高中数学讲义

谢子涵

2368589718@qq.com

2022 年 9 月 7 日———2024 年 9 月 22 日

第四版序言

这一版是我在高中阶段写的最后一个版本。

首先是**板块问题**，这版将会主要分为：集合论与逻辑学基础、代数学、微积分、初等数论、概率论与数理统计、几何学以及组合学这七个部分，这只是相对笼统的表示会涉及到的内容，例如还会有运筹学（管理数学）、复分析等内容；这样区分是为了更好的衔接高中的内容。

第二是**排版问题**，之前三版的文章结构都是 Article（文章）类，而这一版改成了 Book（书）类；这一点主要区别在文章的子级标题上；此外，页面排版也做了调整，一个是区分了奇偶页，另一个是扩大了每一页的空间利用率，这两点改动主要体现在打印该文章时，更符合阅读习惯以及更省钱。

第三是**内容**，我会向竞赛、强基、高考偏难怪以及大学自招的知识点靠拢，一部分过于超纲的内容以及我现阶段无力证明的内容我也会写，例如代数学基本定理、Gosper 算法，这一方面是时间问题，另一方面是前置知识的问题；最后是有关实用性，我新介绍的可能（亦或极大可能）对一般高考无用，但却是我为本科准备的基础，所以请读者斟酌自身情况，有取舍的看待这篇讲义。

第四是一些**记号问题**，关于有一些记号不标准，以及像是一部分是这个记号另一部分是另外的记号；这是前几版遗留下的不规范的记号导致的，例如在书写向量时，如果用单一字母表示向量，应为粗体斜体（具体好像不是这么说的），而我在前几版习惯性的用手写体表示向量，我写的和标准的对照如下

$$\vec{a}, \quad \boldsymbol{a}$$

这种问题因为时间关系，我只能尽量去改。其次还有一种书写不规范，当我在阅读日本教辅以及美国教材时我发现各个国家的“排列数”和“组合数”标记并不统一，规范来讲，组合数应写为

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}$$

总的来说，在之后写的文章都会统一向国际标准的符号靠拢。

第五是附录改动，由于要更加深入的学习，之前的附录更多是将一些附加的知识，但这一版的附录更多是负责记录一些需要反复用到的东西，之前的序言中有累加和与连乘积的介绍，现在我把一些常用的符号都归为附录，读者需要时可查看，特别是某章节中突然出现的新符号，如果在该节都没有定义，那读者可以在附录找找。

第三版序言

这是以中国普通高中应试教育为背景编写的一套讲义，该讲义内容包括：高中数学书上的基础知识点、以及被广泛使用的一、二级结论，以及作为常识的补充内容。

该讲义适用于初学者对于知识体系的构建，入门者对于知识理解的巩固，以及一、二轮复习的查漏补缺；

高中数学笼统的可以分为六个模块：线性代数 & 解析几何、数学分析、复变函数、组合学、概率论、数理统计；如果读者看过这些模块的书，不难知道其实高中数学就只是本科的一个引子，学的知识都是一些基础的东西，那么为什么高中数学难呢？难就难在应试上，不仅要对知识的充分掌握，而且要在有限的时间内判断、思考套路和反套路的题目，这不仅是对应试技巧的把握，还有的是日复一日培养的状态，以及考场上的运气，因此高中数学难在这些方面。

本科的线性代数在高中并不会涉及到，但数学分析和高等代数（内包含线性代数）是本科中最为基础的两门课程，因此解析几何常在线性代数中调用内容来用，所以高中课程里我把线性代数和解析几何捏成一个模块；此外高中的几何内容除了解析几何的内容外还有欧几里得的纯几何内容，由纯几何的定理推导出解析几何的定理。

这讲义我准备写两个版的，一个是顺序版，内容和新版高中数学人教 A 版的排布一样，我想这有利于预习的顺序，而且教师在拿到这本讲义后可以查查有哪些一二级结论没说、哪些方法没讲的，所以这个版本中我会尽量讲课本以及常见的推论（除导数外）；

另一个版本是混序版，内容由以上说过的七个模块组成，更适合于一二轮复习的人使用，这样的排版更加具有知识的连贯性，因此更有利于复习，两个版本内容无异，主要是知识排布，像是顺序版，一些额外的知识需要额外新建章节，其实额外的知识都是些课本上没有，但对于研究其模块应一种“常识”的存在。

另一方面，高中教课书不仅是章与章之间的内容会互相调用，甚至章内的小节间也会相互调用，所以我认为混序版更具有线性的认识。

在应用费曼学习法的过程中，我意识到自己在文字上的不足，定义、定理的描述缺乏揣摩，一些文字显得稚嫩，也许第一次学习中用自己的语言可以阐述，但是以讲义的角度、为了未来的回溯拾遗，我会更倾向于站在前人的肩膀上，用参考文献的描述。

数学的学习中做题刷题是必要的，“弱水三千，只取一瓢”，这本讲义因为篇幅有限，基本上没有例题，不过为了把方法讲的更明白，会有一些简单的例子引入，之后解决较难或较一般的问题，如果读者酷爱刷题，或觉得没有实践会记不牢靠，那么可以看看最后的参考文献部分，或许对有需要的有帮助。

这一版中，内容都是根据教材以及要证的推论而写的，基本上不存在多余的不必要的知识；例如书中这样一个逻辑链条：为什么写“对勾飘带函数”，因为进一步要写“飘带放缩”，为什么要写飘带放缩？因为飘带放缩是证明对数均值不等式高中最常用的方法；为什么要证明对数均值不等式？因为极值点偏移这一类题目中经常会用到对数均值不等式；因此构成了这个逻辑链条。

这本讲义中，虽然会有一些模型，但我一再强调，熟悉模型不过是理解推导过程与理解性质的副产物，读者可以不记住任何结论，但推导过程是思路的体现，运用模型是化归的表现，不可本末倒置。

鉴于本人水平有限，书中错误和不妥之处，敬请读者批评指正。

谢子涵
2023 年 12 月 3 日

[1] [2] [3] [4]

目录

第一部分

集合论基础和逻辑基础

第一章 集合

高中集合内容与数学分析的集合内容大同小异，所以我会直接引用数学分析的内容来写，对于高中没有的内容会有特别标注。

§1.1 集合有关的一些定义

集合

集合是数学中不予定义的原始对象，对于原始概念“集合”虽不予定义，但是可以描述；集合是具有某种特定性质的具体或抽象的对象汇集成的总体，这些对象称为该集合的元素。

一般情况下，用大写字母 $A, B, C, \ldots$ 表示集合，用小写字母 $a, b, c, \ldots$ 表示元素。集合可以通过 $A = \{a, b, c, \ldots\}$ 来表示集合 A 含有元素 $a, b, c, \ldots$

例如：由 2∼5 之间整数组成的集合 A，其中 A 有 2,3,4,5 这四个元素。可表示为 $A = \{2, 3, 4, 5\}$

集合的性质：无序性，互异性以及元素具有确定性。
无序性：元素先后位置不会影响集合，如:$\{a, b, c\} = \{b, c, a\}$
互异性：元素之间互不相同，如不存在 $\{a, a, b\}$
确定性：元素是一个确定的对象，高中数学不研究模糊性质，如不存在元素“高个子”。

空集：不包含任何元素的集合，用 ϕ 表示
全集：含有研究的所有对象的集合

集合的表示方法：
<u>枚举法</u>: 将元素逐一列举，如 $A = \{a, b, c, d\}$
<u>描述法</u>：描述元素具有的性质 P, 如 $A = \{x | x \text{ 具有性质 } P\}$
高中还有特别说明用 Venn 图以及区间可表示集合

§1.1.1 集合与元素、集合的关系

集合与元素的关系：

$$x \in S, (x\text{属于}S)$$
$$x \notin S, (x\text{不属于}S)$$

也有用 $x \overline{\in} S$ 来表示“不属于”关系。

集合与集合的关系：

$S \subset T$, 意为 $\forall x \in S \Rightarrow x \in T$, 称为 S 是 T 的子集 (高中用 $S \subseteq T$)
$S \not\subset T$, 意为 $\exists x \in S \Rightarrow x \notin T$，故 S 不是 T 的子集 (高中用 $S \nsubseteq T$ 表示)
$S = T \iff S \subset T, T \subset S$
$S \subsetneqq T$, 意为 $S \subset T, \exists x \in T, x \notin S$,$S$ 称为 T 的真子集

区间表示：

$$
\begin{aligned}
(a,b) &= \{x|a < x < b\}\text{, 闭区间} \\
[a,b] &= \{x|a \le x \le b\}\text{, 开区间} \\
(a,b] &= \{x|a < x \le b\}\text{, 半开半闭区间}
\end{aligned}
$$

常用集合：$\mathbb{R}$ 实数集,$\mathbb{Q}$ 有理数集,$\mathbb{Z}$ 整数集,$\mathbb{N}$ 自然数集[1]，其中蕴含关系：$\mathbb{R} \supset \mathbb{Q} \supset \mathbb{Z} \supset \mathbb{N}^+$

计算一个集合的子集个数，如果集合 $T = \{a_1, a_2, a_3, \dots, a_n\}$，则共有 2^n 个子集，$2^n - 1$ 个真子集，$2^n - 2$ 个非空真子集，其结论由组合数推出[2]

$$\sum_{i=0}^{n} \binom{n}{i} = 2^n$$

§1.2 集合的运算

集合的基本运算分为交并补差[3]运算 (点击1.2查看)

配上 Venn 图更好理解每个运算。

[1] $\mathbb{N}^+$ 或 $\mathbb{N}^*$ 正整数集
[2] $\binom{n}{i}$ 就是组合数，高中用 C_n^i 表示，上下是相反的
[3] 高中不学差运算

交集：$S\bigcap T=\{x|x\in S$ 且 $x\in T\}$

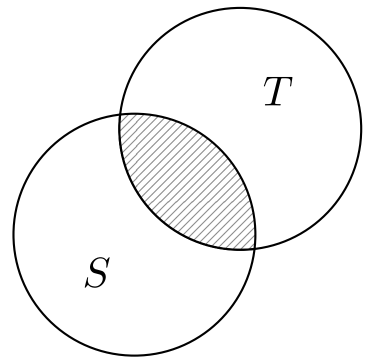


图 1.1: 交集 $S\bigcap T$

并集：$S\bigcup T=\{x|x\in S$ 或 $x\in T\}$

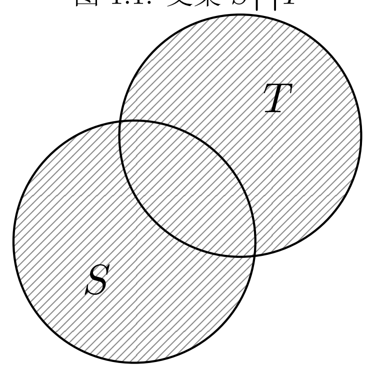


图 1.2: 并集 $S\bigcup T$

补集：S_X^C 或 $\complement_X S=X\backslash S$
通常将 S_X^C 记为 S^C,X 为全集。

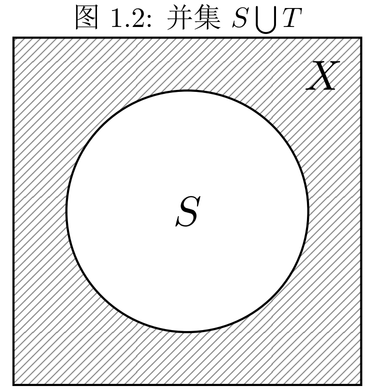


图 1.3: 补集 S_X^C 或 $\complement_X S$

差集：$S\backslash T\{x|x\in S$ 且 $x\notin T\}$

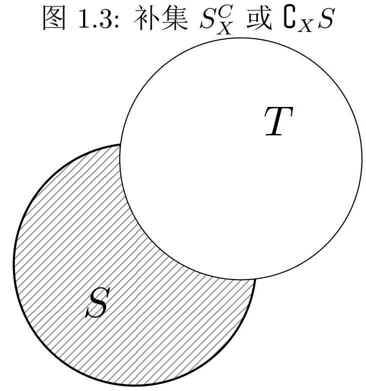


图 1.4: 差集 $S\backslash T$

交换律：

$$A\bigcup B = B\bigcap A$$
$$A\bigcap B = B\bigcap A$$

结合律：

$$A\bigcup\left(B\bigcup C\right) = \left(A\bigcup B\right)\bigcup C$$
$$A\bigcap\left(B\bigcap C\right) = \left(A\bigcap B\right)\bigcap C$$

分配律：

$$A\bigcap\left(B\bigcup C\right) = \left(A\bigcup B\right)\bigcap\left(A\bigcup C\right)$$
$$A\bigcup\left(B\bigcap C\right) = \left(A\bigcup B\right)\bigcap\left(A\bigcup C\right)$$

对偶率[4]

$$\left(A\bigcup B\right)^C = A^C\bigcap B^C$$
$$\overline{A\bigcup B} = \overline{A}\bigcap\overline{B}$$

有限集与无限集：有限个元素组成为有限集，不是有限集的集合为无限集 (高中一般研究有限集)

集合 A 的元素个数，高中常用 $\mathrm{card}A$ 表示，有

$$\mathrm{card}\left(A\bigcup B\right) = \mathrm{card}A + \mathrm{card}B - \mathrm{card}\left(A\bigcap B\right)$$

[4] 又叫德摩根率，有些旧书叫反演率

第二章　数理逻辑基础

§2.1　命题表达式

§2.2　充分条件与必要条件

对于两个事件 p, q, 如果：

1. $p \Rightarrow q$，则称事件 p 是事件 q 的充分条件，而 q 是 p 的必要条件。

2. $p \nRightarrow q$, 则称事件 p 不是事件 q 的充分条件，而 q 不是 p 的必要条件。

3. $p \Leftrightarrow q$, 则称事件 p 是事件 q 的充分必要条件简称充要条件

举个例子：$p : x < 3, q : x < 2$, 不难知道，p 是 q 的必要不充分条件，也就是 $p \nRightarrow q, q \Rightarrow p$

§2.3　全称量词与存在量词

“若 p 则 q”称为一个命题

1.“所有的，任意一个”为全称量词，用 $\forall$ 表示

2.“存在一个，至少有一个”为存在量词，用 $\exists$ 表示

事件 p 的否定用 $\neg p$ 表示，概率论中也有类似表示如概率空间中 A 的反，就是 $\overline{A}$
$p \wedge q$ 为 p 且 q, 可以想为串联电路或交集
$p \vee q$ 为 p 或 q, 可以联想并联电路或并集

1. 互逆命题：“若 p 则 q”的逆命题为“若 q 则 p”

2. 互否命题：“若 p 则 q”的否命题为“若 $\neg p$ 则 $\neg q$”

3. 互为逆否命题：“若 p 则 q”的逆否命题为“若 $\neg q$ 则 $\neg p$”

原命题	逆命题	否命题	逆否命题
真	真	真	真
真	假	假	真
假	真	真	假
假	假	假	假

主要记住原命题与逆否命题真假一致就行了。
对于全称命题的否定：

$$\forall x \in S, p(x)$$

$$\exists x \in S, \neg p(x)$$

同理对特称量词的否定：

$$\exists x \in S, p(x)$$

$$\forall x \in S, \neg p(x)$$

第二部分

代数学

第三章　函数的概念与性质

§3.1　映射

映射是两集合的一种对应关系。

定义

设集合 X,Y，其中 $x \in X, y \in Y$，若通过规则 f，使任意 x，有唯一 y 对应，则称 f 为 X 到 Y 的**映射**，记为

$$
\begin{aligned}
&f: X \to Y \\
&x \mapsto y = f(x)
\end{aligned}
$$

其中 y 称为映射 f 的之下 x 的**像**。

x 称为映射 f 之下 y 的**逆像**或**原像**。

X 称为**定义域**，记为 D_f，Y 为陪域[a]

x 的像 y 的全体称为**值域**，记为 R_f，$R_f \subset Y$。

$$R_f = \{y|y \in Y \wedge y = f(x), x \in X\}$$

[a]陪域 (Codomain) 又称上域、到达域，一般教材中不会特别说明

从定义可以看出构成一个映射的三要素（高中为函数的三要素）：

1. 定义域 X，即 $D_f = X$。
2. 陪域 Y，$R_f \subset Y$，陪域包含值域，限制值域的范围。
3. 对应规则 f，使每一个 $x \in X$，有唯一的 $y = f(x)$。

映射要求元素的像是唯一的，但不要求逆像也具有唯一性，若逆像唯一则有如下规定

若对于映射：$f: X \to Y$，f 的逆像也唯一，则 f 为**单射**，若 $R_f = Y$，则 f 称为**满射**

若又是单射又是满射，则 f 称为**双射**或**一一对应**。

§3.1.1　逆映射

设 $f: X \to Y$ 是单射，有对应关系：

$$
\begin{aligned}
&g: R_f \to X \\
&y \mapsto x\,(f(x) = y)
\end{aligned}
$$

构成 R_f 到 X 的映射，把它称为 f 的逆映射，记为 f^{-1}，其定义域为 $D_{f^{-1}} = R_f$，值域为 $R_{f^{-1}} = X$。

显然只要逆映射 f^{-1} 存在，那它一定就是 R_f 到 X 的双射。

复合映射

设有两个映射

$$
\begin{aligned}
g : X &\to U_1 \\
x &\mapsto u = g(x)
\end{aligned}
$$

和

$$
\begin{aligned}
f : U_2 &\to Y \\
u &\mapsto y = f(u)
\end{aligned}
$$

如果 $R_g \subset U_2 = D_f$，那么就可以构造一个新的对应关系

$$
\begin{aligned}
f \circ g : X &\to Y \\
x &\mapsto y = f\left(g(x)\right)
\end{aligned}
$$

这是一个映射，我们将之称为 f 和 g 的**复合映射**（关键在于 $R_g \subset D_f$ 是否成立）

要注意 $f \circ g$ 与 $g \circ f$ 一般是不同的。显然可以得到两个恒等式：

$$
\begin{aligned}
f \circ f^{-1}(y) = y, y \in R_f \\
f_{-1} \circ f(x) = x, x \in X
\end{aligned}
$$

§3.1.2 一元实函数

设 $X \subset \mathbb{R}, Y \subset \mathbb{R}$ 对于映射

$$
\begin{aligned}
f : X &\to Y \\
x &\mapsto y = f(x)
\end{aligned}
$$

称为一元实函数，简称函数。x 为自变量，y 为因变量，f 函数关系。

§3.2 初等函数

以下六类函数是我们经常接触到的

1. 常数函数：$y = c$
2. 幂函数：$y = x^\alpha (\alpha \in \mathbb{R})$
3. 指数函数：$y = a^x (a > 0 \land a \neq 1)$
4. 对数函数：$y = \log_a x (a > 0 \land a \neq 1)$
5. 三角函数：如 $y = \sin x, y = \cos x, y = \tan x, y = \cot x$ 等
6. 反三角函数：如 $y = \arcsin x, y = \arccos x, y = \arctan x$ 等

它们被统称为**基本初等函数**。

由基本初等函数经过有限次的四则运算与复合运算产生的函数称为**初等函数**。

初等函数的自然定义域是指自变量的最大取值范围

e.g.

$\sin x$ 的自然定义域就是 $\mathbb{R}$，

$\log_a x (a > 0, a \neq 1)$ 的自然定义域就是 $\mathbb{R}^+ = (0, +\infty)$。

当两个函数的 f, D_f 相同时（即对应规则和定义域相同），则它们为同一函数。

§3.3 函数表示

§3.3.1 分段表示

设 A, B 是两个互不相交的实数集合，即 $A \cap B = \varnothing$，且 $A \cup B \neq \varnothing$，$\psi(x), \varphi(x)$ 分别是定义在 A, B 上的函数，则

$$f(x) = \begin{cases} \psi(x) & , \quad x \in A \\ \varphi(x) & , \quad x \in B \end{cases}$$

函数 $f(x)$ 是定义域为 $A \cup B = D_f$ 的函数，这样的函数表示称为函数的**函数的分段表示**。这里只分了两段，但其实可以分为任意有限段，甚至无限多段。

下面举几个非常常见的分段函数：

e.g. 符号函数 $\operatorname{sgn} x$ 定义如下

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

其定义域为 $D_f = \mathbb{R}$，值域为 $R_f = \{-1, 0, 1\}$。

e.g. 高斯函数，也叫取整函数记为 $y = [x]$，定义如下

$$y = [x] = n, n \leqslant x < n + 1, n \in \mathbb{Z}$$

其定义域为 $D_f = \mathbb{R}$，值域为 $R_f = \mathbb{Z}$。

§3.3.2 显式与隐式表示

像 $y = f(x)$，因变量单独放等式的一边，另一边是只含自变量的表达形式，这称为**函数的显示表示**。

而函数的隐式表示，是指通过方程 $F(x, y) = 0$ 来确定变量 y 与 x 的关系，这也是一种重要的表示形式。

当只考虑圆方程的上或下半圆时，$x^2 + y^2 = R^2, y \geqslant 0$（或 $y \leqslant 0$）就是它的隐式表示。

§3.3.3 参数表示

表示 x, y 之间的关系时，需要引入参数 t，通过建立 t 与 x，t 与 y 之间的关系来确定 x, y 之间的关系，即

$$\begin{cases} x = x(t) & , \quad t \in [a, b] \\ y = y(t) & , \quad t \in [a, b] \end{cases}$$

这种方法叫函数的参数表示[1]。

§3.4　函数的简单特性

§3.4.1　有界性

定义

若存在两个常数 m, M，使得函数 $y = f(x), x \in D_f$ 满足

$$m \leqslant f(x) \leqslant M, x \in D_f$$

则称函数 f 在 D_f **有界**，其中 m 为它的**下界**，M 为它的**上界**。

易知一个函数有界时它的上界和下界是不唯一的。

另一个有界函数定义是：

存在常数 $M > 0$ 使函数 $y = f(x), x \in D_f$ 满足 $|f(x)| \leqslant M, x \in D_f$，容易知道这两种定义是等价的。

§3.4.2　单调性

定义

对函数 $y = f(x), x \in D_f$，$\forall x_1, x_2 \in D_f, x_1 \leqslant x_2$ 都有 $f(x_1) \leqslant f(x_2)$，则称函数 f 在 D_f 上**单调增加**，记作 $f \uparrow$ 当 $\leqslant$ 变为 $<$ 时称为**严格单调增加**，记作 f 严格 $\uparrow$，同理可定义单调减少和严格单调减少。

很多函数在自然定义域中并非单调的，但在较小的范围内有单调性，例如 $y = x^2$。

在一个区间的连续的单射函数是单调的，但需要注意的是单调性和连续性既不充分也不必要。

§3.4.3　奇偶性

函数 f 的定义域 D_f 关于原点对称，即 $x \in D_f \iff -x \in D_f$

偶函数

若对一切 $x \in D_f$ 有 $f(-x) = f(x)$ 则 f 是偶函数。

奇函数

若对一切 $x \in D_f$ 有 $-f(x) = f(-x)$ 则 f 是奇函数。

特别的，当 $x = 0$ 处有定义，则 $f(0) = 0$

可以结合图像理解，奇函数关于原点对称，偶函数关于 y 轴对称。

了解了函数的奇偶性，有时我们只需要研究 $D_f \cap [0, +\infty)$ 上 $f(x)$ 的性质，再由对称性可得另一半的性质。

现设 $g(x) = a + f(x)$，其中 $f(x)$ 为奇函数，若函数有最大值 M、最小值 m，那么可以得到一个结论：

$$g(M) + g(m) = 2a$$

The proof is trivial,and is left as an exercise.

[1] tikz 宏包中，绘制函数图像就是用的参数表示方法

由以上结论我们可以见微知著，推导出函数对称性更加一般的结论。

推论

若函数 $f(x)$ 关于直线 $x=c$（c 是某个常数）对称，则任意满足 $\frac{a+b}{2}=c$ 的一对 a,b，都可使下式成立

$$f(a+x)=f(b-x)$$

特别的，当 $c=0$ 时，$b=-a$，即

$$f(a+x)=f(-(a+x))$$

显然 $c=0$ 时 $f(x)$ 是偶函数。◇

这个推论是不难证明的，读者可以函数平移变换的性质自行证明。

类似的，可以得到函数关于某点对称时的推论。

推论

若函数 $f(x)$ 关于点 (a,b) 对称，则有

$$f(a-x)+f(a+x)=2b$$

特别的，当 $a=b=0$ 时，$f(x)$ 是奇函数。◇

定理

函数 $f(x)$ 是奇函数，则 $f'(x)$ 是偶函数；如果 $f(x)$ 是偶函数，则 $f'(x)$ 是奇函数。♠

证明：若 $f(x)$ 为奇函数，则有

$$f(x)=-f(-x)$$

两边求导得：

$$f'(x)=f'(-x)$$

当 $f(x)$ 是偶函数时证明类似。

§3.4.4 周期性

若存在常数 $T>0$ 使得对一切 $x\in D_f$，成立 $f(x+T)=f(x)$，则称函数 f 是周期函数，T 为它的周期，若存在满足上述条件的最小的 T，则称它为最小周期。

但并非每个周期函数都有最小周期的，如迪利克雷函数不存在最小周期。

一个函数是周期函数，那么它的导函数也是一个周期函数，这是显然的

$$f(x)=f(x+T)$$

可以推出：

$$f'(x)=f'(x+T)$$

□

如果已知函数 $f(x)$ 关于 $x=c$ 对称，且关于点 (a,b) 对称，则这个函数是一个周期函数。

§3.4.5　伸缩平移

函数的平移变化可以通过“左加右减，上加下减”这句话来记忆，即

定理

对于 $f(x)$ 的图像有如下性质：

$$\begin{aligned}
f(x)\text{ 向左平移 } a(a>0)\text{ 个单位} &\iff f(x+a)\\
f(x)\text{ 向右平移 } b(b>0)\text{ 个单位} &\iff f(x-b)\\
f(x)\text{ 向上平移 } a(a>0)\text{ 个单位} &\iff f(x)+a\\
f(x)\text{ 向下平移 } b(b>0)\text{ 个单位} &\iff f(x)-b
\end{aligned}$$

♠

上面，我限定了定向平移，也就是 $f(x)$ 向左平移 $a(a>0)$，图像 $f(x)$ 只会向左移动，这是为了更直观的理解字面意思，需要注意尽管 $a<0$ 时，性质也成立，只不过描述的不符合常用语境，

e.g. $f(x)$ 向左平移-4 个单位 $\iff f(x-4) \iff f(x)$ 向右平移 4 个单位。

函数的伸缩有如下性质：

定理

对于 $f(x)$：

$$\begin{aligned}
f(x)\text{ 的横坐标变为原来的 } a(a>1)\text{ 倍} &\iff f(\frac{x}{a})\\
f(x)\text{ 的横坐标变为原来的 } b(1>b>0)\text{ 倍} &\iff f(bx)\\
f(x)\text{ 的纵坐标变为原来的 } a(a>1)\text{ 倍} &\iff af(x)\\
f(x)\text{ 的纵坐标变为原来的 } b(1>b>0)\text{ 倍} &\iff \frac{1}{b}f(x)
\end{aligned}$$

♠

其中变为原来的 $a(a>1)$ 倍通常被描述成，伸长 a 倍；变为原来的 $b(1>b>0)$ 倍通常被描述成，缩短为原来的 b 倍。

§3.5　一些函数

§3.5.1　对勾以及飘带

两个比较常见的函数模型，形如

$$f(x)=ax+\frac{b}{x},\ ab>0$$

的函数叫做**对勾函数**，形如

$$f(x)=ax-\frac{b}{x},\ ab>0$$

的函数叫做**飘带函数**。最简单的对勾函数和飘带函数是当 $a=b=1$ 时，如下图

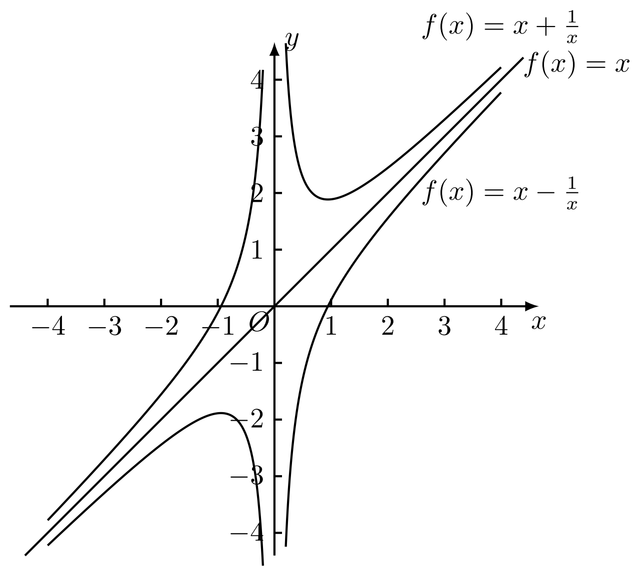


图 3.1: 对勾、飘带函数

以下讨论对勾函数与飘带函数的极值以及渐近线，设 $a, b > 0$（另一种 $a, b < 0$ 的情况类似）

由基本不等式可知，当 $a, b > 0$ 时对勾函数有极值 $f(\pm\sqrt{\frac{b}{a}}) = \pm 2\sqrt{ab}$。

渐近线：构造函数 $f(x) = ax + \frac{b}{x}$，$ab > 0, g(x) = ax$，两函数做差得到

$$f(x) - g(x) = \frac{b}{x}$$

又

$$\lim_{x\to\infty} \frac{b}{x} = 0$$

故 $g(x)$ 是 $f(x)$ 的渐近线，类似的方法可以证飘带函数的渐近线。

第四章　基本初等函数

§4.1　指数

定义

$$x^n = a$$

其中，x 称为 a 的 n 次方根，对上式开 n 次方根

$$\sqrt[n]{a} = x$$

叫根式，n 为根指数，a 为被开方数。

在初中我们以及初步了解了指数的基本运算，不过初中的指数运算仅停留在指数为非负整数的情况，而在高中阶段，我们对指数的取值范围进行扩充，即扩充定义，使其为整个有理数集，因此有如下定义：

定义

$$a^{\frac{m}{n}} = \sqrt[n]{a^m}\ m, n \in \mathbb{Z}, n \neq 0$$

特别的有 $\sqrt[n]{0} = 0$，当指数为负数时，有

$$a^{-n} = \frac{1}{a^n}$$

综上，我们就把指数的定义进行扩充到了有理数，对于

$$a^n = x,\quad n\text{为无理数}$$

其中 x 是一个确定的实数，因此指数的取值范围扩充到实数范围。

运算性质：若 $a > 0, b > 0, r, s \in \mathbb{Q}$ 则

$$a^r a^s = a^{r+s}$$
$$(a^r)^s = a^{rs}$$
$$(ab)^r = a^r b^r$$

由定义当 $r, s \in \mathbb{R}$ 时上述性质也成立。

§4.1.1　指数函数

一般的，形如

$$y = f(x) = a^x \quad (a > 0, a \neq 1, D_f = \mathbb{R})$$

称为**指数函数**，通过指数的运算性质我们可以对其进行分类研究。

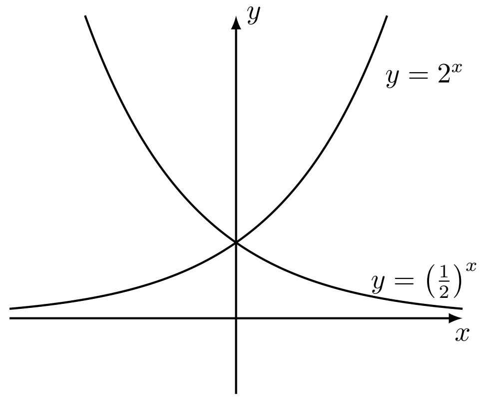


图 4.1: 指数函数

如图指数函数大致分两种情况：$a \in (0,1)$ 和 $a \in (1,+\infty)$。

1. 当 $a \in (0,1)$ 时，$f(x)$ 在 $\mathbb{R}$ 上单调递减，而且当 a 越接近 0 时，函数变化率越大，设有 $a_1, a_2 \in (0,1), a_1 > a_2$，那么有 $f_1(x) < f_2(x), x \in (-\infty, 0)$，和 $f_1(x) > f_2(x), x \in (0,+\infty)$
2. 当 $a \in (1,+\infty)$ 时，$f(x)$ 在 $\mathbb{R}$ 上单调递增，而且当 a 越接近 $+\infty$ 时，函数变化率越大，设有 $a_1, a_2 \in (1,+\infty), a_1 > a_2$，那么有 $f_1(x) < f_2(x), x \in (-\infty, 0)$，和 $f_1(x) > f_2(x), x \in (0,+\infty)$

