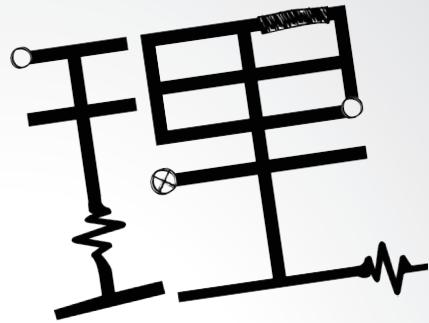
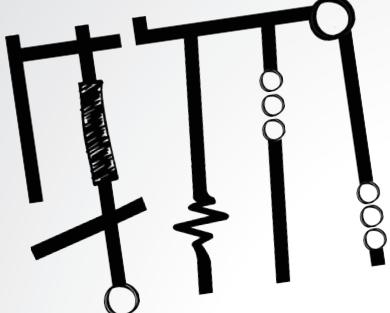


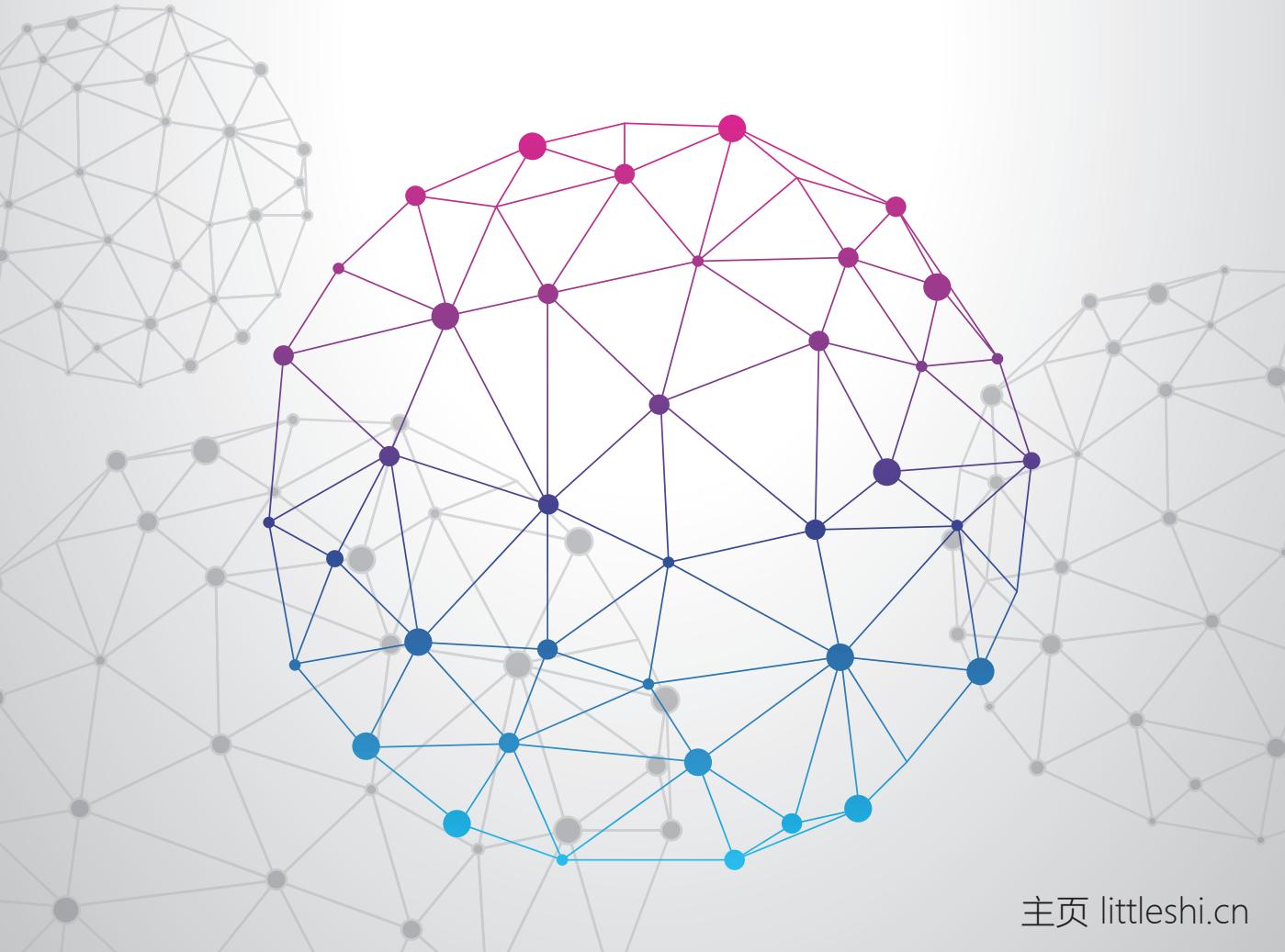
小时



百科

力学与相关数学

小时编



谨以此页向所有尊重版权的读者致敬！

版权声明

由于电子书的种种优势，我一直对电子书有一种独特的感情，希望有一天所有的中文教材都能实现电子化。然而遗憾的是国内的电子教材尚未普及，这里面可能的原因包括版权保护，正版销售平台和设备的普及率低等。

而另一方面，即使使用正规平台销售的电子教材，相对于 PDF 文件而言也还有相当的局限性，包括只能在特定的软件中阅读，注释的种类有限，不支持所有的设备和操作系统，不支持打印等。

考虑到以上的种种原因，我最后决定永久提供本书 PDF 的免费下载，并将捐赠及自觉付款和捐款作为主要的收入方式。这是一个非常艰巨的决定，因为本书的写作需要花费巨大的时间¹ 和金钱²，另一方面，本书的特色之一就是能利用 PDF 文件的普及性，注释功能，超链接功能和检索功能等优势。

在未来一段时间内，本书都会仍处于创作阶段（以下称为非正式版），内容上有许多不完整。对于非正式版，读者可以免费使用，且在不修改文件的条件下自由拷贝，但也鼓励对本书捐款。从本书首个正式版起，将要求每位读者在试读累计 3 小时后以自觉付款的形式购买本书，否则所持有的任何拷贝都视为盗版。

若发现本书内容有误，请先核对本书最新版核对，再与作者联系。若有任何建议也欢迎联系。最后，若以任何方式引用本书内容或插图，请注明出处并给出链接，在任何出版物上使用本书内容需经过作者同意。

本书的最新版本，作者的联系方式及捐款方式见本书网站 littleshicn。

¹保守估计，目前花费的时间在 1000 小时以上

²包括 TeX 排版费，以及购买 Microsoft Office，MathType，Adobe Illustrator，Matlab，Mathematica 等软件的费用

关于本书

1. 内容

与真正的百科（例如维基百科）不同的是，《小时物理百科》系列更偏向于教材而不仅仅是一本供专业人士查阅的工具书。本书的主要面向的人群是：

1. 具有一定的高中数学物理基础，但没有上过任何大学数理课程，想自学大学物理的读者。
2. 正在上相关大学物理课程的读者（作为参考书或预习使用）。
3. 本科已经毕业，想重温或巩固本科内容的读者。
4. 已学过本书内容，但需要经常查阅相关公式定理的读者。

本书作为《小时物理百科》系列的第一本（力学分册），涵盖了物理专业本科生所学的力学课程（通常是第一门专业物理课程）及少许理论力学课程（以下统称为**经典力学**）的内容，以及书中所需的所有超出高考大纲的数学内容（高等数学，线性代数等）。

本书可以帮助读者迅速了解力学的理论框架，配有若干例题，但暂不设习题，不设物理学史等拓展内容，因此不能代替相关课程的教材。

全书分为两个部分，数学部分介绍了力学部分涉及的任何超出高中数学的内容，包括数学拾遗，一元微积分，线性代数，多元微积分与计算物理六章。与标准的本科高等数学教材（如同济大学的《高等数学》和《线性代数》）相比，数学部分忽略了许多细节，也忽略了一些物理部分不需要使用的内容。这样，读者可以尽早开始学习物理而无需花费太多时间在数学上。力学部分主要讲解了牛顿的经典力学，包括质点，质点系与刚体，振动与波动，天体运动与中心力场，理论力学五章。

2. 结构和特点

与其他教材不同，本书更接近于百科的形式，由许多词条构成。一个词条从半页到几页不等，可能包括“预备知识”，“结论”，“推导”，“例”，“应用实例”，“拓展阅读”等不同的组成部分。理论上来说，读者可以直接跳到最感兴趣的词条，如果“预备知识”里面列出的词条都已经掌握，就可以开始学

习该词条³, 否则就先掌握“预备知识”里面的词条, 以此类推. 这种学习方法的好处在于无需按照课程的顺序学习, 因此本书在结构上具有很大的优势. 读者甚至可以根据自己的目标词条, 画出一个知识结构图, 这样就可以对知识结构做到一览无余, 用最高的效率攀登目标词条.

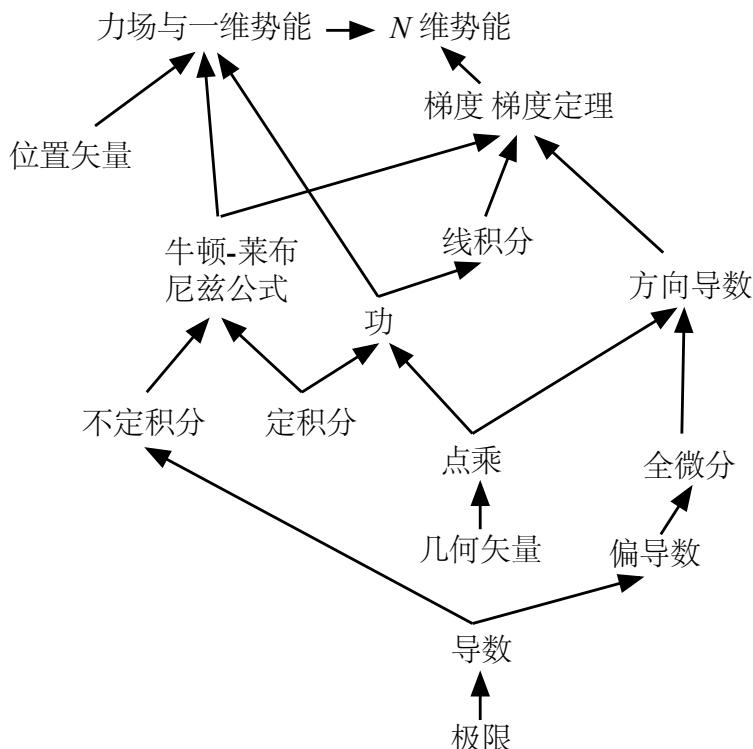


图 1: 示例: 由“预备知识”画出的知识结构图 (目标词条为“力场 势能^[166]”)

需要注意的是, 由于本书内容繁多, 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, 而是应该以每一章节给出的“导航”(第一个词条)为主线来学习, 再根据兴趣和需要阅读其余词条.

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^[34]”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^[34]”的格式, 右上角的方括号是公式所在的页码. 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

³如果“预备知识”出现在词条开始, 则必须先掌握, 如果出现在某个小标题下方, 则只有阅读该部分时需要掌握

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。

4. 更新

本书更新速度较快，为保证阅读质量，请及时下载最新版。下载链接见本书网站 littleshicn.com。如果网页出现故障，请发邮件到 270174408@qq.com 报错。

经典力学及其他物理理论

物理学理论的可证伪性

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。毕竟人们永远也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⁴，这就是物理学理论的可证伪性。

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（如牛顿力学，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 经典力学的价值

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，经典力学是一个错误的理论。

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（叫做量子场论，**Quantum Field Theory**）。

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

⁴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

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⁵。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提出的原理才是正确的，而经典力学的原理就是错误的。

既然原理不对，应用范围又相对窄，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，它比相对论简单得多。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定律的关系。量子力学除了经典力学的范围，还包括了“微观”范围。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，提出的一些概念并不正确，公式也只是一种近似。一些“民间科学家”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”。目前，“相对论量子力学”的理论还并不完善，是许多理论物理学家努力的方向。

⁵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。

目录

第一部分 数学

第一章 数学拾遗

二项式定理 [3] 二项式定理(非整数幂) [4] 三角恒等式 [4] 双曲函数 [6]
充分必要条件 [7] 极坐标系 [9] 柱坐标系 [10] 球坐标系 [11] 球坐标与
直角坐标的转换 [13] 圆锥曲线的极坐标方程 [14] 椭圆的三种定义 [16]
双曲线的三种定义 [18] 复数 [19] 复变函数 [22] 指数函数(复数) [23]
三角函数(复数) [24]

第二章 一元微积分

微积分导航 [28] 极限 [29] 小角正弦极限 [32] 自然对数底 [33] 切线与
割线 [34] 导数 [34] 求导法则 [37] 反函数求导 [39] 基本初等函数的导
数 [41] 导数与函数极值 [44] 用极值点确定函数图像 [46] 一元函数的微
分 [47] 复合函数求导(链式法则) [48] 泰勒展开 [50] 导数与差分 [53]
不定积分 [54] 换元积分法 [56] 分部积分法 [58] 积分表 [59] 定积分 [65]
牛顿—莱布尼兹公式 [69] 常微分方程 [71] 一阶线性微分方程 [72] 二
阶常系数齐次微分方程 [73] 二阶常系数非齐次微分方程 [75] 正交函数
系 [77] 傅里叶级数(三角) [78] 傅里叶级数(指数) [83]

第三章 线性代数

线性代数导航 [87] 几何矢量 [88] 矢量点乘 [91] 正交归一基底 [94] 右
手定则 [95] 矢量叉乘 [96] 矢量叉乘分配律的几何证明 [99] 连续叉乘的
化简 [100] 三矢量的混合积 [101] 平面旋转变换 [102] 线性变换 [104] 矩
阵 [105] 单位正交阵 [109] 平面旋转矩阵 [110] 空间旋转矩阵 [112] 绕轴
旋转矩阵 [112] 行列式 [115]

第四章 多元微积分

偏导数 [118] 二元函数的极值 [119] 最小二乘法 [120] 全微分 [122] 复合

函数的偏导 链式法则 [124] 全导数 [125] 矢量的导数 求导法则 [126] 一元矢量函数的积分 [128] 方向导数 [130] 重积分 [132] 矢量场 [135] 极坐标中单位矢量的偏导 [136] 线积分 [137] 梯度 梯度定理 [139] 散度 散度定理 [143]

第五章 概率与统计

随机变量的变换 [149]

第二部分 力学

第一章 质点

位置矢量 位移 [153] 速度 加速度（一维） [154] 速度 加速度 [156] 匀速圆周运动的速度（几何法） [157] 匀速圆周运动的速度（求导法） [158] 匀速圆周运动的加速度（几何法） [159] 匀速圆周运动的加速度（求导法） [160] 匀加速运动 [160] 极坐标中的速度和加速度 [161] 牛顿运动定律 惯性系 [162] 功 功率 [164] 动能 动能定理（单个质点） [165] 力场 势能 [166] 机械能守恒（单个质点） [170] 动量 动量定理（单个质点） [170] 角动量定理 角动量守恒（单个质点） [171] 简谐振子 [172] 受阻落体 [173] 单摆 [175] 傅科摆 [177] 惯性力 [178] 离心力 [181] 科里奥利力 [182] 地球表面的科里奥利力 [184]

第二章 质点系与刚体

质点系 [187] 质心 质心系 [187] 刚体 [189] 二体系统 [191] 二体碰撞 [192] 质点系的动量 [194] 动量定理 动量守恒 [195] 质点系的动能 柯尼西定理 [196] 力矩 [197] 刚体的静力平衡 [199] 角动量 [200] 角动量定理 角动量守恒 [202] 刚体的绕轴转动 转动惯量 [203] 平行轴定理与垂直轴定理 [205] 常见几何体的转动惯量 [206] 刚体的平面运动方程 [209] 浮力 [211]

第三章 振动与波动

振动的指数形式 [214] 受阻简谐振子 [215] 简谐振子受迫运动 [217] 平面波 [218] 一维波动方程 [221]

第四章 天体运动与中心力场

万有引力 引力势能 [225] 开普勒三定律 [226] 开普勒问题 [228] 反开普勒问题 [229] 卢瑟福散射 [230] 开普勒第一定律的证明 [231] 开普勒第二定律的证明 [233] 开普勒第三定律的证明 [234] 比耐公式 [235]

第三部分 计算物理

第一章 Matlab 及其他工具

计算物理导航 [238] Matlab 简介 [239] Matlab 的变量与矩阵 [242] Matlab 的判断与循环 [253] Matlab 的函数 [257] Matlab 画图 [261] Matlab 的程序调试及其他功能 [263]

第二章 数值验证及常用算法

二项式定理（非整数幂）的数值验证 [268] 二分法 [269] 多区间二分法 [270] 冒泡法 [273] Nelder-Mead 算法 [275] 最小二乘法的数值计算 [279]

第三章 微分方程

简谐振子受迫运动的简单数值计算 [283] 天体运动的简单数值计算 [284] 常微分方程（组）的数值解 [288] 中点法解常微分方程（组） [289] 四阶龙格库塔法 [292] 物理学常数定义 [296]

第一部分

数学

第一章

数学拾遗

二项式定理

预备知识 排列组合

结论

二项式展开公式为

$$(a+b)^n = \sum_{i=0}^n C_n^i a^i b^{n-i} \quad (n \text{ 为整数}) \quad (1)$$

其中

$$C_n^i = \frac{n(n-1)\dots(n-i+1)}{i!} = \frac{n!}{i!(n-i)!} \quad (2)$$

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
- \vdots
- $(a+b)^n$ 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 $a^i b^{n-i}$ ，那么共有 C_n^i 个这样的项。把它们相加得 $C_n^i a^i b^{n-i}$ 。所以

$$(a+b)^n = \sum_{i=0}^n C_n^i a^i b^{n-i} \quad (3)$$

证毕。

二项式定理（非整数幂）

预备知识 二项式定理^[3]

当 a, b 为实数, u 为非零实数时, 有

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

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^[3].

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^[50] 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 $|x| < 1$.

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

所以只要用数值验证

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

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^[268].

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

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

$$\sin^2 x + \cos^2 x = 1 \quad (1)$$

等式两边同除 $\cos^2 x$ 得

$$\tan^2 x + 1 = \sec^2 x \quad (2)$$

两角和公式

$$\sin(x \pm y) = \sin x \cos y \pm \cos x \sin y \quad (3)$$

$$\cos(x \pm y) = \cos x \cos y \mp \sin x \sin y \quad (4)$$

二倍角公式

令式 3 中 $y = x$ 取上号得

$$\sin 2x = 2 \sin x \cos x \quad (5)$$

$$\cos 2x = \cos^2 x - \sin^2 x \quad (6)$$

降幂公式

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

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

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

和差化积公式

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

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

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

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

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} 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.

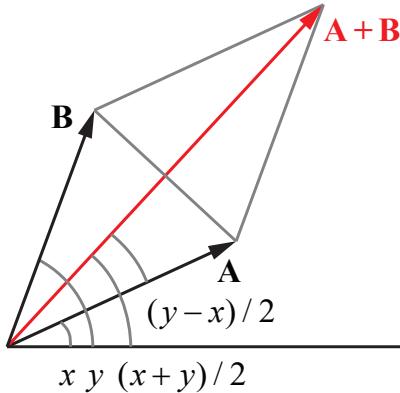


图 1: 和差化积公式推导

双曲函数

预备知识 指数函数

这里介绍三种双曲函数: 双曲正弦函数, 双曲余弦函数和双曲正切函数, 他们的定义分别为

$$\sinh x = \frac{e^x - e^{-x}}{2} \quad (1)$$

$$\cosh x = \frac{e^x + e^{-x}}{2} \quad (2)$$

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3)$$

其中 e 是一个高等数学中常见的常数，叫做自然对数底^[33]。这三个函数的图像如图 1。

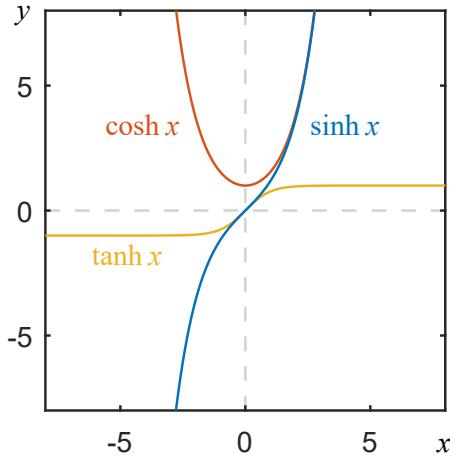


图 1: 三种双曲函数的图像

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

$$x = \sinh y = \frac{e^y - e^{-y}}{2} \quad (4)$$

整理成关于 e^y 的二次方程，得

$$(e^y)^2 - 2xe^y - 1 = 0 \quad (5)$$

解出 e^y 为

$$e^y = x \pm \sqrt{x^2 + 1} \quad (6)$$

由于 $e^y > 0$ ，上式取正号。两边取自然对数，得

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

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.

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邻边垂直. 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。
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 。

极坐标系

预备知识 平面直角坐标系，矢量

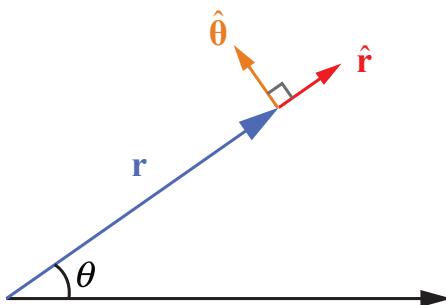


图 1: 极坐标系和两个单位矢量

在平面上取一个点作为极点，过极点的一条轴作为极轴。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 p ）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。若 θ 为负值，则表示从极轴开始延顺时针方向转动 $|\theta|$ 角。 θ 的值通常表示成弧度。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。

一般以坐标名上面加单位矢量符号表示该坐标对应的单位矢量。例如直角坐标系中， \hat{x} （有时也记为 \hat{i} ）是 x 坐标增加方向的单位矢量。所以在极坐标中，定义 \hat{r} 为 r 增加的方向的单位矢量， $\hat{\theta}$ 为 θ 坐标增加方向的单位矢量（即 \hat{r} 逆时针旋转 $\pi/2$ 的方向）。必须注意的是， \hat{r} 与 $\hat{\theta}$ 互相垂直，构成一对单位正交基底，平面上的任意向量都可以正交分解到这两个方向上。通常把 \hat{r} 的方向叫做径向，把 $\hat{\theta}$ 的方向叫做法向。

拓展阅读 极坐标中单位矢量的偏导，正交曲线坐标系

柱坐标系

预备知识 极坐标系^[9]

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 ($r \geq 0$)， θ 代表与 x 轴的夹角， z 与直角坐标系相同。

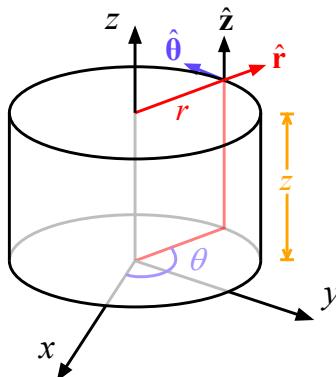


图 1：定义柱坐标系

单位矢量

柱坐标系中的单位矢量如图 1 中的 $\hat{r}, \hat{\theta}, \hat{z}$ 所示。与直角坐标系不同的是，三个单位矢量与具体的坐标相关，不是常矢量。这在矢量求导时非常关键。

与直角坐标系之间的变换

根据三角函数相关定义以及勾股定理，显然有

$$\begin{cases} x = r \cos \theta \\ y = r \sin \theta \\ z = z \end{cases} \quad \begin{cases} r = \sqrt{x^2 + y^2} \\ \theta = \arctan 2(x, y) \\ z = z \end{cases} \quad (1)$$

其中 $\arctan 2(x, y)$ 在第一四象限中就是我们熟知的 $\arctan(y/x)$ ，但 \arctan 无法表示其他象限。为了解决这个问题，一些编程语言中加入了 $\arctan 2$ 。本书将其定义为

$$\arctan 2(x, y) \equiv \begin{cases} \arctan(y/x) & (x > 0) \\ \pi/2 & (x = 0, y > 0) \\ -\pi/2 & (x = 0, y < 0) \\ \arctan(y/x) + \pi & (x < 0, y \geq 0) \\ \arctan(y/x) - \pi & (x < 0, y < 0) \end{cases} \quad (2)$$

球坐标系

预备知识 位矢^[153]，矢量的叉乘^[96]

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 ($r \geq 0$)，即位矢^[153] 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。

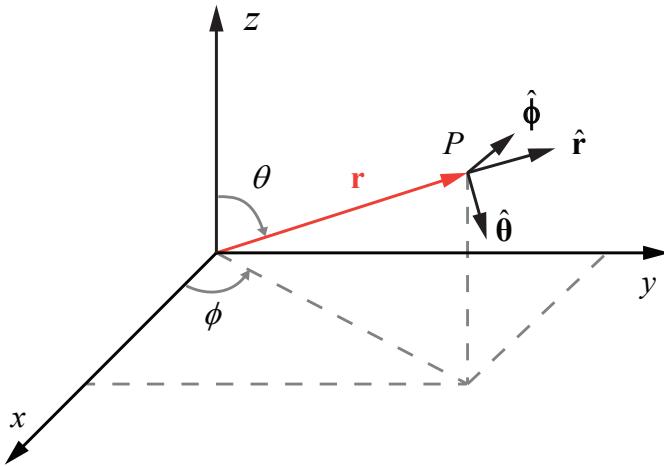


图 1: 球坐标系

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 $\hat{r}, \hat{\theta}, \hat{\phi}$ (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{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

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

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 $\hat{r}(\theta, \phi), \hat{\theta}(\theta, \phi), \hat{\phi}(\phi)$. 例如 P 的球坐标为 $(1, \pi/2, 0)$, 直角坐标为 $(1, 0, 0)$ 时, $\hat{r} = \hat{x}$, $\hat{\theta} = -\hat{z}$, $\hat{\phi} = \hat{y}$. 但是球坐标为 $(1, \pi/2, \pi/2)$, 直角坐标为 $(0, 1, 0)$ 时, $\hat{r} = \hat{y}$, $\hat{\theta} = -\hat{z}$, $\hat{\phi} = -\hat{x}$. 一般地, 对于球坐标为 (r, θ, ϕ) 的点 $\hat{r}, \hat{\theta}, \hat{\phi}$ 与 $\hat{x}, \hat{y}, \hat{z}$ 的关系见球坐标与直角坐标的转换^[13]. 另外注意改变 r 时 $\hat{r}, \hat{\theta}, \hat{\phi}$ 都保持不变, 且 $\hat{\phi}(\phi)$ 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 $\hat{r} \times \hat{\theta} = \hat{\phi}$ (类比直角坐标系的三个单位矢量必须满足 $\hat{x} \times \hat{y} = \hat{z}$, 见矢量的叉乘^[96]). 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^[153]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

$$\mathbf{v} = \omega r \sin \theta \hat{\phi} \quad (2)$$

球坐标与直角坐标的转换

预备知识 球坐标系的定义^[11]

结论

根据球坐标的定义，可得两种坐标之间的变换关系

$$\begin{cases} x = r \sin \theta \cos \phi \\ y = r \sin \theta \sin \phi \\ z = r \cos \theta \end{cases} \quad (1)$$

$$\begin{cases} r = \sqrt{x^2 + y^2 + z^2} \\ \theta = \arccos \frac{z}{\sqrt{x^2 + y^2 + z^2}} \\ \phi = \arctan \frac{y}{x} \end{cases} \quad (2)$$

以及两组基底之间的变换关系

$$\begin{cases} \hat{\mathbf{r}} = \sin \theta \cos \phi \hat{\mathbf{x}} + \sin \theta \sin \phi \hat{\mathbf{y}} + \cos \theta \hat{\mathbf{z}} \\ \hat{\theta} = \cos \theta \cos \phi \hat{\mathbf{x}} + \cos \theta \sin \phi \hat{\mathbf{y}} - \sin \theta \hat{\mathbf{z}} \\ \hat{\phi} = -\sin \phi \hat{\mathbf{x}} + \cos \phi \hat{\mathbf{y}} \end{cases} \quad (3)$$

$$\begin{cases} \hat{\mathbf{x}} = \sin \theta \cos \phi \hat{\mathbf{r}} + \cos \theta \cos \phi \hat{\theta} - \sin \phi \hat{\phi} \\ \hat{\mathbf{y}} = \sin \theta \sin \phi \hat{\mathbf{r}} + \cos \theta \sin \phi \hat{\theta} + \cos \phi \hat{\phi} \\ \hat{\mathbf{z}} = \cos \theta \hat{\mathbf{r}} - \sin \theta \hat{\theta} \end{cases} \quad (4)$$

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 $y = r \sin \theta \cos \phi$, $x = r \sin \theta \sin \phi$, 这就得到了式 1.

在直接坐标系中, 显然有 $r = \sqrt{x^2 + y^2 + z^2}$, 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 $\hat{\mathbf{r}}, \hat{\theta}, \hat{\phi}$ 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

$$(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta) \quad (5)$$

写成矢量的形式, 就是

$$\hat{\mathbf{r}} = \sin \theta \cos \phi \hat{\mathbf{x}} + \sin \theta \sin \phi \hat{\mathbf{y}} + \cos \theta \hat{\mathbf{z}} \quad (6)$$

至于式 3 的第二条式子, 在同一个球坐标 (r, θ, ϕ) 处, $\hat{\theta}$ 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

$$(\cos \theta \cos \phi, \cos \theta \sin \phi, -\sin \theta) \quad (7)$$

$$\hat{\theta} = \cos \theta \cos \phi \hat{\mathbf{x}} + \cos \theta \sin \phi \hat{\mathbf{y}} - \sin \theta \hat{\mathbf{z}} \quad (8)$$

同理, 在同一点 (r, θ, ϕ) 处, $\hat{\phi}$ 的球坐标为 $(1, \pi/2, \phi + \pi/2)$, 得到第三条式子.

下面推导变换式 4. 由于已经知道了变换式 3, 且直角坐标系和球坐标系中的基底都是单位正交基, 所以直接把变换式 3 中的系数写成 3×3 的矩阵形式, 再转置即可得到变换式 4 中的系数矩阵.

圆锥曲线的极坐标方程

预备知识 极坐标的定义 [9]

结论

圆锥曲线的极坐标方程为

$$\rho = \frac{p}{1 - e \cos \theta} \quad (1)$$

其中 p 是通径的一半, e 是离心率.

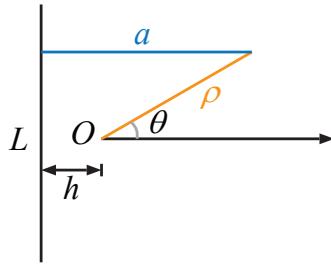


图 1: 由离心率定义圆锥曲线

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 ρ , 到 L 的（垂直）距离为 a , 令常数 $e > 0$, 则所有满足

$$\rho/a = e \quad (2)$$

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 $0 < e < 1$ 时, 曲线是椭圆, $e = 1$ 时是抛物线, $e > 1$ 时是双曲线.

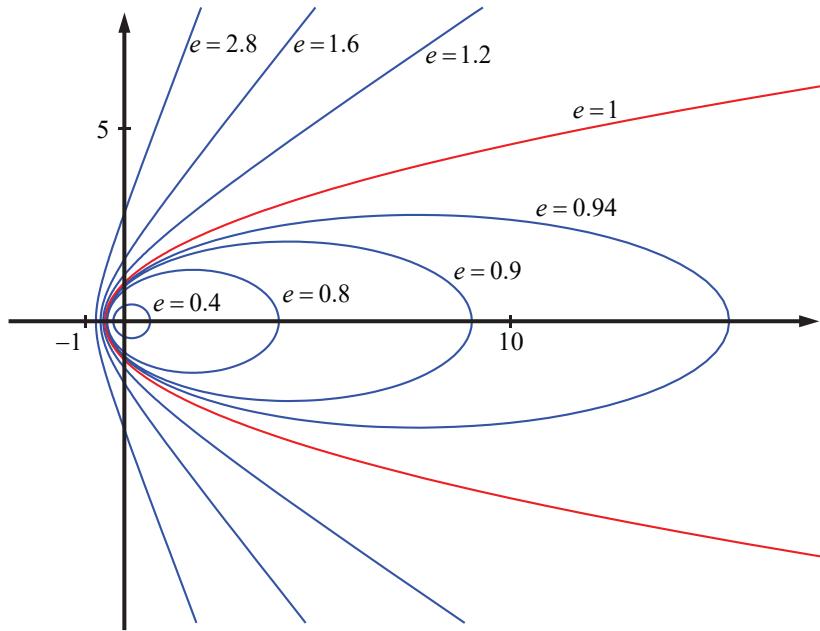
以 O 点为原点, 使极轴垂直于准线 (如上图). 则 $a = h + \rho \cos \theta$, 根据式 2 得

$$\frac{\rho}{h + \rho \cos \theta} = e \quad (3)$$

变形, 得

$$\rho = \frac{eh}{1 - e \cos \theta} \quad (4)$$

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, 令其长度为 $2p$, 那么有 $\rho(\pi/2) = p$. 代入式 4 得 $p = eh$. 所以式 4 又可

图 2: 不同离心率 e 的圆锥曲线

以写为

$$\rho = \frac{p}{1 - e \cos \theta} \quad (5)$$

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.

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 ρ 取负值, 会产生双曲线的左半支 (图中未画出). 左半支上的任意一点同样满足式 2.

椭圆的三种定义

预备知识 圆锥曲线的极坐标方程^[14]

第二种定义

我们前面已经见过椭圆在极坐标中的定义, 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

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，则式 2^[14] 变为

$$\frac{\sqrt{x^2 + y^2}}{x + h} = e \quad (1)$$

其中 h 为 y 轴到准线的距离。两边平方并整理得

$$(1 - e^2) \left(x - \frac{e^2 h}{1 - e^2} \right)^2 + y^2 = \frac{e^2 h^2}{1 - e^2} \quad (2)$$

由此可见，如果我们把椭圆左移 $e^2 h / (1 - e^2)$ ，椭圆将具有

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (3)$$

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

$$c = \frac{e^2 h}{1 - e^2} \quad (4)$$

式 2 和式 3 对比系数得

$$a = \frac{eh}{1 - e^2} \quad b = \frac{eh}{\sqrt{1 - e^2}} \quad (5)$$

不难证明

$$a^2 = b^2 + c^2 \quad (6)$$

以及

$$e = \frac{c}{a} \quad h = \frac{b^2}{c} \quad (7)$$

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

$$e = \frac{r_1}{d_1} = \frac{r_2}{d_2} = \frac{r_1 + r_2}{d_1 + d_2} \quad (8)$$

所以

$$r_1 + r_2 = e(d_1 + d_2) = 2e(c + h) = 2\frac{c}{a} \left(c + \frac{b^2}{c} \right) = 2a \quad (9)$$

证毕。

双曲线的三种定义

预备知识 圆锥曲线的极坐标方程^[14]

第二种定义

双曲线的极坐标方程为 ($e > 1$)

$$r = \frac{p}{1 - e \cos \theta} \quad (1)$$

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

$$\sqrt{x^2 + y^2} = p + ex \quad (2)$$

两边平方且化简得

$$\frac{(e^2 - 1)^2}{p^2} \left(x + \frac{ep}{e^2 - 1} \right)^2 - \frac{e^2 - 1}{p^2} y^2 = 1 \quad (3)$$

把双曲线沿 x 轴正方向移动 c , 可得以下形式

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \quad (4)$$

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

$$a = \frac{p}{e^2 - 1} \quad b = \frac{p}{\sqrt{e^2 - 1}} \quad c = \frac{ep}{e^2 - 1} \quad (5)$$

$$c^2 = a^2 + b^2 \quad (6)$$

用 a, b, c 表示 e, p 有

$$e = \frac{c}{a} \quad p = \frac{b^2}{a} \quad (7)$$

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

$$e = \frac{r_1}{d_1} = \frac{r_2}{d_2} = \frac{r_2 - r_1}{d_2 - d_1} \quad (8)$$

由于两准线之间的距离恒为 $2a^2/c$ ，上式变为

$$r_2 - r_1 = e(d_2 - d_1) = 2a \quad (9)$$

证毕.

渐近线

当 $x, y \rightarrow \infty$ 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

$$\theta_0 = \arctan(b/a) \quad (10)$$

两条渐近线到两个焦点的距离都为

$$c \sin \theta_0 = c \cdot b/c = b \quad (11)$$

复数

预备知识 几何矢量^[88]，三角恒等式^[4]

我们首先定义虚数单位 i 满足 $i^2 = -1$ ，则任何一个复数 z 可以表示为¹

$$z = x + iy \quad (1)$$

其中 x, y 为任意实数，分别被称为复数 z 的**实部** (**real part**) 和**虚部** (**imaginary part**)，记为 $\text{Re}[z]$ 和 $\text{Im}[z]$.

¹为了与变量 i 区分，本书中虚数单位使用正体的 i .

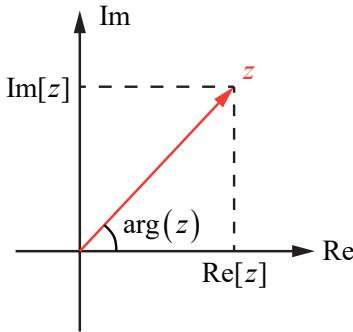


图 1: 复平面与复数

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. 复数的模长定义为对应矢量的模长, 即

$$|z| = \sqrt{\operatorname{Re}[z]^2 + \operatorname{Im}[z]^2} \quad (2)$$

另外我们把矢量与实轴的夹角称为幅角, 记为 $\arg(z)$. 我们可以通过实部和虚部计算幅角

$$\arg(z) = \arctan(\operatorname{Im}[z] / \operatorname{Re}[z]) \quad (3)$$

也可以通过模长和幅角来计算实部与虚部

$$\operatorname{Re}[z] = |z| \cos(\arg z) \quad \operatorname{Im}[z] = |z| \sin(\arg z) \quad (4)$$

在“指数函数(复数)[\[23\]](#)”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

$$z = A(\cos \theta + i \sin \theta) = A e^{i\theta} \quad (5)$$

其中 $A = |z|$, $\theta = \arg z$.

基本运算

共轭

一个复数的共轭等于一个实部相同, 虚部相反的复数

$$z^* = \operatorname{Re}[z] - i \operatorname{Im}[z] \quad (6)$$

所以共轭运算不改变复数的模长, 但将其幅角变为相反数.

加和减

复数的加减就是把两个复数的实部与虚部分别相加减（为了书写方便，这里把复数 z_i 的实部虚部记为 x_i, y_i ）

$$z_1 \pm z_2 = (x_1 \pm x_2) + i(y_1 \pm y_2) \quad (7)$$

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。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

$$\operatorname{Re}[z] = \frac{1}{2}(z + z^*) \quad \operatorname{Im}[z] = \frac{1}{2}(z - z^*) \quad (8)$$

乘法

两个复数相乘定义为（注意应用 $i^2 = -1$ ）

$$z_1 z_2 = (x_1 + iy_1)(x_2 + iy_2) = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1) \quad (9)$$

可以证明复数相乘后，乘积的模长等于两复数模长之积，乘积的幅角等于两复数的幅角之和，即

$$|z_1 z_2| = |z_1| |z_2| \quad (10)$$

$$\arg(z_1 z_2) = \arg(z_1) + \arg(z_2) \quad (11)$$

证明如下：令 $A_i = |z_i|$, $\theta_i = \arg z_i$, 则

$$\begin{aligned} z_1 z_2 &= (A_1 \cos \theta_1 + i A_1 \sin \theta_1)(A_2 \cos \theta_2 + i A_2 \sin \theta_2) \\ &= A_1 A_2 (\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2) + i A_1 A_2 (\cos \theta_1 \sin \theta_2 + \cos \theta_2 \sin \theta_1) \\ &= A_1 A_2 [\cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2)] \end{aligned} \quad (12)$$

其中最后一步用到了两角和公式式 3^[4]。容易看出，最后得到的是一个模长为 $A_1 A_2$, 幅角为 $\theta_1 + \theta_2$ 的复数。

不难证明复数的乘法满足交换律和结合律。

特殊地，一个复数乘以其复共轭可得其模长的平方。

$$z z^* = x^2 + y^2 = |z|^2 \quad (13)$$

除法

令 $z_1 = zz_2$, 则两个复数相除可以记为

$$z = \frac{z_1}{z_2} = \frac{x_1 + iy_1}{x_2 + iy_2} \quad (14)$$

但我们可以希望可以将结果的实部与虚部分开, 于是我们可以在等式两边同时乘以 z_2^* , 即 $z_1 z_2^* = zz_2 z_2^*$, 或

$$z = \frac{z_1 z_2^*}{z_2 z_2^*} = \frac{(x_1 + iy_1)(x_2 - iy_2)}{(x_2 + iy_2)(x_2 - iy_2)} = \frac{x_1 x_2 + y_1 y_2}{x_2^2 + y_2^2} + i \frac{x_2 y_1 - x_1 y_2}{x_2^2 + y_2^2} \quad (15)$$

这个步骤叫做分母有理化.

复变函数

预备知识 复数

复变函数是自变量和因变量都在复数域内取值的函数, 通常表示为

$$w = f(z) \quad (1)$$

若把自变量用 z 的实部 x 和虚部 y 表示, 因变量表示成实部函数 $u(x, y)$ 和虚部函数 $v(x, y)$ 两个函数相加, 则复变函数记为

$$w = u(x, y) + iv(x, y) \quad (2)$$

例如, 复数范围内的指数函数被定义为

$$w = e^z = e^x \cos y + ie^x \sin y \quad (3)$$

由于复变函数的图像比较复杂, 没有必要记忆图像, 只需要知道一些基本的性质即可.

与实变函数的“兼容性”

复变函数中很多函数与我们原来我们学过的函数同名, 只是自变量的范围从实数拓展到了复数. 例如以指数函数, 三角函数, 对数函数, 指数函数, 等.

这些新函数的定义必须要与原来的函数“兼容”，即当自变量被限制在实数范围内取值时，这些函数与原来的函数相同。

例如，当复数范围内的指数函数 $w = e^z = e^x \cos y + ie^x \sin y$ 的自变量只在实数范围取值（即 $y = 0$ ）时，该函数变为我们原来所熟悉的 e^x . 又如，复数范围内正弦函数被定义为

$$\sin z = \sin(x + iy) = \sin x \cosh y + i \cos x \sinh y \quad (4)$$

其中 \sinh 和 \cosh 是双曲正弦和双曲余弦函数。当 $y = 0$ 时，该函数变为 $\sin x$ 所以从这个意义上来说，与实变函数同名的复变函数只是把函数的定义域扩大了。

复变函数的导数

由于复变函数相当于两个实数自变量和两个实数因变量的函数，一般情况下求导变得非常复杂。但如果复变函数在某个域上解析，那么可以在该域上进行求导，得到唯一的导数。对于复数域初等基本函数，求导的结果也和实数域的求导一样。

指数函数（复数）

预备知识 指数函数，复数^[19]，三角恒等式^[4]，导数^[34]

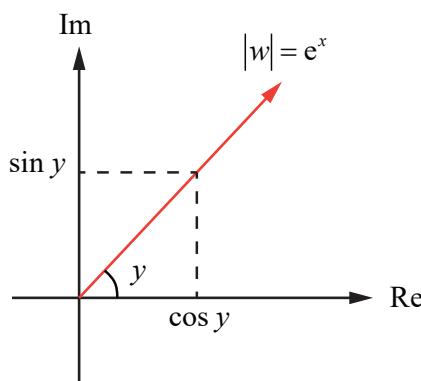


图 1: 复数域中的指数函数

复数域中的指数函数被定义为

$$w = e^z = e^{x+iy} = e^x(\cos y + i \sin y) \quad (1)$$

在复平面上表示这个函数，则指数的实部 x 控制因变量 w 的模长，虚部 y 控制 w 的幅角，如图 1

$$|w| = e^x \quad \arg(w) = y \quad (2)$$

当指数为纯虚数时，式 1 变为著名的欧拉公式

$$e^{ix} = \cos x + i \sin x \quad (3)$$

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^[24]，对于任何复数 z ，都有欧拉公式

$$e^{iz} = \cos z + i \sin z \quad (4)$$

将“三角函数（复数）^[23]”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^[4]），容易证明 e^z 同样满足

$$e^{z_1+z_2} = e^{z_1}e^{z_2} \quad (5)$$

虽然我们还没有系统地学习复变函数求导的概念，我们可以根据式 3 求出一个物理中常见的导数公式

$$\begin{aligned} \frac{d}{dx} e^{iax} &= -a \sin(ax) + ia \cos(ax) \\ &= ia[\cos(ax) + i \sin(ax)] \\ &= ie^{iax} \end{aligned} \quad (6)$$

进一步拓展，令复常数 $z = a + ib$ 得

$$\frac{d}{dx} e^{zx} = \frac{d}{dx}(e^{ax}e^{ibx}) = (a + ib)e^{(a+ib)x} = ze^{zx} \quad (7)$$

可见 e^z 的许多性质与实数域的 e^x 类似。

三角函数（复数）

预备知识 指数函数（复数）^[23]

定义

复数域的正弦函数为

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i} \quad (1)$$

复数域的余弦函数为

$$\cos z = \frac{e^{iz} + e^{-iz}}{2} \quad (2)$$

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

$$\sin z = \frac{e^{ix} - e^{-ix}}{2i} \quad (3)$$

根据欧拉公式^[23]，

$$e^{ix} = \cos x + i \sin x \quad (4)$$

$$e^{-ix} = \cos x - i \sin x \quad (5)$$

代入得

$$\sin z = \frac{(\cos x + i \sin x) - (\cos x - i \sin x)}{2i} = \sin x \quad (6)$$

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

$$\sin(z_1 + z_2) = \sin z_1 \cos z_2 + \cos z_1 \sin z_2 \quad (7)$$

$$\cos(z_1 + z_2) = \cos z_1 \cos z_2 - \sin z_1 \sin z_2 \quad (8)$$

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 iy , 则有

$$\sin z = \sin(x + iy) = \sin x \cos iy + \cos x \sin iy \quad (9)$$

$$\cos z = \cos(x + iy) = \cos x \cos iy - \sin x \sin iy \quad (10)$$

其中

$$\cos(iy) = \frac{e^{-y} + e^y}{2} = \cosh y \quad (11)$$

$$\sin(iy) = \frac{e^{-y} - e^y}{2i} = i \frac{e^y - e^{-y}}{2} = i \sinh y \quad (12)$$

代入得

$$\sin z = \sin(x + iy) = \sin x \cosh y + i \cos x \sinh y \quad (13)$$

$$\cos z = \cos(x + iy) = \cos x \cosh y - i \sin x \sinh y \quad (14)$$

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）。

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，常/偏微分方程和无穷级数。

极限

极限^[29]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 $x \rightarrow \infty$ 时的极限为零， $(1+x)/(2+x)$ 在 $x \rightarrow 0$ 时的极限为 $1/2$ 。

导数

理解极限了以后，导数^[34]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 $\Delta s/\Delta t$ 在 $\Delta t \rightarrow 0$ （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^[159]。学习了矢量求导^[126]以后，就不必再使用这种不成熟的“几何微元法”而直接按照矢量求导法^[160]则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。

积分

高中物理中，位移 s 等于速度 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

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

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^[65]，以上例子中变化的量（速度，密度，力）。在求定积分时，我们需要先求出一个原函数，而巧妙的地方在于，

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^[172]（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）。

中，结合牛顿第二定律^[162] 和胡克定律得到 $ma = F = -kx$ 其中位移 x 可看做关于时间的函数 $x(t)$ ，是未知函数（微分方程的解），加速度是时间的二阶导数 $a(t) = x''(t)$ 。所以微分方程为

本章介绍本书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

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 项是的前位小数近似（不考虑四舍五入），即

$$a_0 = 3, a_1 = 3.1, a_2 = 3.14, a_3 = 3.141, \dots \quad (1)$$

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

$$\lim_{n \rightarrow \infty} a_n = \pi \quad (2)$$

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。 $\lim_{n \rightarrow \infty}$ 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“因变量”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 $n \rightarrow \infty$ 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 $\lim_{n \rightarrow \infty}$ 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。这个定义无需硬记，如果理解了上面的描述，就觉得它理所当然了。

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 ” (any) 表示，把“存在”用“ \exists ” (exist) 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 $\lim_{n \rightarrow \infty} a_n = \pi$.

