北京大学《高等数学》(B) (一) 课程讲义
2024 年秋学期-07 班

赖志坚
北京大学北京国际数学研究中心
最后更新于 2024 年 12 月 18 日

如发现任何错误, 请联系 lai_zhijian@pku.edu.cn, 我会将您的姓名加入致谢名单.

目录

致谢

本讲义改编自北京大学高等数学 (B) 课程指定教材《高等数学》(上册) 李忠, 周建莹编著, 北大出版社. 我们补充了许多其他相关资料, 尤其是国内广泛使用的《高等数学》同济大学版. 本讲义使用的 LaTeX 模板 FunMathArticles 来自网络, 感谢其作者 shzaiz.

感谢以下同学和老师对本讲义的建议和勘误上的帮助: 盛仁杰, 王蕙钰, 张茂林, 谢信豪, 肖博闻, 王昱涵, 楚一飞, 杨启帆, . . .

教材中没制作部分: 第三章, 第 5 节第 4 小节曲线弧的质心与转动惯量; 第 6 节定积分的近似计算. 第四章, 第 7 节曲线的曲率. 第六章, 第 10 节曲面论初步.

赖志坚

0.1 课程相关

0.1.1 基本信息

- 24-25 学年第 1 学期; 年级: 24 级.
- 课程号: 00130201; 课程名称: 高等数学 (B) (一); 学分: 5.0; 上课起止周: 1-16; 班号: 7.
- 任课教师: 赖志坚; 联系方式: lai_zhijian@pku.edu.cn
- 上课时间: 周二 3-4 节, 周五 1-2 节; 教室: 理教 302.
- 期中考试日期: 未定 (后续通知, 预计在周日)
- 期末考试日期: 2024 年 12 月 30 日 (周一上午)
- 最终成绩: 平时成绩 (以作业形式) 占 20%+ 期中考试成绩占 30%+ 期末考试成绩占 50%.
- 教材是《高等数学》李忠, 周建莹编著, 北大出版社. 期中考试内容仅涵盖前 3 章, 期末考试内容仅涵盖后 3 章. 教材, 考试题目, 考试时间和其他班级完全一样.

0.1.2 上课要求

1. 不点名.
2. 课前必须预习. 否则可以不用来听课, 属于浪费自己时间.
3. 每次课结束后会留作业. 每周二上习题课前将上周的作业交给各班习题课老师, 习题课老师对作业打分. 不交, 迟交或者抄袭将影响平时作业成绩.
4. 课本内容太多, 一看就懂的东西会被跳过不讲 (但是课下要自学). 我会讲到很多书上没有的提到的.
5. 及时将意见建议等反馈给任课老师和习题课老师.

0.2 如何学习

0.2.1 学习数学的建议

1. 大学的学习 90% 靠自学而不是课堂. 研究生以上更是 99.99%. 上课只是帮助总结, 验证和引导你的自学成果.
2. 看不懂课本就去网上搜索相关资料. 但是独立地思考和甄别.
   (a) 搜索引擎使用 bing, 如果可以推荐使用 google. 不建议用百度.
   (b) 如果中文找不到就用英文搜索. 英文资料更多更丰富.
   (c) B 站大学 (高等数学 - 哔哩哔哩 _bilibili) 有非常优秀的视频内容.
3. 学会用软件工具去帮助解决理解数学概念, 解决问题.
   (a) 使用简单的在线绘图工具: GeoGebra - the world's favorite, free math tools used by over 100 million students and teachers; Desmos | 免费使用的精美数学工具组.

   (b) 功能丰富商用数学软件: Matlab (北大正版软件共享平台); Mathematica
   (c) 无限可能的低门槛组合: 编程语言 python + 大模型辅助 (例如, DeepSeek; 更多请看 测爆我! 我们决定让全 B 站白嫖, 只为找到最强大模型! _ 哔哩哔哩 _bilibili)

4. 自学课本的建议.
   (a) 某一块内容理解不了就暂时跳过去, 之后再看可能就懂了.
   (b) 不用死磕一本书, 可以在知乎或者 B 站搜索其他推荐的国内外高数教材. 国外一般把本课程直接叫做 Calculus, 即微积分. 《高等数学》(第七版) 同济大学出版社, 是国内最广泛使用的教材, 有着超多的相关资料.
   (c) 英语正式变为一门工具, 可以从用英文学习数学开始.

0.2.2 如何学好大学数学

1. 高数是大学本科最重要的课程之一. 未来你在任何国家参加任何升学考试都是必考科目. 如果未来你从事科研工作, 它则将伴随你一生. 既然要学, 则一开始就好好学.
2. 充分地理解数学中的概念/定义是学习中首要任务.
3. 数学的证明过程往往比结论更重要. 跳过证明过程只看结论然后死记硬背或者直接套用, 是杀鸡取卵的学习方法. 不要让大学成为高中的延续.
4. 学数学就要去质疑. 数学不是别人规定好而你只需要学习和接受的知识和规则. 要敢于提出质疑, 多思考问: 我们为什么需要这个概念 (例如某个数学工具)? 为什么是如此定义, 而不是别的方式? 这个东西它具体有什么应用?
5. 老师也会犯错, 无论他的身份. 不要迷信任何人任何权威. 要相信你自己认真思考后的判断.
6. 高等数学内容较多, 大课基本都是在赶进度, 课下一定要花时间超前学.

第一章　函数和极限

1.1　实数

1.1.1　有理数与无理数

人类最早知道的是 **自然数 (natural number)**: $1, 2, 3, \cdots$. 通常全体自然数用 $\mathbf{N}$ 表示, 称之为自然数集.[1] 由于做加法逆运算[2]的需要, 人们增添了 0 及负整数, 从而将自然数扩充为 **整数 (integer)**. 今后, 我们用 $\mathbf{Z}$ 表示全体整数, 称之为整数集. 乘法的逆运算[3]又导致 **分数 (fraction)** 的产生, 而分数又称为 **有理数 (rational number)**.

定义 1.1.1: 有理数

通常用 $\mathbf{Q}$ 表示全体有理数 (称为有理数集), 即

$$\mathbf{Q} = \left\{ \frac{m}{n} \,\middle|\, m, n \in \mathbf{Z}, n > 0, (m, n) = 1 \right\}$$

其中 (m, n) 表示 m 与 n 的最大公约数[a]. $(m, n) = 1$ 表明 m 与 n 没有大于 1 的公约数, 因而此时 $\dfrac{m}{n}$ 是既约分数.

[a] Greatest common divisor - Wikipedia In mathematics, the greatest common divisor (GCD) of two or more integers, which are not all zero, is the largest positive integer that divides each of the integers. For two integers x, y, the greatest common divisor of x and y is denoted $\gcd(x, y)$. For example, the GCD of 8 and 12 is 4, that is, $\gcd(8, 12) = 4$. 不建议使用课本上的 "仅括号" 记号, 容易有歧义.

为什么不直接使用如下定义?

$$\mathbf{Q} = \left\{ \frac{m}{n} \,\middle|\, m, n \in \mathbf{Z}, n \neq 0 \right\}.$$

回答这个问题等于回答下面几个问题.

1. 定义中为什么要求分母 n 是正的?

   分母为正时, 有理数的符号完全由分子 m 决定, 这样可以避免同一个有理数出现多种不同的表示方式. 例如, 如果我们不要求 $n > 0$, 那么 $\dfrac{1}{2}$ 和 $\dfrac{-1}{-2}$ 都可以作为相同有理数的表示, 再比如 $\dfrac{3}{-5}$ 和 $\dfrac{-3}{5}$. 这些会引入不必要的冗余. 因此, 通过规定分母 n 为正数, 可以保证每个有理数都只有一个标准化的表示形式.

[1] Natural number - Wikipedia In mathematics, the natural numbers are the numbers $0, 1, 2, 3$, etc., possibly excluding 0. Some define the natural numbers as the non-negative integers $0, 1, 2, 3, \ldots$, while others define them as the positive integers $1, 2, 3, \ldots$. Some authors acknowledge both definitions whenever convenient.

[2] 什么是加法的逆运算? $a - a = a + (-a) = 0$.

[3] 什么是乘法的逆运算? $a/a = a \times \dfrac{1}{a} = 1$. 算数中, 为什么不能除以 0? 后者可参考: 都说 1 不能除以 0, 如果我硬要除呢? _ 哔哩哔哩 _bilibili; 为什么不能除以 0?_ 哔哩哔哩 _bilibili

2. 根据前一问的回答, 为什么是要求 “分母 n 是正的”, 而不是要求 “分子 m 是正的”? 他们的效果不是一样吗?

   原则上要求 “分子 m 是正的” 是可以的. 就像中美开车走右侧, 英日开车走左侧. 但是要求 “分母 n 是正的” 这一条件已经暗含地排除了分母 $n = 0$ 的情况, 所以书写更加简洁.

3. 定义中为什么要求是既约分数?

   如果不要求则会导致每个有理数具有多种不同的表示方式. 例如, $\frac{1}{2}$ 可以表示为 $\frac{2}{4}$, $\frac{3}{6}$ 等等. 为了避免这种冗余表示, 通常要求有理数的表示形式为既约分数, 即要求 $\frac{m}{n}$ 中的分子 m 和分母 n 的最大公约数 $(m, n) = 1$. 这种定义确保每个有理数都有唯一的表示形式. 此外, 使用既约分数的定义使得有理数集合的数学结构更加简单, 在研究有理数的性质时, 证明和推导也更为便捷. 比如, 下面的命题 1.1.1 中就用到了这一性质.

总之, 从集合的角度来看, 我们不是不能采用 $\mathbf{Q} = \left\{ \frac{m}{n} \,\middle|\, m, n \in \mathbf{Z}, n \neq 0 \right\}$, 因为两种不同形式的定义得到的集合在根本上是一致的. 只是使用 $\mathbf{Q} = \left\{ \frac{m}{n} \,\middle|\, m, n \in \mathbf{Z}, n > 0, (m, n) = 1 \right\}$ 有诸多的好处.

有理数集 $\mathbf{Q}$ 的一个重要特征是对加减乘除 (除数不为零) 四则运算 **封闭** (closed), 即这个集合之中的任意两个数做上述四种运算时, 其结果仍在这个集合之中. 粗略地说, 对加减乘除四则运算封闭的数的集合叫作 **数域 (field)**. 因此, 有理数集 $\mathbf{Q}$ 是一个数域, 称之为 **有理数域**.

公元前五百多年, 古希腊人发现了等腰直角三角形的腰与斜边没有公度[4], 从而证明了 $\sqrt{2}$ 不是有理数. 这样, 人类首次知道了无理数的存在.

命题 1.1.1

$\sqrt{2}$ 不是有理数.

证明. 用反证法. 假设 $\sqrt{2}$ 是有理数, 这时存在两个正整数 m 及 n, 使得 $(m, n) = 1$, 且

$$\sqrt{2} = \frac{m}{n}$$

对上式两边取平方, 即得到 $2n^2 = m^2$. 这表明 m^2 是偶数. 因此, m 一定是偶数. 设 $m = 2l$ (l 是某个正整数), 则有 $2n^2 = 4l^2$, 因此可得

$$n^2 = 2l^2.$$

这又表明 n^2 是偶数, 从而 n 也是偶数. 既然 m 与 n 均为偶数, 那么 2 就是它们的公约数. 这与 $(m, n) = 1$ 矛盾. 证毕. □

注意 大学以后的数学证明过程都是采用像上面一样的方式书写的, 而不是中学那种数学公式的堆砌. 要养成上面这种书写方式, 把数学符号和公式融入到句子中, 像写小作文一样去写证明. 不要吝啬语言文字, 不要只有数学符号. 为什么大学不用 3 个点表示的 “因为”, “所以” 符号呢? _ 哔哩哔哩 _bilibili

后来人们发现了更多的无理数, 比如 $\sqrt{3}, \sqrt{5}, \cdots$, 以及 π 与 e. 究竟什么是无理数? 在本书中, 我们不打算给出其严格的定义, 而只是把它们形式地视作一个无穷不循环小数. 从中学的数学课本中我们知道, 有理数

[4] “公度” 是一个古代数学概念, 指的是两个量是否存在一个共同的度量单位. 如果两个量可以用某个共同的度量单位来精确测量并表达为整数倍数, 那么这两个量就被称为 “有公度的”. 反之, 如果找不到这样一个共同的度量单位, 那么这两个量就被称为 “无公度的”.

可以表示成有穷小数或无穷循环小数, 比如

$$\frac{6}{5}=1.2,\quad \frac{11}{7}=1.571428571428\cdots$$

反过来, 任何有穷小数或无穷循环小数一定是有理数. 因此, 我们认为无理数等价于无穷不循环小数. 通常我们把有理数与无理数统称为 **实数 (real number)**, 并把全体实数组成的集合记作 $\mathbf{R}$, 称为实数集.

问题 $\mathbf{R}$ 表示实数集, 源于英语单词 “real”. $\mathbf{N}$ 表示自然数集, 源于英语单词 “natural”. 那么为什么用 $\mathbf{Z}$ 表示整数, 用 $\mathbf{Q}$ 表示有理数? 请参考 History of notation of sets: Why $\mathbb{Z}$ and $\mathbb{Q}$ for integers and rationals, but $\mathbb{R}$ and $\mathbb{N}$ for reals and naturals? - Mathematics Stack Exchange

1.1.2 实数集 R 是一个完备的有序数域

结论 1.1.1: 实数集 R 的基本性质

1. $\mathbf{R}$ 是一个数域 (称为实数域): 任意两个实数做加减乘除 (除数不为零) 四则运算后仍然是一个实数.

2. 对乘法和加法满足交换律, 结合律与分配律: 对于任意的 $a,b,c\in\mathbf{R}$, 总有

$$\begin{aligned}&a\cdot b=b\cdot a,\quad a+b=b+a\\&(a\cdot b)\cdot c=a\cdot(b\cdot c),\quad (a+b)+c=a+(b+c)\\&a\cdot(b+c)=a\cdot b+a\cdot c\end{aligned}$$

3. 实数域是一个有序数域. 确切地说, $\mathbf{R}$ 中任意两个不同的数 a 与 b 都有大小关系, 即 $a<b$ 与 $b<a$ 中有且只有一种情况成立[a], 并且这种大小关系在做加法与乘法运算时满足下列关系:

$$\begin{aligned}&a<b,c<d\Longrightarrow a+c<b+d\\&0<a,b<c\Longrightarrow a\cdot b<a\cdot c\end{aligned}$$

   顺便指出, 今后我们用记号 $a\leqslant b$ 表示 $a<b$ 或 $a=b$. 这也是一个常用的记号. 显然, 若 $a\leqslant b$ 且 $b\leqslant a$, 则必有 $a=b$.

4. 实数域具有 **完备性 (completeness)**[b]. 即, **在实数域中, 任意一个单调且有界的数列 (由实数构成的数的序列) 一定有极限存在于实数域中.**

[a]举一个反例, 在复数域 $\mathbf{C}$ 中, 大小关系没有明确定义. 例如, 对于两个复数 $z_1=1+i$ 和 $z_2=2+i$, 无法定义 $z_1<z_2$ 或 $z_2<z_1$, 因为复数没有类似于实数的全序结构.

[b]对于如何描述实数域的完备性, 有许多彼此等价的说法. 本书值采用其中一种作为实数域完备性的一种刻画.

现在我们来解释上述第 4 点.

- 数学中 “**序列 (sequence)**” 一词也叫 “**数列**”, 即数的序列.
- 关于什么是极限, 本章后面的几节中将详细讨论, 目前无须深究, 只要做一个朴素的理解即可: 所谓的常数 l 是序列 $\{a_n\}$ 的极限, 是指当 n 充分大时, a_n 可以任意接近于 l.
- 所谓的序列 $\{a_n\}$ 是 **有界的 (bounded)**, 是指 $a_n(n=1,\ 2,\cdots)$ 的绝对值 $|a_n|$ 有一个公共的上界, 即存在一个正数 M, 使得 $|a_n|\leqslant M(n=1,2,\cdots)$.

> **注释 1.1.1** 比如, $\left\{1+\dfrac{1}{2^n}\right\}$ 中每一项的绝对值均小于 2, 因而它是一个有界序列; 而 $\{n^2 \mid n \in \mathbf{Z}\}$ 则是一个无界序列.

- 若序列 $\{a_n\}$ 中每一项都不超过其后一项, 即 $a_n \leqslant a_{n+1}(n=1,2,\cdots)$, 则称 $\{a_n\}$ **单调递增 (monotonically increasing)**;

- 若序列 $\{a_n\}$ 中每一项都大于或等于其后一项, 即 $a_n \geqslant a_{n+1}(n=1,2,\cdots)$, 则称 $\{a_n\}$ **单调递减 (monotonically decreasing)**.

- 单调递增或单调递减的序列统称为 **单调序列 (monotonic sequence)**.

有理数域 $\mathbf{Q}$ 与实数域 $\mathbf{R}$ 有着实质性的差异. 这主要体现在: *有理数域对极限运算不是封闭的* (即, 有理数序列的极限可能不再是有理数.) 而 *在实数域中单调有界序列总有极限存在*. 实数域 $\mathbf{R}$ 的这一性质通常称为完备性.

例 1.1.1 比如, 对于逼近 $\sqrt{2}$ 的有理数序列 $1.4, 1.41, 1.414, \cdots$, 虽然它是一个单调递增的有界序列, 但是它在有理数域中却没有极限. 而实数域对极限运算是封闭的 (若实数序列有极限, 则其极限仍是实数).

例 1.1.2 更常见的一个例子是:

$$a_n = 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!}$$

这是一个有理数序列, 因为每一项都是有理数的有限和. 然而, 当 $n \to \infty$ 时, 序列趋于欧拉数 $e \approx 2.718$, 但 e 是无理数. 至于为什么, 这在未来会学习到.

这两个例子说明了有理数序列的极限可能会超出有理数的范围.

在本书中, 我们承认实数集 $\mathbf{R}$ 自然具有性质 1, 2, 3, 4, 而无须证明, 它们构成我们今后讨论的基础. 用一句话来概括: 实数集 $\mathbf{R}$ 是一个完备的有序数域. 暂时接受它, 如果想一探究竟, 你需要去学习《数学分析》一课.

1.1.3 数轴与区间

通过将实数集 $\mathbf{R}$ 与数轴上的点建立一一对应关系, 使得每个实数 x 都可以在数轴上找到唯一的对应点. 具体方法是在直线上取定一个原点 O, 一个单位长度和方向, 对于任意实数 x, 正数对应正向的点, 负数对应反向的点, 零对应原点. 这样, 数轴上的每个点都代表一个实数, 实数集 $\mathbf{R}$ 可以视为数轴上的点集.

在微积分中, 我们要用到 **区间 (interval)** 的概念.

定义 1.1.2: 区间

给定两个实数 a, b $(a < b)$. 我们可以定义如下各式各样的区间:

- 数集合 $\{x \mid a \leqslant x \leqslant b\}$ 称作 **闭区间 (closed interval)**, 记作 $[a, b]$.
- 数集合 $\{x \mid a < x < b\}$ 称作 **开区间 (open interval)**, 记作 (a, b).
- 类似地, 可定义半开半闭区间:

$$(a, b] = \{x \mid a < x \leqslant b\}, \quad [a, b) = \{x \mid a \leqslant x < b\}.$$

- 此外, 有时我们也可将整个数轴或实数集 $\mathbf{R}$ 表示成无穷区间 $(-\infty, +\infty)$[a].
- 在某些情况下, 我们还会考虑下列 **无穷区间**:

$$(a, +\infty) = \{x \mid a < x < +\infty\}, \quad (-\infty, b) = \{x \mid -\infty < x < b\}.$$

  据此, 读者可以自己定义无穷区间 $[a, +\infty)$ 及 $(-\infty, b]$.

[a]应该特别强调, 这里 $-\infty$ 与 $+\infty$ 只是两个记号而已, 它们不是两个数, 不能做任何运算.

特别地, 对于开区间 (a, b) 或闭区间 $[a, b]$, 我们定义其 **长度 (length)** 为 $b - a$, 并称点 $c = \dfrac{a+b}{2}$ 为 (a, b) 或 $[a, b]$ 的 **中心 (center)**, 也称 $\dfrac{b-a}{2}$ 为这两个区间的 **半径 (radius)**.

1.1.4 绝对值不等式

在微积分中, 我们经常要使用绝对值不等式来描述变量的变化. 因此, 熟练地运用绝对值不等式是十分重要的. 为此, 我们要复习一下在中学时学过的绝对值的概念. 实数 x 的绝对值记作 $|x|$, 它的定义是

$$|x| = \begin{cases} x, & x \geqslant 0 \\ -x, & x < 0. \end{cases}$$

因此, 实数 x 的绝对值 $|x|$ 总是非负的, 并且代表在数轴上点 x 到原点 O 的距离, 不论实数 x 是正的还是负的, 都是如此. 根据绝对值的定义, 立即可以推出下列命题:

命题 1.1.2: 绝对值函数的基本性质 1

对于任意的 $x, y \in \mathbf{R}$, 我们有:

1. $|x| \geqslant 0$, 其中等号当且仅当 $x = 0$ 时成立;
2. $|x| = |-x|$;
3. $|x + y| \leqslant |x| + |y|$;
4. $|xy| = |x||y|$.

证明. 只证明第 3 点.

- 情况 1: $x \geqslant 0$ 且 $y \geqslant 0$. 在这种情况下, $x+y \geqslant 0$, 因此

  $$|x+y| = x+y = |x|+|y|.$$

  不等式显然成立.

- 情况 2: $x<0$ 且 $y<0$. 在这种情况下, $x+y<0$, 因此

  $$|x+y| = -(x+y) = -x-y = |x|+|y|.$$

  不等式也成立.

- 情况 3: $x \geqslant 0$ 且 $y<0$. 此时 $x+y$ 的符号取决于 x 和 $|y|$ 的大小.

  1. 如果 $x+y \geqslant 0$, 则 $|x+y| = x+y$, 有 $|x+y| = x+y \leqslant x-y = |x|+|y|$.
  2. 如果 $x+y<0$, 则 $|x+y| = -(x+y)$, 有 $|x+y| = -x-y \leqslant x-y = |x|+|y|$.

  不等式也成立.

- 情况 4: $x<0$ 且 $y \geqslant 0$. 完全同情况 3, 只是交换了 x 和 y 的位置.

因此, 在所有情况下, $|x+y| \leqslant |x|+|y|$ 不等式都成立. □

一般来说, 给定两个实数 a 与 b, 则 $a-b$ 的绝对值 $|a-b|$ 在数轴上代表点 a 到点 b 的距离. 在命题 1.1.2 中, 令 $x=a-b, y=b-c$, 立即推出下列命题.

命题 1.1.3: 绝对值函数的基本性质 2

对于任意实数 a, b, c, 我们有:

1. $|a-b| \geqslant 0$, 其中等号当且仅当 $a=b$ 时成立;
2. $|a-b| = |b-a|$;
3. $|a-c| \leqslant |a-b|+|b-c|$.

命题 1.1.3 中的结论 (3) 称作 **三角不等式**. 当 a, b, c 是平面上的三点且不在一条直线上时, 这一结论便是三角形中两边之和大于第三边. 这就是我们称 (3) 为三角不等式的缘由.

今后, 我们会经常用到形如 $|x-a|<r$ 的不等式. 这里, 实数 a, r 是给定的值, 而 x 是变量. 此不等式在数轴上的几何意义是: 点 x 到点 a 的距离小于 r. 于是, 点 x 必然落入以点 a 为中心, r 为半径的区间 $(a-r, a+r)$ 之内, 即 $a-r<x<a+r$.

定义 1.1.3: 邻域

我们定义一类重要的区间:

- 通常我们将区间 $(a-r, a+r)$ 称为点 a 的 r 邻域, 简称 **邻域 (neighbourhood)**, 记为 $U_r(a)$;
- 而将 $U_r(a) \backslash \{a\}$ 称为点 a 的空心 r 邻域 (简称 **空心邻域**), 记为 $\mathring{U}_r(a)$.

有时也将 $U_r(a)$ 和 $\mathring{U}_r(a)$ 分别简记为 $U(a)$ 和 $\mathring{U}(a)$. (通常这么写是因为半径 r 是某个固定的值而没有变化, 或者即使变化了它也对我们关心的问题来说不重要)

总之, 我们有下列命题:

命题 1.1.4

对于任意实数 x, a, r, 其中 $r > 0$, 我们有:

$$|x-a| < r \Leftrightarrow a-r < x < a+r.$$

证明以下的命题是一个不错的练习.

命题 1.1.5

对于任意实数 x, y, 我们有:

$$||x|-|y|| \leqslant |x-y|.$$

证明. 由命题 1.1.2 得到

$$|x| = |x-y+y| \leqslant |x-y| + |y|,$$

于是我们有

$$|x| - |y| \leqslant |x-y|.$$

又由于 $|x-y| = |y-x|$, 在上式中交换 x 与 y 的位置后即得到

$$|y| - |x| \leqslant |x-y|.$$

这样, 我们得到

$$-|x-y| \leqslant |x| - |y| \leqslant |x-y|.$$

即实数 $|x|-|y|$ 在以 $-|x-y|, |x-y|$ 为端点的闭区间之内, 从而 $|x|-|y|$ 的绝对值不超过 $|x-y|$. 证毕. $\square$

1.2 变量与函数

1.2.1 函数的定义

高等数学与初等数学的重要区别在于: 高等数学主要是处理变量的, 而初等数学大体上是处理常量的. 变量是微积分的基本研究对象. 在自然和社会现象中, 常常会看到一个变量依赖于另外一个或几个变量. 下面是一些例子.

- 金属杆的长度 l 依赖于温度 T 的变化:

  $$l = l_0(1+\alpha T),$$

  其中 l_0, α 为常数; 这是描述金属热胀冷缩现象的公式. 在物理学中, 当温度升高时, 金属杆的长度会增加, 这一现象称为热膨胀. 公式中的 l_0 是金属杆在初始温度下的长度, α 是材料的线膨胀系数, 表示材料的膨胀程度. 当温度 T 发生变化时, 金属杆的长度 l 会按照公式 $l = l_0(1+\alpha T)$ 进行调整.

- 一定量气体的体积 V 依赖于温度 T 与压力 p:

  $$V = c\frac{T}{p},$$

  其中 c 为常数; 这是理想气体状态方程的一个变形形式, 用来描述一定量气体的体积 V 如何依赖于温度 T 和压力 p.

- 若某快递员的月收入 x (单位: 元) 不超过 8000 元, 则他每月交纳的个人所得税金额 y (单位: 元) 与 x 有如下关系:

$$y = \begin{cases} 0, & 0 \leqslant x \leqslant 5000 \\ (x - 5000) \times 3\%, & 5000 < x \leqslant 8000. \end{cases}$$

  当收入 x 在 5000 元或以下时, 快递员不需要交纳个人所得税. 当收入 x 超过 5000 元但不超过 8000 元时, 税款按超出部分的 3% 计算. 这是超额累进税制的一种特殊形式.

在上述例子中, 变量 l, V, y 分别由变量 T, p, x 所确定. 因而, 我们称 T, p, x 为 **自变量/输入**, 并称 l, V, y 为 **因变量/输出**. 在这三个例子中, 因变量由自变量的值唯一确定. 变量之间的这种确定的依赖关系称为 **函数关系**.

有些变量之间也有某种依赖关系, 但不是确定的关系. 比如, 某些疾病 (如心脏病, 糖尿病) 与饮食习惯, 运动, 遗传, 环境因素等有关, 但这些因素并不能唯一确定是否会患病. 许多因素的交互作用和随机性决定了疾病的发生风险, 因此不构成确定的函数关系. 在微积分中, 我们只讨论变量之间的那种确定的依赖关系, 即函数关系. 重点在于, 函数关系中没有随机性和随机现象, 输入一旦确定了输出也就被唯一地确定了. 函数的确切定义如下:

定义 1.2.1: 函数

设 x 与 y 是两个变量, 分别在实数集合[a] X 与 Y 中取值. 假如有一种规则[b] f, 使得对于每个数 $x \in X$, 都能找到唯一确定的数 $y \in Y$ 与之相对应, 则我们称 f 是一个 **函数 (function)**, 记作

$$f : X \to Y,$$

并称 X 为 f 的 **定义域 (domain)**, 称 Y 为 f 的 **到达域 (codomain)**. 这里的 x 称为自变量, y 称为因变量. 这时, 也称 y 是 x 的函数. 正确理解该定义的核心要点在于:

- 集合 X 中的 *每个* $x \in X$ 都能按照规则 f, 来找到集合 Y 中的某个元素 $y \in Y$. 如果定义域中某个 x 按照规则 f 找不到对应的 y, 则须把它从定义域中人为地剔除, 否则这个函数不是合法定义的.

- 对于 x 来讲, 对应的 y 有且仅有唯一一个. 意味着, 对于某个固定的 x 来讲, 对应的 y 不会有两个或者多个, 且不会变化或改变. 但是通常, 不同的 x 是可以对应于同一个 y 的.

这里与 x 相对应的 y 称作 f 在点 x 的函数值, 记作 $f(x)$. 而 Y 中一切可能被取到的函数值组成的集合称作 f 的 **值域 (range)**, 记作 $f(X)$. 显然, $f(X)$[c]是 Y 的一个子集合:

$$f(X) := \{f(x) \in Y \mid x \in X\} \subseteq Y.$$

函数 $f : X \to Y$ 通常也记为

$$y = f(x), \quad x \in X,$$

或者简单地记为 $y = f(x)$ 或 $f(x)$.

[a] 指的是由实数构成的集合, 不一定是 $\mathbf{R}$, 但是是 $\mathbf{R}$ 的子集. 学习后面的定义 1.2.3 可知, 把 X, Y 限定为实数集这一条件大可不必.

[b] 这里的 "规则" 是语言, 逻辑, 哲学等角度理解的广泛规则, 通常使用数学表达式来表达, 但也可以不是.

[c] 另一个完全等价的书写是 $f(X) := \{y \in Y \mid 存在某个 x \in X, y = f(x)\}$.

现在, 对定义 1.2.1 做几点解释.

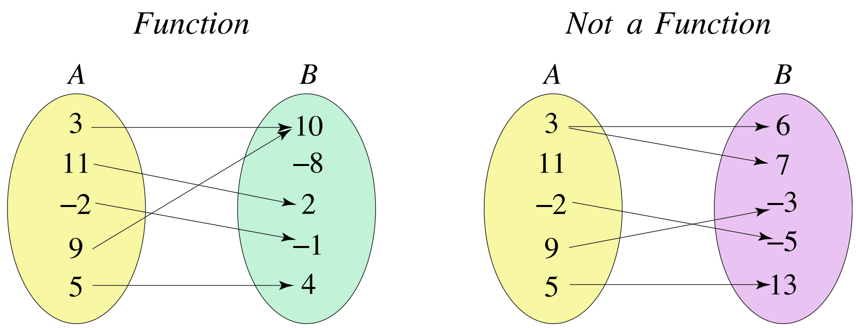


图 1.1: 函数的定义

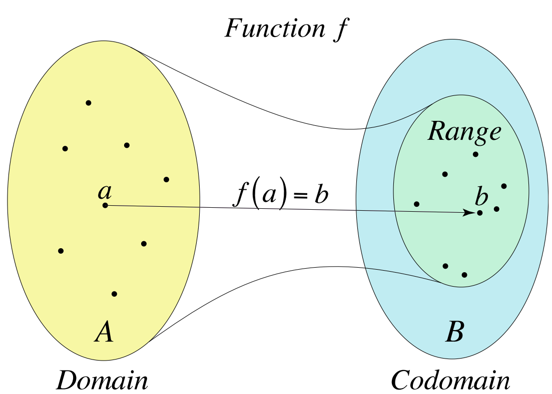


图 1.2: 到达域 (codomain) 和值域 (range)

1. 所有的函数都能通过数学表达式来定义吗?

   通常见到的函数多数是由单个表达式给出的, 比如 $s=\frac{1}{2}gt^2$, 其中 t 到 s 的对应关系是由一个表达式给出的[5]. 有时用一个表达式不够, 要用几个表达式给出, 正像我们在前面关于快递员纳税的例子中看到的. 这种函数通常称作 **分段函数 (piece-wise function)**. 但是, 并不是所有的函数都可以用 (一个或几个) 数学表达式给出的. 比如, 股市上某个交易日某种股票的价格显然是时间的函数. 我们可以用图形表示它的起伏状况, 但无法用一个表达式来表达. (你要能表达出来, 你就是股神了, 因为你能预测未来的股价.)

2. 函数关系是因果关系吗?

   函数关系并不总是等同于因果关系, 虽然在某些情况下它们可能会重合. 比如, 股票价格和时间并不存在因果关系; 再比如, 每个同学都有自己的身份证号码, 根据这个号码, 我就能确定你的姓名. 但是身份证号码不是父母给我起这个名字原因.

3. 函数的定义域是谁来决定的?

   函数的定义域是根据具体函数的定义来确定的. 对于一个由表达式给出的函数, *通常认为 (如果是考试的话这是唯一标准) 其定义域是使得表达式有"数学意义"的自变量的一切值, 我们称其为自然定义域.* 比如, $y=\sqrt{x-1}$ 的定义域是 $\{x \mid x \geqslant 1\}$. 但是, 对于在物理问题或其他问题中提出的函数, 其定义域要根据所讨论的问题的实际应用来确定. 比如 $s=\frac{1}{2}gt^2$ 的定义域是 $\{t \mid t \geqslant 0\}$ 因为时间 $t<0$ 没有物理意义. 谁都可规定函数的定义域, 只要合理即可. 在微积分中所讨论函数的定义域通常是一个闭区间 $[a,b]$, 或开区间 (a,b), 或半开半闭区间 $[a,b),(a,b]$, 有时是无穷区间甚至整个数轴 $\mathbf{R}$.

4. 到达域和值域是一回事吗?

   因变量 y 的变化范围 Y (到达域) 不一定恰好就是函数的值域 $f(X)$, 如图 1.2. 比如, 定义在 $X=\{x \mid x \geqslant 1\}$ 上的函数 $y=f(x)=\sqrt{x-1}$ 的值域是 $\{y \mid y \geqslant 0\}$, 但我们仍然可以认为 y 的变化范围是 $\mathbf{R}$, 并记该函数为 $f: X \to \mathbf{R}$. 其实, 你完全可以根据自己的喜好任意去设定到达域, 只要合理就行. 通常 $\mathbf{R}$ 是我们最常用的到达域, 因为它几乎是万能的. 如果你能找到函数 f 的值域, 然后你就可以把这个值域规定成到达域. 此时你会发现什么都没有受到影响. 但是你不能进步一缩小到达域了, 否则定义域中某些 x 无法取到对应的 y. 我们又发现, 值域是受到定义域的影响, 如果我们改变定义域, 则在同一规则下 f 的值域可能不同. 比如, 定义在 $X=\{x \mid x \geqslant 4\}$ 上的函数 $f(x)=\sqrt{x-1}$ 的值域是 $\{y \mid y \geqslant 2\}$, 而不是 $\{y \mid y \geqslant 0\}$. 所以我们把 f 的值域写作 $f(X)$ 以表示这一相关性.

猜想 1.2.1

既然不同的定义域 X 会得到一个对应的值域 $f(X)$, 那么这种关系是不是一个函数呢? 那么它的定义域和到达域又是什么呢?

例 1.2.1 在中学时, 我们已经接触过序列 (也称数列) $\{a_n\}$, 它是依次排列起来的一串实数:

$$a_1, a_2, \cdots, a_n, \cdots$$

[5]表达式 $s=\frac{1}{2}gt^2$ 是物理中自由落体运动的位移公式之一. s 表示物体的位移 (通常是从初始位置到某一时刻的位置, 单位可以是米). g 表示重力加速度, 在地球表面通常取 $g \approx 9.8\ \mathrm{m/s^2}$. t 表示时间, 从物体开始自由落体运动的时刻算起 (单位是秒).

其中 a_n 称为序列的通项. 对于任意一个正整数 n, 我们都有唯一确定的一个数 a_n 与之对应. 在这种看法之下, 序列 $\{a_n\}$ 便是定义在正整数集 $\mathbf{N}^*$ (全体正整数组成的组合) 上的一个函数.

补充内容: 更具定义, 构成一个函数必须具备以下三个要素：定义域 X, 到达域 Y 和对应的规则 f. 所以说两个函数相等, 就是说构成他们的三个要素都分别相等.

习题 1.2.1 下列各题中, 函数 $f(x)$ 和 $g(x)$ 是否相同[6]? 为什么?

1. $f(x)=\lg x^2, g(x)=2\lg x$;
2. $f(x)=x, g(x)=\sqrt{x^2}$;
3. $f(x)=\sqrt[3]{x^4-x^3}, g(x)=x\sqrt[3]{x-1}$;
4. $f(x)=1, g(x)=\sec^2 x-\tan^2 x$.

1.2.2　基本初等函数

在中学数学课本中, 我们已经遇到过很多函数, 如三角函数, 幂函数, 指数函数, 反三角函数, 对数函数等. 这些函数都是基本初等函数. 通常, 我们把下面六类函数称作 **基本初等函数**:

1. 常数函数: $y=c$ (c 为常数), 即无论自变量 x 为何值, 其函数值总是 c. 显然, 常数函数 $y=c$ 的定义域为 $\mathbf{R}$.
2. 幂函数: $y=x^\alpha$ $(\alpha\neq 0)$.
   (a) 当 α 为正整数时, 其定义域为 $\mathbf{R}$;
   (b) 当 α 为负整数时, 其定义域为 $\mathbf{R}\backslash\{0\}$;
   (c) 当 α 为有理数时, $\alpha=\dfrac{n}{m}$, 其中 $m,n\in\mathbf{Z}, m>0, (m,n)=1$, 这时我们认为
   $$x^{\frac{n}{m}}=(x^n)^{\frac{1}{m}}.$$
   因此, 当 α 为有理数 $\dfrac{n}{m}$ 时, 函数 $y=x^\alpha$ 的定义域依赖于 α 的符号以及分母 m 的奇偶性.
      i. 当 $\alpha>0$, 而 m 为奇数时, 其定义域为 $\mathbf{R}$;
      ii. 当 $\alpha>0$, 而 m 为偶数时, 其定义域为 $[0,+\infty)$.
   (d) $\alpha<0$ 的情况留给读者自己考虑.
   (e) 当 α 为无理数时, x^α 被理解为 x^{α_n} 的极限, 其中 $\alpha_n(n=1,2,\cdots)$ 是任意一串趋向于 α 的有理数. 由于 α_n 表示成分数式时其分母可能出现偶数, 所以要求 x 非负. 因此,
      i. 当 α 为正无理数时, $y=x^\alpha$ 的定义域为 $[0,+\infty)$;
      ii. 当 α 为负无理数时, 其定义域为 $(0,+\infty)$.

   不论上述哪种情况, 幂函数 $y=x^\alpha$ 的定义域总包含 $(0,+\infty)$.
3. 指数函数: $y=a^x$ $(a>0, a\neq 1)$. 指数函数 $y=a^x$ 的定义域为 $\mathbf{R}$.

[6]只有第 3 个相等.

4. 对数函数: $y = \log_a x$ $(a > 0, a \neq 1)$. 对数函数 $y = \log_a x$ 的定义域为 $(0, +\infty)$. 在本教材中, $\ln x$ 表示以 e 为底的自然对数.

5. 三角函数: $y = \sin x, y = \cos x, y = \tan x, y = \cot x, y = \sec x$, $y = \csc x$.

   - 正弦函数 $y = \sin x$ 与余弦函数 $y = \cos x$ 的定义域均为 $\mathbf{R}$,
   - 正切函数 $y = \tan x$ 与正割函数 $y = \sec x$ 的定义域均为 $\mathbf{R} \backslash \left\{ \left(n + \dfrac{1}{2} \right) \pi \middle| \, n \in \mathbf{Z} \right\}$,
   - 余切函数 $y = \cot x$ 与余割函数 $y = \csc x$ 的定义域均为 $\mathbf{R} \backslash \{ n\pi \mid n \in \mathbf{Z} \}$.

   在微积分中, 三角函数中的角度用弧度制表示.

6. 反三角函数: $y = \arcsin x, y = \arccos x, y = \arctan x, y = \operatorname{arccot} x$.

   - 反正弦函数 $y = \arcsin x$ 与反余弦函数 $y = \arccos x$ 的定义域均为 $[-1, 1]$,
   - 反正切函数 $y = \arctan x$ 与反余切函数 $y = \operatorname{arccot} x$ 的定义域均为 $(-\infty, +\infty)$.

我们假定读者通过中学数学课程对上述六类基本初等函数的性质及其图形已有足够的了解. 因此, 本教材略去关于它们的性质的叙述及其图形的描述. 如果想熟悉他们的性质, 强烈建议使用画图工具 Geogebra 自行探索.

现在我们来定义复合函数. 复合函数是本课程重点中的重点.

定义 1.2.2: 复合函数

假定我们有两个函数 $f : X \to Y$ 及 $g : Y^* \to Z$, 并且假定 $f(X) \subset Y^*$. 这时, 对于每个数 $x \in X$, 都有唯一确定的数 $y = f(x) \in Y$ 与之相对应. 对于这个数 $y = f(x)$, 由于它一定属于 Y^*, 因而又有唯一确定的数 $z = g(y) \in Z$ 与之相对应. 这样一来, 我们就建立了一个从 x 到 z 的对应, 从而得到一个新的函数. 这个函数称为函数 f 与 g 的 **复合函数 (composition)**, 记作

$$z = g(f(x)) \text{ 或 } g \circ f(x).$$

例 1.2.2 设函数 $f(x) = \sin x$, $g(y) = \mathrm{e}^y$, 则

$$g(f(x)) = \mathrm{e}^{\sin x}, \quad x \in \mathbf{R}.$$

交换复合的顺序, 则

$$f(g(y)) = \sin(\mathrm{e}^y), \quad y \in \mathbf{R}.$$

请用 GeoGebra 作图并观察这两个复合函数的图像是否一致.

例 1.2.3 设函数 $f(x) = x^2$, $g(y) = \sqrt{y}$, 则

$$g(f(x)) = \sqrt{x^2} = |x|, \quad x \in \mathbf{R}.$$

交换复合的顺序, 则

$$f(g(y)) = y, \quad y \geqslant 0.$$

例 1.2.4 设函数 $f(x)=\sqrt{x}$, $g(x)=\ln(x)$, 则

$$f\circ g(x)=\sqrt{\ln(x)},\quad x\in[1,+\infty).$$

要使得复合函数 $f\circ g(x)=\sqrt{\ln(x)}$ 有定义, 首先需要保证 $\ln(x)\geqslant 0$. 但我们知道, 只有当 $x\geqslant 1$ 时, $\ln(x)\geqslant 0$. 因此, 复合函数的定义域是 $[1,+\infty)$. 交换复合的顺序, 则

$$g\circ f(x)=\ln(\sqrt{x}),\quad x\in(0,+\infty).$$

请用 GeoGebra 作图并观察这两个复合函数的图像是否一致.

我们在定义复合函数时要小心谨慎. 不是所有两个函数之间都可以进行复合操作. 如果 f 的值域和 g 的定义域完全没有交集, 则 $g\circ f(x)$ 完全就没有意义. 同理, 如果 g 的值域和 f 的定义域完全没有交集, 则 $f\circ g(x)$ 完全就没有意义. 一个例子是:

$$f(x)=e^x \text{ 的值域是}(0,+\infty),\ g(x)=\ln(-x) \text{ 的定义域是}(-\infty,0).$$

尝试定义复合函数 $g\circ f(x)=\ln\left(-e^x\right)$ 时则会出现问题. 该函数对于任何 x 都是没有意义的. 但如果我们交换了复合的顺序, 则他们的复合函数可能就有意义了. 还是同样的函数 f 和 g, 我们发现:

$$g(x)=\ln(-x) \text{ 的值域是}\mathbf{R},\ f(x)=e^x \text{ 的定义域是}\mathbf{R}.$$

此时, 尝试定义复合函数 $f\circ g(x)=e^{\ln(-x)}$ 则没有任何问题, 其定义域为 $(-\infty,0)$. 综合以上讨论和例题, 对于复合函数的理解, 我们有如下总结:

- 复合函数的定义依赖于一个函数[7]的值域和另一个函数[8]的定义域有没有交集 (多数情况下, 一个函数的值域可能会完全落入另一个函数的定义域; 也有可能仅有部分的相交). 特别是, 如果这两个集合没有交集, 就无法复合出一个有效的复合函数, 即不存在复合函数.
- 复合函数的结果与复合的顺序息息相关. 不同的顺序其定义域, 值域, 以及函数的对应规则本身都有可能不同. 甚至, 按某个顺序复合, 则其复合函数不存在, 若交换顺序则其复合函数又存在了.
- 复合函数的定义域多数情况下 (但不是全部) 由第一个作用的函数的定义域决定, 正例包括例 1.2.2, 例 1.2.3 和例 1.2.5, 反例是例 1.2.4. 一个很好的练习是去思考为什么不是全部呢? 复合函数的值域通常受到最外侧函数的影响最大. 比如, 我们观察 $\sin\left(\mathrm{e}^x\right)$, 或者 $\sin(f(x))$. 我们不管 f 是什么, 都可以看一眼瞬间就得到一个初步但正确的结论: $\sin(f(x))$ 的值域在 -1 到 1 之间. 这是一个重要的技巧.

例 1.2.5 设函数 $f(x)=\sin x$, $g(x)=\arcsin x$, 求复合函数 $f\circ g(x), g\circ f(x)$ 的定义域和值域, 并画出这两个复合函数的图形.

证明. 我们依次考虑. 见图 1.3.

1. 复合函数 $f\circ g(x)=\sin(\arcsin(x))$

[7]指的是内侧的函数, 即作用于 x 的第一个函数.
[8]指的是外侧的函数, 即作用于 x 的第二个函数.

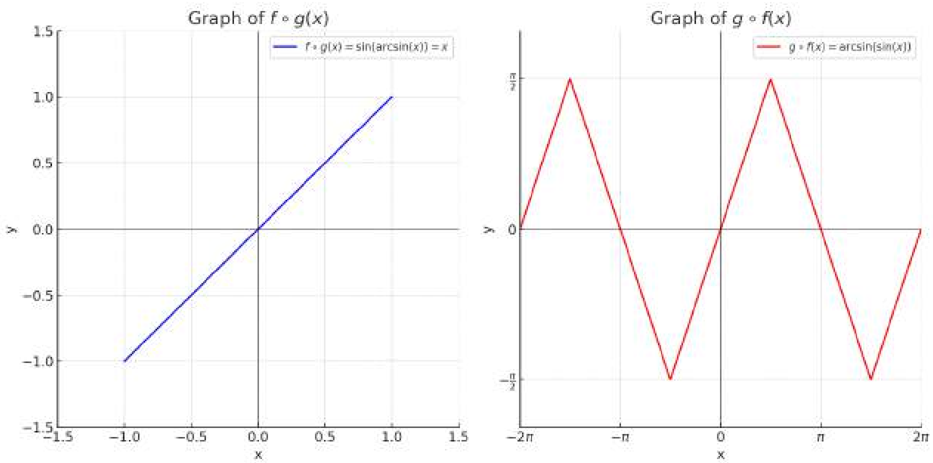


图 1.3: 不同顺序下的复合结果, 见例 1.2.5

- 定义域: 函数 $g(x) = \arcsin(x)$ 的定义域是 $[-1, 1]$, 因此复合函数 $f \circ g(x) = \sin(\arcsin(x))$ 的定义域也是 $[-1, 1]$.
- 值域: 对于任意 $x \in [-1, 1]$, 有 $f \circ g(x) = \sin(\arcsin(x)) = x$, 因此值域也是 $[-1, 1]$.
- 图形: 由于 $f \circ g(x) = x$, 其图像是一条斜率为 1 的直线, 通过原点, 范围在 $[-1, 1]$.

2. 复合函数 $g \circ f(x) = \arcsin(\sin(x))$

- 定义域: 函数 $f(x) = \sin(x)$ 的定义域是 $(-\infty, +\infty)$, 且它的值域是 $[-1, 1]$, 因此复合函数 $g \circ f(x) = \arcsin(\sin(x))$ 的定义域为 $(-\infty, +\infty)$.
- 值域: 由于 $g(x) = \arcsin(x)$ 的值域为 $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, 且函数 $g \circ f(x)$ 是一个以 2π 为周期的周期函数, 所以它的值域为 $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.
- 周期性分析: 此时建议先看后面的例 1.2.10 弄清楚反三角函数的一些事实.
  (a) 当 $x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, 有 $g \circ f(x) = \arcsin(\sin(x)) = x$.
  (b) 当 $x \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$, 有 $g \circ f(x) = \arcsin(\sin(x)) = \pi - x$. 根据例 1.2.10, arcsin 只是定义在 $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ 上 sin 的反函数, 而不是定义在 $\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$ 上 sin 的反函数. 我们做如下变换: 根据三角恒等式, 我们有 $g \circ f(x) = \arcsin(\sin(x)) = \arcsin(\sin(\pi - x))$. 令 $y := \pi - x$, 由 $x \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$ 可以得到, $y \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. 因此, $\arcsin(\sin(y)) = y = \pi - x$. 最后有, $g \circ f(x) = \arcsin(\sin(x)) = \pi - x$.
- 图形: 函数 $g \circ f(x)$ 是一个周期为 2π 的分段函数. 在 $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ 区间内, 其图像为 $y = x$; 在 $\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$ 区间内, 其图像为 $y = \pi - x$, 依次类推.

□

例 1.2.6 证明 $\sin(\arccos x) = \sqrt{1 - x^2}$.

证明. 对于任意的 $x \in [-1,1]$, 记 $y = \arccos x$, 则 $x = \cos y, y \in [0,\pi]$, 且

$$\sin(\arccos x) = \sin y = \sqrt{1-\cos^2 y} = \sqrt{1-x^2}.$$

更多类似的结果请看附录 .3. 这些结论在未来会有用. □

由有限个基本初等函数经过有限次四则运算及复合运算所得到的函数, 称为 **初等函数**. 上面两个例子中的函数都是初等函数. 再比如, $y = \sum\limits_{k=1}^{n} \dfrac{\sin kx}{k}$ 是初等函数. 并非所有函数都是初等函数, 如今后经常会遇到的 **符号函数 (sign function)** 就是一个例子:

$$y = \operatorname{sgn} x = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$$

以后我们将会说明这个函数不是初等函数.

例 1.2.7 设函数 $y = \operatorname{sh} x := \dfrac{\mathrm{e}^x - \mathrm{e}^{-x}}{2}$, $y = \operatorname{ch} x := \dfrac{\mathrm{e}^x + \mathrm{e}^{-x}}{2}$, 它们分别称为 **双曲正弦函数**, **双曲余弦函数**[9]. 他们都是 **双曲函数 (hyperbolic functions)** 的一种. “sh” 是”sinh” 的简写, 发音是类似于单词 shine; “ch” 是”cosh” 的简写, 发音是类似于单词 coach; 词尾的字母 h 就是 hyperbolic. 他们和我们熟悉的 sin, cos 有这类似的下列性质.

1. $\operatorname{ch}^2 x - \operatorname{sh}^2 x = 1$;
2. $\operatorname{sh}(x \pm y) = \operatorname{sh} x \operatorname{ch} y \pm \operatorname{ch} x \operatorname{sh} y$;
3. $\operatorname{ch}(x \pm y) = \operatorname{ch} x \operatorname{ch} y \pm \operatorname{sh} x \operatorname{sh} y$[10].

证明. 上述 3 个等式的证明详见课本 P22 例 6. 其实我们最熟悉的 cos 和 sin 也有类似的表达式. 即下面的恒等式:

$$\cos x = \frac{e^{ix} + e^{-ix}}{2}, \ \sin x = \frac{e^{ix} - e^{-ix}}{2i}.$$

因此, 我们发现 $\cos ix = \operatorname{ch} x$, $\sin ix = i \operatorname{sh} x$. 这也难怪我们会得到类似的性质了. □

例 1.2.8 设函数 $y = x - [x]$, 其中 $[x]$ 表示不超过 x 的最大整数, 例如 $[5] = 5, [\pi] = 3, [-\mathrm{e}] = -3$. 显然, 该函数的值域为 $[0,1)$. 当 $0 \leqslant x < 1$ 时, $y = x - [x] = x$; 当 $1 \leqslant x < 2$ 时, $y = x - [x] = x - 1$; 依此类推. 这个函数也不是初等函数, 但它有很多用途, 在数论中常用到它, 在某些工程计算中也会用到它. 在数学文献中把这个函数记作 $y = \{x\}$, 并把 $\{x\} \equiv x - [x]$ 称为 x 的小数部分. 不过这里我们要提醒读者: 当 $x > 0$ 时, $\{x\}$ 就是人们通常理解的 x 的小数部分, 比如 $\{3.14\} = 0.14$; 但当 $x < 0$ 时, 情形就不同了, 比如 $x = -3.14$, 这时 $[x] = -4$, 故 $\{-3.14\} = -3.14 + 4 = 0.86$, 而不是 -0.14. 容易验证下面两条性质:

[9] 类似地, 还可以定义双曲正切函数 $y = \operatorname{th} x := \dfrac{\operatorname{sh} x}{\operatorname{ch} x}$, 双曲余切函数 $y = \operatorname{coth} x := \dfrac{\operatorname{ch} x}{\operatorname{sh} x}$, 双曲正割函数 $y = \operatorname{sech} x := \dfrac{1}{\operatorname{ch} x}$, 双曲余割函数 $y = \operatorname{csch} x := \dfrac{1}{\operatorname{sh} x}$.

[10] 注意, $\cos(x \pm y) = \cos x \cos y \mp \sin x \sin y$, 所以第 3 个等式稍有不同.

1. 对任意实数 x, 有 $[x] \leqslant x < [x]+1$;

2. $y = x - [x]$ 是以 1 为周期的函数.

例 1.2.9 下面的函数称为狄利克雷 (Dirichlet) 函数, 或称为狄氏函数:

$$\mathrm{D}(x) = \begin{cases} 1, & x \text{ 为有理数}, \\ 0, & x \text{ 为无理数}. \end{cases}$$

人们很难画出这个函数的图形. 因为在任意的小区间中都有有理数和无理数 (见课本习题 1.1 中的第 7 题与第 8 题), 所以其图形是分别分布在 x 轴上及直线 $y=1$ 上 "密密麻麻" 的点. 可见, 函数的图形可以不是由一条或若干条曲线组成的. 狄利克雷函数常被用来澄清某些概念

函数概念的一般化就是集合之间的映射. 下面介绍有关映射的一些术语.

定义 1.2.3: 映射

我们说 $f: E \to F$ 是集合 E 与 F 之间的一个 **映射 (map)**, 如果对于每个元素 $x \in E$ 都有唯一确定的元素 $y \in F$ 与之相对应. 这时, 我们将 y 记作 $f(x)$, 并称之为 x 的 **像 (image)**, 或者像点. 全体像点组成的集合

$$f(E) := \{y \in F \mid \text{ 存在} x \in E, \text{ 使得} y = f(x)\}$$

称为 E 的像集合或值域. 显然, $f(E) \subset F$.

函数是一般映射的特殊情况. 函数与一般映射的区分在于前者要求其定义域与值域都是数集合. 但实际上, 这么区分意义不大. 以后我们说函数就是指映射, 映射就是指函数.

注意 映射又称为算子 (operator). 根据集合 X, Y 的不同情形, 在不同的数学分支中, 映射又有不同的惯用名称. 例如, 从非空集 X 到数集 Y 的映射又称为 X 上的泛函 (functional), 从非空集 X 到它自身的映射又称为 X 上的变换 (transform), 从实数集 (或其子集) X 到实数集 Y 的映射通常称为定义在 X 上的函数 (function).

定义 1.2.4: 满射, 单射, 全射, 反函数

下面关于函数的定义极其重要:

- 若映射 $f: E \to F$ 的值域 $f(E) = F$, 则表明 F 中的每个元素都是一个像点. 这时, 我们称 $f: E \to F$ 为 **满射 (surjection, surjective map)**. 说人话就是: 函数 f 是满射, 当且仅当值域等于到达域.

- 若映射 $f: E \to F$ 具有下列性质, 则称 f 为 **一一映射 (one-to-one)** 或 **单射 (injection, injective map)**[a], 即[b]

$$\forall x_1, x_2 \in E, x_1 \neq x_2 \Longrightarrow f(x_1) \neq f(x_2).$$

- 当映射 $f: E \to F$ 既是满射又是单射时, 它被称为**双射 (bijection, bijective map)** 或 **一一对应 (one-to-one correspondence)**. 在这种情况下, 映射 f 具有以下两个重要性质:

  - 对目标集合 F 的覆盖性 (满射保证): 对于目标集合 F 中的每个元素 y, 都存在至少一个定义域 E 中的元素 x 使得 $f(x) = y$.
  - 唯一性 (单射保证): 对于目标集合 F 中的每个元素 y, 只能有一个唯一的定义域 E 中的元素 x 对应到它, 即 $f(x) = y$.

  因此, 双射中, 对于到达域 F 中的每个元素 y, 都有且仅有唯一的 $x \in E$, 使得 $f(x) = y$. 换句话说, 每个元素都一一对应. 这就自然形成一个自 F 到 E 的映射. 这个映射称为 f 的 **逆映射 (inverse map)**, 或者 **反函数**, 记作 $f^{-1}: F \to E$. 所以 f 是双射也被叫做 f 是 **可逆的 (invertible)**, 即反函数存在.

[a] 如果定义域 E 中只有一个元素, 映射 f 依然被视为单射, 因为在这种情况下, 不存在违反单射定义的情况. 单射的条件在定义域为单个元素时自动满足.

[b] 对于符号 $\forall$, 详见附录 .4.

讨论一个函数是不是满射, 是不是单射, 这与该函数的定义域/到达域密切相关. 脱离具体的定义域/到达域去讨论它是否是满射或者单射, 纯粹没有意义. 我们有以下观察:

1. 如果刻意地选择将值域作为函数的到达域, 那么任何函数都可以变成为满射. 根据函数的定义, 值域不可能是空集, 所以这种操作总是可行的. 反之, 选择一个到达域使得值域成为其真子集, 那么任何函数都可以不是满射.

2. 对于定义域, 通常我们默认选择的是该函数的 "自然定义域" (即数学上有意义的任何输入值), 但是我们仍旧可以刻意地选择某个特殊的区间作为定义域, 此时, 该函数在此定义域上可能就变成了 (或者变不成) 单射. 例 1.2.10 就是一个例子. 一个特殊的例子是常函数 $y = c$ (c 为常数), 无论我们选择任何形式的定义域, 它都变不成单射. 与之相反, 对于线性函数 $y = x$, 无论我们选择任何形式的定义域, 它都是单射.

3. 总结前面两点: 讨论一个函数是不是满射, 这仅与该函数的到达域有关, 并非由函数规则 f 本身决定. 讨论一个函数是不是单射, 这与该函数的定义域和函数规则 f 本身都有关.

4. 通常情况下, 我们遇见的多数函数, 即使其在自然定义域上不是单射, 我们总能找到某个较小的区间, 使得其变成单射. (从图像上看经常表现为单调递减或递增). 如果更进一步刻意地设定值域为到达域, 那么为函数在这个较小的区间上就是可逆的, 我们称之为局部可逆. 下面就是一个例子.

例 1.2.10 设函数 $f(x) = \sin x$. 则 f 在 $\mathbf{R} \to \mathbf{R}$ 上既不是单射也不是满射. 我们知道它的值域是 $[-1, 1]$, 选择值域作为到达域, 则 f 在 $\mathbf{R} \to [-1, 1]$ 上是满射, 却不是单射. 若进一步将 f 的定义域改换成 $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, 也即

$$f : \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \to [-1, 1], \quad x \mapsto \sin x,$$

其中 $x \mapsto \sin x$ 表示点 x 对应的函数值为 $\sin x$[11]. 这时, f 既是满射又单射, 即双射, 因此它还有逆函数存在:

$$f^{-1} : [-1, 1] \to \left[-\frac{\pi}{2}, \frac{\pi}{2}\right], \quad x \mapsto \arcsin x.$$

注意, 如果逆函数存在, 则逆函数的定义域 (值域) 正好是原来的函数的值域 (定义域), 互相颠倒而已. 我们再考虑另外一个定义域上的 sin 函数:

$$f' : \left[\frac{\pi}{2}, \frac{3\pi}{2}\right] \to [-1, 1], \quad x \mapsto \sin x,$$

注意, 虽然 f' 和 f 都是 sin 函数, 但是两者的定义域不同, 其函数图像也不同. *通常我们默认的 arcsin 是定义在区间 $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ 上的 sin 函数的反函数, 而不是定义在别的区间的 sin 函数. 所以 arcsin 并不是 f' 的反函数.*

再次强调, 脱离具体的定义域/到达域去讨论一个函数是不是满射, 是不是单射, 是不是双射 (即反函数是否存在) 的行为是没有意义的.

例 1.2.11 正像大家所熟知的, $y = \ln x$ 是 $x = \mathrm{e}^y$ 的反函数. 由于习惯上以 x 为自变量, y 为因变量, 所以通常说 $y = \ln x$ 是 $y = \mathrm{e}^x$ 的反函数.

可以像定义复合函数一样来定义复合映射, 比如映射 $f : E \to F$ 及 $g : F^* \to G$ (假定 $F \subset F^*$) 的复合映射 $g \circ f$ 是 $E \to G$ 的一个映射, 它将每个元素 $x \in E$ 映为 $g(f(x))$.

以下是更贴近生活, 易于理解的单射和满射的例子:

1. 单射 (Injective) 的特点是每个输入都有唯一的输出, 且不同的输入不会映射到相同的输出.
   (a) 例子 1: 每个联系人都有一个独立的手机号码, 两个不同的联系人不可能有相同的手机号码. 但并不是每个手机号码都必须被使用. 有些号码可能还没有被分配给任何联系人, 这也是单射.
   (b) 例子 2: 每个学生有唯一的学号, 但并不是所有学号都必须分配给学生. 也就是说, 不同的学生有不同的学号, 没有两个学生可以有相同的学号, 这是单射的.
2. 满射 (Surjective) 意味着所有输出都有对应的输入, 即每个输出值都至少有一个输入值与之对应.
   (a) 例子 1: 在一台饮料机上, 每个按钮都对应一种饮料. 有时候可能有多个按钮能选择相同的饮料 (比如两种不同的组合键都出可乐), 但是饮料机中的每种饮料都至少有一个按钮可以选择到. 这里, 按钮到饮料的映射是满射的, 因为每种饮料都能被至少一个按钮选到.
3. 双射 (Bijection) 是单射和满射的结合, 意味着每个输入都有且只有一个输出, 反之亦然.
   (a) 例子 1: 假设有一间教室里有 20 个座位, 20 个学生需要坐下. 如果每个学生只能坐在一个座位上, 且每个座位只能分配给一个学生, 最终每个学生都有一个座位, 每个座位都有一个学生对应, 这就是一个双射的关系.

[11] 符号 $\mapsto$ 这种用法是常见的, 它关心的是函数 f 具体的变化规则. 要区别它和符号 $\to$, 它关心的是定义域和到达域.

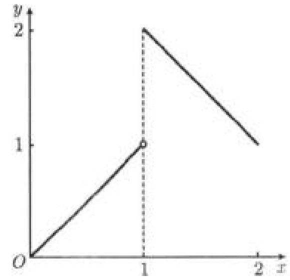


图 1.4: 一一对应的函数不一定是严格单调的

1.2.3 反函数

面前定义的反函数不够严谨, 现在我们重新学习反函数.

定义 1.2.5: 反函数的其他定义方式

反函数的定义有多重等价形式, 下面我们学习常见的两种:

1. 给定 $f: X \to Y$, 若对任意的 $y \in Y$, 方程 $f(x) = y$ 在 $x \in X$ 上有且仅有一解, 则由此定义一个从 Y 到 X 的函数, 称为 f 的反函数, 记做 $f^{-1}: Y \to X$.
2. 给定函数 $y = f(x)$, 其定义域和值域分别记做 X 和 Y, 若存在函数 $g: Y \to X$, 满足
$$g(f(x)) = x, \quad \text{对任意的} x \in X,$$
则称 g 为 f 的反函数, 记做 $f^{-1} = g: Y \to X$.

注意, $y = f(x)$ 与 $y = f^{-1}(x)$ 的图形关于直线 $y = x$ 对称.

命题 1.2.1

$f: X \to Y$ 有反函数的充分必要条件是 f 在 $X \to Y$ 是一一对应的.

例 1.2.12 若 $f(x)$ 在 X 上严格单调, 则 f 在 X 上是一一对应的, 所以 f 的反函数存在. 但是一一对应的函数不一定是严格单调的. 例如函数

$$f(x) = \begin{cases} x, & 0 \leqslant x < 1, \\ 3 - x, & 1 \leqslant x \leqslant 2 \end{cases}$$

在 $[0, 2]$ 上既是单射又是满射 (图 1.4), 所以是一一对应的, 但不是单调函数, 这个函数的反函数仍是它自己.

1.2.4 有界函数

最后, 我们讨论有界函数的概念作为本节的收尾.

- 我们称函数 $f: X \to Y$ 是 **有上界的 (bounded from above)**, 如果存在一个实数 M, 使得

$$f(x) \leqslant M, \quad \forall x \in X.$$

这时, M 就称为 f 的一个 **上界 (upper bound)**.

> **注释 1.2.1** 显然, 函数 $y = \sin x$ 及 $y = -\mathrm{e}^x$ 在其定义域 $\mathbf{R}$ 上是有上界的. 这是因为 $\sin x \leqslant 1$, 而 $-\mathrm{e}^x \leqslant 0$. 但函数 $y = x^2$ 在其定义域 $\mathbf{R}$ 上则是无上界的. 在这里, 1 是 $\sin x$ 的一个上界, 而且任何一个大于 1 的数也是 $\sin x$ 的上界. 可见, 有上界的函数的上界不是唯一的, 可以有无穷多个.

- 类似地, 可定义有下界的函数. 若存在一个实数 N, 使得函数 $f: X \to Y$ 满足

$$f(x) \geqslant N, \quad \forall x \in X,$$

则称 f 是 **有下界的 (bounded from below)**, 并称 N 是 f 的一个 **下界 (lower bound)**.

> **注释 1.2.2** 在前面提到的例子中, 函数 $y = \sin x$ 及 $y = x^2$ 在其定义域 $\mathbf{R}$ 上是有下界的. 这是因为 $\sin x \geqslant -1, x^2 \geqslant 0$. 但函数 $y = -\mathrm{e}^x$ 在其定义域 $\mathbf{R}$ 上是没有下界的. 请读者画出 $y = -\mathrm{e}^x$ 的草图 (用 GeoGebra 画图), 从中立即可看出这一事实.

- 既有上界又有下界的函数称为 **有界函数 (bounded function)**. 换句话说, 我们称 $f: X \to Y$ 是有界函数, 如果存在两个实数 M 与 N, 使得

$$N \leqslant f(x) \leqslant M, \quad \forall x \in X.$$

从直观上看, 一个有界函数的图形是介于两条水平线 $y = M$ 及 $y = N$ 之间的.

命题 1.2.2: 有界函数的等价定义

函数 $f: X \to Y$ 为有界函数的充要条件是, 存在一个常数 C, 使得

$$|f(x)| \leqslant C, \quad \forall x \in X.$$

证明. 必要性: 若函数 f 是有界的, 则存在两个实数 M 和 N, 使得 $N \leqslant f(x) \leqslant M$ 对任意 $x \in X$ 成立. 于是, $|f(x)| \leqslant \max(|M|, |N|)$, 令 $C = \max(|M|, |N|)$, 则有 $|f(x)| \leqslant C$. 充分性: 反之, 若存在常数 C 使得 $|f(x)| \leqslant C$, 则有 $-C \leqslant f(x) \leqslant C$, 故函数 f 是有界的. □

> **注意** 通常, 如果一个定义有多个不同的表述, 且而他们之间是等价的 (即, 互为充分必要条件). 我们则可以选择任何一个喜欢的表述作为定义, 将其作为我们的出发点, 从而得到其他等价的结果作为命题结果. 数学中, **定义 (definition)** 是非常灵活多变. 同一个概念在不同的教材中, 其定义通常会有所差别. 学好数学就是要理解这种差别, 知道它们都是同一个事情的不同表述而已.

命题 1.2.3

设 $f: X \to Y, g: X \to Y$ 是两个有界函数, 则 $f \cdot g$ 也是有界函数.

证明. 由于 f 和 g 都是有界函数和命题 1.2.2, 存在常数 C_1 和 C_2, 使得对于所有 $x \in X$, 有 $|f(x)| \leqslant C_1$ 和 $|g(x)| \leqslant C_2$. 因此, 对于所有 $x \in X$,

$$|f(x) \cdot g(x)| = |f(x)| \cdot |g(x)| \leqslant C_1 \cdot C_2.$$

令 $C = C_1 \cdot C_2$, 则 $|f(x) \cdot g(x)| \leqslant C$ 对所有 $x \in X$ 成立, 说明 $f \cdot g$ 是有界函数. □

注意, 上述命题的结论很容易推广: 若 f 和 g 都是有界函数, 则 $f \pm g$ 也是有界函数. 但这一结论对于 f/g 未必成立. 例如, 定义在区间 $[1, \infty)$ 上的 $f(x) = 1$, $g(x) = 1/x$.

这里我们提醒读者: 一个定义在区间 $[a, b]$ 上函数 $y = f(x)$, 虽然对每个点 $x \in [a, b]$, $f(x)$ 都是一个有限数, 但函数 $y = f(x)$ 在 $[a, b]$ 上仍可能是无界的. 请看下面的例子. 函数

$$y = f(x) = \begin{cases} \dfrac{1}{x}, & 0 < x \leqslant a \\ 0, & x = 0 \end{cases}$$

在区间 $[0, a]$ 上有定义, 但它是一个无界函数.

结论 1.2.1

定义在有界的区间 (无论开或闭) 上的函数, 其函数可能是无界的.

只有对区间以及函数本身的性质做更多的要求 (比如函数的连续性), 才能使得其函数值是有界的. 这是未来章节中的内容.

1.2.5 补充知识

除了有界性, 函数还有 **单调性**, **奇偶性**, **周期性**等常见性质. 这些知识想必在高中阶段就已经学过, 未来如果复习要复习它们, 建议阅读同济版《高等数学》(第七版) 第 1 节内容, 其中有详细介绍. 这里我们不过多赘述, 我们假设读者了解单调性, 奇偶性, 周期性的定义是. 最后, 我们补充一些未来可能需要的内容.

定义 1.2.6: 函数的运算

设函数 $f(x), g(x)$ 的定义域依次为 D_f, D_g, 且 $D = D_f \cap D_g \neq \varnothing$, 则我们可以定义这两个函数的下列运算:

- 和 (差): $(f \pm g)(x) := f(x) \pm g(x), \quad x \in D;$
- 积: $(f \cdot g)(x) := f(x) \cdot g(x), \quad x \in D;$
- 商: $\left(\dfrac{f}{g}\right)(x) := \dfrac{f(x)}{g(x)}, \quad x \in D \setminus \{x \mid g(x) = 0, x \in D\}.$

问题　上述定义的函数的运算, 都是把任意符合要求的 2 个函数 (把它们看成一对) 变成另一个函数. 这一过程本身是不是函数呢?

定理 1.2.1

设函数 $f(x)$ 的定义域为 $(-l,l)$, 则必存在 $(-l,l)$ 上的偶函数 $g(x)$ 及奇函数 $h(x)$, 使得

$$f(x)=g(x)+h(x).$$

证明. 做如下定义,

$$g(x):=\frac{1}{2}[f(x)+f(-x)],\quad h(x):=\frac{1}{2}[f(x)-f(-x)].$$

则

$$\begin{aligned}g(x)+h(x)&=f(x),\\g(-x)&=\frac{1}{2}[f(-x)+f(x)]=g(x),\\h(-x)&=\frac{1}{2}[f(-x)-f(x)]=-h(x).\end{aligned}$$

证毕. □

命题 1.2.4

设下面所考虑的函数都是定义在区间 $(-l,l)$ 上的. 证明:

1. 两个偶函数的和是偶函数, 两个奇函数的和是奇函数;
2. 两个偶函数的乘积是偶函数, 两个奇函数的乘积是偶函数, 偶函数与奇函数的乘积是奇函数.

上面命题证明很容易. 其结论类似于正负号相加和相乘的效果.

1.3 序列极限

极限的概念与理论是微积分的基础. 微积分中的两个基本概念导数 (微商) 与定积分, 都是建立在极限概念的基础之上的. 与其说极限的概念产生于自然现象，不如说它是我们认识某些复杂的量的一种方法. 比如圆的面积, 原本我们不会计算, 但我们会计算圆内接正 n 边形的面积 $(n\geqslant 3)$ ，于是就将圆的面积作为内接正 n 边形的面积当 n 无限增大时的极限。我们将在以后的讨论中看到极限概念在微积分中的重要意义。

我们的讨论涉及两类极限：序列的极限与函数的极限。本节只讨论序列的极限。

1.3.1 序列极限的定义

设 $\{a_n\}$ 是一个给定的序列. 我们关心的是, 在 n 无限增大的过程中通项 a_n 的变化趋势; 与之相对, 我们不关心的是, 序列 $\{a_n\}$ 在早期阶段 (即开始的有限个的项, 如 a_1,a_2,a_3 这个 3 个项) 的表现.

我们先看几个具体的例子. 它们的图示见图 1.5.

1. 考虑序列 $\left\{\dfrac{1}{2^n}\right\}$:

$$\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{16},\cdots,\frac{1}{2^n},\cdots$$

该序列每一项都是大于零的, 然而, 当 n 趋向于无穷大时, 通项 $\dfrac{1}{2^n}$ 可以任意接近于零. 我们说零是这个序列的极限. 在这个例子中, 序列中的项是永远达不到其极限的, 但可以任意地接近于它的极限.

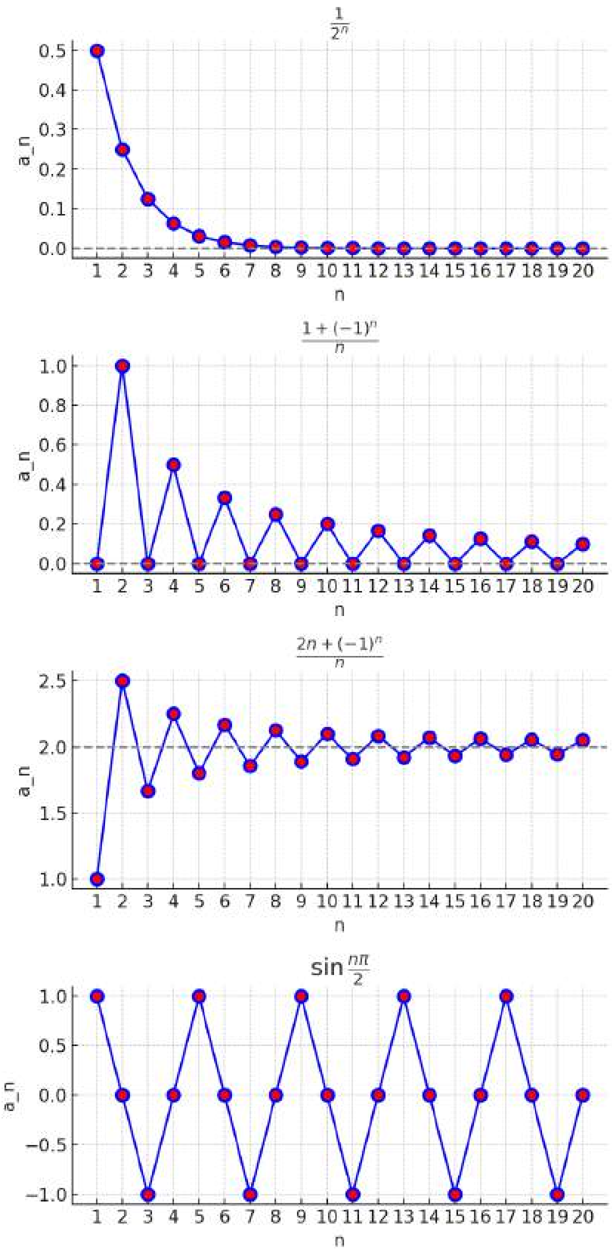


图 1.5: 4 种不同序列的在 n 无限增大的过程中通项 a_n 的变化趋势. 图中的蓝色实线没有任意, 只是为了方便展示他们变化的顺序.

2. 考虑序列 $\left\{\frac{1+(-1)^n}{n}\right\}$:
$$\frac{0}{1},\frac{2}{2},\frac{0}{3},\frac{2}{4},\cdots,\frac{1+(-1)^n}{n},\cdots$$
当 n 无限增大时, 这个序列的通项显然要么本身为零, 要么本身不为零而可任意接近于零. 对于这种情况, 我们仍然把零称作这个序列的极限.

3. 考虑序列 $\left\{\frac{2n+(-1)^n}{n}\right\}$:
$$\frac{1}{1},\frac{5}{2},\frac{5}{3},\frac{9}{4},\frac{9}{5},\cdots,\frac{2n+(-1)^n}{n},\cdots$$
显然, 它的通项可以表示成
$$2+\frac{(-1)^n}{n}.$$
因此, 当 n 无限增大时, 它可以任意接近于 2. 不过, 它不是递减或递增地接近于 2, 而是在 2 的左右摆动. 对于这种情况, 我们仍然称它的极限是 2. 换句话说, 不断摆动地趋向于某个数时, 仍把这个数作为极限.

4. 但是, 并非所有序列都有一个固定的变化趋势. 考虑序列 $\left\{\sin\frac{n\pi}{2}\right\}$:
$$a_n=\sin\frac{n\pi}{2},\quad n=1,2,\cdots$$
当 n 为偶数时, $a_n=0$; 而当 $n=2k+1$ 时, $a_n=(-1)^k$. 可见, 在 n 趋向于无穷大的过程中, a_n 在 $-1,0,1$ 三个值中变来变去, 没有一个固定的趋势. 这时, 我们称 $\{a_n\}$ 没有极限. 注意: 尽管 a_n 大体上确实在围绕着 0 上下变动, 但是却永远不能趋向于 0 或者任意接近于 0.

通过上面这些例子, 我们看到**若序列 $\{a_n\}$ 在 n 趋向于无穷大的过程中有一个确定的趋势, 也就是说, a_n 可任意接近于某个常数 l, 就称 l 为 $\{a_n\}$ 的极限.** 那么什么叫 a_n 可任意接近于 l 呢? 为此, 我们应当考查点 a_n 到点 l 的距离 $|a_n-l|$ 以从严格的数学意义上度量两个量之间的差距. 所谓的 a_n 可任意接近于 l, 即趋向于 l, 是指当 n 充分大时, $|a_n-l|$ 可以任意小. 现在我们以刚才展示过的序列
$$a_n=\frac{2n+(-1)^n}{n}=2+\frac{(-1)^n}{n},\quad n=1,2,\cdots$$
为例. 这时, 极限 $l=2$. 注意到对于任意的 $n=1,2,\cdots$, 我们有
$$|a_n-2|=\left|\frac{(-1)^n}{n}\right|=\frac{1}{n}.$$
因此,

- 当 $n>10$ 时, $|a_n-2|<\frac{1}{10}$;
- 当 $n>10^2$ 时, $|a_n-2|<\frac{1}{10^2}$;
- 一般地, 当 $n>10^k$ 时, $|a_n-2|<\frac{1}{10^k}$. 其中 k 可以是任意大的正整数.

这就是说, $|a_n-2|$ 可以小于任意给定的一个正数 (无论它是多小的正数), 只要[12]其中 n 大于某个正整数.


注释 1.3.1 额外资料: 如何通俗地解释数列极限的定义 | 马同学图解微积分 _ 哔哩哔哩 _bilibili

[12]教材喜欢使用 "xxx, 只要 xxx." 的句式. 我猜测这来自英语 "xxx, if xxx." 习惯就好. 通常可以先读 "只要" 后面的条件再读前面的结果.

下面给出序列极限的正式定义.

定义 1.3.1: 序列的极限

设 $\{a_n\}$ 是一个给定的序列. 若存在一个常数 l, 对于任意给定的正数 ε, 无论它多么小, 都存在一个正整数 $N = N(\varepsilon)$, 使得

$$|a_n - l| < \varepsilon, \text{ 只要 } n > N(\varepsilon),$$

则我们称 $\{a_n\}$ 以 l 为**极限 (limit)**, 或者称 $\{a_n\}$ **收敛 (converge)** 于 l, 记作

$$\lim_{n\to\infty} a_n = l.$$

这时, 也称 n 趋向于无穷大时 a_n 趋向于 l, 记作

$$a_n \to l \quad (n \to \infty).$$

假若序列 $\{a_n\}$ 以某个常数 l 为极限, 则称该序列的**极限存在**.

从直观上来看, 上述定义中的条件实际上是说, 对于任意小的 $\varepsilon > 0$, 都有一个正整数 N, 使得第 N 项之后的所有项 a_n 都满足

$$l - \varepsilon < a_n < l + \varepsilon,$$

即第 N 项之后的所有项 a_n 都落在以 l 为中心以 ϵ 为半径的邻域内. 也就是说, 在 n 无限增大过程中, 总有一个时刻 N, 在此之后点 a_n 到点 l 的距离小于事先任意给定的正数 ε. 显然, 正数 ε 越小, 所要求的 N 就越大.

注意　理解极限定义最重要的一点是: N 要依赖于 ε! N 要依赖于 ε! N 要依赖于 ε! (重要的是说三遍!) 有的时候为了凸显这一事实, 我们会 (不严谨地) 把 N 写作关于 ε 的函数 $N(\varepsilon)$. 即使为了省事不这样子写, 也要牢记它们之间的依赖关系.

显然, 若序列 $\{a_n\}$ 的极限存在, 则其极限值必是唯一的, 因为在 n 无限增大过程中, a_n 不可能同时任意靠近两个不同的数. 下面给出严格证明.

定理 1.3.1: 极限存在的唯一性

若序列的极限存在, 则极限值唯一.

证明. 假设序列 $\{a_n\}$ 的极限存在, 并且它有两个极限 l_1 和 l_2, 即:

$$\lim_{n\to\infty} a_n = l_1 \quad \text{且} \quad \lim_{n\to\infty} a_n = l_2.$$

根据极限的定义, 对于任意 $\varepsilon > 0$, 存在 N_1 和 N_2, 使得当 $n > N_1$ 时, $|a_n - l_1| < \dfrac{\varepsilon}{2}$, 并且当 $n > N_2$ 时, $|a_n - l_2| < \dfrac{\varepsilon}{2}$.[13] 取 $N = \max(N_1, N_2)$, 则对于 $n > N$, 同时满足:

$$|a_n - l_1| < \frac{\varepsilon}{2} \quad \text{和} \quad |a_n - l_2| < \frac{\varepsilon}{2}.$$

[13] 这里有一个小技巧, 定义中的 $\varepsilon > 0$ 是任意的, 所以取 $k\varepsilon$ (k 是任意固定的正数) 都是没有问题的, 因为只要做适当的缩放, $k\varepsilon$ 也是可以取到任意的正数.

由三角不等式得:

$$|l_1 - l_2| = |(l_1 - a_n) + (a_n - l_2)| \le |l_1 - a_n| + |a_n - l_2| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

因为 ε 是任意小的, 所以 $|l_1 - l_2| = 0$, 即 $l_1 = l_2$. 这里用到了一个小引理: 如果对于任意小的实数 ε, 都有 $|c| < \varepsilon$, 那么一定有 $c = 0$. □

根据极限定义可证明下列事实. 所以, 对于一个序列. 我们关心的是在 n 无限增大的过程中通项的变化趋势, 而不是通项在早期阶段 (即开始的有限个的项) 的表现.

命题 1.3.1: 有限截断不影响数列的收敛性质

若 $\lim\limits_{n\to\infty} a_n = l$, 则对于任意正整数 k, 都有 $\lim\limits_{n\to\infty} a_{n+k} = l$.

通常把定义 1.3.1 中的这种严格的定义说法称作 $\varepsilon - N$ 说法. 现在我们用 $\varepsilon - N$ 说法证明几个常见序列的极限.

例 1.3.1 证明: $\lim\limits_{n\to\infty} \dfrac{1}{n} = 0$.

证明. 对于任意给定的 $\varepsilon > 0$, 要使

$$\left|\frac{1}{n} - 0\right| < \varepsilon,$$

只要 $n > \varepsilon^{-1}$. 因此, 我们取 $N = \left[\varepsilon^{-1}\right] + 1$, 即有

$$\left|\frac{1}{n} - 0\right| < \varepsilon, \qquad \text{只要 } n > N.$$

证毕. □

这里 $\left[\varepsilon^{-1}\right]$ 是 ε^{-1} 的整数部分. 当 $\varepsilon > 1$ 时, $\left[\varepsilon^{-1}\right] = 0$, 为了保证 N 为正整数, 我们取 N 等于 $\left[\varepsilon^{-1}\right] + 1$. 从这个简单的例子中我们看到,

- N 是根据 ε 找的. 一般来说, N 要依赖于 ε.
- 另外, 只要找到符合要求的 N 即可, 而不要求它是最小的. 即, 如果我们找到了一个符合要求的 N, 那么任意一个比 N 大的正整数也都是符合要的. 比如, 在这个例子中取 $N = \left[\varepsilon^{-1}\right] + 5$, 自然也是可以的.

例 1.3.2 证明: $\lim\limits_{n\to\infty} \dfrac{1}{n^\alpha} = 0\ (\alpha > 0)$.

证明. 对于任意给定的 $\varepsilon > 0$, 要使

$$\left|\frac{1}{n^\alpha} - 0\right| < \varepsilon$$

只要 $n^\alpha > \varepsilon^{-1}$, 故只要 $n > \sqrt[\alpha]{\varepsilon^{-1}}$. 因此, 我们取 $N = \left[\sqrt[\alpha]{\varepsilon^{-1}}\right] + 1$, 即有

$$\left|\frac{1}{n^\alpha} - 0\right| < \varepsilon, \quad \text{只要 } n > N,$$

证毕. □

例 1.3.3 设 $a>1$ 是给定的实数, 证明:

$$\lim_{n\to\infty}\sqrt[n]{a}=1.$$

证明. 注意到 $a>1$, 我们有 $a^{\frac{1}{n}}>1$. 对于任意给定的 $\varepsilon>0$, 要使 $\left|a^{\frac{1}{n}}-1\right|<\varepsilon$, 只要 $a^{\frac{1}{n}}-1<\varepsilon$, 即只要 $a^{\frac{1}{n}}<1+\varepsilon$. 此式两边取以 a 为底的对数即得 $\dfrac{1}{n}<\log_a(1+\varepsilon)$, 等价于

$$n>\left(\log_a(1+\varepsilon)\right)^{-1}.$$

于是, 我们取 $N=\left[\left(\log_a(1+\varepsilon)\right)^{-1}\right]+1$, 即有

$$\left|a^{\frac{1}{n}}-1\right|<\varepsilon,\ \text{只要 } n>N.$$

证毕. □

> **注释 1.3.2** 类似可证明 $\lim\limits_{n\to\infty}\sqrt[n]{a}=1\ (0<a\leqslant 1)$.

上面的三个例子告诉我们, 证明某个序列的极限为 l, 这一问题可归结为解不等式:

1. 从不等式
$$|a_n-l|<\varepsilon,$$
出发. 我们注意到 a_n 是关于 n 的函数, 或者说关于 n 的数学表达式. 那么, 不等式的左侧 $|a_n-l|$ 也是仅关于 n 的函数. 此时, ε 只出现在了右侧, 且左侧与 ε 无关.
2. 然后, 通过各种解不等式的技巧, 我们希望获得如下形式:
$$n>\ \text{关于 } \varepsilon \text{ 的函数/数学表达式}.$$
此时, n 只出现在了左侧, 且右侧与 n 无关.
3. 但不是求出所有满足不等式的 n, 而是要求找到一个 N, 使得 $n>N$ 时上述不等式成立即可. 归根到底, 我们是在不断地寻找充分条件, 使得 $|a_n-l|<\varepsilon$ 成立. 而不是必要条件.

例 1.3.4 设 q 为常数, $|q|<1$, 证明:

$$\lim_{n\to\infty}q^n=0.$$

证明. 若 $q=0$, 则 $q^n=0$ 对于一切 $n=1,2,\cdots$ 成立. 这时, 显然有

$$\lim_{n\to\infty}q^n=0.$$

假定 $q\neq 0$, 这时 $0<|q|<1$. 对于任意给定的 $\varepsilon>0$, 要使 $|q^n-0|<\varepsilon$, 只要 $n\ln|q|<\ln\varepsilon$, 即只要

$$n>\frac{\ln\varepsilon}{\ln|q|}.$$

这里当 $\varepsilon > 1$ 时, $\dfrac{\ln \varepsilon}{\ln |q|}$ 为负数, 这时可取 $N = 1$; 当 $0 < \varepsilon < 1$ 时, 我们取

$$N = \left[\frac{\ln \varepsilon}{\ln |q|}\right] + 1$$

这样, 当 $n > N$ 时, 便有 $|q^n - 0| < \varepsilon$. 证毕. □

在这个例子中, 我们要区别 ε 的取值范围来决定 N 的取法. 今后为了简便起见, 在证此类题时, 不妨一开始就假定 $0 < \varepsilon < 1$. 这样假定是合理的, 这是因为对于较小的 ε 找到的 N, 一定也适用于较大的 ε.

例 1.3.5 证明: $\lim\limits_{n\to\infty} \dfrac{n+2}{n^2+1} = 0$.

证明. 对于任意给定的 $\varepsilon > 0$, 取 $N = \max\left\{2, \left[\dfrac{2}{\varepsilon}\right]\right\}$, 当 $n > N$ 时, 首先有 $n > 2$, 那么

$$\left|\frac{n+2}{n^2+1} - 0\right| = \frac{n+2}{n^2+1} < \frac{n+n}{n^2} = \frac{2}{n}.$$

其次有 $n > \left[\dfrac{2}{\varepsilon}\right]$, 则 $n > \dfrac{2}{\varepsilon}$, 从而

$$\left|\frac{n+2}{n^2+1} - 0\right| < \frac{2}{n} < \varepsilon.$$

证毕. □

在这个证明过程中, 由 $\left|\dfrac{n+2}{n^2+1} - 0\right| < \varepsilon$ 去找合适的 N 有困难. 为了求得 N, 我们先将 $\left|\dfrac{n+2}{n^2+1} - 0\right|$ 适当放大到 $\dfrac{2}{n}$, 这个放大过程是在 $n > 2$ 的前提条件下进行的; 再根据 $\dfrac{2}{n} < \varepsilon$ 确定 N, 并将前提条件考虑进去. 这种适当放大的方法是常用且有效的.

例 1.3.6 证明: $\lim\limits_{n\to\infty} \dfrac{n^2+3n+10}{3n^2-n+2} = \dfrac{1}{3}$.

证明. 注意到, 当 $n > 3$ 时, 有

$$\left|\frac{n^2+3n+10}{3n^2-n+2} - \frac{1}{3}\right| = \left|\frac{10n+28}{3\left(3n^2-n+2\right)}\right| = \frac{10n+28}{3\left[2n^2+n(n-1)+2\right]} < \frac{10n+10n}{3\cdot 2n^2} = \frac{10}{3n}$$

故对于任意给定的 $\varepsilon > 0$, 取 $N = \max\left\{3, \left[\dfrac{10}{3\varepsilon}\right]\right\}$, 那么

$$\left|\frac{n^2+3n+10}{3n^2-n+2} - \frac{1}{3}\right| < \frac{10}{3n} < \varepsilon, \text{ 只要 } n > N.$$

证毕. □

例 1.3.7 设 $a_n \geqslant 0\ (n = 1, 2, \cdots)$, 且 $\lim\limits_{n\to\infty} a_n = a \geqslant 0$, 证明:

$$\lim_{n\to\infty} \sqrt{a_n} = \sqrt{a}.$$

证明. 若 $a=0$, 对于任意给定的 $\varepsilon>0$, 由 $\lim\limits_{n\to\infty} a_n=0$ 知, 存在一个正整数 N, 使得

$$|a_n-0|<\varepsilon^2,\ \text{只要}\ n>N,$$

那么

$$|\sqrt{a_n}-0|<\varepsilon,\ \text{只要}\ n>N,$$

即得到

$$\lim_{n\to\infty}\sqrt{a_n}=\sqrt{a}$$

下面设 $a>0$. 对于任意给定的 $\varepsilon>0$, 由 $\lim\limits_{n\to\infty} a_n=a$ 知, 存在一个正整数 N, 使得

$$|a_n-a|<\sqrt{a}\cdot\varepsilon,\qquad \text{只要}\ n>N,$$

则只要 $n>N$, 就有

$$\left|\sqrt{a_n}-\sqrt{a}\right|=\frac{|a_n-a|}{\sqrt{a_n}+\sqrt{a}}\leqslant\frac{|a_n-a|}{\sqrt{a}}<\varepsilon$$

于是, 当 $a>0$ 时, 也得到

$$\lim_{n\to\infty}\sqrt{a_n}=\sqrt{a}$$

□

例 1.3.8 证明: $\lim\limits_{n\to\infty} 0.\underbrace{99\cdots9}_{n\text{ 个}9}=1$.

证明. 记 $a_n=0.\underbrace{99\cdots9}_{n\text{ 个}9}$. 我们先将 a_n 表示成分数形式:

$$\begin{aligned}a_n&=\frac{9}{10}+\frac{9}{10^2}+\cdots+\frac{9}{10^n}=\frac{9}{10}\left(1+\frac{1}{10}+\cdots+\frac{1}{10^{n-1}}\right)\\&=\frac{9}{10}\cdot\frac{1-\frac{1}{10^n}}{1-\frac{1}{10}}=1-\frac{1}{10^n}\end{aligned}$$

对于任意给定的 $\varepsilon>0$, 不妨设 $0<\varepsilon<1$, 取 $N=\left[\ln\dfrac{1}{\varepsilon}/\ln 10\right]+1$, 则当 $n>N$ 时, $n>\ln\dfrac{1}{\varepsilon}/\ln 10$, 从而

$$|a_n-1|=\frac{1}{10^n}<\varepsilon$$

则 $\lim\limits_{n\to\infty} a_n=1$. 证毕. □

希望读者记住以上例题中使用的各类技巧, 以后我们会经常用到它们.

1.3.2 夹逼定理

在此之前, 我们遇到的问题都是假设我们先验地知道了某个序列的极限值 l, 然后从定义出发验证论证 l 确实是极限. 但是更多时候我们不知道极限值, 所以想找到它. 下面我们介绍一个很有用的定理, 俗称**夹逼定理/三明治定理**.

定理 1.3.2: 夹逼定理/三明治定理

设 $\{a_n\},\{b_n\},\{c_n\}$ 为三个序列, 且存在一个正整数 N_0, 使得

$$c_n \leqslant a_n \leqslant b_n, \quad \forall n \geqslant N_0.$$

若 $\{c_n\}$ 与 $\{b_n\}$ 都有极限存在, 且都等于 l, 则 $\{a_n\}$ 的极限存在, 且也等于 l.

证明. 根据定理的假定, 对于任意给定的 $\varepsilon > 0$, 存在正整数 N_1 与 N_2, 使得

$$\begin{aligned} |b_n - l| < \varepsilon, &\quad \text{只要 } n > N_1, \\ |c_n - l| < \varepsilon, &\quad \text{只要 } n > N_2, \end{aligned}$$

即

$$\begin{aligned} b_n < l + \varepsilon, &\ \text{只要 } n > N_1, \\ l - \varepsilon < c_n, &\ \text{只要 } n > N_2. \end{aligned}$$

现在, 我们取 $N = \max\{N_0, N_1, N_2\}$, 则有

$$l - \varepsilon < c_n \leqslant a_n \leqslant b_n < l + \varepsilon, \text{ 只要 } n > N,$$

即

$$|a_n - l| < \varepsilon, \text{ 只要 } n > N.$$

证毕. □

这个定理的用途之一是: 当 a_n 的表达式比较复杂, 一时难以处理时, 不妨对它做适当的放大与缩小, 只要适当放大与缩小后的序列有相同的极限, 则序列 $\{a_n\}$ 就也有极限.

例 1.3.9 设 $a > 1$ 是任意给定的常数. 考查序列

$$a_n = \frac{a^n}{n!}, \quad n = 1, 2, \cdots$$

是否有极限.

证明. 乍一看很难对这个序列是否有极限做出判断, 因为分子与分母都会无限增大 (当 n 无限增大时). 但是, 当我们将它改写为下面的形式时, 就立刻能得出结论. 当 $n > [a] + 1$ 时, 有

$$0 \leqslant a_n = \frac{a^n}{n!} = \frac{a^{[a]}}{[a]!} \times \frac{a}{[a]+1} \times \cdots \times \frac{a}{n-1} \times \frac{a}{n} < \frac{a^{[a]}}{[a]!} \times \frac{a}{n},$$

因为中间部分的 $\dfrac{a}{[a]+1} \times \cdots \times \dfrac{a}{n-1}$ 的每一项都在 $(0,1)$ 之间. 则我们有

$$0 \leqslant a_n < \frac{a^{[a]}}{[a]!} \times \frac{a}{n}, \quad n > [a] + 1.$$

注意到 a 给定后 $\dfrac{a^{[a]}}{[a]!}$ 是一个常数, 很容易看出

$$\lim_{n \to \infty} \frac{a^{[a]}}{[a]!} \cdot \frac{a}{n} = 0.$$

另外, a_n 的下界为 0, 又显然 $\lim\limits_{n\to\infty} 0 = 0$, 于是应用定理 1 可得

$$\lim_{n\to\infty}\frac{a^n}{n!}=0.$$

注意这里, 我们把常数 0 当做一个常数数列 $\{a_n\equiv 0\}$, $n=1,2,\ldots$; 这是在使用夹逼定理时的常用技巧. □

注意, 从上面这个例子可以看出, 在使用夹逼定理 1.3.2 时, 如何去构造和找到复合不等式要求的下界数列 $\{c_n\}$ 与上界数列 $\{b_n\}$ 是关键一步. 要小心的是, 我们要求不等式 $c_n\leqslant a_n\leqslant b_n$ 的成立不是对所有 $n=1,2,\cdots$, 而是当 n 充分大时成立. 所以, 在这三个的序列的早期阶段, 即使他们不满足不等式关系也无所谓. 这一点返回来帮助去构造下界数列 $\{c_n\}$ 与上界数列 $\{b_n\}$, 因为我们可以要求 n 大于任何一个固定的常数. 这一构造技巧在下面的例子中可以进一步感受到.

例 1.3.10 设 k 为大于 1 的整数, 证明:

$$\lim_{n\to\infty}\frac{n^{k-1}}{(n-1)(n-2)\cdots(n-k)}=0.$$

证明. 令 $a_n=\dfrac{n^{k-1}}{(n-1)(n-2)\cdots(n-k)}$, 其中 $n>k$. 用 n^k 同除 a_n 的分子与分母, 得

$$0\leqslant a_n=\frac{\frac{1}{n}}{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k}{n}\right)}.$$

当 $n>2k$ 时, 上式右端的分母中因子 $1-\dfrac{i}{n}(i=1,2,\cdots,k)$ 均大于 $\dfrac{1}{2}$, 于是我们得到不等式

$$0\leqslant a_n\leqslant 2^k\cdot\frac{1}{n},\quad n>2k.$$

注意到上述不等式左端构成的序列与右端构成的序列的极限均为零, 由定理 1 即得 $\lim\limits_{n\to\infty}a_n=0$. 证毕. □

例 1.3.11 设 $a>1$ 是一个常数, 证明:

$$\lim_{n\to\infty}\frac{n}{a^n}=0.$$

证明. 令 $h=a-1$, 则 $h>0$, 且

$$\begin{aligned}a^n=&(1+h)^n\\=&1+nh+\frac{n(n-1)}{2!}h^2+\cdots+\frac{n(n-1)\cdots(n-k)}{(k+1)!}h^{k+1}+\cdots+h^n.\end{aligned}$$

因此, 当 $n>1$ 时, $a^n\geqslant\dfrac{n(n-1)}{2}h^2$, 从而

$$0\leqslant\frac{n}{a^n}\leqslant\frac{n}{\frac{1}{2}n(n-1)h^2}=\frac{2}{(n-1)h^2}.$$

应用定理 1 即得到 $\lim\limits_{n\to\infty}\dfrac{n}{a^n}=0$. 证毕. □

注释 1.3.3 利用 $a^n > \dfrac{n(n-1)\cdots(n-k)}{(k+1)!}h^{k+1}$ 及例 10, 可进一步证明, 当 $a>1$ 时, 对于任意正整数 k, 我们有

$$\lim_{n\to\infty}\frac{n^k}{a^n}=0.$$

通过上述例子可以看出, 虽然序列 $\{n\}$ (或 $\{n^k\}$),$\{a^n\}$ $(a>1)$ 及 $\{n!\}$ 都是趋向于无穷大的, 但其趋向于无穷大的" 速度" 不同: $\{n\}$ 较 $\{a^n\}$ 慢, 而 $\{a^n\}$ 较 $\{n!\}$ 慢. 这三种不同的关于 n 的增长速度, 是计算机理论中算法复杂度的重要组成.

在现在为止出现的例题都是极限值等于 0 的情况. 由于这种情况十分常见, 所以直接把其写成一推论记在心中是值得的.

推论 1.3.1: 一个常用的特殊情况的夹逼定理

设 $\{a_n\},\{b_n\}$ 为两个非负序列[a](即所有项都是非负的), 且存在一个正整数 N_0, 使得

$$a_n\leqslant b_n,\quad \forall n\geqslant N_0.$$

若 $\lim\limits_{n\to\infty}b_n=0$, 则 $\lim\limits_{n\to\infty}a_n=0$.

[a]把条件放松一些, 可以仅要求当 n 充分大时, 序列是非负的.

上面的推论中常数 0 也只不过是一个特殊情况, 把它换成任何其他常数 c 也是没有问题的: 设 $\{a_n\},\{b_n\}$ 为两个序列, 其中当 n 充分大时, $c\leqslant a_n$ 恒成立, 且存在一个正整数 N_0, 使得

$$a_n\leqslant b_n,\quad \forall n\geqslant N_0.$$

此时, 若 $\lim\limits_{n\to\infty}b_n=c$, 则 $\lim\limits_{n\to\infty}a_n=c$.

1.3.3 极限不等式

下面两个定理是关于极限不等式的, 它们同样是经常要用的, 有关极限的基本定理.

定理 1.3.3: 极限的保序性 (1)

设序列 $\{a_n\}$ 及 $\{b_n\}$ 分别有极限 l_1 及 l_2, 并且 $l_1>l_2$, 则存在一个正整数 N, 使得

$$a_n>b_n,\ 只要\ n>N.$$

也就是说, 在两个序列的极限存在的条件下, 项数充分大之后, 极限较大的序列的项要大于极限较小的序列的对应项.

证明. 对于任意给定的 $\varepsilon>0$, 存在正整数 N_1 及 N_2, 使得

$$|a_n-l_1|<\varepsilon,\quad 只要\ n>N_1;\qquad |b_n-l_2|<\varepsilon,\quad 只要\ n>N_2.$$

取 $N=\max\{N_1,N_2\}$, 那么当 $n>N$ 时, 有 $|a_n-l_1|<\varepsilon,|b_n-l_2|<\varepsilon$, 即

$$l_1-\varepsilon<a_n,\quad b_n<l_2+\varepsilon.$$

因此, 当 $n > N$ 时, $a_n - b_n > l_1 - l_2 - 2\varepsilon$. 由于 ε 是任意给定的正数, 故事先可取 $0 < \varepsilon < \frac{1}{2}(l_1 - l_2)$. 在这种取法下, 当 $n > N$ 时, $a_n - b_n > 0$, 即 $a_n > b_n$. 证毕. □

注意 特别注意, 定理 1.3.3 中的条件是 $l_1 > l_2$, 而不能是 $l_1 \geq l_2$, 否则结论: 存在一个正整数 N, 使得

$$a_n > b_n, \text{ 只要 } n > N. \quad \text{亦或者} \quad a_n \geqslant b_n, \text{ 只要 } n > N. \tag{1.1}$$

可能不成立. 一个典型的反例是两个序列 $a_n = -\frac{1}{n}$ 和 $b_n = \frac{2}{n}$, 他们的极限都是 0. 但是 $a_n < b_n$, 对于所有 $n = 1, 2, \ldots$

在定理 1.3.3 中令 $b_n = 0$ 即可得到下面的推论.

推论 1.3.2

设序列 $\{a_n\}$ 有极限 l 且 $l > 0$ (或 < 0), 则存在一个正整数 N, 使得当 $n > N$ 时, $a_n > 0$ (或 < 0).

也就是说, 若序列的极限存在且严格大于 0 (严格小于 0), 那么项数充分大之后, 所有项都会严格大于 0 (严格小于 0).

一个稍稍一般化的推论是:

推论 1.3.3

设序列 $\{a_n\}$ 有极限 l, 以及 c 是某个常数. 若 $l > c$ (或 $l < c$), 则存在一个正整数 N, 使得当 $n > N$ 时, $a_n > c$ (或 $a_n < c$).

下面的定理是定理 1.3.3 在某种程度上的反向结论. 注意, 只是某种程上的, 它并不是主张定理 1.3.3 的条件和结论是充分必要条件, 正如我们在注意 (1.1) 所讨论.

定理 1.3.4: 极限的保序性 (2)

设序列 $\{a_n\}$ 及 $\{b_n\}$ 分别有极限 l_1 及 l_2, 并且存在正整数 N_0, 使得 $a_n \geqslant b_n$, 只要 $n > N$. 则

$$l_1 \geqslant l_2.$$

证明. 用反证法. 若 $l_1 < l_2$, 则由定理 2 推出, 存在一个正整数 N, 使得当 $n > N$ 时, $a_n < b_n$. 这与已知条件矛盾. 由此推出 $l_1 \geqslant l_2$. 证毕. □

上述定理告诉我们, 如果从自某项开始, 一个序列的项总是大于或等于另一序列的对应项, 则它们的极限也有同样的大小次序. 不过, 我们要提醒读者, 即使是 a_n 严格大于 b_n, 即

$$a_n > b_n, \quad n = 1, 2, \cdots$$

但仍然只能推出

$$\lim_{n\to\infty} a_n = l_1 \geqslant l_2 = \lim_{n\to\infty} b_n.$$

这里 "=" 是可能发生的. 比如, 对于

$$a_n = \frac{2n+4}{n}, \quad b_n = \frac{2n+1}{n}, \quad n = 1, 2, \cdots \tag{1.2}$$

显然有 $a_n > b_n$ $(n=1,2,\cdots)$, 但它们有相同的极限 2.

为清楚地展示这本小节中两个定理的关系, 我们把它们写到一起:

定理 1.3.5: 极限的保序性 (汇总)

设序列 $\{a_n\}$ 及 $\{b_n\}$ 分别有极限 l_1 及 l_2. 有以下两个结论:

1. 若 $l_1 > l_2$; 则存在一个正整数 N, 使得 $a_n > b_n$, 只要 $n > N$.
2. 若存在一个正整数 N, 使得 $a_n \geqslant b_n$, 只要 $n > N$; 则 $l_1 \geqslant l_2$.

1.3.4　极限的四则运算

从现在开始, 我们尝试用下面的定理试图去找到未知的极限值 (如果存在).

定理 1.3.6: 极限的四则运算

设序列 $\{a_n\}$ 与 $\{b_n\}$ 都有极限, 它们的极限分别为 l_1 与 l_2, 则有

$$\lim_{n\to\infty}(a_n \pm b_n) = l_1 \pm l_2, \quad \lim_{n\to\infty} a_n b_n = l_1 l_2.$$

特别地, 如果 $b_n = c$ 是某个常数序列, 则有

$$\lim_{n\to\infty} c\cdot a_n = c\cdot l_1.$$

并且当 $l_2 \neq 0$ 时, 有

$$\lim_{n\to\infty}\frac{a_n}{b_n} = \frac{l_1}{l_2}.$$

注意　若 $l_2 \neq 0$, 根据推论 1.3.2 告诉我们, 当 n 充分大时, $b_n \neq 0$, 因而 $\dfrac{a_n}{b_n}$ 在 n 充分大时是有意义的. 但是在 n 比较小的时候仍有可能 $b_n = 0$, 此时我们怎么办呢? 结论是: 直接删掉前面前一段序列, 不予考虑. 再次重申: 对于任意一个序列. 我们关心的是在 n 无限增大的过程中通项的变化趋势; 我们不关心的是通项在早期阶段 (即开始的有限个的项) 的表现.

证明. 我们依次证明加 (减), 乘和商.

1. 设 ε 是任意给定的正数. 则对于 $\dfrac{\varepsilon}{2}$ 这个正数, 可以找到正整数 N_1 及 N_2, 使得
$$|a_n - l_1| < \frac{\varepsilon}{2},\quad 只要\ n > N_1;\quad |b_n - l_2| < \frac{\varepsilon}{2},\quad 只要\ n > N_2.$$
取 $N = \max\{N_1, N_2\}$, 则我们有
$$|(a_n \pm b_n) - (l_1 \pm l_2)| = |(a_n - l_1) \pm (b_n - l_2)| \leqslant |a_n - l_1| + |b_n - l_2| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,\quad 只要\ n > N.$$
这就证明了 $\lim\limits_{n\to\infty}(a_n \pm b_n) = l_1 \pm l_2$.

2. 为了证明 $\lim\limits_{n\to\infty} a_n b_n = l_1 l_2$, 我们将 $|a_n b_n - l_1 l_2|$ 做适当放大:
$$|a_n b_n - l_1 l_2| = |(a_n - l_1)\, b_n + l_1\,(b_n - l_2)| \leqslant |b_n|\,|a_n - l_1| + |l_1|\,|b_n - l_2|\,.$$

对于 $\varepsilon_0 = 1$, 我们可以找到一个正整数 N', 使得

$$|b_n - l_2| < 1, \qquad 只要\ n > N',$$

从而 $|b_n| - |l_2| < 1$[14], 只要 $n > N'$, 即

$$|b_n| < 1 + |l_2|, \qquad 只要\ n > N'.$$

因此, 当 $n > N'$ 时, 我们有

$$|a_n b_n - l_1 l_2| \leqslant (1 + |l_2|)\,|a_n - l_1| + |l_1|\,|b_n - l_2|. \tag{1.3}$$

设 ε 是任意给定的正数.

- 根据假定 $\lim\limits_{n\to\infty} a_n = l_1$, 对于正数 $\varepsilon' = \dfrac{\varepsilon}{2\,(1 + |l_2|)}$, 存在一个正整数 N_1, 使得
$$|a_n - l_1| < \varepsilon', \qquad 只要\ n > N_1. \tag{1.4}$$
- 根据假定 $\lim\limits_{n\to\infty} b_n = l_2$, 对于正数 $\varepsilon'' = \dfrac{\varepsilon}{2\,(1 + |l_1|)}$, 存在一个正整数 N_2, 使得
$$|b_n - l_2| < \varepsilon'', \qquad 只要\ n > N_2. \tag{1.5}$$

取 $N = \max\{N_1, N_2, N'\}$, 那么当 $n > N$ 时, 由 (1.3), (1.4), (1.5) 可得

$$|a_n b_n - l_1 l_2| < (1 + |l_2|)\,\varepsilon' + |l_1|\,\varepsilon'' \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

这就证明了 $\lim\limits_{n\to\infty} a_n b_n = l_1 l_2$.

3. 下面证明
$$\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{l_1}{l_2} \quad (l_2 \neq 0).$$
为了证明这个公式, 只要证明
$$\lim_{n\to\infty} \frac{1}{b_n} = \frac{1}{l_2} \quad (l_2 \neq 0).$$
就足够了. 事实上, 若能证明此式, 则要证的公式就可以由已证明的极限乘法公式推出.

根据 $\lim\limits_{n\to\infty} b_n = l_2$ 及 $l_2 \neq 0$ 的假定, 我们有 $\lim\limits_{n\to\infty} b_n l_2 = l_2^2 > 0$, 故依据推论 1.3.3, 存在一个正整数 N_0, 使得
$$b_n l_2 > \frac{1}{2} l_2^2, \ 只要\ n > N_0.$$
因此, 当 $n > N_0$ 时, 我们有
$$\left|\frac{1}{b_n} - \frac{1}{l_2}\right| = \frac{1}{|b_n l_2|}\,|b_n - l_2| \leqslant \frac{2}{l_2^2}\,|b_n - l_2|.$$
对于任意给定的 $\varepsilon > 0$, 取 $\varepsilon' = \dfrac{l_2^2 \varepsilon}{2}$. 根据 $\lim\limits_{n\to\infty} b_n = l_2$, 对于 $\varepsilon' > 0$, 存在一个正整数 N_1, 使得
$$|b_n - l_2| < \varepsilon' = \frac{l_2^2 \varepsilon}{2}, \ 只要\ n > N_1.$$
因此, 当 $n > \max\{N_0, N_1\}$ 时,
$$\left|\frac{1}{b_n} - \frac{1}{l_2}\right| \leqslant \frac{2}{l_2^2}\,|b_n - l_2| < \frac{2}{l_2^2} \cdot \varepsilon' = \varepsilon.$$
即 $\lim\limits_{n\to\infty} \dfrac{1}{b_n} = \dfrac{1}{l_2}$. 证毕.

[14]这里用到了一个绝对值不等式的性质, 可以见教材习题 1.1 的 4 (2).

□

这个定理告诉我们: 在所涉及的序列都有极限的情况下, 极限运算可以与四则运算交换次序. 这为求极限提供了很大方便.

例 1.3.12 求极限 $\lim_{n\to\infty}\frac{n^3+5n+1}{4n^3+8}$.

证明. 原式中的分子与分母都没有极限, 故不能直接应用定理 1.3.6. 但是, 用 n^3 同除分子与分母后即可应用定理 1.3.6:

$$\lim_{n\to\infty}\frac{n^3+5n+1}{4n^3+8}=\lim_{n\to\infty}\frac{1+\frac{5}{n^2}+\frac{1}{n^3}}{4+\frac{8}{n^3}}=\frac{\lim_{n\to\infty}\left(1+\frac{5}{n^2}+\frac{1}{n^3}\right)}{\lim_{n\to\infty}\left(4+\frac{8}{n^3}\right)}=\frac{1}{4}.$$

□

例 1.3.13 求极限 $\lim_{n\to\infty}(\sqrt{n+\sqrt{n}}-\sqrt{n})$.

证明. 因为 $\{\sqrt{n+\sqrt{n}}\}$ 及 $\{\sqrt{n}\}$ 均无极限, 所以我们不能直接应用定理 4 . 这时, 我们做如下变形:

$$\sqrt{n+\sqrt{n}}-\sqrt{n}=\frac{\sqrt{n}}{\sqrt{n+\sqrt{n}}+\sqrt{n}}=\frac{1}{\sqrt{1+\frac{1}{\sqrt{n}}}+1}.$$

这里等式右端分母的极限为 2. 由此推出

$$\lim_{n\to\infty}(\sqrt{n+\sqrt{n}}-\sqrt{n})=\frac{1}{2}.$$

□

显然, 定理 1.3.6 可以被推广到有限次加减乘除四则运算的情况. 但是, 对无限次四则运算不一定成立. 这一点初学者要特别注意. 例如, 考虑序列

$$a_n=\frac{1}{n+1}+\frac{1}{n+2}+\cdots+\frac{1}{n+n},\quad n=1,2,\cdots$$

其中 a_n 的每一项当 $n\to\infty$ 时均趋向于零, 因此有人会误认为 $a_n\to 0\ (n\to 0)$. 但这是不对的, 因为这个和式中共有 n 项, 这里的项数随着 n 增大而无限增多, 从而加法运算的次数不是有限的, 不能使用定理 1.3.6. 另外, 很容易看出

$$a_n=\frac{1}{n+1}+\frac{1}{n+2}+\cdots+\frac{1}{n+n}>\frac{1}{n+n}+\frac{1}{n+n}+\cdots+\frac{1}{n+n}=\frac{n}{n+n}=\frac{1}{2},\quad n=1,2,\cdots$$

可见, 即使 $\{a_n\}$ 有极限, $\{a_n\}$ 也不可能以零为极限.

注释 1.3.4 上面所讨论的通项 a_n 是 n 个项的求和. 当 $n \to \infty$ 时, a_n 会变成无穷多个项的求和. 无穷多个项的求和这一形式叫做**级数 (series)**. 下学期的高数课会讲到级数. 关于级数, 比较反直觉的是: 无穷多个项 (即使每一个项都是严格大于 0) 的和仍然可能是一个定数, 而不是正无穷大. 最有名的结论是:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

尤其是, 对于自然常数 e:

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} = \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots \tag{1.6}$$

你可以用 python 或者 matlab 做数值实验以验证上述结论.

在讨论序列的极限问题时, 常常涉及序列的**子序列 (subsequence)** 概念. **子序列是在原序列中抽出一部分项 (必须抽出无穷多项) 并保持原来项的次序所组成的新序列**. 比如, 我们在序列 $\{a_n\}$ 中抽出其全部奇数项

$$a_1, a_3, a_5, \cdots, a_{2k-1}, \cdots$$

它们就组成一个子序列. 这个子序列中的第 k 项恰好是原序列 $\{a_n\}$ 中的第 $2k-1$ 项.

一般来说, 如果我们在序列 $\{a_n\}$ 中首先挑出 a_{n_1} 作为子序列的第 1 项, 然后在 a_{n_1} 后面又挑一项 a_{n_2} 作为第 2 项 $\cdots\cdots$ 如此下去, 我们就得 $\{a_n\}$ 的一个子序列:

$$a_{n_1}, a_{n_2}, \cdots, a_{n_k}, \cdots$$

该子序列的第 k 项 a_{n_k} 恰好是原序列 $\{a_n\}$ 的第 n_k 项. 在奇数项组成的子序列中 $n_k = 2k-1$.

根据挑选的次序, 不难看出

$$n_k \geqslant k; \quad n_{k_1} < n_{k_2}, \text{如果} k_1 < k_2.$$

定理 1.3.7: 子序列的极限等于原序列的极限

设序列 $\{a_n\}$ 有极限 l, 则它的任意一个子序列 $\{a_{n_k}\}$ 也以 l 为极限.

证明. 对于任意给定的 $\varepsilon > 0$, 由假定可知存在一个正整数 N, 使得

$$|a_n - l| < \varepsilon, \qquad \text{只要 } n > N.$$

特别地, 我们有

$$|a_{n_k} - l| < \varepsilon, \qquad \text{只要 } n_k > N.$$

因 $n_k \geqslant k$, 故当 $k > N$ 时, $n_k > N$, 从而有 $|a_{n_k} - l| < \varepsilon$. 这就证明了 $\{a_{n_k}\}$ 以 l 为极限. 证毕. □

这个定理的意义不仅在于保证了子序列的极限等于原序列的极限, 而且还在于告诉我们:

命题 1.3.2: 通过子序列判别极限不存在

若在一个序列中存在两个子序列, 它们都有极限, 但极限值不同, 那么原序列的极限不存在.

例 1.3.14 例如, 在前面举的例子 $\left\{\sin\dfrac{n\pi}{2}\right\}$ 中, 就可以取出有不同极限的两个子序列 (比如, 常数序列 $\{1\}$ 和 $\{-1\}$), 从而严格证明了它的极限不存在.

1.3.5 一个重要极限

现在我们要介绍一个非常重要极限:

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n=\mathrm{e}.$$

首先证明: 当 $n\to\infty$ 时, 序列 $\left\{\left(1+\dfrac{1}{n}\right)^n\right\}$ 有极限. 为了证明这个序列有极限, 我们首先需证明两个事实:

1. $\left\{\left(1+\dfrac{1}{n}\right)^n\right\}$ 是有界的. 事实上, 由二项式定理 (见附录 .2) 有

$$\begin{aligned}\left(1+\frac{1}{n}\right)^n=&1+n\cdot\frac{1}{n}+\frac{1}{2!}\cdot\frac{n(n-1)}{n^2}+\frac{1}{3!}\cdot\frac{n(n-1)(n-2)}{n^3}+\cdots+\frac{1}{n!}\cdot\frac{n!}{n^n}\\ <&1+1+\frac{1}{2!}+\frac{1}{3!}+\cdots+\frac{1}{n!}\\ <&1+1+\frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\cdots+\frac{1}{n(n-1)}\\ =&1+1+\left(1-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\cdots+\left(\frac{1}{n-1}-\frac{1}{n}\right)\\ =&3-\frac{1}{n}<3.\end{aligned}$$

2. $\left\{\left(1+\dfrac{1}{n}\right)^n\right\}$ 是单调递增的. 这是很容易验证的: 我们将 $\left(1+\dfrac{1}{n}\right)^n$ 与 $\left(1+\dfrac{1}{n+1}\right)^{n+1}$ 展开:

$$\begin{aligned}\left(1+\frac{1}{n}\right)^n&=1+1+\sum_{k=2}^{n}\frac{1}{k!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right),\\ \left(1+\frac{1}{n+1}\right)^{n+1}&=1+1+\sum_{k=2}^{n}\frac{1}{k!}\left(1-\frac{1}{n+1}\right)\cdots\left(1-\frac{k-1}{n+1}\right)+\left(\frac{1}{n+1}\right)^{n+1}\\ &=\left(1+\frac{1}{n}\right)^n+\left(\frac{1}{n+1}\right)^{n+1}.\end{aligned}$$

比较上面两个式子右端两个求和号中的对应项, 显然前者的项较小, 又 $\left(\dfrac{1}{n+1}\right)^{n+1}>0$, 于是

$$\left(1+\frac{1}{n}\right)^n<\left(1+\frac{1}{n+1}\right)^{n+1}.$$

根据实数域的完备性, 在实数域中单调有界序列有极限. 特别地, 单调递增有上界的序列有极限. 因此, 序列 $\left\{\left(1+\dfrac{1}{n}\right)^n\right\}$ 有极限. 通常人们把这个序列的极限记为**自然常数** e. 雅各布 · 伯努利 (Jacob Bernoulli) 在研究复利计算时发现了这个极限, 而欧拉 (Euler) 首先使用 e 来代表这个极限, 并把它作为自然对数的底. e 的引进大大简化了许多计算, 成为数学中最重要的常数之一. 通常来讲, 自然常数 e 的一个标准的严格定义就是

$$\mathrm{e}:=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n.$$

当然, (1.6) 是另一种等价的标准定义. 但是只有未来学习了级数之后才能理解. 根据前面的讨论, 我们知道 e 介于 2 与 3 之间. 可以证明 e 是一个无理数. e 的近似值为

$$\mathrm{e} = 2.71828182845904523536\ldots$$

常数 e 及自然对数的引入为科学计算带来很大方便. 这一点会在今后的讨论中逐步显露出来.

例 1.3.15 求极限 $\lim\limits_{n\to\infty}\left(1+\dfrac{1}{n}\right)^{2n}$.

证明. 由于

$$\left(1+\frac{1}{n}\right)^{2n} = \left[\left(1+\frac{1}{n}\right)^{n}\right]^{2}$$

故有

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{2n} = \mathrm{e}^2.$$

□

受上面例题启发, 我们可以总结出如下推论. 在定理 1.3.6 中, 令 $a_n = b_n$ 然后反复调用极限的乘法运算法则即可得到.

推论 1.3.4: 极限的幂运算

设序列 $\{a_n\}$ 有极限 l, 则有 $\lim\limits_{n\to\infty} a_n^k = l^k$, 其中 k 是任意正整数.

例 1.3.16 求极限 $\lim\limits_{n\to\infty}\left(\dfrac{n+3}{n+1}\right)^{n+1}$.

证明. 我们有

$$\begin{aligned}\left(\frac{n+3}{n+1}\right)^{n+1} &= \left(\frac{n+2}{n+1}\cdot\frac{n+3}{n+2}\right)^{n+1}\\ &= \left(1+\frac{1}{n+1}\right)^{n+1}\left(1+\frac{1}{n+2}\right)^{n+1}\\ &= \left(1+\frac{1}{n+1}\right)^{n+1}\left(1+\frac{1}{n+2}\right)^{n+2}\left(1+\frac{1}{n+2}\right)^{-1}.\end{aligned}$$

根据命题 1.3.1 得到

$$\lim_{n\to\infty}\left(\frac{n+3}{n+1}\right)^{n+1} = \mathrm{e}\cdot\mathrm{e}\cdot 1^{-1} = \mathrm{e}^2.$$

□

例 1.3.17 求极限 $\lim\limits_{n\to\infty}\left(1-\dfrac{1}{n}\right)^{n}$.

证明. 由于

$$\left(1-\frac{1}{n}\right)^n=\left(\frac{n-1}{n}\right)^n=\frac{1}{\left(\frac{n}{n-1}\right)^n}=\frac{1}{\left(1+\frac{1}{n-1}\right)^{n-1}}\cdot\frac{1}{1+\frac{1}{n-1}},$$

故有

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n=\frac{1}{\mathrm{e}}\cdot\frac{1}{1}=\mathrm{e}^{-1}.$$

□

1.3.6 其他重要结论

本小节的结论在教材正文中没有被提及, 但仍然十分重要. 同济版高数在正文中将它们提及.

定理 1.3.8: 极限的取绝对值运算

如果 $\lim\limits_{n\to\infty}a_n=l$, 则 $\lim\limits_{n\to\infty}|a_n|=|l|$.

证明. 由三角不等式性质 (命题 1.1.5) 可知:

$$||a_n|-|l||\leq|a_n-l|.$$

由于 $\lim\limits_{n\to\infty}a_n=l$, 则对于任意 $\varepsilon>0$, 存在正整数 N, 当 $n>N$ 时, 有:

$$||a_n|-|l||\leq|a_n-l|<\varepsilon.$$

由此可得 $\lim\limits_{n\to\infty}|a_n|=|l|$. 证毕. □

上面定理的反方向通常是不成立的. 通常, 即使数列 $\{|a_n|\}$ 有极限, 但数列 $\{a_n\}$ 未必有极限. 例如 $a_n=(-1)^n$.

下面这个定理是显而易见的, 如果序列 $\{a_n\}$ 不是有界的, 那么给定一个任意大的数字 M, 都会存在某个 $|a_n|>M$. 此时, 序列 $\{a_n\}$ 又怎么会收敛到某个值呢?

定理 1.3.9: 收敛序列的有界性

如果序列 $\{a_n\}$ 有极限 l, 则 $\{a_n\}$ 是一个有界序列 (即存在一个常数 M, 使得 $|a_n|\leqslant M$, $n=1,2,\cdots$).

证明. 令 $\varepsilon=1$, 根据极限的定义, 存在正整数 N, 当 $n>N$ 时, 有 $|a_n-l|<1$. 因此

$$|a_n|=|a_n-l+l|\leq|a_n-l|+|l|<1+|l|.$$

即当 $n>N$ 时, 有

$$|a_n|<|l|+1.$$

对于 $n=1,2,\ldots,N$, 这些 a_n 只取有限个值, 因此我们可以设

$$M=\max\{M_1,|l|+1\},$$

其中, $M_1=\max\{|a_1|,|a_2|,\ldots,|a_N|\}$. 则对于所有 $n\geq1$, 有 $|a_n|\leqslant M$. 因此, 序列 $\{a_n\}$ 是有界的. 证毕. □

定理 1.3.10

如果 $\lim_{n\to\infty} a_n = 0$, 且 $\{b_n\}$ 是有界序列, 则 $\lim_{n\to\infty} a_n b_n = 0$.

证明. 已知 $\{b_n\}$ 有界, 即存在常数 M, 使得 $|b_n| \le M$ 对于所有 n 成立. 因为 $\lim_{n\to\infty} a_n = 0$, 对于任意 $\varepsilon > 0$, 存在正整数 N, 使得当 $n > N$ 时, $|a_n| < \frac{\varepsilon}{M}$. 因此, 当 $n > N$ 时:

$$|a_n b_n| \le |a_n| \cdot |b_n| \le |a_n| \cdot M < \frac{\varepsilon}{M} \cdot M = \varepsilon.$$

由此可得 $\lim_{n\to\infty} a_n b_n = 0$. 证毕. □

1.3.7 序列极限, 收敛与发散的总结

下面是各类不同序列极限的定义. 很多是课本没有的.

1. 序列 a_n 的极限是 l, 或者 a_n 收敛于 l:

$$\lim_{n\to\infty} a_n = l \Leftrightarrow (\forall \varepsilon > 0), (\exists \text{正整数} N), (\forall n \in \mathbb{N}), (n > N \Rightarrow |a_n - l| < \varepsilon).$$

   解释: 对于任意小的正数 ε, 总存在一个足够大的正整数 N, 使得当 $n > N$ 时, a_n 与 l 之间的距离小于 ε, 即 a_n 越来越接近 l.

2. 取其前一条的否定即为: 序列 a_n 的极限不是 l, 或者 a_n 不收敛于 l[15]:

$$\lim_{n\to\infty} a_n \ne l \Leftrightarrow (\exists \varepsilon > 0), (\forall \text{正整数} N), (\exists n \in \mathbb{N}), (n > N \wedge |a_n - l| \ge \varepsilon).$$

   解释: 存在某个正数 ε, 无论如何选择正整数 N, 总存在 $n > N$ 时, a_n 与 l 的距离总是大于等于 ε, 即 a_n 不会无限接近 l.

3. 序列 a_n 发散到正无穷大:

$$\lim_{n\to\infty} a_n = +\infty \Leftrightarrow (\forall M > 0), (\exists \text{正整数} N), (\forall n \in \mathbb{N}), (n > N \Rightarrow a_n > M).$$

   解释: 对于任意大的正数 M, 总存在正整数 N, 使得当 $n > N$ 时, a_n 的值大于 M, 即 a_n 越来越大, 趋向于正无穷.

4. 序列 a_n 发散到负无穷大:

$$\lim_{n\to\infty} a_n = -\infty \Leftrightarrow (\forall M < 0), (\exists \text{正整数} N), (\forall n \in \mathbb{N}), (n > N \Rightarrow a_n < M).$$

   解释: 对于任意小的负数 M, 总存在正整数 N, 使得当 $n > N$ 时, a_n 的值小于 M, 即 a_n 越来越小, 趋向于负无穷.

5. 收敛与发散的一些总结:

   - 序列 a_n 收敛意味着存在某个实数 l, 使得 a_n 收敛于 l.
   - 取其否命题即为: 序列 a_n 不收敛, 意味着对于任意实数 l, 序列 a_n 都不收敛于 l.

[15] 但有可能收敛到别的数.

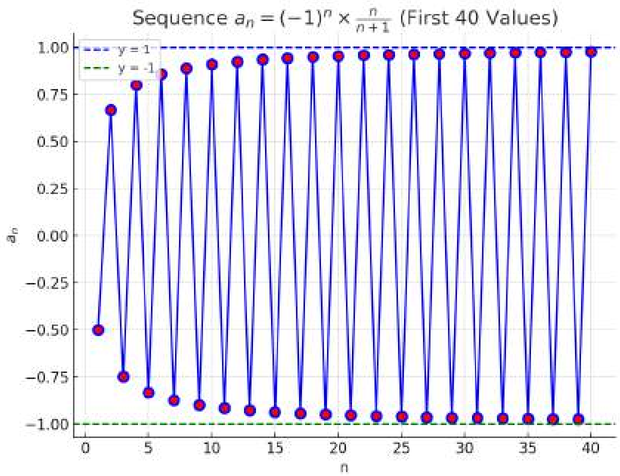


图 1.6: 数列 $a_n = (-1)^n \times \dfrac{n}{n+1}$

- 序列 a_n 不收敛, 即序列 a_n 的极限不存在, 即序列发散.
- 发散到正无穷大或负无穷大只是序列发散的一种特殊情况, 还有另一类序列 (例如反复震荡的序列), 虽然它们不发散到正无穷或负无穷, 但同样不收敛. 考虑数列 $a_n = (-1)^n \times \dfrac{n}{n+1}$. 这个数列既没有极限, 也不发散到无穷, 也不是典型的振荡 (周期性的反复变化, 如 $\{(-1)^n\}$). 我们可以分析一下这个数列的行为 (见图 1.6): 当 n 为奇数时, 数列 a_n 是负数, 趋向于 -1; 当 n 为偶数时, 数列 a_n 是正数, 趋向于 1.

1.4 函数的极限

在上一节中, 我们讨论了序列的极限, 这只是一般函数的极限的特殊情况. 事实上, 序列是定义在正整数集上的一种函数, 只不过其自变量 n 的变化是离散的而已[16].

在考虑序列极限时, 自变量的变化过程只有一种, 即 n 趋向于无穷大, 记作 $n \to \infty$. 但是, 当我们考查函数 $y = f(x)$ 的极限时, 自变量 x 的变化过程是连续的, 并有多种可能性:

1. 实数 a 是某个固定值. 自变量 x 趋向这个固定值.
   (a) x 从点 a 的右侧趋向于点 a, 这时记作 $x \to a + 0$;
   (b) x 从点 a 的左侧趋向于点 a, 记作 $x \to a - 0$;
   (c) x 既可以从点 a 的右侧趋向于点 a, 也可以从点 a 的左侧趋向于点 a, 即 x 同时从点 a 的两侧趋向于点 a, 记作 $x \to a$. 其实就是 $|x - a| \to 0$.
2. 自变量 x 趋向于无穷大.

[16]为什么说序列极限只是函数极限一种特殊情况?

1. 序列是一种特殊的函数.
2. n 的变化是离散的, 只能趋向正无穷; x 的变化是连续的, 可以有 6 种不同的趋向方式.
3. 有时会遇到下面情况, 仅此时两者区别不大:

$$\lim_{n\to\infty} \frac{1}{n}, \quad \lim_{x\to+\infty} \frac{1}{x}. \tag{1.7}$$

但是, 序列极限和函数极限有极其紧密的关联. 见定理 1.4.5.

(a) x 无限制地增大, 记作 $x \to +\infty$;

(b) x 无限制地减小, 记作 $x \to -\infty$;

(c) x 既可以无限制地增大, 也可以无限制地减小, 即 x 沿 x 轴的正向与负向同时无限远离原点, 记作 $x \to \infty$. 其实就是 $|x| \to +\infty$.

在上述 1-a, 1-b, 2-a 和 2-b 情况下, 函数的极限称为单侧极限, 其余两种情况则称为双侧极限. 无论是单侧极限还是双侧极限, 它们的定义非常相似. 在接下来的学习中, 需要特别注意区分它们的共性与差异, 避免混淆.

注意 在未来的章节中, 大多数情况下, 当讨论自变量 x 趋近于某个固定值的极限时, 我们主要关注双侧极限; 而当讨论自变量 x 趋向无穷大时, 我们通常只讨论正无穷大或负无穷大的单侧极限.

注意 无论序列极限还是函数极限, 我们只考虑两个问题: 1. 极限是否存在. 2. 如果存在具体是什么?

1.4.1 自变量趋向点 a 的单侧极限

现在我们讨论前两种种情况, 即 $x \to a+0$ 或 $x \to a-0$ 的情况. 先看几个具体例子 (见图 1.7):

- 符号函数 $y = \operatorname{sgn} x$: 直观上看, 当 $x \to 0+0$ 时, 即当 x 自原点右侧趋向于原点时, $\operatorname{sgn} x$ 趋向于 1; 而当 $x \to 0-0$ 时, 即当 x 自原点左侧趋向于原点时, $\operatorname{sgn} x$ 趋向于一 1. 这时, 我们称符号函数 $y = \operatorname{sgn} x$ 在原点处的右极限为 1, 左极限为 -1.

- 函数 $y = \{x\} := x - [x]$: 显然, 当 $x \to 1+0$ 时, $\{x\}$ 趋向于零; 而当 $x \to 1-0$ 时, $\{x\}$ 趋向于 1. 这时, 我们称函数 $y = \{x\}$ 在点 $x = 1$ 处的右极限为零, 左极限为 1.

- 函数 $y = \sin x$: 当 $x \to \dfrac{\pi}{2}+0$ 时, $\sin x$ 趋向于 1; 而当 $x \to \dfrac{\pi}{2}-0$ 时, 也有 $\sin x$ 趋向于 1, 故函数 $y = \sin x$ 在点 $x = \dfrac{\pi}{2}$ 处的左, 右极限都是 1.

- 函数 $y = \sin \dfrac{1}{x}$: 当 $x \to 0+0$ 或 $x \to 0-0$ 时, 函数值不断地在 -1 及 1 之间摆动, 这时无极限可言.

- 函数 $y = x \sin \dfrac{1}{x}$: 当 x 自原点右侧或左侧趋向于原点时, 虽然函数值也在不断地上下摆动, 但其 "振幅" 却在不断减小, 并且任意接近于 0. 这时, 0 是这个函数在原点处的左, 右极限.

对于序列 $\{a_n\}$ 的极限, 我们有严格的 $\varepsilon - N$ 说法, 其中 ε 用来描述 a_n 的值与极限值之间的偏离程度, 而 N 用来描述序列的项数 n 接近于 ∞ 的程度. 现在讨论函数 $f(x)$ 的极限, 当然仍用 ε 来描述函数值 $f(x)$ 与极限值之间的偏离程度, 而自变量 x 与点 a 的接近程度应该换成用另一个充分小的正数来描述, 通常用 δ 来表示这个正数 (δ 是希腊字母, 读作 delta).

- 所谓的 "x 从点 a 的右侧充分接近于点 a", 可以用 $0 < x - a < \delta$ 来描述, 其中 δ 是充分小或任意小的正数.

- 类似地, "x 从点 a 的左侧充分接近于点 a", 可以用 $0 < a - x < \delta$ 来描述, 其中 δ 是充分小或任意小的正数.

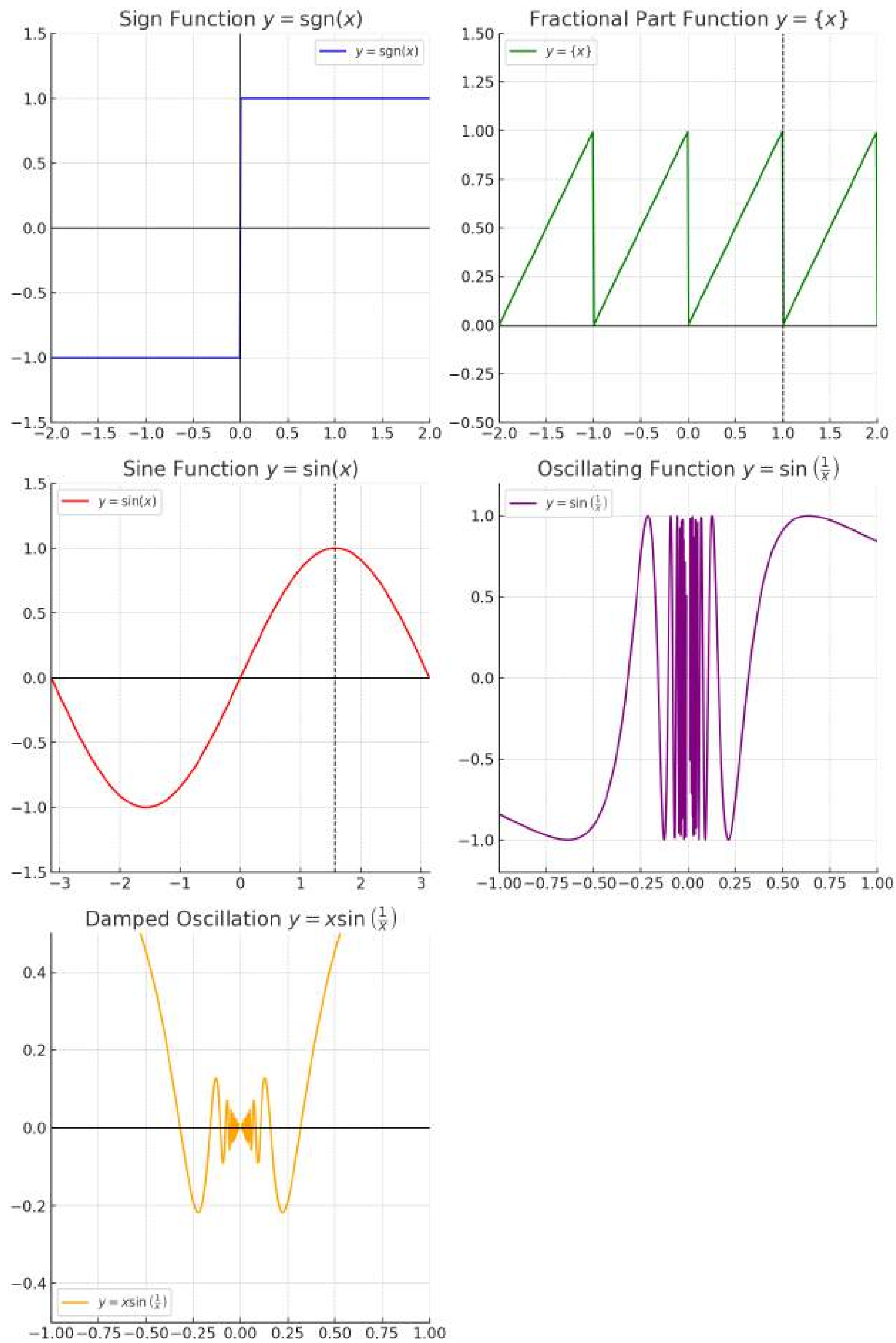


图 1.7: 函数极限的几个具体例子. 或见 desmos.

在这种看法下, 过去对序列极限的 $\varepsilon - N$ 说法, 对函数就换成下面的 $\varepsilon - \delta$ 说法.

定义 1.4.1: 函数的极限: 单侧极限 $x \to a+0$ 和 $x \to a-0$

- 令 $a < b$. 设 $y = f(x)$ 是定义在区间 (a, b) 上的一个函数. 若存在一个常数 l, 对于任意给定的 $\varepsilon > 0$, 无论它多么小, 都存在一个 $\delta = \delta(\varepsilon) > 0$, 使得

$$|f(x) - l| < \varepsilon, \qquad \text{只要 } 0 < x - a < \delta,$$

则我们称, 当 $x \to a+0$ 时, $f(x)$ 以 l 为 **右极限 (right limit)**, 记作

$$\lim_{x \to a+0} f(x) = l,$$

或者说, 当 $x \to a+0$ 时, $f(x)$ 趋向于 l, 记作 $f(x) \to l \ (x \to a+0)$.

- 令 $b < a$. 设 $y = f(x)$ 是定义在区间 (b, a) 上的一个函数. 若存在一个常数 l, 对于任意给定的 $\varepsilon > 0$, 无论它多么小, 都存在一个 $\delta = \delta(\varepsilon) > 0$, 使得

$$|f(x) - l| < \varepsilon, \qquad \text{只要 } 0 < a - x < \delta,$$

则我们称, 当 $x \to a-0$ 时, $f(x)$ 以 l 为 **左极限 (left limit)**, 记作

$$\lim_{x \to a-0} f(x) = l,$$

或者说, 当 $x \to a-0$ 时, $f(x)$ 趋向于 l, 记作 $f(x) \to l \ (x \to a-0)$.

以上两种情况的极限称作 **单侧极限 (one-sided limit)**.

- 对于单侧极限, 只需要函数在点 a 某一侧有定义就足够了, 而且这个单侧的定义域多大或多小都没有关系.

- 上述定义中的 δ 必然是依赖于 ε 的. 通常, ε 越小, 则 δ 也越小.

- 从前面的例子可以看出, 一个函数的右极限与左极限可能不存在, 当存在时可能不相等, 也可能相等.

注释 1.4.1 作弊器: 极限计算器 (symbolab.com) 或者直接画图观察.

1.4.2 自变量趋向点 a 的双侧极限

下面讨论自变量 x 从点 a 的两侧趋向于点 a 的情况. 这种情况下的极限就是双侧极限. 有了单侧极限的讨论, 我们便很容易给出 $x \to a$ 时 $f(x)$ 的极限定义, 只要将前面单侧极限定义中的条件 $0 < x - a < \delta$ 或 $0 < a - x < \delta$ 换成 $0 < |x - a| < \delta$ 即可.

定义 1.4.2: 函数的极限：双侧极限 $x \to a$

设 $y = f(x)$ 是定义在点 a 的某个空心邻域

$$U_r(a) \backslash \{a\} = (a - r, a) \cup (a, a + r)$$

上的一个函数. 若存在一个常数 l, 对于任意给定的 $\varepsilon > 0$, 无论它多么小, 都存在一个 $\delta = \delta(\varepsilon) > 0$, 使得

$$|f(x) - l| < \varepsilon, \qquad \text{只要 } 0 < |x - a| < \delta,$$

则我们称, 当 $x \to a$ 时, $f(x)$ 以 l 为 **极限 (limit)**, 记作

$$\lim_{x \to a} f(x) = l.$$

或者说, 当 $x \to a$ 时, $f(x)$ 趋向于 l, 记作 $f(x) \to l \ (x \to a)$. 当 $x \to a$ 时, 若 $f(x)$ 以某个常数 l 为极限, 则称当 $x \to a$ 时, $f(x)$ 的**极限存在**.

如果熟悉序列极限的定义, 那么就能发现这些定义都是同一个套路.

- 和单侧极限不同, 双侧极限需要函数在点 a 的两侧都有定义, 所以我们取一个空心邻域. 而这个空心邻域多大或多小都没有关系.
- 特别是, 上述定义中的 δ 必然是依赖于 ε 的. 通常, ε 越小, 则 δ 也越小.
- 在上述极限的定义中, 我们没有要求函数 $y = f(x)$ 在点 a 处有定义, 而只是要求它在点 a 的空心邻域内有定义. 这里我们所关心的是, 当自变量 x 自点 a 左, 右两侧趋向于点 a (没有达到点 a) 时, 函数值的变化趋势. 因此, $y = f(x)$ 在点 a 处可以没有定义.
- 即使有定义, $y = f(x)$ 在点 a 处的值 $f(a)$ 与 $x \to a$ 时 $f(x)$ 的极限之间没有必然的联系. 例如, $\lim\limits_{x \to 0} |\operatorname{sgn} x| = 1$, 但是 $|\operatorname{sgn} 0| = 0$. 但是, 我们见到的大量函数 $y = f(x)$ 在某个点 a 处的值 $f(a)$ 与 $x \to a$ 时 $f(x)$ 的极限就是一样的. 此时, 我们称 $y = f(x)$ 在点 a 处是 **连续的**. 函数的连续性是极其极其重要的性质. 尽管我们遇到的大多数函数都是连续函数, 但是不连续的函数也很常见. 函数的连续性是下一节的内容.

从双侧极限和单侧极限的定义中可以看出以下命题.

命题 1.4.1

- 若 $\lim\limits_{x \to a} f(x)$ 存在, 则 $\lim\limits_{x \to a-0} f(x)$ 与 $\lim\limits_{x \to a+0} f(x)$ 也存在并且相等, 都等于 $\lim\limits_{x \to a} f(x)$;
- 反之亦然, 即若 $\lim\limits_{x \to a+0} f(x)$ 与 $\lim\limits_{x \to a-0} f(x)$ 都存在并且相等, 则 $\lim\limits_{x \to a} f(x)$ 也存在并且等于它们.

但是, 若两个单侧极限中有一个不存在, 或虽都存在但不相等, 则双侧极限不存在.


注意 通常我们讲函数的 "极限", 默认指的是双侧极限. 所以, 如果我们说某点处的函数极限不存在, 也仅仅是指双侧极限不存在而已.

定理 1.4.1

如果 $\lim\limits_{x\to a} f(x)$ 存在, 那么这个极限唯一. (证明过程同序列极限.)

例 1.4.1 符号函数 $y = \operatorname{sgn} x$ 在原点 $x = 0$ 处没有双侧极限. 事实上, $\lim\limits_{x\to 0+0} \operatorname{sgn} x = 1$, $\lim\limits_{x\to 0-0} \operatorname{sgn} x = -1$, 故 $y = \operatorname{sgn} x$ 在原点 $x = 0$ 处的双侧极限不存在.

下面我们通过几个具体例子说明如何用 $\varepsilon - \delta$ 说法去证明一个极限式. 下面这些证明题目只是用于帮助读者熟悉函数极限的定义. 实际中最常见的问题是找出某个函数的极限值是什么. 未来我们会学习很多相关的计算技巧. 最常见的方法就是直接代入 $x = a$ 的值到函数中, 如果它有数学意义的话. 比如下面这个例题.

例 1.4.2 证明: $\lim\limits_{x\to 2} \sqrt{3x-2} = 2$.

第一步, 不要着急, 先观察函数的定义域. 发现是 $\left[\dfrac{2}{3}, +\infty\right)$. 这说明该函数在 $x = 2$ 的左右邻域中都是有定义的. 时刻牢记, 在点 a 的 (空心) 邻域内有定义是我们开始工作的前提条件. 假如有人要你找到 $\lim\limits_{x\to 0} \sqrt{3x-2}$ 的极限值, 这显然是没有意义的. 我们还发现对于函数 $\sqrt{3x-2}$, 可以讨论 $x = \dfrac{2}{3}$ 处的右极限, 而无法讨论 $x = \dfrac{2}{3}$ 处的左极限.

证明. 考虑

$$\sqrt{3x-2} - 2 = \frac{3(x-2)}{\sqrt{3x-2}+2} \qquad \left(x \geqslant \frac{2}{3}\right),$$

于是, 我们有

$$|\sqrt{3x-2} - 2| = \left|\frac{3(x-2)}{\sqrt{3x-2}+2}\right| \leqslant \left|\frac{3(x-2)}{2}\right| = \frac{3}{2}|x-2| \qquad \left(x \geqslant \frac{2}{3}\right).$$

对于任意给定的 $\varepsilon > 0$, 要使 $|\sqrt{3x-2} - 2| < \varepsilon$, 只要 $\dfrac{3}{2}|x-2| < \varepsilon$, 即只要 $|x-2| < \dfrac{2}{3}\varepsilon$. 另外, 考虑到函数 $\sqrt{3x-2}$ 的定义域为 $\left[\dfrac{2}{3}, +\infty\right)$. 因此, 应有 $x - 2 \geqslant -\dfrac{4}{3}$. 可见, 为了保证 x 落在 $\sqrt{3x-2}$ 的定义域, 还应要求 $|x-2| < \dfrac{4}{3}$. 这样, 我们取 $\delta = \min\left\{\dfrac{4}{3}, \dfrac{2}{3}\varepsilon\right\}$, 即有

$$|\sqrt{3x-2} - 2| < \varepsilon, \qquad \text{只要 } 0 < |x-2| < \delta.$$

证毕. □

注意 上面的题中, 我考虑到了函数自身的定义域, 最终还要额外求 $|x-2| < \dfrac{4}{3}$; 否则, $f(x)$ 可能会没有定义. 故我们取 $\delta = \min\left\{\dfrac{4}{3}, \dfrac{2}{3}\varepsilon\right\}$. 这种考虑让问题变得更棘手, 其实可以不必担心这一点. 根据极限的定义, 我们关心的是任意小的正数 ε, 而不是比较大的正数 ε. 我们其实可以直接取 $\delta = \dfrac{2}{3}\varepsilon$, 此时, 当 ε 足够小时, $\delta = \dfrac{2}{3}\varepsilon$ 肯定会小于 $\dfrac{4}{3}$. 上面这个题目中, 过于严谨的推导可能会误导读者, 让他们认为函数的特定定义域非常重要, 其实这并不关键. 通常情况下, 我们找到的 $\delta = \delta(\varepsilon)$ 在 $\varepsilon \to 0$ 时也趋于 0. 因此, 当 ε 足够小时, 通过 $0 < |x-a| < \delta$ 这个条件, x 总会被限制在 a 点的某个领域内, 从而确保函数值 $f(x)$ 在这个范围内有定义.

例 1.4.3 证明: $\lim\limits_{x\to2}\dfrac{x^2-x-2}{x^2-4}=\dfrac{3}{4}$.

知道你很急, 但是先别急. 一切和函数的极限相关的问题, 永远是先把函数的定义域搞清楚, 然后检查该函数的在 $x=a$ 的左右两侧是否有定义. 显然, 该函数的定义域是 $\mathbf{R}\setminus\{2,-2\}$. 在 $x=2$ 的两侧 (某个开区间, 其长度无所谓) 是有定义的.

证明. 我们先化简函数:

$$\frac{x^2-x-2}{x^2-4}-\frac{3}{4}=\frac{(x+1)(x-2)}{(x+2)(x-2)}-\frac{3}{4}=\frac{x+1}{x+2}-\frac{3}{4}=\frac{x-2}{4(x+2)}\qquad(x\neq2,-2)$$

对于任意给定的 $\varepsilon>0$, 要使 $\left|\dfrac{x^2-x-2}{x^2-4}-\dfrac{3}{4}\right|<\varepsilon$, 只要 $\left|\dfrac{x-2}{4(x+2)}\right|<\varepsilon$.

由 $\left|\dfrac{x-2}{4(x+2)}\right|<\varepsilon$ 直接解出满足条件的 x 会很困难. 这里我们所关心的是, x 在点 2 附近时函数值的变化趋势, 因此可先约束 x 的范围. 先设 $|x-2|<1$, 此时 $1<x<3$, 进而 $3<x+2<5$, 进而

$$\left|\frac{x-2}{4(x+2)}\right|<\frac{|x-2|}{4\times3}=\frac{|x-2|}{12}.$$

因此只要 x 能满足 $|x-2|<12\varepsilon$, 就有 $\left|\dfrac{x-2}{4(x+2)}\right|<\varepsilon$. 故取 $\delta=\min\{1,12\varepsilon\}$. 当 $0<|x-2|<\delta$ 时, 因为 $\delta\leqslant1$, 所以

$$\left|\frac{x^2-x-2}{x^2-4}-\frac{3}{4}\right|=\left|\frac{x-2}{4(x+2)}\right|<\frac{|x-2|}{12},$$

又 $\delta\leqslant12\varepsilon$, 于是

$$\left|\frac{x^2-x-2}{x^2-4}-\frac{3}{4}\right|<\frac{\delta}{12}\leqslant\varepsilon.$$

证毕. □

注释 1.4.2 **有理式 (rational function)** 是两个多项式的比值 (商). 具体来说, 一个有理式的形式为:

$$f(x)=\frac{P(x)}{Q(x)}.$$

其中, $P(x)$ 和 $Q(x)$ 都是多项式, 且 $Q(x)$ 不为零. 也就是说, 有理式是两个多项式相除的结果. 举例:

1. $\dfrac{2x^2+3x+1}{x-1}$ 是一个有理式, 其中 $P(x)=2x^2+3x+1, Q(x)=x-1$.
2. $\dfrac{1}{x^2-1}$ 也是有理式, 其中 $P(x)=1, Q(x)=x^2-1$.

上题中出现的技巧, 对于适当放大有理式很有启发. 建议多做题多体会.

例 1.4.4 证明: $\lim\limits_{x\to a}\sin x=\sin a$.

证明. 我们有不等式

$$|\sin x-\sin a|=\left|2\sin\frac{x-a}{2}\cos\frac{x+a}{2}\right|\leqslant2\left|\sin\frac{x-a}{2}\right|.$$

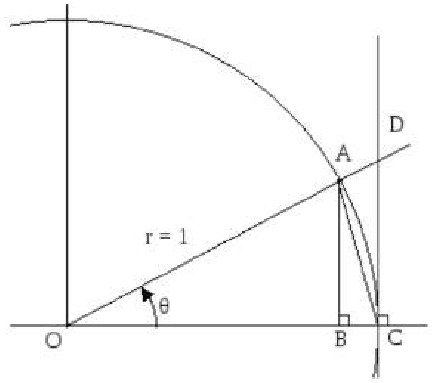


图 1.8: 不等式 $|\sin\theta| \leqslant |\theta|$ 的图示, 其中, $|\sin\theta| = AB$, $|\theta| = \overset{\frown}{AC}$.

利用弧长大于弦长, 由图 1.8 很容易看出 $|\sin\theta| \leqslant |\theta|$. 可见, 由上述不等式有

$$|\sin x - \sin a| \leqslant |x-a|.$$

这样, 要使 $|\sin x - \sin a| < \varepsilon$, 只要 $|x-a| < \varepsilon$ 即可. 对于任意给定的正数 ε, 我们取 $\delta = \varepsilon$, 则有

$$|\sin x - \sin a| < \varepsilon, \quad \text{只要 } 0 < |x-a| < \delta.$$

证毕. □

上题中, 我们是用作图的办法来证明不等式 $|\sin\theta| \leqslant |\theta|$. 这其实不太严谨. 作图可以去印证一个正确的结论, 但正确的结论不能只依靠作图. 所以, 我们来严格证明不等式 $|\sin\theta| \leqslant |\theta|$ 对任意实数 θ 都成立. 本证明内容完全由 AI 大模型 ChatGPT 4o 自动生成. 你只需要输入如下的提问内容:

请你严格证明 `|\sin \theta| \leqslant |\theta|`

后面的数学代码语言是 latex[17], 也就是本讲义的制作语言. 数学是人工智能的核心工具, 现在它又反哺于人类更容易地学习数学.

命题 1.4.2

不等式 $|\sin\theta| \leqslant |\theta|$ 对任意实数 θ 都成立.

证明. 证明步骤:

1. 在 $\theta = 0$ 处成立:
$$\sin 0 = 0 \quad \text{且} \quad |0| = 0 \quad \Rightarrow \quad |\sin 0| = |0|$$
因此, 当 $\theta = 0$ 时, $|\sin\theta| \leqslant |\theta|$ 成立.

2. 在 $\theta \neq 0$ 时, 考察函数: 我们需要证明 $|\sin\theta| \leqslant |\theta|$ 对所有 $\theta \neq 0$ 都成立. 为了简化问题, 我们首先考察 $\theta \geqslant 0$, 因为对于负数情况 $|\sin(-\theta)| = |\sin\theta|$ 以及 $|-\theta| = |\theta|$, 即不等式对负数情况与正数情况相同. 我们现在证明对于 $\theta \geqslant 0$, 有:
$$\sin\theta \leqslant \theta$$

[17] 参考视频 一个非常快速的 Latex 入门教程 _ 哔哩哔哩 _bilibili

3. 比较函数的增长速度: 我们来研究函数 $f(\theta)=\theta-\sin\theta$ 在 $\theta\geqslant 0$ 的性质. 如果能够证明 $f(\theta)\geqslant 0$ 对所有 $\theta\geqslant 0$ 都成立, 则表明 $\sin\theta\leqslant\theta$ 在此范围内成立.

4. 计算导数: 计算 $f(\theta)=\theta-\sin\theta$ 的导数:

$$f'(\theta)=1-\cos\theta$$

由于 $\cos\theta\leqslant 1$, 因此:

$$f'(\theta)\geqslant 0$$

这表明函数 $f(\theta)$ 在 $\theta\geqslant 0$ 是递增的.

5. 初值判断: 计算 $f(0)$:

$$f(0)=0-\sin 0=0$$

由于 $f(\theta)$ 在 $\theta\geqslant 0$ 是递增的, 且 $f(0)=0$, 所以对所有 $\theta\geqslant 0$, 有 $f(\theta)\geqslant 0$, 即:

$$\theta-\sin\theta\geqslant 0\quad\Rightarrow\quad\sin\theta\leqslant\theta$$

6. 结论: 通过上述步骤, 我们得出对于所有实数 θ, 都有:

$$|\sin\theta|\leqslant|\theta|.$$

□

总结: 从上面几个例子中我们看出, 若要证明一个函数 $f(x)$ 当 $x\to a$ 时的极限为 l, 首先要考虑 $|f(x)-l|$, 并对它做适当的变形和放大, 尽可能使之成为依赖于 $|x-a|$ 的一个量, 然后让放大后的表达式小于给定的 ε, 从而决定要找的 $\delta=\delta(\varepsilon)$.

1.4.3 函数极限的定理

有关序列极限的全部定理对于函数的极限都成立, 其证明也类似. 现在我们仅就双侧极限的情况将它们列出而略去其证明.

夹逼定理

定理 1.4.2: 函数极限的夹逼定理

设有 $f(x),g(x)$ 及 $h(x)$ 三个函数定义在点 a 的一个空心邻域内, 并且在这个邻域中满足不等式

$$h(x)\leqslant f(x)\leqslant g(x).$$

假如 $\lim\limits_{x\to a}g(x)=l$ 且 $\lim\limits_{x\to a}h(x)=l$, 则 $\lim\limits_{x\to a}f(x)=l$.

上述定理的表述使用的是双侧极限, 但对于单侧极限也是完全成立的. 只要考虑三个函数在点 a 的一个单侧区间内有定义且满足夹逼的不等式即可.

问题 "有关序列极限的全部定理对于函数的极限都成立, 其证明也类似." 真的类似吗? 你说类似就类似啊. 我不信. 我要自己试着想想看: 能否真的用序列极限的夹逼定理的证明思路来得到定理 1.4.2 的证明.

糊弄积累得的多了, 做什么事情都爱糊弄.

例 1.4.5 证明: $\lim\limits_{x\to 0}\dfrac{\sin x}{x}=1$.

前面已经指出极限 $\lim\limits_{n\to\infty}\left(1+\dfrac{1}{n}\right)^n=\mathrm{e}$ 是一个重要极限. 现在的这个极限 $\lim\limits_{x\to 0}\dfrac{\sin x}{x}=1$ 是微积分中又一个重要极限, 其重要性在后面讨论导数时会看出: 利用它能严格导出 $\sin x$ 的导数为 $\cos x$.

证明. 证我们先考虑 $0<x<\dfrac{\pi}{2}$ 的情况. 为此, 我们作图 1.9, 其中圆半径为 1, 而圆弧 $\widehat{AD}$ 的长度为 x, 即它所对的圆心角的弧度为 x. 一方面, 从图 1.9 中容易看出[18]

$$\sin x<x.$$

另一方面, 扇形 OAD 的面积[19]小于三角形 OCD 的面积, 即

$$\frac{1}{2}x<\frac{1}{2}\tan x.$$

综上所述, 我们有

$$\sin x<x<\frac{\sin x}{\cos x},\quad 0<x<\frac{\pi}{2},$$

即

$$\cos x<\frac{\sin x}{x}<1,\quad 0<x<\frac{\pi}{2}.$$

上式对 $-\dfrac{\pi}{2}<x<0$ 的情况也成立. 这是因为 $\cos x$ 与 $\dfrac{\sin x}{x}$ 都是偶函数, 即当将 x 换成 $-x$ 时其函数值不变. 注意到 $\lim\limits_{x\to 0}\cos x=1$ 及常数函数 $y=1$ 的极限是 1, 由上式及定理 1.4.2 立即得出我们所要的结论. 证毕. □

注意, 在上一道题中, 我们仅在 0 的一个空心邻域, 即 $(-\dfrac{\pi}{2},0)\cup(0,\dfrac{\pi}{2})$ 上时才有不等式关系 $\sin x<x<\dfrac{\sin x}{\cos x}$. 对于其他的一些 x, 该不等式不成立. 但是对于运用定理 1.4.2 来讲, 这就足够了. 再次强调, 我们只需要 $h(x)\leqslant f(x)\leqslant g(x)$ 在 a 空心邻域内成立就够用. 但有时也会遇到三个函数对于所有 x 都有不等式成立, 自是最好不过了.

问题 我们说过, 作图来证明不是好的办法. 请自行用别的方法证明: 当 $0<x<\dfrac{\pi}{2}$ 时, 有 $x<\dfrac{\sin x}{\cos x}$.

例 1.4.6 求极限 $\lim\limits_{x\to 0}x\left[\dfrac{1}{x}\right]$.

[18] 其实命题 1.4.2 已经证明了.

[19] 圆中扇形面积的计算公式是 $S=\dfrac{1}{2}r^2\theta$. 其中, r 表示圆的半径; θ 表示扇形对应的圆心角, 单位为弧度.

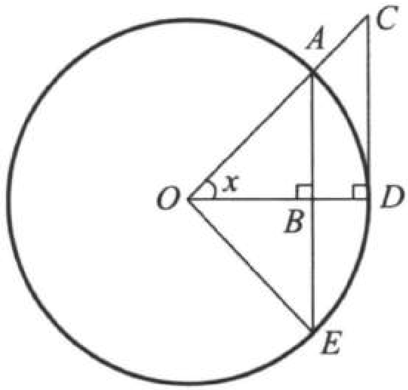


图 1.9: 利用几何图形证明 $\sin x < x < \dfrac{\sin x}{\cos x}$, 当 $0 < x < \dfrac{\pi}{2}$. 其中, 线段 CD 的长度是 $\tan x$.

证明. 一般地, 对于任意的 $x \neq 0$, 我们有

$$[x] \leqslant x < [x] + 1.$$

如果 $x \leftarrow \dfrac{1}{x}$, 则有

$$\left[\frac{1}{x}\right] \leqslant \frac{1}{x} < \left[\frac{1}{x}\right] + 1.$$

两边同时减去 1 并重排, 则有

$$\frac{1}{x} - 1 < \left[\frac{1}{x}\right] \leqslant \frac{1}{x}. \tag{1.8}$$

- 若 $x > 0$, 对 (1.8) 同时乘以 x, 则有 $1 - x < x\left[\dfrac{1}{x}\right] \leqslant 1$. 而 $\lim\limits_{x \to 0+0}(1-x) = 1$. 利用 (单侧极限下的) 夹逼定理, 有

$$\lim_{x \to 0+0} x\left[\frac{1}{x}\right] = 1.$$

- 若 $x < 0$, 对 (1.8) 同时乘以 x, 则 $1 \leqslant x\left[\dfrac{1}{x}\right] < 1 - x$. 同理可得

$$\lim_{x \to 0-0} x\left[\frac{1}{x}\right] = 1.$$

函数 $y = x\left[\dfrac{1}{x}\right]$ 在点 0 处的左极限和右极限都存在且等于 1, 于是 $\lim\limits_{x \to 0} x\left[\dfrac{1}{x}\right] = 1$. 其图像见图 1.10. □

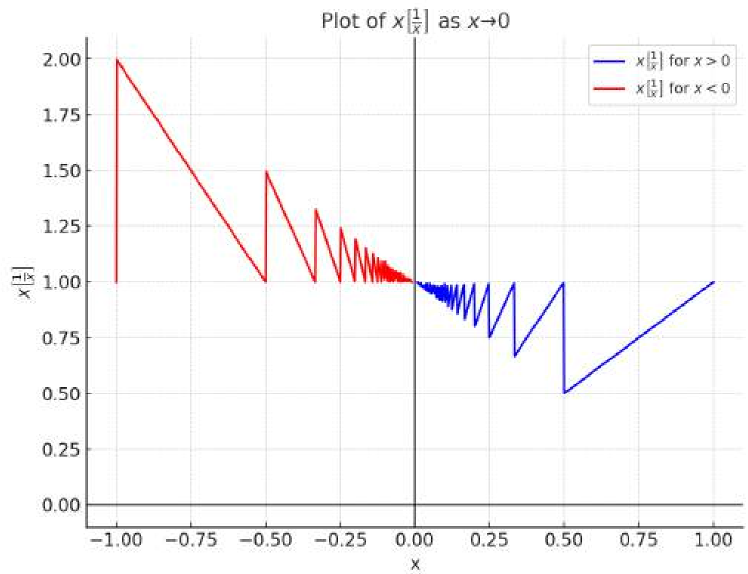


图 1.10: 函数 $x\left[\dfrac{1}{x}\right]$ 的图像. Geogebra 可能画不出来函数 $y=[x]$ ，可以用 python 作图，代码见 plots/-plot_x_1_x.py · 高等数学 B-07 班-资料共享 - Gitee.com.

四则运算

定理 1.4.3: 函数极限的四则运算

设 $f(x)$ 及 $g(x)$ 是定义在点 a 的一个空心邻域内的函数. 若

$$\lim_{x\to a} f(x)=l_1,\quad \lim_{x\to a} g(x)=l_2,$$

则有

$$\lim_{x\to a}(f(x)\pm g(x))=l_1\pm l_2,\quad \lim_{x\to a} f(x)g(x)=l_1 l_2.$$

特别地, 如果 $g(x)=c$ 是某个常函数, 则有

$$\lim_{n\to\infty} cf(x)=cl_1.$$

并且当 $l_2\neq 0$ 时, 有

$$\lim_{x\to a}\frac{f(x)}{g(x)}=\frac{l_1}{l_2}.$$

上述定理的表述使用的是双侧极限, 但对于单侧极限也是完全成立的. 只要考虑三个函数在点 a 的一个单侧区间内有定义且满足夹逼的不等式即可.

注意 自行思考: $\lim\limits_{x\to 0} x\cdot\dfrac{1}{x}=1$, 其中 $\lim\limits_{x\to 0} x=0$ 极限存在, 但是 $\lim\limits_{x\to 0}\dfrac{1}{x}$ 趋向于无穷大. 这个结论是否和上述定理矛盾呢.

例 1.4.7 求极限 $\lim\limits_{x\to 0}\dfrac{\sqrt{1+x}-1}{x+\sin x}$.

首先明确该函数的定义域是 $\{-1,0\}\cup\{0,+\infty\}$. 注意到: 可以通过求 $x+\sin x$ 的导数知道它是单调递增函数, 而 $0+\sin 0=0$. 所以分母部分只有一个零点 (根).

证明. 这个求极限的函数表达式中分子与分母的极限都是零, 因而不能直接应用定理 1.4.3. 先将分子与分母除以 x (注意在 $x\to 0$ 的过程中 $x\neq 0$), 立即得到

$$\frac{\sqrt{1+x}-1}{x+\sin x}=\frac{\frac{\sqrt{1+x}-1}{x}}{1+\frac{\sin x}{x}}.$$

在做了这样的形变后, 分母的极限不再为零, 这时便可应用定理 1.4.3:

- (分母) 当 $x\to 0$ 时, 分母部分 $1+\dfrac{\sin x}{x}\to 2$, 这里直接使用例 1.4.5 的结果.
- (分子) 当 $x\to 0$ 时, 分子部分再次出现分子与分母的极限都是零的情况, 再一次做变形:
$$\frac{\sqrt{1+x}-1}{x}=\frac{1+x-1}{x(\sqrt{1+x}+1)}=\frac{1}{\sqrt{1+x}+1}\to\frac{1}{2}.$$
这里用到 $\sqrt{1+x}\to 1\ (x\to 0)$ 这一事实.

综合以上结果, 于是 $\lim\limits_{x\to 0}\dfrac{\sqrt{1+x}-1}{x+\sin x}=\dfrac{1}{4}$. □

例 1.4.8 求极限 $\lim\limits_{x\to 0}\dfrac{\sin 5x-\tan 3x}{2x}$.

把该函数的定义域完整写下来并不难. 但是聪明的你已经发现了一个事实, 只要我们能够确认在 0 的某个空心邻域, 该函数是有定义的就足够了.

证明. 这个函数表达式中分子与分母的极限都是零, 因而不能直接应用定理 1.4.3. 我们将函数拆分成两部分: $\dfrac{\sin 5x}{2x}-\dfrac{\tan 3x}{2x}$, 然后分别求极限.

- 由 $\dfrac{\sin 5x}{2x}=\dfrac{\sin 5x}{5x}\cdot\dfrac{5}{2}$ 可以推
$$\lim_{x\to 0}\frac{\sin 5x}{2x}=\lim_{x\to 0}\frac{\sin 5x}{5x}\cdot\frac{5}{2}=\lim_{y\to 0}\frac{\sin y}{y}\cdot\frac{5}{2}=\frac{5}{2}.$$
这里我们做了一个变量替换 $y:=5x$, 并且再次直接使用例 1.4.5 的结果[20].
- 同理, 可以推出
$$\begin{aligned}\lim_{x\to 0}\frac{\tan 3x}{2x}&=\lim_{x\to 0}\left(\frac{1}{\cos 3x}\cdot\frac{\sin 3x}{3x}\cdot\frac{3}{2}\right)\\&=\left(\lim_{x\to 0}\frac{1}{\cos 3x}\right)\left(\lim_{x\to 0}\frac{\sin 3x}{3x}\right)\left(\frac{3}{2}\right)\\&=\left(\lim_{y\to 0}\frac{1}{\cos y}\right)\left(\lim_{y\to 0}\frac{\sin y}{y}\right)\left(\frac{3}{2}\right)\\&=\frac{3}{2}.\end{aligned}$$
注意, 上面的第二个等式利用了利用定理 1.4.3. 这里还用到 $\lim\limits_{y\to 0}\cos y=1$.

[20] 为什么说极限 $\lim\limits_{x\to 0}\dfrac{\sin x}{x}=1$ 很重要, 因为很多其他极限最终会归结为这个形式, 然后我们直接使用结果即可.

总结前面的结果, 再次利用定理 1.4.3, 我们有 $\lim\limits_{x\to 0}\dfrac{\sin 5x-\tan 3x}{2x}=\lim\limits_{x\to 0}\dfrac{\sin 5x}{2x}-\lim\limits_{x\to 0}\dfrac{\tan 3x}{2x}=\dfrac{5}{2}-\dfrac{3}{2}=1.$ □

注意 注意, 定理 1.4.3 告诉我们下面的等式成立:

$$\lim_{x\to 0}\left(\frac{1}{\cos 3x}\cdot\frac{\sin 3x}{3x}\cdot\frac{3}{2}\right)=\left(\lim_{x\to 0}\frac{1}{\cos 3x}\right)\left(\lim_{x\to 0}\frac{\sin 3x}{3x}\right)\left(\frac{3}{2}\right).$$

但我们应该牢记心中, 该式子成立的前提是三个极限值 $\lim\limits_{x\to 0}\dfrac{1}{\cos 3x}$, $\lim\limits_{x\to 0}\dfrac{\sin 3x}{3x}$ 和 $\lim\limits_{x\to 0}\dfrac{3}{2}$ 都分别独立存在. 按道理, 我们是从上式的右侧到得了左侧, 但是实际做题时, 我们习惯性的从上式的左侧得到右侧. 所以在使用定理 1.4.3 是要留个心眼儿. 不然可能最后得到错误的结论.

保序性

序列极限中有关极限不等式的定理, 在函数极限中同样成立.

定理 1.4.4: 函数极限的保序性

设 $f(x)$ 及 $g(x)$ 是定义在点 a 的一个空心邻域内的函数, 且

$$\lim_{x\to a}f(x)=l_1,\ \lim_{x\to a}g(x)=l_2.$$

1. 若 $l_1>l_2$, 则存在一个 $\delta>0$, 使得

$$f(x)>g(x),\qquad \text{只要 } 0<|x-a|<\delta.$$

2. 若在这个空心邻域内满足 $f(x)\geqslant g(x)$, 则 $l_1\geqslant l_2$.

推论 1.4.1

设 $\lim\limits_{x\to a}f(x)=l$, 且 $l>0$ (或 $l<0$), 则存在一个 $\delta>0$, 使得

$$f(x)>0\ (\text{或 } f(x)<0),\qquad \text{只要 } 0<|x-a|<\delta.$$

也就是说, 若一个函数当自变量趋向于一点 a 时的极限大于零, 则它在点 a 的某个小的空心邻域内都大于零. 这个结论很容易从定理 1.4.4 的第一个结果中推出, 只要将其中的数 $g(x)$ 取成恒等于零的函数即可. 它的几何意义也是十分明显的.

现在我们来讨论函数极限与序列极限之间的关系. 我们有下面的定理.

函数极限与序列极限

定理 1.4.5: 函数极限与序列极限的关系

设函数 $f(x)$ 在点 a 的一个空心邻域 $\overset{\circ}{U}_r(a)$ 内有定义. 则下面两个条件互为充分必要条件.

1. $\lim\limits_{x\to a} f(x)=l$.
2. 对于在该空心邻域 $\overset{\circ}{U}_r(a)$ 内取值的任一序列 $\{x_n\}$, 都有:

$$\lim_{n\to\infty} x_n = a \Longrightarrow \lim_{n\to\infty} f(x_n)=l.$$

注意, 函数值序列 $\{f(x_n)\}$ 是一个新的序列, 而不是函数. 该定理还可以修改成单侧极限的版本, 则序列 $\{x_n\}$ 定义在 a 的某个单侧开区间中.

注意 上述定理中, 要求 $\{x_n\}$ 是一个在该空心邻域 $\overset{\circ}{U}_r(a)$ 内取值的序列, 意味着对于所有 $n=1,2,\ldots$, 都有 $x_n \in \overset{\circ}{U}_r(a)$. 其实这一条件还可以再放松一下, 即只要求 "当 n 充分大时, $\{x_n\}$ 是一个在该空心邻域 $\overset{\circ}{U}_r(a)$ 内取值的序列". 此时定理的结论仍然成立.

对应于定理 1.4.5 的课本内容只陈述了从 1 到 2 的事实, 这里我们补充了从 2 到 1 的事实. 这样子, 我们就知道为什么序列极限和函数极限有着如此紧闭的联系. 以下仅证明了从 1 到 2 的推导, 反向推导略去.

证明. 根据假定 $\lim\limits_{x\to a} f(x)=l$, 对于任意给定的 $\varepsilon>0$, 存在一个 $\delta>0$, 使得

$$|f(x)-l|<\varepsilon, \quad \text{只要 } 0<|x-a|<\delta.$$

特别地, $|f(x_n)-l|<\varepsilon$, 只要 $0<|x_n-a|<\delta$.

- 根据定理中 $\lim\limits_{n\to\infty} x_n=a$ 的假定, 对于这个 $\delta>0$, 存在一个正整数 $N\equiv N(\delta)$, 使得

$$|x_n-a|<\delta, \quad \text{只要 } n>N.$$

- 注意到序列 $\{x_n\}$ 只在点 a 的空心邻域内取值, 故对于一切 n, 有 $x_n\neq a$, 即有 $|x_n-a|>0$.

于是, 我们有 $0<|x_n-a|<\delta$, 只要 $n>N$. 最终, 我们得到

$$|f(x_n)-l|<\varepsilon, \quad \text{只要 } n>N.$$

证毕. □

特别是, $f(x)$ 在 $x\to 0$ 过程中极限是 l 的充分且必要条件为: 对于 0 某个空心邻域内的任一序列 $\{x_n\}$, 都有 $\lim\limits_{n\to\infty} x_n=0 \Longrightarrow \lim\limits_{n\to\infty} f(x_n)=l$. 定理 1.4.5 为我们提供了一种证明函数极限不存在的方法.

推论 1.4.2

设函数 $f(x)$ 在点 a 的某个空心邻域内有定义. 设有两个序列 $\{x_n'\}$ 与 $\{x_n''\}$, 它们都在该空心邻域内取值, 且当 $n\to\infty$ 时都以 a 为极限. 如果极限 $\lim\limits_{n\to\infty} f(x_n')$ 与 $\lim\limits_{n\to\infty} f(x_n'')$ 都存在, 但不相等, 则当 $x\to a$ 时 $f(x)$ 的极限不存在.

例 1.4.9 当 $x \to 0$ 时, 函数 $\sin \dfrac{1}{x}$ 没有极限. 事实上,

- 取 $x_n' = \dfrac{1}{2n\pi}$, 则 $\sin \dfrac{1}{x_n'} = 0$;
- 取 $x_n'' = \left(2n\pi + \dfrac{\pi}{2}\right)^{-1}$, 则 $\sin \dfrac{1}{x_n''} = 1$.

这时, 显然有 $x_n' \to 0\ (n \to \infty)$, $x_n'' \to 0\ (n \to \infty)$, 但 $\lim\limits_{n\to\infty} \sin \dfrac{1}{x_n'} \neq \lim\limits_{n\to\infty} \sin \dfrac{1}{x_n''}$. 因此, 当 $x \to 0$ 时, 函数 $\sin \dfrac{1}{x}$ 没有极限.

下面这个函数虽然定义在整个 $\mathbf{R}$ 上, 但它在任何一点处的极限都不存在.

例 1.4.10 讨论 $\lim\limits_{x\to a} \mathrm{D}(x)$ 的存在性[21]. 其中, a 是任意实数,

$$\mathrm{D}(x) = \begin{cases} 1, & x \text{ 为有理数}, \\ 0, & x \text{ 为无理数}, \end{cases}$$

是狄利克雷 (Dirichlet) 函数.

利用习题 1.1 中的第 7, 8 题可知, 有理数集与无理数集在数轴 $\mathbf{R}$ 上都是 **处处稠密的 (everywhere dense)**, 即, 设 (a, b) 为任意一个开区间, 则

- (a, b) 中必有有理数;
- (a, b) 中必有无理数.

一言以蔽之就是, 任何一个开区间内都存在有理数和无理数. 这一结论直觉上是十分合理的. 它的证明我们不需要掌握.

因此, 对于任意正整数 n, 开区间 $\left(a - \dfrac{1}{n}, a + \dfrac{1}{n}\right)$ 中总有有理数, 我们任取一个, 记为 x_n'; 也总有无理数, 我们任取一个, 记为 x_n'', 则

$$a - \frac{1}{n} < x_n' < a + \frac{1}{n}, \quad a - \frac{1}{n} < x_n'' < a + \frac{1}{n}.$$

利用序列极限的夹逼定理, 于是有 $x_n' \to a\ (n \to \infty)$, $x_n'' \to a\ (n \to \infty)$. 另一方面, 我们注意到,

$$\begin{aligned} \mathrm{D}(x_n') &= 1, \quad \lim_{n\to\infty} \mathrm{D}(x_n') = 1, \\ \mathrm{D}(x_n'') &= 0, \quad \lim_{n\to\infty} \mathrm{D}(x_n'') = 0. \end{aligned}$$

因此, $\lim\limits_{x\to a} \mathrm{D}(x)$ 不存在. 由于 a 是任意的, 所以我们有结论: 狄利克雷函数处处无极限.

1.4.4 自变量趋向于无穷大时的极限

正如本节开始时所指出的, 自变量 x 趋向于无穷大有三种情况: $x \to +\infty, x \to -\infty$ 及 $x \to \infty$. 前两种情况意义十分明白, 而第三种情况 $x \to \infty$ 则要特别注意它要求自变量 x 沿 x 轴的正向与负向同时无限远离原

[21] 课本的例题讨论的是 $\lim\limits_{x\to a} x\mathrm{D}(x)$, 它只有在 $a = 0$ 处有极限, 其他地方也是处处无极限. 这个两个问题内容相近, 只保留一个在本讲义.

点, 也相当于要求 $|x| \to +\infty$. 因此, $x \to \infty$ 的过程实际上包含了 $x \to +\infty$ 与 $x \to -\infty$ 两个过程, 这一点读者应当注意.

定义 1.4.3: 函数的极限: 单侧极限 $x \to \pm\infty$

- 设函数 $f(x)$ 在区间 $(a, +\infty)$ 内有定义. 若有一个常数 l, 对于任意给定的正数 ε, 都存在一个 $A = A(\varepsilon) > a$, 使得
$$|f(x) - l| < \varepsilon, \quad \text{只要 } x > A,$$
则我们称, 当 $x \to +\infty$ 时, $f(x)$ 以 l 为极限, 记作 $\lim\limits_{x \to +\infty} f(x) = l$.
- 设函数 $f(x)$ 在区间 $(-\infty, a)$ 内有定义. 若有一个常数 l, 对于任意给定的正数 ε, 都存在一个 $A = A(\varepsilon) < a$, 使得
$$|f(x) - l| < \varepsilon, \quad \text{只要 } x < A,$$
则我们称, 当 $x \to -\infty$ 时, $f(x)$ 以 l 为极限, 记作 $\lim\limits_{x \to -\infty} f(x) = l$.

例 1.4.11 证明: $\lim\limits_{x \to +\infty} \mathrm{e}^{-x} = 0$.

证明. 对于任意给定的 $\varepsilon > 0$, 要使 $\left|\mathrm{e}^{-x} - 0\right| < \varepsilon$, 只要 $\mathrm{e}^{-x} < \varepsilon$, 即只要 $x > -\ln \varepsilon$. 因此, 可取 $A = -\ln \varepsilon$, 这时有
$$\left|\mathrm{e}^{-x} - 0\right| < \varepsilon, \quad \text{只要 } x > A.$$
证毕. □

定义 1.4.4: 函数的极限: 双侧极限 $x \to \infty$

令 $a > 0$. 设函数 $f(x)$ 在区间 $\{x : |x| > a\}$ 内有定义. 若有一个常数 l, 对于任意给定的 $\varepsilon > 0$, 存在一个 $A > a$, 使得
$$|f(x) - l| < \varepsilon, \quad \text{只要 } |x| > A,$$
则我们称, 当 $x \to \infty$ 时, $f(x)$ 以 l 为极限, 记作 $\lim\limits_{x \to \infty} f(x) = l$.

从双侧极限和单侧极限的定义中可以看出以下命题.

命题 1.4.3

- 若 $\lim\limits_{x \to \infty} f(x)$ 存在, 则 $\lim\limits_{x \to +\infty} f(x)$ 与 $\lim\limits_{x \to -\infty} f(x)$ 也存在并且相等, 都等于 $\lim\limits_{x \to \infty} f(x)$;
- 反之亦然, 即若 $\lim\limits_{x \to +\infty} f(x)$ 与 $\lim\limits_{x \to -\infty} f(x)$ 都存在并且相等, 则 $\lim\limits_{x \to \infty} f(x)$ 也存在并且等于它们.

但是, 若 $\lim\limits_{x \to +\infty} f(x)$ 与 $\lim\limits_{x \to -\infty} f(x)$ 中有一个不存在, 或它们都存在但不相等, 则 $\lim\limits_{x \to \infty} f(x)$ 不存在. 例如, $\lim\limits_{x \to \infty} \mathrm{e}^{-x}$ 不存在, 因为 $\lim\limits_{x \to -\infty} \mathrm{e}^{-x}$ 不存在. 又如, $\lim\limits_{x \to \infty} \operatorname{sgn} x$ 也不存在, 因为 $\lim\limits_{x \to +\infty} \operatorname{sgn} x = 1$, 而 $\lim\limits_{x \to -\infty} \operatorname{sgn} x =$

-1.

根据定义 1.4.4 很容易看出, 当 $\lim\limits_{x\to\infty} f(x)$ 存在时, $\lim\limits_{y\to 0} f\left(\frac{1}{y}\right)$ 也存在, 且有

$$\lim_{x\to\infty} f(x) = \lim_{y\to 0} f\left(\frac{1}{y}\right);$$

反之亦然.

我们还要指出: 关于极限过程 $x \to a$ 的夹逼定理, 极限四则运算以及极限不等式 (保序性), 对 $x \to +\infty, x \to -\infty$ 以及 $x \to \infty$ 等极限过程也都成立, 这里不再重述.

例 1.4.12 求极限 $\lim\limits_{x\to\infty} \dfrac{x^4+x^2+1}{(x+1)^4}$.

证明. 做如下变形 (分子分母同除以 x^4):

$$\frac{x^4+x^2+1}{(x+1)^4} = \frac{1+\frac{1}{x^2}+\frac{1}{x^4}}{\left(1+\frac{1}{x}\right)^4}.$$

设 n 为正整数. 由于 $\lim\limits_{y\to 0} y^n = 0$, 所以 $\lim\limits_{x\to\infty} \dfrac{1}{x^n} = 0$, 于是由极限四则运算可以得到

$$\lim_{x\to\infty} \frac{x^4+x^2+1}{(x+1)^4} = \frac{\lim_{x\to\infty}\left(1+\frac{1}{x^2}+\frac{1}{x^4}\right)}{\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^4} = 1.$$

□

下面, 我们通过例子的形式来讨论一个重要极限 $\lim\limits_{x\to\infty}\left(1+\dfrac{1}{x}\right)^x = \mathrm{e}$, 它是序列极限 $\lim\limits_{n\to\infty}\left(1+\dfrac{1}{n}\right)^n = \mathrm{e}$ 的推广. 我们可以推出 $\lim\limits_{y\to 0}(1+y)^{\frac{1}{y}} = \mathrm{e}$. 这只要做变量替换 $y = \dfrac{1}{x}$ 即可.

例 1.4.13 证明: $\lim\limits_{x\to\infty}\left(1+\dfrac{1}{x}\right)^x = \mathrm{e}$.

