.

第五章

其他

BLAS 简介

预备知识 C 语言基础

BLAS（basic linear algebra subroutine）是一系列基本**线性代数运算**函数[1]的**接口（interface）**标准．这里的线性代数运算是指例如矢量的线性组合，矩阵乘以矢量，矩阵乘以矩阵等．接口在这里指的是诸如哪个函数名实现什么功能，有几个输入和输出变量，分别是什么．

BLAS 被广泛用于科学计算和工业界，已成为业界标准．在更高级的语言和库中，即使我们不直接使用 BLAS 接口，它们也是通过调用 BLAS 来实现的（如 Matlab 中的各种矩阵运算）．

BLAS 原本是用 Fortran 语言写的，但后来也产生了 C 语言的版本 cBLAS，接口与 Fortran 的略有不同（例如使用指针传递数组），但大同小异．

注意 BLAS 是一个接口的标准而不是某种具体**实现（implementation）**．简单来说，就是不同的作者可以各自写出不同版本的 BLAS 库，实现同样的接口和功能，但每个函数内部的算法可以不同．这些不同导致了不同版本的 BLAS 在不同机器上运行的速度也不同．

BLAS 的官网是 Netlib，可以浏览完整的说明文档以及下载源代码．这个版本的 BLAS 被称为 reference BLAS，运行速度较慢，通常被其他版本用于衡量性能．对于 Intel CPU 的计算机，性能最高的是 Intel 的 MKL （Math Kernel Library） 中提供的 BLAS．MKL 虽然不是一个开源软件，但目前可以免费下载使用．如果想要免费开源的版本，可以尝试 OpenBlas 或者 ATLAS[2]．另外，无论是否使用 MKL，BLAS 的文档都推荐看 MKL 的相关页面．

使用 cBLAS

[1]编程中的函数，不是数学上的函数，在一些编程语言（如 Fortran）中也叫**子程序（subroutine）**．可以简单地认为 Fortran 函数有返回值，子程序没有，如下面的例子中，传进和传出的是指针

[2]至于安装，在 Windows 系统上作者推荐在 Visual Studio 的基础上安装 MKL 或者 Parallel Studio，这些软件都比较大，可能需要较长时间下载安装．在 Linux 系统上，可以直接用 apt-get 等安装开源版本，也可以下载 intel MKL 的 deb 安装包按照提示安装．

预备知识 矩阵的储存[1782]

我们这里介绍一个 cBLAS 的例子[3]．我们来测试矩阵和矢量的乘法函数 gemv，参考文档见这里．

BLAS 可以分为四种类型和三个级别

类型

- single
- double
- complex
- double complex

级别（level）

1. 向量与向量操作（复杂度 n）
2. 矩阵与向量操作（复杂度 n^2）
3. 矩阵与矩阵操作（复杂度 n^3）

根据上述的类型和级别，BLAS 有一套系统的命名规则，我们需要的函数名格式为 cblas_?gemv，其中 cblas_ 是固定前缀，问号表示 s, d, c, z, 中的一个，分别代表单精度（float），双精度（double），单精度复数和双精度复数[4]．ge 表示输入的矩阵是一个一般的矩阵[5]，以行主序或者列主序[1782] 线性储存，mv 表示矩阵与向量操作．

我们接下来以 cblas_zgemv 为例，先来看函数声明．

```
1 void cblas_zgemv (const CBLAS_LAYOUT Layout, const
     CBLAS_TRANSPOSE trans,
2 const MKL_INT m, const MKL_INT n, const void *alpha, const void
     *a,
```

[3]要在 Linux 上编译，见“在 Linux 上编译第一个 C++ 程序[1764]”

[4]注意 C 语言中内置的复数类型和 C++ 标准库中的 std :: complex<> 不同，但由于 cBLAS 用指针 void * 传递复数（任何指针都可以自动转换），所以只要内存中用两个连续的 float 或 double 表示一个复数即可

[5]而不是对称矩阵或者厄米矩阵等，后者可以使用专门的函数使性能提升，例如 symv

```
const MKL_INT lda, const void *x, const MKL_INT incx, const void
    *beta,
void *y, const MKL_INT incy);
```

BLAS 接口给人的第一感觉就是冗长，为什么实现一个简单的功能需要这么多变量？因为这个接口具有相当大的灵活性．例如可以使用一个列主序（行主序）矩阵的子矩阵作为矩阵，又例如可以使用一个列主序（行主序）矩阵的某一行（列）作为矢量；例如可以在做乘法以前对矩阵进行转置[6]；又例如可以把相乘的结果累加到输出矢量已有的值上，而不是直接覆盖．

作为一个最简单的例子，我们可以用以下测试函数，在这个例子中，我们只需要关心几个变量，即 Layout（一个 enum 类型）指定行主序（CblasRowMajor）还是列主序（CblasColMajor），m 和 n 分别指定矩阵的行数和列数，a, x, y 分别是矩阵第一个元的指针．

代码 1: blas_demo.cpp

```
// 计算矩阵—矢量乘法 y = a * x
#include <iostream>
#include <complex>
#include <cblas.h>
int main() {
    using namespace std;
    typedef complex<double> Comp; // 定义复数类型
    int Nr = 2, Nc = 3; // 矩阵行数和列数
    // 矩阵和矢量分配内存
    Comp *a = new Comp [Nr*Nc];
    Comp *x = new Comp [Nc];
    Comp *y = new Comp [Nr];
    Comp alpha(1, 0), beta(0, 0);
    // x 矢量赋值
    for (int i = 0; i < Nc; ++i) {
        x[i] = Comp(i+1., i+2.);
    }
    // a 矩阵赋值
```

[6]其实这个转置操作实现起来并不需要额外的运算，在函数的代码中只需要把列主序（行主序）矩阵看作是行主序（列主序）的即可．所以这么做比事先在内存中将矩阵元调换的方法快许多

```
    for (int i = 0; i < Nr*Nc; ++i) {
        a[i] = Comp(i+1., i+2.);
    }
    // 做乘法
    cblas_zgemv(CblasColMajor, CblasNoTrans, Nr, Nc, &alpha, a,
        Nr, x, 1, &beta, y, 1);

    // 控制行分别输出 x, a, y
    for (int i = 0; i < Nc; ++i) {
        cout << x[i] << "  ";
    }
    cout << "\n" << endl;
    for (int i = 0; i < Nr; ++i) {
        for (int j = 0; j < Nc; ++j) {
            cout << a[i + Nr*j] << "  ";
        }
        cout << endl;
    }
    cout << "\n" << endl;
    for (int i = 0; i < Nr; ++i) {
        cout << y[i] << "  ";
    }
}
```

编译与运行结果

```
$ g++ test_cblas.cpp -cblas
$ ./a.out
(1,2)  (2,3)  (3,4)

(1,2)  (3,4)  (5,6)
(2,3)  (4,5)  (6,7)

(-18,59)  (-21,74)
```

其他参数

- `trans`（也是`enum`）指定预先转置矩阵（`CblasTrans`）或不转置（`CblasNoTrans`）
- `alpha, beta` 用于计算 `y = alpha * a * x + beta y`，上例中我们使用了 `alpha = 1, beta = 0`
- `lda` 叫做 leading dimension，在列主元的情况下用于指定从某一列第 i 个元的索引（index）减去上一列第 i 元的索引．在上例中 `lda` 就是行数．但如果我们需要在一个更大的列（行）主元矩阵中截取一个小矩阵进行计算，那么 `lda` 就是大矩阵的行（列）数．
- `incx` 和 `incy` 用于指定两个矢量的步长（increment），由于上例中两矢量在内存中是连续的，所以步长为 1．但若某个矢量是行（列）主元矩阵中的一列（行），那么步长就是矩阵的列（行）数．

GNU Scientific Library

- 本词条处于草稿阶段．
- 本词条缺少预备知识，初学者可能会遇到困难．

GNUGNU Scientific Library (GSL) 是一个 GNU 开源项目，由 C 语言写成．

安装

在 Debian 系统上可执行 `sudo apt install libgsl-dev` 安装．
第二种方法是自己下载源码编译．

使用

在编译时加上 linker 选项 `-lgsl` 即可．

源码结构

文本文件与字符编码

文本文件

我们这里讨论的是**文本文件（text file）**，也就是 Windows 中为人熟知的 txt 拓展名文件．文本文件只能储存字符，也就是说文件中的信息只能是“有几个字符”以及“每个字符是什么”．所以文本文件本质上不包括字体，字号，下划线等等的其他信息．一个文件是否是文本文件是由它的储存格式决定的，而不是由拓展名决定的．一些其他的文件拓展名如 html，xml，md，以及大部分编程语言的代码文件都是文本文件．

我们把所有非文本文件统称为**二进制文件（binary file）**，因为从原理上来说任何文件都是以二进制的形式储存的．例如 Word 文档保存的 doc 或 docx 拓展名文件就是二进制文件，因为里面用其特定的格式储存了许多其他信息．又例如图片文件 jpg, png，视频文件 mp4 等也都是二进制文件．

文本文件的一个基本的问题是，字符以什么形式保存？从概念上来说，文本文件中每个字符对应一个整数，我们把对应的规则叫做**编码（encoding）**，一个文本文件一般只使用一种编码．例如在著名的 ASCII （读作 asky）编码中，128 个字符被一一对应到 0-127 的整数[7]，这些字符包括大写和小写字母，数字，常见标点，以及一些格式上的符号如空格，换行符，制表符等．ASCII 编码可以满足通常的英语写作需求，但显然不支持其他语言如中文．下文会看到中文的文本文档通常以 UTF-8 （Unicode 的一种）或者 GBK 编码储存．

遗憾的是，一般情况下文本文档中不会声明使用的编码，所以如果储存和打开文件的默认编码方式不一致，就会导致显示乱码．尽管现在大部分系统和软件都用 Unicode 编码保存和打开文本文件，但仍然有一些软件为了兼容性使用以前的编码，例如 Windows 写字板（Notepad，不建议使用）默认将中文保存为 GBK（即中文环境下所谓的 ANSI）编码．

如果经常编辑文本文档（尤其是编程），我们推荐使用一款功能较完善的编辑器，例如 Visual Studio Code （简称 VScode）[8]，Notepad++ 等．这里我们

[7] 另外有 **extended ASCII** 编码对应 0-255 的整数

[8] 注意 Visual Studio Code 不是 Visual Studio，是一款轻量级文本编辑器，可以通过安装插件拓展许多功能．

使用 VScode 编辑器为例进行介绍.

当 VScode 打开一个文本文档时，会尝试自动检测编码，并显示在右下角的状态栏图 1．如果这时编辑器中仍然显示乱码，说明自动检测失败，可以尝



图 1: VScode 的右下角状态栏显示 Tab 的显示尺寸为 4 个空格，编码为 UTF-8，换行符为 LF，文件为普通文本文档

试手动切换不同编码打开文件：点击状态栏上显示的编码，选择“Reopen with Encoding”，选择需要的编码即可．以正确的编码打开后，如果想改为别的编码，就选“Save with Encoding”.

换行符

除了编码外，文本文件的换行符使用也有不同的规范．历史上，人们在使用机械打字机时，当一行打完，打字头停留在页面某一行的最右端．回车键（carrage return，简写为 CR）用于将打字头移动到页面的最左端，而 LF （line feed）键用于将纸张上移一行．在 ASCII 编码中，CR 和 LF 分别对应 13 和 10. Windows 系统沿用了这个传统，所以当我们在写字板程序中按下回车时，分别会在文件中插入 CR 和 LF 两个字符．Linux 系统默认使用单个 LF 换行，而 MacOS 则默认用单个 CR 换行．所以同一个文件在不同系统的默认编辑器中可能会出现全部文字排成一行的情况．一般来说，同一个文本文件应该使用同一个规范.

无论在哪个系统中使用 VScode，它在打开文本文件时都会自动检测换行符并显示在图 1 的状态栏中[9]．若要将改变换行符，点击当前显示的换行符并选择即可.

ASCII 编码

（表格未完成）

其他绝大部分编码（如 Unicode，GBK）的前 128 个数字代表的字符都与 ASCII 相同，所以一个 ASCII 编码的文件用任何编码方式打开的结果都极有可能是相同的.

[9]一般不会检测失败，除非不同的换行符格式混用

Unicode 编码

Unicode 编码可以同时支持许多（人类）语言的字符，还包含一些图形（如箭头，多边形，数学符号等等）和 emoji 表情．它将每个字符和 0-4294967295（即 $2^{32}-1$）范围的整数一一对应．Unicode 分为几种不同的具体编码，UTF-32 是定长度的，即将每个字符都储存为 4 个字节（32 比特）．最常见的 UTF-8 是变长度的，即如果字符对应的数字较小，就储存为 1 个字节（8 比特），随着数字变大，可以最多储存为 4 个字节．UTF-16 也是变长度的，原理和 UTF-8 类似，但一个字符最少对应两个字节．

如今 UTF-8 是使用最广泛的文本文档编码，我们推荐尽可能使用 UTF-8 储存文本文档[10]．由于 UTF-8 最开始的部分是 ASCII 编码，所以如果文件中只含有 ASCII 字符，文件大小和 ASCII 编码的文件是一样的．

Byte Order Mark

在 Unicode 编码的文件开头（很少用于 UTF-32），有一些软件会插入（或需要）2 到 3 个字节的 **BOM（byte order mark）**，这样一是可以表明文件使用 Unicode 编码，二来可以表明字节排列顺序．当一个整数需要用多个字节表示时，不同的系统可能会把这些字节倒置，即所谓的 big endian 和 small endian．

VScode 同样可以自动检测文件是否存在 BOM，如果有则显示在状态栏的编码中，例如“UTF-8 with BOM”．

正则表达式

- 本词条处于草稿阶段．

预备知识 文本文件与字符编码[1818]

[11]在文本文件中搜索内容的时候，有时候想要的是某种格式而不是某些具体的字符，例如要搜索“* 月 * 日消费 ** 元”，找到匹配项后需要选中这几

[10]本书介绍的 Matlab 编程语言，从 R2020a 版本开始默认使用 UTF-8 储存代码文件，在此之前中文系统的代码文件默认以 GBK 储存．而大部分编程语言很早以前就使用 UTF-8 编码了．

[11]参考 Matlab 文档相关页面．

个字（以便进行替换等操作），又或者仅选中“元”前面的数值（以便进行统计等）．理论上我们可以通过编程解决这个问题，但更简单地，可以用一种广为使用的表达式来达到同样的效果，就是下面要介绍的**正则表达式（regular expresion）**.

正则表达式在许多软件中都被支持，例如在常用的文本编辑器（如 VS-code），搜索软件（如 Fileseek），和大部分编程语言（如 C++，python，Matlab）中都有很好的支持.

字符匹配

表 1: 字符匹配

符号	说明
`.`	匹配单个任意字符，包括空格回车等
`[...]`	匹配方括号中的任意一个字符．如果要表示一个范围的字符可以
`[^...]`	匹配任何除方括号中以外的字符
`\w`	等效于 `[a-zA-Z_0-9]`
`\W`	等效于 `[^a-zA-Z_0-9]`
`\s`	空白字符，等效于 `[ \f\n\r\t\v]`（注意第一个字符是空格，剩下的符号是不
`\S`	非空白字符，等效于 `[^ \f\n\r\t\v]`
`\d`	一个数字（digit），等效于 `[0-9]`
`\D`	一个非数字，等效于 `[^0-9]`

- `\a` Alarm (beep)
- `\b` Backspace
- `\f` Form feed
- `\n` New line
- `\r` Carriage return
- `\t` Horizontal tab
- `\v` Vertical tab

- \特殊字符 例如 \? 表示问号，* 表示星号，\\ 表示反斜杠
- \oN 或 \o{N}，用 8 进制指定字符
- \xN 或 \x{N}，用 16 进制指定字符

重复匹配

在表 1 的命令后面可以加上如下 quantifier

- ? 表示 0 或 1 次重复
- + 表示 1 或者若干次重复
- * 表示 0 或者若干次重复
- {n} 表示 n 次重复
- {m,n} 表示 m 到 n 次重复
- {m,} 表示 m 次或以上重复
- 以上 quantifier 后面加 ? 可以匹配尽量少，例如 abc.+?def

Group

- (?:expr) 可以把表达式 group 到一起，例如 (?:abc){2} 搜索"abcabc"
- (expr) 可以把表达式 group 到一起并 match，例如 abc(\d+)def 寻找 abc\d+def 并且 match \d+ 部分
- (expr1|expr2) 匹配其中一个

在 VScode 中，用小括号 group 起来的部分，在替换的时候可以用 $1，$2 等来表示.

LaTeX 结构简介

本文对 LaTeX 文档的结构做一个简要的介绍．LaTeX 是一种**所见非所得**的排版语言，区别于 Word 文档这种**所见即所得**的排版，即用户编辑的是代码，需要经过编译过程才能获得最终的显示效果．小时百科的 PDF 编译使用 TeXlive 套装中的 XeLaTeX 编译器（下载）——简单来说就是支持中文和其他语言的 LaTeX，而在线编辑器则是我们自己开发的．

一个简单完整的 LaTeX 文档如下[12]

代码 1: test.tex

```
\documentclass{article}
\usepackage{amsmath}

\begin{document}
\title{My Title}
\author{My Name}
\maketitle

\section{Introduction}
Some introduction.

\begin{equation}
a^2 + b^2 = c^2 % is this correct?
\end{equation}

\subsection{Subtitle}
Subsection text.

\end{document}
```

编译后效果如图 1．下面我们来解释本例中的代码．

[12]LaTeX 文档通常使用 UTF-8 编码的文本文档．

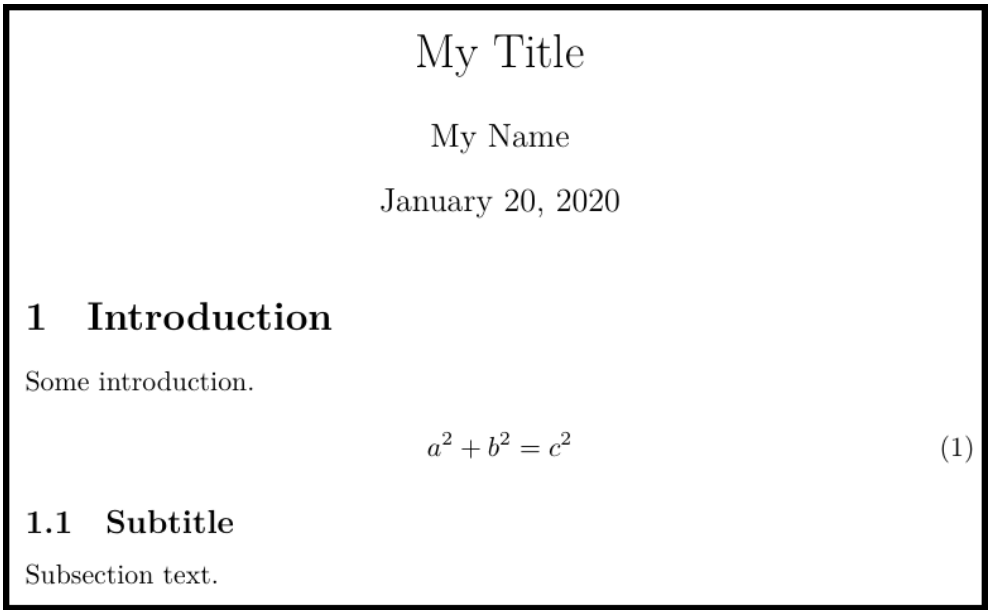


图 1: 排版效果

环境

一个完整的 LaTeX 文档是由许多**环境**构成的，环境的格式如下

```
1 \begin{环境名}[可选设置]
2 ...
3 ...
4 \end{环境名}
```

其中 [可选设置] 通常不需要出现．在一个完整的 LaTeX 文档中，最大的环境是 document 环境（test.tex 第 4 行），文档的所有内容（包括其他环境）都在 document 环境中．在 document 环境之前通常会有一些设置，例如规定文档的类别（第 1 行中的 article），使用一些语言拓展包（即**宏包**，如第 2 行的 amsmath）等．这些设置过于复杂，这里不进行介绍．一般建议直接直接使用现成的模板．

在 document 环境中，我们可以用 \section 和 \subsection 等命令把文章划分成不同的章节和子章节．编译器还可以根据这些命令自动生成目录．在小时百科中，我们使用四级标题，分别是**部分**（\part），**词条**（\section），**节**（\subsection），**子节**（\subsection）．注意这些命令只是用于插入对应的标题，并不构成环境．

另一些常用的环境包括**公式环境**（第 12 行的 equation）和**图片环境**（figure，未给出）.

命令

LaTeX 中命令的格式为反斜杠 \ 加连续的大小写字母.上面的 \maketitle，\begin 和 \section 等都是命令．有的命令可以单独存在，而另一些命令需要一个或多个输入，用花括号表示（如 \begin{...} 和 \section{...}）.

注释

LaTeX 源码中可以用 % 注释内容 的格式注释（第 13 行）．注释是给作者看的笔记或批注，在排版时会被忽略．如果一行中出现了 %，那么这行剩下的内容都会被视为注释．如果需要插入百分号而不是注释，使用 \% 即可.

多文件编译

当文档内容较多时（如一本书），我们可以将主文件中的内容（比如每节）剪切到一个个独立的文件中，再通过 \input{文件名} 命令将这些文档的内容插入主文档中．例如上面的文档可以划分为两个文件，主文件为

代码 2: test.tex

```
1  \documentclass{article}
2  \usepackage{amsmath}
3
4  \begin{document}
5  \title{My Title}
6  \author{My Name}
7  \maketitle
8
9  \input{section1.tex}
10
11 \end{document}
```

子文件为

代码 3: section1.tex

```
\section{Introduction}
Some introduction.

\begin{equation}
a^2 + b^2 = c^2 % is this correct?
\end{equation}

\subsection{Subtitle}
Subsection text.
```

校验和

- 术语可能不严谨
- 需要预备词条：计算机文件：文件名，内容，metadata，文件系统

[13]**校验和（checksum）**是指一类算法，可以将任意字节的数据（以下统称为文件）转换为一个简短的值．我们通常也将其称为文件的**指纹（fingerprint）**或者**哈希值（hash）**[14]．这可以用于检查两个文件的内容是否相同．

通常来说，文件在储存或传输的过程中有一定可能会被损坏，例如网络故障，黑客攻击，或者储存介质的个别字节出现故障．如果我们知道文件原来的校验和，就可以随时重新做一次校验和，如果结果和以前不一样，就说明文件被改变了．这也是为什么一些网站的提供文件下载时同时也会提供各种校验和，以便用户下载以后对照校验和确保文件无误．

碰撞

然而要达到上述目的，理想的情况就是确保任何文件和它的指纹都是一一对应[98] 的．但实际上，我们只能确保内容相同的文件得到同样的指纹，或者说不同的指纹必定对应内容不同的文件．不同的文件有可能对应相同的指纹，我们把这种情况叫做**碰撞（collision/clash）**．

[13]本文参考 Wikipedia 相关页面．

[14]严格来说这些不是同义词，但人们经常混用

哈希碰撞的概率通常很小，但却无法避免．例如某文件有 100 个比特，那么就有可能有 2^{100} 种不同的内容，但如果哈希值只有 10 个比特，那么只可能有 2^{10} 个不同的哈希值，所以必定会出现碰撞．为了减小碰撞的可能性，我们可以使用更好的算法，或者用更长的指纹．

为什么不同的算法会影响碰撞的可能性呢？我们可以考虑一个最简单的算法：把文件中每个字节对应的整数相加得到指纹．这显然是一个不好的算法，因为实际操作中很可能会出现两个文件仅仅存在字节顺序上的不同，例如 `abcd` 和 `dcba` 按照这个算法得到的结果都是一样的．所以好的算法在于能尽量减小实际使用时发生碰撞的概率．

安全性

除了碰撞，一个校验和算法的安全性也十分重要．安全性是指能否创造或修改某个文件，使得它具有指定的指纹．这种操作越难实现，算法就越安全．

例如一个黑客在某文件被传输时将其篡改，为了防止被发现，将其篡改的文件略作不重要的修改使其具有和源文件相同的指纹，那么即使文件的接收者对比指纹，也不会有所察觉，这就使得这个算法不安全．

用于服务器保存密码

网站的服务器中经常使用某种安全的指纹算法保存密码．这是因为如果一个网站的服务器中直接保存用户的密码明文，那么一旦这些明文密码被泄露，那么得到密码的人将可以任意访问用户的数据．

更好的做法是服务器仅仅保存每个用户密码的哈希值，用户每次输入密码时，服务器先将密码转换为哈希值再验证是否正确．这样即使这些哈希值泄露，也不能把他们作为密码直接使用，或者逆向得到密码原文．

常用算法

SHA1（Secure Hash Algorithm 1）著名文件版本控制软件 Git 就是使用 SHA1 散列值来检查文件是否发生变化．这是一个广泛使用的算法，但 SHA1 已被破解，所以安全性无保障．

MD5 是一个更安全的算法．

常用软件

大部分时候我们在命令行中操作，如 linux 的 `sha1sum`, `md5sum` 等命令，但也有一些软件提 GUI 界面，可以通过鼠标右键菜单的方式获得校验和，如压缩软件 7zip.

GitHub Desktop 的简单使用

- 本词条存在未完成的内容.

Git 与 GitHub

要了解 GitHub 首先要了解 Git. Git 是一个文件版本控制程序，通常用于管理程序代码，但也可以用来管理任何文件. 从某种意义上来讲，Git 相当于一个强大的备份软件，它可以保存一个文件夹内所有子文件夹和文件（这个文件夹叫做 **仓库（repository）**，简称 **repo**）的许多不同时间的 **snapshot**（快照，即所有文件某时刻的内容），而且可以对比出每个快 snapshot 相对上一个的变化. 但查看变化仅限文本文档，但 word 和 pdf 可以通过插件先自动转换为文本文档再做对比. Git 还可以给文件夹创造不同的 **branch （分支）**，例如一个程序写到一定的阶段后，我们希望将其向两个不同的方向发展为两个功能略不相同的程序，就可以创建一个新的分支，分别记录这两个程序的发展.

Git 本身是一个基于控制行的程序，最初是为了管理 Linux 操作系统的源代码而创造的，虽然 Git 现在已经有了各种各样的图形界面程序，但控制行版本的功能仍然是最齐全的. 所以一般的 Git 教程都是控制行版本的教程. 这里不介绍控制行[15]，而是介绍一个简单但广泛使用的界面程序 GitHub Desktop，注意该程序只能在 Windows 或 Mac 系统中安装.

Git 的一个重要功能就是可以让许多人合作完成一个工程，这个功能通常需要在一台服务器中安装 Git Server，每个合作者加入时把服务器中的 repo **clone**（克隆）到自己的电脑上. 更常见地是使用一些提供 server 的网站，如

[15] 可以参考官方教程.

github.com．通过 Git 或者 GitHub Desktop 也可以把 github.com 上的开源项目 clone 到本地．

Git 中的每一个快照也叫 **commit**，都是通过手动**提交（commit）**操作完成．我们可以随时查看当前文件夹中的文件与上次 commit 相比是否发生了改动，以及改动了什么内容，也可以选择撤销哪些改动．

创建第一个仓库

在官网下载 GitHub Desktop 并安装打开，可以登录或注册 GitHub 账号．注意 GitHub Enterprise 是企业账号，个人账号是免费注册的．注意免费账号只能建立 public repo（任何人都能访问或下载），创建私密的的 private repo 需要开通会员．目前使用 edu 结尾的学生邮箱注册可以获得免费会员．

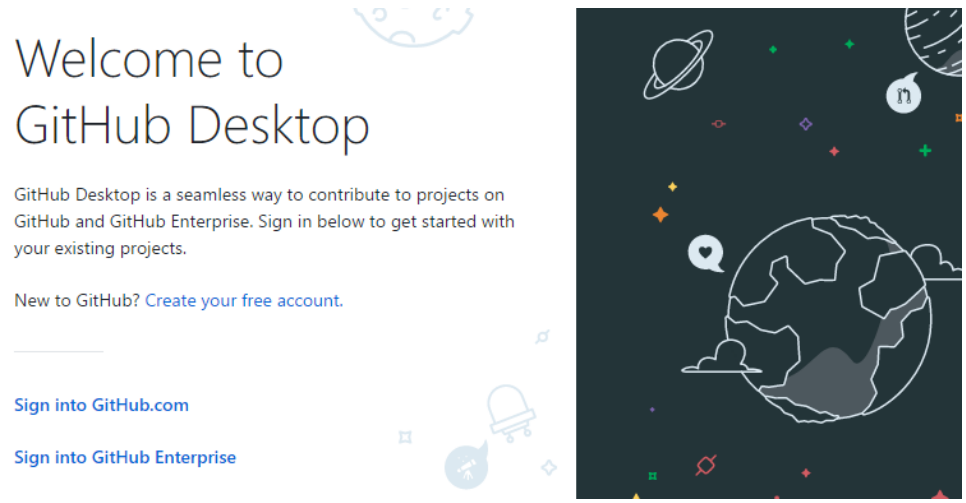


图 1: GitHub Desktop 登录

登录以后可以选择新建一个 repo 或者从 github 上 clone，我们这里选择新建（如图 2）.根据提示输入 repo 的名称 `test`，描述 `learn to use GitHub Desktop`，文件夹储存目录（默认是 `Documents/GitHub/`）．勾选 Initialize this repository with a README．然后就可以确定了．

创建完后界面如图 3 所示

单击 History 可以看到左下角已经有一个 Initial commit 了，所有的 commit 的历史可以在左上角的 History 找到．点击 History 可以看到 Initial commit 中添加了两个文件，`README.md` 和 `.gitattributes`．其中 `README.md` 是一个文本

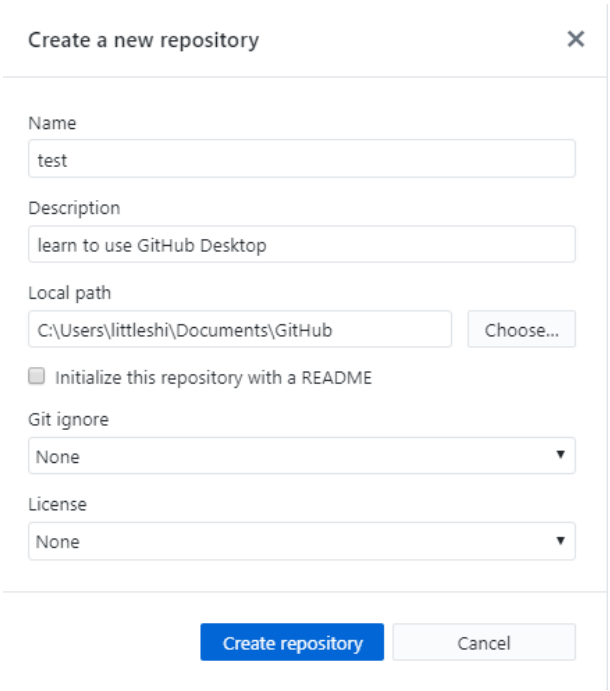


图 2: 新建仓库

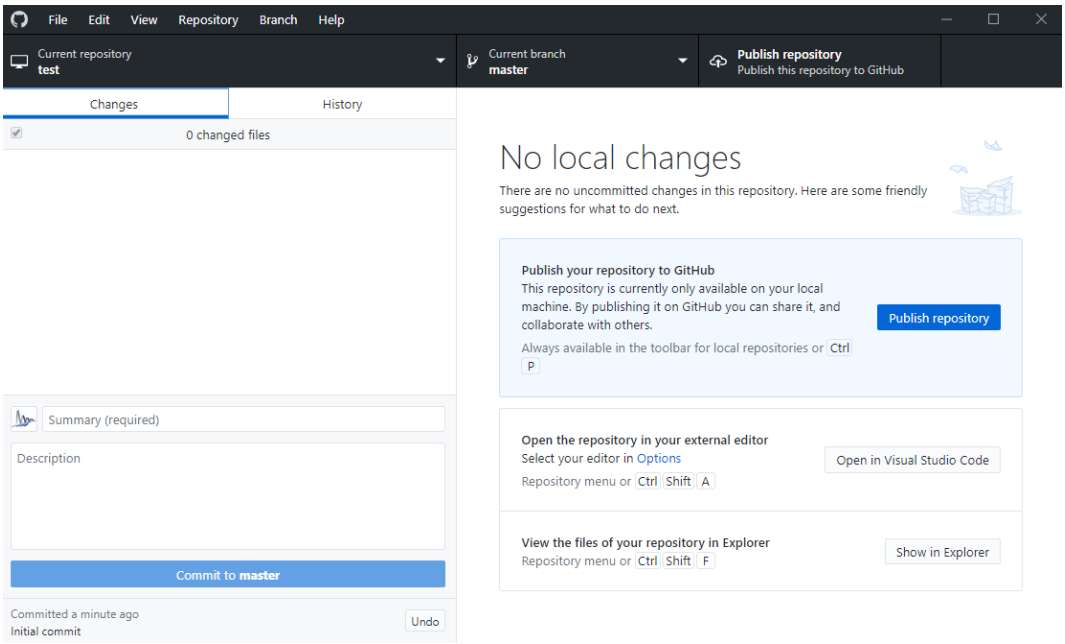


图 3: repo 主界面

文档，可以添加 repo 的描述，使用说明等．`.gitattributes` 文件用于对 Git 做

一些设置，例如如何处理换行[16]．如果新建 repo 的时候没有勾选 Initialize this repository with a README，就不会有 Initial commit．点击左上角的 Changes 返回刚才的界面．

注意左下角有一个 Undo 按钮．如果按下，Initial commit 将被撤销．再来看右上角的 Push 按钮，这个按钮用于把新增的 commit 上传（即 push）到 GitHub．由于我们从来没有 push 过，现在打开 GitHub 网站是看不到刚刚创建的项目的．按下 push，完成后再刷新浏览器就可以在 GitHub 看到了．

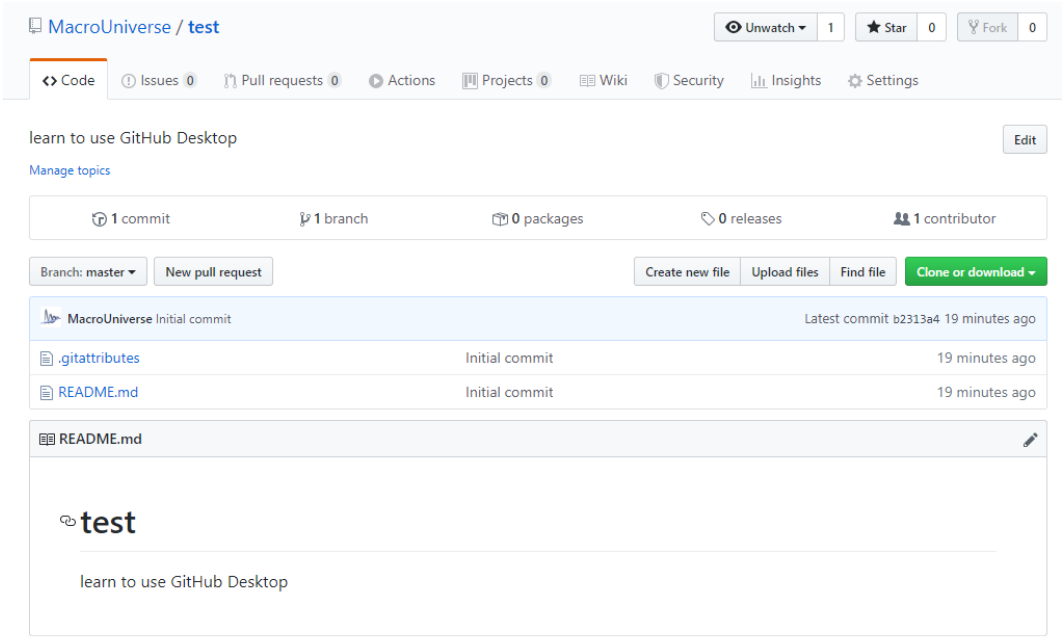


图 4: GitHub 网站上查看 repo

Commit

未完成：基本操作和概念

cuBLAS 库

[16]Windows 系统的换行使用 CR 和 LF 两个字符，而 Linux 系统中只用一个 LF 字符．

- 本词条缺少预备知识，初学者可能会遇到困难.

本词条将大致介绍如何使用 CUBLAS 库，同时演示一个使用 CUBLAS 库进行矩阵乘法的例子.

CUDA 的安装

请参阅 nvidia 官方文档或你的发行版文档.

如：ArchLinux 通过

```
pacman -S cuda
```

一条命令安装.

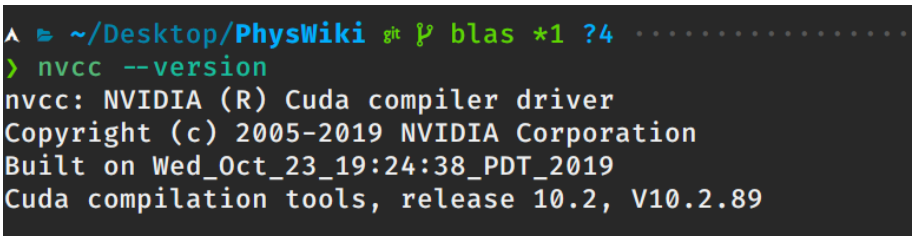


图 1: 测试 nvcc

CUBLAS 内容

CUBLAS 是 CUDA 专门用来解决线性代数运算的库，它分为三个级别（见 blas）

同时该库还包含状态结构和一些功能函数.

CUBLAS 用法

大体分成以下几个步骤：

1. 定义 CUBLAS 库对象
2. 在显存中为待运算的数据以及需要存放结果的变量开辟显存空间（`cudaMalloc` 函数实现）

3. 将待运算的数据传输进显存（cudaMemcpy，cublasSetVector 等函数实现）

4. 调用 CUBLAS 库函数（根据 CUBLAS 手册调用需要的函数）

5. 从显存中获取结果变量（cudaMemcpy，cublasGetVector 等函数实现）

6. 释放申请的显存空间以及 CUBLAS 库对象（cudaFree 及 cublasDestroy 函数实现）

代码示例

使用 CUBLAS 库进行矩阵乘法运算

如果你的文本编辑器是 vim, 推荐使用 ale, ycm, Asynctask 来做 cuda 的语法检查，语义补全，项目管理工具.