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 $\lim_{x \rightarrow +\infty} f(x) = A$, 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 $\lim_{x \rightarrow -\infty} f(x)$ (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 $\forall \varepsilon, \exists \delta$, 当 $|x - x_0| < \delta$ 时, 有 $|f(x) - A| < \varepsilon$ ”.

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

$$\lim_{x \rightarrow 1} 2x + 1 = 3 \quad \lim_{x \rightarrow 2} \frac{x+1}{x+2} = \frac{3}{4} \quad (3)$$

当无穷大与常数相加时, 可以忽略常数, 如

$$\lim_{x \rightarrow +\infty} \frac{x+1}{2x+2} = \lim_{x \rightarrow +\infty} \frac{x}{2x} = \frac{1}{2} \quad (4)$$

无穷小的阶

如果令 $x \rightarrow 0$, 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

$$\lim_{x \rightarrow 0} \frac{\alpha(x)}{x} = 0 \quad (5)$$

那 $\alpha(x)$ 就是 x 的高阶无穷小. 若

$$\lim_{x \rightarrow 0} \frac{\alpha(x)}{x^n} \neq 0 \quad (6)$$

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

$$\mathcal{O}(x^n)x^m = \mathcal{O}(x^{n+m}) \quad (m > -n) \quad (7)$$

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 $f(x) = g(x) + \mathcal{O}(h^{n+1})$ 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 $g(x)$ 精确到 $\mathcal{O}(h^n)$.

小角正弦极限

预备知识 极限^[29]

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）。

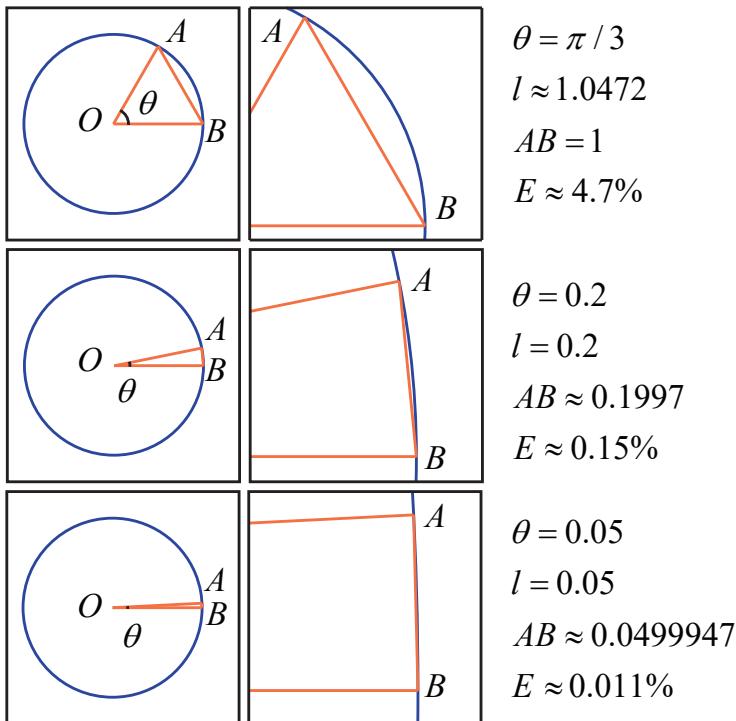


图 1：单位圆中，随着角度不断减小，弧长与线段长度的相对误差也不断减小

由弧长公式得

$$l = R\theta \quad (1)$$

线段 AB 的长度为

$$AB = 2R \sin \frac{\theta}{2} \quad (2)$$

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^[29] 的语言来说, 就是当 θ 趋近于 0 时, 它们的比值趋近于 1². 现在我们可以总结出 $\theta \rightarrow 0$ 时的两个结论

1. 线段长度 AB 趋近于弧长 l , 一般情况下可近似认为 $AB = l$ ³.
2. 代入上面的长度表达式 (式 1), 有

$$1 = \lim_{\theta \rightarrow 0} \frac{AB}{l} = \lim_{\theta \rightarrow 0} \frac{2R \sin(\theta/2)}{R\theta} = \lim_{\theta \rightarrow 0} \frac{\sin(\theta/2)}{\theta/2} \quad (3)$$

令 $x = \theta/2$, 有

$$\lim_{\theta \rightarrow 0} \frac{\sin x}{x} = 1 \quad (4)$$

这是一个非常重要的极限.

自然对数底

预备知识 极限^[29]

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⁴ e .

$$e \equiv \lim_{x \rightarrow 0} (1 + x)^{\frac{1}{x}} = 2.71828 \dots \quad (1)$$

这里仅用数值的方法验证该极限⁵ (表 1).

以 e 为底的对数函数 $\log_e x$ 叫做自然对数, 通常记为

$$\ln x \quad \text{或} \quad \log x \quad (2)$$

²注意这只是一个经验上的总结, 证明参考高等数学教材.

³严格来说, 这是一个一阶近似, 见泰勒级数.

⁴为了与其他变量区分, 本书使用正体字母表示自然对数底.

⁵注意若 x 从负值趋近 0 时该极限同样成立

表 1: 极限 e 数值验证

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

切线与割线

预备知识 极限^[29]

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

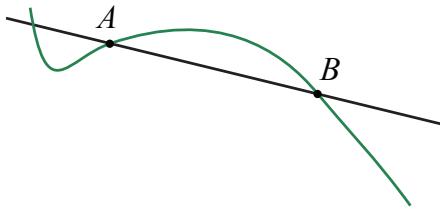


图 1: 割线

点靠近，割线的位置逐渐趋于不变，割线位置的极限^[29] 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。

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。

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。

拓展阅读 多重极限

导数

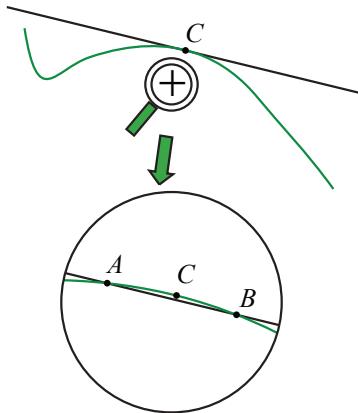
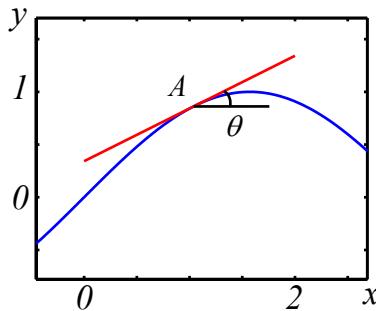


图 2: 割线的极限是切线

预备知识 极限^[29]

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^[34].

图 1: 点 A 的切线

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, $\theta = \tan \theta = 0$.

若函数曲线在 x 的某一开区间每一点都可导, 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)

$$f'(x), [f(x)]', \frac{dy}{dx}, \frac{df}{dx}, \frac{d}{dx} f(x) \quad (1)$$

另外，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。

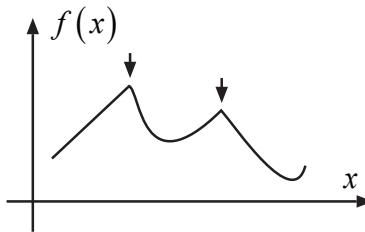


图 2: 棱角处不可导

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 dx （在微积分中，通常把非常小的一段 Δx 记为 dx ， dx 是一个不能分割的整体符号，而不是两个量相乘），竖直边的边长为 dy （当函数递增时， dy 取正值，反之取负值）。根据上面导数的定义， $dy/dx = \tan \theta$ 就是函数的导数。所以导数通常表示为 dy/dx ，导数的倒数则为 dx/dy 。

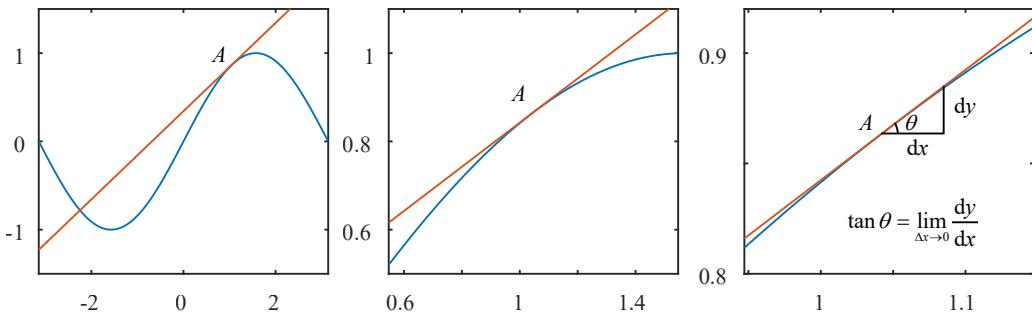


图 3: 将切点放大，会发现切线和曲线在切点附近“重合”

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 $dy = f'(x)dx$ 。这个关系就叫函数的微分（看到这里，你可能会觉得，原来微分这么简单！其实“微积分”基本就是在讲“微分”和“积分”而一元函数的微分及其基本原理的几何理解基

本上就可以认为是上面这张图所表示的).

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^[29]表达出来就是

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \quad (2)$$

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

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x) - f(x - \Delta x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x - \Delta x)}{2\Delta x} \quad (3)$$

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。

应用举例 一维运动的速度定义，一维运动的加速度定义

拓展阅读 基本初等函数的导数^[41]，求导法则，高阶导数

求导法则

预备知识 导数^[34]，基本初等函数的导数^[41]

结论

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进行求导

四则运算

$$[f(x) \pm g(x)]' = f'(x) \pm g'(x) \quad (1)$$

$$[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x) \quad (2)$$

$$\left[\frac{f(x)}{g(x)} \right]' = \frac{f'(x)g(x) - g'(x)f(x)}{g(x)^2} \quad (3)$$

复合函数

$$f[g(x)] = f'[g(x)]g'(x) \quad (4)$$

详细见“一元复合函数求导（链式法则）[\[48\]](#)”

线性

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）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，这里只需要证明求导运算是线性的，即求导是一种**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

$$\begin{aligned} \frac{d}{dx} \sum_i c_i f_i(x) &= \lim_{h \rightarrow 0} \left[\sum_i c_i f_i(x+h) - \sum_i c_i f_i(x) \right] / h \\ &= \sum_i c_i \lim_{h \rightarrow 0} [f_i(x+h) - f_i(x)] / h \\ &= \sum_i c_i f'_i(x) \end{aligned} \quad (5)$$

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

$$f'(x) = 5 \sin' x + 3(x^2)' = 5 \cos x + 3(2x) = 5 \cos x + 6x \quad (6)$$

两函数相乘的导数

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^[34] 得

$$[f(x)g(x)]' = \lim_{h \rightarrow 0} [f(x+h)g(x+h) - f(x)g(x)]/h \quad (7)$$

从几何上来看, 我们可以把 $f(x)g(x)$ 看做一个矩形的面积

反函数求导

预备知识 导数^[34]

结论

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

$$[f^{-1}(x)]' = \frac{1}{f'[f^{-1}(x)]} \quad (1)$$

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

$$g'(x) = \frac{1}{h[g(x)]} \quad (2)$$

反函数存在的条件

函数 $y = f(x)$, 在某个区间 (x_1, x_2) 内连续且单调, 且 x 与 y 一一对应. 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.

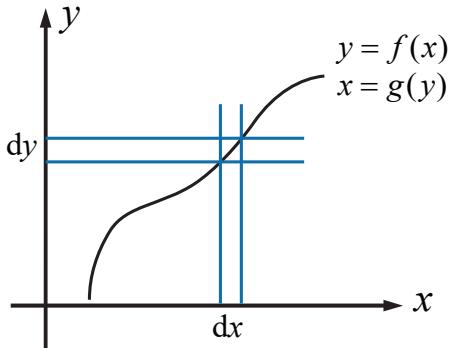


图 1: 在同一点处, $f' = \frac{dy}{dx}$, $g' = \frac{dx}{dy}$, 互为倒数

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

$$dy = f'(x_0) dx \quad (3)$$

同一点也满足 $g(y_0) = x_0$, 且

$$g'(y_0) dy = dx \quad (4)$$

对比式 3 和式 4, 得

$$g'(y_0) = \frac{dx}{dy} = \frac{1}{f'(x_0)} \quad (5)$$

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

$$g'(y_0) = \frac{dx}{dy} = \frac{1}{f'[g(y_0)]} \quad (6)$$

上式中, y_0 可以是 g 函数定义域的任意一点, 所以

$$g'(y) = \frac{1}{f'[g(y)]} \quad (7)$$

或者用习惯上的 x 作为自变量, 得

$$g'(x) = \frac{1}{f'[g(x)]} \quad (8)$$

证毕.

基本初等函数的导数

预备知识 导数^[34]

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

$$x^a \quad (a \in R) \quad (1)$$

- 指数函数

$$a^x \quad (a > 0, a \neq 1) \quad (2)$$

- 对数函数

$$\log_a x \quad (a > 0, a \neq 1) \quad (3)$$

当底为 $a = e$ 时，叫做自然对数函数，记为 $\ln x$.

- 三角函数

$$\sin x \quad \cos x \quad \tan x \quad (4)$$

- 反三角函数

$$\arcsin x \quad \arccos x \quad \arctan x \quad (5)$$

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

$$y = \sqrt{1 - x^2} \quad y = \sin^2 x \quad y = \sqrt{\cot \frac{x}{2}} \quad (6)$$

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

$$(x^a)' = ax^{a-1} \quad (a \in R) \quad (7)$$

三角函数

$$\sin' x = \cos x \quad \cos' x = -\sin x \quad \tan' x = 1/\cos^2 x = \sec^2 x \quad (8)$$

指数函数

$$(a^x)' = \ln(a)a^x \quad (9)$$

特殊地， $(e^x)' = e^x$

对数函数

$$(\log_a x)' = \frac{1}{\ln(a)x} \quad (10)$$

特殊地， $\ln' x = 1/x$.

幂函数证明

由导数的代数定义， $f'(x) = \lim_{h \rightarrow 0} [f(x+h) - f(x)]/h$ ，而

$$(x+h)^a - x^a = x^a[(1+h/x)^a - 1] \quad (11)$$

由于 $h \rightarrow 0$, $h/x \rightarrow 0$. 令 $\varepsilon = h/x$, 由非整数二项式定理^[3],

$$(1+\varepsilon)^a = 1 + a\varepsilon + \frac{a(a-1)}{2!}\varepsilon^2 + \frac{a(a-1)(a-2)}{3!}\varepsilon^3 \dots \quad (12)$$

所以

$$\begin{aligned} (x^a)' &= x^a \lim_{\varepsilon \rightarrow 0} \frac{(1+\varepsilon)^a - 1}{\varepsilon x} \\ &= x^{a-1} \lim_{\varepsilon \rightarrow 0} \left(a + \frac{a(a-1)}{2!}\varepsilon + \frac{a(a-1)(a-2)}{3!}\varepsilon^2 \dots \right) = ax^{a-1} \end{aligned} \quad (13)$$

正弦函数证明

使用三角函数和差化积公式化简极限

$$\sin' x = \lim_{h \rightarrow 0} \frac{\sin(x+h) - \sin x}{h} = \lim_{h \rightarrow 0} \frac{\sin(h/2)}{h/2} \cos\left(x + \frac{h}{2}\right) \quad (14)$$

由小角正弦值极限^[32] 中的结论, 其中

$$\lim_{h \rightarrow 0} \frac{\sin(h/2)}{h/2} = 1 \quad (15)$$

所以

$$\sin' x = \lim_{h \rightarrow 0} \cos\left(x + \frac{h}{2}\right) = \cos x \quad (16)$$

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^[37], 因为 $\tan x = \sin x / \cos x$, 所以

$$\tan' x = \frac{\sin' x \cos x - \cos' x \sin x}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x \quad (17)$$

对数函数证明

先证明 $\ln' x = 1/x$. $\ln(x+h) - \ln x = \ln(1+h/x)$, 所以

$$\ln' x = \lim_{h \rightarrow 0} \frac{\ln(x+h) - \ln x}{h} = \frac{1}{x} \lim_{h \rightarrow 0} \frac{\ln(1+h/x)}{h/x} \quad (18)$$

令 $\varepsilon = h/x$, 则

$$\ln' x = \frac{1}{x} \lim_{\varepsilon \rightarrow 0} \frac{\ln(1+\varepsilon)}{\varepsilon} = \frac{1}{x} \lim_{\varepsilon \rightarrow 0} \ln(1+\varepsilon)^{\frac{1}{\varepsilon}} \quad (19)$$

由自然对数底的定义, $\lim_{\varepsilon \rightarrow 0} (1 + \varepsilon)^{\frac{1}{\varepsilon}} = e$, 所以

$$\ln' x = \frac{\ln e}{x} = \frac{1}{x} \quad (20)$$

再证明 $\log'_a x = 1/(x \ln a)$. 由对数函数的性质 $\log_a b = \ln b / \ln a$

$$\log'_a x = \left(\frac{\ln x}{\ln a} \right)' = \frac{1}{\ln a} \ln' x = \frac{1}{x \ln a} \quad (21)$$

指数函数证明

先证明 $(e^x)' = e^x$. 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 $f(x) = \ln x$, $f'(x) = 1/x$, $f^{-1}(x) = e^x$, 代入反函数的求导法则^[39]

$$[f^{-1}(x)]' = \frac{1}{f'[f^{-1}(x)]} \quad (22)$$

得

$$(e^x)' = \frac{1}{1/e^x} = e^x \quad (23)$$

再证明 $(a^x)' = a^x \ln a$. $(a^x)' = [(e^{\ln a})^x]' = (e^{(\ln a)x})'$. 把 $e^{(\ln a)x}$ 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^[37], $(a^x)' = (\ln a)a^x$

导数与函数极值

预备知识 导数^[34]

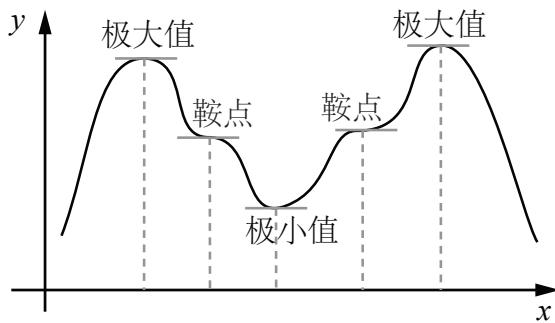


图 1: 导数为零的三种点

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 $f''(x_i) = 0$, 则 x_i 是鞍点, 如果 $f''(x_i) > 0$, x_i 是极小值点.

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.

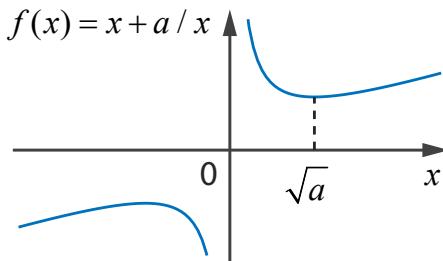


图 2: 例 2 函数图

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, 所以该驻点是一个鞍点.

用极值点确定函数图像

预备知识 导数与函数的极值^[44]

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

$$f(x) = a + \frac{b}{2}x^2 - \frac{c}{4}x^4 + \frac{d}{6}x^6 \quad (a, b, c, d > 0) \quad (1)$$

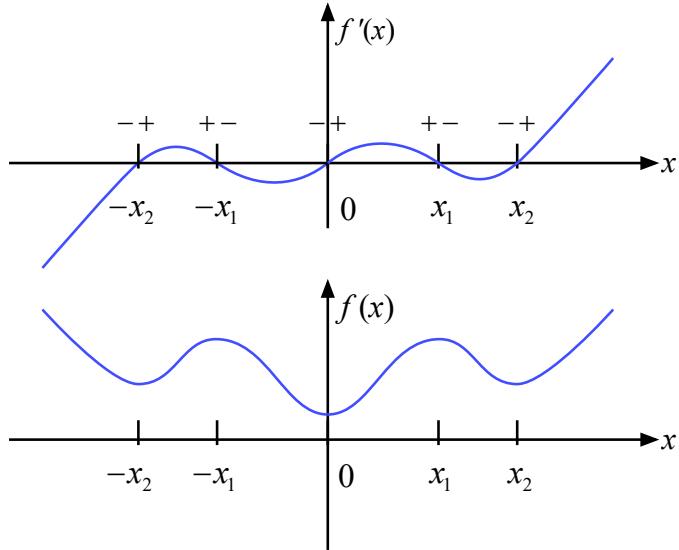


图 1: 导数 $f'(x)$ 与原函数 $f(x)$

先对函数求导，得

$$f'(x) = bx - cx^3 + dx^5 \quad (2)$$

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.

$$bx + cx^3 + dx^5 = x[b + c(x^2) + d(x^2)^2] \quad (3)$$

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_2^2 , $x_1 < x_2$. 此时五个根分别为 $0, \pm x_1, \pm x_2$.

$$f'(x) = d \cdot x(x^2 - x_2^2)(x^2 - x_1^2) = d \cdot x(x + x_1)(x - x_1)(x + x_1)(x - x_2) \quad (4)$$

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^[34]

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 dx , 函数值会相应增加无穷小量 dy . 根据导数的定义^[34] $f'(x) = dy/dx$, 我们将 dy 与 dx 的关系记为

$$dy = f'(x) dx \quad (1)$$

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 $dx \rightarrow 0$ 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

$$\Delta y \approx f'(x) \Delta x \quad (2)$$

注意在近似式中不能出现微分符号 d , 也不能使用等号.

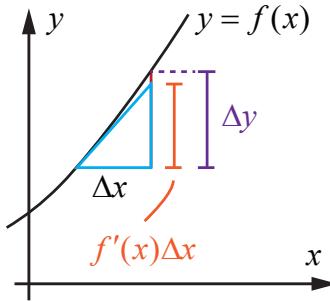


图 1: 微分近似用函数曲线的切线增量 $f'(x)\Delta x$ 来近似函数增量 Δy

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

$$\sigma_V \approx V'(a)\sigma_a = 3a^2\sigma_a \quad (3)$$

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 $dA = 2\pi r dr$. 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 dr 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 $dV = 4\pi r^2 dr$. 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 dr 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^[47]

若有一元函数 $f(x)$ 和 $g(x)$, 我们可以用 $f[g(x)]$ 表示其复合函数, 即把 g 的因变量作为 f 的自变量. 现在假设 f 和 g 在某定义域可导, 且我们已知导函数 f' 和 g' , 如何求复合函数 $f[g(x)]$ 的导数呢?

我们不妨先引入中间变量 u 作为 f 的自变量和 g 的因变量, 即 $y = f(u)$, $u = g(x)$. 现在可以写出微分关系^[47]

$$dy = f'(u) du \quad du = g'(x) dx \quad (1)$$

即 y 的微小变化由 u 的微小变化引起, 而 u 的微小变化又由 x 引起. 代入消去 du , 得

$$dy = f'(u)g'(x) dx = f'[g(x)]g'(x) dx \quad (2)$$

或记为

$$df = \frac{df}{dg} \frac{dg}{dx} dx \quad (3)$$

写成导数的形式为

$$\frac{dy}{dx} = f'[g(x)]g'(x) \quad \text{或} \quad \frac{df}{dx} = \frac{df}{dg} \frac{dg}{dx} \quad (4)$$

这就是一元复合函数求导的链式法则.

对于多重嵌套的情况如 $f[g(h(x))]$, 可以先对 $g[h(x)]$ 求导得 $g'[h(x)]h'(x)$ 再得到

$$\frac{d}{dx} f[g(h(x))] = f'[g(h(x))]g'[h(x)]h'(x) \quad (5)$$

例 1 对函数求导

$$\frac{1}{\sqrt{x^2 + a^2}} \quad (6)$$

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^[41],

$$f'(x) = -\frac{1}{2\sqrt{x^3}} \quad g'(x) = 2x \quad (7)$$

代入式 4, 得

$$\frac{d}{dx} \frac{1}{\sqrt{x^2 + a^2}} = f'[g(x)]g'(x) = -\frac{x}{\sqrt{(x^2 + a^2)^3}} \quad (8)$$

例 2 对函数求导

$$\sin^2 x \quad (9)$$

类似地, 令 $f(x) = x^2$ 再令 $g(x) = \sin(x)$, 上式等于 $f[g(x)]$. 由基本初等函数的导数,

$$f'(x) = 2x \quad g'(x) = \cos(x) \quad (10)$$

代入式4, 得

$$\frac{d}{dx} \frac{1}{\sqrt{x^2 + a^2}} = f'[g(x)]g'(x) = 2 \sin x \cos x \quad (11)$$

一种较灵活的情况是, 当三个变量只有一个自由度⁶时, 任何一个变量都可以看做任何另外一个变量的函数⁷, 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, 速度和加速度分别为 $x(t)$, $v(t) = dx/dt$, $a(t) = dv/dt$, 但若把速度 v 看做复合函数 $v[x(t)]$, 根据链式法则有

$$a = \frac{dv}{dt} = \frac{dv}{dx} \frac{dx}{dt} = \frac{dv}{dx} v \quad (12)$$

写成微分表达式, 有 $a dx = v dv$. 注意到 $d(v^2) = 2v dv$, 代入得

$$d(v^2) = 2a dx \quad (13)$$

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

$$v_2^2 - v_1^2 = 2a(x_2 - x_1) \quad (14)$$

其中 x_1, v_1 和 x_2, v_2 分别是 t_1, t_2 时刻的位置和速度.

泰勒展开

预备知识 高阶导数

若一个函数在某个区间内可以求任意阶的导数 (例如幂函数, 三角函数, 指数函数, 对数函数等), 那么这个函数可以用一个多项式近似, 且总项数 N 越多, 近似得越精确. 令多项式为

$$f(x) \approx c_0 + c_1(x - x_0) + c_2(x - x_0)^2 + \dots = \sum_{n=0}^N c_n(x - x_0)^n \quad (1)$$

⁶即任何一个变量值确定后, 另外两个变量也随之确定

⁷姑且假设不会出现一个自变量对应两个因变量的情况

其中 x_0 是该区间内的任意一点，多项式每一项的系数由函数在 x_0 处的第 n 阶导数求得

$$c_n = \frac{1}{n!} f^{(n)}(x_0) \quad (2)$$

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. 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

$$\sin x = x + \frac{1}{3!}x^3 - \frac{1}{5!}x^5 + \dots \quad (3)$$

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.

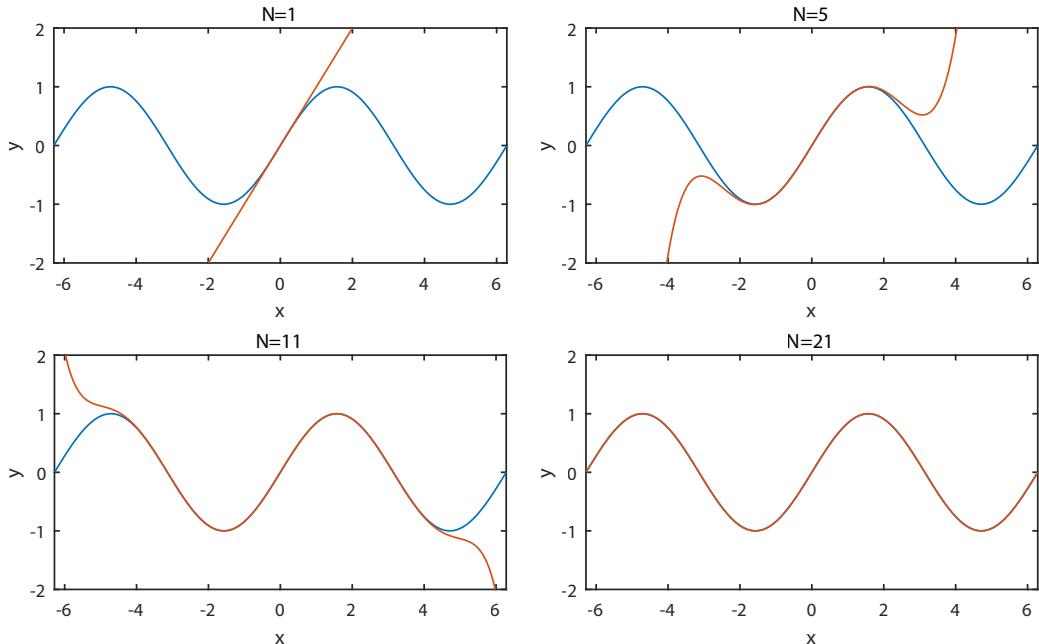


图 1: $\sin x$ 在原点处的泰勒展开的前 N 项求和. 容易看出，求和的项数越多，多项式（橙）与 $\sin x$ （蓝）吻合得越好.

系数公式的推导

我们假设当项数 $N \rightarrow \infty$ 时，总存在一个多项式趋于函数 $f(x)$ ⁸，即

$$f(x) = \sum_{n=0}^N c_n(x - x_0)^n \quad (4)$$

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

$$f'(x_0) = c_1 + \sum_{n=2}^N n c_n (x - x_0)^{n-1} \Big|_{x=x_0} = c_1 \quad (5)$$

如果对式 4 两边在 x_0 处求二阶导数，得

$$f''(x_0) = 2c_2 + \sum_{n=3}^N n(n-1)c_n(x - x_0)^{n-2} \Big|_{x=x_0} = 2c_2 \quad (6)$$

即 $c_2 = f''(x_0)/2!$ 。以此类推，如果对式 4 两边在 x_0 处求 m 阶导数得

$$f^{(m)}(x_0) = m!c_m + \sum_{n=m+1}^N \frac{n!}{(n-m)!} c_n (x - x_0)^{n-m} \Big|_{x=x_0} = m!c_m \quad (7)$$

所以系数公式为

$$c_m = \frac{1}{m!} f^{(m)}(x_0) \quad (8)$$

泰勒展开的存在说明了无穷可导函数的一个重要性质：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。

一些常见函数关于原点的泰勒展开

作为一个求导的练习，请验证以下泰勒展开式

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 \dots \quad (9)$$

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 \dots \quad (10)$$

⁸这叫做多项式的完备性，这里不予证明。

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 \dots \quad (11)$$

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 \dots \quad (12)$$

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 \dots \quad (13)$$

$$\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 \dots \quad (14)$$

泰勒展开与近似

事实上，泰勒展开可以看成是微分近似^[47] 的一种高阶拓展。微分近似中，在某点 x_0 附近有

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) \quad (15)$$

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 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^[29] $\mathcal{O}(x^{n+1})$ 。

导数与差分

预备知识 泰勒展开^[50]

我们在导数的定义^[34] 中已经知道⁹

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x + h/2) - f(x - h/2)}{h} \quad (1)$$

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

$$f(x \pm h/2) = f(x) \pm f'(x) \frac{h}{2} + \frac{1}{2} f''(x) \left(\frac{h}{2}\right)^2 + \mathcal{O}(h^3) \quad (2)$$

⁹以下假设 $f(x)$ 在某区间内处处可导。

代入式 1 得

$$\lim_{h \rightarrow 0} \frac{f'(x)h + \mathcal{O}(h^3)}{h} = f'(x) + \mathcal{O}(h^2) \quad (3)$$

所以用差分代替一阶导数可以精确到 h 的一阶无穷小 $\mathcal{O}(h)$.

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

$$\begin{aligned} f''(x) &= \lim_{l \rightarrow 0} \frac{f'(x + l/2) - f'(x - l/2)}{l} \\ &= \lim_{l \rightarrow 0} \lim_{h \rightarrow 0} \frac{1}{lh} [f(x + l/2 + h/2) - f(x + l/2 - h/2) \\ &\quad - f(x - l/2 + h/2) + f(x - l/2 - h/2)] \end{aligned} \quad (4)$$

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

$$f''(x) = \lim_{h \rightarrow 0} \frac{f(x + h) - 2f(x) + f(x - h)}{h^2} \quad (5)$$

要验证该式成立与否，不妨将 $f(x \pm h)$ 关于 h 做泰勒展开得

$$f(x \pm h) = f(x) \pm f'(x)h + \frac{1}{2}f''(x)h^2 \pm \frac{1}{6}f'''(x)h^3 + \mathcal{O}(h^4) \quad (6)$$

代入式 5 右边得

$$\lim_{h \rightarrow 0} \frac{f(x + h) - 2f(x) + f(x - h)}{h^2} = \lim_{h \rightarrow 0} \frac{f''(x)h^2 + \mathcal{O}(h^4)}{h^2} = f''(x) + \mathcal{O}(h^2) \quad (7)$$

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 $f''(x)$ 同样是精确到一阶无穷小 $\mathcal{O}(h)$.

不定积分

预备知识 导数^[34]，基本初等函数的导数^[41]

对一个函数 $f(x)$ 的**不定积分**（简称积分）可得到另一个函数 $F(x)$ ，叫做 $f(x)$ 的**原函数**。不定积分用符号表示为

$$F(x) = \int f(x) dx \quad (1)$$

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.

$$F'(x) = f(x) \quad (2)$$

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.

证明: 由于常数导数^[34] 为 0, 给原函数加上常数后式 2 仍然成立

$$\frac{d}{dx}[F(x) + C] = f(x) \quad (3)$$

我们可以从几何上来理解该式: 将函数曲线整体在 y 方向平移并不影响某个 x 坐标处函数曲线的斜率.

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

$$\int [a_1 f_1(x) + a_2 f_2(x) \dots] dx = a_1 \int f_1(x) dx + a_2 \int f_2(x) dx \dots \quad (4)$$

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^[56], 包括第一类换元法和第二类换元法.
3. 分部积分法^[58]
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^[59] 中的常用积分部分。

换元积分法

预备知识 不定积分^[54]

第一类换元积分法

由复合函数的求导法则，若令 $F'(x) = f(x)$ ，则

$$\frac{d}{dx}F[u(x)] = f[u(x)]u'(x) \quad (1)$$

由于求导的逆运算是积分，有

$$\int f[u(x)]u'(x) dx = F[u(x)] + C \quad (2)$$

所以如果某个积分可以看成 $\int f[u(x)]u'(x) dx$ 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结构是 $\int f[u(x)]u'(x) dx$ ，只有多练习才能熟能生巧。

例 1

计算

$$\int a \sin(ax + b) dx \quad (3)$$

令 $f(x) = \sin(x)$, $u(x) = ax + b$, 则上式刚好是 $\int f[u(x)]u'(x) dx$ 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

$$F[u(x)] + C = -\cos(ax + b) + C \quad (4)$$

总结到更一般的情况，根据换元积分法，若已知 $\int f(x) dx = F(x) + C$ ，则对于任意常数 a 和 b ，必有 $\int a f(ax + b) dx = F(ax + b)$ 。根据积分的基本性质，两边同除 a ，得

$$\int f(ax + b) dx = \frac{1}{a}F(ax + b) \quad (5)$$

该式使用频率很高，需要熟练掌握。类似例子还有

$$\int \frac{1}{ax+b} dx = \frac{1}{a} \ln(ax+b) + C \quad \int e^{ax+b} dx = \frac{1}{a} e^{ax+b} + C \quad (6)$$

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^[47]）

$$\begin{aligned} \int f[u(x)]u'(x) dx &= \int f[u(x)] d[u(x)] = \int f(u) du = F(u) + C \\ &= F[u(x)] + C \end{aligned} \quad (7)$$

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。

$$\int f(x) dx = \int f[x(t)] d[x(t)] = \int f[x(t)]x'(t) dt \quad (8)$$

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

$$\int \frac{dx}{\sqrt{1-x^2}} \quad (9)$$

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

$$\int \frac{dx}{\sqrt{1-x^2}} = \int \frac{d \sin t}{\sqrt{1-\sin^2 t}} = \int \frac{\cos t dt}{\cos t} = \int 1 dt = t + C = \arcsin x + C \quad (10)$$

验证：根据反函数求导法则^[39]

$$\arcsin' x = \frac{1}{\cos(\arcsin x)} = \frac{1}{\sqrt{1-x^2}} \quad (11)$$

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。

分部积分法

预备知识 不定积分^[54], 牛顿—莱布尼兹公式^[69]

结论

$$\int F(x)g(x) \, dx = F(x)G(x) - \int f(x)G(x) \, dx \quad (1)$$

$$\int_a^b F(x)g(x) \, dx = [F(x)G(x)]_a^b - \int_a^b f(x)G(x) \, dx \quad (2)$$

$$\int f(x)g(x) \, dx = f(x)g^{[1]}(x) - f^{(1)}(x)g^{[2]}(x) + \int f^{(2)}(x)g^{[2]}(x) \, dx \quad (3)$$

推导

令 $f(x) = F'(x)$, $g(x) = G'(x)$, 根据乘法的求导公式

$$[F(x)G(x)]' = f(x)G(x) + F(x)g(x) \quad (4)$$

即

$$F(x)g(x) = [F(x)G(x)]' - f(x)G(x) \quad (5)$$

两边不定积分（积分常数可任取）得

$$\int F(x)g(x) \, dx = F(x)G(x) - \int f(x)G(x) \, dx \quad (6)$$

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

$$\int_a^b F(x)g(x) \, dx = [F(x)G(x)]_a^b - \int_a^b f(x)G(x) \, dx \quad (7)$$

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

$$\int xe^{-x} dx = x(-e^{-x}) - \int 1 \times (-e^{-x}) dx = -xe^{-x} - e^{-x} + C \quad (8)$$

如果直接计算定积分，把“求导项”和“积分项”直接代入式 7 得

$$\int_0^{+\infty} xe^{-x} dx = x(-e^{-x}) \Big|_0^{+\infty} - \int_0^{+\infty} 1 \times (-e^{-x}) dx = 0 - e^{-x} \Big|_0^{+\infty} = 1 \quad (9)$$

连续分布积分

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¹¹ $g^{[n]}(x)$. 则分部积分式 6 可记为

$$\int f(x)g(x) dx = f(x)g^{[1]}(x) - \int f^{(1)}(x)g^{[1]}(x) dx \quad (10)$$

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

$$\int f(x)g(x) dx = f(x)g^{[1]}(x) - f^{(1)}(x)g^{[2]}(x) + \int f^{(2)}(x)g^{[2]}(x) dx \quad (11)$$

再把 $f^{(2)}$ 作为“求导项”， $g^{[2]}$ 作为“积分项”，分布积分得

$$\int f(x)g(x) dx = f(x)g^{[1]}(x) - f^{(1)}(x)g^{[2]}(x) + f^{(2)}(x)g^{[3]}(x) - \int f^{(3)}(x)g^{[3]}(x) dx \quad (12)$$

可以发现若要使用 N 次分部积分，第 $i \leq N$ 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^[54]

¹¹这是我自己的符号

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^[41] 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^[56], 表中任何积分都可以拓展为

$$\int f(ax + b) \, dx = \frac{1}{a} F(ax + b) \quad (1)$$

基本积分表

$$\int x^a \, dx = \frac{1}{a+1} x^{a+1} + C \quad (a \in R, a \neq -1) \quad (2)$$

$$\int \frac{1}{x} \, dx = \ln|x| + C \quad (\text{例 6}) \quad (3)$$

$$\int \cos x \, dx = \sin x + C \quad (4)$$

$$\int \sin x \, dx = -\cos x + C \quad (5)$$

$$\int \tan x \, dx = -\ln|\cos x| + C \quad (\text{例 2}) \quad (6)$$

$$\int \cot x \, dx = \ln|\sin x| + C \quad (\text{例 3}) \quad (7)$$

$$\int \frac{1}{\cos^2 x} \, dx = \tan x + C \quad (8)$$

$$\int \frac{1}{1+x^2} \, dx = \arctan x + C \quad (9)$$

$$\int e^x \, dx = e^x + C \quad (10)$$

$$\int x e^x \, dx = e^x(x-1) + C \quad (\text{例 7}) \quad (11)$$

$$\int a^x \, dx = \frac{1}{\ln a} a^x + C \quad (\text{例 1}) \quad (12)$$

常用积分表

$$\int \sin^2 x \, dx = \frac{1}{2}(x - \sin x \cos x) + C \quad (\text{例 4}) \quad (13)$$

$$\int \cos^2 x \, dx = \frac{1}{2}(x + \sin x \cos x) + C \quad (\text{例 5}) \quad (14)$$

$$\int \sec x \, dx = \ln |\tan x + \sec x| + C \quad (\text{例 11}) \quad (15)$$

$$\int \ln x \, dx = x \ln x - x + C \quad (\text{例 8}) \quad (16)$$

$$\int \sqrt{a^2 - x^2} \, dx = \frac{1}{2} \left(x \sqrt{a^2 - x^2} + a^2 \arcsin \frac{x}{a} \right) + C \quad (\text{例 9}) \quad (17)$$

$$\int \frac{1}{\sqrt{1-x^2}} \, dx = \arcsin(x) + C \quad (\text{例 10}) \quad (18)$$

$$\int \frac{1}{\sqrt{1+x^2}} \, dx = \ln \left(1 + \sqrt{1+x^2} \right) = \sinh^{-1} x + C \quad (\text{例 12}) \quad (19)$$

例 1

$$\int a^x \, dx \quad (20)$$

我们已经知道如何算 e^x 的积分，而 $a = e^{\ln a}$ ，再根据式 1 就有

$$\int e^{\ln(a)x} \, dx = \frac{1}{\ln a} e^{\ln(a)x} + C = \frac{1}{\ln a} a^x + C \quad (21)$$

例 2

$$\int \tan x \, dx \quad (22)$$

这个积分用第一类换元积分法 (式 2^[56])

$$\int f[u(x)]u'(x) \, dx = F[u(x)] + C \quad (23)$$

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

$$\int \tan x \, dx = \int f[u(x)]u'(x) \, dx = F[u(x)] + C = -\ln|\cos x| + C \quad (24)$$

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

$$\int \cot x \, dx = F[u(x)] + C = \ln|\sin x| + C \quad (25)$$

例 4

$$\int \sin^2 x \, dx \quad (26)$$

用降幂公式（式 7^[4]）和不定积分的线性（式 4^[54]）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

$$\begin{aligned} \int \sin^2 x \, dx &= \int \frac{1}{2} \, dx - \frac{1}{2} \int \cos 2x \, dx \\ &= \frac{x}{2} - \frac{1}{4} \sin(2x) = \frac{1}{2}(x - \sin x \cos x) + C \end{aligned} \quad (27)$$

例 5

$$\int \cos^2 x \, dx \quad (28)$$

与例 4 类似，用三角恒等式 $\cos^2(x) = [1 + \cos(2x)]/2$ 得

$$\begin{aligned} \int \cos^2 x \, dx &= \int \frac{1}{2} \, dx + \frac{1}{2} \int \cos(2x) \, dx \\ &= \frac{x}{2} + \frac{1}{4} \sin(2x) = \frac{1}{2}(x + \sin x \cos x) + C \end{aligned} \quad (29)$$

例 6

$$\int \frac{1}{x} \, dx \quad (30)$$

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

$$\int \frac{1}{x} \, dx = \ln|x| + C \quad (31)$$

事实上，由于 $1/x$ 在 $x = 0$ 没有定义，更广义的原函数可以取

$$\int \frac{1}{x} \, dx = \begin{cases} \ln x + C_1 & (x > 0) \\ \ln(-x) + C_2 & (x < 0) \end{cases} \quad (32)$$

其中 C_1 和 C_2 是两个不相同的待定常数.

例 7

$$\int x e^x dx \quad (33)$$

使用用分部积分式 1^[58]

$$\int F(x)g(x) dx = F(x)G(x) - \int f(x)G(x) dx \quad (34)$$

令 $F(x) = x$, 求导得 $f(x) = 1$, 令 $g(x) = e^x$, 由式 10, $G(x) = e^x$. 代入分部积分得

$$\int x e^x dx = x e^x - \int 1 \cdot e^x dx = e^x(x - 1) + C \quad (35)$$

例 8

$$\int \ln x dx \quad (36)$$

方法一： 使用第二类换元法式 8^[56]

$$\int f(x) dx = \int f[x(t)] d[x(t)] = \int f[x(t)]x'(t) dt \quad (37)$$

令¹² $x = e^t$, 求导得 $x'(t) = e^t$, 换元得

$$\int \ln x dx = \int \ln(e^t) e^t dt = \int t e^t dt \quad (38)$$

由例 7 中的分部积分得

$$\int \ln x dx = e^t(t - 1) + C = e^{\ln x}(\ln x - 1) + C = x(\ln x - 1) + C \quad (39)$$

方法二： 直接使用分部积分法式 1^[58], 对常数 1 积分, 对 $\ln x$ 求导, 得

$$\int \ln x dx = x \ln x - \int x \cdot \frac{1}{x} dx = x \ln x - x + C \quad (40)$$

例 9

$$\int \sqrt{a^2 - x^2} dx \quad (41)$$

¹²注意被积函数只在 $x > 0$ 区间有定义, 否则使用 $x = e^t$ 将会自动忽略 $x \leq 0$ 的情况.