证明. 对于任意 x, 一般有

$$[x] \le x < [x]+1.$$

若只考虑 $x > 1$, 上式取倒数后再加 1 可得

$$1+\frac{1}{[x]+1} < 1+\frac{1}{x} \le 1+\frac{1}{[x]}.$$

所以, 我们有

$$\left(1+\frac{1}{[x]+1}\right)^{[x]} \leqslant \left(1+\frac{1}{x}\right)^x \leqslant \left(1+\frac{1}{[x]}\right)^{[x]+1}.$$

注意, 上面这一串不等式成立的原因在于它们的底数和指数都呈现链式的不等式关系.

- 先考虑 $x \to +\infty$ 的极限过程. 记 $n := [x]$. 若 $x \to +\infty$, 则有 $n \to +\infty$. 此时函数极限正好变成了序列极限:

$$\lim_{x\to+\infty}\left(1+\frac{1}{[x]+1}\right)^{[x]} = \lim_{n\to\infty}\left(1+\frac{1}{n+1}\right)^n = \lim_{n\to\infty}\left(1+\frac{1}{n+1}\right)^{n+1}\left(1+\frac{1}{n+1}\right)^{-1} = \mathrm{e},$$

以及,

$$\lim_{x\to+\infty}\left(1+\frac{1}{[x]}\right)^{[x]+1}=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{n+1}=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{n}\left(1+\frac{1}{n}\right)=\mathrm{e}.$$

由夹逼定理即得

$$\lim_{x\to+\infty}\left(1+\frac{1}{x}\right)^{x}=\mathrm{e}.$$

- 现在考虑 $x\to-\infty$ 的极限过程. 令 $y=-x$, 则 $y\to+\infty$. 利用上面所得的结果, 有

$$\begin{aligned}\lim_{x\to-\infty}\left(1+\frac{1}{x}\right)^{x}&=\lim_{y\to+\infty}\left(1-\frac{1}{y}\right)^{-y}\\&=\lim_{y\to+\infty}\left(1+\frac{1}{y-1}\right)^{y}=\lim_{y\to+\infty}\left(1+\frac{1}{y-1}\right)^{y-1}\left(1+\frac{1}{y-1}\right)=\mathrm{e}.\end{aligned}$$

综上, 由 $\lim\limits_{x\to+\infty}\left(1+\frac{1}{x}\right)^{x}=\lim\limits_{x\to-\infty}\left(1+\frac{1}{x}\right)^{x}=\mathrm{e}$ 即得 $\lim\limits_{x\to\infty}\left(1+\frac{1}{x}\right)^{x}=\mathrm{e}$. 证毕. □

现在证明了极限 $\lim\limits_{x\to\infty}\left(1+\frac{1}{x}\right)^{x}=\mathrm{e}$, 其中 e 是自然对数的底. 这使我们不能不提到它的引入者——欧拉. [天才简史 - 欧拉] 神级人物欧拉究竟多厉害? 所有学生的噩梦, 小说都不敢这么写! _ 哔哩哔哩 _bilibili

例 1.4.14 设 k 为正整数, 证明:

1. $\lim\limits_{x\to\infty}\left(1+\frac{1}{x}\right)^{kx}=\mathrm{e}^k$;
2. $\lim\limits_{x\to 0}(1+kx)^{\frac{1}{x}}=\mathrm{e}^k$.

证明. 我们分别考虑.

1. 由

$$\left(1+\frac{1}{x}\right)^{kx}=\left[\left(1+\frac{1}{x}\right)^{x}\right]^{k}$$

及极限的乘法公式, 立即推出结论. 注意, $\lim\limits_{x\to\infty}\left(1+\frac{1}{x}\right)^{x}=e$ 是重要的函数极限, 是值得把它直接的记在心里的.

2. 我们令 $y=\frac{1}{kx}$, 则有 $ky=\frac{1}{x}$, 和

$$(1+kx)^{\frac{1}{x}}=\left(1+\frac{1}{y}\right)^{ky}.$$

注意到当 $x\to 0$ 时, $y\to\infty$, 于是由前面第一问又得

$$\lim_{x\to 0}(1+kx)^{\frac{1}{x}}=\lim_{y\to\infty}\left(1+\frac{1}{y}\right)^{ky}=\mathrm{e}^k.$$

证毕.

□

注释 1.4.3 自变量 x 趋向于无穷大包括 $x\to+\infty, x\to-\infty$ 及 $x\to\infty$. 可以用 $y=\frac{1}{x}$ 做代换, 则它们分别变成 $y\to 0+0, y\to 0-0$ 及 $y\to 0$. 反之亦然. 这种代换方法是非常常见和使用.

1.4.5 无穷大量

最后, 我们介绍无穷大量的概念. 存在一些函数, 其绝对值在自变量的一个特定变化过程中无限制地增大. 比如, 当 $x \to +\infty$ 时, e^x 无限增大. 这种绝对值无限增大的量就称为无穷大量. 无穷大量的确切定义如下:

定义 1.4.5: 无穷大量

设函数 $f(x)$ 在点 x_0 的一个空心邻域内有定义. 若对于任意给定的正数 M, 不论它有多么大, 总存在一个 $\delta = \delta(M) > 0$, 使得

$$|f(x)| > M, \qquad \text{只要 } 0 < |x - x_0| < \delta,$$

则称, 当 $x \to x_0$ 时, $f(x)$ 为 **无穷大量**, 记作

$$\lim_{x \to x_0} f(x) = \infty$$

或 $f(x) \to \infty$ $(x \to x_0)$. 读者可以仿照此定义自己给出自变量在其他变化过程中的无穷大量的定义. 无穷大量又有两种特殊情况: $f(x) \to +\infty \, (x \to x_0)$ 及 $f(x) \to -\infty \, (x \to x_0)$. 它们的含义是清楚的, 读者自己可以给出其确切定义.

显然, 若当 $x \to x_0$ 时, $f(x)$ 是无穷大量, 则当 $x \to x_0$ 时, $f(x)$ 没有极限. 反之并不成立, 见例 1.4.10.

例 1.4.15 证明: 函数 $f(x) = \dfrac{1}{x} \cos \dfrac{\pi}{x}$ 在点 $x = 0$ 的任意一个邻域内都是无界的, 但当 $x \to 0$ 时, $f(x)$ 不是无穷大量. 见 desmos.

这里应该特别提醒读者: 尽管将无穷大量 $f(x)$ 写成 $\lim\limits_{x \to x_0} f(x) = \infty$ 的形式, 但是我们仍然认为当 $x \to x_0$ 时, $f(x)$ 的极限不存在; 所以, 通常关于极限四则运算的定理对于无穷大量而言不成立.

发散到正无穷大或负无穷大只是函数的极限不存在的一种特殊情况, 还有另一类震荡的情况, 如 $\lim\limits_{x \to x_0} \sin(x)$, 不发散到无穷大, 但同样不收敛.

1.5 连续函数

本节要讨论一类重要函数, 即连续函数. 我们所常见的所有基本初等函数在其定义域内都是连续函数. 从直观上看, 所谓的连续函数就是指这样一类函数, 其图形是一条连续的曲线[22]. 连续函数是微积分研究的基本对象. 连续函数的基本性质构成微积分的重要基础.

1.5.1 连续性的定义

我们从具体例子讲起. 从直观上看, 函数 $y = \sin x$ 的图形是一条连续的曲线, 而函数 $y = \operatorname{sgn} x$ 的图形是一条断开的曲线. 那么, 用什么办法去刻画一个函数的连续性呢? 我们注意到, 对于函数 $y = \sin x$, 在任意一点 x_0 处都有

$$\lim_{x \to x_0} \sin x = \sin x_0,$$

即函数 $y = \sin x$ 在任意一点处的双侧极限恰好等于在该点处的函数值. 而函数 $y = \operatorname{sgn} x$ 不具备这样的性质, 它在点 $x = 0$ 处没有双侧极限. 这样, 我们看到一个函数 $y = f(x)$ 在一点 x_0 处的连续性可以用极限

[22]通常, 直线也当成曲线的一种特殊情况, 或者说平凡 (trivial) 情况.

$\lim\limits_{x\to x_0} f(x) = f(x_0)$ 来描述.

定义 1.5.1: 函数的连续性 (3 种常见表达方式)

假定函数 $y=f(x)$ 在区间 (a,b) 内有定义, 且点 $x_0 \in (a,b)$[a]. 若

1. $y=f(x)$ 在点 x_0 处有双侧极限, 且极限值等于函数值 $f(x_0)$, 即

$$\lim_{x\to x_0} f(x) = f(x_0);$$

2. 或者令 $\Delta x := x - x_0$, 则 $x \to x_0$ 等价于 $\Delta x \to 0$. 则上述极限式可重写[b]为

$$\lim_{\Delta x\to 0} f(x_0+\Delta x) = f(x_0);$$

3. (或者直接对 1. 使用 $\varepsilon-\delta$ 说法) 如果对于任意给定的 $\varepsilon>0$, 存在一个 $\delta=\delta(\varepsilon)>0$, 使得

$$|f(x)-f(x_0)|<\varepsilon, \quad \text{只要 } |x-x_0|<\delta;$$

则称 $y=f(x)$ 在点 x_0 处 **连续 (continuous)**. 若 $y=f(x)$ 在 (a,b) 的每一点处都连续, 则称它在 (a,b) 上连续. 这时, 也称 $y=f(x)$ 在 (a,b) 上是 **连续函数 (continuous function)**.

[a] 此时我们不能只考虑点 x_0 的空心邻域了, 否则 $f(x_0)$ 没有定义的话那就进行不下去了.

[b] 写作 $\lim\limits_{\Delta x\to 0} f(x_0-\Delta x) = f(x_0)$ 也是没有问题的, 但是我们一般习惯加上 Δx 而不是减去 Δx.

与函数极限的定义相比较, 我们发现的不同之处仅在于:

- 首先, 这里要求 $f(x)$ 在点 x_0 处有定义, 且以 $f(x_0)$ 作为极限值.
- 其次, 在上面的第 3 点中, 这里没有要求 $0<|x-x_0|$, 即没有要求 $0\neq|x-x_0|$. 这是因为, 当 $x-x_0=0$ 时, 不等式 $|f(x)-f(x_0)|=|f(x_0)-f(x_0)|=0<\varepsilon$ 自然成立.

例 1.5.1

1. 函数 $y=\sin x, y=\cos x, y=\mathrm{e}^x x^n$ (n 为正整数) 在 $\mathbf{R}$ 上是连续函数, 函数 $y=\sqrt{x}$ 在区间 $(0,+\infty)$ 上是连续函数.
2. 函数 $y=[x]$ 在每个整数点处都是不连续的, 而在其他点处都是连续的.
3. 狄利克雷函数任何一点处都不连续, 因为它在任何一点处连极限都没有. 见例题 1.4.10.

再次强调:

- $y=f(x)$ 在点 x_0 处连续意味着, 必须同时满足: $f(x)$ 在 x_0 处有定义, $\lim\limits_{x\to x_0} f(x)$ 存在, 并且极限值等于 $f(x_0)$.
- 函数的连续性是函数在某个点上的局部性质. 故当我们在讨论函数是不是连续的, 首先要明确是在哪个点上讨论其连续性.

- 有的时候, 如果我们简单地说 “这 (不) 是一个连续函数”. 此时, 我们讨论的是该函数在其自然定义域内所有点上的连续性. 说一个函数不是连续函数, 意味着至少在它的某个有定义的点处, 该函数不连续. 例如, 符号函数 $y=\operatorname{sgn}(x)$. 更具体地, 我们会聚焦在定义域的某个区间上, 讨论在这个区间上函数是不是连续函数.

- 对于闭区间的端点处的连续性, 我们默认的说都是左连续或者右连续. 见下面的定义.

定义 1.5.2: 右连续, 左连续

- 设函数 $y=f(x)$ 在以点 x_0 为左侧端点 (包含点 x_0) 的某个区间有定义. 若

$$\lim_{x\to x_0+0} f(x)=f(x_0),$$

  则称 $y=f(x)$ 在点 x_0 处 **右连续**. 上式极限等价于 $\lim\limits_{\Delta x\to 0+0} f\left(x_0+\Delta x\right)=f\left(x_0\right)$.

- 设函数 $y=f(x)$ 在以点 x_0 为右侧端点 (包含点 x_0) 的某个区间有定义. 若

$$\lim_{x\to x_0-0} f(x)=f(x_0),$$

  则称 $y=f(x)$ 在点 x_0 处 **左连续**. 上式极限等价于 $\lim\limits_{\Delta x\to 0-0} f\left(x_0+\Delta x\right)=f\left(x_0\right)$.

显然, $y=f(x)$ 在点 x_0 处连续的充要条件是它在点 x_0 处左连续且右连续. 此时不用额外要求他们各自的极限值相等了, 因此根据左右连续的定义他们肯定相等.

定义 1.5.3: 函数在闭区间上的连续性

假定函数 $y=f(x)$ 在区间 $[a,b]$ 上有定义. 若 $y=f(x)$ 在区间 (a,b) 上连续, 且在点 a 处右连续, 在点 b 处左连续, 则称 $y=f(x)$ 在 $[a,b]$ 上连续, 或称 $y=f(x)$ 在 $[a,b]$ 上是连续函数.

类似地, 我们也可以给出 $y=f(x)$ 在 $[a,b)$ 上连续, 或者 $y=f(x)$ 在 $(a,b]$ 上连续的定义.

例 1.5.2 函数 $y=x-[x]$ 在每个整数点处右连续, 但不左连续. 函数 $y=x-[x]$ 在区间 $\left[0,\dfrac{1}{2}\right]$ 上是连续函数, 但它在区间 $[0,1]$ 上不是连续函数, 因为在左侧的端点处是右连续而不是左连续.

我们把之前的内容总结如下.

结论 1.5.1: 连续性的等价表述

函数 $f(x)$ 在点 x_0 连续, 有如下等价的表述:

1. $\lim\limits_{x\to x_0} f(x) = f(x_0)$. 即必须同时满足 $f(x)$ 在 x_0 处有定义, $\lim\limits_{x\to x_0} f(x)$ 存在, 并且极限值等于 $f(x_0)$.
2. $\forall\varepsilon>0, \exists\delta>0$, 当 $|x-x_0|<\delta$ 时, 恒有 $|f(x)-f(x_0)|<\varepsilon$.
3. $\lim\limits_{x\to x_0+0} f(x) = \lim\limits_{x\to x_0-0} f(x) = f(x_0)$. 即必须同时满足: $f(x)$ 在 x_0 处有定义, 左右极限都存在, 并且极限值等于 $f(x_0)$.
4. 在点 x_0 处左连续且右连续.
5. $\lim\limits_{\Delta x\to 0} f(x_0+\Delta x) = f(x_0)$.

根据以上的定义, 在某点出极限不存在的有下面几种情况, 它们有时可能是交叉存在的 (其他分类的叫法, 见后面的小节 1.5.4):

- 该点处的极限不存在. (这有包含了几点情况, 见上一节的具体讨论.)
- 该点处的函数值不存在, 即函数在该点没有定义.
- 该点处的极限和函数值都存在, 但不相等.
- 在该点处左 (右) 连续, 但不右 (左) 连续.

根据函数极限的四则运算法则, 立即推出函数的四则运算保持函数的连续性, 即有下面的定理成立.

定理 1.5.1: 连续函数的和, 差, 积, 商的连续性

设函数 $y=f(x)$ 及 $y=g(x)$ 在点 x_0 附近有定义, 且在点 x_0 处连续, 则 $y=f(x)\pm g(x), y=f(x)g(x)$ 及 $y=\dfrac{f(x)}{g(x)}\,(g(x_0)\neq 0)$ 在点 x_0 处连续.

由这个定理立即推出下例中的结论.

例 1.5.3 正切函数 $y=\tan x=\dfrac{\sin x}{\cos x}$ 与余切函数 $y=\cot x=\dfrac{\cos x}{\sin x}$ 在各自的定义域内是连续函数.

1.5.2 复合函数的连续性

下面的定理告诉我们: 连续函数之间的复合仍是连续函数.

定理 1.5.2

设函数 $f:(a,b)\to(c,d)$ 在点 x_0 处连续, 而函数 $g:(c,d)\to\mathbf{R}$ 在点 $y_0=f(x_0)$ 处连续, 则复合函数 $g\circ f$ 在点 x_0 处是连续的, 即

$$\lim_{x\to x_0} g(f(x)) = g(f(x_0)).$$

证明. 因为函数 g 在点 $y_0=f(x_0)$ 处连续, 那么对于任意给定的 $\varepsilon>0$, 存在一个 $\delta_1>0$, 使得

$$|g(y)-g(y_0)|<\varepsilon, \quad 只要\ |y-y_0|<\delta_1.$$

因为函数 f 在点 x_0 处连续, 则对于这个 $\delta_1>0$, 则又存在一个 $\delta>0$, 使得[23]

$$|f(x)-f(x_0)|<\delta_1, \quad 只要\ |x-x_0|<\delta.$$

这样, 我们有

$$|g(f(x))-g(f(x_0))|<\varepsilon, \quad 只要\ |x-x_0|<\delta.$$

证毕. □

上述定理告诉我们: 常见的基本初等函数的复合函数, 如 $y=\sin x^n, y=\mathrm{e}^{\sin x}, y=\cos e^x, y=\cos(\sin x)$ 等, 在其自然定义域上都是连续函数. 定理 1.5.2 的证明可以稍做推广, 而写成下面的定理. 注意, 下面定理中的条件更加宽松了.

定理 1.5.3

假定函数 $g:(c,d)\to \boldsymbol{R}$ 在点 $y_0\in(c,d)$ 处连续, 又假定 $\lim\limits_{x\to a} f(x)=y_0$, 那么

$$\lim_{x\to a} g(f(x))=g(y_0).$$

上式可以直接写成下列形式:

$$\lim_{x\to a} g(f(x))=g\left(\lim_{x\to a} f(x)\right).$$

也就是说, 极限运算符号[a] lim 可以与连续函数的符号 g 交换次序.

[a]注意, 这里 $x\to a$ 可以换成 $+\infty, -\infty$ 或 ∞ 等等其他类型.

证明. 由于 $g(y)$ 在 y_0 处连续, 那么对于任意给定的 $\varepsilon>0$, 存在一个 $\delta_1>0$, 使得

$$|g(y)-g(y_0)|<\varepsilon, \quad 只要\ |y-y_0|<\delta_1.$$

又由于 $\lim\limits_{x\to a} f(x)=y_0$, 根据极限的定义, 对于这个 $\delta_1>0$, 则又存在一个 $\delta>0$, 使得

$$|f(x)-y_0|<\delta_1, \quad 只要\ 0<|x-a|<\delta.$$

因此, 当 $0<|x-a|<\delta$ 时, 有 (将 $f(x)=y$ 代入):

$$|g(f(x))-g(y_0)|<\varepsilon.$$

这表明 $\lim\limits_{x\to a} g(f(x))=g(y_0)$. 证毕. □

比较定理 1.5.2 与定理 1.5.3, 立即发现其差别在于: *定理 1.5.3 中没有假定 $f(x)$ 在点 a 处连续, 甚至没有假定它在点 a 处有定义.* 该定理对于求函数的极限有很大帮助.

例 1.5.4 $\lim\limits_{x\to\infty} \cos\left(1+\dfrac{1}{x}\right)^x=\cos\left(\lim\limits_{x\to\infty}\left(1+\dfrac{1}{x}\right)^x\right)=\cos e.$

[23]开始套娃.

例 1.5.5 $\lim\limits_{x\to+\infty}\sin(\sqrt{x+1}-\sqrt{x})=\sin 0=0$. 这里用到了 $\lim\limits_{x\to+\infty}(\sqrt{x+1}-\sqrt{x})=\lim\limits_{x\to+\infty}\dfrac{1}{\sqrt{x+1}+\sqrt{x}}=0$.

例 1.5.6 $\lim\limits_{x\to0}\dfrac{\ln(1+x)}{x}=1$. 事实上, 我们有 $\lim\limits_{x\to0}(1+x)^{\frac{1}{x}}=\lim\limits_{y\to\infty}\left(1+\dfrac{1}{y}\right)^y=\mathrm{e}$. 于是, 由对数函数的连续性 (见下一小节) 及上面的定理 1.5.3 立即有

$$\lim_{x\to0}\frac{\ln(1+x)}{x}=\lim_{x\to0}\ln\left((1+x)^{\frac{1}{x}}\right)=\ln\left(\lim_{x\to0}(1+x)^{\frac{1}{x}}\right)=\ln\mathrm{e}=1.$$

例 1.5.7 证明: 如果函数 $y=f(x)$ 在区间 (a,b) 上连续, 则函数 $y=|f(x)|$ 在 (a,b) 上也连续. 但是其逆命题不成立.

证明. 令 x_0 是 (a,b) 中的任意点. 我们需要证明 $\lim\limits_{x\to x_0}|f(x)|=|f(x_0)|$. 对于任意的 $\epsilon>0$, 由 $f(x)$ 在 x_0 处的连续性, 存在 $\delta>0$, 当 $|x-x_0|<\delta$ 时, $|f(x)-f(x_0)|<\epsilon$. 根据三角不等式,

$$||f(x)|-|f(x_0)||\le|f(x)-f(x_0)|.$$

因此, 当 $|f(x)-f(x_0)|<\epsilon$ 时, 也有

$$||f(x)|-|f(x_0)||<\epsilon.$$

这表明 $\lim\limits_{x\to x_0}|f(x)|=|f(x_0)|$, 即 $|f(x)|$ 在 x_0 处连续. 由于 x_0 是 (a,b) 中的任意点, 因此函数 $y=|f(x)|$ 在区间 (a,b) 上连续.

如果函数 $y=|f(x)|$ 在区间 (a,b) 上连续, 函数 $y=f(x)$ 是否也连续? 不成立. 反例可以说明这个问题. 设

$$f(x)=\begin{cases}1, & x\ge0,\\ -1, & x<0.\end{cases}$$

在这个例子中, 函数 $f(x)$ 在 $x=0$ 处不连续. 但 $|f(x)|$ 是恒等于 1 的, 因此 $|f(x)|$ 是在所有点都连续的. 证毕. □

1.5.3 反函数的连续性

我们先讨论单调函数的概念.

定义 1.5.4: 函数的单调性

设 $f : X \to Y$ 是一个函数. 如果 f 满足下列条件:

$$f(x_1) \leqslant f(x_2), \quad \text{只要 } x_1, x_2 \in X, x_1 < x_2,$$

则我们称 f 是**递增的**. 如果 f 满足更强的条件:

$$f(x_1) < f(x_2), \quad \text{只要 } x_1, x_2 \in X, x_1 < x_2,$$

则我们称 f 是**严格递增的**. 类似地, 可以定义函数 f 递减和严格递减的概念, 只要在上述定义中交换 $f(x_1)$ 与 $f(x_2)$ 的位置而其他叙述不改即可. 一个函数称为**单调** (**严格单调**) 的, 如果它是递增或递减 (严格递增或严格递减) 的.[a]

[a]有些教材中, "递增" 指的就是我们这里的严格递增, 递减也是类似. 注意不同教材中同一术语 (terminology) 的细微差别.

在基本初等函数中:

- 正弦函数和余弦函数在区间 $(-\infty, +\infty)$ 内显然不是单调函数, 但它们分别在区间 $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ 及 $(0, \pi)$ 上是严格单调的;
- 对数函数在区间 $(0, +\infty)$ 内是严格单调的;
- 指数函数在区间 $(-\infty, +\infty)$ 内是严格单调的;
- 常数函数是单调函数, 但不是严格单调的.

定理 1.5.4

设 $f : (a, b) \to (c, d)$ 是双射, 故其反函数 $f^{-1} : (c, d) \to (a, b)$ 存在并且也是双射. 如果 f 是严格单调的[a], 则

1. f^{-1} 在区间 (c, d) 上具有和 f 相同的单调性.
2. f 是区间 (a, b) 上的连续函数.
3. f^{-1} 是区间 (c, d) 上的连续函数.

[a]第二节讲函数时, 我们举过例子说明: 即使 f 是双射, 但它不一定单调. 因此这个单调性条件是有意义的.

证明. 不失一般性, 我们假设 f 是严格单调递增的.

1. 我们现在证明 f^{-1} 是严格单调递增的. 令 $y_1, y_2 \in (c, d)$ 且 $y_1 < y_2$. 因为 f^{-1} 是 f 的反函数, 意味着

$$f(f^{-1}(y_1)) = y_1 \quad \text{且} \quad f(f^{-1}(y_2)) = y_2.$$

由 $y_1 < y_2$ 知道

$$f(f^{-1}(y_1)) < f(f^{-1}(y_2)).$$

因为 f 是严格单调递增的, 所以 f 保持输入值的大小关系. 这意味着 $f^{-1}(y_1) < f^{-1}(y_2)$.

2. 我们现在证明 f 是区间 (a,b) 上的连续函数. 设 $x_0\in(a,b)$ 是任意给定的一点, 令 $y_0:=f(x_0)\in(c,d)$. 又设 ε 是任意给定的正数, 并且充分的小, 能使得

$$c<y_0-\varepsilon<y_0<y_0+\varepsilon<d.$$

由于 f 是严格递增的, 1 告诉我们 f^{-1} 也是严格递增的. 令

$$x_1:=f^{-1}\left(y_0-\varepsilon\right),\quad x_2:=f^{-1}\left(y_0+\varepsilon\right),$$

又因为 $x_0=f^{-1}\left(y_0\right)$, 则我们有 $x_1<x_0<x_2$. 显然, 这时由 f 的严格递增性有

$$f(x_0)-\varepsilon=y_0-\varepsilon=f(x_1)<f(x)<f(x_2)=y_0+\varepsilon=f(x_0)+\varepsilon,\qquad \text{只要 } x_1<x<x_2.$$

取 $\delta=\min\left\{x_2-x_0,x_0-x_1\right\}$, 则 $U_\delta\left(x_0\right)\subset\left(x_1,x_2\right)$, 从而有

$$|f(x)-f(x_0)|<\varepsilon,\qquad \text{只要 } |x-x_0|<\delta.$$

这表明, f 在点 x_0 处是连续的. 由于 x_0 是 (a,b) 内任意的一点, 故 f 在 (a,b) 上连续.

3. 因为 f^{-1} 也严格单调, 因而上述证明也完全适用于 f^{-1}, 故 f^{-1} 在 (c,d) 上也是连续的. 证毕.

□

从上述证明中可以看出, 若将定理 1.5.4 中的区间 (a,b) 与 (c,d) 同时改为闭区间或半开半闭区间, 结论同样成立.

下面我们应用定理 1.5.4 来证明 **基本初等函数在其定义域内的连续性.** 关于基本初等函数, 我们认为它们在中学数学课本中已经有定义, 并且承认已为大家所熟知的性质, 如正弦函数 $y=\sin x$ 在区间 $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ 上严格单调, 且将区间 $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ 映满区间 $[-1,1]$. 本教材中我们承认这些结论而不打算深究其理由.

- 基于正弦函数 $y=\sin x$ 在区间 $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ 上的上述性质, 由定理 4 即推出反正弦函数 $y=\arcsin x$ 在区间 $[-1,1]$ 上的连续性. 一般地, 我们有结论: 反三角函数 $y=\arcsin x, y=\arccos x, y=\arctan x$ 及 $y=\operatorname{arccot} x$ 在各自的定义域内是连续函数.
- 基于指数的单调性及其取值范围, 我们可以得到指数函数与对数函数的连续性. 对数函数 $y=\log_a x$ 与指数函数 $y=a^x$ $(a>0,a\neq1)$ 分别在区间 $(0,+\infty)$ 及 $(-\infty,+\infty)$ 上连续.
- 我们来证明一般幂函数 $y=x^\alpha$ 在区间 $(0,+\infty)$ 内是连续的. 事实上, $y=x^\alpha$ 可以写成

$$y=x^\alpha=\mathrm{e}^{\alpha\ln x}$$

即 $y=x^\alpha$ 可以看作对数函数与指数函数的复合函数. 这样, 由对数函数与指数函数的连续性即推出 $y=x^\alpha$ 的连续性.

总之, 我们证明了, 所有基本初等函数在其定义域内是连续的. 我们知道, 任意一个初等函数都是基本初等函数经过有限次四则运算及复合运算的结果, 而四则运算及复合运算保持函数的连续性不变. 因此, 每个初等函数在其定义域中任意一个区间上都是连续的[24].

例 1.5.8 求极限 $\lim\limits_{x\to+\infty}\left(1+\dfrac{1}{x}\right)^{x+\sqrt{x}}$.

[24]初等函数的定义域可能包含孤立点, 例如 $y=\sqrt{x^2(x-1)(x+1)}$ 的定义域为 $(-\infty,-1]\cup\{0\}\cup[1,+\infty)$, 其中 $x=0$ 就是一个孤立点. 根据现有函数连续性的定义, 我们不能在孤立点谈论函数的连续性. 故这里我们只断言: 初等函数在其定义域中的任意一个区间上是连续的, 这里的区间可以是开区间, 闭区间, 半开半闭区间或无穷区间.

证明. 我们有

$$\left(1+\frac{1}{x}\right)^{x+\sqrt{x}}=\mathrm{e}^{\ln\left[\left(1+\frac{1}{x}\right)^{x+\sqrt{x}}\right]}=\mathrm{e}^{(x+\sqrt{x})\ln\left(1+\frac{1}{x}\right)}=\mathrm{e}^{\left(1+\frac{1}{\sqrt{x}}\right)\cdot\ln\left(1+\frac{1}{x}\right)^{x}}.$$

利用对数函数的连续性, 有

$$\lim_{x\to+\infty}\ln\left(1+\frac{1}{x}\right)^{x}=\ln\left(\lim_{x\to+\infty}\left(1+\frac{1}{x}\right)^{x}\right)=\ln\mathrm{e}=1,$$

故

$$\lim_{x\to+\infty}\left(1+\frac{1}{\sqrt{x}}\right)\cdot\ln\left(1+\frac{1}{x}\right)^{x}=1.$$

再由 e^{x} 的连续性, 最终我们有

$$\lim_{x\to+\infty}\left(1+\frac{1}{x}\right)^{x+\sqrt{x}}=\lim_{x\to+\infty}\mathrm{e}^{\left(1+\frac{1}{\sqrt{x}}\right)\cdot\ln\left(1+\frac{1}{x}\right)^{x}}=\mathrm{e}^{\lim_{x\to+\infty}\left(1+\frac{1}{\sqrt{x}}\right)\cdot\ln\left(1+\frac{1}{x}\right)^{x}}=\mathrm{e}.$$

□

1.5.4　间断点的分类

定义 1.5.5: 间断点

函数 $y=f(x)$ 在一点 x_0 附近有定义但不连续, 则称 x_0 为其一个**间断点 (discontinuity)**. 具体细分为以下两种类型.

1. **第一类间断点**:
   (a) $\lim\limits_{x\to x_0+0}f(x)$ 与 $\lim\limits_{x\to x_0-0}f(x)$ 都存在, 但它们彼此不相等.
   (b) 它们相等, 但不等于函数值 $f(x_0)$. 此时人们可以通过修改在点 x_0 处的函数值而使得函数 $f(x)$ 在点 x_0 处连续, 故这种情况又称 x_0 为 **可去间断点**.
2. **第二类间断点**:
   (a) $\lim\limits_{x\to x_0+0}f(x)$ 与 $\lim\limits_{x\to x_0-0}f(x)$ 中至少有一个不存在.

下面是一些例子.

- 函数 $y=x-[x]$ 的间断点都是第一类间断点;
- 函数 $y=\operatorname{sgn}x^2$ 的间断点 $x=0$ 是第一类间断点, 并且是可去间断点;
- 函数
$$y=\begin{cases}\sin\dfrac{1}{x}, & x\neq 0\\ 0, & x=0\end{cases}$$
的间断点 $x=0$ 是第二类间断点.

例 1.5.9 指出函数 $f(x)=\begin{cases}\dfrac{\sin x}{x(1-x)}, & x\neq 0,1,\\ -1, & x=0,\\ 1, & x=1\end{cases}$ 的间断点及其类型, 若是可去间断点, 请修改在该点处的函数值, 使得 $f(x)$ 在该点处连续.

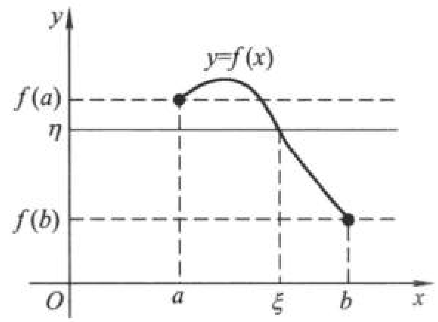


图 1.11: 中值定理的图示

证明. 显然, 当 $x \neq 0,1$ 时, 因为 $\dfrac{\sin x}{x(1-x)}$ 是初等函数, 所以 f 都是连续的, 没有间断点. 只需要考虑 $x=0,1$ 是否是间断点.

- 当 $x=0$. 考虑

$$\lim_{x\to 0}\frac{\sin x}{x(1-x)}=\lim_{x\to 0}\frac{\sin x}{x}\cdot\frac{1}{1-x}=1,\quad f(0)=-1,$$

  所以 $x=0$ 是可去间断点. 故修改点 $x=0$ 处的函数值, 定义 $f(0)=1$, 则 $f(x)$ 在点 $x=0$ 处连续.

- 当 $x=1$. 考虑

$$\lim_{x\to 1+0}\frac{\sin x}{x(1-x)}=\lim_{x\to 1+0}\frac{\sin x}{x}\cdot\frac{1}{1-x}=\sin(1)\cdot\lim_{x\to 1+0}\frac{1}{1-x}=-\infty,$$

  所以函数在 $x=1$ 处的右极限不存在. 同理可得到, $\lim\limits_{x\to 1-0}\dfrac{\sin x}{x(1-x)}=+\infty$. 故函数在 $x=1$ 处的左极限不存在. 于是 $x=1$ 是第二类间断点.

□

1.6 闭区间上连续函数的性质

闭区间上连续函数的基本性质是微分学及积分学中某些理论的基础, 其重要性在今后的讨论中将会逐渐显露出来.

1.6.1 关于连续函数的几个重要定理

现在我们列出四条定理. 它们在直观上是很容易接受的事实.

定理 1.6.1: 介值定理, 中值定理 (intermediate value theorem)

设 $f:[a,b]\to\mathbf{R}$ 是闭区间 $[a,b]$ 上的连续函数, 并且 $f(a)\neq f(b)$, 则

1. 对于任意一个严格[a]位于 $f(a)$ 和 $f(b)$ 之间的值 η, 都存在一点 $\xi\in(a,b)$, 使得
$$f(\xi)=\eta.$$
2. 将 1 换句说, 即函数 $f:[a,b]\to\mathbf{R}$ 可以取得到 $f(a)$ 和 $f(b)$ 之间的所有值. 自然也包括两个端点 $f(a),f(b)$ 本身.

[a]即 $f(b)<\eta<f(a)$ (或 $f(a)<\eta<f(b)$).

上述定理的证明在下一小节 1.6.2. 直观解释如下 (见图 1.11): 设想 x 自点 a 连续地向点 b 移动, 这时平面上的点 $(x,f(x))$ 在一条连续曲线上自点 $(a,f(a))$ 连续地向点 $(b,f(b))$ 移动. 因为这条曲线的端点分别处在直线 $y=\eta$ 的两侧, 故这条曲线必与该直线在某处相交, 而交点的 x 坐标就是所要找的 ξ. 不连续的函数, 如符号函数, 显然没有定理 1.6.1 所说的性质.

我们很容易得到下面这一常用的定理. 它是定理 1.6.1 的一种特殊情况.

定理 1.6.2: 零点存在定理

若函数 $f(x)$ 在闭区间 $[a,b]$ 上连续, 且 $f(a)f(b)<0$, 则存在一点 $\xi\in(a,b)$, 使得 $f(\xi)=0$.

例 1.6.1 证明: 方程 $x-\cos x=0$ 在区间 $\left(0,\dfrac{\pi}{2}\right)$ 中有唯一解.

证明. 证明分两步.

- 我们首先证明解的存在性. 设 $f(x)=x-\cos x$, 则 $f(x)$ 在闭区间 $\left[0,\dfrac{\pi}{2}\right]$ 上连续, 并且
$$f(0)=-1,\quad f\left(\frac{\pi}{2}\right)=\frac{\pi}{2}.$$
利用零点存在定理 1.6.2, 存在一点 $\xi\in\left(0,\dfrac{\pi}{2}\right)$, 使得 $f(\xi)=\eta$, 即 ξ 是方程 $x-\cos x=0$ 在 $\left(0,\dfrac{\pi}{2}\right)$ 中的解.

- 我们其次证明解的唯一性. 假设还有一个 $\xi'\in\left(0,\dfrac{\pi}{2}\right)$ 是方程 $x-\cos x=0$ 的解. 不妨设 $\xi>\xi'$. 因为 $\xi-\cos\xi=0$, $\xi'-\cos\xi'=0$, 所以
$$\xi-\cos\xi=\xi'-\cos\xi',$$
即
$$\xi-\xi'=\cos\xi-\cos\xi'.$$
上式左端大于零, 而 $\cos x$ 在 $\left[0,\dfrac{\pi}{2}\right)$ 上严格递减, 从而上式右端小于零, 得到相互矛盾的结果, 所以假设不成立, 因此 ξ 是方程 $x-\cos x=0$ 在 $\left(0,\dfrac{\pi}{2}\right)$ 中的唯一解.

$\square$

定理 1.6.3: 最大值与最小值定理, 最值定理, 极值定理 (extreme value theorem)

设 $f:[a,b]\to\mathbf{R}$ 是闭区间 $[a,b]$ 上的连续函数, 则它的函数值可以取得最大值与最小值, 即存在 $x_1\in[a,b]$ 及 $x_2\in[a,b]$, 使得

$$f\left(x_1\right)\leqslant f(x)\leqslant f\left(x_2\right),\quad \forall x\in[a,b].$$

证明. 该定理的结论虽然很容易接受, 但证明确实严重超纲. 需要的其他前置知识非常多, 可以不用管. □

由上面马上可以得到下面.

定理 1.6.4: 有界性定理

闭区间上的连续函数都是有界函数.

所谓的函数 $y=f(x)$ 在闭区间 $[a,b]$ 上有界, 是指存在两个常数 M 与 N, 使得 $N\leqslant f(x)\leqslant M$, $\forall x\in[a,b]$.

注意 函数 $y=f(x)$ 在闭区间 $[a,b]$ 上可以取到最大值与最小值. $\Longrightarrow$ 函数 $y=f(x)$ 在闭区间 $[a,b]$ 上有界. 反之, 不一定成立.

作为定理 1.6.3 的反例:

- 开区间上的连续函数可能是无界的, 如函数 $f(x)=\dfrac{1}{x}$, $x\in(0,1)$.
- 开区间上的连续函数即使是有界的, 其函数值也未必有最大值或最小值, 如函数 $f:(0,1)\to\mathbf{R}, x\mapsto x^2$ 在开区间 $(0,1)$ 上连续, 其值域为 $(0,1)$, 但这个集合中无最大值, 也无最小值.
- 有时函数在闭区间上有间断点. 如,

$$y=f(x)=\begin{cases}-x+1, & 0\leqslant x<1,\\ 1, & x=1,\\ -x+3, & 1<x\leqslant 2,\end{cases}$$

  在闭区间 $[0,2]$ 上有间断点 $x=1$, 这函数 $f(x)$ 在闭区间 $[0,2]$ 上虽然有界, 但是既无最大值又无最小值. 因为它不满足连续性.

同时利用中值定理 1.6.1 与最值定理 1.6.3 可以推出, 闭区间上连续函数的值域是一个闭区间. 如下所述.

推论 1.6.1

在闭区间 $[a,b]$ 上连续的函数 $f(x)$ 的值域为闭区间 $[m,M]$, 其中 m 与 M 依次为 $f(x)$ 在 $[a,b]$ 上的最小值与最大值.

下面我们举例说明连续函数的这些定理的用途

例 1.6.2 证明: 实系数多项式

$$P(x)=a_0x^5+a_1x^4+\cdots+a_5\quad(a_0>0)$$

至少有一个实根.

证明. 我们将 $P(x)$ 写成

$$P(x) = x^5\left(a_0 + \frac{a_1}{x} + \cdots + \frac{a_5}{x^5}\right), \quad x \neq 0.$$

当 $x \to \infty$ 时,

$$a_0 + \frac{a_1}{x} + \cdots + \frac{a_5}{x^5} \to a_0 > 0$$

因此, 当 $|x|$ 充分大时, $a_0 + \frac{a_1}{x} + \cdots + \frac{a_5}{x^5} > 0$, 这时 $P(x)$ 与 x^5 (因此 x 也是) 有相同的符号. 由此知

$$P(x) \to +\infty\ (x \to +\infty), \quad P(x) \to -\infty\ (x \to -\infty).$$

这样, 存在一点 b, 使得 $P(b) > 0$; 并且存在一点 a, 使得 $P(a) < 0$. 多项式是初等函数, 因而在 $\mathbf{R}$ 上连续. 根据介值定理, 在 a 与 b 之间存在一点 c, 使得 $P(c) = 0$. 这样, c 就是多项式 $P(x)$ 的实根. 证毕. □

从上述证明立即看出, 任意奇数次实系数多项式至少有一个实根. 下面的命题使用了中值定理来证明. 其中, 连续的假定是必要的. 通常单射的性质并不一定等到单调性, 考虑函数 $f(x)$ 定义为: 当 $x \in \mathbb{Q}$ 时, $f(x) = x$; 当 $x \notin \mathbb{Q}$ 时, $f(x) = -x$.

命题 1.6.1

设函数 $f(x)$ 在区间 (a, b) 上连续. 若 $f(x)$ 在 (a, b) 上是一一映射 (即单射), 即 $x' \neq x'' \Longrightarrow f(x') \neq f(x'')$, 则 $f(x)$ 在 (a, b) 上是严格单调的.

证明. 用反证法. 假设 $f(x)$ 在 (a, b) 上不是严格单调的, 即存在 $x, y, z \in (a, b)$, $x < y < z$, 使得存在以下两种情况之一:

$$f(x) \leq f(y) \text{ 且 } f(y) \geq f(z),$$

(这意味着在 x 到 z 之间, 函数在 y 处达到一个较大值, 因此不能在整个区间上保持递增性.) 或

$$f(x) \geq f(y) \text{ 且 } f(y) \leq f(z).$$

(这表明在 x 到 z 之间, 函数在 y 处达到一个最小值, 因此也不能在整个区间上保持递减性.) 我们先考虑第一种情况 $f(x) \leq f(y)$ 且 $f(y) \geq f(z)$.

如果 $f(x) = f(y)$, 或 $f(y) = f(z)$, 或 $f(x) = f(z)$, 那么这与 f 是一一映射相矛盾. 因此, $f(x), f(y), f(z)$ 这三个值必然互不相等. 所以, 我们有 $f(x) < f(y)$ 且 $f(y) > f(z)$.

- 若 $f(x) < f(z)$, 即
$$f(x) < f(z) < f(y).$$
由于 f 在 (a, b) 上连续, 可以应用中值定理. 因此, 存在 $c \in (x, y)$ 使得 $f(c) = f(z)$. 由于 $z \notin (x, y)$, 所以 $c \neq z$. 因此, f 不是一一映射, 这与假设矛盾.

- 若 $f(x) > f(z)$, 即
$$f(z) < f(x) < f(y).$$
同样, 可以应用中值定理. 因此, 存在 $c \in (y, z)$ 使得 $f(c) = f(x)$. 因此, f 不是一一映射, 这与假设矛盾.

如果我们考虑第二种情况 $f(x) \geq f(y)$ 且 $f(y) \leq f(z)$, 我们也会得到类似的矛盾. □

上个命题告诉我们: 两个区间之间连续的一一满射, 作为函数一定是严格单调的. 因此, 由上一节中的定理 1.5.4, 我们立刻推出下列推论:

推论 1.6.2

若 $f:(a,b)\to(c,d)$ 是区间 (a,b) 上的连续函数, 并且其反函数 f^{-1}: $(c,d)\to(a,b)$ 存在, 那么 f^{-1} 在区间 (c,d) 上是也连续的.

这就是说, 如果一个在区间 (a,b) 上的连续函数有反函数, 则其反函数也连续.

1.6.2 中值定理的证明

最大 (小) 值, 上 (下) 界, 上 (下) 限

定义 1.6.1: 最大值, 最小值

对于任一实数集合 $A\subseteq\mathbf{R}$, 如果实数 a 满足: $a\in A$ 且对于任意的 $x\in A$, 都有 $x\leq a$ (或者 $x\geq a$), 则称 a 为集合 A 的**最大值** (或**最小值**), 记作 $a=\max A$ (或 $a=\min A$).

例如, 集合 $[0,1)$ 的最小值是 0, 但它没有最大值, 1 不是这个集合的最大值, 但它是一个可以取代最大值的数. 这样的数被称为上限. 上限的严格定义将在后文给出.

定义 1.6.2: 上界, 下界

- 如果实数集合 A 的所有元素都小于等于 (或大于等于) 某个实数 a, 则 a 称为集合 A 的 **上界** (或**下界**), 即
$$a \text{ 是 } A \text{ 的上界} \iff \forall x\in A, x\leq a, \tag{1.9}$$
和,
$$a \text{ 是 } A \text{ 的下界} \iff \forall x\in A, x\geq a. \tag{1.10}$$
- 如果实数集合 A 存在上界 (下界), 则称 A 为**上有界** (或**下有界**) 的集合. 如果一个集合同时存在上界和下界, 则称其为**有界**集合.

例 1.6.3

- 1 或比 1 大的数是集合 $[0,1)$ 的上界. 因此, 集合 $[0,1)$ 的所有上界为 $[1,\infty)$. 同时, 集合 $[0,1)$ 的所有下界为 $(-\infty,0]$.
- 集合 $\left\{1,\frac{1}{2},\frac{1}{3},\cdots,\frac{1}{n},\cdots\right\}$ 的所有上界为 $[1,\infty)$, 所有下界为 $(-\infty,0]$.
- 集合 $\{x\mid\sin x=0\}$ 既没有上界也没有下界.

定义 1.6.3: 上限 (上确界, 最小上界), 下限 (下确界, 最大下界)

- 对于实数集合 A, 如果实数 a 是 A 的上界, 并且没有比 a 更小的数是 A 的上界, 即 a 是 A 的最小的上界时, a 称为 A 的 **上限 (supremum)**, 记作 $a = \sup A$. 该定义可以表示为:

$$\sup A = \min\{a \in \mathbf{R} :\ a \text{ 是 } A \text{ 的上界}\}. \tag{1.11}$$

或者

$$a = \sup A \iff (\forall x \in A, x \leq a) \wedge (\forall y < a, \exists x \in A, x > y). \tag{1.12}$$

其中, 陈述 $(\forall y < a, \exists x \in A, x > y)$ 表示任何比 a 小的数 y 都不是 A 的上界.

- 同样地, 对于实数集合 A, 如果实数 a 是 A 的下界, 并且没有比 a 更大的数是 A 的下界, 即 a 是 A 的最大下界时, a 称为 A 的 **下限 (infimum)**, 记作 $a = \inf A$. 该定义可以表示为:

$$\inf A = \max\{a \in \mathbf{R} :\ a \text{ 是 } A \text{ 的下界}\}.$$

或者

$$a = \inf A \iff (\forall x \in A, x \geq a) \wedge (\forall y > a, \exists x \in A, x < y).$$

其中, 陈述 $(\forall y > a, \exists x \in A, x < y)$ 表示任何比 a 大的数 y 都不是 A 的下界.

命题 1.6.2

对于非空的实数集合 A, 如果 $\sup A = a < \infty$ (或者 $\inf A = a > -\infty$), 那么存在一个数列 $\{a_n\}$, 其中 $a_n \in A$, 并且该数列收敛于 a.

证明. 仅证明 $\sup A = a < \infty$ 的情况. 对于任意自然数 n, 有 $a - \dfrac{1}{n} < a$, 故而 $a - \dfrac{1}{n}$ 不是 A 的上界. 因此存在 A 中的元素 x 满足 $a - \dfrac{1}{n} < x$. 任取一个这样子的 x, 记为 a_n. 另一方面, a 也是 A 的上界. 因此, 有以下不等式成立:

$$a - \frac{1}{n} < a_n \leq a \quad (n = 1, 2, \cdots).$$

于是, 夹逼定理告诉我们 $\lim\limits_{n\to\infty} a_n = a$ 成立. □

实数的连续性公理 (再讨论)

公理 (axiom) 是没有经过证明, 但被当作不证自明的一个命题. 因此, 其真实性被视为是理所当然的, 且被当做演绎及推论其他 (理论相关) 事实的起点. 比如, 通过两点, 有且仅有一条直线. 这是几何公理[25]之一. 第一节课中, 我们提到了实数的完备性. 那时我们说: 实数域具有完备性, 即在实数域中, 任意一个单调且有界

[25]五条几何公理:

1. 过相异两点, 能作且只能作一条直线 (直线公理).
2. 一条有限线段可以继续延长.
3. 以任意点为心及任意的距离可以画圆 (圆公理).
4. 凡直角都彼此相等 (角公理).
5. 同平面内一条直线和另外两条直线相交, 若在某一侧的两个内角和小于二直角的和, 则这二直线经无限延长后在这一侧相交.

的数列一定有极限存在, 并该极限值是实数. 其实, 实数的完备性又叫做连续性. 通常我们是从下面这个非常理所应当又浅显的事实从发, 把它作为公理. 它是我们讨论一切实数的性质的出发点.

公理 1.6.1: 实数的连续性公理

1. 上有界的实数集合有上限.
2. 下有界的实数集合有下限.

实数的连续性公理中的 1 和 2 是等价的命题, 因为对于实数集合 A, 如果 A 上有界, 则 $\{-x \mid x \in A\}$ 下有界, 且 $a = \sup A \Longleftrightarrow -a = \inf\{-x \mid x \in A\}$.

定理 1.6.5: 实数的连续性公理的其他等价描述

实数的连续性公理与以下不同命题相等价:

1. 任意的有界的数列, 一定有收敛的子数列.
2. 柯西数列 (Cauchy sequence)[a]是收敛的.
3. 任意一个单调且有界的数列一定有极限存在, 并该极限值是实数.

[a]对于任意给定的正实数 $\varepsilon > 0$, 总存在一个正整数 N, 使得对于所有的自然数 $m, n > N$, 都有 $|x_m - x_n| < \varepsilon$, 则该数列称为柯西数列 (Cauchy sequence).

如果承认实数的连续性公理, 则根据上述定理, 可以保证有界的单调数列和柯西数列的极限值存在. 实际上, 关于实数的极限讨论, 正是基于实数的连续性公理. 此外, 上述定理中的每个命题有时也被称为实数的连续性公理. 不同的教科书有不同的出发点, 需要用到哪个就说哪个.

中值定理的证明

定理 1.6.6: 介值定理, 中值定理 (intermediate value theorem)

设 $f : [a, b] \to \mathbf{R}$ 是闭区间 $[a, b]$ 上的连续函数, 并且 $f(a) \neq f(b)$, 则

1. 对于任意一个严格[a]位于 $f(a)$ 和 $f(b)$ 之间的值 η, 都存在一点 $\xi \in (a, b)$, 使得
$$f(\xi) = \eta.$$
2. 将 1 换句说, 即函数 f 可以取得到 $f(a)$ 和 $f(b)$ 之间的所有值. 自然也包括两个端点 $f(a), f(b)$ 本身.

[a]即 $f(b) < \eta < f(a)$ (或 $f(a) < \eta < f(b)$).

证明. 只需要证明 1. 不失一般性, 假设 $f(a) < \eta < f(b)$. 定义实数集合
$$A := \{x \in [a, b] \mid f(x) < \eta\}.$$
由于 $f(a) < \eta$, 所以 $a \in A$, 因此 A 是非空集合. 根据定义, A 是上有界的. 因此, 根据 **实数的连续性公理 1.6.1**, 集合 A 存在上限, 记作 $\xi := \sup A$. 由于 b 是 A 的上界, 因此最小上界 $\xi \leq b$. 又由于 ξ 是 A 的上界, 且 $a \in A$, 所以有 $a \leq \xi$. 因此有 $a \leq \xi \leq b$.

根据命题 1.6.2, 存在一个数列 $\{x_n\}$, $x_n \in A$, 并且该数列收敛于 ξ. 由于 f 在点 ξ 处连续, 由定理 1.4.5, 故有

$$\lim_{n\to\infty} f(x_n) = f(\xi).$$

并且因为 $f(x_n) < \eta$ 对所有 n 都成立, 根据序列极限的保序性, 我们得到 $f(\xi) \le \eta$. 同时, 若取序列 $x_n' = \xi + \dfrac{1}{n}$, 那么该数列收敛于 ξ. 又由定理 1.4.5 , 有

$$\lim_{n\to\infty} f\left(\xi + \frac{1}{n}\right) = f(\xi).$$

因为 $\xi + \dfrac{1}{n} \notin A$, $f\left(\xi + \dfrac{1}{n}\right) \ge \eta$ $(n = 1, 2, \cdots)$. 根据序列极限的保序性, 得到 $f(\xi) \ge \eta$. 因此, $f(\xi) = \eta$ 成立.

最后, $a \le \xi \le b$ 中等号不能成立, 否则 $f(\xi) = f(a)$ 或者 $f(\xi) = f(b)$, 这与条件相矛盾. □

第二章　微积分的基本概念

本章的主要内容包括: 导数的概念与计算, 微分的概念, 不定积分与定积分, 微积分基本定理 (牛顿 - 莱布尼茨公式). 它们是微积分的核心内容.

2.1　导数

2.1.1　导数的起源

导数的概念最初来自力学上关于瞬时速度的计算以及几何学上关于切线斜率的计算. 我们先看两个典型的例子.

物体沿直线运动的瞬时速度

假设一个物体沿着某条直线做运动. 在该直线上取定坐标系 (将物体运动的起点作为原点) 后, 物体的运动规律可以由函数 $s = s(t)$ 表示, 其中 $s(t)$ 表示 t 时刻物体的位置, 即物体在时间 t 内的位移. 我们的问题是: 若物体的运动规律是已知的, 应该如何计算它在某一 t_0 时刻的瞬时速度 *(velocity)* 呢? 当物体做匀速运动时, 它在任一时刻的速度很容易求出: 用时间去除位移即可. 当物体不是做匀速度运动时, 问题变得困难起来. 它的瞬时速度是指什么? 这就需要有新的观念来处理这个问题.

我们考虑自变量 t 在点 t_0 处的一个改变量 Δt (这里 Δ 是大写希腊字母, 读作 delta), $\Delta t \neq 0$. Δt 可正可负, 通常称为*自变量 t 的增量*. 这样就得到 t_0 时刻附近的另一时刻 $t_0 + \Delta t$, 而物体从 t_0 时刻到 $t_0 + \Delta t$ 时刻的位移应为 $s(t_0 + \Delta t) - s(t_0)$, 它是位移的改变量, 称为*函数或因变量 s 的增量*, 记作 Δs. 那么, 在这段时间内, 物体的平均速度为

$$\bar{v}_{\Delta t} := \frac{\Delta s}{\Delta t} = \frac{s(t_0 + \Delta t) - s(t_0)}{\Delta t}.$$

一般来说, 时间段 Δt 越小, 平均速度 $\bar{v}_{\Delta t}$ 也就越接近于物体在 t_0 时刻的瞬时速度. 因此, 我们可以认为 $\Delta t \to 0$ 时平均速度 $\bar{v}_{\Delta t}$ 的极限就是 t_0 时刻的瞬时速度, 即认为物体在 t_0 时刻的瞬时速度等于

$$v(t_0) := \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \lim_{\Delta t \to 0} \frac{s(t_0 + \Delta t) - s(t_0)}{\Delta t}.$$

例 2.1.1　现在我们以自由落体运动为例. 已知下落物体的运动规律为

$$s(t) = \frac{1}{2} g t^2.$$

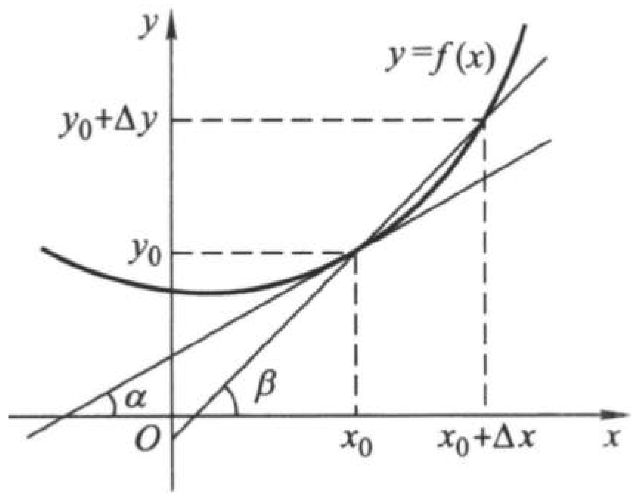


图 2.1: 曲线在一点处的切线

其中 g 是重力加速度. 根据上面的说法, 下落物体在 t 时刻的瞬时速度为

$$\begin{aligned} v(t) &= \lim_{\Delta t \to 0} \frac{\frac{1}{2}g(t+\Delta t)^2 - \frac{1}{2}gt^2}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{\frac{1}{2}g(t^2 + 2t\Delta t + (\Delta t)^2) - \frac{1}{2}gt^2}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{gt\Delta t + \frac{1}{2}g(\Delta t)^2}{\Delta t} \\ &= \lim_{\Delta t \to 0}(gt + \frac{1}{2}g\Delta t) \\ &= gt. \end{aligned}$$

用同样的办法可得到*瞬时加速度*的概念. 将某个时间段 Δt 内 (瞬时) 速度的改变量除以 Δt, 得到平均加速度, 再令 $\Delta t \to 0$, 由平均加速度取极限即得到在 t 时刻的瞬时加速度:

$$a(t) := \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \lim_{\Delta t \to 0} \frac{v(t+\Delta t) - v(t)}{\Delta t}.$$

按照这个公式计算, 自由落体运动在 t 时刻的瞬时加速度为常数 g:

$$a(t) = \lim_{\Delta t \to 0} \frac{g(t+\Delta t) - gt}{\Delta t} = g.$$

曲线在一点处的切线

我们来考虑任一曲线在一点处的切线问题. 当曲线是直线或圆周时, 这个几何问题的解答是明显的. 当曲线是一条任意的曲线时, 这个几何问题的解答就不那么简单了. 有趣的是, *这个几何问题的处理办法却与上述力学问题的处理办法完全相似, 并且在数学表达形式上完全一致.*

我们考虑一个连续函数 $y = f(x)$, 它定义在区间 (a, b) 上. 这时, 它的图形在 Oxy 平面上是一条连续曲线. 设想我们要研究这条曲线在一点 (x_0, y_0) 处的切线, 其中 $y_0 = f(x_0)$. 我们借助极限的方法, 将此切线看作过点 (x_0, y_0) 的一系列割线的极限位置 (见图 2.1).

现在我们考虑该切线的斜率. 在这条曲线上取另外一点 (x, y), 其中 $y = f(x)$. 这时, x 的增量为 $\Delta x := x - x_0$, y 的增量为 $\Delta y := y - y_0 = f(x_0 + \Delta x) - f(x_0)$. 过点 $(x_0, y_0), (x, y)$ 的直线是这条曲线的一条割线,

而该割线的斜率为

$$\tan\beta=\frac{\Delta y}{\Delta x}=\frac{f\left(x_0+\Delta x\right)-f\left(x_0\right)}{\Delta x},$$

其中 β 是该割线与 x 轴正向的夹角. 显然, β 依赖于 Δx, 故记为 $\beta_{\Delta x}$. 当 Δx 趋向于零时, 假如割线有一个极限位置, 那么在这个极限位置上的直线就是切线. 设切线与 x 轴正向的夹角为 α, 就有

$$\tan\alpha=\lim_{\Delta x\to 0}\tan\beta_{\Delta x}=\lim_{\Delta x\to 0}\frac{f\left(x_0+\Delta x\right)-f\left(x_0\right)}{\Delta x}.$$

也就是说, 我们认为切线的斜率是割线斜率的极限. 这样的看法使我们得以将求切线的斜率归结为**求函数的增量与自变量的增量之比的极限.** 在形式上, 这个极限与求瞬时速度时遇到的极限完全相同.

2.1.2　导数的定义

我们还可以举出许多其他例子, 要求人们**根据一个已知的函数关系** $y=f(x)$**, 去计算函数的增量** $\Delta y=f(x+\Delta x)-f(x)$ **与自变量的增量** Δx **之比的极限**. 如果这个极限存在, 则称之为导数或微商. 导数的严格的定义如下:

定义 2.1.1: 导数

设函数 $y=f(x)$ 在一个开区间 (a,b) 内有定义. 对于给定的一点 $x_0\in(a,b)$, 考虑一个 **自变量的增量** $\Delta x\neq 0$ (但可正可负), 并且使得 $x=x_0+\Delta x\in(a,b)$. 定义相应于 Δx 的 **函数的增量** (或者 **因变量的增量**) 为 $\Delta y:=f\left(x_0+\Delta x\right)-f\left(x_0\right)$. 若极限

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\frac{f\left(x_0+\Delta x\right)-f\left(x_0\right)}{\Delta x}=\lim_{x\to x_0}\frac{f\left(x\right)-f\left(x_0\right)}{x-x_0}$$

存在, 则称这个函数在点 x_0 处 **可导**, 并称该极限值为这个函数在点 x_0 处的 **导数 (derivative)** 或**微商**, 记作 $f'\left(x_0\right), y'|_{x=x_0}, \left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=x_0}, \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x=x_0}$ 或 $\left.\frac{\mathrm{d}}{\mathrm{d}x}f(x)\right|_{x=x_0}$.

根据前面的讨论可以看出:

- 函数 $y=f(x)$ 在一点处的导数, 从数量的角度去看, 就是因变量对自变量的变化率. 因变量增量与自变量增量之比 $\frac{\Delta y}{\Delta x}$ 是因变量 y 在以 x_0 和 $x_0+\Delta x$ 为端点的区间上的平均变化率, 而导数 $f'(x_0)$ 则是因变量 y 在点 x_0 处的瞬时变化率. 把下面的公式牢记于心则万变不离其宗:

$$\lim\ \text{平均变化率}=\ \text{瞬时变化率}=\ \text{导数}.$$

- 函数 $y=f(x)$ 在一点处的导数, 从图形上去看, 则是曲线 $y=f(x)$ 在相应点处切线的斜率. 函数图像的切线有一种例外情况, 若曲线在某点处有垂直于 x 轴的切线, 等价于 $\lim\limits_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=+\infty$ 或 $-\infty$.
- 在定义中, 点 $x_0\in(a,b)$ 是一个固定的值. 自变量的增量是以固定点 x_0 为起点, 上下变化的, 并且要求 $x=x_0+\Delta x\in(a,b)$. 只要考虑的增量 Δx 充分接近于 0, 这一要求是容易实现的. 注意增量本身 Δx 成了极限式子中真正的变量.
- 将 x_0 视作常量后, 函数的增量 $\Delta y=\Delta y(\Delta x):=f\left(x_0+\Delta x\right)-f\left(x_0\right)$ 就是一个关于 Δx 的函数. 所以从形式上看, 我们所关心的极限

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}h(\Delta x),$$

其中, 函数 $h(\Delta x)$,

$$\Delta x \mapsto h(\Delta x):=\frac{f\left(x_0+\Delta x\right)-f\left(x_0\right)}{\Delta x},$$

具有一个特定的结构. 简而言之, 我们用函数 f 和某个点 x_0 定义了一个新的 (有特定构造的) 函数 $h(\Delta x)$, 其中 Δx 是变量, 然后求极限 $\lim\limits_{\Delta x \to 0} h(\Delta x)$, 该极限值如果存在的话就是函数 f 在点 x_0 处的导数 $f'\left(x_0\right)$. 在求极限时, 极限变量的符号是 Δx 还是 x, y 或者其他任何符号都无所谓.

- 应当指出, 并非所有函数在其定义域内都是可导的, 即极限 $\lim\limits_{\Delta x \to 0} \dfrac{\Delta y}{\Delta x}$ 未必总存在.

我们要指出, 一般导数定义中的极限是双侧极限. 类似地, 我们可以定义单侧导数.

定义 2.1.2: 右导数, 左导数

- 若单侧极限
$$\lim_{\Delta x \to 0+0} \frac{f\left(x_0+\Delta x\right)-f\left(x_0\right)}{\Delta x}$$
存在, 则称之为 $f(x)$ 在点 x_0 处的右导数, 记为 $f'\left(x_0+0\right)$.

- 类似地, 若单侧极限
$$\lim_{\Delta x \to 0-0} \frac{f\left(x_0+\Delta x\right)-f\left(x_0\right)}{\Delta x}$$
存在, 则称之为 $f(x)$ 在点 x_0 处的左导数, 记为 $f'\left(x_0-0\right)$.

显然, 函数在一点处可导的充要条件是其左, 右导数都存在且相等. 左导数和右导数统称为单侧导数. 如果函数 $f(x)$ 在开区间 (a, b) 内可导, 且 $f'(a+0)$ 及 $f'(b-0)$ 都存在, 那么就说 $f(x)$ 在闭区间 $[a, b]$ 上可导.

例 2.1.2 我们考虑函数 $y=|x|$. 在点 $x=0$ 处, 该函数的右导数为

$$\lim_{\Delta x \to 0+0} \frac{|\Delta x|-0}{\Delta x}=\lim_{\Delta x \to 0+0} \frac{\Delta x}{\Delta x}=1.$$

而左导数为

$$\lim_{\Delta x \to 0-0} \frac{|\Delta x|-0}{\Delta x}=\lim_{\Delta x \to 0-0} \frac{-\Delta x}{\Delta x}=-1.$$

因此, 该函数在点 $x=0$ 处是不可导的.

后面我们还将给出一个函数在某点处左导数或右导数不存在的例子.

例 2.1.3 请指出下列极限是什么函数在哪一点处的导数:

1. $\lim\limits_{\Delta x \to 0} \dfrac{(4+\Delta x)^2-16}{\Delta x}$.

   设 $f(x)=x^2$, 记 $x_0=4$, 则

$$\lim_{\Delta x \to 0} \frac{(4+\Delta x)^2-16}{\Delta x}=\lim_{\Delta x \to 0} \frac{f\left(x_0+\Delta x\right)-f\left(x_0\right)}{\Delta x}.$$

   可见, 此极限是 $f(x)=x^2$ 在点 $x_0=4$ 处的导数.

2. $\lim\limits_{h \to 0} \dfrac{2(5+h)^3 - 2 \times 5^3}{h}$.

   将 h 换成 Δx, 就可看出此极限是 $f(x) = 2x^3$ 在点 $x_0 = 5$ 处的导数. 注意在求极限时, 极限变量的符号是 Δx 还是其他任何符号都无所谓.

3. $\lim\limits_{x \to 3} \dfrac{\frac{2}{x} - \frac{2}{3}}{x - 3}$.

   注意极限已经不是 $\to 0$. 根据导数定义的等价形式 $\lim\limits_{x \to x_0} \dfrac{f(x) - f(x_0)}{x - x_0}$, 一眼便看出此极限是 $f(x) = \dfrac{2}{x}$ 在点 $x_0 = 3$ 处的导数.

4. $\lim\limits_{x \to t} \dfrac{\frac{2}{x} - \frac{2}{t}}{x - t}$.

   同上, 此极限是 $f(x) = \dfrac{2}{x}$ 在点 $x_0 = t$ 处的导数.

例 2.1.4 设函数 $y = f(x)$ 在点 x_0 处可导且 $f'(x_0) = 1$, 求极限 $\lim\limits_{\Delta x \to 0} \dfrac{f(x_0 + \Delta x) - f(x_0 - 2\Delta x)}{\Delta x}$.

证明. 做如下变形:

$$\frac{f(x_0 + \Delta x) - f(x_0 - 2\Delta x)}{\Delta x} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} - (-2) \cdot \frac{f(x_0 - 2\Delta x) - f(x_0)}{-2\Delta x}.$$

因为 $f'(x_0) = 1$, 而

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x},$$

所以利用极限的性质有

$$\lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0 - 2\Delta x)}{\Delta x} = f'(x_0) - (-2) f'(x_0) = 3.$$

$\square$

这里我们用到一个简单的事实: 考虑一个一般的函数 h 和一个非零常数 c. 因为 $\Delta x \to 0$ 等价于 $c \cdot \Delta x \to 0$, 则有

$$\lim_{\Delta x \to 0} h(\Delta x) = \lim_{\Delta x \to 0} h(c \cdot \Delta x) = \lim_{c \cdot \Delta x \to 0} h(c \cdot \Delta x).$$

这是一个常用的技巧. 所以导数的定义式还可以写成 (c 是任意的非零常数):

$$\lim_{\Delta x \to 0} \frac{f(x_0 + c \cdot \Delta x) - f(x_0)}{c \cdot \Delta x} = f'(x_0).$$

注意 上面的例题强调了, 导数归根到底, 就只是具有某种特定构造的函数极限而已. 所以你也可以用函数 f 和某个点 x_0 定义了一个新的函数极限, 并给它属于自己名字, 但是这可能不会像导数那般如此充满现实意义和重要性.

定义 2.1.3: 导函数

若函数 $f(x)$ 在区间 (a,b) 内每一点 x 处都可导, 则称 $f(x)$ 在 (a,b) 内可导. 这时, 每一点 $x\in(a,b)$ 都对应一个导数值 $f'(x)$, 这样便定义了一个新的函数

$$x \mapsto f'(x),$$

称之为 $f(x)$ 的 **导函数**. 在不引起混淆的情况下, 也将导函数简称为导数.

设 $f(x)\equiv c$ (c 为常数), 即 $f(x)$ 为常数函数. 这时, 对于任一点 x 及 $\Delta x\neq 0$, 总有 $\Delta y = f(x+\Delta x) - f(x) = 0$, 因而 $f'(x)\equiv 0$. 简单地说, **常数函数的导数为零**. 这个例子的几何意义是很明显的: 在 Oxy 平面上, 一条水平直线上每一点处的切线都是水平的. 反过来, 若一个函数 $f(x)$ 在其定义域内每一点都有导数 $f'(x)$, 且 $f'(x)$ 恒等于零, 那么它的图形在每一点处的切线都是水平的. 从直观上看, $f(x)$ 的图形一定是一条水平线. 换句话说, 就是 $f(x)$ 恒等于一个常数. 这样, 我们从直观上说明了一个命题.

命题 2.1.1

在一个区间内导数恒等于零的函数当且仅当它是一个常数函数.[a]

[a]此命题的严格证明要用到微分中值定理, 留待以后在第四章中完成.

例 2.1.5 $(\sin x)' = \cos x$.

证明. 事实上, 对于 $y=\sin x$ 以及任意给定的 x 和 $\Delta x\neq 0$, 有

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

(和差化积[1]是经常用的三角恒等式) 因而

$$\begin{aligned}
\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} &= \lim_{\Delta x\to 0} 2\frac{\sin\frac{\Delta x}{2}}{\Delta x}\cos\left(x+\frac{\Delta x}{2}\right) \\
&= \lim_{\Delta x\to 0}\frac{\sin\frac{\Delta x}{2}}{\frac{\Delta x}{2}}\cdot\lim_{\Delta x\to 0}\cos\left(x+\frac{\Delta x}{2}\right) \\
&= \cos x.
\end{aligned}$$

最后一步用到了余弦函数的连续性[2]以及重要极限 $\lim\limits_{x\to 0}\dfrac{\sin x}{x} = 1$. □

[1]
$$\begin{aligned}
\sin\alpha+\sin\beta &= 2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}. \\
\sin\alpha-\sin\beta &= 2\cos\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}. \\
\cos\alpha+\cos\beta &= 2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}. \\
\cos\alpha-\cos\beta &= -2\sin\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}.
\end{aligned}\tag{2.1}$$

[2]连续性告诉我们 $\lim\limits_{\Delta x\to 0} f(x_0+\Delta x) = f(x_0)$.

由上述证明可知, 正弦函数的导数为余弦函数这一简单而优美的事实, 是建立在极限 $\lim\limits_{x\to 0}\dfrac{\sin x}{x}=1$ 的基础之上的. 这里要着重指出: 这一极限公式成立又是建立在正弦函数的自变量 x 采用弧度制的基础之上的. 如果我们不采用弧度制, 而采用其他制, 那时 $\sin x$ 与 x 之比在 $x\to 0$ 时的极限就不再是 1, 而是另外一个常数[3]. 这就会导致正弦函数的导数是余弦函数乘以某个常数, 将会在计算上带来麻烦. 这就是为什么在科学计算中三角函数使用弧度制的原因之一.

例 2.1.6 $(\cos x)'=-\sin x$.

证明. 按照与之前相似的步骤, 我们可以证明 $(\cos x)'=-\sin x$. 对于 $y=\cos x$ 以及任意给定的 x 和 $\Delta x\neq 0$, 有

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

因此

$$\begin{aligned}\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}&=\lim_{\Delta x\to 0}-2\frac{\sin\frac{\Delta x}{2}}{\Delta x}\sin\left(x+\frac{\Delta x}{2}\right)\\&=\lim_{\Delta x\to 0}\frac{\sin\frac{\Delta x}{2}}{\frac{\Delta x}{2}}\cdot\lim_{\Delta x\to 0}\left(-\sin\left(x+\frac{\Delta x}{2}\right)\right)\\&=-\sin x.\end{aligned}$$

□

例 2.1.7 设 m 为一个正整数, 则 $(x^m)'=mx^{m-1}$.

证明. 事实上, 对于 $y=x^m$ 以及任意的 x 和 $\Delta x\neq 0$, 根据二项式定理, 有

$$\Delta y=(x+\Delta x)^m-x^m=mx^{m-1}\Delta x+\frac{1}{2!}m(m-1)x^{m-2}(\Delta x)^2+\cdots+(\Delta x)^m,$$

从而

$$\frac{\Delta y}{\Delta x}=mx^{m-1}+\frac{1}{2!}m(m-1)x^{m-2}\Delta x+\cdots+(\Delta x)^{m-1}.$$

注意到上式中自第二项开始都包含有 Δx 的方幂, 因此它们在 $\Delta x\to 0$ 时趋向于零, 这就得到

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=mx^{m-1}.$$

后面我们还要证明, 当 m 不是正整数而是一般的实数时, 本例子给出的公式依然成立. □

例 2.1.8 $(\mathrm{e}^x)'=\mathrm{e}^x$.

[3]现在考虑使用其他角度单位, 比如角度制 (度数). 在角度制中, 角度 x° 和对应的弧度 x 的关系为 $x=\dfrac{\pi}{180}x^\circ$. 假设我们想在度数制下计算 $\lim\limits_{x^\circ\to 0}\dfrac{\sin x^\circ}{x^\circ}$, 我们需要将角度从度数转换为弧度制:

$$\lim_{x^\circ\to 0}\frac{\sin x^\circ}{x^\circ}=\lim_{x^\circ\to 0}\frac{\sin\left(\frac{\pi}{180}x^\circ\right)}{x^\circ}=\lim_{x^\circ\to 0}\frac{\sin\left(\frac{\pi}{180}x^\circ\right)}{\frac{\pi}{180}x^\circ}\cdot\frac{\pi}{180}.$$

由于我们已经知道弧度制下的极限为 1, 因此 $\lim\limits_{x^\circ\to 0}\dfrac{\sin x^\circ}{x^\circ}=\dfrac{\pi}{180}$. 由此可见, 在角度制下, 计算的结果将会多出一个常数 $\dfrac{\pi}{180}$.

证明. 对于 $y = \mathrm{e}^x$ 以及任意的 x 和 $\Delta x \neq 0$, 有

$$\Delta y = \mathrm{e}^{x+\Delta x} - \mathrm{e}^x = \mathrm{e}^x \left(\mathrm{e}^{\Delta x} - 1\right).$$

(换元法) 令 $\eta = \mathrm{e}^{\Delta x} - 1$, 则当 $\Delta x \to 0$ 时, 显然有 $\eta \to 0$. 另外, 有 $\Delta x = \ln(1+\eta)$. 于是

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \mathrm{e}^x \lim_{\Delta x \to 0} \frac{\mathrm{e}^{\Delta x} - 1}{\Delta x} = \mathrm{e}^x \lim_{\eta \to 0} \frac{\eta}{\ln(1+\eta)} = \mathrm{e}^x \lim_{\eta \to 0} \frac{1}{\ln(1+\eta)^{\frac{1}{\eta}}}.$$

根据之前的结论, 有

$$\lim_{\eta \to 0} (1+\eta)^{\frac{1}{\eta}} = \lim_{\xi \to \infty} \left(1 + \frac{1}{\xi}\right)^{\xi} = \mathrm{e}.$$

再由对数函数的连续性, 并注意到 $\ln x$ 表示以 e 为底的对数, 可得

$$\lim_{\eta \to 0} \frac{1}{\ln(1+\eta)^{\frac{1}{\eta}}} = \frac{1}{\ln(\lim_{\eta \to 0}(1+\eta)^{\frac{1}{\eta}})} = 1.$$

从而立即得到我们所要的结论. □

这是一个非常奇妙的事实: 一个函数的导数值处处等于函数本身的值. 在常见的函数中, 除函数 $y = C\mathrm{e}^x$ (C 为常数) 之外, 没有具有这种性质的函数. 那么, 导函数等于自身的函数都是形如 $y = C\mathrm{e}^x$ (C 为常数) 吗? 答案是肯定的. 这个问题涉及到我们要求解一个**微分方程**[4]

$$f'(x) = f(x),$$

这个方程表示函数的导数等于它自身. 何为微分方程? 普通方程通常涉及未知数, 这些未知数是常数. 例如, 方程 $2x + 3 = 7$ 是一个普通方程, 未知数 x 是一个常数, 通过代数运算可以求出 x 的值. 微分方程涉及的是**未知函数**, 而不仅仅是常数. 微分方程描述的是一个函数及其导数之间的关系. 例如, 微分方程 $f'(x) = f(x)$ 中的未知量是一个函数 $f(x)$, 它的导数 $f'(x)$ 和它本身 $f(x)$ 之间存在一定的关系. 然后我们想找出所有满足这个关系的未知函数 $f(x)$. 再比如, 我们试找到函数 $f(x)$ 使得

$$(f'(x))^2 + (f(x))^2 = 1.$$

显然, 正弦函数和余弦函数满足这个微分方程. 微分方程是下学期的重点.

例 2.1.9 对于一般的指数函数 $y = a^x$ ($a > 0$), 我们有

$$(a^x)' = a^x \ln a.$$

[4]微分方程在工学中扮演着极其重要的角色:

1. 物理建模: 许多物理系统和过程可以用微分方程来描述, 如热传导, 流体动力学, 电磁学等.
2. 控制系统: 在自动控制, 信号处理等领域, 微分方程被广泛用于系统建模和分析.
3. 结构分析: 在土木工程和机械工程中, 微分方程用于分析结构的应力和变形.
4. 电气工程: 电路分析和电磁场理论都依赖于微分方程.
5. 化学工程: 反应动力学和传质过程常用微分方程来描述.

证明. 此结论的证明与上一例题类似 (还是换元法)：我们令 $\eta = a^{\Delta x} - 1$, 这时就有 $\Delta x = \log_a(1+\eta)$, 于是

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = a^x \lim_{\eta \to 0} \frac{\eta}{\log_a(1+\eta)} = a^x \frac{1}{\log_a \mathrm{e}} = a^x \ln a.$$

□

从上面两个例子中我们看到, 以常数 e 为底时指数函数的导数表达式最简单. 也正是由于这个原因, 在一般科学计算中多数采用以 e 为底的对数. 大家知道, 自然对数的引进要归功于欧拉.

根据导数的几何意义并应用直线的点斜式方程, 可知曲线 $y = f(x)$ 在点 $M(x_0, y_0)$ 处的切线方程为

$$y - y_0 = f'(x_0)(x - x_0).$$

过切点 $M(x_0, y_0)$ 且与切线垂直的直线叫做曲线 $y = f(x)$ 在点 M 处的法线. 如果 $f'(x_0) \neq 0$, 法线的斜率为 $-\dfrac{1}{f'(x_0)}$, 从而法线方程为

$$y - y_0 = -\frac{1}{f'(x_0)}(x - x_0).$$

即切线的斜率乘以法线的斜率等于 -1.

例 2.1.10 求等边双曲线 $y = \dfrac{1}{x}$ 在点 $\left(\dfrac{1}{2}, 2\right)$ 处的切线的斜率, 并写出在该点处的切线方程和法线方程.

证明. 根据导数的几何意义知道, 所求切线的斜率为 $k_1 = y'|_{x=\frac{1}{2}}$. 由于 $y' = \left(\dfrac{1}{x}\right)' = -\dfrac{1}{x^2}$, 于是

$$k_1 = -\left.\frac{1}{x^2}\right|_{x=\frac{1}{2}} = -4.$$

从而所求切线方程为 $y - 2 = -4\left(x - \dfrac{1}{2}\right)$, 即

$$4x + y - 4 = 0.$$

所求法线的斜率为 $k_2 = -\dfrac{1}{k_1} = \dfrac{1}{4}$, 于是所求法线方程为 $y - 2 = \dfrac{1}{4}\left(x - \dfrac{1}{2}\right)$, 即

$$2x - 8y + 15 = 0.$$

□

2.1.3 可导与连续性

现在我们来讨论可导与连续的关系.

命题 2.1.2

若函数 $f(x)$ 在一点处可导, 则该函数在这一点处连续.

证明. 假定函数 $f(x)$ 在点 $x=x_0$ 处可导, 这时我们有极限

$$\lim_{\Delta x\to 0}\frac{f\left(x_0+\Delta x\right)-f\left(x_0\right)}{\Delta x}=f'\left(x_0\right).$$

也就是说, 如果我们定义一个函数

$$\eta(\Delta x):=\frac{f\left(x_0+\Delta x\right)-f\left(x_0\right)}{\Delta x}-f'\left(x_0\right),$$

那么当 $\Delta x\to 0$ 时, $\eta(\Delta x)\to 0$. 通过对上面做变形, 当 $\Delta x\to 0$ 时,

$$f\left(x_0+\Delta x\right)-f\left(x_0\right)=f'\left(x_0\right)\Delta x+\eta(\Delta x)\Delta x\to 0,$$

即 $f\left(x_0+\Delta x\right)\to f\left(x_0\right)(\Delta x\to 0)$. 这表明, $f(x)$ 在点 x_0 处连续. □

我们做一个更加直观的解释. 若函数 $f(x)$ 在一点 x_0 处可导, 即

$$f'\left(x_0\right)=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\frac{f\left(x_0+\Delta x\right)-f\left(x_0\right)}{\Delta x}=\lim_{x\to x_0}\frac{f(x)-f\left(x_0\right)}{x-x_0}.$$

注意到右侧三个极限式子的分母 Δx 的极限值都是 0. 所以为了使得整个分式的极限值成为一个常数, 分子部分的极限值也应当是 0. (如果分子的极限值不是零而是其他某个常数甚至是无穷大, 那么整体会发散到无穷大) 所以, 有

$$\Delta y\to 0\ (\Delta x\to 0),\quad \text{或 } f\left(x_0+\Delta x\right)\to f\left(x_0\right)\ (\Delta x\to 0),\quad \text{或 } f(x)\to f\left(x_0\right)\ \left(x\to x_0\right).$$

这些都说明了连续性.

可见, 一个函数 $f(x)$ 若在区间 (a,b) 上可导, 则 $f(x)$ 必在 (a,b) 上连续. 因此, 连续是可导的必要条件. 但是, 函数 $f(x)$ 在点 x_0 处连续并不意味着它在点 x_0 处可导.

- 比如, 函数

  $$y=x^{\frac{1}{3}}$$

  在点 $x=0$ 处是连续的, 但它在该点处是不可导的. 事实上, 在该点处有

  $$\frac{\Delta y}{\Delta x}=\frac{(0+\Delta x)^{\frac{1}{3}}-0^{\frac{1}{3}}}{\Delta x}=\frac{(\Delta x)^{\frac{1}{3}}}{\Delta x}=\frac{1}{(\Delta x)^{\frac{2}{3}}}.$$

  因而当 $\Delta x\to 0$ 时, $\dfrac{\Delta y}{\Delta x}$ 趋向于 ∞, 而不趋向于一个有限数. 实际上, 画图可知, 曲线 $y=x^{\frac{1}{3}}$ 在点 $(0,0)$ 处的切线垂于 x 轴, 因而其与 x 轴正向夹角的正切为 ∞, 导致导数不存在. (具体地, 左导数和右导数都不存在.)

- 另外, 根据例题 2.1.2, 我们已经知道函数

  $$y=|x|$$

  在点 $x=0$ 处也是不可导的, 但它在点 $x=0$ 处是连续的. 画图可知, 曲线 $y=|x|$ 在点 $(0,0)$ 处是一个尖角, 两侧的割线有不同的极限位置, 故在该点处没有切线.

- 下面的函数是另外一个连续而不可导的例子:

  $$y=f(x)=\begin{cases}x\sin\dfrac{1}{x}, & x\neq 0\\ 0, & x=0\end{cases}$$

由于 $|f(x)| \leqslant |x|$, 由夹逼定理可知 $\lim\limits_{x\to 0}|f(x)| = 0$ 因而 $\lim\limits_{x\to 0} f(x) = 0$.[5] 故该函数在点 $x = 0$ 处显然是连续的. 但它在点 $x = 0$ 处是不可导的. 事实上, 在点 $x = 0$ 处,

$$\frac{\Delta y}{\Delta x} = \frac{f(\Delta x) - f(0)}{\Delta x} = \sin\frac{1}{\Delta x}.$$

在 $\Delta x \to 0$ 的过程中没有极限. 画图可知, 在这个例子中, 曲线 $y = f(x)$ 的过点 $(0,0)$ 和 $(0+\Delta x, y_0 + \Delta y)$ $(y_0 = f(0))$ 的割线在 $\Delta x \to 0$ 的过程中摇摆不定. (具体地, 左导数和右导数都不存在.)

我们曾经说过, 初等函数在其有定义的区间内是连续的. 但初等函数可能在其有定义的区间内的个别点处是不可导的. 例如, 上面例子中 $y = x^{\frac{1}{3}}$ 与 $y = |x| = \sqrt{x^2}$ 都是初等函数, 它们在点 $x = 0$ 处均不可导.

魏尔斯特拉斯 (K.Weierstrass)[6] 举出了一个著名例子, 说明一个处处连续的函数可以处处不可导. 关于其图像, 可看视频 维尔斯特拉斯 (Weierstrass) 函数: 处处连续但处处不可导的函数 _ 哔哩哔哩 _bilibili 和 Weierstrass function | Desmos. 魏尔施特拉斯函数的构造是

$$f(x) = \sum_{n=0}^{\infty} a^n \cos\left(b^n \pi x\right),$$

其中 $0 < a < 1$, b 为正的奇数, 使得 $ab > 1 + \frac{3}{2}\pi$. 这里无穷多项的求和 (即, 级数) 我们假定其结果是一个定数, 而不是发散到无穷. 下学期会学习级数.

结论 2.1.1

存在一个处处连续但处处不可导的函数.

例 2.1.11 求出 a 和 b 的值, 使得函数 $f(x) = \begin{cases} ax+b, & x < 2, \\ x^2, & x \geqslant 2 \end{cases}$ 处处可导.

证明. $f(x)$ 在点 $x \neq 2$ 处可导, 故只需讨论 $f(x)$ 在点 $x = 2$ 处的性质. 首先, 要 $f(x)$ 在点 $x = 2$ 处可导, 那么 $f(x)$ 在点 $x = 2$ 处连续, 从而有

$$\lim_{x\to 2+0} f(x) = \lim_{x\to 2+0} x^2 = 4 = \lim_{x\to 2-0} f(x) = \lim_{x\to 2-0}(ax+b) = 2a+b,$$

即 a 和 b 应满足 $2a + b = 4$. 再利用可导性知, 极限 $\lim\limits_{x\to 2+0}\frac{f(x)-f(2)}{x-2}$ 和 $\lim\limits_{x\to 2-0}\frac{f(x)-f(2)}{x-2}$ 都存在且相等. 而

$$\begin{aligned}
\lim_{x\to 2+0}\frac{f(x)-f(2)}{x-2} &= \lim_{x\to 2+0}\frac{x^2-2^2}{x-2} = \lim_{x\to 2+0}(x+2) = 4, \\
\lim_{x\to 2-0}\frac{f(x)-f(2)}{x-2} &= \lim_{x\to 2-0}\frac{(ax+b)-4}{x-2}.
\end{aligned}$$

利用前面得到的条件 $2a + b = 4$, 则有

$$\lim_{x\to 2-0}\frac{f(x)-f(2)}{x-2} = \lim_{x\to 2-0}\frac{(ax+b)-(2a+b)}{x-2} = a,$$

于是 $a = 4$, 再由 $2a + b = 4$ 得 $b = -4$. □

[5]这里使用了一个简单的事实: $\lim\limits_{x\to a}|f(x)| = 0$ 当且仅当 $\lim\limits_{x\to a} f(x) = 0$, 只要使用极限的定义和等式 $|f(x)-0| = |f(x)| = ||f(x)|-0|$ 便可证明. 这一结论也应该纳入工具库, 面对某些情况它非常有用.

[6]连续但处处不可导，魏尔斯特拉斯函数，严谨数学下的妖兽 _ 哔哩哔哩 _bilibili

2.1.4 导数的四则运算

根据导数的定义及极限的四则运算法则, 很容易给出导数的四则运算法则.

定理 2.1.1

设函数 $f(x)$ 及 $g(x)$ 在一个共同的区间 (a,b) 上有定义, 并且在每一点 $x\in(a,b)$ 处都可导, 则我们有

$$(f(x)\pm g(x))' = f'(x)\pm g'(x),$$
$$(f(x)g(x))' = f'(x)g(x)+f(x)g'(x).$$

特别地, 对于任意常数 c, 有

$$(cf(x))' = cf'(x).$$

当 $g(x)\neq 0$ 时,

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x)-f(x)g'(x)}{(g(x))^2}.$$

这里一定要牢记, 商的求导法则中是有顺序的! 是有顺序的! 是有顺序的!

证明.

- 下面证明函数和的导数公式. 根据导数的定义, 函数和的导数可以表示为

$$(f(x)+g(x))' = \lim_{\Delta x\to 0}\frac{(f(x+\Delta x)+g(x+\Delta x))-(f(x)+g(x))}{\Delta x}.$$

将分子展开重组得

$$(f(x+\Delta x)+g(x+\Delta x))-(f(x)+g(x)) = (f(x+\Delta x)-f(x))+(g(x+\Delta x)-g(x)).$$

因此, 导数公式变为

$$(f(x)+g(x))' = \lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}+\lim_{\Delta x\to 0}\frac{g(x+\Delta x)-g(x)}{\Delta x} = f'(x)+g'(x).$$

同理, 函数差的导数公式也可以类似地证明.

- 下面证明函数积的导数公式. 根据导数的定义, 函数积的导数可以表示为

$$(f(x)g(x))' = \lim_{\Delta x\to 0}\frac{f(x+\Delta x)g(x+\Delta x)-f(x)g(x)}{\Delta x}.$$

将分子进行变形

$$f(x+\Delta x)g(x+\Delta x)-f(x)g(x) = (f(x+\Delta x)-f(x))g(x+\Delta x)+f(x)(g(x+\Delta x)-g(x)).$$

因此, 导数公式变为

$$\begin{aligned}(f(x)g(x))' &= \lim_{\Delta x\to 0}\frac{(f(x+\Delta x)-f(x))g(x+\Delta x)+f(x)(g(x+\Delta x)-g(x))}{\Delta x}\\ &= \lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}g(x+\Delta x)+f(x)\cdot\lim_{\Delta x\to 0}\frac{g(x+\Delta x)-g(x)}{\Delta x}\\ &= f'(x)g(x)+f(x)g'(x).\end{aligned}$$

注意到 $g(x)$ 的连续性 $g(x+\Delta x)\to g(x)(\Delta x\to 0)$.

- 下面证明函数商的导数公式. 对于 $g(x)\neq 0$ 的情况, 根据导数的定义, 函数商的导数可以表示为

$$\left(\frac{f(x)}{g(x)}\right)'=\lim_{\Delta x\to 0}\frac{\frac{f(x+\Delta x)}{g(x+\Delta x)}-\frac{f(x)}{g(x)}}{\Delta x}.$$

将分子进行变形

$$\begin{aligned}\frac{f(x+\Delta x)}{g(x+\Delta x)}-\frac{f(x)}{g(x)}&=\frac{f(x+\Delta x)g(x)-f(x)g(x+\Delta x)}{g(x)g(x+\Delta x)}\\&=\frac{(f(x+\Delta x)-f(x))g(x)-f(x)(g(x+\Delta x)-g(x))}{g(x)g(x+\Delta x)}.\end{aligned}$$

因此, 导数公式变为

$$\begin{aligned}\left(\frac{f(x)}{g(x)}\right)'&=\lim_{\Delta x\to 0}\frac{(f(x+\Delta x)-f(x))g(x)-f(x)(g(x+\Delta x)-g(x))}{\Delta x g(x)g(x+\Delta x)}\\&=\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}\frac{g(x)}{g(x)g(x+\Delta x)}-\lim_{\Delta x\to 0}\frac{g(x+\Delta x)-g(x)}{\Delta x}\frac{f(x)}{g(x)g(x+\Delta x)}\\&=\frac{f'(x)g(x)-f(x)g'(x)}{g(x)^2}.\end{aligned}$$

注意到 $g(x)$ 的连续性 $g(x+\Delta x)\to g(x)(\Delta x\to 0)$. 证毕.

□

该定理可推广到任意有限个可导函数的情形. 例如, 设 $u=u(x), v=v(x), w=w(x)$ 均可导, 则有

$$(u+v-w)'=u'+v'-w'.$$

亦有, $(uvw)'=[(uv)w]'=(uv)'w+(uv)w'=(u'v+uv')\,w+uvw'$, 即

$$(uvw)'=u'vw+uv'w+uvw'.$$

例 2.1.12 $(\sin x\cos x)'=\cos^2 x-\sin^2 x=\cos 2x.$

例 2.1.13 $(\tan x)'=\left(\dfrac{\sin x}{\cos x}\right)'=\dfrac{\cos^2 x+\sin^2 x}{\cos^2 x}=\dfrac{1}{\cos^2 x}=\sec^2 x.$

例 2.1.14 $\left(\cos x+x^3+\mathrm{e}^x\right)'=-\sin x+3x^2+\mathrm{e}^x.$

例 2.1.15 设 m 是正整数, 证明

$$\left(x^{-m}\right)'=-mx^{-m-1}.$$

利用函数商的导数公式, 有

$$\left(x^{-m}\right)'=\left(\frac{1}{x^m}\right)'=\frac{(1)'\cdot x^m-1\cdot(x^m)'}{(x^m)^2}=\frac{-mx^{m-1}}{x^{2m}}=-mx^{-m-1}.$$

2.1.5 现实中我们如何计算导数？中心差分法

命题 2.1.3: 导数又一等价定义

设函数 $f(x)$ 在点 x_0 处可导, 证明

$$\lim_{\Delta x\to 0}\frac{f\left(x_0+\Delta x\right)-f\left(x_0-\Delta x\right)}{2\Delta x}=f'\left(x_0\right).$$

证明. 利用导数的定义, 函数 $f(x)$ 在点 x_0 处的导数 $f'(x_0)$ 可以表示为

$$f'(x_0)=\lim_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x}.$$

我们需要证明的极限可以重新表示为

$$\lim_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0-\Delta x)}{2\Delta x}=\lim_{\Delta x\to 0}\left(\frac{f(x_0+\Delta x)-f(x_0)}{2\Delta x}+\frac{f(x_0+(-\Delta x))-f(x_0)}{2(-\Delta x)}\right).$$

根据导数的定义, 上式等于

$$\frac{1}{2}f'(x_0)+\frac{1}{2}f'(x_0)=f'(x_0).$$

证毕. □

上述命题产生了现实中绝大多数应用场景下计算导数的最常用方法 — 中心差分法. 假设我们有一个非常复杂的函数

$$f(x)=\frac{\frac{e^{x^2}\sin(x^3)+\cos\left(x^5\right)}{x^4+1}}{\frac{\cos(x^5)}{x^6+2}}\cdot\frac{\frac{\ln(x^7+3)}{x^8+4}}{\frac{\tan(x^9)}{x^{10}+5}}.$$

这个函数比之前的例子复杂得多, 包含多个商和复杂的嵌套结构. 我们希望在某个点 x_0 处计算其导数 $f'(x_0)$. 由于这个函数非常复杂, 我们无法直接通过解析方法求得其导数 (理论上也不是不行, 但我劝你别算). 因此, 我们可以使用中心差分法来近似计算导数.

结论 2.1.2: 中心差分法

中心差分法的基本思想是通过函数值的差分来近似导数. 具体公式为:

$$f'(x_0)\approx\frac{f(x_0+h)-f(x_0-h)}{2h},$$

其中 h 是一个很小的正数, 通常称为步长. 计算步骤

1. 选择步长 h: 例如, 选择 $h=0.001$.
2. 计算函数值: 计算 $f(x_0+h)$ 和 $f(x_0-h)$.
3. 应用中心差分公式: 将计算得到的函数值代入中心差分公式, 得到近似的导数值.

假设我们希望在 $x_0=1$ 处计算导数 $f'(1)$.

1. 选择步长 $h=0.001$.

2. 计算函数值:

$$f(1+0.001)=\frac{\frac{e^{(1.001)^2}\sin((1.001)^3)+\cos((1.001)^5)}{(1.001)^4+1}}{\frac{\cos((1.001)^5)}{(1.001)^6+2}}\times\frac{\frac{\ln((1.001)^7+3)}{(1.001)^8+4}}{\frac{\tan((1.001)^9)}{(1.001)^{10}+5}}\approx 8.3093$$

$$f(1-0.001)=\frac{\frac{e^{(0.999)^2}\sin((0.999)^3)+\cos((0.999)^5)}{(0.999)^4+1}}{\frac{\cos((0.999)^5)}{(0.999)^6+2}}\times\frac{\frac{\ln((0.999)^7+3)}{(0.999)^8+4}}{\frac{\tan((0.999)^9)}{(0.999)^{10}+5}}\approx 8.4601$$

3. 应用中心差分公式:

$$f'(1)\approx\frac{f(1+0.001)-f(1-0.001)}{2\times 0.001}\approx\frac{8.3093-8.4601}{0.002}=-75.4042.$$

因此, 通过中心差分法, 我们近似得到了 $f'(1)\approx -75.4042$.

现实中, 对于复杂的函数, 我们无法通过解析方法直接求得其导数. 通过中心差分法, 我们可以近似计算导数. 中心差分法在处理复杂函数时具有更高的精度和数值稳定性[7], 因此在实际应用中更为常用. 详细代码见 codes/finite-difference.ipynb · 赖志坚/高等数学 B-07 班-资料共享 - 码云 - 开源中国 (gitee.com) .

命题 2.1.4: 导数又一等价定义

设函数 $f(x)$ 在点 x_0 处可导, 证明

$$\lim_{\Delta x\to 0}\frac{f(x_0)-f(x_0-\Delta x)}{\Delta x}=f'(x_0).$$

证明. 根据导数的定义, 函数 $f(x)$ 在 x_0 处的导数为

$$f'(x_0)=\lim_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h}.$$

令 $\Delta x:=-h$, 则 $\Delta x\to 0$ 等价于 $h\to 0$. 因此, 我们可以将原式改写为

$$\lim_{\Delta x\to 0}\frac{f(x_0-\Delta x)-f(x_0)}{-\Delta x}=\lim_{\Delta x\to 0}\frac{f(x_0)-f(x_0-\Delta x)}{\Delta x}.$$

□

有以上命题导出的 $f'(x_0)\approx\dfrac{f(x_0)-f(x_0-h)}{h}$ 称之为后向差分, 它和前向差分相近. 其误差精度仍旧不如中心差分法.

2.2 复合函数与反函数的导数

现在, 我们讨论复合函数与反函数的导数. 有了前面一节的结果, 再加上复合函数与反函数的求导公式, 我们就会计算所有初等函数的导数.

[7] 为什么不使用原定义式 $\lim_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x}$ 导出的 $f'(x_0)\approx\frac{f(x_0+h)-f(x_0)}{h}$ (称之为前向差分) ? 因为中心差分法通常比前向差分具有更高的精度. 中心差分是通过对称的方式计算差分, 误差项通常是 $O(h^2)$ 级别的, 而前向或后向差分的误差项通常是 $O(h)$ 级别的. 将来我们会用泰勒展开来分析这一事实.

2.2.1 复合函数的导数

定理 2.2.1: 复合函数的导数

设函数 $y=f(x)$ 在区间 (a,b) 上有定义, 且其值域被包含于区间 (A,B), 又设函数 $z=g(y)$ 在 (A,B) 上有定义. 若 $y=f(x)$ 在点 $x_0\in(a,b)$ 处可导, 并且 $z=g(y)$ 在相应的点 $y_0=f(x_0)$ 处可导, 则复合函数 $g(f(x))$ 在点 x_0 处可导, 并且

$$\left.\frac{\mathrm{d}}{\mathrm{d}x}g(f(x))\right|_{x=x_0}=g'(y_0)\,f'(x_0),$$

或写成

$$\left.\frac{\mathrm{d}z}{\mathrm{d}x}\right|_{x=x_0}=\left.\frac{\mathrm{d}z}{\mathrm{d}y}\right|_{y=y_0}\cdot\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=x_0}.$$

证明. 对于任意一点 $x_0\in(a,b)$ 及自变量 x 的一个增量 $\Delta x\neq 0$, 我们有

$$\Delta z=g\left(f\left(x_0+\Delta x\right)\right)-g\left(f\left(x_0\right)\right).$$

要求 $z=g(f(x))$ 在点 x_0 处的导数, 实际上就是要求极限

$$\lim_{\Delta x\to 0}\frac{\Delta z}{\Delta x}=\lim_{\Delta x\to 0}\frac{g\left(f\left(x_0+\Delta x\right)\right)-g\left(f\left(x_0\right)\right)}{\Delta x}.$$

令 $y_0=f\left(x_0\right)$, 则由 $\Delta y=f\left(x_0+\Delta x\right)-f\left(x_0\right)$ 有 $f\left(x_0+\Delta x\right)=y_0+\Delta y$, 于是

$$\Delta z=g\left(y_0+\Delta y\right)-g\left(y_0\right).$$

因 $g(y)$ 在点 y_0 处有导数, 故

$$\lim_{\Delta y\to 0}\frac{g\left(y_0+\Delta y\right)-g\left(y_0\right)}{\Delta y}=g'\left(y_0\right),$$

即

$$\lim_{\Delta y\to 0}\left(\frac{g\left(y_0+\Delta y\right)-g\left(y_0\right)}{\Delta y}-g'\left(y_0\right)\right)=0.$$

令

$$\eta(\Delta y):=\frac{g\left(y_0+\Delta y\right)-g\left(y_0\right)}{\Delta y}-g'\left(y_0\right)\quad(\Delta y\neq 0),$$

则 $\lim\limits_{\Delta y\to 0}\eta(\Delta y)=0$, 且

$$g\left(y_0+\Delta y\right)-g\left(y_0\right)=g'\left(y_0\right)\Delta y+\eta(\Delta y)\Delta y\quad(\Delta y\neq 0).$$

我们补充定义 $\eta(0)=0$, 则上式不仅对 $\Delta y\neq 0$ 成立, 而且对 $\Delta y=0$ 也成立. 这样, 不论 $\Delta y=f\left(x_0+\Delta x\right)-f\left(x_0\right)$ 是否等于零, 当 $\Delta x\neq 0$ 时, 我们总有

$$\frac{\Delta z}{\Delta x}=\frac{g\left(y_0+\Delta y\right)-g\left(y_0\right)}{\Delta x}=\frac{g'\left(y_0\right)\Delta y+\eta(\Delta y)\Delta y}{\Delta x}=g'\left(y_0\right)\frac{\Delta y}{\Delta x}+\eta(\Delta y)\frac{\Delta y}{\Delta x}.\tag{2.2}$$

令 $\Delta x\to 0$, 则上式右端第一项趋向于 $g'\left(y_0\right)f'\left(x_0\right)$, 而第二项趋向于零. 这样, 我们得到复合函数 $z=g(f(x))$ 在点 x_0 处可导, 并且

$$\left.\frac{\mathrm{d}z}{\mathrm{d}x}\right|_{x=x_0}=g'\left(y_0\right)f'\left(x_0\right).$$

□

结论 2.2.1: 复合函数求导公式 — 链式法则 (chain rule)

在定理 2.2.1 中, 如果 $y=f(x)$ 及 $z=g(y)$ 在各自的定义域内可导, 则对于任意一点 $x\in(a,b)$, 我们有

$$\frac{\mathrm{d}}{\mathrm{d}x}g(f(x))=g'(f(x))f'(x),$$

或写成

$$\frac{\mathrm{d}z}{\mathrm{d}x}=\frac{\mathrm{d}z}{\mathrm{d}y}\cdot\frac{\mathrm{d}y}{\mathrm{d}x}.$$

其中 $\frac{\mathrm{d}z}{\mathrm{d}y}$ 表示将 z 视作 y 的函数时对 y 的导数.

这就是复合函数求导公式, 它是导数计算中非常基本的公式, 读者应通过练习熟练掌握它. 复合函数的求导法则可以推广到多个中间变量的情形. 我们以两个中间变量为例, 设 $y=f(u), u=\varphi(v), v=\psi(x)$, 则

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\mathrm{d}y}{\mathrm{d}u}\cdot\frac{\mathrm{d}u}{\mathrm{d}x},$$

而 $\frac{\mathrm{d}u}{\mathrm{d}x}=\frac{\mathrm{d}u}{\mathrm{d}v}\cdot\frac{\mathrm{d}v}{\mathrm{d}x}$, 故复合函数 $y=f\{\varphi[\psi(x)]\}$ 的导数为

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\mathrm{d}y}{\mathrm{d}u}\cdot\frac{\mathrm{d}u}{\mathrm{d}v}\cdot\frac{\mathrm{d}v}{\mathrm{d}x}.$$

当然, 这里假定上式右端所出现的导数在相应处都存在.

例 2.2.1 求函数 $\sin 2x$ 的导数.

证明. 令 $z=\sin y$, 其中 $y=2x$, y 为中间变量, 则

$$\frac{\mathrm{d}z}{\mathrm{d}x}=\frac{\mathrm{d}z}{\mathrm{d}y}\cdot\frac{\mathrm{d}y}{\mathrm{d}x}=\cos y\cdot 2=2\cos 2x.$$

□

例 2.2.2 求函数 e^{x^2} 的导数.

证明. 令 $z=\mathrm{e}^{x^2}$. 引入中间变量 $y=x^2$, 这时 $z=\mathrm{e}^y$, 于是

$$\frac{\mathrm{d}z}{\mathrm{d}x}=\frac{\mathrm{d}z}{\mathrm{d}y}\cdot\frac{\mathrm{d}y}{\mathrm{d}x}=\mathrm{e}^y\cdot 2x=2x\mathrm{e}^{x^2}.$$

□

对初学者而言, 在刚开始时引入中间变量是必要的, 但是做过一定练习之后, 就不必将中间变量明显地写出, 只要心中默记什么是中间变量即可. **重点是, 由外层向里层逐步求导.**

例 2.2.3 求函数 $\left(x^2+3x^3\right)^5$ 的导数.

证明. 在计算这个函数的导数时, 首先我们将括号内的量看作一个中间变量, 并对这个变量的 5 次方幂求导数, 然后乘以这个变量对 x 的导数:

$$\left[\left(x^2+3x^3\right)^5\right]'=5\left(x^2+3x^3\right)^4\left(2x+9x^2\right).$$

□

例 2.2.4 求函数 $\cos 2x\cdot\tan x^2$ 的导数.

证明. 由函数积的导数公式有

$$\left(\cos 2x\cdot\tan x^2\right)'=(\cos 2x)'\tan x^2+\cos 2x\left(\tan x^2\right)'$$

而 $(\cos 2x)'=-\sin 2x\cdot 2$, $\left(\tan x^2\right)'=\sec^2 x^2\cdot 2x$, 故

$$\left(\cos 2x\cdot\tan x^2\right)'=-2\sin 2x\cdot\tan x^2+2x\cos 2x\cdot\sec^2 x^2.$$

□

在一些较复杂的函数表达式中, 用一个中间变量不够, 尚需引入多个. 例如, 求 $\mathrm{e}^{\sin x^2}$ 的导数时, 我们需要两个中间变量 $y=x^2, w=\sin y$, 而将 $\mathrm{e}^{\sin x^2}$ 视作 e^w. 即使是这样的情形, 也不必把两个中间变量写出, 只需一次次地在心中将某部分表达式作为一个整体而反复运用复合函数求导公式就可以了. 在上述例子中, 我们先把 $\sin x^2$ 作为一个整体, 用一次复合函数求导公式, 再将 x^2 作为一个整体, 又用一次复合函数求导公式, 即得出结果:

$$\left(\mathrm{e}^{\sin x^2}\right)'=\mathrm{e}^{\sin x^2}\left(\sin x^2\right)'=\mathrm{e}^{\sin x^2}\cos x^2\left(x^2\right)'=2x\cos x^2\mathrm{e}^{\sin x2}.$$

> **注意** 考试必考求导题, 求导题必然是复合函数的求导. 如果课后作业练习量感觉不够, 建议自行找额外题目练习, 稍稍刷一下题.

2.2.2 反函数的导数

现在我们讨论一个函数的导数与其反函数的导数的关系. 这个定理告诉我们: 一个函数在一点处的导数恰好等于其反函数在对应点处的导数的倒数.

定理 2.2.2: 反函数的导数

设函数 $y=f(x)$ 在区间 (a,b) 内连续, 严格单调且其值域为区间 (A,B), 又设其反函数 $x=g(y)$ 在 (A,B) 内点 y_0 处有导数且不为零, 则 $y=f(x)$ 在对应点 $x_0=g\left(y_0\right)$ 处有导数, 并且

$$f'\left(x_0\right)=\frac{1}{g'\left(y_0\right)}\quad\text{或}\quad f'\left(x_0\right)=\frac{1}{g'\left(f\left(x_0\right)\right)}.$$

(这里 $g'\left(f\left(x_0\right)\right)$ 是 $g(y)$ 先对 y 求导数, 然后将其中的 $y=f\left(x_0\right)$ 代入的结果).

证明. 设 $x_0 \in (a,b)$ 使得 $f\left(x_0\right)=y_0$, 取 $\Delta x \neq 0$ 充分小, 使得 $x_0+\Delta x \in (a,b)$. 这时 $\Delta y=f\left(x_0+\Delta x\right)-f\left(x_0\right) \neq 0$ (若不然, 则 $f\left(x_0+\Delta x\right)=f\left(x_0\right)$, 这与 $f(x)$ 严格单调矛盾). 另外, Δx 又可视作 Δy 的函数, 因为由 $\Delta y=f\left(x_0+\Delta x\right)-f\left(x_0\right)$ 可得

$$y_0+\Delta y=f\left(x_0\right)+\Delta y=f\left(x_0+\Delta x\right). \tag{2.3}$$

上式两边同时取函数 g 得

$$g(y_0+\Delta y)=g(f\left(x_0+\Delta x\right))=x_0+\Delta x. \tag{2.4}$$

因为 $x_0=g\left(y_0\right)$, 对上式变形可得

$$\Delta x=g\left(y_0+\Delta y\right)-g\left(y_0\right). \tag{2.5}$$

再由 $g'\left(y_0\right)$ 存在并不等于零, 立即推出

$$f'\left(x_0\right)=\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}=\lim_{\Delta y \to 0} \frac{1}{\frac{\Delta x}{\Delta y}}=\lim_{\Delta y \to 0} \frac{1}{\frac{g(y_0+\Delta y)-g(y_0)}{\Delta y}}=\frac{1}{g'\left(y_0\right)}.$$

注意, 根据 $f(x)$ 在点 x_0 处连续, 当 $\Delta x \to 0$ 时, $\Delta y \to 0$. 证毕. □

定理 2.2.2 中的公式 $f'\left(x_0\right)=\dfrac{1}{g'\left(y_0\right)}$ 有明显的几何意义: 在 Oxy 平面上, $y=f(x)$ 与 $x=g(y)$ 代表的是同一曲线. 该曲线在一点 (x_0, y_0) 处的切线关于 x 轴的斜率是它关于 y 轴的斜率的倒数.

定理 2.2.2 实在是冗长, 多数情况下我们只需要记住以下结果就够了.

结论 2.2.2: 反函数求导公式

在这个定理 2.2.2 中, 若假定 $g(y)$ 在 (A,B) 内处处有导数且导数不为零, 则有

$$f'(x)=\frac{1}{g'(y)}, \qquad \text{其中 } y=f(x), \forall x \in (a,b),$$

或

$$f'(x)=\frac{1}{g'(f(x))}, \qquad \forall x \in (a,b).$$

这就是反函数求导公式.

上述结论也可以由复合函数求导公式获得. 事实上, 我们有 $g(f(x))=x$. 对此式两边求导数, 得 $g'(f(x))f'(x)=1$. 这样子记忆会更加容易.

例 2.2.5 设函数 $y=f(x)=x^5+2x+1$, 而 $x=f^{-1}(y)$ 是它的反函数, 求 $\left(f^{-1}\right)'(4)$.

证明. 容易求出 $f(1)=4$, 则 $f^{-1}(4)=1$. 利用定理 2.2.2, 有

$$\left(f^{-1}\right)'(4)=\frac{1}{f'(1)}.$$

而 $f'(x)=5x^4+2, f'(1)=7$, 所以 $\left(f^{-1}\right)'(4)=\dfrac{1}{7}$. □

例 2.2.6 求函数 $y=\arcsin x$ $(x\in(-1,1))$ 的导数.

证明. 显然, $y=\arcsin x$ 在区间 $(-1,1)$ 内连续且严格单调, 其值域为区间 $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$, 又知其反函数 $x=\sin y$ 在 $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$ 内的导数为 $(\sin y)'=\cos y\neq 0$. 由定理 2.2.2 立即得到

$$(\arcsin x)'=\frac{1}{\cos y},\qquad \text{其中 } y=\arcsin x,$$

即

$$(\arcsin x)'=\frac{1}{\cos(\arcsin x)},\qquad \forall x\in(-1,1).$$

另外, $\cos(\arcsin x)=\sqrt{1-x^2}$,[8] 故

$$(\arcsin x)'=\frac{1}{\sqrt{1-x^2}},\qquad \forall x\in(-1,1).$$

□

我们已经知道, 正弦函数, 余弦函数, 正切函数, 余切函数等三角函数的导数仍是三角函数. 现在我们看到, 反正弦函数的导数却不再是三角函数或反三角函数, 而是一个无理式.

例 2.2.7 利用与上一道例题类似的方法可以证明

$$(\arccos x)'=-\frac{1}{\sqrt{1-x^2}},\qquad \forall x\in(-1,1).$$

这个公式还可以由大家熟知的公式 (对下式两边同时求导)

$$\arcsin x+\arccos x=\frac{\pi}{2}$$

及上一道例题的结果推出.

例 2.2.8 证明:

$$(\arctan x)'=\frac{1}{1+x^2},\qquad (\operatorname{arccot} x)'=-\frac{1}{1+x^2}.$$

证明. 令 $y=\arctan x$, 则 $x=\tan y$, $y\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$. 根据定理 2.2.2, 有

$$(\arctan x)'=\frac{1}{\sec^2 y},\qquad \text{其中 } y=\arctan x.$$

因为

$$\sec^2 y=1+\tan^2 y=1+\tan^2(\arctan x)=1+x^2.$$

所以

$$(\arctan x)'=\frac{1}{1+x^2}.$$

[8] 这是学习函数那节时的一道练习题

同理可证

$$(\operatorname{arccot} x)' = -\frac{1}{\csc^2 y} = -\frac{1}{1+x^2}.$$

□

这样我们看到, 反正切函数及反余切函数的导数是一个简单的二次有理式.

例 2.2.9 证明:

$$(\ln|x|)' = \frac{1}{x}, \quad x \neq 0.$$

证明. 先讨论 $x > 0$ 的情况. 此时, 函数 $y = \ln x$ $(x > 0)$ 是 $x = \mathrm{e}^y$ 的反函数, 那么根据定理 2.2.2 有

$$(\ln x)' = \frac{1}{\mathrm{e}^y}, \qquad \text{其中 } y = \ln x \ (x > 0),$$

即

$$(\ln x)' = \frac{1}{\mathrm{e}^{\ln x}} = \frac{1}{x} \quad (x > 0).$$

当 $x < 0$ 时, 令 $t = -x$, 这时 $t > 0$ 且 $\ln|x| = \ln t$. 将 t 视作中间变量, 利用复合函数求导公式及刚才的结果即得

$$\frac{\mathrm{d}}{\mathrm{d}x}\ln|x| = \frac{\mathrm{d}(\ln t)}{\mathrm{d}t} \cdot \frac{\mathrm{d}t}{\mathrm{d}x} = \frac{1}{t} \cdot (-1) = \frac{1}{x} \quad (x < 0),$$

因而这时 $\ln|x|$ 的导数仍是 $\dfrac{1}{x}$. □

这样我们看到, 以 e 为底的对数函数的导数竟是一个十分简单的函数——倒数函数.

例 2.2.10 用反函数求导法则来证明 $(\log_a|x|)' = \dfrac{1}{x\ln a}$. $(a > 0, a \neq 1)$.

证明. 先讨论 $x > 0$ 的情况. 此时, $y = \log_a x$ 和 $x = a^y$ 是互为反函数的, 且 $(a^y)' = a^y \ln a$. 应用反函数求导法则, 我们得到 $(\log_a x)' = \dfrac{1}{(a^y)'} = \dfrac{1}{a^y \ln a}$. 注意到 $a^y = x$, 我们可以将表达式改写为 $(\log_a x)' = \dfrac{1}{x\ln a}$. 当 $x < 0$ 时, 证明过程和上一道例题完全一致, 略去. □

例 2.2.11 以前我们曾证明过, 当 m 为非零整数时, $(x^m)' = mx^{m-1}$. 现在我们将这一公式推广到 m 为一般实数的情形. 证明: 设 α 为任意实数, 则有

$$(x^\alpha)' = \alpha x^{\alpha-1} \quad (x > 0).$$

证明. 当 $\alpha = 0$ 时, $x^\alpha \equiv 1$, 故其导数为零, 从而结论成立. 现在设 $\alpha \neq 0$, 这时 $0 < x^\alpha$ 可以写成 $x^\alpha = \mathrm{e}^{\alpha\ln x}$, 故由复合函数求导公式有

$$(x^\alpha)' = \mathrm{e}^{a\ln x}(\alpha\ln x)' = x^\alpha \cdot \alpha \cdot \frac{1}{x} = \alpha x^{\alpha-1} \quad (x > 0).$$

□

连同上一节已知的若干基本初等函数的导数, 至此全部基本初等函数的导数已经求出, 得到如下导数表.

结论 2.2.3: 基本初等函数的导数表

1. $(x^\alpha)' = \alpha x^{\alpha-1}$ (α 为任意实数). 特别地, $(\sqrt{x})' = \dfrac{1}{2\sqrt{x}}$; $\left(\dfrac{1}{x}\right)' = -\dfrac{1}{x^2}$.
2. $(\sin x)' = \cos x, \quad (\cos x)' = -\sin x$;
3. $(\tan x)' = \sec^2 x, \quad (\cot x)' = -\csc^2 x$;
4. $(\arcsin x)' = \dfrac{1}{\sqrt{1-x^2}}, |x| < 1; \quad (\arccos x)' = -\dfrac{1}{\sqrt{1-x^2}}, |x| < 1$;
5. $(\arctan x)' = \dfrac{1}{1+x^2}; \quad (\operatorname{arccot} x)' = -\dfrac{1}{1+x^2}$.
6. $(a^x)' = a^x \ln a$ $(a > 0, a \neq 1)$. 特别地, $(\mathrm{e}^x)' = \mathrm{e}^x$.
7. $(\log_a |x|)' = \dfrac{1}{x \ln a}$ $(a > 0, a \neq 1)$. 特别地, $(\ln |x|)' = \dfrac{1}{x}$.

有了基本初等函数的这一导数表, 再加上导数的四则运算法则及复合函数求导公式, 我们就可以计算所有初等函数的导数. 另外, 我们可以断言, 若初等函数可导, 则其导数仍是初等函数. 基本初等函数的导数表是计算初等函数导数的基本依据, 也是今后求不定积分的基础, 因此要求读者在大量练习中记熟它, 用熟它. **它在微积分计算中的重要性, 犹如初等四则运算中的九九乘法口决表.**

2.2.3 对数求导法

在结束本节前, 我们要指出利用对数求导数在某些情况下会带来方便. 它主要包括了两种不同的情景:

1. 处理一些形如 $g(x)^{f(x)}$ 的幂指函数, 使其变为 $\mathrm{e}^{f(x)\ln g(x)}$.
2. 将乘法运算化为加法运算, 将除法运算化为减法运算, 从而使求导数的过程简化.

例 2.2.12 求函数 $y = x^x$ $(x > 0)$ 的导数.

证明. 我们将这个函数改写为

$$y = \mathrm{e}^{x \ln x}.$$

再用复合函数求导公式得

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \mathrm{e}^{x\ln x} \frac{\mathrm{d}}{\mathrm{d}x}(x \ln x) = \mathrm{e}^{x \ln x}(\ln x + 1) = x^x(\ln x + 1).$$

□

对于一般形式的 **幂指函数**

$$y = u^v \ (u > 0),$$

如果 $u = u(x), v = v(x)$ 都可导, 则可把幂指函数表示为

$$y = \mathrm{e}^{v \ln u}.$$

这样, 便可直接求得

$$y' = \mathrm{e}^{v\ln n}\left(v'\cdot\ln u + v\cdot\frac{u'}{u}\right).$$

例 2.2.13 设函数 $y=\sqrt[3]{\dfrac{(x+1)^2(2-x)}{(3-x)^2(x-4)}}$, 求 y'.

证明. 在函数表达式两边取绝对值并取对数:

$$\ln|y| = \frac{1}{3}(2\ln|x+1| + \ln|2-x| - 2\ln|3-x| - \ln|x-4|).$$

再两边对 x 求导数, 即得

$$\frac{y'}{y} = \frac{1}{3}\left(\frac{2}{x+1} - \frac{1}{2-x} + \frac{2}{3-x} - \frac{1}{x-4}\right).$$

注意, 只有等式两边同时对同一个变量 (这里是 x) 求导数, 上式才能成立. 尤其是左侧, 不是对 y 求导, 而是对于 x 求导. 所有结果是 $\dfrac{y'}{y}$ 而不是 $\dfrac{1}{y}$. 注意, $y'=\dfrac{\mathrm{d}y}{\mathrm{d}x}$. 故有

$$y' = \frac{1}{3}\sqrt[3]{\frac{(x+1)^2(2-x)}{(3-x)^2(x-4)}}\left(\frac{2}{x+1} - \frac{1}{2-x} + \frac{2}{3-x} - \frac{1}{x-4}\right).$$

对于这个函数, 如果用通常办法求导数, 将十分麻烦. □

注意 为什么一开始要取绝对值再取对数呢? 具体来说: 1. 它保证了对数函数的有效性, 使我们能够对任何实数 (无论正负) 取对数, 避免了定义域的限制. 2. 在求导时, 绝对值符号会自然消失 (因为 $(\ln|x|)' = \dfrac{1}{x}$), 这不仅简化了计算, 还保持了结果的准确性.

对数求导法是一种非常有用的求导技巧, 特别适用于复杂的乘除和幂函数. 以下是对数求导法的一般步骤:

1. 给定函数 $y=f(x)$.
2. 两边取绝对值并取自然对数得到 $\ln|y| = \ln|f(x)|$.
3. 利用对数的性质展开右侧 $\ln|f(x)|$, 比如, 乘法变加法: $\ln(AB)=\ln A+\ln B$; 除法变减法: $\ln(A/B) = \ln A - \ln B$; 幂变系数: $\ln(A^n) = n\ln A$.
4. 两边对 x 求导得到 $\dfrac{y'}{y} = \dfrac{\mathrm{d}}{\mathrm{d}x}[\ln|f(x)|]$.
5. 解出 $y' = y\cdot\dfrac{\mathrm{d}}{\mathrm{d}x}[\ln|f(x)|]$.
6. 将原函数 $y=f(x)$ 代回得到最终结果.

2.3 无穷小量与微分

2.3.1 无穷小量

首先, 我们用一个统一的表述来融合函数的极限与数列的极限的定义. 下面定义的两个 "局部" 的理解, 请看视频 无穷小 (无穷小的定义，意义，极限与无穷小) ｜马同学图解微积分 _ 哔哩哔哩 _bilibili.

定义 2.3.1: 统一的极限定义

设函数 $f(x)$ 或数列 $\{a_n\}$ 在 "局部" 有定义. 如果对于 $\forall\epsilon > 0$, 都存在某个 "局部" 使得

$$|f(x) - L| < \epsilon \text{ 或 } |a_n - L| < \epsilon \, (L \in \mathbb{R}).$$

那么就称 L 是函数 $f(x)$ 或数列 $\{a_n\}$ 在此自变量变化过程中的 极限, 记作

$$\lim f(x) = L.$$

因为数列是一种特殊的函数, 所以为了标记的方便, 我们只采用了 $f(x)$ 的书写.

定义 2.3.2: 无穷小量

若 $\lim f(x) = 0$, 则称 $f(x)$ 为 **此自变量变化过程的无穷小量 (infinitesimal)**. 简言之, 所谓的无穷小量, 是指以零为极限的变量. 这里 "变量" 一词指的是函数或者数列.

根据以上的定义, 当说数列 $\{a_n\}$ 是无穷小量时, 就是说它的极限是 0. 因为数列的变量 n 只有趋向正无穷大这一种变化方式, 这非常简单. 但是, 当我们说函数 $f(x)$ 是无穷小量时, 我们首先要弄清楚自变量 x 的具体变化过程[9], 函数 $f(x)$ 究竟是在哪一种自变量 x 的变化过程中它的极限是 0. 例如, 对于函数 $\ln(1+x)$, 当 $x \to 0$ 时, 它是无穷小量; 当 $x \to 1$ 时, 它不是无穷小量; 当 $x \to +\infty$ 时, 它成了无穷大量. 再比如

- 当 $n \to \infty$ 时, 数列 $\dfrac{1}{n}$, q^n $(|q| < 1)$ 及 $\dfrac{1}{n!}$ 都是无穷小量.
- 当 $x \to 0$ 时, 函数 x, $\cos x - 1$ 都是无穷小量.
- 当 $x \to \infty$ 时, 函数 $\dfrac{1}{x}$ 是无穷小量.
- 当 $\Delta x \to 0$ 时, Δx 是无穷小量. (具体使用了什么符号不重要)
- 假定函数 $y = f(x)$ 在点 x_0 处连续, 则当 $\Delta x \to 0$ 时, $\Delta y = f(x_0 + \Delta x) - f(x_0)$ 是无穷小量.

从上面最后两个例子可知道: 假定函数 $y = f(x)$ 在点 x_0 处可导, 那么导数 $f'(x_0)$ 则是两个无穷小量[10] Δy 与 Δx 之比的极限. 像在这里, 我们会省略说明这个无穷小量是 "当 $\Delta x \to 0$ 时" 的无穷小量. 但请读者自己切记一定要搞清楚是哪一个自变量在变化, 以及它的变化趋势是什么.

注释 2.3.1 今后, 我们还会知道, 函数的积分可以看作无限多个无穷小量求和. 因此, 微积分在一定意义上就是对无穷小量进行分析, 微积分有时也称作无穷小分析.

[9] 之前我们总结过函数极限的 6 种不同类型 (即 $x \to a$, $x \to \infty$ 等等). 这里, 我们要求得更加具体. 比如, $x \to 0$ 和 $x \to 1$ 就是两种不同的自变量变化过程.

[10] 之前我们证明过, 函数 $y = f(x)$ 在点 x_0 处可导, 那么它也在该点处连续. 换句话说, 如果 Δy 与 Δx 之比的极限存在, 那么 $\Delta y = f(x_0 + \Delta x) - f(x_0)$ 是无穷小量.

微分的概念是建立在无穷小量概念的基础之上的. 为了讨论微分, 现在我们介绍有关无穷小量的几个概念. 下个命题直接来自于极限的性质.

命题 2.3.1

设 $f(x)$ 及 $g(x)$ 当 $x \to a$ 时是两个无穷小量. 那么, 当 $x \to a$ 时, $f(x) \pm g(x)$ 及 $f(x)g(x)$ 均是无穷小量.

但 $\dfrac{f(x)}{g(x)}$ 却不一定是无穷小量, 它有多种可能性.

定义 2.3.3

设 $f(x)$ 及 $g(x)$ 当 $x \to a$ 时是两个无穷小量. 若 $g(x) \neq 0$,[a] 且

$$\frac{f(x)}{g(x)} \to 1 \quad (x \to a),$$

则称 $f(x)$ 与 $g(x)$ 是 **等价无穷小量**, 记为 $f(x) \sim g(x)$ $(x \to a)$.

[a]指的是当 x 充分接近于 a 时, $g(x) \neq 0$ 总是成立的. 当然, $g(x)$ 作为无穷小量, 其极限肯定是 0.

显然, 根据定义我们有以下结果, 其中 $f(x)$, $g(x)$, $h(x)$ 都是当 $x \to a$ 时的无穷小量:

1. $f(x) \sim f(x)$ $(x \to a)$.
2. 如果 $f(x) \sim g(x)$ $(x \to a)$, 则 $g(x) \sim f(x)$ $(x \to a)$.
3. 如果 $f(x) \sim g(x)$ $(x \to a)$, $g(x) \sim h(x)$ $(x \to a)$, 则 $f(x) \sim h(x)$ $(x \to a)$.

这里以及后面涉及的概念如等价无穷小量, 高阶无穷小量, 同阶无穷小量等, 都只是用来描述两个无穷小量之间的关系. 如果在某个自变量变化的过程中, 函数 $f(x)$ 或 $g(x)$ 本身不是无穷小量, 那么这些讨论便无从谈起. 例如, 考虑

$$\lim_{x \to 0} \frac{x+2}{\mathrm{e}^x + 1} = 1.$$

我们显然不能说 $x+2$ 与 $\mathrm{e}^x + 1$ 是等价无穷小量, 因为它们根本不是无穷小量 (在 $x \to 0$ 时). 因此, 总结如下: **当我们讨论等价, 同阶或高阶无穷小量时, 前提条件是所有涉及的函数或数列在同一个自变量变化的过程中都必须各自是无穷小量, 否则讨论无从展开.**

例 2.3.1 $\sin x \sim x$ $(x \to 0)$.

例 2.3.2 $\ln(1+x) \sim x$ $(x \to 0)$ (见第一章第五节的例题).

所以, 我们又有 $\ln(1+x) \sim \sin x$ $(x \to 0)$.

例 2.3.3 $1 - \cos x \sim \dfrac{1}{2}x^2$ $(x \to 0)$.

证明. 首先确认当 $x \to 0$ 时, $1-\cos x$ 和 $\frac{1}{2}x^2$ 确实都是无穷小量. 然后, 证明下面极限就可以了.

$$\lim_{x\to 0}\frac{1-\cos(x)}{x^2}=\frac{1}{2}.$$

使用二倍角公式 $1-\cos(x)=2\sin^2\left(\frac{x}{2}\right)$, 代入得

$$\lim_{x\to 0}\frac{1-\cos(x)}{x^2}=\lim_{x\to 0}\frac{2\sin^2\left(\frac{x}{2}\right)}{x^2}=\lim_{x\to 0}2\cdot\frac{\sin^2\left(\frac{x}{2}\right)}{\left(\frac{x}{2}\right)^2}\cdot\frac{1}{4}.$$

利用 $\lim\limits_{y\to 0}\frac{\sin(y)}{y}=1$, 因此最终结果为

$$\lim_{x\to 0}\frac{1-\cos(x)}{x^2}=2\cdot 1\cdot\frac{1}{4}=\frac{1}{2}.$$

□

例 2.3.4 当 $x\to 0$, 有 $(1+x)^\alpha-1\sim\alpha x$, 其中 α 为常数.

证明. 证明下面极限就可以了.

$$\lim_{x\to 0}\frac{(1+x)^\alpha-1}{\alpha x}=1.$$

设 $u=(1+x)^\alpha-1$, 则 $(1+x)^\alpha=1+u$. 两边取对数, 有

$$\alpha\ln(1+x)=\ln(1+u).$$

因此, 原式可以改写为

$$\frac{(1+x)^\alpha-1}{\alpha x}=\frac{u}{\ln(1+u)}\cdot\frac{\ln(1+x)}{x}.$$

当 $x\to 0$ 时, $u\to 0$. 根据已知的极限 $\lim\limits_{x\to 0}\frac{\ln(1+x)}{x}=1$, 因此

$$\lim_{x\to 0}\frac{(1+x)^\alpha-1}{\alpha x}=1\times 1=1.$$

证毕. □

根据定义可知, 当 $f(x)\sim g(x)$ $(x\to a)$, 意味着, 当 $x\to a$ 时, 两个量 $f(x), g(x)$ 会不仅会接近于 0, 而且两者会非常地相等. 表 2.1 展示了上面例题中这些等价的无穷小量们会以相同的速度接近于 0. 但是不等价的无穷小量接近于 0 的速度是不同的.

一般来说, 无穷小量趋向于零的速度有可能不同. 比如, 当 $n\to\infty$ 时, $\frac{1}{n!}$ 比 $\frac{1}{n}$ 趋向于零的速度要快. 又比如, 由表 2.1 可知, 当 $x\to 0$ 时, $1-\cos x$ 比 $\ln(1+x)$ 趋向于零的速度要快.

一般来说, 在 $x\to a$ 的过程中, 当一个无穷小量 $\alpha(x)$ 比另一个无穷小量 $\beta(x)$ 趋向于零的速度要快时, 我们称 $\alpha(x)$ 是比 $\beta(x)$ 高阶的无穷小量. 更确切地说:

x	$\sin(x)$	$\ln(1+x)$	$1-\cos(x)$	$\frac{1}{2}x^2$
1.000000e-01	9.983342e-02	9.531018e-02	4.995835e-03	5.000000e-03
1.000000e-02	9.999833e-03	9.950331e-03	4.999958e-05	5.000000e-05
1.000000e-03	9.999998e-04	9.995003e-04	5.000000e-07	5.000000e-07
1.000000e-04	1.000000e-04	9.999500e-05	5.000000e-09	5.000000e-09
1.000000e-05	1.000000e-05	9.999950e-06	5.000000e-11	5.000000e-11
1.000000e-06	1.000000e-06	9.999995e-07	5.000445e-13	5.000000e-13

表 2.1: 无穷小量之间的速度对比. 生成代码见 codes/infinitesimal.ipynb · 赖志坚/高等数学 B-07 班-资料共享 - 码云 - 开源中国 (gitee.com)

定义 2.3.4

设 $\alpha(x)$ 及 $\beta(x)$ 当 $x \to a$ 时是两个无穷小量.

1. 当存在另一个无穷小量 $\eta(x)$ $(x \to a)$, 使得在 a 的某个空心邻域有:

$$\alpha(x) = \eta(x)\beta(x),$$

2. 或者等价地, 当 $\beta(x) \neq 0$ 时,[a]

$$\lim_{x \to a} \frac{\alpha(x)}{\beta(x)} = 0,$$

时, 我们称 $\alpha(x)$ 是比 $\beta(x)$ **高阶的无穷小量**, 记为

$$\alpha(x) = o(\beta(x)) \ (x \to a).$$

[a]指的是当 x 充分接近于 a 时, $\beta(x) \neq 0$ 总是成立的. 当然, $\beta(x)$ 作为无穷小量, 其极限肯定是 0.

例 2.3.5

1. $x \sin x^2 = o\left(x^2\right)$ $(x \to 0)$. 为此, 我们证明

$$\lim_{x \to 0} \frac{x \sin(x^2)}{x^2} = \lim_{x \to 0} x \cdot \frac{\sin(x^2)}{x^2} = 0 \cdot 1 = 0.$$

2. $\mathrm{e}^{-n} = o\left(\frac{1}{n}\right)$ $(n \to \infty)$. 为此, 我们证明

$$\lim_{n \to \infty} \frac{e^{-n}}{\frac{1}{n}} = \lim_{n \to \infty} \frac{n}{e^n} = 0.$$

符号 “o” 是分析学中常用的符号之一, 表达式 $\alpha(x) = o(\beta(x))$ $(x \to a)$ 的含义是存在一个函数 $\eta(x) \to 0$ $(x \to a)$, 使得

$$\alpha(x) = \eta(x)\beta(x),$$

因此单独地使用符号 $o(\beta(x))$ 则代表某一个无穷小量乘以 $\beta(x)$. 借助于符号 “o”, 可得到下面的性质.

命题 2.3.2

$f(x) \sim g(x)$ $(x \to a)$ 的充要条件是 $f(x) = g(x) + o(g(x))$ $(x \to a)$.

证明. 必要性. 设 $f(x) \sim g(x)$ $(x \to a)$, 则

$$\lim_{x \to a} \frac{f(x) - g(x)}{g(x)} = \lim_{x \to a} \left(\frac{f(x)}{g(x)} - 1 \right) = 0.$$

因此

$$f(x) - g(x) = o(g(x))\ (x \to a).$$

充分性. 设 $f(x) - g(x) = o(g(x))$ $(x \to a)$, 即

$$\lim_{x \to a} \frac{f(x) - g(x)}{g(x)} = 0.$$

那么可得出 $\lim\limits_{x \to a} \dfrac{f(x)}{g(x)} = 1$. 故 $f(x) \sim g(x)(x \to a)$. 证毕. □

因为 $f(x) \sim g(x)$ $(x \to a)$ 等价于 $g(x) \sim f(x)$ $(x \to a)$, 所以上述命题的中 f 和 g 完全可以交换位置. 下面引入同阶无穷小量的概念.

定义 2.3.5

设 $\alpha(x)$ 及 $\beta(x)$ 是当 $x \to a$ 时的两个无穷小量.

- 若有
$$\lim_{x \to a} \frac{\alpha(x)}{\beta(x)} = l, \quad l \neq 0,$$
则称 $\alpha(x)$ 与 $\beta(x)$ 为 **同阶无穷小量**. 显然, 等价无穷小是同阶无穷小的特殊情形, 即 $l = 1$ 的情形.

- 当 $x \to a$ 时, 若 $\alpha(x)$ 与 $(x-a)^n$ 为同阶无穷小量, 即,
$$\lim_{x \to a} \frac{\alpha(x)}{(x-a)^n} = l, \quad l \neq 0,$$
则称 $\alpha(x)$ 为 $x-a$ **的** n **阶无穷小量**. 比如, 当 $x \to 0$ 时, $\sin^3 x$ 是 x 的三阶无穷小量, 而 $1 - \cos x$ 是 x 的二阶无穷小量. 注意, 此时 $a = 0$.

例 2.3.6 当 $x \to 0$ 时, $\tan x - \sin x$ 是 x 的三阶无穷小量. 事实上, 我们有

$$\lim_{x \to 0} \frac{\tan x - \sin x}{x^3} = \lim_{x \to 0} \frac{\tan x}{x} \cdot \frac{1 - \cos x}{x^2} = \frac{1}{2}.$$

最后, 我们指出: 若 $x \to a$ 时 $f(x)$ 是一个无穷大量, 则 $x \to a$ 时 $\dfrac{1}{f(x)}$ 是一个无穷小量; 反之, 若 $x \to a$ 时 $\alpha(x)$ 是一个无穷小量, 并且 $\alpha(x) \neq 0$, 则 $x \to a$ 时 $\dfrac{1}{\alpha(x)}$ 是一个无穷大量.

命题 2.3.3

设 $\alpha(x), \tilde{\alpha}(x)$ 及 $\beta(x), \tilde{\beta}(x)$ 是当 $x \to a$ 时的无穷小量. 如果 $\alpha(x) \sim \tilde{\alpha}(x)$, $\beta(x) \sim \tilde{\beta}(x)$, 且 $\lim\limits_{x\to a} \dfrac{\tilde{\beta}(x)}{\tilde{\alpha}(x)}$ 存在, 那么

$$\lim_{x\to a} \frac{\beta(x)}{\alpha(x)} = \lim_{x\to a} \frac{\tilde{\beta}(x)}{\tilde{\alpha}(x)}.$$

证明. 利用极限运算的乘法法则, 直接有

$$\lim_{x\to a} \frac{\beta(x)}{\alpha(x)} = \lim_{x\to a} \left(\frac{\beta(x)}{\tilde{\beta}(x)} \cdot \frac{\tilde{\beta}(x)}{\tilde{\alpha}(x)} \cdot \frac{\tilde{\alpha}(x)}{\alpha(x)} \right) = \lim_{x\to a} \frac{\beta(x)}{\tilde{\beta}(x)} \cdot \lim_{x\to a} \frac{\tilde{\beta}(x)}{\tilde{\alpha}(x)} \cdot \lim_{x\to a} \frac{\tilde{\alpha}(x)}{\alpha(x)} = \lim_{x\to a} \frac{\tilde{\beta}(x)}{\tilde{\alpha}(x)}.$$

□

从上面的证明就能知道, **用等价无穷小量的性质去求极限的本质, 就是在背记和使用极限值为 1 的常用极限.** 所以切记不可以乱用等价无穷小量的替换. 它需要严格符合命题 2.3.3 的条件.

当 $x \to 0$ 时	等价无穷小关系
$\tan x$	$\sim x$
$\sin x$	$\sim x$
$\arcsin x$	$\sim x$
$\arctan x$	$\sim x$
$\ln(1+x)$	$\sim x$
$e^x - 1$	$\sim x$
$1 - \cos x$	$\sim \frac{1}{2}x^2$
$\sec x - 1$	$\sim \frac{1}{2}x^2$
$x - \ln(1+x)$	$\sim \frac{1}{2}x^2$
$\tan x - \sin x$	$\sim \frac{1}{2}x^3$
$\arcsin x - \arctan x$	$\sim \frac{1}{2}x^3$
$\tan x - x$	$\sim \frac{1}{3}x^3$
$x - \arctan x$	$\sim \frac{1}{3}x^3$
$x - \sin x$	$\sim \frac{1}{6}x^3$
$(1+x)^a - 1$	$\sim ax$
$a^x - 1$	$\sim \ln a \times x$

表 2.2: 常见的等价无穷小关系

例 2.3.7 求 $\lim\limits_{x\to 0} \dfrac{\tan 2x}{\sin 5x}$. 当 $x \to 0$ 时, $\tan 2x \sim 2x, \sin 5x \sim 5x$, 所以

$$\lim_{x\to 0} \frac{\tan 2x}{\sin 5x} = \lim_{x\to 0} \frac{2x}{5x} = \frac{2}{5}.$$

例 2.3.8 求 $\lim\limits_{x\to0}\dfrac{\sin x}{x^3+3x}$. 当 $x\to0$ 时, $\sin x\sim x$, 无穷小 x^3+3x 与它本身显然是等价的, 所以

$$\lim_{x\to0}\frac{\sin x}{x^3+3x}=\lim_{x\to0}\frac{x}{x\left(x^2+3\right)}=\lim_{x\to0}\frac{1}{x^2+3}=\frac{1}{3}.$$

例 2.3.9 求 $\lim\limits_{x\to0}\dfrac{\tan x-\sin x}{\sin^3 x}$.

$$\lim_{x\to0}\frac{\tan x-\sin x}{\sin^3 x}=\lim_{x\to0}\frac{\sec x-1}{\sin^2 x}=\lim_{x\to0}\frac{\frac{x^2}{2}}{x^2}=\frac{1}{2}.$$

注意, **在作等价无穷小的代换求极限时, 可以对分子或分母中的一个或若干个因子作代换, 但不能对分子或分母中的某个加项作代换.** 例如, 本题中若将分子中的 $\tan x$, $\sin x$ 均换成 x, 那么分子成为 0, 得出极限为 0, 这就导致错误的结果.

例 2.3.10 设 $\alpha(x)=o(x)(x\to a),\beta(x)=o(x)(x\to a)$, 证明 $\alpha(x)\pm\beta(x)=o(x)(x\to a)$. 上述结果有时可写成: $o(x)\pm o(x)=o(x)(x\to a)$.

证明. 根据定义, $\alpha(x)=o(x)\ (x\to a)$ 意味着存在无穷小量 $\eta_1(x)\ (x\to a)$, 使得 $\alpha(x)=\eta_1(x)x$, 且 $\eta_1(x)\to0\ (x\to a)$. 类似地, $\beta(x)=o(x)\ (x\to a)$ 表示存在无穷小量 $\eta_2(x)\ (x\to a)$, 使得 $\beta(x)=\eta_2(x)x$, 且 $\eta_2(x)\to0\ (x\to a)$. 于是,

$$\alpha(x)\pm\beta(x)=\eta_1(x)x\pm\eta_2(x)x=(\eta_1(x)\pm\eta_2(x))x.$$

定义 $\xi(x)=\eta_1(x)\pm\eta_2(x)$, 显然 $\xi(x)\to0\ (x\to a)$, 因此 $\xi(x)$ 是无穷小量, 从而

$$\alpha(x)\pm\beta(x)=\xi(x)x=o(x)\ (x\to a).$$

□

2.3.2 微分的定义

本小节的内容是我们将来学习多元函数微积分的基础. 它重点在于对微分概念的理解, 以及对于微分和导数关系的认知. 在计算上, 本小节没有任何新东西! 虽然对于应试影响不大, 但是对于微分概念的正确理解是未来很多内容的前提. 我们首先给出函数 $y=f(x)$ 在点 x_0 处可微及其微分的定义.

定义 2.3.6: 微分

设 $y=f(x)$ 在点 x_0 的某个邻域内有定义. 如果存在一个常数 A, 使得

$$f\left(x_0+\Delta x\right)-f\left(x_0\right)=A\Delta x+o(\Delta x)\quad(\Delta x\to0),$$

成立, 则称 $y=f(x)$ 在点 x_0 处 **可微**, 并把关于 Δx 的 **线性函数** $\Delta x\mapsto A\Delta x$ 称作 $y=f(x)$ 在点 x_0 处的 **微分**, 记作

$$\mathrm{d}f:=A\Delta x\ \text{ 或 }\mathrm{d}y:=A\Delta x. \tag{2.6}$$

若函数 $y=f(x)$ 在区间 (a,b) 中的每一点处都可微, 则称它在 (a,b) 内可微.

我们首先解释一下什么是 **线性函数 (linear function)**.

定义 2.3.7: 线性函数

线性函数是指满足以下两个特定性质的函数:

1. 加法性: 对于两个任意输入 x_1 和 x_2, 有
$$f(x_1+x_2)=f(x_1)+f(x_2).$$
2. 齐次性: 对于任意的输入 x 和任意的实数 c, 有
$$f(c\cdot x)=c\cdot f(x).$$

在以上描述中, 我们不指定函数 f 的定义域和到达域. 只要在这些定义域和到达域中, “加法” 和 “标量乘法[11]” 能够被良好定义, 无论它们是什么都可以.

例 2.3.11 以下运算都是线性函数, 其定义域都是函数的集合.

1. 给定某点 a, 考虑函数集合:
$$X:=\{f: \text{函数 } f \text{ 在 } a \text{ 的某个空心领域内有定义且 } \lim_{x\to a} f(x) \text{ 存在}\}.$$
则求极限运算
$$f\mapsto \lim_{x\to a} f(x)$$
是一个 X 到 $\mathbf{R}$ 的线性函数.
2. 给定某个区间 (a,b). 考虑函数集合:
$$X:=\{f: \text{ 函数 } f \text{ 定义在区间 } (a,b) \text{ 上且在该区间可导}\},$$
$$Y:=\{f: \text{ 函数 } f \text{ 定义在区间 } (a,b) \text{ 上}\}.$$
则求导运算
$$f\mapsto f'$$
是一个 X 到 Y 的线性函数.
3. 马上要学习的积分运算也是一种线性函数. 其他数学基础课《线性代数》《概率论与统计》中也到处都是线性函数.

如果我们只考虑形如 $f:\mathbf{R}\to\mathbf{R}$ 的线性函数, 则任何线性函数都可以被表示为
$$f(x)=ax,$$

[11]一般把实数和复数统称为标量 (scalar). 标量乘法不是指标量之间的乘法, 而是某个集合中的所有元素都可以乘以某个标量, 其结果还在这个集合中. 考虑标量为实数. (1) 当考虑实数集合 $\mathbf{R}$ 上的标量乘法, 那么它就是普通的乘法. (2) 考虑所有定义在 $\mathbf{R}\to\mathbf{R}$ 上的连续函数的集合, 如果它的一个元素是 $f(x)$, 那么对于任意实数 k, 函数 $kf(x)$ 也是连续函数, 所以它仍然在该集合中.

其中 a 是一个常数. 证明如下: 假设 $f:\mathbf{R}\to\mathbf{R}$ 是一个线性函数. 取 $x=1$, 则对于任意实数 c, 有

$$f(c)=f(c\cdot 1)=c\cdot f(1).$$

令 $a:=f(1)$, 那么我们得到 $f(c)=a\cdot c$. 因为这里的实数 c 是任意的, 所以用新的符号 x 去替换它, 即 $f(x)=ax$ 对于任意 $x\in\mathbf{R}$.

这意味着线性函数 $f:\mathbf{R}\to\mathbf{R}$ 的图像只能是一条过原点[12]的直线. 比如, $f(x)=2x$ 是一个线性函数. $f(x)=-3x$ 也是一个线性函数. 但要注意 $f(x)=-3x+1$ 不是一个线性函数. 我们发现图像为不经过原点的直线并不满足上面的加法性和齐次性条件 (读者务必自行验证), 因此不是严格的线性函数.

现在, 回到定义 2.3.6, 我知道了由常数 A 确定的 $\mathbf{R}\to\mathbf{R}$ 的函数

$$\Delta x\mapsto A\Delta x$$

确实是一个线性函数. 通常, 我们记作该线性函数为 $\mathrm{d}f=A\Delta x$ 或 $\mathrm{d}y=A\Delta x$. 所以一看到符号 $\mathrm{d}f$ 和 $\mathrm{d}y$, 不要把它们看所以静态的数量, 而是用看做动态的 (关于变量 Δx 的) 函数.

接下里, 我们应该正确理解定义 2.3.6 中的等式

$$\Delta y=f\left(x_0+\Delta x\right)-f\left(x_0\right)=A\Delta x+o(\Delta x)\quad(\Delta x\to 0)$$

的含义. 要牢记, 这不是一个普通意义上的等式 (比如, $2=\dfrac{4}{2}$, $\sin(x)=\cos(x+\dfrac{\pi}{2})$ 等). 上式是一个关于极限的**极限等式** (比如 $\lim\limits_{x\to 0}\sin(x)=0$ 等), 所以要把自变量的变化过程 $(\Delta x\to 0)$ 标记在最右侧以做提示. 该极限等式本质上想表达的是: 以下函数

$$(\text{ 整体看作关于 }\Delta x\text{ 的函数}:=\Delta y-A\Delta x=f\left(x_0+\Delta x\right)-f\left(x_0\right)-A\Delta x)$$

是 Δx 的高阶无穷小量, 即 $o(\Delta x)$. 若完整地展开, 它就是想说

$$\lim_{\Delta x\to 0}\frac{f\left(x_0+\Delta x\right)-f\left(x_0\right)-A\Delta x}{\Delta x}=0.$$

所以我们完全可以等价地改写定义 2.3.6 : 如果存在一个常数 A, 使得

$$\lim_{\Delta x\to 0}\frac{f\left(x_0+\Delta x\right)-f\left(x_0\right)-A\Delta x}{\Delta x}=0.$$

成立, 则称 $y=f(x)$ 在点 x_0 处可微, ... (其他一致) 显然, 这种表达方式不如使用符号 "o" 那么般简便.

定理 2.3.1: 单变量函数下微分和导数是等价的

考虑单变量函数 (也叫一元函数) $y=f(x)$, 则 $y=f(x)$ 在点 x_0 处可微的充分必要条件是 $y=f(x)$ 在点 x_0 处可导. 其中, $\mathrm{d}f=f'\left(x_0\right)\Delta x$.[a]

[a] 如果使用记号 $\mathrm{d}y=f'\left(x_0\right)\Delta x$, 显然, 根据微分的定义, 当 $\Delta x\to 0$ 时, Δy 与 $\mathrm{d}y$ 是等价无穷小, 因为有 $\Delta y=\mathrm{d}y+o(\Delta x)$.

证明. 1. 我们假定 $y=f(x)$ 在点 x_0 处可导. 这时, $y=f(x)$ 在点 x_0 处是连续的, 因而

$$\Delta y=f\left(x_0+\Delta x\right)-f\left(x_0\right)\to 0\quad(\Delta x\to 0).$$

[12] 我们还可以直接从加法性印证. 考虑 $x_1=0$ 和 $x_2=0$ 代入加法性条件得 $f(0+0)=f(0)+f(0)\implies f(0)=2f(0)$. 这意味着 $f(0)=0$.

可见, 当 $\Delta x \to 0$ 时, Δy 是一个无穷小量. 另外, 根据导数的存在性, 我们有

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = f'(x_0).$$

我们进一步考查函数

$$\eta(\Delta x) := \frac{\Delta y}{\Delta x} - f'(x_0).$$

这时, 当 $\Delta x \to 0$ 时, $\eta(\Delta x)$ 也是无穷小量. 这样, 我们得到

$$\Delta y = f'(x_0)\Delta x + \eta(\Delta x)\Delta x.$$

上式告诉我们一个重要事实: 当函数 $y = f(x)$ 在点 x_0 处可导时, 函数的增量

$$\Delta y = f(x_0 + \Delta x) - f(x_0)$$

可以分作两部分, 其中第一部分是一个常数 $f'(x_0)$ 乘以 Δx, 而第二部分是一个比 Δx 高阶的无穷小量[13]. 换句话说, 当 $y = f(x)$ 在点 x_0 处可导时, 函数的增量可以写成

$$\Delta y = f'(x_0)\Delta x + o(\Delta x) \quad (\Delta x \to 0).$$

因此, $y = f(x)$ 在点 x_0 处是可微的, 并且 $\mathrm{d}f = f'(x)\Delta x$.

2. 反过来, 假定存在某个常数 A, 使得

$$f(x_0 + \Delta x) - f(x_0) = A\Delta x + o(\Delta x) \quad (\Delta x \to 0),$$

成立. 两边除以 Δx 并令 $\Delta x \to 0$, 立即推出

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = A.$$

所以 $y = f(x)$ 在点 x_0 处可导, 且 $f'(x_0) = A$.

□

注意 这里我们要提醒读者: 虽然可微与可导是相互等价的, 但**它们从概念上讲是不同的**. 我们还要指出: **可微与可导的等价性也仅限于一元函数 (只含有一个自变量的函数) 的情形**. 在多元函数的情形中, (偏) 导数存在并不意味着可微 (相关内容见第六章).

现在我们再次解释一下定义 2.3.6 中微分概念的意义. 根据定义, 有

$$\Delta y = A\Delta x + o(\Delta x) \quad (\Delta x \to 0).$$

其中 $A\Delta x$ 是 $y = f(x)$ 在点 x_0 处的微分. 首先 $A\Delta x$ 是 Δx 的线性函数, 其次它与 Δy 之差是一个比 Δx 高阶的无穷小量. 因此, 我们说微分是函数增量的线性主要部分.

[13] 由此可见, 当 Δx 很小时, Δy 可以用 $f'(x_0)\Delta x$ 近似地代替.

结论 2.3.1: 微分和导数的关系

假定 $y=f(x)$ 在 (a,b) 内可微, 那么它在点 $x\in(a,b)$ 处的微分应当是线性函数

$$\Delta x\mapsto \mathrm{d}f=f'(x)\Delta x.$$

这时, 微分 $\mathrm{d}f$ 不仅依赖于 Δx, 而且还依赖于点 x, 因为由点 x 确定了该线性函数的系数. 理解上的要点:

- 函数 $y=f(x)$ 在某点处的微分是一个 (线性) 函数, 而不是像导数 $f'(x)$ 一样仅仅是某一个具体的确定数.
- 但一个具体的确定数 $a\in\mathbf{R}$ 可以决定一个 $\mathbf{R}\to\mathbf{R}$ 上的线性函数. 之前我们说过, 任何 $\mathbf{R}\to\mathbf{R}$ 上线性函数函数都可以被表示为
$$f(x)=ax.$$
从图像上可以知道, 不同的斜率 a (即不同的实数 a) 完全地决定了不同的线性函数. 简言之, 每一个实数 a 都和 $\mathbf{R}\to\mathbf{R}$ 上的每一个线性函数是一一对应. 即存在一个函数
$$\varphi:\mathbb{R}\to\{f:\mathbf{R}\to\mathbf{R}\mid f\text{ 是线性函数}\}$$
是双射.
- 函数 $y=f(x)$ 在某点处的微分, 即线性函数 $\Delta x\mapsto \mathrm{d}f=f'(x)\Delta x$ 正好就对应于实数 $f'(x)$. 两者知道其中一个, 就可以知道另外一个. 这也解释了为什么在定理 2.3.1 中说微分和导数是 "等价的".

现在, 请你去看视频: 什么是微分 | 马同学图解微积分 _ 哔哩哔哩 _bilibili

结论 2.3.2: $\mathrm{d}x\equiv\Delta x$

现在考查一个特殊函数 $y=x$. 这时 $\mathrm{d}y=y'\Delta x=\Delta x$, 即

$$\mathrm{d}x\equiv\Delta x.$$

这告诉我们: **对于自变量 x 而言, 它的微分与增量总相等.** 因此, 公式 $\mathrm{d}f=f'(x)\Delta x$ 可改写成下列形式:

$$\mathrm{d}f=f'(x)\mathrm{d}x$$

这种形式给我们带来许多方便, 故通常使用这种形式. 在公式 $\mathrm{d}f=f'(x)\mathrm{d}x$ 中两边用 $\mathrm{d}x$ 去除, 即得

$$\frac{\mathrm{d}f}{\mathrm{d}x}=f'(x).$$

这告诉我们: 函数在一点处的导数是其因变量的微分与自变量的微分之商. 也正是这个原因, 我们才把导数称为 **微商**.

在前面我们引入导数记号时, $\dfrac{\mathrm{d}f}{\mathrm{d}x}$ 是作为一个整体记号表示导数的. 现在有了微分的记号, 就可以把它看作两个微分之商了. 可见, 微商者, 微分之商也.

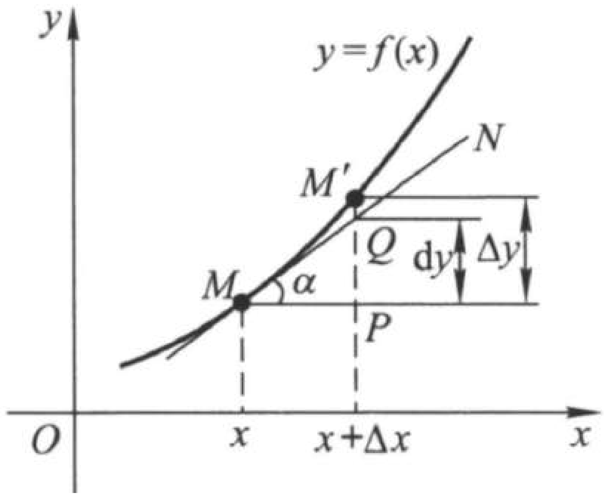


图 2.2: 微分的几何意义.

例 2.3.12 $\mathrm{d}\left(\sin^2 x\right) = 2\sin x\cos x\ \mathrm{d}x$.

例 2.3.13 $\mathrm{d}\left(x\mathrm{e}^x\right) = \mathrm{e}^x(1+x)\mathrm{d}x$.

微分 $\mathrm{d}y$ 的几何意义在图 2.2 中, M 是点 $(x, f(x))$, M' 是点 $(x+\Delta x, f(x+\Delta x))$, P 是点 $(x+\Delta x, f(x))$. 另外, Q 是曲线 $y = f(x)$ 在点 M 处的切线 MN 与直线 $M'P$ 的交点. 注意到切线 MN 的斜率是 $f'(x)$, 不难看到 $\mathrm{d}y = f'(x)\Delta x$ 恰好就是线段 QP 的长度. 以上说明, 切线上两点 (M 与 Q) 的纵坐标之差 $\mathrm{d}y$, 是曲线上相应两点 (M 与 M') 的纵坐标之差 Δy 的主要部分.

> **注释 2.3.2** 在力学, 物理学问题或其他实际问题的讨论中, 经常用 $\mathrm{d}y$ 去替代 Δy, 或者对两者不加区分, 就是因为当 $\Delta x \to 0$ 时, $\Delta y - \mathrm{d}y = o(\Delta x)$. 也就是说, 用 $\mathrm{d}y$ 去替代 Δy 所差的量是一个比 Δx 高阶的无穷小量. 这在实际问题的讨论中一般不会导致谬误.

定理 2.3.2

根据求导运算的规则, 我们可以得到下列求微分运算的规则:

1. $\mathrm{d}(f(x) \pm g(x)) = \mathrm{d}(f(x)) \pm \mathrm{d}(g(x))$;
2. $\mathrm{d}(f(x)g(x)) = g(x)\mathrm{d}(f(x)) + f(x)\mathrm{d}(g(x))$;
3. $\mathrm{d}\left(\dfrac{f(x)}{g(x)}\right) = \dfrac{g(x)\mathrm{d}(f(x)) - f(x)\mathrm{d}(g(x))}{g^2(x)}$.

证明. 现在我们以乘积的微分法则为例加以证明. 根据函数微分的表达式, 有

$$\mathrm{d}(fg) = (fg)'\mathrm{d}x.$$

再根据乘积的求导法则, 有

$$(fg)' = f'g + fg'.$$

于是

$$\mathrm{d}(fg) = (f'g + fg')\,\mathrm{d}x = f'g\ \mathrm{d}x + fg'\mathrm{d}x.$$

由于 $f'\mathrm{d}x = \mathrm{d}f, g'\mathrm{d}x = \mathrm{d}g$, 所以

$$\mathrm{d}(fg) = g\ \mathrm{d}f + f\mathrm{d}g.$$

其他法则都可以用类似方法证明. □

对于基本初等函数, 我们已经有了导数公式, 因此求初等函数的微分不会有任何困难.

例 2.3.14

$$\begin{aligned}
&\mathrm{d}\left(x - \mathrm{e}^x + \arctan x\right) = \mathrm{d}x - \mathrm{d}\left(\mathrm{e}^x\right) + \mathrm{d}(\arctan x) = \mathrm{d}x - \mathrm{e}^x\ \mathrm{d}x + \frac{1}{1+x^2}\ \mathrm{d}x \\
&= \left(1 - \mathrm{e}^x + \frac{1}{1+x^2}\right)\mathrm{d}x.
\end{aligned} \tag{2.7}$$

例 2.3.15

$$\mathrm{d}\left(\mathrm{e}^x \sin x\right) = \sin x\ \mathrm{d}\left(\mathrm{e}^x\right) + \mathrm{e}^x\ \mathrm{d}(\sin x) = \mathrm{e}^x \sin x\ \mathrm{d}x + \mathrm{e}^x \cos x\ \mathrm{d}x = \mathrm{e}^x(\sin x + \cos x)\mathrm{d}x. \tag{2.8}$$

题目做多了就要总结一些小技巧, 方便以后拿来就用. 比如, 所有形如 $y = \mathrm{e}^x f(x)$ 的函数的导数是

$$y' = \mathrm{e}^x\left(f(x)' + f(x)\right).$$

例 2.3.16

$$\begin{aligned}
\mathrm{d}\left(\frac{1+x}{1+x^2}\right) &= \frac{\left(1+x^2\right)\mathrm{d}(1+x) - (1+x)\mathrm{d}\left(1+x^2\right)}{\left(1+x^2\right)^2} \\
&= \frac{\left(1+x^2\right)\mathrm{d}x - (1+x)\cdot 2x\ \mathrm{d}x}{\left(1+x^2\right)^2} \\
&= \frac{1-2x-x^2}{\left(1+x^2\right)^2}\ \mathrm{d}x.
\end{aligned} \tag{2.9}$$

2.4 一阶微分的形式不变性及其应用

本节我们要介绍所谓的一阶微分的形式不变性, 并利用这种不变性给出隐函数及由参数方程所确定函数的求导公式.

2.4.1　一阶微分的形式不变性

一阶微分的形式不变性是在计算复合函数的导数时会用到的一个性质. 设函数 $y = f(x)$ 在区间 (a, b) 内可导, 并且其值域包含于区间 (A, B), 又设函数 $z = g(y)$ 在 (A, B) 内可导.

- 当 x 是自变量时, 复合函数 $z = g(f(x))$ 在 (a, b) 内可导, 且

$$(g(f(x)))' = g'(f(x))f'(x), \quad \forall x \in (a, b).$$

  这样, 我们就得到 $z = g(f(x))$ 的关于 x 的微分:

$$\mathrm{d}z = (g(f(x)))'\mathrm{d}x = g'(f(x))f'(x)\mathrm{d}x = g'(y)\mathrm{d}y,$$

  这里最后的等式, 我们使用了 $\mathrm{d}y = f'(x)\mathrm{d}x$ 和 $y = f(x)$. 注意, 此时, $\mathrm{d}x$ 是微分 $\mathrm{d}z$ 的变量. $\mathrm{d}y$ 是关于变量 $\mathrm{d}x$ 的线性函数, $g'(y) = g'(f(x))$ 是一个标量, 所以 $g'(y)$ 和 $\mathrm{d}y$ 的积 $\mathrm{d}z$ 仍旧是关于变量 $\mathrm{d}x$ 的线性函数.

- (此时不考虑函数的复合) 当 y 是自变量时, $z = g(y)$ 关于 y 的微分同样是

$$\mathrm{d}z = g'(y)\mathrm{d}y.$$

  注意, 此时, $\mathrm{d}y$ 本身是微分 $\mathrm{d}z$ 的变量, 即 $\mathrm{d}z$ 是关于变量 $\mathrm{d}y$ 的线性函数.

最终, 我们在不同的情况下得到**形式上**[14]完全相同的公式: 在前一种情况下, $\mathrm{d}z$ 是关于 x 的微分, 其中 y 不是自变量, 而是 x 的函数; 在后一种情况中, $\mathrm{d}z$ 是关于 y 的微分, y 是自变量.

结论 2.4.1

这就是说, 不论 y 是自变量, 还是中间变量, 当 $z = g(y)$ 时, 公式 $\mathrm{d}z = g'(y)\mathrm{d}y$ 总是成立的. 这就是所谓的 **一阶微分的形式不变性**[a].

[a]本书第六章, 我们将学习全微分 (即多变量/多元函数的微分) 的形式不变形.

一阶微分的形式不变性为微分的计算在形式上带来了方便. 它告诉我们: 在计算一阶微分时不必注意 $\mathrm{d}z$ 是关于哪个自变量的微分, 也不必注意 y 是自变量还是任何形式的中间变量, 公式 $\mathrm{d}z = g'(y)\mathrm{d}y$ 总成立. 比如我们有函数

$$z = \mathrm{e}^{\sin x^2}.$$

那么, 我们可以如下地去选择中间变量 y, 并微分:

1. 取 $y = \mathrm{e}^{\sin x^2}$, 则 $z = y$. 此时, $\mathrm{d}z = 1 \cdot \mathrm{d}y = \mathrm{d}\left(\mathrm{e}^{\sin x^2}\right)$.
2. 取 $y = \sin x^2$, 则 $z = \mathrm{e}^y$. 此时, $\mathrm{d}z = \mathrm{e}^y\mathrm{d}y = \mathrm{e}^{\sin x^2}\ \mathrm{d}\left(\sin x^2\right)$.
3. 取 $y = x^2$, 则 $z = \mathrm{e}^{\sin y}$. 此时, $\mathrm{d}z = \mathrm{e}^{\sin y}\cos y\ \mathrm{d}y = \mathrm{e}^{\sin x^2}\cos x^2\ \mathrm{d}\left(x^2\right)$.
4. 取 $y = x$, 则 $z = \mathrm{e}^{\sin y^2}$. 此时, $\mathrm{d}z = 2y\cos y^2\mathrm{e}^{\sin y^2}\ \mathrm{d}y = 2x\cos x^2\mathrm{e}^{\sin x^2}\ \mathrm{d}x$.

[14]注意, 我们分析了它们只是形式上相同, 但含义上不同. 如前者中 $\mathrm{d}y$ 是关于变量 $\mathrm{d}x$ 的线性函数, 而后者 $\mathrm{d}y$ 本身是一个变量而不是谁的函数.

通常, 我们把这个选择中间变量 y 的过程放在心中, 就直接写出下面一些列的关于微分的等式:

$$\mathrm{d}\left(\mathrm{e}^{\sin x^2}\right)=\mathrm{e}^{\sin x^2}\ \mathrm{d}\left(\sin x^2\right)=\mathrm{e}^{\sin x^2}\cos x^2\ \mathrm{d}\left(x^2\right)\ =2x\cos x^2\mathrm{e}^{\sin x^2}\ \mathrm{d}x.$$

其实, 这就是所谓的 “一阶微分的形式不变性”!!

显然, 所谓的 “一阶微分的形式不变性” 这一技巧, 并没有给复合函数的求导运算带来本质的改变. 无非就是逐步计算, 但对于复合函数求导来讲我们向来就是如此做的. 那么为什么要学它? 一个答案是, 这个技巧在之后计算 (不) 定积分时将非常非常重要. 具体来说, 比如上面的 $z=\mathrm{e}^{\sin x^2}$ 我们有四种不同的选择中间变量的方式, 这导致了不同样式的微分 $\mathrm{d}z$. 选择特定样式的微分对于使用 “分部积分法” (第三章会讲的重点内容) 去计算 (不) 定积分是最最关键的一步.

注意 这种不变性仅对一阶微分成立, 而对高阶微分不成立. 这一点在以后讨论高阶微分 (见小节 2.6.2) 时会看到.

例 2.4.1 求函数 $y=\mathrm{e}^{x^2+\sin^2 x+\sqrt{x}}$ 的微分.

证明. 方法一. 利用复合函数求导公式.

$$\mathrm{d}y=\left(\mathrm{e}^{x^2+\sin^2 x+\sqrt{x}}\right)'\mathrm{d}x=\mathrm{e}^{x^2+\sin^2 x+\sqrt{x}}\left(2x+2\sin x\cos x+\frac{1}{2\sqrt{x}}\right)\mathrm{d}x.$$

方法二. 利用一阶微分的形式不变性. 令 $u=x^2+\sin^2 x+\sqrt{x}$, 则 $y=\mathrm{e}^u$, 从而

$$\mathrm{d}y=\mathrm{e}^u\ \mathrm{d}u.$$

而 $\mathrm{d}u=\left(2x+2\sin x\cos x+\dfrac{1}{2\sqrt{x}}\right)\mathrm{d}x$, 于是

$$\mathrm{d}y=\mathrm{e}^{x^2+\sin x+\sqrt{x}}\left(2x+2\sin x\cos x+\frac{1}{2\sqrt{x}}\right)\mathrm{d}x.$$

练习足够之后, 就可以直接写成:

$$\begin{aligned}
dy&=e^{x^2+\sin^2 x+\sqrt{x}}\cdot\mathrm{d}\left(x^2+\sin^2 x+\sqrt{x}\right) &(2.10)\\
&=e^{x^2+\sin^2 x+\sqrt{x}}\cdot\left(\mathrm{d}\left(x^2\right)+\mathrm{d}\left(\sin^2 x\right)+\mathrm{d}(\sqrt{x})\right) &(2.11)\\
&=e^{x^2+\sin^2 x+\sqrt{x}}\cdot\left(2x\ \mathrm{d}x+2\sin x\cos x\ \mathrm{d}x+\frac{1}{2\sqrt{x}}\ \mathrm{d}x\right) &(2.12)\\
&=e^{x^2+\sin^2 x+\sqrt{x}}\cdot\left(2x+2\sin x\cos x+\frac{1}{2\sqrt{x}}\right)\ \mathrm{d}x. &(2.13)
\end{aligned}$$

上面使用了微分运算的法则, 见定理 2.3.2. □

一阶微分的形式不变性还可以用来求隐函数及由参数方程所确定函数的导数.

2.4.2 隐函数的导数

函数 $y=f(x)$ 表示两个变量 y 与 x 之间的对应关系, 这种对应关系可以用各种不同方式表达. 前面我们遇到的函数, 例如 $y=\sin x$, $y=\ln x+\sqrt{1-x^2}$ 等, 这种函数表达方式的特点是: 等号左端是因变量的符号, 而右端是含有自变量的式子, 当自变量取定义域内任一值时, 由这式子能确定对应的函数值. 用这种方式表达的函数叫做 **显函数**.

在某些问题中, 函数关系 $y=f(x)$ 并不是由显函数的表达式给出, 而是由 x 与 y 的一个方程给出的, 如

$$x^2+y^2=R^2 \quad (R>0);$$

或

$$y-x-\varepsilon\sin y=0 \quad (0<\varepsilon<1);$$

等. 这里第一个方程在 Oxy 平面上代表一个以原点为中心, R 为半径的圆周, 从这个方程中我们可以解得两个连续可微函数:

$$y=\sqrt{R^2-x^2}, \quad y=-\sqrt{R^2-x^2} \quad (-R\leqslant x\leqslant R).$$

(即单位圆的上半部和下半部) 这两个函数都称为由方程 $x^2+y^2=R^2$ 所确定的隐函数. 一般来说, 若函数 $y=f(x)$ 代入一个二元方程 $F(x,y)=0$ 时总使得 $F(x,f(x))\equiv 0$, 则称 $y=f(x)$ 是方程 $F(x,y)=0$ 所确定的 **隐函数 (implicit function)**. 圆周方程的例子说明, 一个方程 $F(x,y)=0$ 可能会含有多个不同的隐函数[15].

有时对于一个给定的方程 $F(x,y)=0$, 要想把 y 从中解出来并非易事, 比如在上面的方程 $y-x-\varepsilon\sin y=0$ 中, 可以证明: 对于任意给定的 x, 有唯一确定的 y 满足这个方程[16], 然而 y 却不能表示成 x 的初等函数[17]. 我们现在的问题是: *在不求解方程 $F(x,y)=0$ 的情况下 (即不用把这个隐函数给写成显函数), 如何计算由该方程所确定隐函数的导数 $\dfrac{\mathrm{d}y}{\mathrm{d}x}$?* 仍以方程

$$y-x-\varepsilon\sin y=0 \quad (0<\varepsilon<1)$$

为例. 对此方程两端求微分, 并根据一阶微分的形式不变性, 我们得到

$$\mathrm{d}y-\mathrm{d}x-\varepsilon\cos y\,\mathrm{d}y=0.$$

这里再次强调, 一阶微分形式不变性告诉我们: 在计算微分时不必注意是关于哪个自变量的微分. 所以从形式上, 我们保持了 $\mathrm{d}y$ 的形式, 因而最终上式中同时出现了 $\mathrm{d}x$ 和 $\mathrm{d}y$. 还记得 $\mathrm{d}x=\Delta x$, 是一个非零的形式上的变量. 通过变形, 我们得到

$$\frac{\mathrm{d}y}{\mathrm{d}x}-1-\varepsilon\cos y\frac{\mathrm{d}y}{\mathrm{d}x}=0,$$

[15]对于隐函数的认知, 目前只掌握这些就足够了, 将来我们会学习 "隐函数存在定理 (implicit function theorem)", 到那时我们会再严格审视这些概念.

[16]我们要证明, 对于给定的 x, 方程 $y-x-\varepsilon\sin y=0$ (其中 $0<\varepsilon<1$) 存在唯一的 y. 固定 x, 然后定义函数

$$g(y)=y-x-\varepsilon\sin y.$$

1. 存在性: $g(y)\to\pm\infty$ 当 $y\to\pm\infty$, 故由零点存在定理, $g(y)=0$ 至少有一个解.
2. 唯一性: 求导数 $g'(y)=1-\varepsilon\cos y$. 由于 $0<\varepsilon<1$ 且 $-1\le\cos y\le 1$, 因此 $g'(y)=1-\varepsilon\cos y>0$. 即, 函数 $g(y)$ 在整个定义域 $\mathbf{R}$ 上严格递增. 所以 $g(y)=0$ 的解是唯一的.

所以, 对于每一个 x, 通过解方程, 我们都能找到一个对应的 y. 这构成了一个隐藏着的函数关系.

[17]证明 y 不能表示成 x 的初等函数是严重超纲的. 但可以确认有很多方程 $F(x,y)=0$ 都不能简单地把隐函数给写出来.

即可解出

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{1}{1-\varepsilon\cos y}.$$

我们其实可以一步到位, 直接对方程两边关于 x 求导数, 最后也能得到相同的结果. 看下题.

例 2.4.2 求由方程 $\mathrm{e}^y+xy-\mathrm{e}=0$ 所确定的隐函数的导数 $\dfrac{\mathrm{d}y}{\mathrm{d}x}$.

证明. 我们把方程两边分别对 x 求导数, 注意 $y=y(x)$. 方程左边对 x 求导得

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\mathrm{e}^y+xy-\mathrm{e}\right)=\mathrm{e}^y\frac{\mathrm{d}y}{\mathrm{d}x}+y+x\frac{\mathrm{d}y}{\mathrm{d}x}.$$

方程右边对 x 求导得 $(0)'=0$. 由于等式两边对 x 的导数相等, 所以

$$\mathrm{e}^y\frac{\mathrm{d}y}{\mathrm{d}x}+y+x\frac{\mathrm{d}y}{\mathrm{d}x}=0,$$

从而

$$\frac{\mathrm{d}y}{\mathrm{d}x}=-\frac{y}{x+\mathrm{e}^y}\quad\left(x+\mathrm{e}^y\neq 0\right).$$

在这个结果中, 分式中的 $y=y(x)$ 是由方程 $\mathrm{e}^y+xy-\mathrm{e}=0$ 所确定的隐函数. □

有人会在计算上题之前就好奇方程 $\mathrm{e}^y+xy-\mathrm{e}=0$ 是否确定了一个隐函数呢? 即对于任意 x, 方程的解 y 是否唯一. 毕竟, 如何这个隐函数压根就不存在, 还求什么导数呢. 这一问题必须等到将来我们会学习 "隐函数存在定理 (implicit function theorem)" 才能解决. 至少本学期不用考虑这个事情, 我们只要求大家会计算隐函数的导数即可.

例 2.4.3 设 $y=f(x)$ 是由方程 $\mathrm{e}^{xy}+x^2y-1=0$ 所确定的隐函数, 求 y'.

证明. 对方程 $\mathrm{e}^{xy}+x^2y-1=0$ 两端关于 x 求导数, 其中注意 y 是 x 的函数, 于是利用复合函数求导公式得

$$\mathrm{e}^{xy}\left(y+xy'\right)+2xy+x^2y'=0.$$

解得

$$y'=-\frac{y\mathrm{e}^{xy}+2xy}{x\mathrm{e}^{xy}+x^2}.$$

□

注意: 在求隐函数的导数时, 上面的例子都出现了导数 $\dfrac{\mathrm{d}y}{\mathrm{d}x}$ 的最后表达式中有时还含有因变量 y. 这是允许且常见的, 因为 y 往往不能表示成 x 的一个明显的表达式. 本小节的总结如下.

结论 2.4.2

若函数 $y=f(x)$ 满足方程

$$F(x,f(x))\equiv 0,$$

则称 $y=f(x)$ 是方程 $F(x,y)=0$ 的隐函数. 求隐函数的导数时, 只要对上面的恒等式两边同时关于自变量 x 求导, 再变形即可. (此时必须把 y 看作是 x 的函数, 即使我们并不知道 $y=f(x)$ 的显式表达式!)

2.4.3 参数方程所确定的函数的导数

由参数方程所确定的函数是指这样的函数, 其中自变量与因变量的关系是由参数方程

$$\begin{cases} x = \varphi(t), \\ y = \psi(t), \end{cases} \quad \alpha \leqslant t \leqslant \beta,$$

所确定的, 这里的变量 t 称作参变量. 这样的函数有时也称作**参变量函数**.

- 比如, 对于参数方程
$$\begin{cases} x = R\cos\theta, \\ y = R\sin\theta, \end{cases} \quad 0 \leqslant \theta \leqslant 2\pi.$$
很容易看出, 它代表 Oxy 平面上的一个圆周, 其圆心在原点, 而半径为 R. 这里参变量 θ 有明显的几何意义, 它是原点到点 (x, y) 的连线与 x 轴正向的夹角.

- 如果将上述参数方程中的 R 略做更换, 就得到椭圆周的参数方程:
$$\begin{cases} x = a\cos\theta \\ y = b\sin\theta \end{cases} \quad 0 \leqslant \theta \leqslant 2\pi, 0 < b < a.$$
这时, a, b 分别为椭圆的长半轴与短半轴, 参变量 θ 的几何意义较之前复杂.

对于由参数方程所确定的函数, 同样可以不求出 y 关于 x 的表达式而直接求出其导数. 这不仅仅是因为一般求函数表达式较复杂, 更重要的原因是参数往往具有物理意义或几何意义, 我们希望所求的导数也由参数表示.

定理 2.4.1: 参数方程所确定的函数求导

设 $x = \varphi(t)$, $y = \psi(t)$ 在 $[\alpha, \beta]$ 上连续, 可导, 且 $\varphi'(t)$ 在 $[\alpha, \beta]$ 上连续, $\varphi'(t) \neq 0\ \forall x \in [\alpha, \beta]$. (这时 $\varphi(t)$ 必严格单调), 则参数式确定的函数 $y = \psi\left(\varphi^{-1}(x)\right)$ 可导, 且

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\psi'(t)}{\varphi'(t)},$$

或者

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\frac{\mathrm{d}y}{\mathrm{d}t}}{\frac{\mathrm{d}x}{\mathrm{d}t}}.$$

需要注意的是, 算出来的导数往往也是关于参数 t 的表达式, 这没有问题.

证明. 有两种证明方法.

1. 设函数 $y = f(x)$ 由参数方程
$$\begin{cases} x = \varphi(t), \\ y = \psi(t), \end{cases} \quad \alpha \leqslant t \leqslant \beta,$$
确定. 这时
$$\begin{cases} \mathrm{d}x = \varphi'(t)\mathrm{d}t, \\ \mathrm{d}y = \psi'(t)\mathrm{d}t. \end{cases}$$
用 $\mathrm{d}x$ 去除 $\mathrm{d}y$ 即得 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{\psi'(t)}{\varphi'(t)}$.

2. 也可以做如下解释: 当求 $\dfrac{\mathrm{d}y}{\mathrm{d}x}$ 时, 我们将 y 看作 x 的函数, 参数 t 是过渡性的中间变量, 即 y 是 t 的函数 $y=\psi(t)$, 而 t 是 x 的函数 ($x=\varphi(t)$ 的反函数). 根据一阶微分的形式不变性, 我们有 $\mathrm{d}y=\psi'(t)\mathrm{d}t$, 而根据反函数求导公式, 又有 $\mathrm{d}t=\dfrac{1}{\varphi'(t)}\cdot\mathrm{d}x$, 于是有上述结果.

$\square$

注意 上述定理的使用条件很多, 对吧. 不用担心, 直接记公式算就行了. 因为本小节和上一小节很不严谨, 所以勿要较真. 隐函数和参数方程的导数只要会计算就足够.

例 2.4.4 求出椭圆 $\begin{cases} x=a\cos\theta \\ y=b\sin\theta \end{cases}$ 在 $\theta=\dfrac{\pi}{4}$ 时的切线与 x 轴正向的夹角 φ.

证明. 根据前面的公式, 有

$$\frac{\mathrm{d}y}{\mathrm{d}x}=-\frac{b}{a}\cot\theta.$$

故 $\theta=\dfrac{\pi}{4}$ 时椭圆周的切线与 x 轴正向的夹角为

$$\varphi=\arctan\left(-\frac{b}{a}\right).$$

$\square$

例 2.4.5 弹道方程在不考虑空气阻力的情况下可以写作:

$$\begin{cases} x=v_0t\cos\alpha, \\ y=v_0t\sin\alpha-\dfrac{1}{2}gt^2. \end{cases}$$

其中 v_0 是炮弹的初始速度, α 为发射角, t 为发射后所经过的时间, g 为重力加速度. 试求炮弹在 t 时刻的运动方向与水平线的夹角 φ.

证明. 显然, 在这个问题中, 所求角的正切即 $\dfrac{\mathrm{d}y}{\mathrm{d}x}$, 故

$$\varphi=\arctan\frac{\mathrm{d}y}{\mathrm{d}x}=\arctan\frac{v_0\sin\alpha-gt}{v_0\cos\alpha}.$$

$\square$

例 2.4.6 设 $y=f(x)$ 是由参数方程 $\begin{cases} x=\ln\left(1+t^2\right) \\ y=t-\arctan t \end{cases}$ 所确定的函数, 求 $\dfrac{\mathrm{d}y}{\mathrm{d}x}$.

证明. $\dfrac{\mathrm{d}y}{\mathrm{d}x}=\dfrac{\frac{\mathrm{d}y}{\mathrm{d}t}}{\frac{\mathrm{d}x}{\mathrm{d}t}}=\dfrac{(t-\arctan t)'}{(\ln(1+t^2))'}=\dfrac{1-\frac{1}{1+t^2}}{\frac{2t}{1+t^2}}=\dfrac{t}{2}.$ $\square$

2.5 微分与近似计算

微分的概念可应用于函数值的近似计算. 假设函数 $y=f(x)$ 在点 x_0 处可微, 那么当 $\Delta x\to 0$ 时, 有

$$f\left(x_0+\Delta x\right)=f\left(x_0\right)+f'\left(x_0\right)\Delta x+o(\Delta x),$$

这里当 Δx 很小时, $o(\Delta x)$ 的值很小. 由此导出一个近似值公式:

$$f\left(x_0+\Delta x\right)\approx f\left(x_0\right)+f'\left(x_0\right)\Delta x.$$

因此, 若我们知道 $f\left(x_0\right)$ 及 $f'\left(x_0\right)$ 的值, 则可计算出 $f\left(x_0+\Delta x\right)$ 的近似值. 利用上述办法 (令 $x_0=0,\Delta x=x$), 可以得到一些常用的近似公式:

$$\mathrm{e}^x\approx 1+x,\quad \ln(1+x)\approx x,\quad (1+x)^\alpha\approx 1+\alpha x\quad (\alpha>0)$$

其中要求 x 的绝对值很小. 看得出来, 它们正是我们之前学习的等价无穷小量.

例 2.5.1 计算 $\sqrt{4.6}$ 和 $\sqrt{8.2}$ 的近似值.

证明. 设 $f(x)=\sqrt{x}$, 则 $f'(x)=\dfrac{1}{2\sqrt{x}}$. 先求 $\sqrt{4.6}$ 的近似值. 取 $x_0=4,\Delta x=0.6$, 这时有

$$\begin{aligned}f(4+0.6)&\approx f(4)+f'(4)\times 0.6,\\ \sqrt{4.6}&\approx\sqrt{4}+\frac{1}{2\sqrt{4}}\times 0.6=2.15.\end{aligned}$$

再求 $\sqrt{8.2}$ 的近似值. 这时, 我们取 $x_0=9,\Delta x=-0.8$, 得

$$\begin{aligned}f(9-0.8)&\approx f(9)+f'(9)\times(-0.8),\\ \sqrt{8.2}&\approx\sqrt{9}+\frac{1}{2\sqrt{9}}\times(-0.8)\approx 2.867.\end{aligned}$$

□

例 2.5.2 $\tan\left(\dfrac{\pi}{4}+0.01\right)$ 的近似值.

证明. 取 $f(x)=\tan x, x_0=\dfrac{\pi}{4},\Delta x=0.01$, 这时我们有

$$\tan\left(\frac{\pi}{4}+0.01\right)\approx\tan\frac{\pi}{4}+\sec^2\frac{\pi}{4}\times 0.01\approx 1+\frac{1}{(\sqrt{2}/2)^2}\times 0.01\approx 1.02.$$

□

在处理某些实际问题时, 如果精密度要求不高, 利用上述办法常常是有效的. 应当指出, 我们目前尚无法给出上述办法所获得的近似值与其精确值之间的误差估计, 以后会给出这种误差估计.