代码 1: cublas_demo.cu

```
1  // CUDA runtime 库 + CUBLAS 库
2  #include "cuda_runtime.h"
3  #include "cublas_v2.h"
4
5  #include <time.h>
6  #include <iostream>
7
8  using namespace std;
9
10 // 定义测试矩阵的维度
11 int const M = 5;
12 int const N = 10;
13
14 int main(void)
15 {
16     // 定义状态变量
17     cublasStatus_t status;
18
19     // 在 内存 中为将要计算的矩阵开辟空间
20     float *h_A = (float*)malloc (N*M*sizeof(float));
```

```
21      float *h_B = (float*)malloc (N*M*sizeof(float));
22
23      // 在 内存 中为将要存放运算结果的矩阵开辟空间
24      float *h_C = (float*)malloc (M*M*sizeof(float));
25
26      // 为待运算矩阵的元素赋予 0-10 范围内的随机数
27      for (int i=0; i<N*M; i++) {
28          h_A[i] = (float)(rand()%10+1);
29          h_B[i] = (float)(rand()%10+1);
30
31      }
32
33      // 打印待测试的矩阵
34      cout << "矩阵 A :" << endl;
35      for (int i=0; i<N*M; i++){
36          cout << h_A[i] << " ";
37          if ((i+1)%N == 0) cout << endl;
38      }
39      cout << endl;
40      cout << "矩阵 B :" << endl;
41      for (int i=0; i<N*M; i++){
42          cout << h_B[i] << " ";
43          if ((i+1)%M == 0) cout << endl;
44      }
45      cout << endl;
46
47      /*
48      ** GPU 计算矩阵相乘
49      */
50
51      // 创建并初始化 CUBLAS 库对象
52      cublasHandle_t handle;
53      status = cublasCreate(&handle);
54
55      if (status != CUBLAS_STATUS_SUCCESS)
56      {
```

```
57          if (status == CUBLAS_STATUS_NOT_INITIALIZED) {
58              cout << "CUBLAS 对象实例化出错" << endl;
59          }
60          getchar ();
61          return EXIT_FAILURE;
62      }
63
64      float *d_A, *d_B, *d_C;
65      // 在 显存 中为将要计算的矩阵开辟空间
66      cudaMalloc (
67          (void**)&d_A,    // 指向开辟的空间的指针
68          N*M * sizeof(float)    //  需要开辟空间的字节数
69      );
70      cudaMalloc (
71          (void**)&d_B,
72          N*M * sizeof(float)
73      );
74
75      // 在 显存 中为将要存放运算结果的矩阵开辟空间
76      cudaMalloc (
77          (void**)&d_C,
78          M*M * sizeof(float)
79      );
80
81      // 将矩阵数据传递进 显存 中已经开辟好了的空间
82      cublasSetVector (
83          N*M,    // 要存入显存的元素个数
84          sizeof(float),    // 每个元素大小
85          h_A,    // 主机端起始地址
86          1,    // 连续元素之间的存储间隔
87          d_A,    // GPU 端起始地址
88          1    // 连续元素之间的存储间隔
89      );
90      cublasSetVector (N*M, sizeof(float), h_B, 1, d_B, 1);
91
92      // 同步函数
```

```
93      cudaThreadSynchronize();
94
95      // 传递进矩阵相乘函数中的参数，具体含义请参考函数手册.
96      float a=1; float b=0;
97      // 矩阵相乘. 该函数必然将数组解析成列优先数组
98      cublasSgemm (
99          handle,    // blas 库对象
100         CUBLAS_OP_T,    // 矩阵 A 属性参数
101         CUBLAS_OP_T,    // 矩阵 B 属性参数
102         M,    // A, C 的行数
103         M,    // B, C 的列数
104         N,    // A 的列数和 B 的行数
105         &a,    // 运算式的 \alpha 值
106         d_A,    // A 在显存中的地址
107         N,    // lda
108         d_B,    // B 在显存中的地址
109         M,    // ldb
110         &b,    // 运算式的 \beta 值
111         d_C,    // C 在显存中的地址(结果矩阵)
112         M    // ldc
113     );
114
115     // 同步函数
116     cudaThreadSynchronize();
117
118     // 从 显存 中取出运算结果至 内存中去
119     cublasGetVector (
120         M*M,    //  要取出元素的个数
121         sizeof(float),    // 每个元素大小
122         d_C,    // GPU 端起始地址
123         1,    // 连续元素之间的存储间隔
124         h_C,    // 主机端起始地址
125         1    // 连续元素之间的存储间隔
126     );
127
128     // 打印运算结果
```

```
129     cout << "计算结果的转置 ( (A*B)的转置 ): " << endl;
130
131     for (int i=0;i<M*M; i++){
132             cout << h_C[i] << " ";
133             if ((i+1)%M == 0) cout << endl;
134     }
135
136     // 清理掉使用过的内存
137     free (h_A); free (h_B); free (h_C); cudaFree (d_A);
138     cudaFree (d_B); cudaFree (d_C);
139
140     // 释放 CUBLAS 库对象
141     cublasDestroy (handle);
142     return 0;
143 }
```



图 2: 编译

cublas 文档

cuda 的文档十分易懂，内有丰富的例子.

我将 cuda 安装在 /opt 下,那么 cublas 库的文档就在 /opt/cuda/doc/cublas 下.

Julia 分形

```
~/Desktop/PhysWiki  master ↓33 +4 !4 ?4
❯ ./a.out
矩阵 A :
4 8 4 7 10 3 1 4 1 3
2 8 3 3 8 10 3 10 4 2
10 2 5 9 6 4 2 7 3 7
6 5 7 7 4 5 3 5 5 4
8 7 9 4 5 3 7 10 7 5

矩阵 B :
7 6 6 3 2
8 10 7 7 7
9 10 1 4 6
3 9 8 7 3
4 10 8 5 1
7 1 4 1 2
6 8 6 10 8
6 6 8 5 1
9 9 5 2 10
1 9 3 7 10

计算结果的转置 ( (A*B)的转置 ):
252 332 298 314 416
378 377 423 399 521
284 323 319 281 362
239 247 271 248 334
177 209 228 243 328
```

图 3: 运行

- 本词条处于草稿阶段.
- 本词条缺少预备知识，初学者可能会遇到困难.

[17] 这里给出一个 Julia 分形的演示程序，运行结果如图 1.

未完成：说明未完成

```
function julia_set
close all;
N = 1000;
scale = 1.5;
x = linspace(-1, 1, N); y = x;
[X, Y] = ndgrid(x, y);
```

[17]参考 Wikipedia 相关页面

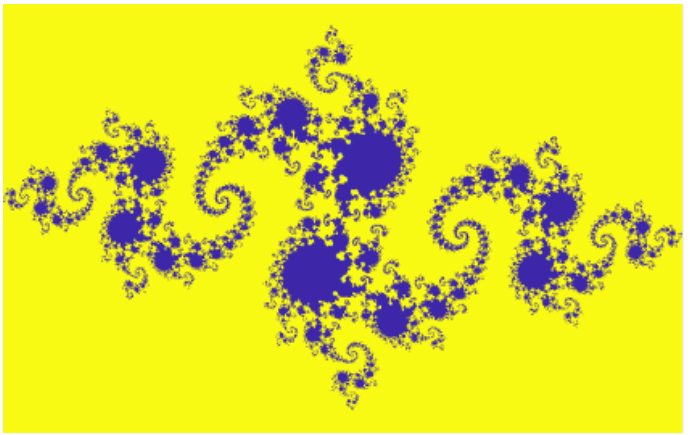


图 1: 运行结果

```
I = julia2(X*scale, Y*scale);
I(I > 1000) = 1000;
surfCart(x, x, I);
caxis([0, 500]);
end

function I = julia2(X, Y)
    c = -0.8 + 0.15745i;
    Z = X + 1i*Y;
    for i=1:80
        Z = Z .* Z + c;
    end
    I = abs(Z); % I = I/max(I(:));
end
```

第六章

数值验证及常用算法

二项式定理（非整数幂）的数值验证

预备知识 二项式定理（非整数）[108]，Matlab 编程基础[1682]

这里介绍一个简单的 Matlab 程序用于计算表 1 [108]．第 7-10 行的循环中，每个循环通过 `x^(ii-1)` 项的系数计算 `x^ii` 项的系数，并将 `x^ii` 项累加到求和结果 `result` 上．

代码 1: binorR.m

```
% 验证二项式定理(非整数幂)
u = -3.5; % 幂
x = 0.6; % |x|<1 使级数收敛
N = 100; % 求和项数
Coeff = 1; % x^ii 项前面的系数
result = 1; % 求和结果
for ii = 1:N
    % 由 x^(ii-1) 项系数计算 x^ii 项系数
    Coeff = Coeff*(u-ii+1) / ii;
    % 将 x^(ii-1) 项累加到求和结果上
    result = result + Coeff * x^ii;
end
format long % 显示全部小数位
disp('求和结果为')
disp(result)
disp('精确结果为')
disp((1+x)^u)
format short % 恢复默认显示
```

二分法

二分法是数值计算中一种求连续一元函数零点的简单方法．先来看一个显然的结论：如果我们知道一个连续函数在某个开区间左端的值和右端的值（假

设都不为 0）乘积小于 0（即正负号不同），那么这个函数在该区间内必有至少一个根．为了进一步缩小这个根所在的区间，我们在区间中点求函数值，如果中点处的函数值与区间左端的函数值同号（相乘大于 0），则函数的根必然在区间中点和区间右端之间，于是我们可以把区间左端移动到区间中点处，再来求新区间的中点．如果区间中点的函数值与区间右端的函数值同号，同理我们也可以把区间右端移动到区间中点处得到新的区间．如果区间中点的函数值恰好为 0，我们便找到了一个根，另一方面，如果区间的长度小于我们对根的精度的要求，那么我们就找到了一个根的近似值．

Matlab 中自带的 fzero 函数如果按照以下格式使用，大致可以认为是二分法

```
>> f = @(x)x-1;
>>fzero(f, [0,2])
ans = 1
```

fzero 的默认精度是 2.2e-16．注意要把一个函数作为其他函数的输入变量，必须使用函数句柄．下面我们给出一个二分法的 Matlab 程序．

代码 1: bisection.m

```
% 二分法求函数的根
function x = bisection(f, int, err)
fl = f(int(1)); fr = f(int(2));
% 两端点是否为 0
if fl == 0
    x = int(1); return;
elseif fr == 0
    x = int(2); return;
end
% 两端点是否同号
if fl * fr > 0
    error('两端点同号');
end
% 主循环
while(int(2) - int(1) > err)
    mid = 0.5*(int(1) + int(2));
    fm = f(mid);
```

```
18      if fm * fl > 0
19          int(1) = mid;
20      elseif fm * fr > 0
21          int(2) = mid;
22      else % fm = 0
23          break;
24      end
25  end
26  x = mid;
27  end
```

我们先来看函数的自变量，与 fzero 类似，该函数的前两个输入变量分别是函数句柄和求根区间，第三个变量是误差值，当区间的长度小于 err 时，函数将输出区间中点作为输出．函数的第 3-13 行做了一些必要的检查，确保区间两端的函数值为异号．第 15 行开始主循环，每循环一次，函数的区间长度减半，直到区间中点处的函数值为 0 或区间长度小于 err 时跳出循环，最后返回区间中点的函数值．

多区间二分法

预备知识 二分法[1841]

这里介绍一种多区间二分法，可以求出连续函数在某区间内几乎全部的根．方法就是把这个区间等分为若干个相等的小区间，然后分别判断这些小区间两端函数值的符号，对所有两端异号的区间使用二分法即可．显然，小区间的个数越多，越有可能找到所有的根．例程如下．

代码 1: bisectionN.m

```
1  function roots = bisectionN(f, int, N)
2  x = linspace(int(1), int(2), N); % 划分区间
3  y = arrayfun(f, x); % 求所有 f(x(ii))
4  figure; plot(x, y, '.-') % 画图
5  title('f(x)')
6  Sign = sign(y);
```

```
7  ind = find(Sign(1:end-1) .* Sign(2:end) <= 0); % 找符合条件的区
       间序号
8  Nroot = numel(ind);
9  roots = zeros(1, Nroot); % 预赋值
10 for ii = 1:Nroot
11     roots(ii) = fzero(f, [x(ind(ii)),x(ind(ii)+1)]);
12 end
13 end
```

函数的前两个输入变量分别是需要求根的函数句柄和求根区间（二元行矢量或列矢量），第三个变量 N 是子区间端点的个数（即子区间的个数加一）. 函数中先求出所有的端点 `x`，以及对应的函数值 `y`，然后画图. 第 6-7 行寻找所有两端异号或有一端为 0 的区间的序号，然后在第 10 行的循环中对这些区间逐个使用二分法. 为了提高运算效率，这里并没有使用“二分法[1841]”中的例程，而是使用了 Matlab 自带的 `fzero` 函数.

`bisectionN` 的画图功能是为了让用户判断是否有可能出现漏根，以下举两个例子说明.

```
>> f = @(x)exp(-0.2*x)*sin(x);
>> roots = bisectionN(f, [0, 15], 50)
roots = 0    3.1416    6.2832    9.4248    12.5664
```

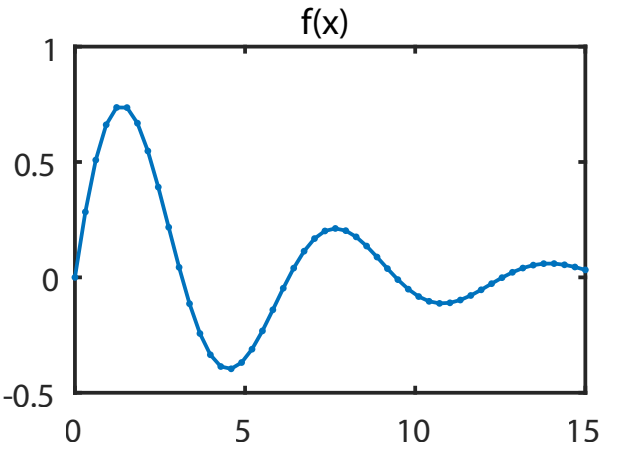


图 1: 运行结果

运行结果如图 1，由于画出的曲线较为光滑，可判断漏根的可能性很小. 再看另一个例子

```
>> f = @(x)sin(1/x);
>> roots = bisectionN(f, [0, 0.3], 50)
roots = 0.0245   0.0398   0.0455   0.0531   0.0637   0.0796   0.1061
   0.1592
```

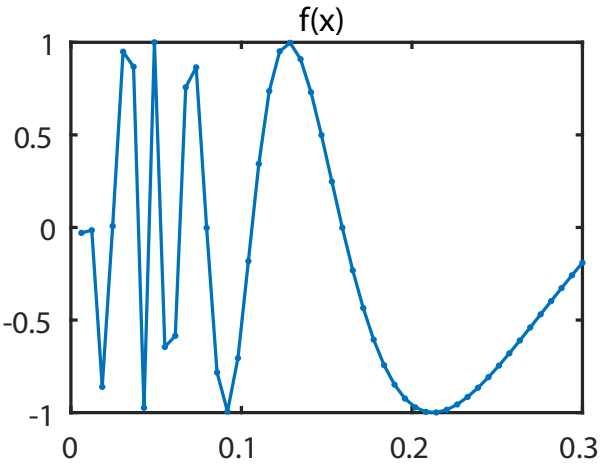


图 2: 运行结果

我们已经知道函数 $\sin(1/x)$ 在该区间上有无数个根，且越接近 $x=0$，相邻根之间的距离越小．运行结果如图 2，可见在区间 $[0, 0.1]$ 内，子区间端点的函数值非常不平滑，极有可能出现漏根．为了求得更多的根，我们可以增加子区间的个数．

冒泡法

预备知识 Matlab 的函数[1700]

我们先来看 Matlab 自带的排序函数 `sort`．假设数列 `age` 是几个小朋友的年龄，`name` 是这几个小朋友对应的名字，现在按年龄从小到大排序如下

```
>> age = [1, 6, 2, 5, 3];
>> name = ['a', 'b', 'c', 'd', 'e'];
>> [age1, order] = sort(x);
age1 = 1 2 3 5 6
order = 1 3 5 4 2
```

```
>> name1 = name(order);
name1 = 'acedb'
```

其中输出变量 order 是排序后每个数字在原来数列中的位置索引，即 name1 等于 name(order). 现在我们用冒泡法实现 sort 的功能. sort 函数默认把数列升序排列，即第二个输入变量默认为 'ascend'. 若要降序排列，可以用 'descend' 作为第二个输入变量.

冒泡法是一种简单的排序算法，但由于效率较低，一般仅作为教学演示用. 在实际使用时还是建议用 sort. 冒泡法的算法为：以升序排列为例，给出一个数列，先把第一个数与第二个进行比较，若第一个数较大，就置换二者的位置（具体操作是，把第一个数的值赋给一个临时变量，再把第二个数的值赋给第一个，最后把临时变量的值赋给第二个），再把第二个数与第三个进行比较，若第二个较大，就置换二者的位置，这样一直进行到最后两个数，完成第一轮. 然后再进行第二轮，第三轮，直到某一轮没有出现置换操作，即可确定排序完成. 至于输出变量 order，我们可以先令 order = 1:N，每置换数列的两个数，就把 order 中对应的两个数也置换即可. 这样，数列与其原来的索引将始终一一对应.

代码 1: bubble.m

```
1  % 冒泡法排序
2  function [x, order] = bubble(x, option)
3  N = numel(x); % 数列个数
4  order = 1:N; % 索引
5  changed = 1; % 是否有置换
6  while(changed == 1)
7      changed = 0;
8      for ii = 1:N-1
9          if x(ii) > x(ii + 1)
10             % 置换
11             changed = 1;
12             temp = x(ii);
13             x(ii) = x(ii + 1);
14             x(ii + 1) = temp;
15             temp = order(ii);
16             order(ii) = order(ii + 1);
```

```
17              order(ii + 1) = temp;
18          end
19      end
20  end
21  % 是否是降序排列
22  if nargin > 1 && option(1) == 'd'
23      x(:) = flipud(x(:));
24      order = fliplr(order);
25  end
26  end
```

第 6 行的循环每循环一次，数列将从头到尾被扫描一遍．每个循环开始时 changed 的值被设为 0，若有任何置换，changed 则变为 1（第 11 行），使 while 的判断条件成立，循环继续．为了使第一个循环发生，在循环前必须把 changed 设为 1．再来看第 9-18 行的判断结构．如果前一个数大于后一个数，则置换发生．注意要置换数列中的两个数，必须要设一个临时变量（temp）．函数的最后，判断输入变量的个数，如果只有一个变量，则默认按照前面的代码升序排列，若第二个变量为 'descend'，则把 x 和 order 翻转一下即可．

高斯消元法程序

预备知识 线性方程组高斯消元法[477]

这里给出一个简单的程序演示高斯消元法的基本步骤．注意在实际应用中，我们解线性方程组一般使用 Matlab 的反斜杠算符：x=A \ y，其中 **A** 是系数矩阵，**y** 是常数列，**x** 是未知量．

代码 1: GaussEli.m

```
1  % 高斯消元法得到梯形系数矩阵
2  % 显示每个行变换的操作和结果
3  % A 为任意尺寸的矩阵
4  % 如果 A 是增广矩阵， 输入 augmented = true， 如果 A 是系数矩
       阵， augmented = false
5  % 输出的 A 是梯形矩阵， q 是一个行矢量， q(ii) 是第 ii 行第一个
       不为零的列标
```

```
% 如果第 ii 行全为零，则 q(ii) = 0

function [A, q] = GaussEli(A, augmented)
% 系数矩阵的行数 Nr 和列数 Nc
[Nr, Nc] = size(A);
if (augmented)
    Nc = Nc - 1;
end
q = zeros(1,Nr); q(1) = 1;
disp(A);
for ii = 1:Nr-1
    % 处理第 ii 行
    if (ii > 1)
        q(ii) = q(ii - 1) + 1;
    end
    for jj = q(ii) : Nc
        % 检查第 jj 列
        if (max(abs(A(ii:end, q(ii)))) > 0)
            % 消元使 A(ii+1:end, q(ii)) 全为 0
            A = eli1(A, ii, q(ii));
            break;
        else
            % 不需要消元，检查下一列
            if (q(ii) < Nc)
                q(ii) = q(ii) + 1;
            else
                return;
            end
        end
    end
end
end

% 做行变换使 A(ii+1:end, q) 全为 0
% A(ii:end,q) 不能全为 0
function A = eli1(A, ii, q)
```

```
% 交换两行，使 A(ii,q) 在 A(ii:end, q) 中最大
[~, iimax] = max(abs(A(ii:end, q)));
iimax = ii + iimax - 1;
if (iimax > ii)
    disp(['交换: r_', num2str(ii), ' <-> r_', num2str(iimax)]);
    temp = A(ii,:);
    A(ii,:) = A(iimax,:);
    A(iimax,:) = temp;
    disp(A);
end

% 用第 ii 行消去所有 A(ii+1:end, q)
for jj = ii+1 : size(A,1)
    if (A(jj,q) == 0)
        continue;
    end
    k = -A(jj,q)/A(ii,q);
    disp(['消元: ', num2str(k), ' * r_', num2str(ii), ' + r_',
       num2str(jj)]);
    A(jj,:) = A(ii,:) * k + A(jj,:);
    disp(A);
end
end
```

与“高斯消元法[477]”中的步骤略有不同的是，该程序在处理每一行 `A(ii,:)` 时都会试图做一个行交换使系数 `A(ii,q(ii))` 的绝对值尽可能大．这是为了减小数值误差：试想如果 `A(ii,q(ii))` 的解析值为 0，但由于数值误差，计算出来是一个很小的数（例如 `1e-15`），那么用其消元时就可能需要将第 `ii` 行乘以一个很大的系数（例如 `1e15`）再加到另一行上，导致程序不稳定．

用例 2 [477] 中的增广矩阵测试程序如下：

```
>> A = [1 1 -1 1; 2 2 -2 1; 1 1 0 2; 2 2 -1 5];
>> y = [3; 7; 3; 4];
>> [A1, q] =GaussEli([A,y],true); q
     1     1    -1     1     3
     2     2    -2     1     7
```

```
     1      1      0      2      3
     2      2     -1      5      4

交换: r_1 <-> r_2
     2      2     -2      1      7
     1      1     -1      1      3
     1      1      0      2      3
     2      2     -1      5      4

消元: -0.5 * r_1 + r_2
    2.0000    2.0000   -2.0000    1.0000    7.0000
         0         0         0    0.5000   -0.5000
    1.0000    1.0000         0    2.0000    3.0000
    2.0000    2.0000   -1.0000    5.0000    4.0000

消元: -0.5 * r_1 + r_3
    2.0000    2.0000   -2.0000    1.0000    7.0000
         0         0         0    0.5000   -0.5000
         0         0    1.0000    1.5000   -0.5000
    2.0000    2.0000   -1.0000    5.0000    4.0000

消元: -1 * r_1 + r_4
    2.0000    2.0000   -2.0000    1.0000    7.0000
         0         0         0    0.5000   -0.5000
         0         0    1.0000    1.5000   -0.5000
         0         0    1.0000    4.0000   -3.0000

交换: r_2 <-> r_3
    2.0000    2.0000   -2.0000    1.0000    7.0000
         0         0    1.0000    1.5000   -0.5000
         0         0         0    0.5000   -0.5000
         0         0    1.0000    4.0000   -3.0000

消元: -1 * r_2 + r_4
    2.0000    2.0000   -2.0000    1.0000    7.0000
         0         0    1.0000    1.5000   -0.5000
```

```
         0         0         0    0.5000   -0.5000
         0         0         0    2.5000   -2.5000

交换: r_3 <-> r_4
    2.0000    2.0000   -2.0000    1.0000    7.0000
         0         0    1.0000    1.5000   -0.5000
         0         0         0    2.5000   -2.5000
         0         0         0    0.5000   -0.5000

消元: -0.2 * r_3 + r_4
    2.0000    2.0000   -2.0000    1.0000    7.0000
         0         0    1.0000    1.5000   -0.5000
         0         0         0    2.5000   -2.5000
         0         0         0         0         0

q =
     1     3     4     0
```

Nelder-Mead 算法

预备知识 Matlab 的函数[1700]

Nelder-Mead 算法是一种求多元函数局部最小值的算法，其优点是不需要函数可导并能较快收敛到局部最小值．Matlab 自带的 `fminsearch` 函数就是使用该算法．对 N 元函数 $f(\mathbf{x})$（这里把函数自变量用 N 维矢量来表示），该算法需要提供函数自变量空间中的一个初始点，$\mathbf{x}_1$，算法从该点出发寻找局部最小值，以下是具体步骤．

我们先根据初始点另外生成 N 个初始点 $\mathbf{x}_2 \dots \mathbf{x}_{N+1}$，使 $\mathbf{x}_{1+i}$ 在第 i 个分量比 $\mathbf{x}_1$ 的大 5%，其他分量保持相同．如果 $\mathbf{x}_1$ 的第 i 个分量为零，那么 $\mathbf{x}_{1+i}$ 的第 i 个分量设为 0.00025．得到 $N+1$ 个初始点后，开始按照以下步骤进行循环，直到满足特定的精度条件时退出循环．

1. 先给这些点按照 $f(\mathbf{x}_i)$ 从小到大的顺序重新排序并重命名，使 i 越大 $f(\mathbf{x}_i)$ 越大．

2. 计算前 N 个点的平均位置为

$$\mathbf{m}=\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \tag{1}$$

3. 计算 $\mathbf{x}_{N+1}$ 关于点 $\mathbf{m}$ 的**反射点**为

$$\mathbf{r}=2\mathbf{m}-\mathbf{x}_{N+1} \tag{2}$$

4. 如果 $f(\mathbf{x}_1)\leqslant f(\mathbf{r})<f(\mathbf{x}_N)$，令 $\mathbf{x}_{N+1}=\mathbf{r}$，并进入下一个循环.

5. 如果 $f(\mathbf{r})<f(\mathbf{x}_1)$，计算**拓展点**为

$$\mathbf{s}=\mathbf{m}+2(\mathbf{m}-\mathbf{x}_{N+1}) \tag{3}$$

如果 $f(\mathbf{s})<f(\mathbf{r})$，令 $\mathbf{x}_{N+1}=\mathbf{s}$ 并进入下一个循环，否则令 $\mathbf{x}_{N+1}=\mathbf{r}$. 并进入下一循环.

6. 如果 $f(\mathbf{x}_N)\leqslant f(\mathbf{r})<f(\mathbf{x}_{N+1})$，令

$$\mathbf{c}_1=\mathbf{m}+(\mathbf{r}-\mathbf{m})/2 \tag{4}$$

如果 $f(\mathbf{c}_1)<f(\mathbf{r})$，令 $\mathbf{x}_{N+1}=\mathbf{c}_1$ 并进入下一循环，否则执行最后一步.

7. 如果 $f(\mathbf{x}_{N+1})\leqslant f(\mathbf{r})$ 令

$$\mathbf{c}_2=\mathbf{m}+(\mathbf{x}_{N+1}-\mathbf{m})/2 \tag{5}$$

如果 $f(\mathbf{c}_2)<f(\mathbf{x}_{N+1})$，令 $\mathbf{x}_{N+1}=\mathbf{c}_2$ 并进入下一循环，否则执行最后一步.

8. 令

$$\mathbf{v}_i=\mathbf{x}_1+(\mathbf{x}_i-\mathbf{x}_1)/2\qquad(i=2\ldots N+1) \tag{6}$$

并用 $\mathbf{v}_i$ 赋值给 $\mathbf{x}_i$，进入下一循环.

观察以上步骤可知，当局部最小值的位置在 $N+1$ 个围成的图形以外时，图形倾向于变大且加速向最小值移动. 当最小值的位置在图形内部时，图形倾向于缩小. 随着循环次数增加，这 $N+1$ 个点最终将向局部最小值聚拢.

我们可以在每个循环的第一步之后计算 $\mathbf{x}_1$ 和 $\mathbf{x}_{N+1}$ 的距离来估算自变量的误差，如果该误差小于某个值，即可结束循环并使用 $\mathbf{x}_1$ 作为最终结果．作为另一种方法，我们也可以在每个循环的第一步之后计算 $f(\mathbf{x}_{N+1}) - f(\mathbf{x}_1)$ 来估算最小值的误差．

以下是该算法的 Matlab 代码．

代码 1: NelderMead.m

```
% f 是函数句柄，只接受一个 N 维行矢量作为输入变量，并返回一个函数值
% x0 是 N 维行矢量，xerr 是一个标量
function [xmin, fmin] = NelderMead(f, x0, xerr)
N = numel(x0); % f 是 N 元函数
x = zeros(N+1,N); % 预赋值
y = zeros(1,N+1);
% 计算 N+1 个初始点
x(1,:) = x0;
for ii = 1:N
    x(ii+1,:) = x(1,:);
    if x(1,ii) == 0
        x(ii+1,ii) = 0.00025;
    else
        x(ii+1,ii) = 1.05 * x(1,ii);
    end
end
% 主循环
for kk = 1:10000
    % 求值并排序
    for ii = 1 : N + 1
        y(ii) = f(x(ii,:));
    end
    [y, order] = sort(y);
    x = x(order,:);
    if norm(x(N+1,:) - x(1,:)) < xerr % 判断误差
        break;
    end
    m = mean(x(1:N,:)); % 平均位置
```

```
    r = 2*m - x(N+1,:); % 反射点
    f_r = f(r);
    if y(1) <= f_r && f_r < y(N) % 第 4 步
        x(N+1,:) = r; continue;
    elseif f_r < y(1) % 第 5 步
        s = m + 2*(m - x(N+1,:));
        if f(s) < f_r
            x(N+1,:) = s;
        else
            x(N+1,:) = r;
        end
        continue;
    elseif f_r < y(N+1) % 第 6 步
        c1 = m + (r - m)*0.5;
        if f(c1) < f_r
            x(N+1,:) = c1; continue;
        end
    else % 第 7 步
        c2 = m + (x(N+1,:) - m)*0.5;
        if f(c2) < y(N+1)
            x(N+1,:) = c2; continue;
        end
    end
    for jj = 2:N+1 % 第 8 步
        x(jj,:) = x(1,:) + (x(jj,:) - x(1,:))*0.5;
    end
end
% 输出变量
xmin = x(1,:);
fmin = f(xmin);
end
```

该程序中有几个需要注意的地方．首先，主循环并没有用 `while 1`，而是用 `for` 循环指定一个最大循环次数．这是为了避免少数情况下可能发生的死循环（例如 $f(\mathbf{x})$ 在某个区域中的值处处相等时）．第二，4-7 步中对 $f(\mathbf{r})$ 的判断有且仅有一个成立，所以我们可以用 `if...elseif...else` 结构来选择．最

后，4-5 步的情况下程序必定会执行 `continue` 语句而跳过第 8 步，只有 6-7 步中的 `if` 判断失败程序才会执行第 8 步.

最小二乘法的数值计算

预备知识 最小二乘法[267]，Nelder-Mead 算法[1851]

我们在“最小二乘法[267]”见到的三个例子中，方差函数都是待定系数的线性组合，这种情况下我们令偏导为零后得到的是线性方程组，便于求解. 然而当方差不是待定系数的线性组合时，得到的方程组往往非常复杂，这时就需要借助数值计算. 相比用数值计算解 N 元的非线性方程组，更简单的方法是直接用数值方法寻找方差函数的极小值（如 Nelder-Mead 算法）. 实践证明，大多数情况下极小值点仅有一个，那就是最小值点.

为了验证结果的正确性，我们先来用数值方法拟合 $A\cos(x+\varphi_0)+C$，并与“最小二乘法[267]”中的方法比较结果.

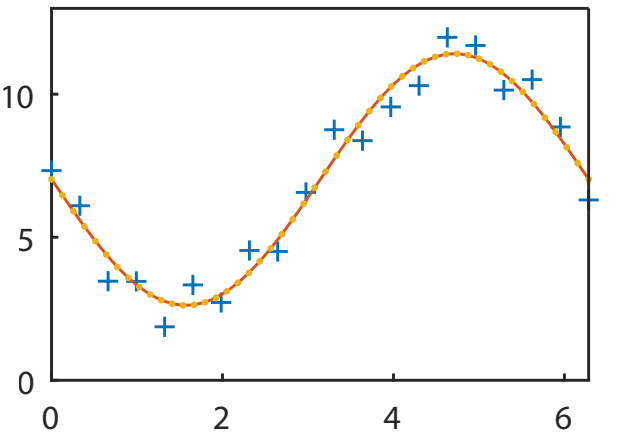


图 1: 运行结果

代码 1: curveFit.m

```
1 function curveFit
2 close all;
3 % 随机生成简谐曲线
4 N = 20;
```

```
x0 = linspace(0, 2*pi, N);
y0 = 5*rand * sin(x0 + 2*pi * rand) + 10 * rand;
y0 = y0 + 2*rand(1,20); % 产生随机误差
scatter(x0, y0, '+'); % 画出散点
hold on;

% Nelder-Mead 求方差最小值点
f = @(x) norm( x(1)*sin(x0 + x(2)) + x(3) - y0 )^2;
c = NelderMead(f, [5, 1, 5], 1e-7);
% 画拟合结果
x = linspace(0, 2*pi, 50);
y1 = c(1) * sin(x + c(2)) + c(3);
plot(x, y1);

% 解线性方程组求方差最小值点
c = sinfit(x0, y0);
% 画拟合结果
y2 = c(1)*cos(x) + c(2)*sin(x) + c(3);
plot(x, y2, '.');
end

% 拟合简谐曲线
% y = C(1)*cos(x) + C(2)*sin(x) + C(3)
function C = sinfit(x, y)
N = numel(x);
cosx = cos(x); sinx = sin(x);
sc = sum(sinx.*cosx);
s = sum(sinx); c = sum(cosx);
% 系数矩阵
M = [sum(cosx.^2), sc          , c;
                sc, sum(sinx.^2), s;
                 c,            s, N];
b = [sum(y.*cosx); sum(y.*sinx); sum(y)];
C = M\b; % 解线性方程组
end
```

运行结果如图 1 所示，可见两种方法拟合出的曲线一致（红色的曲线和黄色的点线）. 注意第 13 行使用了“Nelder-Mead 算法[1851]”中的函数 `NelderMead` 求函数句柄 `f` 的最小值.

数值积分（梯形法）

预备知识 Matlab 的函数[1700]，定积分[198]

本词条介绍一种简单的梯形算法计算数值积分. 如图 1 若要用梯形法计算定积分 $\int_a^b f(x)\,\mathrm{d}x$，则可将区间 $[a,b]$ 划分为 N 个长度为 $h=(b-a)/N$ 的等长的小区间，区间端点从 a 到 b 分别为 $x_1=a, x_2=a+h, \ldots, x_{N+1}=b$.

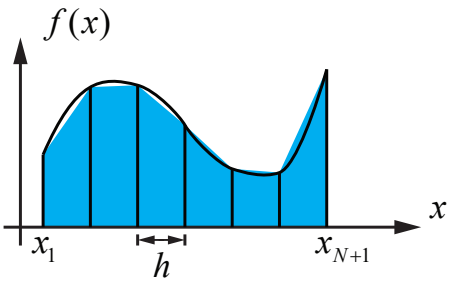


图 1: 梯形法数值积分

接下来将每个区间的被曲线围出的面积用梯形来计算，第 i 个梯形面积为（两底和乘以高除以二）$[f(x_{i+1})+f(x_i)]h/2$. 定积分约等于所有梯形面积

$$\int_a^b f(x)\,\mathrm{d}x \approx \sum_i^N \frac{h}{2}[f(x_{i+1})+f(x_i)] = h\left[\frac{1}{2}f(x_1)+\sum_{i=2}^N f(x_i)+\frac{1}{2}f(x_{N+1})\right] \quad (1)$$

显然，当 N 取越大时，右边的求和就越接近定积分.

这里给出 Matlab 代码

代码 1: trapezoidInt.m

```
1 % 梯形法数值积分
2 % f 是被积函数的函数句柄
3 % [a, b] 为积分区间
4 % N 为子区间个数
```

```
function I = trapezoidInt(f, a, b, N)
x = linspace(a, b, N+1);
y = arrayfun(f, x); % 求所有 y(i) = f(x(i))
h = (b - a)/N;
I = h*(0.5*y(1) + 0.5*y(N+1) + sum(y(2:N)));
end
```

下面来看两个例子，由算法可知，当被积函数具有 c_1x+c_2 的形式时，函数曲线围成的面积可由梯形精确解算，所以即使 `N` 取很小也可以精确到 `double` 类型的精度（16 位左右）.

```
>> format long;
>> trapezoidInt(@(x)x, 0, 1, 4)
ans = 0.500000000000000
```

对其他一些函数，需要较大的 N 才能获得较高的精度. 我们已知 $\sin x$ 在 $[0,\pi]$ 内的积分等于 2，再看数值积分

```
>> trapezoidInt(@sin, 0, pi, 10)
ans = 1.983523537509455
>> trapezoidInt(@sin, 0, pi, 100)
ans = 1.999835503887444
>> trapezoidInt(@sin, 0, pi, 1000)
ans = 1.999998355065662
>> format short;
```

稀疏矩阵

稀疏矩阵（Sparse Matrix）有不同的储存方式（数据结构），这里介绍几种常见的[1].

Banded

Banded 矩阵只储存矩阵主对角线上下的若干条对角线，上带宽和下带宽分别指定主对角线上面和下面有几条对角线，例如三对角矩阵的上带宽和下带

[1]参考 Wikipedia 的 Sparse Matrix 词条

宽都是 1．带内即使有矩阵元为零也必须储存．这样就可以按照 row major 或者 column major 来储存．

Coordinate List （COO）

COO 格式列出非零矩阵元和对应的行标列标．通常将它们储存为三个数组 `a`，`ia`，`ja`，顺序任意．除此之外，有时还需要储存三个数组的长度 `nnz`（none zero）以及矩阵的尺寸．

Compressed Sparse Row （CSR）

也叫 Compressed Row Storage （CRS），这种格式做矩阵与矢量相乘较快．

CRS 格式储存为三个数组 `a`，`ia`，`ja`，非零矩阵元按照 row major 的顺序储存，第 `n` 行矩阵元是 `a(ia(n) : ia(n+1)-1)`，`ja` 是对应矩阵元的列标．

Compressed Sparse Column （CSC）

也叫 Compressed Column Storage （CCS），与 CRS 一样，只是改为 column major．

函数求值

函数求值归根结底，就是找到一个解析公式，用加减乘除表示一个函数．如果算法没有用到近似，有限精度和任意精度都是一样的，只是每个数占的内存不同罢了．任意精度的加减乘除和开根号参考 Numerical Recipes 最后一节．

最常见的展开有三种，分别是泰勒展开，渐进展开，以及连续分数．这些展开在 functions.wolfram.com 都可以找到．

接下来是如何计算级数或连续分数．求级数的方法一般是用

$$f_N(x)=\sum_{n=0}^{N}c_nx^n=(\dots((c_nx+c_{n-1})x+c_{n-2}\dots)x+c_0 \tag{1}$$

这么做误差和直接对多项式求和一样，但计算量却少了很多．将上式中的 x 换成 $1/x$ 就是渐进展开的形式．注意我们可以在计算开始前估计误差，根据精度要求得到我们需要的项数．

显然，$|x|$ 越小时泰勒展开收敛得越快，而 $|x|$ 越大时渐进展开收敛得越快．

再来看连续分数

$$f_N(x) = b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \dots \cfrac{a_N}{b_N}}} \tag{2}$$

也可以表示为

$$f_N(x) = b_0 + \frac{a_1}{b_1+}\frac{a_2}{b_2+}\frac{a_3}{b_3+}\cdots\frac{a_N}{b_N} \tag{3}$$

其中 a_n 和 b_n 都可以是 x 的函数．

乍看之下连续分数只能从右往左求，其实不然．由递归法可以证明

$$f_n = \frac{A_n}{B_n} \tag{4}$$

其中

$$A_n = b_nA_{n-1} + a_nA_{n-2} \qquad B_n = b_nB_{n-1} + a_nB_{n-2} \tag{5}$$

$$A_{-1} = 1 \qquad B_{-1} = 0 \qquad A_0 = b_0 \qquad B_0 = 1 \tag{6}$$

还有一种 Steed's 方法，详见 Numerical Recipes．正向求和的好处是可以判断什么时候开始收敛．据说特定情况下连续分数收敛较快，但具体什么时候用还有待考察．

另外两种不明觉厉的算法分别是**切比雪夫多项式（Chebyshev Polynomial）**和**算数几何平均（Agorithmic-Geometric Mean, AGM**，后者也通常被用于计算高精度的 π，且有二次收敛．

离散傅里叶变换

离散傅里叶变换（Discrete Fourier Transform）（DFT）是一个复数域的线性变换[454]．对两组有序数列 $f_0, f_1, \dots, f_{N-1}$ 和 $g_0, g_2, \dots, g_{N-1}$，正变换和逆变换分别为[2]

$$g_p = \frac{1}{\sqrt{N}}\sum_{q=0}^{N-1}\exp\left(-\frac{2\pi\mathrm{i}}{N}pq\right)f_q \tag{1}$$

[2]工程上的定义常常是正变换没有 $1/\sqrt{N}$ 因子，逆变换的 $1/\sqrt{N}$ 因子变为 $1/N$．这样的好处是节省运算量．本书中使用的定义好处是变换为幺正变换，有保持归一化的特点．

$$f_q = \frac{1}{\sqrt{N}} \sum_{p=0}^{N-1} \exp\left(\frac{2\pi \mathrm{i}}{N} pq\right) g_p \tag{2}$$

一个更常见的名词是**快速傅里叶变换（Fast Fourier Transform）（FFT）**，其定义与离散傅里叶变换一样，只是优化了算法使程序运行更快[3].