使用第二类换元法式 8^[56], 令 $x = a \sin t$ 得

$$\int a \cos t \, d(a \sin t) = a^2 \int \cos^2 t \, dt \quad (42)$$

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

$$\int \sqrt{a^2 - x^2} \, dx = \frac{1}{2} \left(x \sqrt{a^2 - x^2} + a^2 \arcsin \frac{x}{a} \right) + C \quad (43)$$

例 10

$$\int \frac{1}{\sqrt{1 - x^2}} \, dx \quad (44)$$

使用第二类换元法式 8^[56], 令 $x = \sin t$ 得

$$\int \frac{1}{\sqrt{1 - \sin^2 t}} \, d(\sin t) = \int dt = t + C = \arcsin x + C \quad (45)$$

例 11

$$\int \sec x \, dx \quad (46)$$

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^[56]), 令 $u(x) = \sec x + \tan x$, 再使用式 3 即可

$$\begin{aligned} \int \sec x \, dx &= \int \frac{\sec^2 x + \sec x \tan x}{\sec x + \tan x} \, dx = \int \frac{u'(x)}{u} \, dx = \int \frac{1}{u} \, du \\ &= \ln |u| + C = \ln |\sec x + \tan x| + C \end{aligned} \quad (47)$$

例 12

$$\int \frac{1}{\sqrt{1 + x^2}} \, dx \quad (48)$$

使用第二类换元法式 8^[56], 令 $x = \tan t$, 再利用三角恒等式式 2^[4] 和式 3 得

$$\int \frac{1}{\sqrt{1 + \tan^2 t}} \, d(\tan t) = \int \frac{1}{\sec t} \sec^2 t \, dt = \ln |\tan t + \sec t| + C \quad (49)$$

由同一三角恒等式, $\sec t = \sqrt{1 + \tan^2 t} = \sqrt{1 + x^2}$, 所以

$$\int \frac{1}{\sqrt{1 + x^2}} \, dx = \ln \left(x + \sqrt{1 + x^2} \right) + C \quad (50)$$

注意上式中 \ln 后面的绝对值符号消失是因为 $x + \sqrt{1 + x^2} \geq 0$ 恒成立. 另外由“双曲函数^[6]”中例 1 可知上式又等于 $\sinh^{-1} x + C$.

定积分

预备知识 导数^[34], 极限^[29]

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.

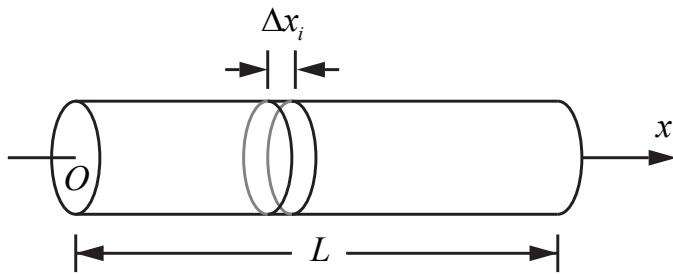


图 1: 密度不均匀的绳子

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 $\rho(\xi_i)$ ($x_i \leq \xi_i \leq x_{i+1}$) 来代替第 i 节的密度, 当每一节足够小时, 可以认为 ξ_i 在 $x_i \leq \xi_i \leq x_{i+1}$ 约束下的取值并不会影响结果. 第 i 小节的质量为

$$\Delta m_i = \rho(\xi_i) \Delta x_i S \quad (1)$$

所以总的质量用求和符号来表示, 就是

$$m = \sum_{i=1}^n \Delta m_i \approx \sum_{i=1}^n \rho(\xi_i) \Delta x_i S = S \sum_{i=1}^n \rho(\xi_i) \Delta x_i \quad (2)$$

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

$$m = \lim_{n \rightarrow \infty} \sum_{i=1}^n \Delta m_i = \lim_{n \rightarrow \infty} \sum_{i=1}^n \rho(\xi_i) \Delta x_i S = S \lim_{n \rightarrow \infty} \sum_{i=1}^n \rho(\xi_i) \Delta x_i \quad (3)$$

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 变为表示微小量的 d ，上式就写为

$$m = \int dm = \int S \rho(x) dx = S \int \rho(x) dx \quad (4)$$

下面先看看 $\int \rho(x) dx$ ，即 $\lim_{n \rightarrow \infty} \sum_{i=1}^n \rho(\xi_i) \Delta x_i$ 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， $\sum_{i=1}^n \rho(\xi_i) \Delta x_i$ 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 $\int \rho(x) dx$ 可以用来表示右图阴影部分的面积。

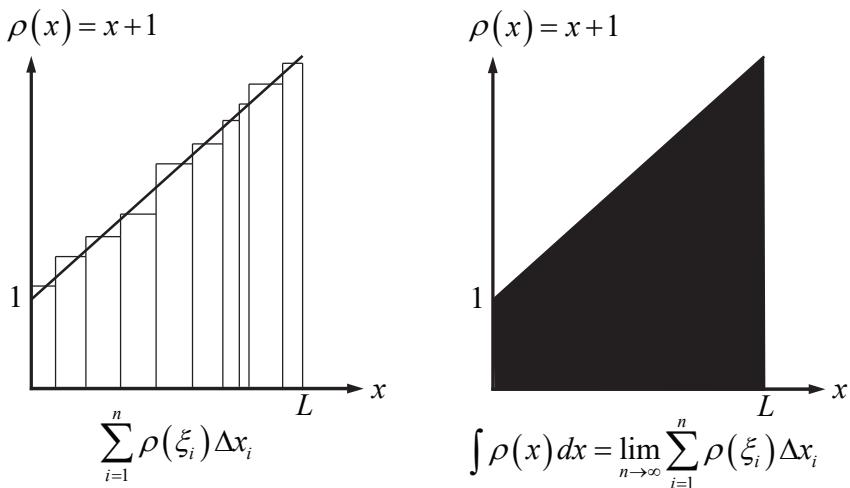


图 2：定积分可以理解为曲线下面的面积，并看做由无限多个无穷窄的矩形组成。

但 $\int \rho(x) dx$ 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

$$\int_0^L \rho(x) dx \quad (5)$$

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^[69] 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

$$y = f(x) = \sqrt{R^2 - x^2} \quad (x \in [-R, R]) \quad (6)$$

将该式定积分再乘以 2 即可得到圆的面积

$$S = 2 \int_{-R}^R f(x) dx = 2 \int_{-R}^R \sqrt{R^2 - x^2} dx \quad (7)$$

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^[47]，每个微小圆环的面积为 $2\pi r dr$ ，所以圆的面积可以用定积分表示为

$$S = \int_0^R 2\pi r dr \quad (8)$$

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^[69]”的例 2。

例 3 球体的表面积

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

$$dS = 2\pi R \sin \theta \cdot R d\theta \quad (9)$$

所以球的表面积可以用定积分记为

$$S = \int_0^\pi 2\pi R \sin \theta \cdot R d\theta = 2\pi R^2 \int_0^\pi \sin \theta d\theta \quad (10)$$

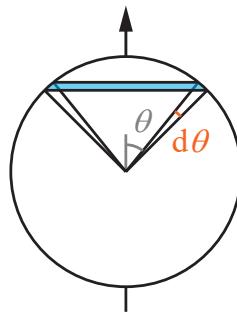


图 3: 将球的表面划分成许多细圆环，每个对应的极角为 $d\theta$

虽然我们还不会计算这个定积分（见“牛顿—莱布尼兹公式^[69]”），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 $dz = -R \sin \theta d\theta$ ，将该式消去式 9 中的 $d\theta$ 得

$$dS = -2\pi R dz \quad (11)$$

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 $d\theta$ 对应正的面积，而正的 $d\theta$ 却对应负的 dz 。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

$$S = \int_{-R}^R 2\pi R dz \quad (12)$$

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^[47]），即 $dV = 4\pi r^2 dr$ 。所以球的体积可用定积分表示为

$$V = \int_0^R 4\pi r^2 dr \quad (13)$$

同样由“牛顿—莱布尼兹公式^[69]”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^[54], 定积分^[65]

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下方的面积, 二者乍看起来没什么联系, 但牛顿—莱布尼兹公式却揭示了两者之间的重要关系.

若 $F(x)$ 是 $f(x)$ 的一个原函数^[54], 则

$$\int_a^b f(x) dx = F(b) - F(a) \quad (1)$$

推导

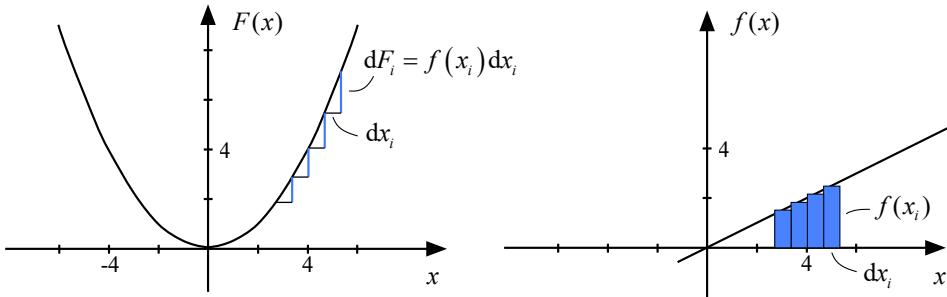


图 1: 右图中 $f(x)$ 的原函数为左图中的 $F(x)$, 当步长趋近 0 时, 右图中的长方形面积趋近于左图中小竖线的长度.

如图 1, 根据定积分^[65] 的定义, 有¹³

$$\int_a^b f(x) dx = \lim_{\Delta x_i \rightarrow 0} \sum_i f(x_i) \Delta x_i \quad (2)$$

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

$$f(x_i) \Delta x_i = F'(x_i) \Delta x_i \approx \Delta F_i = F(x_{i+1}) - F(x_i) \quad (3)$$

¹³这里假设极限存在.

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 $\Delta x_i \rightarrow 0$ 时, 上式取等号. 代回式 1, 有

$$\int_a^b f(x) dx = \lim_{\Delta x_i \rightarrow 0} \sum_i [F(x_{i+1}) - F(x_i)] = F(b) - F(a) \quad (4)$$

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 $\Delta x_i \rightarrow 0$, 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 $F(x)|_a^b$.

例 1 计算定积分

$$\int_{-l}^l \sin^2\left(\frac{n\pi}{l}x\right) dx \quad (5)$$

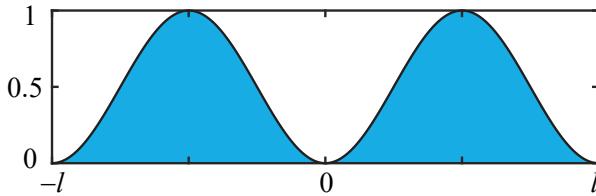


图 2: $y = \sin^2(\pi x/l)$ 的定积分

先计算对应的不定积分. 由积分表^[59] 中的式 13 结合式 1 得不定积分为

$$\int \sin^2\left(\frac{n\pi}{l}x\right) dx = \frac{l}{2n\pi} \left[\frac{n\pi}{l}x - \sin\left(\frac{n\pi}{l}x\right) \cos\left(\frac{n\pi}{l}x\right) \right] \quad (6)$$

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^[65] 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 得

$$\int \sqrt{R^2 - x^2} dx = \frac{1}{2} \left(x \sqrt{R^2 - x^2} + R^2 \arcsin \frac{x}{R} \right) + C \quad (7)$$

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 和牛顿—莱布尼兹公式得

$$\int_0^R 2\pi r \, dr = \pi r^2 \Big|_0^R = \pi R^2 \quad (8)$$

常微分方程

预备知识 简谐振子^[172]

作为一个引入的例子，我们首先看“简谐振子^[172]”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¹⁴），即

$$f(y^{(N)}, y^{(N-1)}, \dots, y, x) = 0 \quad (1)$$

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 $y^{(N-1)}, \dots, y, x$ 可以只出现部分或不出现. 所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 $f(x, C_1, \dots, C_N)$ ，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^[160] 的微分方程为 $d^2y/dt^2 = g$ （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 $y = C_1 + C_2 t + gt^2/2$. 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^[173]”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

二阶线性微分方程

¹⁴这里的“常”强调未知函数只有一个因变量，用于区别多元微积分中的“偏微分方程”.

一阶线性微分方程

预备知识 常微分方程^[71]

具有以下形式的微分方程叫做一阶线性微分方程

$$\frac{dy}{dx} + p(x)y = f(x) \quad (1)$$

一般地，未知函数及其各阶导数都各占一项时，方程就是线性的。另外，如果 $f(x)$ 项不出现，方程就是齐次的，否则就是非齐次的。我们先来看以上方程对应的齐次方程

$$\frac{dy}{dx} + p(x)y = 0 \quad (2)$$

这是一个可分离变量的方程，分离变量得

$$\frac{dy}{y} = -p(x) dx \quad (3)$$

两边积分得

$$\ln|y| = - \int p(x) dx + C \quad (4)$$

两边取自然指数得

$$y = \pm e^C e^{- \int p(x) dx} \quad (5)$$

把 $\pm e^C$ 整体看做一个任意常数 C ，上式变为。

$$y = C e^{- \int p(x) dx} \quad (6)$$

这就是一阶线性齐次微分方程^{式 2} 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 6} 中令 $y_0(x) = \exp(- \int p(x) dx)$ 。假设上式中的 C 是一个函数 $C(x)$ 而不是常数，代入^{式 1} 得

$$C' y_0 + C[y'_0 + p(x)y_0] = f(x) \quad (7)$$

由于 y_0 是齐次解，上式方括号中求和为 0，分离变量得

$$dC = \frac{f(x)}{y_0} dx \quad (8)$$

两边积分得

$$C(x) = \int \frac{f(x)}{y_0} dx \quad (9)$$

所以一阶线性非齐次微分方程的通解为

$$y = y_0 \int \frac{f(x)}{y_0} dx \quad (10)$$

其中

$$y_0(x) = e^{-\int p(x)dx} \quad (11)$$

注意待定常数包含在式 10 的不定积分中，式 11 中的不定积分产生的待定常数在代入式 10 后可消去.

二阶常系数齐次微分方程

预备知识 常微分方程^[71]，指数函数（复数）^[23]

二阶常系数齐次微分方程形式如下

$$ay'' + by' + cy = 0 \quad (1)$$

注意到指数函数 $y = Ce^{rx}$ 第 n 阶导数为 $r^n e^{rx}$ ，不妨尝试把指数函数代入方程，得

$$(ar^2 + br + c)e^{rx} = 0 \quad (2)$$

由于 $e^{rx} \neq 0$ ，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

$$y = C_1 e^{r_1 x} + C_2 e^{r_2 x} \quad (3)$$

2. 有一个重根 r ($b^2 - 4ac = 0$), 方程的通解为

$$y = C_1 e^{rx} + C_2 x e^{rx} \quad (4)$$

3. 有两个纯虚数根 $\pm i\omega_0$ ($b = 0, b^2 - 4ac < 0$), 方程的通解为

$$y = C_1 \cos(\omega_0 x) + C_2 \sin(\omega_0 x) \quad (5)$$

或

$$y = C_1 \cos(\omega_0 x + C_2) \quad (6)$$

其中 $\omega_0 = \sqrt{c/a}$.

4. 有两个复数根 $r \pm i\omega$ ($b \neq 0, b^2 - 4ac < 0$), 方程的通解为

$$y = e^{rx} [C_1 \cos(\omega x) + C_2 \sin(\omega x)] \quad (7)$$

或

$$y = C_1 e^{rx} \cos(\omega x + C_2) \quad (8)$$

其中

$$r = -\frac{b}{2a} \quad \omega = \frac{1}{2a} \sqrt{4ac - b^2} \quad (9)$$

详细推导

情况 1 的结论是显然的, 我们先来看情况 3. 根据 $y = C e^{rx}$ 的假设, 通解应该是

$$y = C_1 e^{i\omega x} + C_2 e^{-i\omega x} \quad (10)$$

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

$$C_1 = \frac{C_3}{2} + \frac{C_4}{2i} \quad C_2 = \frac{C_3}{2} - \frac{C_4}{2i} \quad (11)$$

代入上式得

$$\begin{aligned} y &= C_3 \frac{e^{i\omega x} + e^{-i\omega x}}{2} + C_4 \frac{e^{i\omega x} - e^{-i\omega x}}{2i} \\ &= C_3 \cos(\omega x) + C_4 \sin(\omega x) \end{aligned} \quad (12)$$

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 $\omega \rightarrow 0$ 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

$$\lim_{\omega \rightarrow 0} C_3 e^{rx} \frac{\sin(\omega x)}{\omega x} x = C_3 x e^{rx} \quad (13)$$

这里用到了“小角正弦值极限^[32]”中的结论。

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^[73]

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

$$y'' + by' + cy = f(x) \quad (1)$$

这就是二阶常系数非齐次微分方程。其解为

$$y(x) = C_1 y_1 + C_2 y_2 - y_1 \int \frac{y_2 f}{W} dx + y_2 \int \frac{y_1 f}{W} dx \quad (2)$$

其中 W 可以写成二阶行列式

$$W = \begin{vmatrix} y_1 & y_2 \\ y'_1 & y'_2 \end{vmatrix} = y_1 y'_2 - y'_1 y_2 \quad (3)$$

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。和是对应齐次方程

$$y'' + by' + cy = 0 \quad (4)$$

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

$$y = v_1 y_1 + v_2 y_2 \quad (5)$$

其中，也是关于的函数。对该式两边求导，得

$$y' = v'_1 y_1 + v'_2 y_2 + v_1 y'_1 + v_2 y'_2 \quad (6)$$

为了接下来计算方便，我们规定 v_1, v_2 满足关系¹⁵

$$v'_1 y_1 + v'_2 y_2 = 0 \quad (7)$$

把式 7 代入式 6，得到

$$y' = v_1 y'_1 + v_2 y'_2 \quad (8)$$

继续对求导，得到

$$y'' = v'_1 y'_1 + v'_2 y'_2 + v_1 y''_1 + v_2 y''_2 \quad (9)$$

把式 5 式 8 式 9 代回原方程式 1 得

$$(v'_1 y'_1 + v'_2 y'_2 + v_1 y''_1 + v_2 y''_2) + b(v_1 y'_1 + v_2 y'_2) + c(v_1 y_1 + v_2 y_2) = f \quad (10)$$

化简，得

$$(v'_1 y'_1 + v'_2 y'_2) + v_1(ay''_1 + by'_1 + cy_1) + v_2(ay''_2 + by'_2 + cy_2) = f \quad (11)$$

由于 y_1 和 y_2 都是式 4 的解，式 (9) 化为

$$v'_1 y'_1 + v'_2 y'_2 = f \quad (12)$$

¹⁵ 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。

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

$$\begin{cases} y_1 v'_1 + y_2 v'_2 = 0 \\ y'_1 v'_1 + y'_2 v'_2 = f \end{cases} \quad (13)$$

解得

$$\begin{cases} v'_1 = -y_2 f / W \\ v'_2 = y_1 f / W \end{cases} \quad (14)$$

其中

$$W = y_1 y'_2 - y_2 y'_1 = \begin{vmatrix} y_1 & y_2 \\ y'_1 & y'_2 \end{vmatrix} \quad (15)$$

对（13）的两条式子积分，即可得到

$$v_1 = - \int \frac{y_2 f}{W} dx + C_1 \quad (16)$$

$$v_2 = \int \frac{y_1 f}{W} dx + C_2 \quad (17)$$

(15) (16) 代入 (5) 式，得到方程 (1) 的解为

$$y(x) = C_1 y_1 + C_2 y_2 - y_1 \int \frac{y_2 f}{W} dx + y_2 \int \frac{y_1 f}{W} dx \quad (18)$$

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，我们只能找到没有零点的区间内找到方程式 1 的通解。

拓展阅读 一阶线性非齐次微分方程的通解

正交函数系

预备知识 定积分^[65]

因变量为实数的情况

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

$$\int_a^b f_i(x) f_i(x) dx \neq 0 \quad (1)$$

当整数 $m \neq n$ 时

$$\int_a^b f_m(x) f_n(x) dx = 0 \quad (2)$$

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点乘. 如果两个不同的矢量正交（垂直），则它们的点乘为零. 如果它们的模长不为零，则一个矢量点乘自身不为零.

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

$$\int_a^b f_i(x) f_i(x) dx = 1 \quad (3)$$

那么该正交函数系就是归一的. 其性质可以表示为

$$\int_a^b f_m(x) f_n(x) dx = \delta_{mn} \quad (4)$$

其中 δ_{mn} 是克罗内克 δ 函数 (Kronecker Delta Function).

因变量为复数的情况

若函数系中 $f_i(x)$ 的自变量为实数，因变量为复数，则正交的定义变为

$$\int_a^b f_i^*(x) f_j(x) dx \neq 0 \quad (i \neq j) \quad (5)$$

归一化的定义变为

$$\int_a^b f_i^*(x) f_i(x) dx = 1 \quad (6)$$

正交归一条件可以统一写成

$$\int_a^b f_i^*(x) f_j(x) dx = \delta_{ij} \quad (7)$$

傅里叶级数（三角）

预备知识 几何矢量, 定积分^[65]

结论

满足狄利克雷条件的周期函数 $f(x)$ (周期为 $2l$) 可以使用以下三角函数展开

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi}{l}x\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi}{l}x\right) \quad (1)$$

其中

$$a_n = \frac{1}{l} \int_{-l}^l f(x) \cos\left(\frac{n\pi}{l}x\right) dx \quad (2)$$

$$b_n = \frac{1}{l} \int_{-l}^l f(x) \sin\left(\frac{n\pi}{l}x\right) dx \quad (3)$$

狄利克雷条件: 函数值有限, 存在有限个间断点和有限个极值点.

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 项为零¹⁶. 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

$$f(x) = \frac{1}{2}[f(x) + f(-x)] + \frac{1}{2}[f(x) - f(-x)] \quad (4)$$

$f(x)$ 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.

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. 一般高等数学教材中不证明完备性, 这里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.

¹⁶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.

例 1 方波

首先定义一个方波为

$$f(x) = \begin{cases} 1 & 2k\pi < x \leq (2k+1)\pi \\ -1 & (2k-1)\pi < x \leq 2k\pi \end{cases} \quad (5)$$

注意 $f(x)$ 在 $x = k\pi$ 处存在间断点. 进行傅里叶级数展开得

$$f(x) = \sum_{k=0}^{+\infty} \frac{4}{\pi(2k+1)} \sin((2k+1)x) \quad (6)$$

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.

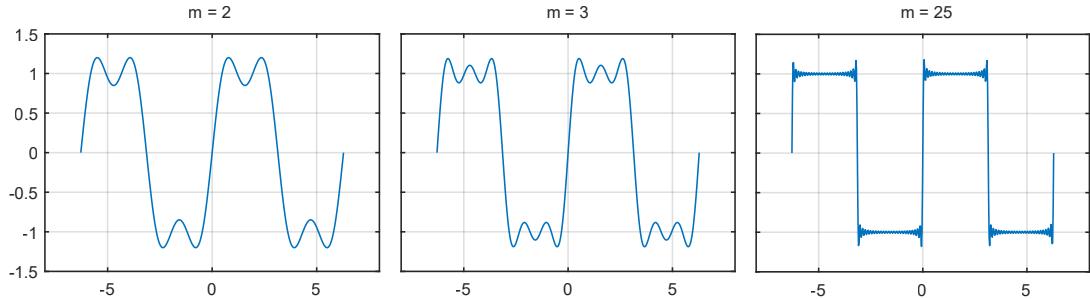


图 1: 有限项傅里叶级数逼近方波

例 2 正弦函数的绝对值

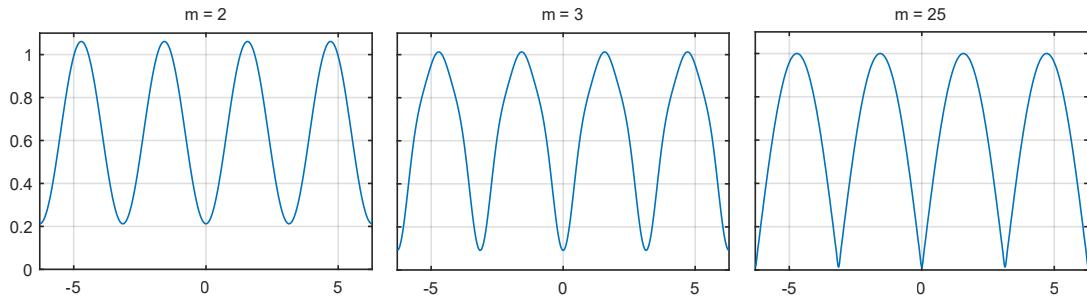
偶函数 $f(x) = |\sin x|$ 存在不光滑的点, 展开成傅里叶级数为

$$f(x) = |\sin x| = \frac{2}{\pi} - \frac{4}{\pi} \sum_{k=1}^{+\infty} \frac{1}{4k^2 - 1} \cos 2kx \quad (7)$$

前 m 项和画图如图 2.

系数公式推导

在泰勒级数^[50]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. 这里我们通过一个重要的类比来讲解, 即几何矢量在正交(不归一)基底上的展开.

图 2: 有限项傅里叶级数逼近 $|\sin(x)|$

给出一组无穷多个函数

$$\frac{1}{2}, \sin \frac{\pi}{l}x, \cos \frac{\pi}{l}x, \sin \frac{2\pi}{l}x, \cos \frac{2\pi}{l}x, \dots \sin \frac{n\pi}{l}x, \cos \frac{n\pi}{l}x \dots \quad (8)$$

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（式 1）。现在把两个任意函数（矢量） $f(x)$ 和 $g(x)$ 的点乘定义为它们的乘积在 $[-l, l]$ 内积分

$$\langle f|g \rangle = \int_{-l}^l f(x)g(x) dx \quad (9)$$

可以证明这组基底正交（即任意两个不同的基底点乘为 0，证明见词条最后）但不归一（某矢量与自身点乘等于 1）。与“几何矢量在正交但不归一的基底上展开”一样，我们只要把函数分别与各个基底点乘，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¹⁷。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。

¹⁷ 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 $f(x)$ 。但更常见的是使用正弦基底，经常在量子力学中使用，因为无限深势阱要求波函数在区间两端消失。

例 3 用正弦基底展开闭区间内的函数

定义区间 $[0, a]$ 内的一个三角形函数如下，先把函数归一化，再用正弦基底做傅里叶展开。

$$f(x) = \begin{cases} \frac{2}{a}x & (0 \leq x \leq \frac{a}{2}) \\ 2 - \frac{2}{a}x & (\frac{a}{2} < x \leq a) \end{cases} \quad (10)$$

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

$$b_n = \frac{2}{a} \int_0^a f(x) \sin\left(\frac{n\pi}{a}x\right) dx \quad (11)$$

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（如图 *） 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，所以

$$\begin{aligned} b_n &= \frac{4}{a} \int_0^{a/2} f(x) \sin\left(\frac{n\pi}{a}x\right) dx \\ &= \frac{8}{a^2} \int_0^{a/2} x \sin\left(\frac{n\pi}{a}x\right) dx \end{aligned} \quad (12)$$

这个积分既可以见例 *，也可以用 Wolfram Alpha 或 Mathematica 完成。结果是

$$b_n = (-1)^{\frac{n-1}{2}} \frac{8}{\pi^2 n^2} \quad (n \text{ 为奇数}) \quad (13)$$

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。

$$\sin \frac{\pi}{l}(x - x_0), \sin \frac{2\pi}{l}(x - x_0), \dots \sin \frac{n\pi}{l}(x - x_0) \dots \quad (14)$$

证明函数基底正交

现在证明任意两个不同的基底点乘等于 0。当 $m \neq n$ 时，有

$$\begin{aligned} &\int_{-\pi}^{\pi} \sin(mx) \sin(nx) dx \\ &= \frac{1}{2} \int_{-\pi}^{\pi} \cos[(m-n)x] dx - \frac{1}{2} \int_{-\pi}^{\pi} \cos[(m+n)x] dx = 0 \end{aligned} \quad (15)$$

$$\begin{aligned} & \int_{-\pi}^{\pi} \cos(mx) \cos(nx) dx \\ &= -\frac{1}{2} \int_{-\pi}^{\pi} \cos[(m-n)x] dx + \frac{1}{2} \int_{-\pi}^{\pi} \cos[(m+n)x] dx = 0 \end{aligned} \tag{16}$$

$$\begin{aligned} & \int_{-\pi}^{\pi} \sin(mx) \cos(nx) dx \\ &= \frac{1}{2} \int_{-\pi}^{\pi} \sin[(m+n)x] dx - \frac{1}{2} \int_{-\pi}^{\pi} \sin[(m-n)x] dx = 0 \end{aligned} \tag{17}$$

$$\int_{-\pi}^{\pi} 1 \cdot \sin x dx = 0 \tag{18}$$

$$\int_{-\pi}^{\pi} 1 \cdot \cos x dx = 0 \tag{19}$$

其中任意一个函数与自己点乘都等于 π (除了常函数 1 积分为 2π).

$$\int_{-\pi}^{\pi} \sin^2(nx) dx = \pi \tag{20}$$

$$\int_{-\pi}^{\pi} \cos^2(nx) dx = \pi \tag{21}$$

$$\int_{-\pi}^{\pi} 1^2 dx = 2\pi \tag{22}$$

傅里叶级数（指数）

结论

预备知识 傅里叶级数（三角）^[78], 欧拉公式^[23]

$f(x)$ 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

$$f(x) = \sum_{n=-\infty}^{+\infty} c_n \exp\left(i \frac{n\pi}{l} x\right) \tag{1}$$

其中

$$c_n = \frac{1}{2l} \int_{-l}^l f(x) \exp\left(-i\frac{n\pi}{l}x\right) dx \quad (2)$$

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^[78]的情况. 这时, 函数基底变为

$$f_n(x) = \exp\left(i\frac{n\pi}{l}x\right) \quad n \in N \quad (3)$$

定义复函数 $f(x)$ 与 $g(x)$ 的点乘为

$$\langle f | g \rangle = \int_{-l}^l f(x)^* g(x) dx \quad (4)$$

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

$$\langle f_m | f_n \rangle = 2l\delta_{mn} \quad (5)$$

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

$$\cos x = \frac{e^{ix} + e^{-ix}}{2} \quad \sin x = \frac{e^{ix} - e^{-ix}}{2i} \quad (6)$$

三角傅里叶级数的系数式 2^[78] 和式 3^[78] 可以用指数傅里叶级数的系数表示

$$\begin{aligned} a_n &= \frac{1}{l} \int_{-l}^l f(x) \cos\left(\frac{n\pi}{l}x\right) dx \\ &= \frac{1}{2l} \int_{-l}^l f(x) \exp\left(i\frac{n\pi}{l}x\right) dx + \frac{1}{2l} \int_{-l}^l f(x) \exp\left(-i\frac{n\pi}{l}x\right) dx \\ &= c_{-n} + c_n \end{aligned} \quad (7)$$

同理,

$$b_n = \frac{c_{-n} - c_n}{i} \quad (8)$$

注意这里全都有 $n \geq 0$. 由以上两式, 也可以解得

$$c_n = \frac{a_n - ib_n}{2} \quad c_{-n} = \frac{a_n + ib_n}{2} \quad (9)$$

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

$$c_{-n} = c_n^* \quad (10)$$

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^[78]，同样根据式 9 可得，两种情况分别对应

$$c_{-n} = c_n = \frac{a_n}{2} \quad c_{-n} = -c_n = i\frac{b_n}{2} \quad (11)$$

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。

第三章

线性代数

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^[88] 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。

矢量的点乘^[91] 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点乘再开方，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**。若两矢量点乘为零，这两个矢量相互**正交**¹。

两矢量叉乘^[96] 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^[94]（所有基底模长

¹对于几何矢量，正交就是方向垂直，不加区分。

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点乘得到.

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². 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^[104]. 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^[110]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

几何矢量

我们来回顾高中的几何矢量, 以下简称为“矢量”. 要强调的是, 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. 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} .

矢量的坐标

我们先在直角坐标系中对矢量的坐标做一个临时的简单的定义(具体的定义要在学习了矢量基底的概念以后才能掌握), 我们姑且先把坐标理解为矢量在 x 和 y 轴上的投影长度. 所以如果平移矢量使起点与坐标原点重合, 那么矢量的坐标就是矢量终点的的坐标.

²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

矢量的加法

如图 1, 两个矢量相加, 既可以使用平行四边形法则, 也可以用三角形法则. 若有多个矢量连续相加, 可以分别把它们首尾相接, 结果就是由起点指向终点的矢量. 容易证明矢量的加法满足加法交换律 $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$, 结合律 $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$.

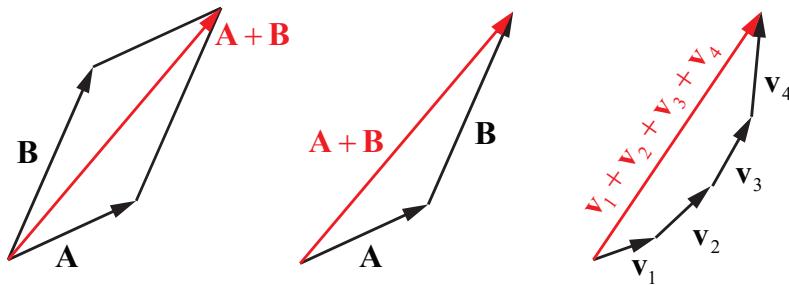


图 1: 矢量的加法

矢量的数乘 共线

如图 2, 一个矢量与一个实数相乘, 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 $\lambda(\mu\mathbf{A}) = (\lambda\mu)\mathbf{A}$.

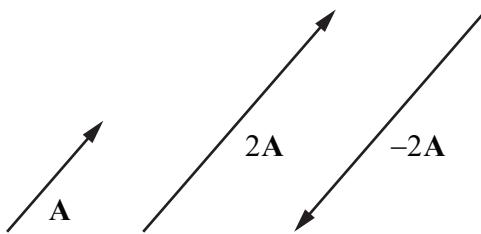


图 2: 矢量的数乘

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^[7] 是, 两矢量方向相同或相反.

矢量的线性组合

把若干矢量 \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

$$\sum_i^N c_i \mathbf{v}_i = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_N \mathbf{v}_N \quad (1)$$

线性相关 线性无关

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的

$$\sum_i^N c_i \mathbf{v}_i = \mathbf{0} \quad (2)$$

这是因为对于任何一个不为零的项 j , 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

$$\mathbf{v}_j = \sum_{i \neq j} \frac{c_i}{c_j} \mathbf{v}_i \quad (3)$$

反之, 如果不存在这样的系数, 这些矢量就是线性无关的.

基底 矢量空间 坐标

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以及它们的加法和数乘运算组成一个一维矢量空间. 一个平面上的所有矢量以及它们的加法和数乘运算, 组成一个二维矢量空间, 二维空间中最多只能找到两个线性无关的矢量. 三维空间同理.

N 维空间中的任意一组线性无关的 N 个矢量 $\beta_1 \dots \beta_N$ 可以作为一组矢量基底, 记为 $\{\beta_i\}$. 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), 即存在不全为零的实数 $c_1 \dots c_{N+1}$ 使下式成立

$$\sum_{i=1}^N c_i \beta_i + c_{N+1} \mathbf{v} = \mathbf{0} \quad (4)$$

我们还可以得知 c_{N+1} 必不为零 (反证法: 如果 $c_{N+1} = 0$, 则可得出基底 $\{\beta_i\}$ 线性相关, 不成立), 所以由式3可知 \mathbf{v} 可用 $\{\beta_i\}$ 的线性组合表示. 令 $x_i = c_i/c_{N+1}$, 该空间中任意矢量 \mathbf{v} 都有

$$\mathbf{v} = \sum_{i=1}^N x_i \beta_i \quad (5)$$

这里的 x_i 就是矢量空间中坐标^{的定义.}

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 x_i 都可以是上式成立, 分别记为 x_i 和 y_i . 那么分别代入上式再把两式相减得到

$$\sum_{i=1}^N (x_i - y_i) \beta_i = \mathbf{0} \quad (6)$$

由于 $(x_i - y_i)$ 不全为零, 得到基底 $\{\beta_i\}$ 线性相关, 而这是不可能的. 证毕.

拓展阅读 矢量的点乘^[91], 正交归一基^[94], 矢量的叉乘^[96]

矢量点乘

预备知识 几何矢量^[88]

几何定义

该词条只讨论几何矢量的点乘, 所以我们先来看几何定义. 注意该定义不需要任何坐标系的概念.

如图1, 两个几何矢量的点乘 (**dot product**)³ 就是把它们的模长相乘, 再乘以它们的夹角的余弦值. 即

$$\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}| |\mathbf{B}| \cos \theta \quad (1)$$

其中 θ 是两个矢量的夹角. 如图, 几何定义中, 既可以把点乘理解为 \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}

³也叫点积, 标量积 (**scalar product**) 或内积 (**inner product**)

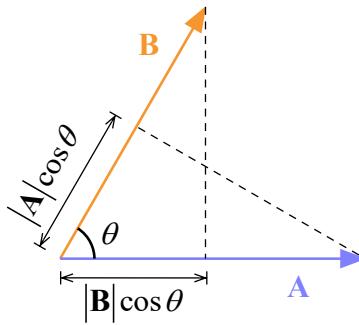


图 1: 点乘的几何定义

的模长⁴. 可见当两矢量模长不变时, 若方向相同, 点乘取最大值 $|\mathbf{A}||\mathbf{B}|$; 若方向相反, 点乘取最小值 $-|\mathbf{A}||\mathbf{B}|$; 若相互垂直, 则点乘为 0.

我们说两个点乘为 0 的矢量互相正交, 对于几何矢量, 正交与垂直没有区别. 几何矢量与自身点乘可得该矢量模长的平方. 我们把模长为 1 的矢量称为 **单位矢量**, 单位矢量与自身点乘等于 1. 我们在矢量上面加上“^”符号表示单位矢量, 如 $\hat{\mathbf{x}}$. 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.

点乘的性质

1. 交换律⁵

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A} \quad (2)$$

2. 分配律⁶

$$\mathbf{A} \cdot (\mathbf{B} + \mathbf{C}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C} \quad (3)$$

注意点乘不满足结合律, 即

$$(\mathbf{A} \cdot \mathbf{B}) \cdot \mathbf{C} \neq \mathbf{A} \cdot (\mathbf{B} \cdot \mathbf{C}) \quad (4)$$

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.

⁴在这种理解下, 若量矢量的夹角为钝角, 投影长度取负值

⁵由式 1 易证

⁶证明见词条最后.

点乘的代数定义

预备知识 正交归一基^[94]

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点乘运算的结果呢? 先用正交归一基^[94] 将两矢量展开

$$\mathbf{A} = A_x \hat{\mathbf{x}} + A_y \hat{\mathbf{y}} \quad \mathbf{B} = B_x \hat{\mathbf{x}} + B_y \hat{\mathbf{y}} \quad (5)$$

所以

$$\mathbf{A} \cdot \mathbf{B} = (A_x \hat{\mathbf{x}} + A_y \hat{\mathbf{y}}) \cdot (B_x \hat{\mathbf{x}} + B_y \hat{\mathbf{y}}) \quad (6)$$

根据分配律式 3, 我们可以把两个括号拆开, 变为 4 个点乘之和.

$$\mathbf{A} \cdot \mathbf{B} = A_x B_x \hat{\mathbf{x}} \cdot \hat{\mathbf{x}} + A_y B_y \hat{\mathbf{y}} \cdot \hat{\mathbf{y}} + A_x B_y \hat{\mathbf{x}} \cdot \hat{\mathbf{y}} + A_y B_x \hat{\mathbf{y}} \cdot \hat{\mathbf{x}} \quad (7)$$

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

$$\mathbf{A} \cdot \mathbf{B} = A_x B_x + A_y B_y \quad (8)$$

同理, 可以在三维直角坐标系 xyz 中把点乘结果用坐标表示

$$\mathbf{A} \cdot \mathbf{B} = A_x B_x + A_y B_y + A_z B_z \quad (9)$$

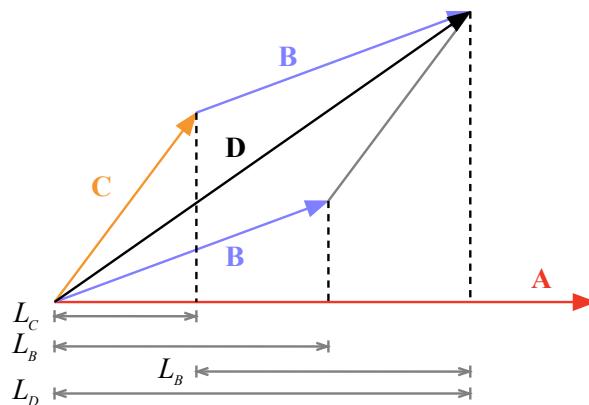


图 2: 点乘分配律的证明

证明点乘的分配律

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 矢量点乘^[91]

我们已经知道了矢量基底的概念, 如果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或单位正交基. 若把这组正交归一基记为 $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 \dots \hat{\mathbf{x}}_n$, 则正交归一可以用点乘表示为

$$\hat{\mathbf{x}}_i \cdot \hat{\mathbf{x}}_j = \delta_{ij} \quad (1)$$

其中 δ_{ij} 是克罗内克 δ 函数 (**Kronecker delta function**), 定义为

$$\delta_{ij} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \quad (2)$$

任意矢量在单位正交基上的展开

$$\mathbf{v} = \sum_{i=1}^n (\mathbf{v} \cdot \hat{\mathbf{x}}_i) \hat{\mathbf{x}}_i = \sum_{i=1}^n v_i \hat{\mathbf{x}}_i \quad (3)$$

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.

$$\mathbf{v} = (\mathbf{v} \cdot \hat{\mathbf{x}}) \hat{\mathbf{x}} + (\mathbf{v} \cdot \hat{\mathbf{y}}) \hat{\mathbf{y}} + (\mathbf{v} \cdot \hat{\mathbf{z}}) \hat{\mathbf{z}} = v_x \hat{\mathbf{x}} + v_y \hat{\mathbf{y}} + v_z \hat{\mathbf{z}} \quad (4)$$

证明

由于任何矢量都可以表示成基底 $\hat{\mathbf{x}}_1 \dots \hat{\mathbf{x}}_n$ 的线性组合, 设

$$\mathbf{v} = \sum_{i=1}^n c_i \hat{\mathbf{x}}_i \quad (5)$$

用 $\hat{\mathbf{x}}_k$ 乘以等式两边，得

$$\mathbf{v} \cdot \hat{\mathbf{x}}_k = \sum_{i=1}^n c_i \hat{\mathbf{x}}_i \cdot \hat{\mathbf{x}}_k = \sum_{i=1}^n c_i \delta_{ik} = c_k \quad (6)$$

所以式 5 中的系数有唯一确定的值 $c_k = \mathbf{v} \cdot \hat{\mathbf{x}}_k$. 证毕.

右手定则

物理中常用右手定则来判断方向，右手定则分为以下两种.

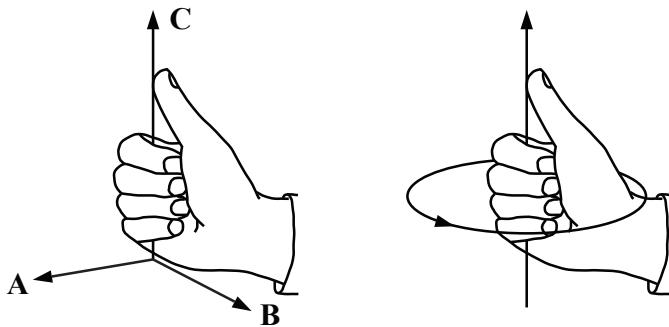


图 1: 两种右手定则

第一种右手定则

假设有两个不共线的矢量，第一个为 \mathbf{A} ，第二个为 \mathbf{B} ，它们可以定义一个平面. 现在我们想根据 \mathbf{A} 和 \mathbf{B} 确定该平面的法向量 \mathbf{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 \mathbf{A} ，再将四指弯向 \mathbf{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”.

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

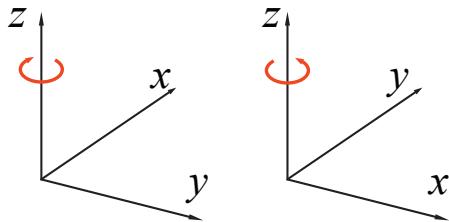


图 2: 左手系 (左) 与右手系 (右)

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.

矢量叉乘

预备知识 右手定则^[95], 三阶行列式^[115]

叉乘的几何定义

两个矢量 \mathbf{A} , \mathbf{B} 的叉乘 (cross product)⁷, 是一个矢量 \mathbf{C} . 叉乘用 “ \times ” 表示, 且不可省略, 即 $\mathbf{A} \times \mathbf{B} = \mathbf{C}$. 要确定一个矢量, 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 ($0 \leq \theta \leq \pi$) 的正弦值相乘.

$$|\mathbf{C}| = |\mathbf{A}| |\mathbf{B}| \sin \theta \quad (1)$$

2. \mathbf{C} 的方向垂直于 \mathbf{A}, \mathbf{B} 所在的平面, 且由右手定则^[95] 决定.

与点乘和数乘不同, 叉乘不满足交换律. 根据几何定义, $\mathbf{B} \times \mathbf{A}$ 与 $\mathbf{A} \times \mathbf{B}$ 模长相等, 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

$$\mathbf{B} \times \mathbf{A} = -\mathbf{A} \times \mathbf{B} \quad (2)$$

⁷也叫叉积 (cross product), 向量积 (vector product) 或矢量积

叉乘与数乘的混合运算

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 \mathbf{A} 和 \mathbf{B} 的模长相乘. 假设 λ 为常数 (标量), 显然有

$$(\lambda\mathbf{A}) \times \mathbf{B} = \mathbf{A} \times (\lambda\mathbf{B}) = \lambda(\mathbf{A} \times \mathbf{B}) \quad (3)$$

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.

$$\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C} \quad (4)$$

由式 2 及上式可以推出

$$(\mathbf{A} + \mathbf{B}) \times \mathbf{C} = -\mathbf{C} \times (\mathbf{A} + \mathbf{B}) = -\mathbf{C} \times \mathbf{A} - \mathbf{C} \times \mathbf{B} = \mathbf{A} \times \mathbf{C} + \mathbf{B} \times \mathbf{C} \quad (5)$$

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^[99]).