2.6 高阶导数与高阶微分

2.6.1 高阶导数

设函数 $y=f(x)$ 在一个区间 (a,b) 上有定义. 若 $y=f(x)$ 在每一点 $x\in(a,b)$ 处都有导数存在, 则其导数 $f'(x)$ 在 (a,b) 上也是一个函数. 自然, 我们可以考虑函数 $f'(x)$ 在点 $x_0\in(a,b)$ 处是否还有导数的问题.

定义 2.6.1

- 如果 $f'(x)$ 在点 x_0 处有导数, 则称此导数为 $y=f(x)$ 在点 x_0 处的 **二阶导数 (second order derivative)**, 记作 $f''(x_0), f^{(2)}(x_0), y^{(2)}\Big|_{x=x_0}, \left.\dfrac{\mathrm{d}^2y}{\mathrm{d}x^2}\right|_{x=x_0}$ 或 $\left.\dfrac{\mathrm{d}^2f}{\mathrm{d}x^2}\right|_{x=x_0}$.
- 类似地, 我们把 $y=f(x)$ 的二阶导数 $f''(x)$ 在点 x_0 处的导数定义为 **三阶导数 (third order derivative)** , 并记作 $f'''(x_0), f^{(3)}(x_0), y^{(3)}\Big|_{x=x_0}, \left.\dfrac{\mathrm{d}^3y}{\mathrm{d}x^3}\right|_{x=x_0}$ 或 $\left.\dfrac{\mathrm{d}^3f}{\mathrm{d}x^3}\right|_{x=x_0}$.
- 依次类推, 可以由 $n-1$ 阶导数定义 n 阶导数. $y=f(x)$ 在点 x_0 处的 n 阶导数记作 $f^{(n)}(x_0), y^{(n)}\Big|_{x=x_0}, \left.\dfrac{\mathrm{d}^ny}{\mathrm{d}x^n}\right|_{x=x_0}$ 或 $\left.\dfrac{\mathrm{d}^nf}{\mathrm{d}x^n}\right|_{x=x_0}$.
- 二阶及二阶以上的导数称为 **高阶导数 (higher-order derivative)**. 相应地, 也称 $f'(x)$ 为 **一阶导数 (first order derivative)**.

在物理学中, 若已知一个质点运动的位移函数为 $s=s(t)$, 那么该质点运动的速度为 $v(t)=s'(t)$, 而其加速度为 $a(t)=v'(t)=s''(t)$.

注意 二阶导数在几何上也有重要意义, 这要留待未来章节专门讲解.

例 2.6.1

$$
\begin{aligned}
(\sin x)' &= \cos x=\sin\left(x+\frac{\pi}{2}\right),\\
(\sin x)'' &= \cos\left(x+\frac{\pi}{2}\right)=\sin(x+\pi),\\
(\sin x)''' &= \cos(x+\pi)=\sin\left(x+\frac{3\pi}{2}\right).
\end{aligned}
\tag{2.14}
$$

一般地, 可以用数学归纳法证明 $\sin x$ 的 n 阶导数为

$$(\sin x)^{(n)}=\sin\left(x+\frac{n\pi}{2}\right).$$

类似地, 可以证明

$$(\cos x)^{(n)}=\cos\left(x+\frac{n\pi}{2}\right).$$

例 2.6.2 可以证明

$$\left(\frac{1}{x}\right)^{(n)}=(-1)^n\frac{n!}{x^{n+1}}.$$

事实上, $\left(\frac{1}{x}\right)' = -\frac{1}{x^2}, \left(\frac{1}{x}\right)'' = \frac{2!}{x^3}$. 若对于 $n = k-1$, 有

$$\left(\frac{1}{x}\right)^{(k-1)} = (-1)^{k-1}\frac{(k-1)!}{x^k},$$

那么对于 $n = k$, 将上式两边求导数, 即得

$$\left(\frac{1}{x}\right)^{(k)} = (-1)^k\frac{k!}{x^{k+1}}.$$

利用上面的结果立即又推出下面例子的结论.

例 2.6.3 $(\ln(1+x))^{(n)} = (-1)^{n-1}\frac{(n-1)!}{(1+x)^n}$, 其中约定 $0! = 1$.

例 2.6.4 求幂函数的 n 阶导数公式. 设 $y = x^\mu$ (μ 是任意常数), 那么

$$\begin{aligned}
&y' = \mu x^{\mu-1},\\
&y'' = \mu(\mu-1)x^{\mu-2},\\
&y''' = \mu(\mu-1)(\mu-2)x^{\mu-3},\\
&y^{(4)} = \mu(\mu-1)(\mu-2)(\mu-3)x^{\mu-4},
\end{aligned}$$

一般地, 可得

$$y^{(n)} = (x^\mu)^{(n)} = \mu(\mu-1)(\mu-2)\cdots(\mu-n+1)x^{\mu-n},$$

当 $\mu = n$ 时, 则得到

$$(x^n)^{(n)} = n(n-1)(n-2)\cdot\cdots\cdot 3\cdot 2\cdot 1 = n!,$$

而

$$(x^n)^{(n+k)} = 0 \quad (k = 1, 2, \cdots).$$

这说明, 当 μ 是正整数时, 其高阶导数最终会消失变为零. 但当 μ 不是正整数时, 其高阶导数并不会消失变为零. (为什么?)

如果函数 $u = u(x)$ 及 $v = v(x)$ 都在点 x 处具有 n 阶导数, 那么显然 $u(x) + v(x)$ 及 $u(x) - v(x)$ 也在点 x 处具有 n 阶导数, 且 $(u \pm v)^{(n)} = u^{(n)} \pm v^{(n)}$. 一般来说, 要求出函数积的 n 阶导数公式并非易事. 但仍旧有下面的规律.

命题 2.6.1: 函数积的高阶求导公式

设函数 $y = f(x)$ 及 $y = g(x)$ 在区间 (a, b) 内有 n 阶导数, 则它们之积的 n 阶导数满足下列公式:

$$(f(x)g(x))^{(n)} = \sum_{k=0}^{n} \mathrm{C}_n^k f^{(k)}(x) g^{(n-k)}(x),$$

其中 $\mathrm{C}_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$, $f^{(0)}(x) = f(x), g^{(0)}(x) = g(x)$.[a] 此公式称为莱布尼茨公式.

[a] 习惯上, 把零阶导数理解为函数本身.

这个公式很像牛顿二项展开式, 只不过是把牛顿二项展开式中方幂的次数换成导数的阶数而已. 当 $n=1$ 时, 上述公式是

$$(f(x)g(x))'=f'(x)g(x)+f(x)g'(x),$$

这是已知的公式. 对上式再一次求导数, 即得

$$(f(x)g(x))^{(2)}=f''(x)g(x)+2f'(x)g'(x)+f(x)g''(x)=\sum_{k=0}^{2}\mathrm{C}_2^k f^{(k)}(x)g^{(2-k)}(x).$$

对于一般情况, 可用数学归纳法证明, 请读者自己完成.

例 2.6.5 设函数 $y=x^2\sin x$, 求 y 的五阶导数.

证明. 注意到 $\left(x^2\right)^{(k)}\equiv 0\ (k>2)$, 由莱布尼茨公式得到

$$\begin{aligned}
y^{(5)}&=x^2(\sin x)^{(5)}+5\left(x^2\right)'(\sin x)^{(4)}+10\left(x^2\right)''(\sin x)^{(3)}\\
&=x^2\sin\left(\frac{\pi}{2}+x\right)+10x\sin x+20\sin\left(x+\frac{3\pi}{2}\right)\\
&=\left(x^2-20\right)\cos x+10x\sin x.
\end{aligned}$$

□

例 2.6.6 设函数 $y=\mathrm{e}^{\sin x}$, 求 $\dfrac{\mathrm{d}^2y}{\mathrm{d}x^2}$.

证明. 利用复合函数求导公式, 得 $\dfrac{\mathrm{d}y}{\mathrm{d}x}=\mathrm{e}^{\sin x}\cos x$. 然后有

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2}=\mathrm{e}^{\sin x}\cos x\cdot\cos x+\mathrm{e}^{\sin x}(-\sin x).$$

这里提醒读者: 要求复合函数的高阶导数, 需重复使用求复合函数一阶导数的方法, 并没有一般公式. □

下面通过例题来说明求隐函数和由参数方程所确定函数的二阶导数的方法. 用同样的方法可求得三阶及三阶以上的导数.

例 2.6.7 设 $y=y(x)$ 是由方程 $x^3+y^3-3xy=0$ 所确定的隐函数, 求 y''.

证明.

1. **方法一: 按本章第四节的方法算出一阶导数, 再按通常办法再次求导.** 方程 $x^3+y^3-3xy=0$ 两端关于 x 求导数, 其中注意 y 是 x 的函数, 得到

$$3x^2+3y^2y'-3y-3xy'=0,$$

解得

$$y'=\frac{x^2-y}{x-y^2}.$$

再对它关于 x 求导数, 其中注意 y 是 x 的函数, 得到

$$y''=\frac{(2x-y')\left(x-y^2\right)-\left(x^2-y\right)(1-2yy')}{\left(x-y^2\right)^2}.$$

将 $y'=\dfrac{x^2-y}{x-y^2}$ 代人上式, 就得到 y'' 的表达式. 大家会觉得 y'' 的表达式过于复杂. 但此表达式中的 x,y 不是没有关系的, 它们满足方程 $x^3+y^3-3xy=0$. 利用此方程, 可适当化简 y'' 的表达式, 最后得到

$$y''=\frac{2xy}{\left(x-y^2\right)^3}.$$

2. **方法二: (比较推荐) 两边分别依次做两次求导运算.** 当我们得到

$$3x^2+3y^2y'-3y-3xy'=0,$$

对上式两端关于 x 求导数, 其中注意 y,y' 是 x 的函数, 得到

$$6x+6yy'\cdot y'+3y^2y''-3y'-3y'-3xy''=0,$$

解得

$$y''=\frac{2x+2yy'\cdot y'-2y'}{x-y^2}.$$

将 $y'=\dfrac{x^2-y}{x-y^2}$ 代人上式, 并利用方程 $x^3+y^3-3xy=0$ 化简, 得到

$$y''=\frac{2xy}{\left(x-y^2\right)^3}.$$

□

例 2.6.8 设函数 $y=f(x)$ 是由参数方程 $\begin{cases} x=t-\sin t\\ y=1-\cos t\end{cases}$ 所确定, 求 $\dfrac{\mathrm{d}^2y}{\mathrm{d}x^2}$.

证明. 使用定理 2.4.1, 有 $\dfrac{\mathrm{d}y}{\mathrm{d}x}=\dfrac{\frac{\mathrm{d}y}{\mathrm{d}t}}{\frac{\mathrm{d}x}{\mathrm{d}t}}=\dfrac{\sin t}{1-\cos t}$. 参数方程所确定函数 $y=f(x)$ 的导数 $\dfrac{\mathrm{d}y}{\mathrm{d}x}$ 的表达式中含有 t, 而 $\dfrac{\mathrm{d}y}{\mathrm{d}x}$ 仍是 x 的函数, 也要由下面这组新的参数方程

$$\begin{cases} x=t-\sin t,\\ \dfrac{\mathrm{d}y}{\mathrm{d}x}=\dfrac{\sin t}{1-\cos t},\end{cases}$$

给出, 因此再次使用定理 2.4.1, 有

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2}=\frac{\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)}{\frac{\mathrm{d}x}{\mathrm{d}t}}=\frac{\frac{\cos t(1-\cos t)-\sin t\sin t}{(1-\cos t)^2}}{1-\cos t}=\frac{\cos t-1}{(1-\cos t)^3}=-\frac{1}{(1-\cos t)^2}.$$

□

2.6.2 高阶微分

现在我们来讲高阶微分, 即二阶及二阶以上的微分.

二阶微分 假定 $y=f(x)$ 有 n 阶导数, 那么它的微分

$$\mathrm{d}f=f'(x)\mathrm{d}x$$

依赖于两个变量: x 与 $\mathrm{d}x$. 我们把 $\mathrm{d}x$ 固定, 则 $\mathrm{d}f$ 就是 x 的函数 (其实就是导函数 $f'(x)$ 多乘上了一个常量 $\mathrm{d}x$). 因此, 又可以对它关于变量 x 求微分, 这就是 **二阶微分**, 记作 d^2f:

$$\mathrm{d}^2f:=\mathrm{d}(\mathrm{d}f)=(f'(x)\mathrm{d}x)'\,\mathrm{d}x=f''(x)\mathrm{d}x^2.$$

这里应注意, 在求二阶微分时, $\mathrm{d}x$ 作为常量. 而记号 $\mathrm{d}x^2$ 表示 $(\mathrm{d}x)^2$.[18] 就像 (一阶) 微分 $\mathrm{d}f$ 视作一次函数[19] $\mathrm{d}x\mapsto f'(x)\mathrm{d}x$. 类似地, 我们固定 x, 把二阶微分 d^2f 视作二次函数

$$\mathrm{d}x\mapsto f''(x)\mathrm{d}x^2.$$

而且, 我们有

$$\frac{\mathrm{d}^2f}{\mathrm{d}x^2}=f''(x).$$

也就是说, 函数的二阶导数是它的二阶微分除以自变量微分的平方. 前面将二阶导数记作 $\dfrac{\mathrm{d}^2f}{\mathrm{d}x^2}$, 也就是这个原因.

n 阶微分 依次类推, 可以得到 **n 阶微分**, 记作 d^nf:

$$\mathrm{d}^nf:=f^{(n)}(x)(\mathrm{d}x)^n=f^{(n)}(x)\mathrm{d}x^n,$$

其中 $\mathrm{d}x^n\equiv(\mathrm{d}x)^n$.[20] 并且有

$$\frac{\mathrm{d}^nf}{\mathrm{d}x^n}=f^{(n)}(x).$$

总之, 函数 $y=f(x)$ 的在某点处的 n 阶微分就等于它对应的 n 阶导数乘以变量 $\mathrm{d}x$ 的 n 次幂, 即 $\mathrm{d}x^n\equiv(\mathrm{d}x)^n$. 注意, n **阶微分仍旧是一个关于变量 $\mathrm{d}x$ 的函数, 而不是一个确定的具体数.** 显然, 当 $n=1$ 时, 一阶微分是线性函数; 当 $n>1$ 时, 它是非线性函数, 而且是简单的 n 阶多项式函数.

高阶微分不具有形式不变性 这里应当提醒读者, 高阶微分不具有形式不变性. 也就是说,

- 当 y 是自变量时, $z=g(y)$ 的二阶微分是

$$\mathrm{d}^2z=g''(y)\mathrm{d}y^2,$$

- 当 y 是 x 的函数, 即 $y=f(x)$ 时, 上述形式不再成立. 事实上, 这时 $z=g(f(x))$, 并且有

$$\mathrm{d}z=g'(f(x))f'(x)\mathrm{d}x,$$

  然后根据二阶微分的定义得 (将 $\mathrm{d}x$ 作为常量, 对它关于变量 x 求微分):

$$\begin{aligned}\mathrm{d}^2z&=\mathrm{d}(g'(f(x))f'(x)\mathrm{d}x)\\&=\mathrm{d}(g'(f(x)))\cdot f'(x)\mathrm{d}x+g'(f(x))\cdot\mathrm{d}(f'(x))\mathrm{d}x\\&=(g''(f(x))\cdot f'(x)\mathrm{d}x)\cdot f'(x)\mathrm{d}x+g'(f(x))\cdot(f''(x)\mathrm{d}x)\mathrm{d}x\\&=g''(f(x))\cdot(f'(x))^2\mathrm{d}x^2+g'(f(x))\cdot f''(x)\mathrm{d}x^2\\&=g''(y)\mathrm{d}y^2+g'(y)\mathrm{d}^2y,\end{aligned}$$

[18] 微分符号 d 的书写习惯要良好. 比如, $\mathrm{d}(x^2)$ 是函数 $y=x^2$ 的微分. 而不能简单写作 $\mathrm{d}x^2$. $\mathrm{d}x^2$ 表示 $(\mathrm{d}x)^2$.

[19] 有时线性函数也叫一次函数.

[20] 类似地, 我们也有 $\mathrm{d}y^n\equiv(\mathrm{d}y)^n$.

上式的最后, 我们用 y 替换 $f(x)$, 并注意到 $\mathrm{d}y = f'(x)\mathrm{d}x$ 和 $\mathrm{d}^2y = f''(x)\mathrm{d}x^2$. 此时, 我们发现这里恰好多出来了一项 $g'(y)\mathrm{d}^2y$. 只有当 y 是 x 的线性函数时, $\mathrm{d}^2y \equiv 0$, 但其他情况下 d^2y 不恒为零.

同样地, 我们可以得到复合函数的三阶微分表达式:

$$\mathrm{d}^3z = g'''(y)\ \mathrm{d}y^3 + 3g''(y)\ \mathrm{d}y\ \mathrm{d}^2y + g'(y)\ \mathrm{d}^3y.$$

从上述结果可以看出, 高阶微分 $\mathrm{d}^2z, \mathrm{d}^3z, \ldots, \mathrm{d}^nz$ 通常是不具有不变性的.

下面的例题中, 我们从高阶微分的定义出发去计算他.

例 2.6.9 设函数 $y = \mathrm{e}^x \cos 2x$, 求二阶微分 d^2y.

证明. 先计算 $\mathrm{d}y = (\mathrm{e}^x \cos 2x - 2\mathrm{e}^x \sin 2x)\,\mathrm{d}x$, 然后计算

$$\begin{aligned}\mathrm{d}^2y &= (\mathrm{e}^x \cos 2x - 2\mathrm{e}^x \sin 2x - 2\mathrm{e}^x \sin 2x - 4\mathrm{e}^x \cos 2x)\,\mathrm{d}x^2 \\ &= (-3\mathrm{e}^x \cos 2x - 4\mathrm{e}^x \sin 2x)\,\mathrm{d}x^2.\end{aligned}$$

□

下面的例题告诉我们, **求 n 阶微分就是求 n 阶导数. 记住这一条就足够了.**

例 2.6.10 求函数 $y = x^4$ 的四阶微分 d^4y.

证明. 首先计算四阶导数:

$$\begin{aligned} y' &= \left(x^4\right)' = 4x^3, \\ y'' &= \left(4x^3\right)' = 4 \cdot 3x^2 = 12x^2, \\ y''' &= \left(12x^2\right)' = 12 \cdot 2x = 24x, \\ y^{(4)} &= (24x)' = 24, \end{aligned}$$

则微分 d^4y 为

$$\mathrm{d}^4y = 24\ \mathrm{d}x^4.$$

□

2.7 不定积分

前面讲述了导数及微分的概念, 现在我们开始讲述微积分中的另一个基本概念 — 积分. 积分有两种: 不定积分与定积分. 本节先讲不定积分.

首先我们引入原函数的概念. 它是导函数的一种 "逆" 的概念.

定义 2.7.1: 原函数

给定一个在区间 (a, b) 上函数 $f(x)$. 若在同一区间上存在另一个函数 $F(x)$ 使得

$$F'(x) = f(x),\ \forall (a, b),$$

则称 $F(x)$ 是 $f(x)$ 在区间 (a, b) 上的一个 **原函数**, 或者叫 **反导函数 (antiderivative)**.

如果我们把求导数看作一种运算 (称为求导运算) 的话, 那么求不定积分则是求导运算的逆运算.

- 求导运算是求出已知函数 $f(x)$ 的导数 $f'(x)$. 求导运算的输入是一个函数, 输出则是另一个函数, 输出函数是输入函数的导函数.

  $$\text{求导运算}: f \mapsto f'.$$

  至此, 我们一直没有讨论过这么一个问题: 给定可导函数 $f(x)$, 那么它的导函数 $f'(x)$ 是唯一吗? 答案是肯定的, 因为函数的定义式来自极限 $f'(x) = \lim\limits_{\Delta x\to 0} \dfrac{f(x+\Delta x)-f(x)}{\Delta x}$, 其中 x 是 (a,b) 内任意的固定点, 我们知道极限如果存在那么极限值是唯一的. 故我们可以把求导运算自身当作一个合法定义的函数 (function) 或者算子.

- 反过来, 已知函数 $f(x)$, 要求出一个函数 $F(x)$, 使得它的导数恰好就是 $f(x)$, 即 $F'(x) = f(x)$, 并且我们还希望求出所有这样的函数 $F(x)$. 即, 所谓的 "不定积分运算" 的输入是一个函数, 输出则是一个函数簇 (即某个以函数为元素的集合, 简称函数簇或函数集合), 输出的函数簇里的每一个函数的导函数都是同一个输入函数 $f(x)$, 也就是说, 输出的函数簇是 $f(x)$ 的所有原函数.

  $$\text{不定积分运算}: f \mapsto \{f \text{ 的所有原函数}\}.$$

  为什么不定积分运算的输出不是某一个具体的原函数, 而是所有原函数的集合呢? 从函数 (function) 的定义出发, 我们发现前者是一对多, 不合法; 后者是一对一, 合法.

为了准确定义不定积分, 我们首先要观察原函数的基本性质. 下面的命题说明: 任何一个函数的原函数都不止一个, 而是有无穷多个. 并且任意两个原函数之间只相差一个常数. 它们之间的关系在图像上看就只是在 Oxy 平面上沿着竖直方向平移而已.

命题 2.7.1

若 $F(x)$ 是 $f(x)$ 在 (a,b) 上的一个原函数, 则对于任意常数 C, 函数 $F(x)+C$ 也是 $f(x)$ 在 (a,b) 上的原函数. 反过来, $f(x)$ 在 (a,b) 上的任意一个原函数都可表示成

$$F(x)+C$$

的形式.

证明. 事实上, 若 $G(x)$ 是 $f(x)$ 在 (a,b) 上的另外一个原函数, 考虑 $G(x)-F(x)$, 它在区间 (a,b) 上的导数为 $(G(x)-F(x))' = G'(x)-F'(x) = f(x)-f(x) \equiv 0$. 前面我们曾指出: 一个区间内导数恒等于零的函数必为常数函数. 因此, $G(x)-F(x) \equiv C$, 其中 C 为常数, 即 $G(x) = F(x)+C$. □

根据以上命题, 若 $F(x)$ 是 $f(x)$ 的一个原函数, 那么

$$\{f \text{ 的所有原函数}\} = \{F(x)+C \mid C \text{ 为任意常数}\}.$$

需要指出, 原则上我们不用特定 $F(x)$ 具体是哪一个原函数. 若 $G(x)$ 是 $f(x)$ 的另外一个原函数, 那么原函数簇可以完全等价地写成

$$\{f \text{ 的所有原函数}\} = \{G(x)+C \mid C \text{ 为任意常数}\}.$$

定义 2.7.2: 不定积分

对于一个给定的函数 $y = f(x)$, 其原函数的一般表达式称为 $f(x)$ 的 **不定积分 (indefinite integral)**, 并记作

$$\int f(x)\ \mathrm{d}x.$$

这里函数 $f(x)$ 称作 **被积函数**, $f(x)\ \mathrm{d}x$ 称为 **被积表达式**. 若已知 $F(x)$ 是 $f(x)$ 的一个原函数, 那么 $F(x)+ C$ (C 为任意常数) 便是 $f(x)$ 的原函数的一般表达式, 即

$$\int f(x)\ \mathrm{d}x := F(x) + C.$$

这里 C 叫做 **积分常数**.

例如, $\sin x$ 是 $\cos x$ 的一个原函数, 因此

$$\int \cos x\ \mathrm{d}x = \sin x + C.$$

又比如, $\frac{1}{2}x^2$ 是 x 的一个原函数, 所以

$$\int x\ \mathrm{d}x = \frac{1}{2}x^2 + C.$$

我们具体探讨一下积分常数 C 的意义. 显然, $\sin x + 1$ 也是 $\cos x$ 的一个原函数, 因此根据定义我还可以写出

$$\int \cos x\ \mathrm{d}x = \sin x + 1 + C.$$

上式原则上是没有问题的, 但是我们通常不会这么做. 一来是, 形式上不如 $\sin x + C$ 那么干净; 二来是, 符号 C 表示的不是一个确定的常数, 而是一个可以取得任意值的符号, 则**任意取值的符号 C 能够 "吞掉" 任何形式的其他常数, 或其他任意取值的量**. 从这个意义上, 我们认为下面这些都是相同的

$$C = 1 + C = 2 + C = \pi + C = C_1 + C_2 = 2C = \frac{C}{3}.$$

是的, 我们应当从集合的视角去理解这个符号 C. 就像我们一开始讨论过的, 若 $F(x)$ 和 $G(x)$ 都是 $f(x)$ 的原函数, 那么集合 $\{f$ 的所有原函数$\}$ 在形式上可以写成不同的集合表达式, 但他们的内涵实质却是一样的:

$$\{f \text{ 的所有原函数}\} = \{F(x) + C \mid C \text{ 为任意常数}\} = \{G(x) + 2C \mid C \text{ 为任意常数}\}.$$

因此, 我们可以这么理解不定积分:

$$\int f(x)\ \mathrm{d}x = \{f \text{ 的所有原函数}\}.$$

然后, 对于右侧的函数簇的一般表达式, 我们选择一个较为简便的 $F(x)$, 然后加一个尾巴 C. 不定积分的所谓 "不定" 就在于这个 C. 忽略这个 C 而写成 $\int f(x)\ \mathrm{d}x = F(x)$ 是错误的.

求不定积分的问题在微积分中是很重要的. 以后我们会看到定积分的计算要归结为求被积函数的原函数, 所以**不定积分的计算是定积分计算的基础.** 不仅如此, 不定积分的计算也是求解常微分方程的基础. 因此, 初学者应当通过做相当数量的题目, 掌握好求不定积分的基本方法, 并熟练运用不定积分的基本公式. 为了求不定积分, 首先要熟记一些重要初等函数的不定积分. 我们将它们列成下表 (称为基本积分表, 这张表是由基本初等函数的导数表反转而成的):

结论 2.7.1: 基本积分表

1. $\int k\ \mathrm{d}x = kx + C$. 特别地, $\int 1\ \mathrm{d}x = x + C$.
2. $\int x^\alpha\ \mathrm{d}x = \frac{1}{\alpha+1}x^{\alpha+1} + C \quad (\alpha \neq -1)$
3. $\int \cos x\ \mathrm{d}x = \sin x + C$
4. $\int \sin x\ \mathrm{d}x = -\cos x + C$
5. $\int \frac{\mathrm{d}x}{\cos^2 x} = \int \sec^2 x\ \mathrm{d}x = \tan x + C$
6. $\int \frac{\mathrm{d}x}{\sin^2 x} = \int \csc^2 x\ \mathrm{d}x = -\cot x + C$
7. $\int \sec x \tan x\ \mathrm{d}x = \sec x + C$
8. $\int \csc x \cot x\ \mathrm{d}x = \csc x + C$
9. $\int \frac{\mathrm{d}x}{\sqrt{1-x^2}} = \arcsin x + C = -\arccos x + C$
10. $\int \frac{\mathrm{d}x}{1+x^2} = \arctan x + C = -\operatorname{arccot} x + C$
11. $\int a^x\ \mathrm{d}x = \frac{1}{\ln a}a^x + C\ (a>0, a\neq 1)$. 特别地, $\int \mathrm{e}^x\ \mathrm{d}x = \mathrm{e}^x + C$.
12. $\int \frac{1}{x}\ \mathrm{d}x = \ln|x| + C$

验证这张表中每个公式的正确性是十分容易的, 只要将公式右端的函数求导数, 看看所得结果是否等于不定积分中的被积函数即可. 除去上述基本积分表之外, 读者还应当掌握基本的求不定积分的法则. 目前我们要求读者掌握下列法则:

定理 2.7.1: 不定积分运算是线性函数

不定积分运算

$$f(x) \mapsto \int f(x)\ \mathrm{d}x$$

是一个线性函数:

1. $\int (f(x) \pm g(x))\ \mathrm{d}x = \int f(x)\ \mathrm{d}x \pm \int g(x)\ \mathrm{d}x$.
2. $\int cf(x)\ \mathrm{d}x = c\int f(x)\ \mathrm{d}x$, 其中 c 为任意常数.

证明. 这两条法则的证明是容易的. 事实上, 一方面, 若 $F(x)$ 及 $G(x)$ 分别为 $f(x)$ 及 $g(x)$ 的一个原函数, 那

么

$$\int f(x)\ \mathrm{d}x = F(x) + C_1,$$

$$\int g(x)\ \mathrm{d}x = G(x) + C_2,$$

其中 C_1 与 C_2 为任意常数. 因此, 我们有

$$\int f(x)\ \mathrm{d}x \pm \int g(x)\ \mathrm{d}x = F(x) \pm G(x) + C.$$

其中 $C = C_1 \pm C_2$. 显然, C 也是一个任意常数. 另一方面,

$$(F(x) \pm G(x))' = f(x) \pm g(x).$$

可见, $F(x) \pm G(x)$ 是 $f(x) \pm g(x)$ 的一个原函数, 因而

$$\int (f(x) \pm g(x))\ \mathrm{d}x = F(x) \pm G(x) + C.$$

于是

$$\int (f(x) \pm g(x))\ \mathrm{d}x = \int f(x)\ \mathrm{d}x \pm \int g(x)\ \mathrm{d}x.$$

按照第一个证明的思路, 我们可以这样来证明第二条法则: 若 $F(x)$ 是 $f(x)$ 的一个原函数, 那么

$$\int f(x)\,\mathrm{d}x = F(x) + C_1,$$

其中 C_1 为任意常数. 故

$$c\int f(x)\,\mathrm{d}x = c(F(x) + C_1) = cF(x) + C,$$

其中 $C = cC_1$, 也是任意常数. 另一方面,

$$(cF(x))' = cF'(x) = cf(x).$$

可见, $cF(x)$ 是 $cf(x)$ 的一个原函数, 因此

$$\int cf(x)\,\mathrm{d}x = cF(x) + C = c\int f(x)\,\mathrm{d}x.$$

这就证明了第二条法则. □

有了上述法则, 我们就可以根据前述基本积分表, 求出许多函数的不定积分.

例 2.7.1 求不定积分 $\displaystyle\int \left(3x^3 + \sin x + \frac{5}{x}\right)\ \mathrm{d}x$.

证明. $\displaystyle\int \left(3x^3 + \sin x + \frac{5}{x}\right)\ \mathrm{d}x = \frac{3}{4}x^4 - \cos x + 5\ln|x| + C.$ □

> **注意** 注意 $\displaystyle\int \frac{1}{x}\ \mathrm{d}x = \ln|x| + C$. 这里必须是 $\ln|x|$ 而不能是 $\ln x$. 否则就是错误的. 这是因为不定积分运算默认是被积函数 $\dfrac{1}{x}$ 的全部自然定义域上去求原函数. 故包含了 $(-\infty, 0)$ 的部分. 所以要写全了.

不定积分运算默认是在被积函数的全部自然定义域上去求原函数. 后面, 我们学习定积分时, 只需要考虑特定区间上被积函数的原函数即可. 这一点在多数情况下并不需要特别注意. 但是积分函数是 $\frac{1}{x}$ 时需要注意.

例 2.7.2 求不定积分 $\displaystyle\int \frac{(x+1)^2}{\sqrt{x}}\,\mathrm{d}x$.

证明.

$$\int \frac{(x+1)^2}{\sqrt{x}}\,\mathrm{d}x = \int \left(x^{\frac{3}{2}} + 2x^{\frac{1}{2}} + x^{-\frac{1}{2}}\right)\,\mathrm{d}x = \frac{2}{5}x^{\frac{5}{2}} + \frac{4}{3}x^{\frac{3}{2}} + 2x^{\frac{1}{2}} + C.$$

□

例 2.7.3 求不定积分 $\displaystyle\int \left(\mathrm{e}^x + \frac{3x^2}{1+x^2}\right)\,\mathrm{d}x$.

证明.

$$\begin{aligned}\int \left(\mathrm{e}^x + \frac{3x^2}{1+x^2}\right)\,\mathrm{d}x &= \int \mathrm{e}^x\,\mathrm{d}x + 3\int \frac{x^2}{1+x^2}\,\mathrm{d}x \\ &= \mathrm{e}^x + C + 3\int \left(1 - \frac{1}{1+x^2}\right)\,\mathrm{d}x \\ &= \mathrm{e}^x + 3x - 3\arctan x + C.\end{aligned}$$

□

做一次不定积分运算可以获得被积函数 $f(x)$ 的全部原函数. 那么如果再对原函数再做一次不定积分运算的结果是什么呢? 答案是, 我们就能获得所有的二阶导数都等于同一个 $f(x)$ 的函数簇. 注意就像下题目所示, 第二次不定积分运算中, 切记不能忘记了第一运算得到的常数 C.

例 2.7.4 设 $f''(x) = 3x+1$, 求 $f(x)$.

证明. 因为 $f''(x) = (f'(x))'$, 而

$$f'(x) = \frac{3}{2}x^2 + x + C_1.$$

所以

$$f(x) = \frac{1}{2}x^3 + \frac{1}{2}x^2 + C_1x + C_2,$$

其中 C_1, C_2 是任意常数. □

注意 两次不定积分运算中会出现两个任意常数符号 C_1, C_2. 小心, 他们没有任何关联, 不可以互相吞并. 上题还可以写成

$$\int \left(\int (3x+1)\,\mathrm{d}x\right)\,\mathrm{d}x = \frac{1}{2}x^3 + \frac{1}{2}x^2 + C_1x + C_2.$$

其中 C_1, C_2 是任意常数. 更进一步, 设 $f'''(x) = 3x+1$, 怎么求 $f(x)$ 呢?

作为求导运算的逆运算, 求不定积分有其直接应用价值. 读者可从下面两个例子中体会到这一点.

例 2.7.5 设 $s=s(t)$ 满足方程

$$\frac{\mathrm{d}^2 s}{\mathrm{d}t^2}=-g,$$

其中 g 为常数, 并且 $s(0)=h_0, s'(0)=v_0$, h_0 及 v_0 为已知常数, 求 $s=s(t)$ 的表达式.

证明. 显然, 如果不考虑 $s(t)$ 及 $s'(t)$ 在点 $t=0$ 处的限制, 那么 $\dfrac{\mathrm{d}s}{\mathrm{d}t}$ 的一般表达式应为

$$\frac{\mathrm{d}s}{\mathrm{d}t}=-\int g\ \mathrm{d}t=-gt+C_1,$$

而 $s=s(t)$ 的一般表达式应为

$$s=\int(-gt+C_1)\,\mathrm{d}t=-\frac{1}{2}gt^2+C_1t+C_2,$$

其中 C_1, C_2 为任意常数. 再考虑到 $s(t)$ 与 $s'(t)$ 在点 $t=0$ 处的**初始条件**, 代入上式中后, 我们有

$$C_2=h_0,\quad C_1=v_0.$$

于是, 最终我们得到

$$s(t)=-\frac{1}{2}gt^2+v_0t+h_0.$$

这就是大家在中学时已熟知的自由落体运动的位移公式. 这里

$$\frac{\mathrm{d}^2 s}{\mathrm{d}t^2}=-g$$

是根据牛顿第二定律列出的微分方程, 而**不定积分为求解该微分方程提供了工具**. 这里 $s(0)=h_0, s'(0)=v_0$, 叫做**微分方程的初始条件**. 初始条件就是来确定最终的函数用的, 它使得我们在不定积分计算后得到的函数簇中获得唯一一个适用于具体场景下的函数. 这是一个十分简单的例子, 但它说明了求不定积分的意义. □

注释 2.7.1 (额外资料) 理论上, 如果知道物体的受力情况并且初始条件已知, 是可以获得物体的全部运动轨迹的. 这是基于经典力学中的牛顿第二定律:

$$\vec{F} = m\vec{a},$$

其中 $\vec{F}$ 是作用在物体上的合力, m 是物体的质量, $\vec{a}$ 是物体的加速度. 因为加速度是速度和位置的二阶导数, 因此只要知道物体的合力和初始条件 (初始位置和初始速度), 理论上可以通过求解相应的微分方程来得到物体的全部运动轨迹. 求解过程概述:

1. 确定受力状态: 通过合力 $\vec{F}$ 确定物体的加速度 $\vec{a} = \dfrac{\vec{F}}{m}$.
2. 求解加速度的积分: 加速度是速度的时间导数, 即
$$\vec{a}(t) = \frac{\mathrm{d}\vec{v}(t)}{\mathrm{d}t}.$$
知道 $\vec{a}(t)$ 后, 对时间积分可以得到速度 $\vec{v}(t)$ 的表达式 (需要初始速度 $\vec{v}_0$ 作为积分常数).
3. 求解速度的积分: 速度是位置的时间导数, 即
$$\vec{v}(t) = \frac{\mathrm{d}\vec{r}(t)}{\mathrm{d}t}.$$
知道 $\vec{v}(t)$ 后, 对时间再次积分可以得到位置 $\vec{r}(t)$ 的表达式 (需要初始位置 $\vec{r}_0$ 作为积分常数).

例 2.7.6 衰变率表示放射性元素随时间减少的速度, 通常是指放射性物质在单位时间内减少的质量. 对于放射性元素, 衰变率通常用微分表达式表示为质量 $m(t)$ 随时间 t 变化的导数的负数, 即:

$$-\frac{\mathrm{d}m}{\mathrm{d}t}.$$

已知放射性元素的衰变率与当时放射性物体的现存质量成正比, 求放射性物体质量的变化规律.

证明. 设放射性物体的质量为 $m = m(t) > 0$, 其中 t 表示时间. 根据衰变规律, 我们得到微分方程

$$-\frac{\mathrm{d}m}{\mathrm{d}t} = km(t),$$

其中 $k > 0$ 为比例常数. 为了更方便地求解, 我们令 $y(t) = \ln m(t)$. 根据链式法则, 有

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{m'(t)}{m(t)}.$$

将 $m'(t)$ 代入得到

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -k.$$

接下来对两边积分, 得到

$$y(t) = -kt + C,$$

其中 C 是积分常数. 还原 $y(t)$ 的定义 $y(t) = \ln m(t)$, 得到

$$\ln m(t) = -kt + C.$$

因此,

$$m(t) = e^C \cdot e^{-kt} = m_0 e^{-kt},$$

如果已知初始条件 $m(0) = m_0$, 则最终的质量随时间的变化规律为:

$$m(t) = m_0 e^{-kt}.$$

这就是放射性物体质量随时间的变化规律. □

上面两道例题都是在求解微分方程. 所谓求解微分方程就是找到所有满足该微分方程的解函数. 如果题目没有给出初始条件, 我们发现方程的解是有无数个. 就像不定积分一样, 这些无数个解函数的函数簇是能写成一般表达式的, 即包含一个或多个可以任意取值的常数 C. (所以微分方程的初始条件就是用来确定一个具体的 C 的; 没有初始条件的话, 就算得到无数个解也没有现实的应用价值.)

例 2.7.7 已知曲线 $y = f(x)$ 上任意一点 $(x, f(x))$ 处的切线斜率为 $4x^3$, 且此曲线通过点 $(1, 2)$, 求此曲线的方程 $y = f(x)$.

证明. 由条件 $f'(x) = 4x^3$ 知, 此曲线的方程满足

$$y = f(x) = \int 4x^3 \, \mathrm{d}x = x^4 + C.$$

又 $f(1) = 2$, 代人上式, 得 $2 = 1 + C, C = 1$, 故所求的曲线方程为

$$y = f(x) = x^4 + 1.$$

□

解关于求不定积分的方法, 我们将在下一章中更系统地讲解. 而关于微分方程的求解, 则在本教材下册中有专门的一章对其进行讨论.

2.8 定积分: 分割-近似-求和-取极限

2.8.1 定积分的概念

从中学数学课本中, 我们已学会计算一些直线所围成图形的面积和圆的面积. 但是, 我们没有一般的方法用以计算一般曲线所围成图形的面积. 现在定积分就给我们提供了一种计算面积的普遍方法.

曲边梯形的面积

假定连续函数 $y = f(x)$ 定义在区间 $[a, b]$ 上, 且其图形在 x 轴上方. 那么, $y = f(x)$ 的图形与直线 $x = a$, $x = b$ 及 x 轴围成一个**曲边梯形** (见图 2.3). 如果能计算这个曲边梯形的面积, 则我们能计算相当一般的曲线所围成图形的面积. 因为在我们常见的图形中, 一般都能分割为若干曲边梯形的并.

在考虑图 2.3 中这个曲边梯形的面积问题时, 我们采取下面的策略: 先求近似值, 再通过取极限以达到精确值. 可借助交互工具 Riemann Sums –GeoGebra 理解下面步骤.

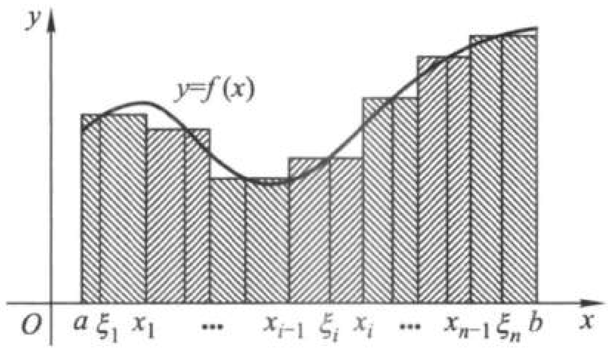


图 2.3: 曲边梯形的面积

1. **分割:** 我们在 $[a, b]$ 上插入一串 (一共 $n+1$ 个) **分点/分割点** $\{x_i \mid i = 0, 1, \cdots, n\}$:

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

注意, 我们通常让闭区间的两端作为第 0 个和第 n 个分割点. 通过这种方式, $n+1$ 个分割点把闭区间 $[a, b]$ 分割成了 n 个相邻的小闭区间. 这种分割, 可以是等距的分割, 也可以不等距.

2. **近似:** 在每个小闭区间 $[x_{i-1}, x_i]$ $(i = 1, 2, \cdots, n)$ 上, 函数 $y = f(x)$ 的图形与直线 $x = x_{i-1}$, $x = x_i$ 及 x 轴相应地形成一个小曲边梯形. 而这些小曲边梯形的真实面积 S_i 可以用小长方形的面积来近似:

$$S_i \approx f(\xi_i)(x_i - x_{i-1}) = f(\xi_i)\Delta x_i,$$

其中, $\Delta x_i := x_i - x_{i-1}$ 是小长方形的底边长度, ξ_i 是 $[x_{i-1}, x_i]$ 中的任意一点 (见图 2.3), 则函数值 $f(\xi_i)$ 就是小长方形的高度. 简单的 ξ_i 的取法有: 取 $[x_{i-1}, x_i]$ 的中点 $\xi_i = \dfrac{1}{2}(x_i + x_{i-1})$, 当然还可以取 $[x_{i-1}, x_i]$ 的左右端点 $\xi_i = x_{i-1}$ 或 $\xi_i = x_i$. 显然, 不同的中间点 ξ_i 的取法会得到不同小长方形, 但它们都可以视作对小曲边梯形的一种近似.

3. **求和:** 这样, 可得整个曲边梯形的真实面积 S 的近似值:

$$S = \sum_{i=1}^{n} S_i \approx \sum_{i=1}^{n} f(\xi_i)\Delta x_i.$$

4. **取极限:** 从直观上看, 当分割点越密 (即分割越细致) 时, 这个近似值就越接近于曲边梯形面积 S 的值. 因此, 自然会把这个近似值的极限值作为曲边梯形面积 S 的值, 即

$$S = \lim_{\lambda \to 0} \sum_{i=1}^{n} f(\xi_i)\Delta x_i,$$

其中 $\lambda = \max\{\Delta x_i \mid i = 1, 2, \cdots, n\}$ 是 Δx_i $(i = 1, 2, \cdots, n)$ 中的最大者. 注意这里随着 $\lambda \to 0$, 必然有 $n \to \infty$.

变力做功

设一个质量为 m 的物体沿直线运动. 假定该物体所受的外力大小可以表示为它的位移 s 的函数 $f(s)$, 方向与运动方向一致. 当函数 $f(s)$ 不是常函数时, 物体受到的力是不断在变化的. 我们要求该物体从 $s = a$ 到 $s = b$ 时外力所做的功 W.

1. **分割:** 同样, 在区间 $[a,b]$ 内插入一串分割点:
$$a=s_0<s_1<s_2<\cdots<s_{n-1}<s_n=b.$$
2. **近似:** 从 s_{i-1} 到 s_i 时外力所做的功近似于常力下的做功:
$$f(\xi_i)(s_i-s_{i-1})=f(\xi_i)\Delta s_i,$$
其中任意取 $\xi_i\in[s_{i-1},s_i]$, $\Delta s_i:=s_i-s_{i-1}$, $i=1,2,\cdots,n$.
3. **求和:** 当分割点增加使分割变细时, 近似值
$$\sum_{i=1}^{n}f(\xi_i)\Delta s_i$$
将越来越接近于我们所求的值.
4. **取极限:** 于是, 问题再一次归结为求如下形式的极限:
$$W=\lim_{\lambda\to 0}\sum_{i=1}^{n}f(\xi_i)\Delta s_i.$$
其中 $\lambda=\max\{\Delta s_i\mid i=1,2,\cdots,n\}$.

变速直线运动的位移

这样的例子还有很多, 比如, 已知做变速 (速度想着怎么变就怎么变, 根本不需要任何规则) 直线运动物体的速度, 求该物体在一定时间内的位移 (或所经过的路程). 这时, 我们同样可以通过这种

(定积分的灵魂) 分割 → 近似代替 → 求和 → 取极限

的步骤, 将问题化成求一个已知函数的某种形式的极限. 这个极限值就是已知函数的定积分.

例 2.8.1 已知一个沿直线运动的质点的瞬时速度为 $v(t)$ $(0\leqslant t\leqslant a)$. 请将该质点自 $t=0$ 时刻到 $t=a$ 时刻的位移写成求和的极限形式.

证明. 使用与之前类似的方法, 我们将时间区间 $[0,a]$ 分成若干小区间, 假设分割点为:
$$0=t_0<t_1<t_2<\cdots<t_{n-1}<t_n=a.$$
在小时间区间 $[t_{i-1},t_i]$ 上, 假设速度保持不变 (即被近似地看作均速直线运动), 取其中一点 $\xi_i\in[t_{i-1},t_i]$ 作为代表点, 这样在小区间 $[t_{i-1},t_i]$ 上的位移可以近似表示为:
$$v(\xi_i)\Delta t_i,$$
其中 $\Delta t_i=t_i-t_{i-1}$ 是该小区间的时间长度. 将所有小区间的位移加总, 我们得到总位移的近似值:
$$\sum_{i=1}^{n}v(\xi_i)\Delta t_i.$$
当我们不断细分区间, 使得分割的最大长度 $\lambda=\max\{\Delta t_i\mid i=1,2,\cdots,n\}$ 趋向于 0 时, 总位移的近似值将趋于精确值. 因此, 质点在 $t=0$ 到 $t=a$ 的位移 s 可以表示为以下的极限形式:
$$s=\lim_{\lambda\to 0}\sum_{i=1}^{n}v(\xi_i)\Delta t_i.$$
□

定积分的定义

现在我们给出定积分的正式定义. 定积分的定义除了下面这种称之为黎曼积分的版本外, 还有其他变体版本. 但是, 它们都是做同一件事情:

(定积分的灵魂) 分割 → 近似代替 → 求和 → 取极限

定义 2.8.1: 定积分/黎曼积分

设 $f(x)$ 是定义在区间 $[a,b]$ 上的函数.

- 对闭区间 $[a,b]$ 插入 $n+1$ 个分割点 x_i $(i=0,1,\cdots,n)$, 使它们满足

$$a=x_0<x_1<x_2<\cdots<x_{n-1}<x_n=b.$$

  我们将其称之为 $[a,b]$ 的一种**分割 (partition of the interval)**, 记为 T. 又记每个小区间的长度

$$\Delta x_i:=x_i-x_{i-1},\ \forall i=1,2,\cdots,n,$$

  以及**所有小区间长度的最大者**:

$$\lambda(T):=\max\left\{\Delta x_i \mid i=1,2,\cdots,n\right\}.$$

  显然, 上面这个量完全由分割 T 决定, 故写成 T 的函数. 该值可以严格衡量一个分割的细致程度.

- 若对于<u>任意的分割</u> T 及<u>任意选取的中间点</u> $\xi_i\in[x_{i-1},x_i]$ $(i=1,2,\cdots,n)$, 极限

$$\lim_{\lambda(T)\to 0}\sum_{i=1}^{n} f\left(\xi_i\right)\Delta x_i$$

  <u>都存在且相等</u>, 则称这个极限值为 $f(x)$ 在 $[a,b]$ 上的 **定积分 (definite integral)**, 记为

$$\int_a^b f(x)\mathrm{d}x.$$

  这里 $f(x)$ 称为 **被积函数**, 而 b 与 a 分别称为 **积分上限**与 **积分下限**, x 称为 **积分变量**, $[a,b]$ 称为 **积分区间**.

- 如果函数 $f(x)$ 的上述极限存在, 则称 $f(x)$ 在区间 $[a,b]$ 上 **可积 (integrable)**.

- 定积分定义中的和式 $\sum\limits_{i=1}^{n} f\left(\xi_i\right)\Delta x_i$ 称为 **黎曼和**, 上述定积分称作 **黎曼积分 (Riemann integral)**, 而可积也称作黎曼可积.

上述定义和之前的例子中, 既然我们想求

$$\sum_{i=1}^{n} f\left(\xi_i\right)\Delta x_i$$

的极限值, 为什么不干脆让定积分定义为极限

$$\lim_{n\to\infty}\sum_{i=1}^{n} f\left(\xi_i\right)\Delta x_i,$$

而是定义为

$$\lim_{\lambda(T)\to 0}\sum_{i=1}^{n} f\left(\xi_i\right)\Delta x_i$$

呢? 即我们使用的是分割 T 的最大区间长度 $\lambda(T)\to 0$ 的极限, 而不是简单地让分割的数量 $n\to\infty$. 这里的关键在于:

$$\lambda(T)\to 0 \ \Rightarrow\ n\to\infty\ (\text{但反之不成立})$$

1. 如果我们固定分割点数 n, 那么无论任何分割 T 都有一个下界:

   $$\frac{b-a}{n}\leq \lambda(T)=\max\left\{\Delta x_i \mid i=1,2,\cdots,n\right\}.$$

   这就阻止了 $\lambda(T)\to 0$. 因此, 如果要让 $\lambda(T)\to 0$, 那么必然会驱使分割数 n 也变得无穷大.

2. 反之, 如果我们只要求 $n\to\infty$, 却没有确保 Δx_i 变小的限制, 那么在不均匀分割的情况下, 某些区间可能始终保持较大, 因此 $\lambda(T)\to 0$ 不一定会实现, 从而无法保证和式 $\displaystyle\sum_{i=1}^{n} f\left(\xi_i\right)\Delta x_i$ 收敛于一个确定的极限值.

综上所述, $\lambda(T)\to 0$ 能够确保在增大 n 的同时, 也使得每个小区间的长度 Δx_i 也逐渐变小. 这保证了分割的细致程度足够高, 黎曼和能收敛到一个确定的数值.

注释 2.8.1 定义 2.8.1 中的极限

$$\lim_{\lambda(T)\to 0}\sum_{i=1}^{n} f\left(\xi_i\right)\Delta x_i$$

不是我们熟悉那种类型, 但它应当有明确定义. 下面我们用 $\varepsilon-\delta$ 说法给出它的定义 (以下内容只是为了严谨, 不是特别重要). 我们说极限 $\displaystyle\lim_{\lambda(T)\to 0}\sum_{i=1}^{n} f\left(\xi_i\right)\Delta x_i$ 存在并等于 I 是指: 对于任意给定的 $\varepsilon>0$, 存在 $\delta>0$, 使得对于任意分割 T 及任意 $\xi_i\in[x_{i-1},x_i]$ $(i=1,2,\cdots,n)$, 都有

$$\left|\sum_{i=1}^{n} f\left(\xi_i\right)\Delta x_i - I\right|<\varepsilon,\qquad \text{只要 } \lambda(T)<\delta.$$

定积分有明显的几何意义:

- 当区间 $[a,b]$ 上的连续函数 $f(x)\geqslant 0$ 时, 其定积分

  $$I=\int_a^b f(x)\mathrm{d}x$$

  恰好代表由曲线 $y=f(x)$, 直线 $x=a, x=b$ 及 x 轴所围成曲边梯形的面积;
- 当 $f(x)\leqslant 0$ 时, 定积分 I 代表相应曲边梯形的面积乘以 -1;
- 当 $f(x)$ 的值在 $[a,b]$ 上有正有负时, 定积分 I 则等于在 x 轴上方和下方的若干曲边梯形面积的**代数和 (在 x 轴上方的曲边梯形面积取正值, 在 x 轴下方的曲边梯形面积取负值)**.

在教材图 2.13 所示的情形中, 有

$$I=\int_a^b f(x)\mathrm{d}x=\int_a^c f(x)\mathrm{d}x+\int_c^d f(x)\mathrm{d}x+\int_d^b f(x)\mathrm{d}x=A_1-A_2+A_3$$

其中 A_1, A_2, A_3 代表图中相应曲边梯形的面积.

例 2.8.2 请给出定积分 $s=\int_0^a v(t)\mathrm{d}t$ 的物理解释, 其中 $v(t)$ 为质点的瞬时速度, 并假定 $v(t)$ 在区间 $[0,b)$ 与 $(c,a]$ 上大于零, 而在区间 $[b,c]$ 上小于或等于零.

证明. 定积分 $s=\int_0^a v(t)\,\mathrm{d}t$ 的物理意义是质点在时间区间 $[0,a]$ 内的净位移. 当 $v(t)>0$ 时, 质点朝一个方向 (例如向前) 运动. 当 $v(t)<0$ 时, 质点则朝相反方向 (例如向后) 运动. 通过计算这三个区间上的定积分并相加, 可以得到质点的净位移, 即质点在时间区间 $[0,a]$ 内相对于起点的位置变化. 总结来说, 定积分 $s=\int_0^a v(t)\,\mathrm{d}t$ 给出的并不是总路程, 而是质点在前进和后退位移相互抵消后的净效果. 因此, 这个定积分可以用来描述质点在 $[0,a]$ 时间段内相对于起始位置的位移变化, 反映出质点在整体上的位置变化方向和大小. □

在定积分的定义中, 要求积分上限 b 大于积分下限 a. 为了今后运算方便, 我们约定:

约定 2.8.1

1. 当 $b=a$ 时, $\int_a^b f(x)\mathrm{d}x=0$.
2. 当 $b<a$ 时, $\int_a^b f(x)\mathrm{d}x=-\int_b^a f(x)\mathrm{d}x$.

这种约定不仅在运算上带来方便, 而且在物理上也是合理的. 比如, 变力从 a 到 b 所做的功是从 b 到 a 所做的功的负值.

并非所有函数都是可积的, 比如**狄利克雷函数** $\mathrm{D}(x)$ **在任意区间** $[a,b]$ **上都不可积**. 事实上, 在 $\mathrm{D}(x)$ 的黎曼和中, 若所有 ξ_i 都取有理数, 则其黎曼和等于 $b-a$; 若所有 ξ_i 都取无理数, 则其黎曼和等于零. 可见, 其黎曼和无极限可言.

关于函数可积性 (即什么样子的函数是可以计算定积分的) 的讨论已超出本教材的要求范围, 这里无法详细讨论. 现在我们不加证明地指出:

命题 2.8.1

闭区间上的连续函数或单调函数都是该区间上的可积函数.

我们还要指出: 一个可积函数一定是有界函数.

例 2.8.3 利用定义求定积分 $\int_0^1 2^x\ \mathrm{d}x$.

证明. 记 $f(x)=2^x$, 则函数 $f(x)$ 在区间 $[0,1]$ 上连续, 从而它在 $[0,1]$ 上可积, 即, 按定义 2.8.1 中的要求, 极限

$$\int_0^1 2^x \ \mathrm{d}x=\lim_{\lambda(T)\to 0}\sum_{i=1}^n f\left(\xi_i\right)\Delta x_i$$

是存在. 于是, 按定义 2.8.1, 对于 $[0,1]$ 的任意分割 T, 任意选取中间点 ξ_i 所得到的黎曼和, 当 $\lambda(T)\to 0$ 时, 都有极限而且都相等. 要求出 $\int_0^1 2^x \ \mathrm{d}x$, 我们只需选择适当的 (方便计算的) 分割和中间点, 使所得到的黎曼和更容易求出极限.

对于任意的正整数 n, 将 $[0,1]$ 进行 n 等分, 得到分割

$$T: x_0=0<x_1<\cdots<x_n=1.$$

其中 $x_i=\dfrac{i}{n}$ $(i=0,1,\cdots,n)$. 则所有的小区间长度都是 $\Delta x_i=\dfrac{1}{n}$, 故而 $\lambda(T)=\dfrac{1}{n}$. 此时, 极限 $\lambda(T)\to 0$ 则等价于 $n\to\infty$. 这种**等分分割**使得我们只需要考虑分割点的个数就足够. 极大地简化了对于 $\lambda(T)$ 的处理. 再次提醒: 我们现在的目的不是根据定义来证明定积分 $\int_0^1 2^x \ \mathrm{d}x$ 的存在性, 它的存在性我们已经用连续性来说明了. 我们现在的目的求出极限值, 所以任何分割 T 和任何中间点 ξ_i 都是允许的.

取中间点 $\xi_i=\dfrac{i}{n}\in[x_{i-1},x_i]$ $(i=1,2,\cdots,n)$ (这里直接取小区间的右端点), 得到黎曼和

$$\sum_{i=1}^n f\left(\xi_i\right)\Delta x_i=\sum_{i=1}^n 2^{\frac{i}{n}}\frac{1}{n}=\frac{1}{n}\cdot\sum_{i=1}^n(2^{\frac{1}{n}})^i=\frac{1}{n}\cdot 2^{\frac{1}{n}}\frac{1-2^{\frac{n}{n}}}{1-2^{\frac{1}{n}}}=\frac{\frac{1}{n}}{2^{\frac{1}{n}}-1}\cdot 2^{\frac{1}{n}}.$$

如前所述, $\lambda(T)\to 0$ 等价于 $n\to\infty$. 另外, 我们有

$$\lim_{n\to\infty}\frac{2^{\frac{1}{n}}-1}{\frac{1}{n}}=\ln 2,\quad \lim_{n\to\infty}2^{\frac{1}{n}}=1,$$

注意第一个极限等价于导数的定义式:

$$\lim_{x\to 0}\frac{2^x-1}{x}=\lim_{x\to 0}\frac{2^x-2^0}{x-0}=\left.(2^x)'\right|_{x=0}=\left.\ln 2\cdot 2^x\right|_{x=0}=\ln 2.$$

最终,

$$\lim_{\lambda(T)\to 0}\sum_{i=1}^n f\left(\xi_i\right)\Delta x_i=\frac{1}{\ln 2}.$$

于是

$$\int_0^1 2^x \ \mathrm{d}x=\frac{1}{\ln 2}.$$

□

显然, 上题的计算非常繁琐. 我们发现 2^x 的一个原函数是 $f(x)=\dfrac{1}{\ln 2}2^x$, 且有

$$f(1)-f(0)=\frac{1}{\ln 2}.$$

求出原函数, 将定积分中的积分上限和下限代入原函数, 然后做差, 就得到我们 (通过复杂的分割, 近似, 求和, 取极限而定义) 的定积分. 这不是巧合, 这就是马上要学习的微积分基本定理. 对定积分的定义要深入理解, 分割, 近似, 求和, 取极限是微积分的核心思想, 这是必须要掌握的. 但是计算则是另外的故事, 一般不用定义出发. 本书的第三章全部都在讲定积分的计算问题.

2.8.2 定积分的性质

上面我们用黎曼和的极限定义了定积分, 并给出了黎曼和极限的严格定义. 尽管黎曼和的极限不同于一般函数的极限, 但过去讲过的有关极限的定理, 如有关极限的四则运算和极限不等式的定理, 对于黎曼和的极限而言依旧成立. 利用这些定理, 立即可推出定积分的若干基本性质:

1. 定积分只依赖于被积函数及积分区间, 而与定积分中所使用的变量符号无关:
$$\int_a^b f(x)\mathrm{d}x = \int_a^b f(t)\mathrm{d}t.$$
2. (定积分的保号性) 设函数 $y = f(x)$ 在区间 $[a, b]$ 上可积, 且 $f(x) \geqslant 0$, 则
$$\int_a^b f(x)\mathrm{d}x \geqslant 0.$$
3. (定积分的保序性) 若函数 $y = f(x)$ 及 $y = g(x)$ 在区间 $[a, b]$ 上可积, 且 $f(x) \geqslant g(x)$, 则
$$\int_a^b f(x)\mathrm{d}x \geqslant \int_a^b g(x)\mathrm{d}x.$$
4. (定积分运算是线性函数) 固定某个闭区间 $[a, b]$, 考虑该区间上任意的可积函数 f, 则定积分运算
$$f \mapsto \int_a^b f(x)\mathrm{d}x \in \mathbf{R}$$
是一个返回实数值的线性函数:
   (a) 设函数 $y = f(x)$ 及 $y = g(x)$ 在区间 $[a, b]$ 上可积, 则函数 $y = f(x) \pm g(x)$ 在 $[a, b]$ 上也可积, 且
$$\int_a^b (f(x) \pm g(x))\mathrm{d}x = \int_a^b f(x)\mathrm{d}x \pm \int_a^b g(x)\mathrm{d}x.$$
   (b) 若函数 $y = f(x)$ 在区间 $[a, b]$ 上可积, 则对于任意常数 c, 函数 $y = cf(x)$ 在 $[a, b]$ 上也可积, 且
$$\int_a^b cf(x)\mathrm{d}x = c\int_a^b f(x)\mathrm{d}x.$$
5. (积分区间的任意可分割) 设函数 $y = f(x)$ 在区间 $[a, b]$ 上可积, 则对于任意 $c \in (a, b), y = f(x)$ 在区间 $[a, c]$ 及 $[c, b]$ 上都可积, 且
$$\int_a^b f(x)\mathrm{d}x = \int_a^c f(x)\mathrm{d}x + \int_c^b f(x)\mathrm{d}x.$$
其实由此可推出更一般的结论: 当 $c \notin [a, b]$ 时, 只要 $f(x)$ 还在区间 $[c, a]$ 或 $[b, c]$ 上可积, 则上述公式仍成立.[21]

[21]事实上, 若 $c < a$, 这时可将 a 看作 $[c, b]$ 内的一点而套用上述公式, 得
$$\int_c^b f(x)\mathrm{d}x = \int_c^a f(x)\mathrm{d}x + \int_a^b f(x)\mathrm{d}x,$$
再移项即得
$$\int_a^b f(x)\mathrm{d}x = \int_c^b f(x)\mathrm{d}x - \int_c^a f(x)\mathrm{d}x = \int_c^b f(x)\mathrm{d}x + \int_a^c f(x)\mathrm{d}x.$$
同理可证, 当 $c > b$ 时, 公式也成立.

6. 设函数 $y=f(x)$ 在区间 $[a,b]$ 上可积, 又设函数 $y=g(x)$ 在 $[a,b]$ 上有定义, 且仅在有限个点处与 $y=f(x)$ 取不同值, 那么 $y=g(x)$ 在 $[a,b]$ 上也可积, 且

$$\int_a^b f(x)\mathrm{d}x=\int_a^b g(x)\mathrm{d}x.$$

这一性质告诉我们: 仅在有限个点处改变函数值并不影响函数的可积性及积分值.[22] 进一步地我们可以推出: 若函数 $y=f(x)$ 在区间 $[a,b]$ 仅有有限个点处不连续.

(a) 第一类间断点: 左右极限都存在，但它们彼此不相等. 其定积分值是: 以这些间断点为边界, 将连续部分分别计算其定积分, 最后求和所得.

(b) 第一类间断点: 左右极限存在且相等, 但不等于函数值 $f(x_0)$, 又称 x_0 为可去间断点. 其定积分值计算可以完全忽略这些点.

7. (定积分的绝对值不等式) 若函数 $y=f(x)$ 在区间 $[a,b]$ 上可积, 则函数 $y=|f(x)|$ 在 $[a,b]$ 上也可积, 且

$$\left|\int_a^b f(x)\mathrm{d}x\right|\leqslant\int_a^b|f(x)|\mathrm{d}x.$$

我们略去这一性质的证明. 画图理解即可.

8. (偶倍奇零) 若 $f(x)$ 在 $[-a,a]$ 上是偶函数, 则

$$\int_{-a}^a f(x)\mathrm{d}x=2\int_0^a f(x)\mathrm{d}x;$$

若 $f(x)$ 在 $[-a,a]$ 上是奇函数, 则

$$\int_{-a}^a f(x)\mathrm{d}x=0;$$

画图理解即可.

2.8.3　定积分的历史

在牛顿与莱布尼茨时代, 定积分并没有严格的定义. 牛顿把求导运算称作 "**流数**术", 而把积分运算作为求导运算的逆运算, 称作 "反流数术". 莱布尼茨则把定积分视作无限个无穷小量之和. 起初他用 omn 表示积分运算, 后来改用 $\int$ 来表示积分运算. 这里的积分号 $\int$ 是 "sum" 的首字母"s" 的拉长. 在莱布尼茨看来, 如果 $\dfrac{\mathrm{d}z}{\mathrm{d}x}=f(x)$, 即 $\mathrm{d}z=f(x)\mathrm{d}x$, 那么

$$z=\int\mathrm{d}z=\int f(x)\mathrm{d}x.$$

(请类比 $\int f(x)\mathrm{d}x$ 和 $\sum_{i=1}^n f(\xi_i)\Delta x_i$) 至今人们仍使用着莱布尼茨这种记号, 它为人们带来很大方便.

关于定积分的明确定义首先是由柯西给出的. 他假定函数 $y=f(x)$ 在区间 $[x_0,x]$ 上连续, 将 $[x_0,x]$ 用有序的分点 $x_0,x_1,\cdots,x_n=x$ 划分成若千小区间, 并考虑和式

$$S=\sum_{i=1}^n f(x_{i-1})(x_i-x_{i-1}).$$

[22]这一性质可以根据定积分的定义直接证明. 事实上, 对于 $[a,b]$ 的同一分割 T 及相同的中间点 ξ_i 的选取, $f(x)$ 的黎曼和与 $g(x)$ 的黎曼和中至多有有限项不同. 当 $\lambda(T)\to 0$ 时, 这有限项之和的极限为零. 因此, 两个黎曼和有相同的极限.

他定义 $f(x)$ 在 $[x_0, x]$ 上的定积分就是分割后小区间长度趋向于零时和式 S 的极限. 柯西证明了闭区间上连续函数的上述定积分是存在的, 并在此基础上对微积分基本定理 (牛顿-莱布尼茨公式) 给予严格的表述与证明.

黎曼研究了一般函数的可积性问题, 把何西关于连续函数定积分的定义加以修改, 其中 $f(x_{i-1})$ 换成 $f(\xi_i), \xi_i$ 是 $[x_{i-1}, x_i]$ 中的任意一点. 因此, 现代形式的定积分定义是由黎曼给出的.

20 世纪初, 法国数学家勒贝格 (H. Lebesgue) 引入了点集合测度的概念, 并在此基础上建立了一种新的积分. 这种积分称为勒贝格积分. 勒贝格积分克服了黎曼积分的某些局限性, 成为近代分析学乃至其他数学领域的基础.

黎曼是 19 世纪最富有开创性的德国数学家, 逝世时年仅 40 岁. 1851 年, 他的博士论文为复变函数与黎曼曲面论奠定了基础, 其中一条定理后来被称为黎曼映射定理, 成为几何函数论的基础. 1854 年, 他在格丁根大学任教就职讲演中提出了全新的几何观念, 这成为现代黎曼几何的发端. 1858 年, 他关于素数分布的论文研究了黎曼函数, 为解析数论奠定了基础. 他所提出的关于黎曼函数零点分布的猜想至今尚未解决. 现代数学的许多分支处处可以看到黎曼的影响.

1. 【天才简史-黎曼】黎曼，为数学而生的天才！他的理论曾启发爱因斯坦 _ 哔哩哔哩 _bilibili

2. 【黎曼猜想】1+2+3+4+...=-1/12？李永乐老师讲黎曼猜想（1）_ 哔哩哔哩 _bilibili

3. 【黎曼几何和广义相对论】爱因斯坦的数学很差吗？什么是罗氏几何和黎曼几何？它们曾经可是数学家的噩梦！ _ 哔哩哔哩 _bilibili

2.9 变上限的定积分

前面我们引入了不定积分与定积分的概念. 本节我们讨论连续函数的变上限的定积分, 从而沟通定积分与不定积分之间的联系.

定义 2.9.1: 变上限积分函数

设函数 $f(x)$ 在区间 $[a, b]$ 上连续. 由上一节的讨论知, 这时 $f(x)$ 在 $[a, b]$ 上可积, 从而对于任意一点 $x \in [a, b]$, $f(x)$ 在子区间 $[a, x]$ 上也是可积的[a]. 则积分上限为 $x \in [a, b]$ 的定积分

$$\int_a^x f(t) \mathrm{d}t$$

称作 f 的 **变上限积分**. 这里, 我们在定积分中把函数 $f(x)$ 的自变量写成 t, 主要是为了避免与积分上限 x 发生混淆. 显然, **变上限积分函数**是关于 x 的一个函数:

$$x \mapsto F_0(x) := \int_a^x f(t) \mathrm{d}t, \ F_0 : [a, b] \to \mathbf{R}$$

[a]因为在任何子区间上也是连续的函数

注意 以后的多数情况下, 我们只考虑在区间 $[a, b]$ 上连续函数 $f(x)$. 因此, 其变上限积分函数 $x \mapsto \displaystyle\int_a^x f(t) \mathrm{d}t$ 总是存在的, 总是可以安全安心地考虑它. 如果不是连续函数, 变上限积分函数的存在性是值得商榷的.

后面我们将要证明: 变上限积分函数 $F_0(x) = \int_a^x f(t)\mathrm{d}t$ 就是 $f(x)$ 的一个原函数. 为此, 我们先证明一个预备性定理 — 积分中值定理.

定理 2.9.1: 积分中值定理

设函数 $f(x)$ 在区间 $[a,b]$ 上连续, 则在 $[a,b]$ 内至少存在一点 c, 使得

$$\int_a^b f(x)\mathrm{d}x = f(c)(b-a).$$

或写成

$$f(c) = \frac{\int_a^b f(x)\mathrm{d}x}{b-a}.$$

证明. 因为 $f(x)$ 在 $[a,b]$ 上连续, 它在 $[a,b]$ 上就有最大值 M 与最小值 m, 即

$$m \leqslant f(x) \leqslant M, \quad \forall x \in [a,b].$$

由上一节中定积分的性质[23]有

$$\int_a^b m\ \mathrm{d}x \leqslant \int_a^b f(x)\mathrm{d}x \leqslant \int_a^b M\ \mathrm{d}x.$$

由课本习题 2.8 的 1. (1) 练习可知 $\int_a^b k\ \mathrm{d}x = k(b-a)$, 其中 k 为常数. 由此推出

$$m \leqslant \frac{\int_a^b f(x)\mathrm{d}x}{b-a} \leqslant M.$$

再由连续函数的中值定理结合在闭区间的极值定理可知[24], 存在 $c \in [a,b]$, 使得

$$f(c) = \frac{\int_a^b f(x)\mathrm{d}x}{b-a},$$

即 $\int_a^b f(x)\mathrm{d}x = f(c)(b-a)$. 证毕. □

从几何上来看, 上述定理的意义是: 以曲线弧 $y = f(x)$ $(a \leqslant x \leqslant b)$ 为曲边的曲边梯形面积, 恰好相当于过该曲线弧上某点 $(c, f(c))$ 处的水平线段所对应的矩形面积 (见图 2.4).

例 2.9.1 设函数 $f(x) = 2^x$, 求实数 c, 使得 $\int_0^1 f(x)\mathrm{d}x = f(c)(1-0)$.

证明. 利用上一节例题的结果 $\int_0^1 2^x\ \mathrm{d}x = \frac{1}{\ln 2}$, 再根据积分中值定理, 存在 $c \in [0,1]$, 使得

$$\frac{1}{\ln 2} = 2^c(1-0).$$

由此解出 $c = -\frac{\ln\ln 2}{\ln 2}$. □

[23]若函数 $y = f(x)$ 及 $y = g(x)$ 在区间 $[a,b]$ 上可积, 且 $f(x) \geqslant g(x)$, 则 $\int_a^b f(x)\mathrm{d}x \geqslant \int_a^b g(x)\mathrm{d}x$.

[24]第一章时, 我们一个推论: 在闭区间 $[a,b]$ 上连续的函数 $f(x)$ 的值域为闭区间 $[m,M]$, 其中 m 与 M 依次为 $f(x)$ 在 $[a,b]$ 上的最小值与最大值.

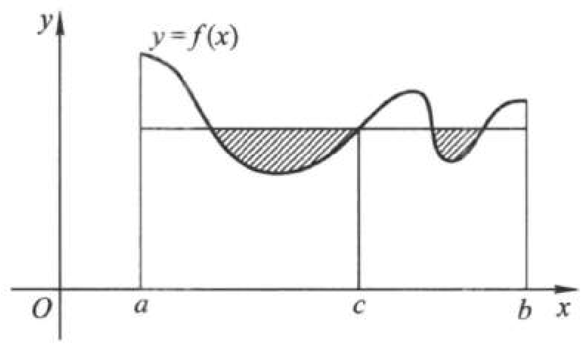


图 2.4: 积分中值定理

例 2.9.2 利用积分中值定理说明: 一个沿直线运动的质点, 在任何一段时间内的平均速度总是等于这段时间内某个时刻的瞬时速度.

证明. 设 $v(t)$ 为质点的瞬时速度函数, 在时间区间 $[t_1, t_2]$ 上连续. 质点在此时间段内走过的距离可表示为 $s = \int_{t_1}^{t_2} v(t)\mathrm{d}t$. 根据积分中值定理, 存在一个时间点 $c \in [t_1, t_2]$, 使得 $\int_{t_1}^{t_2} v(t)\mathrm{d}t = v(c)(t_2 - t_1)$. 这个等式可以改写为 $s = v(c)(t_2 - t_1)$. 两边同时除以时间间隔 $(t_2 - t_1)$, 得到 $\frac{s}{t_2 - t_1} = v(c)$. 等式左边正是质点在该时间段内的平均速度, 右边则是质点在时刻 c 的瞬时速度. 这就证明了, 不管我们选择多长的时间间隔, 总能在这个间隔中找到一个时刻, 使得该时刻的瞬时速度恰好等于整个时间间隔的平均速度, 为我们理解运动过程提供了重要的数学依据. □

定理 2.9.2: 变上限积分函数是一个原函数

设函数 $f(x)$ 在区间 $[a, b]$ 上连续, 则其变上限积分函数

$$x \mapsto F_0(x) := \int_a^x f(t)\mathrm{d}t, \ \ F_0 : [a, b] \to \mathbf{R}$$

是 $[a, b]$ 上的连续函数, 且在 (a, b) 内可导, 并有

$$F_0'(x) = f(x), \quad \forall x \in (a, b).$$

证明. 我们先证明函数 F_0 在 $[a, b]$ 上是连续的.

- 选择一点 $x_0 \in [a, b)$. 考虑其右侧的充分近的某点 $x > x_0$ 使得 $x \in (a, b]$. 根据变上限积分的定义和定积分的性质, 我们有

$$F_0(x) - F_0(x_0) = \int_a^x f(t)\mathrm{d}t - \int_a^{x_0} f(t)\mathrm{d}t = \int_a^x f(t)\mathrm{d}t + \int_{x_0}^a f(t)\mathrm{d}t = \int_{x_0}^x f(t)\mathrm{d}t.$$

  由积分中值定理, 存在常数 $c_x \in [x_0, x]$ 使得

$$F_0(x) - F_0(x_0) = f(c_x)(x - x_0).$$

当 $x \to x_0+0$ 时, 我们反复利用积分中值定理, 总是存在一个依赖于 x 的常数 c_x 使得上式成立. 特别地, 当 $x \to x_0+0$ 时, $c_x \to x_0$, 故有

$$F_0(x)-F_0\left(x_0\right) \to f(x_0)\left(x_0-x_0\right)=0 \quad (x \to x_0+0).$$

由此推出

$$\lim_{x \to x_0+0} F_0(x)=F_0\left(x_0\right).$$

以上证明了 $F_0(x)$ 在点 x_0 处右连续. 由于 x_0 可以是 $[a, b)$ 中的任意一点, 这也就证明了 $F_0(x)$ 在 $[a, b)$ 中每一点处都右连续.

- 选择一点 $x_0 \in (a, b]$. 考虑其左侧的充分近的某点 $x<x_0$ 使得 $x \in [a, b)$. 同理可证

$$\lim_{x \to x_0-0} F_0(x)=F_0\left(x_0\right).$$

故, $F_0(x)$ 在 $(a, b]$ 中每一点处都左连续. 再结合刚才所得 $F_0(x)$ 在 $[a, b)$ 上右连续的结论, 我们便证明了 $F_0(x)$ 在 $[a, b]$ 上的连续性.

接下来, 我们要求 $F_0(x)$ 的导函数. 已知 $F_0(x)$ 在 $[a, b]$ 上是连续的. 设 $x_0, x \in (a, b)$ 且 $x_0 \neq x$, 则由积分中值定理可知在 x 与 x_0 之间存在一点 c, 使得

$$F_0(x)-F_0\left(x_0\right)=\int_{x_0}^{x} f(t) \mathrm{d} t=f(c)\left(x-x_0\right).$$

由此推出

$$\frac{F_0(x)-F_0\left(x_0\right)}{x-x_0}=f(c).$$

当 $x \to x_0$ 时, $c \to x_0$. 于是, 由 $f(x)$ 的连续性可知, 当 $x \to x_0$ 时, $f(c) \to f\left(x_0\right)$, 因而

$$F_0^{\prime}(x_0)=\lim_{x \to x_0} \frac{F_0(x)-F_0\left(x_0\right)}{x-x_0}=f\left(x_0\right).$$

证毕. □

这个定理告诉我们: **一个连续函数的变上限积分就是该函数的一个原函数**. 变上限积分将定积分与不定积分联系在一起. 这个定理是证明微积分基本定理的基础.

推论 2.9.1

任意连续函数都有原函数.

变上限积分有明显的物理意义与几何意义.

- 瞬时速度为 $v(t)$ 的运动质点自 $t=0$ 时刻到 $t=x$ 时刻的位移为

$$s(x)=\int_0^x v(t) \mathrm{d} t.$$

那么, 位移函数的导数就是瞬时速度, 即 $s^{\prime}(x)=v(x)$, 这是再自然不过的事了.

- 当我们把变上限积分

$$S(x)=\int_a^x f(t) \mathrm{d} t$$

看作变动中的曲边梯形的面积时 (见图 2.5), 定理 2.9.2 的几何意义就十分清楚了: 曲边梯形的面积 $S(x)$ 在一点 x 处的变化率 $\dfrac{\mathrm{d} S}{\mathrm{d} x}$ 恰好等于 $f(x)$.

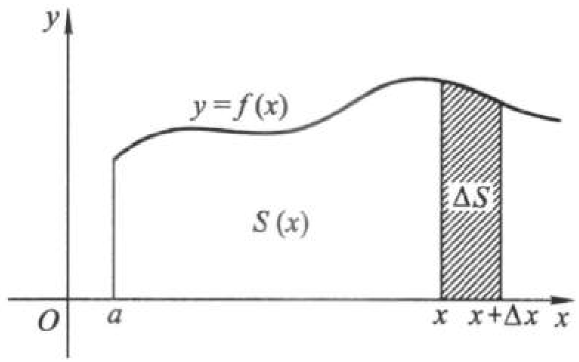


图 2.5: 变上限积分的几何意义

注意 前面讨论中我们只考虑了 $x > a$ 的情况. 实际上, 当 $x < a$ 时, 公式

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x f(t)\mathrm{d}t = f(x)$$

同样成立, 只要其中的函数 $f(x)$ 在所考虑的区间上连续即可.

例 2.9.3 设函数

$$F(x) = \int_0^{2x+1} \mathrm{e}^t \sin 5t \ \mathrm{d}t.$$

求 $F'(x)$.

证明. 我们将 $F(x)$ 视为函数

$$G(y) = \int_0^y \mathrm{e}^t \sin 5t \ \mathrm{d}t$$

与 $y = 2x + 1$ 的复合函数. 于是, 我们有

$$F'(x) = G'(y)y' = \mathrm{e}^y \sin 5y \cdot 2 = 2\mathrm{e}^{2x+1}\sin 5(2x+1).$$

□

设函数 $f(x)$ 在区间 $[a, b]$ 上连续, 我们还需要考虑**变下限积分函数**

$$x \mapsto \int_x^b f(t)\mathrm{d}t.$$

这时, 由于它可以变换成变上限积分:

$$\int_x^b f(t)\mathrm{d}t = -\int_b^x f(t)\mathrm{d}t.$$

所以在 $f(x)$ 连续的条件下有

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_x^b f(t)\mathrm{d}t = -f(x).$$

即, 对于变下限积分函数 $x \mapsto \int_x^b f(t)\mathrm{d}t$ 关于 x 的导函数是 $f(x)$ 的负值. 注意, 这不同于变上限积分函数 $x \mapsto \int_a^x f(t)\mathrm{d}t$ 关于 x 的导函数就是 $f(x)$ 的本身.

例 2.9.4 求函数

$$F(x) = \int_{x^2}^{x} \sqrt{1+t}\ \mathrm{d}t$$

的导数.

证明. 将 $F(x)$ 写成

$$F(x) = \int_0^x \sqrt{1+t}\ \mathrm{d}t + \int_{x^2}^0 \sqrt{1+t}\ \mathrm{d}t = \int_0^x \sqrt{1+t}\ \mathrm{d}t - \int_0^{x^2} \sqrt{1+t}\ \mathrm{d}t,$$

于是我们得到

$$F'(x) = \sqrt{1+x} - 2x\sqrt{1+x^2}.$$

$\square$

将上述题目一般化, 则有

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_{u(x)}^{v(x)} f(t)\mathrm{d}t = f(v(x))\frac{\mathrm{d}v}{\mathrm{d}x} - f(u(x))\frac{\mathrm{d}u}{\mathrm{d}x}.$$

例 2.9.5 设函数 $G(x) = \int_1^x xf(t)\mathrm{d}t$, 其中 $f(x)$ 是区间 $(-\infty, +\infty)$ 上的连续函数, 求 $G'(x)$.

注意, 对于积分运算来讲, 尾巴 “dt” 声明了只有变量 t 对于它而言才是变量. 因而其他的都是常量, 所以能提出积分运算之外. 比如,

$$\int_1^x h(x)f(t)\mathrm{d}t = h(x)\int_1^x f(t)\mathrm{d}t.$$

证明. 将 $G(x)$ 写成

$$G(x) = x\int_1^x f(t)\mathrm{d}t,$$

则利用函数积的求导法则, 有

$$G'(x) = \int_1^x f(t)\mathrm{d}t + xf(x).$$

$\square$

将上题一般化, 可以总结出下面的问题.

例 2.9.6 设函数

$$F(x) = \int_a^{g(x)} h(x)f(t)\,\mathrm{d}t$$

其中 $h(x)$ 和 $g(x)$ 为 x 的可导函数, 且 $f(t)$ 是关于 t 的连续函数. 求 $F(x)$ 关于 x 的导数 $F'(x)$.

证明. 首先, 注意到 $h(x)$ 仅依赖于 x, 而在积分中, t 是积分变量, 因此可以将 $h(x)$ 提出积分符号之外. 因此, 函数 $F(x)$ 可以重新写成:

$$F(x) = h(x)\int_a^{g(x)} f(t)\,\mathrm{d}t.$$

现在我们可以对 $F(x) = h(x)\int_a^{g(x)} f(t)\,\mathrm{d}t$ 使用乘积法则进行求导. 根据乘积法则, 我们有:

$$F'(x) = \frac{\mathrm{d}h(x)}{\mathrm{d}x}\int_a^{g(x)} f(t)\,\mathrm{d}t + h(x)\cdot\frac{\mathrm{d}}{\mathrm{d}x}\left(\int_a^{g(x)} f(t)\,\mathrm{d}t\right).$$

接下来分别计算这两个部分:

1. 由于 $h(x)$ 是关于 x 的可微函数, 对其直接求导得到 $h'(x)$:

$$h'(x)\int_a^{g(x)} f(t)\,\mathrm{d}t.$$

2. 对 $\int_a^{g(x)} f(t)\,\mathrm{d}t$ 关于 x 取导数. 利用复合函数求导的链式法则和定理 2.9.2, 有

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\int_a^{g(x)} f(t)\,\mathrm{d}t\right) = f(g(x))\cdot g'(x).$$

将两个部分合并, 我们得到

$$F'(x) = h'(x)\int_a^{g(x)} f(t)\,\mathrm{d}t + h(x)\cdot f(g(x))\cdot g'(x).$$

□

例 2.9.7 设函数 $F(x) = \int_1^x t^3\ \mathrm{d}t$.

1. 求 $F'(x)$.

2. 求微分方程 $\frac{\mathrm{d}y}{\mathrm{d}x} = x^3$ 满足当 $x = 1$ 时, $y = 0$ 的解.

3. 求定积分 $\int_1^4 t^3\ \mathrm{d}t$.

证明. 1. 利用定理 2.9.2, 得 $F'(x) = x^3$.

2. 所给的方程两端积分, 得

$$y = \int x^3\ \mathrm{d}x = \frac{1}{4}x^4 + C.$$

当 $x = 1$ 时, $y = 0$, 有 $0 = \frac{1}{4} + C$, 故 $C = -\frac{1}{4}$. 于是所求的解为

$$y = \frac{1}{4}x^4 - \frac{1}{4}.$$

3. 我们发现函数

$$F(x) = \int_1^x t^3 \, \mathrm{d}t$$

就是第二问微分方程的解. 首先, 由第一问可知 $y = F(x)$ 满足微分方程 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = x^3$. 其次, 当 $x = 1$ 时, $y = F(1) = \displaystyle\int_1^1 t^3 \, \mathrm{d}t = 0$, 故满足第二问的初始条件. 所以,

$$F(x) = \frac{1}{4}x^4 - \frac{1}{4}.$$

因此, $\displaystyle\int_1^4 t^3 \, \mathrm{d}t = F(4) = \frac{1}{4} \times 4^4 - \frac{1}{4} = \frac{255}{4}$.

□

从上面的例子我们发现

$$\int_1^4 t^3 \, \mathrm{d}t = F(4) - F(1)$$

其中 $F'(x) = x^3$, 即 $F(x)$ 是被积函数 $f(x) = x^3$ 的一个原函数. 这个结论是否对其他函数也成立? 下一节中介绍的微积分基本定理回答了这个问题.

2.10 微积分基本定理

下面我们要证明微积分中最重要的一个定理 — 微积分基本定理. 它将告诉我们: 一个连续函数的定积分可通过该函数的 (任意的) 一个原函数在积分区间两个端点处的值做差来计算. 这样, 定积分的计算就归结为求原函数.

在牛顿与莱布尼茨创立微积分之前, 人们早就知道利用求极限的方法来计算某些图形的面积, 比如我国古代的割圆术. 定积分概念的引入使这种方法得到一般化. 然而, 如果没有一个有效的计算定积分的方法, 那么其应用价值将大受限制.

- 通过习题 2.8 中的第 3 题可以看出, 根据定义计算定积分

  $$\int_0^1 x^2 \, \mathrm{d}x.$$

  这时, 问题归结为求极限

  $$\lim_{n\to\infty} \sum_{k=1}^n \frac{1}{n}(\frac{k}{n})^2 = \lim_{n\to\infty} \frac{1}{n^3} \sum_{k=1}^n k^2.$$

  要用到求和公式

  $$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{1}{6}n(n+1)(2n+1).$$

  但是, 用此方法去计算其他定积分就会遇到很大困难, 甚至束手无策.

- 比如, 计算定积分

  $$\int_0^1 \mathrm{e}^x \sin x \, \mathrm{d}x.$$

  这时, 问题归结为求极限

  $$\lim_{n\to\infty} \sum_{k=1}^n \frac{1}{n}\mathrm{e}^{\frac{k}{n}} \sin\frac{k}{n}.$$

  而这是十分困难的.

牛顿与莱布尼茨建立了一个公式:

$$\int_a^b f(x)\mathrm{d}x = F(b) - F(a).$$

其中 $F(x)$ 是 $f(x)$ 的任意一个原函数. 这为计算定积分给出了一种一般方法. 这个公式把求定积分的问题归结为求原函数的问题.

定理 2.10.1: 微积分基本定理

设函数 $f(x)$ 在闭区间 $[a,b]$ 上连续, 又设 $F(x)$ 是 $f(x)$ 在开区间 (a,b) 上的一个原函数, 即

$$F'(x) = f(x), \quad \forall x \in (a,b).$$

并且设 $F(x)$ 在 $[a,b]$ 上连续[a], 这时我们有

$$\int_a^b f(x)\mathrm{d}x = F(x)\big|_a^b := F(b) - F(a).$$

这个公式称为**牛顿-莱布尼茨公式**. 该公式可以写成

$$\int_a^b \frac{\mathrm{d}F}{\mathrm{d}x}\,\mathrm{d}x = F(b) - F(a) \quad \text{或} \quad \int_a^b \mathrm{d}F = F(x)\big|_a^b.$$

[a]有同学会好奇, 既然可导必然连续, 那么条件 "$F(x)$ 是 $f(x)$ 在开区间 (a,b) 上的一个原函数, 即 $F'(x) = f(x)$, $x \in (a,b)$" 不就说明了 $F(x)$ 的连续性了吗? 确实如此, 但注意这一条件只能反映 $F(x)$ 在整个开区间 (a,b) 上是连续的, 并没有反映左右两个端点的单侧极限情况. 从下面的证明过程中可以看出, 进一步假定 $F(x)$ 在左右两个端点处单侧连续性是必要的.

证明. 设 $F(x)$ 是 $f(x)$ 在开区间 (a,b) 上的一个原函数. 定义一个积分上限函数

$$F_0(x) = \int_a^x f(t)\mathrm{d}t, \quad \forall x \in [a,b].$$

由上一节中的定理 2.9.2 有

$$F_0'(x) = f(x), \quad \forall x \in (a,b).$$

这样, $F'(x) - F_0'(x) \equiv 0$, $\forall x \in (a,b)$, 于是

$$F(x) = F_0(x) + C, \quad \forall x \in (a,b),$$

其中 C 是一个常数. 再由 $F(x)$ 及 $F_0(x)$ 在 $[a,b]$ 上的连续性[25]有

$$F(a) = \lim_{x\to a+0} F(x) = \lim_{x\to a+0}(F_0(x) + C) = F_0(a) + C = C.$$

注意, 上式的第一个等式来自 $F(x)$ 在点 a 处的右连续性; 第二个等式是因为在点 a 的足够小的右侧区间内恒有 $F(x) = F_0(x) + C$; 第三个等式来自 $F(x)$ 在点 a 处的右连续性; 最后等式使用了 $F_0(a) = 0$. 最终我们有

$$F(a) = C.$$

同理可得 $F(b) = F_0(b) + C$. 于是

$$\int_a^b f(t)\mathrm{d}t = F_0(b) = F(b) - C = F(b) - F(a).$$

证毕. □

[25]注意, $F(x)$ 在 $[a,b]$ 上的连续性是本定理的给定条件. 及 $F_0(x)$ 在 $[a,b]$ 上的连续性是使用定理 2.9.2 的结果, 并非本定理一开始给定的.

例 2.10.1 求定积分 $\displaystyle\int_0^1 (\mathrm{e}^x + x)\,\mathrm{d}x$.

证明. 很容易验证 $\left(\mathrm{e}^x + \frac{1}{2}x^2\right)' = \mathrm{e}^x + x$, 也就是说, $\mathrm{e}^x + \frac{1}{2}x^2$ 是 $\mathrm{e}^x + x$ 的一个原函数, 并且它们在区间 $[0,1]$ 上满足微积分基本定理的条件, 即 $\mathrm{e}^x + x$ 和 $\mathrm{e}^x + \frac{1}{2}x^2$ 都在区间 $[0,1]$ 上连续. 于是, 根据牛顿-莱布尼茨公式, 我们得到

$$\int_0^1 (\mathrm{e}^x + x)\,\mathrm{d}x = \left(\mathrm{e}^x + \frac{1}{2}x^2\right)\Big|_0^1 = \mathrm{e} - \frac{1}{2}.$$

□

例 2.10.2 验证

$$F(x) = \frac{1}{2}\mathrm{e}^x(\sin x - \cos x)$$

是 $f(x) = \mathrm{e}^x \sin x$ 的一个原函数, 并计算定积分 $\displaystyle\int_0^1 \mathrm{e}^x \sin x\ \mathrm{d}x$.

证明. 我们有

$$F'(x) = \frac{1}{2}\mathrm{e}^x(\sin x - \cos x) + \frac{1}{2}\mathrm{e}^x(\cos x + \sin x) = \mathrm{e}^x \sin x.$$

可见, $F(x)$ 是 $f(x) = \mathrm{e}^x \sin x$ 的一个原函数. 于是, 我们无须处理复杂的极限而得到

$$\int_0^1 \mathrm{e}^x \sin x\ \mathrm{d}x = F(1) - F(0) = \frac{\mathrm{e}}{2}(\sin 1 - \cos 1) + \frac{1}{2}.$$

当然, 在这个例子中是在已知 $f(x) = \mathrm{e}^x \sin x$ 的原函数 $F(x)$ 的条件下求得定积分值的. 在下一章我们将系统地讲授原函数的求法 (也就是不定积分的求法). □

例 2.10.3 由微积分基本定理不难验证下列公式:

1. $\arcsin x = \displaystyle\int_0^x \frac{\mathrm{d}t}{\sqrt{1-t^2}} \quad (-1 < x < 1)$.
2. $\arctan x = \displaystyle\int_0^x \frac{\mathrm{d}t}{1+t^2}$.
3. $\ln x = \displaystyle\int_1^x \frac{\mathrm{d}t}{t} \quad (x > 0)$.

这三个公式告诉我们一个有趣的事实: 反三角函数及对数函数可以表示成某些简单函数的积分. 特别地, 我们有

$$\frac{\pi}{4} = \int_0^1 \frac{\mathrm{d}t}{1+t^2}.$$

定积分是通过黎曼和形式的极限来定义的. 既然我们已经掌握快速计算定积分的牛顿-莱布尼茨公式, 反过来, 就可以用它来计算某个类似黎曼和形式的极限了. 下面就是一个应用. 这个方法求极限只能只用于类似黎曼和形式的极限问题, 难点在于如何发现, 构造定积分.

例 2.10.4 求极限 $\lim\limits_{n\to\infty}\sum\limits_{k=1}^{n}\left(\frac{k}{n}+1\right)^{3}\frac{1}{n}$.

证明. 设 $f(x)=(x+1)^3$. 将区间 $[0,1]$ 进行 n 等分:

$$x_0=0<x_1<\cdots<x_n=1.$$

其中 $x_k=\frac{k}{n}$ $(k=0,1,\cdots,n)$, 于是 $\Delta x_k=\frac{1}{n}$ $(k=1,2,\cdots,n)$. 取 $\xi_k=\frac{k}{n}\in[x_{k-1},x_k]$ $(k=1,2,\cdots,n)$, 得到黎曼和

$$\sum_{k=1}^{n}f\left(\xi_k\right)\Delta x_k=\sum_{k=1}^{n}\left(\frac{k}{n}+1\right)^{3}\frac{1}{n}.$$

因为 $f(x)=(x+1)^3$ 在 $[0,1]$ 上可积, 并且利用牛顿-莱布尼茨公式有

$$\int_0^1(x+1)^3\ \mathrm{d}x=\left.\frac{1}{4}(x+1)^4\right|_0^1=\frac{1}{4}\left(2^4-1\right)=\frac{15}{4}.$$

所以

$$\lim_{n\to\infty}\sum_{k=1}^{n}\left(\frac{k}{n}+1\right)^{3}\frac{1}{n}=\int_0^1(x+1)^3\ \mathrm{d}x=\frac{15}{4}.$$

□

从上面我们可以发现以下这一般结果. 设 $f(x)$ 是区间 $[0,1]$ 上的连续函数, 则定积分 $\int_0^1 f(x)\ \mathrm{d}x$ 存在. 将区间 $[0,1]$ 进行 n 等分:

$$x_0=0<x_1<\cdots<x_n=1.$$

其中 $x_k=\frac{k}{n}$ $(k=0,1,\cdots,n)$, 产生 $\Delta x_k=\frac{1}{n}$ $(k=1,2,\cdots,n)$. 这种等分的分割方式是最常见的.

1. 若选择了小区间的右端点 $\xi_k=x_k=\frac{k}{n}\in[x_{k-1},x_k]$, 则有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}f(\frac{k}{n})=\int_0^1 f(x)\ \mathrm{d}x.$$

2. 若选择了小区间的左端点 $\xi_k=x_{k-1}=\frac{k-1}{n}\in[x_{k-1},x_k]$, 则有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}f(\frac{k}{n}-\frac{1}{n})=\int_0^1 f(x)\ \mathrm{d}x.$$

3. 若选择了小区间的中点 $\xi_k=\frac{1}{2}(x_{k-1}+x_k)=\frac{2k-1}{2n}=\frac{k}{n}-\frac{1}{2n}\in[x_{k-1},x_k]$, 则有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}f(\frac{k}{n}-\frac{1}{2n})=\int_0^1 f(x)\ \mathrm{d}x.$$

4. 我们还可如此构造中间点: $\xi_k=\frac{k}{n}-\eta_k$. 这里只要令

$$0\le\eta_k\le\frac{1}{n},\ \forall(k=1,2,\cdots,n),$$

   就可以使得 $\xi_k\in[x_{k-1},x_k]$. 比如前面的例子中, $\eta_k\equiv0,\frac{1}{n},\frac{1}{2n}$.

所以凡是形如类似

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}f(\frac{k}{n}-\eta_k)$$

的极限, 都可以多多思考如何选择 ξ_k, 能使其变成一个黎曼和形式. 然后回定积分去计算.

例 2.10.5 将下列极限视作适当函数的黎曼和极限, 然后使用牛顿-莱布尼茨公式求该极限的值:

1. $\lim\limits_{n\to\infty}\sum\limits_{k=1}^{n}\frac{1}{n}\sin\frac{k}{n}$.

2. $\lim\limits_{n\to\infty}\sum\limits_{k=1}^{n}\frac{k^3}{n^4}=\lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{k=1}^{n}(\frac{k}{n})^3$.

3. $\lim\limits_{n\to\infty}\sum\limits_{k=1}^{n}\frac{1}{n+k}=\lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{k=1}^{n}\frac{n}{n+k}=\lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{k=1}^{n}\frac{1}{1+(\frac{k}{n})}$.

4. $\lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{k=1}^{n}\sin\left(\frac{k}{n}-\frac{1}{2\left(n^k\right)}\right)$. 容易验证 $0\le\eta_k=\frac{1}{2\left(n^k\right)}\le\frac{1}{n}$, $\forall(k=1,2,\cdots,n)$,

例 2.10.6 求定积分 $\int_{-1}^{2}|x|[x]\mathrm{d}x$.

证明. 记

$$f(x)=|x|[x]=\begin{cases}x, & -1\leqslant x<0;\\ 0, & 0\leqslant x<1;\\ x, & 1\leqslant x\leqslant 2.\end{cases}$$

该函数在区间 $[-1,2]$ 上只在 $x=1$ 处不连续, 且是第一类间断点 (左右极限存在但不相等), 由定积分的性质知道, $f(x)$ 在区间 $[-1,2]$ 上可积. 但是, 不能直接使用牛顿-莱布尼茨公式计算 $\int_{-1}^{2}f(x)\mathrm{d}x$. 我们将此定积分改成若干子区间上定积分之和:

$$\int_{-1}^{2}f(x)\mathrm{d}x=\int_{-1}^{0}f(x)\mathrm{d}x+\int_{0}^{1}f(x)\mathrm{d}x+\int_{1}^{2}f(x)\mathrm{d}x.$$

然后, 对上式右端各项分别使用牛顿-莱布尼茨公式, 得到

$$\int_{-1}^{2}|x|[x]\mathrm{d}x=\int_{-1}^{0}x\ \mathrm{d}x+\int_{0}^{1}0\cdot\ \mathrm{d}x+\int_{1}^{2}x\ \mathrm{d}x=\frac{1}{2}x^2\Big|_{-1}^{0}+0+\frac{1}{2}x^2\Big|_{1}^{2}=1.$$

□

上题中使用了事实: **分段函数的定积分是其若干个子区间上定积分之和.**

例 2.10.7 求由曲线 $y=x^2$ 及 $y=\sqrt{x}$ 所围成图形的面积.

证明. 首先作图. 很容易看出, 所求的面积为

$$S=\int_0^1\left(\sqrt{x}-x^2\right)\mathrm{d}x.$$

容易验证 $\frac{2}{3}x^{\frac{3}{2}}-\frac{1}{3}x^3$ 是 $\sqrt{x}-x^2$ 的一个原函数, 故

$$S=\int_0^1\left(\sqrt{x}-x^2\right)\mathrm{d}x=\left(\frac{2}{3}x^{\frac{3}{2}}-\frac{1}{3}x^3\right)\Big|_0^1=\frac{1}{3}.$$

在这个例子中, 被积函数的原函数是很容易求出的. 但是, 在很多情况下, 求一个给定函数的原函数就不这样简单了. 在下一章中, 我们将专门讨论求原函数的一般方法. □

利用定积分求被多个曲线包围 (不仅仅是曲面梯形了) 的图形的面积是常见的考题. 关键在于

- 画出正确的函数图像;
- 选择合适的积分变量, 是以横坐标 x 为积分变量还是纵坐标 y 为积分变量非常影响定积分计算难易度;
- 确定正确的被积函数和积分区间.

例 2.10.8 求由曲线 $y^2=4x$ 和直线 $4x-3y=4$ 所围成图形的面积.

证明. 先求出所给曲线与直线的交点. 交点坐标满足方程组

$$\begin{cases} y^2=4x \\ 4x-3y=4 \end{cases}$$

由此解得交点 $(4,4)$, $\left(\frac{1}{4},-1\right)$. 于是, 以纵坐标 y 为积分变量, 所求的面积为

$$S=\int_{-1}^4\left(\frac{3y+4}{4}-\frac{y^2}{4}\right)\mathrm{d}y=\frac{1}{4}\int_{-1}^4\left(3y+4-y^2\right)\mathrm{d}y=\frac{1}{4}\left(\frac{3}{2}y^2+4y-\frac{1}{3}y^3\right)\Big|_{-1}^4=\frac{125}{24}$$

注意, 此例题还可以横坐标 x 为积分变量, 则所求的面积还可以利用如下式子求得:

$$S=\int_0^{\frac{1}{4}}[\sqrt{4x}-(-\sqrt{4x})]\mathrm{d}x+\int_{\frac{1}{4}}^4\left(\sqrt{4x}-\frac{4x-4}{3}\right)\mathrm{d}x$$

其计算结果是一样的, 但显然后者的计算较为麻烦. □

最后, 我们指出: 在使用牛顿-莱布尼茨公式时要注意验证微积分基本定理的条件.

1. 比如, 在使用牛顿-莱布尼茨公式时, 要求被积函数在指定的积分区间上是连续的. 这样子定积分本身才能存在.
2. 又比如, $F'(x)=f(x)$ 在积分区间的内部开区间上处处成立, 不能有一点例外. 否则有可能导致错误结果.

例 2.10.9 验证 $\frac{1}{2}x^2 - \frac{1}{x}$ 是 $x + \frac{1}{x^2}$ 的一个原函数, 并计算定积分

$$\int_2^4 \left(x + \frac{1}{x^2}\right) \mathrm{d}x$$

试问: 等式

$$\int_{-1}^1 \left(x + \frac{1}{x^2}\right) \mathrm{d}x = \left(\frac{1}{2}x^2 - \frac{1}{x}\right)\Bigg|_{-1}^1$$

是否成立? 为什么?

证明. 因为

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{2}x^2 - \frac{1}{x}\right) = x + \frac{1}{x^2}$$

因此 $\frac{1}{2}x^2 - \frac{1}{x}$ 是 $x + \frac{1}{x^2}$ 的一个原函数. 接下来, 计算定积分

$$\int_2^1 \left(x + \frac{1}{x^2}\right) \mathrm{d}x = \left[\frac{1}{2}x^2 - \frac{1}{x}\right]_2^1 = -2.$$

对于等式

$$\int_{-1}^1 \left(x + \frac{1}{x^2}\right) \mathrm{d}x = \left(\frac{1}{2}x^2 - \frac{1}{x}\right)\Bigg|_{-1}^1$$

首先需要注意的是, 函数 $x + \frac{1}{x^2}$ 在 $x = 0$ 处不连续, 因为它都没有定义. 而其原函数在开区间 $(-1, 1)$ 上也不是连续的. 所以不能使用牛顿-莱布尼茨公式. 其实, 若从定义出发, $\int_{-1}^1 \left(x + \frac{1}{x^2}\right) \mathrm{d}x$ 实际上并不收敛. $\square$

第三章　积分的计算及应用

从微积分基本定理中, 我们看到了求原函数 (或者说不定积分) 的重要性. 本章将系统地介绍求不定积分的方法, 同时讨论定积分的计算及应用.

本章的内容多数属于技术性或方法性的, 概念性或理论性内容讨论较少. 对读者而言, 重要的是通过多做练习来掌握求积分 (不定积分和定积分) 的基本方法.

注意

其实本章几乎没有新的概念工具. 就是两个字 "刷题"! 就是干! 熟能生巧!
其实本章几乎没有新的概念工具. 就是两个字 "刷题"! 就是干! 熟能生巧!
其实本章几乎没有新的概念工具. 就是两个字 "刷题"! 就是干! 熟能生巧!

3.1　不定积分的换元法

换元法是常见的求不定积分的一种方法. 我们先讲第一换元法.

3.1.1　第一换元法 (凑微分法)

命题 3.1.1: 第一换元法 (凑微分法)

若 $F(x)$ 是 $f(x)$ 的一个原函数, 则

$$\int f(\varphi(x))\varphi'(x)\ \mathrm{d}x = \int f(\varphi(x))\ \mathrm{d}(\varphi(x)) = F(\varphi(x)) + C.$$

上式从左侧到右侧的转化过程就是使用第一换元法 (凑微分法) 求不定积分的过程. 其核心和难点是上面第一个等式, 即**把被积函数的哪一部分移入微分号 "d" 内** (这一过程就是叫做**凑微分**), 使不定积分化成

$$\int f(\varphi(x))\mathrm{d}(\varphi(x)) = \int f(y)\mathrm{d}y$$

(其中 $y = \varphi(x)$) **而** $f(y)\mathrm{d}y$ **恰好又是某个已知的函数** $F(y)$ **的微分**. 该方法属于换元法 (尽管使用过程中感受不明显), 是因为如果我们从另一个视角看: 设 $f(y)$ 具有原函数, $y = \varphi(x)$ 可导, 则有换元公式

$$\left[\int f(y)\mathrm{d}y\right]_{y=\varphi(x)} = \int f[\varphi(x)]\varphi'(x)\mathrm{d}x.$$

证明. 设 $F'(y) = f(y)$, 函数 $y = \varphi(x)$ 可导. 由复合函数求导公式, 我们得到

$$\frac{\mathrm{d}}{\mathrm{d}x} F(\varphi(x)) = f(\varphi(x))\varphi'(x),$$

于是有

$$\int f(\varphi(x))\varphi'(x)\ \mathrm{d}x = F(\varphi(x)) + C.$$

另外, 对于等式

$$\int f(\varphi(x))\varphi'(x)\ \mathrm{d}x = \int f(\varphi(x))\ \mathrm{d}(\varphi(x)),$$

我们利用了微分的定义:

$$\mathrm{d}(\varphi(x)) = \varphi'(x)\ \mathrm{d}x.$$

这里不用细扣定义. 尽管在不定积分的定义中, 我们从来没有说尾巴 $\mathrm{d}x$ 就是第一章微分概念中所谓的 "自变量的增量", 但是我们确实可以这么去看待它. 所以我们才能把 $\varphi'(x)\ \mathrm{d}x$ 当作一个新的微分 $\mathrm{d}(\varphi(x))$. □

例 3.1.1 求 $\displaystyle\int 2\cos 2x\ \mathrm{d}x$.

证明.

$$\int 2\cos 2x\ \mathrm{d}x = \int \cos 2x \cdot 2\ \mathrm{d}x = \int \cos 2x\ \mathrm{d}(2x) = \sin(2x) + C.$$

□

注意

一定要养成好习惯, 计算完不定积分后要求导验证结果! 尤其是考试中!
一定要养成好习惯, 计算完不定积分后要求导验证结果! 尤其是考试中!
一定要养成好习惯, 计算完不定积分后要求导验证结果! 尤其是考试中!

例 3.1.2 求 $\displaystyle\int \frac{1}{3+2x}\ \mathrm{d}x$.

证明.

$$\int \frac{1}{3+2x}\ \mathrm{d}x = \int \frac{1}{2} \cdot \frac{1}{3+2x}\ \mathrm{d}(3+2x) = \frac{1}{2}\ln|3+2x| + C.$$

□

例 3.1.3 求 $\displaystyle\int \sin(ax+b)\ \mathrm{d}x$, 其中 a, b 为常数且 $a \neq 0$.

证明.

$$\int \sin(ax+b)\ \mathrm{d}x = \frac{1}{a}\int \sin(ax+b)\ \mathrm{d}(ax+b) = -\frac{1}{a}\cos(ax+b) + C.$$

□

例 3.1.4 求 $\int 2x\mathrm{e}^{x^2}\,\mathrm{d}x$.

证明.

$$\int 2x\mathrm{e}^{x^2}\,\mathrm{d}x = \int \mathrm{e}^{x^2}\,\mathrm{d}(x^2) = \mathrm{e}^{x^2} + C.$$

□

例 3.1.5 求 $\int \frac{x\,\mathrm{d}x}{1+x^4}$.

证明. 这里的 $x\,\mathrm{d}x$ 可以看作 $\frac{1}{2}\,\mathrm{d}(x^2)$, 于是

$$\int \frac{x\,\mathrm{d}x}{1+x^4} = \frac{1}{2}\int \frac{\mathrm{d}(x^2)}{1+(x^2)^2} = \frac{1}{2}\arctan x^2 + C.$$

□

例 3.1.6 求 $\int \frac{\mathrm{e}^{3\sqrt{x}}}{\sqrt{x}}\,\mathrm{d}x$.

证明. 由于 $\mathrm{d}(\sqrt{x}) = \frac{1}{2}\frac{\mathrm{d}x}{\sqrt{x}}$, 因此,

$$\int \frac{\mathrm{e}^{3\sqrt{x}}}{\sqrt{x}}\,\mathrm{d}x = 2\int \mathrm{e}^{3\sqrt{x}}\,\mathrm{d}(\sqrt{x}) = \frac{2}{3}\int \mathrm{e}^{3\sqrt{x}}\,\mathrm{d}(3\sqrt{x}) = \frac{2}{3}\mathrm{e}^{3\sqrt{x}} + C.$$

□

例 3.1.7 求 $\int \frac{\mathrm{d}x}{x(1+2\ln x)}$.

证明.

$$\int \frac{\mathrm{d}x}{x(1+2\ln x)} = \int \frac{\mathrm{d}(\ln x)}{1+2\ln x} = \frac{1}{2}\int \frac{\mathrm{d}(1+2\ln x)}{1+2\ln x} = \frac{1}{2}\ln|1+2\ln x| + C.$$

注意, 被积函数函数中的分母部分的 $\ln x$ 告诉我们, 被积函数的定义域不包括 $x<0$ 的部分. 所以在凑微分时有 $\frac{1}{x} = \mathrm{d}(\ln x)$, 而不是 $\frac{1}{x} = \mathrm{d}(\ln|x|)$. □

例 3.1.8 求 $\int \tan x\,\mathrm{d}x$.

证明.

$$\int \tan x\,\mathrm{d}x = \int \frac{\sin x}{\cos x}\,\mathrm{d}x = -\int \frac{\mathrm{d}(\cos x)}{\cos x} = -\ln|\cos x| + C.$$

□

例 3.1.9 求 $\int \sec^6 x \, \mathrm{d}x$.

证明.

$$\begin{aligned}\int \sec^6 x \, \mathrm{d}x &= \int \left(\sec^2 x\right)^2 \sec^2 x \, \mathrm{d}x \\ &= \int \left(1+\tan^2 x\right)^2 \, \mathrm{d}(\tan x) \\ &= \int \left(1+2\tan^2 x+\tan^4 x\right) \, \mathrm{d}(\tan x) \\ &= \tan x + \frac{2}{3}\tan^3 x + \frac{1}{5}\tan^5 x + C.\end{aligned}$$

□

例 3.1.10 求 $\int \tan^5 x \sec^3 x \, \mathrm{d}x$.

证明.

$$\begin{aligned}\int \tan^5 x \sec^3 x \, \mathrm{d}x &= \int \tan^4 x \sec^2 x \cdot \sec x \tan x \, \mathrm{d}x \\ &= \int \left(\sec^2 x - 1\right)^2 \sec^2 x \, \mathrm{d}(\sec x) \\ &= \int \left(\sec^6 x - 2\sec^4 x + \sec^2 x\right) \, \mathrm{d}(\sec x) \\ &= \frac{1}{7}\sec^7 x - \frac{2}{5}\sec^5 x + \frac{1}{3}\sec^3 x + C.\end{aligned}$$

□

例 3.1.11 求 $\int \frac{\mathrm{d}x}{a^2 - x^2} \quad (a > 0)$.

证明. 我们先将被积函数化成两个简单函数之和, 再求不定积分. 首先观察到等式

$$\frac{1}{a^2-x^2} = \frac{1}{2a}\left(\frac{1}{a-x} + \frac{1}{a+x}\right).$$

所以,

$$\begin{aligned}\int \frac{\mathrm{d}x}{a^2-x^2} &= \frac{1}{2a}\int \left(\frac{1}{a-x} + \frac{1}{a+x}\right) \, \mathrm{d}x \\ &= \frac{1}{2a}\int \frac{1}{a-x} \, \mathrm{d}x + \frac{1}{2a}\int \frac{1}{a+x} \, \mathrm{d}x \\ &= \frac{-1}{2a}\int \frac{1}{a-x} \, \mathrm{d}(a-x) + \frac{1}{2a}\int \frac{1}{a+x} \, \mathrm{d}(a+x) \\ &= \frac{-1}{2a}\ln|a-x| + \frac{1}{2a}\ln|a+x| + C \\ &= \frac{1}{2a}\ln\left|\frac{a+x}{a-x}\right| + C.\end{aligned}$$

□

例 3.1.12 求 $\displaystyle\int \frac{\mathrm{d}x}{a^2+x^2} \quad (a \neq 0)$.

证明. 先提出分母中的因子 a^2, 再用凑微分法:

$$\int \frac{\mathrm{d}x}{a^2+x^2} = \frac{1}{a^2}\int \frac{\mathrm{d}x}{1+\left(\frac{x}{a}\right)^2} = \frac{1}{a}\int \frac{1}{1+\left(\frac{x}{a}\right)^2}\,\mathrm{d}\left(\frac{x}{a}\right) = \frac{1}{a}\arctan\frac{x}{a} + C.$$

□

例 3.1.13 求 $\displaystyle\int \frac{\mathrm{d}x}{\sqrt{a^2-x^2}} \quad (a > 0)$.

证明.

$$\int \frac{\mathrm{d}x}{\sqrt{a^2-x^2}} = \int \frac{\mathrm{d}x}{\sqrt{a^2\left[1-\left(\frac{x}{a}\right)^2\right]}} = \int \frac{\frac{1}{a}\,\mathrm{d}x}{\sqrt{1-\left(\frac{x}{a}\right)^2}} = \int \frac{1}{\sqrt{1-\left(\frac{x}{a}\right)^2}}\,\mathrm{d}\left(\frac{x}{a}\right) = \arcsin\frac{x}{a} + C.$$

□

很多不定积分都可化成例 3.1.12 或例 3.1.13 中不定积分的形式, 记住这两个不定积分的结果是有益的.

例 3.1.14 求 $\displaystyle\int \csc x\,\mathrm{d}x = \int \frac{\mathrm{d}x}{\sin x}$.

证明.

$$\begin{aligned}
\int \frac{\mathrm{d}x}{\sin x} &= \int \frac{\mathrm{d}x}{2\sin\frac{x}{2}\cos\frac{x}{2}} \\
&= \int \frac{\mathrm{d}x}{2\tan\frac{x}{2}\cos^2\frac{x}{2}} \\
&= \int \frac{\mathrm{d}\left(\frac{x}{2}\right)}{\tan\frac{x}{2}\cos^2\frac{x}{2}} \quad (\text{第一次凑微分}) \\
&= \int \frac{\mathrm{d}\left(\tan\frac{x}{2}\right)}{\tan\frac{x}{2}} \quad (\text{第二次凑微分}) \\
&= \ln\left|\tan\frac{x}{2}\right| + C.
\end{aligned}$$

或者另一种解法:

$$\int \frac{\mathrm{d}x}{\sin x} = \int \frac{\sin x\,\mathrm{d}x}{\sin^2 x} = -\int \frac{\mathrm{d}(\cos x)}{1-\cos^2 x} = \frac{1}{2}\ln\left|\frac{1-\cos x}{1+\cos x}\right| + C$$

最后一步用到了例 3.1.11 的结果. □

读者或许已发现, 上题中两种解法的最终结果形式上不尽相同, 但是利用半角公式很容易证实它们是完全一致的.

例 3.1.15 求 $\displaystyle\int \sec x\,\mathrm{d}x$.

证明.

$$\int \sec x \, \mathrm{d}x = \int \frac{1}{\cos x} \, \mathrm{d}x = \int \frac{\cos x}{\cos^2 x} \, \mathrm{d}x = \int \frac{1}{1-\sin^2 x} \, \mathrm{d}(\sin x).$$

利用例 3.1.11 的结果, 得

$$\int \sec x \, \mathrm{d}x = \frac{1}{2} \ln \left| \frac{1+\sin x}{1-\sin x} \right| + C.$$

我们可以做进一步化简 (这不是必须的工作, 但做了肯定是加分项):

$$\begin{aligned}\int \sec x \, \mathrm{d}x &= \frac{1}{2} \ln \left| \frac{(1+\sin x)(1+\sin x)}{(1-\sin x)(1+\sin x)} \right| + C \\ &= \frac{1}{2} \ln \left| \frac{(1+\sin x)^2}{\cos^2 x} \right| + C \\ &= \ln \left| \frac{1+\sin x}{\cos x} \right| + C \\ &= \ln |\sec x + \tan x| + C.\end{aligned}$$

□

常见的凑微分总结. (以下内容本质上和微分基本公式和导数基本公式别无二致) 下面中的 f 的原函数则需要读者自己平时多积累了, 这样才能得到最终结果.

1. $\int f(ax+b) \, \mathrm{d}x = \frac{1}{a} \int f(ax+b) \, \mathrm{d}(ax+b) \quad (a \neq 0)$
2. $\int f(ax^n+b)x^{n-1} \, \mathrm{d}x = \frac{1}{na} \int f(ax^n+b) \, \mathrm{d}(ax^n+b) \quad (a \neq 0, n \neq 0)$
3. $\int \frac{f(\ln x)}{x} \, \mathrm{d}x = \int f(\ln x) \, \mathrm{d}(\ln x)$
4. $\int f(\frac{1}{x}) \frac{\mathrm{d}x}{x^2} = -\int f(\frac{1}{x}) \, \mathrm{d}(\frac{1}{x})$
5. $\int f(\sqrt{x}) \frac{\mathrm{d}x}{\sqrt{x}} = 2 \int f(\sqrt{x}) \, \mathrm{d}(\sqrt{x})$
6. $\int f(a^x)a^x \, \mathrm{d}x = \frac{1}{\ln a} \int f(a^x) \, \mathrm{d}(a^x) \quad (a > 0, a \neq 1)$
7. $\int f(e^x)e^x \, \mathrm{d}x = \int f(e^x) \, \mathrm{d}(e^x)$
8. $\int f(\sin x) \cos x \, \mathrm{d}x = \int f(\sin x) \, \mathrm{d}(\sin x)$
9. $\int f(\cos x) \sin x \, \mathrm{d}x = -\int f(\cos x) \, \mathrm{d}(\cos x)$
10. $\int f(\tan x) \sec^2 x \, \mathrm{d}x = \int f(\tan x) \, \mathrm{d}(\tan x)$
11. $\int f(\cot x) \csc^2 x \, \mathrm{d}x = -\int f(\cot x) \, \mathrm{d}(\cot x)$
12. $\int f(\sec x) \sec x \tan x \, \mathrm{d}x = \int f(\sec x) \, \mathrm{d}(\sec x)$
13. $\int f(\csc x) \csc x \cot x \, \mathrm{d}x = -\int f(\csc x) \, \mathrm{d}(\csc x)$

14. $\displaystyle\int \frac{f(\arcsin x)}{\sqrt{1-x^2}}\ \mathrm{d}x = \int f(\arcsin x)\ \mathrm{d}(\arcsin x)$

15. $\displaystyle\int \frac{f(\arctan x)}{1+x^2}\ \mathrm{d}x = \int f(\arctan x)\ \mathrm{d}(\arctan x)$

16. $\displaystyle\int \frac{f(\operatorname{arccot} x)}{1+x^2}\ \mathrm{d}x = -\int f(\operatorname{arccot} x)\ \mathrm{d}(\operatorname{arccot} x)$

17. $\displaystyle\int \frac{f(\arctan \frac{1}{x})}{1+x^2}\ \mathrm{d}x = -\int f(\arctan \frac{1}{x})\ \mathrm{d}(\arctan \frac{1}{x})$

下面再举一些积分的例子, 它们的被积函数中含有三角函数, 在计算这种积分的过程中, 往往要用到一些三角恒等式. (本章的题目严重依赖于高中学习的各类三角恒等式, 请读者自行复习.)

例 3.1.16 求 $\displaystyle\int \sin^3 x \cos^2 x\ \mathrm{d}x$.

证明.

$$\begin{aligned}
\int \sin^3 x \cos^2 x\ \mathrm{d}x &= \int \sin^2 x \cos^2 x \cdot \sin x\ \mathrm{d}x \\
&= -\int \left(1-\cos^2 x\right) \cos^2 x\ \mathrm{d}(\cos x) \\
&= \int \left(\cos^4 x - \cos^2 x\right)\ \mathrm{d}(\cos x) \\
&= \frac{1}{5}\cos^5 x - \frac{1}{3}\cos^3 x + C.
\end{aligned}$$

□

例 3.1.17 求 $\displaystyle\int \sin^2 x \cos^4 x\ \mathrm{d}x$.

证明.

$$\begin{aligned}
\int \sin^2 x \cos^4 x\ \mathrm{d}x &= \frac{1}{8}\int (1-\cos 2x)(1+\cos 2x)^2\ \mathrm{d}x \\
&= \frac{1}{8}\int \left(1+\cos 2x - \cos^2 2x - \cos^3 2x\right)\ \mathrm{d}x \\
&= \frac{1}{8}\int \left(\cos 2x - \cos^3 2x\right)\ \mathrm{d}x + \frac{1}{8}\int \left(1-\cos^2 2x\right)\ \mathrm{d}x \\
&= \frac{1}{8}\int \sin^2 2x \cdot \frac{1}{2}\ \mathrm{d}(\sin 2x) + \frac{1}{8}\int \frac{1}{2}(1-\cos 4x)\ \mathrm{d}x \\
&= \frac{1}{48}\sin^3 2x + \frac{x}{16} - \frac{1}{64}\sin 4x + C.
\end{aligned}$$

□

例 3.1.18 求 $\displaystyle\int \cos^2 x\ \mathrm{d}x$.

证明.

$$\begin{aligned}\int \cos^2 x \, \mathrm{d}x &= \int \frac{1+\cos 2x}{2} \, \mathrm{d}x \\ &= \frac{1}{2}\left(\int \, \mathrm{d}x + \int \cos 2x \, \mathrm{d}x\right) \\ &= \frac{1}{2}\int \, \mathrm{d}x + \frac{1}{4}\int \cos 2x \, \mathrm{d}(2x) \\ &= \frac{x}{2} + \frac{\sin 2x}{4} + C.\end{aligned}$$

□

请读者自行求 $\int \sin^2 x \, \mathrm{d}x$.

例 3.1.19 求 $\int \sin^2 x \cos^5 x \, \mathrm{d}x$.

证明.

$$\begin{aligned}\int \sin^2 x \cos^5 x \, \mathrm{d}x &= \int \sin^2 x \cos^4 x \cos x \, \mathrm{d}x \\ &= \int \sin^2 x \left(1-\sin^2 x\right)^2 \, \mathrm{d}(\sin x) \\ &= \int \left(\sin^2 x - 2\sin^4 x + \sin^6 x\right) \, \mathrm{d}(\sin x) \\ &= \frac{1}{3}\sin^3 x - \frac{2}{5}\sin^5 x + \frac{1}{7}\sin^7 x + C.\end{aligned}$$

□

例 3.1.20 求 $\int \cos 3x \cos 2x \, \mathrm{d}x$.

证明. 利用三角函数的积化和差公式

$$\cos A \cos B = \frac{1}{2}[\cos(A-B) + \cos(A+B)],$$

得

$$\cos 3x \cos 2x = \frac{1}{2}(\cos x + \cos 5x),$$

于是

$$\begin{aligned}\int \cos 3x \cos 2x \, \mathrm{d}x &= \frac{1}{2}\int (\cos x + \cos 5x) \, \mathrm{d}x \\ &= \frac{1}{2}\left(\int \cos x \, \mathrm{d}x + \frac{1}{5}\int \cos 5x \, \mathrm{d}(5x)\right) \\ &= \frac{1}{2}\sin x + \frac{1}{10}\sin 5x + C.\end{aligned}$$

□

例 3.1.21 求 $\int \sin nx \sin mx \, \mathrm{d}x \quad (0 < n < m)$.

证明. 首先应将被积函数化作两个三角函数的和或差, 然后分别求其不定积分. 事实上, 有

$$\begin{aligned}\int \sin nx \sin mx \, \mathrm{d}x &= \frac{1}{2}\int [\cos((m-n)x) - \cos((m+n)x)] \, \mathrm{d}x \\ &= \frac{1}{2}\int \cos((m-n)x) \, \mathrm{d}x - \frac{1}{2}\int \cos((m+n)x) \, \mathrm{d}x \\ &= \frac{1}{2}\left\{\frac{1}{m-n}\int \cos((m-n)x) \, \mathrm{d}[(m-n)x] - \frac{1}{m+n}\int \cos((m+n)x) \, \mathrm{d}[(m+n)x]\right\} \\ &= \frac{1}{2}\left[\frac{1}{m-n}\sin((m-n)x) - \frac{1}{m+n}\sin((m+n)x)\right] + C.\end{aligned}$$

□

3.1.2 第二换元法 (真 ● 换元法)

下面将介绍的第二类换元法是: 适当地选择变量代换 $x = \varphi(t)$, 将积分 $\int f(x) \, \mathrm{d}x$ 化为积分 $\int f[\varphi(t)]\varphi'(t) \, \mathrm{d}t$. 这是另一种形式的变量代换, 换元公式可表达为

$$\int f(x) \, \mathrm{d}x = \int f[\varphi(t)]\varphi'(t) \, \mathrm{d}t.$$

这公式的成立是需要一定条件的. 首先, 等式右边的不定积分要存在, 即 $f[\varphi(t)]\varphi'(t)$ 有原函数; 其次, $\int f[\varphi(t)]\varphi'(t) \, \mathrm{d}t$ 求出后必须用 $x = \varphi(t)$ 的反函数 $t = \varphi^{-1}(x)$ 代回去, 为了保证这反函数存在而且是可导的, 我们假定**直接函数 $x = \varphi(t)$ 在 t 的某一个区间（这区间和所考虑的 x 的积分区间相对应）上是单调的，可导的，并且 $\varphi'(t) \neq 0$.** 归纳上述, 我们给出下面的定理:

定理 3.1.1: 不定积分的第二换元法

欲求不定积分

$$\int f(x) \, \mathrm{d}x \quad (\text{其中 } f \text{ 的定义域是 } x \in D \subseteq \mathbf{R}).$$

现对 x 进行换元操作, 即找到某个函数 $x = \varphi(t)$, 其定义域为 $E \subseteq \mathbf{R}$, 且值域正好等于 D, 即 $\varphi(E) = D$. 此时, 将 $x = \varphi(t)$ 代入上式中, 得到新的关于变量 t 的不定积分:

$$\int f[\varphi(t)] \, \mathrm{d}(\varphi(t)) = \int f[\varphi(t)]\varphi'(t) \, \mathrm{d}t.$$

如果我们选择的换元函数 $x = \varphi(t)$ 满足:

1. 对于 $t \in E$, 函数 $x = \varphi(t)$ 是单调的, 可导, 并且 $\varphi'(t) \neq 0$;
2. 新的关于变量 t 的不定积分存在, 即 $f[\varphi(t)]\varphi'(t)$ 具有原函数;

那么就有

$$\int f(x) \, \mathrm{d}x = \left[\int f[\varphi(t)]\varphi'(t) \, \mathrm{d}t\right]_{t=\varphi^{-1}(x)},$$

其中 $t = \varphi^{-1}(x)$ 是 $x = \varphi(t)$ 的反函数. 注意, 最后在求得关于 t 的原函数后还应当将 t 换成 $\varphi^{-1}(x)$.

证明. 设 $f[\varphi(t)]\varphi'(t)$ 的原函数为 $\Phi(t)$, 记 $\Phi\left[\varphi^{-1}(x)\right] = F(x)$, 利用复合函数及反函数的求导法则, 得到

$$F'(x) = \frac{\mathrm{d}\Phi}{\mathrm{d}t} \cdot \frac{\mathrm{d}t}{\mathrm{d}x} = f[\varphi(t)]\varphi'(t) \cdot \frac{1}{\varphi'(t)} = f[\varphi(t)] = f(x),$$

即 $F(x)$ 是 $f(x)$ 的原函数. 注意, 本定理中的条件 "$x=\varphi(t)$ 是单调的, 可导, 并且 $\varphi'(t)\neq 0$" 的目的在于使用反函数求导法则. 最后有,

$$\int f(x)\ \mathrm{d}x = F(x)+C = \Phi\left[\varphi^{-1}(x)\right]+C = \left[\int f[\varphi(t)]\varphi'(t)\ \mathrm{d}t\right]_{t=\varphi^{-1}(x)}.$$

证毕. □

在使用第二换元法时, 我们要求变量替换函数 $x=\varphi(t)$ 一定有反函数, 并且在最后的结果中要将 t 换成 x 的函数.

例 3.1.22 求 $\displaystyle\int \frac{\mathrm{d}x}{\sqrt{x+1}+1}$.

证明. 根据被积函数的表达式, 可知 x 的取值范围/定义域是 $(-1,+\infty)$. 令

$$x=\varphi(t)=t^2-1,$$

且只考虑定义域 $t>0$. 显然, 画图可知, 对于在 $t>0$ 的区间上, 函数 $x=t^2-1$ 是单调可导的, 且导数 $2t\neq 0$. 并且值域正好就是原来的 x 的取值范围 $(-1,+\infty)$. 这一换元函数 $x=t^2-1$ 满足第二换元法的所有使用条件. 此时, 将 $x=t^2-1$ 代入原来的不定积分可得: (下式第一个等号中, 有 $\sqrt{t^2}=t$ 而不是 $\sqrt{t^2}=|t|$, 这是因为我们考虑是 $t>0$ 范围)

$$\int \frac{\mathrm{d}x}{\sqrt{x+1}+1} = \int \frac{\mathrm{d}(t^2-1)}{t+1} = 2\int \frac{t\ \mathrm{d}t}{t+1} = 2\int\left(1-\frac{1}{t+1}\right)\ \mathrm{d}t = 2(t-\ln(t+1))+C.$$

最后, 我们应当将变量 t 再换成 x 的函数, 即将 $x=t^2-1$ 的反函数 $t=\sqrt{x+1}$ 代入上式, 得到最终结果

$$\int \frac{\mathrm{d}x}{\sqrt{x+1}+1} = 2(\sqrt{x+1}-\ln(\sqrt{x+1}+1))+C.$$

□

上题中, 我们是怎么想到使用 $x=t^2-1$, 而不是别的换元函数呢? 我们如何能未卜先知? 常见的技巧就是令 t 等于被积函数中的某个无理式, 比如, 上面的 $t=\sqrt{x+1}$. 而这个式子 "通常" (这仅是经验之谈! 例外也很多) 就是 $t=\varphi^{-1}(x)$ 的表达式, 然后我们再反推出 $x=\varphi(t)$ 的表达式. 但是, 必须验证这个换元函数是否符合定理 3.1.1 的条件, 否则非常非常容易出错.

结论 3.1.1: 构建换元函数 $x=\varphi(t)$ 的要点

总结一下构建一个合理的换元函数 $x=\varphi(t)$ 的要点:

1. 如果被积函数的中 x 的定义域是 D, 应当人为限定 $x=\varphi(t)$ 的定义为 E 使得 $\varphi(E)=D$.
2. 在这个定义域 E 上 $x=\varphi(t)$ 应当单调可导, 且对于所有 t, 恒有 $\varphi'(t)\neq 0$. **我们的目标就是建立变量 x 的 (自然) 定义域和另外一个变量 t 在某个范围上的双射变换.** 函数 $x=\varphi(t)$ 和 $t=\varphi^{-1}(x)$ 是它们的沟通桥梁. (可以说, x 的定义域是 t 的值域, t 的定义域是 x 的值域.)
3. 满足前面 2 点的换元函数往往不止一种. 而一个好的变量替换 $x=\varphi(t)$, 其目的在于使被积函数由原来的 $f(x)$ 换成 $f(\varphi(t))\varphi'(t)$, 而后者较易于求不定积分.

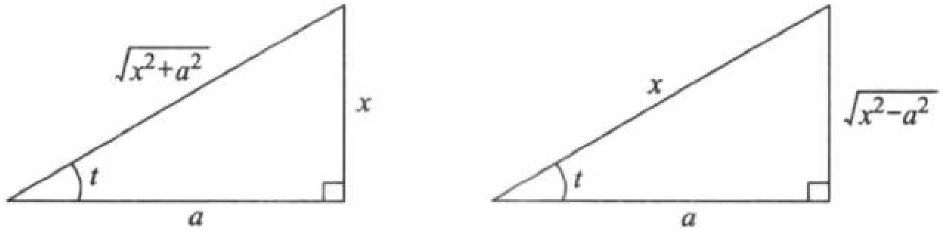


图 3.1: 被积函数含有 $\sqrt{x^2 \pm a^2}$ 时, 可以画出辅助三角形来帮助使用第二类换元法.

例 3.1.23 求 $\int \sqrt{a^2 - x^2}\, \mathrm{d}x \quad (a > 0)$.

证明. 首先由被积函数的表达式可知, x 的取值范围是 $(-a, a)$. 考虑换元函数

$$x = a \sin t, \quad -\frac{\pi}{2} < t < \frac{\pi}{2}.$$

显然, 在该定义域 $(-\frac{\pi}{2} < t < \frac{\pi}{2})$ 上, 函数 $x = a \sin t$ 关于 t 是单调可导的, 且恒有非零导数. 并且该函数的值域正好就是原来的 x 的取值范围 $(-a, a)$. 故这是一个合法的换元函数. 将 $x = a \sin t$ 代入可得

$$\begin{aligned}
\int \sqrt{a^2 - x^2}\, \mathrm{d}x &= \int \sqrt{a^2 - a^2 \sin^2 t} \cdot a \cos t\, \mathrm{d}t \\
&= a^2 \int \cos^2 t\, \mathrm{d}t \quad (\text{ 对于 } -\frac{\pi}{2} < t < \frac{\pi}{2}, \cos t > 0) \\
&= \frac{a^2}{2} \int (1 + \cos 2t)\, \mathrm{d}t \\
&= \frac{a^2}{2} \left(t + \frac{1}{2} \sin 2t \right) + C \\
&= \frac{a^2}{2} (t + \sin t \cos t) + C \\
&= \frac{a^2}{2} \left[\arcsin \frac{x}{a} + \frac{x}{a} \sqrt{1 - \left(\frac{x}{a} \right)^2} \right] + C \\
&= \frac{a^2}{2} \arcsin \frac{x}{a} + \frac{x}{2} \sqrt{a^2 - x^2} + C.
\end{aligned}$$

注意, $x = a \sin t \ (-\frac{\pi}{2} < t < \frac{\pi}{2})$ 的逆函数是 $t = \arcsin \frac{x}{a} \ (-a < x < a)$. □

例 3.1.24 求 $\int \frac{\mathrm{d}x}{\sqrt{x^2 + a^2}} \quad (a > 0)$.

证明. 首先由被积函数的表达式可知, x 的取值范围是 $\mathbf{R}$. 考虑换元函数

$$x = a \tan t, \quad -\frac{\pi}{2} < t < \frac{\pi}{2}.$$

显然, 在该定义域 $(-\frac{\pi}{2} < t < \frac{\pi}{2})$ 上, 函数 $x = a \tan t$ 关于 t 是单调可导的, 且恒有非零导数. 并且该函数的值域正好就是原来的 x 的取值范围 $\mathbf{R}$. 故这是一个合法的换元函数. 将 $x = a \tan t$ 代入可得

$$\int \frac{\mathrm{d}x}{\sqrt{x^2+a^2}} = \int \frac{a\sec^2 t\ \mathrm{d}t}{\sqrt{a^2\tan^2 t + a^2}} = \int \frac{a\sec^2 t\ \mathrm{d}t}{\sqrt{a^2\sec^2 t}} = \int \sec t\ \mathrm{d}t.$$

利用例 3.1.15 的结果, 得

$$\int \frac{\mathrm{d}x}{\sqrt{x^2+a^2}} = \int \sec t\ \mathrm{d}t = \ln|\sec t + \tan t| + C$$

由图 3.1 所示的**辅助三角形**可以看出, 当 $\tan t = \frac{x}{a}$ 时, $\sec t = \frac{\sqrt{x^2+a^2}}{a}$. 代入上式, 得

$$\int \frac{\mathrm{d}x}{\sqrt{x^2+a^2}} = \ln\left|\frac{x}{a} + \frac{\sqrt{x^2+a^2}}{a}\right| + C = \ln\left|x + \sqrt{x^2+a^2}\right| + C.$$

在上式最后一步中, 我们已将 $-\ln a$ 合并到任意常数 C 中. □

上题中最后一步, 将换元函数的逆函数 $t = \varphi^{-1}(x)$ 放回积分结果时, 可以不用那么死板, 因为三角恒等式的诸多关系, 画出辅助三角形可以一步到位. 否则, 将逆函数 $t = \arctan\frac{x}{a}$ 代回得到,

$$\int \frac{\mathrm{d}x}{\sqrt{x^2+a^2}} = \int \sec t\ \mathrm{d}t = \ln|\sec(\arctan\frac{x}{a}) + \frac{x}{a}| + C.$$

除非你能记得住公式 $\cos(\arctan(x)) = \frac{1}{\sqrt{1+x^2}}$, 否则上式就很难继续化简了. 严格说, 上式的结果也是对, 可以选择不做进一步的化简. 这并没有做错题目. 可倘若, 考试的出题人更一步要求你最终的积分结果不能出现任何三角函数符号呢?

例 3.1.25 求 $\int \frac{\mathrm{d}x}{\sqrt{x^2-a^2}}\quad (a>0)$.

证明. 马上看出, x 的取值范围是 $x > a$ 与 $x < -a$. 分两种情况讨论.

1. 先考虑在 $x > a$ 上的不定积分. 考虑换元函数

$$x = a\sec t, \quad 0 < t < \frac{\pi}{2}.$$

显然, 在该定义域上, 函数 $x = a\sec t$ 关于 t 是单调可导的, 且恒有非零导数. 并且该函数的值域正好就是原来的 x 的取值范围 $(a, +\infty)$. 故这是一个合法的换元函数. 将 $x = a\sec t$ 代入可得

$$\begin{aligned}\int \frac{\mathrm{d}x}{\sqrt{x^2-a^2}} &= a\int \frac{1}{\sqrt{a^2\sec^2 t - a^2}}\cdot \sec t\tan t\ \mathrm{d}t \\ &= a\int \frac{1}{\sqrt{a^2\tan^2 t}}\cdot \sec t\tan t\ \mathrm{d}t \\ &= \int \frac{\mathrm{d}t}{\cos t} \quad (\text{对于}\ -0 < t < \frac{\pi}{2}, \tan t > 0) \\ &= \ln|\sec t + \tan t| + C. \quad (\text{利用例 13})\end{aligned}$$

这时令 $x = a\sec t$, 则 t 的变化范围为 $\left(0, \frac{\pi}{2}\right)$, 并有

$$\begin{aligned}\int \frac{\mathrm{d}x}{\sqrt{x^2-a^2}} &= a\int \frac{1}{\sqrt{a^2\tan^2 t}}\cdot \frac{\sin t}{\cos^2 t}\ \mathrm{d}t \\ &= \int \frac{\mathrm{d}t}{\cos t} \quad (\text{利用例 3.1.15}) \\ &= \ln|\sec t + \tan t| + C.\end{aligned}$$

由图 3.1 所示的辅助三角形可以看出, 当 $\sec t = \dfrac{x}{a}$ 时, $\tan t = \dfrac{\sqrt{x^2-a^2}}{a}$. 因此, 我们有

$$\int \frac{\mathrm{d}x}{\sqrt{x^2-a^2}} = \ln\left|\frac{x}{a}+\frac{\sqrt{x^2-a^2}}{a}\right| + C = \ln\left|x+\sqrt{x^2-a^2}\right| + C.$$

2. 再考虑在 $x<-a$ 上的不定积分. 考虑换元函数 $x=-a\sec t$, 这时 t 的定义域同样为 $\left(0,\dfrac{\pi}{2}\right)$, 用完全类似的方法可以求得, 在这种情况下上述结果仍然成立.

□

在前面的三个例题中, 被积函数都是含有 $\sqrt{x^2\pm a^2}$ 的无理式. 这种类型的积分题目特别适合使用第二类换元法, 且一开始就可以用勾股定理做辅助三角形, 通过观察就能知道应该选择什么换元函数.

总结前面两个例题的结果, 我们有

$$\int \frac{\mathrm{d}x}{\sqrt{x^2\pm a^2}} = \ln\left|x+\sqrt{x^2\pm a^2}\right| + C \quad (a>0).$$

这个公式今后会经常用到, 建议读者记住它.

例 3.1.26 求 $\displaystyle\int \frac{2x}{\sqrt{x^2+2x+26}}\,\mathrm{d}x$.

证明.

$$\begin{aligned}
\int \frac{2x}{\sqrt{x^2+2x+26}}\,\mathrm{d}x &= \int \frac{2x+2-2}{\sqrt{x^2+2x+26}}\,\mathrm{d}x \\
&= \int \frac{2x+2}{\sqrt{x^2+2x+26}}\,\mathrm{d}x - 2\int \frac{1}{\sqrt{x^2+2x+26}}\,\mathrm{d}x \\
&= \int \frac{\mathrm{d}\left(x^2+2x+26\right)}{\sqrt{x^2+2x+26}} - 2\int \frac{1}{\sqrt{(x+1)^2+5^2}}\,\mathrm{d}x \\
&= 2\sqrt{x^2+2x+26} - 2\ln\left|x+1+\sqrt{(x+1)^2+5^2}\right| + C \\
&= 2\sqrt{x^2+2x+26} - 2\ln\left|x+1+\sqrt{x^2+2x+26}\right| + C.
\end{aligned}$$

□

例 3.1.27 求 $\displaystyle\int \frac{\sqrt{4-x^2}}{x}\,\mathrm{d}x$.

证明. 首先由被积函数的表达式可知, x 的取值范围是 $(-2,0)\cup(0,2)$. 我们使用两种不同的换元方法.

第一种换元函数. 因为被积函数是含有 $\sqrt{x^2\pm a^2}$ 的无理式, 用勾股定理做辅助三角形, 通过观察可知应该使用换元函数:

$$x = 2\sin t, \quad -\frac{\pi}{2} < t < \frac{\pi}{2}, t \neq 0.$$

显然, 在该定义域 $(-\dfrac{\pi}{2}<t<\dfrac{\pi}{2}, t\neq 0)$ 上, 函数 $x=2\sin t$ 关于 t 是单调可导的, 且恒有非零导数. 并且该函数的值域正好就是原来的 x 的取值范围 $(-2,0)\cup(0,2)$. 则通过辅助三角形易知 $\sqrt{4-x^2}=2\cos t$, $\mathrm{d}x=$

$2\cos t\ \mathrm{d}t$, 从而

$$\begin{aligned}\int \frac{\sqrt{4-x^2}}{x}\ \mathrm{d}x &= \int \frac{2\cos t}{2\sin t}\cdot 2\cos t\ \mathrm{d}t = \int \frac{2\cos^2 t}{\sin t}\ \mathrm{d}t = 2\int \frac{1-\sin^2 t}{\sin t}\ \mathrm{d}t \\ &= 2\int \frac{1}{\sin t}\ \mathrm{d}t - 2\int \sin t\ \mathrm{d}t \\ &= \ln\left|\frac{1-\cos t}{1+\cos t}\right| + 2\cos t + C \\ &= \ln\left|\frac{2-2\cos t}{2+2\cos t}\right| + 2\cos t + C \\ &= \ln\left|\frac{2-\sqrt{4-x^2}}{2+\sqrt{4-x^2}}\right| + \sqrt{4-x^2} + C.\end{aligned}$$

第二种换元函数.（下面过程是简便但不严谨的计算步骤. 尤其是省略分析换元函数的正确性. 对此, 请参考下一题的注意事项.） 或者用另外一种换元方法来求. 令 $t=\sqrt{4-x^2}$, 则 $x^2=4-t^2$, $2x\ \mathrm{d}x=-2t\ \mathrm{d}t$, 从而

$$\begin{aligned}\int \frac{\sqrt{4-x^2}}{x}\ \mathrm{d}x &= \int \frac{\sqrt{4-x^2}}{x^2}x\ \mathrm{d}x = \int \frac{t}{4-t^2}(-t\ \mathrm{d}t) = \int \frac{-t^2}{4-t^2}\ \mathrm{d}t \\ &= \int \frac{4-t^2-4}{4-t^2}\ \mathrm{d}t = \int\ \mathrm{d}t - 4\int \frac{1}{4-t^2}\ \mathrm{d}t.\end{aligned}$$

利用例 3.1.12 的结果, 得

$$\int \frac{\sqrt{4-x^2}}{x}\ \mathrm{d}x = t - 4\cdot\frac{1}{4}\ln\left|\frac{2+t}{2-t}\right| + C = \sqrt{4-x^2} - \ln\left|\frac{2+\sqrt{4-x^2}}{2-\sqrt{4-x^2}}\right| + C.$$

□

下面我们通过例子来介绍一种也很有用的代换 — **倒代换**, 利用它常可消去被积函数的分母中的变量因子 x.

例 3.1.28 求 $\displaystyle\int \frac{\sqrt{a^2-x^2}}{x^4}\ \mathrm{d}x \quad (a\neq 0)$.

证明. 设 $x=\dfrac{1}{t}$, 则 $\mathrm{d}x=-\dfrac{\mathrm{d}t}{t^2}$, 于是

$$\begin{aligned}\int \frac{\sqrt{a^2-x^2}}{x^4}\ \mathrm{d}x &= \int \frac{\sqrt{a^2-\frac{1}{t^2}}\cdot\left(-\frac{\mathrm{d}t}{t^2}\right)}{\frac{1}{t^4}} = -\int \sqrt{a^2-\frac{1}{t^2}}\cdot t^2\ \mathrm{d}t = -\int \sqrt{a^2t^2-1}\cdot\frac{t^2}{|t|}\ \mathrm{d}t \\ &= -\int \left(a^2t^2-1\right)^{\frac{1}{2}}|t|\ \mathrm{d}t,\end{aligned}$$

当 $x>0$ 时, 有

$$\begin{aligned}\int \frac{\sqrt{a^2-x^2}}{x^4}\ \mathrm{d}x &= -\frac{1}{2a^2}\int \left(a^2t^2-1\right)^{\frac{1}{2}}\ \mathrm{d}\left(a^2t^2-1\right) = -\frac{\left(a^2t^2-1\right)^{\frac{3}{2}}}{3a^2} + C \\ &= -\frac{\left(a^2-x^2\right)^{\frac{3}{2}}}{3a^2x^3} + C.\end{aligned}$$

当 $x<0$ 时, 有相同的结果.

注意 之前的例题中, 我们每次都按照定理 3.1.1 所要求的条件严格分析了换元函数的正确性. 但确实这是比较繁琐的过程. 本题演示了省略这一过程的第二类换元法. 考试中这一分析过程不是必须的, 但强烈读者在草纸上写出来, 以确保万无一失. 本例题换元函数的分析过程如下:
“首先由被积函数的表达式可知, x 的取值范围是 $(-|a|,0)\cup(0,|a|)$. 考虑换元函数

$$x=\frac{1}{t},\quad t\in(-\infty,-\frac{1}{|a|})\cup(\frac{1}{|a|},+\infty).$$

显然, 分别在定义域 $(\frac{1}{|a|},+\infty)$ 或 $(-\infty,-\frac{1}{|a|})$ 上, 函数 $\frac{1}{t}$ 关于 t 是单调可导的, 且恒有非零导数. 并且该函数的值域正好就是原来的 x 的取值范围. 故这是一个合法的换元函数. 我们分别考虑当 $x>0$ 时和当 $x<0$ 时, 使用第二类换元法.”

□

3.2　不定积分的分部积分法

分部积分法是又一个求不定积分的常用方法.

定理 3.2.1: 不定积分的分部积分法

设 $u(x)$ 与 $v(x)$ 都是可微函数. 根据函数积的求导公式, 我们有

$$u(x)v'(x)=(u(x)v(x))'-u'(x)v(x)$$

上式两边取不定积分 (积分变量都是 x), 即有

$$\int u(x)v'(x)\mathrm{d}x=u(x)v(x)-\int u'(x)v(x)\mathrm{d}x,$$

或者写成 (把 x 忽略以方便记忆公式)

$$\int u\mathrm{d}v=uv-\int v\mathrm{d}u.$$

这就是所谓的**分部积分公式**. 利用分部积分公式去求不定积分的方法称作分部积分法.

分部积分法的基本想法是: 将求不定积分 $\int u(x)\mathrm{d}(v(x))$ 转化成求不定积分 $\int v(x)\mathrm{d}(u(x))$, 如果不定积分 $\int v(x)\mathrm{d}(u(x))$ 容易求得, 那么原来的不定积分也就随之求出来了.

例 3.2.1 求 $\int x\mathrm{e}^x\,\mathrm{d}x$.

证明. 我们首先把该不定积分写成 $\int x\,\mathrm{d}\mathrm{e}^x$, 然后根据分部积分公式有

$$\int x\mathrm{e}^x\,\mathrm{d}x=\int x\,\mathrm{d}\mathrm{e}^x=x\mathrm{e}^x-\int \mathrm{e}^x\,\mathrm{d}x=x\mathrm{e}^x-\mathrm{e}^x+C.$$

□

在上题中, 如果一开始我们不是将 e^x 移入微分符号 “d” 内, 而是将 x 移入微分符号 “d” 内, 即写成

$$\frac{1}{2}\int e^x \ d(x^2),$$

则根据分部积分公式有

$$\frac{1}{2}\int e^x \ d(x^2) = \frac{1}{2}x^2e^x - \frac{1}{2}\int x^2 \ de^x = \frac{1}{2}x^2e^x - \frac{1}{2}\int x^2e^x \ dx.$$

这样做是徒劳的, 它只会将 x 的方幂次数升高而不会求出任何结果. 由此可见, **选择被积函数的适当部分移入微分符号 “d” 内是分部积分法的关键, 即如何选择 u 和 v (这一过程也是在 “凑微分”). 选择的原则是: 使用分部积分公式之后, 公式的右端得以化简.** 所以我们要从下面的不同表达式中选择适合的一种:

$$xe^x \ dx = x \ de^x = \frac{1}{2}e^x \ d(x^2).$$

使用分部积分法的一个技巧是直接观察分部积分公式

$$\int u \ dv = uv - \int v \ du$$

的最右端 $\int v \ du$ 是否更好算. 特别是在草纸上作演算是时, 可以暂时忽略 uv 这一加项, 还有 $\int v \ du$ 前面的负号. 如果发现 $\int v \ du$ 没有变得更简单, 那么就及时调整, 换一种方法去构造 u 和 v.

例 3.2.2 求 $\int x^2 \cos x \ dx$.

证明.

$$\begin{aligned}\int x^2 \cos x \ dx &= \int x^2 \ d(\sin x)\\ &= x^2 \sin x - \int \sin x \ d(x^2)\\ &= x^2 \sin x - \int \sin x \cdot 2x \ dx\\ &= x^2 \sin x - 2\int x \sin x \ dx.\end{aligned}$$

虽然还没有求出最终结果, 但是用一次分部积分公式就将被积函数中 x 的方幂次数降低了一次. 很自然会想到, 再用一次分部积分公式即可把被积函数中 x 的方幂次数降为零. 事实上, 有

$$\int x \sin x \ dx = -\int x \ d(\cos x) = -x \cos x + \int \cos x \ dx = -x\cos x + \sin x + C.$$

将此结果代入前面得到的等式, 即得

$$\begin{aligned}\int x^2 \cos x \ dx &= x^2 \sin x - 2(-x\cos x + \sin x) + C\\ &= x^2 \sin x + 2x \cos x - 2\sin x + C.\end{aligned}$$

□

上题告诉我们, 同一个问题多次使用分部积分法是没有问题的, 而且也是常见的.

注意

一定要养成好习惯, 计算完不定积分后要求导验证结果! 尤其是考试中!
一定要养成好习惯, 计算完不定积分后要求导验证结果! 尤其是考试中!
一定要养成好习惯, 计算完不定积分后要求导验证结果! 尤其是考试中!

例 3.2.3 求 $\int x^3 \ln x \, \mathrm{d}x$.

证明.

$$\begin{aligned}\int x^3 \ln x \, \mathrm{d}x &= \frac{1}{4}\int \ln x \, \mathrm{d}(x^4) \\ &= \frac{1}{4}x^4 \ln x - \frac{1}{4}\int x^4 \, \mathrm{d}(\ln x) \\ &= \frac{1}{4}x^4 \ln x - \frac{1}{4}\int x^3 \, \mathrm{d}x \\ &= \frac{1}{4}x^4 \ln x - \frac{1}{16}x^4 + C.\end{aligned}$$

这里, 我们把 x^3 先移至微分符号内, 即将 $x^3 \, \mathrm{d}x$ 写成 $\frac{1}{4} \, \mathrm{d}(x^4)$. 虽然 x 的方幂次数升高了, 但是经过分部积分之后, 不定积分的被积表达式变成 $x^4 \, \mathrm{d}(\ln x) = x^4(\ln x)'\mathrm{d}x = x^3 \, \mathrm{d}x$, 对数函数不再出现. 这自然就化简了. □

例 3.2.4 求 $\int \arctan x \, \mathrm{d}x$.

证明. 将 $\arctan x$ 视作分部积分公式中的 $u(x)$, 而将 x 视作 $v(x)$, 这时我们有

$$\begin{aligned}\int \arctan x \, \mathrm{d}x &= x \arctan x - \int x \, \mathrm{d}(\arctan x) \\ &= x \arctan x - \int \frac{x \, \mathrm{d}x}{1+x^2} \\ &= x \arctan x - \frac{1}{2}\int \frac{\mathrm{d}\left(x^2+1\right)}{1+x^2} \\ &= x \arctan x - \frac{1}{2}\ln\left(1+x^2\right) + C.\end{aligned}$$

□

从上题可见, 分部积分法一种非常常见的特殊形式 (将 $v(x) = x$) 就是

$$\int u(x) \, \mathrm{d}x = u(x)x - \int x \, \mathrm{d}(u(x)).$$

特别是, 当被积函数没有任何适当的部分能够移入微分符号 “d” 内时, 上式这个特殊公式至少保证我们对于任何形式的不定积分都可以使用分部积分法. 当然, $\int x \, \mathrm{d}(u(x))$ 好不好解出来则是另外一回事了.

例 3.2.5 求 $\int \arccos x \, \mathrm{d}x$.

证明.

$$\begin{aligned}\int \arccos x \, \mathrm{d}x &= x\arccos x - \int x \, \mathrm{d}(\arccos x)\\ &= x\arccos x + \int \frac{x}{\sqrt{1-x^2}} \, \mathrm{d}x\\ &= x\arccos x - \frac{1}{2}\int \frac{1}{(1-x^2)^{\frac{1}{2}}} \, \mathrm{d}\left(1-x^2\right)\\ &= x\arccos x - \sqrt{1-x^2} + C.\end{aligned}$$

□

例 3.2.6 求 $\int \sqrt{a^2-x^2} \, \mathrm{d}x \quad (a>0)$.[1]

证明. 与上例类似, 可得

$$\begin{aligned}\int \sqrt{a^2-x^2} \, \mathrm{d}x &= x\sqrt{a^2-x^2} + \int \frac{x^2}{\sqrt{a^2-x^2}} \, \mathrm{d}x\\ &= x\sqrt{a^2-x^2} - \int \frac{a^2-x^2}{\sqrt{a^2-x^2}} \, \mathrm{d}x + a^2\int \frac{\mathrm{d}x}{\sqrt{a^2-x^2}}\\ &= x\sqrt{a^2-x^2} - \int \sqrt{a^2-x^2} \, \mathrm{d}x + a^2 \arcsin\frac{x}{a} + C.\end{aligned}$$

上式最后使用了上一节例题的结果 (这个常用的积分结果读者应该背记住)

$$\int \frac{\mathrm{d}x}{\sqrt{a^2-x^2}} = \arcsin\frac{x}{a} + C.$$

通过移项即得

$$\int \sqrt{a^2-x^2} \, \mathrm{d}x = \frac{x}{2}\sqrt{a^2-x^2} + \frac{a^2}{2}\arcsin\frac{x}{a} + C$$

这与上一节中例 3.1.23 所得结果完全一致. □

在上题中, 经过分部积分和适当变形之后, 又出现了我们要求的不定积分, 这时经过移项即得要求的不定积分. 这种情况在许多求不定积分的问题中会遇到.

例 3.2.7 求 $\int \sec^3 t \, \mathrm{d}t$,

[1] 本题在上一节中用第二类换元法解决过, 见例 3.1.23 .

证明.

$$\begin{aligned}\int \sec^3 t \ \mathrm{d}t &= \int \sec t \ \mathrm{d}(\tan t) = \sec t \tan t - \int \tan t \cdot \sec t \tan t \ \mathrm{d}t \\ &= \sec t \tan t - \int \tan^2 t \sec t \ \mathrm{d}t \\ &= \sec t \tan t - \int (\sec^2 t - 1) \sec t \ \mathrm{d}t \\ &= \sec t \tan t - \int \sec^3 t \ \mathrm{d}t + \int \sec t \ \mathrm{d}t \\ &= \sec t \tan t + \ln|\sec t + \tan t| - \int \sec^3 t \ \mathrm{d}t.\end{aligned}$$

通过移项, 得

$$\int \sec^3 t \ \mathrm{d}t = \frac{1}{2} \sec t \tan t + \frac{1}{2} \ln|\sec t + \tan t| + C.$$

所用方法还可用于求 $\int \sec^n t \ \mathrm{d}t \ (n > 1)$. □

问题 求 $\int \sec^n t \ \mathrm{d}t \ (n > 1)$.

例 3.2.8 求 $\int \sqrt{a^2 + x^2} \ \mathrm{d}x \quad (a > 0)$.

证明. 我们用两种办法.

1. 方法一.

$$\begin{aligned}\int \sqrt{a^2 + x^2} \ \mathrm{d}x &= x\sqrt{a^2 + x^2} - \int x \cdot \frac{1}{2} \cdot \frac{2x}{\sqrt{a^2 + x^2}} \ \mathrm{d}x \\ &= x\sqrt{a^2 + x^2} - \int \frac{a^2 + x^2 - a^2}{\sqrt{a^2 + x^2}} \ \mathrm{d}x \\ &= x\sqrt{a^2 + x^2} + a^2 \int \frac{1}{\sqrt{a^2 + x^2}} \ \mathrm{d}x - \int \sqrt{a^2 + x^2} \ \mathrm{d}x.\end{aligned}$$

   移项得

$$\int \sqrt{a^2 + x^2} \ \mathrm{d}x = \frac{x}{2}\sqrt{a^2 + x^2} + \frac{a^2}{2} \int \frac{1}{\sqrt{a^2 + x^2}} \ \mathrm{d}x.$$

   利用上一节中例 3.1.24 的结果, 最终有

$$\int \sqrt{a^2 + x^2} \ \mathrm{d}x = \frac{x}{2}\sqrt{a^2 + x^2} + \frac{a^2}{2} \ln\left|x + \sqrt{a^2 + x^2}\right| + C.$$

2. 方法二. 此例还可以使用上一节中例 3.1.24 相同的换元方法来求: (这省略了对换元函数合法性的讨论) 令 $x = a\tan t \ (-\frac{\pi}{2} < t < \frac{\pi}{2})$, 则通过做辅助三角形可知: $\sqrt{a^2 + x^2} = a\sec t$, $\mathrm{d}x = a\sec^2 t \ \mathrm{d}t$, 从而

$$\int \sqrt{a^2 + x^2} \ \mathrm{d}x = \int a\sec t \cdot a\sec^2 t \ \mathrm{d}t = a^2 \int \sec^3 t \ \mathrm{d}t.$$

   由例 3.2.7 (我们使用了分部积分法求得的) 可知,

$$\int \sec^3 t \ \mathrm{d}t = \frac{1}{2} \sec t \tan t + \frac{1}{2} \ln|\sec t + \tan t| + C.$$

于是最终代回并化简得到,

$$
\begin{aligned}
\int \sqrt{a^2+x^2}\ \mathrm{d}x &= \frac{a^2}{2}\sec t\tan t + \frac{a^2}{2}\ln|\sec t+\tan t| + C \\
&= \frac{1}{2}x\sqrt{a^2+x^2} + \frac{a^2}{2}\ln\left|\frac{x}{a}+\frac{\sqrt{a^2+x^2}}{a}\right| + C \\
&= \frac{1}{2}x\sqrt{a^2+x^2} + \frac{a^2}{2}\ln\left|x+\sqrt{a^2+x^2}\right| + C.
\end{aligned}
$$

□

上题的第二个解法说明了在积分的过程中往往要兼用换元法与分部积分法, 下面再来举一个例子.

例 3.2.9 求 $\int \mathrm{e}^{\sqrt{x}}\ \mathrm{d}x$.

证明. 令 $\sqrt{x}=t$, 则 $x=t^2$, $\mathrm{d}x=2t\ \mathrm{d}t$. 于是

$$\int \mathrm{e}^{\sqrt{x}}\ \mathrm{d}x = 2\int t\mathrm{e}^t\ \mathrm{d}t.$$

利用例 3.2.1 的结果, 并用 $t=\sqrt{x}$ 代回, 便得所求积分:

$$\int \mathrm{e}^{\sqrt{x}}\ \mathrm{d}x = 2\int t\mathrm{e}^t\ \mathrm{d}t = 2\mathrm{e}^t(t-1)+C = 2\mathrm{e}^{\sqrt{x}}(\sqrt{x}-1)+C.$$

□

例 3.2.10 求 $\int \sqrt{x^2-a^2}\ \mathrm{d}x \quad (a>0)$.[2]

证明.

$$
\begin{aligned}
\int \sqrt{x^2-a^2}\ \mathrm{d}x &= x\sqrt{x^2-a^2} - \int x\cdot\frac{1}{2}\cdot\frac{2x}{\sqrt{x^2-a^2}}\ \mathrm{d}x \\
&= x\sqrt{x^2-a^2} - \int \frac{x^2-a^2+a^2}{\sqrt{x^2-a^2}}\ \mathrm{d}x \\
&= x\sqrt{x^2-a^2} - \int \sqrt{x^2-a^2}\ \mathrm{d}x - a^2\int \frac{1}{\sqrt{x^2-a^2}}\ \mathrm{d}x.
\end{aligned}
$$

利用上一节中例 3.1.25 的结果, 再移项得到

$$\int \sqrt{x^2-a^2}\ \mathrm{d}x = \frac{1}{2}x\sqrt{x^2-a^2} - \frac{a^2}{2}\ln\left|x+\sqrt{x^2-a^2}\right| + C.$$

□

[2]注意, 例 3.2.6 是 $\int \sqrt{a^2-x^2}\ \mathrm{d}x = \frac{a^2}{2}\arcsin\frac{x}{a} + \frac{x}{2}\sqrt{a^2-x^2} + C$. 它们是不同的.

结论 3.2.1: 基本积分表 (补充)

除了基本积分表之外, 我们还建议读者记住下列不定积分公式:

1. $\displaystyle\int \tan x \, \mathrm{d}x = -\ln|\cos x| + C$
2. $\displaystyle\int \cot x \, \mathrm{d}x = \ln|\sin x| + C$
3. $\displaystyle\int \sec x \, \mathrm{d}x = \ln|\sec x + \tan x| + C$
4. $\displaystyle\int \csc x \, \mathrm{d}x = \ln|\csc x - \cot x| + C$
5. $\displaystyle\int \frac{\mathrm{d}x}{x^2 + a^2} = \frac{1}{a}\arctan\frac{x}{a} + C \quad (a \neq 0)$
6. $\displaystyle\int \frac{\mathrm{d}x}{x^2 - a^2} = \frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right| + C \quad (a > 0)$
7. $\displaystyle\int \frac{\mathrm{d}x}{\sqrt{a^2 - x^2}} = \arcsin\frac{x}{a} + C \quad (a > 0)$
8. $\displaystyle\int \frac{\mathrm{d}x}{\sqrt{x^2 \pm a^2}} = \ln\left|x + \sqrt{x^2 \pm a^2}\right| + C$

例 3.2.11 求 $\displaystyle\int \mathrm{e}^{ax}\cos bx \, \mathrm{d}x \quad (a, b > 0)$.

证明. 我们先将 e^{ax} 移入微分符号 “d” 内, 再用分部积分:

$$\begin{aligned}\int \mathrm{e}^{ax}\cos bx \, \mathrm{d}x &= \frac{1}{a}\int \cos bx \, \mathrm{d}\left(\mathrm{e}^{ax}\right) = \frac{1}{a}\left(\cos bx \cdot \mathrm{e}^{ax} - \int \mathrm{e}^{ax} \, \mathrm{d}(\cos bx)\right) \\ &= \frac{1}{a}\left(\cos bx \cdot \mathrm{e}^{ax} + b\int \mathrm{e}^{ax}\sin bx \, \mathrm{d}x\right).\end{aligned}$$

对上式最后一个等号右端中的不定积分单独处理, 再用一次分部积分:

$$\int \mathrm{e}^{ax}\sin bx \, \mathrm{d}x = \frac{1}{a}\int \sin bx \, \mathrm{d}\left(\mathrm{e}^{ax}\right) = \frac{1}{a}\left(\sin bx \cdot \mathrm{e}^{ax} - b\int \mathrm{e}^{ax}\cos bx \, \mathrm{d}x\right).$$

我们用了两次分部积分法后得到了原不定积分, 记作

$$J = \int \mathrm{e}^{ax}\cos bx \, \mathrm{d}x.$$

将上式代人前一式便得到关于 J 的一个方程式:

$$\begin{aligned}J &= \frac{1}{a}\left(\cos bx \cdot \mathrm{e}^{ax} + b \cdot \frac{1}{a}\left(\sin bx \cdot \mathrm{e}^{ax} - bJ\right)\right) \\ &= \frac{1}{a}\cos bx \cdot \mathrm{e}^{ax} + \frac{b}{a^2}\sin bx \cdot \mathrm{e}^{ax} - \frac{b^2}{a^2}J \\ &= (a\cos bx + b\sin bx)\frac{1}{a^2}\mathrm{e}^{ax} - \frac{b^2}{a^2}J.\end{aligned}$$

由此方程即可求得

$$\int \mathrm{e}^{ax}\cos bx\ \mathrm{d}x = J = \frac{1}{a^2+b^2}(a\cos bx + b\sin bx)\mathrm{e}^{ax} + C.$$

□

和上题类似的步骤可得

$$\int \mathrm{e}^{ax}\sin bx\ \mathrm{d}x = \frac{1}{a^2+b^2}(a\sin bx - b\cos bx)\mathrm{e}^{ax} + C.$$

还有一类不定积分, 用分部积分法不能直接得出其结果, 但可导出一个递推公式. 我们可由递推公式推出所要的结果.

例 3.2.12 求

$$I_n = \int \frac{\mathrm{d}t}{(t^2+a^2)^n} \quad (n \text{ 为整数且} n > 1, a > 0).$$

证明. 由分部积分法有

$$\begin{aligned} I_n &= \frac{t}{(t^2+a^2)^n} - (-n)\int \frac{t\cdot 2t}{(t^2+a^2)^{n+1}}\ \mathrm{d}t \\ &= \frac{t}{(t^2+a^2)^n} + 2n\int \frac{t^2}{(t^2+a^2)^{n+1}}\ \mathrm{d}t. \end{aligned}$$

另一方面, 观察得到

$$\begin{aligned} I_n &= \int \frac{\mathrm{d}t}{(t^2+a^2)^n} = \int \frac{(t^2+a^2)}{(t^2+a^2)^{n+1}}\mathrm{d}t = \int \frac{t^2}{(t^2+a^2)^{n+1}}\mathrm{d}t + a^2\int \frac{\mathrm{d}t}{(t^2+a^2)^{n+1}} \\ &= \int \frac{t^2}{(t^2+a^2)^{n+1}}\mathrm{d}t + a^2\cdot I_{n+1}. \end{aligned}$$

所有代入上面有

$$\begin{aligned} I_n &= \frac{t}{(t^2+a^2)^n} + 2n\left(I_n - a^2 I_{n+1}\right) \\ &= \frac{t}{(t^2+a^2)^n} + 2nI_n - 2na^2 I_{n+1}. \end{aligned}$$

由此得到递推公式[3]

$$I_{n+1} = \frac{2n-1}{2na^2}I_n + \frac{t}{2na^2\,(t^2+a^2)^n}$$

利用这个递推公式可将 I_n 的下标依次降低, 直到 I_1. 而

$$I_1 = \int \frac{\mathrm{d}t}{t^2+a^2} = \frac{1}{a}\arctan\frac{t}{a} + C.$$

[3]

$$\begin{aligned} I_n &= \frac{t}{(t^2+a^2)^n} + 2nI_n - 2na^2 I_{n+1} \\ \Rightarrow (1-2n)I_n &= \frac{t}{(t^2+a^2)^n} - 2na^2 I_{n+1} \\ \Rightarrow 2na^2 I_{n+1} &= \frac{t}{(t^2+a^2)^n} - (1-2n)I_n \\ \Rightarrow I_{n+1} &= \frac{(2n-1)I_n + \frac{t}{(t^2+a^2)^n}}{2na^2} = \frac{2n-1}{2na^2}I_n + \frac{t}{2na^2\,(t^2+a^2)^n}. \end{aligned}$$

这样就可求得 I_n, 比如

$$I_2=\frac{1}{2a^2}I_1+\frac{t}{2a^2\left(t^2+a^2\right)}=\frac{1}{2a^3}\arctan\frac{t}{a}+\frac{t}{2a^2\left(t^2+a^2\right)}+C.$$

□

无论是分部积分法还是换元法, 它们的适用范围都是有限的. 而且, 在一个求不定积分的问题中应该用什么方法, 这没有一定之规, 只能通过对具体题目的观察和分析, 通过摸索与不断总结经验找到门路. 求不定积分有相当的技巧性, 而熟才能生巧, 只有适当地多做些练习, 才能掌握这些技巧.

注意

就是两个字 "刷题"! 熟能生巧!
就是两个字 "刷题"! 熟能生巧!
就是两个字 "刷题"! 熟能生巧!

在结束本节之前, 我们要指出: 并非所有初等函数的不定积分都是可以求出 (或称 "积出来") 的. 更确切地说, **并非所有初等函数的原函数都是初等函数**.[4] 比如, 人们已经证明

$$\int\frac{\sin x}{x}\,\mathrm{d}x,\int\mathrm{e}^{-x^2}\,\mathrm{d}x,\int\sin x^2\,\mathrm{d}x,\int\cos x^2\,\mathrm{d}x,\int\frac{\cos x}{x}\,\mathrm{d}x,\int\frac{1}{\ln x}\,\mathrm{d}x,$$

等都不能用初等函数表示, 尽管它们的被积函数都十分简单. 那么它们的原函数怎么表示呢? 注意到被积函数都是连续函数, 根据变上限积分函数就是被积函数的原函数这一事实, 可以直接写成, 比如,

$$\int\frac{\sin x}{x}\,\mathrm{d}x=\int_0^x\frac{\sin t}{t}\,\mathrm{d}t+C.$$

这种表达式是没有任何问题的. 未来, 在读者的各个专业领域中, 经常会遇到用定积分来定义的函数, 而不仅仅是由有限次加, 减, 乘, 除以及幂, 指数, 对数, 三角函数和反三角函数的组合构成的函数. 下面就是一些特殊积分和对应的函数 (课外知识, 考试不考).

- $\int\frac{\sin x}{x}\,\mathrm{d}x$: 这被称为 Sine 积分, 记作 $\mathrm{Si}(x)$, 它是一个特殊函数, 用于表示广泛的物理现象, 尤其是在信号处理和光学中.
$$\mathrm{Si}(x)=\int_0^x\frac{\sin t}{t}\,\mathrm{d}t.$$
- $\int\mathrm{e}^{-x^2}\,\mathrm{d}x$: 这是高斯积分, 不能用初等函数表示. 其累积积分为误差函数, 记作 $\mathrm{erf}(x)$, 用于概率, 统计和物理中.
$$\mathrm{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,\mathrm{d}t.$$
- $\int\sin x^2\,\mathrm{d}x$ 和 $\int\cos x^2\,\mathrm{d}x$: 这些积分被称为 Fresnel 积分, 记作 $\mathrm{S}(x)$ 和 $\mathrm{C}(x)$, 常用于衍射和波动光学.
$$\mathrm{S}(x)=\int_0^x\sin(t^2)\,\mathrm{d}t,\quad\mathrm{C}(x)=\int_0^x\cos(t^2)\,\mathrm{d}t.$$

[4]对初等函数来说, 在其定义区间上, 它的原函数一定存在, 但原函数不一定都是初等函数

- $\int \frac{\cos x}{x}\ \mathrm{d}x$: 这是 Cosine Integral, 记作 $\mathrm{Ci}(x)$, 也常见于波动和电磁学领域.

$$\mathrm{Ci}(x) = -\int_x^{\infty} \frac{\cos t}{t}\ \mathrm{d}t.$$

- $\int \frac{1}{\ln x}\ \mathrm{d}x$: 这个积分与 Logarithmic Integral 相关, 记作 $\mathrm{Li}(x)$, 在数论中有重要应用, 比如素数定理中.

$$\mathrm{Li}(x) = \int_0^x \frac{\mathrm{d}t}{\ln t}.$$

这些特殊函数都是扩展初等函数范围的工具, 用来解决许多无法用初等函数表示的积分问题. 它们在物理学, 统计学, 工程学等领域都有广泛的应用.

3.3 有理式的不定积分与有理化方法

我们曾经指出: 有些被积函数的原函数不是初等函数, 即我们无法通过有限步把这些不定积分 "积出来". 那么, 一个很自然的问题是: 到底哪些函数的不定积分是可以 "积出来" 的呢? 本节要告诉读者的是:

1. 自变量 x 的有理式,
2. 三角函数的有理式,
3. 及某些根式的有理式,

的不定积分都是可以 "积出来" 的.

3.3.1 多项式的基本知识

一个**多项式 (polynomial)** $P(x)$ 可以写成

$$P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

其中, a_i 是**系数 (coefficients)**, n 是多项式的最高次数, 称之为多项式的**度 (degree)**, 并且 $a_n \neq 0$. 一个简单的事实是: 如果 $a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \equiv b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0$ $(\forall x \in R)$, 那么对应的系数相等, $a_n = b_n$, $a_{n-1} = b_{n-1}$, $\ldots$, $a_0 = b_0$.

多项式的分类

- 按次数分类:
  - 零次多项式: 常数项, 如 $P(x) = 5$
  - 一次多项式: 线性多项式, 如 $P(x) = 2x + 3$
  - 二次多项式: 如 $P(x) = x^2 - 4x + 4$
  - 三次及更高次数的多项式: 如 $P(x) = x^3 + 2x^2 - 5$
- 按项数分类:
  - 单项式: 只有一个项, 如 $3x^2$
  - 二项式: 两个项, 如 $x^2 + 2$
  - 多项式: 三个或更多项, 如 $x^3 + x^2 + x + 1$

多项式的运算

- 加法: 将相同次数的项相加

$$P(x)+Q(x)=(a_n+b_n)x^n+(a_{n-1}+b_{n-1})x^{n-1}+\cdots+(a_0+b_0).$$

- 减法: 将相同次数的项相减

$$P(x)-Q(x)=(a_n-b_n)x^n+(a_{n-1}-b_{n-1})x^{n-1}+\cdots+(a_0-b_0).$$

- 乘法: 使用分配律

$$P(x)\cdot Q(x)=\sum_{i=0}^{m}\sum_{j=0}^{n}a_ib_jx^{i+j}.$$

  两个多项式相乘时, 乘积多项式的度等于这两个多项式的度之和. 例如, 若 $P(x)=2x^3+x^2+1$ (三次), $Q(x)=x^2-2x+3$ (二次), 则 $P(x)Q(x)$ 的最高次项是 $2x^5$ (五次).

- 除法: 多项式的除法受到算数等式的启发, 即

$$\text{被除数 (dividend)} \ = \ \text{商 (quotient)} \ \times \ \text{除数 (divisor)} \ + \ \text{余数 (remainder)}\ ,$$

  可以使用**多项式长除法**来计算. 例如, 试图计算

$$\frac{27x^3+9x^2-3x-9}{3x-2}.$$

  所谓长除法就是

$$\begin{array}{r|rrrr}
 & 9x^2 & +9x & +5 & \\ \cline{2-5}
3x-2 & 27x^3 & +9x^2 & -3x & -9 \\
 & -27x^3 & +18x^2 & & \\ \cline{2-3}
 & & 27x^2 & -3x & \\
 & & -27x^2 & +18x & \\ \cline{3-4}
 & & & 15x & -9 \\
 & & & -15x & +10 \\ \cline{4-5}
 & & & & 1
\end{array}$$

  因此, 得到

$$\underbrace{27x^3+9x^2-3x-9}_{\text{被除数}}=\underbrace{(9x^2+9x+5)}_{\text{商}}\underbrace{(3x-2)}_{\text{除数}}+\underbrace{1}_{\text{余数}}.$$

  所以,

$$\frac{27x^3+9x^2-3x-9}{3x-2}=(9x^2+9x+5)+\frac{1}{3x-2}.$$

再看几个例题.

例 3.3.1 计算 $\dfrac{x^3-1}{x-1}$.

证明.

$$
\begin{array}{r@{}l}
 & \;x^2+x+1 \\ \cline{2-2}
x-1 & \big)\; x^3 \qquad\quad -1 \\
 & -x^3+x^2 \\ \cline{2-2}
 & \qquad x^2 \\
 & \qquad -x^2+x \\ \cline{2-2}
 & \qquad\qquad x-1 \\
 & \qquad\qquad -x+1 \\ \cline{2-2}
 & \qquad\qquad\quad 0
\end{array}
$$

所以,

$$\frac{x^3-1}{x-1}=x^2+x+1.$$

□

例 3.3.2 计算 $\dfrac{1-2x^2+3x^3}{x^2-2}$.

证明.

$$
\begin{array}{r@{}l}
 & \qquad\qquad 3x-2 \\ \cline{2-2}
x^2-2 & \big)\; 3x^3-2x^2 \qquad +1 \\
 & -3x^3 \qquad +6x \\ \cline{2-2}
 & \qquad -2x^2+6x+1 \\
 & \qquad\; 2x^2 \qquad -4 \\ \cline{2-2}
 & \qquad\qquad\quad 6x-3
\end{array}
$$

因此, 得到

$$1-2x^2+3x^3=(3x-2)(x^2-2)+6x-3.$$

所以,

$$\frac{1-2x^2+3x^3}{x^2-2}=3x-2+\frac{6x-3}{x^2-2}.$$

□

例 3.3.3 用长除法计算多项式的除法:

1. $\dfrac{x^2+5x+17}{x-3}$
2. $\dfrac{4x^3+3x-5}{x^2+1}$
3. $\dfrac{4x^4+2x^2+1}{2x^2+1}$

证明. 这里只给出答案.

1. $\dfrac{x^2+5x+17}{x-3} = (x+8) + \dfrac{41}{x-3}$;

2. $\dfrac{4x^3+3x-5}{x^2+1} = 4x + \dfrac{-x-5}{x^2+1}$;

3. $\dfrac{4x^4+2x^2+1}{2x^2+1} = 2x^2 + \dfrac{1}{2x^2+1}$

$\square$

多项式的长除法实际上证明了下面的定理.

定理 3.3.1: 多项式除法的基本原理

给定任意多项式 $P(x)$ 和非常数多项式 $D(x)$, 总存在唯一的一对多项式 $Q(x)$ (商式) 和 $R(x)$ (余式), 使得

$$P(x) \equiv D(x)Q(x) + R(x), \quad \Leftrightarrow \quad \frac{P(x)}{D(x)} \equiv Q(x) + \frac{R(x)}{D(x)},$$

且满足:

1. 余式 $R(x)$ 的度 (最高次数) 严格小于除式 $D(x)$ 的度.

2. 商式 $Q(x)$ 的度等于被除式 $P(x)$ 的度减去除式 $D(x)$ 的度.

这个定理告诉我们: 任何多项式除法都能得到唯一确定的商和余数; 余数的次数必须比除数低; 商的次数可以通过被除数和除数的次数之差来确定.

一元二次方程的根 解一元二次方程就是找出某个二次多项式的根, 它的一般形式为: $ax^2+bx+c=0$, 其中 $a \neq 0$. 一元二次方程根的判别式为 $\Delta = b^2-4ac$.

1. 若 $\Delta > 0$, 则该方程有两个不相等的实数根: $x_{1,2} = \dfrac{-b \pm \sqrt{b^2-4ac}}{2a}$;

2. 若 $\Delta = 0$, 则该方程有两个相等的实数根: $x_{1,2} = -\dfrac{b}{2a}$;

3. 若 $\Delta < 0$, 则该方程有一对共轭复数根: $x_{1,2} = \dfrac{-b \pm i\sqrt{4ac-b^2}}{2a}$.

可以验证, 无论哪种情况, 都有因式分解

$$ax^2+bx+c = (x-x_1)(x-x_2).$$

由上可知, 在实数范围内求解一元二次方程, 当 $\Delta \geq 0$ 时, 方程才有根 (有两个不等实数根或两个相等实数根); 当 $\Delta < 0$ 时, 方程有两个复数根, 但是在实数范围无解. 但总之, 在复数域上, 一元二次方程永远有 2 个根 (包括重根).

多项式的根 (代数基本定理) 多项式的根是使多项式值为零的 x 的值. 假设有一个 n 次多项式 $P(x)$, 它的形式为:

$$P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

其中 $a_n \neq 0$, 则根是使得 $P(x) = 0$ 成立的 x 值.

定理 3.3.2: 代数基本定理

根据**代数基本定理 (fundamental theorem of algebra)**, 任何一个实系数 n 次多项式在复数范围内一定有 n 个根 (包括重根). 这些根可以分为两大类:

1. 实根: 根是实数, 即多项式在实数范围内的解. 实根可以是单根或重根.
2. 复根: 根是复数 (非实数), 并且复根会以共轭复数的形式成对出现, 即如果 z 是一个复根, 则 $\bar{z}$ 也必然是一个复根.

对于具有 n 个根 (包括重根) 的 n 次多项式 $P(x)$, 我们可以将它写成线性因子的形式:

$$P(x) = a_n\left(x - r_1\right)\left(x - r_2\right)\cdots\left(x - r_n\right)$$

其中 $r_1, r_2, \ldots, r_n$ 是 $P(x)$ 的根, 且可能包括复数根.

实系数的多项式, 如果含有复根, 这些复根会成对出现, 共轭复根对以 **(实数域上的) 不可约二次因式**的形式出现. 下面是一些多项式因式分解的具体例子.

例 3.3.4 考虑多项式

$$P(x) = x^3 - 6x^2 + 11x - 6.$$

求解 $P(x) = 0$, 得到实根 $r_1 = 1$, $r_2 = 2$, $r_3 = 3$. 根据这些根, 可以将多项式分解为

$$P(x) = (x-1)(x-2)(x-3).$$

例 3.3.5 考虑多项式

$$P(x) = x^3 - 2x^2 + 5x - 6.$$

求解方程 $P(x) = 0$ 得到一个实根: $r_1 = 2$, 一对共轭复根: $r_2 = -1 + i$, $r_3 = -1 - i$. 因此, 该多项式可以分解为

$$P(x) = (x-2)(x-(-1+i))(x-(-1-i)).$$

也可写作

$$P(x) = (x-2)(x^2+2x+2).$$

其中 $x^2 + 2x + 2$ 是一个无实根的二次因式. 其中, $x^2 + 2x + 2$ 就是 (实数域上的) 不可约二次因式.

3.3.2 有理式的不定积分

所谓的 x 的**有理式 (rational function)**, 是指两个关于 x 的多项式 (函数) 之比, 即

$$\frac{P(x)}{Q(x)} = \frac{a_0x^n + a_1x^{n-1} + \cdots + a_n}{b_0x^m + b_1x^{m-1} + \cdots + b_m}.$$

该函数的定义域是所有使得分母 $Q(x) \neq 0$ 的实数 x, 即整个实数域中除去 $Q(x)$ 的所有实根. 当 $m > n$, 即分母的度严格大于分子的度时, 上述有理式称为**真分式**; 否则, 当 $m \leq n$, 上述有理式称为**假分式**. 本小节的目

的在于讨论求不定积分

$$\int \frac{P(x)}{Q(x)}\,\mathrm{d}x \quad (\forall x \in R, Q(x) \neq 0)$$

的一般的方法. 首先我们对该问题做出一些假定条件.

约定 3.3.1

要求有理式的不定积分:

$$\int \frac{P(x)}{Q(x)} = \frac{a_0x^n + a_1x^{n-1} + \cdots + a_n}{b_0x^m + b_1x^{m-1} + \cdots + b_m}\,\mathrm{d}x \quad (\forall x \in R, Q(x) \neq 0).$$

通常我们假定:

1. $a_0 \neq 0$, $b_0 \neq 0$.
2. 多项式 $P(x)$ 与 $Q(x)$ 没有公因子.
3. 有理式 $\dfrac{P(x)}{Q(x)}$ 是真分式.

原因如下:

1. 若最高次数项系数为零, 则该项消失, 最高次次数应该降低, 直到遇到非零系数.

2. 若 $P(x)$ 与 $Q(x)$ 有公因子, 则应该同时约去公因子, 以保持被积函数 $\dfrac{P(x)}{Q(x)}$ 不变的同时化简问题. 例如, 欲求

$$\int \frac{x-2}{x^2-4}\,\mathrm{d}x \quad (x \neq \pm 2).$$

可看到, 分子分母有公因子 $x-2$, 故可化简问题:

$$\int \frac{x-2}{(x-2)(x+2)}\,\mathrm{d}x = \int \frac{1}{x+2}\,\mathrm{d}x = \ln|x+2| + C \quad (x \neq \pm 2).$$

但是务必要小心, 同时约去公因子后, 函数的定义域应该不能发生变化. 特别是 $x = 2$ 在最终的积分结果中必须排除 (尽管将 $x = 2$ 代入 $\ln|x+2|$ 是可以计算的). 毕竟, 总不能在求被积函数的原函数的过程中, 求着求着把被积函数的定义域都改变了.