可以看出无论 a 的取何值，其函数的图像都恒过 $(0,1)$。

§4.2 对数

定义一种运算：

定义

若 $a^x = N$ $(a > 0, a \neq 1)$，那么称 x 为以 a 为底 N 的对数，记作

$$\log_a N = x$$

其中 a 叫做**底数**，N 叫做**真数**。

特别的有：

$$\log_a 1 = 0$$
$$\log_a a = 1$$

根据指数的性质，立马可以得出的。

此外我们特别记

$$\log_{10} x = \lg x$$
$$\log_e x = \ln x$$

把以 10 为底的对数称为**常用对数**，记为 $\lg x$；

把以 e 为底的对数称为**自然对数**，记为 $\ln x$。

根据指数的运算性质，立即可得

1. $\log_a(MN) = \log_a M + \log_a N$
2. $\log_a \frac{M}{N} = \log_a M - \log_a N$
3. $\log_a M^N = N \log_a M$
4. 恒等式：$a^{\log_a N} = N$
5. 换底公式：$\log_a b = \frac{\log_c b}{\log_c a}$

证明不难，读者自证。

§4.2.1　对数函数

形如

$$y = f(x) = \log_a x,\ a > 0$$

的函数称为**对数函数**，其定义域为 $D_f = (0, +\infty)$，并且无论 a 取何值时，函数图像都恒过定点 $(1, 0)$。

与指数函数类似，我们对其进行分类讨论。

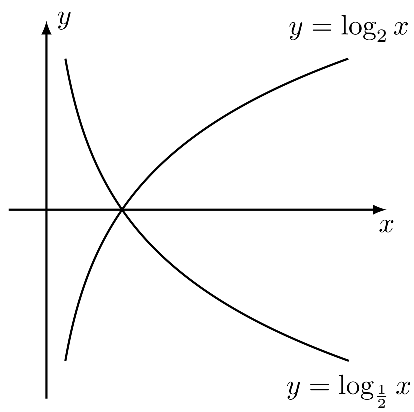


图 4.2: 对数函数

如图对数函数也分两种情况：$a \in (0, 1)$ 和 $a \in (1, +\infty)$。

1. 当 $a \in (0, 1)$ 时，$f(x)$ 在 $(0, +\infty)$ 上单调递减，而且当 a 越接近 0 时，函数变化率越小，设有 $a_1, a_2 \in (0, 1), a_1 > a_2$，那么有 $f_1(x) > f_2(x), x \in (0, 1)$，和 $f_1(x) < f_2(x), x \in (1, +\infty)$。
2. 当 $a \in (1, +\infty)$ 时，$f(x)$ 在 $(0, +\infty)$ 上单调递增，而且当 a 越接近 $+\infty$ 时，函数变化率越小，设有 $a_1, a_2 \in (1, +\infty), a_1 > a_2$，那么有 $f_1(x) > f_2(x), x \in (0, 1)$，和 $f_1(x) < f_2(x), x \in (1, +\infty)$。

§4.3　幂函数（高中内容）

$f(x) = x^\alpha$ (α 是常数)，可知 $f(x)$ 恒过 $(1, 1)$

§4.4 三角函数

§4.4.1 角的一些定义

任意角的概念

一个射线绕端点旋转到另一个位置的图形，由始边、终边以及一个角组成；

规定按逆时针旋转的为正角，顺时针旋转为负角，没旋转就是零角；角度 $\alpha - \beta$ 的计算可以从两个角度理解，一个是 α 终边顺时针转动 β，另一个是 α 终边逆时针转动 $-\beta$，它们是等价的，即：

$$\alpha - \beta = \alpha + (-\beta)$$

与 α 终边相同的角组成集合 S 可表示为：

$$S = \{\beta | \beta = \alpha + k \cdot 2\pi, k \in \mathbb{Z}\}$$

其中角度的单位为弧度制，这一点在下一节介绍。

§4.4.2 弧度制

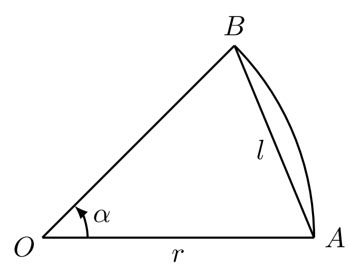


图 4.3: 弧度制

弧度制

定义角度大小的一种表示方式，如图，弧 $AB = l$，$OA = r$，记

$$\alpha = \frac{l}{r}$$

由上式可知 α 的大小，其单位为 rad，由于我们更熟悉角度制，下面展示一些基本的转换：注意到圆的周长 $C = 2\pi r$，因此：

$$\begin{aligned} 360^\circ &= 2\pi \text{ rad} \\ 180^\circ &= \pi \text{ rad} \\ 45^\circ &= \frac{\pi}{4}\pi \text{ rad} \end{aligned}$$

对于弧度制看可以类比 $\tan\alpha = \frac{\sin\alpha}{\cos\alpha}$ 的关系来记忆，即对弧邻边。

由圆的知识可以推出扇形的面积公式。

定理

圆心角为 α 半径为 r 的扇形，记其面积为 S，有：

$$S = \frac{\alpha r^2}{2} = \frac{lr}{2}$$

其中圆心角的单位为弧度制。

证明：记 $\alpha\text{rad} = n^\circ$，易知：

$$S = \pi r^2 \frac{n^\circ}{360^\circ}$$

由于 $\frac{n^\circ}{360^\circ} = \frac{\alpha}{2\pi}$，即证

$$\begin{aligned} S &= \frac{\alpha r^2}{2} \\ &= \frac{lr}{2} \end{aligned}$$

□

§4.4.3 三角函数的定义

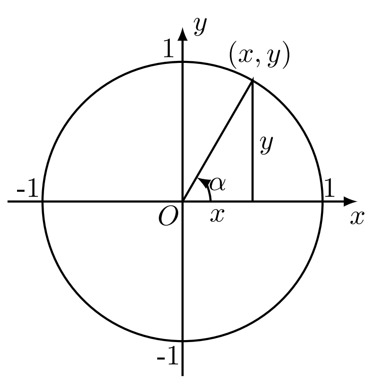


图 4.4: 单位圆

如图，该圆为半径是 1 的圆（称为**单位圆**），其圆上任意一点都可以用 (x, y) 表示，易知

$$x^2 + y^2 = 1$$

若以 x 轴正半轴为始边，由初中知识可知：第一象限中，单位圆上的点可以用 $(\cos\alpha, \sin\alpha)$ 表示，按照这个思路将三角函数的取值范围推广到任意角。

定义

在单位圆上的一点 $P(x, y)$，以 x 轴正半轴为使边，以 OP 为终边的夹角为 α，定义：

$$\cos\alpha = x, \quad \sin\alpha = y, \quad \tan\alpha = \frac{y}{x}$$

分别称为**余弦**、**正弦**、**正切**，此外还有一些高中不常用的表示

$$\cot\alpha = \frac{x}{y},\quad \csc\alpha = \frac{1}{y},\quad \sec\alpha = \frac{1}{x}$$

分别称为**余切**、**余割**、**正割**。

由于 α 绕圆转一圈后点 P 对应的点 P' 的夹角依然为 α，可知

$$\sin(\alpha + 2k\pi) = \sin\alpha\,(k \in \mathbb{Z})$$

同理 $\cos\alpha, \tan\alpha$ 一样。

单位圆上任意一点 P 到圆心 O 的距离为 $r = 1$，因此有 $x^2 + y^2 = 1$，可以推出三个重要的恒等式：

$$\sin^2\alpha + \cos^2\alpha = 1$$
$$\sec^2\alpha - \tan^2\alpha = 1$$
$$\csc^2\alpha - \cot^2\alpha = 1$$

§4.4.4　诱导公式

诱导公式是直观的，它表现了三角函数的几何意义，以及正余弦的互相转换，在常规意义下的诱导公式有以下 3 个内容：

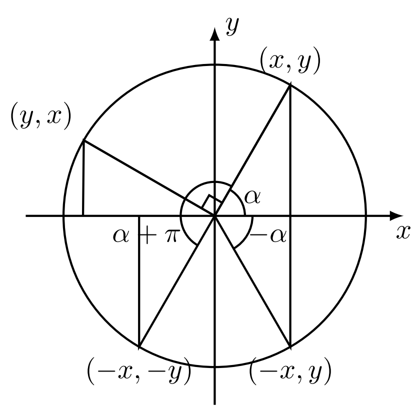


图 4.5: 诱导公式

定理

对于任意 α 都有：

$$\sin(\pi \pm \alpha) = -\sin\alpha$$
$$\sin -\alpha = -\sin\alpha$$
$$\sin(\pi - \alpha) = \sin\alpha$$
$$\sin(\frac{\pi}{2} + \alpha) = \cos\alpha$$

以上是那 $\sin\alpha$ 举例，按照图像理解可以推导出 $\cos\alpha, \tan\alpha$ 的诱导公式。

通过图像，诱导公式是显然的，对于任意角 θ 作的简单变换，都可以把 θ 看作锐角 α 作的简单变换，通过观察锐角 α 变换是比较方便的，就如图示，下面阐述下这三个内容的几何意义。。

1. π 在弧度制中代表着 $180°$，所以三角函数中加或减 π 在几何意义上来说是将角加一个平角，也就是转一个 $180°$。
2. $-\alpha$ 意为着它向顺时针旋转 α，可以看作与 α 关于 x 轴对称。
3. $\frac{\pi}{2}$ 就是 $90°$，这个稍微难理解一点，在转 $90°$ 以后往往会 x, y 互相调换，可以由图理解。

§4.5 函数部分

§4.5.1 正弦函数

定义最基本的正弦函数为

$$f(x) = \sin x$$

它的图像如下图：

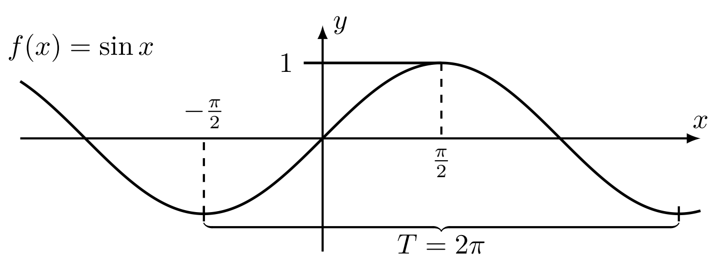


图 4.6: 正弦函数

对于函数 $f(x)$ 有 $D_f = \mathbb{R}, R_f = [-1, 1]$，并且由诱导公式可知这个图像的一些基本性质，例如由

$$\sin\alpha = \sin(2\pi + \alpha)$$

可知 2π 为 $f(x)$ 的一个周期，且不难证 2π 是它的最小正周期；再由

$$\sin -x = -\sin x$$

可知 $f(x)$ 为奇函数。

通过求导或图像或单位圆可知 $f(x)$ 在 $\left(-\frac{\pi}{2} + 2k\pi, \frac{\pi}{2} + 2k\pi\right), k \in \mathbb{Z}$ 上单调递增,在 $\left(\frac{\pi}{2} + 2k\pi, \frac{3\pi}{2} + 2k\pi\right), k \in \mathbb{Z}$ 上单调递减，在 $x = \frac{\pi}{2} + k\pi, k \in \mathbb{Z}$ 取最值，以及对称轴和对称中心。

正弦函数更一般形式是

$$y = A\sin(\omega x + \varphi) + B$$

由函数伸缩平移的知识可知这四个参数的意义

$$Af(\omega x + \varphi) + B = A\sin(\omega x + \varphi) + B$$

其中 A 称为**振幅**，B 决定上下移动量，上加下减，A 与 B 一起决定 $f_{\max}(x), f_{\min}(x)$，也即 $f(x)$ 的值域。

不妨设 $\alpha = \omega x + \varphi$，将 $\omega x + \varphi$ 看作一个整体；这样一来只需研究 $\sin\alpha$ 的性质，其中

$$f(x) = \sin\alpha$$

是最基本的正弦函数，由 $\sin x$ 到 $\sin\alpha$ 的图像，只是一个线性变换。

其中 ω 决定最小正周期 T，因为

$$\sin\omega x = \sin\omega x + 2\pi = \sin\omega\left(x + \frac{2\pi}{\omega}\right)$$

所以 $T = \frac{2\pi}{\omega}$

φ 称为**初相**，决定左右移动量，左加右减，即当 $\varphi > 0$，图像向左，$\varphi < 0$ 向右移。

上文中已经研究了 $y = \sin\alpha$ 的一些基本性质，需要注意的是 $f(x) = \sin\alpha$ 相当于一个复合函数，为了确保增减性一致，需要确保 $\omega > 0$，这样一来，对于 $\sin\alpha$ 的最值位置、单调区间、对称中心以及对称轴位置都可以用 x 表示。

e.g. 求 $\sin(2x+\pi)$ 的单调递增区间，首先换元令 $\alpha = 2x + \pi$，易知对于 $\sin\alpha$ 的单调递增区间为

$$A = \left(-\frac{\pi}{2} + 2k\pi, \frac{\pi}{2} + 2k\pi\right), k \in \mathbb{Z}$$

也即 $2x + \pi \in A$ 上单调递增，故知

$$x \in \left(-\frac{3\pi}{4} + k\pi, -\frac{\pi}{4} + k\pi\right), k \in \mathbb{Z}$$

§4.5.2 余弦函数

最基本的余弦函数为 $f(x) = \cos x$，对于余弦函数我们同样可以采用与正弦函数类似的方法去研究，例如由 $\cos -x = \cos x$ 可知 $f(x)$ 是一个偶函数；

不过我会根据

$$\sin(x + \frac{\pi}{2}) = \cos x$$

将余弦函数转换为正弦函数去研究。

§4.5.3 正切函数

$f(x) = \tan x$ 为最基本的正切函数，可知它的定义域以及值域为

$$D_f = \mathbb{R}\backslash\left\{\beta|\beta = \frac{\pi}{2} + k\pi, (k \in \mathbb{R})\right\}, \quad R_f = \mathbb{R}$$

如图

可知最小正周期为 π，在

$$A = \left(-\frac{\pi}{2} + k\pi, \frac{\pi}{2} + k\pi\right), k \in \mathbb{Z}$$

上单调递增，其中

$$\lim_{x\to\frac{\pi}{2}^-} = +\infty, \quad \lim_{x\to\frac{\pi}{2}^+} = -\infty$$

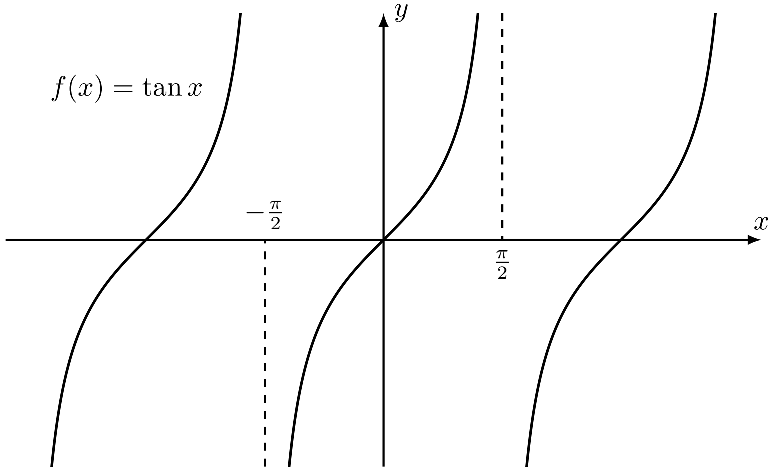


图 4.7: 正切函数

正切函数也有一般形式：

$$f(x) = A\tan(\omega x + \varphi) + B$$

研究它与前文中研究正弦函数的方法类似在此不过多赘述。

§4.6 恒等变换

定理

对于任意 α, β 有

$$\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$$
$$\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \sin\beta\cos\alpha$$
$$\tan(\alpha \pm \beta) = \frac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta}$$

上式被称为**和差角公式**。 ♠

证明：

如图，A, B 为单位圆上的两点，P 为单位圆与 x 轴正半轴的交点，令 $\angle AOP = \alpha, \angle BOP = \beta$，易知：

$$A(\cos\alpha, \sin\alpha), \quad B(\cos\beta, \sin\beta)$$

令 A 绕原点 O 顺时针绕 β，可以得到 C，易知

$$C(\cos(\alpha - \beta), \sin(\alpha - \beta))$$

且

$$\angle COP = \alpha - \beta = \angle AOB$$

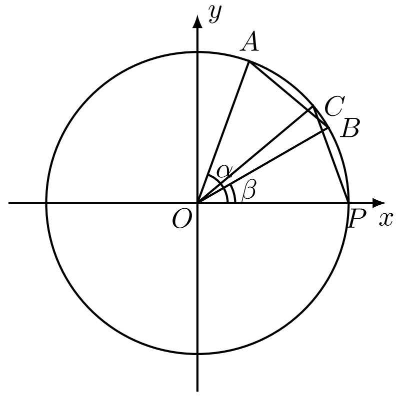


图 4.8: 和差角公式推导

又 $|OA| = |OB| = |OC| = |OP| = 1$，即知：

$$\triangle AOB \cong \triangle COP$$

故 $|AB| = |CP|$，由两点间距离公式得：

$$\begin{aligned}|CP|^2 &= |AB|^2\\(1-\cos(\alpha-\beta))^2+\sin^2(\alpha-\beta) &= (\cos\alpha-\cos\beta)^2+(\sin\alpha-\sin\beta)^2\\-2\cos(\alpha-\beta)+2 &= 2-2\cos\alpha\cos\beta-2\sin\alpha\sin\beta\\\cos(\alpha-\beta) &= \cos\alpha\cos\beta+\sin\alpha\sin\beta\end{aligned}$$

若令 $\beta = -\beta$，即得

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta-\sin\alpha\sin\beta$$

余弦的和差角公式证得。

□

由诱导公式可得

$$\begin{aligned}\cos\left(\alpha-\beta+\frac{\pi}{2}\right) &= \cos\left(\alpha+\frac{\pi}{2}\right)\cos\beta+\sin\left(\alpha+\frac{\pi}{2}\right)\sin\beta\\-\sin(\alpha-\beta) &= -\sin\alpha\cos\beta+\sin\beta\cos\alpha\\\sin(\alpha-\beta) &= \sin\alpha\cos\beta-\sin\beta\cos\alpha\end{aligned}$$

同样的，令 $\beta = -\beta$ 可得：

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta+\sin\beta\cos\alpha$$

即证正弦的和差角公式。

□

最后由

$$\tan(\alpha-\beta) = \frac{\sin(\alpha-\beta)}{\cos(\alpha-\beta)}$$

可得

$$\begin{aligned}\tan(\alpha-\beta) &= \frac{\sin\alpha\cos\beta-\sin\beta\cos\alpha}{\cos\alpha\cos\beta+\sin\alpha\sin\beta} \\ &= \frac{\tan\alpha-\tan\beta}{1+\tan alpha\tan\beta}\end{aligned}$$

这里分子分母上下同时除以了一个 $\cos\alpha\cos\beta$，替换 $\beta=-\beta$ 可得：

$$\tan(\alpha+\beta)=\frac{\tan\alpha+\tan\beta}{1-\tan\alpha\tan\beta}$$

□

推论

对于任意 α 有

$$\begin{aligned}\sin(2\alpha) &= 2\sin\alpha\cos\alpha \\ \cos(2\alpha) &= \cos^2\alpha-\sin^2\alpha \\ \tan(2\alpha) &= \frac{2\tan\alpha}{1-\tan^2\alpha}\end{aligned}$$

上式被称为**二倍角公式**。 ◇

由和角公式，若令 $\alpha=\beta$ 即可得二倍角公式，其中余弦的二倍角有以下几种变形

$$\begin{aligned}\cos(2\alpha) &= 2\cos^2\alpha-1 \\ \cos(2\alpha) &= 1-2\sin^2\alpha\end{aligned}$$

简单移项即可得

$$\frac{\cos(2\alpha)+1}{2}=\cos^2\alpha$$

有人将上式称为**降幂公式**，但我觉得实际作用很小，因为是二倍角公式的副产物，只需记住二倍角公式即可。

推论

对于任意 α 有

$$\begin{aligned}\sin\frac{\alpha}{2} &= \pm\sqrt{\frac{1-\cos\alpha}{2}} \\ \cos\frac{\alpha}{2} &= \pm\sqrt{\frac{1+\cos\alpha}{2}} \\ \tan\frac{\alpha}{2} &= \pm\sqrt{\frac{1-\cos\alpha}{1+\cos\alpha}}=\frac{\sin\alpha}{1+\cos\alpha}=\csc\alpha-\cot\alpha\end{aligned}$$

上式被称为**半角公式**。 ◇

由余弦的二倍角公式即可推出，在此不做证明。

§4.6.1 积化和差，和差化积

积化和差：

$$
\begin{aligned}
\sin\alpha\cos\beta &= \frac{1}{2}\left[\sin\left(\alpha+\beta\right)+\sin\left(\alpha-\beta\right)\right]\\
\cos\alpha\sin\beta &= \frac{1}{2}\left[\sin\left(\alpha+\beta\right)-\sin\left(\alpha-\beta\right)\right]\\
\cos\alpha\cos\beta &= \frac{1}{2}\left[\cos\left(\alpha+\beta\right)+\cos\left(\alpha-\beta\right)\right]\\
\sin\alpha\sin\beta &= -\frac{1}{2}\left[\cos\left(\alpha+\beta\right)-\cos\left(\alpha-\beta\right)\right]
\end{aligned}
$$

和差化积：

$$
\begin{aligned}
\sin\alpha+\sin\beta &= 2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}\\
\sin\alpha-\sin\beta &= 2\cos\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}\\
\cos\alpha+\cos\beta &= 2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}\\
\cos\alpha-\cos\beta &= -2\sin\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}
\end{aligned}
$$

通过和差角公式即可推出，证明留给读者。

辅助角公式

定理

辅助角公式：

$$a\sin\alpha+b\cos\alpha=\sqrt{a^2+b^2}\sin\left(\alpha+\varphi\right)\left(ab\neq 0\right)$$

其中 φ 就是辅助角满足

$$\tan\varphi=\frac{b}{a},\cos\varphi=\frac{a}{\sqrt{a^2+b^2}},\sin\varphi=\frac{b}{\sqrt{a^2+b^2}}$$

证明：

$$
\begin{aligned}
&= a\sin\alpha+b\cos\alpha\\
&= \sqrt{a^2+b^2}\left(\frac{a}{\sqrt{a^2+b^2}}\sin\alpha+\frac{b}{\sqrt{a^2+b^2}}\cos\alpha\right)
\end{aligned}
$$

令

$$\cos\varphi=\frac{a}{\sqrt{a^2+b^2}},\sin\varphi=\frac{b}{\sqrt{a^2+b^2}}$$

可得

$$
\begin{aligned}
&= \sqrt{a^2+b^2}\left(\cos\varphi\sin\alpha+\sin\varphi\cos\alpha\right)\\
&= \sqrt{a^2+b^2}\sin\left(\alpha+\varphi\right)
\end{aligned}
$$

□

在实际运用中，我们不需要知道 φ 的准确值，更多时候辅助角公式是帮我们把正弦和余弦连接起来，转化为更容易研究的基本三角函数。

§4.7 反三角函数

这里简单介绍一下反三角函数的内容。

定义

正弦函数 $y=\sin x$ 的反函数为

$$y=\arcsin x$$

称为**反正弦函数**。

余弦函数 $y=\cos x$ 的反函数为

$$y=\arccos x$$

称为**反余弦函数**。

正切函数 $y=\tan x$ 的反函数为

$$y=\arctan x$$

称为**反正切函数**。

♡

由正弦函数 $y=\sin x$ 是一个周期函数，因此对于它的反函数 $y=\arcsin x$，只取 $R_f\in[-\frac{\pi}{2},\frac{\pi}{2}]$，定义域为 $D_f\in[-1,1]$。

同理，余弦函数的反函数 $y=\arccos x$，其定义域为 $D_f\in[-1,1]$，值域为 $[0,\pi]$。

因正切函数的值域为 $\mathbb{R}$，因此反正切函数的定义域为 $D_f\in\mathbb{R}$，值域为 $R_f\in[-\frac{\pi}{2},\frac{\pi}{2}]$。

第五章　不等式

§5.1　不等式基本性质

用不等号 $(<,>,\leqslant,\geqslant,\neq)$ 表示不等关系的式子叫不等式。

基本事实：对于 $a,b\in\mathbb{R}$ 有

$$
\begin{aligned}
a-b>0 &\iff a>b\\
a-b=0 &\iff a=b\\
a-b<0 &\iff a<b
\end{aligned}
$$

不等式基本性质

1. $a>b \iff b<a$ (对称性)
2. $a>b,b>c \iff a>c$ (传递性)
3. 如果 $a>b$，则 $a+c>b+c$
4. 如果 $a>b,c>0$，则 $ac>bc$
   如果 $a>b,c<0$，则 $ac<bc$
5. 如果 $a>b,c>d$，则 $a+c>b+d$
6. 如果 $a>b>0,c>d>0$，则 $ac>bd$
7. 如果 $a>b>0$，那么 $a^n>b^n\,(n\in\mathbb{N},n\geqslant 1)$
8. 如果 $a>b>0$，那么 $\sqrt[n]{a}>\sqrt[n]{b}\,(n\in\mathbb{N},n\geqslant 2)$
9. $\frac{a}{b}>0 \iff ab>0,\frac{a}{b}<0 \iff ab<0$

这些性质是显然的。

其中不等式的传递性是放缩技巧的本质，第 9 条在高次不等式中常用到的性质。

同解性

同解不等式：如果两个不等式的解集相同，那么这两个不等式叫做同解不等式。

1. 对于任何一个整式 $F(x)$，不等式 $f(x)>g(x)$ 与 $f(x)+F(x)>g(x)+F(x)$ 同解；
2. 若 $m>0$，不等式 $f(x)>g(x)$ 与 $mf(x)>mg(x)$ 同解；
3. 若 $m<0$，不等式 $f(x)>g(x)$ 与 $mf(x)<mg(x)$ 同解；

§5.2 高中不等式的一些处理方法

§5.2.1 一元二次不等式

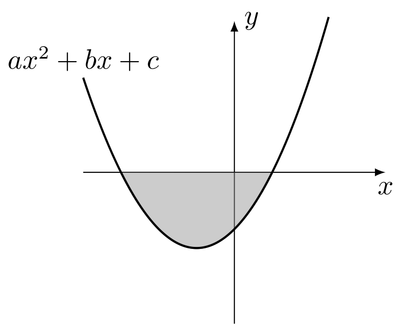


图 5.1: 一元二次不等式

一般的，一元二次不等式

$$ax^2 + bx + c < 0\text{（不等号可以取 }\geqslant, \leqslant, >, <\text{）}$$

画出 ax^2+bx+c 的函数图像可知,其实就是求函数图像在 x 轴以下的部分的 x 取值范围。同理可得 $ax^2+bx+c > 0$ 就是求函数图像在 x 轴以上的部分的 x 取值范围。

由以上分析，我们可以得到一般解法：

1. 求 $\sqrt{\Delta} = \sqrt{b^2 - 4ac}$，这一步是为了判断函数图像与 x 轴是否有交点
2. 通过 a 判断开口方向,$a > 0$ 开口向上，$a < 0$ 开口向下
3. 计算 x 取值范围

注意到判别式 Δ 的取值，当 $\Delta > 0$ 时，这个不等式可因式分解为

$$C(x - a_1)(x - a_2) < 0,\ a_1 \neq a_2, C \neq 0$$

这个 a_1, a_2 就是函数零点，不妨设 $a_1 < a_2$，那么当 $C > 0$ 时，立即得解集 (a_1, a_2)，当 $C < 0$ 时，得解集 $\mathbb{R}\backslash[a_1, a_2]$。

当 $\Delta = 0$ 时，该方程能写为

$$C(x - a)^2 < 0\ , C \neq 0$$

显然，当 $C > 0$ 时，不等式无解，而当 $C < 0$ 时，解集为 $\mathbb{R}\backslash\{a\}$。

当 $\Delta < 0$ 时，方程无法因式分解（在实数域内），由算术基本定理可知，该方程一定可以有解，不过在实数域内无解，这意味着，该方程无零点，也即该方程的函数图像要么全在 x 轴以下，要么全在 x 轴以上，对应的，解集要么是整个实数集，要么无解。

§5.2.2 高次不等式

通过研究一元二次不等式，不难发现讨论时数形结合会有奇效，依照这个思路，我们讨论可因式分解的一元多次不等式。

对于高次不等式，如果这个高次不等式能够因式分解，那么常用到“穿针引线法”来解决，先说这个方法的口诀：自上而下、从右往左、奇穿偶不穿。

我比较排斥例如“口诀、大招、秒杀”一类的字眼，因为它自带特殊性，把一般问题特殊化，深入的推导表面化，尽管是应试学习，学数学也需要究其本质。

实际上“穿针引线法”是对于高次不等式，在代数上与几何上数形结合推导出来的产物，我们先来看看一个简单的式子：

对于不等式：

$$x(x-1)(x-2)>0$$

不难看出，如果我们带入当自变量 $x=0,1,2$，则 $y=x(x-1)(x-2)$ 均为 0，也就是说 $f(x)=y$ 有三个零点，而且不论 x 取何值，均有：

$$x>(x-1)>(x-2)$$

不妨设 $a=x, b=x-1, c=x-2$，可知：

i. $2<x$ 时，$a,b,c>0$

ii. $1<x<2$ 时，$a,b>0, c<0$

iii. $0<x<1$ 时，$a>0, b,c<0$

iv. $x<0$ 时，$a,b,c<0$

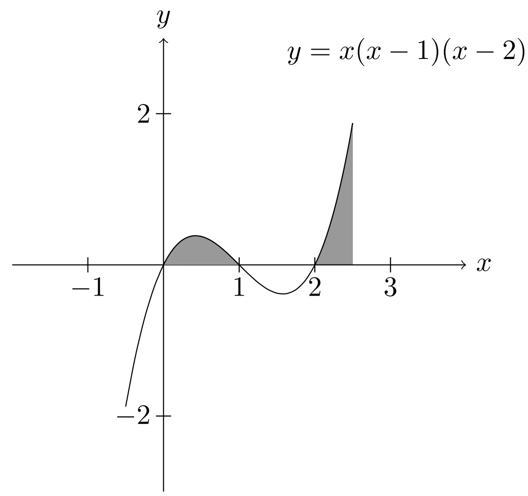


图 5.2: 穿针引线

又因 $y=x(x-1)(x-2)=abc$，由上面得出的四条推论推出：

i. $2<x$ 时，　$a\cdot b\cdot c>0 \iff y>0$

ii. $1<x<2$ 时，$a\cdot b\cdot c<0 \iff y<0$

iii. $0<x<1$ 时，$a\cdot b\cdot c>0 \iff y>0$

iv. $x<0$ 时，　$a\cdot b\cdot c<0 \iff y<0$

结合以上结论我们可以得到 $y>0$ 时 x 的取值范围，也能因此画出图像，阴影部分为取值范围。

我们也可以直接奔结论，将推论进行筛选，这能节省不少时间，如：

$$\because y>0, a>b>c$$
$$\therefore a>0 \quad b,c<0 \quad \text{or} \quad a,b,c>0$$

上式的意思是：若 $y=abc$ 要为正数，且 $a>b>c$，那么只有两种情况满足，也就是 a,b,c 均为正数时，或 a 为正数，b,c 为负数，因此直接求这两个情况的取值范围即可。

以上的例子是由代数的思想推导出的，但我们只要有第一步的推论，其实就可以画出函数图像的大概样子了，结合图像判断得更快。

而穿针引线法其实就是教你怎么画这个函数图像，我来解释一下口诀的具体做法：“从上至下，从右到左”这句意为着作图的方向，“奇穿偶不穿”其中“奇偶”意为因式分解后每个项的指数奇偶，而“奇穿偶不穿”意为当指数为奇数时穿过这个项所代表的零点，而指数为偶数时不穿过。

看文字可能比较抽象，可以尝试用这个方法画上面的例子，其中 a,b,c 都是一次项，即它们的指数都是奇数，因

此都穿过。

例如不等式：

$$\frac{5}{2}x(x-1)^2(x-2)^3<0$$

此时 b 的指数为偶数，b 对应的零点不穿过，a,c 的指数为奇数，因此它们对应的零点穿过，结合图像即可画出，如右图所示。

奇穿偶不穿的原理是什么？不妨以代数的角度来思考，当一个项的指数为偶数时，那么它必定为非负的数，不妨把它看作一个正的常数，只需要注意当 x 取到这个项的零点时因变量 y 也为 0，其他时候尽管把它看作正的常数。