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

$$\hat{\mathbf{x}} \times \hat{\mathbf{y}} = \hat{\mathbf{z}} \quad \hat{\mathbf{y}} \times \hat{\mathbf{z}} = \hat{\mathbf{x}} \quad \hat{\mathbf{z}} \times \hat{\mathbf{x}} = \hat{\mathbf{y}} \quad (6)$$

由关系式 6 可得

$$\hat{\mathbf{y}} \times \hat{\mathbf{x}} = -\hat{\mathbf{z}} \quad \hat{\mathbf{z}} \times \hat{\mathbf{y}} = -\hat{\mathbf{x}} \quad \hat{\mathbf{x}} \times \hat{\mathbf{z}} = -\hat{\mathbf{y}} \quad (7)$$

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

$$\hat{\mathbf{x}} \times \hat{\mathbf{x}} = \mathbf{0} \quad \hat{\mathbf{y}} \times \hat{\mathbf{y}} = \mathbf{0} \quad \hat{\mathbf{z}} \times \hat{\mathbf{z}} = \mathbf{0} \quad (8)$$

式 6, 式 7 和式 8 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

$$\mathbf{A} = a_x \hat{\mathbf{x}} + a_y \hat{\mathbf{y}} + a_z \hat{\mathbf{z}} \quad \mathbf{B} = b_x \hat{\mathbf{x}} + b_y \hat{\mathbf{y}} + b_z \hat{\mathbf{z}} \quad (9)$$

(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 4 式 5), 可得到如下 9 项

$$\begin{aligned}\mathbf{A} \times \mathbf{B} &= (a_x \hat{\mathbf{x}} + a_y \hat{\mathbf{y}} + a_z \hat{\mathbf{z}}) \times (b_x \hat{\mathbf{x}} + b_y \hat{\mathbf{y}} + b_z \hat{\mathbf{z}}) \\ &= + a_x b_x (\hat{\mathbf{x}} \times \hat{\mathbf{x}}) + a_x b_y (\hat{\mathbf{x}} \times \hat{\mathbf{y}}) + a_x b_z (\hat{\mathbf{x}} \times \hat{\mathbf{z}}) \\ &\quad + a_y b_x (\hat{\mathbf{y}} \times \hat{\mathbf{x}}) + a_y b_y (\hat{\mathbf{y}} \times \hat{\mathbf{y}}) + a_y b_z (\hat{\mathbf{y}} \times \hat{\mathbf{z}}) \\ &\quad + a_z b_x (\hat{\mathbf{z}} \times \hat{\mathbf{x}}) + a_z b_y (\hat{\mathbf{z}} \times \hat{\mathbf{y}}) + a_z b_z (\hat{\mathbf{z}} \times \hat{\mathbf{z}})\end{aligned}\tag{10}$$

注意每一项中的运算在式 6, 式 7 和式 8 中都能找到答案, 于是上式化为

$$\mathbf{A} \times \mathbf{B} = (a_y b_z - a_z b_y) \hat{\mathbf{x}} + (a_x b_z - a_z b_x) \hat{\mathbf{y}} + (a_x b_y - a_y b_x) \hat{\mathbf{z}}\tag{11}$$

令 $\mathbf{C} = \mathbf{A} \times \mathbf{B}$, 则 \mathbf{C} 的分量表达式为

$$\begin{cases} c_x = a_y b_z - a_z b_y \\ c_y = a_x b_z - a_z b_x \\ c_z = a_x b_y - a_y b_x \end{cases}\tag{12}$$

式 11 可以用三阶行列式^[115] 表示为

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix}\tag{13}$$

与普通行列式不同的是, 这个行列式中的元有部分是矢量, 所以得出的结果也是矢量.

例 1

空间直角坐标系中三角形的三点分别为 $O(0, 0, 0)$, $A(1, 1, 0)$, $B(-11, 1)$. 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

$$\begin{aligned}\mathbf{a} &= (1, 1, 0) - (0, 0, 0) = (1, 1, 0) \\ \mathbf{b} &= (-11, 1, 1) - (0, 0, 0) = (-11, 1, 1)\end{aligned}\tag{14}$$

三角形的面积为

$$S = \frac{1}{2} ab \sin \theta\tag{15}$$

其中 θ 是 \mathbf{a} 与 \mathbf{b} 的夹角。根据叉乘的几何定义中的式 1

$$S = \frac{1}{2}ab \sin \theta = \frac{1}{2}|\mathbf{a} \times \mathbf{b}| \quad (16)$$

令

$$\mathbf{v} = \mathbf{a} \times \mathbf{b} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ 1 & 1 & 0 \\ -1 & 1 & 1 \end{vmatrix} = \hat{\mathbf{x}} - \hat{\mathbf{y}} + 2\hat{\mathbf{z}} \quad (17)$$

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⁸ 得单位法向量为

$$\hat{\mathbf{v}} = \frac{\mathbf{v}}{|\mathbf{v}|} = \frac{(1, -1, 2)}{\sqrt{6}} = \left(\frac{\sqrt{6}}{6}, -\frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{3} \right) \quad (18)$$

矢量叉乘分配律的几何证明

预备知识 矢量的叉乘^[96]

证明 $\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C}$ 首先令

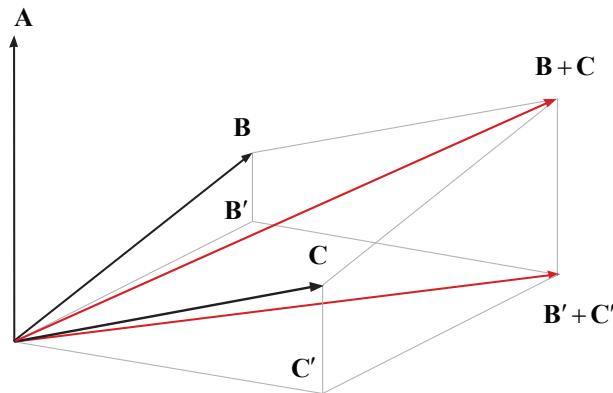


图 1: 把 $\mathbf{B}, \mathbf{C}, \mathbf{D}$ 投影到与 \mathbf{A} 垂直的平面上

$$\mathbf{D} = \mathbf{B} + \mathbf{C} \quad (1)$$

⁸把矢量长度变为 1, 方向不变

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

$$\mathbf{A} \times \mathbf{B} = \mathbf{A} \times \mathbf{B}' \quad (2)$$

这是叉乘的一个基本的性质. 首先，根据叉乘的几何定义^[96]， $\mathbf{A} \times \mathbf{B}$ 与 $\mathbf{A} \times \mathbf{B}'$ 的方向相同. 另外

$$|\mathbf{A} \times \mathbf{B}| = |\mathbf{A}| |\mathbf{B}| \sin \theta_{AB} = |\mathbf{A}| |\mathbf{B}'| = |\mathbf{A} \times \mathbf{B}'| \quad (3)$$

所以二者模长也相等，证毕.

同理有

$$\mathbf{A} \times \mathbf{C} = \mathbf{A} \times \mathbf{C}' \quad (4)$$

$$\mathbf{A} \times \mathbf{D} = \mathbf{A} \times \mathbf{D}' \quad (5)$$

所以，要证明

$$\mathbf{A} \times \mathbf{D} = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C} \quad (6)$$

只需要证明

$$\mathbf{A} \times \mathbf{D}' = \mathbf{A} \times \mathbf{B}' + \mathbf{A} \times \mathbf{C}' \quad (7)$$

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，“ \mathbf{D}' 乘以 $|\mathbf{A}|$ 旋转 90° ”和“ \mathbf{B}' 与 \mathbf{C}' 分别乘以 $|\mathbf{A}|$ 旋转 90° 再相加”结果相同，而这显然成立. 证毕.

连续叉乘的化简

预备知识 矢量的叉乘^[96]

连续两个叉乘的化简也叫 BAC-CAB 定理

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B} \cdot (\mathbf{A} \cdot \mathbf{C}) - \mathbf{C} \cdot (\mathbf{A} \cdot \mathbf{B}) \quad (1)$$

$$(\mathbf{B} \times \mathbf{C}) \times \mathbf{A} = \mathbf{C} \cdot (\mathbf{A} \cdot \mathbf{B}) - \mathbf{B} \cdot (\mathbf{A} \cdot \mathbf{C}) \quad (2)$$

这个定理既可以“暴力破解”，也可以用几何的方法证明。以下对几何原理略作说明，便可解释该公式的结构，方便理解和记忆。考察第一条式子，先计算 $\mathbf{B} \times \mathbf{C}$ （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} ，所以结果还是落到 \mathbf{B} 和 \mathbf{C} 所在平面，所以等式右边是 \mathbf{B} 和 \mathbf{C} 的线性组合。

这里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。

$$\begin{aligned}\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) &= \underbrace{\mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})}_{\text{括号外的矢量在左边}} \\ (\mathbf{B} \times \mathbf{C}) \times \mathbf{A} &= \underbrace{\mathbf{C}(\mathbf{A} \cdot \mathbf{B}) - \mathbf{B}(\mathbf{A} \cdot \mathbf{C})}_{\text{括号外的矢量在左边}}\end{aligned}$$

图 1: 三矢量叉乘的化简

三矢量的混合积

预备知识 矢量的叉乘^[96]

定义

任意的三个矢量，如果其中两个叉乘再与第三个矢量点乘，那么这种运算叫做这三个矢量的混合积

$\mathbf{A} \times \mathbf{B} \cdot \mathbf{C}$ 是矢量 $\mathbf{A}, \mathbf{B}, \mathbf{C}$ 的混合积，且满足

$$\mathbf{A} \times \mathbf{B} \cdot \mathbf{C} = \mathbf{B} \times \mathbf{C} \cdot \mathbf{A} = \mathbf{C} \times \mathbf{A} \cdot \mathbf{B} \tag{1}$$

这个公式可由下图记忆。混合积的方向由叉乘的方向决定，与点乘无关。如果混合积的顺序取与右图相反的方向，根据叉乘的定义，需要在前面加上负号（叉乘不满足乘法交换律！）。下面三式与上面三式互为相反数。

$$\mathbf{C} \times \mathbf{B} \cdot \mathbf{A} = \mathbf{B} \times \mathbf{A} \cdot \mathbf{C} = \mathbf{A} \times \mathbf{C} \cdot \mathbf{B} \tag{2}$$

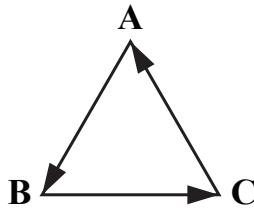


图 1: 式 1 记忆法

另外要注意混合积的方向是由叉乘的次序所决定的，与点乘的次序无关。所以
上式也可记为

$$\mathbf{A} \cdot (\mathbf{C} \times \mathbf{B}) = \mathbf{C} \cdot (\mathbf{B} \times \mathbf{A}) = \mathbf{B} \cdot (\mathbf{A} \times \mathbf{C}) \quad (3)$$

注意括号是必须的， $\mathbf{A} \cdot \mathbf{C} \times \mathbf{B}$ 是错误的式子，因为 $\mathbf{A} \cdot \mathbf{C}$ 是常数，不能与 \mathbf{B}
叉乘。

几何法证明

设三个矢量都以原点作为起点，以三个矢量为棱作平行六面体（如图，）。
从矢量的叉乘^[96]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为底面积乘以高

$$V = |\mathbf{A} \times \mathbf{B}| |\hat{\mathbf{n}} \cdot \mathbf{C}| = |\mathbf{A} \times \mathbf{B} \cdot \mathbf{C}| \quad (4)$$

同理可得对于同一平行六面体

$$V = |\mathbf{B} \times \mathbf{C} \cdot \mathbf{A}| = |\mathbf{C} \times \mathbf{A} \cdot \mathbf{B}| \quad (5)$$

这里只证明了绝对值，证明正负关系并不难，留给读者思考。

代数法证明

不难证明三矢积若展开成分量的形式，等于三个矢量组成的行列式。而
利用行列式中任意两行置换符号改变，即可证明式 1.

平面旋转变换

预备知识 三角恒等式^[4], 极坐标系^[9]

结论

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

$$x' = (\cos \alpha)x + (-\sin \alpha)y \quad (1)$$

$$y' = (\sin \alpha)x + (\cos \alpha)y \quad (2)$$

其逆变换如下, 即已知 $P'(x', y')$ 求 $P(x, y)$

$$x = (\cos \alpha)x' + (\sin \alpha)y' \quad (3)$$

$$y = (-\sin \alpha)x' + (\cos \alpha)y' \quad (4)$$

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

$$x = r \cos \theta \quad y = r \sin \theta \quad (5)$$

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^[4]) 化简如下

$$x' = r \cos(\theta + \alpha) = r \cos \theta \cos \alpha - r \sin \theta \sin \alpha = x \cos \alpha - y \sin \alpha \quad (6)$$

$$y' = r \sin(\theta + \alpha) = r \sin \theta \cos \alpha + r \cos \theta \sin \alpha = x \sin \alpha + y \cos \alpha \quad (7)$$

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

$$x = \cos(-\alpha)x' - \sin(-\alpha)y' = \cos(\alpha)x' + \sin(\alpha)y' \quad (8)$$

$$y = \sin(-\alpha)x' + \cos(-\alpha)y' = -\sin(\alpha)x' + \cos(\alpha)y' \quad (9)$$

证毕.

线性变换

预备知识 平面旋转变换^[102]

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，得到另外几个数的过程就叫线性变换。例如，在直角坐标的旋转变换中，直角坐标系中任意一点 P 的坐标 (x, y) 绕原点旋转角 α 以后的坐标为

$$\begin{cases} x' = (\cos \alpha)x + (-\sin \alpha)y \\ y' = (\sin \alpha)x + (\cos \alpha)y \end{cases} \quad (1)$$

这就是一个典型的线性变换，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。求逆变换的一般方法就是把等号右边的 n 个变量作为未知数，求解 n 元一次方程组即可得到逆变换的表达式。

线性变换的矩阵表示

由 n 个数 $x_1 \dots x_n$ 变换到 m 个数 $y_1 \dots y_m$ 的线性变换的一般形式为

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ y_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{cases} \quad (2)$$

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，就是该变换的矩阵。

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \quad (3)$$

矩阵

预备知识 线性变换^[104]

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 列的矩阵元表示为 A_{ij} 。

矩阵的转置

任意矩阵 \mathbf{A} 的转置（Transpose）记为 \mathbf{A}^T 。转置操作把 \mathbf{A} 的第 i 行变为 \mathbf{A}^T 的第 i 列，相当于把矩阵沿对角线翻转 180° 。即任意矩阵元满足

$$A_{ij}^T = A_{ji} \quad (1)$$

行矢量和列矢量可看做特殊的矩阵，行矢量转置后变为列矢量，反之亦然

$$(x_1, x_2 \dots, x_n)^T = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad (2)$$

为了排版方便，在正文中通常用 $(x_1, x_2 \dots, x_n)^T$ 表示列矢量。另外，若一个矩阵的行数等于列数，这个矩阵就是一个方阵。

矩阵的乘法

矩阵最常见的运算是矩阵的乘法。线性变换^[104]

$$\left\{ \begin{array}{l} y_1 = A_{11}x_1 + A_{12}x_2 + \dots + A_{1n}x_n \\ y_2 = A_{21}x_1 + A_{22}x_2 + \dots + A_{2n}x_n \\ \vdots \\ y_m = A_{m1}x_1 + A_{m2}x_2 + \dots + A_{mn}x_n \end{array} \right. \quad (3)$$

可用矩阵与列矢量的乘法表示为

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad (4)$$

令列矢量 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$, $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$, 系数矩阵为 \mathbf{A} , 上式可记为

$$\mathbf{y} = \mathbf{Ax} \quad (5)$$

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, \dots, x_n 分别相乘再相加, 即点乘^[91] 的代数定义

$$y_i = \sum_{j=1}^n A_{ij} x_j \quad (6)$$

若有 l 个不同的 \mathbf{x} 和 \mathbf{y} , 第 k 个记为 $\mathbf{x}_k = (x_{1k}, \dots, x_{nk})^T$ 和 $\mathbf{y}_k = (y_{1k}, \dots, y_{mk})^T$, 对应的变换为

$$\begin{pmatrix} y_{1k} \\ y_{2k} \\ \vdots \\ y_{mk} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{pmatrix} \begin{pmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{nk} \end{pmatrix} \quad (7)$$

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

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1l} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nl} \end{pmatrix} \quad \mathbf{Y} = \begin{pmatrix} y_{11} & \cdots & y_{1l} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{ml} \end{pmatrix} \quad (8)$$

现在把 l 组线性变换用一条式子表示为

$$\begin{pmatrix} y_{11} & \cdots & y_{1l} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{ml} \end{pmatrix} = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} x_{11} & \cdots & x_{1l} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nl} \end{pmatrix} \quad (9)$$

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 列点乘.

$$\mathbf{Y} = \mathbf{AX} \quad (10)$$

矩阵元公式为

$$Y_{ij} = \sum_{k=1}^n A_{ik} X_{kj} \quad (11)$$

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.

根据定义, 容易证明矩阵乘法满足分配律 $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$, 但一般不满足交换律, 举一个反例:

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \quad (12)$$

矩阵的乘法分配律

下面证明矩阵的乘法分配律

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC} \quad (13)$$

$$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC} \quad (14)$$

令式 13 左边等于矩阵 \mathbf{D} , 则其矩阵元为

$$D_{ij} = \sum_k A_{ik}(B_{kj} + C_{kj}) \quad (15)$$

拆括号得

$$D_{ij} = \sum_k A_{ik}B_{kj} + \sum_k A_{ik}C_{kj} \quad (16)$$

而这恰好是 $\mathbf{AB} + \mathbf{AC}$ 的矩阵元. 证毕. 式 14 的证明类似.

矩阵乘法的结合律

现在来看三个矩阵相乘, 令

$$\mathbf{D} = \mathbf{A}(\mathbf{BC}) \quad (17)$$

这里的括号是为了强调顺序. 即使没有括号, 习惯上也是从右向左计算. \mathbf{D} 的矩阵元为

$$D_{ij} = \sum_l A_{il}(BC)_{lj} = \sum_l A_{il} \left(\sum_k B_{lk} C_{kj} \right) \quad (18)$$

拆括号, 得

$$D_{ij} = \sum_k \sum_l (A_{il} B_{lk} C_{kj}) \quad (19)$$

对 C_{kj} 进行合并同类项, 得

$$D_{ij} = \sum_k \left(\sum_l A_{il} B_{lk} \right) C_{kj} \quad (20)$$

括号中恰好是 \mathbf{A} 乘以 \mathbf{B} 所得矩阵的矩阵元 $(AB)_{ik}$ 所以

$$D_{ij} = \sum_k (AB)_{ik} C_{kj} \quad (21)$$

即

$$\mathbf{D} = (\mathbf{AB})\mathbf{C} \quad (22)$$

证毕.

单位矩阵

单位矩阵就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 2^[94]) 表示为

$$I_{ij} = \delta_{ij} \quad (23)$$

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.

逆矩阵

记 \mathbf{M} 的逆矩阵为 \mathbf{M}^{-1} , 且满足

$$\mathbf{M}^{-1}\mathbf{M} = \mathbf{I} \quad (24)$$

其中 \mathbf{I} 是单位矩阵. 也就是说, 任意一个矩阵 \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.

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

$$\mathbf{y} = \mathbf{M}\mathbf{x} \quad (25)$$

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

$$\mathbf{x} = \mathbf{M}^{-1}\mathbf{y} \quad (26)$$

所以如果要求逆矩阵, 我们只需要把式 25 看做 N 元一次方程组, 解出所有的 x_i 即可得到式 26 进而得到 \mathbf{M}^{-1} . 要注意的是, 只有在 \mathbf{M} 的行数大于等于列数时, 才有可能得到唯一的解, 具体讨论从略.

单位正交阵

预备知识 正交归一基底^[94], 矩阵^[105]

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

$$\sum_k U_{ki}U_{kj} = \delta_{ij} \quad (1)$$

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.

行矢量的性质

单位正交阵的一个性质是其所有的行也同样满足正交归一的条件

$$\sum_k U_{ik}U_{jk} = \delta_{ij} \quad (2)$$

所以我们也可以把定义中的所有“列”字换为“行”字.

现在我们以二维的单位正交阵 \mathbf{U} 为例来证明。我们把它的两列分别记为单位矢量 $\hat{\mathbf{u}}$ 和 $\hat{\mathbf{v}}$ ，则它的矩阵元可以用 $\hat{\mathbf{u}}$, $\hat{\mathbf{v}}$ 和正交归一基底 $\hat{\mathbf{x}} = (1, 0)^T$, $\hat{\mathbf{y}} = (0, 1)^T$ 的点乘来表示

$$\mathbf{U} = \begin{pmatrix} \hat{\mathbf{u}} \cdot \hat{\mathbf{x}} & \hat{\mathbf{v}} \cdot \hat{\mathbf{x}} \\ \hat{\mathbf{u}} \cdot \hat{\mathbf{y}} & \hat{\mathbf{v}} \cdot \hat{\mathbf{y}} \end{pmatrix} \quad (3)$$

由于 $\hat{\mathbf{u}}$, $\hat{\mathbf{v}}$ 和 $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ 一样也是正交归一基底，我们可以把矩阵 \mathbf{U} 的两行分别看做 $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ 在 $\hat{\mathbf{u}}$, $\hat{\mathbf{v}}$ 基底上的坐标，所以 \mathbf{U} 的两行也分别正交归一。证毕。

逆矩阵与转置矩阵

单位正交阵的另一个性质就是其逆矩阵等于其转置矩阵。还是以上面的 \mathbf{U} 为例来证明，我们先来看线性变换

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{u}} \cdot \hat{\mathbf{x}} & \hat{\mathbf{v}} \cdot \hat{\mathbf{x}} \\ \hat{\mathbf{u}} \cdot \hat{\mathbf{y}} & \hat{\mathbf{v}} \cdot \hat{\mathbf{y}} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \quad (4)$$

事实上，该变换就是把一个矢量以 $\hat{\mathbf{u}}$ 和 $\hat{\mathbf{v}}$ 作为基底的坐标 (u, v) 变换为同一个矢量以 $\hat{\mathbf{x}}$ 和 $\hat{\mathbf{y}}$ 作为基底的坐标 (x, y) ，即把矢量 $u\hat{\mathbf{u}} + v\hat{\mathbf{v}}$ 分别在 $\hat{\mathbf{x}}$ 和 $\hat{\mathbf{y}}$ 上投影。

$$\begin{cases} x = (u\hat{\mathbf{u}} + v\hat{\mathbf{v}}) \cdot \hat{\mathbf{x}} \\ y = (u\hat{\mathbf{u}} + v\hat{\mathbf{v}}) \cdot \hat{\mathbf{y}} \end{cases} \quad (5)$$

显然，该变换的逆变换为

$$\begin{cases} u = (x\hat{\mathbf{x}} + y\hat{\mathbf{y}}) \cdot \hat{\mathbf{u}} \\ v = (x\hat{\mathbf{x}} + y\hat{\mathbf{y}}) \cdot \hat{\mathbf{v}} \end{cases} \quad (6)$$

该变换所对应的矩阵就是 \mathbf{M} 的逆矩阵

$$\mathbf{M}^{-1} = \begin{pmatrix} \hat{\mathbf{x}} \cdot \hat{\mathbf{u}} & \hat{\mathbf{y}} \cdot \hat{\mathbf{u}} \\ \hat{\mathbf{x}} \cdot \hat{\mathbf{v}} & \hat{\mathbf{y}} \cdot \hat{\mathbf{v}} \end{pmatrix} \quad (7)$$

对比上式和式 3 可以发现 \mathbf{M} 与 \mathbf{M}^{-1} 互为转置矩阵。证毕。

平面旋转矩阵

预备知识 单位正交阵^[109]

平面旋转变换^[102] 属于线性变换，可以用矩阵 \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，更能帮助理解和记忆。由矩阵乘法的分配律，

$$\mathbf{R}_2(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1\mathbf{R}_2\mathbf{v}_1 + c_2\mathbf{R}_2\mathbf{v}_2 \quad (1)$$

已知单位矢量 $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ 逆时针旋转 θ 为

$$\mathbf{R}_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \quad \mathbf{R}_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix} \quad (2)$$

而任何矢量 $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ 都可以表示成 $\hat{\mathbf{x}}$ 和 $\hat{\mathbf{y}}$ 的线性组合 $x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ，所以

$$\begin{aligned} \mathbf{R}_2 \left[x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right] &= x_1 \mathbf{R}_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \mathbf{R}_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= x_1 \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} + x_2 \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix} \\ &= \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \end{aligned} \quad (3)$$

所以旋转矩阵为

$$\mathbf{R}_2 = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (4)$$

这与平面旋转变换^[102] 得出的结果一致。

逆矩阵

我们既可以使用平面旋转变换^[102] 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。但最方便的是，由于 \mathbf{R}_2 是一个单位正交阵^[109]，我们只需要把矩阵转置即可得到逆矩阵。

$$\mathbf{R}_2^{-1} = \mathbf{R}_2^T = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \quad (5)$$

空间旋转矩阵

预备知识 平面旋转矩阵^[110]

类比平面旋转矩阵^[110]，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。如果直接从转轴和转动角度来定义该矩阵，矩阵比较复杂，这里从略。

若已经知道空间直角坐标系中三个单位正交矢量

$$\hat{\mathbf{x}} = (1, 0, 0)^T \quad \hat{\mathbf{y}} = (0, 1, 0)^T \quad \hat{\mathbf{z}} = (0, 0, 1)^T \quad (1)$$

经过三维旋转矩阵变换以后变为

$$(a_{11}, a_{21}, a_{31})^T \quad (a_{12}, a_{22}, a_{32})^T \quad (a_{13}, a_{23}, a_{33})^T \quad (2)$$

类比平面旋转矩阵^[110]

$$\mathbf{R}_3 = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad (3)$$

绕轴旋转矩阵

预备知识 空间旋转矩阵^[112]，绕轴旋转的线速度

结论

直角坐标系中，某点 $\mathbf{r} = (x, y, z)^T$ 以单位矢量 $\hat{\mathbf{A}} = (A_x, A_y, A_z)^T$ 为轴按右手定则转动 θ 角的得到的点 $\mathbf{r}' = (x', y', z')^T$ 可用矩阵乘法计算

$$\mathbf{r}' = \mathbf{R}_\theta \mathbf{r} \quad (1)$$

其中 \mathbf{R}_θ 为绕轴旋转矩阵

$$\mathbf{R}_\theta = \begin{pmatrix} aA_x^2 + c & aA_xA_y - sA_z & aA_xA_z + sA_y \\ aA_yA_x + sA_z & aA_y^2 + c & aA_yA_z - sA_x \\ aA_zA_x - sA_y & aA_zA_y + sA_x & aA_z^2 + c \end{pmatrix} \quad (2)$$

其中

$$c = \cos \theta \quad s = \sin \theta \quad a = 1 - \cos \theta \quad (3)$$

推导

推导的思路是用 $\hat{\mathbf{A}}$, \mathbf{r} 和 θ 三个以质量经过数乘, 点乘^[91] 和叉乘^[96] 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^[110] 的思路不一样.

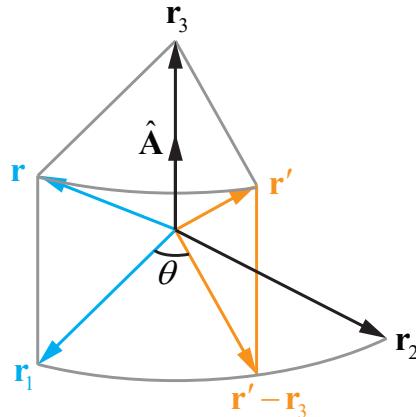


图 1: 绕轴旋转矩阵的推导

如图, \mathbf{r} 绕单位矢量 $\hat{\mathbf{A}}$ 旋转后得到 \mathbf{r}' . \mathbf{r} 在 $\hat{\mathbf{A}}$ 方向的分量为

$$\mathbf{r}_3 = (\hat{\mathbf{A}} \cdot \mathbf{r})\hat{\mathbf{A}} \quad (4)$$

在与 $\hat{\mathbf{A}}$ 垂直方向的分量为

$$\mathbf{r}_1 = \mathbf{r} - \mathbf{r}_3 \quad (5)$$

为了构成一组正交基底, 令

$$\mathbf{r}_2 = \hat{\mathbf{A}} \times \mathbf{r}_1 \quad (6)$$

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 绕旋转 θ 角后的结果

$$\mathbf{r}' - \mathbf{r}_3 = \mathbf{r}_1 \cos \theta + \mathbf{r}_2 \sin \theta \quad (7)$$

即

$$\mathbf{r}' = \mathbf{r}_1 \cos \theta + \mathbf{r}_2 \sin \theta + \mathbf{r}_3 \quad (8)$$

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^[105].

由旋转矩阵推导出匀速圆周运动的线速度

虽然这个公式有更简单的几何方法（见绕轴旋转的线速度），但是这种方法更偏数学一些，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 $\theta = \omega t \rightarrow 0$ ，此时有 $\sin \theta \rightarrow \theta$ 和 $\cos \theta \rightarrow 1$. 旋转矩阵变为

$$\mathbf{R}_\theta = \begin{pmatrix} 1 & -A_z \theta & A_y \theta \\ A_z \theta & 1 & -A_x \theta \\ -A_y \theta & A_x \theta & 1 \end{pmatrix} \quad (9)$$

下面 \mathbf{R}_θ 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}

$$\begin{aligned} \mathbf{s} &= \mathbf{v}t \\ &= \begin{pmatrix} 1 & -A_z \theta & A_y \theta \\ A_z \theta & 1 & -A_x \theta \\ -A_y \theta & A_x \theta & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ &= \theta \begin{pmatrix} 0 & -A_z & A_y \\ A_z & 0 & -A_x \\ -A_y & A_x & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ &= \theta \hat{\mathbf{A}} \times \mathbf{r} = (\boldsymbol{\omega} t) \times \mathbf{r} \end{aligned} \quad (10)$$

两边除以 t ，得 $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$ 这与绕轴转动的线速度中得出的结论一致.

行列式

行列式是线性代数中的一个重要工具，主要用于判断方阵中的所有列向量的线性无关⁹。行列式运算的结果是一个数，若结果不为零，则线性无关，为零则线性相关。物理中经常出现的是二阶和三阶行列式，高阶行列式的计算较为复杂（见词条最后），不必记忆其算法，可通过数学软件¹⁰计算。

二阶行列式的定义

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \quad (1)$$

三阶行列式的定义

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{cases} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{cases} \quad (2)$$

几何理解

N 阶行列式是 N 维空间中平行体的体积，平行体由矩阵的列矢量（或行矢量）定义。例如，二阶行列式代表一个平行四边形的面积（二维体积），平行四边形的四个顶点坐标分别为 $(0, 0)$, (a_{11}, a_{21}) , (a_{12}, a_{22}) 和 $(a_{11} + a_{12}, a_{21} + a_{22})$ 。用几何理解可以很形象地解释下面的性质 1, 3, 4.

行列式的性质

以下性质中，把“列”换成“行”同样成立，这是因为性质 2.

1. 若矩阵的列矢量线性相关，行列式为零，否则不为零。

⁹充分必要条件是所有行向量也线性无关

¹⁰详见 Matlab, Mathematica 和 Wolfram Alpha 的计算方法。

2. 矩阵转置后行列式的值不变.
3. 矩阵的任意一列乘以常数, 行列式的值也要乘以该常数.
4. 把矩阵的第 i 列叠加上“第 j 列乘任意常数”, 行列式的值不变.

以上性质证明略¹¹.

高阶行列式的定义

N 阶行列式 (N 为正整数¹²) 共有 $N!$ 项, 每一项都是 N 个矩阵元的乘积. 这 N 个矩阵元的行数和列数各不相同, 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 ± 号)

$$\prod_{i=1}^N a_{i,P_n(i)} = a_{1,P_n(1)} \cdot a_{2,P_n(2)} \cdots \quad (3)$$

其中列标 $P_n(i)$ 是数列 $1, 2, \dots, N$ 置换 (用某种顺序排列) 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 ± 号. 这由 P_n 的 逆序数 决定, 若逆序数为偶数, 则前面加正号, 奇数则加负号. 逆序数被定义为

$$\sum_{i=2}^N \text{满足 } P_n(i) < P_n(j) (j < i) \text{ 的个数} \quad (4)$$

若根据 P_n 对应的符号定义数列 S_n (取值 1 或 -1), 则 N 阶行列式的公式为

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = \sum_{n=1}^{N!} S_n \prod_{i=1}^N a_{i,P_n(i)} \quad (5)$$

¹¹见同济大学的《线性代数》

¹²一阶行列式定义为 $|a_{11}| = a_{11}$, 虽然几乎从不被使用

第四章

多元微积分

偏导数

预备知识 导数^[34]

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

$$\frac{\partial z}{\partial x} \quad \frac{\partial f}{\partial x} \quad f_x \quad \left(\frac{\partial f}{\partial x} \right)_y \quad (1)$$

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

$$\frac{\partial f}{\partial x} = 2x + 2y \quad \frac{\partial f}{\partial y} = 4y + 2x \quad (2)$$

例 2

对于函数 $z = \sin(y \cos x) + \cos^2 x$

$$\frac{\partial z}{\partial x} = -y \cos(y \cos x) \sin x - 2 \cos x \sin x = -y \cos(y \cos x) - \sin 2x \quad (3)$$

$$\frac{\partial z}{\partial y} = \cos(y \cos x) \cos x \quad (4)$$

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

$$\Delta f \approx \frac{\partial f}{\partial x} \Delta x \quad (5)$$

写成微分关系就是

$$df = \frac{\partial f}{\partial x} dx \quad (y \text{ 不变}) \quad (6)$$

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

$$\frac{\partial^2 f}{\partial x^2} \quad \frac{\partial^2 f}{\partial x \partial y} \quad \frac{\partial^2 f}{\partial y \partial x} \quad \frac{\partial^2 f}{\partial y^2} \quad (7)$$

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。

二元函数的极值

预备知识 导数与函数极值^[44]，偏导数^[118]，方向导数^[130]

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^[130]，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

$$\frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 > 0 \quad (1)$$

则驻点是极值点。如果 $\partial^2 f / \partial x^2$ 和 $\partial^2 f / \partial y^2$ 都大于零¹，则极值为极小值，若都小于零，则极值为极大值。

¹根据式 1，只需验证 $\partial^2 f / \partial x^2$ 或 $\partial^2 f / \partial y^2$ 中的任意一个大于零，另外一个就必定大于零。

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。令某方向为 $\mathbf{n} = \mathbf{x} \cos \theta + \mathbf{y} \sin \theta$ ，由式 9^[130] 得该方向的方向导数为

$$\left(\cos \theta \frac{\partial}{\partial x} + \sin \theta \frac{\partial}{\partial y} \right) f \quad (2)$$

再次求方向导数得二阶方向导数为

$$\left(\cos \theta \frac{\partial}{\partial x} + \sin \theta \frac{\partial}{\partial y} \right)^2 f = \frac{\partial^2 f}{\partial x^2} \cos^2 \theta + 2 \frac{\partial^2 f}{\partial x \partial y} \sin \theta \cos \theta + \frac{\partial^2 f}{\partial y^2} \sin^2 \theta \quad (3)$$

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

$$\frac{\partial^2 f}{\partial y^2} \tan^2 \theta + 2 \frac{\partial^2 f}{\partial x \partial y} \tan \theta + \frac{\partial^2 f}{\partial x^2} > 0 \quad (4)$$

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。

最小二乘法

预备知识 二元函数的极值^[119]

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 ($i = 1 \dots N$) 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。

$$S_2 = \sum_{i=1}^N (ax_i + b - y_i)^2 \quad (1)$$

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.

$$\begin{cases} \left(\sum_i x_i^2 \right) a + \left(\sum_i x_i \right) b = \sum_i x_i y_i \\ \left(\sum_i x_i \right) a + N b = \sum_i y_i \end{cases} \quad (2)$$

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

$$N \sum_i x_i^2 - \left(\sum_i x_i \right)^2 = N \sum_i (x_i - \bar{x})^2 > 0 \quad (3)$$

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 $c_2 x^2 + c_1 x + c_0$ 为例, 方差为

$$S_2 = \sum_{i=1}^N (c_2 x_i^2 + c_1 x_i + c_0 - y_i)^2 \quad (4)$$

分别令方差对 c_2, c_1, c_0 的偏导为零, 得线性方程组

$$\begin{cases} \left(\sum_i x_i^4 \right) c_2 + \left(\sum_i x_i^3 \right) c_1 + \left(\sum_i x_i^2 \right) c_0 = \sum_i y_i x_i^2 \\ \left(\sum_i x_i^3 \right) c_2 + \left(\sum_i x_i^2 \right) c_1 + \left(\sum_i x_i \right) c_0 = \sum_i y_i x_i \\ \left(\sum_i x_i^2 \right) c_2 + \left(\sum_i x_i \right) c_1 + N c_0 = \sum_i y_i \end{cases} \quad (5)$$

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

$$S_2 = \sum_i (c_1 \cos x_i + c_2 \sin x_i + c_3 - y_i)^2 \quad (6)$$

分别令方差对 c_1, c_2, c_3 的偏导为零，得线性方程组

$$\begin{cases} \left(\sum_i \cos^2 x_i \right) c_1 + \left(\sum_i \sin x_i \cos x_i \right) c_2 + \left(\sum_i \cos x_i \right) c_3 = \sum_i y_i \cos x_i \\ \left(\sum_i \sin x_i \cos x_i \right) c_1 + \left(\sum_i \sin^2 x_i \right) c_2 + \left(\sum_i \sin x_i \right) c_3 = \sum_i y_i \sin x_i \\ \left(\sum_i \cos x_i \right) c_1 + \left(\sum_i \sin x_i \right) c_2 + N c_3 = \sum_i y_i \end{cases} \quad (7)$$

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^[118]

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²，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

$$z = c_0 + c_x(x - x_0) + c_y(y - y_0) \quad (1)$$

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

$$c_x = \frac{\partial f}{\partial x} \quad c_y = \frac{\partial f}{\partial y} \quad (2)$$

²光滑的数学定义是，各阶偏导数在区域内连续，即没有断点。

令坐标增量为 $\Delta x \equiv x - x_0$, $\Delta y \equiv y - y_0$, $\Delta z \equiv z - z_0$, 则平面方程变为

$$\Delta z = \frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial y} \Delta y \quad (3)$$

令增量为无穷小, 即

$$dz = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy \quad (4)$$

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 (dx, dy) , 函数的增量等于只向 x 方向移动 dx 的增量加上只向 y 方向移动 dy 的增量. 类似地, N 元函数的全微分关系为

$$dz = \sum_{i=1}^N \frac{\partial f}{\partial x_i} dx_i \quad (5)$$

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^[47] $\Delta y \approx df/dx \cdot \Delta x$, 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

$$\begin{aligned} \Delta z &= f(x_1 + \Delta x_1 \dots x_N + \Delta x_N) - f(x_1 \dots x_N) \\ &\approx \frac{\partial f}{\partial x_1} \Delta x_1 + \dots + \frac{\partial f}{\partial x_N} \Delta x_N \end{aligned} \quad (6)$$

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

$$\sigma_V \approx \frac{\partial V}{\partial a} \sigma_a + \frac{\partial V}{\partial b} \sigma_b + \frac{\partial V}{\partial c} \sigma_c = bc\sigma_a + ac\sigma_b + ab\sigma_c \quad (7)$$

相对不确定度为

$$\frac{\sigma_V}{V} \approx \frac{\sigma_a}{a} + \frac{\sigma_b}{b} + \frac{\sigma_c}{c} \quad (8)$$

复合函数的偏导 链式法则

预备知识 复合函数的偏导 链式法则^[124]

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

$$z(x, y) = f[u(x, y), v(x, y)] \quad (1)$$

那如何求 z 对 x 和 y 的偏微分呢? 我们先来看全微分关系. 首先

$$dz = \frac{\partial f}{\partial u} du + \frac{\partial f}{\partial v} dv \quad (2)$$

而 u 和 v 的微小变化又都是由 x 和 y 的微小变化引起的

$$du = \frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy \quad dv = \frac{\partial v}{\partial x} dx + \frac{\partial v}{\partial y} dy \quad (3)$$

所以

$$\begin{aligned} dz &= \frac{\partial f}{\partial u} \left(\frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy \right) + \frac{\partial f}{\partial v} \left(\frac{\partial v}{\partial x} dx + \frac{\partial v}{\partial y} dy \right) \\ &= \left(\frac{\partial f}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial f}{\partial v} \frac{\partial v}{\partial x} \right) dx + \left(\frac{\partial f}{\partial u} \frac{\partial u}{\partial y} + \frac{\partial f}{\partial v} \frac{\partial v}{\partial y} \right) dy \end{aligned} \quad (4)$$

这就是 z 关于 x 和 y 的全微分关系. 根据定义

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial f}{\partial v} \frac{\partial v}{\partial x} \quad (5)$$

$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial u} \frac{\partial u}{\partial y} + \frac{\partial f}{\partial v} \frac{\partial v}{\partial y} \quad (6)$$

这也叫偏导的链式法则.

通用函数名

物理中常常会出现一种容易混淆的情况³, 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

³ “通用函数名”是我的叫法

例 1

在二维直角坐标系中，定义势能函数为

$$V = f(x, y) = x^2 + y^2 + 2x \quad (7)$$

而若用极坐标描述该势能，则函数变为

$$V = g(r, \theta) = f(r \cos \theta, r \sin \theta) = r^2 + 2r \cos \theta \quad (8)$$

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 7 和 $V(r, \theta)$ 表示式 8。这样后者就有可能被误解为

$$V(r, \theta) = r^2 + \theta^2 + 2r \quad (\text{错}) \quad (9)$$

这就需要从语境中判断是否使用了通用函数名。

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， $\partial V / \partial r$ 默认使用 $V(r, \theta)$ 求偏导。一种更复杂的情况如 $(\partial V / \partial x)_\theta$ 。按照定义⁴，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 7 得

$$V(x, \theta) = x^2(1 + \tan^2 \theta) + 2x \quad (10)$$

现在再对 x 求偏导即可（略）。

全导数

预备知识 全微分^[122]

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。

⁴ 见偏导数^[118] 中的式 1

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^[122]

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial t} dt \quad (1)$$

而根据 x, y 与 t 的微分关系

$$dx = \frac{dx}{dt} dt \quad dy = \frac{dy}{dt} dt \quad (2)$$

代入上式得

$$df = \left(\frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial t} \right) dt \quad (3)$$

而若把 f 看做 t 的一元函数，又应该有全微分关系

$$df = \frac{df}{dt} dt \quad (4)$$

对比以上两式可得 f 关于 t 的全导数为

$$\frac{df}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial t} \quad (5)$$

矢量的导数 求导法则

预备知识 矢量的加减，矢量的数乘，矢量函数，导数^[34]

矢量的导数

若矢量 \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（括号可选择省略）

$$\frac{d\mathbf{v}}{dt} \quad \frac{d}{dt} \mathbf{v} \quad \dot{\mathbf{v}} \quad (1)$$

其定义为（类比导数^[34] 中导数的代数定义）

$$\frac{d\mathbf{v}}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{v}(t + \Delta t) - \mathbf{v}(t)}{\Delta t} \quad (2)$$

唯一与标量函数的导数不同的是，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 $1/\Delta t$ ，结果也是矢量。所以 $d\mathbf{v}/dt$ 也是一个关于 t 的矢量函数。

直角坐标系中对单变量矢量函数求导就是对矢量的各个分量分别求导（见下文“求导法则”）。

应用举例 速度和加速度(矢量)^[156], 匀速圆周运动的速度^[157] 和加速度^[159](几何法)

矢量的偏导数

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.

$$\frac{\partial \mathbf{v}}{\partial x_i} = \lim_{\Delta x_i \rightarrow 0} \frac{\mathbf{v}(x_1 \dots x_i + \Delta x_i \dots x_N) - \mathbf{v}(x_1 \dots x_i \dots x_N)}{\Delta x_i} \quad (3)$$

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”).

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)⁵

$$\frac{d}{dt}[c_1 \mathbf{v}_1(t) + c_2 \mathbf{v}_2(t) + \dots] = c_1 \frac{d\mathbf{v}_1}{dt} + c_2 \frac{d\mathbf{v}_2}{dt} \dots \quad (4)$$

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.

$$\frac{d\mathbf{v}}{dt} = \frac{d}{dt}[v_x(t)\hat{\mathbf{x}}] + \frac{d}{dt}[v_y(t)\hat{\mathbf{y}}] + \frac{d}{dt}[v_z(t)\hat{\mathbf{z}}] = \dot{v}_x(t)\hat{\mathbf{x}} + \dot{v}_y(t)\hat{\mathbf{y}} + \dot{v}_z(t)\hat{\mathbf{z}} \quad (5)$$

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^[136]”.

应用举例 匀速圆周运动的速度^[158] 和加速度^[160] (求导法)

矢量数乘, 点乘或叉乘的求导在形式上都与标量函数的情况类似.

$$\frac{d}{dt}[f(t)\mathbf{v}(t)] = \frac{df}{dt}\mathbf{v} + f \frac{d\mathbf{v}}{dt} \quad (6)$$

$$\frac{d}{dt}[\mathbf{u}(t) \cdot \mathbf{v}(t)] = \frac{d\mathbf{u}}{dt} \cdot \mathbf{v} + \mathbf{u} \cdot \frac{d\mathbf{v}}{dt} \quad (7)$$

⁵以下法则虽然以导数为例, 但对偏导也同样适用.

$$\frac{d}{dt}[\mathbf{u}(t) \times \mathbf{v}(t)] = \frac{d\mathbf{u}}{dt} \times \mathbf{v} + \mathbf{u} \times \frac{d\mathbf{v}}{dt} \quad (8)$$

由定义出发, 不难证明以上三式, 这里以式 7 为例进行证明. 根据点乘定义以及标量函数的求导法则^[37] 有

$$\begin{aligned} \frac{d}{dt}(\mathbf{u} \cdot \mathbf{v}) &= \frac{d}{dt}(u_x v_x + u_y v_y + u_z v_z) \\ &= \left(\frac{du_x}{dt} v_x + u_x \frac{dv_x}{dt} \right) + \left(\frac{du_y}{dt} v_y + u_y \frac{dv_y}{dt} \right) + \left(\frac{du_z}{dt} v_z + u_z \frac{dv_z}{dt} \right) \\ &= \left(\frac{du_x}{dt} v_x + \frac{du_y}{dt} v_y + \frac{du_z}{dt} v_z \right) + \left(u_x \frac{dv_x}{dt} + u_y \frac{dv_y}{dt} + u_z \frac{dv_z}{dt} \right) \\ &= \frac{d\mathbf{u}}{dt} \cdot \mathbf{v} + \mathbf{u} \cdot \frac{d\mathbf{v}}{dt} \end{aligned} \quad (9)$$

应用举例 动量定理^[195], 角动量定理 (单个质点)^[171]

矢量的高阶导数和偏导

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导数.

与标量函数的偏导^[118] 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.