3. 我们只讨论真分式就足够了, 因为**任何一个假分式都可以表示成一个多项式加上一个真分式, 而求多项式的不定积分是容易的**.

对于一个给定的有理真分式 $\dfrac{P(x)}{Q(x)}$, 求不定积分的步骤如下所示.

结论 3.3.1: 有理式不定积分求的一般解法 - 拆! 拆! 拆!

要求有理式的不定积分:

$$\int \frac{P(x)}{Q(x)} = \frac{a_0x^n + a_1x^{n-1} + \cdots + a_n}{b_0x^m + b_1x^{m-1} + \cdots + b_m}\,\mathrm{d}x \quad (\forall x \in R, Q(x) \neq 0).$$

步骤如下:

1. 检查有理式, 使其满足约定 3.3.1 的 3 个条件. 要特别检查是不是真分式. 如若不是, 通过多项式长除法弄出来一个真分式, 然后我们重点关心此真分式的不定积分. 故不是一般性, 假定 $\frac{P(x)}{Q(x)}$ 是真分式.
2. 将真分式 $\frac{P(x)}{Q(x)}$ 表示成部分分式之和.[a] 所谓的**部分分式**, 是特指下列**四种类型**的最简单的真分式:
$$\frac{A}{x-a},\quad \frac{A}{(x-a)^n},\quad \frac{Bx+D}{x^2+px+q},\quad \frac{Bx+D}{(x^2+px+q)^n}.$$
其中 A, B, D, a, p, q 为某些常数, n 为大于等于 2 的整数[b]; 上式后两者的分母中的 x^2+px+q 是没有实根的二次多项式, 也就是说 $p^2-4q<0$.
3. 求得每一个部分分式的不定积分, 这个过程相对容易. (后面会给出一般的求解过程)
4. 将所有部分分式的不定积分的结果汇总求和就是 $\frac{P(x)}{Q(x)}$ 的不定积分.

[a]比如
$$\frac{3x^3+3x^2+2x+2}{x(x+2)(x^2+1)} = \frac{1}{x} + \frac{2}{x+2} + \frac{x-1}{x^2+1}.$$
它们由第一和第二种类型的部分分部求和.

[b]实际上只有两种, 即 $\frac{A}{(x-a)^n}$ 和 $\frac{Bx+D}{(x^2+px+q)^n}$, 只是现在 n 为大于等于 1 的整数. 把 $n=1$ 单独提出来是因为这类真分式最常见最重要.

上述过程的关键步骤是如何将 $\frac{P(x)}{Q(x)}$ 表示成部分分式之和. 要想将一个真分式 $\frac{P(x)}{Q(x)}$ 表示成部分分式之和, 首先第一步是要找出分母有理式 $Q(x)$ 的根. 任何多项式 $Q(x)$ 的根的存在情况, 都可以用下面的语言来统一概述. 特别注意, 多项式可以同时有实数根和复数根, 亦或者仅有实数或复数根. 假定多项式 $Q(x)$ 有以下的根:

1. **(实数根/实根的情况)** 存在 k 个互不相同的实根 $a_1, a_2, \cdots, a_k$, 而它们的**重数（重复次数）**分别是 $n_1, n_2, \cdots, n_k$;
2. **(复数根的情况)** 存在 l 对互不相同的**共轭复根** $\alpha_1 \pm \mathrm{i}\beta_1, \alpha_2 \pm \mathrm{i}\beta_2, \cdots, \alpha_l \pm \mathrm{i}\beta_l$, 其重数分别为 $m_1, m_2, \cdots, m_l$. 注意, 这里假定所有复根的虚部 β_i 都不等于 0; 否则的话, $\alpha_i \pm \mathrm{i}\beta_i = \alpha_i$ 应当是两个相等的实数根, 对此根的假设回归到第一条.

这样, 则我们有 $Q(x)$ 的因式分解形式

$$Q(x) = (x-a_1)^{n_1} \cdots (x-a_k)^{n_k} \left(x^2+p_1x+q_1\right)^{m_1} \cdots \left(x^2+p_lx+q_l\right)^{m_l},$$

其中 $p_i := -2\alpha_i$, $q_i := \alpha_i^2+\beta_i^2$, $p_i^2 < 4q_i$ $(i = 1, 2, \cdots, l)$. 于是, 根据代数中的一个定理 (证明不需要关心, 从

后面的例子中可以发现这是正确的), 有

$$\frac{P(x)}{Q(x)}=\sum_{j=1}^{k}\sum_{n=1}^{n_j}\frac{A_{nj}}{(x-a_j)^n}+\sum_{i=1}^{l}\sum_{m=1}^{m_i}\frac{B_{mi}x+D_{mi}}{(x^2+p_ix+q_i)^m},$$

其中, (这里 A,B,D 的下角标从右往左读)

- $A_{nj}\ (n=1,2,\cdots,n_j,j=1,2,\cdots,k)$,
- $B_{mi},D_{mi}\ (m=1,2,\cdots,m_i,i=1,2,\cdots,l)$,

均为某些常数. 我们可以将上式的双重求和号展开, 写成更为明显的形式:

$$\left.\begin{aligned}\frac{P(x)}{Q(x)}=\ &\frac{A_{11}}{(x-a_1)}+\frac{A_{21}}{(x-a_1)^2}+\cdots+\frac{A_{n_11}}{(x-a_1)^{n_1}}+\\ &\qquad\vdots\\ &+\frac{A_{1k}}{(x-a_k)}+\frac{A_{2k}}{(x-a_k)^2}+\cdots+\frac{A_{n_kk}}{(x-a_k)^{n_k}}\end{aligned}\right\}\text{第一部分求和 }\sum_{j=1}^{k}\sum_{n=1}^{n_j}\frac{A_{nj}}{(x-a_j)^n}$$

$$\left.\begin{aligned}&+\frac{B_{11}x+D_{11}}{x^2+p_1x+q_1}+\cdots+\frac{B_{m_11}x+D_{m_11}}{(x^2+p_1x+q_1)^{m_1}}+\\ &\qquad\vdots\\ &+\frac{B_{1l}x+D_{1l}}{x^2+p_lx+q_l}+\cdots+\frac{B_{m_ll}x+D_{m_ll}}{(x^2+p_lx+q_l)^{m_l}}.\end{aligned}\right\}\text{第二部分求和 }\sum_{i=1}^{l}\sum_{m=1}^{m_i}\frac{B_{mi}x+D_{mi}}{(x^2+p_ix+q_i)^m}$$

我们把以上结论用更加简单的方式描述, 即我们抛开冗余的小角标, 只按照不同的类型去解释:

1. 如果 $Q(x)$ 的因式分解形式中包含因子
$$(x-a)^n,$$
即 $Q(x)$ 有一个 n 重实根 a, 则 $\dfrac{P(x)}{Q(x)}$ 的部分分式中一定包含如下形式的 n 个部分分式之和:
$$\frac{A_1}{x-a}+\frac{A_2}{(x-a)^2}+\cdots+\frac{A_n}{(x-a)^n}.$$

2. 如果 $Q(x)$ 的因式分解形式中包含因子
$$\left(x^2+px+q\right)^m,\quad p^2-4q<0,$$
即 $Q(x)$ 有一对共轭的复数根 $\dfrac{-p\pm i\sqrt{4q-p^2}}{2}$ (通常并没有求出该根的必须, 只要确定这个二次多项式的没有实根即可), 则 $\dfrac{P(x)}{Q(x)}$ 的部分分式中一定包含如下形式的 m 项部分分式之和:
$$\frac{B_1x+D_1}{x^2+px+q}+\frac{B_2x+D_2}{(x^2+px+q)^2}+\cdots+\frac{B_mx+D_m}{(x^2+px+q)^m}.$$

3. 将 $Q(x)$ 的全体因子对应的这些部分分式之和全都加起来, 就等于 $\dfrac{P(x)}{Q(x)}$.

现在我们举例说明上述结论. 比如, 我们考虑某个真分式 $\dfrac{P(x)}{Q(x)}$, 已知其分母 $Q(x)$ 可以表示成如下的因式分解形:

$$Q(x)=a(x-1)(x-2)^2(x^2+1)^3(x^2+x+1).$$

那么, 这个分式可以分解成如下形式的部分分式之和:

$$\begin{aligned}\frac{P(x)}{Q(x)} = &\frac{A_{11}}{x-1} \quad \leftarrow \text{对应因式 } (x-1) \text{ 或一重实根 } x=1\\ &+\frac{A_{12}}{x-2}+\frac{A_{22}}{(x-2)^2} \quad \leftarrow \text{对应因式 } (x-2)^2 \text{ 或二重实根 } x=2\\ &+\frac{B_{11}x+D_{11}}{x^2+1}+\frac{B_{21}x+D_{21}}{(x^2+1)^2}+\frac{B_{31}x+D_{31}}{(x^2+1)^3} \quad \leftarrow \text{对应因式 } (x^2+1)^3 \text{ 或一对共轭复数根 } x=\pm i\\ &+\frac{B_{12}x+D_{12}}{x^2+x+1}. \quad \leftarrow \text{对应因式 } (x^2+x+1) \text{ 或一对共轭复数根 } x=\frac{-1\pm i\sqrt{3}}{2}\end{aligned}$$

以后, 我们会看到这里的常数 A_{ij} $(i,j=1,2)$, B_{ij}, D_{ij} $(i=1,2,j=1,2,3)$ 可以通过比较上式等号两端的系数逐一定出.

以上我们说明了**任意一个真分式都可分解成若干部分分式之和, 并指明了具体的分解方法. 这样一来, 全部的问题归结为求上述四种部分分式的不定积分.**

这里, 我们尝试着带着参数去求解这四种不定积分. 不要去背记最终的结果, 重点是要掌握求解的过程.

1. 对于第一种部分分式, 其不定积分是显然的:

$$\int \frac{A\,\mathrm{d}x}{x-a} = A\ln|x-a|+C.$$

2. 对于第二种部分分式, 其不定积分亦是显然的:

$$\int \frac{A\,\mathrm{d}x}{(x-a)^n} = \frac{A}{1-n}(x-a)^{1-n}+C \quad (\text{正整数 } n \geq 1).$$

3. 第三种部分分式的不定积分也是不难求出的:

$$\begin{aligned}&\int \frac{Bx+D}{x^2+px+q}\,\mathrm{d}x\\ =&\int \frac{Bx+D}{\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)}\,\mathrm{d}x \quad (\text{分母配出平方项})\\ =&B\int \frac{x}{\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)}\,\mathrm{d}x+D\int \frac{1}{\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)}\,\mathrm{d}x \quad (\text{拆分})\\ =&B\int \frac{\left(x+\frac{p}{2}\right)-\frac{p}{2}}{\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)}\,\mathrm{d}x+D\int \frac{1}{\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)}\,\mathrm{d}x \quad (\text{调整第一项被积函数的分子})\\ =&B\int \frac{\left(x+\frac{p}{2}\right)}{\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)}\,\mathrm{d}x+\left(D-\frac{Bp}{2}\right)\int \frac{1}{\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)}\,\mathrm{d}x \quad (\text{把第一项的一部分合并到第二项})\\ =&\frac{B}{2}\int \frac{\mathrm{d}\left[\left(x+\frac{p}{2}\right)^2-\left(q-\frac{p^2}{4}\right)\right]}{\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)}+\left(D-\frac{Bp}{2}\right)\int \frac{1}{\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)}\,\mathrm{d}x \quad (\text{第一项凑微分})\\ =&\frac{B}{2}\ln\left|\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)\right|+\left(D-\frac{Bp}{2}\right)\cdot\frac{1}{\sqrt{q-\frac{p^2}{4}}}\arctan\left(\frac{x+\frac{p}{2}}{\sqrt{q-\frac{p^2}{4}}}\right)+C \quad (\text{两项分别使用凑微分法})\\ =&\frac{B}{2}\ln\left|x^2+px+q\right|+\frac{D-\frac{Bp}{2}}{\sqrt{q-\frac{p^2}{4}}}\arctan\left(\frac{x+\frac{p}{2}}{\sqrt{q-\frac{p^2}{4}}}\right)+C. \quad (\text{化简})\end{aligned}$$

注意, 上面的倒数第二个等式中的第二项, 使用了基本积分表中的

$$\int \frac{\mathrm{d}x}{x^2+a^2} = \frac{1}{a}\arctan\frac{x}{a}+C \quad (a\neq 0).$$

因为 $p^2-4q<0$, 故在我们的问题中 $a=q-\dfrac{p^2}{4}\neq 0$, 正好符合了使用上述积分的条件.

4. 对于第四种部分分式, 我们有

$$
\begin{aligned}
&\int \frac{Bx+D}{(x^2+px+q)^n}\,\mathrm{d}x\\
&=\int \frac{Bx+D}{\left[\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)\right]^n}\,\mathrm{d}x \quad \text{(分母配出平方项)}\\
&\ \ \vdots \quad \text{(省略了和第三种部分分式的积分计算完全类似的步骤)}\\
&=\frac{B}{2}\int \frac{\mathrm{d}\left[\left(x+\frac{p}{2}\right)^2-\left(q-\frac{p^2}{4}\right)\right]}{\left[\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)\right]^n}+\left(D-\frac{Bp}{2}\right)\int \frac{1}{\left[\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)\right]^n}\,\mathrm{d}x\\
&=\frac{B}{2(1-n)}\left[\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)\right]^{1-n}+\left(D-\frac{Bp}{2}\right)\int \frac{1}{\left[\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)\right]^n}\,\mathrm{d}x.
\end{aligned}
$$

其中最后一项不定积分可以通过上一节例题? 的递推公式求解.

以上我们证明了有理式的不定积分是可以 “积出来” 的, 并且提供了求这种不定积分的实际途径.

注意 我们经常要做多项式的因式分解. 读者应该熟悉**十字相乘法**. 详细资料可以参考 利用十字交乘法做因式分解.pdf 或 B 站搜做相关视频.

例 3.3.6 求 $\displaystyle\int \frac{5x+3}{x^3-2x^2-3x}\,\mathrm{d}x$.

证明. 这里被积函数是真分式, 其中分母为 $Q(x)=x^3-2x^2-3x$. 首先姜将 $Q(x)$ 因式分解:

$$Q(x)=x(x^2-2x-3)=x(x+1)(x-3).$$

注意, 第二个等式用了十字相乘法得到. 可见, $Q(x)$ 有三个根: $x=0, x=-1, x=3$, 都是单根. 于是, 设

$$\frac{5x+3}{x(x+1)(x-3)}=\frac{A_1}{x}+\frac{A_2}{x+1}+\frac{A_3}{x-3},$$

其中 A_1, A_2, A_3 为待定常数. 上式右端通分后得到

$$\frac{5x+3}{x(x+1)(x-3)}=\frac{A_1(x+1)(x-3)+A_2x(x-3)+A_3x(x+1)}{x(x+1)(x-3)}.$$

于是, 只关心两边的分子部分, 展开后得到

$$
\begin{aligned}
5x+3&=A_1(x+1)(x-3)+A_2x(x-3)+A_3x(x+1)\\
&=(A_1+A_2+A_3)\,x^2+(-2A_1-3A_2+A_3)\,x-3A_1.
\end{aligned}
$$

由上式两端同次幂的系数相等, 得到方程组:

$$
\begin{cases}
A_1+A_2+A_3=0\\
-2A_1-3A_2+A_3=5\\
-3A_1=3
\end{cases}
$$

由此方程组解得 $A_1=-1, A_2=-\dfrac{1}{2}, A_3=\dfrac{3}{2}$. 这样, 我们有

$$\frac{5x+3}{x(x+1)(x-3)}=\frac{-1}{x}+\frac{-\frac{1}{2}}{x+1}+\frac{\frac{3}{2}}{x-3}.$$

至此, 我们将有理式变成了部分分式之和. 于是

$$\begin{aligned}\int \frac{5x+3}{x^3-2x^2-3x}\,\mathrm{d}x&=\int\left(\frac{-1}{x}+\frac{-\frac{1}{2}}{x+1}+\frac{\frac{3}{2}}{x-3}\right)\mathrm{d}x\\&=-\ln|x|-\frac{1}{2}\ln|x+1|+\frac{3}{2}\ln|x-3|+C.\end{aligned}$$

□

例 3.3.7 求 $\displaystyle\int \frac{x^3+1}{x(x-1)^3}\,\mathrm{d}x$.

证明. 这里被积函数是有理式, 其分母有两个实根: $x=0$ 为单根, 而 $x=1$ 为三重根. 因此, 被积函数的部分分式分解应为

$$\frac{x^3+1}{x(x-1)^3}=\frac{A_{11}}{x}+\frac{A_{12}}{x-1}+\frac{A_{22}}{(x-1)^2}+\frac{A_{32}}{(x-1)^3},$$

其中 $A_{11}, A_{12}, A_{22}, A_{32}$ 为待定常数, 可通过比较系数法确定. 事实上, 上式右端通分后得到

$$\frac{x^3+1}{x(x-1)^3}=\frac{A_{11}(x-1)^3+A_{12}x(x-1)^2+A_{22}x(x-1)+A_{32}x}{x(x-1)^3}.$$

只关心并展开分子得

$$\begin{aligned}x^3+1&=A_{11}(x-1)^3+A_{12}x(x-1)^2+A_{22}x(x-1)+A_{32}x\\&=A_{11}(x^3-3x^2+3x-1)+A_{12}x(x^2-2x+1)+A_{22}x(x-1)+A_{32}x\\&=A_{11}x^3-3A_{11}x^2+3A_{11}x-A_{11}+A_{12}x^3-2A_{12}x^2+A_{12}x+A_{22}x^2-A_{22}x+A_{32}x\\&=(A_{11}+A_{12})x^3+(-3A_{11}-2A_{12}+A_{22})x^2+(3A_{11}+A_{12}-A_{22}+A_{32})x-A_{11}.\end{aligned}$$

令上式两端同次幂的系数相等, 得到一个方程组:

$$\begin{cases}1=A_{11}+A_{12},\\0=-3A_{11}-2A_{12}+A_{22},\\0=3A_{11}+A_{12}-A_{22}+A_{32},\\1=-A_{11}.\end{cases}$$

由此方程组可解得 $A_{11}=-1, A_{12}=2, A_{22}=1, A_{32}=2$. 这样, 我们有

$$\frac{x^3+1}{x(x-1)^3}=\frac{-1}{x}+\frac{2}{x-1}+\frac{1}{(x-1)^2}+\frac{2}{(x-1)^3},$$

于是

$$\begin{aligned}\int\frac{(x^3+1)\;\mathrm{d}x}{x(x-1)^3}&=-\ln|x|+2\ln|x-1|-\frac{1}{x-1}-\frac{1}{(x-1)^2}+C\\&=\ln\frac{|x-1|^2}{|x|}-\frac{x}{(x-1)^2}+C.\end{aligned}$$

□

上一例题中, 我们用比较等式两端 x 的同次幂系数的方法确定出了常数. 这个方法非常地通用, 单也有一些麻烦的地方:

1. 首先要完全展开含着待定系数的"通分后的分子之和". 在上题中, 就是将

$$A_{11}(x-1)^3+A_{12}x(x-1)^2+A_{22}x(x-1)+A_{32}x$$

展开得到

$$\left(A_{11}+A_{12}\right)x^3+\left(-3A_{11}-2A_{12}+A_{22}\right)x^2+\left(3A_{11}+A_{12}-A_{22}+A_{32}\right)x-A_{11}.$$

这个过程颇为, 需要读者有耐心和细心.

2. 比较等式两端 x 的同次幂系数之后, 往往要求解一个含多变量的方程组. 而解方程组总是麻烦的.

有时也可以根据等式两端的性质, 用灵活的方法来确定出这些常数. 比如, 在上一道例题中, 我们可以在等式

$$x^3+1=A_{11}(x-1)^3+A_{12}x(x-1)^2+A_{22}x(x-1)+A_{32}x$$

的两端令 $x=0$, 则立刻得到 $A_{11}=-1$; 令 $x=1$, 则立刻得到 $A_{32}=2$. 之所以选择 0 和 1, 是来自于对右侧的多项式的结构的观察. 取这两个值可以快速令某些项为零. 于是再把这两个系数的结果代回得到

$$x^3+1+(x-1)^3-2x=A_{12}x(x-1)^2+A_{22}x(x-1).$$

化简上式左侧并约去两端的公因子 x, 得到

$$2x^2-3x+1=A_{12}(x-1)^2+A_{22}(x-1).$$

再对上式左侧用十字相乘法, 得到

$$2x^2-3x+1=(2x-1)(x-1)=A_{12}(x-1)^2+A_{22}(x-1),$$

又发现了一个公因子 $(x-1)$, 最终有

$$2x-1=A_{12}(x-1)+A_{22}.$$

由此推出 $A_{12}=2$, $A_{22}=1$.

例 3.3.8 求 $\displaystyle\int\frac{4}{x^3+4x}\,\mathrm{d}x$.

证明. 被积函数的分母 $x^3+4x=x(x^2+4)$ 有实根 $x=0$ 及复根 $x=2\mathrm{i},-2\mathrm{i}$, 都是单根, 故设

$$\frac{4}{x^3+4x}=\frac{A}{x}+\frac{Bx+D}{x^2+4},$$

其中 A,B,D 为待定常数, 于是通分后有

$$4=A\left(x^2+4\right)+Bx^2+Dx.$$

令 $x=0$, 得 $A=1$. 代回上式, 有

$$Bx^2+Dx=4-\left(x^2+4\right)=-x^2,$$

即 $Bx+D=-x$. 由此推出 $B=-1, D=0$. 这样, 我们得到

$$\frac{4}{x^3+4x}=\frac{1}{x}-\frac{x}{x^2+4}.$$

于是

$$\int\frac{4\,\mathrm{d}x}{x^3+4x}=\ln|x|-\frac{1}{2}\ln\left(x^2+4\right)+C.$$

□

例 3.3.9 求 $\displaystyle\int\frac{x^3+x^2+2}{x\left(x^2+2\right)^2}\,\mathrm{d}x$.

证明. 被积函数的分母含有因子 $\left(x^2+2\right)^2$, 另外有一个实单根 $x=0$, 故可设

$$\frac{x^3+x^2+2}{x\left(x^2+2\right)^2}=\frac{A}{x}+\frac{Bx+D}{x^2+2}+\frac{B'x+D'}{\left(x^2+2\right)^2},$$

其中 A,B,D,B',D' 为待定常数, 于是

$$x^3+x^2+2=A\left(x^2+2\right)^2+(Bx+D)x\left(x^2+2\right)+B'x^2+D'x.$$

令 $x=0$, 得 $A=\frac{1}{2}$, 这样有

$$x^3+x^2+2-\frac{1}{2}\left(x^2+2\right)^2=(Bx+D)x\left(x^2+2\right)+B'x^2+D'x,$$

化简得

$$x^2-x-\frac{1}{2}x^3=Bx^3+Dx^2+2Bx+B'x+2D+D'.$$

比较上式两端 x 的同次幂系数, 即得

$$B=-\frac{1}{2},\quad D=1,\quad B'=0,\quad 2D+D'=0,$$

也即 $D'=-2D=-2$. 因此

$$\int\frac{x^3+x^2+2}{x\left(x^2+2\right)^2}\,\mathrm{d}x=\frac{1}{2}\int\frac{1}{x}\,\mathrm{d}x-\int\frac{\frac{1}{2}x-1}{x^2+2}\,\mathrm{d}x-2\int\frac{1}{\left(x^2+2\right)^2}\,\mathrm{d}x.$$

上式的第二个积分有

$$\begin{aligned}\int\frac{\frac{1}{2}x-1}{x^2+2}\,\mathrm{d}x&=\frac{1}{2}\int\frac{x}{x^2+2}\,\mathrm{d}x-\int\frac{1}{x^2+2}\,\mathrm{d}x\\&=\frac{1}{4}\ln\left|x^2+2\right|-\frac{1}{\sqrt{2}}\arctan\left(\frac{x}{\sqrt{2}}\right)+C.\end{aligned}$$

对于上式最后一个不定积分, 可用已知的递推公式 (见上一节中的例?) 来求. 这里直接用分部积分法来求:[5]

$$\begin{aligned}\int\frac{\mathrm{d}x}{\left(x^2+2\right)^2}&=-\int\frac{1}{2x}\,\mathrm{d}\left(\frac{1}{x^2+2}\right)=-\frac{1}{2x\left(x^2+2\right)}-\frac{1}{2}\int\frac{\mathrm{d}x}{x^2\left(x^2+2\right)}\\&=-\frac{1}{2x\left(x^2+2\right)}-\frac{1}{4}\int\left(\frac{1}{x^2}-\frac{1}{x^2+2}\right)\mathrm{d}x\\&=-\frac{1}{2x\left(x^2+2\right)}+\frac{1}{4}\cdot\frac{1}{x}+\frac{1}{4\sqrt{2}}\arctan\frac{x}{\sqrt{2}}+C.\end{aligned}$$

[5]注意, 下面的分析中我们用了一个观察:

$$\frac{1}{x(x+a)}=\frac{1}{x}-\frac{1}{x+a}.$$

其中, x,a 可以是任何东西. 像这类非常简单的小结果要好好总结. 往往用起来比较方便.

汇总所有结果, 最终我们得到

$$\int \frac{x^3+x^2+2}{x\left(x^2+2\right)^2}\,\mathrm{d}x=\frac{1}{2}\ln|x|-\frac{1}{4}\ln\left|x^2+2\right|+\frac{1}{x\left(x^2+2\right)}-\frac{1}{2x}+\frac{1}{2\sqrt{2}}\arctan\frac{x}{\sqrt{2}}+C.$$

□

例 3.3.10 求 $\int \frac{x^3}{x^2+x-2}\,\mathrm{d}x$.

证明. 这里被积函数是一个假分式, 我们先把它表示成一个多项式加上一个真分式的形式. 使用多项式长除法:

$$\begin{array}{r|l}
 & x-1 \\
\hline
x^2+x-2 & x^3 \\
 & -x^3-x^2+2x \\
\hline
 & -x^2+2x \\
 & x^2+x-2 \\
\hline
 & 3x-2
\end{array}$$

因此有,

$$\frac{x^3}{x^2+x-2}=x-1+\frac{3x-2}{x^2+x-2}.$$

再对真分式 $\frac{3x-2}{x^2+x-2}$ 做部分分式分解, 设

$$\frac{3x-2}{x^2+x-2}=\frac{3x-2}{(x-1)(x+2)}=\frac{A_1}{x-1}+\frac{A_2}{x+2},$$

其中 A_1, A_2 为待定常数. 上式两端乘以 $(x-1)(x+2)$, 得到

$$3x-2=A_1(x+2)+A_2(x-1).$$

令 $x=1$, 得 $A_1=\frac{1}{3}$; 令 $x=-2$, 得 $A_2=\frac{8}{3}$. 因此

$$\begin{aligned}
\int \frac{x^3}{x^2+x-2}\,\mathrm{d}x &= \int\left(x-1+\frac{\frac{1}{3}}{x-1}+\frac{\frac{8}{3}}{x+2}\right)\,\mathrm{d}x \\
&= \frac{1}{2}x^2-x+\frac{1}{3}\ln|x-1|+\frac{8}{3}\ln|x+2|+C.
\end{aligned}$$

□

3.3.3 三角函数的有理式的不定积分

下面我们讨论三角函数的有理式的不定积分.

二元多项式和有理式

到现在为止, 我们学习过的多项式其实都是一元多项式 (也称作单变量多项式), 即只关于单个变量 x 的一类函数:

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = \sum_{i=0}^{n} a_i x^i.$$

而**二元多项式** (也称作双变量多项式) 是一元多项式的扩展, 它涉及两个独立的变量 x 和 y, 并包含它们的各个幂次项及其系数. 一般地, 一个二元多项式 $P(x, y)$ 可以写成以下形式:

$$P(x, y) = \sum_{i=0}^{n} \sum_{j=0}^{m} a_{i,j} x^i y^j.$$

其中, x 和 y 是多项式中的两个变量, $a_{i,j}$ 是相应项的系数, 通常是实数或复数, i 和 j 是非负整数, 表示 x 和 y 的幂次, n 和 m 分别是 x 和 y 的最大幂次. 以下都是二元多项式:

1. $P(x, y) = xy$
2. $P(x, y) = x \pm y$
3. $P(x, y) = 3x^2 + 2xy + y^2$
4. $P(x, y) = 5x^3 - 4x^2 y + 2xy^2 + y^3 + 1$
5. $P(x, y) = 7x^4 - 3x^2 y^2 + 6y^3 + 5$

二元多项式的 **度 (degree)** 是其中最高次项的次数. 例如, $x^2y^2 + 3x^3 + 4y$ 的度是 4, 这与项 x^2y^2 的次数相同. $3x^2 + 2xy + y^2$ 的最高次项是 x^2 和 y^2, 所以它的度为 2. 根据度的定义, 我们可以知道:

1. 所有度为 0 的二元多项式都形如
$$P(x, y) = a_{0,0},$$
其中 $a_{0,0}$ 是一个常数, 故这只是一个常数函数.
2. 所有度为 1 的二元多项式都形如
$$P(x, y) = a_{1,0} x + a_{0,1} y + a_{0,0},$$
其中 $a_{1,0}$ 和 $a_{0,1}$ 不同时为零. 例如, $2x + 4y + 5$, $-2x + 3$, x, y, 以及 $x - y$.
3. 所有度为 2 的二元多项式都形如
$$P(x, y) = a_{2,0} x^2 + a_{1,1} xy + a_{0,2} y^2 + a_{1,0} x + a_{0,1} y + a_{0,0},$$
其中 $a_{2,0}$, $a_{1,1}$ 和 $a_{0,2}$ 不同时为零. 例如, $2x^2 + 4xy + 7y^2 + 3x + 2y - 8$, $x^2 - xy$, $x^2 + y^2 - 1$, $xy + x - 3$.

上一小节中涉及的有理式, 我们也只是在讨论单变量 x 的有理式 (rational function), 它被定义为任意两个关于 x 的一元多项式之比, 即

$$\frac{P(x)}{Q(x)}.$$

类似地, 我们也可以定义**关于两个变量 x, y 的有理式**, 它是任意两个关于 x, y 的二元多项式之比, 即

$$\frac{P(x, y)}{Q(x, y)}.$$

三角函数的有理式

一般来说, 设

$$R(u,v)=\frac{P(u,v)}{Q(u,v)}$$

是关于两个变量 u,v 的有理式 (即, 关于 u,v 的两个二元多项式之商). 这告诉我们了一个二元函数:

$$(u,v)\mapsto R(u,v)=\frac{P(u,v)}{Q(u,v)}.$$

我们可以用 $u=\sin x$ 和 $v=\cos x$ 做换元, 然后就可以得到一个只关于 x 的一元函数:

$$x\mapsto R(\sin x,\cos x)=\frac{P(\sin x,\cos x)}{Q(\sin x,\cos x)}.$$

显然, 该函数长成关于 $\sin x$ 和 $\cos x$ 的函数表达式的样子. 比如 $\dfrac{\cos x}{\sin x+\cos x}$, $\dfrac{\cos x^2+\sin x+1}{2\sin x+\cos x}$.

定义 3.3.1: 三角函数的有理式

若某个关于 $\sin x$ 和 $\cos x$ 的函数表达式

$$R(\sin x,\cos x),$$

通过 $\sin x=u$ 和 $\cos x=v$ 换元后的结果

$$R(u,v)$$

是一个关于两个变量 u,v 的有理式 (两个二元多项式之商), 则称 $R(\sin x,\cos x)$ 是一个**三角函数的有理式**. 注意, 由于三角恒等式:

$$\tan x=\frac{\sin x}{\cos x},\quad \cot x=\frac{\cos x}{\sin x},\quad \sec x=\frac{1}{\cos x},\quad \csc x=\frac{1}{\sin x},$$

故所有形如 $R(\sin x,\cos x,\tan x,\cot x,\sec x,\csc x)$ 的函数本质上也应该视作 $R(\sin x,\cos x)$.

以下是一些正例和反例.

1. $\dfrac{\cos^2 x+\sin^5 x}{2\sin x}$ 是一个三角函数的有理式, 因为可以写作

$$\frac{\cos^2 x+\sin^5 x}{2\sin x}=\left[\frac{v^2+u^5}{2u}\right]_{\sin x=u,\cos x=v},$$

而 $\dfrac{v^2+u^5}{2u}$ 是一个二元有理式.

2. $\sqrt{1+\sin^2 x}$ 不是一个三角函数的有理式. 因为

$$\sqrt{1+\sin^2 x}=\left[\sqrt{1+u^2}\right]_{\sin x=u,\cos x=v},$$

而 $\sqrt{1+u^2}$ 中出现了开根号, 这不是一个多项式.

3. $\sin\frac{1}{2}x+\sin x$ 不是一个三角函数的有理式. 二倍角公式有 $\sin^2\frac{1}{2}x=\dfrac{1-\cos x}{2}$, 故

$$\sin\frac{1}{2}x+\sin x=\sqrt{\frac{1-\cos x}{2}}+\sin x.$$

和上一个例子一样, 出现了开根号, 所以这不是一个多项式. (默认将 $\sin x = u, \cos x = v$ 记在心中即可, 可不用显示地写出来)

4. $\dfrac{\cot x}{\sin x + \cos x - 1}$ 是一个三角函数的有理式.

$$\frac{\cot x}{\sin x + \cos x - 1} = \frac{\cos x}{(\sin x + \cos x - 1)\sin x} = \frac{\cos x}{\sin^2 x + \sin x \cos x - \sin x}$$

是一个关于 $\sin x, \cos x$ 的二元有理式.

三角函数的有理式的不定积分

设 $R(\sin x, \cos x)$ 是一个三角函数的有理式, 而我们要求的不定积分是

$$\int R(\sin x, \cos x)\ \mathrm{d}x.$$

对于三角函数的有理式, 求不定积分的一般方法是做变量替换 (即第二类换元法). 设新元 t 为

$$t = \tan\frac{x}{2}.\quad (\text{用第二类换元的话讲, 就是 } t = \varphi^{-1}(x) = \tan\frac{x}{2})$$

也就是说, 令旧元 x 变为

$$x = 2\arctan t.\quad (\text{用第二类换元的话讲, 就是 } x = \varphi(t) = 2\arctan t)$$

这个变量替换俗称**万能替换**[6], 因为它将三角函数的有理式变为 t 的有理式, 而后者总是可积的. 下面, 我们看一下具体如何替换的.

1. 首先, 想办法用 $t = \tan\dfrac{x}{2}$ 去替换掉所有的 $\sin x$ (即把 $\sin x$ 写成一个关于 $t = \tan\dfrac{x}{2}$ 的函数):

$$\sin x = 2\sin\frac{x}{2}\cos\frac{x}{2} = \frac{2\sin\frac{x}{2}\cos\frac{x}{2}}{\cos^2\frac{x}{2} + \sin^2\frac{x}{2}} = \frac{2\tan\frac{x}{2}}{1 + \tan^2\frac{x}{2}}.$$

故我们有

$$\sin x = \frac{2t}{1 + t^2}.$$

2. 然后, 想办法用 $t = \tan\dfrac{x}{2}$ 去替换掉所有的 $\cos x$ (即把 $\cos x$ 写成一个关于 $t = \tan\dfrac{x}{2}$ 的函数): 我们知道三角恒等式 $\cos^2\dfrac{x}{2} + \sin^2\dfrac{x}{2} = 1$, 两边除以 $\cos^2\dfrac{x}{2}$ 得到 $1 + \tan^2\dfrac{x}{2} = \dfrac{1}{\cos^2\frac{x}{2}}$. 所以

$$\cos x = 2\cos^2\frac{x}{2} - 1 = 2\cdot\frac{1}{1 + \tan^2\frac{x}{2}} - 1 = \frac{2 - (1 + \tan^2\frac{x}{2})}{1 + \tan^2\frac{x}{2}} = \frac{1 - \tan^2\frac{x}{2}}{1 + \tan^2\frac{x}{2}}.$$

故我们有

$$\cos x = \frac{1 - t^2}{1 + t^2}.$$

[6]高中的时候, 我们就学习过万能公式:

$$\sin x = \frac{2\tan\frac{x}{2}}{1 + \tan^2\frac{x}{2}},\quad \cos x = \frac{1 - \tan^2\frac{x}{2}}{1 + \tan^2\frac{x}{2}},\quad \tan x = \frac{2\tan\frac{x}{2}}{1 - \tan^2\frac{x}{2}}.$$

我们这里的换元法本质上就是在运用这个公式.

3. 最后, 因为 $x = 2\arctan t$, 所以

$$\mathrm{d}x = \frac{2}{1+t^2}\ \mathrm{d}t.$$

结论 3.3.2

设 $R(\sin x, \cos x)$ 是一个三角函数的有理式. 要求三角函数的有理式的不定积分,

$$\int R(\sin x, \cos x)\ \mathrm{d}x,$$

总可以使用一种特定的第二类换元法操作. 即令 $x = 2\arctan t$ (或令 $t = \tan\dfrac{x}{2}$) 时, 总有

$$\sin x = \frac{2t}{1+t^2}, \quad \cos x = \frac{1-t^2}{1+t^2}, \quad \mathrm{d}x = \frac{2}{1+t^2}\ \mathrm{d}t.$$

(由前两个等式马上能推出

$$\tan x = \frac{2t}{1-t^2},\ \cot x = \frac{1-t^2}{2t},\ \sec x = \frac{1+t^2}{1-t^2},\ \csc x = \frac{1+t^2}{2t}.$$

有时直接使用这些结果也很有用) 所以, 原不定积分

$$\int R(\sin x, \cos x)\ \mathrm{d}x \quad \to \quad \int R\left(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}\right)\frac{2}{1+t^2}\ \mathrm{d}t.$$

这时, 被积函数

$$R\left(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}\right)\frac{2}{1+t^2}$$

变成了 t 的一个 (一元) 有理式, 而对于 (一元) 有理式的不定积分, 我们是会求的 (上一小节讨论它的一般解法). 总结下来, 就是

$$\int R(\sin x, \cos x)\ \mathrm{d}x = \left[\int R\left(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}\right)\frac{2}{1+t^2}\ \mathrm{d}t\right]_{t=\tan\frac{x}{2}}.$$

记着, 既然是第二类换元, 最后一步还是要把旧元 x 给再换回去.

(以下论述是补充材料. 还没学懵就看. 学懵了就以后再看.) 既然 $x = 2\arctan t$ (或令 $t = \tan\dfrac{x}{2}$) 是一种特定的第二类换元法的换元操作. 那么这个换元函数符合当初我们给出的诸多条件吗? 在此之前, 我们讨论一些关于周期函数的事情.

我们说 $f(x)$ 是周期为 T 的函数, 即对于所有 x, 恒有 $f(x+T) = f(x)$. 由于 $\sin x$ 和 $\cos x$ 具有周期 $T = 2\pi$, 因此, 若令 $f(x) = R(\sin x, \cos x)$, 则对于任意 x, 我们有

$$f(x+2\pi) = R(\sin(x+2\pi), \cos(x+2\pi)) = R(\sin x, \cos x) = f(x).$$

所以, 三角函数的有理式也是一个周期函数, 其周期也是 $T = 2\pi$. 于是, 我们有理由只关心被积函数在任意一个周期长度的区间上的原函数. 这里我们特意选择了区间 $x \in [-\pi, \pi]$. 这正是换元函数 $x = 2\arctan t$ ($\forall t \in \mathbf{R}$) 的值域, 也是 $t = \tan\dfrac{x}{2}$ 的定义域. 所以这个换元函数是符合我们一系列条件的 (单调可导, 导数恒不为零)! 可是, 我们最终在此换元法帮助下求得的原函数 $F(x)$ 也只是定义在 $x \in [-\pi, \pi]$ 上的. 一个值得大胆猜的的事实是, 是否原函数 $F(x)$ 也是周期函数, 而且周期也是 2π. 如果是, 那么原函数 $F(x)$ 的定义域就可以不用限定了. 函数 $F(x)$ 也可在所有 x 上有定义, 重要的是函数的表示不用做任何处理. 很可惜, 这并不是事实.

一般地, 如果 $f(x)$ 是周期函数, 它的导函数 $f'(x)$ 也必然是周期函数且具有相同的周期. 但是, 它的原函数却不一定是周期函数. 比如, 常函数 $f(x)=1$ 是周期函数, 但是任意的原函数 $x+C$ 都没有周期性. 再比如,

$$\int \sin^2 x \, \mathrm{d}x = \frac{x}{2} - \frac{\sin(2x)}{4} + C.$$

$\sin^2 x$ 是周期函数, 但 $\dfrac{x}{2} - \dfrac{\sin(2x)}{4}$ 不是. 可看图 周期函数的原函数不是周期函数 (反例) | Desmos. 回到我们关系的 $f(x)=R(\sin x, \cos x)$ 的问题上. 我们已经计算出了它在 $x\in[-\pi,\pi]$ 上原函数 $F(x)$. 其实, 我们并不是真的需要 $F(x)$ 具有周期性. 其实想说 $F(x)$ 在任何周期上的表达式应当是一致的. 我们再换一个区间, 比如 $x\in[\pi,3\pi]$ 上用类似的换元法求其原函数. 一个自然的选择是换元函数

$$x = 2\pi + 2\arctan t, \quad (\forall t\in\mathbf{R}), \quad \text{值域是 } [\pi, 3\pi].$$

也就是

$$t = \tan\left(\frac{x-2\pi}{2}\right) = \tan\left(\frac{x}{2}-\pi\right) = \tan\frac{x}{2}, \quad (\forall x\in[\pi,3\pi]).$$

这不就是我们在 $[-\pi,\pi]$ 上用的换元吗. 所以用完全一样的过程, 我们仍然有

$$\sin x = \frac{2t}{1+t^2}, \quad \cos x = \frac{1-t^2}{1+t^2}, \quad \mathrm{d}x = \ \mathrm{d}(2\pi + 2\arctan t) = \frac{2}{1+t^2} \ \mathrm{d}t.$$

我们还是能把问题转化成

$$\int R(\sin x, \cos x)\mathrm{d}x \to \int R\left(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}\right) \frac{2}{1+t^2} \ \mathrm{d}t.$$

所以, 最终求得的原函数 $F(x)$ 在表达式上, 它和在 $[-\pi,\pi]$ 上求得的 $F(x)$ 不会有任何区别. 只是背后隐含着不同定义域. 类似地, 再换一个区间 $[3\pi, 5\pi]$ 等等, 直到覆盖所有实数轴. 其表达示都是一样的.

例 3.3.11 求 $\displaystyle\int \frac{\cot x \ \mathrm{d}x}{\sin x + \cos x - 1}$.

证明. 令 $t=\tan\dfrac{x}{2}$, 则

$$\mathrm{d}x = \frac{2 \ \mathrm{d}t}{1+t^2}, \quad \sin x = \frac{2t}{1+t^2}, \quad \cos x = \frac{1-t^2}{1+t^2}, \quad \cot x = \frac{1-t^2}{2t}$$

从而

$$
\begin{aligned}
\int \frac{\cot x \ \mathrm{d}x}{\sin x+\cos x-1} &= \int \frac{\frac{1-t^2}{2t}\cdot\frac{2}{1+t^2}\ \mathrm{d}t}{\frac{2t}{1+t^2}+\frac{1-t^2}{1+t^2}-1} \\
&= \int \frac{\frac{1-t^2}{2t}\cdot\frac{2}{1+t^2}\ \mathrm{d}t}{\frac{2t+1-t^2-1-t^2}{1+t^2}} \\
&= \int \frac{\frac{1-t^2}{2t}\cdot\frac{2}{1+t^2}\ \mathrm{d}t}{\frac{2t-2t^2}{1+t^2}} \\
&= \int \frac{\left(1-t^2\right)}{t\cdot 2t(1-t)}\ \mathrm{d}t \\
&= \int \frac{1+t}{2t^2}\ \mathrm{d}t \\
&= \frac{1}{2}\int \frac{1}{t^2}\ \mathrm{d}t+\frac{1}{2}\int \frac{1}{t}\ \mathrm{d}t \\
&= -\frac{1}{2t}+\frac{1}{2}\ln|t|+C \\
&= -\frac{\cos\frac{x}{2}}{2\sin\frac{x}{2}}+\frac{1}{2}\ln\left|\tan\frac{x}{2}\right|+C.
\end{aligned}
$$

□

万能替换虽然对一般的三角函数的有理式总是适用的，但是有时比较麻烦．因此，在一个具体问题面前，如果可以用其他方法求出不定积分，则不一定要用这种替换，如下面各例．

例 3.3.12 求 $\displaystyle\int \frac{\sin x\cos x}{1+\sin^2 x}\ \mathrm{d}x$.

证明．此题中被积函数

$$\frac{\sin x\cos x}{1+\sin^2 x}$$

确实是一个三角函数的有理式．可以使用特定的第二类换元 $t=\tan\dfrac{x}{2}$．但是本题却有更简便的方法．因为被积函数是可写成 $f(\sin x)\cos x$ 的形式，这时很自然想到凑微分法！于是，

$$
\begin{aligned}
\int \frac{\sin x\cos x}{1+\sin^2 x}\ \mathrm{d}x &= \int \frac{\sin x\ \mathrm{d}(\sin x)}{1+\sin^2 x} \\
&= \frac{1}{2}\int \frac{\mathrm{d}\left(1+\sin^2 x\right)}{1+\sin^2 x} \\
&= \frac{1}{2}\ln\left(1+\sin^2 x\right)+C.
\end{aligned}
$$

□

本节的内容都是某种特定形式积分的通用算法介绍．确实比较难理解．它们的一个重要作用是开发自动求积分工具，这类软件都必须基于通用算法，而不可能像人类一样变通．

> **注意** 人类的思维富有创造性和灵活性，可以根据具体问题灵活选择最优解法．但对计算机来说，系统化的通用算法更有价值． 正是因为计算机具有准确性和不知疲倦的特点，它能够通过标准化的步骤，可靠地处理各种形式的积分运算．这种方法虽然可能不是最优雅的，但能确保结果的准确性和过程的可重复性．人类擅长创新和灵活思考，而计算机则在执行规范化流程时表现出色．

例 3.3.13 求 $\displaystyle\int \frac{\cos x\ \mathrm{d}x}{\sin x+\cos x}\quad (\cos x\neq 0)$.

证明. 被积函数仍是一个三角函数的有理式, 可以使用特定的换元函数 $t=\tan\frac{x}{2}$. 但我们再次使用另一个换元函数.

将被积函数的分子, 分母除以 $\cos x$, 并令 $t=\tan x$ (故 $x=\arctan t$), 则有

$$\begin{aligned}
\int \frac{\cos x\ \mathrm{d}x}{\sin x+\cos x} &= \int \frac{\mathrm{d}x}{1+\tan x}\\
&= \int \frac{1}{1+t}\cdot\frac{\mathrm{d}t}{1+t^2}\\
&= \frac{1}{2}\int\left(\frac{1}{1+t}-\frac{t-1}{1+t^2}\right)\ \mathrm{d}t \quad \text{(用上一小节的有理式拆分方法)}\\
&= \frac{1}{2}\int\left(\frac{1}{1+t}-\frac{t}{1+t^2}+\frac{1}{1+t^2}\right)\ \mathrm{d}t\\
&= \frac{1}{2}\left(\ln|1+t|-\frac{1}{2}\ln\left(1+t^2\right)+\arctan t\right)+C\\
&= \frac{1}{2}(\ln|1+\tan x|-\ln|\sec x|+x)+C\\
&= \frac{1}{2}(\ln|\cos x+\sin x|+x)+C.
\end{aligned}$$

□

例 3.3.14 求 $\displaystyle\int \sin^4 x\cos^2 x\ \mathrm{d}x$.

证明. 注意到被积函数的因子都是 $\sin x$ 或 $\cos x$ 的偶次方幂, 我们可以利用倍角公式降低其方幂次数. 于是, 我们有

$$\begin{aligned}
I &= \int \sin^4 x\cos^2 x\ \mathrm{d}x\\
&= \int \left(\frac{1-\cos 2x}{2}\right)^2\frac{1+\cos 2x}{2}\ \mathrm{d}x\\
&= \frac{1}{8}\int\left(\cos^3 2x-\cos^2 2x-\cos 2x+1\right)\ \mathrm{d}x
\end{aligned}$$

我们现在对各个部分分别进行积分.

- 对 $\cos^3 2x$ 项: 使用变量替换 $u=\sin 2x$, 则 $\mathrm{d}u=2\cos 2x\ \mathrm{d}x$,
$$\int\cos^3 2x\ \mathrm{d}x=\frac{1}{2}\int(1-u^2)\,\mathrm{d}u=\frac{1}{2}\left(u-\frac{u^3}{3}\right)=\frac{1}{2}\left(\sin 2x-\frac{1}{3}\sin^3 2x\right)+C.$$
- 对 $\cos^2 2x$ 项:
$$\int\cos^2 2x\ \mathrm{d}x=\frac{1}{2}\int(1+\cos 4x)\ \mathrm{d}x=\frac{1}{2}\left(x+\frac{1}{4}\sin 4x\right)+C.$$
- 对 $\cos 2x$ 项:
$$\int\cos 2x\ \mathrm{d}x=\frac{1}{2}\sin 2x+C.$$

将所有的积分结果代入到原式中:

$$
\begin{aligned}
I &= \frac{1}{8}\left(\frac{1}{2}\left(\sin 2x - \frac{1}{3}\sin^3 2x\right) - \frac{1}{2}\left(x + \frac{1}{4}\sin 4x\right) - \frac{1}{2}\sin 2x + x\right) + C \\
&= \frac{1}{16}\left(x - \frac{1}{3}\sin^3 2x - \frac{1}{4}\sin 4x\right) + C.
\end{aligned}
$$

这就是最后的结果. □

3.3.4 某些根式的不定积分

除了三角函数的有理式的不定积分之外, 某些特殊形式无理式的不定积分也能经过变量替换, 使被积函数变成新变量的有理式, 从而求出其不定积分.

结论 3.3.3

设 $R(u,v)$ 为 u,v 的二元有理式, 我们考虑不定积分

$$\int R(x,\, \sqrt[n]{ax+b})\ \mathrm{d}x$$

其中 a,b 为常数, 且 $a \neq 0$. 这种形式不定积分的被积函数总可以通过变量替换

$$t = \sqrt[n]{ax+b}$$

而有理化. 这时 $x = \dfrac{t^n - b}{a}$, 因而 $\mathrm{d}x = \dfrac{nt^{n-1}}{a}\ \mathrm{d}t$, 于是

$$\int R(x,\, \sqrt[n]{ax+b})\ \mathrm{d}x \quad \to \quad \int R\left(\frac{t^n - b}{a}, t\right)\frac{nt^{n-1}}{a}\ \mathrm{d}t.$$

上式右端的被积函数是 t 的有理式.

例 3.3.15 求 $\displaystyle\int \frac{\mathrm{d}x}{3x + \sqrt[3]{3x+2}}$.

证明. 令 $t = \sqrt[3]{3x+2}$, 这时 $x = \dfrac{1}{3}\left(t^3 - 2\right), \mathrm{d}x = t^2\ \mathrm{d}t$, 并有

$$\int \frac{\mathrm{d}x}{3x + \sqrt[3]{3x+2}} = \int \frac{t^2\ \mathrm{d}t}{t^3 + t - 2}.$$

这样, 原来的不定积分变成 t 的有理式的不定积分. 由于

$$\frac{t^2}{t^3 + t - 2} = \frac{1}{4}\left(\frac{1}{t-1} + \frac{3t+2}{t^2+t+2}\right).$$

根据前面的不定积分公式和有理式的第三种部分分式的积分公式 (前面证明过), 我们得到

$$\int \frac{t^2\ \mathrm{d}t}{t^3+t-2} = \frac{1}{4}\ln|t-1| + \frac{3}{8}\ln\left(t^2+t+2\right) + \frac{1}{4\sqrt{7}}\arctan\frac{2t+1}{\sqrt{7}} + C.$$

将 $t=\sqrt[3]{3x+2}$ 代入上式即得到最终结果:

$$\begin{aligned}\int \frac{\mathrm{d}x}{3x+\sqrt[3]{3x+2}} =& \frac{1}{4}\ln\left|\sqrt[3]{3x+2}-1\right| \\ &+\frac{3}{8}\ln\left(\sqrt[3]{(3x+2)^2}+\sqrt[3]{3x+2}+2\right) \\ &+\frac{1}{4\sqrt{7}}\arctan\frac{2\sqrt[3]{3x+2}+1}{\sqrt{7}}+C.\end{aligned}$$

□

更一般地, 对于形如

$$\int R\left(x,\sqrt[n]{\frac{ax+b}{cx+d}}\right)\,\mathrm{d}x$$

的不定积分 (a,b,c,d 为常数且 $a,c\neq 0$), 可用类似于上面的方法, 通过变量替换 $t=\sqrt[n]{\dfrac{ax+b}{cx+d}}$ 将被积函数有理化.

注意 学到这里, 当初说的第二换元应该检查换元

例 3.3.16 求 $\displaystyle\int x\sqrt{\frac{x-1}{x+1}}\,\mathrm{d}x$.

证明. 令 $t=\sqrt{\dfrac{x-1}{x+1}}$, 这时 $x=\dfrac{1+t^2}{1-t^2}$, $\mathrm{d}x=\dfrac{4t}{\left(1-t^2\right)^2}\,\mathrm{d}t$, 并有

$$\int x\sqrt{\frac{x-1}{x+1}}\,\mathrm{d}x=\int\frac{1+t^2}{1-t^2}\cdot t\cdot\frac{4t}{\left(1-t^2\right)^2}\,\mathrm{d}t=\int\frac{4t^2\left(1+t^2\right)}{\left(1-t^2\right)^3}\,\mathrm{d}t$$

这样, 原来的不定积分变成 t 的有理式的不定积分. 以下便可以按照已知的方法求出关于 t 的原函数, 然后以 $t=\sqrt{\dfrac{x-1}{x+1}}$ 代入即可 (请读者自行完成). □

求不定积分的方法还有许多, 我们不能在这里一一介绍. 我们应指出: 在计算机技术高度发达的今天, 一些功能强大的数学软件包, 不仅仅能进行各种数值计算, 而且能进行某些逻辑符号演算, 其中包括微分运算及积分运算.

但是, 对于初学者而言, 基本积分表, 换元法, 分部积分法以及有理化方法还应掌握, 因为求不定积分是解微分方程和求定积分的基础, 我们总不能对一些非常基本的不定积分也要查书或求助于计算机吧.

3.4 定积分的分部积分法与换元法

牛顿 - 莱布尼茨公式建立了定积分与原函数之间的联系:

$$\int_a^b f(x)\mathrm{d}x=\left.F(x)\right|_a^b,$$

其中 $F(x)$ 是 $f(x)$ 的一个原函数. 因此, 从原则上说, 只要求出被积函数的一个原函数, 即可计算出定积分的值. 但是, 求出原函数有时并非易事, 甚至不可能. 所以, 我们有必要直接讨论有关定积分计算的一些基本方法. 特别是某些特定情况下, 即使不用求出原函数, 也能计算出其定积分.

3.4.1　定积分的分部积分法

不定积分的分部积分公式是

$$\int u(x)v'(x)\mathrm{d}x = u(x)v(x) - \int v(x)u'(x)\mathrm{d}x.$$

由此及牛顿 - 莱布尼茨公式立即推出定积分的分部积分公式.

定理 3.4.1

设函数 $u(x)$ 及 $v(x)$ 在区间 $[a,b]$ 上可导且其导数在 $[a,b]$ 上连续, 则我们有

$$\int_a^b u(x)v'(x)\mathrm{d}x = u(x)v(x)\big|_a^b - \int_a^b v(x)u'(x)\mathrm{d}x.$$

也可以写成如下形式

$$\int_a^b u(x)\mathrm{d}(v(x)) = u(x)v(x)\big|_a^b - \int_a^b v(x)\mathrm{d}(u(x)).$$

这就是定积分的分部积分公式, 它给出了定积分的分部积分法.

例 3.4.1 求定积分 $\displaystyle\int_1^2 x\ln x \,\mathrm{d}x$.

证明. 我们将所求的定积分写成如下形式

$$\int_1^2 x\ln x\,\mathrm{d}x = \frac{1}{2}\int_1^2 \ln x\,\mathrm{d}\left(x^2\right),$$

由定积分的分部积分公式得

$$\begin{aligned}\int_1^2 x\ln x\,\mathrm{d}x &= \frac{1}{2}x^2\ln x\Big|_1^2 - \frac{1}{2}\int_1^2 x^2\,\mathrm{d}(\ln x)\\ &= 2\ln 2 - \frac{1}{2}\int_1^2 x\,\mathrm{d}x\\ &= 2\ln 2 - \frac{1}{4}x^2\Big|_1^2\\ &= 2\ln 2 - \frac{3}{4}.\end{aligned}$$

这个例题当然也可以先用不定积分的分部积分公式求得原函数

$$\int x\ln x\,\mathrm{d}x = \frac{1}{2}x^2\ln x - \frac{1}{4}x^2 + C,$$

再用牛顿 - 莱布尼茨公式, 得

$$\int_1^2 x\ln x\,\mathrm{d}x = \left(\frac{1}{2}x^2\ln x - \frac{1}{4}x^2\right)\Big|_1^2 = 2\ln 2 - \frac{3}{4}.$$

这两种方法本质上是一样的, 但通常情况下我们直接使用定积分的分部积分法. □

利用定积分的分部积分法有时不能一步得出结果, 但它可能导出某个递推公式, 从而推出所求的定积分值. (其实在学习不定积分的分部积分法时我们就遇到过一样的事情) 下面是一个典型的例子.

例 3.4.2 求定积分

$$I_n=\int_0^{\frac{\pi}{2}}\sin^n x\ \mathrm{d}x\quad (n\ 为整数且 n\geqslant 2).$$

证明. 先将 I_n 写成

$$I_n=-\int_0^{\frac{\pi}{2}}\sin^{n-1}x\ \mathrm{d}(\cos x),$$

则由定积分的分部积分公式得

$$\begin{aligned}
I_n&=-\left(\sin^{n-1}x\cos x\right)\Big|_0^{\frac{\pi}{2}}+\int_0^{\frac{\pi}{2}}\cos x\ \mathrm{d}\left(\sin^{n-1}x\right)\\
&=(n-1)\int_0^{\frac{\pi}{2}}\sin^{n-2}x\cos^2 x\ \mathrm{d}x\quad (注意到\ \sin 0=0,\cos\frac{\pi}{2}=0,\ 故第一项直接是\ 0)\\
&=(n-1)\int_0^{\frac{\pi}{2}}\sin^{n-2}x\left(1-\sin^2 x\right)\mathrm{d}x\\
&=(n-1)\int_0^{\frac{\pi}{2}}\left(\sin^{n-2}x-\sin^n x\right)\mathrm{d}x\\
&=(n-1)I_{n-2}-(n-1)I_n.
\end{aligned}$$

这样, 我们解出递推公式

$$I_n=\frac{n-1}{n}I_{n-2}.$$

每用一次这个递推公式, I_n 的下标就减少 2, 故依照 n 的奇偶性最后可能达到 I_0 或 I_1, 令 k 为大于 1 的正整数, 则

$$\begin{aligned}
I_{2k}&=\frac{2k-1}{2k}\cdot\frac{2k-3}{2k-2}\cdot\cdots\cdot\frac{1}{2}I_0,\quad (当\ n=2k),\\
I_{2k+1}&=\frac{2k}{2k+1}\cdot\frac{2k-2}{2k-1}\cdot\cdots\cdot\frac{2}{3}I_1,\quad (当\ n=2k+1).
\end{aligned}$$

其中

$$I_0=\int_0^{\frac{\pi}{2}}(\sin x)^0\ \mathrm{d}x=\int_0^{\frac{\pi}{2}}\ \mathrm{d}x=\frac{\pi}{2},\quad I_1=\int_0^{\frac{\pi}{2}}\sin x\ \mathrm{d}x=(-\cos x)\big|_0^{\frac{\pi}{2}}=1.$$

综合上面的结果可简写成

$$\begin{aligned}
\int_0^{\frac{\pi}{2}}\sin^{2k}x\ \mathrm{d}x&=\frac{(2k-1)!!}{(2k)!!}\cdot\frac{\pi}{2},\quad (当\ n=2k),\\
\int_0^{\frac{\pi}{2}}\sin^{2k+1}x\ \mathrm{d}x&=\frac{(2k)!!}{(2k+1)!!},\quad (当\ n=2k+1).
\end{aligned}$$

这里 $n!!$ 表示不超过 n 的 "**二进阶乘**", 如

$$8!!=8\times 6\times 4\times 2,\quad 7!!=7\times 5\times 3\times 1.$$

二进阶乘的要点是第一项要从最大的正整数 n 开始, 往后每一个项减去 2, 最后一项根据 n 的奇偶性才能确定是 2 还是 1. 不要学通常的阶乘从最小的数 1 开始, 往上加上 2. 这一点容易搞错. □

3.4.2 定积分的换元法

定理 3.4.2: 定积分的换元法

设函数 $f(x)$ 在区间 $[A,B]$ 上连续, 且 $a,b\in[A,B]$. 现欲求定积分

$$\int_a^b f(x)\mathrm{d}x.$$

若有换元函数 $x=\varphi(t)$ 满足:

- 其定义域为区间 $[\alpha,\beta]$, 其在该区间上有连续的导函数[a],
- $\varphi(\alpha)=a,\varphi(\beta)=b$ 且 $\varphi([\alpha,\beta])\subseteq[A,B]$,

则有

$$\int_a^b f(x)\mathrm{d}x=\int_\alpha^\beta f(\varphi(t))\varphi'(t)\mathrm{d}t.$$

这就是定积分的换元法.

[a]这里要求其导函数自身也是连续. 一般而言, 可导函数 (自然是连续函数) 的导函数不一定总是连续的.

证明. 我们知道任意连续函数都有原函数 (前几节我们证明过变上限积分函数就是一个原函数). 所以, 可以设 $F(x)$ 是函数 $f(x)$ 的一个原函数. 因为 $\varphi([\alpha,\beta])\subseteq[A,B]$, 这使得 $F(\varphi(t))$ 有意义, 即可以将其视作定义在区间 $[\alpha,\beta]$ 上的合法函数.

根据条件, 换元函数 $x=\varphi(t)$ 在区间 $[\alpha,\beta]$ 上有导函数 $\varphi'(t)$ 且 $\varphi'(t)$ 是连续的. 由链式法则可知, $F(\varphi(t))$ 是 $f(\varphi(t))\varphi'(t)$ 的一个原函数. 注意, $f(\varphi(t))\varphi'(t)$ 在区间 $[\alpha,\beta]$ 上也是连续的 (因为连续函数的积还是连续函数), 从而确保 $f(\varphi(t))\varphi'(t)$ 是可积函数. 于是, 由微积分基本定理可知

$$\int_\alpha^\beta f(\varphi(t))\varphi'(t)\mathrm{d}t=F(\varphi(t))\big|_\alpha^\beta=F(\varphi(\beta))-F(\varphi(\alpha)).$$

将 $a=\varphi(\alpha),b=\varphi(\beta)$ 代入上式, 则有

$$F(\varphi(\beta))-F(\varphi(\alpha))=F(b)-F(a)=\int_a^b f(x)\mathrm{d}x.$$

这样, 我们得到

$$\int_a^b f(x)\mathrm{d}x=\int_a^\beta f(\varphi(t))\varphi'(t)\mathrm{d}t.$$

证毕. □

这里我们指出:

- **定积分的换元法中, 可以不要求换元函数 $x=\varphi(t)$ 有反函数,** 毕竟无论该方法的证明过程中还是结论我们都没有使用到反函数. 这一点和求不定积分的换元法有所不同. 在求不定积分时, 如果将积分变量 x 通过 $x=\varphi(t)$ 而换成 t, 在求出关于 t 的原函数后, 还应该将 t 反解为 x 的函数 (因为我们要求的是关于 x 的原函数). 但是, 在求定积分时, 这一步可以不要了, 而只需相应地改变其积分上, 下限就可以了.
- **正确地写出换元后的积分上, 下限是十分重要的.** 当所做的变量替换为 $x=\varphi(t)$ 时,
  1. t 的积分下限 α 应与 x 的积分下限 a 对应, 即 $a=\varphi(\alpha)$,

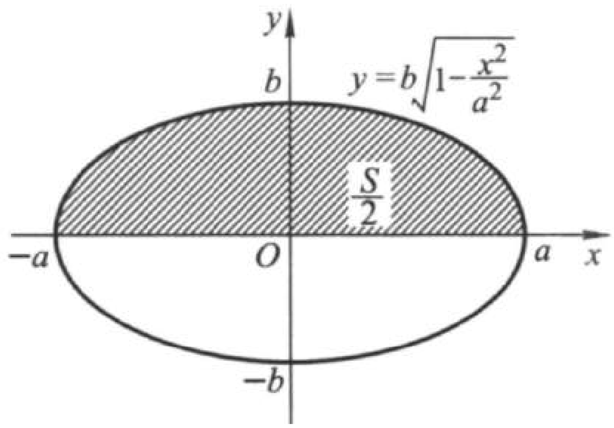


图 3.2: 椭圆

2. t 的积分上限 β 应与 x 的积分上限 b 对应, 即 $b=\varphi(\beta)$,
3. 只需考虑积分上, 下限各自的对应关系, 无须考虑 α,β 中谁大谁小. 这是因为, 该方法的证明的基础是牛顿 - 莱布尼茨公式, 而这一公式并不限定积分上限大于积分下限.

例 3.4.3 求定积分 $\displaystyle\int_0^1 \frac{\sqrt{x}}{1+\sqrt{x}}\,\mathrm{d}x$.

证明. 令 $\sqrt{x}=t\geq 0$, 这时 $x=t^2$ $(0\leq t\leq 1)$, $\mathrm{d}x=2t\ \mathrm{d}t$, 又 $t=0,1$ 时分别与 $x=0,1$ 对应, 故有

$$\begin{aligned}
\int_0^1 \frac{\sqrt{x}}{1+\sqrt{x}}\ \mathrm{d}x &= 2\int_0^1 \frac{t^2\ \mathrm{d}t}{t+1} \\
&= 2\int_0^1 \left(t-1+\frac{1}{t+1}\right)\mathrm{d}t \\
&= 2\left(\frac{t^2}{2}-t+\ln(1+t)\right)\Bigg|_0^1 \\
&= 2\ln 2-1.
\end{aligned}$$

注意, 在我们这个具体的积分区间内, $0\leq t\leq 1$, 所以是 $\ln(1+t)$ 而不是 $\ln|1+t|$. 定积分中求原函数的时候, 一定要清楚积分区间, 有时可以把原函数变得简单.

在本题中, 相当于定理 3.4.2 中 $x=\varphi(t)=t^2, a=0, b=1$. 做换元后积分下限 α 应满足 $\varphi(\alpha)=0$, 因此 $\alpha=0$; 积分上限 β 应满足 $\varphi(\beta)=1$, 这时 β 可取 1 或 -1, 但是在上述变换中令 $t=\sqrt{x}$, 因而 t 应该取非负值, 故 -1 不合要求. 另外, 虽然原则上不要求换元函数是可逆的, 在本题中, 函数 $x=t^2$ 在 $[0,1]$ 上确实是单调的, 故反函数存在. 其实, $x=t^2$ 是将新的积分区域 $[\alpha,\beta]=[0,1]$ 映射回原本的积分区域 $[a,b]=[0,1]$. □

例 3.4.4 计算椭圆 (见图 3.2)

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}\leqslant 1 \quad (a,b>0)$$

的面积 S.

证明. 显然, 所要求的面积可以写成定积分:

$$S=2\int_{-a}^{a} b\sqrt{1-\frac{x^2}{a^2}}\ \mathrm{d}x.$$

画出辅助三角形可知, 应该令 $x=a\sin t$. 则有 $\cos t=\sqrt{1-\frac{x^2}{a^2}}$. 注意到, 如果限定 $-\frac{\pi}{2}\leqslant t\leqslant\frac{\pi}{2}$, 那么换元函数 $x=a\sin t$ 正好将新的积分区域 $[-\frac{\pi}{2},\frac{\pi}{2}]$ 映射回原本的积分区域 $[-a,a]$, 并且两个端点对应 (注意对应次序). 则有

$$\begin{aligned}
S&=2b\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos t\cdot a\cos t\ \mathrm{d}t=2ab\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos^2 t\ \mathrm{d}t\\
&=2ab\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{1}{2}(1+\cos 2t)\mathrm{d}t\\
&=ab\left(t+\frac{\sin 2t}{2}\right)\bigg|_{-\frac{\pi}{2}}^{\frac{\pi}{2}}=ab\pi.
\end{aligned}$$

这就证明了椭圆的面积公式

$$S=ab\pi.$$

这表明, 椭圆的面积等于两个半轴之积乘以 π. 当椭圆的两个半轴相等时, 椭圆就化成圆, 而这时上述公式恰好就是圆的面积公式. □

例 3.4.5 证明: $\int_0^{\frac{\pi}{2}}\cos^n x\ \mathrm{d}x=\int_0^{\frac{\pi}{2}}\sin^n x\ \mathrm{d}x$.

证明. 令换元函数 $x=\frac{\pi}{2}-t$, 观察发现它的图像可知, 它将新的积分区域 $t\in[0,\frac{\pi}{2}]$ 映射回原本的积分区域 $x\in[0,\frac{\pi}{2}]$, 并且两个端点对应, 但此时对应的顺序正好反向: $t=0$ 对应 $x=\frac{\pi}{2}$, $t=\frac{\pi}{2}$ 对应 $x=0$. 则有

$$\int_0^{\frac{\pi}{2}}\cos^n x\ \mathrm{d}x=\int_{\frac{\pi}{2}}^{0}\sin^n t\cdot(-\mathrm{d}t)=\int_0^{\frac{\pi}{2}}\sin^n t\ \mathrm{d}t=\int_0^{\frac{\pi}{2}}\sin^n x\ \mathrm{d}x.$$

由本例题可知, 例 3.4.2 中关于定积分 $\int_0^{\frac{\pi}{2}}\sin^n x\ \mathrm{d}x$ 的记号及递推公式对于定积分 $\int_0^{\frac{\pi}{2}}\cos^n x\ \mathrm{d}x$ 也适用. □

例 3.4.6 求定积分 $\int_0^a x^4\sqrt{a^2-x^2}\ \mathrm{d}x\quad(a>0)$.

证明. 画出辅助三角形可知, 令 $x=a\cos t$ (令 $x=a\cos t$ 也是没有问题的) 并且 $0\leqslant t\leqslant\frac{\pi}{2}$, 注意到 $t=0$ 时 $x=a, t=\frac{\pi}{2}$ 时 $x=0$, 根据积分上, 下限的对应原则, 在积分变量换成 t 后, 积分上限为 0, 而积分下限为 $\frac{\pi}{2}$ (注意: 这里积分下限比积分上限大), 故有

$$\begin{aligned}
\int_0^a x^4\sqrt{a^2-x^2}\ \mathrm{d}x&=\int_{\frac{\pi}{2}}^{0}a^4\cos^4 t\cdot a\sin t\cdot(-a\sin t)\mathrm{d}t\\
&=a^6\int_0^{\frac{\pi}{2}}\cos^4 t\left(1-\cos^2 t\right)\mathrm{d}t\\
&=a^6\left(I_4-I_6\right).
\end{aligned}$$

由例题 3.4.2 中得到了递推公式

$$I_n = \int_0^{\frac{\pi}{2}} \cos^n x \, \mathrm{d}x = \int_0^{\frac{\pi}{2}} \sin^n x \, \mathrm{d}x = \frac{n-1}{n} I_{n-2}.$$

所以, $I_6 = \dfrac{5}{6} I_4$. 则

$$\int_0^a x^4 \sqrt{a^2 - x^2} \, \mathrm{d}x = a^6 \frac{5}{6} I_4 = \frac{\pi}{32} a^6.$$

最后直接使用了例题 3.4.2 的结果来计算 I_4:

$$I_{2k} = \frac{(2k-1)!!}{(2k)!!} \cdot \frac{\pi}{2}.$$

□

做适当的变量替换往往成为某些定积分计算的关键, 下面就是一个有趣的例子, 但该例技巧性很强, 没有一般意义.

例 3.4.7 求定积分 $\displaystyle\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sin^2 x}{1+\mathrm{e}^x} \, \mathrm{d}x$.

证明. 首先, 将原定积分按积分区域拆分:

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sin^2 x}{1+\mathrm{e}^x} \, \mathrm{d}x = \int_{-\frac{\pi}{2}}^{0} \frac{\sin^2 x}{1+\mathrm{e}^x} \, \mathrm{d}x + \int_0^{\frac{\pi}{2}} \frac{\sin^2 x}{1+\mathrm{e}^x} \, \mathrm{d}x.$$

对上式右端的第一个定积分做变量替换 $x = -t$, 我们有

$$\begin{aligned}
\int_{-\frac{\pi}{2}}^{0} \frac{\sin^2 x}{1+\mathrm{e}^x} \, \mathrm{d}x &= \int_{\frac{\pi}{2}}^{0} \frac{\sin^2 t}{1+\mathrm{e}^{-t}} (-\,\mathrm{d}t) \\
&= -\int_{\frac{\pi}{2}}^{0} \frac{\sin^2 t}{1+\mathrm{e}^{-t}} \, \mathrm{d}t \quad (\text{同乘以 } e^t) \\
&= \int_0^{\frac{\pi}{2}} \frac{\mathrm{e}^t}{1+\mathrm{e}^t} \sin^2 t \, \mathrm{d}t \\
&= \int_0^{\frac{\pi}{2}} \frac{\mathrm{e}^x}{1+\mathrm{e}^x} \sin^2 x \, \mathrm{d}x.
\end{aligned}$$

于是, 我们得到

$$\begin{aligned}
\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sin^2 t}{1+\mathrm{e}^t} \, \mathrm{d}t &= \int_0^{\frac{\pi}{2}} \left(\frac{1}{1+\mathrm{e}^x} + \frac{\mathrm{e}^x}{1+\mathrm{e}^x} \right) \sin^2 x \, \mathrm{d}x \\
&= \int_0^{\frac{\pi}{2}} \sin^2 x \, \mathrm{d}x = \frac{\pi}{4}.
\end{aligned}$$

□

3.4.3 偶函数, 奇函数及周期函数的定积分

现在我们通过定积分的换元法导出奇函数, 偶函数及周期函数的定积分的一些特殊性质. 为此, 先回顾一下这些函数的定义.

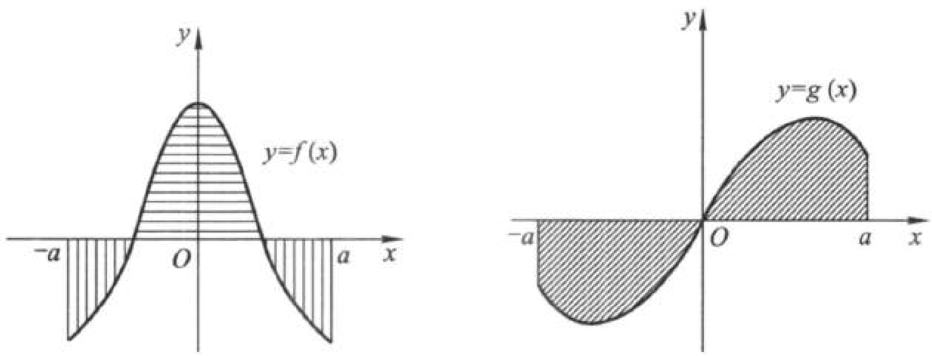


图 3.3: 奇偶函数

一个定义在**对称区间** $[-a,a]$ 上的函数 $f(x)$, 若满足

$$f(x)=f(-x),\quad \forall x\in[-a,a],$$

则称之为**偶函数**; 若满足

$$f(x)=-f(-x),\quad \forall x\in[-a,a],$$

则称之为**奇函数**. 大家熟知的余弦函数 $\cos x$ 是偶函数, 而正弦函数 $\sin x$ 是奇函数. 从函数的图形看, 偶函数关于 y 轴对称, 而奇函数关于原点对称.

命题 3.4.1

设 $f(x)$ 为区间 $[-a,a]$ 上的偶函数, 而 $g(x)$ 为区间 $[-a,a]$ 上的奇函数, 并假定它们都是可积的, 则有

$$\int_{-a}^{a} f(x)\mathrm{d}x=2\int_{0}^{a} f(x)\mathrm{d}x,\quad \int_{-a}^{a} g(x)\mathrm{d}x=0.$$

这个结论的几何意义是十分明显的 (见图 3.3). 在以原点为中心的对称区间上对奇函数积分, 其积分值为零; 在这样的区间上对偶函数积分, 其积分值等于在右半区间上积分值的两倍.

证明. 这个结论的证明也是容易的.

- 事实上,
$$\int_{-a}^{a} f(x)\mathrm{d}x=\int_{0}^{a} f(x)\mathrm{d}x+\int_{-a}^{0} f(x)\mathrm{d}x.$$
因 $f(x)$ 是偶函数, $f(-x)=f(x)$, 故对上式右端的第二个定积分做变量替换 $x=-t$ 后有
$$\int_{-a}^{0} f(x)\mathrm{d}x=\int_{a}^{0} f(-t)\mathrm{d}(-t)=\int_{0}^{a} f(-t)\mathrm{d}t=\int_{0}^{a} f(t)\mathrm{d}t.$$
于是, 我们得到
$$\int_{-a}^{a} f(x)\mathrm{d}x=2\int_{0}^{a} f(x)\mathrm{d}x.$$
- 注意到奇函数 $g(x)=-g(-x)$, 并注意到
$$\int_{-a}^{0} g(x)\mathrm{d}x=-\int_{a}^{0} g(-t)\mathrm{d}t=-\int_{0}^{a} g(t)\mathrm{d}t=-\int_{0}^{a} g(x)\mathrm{d}x,$$

我们可得到

$$\int_{-a}^{a} g(x)\mathrm{d}x = 0.$$

□

利用命题 3.4.1 在计算某些定积分时可简化计算.

例 3.4.8 求定积分 $\displaystyle\int_{-\pi}^{\pi} \frac{\sin x}{\sqrt{1+x^4}}\ \mathrm{d}x$.

证明. 因为 $\dfrac{\sin x}{\sqrt{1+x^4}}$ 是奇函数, 所以 $\displaystyle\int_{-\pi}^{\pi} \frac{\sin x}{\sqrt{1+x^4}}\ \mathrm{d}x = 0$. □

对于上题背记函数的奇偶性, 一个便捷的判断方法是利用下面的结果, 其结论类似于正负号加减乘除的效果. (它的证明是非常容易的)

1. **和的性质**:
    - 两个偶函数的和是偶函数.
    - 两个奇函数的和是奇函数.
2. **乘积的性质**:
    - 两个偶函数的乘积是偶函数.
    - 两个奇函数的乘积是偶函数.
    - 偶函数与奇函数的乘积是奇函数.
3. **商的性质** (假设分母不为零):
    - 两个偶函数的商是偶函数.
    - 两个奇函数的商是偶函数.
    - 偶函数与奇函数的商 (不论分子分母顺序) 是奇函数.

这些结论适用于定义在区间 $(-l, l)$ 上的函数.

例 3.4.9 求定积分 $\displaystyle\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^5 x\ \mathrm{d}x$.

证明. 注意到 $\cos^5 x$ 为偶函数, 我们有

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^5 x\ \mathrm{d}x = 2\int_{0}^{\frac{\pi}{2}} \cos^5 x\ \mathrm{d}x = \frac{16}{15}.$$

最后直接使用了例题 3.4.2 的结果来计算 I_5. □

上题中隐藏着一个简单的事实. 对于任何 x 值, $\cos(x)=\cos(-x)$ 成立, 因此无论 f 是什么函数, 都有

$$f(\cos(-x))=f(\cos(x)).$$

因此, $f(\cos(x))$ 总是偶函数. 更一般地, 如果 $g(x)$ 是一个偶函数, 则对于任意函数 f, 复合函数 $f(g(x))$ 也是偶函数. 这是因为

$$f(g(-x))=f(g(x)).$$

当然定义域必须是对称的. 若 $g(x)$ 是一个奇函数, 则 $f(g(x))$ 不一定是奇函数. 这取决于 f 的具体形式. 假设 f 是奇函数, 我们有

$$f(g(-x))=f(-g(x))=-f(g(x)),$$

因此 $f(g(x))$ 是奇函数. 假设 f 是偶函数, 我们有

$$f(g(-x))=f(-g(x))=f(g(x)),$$

因此 $f(g(x))$ 是偶函数. 总之, 我们有以下结论.

结论 3.4.1

若 $g(x)$ 是偶函数, 则 $f(g(x))$ 对任何 f 都是偶函数. 若 $g(x)$ 是奇函数, 则需要进一步分类:

1. 如果 f 是奇函数, 则 $f(g(x))$ 是奇函数.
2. 如果 f 是偶函数, 则 $f(g(x))$ 是偶函数.
3. 如果 f 既不是奇函数也不是偶函数, 则 $f(g(x))$ 可能既不是奇函数也不是偶函数.

例 3.4.10 定积分 $\displaystyle\int_{-2}^{2}\left(x\sin^4 x+x^3-x^4\right)\mathrm{d}x$.

证明. 注意到 $x\sin^4 x$ 和 x^3 是奇函数, 而 x^4 是偶函数, 则

$$\begin{aligned}\int_{-2}^{2}\left(x\sin^4 x+x^3-x^4\right)\mathrm{d}x&=\int_{-2}^{2}x\sin^4 x\ \mathrm{d}x+\int_{-2}^{2}x^3\ \mathrm{d}x-\int_{-2}^{2}x^4\ \mathrm{d}x\\&=-2\int_{0}^{2}x^4\ \mathrm{d}x=-\left.\frac{2}{5}x^5\right|_0^2=-\frac{64}{5}.\end{aligned}$$

□

设函数 $f(x)$ 在区间 $(-\infty,+\infty)$ 上有定义. 若存在一个数 $T\neq 0$ (理论上可以为负数), 使得对于一切 $x\in(-\infty,+\infty)$, 都有

$$f(x+T)=f(x),$$

则称 $f(x)$ 是一个**周期函数**, 而称 T 为 $f(x)$ 的**周期**.

若 $f(x)$ 是一个以 T 为周期的函数, 则对于任意整数 $n\neq 0, nT$ 也是 $f(x)$ 的周期. 可见, 一个周期函数总有无穷多个周期. 假若周期函数的周期中有最小的正周期, 则称之为**最小（正）周期**.

例 3.4.11 例如, $\sin x$ 以 2π 为其最小周期. 但是, 并非所有周期函数都有最小正周期. 比如, 狄利克雷函数 $\mathrm{D}(x)$ 以任意正的有理数为其周期, 而所有正的有理数组成的集合无最小元素.

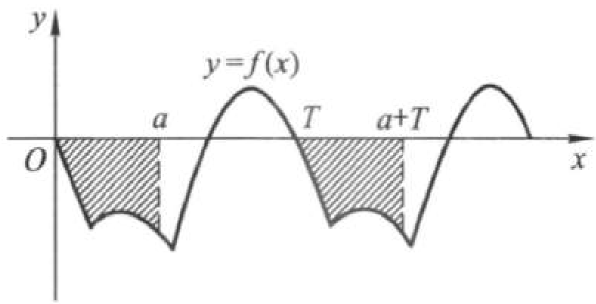


图 3.4: 周期函数的定积分

顺便指出: 周期函数 $y=f(x)$ 可以不要求在区间 $(-\infty,+\infty)$ 上有定义, 而只要其在定义域 X 内满足如下条件即可:

$$\forall x \in X \Longrightarrow x+T \in X,$$

其中 T 是 $f(x)$ 的周期. 为了方便, 今后我们总是假定周期函数在区间 $(-\infty,+\infty)$ 上有定义. 显然如果定义域 X 是某个有长度的区间, 那么上式似乎不能成立. 但看下面的例子.

例 3.4.12 定义在所有整数 $X=\mathbf{Z}$ 上的周期函数 $f(x)$, 当 x 是奇数返回 0, 偶数返回 1, 可以写成如下形式:

$$f(x)=\frac{1+(-1)^x}{2}.$$

该函数 $f(x)$ 以周期 $T=2$ 为周期, 即对于任意整数 x, 都有

$$f(x+2)=f(x).$$

且 $\forall x \in X \Longrightarrow x+T \in X$.

命题 3.4.2

设 $f(x)$ 是区间 $(-\infty,+\infty)$ 上的连续函数, 并且以 T 为周期, 则对于任意实数 a, 如下公式成立:

$$\int_0^T f(x)\mathrm{d}x=\int_a^{a+T} f(x)\mathrm{d}x.$$

这个命题告诉我们: **对于周期函数, 只要积分区间长度等于函数的周期 T, 无论从哪个位置开始积分, 其积分值都相等.** 见图 3.4. (特定积分区间长度为 T, 不限定平移长度)

证明.

$$\begin{aligned}
& \int_0^T f(x)\mathrm{d}x=\int_a^{a+T} f(x)\mathrm{d}x \\
\Leftrightarrow & \int_0^a f(x)\mathrm{d}x+\int_a^T f(x)\mathrm{d}x=\int_a^T f(x)\mathrm{d}x+\int_T^{T+a} f(x)\mathrm{d}x \\
\Leftrightarrow & \int_0^a f(x)\mathrm{d}x=\int_T^{T+a} f(x)\mathrm{d}x
\end{aligned}$$

这个等价推导展示了原命题可以转化为证明

$$\int_0^a f(x)\mathrm{d}x = \int_T^{T+a} f(x)\mathrm{d}x.$$

见图 3.4. 我们先将定积分分为两部分:

$$\int_a^{a+T} f(x)\mathrm{d}x = \int_a^T f(x)\mathrm{d}x + \int_T^{T+a} f(x)\mathrm{d}x.$$

对上式右端的第二个定积分做变量替换 $t = x - T$, 我们得到

$$\int_T^{T+a} f(x)\mathrm{d}x = \int_0^a f(t+T)\mathrm{d}t = \int_0^a f(t)\mathrm{d}t,$$

这里我们用到了函数 $f(x)$ 的周期性. 于是

$$\int_a^{a+T} f(x)\mathrm{d}x = \int_a^T f(x)\mathrm{d}x + \int_0^a f(t)\mathrm{d}t = \int_0^T f(x)\mathrm{d}x.$$

证毕. □

例 3.4.13 求定积分 $\displaystyle\int_1^{1+\pi} |\cos x|\mathrm{d}x.$

证明. $|\cos x|$ 以 π 为周期. 利用命题 3.4.2, 得到

$$\begin{aligned}\int_1^{1+\pi} |\cos x|\mathrm{d}x &= \int_0^{\pi} |\cos x|\mathrm{d}x \\ &= \int_{0-\frac{\pi}{2}}^{\pi-\frac{\pi}{2}} |\cos x|\mathrm{d}x \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos x \ \mathrm{d}x = \sin x|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = 2.\end{aligned}$$

□

命题 3.4.3

设 $f(x)$ 是区间 $(-\infty, +\infty)$ 上的连续函数, 且以 T 为周期. 对于任意实数 a, b, 如下公式成立:

$$\int_{a+T}^{b+T} f(x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x.$$

这个命题告诉我们: **对于周期函数, 将区间整体平移一个或整数个周期, 积分值都是相等的.** (特定平移长度为 T, 不限定积分区间长度)

证明. 利用命题 3.4.2,

$$\begin{aligned}\int_{a+T}^{b+T} f(x)\mathrm{d}x &= \int_{a+T}^{a} f(x)\mathrm{d}x + \int_a^b f(x)\mathrm{d}x + \int_b^{b+T} f(x)\mathrm{d}x \\ &= -\int_0^T f(x)\mathrm{d}x + \int_a^b f(x)\mathrm{d}x + \int_0^T f(x)\mathrm{d}x \\ &= \int_a^b f(x)\mathrm{d}x.\end{aligned}$$

□

例 3.4.14 求定积分 $\int_{\frac{\pi}{2}}^{\pi} |\sin 2x| \mathrm{d}x$.

证明. $|\sin 2x|$ 以 $\frac{\pi}{2}$ 为周期, 于是

$$\begin{aligned}\int_{\frac{\pi}{2}}^{\pi} |\sin 2x| \mathrm{d}x &= \int_{0+\frac{\pi}{2}}^{\frac{\pi}{2}+\frac{\pi}{2}} |\sin 2x| \mathrm{d}x = \int_{0}^{\frac{\pi}{2}} \sin 2x \ \mathrm{d}x \\ &= -\frac{1}{2}\cos 2x\Big|_0^{\frac{\pi}{2}} = 1.\end{aligned}$$

本题的周期和积分区间长度恰好相等, 所以运用命题 3.4.2 或命题 3.4.3 皆可. □

例 3.4.15 求定积分 $\int_0^n (x-[x])\mathrm{d}x$, 其中 n 为正整数, 而 $[x]$ 为不超过 x 的最大整数.

证明. 显然, 被积函数 $x-[x]$ 是以 1 为周期的函数. 根据命题 3.4.2, 我们有

$$\int_0^n (x-[x])\mathrm{d}x = n\int_0^1 (x-[x])\mathrm{d}x.$$

注意到当 $0 \leqslant x < 1$ 时, $[x]=0$, 于是我们得到

$$\int_0^n (x-[x])\mathrm{d}x = n\int_0^1 x \ \mathrm{d}x = \frac{n}{2}.$$

□

问题 关于周期函数的有趣问题

1. 周期函数的导函数也是周期函数吗, 周期是否一致? Is the derivative of a periodic function always periodic? - Mathematics Stack Exchange
2. 周期函数的原函数也是周期函数吗, 周期是否一致? Integrals of Periodic Functions
3. 函数 $f(x)=\sin(x)+\sin(\pi x)$ 的图像长得非常像周期函数, 前者周期为 2π, 后者为 2, 为什么它不是周期函数? (所以任意两个周期函数的和是周期函数吗?) real analysis - Why $\sin(x)+\sin(\pi x)$ is not periodic? - Mathematics Stack Exchange

3.4.4 其他公式

之前我们重点讨论了特殊函数 (奇偶周期) 的定积分. 定积分的计算中还有许多有用的通用计算公式. 它们往往对于被积函数和积分区间的限定要求较少. 搜索和积累这类公式能帮助解决一类型题目. 这里抛砖引玉, 讲几个. 读者要自己认识到这类公式的价值, 自行收集.

例 3.4.16 **任意积分区间都可以单位化** 设函数 $f(x)$ 在区间 $[a,b]$ 上连续, 证明:

$$\int_a^b f(x) \ \mathrm{d}x = (b-a)\int_0^1 f(a+(b-a)x) \ \mathrm{d}x.$$

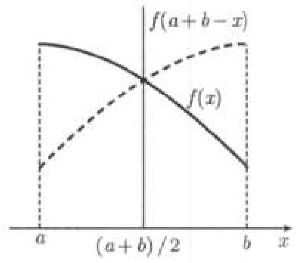


图 3.5: 翻折不不改变定积分的值

这个公式的本质是将积分区间 $[a,b]$ 通过线性变换映射到 $[0,1]$ 上, 同时保持积分值不变. 这个变换方法在定积分计算中非常有用, 因为**它可以将任意区间的积分转化为单位区间 $[0,1]$ 上的积分.**

证明. 让我们先做变量替换, 令 $x = a + (b-a)t$, 当 $t = 0$ 时, $x = a$; 当 $t = 1$ 时, $x = b$. 求导可得: $\mathrm{d}x = (b-a)\ \mathrm{d}t$. 将这个换元代入原积分:

$$\int_a^b f(x)\ \mathrm{d}x = \int_0^1 f(a+(b-a)t)(b-a)\ \mathrm{d}t = (b-a)\int_0^1 f(a+(b-a)t)\ \mathrm{d}t$$

这就证明了原式成立. □

例 3.4.17 **翻转公式** 设 $f(x)$ 在 $[a,b]$ 上连续, 求证

$$\int_a^b f(x)\mathrm{d}x = \int_a^b f(a+b-x)\mathrm{d}x.$$

这个等式的几何意义是将积分区间以区间中点为对称中心翻转, 面积不变 (图 3.5). 事实上, 若记 $g(x) = f(a+b-x)$, 则有

$$g\left(\frac{a+b}{2}-t\right) = f\left(a+b-\left(\frac{a+b}{2}-t\right)\right) = f\left(\frac{a+b}{2}+t\right),\quad 0\leqslant t\leqslant\frac{b-a}{2}.$$

鉴于这个几何意义, 姑且称此公式为**翻转公式**. 当 $a=-b$ 时, 翻转公式简化成:

$$\int_{-b}^b f(x)\mathrm{d}x = \int_{-b}^b f(-x)\mathrm{d}x.$$

证明. 令 $u = a+b-x$, 则 $\mathrm{d}x = -\mathrm{d}u$. 当 $x=a$ 时, $u = a+b-a = b$; 当 $x = b$ 时, $u = a+b-b = a$. 因此, 积分区间 $[a,b]$ 在变换后变为 $[b,a]$. 于是:

$$\int_a^b f(a+b-x)\,\mathrm{d}x = \int_b^a f(u)\cdot(-\mathrm{d}u) = \int_a^b f(u)\,\mathrm{d}u = \int_a^b f(x)\,\mathrm{d}x.$$

这就证明了原式成立. □

例 3.4.18 利用定积分的换元法证明:

$$\int_0^\pi xf(\sin x)\mathrm{d}x = \pi\int_0^{\frac{\pi}{2}} f(\sin x)\mathrm{d}x.$$

证明. 对于上式左侧的积分, 可以将积分区间拆分为 $[0, \frac{\pi}{2}]$ 和 $[\frac{\pi}{2}, \pi]$, 得到:

$$\int_0^{\pi} xf(\sin x)\,\mathrm{d}x = \int_0^{\frac{\pi}{2}} xf(\sin x)\,\mathrm{d}x + \int_{\frac{\pi}{2}}^{\pi} xf(\sin x)\,\mathrm{d}x.$$

对第二个积分做变量替换: 在区间 $[\frac{\pi}{2}, \pi]$ 中, 取变量替换 $u = \pi - x$, 那么 $\mathrm{d}x = -\mathrm{d}u$. 当 $x = \frac{\pi}{2}$ 时, $u = \frac{\pi}{2}$; 当 $x = \pi$ 时, $u = 0$. 因此, 上式右侧第二个积分等于

$$\begin{aligned}\int_{\frac{\pi}{2}}^{\pi} xf(\sin x)\,\mathrm{d}x &= \int_{\frac{\pi}{2}}^{0} (\pi - u)f(\sin u)\,(-\mathrm{d}u) \\ &= \int_0^{\frac{\pi}{2}} (\pi - u)f(\sin u)\,\mathrm{d}u \\ &= \pi\int_0^{\frac{\pi}{2}} f(\sin u)\,\mathrm{d}u - \int_0^{\frac{\pi}{2}} uf(\sin u)\,\mathrm{d}u.\end{aligned}$$

将第二个积分的结果代回一开始的拆分的两项之和中. 其中, $\int_0^{\frac{\pi}{2}} xf(\sin x)\,\mathrm{d}x$ 会被抵消, 因此得到

$$\int_0^{\pi} xf(\sin x)\,\mathrm{d}x = \pi\int_0^{\frac{\pi}{2}} f(\sin x)\,\mathrm{d}x.$$

证毕. □

例 3.4.19 利用上题的结果, 求定积分 $\int_0^{\pi} \frac{x}{2+\sin x}\,\mathrm{d}x$.

证明. 参考视频 粉丝间的定积分计算，这不算难题吧？_ 哔哩哔哩 _bilibili □

3.5 §5 定积分的若干应用

除了我们所熟知的可利用定积分来计算曲边梯形的面积, 变力所做的功之外, 定积分还有许多其他应用.

> **注意** 本节的内容不要严格去推敲, 只要用直觉上认识并感受即可. 本节的证明不严谨, 经常感性地使用近似 $\approx$ 符号. 本节的重点还是计算.

3.5.1 曲线的弧长和弧微分

(定积分的灵魂) 分割 → 近似代替 → 求和 → 取极限

参数方程所定曲线的弧长

设平面上给定一条曲线弧 $\widehat{AB}$, 它的参数方程为

$$\begin{cases} x = x(t), \\ y = y(t), \end{cases} \quad \alpha \le t \le \beta,$$

并且 $t = \alpha$ 对应于端点 A, $t = \beta$ 对应于端点 B (见图 3.6). 我们假定函数 $x(t)$ 及 $y(t)$ 有连续的导数 $x'(t)$ 及 $y'(t)$. 这时, 曲线弧 $\widehat{AB}$ 上的切线可连续变动. 通常把满足这种条件的曲线 (弧) 称为**光滑 (smooth) 曲线**. 光滑曲线总是可以计算弧长的.

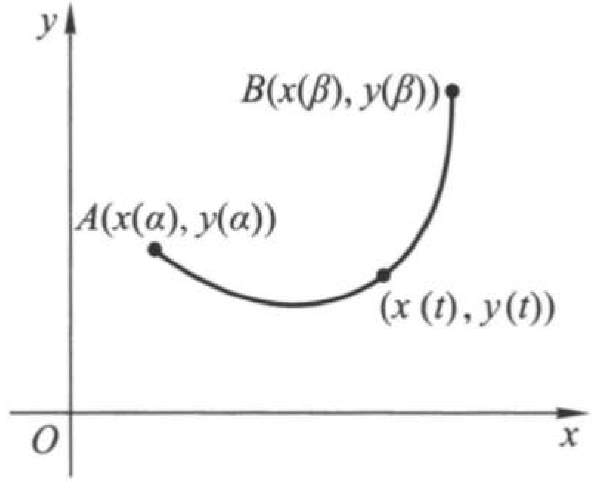


图 3.6: 曲线的弧长

1. **分割:** 我们先将区间 $[\alpha,\beta]$ (可以看作是一小段直线) 分成 n 份, 插入 $n+1$ 个分割点:

$$\alpha=t_0<t_1<\cdots<t_n=\beta.$$

那么在曲线弧 $\widehat{AB}$ 上就可以找到对应的 $n+1$ 个分割点:

$$M_i=\left(x\left(t_i\right), y\left(t_i\right)\right), \quad i=0,1,\cdots,n.$$

注意, 曲线弧 $\widehat{AB}$ 的端点 A 就是 $M_0=(x(\alpha), y(\alpha))$, 端点 B 就是 $M_n=(x(\beta), y(\beta))$, 见图 3.6. 它们把整个曲线弧 $\widehat{AB}$ 也划分成 n 份小曲线弧 $\widehat{M_{i-1}M_i}$ $(i=1,\cdots,n.)$. 至此, 我们利用对于直线线段 $[\alpha,\beta]$ 的分割来实现了对曲线线段 $\widehat{AB}$ 的分割. 显然, 对 $[\alpha,\beta]$ 的分割很细致时, 对于曲线弧 $\widehat{AB}$ 的分割也会很细致.

2. **近似:** 记 Δs_i $(i=1,2,\cdots,n)$ 为小曲线弧 $\widehat{M_{i-1}M_i}$ 的真实长度. 当分割点充分密集且分割的细致程度足够高时, 真实小曲线长度 Δs_i 近似于弦 $M_{i-1}M_i$ 的长度, (截弯取直) 用勾股定理则有

$$\Delta s_i \approx \sqrt{\left(x\left(t_i\right)-x\left(t_{i-1}\right)\right)^2+\left(y\left(t_i\right)-y\left(t_{i-1}\right)\right)^2}.$$

而当分割的细致程度足够高时, t_{i-1} 会与 t_i 充分接近, 此时利用微分的近似计算[7], 有

$$x\left(t_i\right)-x\left(t_{i-1}\right) \approx x'\left(t_{i-1}\right)\left(t_i-t_{i-1}\right),$$
$$y\left(t_i\right)-y\left(t_{i-1}\right) \approx y'\left(t_{i-1}\right)\left(t_i-t_{i-1}\right),$$

把这两个式子代入 Δs_i 近似值的右侧, 有

$$\Delta s_i \approx \sqrt{\left(x'\left(t_{i-1}\right)\right)^2+\left(y'\left(t_{i-1}\right)\right)^2}\Delta t_i,$$

其中 $\Delta t_i:=t_i-t_{i-1}>0$.

3. **求和:** 这样, 可得整个曲线弧 $\widehat{AB}$ 的真实长度 s 的近似值:

$$s=\sum_{i=1}^{n}\Delta s_i \approx \sum_{i=1}^{n}\sqrt{\left(x'\left(t_{i-1}\right)\right)^2+\left(y'\left(t_{i-1}\right)\right)^2}\Delta t_i.$$

[7] 回顾第 2 章的第 5 节, 我们有 $f\left(x_0+\Delta x\right)=f\left(x_0\right)+f'\left(x_0\right)\Delta x+o(\Delta x)$. 故当 Δx 足够小时, 有有 $f\left(x_0+\Delta x\right)\approx f\left(x_0\right)+f'\left(x_0\right)\Delta x$.

4. **取极限:** 令所有小区间长度的最大者 $\lambda = \max\{\Delta t_i \mid i = 1, 2, \cdots, n\}$, 那么当 $\lambda \to 0$ 时, 曲线弧 $\widehat{AB}$ 上的分割点所形成弦 $M_{i-1}M_i$ $(i = 1, 2, \cdots, n)$ 的长度的最大值也趋向于零. 我们有理由将曲线弧 $\widehat{AB}$ 的长度 s 看作下列极限:

$$s = \lim_{\lambda \to 0} \sum_{i=1}^{n} \sqrt{\left(x'\left(t_{i-1}\right)\right)^2 + \left(y'\left(t_{i-1}\right)\right)^2} \Delta t_i,$$

即

$$s = \int_{\alpha}^{\beta} \sqrt{\left(x'(t)\right)^2 + \left(y'(t)\right)^2} \, \mathrm{d}t.$$

设从点 A 至点 $(x(t), y(t))$ 的弧长为 t 的函数:

$$s(t) = \int_{\alpha}^{t} \sqrt{\left(x'(t)\right)^2 + \left(y'(t)\right)^2} \, \mathrm{d}t.$$

$s(t)$ 对 t 的导数为

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \sqrt{\left(x'(t)\right)^2 + \left(y'(t)\right)^2},$$

所以 $s(t)$ 微分 (称为**弧微分**) 为

$$\mathrm{d}s = \sqrt{\left(x'(t)\right)^2 + \left(y'(t)\right)^2} \, \mathrm{d}t.$$

这就是所谓的参数方程下的弧微分公式.

直角坐标系下函数曲线的弧长

当曲线弧 $\widehat{AB}$ 不是由参数方程给出, 而是由直角坐标系下的函数 $y = f(x)$ $(a \leq x \leq b)$ 给出时, x 本身可以替代上述的 t 作为参数, 而写成如下的参数方程形式:

$$\begin{cases} x = t, \\ y = f(t), \end{cases} \quad a \leq t \leq b.$$

这样, 其弧长公式立刻得到为

$$s = \int_a^b \sqrt{1 + \left(f'(t)\right)^2} \, \mathrm{d}t = \int_a^b \sqrt{1 + \left(f'(x)\right)^2} \, \mathrm{d}x = \int_a^b \, \mathrm{d}s,$$

其中, 直角坐标系下函数曲线的的弧微分就是

$$\mathrm{d}s = \sqrt{1 + \left(f'(x)\right)^2} \, \mathrm{d}x.$$

极坐标系下函数曲线的弧长

先复习一些基础知识. 极坐标 (r, θ) 和直角坐标 (x, y) 的转换公式如下.

1. 从极坐标到直角坐标: 给定极坐标 (r, θ), 对应的直角坐标 (x, y) 可以表示为

$$x = r\cos\theta, \quad y = r\sin\theta.$$

2. 从直角坐标到极坐标: 给定直角坐标 (x, y), 对应的极坐标 (r, θ) 可以表示为

$$r = \sqrt{x^2 + y^2}, \quad \theta = \arctan\frac{y}{x}.$$

但注意, θ 的最终结果还需要根据 x 和 y 的符号来决定.

当曲线弧 $\widehat{AB}$ 由 (极坐标系下函数曲线) 极坐标方程

$$r = r(\theta) \quad (\alpha \le \theta \le \beta)$$

给出时, 可将 θ 作为参数而得到转化后的直角坐标系下的参数方程:

$$\begin{cases} x(\theta) = r(\theta)\cos\theta, \\ y(\theta) = r(\theta)\sin\theta, \end{cases} \quad \alpha \le \theta \le \beta.$$

这时对两边关于 θ 求导,

$$\begin{cases} x'(\theta) = r'(\theta)\cos\theta - r(\theta)\sin\theta, \\ y'(\theta) = r'(\theta)\sin\theta + r(\theta)\cos\theta, \end{cases}$$

由此可得

$$(x'(\theta))^2 + (y'(\theta))^2 = r^2(\theta) + (r'(\theta))^2.$$

于是弧长公式是

$$s = \int_a^\beta \sqrt{r^2(\theta) + (r'(\theta))^2}\,\mathrm{d}\theta = \int_a^b \,\mathrm{d}s,$$

其中, 极坐标系下函数曲线的的弧微分就是

$$\mathrm{d}s = \sqrt{r^2(\theta) + (r'(\theta))^2}\,\mathrm{d}\theta.$$

结论 3.5.1: 曲线的弧长和弧微分 (仅计算用结论, 经不起细究)

1. **参数方程所定曲线的弧长:** 已知参数方程 $\begin{cases} x = x(t) \\ y = y(t) \end{cases}$, 若 $x'(t), y'(t)$ 在区间 $[\alpha, \beta]$ 上存在且连续, 那么该参数方程在 $[\alpha, \beta]$ 上的曲线的长度为:

$$s = \int_\alpha^\beta \sqrt{(x'(t))^2 + (y'(t))^2}\,\mathrm{d}t = \int_\alpha^\beta \,\mathrm{d}s,$$

其中 $\mathrm{d}s = \sqrt{(x'(t))^2 + (y'(t))^2}\,\mathrm{d}t$ 就是参数方程下的弧微分.

2. **直角坐标系下函数曲线的弧长:** 若直角坐标系下函数 $y = f(x)$ 在区间 $[a, b]$ 上是光滑的[a], 那么该函数在 $[a, b]$ 上的曲线的长度为:

$$s = \int_a^b \sqrt{1 + (f'(x))^2}\,\mathrm{d}x = \int_a^b \,\mathrm{d}s,$$

其中 $\mathrm{d}s = \sqrt{1 + (f'(x))^2}\,\mathrm{d}x$ 就是直角坐标系下函数曲线的的弧微分.

3. **极坐标系下函数曲线的弧长:** 若极坐标系下函数 $r = r(\theta)$ 在区间 $[\alpha, \beta]$ 上是光滑的, 那么该函数在 $[\alpha, \beta]$ 上的曲线的长度为:

$$s = \int_a^\beta \sqrt{r^2(\theta) + (r'(\theta))^2}\,\mathrm{d}\theta = \int_a^b \,\mathrm{d}s,$$

其中 $\mathrm{d}s = \sqrt{r^2(\theta) + (r'(\theta))^2}\,\mathrm{d}\theta$ 就是极坐标系下函数曲线的的弧微分.

[a]函数的光滑可以认为其存在任意阶的导函数.

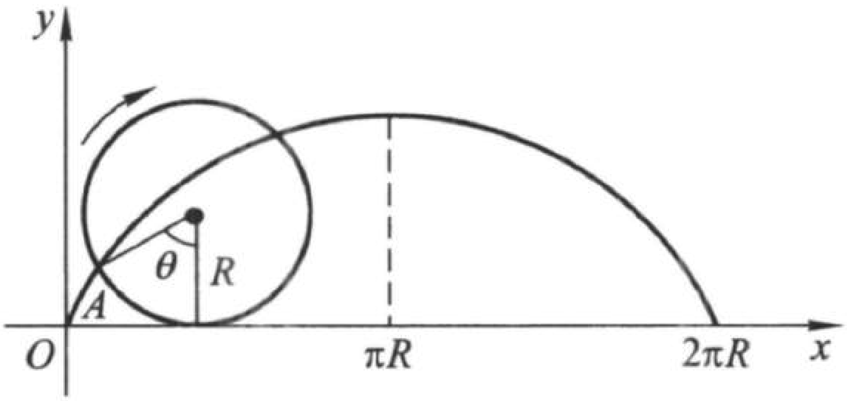


图 3.7: 旋轮线

例 3.5.1 设想在 Oxy 平面上有一个半径为 R 的圆盘, 在其边缘上固定一点 A. 设开始时点 A 与原点重合, 且圆盘与 x 轴相切于该点, 当圆盘沿 x 轴滚动时, 点 A 的轨迹称为**旋轮线, 也叫摆线** (见图 3.7). 求这条旋轮线第一拱 (自开始至点 A 再次遇到 x 轴) 的弧长. 另外动画参考 Roulettes –GeoGebra, Cycloid Demonstration –GeoGebra.

> **注释 3.5.1** 旋轮线指出了固定高度下小球自由下降的最快速的轨迹, 有兴趣的同学可以自行了解. 这是一个非常有趣的问题.
> 最速降线问题：最快下滑路径为什么是旋轮线?（中英字幕）_ 哔哩哔哩 _bilibili
> “旋轮线”是什么? 将小球从等高轨道推下，有趣的现象发生了 _ 哔哩哔哩 _bilibili
> 最速降线问题——光学分析法秒了! _ 哔哩哔哩 _bilibili

证明. 很容易看出这条旋轮线的方程为

$$\begin{cases} x = R(\theta - \sin\theta) \\ y = R(1 - \cos\theta) \end{cases} \quad (\theta \geq 0),$$

其中参数 θ 的几何意义如图 3.7 所示, 第一拱所对应的参数范围应为 $0 \leq \theta \leq 2\pi$, 因此所求的弧长应为

$$\begin{aligned} s &= \int_0^{2\pi} \sqrt{R^2(1-\cos\theta)^2 + R^2\sin^2\theta}\,\mathrm{d}\theta \\ &= R\int_0^{2\pi} \sqrt{2(1-\cos\theta)}\mathrm{d}\theta \\ &= 2R\int_0^{2\pi} \left|\sin\frac{\theta}{2}\right| \mathrm{d}\theta. \end{aligned}$$

注意到当 θ 自 0 变到 2π 时, $\sin\dfrac{\theta}{2} \geq 0$, 故

$$s = 2R\int_0^{2\pi} \sin\frac{\theta}{2}\,\mathrm{d}\theta = 4R\int_0^{\pi} \sin t\,\mathrm{d}t = 8R.$$

□

例 3.5.2 求椭圆周

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \quad (a \geq b>0)$$

的周长. (椭圆周长公式理论上存在的不过它不能用初等函数表示，它是一个与离心率有关的无穷收敛级数.)

证明. 该椭圆周落在第一, 第二象限的部分可用如下函数表示:

$$y=\frac{b}{a} \sqrt{a^2-x^2} \quad (-a \leq x \leq a).$$

这时弧微分应为

$$\mathrm{d} s=\sqrt{1+\left(y^{\prime}\right)^2} \mathrm{~d} x=\sqrt{1+\frac{\frac{b^2}{a^2} x^2}{a^2-x^2}} \mathrm{~d} x=\sqrt{\frac{a^2+\left(\frac{b^2}{a^2}-1\right) x^2}{a^2-x^2}} \mathrm{~d} x,$$

故该椭圆周的周长应为

$$s=2 \int_{-a}^a \sqrt{\frac{a^2+c x^2}{a^2-x^2}} \mathrm{~d} x \quad\left(c=\frac{b^2}{a^2}-1\right).$$

- 当 $a=b$ 时, 该椭圆周退化为圆周, 这时 $c=0$, $a=b$ 是圆的半径, 并有

$$s=2 \int_{-a}^a \frac{a \mathrm{~d} x}{\sqrt{a^2-x^2}}=2 a \int_{-a}^a \frac{\mathrm{d}\left(\frac{x}{a}\right)}{\sqrt{1-\left(\frac{x}{a}\right)^2}}=2 a \arcsin \left.\frac{x}{a}\right|_{-a} ^a=2 \pi a .$$

  这就是通常的圆周长公式.

- 但是, 当 $a \neq b$ 时, 事情就复杂了: 这时, $c \neq 0$, 不定积分

$$\int \sqrt{\frac{a^2+c x^2}{a^2-x^2}} \mathrm{~d} x$$

  的原函数不是初等函数, 这是已经证明了的事实. 这个不定积分称为**椭圆积分**. 无论是在应用上还是在数学基础理论研究上, 椭圆积分都有重要价值.

□

> **注释 3.5.2** 椭圆的周长是很难计算的! 目前还没有找到椭圆周长的一般公式!
> 椭圆的周长与面积 - 科学空间 |Scientific Spaces
> 你可以计算出椭圆的面积，却永远计算不出椭圆的周长 - 哔哩哔哩

3.5.2 某些立体的体积

由函数曲线所定的旋转体的体积

(定积分的灵魂) 分割 → 近似代替 → 求和 → 取极限

设 $y=f(x)$ 是区间 $[a, b]$ 上的一个非负连续函数, 那么它的图形与直线 $x=a, x=b$ 及 x 轴围成一个曲边梯形. 设想将这个曲边梯形绕 x 轴旋转一周, 则得到一个旋转体. 日常见到的许多瓷器及玻璃制品都是旋转体. 我们的问题是: 如何计算这个旋转体的体积与侧面积?

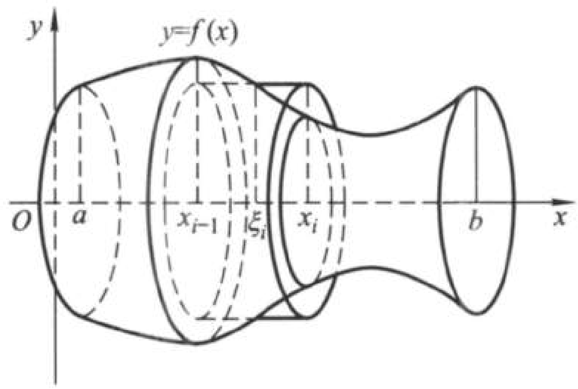


图 3.9

图 3.8: 旋转体的体积

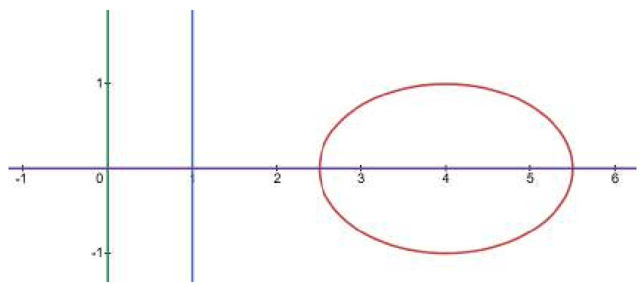


图 3.9: 椭圆周 $4(x-4)^2+9y^2=9$.

我们先讨论如何求这个旋转体的体积. 将区间 $[a,b]$ 用分割点

$$a=x_0<x_1<x_2<\cdots<x_{n-1}<x_n=b$$

分成 n 个小区间 $[x_{i-1},x_i]$, $i=1,2,\cdots,n$. 过各分割点作垂直于 x 轴的平面, 这时就将该旋转体分作 n 块薄片, 令 ΔV_i $(i=1,2,\cdots,n)$ 表示第 i 块薄片的真实体积 (见图 3.8). 在每个小区间 $[x_{i-1},x_i]$ 中取一点 ξ_i, 我们用高度为 Δx_i, 底面半径为 $f(\xi_i)$ 的小圆柱体的体积来作为 ΔV_i 的近似值, 即

$$\Delta V_i\approx\pi\left(f\left(\xi_i\right)\right)^2\Delta x_i.$$

那么所求的体积的近似值为

$$V=\sum_{i=1}^{n}\Delta V_i\approx\pi\sum_{i=1}^{n}\left(f\left(\xi_i\right)\right)^2\Delta x_i.$$

与前面一样, 我们令 $\lambda=\max\{\Delta x_i\mid i=1,2,\cdots,n\}$. 当 $\lambda\to 0$ 时, 我们认为上述作为近似值的和式的极限就是所求的体积, 于是

$$V=\pi\int_a^b(f(x))^2\mathrm{d}x.$$

例 3.5.3 设区域 D 由椭圆周 $4(x-4)^2+9y^2=9$ 所围成, 求 D 绕下列直线旋转一周所形成旋转体的体积 (见图 3.9):

1. x 轴;

2. y 轴;

3. 直线 $x=1$.

证明. 1. 这时应首先根据椭圆周的方程将 y 解为 x 的函数 (只取 x 轴上方对应部分):

$$y=\sqrt{1-\frac{4}{9}(x-4)^2}\quad\left(\frac{5}{2}\leq x\leq\frac{11}{2}\right).$$

所以, 所求的旋转体体积为

$$V_1=\pi\int_{\frac{5}{2}}^{\frac{11}{2}}\left[1-\frac{4}{9}(x-4)^2\right]\mathrm{d}x=\pi\left[x-\frac{4}{27}(x-4)^3\right]\Bigg|_{\frac{5}{2}}^{\frac{11}{2}}=2\pi.$$

2. (目前为止, 我们只讨论了由函数 $y=f(x)$ 在区间 $[a,b]$ 上确定的曲边梯形绕 x 轴旋转一周所得到的旋转体的体积. 类似地, 如果考虑绕 y 轴的旋转, 只需将 y 轴视作自变量即可.) 这时 D 旋转所形成的旋转体是一个**椭圆环体**, 其体积 V_2 可视为右半椭圆周

$$x=4+\frac{3}{2}\sqrt{1-y^2}\quad(-1\leq y\leq 1)$$

与左半椭圆周

$$x=4-\frac{3}{2}\sqrt{1-y^2}\quad(-1\leq y\leq 1)$$

分别绕 y 轴旋转所形成的旋转体体积之差. 注意, 这一我们. 因此, 我们有

$$\begin{aligned}V_2&=\pi\int_{-1}^{1}\left[\left(4+\frac{3}{2}\sqrt{1-y^2}\right)^2-\left(4-\frac{3}{2}\sqrt{1-y^2}\right)^2\right]\mathrm{d}y\\&=24\pi\int_{-1}^{1}\sqrt{1-y^2}\,\mathrm{d}y=12\pi^2.\end{aligned}$$

3. (目前, 我们只讨论了绕 x 轴和 y 轴这类直线的旋转体的体积, 尚未涉及绕任意平面直线的情况. 然而, 对于旋转轴为平行于 x 轴或 y 轴的直线的情况, 我们只需对原函数进行平移即可进行处理.) 这时情形与 (2) 类似, 所求的旋转体体积为

$$\begin{aligned}V_3&=\pi\int_{-1}^{1}\left[\left(4+\frac{3}{2}\sqrt{1-y^2}-1\right)^2-\left(4-\frac{3}{2}\sqrt{1-y^2}-1\right)^2\right]\mathrm{d}y\\&=18\pi\int_{-1}^{1}\sqrt{1-y^2}\,\mathrm{d}y=9\pi^2.\end{aligned}$$

□

例 3.5.4 证明椭圆 $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1$ (其中 $0<b\leq a$) 绕长轴 (x 轴) 和短轴 (y 轴) 旋转形成的椭球的体积分别为 $\dfrac{4}{3}\pi ab^2$ 和 $\dfrac{4}{3}\pi a^2b$. (利用这一一般性结果可以验证上一题第一问的结果.)

证明. 我们将分别计算绕长轴和短轴旋转形成的椭球的体积.

1. **绕长轴旋转的体积.** 首先, 考虑椭圆方程 $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1$. 我们可以解出 y 的表达式:

$$y=b\sqrt{1-\frac{x^2}{a^2}}.$$

当 x 从 $-a$ 到 a 时, 利用**圆盘法** (也就是本小节开头的内容) 计算绕 x 轴旋转形成的椭球体积:

$$V_{\text{长轴}}=\int_{-a}^{a}\pi y^2\,dx=\int_{-a}^{a}\pi\left(b\sqrt{1-\frac{x^2}{a^2}}\right)^2dx=\pi b^2\int_{-a}^{a}\left(1-\frac{x^2}{a^2}\right)dx$$

计算这两个部分的积分:

1. $\displaystyle\int_{-a}^{a}1\,dx=2a.$

2. $\displaystyle\int_{-a}^{a}\frac{x^2}{a^2}\,dx=\frac{1}{a^2}\cdot\left.\frac{x^3}{3}\right|_{-a}^{a}=\frac{1}{a^2}\cdot\left(\frac{a^3}{3}-\left(-\frac{a^3}{3}\right)\right)=\frac{2a^3}{3a^2}=\frac{2a}{3}.$

因此, 得到

$$V_{\text{长轴}}=\pi b^2\left(2a-\frac{2a}{3}\right)=\pi b^2\left(\frac{6a}{3}-\frac{2a}{3}\right)=\frac{4}{3}\pi ab^2.$$

2. **绕短轴旋转的体积.** 可以解出 x 的表达式:

$$x=a\sqrt{1-\frac{y^2}{b^2}}.$$

当 y 从 $-b$ 到 b 时, 利用圆盘法计算绕 y 轴旋转形成的椭球的体积:

$$V_{\text{短轴}}=\int_{-b}^{b}\pi x^2\,dy=\int_{-b}^{b}\pi\left(a\sqrt{1-\frac{y^2}{b^2}}\right)^2dy=\pi a^2\int_{-b}^{b}\left(1-\frac{y^2}{b^2}\right)dy.$$

计算这两个部分的积分:

1. $\displaystyle\int_{-b}^{b}1\,dy=2b.$

2. $\displaystyle\int_{-b}^{b}\frac{y^2}{b^2}\,dy=\frac{1}{b^2}\cdot\left[\frac{y^3}{3}\right]_{-b}^{b}=\frac{1}{b^2}\cdot\left(\frac{b^3}{3}-\left(-\frac{b^3}{3}\right)\right)=\frac{2b^3}{3b^2}=\frac{2b}{3}.$

将结果代入, 我们得到

$$V_{\text{短轴}}=\pi a^2\left(2b-\frac{2b}{3}\right)=\pi a^2\left(\frac{6b}{3}-\frac{2b}{3}\right)==\frac{4}{3}\pi a^2b.$$

□

截面面积已知的立体的体积

如果一个立体不是旋转体, 但是它位于过点 $x=a$, $x=b$ 且垂直于 x 轴的两个平面之间, 并且已知过点 x $(a\le x\le b)$ 且垂直于 x 轴的截面面积为 $A(x)$, 那么使用类似于前面求旋转体体积的方法, 可以求得这个截面面积已知的立体的体积 V. 分割区间 $[a,b]$:

$$a<x_0<x_1<\cdots<x_{n-1}<x_n=b.$$

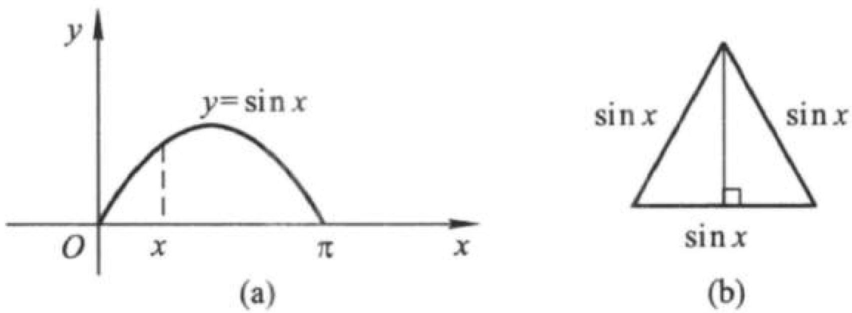


图 3.10: Enter Caption

过各分割点作垂直于 x 轴的截面, 将此立体分作 n 块薄片, ΔV_i $(i=1,2,\cdots,n)$ 表示第 i 块的体积. 在每个小区间 $[x_{i-1},x_i]$ 中任取一点 ξ_i, 则 ΔV_i 近似等于底面面积为 $A(\xi_i)$, 高为 Δx_i 的小柱体体积, 即 $\Delta V_i\approx A(\xi_i)\Delta x_i$, 从而

$$V=\sum_{i=1}^{n}\Delta V_i\approx\sum_{i=1}^{n}A(\xi_i)\Delta x_i.$$

这样, 所求的立体体积为

$$V=\int_a^b A(x)\mathrm{d}x.$$

例 3.5.5 空间一个物体 Ω 放置在 Oxy 平面上, 其底座所占的区域是曲线 $y=\sin x$ $(0\le x\le\pi)$ 和 x 轴围成的图形 [图 3.10 (a)], 垂直于 x 轴的横截面都是等边三角形 [图 3.10 (b)], 求 Ω 的体积.

证明. 对于 $x\in[0,\pi]$, 过点 x 且垂直于 x 轴的截面是边长为 $\sin x$ 的等边三角形, 其面积为

$$A(x)=\frac{1}{2}\cdot\sin x\cdot\frac{\sqrt{3}}{2}\sin x=\frac{\sqrt{3}}{4}\sin^2 x.$$

则 Ω 的体积为

$$V=\int_0^\pi\frac{\sqrt{3}}{4}\sin^2 x\ \mathrm{d}x=\frac{\sqrt{3}}{8}\int_0^\pi(1-\cos 2x)\mathrm{d}x=\frac{\sqrt{3}}{8}\pi.$$

□

以下是本小节的总结.

结论 3.5.2: 某些立体的体积 (仅计算用结论, 经不起细究)

	旋转体	截面积已知
已知条件	连续曲线 $f(x)$ 绕 x 轴旋转	截面积的函数为 $A(x)$
微元	$\pi[f(x)]^2\mathrm{d}x$	$A(x)\mathrm{d}x$
体积	$\int_a^b\pi[f(x)]^2\mathrm{d}x$	$\int_a^b A(x)\mathrm{d}x$

可以发现, 前者是后者的一种特殊情况.

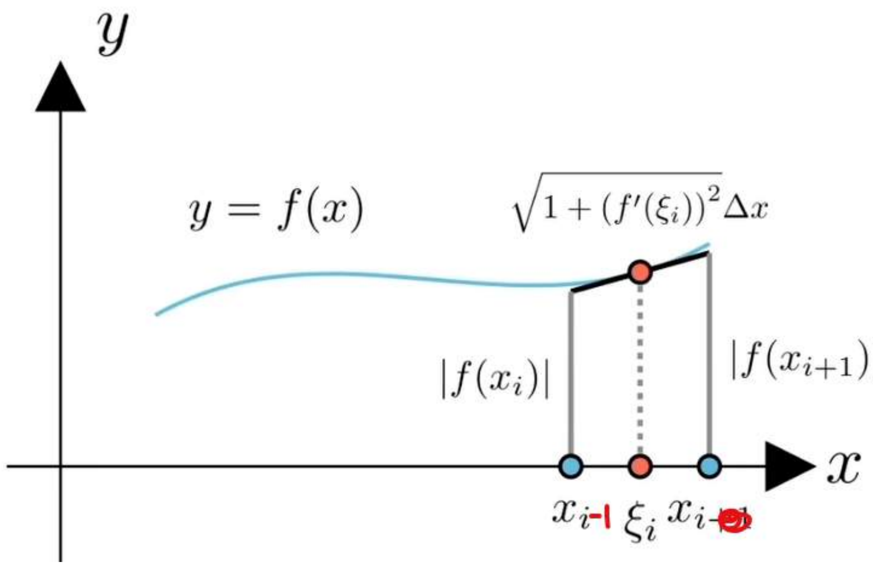


图 3.11: 旋转体的侧面积

3.5.3 旋转体的侧面积

(定积分的灵魂) 分割 $\to$ 近似代替 $\to$ 求和 $\to$ 取极限

> **注意** 课本上本节内容讲得不清楚. 我已经尽力修改了. 还请大家重点参考视频: 如何求旋转体表面积 (旋转体的侧面积) ｜马同学图解微积分 _ 哔哩哔哩 _bilibili

现在我们讨论旋转体侧面积的求法. 我们依然假定旋转体是由曲线弧

$$\Gamma : y = f(x) \quad (a \le x \le b,\ f(x) > 0)$$

与直线 $x = a$, $x = b$ 及 x 轴所围成曲边梯形绕 x 轴旋转一周而形成的. 为了计算这个旋转体的侧面积 (不包括上下两个面, 侧面积只和曲边相关), 我们需要假定函数 $y = f(x)$ 有连续导数 $f'(x)$. 首先, 对区间 $[a, b]$ 进行分割:

$$T : x_0 = a < x_1 < \cdots < x_{n-1} < x_n = b,$$

并考虑其中一个小区间 $[x_{i-1}, x_i]$ 及其所对应的小曲线弧:

$$\Gamma_i : y = f(x) \quad (x_{i-1} \le x \le x_i)\,.$$

我们记 F 为该旋转体的真实侧面积, 而 ΔF_i 为小曲线弧 Γ_i 相应的部分真实侧面积. 那么, 显然 $F = \sum\limits_{i=1}^{n} \Delta F_i$.

我们考虑用某种方式去近似 ΔF_i 的值, 这里我们使用如图 3.11 所示的圆台方案. 我们回顾一下圆台的面积. 设一个标准的圆台的上顶面半径是 r_1, 下顶面半径是 r_2, 圆台的侧边的长度是 l. 则该圆台的侧面积是[8]

$$\pi\,(r_1 + r_2)\,l.$$

[8]参考 圆台侧面积和体积计算公式推导 - 冰雪-悦灵 - 博客园

此公式可以利用两个圆锥的侧面积相减证明得到. 回到上面, 如图 3.11, 我们考虑用圆台的侧面积来近似小曲线弧 Γ_i 相应的部分真实侧面积. 记圆台的侧面积是 A_i, 则有

$$\begin{aligned}\Delta F_i \approx A_i &= \pi\left(r_1+r_2\right) l \\ &\approx \pi\left(f\left(x_{i-1}\right)+f\left(x_i\right)\right) \sqrt{1+\left(f'\left(\xi_i\right)\right)^2} \Delta x \\ &\approx 2 \pi f\left(\xi_i\right) \sqrt{1+\left(f'\left(\xi_i\right)\right)^2} \Delta x \quad(n \rightarrow \infty).\end{aligned}$$

因此,

$$F=\sum_{i=1}^n \Delta F_i \approx \lim_{n \rightarrow \infty} \sum_{i=1}^n A_i=\int_a^b 2 \pi|f(x)| \sqrt{1+\left(f'(x)\right)^2} \mathrm{~d} x .$$

这样, 我们有理由认为该旋转体的侧面积为

$$F=2 \pi \int_a^b f(x) \sqrt{1+\left(f'(x)\right)^2} \mathrm{~d} x .$$

为什么没有使用圆柱的侧面积来做近似? 即, 为什么不用

$$\Delta F_i \approx 2 \pi f\left(\xi_i\right) \Delta x_i .$$

还记得上一小节中, 推导旋转体的体积时, 我们就是用的圆柱的体积来近似的呀. 怎么到了推导旋转体的侧面积是, 我们就变了呢? 这个原因比较复杂, 我们不能展开解释, 但其核心在于近似的误差的阶数不同. 我们本节一开始就说过, 本节的内容非常不严谨, 不要太用力推敲. 因为我们一直在滥用约等于符号 $\approx$. 这个符号表示两边是有误差的, 但是误差的量级/阶数都没有讨论过. 从逻辑上看充满漏洞的. 但目前作为高数 B 的课程, 我们只能暂时先接受这一设定. 未来只要学习严格的数学分析才能搞清楚.

同理可以证明其他情况下的侧面积公式.

结论 3.5.3: 旋转体的侧面积 (仅计算用结论, 经不起细究)

1. **直角坐标系下的函数曲线:** 若函数 $f(x)>0$ 在区间 $[a, b]$ 上光滑, 则该函数在区间 $[a, b]$ 上的曲线绕 x 轴旋转一周所形成的旋转体的侧面积为

$$\int_a^b 2 \pi f(x) \sqrt{1+\left(f'(x)\right)^2} \mathrm{d} x=\int_a^b 2 \pi f(x) \mathrm{~d} s,$$

其中 $\mathrm{d} s=\sqrt{1+\left(f'(x)\right)^2} \mathrm{d} x$ 是直角坐标系下函数曲线的弧微分.

2. **参数方程下的曲线:** 当曲线弧 Γ 由参数方程 $\begin{cases}x=x(t), \\ y=y(t),\end{cases} \quad(\alpha \leq t \leq \beta,\ y(t)>0)$ 给出时, 该旋转体的侧面积为

$$\int_\alpha^\beta 2 \pi y(t) \sqrt{\left(x'(t)\right)^2+\left(y'(t)\right)^2} \mathrm{d} t=\int_a^b 2 \pi y(t) \mathrm{~d} s,$$

其中 $\mathrm{d} s=\sqrt{\left(x'(t)\right)^2+\left(y'(t)\right)^2} \mathrm{d} t$ 是参数方程下的弧微分.

3. **极坐标系下的函数曲线:** 当曲线弧 Γ 由极坐标系下的函数曲线 $r=r(\theta)$ 给出时, 该旋转体的侧面积公式为

$$s=\int_a^\beta 2 \pi r(\theta) \sin \theta \sqrt{r^2(\theta)+\left(r'(\theta)\right)^2} \mathrm{d} \theta=\int_a^b 2 \pi r(\theta) \sin \theta \mathrm{~d} s,$$

其中 $\mathrm{d} s=\sqrt{r^2(\theta)+\left(r'(\theta)\right)^2} \mathrm{d} \theta$ 是极坐标系下函数曲线的弧微分.

例 3.5.6 求抛物线弧段

$$y^2 = 4(4-x) \quad \left(-1 \leq x \leq \frac{7}{2}\right)$$

绕 x 轴旋转一周所形成旋转体的侧面积.

证明. 我们首先从抛物线弧段的方程中将 y 解出, 使之成为 x 的函数:

$$y = \pm 2\sqrt{4-x} \quad \left(-1 \leq x \leq \frac{7}{2}\right).$$

其中符号 "±" 代表着抛物线的上, 下两支. 根据旋转体的侧面积公式中 $y > 0$ 的要求, 我们取正号. 这时, 经过简单计算得到

$$y' = -\frac{1}{\sqrt{4-x}} \quad 与 \quad \sqrt{1+(y')^2} = \sqrt{\frac{5-x}{4-x}}.$$

于是, 所求的侧面积为

$$F = 2\pi \int_{-1}^{\frac{7}{2}} 2\sqrt{4-x} \cdot \sqrt{\frac{5-x}{4-x}}\, \mathrm{d}x = 4\pi \int_{-1}^{\frac{7}{2}} \sqrt{5-x}\, \mathrm{d}x = 14\sqrt{6}\pi.$$

□

第四章　微分中值定理与泰勒公式

从前面几章中我们看到, 微积分的概念及其基本定理有效地解决了物理学及几何学中的许多问题. 本章将进一步讨论微积分在研究函数性态方面的应用, 其核心内容是微分中值定理与泰勒 (Taylor) 公式. 微分中值定理建立了导数与函数值之间的联系, 使我们得以根据导数某些性质去推断函数的性态; 而泰勒公式则告诉我们, 一个函数可以由一个多项式来近似替代. 两者增进了我们对函数性质的理解, 并有重要的应用价值.

作为微分中值定理及泰勒公式的应用, 本章还将讨论未定式的极限, 极值问题, 函数作图, 曲线的曲率等内容.

4.1　微分中值定理

4.1.1　总结: 所有的中值定理

为了读者不和之前的学习的两个中值定理搞混, 我们把它们和本章涉及到所有微分中值定理陈列到一起. 本总结请日后再回顾. 无论是普通的中值定理, 积分中值定理, 还是微分中值定理, 它们都有明显的几何含义, 请对比它们之间几何含义的不同之处. 需要注意的是: 普通的中值定理, 积分中值定理的适用条件, 只要求函数 $f(x)$ 的连续性. 而微分中值定理, 正如其名, 进一步要求函数 $f(x)$ 的可导性. 毕竟所有微分中值定理的等式结果都出现了导数 $f'(x)$, 而普通的中值定理, 积分中值定理则不需要.

1. 中值定理: 设函数 $f(x)$ 在闭区间 $[a,b]$ 上连续. 如果 $f(a) \neq f(b)$, 则对于任意一个严格位于 $f(a)$ 和 $f(b)$ 之间的值 η, 都必存在一点 $c \in (a,b)$, 使得
$$f(c) = \eta.$$
2. 积分中值定理: 设函数 $f(x)$ 在闭区间 $[a,b]$ 上连续, 则必存在一点 $c \in (a,b)$, 使得
$$\int_a^b f(x)\ \mathrm{d}x = f(c)(b-a).$$
3. 微分中值定理 (本节和下一节内容):
   (a) 罗尔 (Rolle) 中值定理 (定理 4.1.1): 设函数 $f(x)$ 在闭区间 $[a,b]$ 上连续, 在开区间 (a,b) 内可导, 如果 $f(a) = f(b)$, 则必存在一点 $c \in (a,b)$, 使得
   $$f'(c) = 0.$$
   (b) 拉格朗日 (Lagrange) 中值定理 (定理 4.1.2), 默认的所谓 “微分中值定理”: 设函数 $f(x)$ 在闭区间 $[a,b]$ 上连续, 在开区间 (a,b) 内可导, 则必存在一点 $c \in (a,b)$, 使得
   $$f'(c) = \frac{f(b)-f(a)}{b-a}.$$

(c) 柯西 (Cauchy) 中值定理 (定理 4.1.4): 设函数 $f(x)$ 与 $g(x)$ 在闭区间 $[a,b]$ 上连续, 在开区间 (a,b) 内可导, 并且恒有 $g'(x)\neq 0$, 则必存在一点 $c\in(a,b)$, 使得

$$\frac{f(b)-f(a)}{g(b)-g(a)}=\frac{f'(c)}{g'(c)}.$$

这三种微分中值定理是依次推广的: 当柯西中值定理中取 g 为恒等函数, 即 $g(x)=x$ 时, 它就变成了拉格朗日中值定理. 当拉格朗日中值定理中添加了 $f(a)=f(b)$ 的条件, 它就变成了罗尔中值定理. 所以, 数学形式上最简单的罗尔中值定理是后两种的一个特殊情况. 然而, 罗尔定理最重要的作用是用于证明后两个定理. 后两个定理的证明过程中, 都是通过构造一个辅助函数, 再对辅助函数作用罗尔定理, 最终就能得到相应的结论.

4.1.2 罗尔中值定理, 拉格朗日中值定理

微分中值定理是微分学中超级重要的定理, 它建立了导数与函数值之间的联系.

在第二章中, 我们曾提到一个简单而基本的命题: 在一个开区间内导数恒等于零的函数当且仅当它是一个常数函数. 该命题的一个方向 (常数函数 $\Rightarrow$ 导数恒为零) 是十分容易证明的: 设 $f(x)\equiv c$ (c 为常数), 这时, 对于任一点 x 及 $\Delta x\neq 0$, 总有 $\Delta y=f(x+\Delta x)-f(x)=0$, 因而直接使用导数的定义, 则有 $f'(x)\equiv 0$. 困难的是命题的另一个方向 (导数恒为零 $\Rightarrow$ 常数函数), 假若不用微分中值定理, 这一结论的证明将是麻烦的.

微分中值定理的一种特殊形式是罗尔 (Rolle) 中值定理. 我们将通过这个罗尔中值定理推导出其他两个微分中值定理.

定理 4.1.1: 罗尔中值定理

设函数 $f(x)$ 在闭区间 $[a,b]$ 上连续, 在开区间 (a,b) 内可导, 如果 $f(a)=f(b)$, 则必存在一点 $c\in(a,b)$, 使得

$$f'(c)=0.$$

证明. 假定 $f(x)$ 是常数函数, 则 (a,b) 中任意一点 x 均满足 $f'(x)=0$, 故 (a,b) 中任意一点都可取作为 c. 这时, 定理的结论显然成立.

现在假定 $f(x)$ 不是常数函数. 因为 $f(x)$ 在闭区间 $[a,b]$ 上连续, 根据闭区间上的极值定理, $f(x)$ 可以在 $[a,b]$ 上取到最大值和最小值. 我们考虑: 如果函数在两个端点处的值 $f(a)=f(b)$ 既是 $f(x)$ 的最大值, 又是 $f(x)$ 的最小值, 则最大值等于最小值, 那么 $f(x)$ 就变成了常数函数. 所以, 函数不能在两个端点处同时取得最大值和最小值. 因此, $f(x)$ 的最大值和最小值中必定至少有一个在开区间 (a,b) 内达到. 图 4.1 就展示了三种允许的可能情况.

可以暂时先假定最大值在 (a,b) 内达到, 如图 4.1 . 也就是说, 存在一点 $c\in(a,b)$, 使得 $f(c)$ 是 $f(x)$ 在 $[a,b]$ 上的最大值:

$$f(x)\leq f(c),\quad \forall x\in[a,b].$$

利用 $f(x)$ 在点 c 处的可导性, 可知极限

$$\lim_{x\to c}\frac{f(x)-f(c)}{x-c}$$

存在, 并等于 $f'(c)$. 我们将证明上述极限等于零, 从而说明 $f'(c)=0$.

1. 我们先考虑右极限 $\displaystyle\lim_{x\to c+0}\frac{f(x)-f(c)}{x-c}$. 注意到对于 c 的右侧某个邻域内的所有 x, 恒有 $x-c>0$, 且 $f(x)-f(c)\leq 0$. 故, 对于 c 的右侧某个邻域内的所有 x, 总有 $\dfrac{f(x)-f(c)}{x-c}\leq 0$. 因此, 根据定理-函数极

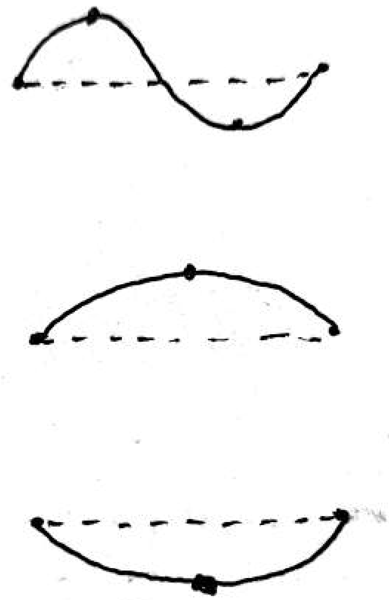


图 4.1: 罗尔定理证明用图

限的保序性, 右极限满足

$$\lim_{x\to c+0}\frac{f(x)-f(c)}{x-c}\le 0.$$

2. 再考虑左极限 $\lim\limits_{x\to c-0}\dfrac{f(x)-f(c)}{x-c}$. 注意到对于 c 的左侧某个邻域内的所有 x, 恒有 $x-c<0$, 且 $f(x)-f(c)\le 0$. 故, 对于 c 的左侧某个邻域内的所有 x, 总有 $\dfrac{f(x)-f(c)}{x-c}\ge 0$. 因此, 根据定理-函数极限的保序性, 右极限满足

$$\lim_{x\to c-0}\frac{f(x)-f(c)}{x-c}\ge 0.$$

3. 由于极限 $\lim\limits_{x\to c}\dfrac{f(x)-f(c)}{x-c}$ 的存在性要求右极限与左极限相等, 于是

$$f'(c)=\lim_{x\to c+0}\frac{f(x)-f(c)}{x-c}\le 0,\quad f'(c)=\lim_{x\to c-0}\frac{f(x)-f(c)}{x-c}\ge 0.$$

若一个数同时满足非负和非正, 则它一定是零. 所以,

$$f'(c)=0.$$

当 $f(x)$ 的最小值在 (a,b) 内达到时, 如图 4.1 . 讨论完全类似. 证毕. □

关于罗尔中值定理的证明, 我们要强调的是, 它本质上依赖于闭区间上的连续函数有最大值与最小值这一定理.

罗尔中值定理的几何意义是: 一条可微曲线弧, 如果两个端点的连线平行于 x 轴 (因为我们要求 $f(a)=f(b)$), 则在该曲线弧上非端点处有一点, 使得过该点的切线也平行于 x 轴 (见图 4.2). 显然, 这是一条关于曲线弧的几何性质的命题. 任何几何性质应该与坐标轴的选取无关. 这意味着, 在图 4.2 中, 固定曲线的图像不动, 如果将整个 x,y 轴以原点为中心, 朝着任意方向旋转, 那么该曲线弧的上述几何性质是不会变得. 但是, 如

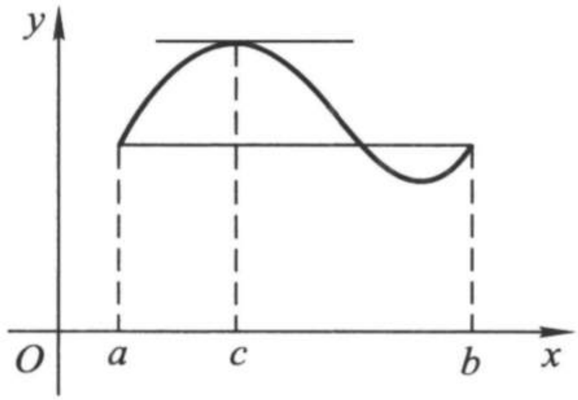


图 4.2: 罗尔中值定理的几何意义

果以新的 x, y 轴下再去考虑这个曲线弧对应的函数, 通常不会再有 $f(a) = f(b)$ 了. 换句话说, 要求端点连线平行于 x 轴不是本质的. 如果我们放弃这个条件, 会自然想到应该成立下列命题: *在一条可微曲线弧上一定存在一点 (非端点), 使得过该点的切线平行于两个端点的连线*. 这一几何命题翻译成分析学的语言就是我们所说的拉格朗日中值定理.

定理 4.1.2: 拉格朗日 (Lagrange) 中值定理, 微分中值定理

设函数 $f(x)$ 在闭区间 $[a, b]$ 上连续, 在开区间 (a, b) 内可导, 则必存在一点 $c \in (a, b)$, 使得

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

上述公式称为**微分中值公式**.

拉格朗日 (Lagrange) 中值定理是我们通常默认的所谓 "微分中值定理". 该公式的右侧是和 c 无关的常数, 如果能够知道导函数 $f'(x)$ 的具体表达式, 则可以通过解方程算出 c.

注意 微分中值定理常用证明套路: 构造一个辅助函数, 该辅助函数满足罗尔中值定理的使用条件. 再对辅助函数作用罗尔定理, 最终变形就能得到相应结果.

证明. 构造辅助函数

$$g(x) := f(x) - \frac{f(b) - f(a)}{b - a}(x - a), \quad \forall x \in [a, b].$$

很容易验证: 函数 $g(x)$ 在 $[a, b]$ 上连续, 在 (a, b) 内可导[1]. 其导数为

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}, \quad \forall x \in (a, b).$$

并且两个端点的值相等:

$$g(a) = g(b) = f(a).$$

[1]想想为什么? g 的构造就是: f 加上一个线性函数, 再加上一个常数.

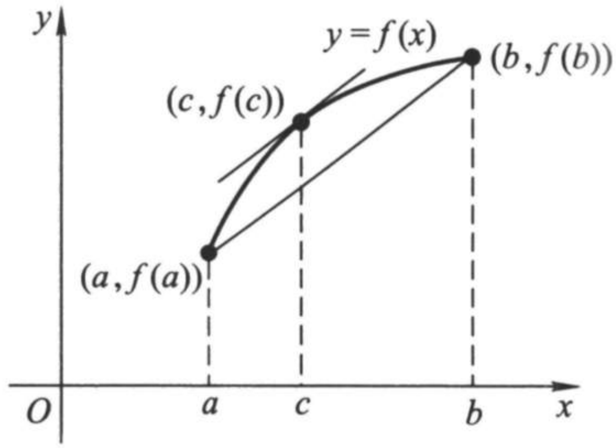


图 4.3: 拉格朗日中值定理的几何意义

于是, $g(x)$ 满足罗尔中值定理的使用条件. 对 $g(x)$ 应用罗尔中值定理就得到: 存在一点 $c \in (a,b)$, 使得 $g'(c) = 0$, 即

$$g'(c) = f'(c) - \frac{f(b)-f(a)}{b-a} = 0.$$

证毕. □

从几何上看, $\dfrac{f(b)-f(a)}{b-a}$ 是函数 $f(x)$ $(a \le x \le b)$ 所代表曲线弧的两个端点 $(a, f(a))$ 与 $(b, f(b))$ 连线的斜率. $f'(c)$ 等于这个值, 表明该曲线弧上点 $(c, f(c))$ 处的切线平行于两个端点的连线 (见图 4.3). 可以想象一下, 在图 4.3 中的整个坐标系逆时针旋转, 直到两个端点的连线和 x 轴平行, 这个不就刚刚的罗尔中值定理吗.

> **注意** 在任何微分中值定理中, 把导数和函数值联系起来的特殊点 c 不一定是唯一的. 我们只说明它的存在性, 并没有要求唯一性. 下面的例题是一个例子.

例 4.1.1 设函数 $f(x) = x^3 - x^2 - x + 1$, 写出 $f(x)$ 在区间 $[-1, 2]$ 上的微分中值公式, 并求出其中的 c.

证明. $f'(x) = 3x^2 - 2x - 1$. 对 $f(x)$ 在区间 $[-1, 2]$ 上应用拉格朗日中值定理, 得到等式

$$\frac{f(2)-f(-1)}{2-(-1)} = f'(c).$$

即 $\dfrac{3-0}{3} = 3c^2 - 2c - 1$, 也即 $3c^2 - 2c - 2 = 0$, 解得 $c = \dfrac{2 \pm \sqrt{4+24}}{6} = \dfrac{1 \pm \sqrt{7}}{3}$. 可见, 有两个 c 值:

$$c_1 = \frac{1-\sqrt{7}}{3}, \quad c_2 = \frac{1+\sqrt{7}}{3}$$

这两个值都在 $(-1, 2)$ 中. 见图 拉格朗日中值定理 - GeoGebra. □

拉格朗日中值定理中的微分中值公式 (当 $a<b$):

$$f'(c)=\frac{f(b)-f(a)}{b-a},\quad \exists c\in(a,b),$$

有非常灵活的数学表达方式. 它们之间都是完全等价的, 但是某些形式使用起来更加方便. 比如, 它可以等价地变形为

$$f(b)-f(a)=f'(c)(b-a),\quad \exists c\in(a,b),$$

或者,

$$f(b)=f(a)+f'(c)(b-a),\quad \exists c\in(a,b).$$

另外, 我们总是可以将事实 $c\in(a,b)$ 写成**参数化形式**:

$$c=a+\theta(b-a),\quad \exists\theta\in(0,1).$$

当参数 $\theta=0$, 则 $c=a$; 当参数 $\theta=1$, 则 $c=b$; 参数 θ 就像一个滑动条控制器, 调整着 c 在区间 (a,b) 上的具体位置. 这是表示一个数在某一个区间的常用等价写法 (小技巧). 则此时, 可以把微分中值公式相应地改写成

$$f'\left(a+\theta(b-a)\right)=\frac{f(b)-f(a)}{b-a},\quad \exists\theta\in(0,1),$$

或者

$$f\left(b\right)=f\left(a\right)+f'\left(a+\theta(b-a)\right)\left(b-a\right),\quad \exists\theta\in(0,1).$$

我们还发现, 参数化形式 $c=a+\theta(b-a)$ 还等于

$$c=(1-\theta)a+\theta b,\quad \exists\theta\in(0,1).$$

或者

$$c=\theta_1a+\theta_2b,\quad \exists\theta_1,\theta_2\in(0,1),\theta_1+\theta_2=1.$$

如果把这些对于事实 $c\in(a,b)$ 的参数化形式放回原来的微分中值公式, 那么我们又会得到很多完全等价的微分中值公式. 这里不再一一列举.

观察微分中值公式 $f'(c)=\dfrac{f(b)-f(a)}{b-a}$ 可以发现, 只要 $a\neq b$, 那么是 $a<b$ 还是 $b<a$ 其实不重要, 中值定理总是成立. 重要的是什么? 是**形式顺序**! 比如, 我们不能写成

$$f'(c)=\frac{f(b)-f(a)}{a-b},\text{ 或者 } f'(c)=\frac{f(a)-f(b)}{b-a},$$

这绝对是错误的. 但是可以写成

$$f'(c)=\frac{f(a)-f(b)}{a-b}.$$

刚刚提到的 c 参数化形式还有一个好处是, 根本不用在乎 $a<b$ 还是 $b<a$. 只要给定两个不等的数 a,b, 那么下列参数化:

$$\begin{aligned}c&=a+\theta(b-a),\quad \exists\theta\in(0,1);\\&=(1-\theta)a+\theta b,\quad \exists\theta\in(0,1);\\c&=\theta_1a+\theta_2b,\quad \exists\theta_1,\theta_2\in(0,1),\theta_1+\theta_2=1;\end{aligned}$$

皆可完美表达 c 严格位于 a,b 之间的事实.

若用 x_0 与 x 分别替代 a 与 b 这两个符号, 则微分中值公式可写成:

$$f(x)=f\left(x_0\right)+f'(c)\left(x-x_0\right), \quad \exists c \text{ 严格位于} x, x_0 \text{ 之间}.$$

如果对 c 参数化, 且进一步引入符号 $\Delta x=x-x_0$ (这和导数那一章节的符号一致), 则我们有**等式**:

$$\begin{gathered} f(x)=f\left(x_0\right)+f'(x_0+\theta \Delta x)\left(x-x_0\right), \quad \exists \theta \in(0,1); \\ f\left(x_0+\Delta x\right)=f\left(x_0\right)+f'\left(x_0+\theta \Delta x\right) \Delta x, \quad \exists \theta \in(0,1). \end{gathered}$$

应当注意, 在这的 x 可以大于 x_0, 也可以小于 x_0, 故而 Δx 可正可负. 在第二章中我们曾证明过, 当函数 $f(x)$ 在一点 x_0 处可导, 即可微时, 我们有**极限等式**:

$$\begin{gathered} f(x)=f\left(x_0\right)+f'\left(x_0\right)\left(x-x_0\right)+o\left(x-x_0\right) \quad\left(x \rightarrow x_0\right); \\ f\left(x_0+\Delta x\right)=f\left(x_0\right)+f'\left(x_0\right) \Delta x+o(\Delta x) \quad(\Delta x \rightarrow 0). \end{gathered}$$

(以上两个就是马上要学习的泰勒展开至 1 阶的情况) 另外, 当 x 非常接近于 x_0, 或 Δx 非常接近于 0 时, 我们有**近似等式**:

$$\begin{gathered} f(x) \approx f\left(x_0\right)+f'\left(x_0\right)\left(x-x_0\right); \\ f\left(x_0+\Delta x\right) \approx f\left(x_0\right)+f'\left(x_0\right) \Delta x. \end{gathered}$$

以上三组常见的关于固定点 x_0 和其增量的等式们长得非常相似. 但请不要混淆它们. 考虑一下它们各自成立的条件是什么?

结论 4.1.1: 记忆微分中值公式

至此, 这么多的微分中值公式到底怎么记忆呢? 核心就是:

$$f(\text{符号 } 1)=f(\text{符号 } 2)+f'(\text{某个位于两个符号之间的量, 可任意参数化}) \times(\text{符号 } 1 \text{ - 符号 } 2).$$

$$f(\text{符号 } 1)-f(\text{符号 } 2)=f'(\text{某个位于两个符号之间的量, 可任意参数化}) \times(\text{符号 } 1 \text{ - 符号 } 2).$$

$$f'(\text{某个位于两个符号之间的量, 可任意参数化})=\frac{f(\text{符号 } 1)-f(\text{符号 } 2)}{(\text{符号 } 1 \text{ - 符号 } 2)}.$$

你会发现, 前面所有的表示式都是上面这样子的. 这里面的 “顺序” 和 “正负号” 一点都不能记错!

利用变形的微分中值公式, 我们可以更加方便地应用它. 比如下面本节开头所讲的推论.

推论 4.1.1

设函数 $f(x)$ 在某个开区间上可导, 且导数恒为零, 则 $f(x)$ 在该区间上是常数函数.

证明. 设开区间为 (A, B), 在 (A, B) 内任意取定一点 a. 对于任意的 $x \in(A, B), x \neq a$, 我们在区间 $[a, x]$ (或 $[x, a]$) 上应用拉格朗日中值定理即得

$$f(x)=f(a)+f'(c)(x-a),$$

其中 c 是介于 a 与 x 之间的一点. 根据假定 $f'(c)=0$, 有 $f(x)=f(a)$. 由于 x 是 (A, B) 内任意的点, 这表明 $f(x)$ 是一个常数函数. 证毕. □

这就补充证明了前两章中反复应用的一个命题: 在一个开区间内导数恒等于零的函数当且仅当它是一个常数函数. 这一命题在讨论不定积分以及微积分基本定理时是一个重要依据[2].

问题 思考题: 试用质点运动的瞬时速度与位移函数的关系来说明罗尔中值定理与拉格朗日中值定理的物理意义.

证明. 1. 罗尔中值定理表明: 如果一个函数在闭区间上连续, 在开区间上可导, 且在端点的函数值相等, 那么至少存在一个点, 在这个点上, 函数的导数为零.

物理背景: 假设一个质点沿着一条直线运动[3], 位移函数为 $x(t)$, 其中 t 是时间. 我们考察质点在时间区间 $[t_1, t_2]$ 上的运动, 假设质点在这两个时间点的位置相同, 即 $x(t_1) = x(t_2)$. 根据罗尔中值定理, 必定存在一个时间 $t_c \in (t_1, t_2)$, 使得质点在这个时刻的瞬时速度为零, 即 $v(t_c) = \dfrac{dx}{dt}(t_c) = 0$.

物理解释: 这个时刻 t_c 可以理解为质点的瞬时速度为零的时刻, 即质点在这个时刻停下来. 假设质点的运动路径是连续且可导的 (即没有剧烈跳跃或突然停止), 那么如果它在起点和终点的位置相同, 必定有一个时刻, 质点的瞬时速度会为零, 这代表着它一定会有一个" 停顿" 时刻.

2. 拉格朗日中值定理表明: 如果一个函数在闭区间上连续, 在开区间上可导, 那么至少存在一个点 $c \in (a, b)$, 使得函数的导数在这个点等于该函数在端点处的增量除以区间的长度.

物理背景: 假设质点的位移函数是 $x(t)$, 我们考察质点在区间 $[t_1, t_2]$ 上的运动, 质点在端点的位移分别为 $x(t_1)$ 和 $x(t_2)$. 根据拉格朗日中值定理, 存在一个时间点 $t_c \in (t_1, t_2)$, 使得瞬时速度 $v(t_c)$ 等于平均速度:

$$v(t_c) = \frac{x(t_2) - x(t_1)}{t_2 - t_1}.$$

物理解释: 瞬时速度 $v(t_c)$ 在某一时刻等于质点在整个时间区间内的平均速度. 换句话说, 尽管质点在运动过程中速度可能变化, 但至少在某一时刻, 它的瞬时速度会等于整个区间内的平均速度. 这个时刻 t_c 可以理解为质点的" 平均速度点", 即在这个点上, 质点的运动速率等于其在整个区间上的平均速率.[4]

□

4.1.3 微分中值定理的应用: 导数与单调性

拉格朗日中值定理在研究函数的性质上有重要的作用, 它是一个" 桥梁", 建立了函数的导数与增量之间的联系. 比如以下重要定理.

[2]可以复习一下, 当时我们确实经常用它.

[3]否则的转一圈也能回到原来的出发点

[4]注意, 在学习积分中值定理的时候, 我们也得到过同样的物理解释, 请对比他们的异同.

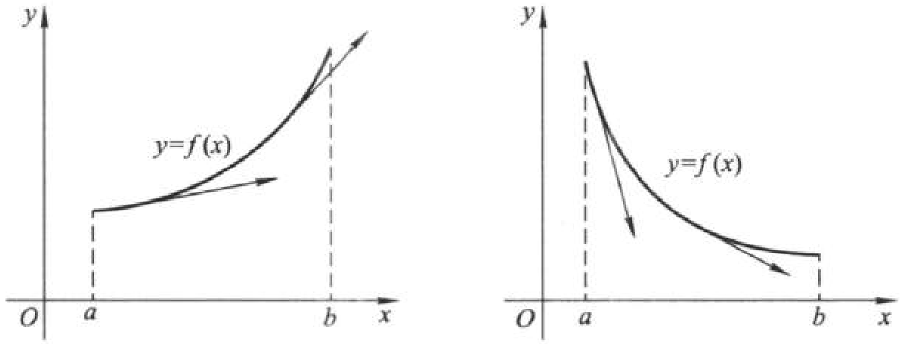


图 4.4: 函数的导数与增量之间的联系

定理 4.1.3: 导数与单调性

设函数 $f(x)$ 在闭区间 $[a,b]$ 上连续, 在开区间 (a,b) 内可导. 若

$$f'(x)>0\ (\text{或}\geq 0),\quad \forall x\in(a,b),$$

则 $f(x)$ 在 $[a,b]$ 上严格递增 (或递增); 若

$$f'(x)<0\ (\text{或}\leq 0),\quad \forall x\in(a,b),$$

则 $f(x)$ 在 $[a,b]$ 上严格递减 (或递减).

我们还应提醒读者注意, 上述定理的逆命题不一定成立. 事实上, 严格递增的函数的导数未必处处大于零. 比如, $y=x^3$ 是严格递增的, 但在点 $x=0$ 处的导数为零.

证明. 若函数 $f(x)$ 在某一个开区间内的导数 $f'(x)$ 总是严格大于零的, 则由拉格朗日中值定理可知, 对于该区间上的任意两点 x_1,x_2, 且 $x_1>x_2$, 有

$$f(x_1)-f(x_2)=f'(c)(x_1-x_2)>0,\quad \exists c\ \text{严格位于}x_1,x_2\ \text{之间}.$$

即 $f(x_1)>f(x_2)$. 也就是说, $f(x)$ 在这个区间上是严格递增的. 完全类似地, 由 $f(x)$ 的导数 $f'(x)$ 处处严格小于零可以推出 $f(x)$ 是严格递减的. 如果将条件放宽到 $f'(x)\geq 0$ (或 ≤ 0), 那么结论也就相应地改成递增 (或递减), 不再是严格递增 (或严格递减). 证毕. □

这个定理的几何意义也是十分清楚的, 见图 4.4. 如果在函数 $f(x)$ 所代表曲线的切线上标出一个箭头, 使切线的方向指向自变量增加的方向, 那么我们就会发现: 当导数 $f'(x)$ 大于零时, 切线指向右上方; 而当导数 $f'(x)$ 小于零时, 切线指向右下方. 切线总是指向右上方时, 意味着相应的函数递增; 而切线总是指向右下方时, 意味着相应的函数递减.

定理 4.1.3 为研究函数的单调性区间提供了有效的办法.

例 4.1.2 设函数 $f(x)=x^3-3x^2+1$, 求 $f(x)$ 的单调性区间.

证明. 由于 $f'(x) = 3x^2 - 6x = 3x(x-2)$, 故只在点 $x = 0$ 及 $x = 2$ 处导数为零; 在区间 $(-\infty, 0)$ 中, $f'(x) > 0, f(x)$ 是严格递增的; 在区间 $(0, 2)$ 中, $f'(x) < 0, f(x)$ 是严格递减的; 在区间 $(2, +\infty)$ 中, $f'(x) > 0$, $f(x)$ 是严格递增的. □

微分中值定理的另一个应用是证明不等式. 注意, 微分中值定理的使用条件非常的宽松. 故而在实际做题中非常好用. 尤其是证明题. 要时常想想看如何套用中值定理会发生什么.

例 4.1.3 证明不等式:

$$\frac{x}{1+x} < \ln(1+x) < x, \quad \forall x > 0.$$

证明. 分别使用两种中值定理证明.

1. **使用微分中值定理证明.** 对于任意取定的 $x > 0$, 我们对函数 $\ln(1+x)$ 在区间 $[0, x]$ 上应用微分中值定理, 则存在 c $(0 < c < x)$, 使得

$$\ln(1+x) - \ln 1 = \ln(1+x) = \frac{1}{1+c}x.$$

将 $0 < c < x$ 分别以其上界和下界代入, 就得到

$$\frac{x}{1+x} < \ln(1+x) < x.$$

2. **使用积分中值定理证明.** 首先观察到[5]

$$\ln(1+x) = \int_0^x \frac{\mathrm{d}t}{1+t}, \quad \forall x > 0.$$

根据积分中值定理[6], 存在某个点 $c \in (0, x)$, 使得

$$\int_0^x \frac{\mathrm{d}t}{1+t} = \frac{1}{1+c} \cdot x.$$

所以

$$\ln(1+x) = \frac{x}{1+c},$$

其中 $c \in (0, x)$. 因此, $1 < 1 + c < 1 + x$, 即,

$$\frac{1}{1+x} < \frac{1}{1+c} < 1.$$

上式同乘以 x. 于是我们得到:

$$\frac{x}{1+x} < \ln(1+x) < x.$$

□

[5] 和微分中值定理不同, 如果构造不出来一个定积分, 则很难用它.

[6] 如果一个函数在区间 $[a, b]$ 上是连续的, 那么存在某个点 $c \in (a, b)$, 使得: $\int_a^b f(t)\,dt = f(c) \cdot (b - a)$.

我们隐隐地感觉到, 微分中值定理和积分中值定理有着某种关联. 我们不严谨地做如下推导, 尝试发现两者之间的关系: 设函数 $f(x)$ 在区间 $[a,b]$ 上连续, $F(x)$ 是其一个原函数. 则有,

$$\begin{aligned}\int_a^b f(x)\mathrm{d}x &= F(b)-F(a) \quad \text{(使用微积分基本定理)}\\ &= F'(c)(b-a) \quad \text{(使用微分中值定理)}\\ &= f(c)(b-a).\end{aligned}$$

最后的结果就是积分中值定理. 但是这个推导过程忽略了微积分基本定理和微分中值定理的使用条件, 比较粗糙且不严谨. 下面的例题的证明才是严谨的.

例 4.1.4 通过应用微分中值定理, 推导出积分中值定理.

证明. 设函数 $f(x)$ 在区间 $[a,b]$ 上连续. 考虑其变上限积分函数

$$F(x)=\int_a^x f(t)\ \mathrm{d}t, \quad x\in[a,b].$$

根据定积分的性质[7], $F(x)$ 是一个在 $[a,b]$ 上连续, 在 (a,b) 上可导的函数, 且:

$$F'(x)=f(x), \quad \forall x\in(a,b).$$

对 $F(x)$ 使用微分中值定理, 则存在一个点 $c\in(a,b)$ 使得:

$$F'(c)=\frac{F(b)-F(a)}{b-a}.$$

另一方面, $F'(c)=f(c)$, 且 $F(b)-F(a)=\displaystyle\int_a^b f(x)\ \mathrm{d}t$. 因此, 我们得到了:

$$f(c)=\frac{\int_a^b f(x)\,dx}{b-a}.$$

这就证明了积分中值定理. □

例 4.1.5 证明不等式: $\mathrm{e}^x>1+x, x\neq 0$.

证明. 令 $f(x)=\mathrm{e}^x-1-x$, 则 $f(x)$ 在 $(-\infty,+\infty)$ 可导, 并且

$$f'(x)=\mathrm{e}^x-1.$$

在 $(0,+\infty)$ 中, $f'(x)>0$, 所以 $f(x)$ 是严格递增的, 则

$$f(x)>f(0)=0, \quad x\in(0,+\infty);$$

在 $(-\infty,0)$ 中, $f'(x)<0$, 所以 $f(x)$ 是严格递减的, 则

$$f(x)>f(0)=0, \quad x\in(-\infty,0).$$

于是对于任意的 $x\neq 0$, 有 $f(x)>0$, 即 $\mathrm{e}^x>1+x$. □

[7]定理 2.9.2: 变上限积分函数是一个原函数

历史的注记: 拉格朗日是法国著名数学家, 他所完成的《分析力学》是牛顿《自然哲学的数学原理》之后的又一部经典力学著作, 其中运用变分原理与分析方法建立了完整的力学体系. 他在数学的众多分支领域有杰出贡献, 被认为是对分析学产生全面影响的数学家之一.

4.1.4 柯西中值定理

柯西 (Cauchy) 中值定理是拉格朗日中值定理的一种推广. 这种推广的主要意义在于给出求未定式极限的洛必达 (L'Hospital) 法则.

定理 4.1.4: 柯西中值定理

设函数 $f(x)$ 与 $g(x)$ 在闭区间 $[a,b]$ 上连续, 在开区间 (a,b) 内可导, 并且恒有 $g'(x)\neq 0$, 则必存在一点 $c\in(a,b)$, 使得
$$\frac{f(b)-f(a)}{g(b)-g(a)}=\frac{f'(c)}{g'(c)}.$$

该公式的左侧是和 c 无关的常数, 如果能够知道 $\dfrac{f'(c)}{g'(c)}$ 的具体表达式, 则可以通过解方程算出 c. 很容易会想到柯西中值定理的结论是分别对 $f(x)$ 及 $g(x)$ 使用拉格朗日中值定理后相除的结果. 其实, 这是不对的, 因为拉格朗日中值定理中的中间点 c 与函数及区间有关, 对 $f(x)$ 及 $g(x)$ 分别使用拉格朗日中值定理时所得的中间点可能不是同一点, 而现在定理中要求的是同一点.

> **注意** 微分中值定理常用证明套路: 构造一个辅助函数, 该辅助函数满足罗尔中值定理的使用条件. 再对辅助函数作用罗尔定理, 最终变形就能得到相应结果.

证明. 先指出一个事实: 已知 $g(x)$ 在闭区间 $[a,b]$ 上连续, 在开区间 (a,b) 内可导. 假设有 $g(a)=g(b)$, 则可以对 $g(x)$ 使用罗尔中值定理, 则
$$g'(c)=0,\quad c\in(a,b).$$
这与定理的条件 $g'(x)\neq 0, \forall x\in(a,b)$ 相矛盾. 故假设不成立, 因而推出 $g(a)\neq g(b)$.

由以上事实, 我们可以构造辅助函数
$$h(x):=f(x)-\frac{f(b)-f(a)}{g(b)-g(a)}(g(x)-g(a)),\quad \forall x\in[a,b].$$
容易验证 $h(x)$ 在 $[a,b]$ 上连续, 在 (a,b) 内可导, 其导数为
$$h'(x)=f'(x)-\frac{f(b)-f(a)}{g(b)-g(a)}g'(x),\quad \forall x\in(a,b).$$
并且两个端点的值相等:
$$h(a)=h(b)=f(a).$$
于是, $h(x)$ 满足罗尔中值定理的使用条件. 对 $h(x)$ 应用罗尔中值定理就得到: 存在一点 $c\in(a,b)$, 使得 $h'(c)=0$, 即
$$h'(c)=f'(c)-\frac{f(b)-f(a)}{g(b)-g(a)}g'(c)=0.$$
证毕. □

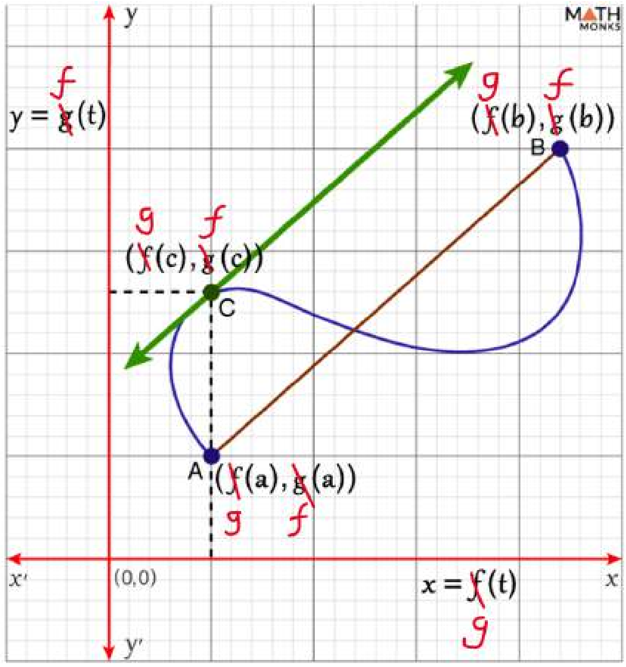


图 4.5: 柯西中值定理的几何意义, 图片来源: Cauchy’s Mean Value Theorem - Proof, Examples, and Applications

柯西中值定理的几何意义与拉格朗日中值定理的几何意义完全相同, 它们都表明: 在一条平面可微曲线弧上有一点, 该点处的切线平行于两个端点的连线. 它们的差别在于: 拉格朗日中值定理中的曲线弧是由 $y=f(x)$ 给出的, 而柯西中值定理中的曲线弧是由参数方程

$$\begin{cases} x = g(t), \\ y = f(t) \end{cases} \quad (a \leqslant t \leqslant b)$$

给出的. 见图 4.5. 事实上, 在柯西中值定理的条件下, 这个参数方程所代表曲线弧的端点连线在 Oxy 平面上的斜率是

$$\frac{f(b)-f(a)}{g(b)-g(a)},$$

而该曲线弧在点 $t=c$ 处的切线斜率是

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{f'(c)}{g'(c)}.$$

4.2 洛必达法则

柯西中值定理的一个重要应用是给出了求未定式极限的洛必达法则.

4.2.1 未定式

现在我们先来解释什么是未定式. 我们考虑两个函数 $f(x)$ 及 $g(x)$. 设它们在一点 a 附近有定义 (在点 a 处可以有定义, 也可以没有定义). 并假设当 $x \to a$ 时, 极限 $\lim\limits_{x\to a} f(x)$ 和 $\lim\limits_{x\to a} g(x)$ 都存在. 我们根据这两者极限分别是否为零, 可以做如下 4 种情况的讨论:

1. 若 $\lim\limits_{x\to a} g(x) \neq 0$, 且 $\lim\limits_{x\to a} f(x) \neq 0$, 则根据函数极限的运算定理[8], 我们有

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)} = \text{两个非零常数之比}.$$

这说明极限存在. 这时求极限是毫无困难的.

2. 若 $\lim\limits_{x\to a} g(x) \neq 0$, 且 $\lim\limits_{x\to a} f(x) = 0$, 则再根据函数极限的运算定理, 我们有

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)} = \frac{0}{\text{某个非零常数}} = 0.$$

这亦说明极限存在. 这时求极限亦是毫无困难的. 结合第一种情况, 只要分母的极限不为零, 那么当分子的极限存在时 (无论是否为零), 这种求商的极限总是非常简单的.

3. 若 $\lim\limits_{x\to a} g(x) = 0$, 且 $\lim\limits_{x\to a} f(x) \neq 0$, 则也容易知道这是一个常数乘以无穷大量:

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \infty.$$

故没有极限, 这时结论也是清楚的.

4. 最困难的是 $\lim\limits_{x\to a} g(x) = 0$ 与 $\lim\limits_{x\to a} f(x) = 0$ 同时发生的情况. 在这种情况下, 当 $x\to a$ 时, 比式 $\dfrac{f(x)}{g(x)}$ 的变化趋势有多种可能性:

    (a) 可能有极限, 如我们最熟悉的

    $$\lim_{x\to 0}\frac{\sin x}{x} = 1;$$

    (a) 也可能没有极限 (例如震荡), 如 (注意分子是无穷小乘上一个有界函数, 故而还是无穷小, 极限为零)

    $$\frac{x\sin\frac{1}{x}}{x} = \sin\frac{1}{x} \quad (x\to 0);$$

    (a) 甚至有可能趋向于 ∞, 如

    $$\frac{x^2+\sin^2 x}{x^3} \to \infty \quad (x\to 0)$$

正因为如此, 最后一种类型的商的极限值得给出一个特别的名字, 即 $\dfrac{0}{0}$ 型未定式. 类似地, $\dfrac{\infty}{\infty}$ 的求极限结果也是多种多样的.

定义 4.2.1: 未定式

- 当 $\lim\limits_{x\to a} g(x) = \lim\limits_{x\to a} f(x) = 0$ 时, 我们称 $\lim\limits_{x\to a}\dfrac{f(x)}{g(x)}$ 是 $\dfrac{0}{0}$ **型未定式**.
- 当 $\lim\limits_{x\to a} g(x) = \lim\limits_{x\to a} f(x) = \infty$ 时, 我们称 $\lim\limits_{x\to a}\dfrac{f(x)}{g(x)}$ 是 $\dfrac{\infty}{\infty}$ **型未定式**.

还有很多其他类型的未定式, 但就目前能遇到的所有的未定式而言, 它们都可以转化成 $\dfrac{0}{0}, \dfrac{\infty}{\infty}$ 这两种未定式其中之一, 然后再做处理.

[8]设 $f(x)$ 及 $g(x)$ 是定义在点 a 的一个空心邻域内的函数. 若 $\lim\limits_{x\to a} f(x) = l_1$, $\lim\limits_{x\to a} g(x) = l_2$, 并且当 $l_2 \neq 0$ 时, 有 $\lim\limits_{x\to a}\dfrac{f(x)}{g(x)} = \dfrac{l_1}{l_2}$.

4.2.2 洛必达法则

洛必达法则为处理 $\dfrac{0}{0}$ 型未定式以及其他类型未定式提供了一种途径.

定理 4.2.1: 洛必达法则 1 ($x \to a$, $\dfrac{0}{0}$ 型)

设函数 $f(x)$ 及 $g(x)$ 在点 a 的一个空心邻域内有定义, 在该空心邻域内可导[a], 并且 $g'(x) \neq 0$. 如果 $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} g(x) = 0$, 并且极限 $\lim\limits_{x\to a} \dfrac{f'(x)}{g'(x)}$ 存在, 那么 $\lim\limits_{x\to a} \dfrac{f(x)}{g(x)}$ 的极限存在, 并且

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

[a]其实稍稍有些废话了, 可导的必要条件之一就是在该点处有定义.

证明. 由定理所给的条件: $f(x)$, $g(x)$ 在点 a 的一个空心邻域内可导, 所以在该空心邻域上连续. 我们补充定义 $f(a) = g(a) = 0$,[9] 于是, 这时 $f(x)$ 和 $g(x)$ 在点 a 的一整个邻域内 (包含点 a 在内) 连续.

设 x 是在该邻域内任意取定的一点, 且 $x \neq a$. 则 $f(x)$, $g(x)$ 在区间 $[a,x]$ 或 $[x,a]$ 上连续, (a,x) 或 (x,a) 上可导. (注意, 根据条件, 它们在点 a 处是否可导并不清楚, 但这也不重要.) 因此可以对 $f(x)$, $g(x)$ 应用柯西中值定理, 则存在介于 a 与 x 之间的一点, 记为 c_x, 使得

$$\frac{f'(c_x)}{g'(c_x)} = \frac{f(x)-f(a)}{g(x)-g(a)} = \frac{f(x)}{g(x)}.$$

因 c_x 介于 a 与 x 之间, 故当 $x \to a$ 时, c_x 也趋向于 a. 令 $x \to a$, 并对上式取极限, 即有

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{c_x\to a} \frac{f'(c_x)}{g'(c_x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

最后一个等式, 我们用形式的符号 x 去替代了 c_x. 证毕. □

显然, 洛必达法则对于其他极限过程, 如 $x \to a+0$ 或 $x \to a-0$, 也是成立的. 这是不需重新证明的, 前面所做的证明完全适用于这两种情况.

例 4.2.1 求极限 $\lim\limits_{x\to 0} \dfrac{\ln(1+x)-x}{x^2}$.

证明. 显然, 这是一个 $\dfrac{0}{0}$ 型未定式, 应用洛必达法则有

$$\lim_{x\to 0} \frac{\ln(1+x)-x}{x^2} = \lim_{x\to 0} \frac{\frac{1}{1+x}-1}{2x} = \lim_{x\to 0} \frac{\frac{-x}{1+x}}{2x} = \frac{1}{2}\lim_{x\to 0} \frac{-1}{1+x} = -\frac{1}{2}.$$

这里提醒读者: 可以按现在的次序书写式子, 但心中应该明白这里的推理步骤 (因果关系) 是反过来的. 严格地讲, 是因为极限 $\lim\limits_{x\to 0} \dfrac{\frac{1}{1+x}-1}{2x}$ 存在, 才肯定了极限 $\lim\limits_{x\to 0} \dfrac{\ln(1+x)-x}{x^2}$ 存在并等于 $-\dfrac{1}{2}$. □

在应用洛必达法则时应当注意以下事项:

1. 首先要验证所讨论的极限是否是未定式. 如果不是未定式, 就不能用洛必达法则.

[9]我们可以这么做. 无论 $f(x)$, $g(x)$ 在 $x = a$ 处有没有定义, 以及有定义时函数值是不是等于 0, 都没有任何关系. 例题中常见的情况是 $f(a) = g(a) = 0$ 本就存在, 不用补充该点处的定义.

2. 其次, 洛必达法则的本质逻辑是:

$$\frac{f'(x)}{g'(x)} \text{ 的极限存在} \Longrightarrow \frac{f(x)}{g(x)} \text{ 的极限存在并与前者相等}$$

但是, 当 $\dfrac{f'(x)}{g'(x)}$ 的极限不存在时, 那我们什么结论都不能去论断, 尤其是不能去说 $\dfrac{f(x)}{g(x)}$ 的极限也不存在. 因为上述关系不是一个充分必要条件. 例如, 对于

$$f(x) = x^2 \sin\frac{1}{x}, \quad g(x) = x.$$

显然, 当 $x \to 0$ 时, $\dfrac{f(x)}{g(x)} = x\sin\dfrac{1}{x} \to 0$, 极限存在. 但是

$$\frac{f'(x)}{g'(x)} = \frac{2x\sin\frac{1}{x} - \cos\frac{1}{x}}{1} = 2x\sin\frac{1}{x} - \cos\frac{1}{x}.$$

当 $x \to 0$ 时, $\dfrac{f'(x)}{g'(x)}$ 没有极限, 因为第一想趋近于 0, 但后一项在震荡.

3. 有些未定式, 对其使用一次洛必达法则后还是未定式, 这时就需使用多次洛必达法则.

例 4.2.2 求极限 $\lim\limits_{x\to 0}\dfrac{e^x - 1 - x - \frac{1}{2}x^2}{x^3}$.

证明. 这是一个 $\dfrac{0}{0}$ 型未定式, 分子与分母求导数后仍是 $\dfrac{0}{0}$ 型未定式, 继续对分子与分母求导数后才归结为非未定式并存在极限. 我们有

$$\begin{aligned}
\lim_{x\to 0}\frac{e^x - 1 - x - \frac{1}{2}x^2}{x^3} &= \lim_{x\to 0}\frac{e^x - 1 - x}{3x^2} \quad (\text{判断是 } \frac{0}{0}, \text{洛一下}) \\
&= \lim_{x\to 0}\frac{e^x - 1}{6x} \quad (\text{判断是 } \frac{0}{0}, \text{再洛一下}) \\
&= \lim_{x\to 0}\frac{e^x}{6} \quad (\text{判断是 } \frac{0}{0}, \text{再洛一下}) \\
&= \frac{1}{6}.
\end{aligned}$$

与关于上一题的说明类似, 这里的**应用洛必达法则时, 推理步骤与式子的书写次序相反**:

- 由于上式中第 4 个极限存在, 才肯定了第 3 个极限存在并与之相等;
- 由于上式中第 3 个极限存在, 才肯定了第 2 个极限存在并与之相等;
- 由于上式中第 2 个极限存在, 才肯定了第 1 个极限存在并与之相等;

通过这样依次反推上去, 我们就可以说明, 第 1 个极限存在并等于最后一个极限. □

假设 $f(x), g(x)$ 的 n 阶导数都存在, 并且对于求极限 $\lim\limits_{x\to a}\dfrac{f(x)}{g(x)}$ 可以连续使用 n 次洛必达法则, 其背后的逻辑链条是

$$\frac{f^{(n)}(x)}{g^{(n)}(x)} \text{ 的极限存在}$$

$$
\begin{aligned}
&\Longrightarrow \frac{f^{(n-1)}(x)}{g^{(n-1)}(x)} \text{ 的极限存在并与前者相等} \\
&\Longrightarrow \frac{f^{(n-2)}(x)}{g^{(n-2)}(x)} \text{ 的极限存在并与前者相等} \\
&\qquad \vdots \\
&\Longrightarrow \frac{f^{(1)}(x)}{g^{(1)}(x)} \text{ 的极限存在并与前者相等} \\
&\Longrightarrow \frac{f(x)}{g(x)} \text{ 的极限存在并与前者相等}
\end{aligned}
$$

请大家熟悉这个连续使用洛必达法则的逻辑链, 下一节我们将用这个来证明泰勒公式.

例 4.2.3 求极限 $\lim\limits_{x\to 0+0} \dfrac{\sqrt{x}-\sin\sqrt{x}}{x^{\frac{3}{2}}}$.

证明. 这是一个 $\dfrac{0}{0}$ 型未定式. 应用洛必达法则, 有

$$
\lim_{x\to 0+0} \frac{\sqrt{x}-\sin\sqrt{x}}{x^{\frac{3}{2}}} = \lim_{x\to 0+0} \frac{\frac{1}{2\sqrt{x}} - \cos\sqrt{x}\cdot\frac{1}{2\sqrt{x}}}{\frac{3}{2}\sqrt{x}} = \lim_{x\to 0+0} \frac{1-\cos\sqrt{x}}{3x}.
$$

这仍然是一个 $\dfrac{0}{0}$ 型未定式. 再使用一次洛必达法则, 得到

$$
\lim_{x\to 0+0} \frac{1-\cos\sqrt{x}}{3x} = \lim_{x\to 0+0} \frac{\sin\sqrt{x}\cdot\frac{1}{2\sqrt{x}}}{3} = \frac{1}{6}\lim_{x\to 0+0} \frac{\sin\sqrt{x}}{\sqrt{x}} = \frac{1}{6}.
$$

故最终有

$$
\lim_{x\to 0+0} \frac{\sqrt{x}-\sin\sqrt{x}}{x^{\frac{3}{2}}} = \frac{1}{6}.
$$

□

注意 洛必达法则不是孤立地被使用的. 大家在做习题的时候, 一定要多多结合之前学习的各种求极限方法. 比如先换元, 然后用洛必达等等.

对于 $x\to\infty$ 的极限过程, 洛必达法则仍然成立, 但定理的叙述与证明需要改写. 下面仅以 $x\to\infty$ 的情况为例给出定理的证明. 对于 $x\to+\infty$ 或 $x\to-\infty$ 的情况, 定理的结论依然成立.

定理 4.2.2: 洛必达法则 2 ($x\to\infty$, $\dfrac{0}{0}$ 型)

设函数 $f(x)$ 及 $g(x)$ 在 $\mathbf{R}\backslash[-A,A]$ 内可导, 并且 $g'(x)\neq 0$, 其中 $A>0$. 如果 $\lim\limits_{x\to\infty} f(x) = \lim\limits_{x\to\infty} g(x) = 0$, 并且极限 $\lim\limits_{x\to\infty}\dfrac{f'(x)}{g'(x)}$ 存在, 那么 $\lim\limits_{x\to\infty}\dfrac{f(x)}{g(x)}$ 的极限存在, 并且

$$
\lim_{x\to\infty}\frac{f(x)}{g(x)} = \lim_{x\to\infty}\frac{f'(x)}{g'(x)}.
$$

证明. 我们的整体思路是通过换元, 将 $x \to \infty$ 变成 $x \to 0$, 从而直接使用之前版本的洛必达法则 - 定理 4.2.1.