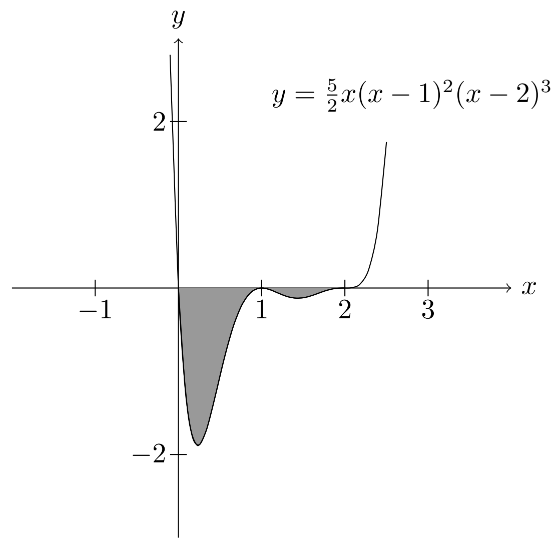


图 5.3: 穿针引线 2

下面我举一个比较复杂的例子，也是比较综合的问题：解不等式

$$\frac{(x^2+1)(x^2-5x+6)(x+4)^3(\frac{2}{\mathrm{e}}x-2)^4}{(x-10)(x^2+2x+1)}\mathrm{e}^x>0$$

求 x 的取值范围。

这道题如果用我们用第一个例子的方法去做，会比较复杂，但用穿针引线法画图，以几何的视角来解，会简单一点点，不易漏情况，读者可当作课后习题。

下面我来讲解一下思路：

第一步用不等式的性质将分式转换为等价形式，以及因式分解：

$$\begin{aligned}
&\frac{(x^2+1)(x^2-5x+6)(x+4)^3(\frac{2}{\mathrm{e}}x-2)^4}{(x-10)(x^2+2x+1)}\mathrm{e}^x>0\\
\iff\ &\left(x^2+1\right)\left(x^2-5x+6\right)(x+4)^3\left(\frac{2}{\mathrm{e}}x-2\right)^4(x-10)\left(x^2+2x+1\right)\mathrm{e}^x>0\\
&\left(x^2+1\right)(x-2)(x-3)(x+4)^3\left(\frac{2}{\mathrm{e}}x-2\right)^4(x-10)(x+1)^2\mathrm{e}^x>0
\end{aligned}$$

第二步将不会影响结果的项排除掉，该式中，(x^2+1) 与 e^x 项恒大于 0，因此不影响结果，再来我们把这一串式子的零点找出来，从小到大依次为：$-4,-1,2,\mathrm{e},3,10$。

第三步，画函数图像判断取值范围；这一步读者可自己画一下，最后得到的解集为：

$$x\in(-\infty,-4)\cup(2,\mathrm{e})\cup(\mathrm{e},3)\cup(10,+\infty)$$

最后我对这个方法的本质进行总结：从代数观点来看，这个方法利用了不等式的基本性质，可以用分类讨

论来说明，如例一中，由基本事实

$$a > b > c$$

通过分类讨论 x 的取值，即可求出不等式的解集。

从几何观点来看，这个方法抓住因式分解可控制零点的特点，巧妙且直观的体现出不等式的解集。

§5.2.3 三角换元

三角换元是很好用也非常灵活的一个方法，它的本质就是三角函数的恒等式：

$$\sin^2\theta + \cos^2\theta = 1$$

将 $\sin\theta, \cos\theta$ 与未知数进行换元，例如已知 $x^2 + y^2 = 1$，求 $x + 2y$ 的取值范围。令:

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

则原式可化为:

$$x + 2y = \sin\theta + 2\cos\theta$$

此时通过辅助角公式可知：

$$x + 2y = \sqrt{5}\sin(\theta + \varphi)$$

其中 $\sin(\theta+\varphi) \in [-1, 1]$，$\varphi$ 满足 $\cos\varphi = \frac{1}{\sqrt{5}}, \sin\varphi = \frac{2}{\sqrt{5}}$，如果这里不知道辅助角怎么推导到这一步的，可以查看一下辅助角公式那一节。

最后可得 $x + 2y \in [-\sqrt{5}, \sqrt{5}]$。

这是显而易见可以用到三角不等式的例子，下面我举的例子都是不容易看出来的，因篇幅有限，有些例题只将思路。

例一：已知 $x^2 - 2xy + 2y^2 = 2$，求 $x^2 + 2y^2$ 取值范围。
这题简单配凑即可换元，条件可转换为 $(x-y)^2 + y^2 = \sqrt{2}^2$，令：

$$\begin{cases} x - y & = \sqrt{2}\sin\theta \\ \quad y & = \sqrt{2}\cos\theta \end{cases}$$

带回原式

$$\begin{aligned} x^2 + 2y^2 &= \left(\sqrt{2}(\sin\theta + \cos\theta)\right)^2 + 2\left(\sqrt{2}\cos\theta\right)^2 \\ &= 2(2 + \sin 2\theta + \cos 2\theta) \\ &= 4 + 2\sqrt{2}\sin(2\theta + \varphi) \end{aligned}$$

最后得 $x^2 + 2y^2 \in [4 - 2\sqrt{2}, 4 + 2\sqrt{2}]$。

例二：已知 $x^2 + 2xy = 1$，求 $x^2 + y^2$ 的取值范围。
这题从已知条件不好求，所以从 $x^2 + y^2$ 开始换元：

不妨设 $x^2+y^2=r^2$，则有：

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

之后将设的 x,y 带到条件中即可解出 r 的取值范围。

例三：已知 $f(x)=\sqrt{3-x}+2\sqrt{x+1}$，求 $f(x)$ 的值域。
这里注意到两个根号式子加起来，可以将 x 给消去，即：

$$\begin{aligned} &(\sqrt{3-x})^2+(\sqrt{x+1})^2 \\ &=3-x+x+1 \\ &=4 \end{aligned}$$

此时令

$$\begin{cases} m=\sqrt{3-x}=2\sin\theta \\ n=\sqrt{x+1}=2\cos\theta \end{cases}$$

带回函数 $f(x)$ 中即可，不过需要注意的是 θ 的取值范围。

§5.3 一些常用的不等式

§5.3.1 基本不等式

由 $(a-b)^2\geqslant 0$ 推出：

$$a^2+b^2\geqslant 2ab$$

当且仅当 $a=b$ 时等号成立

这时如果令 $a=\sqrt{x},b=\sqrt{y}$ 则可得到基本不等式：

基本不等式

对于 $\forall x,y\in\mathbb{R}^+$ 有：

$$x+y\geqslant 2\sqrt{xy}\,(x,y\geqslant 0)$$

这是平均值不等式的一种，在“平均值不等式”会提到。

基本不等式常见变形：

$$\frac{b}{a}+\frac{a}{b}\geqslant 2,(ab>0)$$

两个不定值的最值，通常是通过变形为一方，使另一方为定值。

当基本不等式一侧的和为定值，或另一侧的积为定值时可以用基本不等式来求最值，但很多情况需要灵活的使用定值部分，如下列例题就用到“1 的代换”。

已知 $a+6b=2(a,b>0)$，求 $\frac{4}{a}+\frac{6}{b}$ 的最小值。

$$\begin{aligned}\frac{4}{a}+\frac{6}{b}&=(\frac{4}{a}+\frac{6}{b})\cdot\frac{1}{2}\cdot 2\\&=\frac{1}{2}\cdot(\frac{4}{a}+\frac{6}{b})\cdot(a+6b)\\&=20+\frac{12b}{a}+\frac{3a}{b}\geqslant 2\sqrt{\frac{12b}{a}\cdot\frac{3a}{b}}+20=32\end{aligned}$$

§5.3.2 平均值不等式

设有 n 个正实数 $a_1,a_2,\ldots,a_n$，记：

$$\begin{aligned}\text{算术平均数}A_n&=\frac{a_1+a_2+\cdots+a_n}{n}\\\text{几何平均数}G_n&=\sqrt[n]{a_1a_2\cdots a_n}\\\text{调和平均数}H_n&=\frac{n}{\frac{1}{a_1}+\frac{1}{a_2}+\cdots+\frac{1}{a_n}}\\\text{平方平均数}Q_n&=\sqrt{\frac{a_1^2+a_2^2+\cdots+a_n^2}{n}}\end{aligned}$$

A_n,G_n,H_n,Q_n 它们有如下关系：

$$H_n\leqslant G_n\leqslant A_n\leqslant Q_n$$

特别的，基本不等式就是 $G_2\leqslant A_2$ 的情况。

当 $n=2$ 时可以得到二元均值不等式即：对于任意正实数 a,b 有

$$\frac{2}{\frac{1}{a}+\frac{1}{b}}\leqslant\sqrt{ab}\leqslant\frac{a+b}{2}\leqslant\sqrt{\frac{a^2+b^2}{2}}$$

当且仅当 $a=b$ 时取得等号。

下面用几何方法证明这个不等式链。

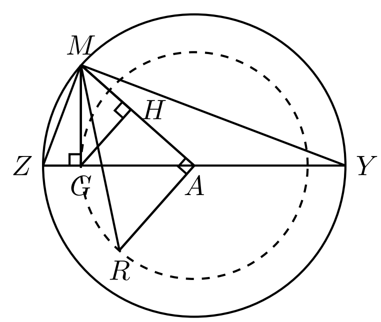


图 5.4: 二元均值不等式

如图在圆 C 中，圆心为 A、直径为 ZY，点 M 是其上任意一点，记过 M 点做 MG 垂直 ZY 垂足为 G，这时不妨设 $a=|ZG|,b=|YG|$，连接 MA，显然 $|MA|$ 为半径，即

$$|MA|=\frac{|ZG|+|YG|}{2}=\frac{a+b}{2}$$

根据射影定理（相似）可得

$$|MG| = \sqrt{|ZG| \cdot |YG|} = \sqrt{ab}$$

M 到直线 ZY 的最短距离为 $|MG|$，即

$$\begin{aligned} |MG| &\leqslant |MA| \\ \sqrt{ab} &\leqslant \frac{a+b}{2} \end{aligned}$$

过点 G 做 GH 垂直于 MA，垂足为 H，显然 $\triangle MHG \sim \triangle MGA$，因此可以求得 $|MH|$

$$\begin{aligned} |MH| &= \frac{|GM| \cdot |GM|}{|MA|} \\ &= \frac{2ab}{a+b} = \frac{2}{\frac{1}{a}+\frac{1}{b}} \end{aligned}$$

以 A 为圆心，$|GA|$ 为半径做圆 C'（虚线所示），过点 A 做 AR 平行于 HG，R 为圆 C' 上的点，连接 MR，根据勾股定理可得 $|MR|$

$$\begin{aligned} |MR| &= \sqrt{|RA|^2 + |MA|^2} \\ &= \sqrt{(b - \frac{a+b}{2})^2 + (\frac{a+b}{2})^2} \\ &= \sqrt{\frac{a^2+b^2}{2}} \end{aligned}$$

显然点 M 到直线 GH 的距离最短的为 $|MH|$ 因此

$$\begin{aligned} |MH| &\leqslant |MG| \\ \frac{2}{\frac{1}{a}+\frac{1}{b}} &\leqslant \sqrt{ab} \end{aligned}$$

同理点 M 到直线 AR 的距离最短为 $|MA|$ 因此

$$\begin{aligned} |MA| &\leqslant |MR| \\ \frac{a+b}{2} &\leqslant \sqrt{\frac{a^2+b^2}{2}} \end{aligned}$$

综上：

$$|MH| \leqslant |MG| \leqslant |MA| \leqslant |MR|$$

即证

$$\frac{2}{\frac{1}{a}+\frac{1}{b}} \leqslant \sqrt{ab} \leqslant \frac{a+b}{2} \leqslant \sqrt{\frac{a^2+b^2}{2}}$$

特别的，当 $a = b$ 时，也即点 M 为半圆弧 ZY 的中点时，点 G 与点 A 重合，上式取等号。

□

§5.3.3 柯西不等式

定理

柯西不等式：

$$\left(\sum_{i=1}^{n} x_i^2\right)\cdot\left(\sum_{i=1}^{n} y_i^2\right) \geqslant \left(\sum_{i=1}^{n} x_i y_i\right)^2$$

当且仅当 $\frac{x_i}{y_i}$ 为定值时等号成立。

♠

下面进行证明：

有两组数据 $x_i, y_i(i=1,2,\ldots,n)$，可以将他们看作 n 维向量 $\overrightarrow{x}, \overrightarrow{y}$ 的坐标，即:

$$\begin{aligned}\overrightarrow{x} &= (x_1, x_2, \ldots, x_n)\\ \overrightarrow{y} &= (y_1, y_2, \ldots, y_n)\end{aligned}$$

由向量的内积可知：

$$\overrightarrow{x}\cdot\overrightarrow{y} = \cos\theta\,|\overrightarrow{x}|\,|\overrightarrow{y}|\,(\theta =< \overrightarrow{x}, \overrightarrow{y} >)$$

其中:

$$\overrightarrow{x}\cdot\overrightarrow{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \sum_{i=1}^{n} x_i y_i$$

$$|\overrightarrow{x}|\cdot|\overrightarrow{y}| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}\cdot\sqrt{y_1^2 + y_2^2 + \cdots + y_n^2} = \sqrt{\sum_{i=1}^{n} x_i^2}\cdot\sqrt{\sum_{i=1}^{n} y_i^2}$$

可得到

$$\cos\theta = \frac{\overrightarrow{x}\cdot\overrightarrow{y}}{|\overrightarrow{x}|\cdot|\overrightarrow{y}|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\cdot\sqrt{\sum_{i=1}^{n} y_i^2}}$$

不难知道，$\theta \in \mathbb{R}$ 时，有 $\cos\theta \in [-1,1]$,$\cos^2\theta \in [0,1]$，所以有

$$\cos^2\theta = \frac{\left(\sum_{i=1}^{n} x_i y_i\right)^2}{\sum_{i=1}^{n} {x_i}^2\cdot\sum_{i=1}^{n} {y_i}^2}$$

可以得到取值范围：

$$0 \leq \frac{\left(\sum_{i=1}^{n} x_i y_i\right)^2}{\sum_{i=1}^{n} x_i^2\cdot\sum_{i=1}^{n} y_i^2} \leq 1$$

最后可以得出柯西不等式:

$$0 \leq \left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \sum_{i=1}^{n} x_i^2\cdot\sum_{i=1}^{n} y_i^2$$

当且仅当 $\frac{x_i}{y_i}$ 为定值时，取等号。

□

我们常用到二、三维柯西不等式，也就是 $n=2$ 或 3 时；一般形式的柯西不等式：对于 $\forall a_i, b_i \in \mathbb{R}, i =$

$1, 2, \ldots, n$，有：

$$\left({a_1}^2 + {a_2}^2 + \cdots + {a_n}^2\right) \cdot \left({b_1}^2 + {b_2}^2 + \cdots + {b_n}^2\right) \geqslant (a_1b_1 + a_2b_2 + \cdots + a_nb_n)^2$$

二维形式：

$$\left({a_1}^2 + {a_2}^2\right) \cdot \left({b_1}^2 + {b_2}^2\right) \geqslant (a_1b_1 + a_2b_2)^2$$

三维形式：

$$\left({a_1}^2 + {a_2}^2 + {a_3}^2\right) \cdot \left({b_1}^2 + {b_2}^2 + {b_3}^2\right) \geqslant (a_1b_1 + a_2b_2 + a_3b_3)^2$$

§5.3.4 三角不等式

三角不等式

对于 $\forall a, b \in \mathbb{R}$ 有：

$$||a| - |b|| \leqslant |a + b| \leqslant |a| + |b|$$

这个不等式有向量的形式、复数的形式，我觉得用向量的方法去记是比较好的，不过我先证明一下最普遍的形式（以上这个）

证：因为对于任意实数 a, b 都有：

$$-|a||b| \leqslant ab \leqslant |a||b|$$

所以

$$|a|^2 - 2|a||b| + |b|^2 \leqslant a^2 + 2ab + b^2 \leqslant |a|^2 + 2|a||b| + |b|^2$$

两边开方后就得到

$$||a| - |b|| \leqslant |a + b| \leqslant |a| + |b|$$

□

这个不等式常用在有关绝对值的运算中，若将 a, b 理解成向量，则 $\overrightarrow{a}, \overrightarrow{b}, \overrightarrow{a+b}$ 构成一个三角形，而三角不等式表示的意义是在三角形中，任意两边之和大于第三边，任意两边之差小于第三边。

若对于 $-|a||b| \leqslant ab \leqslant |a||b|$ 我们给式子乘一个-1，同样可以推出，所以综合一下就是：

$$||a| - |b|| \leqslant |a \pm b| \leqslant |a| + |b|$$

什么时候用 $+$ 什么时候用 $-$，这得具体情况具体分析，如果 $-$ 的形式可以化简式子，那就用 $-$ 的形式。

下面我举个有关绝对值不等式的例子：

已知 $f(x) = |x - a| + |x + 3|$，若 $f(x) > -a$，求 a 的取值范围。

用三角不等式可得：

$$|x - a| + |x + 3| \geqslant |x - a - x - 3| = |a + 3|$$

故只需要解 $|a + 3| > -a$ 的解集即可，最后可以分类讨论或两边平方来解不等式，可以得到 a 的取值为 $a \in (-\frac{3}{2}, \infty)$

§5.3.5 权方和不等式

权方和不等式在高考中常用于一些分式和的最值问题，但高考中基本只会用到权方和二元二次或多元二次的形式，对于更一般的形式以及证明我会放在后面。

权方和不等式

二元二次的形式：设：$x,y,a,b\in\mathbb{R}^+$

$$\frac{a^2}{x}+\frac{b^2}{y}\geqslant\frac{(a+b)^2}{x+y}$$

当且仅当 $\frac{a}{x}=\frac{b}{y}$ 时取等号。

多元二次的形式：设 $x_i,y_i\in\mathbb{R}^+$

$$\frac{x_1^2}{y_1}+\frac{x_2^2}{y_2}+\cdots+\frac{x_n^2}{y_n}\geqslant\frac{(x_1+x_2+\cdots+x_n)^2}{y_1+y_2+\cdots+y_n}$$
$$\sum_{i=1}^{n}\frac{x_i^2}{y_i}\geqslant\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{\sum_{i=1}^{n}y_i}$$

当且仅当 $x_i=ky_i$ 时取等号。

证明：对于二元二次不等式，只需用均值不等式即可证明。

$$\begin{aligned}\frac{a^2}{x}+\frac{b^2}{y}&\geqslant\frac{(a+b)^2}{x+y}\\(x+y)\cdot(\frac{a^2}{x}+\frac{b^2}{y})&\geqslant(a+b)^2\\a^2+\frac{x}{y}b^2+b^2+\frac{y}{x}a^2&\geqslant a^2+b^2+2ab\\\frac{y}{x}a^2+\frac{x}{y}b^2&\geqslant 2ab\qquad(1)\end{aligned}$$

通过均值不等式可知，(1) 式显然成立，QED。

对于多元二次不等式的证明，需要用到柯西不等式：

$$\begin{aligned}\frac{x_1^2}{y_1}+\frac{x_2^2}{y_2}+\cdots+\frac{x_n^2}{y_n}&\geqslant\frac{(x_1+x_2+\cdots+x_n)^2}{y_1+y_2+\cdots+y_n}\\(y_1+y_2+\cdots+y_n)\cdot\left(\frac{x_1^2}{y_1}+\frac{x_2^2}{y_2}+\cdots+\frac{x_n^2}{y_n}\right)&\geqslant(x_1+x_2+\cdots+x_n)^2\\\left(\sum_{i=1}^{n}\sqrt{y_i}^2\right)\cdot\left(\sum_{i=1}^{n}\frac{{x_i}^2}{\sqrt{y_i}^2}\right)&\geqslant\left(\sum_{i=1}^{n}x_i\right)^2\end{aligned}$$

不妨设 $\sqrt{y_i}=a_i,\frac{x_i}{\sqrt{y_i}}=b_i,i=1,2,\ldots,n$，则原式转换为：

$$\left({a_1}^2+{a_2}^2+\cdots+{a_n}^2\right)\cdot\left({b_1}^2+{b_2}^2+\cdots+{b_n}^2\right)\geqslant(a_1b_1+a_2b_2+\cdots+a_nb_n)^2$$

由柯西不等式可知该式显然成立，QED。

下面举两个例子：

已知 $x \in (0,1)$，求 $\frac{9}{x}+\frac{16}{1-x}$ 的最小值。

这题我们可以用“1 的代换”来做，设 $a=x, b=1-x$，又 $a+b=1$，带回原式即可，但用权方和不等式会稍微快那么一点：

由权方和不等式可知：

$$(x+1-x)\left(\frac{9}{x}+\frac{16}{1-x}\right) \geqslant \left(\sqrt{9}+\sqrt{16}\right)^2 = 49$$

例二：已知 $a>1, b>1$，求 $\frac{b^2}{a-1}+\frac{a^2}{b-1}$ 的最小值。

由权方和不等式可知：

$$\frac{b^2}{a-1}+\frac{a^2}{b-1} \geqslant \frac{(a+b)^2}{a+b-2}$$

不妨设 $t=a+b-2$，代入可得

$$\begin{aligned}\frac{(a+b)^2}{a+b-2} &= \frac{(t+2)^2}{t} \\ &= t+4+\frac{4}{t} \geqslant 2\sqrt{4}+4=8\end{aligned}$$

故可得原式的最小值为 8

§5.3.6 对数均值不等式

定理

对数均值不等式 (ALG[a]) 对于互异的两正实数 a, b 有：

$$\frac{a+b}{2} > \frac{a-b}{\ln a-\ln b} > \sqrt{ab}$$

[a]Arithmetic-Logarithmic-Geometric mean inequalities

♠

需要注意的是 $a-b$ 或 $b-a$ 并不影响这个不等式，依据是：

$$\begin{gathered}\frac{a-b}{\ln a-\ln b} = \frac{b-a}{\ln b-\ln a} \\ -(\ln b-\ln a) = \ln a-\ln b\end{gathered}$$

在“极值点偏移”中的题目许多都用到这个不等式，证明这个不等式有许多种方法，高中最常用的方法我放在飘带放缩那一节具体在导数放缩那里，下面来看看用几何直观证明 ALG；

证明：先来证明

$$\frac{a+b}{2} > \frac{a-b}{\ln a-\ln b}$$

如下图

设 a, b 分别是图中 D, C，它们的中点 E 为 $\frac{a+b}{2}$，EF 垂直于 x 轴，F 是 $f(x)$ 上的点，过 F 点作与 $f(x)$

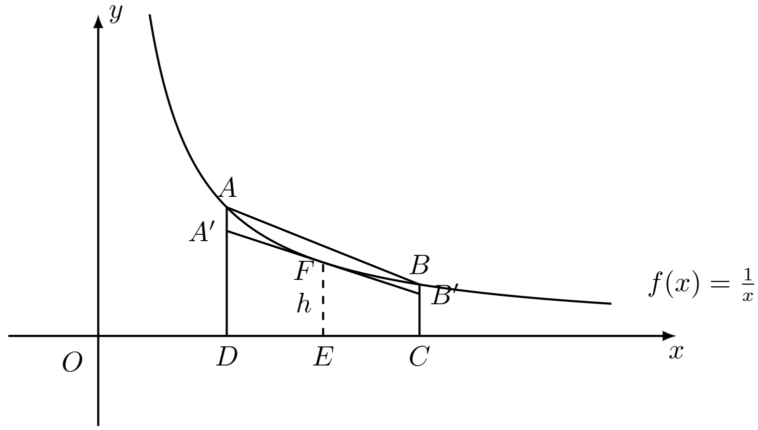


图 5.5: ALG 的几何证明 1

的切线交 AD, BC 于 A', B'；记 $S_1 = \widehat{AB}CD, S_2 = A'B'CD$，（$\widehat{AB}$ 的意思是弧 AB）直观的可以看出 $S_1 > S_2$；不难知道

$$S_1 = \int_a^b \frac{1}{x}\mathrm{d}x = \ln b - \ln a$$

通过梯形的面积公式也很容易知道：

$$\begin{aligned} S_2 &= \frac{(|A'D| + |B'C|)|DC|}{2} \\ &= |FC| \cdot |DC| \\ &= \frac{2(b-a)}{a+b} \end{aligned}$$

因此可以得到

$$\begin{aligned} S_1 &> S_2 \\ \ln b - \ln a &> \frac{2(b-a)}{a+b} \\ \frac{a+b}{2} &> \frac{b-a}{\ln b - \ln a} = \frac{a-b}{\ln a - \ln b} \end{aligned}$$

故左边得证；

现在来证明

$$\frac{a-b}{\ln a - \ln b} > \sqrt{ab}$$

如下图

设 a, b 分别为 D, C，记 E 代表 $\sqrt{ab}$，由于 $0 < a < b$ 不难证

$$a < \sqrt{ab} < b$$

EF 垂直于 x 轴，F 是 $f(x)$ 上的点，连接 AF, BF，记：$S_1 = \widehat{AB}CD, S_3 = S_{ADEF} + S_{BCEF}$，直观的可以看

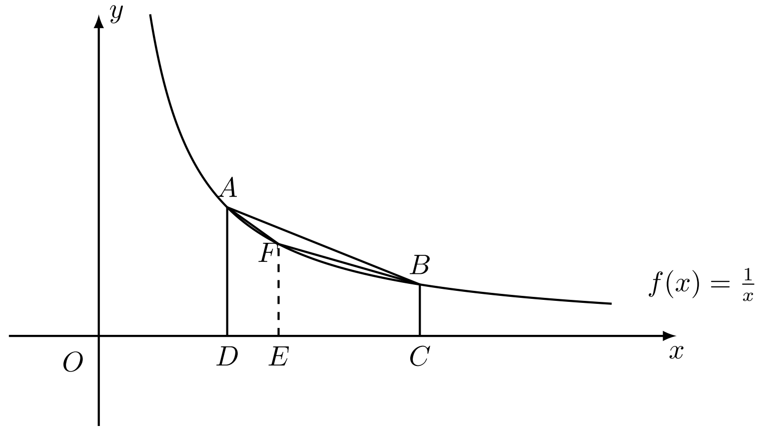


图 5.6: ALG 的几何证明 2

出 $S_3 > S_1$，可以知道

$$
\begin{aligned}
S_1 &= \int_a^b \frac{1}{x}\mathrm{d}x = \ln b - \ln a \\
S_3 &= (\frac{1}{\sqrt{ab}} + \frac{1}{b})(b - \sqrt{ab})\frac{1}{2} + (\frac{1}{\sqrt{ab}} + \frac{1}{a})(\sqrt{ab} - a)\frac{1}{2} \\
S_3 &= \frac{b-a}{\sqrt{ab}}
\end{aligned}
$$

综上：

$$
\begin{aligned}
S_3 &> S_1 \\
\frac{b-a}{\sqrt{ab}} &> \ln b - \ln a \\
\frac{a-b}{\ln a - \ln b} = \frac{b-a}{\ln b - \ln a} &> \sqrt{ab}
\end{aligned}
$$

□

额外的，记梯形 $ABCD$ 的面积为 S_4，易得

$$S_4 = \frac{f(a)+f(b)}{2}(b-a) = \frac{\frac{1}{a}+\frac{1}{b}}{2}(b-a)$$

从图中可以直观的得到

$$S_4 > S_3 > S_1$$

进而推出

$$
\begin{aligned}
S_4 &> S_3 \\
\frac{\frac{1}{a}+\frac{1}{b}}{2}(b-a) &> \frac{b-a}{\sqrt{ab}} \\
\sqrt{ab} &> \frac{2}{\frac{1}{a}+\frac{1}{b}}
\end{aligned}
$$

即调和平均数小于几何平均数，类似的得：

$$S_4 > S_1$$
$$\frac{\frac{1}{a}+\frac{1}{b}}{2}(b-a) > \ln b - \ln a$$
$$\frac{b-a}{\ln b - \ln a} > \frac{2}{\frac{1}{a}+\frac{1}{b}}$$

即得对数平均数大于调和平均数。

§5.4 比较大小的一些方法

比较大小最基本的方法比较法，分为**作差法**和**作商法**
作差法：利用 $a-b>0 \iff a>b$ 思路，将其化为恒等式后判断。
作商法：$a>0, b>0$ 有 $\frac{a}{b}>1 \Rightarrow a>b$，可用 $\frac{a}{b}$ 的值来判断。

构造函数

比大小中最常用的方法就是构造函数然后判断函数的单调性；例如下面的例子：

e.g. 已知 $a=\ln 2+\frac{1}{3}, b=\ln 3+\frac{1}{3}, c=\frac{\mathrm{e}+2}{\mathrm{e}+1}$，判断 a,b,c 的大小关系。注意到

$$c=\frac{\mathrm{e}+2}{\mathrm{e}+1}=\ln\mathrm{e}+\frac{1}{\mathrm{e}+1}$$

因此 a,b,c 都有相同的结构，所以构造函数

$$f(x)=\ln x+\frac{1}{x+1}$$

可知 $a=f(2), b=f(3), c=f(\mathrm{e})$，对 $f(x)$ 求导得

$$f'(x)=\frac{(x+1)^1-x}{x(x+1)^2}=\frac{x^2+x+1}{x(x+1)^2}$$

不难知道，当 $x>0$ 时 $f(x)$ 单调递增，因此可得

$$a<c<b$$

放缩

除了构造函数，我们可以通过一些放缩来判断量的关系，主要的放缩我放在导数那一节了，可以去查看比较常用的放缩式。

§5.4.1 糖水不等式

糖水不等式的核心就是放缩，通常意义下有两种糖水不等式，当然通过直觉理解是最好的，不过我也会写代数证明。

定理

若 $a>b>0,c>0$，则有：

$$\frac{b}{a}<\frac{b+c}{a+c}$$

若 $a_1>b_1>0,a_2>b_2>0,\frac{b_1}{a_1}<\frac{b_2}{a_2}$，则有：

$$\frac{b_1}{a_1}<\frac{b_1+b_2}{a_1+a_2}<\frac{b_2}{a_2}$$

♠

想象一下，现在有一杯糖水质量为 a 克，含糖 b 克，则它的糖水浓度为 $\frac{b}{a}$，现在加一块质量为 c 的糖进去，容易得到，新的甜水浓度为 $\frac{b+c}{a+c}$，生活常识告诉我们，一杯水加入糖肯定会更甜而不是更淡因此成立：

$$\frac{b}{a}<\frac{b+c}{a+c}$$

下面是代数证明：

$$\begin{gathered}\frac{b}{a}<\frac{b+c}{a+c}\\ b(a+c)<a(b+c)\\ bc<ac\\ b<a\end{gathered}$$

由条件，等式显然成立。

□

现在，我们有两杯糖水，一杯浓一杯淡，可以得到 $\frac{b_1}{a_1}<\frac{b_2}{a_2}$，将这两杯糖水混合在一起得到一杯新的糖水，其浓度为 $\frac{b_1+b_2}{a_1+a_2}$，生活经验告诉我们淡的加浓会更浓，而浓的加淡的会更淡，因此不等式成立：

$$\frac{b_1}{a_1}<\frac{b_1+b_2}{a_1+a_2}<\frac{b_2}{a_2}$$

下面进行代数证明：由于左边和右边证明类似，所以只证明左边（其实与上式证明也差不多）

$$\begin{gathered}\frac{b_1}{a_1}<\frac{b_1+b_2}{a_1+a_2}\\ b_1(a_1+a_2)<a_1(b_1+b_2)\\ b_1a_2<a_1b_2\\ \frac{b_1}{a_1}<\frac{b_2}{a_2}\end{gathered}$$

条件可知，该式成立。

□

§5.4.2　一些逼近方法

除了以上的方法外，还有一类方法是思维量少但计算量记忆量都很大的方法，也就是逼近，去求近似值。

泰勒展开

使用来求近似值，需要注意的是泰勒展开只是一种局部的性质，因此在用它进行近似计算时，x 不能远离 x_0，否则效果会比较差。

帕德逼近

，帕德逼近是一种用分式多项式来逼近的方法，效果比泰勒展开要更好，不过计算量很大。

对于 e^x，有如下函数逼近的近似表：

$f(x)=\mathrm{e}^x$	0	1	2
0	1	$x+1$	$\frac{x^2+2x+2}{2}$
1	$\frac{1}{1-x}$	$\frac{2+x}{2-x}$	$\frac{6+4x+x^2}{6-2x}$
2	$\frac{2}{x^2-2x+2}$	$\frac{2x+6}{x^2-4x+6}$	$\frac{x^2+6x+12}{x^2-6x+12}$