离散傅里叶变换与傅里叶变换

预备知识 傅里叶变换（指数函数）[237]

在详细分析 DFT 的性质之前，我们先看看它与解析的傅里叶变换如何对应．函数 $f(x)$ 和 $g(k)$ 间的傅里叶变换为

$$g(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) \mathrm{e}^{-\mathrm{i}kx} \,\mathrm{d}x \tag{3}$$

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} g(k) \mathrm{e}^{\mathrm{i}kx} \,\mathrm{d}k \tag{4}$$

如果 $f(x)$ 和 $g(k)$ 分别只在区间 $[-L_x/2, L_x, 2]$ 和 $[-L_k/2, L_k/2]$ 内不为零，积分就可以只在这两个区间内进行．我们再给两个区间划出 N 个等间距的格点 $\ldots, x_{-1}, x_0, x_1, \ldots$ 和 $\ldots, k_{-1}, k_0, k_1, \ldots$ （$x_0 = k_0 = 0$），并规定相邻格点的间距为

$$\Delta x = L_x/N \qquad \Delta k = L_k/N \tag{5}$$

注意 x 和 k 的首尾格点分别相距 $L_x(N-1)/N$ 和 $L_k(N-1)/N$．若 N 是奇数，我们令中间的格点为 x_0 和 k_0，如果 N 是偶数，我们令中间靠右的格点为 x_0 和 k_0.

现在我们用求和近似上面的积分得[4]

$$g(k_p) = \frac{1}{\sqrt{2\pi}} \sum_q f(x_q) \mathrm{e}^{-\mathrm{i}k_p x_q} \Delta x \tag{6}$$

$$f(x_q) = \frac{1}{\sqrt{2\pi}} \sum_p g(k_p) \mathrm{e}^{\mathrm{i}k_p x_q} \Delta k \tag{7}$$

[3]参考 Numerical Recipes 3ed

[4]事实上，学习了下文的采样定理会发现这里用求和代替积分是没有任何误差的.

式中 $k_p x_q = (\Delta x \Delta k)pq$，对比 DFT 中的指数项，得

$$\Delta x \Delta k = \frac{2\pi}{N} \tag{8}$$

但我们会发现这里的 p, q 可以是负整数，而 DFT 中的 p, q 都是非负整数．但稍加计算就会发现当 p 或 q 是负值时，把它们加上 N，指数项并不改变，例如 $k_{p+N} x_q = k_p x_q + 2\pi$，并不影响指数项．所以，我们只要将所有小于零的格点编号加上 N 并重新排列即可．例如 $N = 4$ 时，x 格点为 x_{-2}, x_{-1}, x_0, x_1，令 $x_2 = x_{-2}, x_3 = x_{-1}$，这四个格点的名字就变为 x_0, x_1, x_2, x_3．现在再来对比 DFT 和式 6 式 7，就只是相差两个常数因子而已了．

式 8 是 DFT 一个重要的性质．结合式 5 得

$$N\Delta x \Delta k = L_x \Delta k = L_k \Delta x = \frac{L_x L_k}{N} = 2\pi \tag{9}$$

离散傅里叶变换只有两个自由度，只要决定 $N, L_x, \Delta x, L_k, \Delta k$ 中的任意两个，就可以完全决定变换公式．注意 L_x 与 Δk 成反比，L_k 与 Δx 成反比，$L_x L_k$ 与 N 成正比．

总结起来，要用 DFT 数值求解一个函数 $f(x)$ 的傅里叶变换，就先用上述方法生成的格点将该函数等间距采样，然后左半和右半调换（负脚标加上 N）得到 f_i，代入 DFT 公式（在程序中使用 FFT）得到 g_i 再左半和右半调换得到 $g(k)$ 的离散点．反傅里叶变换同理．

以后如无特殊说明，默认 DFT 前后各将数列的左半右半交换一次．

变换矩阵

预备知识 酉矩阵

DFT 显然是一个线性变换，我们来看变换矩阵的性质．把变换和逆变换的系数矩阵用 $\mathbf{F}$ 和 $\mathbf{F}^{-1}$ 来表示，令列矢量 $\mathbf{f} = (f_0, f_1, \ldots, f_{N-1})^{\mathrm{T}}$，$\mathbf{g} = (g_0, g_1, \ldots, g_{N-1})^{\mathrm{T}}$，变换和逆变换分别记为

$$\mathbf{g} = \mathbf{F}\mathbf{f} \qquad \mathbf{f} = \mathbf{F}^{-1}\mathbf{g} \tag{10}$$

其中

$$F_{pq} = \frac{1}{\sqrt{N}} \exp\left(-\frac{2\pi \mathrm{i}}{N} pq\right) \tag{11}$$

下面证明，$\mathbf{F}$ 是一个酉矩阵，所以逆变换矩阵就是 $\mathbf{F}$ 的厄米共轭．

$$\mathbf{F}^{-1} = \mathbf{F}^{\dagger} \tag{12}$$

不难验证式 2 的变换矩阵的确等于 $\mathbf{F}^{\dagger}$．

根据酉矩阵的定义，我们需要证明

$$\sum_{p=0}^{N-1} F^{*}_{pq_1} F_{pq_2} = 0 \quad (q_1 \neq q_2) \tag{13}$$

而

$$\sum_{p=0}^{N-1} F^{*}_{pq_1} F_{pq_2} = \frac{1}{N} \sum_{p=0}^{N-1} \exp\left[\frac{2\pi \mathrm{i}}{N}(q_2 - q_1)p\right] \tag{14}$$

注意到求和的每一项在复平面上都对应模长为 $1/N$，幅角为 $(q_2 - q_1)p/N$ 个圆周的矢量，而 N 条矢量恰好向不同方向均匀分布，所以相加为 0．证毕．

采样定理（Sampling Theorem）

从以上的分析中，DFT 似乎只是一种近似，且有种种限制，例如我们只能取关于原点对称的区间，又例如变换只有两个自由度．我们不禁想定义更广义的离散傅里叶变换，使式 6 和式 7 中 x_i 和 k_j 可以取任意区间的任意多个等间距格点．但事实上，这样的定义并不比 DFT 更精确，而且不能用 FFT 算法提高运算速度．

这里给出傅里叶变换的**采样定理**，采样定理的详细表述是：如果 $g(k)$ 不超出 $[-L_k/2, L_k/2]$（即 $f(x)$ 是**有限带宽的**），那么我们只需要用 $\Delta x = 2\pi/L_k$（DFT 恰好满足这个条件）来对 $f(x)$ 采样就可以用以下插值公式精确还原出 $f(x)$[5]

$$f(x) = \sum_{n=-\infty}^{\infty} f(x_n)\,\mathrm{sinc}[\pi(x - x_n)/\Delta x] \tag{15}$$

其中 $\mathrm{sinc}\, x = \sin x/x$ （且定义 $\mathrm{sinc}\, 0 = 1$）．

现在我们将插值公式代入傅里叶变换得

$$g(k) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} f(x_n) \int_{-\infty}^{+\infty} \mathrm{sinc}[\pi(x - x_n)/\Delta x] \mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x \tag{16}$$

[5]其实我并不确定是否需要 $x_0 = 0$ 或者 x_i 的位置有什么其他要求，现在姑且认为 $x_0 = 0$ 好了．

结果是

$$g(k)=\begin{cases}\dfrac{1}{\sqrt{2\pi}}\displaystyle\sum_{n=-\infty}^{\infty}f(x_n)\mathrm{e}^{-\mathrm{i}kx_n}\Delta x & (|k|\leqslant L_k/2)\\ 0 & (|k|>L_k/2)\end{cases}\tag{17}$$

当 $k=k_p$ 时，这正是式 6，所以这里用求和代替积分是没有误差的．不仅如此，该式也可以作为 $g(k)$ 的插值公式．

有时候我们并不知道 $f(x)$ 的带宽，如何决定采样点的最大步长 Δx 呢？一个简单的方法是先选一个较小的 Δx，使 $f(x_i)$ 散点看起来较为连续．这时解析傅里叶变换就可以用 DFT 近似．得到 $g(k_j)$ 散点后，再判断合理的 L_k 并计算相应的 Δx 即可．

由于傅里叶变化和反变换是对称的，所以以上的分析中将 $f(x)$ 和 $g(k)$ 调换同样成立．一个有趣的问题是，是否存在一些情况使 $f(x)$ 和 $g(x)$ 都是有限带宽的？严格来说这种情况是不存在的[6]，但许多时候我们可以近似认为函数在区间外等于零，其中最常见的就是高斯波包的傅里叶变换．

归一化

我们知道傅里叶变换可以保持函数的模长不变，或者说保持函数的归一化．

假设采样定理的条件成立，我们来看如果通过离散点计算函数的模长．以 $f(x)$ 为例，用插值公式来计算模长的平方，得

$$\begin{aligned}&\int_{-\infty}^{\infty}f^*(x)f(x)\,\mathrm{d}x=\\&\qquad\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}f(x_m)^*f(x_n)\int_{-\infty}^{\infty}\operatorname{sinc}\left[\frac{\pi(x-x_m)}{\Delta x}\right]\operatorname{sinc}\left[\frac{\pi(x-x_n)}{\Delta x}\right]\mathrm{d}x\end{aligned}\tag{18}$$

果然，这个积分的结果是

$$\sum_{i=-\infty}^{\infty}|f_i|^2\Delta x\tag{19}$$

即函数的模方等于列矢量的模方乘以步长，这对 $g(k)$ 也同样成立．既然 DFT 是精确的，我们预期变换矩阵 $\mathbf{F}$ 能保证变换前后列矢量的模长相等，而这恰好是酉矩阵的性质．

[6]我们可以用反证法做一个不太严谨的证明：如果这种情况存在，那么 $f(x)$ 就可以用有限个 $\operatorname{sinc}(x)$ 函数的线性组合来表示，然而从定义上来看 $\operatorname{sinc}(x)$ 函数在 x 的区间外不可能恒为零．

插值与添零

上面我们看到，如果 $f(x)$ 是有限带宽的，就可以使用式 15 进行插值，称为 **sinc 插值**[7]．然而在实践中，由于直接用 sinc 插值计算量太大，我们往往用下面介绍的**添零法**代替．

假设我们要对 N 个散点 $f(x_i)$ 插值，可以先做 DFT 得到 N 个 $g(k_j)$，由式 9 可知，要缩小 Δx 而保持 L_x 不变，就要保持 Δk 不变且增加 L_k 和 N．所以我们可以在 $g(k_j)$ 数列的两边添加相同数量的 0，然后再做反 DFT 变换即可[8]．

然而这么做与 sinc 插值得到的结果并不完全相同．sinc 插值得到的 $f(x)$ 往往会超出原来的区间（因为 sinc 函数只以 $\Delta x/x$ 的速度衰减）．所以我们首先需要在误差能接受的范围内取一个更大 L_x（即更小的 Δk）．由于 sinc 插值函数的带宽严格小于 $L_k = 2\pi/\Delta x$，所以 L_k 不需要改变．现在我们可以认为两个函数都是有限带宽的了，要得到“包含所有信息”的 DFT，就要先保持 Δx 不变，对 $f(x_i)$ 在新区间补零，做 DFT 得到 $g(k_j)$．现在两个函数的地位完全平等，都可以用式 15 和式 17 中的任意一个来精确插值．而注意补零插值法就相当于式 17．所以这时再对 $g(k_j)$ 补任意多的零，就可以得到 $f(x)$ 任意密度的插值．在误差范围内，现在得到的 $f(x)$ 就与 sinc 插值得到的相等了．

如果直接按照 DFT 公式计算添零法，程序的速度未必比 sinc 插值要快，但如果用 FFT，那添零法将会更快[9]．

任意区间的 DFT

上面提到 DFT 或 FFT 的另一个限制就是只能在以零点为中心的两个区间之间变换 $f(x)$ 和 $g(k)$．但事实上我们只需要把数据稍作处理就可以在两个任意的区间 $[x_a, x_b]$ 和 $[k_a, k_b]$ 进行变换．

我们先定义两个新函数[10]

$$f_1(x) = f(x + x_0)\mathrm{e}^{-\mathrm{i}k_0 x} \qquad g_1(k) = g(k + k_0)\mathrm{e}^{\mathrm{i}(k+k_0)x_0} \tag{20}$$

[7]注意实现的时候只能对有限项求和．

[8]不难证明如果新的长度是 N 的整数倍，反变换后在老格点处仍能得到同样的 $f(x_i)$．

[9]如果 FFT 中只有部分格点不为零，那么 FFT 理论上可以变得更快，称为 pruned FFT，然而 FFTW 的相关页面中介绍，只要非零格点的个数大于 1%，pruned FFT 都不会有显著的性能提升．

[10]式 20 也可以改为 $f_1(x) = f(x + x_0)\mathrm{e}^{-\mathrm{i}k_0(x+x_0)}$ 和 $g_1(k) = g(k + k_0)\mathrm{e}^{\mathrm{i}kx_0}$

其中 $x_0=(x_a+x_b)/2$，$k_0=(k_a+k_b)/2$．另外定义 $L'_x=x_b-x_a$，$L'_k=k_b-k_a$，$\Delta x'=2\pi/L'_k$，$\Delta k'=2\pi/L'_x$．根据傅里叶变换的平移性质[11]，$f_1(x)$ 的傅里叶变换就是 $g_1(k)$．注意 $f_1(x)$ 和 $g_1(k)$ 的 DFT 区间 $[-L_x/2,L_x/2]$ 和 $[-L_k/2,L_k/2]$ 分别关于原点对称．现在我们只需要做 $f_1(x)$ 和 $g_1(k)$ 间的 DFT，就能得到 $f(x)$ 和 $g(k)$ 间的离散变换，不妨称为**广义 DFT**．

经过简单的推导可以发现广义 DFT 仍然可以用式 6 和式 7 表示，只是 x 和 k 的 N 个格点分别落在 $[x_a,x_b]$ 和 $[k_a,k_b]$ 内[12]．但在程序中，最高效的做法还是先将 $\mathbf{f}$ 先变为 $\mathbf{f}_1$，做 FFT 得到 $\mathbf{g}_1$ 再变为 $\mathbf{g}$．

由采样定理易证，如果 $g(k)$ 只在 $[k_a,k_b]$ 不为零，那么只需用 $\Delta x'=2\pi/L'_k$ 对 $f(x)$ 采样就可以用以下插值公式精确还原出[13] $f(x)$．

$$f(x)=\sum_{n=-\infty}^{\infty}f(x_n)\operatorname{sinc}\left[\frac{\pi(x-x_n)}{\Delta x}\right]\mathrm{e}^{\mathrm{i}k_0(x-x_n)} \tag{21}$$

实际应用中同样也可以用补零法来插值．

广义 DFT 的好处是什么呢？例如有时候非零区间 $[x_a,x_b]$ 或（和）$[k_a,k_b]$ 很窄但却远离原点，则 L_k 或 L_x 仍需取较大的值，这时 DFT 就没有发挥最大效率，因为大部分格点的函数值都是 0．但如果用广义 DFT，就可以大大减少格点数．

例 1　高斯波包

我们现在来用 FFT 数值计算以下高斯波包的傅里叶变换．

$$f(x)=\mathrm{e}^{-x^2}\mathrm{e}^{100\mathrm{i}x} \tag{22}$$

为了便于验证数值结果，解析的傅里叶变换结果是

$$g(k)=\frac{1}{\sqrt{2}}\mathrm{e}^{-(k-100)^2/4} \tag{23}$$

我们分别取 x,k 的区间为 $[-5,5]$ 和 $[90,110]$（区间外的函数值不超过 1.4×10^{-11}）．所以我们采用式 21 的傅里叶变换，令 $x_0=0$，$k_0=100$，则 $L_x\geqslant10$，$L_k\geqslant20$．为了满足 $L_xL_k=2\pi N$（见式 9），N 取最小值 32．所以不妨令 $L_x=10,L_k=2\pi N/L_x=20.106\ldots$．所以 $\Delta x=L_x/N=0.313\ldots$，$\Delta k=L_k/N=2\pi/L_x=0.628\ldots$．现在我们就可以生成 x,k 的格点并且做 FFT 了．

[11] 容易证明，DFT 也精确满足这个性质．

[12] 同样，如果 N 是奇数，x_0,k_0 是中间格点，否则就是中间靠右格点．

[13] 对离散的 $f_1(x)$ 使用之前的插值公式，再由 $f_1(x)$ 求 $f(x)$ 就可以推导出式 21．

离散正弦变换

预备知识 离散傅里叶变换[1860]

由正弦级数（$n=1,2,3\ldots$）

$$f(x)=\sum_n C_n\sin\left(\frac{n\pi}{l}x\right) \tag{1}$$

$$C_n=\frac{2}{l}\int_0^l f(x)\sin\left(\frac{n\pi}{l}x\right)\mathrm{d}x \tag{2}$$

不难推出正弦变换

$$g(k)=\sqrt{\frac{2}{\pi}}\int_0^\infty f(x)\sin(kx)\,\mathrm{d}x \tag{3}$$

$$f(x)=\sqrt{\frac{2}{\pi}}\int_0^\infty g(k)\sin(kx)\,\mathrm{d}k \tag{4}$$

注意这是一个正半轴的变换，且正反变换相同.

正弦变换同样有采样定理，即若 $g(k)$ 的区间为 $[0,L_k]$，那么只需要取 $\Delta x=\pi/L_k$ 对 $f(x)$ 采样即可用以下插值公式精确还原 $f(x)$

$$f(x)=\sum_{n=1}^\infty f(x_n)\frac{2x_n}{x+x_n}\operatorname{sinc}[\pi(x-x_n)/\Delta x] \tag{5}$$

离散正弦变换

把插值公式做正弦变换，得

$$g(k)=\sqrt{\frac{2}{\pi}}\sum_{n=1}^\infty f(x_n)\sin(kx_n)\Delta x \tag{6}$$

现在假设 $f(x)$ 和 $g(k)$ 都只在 $[0,L_x]$ 和 $[0,L_k]$ 内，所以有

$$\Delta xL_k=\Delta kL_x=N\Delta x\Delta k=\frac{L_xL_k}{N}=\pi \tag{7}$$

可得无损的离散正弦变换为

$$g_q=\sum_{p=1}^{N-1}f_p\sin(\pi pq/N) \tag{8}$$

$$f_p = \sum_{q=1}^{N-1} g_q \sin(\pi pq/N) \tag{9}$$

可以证明变换矩阵是对称的单位正交矩阵，所以逆矩阵就是矩阵本身.

第七章

计算机图形学

图像坐标系

下面介绍的这种坐标在图片处理中广泛使用（例如 Matlab 也采用这种定义[1]）.

如图 1，图片中每个像素可以表示为一个正方形，我们 x 方向指向右，y 方向指向下，令左上角的像素的正方形中心坐标为 $(1,1)$，所以坐标原点和坐标轴落在图片外面.

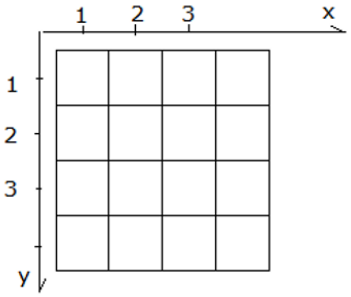


图 1: 图片坐标系

三维投影

- 本词条存在未完成的内容.

[2]当我们在平面上画三维物体时，我们需要某种投影算法把物体上的每个点对应到平面上的一点. 以下介绍两种常用的方法，一种是**平行投影（parallel projection）**，另一种是**透视投影（perspective projection）**.

平行投影

顾名思义，平行投影是指在空间中指定一个方向（如图未完成），将三维物体上的每一点沿着该方向投影到与该方向垂直的平面上. 工程制图中的正视

[1]见 imshow，ginput 等函数

[2]参考 Wikipedia 相关页面.

图，侧视图等都属于平行投影．这种投影的特点是，空间中的任意两条平行线的投影仍然是平行线．平行投影是线性的，即一个线段若伸长若干倍，那么它的投影也会按照同样的比例伸长．

透视投影

当人眼或相机观察一个三维物体时，使用的是透视投影．

未完成：用一张图介绍原理

我们在平面后方取一个固定点 F 称为**焦点**，要把物体上任意一点 P 投影到平面上，先作直线 PF，该直线与平面的交点 P' 就是投影后的点．我们在平面上建立直角坐标系，把平面上离焦点最近的点定义为平面坐标的原点．透视投影不是线性的，但可以保证直线的投影仍然为直线．

为什么透视投影要这样定义？我们先思考一种非常简单的相机，即小孔成像相机．根据光的直线传播，物体上的某点发出的（或反射的）光线只有经过小孔才能投影到孔后面的平面．如果按照这个模型，我们在计算透视投影时应该把焦点定义在屏幕之前，然而这么做有一个缺点就是投影后平面上的像是倒像．所以为了方便起见，我们保持焦点不变，但是把屏幕平移到焦点之前（焦距也不变），容易看出这样做的唯一改变就是把倒像变为正像，即点 P' 的两个平面坐标分别取相反数．

如果我们将相机上的小孔改为焦距为 f 的小凸透镜，并假设物距远大于焦距，那么根据成像公式，凸透镜和平面（即底片）的距离就是焦距 f．根据凸透镜成像原理，物体一点在底片上的像仍然会过凸透镜的中点，所以使用凸透镜的相机和使用小孔成像的相机得到的投影是相同的．至于人眼，人眼成像的原理和相机基本一致，虽然视网膜并不是一个平面，得到的成像是扭曲的，但大脑在处理图象是会自动纠正这种扭曲（例如我们看到的直线仍然是直的）．

计算方法

无论是哪种投影，我们通常建立两个坐标系，一个是**世界坐标系（world frame）** S，一个是**相机坐标系**[1872]**（camera frame）** S'．一个基本的问题就是将一个系里面的坐标变换到另一个系中的坐标．这可以通过空间旋转矩阵[404]，再加上一个平移完成（平移的矢量是两个坐标系原点之间的位移）．

把所有要投影的点 P_i （$i = 1, \ldots, N$）的在世界系中的坐标（假设已知）记为一个 3 行 N 列矩阵，矩阵的第 i 列就是列矢量 $\mathbf{P}_i = (x_i, y_i, z_i)^{\mathrm{T}}$．将 3 乘 3 的旋转矩阵左乘该矩阵，再给每行加上常数进行平移即可．完成后，我们就得到了相机系中的坐标矩阵，每一列是 $\mathbf{P}'_i = (x'_i, y'_i, z'_i)^{\mathrm{T}}$

当我们做平行投影时，可以令相机系的 x-y 平面为投影平面，投影方向为 $-\hat{\mathbf{z}}$．即 (x'_i, y'_i, z'_i) 投影后为 (x'_i, y'_i)．

当我们做透视投影时，可以令原点为焦点，$(0, 0, f)$ 为平面的中点，平面与 z 轴垂直．相机系中的某点 (x'_i, y'_i, z'_i) 投影后变为 $(x'_i f/z'_i, y'_i f/z'_i)$，即每个分量乘以 f/z'_i，使得 z 分量等于 f．

未完成：另外开一篇文章分享平行投影和透视投影的 Matlab 代码

3D 艺术画

见 3D 艺术画[1874]．

相机模型

- 本词条处于草稿阶段．

预备知识 三维投影[1870]，图片坐标系[1870]

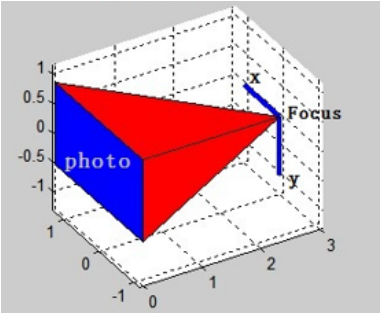


图 1: 相机坐标系

从焦点出发，穿过底片并与底片垂直的射线叫做**光轴**．

以焦点为原点，光轴方向为 z 轴，图片的右方为 x 轴，下方为 y 轴（注意这是常用的右手系[20]）.

真实的相机和这里的相机模型存在误差，通常需要将图片进行纠偏后才能得到符合模型的图像．以后我们讨论的图像一般指纠偏后的图像.

注意图像的中点不一定是光轴和底片的交点.

相机参数：焦距，底片宽度，底片高度，光轴的图像坐标

相机的位置参数：焦点的位置，三个单位矢量在世界坐标系中的坐标（即旋转矩阵，也可以使用四元数表示）

反过来，若已知图片和相机位置和参数，我们可以将图片上的每一点对应到以相机为起点的一条射线．而至于这点具体在射线上的什么位置无从得知．但如果我们从两个不同的位置拍摄同一点，就可以通过两条射线的交点确定该点位置.

若已知相机参数，我们可以通过图片中的任意两点确定它们关于相机的夹角（即两条射线的夹角）.

变换方程

变换有 7 个自由度，相机位置 3 个，相机角度 3 个，相机焦距 1 个.

一般形式为（推导：根据相机模型结合立体几何）

$$\begin{cases} x' = \dfrac{c_4x + c_5y + c_6z + c_7}{c_1x + c_2y + c_3} \\ y' = \dfrac{c_8x + c_9y + c_{10}z + c_{11}}{c_1x + c_2y + c_3} \end{cases} \tag{1}$$

注意这里的 9 个常系数中只有 7 个是独立的．所以理论上我们可以只用 4 个一般位置的点及它们在图片上的坐标解得所有的系数，但这样方程就不是线性的了．所以我们也可以假设 9 个系数都是独立的，并用 6 个或以上一般位置的点得到 12 条（或以上）关于 c_i 的超定齐次线性方程组，并利用超定来减小误差.

如果我们不仅需要得到变换公式，还需要从这些系数里面得到相机的位置和朝向，那么我们可以从立体几何的角度出发，例如见长方形相机定位法[1878]．但这样我们需要的 4 个点就必须是长方形的四个顶点而不是任意点.

平面到平面的变换

注意式 1 不存在逆变换，因为我们把空间中的点投影到平面上时丢失了景深的信息．但如果我们限制所有空间点 (x, y, z) 都在某个平面上（不失一般性，假设它们在 x-y 平面上），那两个平面上的点就有一一对应的关系．

$$\begin{cases} x' = \dfrac{c_4 x + c_5 y + c_6}{c_1 x + c_2 y + c_3} \\ y' = \dfrac{c_7 x + c_8 y + c_9}{c_1 x + c_2 y + c_3} \end{cases} \tag{2}$$

可以证明其逆变换也有相同的形式，虽然这并不显然．

由图像坐标计算射线

- 本词条处于草稿阶段．

预备知识 相机模型[1872]

图片中的每个坐标对应一个射线．

从世界坐标到相机坐标（记为 x', y'）的矩阵为 $\mathbf{R}$，相机焦点位置为 $\mathbf{c}$，焦距为 f．

相机坐标系中，射线的起点为原点，方向为 $\mathbf{v}' = (x', y', f)$．

那么在世界坐标系中，这条射线的起点为 $\mathbf{c}$，方向为 $\mathbf{v} = \mathbf{R}^{\mathrm{T}}\mathbf{v}'$．

计算 3D 艺术画

- 本词条需要更多参考文献．
- 本词条处于草稿阶段．

3D 艺术画

计算

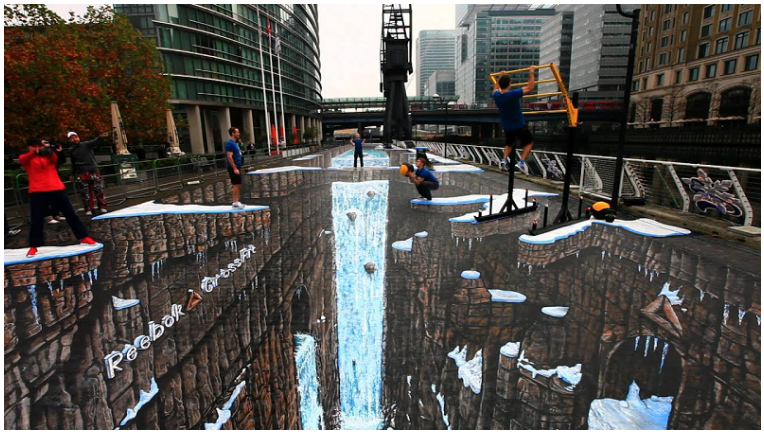


图 1: 2011 年最大的 3D 艺术画，画于伦敦街头，获吉尼斯纪录



图 2: 从不同角度看 3D 艺术画（图 1 由该图中右上角相机拍摄）

预备知识 由图像坐标计算射线[1874]，直线和平面的交点[109]

由预备知识中的两个词条，若给出预期的效果图（即图 1 中的虚构部分），世界坐标系中投影平面的位置，以及相机模型的参数和位置，就可以将效果图上的点一一对应到世界系的射线，再对应到投影平面上.

我们不直接投影到世界系的 xy 平面上，是因为一些 3D 艺术画不仅使用地板，也可能使用墙壁和天花板等多个平面.

代码 1: art3D.m

```
1 % 参数中矢量不必归一化
2
```

```
% === 设置相机参数 ===
imfile = 'test.png'; % 图片位置
cam_pos = [10, -10, 10]; % 世界参考系中相机位置
cam_z = -cam_pos; % 世界参考系中相机朝向
view_ang = 40/180*pi; % 相机水平视角 (单位: 弧度)
cc = [0, 0]; % 光轴相对图片中心的偏移 (单位:  像素),

% === 设置投影平面 ===
plane_o = [0, 0, 0]; % 原点
plane_z = [0, 0, 1]; % z 轴单位矢量
plane_x = [1, 0, 0]; % x 轴单位矢量
% ================

% 相机坐标系
% 相机的 x 轴总是水平
cam_z = cam_z / norm(cam_z);
cam_x = cross([0, 0, -1], cam_z); cam_x = cam_x / norm(cam_x);
cam_y = cross(cam_z, cam_x);
RT = [cam_x', cam_y', cam_z']; % rotation matrix

% 平面坐标系
plane_z = plane_z / norm(plane_z);
plane_y = cross(plane_z, plane_x);
plane_y = plane_y / norm(plane_y);
plane_x = cross(plane_y, plane_z);

% 读取图片
I = imread(imfile);
[N1, N2, ~] = size(I);
f = N2/2/tan(view_ang/2); % 焦距 (单位:  像素)

% 在相机系中生成图片网格 (每个像素都是网格中一个正方形)
[xc, yc] = meshgrid(linspace(0.5, N2+0.5, N2+1), linspace(0.5,
    N1+0.5, N1+1));
xc = xc - N2/2 + 0.5;
yc = yc - N1/2 + 0.5;
```

```
38
39 % 世界系中的射线（对应每个格点）
40 vc = zeros(3, (N1+1)*(N2+1));
41 vc(1,:) = xc(:); vc(2,:) = yc(:); vc(3,:) = f;
42 v = RT * vc;
43
44 % 世界系中射线与平面的交点（参考 https://wuli.wiki/online/LPint.
   html）
45 intersec = ((plane_o - cam_pos) * plane_z') * v ./ (plane_z * v)
   + cam_pos';
46
47 % 将交点转换到平面参考系
48 inter1 = intersec - plane_o';
49 if max(plane_z * inter1) > 1e-14
50     error('intersection not on plane');
51 end
52 xp = plane_x * inter1;
53 yp = plane_y * inter1;
54 xp = reshape(xp, N1+1, N2+1);
55 yp = reshape(yp, N1+1, N2+1);
56
57 % 画图
58 zp = zeros(size(xp));
59 cdata = double(I)/255; cdata(end+1, end+1, 3) = 0;
60 figure; surf(xp, yp, zp, 'CData', cdata, 'LineStyle', 'none');
61 view(0, 90); axis equal;
```

相机的定位

- 本词条处于草稿阶段.

预备知识 相机模型[1872]

若我们已知世界系中若干点 $\mathbf{p}_i$（$i=1,\ldots,N$）的坐标（假设任意 3 个 $\mathbf{p}_i$

不在一条直线上），而一个相机同时拍摄了这些点，我们在一定条件下可以根据得到的图片确定相机的位置和朝向.

若 $N=3$ 且，这个问题一般有 3 个解，但若 $N=4$ 时，就只有一个解. 当 N 更大时，可以用多余的点进一步减小计算误差.

我们可以通过计算得到图片上任意两点的夹角，所以当 $N=3$ 时，我们相当于在解三棱锥：已知三棱锥三条底边的长度以及顶点的三个顶角，求顶点的坐标. 如果把三条棱的长度作为未知数，我们会得到非线性的三元二次方程组，这个方程组最多有三个解.

（方程组未完成）

另一种方法是使用长方形，见“长方形相机定位法[1878]”.

长方形相机定位法

- 本词条处于草稿阶段.

预备知识 相机模型[1872]，直线和平面的交点[109]

如果我们事先不知道相机参数，假设图片已经纠偏，假设光心在图片的中点，我们仅通过一个长方形在图片上的位置判断长方形的边长比，以及相对相机的位置.

思路

找到图片上四边形的两组对边的交点，这两个点关于相机焦点的空间夹角必为直角，由此可以判断相机焦距. 也就是说只要通过图片中的一个长方形就可以确定相机的焦距或者视角！

这两个点对应的射线就是两组平行线的方向，由此可以确定平面的法向量，令长方形的较短边长为 1，两条边分别为世界系的 x 轴和 y 轴，由此可以确定相机的位置和方向.

两条射线的矢量分别为 (x_1,y_1,f) 以及 (x_2,y_2,f).

$$x_1x_2+y_1y_2+f^2=0 \tag{1}$$

$$f = \sqrt{-x_1x_2 - y_1y_2} \tag{2}$$

再根据四条射线和平面的交点公式式 4 [109] 就可以解出长方形两条边长的比例.

解三棱锥 1 （Matlab）

- 本词条处于草稿阶段.

未完成：给出公式

若已知 4 个点的坐标为 $\mathbf{P}_i = (x_i, y_i, z_i)$，且某点 $\mathbf{P} = (x, y, z)$ 到它们的距离 r_i，那么可以唯一地解出 $\mathbf{P}$.

$$(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 = r_i^2 \qquad (i = 1, 2, 3, 4) \tag{1}$$

以下给出 Matlab 代码.

未完成：该程序输入只有三个点，按照右手定则来选取. 应该再写一个程序输入 4 个点，输出唯一的解.

```
% 已知三棱锥底面的三个点坐标 P1,P2,P3, 和三条棱 L1,L2,L3, 求顶点坐标
function P = pmdL32P(L1,L2,L3,P1,P2,P3)

P12 = P2-P1; mag = norm(P12);
C1 = (L1^2+mag^2-L2^2)/(2*L1*mag);% C1 = cos<P1P,P1P2>
P12 = P12/mag;

P13 = P3-P1; mag = norm(P13);
C2 = (L1^2+mag^2-L3^2)/(2*L1*mag);% C2 = cos<P1P,P1P3>
P13 = P13/norm(P13);

% P12和P13的法向量
vn = cross(P12,P13); vn = vn/norm(vn);

```

```
% 求单位投影矢量
M33 = [P12; P13; vn]; % 方程组: xs,ys,zs的点是单位P1P在P1P2,P1P3
    平面上的投影矢量(单位矢量)
C = [C1; C2; 0];      % 第一条: 矢量在P1P2方向的投影 = C1
Pproj = (M33\C)';     % 第二条: 矢量在P1P3方向的投影 = C2
                      % 第三条: 矢量在(nx,ny,nz)方向的投影 = 0
% 还原
a = sqrt(1-norm(Pproj)^2);
P = L1*(a*vn+Pproj)+P1;
end
```

未完成：demo

第八章

微分方程数值解

简谐振子受迫运动的简单数值计算

预备知识 简谐振子受迫运动[1004]，Matlab 编程基础[1682]

以下以弹簧振子的受迫运动为例，介绍一种解 n 阶微分方程的简单方法，以便了解数值解微分方程的基本思想．但是这种方法误差较大，需要大量计算才能获得较精确的数值解．在实际运用中已有更复杂更成熟的算法（参考 MATLAB 常微分方程（组）数值解简介）．

在简谐振子受迫运动[1004] 中，列出的二阶微分方程为

$$m\ddot{y} = \alpha\dot{y} - ky + f(t) \tag{1}$$

若已知初值条件（可代入任意具体数值）$\dot{y}(0) = v_0$，$y(0) = y_0$，且已知驱动力 $f(t)$，此时可以把初值条件代入式 1 求出 $t = 0$ 时的加速度．

$$\ddot{y}(0) = [-\alpha\dot{y}(0) - ky(0) + f(0)]/m \tag{2}$$

接下来的一小段微小的时间 Δt 内（Δt 称为**步长**，步长越小误差越小），根据微分近似，可以算出 $t = \Delta t$ 时刻的状态．

$$y(\Delta t) = y(0) + \Delta y \approx y(0) + \dot{y}(0)\Delta t \tag{3}$$

$$\dot{y}(\Delta t) = \dot{y}(0) + \Delta\dot{y} \approx \dot{y}(0) + \ddot{y}(0)\Delta t \tag{4}$$

微分近似在这里的物理意义是在 Δt 内速度和加速度都近似为常数．把 $y(\Delta t)$ 和 $\dot{y}(\Delta t)$ 再次代入式 1

$$\ddot{y}(\Delta t) = [-\alpha\dot{y}(\Delta t) - ky(\Delta t) + f(\Delta t)]/m \tag{5}$$

再次使用微分近似有

$$y(2\Delta t) = y(\Delta t) + \Delta y \approx y(\Delta t) + \dot{y}(\Delta t)\Delta t \tag{6}$$

$$\dot{y}(2\Delta t) = \dot{y}(\Delta t) + \Delta\dot{y} \approx \dot{y}(\Delta t) + \ddot{y}(\Delta t)\Delta t \tag{7}$$

重复以上各步骤，就可以继续得到 $y(3\Delta t)$，$y(4\Delta t)$ 等的近似值．在 y-t 图中把这些散点连接起来，就得到了 $y(t)$ 的函数图．

代码 1: SHOf.m

```
% 参数设定
m = 0.1; % 质量
k = 1; % 劲度系数
a = 0.03; % 阻尼系数
T = 20; % 停止时间
Nstep = 10000; % 步数
A = 2; w = 3; % f(t)=A*sin(w*t);

dt = T/Nstep; % 计算步长
y2 = zeros(step,1); y1 = y2; y = y2; % 矩阵预赋值
y(1) = 0; y1(1) = 0; % 初值, y1 是 y 的一阶导数

% 迭代循环
for ii = 2:step
    y2(ii) = (-a*y1(ii)-k*y(ii)+2*sin(w*(ii*dt)))/m; % 代入微分
        方程求出 y''.
    y(ii) = y(ii-1) + y1(ii-1)*dt; % y 的微分近似
    y1(ii) = y1(ii-1) + y2(ii-1)*dt; % y' 的微分近似
end

% 画图
t=(0:step-1)*dt;
plot(t,y);
```

运行结果如图 1，可见开始时驱动力不断给弹簧振子补充能量，振幅变大. 当补充的功率等于消耗的功率时，弹簧做稳定振动.