一元矢量函数的积分

预备知识 矢量的导数^[126]

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

$$\mathbf{F}'(t) = \mathbf{f}(t) \quad (1)$$

$$F'_x(t) = f_x(t) \quad F'_y(t) = f_y(t) \quad F'_z(t) = f_z(t) \quad (2)$$

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 \mathbf{C} 。

$$\begin{aligned} \int \mathbf{f}(t) dt &= \hat{\mathbf{x}} \int f_x(t) dt + \hat{\mathbf{y}} \int f_y(t) dt + \hat{\mathbf{z}} \int f_z(t) dt \\ &= [F_x(t) + C_x] \hat{\mathbf{x}} + [F_y(t) + C_y] \hat{\mathbf{y}} + [F_z(t) + C_z] \hat{\mathbf{z}} \\ &= \mathbf{F}(t) + \mathbf{C} \end{aligned} \quad (3)$$

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^[65] 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 $N \rightarrow \infty$ ）时，如果以下极限存在，得到的矢量就是定积分的结果。

$$\int_{t_1}^{t_2} \mathbf{f}(t) dt = \lim_{\Delta t_i \rightarrow 0} \sum_{i=0}^N \mathbf{f}(t_i) \Delta t_i \quad (4)$$

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。

$$\begin{aligned} \int_{t_1}^{t_2} \mathbf{f}(t) dt &= \hat{\mathbf{x}} \lim_{\Delta t_i \rightarrow 0} \sum_{i=0}^N f_x(t_i) \Delta t_i + \hat{\mathbf{y}} \lim_{\Delta t_i \rightarrow 0} \sum_{i=0}^N f_y(t_i) \Delta t_i + \hat{\mathbf{z}} \lim_{\Delta t_i \rightarrow 0} \sum_{i=0}^N f_z(t_i) \Delta t_i \\ &= \hat{\mathbf{x}} \int_{t_1}^{t_2} f_x(t) dt + \hat{\mathbf{y}} \int_{t_1}^{t_2} f_y(t) dt + \hat{\mathbf{z}} \int_{t_1}^{t_2} f_z(t) dt \end{aligned} \quad (5)$$

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^[69], $\mathbf{f}(t)$ 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

$$\begin{aligned} \int_{t_1}^{t_2} \mathbf{f}(t) dt &= \hat{\mathbf{x}}[F_x(t_2) - F_x(t_1)] + \hat{\mathbf{y}}[F_y(t_2) - F_y(t_1)] + \hat{\mathbf{z}}[F_z(t_2) - F_z(t_1)] \\ &= \mathbf{F}(t_2) - \mathbf{F}(t_1) \end{aligned} \quad (6)$$

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

由于质点的速度一时间函数 $\mathbf{v}(t)$ 是位移一时间函数 $\mathbf{r}(t)$ 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

$$\mathbf{r}(t) - \mathbf{r}(t_0) = \int_{t_1}^{t_2} \mathbf{v}(t) dt \quad (7)$$

即

$$\mathbf{r}(t) = \mathbf{r}(t_0) + \int_{t_1}^{t_2} \mathbf{v}(t) dt \quad (8)$$

这是一维情况^{式 5^[154] 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

$$\mathbf{v}(t) = \mathbf{v}(t_0) + \int_{t_1}^{t_2} \mathbf{a}(t) dt \quad (9)$$

应用举例 匀加速运动^[160]

方向导数

预备知识 全微分^[122], 点乘^[91], 正交归一基^[94]

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} 是位矢^[153], 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。

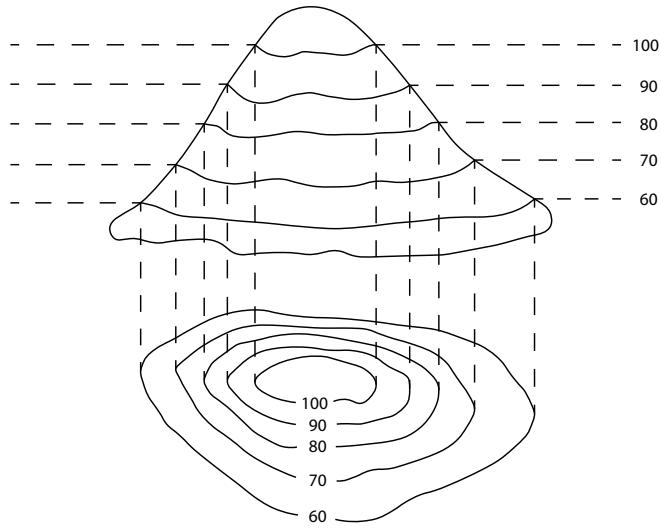


图 1: 等高线

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

$$\frac{df}{dn} \equiv \lim_{\Delta s \rightarrow 0} \frac{f(\mathbf{r} + \hat{\mathbf{n}}\Delta s) - f(\mathbf{r})}{\Delta s} = \frac{df}{ds} \quad (1)$$

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。

由“全微分^[122]”中的结论

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy \quad (2)$$

而现在我们往 $\hat{\mathbf{n}} = (n_x, n_y)$ 方向移动 ds ，所以

$$dx = n_x ds \quad dy = n_y ds \quad (3)$$

代入上式，得

$$df = \left(\frac{\partial f}{\partial x} n_x + \frac{\partial f}{\partial y} n_y \right) ds \quad (4)$$

根据导数与微分的关系（也可以通俗地说“两边同除 ds ”），就得到方向导数

$$\frac{\partial f}{\partial n} = \frac{df}{ds} = \frac{\partial f}{\partial x} n_x + \frac{\partial f}{\partial y} n_y \quad (5)$$

如果使用平面的正交归一基^[94] $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ 写成矢量点乘^[91] 的形式，就是

$$\frac{df}{dn} = \left(\frac{\partial f}{\partial x} \hat{\mathbf{x}} + \frac{\partial f}{\partial y} \hat{\mathbf{y}} \right) \cdot \hat{\mathbf{n}} \quad (6)$$

定义二维直角坐标系中的 **Del 算符** 为

$$\nabla = \hat{\mathbf{x}} \frac{\partial}{\partial x} + \hat{\mathbf{y}} \frac{\partial}{\partial y} \quad (7)$$

其作用在函数上表示

$$\nabla f = \frac{\partial f}{\partial x} \hat{\mathbf{x}} + \frac{\partial f}{\partial y} \hat{\mathbf{y}} \quad (8)$$

则方向导数可以写成相当简洁的形式，即

$$\frac{\partial f}{\partial n} = \nabla f \cdot \hat{\mathbf{n}} \quad (9)$$

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 $\hat{\mathbf{n}} = (n_{x1}, n_{x2} \dots n_{xN})$ 的方向上的微分关系为

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 \dots = \left(\frac{\partial f}{\partial x_1} n_{x1} + \frac{\partial f}{\partial x_2} n_{x2} \dots \right) ds \quad (10)$$

方向导数为

$$\frac{\partial f}{\partial n} = \frac{df}{ds} = \left(\frac{\partial f}{\partial x_1} \hat{\mathbf{x}}_1 + \frac{\partial f}{\partial x_2} \hat{\mathbf{x}}_2 \dots \frac{\partial f}{\partial x_N} \hat{\mathbf{x}}_N \right) \cdot \hat{\mathbf{n}} = \nabla f \cdot \hat{\mathbf{n}} \quad (11)$$

形式与式 9 相同。这里定义了 N 维直角坐标系的 **Del 算符** 为

$$\nabla = \hat{\mathbf{x}}_1 \frac{\partial}{\partial x_1} + \hat{\mathbf{x}}_2 \frac{\partial}{\partial x_2} \dots \hat{\mathbf{x}}_n \frac{\partial}{\partial x_N} \quad (12)$$

重积分

预备知识 定积分^[65]

面积分

面积分也叫二重积分，可以看做一元函数定积分^[65]的一种拓展。从几何上来理解，如果后者是计算一定区间内被积函数曲线与坐标轴之间的面积，那么前者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。

例 1 曲面下的体积

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 。

与定积分的思想一样，我们可以把三角形区域划分成许多更小的区域，每一个区域都对应一个被曲面截取的小柱体，由于我们使用直角坐标系，我们不妨沿 x 和 y 方向划分出许多小矩形，用 (i, j) 给它们编号，每个矩形的面积分别为 $\Delta x_i \Delta y_j$ ，小柱体的体积近似用底乘以高计算得 $\Delta V_{ij} = f(x_i, y_i) \Delta x_i \Delta y_j$ ，其中 x_i, y_i 分别为区间 $\Delta x_i, \Delta y_i$ 内任意一点。所以总体积就可以用极限表示为

$$V = \lim_{\substack{\Delta x_i \rightarrow 0 \\ \Delta y_j \rightarrow 0}} \sum_{i,j} f(x_i, y_j) \Delta x_i \Delta y_j \quad (1)$$

这是一个**双重极限**，类比定积分的定义，用面积分简写上式记为

$$V = \iint_{\mathcal{S}} f(x, y) \, dx \, dy \quad (2)$$

其中 \mathcal{S} 代表以上定义的三角形积分区域。

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

$$S(x) = \int_{y_1(x)}^{y_2(x)} f(x, y) \, dy = \frac{1}{3}[y_2^3(x) - y_1^3(x)] + x^2[y_2(x) - y_1(x)] = \frac{8}{3}x^3 \quad (3)$$

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

$$V = \lim_{\Delta x_i \rightarrow 0} \sum_i S(x_i) \Delta x_i = \int_0^1 S(x) \, dx = \frac{2}{3}(1^4 - 0^4) = \frac{2}{3} \quad (4)$$

要从定义上证明这种方法可行，我们可以把式1写成

$$V = \lim_{\Delta x_i \rightarrow 0} \sum_i \left[\lim_{\Delta y_j \rightarrow 0} \sum_j f(x_i, y_j) \Delta y_j \right] \Delta x_i \quad (5)$$

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

$$V = \int_0^1 \left[\int_{y_1(x)}^{y_2(x)} f(x, y) dy \right] dx \quad (6)$$

另一种等效的书写方式是，把 $\int dx, \int dy$ 看成像 $d/dx, d/dy$ 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

$$V = \int_0^1 dx \int_{y_1(x)}^{y_2(x)} dy f(x, y) \quad (7)$$

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 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

$$V = \lim_{\substack{\Delta x_i \rightarrow 0 \\ \Delta y_j \rightarrow 0 \\ \Delta z_k \rightarrow 0}} \sum_{i,j,k} f(x_i, y_j, z_k) \Delta x_i \Delta y_j \Delta z_k \quad (8)$$

用三重积分可以把这个三重极限表示为

$$V = \iiint_V f(x, y, z) dx dy dz \quad (9)$$

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 $z_1(x, y) = 0$

$$\begin{aligned}\sigma(x, y) &= \int_0^{1-x-y} (x + y + z) dz \\ &= (x + y)[z_2(x, y) - z_1(x, y)] + \frac{1}{2}[z_2^2(x, y) - z_1^2(x, y)] \\ &= (x + y)(1 - x - y) + \frac{1}{2}(1 - x - y)^2 \\ &= \frac{1}{2}(1 - x^2 - y^2) - xy\end{aligned}\tag{10}$$

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 $y_1(x) = 0$

$$\begin{aligned}\lambda(x) &= \int_0^{1-x} \sigma(x, y) dy \\ &= \frac{1}{2}(1 - x^2)(1 - x) - \frac{1}{6}(1 - x)^3 - \frac{1}{2}x(1 - x)^2 \\ &= \frac{x^3}{6} - \frac{x}{2} + \frac{1}{3}\end{aligned}\tag{11}$$

最后把线密度从 1 和 0 积分得到总质量

$$M = \int_0^1 \lambda(x) dx = \left[\frac{1}{24}x^4 - \frac{1}{4}x^2 + \frac{1}{3}x \right]_0^1 = \frac{1}{8}\tag{12}$$

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

$$M = \lim_{\Delta x_i \rightarrow 0} \sum_i \left[\lim_{\Delta y_j \rightarrow 0} \sum_j \left(\lim_{\Delta z_k \rightarrow 0} \sum_k f(x_i, y_j, z_k) \Delta z_k \right) \Delta y_j \right] \Delta x_i\tag{13}$$

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

$$M = \int_0^1 dx \int_0^{1-x} dy \int_0^{1-x-y} dz f(x, y, z)\tag{14}$$

预备知识 球坐标系的定义^[11], 矢量的求导法则^[126]

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.

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

$$\begin{cases} v_x(x, y, z) = \mathbf{v} \cdot \hat{\mathbf{x}} \\ v_y(x, y, z) = \mathbf{v} \cdot \hat{\mathbf{y}} \\ v_z(x, y, z) = \mathbf{v} \cdot \hat{\mathbf{z}} \end{cases} \quad (1)$$

也可以作为单位正交基^[94] 的线性组合写成一个整体

$$\begin{aligned} \mathbf{v} &= (\mathbf{v} \cdot \hat{\mathbf{x}}) \hat{\mathbf{x}} + (\mathbf{v} \cdot \hat{\mathbf{y}}) \hat{\mathbf{y}} + (\mathbf{v} \cdot \hat{\mathbf{z}}) \hat{\mathbf{z}} \\ &= v_x(x, y, z) \hat{\mathbf{x}} + v_y(x, y, z) \hat{\mathbf{y}} + v_z(x, y, z) \hat{\mathbf{z}} \end{aligned} \quad (2)$$

在球坐标系^[11] 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

$$\mathbf{v} = v_r(r, \theta, \phi) \hat{\mathbf{r}} + v_\theta(r, \theta, \phi) \hat{\boldsymbol{\theta}} + v_\phi(r, \theta, \phi) \hat{\boldsymbol{\phi}} \quad (3)$$

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^[126] 进行.

应用举例 力场^[166]

拓展阅读 梯度^[139], 散度^[143], 旋度

极坐标中单位矢量的偏导

预备知识 极坐标系^[9], 矢量的偏导^[126]

与直角坐标系不同的是，极坐标系中的 $\hat{\mathbf{r}}$ 与 $\hat{\theta}$ 都是坐标的函数，即 $\hat{\mathbf{r}} = \hat{\mathbf{r}}(r, \theta)$, $\hat{\theta} = \hat{\theta}(r, \theta)$, 它们对坐标的偏导如下

$$\begin{cases} \frac{\partial \hat{\mathbf{r}}}{\partial r} = 0 \\ \frac{\partial \hat{\mathbf{r}}}{\partial \theta} = \hat{\theta} \end{cases} \quad \begin{cases} \frac{\partial \hat{\theta}}{\partial r} = 0 \\ \frac{\partial \hat{\theta}}{\partial \theta} = -\hat{\mathbf{r}} \end{cases} \quad (1)$$

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

$$\hat{\mathbf{r}} = \cos \theta \hat{\mathbf{x}} + \sin \theta \hat{\mathbf{y}} \quad (2)$$

$$\hat{\theta} = \cos(\theta + \pi/2) \hat{\mathbf{x}} + \sin(\theta + \pi/2) \hat{\mathbf{y}} = -\sin \theta \hat{\mathbf{x}} + \cos \theta \hat{\mathbf{y}} \quad (3)$$

所以

$$\begin{cases} \frac{\partial \hat{\mathbf{r}}}{\partial r} = 0 \\ \frac{\partial \hat{\mathbf{r}}}{\partial \theta} = -\sin \theta \hat{\mathbf{x}} + \cos \theta \hat{\mathbf{y}} = \hat{\theta} \end{cases} \quad \begin{cases} \frac{\partial \hat{\theta}}{\partial r} = 0 \\ \frac{\partial \hat{\theta}}{\partial \theta} = -\cos \theta \hat{\mathbf{x}} - \sin \theta \hat{\mathbf{y}} = -\hat{\mathbf{r}} \end{cases} \quad (4)$$

事实上，由于 $\hat{\mathbf{r}}$ 与 $\hat{\theta}$ 都只是 θ 的函数，也可以把偏导符号改成导数符号

$$\frac{d\hat{\theta}}{d\theta} = -\hat{\mathbf{r}} \quad \frac{d\hat{\mathbf{r}}}{d\theta} = \hat{\theta} \quad (5)$$

线积分

预备知识 功 功率^[164]， 定积分^[65]

在“功 功率^[164]”中，我们大致了解了线积分 $\int_{C_{ab}} \mathbf{F}(\mathbf{r}) \cdot d\mathbf{r}$ 的意义. 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可以表示为⁶ $x(t), y(t), z(t)$, 则曲线上任意一点都唯一对应一个 t 值. 则根据微分关系, 当 t 增加 dt 时, 曲线上的一小段微位移矢量 $d\mathbf{r} = (dx, dy, dz)$ 中

$$dx = x'(t) dt \quad dy = y'(t) dt \quad dz = z'(t) dt \quad (1)$$

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⁷ 则表示为关于 t 的单变量标量函数

$$F_{x_i}(t) = F_{x_i}[x(t), y(t), z(t)] \quad (i = 1, 2, 3) \quad (2)$$

下面将三维空间的线积分转换为三个一元定积分

$$\begin{aligned} \int \mathbf{F}(\mathbf{r}) \cdot d\mathbf{r} &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbf{F}(\mathbf{r}_i) \cdot \Delta \mathbf{r}_i \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n F_x(\mathbf{r}_i) \Delta x_i + \lim_{n \rightarrow \infty} \sum_{i=1}^n F_y(\mathbf{r}_i) \Delta y_i + \lim_{n \rightarrow \infty} \sum_{i=1}^n F_z(\mathbf{r}_i) \Delta z_i \\ &= \int F_x(\mathbf{r}) dx + \int F_y(\mathbf{r}) dy + \int F_z(\mathbf{r}) dz \end{aligned} \quad (3)$$

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

$$\int F_{x_i}(\mathbf{r}) dx_i = \int_a^b F_{x_i}[\mathbf{r}(t)] x'_i(t) dt \quad (i = 1, 2, 3) \quad (4)$$

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 $t, t \in [0, \pi]$.

$$\begin{cases} x(t) = a(1 - \cos t) \\ y(t) = a \sin t \end{cases} \quad \begin{cases} x'(t) = a \sin t \\ y'(t) = a \cos t \end{cases} \quad (5)$$

⁶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.

⁷为了书写简洁, 这里定义 $x_1 \equiv x, x_2 \equiv y, x_3 \equiv z$.

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

$$W = \int \mathbf{F} \cdot d\mathbf{r} = \int_0^\pi \alpha a(1 - \cos t) \cdot a \sin t dt + \int_0^\pi \alpha a \sin t \cdot a \cos t dt \quad (6)$$

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

$$\int_0^\pi a^2 \sin t dt = \frac{\alpha a^2}{2} \quad (7)$$

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 $x(t) = t, y(t) = 0, x \in [0, a]$. 代入式4 得做功为

$$W = \int_0^a \alpha x dx = \frac{\alpha a^2}{2} \quad (8)$$

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.

拓展阅读 梯度 梯度定理^[139]

梯度 梯度定理

预备知识 方向导数^[130]

在方向导数^[130] 中, 我们推出方向导数为

$$\frac{\partial f}{\partial n} = \nabla f \cdot \hat{\mathbf{n}} \quad (1)$$

其中 ∇f 就叫标量函数 f 的梯度⁸. 要注意当且仅当 Del 算符 ∇ 作用在标量函数 (即因变量是一个数而不是矢量) 上时, 可以称其为梯度算符. 这里的

⁸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.

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

$$\nabla f = \sum_{i=1}^N \frac{\partial f}{\partial x_i} \hat{\mathbf{x}}_i \quad (2)$$

其中所有的 $\hat{\mathbf{x}}_i$ 组成直角坐标系的正交归一基^[94], 现在来看全微分^[122]关系

$$df = \sum_{i=1}^N \frac{\partial f}{\partial x_i} dx_i \quad (3)$$

若定义微位移矢量为

$$d\mathbf{r} = \sum_{i=1}^N dx_i \hat{\mathbf{x}}_i \quad (4)$$

式 3 可用势函数的梯度和微位移矢量的点乘表示

$$df = \nabla f \cdot d\mathbf{r} \quad (5)$$

由点乘的几何定义^[91]可知, 从某点出发, 若微位移 $d\mathbf{r}$ 的大小不变, 那么当其方向与梯度方向相同时函数增量 df 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^[34]和微分^[47]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^[122]” 式 6) 记为矢量的形式

$$\Delta f \approx \nabla f \cdot \Delta \mathbf{r} \quad (6)$$

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 $d\mathbf{r} = dx \hat{\mathbf{x}} + dy \hat{\mathbf{y}}$, 由于 $(x + dx, y + dy)$ 仍然在等值线上, 函数增量 $df = 0$. 代入式 5 得

$$\nabla f \cdot d\mathbf{r} = 0 \quad (7)$$

即 $f(x, y)$ 的梯度与 $d\mathbf{r}$ 垂直. 所以 $\nabla f(x, y)$ 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 极坐标系^[9], 柱坐标系^[10], 球坐标系^[11]

我们先写出极坐标中函数 $f(r, \theta)$ 的全微分为

$$df = \frac{\partial f}{\partial r} dr + \frac{\partial f}{\partial \theta} d\theta \quad (8)$$

再写出极坐标中的微位移为

$$dr = dr \hat{\mathbf{r}} + r d\theta \hat{\theta} \quad (9)$$

所以为了满足梯度的定义式 5, 我们可以把式 8 第二项的 $\partial f / \partial \theta$ 除以 r , 第二个因子 $d\theta$ 乘以 r , 分别作为梯度和微位移的两个分量.

$$df = \frac{\partial f}{\partial r} dr + \frac{1}{r} \frac{\partial f}{\partial \theta} r d\theta \quad (10)$$

对比式 5, 式 9 和式 10 可以得出极坐标中的梯度算符为

$$\nabla = \hat{\mathbf{r}} \frac{\partial}{\partial r} + \hat{\theta} \frac{1}{r} \frac{\partial}{\partial \theta} \quad (11)$$

同理, 柱坐标中的微位移与函数 $f(r, \theta, z)$ 的全微分可以分别表示为

$$dr = dr \hat{\mathbf{r}} + r d\theta \hat{\theta} + dz \hat{\mathbf{z}} \quad (12)$$

$$df = \frac{\partial f}{\partial r} dr + \frac{1}{r} \frac{\partial f}{\partial \theta} r d\theta + \frac{\partial f}{\partial z} dz \quad (13)$$

所以柱坐标中的梯度算符为

$$\nabla = \hat{\mathbf{r}} \frac{\partial}{\partial r} + \hat{\theta} \frac{1}{r} \frac{\partial}{\partial \theta} + \hat{\mathbf{z}} \frac{\partial}{\partial z} \quad (14)$$

再来看球坐标, 球坐标中的微位移与 $f(r, \theta, \phi)$ 的全微分可以分别表示为

梯度定理

预备知识 线积分^[137], 牛顿—莱布尼兹公式^[69]

梯度定理: 一个标量函数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

$$\int_{\mathbf{r}_i}^{\mathbf{r}_f} \nabla f(\mathbf{r}) \cdot d\mathbf{r} = f(\mathbf{r}_f) - f(\mathbf{r}_i) \quad (15)$$

梯度定理可以看做是牛顿—莱布尼兹公式^[69]

$$\int_a^b f'(x) dx = f(b) - f(a) \quad (16)$$

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 15 变为式 16). 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 8 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

$$\int_{\mathbf{r}_{i0}}^{\mathbf{r}_{i1}} \nabla f(\mathbf{r}) \cdot d\mathbf{r} \approx \nabla f(\mathbf{r}_i) \cdot \Delta \mathbf{r}_i \quad (17)$$

再利用全微分近似 (式 6), 上式等于

$$\int_{\mathbf{r}_{i0}}^{\mathbf{r}_{i1}} \nabla f(\mathbf{r}) \cdot d\mathbf{r} \approx f(\mathbf{r}_{i1}) - f(\mathbf{r}_{i0}) \quad (18)$$

将所有小段的线积分求和得到总的线积分得 (注意 $\mathbf{r}_{i1} = \mathbf{r}_{(i+1)0}$)

$$\begin{aligned} \int_C \nabla f(\mathbf{r}) \cdot d\mathbf{r} &= \sum_{i=1}^n \int_{\mathbf{r}_{i0}}^{\mathbf{r}_{i1}} \nabla f(\mathbf{r}) \cdot d\mathbf{r} \\ &\approx \sum_{i=1}^n [f(\mathbf{r}_{i1}) - f(\mathbf{r}_{i0})] = f(\mathbf{r}_f) - f(\mathbf{r}_i) \end{aligned} \quad (19)$$

最后取极限 $n \rightarrow \infty$, 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 8，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 $f(\mathbf{r})$ 。

$$f(\mathbf{r}) = f(\mathbf{r}_0) + \int_{\mathbf{r}_0}^{\mathbf{r}} \nabla f(\mathbf{r}) \cdot d\mathbf{r} \quad (20)$$

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^[137]（分别把 x 和 y 作为线积分的参数）分别为

$$\int_{\mathbf{r}_0}^{\mathbf{r}_1} \nabla f(\mathbf{r}) \cdot d\mathbf{r} = \int_{x_0}^x f_x(x, y_0) dx + 0 = F_x(x, y_0) - F_x(x_0, y_0) \quad (21)$$

$$\int_{\mathbf{r}_1}^{\mathbf{r}} \nabla f(\mathbf{r}) \cdot d\mathbf{r} = \int_{y_0}^y f_y(x, y) dy + 0 = F_y(x, y) - F_y(x, y_0) \quad (22)$$

代回式 20 得势函数为

$$\begin{aligned} f(x, y) &= f(x_0, y_0) + F_x(x, y_0) - F_x(x_0, y_0) + F_y(x, y) - F_y(x, y_0) \\ &= F_y(x, y) - F_y(x, y_0) + F_x(x, y_0) + C \end{aligned} \quad (23)$$

其中 C 为待定常数。

应用举例 势能^[166]

散度 散度定理

预备知识 全微分^[122], 矢量场^[135], 重积分^[132], 面积分(曲面), 流密度,

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}

$$\Phi = \oint \mathbf{F} \cdot d\mathbf{a} \quad (1)$$

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 $\nabla \cdot$ 算符⁹记为 (下文将介绍)

$$\nabla \cdot \mathbf{F}(\mathbf{r}) \equiv \lim_{V \rightarrow 0} \frac{\Phi}{V} \quad (2)$$

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¹⁰, 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.

例 2 变速水流

在例 1 中, 流密度场随 x 坐标线性增加, $\mathbf{j} = (j_0 + \alpha x)\hat{\mathbf{x}}$ (α 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

$$\mathbf{F}(x, y, z) = F_x(x, y, z)\hat{\mathbf{x}} + F_y(x, y, z)\hat{\mathbf{y}} + F_z(x, y, z)\hat{\mathbf{z}} \quad (3)$$

⁹ 符号 ∇ 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.

¹⁰ 本书不作证明

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^[122] 得 (为简便书写, 以下的函数值和偏导都默认在 \mathbf{r} 处取值)

$$F_x(x+x', y+y', z+z') \approx F_x + \frac{\partial F_x}{\partial x}x' + \frac{\partial F_x}{\partial y}y' + \frac{\partial F_x}{\partial z}z' \quad (4)$$

由于只有 x 方向的场分量对 Φ_x 有贡献,

$$\begin{aligned} \Phi_x &\approx \int_0^h \int_0^h dy' dz' \times \\ &\left[\left(F_x + \frac{\partial F_x}{\partial x}h + \frac{\partial F_x}{\partial y}y' + \frac{\partial F_x}{\partial z}z' \right) - \left(F_x + \frac{\partial F_x}{\partial x}0 + \frac{\partial F_x}{\partial y}y' + \frac{\partial F_x}{\partial z}z' \right) \right] \\ &= \frac{\partial F_x}{\partial x}h \int_0^h \int_0^h dy' dz' = \frac{\partial F_x}{\partial x}h^3 = \frac{\partial F_x}{\partial x}V \end{aligned} \quad (5)$$

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

$$\Phi \approx \left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \right) V \quad (6)$$

根据定义式 2, 可得直角坐标中的散度公式

$$\nabla \cdot \mathbf{F} = \lim_{V \rightarrow 0} \frac{\Phi}{V} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \quad (7)$$

从形式上, 我们可以引入一个 ∇ 算符, 在直角坐标系中的形式为

$$\nabla = \hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial z} \quad (8)$$

那么 $\nabla \cdot \mathbf{F}(\mathbf{r})$ 从形式上可以看做矢量算符 ∇ 与某点场矢量 $\mathbf{F}(\mathbf{r})$ 的“点乘”. 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

$$\nabla \cdot [C_1 \mathbf{F}_1(\mathbf{r}) + C_2 \mathbf{F}_2(\mathbf{r}) + \dots] = C_1 \nabla \cdot \mathbf{F}_1(\mathbf{r}) + C_2 \nabla \cdot \mathbf{F}_2(\mathbf{r}) + \dots \quad (9)$$

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^[225]) 为

$$\mathbf{g}(\mathbf{r}) = -\frac{Gm}{r^3} \mathbf{r} \quad (10)$$

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

$$\begin{aligned}\nabla \cdot \mathbf{g} &= \frac{Gm}{(x^2 + y^2 + z^2)^{5/2}} [(2x^2 - y^2 - z^2) + (2y^2 - z^2 - x^2) + (2z^2 - x^2 - y^2)] \\ &= 0\end{aligned}\tag{11}$$

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 $\Phi_i \approx \nabla \cdot \mathbf{F}(\mathbf{r}_i)V_i$. 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

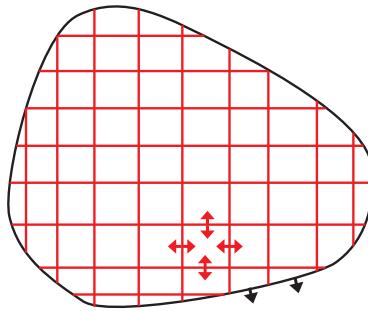


图 1: 证明所有小曲面的通量之和等于大曲面的通量

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

$$\Phi = \sum_i \Phi_i \approx \sum_i \nabla \cdot \mathbf{F}(\mathbf{r}_i)V_i\tag{12}$$

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})

$$\Phi = \int \nabla \cdot \mathbf{F}(\mathbf{r}) dV\tag{13}$$

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.

第五章

概率与统计

随机变量的变换

预备知识 分布函数

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 $f_2(x_2)$.

将两个概率分布的微分形式，有

$$dP = f_1(x_1) dx_1 \quad (1)$$

$$dP = f_2(x_2) dx_2 \quad (2)$$

若将式 1 中的 x_1 替换成 $g(x_2)$ ， dx_1 替换成 $g'(x_2) dx_2$ ，有

$$dP = f_1[g(x_2)]g'(x_2) dx_2 \quad (3)$$

对比式 2，得

$$f_2(x) = f_1[g(x)]g'(x) \quad (4)$$

这样，就求出了 x_2 的分布函数 $f_2(x)$.

例 1

已知 $f_1 x_1 = 3x_1^2$, $x_2 = x_1^2$, 求 x_2 的分布函数.

用 x_2 表示 x_1 得 $x_1 = \sqrt{x_2}$, 代入 $dP = f_1(x_1) dx_1$, 得

$$dP = 3\sqrt{x_2}^2 d(\sqrt{x_2}) = \frac{3}{2}x_2^{3/2} dx_2 \quad (5)$$

所以 x_2 的分布函数为 $f_2(x) = 3x^{3/2}/2$.

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

$$f_1(x_1) dx_1 = f_2(x_2) dx_2 \quad (6)$$

将方程两边积分即可得到两变量所满足的函数关系

$$F_1(x_1) = F_2(x_2) + C \quad (7)$$

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.

这个问题最常见的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

$$x_1 = x_2^2 + C \quad (8)$$

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

$$dP = \frac{2\pi r dr}{\pi} = 2r dr \quad (9)$$

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

$$1 d\xi_1 = 2r dr \quad (10)$$

两边积分得 $\xi_1 = r^2$ ($\xi_1 = 0$ 时 $r = 0$, 所以积分常数为零), 即 $r = \sqrt{\xi_1}$. 这样我们就可以根据给出的两个随机变量分别计算随机点的坐标了.

第二部分

力学

第一章

质点

位置矢量 位移

预备知识 矢量

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

$$\Delta\mathbf{r} = \mathbf{r}(t_2) - \mathbf{r}(t_1) \quad (1)$$

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^[122]，矢量的微分

例 1 证明 $dR = \hat{\mathbf{R}} \cdot d\mathbf{R}$

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正方向的投影.

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

$$dR = \frac{\partial R}{\partial x} dx + \frac{\partial R}{\partial y} dy = \frac{x}{\sqrt{x^2 + y^2}} dx + \frac{y}{\sqrt{x^2 + y^2}} dy \quad (2)$$

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， dx 和 dy 分别为 $d\mathbf{R}$ 的两个分量，根据点乘的定义^[91] 上式变为

$$dR = \hat{\mathbf{R}} \cdot d\mathbf{R} \quad (3)$$

拓展阅读 速度和加速度（矢量）[\[156\]](#)

速度 加速度（一维）

预备知识 位移[\[153\]](#)，基本初等函数的导数[\[41\]](#)，复合函数求导[\[48\]](#)，牛顿—莱布尼兹公式[\[69\]](#)

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带正负号的标量来表示，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。所以下面我们用坐标 x 来表示一维位移，实数 v 和 a 来表示一维速度和加速度。

物理学中，速度和加速度通常指瞬时值。在一维运动中，瞬时速度的定义为一段极短时间 Δt 内质点的位移[\[153\]](#) Δx 除以这段时间，瞬时加速度的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[\[34\]](#) 的定义。用极限符号和导数来表示，就是

$$v(t) = \lim_{\Delta t \rightarrow 0} \frac{x(t + \Delta t) - x(t)}{\Delta t} = \frac{dx(t)}{dt} \quad (1)$$

$$a(t) = \lim_{\Delta t \rightarrow 0} \frac{v(t + \Delta t) - v(t)}{\Delta t} = \frac{dv(t)}{dt} \quad (2)$$

根据高阶导数的定义，加速度就是位矢的二阶导数

$$a(t) = \frac{d^2x(t)}{dt^2} \quad (3)$$

例 1 匀加速运动

已知匀加速运动的位移为 $x(t) = x_0 + v_0 t + at^2/2$ ，注意到这是一个幂函数，求导得到速度为 $v(t) = v_0 + at$ ，再次求导（二阶导数）得到加速度为 $a(t) = a$ 。可见这是一个匀加速运动。

例 2 简谐振动

已知简谐振动的位移函数为 $x(t) = A \cos(\omega t)$ ，运用复合函数求导得速度为 $v(t) = -A\omega \sin(\omega t)$ ，加速度为 $a(t) = -A\omega^2 \cos(\omega t)$ 。

由速度或加速度求位移

既然一维速度是位置的导数（即 $x(t)$ 是速度的原函数）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

$$x(t_2) - x(t_1) = \int_{t_1}^{t_2} v(t) dt \quad (4)$$

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（为了区分积分变量和积分上限，我们把积分变量改成 t' ）

$$x(t) = x_0 + \int_{t_0}^t v(t') dt' \quad (5)$$

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 5 得

$$x(t) = x_0 + \int_{t_0}^t v_0 dt = x_0 + v_0 t \quad (6)$$

与式 4 和式 5 同理，一维速度和加速度之间也有类似关系

$$v(t_2) - v(t_1) = \int_{t_1}^{t_2} a(t) dt \quad (7)$$

$$v(t) = v_0 + \int_{t_0}^t a(t') dt' \quad (8)$$

例 4 匀加速直线运动

若质点在 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 $x(t)$.

我们首先由式 8 得到速度函数为

$$v(t) = v_0 + \int_{t_0}^t a_0 dt' = v_0 + a_0 t \quad (9)$$

然后再次积分得到位置函数

$$x(t) = x_0 + \int_{t_0}^t (v_0 + a_0 t') dt' = x_0 + v_0 t + a_0 t^2 / 2 \quad (10)$$

速度 加速度

预备知识 位置矢量^[153], 速度 加速度（一维）^[154], 矢量的导数^[126], 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

$$\mathbf{v} = \lim_{\Delta t \rightarrow 0} \frac{\Delta\mathbf{r}}{\Delta t} \quad (1)$$

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

$$\Delta\mathbf{r} = \mathbf{r}(t_1 + \Delta t) - \mathbf{r}(t_1) \quad (2)$$

式 2 代入式 1, 得

$$\mathbf{v}(t_1) = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{r}(t_1 + \Delta t) - \mathbf{r}(t_1)}{\Delta t} \quad (3)$$

根据矢量求导^[126] 的定义, 这就是位矢对时间的导数, 即

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} \quad (4)$$

应用举例 匀速圆周运动的速度（求导法）^[158]

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¹，类比速度的定义，加速的定义为

$$\mathbf{a}(t_1) = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{v}(t_1 + \Delta t) - \mathbf{v}(t_1)}{\Delta t} = \frac{d\mathbf{v}}{dt} \quad (5)$$

结合速度的定义，加速度为

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d}{dt} \left(\frac{d\mathbf{r}}{dt} \right) = \frac{d^2\mathbf{r}}{dt^2} \quad (6)$$

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。

应用举例 匀速圆周运动的速加速度（求导法^[34]）

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 ，该如何得到位移一时间函数 $\mathbf{r}(t)$ 呢？类比一维的情况^[154]，我们也可以通过矢量函数的定积分^[128]（见例 1）来求出速度一时间函数进而求出位移一时间函数

$$\mathbf{v}(t) = \mathbf{v}_0 + \int_{t_0}^t \mathbf{a}(t) dt \quad (7)$$

$$\mathbf{r}(t) = \mathbf{r}_0 + \int_{t_0}^t \mathbf{v}(t) dt \quad (8)$$

应用举例 匀加速运动^[160]

匀速圆周运动的速度（几何法）

¹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。

预备知识 小角正弦值极限^[32], 速度的定义^[156]

设一个点 A 做半径为 R 的圆周运动, 恒定的角速度为 ω , 那么经过一段微小时间 Δt 以后, 点 A 转过的角度为 $\Delta\theta = \omega\Delta t$. 粗略地说, 当 ω 不算太大时, $\Delta\theta$ 也是一个微小量. 这样, 根据小角正弦值极限^[32], 当 Δt 趋近于 0 时, 物体在 Δt 内走过的位移长度 (线段的长度) 趋近于弧的长度, 即 $|\Delta s|$ 趋近于 $R\omega\Delta t$.

根据速度的定义

$$\mathbf{v} = \lim_{\Delta t \rightarrow 0} \frac{\Delta \mathbf{s}}{\Delta t} \quad (1)$$

速度的大小为

$$v = \lim_{\Delta t \rightarrow 0} \frac{|\Delta s|}{\Delta t} = \frac{R\omega\Delta t}{\Delta t} = R\omega \quad (2)$$

速度的方向显然与过 A 点的圆的切线重合.

匀速圆周运动的速度 (求导法)

预备知识 矢量的导数^[126]

如图, 在平面直角坐标系 (单位矢量分别为 \hat{x} , \hat{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} , 圆周运动的半径为 R , 角速度为 ω (逆时针为正). $t = 0$ 时刻与 x 轴夹角为 0. 那么任意时刻将位矢 \mathbf{r} 沿着 x 与 y 轴方向分解, 则

$$\mathbf{r} = R(\hat{x} \cos \omega t + \hat{y} \sin \omega t) \quad (1)$$

其中 \hat{x} 是 x 轴正方向的单位矢量, \hat{y} 是 y 轴正方向的单位矢量. 这样, \mathbf{r} 就成了时间 t 的函数, 可直接求导. 根据速度的定义, $\mathbf{v} = d\mathbf{r}/dt$ 即

$$\begin{aligned} \mathbf{v} &= \frac{d}{dt}(\hat{x}R \cos \omega t + \hat{y}R \sin \omega t) \\ &= \hat{x}R \frac{d}{dt} \cos \omega t + \hat{y}R \frac{d}{dt} \sin \omega t \\ &= -\hat{x}R\omega \sin \omega t + \hat{y}R\omega \cos \omega t \\ &= R\omega(-\hat{x} \sin \omega t + \hat{y} \cos \omega t) \end{aligned} \quad (2)$$

对比式 1 和式 2, 比较位矢 \mathbf{r} 和速度矢量 \mathbf{v} , 可以发现速度的大小 $|\mathbf{v}| = R\omega = |\mathbf{r}|\omega$, 多了一个 ω 因子. \mathbf{v} 与 $\hat{\mathbf{x}}$ 轴的夹角比 \mathbf{r} 与 $\hat{\mathbf{x}}$ 轴的夹角大了, 即逆时针转了 90° .

匀速圆周运动的加速度（几何法）

预备知识 匀速圆周运动的速度（几何法）[\[157\]](#), 加速度（矢量）[\[156\]](#)

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[\[156\]](#).

$$\mathbf{a} = \lim_{\Delta t \rightarrow 0} \frac{\Delta \mathbf{v}}{\Delta t} \quad (1)$$

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).

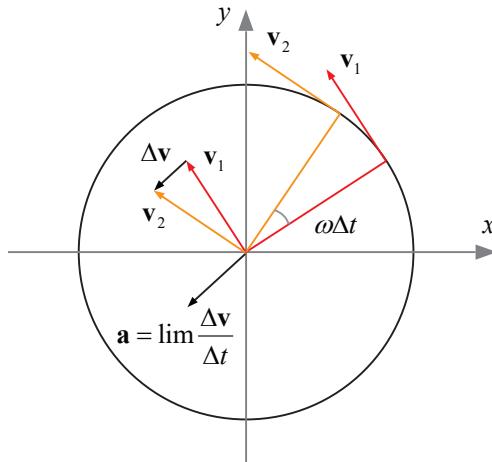


图 1: 把速度矢量移到原点再相减

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[\[32\]](#)” 中的结论, 把 $\Delta \mathbf{v}$ 的长度用弧长近似, 得

$$|\Delta \mathbf{v}| = |\mathbf{v}| |\Delta \theta| = (R\omega) \omega \Delta t \quad (2)$$

所以质点的加速度大小为

$$|\mathbf{a}| = \lim_{\Delta t \rightarrow 0} \frac{(R\omega)\omega\Delta t}{\Delta t} = R\omega^2 \lim_{\Delta t \rightarrow 0} \frac{\Delta t}{\Delta t} = R\omega^2 \quad (3)$$

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, 就得到加速度的矢量形式

$$\mathbf{a} = -\omega^2 \mathbf{r} \quad (4)$$

匀速圆周运动的加速度（求导法）

预备知识 匀速圆周运动的速度（求导法）[\[158\]](#)

在匀速圆周运动的速度（求导法）[\[158\]](#) 中, 令圆周运动的运动方程为

$$\mathbf{r} = R(\hat{\mathbf{x}} \cos \omega t + \hat{\mathbf{y}} \sin \omega t) \quad (1)$$

继而求出速度矢量关于时间的函数是

$$\mathbf{v} = -\hat{\mathbf{x}}R\omega \sin \omega t + \hat{\mathbf{y}}R\omega \cos \omega t \quad (2)$$

根据加速度的定义 $\mathbf{a} = d\mathbf{v}/dt$, 即上式再对时间求导.

$$\begin{aligned} \mathbf{a} &= \frac{d\mathbf{v}}{dt} = -\hat{\mathbf{x}}R\omega \frac{d}{dt} \sin \omega t + \hat{\mathbf{y}}R\omega \frac{d}{dt} \cos \omega t \\ &= -\hat{\mathbf{x}}R\omega^2 \cos \omega t - \hat{\mathbf{y}}R\omega^2 \sin \omega t = -\omega^2 \mathbf{r} \end{aligned} \quad (3)$$

匀加速运动

预备知识 速度 加速度[\[156\]](#)

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. 由“速度 加速度”中的[式 7](#)和[式 8](#), 速度和位移函数分别为

$$\mathbf{v}(t) = \mathbf{v}_0 + \int_{t_0}^t \mathbf{a} dt = \mathbf{v}_0 + \mathbf{a} \cdot (t - t_0) \quad (1)$$

$$\mathbf{r}(t) = \mathbf{r}_0 + \int_{t_0}^t \mathbf{v}(t) dt = \mathbf{r}_0 + \mathbf{v}_0 \cdot (t - t_0) + \frac{1}{2} \mathbf{a} \cdot (t - t_0)^2 \quad (2)$$

自由落体运动

一个最简单的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 $g \approx 9.8 \text{ m/s}^2$ 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

$$\mathbf{v}(t) = \mathbf{g} \cdot (t - t_0) \quad (3)$$

$$\mathbf{r}(t) = \mathbf{r}_0 + \frac{1}{2}\mathbf{g} \cdot (t - t_0)^2 \quad (4)$$

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

$$\mathbf{v}(t) = \mathbf{v}_0 + \mathbf{g} \cdot (t - t_0) \quad (5)$$

$$\mathbf{r}(t) = \mathbf{r}_0 + \mathbf{v}_0 \cdot (t - t_0) + \frac{1}{2}\mathbf{g} \cdot (t - t_0)^2 \quad (6)$$

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。

极坐标中的速度和加速度

预备知识 速度 加速度^[156], 极坐标中单位矢量的偏导^[136]

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

$$\mathbf{v} = \dot{r}\hat{\mathbf{r}} + r\frac{d\hat{\mathbf{r}}}{dt} \quad (1)$$

由链式法则和式 1^[136], 上式中

$$\frac{d\hat{\mathbf{r}}}{dt} = \frac{d\hat{\mathbf{r}}}{d\theta}\dot{\theta} = \dot{\theta}\hat{\theta} \quad (2)$$

所以极坐标中的速度为

$$\mathbf{v} = \dot{r}\hat{\mathbf{r}} + r\dot{\theta}\hat{\boldsymbol{\theta}} \quad (3)$$

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

$$\begin{aligned} \mathbf{a} &= \ddot{r}\hat{\mathbf{r}} + \dot{r}\frac{d\hat{\mathbf{r}}}{dt} + \dot{r}\dot{\theta}\hat{\boldsymbol{\theta}} + r\ddot{\theta}\hat{\boldsymbol{\theta}} + r\dot{\theta}\frac{d\hat{\boldsymbol{\theta}}}{dt} \\ &= \ddot{r}\hat{\mathbf{r}} + \dot{r}\dot{\theta}\hat{\boldsymbol{\theta}} + \dot{r}\dot{\theta}\hat{\boldsymbol{\theta}} + r\ddot{\theta}\hat{\boldsymbol{\theta}} - r\dot{\theta}^2\hat{\mathbf{r}} \\ &= (\ddot{r} - r\dot{\theta}^2)\hat{\mathbf{r}} + (2\dot{r}\dot{\theta} + r\ddot{\theta})\hat{\boldsymbol{\theta}} \end{aligned} \quad (4)$$

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

$$a_\theta = \frac{1}{r} \frac{d}{dt} \left(r^2 \frac{d\theta}{dt} \right) \quad (5)$$

牛顿运动定律 惯性系

预备知识 加速度^[156]

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运动或静止。
- **第二定律** 质点所受合外力等于的质量乘以加速度。
- **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

$$\mathbf{v}_B = \mathbf{v}_A + \mathbf{v}_{AB} \quad (1)$$

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^[178]进行修正。用矢量 \mathbf{F} 表示合力，牛顿第二定律记为

$$\mathbf{F} = m\mathbf{a} \quad (2)$$

事实上，牛顿本人对第二定律的表述使用了动量定理（单个质点）^[170]，记质点的动量为 \mathbf{p} ，则

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} \quad (3)$$

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²，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^[195]。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。

²在狭义相对论中，动量的定义有所不同。

功 功率

预备知识 力场^[166], 矢量的点乘^[91], 定积分^[65]

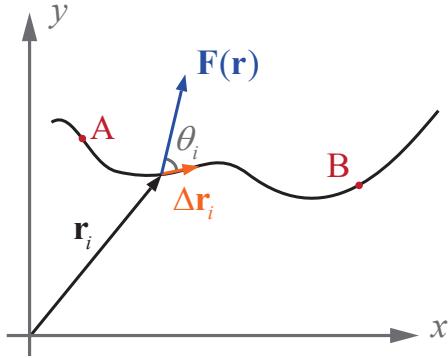


图 1: 在一小段位移中, 把变力看做恒力

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 $\Delta\mathbf{r}_1, \Delta\mathbf{r}_2 \dots \Delta\mathbf{r}_n$ 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

$$\Delta W_i = |\mathbf{F}(\mathbf{r}_i)| |\Delta\mathbf{r}_i| \cos \theta_i \quad (1)$$

上式可以表示成矢量点乘^[91]的形式

$$\Delta W_i = \mathbf{F}(\mathbf{r}_i) \cdot \Delta\mathbf{r}_i \quad (2)$$

把上式对所有的 i 求和, 就得到了做功的近似表达式

$$W_{ab} = \sum_{i=1}^n \Delta W_i \approx \sum_{i=1}^n \mathbf{F}(\mathbf{r}_i) \cdot \Delta\mathbf{r}_i \quad (3)$$

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^[65]中的介绍, 令 $n \rightarrow \infty$, 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.

$$W_{ab} = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbf{F}(\mathbf{r}_i) \cdot \Delta \mathbf{r}_i = \int_{C_{ab}} \mathbf{F}(\mathbf{r}) \cdot d\mathbf{r} \quad (4)$$

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^[137]”.