其中顶栏的 $0,1,2$ 表示分子的次数，侧栏同理。

对于 $\ln(x+1)$，有如下函数逼近的近似表：

$f(x)=\ln(x+1)$	0	1	2
0	$--$	x	$\frac{2x-x^2}{2}$
1	$--$	$\frac{2x}{x+2}$	$\frac{6x+x^2}{4x+6}$
2	$--$	$\frac{12x}{12+6x-x^2}$	$\frac{3x^2+6x}{x^2+6x+6}$

洛朗级数

§5.5 其他

§5.5.1 切比雪夫最佳逼近定理

§5.5.2 曼哈顿距离

第六章　数列

数列这一章，研究了有序数组的关系，在高中里主要研究等差和等比数列，其实数列可以看作是自变量取自然数集的函数，主要考点比较稳定，各种求和考察细心。

§6.1　数列的定义

定义

数列是指按正整数编了号的一串数，表示为 $\{a_n\}$ 或 $\{x_n\}$，其中 a_n 代表第 n 项，a_n 为**通项**，a_1 通常为称为**首项**。

按照数列的项分类，可分为有穷数列和无穷数列。

按照数列的**项的变化趋势**，可分为：

1. 递增数列：从第 2 项起，每一项都大于它的前一项。
2. 递减数列：从第 2 项起，每一项都小于它的前一项。
3. 常数列：各项相等。
4. 摆动数列：无规律。

数列可以看作一个函数：$D_f = \mathbb{N}^+, f(n) = a_n$，因此通项公式可以看作函数 $f(n)$ 的解析式。

下面两个是初学者容易混淆的概念，请注意区分

通项公式：如果数列 $\{a_n\}$ 的第 n 项与序号 n 之间关系可以用一个式子表示，则这个公式叫做该数列的通项公式，通项公式并不一定唯一，也并非所有数列都有通项公式。

通项公式例如：$a_n = n - 1$

递推公式[①]：如果已知数列 $\{a_n\}$ 的第 1 项 (或前几项)，且从第 2 项 (或某一项) 开始任意一项 a_n 与它前一项 a_{n-1} (或前几项) 间关系可以用一个公式表示，那么这个公式就是递推公式；*e.g.* $a_n = 2a_{n+1}$

§6.2　等差数列

等差数列

等差数列：从第 2 项起，每一项与前一项的差等于同一个数；这个数叫做**公差**，用 d 表示。

即：

$$a_n - a_{n-1} = d, n \geqslant 2$$

若 a, b, c 成等差数列，即有 $c = b + d = a + 2d$，可以推出

$$c + a = 2b = 2a + 2d$$

[①]在有些书中称为递归数列。

称 b 为 a 与 c 的**等差中项**。

由定义知道等差数列的递推公式，但在通常情况下，通项公式更加有用，因为不需要知道其前一项，下面给出等差数列的通项公式及其推导：

定理

等差数列 $\{a_n\}$ 的首项记为 a_1，公差为 d，则有

$$a_n = a_1 + (n-1)d$$

以上为等差数列的通项公式。

证明：（累加法）

由定义：$a_n - a_{n-1} = d, n \geqslant 2$，可列出：

$$a_n - a_{n-1} + \cdots + a_3 - a_2 + a_2 - a_1 = \sum_{k=1}^{n-1} d$$

注意到相邻两项可以两两消除，得到：

$$\begin{aligned} a_n - a_1 &= (n-1)d \\ a_n &= a_1 + (n-1)d \end{aligned}$$

□

由于 a_1, d 是常数，所以 a_n 可以看作是一个一次函数，但定义域为 $\mathbb{N}^+$; 反过来说，如果一个数列的通项公式是一个一次函数，那么它就是一个等差数列，其中 d 就是这个一次函数的斜率，当 d 大于 0 时，$d > 0 \iff \{a_n\}$ 是递增数列，小于 0 时为递减数列，等于 0 时为常数列。

对于 $m < n, m \in \mathbb{N}^+$，有

$$\frac{a_n - a_m}{n - m} = d$$

这是显然的；若 $m + n = p + q \, (m, n, p, q \in \mathbb{N}^+)$ 则有：

$$\begin{gathered} m - p = q - n \\ (m-p)d = (q-n)d \\ a_m + a_n = a_p + a_q \end{gathered}$$

特别的，当 $m + n = 2k \, (m, n, k \in \mathbb{N}^+)$ 时，有

$$a_m + a_n = 2a_k$$

这被称为等差中项的一般形式。

下面来介绍及推导等差数列的前 n 项和的公式。

定理

若数列 $\{a_n\}$ 是首项为 a_1，公差为 d 的等差数列，记它的前 n 项和为 S_n，有：

$$S_n = \frac{n(a_1+a_n)}{2} \quad ①$$
$$S_n = na_1 + \frac{n(n-1)}{2}d \quad ②$$

上面两个式子是等价的，① 就是大家所熟知高斯小学推导出的，而 ② 式是在计算中最常用到的形式；若将 a_n 的通项公式带入，可证两式等价。

证明：（倒序相加）

由定义可知

$$S_n = a_1 + a_2 + a_3 + \cdots + a_{n-1} + a_n$$

再将上式倒序写一遍

$$S_n = a_n + a_{n-1} + a_{n-2} + \cdots + a_2 + a_1$$

注意到 $a_1 + a_n = a_2 + a_{n-1} = \cdots = a_i + a_{n+1-i} = 2a_1 + (n-1)d$，则两式相加有

$$2S_n = \sum_{k=1}^{n}(a_k + a_{n+1-k})$$
$$S_n = a_1 n + \frac{n(n-1)}{2}d$$

□

另外，在计算 S_n 时有个小技巧，可证：

$$S_n = na_{\frac{n+1}{2}}$$

当 n 为奇数时，可以非常方便的化简 *e.g.* $S_7 = 7a_4$，当 n 为偶数时，式子也成立，不过 $a_{\frac{7}{2}}$ 看起来比较怪，只需带入通项公式算即可。

当 $\{a_n\}$ 为等差数列，S_n 为 $\{a_n\}$ 的前 n 项和时，数列 $\{S_{(m+1)n} - S_{mn}\}, m \in \mathbb{N}^+$，$n$ 为一正常数，则这个数列也成等差数列。即

$$S_n, S_{2n} - S_n, S_{3n} - S_{2n}, \cdots$$

为等差数列，且公差为 $n^2 d$，其中 d 为 $\{a_n\}$ 的公差。

证明：原题即证

$$S_{(m+1)n} - 2S_{mn} + S_{(m-1)n} = d'$$

其中 d' 为一常数；因为 S_n 为 $\{a_n\}$ 的前 n 项和，那么立即可写出

$$
\begin{aligned}
d' &= S_{(m+1)n} - 2S_{mn} + S_{(m-1)n} \\
d' &= \frac{n((m+1)(mn+n-1) - 2m(mn-1) + (m-1)(nm-n-1))}{2}d \\
d' &= \frac{2n^2}{2}d \\
d' &= n^2 d
\end{aligned}
$$

由已知，n, d 均为常数，即 d' 为常数。需要注意的是 $m = 0$ 时也即 S_0 并无意义，因此以上证明不能证

$$S_{2n} - 2S_n = n^2 d$$

不过上式展开可知显然成立。

□

§6.3　等比数列

等比数列

等比数列：从第 2 项起，每一项与前一项的比等于同一个数；这个常数叫做**公比**，用 q 表示 $(q \neq 0)$。即：

$$\frac{a_n}{a_{n-1}} = q, n \geqslant 2$$

若 a, b, c 成等比数列，即有 $c = bq = aq^2$，由定义：

$$\frac{b}{a} = \frac{c}{b}$$

可以推出

$$ac = b^2$$

称 b 为 a 与 c 的**等比中项**。

定理

记等比数列 $\{a_n\}$ 的首项为 a_1，公比为 q，则有

$$a_n = a_1 q^{n-1}$$

以上为等比数列的通项公式。

证明：（累乘法）有定义，可得：

$$\frac{a_n}{a_{n-1}}\cdot\frac{a_{n-1}}{a_{n-2}}\cdots\frac{a_3}{a_2}\cdot\frac{a_2}{a_1}=\prod_{k=1}^{n-1}q$$
$$\frac{a_n}{a_1}=q^{n-1}$$
$$a_n=a_1q^{n-1}$$

□

由定理看出，等比数列的通项公式可以近似的看作一个指数函数，易知当 $a_1>0,q>1$ 或者 $a_1<0,0<q<1$ 可推出 $\{a_n\}$ 为递增数列;$a_1<0,q>1$ 或者 $a_1>0,0<q<1$ 可推出 $\{a_n\}$ 为递减数列。

若 $m+n=p+q\,(m,n,p,q\in\mathbb{N}^+)$，则 $a_ma_n=a_pa_q$。

这一点的证明是显然的，只需带入等比数列的通项公式即证；特别的当 $m=n$ 时

$$2n=p+q$$
$$a_n^2=a_pa_q$$

即 a_n 是 a_p,a_q 的等比中项。

定理

等比数列的前 n 项和公式：

$$S_n=\begin{cases}na_1 & (q=1)\\ \frac{a_1(1-q^n)}{1-q}=\frac{a_1-a_nq}{1-q} & (q\neq 1)\end{cases}$$

♠

证明：（错位相减）设 $\{a_n\}$ 为等比数列，首项为 a_1 公比为 q，记 S_n 为 $\{a_n\}$ 的前 n 项和。

易知

$$S_n=a_1+a_2+a_3+\cdots+a_n$$
$$qS_n=a_2+a_3+a_4+\cdots+a_{n+1}$$

两式相减得

$$(1-q)S_n=a_1-a_{n+1}$$
$$S_n=\frac{a_1-a_{n+1}}{1-q}$$
$$S_n=a_1\frac{1-q^n}{1-q}$$

上面假设了 $q\neq 1$，当 $q=1$ 时，$a_1=a_2=\cdots=a_n$，亦即数列 $\{a_n\}$ 为常数列，显然

$$S_n=\sum_{k=1}^{n}a_1=na_1$$

综合起来就是：

$$S_n = \begin{cases} na_1 & (q=1) \\ \frac{a_1(1-q^n)}{1-q} = \frac{a_1-a_nq}{1-q} & (q \neq 1) \end{cases}$$

□

与等差数列类似，如果等比数列 $\{a_n\}$ 的前 n 项和为 S_n，那么数列

$$S_n, S_{2n}-S_n, S_{3n}-S_{2n}, \cdots$$

也成等比数列，且公比为 q^n

证明：即证 $\{S_{(m+1)n}-S_{mn}\}$ 为等比数列，易知

$$\begin{aligned} \frac{S_{(m+1)n}-S_{mn}}{S_{mn}-S_{(m-1)n}} &= \frac{1-q^{(m+1)n}-1+q^{mn}}{1-q^{mn}-1+q^{(m-1)n}} \\ &= \frac{q^{mn}-q^{(m+1)n}}{q^{(m-1)n}-q^{mn}} \\ &= \frac{q^{mn}}{q^{(m-1)n}}\frac{1-q^n}{1-q^n} = q^n \end{aligned}$$

且有

$$\begin{aligned} \frac{S_{2n}-S_n}{S_n} &= q^n\frac{1-q^n}{1-q^n} \\ &= q^n \end{aligned}$$

□

§6.4 线性递推数列

在等差、等比数列时，我们通过它们的定义，转换成递推公式，即

$$a_n = a_{n-1}+d, \qquad a_n = qa_{n-1}$$

再通过累加或累乘求出它们的通项公式。我们注意到它们的递推公式都是某一项与该项前一项的线性关系，因此引入定义

定义

给定数列 $\{a_n\}$，它与如果 $a_n, a_{n+1}, \ldots, a_{n+k}$ 可以写成

$$a_{n+k} = p(n) + \sum_{i=0}^{k-1} C_i a_{n+i}$$

则称上式为 k 阶递推式（更准确为 k 常系数线性阶递推式），其中 C_i 为常数。特别的

1. 当 $p(n)=0$ 时，称为 k 阶齐次线性递推式（齐次 k 常系数线性阶递推式）

2. 当 $p(n) \neq 0$ 时，称为 k 阶非齐次线性递推式（非齐次 k 常系数线性阶递推式）

显然，等比数列的递推式是齐次一阶常系数线性递推式，或称一阶齐次递推式；而等差数列的递推式则是非齐次一阶递推式。

一般情况下的 k 阶递线性推式不能像等差或等比一样快速的得到通项公式，下面来讨论求这些递推数列的方法。

§6.4.1 一阶线性递推式

一阶线性递推式可以写成

$$a_{n+1} = q(n)a_n + p(n)$$

下面先来讨论齐次的情况。

累乘求通项

累乘法除了用来求等比的通项外，还可以求齐次的一阶线性递推式，由于是齐次的，即 $p(n) = 0$，那么有

$$a_{n+1} = q(n)a_n$$

可以看出，等比数列是 $q(n)$ 为常数的特殊情况，将 a_n 移到等式左边，即

$$\frac{a_{n+1}}{a_n} = q(n)$$

注意到

$$\begin{cases} \frac{a_2}{a_1} = q(1) \\ \frac{a_3}{a_2} = q(1) \\ \cdots \\ \frac{a_{n+1}}{a_n} = q(n) \end{cases}$$

看出累乘后相邻项可消去，进行累乘

$$\begin{aligned} \prod_{k=1}^{n} \frac{a_{k+1}}{a_k} &= \prod_{k=1}^{n} q(k) \\ \frac{a_{n+1}}{a_1} &= \prod_{k=1}^{n} q(k) \\ a_{n+1} &= a_1 \cdot \prod_{k=1}^{n} q(k) \end{aligned}$$

再将 a_{n+1} 带入递推式即可得到 a_n 的通项公式。

累加求通项

$a_{n+1}=q(n)a_n+p(n)$ 当 $q(n)=1$ 时，是一个特殊的非齐次一阶线性递推式，之前我们已经用累加法求过等比数列的通项公式，事实上可以通过这个思路来求这种形式。不妨设为

$$a_{n+1}=a_n+p(n)$$

其中 $p(n)$ 是有关 n 的函数。

将 a_n 移到等式左边，即左边为"$a_{n+1}-a_n$"，右边为 $p(n)$，注意到

$$\begin{cases} a_1-a_2=p(1) \\ a_2-a_3=p(2) \\ \cdots \\ a_n-a_{n+1}=p(n) \end{cases}$$

若将它们加起来，相邻部分可以相消，即得

$$\begin{aligned} \sum_{k=1}^{n} a_n-a_{n+1} &= \sum_{k=1}^{n} p(k) \\ a_1-a_{n+1} &= \sum_{k=1}^{n} p(k) \\ a_{n+1} &= a_1-\sum_{k=1}^{n} p(k) \end{aligned}$$

其中 a_1 是常数，$\sum_{k=1}^{n} p(k)$ 通常可以表示出来，故可以得到 a_{n+1} 的通项公式，最后将 a_{n+1} 回递推式即可得到通项公式。

如 $a_n-a_{n-1}=f(n)$ 的式子，也是用累加法求，但格外注意的是取值范围应是 $n\geqslant 2$，以及项数、项的关系，这里很容易出错。

不动点

当 $a_{n+1}=q(n)a_n+p(n)$ 中 $q(n)=q,q\neq 1,p(n)=p$ 时，是比上一种更一般的非齐次一阶线性递推式。

求这类递推式的通项，通常有两种做法，一种是求不动点，将它转换成一个新数列的齐次一阶常系数递推式，再累乘来求。另一种是同时除以 q^{n+1}，构造一个新数列，通过累加来求，下面详细介绍一下两种方法。

$$a_{n+1}=qa_n+p \tag{1}$$

要将 ① 式转换成一个齐次一阶线性递推式，由于 a_n 前有一个常数 q，因此设这个新数列为 $b_n=a_n+x$，其中 x 为常数，满足

$$a_{n+1}+x=q(a_n+x)$$

这样一来，b_n 就是以 q 为公比的等比数列，而这个 x 就是数列 a_n 的不动点，求 x 可以通过待定系数法或者不

动点法，不动点法的具体操作如下：将原递推式中 a_{n+1}, a_n 均用 x 代替，即

$$x = qx + p$$

这个一次方程的解就是该数列的不动点。随后得到 $b_n = a_n + x$，通过等比数列的性质得

$$\frac{b_{n+1}}{b_n} = q$$
$$b_n = b_1 q^{n-1}$$

再将 $b_n = a_n + x$ 带入即可解出 a_n 的通项公式。

再来看第二种方法，将 ① 式两边同时除以 q^{n+1} 得到

$$\frac{a_{n+1}}{q^{n+1}} = \frac{a_n}{q^n} + \frac{p}{q^{n+1}}$$

令 $b_n = \frac{a_n}{q^n}$，这样就转为非齐次一阶线性递推式的第一种类型，通过累加即可求。

可以将第二种方法拓展一下，即对于

$$a_{n+1} = Aa_n + Bq^{n+1}$$

其中 A, B, q 均为常数，让等式两边同时除以 q^{n+1}

$$\frac{a_{n+1}}{q^{n+1}} = \frac{A}{q} \cdot \frac{a_n}{q^n} + B$$

此时，令 $b_n = \frac{a_n}{q^n}$，那么有

$$b_{n+1} = \frac{A}{q} b_n + B$$

用不动点或者重复之前的操作即可求得。

当非齐次一阶常系数线性递推式中的 $p(n)$ 为一次函数时，不妨设为

$$a_{n+1} = Aa_n + Bn + C$$

这种类型的整体思路和之前求非线性递推式，但 $p(n) = p$ 的做法是类似的，同样是构造一个新数列，转换为齐次递推式。这时推荐用待定系数法来求，不妨设 $b_n = a_n + xn + y$，它满足

$$\frac{b_{n+1}}{b_n} = \frac{a_{n+1} + x(n+1) + y}{a_n + xn + y} = D$$

整理后对比得出待定的系数 x, y, D，显然 b_n 为一等比数列，通过等比数列的通项，即可求出 a_n 的通项公式。

在处理一阶线性递推式时，可能会遇到“$a_{n+1}a_n$”的结构，称为手拉手模型。处理方法一般是对等式两边同除以 $a_{n+1}a_n$，让它们两个分开，再进行构造。

§6.4.2 二阶线性递推式

这里主要介绍齐次二阶常系数线性递推式的求通项的方法。不妨设这个递推式为

$$a_{n+2} = qa_{n+1} + pa_n \tag{②}$$

求通项的思路是将二阶转换为一阶，再用一阶的方法去求通项；与一阶的不动点法思路类似，将上式转换为一个等比数列，不妨设

$$\frac{a_{n+2} - xa_{n+1}}{a_{n+1} - xa_n} = Q \tag{③}$$

将上式移向后展开，得到

$$a_{n+2} = (Q + x)a_{n+1} - Qxa_n$$

与 ② 式对比后发现，这些常数满足

$$\begin{cases} Q = q - x \\ Q = -\frac{p}{x} \end{cases}$$

将 Q 消去得到 x, p, q 间的关系

$$x^2 - qx - p = 0 \tag{④}$$

上式称为该递推式的**特征根方程**，而这个方程的解称为**特征根**。

当 ④ 有两解时，记为 x_1, x_2，带入 ③ 式中（带入 x_1 或 x_2 都是等价的），设 $b_n = a_{n+1} - xa_n$ 得

$$\frac{b_{n+1}}{b_n} = Q$$

即 $\{b_n\}$ 为等比数列，易得它的通项公式

$$b_n = a_{n+1} - xa_n = (a_2 - xa_1)Q^{n-1}$$

这样就成功的将 a_n 从二阶齐次递推式化为一阶非齐次递推式，按照之前的方法可求得通项。

当特征根方程用重根时，即 $x_1 = x_2$ 时，实际上处理方式是一样的，在此不过多赘述。

当特征根方程无解时，则有可能是周期数列。

e.g. 斐波那契数列（也称兔子数列），它的递推式是二阶齐次常系数线性递推式，可以表示为

$$F_{n+2} = F_{n+1} + F_n$$

且 $F_1 = F_2 = 1$，现求 F_n 的通项公式。

先根据递推式写出它的特征根方程：

$$x^2 - x - 1 = 0$$

可以解出它的两个特征根分别为

$$x_1 = \frac{1 + \sqrt{5}}{2}, \quad x_2 = \frac{1 - \sqrt{5}}{2}$$

故递推式可以写成

$$F_{n+2}-\frac{1+\sqrt{5}}{2}F_{n+1}=\frac{1-\sqrt{5}}{2}F_{n+1}+F_n$$
$$\frac{F_{n+2}-\frac{1+\sqrt{5}}{2}F_{n+1}}{F_{n+1}-\frac{1+\sqrt{5}}{2}F_n}=\frac{1-\sqrt{5}}{2}$$

构造一个新数列使 $p_n=F_{n+1}-\frac{1+\sqrt{5}}{2}F_n$，则根据等比数列性质有

$$\frac{p_{n+1}}{p_n}=\frac{1-\sqrt{5}}{2}$$
$$p_n=p_1\left(\frac{1-\sqrt{5}}{2}\right)^{n-1}=\left(\frac{1-\sqrt{5}}{2}\right)^n$$
$$F_{n+1}-\frac{1+\sqrt{5}}{2}F_n=\left(\frac{1-\sqrt{5}}{2}\right)^n$$

这时将齐次二阶递推式转化为了非齐次一阶递推式，将等式两次同时除以 $\left(\frac{1+\sqrt{5}}{2}\right)^{n+1}$ 得

$$\frac{F_{n+1}}{\left(\frac{1+\sqrt{5}}{2}\right)^{n+1}}-\frac{F_n}{\left(\frac{1+\sqrt{5}}{2}\right)^n}=\frac{\left(\frac{1-\sqrt{5}}{2}\right)^n}{\left(\frac{1+\sqrt{5}}{2}\right)^{n+1}}=\left(\frac{1-\sqrt{5}}{1+\sqrt{5}}\right)^n\cdot\frac{2}{1+\sqrt{5}}$$

构造一个新数列使 $q_n=\frac{F_n}{\left(\frac{1+\sqrt{5}}{2}\right)^n}$，又由等差数列的性质得

$$q_{n+1}-q_n=\left(\frac{1-\sqrt{5}}{1+\sqrt{5}}\right)^n\cdot\frac{2}{1+\sqrt{5}}$$
$$q_{n+1}=q_1+\frac{2}{1+\sqrt{5}}\cdot\sum_{k=1}^{n}\left(\frac{1-\sqrt{5}}{1+\sqrt{5}}\right)^n$$
$$q_{n+1}=\frac{2}{1+\sqrt{5}}+\left(\frac{\left(1+\sqrt{5}\right)^n-\left(1-\sqrt{5}\right)^n}{\sqrt{5}\left(1+\sqrt{5}\right)^n}\cdot\frac{1-\sqrt{5}}{1+\sqrt{5}}\right)$$
$$\left(\frac{1+\sqrt{5}}{2}\right)^n\cdot q_{n+1}=F_{n+1}=\frac{\left(1+\sqrt{5}\right)^{n+1}}{2^{n+1}}\cdot\left(\frac{\left(1+\sqrt{5}\right)^n-\left(1-\sqrt{5}\right)^n}{\sqrt{5}\left(1+\sqrt{5}\right)^n}\right)$$
$$F_{n+1}=\frac{\left(1+\sqrt{5}\right)^{n+1}-\left(1-\sqrt{5}\right)^{n+1}}{2^{n+1}\cdot\sqrt{5}}$$
$$F_n=\frac{\left(1+\sqrt{5}\right)^n-\left(1-\sqrt{5}\right)^n}{2^n\cdot\sqrt{5}}=\frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n-\left(\frac{1-\sqrt{5}}{2}\right)^n\right]$$

即得斐波那契数列的通项公式。

□

§6.4.3 阶差法

根据数列 $\{a_n\}$ 的前 n 项和的定义，有：

$$S_n = a_1 + a_2 + a_3 + \cdots + a_n$$
$$S_{n-1} = a_1 + a_2 + a_3 + \cdots + a_{n-1}$$

两式相减得

$$S_n - S_{n-1} = a_n$$

如果知道 S_n 的表达式，就可以用这个推论求 a_n 的通项公式；**要注意的是**求完以后需验证一步 $S_1 = a_1$。

这种求通项公式的方法是最常见的一种方法，优点在于对于不是等差等比的数列仍可求通项，因为根据定义推出的。

§6.5 数列求和

§6.5.1 错位相减

当我们遇到“等差 $\times$ 等比”型的数列时，这类数列又称为**差比数列**，对于它们的求和一般用错位相减来处理，现设等差数列 $\{a_n\} = a_1 + (n-1)d$，等比数列 $\{b_n\} = b_1 \cdot q^{n-1}$，数列 $\{c_n\} = \{a_n \cdot b_n\}$ 的前 n 项和为 T_n

现在我们来求 T_n 的通项公式：

不难知道

$$\begin{aligned} T_n &= c_1 + c_2 + \cdots + c_n \\ &= a_1 \cdot b_1 + a_2 \cdot b_2 + \cdots + a_n \cdot b_n \\ &= a_1 \cdot b_1 q^0 + a_2 \cdot b_1 q^1 + \cdots + a_n \cdot b_1 q^{n-1} \end{aligned} \qquad ①$$

由 (1) 式，我们两边同时乘公比 q，得到：

$$qT_n = a_1 \cdot b_1 q^1 + a_2 \cdot b_1 q^2 + \cdots + a_n \cdot b_1 q^n \qquad ②$$

将 ① 式和 ② 式进行对比，注意到两式中 q 的指数相同的项相减后得到 d，即 ① − ②

$$\begin{aligned} (1-q)T_n &= a_1 \cdot b_1 q^0 + (a_2 - a_1) b_1 q_1 + (a_3 - a_2) b_1 q_2 + \cdots + (a_n - a_{n-1}) b_1 q_{n-1} - a_n \cdot b_1 q^n \\ &= a_1 \cdot b_1 q^0 + d \sum_{k=2}^{n} b_k - a_n \cdot b_1 q^n \\ &= a_1 \cdot b_1 + d b_2 \frac{1-q^n}{1-q} - a_n \cdot b_1 q^n \end{aligned}$$

综上，最后得到 T_n 的通项公式：

$$T_n = \frac{a_1 \cdot b_1}{(1-q)} + d b_2 \frac{1-q^n}{(1-q)^2} - \frac{a_n \cdot b_1 q^n}{1-q}$$

错位相减注重的是处理的过程，而不是最后的结论。

错位相减并不只是用在差比数列中，当数列为

$$t(k)=\frac{q(k)}{r(k)}$$

其中 $q(k)$ 是一元多项式，次数为 n，$r(k)$ 为等比数列时依旧可以用错位相减来求，且最多需要 n 次错位相减即可求和。证明如下：

设 $T(k)$ 为 $t(k)$ 的前 k 项和，设 $r(k)$ 的公比为 s，那么有

$$T(k)=\frac{q(1)}{r(1)}+\frac{q(2)}{r(2)}+\cdots+\frac{q(k)}{r(k)} \tag{1}$$

以及

$$\frac{1}{s}T(k)=\frac{q(1)}{r(1)}+\frac{q(1)}{r(3)}+\cdots+\frac{q(k)}{r(k+1)} \tag{2}$$

将 ① – ② 得到

$$\frac{s-1}{s}T(k)=\frac{q(1)}{r(1)}-\frac{q(k)}{r(k+1)}+\sum_{i=2}^{k}\frac{q(i)-q(i-1)}{r(i)} \tag{3}$$

注意“累加和”中的式子，不难知道其中 $q(k)-q(k-1)$ 是一个最高次为 $n-1$ 的一元多项式 $p(k)$，若令数列

$$t'(k)=\frac{p(k)}{r(k)}$$

那么累加和中继续出现形如 ③ 式的式子，可以继续将 $n-1$ 次一元多项式用错位相减展开为 $n-2$ 次多项式，循环往复；即最多通过 n 次错位相减，可以将多项式化为零多项式，继而用等比数列求和。

e.g. 设数列 $\{a_n\}$ 通项公式为 $a_n=\frac{n^2+n}{3^n}$，前 n 项和为 S_n，则有

$$S_n=\frac{1^2+1}{3^1}+\frac{2^2+2}{3^2}+\cdots+\frac{n^2+n}{3^n} \tag{4}$$

$$\frac{1}{3}S_n=\frac{1^2+1}{3^2}+\frac{2^2+2}{3^3}+\cdots+\frac{n^2+n}{3^{n+1}} \tag{5}$$

其中 ④ – ⑤ 式得

$$\begin{aligned}
\frac{2}{3}S_n&=\frac{1^2+1}{3^1}-\frac{n^2+n}{3^{n+1}}+\sum_{k=2}^{n}\frac{k^2+k-(k-1)^2-(k-1)}{3^k}\\
\frac{2}{3}S_n&=\frac{1^2+1}{3^1}-\frac{n^2+n}{3^{n+1}}+\sum_{k=2}^{n}\frac{2k}{3^k}\\
\frac{2}{3}S_n&=\frac{1^2+1}{3^1}-\frac{n^2+n}{3^{n+1}}+2\left(\frac{3}{2}\left(\frac{1}{3}-\frac{n}{3^{n+1}}+\sum_{k=2}^{n}\frac{1}{3^k}\right)-\frac{1}{3}\right)\\
\frac{2}{3}S_n&=\frac{1^2+1}{3^1}-\frac{n^2+n}{3^{n+1}}+\frac{3}{2}-\frac{2n+3}{2\cdot 3^n}-\frac{2}{3}\\
S_n&=\frac{9}{4}-\frac{2n^2+8n+9}{4\cdot 3^{n+1}}
\end{aligned}$$

可见，当一元多项式的次数稍微高一点时，这个计算量都是非常大的；因此我推荐用裂项来计算这种式子。

实际上，任意差比数列（公比不为 1）都可以裂项来计算，这一点我会在裂项那一节证明。

§6.5.2 裂项相消

裂项相消一般是把数列的通项公式进行拆分，转换成 $a_n = f(n) - f(n+1)$ 等能够前后抵消的形式；与累加求通项有异曲同工之妙，网上有大量繁复的裂项，但我个人认为没必要去记，裂项考的更多是思想。

先说应用场景：倘若题目给出 a_n 的通项公式，要求 a_n 的前 n 项和，但 a_n 又是一个复杂的式子，这时候就考虑到裂项了。

对于繁杂的公式我一个都不想记，那么该如何去裂项呢？我们不妨逆向思考一下，倘若我们确定是要进行裂项，那么题目一定会有“相消”的部分，即设：

$$a_n = \frac{c}{f(n)f(n+1)}$$

如果把带 n 的看作小的项，带 $n+1$ 看作大的项，因为它一定会“相消”，所以必定会转换成：

$$a_n = C \cdot \left(\frac{1}{\text{小}} - \frac{1}{\text{大}}\right)$$

其中 C 是一个常数，用“小”“大”只是为了说明项的关系，当只隔一项时就可以是：

$$a_n = C \cdot \left(\frac{1}{f(n)} - \frac{1}{f(n+1)}\right)$$

通分后可求出 C

$$\begin{aligned}\frac{c}{f(n)f(n+1)} &= \frac{C\left(f(n+1) - f(n)\right)}{f(n)f(n+1)}\\ C &= \frac{c}{f(n+1) - f(n)}\end{aligned}$$

一般 $\frac{c}{f(n+1)-f(n)}$ 最后都会化为一个常数，如果不是也没啥问题（个人感觉），最后求 a_n 的前 n 项和 S_n 为：

$$\begin{aligned}S_n &= C\left(\frac{1}{f(1)} - \frac{1}{f(2)} + \frac{1}{f(2)} - \frac{1}{f(3)} + \cdots + \frac{1}{f(n)} - \frac{1}{f(n+1)}\right)\\ S_n &= C\left(\frac{1}{f(1)} - \frac{1}{f(n+1)}\right)\end{aligned}$$

这是我做题一直用的方法，目前还没看到过更可靠更一般的方法，虽然根据考点反推有点本末倒置的感觉，但在此类题型中，这种方法很适用。。。

下面给出一些常用的裂项公式

1. 等差数列型：

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}$$

$$\frac{1}{(2n-1)(2n+1)} = \frac{1}{2}\left(\frac{1}{2n-1} - \frac{1}{2n+1}\right)$$

2. 无理型：

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

3. 指数型：

$$\frac{2^n}{(2^n-1)(2^{n+1}-1)}=\frac{1}{2^n-1}-\frac{1}{2^{n+1}-1}$$

这些都可以用我上述的方法推出，这里读者尝试自证。

Gosper 算法

关于裂项，下面介绍一个系统性的方法————**Gosper 算法**，这个方法通常是用来将超几何单项式裂项，从而达到求超几何级数的目的；关于原理我不在这里介绍，下面先介绍流程再举几个例子。