令 (这里是一种隐蔽的换元法)

$$F(x) := f\left(\frac{1}{x}\right), \quad G(x) := g\left(\frac{1}{x}\right),$$

那么函数 $F(x)$ 与 $G(x)$ 在 $(-\delta, \delta)\backslash\{0\}$ 中有定义且可导 (可导性来自于复合函数的性质), 其中 $\delta := \frac{1}{A}$. 根据 $\lim\limits_{x\to\infty} f(x) = \lim\limits_{x\to\infty} g(x) = 0$, 可知

$$\lim_{x\to 0} F(x) = \lim_{x\to 0} f\left(\frac{1}{x}\right) = \lim_{x\to\infty} f(x) = 0,$$
$$\lim_{x\to 0} G(x) = \lim_{x\to 0} g\left(\frac{1}{x}\right) = \lim_{x\to\infty} g(x) = 0.$$

另外, 其导函数分别是

$$F'(x) = -\frac{1}{x^2}\cdot f'\left(\frac{1}{x}\right), \quad G'(x) = -\frac{1}{x^2}\cdot g'\left(\frac{1}{x}\right), \quad \forall x \in (-\delta, \delta)\backslash\{0\}.$$

则,

$$\lim_{x\to 0}\frac{F'(x)}{G'(x)} = \lim_{x\to 0}\frac{-\frac{1}{x^2}\cdot f'\left(\frac{1}{x}\right)}{-\frac{1}{x^2}\cdot g'\left(\frac{1}{x}\right)} = \lim_{x\to 0}\frac{f'\left(\frac{1}{x}\right)}{g'\left(\frac{1}{x}\right)} = \lim_{x\to\infty}\frac{f'(x)}{g'(x)}.$$

条件中极限 $\lim\limits_{x\to\infty}\frac{f'(x)}{g'(x)}$ 的存在性说明了 $\lim\limits_{x\to 0}\frac{F'(x)}{G'(x)}$ 的存在性. 注意到, 由于 $g'(x) \neq 0$, 所以 $G'(x) = -\frac{1}{x^2}\cdot g'\left(\frac{1}{x}\right) \neq 0, \forall x \in (-\delta, \delta)\backslash\{0\}$. 此时, $F(x)$ 与 $G(x)$ 已经完全符合了洛必达法则 - 定理 4.2.1 中的使用条件, 因此我们有

$$\lim_{x\to 0}\frac{F(x)}{G(x)} = \lim_{x\to 0}\frac{F'(x)}{G'(x)}.$$

再换元回去,

$$\lim_{x\to\infty}\frac{f(x)}{g(x)} = \lim_{x\to\infty}\frac{f'(x)}{g'(x)}.$$

证毕. □

例 4.2.4 求极限 $\lim\limits_{x\to+\infty}\frac{\ln\left(1+\frac{1}{x}\right)}{\operatorname{arccot} x}$.

证明. 这是一个 $\frac{0}{0}$ 型未定式. 应用洛必达法则, 有

$$\lim_{x\to+\infty}\frac{\ln\left(1+\frac{1}{x}\right)}{\operatorname{arccot} x} = \lim_{x\to+\infty}\frac{-\frac{1}{x^2}/\left(1+\frac{1}{x}\right)}{-\frac{1}{1+x^2}} = \lim_{x\to+\infty}\frac{1+x^2}{x^2+x} = 1.$$

□

除了 $\frac{0}{0}$ 型未定式之外, 我们之前还定义了 $\frac{\infty}{\infty}$ 型未定式. 顾名思义, 所谓的 $\frac{\infty}{\infty}$ 型未定式, 就是分子与分母都趋向于无穷大的比式的极限. 对于 $\frac{\infty}{\infty}$ 型未定式, 同样有洛必达法则. 现在仅以 $x \to a$ 的极限过程为例, 叙述如下. 其证明较长, 这里从略.

定理 4.2.3: 洛必达法则 3 ($x \to a$, $\dfrac{\infty}{\infty}$ 型)

设函数 $f(x)$ 及 $g(x)$ 在点 a 的一个空心邻域内可导, 并且 $g'(x) \neq 0$. 假若 $\lim\limits_{x\to a} f(x) = \infty$, $\lim\limits_{x\to a} g(x) = \infty$, 并且极限 $\lim\limits_{x\to a} \dfrac{f'(x)}{g'(x)}$ 存在, 那么 $\lim\limits_{x\to a} \dfrac{f(x)}{g(x)}$ 的极限存在, 并且

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

对于所有其他类型的极限过程 ($x \to a+0, x \to a-0, x \to \infty, x \to +\infty$ 或 $x \to -\infty$, 定理 4 给出的 $\dfrac{\infty}{\infty}$ 型未定式的洛必达法则都是成立的, 这里不再一一叙述.

例 4.2.5 求极限 $\lim\limits_{x\to 1-0} \dfrac{\ln\left(\tan\frac{\pi}{2}x\right)}{\ln(1-x)}$.

证明. 这是一个 $\dfrac{\infty}{\infty}$ 型未定式, 利用洛必达法则, 得到

$$\begin{aligned}
\lim_{x\to 1-0} \frac{\ln\left(\tan\frac{\pi}{2}x\right)}{\ln(1-x)} &= \lim_{x\to 1-0} \frac{\frac{1}{\tan\frac{\pi}{2}x} \cdot \sec^2\frac{\pi}{2}x \cdot \frac{\pi}{2}}{\frac{-1}{1-x}} \\
&= \lim_{x\to 1-0} \frac{-\pi(1-x)}{2\sin\frac{\pi}{2}x \cos\frac{\pi}{2}x} \\
&= \lim_{x\to 1-0} \frac{-1}{\sin\frac{\pi}{2}x} \cdot \frac{\frac{\pi}{2}(1-x)}{\sin\frac{\pi}{2}(1-x)} = -1.
\end{aligned}$$

□

例 4.2.6 求极限 $\lim\limits_{x\to +\infty} \dfrac{\ln x}{x^\alpha} \quad (\alpha > 0)$.

证明. 利用洛必达法则, 得到

$$\lim_{x\to+\infty} \frac{\ln x}{x^\alpha} = \lim_{x\to+\infty} \frac{\frac{1}{x}}{\alpha x^{\alpha-1}} = \lim_{x\to+\infty} \frac{1}{\alpha x^\alpha} = 0.$$

□

例 4.2.7 证明: $\lim\limits_{x\to+\infty} \dfrac{P(x)}{\mathrm{e}^x} = 0$, 其中 $P(x)$ 为 x 的一个 n 次多项式, $n > 1$.

证明. 设 $P(x) = a_0x^n + \cdots + a_n$, 则 $P(x)$ 的 n 阶导数 $P^{(n)}(x) = a_0 n!$ 是一个常数. 而 e^x 的 n 阶导数仍是它本身. 由

$$\lim_{x\to+\infty} \frac{P^{(n)}(x)}{\mathrm{e}^x} = 0,$$

根据洛必达法则推出

$$\lim_{x\to+\infty} \frac{P^{(n-1)}(x)}{\mathrm{e}^x} = \lim_{x\to+\infty} \frac{P^{(n)}(x)}{\mathrm{e}^x} = 0.$$

所以, 反复运用洛必达法则, 最后即推出

$$\lim_{x\to+\infty}\frac{P(x)}{\mathrm{e}^x}=\lim_{x\to+\infty}\frac{P'(x)}{\mathrm{e}^x}=\cdots=\lim_{x\to+\infty}\frac{P^{(n)}(x)}{\mathrm{e}^x}=0.$$

证毕.

这里应该指出: 在这个例子中每使用一次洛必达法则, 都应当说明所讨论的极限是一个 $\frac{\infty}{\infty}$ 型未定式. 事实上, 这是很容易说明的: 分母 e^x 总是无穷大量, 而分子 $P^{(k)}(x)$ 是一个 $n-k$ $(1\leqslant k<n)$ 次多项式, 对于每个小于 n 的 k, $P^{(k)}(x)$ 当 $x\to+\infty$ 时也是一个无穷大量. □

除了 $\frac{0}{0}$ 型与 $\frac{\infty}{\infty}$ 型未定式之外, 尚有许多其他类型的未定式, 处理这些类型未定式的原则是设法经过变形后化成 $\frac{0}{0}$ 型或 $\frac{\infty}{\infty}$ 型未定式, 再使用洛必达法则. 比如:

1. $0\cdot\infty$ 型 $\longrightarrow$
   (a) 因为乘以 ∞, 就是除以 0, 故它可以变成 $\frac{0}{0}$ 型;
   (b) 又因为乘以 0, 就是除以 ∞, 它还可以变成 $\frac{\infty}{\infty}$ 型.
2. $\infty-\infty$ 型 $\longrightarrow$ 尝试通分, 化成 $\frac{0}{0}$ 型或 $\frac{\infty}{\infty}$ 型.
3. 0^0 型, ∞^0 型及 1^∞ 型[10] $\longrightarrow$ 管它是什么类型, 不用记. 只要看到幂指函数, 上来就取对数就行了, 得到什么类型是什么类型.

值得说明的是, 对于 $0\cdot\infty$ 型, 我们既可以变成 $\frac{0}{0}$ 型, 也可以变成 $\frac{\infty}{\infty}$ 型. 但是不同变形下应用洛必达法则, 其问题的难度可能不同. 有时会更加复杂, 比如下一题. 所以对于 $0\cdot\infty$ 型要多尝试一下.

例 4.2.8 求极限 $\lim\limits_{x\to0+0}x^x$.

证明. 这是一个 0^0 型未定式, 通过取对数:

$$\ln x^x=x\ln x,$$

1. 可将它化成 $\frac{\infty}{\infty}$ 型未定式:

$$\ln x^x=x\ln x=\frac{\ln x}{\frac{1}{x}},$$

   使用洛必达法则后得

$$\lim_{x\to0+0}\ln x^x=\lim_{x\to0+0}\frac{\frac{1}{x}}{-\frac{1}{x^2}}=\lim_{x\to0+0}-x=0.$$

   再由指数函数的连续性得到

$$\lim_{x\to0+0}x^x=\lim_{x\to0+0}\mathrm{e}^{x\ln x}=\mathrm{e}^{\lim_{x\to0+0}x\ln x}=1.$$

2. 可将它化成 $\frac{0}{0}$ 型未定式:

$$\ln x^x=x\ln x=\frac{x}{\frac{1}{\ln x}},$$

[10]这里的 1 表示, 底数的位置的极限是一个非零常数.

使用洛必达法则后得

$$\lim_{x\to 0+0}\ln x^x=\lim_{x\to 0+0}\frac{1}{-\frac{1}{(\ln x)^2}\cdot\frac{1}{x}}=\lim_{x\to 0+0}-x(\ln x)^2=?$$

这似乎回到了起点. 毕竟我们想求的就是 $\lim\limits_{x\to 0+0}x\ln x$, 怎么还升幂了. 接下来的计算就很难在进行下去了. 但还是注意, $\lim\limits_{x\to 0+0}-x(\ln x)^2$ 的极限肯定是等于 0, 的因为之前我们就算出来了 $\lim\limits_{x\to 0+0}\ln x^x=0$. 所以, 将 $0\cdot\infty$ 型, 变成 $\dfrac{0}{0}$ 型, 还是 $\dfrac{\infty}{\infty}$ 型. 理论上都没有错, 但是计算难度上可能不同.

□

上题的解答过程证明了

$$\lim_{x\to 0+0}x\ln x=0.$$

类似地, 还可以证明 (读者自行证明)

$$\lim_{x\to 0+0}x^\alpha\ln x=0\quad(\alpha>0).$$

另外, 我们还发现了甚至 $\dfrac{0}{0}$ 型和 $\dfrac{\infty}{\infty}$ 型之间也是可以相互转化的. 我们考虑任意两个函数 $f(x),g(x)$, 且 $f'(x)\neq 0,g'(x)\neq 0$, 那么总有

$$\frac{f(x)}{g(x)}=\frac{\frac{1}{g(x)}}{\frac{1}{f(x)}}.$$

当 $\lim f(x)=\lim g(x)=0$ 时, 则 $\lim\dfrac{1}{f(x)}=\lim\dfrac{1}{g(x)}=\infty$; 当 $\lim f(x)=\lim g(x)=\infty$ 时, 则 $\lim\dfrac{1}{f(x)}=\lim\dfrac{1}{g(x)}=0$. 这说明了 $\dfrac{0}{0}$ 型和 $\dfrac{\infty}{\infty}$ 型并不是两个割裂的未定式.

例 4.2.9 求极限 $\lim\limits_{x\to 1}\left(\dfrac{1}{\ln x}-\dfrac{1}{x-1}\right)$.

证明. 这是一个 $\infty-\infty$ 型未定式, 通过通分:

$$\frac{1}{\ln x}-\frac{1}{x-1}=\frac{x-1-\ln x}{(x-1)\ln x}.$$

可将它归结为 $\dfrac{0}{0}$ 型未定式. 使用洛必达法则即得

$$\begin{aligned}\lim_{x\to 1}\left(\frac{1}{\ln x}-\frac{1}{x-1}\right)&=\lim_{x\to 1}\frac{x-1-\ln x}{(x-1)\ln x}=\lim_{x\to 1}\frac{1-\frac{1}{x}}{\ln x+\frac{x-1}{x}}\\&=\lim_{x\to 1}\frac{x-1}{x\ln x+x-1}\\&=\lim_{x\to 1}\frac{1}{\ln x+2}=\frac{1}{2}.\end{aligned}$$

□

例 4.2.10 求极限 $\lim\limits_{x\to 0+0}\left(\dfrac{\sin x}{x}\right)^{\frac{1}{x^2}}$.

证明. 这是一个 1^∞ 型未定式. 设 $y=\left(\dfrac{\sin x}{x}\right)^{\frac{1}{x^2}}$, 通过取对数化成 $\dfrac{0}{0}$ 型未定式:

$$\ln y=\frac{1}{x^2}\ln\frac{\sin x}{x}=\frac{\ln\frac{\sin x}{x}}{x^2}=\frac{\ln(\sin x)-\ln x}{x^2}.$$

使用洛必达法则, 得到

$$\lim_{x\to0+0}\ln y=\lim_{x\to0+0}\frac{\frac{\cos x}{\sin x}-\frac{1}{x}}{2x}=\lim_{x\to0+0}\frac{x\cos x-\sin x}{2x^2\sin x}.$$

这仍是 $\dfrac{0}{0}$ 型未定式, 适当化简并再次使用洛必达法则:

$$\begin{aligned}\lim_{x\to0+0}\ln y&=\lim_{x\to0+0}\frac{x\cos x-\sin x}{2x^2\sin x}\\&=\lim_{x\to0+0}\frac{x\cos x-\sin x}{2x^3}\cdot\frac{x}{\sin x}\\&=\lim_{x\to0+0}\frac{x\cos x-\sin x}{2x^3}\quad(\text{其实就是等价无穷小代换 }\sin x\sim x)\\&=\lim_{x\to0+0}\frac{\cos x-x\sin x-\cos x}{6x^2}\\&=-\frac{1}{6}\lim_{x\to0+0}\frac{\sin x}{x}=-\frac{1}{6}.\end{aligned}$$

由指数函数的连续性得到

$$\lim_{x\to0+0}\left(\frac{\sin x}{x}\right)^{\frac{1}{x^2}}=\lim_{x\to0+0}\mathrm{e}^{\ln y}=\mathrm{e}^{-\frac{1}{6}}.$$

□

例 4.2.11 求极限 $\lim\limits_{x\to\frac{\pi}{2}-0}(\tan x)^{\cos x}$.

证明. 这是一个 ∞^0 型未定式. 设 $y=(\tan x)^{\cos x}$, 则

$$\ln y=\cos x\ln(\tan x)=\frac{\ln(\tan x)}{\sec x},$$

从而问题化为求极限 $\lim\limits_{x\to\frac{\pi}{2}-0}\dfrac{\ln(\tan x)}{\sec x}$. 这是一个 $\dfrac{\infty}{\infty}$ 型未定式. 使用洛必达法则即得

$$\lim_{x\to\frac{\pi}{2}-0}\frac{\ln(\tan x)}{\sec x}=\lim_{x\to\frac{\pi}{2}-0}\frac{\frac{1}{\tan x}\sec^2 x}{\sec x\tan x}=\lim_{x\to\frac{\pi}{2}-0}\frac{\sec x}{\tan^2 x}=\lim_{x\to\frac{\pi}{2}-0}\frac{\frac{1}{\cos x}}{\frac{\sin^2 x}{\cos^2 x}}=\lim_{x\to\frac{\pi}{2}-0}\frac{\cos x}{\sin^2 x}=0.$$

根据指数函数的连续性, 得到

$$\lim_{x\to\frac{\pi}{2}-0}(\tan x)^{\cos x}=\lim_{x\to\frac{\pi}{2}-0}\mathrm{e}^{\ln y}=1.$$

□

4.3 泰勒公式

洛必达法则的另一个重要应用是证明泰勒公式, 而泰勒公式在研究函数性质上无疑是一个极重要的公式.

4.3.1 高次多项式逼近的探索

在第二章中曾证明过, 当函数 $f(x)$ 在一点 x_0 处可导时, 因为等价于可微, 所以我们有极限等式:

$$f(x)=f\left(x_0\right)+f'\left(x_0\right)\left(x-x_0\right)+o\left(x-x_0\right), \quad \left(x\rightarrow x_0\right).$$

这就是说, 当 x 充分靠近 x_0 时, $f(x)$ 的值可以由关于 $\Delta x=x-x_0$ (后面, 我们不用符号 Δx, 而默认把 $x-x_0$ 当作一个整体) 的一个线性函数[11] (也就是一次多项式函数)

$$\left(x-x_0\right)\mapsto f\left(x_0\right)+f'\left(x_0\right)\left(x-x_0\right)$$

近似代替, 其误差是比 $x-x_0$ 高阶的无穷小量. 这一结论启发我们: 用 $x-x_0$ 的高次多项式去逼近函数值 $f(x)$, 可能得到更高的精确度的近似代替. 自然, 这有可能要求 $f(x)$ 有更强的条件, 比如高阶导数的存在性.

我们先来看如何用 $x-x_0$ 的二次多项式来逼近 $f(x)$. 这时, 我们假定 $f(x)$ 在点 x_0 处有二阶导数. 也就是说, 导函数 $f'(x)$ 自身也在点 x_0 附近有定义且在点 x_0 处可导. 于是, 将上面的极限等式可以应用于 $f'(x)$, 即形式地将 $f(x)$ 替换成 $f'(x)$, 将 $f'(x)$ 替换成 $f''(x)$, 于是我们有极限等式:

$$f'(x)=f'\left(x_0\right)+f''\left(x_0\right)\left(x-x_0\right)+o\left(x-x_0\right), \quad \left(x\rightarrow x_0\right).$$

为了得到关于 $f(x)$ 的二次多项式近似式, 我们自然的想法是把以上等式的两边同时取积分, 这样可以预想到左侧的导函数 $f'(x)$ 应该会返还成原函数 $f(x)$. 这里我们不使用不定积分, 而是使用定积分. 设想当 x 充分靠近 x_0, 并固定 x, 考虑定义在 $t\in[x_0,x]$ 的函数:

$$f'(t)=f'\left(x_0\right)+f''\left(x_0\right)\left(t-x_0\right)+o\left(t-x_0\right),$$

然后两边同时对 t 自 x_0 到 x 的区间做定积分, 就得到

$$\int_{x_0}^{x} f'(t)\ \mathrm{d}t=\int_{x_0}^{x} f'(x_0)\ \mathrm{d}t+\int_{x_0}^{x} f''(x_0)(t-x_0)\ \mathrm{d}t+\int_{x_0}^{x} o(t-x_0)\ \mathrm{d}t,$$

进一步展开,

$$f(x)-f\left(x_0\right)=f'\left(x_0\right)\left(x-x_0\right)+\frac{1}{2}f''\left(x_0\right)\left(x-x_0\right)^2+\int_{x_0}^{x} o(t-x_0)\ \mathrm{d}t,$$

即,

$$f(x)=f\left(x_0\right)+f'\left(x_0\right)\left(x-x_0\right)+\frac{1}{2}f''\left(x_0\right)\left(x-x_0\right)^2+\int_{x_0}^{x} o(t-x_0)\ \mathrm{d}t,$$

显然, 右侧的前三项合在一起就是我们想要的 $x-x_0$ 的二次多项式 (应总是把 $x-x_0$ 整体当作一个量). 可最后面还有一个尾巴:

$$\int_{x_0}^{x} o(t-x_0)\ \mathrm{d}t.$$

这如何处理呢? 我们还没有学过怎么处理一个含 o 符号的积分, 我们甚至怀疑上面这个定积分是否是合法定义的? 这里我们做一个简化的处理, 既然 $o(t-x_0)$ 表示的是关于高阶无穷小, 那么我们先任意选择一个满足要求的具体的高阶无穷小, 比如最简单的 $(t-x_0)^2$, 显然有, $(t-x_0)^2=o(t-x_0)$. 于是, 我们用下面的定积分去替代上面那个含 o 符号的积分:

$$\int_{x_0}^{x} o(t-x_0)\ \mathrm{d}t \underset{\text{直觉的替代}}{\longrightarrow} \int_{x_0}^{x}(t-x_0)^2\ \mathrm{d}t=\frac{1}{3}(x-x_0)^3=o\left((x-x_0)^2\right).$$

[11] 虽然严格的线性函数要求不能有额外的常数项 $f\left(x_0\right)$, 但是人们也经常习惯地把加上常数项的线性函数也叫做 "线性函数", 我们自己要清楚这个差别.

所以, 我们得到

$$\int_{x_0}^{x} o(t-x_0)\ \mathrm{d}t = o\left((x-x_0)^2\right).$$

既然被积函数是含 o 符号的, 那么定积分的结果也含 o 符号, 这样的结果直觉上是合理. 把以上对尾巴的处理结果带回原式中, 得到

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2}f''(x_0)(x-x_0)^2 + o\left((x-x_0)^2\right).$$

因为, 以上推导都是在 x 充分靠近 x_0 时才成立的, 所以我们应该在最右边加上极限变量的趋向声明, 最终有:

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2}f''(x_0)(x-x_0)^2 + o\left((x-x_0)^2\right), \quad (x \to x_0).$$

应当严正指出, 以上的推导是不严谨的[12], 所以它不是任何证明, 而只是为了发现 $x - x_0$ 的二次多项式的逼近形式所做的探索. 所谓探索就是不去死磕细节, 而是相对自由地发挥推导. 就像是考试中猜答案, 如果能验证答案确实是对的, 那么我们也就成功了. 下面我们严格证明我们猜测的 "答案" 确实成立.

例 4.3.1 设函数 $f(x)$ 在点 x_0 的某个邻域内有定义, 并在点 x_0 处有 2 阶导数. 定义二次多项式函数

$$T_2(x) := f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2}f''(x_0)(x-x_0)^2.$$

证明 $f(x) = T_2(x) + o\left((x-x_0)^2\right)$, $(x \to x_0)$ 成立.

证明. 实际上就是要证明

$$\lim_{x \to x_0} \frac{f(x) - T_2(x)}{(x-x_0)^2} = 0.$$

上式左端恰好是一个 $\dfrac{0}{0}$ 型未定式. 分子与分母分别求导数之后得到

$$\begin{aligned} \frac{f'(x) - T_2'(x)}{2(x-x_0)} &= \frac{f'(x) - \left(f'(x_0) + f''(x_0)(x-x_0)\right)}{2(x-x_0)} \\ &= \frac{1}{2} \cdot \frac{f'(x) - f'(x_0)}{(x-x_0)} - \frac{1}{2}f''(x_0). \end{aligned}$$

由 $f(x)$ 在点 x_0 处二阶导数的存在性可知 $\displaystyle\lim_{x \to x_0} \frac{f'(x) - f'(x_0)}{x - x_0} = f''(x_0)$. 于是, 利用洛必达法则就证明了

$$\lim_{x \to x_0} \frac{f(x) - T_2(x)}{(x-x_0)^2} = \lim_{x \to x_0} \frac{f'(x) - T_2'(x)}{2(x-x_0)} = 0.$$

□

- 总结一下, 我们证明了当 $f(x)$ 在点 x_0 处有 2 阶导数时, 对于 x_0 的附近, 有二阶多项式近似:

$$\begin{aligned} f(x) = f(x_0) &+ f'(x_0)(x-x_0) + \frac{1}{2}f''(x_0)(x-x_0)^2 \\ &+ o\left((x-x_0)^2\right), \quad (x \to x_0). \end{aligned}$$

  这个二次多项式逼近是比之前的一次多项式逼近更加精确.

[12] 显然我们忽略了很多东西: 比如, 对一个含有 o 尾巴的极限等式, 两边可以取定积分然后结果相等吗?

- 现在我们有理由猜想, 当 $f(x)$ 在点 x_0 处有 3 阶导数时, 对于 x_0 的附近, 有三阶多项式近似:

$$f(x)=f\left(x_0\right)+f'\left(x_0\right)\left(x-x_0\right)+\frac{1}{2}f''\left(x_0\right)\left(x-x_0\right)^2+\frac{1}{3!}f'''\left(x_0\right)\left(x-x_0\right)^3 \\ +o\left(\left(x-x_0\right)^3\right),\quad \left(x\to x_0\right).$$

这个三次多项式逼近是比二次多项式逼近更加精确.

- 进一步猜想, 当 $f(x)$ 在点 x_0 处有 n 阶导数时, 对于 x_0 的附近, 有 n 阶多项式近似:

$$f(x)=f\left(x_0\right)+f'\left(x_0\right)\left(x-x_0\right)+\frac{1}{2!}f''\left(x_0\right)\left(x-x_0\right)^2+\frac{1}{3!}f'''\left(x_0\right)\left(x-x_0\right)^3+\cdots+\frac{1}{n!}f^{(n)}\left(x_0\right)\left(x-x_0\right)^n \\ +o\left(\left(x-x_0\right)^n\right),\quad \left(x\to x_0\right).$$

4.3.2 泰勒公式

本小节以后会大量用到对于小 o 符号的运算操作, 请先看小节 4.3.5.

定理 4.3.1: 泰勒公式 (Taylor's Formula)

设函数 $f(x)$ 在点 x_0 的某个邻域内有定义, 并在点 x_0 处有 n $(n\geqslant 1)$ 阶导数, 则 $f(x)$ 在点 x_0 附近有如下展开式:

$$f(x)=\underbrace{\sum_{k=0}^{n}\frac{f^{(k)}(x_0)}{k!}(x-x_0)^k}_{T_n(x):=}+\underbrace{o\left(\left(x-x_0\right)^n\right)}_{\text{余项}},\quad \left(x\to x_0\right).$$

其中, 关于 $x-x_0$ 或者说关于 x 的多项式

$$T_n(x)=f\left(x_0\right)+f'\left(x_0\right)\left(x-x_0\right)+\frac{1}{2!}f''\left(x_0\right)\left(x-x_0\right)+\frac{1}{3!}f'''\left(x_0\right)\left(x-x_0\right)+ \\ \cdots+\frac{1}{n!}f^{(n)}\left(x_0\right)\left(x-x_0\right)^n$$

称为 (n 阶) **泰勒多项式**, 而 $o\left(\left(x-x_0\right)^n\right)$ 称为**余项 (remainder)**, 或皮亚诺余项. 整个等式整体 (泰勒多项式 + 余项) 称为 $f(x)$ 在点 x_0 处的 (n 阶) **泰勒公式 (Taylor's formula)**. 已知函数 $f(x)$ 的表达式, 将其泰勒公式写出来的过程成为**泰勒展开**. 因为上述这种形式的泰勒公式只告诉我们当 $x\to x_0$ 时函数的性态, 所以这样的泰勒公式也称为**局部泰勒公式**.

证明. 本证明过程个刚才的例题 4.3.1 完全一致. 实际上就是要证明

$$\lim_{x\to x_0}\frac{f(x)-T_n(x)}{\left(x-x_0\right)^n}=0.$$

我们可以连续使用 $n-1$ 次洛必达法则 (每次使用后都是 $\frac{0}{0}$ 型未定式), 我们有

$$\lim_{x\to x_0}\frac{f(x)-T_n(x)}{\left(x-x_0\right)^n}=\lim_{x\to x_0}\frac{f'(x)-T_n'(x)}{n\left(x-x_0\right)^{n-1}}=\cdots=\lim_{x\to x_0}\frac{f^{(n-1)}(x)-T_n^{(n-1)}(x)}{n!\left(x-x_0\right)}.$$

另外, n 次多项式 $T_n(x)$ 的 $n-1$ 阶导函数为一次多项式:

$$T_n^{(n-1)}(x)=f^{(n-1)}\left(x_0\right)+f^{(n)}\left(x_0\right)\left(x-x_0\right).$$

带回得

$$\begin{aligned}
\lim_{x\to x_0}\frac{f(x)-T_n(x)}{(x-x_0)^n} &= \lim_{x\to x_0}\frac{f^{(n-1)}(x)-T_n^{(n-1)}(x)}{n!\,(x-x_0)}\\
&= \lim_{x\to x_0}\frac{f^{(n-1)}(x)-\left(f^{(n-1)}(x_0)+f^{(n)}(x_0)(x-x_0)\right)}{n!\,(x-x_0)}\\
&= \lim_{x\to x_0}\left[\frac{1}{n!}\cdot\frac{f^{(n-1)}(x)-f^{(n-1)}(x_0)}{(x-x_0)}-\frac{1}{n!}f^{(n)}(x_0)\right]\\
&= 0,
\end{aligned}$$

这里最后一步用到 $f(x)$ 在点 x_0 处 n 阶导数的存在性和定义:

$$\lim_{x\to x_0}\frac{f^{(n-1)}(x)-f^{(n-1)}(x_0)}{(x-x_0)}=f^{(n)}(x_0).$$

证毕. □

注释 4.3.1 如果定理 4.3.1 中的 n 趋向正无穷的整数会发生什么? 我们假设函数 $f(x)$ 在点 x_0 的某个邻域内有任意阶导数. 后面我们会知道, 泰勒公式的级数表达式:

$$f(x)=\sum_{k=0}^{\infty}\frac{f^{(k)}(x_0)}{k!}(x-x_0)^k.$$

此时, 余项消失了, 因为当 x 充分靠近 x_0 时, 有 $\lim\limits_{n\to\infty}o\left((x-x_0)^n\right)=0$.

函数 $f(x)$ 在点 $x_0=0$ 处的泰勒公式称为**麦克劳林 (Maclaurin) 公式**:

$$\begin{aligned}
f(x) &= \sum_{k=0}^{n}\frac{f^{(k)}(0)}{k!}x^k+o(x^n),\quad (x\to 0)\\
&= f(0)+f'(0)x+\frac{1}{2!}f''x^2+\frac{1}{3!}f'''(0)x^3+\cdots+\frac{1}{n!}f^{(n)}(0)x^n+o(x^n),\quad (x\to 0).
\end{aligned}$$

最好用语言来记忆泰勒公式, 比如,

$$f(x)=\sum_{k=0}^{n}\left[\frac{f\text{ 在 }x_0\text{ 处的 }k\text{ 阶导}}{k\text{ 的阶乘}}\times(x-x_0)\text{ 的 }k\text{ 次方}\right]+o\left((x-x_0)^n\right).$$

再简单点就是, (这里不需要严谨, 只是为了方便记忆)

$$f(x)=\sum_{k=0}^{n}\left[\frac{k\text{ 阶导}}{k\text{ 阶乘}}\times(x-x_0)\text{ 的 }k\text{ 次方}\right]+o\left((x-x_0)^n\right).$$

特别是在点 $x_0=0$ 处,

$$f(x)=\sum_{k=0}^{n}\left[\frac{k\text{ 阶导}}{k\text{ 阶乘}}\times x\text{ 的 }k\text{ 次方}\right]+o(x^n).$$

现在我们给出一些常见初等函数的局部泰勒公式. 详细情况请参考小节 4.3.4.

例 4.3.2 求函数 $y=\mathrm{e}^x$ 在点 $x=0$ 处的泰勒公式.

证明. 很明显, $y=\mathrm{e}^x$ 的任意阶导数在点 $x=0$ 处的值都是 1, 故有

$$\begin{aligned}\mathrm{e}^x&=1+\frac{x}{1!}+\frac{1}{2!}x^2+\cdots+\frac{1}{n!}x^n+o\left(x^n\right),\quad(x\to 0)\\&=\sum_{k=0}^{n}\frac{x^k}{k!}+o\left(x^n\right),\quad(x\to 0).\end{aligned}$$

□

请在电脑上查看 Taylor polynomial graphs –GeoGebra, 将函数设置为 $f(x)=e^x$, 并调整近似的阶数 n. 看看 n 非常大的时候会发生什么.

例 4.3.3 求函数 $y=\sin x$ 在点 $x=0$ 处的泰勒公式.

证明. 问题的关键在于求出 $y=\sin x$ 在点 $x=0$ 处的 n 阶导数. 为此, 我们将 $(\sin x)'=\cos x$ 写成 $(\sin x)'=\sin\left(\frac{\pi}{2}+x\right)$. 于是, 类似地, 对于一般正整数 n, 有[13]

$$(\sin x)^{(n)}=\sin\left(\frac{n\pi}{2}+x\right).$$

这样, 我们有

$$\begin{aligned}(\sin x)^{(2k)}\Big|_{x=0}&=0,\quad k=1,2,\cdots\\(\sin x)^{(2k+1)}\Big|_{x=0}&=(-1)^k,\quad k=0,1,\cdots\end{aligned}$$

这就是说, 正弦函数在点 $x=0$ 处的泰勒公式中 x 的偶次方幂系数均为零, 而 x 的奇次方幂系数的符号正负相间:

$$\sin x=x-\frac{1}{3!}x^3+\frac{1}{5!}x^5-\cdots+\frac{(-1)^k}{(2k+1)!}x^{2k+1}+o\left(x^{2k+1}\right),\quad(x\to 0).$$

或者认为展开式结束于偶数项:

$$\sin x=x-\frac{1}{3!}x^3+\frac{1}{5!}x^5-\cdots+\frac{(-1)^k}{(2k+1)!}x^{2k+1}+o\left(x^{2k+2}\right),\quad(x\to 0).$$

简言之, 偶数项消失, 只有奇数项, 符号正负交替. 注意, 0 次幂算是偶数项. 我们发现右侧每一个奇次方幂都是奇函数, 它们的和也是奇函数, 这和 $\sin x$ 的性质一致. □

请在电脑上查看 Taylor polynomial graphs –GeoGebra, 将函数设置为 $f(x)=\sin x$, 并调整近似的阶数 n. 看看 n 非常大的时候会发生什么.

例 4.3.4 求函数 $y=\cos x$ 在点 $x=0$ 处的泰勒公式.

证明. 因 $(\cos x)'=-\sin x$, 故当 $n>1$ 时, 根据上一道例题的结果有

$$(\cos x)^{(n)}=-(\sin x)^{(n-1)}=-\sin\left(\frac{n-1}{2}\pi+x\right).$$

[13] 请自行证明 $(\sin x)^{(n)}=\sin\left(\frac{n\pi}{2}+x\right)$.

由此可以归纳出

$$(\cos x)^{(2k+1)}\Big|_{x=0} = 0, \quad k = 0, 1, \cdots$$
$$(\cos x)^{(2k)}\Big|_{x=0} = (-1)^k, \quad k = 1, 2, \cdots$$

于是, $y = \cos x$ 在点 $x = 0$ 处的泰勒公式为

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots + \frac{(-1)^k}{(2k)!}x^{2k} + o\left(x^{2k+1}\right), \quad (x \to 0).$$

或者认为展开式结束于偶数项:

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots + \frac{(-1)^k}{(2k)!}x^{2k} + o\left(x^{2k}\right), \quad (x \to 0).$$

简言之, 奇数项消失, 只有偶数项, 符号正负交替. 注意, 0 次幂算是偶数项. 我们发现右侧每一个偶次方幂都是偶函数, 它们的和也是偶函数, 这和 $\cos x$ 的性质一致. □

请在电脑上查看 Taylor polynomial graphs –GeoGebra, 将函数设置为 $f(x) = \cos x$, 并调整近似的阶数 n. 看看 n 非常大的时候会发生什么.

例 4.3.5 函数 $y = (1+x)^\alpha$ (α 为常数) 在点 $x = 0$ 处的泰勒公式为

$$(1+x)^\alpha = 1 + \frac{\alpha}{1!}x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots + \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}x^n + o\left(x^n\right), \quad (x \to 0).$$

证明. 这一公式的验证留给读者. 我们要指出: 这一公式在形式上很像牛顿二项展开式, 但这里 α 不一定是整数, 也不一定是有理数, 它可以是任意实数. 当 α 是正整数 m 时, 如果展开的阶数 n 等于 m, 那么这时上述公式恰好是 $(1+x)^m$ 的牛顿二项展开式, 因此这一公式是牛顿二项展开式的推广. □

例 4.3.6 证明: 函数 $y = \ln(1+x)$ 在点 $x = 0$ 处的泰勒公式为

$$\begin{aligned}\ln(1+x) &= x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n-1}\frac{x^n}{n} + o\left(x^n\right), \quad (x \to 0)\\ &= \sum_{k=1}^{n}(-1)^{k+1}\frac{x^k}{k} + o\left(x^n\right), \quad (x \to 0).\end{aligned}$$

证明. 为了证明这个公式, 我们应求出 $\ln(1+x)$ 在点 $x = 0$ 处的 n 阶导数. 直接套用第二章中的结果 (或用数学归纳法证明)

$$(\ln(1+x))^{(n)} = (-1)^{n-1}\frac{(n-1)!}{(1+x)^n},$$

并求出它在点 $x = 0$ 处的值:

$$(\ln(1+x))^{(n)}\Big|_{x=0} = (-1)^{n-1}(n-1)!,$$

然后代入一般形式的泰勒公式即得所要证明的公式. 证毕. □

请在电脑上查看 Taylor polynomial graphs –GeoGebra, 将函数设置为 $f(x) = \ln(x+1)$, 并调整近似的阶数 n. 看看 n 非常大的时候会发生什么.

4.3.3 局部泰勒展开式的唯一性定理

设函数 $f(x)$ 在点 $x=x_0$ 处有 n 阶导数. 假定我们已知下列公式成立:

$$\begin{aligned}f(x)=&A_0+A_1(x-x_0)+\cdots+A_n(x-x_0)^n\\&+o((x-x_0)^n)\quad(x\to x_0)\end{aligned}$$

其中 $A_0,A_1,\cdots,A_n$ 为常数. 我们要问: 这个公式是否一定是 $f(x)$ 在点 x_0 处的泰勒公式? 也就是问: 这个公式右端的多项式是否一定是泰勒多项式? 问题的答案是肯定的.

定理 4.3.2: 局部泰勒展开式的唯一性定理

设函数 $f(x)$ 在点 x_0 附近有定义, 并且在点 x_0 处的 n 阶导数存在. 假如有 $n+1$ 个常数 $A_0,A_1,\cdots,A_n$, 使得下式成立:

$$f(x)=A_0+A_1(x-x_0)+\cdots+A_n(x-x_0)^n+o((x-x_0)^n),\quad(x\to x_0),$$

则一定有

$$A_k=\frac{f^{(k)}(x_0)}{k!},\quad k=0,1,\cdots,n,$$

其中 $f^{(0)}(x_0)\equiv f(x_0)$.

证明. 已知极限等式

$$\begin{aligned}f(x)=&A_0+A_1(x-x_0)+A_2(x-x_0)^2+A_3(x-x_0)^3+\cdots+A_n(x-x_0)^n\\&+o((x-x_0)^n),\quad(x\to x_0),\end{aligned}$$

成立. 我们下面依次去证明 $A_k=\dfrac{f^{(k)}(x_0)}{k!},\quad k=0,1,\cdots,n.$

1. 确定 A_0. 当 $x\to x_0$ 时, 直接对上式取极限即得

$$A_0=\lim_{x\to x_0}f(x)=f(x_0)\equiv f^{(0)}(x_0).$$

2. 确定 A_1. 当 $x\neq x_0$ 时, 考虑

$$\frac{f(x)-A_0}{x-x_0}=A_1+A_2(x-x_0)+\cdots+A_n(x-x_0)^{n-1}+o\left((x-x_0)^{n-1}\right),\quad(x\to x_0).$$

   当 $x\to x_0$ 时, 对上式取极限即得

$$A_1=\lim_{x\to x_0}\frac{f(x)-A_0}{x-x_0}=\lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0}=f'(x_0).$$

3. 确定 A_2. 再次当 $x\neq x_0$ 时, 考虑

$$\frac{f(x)-A_0-A_1(x-x_0)}{(x-x_0)^2}=A_2+A_3(x-x_0)+\cdots+A_n(x-x_0)^{n-2}+o\left((x-x_0)^{n-2}\right),\quad(x\to x_0).$$

当 $x \to x_0$ 时, 对上式取极限得

$$\begin{aligned}
A_2 &= \lim_{x \to x_0} \frac{f(x) - A_0 - A_1(x - x_0)}{(x - x_0)^2} \\
&= \lim_{x \to x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{(x - x_0)^2} \\
&= \lim_{x \to x_0} \frac{f'(x) - f'(x_0)}{2(x - x_0)} \quad \left(\frac{0}{0} \text{ 型, 应用洛必达法则}\right) \\
&= \frac{1}{2} f''(x_0).
\end{aligned}$$

4. 依次类推, 最后可以通过 $n-1$ 次洛必达法则证明 $A_n = \dfrac{1}{n!} f^{(n)}(x_0)$.

证毕. □

局部泰勒展开式的唯一性定理 4.3.2 非常非常有用! 这就是说, 不管你是用怎样的办法证明了 $f(x)$ 可以用 $x - x_0$ 的某个 n 次多项式逼近, 其误差当 $x \to x_0$ 时是比 $(x - x_0)^n$ 高阶的无穷小量, 那么这个 $x - x_0$ 的 n 次多项式加上 $o((x - x_0)^n)$ 就一定是 $f(x)$ 在点 x_0 处的泰勒公式. 简而言之, 但凡我们能写出 (不管你用了任何方法野路子):

$$f(x) = \sum_{k=0}^{n} [\text{某些常数} \times (x - x_0) \text{ 的 } k \text{ 次方}] + o((x - x_0)^n),$$

那么这些常数肯定等于相应的 $\dfrac{k \text{ 阶导}}{k \text{ 阶乘}}$, 因而这个表达式是且仅是 $f(x)$ 在点 x_0 处的泰勒展开. 特别当 $x_0 = 0$ 时, 但凡我们能写出

$$f(x) = \sum_{k=0}^{n} [\text{某些常数} \times x \text{ 的 } k \text{ 次方}] + o(x^n),$$

那么这个表达式肯定是 $f(x)$ 在点 $x_0 = 0$ 处的泰勒展开. 有了这一结论, 在求某个函数的局部泰勒公式时, 我们就不一定要通过求函数在某点处的各阶导数来求, 而可以用其他变通的方法了.

例 4.3.7 求函数 $y = \mathrm{e}^{-x^2}$ 在点 $x = 0$ 处的泰勒公式.

证明. 我们已知 $y = \mathrm{e}^x$ 在点 $x = 0$ 处的泰勒公式, 将 $-x^2$ 替代 x 代人 e^x 在点 $x = 0$ 处的泰勒公式中即得

$$\begin{aligned}
\mathrm{e}^{-x^2} &= 1 + (-x^2) + \frac{1}{2!}(-x^2)^2 + \cdots + \frac{1}{n!}(-x^2)^n + o\left(\left(x^2\right)^n\right), \quad (x^2 \to 0) \\
&= 1 - x^2 + \frac{1}{2!}x^4 - \cdots + \frac{(-1)^n}{n!}x^{2n} + o\left(x^{2n}\right), \quad (x \to 0).
\end{aligned}$$

那么根据局部泰勒展开式的唯一性定理 4.3.2, 这个结果就是函数 $y = \mathrm{e}^{-x^2}$ 在点 $x = 0$ 处的泰勒展开. □

局部泰勒公式为求 $\dfrac{0}{0}$ 型未定式极限提供了方便, 使我们不必使用洛必达法则而求得未定式极限.

例 4.3.8 求极限

$$\lim_{x \to 0} \frac{\mathrm{e}^x - 1 - x - \frac{x}{2}\sin x}{\sin x - x\cos x}.$$

证明. 粗略地估计知此式分子与分母都是 x 的三阶无穷小量[14]. 于是, 我们首先将分子与分母展开至 x^3 项. 以下等式都是默认在取极限 $x \to 0$ 时成立, 为了方便起见, 我们没有明确标记, 但读者心中要清楚, 后面我们经常这样子 "偷懒". 对于分子, 对 e^x 和 $\sin x$ 展开至 x^3 项, 有

$$\begin{aligned}(\mathrm{e}^x) - 1 - x - \frac{x}{2}(\sin x) &= \left(1 + x + \frac{x^2}{2} + \frac{x^3}{6} + o\left(x^3\right)\right) - 1 - x - \frac{x}{2}\left(x - \frac{1}{6}x^3 + o\left(x^3\right)\right) \\ &= \frac{1}{6}x^3 + \frac{1}{12}x^4 + o\left(x^4\right) + o\left(x^3\right) \\ &= \frac{1}{6}x^3 + o\left(x^3\right).\end{aligned}$$

对于分母, 对 $\sin x$ 和 $\cos x$ 展开至 x^3 项, 有

$$\begin{aligned}\sin x - x\cos x &= \left(x - \frac{1}{6}x^3 + o\left(x^3\right)\right) - x\left(1 - \frac{1}{2}x^2 + o\left(x^3\right)\right) \\ &= x - \frac{1}{6}x^3 + o\left(x^3\right) - x + \frac{1}{2}x^3 + o\left(x^4\right) \\ &= \frac{1}{3}x^3 + o\left(x^3\right).\end{aligned}$$

所以, 我们有

$$\frac{\mathrm{e}^x - 1 - x - \frac{x}{2}\sin x}{\sin x - x\cos x} = \frac{\frac{1}{6}x^3 + o\left(x^3\right)}{\frac{1}{3}x^3 + o\left(x^3\right)}.$$

从而得到

$$\lim_{x\to 0}\frac{\mathrm{e}^x - 1 - x - \frac{x}{2}\sin x}{\sin x - x\cos x} = \frac{1}{2}.$$

在这个题目中, 我们用到了下述事实:

$$xo\left(x^3\right) = o\left(x^4\right) \text{ 与} o\left(x^4\right) + o\left(x^3\right) = o\left(x^3\right).$$

请读者根据高阶无穷小量的定义自己证明, 并希望读者能熟练运用类似的关系式. □

例 4.3.9 求极限 $\lim\limits_{x\to 0}\dfrac{\sin x - x\cos x}{\sin^3 x}$.

证明. 我们将分子分母中的 $\sin x, \cos x$ 都展开成麦克劳林公式, 只需要展开到三阶就可以了, 因为我们知道 $\sin^3 x$ 与 x^3 差不多. 讲

$$\sin x = x - \frac{x^3}{3!} + o\left(x^3\right), \quad \cos x = 1 - \frac{x^2}{2!} + o\left(x^2\right)$$

将展开后的 $\sin x$ 和 $\cos x$ 代入到分子得到

$$\begin{aligned}\sin x - x\cos x &= \left(x - \frac{x^3}{6} + o\left(x^3\right)\right) - x\left(1 - \frac{x^2}{2} + o\left(x^2\right)\right) \\ &= x - \frac{x^3}{6} + o\left(x^3\right) - x + \frac{x^3}{2} + o\left(x^3\right) \\ &= \frac{x^3}{3} + o\left(x^3\right).\end{aligned}$$

[14]以下定义在介绍同阶无穷小时提到过, 请复习. 若 $f(x)$ 是无穷小量, a 是常数, 如果 $\lim\limits_{x\to a}\dfrac{f(x)}{(x-a)^n} = l \neq 0$, 则称 $f(x)$ 为 $x - a$ 的 n 阶无穷小量. 特别地, 如果 $\lim\limits_{x\to a}\dfrac{f(x)}{x^n} = l \neq 0$, 则称 $f(x)$ 为 x 的 n 阶无穷小量.

现在, 利用等价无穷小替换并将分子代入原式, 我们得到

$$\begin{aligned}\lim_{x\to 0}\frac{\sin x - x\cos x}{\sin^3 x} &= \lim_{x\to 0}\frac{\sin x - x\cos x}{x^3}\cdot\frac{x^3}{\sin^3 x}\\ &= \lim_{x\to 0}\frac{\frac{1}{3}x^3 + o\left(x^3\right)}{x^3}\\ &= \frac{1}{3}.\end{aligned}$$

请读者用洛必达法则计算本题. □

问题 之后做练习题是, 多多思考一下, 洛必达法则和泰勒展开在处理未定型的异同点.

注释 4.3.2 泰勒公式在函数极限计算中的方法探讨-维普期刊中文期刊服务平台 这篇文章总结了用泰勒公式计算极限的两类题型. 请自行下载并认真阅读.

局部泰勒公式也可以用来处理其他类型的未定式.

例 4.3.10 设 m 为整数且 $m>1$, 求极限

$$\lim_{x\to\infty}\left[\left(x^m + x^{m-1}\right)^{\frac{1}{m}} - \left(x^m - x^{m-1}\right)^{\frac{1}{m}}\right].$$

证明. 我们先处理做差的第一项. 观察到

$$\left(x^m + x^{m-1}\right)^{\frac{1}{m}} = \left(x^m\left(1+x^{-1}\right)\right)^{\frac{1}{m}} = x\left(1+\frac{1}{x}\right)^{\frac{1}{m}}.$$

因为当 $x\to\infty$ 时, $\frac{1}{x}\to 0$. 我们可以把 $\frac{1}{x}$ 整体看作一个趋于 0 的量. 当 $x\to 0$, 由 $(1+x)^{\alpha}$ 的局部泰勒公式, 把公式中的 x 替换成 $\frac{1}{x}$, 且 $\alpha=\frac{1}{m}$, 我们有

$$\left(1+\frac{1}{x}\right)^{\frac{1}{m}} = 1+\frac{1}{m}\cdot\frac{1}{x}+o\left(\frac{1}{x}\right),\quad (x\to\infty).$$

因此

$$x\left(1+\frac{1}{x}\right)^{\frac{1}{m}} = x+\frac{1}{m}+o(1)\quad (x\to\infty),$$

其中 $o(1)$ 代表一个无穷小量 $(x\to\infty)$. 这里我们用到 $xo\left(\frac{1}{x}\right)=o(1)$, 这一点读者可以由高阶无穷小量的定义验证.

回到题目, 看做差的第二项. 类似地, 我们有

$$\left(x^m - x^{m-1}\right)^{\frac{1}{m}} = x\left(1-\frac{1}{x}\right)^{\frac{1}{m}} = x-\frac{1}{m}+o(1)\quad (x\to\infty).$$

这样, 最后得到

$$\lim_{x\to\infty}\left[\left(x^m + x^{m-1}\right)^{\frac{1}{m}} - \left(x^m - x^{m-1}\right)^{\frac{1}{m}}\right] = \frac{2}{m}$$

上式我们用到 $o(1)-o(1)=o(1)$. 这是因为, 两个无穷小量之差是一个无穷小量. □

问题 思考题. 求极限

$$\lim_{n\to\infty} n\left[\left(1+\frac{1}{n}\right)^n - \mathrm{e}\right].$$

提示: $\left(1+\dfrac{1}{n}\right)^n = \mathrm{e}^{n\ln\left(1+\frac{1}{n}\right)}$.

无论是求 $\dfrac{0}{0}$ 型未定式极限还是估计一个无穷小量的阶数, 都需要熟记常见的基本初等函数的局部泰勒公式, 尤其是下列一些公式. 详细情况请参考小节 4.3.4.

结论 4.3.1: 常见的基本初等函数的局部泰勒公式

1. $\mathrm{e}^x = 1 + x + \dfrac{1}{2!}x^2 + \cdots + \dfrac{1}{n!}x^n + o\left(x^n\right), \quad (x\to 0).$
2. $\sin x = x - \dfrac{1}{3!}x^3 + \cdots + (-1)^{n-1}\dfrac{x^{2n-1}}{(2n-1)!} + o\left(x^{2n}\right), \quad (x\to 0).$
3. $\cos x = 1 - \dfrac{1}{2!}x^2 + \cdots + (-1)^n\dfrac{x^{2n}}{(2n)!} + o\left(x^{2n+1}\right), \quad (x\to 0).$
4. $(1+x)^\alpha = 1 + \alpha x + \dfrac{\alpha(\alpha-1)}{2!}x^2 + \cdots + \dfrac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}x^n + o\left(x^n\right), \quad (x\to 0).$
5. $\ln(1+x) = x - \dfrac{x^2}{2} + \dfrac{x^3}{3} - \cdots + (-1)^{n-1}\dfrac{x^n}{n} + o\left(x^n\right), \quad (x\to 0).$

一般来说, 在点 $x=0$ 处的泰勒公式中, "o" 中 x 的方幂次数与泰勒多项式的次数相同. 但对于 $\sin x$ 及 $\cos x$, 在点 $x=0$ 处的泰勒公式中, "o" 中 x 的方幂次数却比泰勒多项式的次数高一次 (详细情况请参考小节 4.3.4). 这是因为, 在这两个泰勒多项式中, 与 "o" 中 x 同方幂的系数为零的缘故.

例 4.3.11 求函数 $f(x) = \cos 2x \cdot \ln(1+x)$ 在点 $x=0$ 处的泰勒公式, 至 x^4 项.

证明. 理论上, 把两部分 $\cos 2x$ 和 $\ln(1+x)$ 分别展开至 x^4 项, 然后乘起来就可以了. 比如, $\ln(1+x)$ 展开至 x^4 项:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + o\left(x^4\right).$$

$\cos(2x)$ 展开至 x^4 项:

$$\begin{aligned}\cos(2x) &= 1 - \frac{(2x)^2}{2!} + \frac{(2x)^4}{4!} + o\left((2x)^4\right)\\ &= 1 - \frac{(2x)^2}{2!} + \frac{(2x)^4}{4!} + o\left(x^4\right).\end{aligned}$$

但是我们发现, 若 $\cos(2x)$ 展开至 x^3 项:

$$\begin{aligned}\cos(2x) &= 1 - \frac{(2x)^2}{2!} + o\left((2x)^3\right)\\ &= 1 - \frac{(2x)^2}{2!} + o\left(x^3\right).\end{aligned}$$

用 $(\cos(2x)$ 展开至x^3 项$)\times(\ln(1+x)$ 展开至x^4 项$)$ 就足够推导出结果了, 这是因为把 $o\left(x^3\right)\times(\ln(1+x)$ 展开至x^4 项$)$ 相乘展开后, 每一项都是不小于 4 阶的项, 即

$$o\left(x^3\right)\times\left(x-\frac{1}{2}x^2+\frac{1}{3}x^3-\frac{1}{4}x^4+o\left(x^4\right)\right).$$

这样子就足够了. 于是乎我们简化了问题. 具体操作如下,

$$\begin{aligned}
f(x)=&\underbrace{\left(1-\frac{1}{2}(2x)^2+o\left(x^3\right)\right)}_{(1)}\times\underbrace{\left(x-\frac{1}{2}x^2+\frac{1}{3}x^3-\frac{1}{4}x^4+o\left(x^4\right)\right)}_{(2)}\\
=&\left(x-\frac{1}{2}x^2+\frac{1}{3}x^3-\frac{1}{4}x^4+o\left(x^4\right)\right)\quad((1)\text{ 的 }1\text{ 乘以 }(2))\\
&-\frac{1}{2}(2x)^2\left(x-\frac{1}{2}x^2+o\left(x^2\right)\right)\quad((1)\text{ 的 }-\frac{1}{2}(2x)^2\text{ 乘以 }(2);(2)\text{ 中大于 2 次的项不要})\\
&+o\left(x^3\right)(x+o(x))\quad((1)\text{ 的 }o\left(x^3\right)\text{ 乘以 }(2);\ (2)\text{ 中大于 1 次的项不要})\\
=&x-\frac{1}{2}x^2+\frac{1}{3}x^3-\frac{1}{4}x^4+o\left(x^4\right)\\
&-\frac{1}{2}(2x)^2x+\frac{1}{2}(2x)^2\cdot\frac{1}{2}x^2-\frac{1}{2}(2x)^2o\left(x^4\right)\\
&+o\left(x^4\right)\\
=&x-\frac{1}{2}x^2+\frac{1}{3}x^3-\frac{1}{4}x^4\quad(\text{可以直接把多余的 }o\left(x^4\right)\text{ 丢掉})\\
&-\frac{1}{2}(2x)^2x+\frac{1}{2}(2x)^2\cdot\frac{1}{2}x^2\quad(\text{可以直接把 }-\frac{1}{2}(2x)^2o\left(x^4\right)\text{ 丢掉})\\
&+o\left(x^4\right)\\
=&x-\frac{1}{2}x^2-\frac{5}{3}x^3+\frac{3}{4}x^4+o\left(x^4\right).
\end{aligned}$$

我们再看一下, 用 $(\cos(2x)$ 展开至x 项$)\times(\ln(1+x)$ 展开至x^4 项$)$ 会发生什么? 其中, $\cos(2x)$ 展开至 x 项有

$$\cos(2x)=1+o((2x))=1+o(x).$$

于是,

$$f(x)=(1+o(x))\times\left(x-\frac{1}{2}x^2+\frac{1}{3}x^3-\frac{1}{4}x^4+o\left(x^4\right)\right).$$

上式拆开后, 一定会出现 $o(x)$ 项, 这一项是不能被 $o\left(x^4\right)$, 反而 $o\left(x^4\right)$ 会被 $o(x)$ 吸收. 于是最终的精度只是 2 阶的, 而非 4 阶的. 类似地, 用 $(\cos(2x)$ 展开至x^2 项$)\times(\ln(1+x)$ 展开至x^4 项$)$ 也是一样不能计算到 4 阶. □

局部泰勒公式还可以用来求复杂函数的高阶导数.

例 4.3.12 设函数 $f(x)=\mathrm{e}^{\cos x}$, 求 $f^{(4)}(0)$.

证明. $f(x)$ 可化为 $\mathrm{e}^{\cos x}=\mathrm{e}\cdot\mathrm{e}^{\cos x-1}$. 当 $x\to0$ 时, $\cos x-1\to0$. 先利用 $\cos x$ 在点 $x=0$ 处的泰勒公式得到

$$\cos x-1=-\frac{1}{2}x^2+\frac{1}{4!}x^4+o\left(x^4\right),$$

再代入 e^x 的局部泰勒公式得到

$$\begin{aligned}
\mathrm{e}^{\cos x} &= \mathrm{e}\cdot\mathrm{e}^{-\frac{1}{2}x^2+\frac{1}{4!}x^4+o\left(x^4\right)}\\
&= \mathrm{e}\left[1+\left(-\frac{1}{2}x^2+\frac{1}{4!}x^4+o\left(x^4\right)\right)+\frac{1}{2}\left(-\frac{1}{2}x^2+o\left(x^2\right)\right)^2+o\left(x^4\right)\right]\\
&= \mathrm{e}\left[1+\left(-\frac{1}{2}x^2+\frac{1}{4!}x^4+o\left(x^4\right)\right)+\frac{1}{2}\left(\frac{1}{4}x^4+o\left(x^4\right)\right)+o\left(x^4\right)\right]\\
&= \mathrm{e}\left[1-\frac{1}{2}x^2+\frac{1}{4!}x^4+\frac{1}{2}\cdot\left(\frac{1}{4}x^4\right)+o\left(x^4\right)\right]\\
&= \mathrm{e}-\frac{\mathrm{e}}{2}x^2+\frac{\mathrm{e}}{6}x^4+o\left(x^4\right).
\end{aligned}$$

根据局部泰勒展开式的唯一性, 有 $\dfrac{f^{(4)}(0)}{4!}=\dfrac{\mathrm{e}}{6}$ (对应其 x^4 的系数). 由此得到 $f^{(4)}(0)=4\mathrm{e}$. □

4.3.4　常用的麦克劳林展开的前几阶

读者需要非常熟悉以下几个函数的麦克劳林展开的前几阶. 要求是非常熟悉, 直接就能背写.

1. 指数函数 e^x

泰勒展开为: 整数 $n\geq 0$,

$$\mathrm{e}^x=1+x+\frac{1}{2!}x^2+\cdots+\frac{1}{n!}x^n+o\left(x^n\right),\quad (x\to 0).$$

具体展开到各阶:

- 0 阶: $\mathrm{e}^x=1+o(1)$
- 1 阶: $\mathrm{e}^x=1+x+o(x)$
- 2 阶: $\mathrm{e}^x=1+x+\dfrac{x^2}{2!}+o(x^2)$
- 3 阶: $\mathrm{e}^x=1+x+\dfrac{x^2}{2!}+\dfrac{x^3}{3!}+o(x^3)$
- 4 阶: $\mathrm{e}^x=1+x+\dfrac{x^2}{2!}+\dfrac{x^3}{3!}+\dfrac{x^4}{4!}+o(x^4)$
- 5 阶: $\mathrm{e}^x=1+x+\dfrac{x^2}{2!}+\dfrac{x^3}{3!}+\dfrac{x^4}{4!}+\dfrac{x^5}{5!}+o(x^5)$

2. 正弦函数 $\sin x$

泰勒展开为: 整数 $n\geq 0$,

$$\sin x=x-\frac{1}{3!}x^3+\cdots+(-1)^{n-1}\frac{x^{2n-1}}{(2n-1)!}+o\left(x^{2n}\right),\quad (x\to 0).$$

具体展开到各阶: (注意上述一般公式只能对应下面加粗显示的阶数)

- **0 阶**: $\sin x=o(1)$
- 1 阶: $\sin x=x+o(x)$

- **2 阶**: $\sin x = x + o(x^2)$
- 3 阶: $\sin x = x - \dfrac{x^3}{3!} + o(x^3)$
- **4 阶**: $\sin x = x - \dfrac{x^3}{3!} + o(x^4)$
- 5 阶: $\sin x = x - \dfrac{x^3}{3!} + \dfrac{x^5}{5!} + o(x^5)$
- **6 阶**: $\sin x = x - \dfrac{x^3}{3!} + \dfrac{x^5}{5!} + o(x^6)$
- 7 阶: $\sin x = x - \dfrac{x^3}{3!} + \dfrac{x^5}{5!} - \dfrac{x^7}{7!} + o(x^7)$
- **8 阶**: $\sin x = x - \dfrac{x^3}{3!} + \dfrac{x^5}{5!} - \dfrac{x^7}{7!} + o(x^8)$

可以认为, $\sin x$ 的 $x=0$ 处的展开, 是 1-2, 3-4, 5-6, 7-8, ...为一对, 通常 “o” 中 x 的方幂次数, 我们取阶数更大的后者.

3. 余弦函数 $\cos x$

泰勒展开为: 整数 $n \geq 0$,

$$\cos x = 1 - \frac{1}{2!}x^2 + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + o\left(x^{2n+1}\right), \quad (x \to 0).$$

具体展开到各阶: (注意上述一般公式只能对应下面加粗显示的阶数)

- 0 阶: $\cos x = 1 + o(1)$
- **1 阶**: $\cos x = 1 + o(x)$
- 2 阶: $\cos x = 1 - \dfrac{x^2}{2!} + o(x^2)$
- **3 阶**: $\cos x = 1 - \dfrac{x^2}{2!} + o(x^3)$
- 4 阶: $\cos x = 1 - \dfrac{x^2}{2!} + \dfrac{x^4}{4!} + o(x^4)$
- **5 阶**: $\cos x = 1 - \dfrac{x^2}{2!} + \dfrac{x^4}{4!} + o(x^5)$
- 6 阶: $\cos x = 1 - \dfrac{x^2}{2!} + \dfrac{x^4}{4!} - \dfrac{x^6}{6!} + o(x^6)$
- **7 阶**: $\cos x = 1 - \dfrac{x^2}{2!} + \dfrac{x^4}{4!} - \dfrac{x^6}{6!} + o(x^7)$

可以认为, $\cos x$ 的 $x=0$ 处的展开, 是 0-1, 2-3, 4-5, 6-7, ...为一对, 通常 “o” 中 x 的方幂次数, 我们取阶数更大的后者.

4. 幂函数 $(1+x)^\alpha$

泰勒展开为:

$$(1+x)^\alpha = 1+\alpha x+\frac{\alpha(\alpha-1)}{2!}x^2+\cdots+\frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}x^n+o\left(x^n\right), \quad (x\to 0).$$

具体展开到各阶:

- 0 阶: $(1+x)^\alpha = 1+o(1)$
- 1 阶: $(1+x)^\alpha = 1+\alpha x+o(x)$
- 2 阶: $(1+x)^\alpha = 1+\alpha x+\frac{\alpha(\alpha-1)}{2!}x^2+o(x^2)$
- 3 阶: $(1+x)^\alpha = 1+\alpha x+\frac{\alpha(\alpha-1)}{2!}x^2+\frac{\alpha(\alpha-1)(\alpha-2)}{3!}x^3+o(x^3)$
- 4 阶: $(1+x)^\alpha = 1+\alpha x+\frac{\alpha(\alpha-1)}{2!}x^2+\frac{\alpha(\alpha-1)(\alpha-2)}{3!}x^3+\frac{\alpha(\alpha-1)(\alpha-2)(\alpha-3)}{4!}x^4+o(x^4)$

5. 对数函数 $\ln(1+x)$

泰勒展开为:

$$\ln(1+x) = x-\frac{x^2}{2}+\frac{x^3}{3}-\cdots+(-1)^{n-1}\frac{x^n}{n}+o\left(x^n\right), \quad (x\to 0).$$

具体展开到各阶:

- 0 阶: $\ln(1+x) = 0+o(1)$
- 1 阶: $\ln(1+x) = x+o(x)$
- 2 阶: $\ln(1+x) = x-\frac{x^2}{2}+o(x^2)$
- 3 阶: $\ln(1+x) = x-\frac{x^2}{2}+\frac{x^3}{3}+o(x^3)$
- 4 阶: $\ln(1+x) = x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+o(x^4)$
- 5 阶: $\ln(1+x) = x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\frac{x^5}{5}+o(x^5)$

4.3.5 再次认识符号 $o(\cdot)$

以下内容摘自 Calculus Apostol book[15] 的 P288. 作为理解小 o 的阅读材料.

DEFINITION. Assume $g(x)\neq 0$ for all $x\neq a$ in some interval containing a. The notation

$$f(x)=o(g(x)) \quad \text{as} \quad x\to a$$

means that

$$\lim_{x\to a}\frac{f(x)}{g(x)}=0.$$

[15]另外参考: limits - Property of the little-o notation in Calculus Apostol - Mathematics Stack Exchange; 无穷小与高阶无穷小的解读——函数与函数族 - 知乎

The symbol $f(x)=o(g(x))$ is read "$f(x)$ is little-oh of $g(x)$".

EXAMPLE 1. $f(x)=o(1)$ as $x\to a$ means that $f(x)\to 0$ as $x\to a$.

EXAMPLE 2. $f(x)=o(x)$ as $x\to 0$ means that $\dfrac{f(x)}{x}\to 0$ as $x\to 0$.

An equation of the form $f(x)=h(x)+o(g(x))$ is understood to mean that $f(x)-h(x)=o(g(x))$ or, in other words, $[f(x)-h(x)]/g(x)\to 0$ as $x\to a$.

EXAMPLE 3. We have $\sin x=x+o(x)$ because $\dfrac{\sin x-x}{x}=\dfrac{\sin x}{x}-1\to 0$ as $x\to 0$.

In calculations involving Taylor approximations, it often becomes necessary to combine several terms involving the o-symbol. A few simple rules for manipulating o-symbols are discussed in the next theorem. These cover most situations that arise in practice.

THEOREM 7.8. Algebra of o-symbols. As $x\to a$, we have the following:

(a) $o(g(x))\pm o(g(x))=o(g(x))$.

(b) $o(cg(x))=o(g(x))$ if $c\neq 0$.

(c) $f(x)\cdot o(g(x))=o(f(x)g(x))$.

(d) $o(o(g(x)))=o(g(x))$.

(e) $\dfrac{1}{1+g(x)}=1-g(x)+o(g(x))\quad$ if $g(x)\to 0\quad$ as $x\to a$.

Proof. The statement in part (a) is understood to mean that if $f_1(x)=o(g(x))$ and if $f_2(x)=o(g(x))$, then $f_1(x)\pm f_2(x)=o(g(x))$. But since we have

$$\frac{f_1(x)\pm f_2(x)}{g(x)}=\frac{f_1(x)}{g(x)}\pm\frac{f_2(x)}{g(x)},$$

each term on the right tends to 0 as $x>a$, so part (a) is proved. The statements in (b), (c), and (d) are proved in a similar way. To prove (e), we use the algebraic identity

$$\frac{1}{1+u}=1-u+u\frac{u}{1+u}$$

with u replaced by $g(x)$ and then note that $\dfrac{g(x)}{1+g(x)}\to 0$ as $x\to a$.

EXAMPLE 1. Prove that $\tan x=x+\dfrac{1}{3}x^3+o\left(x^3\right)$ as $x\to 0$.

Solution. We use the Taylor approximations for the sine and cosine. From part (e) of Theorem 7.8, with $g(x)=-\dfrac{1}{2}x^2+o\left(x^3\right)$, we have

$$\frac{1}{\cos x}=\frac{1}{1-\frac{1}{2}x^2+o\left(x^3\right)}=1+\frac{1}{2}x^2+o\left(x^2\right)\quad\text{as}\quad x\to 0.$$

Therefore, we have

$$\tan x=\frac{\sin x}{\cos x}=\left(x-\frac{1}{6}x^3+o\left(x^4\right)\right)\left(1+\frac{1}{2}x^2+o\left(x^2\right)\right)=x+\frac{1}{3}x^3+o\left(x^3\right).$$

4.4 带拉格朗日余项的泰勒公式

以下我们考虑的函数 $f(x)$ 都默认是在 x_0 附近有足够阶导数定义的. 上一节所证明的泰勒公式为

$$\begin{aligned}f(x)=&f\left(x_0\right)+\frac{f'\left(x_0\right)}{1!}\left(x-x_0\right)+\cdots+\frac{f^{(n)}\left(x_0\right)}{n!}\left(x-x_0\right)^n\\&+o\left(\left(x-x_0\right)^n\right),\quad\left(x\to x_0\right).\end{aligned}$$

这个局部泰勒公式的最大意义, 在于如何用一个 (n 阶) 多项式函数 (即泰勒多项式):

$$T_n(x)=f(x_0)+\frac{f'(x_0)}{1!}(x-x_0)+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$$

来逼近或近似原本的函数 $f(x)$ 在 x_0 附近处的表现. 我们知道随着阶数 n 越来越大, 这个 (n 阶) 多项式函数 $T_n(x)$ 的近似效果也就越来越好. 其中, 原本的函数 $f(x)$ 和用于近似的 $T_n(x)$ 之间的误差项是使用皮亚诺余项来表示的:

$$o\left((x-x_0)^n\right),\quad (x\to x_0).$$

但是, 这个余项还是太泛泛了, 它只告诉我们: 它是当 $x\to x_0$ 时比 $(x-x_0)^n$ 高阶的无穷小量, 但并没有具体地告诉我们, 对于任意给定的 x (当然 x_0 是早就固定好的, 这里我们考虑的 x 也都是在 x_0 的附近处取值的变量), 它的数值有多大, 或者说这个误差的上界是多少. 因此, 我们希望对于泰勒公式中的误差项有一个更加具体的估计式或更为明显的表达式. 这就是本节的目的.

还是先回到一般的泰勒公式中, 在阶数 $n=0$ 时, 泰勒公式化简为:

$$f(x)=f(x_0)+o(1),\quad (x\to x_0).$$

其中, $o(1)$ 表示一个当 $x\to x_0$ 时的无穷小量. 此时, 我们用了一个 0 阶多项式, 即常数函数 $y=f(x_0)$, 来近似 $y=f(x)$. 此时的皮亚诺余项就是 $o(1)$. 它几乎没有给我们任何有用的信息, 毕竟我们总有 $f(x)-f(x_0)=o(1)$, 即 $\lim\limits_{x\to x_0}f(x)=f(x_0)$. 另一方面, 拉格朗日中值定理告诉我们:

$$f(x)=f(x_0)+f'(\xi)(x-x_0),$$

其中 ξ 是介于 x_0 与 x 之间的某一点. 尽管我们可能不清楚 ξ 的具体取值, 但是上式是一个普通的等式, 而不是极限等式了. 仔细对比拉格朗日中值定理和 $n=0$ 时的泰勒公式, 我们发现这个公式可以看作在 $n=0$ 时对皮亚诺余项有明显表达式的一个泰勒公式, 这是因为

$$f'(\xi)(x-x_0)=o(1),\quad (x\to x_0),$$

故 $f'(\xi)(x-x_0)$[16] 就是皮亚诺余项 $o(1)$ 有明显表达式的一个特例. 从这一种特殊情况出发, 我们似乎有理由猜想如下结果可能成立.

- 考虑阶数 $n=1$ 时的泰勒公式:

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+o(x-x_0),\quad (x\to x_0).$$

  我们可能有类似的中值定理结果:

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{1}{2}f''(\xi)(x-x_0)^2,$$

  其中 ξ 是介于 x_0 与 x 之间的某一点. 显然,

$$\frac{1}{2}f''(\xi)(x-x_0)^2=o(x-x_0),\quad (x\to x_0).$$

- 考虑阶数 $n=2$ 时的泰勒公式:

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{1}{2}f''(x_0)(x-x_0)^2+o\left((x-x_0)^2\right),\quad (x\to x_0).$$

[16]注意, ξ 是介于 x_0 与 x 之间的某一点, 是依赖于 x 的, 但当 $x\to x_0$ 是, $\xi\to x_0$. 所以 $\lim\limits_{x\to x_0}f'(\xi)(x-x_0)=f'(x_0)(x_0-x_0)=0$.

我们可能有类似的中值定理结果:

$$f(x)=f\left(x_0\right)+f'\left(x_0\right)\left(x-x_0\right)+\frac{1}{2}f''\left(x_0\right)\left(x-x_0\right)^2+\frac{1}{3!}f'''(\xi)\left(x-x_0\right)^3,$$

其中 ξ 是介于 x_0 与 x 之间的某一点. 显然,

$$\frac{1}{3!}f'''(\xi)\left(x-x_0\right)^3=o\left(\left(x-x_0\right)^2\right),\quad\left(x\to x_0\right).$$

- 对于任意阶数 n 的泰勒公式, 我们可能有类似的中值定理结果:

$$f(x)=f\left(x_0\right)+\frac{f'\left(x_0\right)}{1!}\left(x-x_0\right)+\cdots+\frac{f^{(n)}\left(x_0\right)}{n!}\left(x-x_0\right)^n+\frac{f^{(n+1)}(\xi)}{(n+1)!}\left(x-x_0\right)^{n+1},$$

其中 ξ 是介于 x_0 与 x 之间的某一点. 显然,

$$\frac{f^{(n+1)}(\xi)}{(n+1)!}\left(x-x_0\right)^{n+1}=o\left(\left(x-x_0\right)^n\right),\quad\left(x\to x_0\right).$$

以上猜想其实都是正确的. 其证明由以下定理给出.

定理 4.4.1: 带拉格朗日余项的泰勒公式, 泰勒中值定理

设函数 $f(x)$ 在区间 (A,B) 内有 $n+1$ 阶导数, 则对于 (A,B) 中任意取定的一点 x_0 及任意的 $x\in(A,B)$, 有等式:

$$f(x)=\underbrace{\sum_{k=0}^{n}\frac{f^{(k)}\left(x_0\right)}{k!}\left(x-x_0\right)^k}_{\text{泰勒多项式 } T_n(x)}+\underbrace{\frac{f^{(n+1)}(\xi)}{(n+1)!}\left(x-x_0\right)^{n+1}}_{\text{拉格朗日余项 } =o\left(\left(x-x_0\right)^n\right)},$$

其中 ξ 是介于 x_0 与 x 之间的某一点. 这里的余项称为**拉格朗日余项**, 而整个公式称为**带拉格朗日余项的泰勒公式**. 这一公式是拉格朗日中值定理的推广.

证明. 假如 $x=x_0$, 则上述公式显然成立. 因此, 不妨设 $x\neq x_0$. 不是一般性, 设 $x_0<x$ ($x<x_0$ 也是类似的). 我们考虑定义在 $t\in[x_0,x]$ 上的两个关于 t 的函数:

$$\begin{aligned}&F(t)=f(t)-T_n(t)=f(t)-f\left(x_0\right)-\frac{f'\left(x_0\right)}{1!}\left(t-x_0\right)-\cdots-\frac{f^{(n)}\left(x_0\right)}{n!}\left(t-x_0\right)^n,\\&G(t)=\left(t-x_0\right)^{n+1}.\end{aligned}$$

很容易验证

$$\begin{aligned}&F\left(x_0\right)=G\left(x_0\right)=0,\\&F^{(k)}\left(x_0\right)=G^{(k)}\left(x_0\right)=0,\quad k=1,2,\cdots,n.\end{aligned}$$

由于 $G(t)$ 的结构, 我们知道对于任意 $k=0,1,2,\cdots,n$, 和任意 $t\neq x_0$, 都有 $G^{(k+1)}(t)\neq 0$. (注意下面柯西中值定理使用条件)

1. 注意到 $F\left(x_0\right)=G\left(x_0\right)=0$. 对 $F(t),G(t)$ 在区间 (x_0,x) 应用柯西中值定理[17]有

$$\frac{F(x)}{G(x)}=\frac{F(x)-F\left(x_0\right)}{G(x)-G\left(x_0\right)}=\frac{F'\left(x_1\right)}{G'\left(x_1\right)},$$

其中 $x_1\in(x_0,x)$.

[17]复习: 定理 1.4: 柯西中值定理设函数 $f(x)$ 与 $g(x)$ 在闭区间 $[a,b]$ 上连续, 在开区间 (a,b) 内可导, 并且恒有 $g'(x)\neq 0$, 则必存在一点 $c\in(a,b)$, 使得 $\dfrac{f(b)-f(a)}{g(b)-g(a)}=\dfrac{f'(c)}{g'(c)}$.

2. 注意到 $F'(x_0)=G'(x_0)=0$. 对 $F'(t), G'(t)$ 在区间 (x_0,x_1) 应用柯西中值定理有

$$\frac{F'(x_1)}{G'(x_1)}=\frac{F'(x_1)-F'(x_0)}{G'(x_1)-G'(x_0)}=\frac{F''(x_2)}{G''(x_2)},$$

其中 $x_2\in(x_0,x_1)\subseteq(x_0,x)$.

3. 如此下去, 总共使用 $n+1$ 次柯西中值定理[18], 最后得到

$$\frac{F(x)}{G(x)}=\frac{F'(x_1)}{G'(x_1)}=\frac{F''(x_2)}{G''(x_2)}=\cdots=\frac{F^{(n+1)}(x_{n+1})}{G^{(n+1)}(x_{n+1})},$$

其中 $x_k\in(x_0,x)$ 对于 $k=1,2,\cdots,n+1$, 且 $x_0<x_{n+1}<x_n<\cdots<x_2<x_1<x$.

另外, 不难看出

$$F^{(n+1)}(t)=f^{(n+1)}(t),\quad G^{(n+1)}(t)=(n+1)!,$$

故我们得到

$$\frac{F(x)}{G(x)}=\frac{F^{(n+1)}(x_{n+1})}{G^{(n+1)}(x_{n+1})}=\frac{f^{(n+1)}(x_{n+1})}{(n+1)!},$$

即

$$F(x)=\frac{f^{(n+1)}(x_{n+1})}{(n+1)!}(x-x_0)^{n+1}.$$

将 x_{n+1} 换成 ξ, 上式就是要证的公式. 证毕. □

定理 4.4.1 的表述倾向于先去固定一点 x_0, 然后再考虑另外一个动点 x. 但是, 定理 4.4.1 中的 x_0 也可以任意取的, 故我们还可以换一种表述方式, 即设函数 $f(x)$ 在区间 (A,B) 内有 $n+1$ 阶导数, 则对于任意的 $x,y\in(A,B)$ 总有等式:

$$f(x)=\sum_{k=0}^{n}\frac{f^{(k)}(y)}{k!}(x-y)^k+\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-y)^{n+1},$$

其中 ξ 是介于 x 与 y 之间的某一点. 特别是以下几种特殊情况比较重要:

1. ($n=0$ 时) 对于任意的 $x,y\in(A,B)$ 总有等式

$$f(x)=f(y)+f'(\xi)(x-y),\quad \exists\xi\in(x,y).$$

这就是拉格朗日 (Lagrange) 中值定理, 即默认的微分中值定理.

2. ($n=1$ 时) 对于任意的 $x,y\in(A,B)$ 总有等式

$$f(x)=f(y)+f'(y)(x-y)+\frac{1}{2}f''(\xi)(x-y)^2,\quad \exists\xi\in(x,y).$$

拉格朗日中值定理自然扩展, 现在就将二阶导数和函数值建立了关系.

对于很多常见的初等函数, 它们通常都是任意阶可导的, 而且定义域是整个实数轴. 那么上面公式 x,y 就可以任意取值了, 这说明了其应用的广泛性. 从这种意义上, 我们才能感受到定理 4.4.1 也叫做泰勒中值定理的原因, 它是将拉格朗日中值定理中的一阶导数扩张到任意阶导数.

[18]我们每次适用的区间范围逐步缩小, 并且呈现嵌套的子集关系, 最大的区间还是 (x_0,x).

定理 4.4.1 的拉格朗日余项为泰勒公式的误差提供了更准确的估计. 回到定理 4.4.1, 那么对于泰勒多项式的误差就是拉格朗日余项:

$$R_n(x) := f(x) - T_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1}, \quad \forall x \in (A,B).$$

其中, ξ 是介于 x_0 与 x 之间的某一点. 那么我们有如下估计式:

$$|R_n(x)| = \frac{\left|f^{(n+1)}(\xi)\right|}{(n+1)!}|x-x_0|^{n+1} \leq \frac{M_x}{(n+1)!}|x-x_0|^{n+1}, \quad \forall x \in (A,B),$$

其中, 令 M_x 是 $\left|f^{(n+1)}(t)\right|$ 在区间 $[x_0,x]$ 或 $[x,x_0]$ 上的一个上界. 假设 $f^{(n+1)}(t)$ 是连续的, 那么根据极值定理, M_x 肯定是存在的. 如果 $f^{(n+1)}(x)$ 在整个闭区间 $[A,B]$ 上是连续的, 则存在 M 是 $\left|f^{(n+1)}(x)\right|$ 在区间 $[A,B]$ 的一个上界. 此时, 若进一步限定 $|x-x_0|<1$, 我们可以得到一个和具体的 x 无关的误差上界[19]:

$$|R_n(x)| < \frac{M}{(n+1)!}, \quad \forall x \in (x_0-1, x_0+1).$$

这时, 只要我们选择的 n 够大, 那么我们可以获得任意小的误差, 即任意高精度的近似. 重要的是, 这个上界的分母有阶乘 $(n+1)!$, 其变小的速度非常快 (比指数快), 远比常数分子 M 大很多, 见下表.

n	$\frac{1}{n!}$	n	$\frac{1}{n!}$
1	1.000000×10^{0}	2	5.000000×10^{-1}
3	1.666667×10^{-1}	4	4.166667×10^{-2}
5	8.333333×10^{-3}	6	1.388889×10^{-3}
7	1.984127×10^{-4}	8	2.480158×10^{-5}
9	2.755731×10^{-6}	10	2.755731×10^{-7}
11	2.505210×10^{-8}	12	2.086513×10^{-9}
13	1.607488×10^{-10}	14	1.146235×10^{-11}
15	7.634830×10^{-13}	16	4.784694×10^{-14}
17	2.814313×10^{-15}	18	1.562069×10^{-16}
19	8.213800×10^{-18}	20	4.110317×10^{-19}

而且这个误差估计 $|R_n(x)| < \frac{M}{(n+1)!}$ 对于所有 $x_0-1<x<x_0+1$ 都成立. 之前我们用局部泰勒公式时, 也只是笼统地讲当 x 充分靠近 x_0, 泰勒多项式会有很好的近似. 这不利于具体的分析. 现在我们有了非常明确的 x 的一个近似范围了, 而且误差也非常的清楚.

注意 定理 4.4.1 的泰勒公式和上一节的局部泰勒公式有什么不同呢? 其实它们都是泰勒公式, 其重要展开都是相同的泰勒多项式, 只是其在表达误差项或者余项时的方法不同. 一个是用中值定理的形式, 即拉格朗日余项, 其结果表现为普通等式; 另一个则是用高阶无穷小量的形式, 即皮亚诺余项, 其结果表现为极限等式. 前者的优势在于便于量化地分析误差项上界.

和上一节一样, 我们最关心 $x_0=0$ 的情况. 则此时, 定理 4.4.1 可以陈述为: 设函数 $f(x)$ 在一个包含 0 的区间 (A,B) 内有 $n+1$ 阶导数, 则对于任意的 $x \in (A,B)$, 有等式

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!}x^k + \frac{f^{(n+1)}(\xi)}{(n+1)!}x^{n+1},$$

[19]这里严格地讲, 应该是 $\forall x \in (x_0-1, x_0+1) \cap (A,B)$.

其中 ξ 是介于 0 与 x 之间的某一点. 对泰勒多项式有如下误差估计式:

$$\left|f(x)-\sum_{k=0}^{n}\frac{f^{(k)}(0)}{k!}x^k\right|=\frac{\left|f^{(n+1)}(\xi)\right|}{(n+1)!}|x|^{n+1}\le\frac{M}{(n+1)!}|x|^{n+1},\quad\forall x\in(A,B).$$

上式的 M 是函数 $\left|f^{(n+1)}(x)\right|$ 在整个 $[A,B]$ 的上界 (如果存在的话). 显然, 当 x 离 0 太远时, 这个误差可能会变得很大. 本质上, 带拉格朗日余项的泰勒公式还是局部的泰勒公式.

定理 4.4.1 的拉格朗日余项为泰勒公式在函数值的计算方面有重要意义. 事实上, 迄今为止, 各种初等函数的数值表, 如三角函数表, 对数函数表等, 都是根据带拉格朗日余项或其他形式余项的泰勒公式计算来的.

例 4.4.1 设 $-\frac{\pi}{4}<x<\frac{\pi}{4}$, 在使用带拉格朗日余项的泰勒公式计算 $\sin x$ 时, 为了使误差小于 5×10^{-7}, 应在其泰勒公式中取多少项?

证明. 考虑 $-\frac{\pi}{4}<x<\frac{\pi}{4}$. 因为

$$(\sin x)^{(n)}=\sin\left(x+\frac{n}{2}\pi\right),$$

故 $\sin x$ 在 $x_0=0$ 时的带拉格朗日余项的泰勒公式应为 (就是上一节的麦克劳林公式换了余项尾巴而已!)

$$\sin x=x-\frac{x^3}{3!}+\cdots+(-1)^{n-1}\frac{x^{2n-1}}{(2n-1)!}+\frac{(-1)^n\cos\xi}{(2n+1)!}x^{2n+1},$$

其中 ξ 是介于 0 与 x 之间的某一点. 和之前的局部泰勒公式不同的是, 因为拉格朗日余项需要形式化地写出 ξ 处的高阶导数 $f^{(n+1)}(\xi)$, 而通常我们不知道 ξ 的具体取值, 所以一般要知道函数 $f^{(n)}(x)$ 的一般表达式. 其误差有如下不等式关系:

$$\left|\frac{(-1)^n\cos\xi}{(2n+1)!}x^{2n+1}\right|\le\frac{|x|^{2n+1}}{(2n+1)!}\le\frac{\left(\frac{\pi}{4}\right)^{2n+1}}{(2n+1)!}\le\frac{1}{(2n+1)!}.$$

为了使误差小于 5×10^{-7}, 取 $n=5$ 即可, 因为 $\frac{1}{11!}<3\times10^{-8}<5\times10^{-7}$. 也就是说, 对于任意的 $-\frac{\pi}{4}<x<\frac{\pi}{4}$, 总有

$$\left|\sin x-\left(x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\frac{x^9}{9!}\right)\right|<5\times10^{-7}.$$

这个误差在多数应用中已经足够精确了. 若取 $n=6$, 则有 (对于任意的 $-\frac{\pi}{4}<x<\frac{\pi}{4}$)

$$\left|\sin x-\left(x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\frac{x^9}{9!}-\frac{x^{11}}{11!}\right)\right|=\left|\frac{(-1)^6\cos\xi}{(2\times6+1)!}x^{2\times6+1}\right|\le\frac{1}{(2\times6+1)!}=\frac{1}{13!}\approx1.605\times10^{-10}.$$

为此, 我们只需要四则运算即可搞定高精度的 $\sin x$ 的估值. □

上个例题展现了泰勒多项式对高精度的 $\sin x$ 估值的惊人威力. 但是! 它似乎只能用来计算 $-\frac{\pi}{4}<x<\frac{\pi}{4}$ 范围内的 $\sin x$ 值. 如何才能计算整个数轴上的 x 呢? 以下是笔者自己的一个想法:

1. $\sin x$ 是周期函数, 其周期为 2π. 也就是说, 对于任意的 x, 都有 $\sin x=\sin(x+2k\pi), k\in\mathbb{Z}$. 故我们可以将 x 的范围缩小为 $-\pi$ 到 π 之间;

2. 利用三角恒等式 $\sin(\pi-x)=\sin x$, 故我们可以将 x 的范围缩小为 $-\frac{\pi}{2}$ 到 $\frac{\pi}{2}$ 之间;

3. 利用三角恒等式 $\sin x = 2\sin\frac{x}{2}\cos\frac{x}{2} = 2\sin\frac{x}{2}\sqrt{1-\sin^2\frac{x}{2}}$, 故我们可以只考虑如何去计算 $t := \frac{x}{2}$, 其范围正好就是 $-\frac{\pi}{4}$ 到 $\frac{\pi}{4}$;

4. 最终用上一题的结果, 我们能高精地估算 $\sin t\ (-\frac{\pi}{4} < t < \frac{\pi}{4})$. 从而最终能估计一开始的 $\sin x$.

但是需要声明的, 上一题的估计的误差, 不能直接生搬硬套到这里. 当我们用 5 阶泰勒多项式来近似时, 我们对 $u = \sin t\ (-\frac{\pi}{4} < t < \frac{\pi}{4})$ 有小于 1.605×10^{-10} 的误差, 但是我们最后是将含有误差的 u 代入到以下函数

$$h(u) = 2u\sqrt{1-u^2}$$

以获得最终的结果. 可它的最终误差还有待进一步分析, 这超出了本书的范畴. (以上讨论属于计算数学的范畴, 它目的就是要实现真实计算机上高性能高精度的数值计算办法.)

下面是几个常见初等函数的带拉格朗日余项的泰勒公式:

结论 4.4.1: 常见的基本初等函数的带拉格朗日余项的泰勒公式 (默认 $x_0 = 0$)

1. $\mathrm{e}^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \frac{\mathrm{e}^\xi}{(n+1)!}x^{n+1}, \quad (-\infty < x < +\infty).$

2. $\sin x = x - \frac{x^3}{3!} + \cdots + (-1)^{n-1}\frac{x^{2n-1}}{(2n-1)!} + (-1)^n\frac{\cos\xi}{(2n+1)!}x^{2n+1}, \quad (-\infty < x < +\infty).$

3. $\cos x = 1 - \frac{x^2}{2!} + \cdots + (-1)^n\frac{x^{2n}}{(2n)!} + (-1)^{n+1}\frac{\cos\xi}{(2n+2)!}x^{2n+2}, \quad (-\infty < x < +\infty).$

4. $$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots + \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}x^n \\ + \frac{\alpha(\alpha-1)\cdots(\alpha-n)}{(n+1)!}(1+\xi)^{\alpha-n-1}x^{n+1}, \quad (-1 < x < +\infty).$$

5. $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n-1}\frac{x^n}{n} + (-1)^n\frac{x^{n+1}}{(n+1)(1+\xi)^{n+1}}, \quad (-1 < x < +\infty).$

在以上各式中, ξ 是介于 0 与 x 之间的数.

应该指出: 在上述公式中, $(1+x)^\alpha$ 及 $\ln(1+x)$ 的拉格朗日余项相对而言较大. 换句话说, 当我们在泰勒公式中取相同项数 n 时, $\mathrm{e}^x, \sin x$ 及 $\cos x$ 的拉格朗日余项一般较小, 而 $(1+x)^\alpha$ 与 $\ln(1+x)$ 的拉格朗日余项较大. 很明显, $\mathrm{e}^x, \sin x$ 及 $\cos x$ 的拉格朗日余项中有明显的关于 n 的阶乘的分母项, 而 $\ln(1+x)$ 没有阶乘的分母项; $(1+x)^\alpha$ 虽然有阶乘的分母项, 但分子也有阶乘, 两者会相抵消.

泰勒公式对某些函数会失去意义, 例如函数

$$f(x) = \begin{cases} \mathrm{e}^{-\frac{1}{x^2}}, & x \neq 0, \\ 0, & x = 0. \end{cases}$$

显然它在整个 $\mathbf{R}$ 上都是连续的. 先计算它的一阶导数. 首先, 我们验证函数 $f(x)$ 在点 $x = 0$ 处的导数 $f'(0) = 0$. 事实上,

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} = \lim_{\Delta x\to 0}\frac{f(0+\Delta x) - f(0)}{\Delta x} = \lim_{\Delta x\to 0}\frac{\mathrm{e}^{-\frac{1}{(\Delta x)^2}}}{\Delta x},$$

而洛必达法则无法直接应用于这一个 $\dfrac{0}{0}$ 型未定式, 需要将它做一点变形, 变成 $\dfrac{\infty}{\infty}$ 型未定式:

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\frac{\frac{1}{\Delta x}}{\frac{1}{\mathrm{e}^{-\frac{1}{(\Delta x)^2}}}}=\lim_{\Delta x\to 0}\frac{\frac{1}{\Delta x}}{\frac{1}{\mathrm{e}^{(\Delta x)^2}}}=\lim_{t\to\infty}\frac{t}{\mathrm{e}^{t^2}},$$

其中 $t:=\dfrac{1}{\Delta x}$, 当 $\Delta x\to 0$ 时, $t\to\infty$. 利用洛必达法则, 很容易证明

$$f'(0)=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{t\to\infty}\frac{t}{\mathrm{e}^{t^2}}=\lim_{t\to\infty}\frac{1}{2t\mathrm{e}^{t^2}}=0.$$

因此, 我们证明了

$$f'(x)=\begin{cases}\dfrac{2}{x^3}\mathrm{e}^{-\frac{1}{x^2}}, & x\neq 0,\\ 0, & x=0,\end{cases}$$

其次, 用类似于上面的方法, 可以证明 $f''(0)=0$, 并且可以一般地证明 $f(x)$ 在点 $x=0$ 处的各阶导数都等于零, 即

$$f^{(n)}(0)=0,\quad n=1,2,\ldots.$$

因此, 对于任意的正整数 n, $f(x)$ 在点 $x=0$ 处的带拉格朗日余项的泰勒公式是

$$f(x)=0+\frac{f^{(n+1)}(\xi)}{(n+1)!}x^{n+1}.$$

这说明, 此时拉格朗日余项就是函数本身, 或者说泰勒多项式的所有项的系数都是 0, 这在函数值计算上没有告诉我们任何事情. 我们对误差项 (拉格朗日余项) 感兴趣的前提是, 我们首先能拥有一个用于近似为目的的多项式函数. 这个极端的例子表明, *泰勒公式对于函数值的计算也不是万能的. 特别是, 若出现在所关心的 x_0 处的任意阶导数都为 0, 那么泰勒多项式将不复存在, 因而泰勒展开没有意义.*

问题 思考题. 试比较局部泰勒公式与带拉格朗日余项的泰勒公式在意义上与应用上的差别.

证明. 局部泰勒定理和带拉格朗日余项的泰勒公式都用于函数的局部逼近, 但它们在形式和意义, 应用上有一些的差别:

1. 形式上的差别

   (a) 局部泰勒定理: 提供了一个关于函数 $f(x)$ 在 x_0 处的 n 阶多项式近似, 后面加上一个高阶无穷小项 $o\left((x-x_0)^n\right)$, 这意味着误差的增长速度要比 $(x-x_0)^n$ 快, 即该项比 $(x-x_0)^n$ 低阶.

   $$f(x)=f(x_0)+f'(x_0)(x-x_0)+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+o((x-x_0)^n)$$

   这种表示方法没有给出余项的确切表达式, 只是给出了一个 "渐进" 的误差估计, 即对于足够小的 $(x-x_0)$, 误差的增长速度会更快.

   (a) 带拉格朗日余项的泰勒公式: 在泰勒展开式的基础上, 明确给出了余项的具体形式. 余项是通过 $f^{(n+1)}(\xi)$ 来表示的, 其中 ξ 是在 x_0 和 x 之间的某个点. 这种余项的形式提供了一个精确的误差估计.

   $$f(x)=f(x_0)+f'(x_0)(x-x_0)+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1}$$

   这种表示方法给出了余项的具体量度, 能精确反映余项对 $x-x_0$ 的依赖, 并且可以用来估计余项的上界.

2. 意义上的差别

   (a) 局部泰勒定理: 它更侧重于对函数的局部性质的描述, 提供了一个大致的误差估计, 适合用于快速推导和近似分析, 特别是在计算中用于舍去高阶项时. 然而, 它没有提供一个明确的余项形式, 只是通过 o 符号表示误差的快速增长.

   (a) 带拉格朗日余项的泰勒公式: 它的重点是精确描述误差. 通过给出余项的具体形式, 可以根据该余项表达式计算出误差的具体大小, 这对于需要高精度近似的应用特别有用, 尤其是在数值分析, 优化算法中, 余项的量化可以帮助评估误差并决定截断的级数.

3. 应用上的差别

   (a) 局部泰勒定理: 通常用于快速推导和理论分析中, 尤其是在不需要精确控制误差的场合. 例如, 在物理学或工程学中, 当只关心近似解时, 可以使用局部泰勒展开.

   (a) 带拉格朗日余项的泰勒公式: 它的应用范围更广泛, 特别是在数值计算中, 使用明确的余项形式可以确保计算结果的误差控制. 例如, 数值积分, 微分方程的数值解法等需要控制误差的精确度时, 带拉格朗日余项的公式更为适用.

总结: 局部泰勒定理提供了一种关于函数的快速逼近, 不明确给出误差的具体形式, 只关心误差的增长速度; 带拉格朗日余项的泰勒公式则给出了误差的具体度量, 能提供更精确的误差估计, 适用于需要控制精度的应用场景. □

4.5 极值问题

所谓的**极值问题**, 就是**寻求一个给定函数在一定范围内的极值或最值的问题**. 这样的问题也叫做**最优化 (optimization) 问题**. 历史上各式各样的极值问题[20]曾经是数学发展的重要源泉之一, 而某些极值问题的研究与微分学的发展有直接的联系. 现在我们利用微分学的知识来讨论极值问题. 我们先引入一些术语.

定义 4.5.1: 极值, 极值点

我们称一点 x_0 是函数 $f(x)$ 的**极值点**, 如果该函数在点 x_0 附近有定义且存在一个双侧[a]邻域 $U_\delta(x_0)$, 使得

$$f(x) \leqslant f(x_0), \quad \forall x \in U_\delta(x_0),$$

或

$$f(x) \geqslant f(x_0), \quad \forall x \in U_\delta(x_0).$$

在前一种情况下, 称 $f(x_0)$ 的值为**极大值**, 点 x_0 为**极大值点**; 而在后一种情况下, 称 $f(x_0)$ 的值为**极小值**, 点 x_0 为**极小值点**. 极大值与极小值统称为**极值**.

[a]本书的定义要求必须是双侧的, 所以闭区间的端点天然不在考虑范围内.

过去, 我们曾谈论过一个闭区间 $[a,b]$ 上的连续函数的最大值与最小值. 应该指出: 极大值或极小值只是函数在一个局部范围内的最大值或最小值, 在整个区间上它们未必是最大值或最小值. 反过来, 一个函数在闭

[20]我们每时每刻都在考虑如何做出最优的选择. 利润最大, 用料最省, 效率最高, 费用最少, 路线最短, 容积最大等问题. 最优化问题无处不在. 专业内容可参考: 最优化: 建模、算法与理论/最优化计算方法. 我在日本期间经常会参与由高中生主导的最优化应用课题: 高大連携–筑波大学. 现实具体应用场景可以参考商用优化软件: 了解杉数求解器 COPT-杉数科技, 天筹求解器服务-华为云.

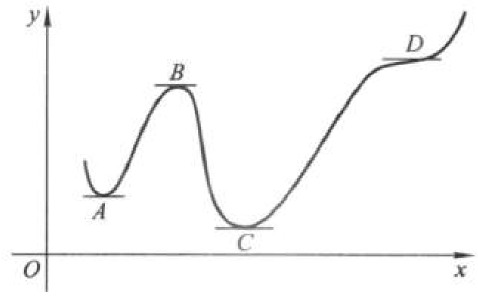


图 4.6: 极值点的一阶必要条件, 费马 (Fermat) 定理.

区间上的最大值或最小值也未必一定是极大值或极小值, 比如若一个函数的最大值或最小值不是在定义区间的内部达到, 而是在端点达到, 这时它就一定不是极大值或极小值, 因为这时不存在极值点所要求的双侧邻域. 但是, 无论如何, 如果函数的最大值或最小值在定义区间内达到, 则它必为极大值或极小值, 这一事实直接使用定义即可得到.

由上面的讨论可看出: **寻求一个函数在一个闭区间上的最大值或最小值, 只要我们找出它在区间内的全部极值点, 再将这些点所对应的函数值与函数在端点处的值加以比较就够了.** 下面的定理给出了极值点的一阶必要条件.

定理 4.5.1: 极值点的一阶必要条件, 费马 (Fermat) 定理

设函数 $f(x)$ 在区间 (a,b) 上有定义. 若 $x_0 \in (a,b)$ 是 $f(x)$ 的极值点, 并且 $f(x)$ 在点 x_0 处可导, 则

$$f'(x_0) = 0.$$

证明. 不失一般性, 我们假定 x_0 是 $f(x)$ 的极大值点. 这时, 存在点 x_0 的一个邻域 $U_\delta(x_0) \subset (a,b)$, 使得 $f(x) \leqslant f(x_0), \forall x \in U_\delta(x_0)$. 这样, 我们有

$$\lim_{x \to x_0+0} \frac{f(x)-f(x_0)}{x-x_0} \leqslant 0$$

和

$$\lim_{x \to x_0-0} \frac{f(x)-f(x_0)}{x-x_0} \geqslant 0$$

$f(x)$ 在点 x_0 处的可导性说明着上述两个单侧极限存在并且都等于 $f'(x_0)$, 于是推出 $f'(x_0)=0$. 证毕.[21] □

这个定理 4.5.1 告诉我们:

1. 对于可导函数而言, 极值点的必要条件是函数在该点处的导数为零. 其几何意义是: 在平面上, 一条可微曲线弧上极值点所对应点处的切线与 x 轴平行, 如图 4.6 中 A, B, C 三点处的切线.

2. 然而, 我们应当注意, 使导数等于零的点不一定就是极值点. 请观察图 4.6 的点 D.

3. 另外, 定理 4.5.1 只适用于可导函数. 回顾极值点的定义, 我们对函数 f 的性质没有任何要求, 它甚至不连续都可以. 绝对值函数在 $x=0$ 处没有导数, 但是却在该点处取得极小值.

[21]本定理的证明过程和之前学习过的罗尔定理的证明是完全一致的. 事实上, 在罗尔定理的找到的 c 就是极值点. 在其证明中, 我们用关于是否是常函数的讨论, 得到其最大值在 (a,b) 内达到. 之后就完全一致了.

定义 4.5.2

通常将导数等于零的点称为**驻点**, 也叫**稳定点 (stationary point)** 或临界点. 即, 导函数的根, 或者方程 $f'(x) = 0$ 的解集.

驻点是一个老的名字, 可能从物理学上得来: 如果将考虑的函数看作一个运动物体的位移函数 (即, x 轴是时间, y 轴是运动的直线), 那么导数等于零意味着其瞬时速度为零, 似乎在这一瞬间运动着的物体停驻于原处. 人们也将导数等于零的点称作稳定点或临界点, 因为在许多现象中这种点对应的是某种稳定状态或临界状态. 在本教材中, 今后我们将导数等于零的点称为稳定点. 根据上面的讨论, **对于一个可导函数, 极值点要在稳定点中寻求. 但是稳定点不一定都是极值点.** 即对于可导函数而言, 定理 4.5.1 只能缩小极值点的范围, 而不能直接告诉你最终答案.

为了确定一个稳定点是否是极值点, 一般要考虑在稳定点左侧及右侧附近函数的性态. 比如, 若在稳定点左侧附近函数递增 (或等价地说, 其导数是非负的), 而在稳定点右侧附近函数递减 (导数是非正的), 这时稳定点一定是极大值点. 即, **只有当在稳定点的左侧及右侧附近, 函数单调性互异, 此时稳定点才是极值点.** 图 4.6 的点 D 就是一个反例.

例 4.5.1 求出函数 $f(x) = x^3 - 6x^2 - 15x + 4$ 的极大值点与极小值点.

证明. 因 $f'(x) = 3x^2 - 12x - 15 = 3(x+1)(x-5)$, 故 $f(x)$ 的稳定点是 $x = -1$ 及 $x = 5$, 并且有 $f'(x) > 0, \forall x < -1; f'(x) < 0, \forall x \in (-1, 5); f'(x) > 0, \forall x > 5$. 因此, $x = -1$ 为极大值点, $x = 5$ 为极小值点. 把上述讨论列成下表将更清晰.

x	$(-\infty, -1)$	-1	$(-1, 5)$	5	$(5, +\infty)$
$f'(x)$	+	0	-	0	+
$f(x)$	$\nearrow$	极大值	$\searrow$	极小值	$\nearrow$

□

一般来说, 若导数在一个稳定点的两侧附近改变其符号, 则该稳定点为极值点. 此时, 导数在该稳定点的附近是单调变化的, 从正数变成负数, 或者从负数变成正数. 因此, 考查函数在稳定点处的二阶导数的符号也是确定稳定点是否是极值点的一个方法.

定理 4.5.2: 极值点的二阶充分条件

设函数 $f(x)$ 在区间 (a, b) 内具有一阶导数, 并且在某点 $x_0 \in (a, b)$ 处存在二阶导数. 如果 $f'(x_0) = 0$, 且该点处的二阶导数有以下情况时:

- $f''(x_0) > 0$, 则 x_0 是一个极小值点.
- $f''(x_0) < 0$, 则 x_0 是一个极大值点.
- $f''(x_0) = 0$, 则无法通过二阶导数判断该点是否为极值点, 可能需要进一步的分析 (如高阶导数测试, 或者通过其他方法判断).

证明. 我们只讨论 $f''(x_0) > 0$ 的情况, 另一种情况的讨论完全类似. 因 $f'(x_0) = 0$, 并且

$$0 < f''(x_0) = \lim_{x \to x_0} \frac{f'(x) - f'(x_0)}{x - x_0} = \lim_{x \to x_0} \frac{f'(x)}{x - x_0},$$

故根据函数极限的保序性[22], 存在一个正数 δ, 使得 $U_\delta(x_0) \subset (a, b)$, 并且

$$\frac{f'(x)}{x - x_0} > 0, \quad \forall x \in U_\delta(x_0) \setminus \{x_0\}.$$

注意以上一个双侧极限. 也就是说, 在点 x_0 的两侧附近 $f'(x)$ 与 $x - x_0$ 的符号相同. 于是, $f'(x)$ 在点 x_0 的左侧附近为负的, 而在点 x_0 的右侧附近为正的. 再依据 (定理) 导数与单调性的关系. 因此, $f(x_0)$ 为极小值. 证毕. □

仍以例 4.5.1 中的函数为例. 因 $f''(x) = 6x - 12$, 故 $f''(-1) = -18 < 0$, $f''(5) = 18 > 0$. 因此, 由定理 4.5.2 再次证实 $x = -1$ 为极大值点, $x = 5$ 为极小值点.

定理 4.5.2 的另外一种证明方法是应用局部泰勒公式. 设 $f'(x_0) = 0$, 那么泰勒公式展开到 2 阶有:

$$\begin{aligned} f(x) &= f(x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + o\left((x - x_0)^2\right) \\ &= f(x_0) + \left(\frac{1}{2} f''(x_0) + \alpha(x)\right)(x - x_0)^2, \quad (x \to x_0), \end{aligned}$$

其中 $\alpha(x)$ 是一个无穷小量 $(x \to x_0)$, 这来自于无穷小量的定义. 我们只讨论 $f''(x_0) > 0$ 的情况, 另一种情况的讨论完全类似. 因此, 当 x 充分靠近 x_0 时,

$$|\alpha(x)| < \frac{1}{2} f''(x_0).$$

从而 $\frac{1}{2} f''(x_0) + \alpha(x) > 0$. 这里 $\alpha(x)$ 具体的符号不重要, 重要的是它的量级. 即使它是负数, 也不能把 $\frac{1}{2} f''(x_0) + \alpha(x)$ 变成负的. 于是, 当 x 充分靠近 x_0 时,

$$\left(\frac{1}{2} f''(x_0) + \alpha(x)\right)(x - x_0)^2 \geqslant 0.$$

由此推出, 当 x 充分靠近 x_0 时, $f(x) \geqslant f(x_0)$. 这个证明的好处在于它可以推广到更一般的情况, 得到如下结论: 设函数 $f(x)$ 在点 x_0 处有 $2n$ 阶导数, 并且在点 x_0 处前 $2n - 1$ 个低阶导数值都为零:

$$f'(x_0) = f''(x_0) = \cdots = f^{(2n-1)}(x_0) = 0.$$

- 若 $f^{(2n)}(x_0) > 0$, 则 x_0 为极小值点.
- 若 $f^{(2n)}(x_0) < 0$, 则 x_0 为极大值点.
- 若 $f^{(2n)}(x_0) = 0$, 则无法判断.

定理 4.5.2 是 $n = 1$ 的一个特例.

命题 4.5.1

若函数 $f(x)$ 在 x_0 处的前 2 阶导数为零, 且 3 阶导数不为零, 则 x_0 一定不是极值点.

[22] 复习: 设 $\lim_{x \to a} f(x) = l$, 且 $l > 0$(或 $l < 0$), 则存在一个 $\delta > 0$, 使得 $f(x) > 0$ (或 $f(x) < 0$), 只要 $0 < |x - a| < \delta$.

证明. 设 $f'(x_0)=0$ 和 $f''(x_0)=0$, 那么泰勒公式展开到 3 阶有:

$$\begin{aligned} f(x) &= f(x_0)+\frac{f^{(3)}(x_0)}{3!}(x-x_0)^3+o\left((x-x_0)^3\right) \\ &= f(x_0)+\left(\frac{f^{(3)}(x_0)}{3!}+\alpha(x)\right)(x-x_0)^3, \quad (x\to x_0), \end{aligned}$$

其中 $\alpha(x)$ 是一个无穷小量 $(x\to x_0)$. 我们只讨论 $f^{(3)}(x_0)>0$ 的情况, 另一种情况的讨论完全类似. 随着 x 充分靠近 x_0, $\alpha(x)$ 相对于 $\dfrac{f^{(3)}(x_0)}{3!}\neq 0$ 是很小的, 因此, 存在一个 x_0 的邻域, 使得

$$\frac{f^{(3)}(x_0)}{3!}+\alpha(x)>0.$$

即三次项 $(x-x_0)^3$ 前的符号总是正的. 我们讨论 x 从不同方向靠近 x_0 的情况:

- 当 x 在 x_0 充分近的右侧时, $(x-x_0)^3>0$, 所以 $f(x_0)<f(x)$.
- 当 x 在 x_0 充分近的左侧时, $(x-x_0)^3<0$, 所以 $f(x)<f(x_0)$.

这表明, 函数 $f(x)$ 在 x_0 处并没有保持极值的性质. 证毕. □

推广到更一般的情况, 得到如下结论: 若函数 $f(x)$ 在稳定点 x_0 处的前 $2n$ 阶导数为零, 而 $2n+1$ 阶导数不为零, 则 x_0 必不是极值点. 命题 4.5.1 是 $n=1$ 的一个特例.

根据费马定理, 若函数在极值点处可导, 则极值点一定是稳定点. 另外, 函数的不可导点也可能是极值点.

例 4.5.2 求函数 $f(x)=x^{\frac{2}{3}}$ 的极值点. 4-5 节例题图 | Desmos

证明. $f(x)$ 在区间 $(-\infty,+\infty)$ 上连续, 且

$$f'(x)=\frac{2}{3}x^{-\frac{1}{3}}=\frac{2}{3}\cdot\frac{1}{\sqrt[3]{x}}, \quad x\neq 0.$$

$f(x)$ 在点 $x=0$ 处不可导. 对于任意的 $x\in(-\infty,0)$, 有 $f'(x)<0$; 对于任意的 $x\in(0,+\infty)$, 有 $f'(x)>0$. 所以, $f(x)$ 在区间 $(-\infty,0]$ 内递减, 在区间 $[0,+\infty)$ 内递增, 进而得出

$$f(x)>f(0), \quad x\neq 0,$$

即 $x=0$ 是 $f(x)$ 的极小值点. □

例 4.5.3 研究函数 $f(x)=x^3\mathrm{e}^{-x}$ 的极值点. 4-5 节例题图 | Desmos

证明. 我们有 $f'(x)=\left(3x^2-x^3\right)\mathrm{e}^{-x}=x\left(3x-x^2\right)\mathrm{e}^{-x}$, 故 $f(x)$ 的所有稳定点为 $x=0$ 及 $x=3$. 容易算得

$$f''(x)=\left(x^3-6x^2+6x\right)\mathrm{e}^{-x}.$$

于是 $f''(0)=0$, $f''(3)=-9\mathrm{e}^{-3}<0$.

1. 根据定理 4.5.2, $x=3$ 是极大值点.

2. 由于 $f''(0)=0$, 故不能判断 $x=0$ 是否是极值点. 我们求 $f(x)$ 的三阶导数. 经过计算得

$$f'''(x)=\left(-x^3+9x^2-18x+6\right)\mathrm{e}^{-x},$$

因此 $f'''(0)=6>0$. 可见, $x=0$ 不是极值点.

□

很多实际问题不是要寻求函数的极大值或极小值, 而是要寻求函数在一个闭区间上的最大值或最小值. 为了解决求最大值或最小值的问题, 一般来说, 这需要求出区间内部的全部极值点, 然后将它们的函数值与区间端点处的函数值加以比较, 从中挑选出最大者或最小者, 即为最大值或最小值.

尽管原则上一般应该如此, 但是在有些实际问题中会出现下述情况: **区间内部只有一个极值点. 在这种特殊情况下, 可以肯定这个极值点必为最值点, 而且如果它是极大值点或极小值点, 则它对应的函数值必定是整个区间上的最大值或最小值**. (画图理解就行, 不必严格证明)

例 4.5.4 求函数 $f(x)=(x-1)\sqrt[3]{x^2}$ 在区间 $[-1,1]$ 上的最大值和最小值. 4-5 节例题图 | Desmos

证明. 我们有

$$f'(x)=x^{\frac{2}{3}}+(x-1)\cdot\frac{2}{3}x^{-\frac{1}{3}}=\frac{5x-2}{3x^{\frac{1}{3}}}.$$

可见, $f(x)$ 的稳定点是 $x=\dfrac{2}{5}$, 不可导点是 $x=0.x=-1$ 和 $x=1$ 是区间端点. $f(x)$ 在上述各点处的值为

$$f\left(\frac{2}{5}\right)=-\frac{3}{5}\sqrt[3]{\frac{4}{25}},\quad f(0)=0,\quad f(-1)=-2,\quad f(1)=0$$

从中选出最大值 0, 故 0 是 $f(x)$ 在 $[-1,1]$ 上的最大值, 最大值点为 $x=0$ 和 $x=1$; 从中选出最小值 -2, 故 -2 是 $f(x)$ 在 $[-1,1]$ 上的最小值, 最小值点为 $x=-1$. □

我不加证明[23]的给出以下符合直觉的结论.

结论 4.5.1

设 $f:\mathbb{R}\to\mathbb{R}$ 是一个连续函数, 满足 $\lim\limits_{x\to+\infty}f(x)=L_1$ 和 $\lim\limits_{x\to-\infty}f(x)=L_2$.

1. 如果 $L_1=L_2$, 则函数 f 在 $\mathbb{R}$ 上存在极值点.
2. 如果 $L_1=L_2=+\infty$, 则函数 f 在 $\mathbb{R}$ 上存在极小值点.
3. 如果 $L_1=L_2=-\infty$, 则函数 f 在 $\mathbb{R}$ 上存在极大值点.

给出一个 $L_1\neq L_2$, 且在 $\mathbb{R}$ 上没有极值点的函数的例子. (比如, $\dfrac{1}{1+e^{-x}}$)

例 4.5.5 费马原理, 最短时间原理 设有甲, 乙两种介质, 其交界面为一平面. 假设光自甲介质中的点 A 发出经过两种介质的交界面而到达乙介质中的点 B. 从物理学上我们知道, 光所走的路径一定满足**折射定律**:

$$\frac{\sin\alpha_0}{\sin\beta_0}=\frac{v_1}{v_2},$$

[23] 可以参考 calculus - Proof about Weierstrass Theorem and limits. - Mathematics Stack Exchange.

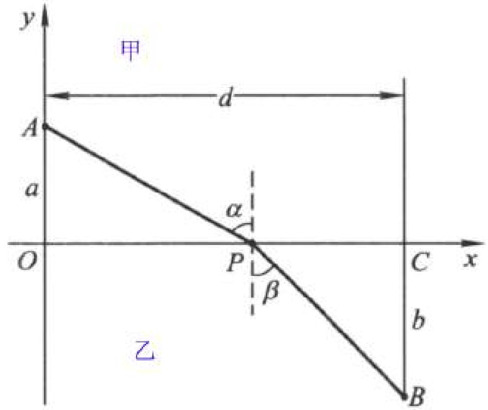


图 4.7: 费马原理: 光所走的路径是时间最少的路径.

其中 α_0 是光的入射角, 而 β_0 是折射角, v_1 与 v_2 分别为光在甲, 乙两种介质中的速度. 证明: **光所走的路径是总时间最少的路径**.

证明. 我们考虑垂直于甲, 乙两种介质交界面且包含 A, B 两点的平面, 并在此平面上取定直角坐标系, 使点 A 落在 y 轴上, 而 x 轴在交界面上 (见图 4.7). 假定点 A 至交界面的距离为 a, 而点 B 至交界面的距离为 b, 点 B 在 x 轴上的垂足 C 至原点 O 的距离为 d.

现在我们假定光从点 $P(x,0)$ 处进入乙介质, 其入射角与折射角分别为 α 与 β, 这时光在甲介质中所走过的路程为 $\sqrt{a^2+x^2}$, 而在乙介质中所走过的路程为 $\sqrt{b^2+(d-x)^2}$, 从而光自点 A 至点 B 所需的总时间为

$$T(x)=\frac{\sqrt{a^2+x^2}}{v_1}+\frac{\sqrt{b^2+(d-x)^2}}{v_2}.$$

注意, 所有参数 a, b, d, v_1, v_2 都是正的. 这里 $T(x)$ 是在整个实轴上有定义的可导函数, 并且当 $x \to +\infty$ 或 $x \to -\infty$ 时, $T(x) \to +\infty$. 因此, $T(x)$ 至少有一个极小值点, 从而又推出它至少有一个稳定点 (因为它处处可导), 即 $T'(x)=0$ 在整个实轴上至少有一个根. 对 $T(x)$ 求一阶与二阶导数, 我们有

$$T'(x)=\frac{x}{v_1\sqrt{a^2+x^2}}-\frac{d-x}{v_2\sqrt{b^2+(d-x)^2}},$$

和

$$T''(x)=\frac{a^2}{v_1\left(\sqrt{a^2+x^2}\right)^3}+\frac{b^2}{v_2\left[\sqrt{b^2+(d-x)^2}\right]^3}.$$

由 $T'(0)<0$, $T'(d)>0$ 以及 $T'(x)$ 的连续性, 可看出 $T'(x)$ 在 $[0,d]$ 上必有 $T'(x)=0$ 的根. 又注意到 $T''(x)>0, \forall x \in \mathbf{R}$, 可知 $T'(x)$ 是单调递增的, 所以 $T'(x)=0$ 有且仅有一个实根. 这样, 存在唯一的点 x_0, 使得 $T'\left(x_0\right)=0$, 即

$$\frac{x_0}{v_1\sqrt{a^2+x_0^2}}=\frac{d-x_0}{v_2\sqrt{b^2+\left(d-x_0\right)^2}},$$

综上所述, 因此 x_0 是 $T(x)$ 的极小值点, 且也是最小值点. 设光在点 x_0 处的入射角为 α_0, 折射角为 β_0, 则上式可写成

$$\frac{\sin\alpha_0}{v_1}=\frac{\sin\beta_0}{v_2},$$

其中, 我们有 (从图中可以知道)

$$\sin\alpha_0 = \frac{x_0}{\sqrt{a^2+x_0^2}}, \quad \sin\beta_0 = \frac{d-x_0}{\sqrt{b^2+(d-x_0)^2}}.$$

这恰好是折射定律的结论. 这样, 我们就证明了光所走的路径是时间最少的路径. 费马原理的核心是: *光的传播路径并非总路径最短, 而是总用时最短的路径.* 折射定律正是描述了光在两种介质交界面上的折射角和入射角的关系, 从而使得在各介质中传播的时间之和最小. □

例 4.5.6 最小二乘法 (极简版) 最小二乘法是一个有广泛应用的有关数据处理的方法. 假设我们要测定某个量的值, 共做了 n 次试验, 测得的数据分别是 $a_1, a_2, \cdots, a_n$. 到底应该取一个怎样的值来代表我们要测定的量呢? 一个简单而有效的方法是: 找这样一个值 x_0, 它使得平方和函数

$$f(x) = \sum_{i=1}^{n} (x-a_i)^2$$

达到最小. 用这样的 x_0 作为我们要测定量的值, 则有较大的把握使误差较小. 现在我们的问题不是讨论最后一句话的合理性 (这要用数理统计的知识说明), 而是讨论怎样求出这样的 x_0.

证明. 上述函数 $f(x)$ 在整个实轴上有定义且连续, 并且当 $x \to +\infty$ 或 $x \to -\infty$ 时, $f(x) \to +\infty$. 因此, 它有极小值. 另外, 它的稳定点很容易由

$$f'(x) = 2\sum_{i=1}^{n}(x-a_i) = 2\left(nx - \sum_{i=1}^{n} a_i\right) = 0$$

求出, 并且不难看出, 它只有一个稳定点:

$$x_0 = \frac{1}{n}(a_1+a_2+\cdots+a_n).$$

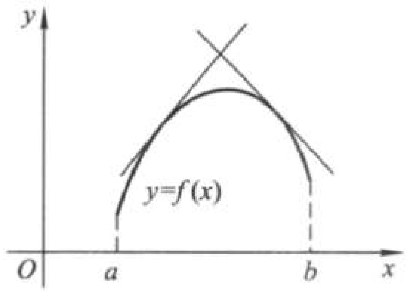
前面我们已经说明这个函数至少有一个极小值点, 且函数在整个实轴上处处可导, 因此这唯一的稳定点只能是极小值点, 并且它对应的极小值也就是该函数的最小值. 故由上式确定的 x_0 即为我们所求的值. □

以上我们利用求平方和函数最小值的方法, 来确定要测定量的估计值 x_0. 这种方法称为最小二乘法. 从讨论结果看出, 用最小二乘法所确定的 x_0, 恰好就是所测得的 n 个数据的算术平均值.

历史的注记. 费马是法国著名数学家, 他先是学习法律的, 后以地方议会的议员为其职业. 他利用公职之余钻研数学, 并在数论, 解析几何, 概率论等方面有重大贡献. 他的著名猜想" 方程

$$x^n + y^n = z^n \quad (n>2)$$

无整数解" 成了各国数学家努力研究的课题, 且直到 20 世纪 90 年代, 即在他去世三百多年后才得以证明. 特别应该指出的是, 费马是微积分的先驱. 他在微积分创立之前, 就提出了求极值点的方法, 其中实际上给出了求曲线切线的方法. 此外, 他与笛卡儿差不多同时得到了解析几何的基本观念. 他淡泊名利, 很少发表自己的著作. 他的许多论述都写在读过的书页的空白处或给友人的信中. 他去世后, 他的儿子将这些遗作汇集成两大卷书出版.



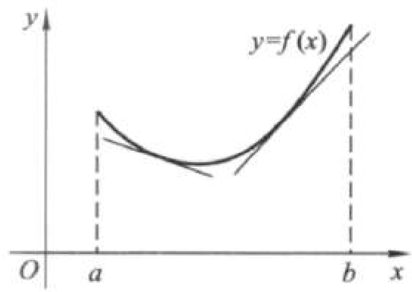


图 4.8: 函数的凸凹性.

4.6 函数的凸凹性与函数作图

我们已经知道函数的一阶导函数的正负号与函数的单调性相关, 自然希望知道函数的二阶导数的正负号反映函数怎样的性质. 我们先看一个很简单的例子. 函数 $y=ax^2$ $(a\neq 0)$ 的二阶导数 $y''=2a$ 是一个常数. 若 $a>0$, 则 $y=ax^2$ 的图形是一条向下凸的抛物线; 若 $a<0$, 则 $y=ax^2$ 的图形是一条向上凸的抛物线. 这个例子启示我们: 函数的二阶导数的正负号可能与函数图形的凸凹性有一定的关系.

4.6.1 函数的凸凹性

定义 4.6.1: 函数的凸凹性

设函数 $f(x)$ 在区间 (a,b) 内可导.[a]

- 我们称 $f(x)$ 在 (a,b) 上是一个**向上凸** (简称凸)[b] 的函数, 如果对于每一点 $x_0\in(a,b)$, 都有

$$f(x)<f\left(x_0\right)+f'\left(x_0\right)\left(x-x_0\right),\quad \forall x\in(a,b), x\neq x_0.$$

- 我们称 $f(x)$ 在 (a,b) 上是一个**向下凸** (简称凹) 的函数, 如果对于每一点 $x_0\in(a,b)$, 都有

$$f(x)>f\left(x_0\right)+f'\left(x_0\right)\left(x-x_0\right),\quad \forall x\in(a,b), x\neq x_0.$$

[a] 本教材只讨论可导函数的凸凹性. 但该定义可以不依赖于导数, 这超出了本书的范围.

[b] 英文的 "凸" 一词 convex, 指的是我们这里定义的 "向下凸 (简称凹)"; 英文的 "凹" 一词 concave, 指的是我们这里定义的 "向上凸 (简称凸)"; 有这种完全相反的翻译. 可能是因为中文是象形文字, 而英文不是. 机器学习中常见的凸函数, 就是 convex function.

现在我们来解释上述不等式的几何意义. 注意到

$$y=f\left(x_0\right)+f'\left(x_0\right)\left(x-x_0\right)$$

是曲线 $y=f(x)$ 的过点 $\left(x_0, f\left(x_0\right)\right)$ 的切线方程, 立刻就发现: 向上凸的条件是要求曲线 $y=f(x)$ 总是要在它的切线的下方, 而向下凸的条件则是要求曲线 $y=f(x)$ 总是落在切线的上方. 这恰好与我们从几何直观上了解的是一致的 (见图 4.8). 另外, 从定义出发可以发现: 如果 $f(x)$ 在 (a,b) 上是向上凸 (向下凸) 的, 那么 $-f(x)$ 在 (a,b) 是向下凸 (向上凸) 的.

定理 4.6.1: 凸凹性的二阶充分必要条件

设函数 $f(x)$ 在区间 (a,b) 内有二阶导数.

- 若对于每一点 $x \in (a,b)$, 都有 $f''(x) > 0$, 则 $f(x)$ 在 (a,b) 上是向下凸的; 反之也成立.
- 若对于每一点 $x \in (a,b)$, 都有 $f''(x) < 0$, 则 $f(x)$ 在 (a,b) 上是向上凸的; 反之也成立.

证明. 设 $x_0 \in (a,b)$ 为任意给定一点. 对于任意一点 $x \in (a,b)$, $x \neq x_0$, 由带拉格朗日余项的泰勒公式有

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!} f''(\xi)(x - x_0)^2,$$

其中 ξ 是介于 x_0 与 x 之间的某一点. 若 $f''(x)$ 在 (a,b) 内总是正的, 则

$$f(x) > f(x_0) + f'(x_0)(x - x_0),$$

即 $f(x)$ 是向下凸的, 反之也成立. 若 $f''(x)$ 在 (a,b) 内总是负的, 则

$$f(x) < f(x_0) + f'(x_0)(x - x_0),$$

即 $f(x)$ 是向上凸的, 反之也成立. 证毕. □

例 4.6.1 请读者自行证明下面函数在其定义域上都是向下凸的.

- e^{ax}, 其中 $a \neq 0$, 定义域是 $\mathbf{R}$
- $-\log(x)$, 定义域是 $(0,+\infty)$
- $x\log(x)$, 定义域是 $(0,+\infty)$
- x^a, 其中 $a \geq 1$ 或 $a \leq 0$, 定义域是 $(0,+\infty)$
- $-x^a$, 其中 $0 \leq a \leq 1$, 定义域是 $(0,+\infty)$

命题 4.6.1

设函数 $f(x)$ 在区间 (a,b) 内可导, 是向下凸的. 则 $f(x)$ 在区间 (a,b) 上的极小值点 (如果存在的话) 都是最小值点.[a]

[a] 因为该性质, 向下凸的函数, 在机器学习中占据核心位置.

证明. 设 x_0 是 $f(x)$ 在区间 (a,b) 上的一个极小值点. 因为函数可导, 由一阶必要条件, 有 $f'(x_0) = 0$. 由于 $f(x)$ 在 (a,b) 上是向下凸的, 对于任意 $x_0 \in (a,b)$, 有:

$$f(x) > f(x_0) + f'(x_0)(x - x_0), \quad \forall x \in (a,b), x \neq x_0.$$

特别, 当 $f'(x_0) = 0$, 我们得到

$$f(x) > f(x_0) \quad \forall x \in (a,b), x \neq x_0.$$

证毕. □

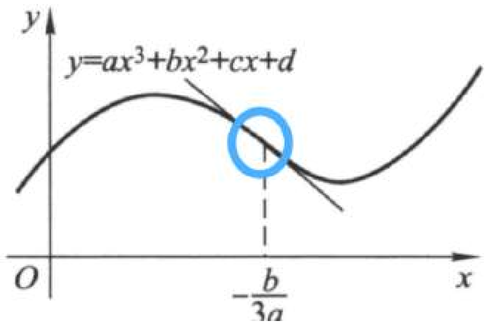


图 4.9: 函数 $y = ax^3 + bx^2 + cx + d$ 的拐点.

上一节的例题 4.5.5 和例题 4.5.6 的目标函数都是向下凸的: 例题 4.5.5 中目标函数的二阶导函数满足 $T''(x) > 0, \forall x \in \mathbf{R}$; 例题 4.5.6 中目标函数的二阶导函数满足 $f''(x) = 2n > 0, \forall x \in \mathbf{R}$. 故它们的稳定点都是最小值点.

类似地, 可以证明: 设函数 $f(x)$ 在区间 (a, b) 内可导, 且 $f(x)$ 在区间 (a, b) 上是向上凸的. $f(x)$ 在区间 (a, b) 上的极大值点 (如果存在的话) 都是最大值点.

例 4.6.2 设函数 $y = ax^3 + bx^2 + cx + d$ $(a, b, c, d$ 为常数且 $a > 0)$, 研究这个函数的凸凹性区间.

证明. 很容易计算得 $y'' = 6ax + 2b$. 于是, 当 $x > -\dfrac{b}{3a}$ 时, $y'' > 0$; 当 $x < -\dfrac{b}{3a}$ 时, $y'' < 0$. 可见, 这个函数在区间 $\left(-\infty, -\dfrac{b}{3a}\right)$ 上向上凸, 而在区间 $\left(-\dfrac{b}{3a}, +\infty\right)$ 上向下凸, 见图 4.9. 这个例子中, 在点 $x = -\dfrac{b}{3a}$ 的两侧, 函数的凸凹性恰好相反, 且对应点处函数曲线 (函数的图形) 的切线穿过函数曲线, 并将函数曲线分作两部分, 而这两部分各自落在切线的一侧. □

定义 4.6.2

若函数 $f(x)$ 在某一点 x_0 的两侧附近的凸凹性相反, 则称该点为**拐点**.

上例中点 $x = -\dfrac{b}{3a}$ 就是一个拐点. 对于有连续的二阶导数的函数而言, 拐点的必要条件是该函数在此点处的二阶导数为零. 注意, 二阶导数尽管存在, 也不一定是连续的, 这里我们需要更强的条件.

定理 4.6.2: 拐点的二阶必要条件

设函数 $f(x)$ 在区间 (a, b) 内有连续的二阶导数. 若点 $c \in (a, b)$ 是 $f(x)$ 的拐点, 则 $f''(c) = 0$.

证明. 用反证法. 设 $f''(c) \neq 0$, 不妨设 $f''(c) > 0$. 由 $f''(x)$ 的连续性, 必存在 c 的一个邻域 $U_\delta(c)$, 使得

$$f''(x) > 0, \quad \forall x \in U_\delta(c).$$

于是, 由定理?? 知, $f(x)$ 在整个 $U_\delta(c)$ 内都是向下凸的, 不论 x 是在 c 的左侧附近还是在右侧附近. 这与 c 是拐点矛盾. 证毕. □

二阶导数为零仅是拐点的必要条件, 不是充分条件. 例如, 函数 $y = x^4$ 在点 $x = 0$ 处的二阶导数为零, 但 $x = 0$ 并不是拐点. (见 4-6 节用图 | Desmos) 要判断一个二阶导数为零的点是否是拐点, 还需要进一步分析.

例 4.6.3 设函数 $f(x)$ 在区间 (a,b) 内有连续的三阶导数. 若存在一点 $c\in(a,b)$, 使得 $f''(c)=0$ 且 $f'''(c)\neq 0$, 则 c 必为 $f(x)$ 的拐点.[24]

证明. 对二阶导函数 $f''(x)$ 自身直接应用一阶泰勒展开:

$$f''(x)=f''(c)+f'''(c)(x-c)+o(x-c),\quad (x\to c).$$

由于 $f''(c)=0$, 上式简化为

$$\begin{aligned}f''(x)&=f'''(c)(x-c)+o(x-c)\\&=\left(f'''(c)+\alpha(x)\right)(x-x_0),\quad (x\to c).\end{aligned}$$

其中 $\alpha(x)$ 是一个无穷小量 $(x\to c)$. 我们只讨论 $f'''(c)>0$ 的情况 ($f'''(c)<0$ 情况的讨论完全类似). 随着 x 充分靠近 $x_0, \alpha(x)$ 相对于 $f'''(c)>0$ 是很小的, 因此, 存在一个 x_0 的邻域, 使得

$$f'''(c)+\alpha(x)>0.$$

拐点的定义是二阶导数在点 c 附近的符号发生变化. 我们来考察 $f''(x)$ 在 c 左右的符号:

- 当 x 向 c 的右侧 (即 $x>c$) 靠近时, $(x-c)>0$, 因此 $f''(x)$ 为正;
- 当 x 向 c 的左侧 (即 $x<c$) 靠近时, $(x-c)<0$, 因此 $f''(x)$ 为负.

因此, 二阶导数 $f''(x)$ 在 c 处的符号发生了变化, 从负变为正或从正变为负, 这证明了 c 是拐点. 证毕. □

4.6.2 函数的渐近线

这里, 我们需要讲一下渐近线的概念. 为了明确起见, 我们讨论 $x\to+\infty$ 的情况. 至于 $x\to-\infty$ 的情况, 完全类似. 只需将所有出现 $+\infty$ 的地方全部改成 $x\to-\infty$ 即可.

定义 4.6.3: $x\to+\infty$ 时的渐近线

设函数 $f(x)$ 定义在区间 $(c,+\infty)$ 上, 我们称直线 $y=ax+b$ 是曲线 $y=f(x)$ 当 $x\to+\infty$ 时的一条渐近线, 如果当 $x\to+\infty$ 时, 点 $(x,f(x))$ 到直线 $y=ax+b$ 的距离趋向于零, 如图 4.10 所示.

定理 4.6.3: $x\to+\infty$ 时的渐近线的充分必要条件

设函数 $f(x)$ 在区间 $(c,+\infty)$ 上有定义, 则直线 $y=ax+b$ 是曲线 $y=f(x)$ 当 $x\to+\infty$ 时的渐近线的充分必要条件是

$$a=\lim_{x\to+\infty}\frac{f(x)}{x},\quad b=\lim_{x\to+\infty}(f(x)-ax).$$

证明.
- **必要性.** 设直线 $y=ax+b$ 是曲线 $y=f(x)$ 当 $x\to+\infty$ 时的渐近线, 又设点 $(x,f(x))$ 到直线 $y=ax+b$ 的距离为 $d(x)$, 则 $d(x)\to 0$ $(x\to+\infty)$. 由图 4.10 可看出 ($d(x)$ 是图中红色框住的部分, 是三角形的斜边.)

[24] 本题的结论和证明都和命题 4.5.1 非常相似. 请对比着看.

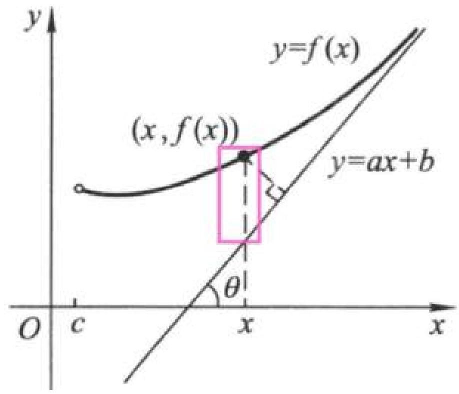


图 4.10: 函数的渐近线.

$$d(x) = |f(x) - (ax+b)|\cos\theta,$$

其中 θ 为渐近线 $y = ax + b$ 与 x 轴正向的夹角 (固定的值). 条件 $d(x) \to 0$ $(x \to +\infty)$ 意味着

$$\lim_{x\to+\infty}(f(x) - ax - b) = 0,$$

即有 $b = \lim\limits_{x\to+\infty}(f(x) - ax)$. 由此可得

$$\lim_{x\to+\infty}\left(\frac{f(x)}{x} - a\right) = \lim_{x\to+\infty}\frac{f(x)-ax}{x} = \lim_{x\to+\infty}\frac{b}{x} = 0,$$

因此, $a = \lim\limits_{x\to+\infty}\dfrac{f(x)}{x}$.

- **充分性.** 条件 $\lim\limits_{x\to+\infty}(f(x) - ax) = b$ 蕴含着 $\lim\limits_{x\to+\infty}(f(x) - ax - b) = 0$, 而后者意味着 $d(x) \to 0$ $(x \to +\infty)$. 证毕.

□

渐近线的分类如下:

1. 一个特殊情况是 $a = 0$, 这时有 $\lim\limits_{x\to+\infty} f(x) = b$. 此时, 我们称渐近线 $y = b$ 为**水平渐近线.**

2. 当 $a \neq 0$ 时, 也称渐近线 $y = ax + b$ 为**斜渐近线**.

3. 另外, 还有一种情况是前面定义所没有包含的: 若当 $x \to c$ (或 $x \to c+0, x \to c-0$) 时, $f(x) \to +\infty$ 或 $f(x) \to -\infty$, 则我们称 $x = c$ 是曲线 $y = f(x)$ 的一条**垂直渐近线**. 这个就是我们之前第一章讨论过的函数极限的一个特例.

例 4.6.4 求曲线 $y = f(x) = \dfrac{x^3+x+1}{(x+1)^2}$ 的渐近线.

证明. 它有没有渐进线, 我们从直觉上无法判断. 最好的办法就是直接套计算公式. 定理 4.6.3 说明: 要先算 a 的值, 再算 b 的值. (如果极限 a 不存在, 那么说明在 $x \to +\infty$ 时函数没有渐近线. 这是因为定理 4.6.3 存在的充分必要条件.) 先计算极限:

$$a = \lim_{x \to +\infty} \frac{f(x)}{x} = \lim_{x \to +\infty} \frac{x^3 + x + 1}{x(x+1)^2} = 1,$$

再计算极限:

$$b = \lim_{x \to +\infty} (f(x) - x) = \lim_{x \to +\infty} \left[\frac{x^3 + x + 1}{(x+1)^2} - x \right] = \lim_{x \to +\infty} \frac{-2x^2 + 1}{(x+1)^2} = -2.$$

立即得到曲线 $y = f(x)$ 的斜渐近线 $y = x - 2$. 见图 4-6 节用图 | Desmos. 又因为 $\lim_{x \to -1} f(x) = -\infty$, 所以 $x = -1$ 是曲线 $y = f(x)$ 的垂直渐近线. □

当然, 并非所有函数的图形当 $x \to +\infty$ 时都一定有渐近线. 对于一般的函数 $f(x)$, 由定理 4.6.3 可看出,

1. 若当 $x \to +\infty$ 时, $\dfrac{f(x)}{x}$ 没有极限,
2. 或者虽然 $\dfrac{f(x)}{x}$ 以 a 为极限, 但 $f(x) - ax$ 没有极限,

这都表明曲线 $y = f(x)$ 当 $x \to +\infty$ 时没有渐近线. 即, 当 $x \to +\infty$ 时有渐近线的充要条件是, 两个极限都存在, 少一个都不行. 例如, 曲线 $y = \mathrm{e}^x$ 当 $x \to +\infty$ 时就没有渐近线, 因为有 $\lim_{x \to +\infty} \dfrac{\mathrm{e}^x}{x} = \infty$. 另一个典型的例子是函数 $f(x) = \sin(x) + x$. 我们来分析一下这个函数的渐近行为:

1. 计算 $\dfrac{f(x)}{x}$ 的极限:

$$\frac{f(x)}{x} = \frac{\sin(x) + x}{x} = \frac{\sin(x)}{x} + 1$$

当 $x \to +\infty$ 时, $\dfrac{\sin(x)}{x} \to 0$, 因此

$$\lim_{x \to +\infty} \frac{f(x)}{x} = 1 = a.$$

2. 计算 $f(x) - ax$ 的极限, 其中 $a = 1$,

$$f(x) - ax = \sin(x) + x - x = \sin(x).$$

显然, $\sin(x)$ 不趋向于一个有限的极限, 它在 $(-1, 1)$ 区间内震荡, 因此

$$\lim_{x \to +\infty} (f(x) - ax) = \lim_{x \to +\infty} \sin(x) \text{ 不存在极限.}$$

因此, 尽管 $\dfrac{f(x)}{x}$ 有极限 $a = 1$, 但 $f(x) - ax = \sin(x)$ 没有极限, 所以函数 $f(x) = \sin(x) + x$ 当 $x \to +\infty$ 时没有渐近线. 见图 4-6 节用图 | Desmos.

4.6.3 函数作图

给定一个函数的表达式, 描绘出它的图形对我们直观地了解该函数的性态当然是十分重要的. 过去, 我们没有别的作图方法, 只有描点法, 即计算若干函数值, 得到平面上的若干点, 然后将它们连成曲线. 一般来说, 要得到较准确的图形, 需要相当多的函数值才可以, 工作量较大. 现在, 利用前面有关函数的单调性, 极值点, 凸凹性, 拐点的讨论, 就能较准确地把握函数的特征, 作出函数的图形. 函数作图的一般步骤如下:

1. 考查函数的定义域及其在定义域内的连续性, 可导性. 如果有间断点, 不可导点, 要将这些点 (一般来说, 这是一些孤立点) 处的函数值计算出来 (如果存在的话), 并描出相应的点. 在间断点处还要弄清它的间断性类别以及左, 右极限, 以便把握函数在间断点附近的性态.
2. 求出函数的导数, 然后找出函数的稳定点, 单调性区间及极值点.
3. 求出函数的二阶导数, 然后确定函数的凸凹性区间及拐点.
4. 当定义域不是有穷区间时, 需要研究 $x \to +\infty$ (或 $x \to -\infty$) 时函数值的变化趋势, 特别是要考虑函数有无渐近线. 如果有渐近线, 应当将其渐近线画出.

下面是一个有关函数作图的例子.

例 4.6.5 试作函数 $y = f(x) = \dfrac{x^2}{x-1}$ 的简图. 见图 4-6 节用图 | Desmos.

证明. 这个函数在点 $x = 1$ 处没有定义, 且

$$\lim_{x \to 1+0} f(x) = +\infty,$$
$$\lim_{x \to 1-0} f(x) = -\infty.$$

因此, $x = 1$ 是这个函数的图形的一条垂直渐近线. 另外, 不难计算得

$$y' = \frac{x(x-2)}{(x-1)^2}.$$

于是, 这个函数有稳定点 $x = 0$ 及 $x = 2$. 求二阶导数得

$$y'' = \frac{2}{(x-1)^3}.$$

可见, 这个函数在 $x > 1$ 时向下凸, 而在 $x < 1$ 时向上凸. 综上, 可列出下表.

x	$(-\infty, 0)$	0	$(0, 1)$	$(1, 2)$	2	$(2, +\infty)$
y'	+	0	-	-	0	+
y''	-	-	-	+	+	+
y	↗ 上凸	极大值	↘ 上凸	↘ 下凸	极小值	↗ 下凸

为了准确起见, 再计算若干点处的函数值, 以描出函数图形上的一些点, 比如可用下表列出的特殊点的函数值描点.

x	-2	-1	0	$\frac{1}{2}$	2	3
y	$-\frac{4}{3}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$	4	$\frac{9}{2}$

最后, 讨论这个函数的图形的斜渐近线. 由于

$$\lim_{x \to \infty} \frac{f(x)}{x} = \lim_{x \to \infty} \frac{x^2}{x(x-1)} = 1,$$
$$\lim_{x \to \infty} (f(x) - x) = \lim_{x \to \infty} \frac{x}{x-1} = 1,$$

根据定理 4.6.3, 当 $x \to \pm\infty$ 时, 这个函数的图形的斜渐近线均为 $y = x + 1$. 根据上面这些讨论便可以描绘出这个函数的图形 (见图 4-6 节用图 | Desmos.). □

第五章　向量代数与空间解析几何

向量也称为矢量 (vector), 它是在几何学, 物理学和力学中经常要用到的一个基本工具. 在本章中, 首先我们介绍向量的概念及其各种运算; 然后, 我们讨论空间解析几何, 其中主要包括空间直线, 空间平面, 二次曲面及其分类. 本章的某些内容是学习多元微积分的预备知识.

5.1　向量代数

5.1.1　向量空间的定义

在物理学中, 我们遇到过像力 (force), 速度 (velocity) 和位移 (displacement) 这样一些量, 它们与普通的量 (标量) 不同, 不仅有大小, 而且有方向. 这种既有大小又有方向的量称为 **向量 (vector)**. 这里我们讨论的向量一般都只限定在平面或空间中.

1. 通常用**有向线段** $\overrightarrow{AB}$ 表示一个向量: 线段 AB 的长度表示向量的大小, 记作 $|\overrightarrow{AB}|$, 称为向量的**模**; 点 A 到点 B 的方向表示向量的方向, 其中 A 叫起始点, B 叫终点. 一般来说, 向量与起始点有关. 但是, 现在我们所讨论的向量是**自由向量**. 也就是说, 我们并不关心其起始点, 而只关心其大小与方向. *两个向量如果大小与方向相同, 则认为它们是同一向量*. 或者说, 一个向量可以自由地平行移动至任何位置. 特别地, 当两个向量彼此平行时, 我们可通过平行移动使它们具有相同的起始点, 这时它们必落在同一直线上. 因此, 彼此平行的两个向量称为**共线**向量.

2. 除了用有向线段表示向量外, 我们还用黑体英文字母表示向量, 例如 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \cdots$. 这时, 它们的模分别记作 $|\boldsymbol{a}|, |\boldsymbol{b}|, |\boldsymbol{c}|, \cdots$.

设一个平面或空间中的向量集合 $\mathbb{V}$, 其中包含所有可能的自由向量. 我们可以写出向量集合如下:

$$\mathbb{V} = \{\boldsymbol{v} \mid \boldsymbol{v} \text{ 是一个自由向量, 具有大小和方向}\}.$$

模为 1 的向量称为**单位向量**. 显然, 在平面上或空间中, 单位向量有无穷多个. 设 $\boldsymbol{a}$ 是任意一个向量. 我们将沿向量 $\boldsymbol{a}$ 的方向的单位向量记作 $\boldsymbol{a}^0$. 与向量 $\boldsymbol{a}$ 大小相同而方向相反的向量记作 $-\boldsymbol{a}$, 称为 $\boldsymbol{a}$ 的**反向量**.

现在我们引进零向量的概念. **零向量**是模为零的向量, 它的起始点与终点重合在一点. 零向量可以以任何方向作为其方向. 零向量用黑体的 $\boldsymbol{0}$ 表示.

下面我们来定义向量的两种基本运算: **向量加法**, **标量乘法**.

定义 5.1.1: 向量加法

我们按照以下定义任意两个向量 $\boldsymbol{a}, \boldsymbol{b}$ 之间的**向量加法**.

1. 任意给定两个非零向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$, 首先, 通过平行移动使 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 有共同起始点; 然后, 按照平行四边形法则, 以 $\boldsymbol{a}, \boldsymbol{b}$ 为相邻两边作平行四边形, 将平行四边形的对角线所形成的与 $\boldsymbol{a}, \boldsymbol{b}$ 同起始点的向量定义为**和向量**, 并记作 $\boldsymbol{a}+\boldsymbol{b}$. 这与按照三角形法则 (将 $\boldsymbol{b}$ 的起始点移至 $\boldsymbol{a}$ 的终点, $\boldsymbol{a}+\boldsymbol{b}$ 就是自 $\boldsymbol{a}$ 的起始点至 $\boldsymbol{b}$ 的终点所决定的向量) 所得的结果是完全一致的.

2. 零向量 $\mathbf{0}$ 与任何向量 $\boldsymbol{a}$ 相加, 其和 $\mathbf{0}+\boldsymbol{a}$ 与 $\boldsymbol{a}+\mathbf{0}$ 均定义为 $\boldsymbol{a}$.

我们可以把向量加法视作为一个函数 $\mathbb{V}\times\mathbb{V}\to\mathbb{V}$, 即

$$(\boldsymbol{a}, \boldsymbol{b}) \mapsto \boldsymbol{a}+\boldsymbol{b}.$$

显然, 这样定义的向量加法运算满足交换律与结合律:

$$\begin{aligned}&\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{b}+\boldsymbol{a},\\&(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}=\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c}).\end{aligned}$$

根据三角形中两边之和大于第三边的事实, 不难验证如下不等式对于任意两个向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 成立 (称为三角不等式):

$$|\boldsymbol{a}+\boldsymbol{b}| \leqslant |\boldsymbol{a}|+|\boldsymbol{b}|.$$

定义 5.1.2: 标量乘法

我们按照以下定义任意一个向量 $\boldsymbol{a}$ 和任意实数 λ 之间的**标量乘法**, 或者叫做**数乘运算**:

1. 任意给定非零向量 $\boldsymbol{a}$, 则 $\lambda\boldsymbol{a}$ 是这样的一个向量, 其模为 $|\lambda||\boldsymbol{a}|$, 而其方向当 $\lambda>0$ 时与 $\boldsymbol{a}$ 一致, 当 $\lambda<0$ 时与 $\boldsymbol{a}$ 相反.

2. 同时, 我们约定 $\lambda\mathbf{0}=\mathbf{0}$.

3. 我们约定 $\boldsymbol{a}-\boldsymbol{b}:=\boldsymbol{a}+(-\boldsymbol{b})$, 其中反向量 $-\boldsymbol{b}\equiv(-1)\boldsymbol{b}$. 这样, 向量之间就有了减法运算. 因为减法运算本质上还是加法和标量乘法, 所以一般我们不把它和加法运算平起平坐.

我们可以把标量乘法视作为一个函数 $\mathbb{R}\times\mathbb{V}\to\mathbb{V}$, 即

$$(\lambda, \boldsymbol{a}) \mapsto \lambda\boldsymbol{a}.$$

注意, 我们只定义了符号 $\lambda\boldsymbol{a}$ 的含义, 但是并没有定义 $\boldsymbol{a}\lambda$. 一定要注意标量乘法的书写顺序. 有了标量乘法之后, 显然, 一个非零向量 $\boldsymbol{a}$ 所确定的单位向量为

$$\boldsymbol{a}^0=\frac{\boldsymbol{a}}{|\boldsymbol{a}|}=\frac{1}{|\boldsymbol{a}|}\boldsymbol{a}.$$

数乘运算显然满足下列规律 (其中 λ, μ 为任意实数):

$$\begin{aligned}&\lambda(\boldsymbol{a}+\boldsymbol{b})=\lambda\boldsymbol{a}+\lambda\boldsymbol{b},\\&(\lambda\mu)\boldsymbol{a}=\lambda(\mu\boldsymbol{a}),\\&(\lambda+\mu)\boldsymbol{a}=\lambda\boldsymbol{a}+\mu\boldsymbol{a}.\end{aligned}$$

显然, 若对于两个向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$, 存在一个实数 λ, 使得

$$\boldsymbol{a} = \lambda \boldsymbol{b} \quad \text{或} \quad \boldsymbol{b} = \lambda \boldsymbol{a},$$

则 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 共线. 我们约定, 零向量与任何向量均共线.

> 问题 我们非常熟悉实数之间的运算, 特别是加减乘除运算. 现在, 既然有了向量之间的加减运算, 能否去定义向量之间的乘法运算和除法运算? 接下来我们会定义两种截然不同的向量之间的乘法. 一个叫做内积, 一个叫做外积. 尽管他们的输入都是任意两个向量, 但是内积的输出是一个实数, 外积的输出才是一个向量. 向量之间没有什么除法操作. 总之, 把向量的运算类比成实数的运算是非常不推荐的. 读者应该认识到向量和标量的代数运算有着本质不同.

5.1.2 向量之间的内积 (点乘)

现在我们来定义两个向量之间的内积. 设 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 是两个非零向量, 用符号 $\langle \boldsymbol{a}, \boldsymbol{b} \rangle$ 来表示它们之间的夹角. 为了确定起见, 我们约定夹角 $\langle \boldsymbol{a}, \boldsymbol{b} \rangle$ 是 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 所夹的在 $[0, \pi]$ 之间的角. 显然,

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \langle \boldsymbol{b}, \boldsymbol{a} \rangle.$$

显然, $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = 0$ 或 π, 当且仅当两个向量共线. $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \dfrac{\pi}{2}$, 当且仅当两个向量垂直.

定义 5.1.3: 内积 (点乘)

我们按照以下定义任意两个向量 $\boldsymbol{a}, \boldsymbol{b}$ 之间的**内积 (inner product)**, 也叫做**点乘 (dot produtc)**, 数量积.

1. 任意给定两个非零向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$, 我们定义 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的内积是

$$\boldsymbol{a} \cdot \boldsymbol{b} = |\boldsymbol{a}||\boldsymbol{b}| \cos\langle \boldsymbol{a}, \boldsymbol{b} \rangle,$$

   其中 $\langle \boldsymbol{a}, \boldsymbol{b} \rangle \in [0, \pi]$ 是向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 之间的夹角, $|\boldsymbol{a}|$ 和 $|\boldsymbol{b}|$ 是它们的模长.

2. 零向量 $\mathbf{0}$ 与任何向量 $\boldsymbol{a}$ 内积均定义为实数零, 即 $\mathbf{0} \cdot \boldsymbol{a} = \boldsymbol{a} \cdot \mathbf{0} = 0$.

我们可以把内积视作为一个函数 $\mathbb{V} \times \mathbb{V} \to \mathbb{R}$, 即

$$(\boldsymbol{a}, \boldsymbol{b}) \mapsto \boldsymbol{a} \cdot \boldsymbol{b}.$$

在物理学中, 一个恒力 $\boldsymbol{F}$ 沿位移 $\boldsymbol{s}$ 所做的功就是 $\boldsymbol{F} \cdot \boldsymbol{s}$. 根据定义, 给定任意向量 $\boldsymbol{a}$, 都有

$$\boldsymbol{a} \cdot \boldsymbol{a} = |\boldsymbol{a}||\boldsymbol{a}| \cos\langle \boldsymbol{a}, \boldsymbol{a} \rangle = |\boldsymbol{a}|^2 \geq 0.$$

根据点乘运算的定义, 可以验证点乘运算满足下列规律 (其中 λ 为任意实数):

1. $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{b} \cdot \boldsymbol{a}$ (交换律);
2. $(\lambda \boldsymbol{a}) \cdot \boldsymbol{b} = \lambda(\boldsymbol{a} \cdot \boldsymbol{b})$ (与数乘的结合律);
3. $(\boldsymbol{a} + \boldsymbol{b}) \cdot \boldsymbol{c} = \boldsymbol{a} \cdot \boldsymbol{c} + \boldsymbol{b} \cdot \boldsymbol{c}$ (分配律).

结合第一点和第三点, 可知 $\boldsymbol{a}\cdot(\boldsymbol{b}+\boldsymbol{c})=\boldsymbol{a}\cdot\boldsymbol{b}+\boldsymbol{a}\cdot\boldsymbol{c}$. 内积 $\boldsymbol{a}\cdot\boldsymbol{b}$ 可以反映两个向量之间的夹角信息. 根据定义有

$$\cos\langle\boldsymbol{a},\boldsymbol{b}\rangle=\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|},\quad \text{或}\quad \langle\boldsymbol{a},\boldsymbol{b}\rangle=\arccos\left(\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|}\right).$$

注意, cos 函数在 $[0,\pi]$ 上是严格单调的, 所以 $\cos\langle\boldsymbol{a},\boldsymbol{b}\rangle$ 本身就可以反应夹角 $\langle\boldsymbol{a},\boldsymbol{b}\rangle$ 的全部信息. 根据定义, $\cos\langle\boldsymbol{a},\boldsymbol{b}\rangle$ 的正负性和内积 $\boldsymbol{a}\cdot\boldsymbol{b}$ 总是一致的, 所以内积的正负性就对应着夹角是锐角, 直角还是钝角:

- $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的夹角小于 $\frac{\pi}{2}\iff\boldsymbol{a}\cdot\boldsymbol{b}>0$
- $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的夹角等于 $\frac{\pi}{2}\iff\boldsymbol{a}\cdot\boldsymbol{b}=0$
- $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的夹角大于 $\frac{\pi}{2}\iff\boldsymbol{a}\cdot\boldsymbol{b}<0$

因此, 向量点乘运算的引入为描述两个向量之间的垂直关系带来了方便. 我们有以下结论.

结论 5.1.1: 内积判断是否垂直

$$\boldsymbol{a}\text{ 与 }\boldsymbol{b}\text{ 垂直}\iff\boldsymbol{a}\cdot\boldsymbol{b}=0.$$

内积也可以用来描述一个向量在另一个向量上的投影. 考虑向量 $\boldsymbol{a}$ 在向量 $\boldsymbol{b}$ 上的投影, 则定义投影分量[1]为

$$|\boldsymbol{a}|\cos\langle\boldsymbol{a},\boldsymbol{b}\rangle=\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{b}|}.$$

这个表达式给出了 $\boldsymbol{a}$ 沿着 $\boldsymbol{b}$ 的方向的有效分量. 我们还可以考虑一个新向量,

$$\mathrm{proj}_{\boldsymbol{b}}\boldsymbol{a}:=\left(\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{b}|^2}\right)\boldsymbol{b}=\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{b}|}\times\frac{\boldsymbol{b}}{|\boldsymbol{b}|}=\text{ 投影分量}\times\boldsymbol{b}\text{ 所确定的单位向量},$$

这表示 $\boldsymbol{a}$ 在 $\boldsymbol{b}$ 方向上的**投影 (projection) 向量**. 请自行画图, 并当投影分量为负值时意味着什么. 注意到该投影向量的模等于 $\dfrac{|\boldsymbol{a}\cdot\boldsymbol{b}|}{|\boldsymbol{b}|}$.

5.1.3 向量之间的叉乘 (外积)

向量之间的乘法运算除去点乘运算之外, 尚有叉乘运算. 两个向量点乘运算的结果是一个数量, 而现在要定义的叉乘运算却不同: 两个向量叉乘运算的结果是一个新向量.

[1] 当夹角是钝角时, 其值允许为负.

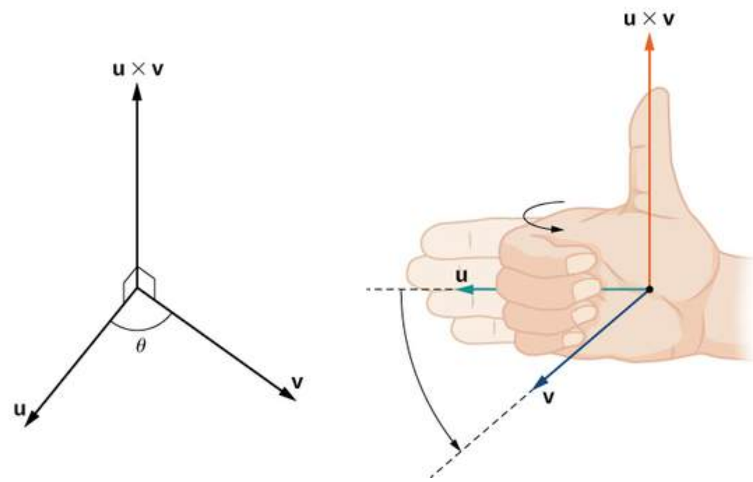


图 5.1: $\mathbf{u}\times\mathbf{v}$ 的方向由右手法则决定. 图片来自 12.4: The Cross Product - Mathematics LibreTexts

定义 5.1.4: 外积 (叉乘)

我们按照以下定义任意两个向量 $\boldsymbol{a},\boldsymbol{b}$ 之间的**叉乘 (cross produtc)**, 也叫做**外积 (outter product)**.[a]

1. 设 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 是两个不共线的向量[b], 我们定义 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的叉乘运算如下:
   (a) 叉乘的结果 $\boldsymbol{a}\times\boldsymbol{b}$ 是一个新的向量,
   (b) 其垂直于 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 所决定的平面 (把 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的起点放在一起),
   (c) 其方向根据*右手系法则*决定, 即用右手的四指从 $\boldsymbol{a}$ 握向 $\boldsymbol{b}$ 时, 拇指的方向作为 $\boldsymbol{a}\times\boldsymbol{b}$ 的方向 (见图 5.1),
   (d) 其大小定义为
$$|\boldsymbol{a}\times\boldsymbol{b}|=|\boldsymbol{a}||\boldsymbol{b}|\sin\langle\boldsymbol{a},\boldsymbol{b}\rangle.$$
2. 若 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 共线, 则我们约定 $\boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{b}\times\boldsymbol{a}=\boldsymbol{0}$.

我们可以把外积视作为一个函数 $\mathbb{V}\times\mathbb{V}\to\mathbb{V}$, 即
$$(\boldsymbol{a},\boldsymbol{b})\mapsto\boldsymbol{a}\times\boldsymbol{b}.$$
从函数视角, 外积的定义域和到达域和向量加法一样, 但是规则完全不同.

[a] 注意叉乘的定义迫使我们必须在空间中考虑向量.
[b] 不共线天然地排出了零向量, 因为零向量和任何向量都共线.

显然, 根据叉乘运算的定义, 不论 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 是否共线, $|\boldsymbol{a}\times\boldsymbol{b}|$ 总等于 $|\boldsymbol{a}||\boldsymbol{b}|\sin\langle\boldsymbol{a},\boldsymbol{b}\rangle$. 特别是当 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 共线时, $\langle\boldsymbol{a},\boldsymbol{b}\rangle$ 是 0 或 π. 可以验证, 以上方式定义的新向量 $\boldsymbol{a}\times\boldsymbol{b}$ 对于任意一对向量 $\boldsymbol{a},\boldsymbol{b}$ 都是唯一存在的, 因此是良好定义的. 我们称 $\boldsymbol{a}\times\boldsymbol{b}$ 为 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的外积.

定义 5.1.4 对叉乘的定义方式, 非常的几何, 这一点和常见的数学定义很不同. 叉乘是物理学中常见的工具. 但因为它的定义严重依赖于几何空间, 所以它一般只是三维空间中两个向量之间的一种运算操作.

当 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 不共线时, $\boldsymbol{a}\times\boldsymbol{b}$ 的模 $|\boldsymbol{a}\times\boldsymbol{b}|$ 恰好是以 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 为邻边的平行四边形的面积. 见动画 Cross

Product and Area Visualization –GeoGebra. 叉乘运算满足以下规律 (其中 λ 为任意实数):

1. $\boldsymbol{a}\times\boldsymbol{b}=-(\boldsymbol{b}\times\boldsymbol{a})$ (反交换律);
2. $\lambda(\boldsymbol{a}\times\boldsymbol{b})=(\lambda\boldsymbol{a})\times\boldsymbol{b}=\boldsymbol{a}\times(\lambda\boldsymbol{b})$ (与数乘的结合律);[2]
3. $(\boldsymbol{a}+\boldsymbol{b})\times\boldsymbol{c}=\boldsymbol{a}\times\boldsymbol{c}+\boldsymbol{b}\times\boldsymbol{c}$ (分配律).

第一点告诉我们, 叉乘是依赖于顺序的. $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的叉乘不等于 $\boldsymbol{b}$ 与 $\boldsymbol{a}$; 与之对比, 内积的结果是不依赖于顺序的. 第三点的分配律参考动画 叉乘的分配律 - GeoGebra 理解即可, 其证明比较复杂, 有兴趣的读者可参考资料 4100AWL/Thomas_APPp001-034. 下一节, 我们将通过代数方法来验证分配律的正确性 (见例题 5.2.7). 此外, 既然叉乘是有顺序的, 那么分配律是否也有顺序呢? 实则不然, 因为

$$\begin{aligned}\boldsymbol{a}\times(\boldsymbol{b}+\boldsymbol{c})&=-((\boldsymbol{b}+\boldsymbol{c})\times\boldsymbol{a})\\&=-(\boldsymbol{b}\times\boldsymbol{a}+\boldsymbol{c}\times\boldsymbol{a})\\&=-\boldsymbol{b}\times\boldsymbol{a}-\boldsymbol{c}\times\boldsymbol{a}\\&=\boldsymbol{a}\times\boldsymbol{b}+\boldsymbol{a}\times\boldsymbol{c}.\end{aligned}$$

所以分配律在第二个向量位置上仍然成立, 即,

$$\boldsymbol{a}\times(\boldsymbol{b}+\boldsymbol{c})=\boldsymbol{a}\times\boldsymbol{b}+\boldsymbol{a}\times\boldsymbol{c}.$$

另外, 向量的叉乘运算不满足交换律. 不仅如此, 这种乘法运算还不满足结合律, 即一般等式 $(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}=\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c})$ 不成立. 参考 叉乘的不满足结合律 - GeoGebra 理解即可.

我们已经看到向量点乘运算的引入为描述向量之间的垂直关系带来方便, 现在向量叉乘运算的引入则为描述向量共线带来了方便. 我们有以下结论.

结论 5.1.2: 外积判断是否共线

$$\boldsymbol{a} \text{ 与 } \boldsymbol{b} \text{ 共线} \iff \boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{0}.$$

任意向量都和自身共线, 故有 $\boldsymbol{a}\times\boldsymbol{a}=\boldsymbol{0}$.

5.1.4 向量之间的混合积

定义 5.1.5

设 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 为任意的三个向量, 那么

$$\boldsymbol{a}\cdot(\boldsymbol{b}\times\boldsymbol{c})$$

是一个实数, 称之为 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 的混合积. 这种先做叉乘再做点乘的运算称为**混合积 (mixed product)**, 也叫三重积 (triple product).

问题 先做叉乘再做点乘的运算称为混合积, 那么先点乘做再做叉乘的运算是否存在呢?

[2]于是我们可以没有歧义地写作 $\boldsymbol{a}\times\boldsymbol{b}=-\boldsymbol{b}\times\boldsymbol{a}$.

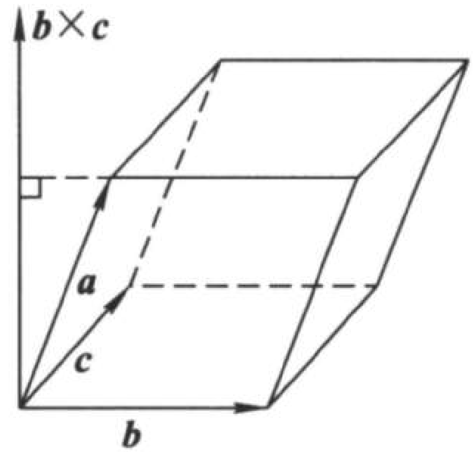


图 5.2: $\boldsymbol{a}\cdot(\boldsymbol{b}\times\boldsymbol{c})$ 的几何意义.

若三个非零向量 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 不共面, 则以它们为棱构成一个平行六面体. 见图 5.2 和动画 Scalar triple product –GeoGebra. 根据之前的讨论, $|\boldsymbol{b}\times\boldsymbol{c}|$ 是这个平行六面体的一个底面积 (该底就是向量 $\boldsymbol{b},\boldsymbol{c}$ 形成的平行四边形), 而 $\boldsymbol{a}$ 在叉乘 $\boldsymbol{b}\times\boldsymbol{c}$ 上的投影向量的模恰好就是这个平行六面体的高. 而根据内积的定义, 我们有

$$|\boldsymbol{a}\cdot(\boldsymbol{b}\times\boldsymbol{c})|=|(\boldsymbol{b}\times\boldsymbol{c})\cdot\boldsymbol{a}|=|\boldsymbol{b}\times\boldsymbol{c}|\underbrace{|\boldsymbol{a}|\,|\cos\langle\boldsymbol{b}\times\boldsymbol{c},\boldsymbol{a}\rangle|}_{\text{以}\boldsymbol{b},\boldsymbol{c}\text{为底的高}}.$$

综上所得, 混合积的绝对值等于由 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 所形成的平行六面体的体积. 混合积运算具有下列性质:

- 混合积在其三个运算数 $\mathbf{a},\mathbf{b},\mathbf{c}$ 的循环移位[3]下保持不变:

  $$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})=\mathbf{b}\cdot(\mathbf{c}\times\mathbf{a})=\mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}).$$

  从几何意义上, 就是在计算平行六面体时, 选择了不同的底面.

- 在不重新排列运算数 $\mathbf{a},\mathbf{b},\mathbf{c}$ 的情况下交换运算符的位置不会改变混合积的值. 这是由上一性质以及点积的交换律得出的:

  $$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})=(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}.$$

- 交换三个运算数中的任意两个会使混合积的值取负. 这是由循环移位性质以及叉乘的反交换律得出的:

  $$\begin{aligned}\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})&=-\mathbf{a}\cdot(\mathbf{c}\times\mathbf{b})\\&=-\mathbf{b}\cdot(\mathbf{a}\times\mathbf{c})\\&=-\mathbf{c}\cdot(\mathbf{b}\times\mathbf{a}).\end{aligned}$$

根据 “三个非零向量的混合积的绝对值是这三个向量形成的平行六面体的体积” 这一性质, 我们有以下结论.

结论 5.1.3: 混合积判断是否共面

$$\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}\text{ 共面}\iff\boldsymbol{a}\cdot(\boldsymbol{b}\times\boldsymbol{c})=0.$$

[3]即, $abc\to bca\to cab\to abc$

5.1.5 广义的向量空间

本小节是额外学习资料. 考试不考.

向量空间 (Vector Space), 也称为线性空间 (Linear Space), 是一个集合, 其中的元素称为向量[4]. 《线性代数》就是研究线性空间和线性函数的一门工具, 其应用极其极其的广泛. 例如, 电脑文件的压缩 zip, 美颜修图的算法, 等等都直接使用线性代数的知识. 本节的目的就是带读者初识一下什么是线性空间.

> **注释 5.1.1** Interactive Linear Algebra 是非常容易阅读和上手的可交互式线性代数教材. 此外, 但凡哪本线性代数的第一章就是讲行列式的都可以扔了.

向量空间的定义

设 V 是一个集合, 首先我们必须在 V 上定义两个运算:

1. 向量加法: 对任意两个元素 $\mathbf{u}, \mathbf{v} \in V$, 可以进行加法运算, 得到一个新的元素 $\mathbf{u}+\mathbf{v} \in V$.
2. 标量乘法: 对任意标量 $\alpha \in \mathbb{R}$, 以及任意元素 $\mathbf{v} \in V$, 可以进行标量乘法, 得到一个新的元素 $\alpha\mathbf{v} \in V$.

以上两种运算在不同集合上有各自的实现方式, “加法”, “乘法” 只是一个名字, 它不一定是我们熟悉的加乘. **任何一个集合 V, 如果定义了向量加法和标量乘法, 且这两个运算满足以下 8 条公理, 那么集合 V 就称为一个向量空间, 或线性空间. 此时, V 中的元素称为向量.** 这 8 条公理如下所示.

1. 加法交换律: 对于任意 $\mathbf{u}, \mathbf{v} \in V$, 有
$$\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}.$$
2. 加法结合律: 对于任意 $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, 有
$$(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w}).$$
3. 存在加法单位元: 存在一个特殊的元素 $\mathbf{0} \in V$, 使得对于任意 $\mathbf{v} \in V$, 有
$$\mathbf{v}+\mathbf{0}=\mathbf{v}.$$
4. 存在加法逆元: 对于每个 $\mathbf{v} \in V$, 存在一个唯一与之对应的元素 $-\mathbf{v} \in V$, 使得
$$\mathbf{v}+(-\mathbf{v})=\mathbf{0}.$$
5. 标量乘法结合律: 对于任意标量 $\alpha, \beta \in \mathbb{R}$ 和 $\mathbf{v} \in V$, 有
$$\alpha(\beta\mathbf{v})=(\alpha\beta)\mathbf{v}.$$
6. 标量乘法分配律 (对向量加法): 对于任意标量 $\alpha \in \mathbb{R}$ 和 $\mathbf{u}, \mathbf{v} \in V$, 有
$$\alpha(\mathbf{u}+\mathbf{v})=\alpha\mathbf{u}+\alpha\mathbf{v}.$$
7. 标量乘法分配律 (对标量加法): 对于任意标量 $\alpha, \beta \in \mathbb{R}$ 和 $\mathbf{v} \in V$, 有
$$(\alpha+\beta)\mathbf{v}=\alpha\mathbf{v}+\beta\mathbf{v}.$$

[4] 这里的 “向量” 一词不是前面的向量, 它的内涵要比 “有大小的方向的量” 要广义地多.

8. 单位标量的作用: 对于任意向量 $\mathbf{v} \in V$, 有
$$1 \cdot \mathbf{v} = \mathbf{v}.$$

读者已经发现了, 我们前面所讨论的 "平面或空间中有大小和方向的向量" 的集合:

$$\mathbb{V} = \{\boldsymbol{v} \mid \boldsymbol{v} \text{ 是一个自由向量, 具有大小和方向}\},$$

正是一个特殊的向量空间. 一个向量空间的定义只需要集合自身, 再配套上一组向量加法和标量乘法的运算. 前面所定义的内积和外积的运算, 是独属于向量空间 $\mathbb{V}$ 上有意义的额外操作.

向量空间的例子

1. 欧几里得空间 $\mathbb{R}^n$

- 定义: 欧几里得空间是最常见的向量空间之一, 表示的是 n-维实数空间. 空间中的元素是所有 n-维列向量 (或者说是 n-元组), 其形式为 $(x_1, x_2, \ldots, x_n)$, 其中每个 $x_i \in \mathbb{R}$[5], 即每个分量都是一个实数.
- 向量加法和标量乘法: 在欧几里得空间中, 向量加法和标量乘法分别遵循以下规则:
  - 向量加法: $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ 和 $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ 的和是
  $$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n).$$
  - 标量乘法: 标量 α 和向量 $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ 的乘积是
  $$\alpha\mathbf{u} = (\alpha u_1, \alpha u_2, \ldots, \alpha u_n).$$
- 加法单位元: 在 $\mathbb{R}^n$ 中, 加法单位元是零向量 $\mathbf{0} = (0, 0, \ldots, 0)$. 对于任意向量 $\mathbf{v} = (v_1, v_2, \ldots, v_n)$, 都有
$$\mathbf{v} + \mathbf{0} = \mathbf{v}.$$
- 加法逆元: 对于任意向量 $\mathbf{v} = (v_1, v_2, \ldots, v_n)$, 加法逆元是向量 $-\mathbf{v} = (-v_1, -v_2, \ldots, -v_n)$, 使得
$$\mathbf{v} + (-\mathbf{v}) = \mathbf{0}.$$
- 举例:
  - $\mathbb{R}^2$: 二维平面中的向量空间. 例如, 向量 $\mathbf{u} = (1, 2)$ 和 $\mathbf{v} = (3, 4)$ 可以加法得到
  $$\mathbf{u} + \mathbf{v} = (1 + 3, 2 + 4) = (4, 6).$$
  零向量是 $\mathbf{0} = (0, 0)$, 加法逆元是 $-\mathbf{u} = (-1, -2)$.
  - $\mathbb{R}^3$: 三维空间中的向量空间. 例如, $\mathbf{u} = (1, 2, 3)$ 和 $\mathbf{v} = (4, 5, 6)$ 可以加法得到
  $$\mathbf{u} + \mathbf{v} = (1 + 4, 2 + 5, 3 + 6) = (5, 7, 9).$$
  零向量是 $\mathbf{0} = (0, 0, 0)$, 加法逆元是 $-\mathbf{u} = (-1, -2, -3)$.
- 欧几里得空间符合所有向量空间的公理, 包括加法交换律, 加法结合律, 标量乘法结合律等, 因此是一个向量空间.

[5]本讲义, 将实数域的符号有时写做 $\mathbb{R}$, 也有时写做 $\mathbf{R}$, 只是单纯地没时间统一起来. 勿怪.

2. 多项式空间

- 定义: 多项式空间是由所有次数不超过 n 的多项式组成的向量空间. 设有多项式空间 P_n, 其中的元素是所有次数不超过 n 的多项式, 例如 $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, 其中 $a_0, a_1, \ldots, a_n$ 是实数系数.
- 向量加法和标量乘法: 在多项式空间中, 向量加法和标量乘法按多项式的常规加法和标量乘法定义:
  - 加法: 两个多项式 $p(x) = a_0 + a_1x + \ldots$ 和 $q(x) = b_0 + b_1x + \ldots$ 的和是
  $$p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + \ldots.$$
  - 标量乘法: 标量 α 和多项式 $p(x) = a_0 + a_1x + \ldots$ 的乘积是
  $$\alpha p(x) = \alpha a_0 + \alpha a_1 x + \ldots.$$
- 加法单位元: 在多项式空间中, 加法单位元是零多项式 $p(x) = 0$, 它使得对于任意多项式 $q(x)$, 都有
$$q(x) + 0 = q(x).$$
- 加法逆元: 对于任意多项式 $p(x) = a_0 + a_1x + \ldots$, 加法逆元是多项式 $-p(x) = -a_0 - a_1x - \ldots$, 使得
$$p(x) + (-p(x)) = 0.$$
- 举例:
  - P_2 中的一个向量是多项式 $p(x) = 2 + 3x + x^2$, 另一个向量 $q(x) = 1 - x + 2x^2$, 它们的和是:
  $$p(x) + q(x) = (2 + 1) + (3 - 1)x + (1 + 2)x^2 = 3 + 2x + 3x^2.$$
  零多项式是 0, 加法逆元是 $-p(x) = -2 - 3x - x^2$.
- 多项式加法和标量乘法满足向量空间的所有公理, 包括加法交换律, 加法结合律, 标量乘法分配律等.

3. 函数空间

- 定义: 函数空间是由在某个区间上的所有连续函数组成的向量空间. 假设 I 是一个区间, 函数空间 $F(I)$ 包含所有在区间 I 上连续的实值函数. 每个元素是一个函数, 例如 $f(x)$, $g(x)$ 等.
- 向量加法和标量乘法: 在函数空间中, 向量加法和标量乘法是按函数的加法和标量乘法进行的:
  - 加法: 对于两个函数 $f(x)$ 和 $g(x)$, 它们的和是 $f(x) + g(x)$, 即将它们的值逐点相加:
  $$(f + g)(x) = f(x) + g(x).$$
  - 标量乘法: 标量 α 和函数 $f(x)$ 的乘积是 $\alpha f(x)$, 即将函数的值逐点乘以标量:
  $$(\alpha f)(x) = \alpha f(x).$$
- 加法单位元: 在函数空间中, 加法单位元是零函数 $f(x) = 0$, 它使得对于任意函数 $g(x)$, 都有
$$g(x) + 0 = g(x).$$

- 加法逆元: 对于任意函数 $f(x)$, 加法逆元是函数 $-f(x)$, 使得

$$f(x) + (-f(x)) = 0.$$

- 举例:
  - 假设区间 $I = [0, 1]$, 那么 $f(x) = x^2$ 和 $g(x) = \sin x$ 是 I 上的两个函数, 它们的和是:

  $$(f + g)(x) = f(x) + g(x) = x^2 + \sin x.$$

    零函数是 0, 加法逆元是 $-f(x) = -x^2$, $-g(x) = -\sin x$.
  - 对于标量乘法, 假设标量 $\alpha = 3$, 则 $\alpha f(x) = 3x^2$.
- 函数空间中的加法和标量乘法符合向量空间的公理. 特别地, 函数的加法和标量乘法都满足交换律, 结合律以及分配律等.

这三个例子展示了向量空间的不同形式, 虽然它们的元素和操作规则不同, 但它们都满足向量空间的公理 (加法和标量乘法的性质). 这些例子中的元素 (向量) 可以是实数, 函数, 甚至更复杂的对象, 只要它们遵循加法和标量乘法的规定, 就构成了一个向量空间.

5.2　向量的空间坐标

5.2.1　向量的坐标表示

本小节的整体叙事逻辑如下. 上一节中我们讨论的向量空间是

$$\mathbb{V} = \{\boldsymbol{v} \mid \boldsymbol{v}\ \text{是一个自由向量, 具有大小和方向}\}.$$

此时, 我们定义由三个实数构成的所有有序数组的集合, 记为 $\mathbf{R}^3$, 即

$$\mathbf{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbf{R}\}.$$

接下来, 我们将证明以下两点:

1. 空间中全体点 $\xLeftrightarrow{\text{一一对应}}$ $\mathbf{R}^3$ (通过建立空间直角坐标系)
2. 空间中全体点 $\xLeftrightarrow{\text{一一对应}}$ $\mathbb{V}$ (通过将向量 $\boldsymbol{v}$ 的起点移至坐标原点 O)

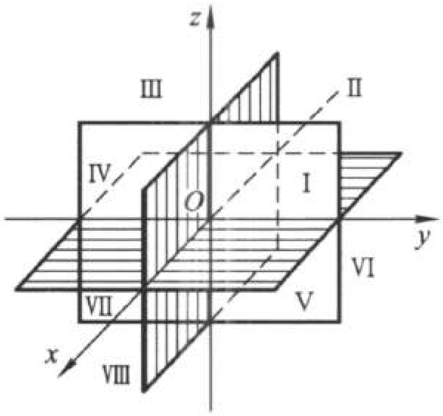
最终我们的目的在于得到向量的坐标表示, 即证明

$$\mathbb{V} \xLeftrightarrow{\text{一一对应}} \mathbf{R}^3.$$

空间中全体点和三元有序数组 $\mathbf{R}^3$ 的一一对应

平面直角坐标系 Oxy 使得平面上的全体点与二元有序数组 $\mathbf{R}^2 = \{(x, y) \mid x, y \in \mathbf{R}\}$ 建立了一一对应的关系. 因此, 我们通常把平面上的全体点与 $\mathbf{R}^2$ 等同起来, 这正像我们过去把 $\mathbf{R}$ 与数轴等同起来一样. 类似地, 我们可以利用空间直角坐标系 $Oxyz$ 使得空间中的全体点与三元有序数组 $\mathbf{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbf{R}\}$ 建立一一对应的关系.

为此, 首先我们需要在空间中引进*直角坐标系*. 具体做法如下:



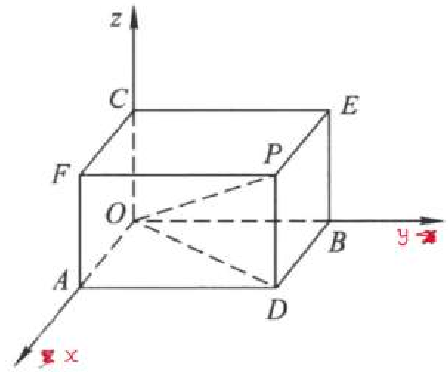


图 5.3: 空间直角坐标系.

- 在空间中取定一点 O, 过该点作三条相互垂直的直线, 并在三条直线上取定相同的单位长度 (即刻度一致) 和各自的方向 (即规定哪里是正向), 使它们成为三条数轴. 然后, 按右手系法则[6]将这三条数轴分别称作 x 轴, y 轴及 z 轴: 将右手握着 z 轴, 使拇指指向 z 轴的正向, 这时其他手指的指向是从 x 轴的正向到 y 轴的正向. 这样, 我们就在空间中建立了一个直角坐标系, 记为 $Oxyz$.
- 点 O 称为坐标原点 (简称原点).
- x 轴, y 轴, z 轴分别称为横轴, 纵轴, 立轴, 它们统称为坐标轴.
- 每两条坐标轴决定一个平面, 称作坐标平面, 其中 x 轴与 y 轴决定的坐标平面称为 Oxy 平面, 相应地还有 Oyz 平面和 Ozx 平面.[7]
- 三个坐标平面将空间分作八部分, 每一部分称作一个卦限, 如图 5.3 所示.

有了空间直角坐标系 $Oxyz$ 后, 我们有以下的观察.

1. 设 P 是空间中的任意一点. 通过点 P 作三个平面, 分别垂直于三条坐标轴, 得到的三个平面与 x 轴, y 轴和 z 轴分别交于点 A, B 和 C (见图 5.3). 这里用到一个基本的几何知识: 过某一点并且与给定直线垂直的平面是唯一的. 换句话说, 点 P 在这三条坐标轴上投影, 得到的投影点分别为 A, B 和 C (见图 5.3). 接着, 我们定义点 A, B, C 在各自坐标轴上的坐标分别为 x, y, z. 因此, 对于空间中任意给定的点 P, 我们能够唯一确定一个三元有序数组 (x, y, z). 我们称 (x, y, z) 为点 P 的坐标, 其中 x, y, z 分别称作点 P 的横坐标, 纵坐标和立坐标.

2. 反过来, 对于任意给定的三元有序数组 (x, y, z), 我们可以唯一确定空间中的一个点 P. 具体方法是: 首先, 在 x 轴, y 轴和 z 轴上分别取坐标为 x, y, z 的三个点 A, B 和 C. 然后, 分别通过这三个点作与各自坐标轴垂直的平面. 由于这三个坐标轴是相互垂直的, 因此这三个平面相交的部分只能是一个点, 这个交点即为点 P (见图 5.3).

这样, 我们便把空间中的每一点都对应于一个由三个实数构成的有序数组.

[6]这里的右手系法则和前一节中定义叉乘时的一样. 这意味还有别的顺序来安排这三条数轴. 但我们从来不用.

[7]一般不写做 Oyx 平面, Ozy 平面和 Oxz 平面. 这里隐藏这一个习惯: 如果只考虑坐标平面 Oxy 平面, 则 x 是自变量, y 是因变量. 如果只考虑坐标平面 Oyz 平面, 则 y 是自变量, z 是因变量. 如果只考虑坐标平面 Ozx 平面, 则 z 是自变量, x 是因变量.

结论 5.2.1: 空间中全体点和三元有序数组 R^3 的一一对应

我们把全体三元有序数组的集合记为 $\mathbf{R}^3$, 即

$$\mathbf{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbf{R}\}.$$

这样, 根据前面的说明, 空间中全体点和 $\mathbf{R}^3$ 是一一对应. 通常我们将空间中的全体点与 $\mathbf{R}^3$ 等同起来.

向量空间 $\mathbb{V}$ 和三元有序数组 R^3 的一一对应

我们注意到, 空点中的全体点和向量空间 $\mathbb{V} = \{\boldsymbol{v} \mid \boldsymbol{v}$ 是一个自由向量, 具有大小和方向 $\}$ 有着非常自然的一一对应关系:

- 给定空间中的任意一点 P, 可以确定一个向量 $\overrightarrow{OP} \in \mathbb{V}$. 因为是自由向量, $\overrightarrow{OP}$ 可以在空间任意平移.
- 给定任意向量 $\boldsymbol{v} \in \mathbb{V}$, 通过平移, 将其起点移至坐标原点 O, 则可以确定一个终点. 此终点自然是空间中的一点.