天体运动的简单数值计算

预备知识 万有引力[1027]，弹簧振子受迫运动的简单数值计算[1882]

下面我们来用一种极其简单的算法对单个天体在中心天体的万有引力作用下的运动进行数值计算. 事实上该问题存在解析解（见开普勒三定律[1033]），所以以下的算法只是用于演示数值解常微分方程的大致原理. 这种方法可以轻

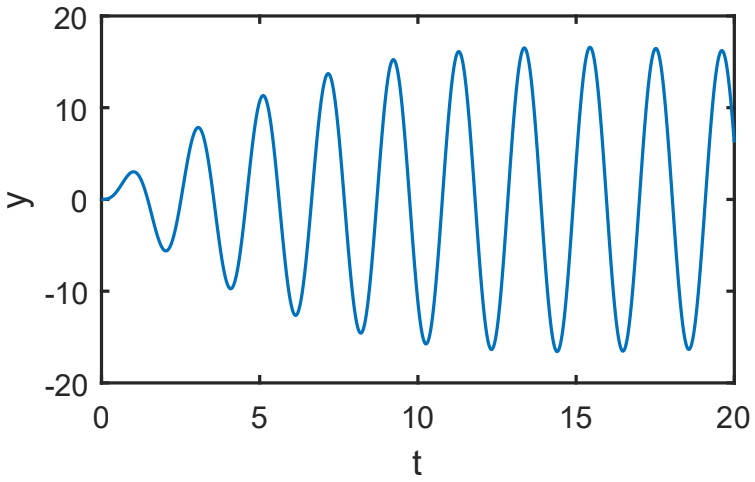


图 1: 运行结果

易地拓展到多个天体的情况，而多体情况没有一般的解析解．一种效率更高的常见常微分方程数值算法可参考“四阶龙格库塔法[1891]”.

直角坐标系中，设中心天体质量为 M，固定在原点不动．根据牛顿万有引力定律，质量为 m 的行星受到中心天体的力为

$$\mathbf{F} = -G\frac{Mm}{r^2}\hat{\mathbf{r}} = -G\frac{Mm}{r^3}\mathbf{r} \tag{1}$$

其中 $\mathbf{r}$ 为行星的位矢（设行星在 xy 平面上运动）．根据牛顿第二定律[899]，加速度为

$$\mathbf{a} = \frac{\mathbf{F}}{m} = -G\frac{M}{r^3}\mathbf{r} \tag{2}$$

用 $r = \sqrt{x^2+y^2}$ 以及 $\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}}$，其中 x, y 看成 t 的函数．考虑到 $a_x = \ddot{x}$，$a_y = \ddot{y}$，可以列出二阶微分方程组（变量上方两点表示关于时间的二阶导数）

$$\begin{cases} \ddot{x} = -\dfrac{GM}{(x^2+y^2)^{3/2}}x \\ \ddot{y} = -\dfrac{GM}{(x^2+y^2)^{3/2}}y \end{cases} \tag{3}$$

假设已知初值条件 $x(0) = x_0$，$y(0) = y_0$，$\dot{x}(0) = v_{x0}$，$\dot{y}(0) = v_{y0}$．下面用“弹簧振子受迫振动的简单数值计算[1882]”中类似的方法求接下来行星的运动轨迹．

1．将初始条件代入式 3，得到初始加速度

$$\begin{cases} \ddot{x}(0) = -\dfrac{GM}{(x_0^2+y_0^2)^{3/2}}x_0 \\ \ddot{y}(0) = -\dfrac{GM}{(x_0^2+y_0^2)^{3/2}}y_0 \end{cases} \tag{4}$$

2．设经过一段极微小的时间步长 Δt（例如 0.0001，数值越小误差越小），根据微分近似（微分近似在这里的物理意义是在 Δt 内速度和加速度都近似为常数）

$$\begin{cases} \dot{x}(\Delta t) \approx \ddot{x}(0)\Delta t + x(0) \\ \dot{y}(\Delta t) \approx \ddot{y}(0)\Delta t + \dot{y}(0) \end{cases} \qquad \begin{cases} x(\Delta t) \approx \dot{x}(0)\Delta t + x(0) \\ y(\Delta t) \approx \dot{y}(0)\Delta t + y(0) \end{cases} \tag{5}$$

3．把 $x(\Delta t), y(\Delta t)$ 再次代入式 3，得到 $x''(\Delta t), y''(\Delta t)$，再次利用微分近似求出 $x(2\Delta t), y(2\Delta t), x'(2\Delta t), y'(2\Delta t)\ldots$ 如此循环下去就可以得到每隔 Δt 的数值解．

代码 1: kepler.m

```
% 参数设定
GM = 1; % 万有引力常数乘以中心天体质量
x0 = 1; y0 = 0; % 初始位置
vx0 = 0; vy0 = 0.7; % 初始速度
T = 4; Nstep = 4000; % 总时间和步数
dt = T/Nstep; % 步长

% 矩阵预赋值
x = nan(Nstep,1); y = x;
x1 = x;  y1 = x;
x2 = x; y2 = x;

% 初始位置，速度，加速度
x(1) = x0; y(1) = y0; % 初位置
x1(1) = vx0; y1(1) = vy0; % 初速度
x2(1) = -GM*x(1)/(x(1)^2+y(1)^2)^(3/2); % 代入方程得到 x''(0)
y2(1) = -GM*y(1)/(x(1)^2+y(1)^2)^(3/2); % 代入方程得到 y''(0)

% 迭代循环
for ii = 2:Nstep
    x(ii) = x(ii-1)+x1(ii-1)*dt; % x的微分
    y(ii) = y(ii-1)+y1(ii-1)*dt; % y的微分

    x1(ii) = x1(ii-1)+x2(ii-1)*dt; % x' 的微分
```

```
    y1(ii) = y1(ii-1)+y2(ii-1)*dt; % y' 的微分

    x2(ii) = -GM*x(ii)/(x(ii)^2+y(ii)^2)^(3/2); % 代入微分方程求出 x''
    y2(ii) = -GM*y(ii)/(x(ii)^2+y(ii)^2)^(3/2); % 代入微分方程求出 y''
end

% 画图
plot(x,y); % 画行星轨道
axis equal; % xy坐标长度一致
hold on; % 继续画图
scatter(0,0); % 标出中心天体
```

程序运行结果如图 1 所示．注意行星轨道并不是一个闭合的椭圆，这是由于这种算法误差较大，为了减小误差，可以增加程序中 `Nstep` 的值．

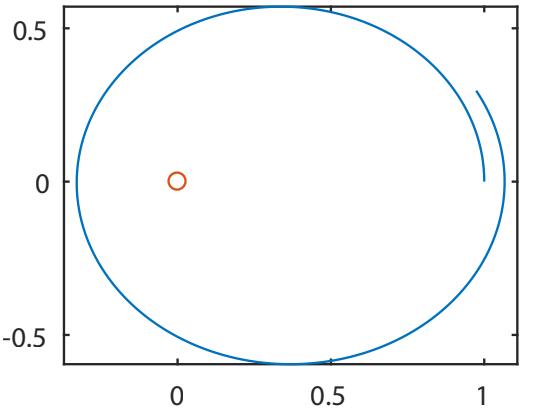


图 1: 运行结果

常微分方程（组）的数值解

预备知识 天体运动的简单数值计算[1883]

数值解微分方程或微分方程组时，一般需要先把微分方程（组）化成一阶微分方程组．例如在“弹簧振子受迫运动的简单数值计算[1882]”中列出的微分

方程为

$$my''=\alpha y'-ky+f(t) \tag{1}$$

新增变量，令 $v=y'$，则可变为一阶微分方程组

$$\begin{cases} v'=[-\alpha u-ky+f(t)]/m \\ y'=v \end{cases} \tag{2}$$

方程组中，t 是自变量，y 和 u 是 t 的函数．给出某时刻的 $y(t),u(t),t$ 就可以通过方程组求出 $u'(t)$ 和 $y'(t)$ 的数值．

又例如在“天体运动的简单数值计算[1883]”中，列出的二阶微分方程组为

$$\begin{cases} x''=-\dfrac{GM}{(x^2+y^2)^{3/2}}x \\ y''=-\dfrac{GM}{(x^2+y^2)^{3/2}}y \end{cases} \tag{3}$$

新增变量，令 $v_x=x'$，$v_y=y'$，上式也可变为一阶微分方程组

$$\begin{cases} x'=v_x \\ v_x'=-GMx/(x^2+y^2)^{3/2} \\ y'=v_y \\ v_y'=-GMy/(x^2+y^2)^{3/2} \end{cases} \tag{4}$$

该式中同样 t 是自变量，其他都是 t 的函数，给出某时刻的 x,y,v_x,v_y,t 就可以由该方程组求出各函数的一阶导数．

在以上两个例子中，我们使用了一种较为原始的方法（微分近似）．这种方法相当于把某时刻 t 的各函数值代入一阶微分方程组，得到 t 时刻各函数的一阶导数，再通过微分近似由这些一阶导数来计算其 $[t,t+\Delta t]$ 时间内的增量，得到 $t+\Delta t$ 时刻的各函数值，再代入一阶方程组得到 $t+\Delta t$ 时刻各函数的一阶导数，如此一直循环，得到各函数每隔 Δt 时间的值．这种方法叫做**欧拉法**．对一阶常微分方程 $y'(t)=f(y,t)$，令 $h=\Delta t$，$t_n=t_0+nh$，$y_n=y(t_n)$ 欧拉法可以表示为

$$y_{n+1}=y_n+hf(y_n,t_n) \tag{5}$$

以下介绍三种更精确的算法，在步长（Δt）相同时它们的精确度递增．注意虽然大多数问题中自变量是时间，但这些方法适用于大多数单个自变量的微分方程（组）．

- 中点法[1888]
- 四阶龙格库塔法[1891]
- Matlab 中的 Ode45 算法

中点法解常微分方程（组）

预备知识 常微分方程（组）的数值解[1886]

我们先来尝试用欧拉法解一阶微分方程

$$y'(t) = y \tag{1}$$

令初始条件为 $y(0) = 1$．令步长为 $h = 0.5$，步数为 5，结果如图 1 所示（代码见词条最后）.

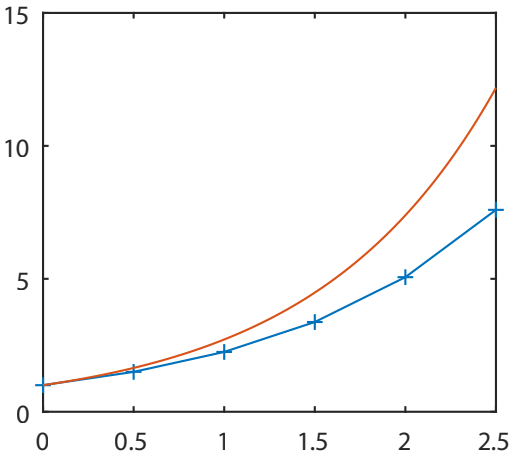


图 1: 欧拉法数值解（蓝）和解析解（红）

我们知道该方程的解析解为 $y = \mathrm{e}^t$．对比数值解和解析解，不难分析出误差产生的原因：我们仅用每段步长区间左端的导数预测整个区间的曲线增量．如果我们能利用每个区间中点的导数计算整个区间的增量，这个预测将会比欧拉法更精确.

考虑微分方程 $y'(t)=f(y,t)$ 在区间 $[t_n,t_{n+1}]$ 的曲线，若我们已知区间左端的函数值为 y_n，我们可以先用微分近似估计曲线中点的函数值为

$$y\left(t_n+\frac{h}{2}\right)=y_n+\frac{h}{2}f(y_n,t_n) \tag{2}$$

然后再求出这个近似中点的导数为

$$y'\left(t_n+\frac{h}{2}\right)=f\left[y_n+\frac{h}{2}f(t_n,y_n),t_n+\frac{h}{2}\right] \tag{3}$$

最后我们利用这个导数估算该区间的曲线增量

$$y_{n+1}=y_n+hy'\left(t_n+\frac{h}{2}\right)=y_n+hf\left[y_n+\frac{h}{2}f(t_n,y_n),t_n+\frac{h}{2}\right] \tag{4}$$

这就是解常微分方程的**中点法**.

我们再来用中点法取同样的步长计算式 1，结果如图 2 所示.

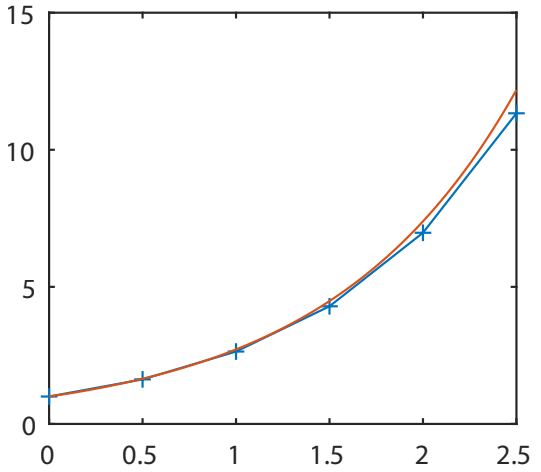


图 2: 中点法数值解（蓝）和解析解（红）

可见虽然中点法每一步的计算过程比欧拉法稍微复杂一些，但精度却大大地提高了.

中点法同样适用于微分方程组，例如对于常微分方程组

$$\begin{cases}x'(t)=f(x,y,t)\\y'(t)=g(x,y,t)\end{cases} \tag{5}$$

首先计算近似中点为

$$\begin{cases} x_{n+1/2} = x_n + \frac{h}{2} f(x_n, y_n, t_n) \\ y_{n+1/2} = y_n + \frac{h}{2} g(x_n, y_n, t_n) \end{cases} \tag{6}$$

然后有

$$\begin{cases} x_{n+1} = x_n + hf\big(x_{n+1/2}, y_{n+1/2}, t_n + h/2\big) \\ y_{n+1} = y_n + hg\big(x_{n+1/2}, y_{n+1/2}, t_n + h/2\big) \end{cases} \tag{7}$$

代码 1: odeMid.m

```
% 设置参数
N = 6;
h = 0.5;
t = linspace(0, (N-1)*h, N); % 自变量
t0 = linspace(0, (N-1)*h, 100); % 用于画图

% 欧拉法
y = zeros(1,N); % 预赋值
y(1) = 1; % 初值
for ii = 1:N-1
    y(ii+1) = y(ii) + h*y(ii);
end

% 画图
figure;
plot(x,y,'+-');
hold on;
plot(t0, exp(t0));

%中点法
y = zeros(1,N);
y(1) = 1;
for ii = 1:N-1
    y(ii+1) = y(ii) + h*(y(ii) + 0.5*h*y(ii));
end

```

```
% 画图
figure;
plot(x,y,'+-');
hold on;
plot(x0, exp(x0));
```

四阶龙格库塔法

预备知识 中点法解常微分方程（组）[1888]

龙格库塔法是一类数值解微分方程的算法，其中较常见的是**四阶龙格库塔法**．这里不进行推导，仅仅给出公式如下（y_n, t_n, h 的定义类比式 5 [1886]）

$$y_{n+1} = y_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) \tag{1}$$

其中

$$\begin{aligned} k_1 &= f(y_n, t_n) & k_2 &= f\left(y_n + h\frac{k_1}{2}, t_n + \frac{h}{2}\right) \\ k_3 &= f\left(y_n + h\frac{k_2}{2}, t_n + \frac{h}{2}\right) & k_4 &= f(y_n + hk_3, t_n + h) \end{aligned} \tag{2}$$

由以上两式，不难把该算法拓展到方程组的情况．对于 N 元微分方程组

$$\begin{cases} y_1'(t) = f_1(y_1, \ldots, y_N, t) \\ y_2'(t) = f_2(y_1, \ldots, y_N, t) \\ \quad\vdots \\ y_N'(t) = f_N(y_1, \ldots, y_N, t) \end{cases} \tag{3}$$

我们可以把该式记为矢量函数的形式

$$\mathbf{y}'(t) = \mathbf{f}(\mathbf{y}, t) \tag{4}$$

现在我们仅需要把式 1 和 式 2 中的所有 y_i 和 k_i 都变为 N 维列矢量 $\mathbf{y}_i$ 和 $\mathbf{k}_i$ 即可将微分方程拓展为微分方程组．

例程

到目前为止，我们每求一个微分方程的数值解都要重新写一次程序，对于一些较为复杂的算法这样做效率较低. 我们这里不妨把四阶龙格库塔法写到一个单独的函数文件 odeRK4.m 中，当我们要解某个特定的方程时，只需把式 2 中的 $f(y,t)$ 作为自变量输入即可解出 $y(t)$.

代码 1: odeRK4.m

```
1  function [Y, t] = odeRK4(f, tspan, Y0, Nt)
2  Nvar = numel(Y0);  % 因变量的个数
3  dt = (tspan(2) - tspan(1)) / (Nt-1); % 计算步长
4  Y = zeros(Nvar, Nt); % 预赋值
5  Y(:, 1) = Y0(:); % 初值
6  t = linspace(tspan(1), tspan(2), Nt);
7
8  for ii=1:Nt-1
9      K1 = f(Y(:,ii)           , t(ii)      );
10     K2 = f(Y(:,ii)+K1*dt/2   , t(ii)+dt/2 );
11     K3 = f(Y(:,ii)+K2*dt/2   , t(ii)+dt/2 );
12     K4 = f(Y(:,ii)+K3*dt     , t(ii)+dt   );
13     Y(:,ii+1) = Y(:,ii) + dt/6 * (K1+2*K2+2*K3+K4);
14 end
15 end
```

我们先来看第 1 行的函数声明，输入变量中，`f` 是式 4 中 $\mathbf{f}(\mathbf{y},t)$ 的函数句柄[1700]，`tspan` 是一个 2℄1 的列矢量，`tspan(1)` 是初始时间，`tspan(2)` 是终止时间，`Y0` 是一个列矢量，`Y0(ii)` 是第 `ii` 个因变量的初始值，`Nt` 是 t_n 的个数，`tspan` 定义的时间区间被等分为 `Nt - 1` 个小区间. 因变量中，`Y` 的行数是因变量的个数，列数是 `Nt`，`t` 是一个行矢量，由第 6 行定义，`Y(ii, jj)` 就是第 `ii` 个变量在 `t(jj)` 时刻的值. 第 5 行把初值 `Y0` 赋给 `Y` 的第 1 列，第 8-14 行的循环根据式 1 和式 2 的矢量形式由 `Y` 的第 `ii` 列（$\mathbf{y}_i$）求第 `ii+1` 列（$\mathbf{y}_{i+1}$）.

我们先来用这个函数来计算“天体运动的简单数值计算[1883]”中的问题.

我们令因变量 $\mathbf{y}$ 的四个分量依次为一阶方程组（式 4 [1886]）

$$\begin{cases} x' = v_x \\ y' = v_y \\ v_x' = -GMx/(x^2+y^2)^{3/2} \\ v_y' = -GMy/(x^2+y^2)^{3/2} \end{cases} \tag{5}$$

中的 x, y, v_x, v_y. 程序代码如下

代码 2: keplerRK4.m

```
function keplerRK4
% 参数设定
GM = 1; % 万有引力常数乘以中心天体质量
x0 = 1; y0 = 0; % 初始位置
vx0 = 0; vy0 = 0.7; % 初始速度
tspan = [0; 4]; % 总时间和步数
Nt = 100; % 步数

Y0 = [x0; y0; vx0; vy0]; % 因变量初值
f = @(Y, t)fun(Y, t, GM);
[Y,~] = odeRK4(f, tspan, Y0, Nt);

% 画图
figure; hold on;
plot(Y(1,:), Y(2,:));
scatter(0, 0);
axis equal;
end

function Y1 = fun(Y, ~, GM)
% 因变量
x = Y(1); y = Y(2);
vx = Y(3); vy = Y(4);
Y1 = zeros(4,1); % 预赋值
Y1(1) = vx;
Y1(2) = vy;
```

```
temp = -GM /(x^2 + y^2)^(3/2);
Y1(3) = temp * x;
Y1(4) = temp * y;
end
```

运行结果如图 1 所示.

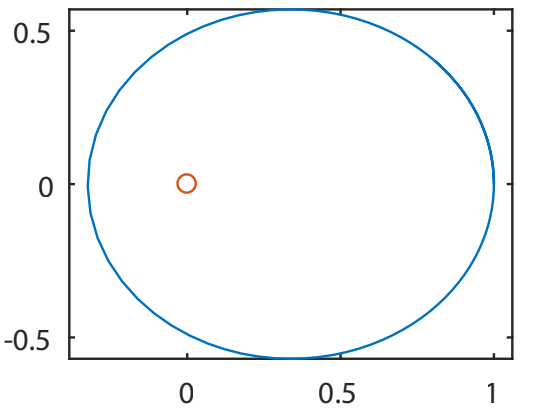


图 1: 运行结果

我们先来看函数 fun （20 行），这个函数就相当于式 5. 第一个输入变量 Y 是一个列矢量，是 $\mathbf{y}_n$ 的值，第二个输入变量是 t，但由于式 5 中没有出现 t，我们用波浪线代替. 第三个输入变量是参数 GM，即万有引力常数和中心天体质量之积. 输出变量 Y1 是一个列矢量，是 $\mathbf{y}'_n$ 的值.

再来看主函数 KeplerRK4 （第 1 行），参数设定中除了步数从 4000 变为了 100，其他都和“天体运动的简单数值计算[1883]”中的程序一样，然而这里运行结果却精确得多（曲线几乎闭合），可见这种算法的优越性.

主函数第 10 行中将 fun(Y, t, GM) 变为函数句柄 f(Y, t)，这样 GM 就可以在“参数设定”中设置，而不用在 fun 函数内部设置. 第 11 行调用了上文中的 odeRK4 函数解方程组，由于我们在画图时不需要用 t，所以把第二个输出变量改为波浪线.

刚体转动数值模拟

预备知识 刚体运动方程（四元数）[981]，四阶龙格库塔法[1891]

这里给出使用 4 阶龙格库塔法模拟刚体绕固定点转动的 Matlab 代码．程序中给出的默认参数用于模拟一个初始静止的长方体受到一个 $(1,1,1)$ 方向的大小为 0.05 的恒力矩时的加速转动．

运行结果见互动演示，截图如图 1．

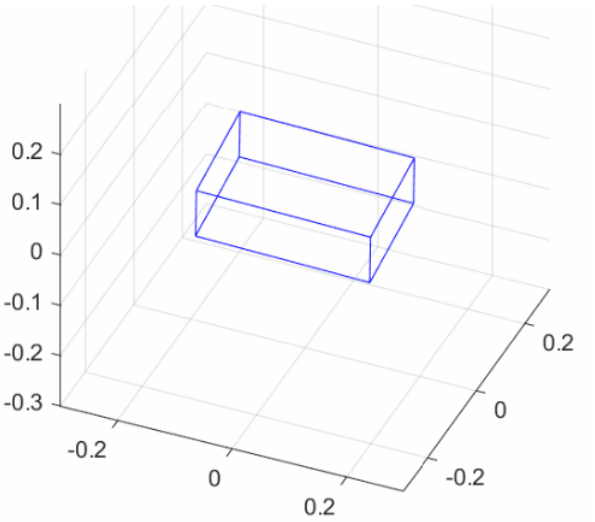


图 1: 动画截图

为了验证数值解的正确性，我们在生成动画之前用 `verify` 函数验证角动量定理，也就是把力矩函数 `tau(t)` 做数值积分，再与数值解中的角动量进行对比，结果如图 2．

代码 1: rigBdRot.m

```
% 刚体绕固定点转动的数值计算

function rigid_body_rotation
close all;
% === 设置长方体 ===
a = 0.3; b = 0.2; c = 0.1; M = 1; % 边长和质量
I0 = M/12 * diag([b^2 + c^2, a^2 + c^2, a^2 + b^2]); % 惯性张量
w0 = [0; 0; 0]; % 初始角速度
q0 = [1; 0; 0; 0]; % 初始朝向（四元数）
tau = @(t) 0.05*[1; 1; 1]/sqrt(3); % 力矩函数
```

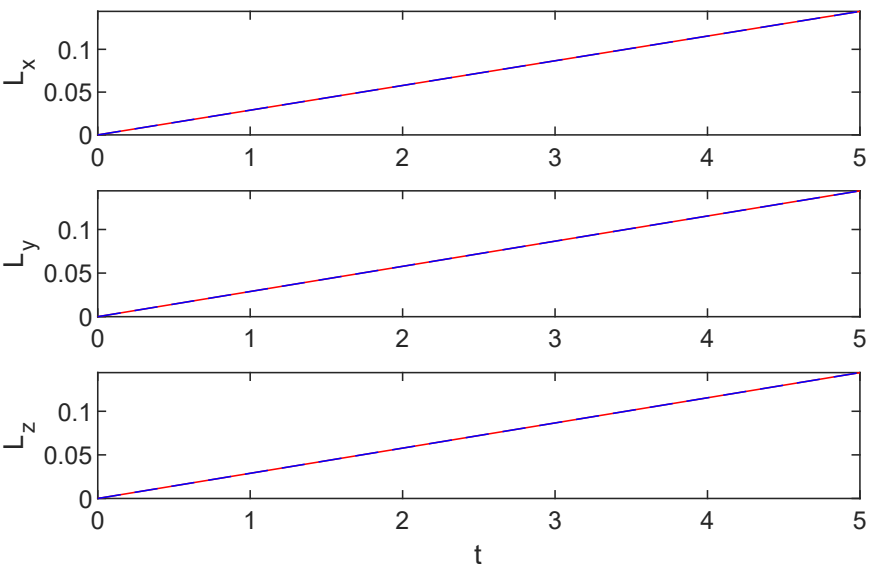


图 2: 验证角动量定理，三个图分别是角动量的三个分量，蓝线是力矩积分，红色是数值解中计算的角动量

```
tmin = 0; tmax = 5; % 起始时间
Nt = 501; % 时间步数
% =================

% 解微分方程
invI0 = inv(I0);
Y0 = [q0; w0];
f = @(Y,t)ode_fun(Y, t, I0, invI0, tau);
[Y, t] = odeRK4(f, [tmin, tmax], Y0, Nt);

% 验证角动量定理
verify(Y, t, I0, tau);

% 播放动画
figure; hold on; grid on; axis equal; view(23, 36);
axis([-0.3, 0.3, -0.3, 0.3, -0.3, 0.3]);
for it = 1:Nt
    q = Y(1:4, it);
    R = quat2mat(q);
```

```
    clf; hold on; grid on; axis equal; view(23, 36);
    axis([-0.3, 0.3, -0.3, 0.3, -0.3, 0.3]);
    plot_cube(R, a, b, c);
    title(['t = ' num2str(t(it), '%.3f')]);
    drawnow();
    % 取消注释可将每一帧保存为 png 图片（当前目录下）
    % saveas(gcf, [num2str(it) '.png']);
end
end

% 微分方程求导函数
function dY = ode_fun(t, Y, I0, invI0, tau)
q = Y(1:4); w = Y(5:7);
q = q / norm(q);
R = quat2mat(q);
RT = R';
dw = R*invI0*RT*(tau(t) - cross(w, R*I0*RT*w));
dq = 0.5*quat_mul([0; w], q);
dY = [dq; dw];
end

% 两个四元数相乘
function out = quat_mul(q1, q2)
s1 = q1(1); v1 = q1(2:4);
s2 = q2(1); v2 = q2(2:4);
out = [s1*s2 - dot(v1,v2); s1*v2 + s2*v1 + cross(v1, v2)];
end

% 由四元数 q 求旋转矩阵 R
function R = quat2mat(q)
s = q(1); vx = q(2); vy = q(3); vz = q(4);
R = [1 - 2*vy^2 - 2*vz^2, 2*vx*vy - 2*s*vz, 2*vx*vz + 2*s*vy;
    2*vx*vy + 2*s*vz, 1 - 2*vx^2 - 2*vz^2, 2*vy*vz - 2*s*vx;
    2*vx*vz - 2*s*vy, 2*vy*vz + 2*s*vx, 1 - 2*vx^2 - 2*vy^2];
end

```

```
% 画长方体（只用一条线）
function plot_cube(R, a, b, c)
x = [-1,  1,  1, -1, -1,  1,  1, -1] * a/2;
y = [-1, -1,  1,  1, -1, -1,  1,  1] * b/2;
z = [-1, -1, -1, -1,  1,  1,  1,  1] * c/2;
P = R * [x; y; z];
order = [1, 2, 3, 4, 1, 5, 6, 7, 8, 5, 6, 2, 3, 7, 8, 4];
plot3(P(1, order), P(2, order), P(3, order), 'b');
end

% 验证角动量定理
function verify(Y, t, I0, tau)
Nt = numel(t);
L = zeros(3, Nt); % 解的角动量
for it = 1:Nt
    q = Y(1:4, it); w = Y(5:7, it);
    R = quat2mat(q);
    L(:, it) = R*I0*R'*w;
end

L0 = zeros(3, Nt); % 力矩积分的角动量
L0(:, 1) = L(:, 1);
for it = 2:Nt
    L0(:, it) = L0(:, it-1) + ...
        integral(tau, t(it-1), t(it), 'ArrayValued', true);
end

figure;
for i = 1:3
    subplot(3, 1, i);
    plot(t, L(i, :), 'r'); hold on;
    plot(t, L0(i, :), 'b--');
end
subplot(3, 1, 1); ylabel L_x;
subplot(3, 1, 2); ylabel L_y;
subplot(3, 1, 3); ylabel L_z;
```

```
xlabel t;
end

% 四阶龙格库塔解微分方程 https://wuli.wiki/online/OdeRK4.html
function [Y,t]=odeRK4(f,tspan,Y0,Nt)
Nvar=numel(Y0);
dt=(tspan(2)-tspan(1))/(Nt-1);
Y=zeros(Nvar,Nt);
Y(:,1)=Y0(:);
t=linspace(tspan(1),tspan(2),Nt);
for ii=1:Nt-1
    K1=f( t(ii),         Y(:,ii)           );
    K2=f( t(ii)+dt/2,    Y(:,ii)+K1*dt/2   );
    K3=f( t(ii)+dt/2,    Y(:,ii)+K2*dt/2   );
    K4=f( t(ii)+dt,      Y(:,ii)+K3*dt     );
    Y(:,ii+1)=Y(:,ii)+dt/6*(K1+2*K2+2*K3+K4);
end
end
```

第九章

偏微分方程数值解

一维波动方程的数值解

预备知识 一维波动方程[1017]，Matlab 的判断与循环[1696]

[1]已知一维的波动方程为（式 5 [1017]）

$$\frac{\partial^2 y}{\partial x^2}-\frac{1}{c^2}\frac{\partial^2 y}{\partial t^2}=0 \tag{1}$$

其中 $y(x,t)$ 是坐标和时间的函数.这里介绍一个简单的**有限差分（finite difference）**法，即把空间 x 和时间 t 划分成等距离的网格 $x_1,\ldots,x_{Nx}$ 和 $t_1,\ldots,t_{Nt}$，步长分别为 Δx 和 Δt．我们将每个格点处的函数值记为 $y_{i,n}=y(x_i,t_n)$

有了网格以后，我们可以用有限差分表示二阶导数（式 5 [185]）得

$$\frac{y_{i-1,n}-2y_{i,n}+y_{i+1,n}}{\Delta x^2}-\frac{1}{c^2}\frac{y_{i,n-1}-2y_{i,n}+y_{i,n+1}}{\Delta t^2}=0 \tag{2}$$

整理得

$$y_{i,n+1}=2y_{i,n}-y_{i,n-1}+C^2(y_{i-1,n}-2y_{i,n}+y_{i+1,n}) \tag{3}$$

其中

$$C=c\frac{\Delta t}{\Delta x} \tag{4}$$

是一个无量纲得常数．也就是说，当我们已知 t_n 和 t_{n-1} 时刻的波函数，就可以得到 t_{n+1} 时刻的波函数.

边界条件

如果令网格边界处 $y=0$ （Dirichlet 边界条件）则会发生全反射且有半波损失．若边界处使用 $\partial y/\partial x=0$（Neumann 边界条件）同样会发生全反射但没有半波损失.

另外有一种很有用的边界条件叫 open boundary condition 或者 radiation boundary condition[2]，可以使波动完全不发生反射

$$\frac{\partial y}{\partial t}+c\frac{\partial y}{\partial x}=0 \tag{5}$$

要验证该条件，将任何速度为 c 的平面波代入发现都可以满足.

[1]本文参考了 BCB

[2]查看原文

Matlab 程序

程序如下，运行结果（动画）见 wuli.wiki/apps/wavBC.html．截图见图 1．

代码 1: wave1D.m

```
% 一维波动方程数值解
% 参考 https://wuli.wiki/online/W1dNum.html

function wave1D
close all;

% ==== 参数 ====
c = 1; % 波速
xmin = -4; xmax = 6; Nx = 1200; % 空间格点
tmin = 0; tmax = 12; Nt = 2401; % 时间格点
k = 6; Ncyc = 5; % 初始波包的波数和周期数
bc = 'd'; % 边界条件: [d] Dirichlet, [n] Neumann, [o] Open
% ================

x = linspace(xmin, xmax, Nx)';
t = linspace(tmin, tmax, Nt);
dx = (xmax - xmin)/(Nx - 1);
dt = (tmax - tmin)/(Nt - 1);
C = c*dt/dx; C2 = C*C;
y = zeros(Nx, Nt);
y(:,1) = y0(x, k, Ncyc);
y(:,2) = y0(x - c*dt, k, Ncyc);

% 二阶差分（边界元设为 0）
D2 = @(v) [0;
    v(1:end-2) - 2*v(2:end-1) + v(3:end);
    0];

figure;
for n = 2:Nt - 1
    y(:,n+1) = 2*y(:,n) - y(:,n-1) + C2*D2(y(:,n));
```

```
    y = bc_set(y, n+1, bc, c, dx, dt);
    if (mod(n, 8) == 0)
        clf; plot(x, y(:,n+1)); axis([xmin,xmax,-2,2]);
        hold on; scatter([xmin,xmax], [y(1,n+1), y(end,n+1)]);
        if bc == 'd'
            title(['Dirichlet B.C.  t = ', num2str(t(n+1), '%.2f')]);
        elseif bc == 'n'
            title(['Neumann B.C.  t = ', num2str(t(n+1), '%.2f')]);
        else
            title(['Open B.C.  t = ', num2str(t(n+1), '%.2f')]);
        end
        xlabel x; ylabel y;
        drawnow;
        % saveas(gcf, [bc 'wv' num2str(n) '.png']); % 保存图片文件
    end
end
end

% 初始波包
% sin^2 波形
function y = y0(x, k, Ncyc)
T = 2*pi/k;
L = T*Ncyc/2;
y = zeros(size(x));
k0 = k / Ncyc / 2;
for i = 1:numel(x)
    xx = x(i);
    if abs(xx) <= L
        y(i) = cos(k0*xx)^2 * sin(k*xx);
    end
end
end

```

```
% 处理边界值
function y = bc_set(y, n, bc, c, dx, dt)
if bc == 'd' % Dirichlet
    y(1, n) = 0;
    y(end, n) = 0;
elseif bc == 'n' % Neumann
    y(1, n) = y(2, n);
    y(end, n) = y(end-1, n);
elseif bc == 'o' % Open
    y(end, n) = y(end-1, n) - 1/c * (y(end-1, n) - y(end-1, n-1)
        )*dx/dt;
end
end
```

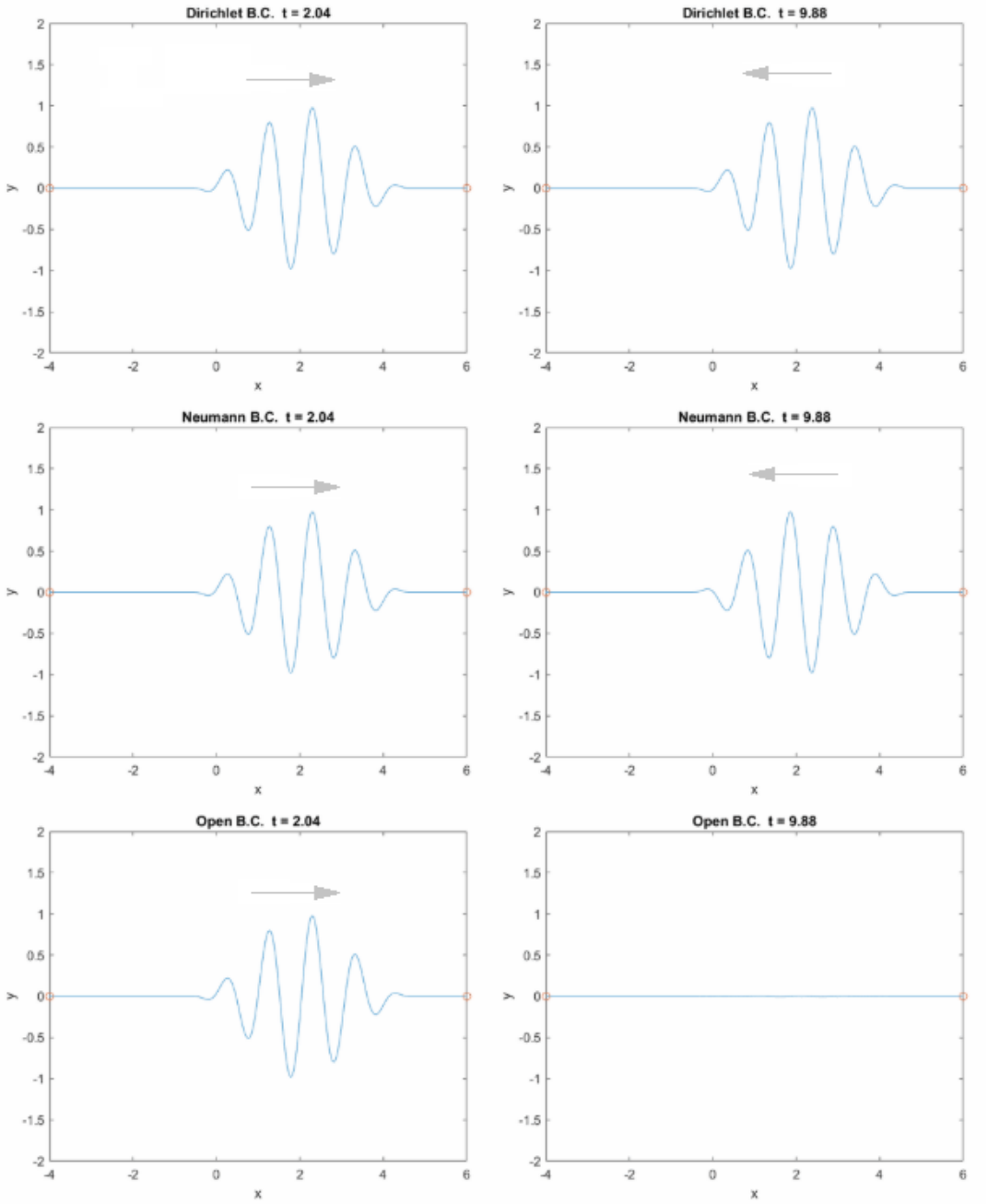


图 1: 运行结果截图，Dirichlet 边界反射后发生相位反转（半波损失），Neumann 边界条件反射后无半波损失，Open 边界条件无反射

第十章

一维薛定谔方程数值解

一维势能束缚态数值解（试射法）

[1]我们介绍用试射法解定态薛定谔方程（TISE，Time Independent Schrodinger Equation），其中 $V(x)$ 已知

$$-\frac{1}{2m}\psi(x)+V(x)\psi(x)=E\psi(x) \tag{1}$$

要使用 ODE 解算器（如 Matlab 的 ode45），需要先将上式化为一阶微分方程组

$$\begin{cases}\psi'(x)=v(x)\\ v'(x)=2m[V(x)-E]\psi(x)\end{cases} \tag{2}$$

对称势能

先讨论偶函数势能的情况．当 $V(x)$ 为偶函数时，定态波函数为奇函数或偶函数．解 ODE 应该从中点出发，前者可用初始条件 $\psi(0)=0$，$\psi'(0)=1$，后者可用初始条件 $\psi(0)=1$，$\psi'(0)=0$．这样解出来的波函数一般不满足归一化条件，需要在最后归一化．

现在可用定步长的欧拉法或者龙格库塔四阶法来解 $x\in[0,x_{max}]$ 的薛定谔方程．显然程序中不能选取 $x\in[0,+\infty]$，但是要保证 x_{max} 足够大，使解出薛定谔方程后有 $\psi(x_{max})\approx 0$，$\psi'(x_{max})\approx 0$．