设我们要进行裂项的数列为 $t(k)$，首先将 $t(k+1)$ 与 $t(k)$ 作比

$$\frac{t(k+1)}{t(k)}=\frac{p(k+1)q(k)}{p(k)r(k+1)} \quad ①$$

作比之后通常会得到分子和分母为多项式的情况,上式的等号右边我们设了三个新的多项式数列,即 $p(k),q(k),t(k)$。

我们先不管 $p(k)$，不妨设 $p(k)=1$，那么 ① 式可以写成

$$\frac{t(k+1)}{t(k)}=\frac{q(k)}{r(k+1)}$$

由于分子分母均为多项式，不妨设分子的多项式为 $q(k)$，分母的多项式为 $r(k+1)$，根据代数学基本定理我们可以将 $q(k),r(k)$ 表示

$$\begin{cases} q(k) &= C_q(k+q_1)(k+q_2)(k+q_3)\cdots(k+q_n) \\ r(k) &= C_r(k+r_1)(k+r_2)(k+r_3)\cdots(k+r_m) \end{cases}$$

如果对于任意 $q_i-r_j,\ i\in 1,2,3,\ldots,n,\ j\in 1,2,3,\ldots,m$ 都不为正整数，那么就可以进行下一步；否则我们令

$$p(k)=(k+r_j+1)(k+r_j+2)\cdots(k+q_i-1)=\prod_{n=1}^{q_i-r_j-1}(k+r_j+n)$$

当有多组 q_i-r_j 为正整数时（每个 q_i 只需任意找到一个 r_j，这一点会在例子中体现），需要将它们乘起来，再把对应 $q(k),r(k)$ 中相同的项给消去。

下面是一个检验的环节，来检查这个数列 $t(k)$ 是否能用 Gosper 算法裂项，记多项式 $p(k),q(k),r(k)$ 的最高次数分别为 d_p,d_q,d_r，特别的当多项式为零时，记为 -1，如 $p(k)=0$ 时，记 $d_p=-1$。接下来将 $q(k),r(k)$ 相减以及相加

$$\begin{cases} Q(k) &= q(k)-r(k) \\ R(k) &= q(k)+r(k) \end{cases}$$

这里最主要的是看 $Q(k),R(k)$ 的最高次数，记它们的最高次数为 d_Q,d_R，并进行以下步骤

$$\begin{cases} d_Q\geqslant d_R & \Rightarrow d=d_p-d_Q \\ d_Q<d_R & \Rightarrow d=d_p-d_R+1 \end{cases}$$

这样一来我们会得到一个常数 d，下面进行 Gosper 判别

$$\begin{cases} d \geqslant 0 & \text{可以进行裂项} \\ d < 0 & \text{不可以用 Gosper 裂项} \end{cases}$$

进行判别后，如果可以用 Gosper 裂项，那么列出 Gosper 方程：

$$p(k) = q(k)s(k+1) - r(k)s(k)$$

其中 $p(k), q(k), r(k)$ 我们都已经知道了，只需求出多项式 $s(k)$，其中

$$s(k) = s_d k^d + s_{d-1} k^{d-1} + \cdots + s_2 k^2 + s_1 k + s_0$$

也即 $s(k)$ 为 d 次多项式，可以通过待定系数法带入 Gosper 方程，对比系数即可解出 $s(k)$。

最后一步

$$t(k) = T(k+1) - T(k)$$

其中 $T(k)$ 就是我们要求的，它可以表示为

$$T(k) = \frac{r(k)s(k)t(k)}{p(k)}$$

下面来举几个例子：

e.g. 记数列 $t(k)$ 为

$$t(k) = (-1)^{k+1} \frac{3^k(8k-2)}{(2k-1)(2k+1)}$$

求数列 $t(k)$ 的前 n 项和 S_n。

首先把 $t(k+1)$ 写出来

$$t(k+1) = (-1)^{k+2} \frac{3^{k+1}(8k+6)}{(2k+1)(2k+3)}$$

将 $t(k+1)$ 与 $t(k)$ 作比，并设多项式 $p(k), q(k), r(k)$

$$\frac{t(k+1)}{t(k)} = \frac{-3(k+\frac{3}{4})(k-\frac{1}{2})}{(k-\frac{1}{4})(k+\frac{3}{2})} = \frac{p(k+1)q(k)}{p(k)r(k+1)}$$

上式中，我把括号内的一次项系数提出，即一次项系数都为 1，之后不妨先设 $p(k)=1$，则

$$\begin{cases} q(k) & = -3\left(k+\frac{3}{4}\right)\left(k-\frac{1}{2}\right) \\ r(k+1) & = \left(k-\frac{1}{4}\right)\left(k+\frac{3}{2}\right) \end{cases}$$

注意 $r(k)$ 应为

$$r(k) = \left(k - \frac{5}{4}\right)\left(k + \frac{1}{2}\right)$$

将 $q_i - r_j,\ i \in 1,2,3,\dots,n,\ j \in 1,2,3,\dots,m$ 依次对照，发现仅有一对为正整数，即

$$q_1 - r_1 = \frac{3}{4} - \left(-\frac{5}{4}\right) = 2$$

那么 $p(k)$ 将为（这里 $k + r_1 + 1 = k + r_1 + (q_1 - r_1 - 1)$）

$$p(k) = k + r_1 + 1 = k + q_1 - 1 = k - \frac{1}{4}$$

也即 $p(k+1) = k + \frac{3}{4}$，这时将 $p(k+1), q(k)$ 对照，$p(k), r(k+1)$ 对照，将相同的式子消掉，保证 $\frac{t(k+1)}{t(k)}$ 不变，那么有

$$\begin{cases} p(k) & = k - \frac{1}{4} \\ q(k) & = -3(k - \frac{1}{2}) \\ r(k) & = k + \frac{1}{2} \end{cases}$$

现在进行检验，记多项式 $p(k), q(k), r(k)$ 的最高次数分别为 d_p, d_q, d_r，显然

$$d_p = d_q = d_r = 1$$

记

$$\begin{cases} Q(k) & = q(k) - r(k) = -4k + 1 \\ R(k) & = q(k) + r(k) = -2k + 2 \end{cases}$$

可见 $Q(k), R(k)$ 的最高次数均为 1，也即 $d_Q = d_R$，那么常数 d 为

$$d = d_p - d_Q = 0 \qquad ②$$

下面列出 Gosper 方程式

$$p(k) = q(k)s(k+1) - r(k)s(k)$$

其中 $s(k)$ 由 ② 式可知为一常数，不妨记 $s(k) = s_0$，带入

$$k - \frac{1}{4} = s_0(-4k + 1)$$

解出 $s(k) = -\frac{1}{4}$，最后可得

$$t(k) = T(k+1) - T(k)$$

其中

$$\begin{aligned} T(k) = \frac{r(k)s(k)t(k)}{p(k)} &= \frac{-\frac{1}{4}(k + \frac{1}{2})}{k - \frac{1}{4}}(-1)^{k+1}\frac{3^k(8k-2)}{(2k-1)(2k+1)} \\ &= \frac{1}{(2k-1)-\frac{1}{3}^{k}} \end{aligned}$$

综上可得

$$t(k) = \frac{1}{(2k+1)-\frac{1}{3}^{k+1}} - \frac{1}{(2k-1)-\frac{1}{3}^{k}}$$

而 $t(k)$ 的前 n 项和为

$$\begin{aligned}S_n=\sum_{k=1}^{n}t(k)&=T(2)-T(1)+T(3)-T(2)+\cdots+T(n+1)-T(n)\\&=\frac{1}{(2k+1)-\frac{1}{3}^{k+1}}-3\end{aligned}$$

在介绍错位相减时，我提到了任何公比不为 1 的差比数列都可裂项，下面用 Gosper 算法证明一下；

证明：不妨设差比数列为 $t(k)=(ak+b)mn^{k-1}$，其中 a 是等差部分的公差，$n(n\neq 1)$ 为公比，则

$$\frac{t(k+1)}{t(k)}=\frac{(k+\frac{b}{a}+1)n}{k+\frac{b}{a}}=\frac{p(k+1)q(k)}{q(k)r(k+1)}$$

显然 $\frac{b}{a}+1-\frac{b}{a}+1=2$，则设

$$\begin{cases}p(k)&=k+\frac{b}{a}\\q(k)&=n\\r(k)&=1\end{cases}$$

故可得 $d_p=1,d_q=0,d_r=0$，由于 $n\neq 1$ 确保了 $Q(k)\neq 0$：

$$\begin{cases}Q(k)&=n-1,\quad d_Q=0\\R(k)&=n+1,\quad d_R=0\end{cases}$$

那么 $d=1$，这意味着所有公比不为 1 的差比数列都可裂，进一步可得

$$\begin{aligned}k+\frac{b}{a}&=ns_1k+ns_1+ns_0-s_1-s_0\\\Rightarrow s(k)&=\frac{k}{n}+\frac{1}{n(n-1)}\end{aligned}$$

可求 $T(k)$

$$T(k)=\frac{r(k)s(k)t(k)}{p(k)}=\frac{amn^k}{n}+\frac{amn^{k-2}}{n-1}$$

□

注意到 Gosper 算法是将 $t(k)$ 裂为 $T(k+1)-T(k)$ 的形式，这在求和上非常方便，但也会出现有一些形如 $\frac{1}{k(k+2)}$ 的式子，这最直观（容易）的裂开显然是

$$\frac{1}{k(k+2)}=2\left(\frac{1}{k}-\frac{1}{k+2}\right)$$

也就是 $t(k)=T(k+2)-T(k)$，这样求和会有一些多的项出现，即

$$\sum_{k=1}^{n}t(k)=T(k+2)+T(k+1)-T(2)-T(1)$$

这种式子用 Gosper 算法也能裂，就如

$$\frac{1}{k(k+2)}=\frac{2k+1}{2(k+1)k}-\frac{2k+3}{2(k+2)(k+1)}$$

§6.5.3 分组求和

如果对于一个数列 $\{a_n\}$，其中 $a_n=b_n+c_n$，倘若要求 $\{a_n\}$ 的前 n 项和 S_n，我们可以分别求 $\{b_n\}$ 的前 n 项和 T_n 与 $\{c_n\}$ 的前 n 项和 H_n，然后把它们相加，即 $S_n=T_n=H_n$，如果觉得抽象，可以看看以下证明：

$$\begin{aligned}
&S_n=a_1+a_2+\cdots+a_n\\
&S_n=(b_1+c_1)+(b_2+c_2)+\cdots+(b_n+c_n)\\
&S_n=(b_1+b_2+\cdots+b_n)+(c_1+c_2+\cdots+c_n)\\
&S_n=T_n+H_n
\end{aligned}$$

分组求和其实考察的更多的是分类的思想，有时数列的通项为周期函数时，也可以进行分类处理，所以要理解分组求和就是要有分类的思想。

§6.5.4 倒序相加

若要计算一个数列的前 n 项和 S_n，先将 S_n 的没一项正序排一遍，再逆序排一遍，若发现首项和末尾项有关系，且第 n 项与倒数第 n 项都有一样的关系，这时就可以用倒序相加的方法；

e.g. 已知 $f(x)=\frac{4^x}{4^x+2}$，现求 $\sum_{i=1}^{2022}f(\frac{i}{2023})$。

令：

$$S=\sum_{i=1}^{2022}f(\frac{i}{2023})$$

首先正序、逆序列一遍：

$$\begin{aligned}
S&=f(\frac{1}{2023})+f(\frac{2}{2023})+\cdots+f(\frac{2022}{2023})\\
S&=f(\frac{2022}{2023})+f(\frac{2021}{2023})+\cdots+f(\frac{1}{2023})
\end{aligned}$$

两式相加得到

$$2S=[f(\frac{1}{2023})+f(\frac{2022}{2023})]+[f(\frac{2}{2023})+f(\frac{2021}{2023})]+\cdots+[f(\frac{2022}{2023})+f(\frac{1}{2023})]$$

注意到分组后方括号内函数的变量和始终为 1，例如 $\frac{1}{2023}+\frac{2022}{2023}=1$；这时我们来讨论 $f(x)+f(1-x)$ 的值：

$$\begin{aligned}
f(x)+f(1-x)&=\frac{4^x}{4^x+2}+\frac{\frac{4}{4^x}}{\frac{4+2\cdot4^x}{4^x}}\\
&=\frac{4^x}{4^x+2}+\frac{2}{4^x+2}=1
\end{aligned}$$

即知

$$2S = 2022$$
$$S = 1011$$

§6.6 数学归纳法

这一章主要介绍证明有关自然数的命题的方法——数学归纳法

第一数学归纳法

设 $P(n)$ 是一个含有自然数 n 的命题，则利用第一数学归纳法的证明步骤如下：

1. 验证 $n = n_0$ $(n_0 \in \mathbb{N}^+)$ 时，命题 $P(n_0)$ 成立；

2. 假设 $n = k$ $(k \in \mathbb{N}^+, k \geqslant n_0)$ 时，命题 $P(k)$ 成立，能推出 $n = k+1$ 时，命题 $P(k+1)$ 也成立。

根据上述两个步骤，知对一切自然数 n $(n \geqslant n_0)$，命题 $P(n)$ 成立。

第一步要证明 $P(n_0)$ 成立，称为**奠基步骤**，是论证命题的基础，第二步称为**归纳步骤**，它判断命题的正确性是否从特殊推广到一般。这两个步骤缺一不可。

§6.7 补充

§6.7.1 算两次

算两次原理，也叫富比尼原理，在组合会着重讲（强基），这里解释两个用算两次原理推导出的数列和。

e.g. 已知数列 $\{a_n\}$，其通项公式为：$a_n = n^2$，求这个数列的前 n 项和 S_n。

如若直接求和，那是看不出什么规律的，因此构造一个数列 $b_n = n(n+1)$，其前 n 项和为 T_n，显然

$$T_n = S_n + \sum_{k=1}^{n} k = S_n + \frac{n^2+n}{2} \qquad ①$$

另一方面，对 b_n 裂项，即等价为

$$b_n = n(n+1) = \frac{1}{3}\left(n(n+1)(n+2) - n(n+1)(n-1)\right)$$

那么 T_n 也可以表示为：

$$T_n = \frac{1}{3}\left(1 \times 2 \times 3 - 1 \times 2 \times 0 + 2 \times 3 \times 4 - 1 \times 2 \times 3 + \cdots + n(n+1)(n+2) - n(n+1)(n-1)\right)$$
$$T_n = \frac{n(n+1)(n+2)}{3} \qquad ②$$

结合 ①, ② 易得

$$
\begin{aligned}
n(n+1)(n+2) &= 3S_n + \frac{3n^2+3n}{2} \\
S_n &= \frac{2n(n+1)(n+2)}{6} - \frac{3n^2+3n}{6} \\
S_n &= \frac{n(n+1)(2n+1)}{6}
\end{aligned}
$$

这个结论可以靠数学归纳来证，因此可以给出一个没啥用的推论：

推论

对于数列 $\{a_n\}$，如果其通项公式为

$$a_n = kn^2 + ln + m$$

那么这个数列的前 n 项和为

$$
\begin{aligned}
S_n &= \frac{kn(n+1)(2n+1) + 3ln(n+1) + 6nm}{6} \\
&= \frac{n(n+1)(2kn+k+3l) + 6nm}{6}
\end{aligned}
$$

◇

证明是显然的，分组求和即可这里不多赘述。

e.g. 已知数列 $\{a_n\}$，其通项公式为：$a_n = n^3$，求这个数列的前 n 项和 S_n。这个命题是非常有名的，有非常多的几何证明方法，我这里还是用代数来证明一下，下面先直接给出结论：

$$
\begin{aligned}
\sum_{k=1}^{n} k^3 &= \left(\sum_{k=1}^{n} k\right)^2 \\
\sum_{k=1}^{n} k^3 &= \frac{n^2(n+1)^2}{4}
\end{aligned}
$$

证明：注意到

$$
\begin{aligned}
(n+1)^4 - n^4 &= \left((n+1)^2 - n^2\right)\left((n+1)^2 + n^2\right) = 4n^3 + 6n^2 + 4n + 1 \\
n^3 &= \frac{1}{4}\left((n+1)^4 - n^4\right) - \frac{3}{2}n^2 - n - \frac{1}{4}
\end{aligned}
$$

其中 $(n+1)^4 - n^4$ 即是裂项，$-\frac{3}{2}n^2 - n - \frac{1}{4}$ 也可求和，故

$$
\begin{aligned}
\sum_{k=1}^{n} k^3 &= \frac{1}{4}\sum_{k=1}^{n}\left((k+1)^4 - k^4\right) - \sum_{k=1}^{n}\left(\frac{3}{2}k^2 + k + \frac{1}{4}\right) \\
&= \frac{(n+1)^4 - 1}{4} - \frac{n(n+1)(2n+1)}{4} - \frac{n^2+n}{2} - \frac{n}{4} \\
&= \frac{n^2(n+1)^2}{4}
\end{aligned}
$$

□

§6.8 奇偶项

§6.8.1 通项

我们从一道简单的例题引入数列奇偶项的问题。

e.g. 已知数列 $\{a_n\}$ 满足

$$a_{n+1} = \begin{cases} a_n + 1, & n = 2k-1 \\ a_n + 2, & n = 2k \end{cases} \quad k \in \mathbb{N}^+$$

且 $a_1 = 1$，求 a_n 的通项公式。

只需要分情况讨论，当 n 为奇数时，$n+2$ 也为奇数，有

$$a_{n+2} = a_{n+1} + 2 = a_n + 3$$

由上式即可知 a_n 奇数项的递推关系，现构造一个新数列 $\{b_n\}$，使得 b_n 与 a_n 的奇数项一一对应，即

$$\begin{cases} b_1 & = a_1 \\ b_2 & = a_3 \\ b_3 & = a_5 \\ \cdots & \cdots \\ b_{\frac{n+1}{2}} & = a_n \end{cases}$$

则数列 $\{b_n\}$ 是以 $b_1 = 1$ 为首项，3 为公差的等差数列，那么有

$$b_n = 3n - 2 \iff a_n = b_{\frac{n+1}{2}} = \frac{3n-1}{2}$$

因此当 n 为奇数时 a_n 的通项公式如上。对于偶数项，我们仍然可以用以上的方法，不过更简便的方法是：当 n 为偶数时，$n+1$ 为奇数，那么

$$a_n = a_{n+1} - 2 = \frac{3n-2}{2}$$

综上

$$a_n = \begin{cases} \frac{3n-1}{2}, & n = 2k-1 \\ \frac{3n-2}{2}, & n = 2k \end{cases} \quad k \in \mathbb{N}^+$$

可见奇偶项的基本思路就是分类讨论。下面再来看一道

e.g. 数列 $\{a_n\}$ 满足 $a_1 = 1$，且 $a_n a_{n+1} = 2^{-n}$, $n \in \mathbb{N}^+$ 求数列 $\{a_n\}$ 的通项公式。

这题虽然没有说明奇偶项，但与之前的思路一样，写出隔项的递推关系。

$$a_{n+2} = \frac{1}{2^{n+1} a_{n+1}} = \frac{a_n}{2} \iff 2a_{n+2} = a_n$$

构造新数列 b_n

$$\begin{cases} b_1 & = a_1 \\ b_2 & = a_3 \\ b_3 & = a_5 \\ \cdots & \cdots \\ b_{\frac{n+1}{2}} & = a_n \end{cases}$$

即知 $\{b_n\}$ 是首项为 1，公比为 $\frac{1}{2}$ 的等比数列，那么

$$b_n = 2^{1-n} \iff a_n = b_{\frac{n+1}{2}} = \frac{1}{2^{\frac{n-1}{2}}}$$

当 n 为偶数时，$n+1$ 为奇数，即有：

$$a_n = 2^{\frac{n}{2}-n} = \frac{1}{2^{\frac{n}{2}}}$$

综上有

$$a_n = \begin{cases} \frac{1}{2^{\frac{n-1}{2}}}, & n = 2k-1 \\ \frac{1}{2^{\frac{n}{2}}}, & n = 2k \end{cases} \quad k \in \mathbb{N}^+$$

§6.8.2 求和

对于奇偶项的求和，通常采用分组求和，例如将奇数项和偶数项分别求和，再加起来；再例如相邻几项分为一组进行求和，下面逐一讨论。

（一）奇偶分类求和。

e.g. 已知数列 $\{a_n\}$ 满足

$$a_{n+1} = \begin{cases} a_n + 1, & n = 2k-1 \\ a_n + 2, & n = 2k \end{cases} \quad k \in \mathbb{N}^+$$

且 $a_1 = 1$，S_n 为 $\{a_n\}$ 的前 n 项和，求 S_n。

由前文已知 a_n 的通项为：

$$a_n = \begin{cases} \frac{3n-1}{2}, & n = 2k-1 \\ \frac{3n-2}{2}, & n = 2k \end{cases} \quad k \in \mathbb{N}^+$$

对于 S_n 我们可以进行如下分组：当 n 为奇数时

$$\begin{aligned} S_n &= a_1 + a_2 + a_3 + \cdots + a_n \\ &= (a_1 + a_3 + a_5 + \cdots + a_n) + (a_2 + a_4 + a_6 + \cdots + a_{n-1}) \end{aligned}$$

与求通项类似，构造 b_n 与 a_n 的奇数项一一对应，那么有

$$a_1 + a_3 + a_5 + \cdots + a_n = b_1 + b_2 + b_3 + \cdots + b_{\frac{n+1}{2}}$$

这一步主要是为了明确项数，易得：

$$
\begin{aligned}
a_1 + a_3 + a_5 + \cdots + a_n &= \sum_{k=1}^{\frac{n+1}{2}} (3k-2) \\
&= \frac{(n+1)(a_1+a_n)}{4} \\
&= \frac{(n+1)(3n+1)}{8}
\end{aligned}
$$

再来，对于偶数项

$$
\begin{aligned}
a_2 + a_4 + a_6 + \cdots + a_{n-1} &= a_3 + a_5 + a_7 + \cdots + a_n + 1 - n \\
&= \frac{(n+1)(3n+1)}{8} - 1 + 1 - n
\end{aligned}
$$

故 S_n 为

$$S_n = \frac{3n^2+1}{4}$$

当 n 为偶数时 S_n 为

$$S_n = S_{n+1} - a_{n+1} = \frac{3n^2}{4}$$

（二）邻项分组

另一种思路是将相邻两项分组，即当 n 为奇数时

$$S_n = (a_1 + a_2) + (a_3 + a_4) + \cdots + a_n$$

记

$$
\begin{cases}
c_1 & = a_1 + a_2 \\
c_2 & = a_3 + a_4 \\
c_3 & = a_5 + a_6 \\
\cdots & \cdots \\
c_{\frac{n-1}{2}} & = a_{n-2} + a_{n-1}
\end{cases}
$$

注意到 a_n 的相邻两项奇偶性相反，那么易得 c_n 是以 c_1 为首项，6 为公差的等差数列

$$c_n = 3 + 6(n-1) = 6n - 3$$

即得

$$
\begin{aligned}
S_n &= a_n + \sum_{k=1}^{\frac{n-1}{2}} (6k-3) \\
&= \frac{(n-1)(3+6n-3)}{4} + \frac{3n-1}{2} \\
&= \frac{3n^2-1}{2}
\end{aligned}
$$

对于偶数项不多赘述。

这两种方法各有优点，对于第一种，在求一些相邻项没太大关系的数列时比较方便，如

$$a_n = \begin{cases} n, & n = 2k-1 \\ 2^n, & n = 2k \end{cases} \quad k \in \mathbb{N}^+$$

此外这种方法在求带 $(-1)^n$ 这种项时非常方便；而另一种方法在求解相邻项有关时，有较大优势，对于带 $(-1)^n$ 项的数列也比较好求。

e.g.

§6.8.3 归一表示

在讨论奇偶项时，往往需要说明 n 的奇偶性，那么有没有一个的通项公式可以统一的表示所有情况呢？下面先来看一例：

e.g. 设数列 $\{a_n\}$ 满足 $a_{n+1} + a_n = 4n - 3,\ n \in \mathbb{N}^+$，且 $a_1 = 2$，求 a_n 的通项公式。由定义有

$$a_{n+2} = 4n + 1 - a_{n+1} = 4 + a_n$$

当 n 为奇数时，构造 b_n 使其与 a_n 的奇数项一一对应，易知 b_n 是首项为 $b_1 = a_1 = 2$，公差为 4 的等差数列，则

$$b_n = 4n - 2 \iff a_n = b_{\frac{n+1}{2}} = 2n$$

当 n 为偶数时，有

$$a_n = 4n - 3 - a_{n+1} = 2n - 5$$

综上

$$a_n = \begin{cases} 2n, & n = 2k-1 \\ 2n-5, & n = 2k \end{cases} \quad k \in \mathbb{N}^+ \qquad ①$$

另一方面数列 $\{a_n\}$ 可以表示为

$$a_{n+1} + a_n = 4n - 3 \iff \frac{2a_{n+1} - 4(n+1) + 5}{2a_n - 4n + 5} = -1$$

即得

$$2a_n - 4n + 5 = 5(-1)^{n-1} \iff a_n = \frac{5(-1)^{n-1} + 4n - 5}{2} \qquad ②$$

可以发现 ② 式是 ① 式的统一表示，因为不需要去考虑 n 的奇偶性。

通过 ② 式求和，也远比 ① 求和方便很多，那么如何进行归一表示呢？下面是我个人的一些思考：

一般地，对于一个数列 a_n，它的奇偶通项分别设为 b_n, c_n，那么

$$a_n = \begin{cases} b_n, & n = 2k-1 \\ c_n, & n = 2k \end{cases} \quad k \in \mathbb{N}^+$$

当 n 为奇数时 $a_n = b_n$，当 n 为偶数时 $a_n = c_n$，非常自然的想法是：当 n 为奇数时 b_n 工作，否则 c_n 工作。不妨构造一个辅助数列 $\{h_n\}$，它满足

$$\{1, 0, 1, 0, 1, 0, 1, 0, \cdots\}$$

再来将 b_n,c_n 的定义域拓展到正整数，即得 a_n 的统一表示：

$$a_n=h_nb_n+h_{n+1}c_n \tag{③}$$

对于 h_n 的构造，最简单的是 $h_n=\frac{1-(-1)^n}{2}$，那么 ③ 表示为

$$\begin{aligned}a_n&=\frac{1-(-1)^n}{2}b_n+\frac{1-(-1)^{n+1}}{2}c_n\\&=\frac{b_n+c_n+(-1)^n(c_n-b_n)}{2}\end{aligned}$$

通过上式对照 ①,②，会发现它们其实是“等价”的，不过值得注意的是如果把 $h_n=\frac{1-(-1)^n}{2}$ 看作一个函数，它在实数域上不连续，不过这并不影响什么。现在对 a_n 的前 n 项进行求和

$$\begin{aligned}\sum_{k=1}^{n}a_k&=\sum_{k=1}^{n}\left(\frac{b_k+c_k+(-1)^k(c_k-b_k)}{2}\right)\\&=\frac{1}{2}\sum_{k=1}^{n}(b_k+c_k)+\frac{1}{2}\sum_{k=1}^{n}\left((-1)^k(c_k-b_k)\right)\\&=\frac{1}{2}\sum_{k=1}^{n}(b_k+c_k)+\frac{1}{2}\left(\frac{(-1)^n-1}{2}(c_n-b_n)+\sum_{k=1}^{n-1}\left(\frac{(-1)^k-1}{2}\right)(c_k-c_{k+1}+b_{k+1}-b_k)\right)\end{aligned}$$

最后一步用了一个阿贝尔变换，可见当 b_n,c_n 为等差数列时，上式是很好求和的，可以完全无视奇偶项的条件。

此外，对于 h_n 的选择是多样的，例如

$$h_n=\frac{1}{2}\left(1+\cos\left(\pi x\right)\right)$$

因此，a_n 的通项公式并不唯一。

§6.8.4 奇偶的推广——模 m 的剩余类

讨论数列 $\{a_n\}$ 的奇偶项，在更广的层面来说就是讨论模 2 的剩余类，剩余类可以看作整数集的划分，给定正整数 m，可以将整数分成 m 个集合，记作 $K_0,K_1,K_2,\ldots,K_{m-1}$，当 $n\in K_r,r=1,2,3,\ldots,m-1$ 时，满足通项 $C_{r,n}$，则

$$a_n=\begin{cases}C_{0,n}, & n\in K_0\\C_{1,n}, & n\in K_1\\C_{2,n}, & n\in K_2\\\cdots & \cdots\\C_{m-1,n}, & n\in K_{m-1}\end{cases}$$

那么对于 a_n 是否存在统一表示？

与之前类似，构造一个数列 h_n，满足

$$\{1,0,0,0,\ldots,1,0,0,0,\ldots\}$$

其中 $h_1 = h_{1+km} = 1, k = 1, 2, 3, \cdots$，此外的项均为 0，通过 h_n 可将 a_n 表示为

$$a_n = \sum_{k=0}^{m-1} (h_{n+k} C_{k,n})$$

现在的问题是如何构造 h_n 数列，最开始我的想法是通过牛顿插值，得到有限项（m 项）的通项，即

$$h_n = 1 + \sum_{k=1}^{m-1} \frac{(-1)^k}{k!} \prod_{i=1}^{k} (n-i)$$

在 $n = 1, 2, \ldots, m$ 时，h_n 满足条件，例如

$$h_n = 1 - (n-1) + \frac{(n-1)(n-2)}{2} - \frac{(n-1)(n-2)(n-3)}{6} + \frac{(n-1)(n-2)(n-3)(n-4)}{24}$$

当 $n = 1, 2, 3, 4, 5$ 时，满足 $\{1, 0, 0, 0, 0\}$，再来只需将 n 替换为周期为 5，且满足 $\{1, 2, 3, 4, 5\}$ 的数列即可。

这样其实没必要，直接构造周期为 m 的数列，然后通过特征方程求解即可。实际上操作也挺繁琐的，对于 h_n 的构造仅当 n 为一些特殊数时，h_n 构造出的通项会比较简洁。

第七章　杂类

§7.1　反函数

设函数 $f(x)=y$，若得到一个函数 $g(y)=x$，这样的函数叫 $f(x)=y$ 的反函数，记作 $x=f^{-1}(y)$ 或 $arcf(x)=g(x)$, 其中有 $D_g=R_f, D_f=R_g$，一般存在反函数的条件是原函数是一一对应的。

§7.1.1　函数与反函数的对称轴

证明：原函数的反函数与原函数关于 $y=x$ 对称

设函数 $F(x,y)=0$，它的反函数为 $F^{-1}(x,y)=0=F(y,x)$

对于点 $P(a,b)$，有 $a\in R_{F^{-1}}, b\in D_{F^{-1}}$ ，因此 $Q(b,a),P(a,b)$ 均有意义，现在证明 P,Q 关于 $y=x$ 对称：

要证明 P,Q 关于 $y=x$ 对称，只需要证明两部分：

1. P,Q 到对称轴 $y=x$ 的距离相等；
2. 过 P,Q 的直线与 $y=x$ 垂直。

令：

1. $a\neq b$
2. $l: y=x$
3. $P(a,b)$ 到 l 的距离为 d_1
4. $Q(b,a)$ 到 l 的距离为 d_2
5. 过 P,Q 的直线为 l_{PQ}

$$
\begin{aligned}
d_1 &= \frac{|b-a|}{\sqrt{2}} \\
d_2 &= \frac{|a-b|}{\sqrt{2}} \\
d_1 &= d_2
\end{aligned}
$$

得到 $d_1=d_2$，接下来证明过 PQ 的直线与 l 垂直

$$
\begin{aligned}
l_{PQ}: y-b &= \frac{b-a}{a-b}(x-a) \\
k_{PQ} &= \frac{b-a}{a-b} = -\frac{b-a}{b-a} = -1 \\
l: y &= x \\
k &= 1
\end{aligned}
$$

因此

$$\begin{aligned} k \cdot k_{PQ} &= -1 \\ &\Longrightarrow l \perp l_{PQ} \end{aligned}$$

故 P, Q 关于直线 $l : y = x$ 对称。

□

这个结论可以推广到二元方程 $F(x,y) = 0$ 和 $F(y,x) = 0$ 关于直线 $y = x$ 对称。这个推论的好处在于方程有 $D_F = R_{F^{-1}}, R_F = D_{F^{-1}}$，而反观映射，当且仅当映射是单射的时候有 $D_F = R_{F^{-1}}, R_F = D_{F^{-1}}$ 成立。