这样就对空间中的点集合与空间中的全体向量建立了一个一一对应. 在上一小节, 我们定义了空间中任意点的坐标 (x, y, z). 于是, 通过刚才的讨论, 我们就可以定义向量的坐标为其对应的空间中点的坐标. 具体定义如下.

定义 5.2.1: 向量的坐标表示

1. 给定任意向量 $\overrightarrow{OP}$, 其终点 P 的坐标 (x, y, z) 就是向量 $\overrightarrow{OP}$ 的坐标.
2. 给定任意向量 $\boldsymbol{a}$, 通过平移, 将其起点移至坐标原点 O, 则其终点的坐标 (x, y, z) 就是向量 $\boldsymbol{a}$ 的坐标.

现在我们来解释向量 $\overrightarrow{OP}$ 的坐标 (x, y, z) 的意义. 为此, 我们分别在 x 轴, y 轴, z 轴上取定单位向量 $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$, 使其方向恰与各自坐标轴的正向相同. 称 $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ 为坐标向量. 这时, 根据点 P 的坐标的定义不难看出, $x\boldsymbol{i} + y\boldsymbol{j}$ 是向量 $\overrightarrow{OP}$ 在 Oxy 平面上的投影向量, 而 $x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$ 则恰好是 $\overrightarrow{OP}$ (见图 5.4), 即

$$\overrightarrow{OP} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}.$$

反过来, 若 $\overrightarrow{OP}$ 可以表示成

$$\overrightarrow{OP} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k},$$

则 $\overrightarrow{OP}$ 的坐标一定是 (x, y, z). 根据前面的讨论, 向量 $\boldsymbol{a}$ 的坐标为 (x, y, z) 的充要条件是 $\boldsymbol{a}$ 可以表示成

$$\boldsymbol{a} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}.$$

这个式子称为 $\boldsymbol{a}$ **关于三个坐标向量的分解式**.

结论 5.2.2: 向量的坐标表示和分解式

分别在 x 轴, y 轴, z 轴上取定单位向量 $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$, 使其方向恰与各自坐标轴的正向相同. 称 $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ 为**坐标向量**. 今后, 我们把向量 $\boldsymbol{a}$ 与其坐标 (x, y, z) 等同起来, 直接写成 $\boldsymbol{a} = (x, y, z)$. 即,

$$\boldsymbol{a} = (x, y, z) \Longleftrightarrow \boldsymbol{a} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}.$$

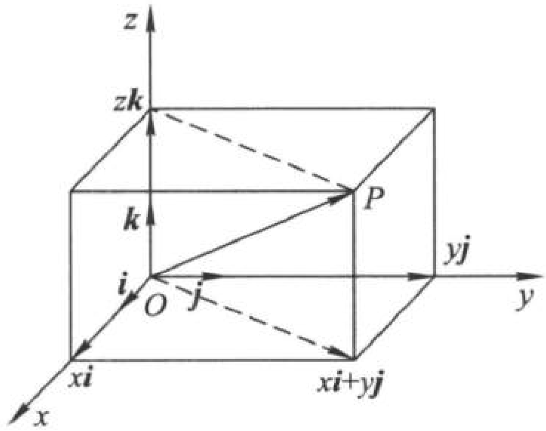


图 5.4: 向量的坐标表示

既然向量自身可以通过坐标来表示了, 那么上一节中定义的诸多向量的运算也就可以用坐标之间的运算来表示, 这极大地减少了对几何的依赖性. 首先, 我们给出向量加法和标量乘法运算的坐标表示.

命题 5.2.1: 向量加法, 标量乘法的坐标表示

1. 设向量 $\boldsymbol{a}=(x_1,y_1,z_1)$, $\boldsymbol{b}=(x_2,y_2,z_2)$, 则有

$$\boldsymbol{a}+\boldsymbol{b}=(x_1+x_2,y_1+y_2,z_1+z_2).$$

2. 另外, 若向量 $\boldsymbol{a}=(x,y,z)$, 则对于任意实数 λ, 有

$$\lambda\boldsymbol{a}=(\lambda x,\lambda y,\lambda z).$$

证明. 1. 向量 $\boldsymbol{a}=(x_1,y_1,z_1)$, $\boldsymbol{b}=(x_2,y_2,z_2)$, 则有

$$\boldsymbol{a}=x_1\boldsymbol{i}+y_1\boldsymbol{j}+z_1\boldsymbol{k},\quad \boldsymbol{b}=x_2\boldsymbol{i}+y_2\boldsymbol{j}+z_2\boldsymbol{k}.$$

故由向量加法运算的性质有

$$\boldsymbol{a}+\boldsymbol{b}=(x_1+x_2)\,\boldsymbol{i}+(y_1+y_2)\,\boldsymbol{j}+(z_1+z_2)\,\boldsymbol{k}.$$

可见, $(x_1+x_2,y_1+y_2,z_1+z_2)$ 是 $\boldsymbol{a}+\boldsymbol{b}$ 的坐标.

2. 向量 $\boldsymbol{a}=(x,y,z)$, 则有

$$\boldsymbol{a}=x\boldsymbol{i}+y\boldsymbol{j}+z\boldsymbol{k}.$$

故由标量乘法的法则有 $\lambda\boldsymbol{a}=(\lambda x)\boldsymbol{i}+(\lambda y)\boldsymbol{j}+(\lambda z)\boldsymbol{k}$. 所以, $\lambda\boldsymbol{a}=(\lambda x,\lambda y,\lambda z)$.

□

5.2.2 内积的坐标表示

接下来我们考虑内积的坐标表示. 设向量 $\boldsymbol{a}=(x_1,y_1,z_1)$, $\boldsymbol{b}=(x_2,y_2,z_2)$, 即

$$\boldsymbol{a}=x_1\boldsymbol{i}+y_1\boldsymbol{j}+z_1\boldsymbol{k},\quad \boldsymbol{b}=x_2\boldsymbol{i}+y_2\boldsymbol{j}+z_2\boldsymbol{k}.$$

注意到点乘运算的性质以及对于三个坐标向量有

$$\boldsymbol{i}\cdot\boldsymbol{i}=\boldsymbol{j}\cdot\boldsymbol{j}=\boldsymbol{k}\cdot\boldsymbol{k}=1,\quad \boldsymbol{i}\cdot\boldsymbol{j}=\boldsymbol{j}\cdot\boldsymbol{k}=\boldsymbol{k}\cdot\boldsymbol{i}=0,$$

即得

$$\boldsymbol{a}\cdot\boldsymbol{b}=(x_1\boldsymbol{i}+y_1\boldsymbol{j}+z_1\boldsymbol{k})\cdot(x_2\boldsymbol{i}+y_2\boldsymbol{j}+z_2\boldsymbol{k})=x_1x_2+y_1y_2+z_1z_2.$$

这就是说, 两个向量的内积是它们的对应坐标分量的乘积之和.

命题 5.2.2: 内积的坐标表示

设向量 $\boldsymbol{a}=(x_1,y_1,z_1)$, $\boldsymbol{b}=(x_2,y_2,z_2)$, 则有

$$\boldsymbol{a}\cdot\boldsymbol{b}=x_1x_2+y_1y_2+z_1z_2.$$

内积的坐标表示有很多用途. 设向量 $\boldsymbol{a}=(x,y,z)$. 由定义有

$$\boldsymbol{a}\cdot\boldsymbol{a}=|\boldsymbol{a}|^2\cos\langle\boldsymbol{a},\boldsymbol{a}\rangle=|\boldsymbol{a}|^2.$$

若记 $\boldsymbol{a}\cdot\boldsymbol{a}=\boldsymbol{a}^2$, 则有 $\boldsymbol{a}^2=|\boldsymbol{a}|^2$. 由上式可得

$$|\boldsymbol{a}|=\sqrt{\boldsymbol{a}\cdot\boldsymbol{a}}=\sqrt{x^2+y^2+z^2}.$$

若 $\boldsymbol{a}\neq\boldsymbol{0}$, 由此可推出沿 $\boldsymbol{a}$ 方向的单位向量 $\boldsymbol{a}^0$ 的坐标表示:

$$\boldsymbol{a}^0=\frac{1}{|\boldsymbol{a}|}\boldsymbol{a}=\frac{1}{\sqrt{x^2+y^2+z^2}}(x,y,z)=\left(\frac{x}{\sqrt{x^2+y^2+z^2}},\frac{y}{\sqrt{x^2+y^2+z^2}},\frac{z}{\sqrt{x^2+y^2+z^2}}\right).$$

例 5.2.1 设点 P_1 的坐标为 (x_1,y_1,z_1), 而点 P_2 的坐标为 (x_2,y_2,z_2), 证明: 点 P_1 到点 P_2 的距离为

$$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}.$$

证明. 因为 $\overrightarrow{P_1P_2}=\overrightarrow{OP_2}-\overrightarrow{OP_1}$, 所以向量 $\overrightarrow{P_1P_2}$ 的坐标等于向量 $\overrightarrow{OP_2}-\overrightarrow{OP_1}$ 的坐标, 即 $(x_2-x_1,y_2-y_1,z_2-z_1)$, 因而 $\overrightarrow{P_1P_2}$ 的模为

$$\left|\overrightarrow{P_1P_2}\right|=\sqrt{\overrightarrow{P_1P_2}\cdot\overrightarrow{P_1P_2}}=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2},$$

$\left|\overrightarrow{P_1P_2}\right|$ 也就是点 P_1 到点 P_2 的距离. 证毕. □

例 5.2.2 已知空间中有三点 $A(1,0,1)$, $B(0,1,1)$, $C(1,-1,1)$, 求 $\overrightarrow{AB}$ 与 $\overrightarrow{AC}$ 之间的夹角.

证明. 由已知条件可计算出 $\overrightarrow{AB}=\overrightarrow{OB}-\overrightarrow{OA}=(-1,1,0)$, $\overrightarrow{AC}=\overrightarrow{OC}-\overrightarrow{OA}=(0,-1,0)$. 设 $\theta=\langle\overrightarrow{AB},\overrightarrow{AC}\rangle$, 那么

$$\cos\theta=\frac{\overrightarrow{AB}\cdot\overrightarrow{AC}}{|\overrightarrow{AB}||\overrightarrow{AC}|}.$$

而 $\overrightarrow{AB}\cdot\overrightarrow{AC}=-1$, $|\overrightarrow{AB}|=\sqrt{2}$, $|\overrightarrow{AC}|=1$, 故 $\cos\theta=-\dfrac{\sqrt{2}}{2}$, $\theta=\dfrac{3\pi}{4}$. □

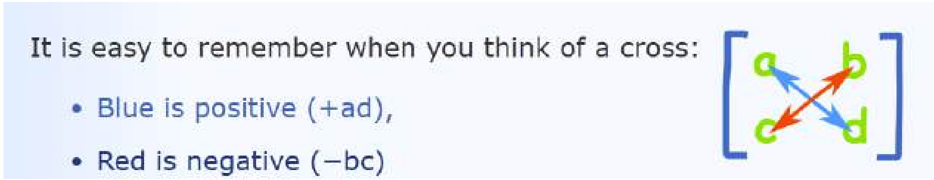


图 5.5: Determinant of a Matrix

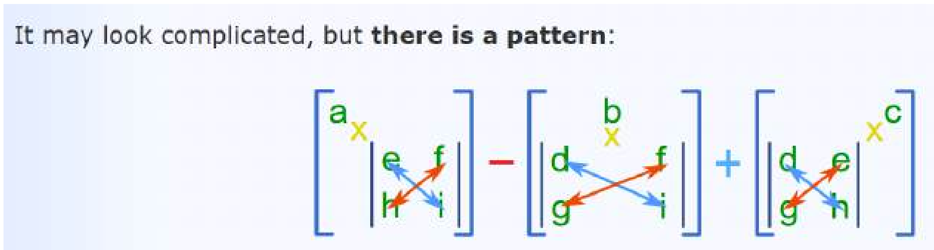


图 5.6: Determinant of a Matrix

5.2.3 外积的坐标表示

现在我们讨论外积的坐标表示. 为此, 我们需要行列式的概念.

只需要一点的行列式知识

下面所有小写字母都是实数. 形如

$$\begin{vmatrix} x & y \\ x_1 & y_1 \end{vmatrix}$$

的数学式称为**二阶行列式**, 我们规定这个数学式的值等于 $xy_1 - yx_1$, 即 (见图 5.5)

$$\begin{vmatrix} x & y \\ x_1 & y_1 \end{vmatrix} := xy_1 - yx_1.$$

例 5.2.3

$$\begin{vmatrix} 4 & 6 \\ 3 & 8 \end{vmatrix} = 4 \times 8 - 6 \times 3 = 32 - 18 = 14.$$

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = 1 \times 4 - 2 \times 3 = 4 - 6 = -2.$$

三阶行列式是指

$$\begin{vmatrix} x & y & z \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix},$$

它的值定义为 (见图 5.6)

$$\begin{vmatrix} x & y & z \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix} := x \begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix} - y \begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix} + z \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}.$$

所以三阶行列式可以归结为计算三个二阶行列式, 而二阶行列式的计算是非常简单的. 注意, 中间的那个 y 前面必须是负号!

例 5.2.4

$$\begin{aligned} \begin{vmatrix} 6 & 1 & 1 \\ 4 & -2 & 5 \\ 2 & 8 & 7 \end{vmatrix} &= 6 \times (-2 \times 7 - 5 \times 8) - 1 \times (4 \times 7 - 5 \times 2) + 1 \times (4 \times 8 - (-2 \times 2)) \\ &= 6 \times (-54) - 1 \times (18) + 1 \times (36) = -306. \end{aligned}$$

对于行列式, 在线性代数中有专门详细的讨论, 现在我们只要求读者记住二阶与三阶行列式的上述计算公式. 误区: 不要以为看到了和绝对值符号类似的竖杠 $|$ 就把行列式和绝对值混为一谈. 可以说它们者没有丝毫关系. 行列式的值可正可负.

外积的坐标表示

令 $\boldsymbol{i}, \boldsymbol{j}$ 和 $\boldsymbol{k}$ 是分别沿 x 轴, y 轴和 z 轴的正向的坐标向量. 通过叉乘的定义, 有以下结果:

$$\begin{aligned} &\boldsymbol{i} \times \boldsymbol{i} = \boldsymbol{0}, \quad \boldsymbol{j} \times \boldsymbol{j} = \boldsymbol{0}, \quad \boldsymbol{k} \times \boldsymbol{k} = \boldsymbol{0}, \\ &\boldsymbol{i} \times \boldsymbol{j} = \boldsymbol{k}, \quad \boldsymbol{j} \times \boldsymbol{k} = \boldsymbol{i}, \quad \boldsymbol{k} \times \boldsymbol{i} = \boldsymbol{j}. \end{aligned}$$

小心第二行叉乘的顺序, 即使是简单的坐标向量, 叉乘的顺序也不能随意调换. 显然, 正是因为当初空间直角坐标系 $Oxyz$ 是通过右手系法则确定的各坐标轴正向的顺序, 此时的各个坐标向量的两两之间叉乘才能有如此结果. 一个记忆技巧是, 坐标向量的叉乘结果保持循环位移不变: $ijk \to jki \to kij \to ijk$.

设向量 $\boldsymbol{a} = (x_1, y_1, z_1)$, $\boldsymbol{b} = (x_2, y_2, z_2)$, 这时利用叉乘运算的分配律和上面的结果, 我们有

$$\begin{aligned} \boldsymbol{a} \times \boldsymbol{b} &= (x_1\boldsymbol{i} + y_1\boldsymbol{j} + z_1\boldsymbol{k}) \times (x_2\boldsymbol{i} + y_2\boldsymbol{j} + z_2\boldsymbol{k}) \\ &= x_1\boldsymbol{i} \times (x_2\boldsymbol{i} + y_2\boldsymbol{j} + z_2\boldsymbol{k}) + y_1\boldsymbol{j} \times (x_2\boldsymbol{i} + y_2\boldsymbol{j} + z_2\boldsymbol{k}) + z_1\boldsymbol{k} \times (x_2\boldsymbol{i} + y_2\boldsymbol{j} + z_2\boldsymbol{k}) \\ &= (x_1y_2\boldsymbol{k} - x_1z_2\boldsymbol{j}) + (-y_1x_2\boldsymbol{k} + y_1z_2\boldsymbol{i}) + (z_1x_2\boldsymbol{j} - z_1y_2\boldsymbol{i}) \\ &= (y_1z_2 - z_1y_2)\,\boldsymbol{i} + (z_1x_2 - x_1z_2)\,\boldsymbol{j} + (x_1y_2 - y_1x_2)\,\boldsymbol{k}. \end{aligned}$$

可以将最后一行结果形式地写成三阶行列式更便于记忆:

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix} = \begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix} \boldsymbol{i} - \begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix} \boldsymbol{j} + \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} \boldsymbol{k}.$$

将此行列式中的三个坐标向量 $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ 形式地代替三阶行列式定义中的第一行 x, y, z, 就是已证得的公式结果. 所谓 “形式地”, 就是明知 $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ 不是数, 而是向量; 但还是把它们当作数来处理.

命题 5.2.3: 外积的坐标表示

设向量 $\boldsymbol{a}=(x_1,y_1,z_1)$, $\boldsymbol{b}=(x_2,y_2,z_2)$, 则有

$$\boldsymbol{a}\times\boldsymbol{b}=\begin{vmatrix}\boldsymbol{i}&\boldsymbol{j}&\boldsymbol{k}\\x_1&y_1&z_1\\x_2&y_2&z_2\end{vmatrix}=(y_1z_2-z_1y_2)\,\boldsymbol{i}+(z_1x_2-x_1z_2)\,\boldsymbol{j}+(x_1y_2-y_1x_2)\,\boldsymbol{k},$$

或

$$\boldsymbol{a}\times\boldsymbol{b}=(y_1z_2-z_1y_2,z_1x_2-x_1z_2,x_1y_2-y_1x_2)\,.$$

例 5.2.5 设向量 $\boldsymbol{a}=(1,0,2)$, $\boldsymbol{b}=(2,-1,1)$, 求一个单位向量 $\boldsymbol{c}$, 使它同时垂直于 $\boldsymbol{a}$ 与 $\boldsymbol{b}$, 且 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 构成右手系. (说了半天就是求 $\boldsymbol{a}\times\boldsymbol{b}$ 的单位向量)

证明. 显然, 所求的向量为

$$\boldsymbol{c}=\frac{\boldsymbol{a}\times\boldsymbol{b}}{|\boldsymbol{a}\times\boldsymbol{b}|}.$$

根据公式, 有

$$\boldsymbol{a}\times\boldsymbol{b}=\begin{vmatrix}\boldsymbol{i}&\boldsymbol{j}&\boldsymbol{k}\\1&0&2\\2&-1&1\end{vmatrix}=2\boldsymbol{i}+3\boldsymbol{j}-\boldsymbol{k}=(2,3,-1).$$

由此计算出 $|\boldsymbol{a}\times\boldsymbol{b}|=\sqrt{14}$, 故

$$\boldsymbol{c}=\left(\frac{2}{\sqrt{14}},\frac{3}{\sqrt{14}},\frac{-1}{\sqrt{14}}\right).$$

□

外积的坐标表示公式为判断两个向量是否共线提供了方法.

例 5.2.6 设向量 $\boldsymbol{a}=(5,6,0)$, $\boldsymbol{b}=(1,2,3)$, 问: $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 是否共线?

证明. 因为

$$\boldsymbol{a}\times\boldsymbol{b}=\begin{vmatrix}\boldsymbol{i}&\boldsymbol{j}&\boldsymbol{k}\\5&6&0\\1&2&3\end{vmatrix}=18\boldsymbol{i}-15\boldsymbol{j}+4\boldsymbol{k}\neq\boldsymbol{0},$$

所以 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 不共线. □

例 5.2.7 用叉乘的坐标表示验证叉乘的分配律:

$$(\boldsymbol{a}+\boldsymbol{b})\times\boldsymbol{c}=\boldsymbol{a}\times\boldsymbol{c}+\boldsymbol{b}\times\boldsymbol{c}.$$

证明. 假设有 $\boldsymbol{a}=(x_1,y_1,z_1),\boldsymbol{b}=(x_2,y_2,z_2),\boldsymbol{c}=(x_3,y_3,z_3)$. 那么有

$$(\boldsymbol{a}+\boldsymbol{b})\times\boldsymbol{c}=\begin{vmatrix}\boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k}\\ x_1+x_2 & y_1+y_2 & z_1+z_2\\ x_3 & y_3 & z_3\end{vmatrix}.$$

展开这个行列式得到

$$\begin{aligned}(\boldsymbol{a}+\boldsymbol{b})\times\boldsymbol{c}&=((y_1+y_2)z_3-(z_1+z_2)y_3)\,\boldsymbol{i}-((x_1+x_2)z_3-(z_1+z_2)x_3)\,\boldsymbol{j}+((x_1+x_2)y_3-(y_1+y_2)x_3)\,\boldsymbol{k}\\&=(y_1z_3-z_1y_3+y_2z_3-z_2y_3)\,\boldsymbol{i}-(x_1z_3-z_1x_3+x_2z_3-z_2x_3)\,\boldsymbol{j}+(x_1y_3-y_1x_3+x_2y_3-y_2x_3)\,\boldsymbol{k}\\&=(y_1z_3-z_1y_3)\,\boldsymbol{i}-(x_1z_3-z_1x_3)\,\boldsymbol{j}+(x_1y_3-y_1x_3)\,\boldsymbol{k}\\&\quad+(y_2z_3-z_2y_3)\,\boldsymbol{i}-(x_2z_3-z_2x_3)\,\boldsymbol{j}+(x_2y_3-y_2x_3)\,\boldsymbol{k}\\&=\boldsymbol{a}\times\boldsymbol{c}+\boldsymbol{b}\times\boldsymbol{c}.\end{aligned}$$

最后一行我们用了实事:

$$\boldsymbol{a}\times\boldsymbol{c}=\begin{vmatrix}\boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k}\\ x_1 & y_1 & z_1\\ x_3 & y_3 & z_3\end{vmatrix}=(y_1z_3-z_1y_3)\,\boldsymbol{i}-(x_1z_3-z_1x_3)\,\boldsymbol{j}+(x_1y_3-y_1x_3)\,\boldsymbol{k},$$

$$\boldsymbol{b}\times\boldsymbol{c}=\begin{vmatrix}\boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k}\\ x_2 & y_2 & z_2\\ x_3 & y_3 & z_3\end{vmatrix}=(y_2z_3-z_2y_3)\,\boldsymbol{i}-(x_2z_3-z_2x_3)\,\boldsymbol{j}+(x_2y_3-y_2x_3)\,\boldsymbol{k}.$$

证毕. 如果分配律只用行列式来表示, 那么 $(\boldsymbol{a}+\boldsymbol{b})\times\boldsymbol{c}=\boldsymbol{a}\times\boldsymbol{c}+\boldsymbol{b}\times\boldsymbol{c}$ 就意味着

$$\begin{vmatrix}\boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k}\\ x_1+x_2 & y_1+y_2 & z_1+z_2\\ x_3 & y_3 & z_3\end{vmatrix}=\begin{vmatrix}\boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k}\\ x_1 & y_1 & z_1\\ x_3 & y_3 & z_3\end{vmatrix}+\begin{vmatrix}\boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k}\\ x_2 & y_2 & z_2\\ x_3 & y_3 & z_3\end{vmatrix}.$$

这一结果说明行列式关于某一行具有线性可加性. 这正是线性代数中的理论.[8] □

[8] 学过或正在学线性代数的同学可以复习一下 (没学的同学放心跳过, 不可能考的!): 行列式可以通过以下三个关键性质来表征. 为了便于说明, 我们将一个 $n\times n$ 矩阵 A 看作由其 n 列构成, 并以列变换操作为主 (行变换是一样的), 表示为

$$A=(a_1,a_2,\ldots,a_n),$$

其中列向量 a_i (对于每个 i) 表示矩阵中第 i 列的元素.

1. 单位矩阵的行列式为 1: 即 $\det(I)=1$, 其中 I 为单位矩阵.
2. 行列式具有多重线性性: 如果矩阵 A 的第 j 列 a_j 可以表示为两个列向量 v 和 w 的线性组合 $a_j=r\cdot v+w$, 其中 r 是一个常数, 则矩阵 A 的行列式可以按如下方式表示为这两个行列式的线性组合:

$$|A|=|a_1,\ldots,a_{j-1},r\cdot v+w,a_{j+1},\ldots,a_n|=r\cdot|a_1,\ldots,v,\ldots,a_n|+|a_1,\ldots,w,\ldots,a_n|$$

3. 行列式具有交替性: 当矩阵的两列相同 (即两列向量完全相同) 时, 其行列式为 0:

$$|a_1,\ldots,v,\ldots,v,\ldots,a_n|=0$$

这个性质表明, 行列式在列向量交换时会发生符号变化 (把交换的两个行列式求和). 如果两列相同, 则行列式为零.

5.2.4 混合积的坐标表示

有了内积和外积的坐标表示公式之后, 混合积的坐标表示公式就十分容易得到了. 设向量 $\boldsymbol{a}=(x_1,y_1,z_1)$, $\boldsymbol{b}=(x_2,y_2,z_2)$, $\boldsymbol{c}=(x_3,y_3,z_3)$, 则混合积 $\boldsymbol{c}\cdot(\boldsymbol{a}\times\boldsymbol{b})$ 应该等于 $\boldsymbol{c}$ 与 $\boldsymbol{a}\times\boldsymbol{b}$ 的对应坐标乘积之和. 通过前面, 我们知道

$$\boldsymbol{a}\times\boldsymbol{b}=(y_1z_2-z_1y_2)\,\boldsymbol{i}+(z_1x_2-x_1z_2)\,\boldsymbol{j}+(x_1y_2-y_1x_2)\,\boldsymbol{k},$$

于是 $\boldsymbol{a}\times\boldsymbol{b}$ 的坐标表示就是 $\boldsymbol{a}\times\boldsymbol{b}=(y_1z_2-z_1y_2,z_1x_2-x_1z_2,x_1y_2-y_1x_2)$. 所以

$$\begin{aligned}\boldsymbol{c}\cdot(\boldsymbol{a}\times\boldsymbol{b})&=(y_1z_2-z_1y_2)\,x_3+(z_1x_2-x_1z_2)\,y_3+(x_1y_2-y_1x_2)\,z_3\\&=x_3\begin{vmatrix}y_1&z_1\\y_2&z_2\end{vmatrix}-y_3\begin{vmatrix}x_1&z_1\\x_2&z_2\end{vmatrix}+z_3\begin{vmatrix}x_1&y_1\\x_2&y_2\end{vmatrix}\\&=\begin{vmatrix}x_3&y_3&z_3\\x_1&y_1&z_1\\x_2&y_2&z_2\end{vmatrix}.\end{aligned}$$

这就是三个向量混合积的坐标表示公式.

命题 5.2.4: 混合积的坐标表示

设向量 $\boldsymbol{a}=(x_1,y_1,z_1)$, $\boldsymbol{b}=(x_2,y_2,z_2)$, $\boldsymbol{c}=(x_3,y_3,z_3)$, 则有

$$\boldsymbol{c}\cdot(\boldsymbol{a}\times\boldsymbol{b})=\begin{vmatrix}x_3&y_3&z_3\\x_1&y_1&z_1\\x_2&y_2&z_2\end{vmatrix}.$$

注意, 三阶行列式的每一行从上到下, 就是混合积 $\boldsymbol{c}\cdot(\boldsymbol{a}\times\boldsymbol{b})$ 中从左到右的向量的坐标. 于是可以这么记忆

$$\boldsymbol{c}\cdot(\boldsymbol{a}\times\boldsymbol{b})=\begin{vmatrix}-&\boldsymbol{c}\text{ 的坐标}&-\\-&\boldsymbol{a}\text{ 的坐标}&-\\-&\boldsymbol{b}\text{ 的坐标}&-\end{vmatrix},$$

其中 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 是任意的向量. 若 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 为三个不共面的非零向量, 则上式右端的绝对值是以 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 为棱的平行六面体的体积. 另外, 如果读者学习了线性代数, 则之前提高的混合积的性质, 比如 $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})=\mathbf{b}\cdot(\mathbf{c}\times\mathbf{a})=\mathbf{c}\cdot(\mathbf{a}\times\mathbf{b})$ 就可以自然的用行列式的性质来解释了.

例 5.2.8 判断下列三个向量是否共面: $\boldsymbol{a}=(3,0,5),\boldsymbol{b}=(1,2,3),\boldsymbol{c}=(5,4,11)$.

证明. 计算混合积 $\boldsymbol{a}\cdot(\boldsymbol{b}\times\boldsymbol{c})$ 可得

$$\boldsymbol{a}\cdot(\boldsymbol{b}\times\boldsymbol{c})=\begin{vmatrix}3&0&5\\1&2&3\\5&4&11\end{vmatrix}=0,$$

可见 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 是共面的. □

5.2.5 方向余弦

在结束本节前, 我们要介绍一个以后会常用的概念——方向余弦. 见动画 direction cosines –GeoGebra.

定义 5.2.2: 方向余弦

设 $\boldsymbol{a}$ 是一个非零向量, 它与 x 轴, y 轴, z 轴正向的夹角分别为 $\alpha, \beta, \gamma \in [0, \pi]$, 我们称 α, β, γ 为 $\boldsymbol{a}$ 的**方向角**, 而称 $\cos\alpha, \cos\beta, \cos\gamma$ 为 $\boldsymbol{a}$ 的**方向余弦 (direction cosines)**.

显然, 根据向量的方向余弦与点乘运算的定义, 我们有 (注意坐标向量都是单位向量)

$$\cos\alpha = \frac{\boldsymbol{a}\cdot\boldsymbol{i}}{|\boldsymbol{a}|}, \quad \cos\beta = \frac{\boldsymbol{a}\cdot\boldsymbol{j}}{|\boldsymbol{a}|}, \quad \cos\gamma = \frac{\boldsymbol{a}\cdot\boldsymbol{k}}{|\boldsymbol{a}|}.$$

设向量 $\boldsymbol{a}$ 的坐标为 (x, y, z). 注意到三个坐标向量的坐标表示就是

$$\begin{aligned} \boldsymbol{i} &= (1,0,0), \\ \boldsymbol{j} &= (0,1,0), \\ \boldsymbol{k} &= (0,0,1). \end{aligned}$$

因此, $\boldsymbol{a}\cdot\boldsymbol{i} = x$, $\boldsymbol{a}\cdot\boldsymbol{j} = y$, $\boldsymbol{a}\cdot\boldsymbol{k} = z$. 又因为 $|\boldsymbol{a}| = \sqrt{x^2+y^2+z^2}$, 于是

$$\cos\alpha = \frac{x}{\sqrt{x^2+y^2+z^2}}, \quad \cos\beta = \frac{y}{\sqrt{x^2+y^2+z^2}}, \quad \cos\gamma = \frac{z}{\sqrt{x^2+y^2+z^2}}.$$

可见, 一个向量的三个方向余弦的平方和为 1:

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1.$$

由此还不难看出, 单位向量

$$\begin{aligned} \boldsymbol{a}^0 = \frac{\boldsymbol{a}}{|\boldsymbol{a}|} &= \frac{1}{\sqrt{x^2+y^2+z^2}}(x\boldsymbol{i}+y\boldsymbol{j}+z\boldsymbol{k}), \\ &= \cos\alpha\boldsymbol{i} + \cos\beta\boldsymbol{j} + \cos\gamma\boldsymbol{k}, \end{aligned}$$

换句话说, $\boldsymbol{a}$ 的方向余弦 $\cos\alpha, \cos\beta, \cos\gamma$ 实际上就是单位向量 $\boldsymbol{a}^0 = \dfrac{\boldsymbol{a}}{|\boldsymbol{a}|}$ 的坐标, 即

$$\boldsymbol{a}^0 = (\cos\alpha, \cos\beta, \cos\gamma).$$

其实, 前面讲内积的坐标表示的时候, 我们就得到过这一结果.

5.3 空间中平面与直线的方程

注意 本节的内容多且杂, 又比较重要. 如果只是背记公式结果, 那么期末考试多半凉凉. 唯一的途径是仔细思考背后的几何常识和以及如何运用内积, 叉乘或标量乘法来表现垂直和平行 (共线) 关系的一整套处理逻辑. 然后就是多做练习.

假定我们在空间中取定了直角坐标系 $Oxyz$ 及相应的三个坐标向量 $\boldsymbol{i}$, j, k. 现在我们来讨论空间中平面与直线所满足的方程.

考虑空间中的一个几何图形 F, 比如直线, 平面, 曲面, 曲线等等. 既然任何几何图形都是由空间中的点构成的, 我们自然可以把图形 F 视作一个集合, 并写做 $F \subseteq \mathbb{R}^3$. 此时给定一个方程 $f(x,y,z)=0$ 或一个方程组

$$\begin{cases} f(x,y,z)=0, \\ g(x,y,z)=0. \end{cases}$$

如果满足该方程（组）的所有点 (x,y,z) 当且仅当它是图形 F 上的点，也就是说集合 F 恰好构成了该方程（组）的解集（即，所有解的集合），那么我们称该方程（组）为此图形 F 的方程（组）. 根据图形 F 的具体形状, 这些方程或方程组称为直线方程, 平面方程, 球面方程, xx 曲面方程等等.

例 5.3.1 例如, 考虑方程

$$f_1(x,y,z)=x+y+z=0.$$

该方程的解集可以直接写做

$$\{(x,y,z)\in\mathbb{R}^3 \mid x+y+z=0\}.$$

几何上, 这表示的是三维空间中的一个平面, 所有满足该条件的点都位于该平面上. 为了解这个方程, 可以通过任意选取 x 和 y 的值来确定 $z=-x-y$ 的值, 故上述解集还可以写做

$$\{(x,y,-x-y)\in\mathbb{R}^3 \mid \forall x,y\in\mathbb{R}\}.$$

例如,

当 $x=0$, $y=0$, 则 $z=0$, 点为 $(0,0,0)$.

当 $x=1$, $y=0$, 则 $z=-1$, 点为 $(1,0,-1)$.

当 $x=0$, $y=1$, 则 $z=-1$, 点为 $(0,1,-1)$.

这些点都满足 $x+y+z=0$, 它们是同一个平面上的点. 任何一个平面应当有两个自由度, 就好比常见的 Oxy 平面, 正因为 x,y 可以任意地取值, 所以它才能伸展成一个平面. 上述讨论的平面尽管在当下的坐标系中是 “倾斜” 的, 但作为平面的几何性质不依赖于坐标系. 我们通过解方程, 还是能找出该方程的解集的自由度. 本例中, 这个两个自由度 x,y 可以任意地伸展出空间中 “倾斜” 平面上点的坐标.

再看方程

$$f_2(x,y,z)=2x+2y+2z=0.$$

显然, f_2 和 f_1 有着完全相同的解集, 所以它们都是同一个平面的方程. 再看方程

$$f_3(x,y,z)=\begin{vmatrix} x & y & z \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{vmatrix}=0.$$

将行列式展开后, $f_3=f_1$. 这表明了通常图形 F 在纯几何视角是确定的, 但其对应的方程通常不唯一, 可以有多种表达式, 只是它们应当有相同的解集.

5.3.1 平面的方程

根据已知条件的不同, 平面方程可有多种不同的形式. 包括: 点法式方程, 一般方程, 三点式方程和截距式方程.

平面的点法式方程

几何常识: 一条直线和一个平面内的任何一条直线都垂直, 就说这条直线和这个平面互相垂直, 直线叫做平面的垂线, 平面叫做直线的垂面, 垂线和平面一定相交, 交点叫做垂足. 过一点并且和一条直线垂直的平面有且仅有一个. 受这个几何实事的启发, 我们下面推出平面的点法式方程.

定义 5.3.1: 平面的法向量

给定一个平面, 任何一个垂直于该平面的非零向量称为该平面的**法向量 (normal vector)**. 一个平面有多个法向量, 但这些法向量都一定共线.

给定一个平面, 设其法向量为 $\boldsymbol{n} = (A, B, C)$, 其中 A, B, C 不是全为零. 假设该平面上有一个已知点 $P_0(x_0, y_0, z_0)$. 现在, 我们想要描述平面上的任意一点 $P(x, y, z)$. 对于平面上的任意一点 $P(x, y, z)$, 由法向量的定义可知, 点 P_0 与 P 之间的向量 $\overrightarrow{P_0P}$ 与法向量 $\boldsymbol{n}$ 一定是垂直的. 也就是说,

$$\boldsymbol{n} \cdot \overrightarrow{P_0P} = 0.$$

将 $\overrightarrow{P_0P} = (x - x_0, y - y_0, z - z_0)$ 代入, 得到

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0.$$

因此, 对于该平面上的任意一点 $P(x, y, z)$, 它的坐标必须满足这个方程. 反过来, 如果一个空间中的点 $P(x, y, z)$ 满足上述方程, 那么该点必定位于平面上. 这个方程是由平面的一个法向量及平面上一个已知点的坐标确定的, 故通常称之为**平面的点法式方程**. 这个方程可以化成

$$Ax + By + Cz + D = 0$$

的形式, 其中 $D = -(Ax_0 + By_0 + Cz_0)$ 是常数, x, y, z 的系数 A, B, C 依次是法向量的三个坐标分量.

如果考虑平面上其他的已知点坐标. 假设我们还知道此平面上另外一个点 $P_1(x_1, y_1, z_1)$. 既然点 P_1 在平面上, 那么它就满足上面的点法式方程, 故有

$$Ax_1 + By_1 + Cz_1 + D = 0,$$

即,

$$Ax_1 + By_1 + Cz_1 = Ax_0 + By_0 + Cz_0.$$

如果一开始, 选择用已知点 $P_1(x_1, y_1, z_1)$ 来构建点法式方程, 那么在法向量 $\boldsymbol{n} = (A, B, C)$ 不变的条件下, 我们得到新的点法式方程为

$$A(x - x_1) + B(y - y_1) + C(z - z_1) = 0.$$

化简后得

$$Ax + By + Cz - (Ax_1 + By_1 + Cz_1) = 0.$$

显然,

$$Ax + By + Cz - (Ax_0 + By_0 + Cz_0) = Ax + By + Cz + D = 0.$$

这说明只要已知点是平面上的点 (无论具体是 P_0 还是 P_1), 那么点法式方程化简后出现的常数项 D 是不会变的. 虽然 $D = -(Ax_0 + By_0 + Cz_0)$ 的定义乍一看依赖于已知点 $P_0(x_0, y_0, z_0)$ 的坐标, 但换成另外一个平面上的已知点, D 也不会变. 因此, 给定法向量, 选择平面上任意的已知点来构建点法式方程, 将得到的方程化简后, 其最终结果一定是一样的.

如果考虑平面其他的法向量. 若平面的法向量为 $\boldsymbol{n}=(A,B,C)$, 其中 A,B,C 不是全为零. 那么任何与 $\boldsymbol{n}$ 共线的非零向量也是该平面的法向量, 可以写做

$$\lambda\boldsymbol{n}=(\lambda A,\lambda B,\lambda C),\quad \forall\lambda\neq 0.$$

如果我们选定这样子的法向量和已知点 $P_0(x_0,y_0,z_0)$ 来构建点法式方程, 会得到

$$\lambda A(x-x_0)+\lambda B(y-y_0)+\lambda C(z-z_0)=0.$$

两边可以同时消去 λ, 就得到了和原来一样的方程. 或者不用消去 λ, 方程的解集没有任何改变, 所以此时的方程仍旧是该平面的方程. 我们将以上讨论结果总结如下.

结论 5.3.1: 平面的点法式方程

已知平面上任意一点 $P_0(x_0,y_0,z_0)$ 和其任意法向量 $\boldsymbol{n}=(A,B,C)$, 则能唯一确定一个平面, 其**平面的点法式方程**为

$$A(x-x_0)+B(y-y_0)+C(z-z_0)=0.$$

这个方程本质上不依赖于 P_0 和 $\boldsymbol{n}$ 的具体坐标:

1. 选择平面上不同的已知点 $P_0(x_0,y_0,z_0)$ 不会影响上述方程化简后的最终结果.
2. 选择平面上不同的法向量 $\boldsymbol{n}=(A,B,C)$ 不会影响方程的解集, 不同法向量的差异等价于方程两边同时乘以一个非零常数.

例 5.3.2 已知一个平面的法向量为 $(2,3,4)$, 该平面上一点的坐标为 $(1,1,1)$, 则该平面的点法式方程是

$$2(x-1)+3(y-1)+4(z-1)=0,$$

化简后就是

$$2x+3y+4z-9=0.$$

1. 还可以知道平面上有另外一点 $(-1,1,2)$, 它满足 $2\times(-1)+3\times1+4\times2-9=0$. 则用此点构建的点法式方程是

   $$2(x+1)+3(y-1)+4(z-2)=0,$$

   化简后结果还是不变:

   $$2x+3y+4z-9=0.$$

2. 如果我们考虑的法向量是 $(1,\frac{3}{2},2)$, 已知点还是 $(1,1,1)$, 则该平面的点法式方程是

   $$(x-1)+\frac{3}{2}(y-1)+2(z-1)=0,$$

   化简后就是

   $$x+\frac{3}{2}y+2z-\frac{9}{2}=0.$$

   它就是原方程的两边同除以 2, 故解集不变. 此方程还是同一个平面的方程, 只不过它直接对应的法向量是 $(1,\frac{3}{2},2)$.

平面的一般方程

(平面 $\mapsto$ 三元一次方程) 上一小节的讨论使我们看到：给一个平面，通过平面上的任意一点和它的法向量，建立点法式方程以后化简它，那么就得到了一个关于 x, y, z 的线性方程 *(三元一次方程)*

$$Ax + By + Cz + D = 0,$$

其中系数 A, B, C 不全为零. 也就是说，空间中一个平面的方程就是形如上式的三元一次方程. 该方程的系数恰是该平面法向量的三个坐标分量.

(三元一次方程 $\mapsto$ 平面) 反过来，任意给定这样一个三元一次方程，它必是某个平面的方程. 事实上，对于任意给定的方程，

$$Ax + By + Cz + D = 0,$$

其中系数 A, B, C 不全为零. 我们很容易找到它的一个解 (x_0, y_0, z_0)，也就是找到一点 (x_0, y_0, z_0)，满足

$$Ax_0 + By_0 + Cz_0 + D = 0.$$

比如，我们假设当 $A \neq 0$，则点 $\left(-\dfrac{D}{A}, 0, 0\right)$ 即为一个解. 将方程 $Ax + By + Cz + D = 0$ 两边同时减去上式两边，得到

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0.$$

注意，方程的解集从未变化. 而上式恰好是由点 (x_0, y_0, z_0) 及法向量 (A, B, C) 所决定的平面的点法式方程. 因此，$Ax + By + Cz + D = 0$ 对应着某个平面. *注意，本讨论也告诉了我们如何将平面的一般方程化成点法式方程的步骤.* 显然，你所找到的平面上的不同，点法式方程的表现形式也不同. 我们将以上讨论结果总结如下.

结论 5.3.2: 平面的一般方程

空间中的平面与三元一次方程有一一对应[a]的关系，今后我们将三元一次方程 (其中系数 A, B, C 不全为零)

$$Ax + By + Cz + D = 0$$

称为**平面的一般方程**. 坐标 (A, B, C) 就是该平面的一个法向量的坐标.

[a] 我们把具有标量倍数关系的方程看作是一个方程：比如 $x + y + z + 1 = 0$ 与 $2x + 2y + 2z + 2 = 0$.

例 5.3.3 将平面的一般方程 $3x + 4y + 6z = 1$ 化成点法式方程.

证明. 先在平面上任意选定一点，比如 $(-3, 1, 1)$. (只需要固定 x, y, z 中任意两个值，就可以求出剩余的那个) 这样，所给的方程可以化成

$$3(x + 3) + 4(x - 1) + 6(y - 1) = 0.$$

这里法向量为 $\boldsymbol{n} = (3, 4, 6)$. $\square$

例 5.3.4 **点到平面的距离.** 已知一个平面的方程为

$$Ax + By + Cz + D = 0,$$

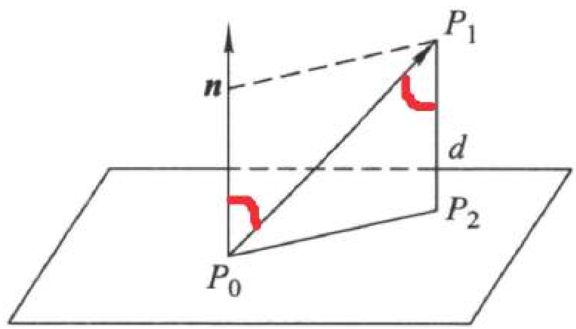


图 5.7: 计算点到平面的距离

其中 A, B, C 不全为零. 那么空间中任意点 $P_1(x_1, y_1, z_1)$ 到该平面的距离等于

$$\frac{|Ax_1 + By_1 + Cz_1 + D|}{\sqrt{A^2 + B^2 + C^2}}.$$

证明. 从平面方程中马上读出来其法向量 $\boldsymbol{n} = (A, B, C)$. 过点 P_1 作该平面的垂线, 设垂足为点 P_2, 那么所求的距离即为点 P_1 到点 P_2 的距离. 显然, 由图 5.7 可知直线 P_1P_2 垂直于平面, 所以和法向量 $\boldsymbol{n} = (A, B, C)$ 平行.

又设 $P_0(x_0, y_0, z_0)$ 为该平面上任意一点, 则有

$$D = -Ax_0 - By_0 - Cz_0.$$

考虑将法向量 $\boldsymbol{n}$ 的起点移到点 P_0. 那么, 点 P_1 到点 P_2 的距离就等于向量 $\overrightarrow{P_0P_1}$ 往法向量 $\boldsymbol{n}$ 上投影分量[9]的绝对值 (见图 5.7). 于是,

$$\begin{aligned}
P_1 \text{ 到该平面的距离} &= \left|\overrightarrow{P_0P_1}\right| \left|\cos\left\langle \boldsymbol{n}, \overrightarrow{P_0P_1}\right\rangle\right| \\
&= \frac{\left|\overrightarrow{P_0P_1} \cdot \boldsymbol{n}\right|}{|\boldsymbol{n}|} \\
&= \frac{|A(x_1 - x_0) + B(y_1 - y_0) + C(z_1 - z_0)|}{\sqrt{A^2 + B^2 + C^2}} \\
&= \frac{|Ax_1 + By_1 + Cz_1 + D|}{\sqrt{A^2 + B^2 + C^2}}.
\end{aligned}$$

证毕.

显然, 如果 $P_1(x_1, y_1, z_1)$ 就在平面上, 那么分子就是 0. 而且, $P_0(x_0, y_0, z_0)$ 只是一个辅助点, 也没有出现在最终的结果中. 并且, 任何和 $\boldsymbol{n} = (A, B, C)$ 共线的向量都可以当作法向量, 从上面结果我们发现

$$\frac{\left|\overrightarrow{P_0P_1} \cdot \lambda\boldsymbol{n}\right|}{|\lambda\boldsymbol{n}|} = \frac{\left|\overrightarrow{P_0P_1} \cdot \boldsymbol{n}\right|}{|\boldsymbol{n}|}.$$

□

[9]复习: 内积也可以用来描述一个向量在另一个向量上的投影. 考虑向量 $\boldsymbol{a}$ 在向量 $\boldsymbol{b}$ 上的投影, 则定义投影分量为 $|\boldsymbol{a}|\cos\langle\boldsymbol{a}, \boldsymbol{b}\rangle = \dfrac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{b}|}$.

不要死记硬背公式结果, 还是应该抓住文字描述的核心逻辑: 比如, 上一道例题告诉我们, 点到平面的距离, 就等于该点和平面上任意一点连成的向量在平面法向量上的投影分量的绝对值, 即

$$P_1 \text{ 到平面的距离} = \frac{\left|\overrightarrow{P_0P_1} \cdot \boldsymbol{n}\right|}{|\boldsymbol{n}|}.$$

而一看到 "投影", 就要想到运用内积的定义. 再用画图辅助自己. 这样子, 才能以不变应万变.

平面的三点式方程

几何常识告诉我们: 空间中任意不共线的三点可以唯一确定一个平面. 给定三点的坐标:

$$P_1\left(x_1, y_1, z_1\right), \quad P_2\left(x_2, y_2, z_2\right), \quad P_3\left(x_3, y_3, z_3\right).$$

于是, 我们想直接用这三个点的坐标写出其确定的平面方程. 假定给定的三点不共线, 这也就是说, 如果我们考虑以 P_1 为起点, 分别以 P_2, P_3 为终点的两个向量 $\overrightarrow{P_1P_2}$ 与 $\overrightarrow{P_1P_3}$, 那么这两个向量不共线, 即

$$\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3} \neq \mathbf{0}.$$

重要的是, 根据叉乘的右手系法则定义, 向量 $\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3}$ 正好就是三点所决定的平面的法向量. 对于平面上的任意一点 $P(x, y, z)$, 点 P_1 与 P 之间的向量 $\overrightarrow{P_1P}$ 与平面的法向量一定是垂直的, 即满足混合积

$$\overrightarrow{P_1P} \cdot \left(\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3}\right) = 0.$$

那么, 将以下向量坐标:

$$\begin{aligned}
\overrightarrow{P_1P} &= \left(x - x_1, y - y_1, z - z_1\right), \\
\overrightarrow{P_1P_3} &= \left(x_2 - x_1, y_2 - y_1, z_2 - z_1\right), \\
\overrightarrow{P_1P_3} &= \left(x_3 - x_1, y_3 - y_1, z_3 - z_1\right),
\end{aligned}$$

代入到混合积的坐标表示中, 我们有

$$\begin{vmatrix}
x - x_1 & y - y_1 & z - z_1 \\
x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\
x_3 - x_1 & y_3 - y_1 & z_3 - z_1
\end{vmatrix} = 0.$$

这个方程称为平面的三点式方程. 将此行列式按照第一行展开, 即得到一个关于 x, y, z 的线性方程 (三元一次方程). 现将以上结论汇总如下.

结论 5.3.3: 平面的三点式方程

已知空间中不共线的三点坐标 $P_1\left(x_1, y_1, z_1\right), P_2\left(x_2, y_2, z_2\right), P_3\left(x_3, y_3, z_3\right)$, 则能唯一确定一个平面, 其**平面的三点式方程**为

$$\overrightarrow{P_1P} \cdot \left(\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3}\right) = 0,$$

即,

$$\begin{vmatrix}
x - x_1 & y - y_1 & z - z_1 \\
x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\
x_3 - x_1 & y_3 - y_1 & z_3 - z_1
\end{vmatrix} = 0.$$

这个方程本质上不依赖平面上三点的具体坐标. 对于任意给定的平面上不共线的三点坐标, 将对应的三点式方程化解后, 最终结果都是一致的.

在利用三点式方程式时, 我们没有必要事先验证三点是否不共线. 如果它们共线, 那么一定有 $\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3} = \mathbf{0}$, 则

$$\overrightarrow{P_1P} \cdot \left(\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3}\right) = \overrightarrow{P_1P} \cdot \mathbf{0} \equiv 0.$$

也就是说, 三点式方程的左侧的 3 阶行列式一展开结果就是 0, 最终得到 $0 = 0$ 的无用方程. 请读者尝试对三点 $P_1(0,0,1), P_2(0,0,2), P_3(0,0,3)$ 应用三点式方程, 看看结果如何.

从前面的分析中可知, 三点 $P_1(x_1, y_1, z_1)$, $P_2(x_2, y_2, z_2)$, $P_3(x_3, y_3, z_3)$ 的角色应该是可以任意互换的. 比如, 我们还可以考虑混合积

$$\overrightarrow{P_2P} \cdot \left(\overrightarrow{P_2P_1} \times \overrightarrow{P_2P_3}\right) = 0, \text{ 或者 } \overrightarrow{P_3P} \cdot \left(\overrightarrow{P_3P_1} \times \overrightarrow{P_3P_2}\right) = 0.$$

将其写成行列式的形式后展开, 最终结果肯定也是一致的. 请读者以下题为例, 自行验证.

例 5.3.5 已知三点 $P_1(0,0,1)$, $P_2(1,1,0)$, $P_3(1,0,1)$, 求过这三点的平面方程.

证明. 所求平面的三点式方程是

$$\begin{aligned} 0 &= \begin{vmatrix} x-0 & y-0 & z-1 \\ 1 & 1 & -1 \\ 1 & 0 & 0 \end{vmatrix} \\ &= x(1\cdot 0 - (-1)\cdot 0) - y(1\cdot 0 - (-1)\cdot 1) + (z-1)(1\cdot 0 - 1\cdot 1) \\ &= -y - (z-1). \end{aligned}$$

即

$$y + z - 1 = 0.$$

即便重新取另一组满足不共线条件的三点, 最终计算出的平面方程仍与当前平面一致. 例如, 考虑另外三点 $Q_1(0,1,0), Q_2(1,2,-1), Q_3(-1,0,1)$. 代入到三点式方程中, 则有

$$\begin{aligned} 0 &= \begin{vmatrix} x-0 & y-1 & z-0 \\ 1 & 1 & -1 \\ 2 & -1 & 1 \end{vmatrix} \\ &= x(1\cdot 1 - (-1)\cdot(-1)) - (y-1)(1\cdot 1 - 2\cdot(-1)) + z(1\cdot(-1) - 2\cdot 1) \\ &= -3(y-1) - 3z. \end{aligned}$$

即 $y + z - 1 = 0$. □

平面的截距式方程

已知一个平面的一般方程 (其中系数 A, B, C 不全为零)

$$Ax + By + Cz + D = 0.$$

如何在空间中确定它的位置?[10] 这是一个很基本的问题. 下面我们讨论这个问题.

[10]当然是用 geogebra 或者 desmos. 这里假设我没电脑工具.

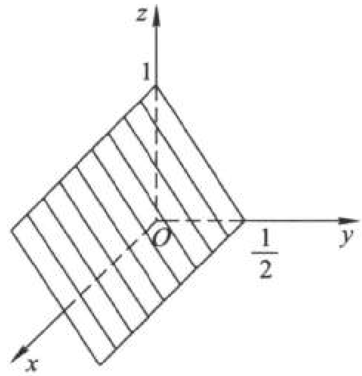


图 5.8: 方程 $0x+By+Cz+D=0$ 所表示的平面

1. 当 A,B,C 中有两个为零时. 比如 $A=B=0$, 这时平面的一般方程变成

$$0x+0y+Cz+D=Cz+D=0.$$

它的解是 $z=-\dfrac{D}{C}$. 更确切地说, 作为三元一次方程, 它的解是

$$\left(x,y,-\frac{D}{C}\right),$$

其中 x,y 可取任意值. 它在空间中代表一个平行于 Oxy 平面的平面, 且在 z 轴上的截距为 $-\dfrac{D}{C}$.

2. 当 A,B,C 中只有一个为零时. 比如 $A=0$, 这意味着法向量 $(0,B,C)$ 与 x 轴垂直, 这时平面的一般方程变成

$$0x+By+Cz+D=By+Cz+D=0.$$

我们在 Oyz 平面上画出直线 $By+Cz+D=0$, 然后将此直线沿平行于 x 轴的方向移动, 即得到方程 $0x+By+Cz+D=0$ 所表示的平面. 比如, 方程 $2y+z-1=0$ 表示的平面如图 5.8 所示.

> **注释 5.3.1** 小技巧: 平面的一般方程 $Ax+By+Cz+D=0$ 中 A,B,C 哪个系数等于零, 说明平面和对应的坐标轴不相交, 没有截距.

3. 当三个系数 A,B,C 均不为零时. 我们可求出平面 $Ax+By+Cz+D=0$ 在三条坐标轴上的截距. 事实上, 该平面在 x 轴, y 轴, z 轴上的截距分别是 $-\dfrac{D}{A},-\dfrac{D}{B},-\dfrac{D}{C}$. 它们依次对应着平面和三条坐标轴的三个交点:

$$\left(-\frac{D}{A},0,0\right),\quad \left(0,-\frac{D}{B},0\right),\quad \left(0,0,-\frac{D}{C}\right).$$

1. 如果 $D=0$, 则三个截距均零, 即平面过原点. 如果 $D\neq 0$, 则三个截距均不为零, 而这时平面方程

$$Ax+By+Cz=-D.$$

可变形成为

$$\frac{Ax}{-D}+\frac{By}{-D}+\frac{Cz}{-D}=1.$$

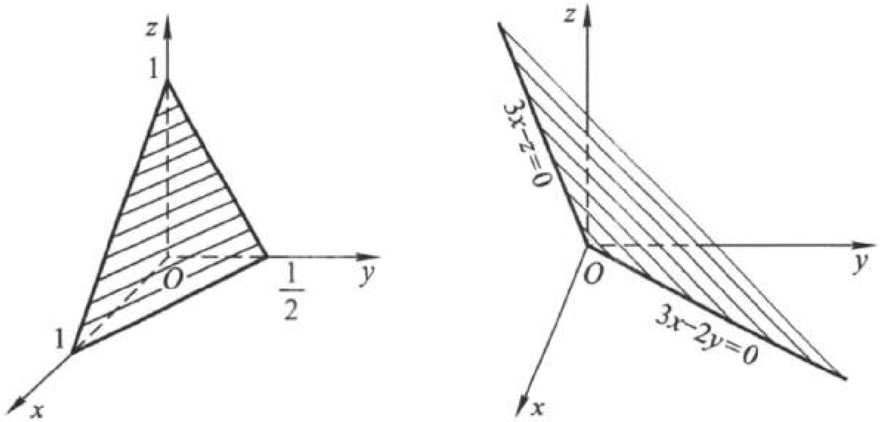


图 5.9: 平面的截距式方程

有因为 A, B, C 均不为零, 故可再变形为

$$\frac{x}{-\frac{D}{A}}+\frac{y}{-\frac{D}{B}}+\frac{z}{-\frac{D}{C}}=1.$$

这个方程称为平面的**截距式方程**, 从中可看出各截距与三个一次项系数的关系. 在三条坐标轴上的截距知道了, 平面在空间中的位置也就容易想象了.

还以通过三点式方程来验证以上结果. 对于三点 $P_1\left(-\dfrac{D}{A},0,0\right)$, $P_2\left(0,-\dfrac{D}{B},0\right)$, $P_3\left(0,0,-\dfrac{D}{C}\right)$, 我们利用三点式方程来求过这三点的平面方程. 将点 P_1, P_2, P_3 的坐标代入得

$$\begin{aligned}0=\begin{vmatrix} x+\dfrac{D}{A} & y-0 & z-0\\ \dfrac{D}{A} & -\dfrac{D}{B} & 0\\ \dfrac{D}{A} & 0 & -\dfrac{D}{C}\end{vmatrix}&=(x+\frac{D}{A})\begin{vmatrix}-\dfrac{D}{B} & 0\\ 0 & -\dfrac{D}{C}\end{vmatrix}-y\begin{vmatrix}\dfrac{D}{A} & 0\\ \dfrac{D}{A} & -\dfrac{D}{C}\end{vmatrix}+z\begin{vmatrix}\dfrac{D}{A} & -\dfrac{D}{B}\\ \dfrac{D}{A} & 0\end{vmatrix}\\ &=(x+\frac{D}{A})\cdot\frac{D^2}{BC}-y\cdot\left(-\frac{D^2}{AC}\right)+z\cdot\frac{D^2}{AB}.\end{aligned}$$

化简后为

$$\frac{D^2}{BC}x+\frac{D^3}{ABC}+\frac{D^2}{AC}y+\frac{D^2}{AB}z=\frac{D^2}{ABC}\left(Ax+By+Cz+D\right)=0.$$

由于 $\dfrac{D^2}{ABC}\neq 0$, 最终平面方程为

$$Ax+By+Cz+D=0.$$

(注意, 以上每一次变形化简的标准是, 不改变方程的解集.)

2. 若 $D=0$, 则表明平面通过原点. 在这种情况下, 为了确定平面的位置, 最好画出平面与两个坐标平面 (比如 Oxy 平面及 Oyz 平面) 的交线. 根据这两条交线即可确定平面的位置. 如图 5.9 所示.

例 5.3.6 方程 $x+2y+z-1=0$ 表示的平面在 x 轴, y 轴, z 轴上的截距分别是 $1, \dfrac{1}{2}, 1$. 该平面在第一卦限内的部分如图 5.9 所示.

例 5.3.7 方程 $3x-2y-z=0$ 表示的平面通过原点, 它与 Oxy 平面 (平面 $z=0$) 的交线是

$$3x-2y=0,\quad z=0.$$

而它与 Ozx 平面的交线是

$$3x-z=0,\quad y=0.$$

在 Oxy 平面上画出直线 $3x-2y=0$, 并在 Ozx 平面上画出直线 $3x-z=0$, 这时所讨论的平面就是通过这两条直线的平面, 如图 5.9 所示.

两个平面之间的位置关系

下面我们将讨论如何根据两个平面的方程来确定这两个平面之间的位置关系以及它们相交时的夹角.

1. 平面重合. 考虑两个平面方程:

$$A_1x+B_1y+C_1z+D_1=0,$$

$$A_2x+B_2y+C_2z+D_2=0.$$

容易知道, 这两个平面重合等价于 A_1,B_1,C_1,D_1 与 A_2,B_2,C_2,D_2 成比例, 因为它们方程的解集保持不变. 两个四元组 (A_1,B_1,C_1,D_1) 和 (A_2,B_2,C_2,D_2) 成比例, 是指这两个四元组之间存在一个比例系数 $k\neq 0$, 使得:

$$A_2=kA_1,\quad B_2=kB_1,\quad C_2=kC_1,\quad D_2=kD_1.$$

换句话说, 它们之间满足以下关系 (假设分母都不为零):

$$\frac{A_2}{A_1}=\frac{B_2}{B_1}=\frac{C_2}{C_1}=\frac{D_2}{D_1}=k.$$

接下来, 我们假设 A_1,B_1,C_1,D_1 与 A_2,B_2,C_2,D_2 不成比例, 故两个平面不重合.

2. 平面平行. 我们知道三元一次方程

$$Ax+By+Cz+D=0$$

所代表平面的法向量是 (A,B,C). 几何常识告诉我们, 给定两个不重合的平面, 它们的垂线相互平行当且仅当这两个平面平行. 由此可见, 平行的充要条件是各自的法向量 (A_1,B_1,C_1) 与 (A_2,B_2,C_2) 共线, 即存在一个实数 λ, 使得 $\lambda(A_1,B_1,C_1)=(A_2,B_2,C_2)$. 换句话说, 这两个平面平行的充要条件是 A_1,B_1,C_1 与 A_2,B_2,C_2 成比例.

3. 平面垂直. 我们定义两个平面

$$A_1x+B_1y+C_1z+D_1=0,$$

$$A_2x+B_2y+C_2z+D_2=0,$$

之间的夹角就是它们的法向量 $\boldsymbol{n}_1=(A_1,B_1,C_1)$ 与 $\boldsymbol{n}_2=(A_2,B_2,C_2)$ 之间的夹角.[11] 设这个夹角为 θ, 那么

$$\cos\theta=\frac{\boldsymbol{n}_1\cdot\boldsymbol{n}_2}{|\boldsymbol{n}_1|\,|\boldsymbol{n}_2|},$$

[11]严格地讲, 两平面的夹角被定义为 $\langle\boldsymbol{n}_1,\boldsymbol{n}_2\rangle$ 和 $\langle-\boldsymbol{n}_1,\boldsymbol{n}_2\rangle=\langle\boldsymbol{n}_1,-\boldsymbol{n}_2\rangle=\pi-\langle\boldsymbol{n}_1,\boldsymbol{n}_2\rangle$ 两者中的锐角或直角. 之所以这么定义, 是因为平面法向量的定义是任何垂直于平面的向量, 故 $\pm\boldsymbol{n}_1,\pm\boldsymbol{n}_2$ 可同时是法向量. 这就会导致出两个夹角, 但和是 π. 所以我们默认选择小于等于 $\frac{\pi}{2}$ 即可. 还有一种定义方式, 就是限定法向量满足 $\langle\boldsymbol{n}_1,\boldsymbol{n}_2\rangle\geq 0$, 否则的话应当给其中任意一个平面的法向量取负号, 比如, 让 $\boldsymbol{n}_1\leftarrow(-\boldsymbol{n}_1)$. 这样永远能保证夹角 $\theta=\langle\boldsymbol{n}_1,\boldsymbol{n}_2\rangle$ 是锐角或直角. 两条直线的夹角也约定是其中锐角或直角的那个.

即

$$\cos\theta = \frac{A_1A_2 + B_1B_2 + C_1C_2}{\sqrt{A_1^2 + B_1^2 + C_1^2}\sqrt{A_2^2 + B_2^2 + C_2^2}}.$$

由此立即推出, 上述两个平面相互垂直的充要条件是

$$A_1A_2 + B_1B_2 + C_1C_2 = 0.$$

结论 5.3.4: 两个平面之间的位置关系

给定两个平面的方程

$$A_1x + B_1y + C_1z + D_1 = 0,$$
$$A_2x + B_2y + C_2z + D_2 = 0,$$

那么,

1. 两个平面重合 $\Longleftrightarrow$ **(就是全部系数成比例)** A_1, B_1, C_1, D_1 与 A_2, B_2, C_2, D_2 成比例.
2. 两个平面平行 $\Longleftrightarrow$ **(就是法向量平行)** A_1, B_1, C_1, D_1 与 A_2, B_2, C_2, D_2 不成比例, 但是 A_1, B_1, C_1 与 A_2, B_2, C_2 成比例.
3. 两个平面垂直 $\Longleftrightarrow$ **(就是法向量垂直)** $A_1A_2 + B_1B_2 + C_1C_2 = 0$.

例 5.3.8 试确定常数 l 与 k, 使平面

$$x + ly + kz = 1$$

与平面 $x + y - z = 8$ 垂直, 且过点 $\left(1, 1, -\dfrac{2}{3}\right)$.

证明. 两个平面相互垂直的条件是它们的法向量垂直. 设两个平面的法向量分别为:

- 平面 $x + ly + kz = 1$ 的法向量为 $\boldsymbol{n}_1 = (1, l, k)$;
- 平面 $x + y - z = 8$ 的法向量为 $\boldsymbol{n}_2 = (1, 1, -1)$.

法向量垂直的条件是它们的点积为零, 即 $\boldsymbol{n}_1 \cdot \boldsymbol{n}_2 = 0$. 故有

$$1 \cdot 1 + l \cdot 1 + k \cdot (-1) = 0,$$

化简得

$$1 + l - k = 0.$$

要求点 $\left(1, 1, -\dfrac{2}{3}\right)$ 在平面 $x + ly + kz = 1$ 上, 代入点的坐标 $(x, y, z) = \left(1, 1, -\dfrac{2}{3}\right)$ 到平面方程 $x + ly + kz = 1$, 有

$$1 + l \cdot 1 + k \cdot \left(-\frac{2}{3}\right) = 1,$$

化简得

$$l - \frac{2}{3}k = 0.$$

联立上面得到的两个关于 l 与 k 的方程, 解得 $l = 2$, $k = 3$. □

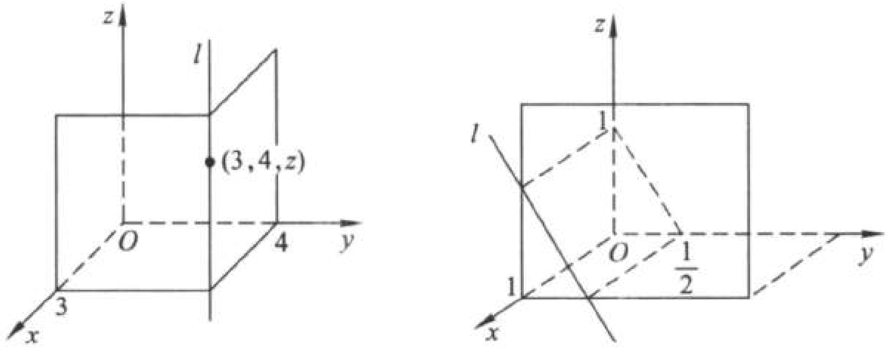


图 5.10: Enter Caption

例 5.3.9 两平行平面的距离 两平行平面方程为

$$\Pi_1 : Ax + By + Cz + D_1 = 0,$$
$$\Pi_2 : Ax + By + Cz + D_2 = 0,$$

几何知识表明: *两平行平面的距离就是平面上任意一点到另外一个平面的距离*. 故在 Π_1 上任意取一点 $P\,(x_0, y_0, z_0)$, 则 P 到在 Π_2 的距离为 (见例题 5.3.4)

$$d = \frac{|Ax_0 + By_0 + Cz_0 + D_2|}{\sqrt{A^2 + B^2 + C^2}},$$

由于 P 在 Π_1 上, 所以 $Ax_0 + By_0 + Cz_0 = -D_1$. 于是所求距离可化简为

$$h = \frac{|D_2 - D_1|}{\sqrt{A^2 + B^2 + C^2}}.$$

(不要背公式, 要背思路和逻辑! 这个立体最重要的是思想: 面和面的距离变成点和面的距离.)

5.3.2 直线的方程

现在我们转入讨论空间中直线的方程. 根据已知条件的不同, 直线方程可有多种不同的形式. 包括: 一般方程 (两面式方程), 参数方程和标准方程.

直线的一般方程 (两面式方程)

我们知道, 在 Oxy 平面上一条直线的方程是关于 x, y 的一个一次方程. 但是, 到了空间的情形, 一条直线的方程就要通过两个平面方程来描述. 事实上, 直线总可以看成两个不平行平面的交线. 用方程的语言来说, 一条直线上的点 (x, y, z) 应当满足一个联立方程组

$$\begin{cases} A_1x + B_1y + C_1z + D_1 = 0, \\ A_2x + B_2y + C_2z + D_2 = 0, \end{cases}$$

其中 A_1, B_1, C_1 与 A_2, B_2, C_2 不成比例. 反过来, 这个联立方程组的解 (x, y, z) 均在两个方程所代表平面的交线上. 我们把形如 x, y, z 的一次联立方程组的直线方程称作**直线的一般方程**或**两面式方程**.

例 5.3.10 联立方程组

$$\begin{cases} x-3=0 \\ y-4=0 \end{cases}$$

的解是 $(3,4,z), \forall z \in \mathbf{R}$, 其图形是平面 $x-3=0$ 与平面 $y-4=0$ 的交线 l, 它平行于 z 轴, 如图 5.10 所示.

例 5.3.11 联立方程组

$$\begin{cases} x-1=0 \\ 2y+z-1=0 \end{cases}$$

代表平面 $x-1=0$ 与平面 $2y+z-1=0$ 的交线 l (见图 5.10).

说白了, 直线的一般方程就是把交线等于给定直线的两个平面的方程凑到了一起. 显然, 几何观察可知: 可以有不同的两个平面, 但它们的交线相同. 而且这样子的一对平面有无穷多. 所以直线的一般方程也可以有多种形式. 例如, 在例题 5.3.10 中, 我们可以通过一些操作使得方程组

$$\begin{cases} x-3=0 \\ y-4=0 \end{cases}$$

变形, 但是其解集不变. 比如把第二行加到第一行上, 得到

$$\begin{cases} x+y-7=0 \\ y-4=0 \end{cases}$$

显然, 解集不变. 接着还可以把第二行两边同乘以 2, 得到

$$\begin{cases} x+y-7=0 \\ 2y-8=0 \end{cases}$$

显然, 解集还是不变. 接着还可以把第二行减去第一行, 得到

$$\begin{cases} x+y-7=0 \\ -x+y-1=0 \end{cases}$$

显然, 解集还是不变. 以上我们变形得到的每一组方程都表示着两个不同的平面, 可因为解集没有变化, 这意味着他们的交线从未改变, 还是如图 5.10 所示. 请在 直线的一般方程 | Desmos 上画图, 并观察它们的交线.

鉴于直线的一般方程 (两面式) 这种严重依赖于平面表达式的问题, 下面马上介绍的直线的标准方程更让人喜欢. 它直接表明了直线的方向和其经过的一点. 不再需要经过该直线的任何平面信息.

直线的参数方程与标准方程

除了一般方程之外, 直线尚有参数方程与标准方程. 假如已知一点 $P_0(x_0, y_0, z_0)$, 又已知一个非零向量 $\boldsymbol{e}=(a,b,c)$, 那么过点 P_0 且方向与向量 $\boldsymbol{e}$ 平行的直线有且仅有一条. 我们想通过 P_0 和向量 $\boldsymbol{e}$ 来确定该直线方程, 如何可以我们想进一步求出它的一般方程 (就能知道它是哪两个平面的交线).

考虑满足这样要求的直线, 那么上任意一点 $P(x,y,z)$ 都使得向量 $\overrightarrow{P_0P}$ 与 $\boldsymbol{e}$ 共线, 即存在一个实数 t, 使得

$$\overrightarrow{P_0P} = t\boldsymbol{e}, \quad \forall t \in \mathbf{R}.$$

写成坐标形式即有

$$\begin{cases} x - x_0 = ta, \\ y - y_0 = tb, \\ z - z_0 = tc, \quad \forall t \in \mathbf{R}. \end{cases}$$

即,

$$\begin{cases} x = ta + x_0, \\ y = tb + y_0, \\ z = tc + z_0. \quad \forall t \in \mathbf{R}. \end{cases}$$

这里的实数 t 是一个依赖于于具体点 $P(x, y, z)$ 的参数, 但是因为点 P 是在直线上任意取的, 故参数 t 也能取任意实数, 故写作 $\forall t \in \mathbf{R}$. 反过来, 对于每个实数 t, 由这组方程所决定的点 (x, y, z) 必定在过点 P_0 且方向与向量 e 平行的直线上. 这组方程称为**直线的参数方程**, 其中非零向量 $\boldsymbol{e} = (a, b, c)$ 称为**直线的方向向量**.

结论 5.3.5: 直线的参数方程与标准方程

假如已知直线上一点 $P_0\left(x_0, y_0, z_0\right)$, 和其方向向量 $\boldsymbol{e} = (a, b, c)$. 设该直线上任意一点 $P(x, y, z)$, 则

$$\overrightarrow{P_0 P} = t\boldsymbol{e}, \quad \forall t \in \mathbf{R}.$$

称为直线的参数方程. 从参数方程出发, 消去参数 t, 又可将直线的方程**形式上写成**:

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}.$$

上式通常称为**直线的标准方程 (假的方程)**.

直线的标准方程只是一个形式上的写法, 最好不要把它当成一个正常的方程. 应当把直线的标准方程看作是一个 "特定的标记写法", 它只是我们从参数方程推导得出一般方程 (两面式方程) 的一种手段, 或说 "中转站", 或说临时记号. 它是一种假的方程, 是因为标准方程中的分母允许出现 0. (尽管方向向量 $\boldsymbol{e} = (a, b, c)$ 是非零向量, 但是不排除各别坐标等于零.) 下面我们具体来分析所有可能情况:

1. 当分母中的 a, b, c 均不为零时. 此时, 我们可以把直线的标准方程拆分为等价的联立方程组:[12]

$$\begin{cases} \dfrac{x - x_0}{a} = \dfrac{y - y_0}{b}, \\ \dfrac{x - x_0}{a} = \dfrac{z - z_0}{c}. \end{cases}$$

于是可以化成

$$\begin{cases} a\left(y - y_0\right) - b\left(x - x_0\right) = 0, \\ a\left(z - z_0\right) - c\left(x - x_0\right) = 0. \end{cases}$$

再次化简后, 就可以得到直线的一般方程 (两面式方程). 这说明, 当 a, b, c 均不为零时, 我们只需正常推导, 就可以把标准 $\dfrac{x - x_0}{a} = \dfrac{y - y_0}{b} = \dfrac{z - z_0}{c}$ 化成一般方程. 这个过程中没有任何特殊情况, 此时的标准方程可以当成普通的方程组来处理, 而下面内容才是要重点小心的.

[12] 具体拆分的路数自由选择: 还可以有 $\begin{cases} \dfrac{x - x_0}{a} = \dfrac{y - y_0}{b}, \\ \dfrac{y - y_0}{b} = \dfrac{z - z_0}{c}. \end{cases}$ 或者 $\begin{cases} \dfrac{x - x_0}{a} = \dfrac{z - z_0}{c}, \\ \dfrac{y - y_0}{b} = \dfrac{z - z_0}{c}. \end{cases}$ 但最终结果没有区别.

2. 当分母中 $a=0$ 而 b,c 不为零时. 此时, 标准方程就变成了

$$\frac{x-x_0}{0}=\frac{y-y_0}{b}=\frac{z-z_0}{c}.$$

显然, $\dfrac{x-x_0}{0}$ 是没有意义的, 所以我们抛去它, 只留下有意义的 $\dfrac{y-y_0}{b}=\dfrac{z-z_0}{c}$. 它可以化成等价的方程

$$b\left(z-z_0\right)-c\left(y-y_0\right)=0.$$

事情尚未结束, 我们还是要想办法处理 $a=0$ 的情况. 回到直线的参数方程中, 把 $a=0$ 代入到参数方程的第一个等式中, 就得到

$$x-x_0=0.$$

综合以上考虑, 我们就直接约定 (或说定义) 标准方程这一记号:

$$\text{整体看作是一种标记写法} \rightarrow \left(\frac{x-x_0}{0}=\frac{y-y_0}{b}=\frac{z-z_0}{c}\right)$$

就代表联立方程组:

$$\begin{cases} x-x_0=0, \\ b\left(z-z_0\right)-c\left(y-y_0\right)=0. \end{cases}$$

再次化简后, 就可以得到直线的一般方程. 其他情况 (当 $b=0$ 而 a,c 不为零; 当 $c=0$ 而 a,b 不为零) 做类似的相应约定. 全部汇总就是:

$$\left(\frac{x-x_0}{0}=\frac{y-y_0}{b}=\frac{z-z_0}{c}\right) \xRightarrow{\text{定义为}} \begin{cases} x-x_0=0, \\ b\left(z-z_0\right)-c\left(y-y_0\right)=0. \end{cases}$$

$$\left(\frac{x-x_0}{a}=\frac{y-y_0}{0}=\frac{z-z_0}{c}\right) \xRightarrow{\text{定义为}} \begin{cases} y-y_0=0, \\ a\left(z-z_0\right)-c\left(x-x_0\right)=0. \end{cases}$$

$$\left(\frac{x-x_0}{a}=\frac{y-y_0}{b}=\frac{z-z_0}{0}\right) \xRightarrow{\text{定义为}} \begin{cases} z-z_0=0, \\ a\left(y-y_0\right)-b\left(x-x_0\right)=0. \end{cases}$$

3. 当分母中 $a=b=0$ 而 c 不为零时. 此时, 标准方程就变成了

$$\frac{x-x_0}{0}=\frac{y-y_0}{0}=\frac{z-z_0}{c}.$$

显然, $\dfrac{x-x_0}{0}$, $\dfrac{y-y_0}{0}$ 都是没有意义的, 所以我们抛去它们. 而 $\dfrac{z-z_0}{c}$ 独木难支, 也没有留下的必要. 此时, 应当直接把 $a=b=0$ 代回到参数方程中, 就得到

$$\begin{cases} x-x_0=0, \\ y-y_0=0, \\ z=tc+z_0, \quad \forall t\in\mathbf{R}. \end{cases}$$

注意到参数 t 只存在于最后一行, 且 $c\neq 0$, 所以函数 $t\mapsto z=tc+z_0$ 能够取到任何实数值. 也就是说, 对于直线上的任意点 $P(x,y,z)$ 的 z 分量, 我们没有任何约束. 故, 可以抛弃掉最后一行, 只留下第一行和第二行方程. 综合以上考虑, 我们就直接约定 (或说定义) 标准方程:

$$\text{整体看作是一种标记写法} \rightarrow \left(\frac{x-x_0}{0}=\frac{y-y_0}{0}=\frac{z-z_0}{c}\right)$$

就代表联立方程组:

$$\begin{cases} x-x_0=0, \\ y-y_0=0. \end{cases}$$

这样的约定显然是合理的. 当 $a=b=0$ 时, 意味着直线的方向平行于向量 $\boldsymbol{e}=(0,0,c)$, 即平行于 z 轴, 因此过点 (x_0,y_0,z_0) 且方向平行于向量 (a,b,c) 的直线应该是平面 $x-x_0=0$ 与 $y-y_0=0$ 的交线. 其他情况 (当 $a=c=0$ 而 b 不为零; 当 $b=c=0$ 而 a 不为零) 做类似的相应约定. 全部汇总就是:

$$\left(\frac{x-x_0}{a}=\frac{y-y_0}{0}=\frac{z-z_0}{0}\right) \xlongequal{\text{定义为}} \begin{cases} y-y_0=0, \\ z-z_0=0. \end{cases}$$

$$\left(\frac{x-x_0}{0}=\frac{y-y_0}{b}=\frac{z-z_0}{0}\right) \xlongequal{\text{定义为}} \begin{cases} x-x_0=0, \\ z-z_0=0. \end{cases}$$

$$\left(\frac{x-x_0}{0}=\frac{y-y_0}{0}=\frac{z-z_0}{c}\right) \xlongequal{\text{定义为}} \begin{cases} x-x_0=0, \\ y-y_0=0. \end{cases}$$

上述说明同时告诉我们如何将一条直线的标准方程化成其一般方程. 反过来, 如果已知一条直线的一般方程

$$\begin{cases} A_1x+B_1y+C_1z+D_1=0, \\ A_2x+B_2y+C_2z+D_2=0, \end{cases}$$

其中 A_1,B_1,C_1 与 A_2,B_2,C_2 不成比例, 那么如何确定其标准方程呢? 显然, 我们只需要找到该直线上的任意一点的坐标, 以及和直线方向平行的一个向量 (方向向量) 即可写出参数方程, 进而得到标准方程. 具体地说:

1. **确定该直线上的任意一点的坐标.** 首先要求出这个联立方程组的一个解 (x_0,y_0,z_0). 通常这是容易的. 我们发现, 我们可以从三个变量 x,y,z 中任意选择一个, 假如是 x. 然后给 x 任意赋值, 比如 $x=0$. 将 $x=0$ 代回方程组得到:

$$\begin{cases} B_1y+C_1z=-D_1, \\ B_2y+C_2z=-D_2. \end{cases}$$

   而这是一个简单地二元一次方程组. 我们通常可以解出来.[13] 如果运气不太好, 最后解不出来的话, 那么只要重新给三个变量 x,y,z 中任意一个赋任意值, 重复这个过程直到能算出来.

2. **确定该直线的方向向量.** 其次, 要确定这条直线的方向. 几何知识表明: *作为两个平面的交线, 这条直线一定同时垂直于这两个平面各自的法线.* (该直线同时存在于两个平面中, 而各自的法向量垂直于各个平面内任意的直线.) 因此, 这条直线的方向向量为

$$\boldsymbol{e}=\boldsymbol{n}_1\times\boldsymbol{n}_2=(A_1,B_1,C_1)\times(A_2,B_2,C_2)\,.$$

3. **写出标准方程.** 最后, 根据这条直线上的一个已知点 $P_0(x_0,y_0,z_0)$ 和方向向量 $\boldsymbol{e}=\boldsymbol{n}_1\times\boldsymbol{n}_2$ 就可得出这条直线的标准方程.

[13]线性代数告诉我们, 这个二元一方程组有且仅有唯一解的充要条件是 2 阶行列式 $\begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix}\neq 0$. 且这和右侧的参数 D_1,D_2 无关.

结论 5.3.6: 直线的方程 - 本小节总结

- 若直线是某两个已知平面的交线. 则可以写出直线的一般方程 (两面式方程).
- 已知直线上一点坐标和其方向向量, 则可以写出直线的参数方程 $\to$ 标准方程.
- 从直线的标准方程可以获得直线的一般方程 (两面式方程);
- 从直线的一般方程 (两面式方程) 可以得到直线上一点坐标和其方向向量, 从而得到直线的标准方程.

例 5.3.12 将直线的一般方程

$$\begin{cases} x-y+z-1=0 \\ 2x+y-3z=0 \end{cases}$$

化成标准方程.

证明. 为了找到直线上一点, 我们假设 $z=0$, 将 $z=0$ 代入两方程组:

$$\begin{cases} x-y=1, \\ 2x+y=0. \end{cases}$$

解出: $x=\frac{1}{3}$, $y=-\frac{2}{3}$. 因此, 直线上的一点为 $P\left(\frac{1}{3},-\frac{2}{3},0\right)$. 而该直线的方向向量为

$$(1,-1,1)\times(2,1,-3)=(2,5,3),$$

于是该直线的标准方程为

$$\frac{x-\frac{1}{3}}{2}=\frac{y+\frac{2}{3}}{5}=\frac{z}{3}.$$

□

例 5.3.13 设直线 L 过点 P_0 且其方向向量为 v, 证明: 直线 L 外一点 P_1 到直线 L 的距离 d 可表示为

$$d=\frac{\left|\overrightarrow{P_0P_1}\times v\right|}{|\boldsymbol{v}|}.$$

证明. 画图, 并且从叉乘的模的定义就能得到此结果. 请读者自己证明. □

补充: 直线相关的垂直和平行关系

前文我们讨论了两个平面之间垂直和平行的关系. 接下来, 我们将添加涉及直线的垂直和平行位置关系. 它们的共性在于: 直线的方向由其方向向量决定, 而平面的方向由其法向量决定. 因此, 所有类型的垂直和平行关系本质上都可以归结为两个法向量之间, 两个方向向量之间, 或者方向向量与法向量之间的垂直或平行关系.

两直线的垂直和平行关系. 设直线 L_1 和 L_2 的方向向量依次为 $\boldsymbol{e}_1=(m_1,n_1,p_1)$ 和 $\boldsymbol{e}_2=(m_2,n_2,p_2)$.

1. 两直线 L_1 和 L_2 互相垂直相当于
$$\boldsymbol{e}_1\cdot\boldsymbol{e}_2=0,$$
即, $m_1m_2+n_1n_2+p_1p_2=0$.

2. 两直线 L_1 和 L_2 互相平行 (不重合的前提下) 相当于
$$\boldsymbol{e}_1=\lambda\boldsymbol{e}_2,\quad \exists\lambda\in\mathbf{R},$$
即, $\dfrac{m_1}{m_2}=\dfrac{n_1}{n_2}=\dfrac{p_1}{p_2}$.

直线和平面的垂直和平行关系. 设直线的方向向量为 $\boldsymbol{e}=(m,n,p)$, 平面的法线向量为 $\boldsymbol{n}=(A,B,C)$.

1. 因为直线与平面垂直相当于直线的方向向量与平面的法线向量平行, 所以, 直线与平面垂直相当于
$$\boldsymbol{e}=\lambda\boldsymbol{n},\quad \exists\lambda\in\mathbf{R},$$
即, $\dfrac{A}{m}=\dfrac{B}{n}=\dfrac{C}{p}$.

2. 因为直线与平面平行或直线在平面上相当于直线的方向向量与平面的法线向量垂直, 所以, 直线与平面平行或直线在平面上相当于
$$\boldsymbol{e}\cdot\boldsymbol{n}=0,$$
即, $Am+Bn+Cp=0$.

注释 5.3.2 本章第一节我们说过, 叉乘可以判断是否共线:

$$\boldsymbol{a}\text{ 与}\boldsymbol{b}\text{ 共线}\iff\boldsymbol{a}\times\boldsymbol{b}=\mathbf{0}.$$

此时, 你突然发现本节并没有用叉乘来判定共线, 而是用

$$\boldsymbol{a}\text{ 与}\boldsymbol{b}\text{ 共线}\iff\boldsymbol{a}=\lambda\boldsymbol{b},\text{ 或 }\boldsymbol{b}=\lambda\boldsymbol{a},\quad\exists\lambda\in\mathbf{R}.$$

其实它们是一样的. 设向量 $\boldsymbol{a}=(x_1,y_1,z_1)$, $\boldsymbol{b}=(x_2,y_2,z_2)$. 利用叉乘的坐标表示, 有方程

$$\boldsymbol{a}\times\boldsymbol{b}=\begin{vmatrix}\boldsymbol{i}&\boldsymbol{j}&\boldsymbol{k}\\x_1&y_1&z_1\\x_2&y_2&z_2\end{vmatrix}=\begin{vmatrix}y_1&z_1\\y_2&z_2\end{vmatrix}\boldsymbol{i}-\begin{vmatrix}x_1&z_1\\x_2&z_2\end{vmatrix}\boldsymbol{j}+\begin{vmatrix}x_1&y_1\\x_2&y_2\end{vmatrix}\boldsymbol{k}=\mathbf{0}.$$

这个式子成立, 当且仅当有

$$\begin{cases}\begin{vmatrix}y_1&z_1\\y_2&z_2\end{vmatrix}=y_1z_2-z_1y_2=0\iff\dfrac{y_1}{y_2}=\dfrac{z_1}{z_2},\\\begin{vmatrix}x_1&z_1\\x_2&z_2\end{vmatrix}=z_1x_2-x_1z_2=0\iff\dfrac{x_1}{x_2}=\dfrac{z_1}{z_2},\\\begin{vmatrix}x_1&y_1\\x_2&y_2\end{vmatrix}=x_1y_2-y_1x_2=0\iff\dfrac{x_1}{x_2}=\dfrac{y_1}{y_2}.\end{cases}$$

在结束这一节时我们指出: 这一节中公式较多, 但都无须死记硬背, 重要的是理解平面方程中系数的几何意义, 特别是要知道平面方程

$$Ax + By + Cz + D = 0$$

中前三个系数 A, B, C 组成的向量 (A, B, C) 是平面的法向量. 这一节中的许多公式都是根据这一事实再利用点乘与叉乘运算得到的.

5.4 二次曲面

在上一节中, 我们已经看到一个三元一次方程在空间中代表一个平面. 在本节中, 我们要讨论一个三元二次方程所代表的曲面.

定义 5.4.1: 二次曲面

三元二次方程的一般形式是

$$Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fzx + Gx + Hy + Iz + J = 0,$$

其中 $A, B, \cdots, J$ 是常数, 且 A, B, C 不全为零 (否则和二次没有关系了). 在空间中取定一个直角坐标系 $Oxyz$ 之后, 任意给定这样的一个方程, 除去个别情况之外, 满足这个方程的全体点 (x, y, z) 会形成一个曲面. 这种曲面称为**二次曲面 (quadric surface)**. 请任意调整参数并观察曲面的图像: 二次曲面的一般方程 | Desmos.

先简单地说一下什么是**坐标变换 (coordinate transform)**. 在空间中取定一个坐标系 $Oxyz$ 其实有无数种办法. 而直角坐标系只是一种特殊的情况: 它要求三个坐标轴相互正交 (垂直), 且三个轴的刻度 (单位长度) 相同. 其实一般的坐标系, 可以打破这些规则. 一般地, 空间中任意三个不共面且相交于同一点的三条直线, 在指定了轴上的刻度和正的方向后, 都可以视为一个空间坐标系. 不同坐标系主要体现在: 1. 如何决定坐标原点 O 的位置; 2. 如何决定坐标轴上单位长度的具体尺度 (可以互不相等); 3. 如何决定三个坐标轴两两之间的角度大小. 所谓坐标变换, 就是讲一个坐标系转换成另一坐标系.

从几何视角看, 一个曲面自始至终和坐标系无关, 它本身是不会变换的. 可以预见, 同一个曲面在不同的坐标系下的曲面方程的表达式通常是不同的. 有的坐标系下, 它的方程很复杂; 但有的坐标系下, 它的方程又很简单. 通过变换坐标系, 把复杂的曲面方程变成简单的曲面方程, 这是坐标变换的重要意义.

可想象一下平面坐标系中一条直线 $y = kx$, 其中 $k \neq 0$. 我们固定原点不变, 只需要把 x, y 轴绕原点做旋转, 那么做这条直线的表达式就可能变成 $y = 0$, 或者 $x = 0$. 这就是一个简单的例子. 几何上看, 这个直线从未动过, 动的是坐标系. 空间中的也是类似的, 任何一个平面, 其一般方程是 $Ax + By + Cz + D = 0$ (假设在直角坐标系下). 那么, 只需要讲坐标原点移动到这个平面上, 那么一定有 $D = 0$. 然后整个坐标系绕着原点旋转, 总是能讲该平面跟和 Oxy 平面重合. 在此新坐标系下, 平面方程就是 $z = 0$. 真正掌握坐标变换需要学习线性代数, 这里我们不展开讲解.[14]

[14]超纲内容: 关于坐标变换, 可以参考 【熟肉】线性代数的本质 - 09 - 基变换 _ 哔哩哔哩 _bilibili.

结论 5.4.1

从几何或代数上可以证明, 对于任意给定的一个三元二次方程, 通过一个适当的坐标变换, 一般可将它化成下面 9 种典型方程其中之一. 这些方程大多数只剩下了二次项, 因此形式上比原来的一般方程简单地多.

下面分别介绍这 9 种典型方程及它们所表示的曲面.

注意 本节内容和前面没什么关系. 就是记一记 9 种二次曲线的方程表达式, 名字和大致图形. 阅读完本节内容后, 请再参考 B 站各类视频, 总结记忆规律. 比如 5 分钟记住! 单叶双曲面? 双叶双曲面? 二次曲面不再忘记 | 考研数学 _ 哔哩哔哩 _bilibili , 理解性记忆常见空间曲面 _ 哔哩哔哩 _bilibili

公式	Desmos 链接
(1) 椭圆锥面: $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0\ (a,b,c>0)$	椭圆锥面 | Desmos
(2) 椭球面: $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1\ (a,b,c>0)$	椭球面 | Desmos
(3) 单叶双曲面: $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1\ (a,b,c>0)$	单叶双曲面 | Desmos
(4) 双叶双曲面: $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1\ (a,b,c>0)$	双叶双曲面 | Desmos
(5) 椭圆柱面: $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1\ (a,b>0)$	椭圆柱面 | Desmos
(6) 双曲柱面: $\frac{x^2}{a^2}-\frac{y^2}{b^2}=1\ (a,b>0)$	双曲柱面 | Desmos
(7) 椭圆抛物面: $\frac{x^2}{a^2}+\frac{y^2}{b^2}-z=0\ (a,b>0)$	椭圆抛物面 | Desmos
(8) 双曲抛物面: $\frac{x^2}{a^2}-\frac{y^2}{b^2}-z=0\ (a,b>0)$	双曲抛物面 | Desmos
(9) 抛物柱面: $\frac{x^2}{a^2}-y=0\ (a>0)$	抛物柱面 | Desmos

表 5.1: 9 种曲面及其 Desmos 链接. **一定要通过动画演示来理解!**

1. 椭圆锥面: $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0\ (a,b,c>0)$. 显然, 这个曲面与平行于 Oxy 平面的平面 $z=z_0\ (z_0\neq 0)$ 相截的截痕是一个椭圆周:

$$\begin{cases} \frac{x^2}{a^2}+\frac{y^2}{b^2}=\frac{z_0^2}{c^2}, \\ z=z_0. \end{cases}$$

而这个椭圆周的两个半轴分别是 $\frac{a}{c}|z_0|$ 与 $\frac{b}{c}|z_0|$. 也就是说, 该椭圆周的半轴与 $|z_0|$ 成正比 (见图 5.20). 另外, 这个曲面与 Oyz 平面相截的截痕是两条直线:

$$\begin{cases} \frac{y}{b} = \frac{z}{c}, \\ x = 0, \end{cases} \quad \begin{cases} \frac{y}{b} = -\frac{z}{c}, \\ x = 0. \end{cases}$$

而与 Ozx 平面相截的截痕也是两条直线:

$$\begin{cases} \frac{x}{a} = \frac{z}{c}, \\ y = 0, \end{cases} \quad \begin{cases} \frac{x}{a} = -\frac{z}{c} \\ y = 0. \end{cases}$$

我们之所以称这个曲面为椭圆锥面, 是因为这个曲面由一束过原点的直线所组成. 事实上, 若点 (x_0, y_0, z_0) (不是原点) 在该曲面上, 则由原点与点 (x_0, y_0, z_0) 所决定的直线均在该曲面上 (请读者自行验证).

2. 椭球面: $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ $(a, b, c > 0)$. 这个曲面是可以装在长立方体

$$\{(x, y, z) |\ |x| \leqslant a, |y| \leqslant b, |z| \leqslant c\}$$

内的, 而且与 x 轴, y 轴, z 轴交点的相应坐标分量分别是 $\pm a, \pm b, \pm c$ (见图 5.20). 用平行于 Oxy 平面的平面 $z = z_0$ $(0 \leq |z_0| < c)$ 去截这个曲面, 得到的截痕是一个椭圆周:

$$\begin{cases} \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 - \frac{z_0^2}{c^2}, \\ z = z_0. \end{cases}$$

若用平行于其他坐标平面的平面去截这个曲面, 结果类似. 椭球面的参数方程为

$$\begin{cases} x = a \sin\varphi \cos\theta, \\ y = b \sin\varphi \sin\theta, \\ z = c \cos\varphi, \end{cases}$$

其中 $0 \leqslant \theta < 2\pi$, $0 \leqslant \varphi \leqslant \pi$. 请将上述参数表达代入原方程中验证. 当 $a = b = c = R$ 时, 椭球面变成球面

$$x^2 + y^2 + z^2 = R^2.$$

3. 单叶双曲面: $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$ $(a, b, c > 0)$. 用平面 $z = z_0$ 去截这个曲面, 则截痕是一个椭圆周:

$$\begin{cases} \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 + \frac{z_0^2}{c^2}, \\ z = z_0. \end{cases}$$

其两个半轴分别为 $a\sqrt{1 + \frac{z_0^2}{c^2}}$ 与 $b\sqrt{1 + \frac{z_0^2}{c^2}}$, 它们随着 z_0 趋向于无穷大而趋向于正无穷大 (见图 5.20).

用平面 $x = x_0$ 或 $y = y_0$ 去截这个曲面, 截痕是双曲线 ($|x_0| \neq a, |y_0| \neq b$ 时), 或者是两条直线 ($|x_0| = a$ 或 $|y_0| = b$ 时). 我们之所以称这个曲面为单叶双曲面, 是因为它是连在一起的一个曲面. 而下面将要讨论的双叶双曲面是彼此不连通的两个曲面.

4. 双叶双曲面: $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}-\dfrac{z^2}{c^2}=-1\ (a,b,c>0)$. 显然, 当 $|z|<c$ 时, 这个曲面的方程无解. 因此, 在平面 $z=c$ 与 $z=-c$ 之间没有这个曲面上的点; 而在这两个平面上, 各只有这个曲面上的一个点. 用平面 $z=z_0\,(|z_0|>c)$ 去截这个曲面时, 截痕是一个椭圆周:

$$\begin{cases} \dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=\dfrac{z_0^2}{c^2}-1, \\ z=z_0. \end{cases}$$

可见, 这个曲面由上, 下两部分组成. 若用平面 $x=x_0$ 或 $y=y_0$ 去截这个曲面, 则截痕总是双曲线 (见图 5.20).

5. 椭圆柱面: $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1\ (a,b>0)$. 这个曲面在 Oxy 平面上的投影是一个椭圆周. 若 (x_0,y_0) 满足上述方程, 则对于任意的 z, 点 (x_0,y_0,z) 均在这个曲面上, 因此过点 $(x_0,y_0,0)$ 且平行于 z 轴的直线落在这个曲面上. 这个曲面也可以看作由平行于 z 轴的动直线沿 Oxy 平面上的一个椭圆周平行移动的结果 (见图 5.20).

定义 5.4.2: 柱面

一般来说, 由平行于某给定方向的动直线沿着给定的曲线移动所得到的曲面称作**柱面**, 其中动直线称作**母线**, 给定的曲线称作**准线**. 柱面–GeoGebra.

上述椭圆柱面的准线为椭圆周, 其母线平行于 z 轴.

6. 双曲柱面: $\dfrac{x^2}{a^2}-\dfrac{y^2}{b^2}=1\ (a,b>0)$. 这个曲面是以 Oxy 平面上的双曲线

$$\begin{cases} \dfrac{x^2}{a^2}-\dfrac{y^2}{b^2}=1, \\ z=0. \end{cases}$$

为准线的柱面, 其母线平行于 z 轴 (见图 5.20).

7. 椭圆抛物面: $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}-z=0\ (a,b>0)$. 当 $z<0$ 时, 这个曲面的方程无解, 故这个曲面只在 $z\geqslant 0$ 时存在. 对于任意的 $z_0>0$, 这个曲面与平面 $z=z_0$ 相截的截痕是一个椭圆周, 而这个曲面与任意平面 $x=x_0$ 或 $y=y_0$ 相截, 其截痕是抛物线, 如图 5.20 所示.

8. 双曲抛物面: $\dfrac{x^2}{a^2}-\dfrac{y^2}{b^2}-z=0\ (a,b>0)$. 用任意一个平行于 Oxy 平面的平面 $z=z_0\,(z_0\neq 0)$ 去截这个曲面时, 截痕都是双曲线:

$$\begin{cases} \dfrac{x^2}{a^2}-\dfrac{y^2}{b^2}=z_0, \\ z=z_0. \end{cases}$$

不过, 当 $z_0>0$ 或 $z_0<0$ 时, 该双曲线的实轴平行于 x 轴或 y 轴:

1. 当 $z_0>0$ 时, 上述双曲线的方程可化为

$$\begin{cases} \dfrac{x^2}{\left(a\sqrt{z_0}\right)^2}-\dfrac{y^2}{\left(b\sqrt{z_0}\right)^2}=1 \\ z=z_0 \end{cases}$$

这时该双曲线在 Oxy 平面上的投影以 x 轴为实轴[15];

[15]复习: 1-211126114630.pdf

2. 当 $z_0 < 0$ 时, 上述双曲线的方程为

$$\begin{cases} -\dfrac{x^2}{(a\sqrt{-z_0})^2} + \dfrac{y^2}{(b\sqrt{-z_0})^2} = 1 \\ z = z_0. \end{cases}$$

这时该双曲线在 Oxy 平面上的投影以 y 轴为实轴.

3. 这个曲面与平面 $z = 0$ 的交线恰是两条直线.

用平行于 Ozx 平面的平面 $y = y_0$ 和平行于 Oyz 平面的平面 $x = x_0$ 去截这个曲面时, 截痕均是抛物线. 前者的截痕是抛物线

$$\begin{cases} z = \dfrac{x^2}{a^2} - \dfrac{y_0^2}{b^2}, \\ y = y_0. \end{cases} \qquad \text{(开口向上)}$$

后者的截痕是抛物线

$$\begin{cases} z = \dfrac{x_0^2}{a^2} - \dfrac{y^2}{b^2}, \\ x = x_0. \end{cases} \qquad \text{(开口向下)}$$

双曲抛物面形如一个马鞍, 因此有时也称之为马鞍面 (见图 5.20).

9. 抛物柱面: $\dfrac{x^2}{a^2} - y = 0\ (a > 0)$. 这是一个以 Oxy 平面上的抛物线

$$\begin{cases} y = \dfrac{x^2}{a^2} \\ z = 0. \end{cases}$$

为准线的柱面, 其母线平行于 z 轴, 见图 5.20.

5.5 空间曲线的切线与弧长

本节讨论一般的空间曲线, 其主要工具是向量及微分学.

5.5.1 空间曲线

定义 5.5.1: 曲线

一般来说, 空间**曲线 (curve)** 是指从区间 $[a, b]$ 到空间 $\mathbf{R}^3$ 的一个连续映射 γ 的像 *(image)*[a]. 这里所谓的 $[a, b]$ 到 $\mathbf{R}^3$ 的连续映射

$$\gamma : [a, b] \to \mathbf{R}^3, \quad t \mapsto (x(t), y(t), z(t)),$$

是指对于每一点 $t \in [a, b]$, 都有唯一确定的点, 记作

$$\gamma(t) = (x(t), y(t), z(t)) \in \mathbf{R}^3.$$

并且当 t 在 $[a, b]$ 上连续变动时, 像点 $(x(t), y(t), z(t))$ 也连续变动. 等价地说, $x = x(t)$, $y = y(t)$, $z = z(t)$ 是 $[a, b]$ 上的三个连续实数值函数.

[a] image 的概念来自函数 function 的定义, 请自行复习.

当我们将 t 看作时间时, 那么空间曲线就可以理解为一个质点在空间中运动的轨迹 (见图 5.21). 类似的, 我们之前学习过的平面中的参数方程, 也就是平面中的曲线. 一条空间曲线可以用坐标形式的方程给出:

$$\begin{cases} x = x(t), \\ y = y(t), \quad a \leqslant t \leqslant b, \\ z = z(t), \end{cases}$$

这称之为该曲线的参数方程, 其中 t 称为参数. 点 a 与 b 所对应的点 $(x(a), y(a), z(a))$ 与 $(x(b), y(b), z(b))$ 都称作空间曲线的端点. 参数 t 也可以在一个开区间 (a, b) 上变动, 这时一般不能谈论曲线的端点.

例 5.5.1 考虑参数方程

$$\begin{cases} x = R\cos t, \\ y = R\sin t, \quad 0 \leqslant t \leqslant 2\pi \\ z = bt, \end{cases}$$

显然它表示柱面 $x^2 + y^2 = R^2$ 上的一条曲线, 其中 R, b 为常数且 $R > 0$ (见图 5.22).

假定在空间中取定了直角坐标系 $Oxyz$ 及相应的坐标向量 $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$, 这时空间中任何一点 $P(x, y, z)$ 与原点 O 连线构成的向量为 $\overrightarrow{OP} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$. 为了方便起见, 我们将 $\overrightarrow{OP}$ 记作 $\boldsymbol{r}$, 即有

$$\boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}.$$

今后, 我们将这个向量 $\boldsymbol{r}$ 看作点 $P(x, y, z)$ 的向量表示. 既然空间中的点 $P(x, y, z)$ 可以由向量 $\boldsymbol{r} = \overrightarrow{OP}$ 表示, 那么一条空间曲线

$$\begin{cases} x = x(t), \\ y = y(t), \quad a \leqslant t \leqslant b, \\ z = z(t), \end{cases}$$

就可以由一个向量函数 $\boldsymbol{r} = \boldsymbol{r}(t)$ $(a \leqslant t \leqslant b)$ 表示, 其中

$$\boldsymbol{r}(t) = x(t)\boldsymbol{i} + y(t)\boldsymbol{j} + z(t)\boldsymbol{k}, \quad a \leqslant t \leqslant b,$$

这就是**空间曲线的向量表示**. 这里再解释一下所谓 “向量函数”, 是指它的像 image 是向量, 也就是说其到达域是向量空间

$$\mathbb{V} = \{\boldsymbol{v} \mid \boldsymbol{v} \text{ 是一个自由向量, 具有大小和方向}\}.$$

也就是, $\boldsymbol{r} : [a, b] \to \mathbb{V} \cong \mathbf{R}^3$, $t \mapsto \boldsymbol{r}(t) \in \mathbb{V} \cong \mathbf{R}^3$.

注释 5.5.1 复习: 我们已经证明:

1. 空间中全体点 $\xLeftrightarrow{\text{一一对应}}$ $\mathbf{R}^3$ (通过建立空间直角坐标系)
2. 空间中全体点 $\xLeftrightarrow{\text{一一对应}}$ $\mathbb{V}$ (通过将向量 $\boldsymbol{v}$ 的起点移至坐标原点 O)

即证明了 $\mathbb{V} \xLeftrightarrow{\text{一一对应}} \mathbf{R}^3$, 通常我们写做 $\mathbb{V} \cong \mathbf{R}^3$.

定义 5.5.2: 光滑曲线

我们称空间曲线

$$\begin{cases} x = x(t) \\ y = y(t), \quad a \leqslant t \leqslant b \\ z = z(t) \end{cases}$$

是一条光滑曲线，如果 $x(t), y(t), z(t)$ 在区间 $[a, b]$ 上有连续的导数，并且对于每一点 $t \in [a, b]$，$x'(t), y'(t), z'(t)$ 不同时为零.

这里要求 $x'(t), y'(t), z'(t)$ 不同时为零，实际上就相当于要求曲线在每一点都有切线，而 $x'(t), y'(t), z'(t)$ 的连续性则表明曲线的切线随参数 t 而连续变动. 具体解释如下:

事实上, 我们知道切线是割线的极限位置 (见图 5.23). 过一点 $\boldsymbol{r}(t)$ 的切线的方向应当是

$$\begin{aligned} \boldsymbol{r}'(t) &:= \lim_{\Delta t \to 0} \frac{\boldsymbol{r}(t+\Delta t) - \boldsymbol{r}(t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{x(t+\Delta t) - x(t)}{\Delta t}\boldsymbol{i} + \lim_{\Delta t \to 0} \frac{y(t+\Delta t) - y(t)}{\Delta t}\boldsymbol{j} + \lim_{\Delta t \to 0} \frac{z(t+\Delta t) - z(t)}{\Delta t}\boldsymbol{k} \\ &= x'(t)\boldsymbol{i} + y'(t)\boldsymbol{j} + z'(t)\boldsymbol{k}. \end{aligned}$$

可见, $x'(t), y'(t), z'(t)$ 不同时为零的假定保证了极限 (注意这是一个向量)

$$\boldsymbol{r}'(t)$$

不退化成一个零向量, 从而在点 $\boldsymbol{r}(t)$ 处有确定的切线方向. 现在我们讨论光滑曲线的切线方程及弧长的计算公式.

5.5.2 光滑曲线的切线和法平面方程

复习: 给定直线上任意一点 P_0 和其方向向量 $\boldsymbol{e}$, 那么直线上任意一点 P 都满足 (其中 u 是参数)

$$\overrightarrow{P_0P} = u\boldsymbol{e}, \quad u \in \mathbf{R}.$$

如果选定坐标原点 O, 则 $\overrightarrow{P_0P} = \overrightarrow{OP} - \overrightarrow{OP_0}$. 上式可以写做

$$\overrightarrow{OP} = \overrightarrow{OP_0} + u\boldsymbol{e}, \quad u \in \mathbf{R}.$$

此时, 考虑光滑曲线 $\boldsymbol{r} = \boldsymbol{r}(t)$ $(a \leqslant t \leqslant b)$ 在点 $\boldsymbol{r}(t_0) = (x(t_0), y(t_0), z(t_0))$ 处的切线方程. 显然, 切线的方向向量是 $\boldsymbol{r}'(t_0)$, 故方程是 (其中 u 是参数)

$$\boldsymbol{r} = \boldsymbol{r}(t_0) + u\boldsymbol{r}'(t_0), \quad u \in \mathbf{R},$$

其中, 该点处的切向量为 $\boldsymbol{r}'(t_0) = (x'(t_0), y'(t_0), z'(t_0))$. 写成参数方程的形式为

$$\begin{cases} x = x(t_0) + ux'(t_0), \\ y = y(t_0) + uy'(t_0), \\ z = z(t_0) + uz'(t_0). \end{cases}$$

由此易得该曲线在点 $\boldsymbol{r}(t_0)$ 处的切线的标准方程是

$$\frac{x - x(t_0)}{x'(t_0)} = \frac{y - y(t_0)}{y'(t_0)} = \frac{z - z(t_0)}{z'(t_0)}.$$

显然, 当 $\Delta t>0$ 且充分小时, $\boldsymbol{r}\left(t_0+\Delta t\right)-\boldsymbol{r}\left(t_0\right)$ 的方向与 $\boldsymbol{r}'\left(t_0\right)$ 的方向接近于一致. 因此, $\boldsymbol{r}'\left(t_0\right)$ 作为曲线 $\boldsymbol{r}=\boldsymbol{r}(t)$ $(a \leqslant t \leqslant b)$ 在点 $\boldsymbol{r}\left(t_0\right)$ 处的切线方向, 它指向参数 t 增加的方向. 见图 5.23.

过点 $\boldsymbol{r}\left(t_0\right)=\left(x\left(t_0\right), y\left(t_0\right), z\left(t_0\right)\right)$ 且与该点处的切向量 $\boldsymbol{r}'\left(t_0\right)=\left(x'\left(t_0\right), y'\left(t_0\right), z'\left(t_0\right)\right)$ 垂直的平面称作曲线 $\boldsymbol{r}=\boldsymbol{r}(t)$ $(a \leqslant t \leqslant b)$ 在该点处的法平面. 则这个法平面的点法式方程是:

$$x'\left(t_0\right)\left(x-x\left(t_0\right)\right)+y'\left(t_0\right)\left(y-y\left(t_0\right)\right)+z'\left(t_0\right)\left(z-z\left(t_0\right)\right)=0.$$

例 5.5.2 求曲线

$$\begin{cases} x=a\cos t \\ y=a\sin t, \\ z=bt \end{cases}$$

在 $\left(0, a, \dfrac{b\pi}{2}\right)$ 处的切线方程及法平面方程.

证明. 点 $\left(0, a, \dfrac{b\pi}{2}\right)$ 对应于 $t=\dfrac{\pi}{2}$. 在该点处, 切向量为 $(-a, 0, b)$, 于是切线方程为

$$\frac{x-0}{-a}=\frac{y-a}{0}=\frac{z-\frac{b\pi}{2}}{b},$$

而法平面方程为

$$-a(x-0)+b\left(z-\frac{\pi b}{2}\right)=0.$$

□

5.5.3 光滑曲线的弧长的计算公式

光滑曲线总是可以谈论弧长的. 设 $\boldsymbol{r}=\boldsymbol{r}(t)$ $(a \leqslant t \leqslant b)$ 是一条光滑曲线. 这条曲线的弧长 s 应当是用折线逼近曲线时折线长度的极限, 即

$$s=\lim_{i \rightarrow 0} \sum_{i=1}^{n}\left|\boldsymbol{r}\left(t_i\right)-\boldsymbol{r}\left(t_{i-1}\right)\right|,$$

其中 $a=t_0<t_1<\cdots<t_{n-1}<t_n=b$ 是对区间 $[a, b]$ 的任意一种分割, $\lambda=\max\left\{t_i-t_{i-1} \mid i=1,2, \cdots, n\right\}$. 显然, 有

$$\begin{aligned}
\left|\boldsymbol{r}\left(t_i\right)-\boldsymbol{r}\left(t_{i-1}\right)\right| &=\sqrt{\left(x\left(t_i\right)-x\left(t_{i-1}\right)\right)^2+\left(y\left(t_i\right)-y\left(t_{i-1}\right)\right)^2+\left(z\left(t_i\right)-z\left(t_{i-1}\right)\right)^2} \\
&\approx \sqrt{\left(x'\left(t_{i-1}\right)\right)^2+\left(y'\left(t_{i-1}\right)\right)^2+\left(z'\left(t_{i-1}\right)\right)^2}\left(t_i-t_{i-1}\right) \\
&=\left|\boldsymbol{r}'\left(t_{i-1}\right)\right| \Delta t_i,
\end{aligned}$$

所以

$$\sum_{i=1}^{n}\left|\boldsymbol{r}\left(t_i\right)-\boldsymbol{r}\left(t_{i-1}\right)\right| \approx \sum_{i=1}^{n}\left|\boldsymbol{r}'\left(t_{i-1}\right)\right| \Delta t_i.$$

可以严格证明, 在曲线光滑的条件下上式两个和式有相同的极限. 于是, 我们得到曲线弧长的计算公式.[16] 参考 Arc Length of a Curve (2D and 3D) –GeoGebra.

[16] 我们目前没法严谨证明.

结论 5.5.1: 光滑曲线的弧长的计算公式

设 $\boldsymbol{r}=\boldsymbol{r}(t)$ $(a\leqslant t\leqslant b)$ 是一条光滑曲线. 这条曲线的弧长 s 为

$$s=\int_a^b|\boldsymbol{r}'(t)|\,\mathrm{d}t \quad 或 \quad s=\int_a^b\sqrt{(x'(t))^2+(y'(t))^2+(z'(t))^2}\,\mathrm{d}t.$$

如果不是计算整条曲线的弧长, 而是计算自端点 $\boldsymbol{r}(a)$ 到点 $\boldsymbol{r}(t)$ 的弧长 $s(t)$, 那么

$$s(t)=\int_a^t|\boldsymbol{r}'(t)|\,\mathrm{d}t=\int_a^t\sqrt{(x'(t))^2+(y'(t))^2+(z'(t))^2}\,\mathrm{d}t.$$

由 $|\boldsymbol{r}'(t)|$ 的连续性可知, 函数 $s=s(t)$ 对 t 是可微的, 且

$$\mathrm{d}s=|\boldsymbol{r}'(t)|\,\mathrm{d}t \quad 或 \quad \mathrm{d}s=\sqrt{(x'(t))^2+(y'(t))^2+(z'(t))^2}\,\mathrm{d}t.$$

这就是所谓的弧微分公式, 它是平面曲线弧微分公式的推广.

例 5.5.3 求曲线

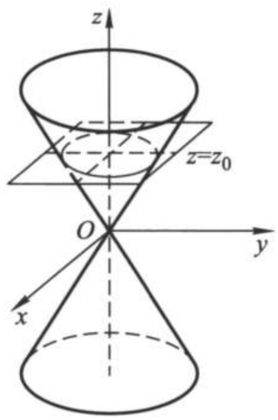
$$\begin{cases} x=a\cos t \\ y=a\sin t, \\ z=bt \end{cases} \quad (a>0,b>0,0\leqslant t\leqslant\pi)$$

的弧长 s. Arc Length of a Curve (2D and 3D) –GeoGebra.

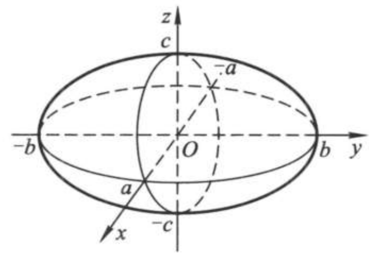
证明. 根据曲线弧长的计算公式, 我们有

$$s=\int_0^\pi\sqrt{a^2\sin^2t+a^2\cos^2t+b^2}\,\mathrm{d}t=\pi\sqrt{a^2+b^2}.$$

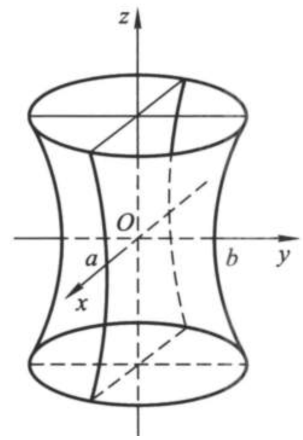
□



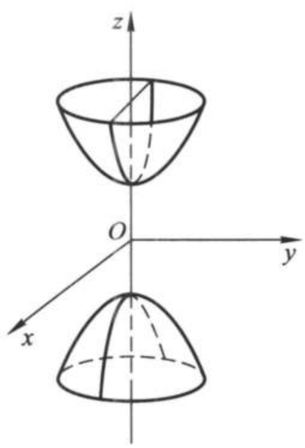
椭圆锥面 $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0.$



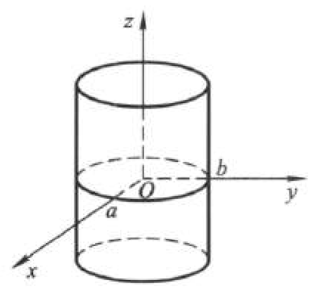
椭球面 $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1.$



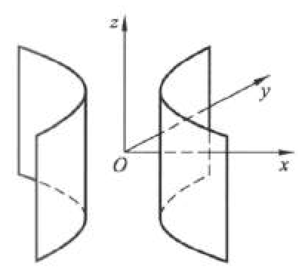
单叶双曲面 $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1$



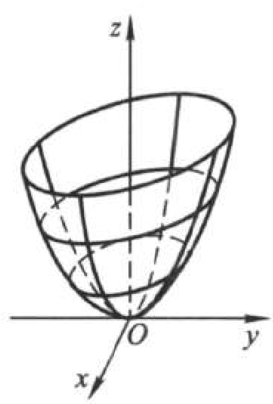
双叶双曲面 $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1.$



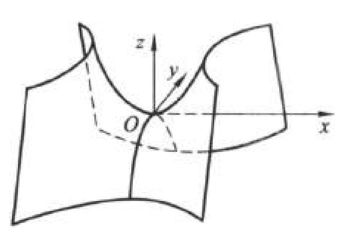
椭圆柱面 $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1.$



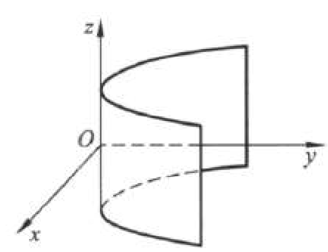
双曲柱面 $\frac{x^2}{a^2}-\frac{y^2}{b^2}=1.$



椭圆抛物面 $\frac{x^2}{a^2}+\frac{y^2}{b^2}-z=0.$



双曲抛物面 $\frac{x^2}{a^2}-\frac{y^2}{b^2}-z=0.$



抛物柱面 $\frac{x^2}{a^2}-y=0.$

图 5.20: 9 种二曲线.

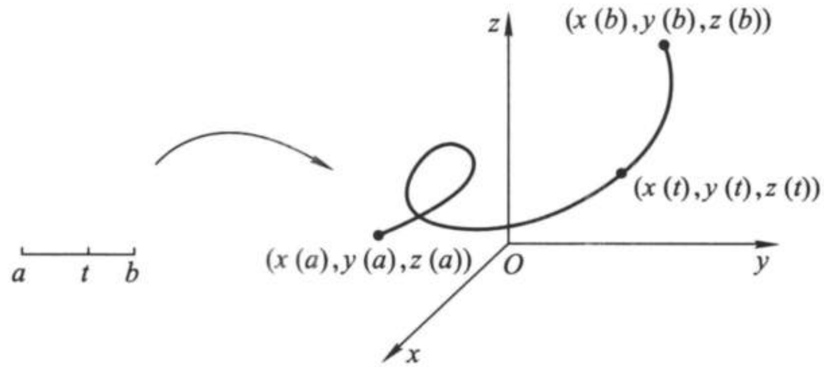


图 5.21: 空间曲线

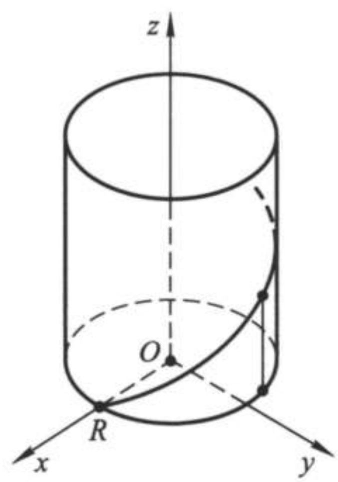


图 5.22: 柱面 $x^2 + y^2 = R^2$ 上的一条曲线.

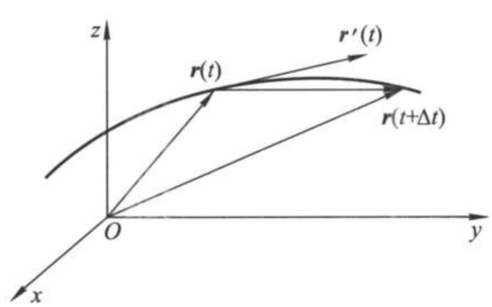


图 5.23: Enter Caption

第六章　多元函数微分学

在前面几章中, 我们讨论了只含有一个自变量的函数的微积分, 即*一元函数微积分*. 但是, 许多问题中所涉及的函数是含有多于一个自变量的函数, 即*多元函数*[1]. 因此, 将一元函数微分学与积分学推广到多元函数的情形是十分必要的, 也是十分自然的. 本章中我们先讨论多元函数微分学, 而将多元函数积分学留到下一学期.

多元函数微分学是以一元函数微分学为基础的, 两者之间既有许多类似之处, 又在某些地方有实质性差别. 对于这一点, 读者应当在学习中特别留意.

虽然一元函数微分学与多元函数微分学有某些实质性差别, 但二元函数微分学与三元及三元以上函数微分学并无实质性差别. 因此, 为了简单起见, 在本章的大部分叙述中以讨论二元函数为主. 对于含有更多个自变量的多元函数的情况, 我们相信读者自己能够得出相应的结论.[2]

6.1　多元函数

6.1.1　多元函数的概念

多元函数就是含有两个及两个以上自变量的函数.

- (余弦定理)[3] 例如, 三角形一边的长度 c 是另外两边的长度 a 与 b 及其夹角 θ 的函数:

$$c = \sqrt{a^2 + b^2 - 2ab\cos\theta}.$$

[1]当初 GPT-3 的参数量就已经达到了 175B (1750 亿). 所谓 “training (训练)” 大语言模型, 就是在对一个含有 175B 个自变量的函数 (称为损失函数) 求其极小值的过程. 这个过程需要反复迭代, 所以也叫训练, 就像人不断地重复才能逐渐会学新东西一样.

[2]在没有学习线性代数之前, 对于 3 个变量以上的微积分的讨论, 从数学符号上看是非常冗杂的. 将来同学们学习线性代数之后, 再去看进阶版的微积分教材.

[3]余弦定理是三角形中三边长度与一个角的余弦值 (cos) 的数学式, 余弦定理指的是 (见图 6.1):

$$c^2 = a^2 + b^2 - 2ab\cos\gamma.$$

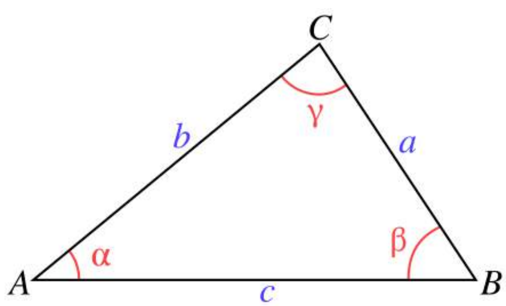


图 6.1: 余弦定理

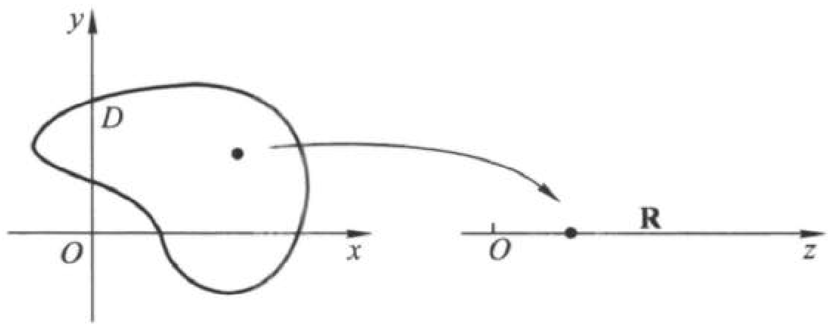


图 6.2: 二元函数

- 有一定质量理想气体的压强 p 是其体积 V 及温度 T 的函数:

$$p = k\frac{T}{V} \quad (k \text{ 为常数}).$$

这里 c 是三个自变量的函数, 而 p 是两个自变量的函数.

为了描述多个自变量，我们需要将它们按照一定次序加以排列而形成一个有序数组的形式. 例如, 在前面的例子中, 自变量形成的有序数组分别是 (a, b, θ) 与 (T, V).

对二元函数和三元函数可做一个几何上的解释: 我们将两个自变量所形成的有序数组, 如上面的 (T, V), 看作平面上的一个点; 而将三个自变量所形成的有序数组, 如上面的 (a, b, θ), 看作空间中的一个点. 当一个二元函数的两个自变量在一定的允许范围内变化时, 相应的有序数组对应于平面上的某个点集合. 在这种看法下, 一个二元函数实质上就是平面上某个点集合 D 到实数集 $\mathbf{R}$ 的一个映射 (见图 6.2). 同样, 一个三元函数实质上就是空间中某个点集合 Ω 到实数集 $\mathbf{R}$ 的一个映射.

今后, 我们将全体二元有序实数组 (x, y) 所组成的集合记作 $\mathbf{R}^2$, 即

$$\mathbf{R}^2 = \{(x, y) \mid x, y \in \mathbf{R}\}.$$

同时, 我们将全体三元有序实数组 (x, y, z) 所组成的集合记作 $\mathbf{R}^3$, 即

$$\mathbf{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbf{R}\}.$$

过去, 我们把实数集 $\mathbf{R}$ 与具有坐标的直线 (数轴) 上的全体点等同起来. 今后, 我们将把 $\mathbf{R}^2$ 和 $\mathbf{R}^3$ 分别与具有坐标的平面和空间中的全体点等同起来. (这个事情已经重复 n 多遍了.)

其中 c 是 γ 角的对边, 而 a 和 b 是 γ 角的邻边. 同样, 也可以将其改为:

$$b^2 = c^2 + a^2 - 2ca\cos\beta,$$
$$a^2 = b^2 + c^2 - 2bc\cos\alpha.$$

下面我们用向量的语言来证明余弦定理. 请先自行复习向量加法的三角形法则. 让 $\mathbf{v}$ 和 $\mathbf{w}$ 是不共线的两个向量. 角度 θ 是 $\mathbf{v}$ 和 $\mathbf{w}$ 之间夹角. 余弦定理就是要证明:

$$|\mathbf{v} - \mathbf{w}|^2 = |\mathbf{v}|^2 + |\mathbf{w}|^2 - 2|\mathbf{v}||\mathbf{w}|\cos\theta.$$

现在观察到:

$$\begin{aligned}
|\mathbf{v} - \mathbf{w}|^2 &= (\mathbf{v} - \mathbf{w}) \cdot (\mathbf{v} - \mathbf{w}) && \text{(向量与自身的内积)} \\
&= \mathbf{v} \cdot \mathbf{v} - 2(\mathbf{v} \cdot \mathbf{w}) + \mathbf{w} \cdot \mathbf{w} && \text{(内积的分配性)} \\
&= |\mathbf{v}|^2 + |\mathbf{w}|^2 - 2(\mathbf{v} \cdot \mathbf{w}) && \text{(向量与自身的内积)} \\
&= |\mathbf{v}|^2 + |\mathbf{w}|^2 - 2|\mathbf{v}||\mathbf{w}|\cos\theta. && \text{(内积的定义)}
\end{aligned}$$

这正是我们想要的结果.

在各种实际问题中，我们还会遇到自变量多于三个的函数. 比如，许多物理量不仅依赖于物体在空间中的位置 (这需要用三个变量来描述)，而且还依赖于时间，像这样的物理量则是一个含有四个自变量的函数，称为四元函数.

定义 6.1.1: $\mathbf{R}^n$

一般来说，为了描述 n 元函数的自变量，需要考虑 n 元有序实数组

$$(x_1, x_2, \cdots, x_n)$$

以及由全体这种数组所组成的集合

$$\mathbf{R}^n = \{(x_1, \cdots, x_n) \mid x_j \in \mathbf{R},\ j = 1, \cdots, n\}.$$

我们把每个这样的数组 $(x_1, \cdots, x_n)$ 称作 $\mathbf{R}^n$ 中的一个点.

现在我们给出多元函数的正式定义.

定义 6.1.2: n 元函数，n 元实值函数

- 设有一个集合 $D \subset \mathbf{R}^n$. 若对于 D 中每一点 $(x_1, \cdots, x_n)$，按照一定的法则 f，都有唯一确定的数 $u \in \mathbf{R}$ 与之相对应，则称 f 是定义在 D 上的一个 n 元函数 (简称函数).
- 这里 D 称为 f 的定义域.
- 与点 $(x_1, \cdots, x_n)$ 相对应的数 u 称为 f 在点 $(x_1, \cdots, x_n)$ 处的值，或者像 (image)，并记为 $f(x_1, \cdots, x_n)$.
- 全体函数值组成的集合

  $$\mathbf{R} \supseteq f(D) = \{f(x_1, \cdots, x_n) \mid (x_1, \cdots, x_n) \in D\}$$

  称为 f 的值域.
- 通常把点 $(x_1, \cdots, x_n)$ 称作函数 f 的自变量，而把 u 称作函数 f 的因变量. 这时，我们也称 u 是 $(x_1, \cdots, x_n)$ 的函数，记作

  $$u = f(x_1, \cdots, x_n).$$
- 我们经常写做:

  $$f : D \subset \mathbf{R}^n \to \mathbf{R}, \quad (x_1, \cdots, x_n) \mapsto f(x_1, \cdots, x_n)$$

显然，在前面关于三角形边长 (余弦定理) 的例子中，c 作为 (a, b, θ) 的函数，其定义域为

$$\{(a, b, \theta) \mid a, b > 0,\ 0 < \theta < \pi\}$$

而其值域则是 $(0, +\infty)$.

像一元函数情形一样，多元函数在很多情况下是由一个表达式给出，这时函数的 (自然) 定义域 D 是一切使得表达式有数学意义的点 $(x_1, \cdots, x_n)$ 所组成的集合.

注释 6.1.1 因为习惯上将 $\mathbf{R}^2$ 中的点用 (x,y) 表示, 而 $\mathbf{R}^3$ 中的点用 (x,y,z) 表示, 所以我们习惯分别用 $z=f(x,y)$ 与 $u=f(x,y,z)$ 表示通常的二元函数与三元函数.

前面定义多元函数时, n 元函数是作为 $\mathbf{R}^n$ 中的一个集合 D 到 $\mathbf{R}$ 的映射给出的. 但是, 二元函数可以有更为直观的几何解释, 这就是**函数图形/图像 (graph of function)** 的概念. 函数的图像使得我们能够以几何的方式直观地理解函数的整个性态.

1. 我们已经熟悉将一个一元函数 $y=f(x)$ 视作平面上的一个图像. 更确切地说, 将它视作平面上一切满足 $y=f(x)$ 的点 (x,y) 组成的集合. 数学上表示就是, 集合

$$\{(x,y) \mid y=f(x), x\in D\}$$

称为函数 $y=f(x)$ 的图像. 在 Desmos | 图形计算器 中输入任何一个一元函数的表达式, 就会立马绘制出它的图像. 请自行尝试.

2. 对于二元函数 $z=f(x,y)$, 也可以有类似的解释. 在取定直角坐标系 $Oxyz$ 的空间中, 集合

$$\{(x,y,z) \mid z=f(x,y), (x,y)\in D\}$$

称为函数 $z=f(x,y)$ 的图像, 这里 D 是该函数的定义域. 在 Desmos | 3D 图形计算器 中输入任何一个二元函数的表达式, 就会立马绘制出它的图像. 请自行尝试.

3. 对于 n $(n\geqslant 3)$ 元函数 $f(x_1,x_2,\ldots,x_n)$, 其图像的定义可以类比于一元函数和二元函数的情形. 具体地, 给定一个 n 元函数 $f: D\subseteq\mathbb{R}^n\to\mathbb{R}$, 其图像为:

$$\{(x_1,x_2,\ldots,x_n,z) \mid z=f(x_1,x_2,\ldots,x_n), (x_1,x_2,\ldots,x_n)\in D\}.$$

n 元函数的图形需要借助于 $n+1$ 维空间, 故这时函数的图形就只能想象而不能画出了, 这对于我们研究 n 元函数而言并无实际用途.

例 6.1.1 函数 $z=\sqrt{r^2-(x^2+y^2)}$ 的定义域为

$$D=\left\{(x,y) \mid x^2+y^2\leqslant r^2\right\}.$$

它是 Oxy 平面上的一个圆盘 (圆周及其内部), 其圆心为原点, 半径为 r. 这个函数的图像则是位于 Oxy 平面上方的半个球面.

例 6.1.2 函数

$$z=\sqrt{y-x^2}+\ln\left(4-\left(x^2+y^2\right)\right)$$

的定义域是一切满足

$$\begin{cases} y\geqslant x^2, \\ x^2+y^2<4 \end{cases}$$

的点 (x,y) 组成的集合. 第六章例 2 图 | Desmos 在 Oxy 平面上, 这个集合中的点位于抛物线 $y=x^2$ 上或其上方且在圆周 $x^2+y^2=4$ 的内部, 如图 6.3 中阴影部分所示. 这个函数的图像是空间中的一个曲面, 它在 Oxy 平面上的投影恰是图 6.3 中阴影部分.

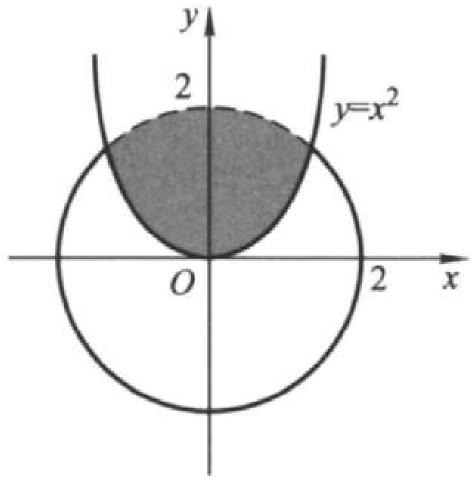


图 6.3: 二元函数 $z=\sqrt{y-x^2}+\ln\left(4-\left(x^2+y^2\right)\right)$ 的定义域.

注意　确定二元函数的定义域是非常重要的技能. 特别是, 我们往往要求明确其定义域在平面上的对应的几何区域. 这是未来很多内容的基础.

6.1.2　$\mathbf{R}^n$ 中的集合到 $\mathbf{R}^m$ 的映射

上面讨论的 n 元函数实质上就是 $\mathbf{R}^n$ 中一个集合到 $\mathbf{R}$ 的映射. 这个概念的一般化就是 $\mathbf{R}^n$ 中的一个集合到 $\mathbf{R}^m$ 的映射. 正整数 n 和 m 可以相等也可以不等. 这种映射在数学或物理学中是经常用到的. 我们先看两个例子.

例 6.1.3　考虑平面曲线的参数方程

$$\begin{cases} x=\varphi(t), \\ y=\psi(t), \end{cases} \quad (\alpha \leqslant t \leqslant \beta),$$

其中 $\varphi(t)$ 与 $\psi(t)$ 是区间 $[\alpha,\beta]$ 上的连续函数. 这实质上是 $\mathbf{R}$ 中的 $[\alpha,\beta]$ 到 $\mathbf{R}^2$ 的一个映射. 我们把这一映射用符号 f 表示: $f:[\alpha,\beta]\to\mathbf{R}^2$. 但是, 这时与函数[4]不同, 对于每个 $t\in[\alpha,\beta]$, $f(t)$ 不再是一个值, 而应是 $\mathbf{R}^2$ 中一个点. 因此, $\boldsymbol{f}(t)$ 应由两个坐标刻画, 即

$$\boldsymbol{f}(t)=(\varphi(t),\psi(t)).$$

请读者回顾上一张最后一节的空间中的曲线, 显然, 它是一个 $\mathbf{R}$ 中的 $[\alpha,\beta]$ 到 $\mathbf{R}^3$ 的一个映射.

例 6.1.4　考虑平面上的坐标变换 (旋转变换)

$$\begin{cases} u=x\cos\alpha-y\sin\alpha, \\ v=x\sin\alpha+y\cos\alpha, \end{cases}$$

[4]本书经常把输出为实数的函数就叫做 “函数”, 而把输出为高维欧式空间中点的函数叫做 “映射”. 但我个人不喜爱这么去区分它, 并没有实际意义. 任何函数只需要定义域和到达域书写清楚了, 就不需要额外的称呼.

其中 α 为一个固定的度数. 映射 $(x, y) \mapsto (u, v)$ 是 $\mathbf{R}^2$ 到 $\mathbf{R}^2$ 的一个映射, 这里 u 与 v 是两个二元函数. 它的几何意义是, 将点 (x, y) 沿着原点, 逆时针方向旋转 α 角度 (弧度制), 得到点就是 (u, v). 直角坐標的旋转变换–GeoGebra.

定义 6.1.3: $\mathbf{R}^n$ 中的集合到 $\mathbf{R}^m$ 的映射

- 设 D 为 $\mathbf{R}^n$ 中的一个集合, 又设 f 是 $D \to \mathbf{R}^m$ 的一个映射, 则对于 D 中的每一点 $(x_1, \cdots, x_n)$, 在 $\mathbf{R}^m$ 中都有唯一确定的点 $(y_1, \cdots, y_m)$ 与之相对应.
- 这里 y_j $(j = 1, 2, \cdots, m)$ 都是由 $(x_1, \cdots, x_n)$ 所确定的, 所以 y_j $(j = 1, 2, \cdots, m)$ 是 $(x_1, \cdots, x_n)$ 的函数, 设其为 $f_j(x_1, \cdots, x_n)$. 这样一来, 上述映射 $\boldsymbol{f}$: $D \to \mathbf{R}^m$ 相当于 m 个 n 元函数的有序排列:
$$\begin{cases} y_1 = f_1(x_1, \cdots, x_n), \\ y_2 = f_2(x_1, \cdots, x_n), \\ \quad \vdots \\ y_m = f_m(x_1, \cdots, x_n). \end{cases}$$
- 因此, $\mathbf{R}^n$ 中的一个集合到 $\mathbf{R}^m$ 的映射可用 m 个有序的 n 元 *(实值)* 函数表示, 其中第 j $(j = 1, 2, \cdots, m)$ 个 n 元函数 f_j 称作**映射的第 j 个分量**.

不仅坐标变换 (例 6.1.4) 是这类映射, 而且只要我们在讨论问题时需要引用若干新变量去替换若干旧变量, 这时新, 旧变量之间就形成了一种对应关系, 这种对应关系一般来说就是 $\mathbf{R}^n$ 中的一个集合到 $\mathbf{R}^m$ 的映射, 这里 n 是旧变量的个数, 而 m 是新变量的个数.

6.1.3 $\mathbf{R}^n$ 中的距离, 邻域及开集

回顾一元函数极限的定义, 立即发现邻域的概念在其中占有重要地位. 因此, 要想对于 n 元函数建立极限概念, 首先应该在 $\mathbf{R}^n$ 中建立邻域的概念.

$\mathbf{R}^n$ 中的距离

我们回顾一下, 数轴 $\mathbf{R}$ 上邻域概念的建立是基于数轴 $\mathbf{R}$ 上两点之间的距离的. 因此, 要在 $\mathbf{R}^n$ 中建立邻域的概念, 办法之一是先在 $\mathbf{R}^n$ 中建立距离的概念.

- 我们知道, 在数轴 $\mathbf{R}$ 上点 x 到 x_0 的距离为
$$|x - x_0| = \sqrt{(x - x_0)^2},$$
- 在平面 $\mathbf{R}^2$ 中点 $P(x, y)$ 到点 $P_0(x_0, y_0)$ 的距离为
$$\sqrt{(x - x_0)^2 + (y - y_0)^2},$$
- 而在空间 $\mathbf{R}^3$ 中点 $P(x, y, z)$ 到点 $P_0(x_0, y_0, z_0)$ 的距离为
$$\sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}.$$

- 因此, 很自然地将 $\mathbf{R}^n$ 中的点 $P\left(x_1, \cdots, x_n\right)$ 到点 $P_0\left(x_1^0, \cdots, x_n^0\right)$ 的距离定义为

$$\sqrt{\sum_{i=1}^{n}\left(x_i-x_i^0\right)^2}.$$

定义 6.1.4: 距离函数

$\mathbf{R}^n$ 中的任意两点 $P\left(x_1, \cdots, x_n\right)$ 到点 $P_0\left(x_1^0, \cdots, x_n^0\right)$ 的**距离 (distance)** 定义为:

$$d: \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}, \quad d(P, P_0) := \sqrt{\sum_{i=1}^{n}\left(x_i-x_i^0\right)^2}.$$

距离函数 d 满足下列条件: $\forall P, Q, R \in \mathbf{R}^n$,

1. $d(P, Q) \geqslant 0$, 当且仅当 $P=Q$ 时等号成立;
2. $d(P, Q)=d(Q, P)$;
3. $d(P, Q) \leqslant d(P, R)+d(R, Q)$.

前面两条性质显然成立, 而第三条性质的证明将放到下一小节. 第三条中的不等式称作三角不等式, 在空间 $\mathbf{R}^3$ 中它的几何解释就是三角形的两边之和大于第三边.

一个问题是, 刻画任意两点之间距离的方式只有如上定义的一种吗? 其实不是的, 以 $\mathbf{R}^2$ 的情况为例, 我们有理由认为下面两类函数也可以是刻画任意两点差异的有效方式:

1. $d\left(\left(x_1, y_1\right),\left(x_2, y_2\right)\right)=\left|x_2-x_1\right|+\left|y_2-y_1\right|$.
2. $d\left(\left(x_1, y_1\right),\left(x_2, y_2\right)\right)=\max \left\{\left|x_2-x_1\right|,\left|y_2-y_1\right|\right\}$.

我们可以证明这两种距离函数, 仍旧符合定义 6.1.4 中距离函数 d 的三个性质. 这告诉我们, 距离函数的具体定义形式只是表象, 核心是要求用它刻画的两点之间差异时, 能够满足定义 6.1.4 中的三个性质. 这 3 条性质才是未来一切的基础. 我们通常把定义 6.1.4 中距离函数称为欧几里得距离.

欧几里得空间, 欧式空间 $\mathbf{R}^n$

像在空间 $\mathbf{R}^3$ 中一样, 我们把 $\mathbf{R}^n$ 中的每一点 $\left(x_1, \cdots, x_n\right)$ 同时也视作一个向量, 并定义两个向量 $\boldsymbol{a}=\left(x_1, \cdots, x_n\right)$ 与 $\boldsymbol{b}=\left(y_1, \cdots, y_n\right)$ 的加法运算

$$\boldsymbol{a}+\boldsymbol{b}=\left(x_1+y_1, \cdots, x_n+y_n\right),$$

以及数乘运算

$$\lambda \boldsymbol{a}=\left(\lambda x_1, \cdots, \lambda x_n\right), \quad \forall \lambda \in \mathbf{R}.$$

此外, 我们也可以定义两个向量的内积:

$$\boldsymbol{a} \cdot \boldsymbol{b}=\sum_{i=1}^{n} x_i y_i,$$

并规定

$$|\boldsymbol{a}|:=\sqrt{\boldsymbol{a} \cdot \boldsymbol{a}}$$

作为向量 $\boldsymbol{a}$ 的模 (也叫**范数 (norm)**). 上述定义的内积也满足下列规律 (其中 λ 为任意实数):

1. $\boldsymbol{a}\cdot\boldsymbol{b}=\boldsymbol{b}\cdot\boldsymbol{a}$ (交换律);
2. $(\lambda\boldsymbol{a})\cdot\boldsymbol{b}=\lambda(\boldsymbol{a}\cdot\boldsymbol{b})$ (与数乘的结合律);
3. $(\boldsymbol{a}+\boldsymbol{b})\cdot\boldsymbol{c}=\boldsymbol{a}\cdot\boldsymbol{c}+\boldsymbol{b}\cdot\boldsymbol{c}$ (分配律).

更加重要的是, $\forall\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}\in\mathbf{R}^n$, 我们有以下结论:

1. $|\boldsymbol{a}\cdot\boldsymbol{b}|\leqslant|\boldsymbol{a}||\boldsymbol{b}|$. (柯西-施瓦茨不等式)[5]
2. $|\boldsymbol{a}+\boldsymbol{b}|\leqslant|\boldsymbol{a}|+|\boldsymbol{b}|$.
3. $|\boldsymbol{a}-\boldsymbol{b}|\leqslant|\boldsymbol{a}-\boldsymbol{\gamma}|+|\boldsymbol{\gamma}-\boldsymbol{b}|$.

将点 $P(x_1,\cdots,x_n)$ 及 $Q(y_1,\cdots,y_n)$ 分别看成向量 $\boldsymbol{a}$ 及 $\boldsymbol{b}$, 则点 P 到点 Q 的欧几里得距离为

$$d(P,Q)=|\boldsymbol{a}-\boldsymbol{b}|.$$

由此, 使用第 3 项结论可导出定义 6.1.4 中三角不等式. 现在我们证明上述的 3 个结论.

证明. 1. 构造一个一元二次函数:

$$\begin{aligned}0\le f(t)&=(\boldsymbol{a}+t\boldsymbol{b})\cdot(\boldsymbol{a}+t\boldsymbol{b})\\&=\boldsymbol{a}\cdot\boldsymbol{a}+2(\boldsymbol{a}\cdot\boldsymbol{b})t+(\boldsymbol{b}\cdot\boldsymbol{b})t^2\\&=|\boldsymbol{a}|^2+2(\boldsymbol{a}\cdot\boldsymbol{b})t+|\boldsymbol{b}|^2t^2.\end{aligned}$$

则该函数的判别式 $\Delta=4(\boldsymbol{a}\cdot\boldsymbol{b})^2-4|\boldsymbol{a}|^2|\boldsymbol{b}|^2\leqslant0$. 即, $|\boldsymbol{a}\cdot\boldsymbol{b}|\leqslant|\boldsymbol{a}||\boldsymbol{b}|$.

2. 将不等式 $|\boldsymbol{a}+\boldsymbol{b}|\leqslant|\boldsymbol{a}|+|\boldsymbol{b}|$ 两边开平方:

$$|\boldsymbol{a}+\boldsymbol{b}|^2\leqslant(|\boldsymbol{a}|+|\boldsymbol{b}|)^2.$$

展开左侧得

$$|\boldsymbol{a}+\boldsymbol{b}|^2=(\boldsymbol{a}+\boldsymbol{b})\cdot(\boldsymbol{a}+\boldsymbol{b})=|\boldsymbol{a}|^2+2(\boldsymbol{a}\cdot\boldsymbol{b})+|\boldsymbol{b}|^2.$$

展开右侧得

$$(|\boldsymbol{a}|+|\boldsymbol{b}|)^2=|\boldsymbol{a}|^2+2|\boldsymbol{a}||\boldsymbol{b}|+|\boldsymbol{b}|^2.$$

因此, 可以消去不等式两侧的 $|\boldsymbol{a}|^2$ 和 $|\boldsymbol{b}|^2$, 得到

$$2\boldsymbol{a}\cdot\boldsymbol{b}\leqslant2|\boldsymbol{a}||\boldsymbol{b}|.$$

这正是第 1 项的结果. 由此可以反推回第 2 项结论.

3. 直接使用第 2 项即可得到, 只需令其中 $\boldsymbol{a}\leftarrow\boldsymbol{a}-\boldsymbol{\gamma}$, $\boldsymbol{b}\leftarrow\boldsymbol{\gamma}-\boldsymbol{b}$, 则 $\boldsymbol{a}+\boldsymbol{b}\leftarrow\boldsymbol{a}-\boldsymbol{b}$.

□

[5]如果不用向量的语言, 就一般而言, 柯西-施瓦茨不等式表述如下: 设 $x_1,x_2,\cdots,x_n$ 和 $y_1,y_2,\cdots,y_n$ 为任意实数, 则恒有不等式

$$\left(\sum_{i=1}^n x_iy_i\right)^2\leqslant\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n y_i^2\right),$$

当且仅当 $\dfrac{x_1}{y_1}=\dfrac{x_2}{y_2}=\cdots=\dfrac{x_n}{y_n}$ 时 (当某 $y_i=0$ 时, 认为 $x_i=0$) 等号成立. 这个不等式非常非常有用!

定义 6.1.5: 欧几里得空间, 欧式空间 $\mathbf{R}^n$

当我们把 $\mathbf{R}^n$ 看成一个向量空间 (线性空间) 并定义了上述加法和数乘运算以及内积后, $\mathbf{R}^n$ 则被称为**欧氏空间**, 或更确切地说, 被称为 n 维欧氏空间.

$\mathbf{R}^n$ 中的邻域

在 $\mathbf{R}^n$ 中定义了距离之后, 我们就可以在 $\mathbf{R}^n$ 中定义一点的邻域了.

定义 6.1.6: 邻域

设 $P_0 \in \mathbf{R}^n$ 为给定的一点, r 是给定的正数. 我们定义点 P_0 的 r 邻域 (简称**邻域**) 是指集合

$$U_r(P_0) = \{P \in \mathbf{R}^n \mid d(P_0, P) < r\}.$$

在不强调 r 的时候, 也用 $U(P_0)$ 表示点 P_0 的某个邻域. 总有 $P_0 \in U_r(P_0)$.

1. 当 $n = 1$ 时, 点 $P_0 \in \mathbf{R}$ 的 r 邻域是以该点为中点, r 为半径的开区间.
2. 当 $n = 2$ 时, 点 $P_0 \in \mathbf{R}^2$ 的 r 邻域是以 P_0 为中心, r 为半径的圆的内部 (不包含边界), 通常也称开圆.
3. 当 $n = 3$ 时, 点 $P_0 \in \mathbf{R}^3$ 的 r 邻域就是以点 P_0 为中心, r 为半径的球的内部 (不包含球面), 通常也称开球.

$\mathbf{R}^n$ 中的开集, 区域

在一元函数微积分中, 我们遇到的函数大多数都定义在一个区间上, 而区间是数轴上的 "一段". 但在多元函数微积分中函数的定义域较复杂些. 我们将用区域的概念来替代过去的开区间. 现在我们来定义开集及区域的概念.

定义 6.1.7: 内点, 外点与边界点

设 $E \subset \mathbf{R}^n$ 是一个给定的集合. 根据集合 E, 我们将 $\mathbf{R}^n$ 中的点分作三类: E 的内点, 外点与边界点.

1. 一点 $P \in \mathbf{R}^n$ 称为 E 的**内点**, 如果存在一个正数 r, 使得点 P 的 r 邻域整个被包含于 E:
$$U_r(P) \subset E.$$
2. 一点 $P \in \mathbf{R}^n$ 称为 E 的**外点**, 如果存在一个正数 r, 使得点 P 的 r 邻域与 E 不交 (这里 $\varnothing$ 表示空集):
$$U_r(P) \cap E = \varnothing,$$
3. 既非内点又非外点的点称为 E 的**边界点**. 一点 $P \in \mathbf{R}^n$ 是 E 的边界点, 当且仅当对于任意的正数 r, 点 P 的 r 邻域 $U_r(P)$ 中既有 E 中的点, 又有不属于 E 中的点. 今后, 我们用 ∂E 表示 E 的全体边界点组成的集合, 称为 E 的**边界**.

注意, 根据上述定义, $\mathbf{R}^n$ 中的任何点都将是且仅是 E 的内点, 外点与边界点这三种情况中的其中一类.

例 6.1.5 设集合 R 是平面 $\mathbf{R}^2$ 上一个矩形的内部:

$$R = \{(x,y) \mid -a < x < a, -b < y < b\},$$

其中 a,b 是正的常数, 则原点 $(0,0)$ 是 R 的一个内点, 点 (a,b) 是 R 的一个边界点, 点 $(2a,2b)$ 是 R 的一个外点. 更一般地说, 这个矩形内的每一点都是 R 的内点, 它的四条边上的每一点都是 R 的边界点, 而它之外 (不含边) 任意一点都是 R 的外点.

例 6.1.6 设 R_1 是一个矩形的内部及其四条边:

$$R_1 = \{(x,y) \mid -a \leqslant x \leqslant a, -b \leqslant y \leqslant b\},$$

其中 a,b 是正的常数. 显然, R_1 与例 6.1.5 中的 R 有相同的内点, 外点及边界点. R_1 区别于 R 的地方是 R_1 包含其全部边界点.

根据定义很容易看出,

- 一个集合 E 的全部内点都属于 E, 因为任何 E 的内点 P 都有 $P \in U_r(P) \subset E$.
- 一个集合 E 的全部外点都不属于 E, 因为任何 E 的外点 P 都有 $P \in U_r(P) \cap E = \varnothing$. 假如, $P \in E$, 则 $U_r(P) \cap E$ 不可能是空集. 所以只能有 $P \notin E$.
- 对于 E 的一个边界点, 则有两种可能性: 或者属于 E, 或者不属于 E.

定义 6.1.8: 开集, 闭集

- 一个集合 E 称为**开集**, 如果它的每一点都是内点. 集合 E 是开集的充要条件是 E 中没有边界点.
- 一个集合 E 称为**闭集**, 如果它包含着它的全部边界点.

在上面给, 我们声称 "E 是开集 $\Longleftrightarrow$ E 中没有边界点". 首先看一个方向: "E 是开集 $\Longrightarrow$ E 中没有边界点". 取其逆否命题: E 中存在边界点 $\Longrightarrow$ E 不是开集. 这显然是成立的. 我们再看另一个方向: "E 中没有边界点 $\Longrightarrow$ E 是开集". E 中没有边界点意味着, E 中的点只能是内点或者外点. 但是外点天然是被排除在集合当中的. 所以 E 中的点只能是内点, 故 E 是开集.

显然, 平面上任意一个矩形的内部, 任意一个圆的内部都是平面 $\mathbf{R}^2$ 中的开集. 例如, 例 6.1.6 中的 R_1 就是一个闭集.

开集与闭集恰好是两种极端情况, 前者不包含任何边界点, 而后者包含其全部边界点. 显然, 有些集合介于两者之间: 包含部分但非全部边界点, 例如

$$R_2 = \{(x,y) \mid -a < x \leqslant a, -b \leqslant y < b\}$$

(建议读者自己画出这个集合的图形). 平面 $\mathbf{R}^2$ 中这样的集合既不是开集, 也不是闭集. 就像开区间和闭区间, $\mathbf{R}$ 中存在既不是开区间也不是闭区间的半开半闭区间.

定义 6.1.9: 开区域, 闭区域

- 设 E 是 $\mathbf{R}^n$ 中的一个开集. 我们说 E 是**连通的 (connected)**, 如果 E 中任意两点都可以用一条落在 E 中的曲线相连接.
- $\mathbf{R}^n$ 中连通的非空开集称为 $\mathbf{R}^n$ 中的 (开) **区域 (region)**. "开集 + 非空 + 连通" = "区域"
- 设 G 是一个区域, 那么集合 $G \cup \partial G$ 是一个闭集. 今后, 我们记集合 $G \cup \partial G$ 为 $\bar{G}$, 并称之为**闭区域 (closed region)**.

- 在例 6.1.5 中, 集合 R 是平面 $\mathbf{R}^2$ 中一个区域.
- 集合 $E = \left\{(x, y) \mid x^2 + y^2 < 1 \text{ 或 } x^2 + y^2 > 2\right\}$ 虽然是开集, 但不是连通的, 因而不是区域.
- 开区间 (a, b) 是数轴 $\mathbf{R}$ 中的一个区域. 可见, $\mathbf{R}^n$ 中区域的概念是数轴 $\mathbf{R}$ 上开区间的推广.
- 例 6.1.6 中的集合 R_1 就是平面 $\mathbf{R}^2$ 中的一个闭区域.

定义 6.1.10: 有界集合, 无界集合

- 集合 $E \subset \mathbf{R}^n$ 称为**有界集合**, 如果存在一个正数 ρ, 使得 E 包含于以原点为中心, ρ 为半径的球内. 换句话说, 就是存在一个常数 $\rho > 0$, 使得

$$\max_{P \in E} d(O, P) < \rho.$$

- 如果不存在这样正数 ρ, 则称 E 为**无界集合**.

区域及闭区域有可能是无界的. 如平面 $\mathbf{R}^2$ 中的带形的区域

$$S = \{(x, y) \mid 0 < y < 1\}$$

及其相应的闭区域

$$\bar{S} = \{(x, y) \mid 0 \leqslant y \leqslant 1\}$$

都是无界的.

在研究多元函数微积分时,

- 我们对于区域有特殊的兴趣, 它相当于数轴 $\mathbf{R}$ 上的开区间;
- 我们也对有界闭区域有特殊的兴趣, 尤其在讨论连续函数的性质时, 它相当于数轴 $\mathbf{R}$ 上的闭区间.

6.2　多元函数的极限

本节及下面几节的叙述主要以二元函数为例, 但其中的定义与定理可以很自然地推广到一般多元函数的情况.

6.2.1 二元函数极限的概念

首先复习一下一元函数的极限概念 (仅以 $x \to a$ 为例). 设 $f(x)$ 是定义在点 x_0 的某个空心邻域

$$U_r(x_0)\backslash\{x_0\} = (x_0 - r, x_0) \cup (x_0, x_0 + r)$$

上的一个函数. 若存在一个常数 A, 对于任意给定的 $\varepsilon > 0$, 无论它多么小, 都存在一个 $\delta = \delta(\varepsilon) > 0$, 使得

$$|f(x) - l| < \varepsilon, \quad \text{只要 } 0 < |x - x_0| < \delta,$$

则我们称, 当 $x \to x_0$ 时, $f(x)$ 以 A 为极限, 记作 $\lim\limits_{x \to x_0} f(x) = A$.

二元函数的极限与一元函数的极限有许多类似之处, 但也有其复杂之处. 像一元函数的情况一样, 二元函数在一点处的极限, 就是指当动点任意接近于该点时, 动点的函数值所趋向的值. 其复杂之处在于, 动点以更高的"自由度"趋向于一个定点, 不像一元函数情形中那样动点只沿 x 轴至多从左, 右两侧趋向于定点. 为了刻画二元变量的动点如何接近于二元的定点, 我们需要用到上一节的距离函数以及由它导出领域等概念.

给定 $\mathbf{R}^2$ 中的一定点 (x_0, y_0), 我们说 (x_0, y_0) 的空心邻域是指它的某个半径为 r 的邻域, 且挖去点 (x_0, y_0) 后所成的集合, 即

$$U_r(x_0, y_0)\backslash\{(x_0, y_0)\} = \left\{(x, y) \in \mathbf{R}^2 \mid 0 < d\left((x, y), (x_0, y_0)\right) < r\right\}.$$

注意, 为了挖去点 (x_0, y_0), 只需要令 $0 < d\left((x, y), (x_0, y_0)\right)$ 即可, 因为 $0 = d\left((x, y), (x_0, y_0)\right)$ 当且仅当 $(x, y) = (x_0, y_0)$. 通常, 我们使用的距离 d 是欧几里得距离, 故上式可具体写做

$$U_r(x_0, y_0)\backslash\{(x_0, y_0)\} = \left\{(x, y) \in \mathbf{R}^2 \mid 0 < \sqrt{(x - x_0)^2 + (y - y_0)^2} < r\right\}.$$

如果记定点 (x_0, y_0) 为 P_0, 而记动点 (x, y) 为 P. 那么, 点 P_0 的空心邻域可以写做

$$U_r(P_0)\backslash\{P_0\} = \left\{P \in \mathbf{R}^2 \mid 0 < d\left(P, P_0\right) < r\right\}.$$

当点 $P_0, P \in \mathbf{R}^n$ 而不是 $P_0, P \in \mathbf{R}^2$, 空心领域的定义是完全一样的, 即

$$U_r(P_0)\backslash\{P_0\} = \{P \in \mathbf{R}^n \mid 0 < d\left(P, P_0\right) < r\}.$$

如果没有特别声明, 我们遇到的所有距离函数 d 都默认是欧几里得距离. 现在我们给出二元函数极限的正式定义, 它也是基于 $\varepsilon - \delta$ 说法的, 其核心思量和一元函数的情况完全一致.

定义 6.2.1: 多元函数的极限（使用欧几里得距离）

(具体版) 设二元函数 $f(x,y)$ 在点 (x_0,y_0) 的某个空心邻域上有定义. 若存在一个常数 A, 对于任意给定的 $\varepsilon>0$, 无论它多么小, 都存在一个 $\delta=\delta(\varepsilon)>0$, 使得

$$|f(x,y)-A|<\varepsilon,\quad \text{只要 } 0<\sqrt{(x-x_0)^2+(y-y_0)^2}<\delta,$$

则我们称, 当 $(x,y)\to(x_0,y_0)$ 时, $f(x,y)$ 以 A 为极限 (limit), 记作

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y)=A,\quad \text{或}\ \lim_{\substack{x\to x_0\\ y\to y_0}} f(x,y)=A.$$

(简洁版) 记定点 (x_0,y_0) 为 P_0, 而动点 (x,y) 为 P, 定义的简洁版形式是: 设函数 $f(P)$ 在点 P_0 的某个空心邻域上有定义. 若存在一个常数 A, 对于任意给定的 $\varepsilon>0$, 无论它多么小, 都存在一个 $\delta=\delta(\varepsilon)>0$, 使得

$$|f(P)-A|<\varepsilon,\quad \text{只要 } 0<d(P,P_0)<\delta,$$

换句话说, 只要动点 P 在点 P_0 的空心 δ 邻域内, 那么其值 $f(P)$ 就落在实数 A 的 ε 邻域内, 即

$$f\left(U_\delta(P_0)\setminus\{P_0\}\right)\subseteq U_\varepsilon(A),$$

则我们称, 当 $P\to P_0$ 时, $f(P)$ 以 A 为极限, 记作 $\lim\limits_{P\to P_0} f(P)=A.$[a]

[a]和一元函数一样, 如果 $\lim\limits_{P\to P_0} f(P)$ 存在, 那么这个极限唯一. 证明略.

简洁版相较于具体版, 只是把关于不等式的叙述可以改用距离函数或者邻域的符号叙述而已. 注意, 因为 A 是一个实数, 所以领域 $U_\varepsilon(A)$ 就是所有满足 $|z-A|<\varepsilon$ 的实数值 z 的集合罢了. 注意, 上述定义中的简洁版可以扩张到 $P_0,P\in\mathbf{R}^n$ 而不需要修改任何一个字.

例 6.2.1 考虑函数 $f(x,y)=\dfrac{x\sin y}{\sqrt{x^2+y^2}}$, 证明 $\lim\limits_{(x,y)\to(0,0)} f(x,y)=0.$

证明. 和第一章时求极限一样. 遇到求极限的题目, 第零步永远是确定函数的定义域. 我们发现其定义域是 $\mathbf{R}^2\setminus\{(0,0)\}$. 这说明该函数在 $(0,0)$ 的任何空心邻域中都是有定义的.

从定义出发, 我们要探索不等式

$$|f(x,y)-0|=\frac{|x\sin y|}{\sqrt{x^2+y^2}}<\varepsilon.$$

而条件是当 $0<\sqrt{(x-0)^2+(y-0)^2}<\delta$, 即 $0<\sqrt{x^2+y^2}<\delta=\delta(\varepsilon)$, 其中要将 δ 写成关于 ε 的函数. 故我们想对

$$\frac{|x\sin y|}{\sqrt{x^2+y^2}}$$

做适当的放大, 使得上式小于某个关于 $\sqrt{x^2+y^2}$ (把 $t=\sqrt{x^2+y^2}$ 视作一个整体) 的函数, 然后反推出 δ. 具体地, 我们发现恒有不等式[6]

$$|x\sin y|\le|xy|\le\frac{1}{2}\left(x^2+y^2\right),$$

[6]当初在第一章讲函数的极限时, 我们证明过命题 1.4.2: 不等式 $|\sin\theta|\le|\theta|$ 对任意实数 θ 都成立. 下面我们经常用到这一结果.

所以,

$$\frac{|x\sin y|}{\sqrt{x^2+y^2}} \le \frac{\frac{1}{2}\left(x^2+y^2\right)}{\sqrt{x^2+y^2}} = \frac{1}{2}\sqrt{x^2+y^2} = g(\sqrt{x^2+y^2}).$$

因此, 对于任意给定的 $\varepsilon > 0$, 取 $\delta = 2\varepsilon$, 则当 $0 < \sqrt{(x-0)^2+(y-0)^2} < \delta$ 时, 便有

$$|f(x,y)-0| < \varepsilon.$$

这就证明了当 $(x,y) \to (0,0)$ 时, $f(x,y)$ 以零为极限, 即

$$\lim_{(x,y)\to(0,0)} f(x,y) = 0.$$

证毕. □

注意 多元函数的极限的难点 (如何放大不等式的链条) 本质上和一元函数的极限是一样的. 不熟练的同学还是顺带复习一下一元函数极限的定义证明题.

上一节我们提到过, 刻画 $\mathbf{R}^n$ 中的任意两点 $P\left(x_1,\cdots,x_n\right)$ 和 $P_0\left(x_1^0,\cdots,x_n^0\right)$ 之间差异性的距离函数 $d:\mathbf{R}^n\times\mathbf{R}^n\to\mathbf{R}$ 不只有最常用的欧几里得距离

$$d(P,P_0) = \sqrt{\sum_{i=1}^{n}\left(x_i-x_i^0\right)^2}$$

这一种方式. 还有**切比雪夫距离 (Chebyshev distance)**:

$$d(P,P_0) = \max_{i=1,\ldots,n}\left\{\left|x_i-x_i^0\right|\right\},$$

和**曼哈顿距离 (Manhattan distance)**: $d(P,P_0) = \sum_{i=1}^{n}\left|x_i-x_i^0\right|$. 我们发现这些距离在 $n=1$ 时, 都退化成 $|x-x^0|$. 事实上, 但凡满足以下三个性质的函数 $d:\mathbf{R}^n\times\mathbf{R}^n\to\mathbf{R}$ 都可以称作 $\mathbf{R}^n$ 中的距离函数: 对于所有点 $P,Q,R\in\mathbf{R}^n$,

1. $d(P,Q)\ge 0$, 当且仅当 $P=Q$ 时等号成立. (正定性)
2. $d(P,Q)=d(Q,P)$. (对称性)
3. $d(P,Q)\le d(P,R)+d(R,Q)$. (三角不等式)

更多距离函数可以参考 Distance Metrics.

一个自然而然的问题是: 定义 6.2.1 中对多元函数的极限定义显示是非常依赖于领域的概念, 而领域又依赖于具体选定的距离函数. 那么如果选择不同的距离函数, 是否有可能使得多元函数的极限值不同呢? 甚至会出现使用欧几里得距离时极限存在, 而用别的距离时极限不存在的情况吗? 下面的内容告诉我们, 任何距离函数导出的极限概念都是等价的.

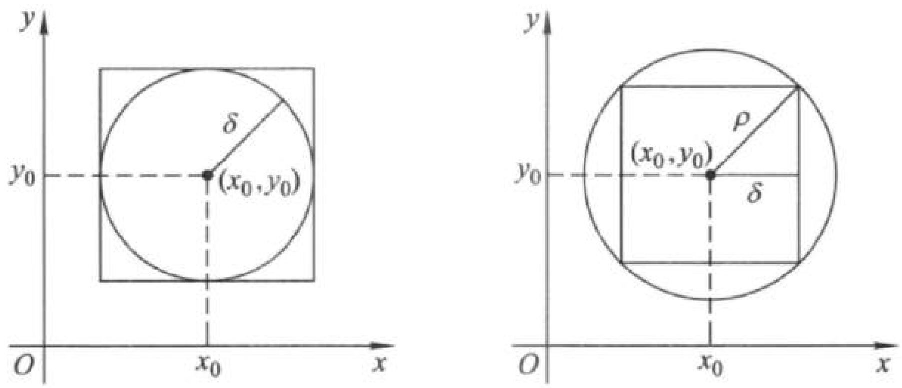


图 6.4: 证明用图: 定义 6.2.2 与定义 6.2.1 是等价的.

定义 6.2.2: 多元函数的极限 (使用切比雪夫距离)

(具体版) 设二元函数 $f(x,y)$ 在点 (x_0,y_0) 的某个空心邻域上有定义. 若存在一个常数 A, 对于任意给定的 $\varepsilon>0$, 无论它多么小, 都存在一个 $\delta=\delta(\varepsilon)>0$, 使得[a]

$$|f(x,y)-A|<\varepsilon, \quad 只要\ |x-x_0|<\delta,\ |y-y_0|<\delta,\ (x,y)\neq(x_0,y_0),$$

则我们称, 当 $(x,y)\to(x_0,y_0)$ 时, $f(x,y)$ 以 A 为极限, 记作

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y)=A, \quad 或\ \lim_{\substack{x\to x_0\\ y\to y_0}} f(x,y)=A.$$

(简洁版) 记定点 (x_0,y_0) 为 P_0, 而动点 (x,y) 为 P, 定义的简洁版形式是: … 略 …. 这里和定义 6.2.1 中的简洁版完全一致, 只需将距离 d 替换成切比雪夫距离.

[a] "只要" 后面的条件可以等价地改写成: $0<|x-x_0|<\delta$, $0<|y-y_0|<\delta$, 或者 $\max\{|x-x_0|,|y-y_0|\}<\delta$.

定理 6.2.1: 多元函数的极限概念不依赖于距离函数

定义 6.2.2 与定义 6.2.1 是等价的. 一般而言, 无论使用哪种距离函数, 只要它能够满足距离函数定义中的三条性质, 那么它推导出来的极限概念皆是等价的, 这包括: 极限的存在性; 如果存在, 极限值相等.

证明. 我们只考虑二元函数的情况, 并且只证明欧几里距离和切比雪夫距离导出的极限的等价性. 推广到任意多元函数, 以及任意两个距离函数的证明思路是差不多的.

1. (在定义 6.2.2 的意义下, $f(x,y)$ 以 A 为极限 $\Longrightarrow$ 在定义 6.2.1 的意义下, $f(x,y)$ 也以 A 为极限): 设 $f(x,y)$ 在定义 6.2.2 的意义下以 A 为极限, 那么对于任意给定的 $\varepsilon>0$, 都存在一个 $\delta>0$, 使得当

$$|x-x_0|<\delta, \quad |y-y_0|<\delta, \quad 且 \quad (x,y)\neq(x_0,y_0)$$

时, 就有 $|f(x,y)-A|<\varepsilon$. 这样的 δ 一定满足定义 6.2.1 中的要求, 因为当

$$\sqrt{(x-x_0)^2+(y-y_0)^2}<\delta$$

时, 必有 $|x-x_0|<\delta$ 与 $|y-y_0|<\delta$ (见图 6.4 的左图). 这样, 我们有

$$|f(x,y)-A|<\varepsilon, \quad 只要 \quad 0<\sqrt{(x-x_0)^2+(y-y_0)^2}<\delta.$$

可见, 在定义 6.2.1 的意义下, $f(x, y)$ 也以 A 为极限.

2. (在定义 6.2.1 的意义下, $f(x, y)$ 以 A 为极限 $\Longrightarrow$ 在定义 6.2.2 的意义下, $f(x, y)$ 也以 A 为极限): 假定 $f(x, y)$ 按照定义 6.2.1 当 $(x, y) \to (x_0, y_0)$ 时以 A 为极限, 那么对于任意给定的 $\varepsilon > 0$, 都存在一个 $\rho(\varepsilon) > 0$, 使得当 $0 < \sqrt{(x - x_0)^2 + (y - y_0)^2} < \rho(\varepsilon)$ 时, 就有

$$|f(x, y) - A| < \varepsilon.$$

取 $\delta(\varepsilon) := \dfrac{\rho(\varepsilon)}{\sqrt{2}}$, 则当 $|x - x_0| < \delta(\varepsilon)$, $|y - y_0| < \delta(\varepsilon)$, 且 $(x, y) \neq (x_0, y_0)$ 时, 就有 (见图 6.4 的右图)

$$0 < \sqrt{(x - x_0)^2 + (y - y_0)^2} < \rho(\varepsilon)$$

从而推出 $|f(x, y) - A| < \varepsilon$. 这表明, $f(x, y)$ 依照定义 6.2.2 也以 A 为极限. 证毕.

□

在有些情况下, 可能使用定义 6.2.1 较为方便 (如例 6.2.1); 而在另外一些情况下, 可能使用定义 6.2.2 较为方便 (见下题). 在讨论具体问题时使用哪种定义, 这要视问题中所涉及函数的表达式而定.

例 6.2.2 设函数 $f(x, y) = x^2 + \sin y$, 证明: 当 $(x, y) \to (3, 0)$ 时, $f(x, y)$ 的极限为 9.

证明. 该函数在 $\mathbf{R}^2$ 中都是有定义的. 我们知道

$$|f(x, y) - 9| = \left|x^2 - 9 + \sin y\right| \le \left|x^2 - 9\right| + |\sin y| \le |x + 3||x - 3| + |y|.$$

这样, 当 $|x - 3| < 1$ 时, 有 $2 < x < 4$, 故 $|x + 3| < 7$. 因此, 当 $|x - 3| < 1$ 时,

$$|x + 3||x - 3| < 7|x - 3|.$$

于是, 对于任意给定的 $\varepsilon > 0$, 取 $\delta = \delta(\varepsilon) = \min\left\{1, \dfrac{1}{8}\varepsilon\right\}$, 则当 $|x - 3| < \delta$, $|y - 0| < \delta$ 时, 便有

$$|f(x, y) - 9| \le |x + 3||x - 3| + |y| < 7|x - 3| + |y| < \frac{7}{8}\varepsilon + \frac{1}{8}\varepsilon = \varepsilon.$$

这就证明了我们的结论. 证毕. □

注意我们在例 6.2.1 中用了定义 6.2.2, 而在例 6.2.2 中用了定义 6.2.1. 综合前面众多的极限概念, 最方便就是记住以下定义形式: 设函数 $f(P)$ 在定点 P_0 的某个空心邻域上有定义. 若存在一个常数 A, 对于任意给定的 $\varepsilon > 0$, 无论它多么小, 都存在一个 $\delta = \delta(\varepsilon) > 0$, 使得

$$|f(P) - A| < \varepsilon, \qquad 只要\ 0 < d(P, P_0) < \delta,$$

则我们称 $\lim\limits_{P \to P_0} f(P) = A$. 这个定义无论使用哪种具体距离函数 d, 还是对于一元函数, 二元函数, 任意 n 元函数都是不需要修改任意一个字的.

注意 到目前为止我们只是在讨论到达域是实数的多元函数. 那么如果到达域也是多元的欧式空间 $\mathbf{R}^m$ 的话, 如何定义极限呢?

给定区域 $D \subseteq \mathbf{R}^n$, 设函数 $f: D \subseteq \mathbf{R}^n \to \mathbf{R}^m$. 又设 P_0 是区域 D 中的一点. 若存在一个常向量 $A \in \mathbf{R}^m$, 对于任意给定的 $\varepsilon > 0$, 无论它多么小, 都存在一个 $\delta = \delta(\varepsilon) > 0$, 使得

$$d\left(f(P), A\right) < \varepsilon, \qquad \text{只要 } 0 < d\left(P, P_0\right) < \delta,$$

则我们称 $\lim\limits_{P \to P_0} f(P) = A$. 这里右侧的 d 表示 $\mathbf{R}^n$ 中两点之间的距离, 而左侧的 d 表示 $\mathbf{R}^m$ 中两点之间的距离. 以上不等式关系还可以写成集合的形式:

$$f\left(U_\delta\left(P_0\right) \setminus \{P_0\}\right) \subseteq U_\varepsilon(A).$$

这种十分抽象的表示式我们在定义 6.2.1 也使用过.

注意 本书提到 (多元) 函数一词, 默认都是实数值函数或实值函数, 即到达域是实数域 $\mathbf{R}$. 本书对于到达域是多元欧式空间 $\mathbf{R}^m$ 的函数, 习惯上叫做 "映射". (这些习惯每本书都不同)

尽管二元函数极限的定义与一元函数极限的定义非常类似, 但是我们要特别指出: 点 $P(x, y)$ 趋向于点 $P_0\left(x_0, y_0\right)$ 的方式很多 (无穷多):

1. 点 P 可以沿直线趋向于点 P_0,
2. 也可以沿折线趋向于点 P_0,
3. 甚至还可以沿任意一条曲线趋向于点 P_0.

我们称**当** $(x, y) \to (x_0, y_0)$ **时,** $f(x, y)$ **以** A **为极限是指: 不论点** (x, y) **沿什么路径趋向于点** (x_0, y_0), $f(x, y)$ **的值都是趋向于同一个常数** A. 所以, 二元函数的极限将比一元函数的极限显得复杂. 但这也从反面提供了证明二元函数极限不存在的一个方法: 只要能指出两条不同的路径, 当点 P 沿这两条路径趋向于点 P_0 时, $f(x, y)$ 趋向于不同的常数, 则我们就可以断言当 $P \to P_0$ 时, $f(x, y)$ 没有极限.

例 6.2.3 当 $(x, y) \to (0, 0)$ 时, 函数 $f(x, y) = \dfrac{|x|}{\sqrt{x^2 + y^2}}$ 是否有极限? 参考图像 二元函数极限不存在 | Desmos.

证明. 我们尝试让点 (x, y) 可以沿直线趋向于点 $(0, 0)$. 令 $y = kx$, 其中 k 是任意固定的常数. 显然, 当 $x \to 0$ 时, $y \to 0$. 换句话说, 我们考虑点 (x, y) 沿直线 $y = kx$ 趋向于点 $(0, 0)$. 在这种限制下我们可以简化函数, 有

$$f(x, y) = \frac{|x|}{\sqrt{x^2 + k^2 x^2}} = \frac{1}{\sqrt{1 + k^2}}.$$

这样一来, 我们看出, 当点 (x, y) 沿直线 $y = kx$ 趋向于点 $(0, 0)$ 时, $f(x, y)$ 变成了常函数, 所以它以 $\dfrac{1}{\sqrt{1 + k^2}}$ 为极限. 因此, 沿不同斜率的直线趋向于点 $(0, 0)$ 时, $f(x, y)$ 趋向于不同的常数. 由此推出, 当 $(x, y) \to (0, 0)$ 时, $f(x, y)$ 没有极限. 证毕. □

若当 (x,y) 沿任何直线 $y=kx$ 趋向于点 $(0,0)$ 时, $f(x,y)$ 都趋向于同一常数, 能否断定当 $(x,y)\to(0,0)$ 时, $f(x,y)$ 有极限? 显然不行. 见下题.

例 6.2.4 设函数

$$f(x,y)=\frac{x^4y^4}{(x^2+y^4)^3},$$

证明: 当动点 (x,y) 沿任何直线 $y=kx$ 趋向于点 $(0,0)$ 时, $f(x,y)$ 都趋向于零; 而当点 (x,y) 沿曲线 $y=\sqrt{x}$ 趋向于点 $(0,0)$ 时, $f(x,y)$ 趋向于 $\frac{1}{8}$. 由此推出, 当 $(x,y)\to(0,0)$ 时, $f(x,y)$ 没有极限.

证明. 1. 沿直线 $y=kx$ 时的极限. 假设点 (x,y) 沿直线 $y=kx$ 趋向于原点 $(0,0)$, 那么我们可以将 $y=kx$ 代入 $f(x,y)$ 中, 得到

$$f(x,y)=\frac{x^4(kx)^4}{(x^2+(kx)^4)^3}=\frac{k^4x^8}{(x^2+k^4x^4)^3}.$$

我们继续简化为

$$f(x,y)=\frac{k^4x^8}{(x^2(1+k^4x^2))^3}=\frac{k^4x^8}{x^6(1+k^4x^2)^3}=\frac{k^4x^2}{(1+k^4x^2)^3}.$$

注意到分母比分子更高阶的多项式. 当 $x\to0$ 时, 显然 $f(x,y)\to0$. 所以, 沿着任何直线 $y=kx$ 趋向原点时, $f(x,y)$ 的极限是 0.

2. 沿曲线 $y=\sqrt{x}$ 时的极限. 现在考虑点 (x,y) 沿曲线 $y=\sqrt{x}$ 趋向原点 $(0,0)$. 将 $y=\sqrt{x}$ 代入 $f(x,y)$ 中, 得到

$$f(x,y)=\frac{x^4(\sqrt{x})^4}{(x^2+(\sqrt{x})^4)^3}=\frac{x^4x^2}{(x^2+x^2)^3}=\frac{x^6}{(2x^2)^3}=\frac{x^6}{8x^6}=\frac{1}{8}.$$

因此, 当点 (x,y) 沿曲线 $y=\sqrt{x}$ 趋向原点时, $f(x,y)$ 的极限是 $\frac{1}{8}$.

□

6.2.2 二元函数极限的运算法则与基本性质

有关一元函数极限的主要定理, 如关于极限的四则运算, 不等式 (保序性) 的定理以及夹逼定理, 对二元函数 (任意多元函数) 情形都成立. 本节内容建议和第一章对比着看. 证明完全类似于一元函数的情况, 故从略.

定理 6.2.2: 多元函数极限的四则运算

设函数 $f(x,y)$ 及 $g(x,y)$ 在点 (x_0,y_0) 的一个空心邻域内有定义. 如果

$$\lim_{(x,y)\to(x_0,y_0)}f(x,y)=A,\quad \lim_{(x,y)\to(x_0,y_0)}g(x,y)=B,$$

则有:

1. $\lim\limits_{(x,y)\to(x_0,y_0)}(f(x,y)\pm g(x,y))=A\pm B$;
2. $\lim\limits_{(x,y)\to(x_0,y_0)}f(x,y)g(x,y)=AB$;
3. 当 $B\neq0$ 时, $\lim\limits_{(x,y)\to(x_0,y_0)}\dfrac{f(x,y)}{g(x,y)}=\dfrac{A}{B}$.

例 6.2.5 求极限 $\lim\limits_{(x,y)\to(1,2)} \dfrac{2x^2+y}{xy}$.

证明. 利用一元函数的极限容易知道

$$\lim_{(x,y)\to(1,2)} x=1, \quad \lim_{(x,y)\to(1,2)} y=2.$$

根据定理 6.2.2, 有

$$\lim_{(x,y)\to(1,2)} \frac{2x^2+y}{xy}=\frac{2\times 1^2+2}{1\times 2}=2.$$

□

上题告诉我们, 和一元情况一样, 通常能把函数值在该点处直接算出的, 其结果就是极限值. 这就是二元函数的连续性, 我们下一节讲.

定理 6.2.3: 多元函数极限的保序性

设函数 $f(x,y)$ 及 $g(x,y)$ 在点 (x_0,y_0) 的一个空心邻域内有定义, 且

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y)=A, \quad \lim_{(x,y)\to(x_0,y_0)} g(x,y)=B,$$

如果若在这个空心邻域内满足 $f(x,y)\geq g(x,y)$, 那么 $A\geq B$.

这个定理告诉我们: 像一元函数的情况一样, 较大函数的极限大于或等于较小函数的极限. 用定理 6.2.3 可以直接推出来下面的定理.

定理 6.2.4: 多元函数极限的夹逼定理

设三个函数 $f(x,y), g(x,y)$ 及 $h(x,y)$ 在点 (x_0,y_0) 的一个空心邻域内有定义, 并且在该空心邻域上满足不等式

$$h(x,y)\leq f(x,y)\leq g(x,y).$$

如果 $\lim\limits_{(x,y)\to(x_0,y_0)} g(x,y)=A$ 且 $\lim\limits_{(x,y)\to(x_0,y_0)} h(x,y)=A$, 那么 $\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y)=A$.

这个定理是一元函数的夹逼定理的推广. 一个显然的推论是: 如果 $\lim\limits_{(x,y)\to(x_0,y_0)} |f(x,y)|=0$, 那么 $\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y)=0$. 夹逼定理的另一推论是: 设两个非负函数 $f(x,y)$, $g(x,y)$ 在点 (x_0,y_0) 的一个空心邻域内有定义, 并且满足不等式

$$f(x,y)\leq g(x,y).$$

如果 $\lim\limits_{(x,y)\to(x_0,y_0)} g(x,y)=0$, 那么 $\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y)=0$.

例 6.2.6 求极限 $\lim\limits_{(x,y)\to(0,0)} \dfrac{\sin\left(x^3+y^4\right)}{x^2+y^2}$.

证明. 对于任意的 $(x,y)\neq(0,0)$, 有 (下面用到了不等式 $|\sin\theta|\leq|\theta|$)

$$0\leq\left|\frac{\sin\left(x^3+y^4\right)}{x^2+y^2}\right|\leq\frac{\left|x^3+y^4\right|}{x^2+y^2}\leq\frac{\left|x^3\right|}{x^2+y^2}+\frac{\left|y^4\right|}{x^2+y^2}=\frac{x^2}{x^2+y^2}|x|+\frac{y^2}{x^2+y^2}\left|y^2\right|\leq|x|+\left|y^2\right|.$$

而 $\lim\limits_{(x,y)\to(0,0)}|x|=0$, $\lim\limits_{(x,y)\to(0,0)}\left|y^2\right|=0$, 利用夹逼定理得到

$$\lim_{(x,y)\to(0,0)}\left|\frac{\sin\left(x^3+y^4\right)}{x^2+y^2}\right|=0.$$

完全类似于一元函数的情况, 由此可以推出

$$\lim_{(x,y)\to(0,0)}\frac{\sin\left(x^3+y^4\right)}{x^2+y^2}=0.$$

□

6.2.3 二元函数的复合函数的极限定理

一元函数之间的复合, 只需要考虑一种基本情况:

$$x\longrightarrow g(x)\longrightarrow f(g(x))$$

简写 $\mathbf{R}\to\mathbf{R}\to\mathbf{R}$. 多元函数的复合函数要比一元函数的复合函数更加复杂. 对于二元函数而言, 至少有如下三种复合情况:

1. **(情况 1)** f 是 (x,y) 的函数, 而 x 与 y 分别是自变量 t 的函数, 比如 $x=g(t),y=h(t)$, 这时复合函数 $f(g(t),h(t))$ 是一个一元函数.

$$t\longrightarrow(g(t),h(t))\longrightarrow f(g(t),h(t))$$

简写 $\mathbf{R}\to\mathbf{R}^2\to\mathbf{R}$.