初始条件

如果势能不是对称的，我们一般从其中一端作为起始点解微分方程．若某个小区间 $V(x)$ 看做常数，解为

$$\psi(x)=\begin{cases}\sin(kx) & (E>V)\\ \exp(\kappa x) & (E<V)\end{cases} \tag{3}$$

其中

$$k=\sqrt{2m(E-V)}\qquad \kappa=\sqrt{2m(V-E)} \tag{4}$$

[1]本词条使用原子单位

所以 ODE 初始条件选择

$$(\psi(x_0),\psi'(x_0))=\begin{cases}(0,1) & (E>V(x_0))\\ (\epsilon,\kappa\epsilon) & (E<V(x_0))\end{cases} \tag{5}$$

其中 ϵ 是一个很小的数．第二种情况中，解析初始条件应该是 $(0,0)$，但数值解中显然不能用这个．注意最后归一化以后要检验是否两端都满足 $\psi(x_0)\ll 1$．

甩尾法

如何求得本征值 E 呢？一种简单的方法是试射法．顾名思义，取不同离散的 E，用一定的条件判断对这些 E 波函数在 x_{max} 处是否满足边界条件 $\psi(x_{max})\approx 0$[2]．

一般来说，若 E 略大于某本征值 E_n 时，会有 $\psi(x_{max})>0$，略小时会有 $\psi(x_{max})>0$．所以画出 $\psi(x_{max})$ 关于 E 的图像，用目测法找到零点即可．若需要更精确的本征值，可选取一个更小的区间，再次画图．

中点匹配

从数值误差来说，从指数末端出射时误差远远大于从指数末端入射．所以更好的办法是从两端 x_L,x_R 分别入射后在某个中点 x_M 处匹配．

简单的匹配可以采用**对数导数（log derivative）**来计算匹配误差，即

$$\mathrm{err}(E)=\frac{\psi_L'(x_M)}{\psi_L(x_M)}-\frac{\psi_R'(x_M)}{\psi_R(x_M)} \tag{6}$$

用多区间二分法即可得到定态能量．

但是这么做有两个缺点，一是 $\psi(x_M)$ 为零或很小的时候，err 有可能不稳定，二是用多区间二分法解 $\mathrm{err}(E)=0$ 有可能出现不收敛的解，因为该函数存在断点（想象 $\sin(kx_M)$ 的对数导数随 k 变化的情况）．

为了解决这个问题，我们可以构造一个性质更好的函数来代替对数导数，即用 θ 表示函数 $\sin x$ 在 $x=\theta$ 处的对数导数

$$\theta=\mathrm{Arctan}(y,y') \tag{7}$$

[2]如果 x_{max} 足够大，只需满足这一条件即可自动满足 $\psi'(x_{max})\approx 0$

注意对数导数是关于 θ 的周期函数，周期为 π．只要保证两个 θ 相同或相差 $n\pi$ 就可以保证对数导数相同．

要构造关于 θ 的误差函数，我们可以先计算两个 θ 之差

$$\theta = \theta_2 - \theta_1 \qquad (\theta \in [-2\pi, 2\pi]) \tag{8}$$

误差函数 $\mathrm{err}(\theta)$ 满足一些条件：

1. $\theta = n\pi$ 为零点（误差函数为零当且仅当对数导数相等）
2. 零点两侧异号（便于用二分法或类似算法求根）
3. 函数连续（连续性自然是好的）
4. 周期为 2π （Arctan 函数会使 θ 在 0 到 2π 之间突变，这样可以保证误差函数连续）

一个显然的误差函数是满足这些条件的三角波（也可以用正弦函数，但是三角波的计算量要小得多）．

现在再用多区间二分法[1843] 就可以解出所有正确的束缚态能量了．

检查漏解

用多区间二分法时，如果区间划分不够细，使一个区间中存在多个解，就会产生漏解．检查是否漏解可以用波函数节点的数量来判断．如果波函数的节点数是逐个递增的，则没有漏解．也就是说第 n 个能级的波函数需要有 $n-1$ 个节点（不包括两端）．

Matlab 代码

完整的代码（bnd_shoot）见下文，作为一个使用的例子，我们来解一维简谐振子[1344]．

代码 1: bndShootDemo.m

```
close all;
% =====  设置参数  ========
V = @(x) 0.5*x.*x; % 势能
```

```
mass = 1; % 质量
xmin = -10; xmid = 0; xmax = 10; % 端点和中点
Espan = [0.1, 10]; % 能量区间
Eresolution = 100; % 二分法区间数
NstepMin = 1000; % ode 解算器最少步数
% ========================

% 画势能曲线
x = linspace(xmin, xmax, 1000);
figure; plot(x, V(x));
xlabel x; title V;

[Eng, X, Psi] = bndShoot(V, xmin, xmid, xmax, mass, Espan, ....
    Eresolution, odeset('RelTol',1e-6, 'MaxStep',abs(xmax-xmin)/
       NstepMin),...
    true);
```

结果如下

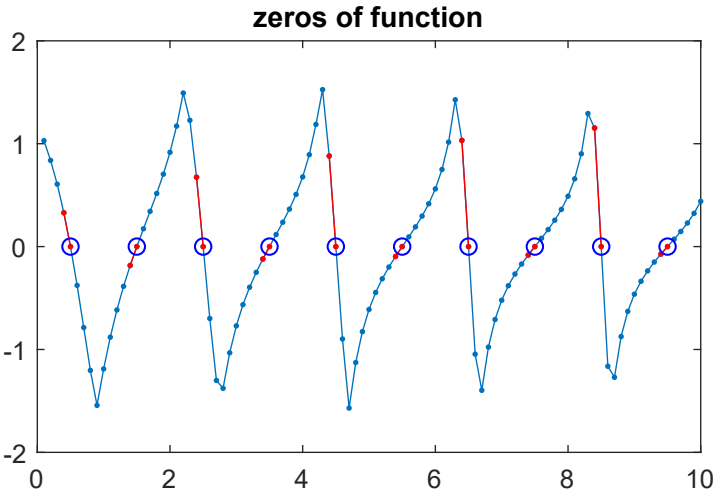


图 1: 误差函数零点

代码 2: bndShoot.m

```
% 求解一维定态薛定谔方程的能级
% 从两端试射
```

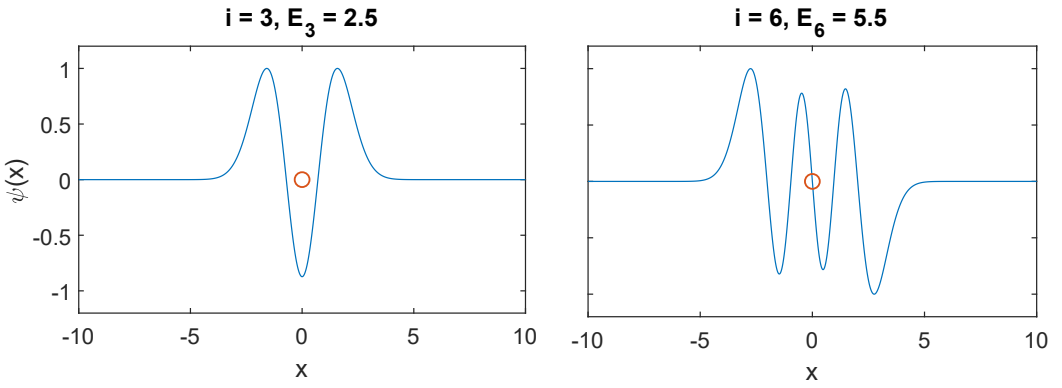


图 2: 解出的波函数，这里只给出了第 3 和第 6 个能级，红点表示左右波函数匹配的位置

```
% varargin 是 ode45 解算器的选项
% Eng(i) 是第 i 个能级的能量
% psi(:,i) 是第 i 个能级的波函数
function [Eng, X, Psi] = bndShoot(V, xmin, xmid, xmax, mass, ...
    Espan, EResolution, odeOpt, plot_flag)

% 多区间二分法解出能级和波函数
trial_fun = @(E) bound_shooting_trial(E, V, xmin, xmid, xmax,
   mass, odeOpt);
Eng = fzeroN(trial_fun, Espan, EResolution);

% 输出,  检查节点个数
X = cell(1,numel(Eng)); Psi = X;
for i = 1:numel(Eng)
    [~, X{i}, Psi{i}] = trial_fun(Eng(i));
    if numzero(Psi{i}) > i - 1
        warning(['missed bound state ', num2str(i),...
            ', increase Eresolution!']);
    elseif numzero(Psi{i}) < i - 1
        error('duplicate bound state found!');
    end

    if plot_flag
        x = X{i}; psi = Psi{i};
```

```
        figure; plot(x,psi); axis([x(1),x(end),-1.2,1.2]);
        hold on;
        xlabel('x'); ylabel('\psi(x)');
        Nzeros = numzero(psi);
        title(['i = ', num2str(i),...
            ', E_{', num2str(Nzeros+1), '} = ',  num2str(Eng(i))
               ]);
        scatter(xmid,0);
    end
end
end

% 单个能量试射
% 输出误差函数和其他信息
function [err, x, psi, th1, th2] = bound_shooting_trial(E, V,
    ...
     xmin, xmid, xmax, mass, odeOpt)
dy = 1e-16; % dy = psi(xmin) 和 psi(xmax)
[x1, Y1] = bound_shooting_left(E, V, [xmin,xmid], mass, dy,
    odeOpt);
[x2, Y2] = bound_shooting_right(E, V, [xmax,xmid], mass, dy,
    odeOpt);

% 误差
[err, th1, th2] = fmatch_err(Y1(end,1),Y1(end,2),Y2(end,1),Y2(
    end,2));

psi1_max = max(abs(Y1(:,1)));
psi2_max = max(abs(Y2(:,1)));

% 波函数 (max(psi) = 1)
if nargout > 1
    x = [x1; flip(x2(1:end-1))].';
end
if nargout > 2
    v1 = Y1(end,:);
```

```
    v2 = Y2(end,:);
    scale = sign(dot(v1,v2))*norm(v1)/norm(v2);
    if psi1_max >= psi2_max * abs(scale)
        psi = [
            Y1(:,1).' / psi1_max,  .....
            flip(Y2(1:end-1,1)).' * (scale / psi1_max)];
    else
        psi = [
            Y1(:,1).' / (scale * psi2_max),  .....
            flip(Y2(1:end-1,1)).' / psi2_max];
    end
end
end

% 从左边入射（使用 ode45 解算器）
function [x, Y] = bound_shooting_left(E, V, xspan, mass, dy,
   odeOpt)
if V(xspan(1)) > E
    Y0 = [1; sqrt(2*mass*(V(xspan(1)) - E))] * dy;
else
    warning('起始点 E > V,  确定吗?');
    Y0 = [0; 1];
end
odefun = @(x,Y) TISE_odefun(x,Y,E,V,mass);
[x,Y] = ode45(odefun, xspan, Y0, odeOpt);
end

% 从右边入射
function [x, Y] = bound_shooting_right(E, V, xspan, mass, dy,
   odeOpt)
if V(xspan(1)) > E
    Y0 = [1; -sqrt(2*mass*(V(xspan(1)) - E))] * dy;
else
    warning('起始点 E > V,  确定吗?');
    Y0 = [0; 1];
end
```

```
odefun = @(x,Y) TISE_odefun(x,Y,E,V,mass);
[x,Y] = ode45(odefun, xspan, Y0, odeOpt);
end

% ode 函数
% 由 Y = [psi(x); psi'(x)] 求 dY(x) = [psi'(x), psi''(x)]
function dY =TISE_odefun(x,Y,E,V,mass)
dY(2,1) = 2*mass*(V(x)-E)*Y(1);
dY(1) = Y(2);
end

% 多区间二分法求函数零点
function roots = fzeroN(fun,interval,resolution,converge,options
    )
if ~exist('converge','var') || isempty(converge)
    converge = false;
end
if ~exist('options','var')
    options = optimset;
end
x=linspace(interval(1),interval(end), resolution);
y=arrayfun(fun,x);
figure; plot(x,y,'.-'); hold on;
title('the function to be found zeros of');
[Nind, ind] = numzero(y);

roots=zeros(1, Nind); kk = 0;
for ii=1:Nind
    plot([x(ind(ii)),x(ind(ii)+1)], ...
        [fun(x(ind(ii))),fun(x(ind(ii)+1))], '.-r');
    root = fzero(fun,[x(ind(ii)),x(ind(ii)+1)], options);
    if ~converge || converge && iszero(fun,root,options.TolX)
        kk = kk + 1;
        roots(kk) = root;
        scatter(root, fun(root), 'b');
    end
```

```
end
roots = roots(1:kk);
end

% 计算两函数在某点匹配的误差
function [err, th1, th2] = fmatch_err(y1,dy1,y2,dy2)
th1 = atan2(y1,dy1);
th2 = atan2(y2,dy2);
err = mod(th2 - th1, 2*pi);
if err > 0.5*pi && err <= 1.5*pi
    err = pi - err;
elseif err > 1.5*pi
    err = -2*pi + err;
end
end

% 数函数的零点
function [N, ind] = numzero(y)
Sign=sign(y);
ind=find(Sign(1:end-1).*Sign(2:end)<0);
N = numel(ind);
end
```

无限深势阱中的高斯波包 Matlab 代码

- 本词条处于草稿阶段.

预备知识　无限深势阱中的高斯波包[1337]，Matlab 的程序调试及其他功能[1710]

代码 1: WpkISW.m

```
% 无限深势阱中的波包

```

```
close all; clear; % 关闭所有画图，  清空所有变量

% === 参数 ===
L = 100; % 势阱区间 [0, L]
sig_x = 4; % 波包宽度
x0 = 50; % 波包初始位置
k0 = 2; % 波包初始动量
Nbasis = 100; % 基底数量
Nx = 501; % 画图格点数
tmax = 100; % 演化时间 [0, tmax]
Nt = 201; % 演化步数
% ============

sig_k = 1/(2*sig_x); % 动量谱宽度
Ax = 1/(2*pi*sig_x^2)^0.25; % x 归一化系数
Ak = 1/(2*pi*sig_k^2)^0.25; % k 归一化系数
% x 和 k 的波函数
f = @(x) Ax * exp(-((x - x0)/(2*sig_x)).^2) .* exp(1i*k0*x);
g = @(k) Ak * exp(-((k - k0)/(2*sig_k)).^2) .* exp(-1i*x0*(k-k0)
    );
% 位置，时间动量格点
x = linspace(0, L, Nx);
t = linspace(0, tmax, Nt);
% 初始波包投影系数
k = (1:Nbasis) * pi / L;
coeff = 1i*sqrt(pi/L)*(g(k) - g(-k));
% 画图判断系数是否收敛
coeff2 = abs(coeff).^2;
figure; plot(coeff2);
title(['sum(|coeff|^2) = ' num2str(sum(coeff2))]);

figure;
for it = 1:Nt
    % 计算 t(it) 时刻的波函数
    psi = zeros(1, Nx);
    for i = 1:Nbasis
```

```
        Eng = 0.5*k(i)^2;
        psi = psi + coeff(i) * exp(-1i*Eng*t(it)) * sqrt(2/L) *
            sin(i*pi*x/L);
    end
    % 画出波函数实部和虚部
    clf; plot(x, real(psi));
    hold on; plot(x, imag(psi));
    xlabel x; ylabel \Psi(x);
    title(['t = ' num2str(t(it), '%.1f')]);
    axis([0, L, -0.5, 0.5]);
    drawnow;
    % 取消注释可将每一帧保存为 png 图片（当前目录下）
    saveas(gcf, [num2str(it) '.png']);
end
```

方势垒定态波函数程序

预备知识　方势垒[1365]

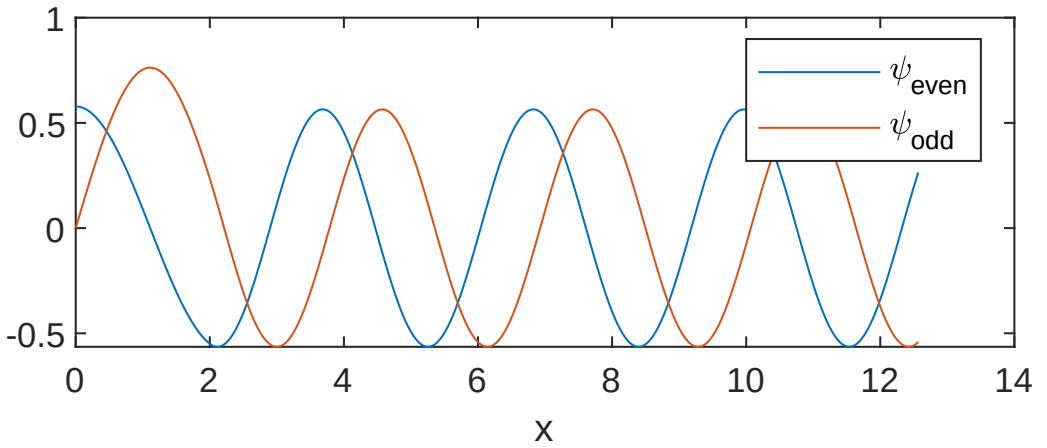


图 1: FSB.m 画图结果

代码 1: FSB.m

```
% m 质量, V0 势垒, E 能量, L 势垒半宽, odd 奇函数 true 偶函数 false
% A 中间区间系数, C 右边区间 cos 系数, D 右边区间 sin 系数

function [psi,A,C,D] = FSB(x, E,L,m,V0,odd)
k = sqrt(2*m*E);
psi = zeros(size(x));
m1 = (x < -L); m2 = (abs(x) <= L); m3 = (x > L);
if (E == V0), E = V0 + eps; end

if E >= V0
    b = sqrt(2*m*(E - V0));
    if ~odd
        A1 = 1/sqrt(pi*(cos(b*L)^2+b^2*sin(b*L)^2/k^2));
        C1 = A1*(cos(k*L)*cos(b*L) + b/k*sin(k*L)*sin(b*L));
        D1 = A1*(-b/k*cos(k*L)*sin(b*L) + sin(k*L)*cos(b*L));
        A = A1; C = C1; D = D1;
        psi(m1) = C1*cos(k*x(m1)) - D1*sin(k*x(m1));
        psi(m2) = A1*cos(b*x(m2));
        psi(m3) = C1*cos(k*x(m3)) + D1*sin(k*x(m3));
    else
        B2 = 1/sqrt(pi*(sin(b*L)^2+b^2*cos(b*L)^2/k^2));
        C2 = B2*(cos(k*L)*sin(b*L) - b/k*sin(k*L)*cos(b*L));
        D2 = B2*(b/k*cos(k*L)*cos(b*L) + sin(k*L) * sin(b*L));
        A = B2; C = C2; D = D2;
        psi(m1) =-C2*cos(k*x(m1)) + D2*sin(k*x(m1));
        psi(m2) = B2*sin(b*x(m2));
        psi(m3) = C2*cos(k*x(m3)) + D2*sin(k*x(m3));
    end
else
    ka = sqrt(2*m*(V0-E));
    if ~odd
        A1 = 1/sqrt(pi*(cosh(ka*L)^2+(ka/k)^2*sinh(k*L)^2));
        C1 = A1*(cosh(ka*L)*cos(k*L) - ka/k*sinh(ka*L)*sin(k*L))
            ;
        D1 = A1*(cosh(ka*L)*sin(k*L) + ka/k*sinh(ka*L)*cos(k*L))
```

```
            ;
        A = A1; C = C1; D = D1;
        psi(m1) = C1*cos(k*x(m1)) - D1*sin(k*x(m1));
        psi(m2) = A1*cosh(ka*x(m2));
        psi(m3) = C1*cos(k*x(m3)) + D1*sin(k*x(m3));
    else
        B2 = 1/sqrt(pi*(cosh(ka*L)^2+(ka/k)^2*sinh(k*L)^2));
        C2 = B2*(sinh(ka*L)*cos(k*L) - ka/k*cosh(ka*L)*sin(k*L))
            ;
        D2 = B2*(sinh(ka*L)*sin(k*L) + ka/k*cosh(ka*L)*cos(k*L))
            ;
        A = B2; C = C2; D = D2;
        psi(m1) =-C2*cos(k*x(m1)) + D2*sin(k*x(m1));
        psi(m2) = B2*sinh(ka*x(m2));
        psi(m3) = C2*cos(k*x(m3)) + D2*sin(k*x(m3));
    end
end
end
```

画图

```
% === 设置参数 =====
m = 1; V0 = 1; E = 2; L = 2;
% =================
x = linspace(-5, 5, 1000);
psi_e = FSB(x, E, L, m, V0, 0);
psi_o = FSB(x, E, L, m, V0, 1);
figure; plot(x, real(psi_e)); hold on; plot(x, real(psi_o));
xlabel x; legend('\psi_{even}', '\psi_{odd}');
```

结果如图 1.

行波

代码 2: FSB2.m

```
% m 质量, V0 势垒, E 能量, L 势垒半宽, odd 奇函数 true 偶函数
   false
```

```
% A,B 左区间系数， C,D 中区间系数， E,F 右区间
% pm = 1: 从左入射; pm = -1: 从右边入射
function [psi, A,B,C,D,E,F] = FSB2(x, E, L, m, V0, pm)
k = sqrt(2*m*E);
psi = zeros(size(x));
m1 = (x < -L); m2 = (abs(x) <= L); m3 = (x > L);
if (E == V0), E = V0 + eps; end
if E >= V0
    b = sqrt(2*m*(E-V0));
    N = ((k^2+b^2)*sin(2*b*L)-2i*k*b*cos(2*b*L))*exp(-1i*k*L)...
         /sqrt(2*pi)/((k^2-b^2)^2*sin(2*b*L)^2 + 4*k^2*b^2);
    A = 1/sqrt(2*pi);
    B = N*(k^2-b^2)*sin(2*b*L)*exp(-1i*k*L);
    C = N*1i*k*(b+k)*exp(-1i*b*L);
    D = N*1i*k*(b-k)*exp(1i*b*L);
    E = N*2i*k*b*exp(-1i*k*L);
    F = 0;
    if pm < 0
        tmp = A; A = F; F = tmp;
        tmp = B; B = E; E = tmp;
        tmp = C; C = D; D = tmp;
    end
    psi(m1) = A*exp(1i*k*x(m1)) + B*exp(-1i*k*x(m1));
    psi(m2) = C*exp(1i*b*x(m2)) + D*exp(-1i*b*x(m2));
    psi(m3) = E*exp(1i*k*x(m3)) + F*exp(-1i*k*x(m3));
else
    ka = sqrt(2*m*(V0-E));
    N = ((k^2-ka^2)*sinh(2*ka*L)-2i*k*ka*cosh(2*ka*L))*exp(-1i*k
       *L)...
         /sqrt(2*pi)/((k^2+ka^2)^2*sinh(2*ka*L)^2 + 4*k^2*ka^2);
    A = 1/sqrt(2*pi);
    B = N*(k^2+ka^2)*sinh(2*ka*L)*exp(-1i*k*L);
    C = N*k*(1i*ka-k)*exp(-ka*L);
    D = N*k*(1i*ka+k)*exp(ka*L);
    E = N*2i*k*ka*exp(-1i*k*L);
    F = 0;
```

```
    if pm < 0
        tmp = A; A = F; F = tmp;
        tmp = B; B = E; E = tmp;
        tmp = C; C = D; D = tmp;
    end
    psi(m1) = A*exp(1i*k*x(m1)) + B*exp(-1i*k*x(m1));
    psi(m2) = C*exp(ka*x(m2)) + D*exp(-ka*x(m2));
    psi(m3) = E*exp(1i*k*x(m3)) + F*exp(-1i*k*x(m3));
end
end
```

画图

```
% === 设置参数 =====
m = 1; V0 = 0.9; E = 1.1; L = 2;
% =================
x = linspace(-5, 5, 1000);
psi = FSB2(x, E, L, m, V0, 1);
figure; plot(x, real(psi)); hold on; plot(x, imag(psi));
xlabel x;
```

第十一章

氢原子薛定谔方程数值解

氢原子波函数画图程序

- 本词条处于草稿阶段.

运行结果见图 1 [1536]. 需要“球谐函数数值计算[1924]”中的 SphHarm 函数. 以及 surfSph 函数.

```
Z = 1;
Rnl = @(Z,n,l,r)sqrt((2*Z/n)^3*factorial(n-l-1)/(2*n*factorial(n+l))) *...
      (2*Z*r/n).^l .* laguerreL(n-l-1, 2*l+1, 2*Z*r/n) .* exp(-Z*r/n);

r = linspace(0, 40, 1000);
th = linspace(0, 2*pi, 361);
ph = zeros(size(th));
thresh = [0.3183 7.28e-4 1e-4 1e-5 1e-6];
for n = 4
    for l = 0:n-1
        for m = 0 : l
            figure; set(gcf, 'units','normalized', ...
                'outerposition',[0 0 0.5 1]);
            set(gca, 'color', 'k');
            grid off; hold on;
            r2R2 = @(r) r.^2 .* abs(Rnl(Z,n,l,r)).^2;
            trial = @(rmax) integral(r2R2, 0, rmax) - 0.95;
            r_cut = fzero(trial, [0.1, 100]);
            tmp = linspace(0, 2*pi, 500);
            plot3(r_cut*cos(tmp), -3*ones(size(tmp)), ...
                r_cut*sin(tmp), 'w');
            prob_cut = abs(Rnl(Z,n,l,r_cut))^2 * 5;

            prob = abs(Rnl(Z,n,l,r).' .* SphHarm(l,m,th,ph)).^2;
            surfSph(r, th, 0, prob); view(0, 0);
            caxis([0, prob_cut]);
```

```
            title([num2str(n) ', ' num2str(l) ', ' num2str(m)]);
            disp([num2str(n) ', ' num2str(l) ', ' num2str(m)]);
            % 未完成: colormap_hydrogen, 使用默认 colormap
            % rgb = colormap_hydrogen(linspace(0,1,100));
            % colormap(rgb);
        end
    end
end
```

球谐函数数值计算

预备知识 球谐函数[761]

目前Matlab并不直接提供球谐函数,而只提供连带勒让德多项式(legendre).以下给出 Matlab 中计算球谐函数的程序.

代码 1: Ylm.m

```
% 计算球谐函数 Y_{lm}
% 使用 Condon-Shortley 相位(和 Mathematica 相同)
% l, m 必须是整数标量
% 角度 Th 和 Ph 可以是任意实数
% Th 和 Ph 可以是矩阵, 尺寸必须相同, 输出 Y_lm 的尺寸也相同
function Y_lm = SphHarm(l,m,Th,Ph)
mark = mod(Th,2*pi) > pi;
Ph(mark) = Ph(mark) + pi;
Sign = sign(m);
m = abs(m);
P_lm = legendre(l,cos(Th(:)));
P_lm = P_lm(m+1,:);

% 归一化因子
Y_lm = zeros(size(Th));
N_lm = sqrt((2*l+1)*factorial(l-m)./(4*pi*factorial(l+m)));
if Sign >= 0
```

```
    Y_lm(:) = N_lm * P_lm' .* exp(1i*m*Ph(:));
else
    Y_lm(:) = (-1)^m * N_lm * P_lm' .* exp(-1i*m*Ph(:));
end
end
```

Crank-Nicolson 算法（一维）

薛定谔方程为

$$-\frac{1}{2}\frac{\partial^2\psi}{\partial x^2}+V\psi=\mathrm{i}\frac{\partial\psi}{\partial t} \tag{1}$$

用 Crank-Nicolson 或 Caley scheme[1] 得到的结果是

$$\left(1+\frac{\mathrm{i}}{2}\mathbf{H}\Delta t\right)\psi^{n+1}=\left(1-\frac{\mathrm{i}}{2}\mathbf{H}\Delta t\right)\psi^{n} \tag{2}$$

其中二阶导数用三点差分计算，得

$$\begin{aligned}\psi_i^{n+1}-\psi_i^n=&\frac{\mathrm{i}\Delta t}{4\Delta x^2}(\psi_{i-1}^n-2\psi_i^n+\psi_{i+1}^n+\psi_{i-1}^{n+1}-2\psi_i^{n+1}+\psi_{i+1}^{n+1})\\&-\frac{\mathrm{i}\Delta t}{2}(V_i^n\psi_i^n+V_i^{n+1}\psi_i^{n+1})\end{aligned} \tag{3}$$

令 $\alpha=\mathrm{i}\Delta t/(4\Delta x^2),\beta=\mathrm{i}\Delta t/2$，整理可得

$$\begin{aligned}&-\alpha\psi_{i-1}^{n+1}+(1+2\alpha+\beta V_i^{n+1})\psi_i^{n+1}-\alpha\psi_{i+1}^{n+1}\\=&\alpha\psi_{i-1}^n+(1-2\alpha-\beta V_i^n)\psi_i^n+\alpha\psi_{i+1}^n\end{aligned} \tag{4}$$

其中 $\psi_i^n=\psi(x_i,t_n)$，$V_i^n=V(x_i,t_n)$.

我们把一个区间划分成 N_x-1 段等长的区间，并令 N_x 个格点为 $x_1\dots x_{N_x}$. 最简单的边界条件是取 $\psi(x_1)=\psi(x_{N_x})=0$. 这样式 4 中的 i 可以取 $i=2\dots N_x-1$，得到 N_x-2 条式子，其中只有 $\psi_2^{n+1}\dots\psi_{N_x-1}^{n+1}$ 这 N_x-2 个未知量，每条式子最多包含连续 3 个未知量. 将线性方程用矩阵表示，就可以得到一个三对角矩阵（第一行和最后一行只有两个系数）.

但事实上，还可以继续减少计算量. 将式 2 整理后得

$$\left(\frac{1}{2}+\frac{\mathrm{i}}{4}\mathbf{H}\Delta t\right)\left(\psi^{n+1}+\psi^n\right)=\psi^n \tag{5}$$

解这个方程，再减去 ψ^n 即可.

[1] 二者是一回事，见 Numerical Recipes 19.2

虚时间

使用虚时间后，式 2 和式 5 分别变为

$$\left(1+\frac{1}{2}\mathbf{H}\Delta t\right)\psi^{n+1}=\left(1-\frac{1}{2}\mathbf{H}\Delta t\right)\psi^{n} \tag{6}$$

$$\left(\frac{1}{2}+\frac{1}{4}\mathbf{H}\Delta t\right)\left(\psi^{n+1}+\psi^{n}\right)=\psi^{n} \tag{7}$$

（代码未完成）

Gauss-Lobatto 积分

- 本词条需要更多讲解，便于帮助理解.

预备知识 定积分[198]

高斯积分（Gauss quadrature）可以用求和来近似表示积分

$$\int f(x)\,\mathrm{d}x\approx\sum_{i}w_{i}f(x_{i}) \tag{1}$$

对于某区间的一组 x_i, w_i，如果它是 N 阶的，那么当 $f(x)$ 是小于等于 N 阶的多项式时，上式精确成立.

Gauss-Lobatto 积分

Gauss-Lobatto 积分中令采样点（包括两个端点）的个数为 N，如果被积函数是 $2N-3$ 阶多项式（$f(x)=x^{2n-3}+\dots$），则积分没有误差.

注意 Gauss-Lobatto 积分是对称的

$$x_j=-x_{N-j+1}\qquad w_{0j}=w_{0,N-j+1} \tag{2}$$

正交归一基底

每个基底都是 $N-1$ 阶多项式，由于阶数和零点数一样，多项式可以唯一确定，即拉格朗日插值多项式

$$\begin{aligned} p_n(x) &= \prod_{i=1}^{n-1} \frac{x-x_i}{x_n-x_i} \prod_{i=n+1}^{N} \frac{x-x_i}{x_n-x_i} \\ &= \frac{x-x_1}{x_n-x_1} \times \cdots \times \frac{x-x_{n-1}}{x_n-x_{n-1}} \frac{x-x-n+1}{x_n-x_{n+1}} \cdots \frac{x-x_N}{x_n-x_N} \end{aligned} \tag{3}$$

$$p_n(x_{n'}) = \delta_{n,n'} \tag{4}$$

由于任意两个基底乘积只是 $2N-2$ 阶的多项式，所以可以用求和精确代替其积分.

$$\int_{-1}^{1} p_i(x)p_j(x)\,\mathrm{d}x = \sum_k w_k p_i(x_k)p_j(x_k) = w_i\delta_{ij} \tag{5}$$

所以 N 个正交归一基底就是

$$f_n(x) = \frac{1}{\sqrt{w_n}} p_n(x) \tag{6}$$

满足

$$f_i(x_j) = \frac{1}{\sqrt{w_i}}\delta_{ij} \tag{7}$$

另一种等效的表示方法是利用 N 阶 Gauss-Lobatto 数值积分对应的多项式，$N-1$ 阶勒让德多项式的导数，$P'_{N-1}(y)$，来构建满足条件的多项式. 根据定义，其 $N-2$ 个零点分别为 $y_2, y_3 \ldots y_{N-1}$，为了加入 $y_1=-1$ 与 $y_n=1$ 这两个零点，将其变为 N 阶多项式

$$(1-y^2)P'_{N-1}(y) \tag{8}$$

然而，式 2 要求 $p_n(y_n)=1$，所以我们将式 8 除以它自己在 y_n 处的切线，在 $y=y_n$ 处形成极限类型 $0/0=1$ 即可得到阶多项式 $u_n(y)$.

$$u_n(y) = \frac{(1-y^2)P'_{N-1}(y)}{[(1-y^2)P'_{N-1}(y)]'_y} = y_n(y-y_n) \tag{9}$$

该式与式 3 事实上是完全相同的多项式，因为所有具有 $N-1$ 个零点的 $N-1$ 阶多项式都可以因式分解成式 3 的形式乘以一个待定常数. 用式 9 便于快速地展开多项式（因为勒让德多项式的系数可以直接通过公式计算）.

FEDVR 网格

预备知识 高斯积分（Gauss-Lobatto）[1926]

FEDVR 基底

现在以一维 FEDVR 为例，把整个区间划分成 N_e 个**有限元**，第 i 个有限元的区间为 $[x_i, x_{i+1}]$．每个有限元内进一步加入格点，令 x_{ij} $(j = 1 \ldots N)$ 为第 i 有限元的 N 阶 Gauss-Lobatto 数值积分的 N 个采样点．这样，$x_{j1} = x_j$，$x_{jN} = x_{j+1}$．

现在来定义 FEDVR 基底 $u_{ij}(x)$．令 a_i 是第 i 个 FE 长度的一半，b_i 是第 i 个 FE 的中点坐标，把式 6 [1926] 中的 $f_n(x)$（x 超出 $[-1,1]$ 时令 $f_n(x) = 0$）记为 $f_n(t)$，并把 $t \in [-1,1]$ 线性依次映射到每个有限元的区间 $x \in [x_i, x_{i+1}]$ 上，有

$$x_{ij} = a_i t_j + b_i \tag{1}$$

$$x = a_i t + b_i \qquad (x_{i1} \leqslant x \leqslant x_{iN}) \tag{2}$$

则基底的定义为

$$u_{ij}(x) = \begin{cases} \dfrac{1}{\sqrt{a_i}} f_j\left(\dfrac{x - b_i}{a_i}\right) & (1 < j < N) \\ \dfrac{1}{\sqrt{a_i + a_{i+1}}}\left[f_N\left(\dfrac{x - b_i}{a_i}\right) + f_1\left(\dfrac{x - b_{i+1}}{a_{i+1}}\right)\right] & (j = N) \end{cases} \tag{3}$$

易证正交归一关系

$$\int_{x_1}^{x_{N_e+1}} u_{ij}(x) u_{i'j'}(x)\,\mathrm{d}x = \delta_{ii'}\delta_{jj'} \tag{4}$$

在总区间的两端我们一般采用函数值为零的边界条件，所以不定义 $u_{11}(x)$ 和 $u_{N_e,N}(x)$．这样，我们最终共有 $N_e(N-1) - 1$ 个节点 x_{ij} 和对应的基底 $u_{ij}(x)$．令 w_{0i} 为式 6 [1926] 中的 w_i，则每个基底在每个节点处的函数值为

$$u_{ij}(x_{i'j'}) = \delta_{ii'}\delta_{jj'} \times \begin{cases} \dfrac{1}{\sqrt{a_i w_{0j}}} & (1 < j < N) \\ \dfrac{1}{\sqrt{(a_i + a_{i+1}) w_{0N}}} & (j = N) \end{cases} \tag{5}$$

易证，若要对整个区间积分，所需权重为

$$w_{ij}=\frac{1}{u_{ij}^2(x_{ij})}=\begin{cases}a_i w_{0j} & (1<j<N)\\ (a_i+a_{i+1})w_{0N} & (j=N)\end{cases} \tag{6}$$

但如果只需要在某个 FE 区间内做积分，只需用 $w_{ij}=a_i w_{0j}$ 即可.

一个函数 $f(x)$ 用基底展开为

$$\langle u_{ij}(x)|f(x)\rangle=\int_{x_1}^{x_{N_e+1}}u_{ij}(x)f(x)\,\mathrm{d}x\approx\sqrt{w_{ij}}f(x_{ij}) \tag{7}$$

把所有基底按对应的 x_{ij} 从小到大排序. $f(x)$ 的展开系数就可以记为一个列矢量.

动能算符

使用分部积分可以把动能矩阵元表示为两个基底的导函数的内积[2]

$$\begin{aligned}K_{ij}&=-\frac{1}{2}\langle u_i|\frac{\mathrm{d}^2}{\mathrm{d}x^2}|u_j\rangle=-\frac{1}{2}\int u_i(x)u_j''(x)\,\mathrm{d}x\\&=-\frac{1}{2}\,u_i(x)u_j'(x)\Big|_{x_1}^{x_2}+\frac{1}{2}\int u_i'(x)u_j'(x)\,\mathrm{d}x\\&=\frac{1}{2}\left\langle u_i'(x)\middle|u_j'(x)\right\rangle\end{aligned} \tag{8}$$

可以看出 $\mathbf{K}$ 是实对称矩阵. 这个积分可以精确地用求和表示，而且只有同一个 FE 里面的 $u_i(x),u_j(x)$ 才能使积分不为零（bridge function 属于两个 FE），所以就得到了几乎是 block diagonalized 的矩阵，只是每一个 block 左上角和右下角的一个矩阵元与其他 block 重叠.

那如何计算基底函数的导数呢？可以通过计算勒让德插值多项式的导数得到

$$f_i'(t_j)=\frac{1}{\sqrt{w_{0i}}}\times\begin{cases}\dfrac{t_j-t_1}{t_i-t_1}\dfrac{t_j-t_2}{t_i-t_2}\cdots\dfrac{1}{t_i-t_j}\cdots\dfrac{t_j-t_{i-1}}{t_i-t_{i-1}}\dfrac{t_j-t_{i+1}}{t_i-t_{i+1}}\cdots\dfrac{t_j-t_N}{t_i-t_N} & (i\neq j)\\ \dfrac{1}{t_i-t_1}+\cdots+\dfrac{1}{t_i-t_{i-1}}+\dfrac{1}{t_i-t_{i+1}}+\ldots\dfrac{1}{t_i-t_N} & (i=j)\end{cases} \tag{9}$$

[2]未完成：这么说不严谨！因为并不是所有基底都是光滑的，分部积分的常数项也未必能消掉.

代入得

$$u'_{ij}(x_{ij'}) = \frac{1}{a_i^{3/2}} f'_j(t_{j'}) \qquad (1 < j < N) \tag{10}$$

$$u'_{iN}(x_{ij'}) = \frac{1}{a_i\sqrt{a_i + a_{i+1}}} f'_N(t_{j'}) \tag{11}$$

$$u'_{iN}(x_{i+1,j'}) = \frac{1}{a_{i+1}\sqrt{a_i + a_{i+1}}} f'_1(t_{j'}) \tag{12}$$

注意 u'_{iN} 在 x_{iN} 处的左右导数不相等.