力的功率

功率 (瞬时) 的定义为做功的变化率, 即

$$P = \lim_{\Delta t \rightarrow 0} \frac{\Delta W}{\Delta t} = \frac{dW}{dt} \quad (5)$$

根据式 2, 力的功率为

$$P = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{F}(\mathbf{r}_i) \cdot \Delta \mathbf{r}_i}{\Delta t_i} = \mathbf{F} \cdot \frac{d\mathbf{r}}{dt} = \mathbf{F} \cdot \mathbf{v} \quad (6)$$

动能 动能定理 (单个质点)

预备知识 功 功率^[164], 牛顿第二定律^[162]

令质点的质量为 m , 速度为 \mathbf{v} , 则质点的动能定义为

$$E_k = \frac{1}{2} m \mathbf{v}^2 = \frac{1}{2} m v^2 \quad (1)$$

质点的动能定理是, 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^[164] 为

$$P = \frac{dW}{dt} = \mathbf{F} \cdot \frac{d\mathbf{r}}{dt} = \mathbf{F} \cdot \mathbf{v} \quad (2)$$

再来看动能的变化率

$$\frac{d}{dt} E_k = \frac{1}{2} m \frac{d}{dt} (\mathbf{v} \cdot \mathbf{v}) \quad (3)$$

由矢量点乘的求导^[126]式 7, $\mathbf{dv} \cdot \mathbf{v}/dt = 2\mathbf{v} \cdot d\mathbf{v}/dt = 2\mathbf{v} \cdot \mathbf{a}$, 上式变为

$$\frac{d}{dt} E_k = m\mathbf{a} \cdot \mathbf{v} = \mathbf{F} \cdot \mathbf{v} \quad (4)$$

最后一步使用了牛顿第二定律. 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 位置矢量^[153], 矢量场^[135], 功^[164], 牛顿—莱布尼兹公式^[69]

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 $\mathbf{F} = mg$). 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 . 如果用矢量的方法表示, 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^[153]) 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

$$\mathbf{F}(\mathbf{r}) = -G \frac{M}{r^2} \hat{\mathbf{r}} \quad (1)$$

若位矢用 \mathbf{r} 来表示 ($\mathbf{r} = r\hat{\mathbf{r}}$), 则

$$\mathbf{F}(\mathbf{r}) = -G \frac{M}{r^3} \mathbf{r} \quad (2)$$

现在变换到直角坐标系中, 有

$$\begin{cases} \mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \\ r = \sqrt{x^2 + y^2 + z^2} \end{cases} \quad (3)$$

代入上式，展开得

$$\mathbf{F}(\mathbf{r}) = -\frac{GMx}{(x^2 + y^2 + z^2)^{3/2}} \hat{\mathbf{x}} - \frac{GMy}{(x^2 + y^2 + z^2)^{3/2}} \hat{\mathbf{y}} - \frac{GMz}{(x^2 + y^2 + z^2)^{3/2}} \hat{\mathbf{z}} \quad (4)$$

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^[164]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

$$\int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{F}(\mathbf{r}) \cdot d\mathbf{r} = V(\mathbf{r}_1) - V(\mathbf{r}_2) \quad (5)$$

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

$$\mathbf{F}(x) = F(x) \hat{\mathbf{x}} \quad (6)$$

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

$$\int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{F}(\mathbf{r}) \cdot d\mathbf{r} = \int_a^b F(x) dx \quad (7)$$

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^[69]，势能函数恰好就是 $F(x)$ 的负原函数乘以，所以 $F(x)$ 是 $V(x)$ 负导函数。

$$V = - \int F(x) dx \quad F(x) = - \frac{dV(x)}{dx} \quad (8)$$

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{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

$$V(x) = - \int (-kx) dx = \frac{1}{2} kx^2 - C \quad (9)$$

为了确定待定常数, 代入 $V(0) = 0$, 解得 $C = 0$. 所以所求势能为

$$V(x) = \frac{1}{2} kx^2 \quad (10)$$

多维势能函数

预备知识 梯度定理^[139]

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^[139] 得

$$\int_{\mathbf{r}_i}^{\mathbf{r}_f} \nabla [-V(\mathbf{r})] \cdot d\mathbf{l} = V(\mathbf{r}_i) - V(\mathbf{r}_f) \quad (11)$$

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

$$\mathbf{F}(\mathbf{r}) = -\nabla V(\mathbf{r}) \quad (12)$$

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

$$F_x(\mathbf{r}) = -\frac{\partial V(\mathbf{r})}{\partial x} \quad F_y(\mathbf{r}) = -\frac{\partial V(\mathbf{r})}{\partial y} \quad \dots \quad (13)$$

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.

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

$$\begin{aligned}\mathbf{F}(\mathbf{r}) &= -\frac{\partial V}{\partial x}\hat{\mathbf{x}} - \frac{\partial V}{\partial y}\hat{\mathbf{y}} \\ &= -[(k_1 + k_2)x + (k_1 - k_2)y]\hat{\mathbf{x}} - [(k_1 - k_2)x + (k_1 + k_2)y]\hat{\mathbf{y}}\end{aligned}\quad (14)$$

现在我们根据“梯度定理^[139]”中的式 23 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

$$G_x(x, y) = -\frac{1}{2}(k_1 + k_2)x^2 - (k_1 - k_2)xy \quad (15)$$

$$G_y(x, y) = -(k_1 - k_2)xy - \frac{1}{2}(k_1 + k_2)y^2 \quad (16)$$

代入得（注意这里的场是势能函数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）

$$\begin{aligned}V(x, y) &= -G_y(x, y) + G_y(x, y_0) - G_x(x, y_0) + C \\ &= \frac{1}{2}(k_1 + k_2)x^2 + (k_1 - k_2)xy + \frac{1}{2}(k_1 + k_2)y^2 + C \\ &= \frac{1}{2}k_1(x + y)^2 + \frac{1}{2}k_2(x - y)^2 + C\end{aligned}\quad (17)$$

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 dt 内，两质点分别移动了 $d\mathbf{r}_A$ ，和 $d\mathbf{r}_B$ ，则相互作用力对二者做功为

$$dW = \mathbf{F} \cdot d\mathbf{r}_B + (-\mathbf{F}) \cdot d\mathbf{r}_A = \mathbf{F} \cdot d\mathbf{R} = F(R)\hat{\mathbf{R}} \cdot d\mathbf{R} = F(R) dR \quad (18)$$

（最后一步的证明见“位置矢量^[153]”中的例 1）定积分得

$$W = \int_{R_1}^{R_2} F(R) dR \quad (19)$$

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

$$W = V(R_2) - V(R_1) \quad (20)$$

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^[165]，势能^[166]

若质点只受不随时间变化的保守力作用³，那么物体从在某段时间从 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

$$W_{AB} = V(\mathbf{r}_A) - V(\mathbf{r}_B) \quad (1)$$

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

$$W_{AB} = E_{kB} - E_{kA} = \frac{1}{2}mv_B^2 - \frac{1}{2}mv_A^2 \quad (2)$$

结合以上两式，得

$$E_{kA} + V(\mathbf{r}_A) = E_{kB} + V(\mathbf{r}_B) \quad (3)$$

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^[162]，矢量的导数^[126]

令质点质量为 m ，速度为 \mathbf{v} ，定义其动量为

$$\mathbf{p} = m\mathbf{v} \quad (1)$$

注意动量是矢量，与速度（矢量）的方向相同，且取决于坐标系。

³即势能函数不随时间变化

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}

$$\frac{d\mathbf{p}}{dt} = m \frac{d\mathbf{v}}{dt} = m\mathbf{a} \quad (2)$$

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

$$\frac{d\mathbf{p}}{dt} = \mathbf{F} \quad (3)$$

这就是动量定理的微分形式, 即动量的变化率等合外力. 也可以写成微分形式

$$d\mathbf{p} = \mathbf{F} dt \quad (4)$$

即微小时间内的动量变化等于力乘以微小时间.

现在用定积分^[65] 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 $\mathbf{F}(t_i)$

$$\mathbf{p}(t_2) - \mathbf{p}(t_1) = \sum_{i=1}^N \Delta\mathbf{p}_i = \sum_{i=1}^N \mathbf{F}(t_i) \Delta t_i \quad (5)$$

当 $N \rightarrow \infty, \Delta t \rightarrow 0$ 时该式可以用定积分(矢量函数)表示⁴

$$\mathbf{p}(t_2) - \mathbf{p}(t_1) = \int_{t_1}^{t_2} \mathbf{F}(t) dt \quad (6)$$

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 $(t_2 - t_1)\mathbf{F}$.

角动量定理 角动量守恒 (单个质点)

预备知识 角动量^[200], 牛顿第二定律^[162], 力矩^[197]

一个质点的质量为 m , 某时刻速度为 \mathbf{v} . 则其动量为 $\mathbf{p} = m\mathbf{v}$. 在三维空间中指定一点 O 为参考点, O 点到质点的矢量为 \mathbf{r} . 根据定义, 该质点关于 O 点的角动量为 $\mathbf{L} = \mathbf{r} \times \mathbf{p}$. 这个质点在该时刻受到的力矩为 \mathbf{M} , 可以证明

$$\frac{d\mathbf{L}}{dt} = \mathbf{M} \quad (1)$$

⁴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”

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 $d\mathbf{L}/dt = \mathbf{0}$ ，即角动量不随时间变化. 这个现象叫做（单个质点的）角动量守恒. 由力矩的定义， $\mathbf{M} = \mathbf{r} \times \mathbf{F}$ ，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.
2. \mathbf{F} 与 \mathbf{r} 同向，即质点只受关于 O 点的有心力.

单个质点的角动量定理证明

质点的速度为 $\mathbf{v} = d\mathbf{r}/dt$ ，加速度为 $\mathbf{a} = dv/dt$ ，叉乘的求导法则与标量乘法求导类似，牛顿第二定律为 $\mathbf{F} = m\mathbf{a}$ ，两个同方向矢量叉乘为零，

$$\begin{aligned}\frac{d\mathbf{L}}{dt} &= \frac{d(\mathbf{r} \times \mathbf{p})}{dt} = m \frac{d(\mathbf{r} \times \mathbf{v})}{dt} = m \left(\frac{d\mathbf{r}}{dt} \times \mathbf{v} + \mathbf{r} \times \frac{d\mathbf{v}}{dt} \right) \\ &= m(\mathbf{v} \times \mathbf{v} + \mathbf{r} \times \mathbf{a}) = \mathbf{r} \times (m\mathbf{a}) \\ &= \mathbf{r} \times \mathbf{F} = \mathbf{M}\end{aligned}\tag{2}$$

简谐振子

预备知识 胡克定律，牛顿第二定律^[162]

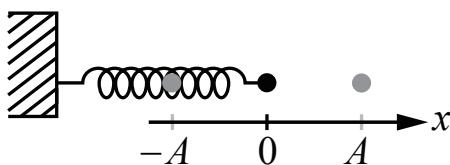


图 1: 简谐振子模型

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^[162] $F = ma = m\ddot{x}$ (\ddot{x} 代表对时间的二阶导数). 两式消去 F , 得

$$m\ddot{x} = -kx \quad (1)$$

这是一个单变量函数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. 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⁵, 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, 弹簧的震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)

$$x = A \cos(\omega t + \varphi_0) \quad (\omega = \sqrt{k/m}) \quad (2)$$

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^[73]), 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 (或简谐震动). 其中 A 为振幅, $\omega t + \varphi_0$ 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 $A \cos \varphi_0 = 0$, $-A\omega \sin \varphi_0 = v_0$, 解得 $\varphi_0 = \pi/2$, $A = -v_0\omega$. 所以

$$x = -v_0\omega \cos\left(\omega t + \frac{\pi}{2}\right) = v_0\omega \sin \omega t \quad (\omega = \sqrt{k/m}) \quad (3)$$

受阻落体

预备知识 匀加速运动^[160]

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^[162] 可以列出动力学方程 (假设向

⁵ $\sin t$ 也有同样的性质, 所以以下讨论对 $\sin t$ 也成立

下为正方向)

$$ma = F = mg - \alpha v \quad (1)$$

考虑到加速度是速度的导数, 上式变为

$$\frac{dv}{dt} = g - \frac{\alpha}{m}v \quad (2)$$

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^[71]. 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 dt 并整理得

$$\frac{1}{g - \alpha v/m} dv = dt \quad (3)$$

这样我们就得到了 v 和 t 的微分^[47] 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

$$F(v) = G(t) \quad (4)$$

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

$$f(v) dv = g(t) dt \quad (5)$$

对比式 3 可得 $f(v) = 1/(g - \alpha v/m)$ 和 $g(t) = 1$, 把二者做不定积分得原函数. 首先显然 $G(t) = t + C_1$. 对 $f(v)$ 积分可用“积分表^[59]”中的式 1 和式 3 得

$$F(v) = -\frac{m}{\alpha} \ln \left| g - \frac{\alpha}{m}v \right| + C_2 = -\frac{m}{\alpha} \ln \left(g - \frac{\alpha}{m}v \right) + C_2 \quad (6)$$

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 dv/dt 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

$$v = \frac{m}{\alpha} \left(g - e^{-\alpha C/m} e^{-\alpha t/m} \right) \quad (7)$$

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

$$v(t) = \frac{mg}{\alpha} (1 - e^{-\alpha t/m}) \quad (8)$$

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 $t \rightarrow +\infty$ 时， $v(t) \rightarrow mg/\alpha$. 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^[172]，匀速圆周运动^[160]

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.

$$T - mg \cos \theta = ma_c \quad (1)$$

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.

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

$$\ddot{\theta} = a_\theta/L \quad (2)$$

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

$$-mg \sin \theta = ma_\theta = -m\ddot{\theta}L \quad (3)$$

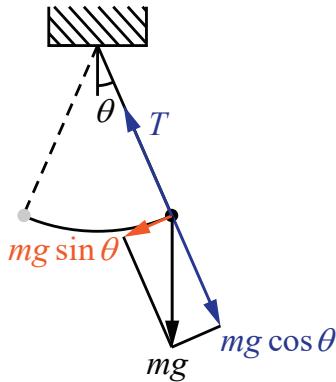


图 1: 单摆

两边消去质量可得摆角 θ 关于时间的二阶微分方程.

$$L\ddot{\theta} = -g \sin \theta \quad (4)$$

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^[32]

遗憾的是, 式 4 的解并不能用有限个基本初等函数表示. 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 $\theta \rightarrow 0$ 时, 我们可以把式 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

$$\ddot{s} = -\frac{g}{L}s \quad (5)$$

观察该式可以发现其结构与简谐振子的微分方程 (式 1^[172]) 非常相似. 用同样的方法, 可得通解为

$$s = A \cos(\omega t + \varphi_0) \quad (\omega = \sqrt{g/L}) \quad (6)$$

考虑到 $\theta \rightarrow 0$ 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 .

傅科摆

预备知识 单摆^[175]

傅科摆是首个直接证明自转的实验，首先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

$$\omega = \omega_0 \sin \alpha \quad (1)$$

傅科摆角速度的一种几何推导

预备知识 矢量叉乘^[96]

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 $d\theta$ ，有

$$d\mathbf{P} = \hat{\mathbf{M}} \times \mathbf{P} d\theta \quad (2)$$

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 ds ($ds \cdot \mathbf{R} = 0$) 后，由式 2 可得

$$d\hat{\mathbf{A}} = \hat{\mathbf{M}} \times \hat{\mathbf{A}} \cdot d\theta = \frac{\mathbf{R} \times ds}{|\mathbf{R} \times ds|} \times \hat{\mathbf{A}} \frac{ds}{R} \quad (3)$$

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 $d\theta$ 时，上式中 $ds = \hat{\mathbf{k}} \times \mathbf{R} d\theta$ ，而地面上的标记 $\hat{\mathbf{B}}$ 也围绕地轴转动，所以 $d\hat{\mathbf{B}} = \hat{\mathbf{k}} \times \hat{\mathbf{A}} d\theta$ 。

下面计算 $d\hat{\mathbf{A}} - d\hat{\mathbf{B}}$. 因为 $\mathbf{R} \cdot d\mathbf{s} = 0$, 所以 $|\mathbf{R} \times d\mathbf{s}| = R ds$, 所以

$$\begin{aligned} d\hat{\mathbf{A}} &= \frac{\mathbf{R} \times d\mathbf{s}}{R^2} \times \hat{\mathbf{A}} = \frac{1}{R^2} \mathbf{R} \times (\hat{\mathbf{k}} \times \mathbf{R} d\theta) \times \hat{\mathbf{A}} \\ &= \hat{\mathbf{R}} \times (\hat{\mathbf{k}} \times \hat{\mathbf{R}}) \times \hat{\mathbf{A}} d\theta = [(\hat{\mathbf{R}} \cdot \hat{\mathbf{R}})\hat{\mathbf{k}} - (\hat{\mathbf{R}} \cdot \hat{\mathbf{k}})\hat{\mathbf{R}}] \times \hat{\mathbf{A}} d\theta \\ &= (\hat{\mathbf{k}} - \hat{\mathbf{R}} \sin \alpha) \times \hat{\mathbf{A}} d\theta \end{aligned} \quad (4)$$

$$d\hat{\mathbf{A}} - d\hat{\mathbf{B}} = (\hat{\mathbf{k}} - \hat{\mathbf{R}} \sin \alpha) \times \hat{\mathbf{A}} d\theta - \hat{\mathbf{k}} \times \hat{\mathbf{A}} d\theta = -\sin \alpha \hat{\mathbf{R}} \times \hat{\mathbf{A}} d\theta \quad (5)$$

所以地球转过 $d\theta$ 角以后, $\hat{\mathbf{A}}$ 与 $\hat{\mathbf{B}}$ 之间的夹角为

$$d\gamma = |d\hat{\mathbf{A}} - d\hat{\mathbf{B}}| = \sin \alpha d\theta \quad (6)$$

两边除以 dt 得角速度

$$\omega = \omega_0 \sin \alpha \quad (7)$$

惯性力

预备知识 牛顿第二定律^[162], 矩阵^[105]

若一个质点 m 某时刻在惯性系 xyz 中的加速度为 \mathbf{a}_{xyz} , 在另一个非惯性系 abc 中的加速度为 \mathbf{a}_{abc} , 则可假设质点受到一个惯性力

$$\mathbf{f} = m(\mathbf{a}_{abc} - \mathbf{a}_{xyz}) \quad (1)$$

使得牛顿运动定律在非惯性系 abc 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“额外的重力”. 为了使自己保持静止, 电梯不得不给人额外的支持力.

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, 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 abc 相对于惯性系 xyz 没有旋转只有平移, 且 t 时刻的相对加速度为 $\mathbf{a}(t)$. 这样, abc 中的任何一个静止点相对于 xyz 系的加速度都是 $\mathbf{a}(t)$. 设 abc 系中有一质点 m , 相对于 abc 系的加速度为 $\mathbf{a}_{abc}(t)$, 那么在惯性系中质点的加速度为 $\mathbf{a}_{xyz} = \mathbf{a}_{abc} + \mathbf{a}$, 即**绝对加速度**⁶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!)

$$\mathbf{F} = m\mathbf{a}_{xyz} = m(\mathbf{a} + \mathbf{a}_{abc}) \quad (2)$$

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

$$\mathbf{F} + \mathbf{f} = m\mathbf{a}_{abc} \quad (3)$$

所以

$$\mathbf{f} = m\mathbf{a}_{abc} - \mathbf{F} = -m\mathbf{a} \quad (4)$$

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. 如果人站在一个秤上, 弹簧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

⁶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.

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。这时并不能类比上面直接得到 $\mathbf{f} = -m\mathbf{a}$ 的结论，即使这里把加速度变为位置和时间的函数（对于非平动参考系，静止点相对于惯性系的加速度显然也取决于位置）。这是因为，对非平动参考系中，并不满足绝对加速度等于相对加速度加牵连加速度，即并不满足 $\mathbf{a}_{xyz} = \mathbf{a}_{abc} + \mathbf{a}$ （见加速度叠加）。对任何非惯性系都成立的普适结论只有

$$\mathbf{f} = m\mathbf{a}_{abc} - \mathbf{F} = m\mathbf{a}_{abc} - m\mathbf{a}_{xyz} = m(\mathbf{a}_{abc} - \mathbf{a}_{xyz}) \quad (5)$$

即，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。

计算惯性力的坐标法（适用于任何非惯性系）

1. 求出所选惯性系和非惯性系之间的坐标变换和逆变换关系

$$\begin{cases} x = f_1(a, b, c, t) \\ y = f_2(a, b, c, t) \\ z = f_3(a, b, c, t) \end{cases} \quad \begin{cases} a = g_1(x, y, z, t) \\ b = g_2(x, y, z, t) \\ c = g_3(x, y, z, t) \end{cases} \quad (6)$$

2. 对 x, y, z 求时间的二阶全导数^[125]，得出惯性系中质点的加速度（用 a, b, c, t 表示）。

$$\begin{cases} \ddot{x} = \frac{d^2}{dt^2} f_1(a, b, c, t) \\ \ddot{y} = \frac{d^2}{dt^2} f_2(a, b, c, t) \\ \ddot{z} = \frac{d^2}{dt^2} f_3(a, b, c, t) \end{cases} \quad (7)$$

3. 把 xyz 系中的加速度矢量⁷ $(\ddot{x}, \ddot{y}, \ddot{z})^T$ 用 g_1, g_2, g_3 逆变换到 abc 系中表示，

$$\begin{pmatrix} \ddot{x} \\ \ddot{y} \\ \ddot{z} \end{pmatrix}_{abc} = \begin{pmatrix} g_1(\ddot{x}, \ddot{y}, \ddot{z}, t) \\ g_2(\ddot{x}, \ddot{y}, \ddot{z}, t) \\ g_3(\ddot{x}, \ddot{y}, \ddot{z}, t) \end{pmatrix} - \begin{pmatrix} g_1(0, 0, 0, t) \\ g_2(0, 0, 0, t) \\ g_3(0, 0, 0, t) \end{pmatrix} \quad (8)$$

⁷上标 T 表示矩阵转置，将行矢量变为列矢量

4. 质点受到的惯性力为

$$\mathbf{f} = m(\mathbf{a}_{abc} - \mathbf{a}_{xyz}) = m \begin{pmatrix} \ddot{a} \\ \ddot{b} \\ \ddot{c} \end{pmatrix} - m \begin{pmatrix} \ddot{x} \\ \ddot{y} \\ \ddot{z} \end{pmatrix}_{abc} \quad (9)$$

注意第三步中的坐标变换是必须的，因为 $(\ddot{a}, \ddot{b}, \ddot{c})^T$ 是 abc 系中的矢量坐标， $(\ddot{x}, \ddot{y}, \ddot{z})^T$ 是 xyz 系中的矢量坐标，直接相减没有任何物理意义。必须把 $(\ddot{x}, \ddot{y}, \ddot{z})^T$ 变换到 abc 系中表示，或者把 $(\ddot{a}, \ddot{b}, \ddot{c})^T$ 变换到 xyz 系中表示才可相减。由于惯性力是非惯性系中的假想力，所以一般把所有矢量用非惯性系中的坐标表示。

离心力

预备知识 匀速圆周运动的加速度^{[159][160]}，惯性力^[178]

令参考系 abc 和 xyz 的 c 轴和 z 轴式中重合。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（离心力）。

令质点的坐标 (a, b, c) 离 c 轴的距离为 $r = \sqrt{b^2 + c^2}$ ，对应的径向矢量为 $\mathbf{r} = (a, b, 0)^T$ 。在 xyz 系中，质点做匀速圆周运动，相对于 xyz 系的加速度（绝对加速度）（用 abc 系的坐标表示）为

$$\mathbf{a}_{xyz} = -\omega^2 \mathbf{r} = -\omega^2 \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} \quad (1)$$

质点相对于 abc 系静止，相对加速度为零

$$\mathbf{a}_{abc} = \mathbf{0} \quad (2)$$

所以由惯性力^[178] 中的结论，惯性力为

$$\mathbf{f} = m(\mathbf{a}_{abc} - \mathbf{a}_{xyz}) = m\omega^2 \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} \quad (3)$$

注意离心力向外，与直觉相符。

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^[182]。

科里奥利力

预备知识 惯性力^[178]，离心力^[181]，平面旋转矩阵，矢量的叉乘^[96]

结论

科里奥利力 (Coriolis Force) 是匀速旋转的参考系中的由相对速度产生的惯性力。

$$\mathbf{F}_{cori} = 2m\mathbf{v}_{abc} \times \boldsymbol{\omega} \quad (1)$$

其中 \mathbf{v}_{abc} 是质点相对于旋转系的瞬时速度， $\boldsymbol{\omega}$ 是旋转系的恒定角速度矢量。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。其中是地球的自转角速度。

推导

这里的主要思路是按照惯性力^[178] 中的坐标法。设空间中存在一个惯性系 xyz 和一个非惯性系 abc 相对于 xyz 绕 z 轴以角速度 $\boldsymbol{\omega}$ 逆时针匀速旋转（右手定则）。由于 z 轴和 c 轴始终重合 ($z = c$)，只需要考虑 x, y 坐标和 a, b 坐标之间的关系即可。

令平面旋转矩阵为

$$\mathbf{R}(\theta) \equiv \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (2)$$

其意义是把坐标逆时针旋转角 θ 。则其逆矩阵等于转置矩阵

$$\mathbf{R}(-\theta) = \mathbf{R}^T(\theta) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \quad (3)$$

意义是顺时针旋转 θ 角。所以两坐标系之间的坐标变换为

$$\begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{R}^T(\omega t) \begin{pmatrix} x \\ y \end{pmatrix} \quad \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{R}(\omega t) \begin{pmatrix} a \\ b \end{pmatrix} \quad (4)$$

为了得到质点在惯性系中的加速度，对上式的 $(x, y)^T$ 求二阶全导数得⁸

$$\begin{pmatrix} \ddot{x} \\ \ddot{y} \end{pmatrix} = \ddot{\mathbf{R}}(\omega t) \begin{pmatrix} a \\ b \end{pmatrix} + 2\dot{\mathbf{R}}(\omega t) \begin{pmatrix} \dot{a} \\ \dot{b} \end{pmatrix} + \mathbf{R}(\omega t) \begin{pmatrix} \ddot{a} \\ \ddot{b} \end{pmatrix} \quad (5)$$

其中⁹

$$\dot{\mathbf{R}}(\omega t) = \omega \begin{pmatrix} \cos(\omega t + \pi/2) & -\sin(\omega t + \pi/2) \\ \sin(\omega t + \pi/2) & \cos(\omega t + \pi/2) \end{pmatrix} = \omega \mathbf{R}(\omega t + \pi/2) \quad (6)$$

$$\ddot{\mathbf{R}}(\omega t) = -\omega^2 \mathbf{R}(\omega t) \quad (7)$$

代入式 5 得

$$\begin{pmatrix} \ddot{x} \\ \ddot{y} \end{pmatrix} = -\omega^2 \mathbf{R}(\omega t) \begin{pmatrix} a \\ b \end{pmatrix} + 2\omega \mathbf{R}(\omega t + \pi/2) \begin{pmatrix} \dot{a} \\ \dot{b} \end{pmatrix} + \mathbf{R}(\omega t) \begin{pmatrix} \ddot{a} \\ \ddot{b} \end{pmatrix} \quad (8)$$

上式中的每一项都是 xyz 参考系中的坐标。所有坐标顺时针旋转 ωt ，得到 abc 参考系中的坐标。首先由旋转矩阵的几何意义得 $\mathbf{R}(\theta_1)\mathbf{R}(\theta_2) = \mathbf{R}(\theta_1 + \theta_2)$

$$\begin{pmatrix} \ddot{x} \\ \ddot{y} \end{pmatrix}_{abc} = \mathbf{R}(-\omega t) \begin{pmatrix} \ddot{x} \\ \ddot{y} \end{pmatrix} = -\omega^2 \begin{pmatrix} a \\ b \end{pmatrix} + 2\omega \mathbf{R}(\pi/2) \begin{pmatrix} \dot{a} \\ \dot{b} \end{pmatrix} + \begin{pmatrix} \ddot{a} \\ \ddot{b} \end{pmatrix} \quad (9)$$

所以旋转参考系中的总惯性力（式 9^[178]）为（用 abc 系中的坐标表示）

$$\mathbf{f} = m(\mathbf{a}_{abc} - \mathbf{a}_{xyz}) = \begin{pmatrix} \ddot{a} \\ \ddot{b} \end{pmatrix} - \begin{pmatrix} \ddot{x} \\ \ddot{y} \end{pmatrix}_{abc} = m\omega^2 \begin{pmatrix} a \\ b \end{pmatrix} - 2m\omega \mathbf{R}(\pi/2) \begin{pmatrix} \dot{a} \\ \dot{b} \end{pmatrix} \quad (10)$$

其中第一项是已知的离心力^[181]

$$m\omega^2 \begin{pmatrix} a \\ b \end{pmatrix} = m\omega^2 \mathbf{r} \quad (11)$$

⁸本词条中，某个量上方加一点表示对时间的一阶导数，两点表示对时间的二阶导数。

⁹式 6 和式 7 相当于用矩阵推导了圆周运动的速度和加速度公式^{[157][159]}。

我们将第二项定义为科里奥利力. 由叉乘的几何意义^[96], 科里奥利力可以用叉乘记为

$$-2m\omega \mathbf{R}(\pi/2) \begin{pmatrix} \dot{a} \\ \dot{b} \end{pmatrix} = 2m\mathbf{v}_{abc} \times \boldsymbol{\omega} \quad (12)$$

其中 $\boldsymbol{\omega}$ 是 abc 系旋转的角速度矢量, \mathbf{v}_{abc} 是质点相对于 abc 系的速度. 最后, 矢量形式的式 10 为

$$\mathbf{f} = m\omega^2 \mathbf{r} + 2m\mathbf{v}_{abc} \times \boldsymbol{\omega} \quad (13)$$

地球表面的科里奥利力

预备知识 科里奥利力^[182]

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¹⁰). 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

$$\hat{\mathbf{v}} = \sin \phi \hat{\mathbf{y}} + \cos \phi \hat{\mathbf{z}}' = -\cos \phi \sin \theta \hat{\mathbf{x}} + \sin \phi \hat{\mathbf{y}} + \cos \phi \cos \theta \hat{\mathbf{z}} \quad (1)$$

现在可以计算科里奥利力

$$\mathbf{F}_{col} = 2mv\omega \hat{\mathbf{v}} \times \hat{\mathbf{z}} = 2mv\omega (\cos \phi \sin \theta \hat{\mathbf{y}} + \sin \phi \hat{\mathbf{z}}) \quad (2)$$

其向北, 向东, 向上的分量分别为

$$\mathbf{F}_{col} \cdot \hat{\mathbf{z}}' = -2mv\omega \sin \theta \sin \phi \quad \mathbf{F}_{col} \cdot \hat{\mathbf{y}} = 2mv\omega \sin \theta \cos \phi \quad (3)$$

$$\mathbf{F}_{col} \cdot \hat{\mathbf{x}}' = 2mv\omega \sin \phi \cos \theta \quad (4)$$

¹⁰但严格来说地球由于受离心力, 赤道宽, 两极窄.

可以证明水平分力可以表示为

$$\mathbf{F}_{col}^{\parallel} = (\mathbf{F}_{col} \cdot \hat{\mathbf{y}}) \hat{\mathbf{y}} + (\mathbf{F}_{col} \cdot \hat{\mathbf{z}}') \hat{\mathbf{z}}' = 2m\mathbf{v} \times \boldsymbol{\omega}' \quad (5)$$

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

$$\begin{aligned} F_{col} &= 2 \times 30,000 \text{ kg} \times \frac{300,000 \text{ m}}{3600 \text{ s}} \times \frac{2\pi}{24 \text{ h} \times 3600 \text{ s}} \times \sin \frac{\pi}{6} \\ &= 60.32 \text{ N} \end{aligned} \quad (6)$$

第二章

质点系与刚体

质点系

预备知识 牛顿第三定律^[162]

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，称为质点系。

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，将外力和内力分别记为 \mathbf{F}_i^{out} 和 \mathbf{F}_i^{in} 。

系统中所有质点所受的合力等于合内力加合外力

$$\mathbf{F}_{tot} = \mathbf{F}_{tot}^{in} + \mathbf{F}_{tot}^{out} = \sum_i^N \mathbf{F}_i^{in} + \sum_i^N \mathbf{F}_i^{out} \quad (1)$$

若将第 j 个质点对第 i 个质点的内力记为 $\mathbf{F}_{j \rightarrow i}$ 则上式中

$$\sum_i \mathbf{F}_i^{in} = \sum_{i,j}^{i \neq j} \mathbf{F}_{j \rightarrow i} \quad (2)$$

任意两个质点 k 和 l 对该求和的贡献是一对相互作用力 $\mathbf{F}_{k \rightarrow l} + \mathbf{F}_{l \rightarrow k}$ ，而根据牛顿第三定律，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

$$\mathbf{F}_{tot} = \mathbf{F}_{tot}^{out} = \sum_i^N \mathbf{F}_i^{out} \quad (3)$$

质心 质心系

预备知识 体积分^[132]，质点系^[187]

质心的定义

对质点系，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

$$\mathbf{r}_c = \frac{1}{M} \sum_i m_i \mathbf{r}_i \quad (1)$$

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

$$M = \int \rho(\mathbf{r}) dV \quad (2)$$

质心定义为

$$\mathbf{r}_c = \frac{1}{M} \int \mathbf{r} \rho(\mathbf{r}) dV \quad (3)$$

以下的讨论都对质点系进行，连续质量分布可看做由许多体积微元组成，也可看做质点系。

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？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

$$\mathbf{r}_{Ac} = \mathbf{r}_{AB} + \mathbf{r}_{Bc} \quad (4)$$

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

$$\mathbf{r}_{Ac} = \frac{1}{M} \sum_i m_i \mathbf{r}_{Ai} \quad \mathbf{r}_{Bc} = \frac{1}{M} \sum_i m_i \mathbf{r}_{Bi} \quad (5)$$

由位矢的坐标系变换， $\mathbf{r}_{Ai} = \mathbf{r}_{AB} + \mathbf{r}_{Bi}$ ，所以

$$\mathbf{r}_{Ac} = \frac{1}{M} \sum_i m_i (\mathbf{r}_{AB} + \mathbf{r}_{Bi}) = \mathbf{r}_{AB} + \frac{1}{M} \sum_i m_i \mathbf{r}_{Bi} = \mathbf{r}_{AB} + \mathbf{r}_{Bc} \quad (6)$$

质心系

定义质点系的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 4），在质心系中计算质心（式 1）仍然落在原点，即

$$\sum_i m_i \mathbf{r}_{ci} = \mathbf{0} \quad (7)$$

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，质心系才是惯性系。在非惯性系中，每个质点受惯性力。

质心系中总动量

把式 7 两边对时间求导，得

$$\sum_i m_i \mathbf{v}_{ci} = \mathbf{0} \quad (8)$$

注意到等式左边恰好为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**.

刚体

预备知识 矢量叉乘^[96]，球坐标系^[11]，质心^[187]

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。

在没有任何约束的情况下，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 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 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 6 个自由度。

我们可以这么划分 6 个自由度：令其中 3 个决定刚体质心的位置，2 个决定过质心的某条轴的朝向（球坐标中的两个角度），1 个决定刚体绕这条轴旋转的角度。

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 . 我们不妨再定义角速度 ω 关于时间的导数（即 θ 关于时间的二阶导数）为角加速度，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

$$\theta = \theta_0 + \omega t + \frac{1}{2}\alpha t^2 \quad (1)$$

$$\omega_1^2 - \omega_0^2 = 2\alpha\theta \quad (2)$$

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^[95] 来判断.

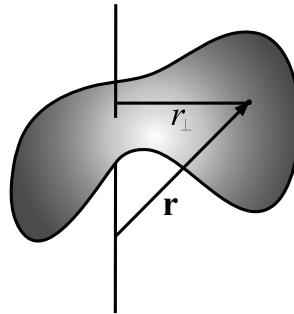


图 1: 刚体绕轴旋转时任意一点的线速度

我们再来考虑刚体旋转时其中任意一点的瞬时速度（如图 1). 显然刚体绕轴旋转时，刚体中不在轴上的点都做圆周运动，速度大小为 $\omega \times r_{\perp}$. 其中 r_{\perp} 是该点到转轴的垂直距离. 如果令 \mathbf{r} 为转轴上任意一点指向刚体中某点的矢量，根据矢量叉乘^[96] 的几何定义，该点的瞬时线速度可以表示为

$$\mathbf{v} = \mathbf{r} \times \boldsymbol{\omega} \quad (3)$$

二体系统

预备知识 质心 质心系^[187]

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都会做匀速直线运动.

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

$$\mathbf{R} = \mathbf{r}_B - \mathbf{r}_A \quad (1)$$

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

$$m_A \mathbf{r}_A + m_B \mathbf{r}_B = \mathbf{0} \quad (2)$$

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, 所以不受外力的二体系统只有三个自由度

$$\mathbf{r}_A = \frac{m_B}{m_A + m_B} \mathbf{R} \quad \mathbf{r}_B = \frac{m_A}{m_A + m_B} \mathbf{R} \quad (3)$$

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

$$\ddot{\mathbf{R}} = \ddot{\mathbf{r}}_B - \ddot{\mathbf{r}}_A = \frac{m_A + m_B}{m_A m_B} \mathbf{F} \quad (4)$$

若定义两质点的约化质量为

$$\mu = \frac{m_A m_B}{m_A + m_B} \quad (5)$$

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

$$\mathbf{F} = \mu \ddot{\mathbf{R}} \quad (6)$$

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 的作用力可以看成原点对等效质点的有心力.

机械能守恒

再来看系统的动能. 使用式3把系统在质心系中的总动能用相对位矢表示得

$$E_k = \frac{1}{2}(m_A \dot{\mathbf{r}}_A^2 + m_B \dot{\mathbf{r}}_B^2) = \frac{1}{2} \frac{m_A m_B}{m_A + m_B} \dot{\mathbf{R}}^2 = \frac{1}{2} \mu \dot{\mathbf{R}}^2 \quad (7)$$

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

$$\mathbf{F}(\mathbf{R}) = F(R) \hat{\mathbf{R}} \quad (8)$$

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^[166]”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

$$E = \frac{1}{2} \mu \dot{\mathbf{R}}^2 + V(R) \quad (9)$$

由于系统不受外力, 机械能守恒.

二体碰撞

预备知识 二体系统^[191]

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 $x_c = (m_1 x_1 + m_2 x_2) / (m_1 + m_2)$, 等式两边对时间 t 求导, 得质心的速度为

$$v_c = (m_1 v_1 + m_2 v_2) / (m_1 + m_2) \quad (1)$$

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 (二体系统^[191]式 3), 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

$$v_{c10} = v_{10} - v_c \quad v_{c20} = v_{20} - v_c \quad (2)$$

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^[187]式 8), 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

$$v_1 = v_c + (-v_{c10}) = 2v_c - v_{10} \quad v_2 = v_c + (-v_{c20}) = 2v_c - v_{20} \quad (3)$$

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

$$\begin{aligned} v_1 &= v_c + (-\alpha v_{c10}) = (1 + \alpha)v_c - \alpha v_{10} \\ v_2 &= v_c + (-\alpha v_{c20}) = (1 + \alpha)v_c - \alpha v_{20} \end{aligned} \quad (4)$$

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^[187]式 8), 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

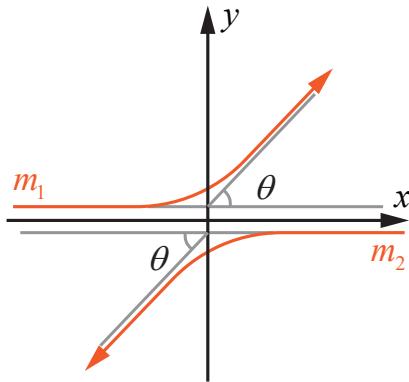


图 1: 质心系中的二体碰撞

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

$$\mathbf{v}_{c10} = (v_{10} - v_c) \hat{\mathbf{x}} \quad \mathbf{v}_{c20} = (-v_{20} - v_c) \hat{\mathbf{x}} \quad (5)$$

其中质心速度为

$$\mathbf{v}_c = (m_1 v_1 + m_2 v_2) \hat{\mathbf{x}} / (m_1 + m_2) \quad (6)$$

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

$$\mathbf{v}_{c1} = (v_{10} - v_c)(\hat{\mathbf{x}} \cos \theta + \hat{\mathbf{y}} \sin \theta) \quad \mathbf{v}_{c2} = -(v_{20} + v_c)(\hat{\mathbf{x}} \cos \theta + \hat{\mathbf{y}} \sin \theta) \quad (7)$$

最后再变换到实验系, 得末速度分别为

$$\mathbf{v}_1 = \mathbf{v}_c + \mathbf{v}_{c1} \quad \mathbf{v}_2 = \mathbf{v}_c + \mathbf{v}_{c2} \quad (8)$$

质点系的动量

预备知识 质心系^[187]

质点系的总动量为

$$\mathbf{P} = \sum_i m_i \mathbf{v}_i = \sum_i m_i (\mathbf{v}_c + \mathbf{v}_{ci}) = M \mathbf{v}_c + \sum_i m_i \mathbf{v}_{ci} \quad (1)$$

其中 M 为质点系的总质量. 注意到质心系中的动量为零, $\sum_i m_i \mathbf{v}_{ci} = \mathbf{0}$

$$\mathbf{P} = M \mathbf{v}_c \quad (2)$$

所以质点系的动量等于质心的动量, 即总质量乘以质心的速度.

动量定理 动量守恒

预备知识 动量 动量定理（单个质点）^[170], 质点系^[187]

结论

系统总动量的变化率等于合外力, 所以合外力为零时系统总动量守恒.

推导

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^[170],

$$\frac{d}{dt} \mathbf{p}_i = \mathbf{F}_i^{in} + \mathbf{F}_i^{out} \quad (1)$$

总动量的变化率为

$$\frac{d\mathbf{P}}{dt} = \sum_i \frac{d}{dt} \mathbf{p}_i = \sum_i \mathbf{F}_i^{in} + \sum_i \mathbf{F}_i^{out} \quad (2)$$

由“质点系^[187]”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

$$\frac{d\mathbf{P}}{dt} = \sum_i \mathbf{F}_i^{out} \quad (3)$$

质点系的动能 柯尼西定理

预备知识 质点系^[187], 质点系的动量^[194]

柯尼西定理

某参考系中质点系的动能等于该参考系中其的质心的动能加上质心系中质点系的动能, 即

$$E_k = \frac{1}{2}Mv_c^2 + \frac{1}{2} \sum_i m_i v_{ci}^2 \quad (1)$$

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

$$E_k = \frac{1}{2}m\omega^2 R^2 + \frac{1}{2} \sum_i m_i \omega^2 R^2 = m\omega^2 R^2 \quad (2)$$

柯尼西定理的证明

在当前参考系中, 第 i 个质点的运动速度为

$$\mathbf{v}_i = \mathbf{v}_c + \mathbf{v}_{ci} \quad (3)$$

于是有

$$\begin{aligned} E_k &= \frac{1}{2} \sum_i m_i \mathbf{v}_i^2 = \frac{1}{2} \sum_i m_i (\mathbf{v}_c + \mathbf{v}_{ci})^2 \\ &= \frac{1}{2} \sum_i m_i \mathbf{v}_c^2 + \frac{1}{2} \sum_i m_i \mathbf{v}_{ci}^2 + \sum_i m_i \mathbf{v}_c \cdot \mathbf{v}_{ci} \end{aligned} \quad (4)$$

现在只需证明 $\sum_i m_i \mathbf{v}_c \cdot \mathbf{v}_{ci} = 0$ 即可。考虑到

$$\sum_i m_i \mathbf{v}_c \cdot \mathbf{v}_{ci} = \mathbf{v}_c \cdot \sum_i m_i \mathbf{v}_{ci} \quad (5)$$

而质心系中的质点系动量为零（式 8^[187]），所以

$$\sum_i m_i \mathbf{v}_{ci} = \mathbf{0} \quad (6)$$

证毕。

力矩

预备知识 矢量的叉乘^[96]

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），就可以计算物体受到的力矩。

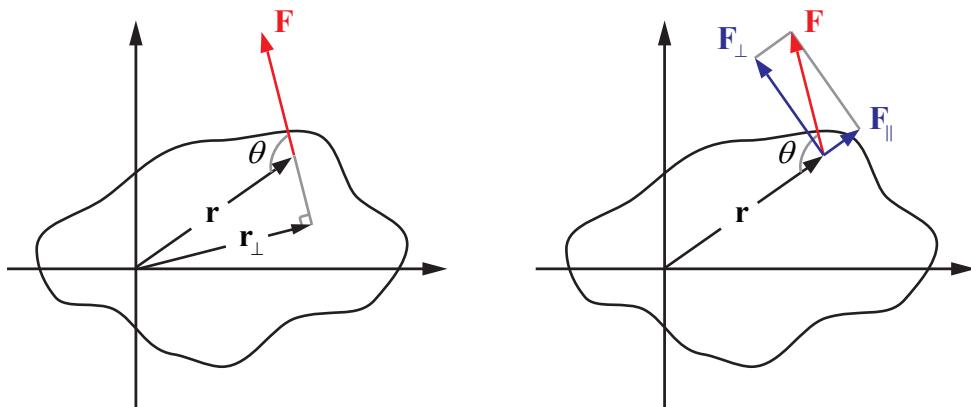


图 1：力矩的两种几何理解

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 M 表示）为

$$M = |\mathbf{r}_\perp| |\mathbf{F}| = |\mathbf{r}| |\mathbf{F}| \sin \theta \quad (1)$$

其中 θ 是 \mathbf{r} 与 \mathbf{F} 的夹角 ($\theta < \pi$). 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

$$M = |\mathbf{r}| |\mathbf{F}_\perp| = |\mathbf{r}| |\mathbf{F}| \sin \theta \quad (2)$$

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. 这种定义把力矩假定为一种标量. 另一种是根据叉乘的定义^[96], 规定力矩为矢量, 且

$$\mathbf{M} = \mathbf{r} \times \mathbf{F} \quad (3)$$

显然, 由这种定义, 力矩大小还是 $M = |\mathbf{r}| |\mathbf{F}| \sin \theta$, 但是逆时针力矩的方向垂直纸面指向读者, 顺时针则相反.