2. **(情况 2)** f 是 (x,y) 的函数, 而 x 与 y 分别是自变量 (u,v) 的函数, 比如 $x=g(u,v),y=h(u,v)$, 这时复合函数 $f(g(u,v),h(u,v))$ 是一个二元函数[7].

$$(u,v)\longrightarrow(g(u,v),h(u,v))\longrightarrow f(g(u,v),h(u,v))$$

简写 $\mathbf{R}^2\to\mathbf{R}^2\to\mathbf{R}$.

3. **(情况 3)** f 是 u 的函数, 而 u 是 (x,y) 的函数, 比如 $u=g(x,y)$, 这时复合函数 $f(g(x,y))$ 仍是一个二元函数.

$$(x,y)\longrightarrow g(x,y)\longrightarrow f(g(x,y))$$

简写 $\mathbf{R}^2\to\mathbf{R}\to\mathbf{R}$.

可见, 二元函数复合的结果中, 自变量个数可能与中间变量个数不同.

当我们考虑一元函数时, 其复合函数的极限可以通过不断地嵌套各个环节的函数极限来求得. 这就是一元函数的复合函数的极限定理: 即, 如果

$$\lim_{x\to x_0}g(x)=u_0,\quad \lim_{u\to u_0}f(u)=L,$$

[7] 当然, x 与 y 也可能是 n $(n\geq3)$ 元函数, 这时复合函数便是一个 n 元函数.

那么
$$\lim_{x \to x_0} f(g(x)) = L.$$
也就是说, 如果

1. 当 $x \to x_0$ 时, $g(x) \to u_0$;
2. 当 $u \to u_0$ 时, $f(u) \to L$.

则当 $x \to x_0$ 时, $f(g(x)) \to L$. 这种写法更加清晰. 我们发现, 把最开始的条件 (当 $x \to x_0$ 时), 和最终的结果 ($f(u) \to L$) 放到一块, 这就是整个我们想要的结论了. 只不过, 还需要把 $f(u)$ 修改成关于 x 的复合形式 $f(g(x))$. 我们可以这么理解:

- 当 $x \to x_0$ 时 (自变量趋向), 则 $g(x) \to u_0$ (函数值趋向);
- 此时, 把 $g(x)$ 整体当作一个中间变量 u, 则有 $u = g(x) \to u_0$ (新的自变量趋向);
- 那么最后有 $f[g(x)] \to L$ (新的函数值趋向).

换句话说, **复合函数的极限可以视为 "逐层嵌套" 每个环节的极限过程: 上一个环节的函数值趋向, 变为下一个环节的自变量趋向, 然后导出新的函数值趋向.** 不断反复这一过程, 形成多米诺骨牌效应.

上述结果可以自然地扩张到多元函数的复合当中. 但是, 二元函数的复合函数的极限定理在叙述上要比一元函数的情形复杂些. 这里, 我们只叙述上述的 (情况 2) 和 (情况 3) 的复合函数的极限定理. 至于 (情况 1) 情况及其他情况, 读者可以举一反三得到相应的结论. 但请读者牢记于心: *二元函数的复合函数的极限定理本质上还是 "逐层嵌套" 这一套东西*. 上一个环节的函数值趋向, 变为下一个环节的自变量趋向, 然后导出新的函数值趋向, 不断反复.

现在, 我们考虑第 2 种复合情况 $\mathbf{R}^2 \to \mathbf{R}^2 \to \mathbf{R}$:
$$(u, v) \longrightarrow (g(u, v), h(u, v)) \longrightarrow f(g(u, v), h(u, v)),$$
则有如下定理.

定理 6.2.5: 多元复合函数的极限定理 (情况 2)

设 $x = g(u, v)$ 及 $y = h(u, v)$ 是定义在点 (u_0, v_0) 的一个空心邻域内的二元函数, 并且有极限
$$\lim_{(u,v)\to(u_0,v_0)} g(u, v) = x_0, \qquad \lim_{(u,v)\to(u_0,v_0)} h(u, v) = y_0.$$
又设 $f(x, y)$ 是定义在点 (x_0, y_0) 的一个空心邻域内的二元函数, 并且有极限
$$\lim_{(x,y)\to(x_0,y_0)} f(x, y) = A.$$
那么, 复合得到的二元函数 $z = f(g(u, v), h(u, v))$ 在点 (u_0, v_0) 的一个空心邻域内良好定义, 且有
$$\lim_{(u,v)\to(u_0,v_0)} f(g(u, v), h(u, v)) = A.$$

也就是说, 如果 (情况 2)

1. 当 $(u, v) \to (u_0, v_0)$ 时, $(g(u, v), h(u, v)) \to (x_0, y_0)$;

2. 当 $(x,y)\to(x_0,y_0)$ 时, $f(x,y)\to A$.

则 $(u,v)\to(u_0,v_0)$ 时, $f(g(u,v),h(u,v))\to A$.

现在, 我们考虑第 3 种复合情况 $\mathbf{R}^2\to\mathbf{R}\to\mathbf{R}$:

$$(x,y)\longrightarrow g(x,y)\longrightarrow f(g(x,y)),$$

则有如下定理.

定理 6.2.6: 多元复合函数的极限定理 (情况 3)

设 $u=g(x,y)$ 是定义在点 (x_0,y_0) 的一个空心邻域内的二元函数, 并且有极限

$$\lim_{(x,y)\to(x_0,y_0)} g(x,y)=u_0.$$

又设 $z=f(u)$ 是定义在点 u_0 的一个空心邻域内的一元函数, 并且有极限

$$\lim_{u\to u_0} f(u)=A.$$

那么, 复合得到的二元函数 $z=f(g(x,y))$ 在点 (x_0,y_0) 的一个空心邻域内良好定义, 且有

$$\lim_{(x,y)\to(x_0,y_0)} f(g(x,y))=A.$$

也就是说, 如果

1. 当 $(x,y)\to(x_0,y_0)$ 时, $g(x,y)\to u_0$;

2. 当 $u\to u_0$ 时, $f(u)\to A$.

则当 $(x,y)\to(x_0,y_0)$ 时, $f(g(x,y))\to A$.

例 6.2.7 证明 $\lim\limits_{(x,y)\to(0,0)}\left(1+x^2+y^2\right)^{\frac{1}{x^2+y^2}}=\mathrm{e}$.

证明. 令 $u=x^2+y^2$, 则

$$\left(1+x^2+y^2\right)^{\frac{1}{x^2+y^2}}=(1+u)^{\frac{1}{u}}.$$

于是, 由 $\lim\limits_{(x,y)\to(0,0)} x^2+y^2=0$, $\lim\limits_{u\to0}(1+u)^{\frac{1}{u}}=\mathrm{e}$ 以及由复合函数的极限定理, 立即推出

$$\lim_{(x,y)\to(0,0)}\left(1+x^2+y^2\right)^{\frac{1}{x^2+y^2}}=\mathrm{e}$$

证毕. □

例 6.2.8 求极限 $\lim\limits_{(u,v)\to(0,0)}\left(\dfrac{\sin 2\left(u^2+v^2\right)}{u^2+v^2}\right)^{\frac{u\sin v}{\sqrt{u^2+v^2}}}$.

证明. 令

$$x(u,v)=\frac{\sin 2\left(u^2+v^2\right)}{u^2+v^2},\quad y(u,v)=\frac{u\sin v}{\sqrt{u^2+v^2}},$$

先求极限

$$\lim_{(u,v)\to(0,0)} x(u,v)=\lim_{(u,v)\to(0,0)}\frac{\sin 2\left(u^2+v^2\right)}{u^2+v^2}.$$

当 $(u,v)\to(0,0)$ 时, $u^2+v^2\to 0$. 将 $t=u^2+v^2\to 0$ 作为整体考虑, 则

$$\frac{\sin 2\left(u^2+v^2\right)}{u^2+v^2}=\frac{\sin 2t}{t}=2\frac{\sin 2t}{2t}\to 2.$$

于是, $\lim\limits_{(u,v)\to(0,0)} x(u,v)=2$. 另外, 由例 6.2.1 可知

$$\lim_{(u,v)\to(0,0)} y(u,v)=0.$$

最后, 由复合函数的极限定理我们得到

$$\lim_{(u,v)\to(0,0)}\left(\frac{\sin 2\left(u^2+v^2\right)}{u^2+v^2}\right)^{\frac{u\sin v}{\sqrt{u^2+v^2}}}=\lim_{(x,y)\to(2,0)} x^y=2^0=1.$$

证毕. □

6.2.4 累次极限与全面极限

这一段内容是关于二元函数极限概念的一个附注. 随便读一读即可.

一个二元函数 $f(x,y)$ 当其中的 y 任意固定时就成为一个一元函数 (自变量 x 的函数), 这时自然可以对 x 取极限. 我们假定对于任意固定的 y, 下列极限存在:

$$\lim_{x\to x_0} f(x,y)=: g(y).$$

(这个极限一般来说依赖于被固定的 y, 故将它记为 $g(y)$). 如果当 $y\to y_0$ 时, $g(y)$ 也有极限, 那么我们称

$$\lim_{y\to y_0} g(y)=\lim_{y\to y_0}\lim_{x\to x_0} f(x,y)$$

是 $f(x,y)$ 的**累次极限**.

例 6.2.9 比如函数 $f(x,y)=x\sin(x+y)$, 这时

$$\lim_{x\to\pi} f(x,y)=\pi\sin(\pi+y),$$

并且

$$\lim_{y\to 1}\lim_{x\to\pi} f(x,y)=\lim_{y\to 1}\pi\sin(\pi+y)=\pi\sin(\pi+1).$$

与上面完全类似, 交换 x,y 求极限时的顺序, 就可以考虑另一种累次极限

$$\lim_{x\to x_0}\lim_{y\to y_0} f(x,y).$$

这样, 对二元函数而言可以有两种累次极限. 为了与累次极限相区别, 我们将前面定义的当 $(x,y)\to(x_0,y_0)$ 时的函数极限称为**全面极限**.

例 6.2.10 函数

$$f(x,y)=\frac{xy}{x^2+y^2}$$

的两个累次极限 $\lim\limits_{y\to0}\lim\limits_{x\to0}f(x,y)$ 及 $\lim\limits_{x\to0}\lim\limits_{y\to0}f(x,y)$ 显然都是零. 但是, 当 $(x,y)\to(0,0)$ 时, 上述函数的全面极限并不存在 (考虑沿不同斜率的直线趋向于原点即可).

一般来说, 累次极限与全面极限之间没有什么必然联系: 全面极限存在并不意味着累次极限存在; 反过来, 两个累次极限都存在且相等也不能保证全面极限存在.

例 6.2.11 例如, 对于函数

$$f(x,y)=\begin{cases}(x+y)\sin\dfrac{1}{x}, & x\neq0,\\ 0, & x=0,\end{cases}$$

当 $(x,y)\to(0,0)$ 时, 其全面极限为零, 但它的累次极限 $\lim\limits_{y\to0}\lim\limits_{x\to0}f(x,y)$ 并不存在. 另外, 函数

$$f(x,y)=\frac{|xy|^{\frac{1}{2}}}{\sqrt{x^2+y^2}}$$

的两个累次极限 $\lim\limits_{x\to0}\lim\limits_{y\to0}f(x,y)$ 与 $\lim\limits_{y\to0}\lim\limits_{x\to0}f(x,y)$ 都存在且是零. 但是, 当 $(x,y)\to(0,0)$ 时, 该函数的全面极限并不存在 (证明方法与例 6.2.2 相同).

总之, 累次极限与全面极限是两个不同的概念. 在求全面极限时不可用累次极限来替代, 那样做可能导致错误结果.

6.3 多元函数的连续性

6.3.1 二元函数连续性的定义

我们只给出二元函数连续性的定义, 至于一般多元函数连续性的定义完全与二元函数的情形类似.

定义 6.3.1: 二元函数的连续性

设函数 $f(x,y)$ 在点 (x_0,y_0) 的一个邻域内有定义. 若当 $(x,y)\to(x_0,y_0)$ 时, $f(x,y)$ 有极限, 并且其极限等于函数值 $f(x_0,y_0)$, 即

$$\lim_{(x,y)\to(x_0,y_0)}f(x,y)=f(x_0,y_0),$$

则称 $f(x,y)$ 在点 (x_0,y_0) 处连续. 如果函数 $f(x,y)$ 在区域 D 内有定义且在 D 内每一点处都连续, 则称 $f(x,y)$ 在区域 D 内连续.

函数 $f(x,y)$ 在点 (x_0,y_0) 处的连续性也可以用 $\varepsilon-\delta$ 说法 (就是直接套用极限的定义) 严格叙述: 若对于任意给定的 $\varepsilon>0$, 都存在一个 $\delta>0$, 使得

$$|f(x,y)-f(x_0,y_0)|<\varepsilon,\qquad \text{只要 } 0<d((x,y),(x_0,y_0))<\delta,$$

则称 $f(x,y)$ 在点 (x_0,y_0) 处连续. 显然, 像在前一节讨论极限时那样, 上述定义中的距离不等式条件可以写成:

- $0 < \sqrt{(x-x_0)^2+(y-y_0)^2} < \delta$.
- $0 < |x-x_0| < \delta$, $0 < |y-y_0| < \delta$.

它们所定义的连续性是彼此等价的.

例 6.3.1 证明: 函数 $f(x,y)=\sin(x+y)+|x+y+1|$ 在 $\mathbf{R}^2$ 中任意一点处连续.

证明. 设 (x_0,y_0) 是 $\mathbf{R}^2$ 中任意给定的一点. 那么, 对于任意一点 $(x,y)\in\mathbf{R}^2$, 利用和差化积公式 $\sin\alpha-\sin\beta=2\cos\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}$, 我们有 (函数的前半部分)

$$\begin{aligned}
|\sin(x+y)-\sin(x_0+y_0)| &= 2\left|\sin\frac{(x+y)-(x_0+y_0)}{2}\cos\frac{(x+y)+(x_0+y_0)}{2}\right| \\
&= 2\left|\sin\frac{(x+y)-(x_0+y_0)}{2}\right|\left|\cos\frac{(x+y)+(x_0+y_0)}{2}\right| \\
&\le 2\left|\sin\frac{(x+y)-(x_0+y_0)}{2}\right| \\
&\le 2\left|\frac{(x+y)-(x_0+y_0)}{2}\right| \\
&= |(x-x_0)+(y-y_0)| \\
&\le |x-x_0|+|y-y_0|.
\end{aligned}$$

另一方面, 利用不等式 $||a|-|b||\le|a-b|$[8], 我们有 (函数的后半部分)

$$\begin{aligned}
||x+y+1|-|x_0+y_0+1|| &\le |(x+y+1)-(x_0+y_0+1)| \\
&\le |x-x_0|+|y-y_0|.
\end{aligned}$$

因此综合上面两个结果, 最后有

$$|f(x,y)-f(x_0,y_0)|\le 2|x-x_0|+2|y-y_0|.$$

这个放大的不等式说明我们应该使用切比雪夫距离, 而不是欧几里得距离. 所以, 对于任意给定的 $\varepsilon>0$, 取 $\delta=\frac{\varepsilon}{4}$, 则当

$$|x-x_0|<\delta,\quad |y-y_0|<\delta$$

时, 便有

$$|f(x,y)-f(x_0,y_0)|<\varepsilon.$$

根据函数 $f(x,y)$ 在点 (x_0,y_0) 处连续的等价定义, 可知 $f(x,y)$ 在 $\mathbf{R}^2$ 中任意一点 (x_0,y_0) 处连续. 证毕. □

例 6.3.2 设函数

$$f(x,y)=\begin{cases}\dfrac{x^2y^2}{x^4+y^4}, & (x,y)\neq(0,0),\\ 0, & (x,y)=(0,0)\end{cases}$$

问: $f(x,y)$ 在何处连续?

[8] 本学期第一节课我们有讲过命题 1.1.5: 对于任意实数 x,y, 我们有 $||x|-|y||\leqslant|x-y|$.

证明. 1. 设 (x_0, y_0) 是任意给定的一点. 当 $(x_0, y_0) \neq (0,0)$ 时, $x_0^4 + y_0^4 \neq 0$, 并很容易直接证明

$$\lim_{(x,y)\to(x_0,y_0)} \left(x^4 + y^4\right) = x_0^4 + y_0^4,$$
$$\lim_{(x,y)\to(x_0,y_0)} x^2y^2 = x_0^2y_0^2.$$

应用极限的四则运算法则, 我们有

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = \frac{x_0^2y_0^2}{x_0^4+y_0^4} = f\left(x_0, y_0\right).$$

因此, $f(x,y)$ 在点 (x_0, y_0) 处连续.

2. 当 $(x_0, y_0) = (0,0)$ 时, 可以按照类似于上一节例题的方法证明 $f(x,y)$ 的极限不存在. 事实上, 当点 (x,y) 沿直线 $y = kx$ (k 为常数) 趋向于点 $(0,0)$ 时, 在这种限制下我们可以简化函数, 有

$$f(x,y) = \frac{x^2(kx)^2}{x^4+(kx)^4} = \frac{k^2}{1+k^4}.$$

当 k 取不同值时, 上述极限的值也不同. 可见, $f(x,y)$ 在点 $(0,0)$ 处不连续.

□

6.3.2 关于二元函数连续性的几个定理

我们知道, 对于一元函数, 四则运算及复合运算均保持函数的连续性. 在多元函数中, 这些结论依然成立. 下面两个定理完全可以由上一节中相应的定理推出.

定理 6.3.1: 连续函数的和, 差, 积, 商的连续性

设两个二元函数 $f(x,y)$ 及 $g(x,y)$ 在一点 (x_0, y_0) 处连续, 则函数 $f(x,y) \pm g(x,y)$ 及 $f(x,y)g(x,y)$ 在点 (x_0, y_0) 处也连续. 此外, 若 $g\left(x_0, y_0\right) \neq 0$, 则函数 $\dfrac{f(x,y)}{g(x,y)}$ 在点 (x_0, y_0) 处连续.

下面讨论复合函数的连续性. 我们只以上一节的 (情况 3) 为例: 考虑 f 是 u 的函数, 而 u 是 (x,y) 的函数, 比如 $u = g(x,y)$, 这时复合函数 $f(g(x,y))$ 仍是一个二元函数.

$$(x,y) \longrightarrow g(x,y) \longrightarrow f(g(x,y))$$

简写 $\mathbf{R}^2 \to \mathbf{R} \to \mathbf{R}$. 下面的定理给出了 (情况 3) 时复合函数的连续性. 只需要记住, *通常而言, 任意的连续函数相复合, 无论是哪一类复合情况 (比如上一节的 (情况 1), (情况 2) 和 (情况 3), 又或者别的), 最终的复合函数也保持连续性.*

定理 6.3.2: 复合函数的连续性 (情况 3: $\mathbf{R}^2 \to \mathbf{R} \to \mathbf{R}$)

设函数 $u = g(x,y)$ 在点 (x_0, y_0) 附近有定义且在点 (x_0, y_0) 处连续, 又设函数 $z = f(u)$ 在点 $u_0 = g\left(x_0, y_0\right)$ 附近有定义且在点 u_0 处连续, 则复合函数 $z = f(g(x,y))$ 在点 (x_0, y_0) 处连续.

有了上面这两个定理, 在讨论函数的连续性时就方便多了, 尤其是对于常见的初等函数而言, 就可免去直接从定义证明之苦. 我们说一个函数 $f(x,y)$ 是**二元初等函数**, *如果它是从自变量 x 与 y 出发进行有限次四则*

运算或者复合以一元初等函数的结果. 类似地, 可定义三元及三元以上初等函数. 例如, 在前面的例题中二元函数

$$f(x,y)=\sin(x+y)+|x+y+1|$$

可以看成 $t=x+y$ 复合以 $\sin t$, 再加上 $|x+y+1|$ 的结果, 因而是 x 与 y 的二元初等函数. 根据一元初等函数的连续性以及定理 6.3.1 与定理 6.3.2, 我们有下面的定理.

定理 6.3.3

任意的二元初等函数 $f(x,y)$ 在其有定义的区域内是连续的.

根据以上定理, 今后我们在讨论二元初等函数 $f(x,y)$ 的极限时, 只要点 (x_0,y_0) 在 $f(x,y)$ 的有定义的区域内, 那么当 $(x,y)\to(x_0,y_0)$ 时, $f(x,y)$ 的极限值就是 $f(x_0,y_0)$, 比如

$$\lim_{(x,y)\to(0,1)}\frac{\cos\left(x^2+y^2\right)\arctan\left(x^3+y^3\right)}{\left(x^2+y^2\right)^{\frac{3}{2}}}=\cos 1\cdot\arctan 1=\frac{\pi}{4}\cos 1.$$

6.3.3　映射的连续性

注意　本书提到 (多元) 函数一词, 默认都是实数值函数或实值函数, 即到达域是实数域 $\mathbf{R}$. 本书对于到达域是多元欧式空间 $\mathbf{R}^m$ 的函数, 习惯上叫做 “映射”.

上一节课我们提到过如何定义任意形如 $f:\mathbf{R}^n\to\mathbf{R}^m$ 的映射的极限. 给定区域 $D\subseteq\mathbf{R}^n$, 设映射 $f:D\to\mathbf{R}^m$, 又设 $P_0\in D$. 若存在一个常向量 $A\in\mathbf{R}^m$, 对于任意给定的 $\varepsilon>0$, 都存在一个 $\delta=\delta(\varepsilon)>0$, 使得

$$d(f(P),A)<\varepsilon,\qquad 只要\ 0<d(P,P_0)<\delta,$$

则我们称 $\lim\limits_{P\to P_0}f(P)=A$. 以上不等式关系还可以写成:

$$f\left(U_\delta\left(P_0\right)\setminus\{P_0\}\right)\subseteq U_\varepsilon(A).$$

容易想象, 对于这一类映射的连续性, 和之前完全一样, 就是当

$$\lim_{P\to P_0}f(P)=f(P_0)$$

时, 我们称 f 在点 P_0 处连续. 因此, 我们可以给出如下任意形如 $f:\mathbf{R}^n\to\mathbf{R}^m$ 的映射的连续性概念. 它包含了定义 6.3.1 这一特殊情况.

定义 6.3.2: 映射的连续性

给定区域 $D\subseteq\mathbf{R}^n$, 设映射 $f:D\to\mathbf{R}^m$, 又设 $P_0\in D$. 若当 $P\to P_0$ 时, $f(P)\in\mathbf{R}^m$ 有极限, 并且其极限等于常向量 $f(P_0)\in\mathbf{R}^m$, 即

$$\lim_{P\to P_0}f(P)=f(P_0),$$

则我们称 f 在点 P_0 处连续. 如果 f 在区域 D 中每一点处都连续, 则称 f 在 D 内连续.

映射 $f:D\to\mathbf{R}^m$ 在点 P_0 处连续, 是指对于任意给定的 $\varepsilon>0$, 都存在一个 $\delta>0$, 使得当 $d(P,P_0)<\delta$ 时, 就有

$$d(f(P),f(P_0))<\varepsilon,$$

如果用集合符号, 以上还可以写成: 如果对于任意给定的 $\varepsilon>0$, 都存在一个 $\delta>0$, 使得

$$f\left(U_\delta(P_0)\right)\subset U_\varepsilon\left(f(P_0)\right).$$

当 $m=1$ 时, 上述映射就是一个函数, 与已经定义过的多元函数连续性 (定义 6.3.1) 完全一致. 但当 $m>1$ 时, 上述映射 $f:D\to\mathbf{R}^m$ 就要用 m 个函数来表示才成. 设 f_j $(j=1,2,\cdots,m)$ 是映射 f 的第 j 个分量, 即

$$f(P)=(f_1(P),f_2(P),\ldots,f_m(P))\in\mathbf{R}^m.$$

上述定义中的条件即可写成

$$\sqrt{\sum_{j=1}^m\left(f_j(P)-f_j(P_0)\right)^2}<\varepsilon,\quad \text{只要 } d\left(P,P_0\right)<\delta.$$

映射 $f:D\to\mathbf{R}^m$ 的连续性也可以由 f 的每个分量 f_j 的连续性来刻画.

1. 若映射 $f:D\to\mathbf{R}^m$ 在点 $P_0\in D$ 处连续, 则对于任意给定的 $\varepsilon>0$, 都存在一个 $\delta>0$, 使得当 $d\left(P,P_0\right)<\delta$ 时, 就有

$$|f_j(P)-f_j(P_0)|<\sqrt{\sum_{j=1}^m\left(f_j(P)-f_j(P_0)\right)^2}=d\left(f(P),f(P_0)\right)<\varepsilon,\quad j=1,2,\cdots,m.$$

这立马说明了每个分量 f_j 在点 P_0 的连续性.

2. 反之, 如果映射 $f:D\to\mathbf{R}^m$ 的每个分量函数 f_j 在点 $P_0\in D$ 处连续, 那么映射 f 在点 P_0 处也是连续的. 证明: 假设每个分量函数 f_j 在点 P_0 处连续. 根据定义, 则对于任意给定的 $\varepsilon>0$, 对于每个 $j=1,2,\ldots,m$, 存在对应的 $\delta_j>0$, 使得当 $d(P,P_0)<\delta_j$ 时, 有

$$|f_j(P)-f_j(P_0)|<\frac{\varepsilon}{\sqrt{m}}.$$

这样, 当 $d(P,P_0)<\delta:=\min\{\delta_1,\delta_2,\ldots,\delta_m\}$ 时, 对于所有 $j=1,2,\ldots,m$, 上述不等式也都成立. 接下来, 可以得到

$$d\left(f(P),f(P_0)\right)=\sqrt{\sum_{j=1}^m\left(f_j(P)-f_j(P_0)\right)^2}<\sqrt{\sum_{j=1}^m\left(\frac{\varepsilon}{\sqrt{m}}\right)^2}=\sqrt{m\cdot\frac{\varepsilon^2}{m}}=\varepsilon.$$

因此, 当 $d(P,P_0)<\delta$ 时, 有 $d\left(f(P),f(P_0)\right)<\varepsilon$, 这正是映射 f 在点 P_0 处连续的定义.

最终我们有以下命题.

命题 6.3.1: 映射连续等价于每一个分量函数连续

映射 $f:D\to\mathbf{R}^m$ 在点 $P_0\in D$ 处连续的充要条件是 f 的每个分量 f_j 在点 P_0 处都连续.

例 6.3.3 考虑映射

$$f:\mathbf{R}^2\to\mathbf{R}^2,$$
$$(x,y)\mapsto(u,v),$$

其中 $u=\sin x+\cos y$, $v=\mathrm{e}^x\sin y$. 这里 f 的第一个分量就是 $u=f_1(x,y)=\sin x+\cos y$, 而第二个分量是 $v=f_2(x,y)=\mathrm{e}^x\sin y$. 这两个分量都是 (x,y) 的连续函数, 因而映射 f 在 $\mathbf{R}^2$ 上是连续的.

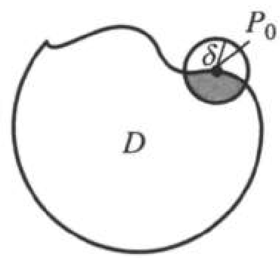


图 6.5: 边界点和闭区域上的连续性.

6.3.4 有界闭区域上连续 (实值) 函数的定理

在第一章中, 我们指出了在有界闭区间上连续的一元函数有一些重要的性质 (介值定理, 最值定理, 有界性定理). 我们说过, 闭区域是闭区间在高维空间中的对应物. 现在我们指出: 对于在有界闭区域上连续的多元函数, 这些重要的性质同样成立.

设 D 是 $\mathbf{R}^n$ (特别是 $n \geq 2$) 中的一个区域, 则 $\bar{D}$ 是一个闭区域. 假定函数 f 在 $\bar{D}$ 上有定义, 怎样才称 f 在闭区域 $\bar{D}$ 上连续呢? 我们知道闭区域里只有两类点:

1. 一类是内点, 这些内点的集合就区域 D. 这些点都存在一个完整的小邻域, 使得函数 f 在该小邻域上处处有定义. 那么就可以用前面的连续性定义了.

2. 但还有一类是区域 D 的边界点 (见图 6.5). 设 P_0 是 D 的一个边界点, 这时点 P_0 的任意一个邻域 (无论半径有多小) 都有 D 中的点, 也有非 D 中的点. 因此, f 在点 P_0 的任何一个小邻域中都不全是有定义的. 考虑到这一点, 现有的连续性定义方式就不能适用了. 我们需要对原来连续性的定义做一个的小小修改: 只需要考虑这个小邻域中有 f 的定义的那一部分就足够了.

多元函数 f 在边界点处的连续性定义如下所示.

定义 6.3.3: 边界点和闭区域上的连续性

设 D 是 $\mathbf{R}^n$ 中的一个区域, 则 $\bar{D}$ 是一个闭区域. 我们称 f 在 $\bar{D}$ 的*边界点* P_0 处连续是指, 如果对于任意给定的 $\varepsilon > 0$, 都存在一个 $\delta > 0$, 使得当 $P \in U_\delta(P_0) \cap \bar{D}$ 时, 便有

$$|f(P) - f(P_0)| < \varepsilon,$$

即

$$f\left(U_\delta(P_0) \cap \bar{D}\right) \subset U_\varepsilon\left(f(P_0)\right).$$

我们称函数 $f: \bar{D} \to \mathbf{R}$ 在整个闭区域 $\bar{D}$ 上连续, 如果 f 在 $\bar{D}$ 上每一点处都连续.

换句话说, 这要求当 P 落在集合 $U_\delta(P_0) \cap \bar{D}$ 中时, $f(P)$ 与 $f(P_0)$ 之差的绝对值小于 ε. 图 6.5 中阴影部分表示了集合 $U_\delta(P_0) \cap \bar{D}$. 有了多元函数在闭区域上连续的定义后, 就可以叙述我们的定理了.

定理 6.3.4: 介值定理

如果多元函数 $f: \bar{D} \subseteq \mathbf{R}^n \to \mathbf{R}$ 在有界闭区域 $\bar{D}$ 上连续, 并假定 M 与 m 分别是 f 在 $\bar{D}$ 上的最大值与最小值, 则对于任意的 η $(m \leq \eta \leq M)$, 存在一点 $P_0 \in \bar{D}$, 使得 $f(P_0) = \eta$.

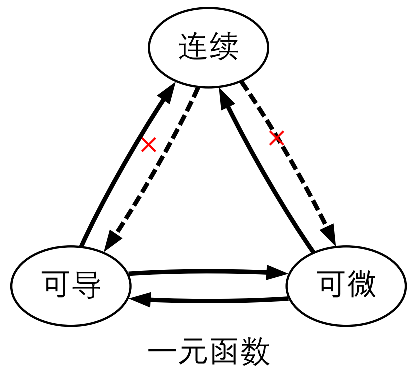


图 6.6: 一元函数连续, 导数, 微分的关系图.

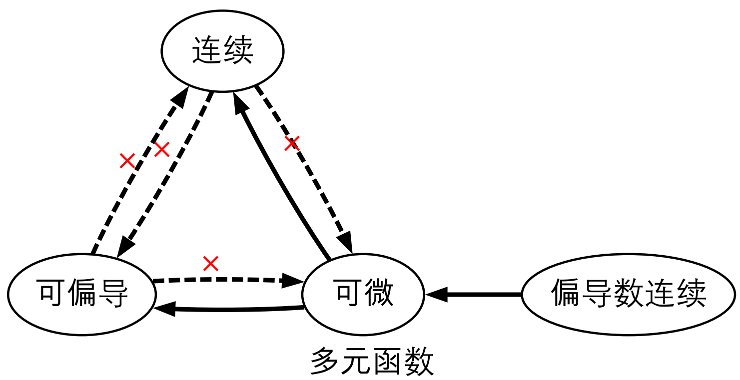


图 6.7: 多元函数连续, 偏导, (全) 微分的关系图.

定理 6.3.5: 最大值与最小值定理

如果多元函数 $f:\bar{D}\subseteq\mathbf{R}^n\to\mathbf{R}$ 在有界闭区域 $\bar{D}$ 上连续, 则 f 在 $\bar{D}$ 上能够取得到最大值与最小值, 即存在点 $P_1,P_2\in\bar{D}$, 使得

$$f\left(P_1\right)\leq f(P)\leq f\left(P_2\right),\quad \forall P\in\bar{D}.$$

定理 6.3.6: 有界性定理

如果多元函数 $f:\bar{D}\subseteq\mathbf{R}^n\to\mathbf{R}$ 在有界闭区域 $\bar{D}$ 上连续, 则 f 在 $\bar{D}$ 上有界, 即存在常数 $M>0$, 使得

$$|f(P)|\leq M,\quad \forall P\in\bar{D}.$$

6.4 偏导数与全微分

本节总结如图 6.7. 对于图 6.7 中多元函数的结果, 其根据如下:

- 可微 $\Rightarrow$ 连续, 见定理 6.4.2.
- 可微 $\Rightarrow$ 可偏导, 见定理 6.4.3.

- 偏导数连续 $\Rightarrow$ 可微, 见定理 6.4.4.

- 可偏导 $\not\Rightarrow$ 可微, 见反例 6.4.15.

- 可偏导 $\not\Rightarrow$ 连续, 见反例 6.4.8, 反例 6.4.15.

- 连续 $\not\Rightarrow$ 可偏导, 见反例 6.4.9.

- 连续 $\not\Rightarrow$ 可微, 见反例 6.4.9.

6.4.1 偏导数

考虑一个二元函数 $z=f(x,y)$.

- 当我们将自变量 y 固定时, $z=f(x,y)$ 就是关于 x 的一个一元函数, 即

  $$x\mapsto f(x,y),\ \text{其中 } y \text{ 是固定的, 视作常数.}$$

  这时, 我们自然可以考虑它关于 x 的导数. 这样求得的对 x 的导数称作 $z=f(x,y)$ 关于 x 的偏导数.

- 类似地, 当我们将自变量 x 固定时, $z=f(x,y)$ 就是关于 y 的一个一元函数, 即

  $$y\mapsto f(x,y),\ \text{其中 } x \text{ 是固定的, 视作常数.}$$

  这时, 我们同样可以考虑它关于 y 的导数. 这样求得的对 y 的导数称作 $z=f(x,y)$ 关于 y 的偏导数.

我们知道导数是由因变量的增量与自变量的增量之比的极限定义的: 一个普通的一元函数 $f(u)$ 在 $u=u_0$ 处的导数被定义为

$$f'(u_0)=\frac{\mathrm{d}}{\mathrm{d}u}f(u)\bigg|_{u=u_0}=\lim_{\Delta u\to 0}\frac{f(u_0+\Delta u)-f(u_0)}{\Delta u}.$$

因此, 刚才讨论的偏导数的正式定义就可以讨论上面的极限给出, 如下所示.

定义 6.4.1: 偏导数

设二元函数 $z=f(x,y)$ 在点 (x_0,y_0) 的某个邻域内有定义.

1. 将 y 固定为 y_0. 考虑一元函数 $x\mapsto f(x,y_0)$, 如果它在 $x=x_0$ 处的导数, 即

$$\left.\frac{\mathrm{d}}{\mathrm{d}x}f(x,y_0)\right|_{x=x_0}=\lim_{\Delta x\to 0}\frac{f\left(x_0+\Delta x,y_0\right)-f\left(x_0,y_0\right)}{\Delta x}$$

存在, 则称该极限值为 $z=f(x,y)$ 在点 (x_0,y_0) 处关于 x 的 (一阶) **偏导数 (partial derivative)**, 也叫偏微商, 记作

$$f_x\left(x_0,y_0\right),\quad \frac{\partial f\left(x_0,y_0\right)}{\partial x},\quad \frac{\partial f}{\partial x}\left(x_0,y_0\right),\quad \left.\frac{\partial z}{\partial x}\right|_{\left(x_0,y_0\right)}\quad 或\quad \left.z_x\right|_{\left(x_0,y_0\right)}.$$

符号 ∂ 中文习惯读作 “偏”, 英文习惯读作 “partial”.

2. 类似地, 将 x 固定为 x_0. 考虑一元函数 $y\mapsto f(x_0,y)$, 如果它在 $y=y_0$ 处的导数, 即

$$\left.\frac{\mathrm{d}}{\mathrm{d}y}f(x_0,y)\right|_{y=y_0}=\lim_{\Delta y\to 0}\frac{f\left(x_0,y_0+\Delta y\right)-f\left(x_0,y_0\right)}{\Delta y}$$

存在, 则称它为 $z=f(x,y)$ 在点 (x_0,y_0) 处关于 y 的 *(一阶)* 偏导数, 记作

$$f_y\left(x_0,y_0\right),\quad \frac{\partial f\left(x_0,y_0\right)}{\partial y},\quad \frac{\partial f}{\partial y}\left(x_0,y_0\right),\quad \left.\frac{\partial z}{\partial y}\right|_{\left(x_0,y_0\right)}\quad 或\quad \left.z_y\right|_{\left(x_0,y_0\right)}.$$

这两个值 $f_x\left(x_0,y_0\right)$ 和 $f_y\left(x_0,y_0\right)$ 统称为 $z=f(x,y)$ 在点 (x_0,y_0) 处的偏导数, 具体以 “关于变量 x 或者 y 的偏导数” 来区分.

从上述定义可以看出: 偏导数 $f_x\left(x_0,y_0\right)$ 就是 $f\left(x,y_0\right)$ 作为 x 的一元函数在点 x_0 处的导数, 而偏导数 $f_y\left(x_0,y_0\right)$ 是 $f\left(x_0,y\right)$ 作为 y 的一元函数在点 y_0 处的导数.

例 6.4.1 计算 $f(x,y)=x^2y^3$ 在点 $(1,2)$ 处的偏导数.

证明. 我们需要分别计算在点 $(1,2)$ 处关于 x 和 y 的偏导数. 首先, 将 y 固定为 2, 即考虑函数 $f(x,2)=x^2 2^3=8x^2$. 则

$$\frac{\partial f}{\partial x}(1,2)=\left.\frac{\mathrm{d}}{\mathrm{d}x}f(x,2)\right|_{x=1}=\left.16x\right|_{x=1}=16.$$

接下来, 将 x 固定为 1, 即考虑函数 $f(1,y)=y^3$. 则

$$\frac{\partial f}{\partial y}(1,2)=\left.\frac{\mathrm{d}}{\mathrm{d}y}f(1,y)\right|_{y=2}=\left.3y^2\right|_{y=2}=12.$$

□

在定义 6.4.1 中, 我们使用了 x_0 或 y_0 表示 x 或 y 的某个固定值, 使它们区别于变动中的 x 或 y. 这在定义偏导数的过程中是必不可少的. 但在建立了偏导数的概念之后, 若计算函数 $z=f(x,y)$ 在任意点 (x,y) 的偏导数

$$f_x=f_x(x,y)\quad 或写作\ \frac{\partial f}{\partial x}=\frac{\partial f}{\partial x}(x,y)$$

只需我们在心中记住, 应将 $f(x,y)$ 的表达式中的 y 看作常数; 而在计算

$$f_y = f_y(x,y) \ \text{ 或写作 } \frac{\partial f}{\partial y} = \frac{\partial f}{\partial y}(x,y)$$

时, 应在心中将 $f(x,y)$ 的表达式中的 x 看作常数. 如下题所示.

例 6.4.2 设函数 $f(x,y) = x^2 + y^2 + xy + \mathrm{e}^x \cos y$, 求偏导数 $\dfrac{\partial f}{\partial x}$ 及 $\dfrac{\partial f}{\partial y}$.

证明. 根据前面的说明, 求 $\dfrac{\partial f}{\partial x}$ 时应将 $f(x,y)$ 的表达式中的 y 视作常数. 那么, 我们有

$$\frac{\partial f}{\partial x}(x,y) = 2x + y + \mathrm{e}^x \cos y.$$

求 $\dfrac{\partial f}{\partial y}$ 时, 应将 $f(x,y)$ 的表达式中的 x 视作常数, 可得到

$$\frac{\partial f}{\partial y}(x,y) = 2y + x - \mathrm{e}^x \sin y.$$

□

由前面的例子可以看出: 我们在计算一个二元初等函数的偏导数时, 有一元函数的导数公式就足够了, 无须添加新公式, 关键是要弄清楚将哪一个变量视作常数, 哪一个变量视作变量. 总之, 计算偏导数, 就是计算一个带着其他字母常数的一元函数导数. 随着求偏导时我们关心的变量改变, 所谓 "变量" 和 "常数" 的角色会不断地调整切换.

例 6.4.3 设函数 $f(x,y) = x^y \ (x,y>0)$, 求 $\dfrac{\partial f}{\partial x}$ 与 $\dfrac{\partial f}{\partial y}$.

证明. 当我们把 y 视作常数时, x^y 是 x 的幂函数, 因此得到

$$\frac{\partial f}{\partial x}(x,y) = yx^{y-1}.$$

当我们把 x 看作常数时, x^y 是 y 的指数函数, 因此有

$$\frac{\partial f}{\partial y}(x,y) = x^y \ln x.$$

□

导数也叫做导函数, 这是因为当我们有了导数 $f'(x)$ 的表达式后, 就可以计算在任意点处的导数值, 故导数可以视作一个一元函数 $x \mapsto f'(x)$. 类似地, 在上面的题目中, 我们得到了二元函数 $f(x,y)$ 在任意点 (x,y) 处的两个偏导数的表达式, 并且可以将它们再次视作两个新的二元函数 (以例 6.4.2 的结果为例):

$$\frac{\partial f}{\partial x} : \mathbf{R}^2 \to \mathbf{R}, \quad (x,y) \mapsto 2x + y + \mathrm{e}^x \cos y,$$
$$\frac{\partial f}{\partial y} : \mathbf{R}^2 \to \mathbf{R}, \quad (x,y) \mapsto 2y + x - \mathrm{e}^x \sin y.$$

它可以帮我们计算 $f(x,y)$ 在任意 $(x,y) = (x_0,y_0)$ 处的分别关于 x 变量和 y 变量的偏导数. 从这个角度看, 我们可以把偏导数也叫做**偏导函数**. 偏导数是由一个二元函数得到的, 而偏导数 $\dfrac{\partial f}{\partial x}$ 或 $\dfrac{\partial f}{\partial y}$ 本身, 也可以视作

一个二元函数. 总之, 一元函数的导 (函) 数是 1 个一元函数, 而二元函数的偏导 (函) 函数则是 2 个二元函数. (下面, 当我们把偏导数扩张到任意元情况时, n 元函数的偏导 (函) 函数则是 n 个 n 元函数.)

例 6.4.4 设函数 $z = f(x,y) = x\ln\left(x^2+y^2\right)$, 求 $\left.\dfrac{\partial z}{\partial x}\right|_{(1,2)}$.

证明. 此题有两种求解方法.

1. 方法一. 先将 $y=2$ 代入该函数的表达式, 得到
$$z = f(x,2) = x\ln\left(x^2+4\right).$$
然后, 将 $f(x,2)$ 作为 x 的一元函数求导数, 即得到
$$\left.\frac{\partial z}{\partial x}\right|_{(1,2)} = \left.\frac{\mathrm{d}}{\mathrm{d}x}f(x,2)\right|_{x=1} = \left.\left(\ln\left(x^2+4\right)+\frac{2x^2}{x^2+4}\right)\right|_{x=1} = \ln 5+\frac{2}{5}.$$
2. 方法二. 求出 $\dfrac{\partial z}{\partial x}$ 在任意一点 (x,y) 处的表达式, 再用 $x=1$ 及 $y=2$ 代入. 事实上, 很容易计算得
$$\frac{\partial z}{\partial x} = \ln\left(x^2+y^2\right)+x\cdot\frac{2x}{x^2+y^2}.$$
这样, 就有
$$\left.\frac{\partial z}{\partial x}\right|_{(1,2)} = \ln 5+\frac{2}{5}.$$

□

上一道例题中, 我们用了两种办法计算某个具体点的偏导数. 很难说这两种方法中哪一个更简便些. 如果在一个具体题目中先代入某个自变量的值会化简表达式, 这时利用第一种方法就会简便些. 下面的例子就说明了这一点.

例 6.4.5 设函数 $z = \arctan\dfrac{(x-2)y+y^2}{xy+(x-2)^2y^3}$, 求 $\left.\dfrac{\partial z}{\partial y}\right|_{(2,0)}$.

证明. 我们先求得 $z(2,y) = \arctan\dfrac{y}{2}$, 于是
$$\left.\frac{\partial z}{\partial y}\right|_{(2,0)} = \left.\frac{\mathrm{d}z(2,y)}{\mathrm{d}y}\right|_{y=0} = \left.\frac{\frac{1}{2}}{1+\frac{y^2}{4}}\right|_{y=0} = \frac{1}{2}.$$
显然, 本题只要求计算在某个具体点出的偏导数. 如果直接把 x 视作常数, 计算偏导数 $\dfrac{\partial z}{\partial y}(x,y)$ 的一般表达式无疑是一个巨大的工作. □

例 6.4.6 设函数
$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2}, & (x,y)\neq(0,0) \\ 0, & (x,y)=(0,0) \end{cases}$$
求 $f_x(x,y)$ 和 $f_y(x,y)$.

证明. 显然要根据 $(x,y)=(0,0)$ 与否来分类计算.

1. 对于任意的 $(x,y)\neq(0,0)$, 有

$$f_x(x,y)=\frac{y\left(x^2+y^2\right)-xy\cdot 2x}{\left(x^2+y^2\right)^2}=\frac{y^3-x^2y}{\left(x^2+y^2\right)^2},$$
$$f_y(x,y)=\frac{x\left(x^2+y^2\right)-xy\cdot 2y}{\left(x^2+y^2\right)^2}=\frac{x^3-xy^2}{\left(x^2+y^2\right)^2}.$$

   这里有一个今后常用的技巧, 如果一个二元函数关于其自变量是对称的, 即对于任意有自然定义的 x, y, 恒有

$$f(x,y)=f(y,x),$$

   那么给定函数 f 的一个偏导数, 只需将其表达式中的 x,y 互换位置就能得到它的另一个偏导数的表达式.

2. 在点 $(0,0)$ 处, 只能根据偏导数的定义计算. 此时, 有

$$f_x(0,0)=\lim_{\Delta x\to 0}\frac{f(0+\Delta x,0)-f(0,0)}{\Delta x}=\lim_{\Delta x\to 0}\frac{0-0}{\Delta x}=0,$$
$$f_y(0,0)=\lim_{\Delta y\to 0}\frac{f(0,0+\Delta y)-f(0,0)}{\Delta y}=\lim_{\Delta x\to 0}\frac{0-0}{\Delta y}=0.$$

综上,

$$f_x(x,y)=\begin{cases}\dfrac{y^3-x^2y}{\left(x^2+y^2\right)^2}, & (x,y)\neq(0,0),\\ 0, & (x,y)=(0,0),\end{cases}$$

以及

$$f_y(x,y)=\begin{cases}\dfrac{x^3-xy^2}{\left(x^2+y^2\right)^2}, & (x,y)\neq(0,0),\\ 0, & (x,y)=(0,0).\end{cases}$$

□

二元函数的偏导数概念可以很自然地推广到 n 元函数 $(n\geq 3)$ 的情形. 比如, 对于三元函数 $u=u(x,y,z)$, 其关于 x 的偏导数即是固定自变量 y 和 z 后, 将 u 视为关于 x 的一元函数, 并对其求导. 其他偏导数的定义也类似.

结论 6.4.1: n 元函数的偏导数

理解偏导数的核心在于: 对于一个 n 元函数 $f(x_1,x_2,\ldots,x_n)$, 其对第 i 个变量 x_i 的偏导数是在固定其他所有变量不变的前提下, 将 x_i 视为唯一可变的自变量, 并将函数看作关于 x_i 的一元函数, 对其进行求导.

例 6.4.7 设函数 $u=\left(x^2+y^2\right)z^2+\sin x^2$, 求 $\dfrac{\partial u}{\partial x}, \dfrac{\partial u}{\partial y}$ 及 $\dfrac{\partial u}{\partial z}$.

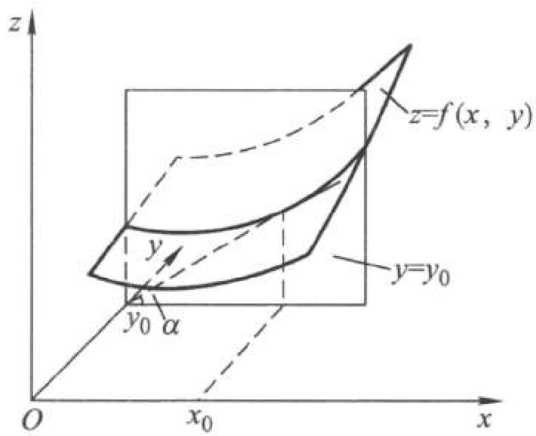


图 6.8: 偏导数的几何意义.

证明. 先求 $\dfrac{\partial u}{\partial x}$, 这时应将函数表达式中的 y 与 z 均视作常数, 于是

$$\frac{\partial u}{\partial x}=2xz^2+2x\cos x^2$$

类似地, 我们得到

$$\frac{\partial u}{\partial y}=2yz^2,\quad \frac{\partial u}{\partial z}=2z\left(x^2+y^2\right).$$

□

下面我们讨论偏导数的几何意义. 参考 偏導數–GeoGebra.

1. 由于 $f_x\left(x_0,y_0\right)=\left.\dfrac{\mathrm{d}f\left(x,y_0\right)}{\mathrm{d}x}\right|_{x=x_0}$, 所以对于连续函数 $f(x,y)$, 在几何上, $f_x\left(x_0,y_0\right)$ 的表示一条平面中的曲线

$$l_1:\begin{cases} z=f(x,y),\\ y=y_0,\end{cases}$$

在点 $x=x_0$ 处的切线, 关于 x 轴的斜率 (见图 6.8). 这里 l_1 实际上是曲面 $z=f(x,y)$ 与平面 $y=y_0$ (垂直于 y 轴) 的交线.

2. 同理, $f_y\left(x_0,y_0\right)$ 表示平面曲线

$$l_2:\begin{cases} z=f(x,y),\\ x=x_0,\end{cases}$$

在点 $y=y_0$ 处的切线关于 y 轴的斜率. 这里 l_2 是曲面 $z=f(x,y)$ 与平面 $x=x_0$ (垂直于 x 轴) 的交线.

因此, $f_x\left(x_0,y_0\right)$ 与 $f_y\left(x_0,y_0\right)$ 是曲面 $z=f(x,y)$ 在点 $\left(x_0,y_0\right)$ 处分别沿着 x 轴方向与 y 轴方向的切线斜率.

问题 上述讨论涉及了两个特殊的平面, 它们垂直于 Oxy 平面, 且都通过点 $\left(x_0,y_0\right)$. 但是一个平行于 x 轴方向和一个平行于 y 轴方向, 自然地, 若我们考虑其他垂直于 Oxy 平面且经过点 $\left(x_0,y_0\right)$ 的平面, 显然它们也会与二元函数 $z=f(x,y)$ 对应的曲面相交, 且相交部分又是一条平面曲线. 那么, 是否可以讨论该曲线在点 $\left(x_0,y_0\right)$ 处的切线斜率呢? 答案是可以的. 这就引出了方向导数的概念, 它可以看作是偏导数的扩展. 这部分将在后面介绍.

我们知道, 对于一元函数 $y=f(x)$, 若它在一点 x_0 处有导数, 则意味它在该点处一定是连续的. 对于二元函数, 它在一点 (x_0,y_0) 处关于 x 与 y 的偏导数都存在时, 不能推出它在该点处连续.

例 6.4.8 考虑二元符号函数 $z=\operatorname{sgn}(xy)$, 其定义为:

$$z=\operatorname{sgn}(xy)=\begin{cases}1 & \text{if } xy>0,\\ 0 & \text{if } xy=0,\\ -1 & \text{if } xy<0.\end{cases}$$

见图 6.9 和动画 二元函数偏导都存在的点不一定连续 | Desmos. 我们来看该函数在 $(0,0)$ 处的偏导数: 将 y 固定为 0, 此时, $\operatorname{sgn}(x\cdot 0)=\operatorname{sgn}(0)=0$. 所以,

$$\frac{\partial}{\partial x}\operatorname{sgn}(xy)\bigg|_{(x,y)=(0,0)}=0.$$

将 x 固定为 0, 此时, $\operatorname{sgn}(0\cdot y)=\operatorname{sgn}(0)=0$. 所以

$$\frac{\partial}{\partial y}\operatorname{sgn}(xy)\bigg|_{(x,y)=(0,0)}=0.$$

因而 $\operatorname{sgn}(xy)$ 在点 $(0,0)$ 处的两个偏导数都存在并等于零. 然而, $\operatorname{sgn}(xy)$ 在点 $(0,0)$ 处不连续, 例如: 沿着 $y=x$ 接近原点时, $xy=x^2>0$, 所以 $\operatorname{sgn}(xy)=1$; 沿着 $y=-x$ 接近原点时, $xy=-x^2<0$, 所以 $\operatorname{sgn}(xy)=-1$. 这就说明, 两个偏导数存在, 不能说明其连续性.

例 6.4.9 证明 $f(x,y)=|x|+|y|$ 在 $(0,0)$ 处连续, 但偏导数 $f_x(0,0)$ 和 $f_y(0,0)$ 都不存在. 因为, 函数 $f(x,0)=|x|$ 在 $x=0$ 处不可导. 同理, 函数 $f(0,y)=|y|$ 在 $y=0$ 处不可导. 所以它两个偏导数都不存在.

上面的两个例子说明了, 对于多元函数而言, 在某点处连续和在某点处可导这二者之间没有什么关系.

6.4.2 高阶偏导数

先复习一下导数和高阶导数的概念. 我们用记号 $F(\mathbf{R})$ 来表示所有定义在实数域 $\mathbf{R}$ 上的一元函数的集合:

$$F(\mathbf{R})=\{f:\mathbf{R}\to\mathbf{R}\mid f(x)\in\mathbf{R}\}.$$

在这里, 我们并不要求每个函数 $f(x)$ 的定义域必须是整个 $\mathbf{R}$. 我们把定义域写成 $\mathbf{R}$, 只是为了强调 f 是一个一元函数, 即其输入是单个变量, 而不是多元变量. 为了简洁起见, 我们继续考虑 $F(\mathbf{R})$ 作为讨论的集合.

假设我们遇到的一元函数的导数都一律存在. 那么, 求导的过程可以看作是一个从 $F(\mathbf{R})$ 到 $F(\mathbf{R})$ 的映射, 写做:

$$\frac{\mathrm{d}}{\mathrm{d}x}\equiv\frac{\mathrm{d}(\cdot)}{\mathrm{d}x}\equiv\frac{\mathrm{d}}{\mathrm{d}x}(\cdot):F(\mathbf{R})\to F(\mathbf{R}),\quad f\mapsto\frac{\mathrm{d}f}{\mathrm{d}x}=\frac{\mathrm{d}}{\mathrm{d}x}f.$$

例如, $f(x)=\sin(x)\in F(\mathbf{R})$. 对其求导,

$$\frac{\mathrm{d}}{\mathrm{d}x}f(x)=\frac{\mathrm{d}}{\mathrm{d}x}\sin(x)=\cos(x)\in F(\mathbf{R}).$$

即我们通过某种规则, 对于给定的输入是正弦函数, 我们输出余弦函数. 这种输入和输出本身就是函数的函数, 我们习惯上称其为**算子 (operator)**. 所以, 求导运算, 或者说导数的记号 $\dfrac{\mathrm{d}}{\mathrm{d}x}$ 其实是一种是算子.

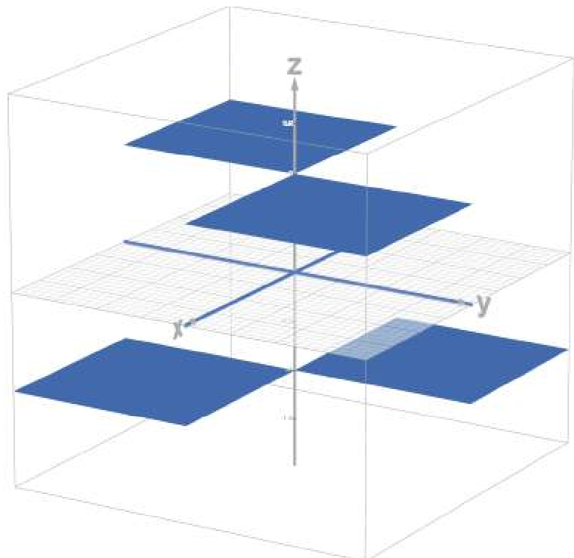


图 6.9: 函数 $\mathrm{sgn}(xy)$ 的图像.

既然, $\dfrac{\mathrm{d}}{\mathrm{d}x}$ 的本质是一个函数, 那么它就可以进行函数复合的操作. 那么连续复合 2 次 $\dfrac{\mathrm{d}}{\mathrm{d}x}$ 得到就是求二阶导数的算子, 写做:

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} \equiv \frac{\mathrm{d}}{\mathrm{d}x} \circ \frac{\mathrm{d}}{\mathrm{d}x} : F(\mathbf{R}) \to F(\mathbf{R}), \quad f \mapsto \frac{\mathrm{d}^2 f}{\mathrm{d}x^2} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}f}{\mathrm{d}x}\right).$$

注意, 算子的定义域和到达域一直是一元函数集合. 类似的, 任意高阶导数的算子就是将基础算子 $\dfrac{\mathrm{d}}{\mathrm{d}x}$ 不断复合得到.

上一小节, 我们提到: 偏导数是由一个二元函数 $f(x,y)$ 得到的, 而偏导数 $\dfrac{\partial f}{\partial x}(x,y)$ 或 $\dfrac{\partial f}{\partial y}(x,y)$ 本身, 又可以视作一个二元函数, 这就产生了一个算子. 类似地, 我们用记号 $F(\mathbf{R}^2)$ 来表示全体在 $\mathbf{R}^2$ 上有定义的二元函数的集合:

$$F(\mathbf{R}^2) = \left\{ f : \mathbf{R}^2 \to \mathbf{R},\ f(x,y) \in \mathbf{R} \right\}.$$

我们假设遇到的二元函数的偏导数都一律存在. 那么, 偏导数记号 $\dfrac{\partial}{\partial x}$ 或 $\dfrac{\partial}{\partial y}$ 本身, 就能视作是一个从 $F(\mathbf{R}^2)$ 到 $F(\mathbf{R}^2)$ 的函数, 写做:

$$\begin{aligned} \frac{\partial}{\partial x} &\equiv \frac{\partial(\cdot)}{\partial x} \equiv \frac{\partial}{\partial x}(\cdot) : F(\mathbf{R}^2) \to F(\mathbf{R}^2), \quad f \mapsto \frac{\partial f}{\partial x}, \\ \frac{\partial}{\partial y} &\equiv \frac{\partial(\cdot)}{\partial y} \equiv \frac{\partial}{\partial y}(\cdot) : F(\mathbf{R}^2) \to F(\mathbf{R}^2), \quad f \mapsto \frac{\partial f}{\partial y}. \end{aligned}$$

所以, 求偏导数的操作, 或者说偏导数记号 $\dfrac{\partial}{\partial x}$ 或 $\dfrac{\partial}{\partial y}$ 都是算子. 例如, 考虑 $f(x,y) = x^2y^3 \in F(\mathbf{R}^2)$,[9] 求其 2 个偏导数得到

$$\frac{\partial f}{\partial x}(x,y) = 2xy^3 \in F(\mathbf{R}^2), \quad \frac{\partial f}{\partial y}(x,y) = 3x^2y^2 \in F(\mathbf{R}^2).$$

即我们通过某种规则, 对于输入的二元函数, 我们输出则是另外一个二元函数.

自然地, 我们也可以对偏导数算子 $\dfrac{\partial}{\partial x}$ 和/或 $\dfrac{\partial}{\partial y}$ 进行复合操作. 注意此时我们的一阶偏导数有两种类型的算子, 而函数复合是有顺序的, 所以我们一共有以下四种类型的组合:

[9]我们并不要求每个函数 $f(x,y)$ 的定义域必须是整个 $\mathbf{R}^2$. 只是想强调 f 是一个二元函数.

1. $\dfrac{\partial^2}{\partial x^2} := \dfrac{\partial}{\partial x} \circ \dfrac{\partial}{\partial x} : F(\mathbf{R}^2) \to F(\mathbf{R}^2), \quad f \mapsto \dfrac{\partial f}{\partial x} \mapsto \dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial x}\right) = \dfrac{\partial^2 f}{\partial x^2}.$

2. $\dfrac{\partial^2}{\partial x \partial y} := \dfrac{\partial}{\partial y} \circ \dfrac{\partial}{\partial x} : F(\mathbf{R}^2) \to F(\mathbf{R}^2), \quad f \mapsto \dfrac{\partial f}{\partial x} \mapsto \dfrac{\partial}{\partial y}\left(\dfrac{\partial f}{\partial x}\right) = \dfrac{\partial^2 f}{\partial x \partial y}$

3. $\dfrac{\partial^2}{\partial y \partial x} := \dfrac{\partial}{\partial x} \circ \dfrac{\partial}{\partial y} : F(\mathbf{R}^2) \to F(\mathbf{R}^2), \quad f \mapsto \dfrac{\partial f}{\partial y} \mapsto \dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial y}\right) = \dfrac{\partial^2 f}{\partial y \partial x}$

4. $\dfrac{\partial^2}{\partial y^2} := \dfrac{\partial}{\partial y} \circ \dfrac{\partial}{\partial y} : F(\mathbf{R}^2) \to F(\mathbf{R}^2), \quad f \mapsto \dfrac{\partial f}{\partial y} \mapsto \dfrac{\partial}{\partial y}\left(\dfrac{\partial f}{\partial y}\right) = \dfrac{\partial^2 f}{\partial y^2}$

这就是二阶偏导数. 下面给出正式定义.

定义 6.4.2: 二阶偏导数

设函数 $z = f(x,y)$ 在区域 D 内有偏导数 $f_x(x,y)$ 与 $f_y(x,y)$. 若这两个偏导函数自身在区域 D 内也有偏导数, 那么它们的偏导数就称作 $z = f(x,y)$ 的二阶偏导数. 根据每次偏导的计算方向不同, $z = f(x,y)$ 的二阶偏导数有四种, 分别记作:

1. $f_{xx}, \dfrac{\partial^2 f}{\partial x^2}, \dfrac{\partial^2 z}{\partial x^2}$ 或 z_{xx};

2. $f_{xy}, \dfrac{\partial^2 f}{\partial x \partial y}, \dfrac{\partial^2 z}{\partial x \partial y}$ 或 z_{xy};

3. $f_{yx}, \dfrac{\partial^2 f}{\partial y \partial x}, \dfrac{\partial^2 z}{\partial y \partial x}$ 或 z_{yx};

4. $f_{yy}, \dfrac{\partial^2 f}{\partial y^2}, \dfrac{\partial^2 z}{\partial y^2}$ 或 z_{yy};

以上四个都是 $z = f(x,y)$ 的**二阶偏导数**. 其中, $f_{xy}(x,y)$ 及 $f_{yx}(x,y)$ 称作**二阶混合偏导数**.

对于二阶混合偏导数有个记法问题. 根据通常的习惯, $f_{xy}(x,y)$ (或 $\dfrac{\partial^2 f}{\partial x \partial y}$) 表示先对 x 求偏导数, 再对 y 求偏导数, 而 $f_{yx}(x,y)$ 则是先对 y 求偏导数, 再对 x 求偏导数. 本书的这个习惯和复合函数的顺序正好相反.[10]

例 6.4.10 设函数 $z = y\mathrm{e}^{xy}$, 求该函数的二阶偏导数.

证明. $z_x = y^2\mathrm{e}^{xy}, z_y = (1+xy)\mathrm{e}^{xy}, z_{xx} = y^3\mathrm{e}^{xy}$,

$$\begin{aligned} z_{xy} &= \left(2y + xy^2\right)\mathrm{e}^{xy}, \\ z_{yx} &= [y + (1+xy)y]\mathrm{e}^{xy} = \left(2y + xy^2\right)\mathrm{e}^{xy}, \\ z_{yy} &= [x + x(1+xy)]\mathrm{e}^{xy} = \left(2x + x^2 y\right)\mathrm{e}^{xy}. \end{aligned}$$

在本例中, 我们看到 $z_{xy} = z_{yx}$, 即两个二阶混合偏导数相等. □

[10] 本书和同济高数的记号都是: $\dfrac{\partial^2 f}{\partial x \partial y} = \dfrac{\partial}{\partial y}\left(\dfrac{\partial f}{\partial x}\right)$, $\dfrac{\partial^2 f}{\partial y \partial x} = \dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial y}\right)$. 但是就我学习过的日文英文教材, 大多写成 $\dfrac{\partial^2 f}{\partial x \partial y} = \dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial y}\right)$, $\dfrac{\partial^2 f}{\partial y \partial x} = \dfrac{\partial}{\partial y}\left(\dfrac{\partial f}{\partial x}\right)$, 正好反的. 这正好就是函数复合的顺序. 我们目前就按照本书的习惯去写, 只要心中记着这是就行了.

在某些情况下, 两个二阶混合偏导数可能不相等 (见习题 6.4 中的第 16 题). 二阶混合偏导数相等是有条件的.

定理 6.4.1: 克莱罗定理

若函数 $f(x,y)$ 的两个二阶混合偏导数 $f_{yx}(x,y)$ 和 $f_{xy}(x,y)$ 在区域 D 内连续, 则在该区域内这两个二阶混合偏导数必相等, 即

$$f_{xy}(x,y)=f_{yx}(x,y),\quad \forall (x,y)\in D.$$

这个定理的证明较长, 故略去. 重要的是读者应该记住: 两个二阶混合偏导数在一个区域内连续可以保证它们彼此相等. 在例题 6.4.10, 就是 z_{xy} 与 z_{yx} 是关于 (x,y) 的连续函数.

定义了二阶偏导数之后, 自然可以考虑二阶偏导数的偏导数, 即得到三阶偏导数. 一个二元函数 $f(x,y)$ 有四个二阶偏导数, 而每个二阶偏导数又可以分别对 x 与 y 求偏导数, 因此一共有八个三阶偏导数. 这里我们就不一一写出了. 以此类推, 我们可以定义各阶偏导数.

定义 6.4.3

今后, 我们用 $C^n(D)$ 表示区域 D 内全体函数 $f(x,y)$ 组成的集合, 其中 $f(x,y)$ 在 D 内有 n 阶偏导数且每个 n 阶偏导数都在 D 内连续.

有了这个记号之后, 由定理 6.4.1 便可以推出:

推论 6.4.1

若 $f(x,y)\in C^2(D)$, 则在 D 内有 $f_{xy}(x,y)=f_{yx}(x,y)$.

例 6.4.11 证明: 函数 $z=\ln\sqrt{x^2+y^2}$ 满足平面拉普拉斯 *(Laplace)* 方程

$$\frac{\partial^2 z}{\partial x^2}+\frac{\partial^2 z}{\partial y^2}=0.$$

证明. 我们有

$$\frac{\partial z}{\partial x}=\frac{x}{x^2+y^2},\quad \frac{\partial^2 z}{\partial x^2}=\frac{\left(x^2+y^2\right)-x\cdot 2x}{\left(x^2+y^2\right)^2}=\frac{y^2-x^2}{\left(x^2+y^2\right)^2}.$$

由该函数关于自变量的对称性可知, 函数 $\dfrac{\partial^2 z}{\partial y^2}$ 就是将 $\dfrac{\partial^2 z}{\partial x^2}$ 表达式中 x,y 交换位置即可得到. 故,

$$\frac{\partial^2 z}{\partial y^2}=\frac{x^2-y^2}{\left(x^2+y^2\right)^2}.$$

于是, $\dfrac{\partial^2 z}{\partial x^2}+\dfrac{\partial^2 z}{\partial y^2}=0$. 证毕. □

一个含有未知函数偏导数的方程称作**偏微分方程 (partial differential equation, PDE)**. 显然, 上述拉普拉斯方程是一个偏微分方程. 诸多物理现象可以通过拉普拉斯方程来描述. 上例告诉我们: $z=\ln\sqrt{x^2+y^2}$ 是平面拉普拉斯方程的一个解.

例 6.4.12 证明: 函数 $u=\dfrac{1}{\sqrt{x^2+y^2+z^2}}$ 满足空间拉普拉斯方程

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}=0.$$

证明. 可求出

$$\begin{aligned}
\frac{\partial u}{\partial x}&=-\frac{x}{(x^2+y^2+z^2)^{\frac{3}{2}}},\\
\frac{\partial^2 u}{\partial x^2}&=-\frac{1}{(x^2+y^2+z^2)^{\frac{3}{2}}}+\frac{3x^2}{(x^2+y^2+z^2)^{\frac{5}{2}}}.
\end{aligned}$$

由函数关于自变量的对称性可知

$$\begin{aligned}
\frac{\partial^2 u}{\partial y^2}&=-\frac{1}{(x^2+y^2+z^2)^{\frac{3}{2}}}+\frac{3y^2}{(x^2+y^2+z^2)^{\frac{5}{2}}},\\
\frac{\partial^2 u}{\partial z^2}&=-\frac{1}{(x^2+y^2+z^2)^{\frac{3}{2}}}+\frac{3z^2}{(x^2+y^2+z^2)^{\frac{5}{2}}}.
\end{aligned}$$

于是, 不难得出

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}=-\frac{3}{(x^2+y^2+z^2)^{\frac{3}{2}}}+\frac{3\left(x^2+y^2+z^2\right)}{(x^2+y^2+z^2)^{\frac{5}{2}}}=0.$$

证毕. □

若引进算子符号

$$\Delta=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2},$$

则上个例题中的方程可写成 $\Delta u=0$. Δ 称为拉普拉斯算子, 它是一个从 $F(\mathbf{R}^3)$ 到 $F(\mathbf{R}^3)$ 的映射, 其中 $F(\mathbf{R}^3)$ 来表示三元实数值函数的全体集合. 上一题说明, 对于三元函数 $u=\dfrac{1}{\sqrt{x^2+y^2+z^2}}$, 其拉普拉斯算子的返回值是值为零的常函数.

例 6.4.13 设函数 $f(x,y,z,w)=z\mathrm{e}^{w^2+x^2+y^2}$, 求 $\dfrac{\partial^2 f}{\partial x\partial w}$, $\dfrac{\partial^2 f}{\partial w\partial z}$ 及 $\dfrac{\partial^2 f}{\partial z\partial x}$.

证明. 先求一阶偏导数:

$$\begin{aligned}
\frac{\partial f}{\partial x}&=z\mathrm{e}^{w^2+x^2+y^2}\cdot 2x=2xz\mathrm{e}^{w^2+x^2+y^2},\\
\frac{\partial f}{\partial w}&=z\mathrm{e}^{w^2+x^2+y^2}\cdot 2w=2wz\mathrm{e}^{w^2+x^2+y^2},\\
\frac{\partial f}{\partial z}&=\mathrm{e}^{w^2+x^2+y^2}
\end{aligned}$$

再求二阶偏导数:

$$\begin{aligned}
\frac{\partial^2 f}{\partial x\partial w}&=\frac{\partial}{\partial w}\left(2xz\mathrm{e}^{w^2+x^2+y^2}\right)=2xz\mathrm{e}^{w^2+x^2+y^2}\cdot 2w=4xzw\mathrm{e}^{w^2+x^2+y^2},\\
\frac{\partial^2 f}{\partial w\partial z}&=\frac{\partial}{\partial z}\left(2wz\mathrm{e}^{w^2+x^2+y^2}\right)=2w\mathrm{e}^{w^2+x^2+y^2},\\
\frac{\partial^2 f}{\partial z\partial x}&=\frac{\partial}{\partial x}\left(\mathrm{e}^{w^2+x^2+y^2}\right)=2x\mathrm{e}^{w^2+x^2+y^2}.
\end{aligned}$$

□

6.4.3 多元实数值的线性函数

本小节是开始全微分前的重要补充知识. 就一般而言, 线性函数 (linear function) 是指满足以下两个特性的函数或算子:

1. 对于两个任意输入 x_1 和 x_2, 有 $f(x_1+x_2)=f(x_1)+f(x_2)$.
2. 对于任意的输入 x 和任意的实数 λ, 有 $f(\lambda x)=\lambda f(x)$.

之前我们讲过, 求导数, 求 (不) 定积分等等抽象操作也都具备线性函数的身影. 上面涉及到的加法和标量乘法要视具体的向量空间才能确定. 但现在, 我们关心的是从 $\mathbf{R}^n$ 到 $\mathbf{R}$ 的多元函数是线性函数时, 会发生什么.

先考虑 $n=1$ 时. 在第一章介绍微分时我们提到过, 如果 $f:\mathbf{R}\to\mathbf{R}$ 是线性函数, 则它一定可以被表示为

$$f(x)=ax,$$

其中 a 是一个常数. 我们更在意高维的情况. 考虑 $n=2$, 假设我们有一个线性函数 $f:\mathbf{R}^2\to\mathbf{R}$, 具体而言它满足以下两个条件:

1. 对于任意 $(x_1,y_1),(x_2,y_2)\in\mathbf{R}^2$, 有 $f((x_1,y_1)+(x_2,y_2))=f(x_1,y_1)+f(x_2,y_2)$.
2. 对于任意 $\lambda\in\mathbf{R}$ 和 $(x,y)\in\mathbf{R}^2$, 有 $f(\lambda x,\lambda y)=\lambda f(x,y)$.

我们现在证明它一定可以表示为

$$f(x,y)=ax+by$$

的形式, 其中 a 和 b 是常数. 如果定义动向量 $P=(x,y)$ 和常向量 $\boldsymbol{c}=(a,b)$, 那么该线性函数可以写做内积的形式:

$$f(P)=P\cdot\boldsymbol{c},\quad \forall P\in\mathbf{R}^2.$$

其证明如下.

证明. 考虑 $\mathbf{R}^2$ 中的两个特殊向量 $(1,0)$ 和 $(0,1)$ (称之为**基底向量**), 并令 $a:=f(1,0)$ 和 $b:=f(0,1)$. 注意到任意的 $(x,y)\in\mathbf{R}^2$ 都可以表示为 $(x,y)=x(1,0)+y(0,1)$. 于是, 利用线性函数的性质可得

$$f(x,y)=f(x(1,0)+y(0,1))=f(x(1,0))+f(y(0,1))=xf(1,0)+yf(0,1).$$

这就是说明了 $f(x,y)=ax+by$. 证毕. □

上面的实事很容易扩展到 $\mathbf{R}^n$ 中, 其证明和上述过程完全一致, 只需要将任意向量拆解为基底向量的组合即可. 比如, 任意的 $(x,y,z)\in\mathbf{R}^3$ 都可以表示为 $(x,y,z)=x(1,0,0)+y(0,1,0)+z(0,0,1)$, 其中 $(1,0,0),(0,1,0),(0,0,1)$ 称作 $\mathbf{R}^3$ 空间的基底向量. $\mathbf{R}^n$ 空间的 n 个基底向量.

命题 6.4.1: 多元实数值的线性函数

如果一个多元函数 $f:\mathbf{R}^n\to\mathbf{R}$ 是线性函数, 那么一定存在某个常向量 $\boldsymbol{c}=(c_1,\dots,c_n)\in\mathbf{R}^n$, 使得对于任意的 $P=(x_1,\dots,x_n)\in\mathbf{R}^n$, 总有

$$f(P)=P\cdot\boldsymbol{c}=\sum_{i=1}^{n}c_ix_i.$$

所以, 任何形如 $\displaystyle\sum_{i=1}^{n}c_ix_i$ 的 n 元函数当且仅当它是 n 元实值的线性函数.

总结一下, 对于单变量 x 而言, 我们说关于 x 的线性函数就是指函数

$$x \mapsto ax,$$

其中 a 是某常数. 对于两个变量 x,y 而言, 我们说关于 x,y 的线性函数就是指二元函数

$$(x,y) \mapsto ax+by,$$

其中 a,b 是某常数. 那么我们说关于三个变量 x,y,z 的线性函数就是指三元函数

$$(x,y,z) \mapsto ax+by+cz,$$

其中 a,b,c 是某常数. 结合下一小节, 我们知道: **任何 n 元函数的在某点处的全微分 (如果存在的话) 就是一个 n 元实值的线性函数.**

6.4.4 全微分

多元函数的全微分是一元函数的微分的推广.

对于一元函数, 当我们考虑函数增量的近似表达式时, 微分是函数增量的线性主要部分. 设一元函数 $y=f(x)$ 在点 x_0 的某个邻域内有定义. 如果存在一个常数 A, 使得

$$\Delta y=f\left(x_0+\Delta x\right)-f\left(x_0\right)=A\Delta x+o(\Delta x), \quad (\Delta x \to 0),$$

成立, 则称 $y=f(x)$ 在点 x_0 处可微, 并把关于 Δx 的线性函数

$$\Delta x \mapsto A\Delta x$$

称作 $y=f(x)$ 在点 x_0 处的微分, 记作 $\mathrm{d}y=A\Delta x$. 注意 $\Delta x \to 0$ 等价于 $|\Delta x| \to 0$. 而 $|\Delta x|$ 刻画着自变量增量的模. 当 $|\Delta x|$ 充分小时, 就有函数增量 Δy 的近似表达式:

$$\Delta y \approx \mathrm{d}y=A\Delta x.$$

因为右侧微分是一个关于 Δx 的线性函数, 所以我们说微分是函数增量的线性主要部分. 在这种情况下, 我们还证明过 $A=f'(x_0)$, 特别是可微等价于可导.[11]

对于一般的二元函数甚至多元函数, 我们也希望能有类似的 "微分" 的概念. 我们大胆地从一元函数的微分公式中猜测出一种范式:

$$\boxed{\text{函数值增量} = \text{关于“自变量增量”的线性函数} + o(\text{自变量增量的模}), \quad (\text{自变量增量的模} \to 0).} \tag{6.1}$$

或者说有以下极限成立:

$$\frac{\text{函数值增量} - \text{关于“自变量增量”的线性函数}}{\text{自变量增量的模}} \to 0, \quad (\text{自变量增量的模} \to 0).$$

下面我们以二元函数为例, 尝试着去构建这样子的范式.

对于二元函数 $z=f(x,y)$, 当自变量 x,y 分别有增量 $\Delta x, \Delta y$ 时, 函数 z 的增量 (称为**全增量**) 也就是

$$\Delta z := f(x+\Delta x, y+\Delta y)-f(x,y).$$

[11] 下面马上就能知道, 可微等价于可导, 是存在于讨论一元函数时. 多元函数不存在等价关系.

注意, 此时自变量是二元向量 (x,y), 所以自变量增量不再是单个实数, 而是向量 $(\Delta x,\Delta y)\in\mathbf{R}^2$. 为了表示这个自变量增量 $(\Delta x,\Delta y)$ 整体的大小, 我们使用它的模来描述:

$$\rho:=|(\Delta x,\Delta y)|=\sqrt{(\Delta x)^2+(\Delta y)^2}.$$

显然, $\rho\to 0$ 当且仅当 $\Delta x\to 0$ 且 $\Delta y\to 0$. 现在我们能够用一个标量 ρ 同时来刻画两个增量 $\Delta x,\Delta y$ 趋向于零的极限情况了. 这部分的讨论扩到到任意多元函数中也是非常自然的. 最后只剩下对 "关于自变量增量 $\Delta x,\Delta y$ " 的线性函数的讨论了. 前一小节的内容说明, 它一定这样子的函数:

$$(\Delta x,\Delta y)\mapsto A\Delta x+B\Delta y,$$

其中 A,B 是某些常数. 根据命题 6.4.1, 对于三元或以上的, 我们也可以考虑关于 "自变量增量" 的线性函数. 总结一下要点就是:

- 因为自变量是多元的向量, 所以考虑的自变量增量就是相同维度的向量.[12]
- 使用自变量增量的模 (这单个量) $\rho\to 0$ 可以精确刻画增量的多个不同分量趋向于零的情况.
- 关于 "自变量增量" 的线性函数, 无论它是几维的向量, 命题 6.4.1 都告诉了我们它的具体形式.

至此, 我们可以给出二元函数微分的正式定义. 它完全符合范式 (6.1).

定义 6.4.4: 二元函数的微分, 全微分

设函数 $z=f(x,y)$ 在点 (x_0,y_0) 的某个邻域内有定义. 若 $z=f(x,y)$ 的全增量

$$\Delta z=f\left(x_0+\Delta x,y_0+\Delta y\right)-f\left(x_0,y_0\right)$$

可写成极限等式:

$$\Delta z=A\Delta x+B\Delta y+o(\rho),\quad(\rho\to 0),$$

其中

- 常数 A,B 只与点 (x_0,y_0) 有关而与自变量的增量 $\Delta x,\Delta y$ 无关,
- $\rho:=\sqrt{(\Delta x)^2+(\Delta y)^2}$.

此时, 我们称 $z=f(x,y)$ 在点 (x_0,y_0) 处 **可微**, 并把关于两个自变量增量 $\Delta x,\Delta y$ 的线性函数

$$(\Delta x,\Delta y)\mapsto A\Delta x+B\Delta y,$$

称作 $z=f(x,y)$ 在点 (x_0,y_0) 处的 **全微分 (total differential)**, 记作 $\mathrm{d}z$ 或 $\mathrm{d}f$, 即

$$\mathrm{d}z\equiv\mathrm{d}z(\Delta x,\Delta y)=A\Delta x+B\Delta y.$$

当函数 $z=f(x,y)$ 在区域 D 内每一点处都可微时, 称 $z=f(x,y)$ 在 D 内可微.

由上述定义看出: 全微分 $\mathrm{d}z$ 是 Δx 与 Δy 的线性函数; 当 $\rho\to 0$ 时, $\mathrm{d}z$ 与 Δz 的差是 $o(\rho)$, 即

$$\Delta z=\mathrm{d}z+o(\rho)\quad(\rho\to 0),$$

[12]但因变量及其增量还是实数.

当 $\rho=\sqrt{(\Delta x)^2+(\Delta y)^2}$ 充分小时, 就有函数增量 Δz 的近似表达式:

$$\Delta z \approx \mathrm{d}z = A\Delta x + B\Delta y.$$

因此, 我们说 z 的全微分 $\mathrm{d}z$ 是全增量 Δz 的线性主要部分. 全微分的几何意义就是在用平面来近似表示点 (x_0,y_0) 处附近的曲面. 这个平面就是此处的切平面. (请读者结合第二章微分的几何意义来自行思考全微分的几何意义.)

例 6.4.14 证明函数 $z(x,y)=xy^2$ 在任意点 (x_0,y_0) 处都是可微的, 并写出它的微分表达式.

证明. 函数 $z(x,y)=xy^2$ 在点 (x_0,y_0) 处的全增量为

$$\Delta z = z(x_0+\Delta x, y_0+\Delta y) - z(x_0,y_0).$$

先计算

$$\begin{aligned} z(x_0+\Delta x, y_0+\Delta y) &= (x_0+\Delta x)(y_0+\Delta y)^2 = (x_0+\Delta x)(y_0^2+2y_0\Delta y+(\Delta y)^2) \\ &= x_0y_0^2 + x_0\cdot 2y_0\Delta y + x_0\cdot(\Delta y)^2 + \Delta x\cdot y_0^2 + \Delta x\cdot 2y_0\Delta y + \Delta x\cdot(\Delta y)^2. \end{aligned}$$

最后得到整理得到

$$\Delta z = \underbrace{y_0^2\Delta x + 2x_0y_0\Delta y}_{\text{关于 }\Delta x\text{ 与 }\Delta y\text{ 的线性函数}} + \underbrace{x_0(\Delta y)^2 + 2y_0\Delta x\Delta y + \Delta x(\Delta y)^2}_{\text{关于 }\Delta x\text{ 与 }\Delta y\text{ 的二元高次多项式}}.$$

显然, 前面两项是关于 Δx 与 Δy 的线性函数, 而后面三项是关于 Δx 与 Δy 的二次以上的二元多项式, 当 Δx 与 Δy 很小时, 可以预见后三项比前两项要小得多.

现在我们要严格证明上式等式的右侧中非线性的部分满足

$$E = E(\Delta x,\Delta y) := x_0(\Delta y)^2 + 2y_0\Delta x\Delta y + \Delta x(\Delta y)^2 = o(\rho),$$

其中 $\rho=\sqrt{(\Delta x)^2+(\Delta y)^2}$. 为证 $E=o(\rho)$, 我们需要证明一个二元极限

$$\lim_{(\Delta x,\Delta y)\to(0,0)}\frac{E}{\rho}=0.$$

对 E 取绝对值后应用三角不等式有

$$|E| \le |x_0|(\Delta y)^2 + 2|y_0||\Delta x||\Delta y| + |\Delta x|(\Delta y)^2.$$

由于 $\rho=\sqrt{(\Delta x)^2+(\Delta y)^2}$, 我们有 $|\Delta x|\le\rho$ 且 $|\Delta y|\le\rho$. 因此, 可以进一步估计 $|E|$ 的上界:

$$|E| \le |x_0|\rho^2 + 2|y_0|\rho^2 + \rho\cdot\rho^2 = (|x_0|+2|y_0|)\rho^2 + \rho^3.$$

将上述不等式两边除以 ρ 得到

$$\frac{|E|}{\rho} \le (|x_0|+2|y_0|)\rho + \rho^2.$$

于是, 当 $(\Delta x,\Delta y)\to(0,0)$ 就有 $\rho\to 0$, 则上式的右侧也趋向于 0. 根据二元函数极限的夹逼定理可得

$$\lim_{(\Delta x,\Delta y)\to(0,0)}\frac{E}{\rho}=0.$$

这就证明了 $E=o(\rho)$.

最终, 函数 $z(x,y)=xy^2$ 在点 (x_0,y_0) 处的全增量可表示为

$$\Delta z=y_0^2\Delta x+2x_0y_0\Delta y+o(\rho),\quad(\rho\to 0).$$

这证明了函数 $z(x,y)=xy^2$ 在任意点 (x_0,y_0) 处都是可微分的, 且它全微分就是二元线性函数

$$\mathrm{d}z=\mathrm{d}z(\Delta x,\Delta y)=y_0^2\Delta x+2x_0y_0,$$

其中, 定义 6.4.4 中的 $A=y_0^2, B=2x_0y_0$. 当 $(\Delta x,\Delta y)$ 充分接近 $(0,0)$ 时, 有

$$\Delta z\approx\mathrm{d}z=y_0^2\Delta x+2x_0y_0\Delta y,$$

也即 Δz 可用 Δx 与 Δy 的一个线性函数近似代替. 这个线性函数的图像就是空间中过坐标原点的平面. □

利用全微分的定义, 马上可以得到连续性.

定理 6.4.2: 可微的必要条件: 可微 $\Rightarrow$ 连续

若函数 $f(x,y)$ 在点 (x_0,y_0) 处可微, 则 $f(x,y)$ 在点 (x_0,y_0) 处必连续.

证明. 设函数 $z=f(x,y)$ 在点 (x_0,y_0) 处可微. 那么, 由

$$\Delta z=f\left(x_0+\Delta x,y_0+\Delta y\right)-f\left(x_0,y_0\right)=A\Delta x+B\Delta y+o(\rho),\quad(\rho\to 0),$$

立即推出, 当 $(\Delta x,\Delta y)\to(0,0)$ 时, $\rho\to 0$, 则 $\Delta z\to 0$, 即

$$\lim_{(\Delta x,\Delta y)\to(0,0)}f\left(x_0+\Delta x,y_0+\Delta y\right)=f\left(x_0,y_0\right).$$

这表明, $z=f(x,y)$ 在点 (x_0,y_0) 处连续. □

在一元函数中, 可微意味着可导. 那么, 在多元函数中, 可微是否意味着偏导数存在呢? 回答是肯定的.

定理 6.4.3: 可微的必要条件: 可微 $\Rightarrow$ 可偏导

若函数 $z=f(x,y)$ 在点 (x_0,y_0) 处可微,

1. 则它在点 (x_0,y_0) 处的两个偏导数存在, 且 $f_x\left(x_0,y_0\right)=A$, $f_y\left(x_0,y_0\right)=B$, 其中 A,B 是定义 6.4.4 中的两个常数.

2. 由第 1 点立马推出: 它在点 (x_0,y_0) 处的全微分可写成

$$\mathrm{d}z=f_x\left(x_0,y_0\right)\Delta x+f_y\left(x_0,y_0\right)\Delta y.$$

证明. 已知 $z=f(x,y)$ 在点 (x_0,y_0) 处可微, 即存在常数 A 与 B, 使得 $\Delta z=A\Delta x+B\Delta y+o(\rho)$, $(\rho\to 0)$. 令 $\Delta y=0$, 这时 $\rho=|\Delta x|$, 并有

$$\Delta z=f\left(x_0+\Delta x,y_0\right)-f\left(x_0,y_0\right)=A\Delta x+o(|\Delta x|),\quad\Delta x\to 0.$$

由此得

$$\frac{f\left(x_0+\Delta x,y_0\right)-f\left(x_0,y_0\right)}{\Delta x}=A+\frac{o(|\Delta x|)}{\Delta x},\quad\Delta x\to 0.$$

取极限得

$$\lim_{\Delta x\to 0}\frac{f\left(x_0+\Delta x,y_0\right)-f\left(x_0,y_0\right)}{\Delta x}=A.$$

上式意味着 $z=f(x,y)$ 在点 (x_0,y_0) 处关于 x 的偏导数存在且等于 A, 即

$$f_x\left(x_0,y_0\right)=A.$$

同理可证 $f_y\left(x_0,y_0\right)=B$. 证毕. □

若 $z=f(x,y)$ 在区域 D 内可微, 则它在 D 内任意一点 (x,y) 处的全微分可写成

$$\mathrm{d}z=f_x(x,y)\Delta x+f_y(x,y)\Delta y.$$

与一元函数的微分类似, 二元函数自变量的全微分就等于其各自的增量, 即 $\mathrm{d}x\equiv\Delta x,\mathrm{d}y\equiv\Delta y$. 这样, 全微分公式又可写成

$$\mathrm{d}z=f_x(x,y)\mathrm{d}x+f_y(x,y)\mathrm{d}y,$$

或

$$\mathrm{d}f=\frac{\partial f}{\partial x}\,\mathrm{d}x+\frac{\partial f}{\partial y}\,\mathrm{d}y.$$

这样, 定理 6.4.3 不仅证明了二元函数可微的必要条件是偏导数存在, 而且还给出了计算全微分的公式.

现在一个自然的问题是: 由偏导数存在是否可以推出可微? 答案是否定的. 因为, 多元函数在某点处的偏导数存在也并不意味着函数在该点处连续. 在例题 6.4.8 中就讨论过了. 下面的例子又验证了这一点.

例 6.4.15 考虑函数

$$f(x,y)=\begin{cases}\dfrac{xy}{x^2+y^2}, & (x,y)\neq(0,0),\\ 0, & (x,y)=(0,0)\end{cases}$$

- 例题 6.4.6 说明它在点 $(0,0)$ 处的偏导数存在, 即 $f_x(0,0)=f_y(0,0)=0$.
- 但它在点 $(0,0)$ 处并不连续. 这是因为 $f(x,y)$ 在点 $(0,0)$ 处连极限都不存在, 它自然也不可能连续. 这一点也很好说明, 之前讲二元函数的极限的讲义中, 就反复出现这种类型的函数在 $(0,0)$ 的极限不存在的讨论. 当点 (x,y) 沿直线 $y=kx$ (k 为常数) 趋向于点 $(0,0)$ 时, 我们可以简化函数, 有

  $$f(x,y)=\frac{x(kx)}{x^2+(kx)^2}=\frac{k}{1+k^2}.$$

  当 k 取不同值时, 上述极限的值也不同. 可见, $f(x,y)$ 在点 $(0,0)$ 处连极限都不存在, 所以也不连续.
- 此外, 既然连续是可微的必要条件. 由此可见, 这个函数在点 $(0,0)$ 处不可微. (如果可微的话, 肯定连续.)

偏导数存在但不可微的例子表明, 一元函数与多元函数在可导与可微的关系上有某种实质性差异. 既然偏导数存在尚不能保证可微, 我们自然希望知道在怎样的条件下才能保证函数的可微性. 下面的定理对此给出了回答.[13]

[13]定理 6.4.4 说明了偏导函数的连续是很好的性质. 但注意, 是存在偏导函数不连续的例子的. 设

$$f(x,y)=\begin{cases}(x^2+y^2)\sin\left(\dfrac{1}{\sqrt{x^2+y^2}}\right), & 当(x,y)\neq(0,0),\\ 0, & 当(x,y)=(0,0).\end{cases}$$

定理 6.4.4: 可微的充分条件: 偏导数连续 $\Rightarrow$ 可微

若函数 $z=f(x,y)$ 的偏导数 $f_x(x,y)$ 与 $f_y(x,y)$ 在点 (x_0,y_0) 的某个邻域内存在, 并且这两个偏导数在点 (x_0,y_0) 处连续, 则 $z=f(x,y)$ 在点 (x_0,y_0) 处可微.

证明. 考虑全增量, 我们有

$$\begin{aligned}\Delta z &= f\left(x_0+\Delta x, y_0+\Delta y\right)-f\left(x_0, y_0\right) \\ &= \left[f\left(x_0+\Delta x, y_0+\Delta y\right)-f\left(x_0, y_0+\Delta y\right)\right]+\left[f\left(x_0, y_0+\Delta y\right)-f\left(x_0, y_0\right)\right]\end{aligned}$$

下面, 我们对上面的两个相加项分别应用拉格朗日中值定理.

注意到, 偏导函数 $f_x(x,y)$ 是良好地定义在点 (x_0,y_0) 的某个邻域内的. 故只要我们考虑充分小的 $\Delta x,\Delta y$, 就能使得点 $(x_0+\Delta x, y_0+\Delta y)$ 还是落在这个邻域之内. 此时, 我们固定第二个变量位置为 $y_0+\Delta y$, 而考虑平面上该邻域内部的某一小线段:

$$\left(x, y_0+\Delta y\right), \quad x_0<x<x_0+\Delta x.$$

并考虑定义在以上区间的关于 x 的一元函数 $f_x\left(\cdot, y_0+\Delta y\right)$, 即

$$x \mapsto f_x(x, y_0+\Delta y)=\frac{\mathrm{d}}{\mathrm{d} x} f\left(x, y_0+\Delta y\right), \quad x_0<x<x_0+\Delta x.$$

根据一元函数导数的性质, 函数 $f\left(\cdot, y_0+\Delta y\right)$ 在区间 $[x_0, x_0+\Delta x]$ 上是连续的, 且在 $(x_0, x_0+\Delta x)$ 上是可导的. 于是我们应用拉格朗日中值定理[14], 得到

$$f\left(x_0+\Delta x, y_0+\Delta y\right)-f\left(x_0, y_0+\Delta y\right)=f_x\left(x_0+\theta_1 \Delta x, y_0+\Delta y\right) \Delta x \quad\left(\exists 0<\theta_1<1\right).$$

类似地, 我们可以考虑关于 y 的一元函数 $f_y\left(x_0, \cdot\right)$, 即

$$y \mapsto f_y(x_0, y)=\frac{\mathrm{d}}{\mathrm{d} y} f\left(x_0, y\right), \quad y_0<y<y_0+\Delta y.$$

对函数 $f\left(x_0, \cdot\right)$ 在区间 $(y_0, y_0+\Delta y)$ 上应用拉格朗日中值定理, 得到

$$f\left(x_0, y_0+\Delta y\right)-f\left(x_0, y_0\right)=f_y\left(x_0, y_0+\theta_2 \Delta y\right) \Delta y \quad\left(\exists 0<\theta_2<1\right).$$

最后, 结合以上结果并代回到全增量, 我们有

$$\Delta z=f_x\left(x_0+\theta_1 \Delta x, y_0+\Delta y\right) \Delta x+f_y\left(x_0, y_0+\theta_2 \Delta y\right) \Delta y \quad\left(\exists 0<\theta_1, \theta_2<1\right).$$

由于 $f_x(x,y), f_y(x,y)$ 在点 (x_0,y_0) 处连续 [注意! 此时我们用到了本定理的重要条件.], 故有

$$\begin{aligned} f_x\left(x_0+\theta_1 \Delta x, y_0+\Delta y\right) &= f_x\left(x_0, y_0\right)+\alpha_1, \\ f_y\left(x_0, y_0+\theta_2 \Delta y\right) &= f_y\left(x_0, y_0\right)+\alpha_2,\end{aligned}$$

其中, 当 $\rho=\sqrt{(\Delta x)^2+(\Delta y)^2} \rightarrow 0$ 时, $\alpha_1 \rightarrow 0, \alpha_2 \rightarrow 0$. 我们不需要知道无穷小量 α_1, α_2 的具体表达式. 将上面两式代入 Δz 的表达式, 得

$$\Delta z=f_x\left(x_0, y_0\right) \Delta x+f_y\left(x_0, y_0\right) \Delta y+\alpha_1 \Delta x+\alpha_2 \Delta y.$$

则对于所有 $(x,y)\in\mathbf{R}^2$, 函数 $f(x,y)$ 的偏导数存在, 这意味函数 $\dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial x}$ 在整个 $\mathbf{R}^2$ 上都有定义. 但 $\dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial x}$ 在点 $(0,0)$ 处不连续. 本节学时不够, 如果有兴趣可以阅读 A differentiable function with discontinuous partial derivatives - Math Insight.

[14] 这种变形表示的拉格朗日中值定理, 在我们的讲义中讲过.

最后我们要证明: 当 $\rho \to 0$ 时, $\alpha_1 \Delta x + \alpha_2 \Delta y = o(\rho)$. 事实上, 有

$$0 \leq \left|\frac{\alpha_1 \Delta x + \alpha_2 \Delta y}{\rho}\right| \leq |\alpha_1| \left|\frac{\Delta x}{\rho}\right| + |\alpha_2| \left|\frac{\Delta y}{\rho}\right| \leq |\alpha_1| + |\alpha_2|,$$

于是由夹逼定理立得

$$\lim_{\rho \to 0} \frac{\alpha_1 \Delta x + \alpha_2 \Delta y}{\rho} = 0,$$

即当 $\rho \to 0$ 时, $\alpha_1 \Delta x + \alpha_2 \Delta y = o(\rho)$. 因此

$$\Delta z = f_x\left(x_0, y_0\right) \Delta x + f_y\left(x_0, y_0\right) \Delta y + o(\rho), \quad \rho \to 0.$$

这就证明了 $z = f(x, y)$ 在点 (x_0, y_0) 处可微. 证毕. □

将定理 6.4.4 对当个点的结论应用于区域内所有点, 则有一下推论.

推论 6.4.2

若 D 是 $\mathbf{R}^2$ 中的一个区域, 而 $f(x, y) \in C^1(D)$, 即 $f(x, y)$ 在 D 中有连续的一阶偏导数, 则 $f(x, y)$ 在 D 内可微.

类似地, 我们可以定义 n $(n \geq 3)$ 元函数的可微性, 并且可以证明有相应于定理 6.4.2, 定理 6.4.3 和定理 6.4.4 的结论及推论成立. 具体见小节 6.4.4.

我们遇到的多元函数大多数是**多元初等函数**. 一个多元初等函数的偏导数若存在的话, 也是多元初等函数, 而多元初等函数在其有定义的区域内是连续的. 所以, 对于多元初等函数而言, 在其有定的区域内只要偏导数存在就一定可微.

例 6.4.16 设函数 $z = x^y$ $(x, y > 0)$, 求 $\mathrm{d}z$.

证明. $\mathrm{d}z = \dfrac{\partial z}{\partial x}\,\mathrm{d}x + \dfrac{\partial z}{\partial y}\,\mathrm{d}y = yx^{y-1}\,\mathrm{d}x + x^y \ln x\,\mathrm{d}y.$ □

例 6.4.17 设函数 $z = \mathrm{e}^x \cos\left(x^2 + y^2\right)$, 求 $\mathrm{d}z$.

证明. $\mathrm{d}z = \dfrac{\partial z}{\partial x}\,\mathrm{d}x + \dfrac{\partial z}{\partial y}\,\mathrm{d}y = \left(\mathrm{e}^x \cos\left(x^2+y^2\right) - 2x\mathrm{e}^x \sin\left(x^2+y^2\right)\right)\mathrm{d}x - 2y\mathrm{e}^x \sin\left(x^2+y^2\right)\mathrm{d}y.$ □

例 6.4.18 已知函数 $z(x, y)$ 的全微分为

$$\mathrm{d}z = \left(6xy - y^2\right)\mathrm{d}x + \left(3x^2 - 2xy\right)\mathrm{d}y,$$

求 $z(x, y)$ 的表达式.

证明. 由已知及全微分的公式有

$$\frac{\partial z}{\partial x}=6xy-y^2,\quad \frac{\partial z}{\partial y}=3x^2-2xy.$$

从 $\dfrac{\partial z}{\partial x}=6xy-y^2$ 可推出 (求原函数)

$$z(x,y)=3x^2y-xy^2+\phi(y),$$

其中 $\phi(y)$ 是关于 y 的一元函数. 再求关于 y 的偏导数, 得到

$$\frac{\partial z}{\partial y}=3x^2-2xy+\phi'(y),$$

而 $\dfrac{\partial z}{\partial y}=3x^2-2xy$, 则 $\phi'(y)=0$, 即 $\phi(y)=C$, 其中 C 是任意常数. 于是

$$z(x,y)=3x^2y-xy^2+C.$$

□

二元函数全微分的概念和全部结论, 都可推广到三元及三元以上函数上. 特别地, 对于区域 $\Omega\subset\mathbf{R}^3$ 中处处可微的三元函数 $u=u(x,y,z)$, 它的全微分公式为

$$\mathrm{d}u=\frac{\partial u}{\partial x}\,\mathrm{d}x+\frac{\partial u}{\partial y}\,\mathrm{d}y+\frac{\partial u}{\partial z}\,\mathrm{d}z.$$

具体见小节 6.4.4.

例 6.4.19 求函数 $u=\left(\dfrac{y}{x}\right)^z$ 在点 $(1,2,-1)$ 处的全微分.

证明. 该函数的偏导数为

$$\frac{\partial u}{\partial x}=z\left(\frac{y}{x}\right)^{z-1}\left(-\frac{y}{x^2}\right),\quad \frac{\partial u}{\partial y}=z\left(\frac{y}{x}\right)^{z-1}\frac{1}{x},\quad \frac{\partial u}{\partial z}=\left(\frac{y}{x}\right)^z\ln\frac{y}{x}.$$

它们在点 $(1,2,-1)$ 的某个邻域内都连续, 故该函数在点 $(1,2,-1)$ 处可微. 另外, 有

$$\left.\frac{\partial u}{\partial x}\right|_{(1,2,-1)}=\frac{1}{2},\quad \left.\frac{\partial u}{\partial y}\right|_{(1,2,-1)}=-\frac{1}{4},\quad \left.\frac{\partial u}{\partial z}\right|_{(1,2,-1)}=\frac{1}{2}\ln 2.$$

所以我们得到

$$\mathrm{d}u|_{(1,2,-1)}=\frac{1}{2}\,\mathrm{d}x-\frac{1}{4}\,\mathrm{d}y+\frac{1}{2}\ln 2\,\mathrm{d}z.$$

□

三元函数的全微分

模仿前面方式, 将全微分扩张到三元函数中. 它仍旧是范式

$$\boxed{\text{函数值增量}=\text{关于“自变量增量”的线性函数}+o(\text{自变量增量的模}),\quad(\text{自变量增量的模}\to 0).}$$

定义 设函数 $z=f(x,y,z)$ 在点 (x_0,y_0,z_0) 的某个邻域内有定义. 若 $z=f(x,y,z)$ 的全增量

$$\Delta z=f(x_0+\Delta x,y_0+\Delta y,z_0+\Delta z)-f(x_0,y_0,z_0)$$

可写成极限等式:

$$\Delta z=A\Delta x+B\Delta y+C\Delta z+o(\rho),\quad(\rho\to 0),$$

其中,

- 常数 A,B,C 只与点 (x_0,y_0,z_0) 有关, 而与自变量的增量 $\Delta x,\Delta y,\Delta z$ 无关,
- $\rho:=\sqrt{(\Delta x)^2+(\Delta y)^2+(\Delta z)^2}$.

此时, 我们称 $z=f(x,y,z)$ 在点 (x_0,y_0,z_0) 处**可微**, 并把关于三个自变量增量 $\Delta x,\Delta y,\Delta z$ 的线性函数

$$(\Delta x,\Delta y,\Delta z)\mapsto A\Delta x+B\Delta y+C\Delta z$$

称作 $z=f(x,y,z)$ 在点 (x_0,y_0,z_0) 处的**全微分 (total differential)**, 记作 $\mathrm{d}z$ 或 $\mathrm{d}f$, 即

$$\mathrm{d}z\equiv\mathrm{d}z(\Delta x,\Delta y,\Delta z)=A\Delta x+B\Delta y+C\Delta z.$$

当函数 $z=f(x,y,z)$ 在区域 $D\in\mathbf{R}^3$ 内每一点处都可微时, 称 $z=f(x,y,z)$ 在 D 内可微.

由上述定义看出, 全微分 $\mathrm{d}z$ 是 Δx, Δy 和 Δz 的线性函数; 当 $\rho\to 0$ 时, $\mathrm{d}z$ 与 Δz 的差是 $o(\rho)$, 即

$$\Delta z=\mathrm{d}z+o(\rho)\quad(\rho\to 0),$$

其中 $\rho=\sqrt{(\Delta x)^2+(\Delta y)^2+(\Delta z)^2}$. 当 ρ 足够小时, 函数增量 Δz 的近似表达式为:

$$\Delta z\approx\mathrm{d}z=A\Delta x+B\Delta y+C\Delta z.$$

因此, 我们可以说, 三元函数 $z=f(x,y,z)$ 的全微分 $\mathrm{d}z$ 是全增量 Δz 的线性主要部分.

可微的必要条件: 可微 $\Rightarrow$ 连续 若函数 $f(x,y,z)$ 在点 (x_0,y_0,z_0) 处可微, 则 $f(x,y,z)$ 在点 (x_0,y_0,z_0) 处必连续.

可微的必要条件: 可微 $\Rightarrow$ 可偏导 若函数 $z=f(x,y,z)$ 在点 (x_0,y_0,z_0) 处可微,

1. 则它在点 (x_0,y_0,z_0) 处的三个偏导数存在, 且

$$f_x(x_0,y_0,z_0)=A,\quad f_y(x_0,y_0,z_0)=B,\quad f_z(x_0,y_0,z_0)=C,$$

其中 A,B,C 是定义中全微分的常数.

2. 由第 1 点立马推出: 它在点 (x_0,y_0,z_0) 处的全微分可写成

$$\mathrm{d}z=f_x(x_0,y_0,z_0)\Delta x+f_y(x_0,y_0,z_0)\Delta y+f_z(x_0,y_0,z_0)\Delta z.$$

若 $z=f(x,y,z)$ 在区域 D 内可微, 则它在 D 内任意一点 (x,y,z) 处的全微分可写成

$$\mathrm{d}z=f_x(x,y,z)\Delta x+f_y(x,y,z)\Delta y+f_z(x,y,z)\Delta z.$$

与一元函数的微分类似, 三元函数自变量的全微分就等于其各自的增量, 即 $\mathrm{d}x \equiv \Delta x$, $\mathrm{d}y \equiv \Delta y$, $\mathrm{d}z \equiv \Delta z$. 这样, 全微分公式可写成

$$\mathrm{d}z = f_x(x,y,z)\mathrm{d}x + f_y(x,y,z)\mathrm{d}y + f_z(x,y,z)\mathrm{d}z,$$

或

$$\mathrm{d}f = \frac{\partial f}{\partial x}\mathrm{d}x + \frac{\partial f}{\partial y}\mathrm{d}y + \frac{\partial f}{\partial z}\mathrm{d}z.$$

因此, 定理不仅证明了三元函数可微的必要条件是偏导数存在, 而且还给出了计算全微分的公式.

可微的充分条件: 偏导数连续 $\Rightarrow$ 可微 若函数 $z = f(x,y,z)$ 的偏导数 $f_x(x,y,z)$, $f_y(x,y,z)$ 和 $f_z(x,y,z)$ 在点 (x_0,y_0,z_0) 的某个邻域内存在, 并且这三个偏导数在点 (x_0,y_0,z_0) 处连续, 则 $z = f(x,y,z)$ 在点 (x_0,y_0,z_0) 处可微.

将单个点的结论应用于区域内所有点, 则得到以下推论: 若 D 是 $\mathbf{R}^3$ 中的一个区域, 而 $f(x,y,z) \in C^1(D)$, 即 $f(x,y,z)$ 在 D 中有连续的一阶偏导数 $f_x(x,y,z)$, $f_y(x,y,z)$ 和 $f_z(x,y,z)$, 则 $f(x,y,z)$ 在 D 内可微.

6.5 复合函数微分法 • 一阶全微分的形式不变性与高阶微分

6.5.1 复合函数微分法

在讨论多元复合函数的极限时, 我们已经知道多元函数的复合有多种多样的形式. 现在我们仅就一种较典型的复合形式 (2-2-1 型[15]), 给出求偏导数的公式定理的证明, 其他情况可以类推得到结果.

2-2-1 型

定理 6.5.1: 多元函数复合的链式法则 (Chain rule) (2-2-1 型)

如果

- 函数 $u = \varphi(x,y)$ 和 $v = \psi(x,y)$ 关于 x 与 y 的偏导数存在;
- 函数 $z = f(u,v)$ 关于 u 与 v 的偏导数存在, 且连续;

那么, 复合函数 $z = f(\varphi(x,y),\psi(x,y))$ 关于 x 与 y 的偏导数也存在, 并且有计算公式

$$\frac{\partial z}{\partial x} = \frac{\partial f}{\partial u}\cdot\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\cdot\frac{\partial v}{\partial x},$$
$$\frac{\partial z}{\partial y} = \frac{\partial f}{\partial u}\cdot\frac{\partial u}{\partial y} + \frac{\partial f}{\partial v}\cdot\frac{\partial v}{\partial y}.$$

这两个公式称为求复合函数偏导数的**链式法则** (Chain rule).

如何记住链式法则公式呢? 可以按以下方法一步步观察:

1. $z = f(\underset{\blacktriangle}{u},v)$ 且 $u = \varphi(\underset{\blacktriangle}{x},y) \to$ 写出 $\dfrac{\partial f}{\partial u}\cdot\dfrac{\partial u}{\partial x}$
2. $z = f(u,\underset{\blacktriangle}{v})$ 且 $v = \psi(\underset{\blacktriangle}{x},y) \to$ 写出 $\dfrac{\partial f}{\partial v}\cdot\dfrac{\partial v}{\partial x}$

[15]这种说法是作者本人定义的, 不是通用的叫法. 说一个复合函数是 “n-m-1 型”, 是指自变量的复合路径是 $\mathbf{R}^n \to \mathbf{R}^m \to \mathbf{R}$.

然后把所有项求和就是最终的 $\dfrac{\partial z}{\partial x}$. 同理, $\dfrac{\partial z}{\partial y}$ 也是一样. 也可以参考图 6.10 中变量之间的关系.

证明. 我们只证明第一个公式

$$\frac{\partial z}{\partial x}=\frac{\partial f}{\partial u}\cdot\frac{\partial u}{\partial x}+\frac{\partial f}{\partial v}\cdot\frac{\partial v}{\partial x}.$$

类似的讨论可以证明另一个公式. 设 (x,y) 为给定的点, (u,v) 为其相应的点, 即 $u=\varphi(x,y), v=\psi(x,y)$. 这里我们强调 u,v 是整个复合函数 $z=f(\varphi(x,y),\psi(x,y))$ 的中间变量.

由于 $z=f(u,v)$ 关于 u 与 v 的偏导数存在且连续, 所以由上一节的定理 (偏导数连续 $\Rightarrow$ 可微) 可知 $z=f(u,v)$ 在点 (u,v) 处可微, 故我们有关于全增量的极限等式

$$\Delta z=f(u+\Delta u,v+\Delta v)-f(u,v)=\frac{\partial f}{\partial u}\Delta u+\frac{\partial f}{\partial v}\Delta v+o(\rho),\quad(\rho\to0),$$

其中 $\rho=\sqrt{(\Delta u)^2+(\Delta v)^2}$. 根据高阶无穷小的定义, 可以将 $o(\rho)$ 表示为

$$\alpha\cdot\rho=\alpha\cdot\sqrt{(\Delta u)^2+(\Delta v)^2},$$

其中, 无穷小量 $\alpha=\alpha(\Delta u,\Delta v)\to0$ $(\rho\to0)$. 注意到 $\rho\to0$ 当且仅当 $\Delta u\to0,\Delta v\to0$. 因此, 可以将全增量的极限等式改写成

$$\Delta z=\frac{\partial f}{\partial u}\Delta u+\frac{\partial f}{\partial v}\Delta v+\alpha\sqrt{(\Delta u)^2+(\Delta v)^2},\quad(\Delta u\to0,\Delta v\to0),\tag{6.2}$$

注意到上述关系对于一切可能的中间变量的增量 Δu 及 Δv 均成立, 只要最终可以使得 $\Delta u\to0,\Delta v\to0$. 特别地, 对于仅仅由单个自变量 x 的增量 Δx 引起的中间变量的增量

$$\begin{aligned}\Delta u&=\varphi(x+\Delta x,y)-\varphi(x,y)=\varphi(x+\Delta x,y)-u,\\ \Delta v&=\psi(x+\Delta x,y)-\psi(x,y)=\psi(x+\Delta x,y)-v,\end{aligned}$$

也是成立的. 这种增量称为**偏增量** (因为我们并没有考虑自变量 y 对应的增量)[16]. 当 Δu 与 Δv 是这种偏增量时, 它们作为 Δx 的函数 (即固定住 x,y) 满足

$$\Delta u\to0,\Delta v\to0,\quad(\Delta x\to0),$$

从而

$$\alpha(\Delta u,\Delta v)\to0,\quad(\Delta x\to0).$$

而且, 因为

$$\begin{aligned}\Delta u+u&=\varphi(x+\Delta x,y),\\ \Delta v+v&=\psi(x+\Delta x,y),\end{aligned}$$

这时,

$$\begin{aligned}\Delta z&=f(u+\Delta u,v+\Delta v)-f(u,v)\\&=f(\varphi(x+\Delta x,y),\psi(x+\Delta x,y))-f(\varphi(x,y),\psi(x,y))\end{aligned}$$

正好就是复合函数 $z=f(\varphi(x,y),\psi(x,y))$ 由 Δx 引起的偏增量. 因此, 根据偏导数的定义, 复合函数关于 x 的偏导数 $\dfrac{\partial z}{\partial x}$ 就等于极限

$$\lim_{\Delta x\to0}\frac{\Delta z}{\Delta x}=\lim_{\Delta x\to0}\frac{f(\varphi(x+\Delta x,y),\psi(x+\Delta x,y))-f(\varphi(x,y),\psi(x,y))}{\Delta x}.$$

[16] 偏增量不是新鲜东西, 早在最开始偏导数的定义时, 极限的分子就是偏增量. $\dfrac{\partial f}{\partial x}(x_0,y_0)=\dfrac{\mathrm{d}}{\mathrm{d}x}f(x,y_0)\Big|_{x=x_0}=\lim\limits_{\Delta x\to0}\dfrac{f(x_0+\Delta x,y_0)-f(x_0,y_0)}{\Delta x}$.

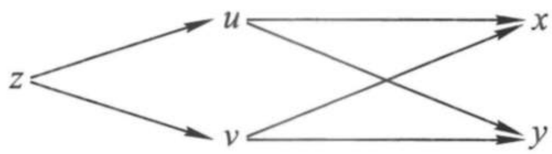


图 6.10: 定理 6.5.1 中 2-2-1 型函数复合示意图.

下面我们要求出极限 $\lim\limits_{\Delta x\to 0}\dfrac{\Delta z}{\Delta x}$ 的表达式.

由 (6.2) 式可得

$$\frac{\Delta z}{\Delta x}=\frac{\partial f}{\partial u}\cdot\frac{\Delta u}{\Delta x}+\frac{\partial f}{\partial v}\cdot\frac{\Delta v}{\Delta x}+\alpha\frac{\sqrt{(\Delta u)^2+(\Delta v)^2}}{\Delta x},$$

并且当 $\Delta x\to 0$ 时, 有

$$\frac{\Delta u}{\Delta x}\to\frac{\partial u}{\partial x},\quad \frac{\Delta v}{\Delta x}\to\frac{\partial v}{\partial x}.$$

所以, 为了证明链式法则中的第一个公式, 剩下的工作就只需证明

$$\alpha\frac{\sqrt{(\Delta u)^2+(\Delta v)^2}}{\Delta x}\to 0,\quad (\Delta x\to 0).$$

事实上, 其绝对值满足

$$\left|\alpha\frac{\sqrt{(\Delta u)^2+(\Delta v)^2}}{\Delta x}\right|=\left|\alpha\sqrt{\left(\frac{\Delta u}{\Delta x}\right)^2+\left(\frac{\Delta v}{\Delta x}\right)^2}\right|\to 0\cdot\sqrt{\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial v}{\partial x}\right)^2}=0,\quad (\Delta x\to 0),$$

故其本身当 $\Delta x\to 0$ 时也趋向于零. 总之, 我们证明了

$$\frac{\partial z}{\partial x}=\frac{\partial f}{\partial u}\cdot\frac{\partial u}{\partial x}+\frac{\partial f}{\partial v}\cdot\frac{\partial v}{\partial x}.$$

证毕. □

在许多问题中, 常常需要引入变量替换. 链式法则的意义在于告诉我们如何在变量替换后计算偏导数. 在使用链式法则时, 在不致混淆的条件下, $\dfrac{\partial f}{\partial u}$ 及 $\dfrac{\partial f}{\partial v}$ 也可以分别用 $\dfrac{\partial z}{\partial u}$ 及 $\dfrac{\partial z}{\partial v}$ 表示. 例如下题, 我们没有使用函数句柄 f 而是用输出值的符号 z 来表示函数 $z=z(u,v)$.

例 6.5.1 设函数 $z=\mathrm{e}^{xy}\sin(x+y)$, 求 $\dfrac{\partial z}{\partial x}$ 及 $\dfrac{\partial z}{\partial y}$.

证明. 令 $u=xy, v=x+y$, 这时 $z=\mathrm{e}^u\sin v$. 根据链式法则, 我们有

$$\begin{aligned}
\frac{\partial z}{\partial x}&=\frac{\partial z}{\partial u}\cdot\frac{\partial u}{\partial x}+\frac{\partial z}{\partial v}\cdot\frac{\partial v}{\partial x}=\mathrm{e}^u\sin v\cdot y+\mathrm{e}^u\cos v\cdot 1=\mathrm{e}^{xy}(y\sin(x+y)+\cos(x+y)),\\
\frac{\partial z}{\partial y}&=\frac{\partial z}{\partial u}\cdot\frac{\partial u}{\partial y}+\frac{\partial z}{\partial v}\cdot\frac{\partial v}{\partial y}=\mathrm{e}^u\sin v\cdot x+\mathrm{e}^u\cos v\cdot 1=\mathrm{e}^{xy}(x\sin(x+y)+\cos(x+y)).
\end{aligned}$$

□

关于上一题的一些讨论:

- 和一元函数的链式法则一样, 最终我们的结果不应该包含中间变量的符号. 要将 $u=xy, v=x+y$ 代进去, 最后形成只关于 x,y 的表达式.

- 另外, 上一题只是一个示范的目的, 我们可以不用设置中间变量, 而直接求其偏导, 其结果是一样的.

- 还有, 我们发现函数 $z = \mathrm{e}^{xy}\sin(x+y)$ 关于起自变量是对称的, 所以 $\dfrac{\partial z}{\partial y}$ 可以直接由 $\dfrac{\partial z}{\partial x}$ 将其两个变量互换位置得到, 这是多元函数求偏导常用的技巧.

- 当我们练习足够时, 可以不用显式地写出公式

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial u}\cdot\frac{\partial u}{\partial x} + \frac{\partial z}{\partial v}\cdot\frac{\partial v}{\partial x}, \quad \frac{\partial z}{\partial y} = \frac{\partial z}{\partial u}\cdot\frac{\partial u}{\partial y} + \frac{\partial z}{\partial v}\cdot\frac{\partial v}{\partial y}.$$

心中清楚即可. 比如下一道题.

例 6.5.2 设函数 $z = f(u,v) = v\ln u$, 其中 $u = x^2+y^2, v = \dfrac{y}{x}$, 求 $\dfrac{\partial z}{\partial x}$ 及 $\dfrac{\partial z}{\partial y}$.

证明. 根据链式法则, 我们有

$$\begin{aligned}
\frac{\partial z}{\partial x} &= \frac{v}{u}\cdot 2x - \ln u\cdot\frac{y}{x^2} = \frac{2y}{(x^2+y^2)} - \frac{y}{x^2}\ln\left(x^2+y^2\right), \\
\frac{\partial z}{\partial y} &= \frac{v}{u}\cdot 2y + \ln u\cdot\frac{1}{x} = \frac{2y^2}{x\left(x^2+y^2\right)} + \frac{1}{x}\ln\left(x^2+y^2\right).
\end{aligned}$$

□

很多时候我们并不知道复合函数中每一个环节上所有函数的具体表达式, 或者说, 我们不在意它的表达式是什么. 此时链式法则仍然可以适用: 如果链式法则的公式中不清楚具体的偏导表达式, 那么就用通常的偏导符号; 如果知道其具体的偏导表达式, 就直接代入. 比如下一题.

例 6.5.3 设函数 $z = f(x,y)$ 有连续的一阶偏导数, 且 $x = r\cos\theta, y = r\sin\theta$, 求 $\dfrac{\partial z}{\partial r}$ 及 $\dfrac{\partial z}{\partial \theta}$, 并证明:[17]

$$\left(\frac{\partial z}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial z}{\partial \theta}\right)^2 = \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2.$$

证明. 由链式法则, 我们得到

$$\begin{aligned}
\frac{\partial z}{\partial r} &= \frac{\partial z}{\partial x}\cdot\frac{\partial x}{\partial r} + \frac{\partial z}{\partial y}\cdot\frac{\partial y}{\partial r} = \cos\theta\frac{\partial z}{\partial x} + \sin\theta\frac{\partial z}{\partial y}, \\
\frac{\partial z}{\partial \theta} &= \frac{\partial z}{\partial x}\cdot\frac{\partial x}{\partial \theta} + \frac{\partial z}{\partial y}\cdot\frac{\partial y}{\partial \theta} = -r\sin\theta\frac{\partial z}{\partial x} + r\cos\theta\frac{\partial z}{\partial y}.
\end{aligned}$$

[17] 我们应当清楚, $\dfrac{\partial z}{\partial r}, \dfrac{\partial z}{\partial \theta}$ 中的 z 是关于 r,θ 的复合函数 $z = f(r\cos\theta, r\sin\theta)$ 的偏导, 而 $\dfrac{\partial z}{\partial x}, \dfrac{\partial z}{\partial y}$ 中的 z 是关于 x,y 的函数 $z = f(x,y)$ 的偏导. 这种歧义在多元函数求偏导中经常遇到. 只要我们自己清楚即可.

所以

$$
\begin{aligned}
\left(\frac{\partial z}{\partial r}\right)^2+\left(\frac{1}{r}\cdot\frac{\partial z}{\partial \theta}\right)^2 &= \left(\cos\theta\frac{\partial z}{\partial x}+\sin\theta\frac{\partial z}{\partial y}\right)^2+\left(-\sin\theta\frac{\partial z}{\partial x}+\cos\theta\frac{\partial z}{\partial y}\right)^2 \\
&= \left[\cos^2\theta\left(\frac{\partial z}{\partial x}\right)^2+2\cos\theta\sin\theta\frac{\partial z}{\partial x}\frac{\partial z}{\partial y}+\sin^2\theta\left(\frac{\partial z}{\partial y}\right)^2\right] \\
&\quad+\left[\sin^2\theta\left(\frac{\partial z}{\partial x}\right)^2-2\sin\theta\cos\theta\frac{\partial z}{\partial x}\frac{\partial z}{\partial y}+\cos^2\theta\left(\frac{\partial z}{\partial y}\right)^2\right] \\
&= \left(\frac{\partial z}{\partial x}\right)^2+\left(\frac{\partial z}{\partial y}\right)^2.
\end{aligned}
$$

□

在上述讨论中, 中间变量 u, v 及自变量 x, y 的个数都是两个. 其实, 链式法则不限于这种情况. 有各种可能性, 链式法则也随之有各种相应的形式.

1-2-1 型

考虑 $z=f(u,v)$, 而 u 及 v 均是 x 的一元函数的情况, 即 1-2-1 型,

$$z=f(u,v),\quad u=\varphi(x),\quad v=\psi(x).$$

复合函数 $z=f(\varphi(x),\psi(x))$ 是 x 的一元函数. 这时, 链式法则是

$$\frac{\mathrm{d}z}{\mathrm{d}x}=\frac{\partial f}{\partial u}\cdot\frac{\mathrm{d}u}{\mathrm{d}x}+\frac{\partial f}{\partial v}\cdot\frac{\mathrm{d}v}{\mathrm{d}x}=\frac{\partial f}{\partial u}\varphi'(x)+\frac{\partial f}{\partial v}\psi'(x).$$

这个公式成立的条件是 $f(u,v)$ 可微, 而 $\varphi'(x)$ 及 $\psi'(x)$ 存在. 可以按以下方法一步步观察:

1. $z=f(\underset{\blacktriangle}{u},v)$, 且 $u=\varphi(\underset{\blacktriangle}{x})\to$ 写出 $\dfrac{\partial f}{\partial u}\cdot\dfrac{\mathrm{d}u}{\mathrm{d}x}$
2. $z=f(u,\underset{\blacktriangle}{v})$, 且 $v=\psi(\underset{\blacktriangle}{x})\to$ 写出 $\dfrac{\partial f}{\partial v}\cdot\dfrac{\mathrm{d}v}{\mathrm{d}x}$

然后把所有项求和就是最终的 $\dfrac{\mathrm{d}z}{\mathrm{d}x}$.

例 6.5.4 设函数 $u=\phi(x)$ 及 $v=\psi(x)$ 在区间 (α,β) 内可微, 且 $\phi(x), \psi(x)>0$, 求 $\dfrac{\mathrm{d}}{\mathrm{d}x}\phi(x)^{\psi(x)}$.

证明. 设 $f(u,v)=u^v\ (u,v>0)$, 则

$$f_u(u,v)=vu^{v-1},\quad f_v(u,v)=u^v\ln u.$$

由于 $\phi(x)^{\psi(x)}=f(\phi(x),\psi(x))$, 所以根据链式法则可得

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\phi(x)^{\psi(x)} &= vu^{v-1}\phi'(x)+u^v\ln u\cdot\psi'(x) \\
&= \psi(x)\phi(x)^{\psi(x)-1}\phi'(x)+\phi(x)^{\psi(x)}\ln\phi(x)\cdot\psi'(x)
\end{aligned}
$$

上面不要忘记最后变成关于 x 的表达式. 另外, 按照一元复合函数的求导方法, 则可先将 $\phi(x)^{\psi(x)}$ 写成 $\mathrm{e}^{\psi(x)\ln\phi(x)}$, 再求导数. 这种方法所得结果是一致的. □

1-3-1 型

考虑 $z=f(u,v,w)$, 其中 u,v,w 都是 x 的一元函数, 即

$$u=\varphi(x),\quad v=\psi(x),\quad w=\xi(x).$$

因此, 复合函数 $z=f(\varphi(x),\psi(x),\xi(x))$ 是 x 的一元函数. 根据链式法则, 我们可以求 z 对 x 的导数. 这时, 链式法则是

$$\frac{\mathrm{d}z}{\mathrm{d}x}=\frac{\partial f}{\partial u}\cdot\frac{\mathrm{d}u}{\mathrm{d}x}+\frac{\partial f}{\partial v}\cdot\frac{\mathrm{d}v}{\mathrm{d}x}+\frac{\partial f}{\partial w}\cdot\frac{\mathrm{d}w}{\mathrm{d}x}=\frac{\partial f}{\partial u}\cdot\varphi'(x)+\frac{\partial f}{\partial v}\cdot\psi'(x)+\frac{\partial f}{\partial w}\cdot\xi'(x).$$

其中 $f(u,v,w)$ 必须是可微的, 并且 $\varphi'(x),\psi'(x),\xi'(x)$ 存在. 可以按以下方法一步步观察:

1. $z=f(\underset{\blacktriangle}{u},v,w)$, 且 $u=\varphi(\underset{\blacktriangle}{x})\to$ 写出 $\dfrac{\partial f}{\partial u}\cdot\dfrac{\mathrm{d}u}{\mathrm{d}x}$
2. $z=f(u,\underset{\blacktriangle}{v},w)$, 且 $v=\psi(\underset{\blacktriangle}{x})\to$ 写出 $\dfrac{\partial f}{\partial v}\cdot\dfrac{\mathrm{d}v}{\mathrm{d}x}$
3. $z=f(u,v,\underset{\blacktriangle}{w})$, 且 $w=\xi(\underset{\blacktriangle}{x})\to$ 写出 $\dfrac{\partial f}{\partial w}\cdot\dfrac{\mathrm{d}w}{\mathrm{d}x}$

然后把所有项求和就是最终的 $\dfrac{\mathrm{d}z}{\mathrm{d}x}$.

例 6.5.5 设函数 $z=uv+vw+uw$, 其中 $u=x^2,v=1-x^2,w=1-x$, 求 $\dfrac{\mathrm{d}z}{\mathrm{d}x}$.

证明. 根据链式法则, 我们有

$$\frac{\mathrm{d}z}{\mathrm{d}x}=(v+w)\cdot 2x+(u+w)\cdot(-2x)+(v+u)\cdot(-1)=-4x^3+2x-1.$$

□

2-3-1 型

考虑 $z=f(u,v,w)$, 而 u,v 及 w 均是 (x,y) 的二元函数的情况, 即 2-3-1 型,

$$u=u(x,y),\quad v=v(x,y),\quad w=w(x,y).$$

复合函数 $z=f(u(x,y),v(x,y),w(x,y))$ 是 (x,y) 的二元函数. 这时, 链式法则是

$$\frac{\partial z}{\partial x}=\frac{\partial f}{\partial u}\cdot\frac{\partial u}{\partial x}+\frac{\partial f}{\partial v}\cdot\frac{\partial v}{\partial x}+\frac{\partial f}{\partial w}\cdot\frac{\partial w}{\partial x},$$
$$\frac{\partial z}{\partial y}=\frac{\partial f}{\partial u}\cdot\frac{\partial u}{\partial y}+\frac{\partial f}{\partial v}\cdot\frac{\partial v}{\partial y}+\frac{\partial f}{\partial w}\cdot\frac{\partial w}{\partial y}.$$

同样, 使用这种链式法则要求 $f(u,v,w)$ 的可微性以及中间变量 (作为自变量的函数) 的偏导数的存在性.[18] 可以按以下方法一步步观察:

1. $z=f(\underset{\blacktriangle}{u}(\underset{\blacktriangle}{x},y),v(x,y),w(x,y))\to$ 写出 $\dfrac{\partial f}{\partial u}\cdot\dfrac{\partial u}{\partial x}$
2. $z=f(u(x,y),\underset{\blacktriangle}{v}(\underset{\blacktriangle}{x},y),w(x,y))\to$ 写出 $\dfrac{\partial f}{\partial v}\cdot\dfrac{\partial v}{\partial x}$

[18] 以我们高数 B 的难度, 不需要在意这些细节, 本节的内容重点是应用链式法则. 大胆地用就行了.

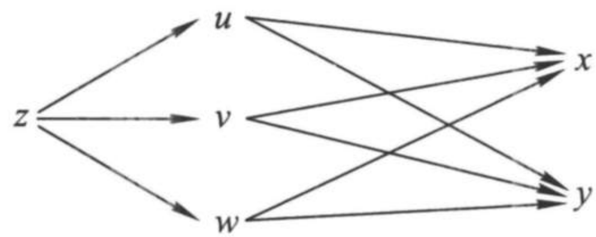


图 6.11: 2-3-1 型函数复合示意图.

3. $z=f(u(x,y),v(x,y),w(x,y)) \to$ 写出 $\dfrac{\partial f}{\partial w}\cdot\dfrac{\partial w}{\partial x}$

然后把所有项求和就是最终的 $\dfrac{\partial z}{\partial x}$. 同理, $\dfrac{\partial z}{\partial y}$ 也是一样. 可用图解的方式表示这种链式法则中变量之间的关系, 如图 6.11 所示.

例 6.5.6 设函数 $z=f(u,v,w)$, 其中 $u=\mathrm{e}^x, v=xy, w=y\sin x$, 求 $\dfrac{\partial z}{\partial x}$ 及 $\dfrac{\partial z}{\partial y}$.

证明. 根据链式法则, 我们有

$$\begin{aligned}\frac{\partial z}{\partial x}&=\frac{\partial f}{\partial u}\mathrm{e}^x+\frac{\partial f}{\partial v}y+\frac{\partial f}{\partial w}y\cos x,\\ \frac{\partial z}{\partial y}&=\frac{\partial f}{\partial u}\cdot 0+\frac{\partial f}{\partial v}x+\frac{\partial f}{\partial w}\sin x=\frac{\partial f}{\partial v}x+\frac{\partial f}{\partial w}\sin x\end{aligned}$$

□

n-m-1 型 (n,m ≥ 1)

读者很容易把链式法则推广到任意多个 (m 个) 中间变量以及任意多个 (n 个) 自变量的情况. 一般来说, 对于复合函数

$$z=f\left(u_1(x_1,\ldots,x_n),u_2(x_1,\ldots,x_n),\ldots,u_m(x_1,\ldots,x_n)\right),$$

其中 z 是 m 个中间变量 $u_1,u_2,\ldots,u_m$ 的函数, 而每个中间变量 u_i 都是 n 个自变量 $x_1,\ldots,x_n$ 的函数. 我们可以利用链式法则来求 z 对每个自变量 x_j $(j=1,\ldots,n)$ 的偏导数. 使用条件是:

1. f 对 $u_1,u_2,\ldots,u_m$ 都是可微的;
2. 每个 u_i 都关于 $x_1,\ldots,x_n$ 的偏导数都存在.

那么. 对于任意 $j\in\{1,2,\ldots,n\}$, 复合函数 z 对 x_j 的偏导数为:

$$\frac{\partial z}{\partial x_j}=\sum_{i=1}^{m}\frac{\partial f}{\partial u_i}\cdot\frac{\partial u_i}{\partial x_j}.$$

对于每个 x_j, 首先计算 f 对每个中间变量 u_i 的偏导数, 再乘以该中间变量对 x_j 的偏导数, 最后对所有中间变量 i 求和. 这样, 就得到了复合函数的链式法则的一般形式. 注意, 当 n=1 时, 偏导数会退化成导数. 上述公式涵盖了前面讨论的所有类型.

特殊情况

还有一种情况也值得提及: 比如, $z=f(x,y,w)$, 而 w 又是 (x,y) 的函数, 即 $w=w(x,y)$. 重点是, 这里 x,y 既是中间变量, 又是自变量. 此时, 这种类型本质上还是 2-3-1 型, 因为我们可以把 x,y 视作两个具体的表达式, 它们都是关于二元变量 (x,y) 的函数. 显然, 二元函数

$$(x,y)\mapsto x$$

关于自变量 x 的偏导恒为 1, 关于自变量 y 的偏导恒为 0. 二元函数 $(x,y)\mapsto y$ 的偏导也有类似结果. 在求复合函数

$$z=f(x,y,w(x,y))$$

作为 (x,y) 的函数的偏导数时, 仍然可用链式法则, 即[19]

$$\begin{aligned}\frac{\partial z}{\partial x}&=\frac{\partial f}{\partial x}\cdot 1+\frac{\partial f}{\partial y}\cdot 0+\frac{\partial f}{\partial w}\cdot\frac{\partial w}{\partial x}=\frac{\partial f}{\partial x}+\frac{\partial f}{\partial w}\cdot\frac{\partial w}{\partial x}\\ \frac{\partial z}{\partial y}&=\frac{\partial f}{\partial x}\cdot 0+\frac{\partial f}{\partial y}\cdot 1+\frac{\partial f}{\partial w}\cdot\frac{\partial w}{\partial y}=\frac{\partial f}{\partial y}+\frac{\partial f}{\partial w}\cdot\frac{\partial w}{\partial y}.\end{aligned}$$

另一途径是, 无论是 x,y 是中间变量, 还是自变量, 我们可以按以下方法一步步观察:

1. $z=f(\underset{\blacktriangle}{x},y,w(x,y))\to$ 直接写出 $\dfrac{\partial f}{\partial x}$
2. $z=f(x,\underset{\blacktriangle}{y},w(x,y))\to$ 和 x 一点关系都没有, 直接跳过
3. $z=f(x,y,\underset{\blacktriangle}{w}(\underset{\blacktriangle}{x},y))\to$ 写出 $\dfrac{\partial f}{\partial w}\cdot\dfrac{\partial w}{\partial x}$

然后把所有项求和就是最终的 $\dfrac{\partial z}{\partial x}$. 同理, $\dfrac{\partial z}{\partial y}$ 也是一样.

例 6.5.7 设函数 $z=f(x,y,w)$ 在 $\mathbf{R}^3$ 中有连续的一阶偏导数, 求复合函数 $z=f\left(x,y,x^2y\right)$ 的偏导数 $\dfrac{\partial z}{\partial x}$ 及 $\dfrac{\partial z}{\partial y}$.

证明. 令 $w=x^2y$, 则

$$\frac{\partial z}{\partial x}=\frac{\partial f}{\partial x}+\frac{\partial f}{\partial w}\cdot 2xy,\quad \frac{\partial z}{\partial y}=\frac{\partial f}{\partial y}+\frac{\partial f}{\partial w}x^2.$$

□

在有些书中约定用 f_1',f_2',f_3' 分别表示 f 关于其第一, 二, 三个变量的偏导数; 用 f_{12}'' 表示先对其第一个变量求偏导数, 然后对其第二个变量求偏导数; 以此类推. 这种约定有其方便之处, 特别是在题目中没有给出 f 的变量名称的时候, 例如 $z=f\left(x^2,xy,x^2+y^2\right)$, 这时

$$\begin{aligned}\frac{\partial z}{\partial x}&=f_1'\cdot 2x+f_2'\cdot y+f_3'\cdot 2x=2x\left(f_1'+f_3'\right)+yf_2',\\ \frac{\partial z}{\partial y}&=f_1'\cdot 0+f_2'\cdot x+f_3'\cdot 2y=xf_2'+2yf_3'.\end{aligned}$$

[19] 这里应当指出: $\dfrac{\partial f}{\partial x}$ 及 $\dfrac{\partial f}{\partial y}$ 是指 f 作为 (x,y,w) 的函数时的偏导数, 即 f 分别关于其第一个变量 x 及第二个变量 y 的偏导数, 而 $\dfrac{\partial z}{\partial x}$ 及 $\dfrac{\partial z}{\partial y}$ 则是复合函数 $z=f(x,y,w(x,y))$ 分别关于自变量 x 及 y 的偏导数. 在这种情况下, $\dfrac{\partial z}{\partial x}$ 就不能再写成 $\dfrac{\partial f}{\partial x}$, 否则会引起混淆.

6.5.2 一阶全微分的形式不变性

与一元函数的一阶微分的形式不变性类似, 多元函数也有一阶全微分的形式不变性.

定理 6.5.2: 一阶全微分的形式不变性

已知当 u,v 是自变量时, 函数 $z=f(u,v)$ 在点 (u,v) 处的全微分表示为

$$\mathrm{d}z=f_u\ \mathrm{d}u+f_v\ \mathrm{d}v,$$

其中 $\mathrm{d}u$, $\mathrm{d}v$ 分别是自变量 u,v 的增量.
此时, 若假设函数 $z=f(u,v)$, $u=u(x,y)$, $v=v(x,y)$ 都有连续的一阶偏导数, 则复合函数 $z=f(u(x,y),v(x,y))$ 在点 (x,y) 处的全微分仍然可表示为

$$\mathrm{d}z=f_u\ \mathrm{d}u+f_v\ \mathrm{d}v,$$

其中 $\mathrm{d}u\equiv\mathrm{d}u(\mathrm{d}x,\mathrm{d}y)$, $\mathrm{d}v\equiv\mathrm{d}v(\mathrm{d}x,\mathrm{d}y)$ 分别是 $u=u(x,y),v=v(x,y)$ 在点 (x,y) 处的微分.[a]

[a]因此他们关于 $\mathrm{d}x,\mathrm{d}y$ 的线性函数, 而非自变量增量.

上述定理表明, 不论 u,v 是中间变量, 还是自变量, 全微分 $\mathrm{d}z$ 都可表示成相同的形式. 这就是所谓的一阶全微分的形式不变性.

证明. 由所给条件及链式法则可知, 复合函数 $z=f(u(x,y),v(x,y))$ 在点 (x,y) 处有连续的一阶偏导数, 由上一节的定理 (偏导数连续 $\Rightarrow$ 可微), 因而复合函数在点 (x,y) 处可微, 并且有

$$\begin{aligned}\mathrm{d}z=\frac{\partial z}{\partial x}\ \mathrm{d}x+\frac{\partial z}{\partial y}\ \mathrm{d}y&=\left(f_u\frac{\partial u}{\partial x}+f_v\frac{\partial v}{\partial x}\right)\ \mathrm{d}x+\left(f_u\frac{\partial u}{\partial y}+f_v\frac{\partial v}{\partial y}\right)\ \mathrm{d}y\\&=f_u\left(\frac{\partial u}{\partial x}\ \mathrm{d}x+\frac{\partial u}{\partial y}\ \mathrm{d}y\right)+f_v\left(\frac{\partial v}{\partial x}\ \mathrm{d}x+\frac{\partial v}{\partial y}\ \mathrm{d}y\right)\\&=f_u\ \mathrm{d}u+f_v\ \mathrm{d}v.\end{aligned}$$

证毕. □

作为一阶全微分的形式不变性的应用, 我们给出下列公式.

结论 6.5.1: 全微分的运算法则

设 $u=u(x,y),v=v(x,y)$ 是可微二元函数, $f=f(x)$ 是一元可微函数, 则有

1. $\mathrm{d}(u\pm v)=\mathrm{d}u\pm\mathrm{d}v$;
2. $\mathrm{d}(cu)=c\ \mathrm{d}u$ (c 为常数);
3. $\mathrm{d}(uv)=v\ \mathrm{d}u+u\ \mathrm{d}v$;
4. $\mathrm{d}(\frac{u}{v})=\dfrac{v\ \mathrm{d}u-u\ \mathrm{d}v}{v^2}$ $(v\neq 0)$;
5. $\mathrm{d}(f(u))=f'(u)\ \mathrm{d}u$.

上述结论中的公式的证明是容易的. 事实上, 在定理 6.5.2 中, 分别令 $z = f(u,v) = u \pm v, cu, uv, \dfrac{u}{v}$ 或 $f(u)$, 就能立即推出上述公式. 一阶全微分的形式不变性及上述全微分公式为全微分的计算提供了一种新途径. 见下题.

例 6.5.8 设 $z = (x-y)^{x^2+y^2}$ $(x>y)$, 求全微分 $\mathrm{d}z$.

证明. 令 $u = x-y, v = x^2+y^2$, 则 $z = u^v$. 我们可以考虑 $z = z(u,v) = u^v$, 由全微分的形式不变性得

$$\mathrm{d}z = \mathrm{d}\left(u^v\right) = \frac{\partial}{\partial u}\left(u^v\right)\mathrm{d}u + \frac{\partial}{\partial v}\left(u^v\right)\mathrm{d}v = vu^{v-1}\ \mathrm{d}u + u^v \ln u\ \mathrm{d}v,$$

其中, $\mathrm{d}u = \mathrm{d}(x-y) = \mathrm{d}x - \mathrm{d}y$, $\mathrm{d}v = \mathrm{d}\left(x^2+y^2\right) = 2x\ \mathrm{d}x + 2y\ \mathrm{d}y$. 代回并整理得

$$\mathrm{d}z = (x-y)^{x^2+y^2}\left[\left(2x\ln(x-y) + \frac{x^2+y^2}{x-y}\right)\mathrm{d}x + \left(2y\ln(x-y) - \frac{x^2+y^2}{x-y}\right)\mathrm{d}y\right].$$

□

上一道题中, 可以从最后的全微分中直接读出

$$\begin{aligned}
\frac{\partial z}{\partial x} &= (x-y)^{x^2+y^2}\left(2x\ln(x-y) + \frac{x^2+y^2}{x-y}\right),\\
\frac{\partial z}{\partial y} &= (x-y)^{x^2+y^2}\left(2y\ln(x-y) - \frac{x^2+y^2}{x-y}\right).
\end{aligned}$$

于是一阶全微分的形式不变形提供了另一种求偏导数的方法. 对于三元函数 $u = u(x,y,z), v = v(x,y,z)$, 本小节所有结论 (特别是结论 6.5.1) 也都成立. 见下题.

例 6.5.9 设函数 $u = \sin\left(x^2+y^2\right) + \mathrm{e}^{xz}$, 求该函数在点 $(1,0,1)$ 处的全微分 $\mathrm{d}u$.

证明. 由一阶全微分的形式不变性和结论 6.5.1 中的公式, 我们有

$$\begin{aligned}
\mathrm{d}u &= \mathrm{d}\left(\sin\left(x^2+y^2\right) + \mathrm{e}^{xz}\right)\\
&= \mathrm{d}\left(\sin\left(x^2+y^2\right)\right) + \ \mathrm{d}\left(\mathrm{e}^{xz}\right)\\
&= \cos\left(x^2+y^2\right)\ \mathrm{d}\left(x^2+y^2\right) + \mathrm{e}^{xz}\ \mathrm{d}(xz)\\
&= \cos\left(x^2+y^2\right)\left(\mathrm{d}x^2 + \ \mathrm{d}y^2\right) + +\mathrm{e}^{xz}(z\ \mathrm{d}x + x\ \mathrm{d}z)\\
&= \cos\left(x^2+y^2\right)(2x\ \mathrm{d}x + 2y\ \mathrm{d}y) + \mathrm{e}^{xz}(z\ \mathrm{d}x + x\ \mathrm{d}z).
\end{aligned}$$

将 $x=1, y=0, z=1$ 代入上式, 我们得到

$$\mathrm{d}u = \cos 1\cdot 2\ \mathrm{d}x + \mathrm{e}(\mathrm{d}x + \mathrm{d}z) = (2\cos 1 + \mathrm{e})\ \mathrm{d}x + \mathrm{e}\ \mathrm{d}z.$$

□

6.5.3 高阶微分

假设本小节所讨论的函数 $f(x,y)$ 任意阶的混合偏导如果存在的话, 那么都是连续的.

二阶微分 已知函数 $f(x,y)$ 的全微分为

$$\mathrm{d}f = f_x(x,y)\ \mathrm{d}x + f_y(x,y)\ \mathrm{d}y.$$

如果把点 (x,y) 固定, 则它是一个关于 $\mathrm{d}x, \mathrm{d}y$ 的二元线性函数, 我们也可以把它视作是关于 $\mathrm{d}x, \mathrm{d}y$ 的次数皆为 1 的二元多项式函数. 当我们把 $\mathrm{d}x$ 与 $\mathrm{d}y$ 固定时, 上述全微分 $\mathrm{d}f$ 又可以看作是关于 (x,y) 的二元函数, 因为它是 $f_x(x,y), f_y(x,y)$ 的线性组合[20]. 于是当 $f_x(x,y), f_y(x,y)$ 都可微时, 则又可以对 $\mathrm{d}f$ 求全微分, 即

$$\begin{aligned}
\mathrm{d}(\mathrm{d}f) &= \frac{\partial}{\partial x}(\mathrm{d}f)\ \mathrm{d}x + \frac{\partial}{\partial y}(\mathrm{d}f)\ \mathrm{d}y \\
&= \frac{\partial}{\partial x}\left(f_x(x,y)\ \mathrm{d}x + f_y(x,y)\ \mathrm{d}y\right)\ \mathrm{d}x + \frac{\partial}{\partial y}\left(f_x(x,y)\ \mathrm{d}x + f_y(x,y)\ \mathrm{d}y\right)\ \mathrm{d}y \\
&= f_{xx}(x,y)\ \mathrm{d}x^2 + f_{yx}(x,y)\ \mathrm{d}y\ \mathrm{d}x + f_{xy}(x,y)\ \mathrm{d}x\ \mathrm{d}y + f_{yy}(x,y)\ \mathrm{d}y^2, \qquad (6.3)
\end{aligned}$$

这里 $\mathrm{d}x^2 = (\mathrm{d}x)^2,\ \mathrm{d}y^2 = (\mathrm{d}y)^2$.[21] 此时, 我们把 $\mathrm{d}(\mathrm{d}f)$ 称作 $f(x,y)$ 的**二阶微分**, 记作 $\mathrm{d}^2 f$. 此时如果再把点 (x,y) 固定, 我们可以把 $\mathrm{d}^2 f$ 视作是关于 $\mathrm{d}x, \mathrm{d}y$ 的次数皆为 2 的二元多项式函数. 当 $f(x,y)$ 的二阶混合偏导数连续, 则 $f_{xy}(x,y) = f_{yx}(x,y)$. 于是, 我们有如下结果

$$\mathrm{d}^2 f = f_{xx}(x,y)\ \mathrm{d}x^2 + 2f_{xy}(x,y)\ \mathrm{d}x\ \mathrm{d}y + f_{yy}(x,y)\ \mathrm{d}y^2.$$

观察上式发现, 可以将 $\mathrm{d}^2 f$ 写成

$$\mathrm{d}^2 f = \left(\mathrm{d}x\frac{\partial}{\partial x} + \mathrm{d}y\frac{\partial}{\partial y}\right)^2 f,$$

这里 $\mathrm{d}^2 = \left(\mathrm{d}x\frac{\partial}{\partial x} + \mathrm{d}y\frac{\partial}{\partial y}\right)^2$ 可按照牛顿二项式展开[22] 为

$$\mathrm{d}^2 = \ \mathrm{d}x^2\frac{\partial^2}{\partial x^2} + 2\ \mathrm{d}x\ \mathrm{d}y\frac{\partial^2}{\partial x\partial y} + \mathrm{d}y^2\frac{\partial^2}{\partial y^2}.$$

把这个算子作用于二元函数 f 的结果就等于 $\mathrm{d}^2 f$.

三阶微分 同理, 当我们把 $\mathrm{d}x$ 与 $\mathrm{d}y$ 固定时, 式 (6.3) 中的 $\mathrm{d}^2 f$ 是四个二元函数

$$f_{xx}(x,y), f_{yx}(x,y), f_{xy}(x,y), f_{yy}(x,y)$$

的线性组合, 因此 $\mathrm{d}^2 f$ 可视作 (x,y) 的二元函数. 如果这些二阶偏导数又可微, 则 $\mathrm{d}^2 f$ 的微分 $\mathrm{d}\left(\mathrm{d}^2 f\right)$ 便是 $f(x,y)$ 的**三阶微分**, 记作 $\mathrm{d}^3 f$. 具体而言, 给定二阶微分为

$$\mathrm{d}^2 f = f_{xx}(x,y)\,\mathrm{d}x^2 + f_{yx}(x,y)\,\mathrm{d}x\,\mathrm{d}y + f_{xy}(x,y)\,\mathrm{d}x\,\mathrm{d}y + f_{yy}(x,y)\,\mathrm{d}y^2.$$

现对 $\mathrm{d}^2 f$ 进行全微分. 每一项都涉及到不同的微分变量 ($\mathrm{d}x$ 和 $\mathrm{d}y$), 因此需要分别对它们进行微分.

---

[20] n 个量的线性组合是指将这 n 个量与一组常数 (称为系数) 相乘并相加得到的表达式. 具体来说,

$$c_1 a_1 + c_2 a_2 + \cdots + c_n a_n,$$

其中, $c_1, c_2, \ldots, c_n$ 是任意的常数 (系数). 这就是 $a_1, a_2, \ldots, a_n$ 的线性组合.

[21] 这是大家通用的习惯记法. 要注意 $\mathrm{d}x^2$ 不是 x^2 的微分, $\mathrm{d}y^2$ 也不是 y^2 的微分. 这和一元函数的高阶微分符号一致.

[22] 最常用的前几个牛顿二项式展开为:

$$\begin{aligned}
(a+b)^2 &= a^2 + 2ab + b^2, \\
(a+b)^3 &= a^3 + 3a^2 b + 3ab^2 + b^3, \\
(a+b)^4 &= a^4 + 4a^3 b + 6a^2 b^2 + 4ab^3 + b^4.
\end{aligned}$$

1. 对 $f_{xx}(x,y)\,\mathrm{d}x^2$ 求全微分:

$$\mathrm{d}\left(f_{xx}(x,y)\,\mathrm{d}x^2\right)=f_{xxx}(x,y)\,\mathrm{d}x^3+2f_{xxy}(x,y)\,\mathrm{d}x^2\,\mathrm{d}y.$$

2. 对 $f_{yx}(x,y)\,\mathrm{d}y\,\mathrm{d}x$ 求全微分:

$$\mathrm{d}\left(f_{yx}(x,y)\,\mathrm{d}y\,\mathrm{d}x\right)=f_{yxx}(x,y)\,\mathrm{d}x^2\,\mathrm{d}y+f_{yxy}(x,y)\,\mathrm{d}x\,\mathrm{d}y^2.$$

3. 对 $f_{xy}(x,y)\,\mathrm{d}x\,\mathrm{d}y$ 求全微分:

$$\mathrm{d}\left(f_{xy}(x,y)\,\mathrm{d}x\,\mathrm{d}y\right)=f_{xyx}(x,y)\,\mathrm{d}x^2\,\mathrm{d}y+f_{xyy}(x,y)\,\mathrm{d}x\,\mathrm{d}y^2.$$

4. 对 $f_{yy}(x,y)\,\mathrm{d}y^2$ 求全微分:

$$\mathrm{d}\left(f_{yy}(x,y)\,\mathrm{d}y^2\right)=f_{yyx}(x,y)\,\mathrm{d}x\,\mathrm{d}y^2+f_{yyy}(x,y)\,\mathrm{d}y^3.$$

上一节, 我们证明过当 $f(x,y)$ 的二阶混合偏导数连续, 则有 $f_{xy}(x,y)=f_{yx}(x,y)$. 类似地结果可以扩张到 $f(x,y)$ 的**三阶混合偏导数** (下面这讨论的 6 种都是三阶混合偏导数), 但稍微复杂些, 需要根据 x,y 的重复度分类讨论:

1. 对 x 求 2 次偏导, 且对 y 求 1 次偏导. 下面的三个三阶混合偏导数当它们各自都连续时, 则它们是相等的函数.

$$f_{xxy}(x,y)=f_{yxx}(x,y)=f_{xyx}(x,y),\quad\forall(x,y).$$

2. 对 x 求 1 次偏导, 且对 y 求 2 次偏导. 下面的三个三阶混合偏导数当它们各自都连续时, 则它们是相等的函数.

$$f_{yxy}(x,y)=f_{xyy}(x,y)=f_{yyx}(x,y),\quad\forall(x,y).$$

把以上所有结果组合在一起, 就得到三阶微分 d^3f 的最终形式:

$$\mathrm{d}^3f=f_{xxx}(x,y)\mathrm{d}x^3+3f_{xxy}(x,y)\mathrm{d}x^2\ \mathrm{d}y+3f_{xyy}(x,y)\mathrm{d}x\ \mathrm{d}y^2+f_{yyy}(x,y)\mathrm{d}y^3.$$

再把点 (x,y) 固定, 我们可以把 d^3f 视作是关于 $\mathrm{d}x,\mathrm{d}y$ 的次数皆为 3 的二元多项式函数. 不严验证,

$$\mathrm{d}^3f=\left(\mathrm{d}x\frac{\partial}{\partial x}+\mathrm{d}y\frac{\partial}{\partial y}\right)^3f=\frac{\partial^3f}{\partial x^3}\ \mathrm{d}x^3+3\frac{\partial^3f}{\partial x^2\partial y}\ \mathrm{d}x^2\ \mathrm{d}y+3\frac{\partial^3f}{\partial x\partial y^2}\ \mathrm{d}x\ \mathrm{d}y^2+\frac{\partial^3f}{\partial y^3}\ \mathrm{d}y^3.$$

因此, 我们仍旧可以用牛顿二项式展开来书写三阶微分.

高阶微分　二阶及二阶以上微分称为**高阶微分**. 总之, 我们有以下结论.

结论 6.5.2: 高阶微分公式

一般地, 当 $f(x,y)\in C^n(D)$ 时, $f(x,y)$ 在区域 D 内的 n $(n\geq 1)$ 阶微分为

$$\mathrm{d}^n f=\left(\mathrm{d}x\frac{\partial}{\partial x}+\mathrm{d}y\frac{\partial}{\partial y}\right)^n f.$$

利用牛顿二项式展开式

$$(a+b)^n=\sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k,\quad \binom{n}{k}=\frac{n!}{k!(n-k)!},$$

则有

$$\begin{aligned}\mathrm{d}^n f&=\left(\sum_{k=0}^{n}\binom{n}{k}\mathrm{d}x^{n-k}\mathrm{d}y^k\frac{\partial^n}{\partial x^{n-k}\partial y^k}\right)f\\&=\sum_{k=0}^{n}\binom{n}{k}\frac{\partial^n f}{\partial x^{n-k}\partial y^k}\cdot\mathrm{d}x^{n-k}\mathrm{d}y^k.\end{aligned}$$

对三元函数来说, 也有类似于二元函数的高阶微分. 当 $f(x,y,z)\in C^2(D)$ 时, 二阶微分就是

$$\begin{aligned}\mathrm{d}^2 f&=\left(\mathrm{d}x\frac{\partial}{\partial x}+\mathrm{d}y\frac{\partial}{\partial y}+\mathrm{d}z\frac{\partial}{\partial z}\right)^2 f\\&=f_{xx}\ \mathrm{d}x^2+f_{yy}\ \mathrm{d}y^2+f_{zz}\ \mathrm{d}z^2+2f_{xy}\ \mathrm{d}x\ \mathrm{d}y+2f_{yz}\ \mathrm{d}y\ \mathrm{d}z+2f_{xz}\ \mathrm{d}x\ \mathrm{d}z.\end{aligned}$$

注意上面我们使用了混合偏导数的连续性. 一般地, 当 $f(x,y,z)\in C^n(D)$ 时, $f(x,y,z)$ 在区域 D 内的 n 阶微分为

$$\mathrm{d}^n f=\left(\mathrm{d}x\frac{\partial}{\partial x}+\mathrm{d}y\frac{\partial}{\partial y}+\mathrm{d}z\frac{\partial}{\partial z}\right)^n f.$$

更多元函数的结果以此类推.

高阶微分不具有形式不变性, 因此高阶微分的上述公式只对 x,y (或 x,y,z) 作为自变量成立, 这一点与一元函数的情况是一致的. 高阶微分的概念将在后面介绍的多元函数的泰勒公式中得到应用.

例 6.5.10 设函数 $f(x,y)=\mathrm{e}^x y^2$, 求 $\mathrm{d}^3 f$.

证明. 根据高阶微分公式, 我们有

$$\mathrm{d}^3 f=\frac{\partial^3 f}{\partial x^3}\ \mathrm{d}x^3+3\frac{\partial^3 f}{\partial x^2\partial y}\ \mathrm{d}x^2\ \mathrm{d}y+3\frac{\partial^3 f}{\partial x\partial y^2}\ \mathrm{d}x\ \mathrm{d}y^2+\frac{\partial^3 f}{\partial y^3}\ \mathrm{d}y^3.$$

注意到

$$\frac{\partial^3 f}{\partial x^3}=\mathrm{e}^x y^2,\quad \frac{\partial^3 f}{\partial x^2\partial y}=2\mathrm{e}^x y,\quad \frac{\partial^3 f}{\partial x\partial y^2}=2\mathrm{e}^x,\quad \frac{\partial^3 f}{\partial y^3}=0,$$

所以

$$\mathrm{d}^3 f=\mathrm{e}^x y^2\ \mathrm{d}x^3+6\mathrm{e}^x y\ \mathrm{d}x^2\ \mathrm{d}y+6\mathrm{e}^x\ \mathrm{d}x\ \mathrm{d}y^2.$$

□

例 6.5.11 设函数 $f(x,y,z)=xy^2+z^3-xyz$, 求 $\mathrm{d}^2 f$.

证明. 先求出所有一阶偏导数和二阶偏导数:

$$
\begin{aligned}
&f_x(x,y,z)=y^2-yz,\quad f_y(x,y,z)=2xy-xz,\quad f_z(x,y,z)=3z^2-xy,\\
&f_{xx}(x,y,z)=0,\quad f_{xy}(x,y,z)=2y-z,\quad f_{xz}(x,y,z)=-y,\\
&f_{yy}(x,y,z)=2x,\quad f_{yz}(x,y,z)=-x,\\
&f_{zz}(x,y,z)=6z.
\end{aligned}
$$

于是

$$
\begin{aligned}
\mathrm{d}^2 f&=f_{xx}\,\mathrm{d}x^2+f_{yy}\,\mathrm{d}y^2+f_{zz}\,\mathrm{d}z^2+2f_{xy}\,\mathrm{d}x\,\mathrm{d}y+2f_{yz}\,\mathrm{d}y\,\mathrm{d}z+2f_{xz}\,\mathrm{d}x\,\mathrm{d}z\\
&=2x\,\mathrm{d}y^2+6z\,\mathrm{d}z^2+2(2y-z)\,\mathrm{d}x\,\mathrm{d}y-2x\,\mathrm{d}y\,\mathrm{d}z-2y\,\mathrm{d}x\,\mathrm{d}z.
\end{aligned}
$$

□

注意 二元函数的高阶微分和三元函数的高阶微分展开后的形式上有重大不同, 请不要搞混了.

6.6 方向导数与梯度

6.6.1 方向导数

我们知道, 偏导数 $\dfrac{\partial f}{\partial x}$ 与 $\dfrac{\partial f}{\partial y}$ 实际上就是函数 $f(x,y)$ 沿相应的坐标轴正向的变化率. 坐标轴的正向仅是两个特殊的方向而已, 所以很自然地可以考虑函数沿其他方向的变化率. 这就导致方向导数概念的引入. 我们以二元函数为例, 自变量的移动是在平面上进行的, 故方向也是指平面上的方向. 首先, 我们要明确一下如何讨论平面上的方向.

设二维向量 $\boldsymbol{l}=(l_1,l_2)$ 是 Oxy 平面上任意一个非零向量, 将其起点放到原点, 则终点的位置就指明着平面上的某个移动方向. 这里, 我们关心的只是这个向量所代表的方向, 而不希望它本身的大小 (模) $|\boldsymbol{l}|=\sqrt{l_1^2+l_2^2}$ 可能会招来没必要的麻烦. 因此, 相比考虑向量 $\boldsymbol{l}=(l_1,l_2)$ 本身, 我们更感兴趣的是和它同方向上的单位向量 $\boldsymbol{l}_0$, 容易算出

$$
\boldsymbol{l}_0=\frac{\boldsymbol{l}}{|\boldsymbol{l}|}=\frac{1}{\sqrt{l_1^2+l_2^2}}(l_1,l_2)=\left(\frac{l_1}{\sqrt{l_1^2+l_2^2}},\frac{l_2}{\sqrt{l_1^2+l_2^2}}\right).
$$

我们发现, 一切和 $\boldsymbol{l}$ 同向的向量 $\lambda\boldsymbol{l}=(\lambda l_1,\lambda l_2)$ (其中 λ 是任意严格大于零的实数), 都会导出完全相同的单位向量, 这是因为

$$
(\lambda\boldsymbol{l})_0=\frac{\lambda\boldsymbol{l}}{|\lambda\boldsymbol{l}|}=\frac{\lambda\boldsymbol{l}}{\lambda|\boldsymbol{l}|}=\frac{\boldsymbol{l}}{|\boldsymbol{l}|}=\boldsymbol{l}_0.
$$

因此, 单位向量 $\boldsymbol{l}_0$ 很好地描述了一个向量的方向性, 而不受到向量大小的影响. 它代表着唯一一个方向. 任何不同的两个单位向量必然指向着不同的方向, 而它们的模都是单位 1.

我们知道单位向量还可以用方向余弦[23]来表示. 令 α,β 分别是二维向量 $\boldsymbol{l}=(l_1,l_2)$ 和 Oxy 平面的 x,y 轴的正向的夹角 (范围在 $[0,\pi]$), 则 $\boldsymbol{l}$ 方向上的单位向量 $\boldsymbol{l}_0$ 还可以表示为

$$
\boldsymbol{l}_0=(\cos\alpha,\cos\beta),
$$

[23]第五章第一节的介绍方向余弦的时候, 我们处理的是三维向量, 但是其相关的所有结果可以自然地对应到二维向量.

其中, $\cos\alpha, \cos\beta$ 就是二维向量 $\boldsymbol{l}$ 的方向余弦. 设平面上的坐标向量为 $\boldsymbol{i}=(1,0), \boldsymbol{j}=(0,1)$, 根据点乘运算的定义, 我们有

$$\cos\alpha=\frac{\boldsymbol{l}\cdot\boldsymbol{i}}{|\boldsymbol{l}|},\quad \cos\beta=\frac{\boldsymbol{l}\cdot\boldsymbol{j}}{|\boldsymbol{l}|}.$$

于是, 毫不意外地有如下结果:

$$\cos\alpha=\frac{l_1}{\sqrt{l_1^2+l_2^2}},\quad \cos\beta=\frac{l_1}{\sqrt{l_1^2+l_2^2}}.$$

因此 $\cos\alpha, \cos\beta$ 作为方向余弦, 一定有恒等式[24]

$$\cos^2\alpha+\cos^2\beta=1.$$

这再次印证了二维向量 $\boldsymbol{l}$ 的方向余弦所确定的向量 $(\cos\alpha, \cos\beta)$ 就是和它同方向上的单位向量 $\boldsymbol{l}_0$. 注意, 方向余弦本身的定义就无关乎 $\boldsymbol{l}$ 的大小, 而只和 $\boldsymbol{l}$ 的方向相关.

总结一下: 设 $\boldsymbol{l}$ 是给定的一个平面上的方向, 当我们讨论其方向余弦组成的向量 $(\cos\alpha, \cos\beta)$ 时, 就是在讨论和它同方向上的单位向量 $\boldsymbol{l}_0$. 方向余弦的描述的办法在某些场合会方便一些, 但是通常在计算时, 要明确知道下面这些都是同一个东西:

$$(\cos\alpha, \cos\beta)\equiv\boldsymbol{l}_0\equiv\frac{\boldsymbol{l}}{|\boldsymbol{l}|}.$$

定义 6.6.1: 方向导数

设函数 $z=f(x,y)$ 在点 $P_0(x_0,y_0)$ 的一个邻域内有定义. 又设 $\boldsymbol{l}$ 是 Oxy 平面上给定的一个方向.

- 考虑 $\boldsymbol{l}$ 的方向余弦为 $(\cos\alpha, \cos\beta)$, 若极限
$$\lim_{t\to 0}\frac{f(x_0+t\cos\alpha, y_0+t\cos\beta)-f(x_0,y_0)}{t} \tag{6.4}$$
- (换一种描述方法) 若一元函数[a] $t\mapsto f(P_0+t\cdot\boldsymbol{l}_0)$ 在 $t=0$ 处的极限
$$\left.\frac{\mathrm{d}}{\mathrm{d}t}f(P_0+t\cdot\boldsymbol{l}_0)\right|_{t=0}=\lim_{t\to 0}\frac{f(P_0+t\cdot\boldsymbol{l}_0)-f(P_0)}{t} \tag{6.5}$$

存在, 则称此极限值为 $z=f(x,y)$ 在点 P_0 处沿方向 $\boldsymbol{l}$ 的**方向导数 (directional derivative)**, 记作

$$\left.\frac{\partial z}{\partial\boldsymbol{l}}\right|_{(x_0,y_0)},\quad \left.\frac{\partial f}{\partial\boldsymbol{l}}\right|_{(x_0,y_0)},\quad \left.\frac{\partial z}{\partial\boldsymbol{l}}\right|_{P_0},\quad 或\quad \left.\frac{\partial f}{\partial\boldsymbol{l}}\right|_{P_0}.$$

[a] 1-2-1 型. t 的定义域是包含 0 的一个开区间, 使得移动后的新点 $P_0+t\cdot\boldsymbol{l}_0$ 还落在点 P_0 的邻域内.

关于以上定义的要点:

1. 极限式 (6.8) 和 (6.9) 是等价的. 过点 $P_0(x_0,y_0)$ 沿给定的方向 $\boldsymbol{l}$ 作一条直线 L (见图 6.12), 则直线 L (取其方向向量为 $\boldsymbol{l}_0=(\cos\alpha,\cos\beta)$) 的参数方程为

$$\begin{cases} x=x_0+t\cos\alpha, \\ y=y_0+t\cos\beta, \end{cases}\quad (-\infty<t<+\infty),$$

[24] 在平面中, 当 α 为锐角时, 有 $\alpha+\beta=\frac{\pi}{2}$, 于是 $\cos\beta=\cos(\frac{\pi}{2}-\alpha)=\sin\alpha$. 当 α 为钝角时, 有 $\alpha-\beta=\frac{\pi}{2}$, 于是 $\cos\beta=\cos(\alpha-\frac{\pi}{2})=\cos(\frac{\pi}{2}-\alpha)=\sin\alpha$. 所以, 恒等式 $\cos^2\alpha+\cos^2\beta=1$ 也就是 $\cos^2\alpha+\sin^2\alpha=1$.

其中 t 为参数. 对于任意的 t, 相应的点记为

$$P_t := P_0 + t \cdot \boldsymbol{l}_0 = (x_0 + t\cos\alpha, y_0 + t\cos\beta)\,.$$

那么定义中的极限式 (6.8) 便可写成

$$\lim_{t\to 0}\frac{f(P_t)-f(P_0)}{t} = \lim_{t\to 0}\frac{f(P_0 + t\cdot \boldsymbol{l}_0)-f(P_0)}{t}.$$

这就是极限式 (6.9). 另外, 定点 $P_0 = (x_0, y_0)$ 和动点 P_t 之间的欧几里得距离为

$$\begin{aligned} d(P_0, P_t) &= \sqrt{(x_0 + t\cos\alpha - x_0)^2 + (y_0 + t\cos\beta - y_0)^2} \\ &= \sqrt{t^2(\cos^2\alpha + \cos^2\beta)} = |t|. \end{aligned}$$

于是, t 的值实际上是点 P_0 到点 P_t 的**有向距离**:

(a) 当 $\overrightarrow{P_0P_t}$ 的方向与方向 $\boldsymbol{l}$ 一致时, 即 $t > 0$, 则点 P_0 到点 P_t 的距离就是 t;

(b) 而当 $\overrightarrow{P_0P_1}$ 的方向与方向 $\boldsymbol{l}$ 相反时, 即 $t < 0$, 则点 P_0 到点 P_t 的距离是 $-t$.

总之, 方向导数实际上表示的是函数 $z = f(x, y)$ 在定点 P_0 出发, 沿给定方向 $\boldsymbol{l}$ 移动时, 函数值的瞬时变化率:

$$\left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{P_0} := \left.\frac{\mathrm{d}}{\mathrm{d}t} f(P_0 + t\cdot \boldsymbol{l}_0)\right|_{t=0}.$$

从几何意义上来看, 方向导数就是函数 $z = f(x, y)$ 图像曲面与 “经过定点 P_0 且与方向 $\boldsymbol{l}$ 平行的平面” 相交所形成的曲线, 在点 P_0 处的切线的斜率. 可以参考 Directional Derivative –GeoGebra.

2. 方向导数需要考虑三个要素: 1. 多元函数 f 自身; 2. 我们所关心的固定的一个起点 P_0; 3. 从 P_0 出发, 朝着定义域的哪个方向 $\boldsymbol{l}$ 移动. 这三个要素共同确定了方向导数

$$\mathbf{R} \ni \left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{P_0} := \left.\frac{\mathrm{d}}{\mathrm{d}t} f(P_0 + t\cdot \boldsymbol{l}_0)\right|_{t=0}.$$

当我们固定方向 $\boldsymbol{l}$ 时, 考虑二元函数 $\dfrac{\partial f}{\partial \boldsymbol{l}}$, 其定义为

$$P \mapsto \left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{P}.$$

它给出了函数 f 在任意点处沿着固定方向 $\boldsymbol{l}$ 移动时的函数值的瞬时变化率.

3. 因为限定为同方向的单位向量, 所以只要方向相同, 在其他不变的情况下, 方向导数就是相同的. (注意, 在有的教材中, 会不限定方向是单位向量. 这只是增加了一些没必要的麻烦.)

4. 上述的定义只讨论了二元函数 $f(x, y)$. 所谓的方向导数的 “方向” 一词, 是自变量在其所处空间的移动方向. 于是, 当讨论三元函数 $f(x, y, z)$ 的方向导数时, 方向 $\boldsymbol{l}$ 就是三维向量, 上述定义的极限式子 (6.8) 就要变成

$$\lim_{t\to 0}\frac{f(x_0 + t\cos\alpha, y_0 + t\cos\beta, z_0 + t\cos\gamma) - f(x_0, y_0, z_0)}{t}.$$

$\cos\alpha, \cos\beta, \cos\gamma$ 为 $\boldsymbol{l}$ 的方向余弦. 更多元的情况以此类推. 但是对于任意多元的函数, 按照极限式子 (6.9) 的方式描述的话, 则定义不需要更改任意一个字.

5. 方向导数是偏导数概念的推广, 而偏导数只是方向导数的特例.

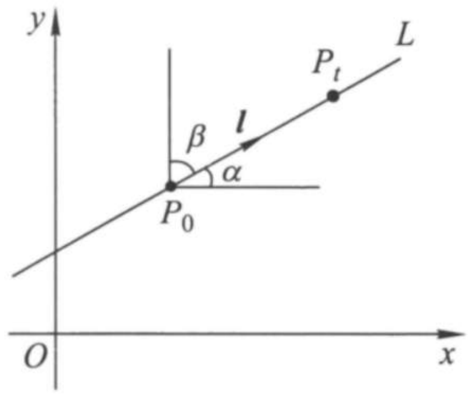


图 6.12: 方向导数定义中, 自变量的移动.

(a) 当 $\alpha=0, \beta=\frac{\pi}{2}$ 时 (也就是所 $\boldsymbol{l}=(l_1, 0), l_1>0$), 则有 $\frac{\partial f}{\partial \boldsymbol{l}}=\frac{\partial f}{\partial x}$;

(b) 当 $\alpha=\frac{\pi}{2}, \beta=0$ 时 (也就是所 $\boldsymbol{l}=(0, l_2), l_2>0$), 则有 $\frac{\partial f}{\partial \boldsymbol{l}}=\frac{\partial f}{\partial y}$.

这从上述定义中看得十分清楚.

下面的定理 (可微 $\Rightarrow$ 任意方向可导) 给出了方向导数存在的一个*充分 (但不必要)* 条件及它的计算公式. 既然偏导数只是方向导数的特例, 故之前的定理 (可微 $\Rightarrow$ 可偏导) 就是下面定理的一个特例.

定理 6.6.1: 可微 $\Rightarrow$ 任意方向可导

若函数 $f(x, y)$ 在点 $P_0\left(x_0, y_0\right)$ 处可微, 则 $f(x, y)$ 在该点处沿任意一个方向 $\boldsymbol{l}$ (其方向余弦设为 $\cos \alpha, \cos \beta$) 的方向导数均存在, 并且

$$\left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{P_0}=f_x\left(x_0, y_0\right) \cos \alpha+f_y\left(x_0, y_0\right) \cos \beta .$$

如果我们引入向量[a]

$$\boldsymbol{g}=\left.\operatorname{grad} f\right|_{P_0}:=\left(f_x\left(x_0, y_0\right), f_y\left(x_0, y_0\right)\right),$$

并将它称为函数 $f(x, y)$ 在点 P_0 处的 **梯度 (gradient)**, 则上述等式就是一个内积:

$$\left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{P_0}=\boldsymbol{g} \cdot \boldsymbol{l}_0=\boldsymbol{g} \cdot \frac{\boldsymbol{l}}{|\boldsymbol{l}|} .$$

[a]在非常多的教材中, 梯度写作 $\nabla f(P_0)$, 符号 ∇ 读作 nabla.

证明. 令 $\cos \alpha, \cos \beta$ 为方向 $\boldsymbol{l}$ 的方向余弦, 则 $\boldsymbol{l}_0=(\cos \alpha, \cos \beta)$. 在方向 $\boldsymbol{l}$ 上任取一点

$$P_t\left(x_0+t \cos \alpha, y_0+t \sin \alpha\right) .$$

因函数 $f(x, y)$ 在点 $\left(x_0, y_0\right)$ 处可微, 故该函数的增量可表示为

$$\begin{aligned}
f\left(P_t\right)-f\left(P_0\right) & =f\left(x_0+t \cos \alpha, y_0+t \cos \beta\right)-f\left(x_0, y_0\right) \\
& =f_x\left(x_0, y_0\right) t \cos \alpha+f_y\left(x_0, y_0\right) t \cos \beta+o(\rho)
\end{aligned}$$

其中 ρ 是点 P_t 与 P_0 之间的欧式距离. 显然, 根据之前的讨论, 有 $\rho = |t|$. 因此, $o(\rho)$ 等于 $o(|t|)$. 于是, 我们有

$$\begin{aligned}\left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{P_0} &= \lim_{t\to 0}\frac{f\left(P_t\right)-f\left(P_0\right)}{t}\\ &= \lim_{t\to 0}\frac{f\left(x_0+t\cos\alpha, y_0+t\cos\beta\right)-f\left(x_0,y_0\right)}{t}\\ &= \lim_{t\to 0}\frac{f_x\left(x_0,y_0\right)t\cos\alpha+f_y\left(x_0,y_0\right)t\cos\beta+o(|t|)}{t}\\ &= f_x\left(x_0,y_0\right)\cos\alpha+f_y\left(x_0,y_0\right)\cos\beta.\end{aligned}$$

还可以使用复合函数的链式法则证明. 根据定义, 方向导数是一元函数 $t\mapsto f\left(P_0+t\cdot\boldsymbol{l}_0\right)$ 在 $t=0$ 处的极限, 即

$$\left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{P_0} = \left.\frac{\mathrm{d}}{\mathrm{d}t}f\left(P_0+t\cdot\boldsymbol{l}_0\right)\right|_{t=0}.$$

设

$$z=f(x,y),\quad x=x_0+t\cos\alpha,\quad y=y_0+t\sin\alpha,$$

则有

$$\begin{aligned}\left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{P_0} = \left.\frac{\mathrm{d}z}{\mathrm{d}t}\right|_{t=0} &= \left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0)}\cdot\left.\frac{\mathrm{d}x}{\mathrm{d}t}\right|_{t=0}+\left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0)}\cdot\left.\frac{\mathrm{d}y}{\mathrm{d}t}\right|_{t=0}\\ &= f_x\left(x_0,y_0\right)\cos\alpha+f_y\left(x_0,y_0\right)\cos\beta.\end{aligned}$$

证毕. □

上述定理告诉我们: *在函数可微的条件下, 任何方向导数都存在, 并且可以通过函数在该点 (P_0) 处的梯度和单位方向 $\boldsymbol{l}_0$ 的内积来计算.* 一个二元函数在点 $P_0\left(x_0,y_0\right)$ 处的梯度

$$\operatorname{grad} f|_{P_0} := \left(f_x\left(x_0,y_0\right), f_y\left(x_0,y_0\right)\right)\in\mathbf{R}^2$$

是一个二维向量. 梯度的概念可以推广到 n $(n\geq 3)$ 元函数中去. 比如, 对于三元函数 $f(x,y,z)$, 它在一点 $P_0\left(x_0,y_0,z_0\right)$ 处的梯度

$$\operatorname{grad} f|_{P_0} := \left(f_x\left(x_0,y_0,z_0\right), f_y\left(x_0,y_0,z_0\right), f_z\left(x_0,y_0,z_0\right)\right)\in\mathbf{R}^3$$

就是一个三维向量. 一般地, n 元函数 $f(x_1,\ldots,x_n)$ 的点 $P_0\in\mathbf{R}^n$ 处的梯度就是

$$\operatorname{grad} f|_{P_0} := \left(\frac{\partial f}{\partial x_1}(P_0), \frac{\partial f}{\partial x_2}(P_0), \cdots, \frac{\partial f}{\partial x_n}(P_0)\right)\in\mathbf{R}^n.$$

总之, **梯度就是将多元函数的所有（一阶）偏导数排列形成的一个向量. n 元函数在任何一点 $P_0\in\mathbf{R}^n$ 处的梯度就一定一定一定是 n 维向量.**

如果我们想考虑函数 f 在任意点 $P\left(x,y\right)$ 处的梯度, 那么我们就到了一个二元到二元的函数:

$$\operatorname{grad} f:\mathbf{R}^2\to\mathbf{R}^2,\quad P\mapsto \operatorname{grad} f|_P := \left(f_x\left(x,y\right), f_y\left(x,y\right)\right).$$

例如, 函数 $f(x,y)=xy+\dfrac{y^2}{2}$ 在任意点 (x,y) 处的梯度就是

$$\operatorname{grad} f|_{(x,y)} = (y, x+y),$$

这就出现了一个函数 $\operatorname{grad} f : \mathbf{R}^2 \to \mathbf{R}^2$, 定义为

$$(x, y) \mapsto (y, x + y).$$

尽管我们不可能想象函数 $\operatorname{grad} f$ 的图像, 但我们还是能选择一些点, 计算他们的梯度并展示出来. 见图 6.13. 我们把这种样子的函数 $\operatorname{grad} f : \mathbf{R}^2 \to \mathbf{R}^2$ 叫做梯度场, 它是向量场 (vector field) 的一种特殊情况.[25] 一般地, n 元函数 $f(x_1, \ldots, x_n)$ 的任意点 $P \in \mathbf{R}^n$ 处的梯度就定义了 n 元到 n 元的函数:

$$\operatorname{grad} f : \mathbf{R}^n \to \mathbf{R}^n, \quad P \mapsto \operatorname{grad} f|_P = \left(\frac{\partial f}{\partial x_1}(P), \frac{\partial f}{\partial x_2}(P), \cdots, \frac{\partial f}{\partial x_n}(P) \right).$$

问题 请在 Vector Fields –GeoGebra 中输入任意的一个二元函数的梯度, 并观察它们整体的梯度是否有某种趋势.

由定理 6.6.1 很容易看出: 当两个方向 $\boldsymbol{l}^1$ 与 $\boldsymbol{l}^2$ 恰好相反时, 函数 $f(x, y)$ 沿这两个方向的方向导数相差一个负号, 即

$$\frac{\partial f}{\partial \boldsymbol{l}^1} = -\frac{\partial f}{\partial \boldsymbol{l}^2}.$$

但这一事实更可以直接从定义出发得到. 若方向 $\boldsymbol{l}^1$ 与 $\boldsymbol{l}^2$ 恰好相反, 则其各自对应的单位向量满足 $\boldsymbol{l}_0^1 = -\boldsymbol{l}_0^2$. 根据定义,

$$\frac{\partial f}{\partial \boldsymbol{l}^1} = \lim_{t \to 0} \frac{f\left(P_0 + t \cdot \boldsymbol{l}_0^1\right) - f\left(P_0\right)}{t} \tag{6.6}$$

$$= -\left(\lim_{t \to 0} \frac{f\left(P_0 + (-t) \cdot \boldsymbol{l}_0^2\right) - f\left(P_0\right)}{-t} \right). \tag{6.7}$$

注意到 $t \to 0$ 等价于 $-t \to 0$. 令 $t' = -t$, 则

$$\frac{\partial f}{\partial \boldsymbol{l}^1} = -\left(\lim_{t' \to 0} \frac{f\left(P_0 + t' \cdot \boldsymbol{l}_0^2\right) - f\left(P_0\right)}{t'} \right) = -\frac{\partial f}{\partial \boldsymbol{l}_2}.$$

设函数 $z = f(x, y)$ 在点 $P_0\left(x_0, y_0\right)$ 的一个邻域内有定义. 又设 $\boldsymbol{l}$ 是 Oxy 平面上给定的一个方向.

- 考虑 $\boldsymbol{l}$ 的方向余弦为 $(\cos\alpha, \cos\beta)$, 若极限

$$\lim_{t \to 0} \frac{f\left(x_0 + t\cos\alpha, y_0 + t\cos\beta\right) - f\left(x_0, y_0\right)}{t} \tag{6.8}$$

- (换一种描述方法) 若一元函数[26] $t \mapsto f(P_0 + t \cdot \boldsymbol{l}_0)$ 在 $t = 0$ 处的极限

$$\left.\frac{\mathrm{d}}{\mathrm{d}t} f(P_0 + t \cdot \boldsymbol{l}_0)\right|_{t=0} = \lim_{t \to 0} \frac{f\left(P_0 + t \cdot \boldsymbol{l}_0\right) - f\left(P_0\right)}{t} \tag{6.9}$$

存在, 则称此极限值为 $z = f(x, y)$ 在点 P_0 处沿方向 $\boldsymbol{l}$ 的**方向导数 (directional derivative)**, 记作

$$\left.\frac{\partial z}{\partial \boldsymbol{l}}\right|_{(x_0, y_0)}, \quad \left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{(x_0, y_0)}, \quad \left.\frac{\partial z}{\partial \boldsymbol{l}}\right|_{P_0}, \quad \text{或} \quad \left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{P_0}.$$

[25] 如果一个向量场 (也就是 $\mathbf{R}^2 \to \mathbf{R}^2$ 的映射) 是某个标量函数的梯度, 那么便称为保守向量场 (conservative vector field). 这暗含这一个实事, 不是所有的 $\mathbf{R}^2 \to \mathbf{R}^2$ 的映射都是某个标量函数的梯度. 比如, 你无论如何找不到一个二元函数 f, 使得 $\operatorname{grad} f|_{(x,y)} = (y, -x + y)$. 这一块知识是下学期的曲线积分中会讲到.

[26] 1-2-1 型. t 的定义域是包含 0 的一个开区间, 使得移动后的新点 $P_0 + t \cdot \boldsymbol{l}_0$ 还落在点 P_0 的邻域内.

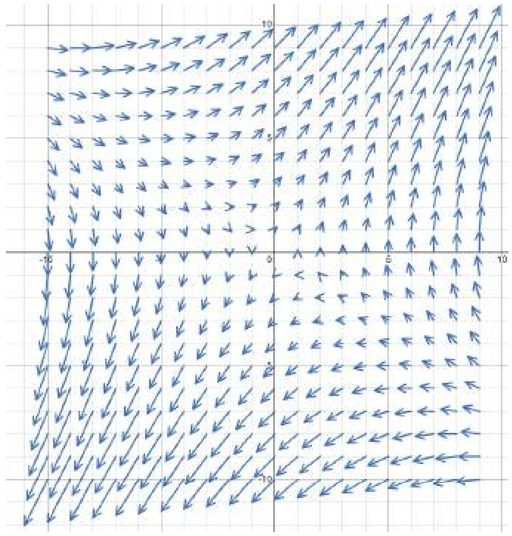


图 6.13: 梯度场 $\operatorname{grad} f:\mathbf{R}^2\to\mathbf{R}^2,\ (x,y)\mapsto(y,x+y)$.

例 6.6.1 设函数 $f(x,y)$ 在点 $P_0(2,4)$ 处可微, 且

$$f_x(2,4)=-3,\quad f_y(2,4)=8$$

求 $f(x,y)$ 在点 P_0 处沿从点 P_0 到点 $P(5,0)$ 方向的方向导数.