现在就可以求动能矩阵的矩阵元了（由于是对称矩阵，我们只列出半边），先看不含 bridge function 的矩阵元

$$\begin{aligned} K_{(im),(in)} &= \frac{1}{2}\int_{x_{i1}}^{x_{iN}} u'_{im}(x)u'_{in}(x)\,\mathrm{d}x \\ &= \frac{1}{2}\sum_k a_i w_{0k} u'_{im}(x_{ik})u'_{in}(x_{ik}) \\ &= \frac{1}{2a_i^2}\sum_k w_{0k} f'_m(t_k)f'_n(t_k) \qquad (1 < m \leqslant n < N) \end{aligned} \tag{13}$$

再看含 bridge function 的矩阵元

$$\begin{aligned} K_{(im),(iN)} &= \frac{1}{2}\int_{x_{i1}}^{x_{iN}} u'_{im}(x)u'_{iN}(x)\,\mathrm{d}x \\ &= \frac{1}{2a_i^{3/2}\sqrt{a_i + a_{i+1}}}\sum_k w_{0k} f'_m(t_k)f'_N(t_k) \qquad (1 < m < N) \end{aligned} \tag{14}$$

$$\begin{aligned} K_{(iN),(iN)} &= \frac{1}{2}\int_{x_{i1}}^{x_{iN}} u'_{iN}(x)^2\,\mathrm{d}x + \frac{1}{2}\int_{x_{i+1,1}}^{x_{i+1,N}} u'_{i+1,1}(x)^2\,\mathrm{d}x \\ &= \frac{1}{2(a_i + a_{i+1})}\sum_k w_{0k}\left[\frac{1}{a_i} f'_N(t_k)^2 + \frac{1}{a_{i+1}} f'_1(t_k)^2\right] \end{aligned} \tag{15}$$

$$\begin{aligned} K_{(iN),(i+1,n)} &= \frac{1}{2}\int_{x_{i+1,1}}^{x_{i+1,N}} u'_{i,N}(x)u'_{i+1,n}(x)\,\mathrm{d}x \\ &= \frac{1}{2a_{i+1}^{3/2}\sqrt{a_i + a_{i+1}}}\sum_k w_{0k} f'_1(t_k)f'_n(t_k) \qquad (1 < n < N) \end{aligned} \tag{16}$$

$$\begin{aligned} K_{(iN),(i+1,N)} &= \frac{1}{2}\int_{x_{i+1,1}}^{x_{i+1,N}} u'_{i,N}(x)u'_{i+1,N}(x)\,\mathrm{d}x \\ &= \frac{1}{2a_{i+1}\sqrt{(a_i + a_{i+1})(a_{i+1} + a_{i+2})}}\sum_k w_{0k} f'_1(t_k)f'_N(t_k) \end{aligned} \tag{17}$$

势能矩阵

$$V_{\alpha\beta} = \int_{x_1}^{x_{N_e+1}} u_n^j(x) u_{n'}^{j'}(x) V(x)\,\mathrm{d}x = \tag{18}$$

Lanczos 算法

背景介绍

[3]在计算含时薛定谔方程时，若已知某时刻 t 的波函数 $\Psi(\mathbf{r},t)$，要求 $\Psi(\mathbf{r},t+\Delta t)$，通常使用传播算符

$$U(\Delta t) = \exp\left(-\frac{\mathrm{i}}{\hbar}\int_t^{t+\Delta t} \mathrm{d}t \cdot H(t)\right) \approx \exp(-\mathrm{i}H(t+\Delta t/2)\Delta t) \tag{1}$$

其中 H 为哈密顿算符．若 H 不含时，约等号变为等号

$$\Psi(\mathbf{r},t+\Delta t) = \exp(-\mathrm{i}H\Delta t/\hbar)\Psi(\mathbf{r},t) \tag{2}$$

通常情况下，用有限个正交归一基底 $\Psi_0,\ldots\Psi_{K-1}$ 近似展开 $\Psi(\mathbf{r},t+\Delta t)$，这时，$H$ 可以表示成矩阵的形式．

$$H_{ij} = \langle\Psi_i|\,H\,|\Psi_j\rangle \tag{3}$$

这样，幺正算符 $\exp(-\mathrm{i}H\Delta t/\hbar)$ 也可以表示成矩阵．根据定义

$$\exp(-\mathrm{i}H\Delta t/\hbar) = 1 + (-\mathrm{i}H\Delta t/\hbar) + \frac{1}{2!}(-\mathrm{i}H\Delta t/\hbar)^2\ldots \tag{4}$$

此时若把 H 对角化（求解本征方程），得到本征矢列矩阵 $\mathbf{P}$（单位正交阵）以及本征值矩阵 $\mathbf{\Lambda}$（对角矩阵），则可进行基底变换变到 H 矩阵的本征空间求解上式

$$\begin{aligned}\exp(-\mathrm{i}H\Delta t/\hbar) &= \mathbf{P}\times\mathrm{diag}(\mathrm{e}^{-\mathrm{iE_1t/\hbar}},\mathrm{e}^{-\mathrm{iE_2t/\hbar}}\ldots\mathrm{e}^{-\mathrm{iE_Nt/\hbar}})\times\mathbf{P}^{-1}\\ &= \sum_{j=0}^{K-1} P_{ij}\mathrm{e}^{-\mathrm{i}E_jt/\hbar}P_{jk}^{-1} = \sum_{j=0}^{K-1} P_{ij}P_{kj}\mathrm{e}^{-\mathrm{i}E_jt/\hbar}\end{aligned} \tag{5}$$

[3]参考 Renate Pazourek 的论文

要说明的是，这种算法的误差（除数值计算误差）来源于两个方面，第一是用 $\exp(\mathrm{i}H\Delta t/\hbar)$ 代替 $\exp\left(-\frac{\mathrm{i}}{\hbar}\int_t^{t+\Delta t}\mathrm{d}t\cdot H(t)\right)$，但如果不含时就没有该误差，第二是有限数量的基底不具有完备性，这个误差可以随着基底数量增加而减小.

Lanczos 算法

在以上的算法中，选取施密特正交归一化的 Krilov 基底作为基底. 令 $\Psi_0\equiv\Psi(\mathbf{r},t)$，$K$ 阶的 Krilov 基底为

$$\left\{|\Psi_0\rangle, H|\Psi_0\rangle, H^2|\Psi_0\rangle\ldots H^{K-1}|\Psi_0\rangle\right\} \tag{6}$$

我们把正交归一化后的基底记为

$$\{|\Psi_0\rangle, |\Psi_1\rangle, |\Psi_2\rangle\ldots|\Psi_{K-1}\rangle\} \tag{7}$$

这样，用其展开哈密顿算符

$$H_{ij}=\langle\Psi_i|H|\Psi_j\rangle \tag{8}$$

在算法上，本来我们可以按部就班地按照以上步骤做，然而由于 Krilov 基底的性质，可以通过一些定理（见词条最后）大大减少计算量. 正交归一化步骤如下

1. 如果 $\Psi(\mathbf{r},t)$ 没有归一化,将其进行归一化 $|\Psi_0\rangle=\Psi(\mathbf{r},t)/\sqrt{\langle\Psi(\mathbf{r},t)|\Psi(\mathbf{r},t)\rangle}$.
2. 把 Ψ_1 进行施密特正交化，令

$$|\tilde{\Psi}_1\rangle=(1-|\Psi_0\rangle\langle\Psi_0|)H|\Psi_0\rangle \qquad |\Psi_1\rangle=|\tilde{\Psi}_1\rangle/\sqrt{\langle\tilde{\Psi}_1|\tilde{\Psi}_1\rangle} \tag{9}$$

3. 现在起我们令 $\beta_j=\sqrt{\langle\tilde{\Psi}_j|\tilde{\Psi}_j\rangle}$.

定理 1

对以上定义的基底

$$(1-|\Psi_j\rangle\langle\Psi_j|-|\Psi_{j-1}\rangle\langle\Psi_{j-1}|-\ldots-|\Psi_0\rangle\langle\Psi_0|)H^{j+1}|\Psi_0\rangle \tag{10}$$

与

$$(1-|\Psi_j\rangle\langle\Psi_j|-|\Psi_{j-1}\rangle\langle\Psi_{j-1}|-\ldots-|\Psi_0\rangle\langle\Psi_0|)H|\Psi_j\rangle \tag{11}$$

共线. 故现在起采用后者进行正交化.

定理 2

$$\langle\Psi_{j-1}|\,H\,|\Psi_j\rangle=\sqrt{\langle\tilde{\Psi}_j|\tilde{\Psi}_j\rangle}=\beta_j \tag{12}$$

4. 把 $H^2\,|\Psi_0\rangle$ 进行施密特正交化，令

$$\alpha_i=\langle\Psi_i|H|\Psi_i\rangle \tag{13}$$

根据两个定理，正交化结果也可以写成

$$\begin{aligned}|\tilde{\Psi}_2\rangle&=(1-|\Psi_1\rangle\langle\Psi_1|-|\Psi_0\rangle\langle\Psi_0|)H\,|\Psi_1\rangle\\&=H\,|\Psi_1\rangle-\alpha_1\,|\Psi_1\rangle-\beta_1\,|\Psi_0\rangle\end{aligned} \tag{14}$$

5. 再次归一化 $|\Psi_2\rangle=|\tilde{\Psi}_2\rangle/\sqrt{\langle\tilde{\Psi}_2|\tilde{\Psi}_2\rangle}=|\tilde{\Psi}_2\rangle/\beta_2$

定理 3

$$\begin{aligned}|\tilde{\Psi}_{j+1}\rangle&=(1-|\Psi_j\rangle\langle\Psi_j|-\ldots-|\Psi_0\rangle\langle\Psi_0|)H\,|\Psi_j\rangle\\&=(1-|\Psi_j\rangle\langle\Psi_j|-|\Psi_{j-1}\rangle\langle\Psi_{j-1}|)H\,|\Psi_j\rangle\\&=H\,|\Psi_j\rangle-\alpha_j\,|\Psi_j\rangle-\beta_j\,|\Psi_{j-1}\rangle\end{aligned} \tag{15}$$

这就是最简洁的正交化公式．比起最原始的正交化

$$|\tilde{\Psi}_{j+1}\rangle=(1-|\Psi_j\rangle\langle\Psi_j|-\ldots-|\Psi_0\rangle\langle\Psi_0|)H^{j+1}\,|\Psi_0\rangle \tag{16}$$

定理 3 只用到了一个矩阵-矢量乘法，两个矢量数乘和两个矢量减法．

接下来只要对 $j=2,4,5...K-1$ 不断重复定理 3 中的正交化和归一化，就可以得到所有正交归一基底 $|\Psi_j\rangle$．

现在我们来求该基底下的 **H** 矩阵．根据 α_i 的定义以及定理 2 可以发现它们分别是矩阵的主对角元和副对角元．

定理 4

在正交归一基底 $|\Psi_j\rangle$ 中，矩阵 $\mathbf{H}$ 是对称三对角矩阵．所以我们已经顺便求出了所有的矩阵元！

$$\mathbf{H}=\begin{pmatrix}\alpha_0 & \beta_1 & & & \\ \beta_1 & \alpha_1 & \beta_2 & & \\ & \beta_2 & \alpha_2 & \ddots & \\ & & \ddots & \ddots & \beta_{K-1} \\ & & & \beta_{K-1} & \alpha_{K-1}\end{pmatrix} \tag{17}$$

该矩阵具有维数小，易求本征问题的优势．

定理 1 证明

根据施密特正交化的性质，Krilov 基底的前 j 项的与 $|\Psi_0\rangle\dots|\Psi_j\rangle$ 展开同一空间．所以 $H^j|\Psi_0\rangle=c_j|\Psi_j\rangle+\ldots+c_0|\Psi_0\rangle$，所以（以下所有系数只是用来表示线性组合，具体值不重要）

$$\begin{aligned}H^{j+1}|\Psi_0\rangle &= c_jH|\Psi_j\rangle+H(c_{j-1}|\Psi_{j-1}\rangle\cdots+c_0|\Psi_0\rangle)\\ &= c_jH|\Psi_j\rangle+H\left(c'_{j-1}H^{j-1}|\Psi_0\rangle+c'_{j-2}H^{j-2}|\Psi_0\rangle\cdots+c'_0|\Psi_0\rangle\right)\\ &= c_jH|\Psi_j\rangle+\left(c'_{j-1}H^{j}|\Psi_0\rangle+c'_{j-2}H^{j-1}|\Psi_0\rangle\cdots+c'_0|\Psi_0\rangle\right)\\ &= c_jH|\Psi_j\rangle+\left(c''_{j-1}|\Psi_j\rangle+c''_{j-2}|\Psi_{j-1}\rangle+\ldots+c''_0|\Psi_0\rangle\right)\end{aligned} \tag{18}$$

把 $H^{j+1}|\Psi_0\rangle$ 施密特正交归一化

$$\begin{aligned}&(1-|\Psi_j\rangle\langle\Psi_j|-\ldots-|\Psi_0\rangle\langle\Psi_0|)H^{j+1}|\Psi_0\rangle\\ &=c_j(1-|\Psi_j\rangle\langle\Psi_j|-\ldots-|\Psi_0\rangle\langle\Psi_0|)H|\Psi_j\rangle\\ &\quad+(1-|\Psi_j\rangle\langle\Psi_j|-\ldots-|\Psi_0\rangle\langle\Psi_0|)\left(c''_{j-1}|\Psi_j\rangle+c''_{j-2}|\Psi_{j-1}\rangle+\ldots+c''_0|\Psi_0\rangle\right)\\ &=c_j(1-|\Psi_j\rangle\langle\Psi_j|-\ldots-|\Psi_0\rangle\langle\Psi_0|)H|\Psi_j\rangle\end{aligned} \tag{19}$$

定理 2 证明

要证 $\langle\Psi_{j-1}|H|\Psi_j\rangle=\sqrt{\langle\tilde{\Psi}_j|\tilde{\Psi}_j\rangle}$，即证 $\langle\Psi_{j-1}|H|\tilde{\Psi}_j\rangle=\langle\tilde{\Psi}_j|\tilde{\Psi}_j\rangle$，即证 $\langle\tilde{\Psi}_j|\tilde{\Psi}_j\rangle=\langle\tilde{\Psi}_j|H\Psi_{j-1}\rangle$ 而

$$\langle\tilde{\Psi}_j|\tilde{\Psi}_j\rangle=\langle\tilde{\Psi}_j|(1-|\Psi_{j-1}\rangle\langle\Psi_{j-1}|-\ldots-|\Psi_0\rangle\langle\Psi_0|)H|\Psi_{j-1}\rangle \tag{20}$$

由于上式中 $\langle\tilde{\Psi}_j|\left(|\Psi_{j-1}\rangle\langle\Psi_{j-1}|-\ldots-|\Psi_0\rangle\langle\Psi_0|\right)=0$

$$\langle\tilde{\Psi}_j|\tilde{\Psi}_j\rangle=\langle\tilde{\Psi}_j|H|\Psi_{j-1}\rangle \tag{21}$$

定理 3 证明

要证明定理 3，即证，对 $n\leqslant j-2$，有 $\langle\Psi_n|\,H\,|\Psi_j\rangle=0$，$H$ 为厄米算符时即证 $\langle\Psi_j|\,H\,|\Psi_n\rangle=0$，即证对 $m\geqslant j+2$，$\langle\Psi_m|\,H\,|\Psi_j\rangle=0$．在定理 1 的证明类似，我们知道

$$\begin{aligned}H\,|\Psi_j\rangle&=H\left(c_jH^j\,|\Psi_0\rangle+\ldots+c_0\,|\Psi_0\rangle\right)=c_jH^{j+1}\,|\Psi_0\rangle+\ldots+c_0H\,|\Psi_0\rangle\\&=c'_j\,|\Psi_{j+1}\rangle+\ldots+c'_0\,|\Psi_0\rangle\end{aligned} \tag{22}$$

所以对 $m\geqslant j+2$ 有

$$\langle\Psi_m|\,H\,|\Psi_j\rangle=\langle\Psi_m|\left(c'_j\,|\Psi_{j+1}\rangle+\ldots+c'_0\,|\Psi_0\rangle\right)=0 \tag{23}$$

定理 4 证明

首先证明 $\mathbf{H}$ 是三对角矩阵．即证 $m\geqslant j+2$ 或 $m\leqslant j-2$ 时 $\langle\Psi_m|\,H\,|\Psi_j\rangle=0$．$m\geqslant j+2$ 的情况在证明 3 中已经证明，只需证明另一种情况．对厄米矩阵，$\langle H\Psi_m|\Psi_j\rangle=\langle\Psi_m|H|\Psi_j\rangle=0$，取复共轭，即 $\langle\Psi_j|H|\Psi_m\rangle=0$．可见左边的角标的确小于等于右边的角标减 2，证毕．

然后证明 $\mathbf{H}$ 是实矩阵，考虑到厄米矩阵 $H_{mn}=H^*_{nm}$ 的性质，只需要证明 $H_{j,j+1}$ 是实数．对任意 $j=2\ldots K-1$

$$\begin{aligned}H_{j,j+1}&=\langle\Psi_j|H|\Psi_{j+1}\rangle=\langle\Psi_j|\,H(H\,|\Psi_j\rangle-\alpha_j\,|\Psi_j\rangle-\beta_j\,|\Psi_{j-1}\rangle)\\&=\langle\Psi_j|H^2|\Psi_j\rangle-\alpha_j\,\langle\Psi_j|\,H\,|\Psi_j\rangle-\beta_j\,\langle\Psi_j|\,H\,|\Psi_{j-1}\rangle\end{aligned} \tag{24}$$

易证 $\langle\Psi_j|H^2|\Psi_j\rangle$，$\alpha_j$ 及 β_j 是实数，只需要证明 $\langle\Psi_j|H|\Psi_{j-1}\rangle$ 是实数即可，即证 $\langle\Psi_{j-1}|H|\Psi_j\rangle$ 是实数，重复上述论证，即证 $\langle\Psi_{j-2}|H|\Psi_{j-1}\rangle$ 是实数，……，即证 $\langle\Psi_1|H|\Psi_0\rangle$ 是实数，即证 $\langle\Psi_0|H|\Psi_1\rangle$ 是实数，即证 $\langle\Psi_0|H|\tilde{\Psi}_1\rangle$ 是实数．而

$$\langle\Psi_0|H|\tilde{\Psi}_1\rangle=\langle\Psi_0|\,H(H\,|\Psi_0\rangle-\alpha_0\,|\Psi_0\rangle)=\langle\Psi_0|H^2|\Psi_0\rangle-\alpha_0\,\langle\Psi_0|H|\Psi_0\rangle \tag{25}$$

显然是实数．以上的思路是数学归纳法的逆过程．

指数格点

预备知识 等比数列[112]

在氢原子的数值计算中，我们有时候在某个半径内使用指数格点，即第 $n+1$ 个 FE 的长度是第 n 个 FE 长度的 α 倍．如果第一个 FE 的长度为 d，那么有

$$x_i = d\frac{\alpha^i - 1}{\alpha - 1} \qquad (i = 0, 1, \dots) \tag{1}$$

如果我们需要把格点变密或者变疏，一种幼稚的想法是把 d 变大而 α 保持不变．但这么做以后会发现，在不同的位置，FE 并没有等比例地变化．

想象我们画出 x_i 的连续函数图．严格的做法是把该图在 x 方向上缩放 β 即可（$\beta > 1$ 就是越疏，$\beta < 1$ 就是越密）．缩放后，式 1 变为

$$x_i = d\frac{\alpha^{\beta\cdot i} - 1}{\alpha - 1} = d_1\frac{\alpha_1^i - 1}{\alpha_1 - 1} \qquad (i = 0, 1, \dots) \tag{2}$$

其中

$$\alpha_1 = \alpha^\beta \qquad d_1 = d\frac{\alpha^\beta - 1}{\alpha - 1} \tag{3}$$

虚时间法求基态波函数

这里介绍一种求解基态波函数的数值方法．如果薛定谔方程中势能不含时间，用分离变量法解薛定谔方程的结果是

$$\Psi(\mathbf{x}, t) = \sum_i \psi_i(\mathbf{x}) \mathrm{e}^{-\mathrm{i}E_i t} \tag{1}$$

其中 $\psi_i(\mathbf{x})$ 是能量为 E_i 的能量本征态．

现在若要求基态，我们可以用虚数时间，即 $t' = -\mathrm{i}t$，使得含时波函数变为

$$\Psi(\mathbf{x}, t) = \sum_i \psi_i(\mathbf{x}) \mathrm{e}^{-E_i t} \tag{2}$$

这样，激发态衰减的速度就都比基态要快，当 $t \to +\infty$ 的时候，就只剩下基态波函数了．

假设我们有一个求解 TDSE 的数值方法，那么我们只需要用其求解方程

$$\mathbf{H}\Psi = -\frac{\partial \Psi}{\partial t} \tag{3}$$

然后每个循环对波函数进行归一化即可．因为该方程的分离变量解就是式 2．

电磁场中氢原子薛定谔方程数值解

预备知识 薛定谔方程[1331]，类氢原子的定态波函数[1536]

本文使用原子单位制[1315]．我们希望求解氢原子在电场中的薛定谔方程

$$-\frac{1}{2m}\nabla^2\Psi + V(\mathbf{r},t)\Psi = \mathrm{i}\frac{\partial}{\partial t}\Psi \tag{1}$$

虽然最直观的方法是使用直角坐标，但计算效率很低．实际中一般使用球坐标系，用球谐函数展开波函数．如果哈密顿算符是轴对称的（例如线偏振电场），那么只需要 $m = 0$ 的球谐函数，即勒让德多项式．

$$\Psi(\mathbf{r},t) = \frac{1}{r}\sum_{l,m}\psi_{l,m}(r)Y_{l,m}(\mathbf{r}) \tag{2}$$

其中 $\psi_{l,m}(r)$ 是约化径向波函数（scaled radial wave function），式中每一项叫做一个**分波（partial wave）**．

在薛定谔方程中，一种简单的做法是使用以下势能表示库仑势能以及电磁场对电子的作用

$$V(\mathbf{r},t) = -\frac{Z}{r} - q\mathbf{E}(t)\cdot\mathbf{r} \tag{3}$$

在原子单位中电子电荷 $q = -1$，该势能使用了长度规范[1437]，此外我们还可以使用速度规范速度规范[1438]，也是等效的．在长度规范中，我们近似认为原子的长度远小于电磁波的波长，所以电磁波不随位置变化．

线偏振光

若我们取电场极化方向为 $\hat{\mathbf{z}}$，则角动量 L_z 是一个守恒量．假设初始波函数关于 $\hat{\mathbf{z}}$ 轴对称，那么在波函数的整个演化（propagation）过程中，我们只需

要 $m=0$ 的球谐函数展开波函数，即

$$\Psi(\mathbf{r},t)=\frac{1}{r}\sum_{l'}\psi_{l'}(r)Y_{l',0}(\hat{\mathbf{r}}) \tag{4}$$

另外薛定谔方程中 $V(\mathbf{r},t)=\mathbf{E}(t)\cdot\mathbf{r}=E(t)z$，进而可以用球谐函数表示（式 9 [761]）

$$V(\mathbf{r},t)=\sqrt{\frac{4\pi}{3}}E(t)r\cdot Y_{1,0}(\hat{\mathbf{r}}) \tag{5}$$

以上两式代入薛定谔方程，再把每一项与 $Y_{l,0}(\hat{\mathbf{r}})$ 做内积（放入积分 $\int Y_{1,0}^{*}(\hat{\mathbf{r}})\,\mathrm{d}\Omega$ 中）可得一系列耦合的方程

$$H_0\psi_l(r)+\sum_{l'=0}^{\infty}E(t)rF_{l,l'}\psi_{l'}(r)=\mathrm{i}\frac{\partial\psi_l}{\partial t}\quad(l=0,1,\dots) \tag{6}$$

其中无场哈密顿算符为

$$H_0=-\frac{1}{2m}\frac{\mathrm{d}^2}{\mathrm{d}r^2}-\frac{Z}{r}+\frac{l(l+1)}{2mr^2} \tag{7}$$

矩阵 $\mathbf{F}$ 就是跃迁偶极子矩阵的角向积分（式 3 [1449]）

$$\begin{aligned}F_{l,l'}&=\sqrt{\frac{4\pi}{3}}\langle Y_{l,0}|Y_{1,0}|Y_{l',0}\rangle\\&=\sqrt{(2l+1)(2l'+1)}\begin{pmatrix}l&1&l'\\0&0&0\end{pmatrix}^2\\&=\sqrt{\frac{2l+1}{2l'+1}}\begin{bmatrix}l&1&l'\\0&0&0\end{bmatrix}^2\end{aligned} \tag{8}$$

可见，当没有外场的时候每一个项（即每一个分波）都可以独立演化，而电场将不同的分波耦合起来．根据氢原子的选择定则 $\Delta l=\pm1$（式 5 [1449]），矩阵 $\mathbf{F}$ 中除了两条副对角线上的元都为零．另外由式 2 [855] 易得 $\mathbf{F}$ 是一个对称矩阵．

任意偏振光

将所有 (l,m) 按某种顺序排列，例如

$$(0,0),\ (1,-1),\ (1,0),\ (1,1),\ (2,-2),\dots \tag{9}$$

并将他们编号为 $\lambda = 1, 2, \ldots$，那么可以把式 2 记为

$$\Psi(\mathbf{r}, t) = \frac{1}{r}\sum_{\lambda} \psi_\lambda(r) Y_\lambda(\mathbf{r}) \tag{10}$$

势能函数为 $V(\mathbf{r}, t) = \mathbf{E}(t) \cdot \mathbf{r} = E_x x + E_y y + E_z z$．式 6 的径向耦合方程拓展为

$$\begin{aligned} H_0\psi_\lambda(r) + \sum_{\lambda'} r\left[E_x(t)F^x_{\lambda,\lambda'} + E_y(t)F^y_{\lambda,\lambda'} + E_z(t)F^z_{\lambda,\lambda'}\right]\psi_{\lambda'}(r) \\ = \mathrm{i}\frac{\partial \psi_\lambda}{\partial t} \quad (\lambda = 0, 1, \ldots) \end{aligned} \tag{11}$$

三个耦合矩阵分别为

$$F^x_{\lambda,\lambda'} = \sqrt{\frac{2\pi}{3}} E_x(\langle Y_{l,m}|Y_{1,-1}|Y_{l',m'}\rangle - \langle Y_{l,m}|Y_{1,1}|Y_{l',m'}\rangle) \tag{12}$$

$$F^y_{\lambda,\lambda'} = \mathrm{i}\sqrt{\frac{2\pi}{3}} E_y(\langle Y_{l,m}|Y_{1,-1}|Y_{l',m'}\rangle + \langle Y_{l,m}|Y_{1,1}|Y_{l',m'}\rangle) \tag{13}$$

$$F^z_{\lambda,\lambda'} = \sqrt{\frac{4\pi}{3}} E_z \langle Y_{l,m}|Y_{1,0}|Y_{l',m'}\rangle \tag{14}$$

其中（式 3 [1449]）

$$\langle Y_{l,m}|Y_{1,m_1}|Y_{l',m'}\rangle = (-1)^m \sqrt{\frac{3(2l+1)(2l'+1)}{4\pi}} \begin{pmatrix} l & 1 & l' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & 1 & l' \\ -m & m_1 & m' \end{pmatrix} \tag{15}$$

算符拆分

在实际的程序中，我们可以把演化子 $\exp(-\mathrm{i}H\Delta t)$ 拆成两项．虽然这么做会引入一定的误差，但是却大大提高了效率

$$\exp(-\mathrm{i}H\Delta t) = \exp\left(-\mathrm{i}H_0\frac{\Delta t}{2}\right)\exp(-\mathrm{i}F\Delta t)\exp\left(-\mathrm{i}H_0\frac{\Delta t}{2}\right) + \mathcal{O}\left(\Delta t^3\right) \tag{16}$$

也就是说，在每个时间步长 Δt 中，我们可以把波函数先根据

$$H_0\psi_l(r) = \mathrm{i}\frac{\partial \psi_l}{\partial t} \tag{17}$$

演化 $\Delta t/2$，其中 t_{mid} 取这段时间的中点．再对每个 $\mathbf{r}$ 根据

$$\sum_{l'} E(t_{mid}) r F_{l,l'} \psi_{l'}(r) = \mathrm{i}\frac{\partial \psi_l}{\partial t} \tag{18}$$

演化 Δt．最后再根据式 17 演化 $\Delta t/2$．

理论上也可以把演化子划分成两项 $\exp(-\mathrm{i}F\Delta t)\exp(-\mathrm{i}H_0\Delta t)$ 但据说式 16 更精确．至于相邻两步之间产生的 $\exp(-\mathrm{i}H_0\Delta t/2)\exp(-\mathrm{i}H_0\Delta t/2)$ 是否可以合并为 $\exp(-\mathrm{i}H_0\Delta t)$，取决于所使用的算法这么做以后是否会引入更大误差．

网格和演化算法

预备知识 Crank-Nicolson 算法（一维）[1925]

未完成：以下内容应该放在一维薛定谔方程里面讲解

可以使用二维数组储存波函数，每一列（或行）是一个分波的 $\psi_l(r)$．径向网格可以使用等间距网格，但 FEDVR 网格[1928] 效率要更高．

演化可以并使用 Crank-Nicolson 算法 Crank-Nicolson 算法（一维）[1925] 演化．但是 Lanczos 算法[1931] 效率更高．

氢原子球坐标数值解 TDSE 2

- 本词条处于草稿阶段．

预备知识 电磁场中氢原子薛定谔方程数值解[1937]，张量积空间[513]

相比于“电磁场中氢原子薛定谔方程数值解[1937]”，这里的描述更数学．

总波函数在球谐基底上展开（我们把 (l,m) 按一定顺序排序后的序号记为 λ）

$$\Psi(\mathbf{r},t)=\sum_{\lambda}|\psi_\lambda(t)\rangle\otimes|Y_\lambda(\hat{\mathbf{r}})\rangle \tag{1}$$

哈密顿算符为

$$H=H_0+V_F(t)=K_r\otimes I+\frac{1}{2r^2}\otimes L^2+V(r)\otimes I+V_F(t) \tag{2}$$

耦合方程组为

$$\sum_{\lambda'}\langle\lambda|H|\lambda'\rangle\,|\psi_{\lambda'}(t)\rangle=-\mathrm{i}\frac{\partial}{\partial t}|\psi_\lambda(t)\rangle \tag{3}$$

其中每个矩阵元 $\langle\lambda|H|\lambda'\rangle$ 都是一个关于 r 空间的算符（张量空间的算符如果关于一个小空间的基底求矩阵元，那么每个矩阵元都会是另一个小空间中的算符）.

无外场的哈密顿算符的 $\langle\lambda|H_0|\lambda'\rangle$ 将会是对角的，即每个 $\psi_\lambda(r,t)$ 都会独立传播，传播子为 $\exp(-\mathrm{i}\langle\lambda|H|\lambda\rangle\Delta t)$.

算符拆分（split operators）

无论用什么传播算法，传播子总可以记为 $\exp(-\mathrm{i}H\Delta t)$，这里近似 Δt 内 H 不随时间变化. 如果 H 可以拆分为几个算符之和（式 2），$\exp(-\mathrm{i}H\Delta t)$ 不一定能拆分成几个传播子的乘积（因为这些项不互相对易），但 Δt 很小时会近似成立.

一种精度比较高的拆分方法是将不含外场（field free）的 H_0 和含外场的 V_F 分开传播

$$\exp(-\mathrm{i}H\Delta t)=\exp\left(-\mathrm{i}H_0\frac{\Delta t}{2}\right)\exp(-\mathrm{i}V_F\Delta t)\exp\left(-\mathrm{i}H_0\frac{\Delta t}{2}\right)+\mathcal{O}\left(\Delta t^3\right)\tag{4}$$

由于 $\langle\lambda|H_0|\lambda'\rangle$ 都是对角的，根据式 3，$\exp(-\mathrm{i}H_0\Delta t/2)$ 作用在总波函数上其实就相当于 $\exp(-\mathrm{i}\langle\lambda|H_0|\lambda\rangle\Delta t)$ 分别作用在每个 $\psi_\lambda(r)$ 上.

线偏振外场

在 length gauge 下，如果有只延 z 方向的外场，那么

$$V_F(t)=\mathbf{E}(t)\cdot\mathbf{r}=E(t)z=E(t)r\otimes Y_{10}(\hat{\mathbf{r}})\tag{5}$$

所以矩阵元为

$$\langle\lambda|V_F(t)|\lambda'\rangle=E(t)r\langle\lambda|Y_{10}|\lambda'\rangle\tag{6}$$

如果我们在 r 空间中取类似 $\delta(r-r_i)$ 的基底 $|r_i\rangle$（例如等间距基底或 FEDVR 基底），那么 V_F 也会有一个很好的性质就是它们可以表示为 r 空间的对角矩阵（矩阵元为 r_i）和角向空间中的一个算符（矩阵）的张量积.

与式 1 相反，将总波函以不同径向基底拆分成若干个角向波函数

$$\Psi(\mathbf{r},t)=\sum_i|r_i\rangle\otimes f_i(\hat{\mathbf{r}})\tag{7}$$

将 V_F 作用在总波函数上得

$$(r \otimes Y_{10}) \sum_i |r_i\rangle \otimes f_i(\hat{\mathbf{r}}) = \sum_i r_i |r_i\rangle \otimes [Y_{10} f_i(\hat{\mathbf{r}})] \tag{8}$$

所以我们只需要对每个 $f_i(\hat{\mathbf{r}})$ 使用传播子 $\exp[-\mathrm{i}(r_i Y_{10})\Delta t]$ 进行传播即可．容易证明

$$\exp\left(\sum_i |i\rangle\langle i| \otimes \hat{B}_i\right) = \prod_i |i\rangle\langle i| \otimes \mathrm{e}^{\hat{B}_i} \tag{9}$$

$$\exp\left(\sum_i \hat{A}_i \otimes |i\rangle\langle i|\right) = \prod_i \mathrm{e}^{\hat{B}_i} \otimes |i\rangle\langle i| \tag{10}$$

选择定则起到的作用

在对每个 $f_i(\hat{\mathbf{r}})$ 使用传播子 $\exp[-\mathrm{i}(r_i Y_{10})\Delta t]$ 的时候，由于 expokit 只需要用户提供矩阵 $\langle\lambda|Y_{10}|\lambda'\rangle$ 与列矢量相乘的 implementation．如果知道选择定则，即那些矩阵元为 0，我们就可以使用 sparse matrix 与列矢量的乘法从而提高计算效率．

例如对于线偏振光，只选取 $m = 0$ 的基底，选择定则要求 $\Delta l = \pm 1$，$\langle\lambda|Y_{10}|\lambda'\rangle$ 只是一个三对角矩阵且对角元为 0．

选择定则对于氢原子其实并没有太大的性能提升，因为实践表明两个径向传播子 $\exp(-\mathrm{i}H_0\Delta t/2)$ 的传播才是最耗时的．而对于氦原子，使用选择定则可能有更明显的优势．

第十二章

氦原子薛定谔方程数值解

氦原子数值解 TDSE 笔记

预备知识 广义球谐函数[771]

波函数展开

使用角向基底为两个电子总轨道角动量和 z 分量的本征态[1403]，即广义球谐函数 $\mathcal{Y}_{l_1,l_2}^{L,M}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)$.

总波函数所在的空间可以看做角向空间和径向空间的张量积空间. 现在有了角向空间的基底，总波函数就可以在该基底上展开

$$|\Psi\rangle=\sum_{\lambda}|R_\lambda\rangle\,|\mathcal{Y}_\lambda\rangle \tag{1}$$

其中 λ 是将所有的 (l_1,l_2,L,M) 组合排序后的序号. $|R_\lambda\rangle$ 是二维径向波函数

$$|R_\lambda\rangle=\frac{1}{r_1r_2}\psi_\lambda(r_1,r_2) \tag{2}$$

即总波函数为

$$\Psi(\mathbf{r}_1,\mathbf{r}_2)=\sum_{L,M,l_1,l_2}\frac{1}{r_1r_2}\psi_{l_1,l_2}^{L,M}(r_1,r_2)\mathcal{Y}_{l_1,l_2}^{L,M}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2) \tag{3}$$

由张量积空间[513] 中的结论，我们可以求哈密顿算符 H 关于角向基底的“矩阵元” $H_{\lambda,\lambda'}$，每个矩阵元是径向空间中的一个算符. 列出薛定谔方程的“矩阵形式”，就得到了一组 couple 的径向波函数的薛定谔方程

$$\sum_{\lambda'}H_{\lambda,\lambda'}\,|R_{\lambda'}\rangle=\mathrm{i}\frac{\partial}{\partial t}\,|R_\lambda\rangle \tag{4}$$

总哈密顿可以表示为

$$H=H_1+H_2+V_{12}+V_{F1}+V_{F2} \tag{5}$$

其中 $H_i\ \ (i=1,2)$ 是单个电子的哈密顿算符，对应对角矩阵

$$H_i=K_{ri}+\frac{L_i^2}{2mr_i^2}-\frac{2}{r_i}\qquad(i=1,2) \tag{6}$$

其中第二项是对角矩阵是因为 $\mathcal{Y}$ 基底是 L_i^2 的本征函数. V_{12} 是两电子之间的库仑势能，V_{Fi} 是电场对每个电子的作用势能.

波函数的对称性

我们一般假设电子自旋态恒为 singlet （反对称），所以总波函数满足粒子交换对称 $\Psi(\mathbf{r}_1,\mathbf{r}_2)=\Psi(\mathbf{r}_2,\mathbf{r}_1)$. 由式 4 [771] 可以导出分波的对称性

$$\psi^{L,M}_{l_1,l_2}(r_1,r_2)=(-1)^{l_1+l_2-L}\psi^{L,M}_{l_2,l_1}(r_2,r_1) \tag{7}$$

理论上用该式可以在程序中减少一半重复的分波，并且分波的 coupling 演化也可以得到精简，但由于编程麻烦暂时没有实现.

另外可以证明（其实是数值计算发现的）l_1+l_2-L 为奇数的分波（奇分波）和 l_1+l_2-L 为偶数的分波（偶分波）在线偏振电场中是独立演化的，也就是 V_{12} 和 z_1,z_2 不会把奇偶性不同的分波耦合起来. 所以如果氦原子从基态（只包含偶分波：$l_1=l_2,L=0$）开始在线偏振电场中演化，那么奇数分波的项将始终为零，可以在程序中去除.