空间力矩

若物体受的若干个受力, 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, 且仍然定义为式 3. 总力矩等于每个力所产生的力矩的矢量叠加.

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.

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

$$\mathbf{M}_A = \sum_i \mathbf{r}_{Ai} \times \mathbf{F}_i = \mathbf{0} \quad (4)$$

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

$$\begin{aligned} \mathbf{M}_B &= \sum_i (\mathbf{r}_{Ai} + \mathbf{r}_{BA}) \times \mathbf{F}_i = \sum_i \mathbf{r}_{Ai} \times \mathbf{F}_i + \sum_i \mathbf{r}_{BA} \times \mathbf{F}_i \\ &= \mathbf{M}_A + \mathbf{r}_{BA} \times \sum_i \mathbf{F}_i \end{aligned} \quad (5)$$

其中最后两步使用了叉乘的分配律^[96]. 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.

拓展阅读 角动量定理角动量守恒（单个质点）[\[171\]](#)

刚体的静力平衡

预备知识 动量定理[\[195\]](#)，角动量定理[\[202\]](#)

结论

在任意参考系中，若刚体所受的合外力，合外力矩都为零，则刚体质心不动或匀速运动，且刚体没有转动或绕质心做匀速转动.

推导

把刚体看做由许多质点组成，合外力为零时刚体动量守恒，而动量等于质心的动量[\[194\]](#) $\mathbf{p}_c = M_c \mathbf{v}_c$ ，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[\[202\]](#)，而角动量等于质心的角动量 $\mathbf{L}_c = \mathbf{r}_c \times \mathbf{p}_c$ 加质心系中的角动量（式 5[\[200\]](#)）。当质心匀速运动或不动时质心的角动量不变，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.

例 1

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.

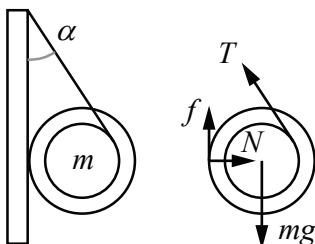


图 1: 线轴的平衡

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

$$\begin{cases} f \leq \mu N & (\text{摩擦系数}) \\ N - T \sin \alpha = 0 & (\text{水平方向受力平衡}) \\ T \cos \alpha + f - mg = 0 & (\text{竖直方向受力平衡}) \\ Tr - fR = 0 & (\text{力矩平衡}) \end{cases} \quad (1)$$

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

$$\sin \alpha \geq \frac{r}{\mu R} \quad (2)$$

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 动量的定义，矢量的叉乘^[96]

单质点的角动量

一个质点的质量为 m ，某时刻速度为 \mathbf{v} ，则其动量为 $\mathbf{p} = m\mathbf{v}$. 在三维空间中建立惯性系^[162]，原点为 O ， O 点到质点的位置矢量为 \mathbf{r} . 定义该质点关于 O 点的角动量为

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} = m\mathbf{r} \times \mathbf{v} \quad (1)$$

由叉乘的几何定义^[96] 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 $L = rp$.

角动量叠加

角动量是矢量，若把系统看做质点系，则总角动量等于所有质点的角动量矢量叠加。

$$\mathbf{L} = \sum_i \mathbf{L}_i = \sum_i \mathbf{r}_i \times \mathbf{p}_i \quad (2)$$

角动量的坐标系变换

可类比力矩的坐标系变换（式 5），坐标系 A 中总角动量为

$$\mathbf{L}_A = \sum_i \mathbf{r}_{Ai} \times \mathbf{p}_i \quad (3)$$

变换到坐标系 B 中，总角动量为

$$\mathbf{L}_B = \sum_i (\mathbf{r}_{BA} + \mathbf{r}_{Ai}) \times \mathbf{p}_i = \mathbf{r}_{BA} \times \sum_i \mathbf{p}_i + \mathbf{L}_A \quad (4)$$

角动量的分解

质心系中的角动量为

$$\mathbf{L}_0 = \sum_i \mathbf{r}_{ci} \times \mathbf{p}_i \quad (5)$$

定义质心角动量为“质心处具有系统总质量 M 的质点的角动量”（类比质心动量的定义）

$$\mathbf{L}_c = \mathbf{r}_c \times (M\mathbf{v}_c) = \mathbf{r}_c \times \mathbf{p}_c \quad (6)$$

现在我们变换到任意坐标系中，令总角动量为 \mathbf{L} ，由式 4 得

$$\mathbf{L} = \mathbf{r}_c \times \sum_i \mathbf{p}_i + \mathbf{L}_0 \quad (7)$$

由于系统总动量 $\sum_i \mathbf{p}_i$ 等于质心动量 \mathbf{p}_c ，右边第一项等于质心角动量式 6。最后得到

$$\mathbf{L} = \mathbf{L}_c + \mathbf{L}_0 \quad (8)$$

所以任何坐标系中，系统的总角动量等于质心角动量加质心系中的角动量。

角动量定理 角动量守恒

预备知识 角动量定理 角动量守恒（单个质点）[\[171\]](#)，牛顿第三定律[\[162\]](#)

角动量定理可以表示为

$$\frac{d\mathbf{L}}{dt} = \mathbf{M} \quad (1)$$

即系统的角动量对时间的变化率等于所受外力的合力矩。系统可以包括任意选定的若干物体。

推导

推导可类比动量定理[\[195\]](#)。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 个质点的角动量为 \mathbf{L}_i ，力矩为 \mathbf{M}_i ，单个质点的角动量定理[\[171\]](#) 为

$$\frac{d\mathbf{L}_i}{dt} = \mathbf{M}_i = \mathbf{M}_i^{in} + \mathbf{M}_i^{out} \quad (2)$$

其中 \mathbf{M}_i^{in} 和 \mathbf{M}_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

$$\frac{d\mathbf{L}}{dt} = \sum_i \frac{d\mathbf{L}_i}{dt} = \sum_i \mathbf{M}_i^{in} + \sum_i \mathbf{M}_i^{out} \quad (3)$$

现在我们只需证明质点系的合内力矩为零即可

$$\sum_i \mathbf{M}_i^{in} = \sum_i \mathbf{r}_i \times \sum_j \mathbf{F}_{j \rightarrow i} = \sum_{i,j} \mathbf{r}_i \times \mathbf{F}_{j \rightarrow i} \quad (4)$$

其中 $\mathbf{F}_{j \rightarrow i}$ 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

$$\mathbf{r}_k \times \mathbf{F}_{l \rightarrow k} + \mathbf{r}_l \times \mathbf{F}_{k \rightarrow l} \equiv \mathbf{M}_{l \rightarrow k} + \mathbf{M}_{k \rightarrow l} \quad (5)$$

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 1，根据定义，力矩的大小等于力力的模长乘以力臂的长度[\[197\]](#)，而一对相互作用力的大小相同，又由于二者共线，力臂也重

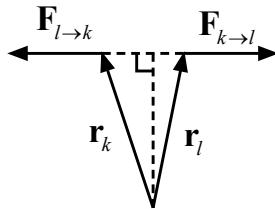


图 1: 两质点的相互作用力对总力矩贡献为零

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.

再看代数的方法: 我们先沿着两质点的连线写出相互作用力 $\mathbf{F}_{l \rightarrow k} = \alpha(\mathbf{r}_k - \mathbf{r}_l)$, $\mathbf{F}_{k \rightarrow l} = \alpha(\mathbf{r}_l - \mathbf{r}_k)$, 直接计算总角动量得

$$\mathbf{r}_k \times (\mathbf{r}_k - \mathbf{r}_l)\alpha + \mathbf{r}_l \times (\mathbf{r}_l - \mathbf{r}_k)\alpha = 0 \quad (6)$$

刚体的绕轴转动 转动惯量

预备知识 刚体^[189], 角动量定理^[202]

设刚体绕光滑轴转动. 这里令轴的方向为 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, 角动量的 z 分量守恒.

对于单个质点, $L_z = (\mathbf{r} \times \mathbf{p}) \cdot \hat{\mathbf{z}}$. 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

$$L_z = |\mathbf{r}_\perp| |\mathbf{p}| = m r_\perp v = m r_\perp^2 \omega \quad (1)$$

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

$$L_z = \omega \sum_i m_i r_{\perp i}^2 \quad (2)$$

用积分写成

$$L_z = \omega \int r_\perp^2 dm = \omega \int r_\perp^2 \rho dV \quad (3)$$

定义刚体的绕轴转动惯量为

$$I = \int r_{\perp}^2 dm \quad (4)$$

(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为

$$L_z = I\omega \quad (5)$$

现在来看“角动量定理^[202]”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

$$\frac{dL_z}{dt} = M_z \quad (6)$$

将式 5 代入式 6, 并利用角加速度的定义得

$$I\alpha = M_z \quad (7)$$

这就是刚体绕轴转动的动力学方程, 其形式可类比牛顿第二定律^[162].

例 1 刚体摆

如图 1, 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.

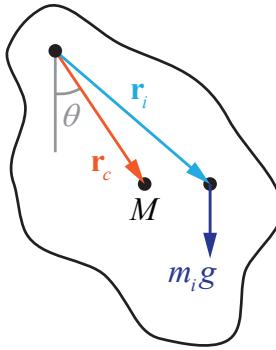


图 1: 刚体摆

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)

$$\begin{aligned} \mathbf{M} &= \sum_i \mathbf{r}_i \times (m_i \mathbf{g}) = \left(\sum_i m_i \mathbf{r}_i \right) \times \mathbf{g} = M \mathbf{r}_c \times \mathbf{g} \\ &= M g r_c \sin \theta \end{aligned} \quad (8)$$

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 7 得刚体摆的运动方程为

$$I\ddot{\theta} = Mgr_c \sin \theta \quad (9)$$

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^[175]。

平行轴定理与垂直轴定理

预备知识 转动惯量^[203]

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

$$I = I_0 + MR^2 \quad (1)$$

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

$$I = \sum_i m_i (\mathbf{R} + \mathbf{r}_i)^2 = R^2 \sum_i m_i + \sum_i m_i r_i^2 + 2\mathbf{r}_0 \cdot \sum_i m_i \mathbf{r}_i \quad (2)$$

由于质心轴经过刚体的质心，上式最后一项中的求和为零，而上式右边第二项恰好是 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

$$I = I_x + I_y \quad (3)$$

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 $(x_i, y_i, 0)$

$$I = \sum_i m_i(x_i^2 + y_i^2) = \sum_i m_i x_i^2 + \sum_i m_i y_i^2 = I_x + I_y \quad (4)$$

证毕。

常见几何体的转动惯量

预备知识 转动惯量^[203]

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

$$I = \sum_i m_i R^2 = MR^2 \quad (1)$$

细棒（端点轴）

细棒的线密度为 $\lambda = M/L$ ，如果划分成长度为 dr 的小段，第 i 段距离转轴 r_i

细棒（中心轴）薄长方体（共面轴）

细棒（中心轴）可以看做两个等质量的细棒（端点轴），质量分都为 M_1 ，每个具有转动惯量 $M_1 R^2/3$ ，总转动惯量为 $2M_1 R^2/3 = MR^2/3$ 。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）可以看成许多许多细棒（中心轴）组成，所以转动惯量的系数仍然为 $1/3$ 。注意一些教材中使用细棒的总长度 $L = 2R$ ，则转动惯量为 $ML^2/12$ 。

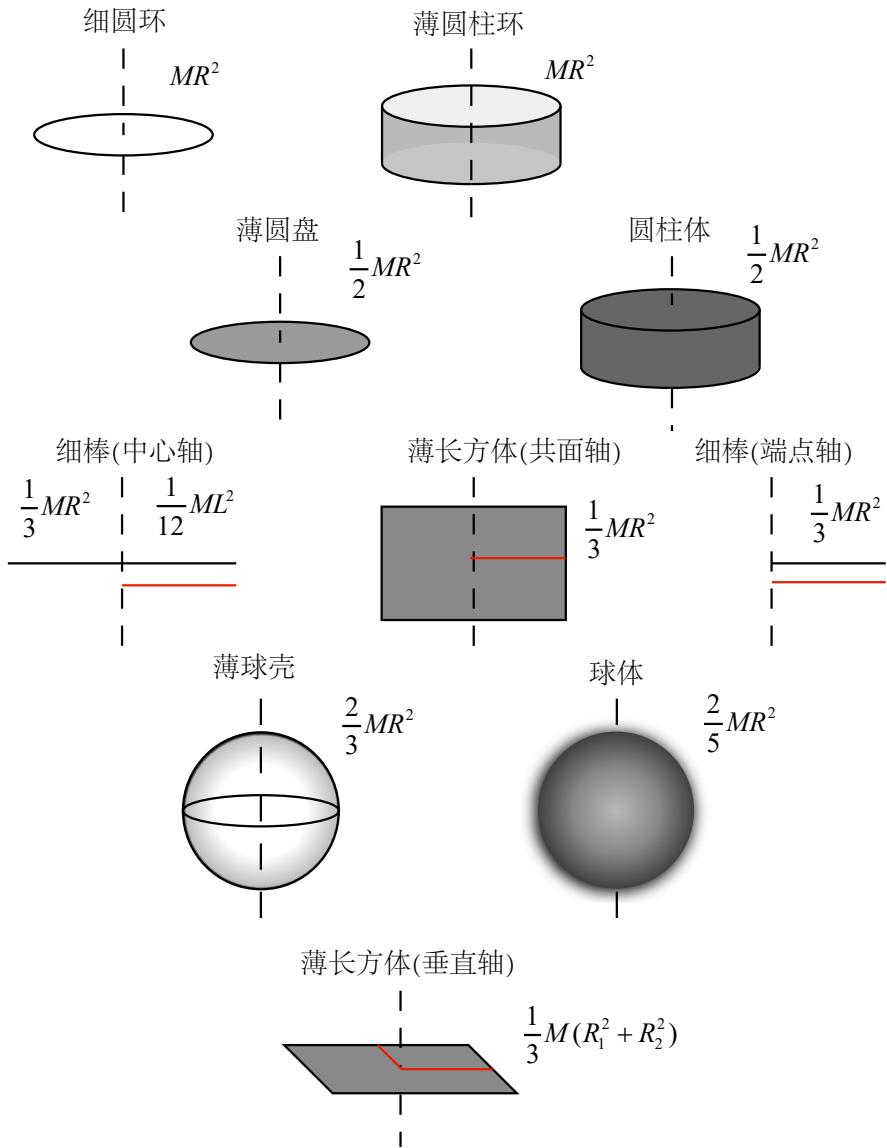


图 1: 常见几何体的转动惯量, 虚线为转轴, 物体质量 M 均匀分布, R 为几何体的半径或红线标注的长度.

薄圆盘 圆柱

薄圆盘可以看做许多宽度为 dr 的细圆环组成¹, 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 $2\pi r_i dr$, 总转动惯量为

$$I = \sum_i r_i^2 dm_i = \sum_i r_i^2 \cdot \sigma \cdot 2\pi r_i dr = 2\pi\sigma \sum_i r_i^3 dr = 2\pi\sigma \int_0^R r^3 dr \quad (2)$$

也可以在极坐标中直接根据定义写出积分

$$I = \int r^2 \sigma ds = \int_0^{2\pi} \int_0^R \sigma r^2 \cdot r dr d\theta = 2\pi\sigma \int_0^R r^3 dr = \frac{1}{2}\sigma\pi R^2 R^2 = \frac{1}{2}MR^2 \quad (3)$$

圆柱可看做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 $R d\theta$, 面积为 $ds_i = 2\pi R \sin \theta_i \cdot R d\theta$, 半径为 $r_i = R \sin \theta_i$, 总转动惯量为

$$\begin{aligned} I &= \sum_i r_i^2 dm_i = \sum_i R^2 \sin^2 \theta_i \cdot \sigma \cdot 2\pi R \sin \theta_i \cdot R d\theta \\ &= 2\pi\sigma R^4 \sum_i \sin^3 \theta_i d\theta = 2\pi\sigma R^4 \int_0^\pi \sin^3 \theta d\theta \end{aligned} \quad (4)$$

也可以在球坐标中直接写出球面积分

$$\begin{aligned} I &= \int r^2 \sigma ds = \int_0^{2\pi} \int_0^\pi (R \sin \theta)^2 \sigma R^2 \sin \theta d\theta d\phi = 2\pi\sigma R^4 \int_0^\pi \sin^3 \theta d\theta \\ &= 2\pi\sigma R^4 \int_{-1}^1 (1 - \cos^2 \theta) d\cos \theta = \frac{2}{3}(\sigma 4\pi R^2) R^2 = \frac{2}{3}MR^2 \end{aligned} \quad (5)$$

其中对 θ 的积分使用了换元积分法.

¹然而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_i ，厚度为 dr ，体积为 $4\pi r_i^2 dr$ ，总转动惯量为

$$I = \sum_i \frac{2}{3} m_i r_i^2 = \frac{2}{3} \sum_i \rho V_i r_i^2 = \frac{2M}{R^3} \sum_i r_i^4 dr = \frac{2M}{R^3} \int_0^{+\infty} r^4 dr \quad (6)$$

也可以在球坐标中直接体积分

$$\begin{aligned} I &= \int (r \sin \theta)^2 dm = \int_0^{2\pi} \int_0^\pi \int_0^R (r \sin \theta)^2 \sigma r^2 \sin \theta dr d\theta d\phi \\ &= \frac{3M}{2R^3} \int_0^\pi \sin^3 \theta d\theta \int_0^R r^4 dr = \frac{2M}{R^3} \int_0^R r^4 dr = \frac{2}{5} MR^2 \end{aligned} \quad (7)$$

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 $MR_1^2/3$ 和 $MR_2^2/3$ ，使用平行轴定理可得关于垂直轴的转动惯量为二者之和

$$I = \frac{1}{3}(MR_1^2 + MR_2^2) \quad (8)$$

刚体的平面运动方程

预备知识 动量定理^[195]，角动量定理^[202]

结论

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

$$M\mathbf{a}_c = \sum_i \mathbf{F}_i \quad (1)$$

$$I\boldsymbol{\alpha} = \sum_i (\mathbf{r}_i - \mathbf{r}_c) \times \mathbf{F}_i \quad (2)$$

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，在任意惯性系中，由动量定理，刚体总动量，即质心动量 \mathbf{p}_c 的变化率为

$$\frac{d}{dt} \mathbf{p}_c = M \mathbf{a}_c = \sum_i \mathbf{F}_i \quad (3)$$

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 $\mathbf{L}_c = I\boldsymbol{\omega}$ ，角动量变化率为²

$$\frac{d\mathbf{L}_c}{dt} = I \frac{d\boldsymbol{\omega}}{dt} = I\boldsymbol{\alpha} \quad (4)$$

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 $-m_i \mathbf{a}_c$ 产生的合力矩为零

$$\sum_i \mathbf{r}_{ci} \times (-m_i \mathbf{a}_c) = \mathbf{a}_c \times \sum_i m_i \mathbf{r}_{ci} = \mathbf{0} \quad (5)$$

现在我们可以继续角动量定理^[202] 得

$$I\boldsymbol{\alpha} = \sum_i \mathbf{r}_{ci} \times \mathbf{F}_i \quad (6)$$

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

$$I\boldsymbol{\alpha} = \sum_i (\mathbf{r}_i - \mathbf{r}_c) \times \mathbf{F}_i \quad (7)$$

²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。

例 1

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^[203]”中例 1 可知细棒绕其质心做单摆运动.

浮力

等效法

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

$$F = \rho_0 V_0 g \quad (1)$$

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.

散度法

预备知识 散度定理^[143]

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

$$P = -\rho_0 g z \quad (2)$$

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

$$\Delta \mathbf{F}_i = -P \Delta \mathbf{S}_i = \rho_0 g z \Delta \mathbf{S}_i \quad (3)$$

现在把所有面元所受的压力求和，并用曲面积分表示为

$$\mathbf{F} = \oint \rho_0 g z \, d\mathbf{S} \quad (4)$$

这就是物体所受的浮力。我们先计算其 z 分量，等式两边点乘 $\hat{\mathbf{z}}$ 得

$$F_z = \oint (\rho_0 g z \hat{\mathbf{z}}) \cdot d\mathbf{S} \quad (5)$$

我们可以把括号内的矢量看做一个矢量场 \mathbf{A} ，其散度为

$$\nabla \cdot \mathbf{A} = \frac{\partial}{\partial z} (\rho_0 g z) = \rho_0 g \quad (6)$$

对上式应用散度定理^[143]，得

$$F_z = \int \nabla \cdot \mathbf{A} \, dV = \rho_0 g V_0 \quad (7)$$

再来计算 x 和 y 分量的浮力，由于

$$\nabla \cdot (\rho_0 g z \hat{\mathbf{x}}) = \nabla \cdot (\rho_0 g z \hat{\mathbf{y}}) = 0 \quad (8)$$

两个水平分量为零。可见该结论与“等效法”中得出的一致。

第三章

振动与波动

振动的指数形式

预备知识 简谐振子^[172], 二阶常系数齐次微分方程^[73]

简谐振子的微分方程

$$m\ddot{x} = -kx \quad (1)$$

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

$$x(t) = C_1 e^{i\omega t} + C_2 e^{-i\omega t} \quad (2)$$

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

$$\tilde{x}(t) = \tilde{A} e^{-i\omega t} \quad (3)$$

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 $\tilde{A} = A e^{-i\varphi_0}$, 代入式 3, 再取实部得

$$x(t) = \operatorname{Re}[\tilde{x}(t)] = A \operatorname{Re}[e^{-i(\omega t + \varphi_0)}] = A \cos(\omega t + \varphi_0) \quad (4)$$

振动的叠加

这里举一个例子说明使用指数函数比三角函数方便. 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

$$\begin{aligned} \sum_i x_i(t) &= \sum_i [A_i \cos \varphi_{0i} \cos(\omega t) - A_i \sin \varphi_{0i} \sin(\omega t)] \\ &= \sum_i (A_i \cos \varphi_{0i}) \cos(\omega t) - \sum_i (A_i \sin \varphi_{0i}) \sin(\omega t) \end{aligned} \quad (5)$$

¹ 式 3 中 $e^{-i\omega t}$ 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 $\varphi_0 = \arg(\tilde{A})$

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

$$\sum_i x_i(t) = A[\cos \varphi_0 \cos(\omega t) - \sin \varphi_0 \sin(\omega t)] = A \cos(\omega t + \varphi_0) \quad (6)$$

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 $\tilde{x}_i(t) = \tilde{A}_i e^{-i\omega t}$, 其中 $\tilde{A}_i = A_i e^{-i\omega \varphi_{0i}}$. 求和得

$$\sum_i \tilde{x}_i(t) = \left(\sum_i \tilde{A}_i \right) e^{-i\omega t} \quad (7)$$

令 $\tilde{A} = \sum_i \tilde{A}_i$, $A = |\tilde{A}|$, $\varphi_0 = -\arg(\tilde{A})$, 则最后结果为

$$\sum_i \tilde{x}_i(t) = \tilde{A} e^{-i\omega t} \quad (8)$$

$$\sum_i x_i(t) = \operatorname{Re} [\tilde{A} e^{-i\omega t}] = A \cos(\omega t + \varphi_0) \quad (9)$$

不难证明上式的 A 和 φ_0 与式 6 得到的 A 和 φ_0 相同, 但是这里的推导的过程却更为简洁.

受阻简谐振子

预备知识 简谐振子^[172], 二阶常系数齐次微分方程的^[75]

结论

两个复数根 ($\alpha^2 - 4km < 0$, 最常讨论和应用的情况)

$$y = C_1 e^{rx} \cos(\omega x + C_2) \quad (1)$$

其中

$$r = -\frac{\alpha}{2m} \quad \omega = \sqrt{\omega_0^2 - \frac{\alpha^2}{4m^2}} \quad (2)$$

其中 $\omega_0 = \sqrt{k/m}$ 为固有振动频率 ($\alpha = 0$ 时的振动频率) 为, r 和 ω 满足

$$r^2 + \omega^2 = \omega_0^2 \quad (3)$$

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 ($\alpha \neq 0$).

$$my'' = -\alpha y' - ky \quad (4)$$

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^[75]，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

$$y = C_1 e^{r_1 x} + C_2 e^{r_2 x} \quad (5)$$

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

$$y = C_1 e^{rx} + C_2 x e^{rx} \quad (6)$$

这是质点振动与不震动的临界点。

3. 两个复数根 r_1, r_2 ($\alpha^2 - 4km < 0$ ，最常讨论和应用的情况)

$$y = C_1 e^{rx} \cos(\omega x + C_2) \quad (7)$$

其中

$$r = -\frac{\alpha}{2m} \quad \omega = \frac{1}{2m} \sqrt{4mk - \alpha^2} = \sqrt{\omega_0^2 - \frac{\alpha^2}{4m^2}} \quad (8)$$

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^[172]的振动频率）

$$\omega_0 = \sqrt{\frac{k}{m}}, \quad \gamma = \frac{\alpha}{2\sqrt{mk}} \quad (9)$$

则

$$r = -\omega_0 \gamma, \quad \omega = \omega_0 \sqrt{1 - \gamma^2} \quad (10)$$

满足 $r^2 + \omega^2 = \omega_0^2$.

简谐振子受迫运动

预备知识 受阻简谐振子^[215], 振动的指数形式^[214]

在受阻弹簧振子的基础上, 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.

$$my'' = -\alpha y' - ky + f(t) \quad (1)$$

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

$$\tilde{y}(t) = \tilde{A}e^{-i\omega t} \quad \tilde{f}(t) = \tilde{B}e^{-i\omega t} \quad (2)$$

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

$$\tilde{A} = \frac{\tilde{B}}{k - \omega^2 m - i\alpha\omega} \quad (3)$$

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

$$\tilde{A} = \frac{\tilde{B}}{m(\omega_0^2 - \omega^2) - i\alpha\omega} \quad (4)$$

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

$$A = \frac{B}{\sqrt{(\omega^2 - \omega_0^2)^2 m^2 + \alpha^2 \omega^2}} \quad (5)$$

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)

$$\varphi_0 = -\arg \tilde{A} = \arctan 2[m(\omega_0^2 - \omega^2), -\alpha\omega] \quad (6)$$

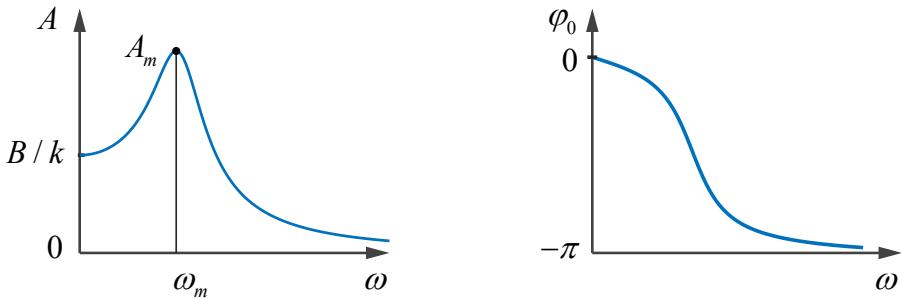


图 1: 幅频曲线和相频曲线

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

$$\omega_m = \sqrt{\omega_0^2 - \frac{\alpha^2}{2m^2}} \quad (7)$$

幅频曲线可改写为

$$A = \frac{B}{m\sqrt{(\omega^2 - \omega_m^2)^2 + (\omega_0^4 - \omega_m^4)}} \quad (8)$$

所以 A 的最大值为

$$A_m = \frac{B}{m\sqrt{\omega_0^4 - \omega_m^4}} \quad (9)$$

由式 8 可得，当 $\omega = 0$ 和 $\omega \rightarrow +\infty$ 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 $\omega \rightarrow +\infty$ 时相位恰好落后 π 。

平面波

预备知识 矢量点乘^[91]，简谐振子^[172]

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，任何时刻 t ，弦的波函数（即形状）可以用 $y(x, t)$ 来描述。若

$$y(x, t) = A \cos(kx - \omega t + \varphi_0) \quad (1)$$

则我们把这个波函数称为平面波，如图 1 所示⁴.

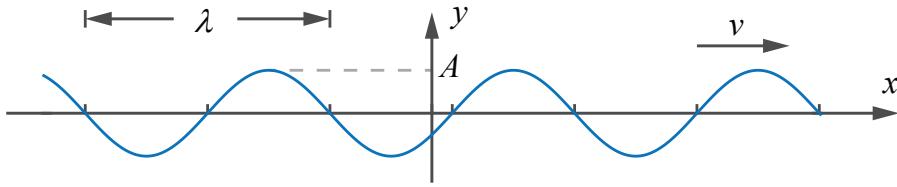


图 1: 平面波

我们定义图中的 A 为振幅，定义一个周期⁵为波长，记为 λ . 与波长一一对应的一个量是式 1 中的 k ，称为波数. 波长与波数的关系可以类比简谐振子^[172]的角频率 ω 与周期 T 的关系，即

$$k = \frac{2\pi}{\lambda} \quad (2)$$

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⁶ $-kx - \varphi_0$.

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 v . 把式 1 稍作整理得

$$y(x, t) = A \cos \left[k \left(x - \frac{\omega}{k} t \right) + \varphi_0 \right] \quad (3)$$

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

$$v = \frac{\omega}{k} \quad (4)$$

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

$$v = \frac{\lambda}{T} \quad (5)$$

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

$$v = \lambda f \quad (6)$$

⁴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.

⁵不是时间周期而是空间周期

⁶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.

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，只是因变量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 ）

$$\xi = A \cos(kx - \omega t + \varphi_0) \quad (7)$$

二维和三维的平面波

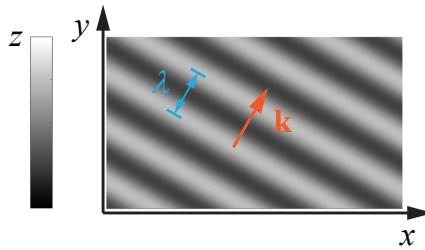


图 2: 二维平面波

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 \mathbf{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。于是我们可以用点乘来表示相位随空间的变化

$$\Delta\varphi = \mathbf{k} \cdot \Delta\mathbf{r} = k_x \Delta x + k_y \Delta y \quad (8)$$

于是我们可以写出波函数为

$$z = A \cos(\mathbf{k} \cdot \mathbf{r} - \omega t + \varphi_0) \quad (9)$$

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

$$\mathbf{s} = \mathbf{A} \cos(\mathbf{k} \cdot \mathbf{r} - \omega t + \varphi_0) \quad (10)$$

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

预备知识 振动的指数形式^[214]

用复数表示波函数, 往往可以化简书写和计算. 类比式 3^[214], 我们可以把平面波表示为指数形式⁷

$$\tilde{s} = \mathbf{A} e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t + \varphi_0)} \quad (11)$$

注意只有实部表示质点的位移, 虚部无物理意义.

一维波动方程

预备知识 简谐振子^[172], 导数与差分^[53], 平面波^[218]

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**.

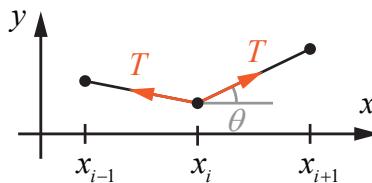


图 1: “微元法”分析弦的波动

我们用“微元法”的思想, 把弦划分为若干小线段(每个线段不弯曲), 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

⁷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.

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

$$\begin{cases} F_x = T \cos \theta \approx T \\ F_y = T \sin \theta \approx T \tan \theta = T \frac{y(x_i + h) - y(x_i)}{h} \end{cases} \quad (1)$$

同理, 来自 m_{i-1} 的拉力的两个分量为

$$\begin{cases} F'_x \approx -T \\ F'_y \approx T \frac{y(x_i - h) - y(x_i)}{h} \end{cases} \quad (2)$$

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

$$F_y = T \frac{y(x_i + h) - 2y(x_i) + y(x_i - h)}{h} \quad (3)$$

结合牛顿定律, 并代入 $m_i = \lambda h$, 有

$$\lambda \frac{\partial^2 y}{\partial t^2} = T \frac{y(x_i + h) - 2y(x_i) + y(x_i - h)}{h^2} \quad (4)$$

注意到 h 很小, 由式 5^[53] 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

$$\frac{\partial^2 y}{\partial x^2} - \frac{\lambda}{T} \frac{\partial^2 y}{\partial t^2} = 0 \quad (5)$$

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，我们理可以列出 m_i 的运动方程为

$$\lambda h \frac{\partial^2 \xi}{\partial t^2} = F_i = \frac{k}{h} [\xi(x_i + h) - 2\xi(x_i) + \xi(x_i - h)] \quad (6)$$

同样利用式 5^[53] 得波动方程为

$$\frac{\partial^2 \xi}{\partial x^2} - \frac{\lambda}{k} \frac{\partial^2 \xi}{\partial t^2} \quad (7)$$

方程的解

第四章

天体运动与中心力场

万有引力 引力势能

预备知识 牛顿运动定律^[162]

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

$$\mathbf{F}_{12} = -\frac{Gm_1m_2}{r_{12}^2}\hat{\mathbf{r}}_{12} = -\frac{Gm_1m_2}{|\mathbf{r}_2 - \mathbf{r}_1|^3}(\mathbf{r}_2 - \mathbf{r}_1) \quad (1)$$

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

$$\mathbf{g}(\mathbf{r}) = -\frac{Gm_1}{|\mathbf{r} - \mathbf{r}_1|^3}(\mathbf{r} - \mathbf{r}_1) \quad (2)$$

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^[166], 球坐标系中的梯度算符

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^[166]. 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^[164]可以用线积分^[137]表示 (以 r 作为参数)

$$W = \int_{\mathcal{L}} \mathbf{F}(\mathbf{r}) \cdot d\mathbf{r} = \int_{r_1}^{r_2} F(r) dr \quad (3)$$

其中第二步是因为

$$\mathbf{F}(\mathbf{r}) \cdot d\mathbf{r} = F(r)\hat{\mathbf{r}} \cdot d\mathbf{r} = F(r)\hat{\mathbf{r}} \cdot (dr\hat{\mathbf{r}} + r d\theta \hat{\boldsymbol{\theta}}) = F(r) dr \quad (4)$$

显然线积分的结果至于初末位置（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^[166] 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

$$V(\mathbf{r}_1) - V(\mathbf{r}_2) = \int_{r_1}^{r_2} F(r) dr = -GMm \int_{r_1}^{r_2} \frac{1}{r^2} dr = -GMm \left(\frac{1}{r_1} - \frac{1}{r_2} \right) \quad (5)$$

可见任意位置的势能函数可以取

$$V(\mathbf{r}) = V(r) = -\frac{GMm}{r} \quad (6)$$

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

$$V = - \int F(r) dr \quad (7)$$

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得

$$\mathbf{F} = \nabla V = GMm \left(\hat{\mathbf{r}} \frac{\partial}{\partial r} + \hat{\theta} \frac{1}{r} \frac{\partial}{\partial \theta} + \hat{\phi} \frac{1}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \frac{1}{r} = -\frac{GMm}{r^2} \hat{\mathbf{r}} \quad (8)$$

也可以使用直角坐标的梯度

$$\mathbf{F} = \nabla V = GMm \left(\hat{\mathbf{x}} \frac{\partial}{\partial x} + \hat{\mathbf{y}} \frac{\partial}{\partial y} + \hat{\mathbf{z}} \frac{\partial}{\partial z} \right) \frac{1}{\sqrt{x^2 + y^2 + z^2}} \quad (9)$$

以 x 分量为例

$$\frac{\partial}{\partial x} (x^2 + y^2 + z^2)^{-1/2} = -x(x^2 + y^2 + z^2)^{-3/2} = -\frac{x}{r^3} \quad (10)$$

另外两个分量类似可得 $-y/r^3$ 和 $-z/r^3$ ，代入式 9 得

$$\mathbf{F} = \nabla V = -\frac{GMm}{r^3} (x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}) = -\frac{GMm}{r^2} \frac{\mathbf{r}}{r} = -\frac{GMm}{r^2} \hat{\mathbf{r}} \quad (11)$$

开普勒三定律

预备知识 椭圆^[16]

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^[162] 和万有引力定律^[225]，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。

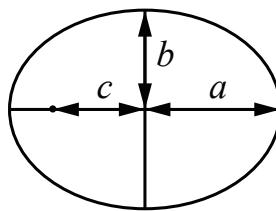


图 1: 开普勒第一定律

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。

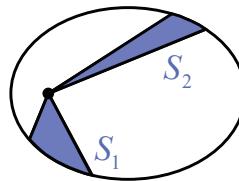


图 2: 开普勒第二定律

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^[284]
- 开普勒第一定律的证明^[231]

- 开普勒第二定律的证明^[233]
- 开普勒第三定律的证明^[234]

开普勒问题

预备知识 椭圆的三种定义^[16], 双曲线的三种定义^[18], 机械能守恒(单个质点)^[170], 角动量守恒(单个质点)^[171], 万有引力^[225], 开普勒三定律^[226]

结论

若已知中心天体和行星的质量, 假设中心天体不动, 能量和角动量可以唯一地确定行星轨道. 当能量小于零时轨道为椭圆 ($e < 1$)

$$E = -\frac{GMm}{2a} \quad L = bm\sqrt{\frac{GM}{a}} \quad (1)$$

能量等于零时轨道为抛物线 ($e = 1$), 焦距与角动量的关系为 (p 为焦距)

$$L = m\sqrt{2GMp} \quad (2)$$

能量大于零时轨道为双曲线 ($e > 1$).

$$E = \frac{GMm}{2a} \quad L = b\sqrt{2E/m} \quad (3)$$

根据圆锥曲线的性质, 对椭圆有 $a^2 = b^2 + c^2$, 双曲线有 $c^2 = a^2 + b^2$ 现在我们假设已知轨道是椭圆, 抛物线和双曲线的一种(见开普勒第一定律的证明^[231]), 证明以上关系

椭圆轨道

令椭圆轨道距离焦点的最近和最远距离分别为 r_1 和 r_2 , 总能量(动能加势能)守恒

$$\frac{1}{2}mv_1^2 - \frac{GMm}{r_1} = \frac{1}{2}mv_2^2 - \frac{GMm}{r_2} \quad (4)$$

总角动量守恒

$$mv_1r_1 = mv_2r_2 \quad (5)$$

把式 5 中的 v_2 代入式 4, 可得

$$v_1^2 = \frac{2GM}{r_1 + r_2} \frac{r_2}{r_1} \quad (6)$$

再代入式 4 的左边, 并使用 $r_1 + r_2 = 2a$ 得到总能量 $E = -GMm/(2a)$. 把式 6 代入式 5 的左边, 并使用 $r_1 r_2 = (a+c)(a-c) = b^2$ 得总角动量 $L = bm\sqrt{GM/a}$.

双曲线

注意双曲线轨道只可能是双曲线离中心天体所在焦点较近的一支, 另一支这里没有物理意义 (当引力变为斥力, 只能取双曲线的另一支)

总能量守恒

$$\frac{1}{2}mv_0^2 = \frac{1}{2}mv_1^2 - \frac{GMm}{r_1} \quad (7)$$

总角动量守恒

$$mv_0 b = mv_1 r_1 \quad (8)$$

抛物线

抛物线轨道离焦点的最近距离为焦距 p ,

反开普勒问题

预备知识 双曲线的三种等效定义 [18], 天体运动 [228]

考虑电荷为 q , 质量为 m 的轻质点从无穷远处以 v_i 入射到电荷为 Q (同种电荷) 的重质点 (假设始终静止). 轻质点的运动轨迹为双曲线的一支, 重质点的位置为离轨道较远的焦点 (证明见下文).

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 .

现在推导轨道参数 a, b 与能量 E 和角动量 L 的关系. 轨道离焦点的最近距离为

$$r_0 = a + c \quad (1)$$

对应的速度为 v_0 , 由能量守恒得

$$\frac{1}{2}mv_i^2 = \frac{1}{2}mv_0^2 + \frac{kQq}{r_0} \quad (2)$$

由角动量守恒得

$$mv_i b = mv_0 r_0 \quad (3)$$

双曲线的三个参数满足

$$a^2 + b^2 = c^2 \quad (4)$$

联立式 1 到式 4 解得

$$E = \frac{kQq}{2a} \quad L = b\sqrt{\frac{2E}{m}} = b\sqrt{\frac{kQq}{ma}} \quad (5)$$

注意若把平方反比力的系数 kQq 换成 GMm , 上式的结论与行星的双曲线轨道一致.

轨道方程推导

卢瑟福散射

预备知识 反开普勒问题^[229], 散射

看做经典散射, 微分截面等于

$$\frac{d\sigma}{d\Omega} = \frac{b \, db \, d\phi}{\sin \theta \, d\theta \, d\phi} = \frac{b}{\sin \theta} \frac{db}{d\theta} \quad (1)$$

由双曲线性质, 偏射角满足

$$\cot \frac{\theta}{2} = \frac{b}{a} \quad (2)$$

由反开普勒问题 $E = kQq/(2a)$, 消去 a 得

$$b = \frac{kQq}{2E} \cot \frac{\theta}{2} \quad (3)$$

求导代入微分截面得

$$\frac{d\sigma}{d\Omega} = \left[\frac{kQq}{4E \sin^2(\theta/2)} \right]^2 \quad (4)$$

开普勒第一定律的证明

预备知识 圆锥曲线的极坐标方程^[14], 极坐标加速度^[161], 牛顿第二定律^[162], 二阶常系数非齐次微分方程的通解^[75]

数学模型

由于太阳质量远大于其他行星, 近似认为太阳不动. 由于太阳和行星相对于行星轨道来说大小可以忽略, 把它们当做质点(另见球体的平方反比力). 以太阳为原点建立平面极坐标系, 行星在该平面上运动, 且仅受万有引力一个外力. 现证明行星的运动轨迹是椭圆, 且焦点在原点.

结论

行星轨道是以中心天体为焦点的任意圆锥曲线¹. 极坐标中, 圆锥曲线的方程^[14]为

$$r = \frac{p}{1 - e \cos \theta} \quad (1)$$

令太阳中心天体在坐标原点, 则行星沿该轨道运行.

证明

极坐标中径向和角向加速度公式分别为

$$a_r = \frac{d^2 r}{dt^2} - r \left(\frac{d\theta}{dt} \right)^2 \quad (2)$$

$$a_\theta = \frac{1}{r} \frac{d}{dt} \left(r^2 \frac{d\theta}{dt} \right) \quad (3)$$

根据牛顿第二定律和万有引力定律, 由于行星只受到沿径向的万有引力, 则有

$$ma_r = -G \frac{Mm}{r^2} \quad (4)$$

¹所以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.

$$ma_\theta = 0 \quad (5)$$

在式 4 和式 5 中同除 m , 代入式 2 和式 3 得

$$\frac{d^2r}{dt^2} - r \left(\frac{d\theta}{dt} \right)^2 = -\frac{GM}{r^2} \quad (6)$$

$$\frac{d}{dt} \left(r^2 \frac{d\theta}{dt} \right) = 0 \quad (7)$$

现在用式 6, 式 7 消去 t . 式 7 括号内部不随时间变化, 令其等于常数 h

$$r^2 \frac{d\theta}{dt} = h \quad (8)$$

其中 h 为任意常数. 我们想得到 r 关于 θ 的微分方程, 就要把所有的时间导数消去. 首先可以把 r 看做复合函数 $r[\theta(t)]$, 再用链式法则^[48] 把式 6 的第一项用 $d\theta/dt$ 表示

$$\begin{aligned} \frac{d^2r}{dt^2} &= \frac{d}{dt} \left(\frac{dr}{dt} \right) = \frac{d}{dt} \left(\frac{dr}{d\theta} \frac{d\theta}{dt} \right) = \frac{d}{d\theta} \left(\frac{dr}{d\theta} \right) \left(\frac{d\theta}{dt} \right)^2 + \frac{dr}{d\theta} \frac{d^2\theta}{dt^2} \\ &= \frac{d^2r}{d\theta^2} \left(\frac{d\theta}{dt} \right)^2 + \frac{dr}{d\theta} \frac{d}{d\theta} \left(\frac{d\theta}{dt} \right) \frac{d\theta}{dt} \end{aligned} \quad (9)$$

然后把式 8 两边同除 r^2 并代入式 6 消去所有时间导数, 得到 r 关于 θ 的微分方程

$$\frac{d^2r}{d\theta^2} \left(\frac{h}{r^2} \right)^2 + \frac{dr}{d\theta} \frac{d}{d\theta} \left(\frac{h}{r^2} \right) \frac{h}{r^2} - r \left(\frac{h}{r^2} \right)^2 = -\frac{GM}{r^2} \quad (10)$$

化简得

$$\frac{h^2}{r^4} \left[\frac{d^2r}{d\theta^2} + r^2 \frac{dr}{d\theta} \frac{d}{d\theta} \left(\frac{1}{r^2} \right) - r \right] = -\frac{GM}{r^2} \quad (11)$$

这就是 r 关于 θ 的微分方程

将式 1 代入, 可验证式 1 是该方程的解. 当然, 在事先不知道轨道方程的情况下, 也可以直接解该方程. 令

$$u \equiv \frac{1}{r} \quad (12)$$

代入式 11, 得到 u 关于 θ 的微分方程

$$h^2 \left(\frac{d^2u}{d\theta^2} + u \right) = GM \quad (13)$$

这就是比耐公式，是一个二阶常系数非齐次微分方程，通解^[75] 为

$$u = \frac{1}{p} [1 - e \cos(\theta + \varphi)] \quad (14)$$

其中

$$p = \frac{h^2}{GM} \quad (15)$$

e, φ, h 为任意常数。写成关于 r 的函数，得到圆锥曲线

$$r = \frac{p}{1 - e \cos \theta} \quad (16)$$

证毕。

拓展阅读 比耐公式^[235]

开普勒第二定律的证明

预备知识 开普勒第二定律^[226]，角动量守恒（单个质点）^[171]

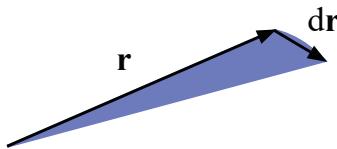


图 1: 微小时间 dt 扫过的面积

令行星的位矢为 \mathbf{r} ，在很小一段时间 dt 内移动了 $d\mathbf{r}$ ，于是扫过的面积就是以 \mathbf{r} 和 $d\mathbf{r}$ 为两条边的三角形的面积 $dS = |\mathbf{r}| |d\mathbf{r}| (\sin \theta) / 2$ ，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，则根据叉乘的定义有 $d\mathbf{S} = \mathbf{r} \times d\mathbf{r} / 2$ 。两边除以 dt ，得扫过面积的速率为

$$\frac{d\mathbf{S}}{dt} = \frac{1}{2} \mathbf{r} \times \frac{d\mathbf{r}}{dt} = \frac{1}{2} \mathbf{r} \times \mathbf{v} \quad (1)$$

由“角动量守恒^[171]”，行星角动量不变。 $\mathbf{r} \times (m\mathbf{v}) = \mathbf{L}$ （常矢量，垂直运动平面）所以 $d\mathbf{S}/dt = \mathbf{L}/2m$ 也是常矢量。写成标量形式，即

$$\frac{dS}{dt} = \frac{L}{2m} \quad (2)$$

所以从任意时刻 t_0 开始, 在一段时间 Δt 内行星扫过的面积为

$$\Delta S = \frac{L}{2m} \Delta t \quad (3)$$

只和时间间隔有关, 而与起始位置无关.