举个例子，抛物线 $x^2 = 2py(p > 0)$ 和 $y^2 = 2px(p > 0)$ 关于直线 $y = x$ 对称。

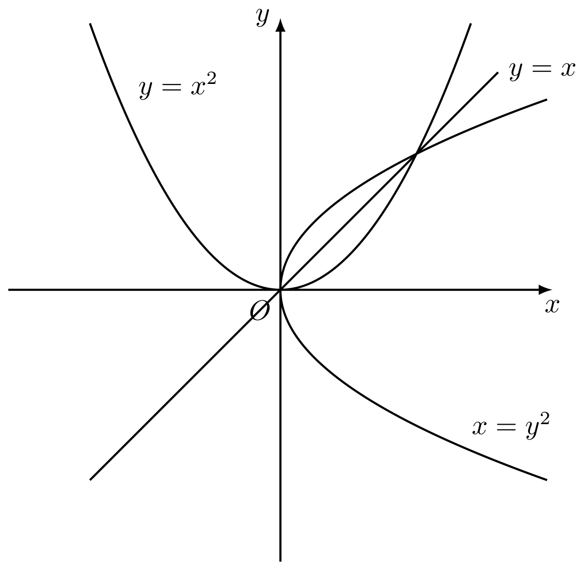


图 7.1: 函数与反函数对称关系

§7.2 方程的近似求解

$f(x) = 0$ 时，x 的值叫做零点，求方程 $f(x) = 0$ 的的解，是实际情况中常遇到的问题，以下说说高中中常用的方法。

§7.2.1 零点存在定理

零点存在定理

高中的定义: 如果函数 $y = f(x)$ 在区间 $[a, b]$ 上的图像是连续不断的一条曲线，并且有 $f(a) \cdot f(b) < 0$，那么，函数 $y = f(x)$ 在区间 (a, b) 内有零点，即存在 $c \in (a, b)$，使得 $f(c) = 0$，这个 c 也就是方程 $f(x) = 0$ 的根。

数学分析中的定义： 若函数 $f(x)$ 在闭区间 $[a, b]$ 连续，且 $f(a) \cdot f(b) < 0$，则一定存在 $\xi \in (a, b)$，使 $f(\xi) = 0$。

顺带一提连续函数的定义：若函数 $f(x)$ 在区间 (a, b) 的每一点都连续，则称函数 $f(x)$ 在开区间 (a, b) 上

连续；

以及函数 $f(x)$ 在 x_0 处连续的定义 (文字语言，$\varepsilon - N$ 语言太难懂了就不写了)：设函数 $f(x)$ 在点 x_0 的某个邻域中有定义，且成立

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

则称函数 $f(x)$ 在 x_0 连续，而称 x_0 是函数 $f(x)$ 的连续点。

§7.2.2 二分法

二分法本质是利用零点存在定理，将开区间一分为二，不断逼近零点。

设 $f(x)$ 在 $[a, b]$ 中连续，且成立

$$f(a) \cdot f(b) < 0$$

那么在 $[a, b]$ 至少纯在着 $f(x)$ 的一个解 x_0，假定 $f(x)$ 在 $[a, b]$ 上只有这一个解，我们希望求出它的近似值 $\tilde{x}$，满足

$$|\tilde{x} - x_0| \leqslant \varepsilon_0$$

这里 ε_0 是预先给定的精度要求，如 10^{-5} 等等，那么可以进行：

1. 记 $[a_1, b_1] = [a, b]$；取 x_1 为 $[a_1, b_1]$ 的中点，即 $x_1 = \frac{a_1+b_1}{2}$；
2. 计算 $f(x_1)$：

   若 $f(x_1) = 0$，那么 x_1 即是方程 $f(x)$ 的解，取 $\tilde{x} = x_1$，计算结束。
3. 否则，按如下规则得到区间 $[a_2, b_2]$：

   (a) 若 $f(x_1) \cdot f(b_1) < 0$，此时 $f(x)$ 的解在 $[x_1, b_1]$ 中，取 $a_2 = x_1, b_2 = b_1$。

   (b) 若 $f(x_1) \cdot f(b_1) > 0$，此时 $f(a_1) \cdot f(x_1) < 0$，因此 $f(x)$ 的解在 $[a_1, x_1]$ 中，取 $a_2 = a_1, b_2 = x_1$。

   易知 $x_0 \in [a_2, b_2]$，且 $[a_2, b_2]$ 的长度是 $[a_1, b_1]$ 的一半。
4. 取 x_2 的中点。
5. 类似的，若 x_2 是方程的解，计算结束；否则可以得到 $[a_3, b_3]$。
6. 重复上述过程。

§7.2.3 牛顿迭代法

数值计算中常用的求近似值的方法是**迭代法**，先将原来的方程

$$f(x) = 0$$

化为等价形式

$$x = F(x)$$

所谓“等价”是指若 x^* 是方程 $f(x) = 0$ 的解，则成立

$$x^* = F(x^*)$$

反之亦然，这里的 $F(x)$ 称为**迭代函数**。

取一个适当的初始值 x_0，按

$$x_{k+1} = F(x_k), k = 0, 1, 2, \cdots$$

产生序列 $\{x_k\}$（设每个 x_k 都属于 $F(x)$ 的定义域），这样的计算过程称为**迭代**。若在理论上成立

$$x_k \to x^* \ (k \to \infty)$$

那么显然 x^* 就是原方程的解，因此只要在迭代的过程中，选取某个合适的 x_k 作为 $\tilde{x}$，就得到原方程的近似解了。理论上，所选取的 x_k 应满足精度需求

$$|x_k - x^*| \leqslant \varepsilon_0$$

但 x^* 是不知道的，所以实际计算时往往采用比较相邻两次的迭代值是否满足

$$|x_{k+1} - x_k| \leqslant \varepsilon_0$$

来决定迭代是否继续下去。

构造迭代函数的方法有很多，例如，最简单的可以取成

$$F(x) = x - f(x)$$

下面利用泰勒公式来举一个例子。

设 $f(x)$ 在含有 x^* 的某个区间 $[a, b]$ 中具有二阶连续导数，且对于每个 $x \in [a, b]$，都有 $f'(x) \neq 0$。作出 $f(x)$ 在 x 处的泰勒公式，由于 x^* 是方程 $f(x) = 0$ 的解，则有

$$f(x^*) = f(x) + f'(x)(x^* - x) + f''(\xi)\frac{(x^* - x)^2}{2} = 0$$

也即

$$x^* = x - \frac{f(x)}{f'(x)} - \frac{f''(\xi)}{f'(x)} \cdot \frac{(x^* - x)^2}{2}$$

当 $x \to x^*$ 时，上式的最后一项是趋向于零的，因此有

$$\lim_{x \to x^*} \left[x - \frac{f(x)}{f'(x)}\right] = x^* - \frac{f(x^*)}{f'(x^*)} = x^*$$

这样，

$$F(x) = x - \frac{f(x)}{f'(x)}$$

就是一个满足 $x^* = F(x^*)$ 要求的迭代函数，由此得到迭代公式

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)},\ k = 0, 1, 2, \cdots$$

这就是著名的牛顿迭代法。

牛顿迭代法的几何意义明显，求解 $f(x) = 0$ 实际上是求曲线 $y = f(x)$ 与 x 轴的交点横坐标，曲线在

$x = x_k$ 的切线方程为

$$y = f'(x_k)(x - x_k) + f(x_k)$$

它与 x 轴的交点的横坐标恰为

$$x = x_k - \frac{f(x_k)}{f'(x_k)} = x_{k+1}$$

也就是说，牛顿迭代法的实质是通过一系列的切线与 x 轴的交点的横坐标，来逼近曲线与 x 轴的交点的横坐标，所以牛顿迭代法也叫牛顿切线法；

我们不加证明地给出如下结论：

定理

设 $f(x)$ 在 $[a,b]$ 中有二阶连续导数，且满足条件

1. $f(a) \cdot f(b) < 0$；
2. $f'(x)$ 在 (a,b) 保号；
3. $f''(x)$ 在 (a,b) 保号。

取 x_0 是 a 和 b 中满足

$$f(x_0) \cdot f''(x_0) > 0$$

的那一个点，则以 x_0 为初值的牛顿迭代过程

$$x_{k+1} = x_k - \frac{x_k}{f'(x_k)},\ k = 0, 1, 2, \cdots$$

产生序列 $\{x_k\}$ 单调收敛于方程

$$f(x) = 0$$

在 $[a,b]$ 中的唯一解。

♠

这里条件 (1) 保证了 $f(x)$ 在 (a,b) 有解，条件 (2) 表明 $f(x)$ 在 $[a,b]$ 严格单调，因此 $f(x)$ 在 (a,b) 中的解是唯一的，而条件 (3) 表示 $f(x)$ 在 $[a,b]$ 中保持固定的凸性，这保证了每一个 x_{k+1} 都在同一方向上比 x_k 更靠近 x^*，因此 $\{x_k\}$ 是一个单调有界数列，它必有极限，这个极限当然就是 x^*。

§7.3 一些补充

§7.3.1 二重根式

二重根式仅作补充知识，我在做题的过程中发现自己的答案和标答不一样，例如我解出 $\sqrt{2+\sqrt{3}}$，而标答是

$$\frac{\sqrt{6}}{2} + \frac{\sqrt{2}}{2}$$

通过做差发现其实它们是相等的，这种情况我就遇到过一次，我通过完全平方公式找出了嵌套二重根式化简的一般方法；虽然这个东西不在考纲，但我看日本高中数学的课程以及 AMC8 中都有这种题，因此我整理了资料简单讲讲。

定理

$$\sqrt{a\pm k\sqrt{b}}=\sqrt{\frac{a+\sqrt{a^2-k^2b}}{2}}\pm\sqrt{\frac{a-\sqrt{a^2-k^2b}}{2}}$$

通过研究判别式 $\Delta=a^2-k^2b$ 可分为几种情况：

1. 当 $\Delta<0$ 时，$\sqrt{a\pm k\sqrt{b}}$ 在实数域内无法分解；
2. 当 $\Delta\geqslant 0$ 时，$\sqrt{a\pm k\sqrt{b}}$ 在实数域内可以分解。

特别的当 Δ 为完全平方数时才能达到“化简”的目的，否则依旧是一个二重根式。

证明：设 $a,b,k\in\mathbb{Q}^+$

$$\begin{aligned}\sqrt{a\pm k\sqrt{b}}=\sqrt{a\pm 2\sqrt{\frac{bk^2}{4}}}&=\sqrt{x_1}\pm\sqrt{x_2}\\ a\pm 2\sqrt{\frac{bk^2}{4}}&=x_1+x_2\pm 2\sqrt{x_1x_2}\end{aligned}$$

得到 $x_1+x_2=a,x_1x_2=\frac{bk^2}{4}$，因此构造函数

$$f(x)=x^2-ax+\frac{bk^2}{4}$$

其中 x_1,x_2 是 $f(x)=0$ 的两个解，最后由求根公式即得：

$$\sqrt{a\pm k\sqrt{b}}=\sqrt{x_1}\pm\sqrt{x_2}=\sqrt{\frac{a+\sqrt{a^2-k^2b}}{2}}\pm\sqrt{\frac{a-\sqrt{a^2-k^2b}}{2}}$$

对于应试，只需要知道这个结论的核心是：任意一个能开放为有理数的数必然是某个有理数的平方即：

$$\sqrt{a\pm k\sqrt{b}}=\sqrt{x^2}$$

在实战中，我最常用的方法是共轭根式，下面结合例题看看操作过程：

e.g. 化简 $\sqrt{2+\sqrt{3}}$

令 $x^2=2+\sqrt{3}$，即求 x，不妨设

$$a^2+b^2=2,\quad 2ab=\sqrt{3}$$

这样一来 $x=a+b$，可以列出 x 的共轭根式：

$$(a-b)^2=2-\sqrt{3}$$

因此有

$$
\begin{aligned}
(a+b)(a-b) &= \sqrt{2+\sqrt{3}}\cdot\sqrt{2-\sqrt{3}} \\
a^2-b^2 &= \sqrt{4-3} = 1
\end{aligned}
$$

联立解出 a, b

$$
\begin{aligned}
a^2 &= \frac{3}{2} \\
b^2 &= \frac{1}{2}
\end{aligned}
$$

综上

$$
a+b = x = \frac{\sqrt{6}+\sqrt{2}}{2}
$$

第八章　线性方程组

这一章中，将会初步介绍有关线性方程组的知识。

§8.1　线性方程组与矩阵消元法

在实际问题中，往往会因为许多不确定的因素导致某些关系表示“均匀”的，即线性的，如“学习好”和“戴眼镜”这两个事件，是由多个因素决定的，它们可能有相关关系，但却没有确定关系。

现在我们讨论最简单（最理想化）的均匀变化的关系，例如某种食品蛋白质含量为 15%，那么 2000kg 的这种食品含蛋白质

$$2000\text{kg} \times 15\% = 300\text{kg}$$

在均匀变化的问题中，列出的是一次方程组。我们看下面这一则例子：

e.g.

第九章　行列式

第十章　矩阵的运算

第三部分

数学分析

第十一章　数列极限

§11.1　实数系的连续性

§11.1.1　实数系

这节我们初步了解一下实数系的连续性，首先介绍一下集合的封闭性。

定义

若一个集合中任意两个元素进行了某种运算，其结果仍旧属于这个集合，那么我们称该集合对这种运算是**封闭的**。

显然，若 $m,n\in\mathbb{N}$，那么 $m+n$ 或 mn 都属于集合 $\mathbb{N}$，故集合 $\mathbb{N}$ 对于 $+,\times$ 封闭。

自然数集是最先出现的数集（相较于我们接下来讨论的数集），随着人们的需要，出现了整数集。显然，整数集 $\mathbb{Z}$ 对于运算 $+,-,\times$ 均封闭，整数集是自然数集的扩充。

将整数集继续扩充，我们有有理数集 $\mathbb{Q}=\left\{x|x=\frac{q}{p},p\in\mathbb{N}^{+},q\in\mathbb{Z}\right\}$。显然，有理数集 $\mathbb{Q}$ 对于 $+,-,\times,\div$ 都是封闭的。

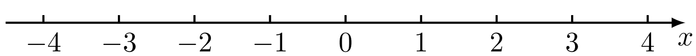


图 11.1: 一条数轴

现在，我们取一条数轴，在这个数轴上，整数集 $\mathbb{Z}$ 的每一个元素都可以找到对应点，这些点称为**整数点**。因为它们之间的最小间隔为 1，所有我们称整数系 $\mathbb{Z}$ 具有“**离散性**”。

对于有理数集 $\mathbb{Q}$，显然其元素也能在这个数轴上找到对应点，这些点叫做**有理点**。容易知道，任取一段大于零的长度，其上都有无穷多个有理点，亦即不存在没“有理点”的区域，我们称有理数系 $\mathbb{Q}$ 具有“**稠密性**”。

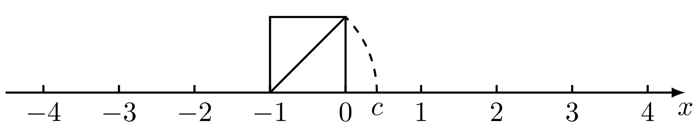


图 11.2: 希伯索斯 Hippasus 发现的 $\sqrt{2}$

不过有理数系还不够“完美”，现在假设有一个边长为 1 的正方形在该数轴上，记它对角线的长度为 c，这个 c 无法用有理数来表示。可以通过反证法来证明：根据勾股定理

$$c^2=1+1=2$$

假定 c 为有理数，不妨设 $c=\frac{q}{p},q\in\mathbb{Z},p\in\mathbb{N}^{+}$，并且 p,q 互素，那么有

$$q^2=2p^2$$

注意到若 k 为偶数，则 $k=2n$，其平方为 $k^2=4n^2$，即 $2\mid k^2$；当 k 为奇数，则 $k=2n+1$，其平方 $k^2=4n^2+4n+1$，显然 $2\nmid k^2$。

故我们可以知道奇数的平方为奇数，偶数的平方为偶数，则 q 是偶数，设 $q=2r$，故

$$4r^2=2p^2\Rightarrow 2r^2=p^2$$

也就是所 p 也是偶数，则 p,q 至少有一素因数 2，这与 p,q 互素的假定矛盾，故 c 不是有理数。

□

也就是说，有理数集 $\mathbb{Q}$ 对开放这个运算不封闭。也因此，有理数点虽然在数轴上密密麻麻，但并不充满整个数轴，其中还有许多“空隙”，例如上面讨论中的 c 就位于这些“空隙”中。

我们注意到有理数可以表示为有限小数或者无限循环小数，很自然想到扩充有理数集，可以把无限不循环小数（无理数）给归纳进来；因此定义

$$\mathbb{R}=\{x|x\text{为有理数或无理数}\}$$

上面这个集合称为**实数集**，下面会介绍到，这些补充进去的无理数对应的**无理点**，确实填充了有理点在数轴上留下的空隙。

这样一来，每个实数都可以在数轴上找到对应点，而数轴上每个点都可以找到唯一一个实数与之对应。实数集的这一性质称为实数系 $\mathbb{R}$ 的连续性。为了强调实数系所特有的这种连续性，$\mathbb{R}$ 又被称为**实数连续统**。

§11.1.2　最大数与最小数

下面讨论实数集的各种子集，简称数集。

设 S 为一数集，如果 $\exists\,\xi\in S$，使得 $\forall\,x\in S$，有 $x\leqslant\xi$，则称 ξ 是数集 S 的最大数，记为 $\xi=\max S$；

如果 $\exists\,\eta\in S$，使得 $\forall\,x\in S$，有 $x\geqslant\eta$，则称 η 是数集 S 的最小数，记为 $\eta=\min S$。

当数集 S 为非空有限集，即 S 只有有限个数时，显然 $\min S,\max S$ 均存在，且 $\max S$ 是有限个数中的最大者，$\min S$ 是有限个数中的最小者。但当 S 为无限集时，最大数和最小数有可能不存在。

e.g. 集合 $B=\{x|0\leqslant x<1\}$ 没有最大值。下面用反证法来证明。

证明：假设 β 是该集合的最大值，那么它满足 $\beta\in B$，可知 $\beta'=\frac{1+\beta}{2}\in B$，这与 $\beta>\beta'$ 矛盾，故命题得证。

□

§11.1.3　上确界以及下确界

参考函数有界性的定义，我们定义一个非空数集 S 的上下界。

定义

设 S 是一个非空数集，如果 $\exists\,M\in\mathbb{R}$，使得 $\forall\,x\in S$，都有 $x\leqslant M$，则称 M 为 S 的一个**上界**；如果 $\exists\,m\in\mathbb{R}$，使得 $\forall\,x\in S$，都有 $x\geqslant m$，则称 m 为 S 的一个**下界**。当数集 S 即有上界又有下界时，称为**有界集**。

显然，S 为有界集等价于 $\exists X>0$ 使得 $\forall x\in S$，有 $|x|\leqslant X$。

设数集 S 有上界，记 U 为 S 全体上界的集合，显然 U 没有最大数，下面将证明：U 一定有最小数。设

U 的最小数为 β，称 β 为数集 S 的**上确界**，即最小上确界，记作

$$\beta = \sup S$$

显然上确界 β 有如下性质：

1. β 是数集 S 的上界：$\forall\, x \in S$ 都有 $x \leqslant \beta$。
2. 任何小于 β 的数都不是 S 的上界：$\forall\, \varepsilon > 0, \exists\, x \in S$，使得 $x > \beta - \varepsilon$。

记 L 为 S 全体下界的集合，显然 L 无最小数。同样可以证明：L 有最大数。设 α 为 L 的最大数，称 α 为 S 的下确界，即最大下确界，记作

$$\alpha = \inf S$$

下确界 α 也有如下性质：

1. α 是数集 S 的下界：$\forall\, x \in S$ 都有 $x \geqslant \alpha$。
2. 任何大于 α 的数都不是 S 的下界：$\forall\, \varepsilon > 0, \exists\, x \in S$，使得 $x < \alpha + \varepsilon$。

确界存在定理

非空有上界的数集必有上确界；非空有下界的数集必有下确界。♠

这个定理又叫做**实数系连续性定理**。

证明：任何一个实数 x 都能表示为

$$x = [x] + \{x\}$$

其中 $[x]$ 为 x 的整数部分，$\{x\}$ 为 x 的非负小数部分。将 $\{x\}$ 表示为无限小数的形式：

$$\{x\} = 0.a_1a_2a_3\cdots a_n\cdots$$

其中 $a_k, k \in \mathbb{N}$ 是数字 $0, 1, 2, \ldots, 9$ 中的一个，若 $\{x\}$ 为有限小数，则在后面接上无穷个 0。这称为实数的**无限小数表示**。

注意到无限小数 $0.a_1a_2a_3\cdots a_p000\cdots$ 于无限小数 $0.a_1a_2a_3\cdots(a_p-1)999\cdots$ 是相等的，为了保持表示的唯一性，我们约定无限小数的表示不出现后者，这样一来任何一个实数集合 S 就可以由一个确定的无限小数的集合表示：

$$\{a_0 + 0.a_1a_2a_3\cdots a_n\cdots | a_0 = [x], 0.a_1a_2a_3\cdots a_n\cdots = \{x\}, x \in S\}$$

设该数集 S 有上界，并令 S 中元素的整数部分最大者为 α_0，并记

$$S_0 = \left\{x | x \in S\text{并且}[x] = \alpha_0\right\}$$

显然 S_0 不是空集，并且对于任意 $x \in S$，如果 $x \notin S_0$，则说明 $x < \alpha_0$。

接着再来考察 S_0 中的元素的无限小数表示中的第一位小数的数字，记它们中最大的数字为 α_1，并记集合

$$S_1 = \left\{x | x \in S\text{并且}x\text{的第一位小数位}\alpha_1\right\}$$

显然 S_1 也不是空集，并且对于任意 $x \in S$，如果 $x \notin S_1$ 那么说明 $x < \alpha_0 + 0.\alpha_1$。

一般地，考察数集 S_{n-1} 中元素的无限小数表示中的第 n 位小数的数字，令它们中的最大数位 α_n，并记集合

$$S_n = \left\{x | x \in S_{n-1}\text{并且}x\text{的第 } n \text{ 位小数位}\alpha_n\right\}$$

显然 S_n 不是空集，并且对于任意 $x \in S$，如果 $x \notin S_n$ 那么说明 $x < \alpha_0 + 0.\alpha_1\alpha_2\cdots\alpha_n$。

照这样做下去，我们可以得到一系列非空数集：

$$S \supset S_0 \supset S_1 \supset \cdots \supset S_n \supset \cdots$$

以及一列数 $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$，它们满足

$$\begin{aligned}&\alpha_0 \in \mathbb{Z}\\&\alpha_k \in 0, 1, 2, \ldots, 9, k \in \mathbb{N}^+\end{aligned}$$

令

$$\beta = \alpha_0 + 0.\alpha_1\alpha_2\alpha_3\cdots\alpha_n\cdots$$

下面分两步证明 β 是数集 S 的上确界。

（一）、设 $x \in S$，则要么存在整数 $n_0 \geqslant 0$，使得 $x \notin S$；要么对任何整数 $n \geqslant 0$ 有 $x \in S_n$。

若 $x \notin S_{n_0}$，则有

$$x < \alpha_0 + 0.\alpha_1\alpha_2\alpha_3\cdots\alpha_{n_0} \leqslant \beta$$

若 $x \in S_n(n \in \mathbb{Z})$，有 S_n 的定义逐个比较 x 与 β 的整数部分以及每一位小数，即知 $x = \beta$。故对于任意 $x \in S$ 都有 $x \leqslant \beta$，即 β 是数集 S 的上确界。

（二）、对于任意给定的 $\varepsilon > 0$，只要将自然数 n_0 取得充分大，就有

$$\frac{1}{10^{n_0}} < \varepsilon$$

取 $x_0 \in S_{n_0}$，则 β 与 x_0 的整数部分以及前 n_0 位小数是相同的，所以

$$\beta - x_0 \leqslant \frac{1}{10^{n_0}} < \varepsilon$$

即知

$$x_0 > \beta - \varepsilon$$

即任何小于 β 的数 $\beta - \varepsilon$ 都不是数集 S 的上界。

同样的方法可以证明非空有下界的数集必有下确界。

□

关于数集的上/下确界有以下的唯一性定理。

定理

非空有界数集的上/下确界是唯一的。♠

证明：

假定非空有上界的数集 S 的上确界为 $\sup S = \beta$ 以及 $\sup S = \beta'$，且 $\beta > \beta'$。根据上确界的性质有：

$$x \leqslant \beta', \forall x \in S$$

记 $\varepsilon = \frac{\beta-\beta'}{2} > 0$，由于 β 是数集 S 的一个上确界，则存在一个 $x_0 \in S$，使得 $\beta - \varepsilon < x_0 \leqslant \beta$，故

$$x_0 > \frac{\beta + \beta'}{2} > \beta$$

这与 $\sup S = \beta'$ 矛盾，故 $\beta \leqslant \beta'$。

同样的方法我们可以证明 $\beta' \leqslant \beta$，综上即证 $\beta = \beta'$。

用同样的方法可以证明非空有下界数集的下确界唯一。

□

§11.1.4 Dedekind 切割定理

§11.2　数列极限

§11.2.1　数列极限的定义

数列的基础概念已经在前文中介绍过了，下面不多对基础概念赘述。

在数学中，如果要计算一个无法直接求得的值，例如圆周率 π，我们常常用到**逼近**的方法，即计算一个较容易求得的、同时又精度较高的数作为它的近似值。

下面我们从数列极限开始了解极限的概念以及严格的定义。

定义

设 $\{x_n\}$ 是一给定数列，a 是一个实常数。如果对于任意给定的 $\varepsilon > 0$，可以找到正整数 N，使得当 $n > N$ 时，成立

$$|x_n - a| < \varepsilon$$

则称数列 $\{x_n\}$ **收敛于**a（或 a 是数列 $\{x_n\}$ 的**极限**），记为

$$\lim_{n\to\infty} x_n = a$$

有时也记为

$$x_n \to a(n \to \infty)$$

如果不存在实数 a，使得 $\{x_n\}$ 收敛于 a，则称数列 $\{x_n\}$ **发散**。

下面我们来看一下数列极限的几何意义。

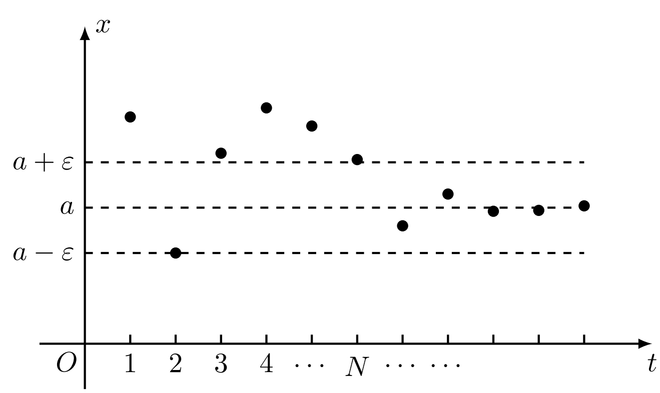


图 11.3: 数列极限的几何意义

将数列看作一个定义域为正整数集的函数，设

$$x = f(t)$$

在平面直角坐标系 Otx 的 x 轴上取 a 为中心，ε 为半径的一个开区间 $(a-\varepsilon, a+\varepsilon)$，称它为点 a 的 ε **邻域**，记为 $O(a,\varepsilon)$：

$$O(a,\varepsilon) = \{x | a - \varepsilon < x < a + \varepsilon\}$$

根据数列极限的定义，当 $n > N$ 时，成立 $|x_n - a| < \varepsilon$，这表示从第 $N+1$ 项起所有的项都落在点 a 的 ε 邻

域中，即

$$x_n \in O(a, \varepsilon), n > N$$

注意到 ε 的取值可以使得图中的上下两条线变化，当 ε 取越小时，它们靠得越近。但不管靠得多么近，总会有一项使得数列从那一项之后的每一项全部若在邻域中。

e.g. 证明数列 $\left\{\frac{n}{n+1919}\right\}$ 的极限为 1。

对于任意给定的 $\varepsilon > 0$，若要满足

$$\left|\frac{n}{n+1919} - 1\right| = \frac{1919}{n+1919} < \varepsilon$$

只需满足

$$n > \frac{1919}{\varepsilon} - 1919$$

显然 N 只需取任意大于 $\frac{1919}{\varepsilon} - 1919$ 的数即可，例如取

$$N = \left[\frac{1919}{\varepsilon}\right] + 1 > \frac{1919}{\varepsilon} - 1919$$

则当 $n > N$ 时，必有 $n > \frac{1919}{\varepsilon} - 1919$ 成立，于是有

$$\frac{1919}{n+1919} < \varepsilon$$

即证数列 $\left\{\frac{n}{n+1919}\right\}$ 的极限为 1。

□

在收敛的数列中，我们称极限为 0 的数列为**无穷小量**，例如 $\left\{\frac{1}{n}\right\}$；需要注意的是无穷小量是一个变量，而不是一个“非常小的量”。

由定义我们可以直接得到

$$\lim_{n\to\infty} x_n = a \iff \{x_n - a\}\text{是无穷小量}$$

§11.2.2 数列极限的性质

下面讨论数列数列的一些简单性质：

1. 极限的唯一性。

2. 数列的有界性。

3. 数列的保序性。

4. 极限的夹逼性。

虽然这些性质从几何直观上是显然正确的，但我们还是需要从代数的角度严格证明这些性质。

定理

收敛数列的极限必唯一。 ♠

证明：假设 $\{x_n\}$ 有极限 a, b，则根据定义有

$$\forall \varepsilon > 0,\ \exists N_1,\ \forall n > N_1 : |x_n - a| < \frac{\varepsilon}{2}$$

并且有

$$\exists N_2,\ \forall n > N_2 : |x_n - a| < \frac{\varepsilon}{2}$$

取 $N \in \max\{N_1, N_2\}$，由三角不等式，对于 $\forall n > N$:

$$\begin{aligned} |a - b| &= |a - x_n + x_n - b| \\ &\leqslant |x_n - a| + |x_n - b| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

由 ε 可任意接近于 0 得知，$a = b$。

□

第十二章　微分

导数涉及到极限概念，是高等数学的东西，高中在没有函数极限的基础上直接就学导数，导致部分公式需要靠背，就比如说链式法则就是用微分推得，但是即便是高中教课书也没明确给出链式法则推到过程，只是用“我们发现”这种借口搪塞，链式法则在高中就是一个比较容易忘记搞混的点，又例如洛必达法则的滥用，即便在小题中，洛必达法则也不是永远有效，所以像是很多微分中值定理，我以前推出来用函数二阶导判断其凹凸性的结论，我到现在还记得，所以如果学有余力的同学，可以拿一本数学分析从数列极限到函数极限再到微分和微分中值定理都补一补。

我不会从头开始讲，但为了更好理解一些导数中的结论，我会讲一点微分便于理解。

§12.1　微分的概念

一个给定函数 $f(x)$ 在点 x 附近有定义，若 $f(x)$ 的自变量 x 处产生了某个增量 Δx 变成了 $x+\Delta x$，那么它的函数值也相应的产生了一个增量 $\Delta y = f(x+\Delta x) - f(x)$；这里 $\Delta x, \Delta y$ 分别是自变量和应变量的差分。

定义：对函数 $y = f(x)$ 定义域中的一点 x_0，若存在一个只与 x_0 有关，而与 Δx 无关的数 $g(x_0)$，使得当 $\Delta x \to 0$ 时恒成立关系式

$$\Delta y = g(x_0)\Delta x + o(\Delta x)$$

则称 $f(x)$ 在 x_0 处的微分存在，或称 $f(x)$ 在 x_0 处可微。

其中 $o(x_0)$ 是高阶无穷小量，$g(x)\Delta x$ 是 Δy 的线性主要部分，因为 $g(x)\Delta x$ 是低阶无穷小量，在加减时起主要影响。

当 $f(x)$ 在 x 处是可微的且 $\Delta x \to 0$ 时，我们将 Δx 称为自变量的微分，记作 $\mathrm{d}x$，而将 Δy 的线性主要部分 $g(x)\mathrm{d}x$ 称为因变量的微分，记作 $\mathrm{d}y$ 就有了

$$\mathrm{d}y = g(x)\mathrm{d}x$$

例如：$y = f(x) = x^2$，可知

$$\Delta y = f(x+\Delta x) - f(x)$$
$$\Delta y = 2x\Delta x + \Delta x^2$$

其中 Δx^2 是高阶无穷小量，$g(x) = 2x$，则 $\mathrm{d}y = \mathrm{d}(x^2) = 2x\mathrm{d}x$。

可微必连续，但连续不一定可微

§12.2　导数的概念

若 $f(x)$ 在 x_0 处可微，则有关系式

$$\Delta y = g(x_0)\Delta x + o(\Delta x)$$

由于 $g(x_0)$ 与 Δx 无关，即知 $g(x_0)$ 是当 $\Delta x \to 0$ 时，因变量差分与自变量差分的比值 $\frac{\Delta y}{\Delta x}$ 的极限值。