电子相互作用项

$$\begin{aligned}V_{12}=\frac{1}{|\mathbf{r}_2-\mathbf{r}_1|}&=4\pi\sum_{l=0}^{\infty}\frac{1}{2l+1}\frac{r_<^l}{r_>^{l+1}}\sum_{m=-l}^{l}Y^*_{l,m}(\hat{\mathbf{r}}_1)Y_{l,m}(\hat{\mathbf{r}}_2)\\&=4\pi\sum_{l=0}^{\infty}\frac{(-1)^l}{\sqrt{2l+1}}\frac{r_<^l}{r_>^{l+1}}\mathcal{Y}^{0,0}_{l,l}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)\end{aligned} \tag{8}$$

这就相当于将张量积空间中的线性算符分解成径向和角向算符的张量积之和. 第二步使用了式 9 [852]，即

$$\mathcal{Y}^{0,0}_{l,l}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)=\frac{(-1)^l}{\sqrt{2l+1}}\sum_{m=-l}^{l}Y^*_{lm}(\hat{\mathbf{r}}_1)Y_{lm}(\hat{\mathbf{r}}_2) \tag{9}$$

要计算矩阵元 $\langle\mathcal{Y}_\lambda|V_{12}|\mathcal{Y}_{\lambda'}\rangle$，就要对 6 个球谐函数的乘积的线性组合做两次角向积分. 使用式 12 [852] 可以将上式表示为六个 CG 系数相乘的线性组合

（二重求和）.

$$
\begin{aligned}
&\left\langle \mathcal{Y}_{l_1 l_2}^{LM} \middle| Y_{lm}^*(\hat{\mathbf{r}}_1) Y_{lm}(\hat{\mathbf{r}}_2) \middle| \mathcal{Y}_{l_1' l_2'}^{L'M'} \right\rangle \\
&= \sum_{m_1', m_2'} \sum_{m_1, m_2} \begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{bmatrix} \begin{bmatrix} l_1' & l_2' & L' \\ m_1' & m_2' & M' \end{bmatrix} \times \\
&\quad \iint Y_{l_1 m_1}^*(\hat{\mathbf{r}}_1) Y_{l_2 m_2}^*(\hat{\mathbf{r}}_2) Y_{lm}^*(\hat{\mathbf{r}}_1) Y_{lm}(\hat{\mathbf{r}}_2) Y_{l_1' m_1'}(\hat{\mathbf{r}}_1) Y_{l_2' m_2'}(\hat{\mathbf{r}}_2) \, \mathrm{d}\Omega_1 \, \mathrm{d}\Omega_2 \\
&= \sum_{m_1', m_2'} \sum_{m_1, m_2} (-1)^{m_1+m_2+m} \begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{bmatrix} \begin{bmatrix} l_1' & l_2' & L' \\ m_1' & m_2' & M' \end{bmatrix} \times \\
&\quad \int Y_{l_1, -m_1}(\hat{\mathbf{r}}_1) Y_{l, -m}(\hat{\mathbf{r}}_1) Y_{l_1' m_1'}(\hat{\mathbf{r}}_1) \, \mathrm{d}\Omega_1 \int Y_{l_2, -m_2}(\hat{\mathbf{r}}_2) Y_{lm}(\hat{\mathbf{r}}_2) Y_{l_2' m_2'}(\hat{\mathbf{r}}_2) \, \mathrm{d}\Omega_2 \\
&= \frac{2l+1}{4\pi} \sqrt{(2l_1+1)(2l_1'+1)(2l_2+1)(2l_2'+1)} \sum_{m_1', m_2'} \sum_{m_1, m_2} (-1)^{m_1+m_2+m} \times \\
&\quad \begin{pmatrix} l_1 & l & l_1' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & l & l_1' \\ -m_1 & -m & m_1' \end{pmatrix} \begin{pmatrix} l_2 & l & l_2' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_2 & l & l_2' \\ -m_2 & m & m_2' \end{pmatrix} \times \\
&\quad \begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{bmatrix} \begin{bmatrix} l_1' & l_2' & L' \\ m_1' & m_2' & M' \end{bmatrix}
\end{aligned} \tag{10}
$$

现在试图对 m 求和并使用选择定理（3j 符号第二行之和为 0），得（注意 m 的范围）该求和中最多只有一项不为 0，所以

$$
\begin{aligned}
&\left\langle \mathcal{Y}_{l_1 l_2}^{LM} \middle| \mathcal{Y}_{l,l}^{0,0} \middle| \mathcal{Y}_{l_1' l_2'}^{L'M'} \right\rangle = \\
&\delta_{M,M'} (-1)^l \frac{\sqrt{2l+1}}{4\pi} \sqrt{(2l_1+1)(2l_1'+1)(2l_2+1)(2l_2'+1)} \times \\
&\sum_{m_1', m_2'} \sum_{m_1, m_2} (-1)^{m_1'+m_2} \begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{bmatrix} \begin{bmatrix} l_1' & l_2' & L' \\ m_1' & m_2' & M \end{bmatrix} \begin{pmatrix} l_1 & l & l_1' \\ 0 & 0 & 0 \end{pmatrix} \times \\
&\begin{pmatrix} l_1 & l & l_1' \\ -m_1 & m_1 - m_1' & m_1' \end{pmatrix} \begin{pmatrix} l_2 & l & l_2' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_2 & l & l_2' \\ -m_2 & m_2 - m_2' & m_2' \end{pmatrix}
\end{aligned} \tag{11}
$$

其中对 m_1, m_2 的求和要求

$$
m_1 + m_2 = m_1' + m_2' = M \qquad |m_1' - m_1| \leqslant l \qquad |m_2' - m_2| \leqslant l \tag{12}
$$

由于 CG 系数都是实数，这必定是一个实数矩阵．选择定则如下

- 根据总角动量守恒（$1/r_{12}$ 算符与 L, L_z 算符对易），$L \neq L'$ 或 $M \neq M'$ 时，矩阵元为零.

- 根据 parity CG coefficients，当 $l_1' + l + l_1$ 或 $l_2' + l + l_2$ 为奇数时，上式等于 0.

- 根据 $\Pi\mathcal{Y} = (-1)^{l_1+l_2}\mathcal{Y}$，考虑到 $1/r_{12}$ 具有偶宇称，所以 $l_1 + l_2$ 与 $l_1' + l_2'$ 的奇偶性必须不同.

爱华的论文中将式 11 用 9j 符号表示，但仅限于 $M \neq M'$. 然而经过数值验证，我发现应该可以将爱华的公式拓展为

$$\left\langle \mathcal{Y}^{LM}_{l_1 l_2} \middle| \mathcal{Y}^{00}_{ll} \middle| \mathcal{Y}^{L'M'}_{l_1' l_2'} \right\rangle = \frac{2l+1}{4\pi}\delta_{L,L'}\delta_{M,M'}\sqrt{(2l_1'+1)(2l_2'+1)(2L+1)}\times \begin{bmatrix} l & l_1' & l_1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} l & l_2' & l_2 \\ 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} l & l_1' & l_1 \\ l & l_2' & l_2 \\ 0 & L & L \end{Bmatrix} \tag{13}$$

注意当 $M = M'$ 时，结果与 M 无关.

泊松积分

按照FEDVR基底的思路，我们会自然地想把 V_{12} 算符展开后的 $\langle r_i r_j | r_<^l / r_>^l | r_{i'} r_{j'} \rangle$ 用高斯 Lobatto[1926] 积分来计算，然而这么做并不精确，因为 $r_<^l / r_>^l$ 用有限阶多项式展开会有误差. 使用泊松积分法可以减小计算误差，且仍然保持对角矩阵的形式，即

$$\langle r_i r_j | \frac{r_<^l}{r_>^{l+1}} | r_{i'} r_{j'} \rangle \propto \delta_{i,i'}\delta_{j,j'} \tag{14}$$

根据我的猜测（有空验证一下！），泊松积分积分的结果应该和直接用数值积分的结果是一样的.

$$\langle r_i r_j | \frac{r_<^l}{r_>^{l+1}} | r_{i'} r_{j'} \rangle = \delta_{i,i'}\delta_{j,j'} \left[\frac{2l+1}{r_i r_j \sqrt{\omega_i \omega_j}} (2\mathbf{T}_l)^{-1}_{i,j} + \frac{r_i^l r_j^l}{r_{max}^{2l+1}} \right] \tag{15}$$

推导参考 McCurdy, Solving the three-body Coulomb breakup problem using exterior complex scaling 式 68．其中 $\mathbf{T}_l$ 是单电子动能算符的矩阵

$$\begin{aligned}(T_l)_{ij} &= \langle\phi_i|\left[-\frac{1}{2m}\frac{\mathrm{d}^2}{\mathrm{d}r^2}+\frac{l(l+1)}{2mr}\right]|\phi_j\rangle \\ &= -\frac{1}{2m}\langle\phi_i|\frac{\mathrm{d}^2}{\mathrm{d}r^2}|\phi_j\rangle+\delta_{i,j}\frac{l(l+1)}{2mr_i^2}\end{aligned} \tag{16}$$

其中 ϕ_i 是一维 FEDVR 格点在 r_i 处的正交归一基底[1928]，r_{max} 是计算 T_l 时最后一个 FE 的边界．虽然式 15 中含有 r_{max}，但结果却与 r_{max} 无关，原因是 T_l 也和 r_{max} 相关，两项中的 r_{max} 抵消了．

电场作用项

对每个电子，

$$V_F = \mathbf{E}\cdot\mathbf{r} = E_x x + E_y y + E_z z \tag{17}$$

与氢原子的情况相同，有

$$x=\sqrt{\frac{2\pi}{3}}r(Y_{1,-1}-Y_{1,1}) \qquad y=\mathrm{i}\sqrt{\frac{2\pi}{3}}r(Y_{1,-1}+Y_{1,1}) \qquad z=2\sqrt{\frac{\pi}{3}}rY_{1,0} \tag{18}$$

对线偏振电场我们只需要最后一项，以第一个电子为例，

$$
\begin{aligned}
\left\langle \mathcal{Y}^{LM}_{l_1l_2}\middle|z_1\middle|\mathcal{Y}^{L'M'}_{l_1'l_2'}\right\rangle &= 2\sqrt{\frac{\pi}{3}}r\sum_{m_1',m_2'}\sum_{m_1,m_2}\begin{bmatrix} l_1 & l_2 & L\\ m_1 & m_2 & M\end{bmatrix}\begin{bmatrix} l_1' & l_2 & L'\\ m_1' & m_2' & M'\end{bmatrix}\times\\
&\iint Y^*_{l_1m_1}(\hat{\mathbf{r}}_1)Y^*_{l_2m_2}(\hat{\mathbf{r}}_2)Y_{10}(\hat{\mathbf{r}}_1)Y_{l_1'm_1'}(\hat{\mathbf{r}}_1)Y_{l_2'm_2'}(\hat{\mathbf{r}}_2)\,\mathrm{d}\Omega_1\,\mathrm{d}\Omega_2\\
&= 2\sqrt{\frac{\pi}{3}}\delta_{l_2,l_2'}r\sum_{m_1,m_2,m_1'}\begin{bmatrix} l_1 & l_2 & L\\ m_1 & m_2 & M\end{bmatrix}\begin{bmatrix} l_1' & l_2' & L'\\ m_1' & m_2 & M'\end{bmatrix}\times\\
&\iint Y^*_{l_1m_1}(\hat{\mathbf{r}}_1)Y_{10}(\hat{\mathbf{r}}_1)Y_{l_1'm_1'}(\hat{\mathbf{r}}_1)\,\mathrm{d}\Omega_1\\
&= \delta_{l_2,l_2'}\sqrt{(2l_1+1)(2l_1'+1)}r\sum_{m_1,m_2,m_1'}(-1)^{m_1}\times\\
&\begin{bmatrix} l_1 & l_2 & L\\ m_1 & m_2 & M\end{bmatrix}\begin{bmatrix} l_1' & l_2 & L'\\ m_1' & m_2 & M'\end{bmatrix}\begin{pmatrix} l_1 & 1 & l_1'\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix} l_1 & 1 & l_1'\\ -m_1 & 0 & m_1'\end{pmatrix}\\
&= \delta_{l_2,l_2'}\delta_{M,M'}\sqrt{(2l_1+1)(2l_1'+1)}r\sum_{m_1,m_2}(-1)^{m_1}\times\\
&\begin{bmatrix} l_1 & l_2 & L\\ m_1 & m_2 & M\end{bmatrix}\begin{bmatrix} l_1' & l_2 & L'\\ m_1 & m_2 & M\end{bmatrix}\begin{pmatrix} l_1 & 1 & l_1'\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix} l_1 & 1 & l_1'\\ -m_1 & 0 & m_1\end{pmatrix}
\end{aligned}
\tag{19}
$$

最后一步使用了 3j 符号中 $m_1+m_2+m_3=0$，以及 CG 系数中 $m_1+m_2=M$ 的性质．两个 delta 函数验证了对第一个电子作用的 z 方向电场不会影响第二个电子的角动量以及总角动量的 z 分量. 推导中还陆续得出了 $m_1'=m_1, m_2'=m_2$ 的结论，即该电场同样不会影响每个电子的角动量 z 分量（所以对 m_1', m_2' 的求和消去了）. 最后对 m_1, m_2 的求和中，要求 $m_1+m_2=M$，且 $|m_1|<\min\{l_1,l_1'\}$.

同理，对第二个电子有

$$
\begin{aligned}
\left\langle \mathcal{Y}^{LM}_{l_1l_2}\middle|z_2\middle|\mathcal{Y}^{L'M'}_{l_1'l_2'}\right\rangle &= \delta_{l_1,l_1'}\delta_{M,M'}\sqrt{(2l_2+1)(2l_2'+1)}r\sum_{m_1,m_2}(-1)^{m_2}\times\\
&\begin{bmatrix} l_1 & l_2 & L\\ m_1 & m_2 & M\end{bmatrix}\begin{bmatrix} l_1 & l_2' & L'\\ m_1 & m_2 & M\end{bmatrix}\begin{pmatrix} l_2 & 1 & l_2'\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix} l_2 & 1 & l_2'\\ -m_2 & 0 & m_2\end{pmatrix}
\end{aligned}
\tag{20}
$$

在爱华的论文中，$M=M'=0$ 时用 9j 符号表示为（数值验证相同）

$$\left\langle \mathcal{Y}_{l_1 l_2}^{LM} \middle| z_1 \middle| \mathcal{Y}_{l_1' l_2'}^{L'M'} \right\rangle = r\sqrt{9(2l_1'+1)(2l_2'+1)(2L'+1)/(4\pi)^2} \times \begin{bmatrix} 1 & l_1' & l_1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & l_2' & l_2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & L' & L \\ 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} 1 & l_1' & l_1 \\ 0 & l_2' & l_2 \\ 1 & L' & L \end{Bmatrix} \tag{21}$$

$$\left\langle \mathcal{Y}_{l_1 l_2}^{LM} \middle| z_2 \middle| \mathcal{Y}_{l_1' l_2'}^{L'M'} \right\rangle = r\sqrt{9(2l_1'+1)(2l_2'+1)(2L'+1)/(4\pi)^2} \times \begin{bmatrix} 0 & l_1' & l_1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & l_2' & l_2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & L' & L \\ 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} 0 & l_1' & l_1 \\ 1 & l_2' & l_2 \\ 1 & L' & L \end{Bmatrix} \tag{22}$$

非线性偏振电场

对非线性偏振电场，式 17 的前两项不为零，但只和 z 分量的结果大同小异（只有最后一个 3j 系数变为两项，再除以 $\sqrt{2}$ 以归一化）

$$\left\langle \mathcal{Y}_{l_1 l_2}^{LM} \middle| x_1 \middle| \mathcal{Y}_{l_1' l_2'}^{L'M'} \right\rangle = \frac{\delta_{l_2,l_2'}\delta_{M,M'}}{\sqrt{2}} \sqrt{(2l_1+1)(2l_1'+1)}r\times \sum_{m_1,m_2} (-1)^{m_1} \begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{bmatrix} \begin{bmatrix} l_1' & l_2 & L' \\ m_1 & m_2 & M \end{bmatrix} \times \begin{pmatrix} l_1 & 1 & l_1' \\ 0 & 0 & 0 \end{pmatrix} \left[\begin{pmatrix} l_1 & 1 & l_1' \\ -m_1 & -1 & m_1 \end{pmatrix} - \begin{pmatrix} l_1 & 1 & l_1' \\ -m_1 & 1 & m_1 \end{pmatrix} \right] \tag{23}$$

$$\left\langle \mathcal{Y}_{l_1 l_2}^{LM} \middle| y_1 \middle| \mathcal{Y}_{l_1' l_2'}^{L'M'} \right\rangle = \frac{\mathrm{i}\delta_{l_2,l_2'}\delta_{M,M'}}{\sqrt{2}} \sqrt{(2l_1+1)(2l_1'+1)}r\times \sum_{m_1,m_2} (-1)^{m_1} \begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{bmatrix} \begin{bmatrix} l_1' & l_2 & L' \\ m_1 & m_2 & M \end{bmatrix} \times \begin{pmatrix} l_1 & 1 & l_1' \\ 0 & 0 & 0 \end{pmatrix} \left[\begin{pmatrix} l_1 & 1 & l_1' \\ -m_1 & -1 & m_1 \end{pmatrix} + \begin{pmatrix} l_1 & 1 & l_1' \\ -m_1 & 1 & m_1 \end{pmatrix} \right] \tag{24}$$

$$\left\langle \mathcal{Y}^{LM}_{l_1 l_2} \middle| x_2 \middle| \mathcal{Y}^{L'M'}_{l_1' l_2'} \right\rangle = \frac{\delta_{l_1,l_1'}\delta_{M,M'}}{\sqrt{2}} \sqrt{(2l_2+1)(2l_2'+1)}r \sum_{m_1,m_2} (-1)^{m_2} \begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{bmatrix} \begin{bmatrix} l_1 & l_2' & L' \\ m_1 & m_2 & M \end{bmatrix} \times \begin{pmatrix} l_2 & 1 & l_2' \\ 0 & 0 & 0 \end{pmatrix} \left[\begin{pmatrix} l_2 & 1 & l_2' \\ -m_2 & -1 & m_2 \end{pmatrix} - \begin{pmatrix} l_2 & 1 & l_2' \\ -m_2 & 0 & m_1 \end{pmatrix} \right] \tag{25}$$

$$\left\langle \mathcal{Y}^{LM}_{l_1 l_2} \middle| y_2 \middle| \mathcal{Y}^{L'M'}_{l_1' l_2'} \right\rangle = \frac{\mathrm{i}\delta_{l_1,l_1'}\delta_{M,M'}}{\sqrt{2}} \sqrt{(2l_2+1)(2l_2'+1)}r \sum_{m_1,m_2} (-1)^{m_2} \begin{bmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{bmatrix} \begin{bmatrix} l_1 & l_2' & L' \\ m_1 & m_2 & M \end{bmatrix} \times \begin{pmatrix} l_2 & 1 & l_2' \\ 0 & 0 & 0 \end{pmatrix} \left[\begin{pmatrix} l_2 & 1 & l_2' \\ -m_2 & -1 & m_2 \end{pmatrix} + \begin{pmatrix} l_2 & 1 & l_2' \\ -m_2 & 0 & m_1 \end{pmatrix} \right] \tag{26}$$

束缚态

能量较低的束缚态可以轻易用虚时间演化得到，即把 $\exp(-\mathrm{i}H\Delta t)$ 替换为 $\exp(-H\Delta t)$，这相当于把 Δt 变为虚数 $-\mathrm{i}\Delta t$，因此得名．氦原子基态只包含 $L=0$ 的分波，这要求 $l_1=l_2=0,1,2,\ldots$（参考图 2 [1403]）．

波函数演化

演化（propagation）同样使用 split operator

$$\exp(-\mathrm{i}H\Delta t) = \exp\left(-\mathrm{i}H_0\frac{\Delta t}{2}\right)\exp(-\mathrm{i}H_{int}\Delta t)\exp\left(-\mathrm{i}H_0\frac{\Delta t}{2}\right) + \mathcal{O}\left(\Delta t^3\right) \tag{27}$$

只是这里的 $H_{int}=V_{12}+V_F$ 是所有 $\langle\mathcal{Y}|\,|\mathcal{Y}\rangle$ 作用后为非对角（算符）矩阵的项，即电子的相互作用和电场对每个电子的作用．

$H_0=H_1+H_2$ 由于 H_1 和 H_2 对易，我们可以将它们独立演化

$$\exp\left(-\mathrm{i}H_0\frac{\Delta t}{2}\right) = \exp\left(-\mathrm{i}H_1\frac{\Delta t}{2}\right)\exp\left(-\mathrm{i}H_2\frac{\Delta t}{2}\right) \tag{28}$$

也就是对每个 partial wave 的二维网格，用类似氢原子的方法演化每一列再演化每一行．

根据以上的推导可以看出 H_{int} 在展开成径向算符与角向算符的张量积的线性组合后，径向算符都是关于 r 的函数，这对 FEDVR 基底来说都是对角矩阵（即使使用泊松积分）. 所以在演化 $\exp(-\mathrm{i}H_{int}\Delta t)$ 时，每个二维 FEDVR 格点都可以独立进行，给大量并行提供了可能性. 总结起来，程序中的**三维波函数矩阵的每个方向的演化都可以每列独立进行**.

在演化 $\exp(-\mathrm{i}H_{int}\Delta t)$ 的时候，爱华的做法是进一步 split 成三个 operators

$$\exp(-\mathrm{i}H_{int}\Delta t) = \exp\left(-\mathrm{i}V_F\frac{\Delta t}{2}\right)\exp(-\mathrm{i}V_{12}\Delta t)\exp\left(-\mathrm{i}V_F\frac{\Delta t}{2}\right) \tag{29}$$

这是因为 $V_F = V_{F1} + V_{F2}$ 是含时的，和电场分量 E_z 成正比.

爱华进一步分割的原因是，他的代码中预先储存了 V_F 和 V_{12} 对角化后的本征值和本征矢矩阵，如果直接对 H_{int} 对角化，那么就需要每一步都重新对角化一次. 预先储存所需的内存和分波数的平方成正比，然而当我们需要很多分波的时候，内存就远远不够了. 为了解决内存问题，我径向和角向都使用 Crank-Nicolson 演化，而且每个时间步长对每个格点重新生成 $V_F + V_{12}$ 矩阵并一次演化不需要分成三次. 对于波函数绝对值低于某个阈值的格点，我们甚至可以忽略不演化.

另外一点，如果只关心 single ionization，可以用一个长方形网格（见 Ossiander 2017 Nat. Phys.）. 但要注意在短边加入 absorber 防止反射.）

分波收敛分析

使用 $l_{1,max}, l_{2,max}, L_{max}$ 来生成分波列表有时候会有较大的浪费，会出现许多概率极小的分波. 最好的办法是把各个分波的概率画成图，从图中可以轻易看出分波是否收敛以及是否有浪费.

分波误差分析（原创）

我们截取有限个分波时，在分波演化（式 29）的过程中，会引入新的误差. 截取以后，V_F, V_{12} 仍然是厄米矩阵，所以他们的传播子同样是幺正的，所以不能用波函数的模长来估计误差. 所以可以用另一种方法来估计误差. 误差可以每隔几步估算一次，令 $\mathbf{V}'_F, \mathbf{V}'_{12}$ 为两倍基底数量的矩阵，那么可以计算 $(\mathrm{i}V'_F\Delta t)\mathbf{v}$ 比基底多出来的几个分波，对每个分波的绝对值进行求和.

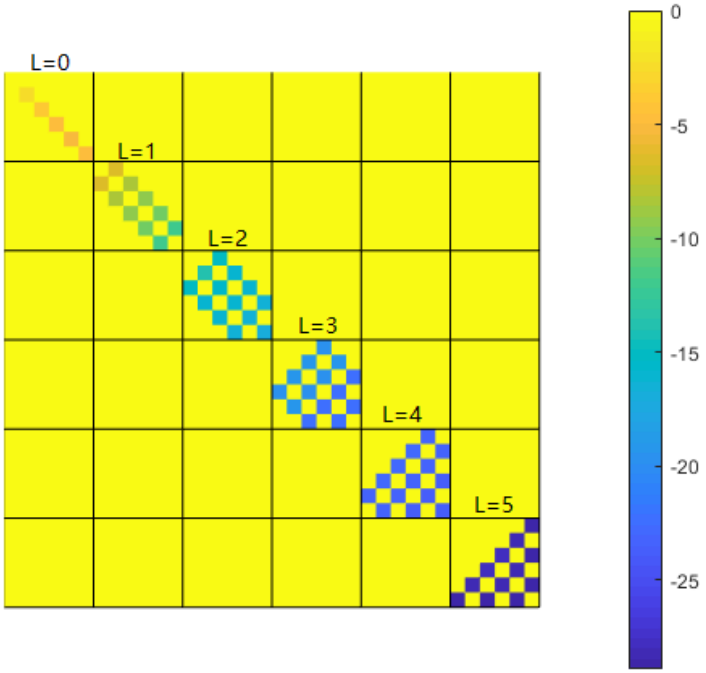


图 1: $l_{1,max}, l_{2,max}, L_{max} = 5$，可以看出 L 较小时需要更大 l_1, l_2，而 L 较大时浪费严重．理想情况是每个 L 中最小概率的分波都差不多．

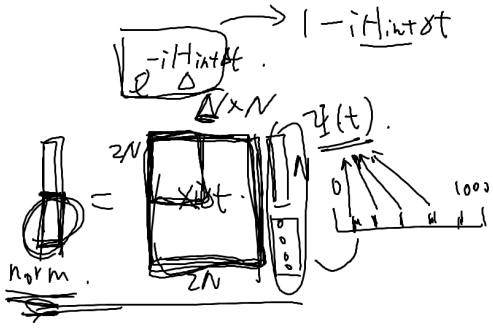


图 2: 分波误差分析

束缚态

理论上可以数值解总哈密顿矩阵的本征值，但实际上这个矩阵大到几乎不可能解出（事实上我们从来不会把总哈密顿矩阵直接储存，而是储存 split operator）．我们使用虚时间演化（imaginary time propagation）[1936]．

虚时演化的初始态我用的是两个 He+ 的基态相乘（实数波函数），和爱华

略有不同，不过这影响不大．由于我们的 split operator 都是实数矩阵，使用虚时间后其指数函数也是实数矩阵，所以虚时间演化得到的基态波函数也一定会是纯实数．我们只需用 $L = 0, M = 0$ 的所有 partial waves 即可．

氦原子波函数数值分析

预备知识 氦原子数值解 TDSE[1944]，库仑波函数[1431]，全同粒子[1454]

氦原子基本数据

- **parahelium**: singlet spin, 对称波函数，存在 1s 基态（我们使用的）．**orthohelium**: triplet spin, 反对称波函数，不存在 1s 基态，但其他束缚态的能量相对较低．
- **基态能量**（数值计算）：-2.9037244au （-79.014366 eV）
- **1s2s 能量**：-2.14597，**1s3s 能量**：-2.06127，**1s4s 能量**：-2.03359 （参考 Feist 毕业论文）
- **氦离子**基态能量：-2au（-54.422772eV）第一激发态 -0.5au（-13.605693eV）第二激发态 -2/9=-0.22...au（-6.0469747eV）
- **direct** （n=1）单电离 threshold：0.9037244au （24.591594eV）
- **shakeup** （n=2）单电离 threshold：2.4037244au （65.408673eV）
- **shakeup** （n=3）单电离 threshold：2.68150218au （72.967391eV）

90eV XUV 电离

- 90eV 的 xuv 光子能量为 3.3074390au
- 90eV xuv 的 direct（n=1）电离后电子的末动能为 2.4037146au（65.408405eV），末动量为 2.1925850au

- 90eV xuv 的 shakeup（n=2）电离后电子末动能为 0.9037146au（24.591326eV），末动量为 1.3444066au

- 90eV xuv 的 shakeup（n=3）电离后电子末动能为 0.62593682au（17.032608eV），末动量为 1.1188716au

能量

用虚时间法得到基态了以后，我们可以用平均能量来计算基态能量

$$E = \langle \Psi | H_0 | \Psi \rangle \tag{1}$$

动量谱（6 维）

这里考虑的是双电离，因为我们将要把两个电子都投影到连续态，如果任何一个电子处于基态，那么投影将为 0. 氢原子的散射态是库仑波函数[1431]，而氦原子的（双电子）散射态没有解析解，于是就假设时间足够大时，两电子间没有相互作用，就可以直接把两个库仑波函数做张量积再对称/反对称化即可. 而这样一来，散射态和束缚态之间显然就不正交了，所以分析时需要将所有束缚态去掉[1]. 爱华的方法是直接减去基态的投影，然而这样一来激发态的分量就仍然存在了. 我认为如果电离波包和未电离的波包已经分离的话，可以像做氢原子那样直接把一定半径内的波函数直接挖掉.

库仑波函数（见式 11 [1431]，我们只需要平面波出射）为（$\eta = Z/k, Z = -2$）

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sqrt{\frac{2}{\pi}} \sum_{l,m} \frac{\mathrm{i}^l}{rk} \mathrm{e}^{-\mathrm{i}\sigma_l(\eta)} F_l(\eta, kr) Y^*_{l,m}(\hat{\mathbf{k}}) Y_{l,m}(\hat{\mathbf{r}}) \tag{2}$$

令对称化算符为

$$\hat{S} f(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} [f(\mathbf{r}_1, \mathbf{r}_2) + f(\mathbf{r}_2, \mathbf{r}_1)] \tag{3}$$

[1]理论上束缚态不去掉也没关系，因为单电子的库仑球波函数和束缚态本来就是正交的. 这么做可能是考虑到数值误差？

那么对称化的散射态为

$$\begin{aligned}&\psi_{\mathbf{k}_1,\mathbf{k}_2}(\mathbf{r}_1,\mathbf{r}_2)=\hat{S}[\psi_{\mathbf{k}_1}(\mathbf{r}_1)\psi_{\mathbf{k}_2}(\mathbf{r}_2)]\\&=\hat{S}\frac{2}{\pi}\sum_{l_1,m_1}\sum_{l_2,m_2}\frac{\mathrm{i}^{l_1+l_2}}{r_1r_2k_1k_2}\times\\&[(\mathrm{e}^{-\mathrm{i}[\sigma_{l_1}(\eta_1)+\sigma_{l_2}(\eta_2)]}F_{l_1}(\eta_1,k_1r_1)F_{l_2}(\eta_2,k_2r_2)Y^*_{l_1,m_1}(\hat{\mathbf{k}}_1)Y^*_{l_2,m_2}(\hat{\mathbf{k}}_2))]\times\\&Y_{l_1,m_1}(\hat{\mathbf{r}}_1)Y_{l_2,m_2}(\hat{\mathbf{r}}_2)\end{aligned}\tag{4}$$

波函数（假设已经满足交换对称）为

$$\begin{aligned}\Psi(\mathbf{r}_1,\mathbf{r}_2)&=\sum_{L,M,l_1,l_2}\frac{1}{r_1r_2}\psi^{L,M}_{l_1,l_2}(r_1,r_2)\mathcal{Y}^{L,M}_{l_1,l_2}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2)\\&=\sum_{L,M,l_1,l_2}\frac{1}{r_1r_2}\psi^{L,M}_{l_1,l_2}(r_1,r_2)\sum_{m_1,m_2}\begin{bmatrix}l_1&l_2&L\\m_1&m_2&M\end{bmatrix}Y_{l_1,m_1}(\hat{\mathbf{r}}_1)Y_{l_2,m_2}(\hat{\mathbf{r}}_2)\end{aligned}\tag{5}$$

投影得（过程略，对称化算符变为 $\sqrt{2}$，两个 $Y_{l_1,m_1}(\hat{\mathbf{r}}_1)Y_{l_2,m_2}(\hat{\mathbf{r}}_2)$ 项积分会产生 2 个 δ 函数）

$$\langle\psi_{\mathbf{k}_1,\mathbf{k}_2}|\Psi\rangle=\frac{2\sqrt{2}}{\pi k_1k_2}\sum_{L,M,l_1,l_2}(-\mathrm{i})^{l_1+l_2}\mathrm{e}^{\mathrm{i}[\sigma_{l_1}(\eta_1)+\sigma_{l_2}(\eta_2)]}I^{L,M}_{l_1,l_2}(k_1,k_2)\mathcal{Y}^{L,M}_{l_1,l_2}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2)\tag{6}$$

其中

$$I^{L,M}_{l_1,l_2}(k_1,k_2)=\int_0^\infty\int_0^\infty F_{l_1}(\eta_1,k_1r_1)F_{l_2}(\eta_2,k_2r_2)\psi^{L,M}_{l_1,l_2}(r_1,r_2)\,\mathrm{d}r_1\,\mathrm{d}r_2\tag{7}$$

投影的平方就是动量分布

$$P(\mathbf{k}_1,\mathbf{k}_2)=|\langle\psi_{\mathbf{k}_1,\mathbf{k}_2}|\Psi\rangle|^2\tag{8}$$

如果固定两个矢量的方向，就可以画出一个二维动量/能量分布图了（注意要乘 $k_1^2k_2^2$）.

径向概率分布

$$P(r_1,r_2)=2\iint|\Psi(\mathbf{r}_1,\mathbf{r}_2)|^2r_1^2r_2^2\,\mathrm{d}\Omega_1\,\mathrm{d}\Omega_2\tag{9}$$

类比氢原子的情况（式 15 [1536]），令 λ 代表 l_1, l_2, L, M

$$\begin{aligned}
&\iint P(r_1, r_2)\,\mathrm{d}r_1\,\mathrm{d}r_2 \\
=&\iint 2\iint \left|\frac{1}{r_1 r_2}\sum_{\lambda}\psi_\lambda(r_1, r_2)\mathcal{Y}_\lambda(\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2)\right|^2 \mathrm{d}\Omega_1\,\mathrm{d}\Omega_2\, r_1^2 r_2^2\,\mathrm{d}r_1\,\mathrm{d}r_2 \\
=&\iint 2\sum_{\lambda}\left|\psi_\lambda(r_1, r_2)\right|^2 \mathrm{d}r_1\,\mathrm{d}r_2 = 2
\end{aligned} \tag{10}$$

所以

$$P(r_1, r_2) = 2\sum_{\lambda}\left|\psi_\lambda(r_1, r_2)\right|^2 \tag{11}$$

这就是 (r_1, r_2) 子空间中的总模长，等于子空间中每个基底的系数的模方和.

能量分布

先继续考虑双电离，式 8 中的概率归一化为（因为对全空间积分重复计算了）

$$\int P(\mathbf{k}_1, \mathbf{k}_2)\cdot k_1^2\,\mathrm{d}\Omega_{k1}\,\mathrm{d}k_1\cdot k_2^2\,\mathrm{d}\Omega_{k2}\,\mathrm{d}k_2 = 2 \tag{12}$$

所以动量绝对值的分布函数为

$$P(k_1, k_2) = k_1^2 k_2^2\int P(\mathbf{k}_1, \mathbf{k}_2)\,\mathrm{d}\Omega_{k1}\,\mathrm{d}\Omega_{k2} \tag{13}$$

能量分布函数为

$$P(E_1, E_2) = \frac{1}{k_1 k_2}P(k_1, k_2) \tag{14}$$

所以只需要将式 8 代入式 13 得

$$P(k_1, k_2) = \frac{8}{\pi^2}\sum_{L,M,l_1,l_2}\left|I_{l_1,l_2}^{L,M}(k_1, k_2)\right|^2 \tag{15}$$

其实该式的推导大可直接用球面波而不是投影到平面波再做角向积分，过程类比氢原子[1539].

双电离概率和截面

双电离总概率为

$$P_{DI}=\frac{1}{2}\int P(k_1,k_2)\,\mathrm{d}k_1\,\mathrm{d}k_2=\frac{1}{2}\int P(k_1,k_2)\,\mathrm{d}k_1\,\mathrm{d}k_2 \tag{16}$$

这里有一个假设就是两个库仑函数的张量积空间就是双电离的本征态空间，即双电子库伦函数所在的空间. 这从物理直觉上应该是成立的，但不知如何证明.

令整个波包中平均单位面积入射的光子数为 n，双电离总截面 σ 的定义是（σn 等于双电离总概率）

$$\sigma=\frac{P_{DI}}{n} \tag{17}$$

JAD（joint angular distribution）

这里的 JAD 并不是总的方向分布，而是某种能量分割的方向分布，即 $k_2-k_1=\Delta k$，所以积分中要有一个 delta 函数.

$$\begin{aligned}P_{JAD}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2)&=\int_0^\infty\int_0^\infty P(\mathbf{k}_1,\mathbf{k}_2)k_1^2k_2^2\delta(k_1-k_2+\Delta k)\,\mathrm{d}k_1\,\mathrm{d}k_2\\&=\frac{8}{\pi^2}\int_0^\infty\int_0^\infty\left|f(k_1,k_2,\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2)\right|^2\delta(k_1-k_2+\Delta k)\,\mathrm{d}k_1\,\mathrm{d}k_2\\&=\frac{8}{\pi^2}\int_0^\infty\left|f(k_1,k_1+\Delta k,\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2)\right|^2\mathrm{d}k_1\end{aligned} \tag{18}$$

其中 f 就是式 6 中的求和

$$f(k_1,k_2,\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2)=\sum_{L,M,l_1,l_2}(-\mathrm{i})^{l_1+l_2}\mathrm{e}^{\mathrm{i}[\sigma_{l_1}(\eta_1)+\sigma_{l_2}(\eta_2)]}I_{l_1,l_2}^{L,M}(k_1,k_2)\mathcal{Y}_{l_1,l_2}^{L,M}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2) \tag{19}$$

一般在同一平面上取 $\hat{\mathbf{k}}_1$ 和 $\hat{\mathbf{k}}_2$（$\phi_1=\phi_2$ 为常数，例如 0），这样就可以画一个横纵坐标分别为 θ_1 和 θ_2 的二维分布图 $P_{JAD}(\theta_1,\theta_2)$.

论文中常用一个参数 ε 来衡量能量分割，定义为较小的光电子末动能与总末动能之比

$$\varepsilon=\frac{\min\{E_1,E_2\}}{E_1+E_2}\qquad(0\leqslant\varepsilon\leqslant0.5) \tag{20}$$

从光子的角度理解 JAD：总能量 $E=k_1^2/2+k_2^2/2$，再加上 $k_2-k_1=\Delta k$ 就把两个光子的能量确定了，剩下的自由度只有角度. 但实际上从 $P(k_1,k_2)$ 的图中可以看出，$E=k_1^2/2+k_2^2/2$ 所表示的圆弧是有宽度的，所以相当于对这个宽度平均了一下.

TDCS（triple differential cross section）

类比微分截面的定义，二重微分截面可以定义为

$$\frac{\partial^2\sigma}{\partial\Omega_1\partial\Omega_2}=\frac{1}{n}\int P(\mathbf{k}_1,\mathbf{k}_2)k_1^2k_2^2\,\mathrm{d}k_1\,\mathrm{d}k_2 \tag{21}$$

三重微分截面可以定义为

$$\frac{\partial^3\sigma}{\partial E_1\partial\Omega_1\partial\Omega_2}=\frac{k_1}{n}\int P(\mathbf{k}_1,\mathbf{k}_2)k_2^2\,\mathrm{d}k_2 \tag{22}$$

爱华的论文中，固定 $\mathbf{k}_1$，令 $\mathbf{k}_2$ 在包含 $\mathbf{k}_1$ 的某平面上旋转，就得到了那个角向分布的实验图.

单光子吸收

类比氢原子吸收一个光子（一阶微扰）后有 $\Delta l=\pm1$ （可以想象成光子带有角动量量子数 1），氦原子从基态吸收一个光子后总角动量也会增加 1. 所以要考虑单电子吸收，我们可以只观察 $L=1$ 的所有分波，甚至在传播时可以用所谓的“weak field TDSE”[2]

$$\mathrm{i}\frac{\partial\Psi^{1P}}{\partial t}=H_0\Psi^{1P}+H_E\Psi^{1S}\mathrm{e}^{-\mathrm{i}E_0t} \tag{23}$$

其中 H_0 是无场的哈密顿，$H_E=E(t)(z_1+z_2)$ 是电场的哈密顿.

$$\Psi^{1P}(\mathbf{r}_1,\mathbf{r}_2,t)=\sum_{l_1,l_2}\frac{1}{r_1r_2}\psi^{1,0}_{l_1,l_2}(r_1,r_2,t)\mathcal{Y}^{1,0}_{l_1,l_2}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2) \tag{24}$$

吸收一个光子的概率就是对 $L=1$ 所有分波进行模方积分求和. 然而如果做 Streaking 的话，通常会吸收多个红外光子，所以还是需要许多分波.

角向基底变换

除了使用广义球谐函数作为角向基底外，另一个自然的选择就是两个球谐函数的乘积. 如果令总波函数为

$$\Psi(\mathbf{r}_1,\mathbf{r}_2)=\frac{1}{r_1r_2}\sum_{l_1,m_1,l_2,m_2}\psi^{m_1,m_2}_{l_1,l_2}(r_1,r_2)Y_{l_1,m_1}(\hat{\mathbf{r}}_1)Y_{l_2,m_2}(\hat{\mathbf{r}}_2) \tag{25}$$

[2]参考 Colgan 2001, J Phys. B: 34 (2001) L457-L466

那么由式 3 [1944] 和式 1 [771] 得

$$\psi_{l_1,l_2}^{m_1,m_2}(r_1,r_2)=\sum_{L,M}\begin{bmatrix}l_1 & l_2 & L\\ m_1 & m_2 & M\end{bmatrix}\psi_{l_1,l_2}^{L,M}(r_1,r_2) \tag{26}$$

若我们有一个 l_1,l_2,L,M 的列表和一个 l_1,l_2,m_1,m_2 的列表，该变换就可以用矩阵表示.