开普勒第三定律的证明

预备知识 开普勒第一定律证明[\[231\]](#), 开普勒第二定律证明[\[233\]](#)

在开普勒第一定律证明[\[231\]](#) 中, 我们得出行星轨道的极坐标方程为

$$r = \frac{p}{1 - e \cos \theta} \quad (1)$$

其中 $p = h^2/(GM)$, h 是行星的角动量与质量之比, G 是引力常数, M 是中心天体的质量. e 是圆锥曲线的离心率, 当 $0 \leq e < 1$ 时, 轨道是椭圆, 取等号时, 轨道是圆的. 椭圆的半长轴为

$$a = \frac{1}{2}[r(0) + r(\pi)] = \frac{1}{1 - e^2} \frac{h^2}{GM} \quad (2)$$

椭圆的半短轴为

$$b = a\sqrt{1 - e^2} = \frac{1}{\sqrt{1 - e^2}} \frac{h^2}{GM} \quad (3)$$

椭圆的面积为

$$S = \pi ab = \frac{\pi}{\sqrt{(1 - e^2)^3}} \frac{h^4}{(GM)^2} \quad (4)$$

由“开普勒第二定律证明[\[233\]](#)”中的结论, 单位时间扫过的面积为

$$\frac{dS}{dt} = \frac{L}{2m} = \frac{h}{2} \quad (5)$$

其中 L 是角动量, m 是行星的质量. 所以行星的周期为

$$T = \frac{S}{dS/dt} = \frac{2\pi}{\sqrt{(1 - e^2)^3}} \frac{h^3}{(GM)^2} \quad (6)$$

所以由式2与式6得 $T^2/a^3 = 4\pi^2/(GM)$ 是一个常数, 但注意这个常数与中心天体的质量有关, 所以不同中心天体的比例常数不同. 这个比例系数不用记忆, 需要用时, 只需用高中的知识计算天体圆周运动时的特例即可. 圆作为特殊的椭圆, 其半长轴就是半径.

比耐公式

预备知识 开普勒第一定律证明[\[231\]](#)

若质量为 m 的质点受有心力 $f(r)$ (规定引力为负, 斥力为正) 时, 若运动轨道用极坐标方程表示为 $r(\theta)$. 且令 $u(\theta) = 1/r(\theta)$, 且把 $f(r)$ 中的 r 替换成 u , 令其为 $f(r) = f(1/u) = F(u)$. 则有以下微分方程.

$$h^2 u^2 \left(\frac{d^2 u}{d\theta^2} + u \right) = -\frac{F(u)}{m} \quad (1)$$

这方程叫做**比耐公式 (Binet Equation)**. 其中 $h = r^2 d\theta/dt$ 是质点对力心的角动量比质量, 对一个质点的一条轨道来说, 是一个常数.

推导

在开普勒第一定律证明中, 式 (14) 列出了 u (行星到中心天体的距离的倒数) 与 θ (行星围绕中心天体转过的角度) 的微分方程.

$$h^2 u^2 \left(\frac{d^2 u}{d\theta^2} + u \right) = GM u^2 \quad (2)$$

注意等号右边是行星受到中心天体的万有引力. 一个质点受到有心力 $F(u)$, 只有两个可能的方向, 现规定向外为正, 向内为负. 所以式 2 变为

$$h^2 u^2 \left(\frac{d^2 u}{d\theta^2} + u \right) = -\frac{F(u)}{m} \quad (3)$$

事实上, 当 $F(u)$ 是其他函数时, 例如大小与 r 成正比 (橡皮绳), 经过同样的推导, 也可以得到式 3. 所以式 3 描述了一个质点受到有心力时的运动情况, 通过解这个方程, 就可以求出质点的运动轨迹.

第三部分

计算物理

第一章

Matlab 及其他工具

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，例程使用 Matlab 代码，且尽量使用简单的语法，这样即使读者使用的编程语言不是 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 Mathematica 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 “<https://www.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^[269] 和多区间二分法^[270]，用于排序的冒泡法^[273]，用于搜索多元函数局部最小值的 Nelder-Mead 算法^[275]，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^[288].

我们先用一阶微分近似的方法^{[283][284]} 举例介绍一般的思路，再介绍更精确的中点法^[289] 和四阶龙格—库塔法^[292].

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 Development 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.

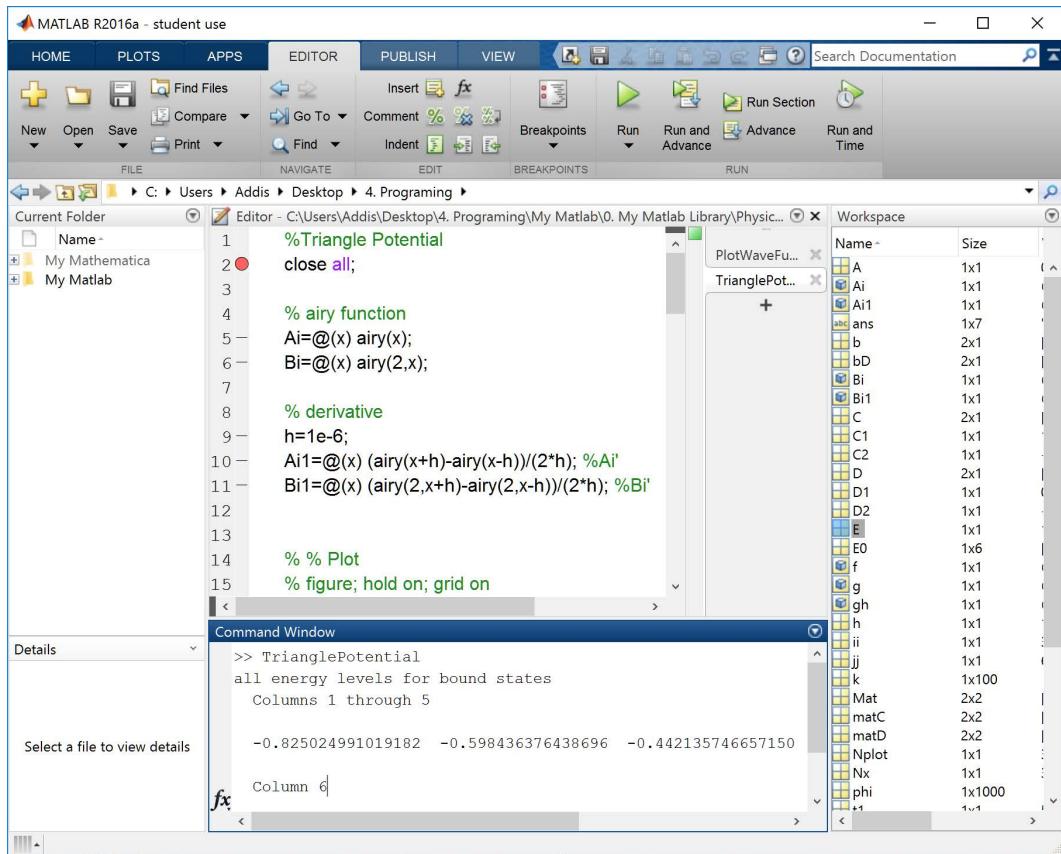


图 1: Matlab 的 IDE 界面

Editor 用于编辑代码，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⁵添加到 Matlab 的搜索路径下。

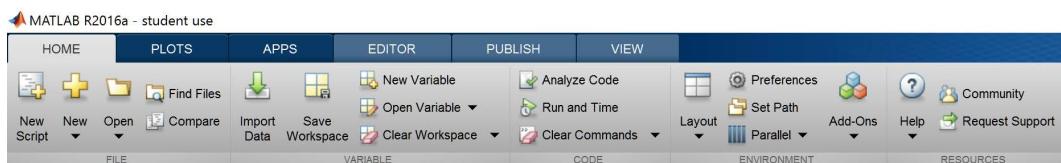


图 2: Home 菜单

⁵在英文界面下 Matlab 不能识别中文目录，建议用英文命名文件夹。

如图 2, Home 菜单中的 Set Path 按钮可以设置 Matlab 的搜索路径. 点开后用 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 littleshi.cn/PhysWiki 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^[242] 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, 注意是小写.

`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[\[239\]](#), 自然对数底[\[33\]](#), 矩阵[\[105\]](#)

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. 另外两个特殊的变量是 `pi` (圆周率) 和 `i` (虚数单位), 一般也不要对他们赋值. 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 $N \times N$ 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. 同规格的尺寸可以进行 + 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
```

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^[115], 用 “\” 算符可以解线性方程组 $\mathbf{Ax} = \mathbf{b}$. 如⁶

```
>> A = [1,2; 3,4]; det(A)
ans = -2
>> x = [-1; 1]; b = A*x
b =
    1
    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).

⁶若 $\det(\mathbf{A}) = 0$, 即方程组存在无数个解的时候, \ 算符只能得到一个特解.

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。

字符串

字符串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 $1 \times N$ 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串合并，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
```

在控制行中，输出的 logical 与 0 各占一行，这里为了节约空间将其写成一行。

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[\[242\]](#)

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）⁷。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）。

⁷ 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。

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^[239]）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

```
1 a = rand(1,1); b = 0.5;
2 if a > b % a 较大
3     disp('a is larger');
4 else % b 较大
5     disp('b is larger');
6 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

```
1 a = rand(1,1);
2 if a > 0.9
3     disp('a in (0.9, 1]');
4 elseif a > 0.6
5     disp('a in (0.6, 0.9]');
6 elseif a > 0.3
7     disp('a in (0.3, 0.6]');
8 else
9     disp('a in [0, 0.3]');
10 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

```
1 for ii = 1:3  
2     disp(['ii^2 = ' num2str(ii^2)]);  
3 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

```
1 Nx = 5;  
2 x = zeros(1, Nx); % 预赋值  
3 x(1) = 2;  
4 for ii = 2:numel(x)  
5     x(ii) = x(ii-1)^2;  
6 end  
7 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 ($a_1 = 1, a_2 = 1, a_{n+1} = a_n + a_{n-1}$)。

fibonacci.m

```
1 a1 = 1; disp(a1);
2 a2 = 1; disp(a2);
3 a3 = a1 + a2;
4 while a3 <= 100
5     disp(a3);
6     a1 = a2;
7     a2 = a3;
8     a3 = a1 + a2;
9 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

fibonacciOdd.m

```
1 a1 = 1; a2 = 1;
2 disp(a1); disp(a2);
3 a3 = a1 + a2;
```

```
4 while 1
5     a1 = a2;
6     a2 = a3;
7     a3 = a1 + a2;
8     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 $a3 \leq 100$ 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[\[253\]](#)

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

```

1 f = @(x) x^2 + 1/x;
2 x = [1,2; 3,4];
3 ans = zeros(size(x));
4 for ii = 1:numel(x)
5     ans(ii) = f(x(ii));
6 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

[< 输出变量 1>, < 输出变量 2>, ...] = < 函数名 >(< 输入变量 1>, < 输入变量 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

```
1 function y = myfun(x, A, phi, y0)
2 if nargin < 4
3     y0 = 0;
4     if nargin < 3
5         phi = 0;
6         if nargin < 2
7             A = 1;
8         end
9     end
10 end
11 y = A*cos(x + phi) + y0;
12 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^[261]”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^[257]

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.

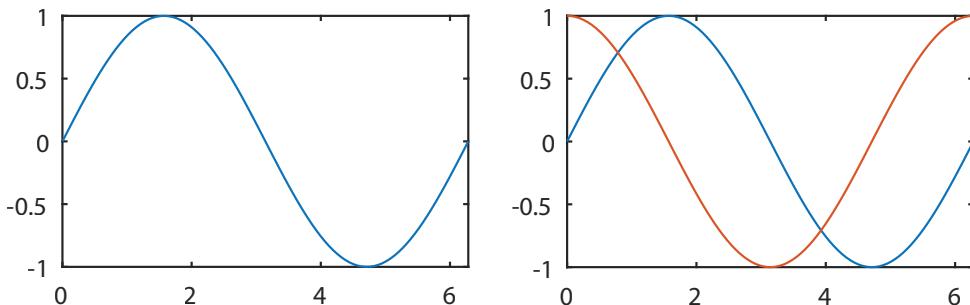


图 1: plot 函数

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

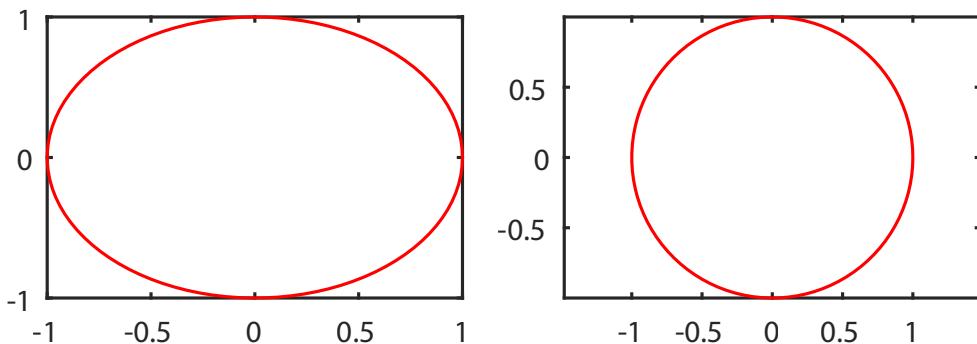


图 2: 红色的单位圆

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。

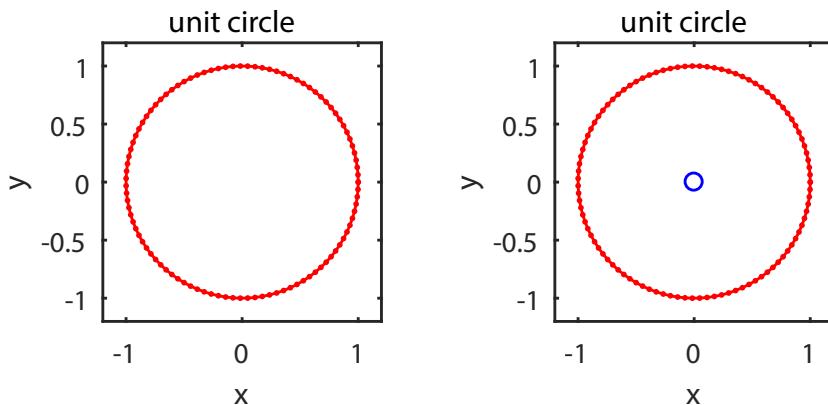


图 3: 红色的单位圆

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

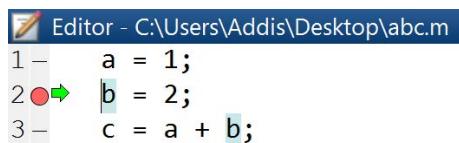
```
>> hold on; scatter(0, 0, 'b');
```

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。

预备知识 Matlab 的函数[\[257\]](#)

程序调试



Editor - C:\Users\Addis\Desktop\abc.m

```
1 - a = 1;
2 → b = 2;
3 - c = a + b;
```

The screenshot shows the MATLAB Editor window with the file 'abc.m' open. Line 2, which contains the assignment 'b = 2;', has a red circular breakpoint marker placed before it. The code consists of three lines: 'a = 1;', 'b = 2;', and 'c = a + b;'. The line with the breakpoint is highlighted in light gray.

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

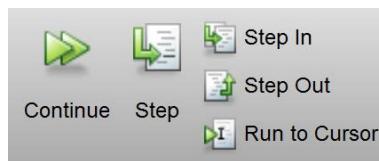


图 2: Step 菜单

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

```
1 function myfun(char)
2 if char == 'w'
3     warning('this is a warning');
4 elseif char == 'e'
5     error('this is an error');
6 else
7     disp('Hello World!');
8 end
9 end
```

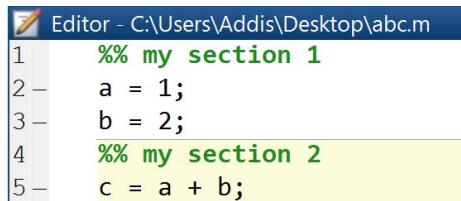
当 error 或 warning 函数被执行时，控制行不但会输出对应的字符串，还会输出它们所在的文件和行号，如

```
>> myfun w
Warning: this is a warning
> In myfun (line 3)
```

单击带下划线的单词，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）。



```
Editor - C:\Users\Addis\Desktop\abc.m
1    %% my section 1
2 -    a = 1;
3 -    b = 2;
4    %% my section 2
5 -    c = a + b;
```

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

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）[\[4\]](#), Matlab 编程基础[\[239\]](#)

这里介绍一个简单的 Matlab 程序用于计算表 1[\[4\]](#). 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.

binomial.m

```
1 % 验证二项式定理(非整数幂)
2 u = -3.5; % 幂
3 x = 0.6; % |x|<1 使级数收敛
4 N = 100; % 求和项数
5 Coeff = 1; % x^ii 项前面的系数
6 result = 1; % 求和结果
7 for ii = 1:N
8     % 由 x^(ii-1) 项系数计算 x^ii 项系数
9     Coeff = Coeff*(u-ii+1) / ii;
10    % 将 x^(ii-1) 项累加到求和结果上
11    result = result + Coeff * x^ii;
12 end
13 disp('求和结果为')
14 disp(result)
15 format short % 恢复默认显示
16 disp('精确结果为')
17 format long % 显示全部小数位
18 disp((1+x)^u)
```

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, 再来求新区间的终点. 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.

bisection.m

```
1 % 二分法求函数的根
2 function x = bisection(f, int, err)
3 fl = f(int(1)); fr = f(int(2));
4 % 两端点是否为 0
5 if fl == 0
6     x = int(1); return;
7 elseif fr == 0
8     x = int(2); return;
```

```
9 end
10 % 两端点是否同号
11 if fl * fr > 0
12     error('两端点同号');
13 end
14 % 主循环
15 while(int(2) - int(1) > err)
16     mid = 0.5*(int(1) + int(2));
17     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，知道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[\[269\]](#)

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, 小区间个数越多, 越有可能找到所有的根. 例程如下.

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^[269]”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

```

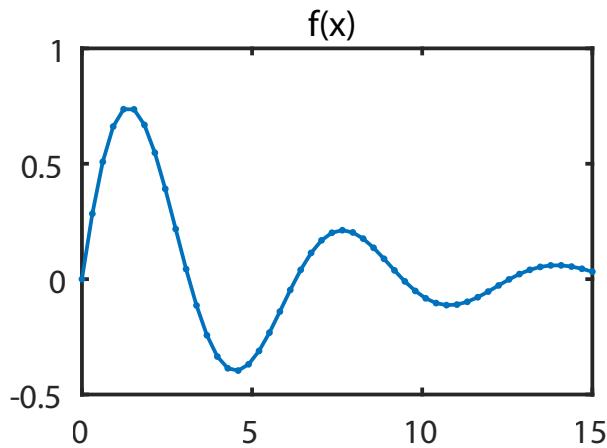


图 1: 运行结果

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
```

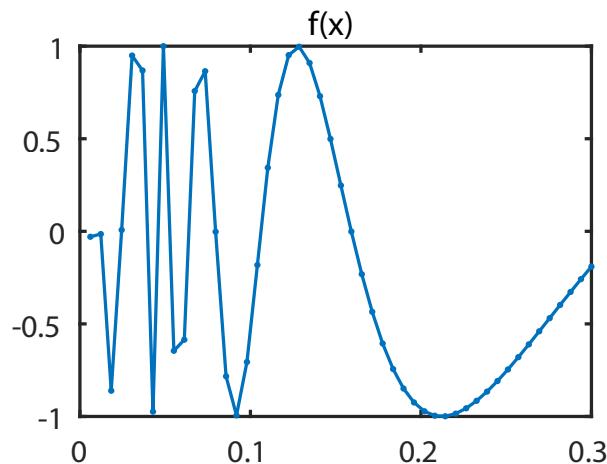


图 2: 运行结果

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[\[257\]](#)

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, 效率没有 `sort` 中的算法快, 所以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以一一对应.

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
% 是否是降序排列
21 if nargin > 1 && option(1) == 'd'
22     x(:) = flipud(x(:));
23     order = fliplr(order);
24 end
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。

Nelder-Mead 算法

预备知识 Matlab

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。以下讲解具体算法。

我们先根据初始点另外生成另外 N 个初始点 $\mathbf{x}_2 \dots \mathbf{x}_{N+1}$ ，使 \mathbf{x}_{1+i} 在第的第 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

$$\mathbf{m} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \quad (1)$$

3. 计算 \mathbf{x}_{N+1} 关于点 \mathbf{m} 的反射点为

$$\mathbf{r} = 2\mathbf{m} - \mathbf{x}_{N+1} \quad (2)$$

4. 如果 $f(\mathbf{x}_1) \leq f(\mathbf{r}) < f(\mathbf{x}_N)$ ，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

$$\mathbf{s} = \mathbf{m} + 2(\mathbf{m} - \mathbf{x}_{N+1}) \quad (3)$$

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 $f(\mathbf{x}_N) \leq f(\mathbf{r}) < f(\mathbf{x}_{N+1})$, 令

$$\mathbf{c}_1 = \mathbf{m} + (\mathbf{r} - \mathbf{m})/2 \quad (4)$$

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 $f(\mathbf{x}_{N+1}) \leq f(\mathbf{r})$ 令

$$\mathbf{c}_2 = \mathbf{m} + (\mathbf{x}_{N+1} - \mathbf{m})/2 \quad (5)$$

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

$$\mathbf{v}_i = \mathbf{x}_1 + (\mathbf{x}_i - \mathbf{x}_1)/2 \quad (i = 2 \dots N+1) \quad (6)$$

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.

NelderMead.m

```

1 % 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
2 % x0 是 N 维行矢量, xerr 是一个标量
3 function [xmin, fmin] = NelderMead(f, x0, xerr)
4 N = numel(x0); % f 是 N 元函数
5 x = zeros(N+1, N); % 预赋值
6 y = zeros(1, N+1);

```

```
7 % 计算 N+1 个初始点
8 x(1,:) = x0;
9 for ii = 1:N
10    x(ii+1,:) = x(1,:);
11    if x(1,ii) == 0
12        x(ii+1,ii) = 0.00025;
13    else
14        x(ii+1,ii) = 1.05 * x(1,ii);
15    end
16 end
17 % 主循环
18 for kk = 1:10000
19     % 求值并排序
20     for ii = 1 : N + 1
21         y(ii) = f(x(ii,:));
22     end
23     [y, order] = sort(y);
24     x = x(order,:);
25     if norm(x(N+1,:) - x(1,:)) < xerr % 判断误差
26         break;
27     end
28     m = mean(x(1:N,:)); % 平均位置
29     r = 2*m - x(N+1,:); % 反射点
30     f_r = f(r);
31     if y(1) <= f_r && f_r < y(N) % 第 4 步
32         x(N+1,:) = r; continue;
33     elseif f_r < y(1) % 第 5 步
34         s = m + 2*(m - x(N+1,:));
35         if f(s) < f_r
36             x(N+1,:) = s;
37         else
```

```
38 x(N+1,:) = r;
39 end
40 continue;
41 elseif f_r < y(N+1) % 第 6 步
42 c1 = m + (r - m)*0.5;
43 if f(c1) < f_r
44     x(N+1,:) = c1; continue;
45 end
46 else % 第 7 步
47     c2 = m + (x(N+1,:)-m)*0.5;
48     if f(c2) < y(N+1)
49         x(N+1,:) = c2; continue;
50     end
51 end
52 for jj = 2:N+1 % 第 8 步
53     x(jj,:) = x(1,:)+ (x(jj,:)-x(1,:))*0.5;
54 end
55 end
% 输出变量
56 xmin = x(1,:);
57 fmin = f(xmin);
58 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^[120], Nelder-Mead 算法^[275]

我们在“最小二乘法^[120]”中见到的三个例子中，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^[120]”中的方法比较结果。

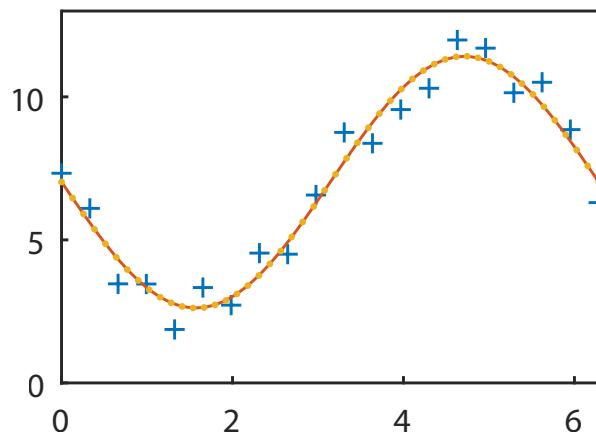


图 1: 运行结果

curveFit.m

```
1 function curveFit
2 close all;
3 % 随机生成简谐曲线
4 N = 20;
5 x0 = linspace(0, 2*pi, N);
```

```
6 y0 = 5*rand * sin(x0 + 2*pi * rand) + 10 * rand;
7 y0 = y0 + 2*rand(1,20); % 产生随机误差
8 scatter(x0, y0, '+'); % 画出散点
9 hold on;
10
11 % Nelder-Mead 求方差最小值点
12 f = @(x) norm( x(1)*sin(x0 + x(2)) + x(3) - y0 );
13 c = NelderMead(f, [5, 1, 5], 1e-7);
14 % 画拟合结果
15 x = linspace(0, 2*pi, 50);
16 y1 = c(1) * sin(x + c(2)) + c(3);
17 plot(x, y1);
18
19 % 解线性方程组求方差最小值点
20 c = sinfit(x0, y0);
21 % 画拟合结果
22 y2 = c(1)*cos(x) + c(2)*sin(x) + c(3);
23 plot(x, y2, '.');
24 end
25
26 % 拟合简谐曲线
27 % y = C(1)*cos(x) + C(2)*sin(x) + C(3)
28 function C = sinfit(x, y)
29 N = numel(x);
30 cosx = cos(x); sinx = sin(x);
31 sc = sum(sinx.*cosx);
32 s = sum(sinx); c = sum(cosx);
33 % 系数矩阵
34 M = [sum(cosx.^2), sc, c;
35 sc, sum(sinx.^2), s;
36 c, s, N];
```

```
37 b = [sum(y.*cosx); sum(y.*sinx); sum(y)];  
38 C = M\b; % 解线性方程组  
39 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^[275]”中的函数 NelderMead 求函数句柄 f 的最小值。

第三章

微分方程

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^[217], 微分近似^[47], Matlab 编程基础^[239]

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^[217] 中, 列出的二阶微分方程为

$$m\ddot{y} = \alpha\dot{y} - ky + f(t) \quad (1)$$

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.

$$\ddot{y}(0) = [-\alpha\dot{y}(0) - ky(0) + f(0)]/m \quad (2)$$

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.

$$y(\Delta t) = y(0) + \Delta y \approx y(0) + \dot{y}(0)\Delta t \quad (3)$$

$$\dot{y}(\Delta t) = \dot{y}(0) + \Delta \dot{y} \approx \dot{y}(0) + \ddot{y}(0)\Delta t \quad (4)$$

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

$$\ddot{y}(\Delta t) = [-\alpha\dot{y}(\Delta t) - ky(\Delta t) + f(\Delta t)]/m \quad (5)$$

再次使用微分近似有

$$y(2\Delta t) = y(\Delta t) + \Delta y \approx y(\Delta t) + \dot{y}(\Delta t)\Delta t \quad (6)$$

$$\dot{y}(2\Delta t) = \dot{y}(\Delta t) + \Delta \dot{y} \approx \dot{y}(\Delta t) + \ddot{y}(\Delta t)\Delta t \quad (7)$$

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.

SHOf.m

```
1 % 参数设定
2 m = 0.1; % 质量
3 k = 1; % 劲度系数
4 a = 0.03; % 阻尼系数
5 T = 20; % 停止时间
6 Nstep = 10000; % 步数
7 A = 2; w = 3; % f(t)=A*sin(w*t);
8
9 dt = T/Nstep; % 计算步长
10 y2 = zeros(step,1); y1 = y2; y = y2; % 矩阵预赋值
11 y(1) = 0; y1(1) = 0; % 初值, y1 是 y 的一阶导数
12
13 % 迭代循环
14 for ii = 2:step
15     y2(ii) = (-a*y1(ii)-k*y(ii)+2*sin(w*(ii*dt)))/m; %
16         代入微分方程求出 y''.
17     y(ii) = y(ii-1) + y1(ii-1)*dt; % y 的微分近似
18     y1(ii) = y1(ii-1) + y2(ii-1)*dt; % y' 的微分近似
19 end
20
21 % 画图
22 t=(0:step-1)*dt;
23 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

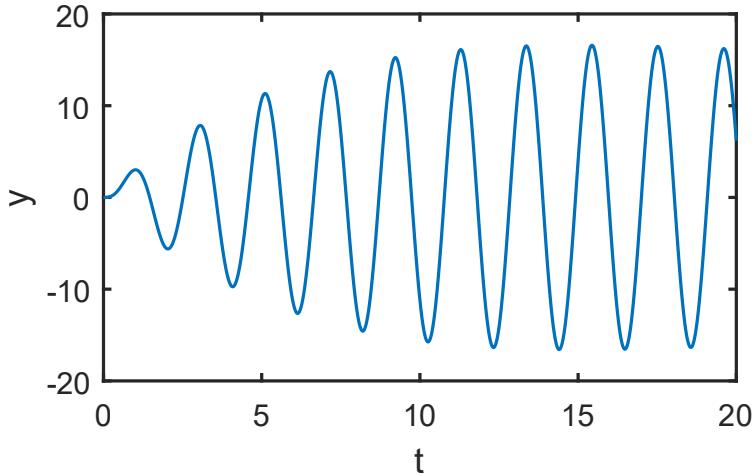


图 1: 运行结果

预备知识 万有引力^[225], 弹簧振子受迫运动的简单数值计算^[283]

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

$$\mathbf{F} = -G \frac{Mm}{r^2} \mathbf{r} = -G \frac{Mm}{r^2} \mathbf{r} \quad (1)$$

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^[162], 加速度为

$$\mathbf{a} = \frac{\mathbf{F}}{m} = -G \frac{M}{r^3} \mathbf{r} \quad (2)$$

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 $a_x = \frac{d^2x}{dt^2}$, $a_y = \frac{d^2y}{dt^2}$, 可以列出二阶微分方程组

$$\begin{cases} \frac{d^2x}{dt^2} = -\frac{GM}{(x^2 + y^2)^{3/2}} x \\ \frac{d^2y}{dt^2} = -\frac{GM}{(x^2 + y^2)^{3/2}} y \end{cases} \quad (3)$$

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“簧振子受迫振动的简单数值计算^[283]”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

$$\begin{cases} \ddot{x}(0) = -\frac{GM}{(x_0^2 + y_0^2)^{3/2}} x_0 \\ \ddot{y}(0) = -\frac{GM}{(x_0^2 + y_0^2)^{3/2}} y_0 \end{cases} \quad (4)$$

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)

$$\begin{cases} \dot{x}(\Delta t) \approx \ddot{x}(0)\Delta t + x(0) \\ \dot{y}(\Delta t) \approx \ddot{y}(0)\Delta t + y(0) \end{cases} \quad \begin{cases} x(\Delta t) \approx \dot{x}(0)\Delta t + x(0) \\ y(\Delta t) \approx \dot{y}(0)\Delta t + y(0) \end{cases} \quad (5)$$

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 $x(2\Delta t), y(2\Delta t), x'(2\Delta t), y'(2\Delta t) \dots$ 如此循环下去就可以得到每隔 Δt 的数值解.

kepler.m

```

1 % 参数设定
2 GM = 1; % 万有引力常数乘以中心天体质量
3 x0 = 1; y0 = 0; % 初始位置
4 vx0 = 0; vy0 = 0.7; % 初始速度
5 T = 4; Nstep = 4000; % 总时间和步数
6 dt = T/Nstep; % 步长
7
8 % 矩阵预赋值
9 x = nan(Nstep,1); y = x;
10 x1 = x; y1 = x;
11 x2 = x; y2 = x;
12
13 % 初始位置, 速度, 加速度
14 x(1) = x0; y(1) = y0; % 初位置
15 x1(1) = vx0; y1(1) = vy0; % 初速度
16 x2(1) = -GM*x(1)/(x(1)^2+y(1)^2)^(3/2); % 代入方程得到 x
17 y2(1) = -GM*y(1)/(x(1)^2+y(1)^2)^(3/2); % 代入方程得到 y
18
19 % 迭代循环

```

```

20 for ii = 2:Nstep
21     x(ii) = x(ii-1)+x1(ii-1)*dt; % x的微分
22     y(ii) = y(ii-1)+y1(ii-1)*dt; % y的微分
23
24     x1(ii) = x1(ii-1)+x2(ii-1)*dt; % x' 的微分
25     y1(ii) = y1(ii-1)+y2(ii-1)*dt; % y' 的微分
26
27     x2(ii) = -GM*x(ii)/(x(ii)^2+y(ii)^2)^(3/2); % 代入微
28         分方程求出 x'''
29     y2(ii) = -GM*y(ii)/(x(ii)^2+y(ii)^2)^(3/2); % 代入微
30         分方程求出 y'''
31 end
32
33 % 画图
34 plot(x,y); % 画行星轨道
35 axis equal; % xy坐标长度一致
36 hold on; % 继续画图
37 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.

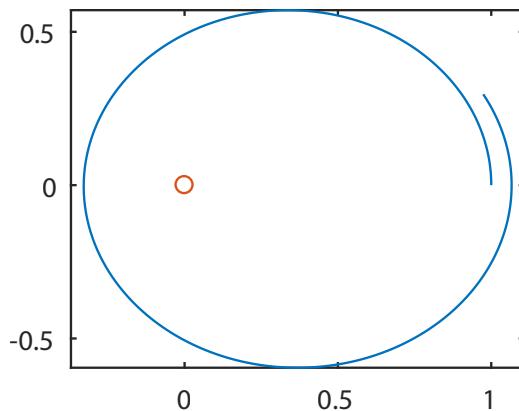


图 1: 运行结果

常微分方程（组）的数值解

预备知识 弹簧振子受迫运动的简单数值计算^[283]，天体运动的简单数值计算^[284]

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^[283]”中列出的微分方程为

$$my'' = \alpha y' - ky + f(t) \quad (1)$$

新增变量，令 $v = y'$ ，则可变为一阶微分方程组

$$\begin{cases} v' = [-\alpha u - ky + f(t)]/m \\ y' = v \end{cases} \quad (2)$$

方程组中， t 是自变量， y 和 u 是因变量。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^[284]”中，列出的二阶微分方程组为

$$\begin{cases} x'' = -\frac{GM}{(x^2 + y^2)^{3/2}}x \\ y'' = -\frac{GM}{(x^2 + y^2)^{3/2}}y \end{cases} \quad (3)$$

新增变量，令 $v_x = x'$, $v_y = y'$ ，上式也可变为一阶微分方程组

$$\begin{cases} x' = v_x \\ v'_x = -GMx/(x^2 + y^2)^{3/2} \\ y' = v_y \\ v'_y = -GMy/(x^2 + y^2)^{3/2} \end{cases} \quad (4)$$

该式中同样 t 是自变量，其他都是因变量，给出某时刻的 x, y, v_x, v_y, t 就可以由该方程组求出因变量的一阶导数。

在以上两个词条中，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 的各因变量代入一阶微分方程组，得到 t 时刻各因变量

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 $t + \Delta t$ 时刻的各因变量，再代入一阶方程组得到 $t + \Delta t$ 时刻各因变量的一阶导数，如此一直循环，得到各因变量每隔 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

$$y_{n+1} = y_n + h f(y_n, t_n) \quad (5)$$

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[\[289\]](#)
- 四阶龙格库塔法[\[292\]](#)
- 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[\[288\]](#)

我们先来尝试用欧拉法解一阶微分方程

$$y'(t) = y \quad (1)$$

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）。

由“”我们知道该方程的解析解为 $y = e^t$ 。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

$$y\left(t_n + \frac{h}{2}\right) = y_n + \frac{h}{2}f(y_n, t_n) \quad (2)$$

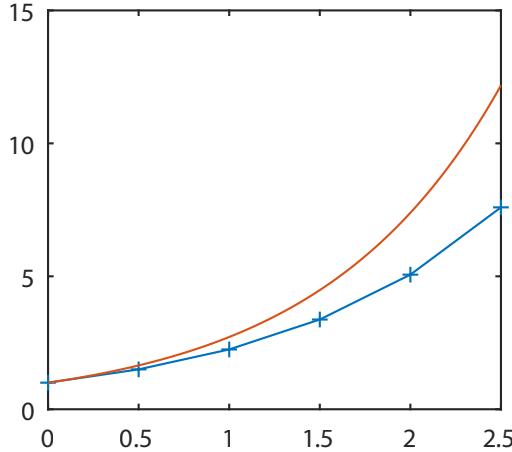


图 1: 欧拉法数值解 (蓝) 和解析解 (红)

然后再求出这个近似中点的导数为

$$y'\left(t_n + \frac{h}{2}\right) = f\left[y_n + \frac{h}{2}f(t_n, y_n), t_n + \frac{h}{2}\right] \quad (3)$$

最后我们利用这个导数估算该区间的曲线增量

$$y_{n+1} = y_n + hy'\left(t_n + \frac{h}{2}\right) = y_n + hf\left[y_n + \frac{h}{2}f(t_n, y_n), t_n + \frac{h}{2}\right] \quad (4)$$

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.

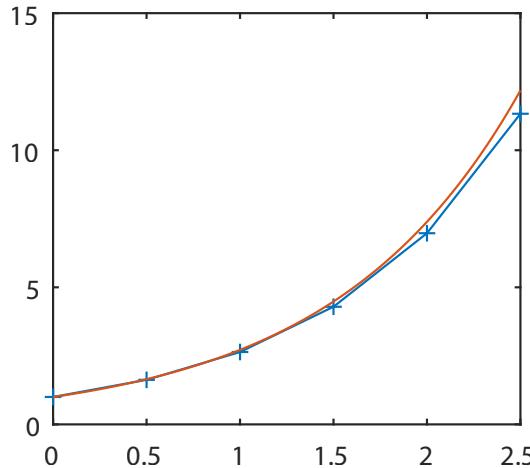


图 2: 中点法数值解 (蓝) 和解析解 (红)

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

$$\begin{cases} x'(t) = f(x, y, t) \\ y'(t) = g(x, y, t) \end{cases} \quad (5)$$

首先计算近似中点为

$$\begin{cases} x_{n+1/2} = x_n + hf(x_n, y_n, t_n) \\ y_{n+1/2} = y_n + hg(x_n, y_n, t_n) \end{cases} \quad (6)$$

然后有

$$\begin{cases} x_{n+1} = x_n + hf\left(x_{n+1/2}, y_{n+1/2}, \frac{h}{2}\right) \\ y_{n+1} = y_n + hg\left(x_{n+1/2}, y_{n+1/2}, \frac{h}{2}\right) \end{cases} \quad (7)$$

odeMid.m

```

1 % 设置参数
2 N = 6;
3 h = 0.5;
4 t = linspace(0, (N-1)*h, N); % 自变量
5 t0 = linspace(0, (N-1)*h, 100); % 用于画图
6
7 % 欧拉法
8 y = zeros(1,N); % 预赋值
9 y(1) = 1; % 初值
10 for ii = 1:N-1
11     y(ii+1) = y(ii) + h*y(ii);
12 end
13
14 % 画图
15 figure;
```

```

16 plot(x,y,'+-');
17 hold on;
18 plot(t0, exp(t0));
19
20 % 中点法
21 y = zeros(1,N);
22 y(1) = 1;
23 for ii = 1:N-1
24     y(ii+1) = y(ii) + h*(y(ii) + 0.5*h*y(ii));
25 end
26
27 % 画图
28 figure;
29 plot(x,y,'+-');
30 hold on;
31 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）[\[289\]](#)

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](#)[\[288\]](#)）

$$y_{n+1} = y_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) \quad (1)$$

其中

$$\begin{aligned} k_1 &= f(y_n, t_n) & k_2 &= f\left(y_n + h\frac{k_1}{2}, t_n + \frac{h}{2}\right) \\ k_3 &= f\left(y_n + h\frac{k_2}{2}, t_n + \frac{h}{2}\right) & k_4 &= f(y_n + hk_3, t_n + h) \end{aligned} \quad (2)$$

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

$$\begin{cases} y'_1(t) = f_1(y_1, \dots, y_N, t) \\ y'_2(t) = f_2(y_1, \dots, y_N, t) \\ \vdots \\ y'_N(t) = f_N(y_1, \dots, y_N, t) \end{cases} \quad (3)$$

我们可以把该式记为矢量函数的形式

$$\mathbf{y}'(t) = \mathbf{f}(\mathbf{y}, t) \quad (4)$$

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 $y(t)$ 。

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 $f(y, t)$ 的函数句柄^[257], $tspan$ 是一个 2×1 的列矢量, $tspan(1)$ 是初始时间, $tspan(2)$ 是终止时间, Y_0 是一个列矢量, $Y_0(ii)$ 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 $Nt - 1$ 个小时. 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 Y_0 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 (y_i) 求第 $ii+1$ 列 (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^[284]”中的问题. 我们令因变量 y 的四个分量依次为一阶方程组 (式 4^[288])

$$\begin{cases} x' = v_x \\ y' = v_y \\ v'_x = -GMx/(x^2 + y^2)^{3/2} \\ v'_y = -GMy/(x^2 + y^2)^{3/2} \end{cases} \quad (5)$$

中的 x, y, v_x, v_y . 程序代码如下

keplerRK4.m

```

1 function keplerRK4
2 % 参数设定
3 GM = 1; % 万有引力常数乘以中心天体质量
4 x0 = 1; y0 = 0; % 初始位置
5 vx0 = 0; vy0 = 0.7; % 初始速度
6 tspan = [0; 4]; % 总时间和步数
7 Nt = 100; % 步数
8
9 Y0 = [x0; y0; vx0; vy0]; % 因变量初值

```

```

10 f = @(Y, t) fun(Y, t, GM);
11 [Y, ~] = odeRK4(f, tspan, Y0, Nt);
12
13 % 画图
14 figure; hold on;
15 plot(Y(1, :), Y(2, :));
16 scatter(0, 0);
17 axis equal;
18 end
19
20 function Y1 = fun(Y, ~, GM)
21 % 因变量
22 x = Y(1); y = Y(2);
23 vx = Y(3); vy = Y(4);
24 Y1 = zeros(4, 1); % 预赋值
25 Y1(1) = vx;
26 Y1(2) = vy;
27 temp = -GM / (x^2 + y^2)^(3/2);
28 Y1(3) = temp * x;
29 Y1(4) = temp * y;
30 end

```

运行结果如图 1 所示。

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^[284]”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

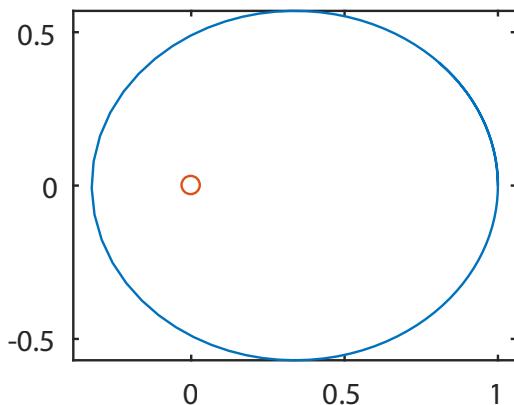


图 1: 运行结果

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.

物理学常数定义

国际单位的定义 (SI Units)

物理单位的一个重要性质就是可测量，至少理论上可测量. 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.

秒 (s) 的定义

铯原子 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.

米 (m) 的定义

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，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.

光速 (c) 的定义

$$c = 299792458 \text{ m/s} \quad (1)$$

说明：根据米的定义，一秒中光可以在真空中传播 299792459 m.

千克 (kg) 的定义

等于国际公斤原器的质量说明：千克是现有的唯一一个由特定的物品所定义单位。国际公斤原器是国际计量大会制造的，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。国际计量大会最终在 2014 年决定原则上千克应该由普朗克常数所决定，但是最终的定义再次被推迟。

牛顿 (N) 的定义

等于使 1 kg 物体获得 1 m/s^2 加速度的力。

真空磁导率 (μ_0) 的定义

$$\mu_0 = 4\pi \times 10^{-7} \text{ N/A}^2 \quad (2)$$

真空介电常数 (ϵ_0) 的定义

$$\epsilon_0 = 1/(c^2 \mu_0) = 8.8541878176 \dots \times 10^{-12} \text{ F/m} \quad (3)$$

说明：其中 c 是光速， μ_0 是真空磁导率。根据该定义， $c = 1/\sqrt{\epsilon_0 \mu_0}$.

库伦 (C) 的定义

若真空中两个相同的点电荷相距一米，产生的相互作用力为 $1/(4\pi\epsilon_0)$ ，则该点电荷为 1 库伦。

安培 (A) 的定义

以下两种定义等效：

1. 每秒钟经过横截面的电荷量为 1 库伦的电流就是 1 安培.
2. 两根相距一米的无限长平行细导线流入 1 安培电流后，相互作用力是 $2 \times 10^{-7} \text{ N/m}$.