定义：若函数 $y = f(x)$ 在其定义域中的一点 x_0 处极限

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

则称 $f(x)$ 在 x_0 处可导，并称这个极限值为 $f(x)$ 在 x_0 处的导数，记为 $f'(x_0)$。

若 $f(x)$ 在一区间每一点可导，则称 $f(x)$ 在该区间可导 $f(x)$ 在 x_0 处的导数还有如下等价定义

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

函数 $f(x)$ 的所有可导点的集合是 $f(x)$ 定义域的子集，上面的定义可以构成一个新的函数 $f'(x)$，称它为 $f(x)$ 的导函数。可知 $D_{f'} \subseteq D_f$

由微分关系式可知其 $g(x)$ 就是 $f'(x)$ 故有

$$\Delta y = f'(x)\Delta x + o(\Delta x)$$

$$\mathrm{d}y = f'(x)\mathrm{d}x$$

因此导数也叫微商。

对于一元函数，可微 $\iff$ 可导

§12.3　函数的性质与导数的关系

导数的几何意义：曲线的切线斜率就是极限

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

的值，即 $f(x)$ 在 x 处的导数值。

不难知道，当导数值小于 0 时函数单调递减，导数值大于 0 时函数单调递增；我们可以有这一性质来简单的画一个函数的草图。

定义

函数的极值，定义：设函数 $f(x)$ 在 (a,b) 有定义，$x_0 \in (a,b)$，如果存在点 x_0 的某一个邻域 $O(x_0,\delta) \subset (a,b)$ 使得：

$$f(x) \leqslant f(x_0), x \in O(x_0,\delta)$$

则称 x_0 是 $f(x)$ 的一个极大值点，$f(x_0)$ 称为相应的极大值，类似的有极小值点，它们统称极值；极值

是相对与附近的，是一个局部性质。

有 x_0 是 $f(x)$ 的一个极值点，则 $f'(x_0)=0$。

§12.4 常用导数及运算法则

计算一个函数的导函数的运算称为对这个函数**求导**。

通过定义可以直接对一些简单函数求导，下面举几个例子：

e.g. 求 $y=\sin x$ 的导函数。

因为

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

由 $\cos x$ 的连续性以及 $\sin\frac{\Delta x}{2}\approx\frac{\Delta x}{2}(\Delta x\to 0)$，可知

$$\begin{aligned}&\lim_{\Delta x\to 0}\frac{\sin(x+\Delta x)-\sin x}{\Delta x}\\=&\lim_{\Delta x\to 0}\left(x+\frac{\Delta x}{2}\right)\cdot\lim_{\Delta x\to 0}\frac{\sin\frac{\Delta x}{2}}{\frac{\Delta x}{2}}\\=&\cos x\end{aligned}$$

即得

$$(\sin x)'=\cos x$$

导数的四则运算法则：设 $f(x),g(x)$ 在某一区间上都是可导的，则对于任意常数 c_1,c_2，它们的线性组合 $c_1f(x)+c_2g(x)$ 也在该区间上可导，且满足如下线性运算关系：

$$[c_1f(x)+c_2g(x)]'=c_1f'(x)+c_2g'(x)$$

$f(x),g(x)$ 它们的积函数也在该区间上可导，满足：

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

$g(x)$ 在某区间可导，且 $g(x)\neq 0$，则它的倒数也在该区间可导：

$$\left[\frac{1}{g(x)}\right]'=-\frac{g'(x)}{[g(x)]^2}$$

推论：

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

反函数求导法则:若 $f(x)$ 在 (a,b) 上连续、严格单调递增、可导并且 $f'(x)\neq 0$,记 $\alpha=\min(f(a+),f(b-)),\beta=$

$\max(f(a+), f(b-))$，则它们的反函数 $x = f^{-1}(y)$ 在 (α, β) 上可导，且有：

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

可以由这些定理推出基本初等函数的导数公式，这些没什么技巧看眼熟就行，实在不行自己再推一遍。

多个函数的导数四则运算很少会用到，因为考试的话计算量太大了，但我还是写一写

$$\left[\sum_{i=1}^{n} c_i f_i(x)\right]' = \sum_{i=1}^{n} c_i f_i'(x)$$

其中 $c_i(i = 1, 2, 3, \ldots, n)$ 为常数

$$\left[\prod_{i=1}^{n} f_i(x)\right]' = \sum_{j=1}^{n} \left\{ f_j'(x) \prod_{i=1, i\neq j}^{n} f_i(x) \right\}$$

复合函数是我们遇到的更多的情况，所有我们研究复合函数求导。

设函数 $u = g(x)$ 在 $x = x_0$ 可导，而函数 $y = f(u)$ 在 $u = u_0 = g(x_0)$ 处可导，则复合函数 $y = f(g(x))$ 在 $x = x_0$ 可导，且有

$$[f(g(x))]'_{x=x_0} = f'(u_0)g'(x_0) = f'(g(x_0))g'(x_0)$$

复合函数的求导规则可以写成

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}u} \cdot \frac{\mathrm{d}u}{\mathrm{d}x}$$

一般称为**链式法则**，从微分的角度更好理解链式法则的原理。

通过链式法则我们可以对基本初等函数的复合函数进行求导，但是会遇到一些基本初等函数之外的情况：

形如

$$y = f(x) = u(x)^{v(x)}$$

的函数称为**幂指函数**，对于幂指函数的求导，常用的方法叫做“对数求导法”，求导时先对两边取对数，令

$$z(x) = \ln y = v(x) \ln u(x)$$

分别求导：

$$z'(x) = [v(x) \ln u(x)]' = v'(x) \ln u(x) + v(x) \frac{u'(x)}{u(x)}$$

另一方面，将 $z(x) = \ln y$ 看作复合函数，则有

$$z'(x) = \frac{y'}{y}$$

综合这两式，可得到：

$$\begin{aligned} y' &= y\left[v'(x)\ln u(x) + v(x)\frac{u'(x)}{u(x)}\right] \\ &= u(x)^{v(x)}\left[v'(x)\ln u(x) + v(x)\frac{u'(x)}{u(x)}\right] \end{aligned}$$

下面举一个例子，设 $y = f(x) = x^x$，求 y'：

$$\begin{aligned} \ln y &= x\ln x \\ y' &= x^x \cdot (\ln x + 1) \end{aligned}$$

下面列出基本初等函数的导数和微分的公式：

$$\begin{aligned} &(C)' = 0 \\ &(x^\alpha)' = \alpha x^{\alpha-1} \\ &(\sin x)' = \cos x \\ &(\cos x)' = -\sin x \\ &(\tan x)' = \sec^2 x \\ &(\cot x)' = -\csc^2 x \\ &(\sec x)' = \tan x \sec x \\ &(\csc x)' = -\cot x \csc x \\ &(\arcsin x)' = \frac{1}{\sqrt{1-x^2}} \\ &(\arccos x)' = -\frac{1}{\sqrt{1-x^2}} \\ &(\arctan x)' = \frac{1}{1+x^2} \\ &(a^x)' = \ln a \cdot a^x \\ &(\log_a x)' = \frac{1}{\ln a} \cdot \frac{1}{x} \end{aligned}$$

特别的

$$\begin{aligned} (\mathrm{e}^x)' &= \mathrm{e}^x \\ (\ln x)' &= \frac{1}{x} \end{aligned}$$

$$\begin{aligned} &\mathrm{d}(C) = 0 \\ &\mathrm{d}(x^\alpha) = \alpha x^{\alpha-1}\mathrm{d}x \\ &\mathrm{d}(\sin x) = \cos x\mathrm{d}x \\ &\mathrm{d}(\cos x) = -\sin x\mathrm{d}x \\ &\mathrm{d}(\tan x) = \sec^2 x\mathrm{d}x \\ &\mathrm{d}(\cot x) = -\csc^2 x\mathrm{d}x \\ &\mathrm{d}(\sec x) = \tan x \sec x\mathrm{d}x \\ &\mathrm{d}(\csc x) = -\cot x \csc x\mathrm{d}x \\ &\mathrm{d}(\arcsin x) = \frac{\mathrm{d}x}{\sqrt{1-x^2}} \\ &\mathrm{d}(\arccos x) = -\frac{\mathrm{d}x}{\sqrt{1-x^2}} \\ &\mathrm{d}(\arctan x) = \frac{\mathrm{d}x}{1+x^2} \\ &\mathrm{d}(a^x) = \ln a \cdot a^x\mathrm{d}x \\ &\mathrm{d}(\log_a x) = \frac{1}{\ln a} \cdot \frac{\mathrm{d}x}{x} \end{aligned}$$

特别的

$$\begin{aligned} \mathrm{d}(\mathrm{e}^x) &= \mathrm{e}^x\mathrm{d}x \\ \mathrm{d}(\ln x) &= \frac{\mathrm{d}x}{x} \end{aligned}$$

§12.5 函数的凸性

定义

设函数 $f(x)$ 在区间 I 上定义，若对 I 中的任意两点 x_1 和 x_2，和任意 $\lambda \in (0,1)$，都有

$$f\left(\lambda x_1 + (1-\lambda)\, x_2\right) \le \lambda f(x_1) + (1-\lambda) f(x_2)$$

则称 $f(x)$ 是 I 上的下凸函数[a]。

若不等号严格成立，则称 $f(x)$ 在 I 上是严格下凸函数。

[a]下凸函数也叫凹函数

类似的可以给出上凸函数的定义。

二阶导数与凸性的关系

设函数 $f(x)$ 在区间 I 上二阶可导，则 $f(x)$ 在区间 I 上是下凸函数的充分必要条件是：对于任意 $x \in I$ 有 $f''(x) \ge 0$。

特别的，若对于任意 $x \in I$ 有 $f''(x) > 0$，则 $f(x)$ 在 I 上是严格下凸函数。

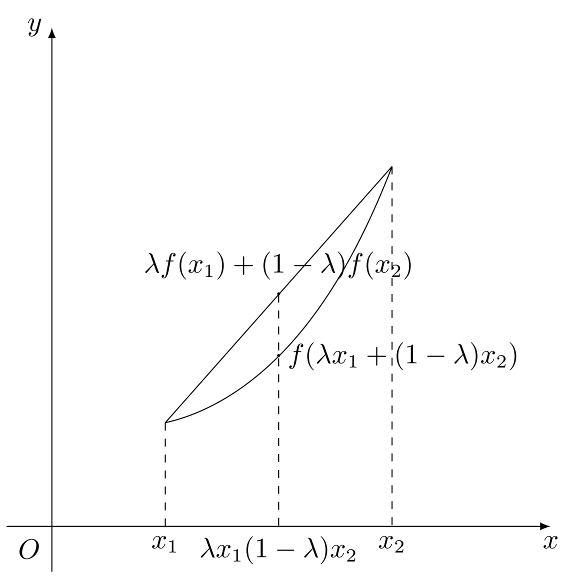


图 12.1: 二阶导数与凸性的关系

证明：必要性：

§12.6 微分中值定理

§12.6.1 拉格朗日中值

引理

罗尔 (Rolle) 定理： 设函数 $f(x)$ 在闭区间 $[a,b]$ 上连续，在开区间 (a,b) 上可导，且 $f(a)=f(b)$，则至少存在一点 $\xi\in(a,b)$，使得

$$f'(\xi)=0$$

♣

证明由闭区间上连续函数的性质，存在 $\xi,\eta\in(a,b)$ 满足：

$$f(\xi)=M,\ f(\eta)=m$$

其中 M,m 分别是 $f(x)$ 在 (a,b) 的最大值和最小值；现在分两种情况：

1. $M=m$，此时 $f(x)$ 在 $[a,b]$ 上恒为常数，结果显然成立。
2. $M>m$，这时 M,m 中至少有一个与 $f(a)$ 不相同，不妨设：

$$M=f(\xi)>f(a)=f(b)$$

因此 ξ 是极大值点，有

$$f'(\xi)=0$$

□

这个定理由几何直观是显然的，罗尔定理要满足三个条件，及 $[a,b]$ 上连续、(a,b) 上可导以及 $f(a)=f(b)$；三个条件缺一不可，我们来看看更一般的情况，若前两个条件满足，但第三个条件不成立，即 $f(a)\neq f(b)$，通过几何直观不难知道，必然至少存在一点 $\xi\in(a,b)$ 使得它的斜率 $f'(\xi)$ 与 $(a,f(a)),(b,f(b))$ 的割线斜率相同，这就是拉格朗日中值定理；

定理

拉格朗日 (Lagrange) 中值定理 设函数 $f(x)$ 在闭区间 $[a,b]$ 上连续，且在开区间 (a,b) 上可导，则至少存在一点 $\xi\in(a,b)$ 使得

$$f'(\xi)=\frac{f(b)-f(a)}{b-a}$$

♠

证明：作辅助函数

$$\varphi(x)=f(x)-f(a)-\frac{f(b)-f(a)}{b-a}(x-a),\ x\in(a,b)$$

由于 $f(x)$ 在闭区间 $[a,b]$ 上连续，开区间 (a,b) 上可导；因此函数 $\varphi(x)$ 也在闭区间 $[a,b]$ 上连续，开区间 (a,b) 上可导，并且有：

$$\varphi(a)=\varphi(b)=0$$

由罗尔定理，至少存在一点 $\xi\in(a,b)$，使得 $\varphi'(\xi)=0$，整理得到：

$$f'(\xi)=\frac{f(b)-f(a)}{b-a}$$

□

最后，给出一个更一般的中值定理

定理

柯西中值定理 设 $f(x)$ 和 $g(x)$ 都在闭区间 $[a,b]$ 上连续，在开区间 (a,b) 上可导，且对于任意 $x\in(a,b), g'(x)\neq 0$。则至少存在一点 $\xi\in(a,b)$，使得

$$\frac{f'(\xi)}{g'(\xi)}=\frac{f(b)-f(a)}{g(b)-g(a)}$$

♠

显然，当 $g(x)=x$ 时，上式为拉格朗日公式，所以拉格朗日中值定理是柯西中值定理的特殊情况；柯西中值定理本书不做证明。

§12.6.2 洛必达法则

在计算分式极限时，时常会遇到分子分母都趋近于零或都趋于无穷大的情况，这时可以使用洛必达法则；当分子分母趋近于零时，我们称这种类型的极限为 $\frac{0}{0}$ 待定型，都趋近于无穷大时称为 $\frac{\infty}{\infty}$ 待定型，简称 $\frac{0}{0}$ 型和 $\frac{\infty}{\infty}$ 型；除了这两种类型以外，还 $0\cdot\infty$ 型、$\infty\pm\infty$ 型、∞^0 型、1^∞ 型、0^0 型等几种，这后面几种都可以化为 $\frac{0}{0}$ 型或 $\frac{\infty}{\infty}$ 型，所以主要讨论如何求这两种类型的极限。

定理

洛必达 (L'Hospital) 法则 设函数 $f(x),g(x)$ 在 $(a,a+d]$ 上可导（d 是某个正常数），且 $g'(x)\neq 0$。若此时有

$$\lim_{x\to a^+}f(x)=\lim_{x\to a^+}g(x)=0$$

或

$$\lim_{x\to a^+}g(x)=\infty$$

且 $\lim_{x\to a^+}\frac{f'(x)}{g'(x)}$ 存在（可以是有限数或 ∞），则成立

$$\lim_{x\to a^+}\frac{f(x)}{g(x)}=\lim_{x\to a^+}\frac{f'(x)}{g'(x)}$$

♠

证明：这里仅证明 $\lim_{x\to a^+}\frac{f'(x)}{g'(x)}=A$，$A$ 为有限数的情况；

先证明 $\lim_{x\to a^+}f(x)=\lim_{x\to a^+}g(x)=0$ 的情况。

由于函数在 $x=a$ 处的值与 $x\to a^+$ 时的极限无关，因此可以补充定义

$$f(a)=g(a)=0$$

使得 $f(x)$ 和 $g(x)$ 在 $[a,a+d]$ 上连续。这样经过补充定义后的函数 $f(x)$ 和 $g(x)$ 在 $[a,a+d]$ 上满足柯西中值的条件，因而对于任意 $x\in(a,a+d)$，存在 $\xi\in(a,a+d)$，满足

$$\frac{f(x)}{g(x)}=\frac{f(x)-f(a)}{g(x)-g(a)}=\frac{f'(\xi)}{g'(\xi)}$$

当 $x \to a^+$ 时显然有 $\xi \to a^+$。由于 $\lim_{x\to a^+}\frac{f'(\xi)}{g'(\xi)}$ 存在，两端令 $x \to a^+$，即有

$$\lim_{x\to a^+}\frac{f(x)}{g(x)} = \lim_{\xi\to a^+}\frac{f'(\xi)}{g'(\xi)} = \lim_{x\to a^+}\frac{f'(x)}{g'(x)}$$

下面证明 $\lim_{x\to a^+} g(x) = \infty$ 时的情况。

记 x_0 是 $(a, a+d]$ 中任意一个固定点，则当 $x \neq x_0$ 时，$\frac{f(x)}{g(x)}$ 可以改写为

$$\begin{aligned}\frac{f(x)}{g(x)} &= \frac{f(x)-f(x_0)}{g(x)} + \frac{f(x_0)}{g(x)}\\ &= \frac{g(x)-g(x_0)}{g(x)}\cdot\frac{f(x)-f(x_0)}{g(x)-g(x_0)} + \frac{f(x_0)}{g(x)}\\ &= \left[1-\frac{g(x_0)}{g(x)}\right]\frac{f(x)-f(x_0)}{g(x)-g(x_0)} + \frac{f(x_0)}{g(x)}\end{aligned}$$

于是，

$$\begin{aligned}\left|\frac{f(x)}{g(x)} - A\right| &= \left|\left[1-\frac{g(x_0)}{g(x)}\right]\frac{f(x)-f(x_0)}{g(x)-g(x_0)} + \frac{f(x_0)}{g(x)} - A\right|\\ &\leqslant \left|1-\frac{g(x_0)}{g(x)}\right|\cdot\left|\frac{f(x)-f(x_0)}{g(x)-g(x_0)} - A\right| + \left|\frac{f(x_0)-Ag(x_0)}{g(x)}\right|\end{aligned}$$

因为 $\lim_{x\to a^+}\frac{f'(x)}{g'(x)} = A$，所以对于任意 $\varepsilon > 0$，存在 $\rho > 0(\rho < d)$，当 $0 < x-a < \rho$ 时，

$$\left|\frac{f'(x)}{g'(x)} - A\right| < \varepsilon$$

取 $x_0 = a+\rho$，有柯西中值定理，对于任意的 $x \in (a, x_0)$，存在 $\xi \in (x, x_0) \subset (a, a+\rho)$ 满足：

$$\frac{f(x)-f(x_0)}{g(x)-g(x_0)} = \frac{f'(\xi)}{g'(\xi)}$$

于是得到

$$\left|\frac{f(x)-f(x_0)}{g(x)-g(x_0)} - A\right| = \left|\frac{f'(\xi)}{g'(\xi)} - A\right| < \varepsilon$$

又因为 $\lim_{x\to a^+} g(x) = \infty$，所以可以找到正数 $\delta < \rho$，当 $0 < x-a < \delta$ 时，成立

$$\left|1-\frac{g(x_0)}{g(x)}\right| < 2,\quad \left|\frac{f(x_0)-Ag(x_0)}{g(x)}\right| < \varepsilon$$

综上所述，即知对于任意 $\varepsilon > 0$，存在 $\delta > 0$，当 $0 < x-a < \delta$ 时，

$$\begin{aligned}\left|\frac{f(x)}{g(x)} - A\right| &\leqslant \left|1-\frac{g(x_0)}{g(x)}\right|\cdot\left|\frac{f(x)-f(x_0)}{g(x)-g(x_0)} - A\right| + \left|\frac{f(x_0)-Ag(x_0)}{g(x)}\right|\\ &< 2\varepsilon + \varepsilon = 3\varepsilon\end{aligned}$$

由定义即得

$$\lim_{x\to a^+}\frac{f(x)}{g(x)} = A = \lim_{x\to a^+}\frac{f'(x)}{g'(x)}$$

□

§12.6.3　泰勒公式以及麦克劳林公式

本节研究的是用一个多项式近似替代一个复杂的函数，主要讨论泰勒公式; 当 $f(x)$ 在 x_0 处 n 阶可导时，有如下较精确的估计：

定理

带皮亚诺 (Peano) 余弦的泰勒 (Taylor) 公式 设 $f(x)$ 在 x_0 处有 n 阶导数，则存在 x_0 的一个邻域，对于该邻域中的任一点 x，成立

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+r_n(x)$$

其中余项 $r_n(x)$ 满足：

$$r_n(x)=o((x-x_0)^2)$$

上述公式被称为 $f(x)$ 在 $x=x_0$ 处的**带 Peano 余项的 Taylor 公式**，它的前 $n+1$ 项的组成的多项式

$$p_n(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$$

称为 $f(x)$ 的 n 次 **Taylor 多项式**，余项 $r_n(x)=o((x-x_0)^n)$ 称为 **Peano 余项**

♠

证明：考虑 $r_n(x)=f(x)-\sum_{k=1}^{n}\frac{1}{k!}f^{(k)}(x_0)(x-x_0)^k$，只需证明 $r_n(x)=o((x-x_0)^n)$。显然

$$r_n(x_0)=r_n'(x_0)=r_n''(x_0)=\cdots=r_n^{(n-1)}(x_0)=0$$

反复应用洛必达法则，可得

$$\begin{aligned}
&\lim_{x\to x_0}\frac{r_n(x)}{(x-x_0)^n}=\lim_{x\to x_0}\frac{r_n'(x)}{n(x-x_0)^{n-1}}=\lim_{x\to x_0}\frac{r_n''(x)}{n(n-1)(x-x_0)^{n-2}}=\\
=&\lim_{x\to x_0}\frac{r_n^{n-1}(x)}{n(n-1)\cdots2\cdot(x-x_0)}=\frac{1}{n!}\lim_{x\to x_0}\left[\frac{f^{(n-1)}(x)-f^{(n-1)}(x_0)-f^{(n)}(x_0)(x-x_0)}{x-x_0}\right]\\
=&\frac{1}{n!}\lim_{x\to x_0}\left[\frac{f^{(n-1)}(x)-f^{(n-1)}(x_0)}{x-x_0}-f^{(n)}(x_0)\right]=\frac{1}{n!}\left[f^{(n)}(x_0)-f^{(n)}(x_0)\right]=0
\end{aligned}$$

因此

$$r_n(x)=o((x-x_0)^n)$$

□

另一种常见的余项由下列定理给出，它是余项的一个定量化的形式。

定理

带 Lagrange 余项的 Taylor 公式 设 $f(x)$ 在 $[a,b]$ 上具有 n 阶连续导数，且在 (a,b) 上有 $n+1$

阶导数。设 $x_0 \in [a,b]$ 为一定点，则对于任意 $x \in [a,b]$，成立

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + r_n(x)$$

其中余项 $r_n(x)$ 满足：

$$r_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1}$$

ξ 在 x 和 x_0 之间，上述公式称为 $f(x)$ 在 $x=x_0$ 处的**带 Lagrange 余项的 Taylor 公式**。余项称为拉格朗日余项。

证明本书不给出；显然，当 $x \to x_0$ 时，带拉格朗日余项的泰勒公式涵盖了皮亚诺余项的泰勒公式，即前者的结论强于后者，但采用皮亚诺余项的泰勒公式时，对 $f(x)$ 的要求也比采用拉格朗日余项的泰勒公式时稍弱一些。

现在我们讨论函数在 $x=0$ 处的泰勒公式，$f(x)$ 在 $x=0$ 处的泰勒公式：

$$f(x) = f(0) + f'(0)(x-0) + \frac{f''(0)}{2!}(x-0)^2 + \cdots + \frac{f^{(n)}(0)}{n!}(x-0)^n + r_n(x)$$

其中 $r_n(x)$ 由皮亚诺余项的与拉格朗日余项两种表示形式，既有 $r_n(x) = o(x^n)$，或 $r_n(x) = \frac{f^{(n+1)}(\theta x)}{(n+1)!}x^{n+1}, \theta \in (0,1)$。函数 $f(x)$ 在 $x=0$ 处的泰勒公式又称为函数 $f(x)$ 的 **Maclaurin 公式**（麦克劳林公式），以下我们来推导几个基本初等函数在 $x=0$ 处的泰勒公式。

求 $f(x) = \mathrm{e}^x$ 在 $x=0$ 处的泰勒公式：对于 $f(x) = \mathrm{e}^x$ 有

$$f(x) = f'(x) = f''(x) = \cdots = f^{(n)}(x) = \mathrm{e}^x$$

于是得到

$$f(0) = f'(0) = f''(0) = \cdots = f^{(n)}(0) = 1$$

因此，得到 e^x 在 $x=0$ 处的泰勒公式：

$$\mathrm{e}^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + r_n(x)$$

它的皮亚诺余项和拉格朗日余项分别为 $r_n(x) = o(x^n), r_n(x) = \frac{\mathrm{e}^{\theta x}}{(n+1)!}x^{n+1}$，$\theta \in (0,1)$。

引理

设函数 $f(x)$ 在 x_0 的某个邻域有 $n+2$ 阶导数存在，则它的 $n+1$ 次多项式的导数恰为 $f'(x)$ 的 n 次泰勒多项式。♣

这个引理不做证明，通过引理可以求得 $\ln(1+x)$ 在 $x=0$ 处的泰勒公式。

§12.7 一些高考题和一些方法

对于高考，因为高中学的内容（至少课本上）是相当少的，高考的很多题有一些固定套路，这一节我们来讨论，分为“一些题”以及“一些处理方法”两个小节来说，不过处理方法和题型不能完全分开，有一些题型

本身就代表一个方法。

§12.7.1 隐零点

隐零点问题，又叫虚设零点、设而不求、零点代换等；
在很多问题中，常常需要我们求一个函数的极值，我们可能会遇到这样一种情况：对原函数求导后，因为含参或是超越方程，又或是两个都有的情况，导致导函数的零点无法直接求得，进而无法得到极值点，得到极值。**这种不能直接解出零点的题目称为隐零点问题**，那应该如何处理呢？通常我们用关于导函数零点 x_0 的表达式带回原函数，消除参数，将一个二元问题转换为一元问题去研究，然后再对改造后的原函数进行下一步的操作（依照题目的问题），

因此隐零点问题通常会用到多种技巧，设而不求只是其中的一步而已，比较套路的处理方式。

此外，需注意零点个数以及零点存在问题，只需求导判断单调性和用零点存在定理即可处理这两个问题。

e.g. 证明 $\mathrm{e}^x - \ln x > 2$
令：$f(x) = \mathrm{e}^x - \ln x$，求导得

$$f'(x) = \mathrm{e}^x - \frac{1}{x}$$

显然 x_0 在满足：$\mathrm{e}^{x_0} = \frac{1}{x_0}$ 时有极值，注意到 $x_0 > 0$，则

$$f''(x_0) = \mathrm{e}^{x_0} + \frac{1}{x_0^2} = \frac{x_0 + 1}{x_0^2} > 0$$

得到 $f(x_0)$ 为函数极小值，由于 $f'(x_0) = 0$ 是超越方程解不出精确解，因此带入推导出的 $\mathrm{e}^{x_0} = \frac{1}{x_0}$ 到 $f(x)$ 中，可以得到：

$$f(x_0) = \mathrm{e}^{x_0} - \ln x_0 = \frac{1}{x_0} + x_0 \geqslant 2$$

当且仅当 $x_0 = 1$ 时不等式取等号，显然 $x_0 \neq 1$（$\mathrm{e}^1 \neq \frac{1}{1}$）因此不等式不能取等。

□

这个问题通过放缩也能求证；证明过程中，在二阶导判断极值那一步就使用了零点代换的条件。

§12.7.2 一些同构

[①]

朗博同构

朗博同构就是通过对数和指数的转换，找到相同结构后换元的类型，例如：

e.g. 求函数最小值

$$f(x) = x\mathrm{e}^x - x - \ln x$$

注意到：

$$x = \mathrm{e}^{\ln x} \tag{1}$$

[①]为了避免歧义，这里说的同构指得是同一结构，与抽象代数中的同构无关

因此构造：$u = x + \ln x$

$$g(u) = f(x) = \mathrm{e}^u - u$$

最后放缩可得 $g(u) \geqslant 1$。

请注意 (1) 式，这是朗博同构的核心。

§12.7.3 极值点偏移

首先介绍极值点偏移的背景，“极值点偏移”，顾名思义就是函数非对称所产生的性质，为了了解极值点偏移，我们先来看看极值点不偏移的情况：

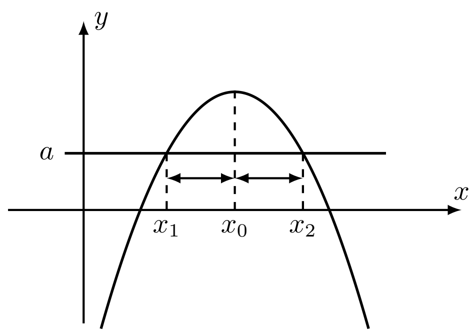


图 12.2: 极值点不偏移

如图所示，上图为一个二次函数，有极大值点 x_0，设 $x_1 \neq x_2$，因二次函数具有对称性，我们知道 x_1, x_2 在 x_0 的两侧，因此对于任意：

$$f(x_1) = f(x_2) = a < f(x_0)$$

都有 $x_1 + x_2 = 2x_0$，这一点从图像来看是显然的，也就是说 x_0 是 x_1, x_2 的中点。

然而对于更一般的函数，并不一定有对称性，因此导致**极值点偏移**，可分为左偏或右偏，如下图

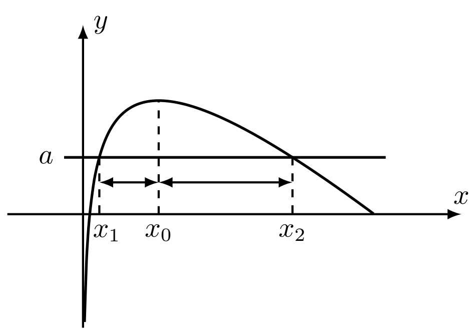


图 12.3: 极值点左偏

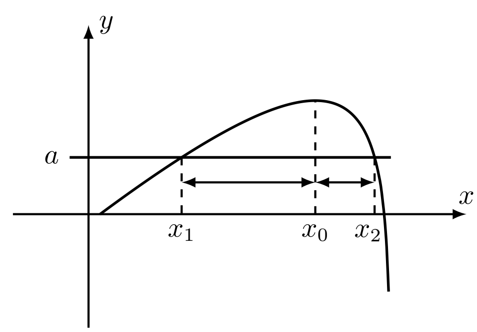


图 12.4: 极值点右偏

由图立马可以得出 $x_1 + x_2 > 2x_0$ 或 $x_1 + x_2 < 2x_0$；

了解了极值点偏移的基本定义后，我们看看以极值点偏移为背景的题目，这一类题目通常是求出极值点后，要求证明 $x_1 + x_2 > 2x_0$ 或 $x_1 + x_2 < 2x_0$，我先用一道经典的例题来讲解最基本的两个方法

e.g. 已知 $f(x) = \frac{x}{\mathrm{e}^x}$ 若 $x_1 \neq x_2$ 且 $f(x_1) = f(x_2)$，证明：$x_1 + x_2 > 2$

首先判断函数 $f(x)$ 的基本性质，通过求导可知：

$$f'(x) = \frac{1-x}{e^x}$$

可知函数 $f(x)$ 在 $x=1$ 处有极小值，即证极值点左偏；

法 1：**构造差函数**

在极值点偏移的背景中我们已经知道了，偏移是由于不对称导致的，因此构造一个关于极值点对称的函数，与原函数做差，这就意为这把本在极值点 x_0 两侧的 x_1, x_2 放到了同侧比较，具体操作如下：构造关于极值点对称的函数

$$g(x) = f(2-x)$$

（这一步看不懂的可以去函数对称性那节看看），让它与原函数做差，构造一个新函数（事实上可以一步到位直接构造一下差函数）

$$F(x) = f(x) - f(2-x) = \frac{x + (x-2)e^{2x-2}}{e^x}$$

不妨设 $x_2 > x_1$，易知 $x_2 > 1 > x_1$，对 $F(x)$ 求导判断单调性，不过由于分母 e^x 恒正，因此只求导分子部分：

$$h(x) = x + (x-2)e^{2x-2}$$
$$h'(x) = (2x-3)e^{2x-2} + 1$$
$$h''(x) = 4(x-1)e^{2x-2}$$