单电子概率分布

由例 1 [1454]，在 $\mathbf{r}_1$ 位置发现任意一个粒子的概率密度为

$$P(\mathbf{r}_1)=2\int|\Psi(\mathbf{r}_1,\mathbf{r}_2)|^2\,\mathrm{d}^3r_2 \tag{27}$$

若使用式 25 形式的波函数[3]，这个积分的形式就变得相当简单，先令 $\mathbf{r}_1$ 为常数，计算关于 r_2 的径向波函数

$$\begin{aligned}\psi_{l_2,m_2}(\mathbf{r}_1,r_2)&=\frac{1}{r_1}\sum_{l_1,m_1}\psi_{l_1,l_2}^{m_1,m_2}(r_1,r_2)Y_{l_1,m_1}(\hat{\mathbf{r}}_1)\\&=\frac{1}{r_1}\sum_{L,M,l_1}\begin{bmatrix}l_1 & l_2 & L\\ M-m_2 & m_2 & M\end{bmatrix}\psi_{l_1,l_2}^{L,M}(r_1,r_2)Y_{l_1,m_1}(\hat{\mathbf{r}}_1)\end{aligned} \tag{28}$$

这样式 25 就变为

$$\Psi(\mathbf{r}_1,\mathbf{r}_2)=\frac{1}{r_2}\sum_{l_2,m_2}\psi_{l_2,m_2}(\mathbf{r}_1,r_2)Y_{l_2,m_2}(\hat{\mathbf{r}}_2) \tag{29}$$

单电子概率为

$$\begin{aligned}P(\mathbf{r}_1)&=2\sum_{l_2,m_2}\int|\psi_{l_2,m_2}(\mathbf{r}_1,r_2)|^2\,\mathrm{d}r_2\\&=\frac{2}{r_1^2}\sum_{l_2,m_2}\int\left|\sum_{l_1,m_1}\psi_{l_1,l_2}^{m_1,m_2}(r_1,r_2)Y_{l_1,m_1}(\hat{\mathbf{r}}_1)\right|^2\mathrm{d}r_2\end{aligned} \tag{30}$$

[3]在程序中不一定需要先这么做，只需要将式 26 代入以下各式，在将 L,M,l_1,l_2 的求和提到最外面即可

单电离动量谱

在 $P(r_1, r_2)$ 分布中，单电离一般是可以与双电离区分开的，只需要取一个 box 即可：$r_1 < r_{1,max}$ ， $r_2 > r_{2,min}$

$$\langle \delta_{\mathbf{r}_1} | \langle C_{l_2,m_2}(k_2) | \Psi \rangle = \frac{1}{r_1} \sum_{l_1,m_1} \psi_{l_1,m_1}(r_1) \cdot Y_{l_1,m_1}(\hat{\mathbf{r}}_1) \tag{31}$$

$$\psi_{l_1,m_1}(r_1) = \int_{r_{2,min}}^{+\infty} \sqrt{\frac{2}{\pi}} F_{l_2}(\eta_2, k_2 r_2) \psi_{l_1,l_2}^{m_1,m_2}(r_1, r_2)\, \mathrm{d}r_2 \tag{32}$$

然而 streaking trace 需要的不是上式的概率而是某个方向的概率.我们使用式 13 [1431] 做线性组合即可

$$\langle \delta_{\mathbf{r}_1} | \langle \psi_{\mathbf{k}_2}^{(-)} | \Psi \rangle = \sum_{l_2,m_2} \frac{\mathrm{i}^{-l_2}}{k_2} \mathrm{e}^{\mathrm{i}\sigma_{l_2}} Y_{l_2,m_2}(\hat{\mathbf{k}}_2) \langle \delta_{\mathbf{r}_1} | \langle C_{l_2,m_2}(k_2) | \Psi \rangle \tag{33}$$

所以（式 31 代入式 33，整理成 $1/r_1 \sum \ldots Y_{l_1,m_1}(\mathbf{r}_1)$ 形式）

$$\begin{aligned} P(\mathbf{k}_2) &= \int_0^{r_{1,max}} \left| \langle \delta_{\mathbf{r}_1} | \langle \psi_{\mathbf{k}_2}^{(-)} | \Psi \rangle \right|^2 \mathrm{d}^3 r_1 \\ &= \sum_{l_1,m_1} \int_0^{r_{1,max}} \left| \sum_{l_2,m_2} \frac{\mathrm{i}^{-l_2}}{k_2} \mathrm{e}^{-\mathrm{i}\sigma_{l_2}} Y_{l_2,m_2}(\hat{\mathbf{k}}_2) \cdot \psi_{l_1,m_1}(r_1) \right|^2 \mathrm{d}r_1 \end{aligned} \tag{34}$$

Shakeup 电离

令归一化的基态为

$$\phi_{n_1,l_1,m_1}(\mathbf{r}_1) = R_{n_1,l_1}(r_1) Y_{l_1,m_1}(\hat{\mathbf{r}}_1) \tag{35}$$

归一化的散射态为

$$C_{l_2,m_2}(k_2, \mathbf{r}_2) = \frac{1}{r_2} \sqrt{\frac{2}{\pi}} F_{l_2}(\eta_2, k_2 r_2) Y_{l_2,m_2}(\hat{\mathbf{r}}_2) \tag{36}$$

以下的计算和单电子概率分布很类似，我们同样不需要太在意粒子交换对称的问题，把内积乘以 $\sqrt{2}$，概率乘以 2 即可．同样先把总波函数转换为式 25 的形式，投影得到

$$\langle \phi_{n_1,l_1,m_1} | \langle C_{l_2,m_2}(k_2) | \Psi \rangle = \sqrt{\frac{2}{\pi}} \iint r_1 R_{n_1,l_1}(r_1) F_{l_2}(\eta_2, k_2 r_2) \psi_{l_1,l_2}^{m_1,m_2}(r_1, r_2)\, \mathrm{d}r_1\, \mathrm{d}r_2 \tag{37}$$

然而，streaking trace 需要的不是上式的概率而是某个方向的概率．我们只需使用式 13 [1431] 做线性组合即可

$$\langle\phi_{n_1,l_1,m_1}|\,\langle\psi_{\mathbf{k}_2}^{(-)}|\Psi\rangle=\sum_{l_2,m_2}\frac{\mathrm{i}^{-l_2}}{k_2}\mathrm{e}^{\mathrm{i}\sigma_{l_2}}Y_{l_2,m_2}(\hat{\mathbf{k}}_2)\,\langle\phi_{n_1,l_1,m_1}|\,\langle C_{l_2,m_2}(k_2)|\Psi\rangle \tag{38}$$

$$P(n_1,l_1,m_1,\mathbf{k}_2)=\left|\langle\phi_{n_1,l_1,m_1}|\,\langle\psi_{\mathbf{k}_2}^{(-)}|\Psi\rangle\right|^2 \tag{39}$$

当 $M\equiv 0$ 时这里对 m_2 求和其实只有一项就是 $m_2=-m_1$，也是一个常数．理论上，如果把式 39 对所有 n_1,l_1,m_1 求和，应该和式 34 相等．

考虑探测器的立体角

$P(n_1,l_1,m_1,\mathbf{k}_2)$ 中，探测器实际捕获的是一个立体角为 Ω 的圆锥中的光电子，如果假设 $\Omega\to 0$，总概率为

$$\int P(n_1,l_1,m_1,\mathbf{k}_2)k_2^2\Omega\,\mathrm{d}k_2 \tag{40}$$

所以实际测到的光电子动量谱应该是

$$\Omega k_2^2P(n_1,l_1,m_1,k_2) \tag{41}$$

单电离的相对延迟分析

Osiander/Pazourek 对延迟的分析的解释是：单电离的瞬间，氦离子存在 electric dipole （几乎不随时间变化），对电离的电子产生了吸引力的作用．但似乎并没有解释如何从波函数画出这个 dipole．

显然，这不能用上述的单电子概率分布来画，因为在当前计算中电离率只有 1e-6 到 1e-5 的数量级．这说明总波函数主要还处于束缚态．我认为正确的方法是计算位置分布 $P(\mathbf{r}_1)=P(\mathbf{r}_1,\mathbf{r}_2)$，其中 $\mathbf{r}_2$ 可以取电离波包的平均位置或者对电离波包的所有位置积分

$$P(\mathbf{r}_1)=\int P(\mathbf{r}_1,\mathbf{r}_2)\,\mathrm{d}^3r_2 \tag{42}$$

或者，也可以计算 $P(\mathbf{r}_1)=P(\mathbf{r}_1,\mathbf{k}_2)$，其中 $\mathbf{k}_2$ 取电离波包的中心动量或者对波包的所有动量积分（要确保只包含电离波包）

$$P(\mathbf{r}_1)=\int P(\mathbf{r}_1,\mathbf{k}_2)\,\mathrm{d}^3k_2 \tag{43}$$

又或者对 streaking trace 上的某条竖线的 shakeup 或 direct 区间积分.

$$P(\mathbf{r}_1) = \int P(\mathbf{r}_1, \mathbf{k}_2)\, \mathrm{d}k_2 \tag{44}$$

我认为最后的做法是最靠谱的，毕竟 streaking trace 是动量谱.

下面讨论如何计算 $P(\mathbf{r}_1, \mathbf{k}_2)$. 可以先将波函数写成

$$\Psi(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{r_1} \sum_{l_1,m_1} \left[\frac{1}{r_2} \sum_{l_2,m_2} \psi_{l_1,l_2}^{m_1,m_2}(r_1, r_2) Y_{l_2,m_2}(\hat{\mathbf{r}}_2) \right] Y_{l_1,m_1}(\hat{\mathbf{r}}_1) \tag{45}$$

中括号中的内容逐项投影到库仑平面波 $|\mathbf{k}_2\rangle$ 上得 $\psi_{l_1,m_1}(r_1)$，按照式 14 [1431] 的方法即可. 于是

$$P(\mathbf{r}_1, \mathbf{k}_2) = 2 \left| \frac{1}{r_1} \sum_{l_1,m_1} \psi_{l_1,m_1}(r_1) Y_{l_1,m_1}(\hat{\mathbf{r}}_1) \right|^2 \tag{46}$$

乘以 2 是考虑了全同粒子.

计算结果：没有看出 dipole，很可能因为计算投影 $\psi_{l_1,m_1}(r_1)$ 的时候原子核附近的电子云贡献了很大一部分. 可能需要切除原子核以后再投影才行.

其实切除的话还不如直接计算 $P(\mathbf{r}_1, \mathbf{r}_2)$ 分布，保持 $\mathbf{r}_2$ 不变（可以取光电子波包的正中心）. 直接用式 45 的 2 倍模方即可（固定 $\mathbf{r}_2$）.

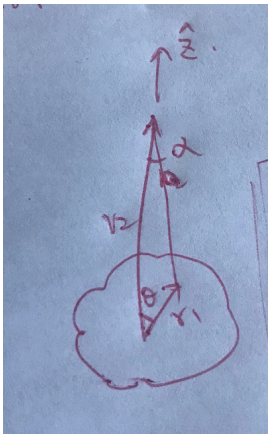


图 1: 经典力学计算延迟

用经典电磁学和力学来试图计算延迟：$P(\mathbf{r}_2)$ 位置分布是两个环形的 dipole，外面的环的代表 direct 电离，里面的环代表 shake-up. 那我们可以分别取 r2 为

z 方向外环中心和内环中心的位置，在每个时间 t 分别计算两个位置的 $P(\mathbf{r}_1)$ 分布，计算该电荷分布对 $\mathbf{r}$ 处的引力．然后比较该引力对两个经典力学轨迹的时间延迟只差即可．受力计算：

$$
\begin{aligned}
F_z(\mathbf{r}_2) &= \iiint \frac{P(\mathbf{r}_1)\cos\alpha}{r_{12}^2} r_1^2 \sin\theta_1 \,\mathrm{d}r_1 \,\mathrm{d}\theta_1 \,\mathrm{d}\phi_1 \\
r_{12}^2 &= r_1^2 + r_2^2 - 2r_1 r_2 \cos\theta_1 \\
\cos\alpha &= \frac{r_2^2 + r_{12}^2 - r_1^2}{2r_2 r_{12}}
\end{aligned}
\tag{47}
$$

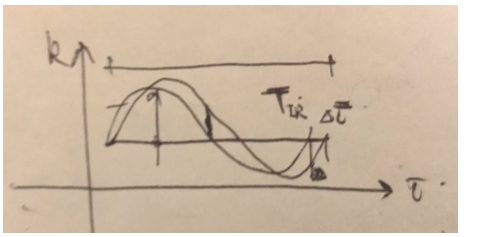


图 2: 估计 Δk

$$
\Delta k_max = k_0 \sin(2\pi\tau/T_{IR}) \approx 2\pi k_0 \tau / T_{IR} \tag{48}
$$

在 test800 中 $k_0 \approx 0.042$，$\tau = 16$，$T_{IR} = 110.245$．得 $\Delta k_max \approx 0.038$.

第二十部分

附录

第一章

附录

1966

小时百科符号与规范

这里列出本书中一些可能会产生歧义的符号以及规范，并给出对应词条的链接，如需补充，请与管理员协商.

数学通用

- 数域用双线字母表示，如实数域 $\mathbb{R}$、复数域 $\mathbb{C}$.
- 定义、定理等环境中的内容必须是严谨的，否则不使用.“证明”也必须是严谨的，否则用“不严谨的证明”或“幼稚的证明”.
- 较长的整数可以使用逗号对每三位进行分割，如 $2,456,789$.
- 集合[92] 中 $\subseteq$ 和 $\subset$ 表示子集，$\subsetneq$ 表示真子集，但应该尽量避免使用 $\subset$.
- 映射[98] 中单射，满射，双射的定义使用定义 1 [98].
- 虚数单位[142] 用正体的 i.
- 四元数中的三个虚数单位用 $\mathrm{i},\mathrm{j},\mathrm{k}$.
- 自然对数底[153] 用正体的 e.

线性代数

- 同义词：“矢量”、“向量”[370]；“矢量空间”、“线性空间”[444].
- 同义词：“线性映射”、“线性变换”、“线性算符”[451].
- 使用粗体与正体表示矩阵[396]，矩阵一般用大写（如 $\mathbf{M}$）.特殊地，列向量可以用小写（如线性方程组记为 $\mathbf{Ax}=\mathbf{b}$）.一个例外是当矩阵符号是希腊字母时，使用粗斜体而不是粗正体（如 $\boldsymbol{\beta}$）.这与 Wikipedia 的规范一致（例子）.
- 由于习惯原因，在不需要与列向量区分的情况下几何矢量[370] 也可以采用粗体加正体（大小写均可，如 $\mathbf{a},\mathbf{R}$）.

- 单位几何矢量在粗体与正体矢量上方加 hat 表示，如 $\hat{\mathbf{x}}$.
- 几何矢量的点成（内积）记为 $\mathbf{u}\cdot\mathbf{v}$，叉乘记为 $\mathbf{u}\times\mathbf{v}$.
- 矢量的坐标表示为实数或复数的笛卡尔积式 2 [92]（如 $\mathbb{R}^N, \mathbb{C}^N$ 或 (c_1, c_2, c_3)），也可以表示为列向量（如 $\mathbf{c}, \mathbf{x}$）以便和矩阵相乘.
- 为了排版美观，列向量出现在正文中时可以记为行向量的转置如 $\begin{pmatrix}1 & 2 & 3\end{pmatrix}^{\mathrm{T}}$ 注意没有逗号.
- 当需要区分矢量和坐标的不同时，矢量空间中的矢量不可以使用粗体正体，而是直接用普通变量字母（如 x, y, u, v），这时内积[471] 可以记为 $\langle u, v\rangle$.
- 也可以用狄拉克符号表示任意矢量空间的矢量，如 $|v\rangle$，对偶矢量如 $\langle v|$，内积如 $\langle u|v\rangle$.

数学分析

- 定义度量空间中序列**收敛**当且仅当它是柯西序列（定义 2 [790]）.

微分几何

- 同义词：“逆变向量（contravariant vectors）”、“切向量（tangent vectors）”、“$(1,0)$ 型张量”
- 同义词：“协变向量（covariant vectors）”、“余切向量（cotangent vectors）”、“切向量的对偶向量（dual vectors）”、“线性映射”、“线性 1-形式（linear 1-forms）”、“$(0,1)$ 型张量”
- 同义词：“张量（tensors）”、“多重线性映射（multilinear maps）”

物理通用

- 如无声明词条默认使用国际单位制[1973]，若使用其他单位制，要在每个词条开头用脚注声明“本文使用 xx 单位制”. 例如原子单位[1315]，厘米—克—秒[1975]，或高斯单位制[1976].
- 若要区分能量和电场，用 E 表示能量，$\mathcal{E}$ 表示电场.

电动力学

- 使用 ϵ 而不是 ε 表示电介质常量.

相对论

- 用 γ 表示 $1/\sqrt{1-v^2/c^2}$.
- 用 β 表示 v/c.

统计力学

- Ξ 表示配分函数.
- k_B 表示玻尔兹曼常数.
- 如果温度使用能量单位，意思是该温度（开尔文温标）乘以玻尔兹曼常数 k_BT.

常见物理量

预备知识 速度 加速度[881]，通量[297]

我们已知速度和加速度都由极限定义，事实上，许多其他物理量同样通过极限来定义.

密度

我们常把**质量密度**简称为**密度**，记为 ρ，定义为

$$\rho = \lim_{\Delta V \to 0} \frac{\Delta m}{\Delta V} \tag{1}$$

其中 Δm 是体积元 ΔV 内的质量. 当我们讨论的是宏观物体的密度时，这个极限的理解是，ΔV 远小于宏观尺度，却远大于微观尺度. 因为微观粒子（质子，电子等）的体积非常小，它们之间有很大的空间，如果按照数学定义取无限小，很大概率会得到密度为零.

类似地我们可以定义其他物理量的密度，如**电荷密度**（将上式的质量 m 换成净电荷 q），粒子数密度（将 m 换成粒子数 N）等.

流密度

（未完成）

国际单位制

[1]**国际单位制（SI Units）**是一套国际通用的单位标准.

本文采用 2019 年 5 月 20 日开始生效的新国际单位标准，该标准中的数值也被称为 2018 CODATA 推荐值. CODATA 是科学技术数据委员会（Committee on Data for Science and Technology）的简称. 在新标准中，所有基本单位都可以通过 7 个预定义的物理常数来测定（表 1 [1973]）. 以下的数值除了有特殊说明，都是精确值（无限位小数用省略号表示），不存在误差.

时间：秒（s）

一秒等于铯（Cs）原子 133 基态的超精细能级之间的跃迁辐射的电磁波周期的 $9,192,631,770$ 倍.

说明：我们知道原子中的电子具有不同的能级，当电子从一个能级跃迁到一个更低的能级时，会放出一个光子. 光子的频率为 $\nu = \varepsilon/h$，其中 ε 是光子的能量，h 为普朗克常数.

长度：米（m）

真空中，光在 $1/299792458$ 秒内传播的距离.

说明：由于真空中的光速是物质和信息能传播的最快速度（见狭义相对论[1253] 相关内容），且在任何参考系中都相同，所以可以作为一个精确的标准. 结合秒的定义，就可以通过实验得到一米的长度. 根据米的定义，一秒中光可以在真空中传播

$$c = 299792458\ \mathrm{m/s} \tag{1}$$

[1]参考 Wikipedia 相关页面.

质量：千克（kg）

千克的定义需要使得普朗克常数精确等于 $h = 6.62607015 \times 10^{-34}$ Js.

说明：2019 年 5 月开始，千克根据普朗克常数定义（见量子力学[??] 相关内容）. 这个定义可以类比“米”的定义（使光速精确地等于 299792459 m/s）：h 的单位 Js 也可以表示为 $\mathrm{kg \cdot m^2/s}$，我们已经定义了“米”和“秒”，所以通过测量普朗克常数，我们就可以定义千克.

历史上，千克最初在 1795 年被定义为一升水的质量，但在实际操作中会遇到许多困难使结果不太精确. 1799 年使用国际公斤原器的质量来定义，并复制若干份分别存放，但经过长时间后被发现和复制品存在细微误差.

力：牛顿（N）

等效于 $\mathrm{kg \cdot ms^{-2}}$. 等于使 1 kg 物体获得 $1\ \mathrm{m/s^2}$ 加速度的力.

说明：该定义符合牛顿第二定律（式 2 [899]）.

压强：帕斯卡（Pa）

使 $1\ \mathrm{m^2}$ 面积受力为 1 N 的压强.

能量：焦耳（J）

等效于 $\mathrm{kg \cdot ms^{-2}}$. 1 N 的恒力将受力物体沿力的方向移动 1 m，做功为 1 J.

电荷：库仑（C）

库伦的定义使得每个电子的电荷精确等于 $1.602176634 \times 10^{-10}$ C.

电势/电压：伏特（V）

等效于 J/C 或者 $\mathrm{kg \cdot m^2 s^{-2} C^{-1}}$. 伏特的定义使得 1 C 的电荷增加 1 V 电势，需要 1 J 的能量.

电场

等效于 $\mathrm{V/m}$ 或 $\mathrm{N/C}$ 或 $\mathrm{kg\cdot ms^{-2}C^{-1}}$.

电容：法拉（F）

等效于 C/V 或者 $\mathrm{C^2s^2kg^{-1}m^{-2}}$. 法拉是电容量的单位，一个 1 F 的电容器两端施加 1 V 电压，可以储存 1 C 净电荷.

电容率

单位 $\mathrm{F/m}$ 或者 $\mathrm{C^2s^2kg^{-1}m^{-3}}$.

电流：安培（A）

每秒钟流过某截面的的静电荷等于一库仑库仑.

磁场：特斯拉（T）

等效于 $\mathrm{kgC^{-1}s^{-1}}$. 可以由洛伦兹力（式 1 [1180]）或安培力（式 1 [1186]）来定义.

电感：亨利（H）

等效于 $\mathrm{kg\cdot m^2\cdot C^{-2}}$（未完成）

磁导率

单位 $\mathrm{H/m}$ 或者 $\mathrm{kg\cdot m\cdot C^{-2}}$.

温度：开尔文（K）

开尔文温度的定义应该使得玻尔兹曼常数精确等于

$$k_B = 1.3806505 \times 10^{-23}\ \mathrm{J/K} \tag{2}$$

例如，理想气体中分子平均动能为 $\langle E_k\rangle = 3k_BT/2$，理论上我们可以根据测其动能来定义温度.

物理学常数

预备知识 国际单位制[1970]

精确定义的常数

[2]2019 年 5 月 20 日生效的国际单位制[1970] 中精确定义了 7 个基本常数，每个基本单位的大小都可以由这些常数的定义测量出来. 如果某个数值最后两位有误差，我们把不确定度写在括号中，例如 1.23(45) 表示 1.23 ± 0.45. 另外约

表 1: 精确定义的常数

符号	精确值	名称
ν_{Cs}	$9,192,631,770$ Hz	铯原子 133 基态的超精细能级之间的跃迁辐射的电磁波频
c	$299,792,458$ m/s	真空中的光速
h	$6.62607015 \times 10^{-34}$ Js	普朗克常数
e	$1.602176634 \times 10^{-19}$ C	元电荷
k_B	1.380649×10^{-23} J/K	玻尔兹曼常数
N_A	$6.02214076 \times 10^{23}$	阿伏伽德罗常数
K_{cd}	683 Im/W	540 THz 电磁波的照射效率

化普朗克常数定义为 $\hbar = h/(2\pi)$.

未完成：Kcd 这个常数有什么用？

力学

万有引力常数[1027]

$$G \approx 6.67430(15) \times 10^{-11}\ \mathrm{m^3 kg^{-1} s^{-2}} \tag{1}$$

[2]本词条参考 Wikipedia 相关页面，以及 NIST 的 2018 CODATA 常数表

电动力学

真空磁导率[1175]

$$\mu_0 \approx 1.25663706212(19) \times 10^{-6}\ \mathrm{H/m} \tag{2}$$

在 2019 年更新前，它被定义为 $4\pi \times 10^{-7}\ \mathrm{H/m}$．现在 μ_0 可以通过测量精细结构常数[1321] 获得．

真空介电常数[1123]

$$\epsilon_0 = 1/(\mu_0 c^2) \approx 8.8541878128(13) \times 10^{-12} \tag{3}$$

相关词条：电容[1145]．在 2019 年更新前，它被定义为 $1/(\mu_0 c^2)$．

量子力学

玻尔半径[1313]

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{m_e e^2} = \frac{\hbar}{\alpha m_e c} \approx 5.29177210903(80) \times 10^{-11}\ \mathrm{m} \tag{4}$$

精细结构常数[1321]

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} \approx 7.2973525693(11) \times 10^{-3} \tag{5}$$

玻尔磁子[1243]

$$\mu_B = \frac{e\hbar}{2m_e} = 9.2740100783(28) \times 10^{-24}\ \mathrm{J/T} \tag{6}$$

原子质量单位（仍然由 C12 定义）

$$m_u = 1.66053906660(50) \times 10^{-27}\ \mathrm{kg} \tag{7}$$

电子质量

$$m_e = 9.1093837015(28) \times 10^{-31} \tag{8}$$

质子质量

$$m_p = 1.67262192369(51) \times 10^{-27}\ \mathrm{kg} \tag{9}$$

中子质量

$$m_n = 1.67492749804(95) \times 10^{-27}\ \mathrm{kg} \tag{10}$$

统计力学

理想气体常数（精确）[1560]

$$R = k_B N_A = 8.31447165136438\ \mathrm{J/K} \tag{11}$$

斯特藩—玻尔兹曼常数（精确）[1595]

$$\sigma = \frac{2\pi^5 k_B^4}{15c^2h^3} = 5.670374419\cdots\times 10^{-8}\ \mathrm{Wm^{-2}K^{-4}} \tag{12}$$

国际单位制词头

预备知识 国际单位制[1973]

[3]在使用国际单位时，我们会在一些单位符号或全称前面加上一个表示数量级的前缀，正式的名字叫做**国际单位制词头（metric prefix）**以方便书写．如长度单位 m（米）可以加不同的词头拓展成更小的单位 cm（厘米，centimeter），mm（毫米，millimeter），km（千米，kilometer）．

国际单位制共有 20 个词头，下面我们来看每个词头的全称和代表的数量级（表 1）．

厘米—克—秒单位制

- 本词条处于草稿阶段．
- 高斯单位制和 CGS 单位有什么区别？

预备知识 国际单位制[1970]，物理量和单位转换[873]

[4]**厘米—克—秒单位制（centimetre-gram-second system of units）**也简记为 CGS 单位制．我们将从 CGS 到 SI 单位制之间的转换常数记为 $\beta_{\text{物理量}}$，CGS

[3]参考 Wikipedia 相关页面．

[4]参考 Wikipedia 相关页面．

表 1: 词头列表

全称	词头	数量级	全称	词头	数量级
分（deci）	d	10^{-1}	十（deca）	da	10^{1}
厘（centi）	c	10^{-2}	百（hecto）	h	10^{2}
毫（milli）	m	10^{-3}	千（killo）	k	10^{3}
微（micro）	μ	10^{-6}	兆（mega）	M	10^{6}
纳（nano）	n	10^{-9}	吉（giga）	G	10^{9}
皮（pico）	p	10^{-12}	太（tera）	T	10^{12}
飞（femto）	f	10^{-15}	拍（peta）	P	10^{15}
阿（atto）	a	10^{-18}	艾（exa）	E	10^{18}
仄（zepto）	z	10^{-21}	泽（zetta）	Z	10^{21}
幺（yocto）	y	10^{-24}	尧（yotta）	Y	10^{24}

单位制的物理量符号用角标 c 加以区分．例如 $x = \beta_x x_c$ 其中 x 是国际单位制的物理量，x_c 是 CGS 单位的物理量．

则 $\beta_x = 0.01\ \mathrm{m/cm}$，$\beta_m = 0.001\ \mathrm{kg/\ g}$，$\beta_t = 1$．为了满足 $a_c = \ddot{x}_c$，有

$$\beta_a = \beta_x/\beta_s^2 = \beta_x = 0.01\ \mathrm{m/cm} \tag{1}$$

为了满足牛顿定律 $F_c = m_c a_c$ 得

$$\beta_F = \beta_m \beta_x/\beta_s^2 = \beta_m \beta_x = 10^{-5}\ \frac{\mathrm{kg \cdot m}}{\mathrm{g \cdot cm}} \tag{2}$$

为了满足 $E = Fs$，得

$$\beta_E = \beta_F \beta_x = \beta_m \beta_x^2 = 10^{-7}\ \frac{\mathrm{kg \cdot m^2}}{\mathrm{g \cdot cm^2}} \tag{3}$$

力的单位叫做**达因（dyne）**，记作 dyn．注意以上的转换常数本质上全部等于 1．

高斯单位制

预备知识 厘米—克—秒单位制[1975]

高斯单位制是在 CGS 单位制[1975] 的基础上添加了一些电磁学相关的单位. 以下先来定义各个物理量的转换常数，已知 CGS 的转换常数为

$$\begin{aligned} \beta_x &= 0.01\ \mathrm{m/cm} & \beta_m &= 0.001\ \mathrm{kg/\ g} \\ \beta_t &= 1 & \beta_F &= \beta_x\beta_m = 10^{-5}\ \mathrm{N/dyn} \end{aligned} \tag{1}$$

电荷

为了满足式 5，与国际单位相比，高斯单位制中电荷的量纲发生变化，单位为 $\sqrt{\mathrm{cm^3 g}}/\mathrm{s}$，为了方便我们不妨记为[5] C_g. 则电荷的转换常数为

$$\beta_q = \sqrt{4\pi\epsilon_0\beta_F}\beta_x \approx 3.3356409510736 \times 10^{-10}\ \mathrm{C/C_g} \tag{2}$$

假如把 C 和 $\mathrm{C_g}$ 看成是相同的量纲，则以下除了磁场 β_B 外其他转换常数都是无量纲的 1，β_B 量纲为速度分之一.

电场

为满足式 6，电场单位为 $\sqrt{\mathrm{g/cm}}/\mathrm{s} = \mathrm{C_g/cm^2}$. 转换常数为

$$\beta_{\mathcal{E}} = \frac{\beta_F}{\beta_q} = \sqrt{\frac{\beta_m}{4\pi\epsilon_0\beta_x}} \approx 2.997924580815998 \times 10^4\ \frac{\mathrm{C_g \cdot m \cdot kg}}{\mathrm{C \cdot cm \cdot g}} \tag{3}$$

若采用 2020 年 5 月以前的国际单位标准[1970]，数值上等于 $10^{-4}c_s$，c_s 是国际单位的光速.

磁场

为满足式 6，电场和磁场应具有相同的单位 $\mathrm{C_g/cm^2}$. 转换常数为[6]

$$\beta_B = \frac{\beta_m\beta_x}{c_s\beta_q} = \frac{\beta_{\mathcal{E}}}{c_s} \approx 1.000000000272 \times 10^{-4}\ \frac{\mathrm{C_g \cdot kg \cdot s}}{\mathrm{C \cdot g \cdot m}} \tag{4}$$

其中 c_s 是国际单位的光速. 若采用旧国际单位，该数值精确等于 10^{-4}.

[5] C_g 是笔者发明的记号，方便理解记忆.

[6] 2020 新国际单位标准[1970] 以前，这个数值精确等于 1/100, 新标准需要乘以 $\sqrt{\mu_0/(4\times10^{-7}\pi)}$ 的数值.

高斯单位制公式

高斯单位制下的电磁学公式比其国际单位要更简洁对称．其中 c_c 为 CGS 单位下的光速[7]．

$$F = \frac{q_1 q_2}{r^2} \qquad \text{（库伦定律）} \tag{5}$$

$$\mathbf{F} = q\mathbf{E} + \frac{q}{c_c}\mathbf{v} \times \mathbf{B} \qquad \text{（广义洛伦兹力）} \tag{6}$$

$$\boldsymbol{\nabla} \times \mathbf{A} = \mathbf{B} \qquad \text{（磁矢势）} \tag{7}$$

$$\mathbf{E} = -\boldsymbol{\nabla}\varphi - \frac{1}{c_c}\frac{\partial \mathbf{A}}{\partial t} \qquad \text{（标量势）} \tag{8}$$

$$\mathbf{A} = \mathbf{A}' + \boldsymbol{\nabla}\chi \qquad \varphi = \varphi' - \frac{1}{c_c}\frac{\partial \chi}{\partial t} \qquad \text{（规范变换）} \tag{9}$$

真空中的平面电磁波[1230] 若用高斯单位表示，有 $E_0 = B_0$．

$$\rho_E = \frac{1}{8\pi}(\mathbf{E}^2 + \mathbf{B}^2) \qquad \text{（场能量密度）} \tag{10}$$

$$\mathbf{s} = \frac{c_c}{4\pi}\mathbf{E} \times \mathbf{B} \qquad \text{（坡印廷矢量）} \tag{11}$$

$$\epsilon_0 = \frac{1}{4\pi} \qquad \mu_0 = 4\pi \tag{12}$$

转换常数为 $\beta_\epsilon = 4\pi\epsilon_0$，$\beta_\mu = \mu_0/(4\pi)$．

小时百科图标

- 本词条存在未完成的内容．

小时百科的图标是由 4 条丝带组成 3D 立体图，从数学上来看，每条丝带的形状分别是一个高斯分布（正态分布）[731]；物理上，四条丝带可以分别看作量子力学中自由粒子的高斯波函数随时间演化[1346]；艺术上，可以寓意为海浪．

目前图标分为立体版本（图 1 ）和扁平化版本（图 2 ）．

[7]事实上，如果高斯单位直接建立在国际单位制上（式 1 中的转换常数全部改为 1），以下公式同样成立（c_c 也要换成 c_s）．

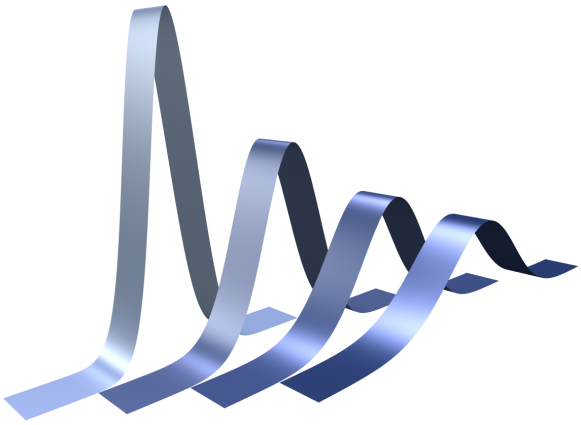


图 1: 立体图标，完成于 2014 年 10 月

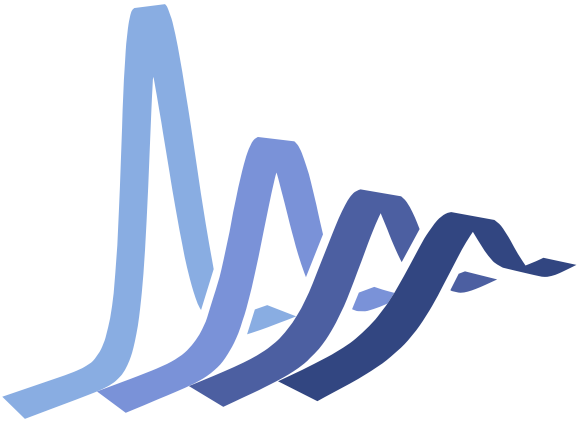


图 2: 扁平化图标，完成于 2020 年 9 月

参数

图标使用透视投影[1870]，每个带的长宽高分别在图 3 中标出.

未完成：量子力学参数，相机参数，光源材质参数，Matlab 代码

参考文献

[1] J. Hass, C. Heil, M. Weir, *Thomas' Cauculus* 14ed

[2] 同济大学数学系. 高等数学第六版

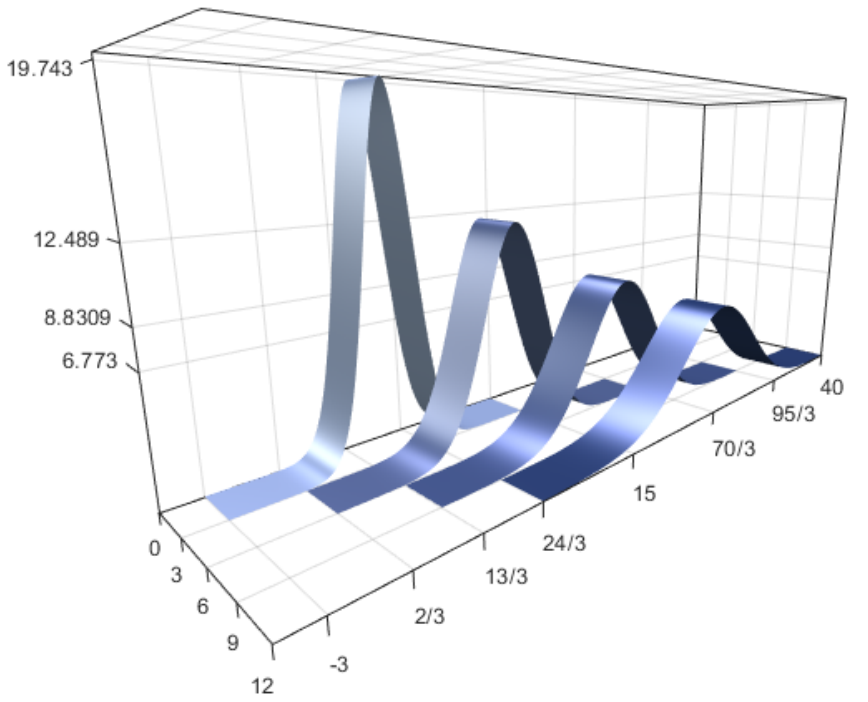


图 3: 图标参数

[3] Sheldon Axler. *Linear Algebra Done Right* 3ed

[4] 同济大学数学系. 线性代数第五版

[5] 梁灿彬, 周彬. 微分几何与广义相对论第二版

[6] Loring W. Tu. *Differential Geometry*: *Connections, Curvature, and Characteristic Classes*, GTM 275, Springer press.

[7] Arfken, Weber, Harris. *Mathematical Methods for Physicists - A Comprehensive Guide* 7ed

[8] Walter Rudin. *Principle of Mathematical Analysis*

[9] Eberhard Zeidler. *Applied Functional Analysis - Applications to Mathematical Physics*

[10] Herbert Goldstein. *Classical Mechanics* 3ed

[11] 赵凯华, 罗蔚茵. 新概念物理教程力学第二版

[12] David Griffiths, *Introduction to Electrodynamics*, 4ed

[13] 赵凯华, 陈熙谋. 新概念物理教程电磁学第二版

[14] David Griffiths, *Introduction to Quantum Mechanics*, 4ed

[15] 赵凯华, 罗蔚茵. 新概念物理教程量子物理第二版

[16] R. Shankar. *Principles of Quantum Mechanics* 2ed

[17] Eugen Merzbacher. *Quantum Mechanics* 3ed

[18] J.J. Sakurai. *Modern Quantum Mechanics* Revised Edition

[19] Brandsen, *Physics of Atoms and Molecules*, 2ed

[20] Daniel V. Schroeder, *An Introduction to Thermal Physics*

[21] 赵凯华, 罗蔚茵. 新概念物理教程热学第二版

[22] P. Peebles *Principles of Physical Cosmology*

[23] Peskin. *An introduction To Quantum Field Theory*

[24] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*，Cambridge University Press.

[25] John Baez, Javier P. Muniain. *Gauge Fields, Knots and Gravity*, Series on Knots and Everthing-Vol. 4, World Scientific press. ISBN-13: 978-981-02-2034-1.

[26] 小时百科志愿者. 小时百科. https://wuli.wiki.