证明. 因为 $\overrightarrow{P_0P}=(3,-4)$, 所以 $\overrightarrow{P_0P}$ 的方向余弦为

$$\cos\alpha=\frac{3}{5},\quad \cos\beta=-\frac{4}{5}.$$

根据上述定理, $f(x,y)$ 在点 P_0 处沿从点 P_0 到点 P 方向, 即方向 $\boldsymbol{l}=\overrightarrow{P_0P}$ 的方向导数为

$$\left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{(2,4)}=f_x(2,4)\cos\alpha+f_y(2,4)\cos\beta=(-3)\times\frac{3}{5}+8\times\left(-\frac{4}{5}\right)=-\frac{41}{5}.$$

□

例 6.6.2 求函数 $f(x,y)=x^3y$ 在点 $P_0(1,2)$ 处沿从点 P_0 到点 $P(1+\sqrt{3},3)$ 方向的方向导数.

证明. 首先, 计算 $f(x,y)$ 在点 P_0 处的偏导数:

$$\left.\frac{\partial f}{\partial x}\right|_{(1,2)}=\left.3x^2y\right|_{(1,2)}=6,\quad \left.\frac{\partial f}{\partial y}\right|_{(1,2)}=1.$$

其次, 计算从点 P_0 到点 P 方向 $\boldsymbol{l}=\overrightarrow{P_0P}$ 的方向余弦 $\cos\alpha,\cos\beta$. 因为 $\overrightarrow{P_0P}=(\sqrt{3},1)$, 所以

$$(\cos\alpha,\cos\beta)=\left(\frac{\sqrt{3}}{2},\frac{1}{2}\right).$$

于是, 我们得到在点 P_0 处沿方向 $\boldsymbol{l}=\overrightarrow{P_0P}$ 的方向导数为

$$\left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{(1,2)}=6\times\frac{\sqrt{3}}{2}+1\times\frac{1}{2}=3\sqrt{3}+\frac{1}{2}.$$

□

对于三元函数 $u=f(x,y,z)$, 可类似地定义它在点 $P_0\left(x_0,y_0,z_0\right)$ 处沿某一方向 $\boldsymbol{l}$ 的方向导数. 若方向 $\boldsymbol{l}$ 的方向余弦为 $\cos\alpha,\cos\beta,\cos\gamma$, 并且 $f(x,y,z)$ 在点 P_0 处可微, 则有计算公式

$$\left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{P_0}=f_x\left(x_0,y_0,z_0\right)\cos\alpha+f_y\left(x_0,y_0,z_0\right)\cos\beta+f_z\left(x_0,y_0,z_0\right)\cos\gamma.$$

如果我们引入 $f(x,y,z)$ 在点 P_0 处的梯度

$$\boldsymbol{g}=\left.\operatorname{grad} f\right|_{P_0}:=\left(f_x\left(x_0,y_0,z_0\right),f_y\left(x_0,y_0,z_0\right),f_z\left(x_0,y_0,z_0\right)\right),$$

则上述等式仍旧是

$$\left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{P_0}=\boldsymbol{g}\cdot\boldsymbol{l}_0=\boldsymbol{g}\cdot\frac{\boldsymbol{l}}{|\boldsymbol{l}|}.$$

这个内积的结果对于 n 元函数都成立. 凸显了梯度的重要性.

例 6.6.3 设函数 $u=xy+yz+zx$, 又设向量 $\boldsymbol{l}$ 的坐标为 $(1,3,1)$, 求函数 u 在点 $(1,1,1)$ 处沿 l 方向的方向导数 $\left.\frac{\partial u}{\partial \boldsymbol{l}}\right|_{(1,1,1)}$.

证明. 先求 l 的方向余弦 $\cos\alpha,\cos\beta,\cos\gamma$. 根据方向余弦的定义, 有

$$(\cos\alpha,\cos\beta,\cos\gamma)=\frac{\boldsymbol{l}}{|\boldsymbol{l}|}=\left(\frac{1}{\sqrt{11}},\frac{3}{\sqrt{11}},\frac{1}{\sqrt{11}}\right).$$

这样, 我们得到

$$\begin{aligned}\left.\frac{\partial u}{\partial \boldsymbol{l}}\right|_{(1,1,1)}&=\left.\frac{\partial u}{\partial x}\right|_{(1,1,1)}\cos\alpha+\left.\frac{\partial u}{\partial y}\right|_{(1,1,1)}\cos\beta+\left.\frac{\partial u}{\partial z}\right|_{(1,1,1)}\cos\gamma\\&=2\times\frac{1}{\sqrt{11}}+2\times\frac{3}{\sqrt{11}}+2\times\frac{1}{\sqrt{11}}=\frac{10}{\sqrt{11}}.\end{aligned}$$

□

下面的例子说明即使方向导数都存在, 但不能说明可微. 因此, 定理 6.6.1 的反面是不成立的.

例 6.6.4 可以证明下列函数在点 $(0,0)$ 处的任何方向的方向导数都存在, 但是在点 $(0,0)$ 处是不可微的.

$$f(x,y)=\begin{cases}0, & \text{当}(x,y)=(0,0),\\ \dfrac{x^3}{x^2+y^2}, & \text{当}(x,y)\neq(0,0).\end{cases}$$

其图像参考 任意方向可导但不可微 | Desmos. 考虑 $\boldsymbol{l}$ 的方向余弦为 $(\cos\alpha,\cos\beta)$, 则在点 $(0,0)$ 处的方向导数为

$$\left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{(0,0)}=\lim_{t\to0}\frac{f\left(t\cos\alpha,t\cos\beta\right)-f\left(0,0\right)}{t}=\frac{\frac{t^3\cos^3\alpha}{t^2\cos^2\alpha+t^2\cos^2\beta}}{t}=\frac{\cos^3\alpha}{\cos^2\alpha+\cos^2\beta}.$$

根据定理 6.6.1, 如果函数在点 $(0,0)$ 处可微, 那么 $\left.\dfrac{\partial f}{\partial \boldsymbol{l}}\right|_{(0,0)}$ 就一定是关于 $\cos\alpha, \cos\beta$ 的线性函数. 而上述表现的是非线性函数, 这就导致了矛盾. 因此函数函数在点 $(0,0)$ 处是不可微的. 从函数的图像上看, 可微分要求的是曲面在点 $(0,0)$ 处存在一个切平面. 从几何上看, 上述函数的图像在点 $(0,0)$ 处是无法存在这样的切平面的.

6.6.2　梯度的意义

函数在一点处沿某一方向的方向导数, 反映了函数沿该方向的变化率, 方向不同, 变化率一般也不同. 现在确定一点, 考虑不同方向的方向导数.

- 如果某一方向的方向导数为正, 则沿着该方向前进时, 函数值会变大. 且方向导数的绝对值越大, 其增加的速度也就越快. (可以参考 Directional Derivative –GeoGebra)

- 如果某一方向的方向导数为负, 则沿着该方向前进时, 函数值会变小. 且方向导数的绝对值越大, 其减少的速度也就越快.

我们自然提出这样一个问题: 在同一点处的所有方向导数中, 是否有最大值或最小值? 如果存在, 是正还是负? 而且沿怎样的方向才能取得最大值或最小值? 下面来讨论这个问题.

设函数 $f(x,y)$ 在点 $P_0\left(x_0,y_0\right)$ 处有连续的一阶偏导数 (因而可微). 由本节中的定理我们可知, $f(x,y)$ 在点 P_0 处沿任意一个方向 $\boldsymbol{l}$ 的方向导数都存在, 且可表示为

$$\left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{P_0} = \boldsymbol{g}\cdot\boldsymbol{l}_0.$$

其中, 取方向 $\boldsymbol{l}$ 的单位向量为 $\boldsymbol{l}_0=(\cos\alpha,\cos\beta)$, 而函数在点 $P_0\left(x_0,y_0\right)$ 处的梯度 (假设 $\boldsymbol{g}$ 不是零向量)

$$\boldsymbol{g} = \left.\operatorname{grad} f\right|_{P_0} = \left(f_x\left(x_0,y_0\right), f_y\left(x_0,y_0\right)\right).$$

这里当点 P_0 给定后梯度 $\boldsymbol{g}$ 为一个固定向量, 但 $\boldsymbol{l}_0$ 是随 $\boldsymbol{l}$ 变化而变化的. 我们进一步将上式改写为如下形式:

$$\left.\frac{\partial f}{\partial \boldsymbol{l}}\right|_{P_0} = |\boldsymbol{g}|\left|\boldsymbol{l}_0\right|\cos\left\langle\boldsymbol{g},\boldsymbol{l}_0\right\rangle = |\boldsymbol{g}|\cos\left\langle\boldsymbol{g},\boldsymbol{l}_0\right\rangle,$$

其中 $\left\langle\boldsymbol{g},\boldsymbol{l}_0\right\rangle$ 表示梯度 $\boldsymbol{g}$ 与 $\boldsymbol{l}_0$ 之间的夹角. 很明显, 有以下观察:

- 当 $\left\langle\boldsymbol{g},\boldsymbol{l}_0\right\rangle=0$ 时, 方向导数 $\left.\dfrac{\partial f}{\partial \boldsymbol{l}}\right|_{P_0}$ 达到最大, 其值为 $|\boldsymbol{g}|>0$. 这就是说, 当方向 $\boldsymbol{l}$ 与梯度向量 $\boldsymbol{g}$ 的方向一致时, 函数沿方向 $\boldsymbol{l}$ 的方向导数最大, 且其最大值为 $|\boldsymbol{g}|$. 换句话说, **梯度的方向是函数值增加最快的方向.**

- 当 $\left\langle\boldsymbol{g},\boldsymbol{l}_0\right\rangle=\pi$ 时, 方向导数 $\left.\dfrac{\partial f}{\partial \boldsymbol{l}}\right|_{P_0}$ 达到最小, 其值为 $-|\boldsymbol{g}|<0$. 这就是说, 当方向 $\boldsymbol{l}$ 与梯度向量 $\boldsymbol{g}$ 的方向相反时, 函数沿方向 $\boldsymbol{l}$ 的方向导数最小, 且其最小值为 $-|\boldsymbol{g}|$. 换句话说, **负梯度的方向是函数值减少最快的方向.**

我们用下面一段话来概括前面的讨论:

结论 6.6.1

函数 $z=f(x,y)$ 在点 $P_0(x_0,y_0)$ 处的梯度 $\operatorname{grad} f|_{P_0}$ 是这样一个向量:

1. 它的方向是使方向导数达到最大值的方向, 它的模就是方向导数的最大值.
2. 它的反方向是使方向导数达到最小值的方向, 它的模取负号就是方向导数的最小值.

如果函数在某点处的梯度是零向量会发生什么? 这说明了函数在该点处往任意方向移动, 其函数值的变化率都是 0; 它既没有增加的趋势, 也没有减少的趋势. 一元函数的极值条件告诉我们, 函数在某点处的导数为 0 是称为极值点的必要条件. 这本章最后我们会讲这一事实扩张到二元函数的求极值问题中, 即可微函数的极值点的必要条件是在该点处梯度为零. 请在 Directional Derivative –GeoGebra 将点 P 直接拖拽到原点, 观察它的梯度和原点处的函数图像特征.

例 6.6.5 设函数 $f(x,y,z)=xyz$, 求 $\operatorname{grad} f|_{(1,2,3)}$.

证明. 根据梯度的计算公式, 我们有

$$\operatorname{grad} f=(yz,zx,xy),$$

因此

$$\operatorname{grad} f|_{(1,2,3)}=(6,3,2).$$

□

注释 6.6.1 在物理学中, 一个最简单的典型例子是点电荷所产生的电势 V 的梯度恰好是其电场强度 $\boldsymbol{E}$ 的负值, 即

$$\operatorname{grad} V=-\boldsymbol{E}.$$

事实上, 设电荷 q 放在原点, 则在点 (x,y,z) 处的电势为

$$V=\frac{q}{4\pi\varepsilon}\cdot\frac{1}{\sqrt{x^2+y^2+z^2}}.$$

其中 ε 为介电常量. 简单的计算表明

$$\operatorname{grad} V=(V_x,V_y,V_z)=\frac{q}{4\pi\varepsilon}\cdot\frac{-1}{\left(\sqrt{x^2+y^2+z^2}\right)^3}(x,y,z).$$

而上式最后的结果恰好是点 (x,y,z) 处的电场强度 $\boldsymbol{E}$ 的负值. $\operatorname{grad} V=-\boldsymbol{E}$ 告诉了我们这样一个事实: 沿着与电场强度相反的方向, 电势增加最快. 将来在讨论场论时读者会进一步看到梯度这一概念在场论中的意义.

现在我们列出有关梯度计算的一些运算法则及公式, 以供读者在必要时查阅: 设 $u=u(x,y), v=v(x,y)$ 是可微二元函数, 则有

1. $\operatorname{grad}(u\pm v)=\operatorname{grad} u\pm\operatorname{grad} v$;
2. $\operatorname{grad}(uv)=v\operatorname{grad} u+u\operatorname{grad} v$;

3. $\operatorname{grad}\dfrac{u}{v}=\dfrac{1}{v^2}(v\operatorname{grad}u-u\operatorname{grad}v), v\neq 0;$

4. $\operatorname{grad}(f(u,v))=\dfrac{\partial f}{\partial u}\operatorname{grad}u+\dfrac{\partial f}{\partial v}\operatorname{grad}v$, 其中 $f(u,v)$ 有连续的一阶偏导数.

这些公式的证明并不难, 读者不妨自己试着证明其中的一两个. 这些公式并不是最基本的数学公式, 无须特别记忆它们. 我们要强调指出: 一个函数 f 经过梯度运算之后得到的梯度 $\operatorname{grad}f$ 是一个向量. 因此, 上述公式 1 至 4 均是**向量等式**, 也就是说, 等式之两端都是向量. 此外, 在上述公式中的

$$v\operatorname{grad}u \quad 及 \quad \frac{\partial f}{\partial u}\operatorname{grad}u$$

等均理解为向量 $\operatorname{grad}u$ 与数的乘积.

6.7 多元函数的微分中值定理与泰勒公式

我们知道, 一元函数的微分中值定理及泰勒公式在研究函数性态上起着重要作用. 这些定理与公式同样可以推广到多元函数的情况, 并用来研究多元函数的性态.

6.7.1 多元函数的微分中值定理

我们先复习一下一元函数的拉格朗日中值定理: 设函数 $f(x)$ 在闭区间 $[a,b]$ 上连续, 在开区间 (a,b) 内可导, 则必存在一点 $c\in(a,b)$, 使得

$$f(b)=f(a)+f'(c)(b-a).$$

上述实事称为一元函数的拉格朗日中值定理. 我们总是可以将存在一点 $c\in(a,b)$ (即 点 c 位于点 a 与 b 连线上) 写成参数化形式:

$$c=a+\theta(b-a), \quad \exists\theta\in(0,1).$$

若用 x_0 与 x 分别替代 a 与 b 这两个符号, 并且进一步引入符号 $\Delta x=x-x_0$, 则我们有

$$f\left(x_0+\Delta x\right)=f\left(x_0\right)+f'\left(x_0+\theta\Delta x\right)\Delta x, \quad \exists\theta\in(0,1).$$

无论使用什么记号, 一元函数的微分中值公式的核心 “范式” 就是

$$f\left(\text{点 }1\right)-f\left(\text{点 }2\right)=\left(\text{两点连线上某个点的导数}\right)\cdot\left(\text{点 1 - 点 2}\right).$$

接下来, 我们将这种 “范式” 推广到任意多元函数上.

定理 6.7.1: 二元函数的拉格朗日中值定理

设函数 $z=f(x,y)$ 在区域 D 内有连续的一阶偏导数, 又设 D 中有两个点 $P_0\left(x_0,y_0\right)$ 与 $P_1\left(x_0+\Delta x,y_0+\Delta y\right)$, 并且点 P_0 到点 P_1 的线段 $P_0P_1\subset D$, 则存在 θ $(0<\theta<1)$, 使得 [a]

$$f\left(x_0+\Delta x,y_0+\Delta y\right)=f\left(x_0,y_0\right)+\frac{\partial f}{\partial x}\left(x_0+\theta\Delta x,y_0+\theta\Delta y\right)\Delta x+\frac{\partial f}{\partial y}\left(x_0+\theta\Delta x,y_0+\theta\Delta y\right)\Delta y.$$

[a] 马上我们会用向量语言来做一个更加简洁的描述, 见定理 6.7.2.

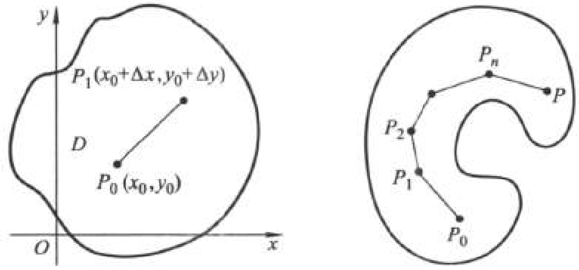


图 6.14: Enter Caption

证明. 设落在点 P_0 与 P_1 的连线上的任意动点为

$$P_t=(x_t,y_t)=(x_0+t\Delta x,y_0+t\Delta y)\in D,$$

其中实数 t $(0\leq t\leq 1)$ 是参数 (见图 6.14). 则我们可以考虑一元复合函数 $t\mapsto\varphi(t)$, 其被定义为动点 P_t 的函数值:

$$\varphi(t)=f\left(x_t,y_t\right),\quad x_t=x_0+t\Delta x,\quad y_t=y_0+t\Delta y.$$

显然, 上式是一个 "1-2-1 型" 的复合函数. 根据本定理的假定可知, 外层的 $f(x,y)$ 在 D 内可微, 而中间变量函数 $t\mapsto x_t$ 和 $t\mapsto y_t$ 又都是关于 t 是可导的, 且其导数分别是定值 $\Delta x,\Delta y$. 根据上一节关于复合函数 ("1-2-1 型" 那一小节) 的链式法则可知, 一元函数 $t\mapsto\varphi(t)$ 是可导的 (等价于可微), 且有公式

$$\varphi'(t)=\left.\frac{\partial f}{\partial x}\right|_{x_t}\cdot\left.\frac{\mathrm{d}x_t}{\mathrm{d}t}\right|_t+\left.\frac{\partial f}{\partial y}\right|_{y_t}\cdot\left.\frac{\mathrm{d}y_t}{\mathrm{d}t}\right|_t=\left.\frac{\partial f}{\partial x}\right|_{x_t}\Delta x+\left.\frac{\partial f}{\partial y}\right|_{y_t}\Delta y,$$

或展开写做

$$\varphi'(t)=\frac{\partial f}{\partial x}\left(x_0+t\Delta x,y_0+t\Delta y\right)\Delta x+\frac{\partial f}{\partial y}\left(x_0+t\Delta x,y_0+t\Delta y\right)\Delta y.$$

另外, 对一元函数 $\varphi(t)$ 使用一元函数的拉格朗日中值定理[27], 可以知道, 存在一个 θ $(0<\theta<1)$, 使得

$$\varphi(1)-\varphi(0)=\varphi'(\theta),$$

即

$$f\left(x_0+\Delta x,y_0+\Delta y\right)-f\left(x_0,y_0\right)=\left.\frac{\partial f}{\partial x}\right|_{x_\theta}\Delta x+\left.\frac{\partial f}{\partial y}\right|_{y_\theta}\Delta y.$$

这就是最终的结果. 证毕. □

假设 $f(x,y,z)$ 是一个三元函数, 且在某区域 D 内具有连续的一阶偏导数. 取两个点 $P_0=(x_0,y_0,z_0)$ 和 $P_1=(x_0+\Delta x,y_0+\Delta y,z_0+\Delta z)$, 并且点 P_0 到点 P_1 的线段 $P_0P_1\subset D$, 则存在 $0<\theta<1$, 使得

$$f(P_1)=f(P_0)+\frac{\partial f}{\partial x}(P_\theta)\Delta x+\frac{\partial f}{\partial y}(P_\theta)\Delta y+\frac{\partial f}{\partial z}(P_\theta)\Delta z,$$

其中, $P_\theta=(x_0+\theta\Delta x,y_0+\theta\Delta y,z_0+\theta\Delta z)$ 是线段 P_0P_1 上的某个点. 这就是三元函数的拉格朗日中值定理的具体形式, 其证明和上述定理完全相同. 若将其推广到任意 n 元函数, 可以利用向量的表示方法给出如下简洁的描述.

[27] 本小节所有扩张到多元函数的定理都是用这一个套路: 通过函数复合, 构造出一个一元函数, 然后对该函数使用一元版本的各种定理. 最将其结论用函数复合展开写, 最后就是多元函数版本所对应的定理结果了.

定理 6.7.2: n 元函数的拉格朗日中值定理

设 n $(n \geq 1)$ 元函数 $z = f(x_1, x_2, \ldots, x_n)$ 在区域 $D \subseteq \mathbf{R}^n$ 内具有连续的一阶偏导数. 设 D 中有两个点 (n 维向量) P_0 与 P_1, 并且连接点 P_0 与 P_1 的线段 $P_0P_1 \subset D$. 则存在 θ $(0 < \theta < 1)$, 使得

$$f(P_1) = f(P_0) + \operatorname{grad} f|_{P_\theta} \cdot \Delta\mathbf{x} = f(P_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(P_\theta)\Delta x_i,$$

其中, $P_\theta := P_0 + \theta\Delta\mathbf{x}$, $\Delta\mathbf{x} := P_1 - P_0$.

总之, n 元函数的拉格朗日中值定理说明了, 对于在 $\mathbf{R}^n$ 的某区域内具有连续一阶偏导数的多元函数 f, 如果 P_0 与 P_1 的连线 P_0P_1 还在该区域内, 那么存在连线 P_0P_1 上的某点 P_θ, 使得

$$f(P_1) - f(P_0) = \operatorname{grad} f|_{P_\theta} \cdot (P_1 - P_0).$$

注意, 以上的定理覆盖了一元函数的拉格朗日中值定理. 至此, 这么多的微分不同版本的拉格朗日中值公式符合范式:

$$\boxed{f\,(\text{点 }1) - f\,(\text{点 }2) = (\text{两点连线上某个点的梯度}) \cdot (\text{点 }1\text{ - 点 }2)\,.}$$

即, **任意两点之间的连线上存在一个中间点，使得函数的增量等于该中间点处梯度向量与两点差向量的内积. 注意这里的顺序和正负号不能记错!**

利用一元函数的拉格朗日中值定理, 我们曾经证明过: 若一元函数 $f(x)$ 在某个开区间上导数恒为零, 则为 $f(x)$ 在该区间上是常数函数. 我们将这一结论扩张到多元函数中, 即梯度恒为零的函数是常数函数.

推论 6.7.1: 梯度恒为零的函数是常数函数

若函数 $z = f(x, y)$ 在区域 D 内有连续的一阶偏导数, 并且满足 $\dfrac{\partial f}{\partial x} \equiv 0, \dfrac{\partial f}{\partial y} \equiv 0$, 则 $f(x, y)$ 在 D 内为一个常数.

证明. 在 D 内任意取定一点 $P_0\,(x_0, y_0)$. 另外再考虑 D 内任意的动点 $P(x, y)$.

- 假设点 P 与 P_0 之间的连线 P_0P 都在 D 内, 则可以直接应用多元函数的拉格朗日中值定理, 有

  $$f(P) - f\left(P_0\right) = f_x\left(P_\theta\right) h + f_y\left(P_\theta\right) k = 0,$$

  其中 $h = x - x_0, k = y - y_0$, P_θ 为线段 P_0P 上的某点. 这样, $f(P) = f\left(P_0\right)$.

- 假设线段 P_0P 不全包含在 D 内. 注意到 D 是区域, 根据区域的定义 (即非空的连通的开集合), D 中任意两点都可以通过用一条落在 D 中的曲线相连接. 显然, 则必存在折线 [28] $P_0P_1P_2\cdots P_nP \subset D$ (见图 6.14). 于是, 由对于每一小段线段使用前面讨论的情况, 总体看就有

  $$f(P) = f\left(P_n\right) = f\left(P_{n-1}\right) = \cdots = f\left(P_1\right) = f\left(P_0\right).$$

  总之, 对于任意的 $P \in D$, 我们证明了 $f(P) = f\left(P_0\right)$. 这就证明了 $f(x, y)$ 在 D 内为一个常数. 证毕.

□

[28] 关于为什么一定是折线的讨论可以忽略, 这超出课程范围. 但这在直觉上容易接受的.

6.7.2 多元函数的泰勒公式

使用高阶微分记号的一元函数的泰勒公式

假定一元函数 $f(x)$ 有 n 阶导数. 它的一阶微分通常写做 $\mathrm{d}f = f'(x)\mathrm{d}x$. 但是, 有时候为了强调这个微分是在哪个点处的微分, 比如点 x_0, 我们引入记号:

$$\mathrm{d}f\big|_{x_0} = \mathrm{d}f(x_0) = f'(x_0)\mathrm{d}x.$$

类似地, 在点 x_0 处的 n $(n \geq 1)$ 阶微分记作

$$\mathrm{d}^n f\big|_{x_0} = \mathrm{d}^n f(x_0) = f^{(n)}(x_0)\mathrm{d}x^n.$$

类似于零阶导数的符号 $f^{(0)}(x) = f(x)$. 我们记零阶微分符号

$$\mathrm{d}^0 f\big|_{x_0} = \mathrm{d}^0 f(x_0) = f^{(0)}(x_0)\mathrm{d}x^0 = f(x_0).$$

利用这种高阶微分 $\mathrm{d}^n f$ $(n \geq 0)$ 的符号, 我们就可以将以前学习过的一元函数的泰勒公式重新描述成一个简洁的样子.

设一元函数 $f(x)$ 在某个开区间 D 内有 $n+1$ 阶导数, 则对于任意的 $x_0, x \in D$, 有等式

$$f(x) = \underbrace{\sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k}_{\text{一元函数的泰勒多项式}} + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1},$$

其中 ξ 是介于 x_0 与 x 之间的某一点. 这里的余项为拉格朗日余项, 而整个公式为带拉格朗日余项的泰勒公式. 令 $\mathrm{d}x \equiv \Delta x = x - x_0$, 则上述泰勒公式可以写做

$$\begin{aligned}
f(x_0+\Delta x) &= \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(x_0) \cdot (\Delta x)^k + \frac{1}{(n+1)!} f^{(n+1)}(\xi) \cdot (\Delta x)^{n+1}, \\
&= \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(x_0)\, \mathrm{d}x^k + \frac{1}{(n+1)!} f^{(n+1)}(\xi)\mathrm{d}x^{n+1}, \\
&= \sum_{k=0}^{n} \frac{1}{k!} \mathrm{d}^k f(x_0) + \frac{1}{(n+1)!} \mathrm{d}^{n+1} f(\xi),
\end{aligned}$$

或者展开写做

$$f(x_0+\Delta x) = f(x_0) + \frac{1}{1!}\mathrm{d}f(x_0) + \frac{1}{2!}\mathrm{d}^2 f(x_0) + \cdots + \frac{1}{n!}\mathrm{d}^n f(x_0) + \frac{1}{(n+1)!}\mathrm{d}^{n+1} f(\xi).$$

若把右侧每一项的高阶微分展开写, 就是原来形式的泰勒公式. 若参数化表示 $\xi = x_0 + \theta\Delta x$, 其中 $0 < \theta < 1$, 则我们有以下结论.

结论 6.7.1: 使用高阶微分记号的一元函数的 (带拉格朗日余项的) 泰勒公式

设一元函数 $f(x)$ 在某个开区间 D 内有 $n+1$ 阶导数, 对于任意的 $x_0, x_0+\Delta x \in D$, 则存在 $0 < \theta < 1$, 使得

$$\begin{aligned}
f(x_0+\Delta x) &= f(x_0) + \frac{1}{1!}\mathrm{d}f(x_0) + \frac{1}{2!}\mathrm{d}^2 f(x_0) + \cdots + \frac{1}{n!}\mathrm{d}^n f(x_0) + \frac{1}{(n+1)!}\mathrm{d}^{n+1} f(x_0+\theta\Delta x) \\
&= \sum_{k=0}^{n} \frac{1}{k!}\mathrm{d}^k f(x_0) + \frac{1}{(n+1)!}\mathrm{d}^{n+1} f(x_0+\theta\Delta x).
\end{aligned}$$

另外, 当带拉格朗日余项的泰勒公式的项数 n 固定而令 $\Delta x \to 0$, 则有

$$\begin{aligned} f(x_0+\Delta x) &= f(x_0) + \frac{1}{1!}\mathrm{d}f(x_0) + \frac{1}{2!}\mathrm{d}^2 f(x_0) + \cdots + \frac{1}{n!}\mathrm{d}^n f(x_0) + o(\Delta x^n) \\ &= \sum_{k=0}^{n} \frac{1}{k!}\mathrm{d}^k f(x_0) + o(\Delta x^n). \end{aligned}$$

于是, 我们便得到使用高阶微分记号的一元函数的 (带佩亚诺余项的) 泰勒公式.

注意 我们其实就是把

$$\mathrm{d}x^k = (\Delta x)^k = (x-x_0)^k$$

给藏到了微分符号 $\mathrm{d}^k f$ 里面当中. 而这里的 $\mathrm{d}x = \Delta x$ 不是可以任意变动的, 它要和等式左侧中的 Δx 和拉格朗日余项中的 Δx 保持一致. 注意, 我们完全混用符号 $\mathrm{d}x = \Delta x$.

上述内容的目的是让读者熟悉使用高阶微分记号来描述泰勒公式或泰勒多项式. 因为接下来, 我们将使用高阶 (全) 微分记号的办法, 将泰勒公式推广到任意多元函数中.

二元函数的泰勒公式

设 $D \subset \mathbf{R}^2$ 为一个区域, 而函数 $f(x,y) \in C^{n+1}(D)$, 又设 $P_0(x_0,y_0) \in D$. 它的一阶微分通常写做 $\mathrm{d}f = f_x(x,y)\mathrm{d}x + f_y(x,y)\mathrm{d}y$. 为了强调这个微分是在哪个点 (比如 $P_0(x_0,y_0)$) 处的微分, 我们引入记号:

$$\mathrm{d}f|_{P_0} = \mathrm{d}f(x_0,y_0) = f_x(x_0,y_0)\mathrm{d}x + f_y(x_0,y_0)\mathrm{d}y.$$

类似地, 在点 $P_0(x_0,y_0)$ 处的 2 阶微分记作

$$\mathrm{d}^2 f|_{P_0} = \mathrm{d}^2 f(x_0,y_0) = f_{xx}(x_0,y_0)\mathrm{d}x^2 + 2f_{xy}(x_0,y_0)\mathrm{d}x\mathrm{d}y + f_{yy}(x_0,y_0)\mathrm{d}y^2.$$

一般地, n 阶微分是

$$\mathrm{d}^n f = \left(\mathrm{d}x\frac{\partial}{\partial x} + \mathrm{d}y\frac{\partial}{\partial y}\right)^n f.$$

利用牛顿二项式展开式

$$(a+b)^n = \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k, \quad \binom{n}{k} = \frac{n!}{k!(n-k)!},$$

则有

$$\begin{aligned} \mathrm{d}^n f &= \left(\sum_{k=0}^{n}\binom{n}{k}\mathrm{d}x^{n-k}\mathrm{d}y^k\frac{\partial^n}{\partial x^{n-k}\partial y^k}\right) f \\ &= \sum_{k=0}^{n}\binom{n}{k}\frac{\partial^n f}{\partial x^{n-k}\partial y^k}\cdot \mathrm{d}x^{n-k}\mathrm{d}y^k. \end{aligned}$$

于是在点 $P_0(x_0,y_0)$ 处的 n 阶微分就是

$$\mathrm{d}^n f|_{P_0} = \mathrm{d}^n f(x_0,y_0) = \sum_{k=0}^{n}\binom{n}{k}\frac{\partial^n f}{\partial x^{n-k}\partial y^k}(x_0,y_0)\cdot \mathrm{d}x^{n-k}\mathrm{d}y^k.$$

这是关于 $\mathrm{d}x, \mathrm{d}y$ 的次数皆为 n 的二元多项式. 类似于一元函数, 我们引入零阶微分符号

$$\mathrm{d}^0 f|_{P_0} = \mathrm{d}^0 f(x_0,y_0) = f(x_0,y_0).$$

利用这种高阶微分 $\mathrm{d}^n f\ (n \geq 0)$ 的符号, 我们就可以简洁地描述的二元函数的泰勒公式.

定理 6.7.3: 二元函数的 (带拉格朗日余项的) 泰勒公式

设 $D \subset \mathbf{R}^2$ 为一个区域, 而函数 $f(x,y) \in C^{n+1}(D)$, 又设 $P_0(x_0,y_0) \in D$, $P_1(x_0+\Delta x, y_0+\Delta y) \in D$, 并且点 P_0 与 P_1 之间的连线 $P_0P_1 \subset D$. 则存在 $0<\theta<1$, 使得

$$\begin{aligned}f(x_0+\Delta x, y_0+\Delta y) = f(x_0,y_0) + \frac{1}{1!}\mathrm{d}f(x_0,y_0) + \frac{1}{2!}\mathrm{d}^2 f(x_0,y_0) + \cdots + \frac{1}{n!}\mathrm{d}^n f(x_0,y_0) \\ + \frac{1}{(n+1)!}\mathrm{d}^{n+1} f(x_0+\theta\Delta x, y_0+\theta\Delta y), \\ = \sum_{k=0}^{n} \frac{1}{k!}\mathrm{d}^k f(x_0,y_0) + \frac{1}{(n+1)!}\mathrm{d}^{n+1} f(x_0+\theta\Delta x, y_0+\theta\Delta y),\end{aligned}$$

其中 $\mathrm{d}^k f\ (k=0,1,\cdots,n+1)$ 是 $f(x,y)$ 的 k 阶微分, 最后一项

$$R_n = \frac{1}{(n+1)!}\mathrm{d}^{n+1} f(x_0+\theta\Delta x, y_0+\theta\Delta y)$$

称为拉格朗日余项.

研究一元函数时, 我们知道带拉格朗日余项的泰勒公式中, 当 $n=0$ 就会退化成通常的拉格朗日中值定理. 令上述定理的 $n=0$, 就有

$$f(x_0+\Delta x, y_0+\Delta y) = f(x_0,y_0) + \mathrm{d}f(x_0+\theta\Delta x, y_0+\theta\Delta y).$$

展开右侧的微分, 就是

$$f(x_0+\Delta x, y_0+\Delta y) = f(x_0,y_0) + f_x(x_0+\theta\Delta x, y_0+\theta\Delta y)\,\mathrm{d}x + f_y(x_0+\theta\Delta x, y_0+\theta\Delta y)\,\mathrm{d}y.$$

这恰是定理 6.7.1 的结果. 所以, **无论一元函数还是多元函数, 其带拉格朗日余项的泰勒公式都是拉格朗日中值定理的高阶推广.**

证明. 证明过程完全类似于定理 6.7.1. 构造一元函数 $\varphi(t) = f(x_0+t\Delta x, y_0+t\Delta y)$, 则对 $\varphi(t)$ 在区间 $[0,1]$ 内应用一元函数的带拉格朗日余项的泰勒公式, 则存在 $0<\theta<1$, 使得

$$\varphi(1) = \varphi(0) + \frac{1}{1!}\varphi'(0) + \frac{1}{2!}\varphi''(0) + \cdots + \frac{1}{n!}\varphi^{(n)}(0) + \frac{1}{(n+1)!}\varphi^{(n+1)}(\theta).$$

然后, 我们依次计算上述右侧的每一项的表达式.

1. 证明 $\varphi'(0) = \mathrm{d}f(x_0,y_0)$. 显然, 由链式法则有 (牢记 $\mathrm{d}x \equiv \Delta x, \mathrm{d}y \equiv \Delta y$)

$$\begin{aligned}\varphi'(t) &= \frac{\partial f}{\partial x}(x_0+t\Delta x, y_0+t\Delta y)\,\Delta x + \frac{\partial f}{\partial y}(x_0+t\Delta x, y_0+t\Delta y)\,\Delta y \\ &= \left(\mathrm{d}x\frac{\partial}{\partial x} + \mathrm{d}y\frac{\partial}{\partial y}\right) f\bigg|_{(x_0+t\Delta x, y_0+t\Delta y)} \\ &= \mathrm{d}f(x_0+t\Delta x, y_0+t\Delta y).\end{aligned}$$

   故代入 $t=0$ 有 $\varphi'(0) = \mathrm{d}f(x_0,y_0)$.

2. 证明 $\varphi''(0) = \mathrm{d}^2 f(x_0,y_0)$. 继续对 $\varphi'(t)$ 求导:

(a) 对 $\varphi'(t)$ 的第一项进行求导得到

$$\frac{\partial^2 f}{\partial x^2}\left(x_0+t\Delta x, y_0+t\Delta y\right)\Delta x^2+\frac{\partial^2 f}{\partial x\partial y}\left(x_0+t\Delta x, y_0+t\Delta y\right)\Delta x\Delta y.$$

(b) 对 $\varphi'(t)$ 的第二项进行求导得到

$$\frac{\partial^2 f}{\partial y\partial x}\left(x_0+t\Delta x, y_0+t\Delta y\right)\Delta y\Delta x+\frac{\partial^2 f}{\partial y^2}\left(x_0+t\Delta x, y_0+t\Delta y\right)\Delta y^2.$$

将两部分相加, 得到二阶导数

$$\begin{aligned}\varphi''(t)&=\frac{\partial^2 f}{\partial x^2}\left(x_0+t\Delta x, y_0+t\Delta y\right)\Delta x^2+2\frac{\partial^2 f}{\partial x\partial y}\left(x_0+t\Delta x, y_0+t\Delta y\right)\Delta x\Delta y\\&\quad+\frac{\partial^2 f}{\partial y^2}\left(x_0+t\Delta x, y_0+t\Delta y\right)\Delta y^2\\&=\left(\mathrm{d}x\frac{\partial}{\partial x}+\mathrm{d}y\frac{\partial}{\partial y}\right)^2 f\bigg|_{(x_0+t\Delta x, y_0+t\Delta y)}\\&=\mathrm{d}^2 f\left(x_0+t\Delta x, y_0+t\Delta y\right).\end{aligned}$$

故代入 $t=0$ 有 $\varphi''(0)=\mathrm{d}^2 f\left(x_0, y_0\right)$.

3. 递推地得到, 对于任意正整数 k, 我们发现

$$\varphi^{(k)}(t)=\left(\mathrm{d}x\frac{\partial}{\partial x}+\mathrm{d}y\frac{\partial}{\partial y}\right)^k f\bigg|_{(x_0+t\Delta x, y_0+t\Delta y)}=\mathrm{d}^k f\left(x_0+t\Delta x, y_0+t\Delta y\right).$$

故代入 $t=0$ 有

$$\varphi^{(k)}(0)=\mathrm{d}^k f\left(x_0, y_0\right), \quad 当\ k=1,2,\cdots,n,$$

而且有

$$\varphi^{(n+1)}(\theta)=\mathrm{d}^{n+1} f\left(x_0+\theta\Delta x, y_0+\theta\Delta y\right).$$

将这些结果代入 $\varphi(t)$ 的泰勒公式即得要证的结论. 证毕. □

注意 同样地, 在定理 6.7.3 中, 我们就是把

$$\mathrm{d}x^k=(\Delta x)^k=\left(x-x_0\right)^k, \quad \mathrm{d}y^k=(\Delta y)^k=\left(y-y_0\right)^k,$$

给藏到了微分符号 $\mathrm{d}^k f$ 里面当中. 而这里的 $\mathrm{d}x, \mathrm{d}y$ 不是可以任意变动的, 它要和等式左侧中的 $\Delta x, \Delta y$ 和拉格朗日余项中的 $\Delta x, \Delta y$ 保持一致.

定理 6.7.3 在二元函数的数值计算上有重要应用价值. 二元函数的带拉格朗日余项的泰勒公式的最后一项

$$R_n=\frac{1}{(n+1)!}\mathrm{d}^{n+1} f\left(x_0+\theta\Delta x, y_0+\theta\Delta y\right)$$

为拉格朗日余项, 它可以用 $n+1$ 阶偏导数来估计. 比如, 若假定 $f(x,y)$ 的 $n+1$ 阶偏导数有界, 即存在常数 $M>0$, 使得[29]

$$\left|\frac{\partial^{n+1} f}{\partial x^{(n+1)-k}\partial y^k}\right|\le M, \quad k=0,1,\cdots,n+1,$$

[29] 准确地讲, 是指对于定义域内所有的点 (x,y), 都有 $\left|\frac{\partial^{n+1} f}{\partial x^{(n+1)-k}\partial y^k}(x,y)\right|\le M, \quad k=0,1,\cdots,n+1,$

那么我们有

$$
\begin{aligned}
|R_n| &= \frac{1}{(n+1)!}\left|\mathrm{d}^{n+1} f\left(x_0+\theta \Delta x, y_0+\theta \Delta y\right)\right| \\
&= \frac{1}{(n+1)!}\left|\sum_{k=0}^{n+1}\binom{n+1}{k} \frac{\partial^{n+1} f}{\partial x^{(n+1)-k} \partial y^k}\left(x_0+\theta \Delta x, y_0+\theta \Delta y\right) \cdot \ \mathrm{d} x^{(n+1)-k} \ \mathrm{d} y^k\right| \\
&\leq \frac{M}{(n+1)!}\left|\sum_{k=0}^{n+1}\binom{n+1}{k} \ \mathrm{d} x^{(n+1)-k} \ \mathrm{d} y^k\right| \\
&\leq \frac{M}{(n+1)!} \sum_{k=0}^{n+1}\binom{n+1}{k}|\mathrm{d} x|^{(n+1)-k}|\mathrm{d} y|^k \\
&= \frac{M}{(n+1)!}(|\mathrm{d} x|+|\mathrm{d} y|)^{n+1} \\
&= \frac{M}{(n+1)!}(|\Delta x|+|\Delta y|)^{n+1}.
\end{aligned}
$$

最终有

$$|R_n| \leq \frac{M}{(n+1)!}(|\Delta x|+|\Delta y|)^{n+1}.$$

令 $\rho=\sqrt{(\Delta x)^2+(\Delta y)^2}$, 则 $\dfrac{|\Delta x|}{\rho} \leq 1, \dfrac{|\Delta y|}{\rho} \leq 1$. 于是, 得到

$$|R_n| \leq \frac{M}{(n+1)!} \rho^{n+1}\left(\frac{|\Delta x|}{\rho}+\frac{|\Delta y|}{\rho}\right)^{n+1} \leq \frac{2^{n+1} M}{(n+1)!} \rho^{n+1}.$$

当固定 $\Delta x, \Delta y$ 时, ρ 是一个常数, 于是由上式不难看出

$$R_n \rightarrow 0 \quad(n \rightarrow \infty).$$

这就是说, 在上述条件下, 当带拉格朗日余项的泰勒公式的项数充分增大时, 其余项则可以任意小.

另外, 当带拉格朗日余项的泰勒公式的项数 n 固定而令 $\rho \rightarrow 0$ 时, 有

$$\frac{|R_n|}{\rho^n} \leq \frac{2^{n+1} M \rho^{n+1}}{(n+1)!\rho^n}=\frac{2^{n+1} M}{(n+1)!} \rho \rightarrow 0,$$

即

$$R_n=o\left(\rho^n\right) \quad(\rho \rightarrow 0).$$

于是, 我们便得到二元函数的带佩亚诺余项的泰勒公式:

$$f\left(x_0+\Delta x, y_0+\Delta y\right)=\sum_{k=0}^n \frac{1}{k!} \mathrm{d}^k f\left(x_0, y_0\right)+o\left(\rho^n\right), \quad(\rho \rightarrow 0),$$

其中 $\rho=\sqrt{(\Delta x)^2+(\Delta y)^2}$.

根据高阶微分的定义, 不难看出

$$\sum_{k=0}^n \frac{1}{k!} \mathrm{d}^k f\left(x_0, y_0\right)=f\left(x_0, y_0\right)+\frac{1}{1!} \mathrm{d} f\left(x_0, y_0\right)+\frac{1}{2!} \mathrm{d}^2 f\left(x_0, y_0\right)+\cdots+\frac{1}{n!} \mathrm{d}^n f\left(x_0, y_0\right)$$

是 Δx 与 Δy 的 n 次二元多项式 (其最高次为 n, 但同时可能包含着低次项), 其系数是 $f(x, y)$ 在点 (x_0, y_0) 处的各阶偏导数. 这个多项式称作 (n 阶) 泰勒多项式.

1. 当 $n=1$ 时, $f(x,y)$ 在点 (x_0,y_0) 处的一阶泰勒多项式是

$$f\left(x_0,y_0\right)+f_x\left(x_0,y_0\right)\Delta x+f_y\left(x_0,y_0\right)\Delta y.$$

2. 当 $n=2$ 时, $f(x,y)$ 在点 (x_0,y_0) 处的二阶泰勒多项式是

$$\begin{aligned}&f\left(x_0,y_0\right)+f_x\left(x_0,y_0\right)\Delta x+f_y\left(x_0,y_0\right)\Delta y\\&\quad+\frac{1}{2}\left[f_{xx}\left(x_0,y_0\right)(\Delta x)^2+2f_{xy}\left(x_0,y_0\right)\Delta x\Delta y+f_{yy}\left(x_0,y_0\right)(\Delta y)^2\right].\end{aligned}$$

3. 当 $n=3$ 时, $f(x,y)$ 在点 (x_0,y_0) 处的三阶泰勒多项式是

$$\begin{aligned}&f\left(x_0,y_0\right)+f_x\left(x_0,y_0\right)\Delta x+f_y\left(x_0,y_0\right)\Delta y\\&\quad+\frac{1}{2}\left[f_{xx}\left(x_0,y_0\right)(\Delta x)^2+2f_{xy}\left(x_0,y_0\right)\Delta x\Delta y+f_{yy}\left(x_0,y_0\right)(\Delta y)^2\right]\\&\quad+\frac{1}{6}\left[f_{xxx}\left(x_0,y_0\right)(\Delta x)^3+3f_{xxy}\left(x_0,y_0\right)(\Delta x)^2\Delta y+3f_{xyy}\left(x_0,y_0\right)\Delta x(\Delta y)^2+f_{yyy}\left(x_0,y_0\right)(\Delta y)^3\right].\end{aligned}$$

例 6.7.1 求函数 $f(x,y)=\sin\left(\frac{\pi}{2}x^2y\right)$ 在点 $(1,1)$ 处的二阶泰勒多项式及带佩亚诺余项的泰勒公式. 其近似效果见 多元函数的泰勒多项式例 1 | Desmos.

证明. 先计算 $f(x,y)$ 在点 $(1,1)$ 处的函数值及一阶和二阶偏导数:

$$\begin{aligned}&f(1,1)=1,\\&\left.\frac{\partial f}{\partial x}\right|_{(1,1)}=0,\quad\left.\frac{\partial f}{\partial y}\right|_{(1,1)}=0,\\&\left.\frac{\partial^2 f}{\partial x^2}\right|_{(1,1)}=-\pi^2,\quad\left.\frac{\partial^2 f}{\partial x\partial y}\right|_{(1,1)}=-\frac{\pi^2}{2},\quad\left.\frac{\partial^2 f}{\partial y^2}\right|_{(1,1)}=-\frac{\pi^2}{4}.\end{aligned}$$

因此, 若令 $\Delta x=x-1,\Delta y=y-1$, 则有带佩亚诺余项的二阶泰勒公式

$$f(1+\Delta x,1+\Delta y)=1+\frac{1}{2!}\left[-\pi^2(\Delta x)^2-\frac{\pi^2}{2}2\Delta x\Delta y+-\frac{\pi^2}{4}(\Delta y)^2\right]+o\left(\rho^2\right),$$

即

$$\sin\left(\frac{\pi}{2}x^2y\right)=1-\frac{\pi^2}{2}\left[(x-1)^2+(x-1)(y-1)+\frac{1}{4}(y-1)^2\right]+o\left((x-1)^2+(y-1)^2\right)\quad(x\to1,y\to1).$$

其中, 函数 $1-\frac{\pi^2}{2}\left[(x-1)^2+(x-1)(y-1)+\frac{1}{4}(y-1)^2\right]$ 就是在点 $(1,1)$ 处的二阶泰勒多项式. □

像上一题目所展示的, **我们通常需要求二元函数在某点 (x_0,y_0) 处的泰勒多项式.** 在定理 6.7.3 中, 令 $x=x_0+\Delta x,y=y_0+\Delta y$, 则有

$$f\left(x,y\right)=\sum_{k=0}^{n}\frac{1}{k!}\mathrm{d}^kf\left(x_0,y_0\right)+R_n,$$

注意, 右侧的微分展开后, 应代入 $\Delta x=x-x_0,\Delta y=y-y_0$. 最后的泰勒多项式 $\sum_{k=0}^{n}\frac{1}{k!}\mathrm{d}^kf\left(x_0,y_0\right)$ 就是关于 x,y 的函数表达式. 再看下题.

例 6.7.2 求函数 $f(x,y)=x^2y-xy^2$ 在点 $(2,1)$ 处的三阶泰勒公式. 其近似效果见 多元函数的泰勒多项式例 2 | Desmos.

证明. 先计算 $f(x,y)$ 的一阶, 二阶和三阶偏导数:

$$\begin{aligned}
&f_x(x,y)=2xy-y^2, \quad f_y(x,y)=x^2-2xy;\\
&f_{xx}(x,y)=2y, \quad f_{xy}(x,y)=2x-2y, \quad f_{yy}(x,y)=-2x;\\
&f_{xxx}(x,y)=0, \quad f_{xxy}(x,y)=2, \quad f_{xyy}(x,y)=-2, \quad f_{yyy}(x,y)=0.
\end{aligned}$$

再求出 $f(x,y)$ 及其一阶, 二阶和三阶偏导数在点 $(2,1)$ 处的值:

$$\begin{aligned}
&f(2,1)=2;\\
&f_x(2,1)=3, \quad f_y(2,1)=0;\\
&f_{xx}(2,1)=2, \quad f_{xy}(2,1)=2, \quad f_{yy}(2,1)=-4;\\
&f_{xxx}(x,1)=0, \quad f_{xxy}(2,1)=2, \quad f_{xyy}(2,1)=-2, \quad f_{yyy}(2,1)=0.
\end{aligned}$$

所以在点 $(2,1)$ 处的三阶泰勒公式是

$$\begin{aligned}
f(x,y)=2+3(x-2)&+\frac{1}{2!}\left[2(x-2)^2+2\times 2(x-2)(y-1)-2(y-1)^2\right]\\
&+\frac{1}{3!}\left[3\times 2(x-2)^2(y-1)+3\times(-2)(x-2)(y-1)^2\right]\\
&=2+3(x-2)+(x-2)^2+2(x-2)(y-1)-2(y-1)^2\\
&\quad+(x-2)^2(y-1)-(x-2)(y-1)^2.
\end{aligned}$$

上式最后化简后, 会发现就是 x^2y-xy^2.[30] □

多元函数的泰勒公式

设 $D\subset\mathbf{R}^3$ 为一个区域, 而函数 $f(x,y,z)\in C^{n+1}(D)$, 又设两点

$$P_0\left(x_0,y_0,z_0\right)\in D, \quad P_1\left(x_0+\Delta x,y_0+\Delta y,z_0+\Delta z\right)\in D,$$

并且点 P_0 与 P_1 之间的连线 $P_0P_1\subset D$. 则存在 $0<\theta<1$, 使得

$$f(P_1)=\sum_{k=0}^{n}\frac{1}{k!}\mathrm{d}^kf\left(P_0\right)+\frac{1}{(n+1)!}\mathrm{d}^{n+1}f\left(P_\theta\right),$$

其中, $P_\theta=(x_0+\theta\Delta x,y_0+\theta\Delta y,z_0+\theta\Delta z)$ 是线段 P_0P_1 上的某个点. 这就是三元函数的泰勒公式. 若将其推广到任意 m 元函数.

[30]一般地, 可以证明: 一个 n 阶多项式函数在任何一个点处的 n 泰勒多项式就是函数自身.

定理 6.7.4: m 元函数的（带拉格朗日余项的）泰勒公式

设 m $(m \geq 1)$ 元函数 $z = f(x_1, x_2, \ldots, x_m)$ 在区域 $D \subseteq \mathbf{R}^m$ 内具有连续的一阶偏导数. 设 D 中有两个点 (m 维向量) P_0 与 P_1, 并且连接点 P_0 与 P_1 的线段 $P_0P_1 \subset D$. 则存在 θ $(0 < \theta < 1)$, 使得

$$f(P_1) = \sum_{k=0}^{n} \frac{1}{k!} \mathrm{d}^k f\left(P_0\right) + \frac{1}{(n+1)!} \mathrm{d}^{n+1} f\left(P_\theta\right),$$

其中, $P_\theta := P_0 + \theta \Delta \mathbf{x}$, $\Delta \mathbf{x} := P_1 - P_0$; $\mathrm{d}^k f$ $(k = 0, 1, \cdots, n+1)$ 是 $z = f(x_1, x_2, \ldots, x_m)$ 的 k 阶微分:

$$\mathrm{d}^k f = \left(\sum_{i=1}^{m} \mathrm{d} x_i \frac{\partial}{\partial x_i}\right)^k f.$$

泰勒多项式的唯一性

像一元函数的情况一样, 多元函数的泰勒多项式也有唯一性定理成立.

定理 6.7.5: 二元函数的泰勒多项式的唯一性定理

若 $P_n(\Delta x, \Delta y)$ 是 Δx 与 Δy 的 n 次多项式, 并且有

$$f\left(x_0 + \Delta x, y_0 + \Delta y\right) = P_n(\Delta x, \Delta y) + o\left(\rho^n\right), \quad (\rho \to 0),$$

其中 $\rho = \sqrt{(\Delta x)^2 + (\Delta y)^2}$. 则 $P_n(\Delta x, \Delta y)$ 是函数 $z = f(x, y)$ 在点 (x_0, y_0) 处的泰勒多项式.

证明省略. 因此, 一个函数的泰勒多项式可由其他途径求得, 而不一定非计算各阶偏导数不可.


例 6.7.3 在点 $(0, 0)$ 的邻域内, 将函数 $f(x, y) = \mathrm{e}^x \cos y$ 按带佩亚诺余项的泰勒公式展开至二次项. 其近似效果见 多元函数的泰勒多项式例 3 | Desmos.

证明. 已知两个一元函数在 0 附近的二阶泰勒展开分别是

$$\mathrm{e}^x = 1 + x + \frac{1}{2} x^2 + o\left(x^2\right), \quad x \to 0,$$
$$\cos y = 1 - \frac{1}{2} y^2 + o\left(y^2\right), \quad y \to 0.$$

因而由上两式相乘可得

$$\begin{aligned}
\mathrm{e}^x \cos y &= \left(1 + x + \frac{1}{2} x^2 + o\left(x^2\right)\right)\left(1 - \frac{1}{2} y^2 + o\left(y^2\right)\right) \\
&= 1 - \frac{1}{2} y^2 + o\left(y^2\right) \\
&\quad + x - \frac{1}{2} x y^2 + x o\left(y^2\right) \\
&\quad + \frac{1}{2} x^2 - \frac{1}{4} x^2 y^2 + \frac{1}{2} x^2 o\left(y^2\right) \\
&\quad + o\left(x^2\right) - o\left(x^2\right) - \frac{1}{2} y^2 o\left(x^2\right) o\left(y^2\right),
\end{aligned}$$

可以证明, 当 $\rho \to 0$ 时, 上面的项满足

- $o\left(x^2\right)=o\left(\rho^2\right), o\left(y^2\right)=o\left(\rho^2\right), o\left(x^2\right)o\left(y^2\right)=o\left(\rho^2\right)$;
- $xy^2=o\left(\rho^2\right), x^2y^2=o\left(\rho^2\right)$;
- $xo\left(y^2\right)=o\left(\rho^2\right), x^2o\left(y^2\right)=o\left(\rho^2\right), y^2o\left(y^2\right)=o\left(\rho^2\right)$;

整理后的结果为

$$\mathrm{e}^x\cos y=1+x+\frac{1}{2}x^2-\frac{1}{2}y^2+o\left(\rho^2\right),\quad \rho\to 0.$$

由泰勒多项式的唯一性知上式即为所求. □

注意 不建议考试中用这种方法. 非常容易出错. 中途需要验证很多二元函数的极限.

例 6.7.4 求函数 $f(x,y)=\mathrm{e}^{-x}\ln(1+x+y)$ 在点 $(0,0)$ 处的二阶带佩亚诺余项的泰勒公式, 并利用此公式求 $f_{xx}(0,0)$ 及 $f_{xy}(0,0)$. 其近似效果见 多元函数的泰勒多项式例 4 | Desmos.

证明. 当 $x\to 0$ 时, 有

$$\mathrm{e}^{-x}=1-x+\frac{1}{2!}x^2+o\left(x^2\right).$$

当 $(x,y)\to(0,0)$ 时, 有 (将 $x+y$ 看作一个整体)

$$\ln(1+x+y)=(x+y)-\frac{1}{2}(x+y)^2+o\left((x+y)^2\right).$$

令 $\rho=\sqrt{x^2+y^2}$, 注意到当 $\rho\to 0$ 时, $o\left(x^2\right)=o\left(\rho^2\right), o\left((x+y)^2\right)=o\left(\rho^2\right)$[31], 则

$$\begin{aligned}\mathrm{e}^{-x}\ln(1+x+y)&=\left(1-x+\frac{1}{2!}x^2+o\left(x^2\right)\right)\times\left[(x+y)-\frac{1}{2}(x+y)^2+o\left((x+y)^2\right)\right]\\&=(x+y)-\frac{1}{2}(x+y)^2-x(x+y)+o\left(\rho^2\right)\\&=x+y-\frac{3}{2}x^2-2xy-\frac{1}{2}y^2+o\left(\rho^2\right).\end{aligned}$$

由泰勒多项式的唯一性知上式即为所求. 根据上式有

$$\frac{1}{2!}f_{xx}(0,0)=-\frac{3}{2},\quad \frac{1}{2!}\cdot 2f_{xy}(0,0)=-2,$$

于是得到

$$f_{xx}(0,0)=-3,\quad f_{xy}(0,0)=-2.$$

□

注意 不建议考试中用这种方法. 非常容易出错. 中途需要验证很多二元函数的极限.

[31] 需要利用不等式 $\frac{(x+y)^2}{x^2+y^2}\le 2$. 提示: $\frac{(x+y)^2}{x^2+y^2}=\frac{x^2+2xy+y^2}{x^2+y^2}=1+\frac{2xy}{x^2+y^2}$, 而 $\frac{2xy}{x^2+y^2}\le 1$ 因为 $(x-y)^2\ge 0$.

6.8 隐函数存在定理 • 逆函数存在定理

第二章曾涉及一些具体二元函数方程 $F(x,y)=0$ 所确定隐函数的求导方法. 但当时我们并没有讨论隐函数的存在性问题, 毕竟如果隐函数都不存在的话, 那么求隐函数的导数是滑稽的. 现在我们有足够的知识储备来讨论隐函数的存在性问题, 及其存在的话如果求导.

6.8.1 隐函数的类型

我们根据方程的类型做如下关于隐函数问题类型的定义和讨论.

1-1 型隐函数

所谓的由一个二元函数方程[32]

$$F:\mathbf{R}^2\to\mathbf{R},\quad F(x;y)=0,$$

所确定的隐函数, 是指这样的一元函数 $y=f(x)$, 其存在某个定义域 (开区间) $D\subseteq\mathbf{R}$, 使得

$$F(x;f(x))\equiv 0,\quad \forall x\in D.$$

我们好奇隐函数 $y=f(x)$ 的存在条件, 若存在的话如何求其导数.

2-1 型隐函数

隐函数的概念也不仅限于二元函数方程, 还可以考虑一个三元及三元以上函数的方程. 比如一个三元函数方程

$$F:\mathbf{R}^3\to\mathbf{R},\quad F(x,y;z)=0.$$

这时, 所确定的隐函数 $z=f(x,y)$ 是一个二元函数. 所谓的 $z=f(x,y)$ 是上述方程所确定的一个隐函数, 是指存在某个区域 $D\subseteq\mathbf{R}^2$, 使得

$$F(x,y;f(x,y))\equiv 0,\quad \forall (x,y)\in D.$$

我们好奇隐函数 $z=f(x,y)$ 的存在条件, 若存在的话如何求其两个偏导数.

1-2 型隐函数

问题的提法还可以更为一般. 前面只涉及一个方程, 现在考虑方程组的情况 (至少两个方程及以上). 比如

$$\begin{cases} F(x;u,v)=0,\\ G(x;u,v)=0.\end{cases}$$

如果 x 的值给定, 那么上述方程组往往能解出一对 (u,v) 的值. 这时, 所确定的隐函数

$$x\mapsto(u,v),\quad u=u(x),\quad v=v(x),$$

[32]通常方程都是非线性方程. 我们说一个方程是非线性方程, 是指方程关于未知变量的表达式不是一个线性函数 (即不是未知量的线性组合). 如果是一个线性方程, 其所确定的隐函数是非常简单的, 比如 $ax+by=0$, 则隐函数 $y=-\dfrac{a}{b}x$ 的存在条件是 $b\neq 0$, 且其导数是 $-\dfrac{a}{b}$. 更多元情况的线性方程也一样容易讨论.

是一个一元到二元函数. 所谓的 $x \mapsto (u,v)$ 是上述方程组所确定的一个隐函数, 是指存在某个区间 $D \subseteq \mathbf{R}$ 内使得

$$\begin{cases} F(x;u(x),v(x)) \equiv 0, \\ G(x;u(x),v(x)) \equiv 0, \end{cases} \quad \forall x \in D.$$

我们好奇隐函数 $u = u(x), v = v(x)$ 的存在条件, 若存在的话如何求其导数. 值得注意的是, 方程组的情况仍可用单个函数来表示. 若定义

$$H : \mathbf{R}^3 \to \mathbf{R}^2, \quad H(x;u,v) = (F(x;u,v), G(x;u,v)),$$

则方程组就写做

$$H(x;u,v) = \mathbf{0} \in \mathbf{R}^2.$$

我们希望找到隐函数

$$\mathbf{g} : D \subseteq \mathbf{R} \to \mathbf{R}^2, \quad \mathbf{g}(x) = (u(x), v(x)),$$

使得

$$H(x;\mathbf{g}(x)) \equiv \mathbf{0}, \quad \forall x \in D.$$

2-2 型隐函数

再比如方程组

$$\begin{cases} F(x,y;u,v) = 0, \\ G(x,y;u,v) = 0. \end{cases}$$

如果给定一对 (x,y) 的值, 那么上述方程组往往能解出一对 (u,v) 的值. 这时, 所确定的隐函数

$$(x,y) \mapsto (u,v), \quad u = u(x,y), \quad v = v(x,y),$$

是一个二元到二元函数. 所谓的 $(x,y) \mapsto (u,v)$ 是上述方程组所确定的一个隐函数, 是指存在某个区域 $D \subseteq \mathbf{R}^2$ 内使得

$$\begin{cases} F(x,y;u(x,y),v(x,y)) \equiv 0, \\ G(x,y;u(x,y),v(x,y)) \equiv 0, \end{cases} \quad \forall (x,y) \in D.$$

我们好奇隐函数 $u = u(x,y), v = v(x,y)$ 的存在条件, 若存在的话如何求其各自的偏导数. 此时的方程组仍可用单个函数来表示. 若定义

$$H : \mathbf{R}^4 \to \mathbf{R}^2, \quad H(x,y;u,v) = (F(x,y;u,v), G(x,y;u,v)),$$

则方程组就写做

$$H(x,y;u,v) = \mathbf{0} \in \mathbf{R}^2.$$

我们希望找到隐函数

$$\mathbf{g} : D \subseteq \mathbf{R}^2 \to \mathbf{R}^2, \quad \mathbf{g}(x,y) = (u(x,y), v(x,y)),$$

使得

$$H(x,y;\mathbf{g}(x,y)) \equiv \mathbf{0}, \quad \forall x \in D.$$

n-m 型隐函数

一般的隐函数存在问题可以通过如下的向量语言来表达. 设 n, m 是任意正整数. 设 $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{R}^n$ 为第一部分的向量; $\mathbf{y} = (y_1, \ldots, y_m) \in \mathbf{R}^m$ 为第二部分的向量. 给定**一个变量个数大于方程个数的方程(组)**

$$H : \mathbf{R}^{n+m} \to \mathbf{R}^m, \quad H(\mathbf{x}; \mathbf{y}) = \mathbf{0}.$$

对于上述方程, 如果代入一个具体的 $\mathbf{x} \in \mathbf{R}^n$ 到方程 $H(\mathbf{x}; \mathbf{y}) = \mathbf{0}$ 中, 则得到一个含有 m 个待求解变量 y_i 的 m 个联立方程, 这通常能解出一个确定的 $\mathbf{y}$. 于是通过这种方式解出来的 $\mathbf{y}$ 是依赖于 $\mathbf{x}$ 的, 这是一个函数关系, 我们定义为 $\mathbf{g}(\mathbf{x}) = \mathbf{y}$. 一般地, 所谓 "**隐函数问题**", 就是我们希望找到一个函数 $\mathbf{g} : D \subseteq \mathbf{R}^n \to \mathbf{R}^m$ 使得方程

$$H(\mathbf{x}; \mathbf{g}(\mathbf{x})) \equiv \mathbf{0}, \quad \forall \mathbf{x} \in D.$$

隐函数问题具体包含了 2 点:

1. 隐函数 $\mathbf{g} = (g_1, \ldots, g_m)$ 是否存在? 为了使其存在, 需要什么条件? 我们不期盼能获得得到隐函数 $\mathbf{g} = (g_1, \ldots, g_m)$ 的显式表达式, 因为多数情况都是不能可以的. 所以知道其存在性就足够了.

2. 如果存在的话, 如何求其各个成分的 n 元函数 $g_i(x_1, \ldots, x_n)$ $(\forall i = 1, \ldots, m)$ 的偏导数呢? (若 $n = 1$ 的话, 就是求导数)

通过方程的维度 n, m 来确定隐函数问题的类型非常关键. 比如前讨论的四种情况 (1-1 型; 2-1 型; 1-2 型; 2-2 型) 就分别对应的 $n = 1, m = 1$; $n = 2, m = 1$; $n = 1, m = 2$; $n = 2, m = 2$. 下面是一些技巧心得:

- 方程组中总的变量的个数一定要多于方程的个数, 这是一切的前提. 也就说, 函数 H 一定是从高维空间 ($n + m$ 维) 到低维空间 (m 维) 的映射.

- 函数 H 的输入的个数减去输出的个数, 就是隐函数的输入的个数.

- 给定的方程组包含几个 (m 个) 方程, 则隐函数的输出就是几个 (m 个). 或者说, 函数 H 输出的个数等于隐函数输出的个数.

6.8.2　一个方程的隐函数存在定理 (m=1)

下面, 我们只讨论一些特定情况的隐函数的存在定理. 一般的 n-m 型隐函数的存在性定理因为缺乏线性代数知识很难清晰地描述.

定理 6.8.1: 隐函数存在定理 1 (1-1 型隐函数)

设函数 $F(x,y)$ 在点 $P_0(x_0,y_0)$ 的某个邻域内有定义, 并且满足

1. $F(x_0,y_0)=0$,
2. $F_x(x,y)$ 及 $F_y(x,y)$ 连续 (有连续的一阶偏导数),
3. $F_y(x_0,y_0)\neq 0$,

则

1. 在点 x_0 的某个开区间 $D\subseteq \mathbf{R}$ 内存在唯一的隐函数 $y=f(x)$, 即
$$F(x,f(x))\equiv 0,\quad \forall x\in D,$$
使得 $y_0=f(x_0)$.
2. 而且, 这样子的隐函数 $y=f(x)$ 在 D 内有连续的导数, 其计算公式是
$$\frac{\mathrm{d}f}{\mathrm{d}x}=-\frac{F_x(x,y)}{F_y(x,y)}.$$

我们略去定理的证明. 条件 $F_y(x_0,y_0)\neq 0$ 对于从方程 $F(x,y)=0$ 中将 y 解成 x 的函数而言是十分重要的. 其实, 这也可以从一个具体例子中看出. 我们考虑一个线性方程

$$ax+by=0,$$

其中 a 与 b 为常数. 这时, 我们的函数为 $F(x,y)=ax+by$. 要想从这个方程中将 y 解成 x 的函数, 即 $y=-\dfrac{a}{b}x$, 则要求 $b\neq 0$, 而 b 恰好是 $F_y(x,y)$. 另外, 隐函数应当满足条件 $y_0=f(x_0)$, 则否可能不唯一. 比如考虑, $x^2+y^2=1$ 确定的隐函数, 且 $-1<x_0<1$. 显然, 任何 x_0 的邻域内存在位于 x 轴上下的两个隐函数.

隐函数存在定理不仅证明了隐函数的存在性, 还提供了隐函数的求导公式. **(隐式求导)** 对方程

$$F(x,f(x))\equiv 0,$$

两边同时关于 x 求导 (如果用 y 来表示 $f(x)$ 的话, 切记把 y 看作是 x 的函数), 使用链式法则得到

$$\frac{\mathrm{d}}{\mathrm{d}x}F(x,f(x))=F_x(x,f(x))+F_y(x,f(x))\cdot f'(x)=0,$$

我们可以解得

$$f'(x)=-\frac{F_x(x,f(x))}{F_y(x,f(x))}=-\frac{F_x(x,y)}{F_y(x,y)},$$

注意, 隐函数 $f(x)$ 的显式表达式通常是写不出来的, 所以我们直接用 $y=f(x)$ 替代. 以前我们就讲过, 隐函数的导数的表达式中包含着因变量 y 是非常正常的. 那么我们就真的算不出来导数了吗? 非也. 我们完全可以计算一个具体的点 $x_1\in D$ 的导数

$$f'(x_1)=-\frac{F_x(x_1,y_1)}{F_y(x_1,y_1)},$$

其中 y_1 的值是通过解方程 $F(x_1,y)=0$ 得到的 $y=y_1$. 注意这个关于 y 的方程一般是没有解析解的 (对比二元一次方程是典型的有解析解的方程), 我们只能通过其他办法获得一个近似解 $y\approx y_1$. 然后再把近似解放

入上述公式中获得导数值. 因此, 隐函数的导数计算往往是通过数值办法, 于是不可避免的有计算误差. 但这是应用中往往是能接受的.

例 6.8.1 求方程

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \quad (a,b>0)$$

在点 $\left(\frac{a}{\sqrt{2}},\frac{b}{\sqrt{2}}\right)$ 附近所确定隐函数 $y=y(x)$ 的导数[33]及 $y'\left(\frac{a}{\sqrt{2}}\right)$.

证明. 令 $F(x,y)=\frac{x^2}{a^2}+\frac{y^2}{b^2}-1$, 则

$$F\left(\frac{a}{\sqrt{2}},\frac{b}{\sqrt{2}}\right)=0,$$

并且 $F_x(x,y)=\frac{2x}{a^2}$, $F_y(x,y)=\frac{2y}{b^2}$ 是连续函数. 又

$$F_y\left(\frac{a}{\sqrt{2}},\frac{b}{\sqrt{2}}\right)=\frac{\sqrt{2}}{b}\neq 0.$$

因此该方程在点 $\left(\frac{a}{\sqrt{2}},\frac{b}{\sqrt{2}}\right)$ 附近所确定的隐函数是存在的, 它满足隐函数存在定理的所有条件. 隐函数 $y=y(x)$ 是可微的, 并且有

$$y'(x)=-\frac{F_x(x,y)}{F_y(x,y)}=-\frac{b^2x}{a^2y},$$

从而

$$y'\left(\frac{a}{\sqrt{2}}\right)=-\frac{b^2}{a^2}\cdot\frac{a/\sqrt{2}}{b/\sqrt{2}}=-\frac{b}{a}.$$

这便是椭圆周 $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$ 在点 $\left(\frac{a}{\sqrt{2}},\frac{b}{\sqrt{2}}\right)$ 处的切线斜率. □

隐函数存在定理可以推广到三元及三元以上函数的情况. 比如, 在三元函数的情况下, 隐函数方程为

$$F(x,y,z)=0$$

我们希望从这个方程中将变量 z 解成 (x,y) 的函数. 对此, 有下面的定理. 定理的证明从略.

[33]这种类型就是第二章时学习的类型.

定理 6.8.2: 隐函数存在定理 2 (2-1 型隐函数)

设函数 $F(x,y,z)$ 在点 $P_0\left(x_0,y_0,z_0\right)$ 的某个邻域内有定义, 并且满足

1. $F\left(x_0,y_0,z_0\right)=0$,
2. $F_x(x,y,z)$ 及 $F_y(x,y,z)$ 连续 (有连续的一阶偏导数),
3. $F_z\left(x_0,y_0,z_0\right)\neq 0$,

则

1. 在点 (x_0,y_0) 的某个邻域 $D\subseteq\mathbf{R}^2$ 内存在唯一的隐函数 $z=z(x,y)$, 即
$$F(x,y,z(x,y))\equiv 0,\quad \forall(x,y)\in D,$$
使得 $z_0=z\left(x_0,y_0\right)$.
2. 而且, 这样子的隐函数 $z(x,y)$ 在 D 内有连续的一阶偏导数, 其计算公式是
$$\frac{\partial z}{\partial x}=-\frac{F_x(x,y,z)}{F_z(x,y,z)},\quad \frac{\partial z}{\partial y}=-\frac{F_y(x,y,z)}{F_z(x,y,z)}.$$

一般来说, 我们无法给出隐函数的具体表达式, 但其偏导数可以按定理 6.8.2 中提供的公式计算. 我们没有记忆该公式的必要. 看看这个公式是怎么来的: **(隐式求导)** 我们对 $F(x,y,z(x,y))\equiv 0$, 两边同时分别关于 x 和 y 进行求导, 来获得 $z(x,y)$ 的偏导数.

- 使用链式法则对 x 求导, 得到
$$\frac{\partial}{\partial x}F(x,y,z(x,y))=F_x(x,y,z(x,y))+F_z(x,y,z(x,y))\cdot\frac{\partial z}{\partial x}=0,$$
解得
$$\frac{\partial z}{\partial x}=-\frac{F_x(x,y,z)}{F_z(x,y,z)}.$$
- 同样地, 使用链式法则对 y 求导, 得到
$$\frac{\partial}{\partial y}F(x,y,z(x,y))=F_y(x,y,z(x,y))+F_z(x,y,z(x,y))\cdot\frac{\partial z}{\partial y}=0,$$
解得
$$\frac{\partial z}{\partial y}=-\frac{F_y(x,y,z)}{F_z(x,y,z)}.$$

因此, 隐函数 $z=z(x,y)$ 的偏导数公式就是定理 6.8.2 给出的. 上面的推理告诉**我们无需记忆公式, 只需要使用隐式求导办法, 就可以解出感兴趣的偏导.**

例 6.8.2 求由方程 $xy+yz+\mathrm{e}^{rz}=3$ 所确定隐函数 $z(x,y)$ 的偏导数 $\dfrac{\partial z}{\partial x}$ 及 $\dfrac{\partial z}{\partial y}$. (这中类型的题目忽略隐函数的存在性讨论, 考察的就是一个套公式)

证明. 1. 方法一. **直接套用现成的公式**. 令

$$F(x,y,z) = xy + yz + \mathrm{e}^{xz} - 3,$$

这时

$$\frac{\partial F}{\partial x} = y + z\mathrm{e}^{xz}, \quad \frac{\partial F}{\partial y} = x + z, \quad \frac{\partial F}{\partial z} = y + x\mathrm{e}^{xz}.$$

由此可见, 当考虑的某个点 (x,y,z) 满足 $y + x\mathrm{e}^{xz} \neq 0$ 时, 对于满足方程 $F(x,y,z) = 0$ 的该点 (x,y,z), 可以确定一个隐函数 $z = z(x,y)$. 由定理 6.8.2 可知

$$\frac{\partial z}{\partial x} = -\frac{F_x(x,y,z)}{F_z(x,y,z)} = -\frac{y + z\mathrm{e}^{xz}}{y + x\mathrm{e}^{xz}},$$
$$\frac{\partial z}{\partial y} = -\frac{F_y(x,y,z)}{F_z(x,y,z)} = -\frac{x + z}{y + x\mathrm{e}^{xz}}.$$

2. 方法二. **隐式求导**[34]. 在方程

$$xy + yz + \mathrm{e}^{xz} = 3$$

中将 z 视作 (x,y) 的函数, 这时该方程变成关于 x,y 的恒等式. 对该方程两端关于 x 求偏导数 (注意其中的 z 是 (x,y) 的函数), 即得

$$y + yz_x + \mathrm{e}^{xz}(z + xz_x) = 0.$$

由此解出

$$\frac{\partial z}{\partial x} = -\frac{y + z\mathrm{e}^{xz}}{y + x\mathrm{e}^{xz}}.$$

同理, 在所给的方程两端对 y 求偏导数, 可解出

$$\frac{\partial z}{\partial y} = -\frac{x + z}{y + x\mathrm{e}^{xz}}.$$

可见, 所得结果与前面相同.

3. 方法三. **利用一阶全微分的形式不变性**. 对所给方程两端求全微分[35], 得到

$$\mathrm{d}(xy) + \mathrm{d}(yz) + \mathrm{d}\mathrm{e}^{xz} = 0,$$

即

$$y\ \mathrm{d}x + x\ \mathrm{d}y + z\ \mathrm{d}y + y\ \mathrm{d}z + \mathrm{e}^{xz}(z\ \mathrm{d}x + x\ \mathrm{d}z) = 0,$$

移项后得到 $z(x,y)$ 的全微分

$$\mathrm{d}z = -\frac{y + z\mathrm{e}^{xz}}{y + x\mathrm{e}^{xz}}\ \mathrm{d}x - \frac{x + z}{y + x\mathrm{e}^{xz}}\ \mathrm{d}y,$$

从而读出了 $z(x,y)$ 的偏导数:

$$\frac{\partial z}{\partial x} = -\frac{y + z\mathrm{e}^{xz}}{y + x\mathrm{e}^{xz}}, \quad \frac{\partial z}{\partial y} = -\frac{x + z}{y + x\mathrm{e}^{xz}}.$$

□

[34]隐式求导方法是一通百通的办法.

[35]注意, 我们是对三元函数 $F(x,y,z) = xy + yz + \mathrm{e}^{xz} - 3$ 求的全微分. 所以变量 x,y,z 地位相同, 也不要把谁当作谁的函数. 这里有一个技巧: 这里求全微分 $\mathrm{d}(xy)$ 时, 严格来讲 xy 也是关于 x,y,z 的三元函数. 不过一求微分, $\mathrm{d}z$ 项就消失了. 故, 每一项的微分, 只需要计算该项中包含的变量即可. 所以, $\mathrm{d}(xy) = y\ \mathrm{d}x + x\ \mathrm{d}y$, $\mathrm{d}(yz) = z\ \mathrm{d}y + y\ \mathrm{d}z$, $\mathrm{d}(\mathrm{e}^{xz}) = \mathrm{e}^{xz}(z\ \mathrm{d}x + x\ \mathrm{d}z)$.

例 6.8.3 求由方程

$$F(x-y, y-z)=0$$

所确定隐函数 $z=z(x,y)$ 的偏导数 $\dfrac{\partial z}{\partial x}$ 及 $\dfrac{\partial z}{\partial y}$.

证明. 这个例子的计算中有一个符号记法问题. 为了避免混淆, 我们用 F_1' 表示 F 作为二元函数 $F(u,v)$ 时对其第一个自变量求偏导数, 而用 F_2' 表示对其第二个自变量求偏导数. 在这种记法下, 利用复合函数求导公式, 我们有

$$F_x=F_1', \quad F_y=-F_1'+F_2', \quad F_z=-F_2'.$$

因此

$$\frac{\partial z}{\partial x}=\frac{F_1'}{F_2'}, \quad \frac{\partial z}{\partial y}=\frac{-F_1'+F_2'}{F_2'}.$$

如果不引入 F_1' 及 F_2' 的记法, 也可以通过引入中间变量 $u=x-y$ 及 $v=y-z$ 来求解. 这时, 有

$$F_x=F_u, \quad F_y=-F_u+F_v, \quad F_z=-F_v.$$

□

例 6.8.4 设 $z(x,y)$ 是由方程 $\mathrm{e}^x\sin y+yz+\mathrm{e}^z+5=0$ 所确定的隐函数, 求 $\dfrac{\partial z}{\partial x}$ 和 $\dfrac{\partial^2 z}{\partial x\partial y}$.

证明. 1. 计算 $\dfrac{\partial z}{\partial x}$. 对所给的方程两端关于 x 求偏导数, 其中 y 视作常数, z 视作 (x,y) 的函数, 得到

$$\mathrm{e}^x\sin y+y\frac{\partial z}{\partial x}+\mathrm{e}^z\frac{\partial z}{\partial x}=0, \tag{6.10}$$

由此解出

$$\frac{\partial z}{\partial x}=-\frac{\mathrm{e}^x\sin y}{y+\mathrm{e}^z}.$$

2. 计算 $\dfrac{\partial z}{\partial y}$.[36] 对所给的方程两端再关于 y 求偏导数, 其中 x 视作常数, z 视作 (x,y) 的函数, 得到

$$\mathrm{e}^x\cos y+z+y\frac{\partial z}{\partial y}+\mathrm{e}^z\frac{\partial z}{\partial y}=0$$

由此解出

$$\frac{\partial z}{\partial y}=-\frac{\mathrm{e}^x\cos y+z}{y+\mathrm{e}^z}.$$

3. 计算 $\dfrac{\partial^2 z}{\partial x\partial y}$. 对 $\dfrac{\partial z}{\partial x}=-\dfrac{\mathrm{e}^x\sin y}{y+\mathrm{e}^z}$ 两端关于 y 求偏导数, x 视作常数, z 视作 (x,y) 的函数, 得到

$$\frac{\partial^2 z}{\partial x\partial y}=-\frac{\mathrm{e}^x\cos y\,(y+\mathrm{e}^z)-\mathrm{e}^x\sin y\left(1+\mathrm{e}^z\frac{\partial z}{\partial y}\right)}{(y+\mathrm{e}^z)^2}.$$

[36] 尽管题目没有要求计算 $\dfrac{\partial z}{\partial y}$, 但是很多问题的二阶偏导都需要全部一阶偏导的表达式.

将 $\dfrac{\partial z}{\partial y}=-\dfrac{\mathrm{e}^x\cos y+z}{y+\mathrm{e}^z}$ 代入上式, 即得

$$\frac{\partial^2 z}{\partial x\partial y}=-\frac{\mathrm{e}^x\cos y\,(y+\mathrm{e}^z)^2-\mathrm{e}^x\sin y\,(y+\mathrm{e}^z)+\mathrm{e}^z\cdot\mathrm{e}^x\sin y\,(\mathrm{e}^x\cos y+z)}{(y+\mathrm{e}^z)^3}.$$

4. 在求 $\dfrac{\partial^2 z}{\partial x\partial y}$ 时, 还可以采用下述方法: 方程 (6.10) 两端关于 y 求偏导数, 其中 x 视作常数, z 和 $\dfrac{\partial z}{\partial x}$ 都视作 (x,y) 的函数, 得到

$$\mathrm{e}^x\cos y+\frac{\partial z}{\partial x}+y\frac{\partial^2 z}{\partial x\partial y}+\mathrm{e}^z\frac{\partial z}{\partial y}\cdot\frac{\partial z}{\partial x}+\mathrm{e}^z\frac{\partial^2 z}{\partial x\partial y}=0,$$

进而解得

$$\frac{\partial^2 z}{\partial x\partial y}=-\frac{\mathrm{e}^x\cos y+\frac{\partial z}{\partial x}+\mathrm{e}^z\frac{\partial z}{\partial y}\cdot\frac{\partial z}{\partial x}}{y+\mathrm{e}^z}.$$

将前面求得的 $\dfrac{\partial z}{\partial x}=-\dfrac{\mathrm{e}^x\sin y}{y+\mathrm{e}^z}$ 和 $\dfrac{\partial z}{\partial y}=-\dfrac{\mathrm{e}^x\cos y+z}{y+\mathrm{e}^z}$ 代入上式, 化简后所得结果与前面相同.

□

6.8.3　2 x 2 线性方程组的克拉默法则

上一小节介绍了一个方程的隐函数存在定理 (m=1), 在讨论两个方程的隐函数存在定理 (m=2) 之前, 我们先介绍线性代数中的一个知识点 — 关于二元线性方程组的克拉默 (Cramer) 法则. 考虑二元线性方程组

$$a_1x+b_1y=c_1, \tag{1}$$
$$a_2x+b_2y=c_2, \tag{2}$$

其中 a_1,b_1,c_1,a_2,b_2,c_2 为常数. 此方程组可以写成 **矩阵 (matrix)** 形式:

$$\begin{bmatrix} a_1 & b_1\\ a_2 & b_2\end{bmatrix}\begin{bmatrix} x\\ y\end{bmatrix}=\begin{bmatrix} c_1\\ c_2\end{bmatrix}.$$

其中, 2 行 2 列 (也叫 2 x 2) 的矩阵

$$\begin{bmatrix} a_1 & b_1\\ a_2 & b_2\end{bmatrix}$$

叫做方程组的系数矩阵, $\begin{bmatrix} c_1\\ c_2\end{bmatrix}$ 叫做方程组的常数列向量[37], $\begin{bmatrix} x\\ y\end{bmatrix}$ 是待求解的未知量的向量. **矩阵形式的写方式能够精简且准确地表达一个线性方程组的全部信息.** 求解该方程的一般思路是: 通过行变换消去其中一个变量, 然后解出另一个变量. 具体操作如下:

1. 假设我们想求解 x, 可以通过消去 y 来求解. 首先, 我们将方程 (2) 乘以方程 (1) 中 y 的系数的相反数 $-b_1$, 得到

$$-b_1a_2x-b_1b_2y=-b_1c_2.$$

再将方程 (1) 乘以方程 (2) 中 y 的系数 b_2,

$$b_2a_1x+b_2b_1y=b_2c_1.$$

[37]这里仅仅是把向量竖着写而已.

然后将两式相加. 这样, y 就被消去, 得到

$$b_2a_1x - b_1a_2x = b_2c_1 - b_1c_2,$$

假设 $a_1b_2 - a_2b_1 \neq 0$, 则可以解出 x 为

$$x = \frac{c_1b_2 - c_2b_1}{a_1b_2 - a_2b_1} = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}.$$

2. 接下来, 若要解出 y, 我们需要消去 x. 类似地, 我们将方程 (1) 乘以 a_2, 将方程 (2) 乘以 $-a_1$, 然后加在一起, 得到

$$a_2b_1y - a_1b_2y = a_2c_1 - a_1c_2.$$

假设 $a_1b_2 - a_2b_1 \neq 0$, 则可以解出 y 为

$$y = \frac{a_2c_1 - a_1c_2}{a_2b_1 - a_1b_2} = \frac{a_1c_2 - a_2c_1}{a_1b_2 - a_2b_1} = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}.$$

注意到, 求解 x 和 y 时, 两者的分母都是系数矩阵的行列式. 这种通过行列式来解线性方程组的方法叫做**克拉默法则 (Cramer's rule)**. 我们将以上内容总结成下面的定理.

定理 6.8.3: 2 x 2 线性方程组的克拉默法则 (Cramer's rule)

试求解二元线性方程组

$$a_1x + b_1y = c_1,$$
$$a_2x + b_2y = c_2,$$

其中 $a_1, b_1, c_1, a_2, b_2, c_2$ 为常数. 我们引入系数矩阵所对应的行列式的符号为

$$\Delta = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}.$$

此二元线性方程组有唯一解的充要条件是其系数行列式 $\Delta \neq 0$. 当 $\Delta \neq 0$ 时, 其唯一解可以表示为

$$x = \frac{\Delta_1}{\Delta}, \quad y = \frac{\Delta_2}{\Delta},$$

其中, 分子部分的行列式为

$$\Delta_1 = \begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}, \quad \Delta_2 = \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}.$$

克拉默法则可以为我们提供方程组的唯一解, 前提是解存在. 克拉默法则的关键在于将感兴趣的变量列替换为常数列, 然后计算行列式. 最后, 我们可以将 x 和 y 表达为两个行列式的商. 克拉默法则可以扩张到有 n 个未知数, n 个方程所组成的线性方程组. 目前暂时还不需要实习这块的完整知识.

例 6.8.5 将下面的方程组改写成矩阵形式, 并且用克拉默法则求解.

$$\begin{cases} 2x+3y &= 8, \\ 4x-y &= 2. \end{cases}$$

证明. 此方程组的矩阵形式是

$$\begin{bmatrix} 2 & 3 \\ 4 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \end{bmatrix}.$$

其中, 方程组的系数矩阵和常数列向量为

$$A = \begin{bmatrix} 2 & 3 \\ 4 & -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 8 \\ 2 \end{bmatrix}.$$

系数矩阵的行列式 $\Delta = |A|$ 为

$$\Delta = \begin{vmatrix} 2 & 3 \\ 4 & -1 \end{vmatrix} = -2-12 = -14.$$

因为 $\Delta \neq 0$ 故有唯一解存在. 接下来, 我们计算 Δ_1 和 Δ_2. 我们将系数矩阵的第一列替换为常数列 $\mathbf{b}$:

$$\Delta_1 = \begin{vmatrix} 8 & 3 \\ 2 & -1 \end{vmatrix} = -8-6 = -14.$$

我们将系数矩阵的第二列替换为常数列 $\mathbf{b}$:

$$\Delta_2 = \begin{vmatrix} 2 & 8 \\ 4 & 2 \end{vmatrix} = 4-32 = -28.$$

根据克拉默法则, 求解 x 和 y 为

$$x = \frac{\Delta_1}{\Delta} = \frac{-14}{-14} = 1, \quad y = \frac{\Delta_2}{\Delta} = \frac{-28}{-14} = 2.$$

我们得到 $x=1$ 和 $y=2$. 将解代入原方程组中, 发现确实成立. 证毕. □

6.8.4 两个方程的隐函数存在定理 (m=2)

现在我们可以开始讨论由多个函数方程所确定的隐函数了. 比如, 在什么条件下, 由两个函数方程 (1-2 型隐函数)

$$\begin{cases} F(x;u,v) = 0, \\ G(x;u,v) = 0, \end{cases}$$

可确定隐函数 $u=u(x)$ 及 $v=v(x)$? 在前面的讨论中, 要想由 $F(x,y)=0$ 将 y 解为 x 的函数, 我们需要条件 $F_y(x,y) \neq 0$. 现在, 在多个函数方程中, 相当于 $F_y(x,y) \neq 0$ 的条件应当是什么? 为此, 不妨考查一下最简单的情况, 因为简单情况常常给人以启迪. 设

$$F(x,u,v) = ex+au+bv, \quad G(x,u,v) = fx+cu+dv,$$

其中 a,b,c,d,e,f 为常数. 这时, 我们的问题相当于解关于 u,v 的二元线性方程组

$$\begin{cases} au+bv = -ex, \\ cu+dv = -fx. \end{cases}$$

上式的右侧由 x 的值来确定, 给定 x 后就可以解出一对 (u,v) 的值. 由线性代数的理论知, 这个方程组有唯一解的充要条件是其系数行列式

$$\Delta = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \neq 0.$$

注意到偏导数 $F_u = a, F_v = b, G_u = c, G_v = d$, 由此我们猜想当 $F(x,u,v)$ 及 $G(x,u,v)$ 是一般函数时, 也需要如下条件:

$$\begin{vmatrix} F_u & F_v \\ G_u & G_v \end{vmatrix} \neq 0.$$

为了准确描述以上内容, 我们引入如下的概念.

1. **函数** $F(x,u,v), G(x,u,v)$[38] **(关于全部自变量** x,u,v **) 的雅可比矩阵 (Jacobian matrix)** 是函数的所有一阶偏导数以一定方式排列成的矩阵, 也就是 2 行 3 列的矩阵:

$$\frac{\partial(F,G)}{\partial(x,u,v)} = \begin{bmatrix} \dfrac{\partial F}{\partial x} & \dfrac{\partial F}{\partial u} & \dfrac{\partial F}{\partial v} \\ \dfrac{\partial G}{\partial x} & \dfrac{\partial G}{\partial u} & \dfrac{\partial G}{\partial v} \end{bmatrix} = \begin{bmatrix} F_x & F_u & F_v \\ G_x & G_u & G_v \end{bmatrix} = \begin{bmatrix} \text{第一个函数 } F \text{ 的梯度} \\ \text{第二个函数 } G \text{ 的梯度} \end{bmatrix}.$$

2. 如果我们只对部分变量的感兴趣, 比如变量 u,v, 则还可以定义**函数** $F(x,u,v), G(x,u,v)$ **关于部分变量** u,v **的雅可比矩阵**, 它就是上述完整矩阵的一个部分 (把第 2,3 列提取了出来), 即 2 x 2 矩阵:

$$\frac{\partial(F,G)}{\partial(u,v)} = \begin{bmatrix} \dfrac{\partial F}{\partial u} & \dfrac{\partial F}{\partial v} \\ \dfrac{\partial G}{\partial u} & \dfrac{\partial G}{\partial v} \end{bmatrix} = \begin{bmatrix} F_u & F_v \\ G_u & G_v \end{bmatrix}.$$

   另外, 这个 2 x 2 矩阵的行列式 (的值) 则称作**函数** $F(x,u,v), G(x,u,v)$ **关于** u,v **的雅可比行列式 (Jacobian determinant)**, 记作

$$J = \frac{\mathrm{D}(F,G)}{\mathrm{D}(u,v)} := \left|\frac{\partial(F,G)}{\partial(u,v)}\right| = \begin{vmatrix} F_u & F_v \\ G_u & G_v \end{vmatrix}.$$

3. 类似地, 则还可以定义**函数** $F(x,u,v), G(x,u,v)$ **关于变量** x **的雅可比矩阵**, 它就是上述完整雅可比矩阵的第 1 列, 即 2 x 1 矩阵:

$$\frac{\partial(F,G)}{\partial x} = \begin{bmatrix} \dfrac{\partial F}{\partial x} \\ \dfrac{\partial G}{\partial x} \end{bmatrix}.$$

   因为完整的雅可比矩阵每一列对应的都是同一个变量的偏导, 所以就是提取对应的列即可. 注意, 此时的矩阵不是一个正方形矩阵 (简称 "方阵"). **行列式的概念只对于类似 1x1, 2x2, 3x3 等这种方阵才存在, 其他任何形式的非方阵的矩阵绝无讨论行列式的可能.**

[38] 如果只考虑单个函数 $F(x,u,v)$, 它的雅可比矩阵就是梯度. 梯度是多元实数值函数的概念, 如果考虑多元到多元的一个映射, 那么该映射的每一个分量函数都各自拥有梯度. 所谓多元到多元映射的雅可比矩阵, 就是把所有分量函数的梯度排列到了一起的矩阵而已. 这里讨论的函数 $F(x,u,v), G(x,u,v)$ 就是三元到二元的一个映射.

注意, **以上所有偏导数都是在同一个点处的偏导数的值, 所以雅可比矩阵或行列式都是依赖于定义域上具体的点的概念**. 见下题.

例 6.8.6 考虑以下两个函数:

$$F(x,u,v)=x^2+uv,\quad G(x,u,v)=e^{xu}+v^2.$$

函数 $F(x,u,v)$ 和 $G(x,u,v)$ 关于全部自变量 x,u,v 的雅可比矩阵为

$$\begin{bmatrix} F_x & F_u & F_v \\ G_x & G_u & G_v \end{bmatrix}=\begin{bmatrix} 2x & v & u \\ ue^{xu} & xe^{xu} & 2v \end{bmatrix}.$$

看的出来, 每一个点 (x,u,v) 都决定了一个该点处的雅可比矩阵. 函数 $F(x,u,v)$ 和 $G(x,u,v)$ 关于变量 u 和 v 的雅可比行列式为

$$J=\frac{\mathrm{D}(F,G)}{\mathrm{D}(u,v)}=\begin{vmatrix} v & u \\ xe^{xu} & 2v \end{vmatrix}=2v^2-xue^{xu}.$$

如果考虑点 $(x,u,v)=(0,1,0)$, 则

$$\left.\frac{\mathrm{D}(F,G)}{\mathrm{D}(u,v)}\right|_{(0,1,0)}=0.$$

如果考虑点 $(x,u,v)=(1,1,1)$, 则

$$\left.\frac{\mathrm{D}(F,G)}{\mathrm{D}(u,v)}\right|_{(1,1,1)}=2-e\neq 0.$$

这些具体点展示了雅可比行列式在不同位置的行为.

事实表明, 由解二元线性方程组而得来的猜测是正确的. 我们有下面的定理. 定理的证明从略.

定理 6.8.4: 隐函数存在定理 3 (1-2 型隐函数)

设函数 $F(x,u,v)$ 及 $G(x,u,v)$ 在点 $P_0(x_0,y_0,z_0)$ 的某个邻域内有定义, 并且满足

1. $F(x_0,u_0,v_0)=0, G(x_0,u_0,v_0)=0$,
2. $F(x,u,v)$ 及 $G(x,u,v)$ 有连续的一阶偏导数,
3. $F(x,u,v), G(x,u,v)$ 关于 u,v 的雅可比行列式

$$J=\frac{\mathrm{D}(F,G)}{\mathrm{D}(u,v)}=\begin{vmatrix} F_u & F_v \\ G_u & G_v \end{vmatrix}$$

   在点 (x_0,u_0,v_0) 处不等于零,

则

1. 在 x_0 的某个邻域 $D\subseteq\mathbf{R}$ 内存在唯一的隐函数 $u=u(x)$ 及 $v=v(x)$, 即

$$\begin{cases} F(x,u(x),v(x))\equiv 0, \\ G(x,u(x),v(x))\equiv 0, \end{cases} \quad \forall x\in D,$$

   使得 $u_0=u(x_0), v_0=u(x_0)$.
2. 而且, 这样子的隐函数 $u=u(x)$ 及 $v=v(x)$ 在 D 内有连续的导数, 其导数由以下线性方程组解出:

$$\begin{bmatrix} \dfrac{\partial F}{\partial u} & \dfrac{\partial F}{\partial v} \\ \dfrac{\partial G}{\partial u} & \dfrac{\partial G}{\partial v} \end{bmatrix}\begin{bmatrix} \dfrac{\mathrm{d}u}{\mathrm{d}x} \\ \dfrac{\mathrm{d}v}{\mathrm{d}x} \end{bmatrix}=-\begin{bmatrix} \dfrac{\partial F}{\partial x} \\ \dfrac{\partial G}{\partial x} \end{bmatrix}.$$

定理中最后的方程组是以 u,v 的导数 $\dfrac{\mathrm{d}u}{\mathrm{d}x},\dfrac{\mathrm{d}v}{\mathrm{d}x}$ 为未知量的一个线性方程组, 其系数矩阵恰好是函数 $F(x,u,v), G(x,u,v)$ 关于 u,v 的雅可比矩阵. 展开写, 就是

$$\begin{cases} \dfrac{\partial F}{\partial x}+\dfrac{\partial F}{\partial u}\cdot\dfrac{\mathrm{d}u}{\mathrm{d}x}+\dfrac{\partial F}{\partial v}\cdot\dfrac{\mathrm{d}v}{\mathrm{d}x}=0, \\ \dfrac{\partial G}{\partial x}+\dfrac{\partial G}{\partial u}\cdot\dfrac{\mathrm{d}u}{\mathrm{d}x}+\dfrac{\partial G}{\partial v}\cdot\dfrac{\mathrm{d}v}{\mathrm{d}x}=0. \end{cases}$$

这个方程组无须记忆, 但应该知道它是怎样得来的: **(隐式求导)** 它是对恒等式

$$\begin{cases} F(x,u(x),v(x))\equiv 0, \\ G(x,u(x),v(x))\equiv 0, \end{cases}$$

对方程两端求导数而得来的.

比定理 6.8.4 更为一般的是将其中的 x 换成多个自变量, 此时解出的 u 及 v 则都是多元函数, 而定理的条件与结论完全类似. 比如, 我们考虑方程组 (2-2 型隐函数)

$$\begin{cases} F(x,y;u,v)=0, \\ G(x,y;u,v)=0. \end{cases}$$

除了要求点 (x_0, y_0, u_0, v_0) 满足这个方程组之外, 同样还应要求函数 $F(x, y, u, v)$ 及 $G(x, y, u, v)$ 在该点的某个邻域内有连续的一阶偏导数, 且 $F(x, y, u, v), G(x, y, u, v)$ 关于 u, v 的雅可比行列式在该点处不等于零. 这时, 在点 (x_0, y_0) 附近可以将 u, v 解为 (x, y) 的函数. 这些与定理 6.8.4 的叙述都类似. 所不同的是, 这时 u, v 各有两个偏导数, 它们分别由两个线性方程组解出. 如下所示.

1. 线性方程组
$$\begin{bmatrix} \dfrac{\partial F}{\partial u} & \dfrac{\partial F}{\partial v} \\ \dfrac{\partial G}{\partial u} & \dfrac{\partial G}{\partial v} \end{bmatrix} \begin{bmatrix} \dfrac{\mathrm{d}u}{\mathrm{d}x} \\ \dfrac{\mathrm{d}v}{\mathrm{d}x} \end{bmatrix} = - \begin{bmatrix} \dfrac{\partial F}{\partial x} \\ \dfrac{\partial G}{\partial x} \end{bmatrix},$$
   或者
$$\begin{cases} \dfrac{\partial F}{\partial x} + \dfrac{\partial F}{\partial u} \cdot \dfrac{\partial u}{\partial x} + \dfrac{\partial F}{\partial v} \cdot \dfrac{\partial v}{\partial x} = 0, \\ \dfrac{\partial G}{\partial x} + \dfrac{\partial G}{\partial u} \cdot \dfrac{\partial u}{\partial x} + \dfrac{\partial G}{\partial v} \cdot \dfrac{\partial v}{\partial x} = 0, \end{cases}$$
   同时解出 $\dfrac{\mathrm{d}u}{\mathrm{d}x}, \dfrac{\mathrm{d}v}{\mathrm{d}x}$.
2. 及线性方程组
$$\begin{bmatrix} \dfrac{\partial F}{\partial u} & \dfrac{\partial F}{\partial v} \\ \dfrac{\partial G}{\partial u} & \dfrac{\partial G}{\partial v} \end{bmatrix} \begin{bmatrix} \dfrac{\mathrm{d}u}{\mathrm{d}y} \\ \dfrac{\mathrm{d}v}{\mathrm{d}y} \end{bmatrix} = - \begin{bmatrix} \dfrac{\partial F}{\partial y} \\ \dfrac{\partial G}{\partial y} \end{bmatrix},$$
   或者
$$\begin{cases} \dfrac{\partial F}{\partial y} + \dfrac{\partial F}{\partial u} \cdot \dfrac{\partial u}{\partial y} + \dfrac{\partial F}{\partial v} \cdot \dfrac{\partial v}{\partial y} = 0, \\ \dfrac{\partial G}{\partial y} + \dfrac{\partial G}{\partial u} \cdot \dfrac{\partial u}{\partial y} + \dfrac{\partial G}{\partial v} \cdot \dfrac{\partial v}{\partial y} = 0, \end{cases}$$
   同时解出 $\dfrac{\mathrm{d}u}{\mathrm{d}y}, \dfrac{\mathrm{d}v}{\mathrm{d}y}$.

重要的是, 这两个线性方程组的系数矩阵完全一致, 只是右边的常数项不同而已.[39] 同样, 上面两个方程组也无须专门记忆, 只要知道 **(隐式求导)** 它们是对恒等式
$$\begin{cases} F(x, y, u(x, y), v(x, y)) \equiv 0, \\ G(x, y, u(x, y), v(x, y)) \equiv 0, \end{cases}$$
分别关于 x 与 y 求偏导数得来即可.

例 6.8.7 由方程组
$$\begin{cases} x^2 + y^2 - uv = 0, \\ xy + u^2 - v^2 = 0 \end{cases}$$
能否确定 u 和 v 为 (x, y) 的函数? 在能确定隐函数的条件下, 求 u_x, v_x, u_y 及 v_y.

[39] 只有学习了线性代数的矩阵乘法和逆矩阵等概念后, 才能有简洁的表达方法.

证明. 令 $F(x,y,u,v)=x^2+y^2-uv$, $G(x,y,u,v)=xy+u^2-v^2$, 则有

$$\frac{\mathrm{D}(F,G)}{\mathrm{D}(u,v)}=\begin{vmatrix}-v & -u\\ 2u & -2v\end{vmatrix}=2\left(u^2+v^2\right).$$

当 $(x,y)\neq(0,0)$ 时, 满足所给方程组的 u,v 不同时为零[40], 也就有 $\dfrac{\mathrm{D}(F,G)}{\mathrm{D}(u,v)}\neq 0$, 从而在点 (x,y) 的某个邻域内能确定隐函数 $u=u(x,y)$ 及 $v=v(x,y)$.

1. 为了求偏导数, 先将所给方程组中的方程两端对 x 求偏导数 (将其中的 u,v 看作 (x,y) 的函数), 得到

$$\begin{cases}2x-u_xv-uv_x=0,\\ y+2uu_x-2vv_x=0.\end{cases}$$

将上述线性方程组表示为矩阵乘法的形式, 即

$$\begin{pmatrix}-v & -u\\ 2u & -2v\end{pmatrix}\begin{pmatrix}u_x\\ v_x\end{pmatrix}=\begin{pmatrix}-2x\\ -y\end{pmatrix}.$$

由克拉默法则可得此方程组的解为

$$u_x=\frac{4xv-yu}{2\left(u^2+v^2\right)},\quad v_x=\frac{4xu+vy}{2\left(u^2+v^2\right)}.$$

2. 类似地, 对所给方程组中的方程两端关于 y 求偏导数, 得到

$$\begin{cases}2y-u_yv-uv_y=0,\\ x+2uu_y-2vv_y=0.\end{cases}$$

将上述线性方程组表示为矩阵乘法的形式, 即

$$\begin{pmatrix}-v & -u\\ 2u & -2v\end{pmatrix}\begin{pmatrix}u_y\\ v_y\end{pmatrix}=\begin{pmatrix}-2y\\ -x\end{pmatrix}.$$

由这个方程组可解出

$$u_y=\frac{4yv-xu}{2\left(u^2+v^2\right)},\quad v_y=\frac{xv+4yu}{2\left(u^2+v^2\right)}.$$

□

例 6.8.8 设 $u(x,y)$ 和 $v(x,y)$ 是由方程组 $\begin{cases}x=\mathrm{e}^u+v,\\ xy=\mathrm{e}^u+u\end{cases}$ 所确定的隐函数, 求 u_x 和 u_{xx}.

证明. 对所给方程组中的方程两端关于 x 求偏导数, 其中 y 视作常数, u 和 v 都视作 (x,y) 的函数, 得到

$$\begin{cases}1=\mathrm{e}^uu_x+v_x,\\ y=\mathrm{e}^uu_x+u_x,\end{cases}\tag{6.11}$$

解得[41]

$$u_x=\frac{y}{\mathrm{e}^u+1}.$$

[40]如果 u,v 同时为零, 则第一个方恒告诉我们 x,y 也同时为零.

[41]这里上下做差, v_x 就消掉了, 非常简单. 如果不能一看出来解法, 还是建议写成矩阵形式, 然后克拉默法则, 保证万无一失.

对方程组 (6.11) 中的方程两端再关于 x 求偏导数, 其中 y 视作常数, u, v, u_x 和 v_x 都视作 (x,y) 的函数, 得到

$$\begin{cases} 0=\mathrm{e}^u u_x u_x+\mathrm{e}^u u_{xx}+v_{xx}, \\ 0=\mathrm{e}^u u_x u_x+\mathrm{e}^u u_{xx}+u_{xx}, \end{cases}$$

解得

$$u_{xx}=\frac{-\mathrm{e}^u u_x \cdot u_x}{\mathrm{e}^u+1}.$$

将 $u_x=\dfrac{y}{\mathrm{e}^u+1}$ 代入上式, 即得

$$u_{xx}=-\frac{\mathrm{e}^u y^2}{\left(\mathrm{e}^u+1\right)^3}.$$

□

6.8.5　逆函数存在定理

方程组的隐函数存在定理的一个重要应用就是给出逆映射的存在性定理. 我们考虑 Ouv 平面的一个区域到 Oxy 平面的映射: $(u,v) \mapsto (x,y)$, 其中

$$\begin{cases} x=x(u,v), \\ y=y(u,v). \end{cases} \tag{6.12}$$

我们的问题是: 在怎样的条件下, 这个映射有逆映射存在? 所谓的逆映射, 是指在一定的范围之内, 对于给定的 (x,y), 根据方程组 (6.12) 能求得相应唯一的解 (u,v). 如果令

$$F(x,y,u,v):=x-x(u,v), \quad G(x,y,u,v):=y-y(u,v),$$

那么方程组 (6.12) 可以写成

$$\begin{cases} F(x,y,u,v)=0, \\ G(x,y,u,v)=0. \end{cases} \tag{6.13}$$

而求逆映射的问题也归结为由方程组 (6.13) 将其中的 u 和 v 解成 (x,y) 的函数的问题. 根据隐函数存在定理, 首先应当有一点 (x_0,y_0,u_0,v_0) 满足方程组 (6.13), 即

$$\begin{cases} x_0-x\left(u_0,v_0\right)=0, \\ y_0-y\left(u_0,v_0\right)=0. \end{cases}$$

其次, 应当要求 $F(x,y,u,v)$ 及 $G(x,y,u,v)$ 在点 (x_0,y_0,u_0,v_0) 的某个邻域内有连续的一阶偏导数, 且

$$\left.\frac{\mathrm{D}(F,G)}{\mathrm{D}(u,v)}\right|_{(x_0,y_0,u_0,v_0)} \neq 0.$$

显然, 只要 $x(u,v)$ 及 $y(u,v)$ 有连续的一阶偏导数就保证了 $F(x,y,u,v)$ 及 $G(x,y,u,v)$ 有连续的一阶偏导数. 根据 $F(x,y,u,v)$ 与 $G(x,y,u,v)$ 的定义, 很容易看出

$$\frac{\mathrm{D}(F,G)}{\mathrm{D}(u,v)}=\left|\begin{array}{ll} F_u & F_v \\ G_u & G_v \end{array}\right|=\left|\begin{array}{ll} x_u & x_v \\ y_u & y_v \end{array}\right|=\frac{\mathrm{D}(x,y)}{\mathrm{D}(u,v)},$$

其中 x 和 y 分别表示函数 $x(u,v)$ 和 $y(u,v)$. 这样一来, 只需

$$\left.\frac{\mathrm{D}(x,y)}{\mathrm{D}(u,v)}\right|_{(u_0,v_0)} \neq 0,$$

即可保证 $\left.\dfrac{\mathrm{D}(F,G)}{\mathrm{D}(u,v)}\right|_{(x_0 \cdot y_0, u_0 \cdot v_0)} \neq 0$. 在这些条件下, 在点 (x_0, y_0) 的一个邻域内存在一对函数 $u = u(x,y)$ 及 $v = v(x,y)$, 使得

$$u_0 = u\left(x_0, y_0\right), \quad v_0 = v\left(x_0, y_0\right),$$

并且它们所决定的映射 $(x,y) \mapsto (u,v)$ 是函数对 $x = x(u,v), y = y(u,v)$ 所决定的映射 $(u,v) \mapsto (x,y)$ 的逆映射. 这样, 我们证明了如下定理.

定理 6.8.5: 逆函数存在定理 ($\mathrm{R}^2 \to \mathrm{R}^2$ 型)

设函数 $x = x(u,v)$ 及 $y = y(u,v)$ 在点 (u_0, v_0) 的某个邻域内有定义, 并且满足

1. $x_0 = x\left(u_0, v_0\right), y_0 = y\left(u_0, v_0\right)$,
2. $x = x(u,v)$ 及 $y = y(u,v)$ 有连续的一阶偏导数,
3. $x = x(u,v), y = y(u,v)$ 关于 u, v 的雅可比行列式在点 (u_0, v_0) 处不等于零:

$$\left.\frac{\mathrm{D}(x,y)}{\mathrm{D}(u,v)}\right|_{(u_0,v_0)} \neq 0,$$

则在点 (x_0, y_0) 的一个邻域内存在一对函数 $u = u(x,y)$ 及 $v = v(x,y)$, 使得 $u_0 = u\left(x_0, y_0\right), v_0 = v\left(x_0, y_0\right)$, 并且映射 $(x,y) \mapsto (u(x,y), v(x,y))$ 是映射 $(u,v) \mapsto (x(u,v), y(u,v))$ 的逆映射.

我们知道, 逆映射的存在性意味着映射的一一性 (双射). 上述讨论使我们得出如下推论:

推论 6.8.1: 雅可比行列式非零 $\Rightarrow$ 局部可逆, 局部双射

若一个连续可微映射的雅可比行列式在一点处不等于零, 则这个映射在该点附近是一一映射 (双射).

在一元函数微分学中我们知道, 若一个函数有连续的导数且其导数在一点处不等于零, 则这个函数在该点附近是严格单调的 (从而是一一映射). 上述推论就是一元函数的这条性质的推广, 而雅可比行列式相当于一元函数的导数. 在一元函数中, 反函数在一点处的导数是原来函数在相应点处的导数的倒数. 对于映射的雅可比行列式, 也有相同的性质, 也就是说,

$$\left.\frac{\mathrm{D}(u,v)}{\mathrm{D}(x,y)}\right|_{(x_0,y_0)} = \frac{1}{\left.\frac{\mathrm{D}(x,y)}{\mathrm{D}(u,v)}\right|_{(u_0,v_0)}}.$$

即, 多元映射中雅可比行列式与其逆映射的雅可比行列式具有倒数关系.

事实上, 若 $u = u(x,y), v = v(x,y)$ 是 $x = x(u,v), y = y(u,v)$ 的逆映射, 则有

$$x \equiv x(u(x,y), v(x,y)),$$
$$y \equiv y(u(x,y), v(x,y)).$$

1. 对这两个恒等式两端关于 x 求偏导数, 即有

$$1 = x_u u_x + x_v v_x,$$
$$0 = y_u u_x + y_v v_x.$$

由此用克拉默法则可解出

$$u_x = \frac{y_v}{J}, \quad v_x = -\frac{y_u}{J},$$

其中 $J = \dfrac{\mathrm{D}(x,y)}{\mathrm{D}(u,v)}$.

2. 再对上述两个恒等式两端关于 y 求偏导数, 类似地, 又得

$$u_y = -\frac{x_v}{J}, \quad v_y = \frac{x_u}{J}.$$

这样, 我们有

$$\frac{\mathrm{D}(u,v)}{\mathrm{D}(x,y)} = \begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix} = \frac{1}{J^2}\begin{vmatrix} y_v & -x_v \\ -y_u & x_u \end{vmatrix} = \frac{1}{J}.$$

例 6.8.9 下面是一个具体例子, 说明雅可比行列式的倒数关系. 设映射

$$\begin{cases} x = ue^v, \\ y = v. \end{cases}$$

则其在任意点 (u,v) 处的雅可比行列式

$$J = \frac{\mathrm{D}(x,y)}{\mathrm{D}(u,v)} = \begin{vmatrix} e^v & ue^v \\ 0 & 1 \end{vmatrix} = e^v \neq 0.$$

因此, 在任意点 (u,v) 处, 雅可比行列式 J 不为零, 满足逆函数定理的条件. 即该函数在任意点处都是局部可逆的. 事实上, 它是全局可逆的, 其逆函数容易解出为

$$\begin{cases} u = xe^{-y}, \\ v = y. \end{cases}$$

此时, 逆映射的雅可比行列式

$$J' = \frac{\mathrm{D}(u,v)}{\mathrm{D}(x,y)} = \begin{vmatrix} e^{-y} & -xe^{-y} \\ 0 & 1 \end{vmatrix} = e^{-y}.$$

显然, 有倒数关系

$$J' = \frac{1}{J}.$$

例 6.8.10 证明: 极坐标变换

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

的雅可比行列式满足

$$\frac{\mathrm{D}(x,y)}{\mathrm{D}(r,\theta)} = r.$$

证明. 根据雅可比行列式的定义, 有

$$\frac{\mathrm{D}(x,y)}{\mathrm{D}(r,\theta)}=\begin{vmatrix} x_r & x_\theta \\ y_r & y_\theta \end{vmatrix}=\begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix}=r.$$

证毕. □

例 6.8.11 设 $u=u(x,y)$ 及 $v=v(x,y)$ 是由方程组

$$\begin{cases} u^2-v=3x+y, \\ u-2v^2=x-2y \end{cases}$$

所确定的隐函数, 求 $\dfrac{\partial u}{\partial x}$ 及 $\dfrac{\partial v}{\partial x}$.

证明. 将所给方程组中的两个方程左端的 u 及 v 视作 (x,y) 的函数, 然后对这两个方程两端关于 x 求偏导数, 即得

$$\begin{cases} 2uu_x-v_x=3, \\ u_x-4vv_x=1. \end{cases}$$

由此用克拉默法则解出

$$\frac{\partial u}{\partial x}=\frac{1-12v}{1-8uv},\quad \frac{\partial v}{\partial x}=\frac{2u-3}{1-8uv},$$

其中 $uv\neq\dfrac{1}{8}$. □

上述关于逆映射的存在性定理及雅可比行列式的概念完全可以推广到 n 个变量的情况. 设区域 $D\subset\mathbf{R}^n$, $f:D\to\mathbf{R}^n$ 是 D 到 $\mathbf{R}^n$ 的映射[42], 其分量为 $u_j=u_j\left(x_1,\cdots,x_n\right)$, 那么映射 f 的雅可比行列式定义为

$$J=\frac{\mathrm{D}\left(u_1,\cdots,u_n\right)}{\mathrm{D}\left(x_1,\cdots,x_n\right)}=\begin{vmatrix} \dfrac{\partial u_1}{\partial x_1} & \cdots & \dfrac{\partial u_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial u_n}{\partial x_n} & \cdots & \dfrac{\partial u_n}{\partial x_n} \end{vmatrix}.$$

有了映射 f 的雅可比行列式概念, 逆映射的存在性即可推广到 n 个变量的情况, 即, 如果映射 f 的在点 $P\in D$ 处的雅可比行列式不等于零, 则存在点 P 的某个邻域, 使得映射 f 是双射.

例 6.8.12 显然, 线性变换

$$y_1=a_{11}x_1+a_{12}x_2+a_{13}x_3$$

$$y_2=a_{21}x_1+a_{22}x_2+a_{23}x_3$$

$$y_3=a_{31}x_1+a_{32}x_2+a_{33}x_3$$

的雅可比行列式就是系数行列式, 其中 $a_{ij}\ (i,j=1,2,3)$ 为常数.

[42]尽管也存在 “输入和输出个数不同的函数也有逆函数” 的情况, 但这些函数实在是非主流, 故我们只考虑输入和输出个数一致的函数是否存在逆函数.

例 6.8.13 球坐标变换

$$\begin{cases} x = r\cos\theta\sin\varphi, \\ y = r\sin\theta\sin\varphi, \\ z = r\cos\varphi, \end{cases} \quad (0 \leq \theta < 2\pi, 0 \leq \varphi \leq \pi).$$

的雅可比行列式为 $J = \dfrac{\mathrm{D}(x,y,z)}{\mathrm{D}(r,\theta,\varphi)} = -r^2\sin\varphi$.[43]

证明. 我们有雅可比矩阵:

$$\frac{\partial(x,y,z)}{\partial(r,\varphi,\theta)} = \begin{pmatrix} \cos\theta\sin\varphi & -r\sin\theta\sin\varphi & r\cos\theta\cos\varphi \\ \sin\theta\sin\varphi & r\cos\theta\sin\varphi & r\sin\theta\cos\varphi \\ \cos\varphi & 0 & -r\sin\varphi. \end{pmatrix}.$$

我们需要计算这个 3×3 矩阵的行列式 J. 按照第一行展开, 行列式的展开公式为

$$J = a_{11}\cdot\det(M_{11}) - a_{12}\cdot\det(M_{12}) + a_{13}\cdot\det(M_{13}),$$

其中 a_{1j} 是第一行的元素, M_{1j} 是去掉第一行和第 j 列后的 2×2 子矩阵.

1. 步骤 1: 确定 a_{11}, a_{12}, a_{13}.

$$a_{11} = \cos\theta\sin\varphi, \quad a_{12} = -r\sin\theta\sin\varphi, \quad a_{13} = r\cos\theta\cos\varphi.$$

2. 步骤 2: 计算各子矩阵的行列式.

   (a) 子矩阵 M_{11} (去掉第一行和第一列):

$$M_{11} = \begin{pmatrix} r\cos\theta\sin\varphi & r\sin\theta\cos\varphi \\ 0 & -r\sin\varphi \end{pmatrix},$$

$$\det(M_{11}) = (r\cos\theta\sin\varphi)\cdot(-r\sin\varphi) - (r\sin\theta\cos\varphi)\cdot 0 = -r^2\cos\theta\sin^2\varphi.$$

   (b) 子矩阵 M_{12} (去掉第一行和第二列):

$$M_{12} = \begin{pmatrix} \sin\theta\sin\varphi & r\sin\theta\cos\varphi \\ \cos\varphi & -r\sin\varphi \end{pmatrix},$$

$$\det(M_{12}) = (\sin\theta\sin\varphi)\cdot(-r\sin\varphi) - (r\sin\theta\cos\varphi)\cdot\cos\varphi = -r\sin\theta.$$

   (c) 子矩阵 M_{13} (去掉第一行和第三列):

$$M_{13} = \begin{pmatrix} \sin\theta\sin\varphi & r\cos\theta\sin\varphi \\ \cos\varphi & 0 \end{pmatrix},$$

$$\det(M_{13}) = (\sin\theta\sin\varphi)\cdot 0 - (r\cos\theta\sin\varphi)\cdot\cos\varphi = -r\cos\theta\sin\varphi\cos\varphi.$$

[43] 这里变量求导的顺序是 r,φ,θ 时, 最后结果会变号, 即 $J = \dfrac{\mathrm{D}(x,y,z)}{\mathrm{D}(r,\varphi,\theta)} = r^2\sin\varphi$.

3. 步骤 3: 代入行列式展开公式

$$
\begin{aligned}
J &= a_{11}\cdot\det(M_{11}) - a_{12}\cdot\det(M_{12}) + a_{13}\cdot\det(M_{13}) \\
&= (\cos\theta\sin\varphi)\cdot(-r^2\cos\theta\sin^2\varphi) - (-r\sin\theta\sin\varphi)\cdot(-r\sin\theta) + (r\cos\theta\cos\varphi)\cdot(-r\cos\theta\sin\varphi\cos\varphi) \\
&= -r^2\cos^2\theta\sin^3\varphi - r^2\sin^2\theta\sin\varphi - r^2\cos^2\theta\sin\varphi\cos^2\varphi \\
&= -r^2\sin\varphi\left(\cos^2\theta\sin^2\varphi + \sin^2\theta + \cos^2\theta\cos^2\varphi\right) \\
&= -r^2\sin\varphi\cdot 1 \\
&= -r^2\sin\varphi.
\end{aligned}
$$

□

雅可比行列式有明显的几何意义, 并在重积分的计算中扮演重要角色.

6.9 极值问题

6.9.1 二元函数极值问题

与一元函数的情况类似, 对多元函数也可定义极值, 极值点及稳定点等概念. 下面以二元函数为例介绍这些概念.

定义 6.9.1: 二元函数的极值点

设 $f: D\subseteq \mathbf{R}^2 \to \mathbf{R}$ 是在区域 $D\subseteq\mathbf{R}^2$ 内定义的函数, 且 $(x_0,y_0)\in D$ 是内点. 如果存在 (x_0,y_0) 的一个邻域 $U\subseteq D$, 使得对该邻域内的任意点 $(x,y)\in U$,

1. 都有
$$f(x,y)\le f(x_0,y_0),$$
则称 $f(x_0,y_0)$ 为 $f(x,y)$ 在 (x_0,y_0) 处的极大值, 并称 (x_0,y_0) 为极大值点.

2. 都有
$$f(x,y)\ge f(x_0,y_0),$$
则称 $f(x_0,y_0)$ 为 $f(x,y)$ 在 (x_0,y_0) 处的极小值, 并称 (x_0,y_0) 为极小值点.

极大值与极小值统称为极值, 极大值点与极小值点统称为极值点.

例 6.9.1 函数 $f(x,y)=(x-1)^2+(y-2)^2$ 在点 $(1,2)$ 处的值为极小值. 首先, 计算在点 $(1,2)$ 处的函数值

$$f(1,2)=(1-1)^2+(2-2)^2=0.$$

接着, 考虑点 $(1,2)$ 邻近的其他点 (x,y). 其实对于任意点 $(x,y)\in\mathbf{R}^2$ 都有

$$f(x,y)=(x-1)^2+(y-2)^2\ge 0=f(1,2).$$

因此, $f(1,2)=0$ 是一个极小值, 而 $(1,2)$ 是一个极小值点.

根据定义, **不是极值点**的意思是, 给定一个点 $(x_0, y_0) \in D$, 该点既不是极大值点, 也不是极小值点. 换句话说, 在该点的任何邻域内, 存在使得函数值大于 $f(x_0, y_0)$ 的点, 也存在使得函数值小于 $f(x_0, y_0)$ 的点.

例 6.9.2 设函数 $f(x, y) = xy$. 首先, 计算 $f(0, 0) = 0$. 可以证明, 点 $(0, 0)$ 并不是极值点. 比如, 在点 $(0, 0)$ 的任意邻域内都包含着两点 (ϵ, ϵ) 和 $(\epsilon, -\epsilon)$, 这里 ϵ 是充分小的正数, 则 $f(\epsilon, \epsilon) = \epsilon^2 > 0$, $f(\epsilon, -\epsilon) = -\epsilon^2 < 0$.

像一元函数的情况一样, 二元函数的极值概念是局部性的. 这意味着极值指的是在某一点附近的最大值或最小值, 而不是在整个定义域中的最大值或最小值. 因此, 一个点可以是局部极值点, 但不一定是全局极值点, 也就是整个定义域中的最大 (小) 值点.

例 6.9.3 考虑函数 $f(x, y) = y^2 + x^2 - x^4$, 定义在 $\mathbf{R}^2$ 上. 我们将证明该函数在点 $(0, 0)$ 处取得局部极小值, 但该点不是全局极小值点. 见图 6.15.

1. 首先, 计算函数在点 $(0, 0)$ 处的值:

   $$f(0, 0) = 0^2 + 0^2 - 0^4 = 0.$$

   接下来, 考虑点 $(0, 0)$ 的一个矩形邻域[44] U_ϵ, 其中 $0 < \epsilon < 1$, 定义为

   $$U_\epsilon = \{(x, y) \in \mathbf{R}^2 \mid |x| < \epsilon, |y| < \epsilon\}.$$

   值得注意的是, 当 $0 < |x| < 1$ 时, 有 $x^4 < x^2$. 因此, 对于任意 $(x, y) \in U_\epsilon$ 且 $(x, y) \neq (0, 0)$, 我们有

   $$f(x, y) = y^2 + x^2 - x^4 > y^2 + x^2 - x^2 = y^2 > 0.$$

   这说明, 在点 $(0, 0)$ 的邻域 U_ϵ 内, 所有点的函数值都满足

   $$f(x, y) \geq 0 = f(0, 0).$$

   因此, 点 $(0, 0)$ 是一个局部极小值点, 并且 $f(0, 0) = 0$ 是该点的局部极小值.

2. 接下来, 考虑当 $y = 0$ 时的情形. 此时,

   $$f(x, 0) = x^2 - x^4 = x^2(1 - x^2).$$

   可以看到, 当 $|x| > 1$ 时, 函数值 $f(x, 0)$ 会变得负值, 并且当 $x \to \infty$ 时, $f(x, 0) \to -\infty$. 这说明函数不存全局最小值.

综上所述, 虽然 $(0, 0)$ 是一个局部极小值点, 且 $f(0, 0) = 0$ 为局部极小值, 但由于在其他区域存在比 $f(0, 0)$ 更小的函数值, $(0, 0)$ 不是全局极小值点.

一阶必要条件

与一元函数类似, 可以利用二元函数的偏导数来给出极值点的必要条件与充分条件.

[44] 其实, 这正是通过切比雪夫距离函数 $d((x_1, y_1), (x_2, y_2)) = \max\{|x_2 - x_1|, |y_2 - y_1|\}$ 推导出来的邻域.

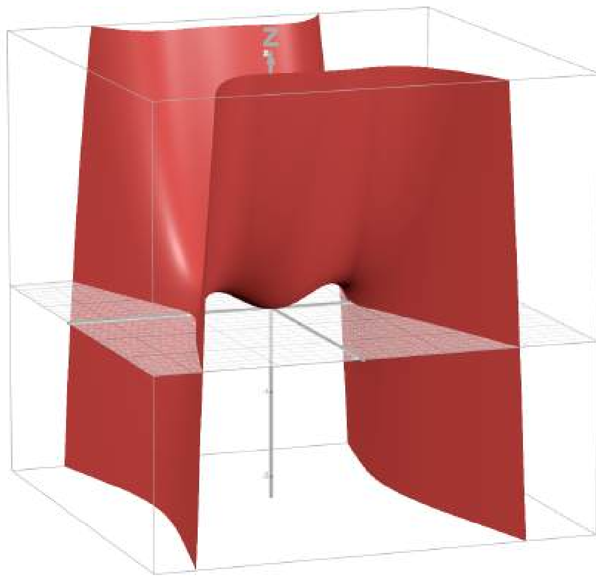


图 6.15: 函数 $f(x,y)=y^2+x^2-x^4$ 在点 $(0,0)$ 处取得局部极小值, 但该点不是全局极小值点.

定理 6.9.1: 二元函数极值点的一阶必要条件

若函数 $f(x,y)$ 在点 (x_0,y_0) 处达到极值, 并且 $f_x(x_0,y_0)$ 与 $f_y(x_0,y_0)$ 存在, 则必有

$$f_x(x_0,y_0)=0,\quad f_y(x_0,y_0)=0.$$

或者, 使用梯度符号表示

$$\operatorname{grad} f(x_0,y_0)=(0,0).$$

证明. 不妨设 $f(x_0,y_0)$ 为极大值. 由极大值的定义, 对于点 (x_0,y_0) 的某一邻域内的一切点 (x,y), 都有

$$f(x,y)\le f(x_0,y_0).$$

特别地, 若只固定 $y=y_0$, 则有

$$f(x,y_0)\le f(x_0,y_0).$$

因此, x_0 是一元函数 $x\mapsto f(x,y_0)$ 的极大值点. 由一元函数极值点的必要条件有

$$\frac{\mathrm{d}}{\mathrm{d}x}f(x,y_0)\Big|_{x_0}=0.$$

由偏导数的定义, 上式意味着 $f_x(x_0,y_0)=0$. 同理可证 $f_y(x_0,y_0)=0$. 证毕. □

例 6.9.4 我们可以使用例 6.9.3 中的函数 $f(x,y)=y^2+x^2-x^4$ 来演示一阶必要条件. 首先, 我们计算该函数在一般点 (x,y) 处的偏导数:

$$f_x(x,y)=2x-4x^3,\quad f_y(x,y)=2y.$$

显然, $f_x(0,0)=f_y(0,0)=0$.

上述结论说明, 极值点处如果可微, 那么其任意方向的方向导数都是零 (方向导数等于梯度和方向向量的内积). 从几何意义上看, 函数在该点的切平面是水平于 Oxy 平面的. 因此, 极值点在几何上表现为曲面上的一个" 平顶" 或" 平谷". 将上述结论推广到一般情况, 可以得到以下定理.

定理 6.9.2: n 元函数的极值的一阶必要条件

如果 $f: D \subseteq \mathbf{R}^n \to \mathbf{R}$ 在点 $\mathbf{x}_0 \in D$ 处取得极值, 且在点 $\mathbf{x}_0$ 处的偏导数都存在, 则在该点的所有偏导数均为零, 即

$$\frac{\partial f}{\partial x_1}(\mathbf{x}_0) = \frac{\partial f}{\partial x_2}(\mathbf{x}_0) = \cdots = \frac{\partial f}{\partial x_n}(\mathbf{x}_0) = 0.$$

或者, 使用梯度符号表示

$$\operatorname{grad} f(\mathbf{x}_0) = \mathbf{0}.$$

总之, **梯度为零是可微[45]函数的极值点的必要条件.** 若函数 $f(x,y)$ 在区域 D 内有偏导数, 则根据定理 6.9.2, $f(x,y)$ 在 D 内的极值点必满足方程组

$$\begin{cases} f_x(x,y) = 0, \\ f_y(x,y) = 0. \end{cases}$$

我们称满足这个方程组的点为 $f(x,y)$ 的**稳定点**. 因此, 当 $f(x,y)$ 在 D 内有偏导数时, 其极值点必定是其稳定点. 但函数的稳定点未必是函数的极值点. 比如, 在前面的例子 6.9.2 中, 对于 $f(x,y) = xy$ 可列出方程

$$\begin{cases} f_x(x,y) = y = 0, \\ f_y(x,y) = x = 0. \end{cases}$$

故, $(0,0)$ 是其唯一的稳定点, 但例子 6.9.2 告诉我们该点不是极值点. 另外, 如果函数在某点如果不可偏导, 它仍可能是一个极值点. 例如, 考虑函数 $f(x,y) = \max(|x|,|y|)$, 定义在 $\mathbf{R}^2$ 上. 显然, $(0,0)$ 是一个局部极小值点. 但是, 函数 $f(x,y)$ 在点 $(0,0)$ 处是不可偏导的. 切记, 极值的一阶必要条件 (零梯度) 只是针对可微或者可偏导函数而言的, 对于不可微或不可偏导函数而言, 则需要具体分析才行.

二阶充分条件

在一元函数的情形中, 判别一个稳定点是否为极值点要借助于函数的二阶导数. 在多元函数的情形中, 则要借助于二阶偏导数, 但比一元函数的情形要复杂些. 现在引入一个新的记号. 设函数 $f(x,y)$ 在点 $P_0(x_0,y_0)$ 的邻域 $U(P_0)$ 内具有连续的二阶偏导数. 函数在点 (x_0,y_0) 处的**黑塞矩阵 (Hessian matrix)** (也译作海森矩阵、海瑟矩阵、海塞矩阵等) 定义为

$$H_f|_{(x_0,y_0)} = H_f(x_0,y_0) = \begin{pmatrix} f_{xx}(x_0,y_0) & f_{xy}(x_0,y_0) \\ f_{xy}(x_0,y_0) & f_{yy}(x_0,y_0) \end{pmatrix}.$$

若引入记号 $f_{xx}(x_0,y_0) = A$, $f_{xy}(x_0,y_0) = B$, 和 $f_{yy}(x_0,y_0) = C$, 则黑塞矩阵可以简洁地写作

$$H_f(x_0,y_0) = \begin{pmatrix} A & B \\ B & C \end{pmatrix}.$$

显然, 黑塞矩阵是一个 2×2 的对称矩阵. 这个矩阵形式的表示方法不仅简洁, 而且能够更好地处理多变量函数的极值判别问题.

[45]可微自然可偏导.

例 6.9.5 例如, 我们考虑以下二元函数

$$f(x,y)=x^3+y^3-3xy.$$

接下来, 我们计算函数的二阶偏导数, 以构成黑塞矩阵的元素

$$f_{xx}=\frac{\partial^2 f}{\partial x^2}=6x,\quad f_{xy}=\frac{\partial^2 f}{\partial x\partial y}=-3,\quad f_{yy}=\frac{\partial^2 f}{\partial y^2}=6y.$$

则函数 $f(x,y)$ 在任意点 (x_0,y_0) 处的黑塞矩阵为

$$H_f\left(x_0,y_0\right)=\begin{pmatrix} f_{xx}\left(x_0,y_0\right) & f_{xy}\left(x_0,y_0\right) \\ f_{xy}\left(x_0,y_0\right) & f_{yy}\left(x_0,y_0\right)\end{pmatrix}=\begin{pmatrix} 6x_0 & -3 \\ -3 & 6y_0\end{pmatrix}.$$

在点 $(x_0,y_0)=(0,0)$ 处, 计算黑塞矩阵

$$H_f(0,0)=\begin{pmatrix} 0 & -3 \\ -3 & 0\end{pmatrix}.$$

随着关心的点的变动, 就这形成了一个点到矩阵的映射: $(x,y)\mapsto H_f(x,y)$.

顺便提一下. 黑塞矩阵可以自然地扩张到更多元的情况. 设 $f(x,y,z)$ 是一个三元函数, 即它是一个以 x, y, z 为自变量的函数. 黑塞矩阵 $H_f(x_0,y_0,z_0)$ 是一个 3×3 的对称矩阵, 其中的元素是 $f(x,y,z)$ 在点 (x_0,y_0,z_0) 处的二阶偏导数. 具体来说,

$$H_f(x_0,y_0,z_0)=\begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2}(x_0,y_0,z_0) & \dfrac{\partial^2 f}{\partial x\partial y}(x_0,y_0,z_0) & \dfrac{\partial^2 f}{\partial x\partial z}(x_0,y_0,z_0) \\ \dfrac{\partial^2 f}{\partial y\partial x}(x_0,y_0,z_0) & \dfrac{\partial^2 f}{\partial y^2}(x_0,y_0,z_0) & \dfrac{\partial^2 f}{\partial y\partial z}(x_0,y_0,z_0) \\ \dfrac{\partial^2 f}{\partial z\partial x}(x_0,y_0,z_0) & \dfrac{\partial^2 f}{\partial z\partial y}(x_0,y_0,z_0) & \dfrac{\partial^2 f}{\partial z^2}(x_0,y_0,z_0)\end{pmatrix}.$$

其中, 对角线元素是函数 $f(x,y,z)$ 关于 x, y, z 的二阶偏导数. 例如, 矩阵的 $(1,1)$ 元素是 $\dfrac{\partial^2 f}{\partial x^2}$, 表示函数对 x 的二阶导数. 非对角线元素则是混合偏导数. 例如, 矩阵的 $(1,2)$ 元素是 $\dfrac{\partial^2 f}{\partial x\partial y}$, 表示函数对 x 和 y 的混合二阶导数. 由于偏导数的对称性, $\dfrac{\partial^2 f}{\partial x\partial y}=\dfrac{\partial^2 f}{\partial y\partial x}$, 因此黑塞矩阵是对称的. 例如, 函数 $f(x,y,z)=x^2y+y^2z+z^2x$ 在点 $(1,2,3)$ 处的黑塞矩阵为

$$H_f(1,2,3)=\begin{pmatrix} 4 & 2 & 6 \\ 2 & 6 & 4 \\ 6 & 4 & 2\end{pmatrix}.$$

随着关心的点的变动, 就这形成了一个点到矩阵的映射: $(x,y,z)\mapsto H_f(x,y,z)$.

定理 6.9.3: 二元函数极值点的二阶判断条件 (二阶充分条件)

设函数 $f(x,y)$ 在点 $P_0\left(x_0,y_0\right)$ 的一个邻域 $U\left(P_0\right)$ 内具有连续的二阶偏导数, 并且该点是一个稳定点 (即 $f_x\left(x_0,y_0\right)=0$ 且 $f_y\left(x_0,y_0\right)=0$). 定义在点 $\left(x_0,y_0\right)$ 处的黑塞矩阵为

$$H \equiv H_f(x_0,y_0)=\begin{pmatrix} A & B \\ B & C \end{pmatrix},$$

其中, $f_{xx}\left(x_0,y_0\right)=A, f_{xy}\left(x_0,y_0\right)=B$, 和 $f_{yy}\left(x_0,y_0\right)=C$. 通过使用黑塞矩阵 H 及其行列式 $\det(H)=AC-B^2$, 我们可以将二阶判断条件表达如下:

(i) 如果 $\det(H)>0$ 且 $A>0$, 则 $f(x,y)$ 在 $P_0\left(x_0,y_0\right)$ 处取得局部极小值.

(ii) 如果 $\det(H)>0$ 且 $A<0$, 则 $f(x,y)$ 在 $P_0\left(x_0,y_0\right)$ 处取得局部极大值.

(iii) 如果 $\det(H)<0$, 则 $P_0\left(x_0,y_0\right)$ 是一个鞍点, 即该点既不是极大值也不是极小值.

(iv) 如果 $\det(H)=0$, 则无法论断, 需要进一步分析.

注意 如果 $\det(H)>0$ 且 $A=0$, 那会发生什么呢? 同样地, 我无法论断, 需要进一步分析吗? 答案是, 其实这根本不会出现. 当 $\det(H)>0$, 则 $AC-B^2>0$; 此时若 $A=0$, 则 $-B^2>0$, 这是不可能的. 总之, 条件 $\det(H)>0$ 则必然导致 $A\neq 0$. 三元及以上的函数极值点的二阶判断条件我们不讨论, 这需要学习非常多的线性代数知识作为预备条件. 但有一点, 它绝不是上述二元函数版本的简单扩张.

上述定理说明了对于二元函数, 如果一个稳定点的黑塞矩阵的行列式 (1) 大于零, 则该点一定是极值点 (极大还是极小根据左上角 A 的符号继续判断); (2) 小于零, 则该点一定不是极值点; (3) 等于零, 则什么也无法论断.

证明上述定理前, 我们需要学一点线性代数的基本知识. **二次型 (quadratic form)** 是指一种特殊类型的多项式, 它包含变量的二次幂和二次项的系数. 具体来说, 二次型是形如:

$$Q\left(x_1,x_2,\ldots,x_n\right)=\sum_{i=1}^{n} a_i x_i^2+\sum_{i\neq j} b_{ij}x_ix_j$$

的表达式, 其中 $x_1,x_2,\ldots,x_n$ 是变量, a_i 和 b_{ij} 是常数系数. 例如, 对于两个变量 x 和 y, 一个二次型可以表示为[46]

$$Q(x,y)=ax^2+2bxy+cy^2.$$

根据线性代数中二次型的理论的知识: 当 $ac>b^2$ 时, 二次型 $Q(x,y)=ax^2+2bxy+cy^2$ 对一切不全为零的 x,y 保持符号, 且其符号与 a 的符号相同. 这一结论可以由配方的方法直接证明. 事实上, 当 $ac>b^2$ 时, $ac\neq 0$, 这时

$$Q(x,y)=ax^2+2bxy+cy^2=\frac{1}{a}\left[(ax+by)^2+\left(ac-b^2\right)y^2\right].$$

当 x,y 不全为零时, 上式右端方括号内的值总大于零.

证明. 1. **首先一起考虑前面两种情况 (i) 和 (ii), 假设** $\det(H)>0$ **且** $A\neq 0$**.** 设 (x,y) 为点 P_0 的邻域 $U(P_0)$ 内的任意一点, 令 $x=x_0+\Delta x$, $y=y_0+\Delta y$. 根据多元函数的泰勒公式, 并结合点 P_0 的梯度为

[46] 对于三个变量的二次型, 可以表示为 $Q(x,y,z)=a_1x^2+a_2y^2+a_3z^2+2b_1xy+2b_2xz+2b_3yz$.

零的事实, 我们可以将 $f(x_0+\Delta x, y_0+\Delta y)$ 展开为二阶的泰勒展开式 (带拉格朗日余项):

$$f(x_0+\Delta x, y_0+\Delta y)=f(x_0,y_0)+\frac{1}{2}\left[\frac{\partial^2 f}{\partial x^2}(P_\theta)(\Delta x)^2+2\frac{\partial^2 f}{\partial x\partial y}(P_\theta)\Delta x\Delta y+\frac{\partial^2 f}{\partial y^2}(P_\theta)(\Delta y)^2\right],$$

其中 $P_\theta=(x_0+\theta\Delta x, y_0+\theta\Delta y)$, 且 $0<\theta<1$. 为了简便起见, 定义

$$\widetilde{A}=\frac{\partial^2 f}{\partial x^2}(P_\theta),\quad \widetilde{B}=\frac{\partial^2 f}{\partial x\partial y}(P_\theta),\quad \widetilde{C}=\frac{\partial^2 f}{\partial y^2}(P_\theta),$$

则泰勒公式可以改写为

$$f(x_0+\Delta x, y_0+\Delta y)=f(x_0,y_0)+\frac{1}{2}\left[\widetilde{A}(\Delta x)^2+2\widetilde{B}\Delta x\Delta y+\widetilde{C}(\Delta y)^2\right].$$

从中我们可以看出, 为了判断 $f(x_0,y_0)$ 是否为极值, 只需研究二次型

$$\widetilde{A}(\Delta x)^2+2\widetilde{B}\Delta x\Delta y+\widetilde{C}(\Delta y)^2=2\left(f(x_0+\Delta x, y_0+\Delta y)-f(x_0,y_0)\right)$$

对于所有足够小的 Δx 和 Δy (不全为零) 是否恒为正或恒为负.

根据二阶偏导数连续的假设, 当 $\Delta x\to 0$ 且 $\Delta y\to 0$ 时, 点 P_θ 会收敛到 P_0, 即

$$\widetilde{A}\to A,\quad \widetilde{B}\to B,\quad \widetilde{C}\to C,$$

因此, 当 Δx 和 Δy 足够小时, $\widetilde{A}$ 与 A 会具有相同的符号. 进一步地, 考虑到

$$(\widetilde{A}\widetilde{C}-\widetilde{B}^2)\to(AC-B^2)=\det(H)>0,$$

根据极限的保号性, 可以得出结论: 当 Δx 和 Δy 足够小时, 总有 $\widetilde{B}^2<\widetilde{A}\widetilde{C}$. 于是, 利用线性代数的知识, 就有下面的讨论结果.

(a) 当 $A>0$ 时, 对于一切充分小 (不全为零) 的 Δx 和 Δy, 有 $\widetilde{A}>0$, 因此二次型

$$\widetilde{A}(\Delta x)^2+2\widetilde{B}\Delta x\Delta y+\widetilde{C}(\Delta y)^2>0,$$

由此得到 $f(x_0+\Delta x, y_0+\Delta y)>f(x_0,y_0)$, 即 $f(x_0,y_0)$ 是极小值. 这就证明了 (i).

(b) 当 $A<0$ 时, 对于一切充分小 (不全为零) 的 Δx 和 Δy, 有 $\widetilde{A}<0$, 因此二次型

$$\widetilde{A}(\Delta x)^2+2\widetilde{B}\Delta x\Delta y+\widetilde{C}(\Delta y)^2<0,$$

由此得到 $f(x_0+\Delta x, y_0+\Delta y)<f(x_0,y_0)$, 即 $f(x_0,y_0)$ 是极大值. 这就证明了 (ii).

2. **现在考虑第三种情况 (iii), 假设** $\det(H)<0$**, 即** $B^2>AC$**.** 下面, 我们根据 A,C 中零的分布情况做分类讨论.

(a) **(Case 1. 非零的个数** ≥ 1**)** 若 A,C 中至少有一个不等于零. 不妨设 $A\neq 0$. 展开至二阶的带佩亚诺余项的泰勒公式为 (默认取极限 $\rho\to 0$)

$$f\left(x_0+\Delta x, y_0+\Delta y\right)=f\left(x_0,y_0\right)+\frac{1}{2}\left[A(\Delta x)^2+2B\Delta x\Delta y+C(\Delta y)^2\right]+o\left(\rho^2\right),$$

其中 $\rho=\sqrt{(\Delta x)^2+(\Delta y)^2}$. 由于 $A\neq 0$, 通过代数变换, 我们将上式改写为

$$f\left(x_0+\Delta x, y_0+\Delta y\right)=f\left(x_0,y_0\right)+\frac{1}{2}\cdot\frac{1}{A}\left[(A\Delta x+B\Delta y)^2-(\Delta y)^2\left(B^2-AC\right)\right]+o\left(\rho^2\right).\quad(6.14)$$

i. 我们现在考虑特定的 $(\Delta x, \Delta y)$, 其中 $\Delta x = -\dfrac{B}{A}\Delta y$, 则此时

$$o\left(\rho^2\right) = o\left((\Delta x)^2 + (\Delta y)^2\right) = o\left(\left(1 + \left(\frac{B}{A}\right)^2\right)(\Delta y)^2\right) = o\left((\Delta y)^2\right).$$

将 $\Delta x = -\dfrac{B}{A}\Delta y$ 代入到式子 (6.14) 中, 得到

$$\begin{aligned} f\left(x_0 - \frac{B}{A}\Delta y, y_0 + \Delta y\right) &= f(x_0, y_0) + \frac{1}{2}\cdot\frac{1}{A}\left[\left(A\cdot -\frac{B}{A}\Delta y + B\Delta y\right)^2 - (\Delta y)^2\left(B^2 - AC\right)\right] + o\left(\rho^2\right) \\ &= f(x_0, y_0) + \frac{1}{2}\cdot\frac{1}{A}\left[-(\Delta y)^2\left(B^2 - AC\right)\right] + o\left((\Delta y)^2\right) \\ &= f(x_0, y_0) - \frac{1}{2}\cdot\frac{B^2 - AC}{A}(\Delta y)^2 + o\left((\Delta y)^2\right). \end{aligned}$$

根据高阶无穷小的定义, 可以将 $o\left((\Delta y)^2\right)$ 替换成 $\alpha\cdot(\Delta y)^2$, 其中 α 代表一个无穷小量 (也可以直接用 $o(1)$ 表示一个无穷小量). 于是

$$\begin{aligned} f\left(x_0 - \frac{B}{A}\Delta y, y_0 + \Delta y\right) &= f(x_0, y_0) - \frac{1}{2}\cdot\frac{B^2 - AC}{A}(\Delta y)^2 + \alpha\cdot(\Delta y)^2 \\ &= f(x_0, y_0) - \left(\frac{1}{2}\cdot\frac{B^2 - AC}{A} + \alpha\right)(\Delta y)^2. \end{aligned}$$

于是

$$-\left(\frac{1}{2}\cdot\frac{B^2 - AC}{A} + \alpha\right)(\Delta y)^2 = f\left(x_0 - \frac{B}{A}\Delta y, y_0 + \Delta y\right) - f(x_0, y_0).$$

当 Δy 足够小时, $\Delta x = -\dfrac{B}{A}\Delta y$ 也会变得非常小, 因此最终 ρ 也会变得足够小. 由此可见, 当 Δy 充分小时, 表达式

$$f\left(x_0 - \frac{B}{A}\Delta y, y_0 + \Delta y\right) - f(x_0, y_0)$$

的符号与 $\dfrac{B^2 - AC}{A}$ 的符号相反, 进而与 A 的符号相反. 注意, 给定条件是 $\det(H) < 0$, 即 $B^2 > AC$, 因此 $\dfrac{B^2 - AC}{A}$ 的符号与分母 A 的符号相同.

ii. 另外, 我们考虑另一种特定的 $(\Delta x, \Delta y)$, 其中 $\Delta y = 0$. 将 $\Delta y = 0$ 代入到式子 (6.14) 中, 得到

$$\begin{aligned} f(x_0 + \Delta x, y_0) &= f(x_0, y_0) + \frac{1}{2}A(\Delta x)^2 + o\left((\Delta x)^2\right) \\ &= f(x_0, y_0) + \left(\frac{1}{2}A + o(1)\right)(\Delta x)^2. \end{aligned}$$

可见, 当 Δx 充分小时, $f(x_0 + \Delta x, y_0) - f(x_0, y_0)$ 的符号与 A 的符号相同. 即, 在该点的任何邻域内, 既存在使得函数值大于 $f(x_0, y_0)$ 的点, 也存在使得函数值小于 $f(x_0, y_0)$ 的点. 所以它必然不是极值点.

总之, $f(x, y) - f(x_0, y_0)$ 在点 (x_0, y_0) 的任何小邻域内总能取到相反的符号. 可见, $f(x_0, y_0)$ 不可能是极值. 对于 $C \neq 0$ 的情况, 证明完全类似.

(b) **(Case 2, 取 Case 1 的否定: 即非零的个数 <1, 即非零的个数 =0, 即全部是零)** 当 $A = C = 0$ 时, $B \neq 0$, 这时二阶的带佩亚诺余项的泰勒公式为 (极限 $\rho \to 0$)

$$f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) = B\Delta x\Delta y + o\left(\rho^2\right).$$

i. 特别地, 令 $\Delta y=\Delta x$, 则有

$$f\left(x_0+\Delta x, y_0+\Delta x\right)-f\left(x_0, y_0\right)=B(\Delta x)^2+o\left((\Delta x)^2\right)=(B+o(1))(\Delta x)^2.$$

ii. 而令 $\Delta y=-\Delta x$, 则有

$$f\left(x_0+\Delta x, y_0-\Delta x\right)-f\left(x_0, y_0\right)=-B(\Delta x)^2+o\left((\Delta x)^2\right)=(-B+o(1))(\Delta x)^2.$$

这里, 前一个等式告诉我们: 当 Δx 充分小时, $f\left(x_0+\Delta x, y_0+\Delta x\right)-f\left(x_0, y_0\right)$ 与 B 有相同的符号; 而后一个等式告诉我们: 当 Δx 充分小时, $f\left(x_0+\Delta x, y_0-\Delta x\right)-f\left(x_0, y_0\right)$ 与 B 有相反的符号. 这就证明了 $f\left(x_0, y_0\right)$ 不是极值.

以上证明了定理的第三种情况 (iii), 当 $B^2>AC$ 时, $f\left(x_0, y_0\right)$ 不是极值.

3. **现在考虑第四种情况 (iv), 假设** $\det(H)=0$**, 即** $B^2=AC$**.** 此时, $f\left(x_0, y_0\right)$ 可能是极值, 也可能不是极值. 这只要举出适当的例子就足够了:

   (a) 对于函数 $f(x, y)=xy$, 其在任意点的二阶偏导为

   $$f_{xx}=\frac{\partial}{\partial x}(y)=0, \quad f_{yy}=\frac{\partial}{\partial y}(x)=0, \quad f_{xy}=\frac{\partial}{\partial y}(y)=1.$$

   于是在点 $(0,0)$ 处的黑塞矩阵为

   $$H_f(0,0)=\left(\begin{array}{cc}f_{xx}(0,0) & f_{xy}(0,0) \\ f_{xy}(0,0) & f_{yy}(0,0)\end{array}\right)=\left(\begin{array}{cc}0 & 1 \\ 1 & 0\end{array}\right),$$

   所以满足 $B^2=AC$. 但是, 例题 6.9.2 说明了 $(0,0)$ 不是 $f(x, y)=xy$ 的极值点.

   (b) 对于函数 $f(x, y)=(x+y)^2$, 其在任意点的二阶偏导为

   $$f_{xx}=\frac{\partial}{\partial x}(2(x+y))=2, \quad f_{yy}=\frac{\partial}{\partial y}(2(x+y))=2, \quad f_{xy}=\frac{\partial}{\partial y}(2(x+y))=2.$$

   于是在点 $(0,0)$ 处的黑塞矩阵为

   $$H_f(0,0)=\left(\begin{array}{cc}f_{xx}(0,0) & f_{xy}(0,0) \\ f_{xy}(0,0) & f_{yy}(0,0)\end{array}\right)=\left(\begin{array}{cc}2 & 2 \\ 2 & 2\end{array}\right),$$

   所以满足 $B^2=AC$. 显然, $(0,0)$ 是 $f(x, y)=(x+y)^2$ 的极值点.

以上说明了当 $\det(H)=0$ 时, 我们什么都无法论断.

□

例 6.9.6 求函数 $f(x, y)=\frac{1}{3}\left(x^3+y^3\right)+xy$ 的稳定点, 并判别其是否是极值点.

证明. 首先, 计算函数 $f(x, y)$ 的一阶偏导数:

$$f_x=x^2+y, \quad f_y=y^2+x.$$

于是, $f(x,y)$ 的稳定点应该满足方程组

$$\begin{cases} x^2+y=0, \\ y^2+x=0. \end{cases}$$

这个方程组有两个解[47], 其对应的点为 $(0,0)$ 及 $(-1,-1)$, 它们就是全体稳定点. 接着, 计算 $f(x,y)$ 的二阶偏导数:

$$f_{xx}=\frac{\partial}{\partial x}\left(x^2+y\right)=2x,\quad f_{yy}=\frac{\partial}{\partial y}\left(y^2+x\right)=2y,\quad f_{xy}=\frac{\partial}{\partial y}\left(x^2+y\right)=1.$$

因此, 其黑塞矩阵为

$$H_f(x,y)=\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}=\begin{pmatrix} 2x & 1 \\ 1 & 2y \end{pmatrix}.$$

下面, 进一步评估在稳定点处的黑塞矩阵的情况.

1. 在点 $(0,0)$ 处有
$$H_f(0,0)=\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$
计算行列式
$$\det(H_f(0,0))=(0)(0)-(1)^2=-1<0.$$
故, 点 $(0,0)$ 是一个鞍点, 即既不是极大值点也不是极小值点.

2. 在点 $(-1,-1)$ 处有
$$H_f(-1,-1)=\begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}.$$
计算行列式
$$\det(H_f(-1,-1))=(-2)(-2)-(1)^2=4-1=3>0.$$
并且 $f_{xx}(-1,-1)=-2<0$. 故, 点 $(-1,-1)$ 是一个极大值点.

□

例 6.9.7 求函数 $f(x,y)=x^4+y^4-x^2-2xy-y^2$ 的稳定点, 并判别它们是否是极值点.

证明. 首先, 计算函数 $f(x,y)$ 的一阶偏导数:

$$f_x=4x^3-2x-2y,\quad f_y=4y^3-2x-2y.$$

于是, $f(x,y)$ 的稳定点应满足方程组

$$\begin{cases} 4x^3-2x-2y=0, \\ 4y^3-2x-2y=0. \end{cases}$$

为了求解这个方程组[48], 我们观察到两个式子说明了

$$4x^3=4y^3=2x+2y,$$

[47] 根据梯度为零建立的方程组通常是非线性, 其解法而言, 目前基本靠 “猜测” 和各类技巧. 实现中, 我们都是利用计算机来获得其近似的数值解.

[48] 根据梯度为零建立的方程组通常是非线性, 其解法而言, 目前基本靠 “猜测” 和各类技巧.

于是就有 $x=y$. 将 $x=y$ 代入方程组得到

$$4x^3-2x-2x=4x^3-4x=0 \quad\Rightarrow\quad 4x(x^2-1)=0.$$

因此, $x=0$ 或 $x=\pm1$. 对应的 y 值也是相同的. 因此, 函数 $f(x,y)$ 的稳定点包括

$$(1,1),\quad (-1,-1),\quad (0,0).$$

接下来, 计算 $f(x,y)$ 的二阶偏导数:

$$f_{xx}=12x^2-2,\quad f_{yy}=12y^2-2,\quad f_{xy}=-2.$$

因此, $f(x,y)$ 的黑塞矩阵为

$$H_f(x,y)=\begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy}\end{pmatrix}=\begin{pmatrix} 12x^2-2 & -2 \\ -2 & 12y^2-2\end{pmatrix}.$$

现在, 分别评估在各稳定点处的黑塞矩阵.

1. 在点 $(1,1)$ 处:

   $$H_f(1,1)=\begin{pmatrix} 10 & -2 \\ -2 & 10\end{pmatrix}.$$

   计算行列式

   $$\det(H_f(1,1))=(10)(10)-(-2)^2=100-4=96>0.$$

   并且 $f_{xx}(1,1)=10>0$. 因此, 点 $(1,1)$ 是一个极小值点.

2. 在点 $(-1,-1)$ 处:

   $$H_f(-1,-1)=\begin{pmatrix} 10 & -2 \\ -2 & 10\end{pmatrix}.$$

   计算行列式

   $$\det(H_f(-1,-1))=(10)(10)-(-2)^2=100-4=96>0.$$

   并且 $f_{xx}(-1,-1)=10>0$. 因此, 点 $(-1,-1)$ 也是一个极小值点.

3. 在点 $(0,0)$ 处:

   $$H_f(0,0)=\begin{pmatrix} -2 & -2 \\ -2 & -2\end{pmatrix}.$$

   计算行列式

   $$\det(H_f(0,0))=(-2)(-2)-(-2)^2=4-4=0.$$

   由于行列式为零, 所以我们什么都不能确定. 但是在 $(0,0)$ 附近具体分析函数表达式的特征可发现: 在直线 $y=-x$ $(0<x<1)$ 上, $f(x,-x)=2x^4>0$; 而在直线 $y=0$ $(0<x<1)$ 上, $f(x,0)=x^4-x^2=x^2\left(x^2-1\right)<0$. 又由于 $f(0,0)=0$, 故 $(0,0)$ 不是 $f(x,y)$ 的极值点.

综上所述, 函数 $f(x,y)$ 的稳定点为 $(0,0)$, $(1,1)$ 和 $(-1,-1)$. 其中, $(1,1)$ 和 $(-1,-1)$ 是极小值点, 而 $(0,0)$ 是一个鞍点, 不是极值点. □

上一道题的最后一种情况告诉了我们一种思路：在一个稳定点 (x_0, y_0) 处，当 $B^2 = AC$ 时，一般不能断定该稳定点是否为极值点. 但是，在一个具体的问题中，可以考查函数值在过点 (x_0, y_0) 的两条直线（或曲线）上的变化，如果在这两条直线（或曲线）上，在点 (x_0, y_0) 的充分小邻域内，$f(x, y) - f(x_0, y_0)$ 的符号不相同，就可以判定 (x_0, y_0) 不是极值点.

例 6.9.8 利用上述提示证明函数 $f(x, y) = xy$ 在点 $(0, 0)$ 不是极值点. 见例题 6.9.2.

证明. 我们考查函数值在经过点 $(0, 0)$ 的两条直线上的变化. 选择两条直线 $y = x$ 和 $y = -x$.

1. 将 $y = x$ 代入函数 $f(x, y)$，得

$$f(x, x) = x \cdot x = x^2.$$

在 $(0, 0)$ 的任意充分小邻域内，当 $x \neq 0$ 时，$f(x, x) = x^2 > 0$. 即，$f(x, x) - f(0, 0) = x^2 \geq 0$.

2. 将 $y = -x$ 代入函数 $f(x, y)$，得

$$f(x, -x) = x \cdot (-x) = -x^2.$$

在 $(0, 0)$ 的任意充分小邻域内，当 $x \neq 0$ 时，$f(x, -x) = -x^2 < 0$. 即，$f(x, -x) - f(0, 0) = -x^2 \leq 0$.

由此可见，在点 $(0, 0)$ 的任意充分小邻域内，$f(x, y) - f(0, 0)$ 在两条直线 $y = x$ 和 $y = -x$ 上的符号相反. 因此，点 $(0, 0)$ 不是极值点. □

例 6.9.9 自行证明：函数 $f(x, y) = x^3 - 3xy^2$ 在点 $(0, 0)$ 不是极值点.（提示：沿一条直线 $y = x$ 讨论足矣.）

6.9.2　二元函数的最值问题

与一元函数类似，可以利用二元函数的极值来求它的最大值与最小值. 我们知道，当函数 $f(x, y)$ 在闭区域 D 上连续时，它在 D 上就有最大值与最小值. 为了求最大（或小）值，应把所有的极大（或小）值与函数在区域边界上的值做比较，其中最大（或小）者就是最大（或小）值. 但这种做法比较复杂.（这通常行不通，跟一元情况不同，二元的定义域上边界点的个数往往是无穷多的.）

在实际应用中，常遇到下列较简单的特殊情况：*一方面，根据实际问题的性质，可以断言 $f(x, y)$ 在 D 的内部必有最大（或小）值；另一方面，又能求出 $f(x, y)$ 在 D 的内部只有唯一的一个极值点 (x_0, y_0).* 在这种情况下，当 $f(x_0, y_0)$ 是极大（或小）值时，它也就是 $f(x, y)$ 在 D 上的最大（或小）值.

例 6.9.10 最小二乘法 已知变量 y 是变量 x 的函数，由实验测得当 x 取 n 个不同的值 $x_1, x_2, \cdots, x_n$ 时，对应的 y 值分别为 $y_1, y_2, \cdots, y_n$. 试据此求一个最佳线性近似公式：

$$y = ax + b, \quad a, b \text{ 为待定常数},$$

使利用这个近似公式算出的 y 值与实验所得值的误差平方和

$$u(a, b) = \sum_{i=1}^{n} (ax_i + b - y_i)^2$$

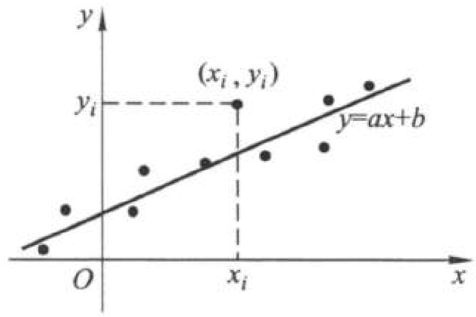


图 6.14

图 6.16: 最小二乘法

最小. 这里 $(ax_i+b-y_i)^2$ 表示点 (x_i,y_i) 的纵坐标与直线 $y=ax+b$ 上 x_i 对应点的纵坐标之差的平方 (见图 6.16). 如果所有数据均落在这条直线上, 那么 $u(a,b)=0$. 因此, 量 $u(a,b)$ 在一定程度上反映了数据组偏离直线 $y=ax+b$ 的大小.

证明. 该问题就是求二元函数 $u(a,b)$ 的最小值点. 因 $u(a,b)$ 在全平面上可微, 故其极值点必是稳定点. 考虑方程组

$$\begin{cases} u_a=\dfrac{\partial u}{\partial a}=\displaystyle\sum_{i=1}^{n}2\left(ax_i+b-y_i\right)x_i=0, \\ u_b=\dfrac{\partial u}{\partial b}=\displaystyle\sum_{i=1}^{n}2\left(ax_i+b-y_i\right)=0, \end{cases}$$

即

$$\begin{cases} \left(\displaystyle\sum_{i=1}^{n}x_i^2\right)a+\left(\displaystyle\sum_{i=1}^{n}x_i\right)b=\displaystyle\sum_{i=1}^{n}x_iy_i, \\ \left(\displaystyle\sum_{i=1}^{n}x_i\right)a+nb=\displaystyle\sum_{i=1}^{n}y_i, \end{cases}$$

这里 a,b 为未知量. 引入记号

- $S_x=\displaystyle\sum_{i=1}^{n}x_i$ (所有 x_i 的和)
- $S_{x^2}=\displaystyle\sum_{i=1}^{n}x_i^2$ (所有 x_i^2 的和)
- $S_y=\displaystyle\sum_{i=1}^{n}y_i$ (所有 y_i 的和)
- $S_{xy}=\displaystyle\sum_{i=1}^{n}x_iy_i$ (所有 x_iy_i 的和)

这样, 方程组可以写成

$$\begin{cases} S_{x^2}a+S_xb=S_{xy}, \\ S_xa+nb=S_y. \end{cases}$$

我们可以继续写成矩阵形式:

$$\begin{pmatrix} S_{x^2} & S_x \\ S_x & n \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} S_{xy} \\ S_y \end{pmatrix}.$$

用数学归纳法可证该方程组的系数行列式

$$nS_{x^2} - S_x^2 = n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{\substack{i=1 \\ i<j\le n}}^{n-1} (x_i - x_j)^2 > 0,$$

因而该方程组有唯一解 (a_0, b_0), 即 $u(a,b)$ 有唯一的稳定点.

接着计算, $u(a,b)$ 的二阶偏导,

$$u_{aa} = 2\sum_{i=1}^{n} x_i^2 = 2S_{x^2}, \quad u_{ab} = 2\sum_{i=1}^{n} x_i = 2S_x, \quad u_{bb} = 2n.$$

于是, 在点 (a_0, b_0) 处有黑塞矩阵

$$H = \begin{pmatrix} 2S_{x^2} & 2S_x \\ 2S_x & 2n \end{pmatrix}.$$

计算 H 的行列式, 得

$$\det(H) = (2S_{x^2})(2n) - (2S_x)(2S_x) = 4S_{x^2}n - 4S_x^2 = 4\left(nS_{x^2} - S_x^2\right) > 0.$$

(上式最后的结果就是刚刚用数学归纳法证明的东西) 又因为 $S_x = \sum_{i=1}^{n} x_i > 0$. 因而 (a_0, b_0) 是极小值点. 根据前面的讨论, 它也就是最小值点. 于是, 我们得到最佳线性近似公式

$$y = a_0 x + b_0,$$

其中 a_0, b_0 为上述方程组的解.

□

6.9.3　条件极值初步

之前讨论的所有极值问题或者最值问题, 对于自变量没有任何约束条件, 即可以取自然定义域上的任何可能变量, 这类问题叫做**无约束的最优化问题 (unconstrained optimization problem)**. 相对而言, 在自变量满足一定约束条件下求函数极值的问题, 称为条件极值问题, 或者**有约束的最优化问题 (constrained optimization problem)**.

例 6.9.11　比如, 求函数

$$z = x^2 + y^2$$

在约束条件 $x + y = 1$ 下的极小值, 这就是一个条件极值问题. 显然, 在这个具体问题中, 我们可以从约束条件 $x + y = 1$ 中将 y 解成 x 的函数, 然后代入函数的表达式中, 将问题化作一个一元函数极值问题. 然而, 一般来说, 当约束条件较复杂时, 将 y 解成 x 的函数在实际上未必总能做到. 有时即使能做到, 其计算也可能很复杂. 因此, 给出一种方法来避免这个过程是必要的.

下面我们介绍求条件极值的一种方法, 此方法称为**拉格朗日乘子法**. 使用这种方法可以避免从约束条件中求解隐函数的过程. 为了简便起见, 我们先讨论二元函数的条件极值问题. (本小节以下内容均是非正式推导过程, 不必太严谨. 优化问题是足矣写出一本几百页的教材, 我们这里的目的就是做个初步介绍.)

$$\boxed{\text{问题: 考虑函数 } z=f(x,y) \text{ 在约束条件 } \varphi(x,y)=0 \text{ 下的极值.}}$$

设 $f(x,y)$ 与 $\varphi(x,y)$ 都有连续的一阶偏导数, 并且 $\varphi_y(x,y)\neq 0$. 设想由所给的约束条件确定出隐函数 $y=y(x)$, 将它代入函数的表达式得到

$$z=f(x,y(x)).$$

这样, 条件极值问题就转化成求函数 $z=f(x,y(x))$ 的普通无约束的极值问题. 现在, 先求其稳定点, 即求解方程

$$\frac{\mathrm{d}z}{\mathrm{d}x}=f_x+f_y y'(x)=0.$$

由隐函数求导公式知 $y'(x)=-\dfrac{\varphi_x}{\varphi_y}$. 代入上式, 得到

$$f_x-f_y\frac{\varphi_x}{\varphi_y}=0,$$

即

$$\frac{f_x}{\varphi_x}=\frac{f_y}{\varphi_y}.$$

令 $\dfrac{f_y}{\varphi_y}=-\lambda$, 则条件极值点必须满足方程组

$$\begin{cases} f_x+\lambda\varphi_x=0,\\ f_y+\lambda\varphi_y=0,\\ \varphi(x,y)=0.\end{cases}$$

这就是**条件极值的一阶必要条件**. 为了便于记忆, 引进**辅助函数** (这种函数称作该极值问题的**拉格朗日函数**)

$$F(x,y,\lambda)=f(x,y)+\lambda\varphi(x,y),$$

则上述方程组恰好就是三元函数 $F(x,y,\lambda)$ 的关于变量 (x,y,λ) 普通极值点必须满足的条件. 注意, 这里新变量 λ 是用来辅助求解问题的, 叫做**拉格朗日乘子**. 尽管我们只对于变量 x,y 感兴趣, 但是辅助变量 λ 对我们的帮助也是必不可少的. 请大家不要对比感到奇怪.

根据以上讨论可知, **约束条件下的稳定点 (x,y) 及其相应的 λ, 一定是 $F(x,y,\lambda)$ 的稳定点**, 因此条件极值问题可归结为: 先求拉格朗日函数 $F(x,y,\lambda)$ 的稳定点, 再判断稳定点是否为极值点的问题. 总结以上讨论, 得到下面的结论.

结论 6.9.1: 拉格朗日乘子法 - 二元目标函数, 一个等式约束条件

要求二元函数 $f(x,y)$ 在约束条件 $\varphi(x,y)=0$ 下的极值点, 作辅助用的拉格朗日函数

$$F(x,y,\lambda)=f(x,y)+\lambda\varphi(x,y).$$

解方程组

$$\begin{cases} F_x=f_x+\lambda\varphi_x=0, \\ F_y=f_y+\lambda\varphi_y=0, \\ F_\lambda=\varphi(x,y)=0, \end{cases}$$

即可得约束条件下的稳定点. 再结合实际情况, 可判断所得稳定点中哪些是条件极大值点, 哪些是条件极小值点.

对于上述的方程组, 设法消去 λ 而得到关于 (x,y) 的解, 其便是约束条件下 $f(x,y)$ 的稳定点. 再判断所得的稳定点是否是条件极值点即可. 至于如何判断所得的稳定点是否是条件极值点, 我们不讨论一般的方法, 但在许多实际问题中往往可由问题本身的性质来判定. 比如, 在某些应用题中, 我们可以根据实际问题的性质断言最大 (或小) 值一定存在, 其稳定点却只有一个, 这时我们即可断定该稳定点必为极大 (或小) 值点, 也就是最大 (或小) 值点.

例 6.9.12 求平面上直线 $x+y=1$ 到原点 $(0,0)$ 距离最近的点.

证明. 我们的问题可以转化为在约束条件 $x+y=1$ 下求解目标函数 (距离) 最小的问题. 原点到点 (x,y) 的距离为 $d(x,y)=\sqrt{x^2+y^2}$. 为了方便计算, 我们可以最小化距离的平方

$$f(x,y)=x^2+y^2.$$

约束条件是直线 $x+y=1$, 即

$$\varphi(x,y)=x+y-1=0.$$

构造拉格朗日函数为

$$F(x,y,\lambda)=x^2+y^2+\lambda(x+y-1).$$

根据拉格朗日乘数法, 要求解以下方程组

$$\frac{\partial F}{\partial x}=0,\quad \frac{\partial F}{\partial y}=0,\quad \frac{\partial F}{\partial \lambda}=0.$$

即,

$$\begin{cases} 2x+\lambda=0, \\ 2y+\lambda=0, \\ x+y-1=0. \end{cases}$$

从第一和第二个方程, 我们得到

$$-2x=-2y \quad\Rightarrow\quad x=y.$$

将 $x=y$ 代入第三个方程, 解出

$$x=\frac{1}{2},\quad y=\frac{1}{2}.$$

结果实际, 显然, 直线 $x+y=1$ 到原点 $(0,0)$ 距离最近的点正是 $\left(\frac{1}{2},\frac{1}{2}\right)$. 我们得到了这个点的坐标 $\left(\frac{1}{2},\frac{1}{2}\right)$, 并且它正是直线 $x+y=1$ 与原点的连线的垂足. □

结论 6.9.2: 拉格朗日乘子法 - 三元目标函数, 一个等式约束条件

类似地, 要求三元函数 $f(x,y,z)$ 在约束条件 $\varphi(x,y,z)=0$ 下的极值点, 作辅助用的拉格朗日函数

$$F(x,y,z,\lambda)=f(x,y,z)+\lambda\varphi(x,y,z).$$

解方程组

$$\begin{cases} F_x=f_x+\lambda\varphi_x=0,\\ F_y=f_y+\lambda\varphi_y=0,\\ F_z=f_z+\lambda\varphi_z=0,\\ F_\lambda=\varphi(x,y,z)=0, \end{cases}$$

即可得约束条件下的稳定点. 再结合实际情况, 可判断所得稳定点中哪些是条件极大值点, 哪些是条件极小值点.

例 6.9.13 求函数 $f(x,y,z)=xyz$ 在球面 $x^2+y^2+z^2=R^2$ $(x,y,z>0;\ R>0)$ 上的最大值.

证明. 当 $x,y,z>0$ 时, $xyz>0$, 且当 (x,y,z) 趋向于球面 $x^2+y^2+z^2=R^2$ 在第一卦限中的三条边界

$$\begin{cases} y^2+z^2=R^2,\\ x=0, \end{cases}\quad \begin{cases} z^2+x^2=R^2,\\ y=0, \end{cases}\quad \text{或}\quad \begin{cases} x^2+y^2=R^2,\\ z=0, \end{cases}$$

时, $f(x,y,z)\to 0$. 故 $f(x,y,z)$ 在球面内部必达到最大值.

令 $F(x,y,z)=xyz+\lambda\left(x^2+y^2+z^2-R^2\right)$, 并解方程组

$$\begin{cases} F_x=yz+2x\lambda=0,\\ F_y=xz+2y\lambda=0,\\ F_z=xy+2z\lambda=0,\\ F_\lambda=x^2+y^2+z^2-R^2=0. \end{cases}$$

由方程组的前三个方程得

$$\lambda=-\frac{yz}{2x}=-\frac{xz}{2y}=-\frac{xy}{2z}.$$

由此推得

$$x^2=y^2=z^2,$$

再代入方程组中的第四个方程得 $3x^2=R^2$. 由此求得唯一稳定点

$$(x_0,y_0,z_0)=\left(\frac{R}{\sqrt{3}},\frac{R}{\sqrt{3}},\frac{R}{\sqrt{3}}\right).$$

我们已断定函数的最大值必在球面内部达到, 故这唯一的稳定点就是最大值点, 最大值为 $f\left(\frac{R}{\sqrt{3}},\frac{R}{\sqrt{3}},\frac{R}{\sqrt{3}}\right)=$

$\dfrac{R^3}{3\sqrt{3}}$, 即当 $x^2+y^2+z^2=R^2\ (x,y,z>0)$ 时, $xyz \le \dfrac{R^3}{3\sqrt{3}}$.[49] □

结论 6.9.3: 拉格朗日乘子法 - n 元目标函数, k 个等式约束条件

对于一般的有若干约束条件的多元函数条件极值问题, 也有相应的拉格朗日乘子法: 为了求函数 $f(x_1,\cdots,x_n)$ 在约束条件 $\varphi_1(x_1,\cdots,x_n)=0$, $\varphi_2(x_1,\cdots,x_n)=0,\cdots,\varphi_k(x_1,\cdots,x_n)=0$ 下的极值, 我们作辅助函数

$$F(x_1,\cdots,x_n,\lambda_1,\cdots,\lambda_k)=f(x_1,\cdots,x_n)+\lambda_1\varphi_1(x_1,\cdots,x_n)+\cdots+\lambda_k\varphi_k(x_1,\cdots,x_n).$$

解方程组

$$\begin{cases} F_{x_1}=f_{x_1}+\lambda_1\dfrac{\partial\varphi_1}{\partial x_1}+\cdots+\lambda_k\dfrac{\partial\varphi_k}{\partial x_1}=0, \\ \quad\vdots \\ F_{x_n}=f_{x_n}+\lambda_1\dfrac{\partial\varphi_1}{\partial x_n}+\cdots+\lambda_k\dfrac{\partial\varphi_k}{\partial x_n}=0, \\ F_{\lambda_1}=\varphi_1(x_1,\cdots,x_n)=0, \\ \quad\vdots \\ F_{\lambda_k}=\varphi_k(x_1,\cdots,x_n)=0, \end{cases}$$

即可得约束条件下的稳定点. 再结合实际情况, 可判断所得稳定点中哪些是条件极大值点, 哪些是条件极小值点.

例 6.9.14 平面 $x+y+z=1$ 截圆柱面 $x^2+y^2=1$ 的截痕是一个粗圆周. 求此椭圆周上离原点最近及最远的点.

证明. 空间中任意一点 (x,y,z) 到原点的距离为 $\sqrt{x^2+y^2+z^2}$. 该问题可化为求函数

$$f(x,y,z)=x^2+y^2+z^2$$

在约束条件

$$\begin{cases} x+y+z=1, \\ x^2+y^2=1, \end{cases}$$

下的最小值点及最大值点. 作辅助函数

$$F(x,y,z,\lambda_1,\lambda_2)=x^2+y^2+z^2+\lambda_1(x+y+z-1)+\lambda_2\left(x^2+y^2-1\right).$$

---

[49] 顺便指出, 以 $R=\sqrt{x^2+y^2+z^2}$ 代入上式, 得到 $xyz\le\left(\dfrac{x^2+y^2+z^2}{3}\right)^{\frac{3}{2}}$, 即 $x^2y^2z^2\le\left(\dfrac{x^2+y^2+z^2}{3}\right)^3$. 令 $x^2=a, y^2=b, z^2=c$, 即得

$$\sqrt[3]{abc}\le\frac{a+b+c}{3}.$$

上式说明, 三个正数 a,b,c 的几何平均值小于或等于其算术平均值.

求解方程组

$$\begin{cases} F_x = 2x + \lambda_1 + 2x\lambda_2 = 0, \\ F_y = 2y + \lambda_1 + 2y\lambda_2 = 0, \\ F_z = 2z + \lambda_1 = 0, \\ F_{\lambda_1} = x + y + z - 1 = 0, \\ F_{\lambda_2} = x^2 + y^2 - 1 = 0. \end{cases}$$

由方程组的前三个方程可推出

$$\begin{cases} x\left(1+\lambda_2\right) = z, \\ y\left(1+\lambda_2\right) = z. \end{cases}$$

下面分两种情况求解:

1. $\lambda_2 = -1$. 这时, 可推出 $z = 0$, 再由方程组的后两个方程即可推出 $x = 1, y = 0$ 或 $x = 0, y = 1$, 从而求得两个稳定点

$$(1,0,0) \text{ 及 } (0,1,0).$$

2. $\lambda_2 \neq -1$. 这时, 由方程组可推出

$$x = y = \frac{z}{1+\lambda_2},$$

再由方程组的后两个方程可得两个稳定点

$$\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 1-\sqrt{2}\right) \quad \text{及} \quad \left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 1+\sqrt{2}\right).$$

比较四个稳定点到原点的距离可判断: 点 $(1,0,0)$ 及 $(0,1,0)$ 离原点最近, 而点 $\left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 1+\sqrt{2}\right)$ 离原点最远. □

我们注意到, 本节讨论的约束条件全是通过等式约束 $\varphi(x,y) = 0$ 给出来的. 现实中, 我们还需要考虑不等式约束条件 $\varphi(x,y) \geq 0$. 比如, 假设我们要最小化 $f(x,y) = 4x + 3y$, 同时满足不等式约束 $x + y \geq 2$ 和 $x, y \geq 0$. 目前为止, 本书不讨论这种不等式约束的极值问题. 这类问题无法通过拉格朗日乘子法直接求解. 换句话说, **拉格朗日乘子法仅是面对等式约束下极值问题的求解办法**.

第七章 附录

.1 希腊字母简表

本表格字母内容来自《现代汉语词典》(2002 增补本)

字母名称	英语音标	大写	小写	字母名称	英语音标	大写	小写
alpha	/'ælfə/	A	α	nu	/nju:/	N	ν
beta	/'bi.tə/ 或 /'beita/	B	β	xi	希腊 /ksi/; 英美/'zai/ 或/'ksai/	Ξ	ξ
gamma	/'gæmə/	Γ	γ	omicron	/əu'maikrən/ 或 /'amikran/	O	o
delta	/'deltə/	Δ	δ	pi	/pai/	Π	π
epsilon	/'epsilon/	E	ε	rho	/rəu/	P	ρ
zeta	/'zi:tə/	Z	ζ	sigma	/'sigmə/	Σ	σ, ς
eta	/'i:tə/	H	η	tau	/to:/ 或 /tav/	T	τ
theta	/'θi:tə/	Θ	θ	upsilon	/'ipsilon/ 或 /'upsilən/	Υ	υ
iota	/ai'əutə/	I	ι	phi	/fai/	Φ	φ
kappa	/'kæpə/	K	κ	chi	/kai/	X	χ
lambda	/'læmdə/	Λ	λ	psi	/psai/	Ψ	ψ
mu	/mju:/	M	μ	omega	/'əumigə/ 或 /o'mega/	Ω	ω

表 1: 希腊字母表及其英语音标

.2 二项式定理

本附录内容摘自 二项式定理 - 维基百科.

二项式定理 (Binomial theorem) 描述了二项式的幂的代数展开. 根据该定理, 可以将两个数之和的整数次幂诸如 $(x+y)^n$ 展开为类似 ax^by^c 项之和的恒等式, 其中 b, c 均为非负整数且 $b+c=n$. 系数 a 是依赖于 n 和 b 的正整数. 当某项的指数为 0 时, 通常略去不写. 例如:

$$(x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4.$$

根据此定理, 可以将 $x+y$ 的任意次幂展开成和的形式

$$(x+y)^n = \binom{n}{0}x^ny^0 + \binom{n}{1}x^{n-1}y^1 + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n-1}x^1y^{n-1} + \binom{n}{n}x^0y^n$$

其中每个 $\binom{n}{k}$ 为一个称作二项式系数的特定正整数, 其等于 $\frac{n!}{k!(n-k)!}$. 这个公式也称二项式公式或二项恒等式. 使用求和符号, 可以把它写作

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

后面的表达式只是将根据 x 与 y 的对称性得出的, 通过比较发现公式中的二项式系数也是对称的. 二项式定理的一个变形是用 1 来代换 y 得到的, 所以它只涉及一个变量. 在这种形式中, 公式写作

$$(1+x)^n = \binom{n}{0} x^0 + \binom{n}{1} x^1 + \binom{n}{2} x^2 + \cdots + \binom{n}{n-1} x^{n-1} + \binom{n}{n} x^n.$$

.3 三角函数和反三角函数之间的关系

本附录内容摘自 Inverse trigonometric functions - Wikipedia.

θ	$\sin(\theta)$	$\cos(\theta)$	$\tan(\theta)$
$\arcsin(x)$	$\sin(\arcsin(x)) = x$	$\cos(\arcsin(x)) = \sqrt{1-x^2}$	$\tan(\arcsin(x)) = \frac{x}{\sqrt{1-x^2}}$
$\arccos(x)$	$\sin(\arccos(x)) = \sqrt{1-x^2}$	$\cos(\arccos(x)) = x$	$\tan(\arccos(x)) = \frac{\sqrt{1-x^2}}{x}$
$\arctan(x)$	$\sin(\arctan(x)) = \frac{x}{\sqrt{1+x^2}}$	$\cos(\arctan(x)) = \frac{1}{\sqrt{1+x^2}}$	$\tan(\arctan(x)) = x$
$\text{arccot}(x)$	$\sin(\text{arccot}(x)) = \frac{1}{\sqrt{1+x^2}}$	$\cos(\text{arccot}(x)) = \frac{x}{\sqrt{1+x^2}}$	$\tan(\text{arccot}(x)) = \frac{1}{x}$
$\text{arcsec}(x)$	$\sin(\text{arcsec}(x)) = \frac{\sqrt{x^2-1}}{\lvert x\rvert}$	$\cos(\text{arcsec}(x)) = \frac{1}{x}$	$\tan(\text{arcsec}(x)) = \text{sgn}(x)\sqrt{x^2-1}$
$\text{arccsc}(x)$	$\sin(\text{arccsc}(x)) = \frac{1}{x}$	$\cos(\text{arccsc}(x)) = \frac{\sqrt{x^2-1}}{\lvert x\rvert}$	$\tan(\text{arccsc}(x)) = \frac{\text{sgn}(x)}{\sqrt{x^2-1}}$

.4 符号 ∀ 和 ∃

1. ∀ (全称量化符号) 读作 “for all” 或 “对所有”. 表示某个命题对集合中的所有元素都成立. 例如: $\forall x \in \mathbb{R}, x^2 \geq 0$. 意思是: 对于所有实数 x, $x^2 \geq 0$ 成立. 这个符号常用于陈述普遍规律或命题.

2. ∃ (存在量化符号) 读作 “there exists” 或 “存在”. 表示在某个集合中至少存在一个元素使得某个命题成立. 例如: $\exists x \in \mathbb{R}, x^2 = 1$. 意思是: 存在一个实数 x, 使得 $x^2 = 1$. 常用于表达某个命题在至少一个特定元素上成立. 数学证明中, 使用这个符号来说明某种性质的存在性.

3. 嵌套使用: $\forall x \in \mathbb{R}, x \geq 0, \exists y \in \mathbb{R}, y^2 = x$. 意思是: 对于每一个正数 x, 都存在一个实数 y, 使得 $y^2 = x$.