易知，$h'(x)$ 在 $x = 1$ 有极小值，又 $h'(1) = 0$，因此 $h'(x) \geqslant h'(1) = 0$，$h(x)$ 在定义域单调递增，注意到 $h(1) = 0$，因此当 $x > 1$ 恒有：

$$f(x) > f(2-x)$$

由于 $x_2 > 1$，带入得

$$f(x_1) = f(x_2) > f(2-x_2)$$

由于 $x_1, 2-x_2 < 1$，且 $f(x)$ 在 $(-\infty, 0)$ 上单调递减，因此有

$$x_1 < 2 - x_2$$
$$x_1 + x_2 < 2$$

□

法 2：**ALG**

已知

$$f(x_1) = f(x_2)$$
$$\frac{x_1}{e^{x_1}} = \frac{x_2}{e^{x_2}}$$

两边取对数得

$$\begin{aligned}\ln x_1 - x_1 &= \ln x_2 - x_2 \\ \ln x_1 - \ln x_2 &= x_1 - x_2\end{aligned}$$

注意到：

$$\lim_{x\to+\infty}\frac{x}{\mathrm{e}^x}=\lim_{x\to+\infty}\frac{1}{\mathrm{e}^x}=0$$

且 $f(0)=0$，可知 $0<x_1<1<x_2$，故可以用 ALG

$$\begin{aligned}\frac{x_1+x_2}{2}>\frac{x_1-x_2}{\ln x_1-\ln x_2}&=1\\ x_1+x_2&>2\end{aligned}$$

□

可以看出，用对数均值不等式时非常的无脑，直接带进去即可，不像法 1 一样需要多次求导判断，不过在高考中对数均值不等式需要证明，也就是说需要构造函数去证明，总的来说是相对思考量小的。

下面介绍一个偏移引理，引理以及证明来自知乎用户虚调子，这是极值点偏移的充分条件

引理

偏移引理：若连续光滑函数 $f(x)$ 在某段开区间 A 上只有一个极大值点 x_0（A 包含于 $f(x)$ 的收敛域内），使得极值点左侧单调增加，右侧单调减少，且存在 $x_1<x_0<x_2\in A, f(x_1)=f(x_2)$，如果 $f(x)$ 的**全体奇数阶导**在极大值点处的值非负（但也不全为 0，下同）即

$$f^{2k+1}(x_0)\geqslant 0, k\in\mathbb{N}$$

则恒有 $x_1+x_2>2x_0$；如果 $f(x)$ 的**全体奇数阶导**在极大值点处的值非正，也即：

$$f^{2k+1}(x_0)\leqslant 0, k\in\mathbb{N}$$

则恒有 $x_1+x_2<2x_0$。

如果 x_0 为极小值点，那么结论变号即可。♣

证明：构造

$$G(x)=f(x)-f(2x_0-x)$$

考虑 $G(x)$ 在极值点处的泰勒展开

$$G(x)=2\sum_{k=0}^{\infty}\frac{f^{2k+1}(x_0)}{(2k+1)!}(x-x_0)^{2k+1}$$

若 $f^{2k+1}(x_0) \geqslant 0$，取 $x = x_1 < x_0$

$$\begin{aligned} G(x_1) < 0 &\iff f(x_1) < f(2x_0 - x_1) \\ &\iff f(x_2) < f(2x_0 - x_1) \\ &\iff x_2 > 2x_0 - x_1 \\ &\iff x_1 + x_2 > 2x_0 \end{aligned}$$

另一种情况同理可证。

□

这个引理也侧面印证了某些人所说的，“函数三阶导控制极值点偏移”。

§12.7.4 变换主元

对于<u>含参证明恒成立</u>，且知道参数范围的情况下通常可以用变换主元，变换主元其实就是将变量和参数的位置调换，以求化简式子，证明恒成立。下面举个例子：

e.g. 已知函数 $f(x) = \frac{ax^2+x-1}{\mathrm{e}^x}$，证明：当 $a \geqslant 1$ 时，$f(x) + \mathrm{e} \geqslant 0$

先简单变换一下：

$$\begin{aligned} f(x) + \mathrm{e} \geqslant 0 \\ ax^2 + x - 1 + \mathrm{e}^{x+1} \geqslant 0 \end{aligned}$$

这时构造有关 x 函数 $h(x)$ 不如构造有关 a 的函数 $g(a)$（因为 $h(x)$ 有二次有一次又有超越方程的），构造 $g(a) = ax^2 + x - 1 + \mathrm{e}^{x+1}$，这时不难发现 $g(a)$ 是一个一次函数，有最小值：

$$g_{\min}(a) = g(1) = x^2 + x - 1 + \mathrm{e}^{x+1}$$

这时只剩下一个变量了，通过放缩即可证明

$$x^2 + x - 1 + \mathrm{e}^{x+1} \geqslant x^2 + 2x + 1 \geqslant 0$$

当 $a = 1, x = -1$ 时取等，当 $x = 0$ 时 $g(a) = \mathrm{e} - 1 \geqslant 0$ 恒成立。

□

从上例可看出，变换主元的目的是化简原式以及消参。

§12.7.5 必要性探路

必要性探路，主要用在有关含参且已知恒等式时，求参数范围的题目；先探查参数取值范围，但这只是必要性，探查出参数取值范围后，只需要证明充分性即可；

必要性探路往往和端点效应一起出现，先来介绍端点效应，假设连续函数 $f(x)$，在 $x \in (m, +\infty)$（左边开闭都可以）上满足恒等式 $f(x) \geqslant 0$，且 $f(m) = 0$，这时如果 $f'(x) > 0$，即可以推出 $f(m)$ 是最小值，进而得到参数的取值范围；其实只要满足 $f(x)$ 在区间 $[m, +\infty)$ 单调（只可能是增）；即可推出 $f(m)$ 是最小值。

以上讨论的情况是最为理想的，往往日常遇到的问题要复杂的多，因此端点效应在实际解决的过程中分为好几种，接下去会讨论，首先讨论最常见的端点效应：

端点效应

设函数 $f(x)$ 中含参数 a，恒成立 $f(x) \geqslant 0$，且 $x \geqslant m$（或 $x > m$），且端点值有 $f(m) = 0$ 若 $f'(x)$ 在定义域内单调且含参，这时只要使 $f'(m) \geqslant 0$ 可求出参数范围；

倘若 $f'(x)$ 含参不单调，那就需要求 $f''(x)$ 来判断，如果 $f''(x)$ 含参单调，那么只需令 $f''(m) \geqslant 0$ 即可求参。总的来说，当 $f(x)$ 某一阶导数是**单调且含参**时，就可以通过判断单调性来求参。可以结合下图理解。

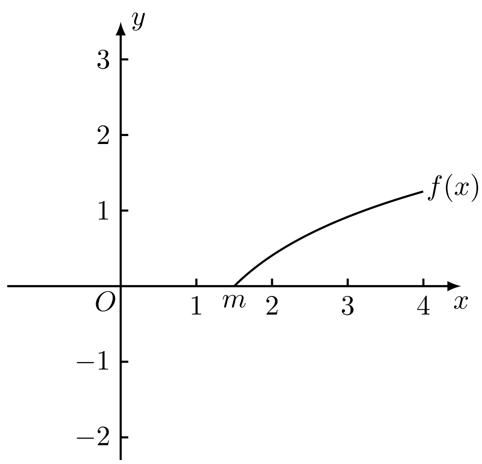


图 12.5: 端点效应

e.g. 设 $f(x) = (1 - x^2)\mathrm{e}^x$，当 $x \geqslant 0$ 时 $f(x) \leqslant 1 + ax$ 恒成立，求 a 的取值范围：

首先把不等式移到一边构造一个新函数：

$$1 + ax - (1 - x^2)\mathrm{e}^x \geqslant 0$$

构造 $g(x) = 1 + ax - (1 - x^2)\mathrm{e}^x$，注意到 $g(0) = 0$，对 $g(x)$ 求导

$$g'(x) = a + (x^2 + 2x - 1)\mathrm{e}^x$$

这时无法判断直接 $g'(x)$ 的单调性，因此求二阶导

$$g''(x) = (x^2 + 4x + 1)\mathrm{e}^x$$

不难知道，当 $x \geqslant 0$ 时 $g''(x) > 0$，即 $g(x)$ 单调递增；易知 $g'(0) \geqslant 0$，即可求出参数范围

$$g'(0) = a - 1 \geqslant 0$$

这时我们就可以分类讨论了：

1. 当 $a \geqslant 1$ 时；

$$g'(x) \geqslant g'(0) \geqslant 0$$

因此 $g(x)$ 单调递增，即

$$g(x) \geqslant g(0) = 0$$

成立。

2. 当 $a<1$ 时；

$$g'(0)<0$$

又 $g'(x)$ 单调递增，则存在 $x_0>0$ 使得 $g'(x_0)=0$，说明 $g(x)$ 在 $x\in(0,x_0)$ 上单调递减；因为 $g(0)=0$，所以存在 $x_1>0$ 且 $g(x_1)<0$，与恒等式矛盾。

事实上，我们可以通过函数极限的局部保号性（或叫保序性）来从代数方面理解端点效应。

定理

局部保号性：若 $\lim_{x\to x_0}f(x)=A,\lim_{x\to x_0}g(x)=B$，且 $A>B$，则存在 $\delta>0$，当 $0<|x-x_0|<\delta$ 时，成立：

$$f(x)>g(x)$$

♠

证明：取 $\varepsilon_0=\frac{A-B}{2}>0$。由 $\lim_{x\to x_0}f(x)=A,\exists\delta_1>0,\forall x(0<|x-x_0|<\delta_1):$

$$|f(x)-A|<\varepsilon_0$$

从而 $\frac{A+B}{2}<f(x)$；由 $\lim_{x\to x_0}g(x)=B,\exists\delta_2>0,\forall x(0<|x-x_0|<\delta_2):$

$$|g(x)-B|<\varepsilon_0$$

从而 $g(x)<\frac{A+B}{2}$，取 $\delta=\min\{\delta_1,\delta_2\}$，当 $0<|x-x_0|<\delta$，成立

$$g(x)<\frac{A+B}{2}<f(x)$$

□

推论

若 $\lim_{x\to x_0}f(x)=A\neq 0$，则存在 $\delta>0$，当 $0<|x-x_0|<\delta$ 时，成立

$$|f(x)|>\frac{|A|}{2}>0$$

◇

证明：由 $\lim_{x\to x_0}f(x)=A$ 以及

$$||f(x)|-|A||\leqslant|f(x)-A|$$

可知 $\lim_{x\to x_0}|f(x)|=|A|$。令 $g(x)=\frac{|A|}{2}$，由 $\frac{|A|}{2}<|A|$ 以及局部保号性可知存在 $\lambda>0$，当 $0<|x-x_0|<\delta$ 时，成立

$$|f(x)|>\frac{|A|}{2}>0$$

□

内点效应

内点效应，又称端点效应失效，其实结合图像容易理解；

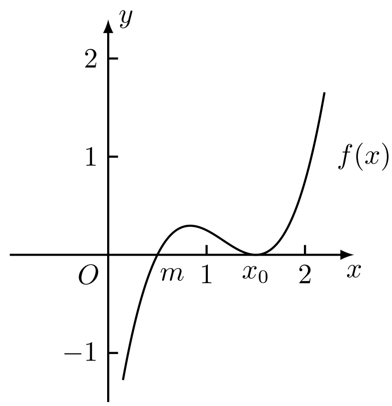


图 12.6: 内点效应

当函数 $f(x)$ 在 $x \geqslant m$ 上恒成立 $f(x) \geqslant 0$，且 $f(m)=0$（这个条件不必要），如果存在一点 $x_0 > m, f'(x_0) = 0$，意味着 $f(x)$ 在 $x \geqslant m$ 上不是单调的，这个极小值点 x_0 称为内点，这时不能按照端点效应的情况来讨论，而是要使 $f(x_0) \geqslant 0$；当然若函数 $f(x)$ 在定义域内不单调，也不一定有极小值，例如

$$\lim_{x\to+\infty} \frac{1}{x} + a = a$$

这时参数 a 就仅仅是位置参数；大多数情况我们都会遇到有极小值点 x_0 的情况；

解决内点问题，最常用的方法就是分离参数，通过 $f(x_0) \geqslant 0$ 进行参数分离，然后找到极小值点（内点 x_0）。

e.g. 已知不等式 $\mathrm{e}^x - 1 \geqslant kx + \ln x$ 在 $x \in (0, +\infty)$ 恒成立，求 k 的最大值。

先来分析，端点 $x=0$ 因为定义域所以取不到，所以分离参数来解决：原式等价于

$$\frac{\mathrm{e}^x - 1 - \ln x}{x} \geqslant k$$

构造函数并求导

$$g(x) = \frac{\mathrm{e}^x - 1 - \ln x}{x}$$
$$g'(x) = \frac{(x-1)\mathrm{e}^x + \ln x}{x^2}$$

这里可以看出 $g'(1) = 0$，但为了探究 $g'(x)$ 的单调性以及零点是否唯一，构造函数 $h(x)$ 并求导

$$h(x) = (x-1)\mathrm{e}^x + \ln x$$
$$h'(x) = \frac{x^2\mathrm{e}^x + 1}{x}$$

因为 $2x, h'(x)$ 在 $x > 0$ 上恒大于零，因此 $g'(x)$ 有唯一零点，即是 $g(x)$ 的极小值点，故

$$k \leqslant g(1) = \mathrm{e} - 1$$

综上可得 k 的最大值为 $\mathrm{e} - 1$。

§12.7.6 放缩

广义上的放缩，本质就是利用不等式的传递性[2]，但本节所说的放缩是一种化归的技巧，通过麦克劳林公式去余项得到一些简单的不等式，再通过换元变形得到一些复杂的形式；

高中没学麦克劳林公式，因此我们通常会化归为两个最基本的不等式，通过简单的求导可证。

我们将 e^x 的麦克劳林公式只保留前两项得到

$$\mathrm{e}^x \geqslant x+1 \tag{1}$$

这时如果令 $x=\ln x$ 可以得到

$$\ln x \leqslant x-1 \tag{2}$$

以上两个不等式是最基本的两个，有人称为母不等式，通过求导即可证明。

倘若将 $x-1=x$ 的代换带入 (1) 式即可得

$$\mathrm{e}^x \geqslant \mathrm{e}x$$

若将 $\frac{1}{x}=x$ 的代换带入 (2) 式即可得

$$\ln x \geqslant 1-\frac{1}{x}$$

这种代换还可以继续，但比较常用的也就以上几个，以及麦克劳林多展开一些：

$$\mathrm{e}^x \geqslant 1+x+\frac{x^2}{2!}$$

等等诸如此类的放缩；

飘带放缩

飘带放缩是比较典型的模型，作图可以直观的感受，如图：

在 $x>1$ 时，有：

$$h(x)=\frac{1}{2}\left(x-\frac{1}{x}\right)>f(x)=\ln x>g(x)=\frac{2(x-1)}{x+1}$$

而在 $x\in(0,1)$ 时不等式取反向，即

$$h(x)<f(x)<g(x)$$

不难看出 $h(x)$ 就是飘带函数；以上飘带放缩的结论只需移向作差构造新函数求导即可证明，令

$$F(x)=h(x)-f(x)=\frac{1}{2}(x-\frac{1}{x})-\ln x$$

求导得

$$\begin{aligned}F'(x)&=\frac{1}{2}+\frac{1}{2x^2}-\frac{1}{x}\\&=\frac{(x-1)^2}{2x^2}\end{aligned}$$

---

[2] $a>b, b>c$ 则 $a>c$

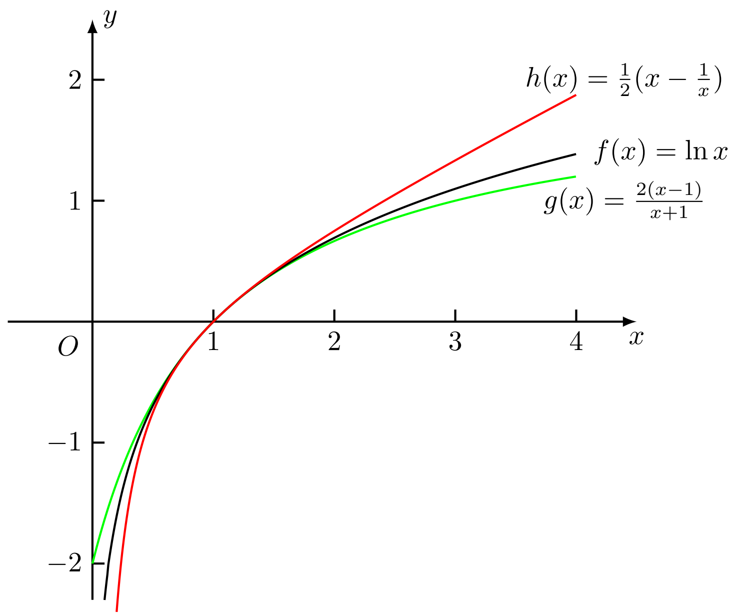


图 12.7: 飘带放缩

可知，$F'(x)$ 在 $x>0$ 时恒大于 0，即原函数 $F(x)$ 在 $x>0$ 上单调递增；注意到 $F(1)=0$，因此当 $x>1$ 时 $F(x)>0$ 即

$$h(x)>f(x)$$

当 $x\in(0,1)$ 时，有 $F(x)<0$，即

$$h(x)<f(x)$$

当且仅当 $x=1$ 时，有 $h(x)=f(x)$。

飘带放缩最常见的用法是用来证明对数均值不等式，这也是高中最常用来证明对数均值不等式的方法，如下：

飘带放缩证明 ALG 证明：

$$\frac{a+b}{2}>\frac{a-b}{\ln a-\ln b}>\sqrt{ab}$$

先证左边，设 $a,b>0$ 且 $a>b$，要证明 $\frac{a+b}{2}>\frac{a-b}{\ln a-\ln b}$ 即证：

$$\ln\frac{a}{b}>2\frac{a-b}{a+b}$$
$$\ln\frac{a}{b}>2\frac{\frac{a}{b}-1}{\frac{a}{b}+1}$$

令 $t=\frac{a}{b},t>1$ 即证：

$$\ln t>2\frac{t-1}{t+1}$$

作差构造函数并求导

$$f(t) = \ln t - 2\frac{t-1}{t+1}$$
$$f'(t) = \frac{(t-1)^2}{t(t+1)^2}$$

可以看出来当 $t > 0$ 是 $f'(t) > 0$，又因 $f(1) = 0$，即证：当 $t > 1$ 时

$$\ln t > 2\frac{t-1}{t+1}$$
$$\ln\frac{a}{b} > 2\frac{a-b}{a+b}$$
$$\frac{a+b}{2} > \frac{a-b}{\ln a - \ln b}$$

当 $b > a, 0 < \frac{a}{b} < 1$ 时，有

$$\ln t < 2\frac{t-1}{t+1}$$

带入 $t = \frac{a}{b}$，注意到 $t \in (0,1), \ln t < 0$ 可得：

$$\ln\frac{a}{b} < 2\frac{a-b}{a+b}$$
$$\frac{a+b}{2} > \frac{a-b}{\ln a - \ln b}$$

再来证明不等式右边：令 $a > b$ 即证

$$\ln\frac{a}{b} < \frac{a-b}{\sqrt{ab}}$$
$$\ln\frac{a}{b} < \frac{\frac{a}{b}-1}{\sqrt{\frac{a}{b}}}$$

令 $t^2 = \frac{a}{b}$ 即证：

$$\ln t < \frac{t^2-1}{2t}$$

同样的作差构造并求导

$$f(t) = \ln t - \frac{t^2-1}{2t}$$
$$f'(t) = \frac{-t^2+t-1}{t^2}$$

不难知道，$f'(t)$ 在 $t > 0$ 时恒成立 $f'(t) < 0$，又因 $f(1) = 0$，因此可得当 $t > 1$ 时有：

$$f(t) < 0$$
$$\ln t < \frac{t^2-1}{2t}$$
$$\sqrt{ab} < \frac{a-b}{\ln a - \ln b}$$

若令 $0 < a < b$ 可以得到同样的结果；

$\square$

第十三章　微分中值定理

第十四章　不定积分

第十五章　复数

复数这一节基本算高考最简单的一节，但复数本身不简单，在数学竞赛里平面几何的一些证明，以及一些在三维空间定位的四元数都属于复数的拓展，在高中我们会学习一些复数的定义与四则运算，以及三角形式；此外，高中课本上拓展部分也会有，例如棣莫弗定理；本章在定义等内容上摘抄了部分复分析的内容 [?]。

§15.1　复数的代数表示

定义

形如

$$z = x + y\mathrm{i} \tag{1}$$

这样的结合叫做复数，其中 $x, y \in \mathbb{R}$，而 i 被称为虚数单位，定义

$$\mathrm{i}^2 = -1$$

在 (1) 式中，x 被称为实部，y 称为虚部，有记号：

$$\mathrm{Re}\ z = x,\ \mathrm{Im}\ z = y$$

于是 (1) 式可以改写为：

$$z = \mathrm{Re}\ z + \mathrm{i}\mathrm{Im}\ z$$

所有的复数都属于复数集 $\mathbb{C}$，其中：

$$\mathbb{C} = \{a + b\mathrm{i} | a, b \in \mathbb{R}\}$$

复数 $z = x + y\mathrm{i}$ 这一形式称为复数的**代数形式**，当一个复数的虚部为零时称为**实数**，而虚部不为零时称为**虚数**, 特别的，当实部为零，虚部不为零时称为纯虚数。

即对于复数 $z = x + y\mathrm{i}$ 可以进行以下分类：

1. 当 $\mathrm{Im}\ z = 0$ 时，z 是实数；
2. 当 $\mathrm{Im}\ z \neq 0$ 时，z 是虚数；
3. 当 $\mathrm{Im}\ z \neq 0$ 且 $\mathrm{Re}\ z = 0$ 时，z 是纯虚数。

相等，两个复数当且仅当实部和虚部分别相等的时候，才被认为是相等的；即

$$a + b\mathrm{i} = c + d\mathrm{i}\ (a, b, c, d \in \mathbb{R}) \iff a = c, b = d$$

不相等，记号 $\neq$ 在运用于复数时，用作等式的否定，即对于两个复数 z_1, z_2，它们的实部或虚部至少有一个是不像等的；而“$>$”和“$<$”不直接用于虚数。

复平面

复数 $z = x + y\mathrm{i}$ 可以和实数对 (x, y) 一一对应，由解析几何的知识可知 (x, y) 又和坐标平面 Oxy 一一对应，因此复数 $z = x + y\mathrm{i}$ 与坐标平面 Oxy 上的点一一对应，Oxy 表示的坐标平面称为复平面，不难知道实数在直线 Ox 上，纯虚数在直线 Oy 上，因此 x 轴被称为实轴，y 轴被称为虚轴。

复数的四则运算

“与实数一样。”

不难证明加法的各项运算成立

1. 交换律：
$$z_1 + z_2 = z_2 + z_1$$
2. 结合律：
$$(z_1 + z_2) + z_3 = z_1 + (z_2 + z_3)$$

而减法，减去某数就是加上它的的负数。

对于乘法，仍可证明

1. 交换律：
$$z_1 z_2 = z_2 z_1$$
2. 结合律：
$$(z_1 z_2) z_3 = z_1 (z_2 z_3)$$
3. 分配律：
$$z_1 (z_2 + z_3) = z_1 z_2 + z_1 z_3$$

对于除法，将 $z = \frac{z_1}{z_2} = \frac{x+y\mathrm{i}}{a+b\mathrm{i}}$ 转换为 $z = \varphi + \lambda\mathrm{i}$，$\varphi, \lambda \in \mathbb{R}$，只需利用平方差公式

$$(a + b)(a - b) = a^2 - b^2$$

推出

$$\begin{aligned} z &= \frac{(x + y\mathrm{i})(a + b\mathrm{i})}{a^2 + b^2} \\ &= \frac{xa - yb}{a^2 + b^2} + \frac{xb + ya}{a^2 + b^2} \end{aligned}$$

共轭复数：

把实部相等，虚部为相反数的两个复数互为共轭复数，在复数上方打一横表示它的共轭复数；如下

$$z = a + b\mathrm{i}$$
$$\overline{z} = a - b\mathrm{i}$$

不难看出 $\overline{z}$ 和 z 互为共轭复数，当 z 是实数时有 $z = \overline{z}$；此外也不难证明一些其他的性质：

1. 两共轭复数之和为实数

$$z+\overline{z}=2\mathrm{Re}\ z$$

2. 两共轭复数之差为纯虚数

$$z-\overline{z}=2\mathrm{Im}\ z$$

3. 两共轭复数之积为大于零的实数

$$z\overline{z}=x^2+y^2\geqslant 0$$

§15.2　复数的三角表示

设 $z=x+y\mathrm{i}$ 是一不为零的复数，即 x,y 不同时为零，引入极坐标来替代直角坐标，有实变量 θ 和 r 的方程组

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

当 $r\geqslant 0$ 时这个方程是有解的，即得

$$r=\sqrt{x^2+y^2}$$

这里平方根取的是正值，并有

$$\begin{cases} \cos\theta=\frac{x}{\sqrt{x^2+y^2}} \\ \sin\theta=\frac{y}{\sqrt{x^2+y^2}} \end{cases} \tag{2}$$

于是，θ 被除相差 $2k\pi,\ k\in\mathbb{Z}$ 外被唯一确定，这个唯一定义的正数 r 被称为复数 z 的**模**（或绝对值），任意满足方程组 (2) 的 θ 被称为 z 的**辐角**

任意复数都有模长，用绝对值符号来表示：

$$|z|=|a+b\mathrm{i}|=r=\sqrt{a^2+b^2}\,(r\geqslant 0,r\in\mathbb{R})$$

除 0 外，任何复数都有无限多个辐角，当 $z=0$ 辐角就失去意义，记辐角为

$$\theta=\arg z$$

由不等式

$$0\leqslant\theta<2\pi$$

来规定辐角的主值。

若将 $z=x+y\mathrm{i}$ 中的 x,y 改用 θ,r 来表示，则得到

$$z=r\left(\cos\theta+\mathrm{i}\sin\theta\right)$$

上式即为复数的三角形式，模和辐角的几何意义是不难明白的，模是 z 到原点的距离，而辐角则是矢量 $\overrightarrow{OZ}$ 与 Ox 正方向之间的任何一个交角，角的大小按逆时针计算。

如果我们把复数 z 看作一个向量，则不难推出复数间的加减法满足“三角形法则”，我们再来看乘法与除

法，用复数的三角形式来表示：设 $z_1 = r_1(\cos\theta_1 + \mathrm{i}\sin\theta_1), z_2 = r_2(\cos\theta_2 + \mathrm{i}\sin\theta_2)$，则有

$$z_1z_2 = r_1r_2\left(\cos(\theta_1+\theta_2) + \mathrm{i}\sin(\theta_1+\theta_2)\right)$$

证明：

$$\begin{aligned}
z_1z_2 &= r_1\left(\cos\theta_1 + \mathrm{i}\sin\theta_1\right)\cdot r_2\left(\cos\theta_2 + \mathrm{i}\sin\theta_2\right)\\
z_1z_2 &= r_1r_2\left(\cos\theta_1\cos\theta_2 + \mathrm{i}\sin\theta_2\cos\theta_1 + \mathrm{i}\sin\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2\right)\\
z_1z_2 &= r_1r_2\left(\cos(\theta_1+\theta_2) + \mathrm{i}\sin(\theta_1+\theta_2)\right)
\end{aligned}$$

□

复数的三角表示除法有：

$$\frac{z_1}{z_2} = \frac{r_1}{r_2}\left(\cos(\theta_1-\theta_2) + \mathrm{i}\sin(\theta_1-\theta_2)\right)$$

证明：

$$\begin{aligned}
\frac{z_1}{z_2} &= \frac{r_1\left(\cos\theta_1 + \mathrm{i}\sin\theta_1\right)}{r_2\left(\cos\theta_2 + \mathrm{i}\sin\theta_2\right)}\\
\frac{z_1}{z_2} &= \frac{r_1}{r_2}\cdot\frac{\left(\cos\theta_1 + \mathrm{i}\sin\theta_1\right)\left(\cos\theta_2 - \mathrm{i}\sin\theta_2\right)}{\cos^2\theta_2 + \sin^2\theta_2}\\
\frac{z_1}{z_2} &= \frac{r_1}{r_2}\left(\cos\theta_1\cos\theta_2 - \mathrm{i}\sin\theta_2\cos\theta_1 + \mathrm{i}\sin\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\right)\\
\frac{z_1}{z_2} &= \frac{r_1}{r_2}\left(\cos(\theta_1-\theta_2) + \mathrm{i}\sin(\theta_1-\theta_2)\right)
\end{aligned}$$

□

§15.2.1　棣莫弗定理

定理

棣莫弗定理

$$\left(\cos\theta + \mathrm{i}\sin\theta\right)^n = \cos n\theta + \mathrm{i}\sin n\theta,\ \ n = 1, 2, 3, \cdots$$

棣莫弗定理使用复数的指数形式推导非常直观，但下面我们用已学过的三角形式来证明棣莫弗定理。

证明：我们已经知道

$$z_1z_2 = r_1r_2\left(\cos(\theta_1+\theta_2) + \mathrm{i}\sin(\theta_1+\theta_2)\right)$$

根据数学归纳法不难证多个复数乘积的推广：

$$z_1z_2\cdots z_n = r_1r_2\cdots r_n\left(\cos(\theta_1+\theta_2+\cdots+\theta_n) + \mathrm{i}\sin(\theta_1+\theta_2+\cdots+\theta_n)\right)$$

当 $z_1 = z_2 = \cdots = z_n$ 时，有：

$$z_1 z_2 \cdots z_n = r^n (\cos n\theta + \mathrm{i} \sin n\theta)$$
$$(\cos\theta + \mathrm{i}\sin\theta)^n = \cos n\theta + \mathrm{i}\sin n\theta$$

□

棣莫弗定理对于研究倍角的三角恒等式是一个方便的工具，下面举个例子：

把 $\sin 3\theta$ 用 $\cos\theta, \sin\theta$ 来表示：

$$\sin 3\theta = \mathrm{Im}\,(\cos 3\theta + \mathrm{i}\sin 3\theta) = \mathrm{Im}\,(\cos\theta + \mathrm{i}\sin\theta)^3$$

令 $a = \sin\theta, b = \cos\theta$，通过二项式定理可知

$$\begin{aligned}\sin 3\theta &= \mathrm{Im}\,(\cos\theta + \mathrm{i}\sin\theta)^3\\ &= C_3^1 b^2 a - C_3^3 a^3\\ &= 3\cos^2\theta\sin\theta - \sin^3\theta\end{aligned}$$

这时已经满足常用时的需求了，但通过三角函数的恒等式可以继续化简得

$$\sin 3\theta = 3\sin\theta - 4\sin^3\theta$$

§15.3 复数的指数表示

定义

如果 $z = x + y\mathrm{i}$，则定义 e^z 为复数：

$$\mathrm{e}^{z_1} = \mathrm{e}^x(\cos y + \mathrm{i}\sin y)$$

为何如此定义已经超出本书的范围，在此不做讲解。

由定义，若 $x = 0$，则可以得到著名的欧拉公式：

$$\mathrm{e}^{y\mathrm{i}} = \cos y + \mathrm{i}\sin y \tag{3}$$

不难看出，若 $z_1 = x_1 + y_1\mathrm{i}, z_2 = x_2 + y_2\mathrm{i}$

$$\mathrm{e}^{z_1}\cdot\mathrm{e}^{z_2} = \mathrm{e}^{z_1+z_2}$$
$$\frac{\mathrm{e}^{z_1}}{\mathrm{e}^{z_2}} = \mathrm{e}^{z_1-z_2}$$

若令 (3) 式中的 $y = \theta$，结合复数的三角表示可以得到

$$z = r\mathrm{e}^{\mathrm{i}\theta}$$

上式被称为复数的指数形式，其中 $\theta \in \mathbb{R}, r \geqslant 0$。

§15.4　代数基本定理

定理

代数基本定理 每一个次数大于 0 的复系数多项式至少有一个复根

这个定理的证明超出了本书的范围，在此不做证明，顺带一提这个定理的第一个严格证明是高斯写的博士论文（1799 年）后面他又给出 4 个证明。

定理

复系数多项式唯一因式分解定理 每一个次数大于 0 的复系数多项式在复数域上都可以唯一地分解成一次因式的乘积。

通过上面的定理可推出

推论

每一个 n 次 $(n>0)$ 复系数多项式恰有 n 个复根（重根按重数计算）

§15.5　补充

请读者证明：

二元一次方程 $ax^2+bx+c=0$ 中，如果其中一个根是复数 z，则另一个根是它的共轭复数 $\overline{z}$, 仅在 $\varDelta\leqslant 0$ 时由此结论。

如果对于复数 $z=r(\cos\theta+\mathrm{i}\sin\theta)$，有 $z\neq 1, z^n=1$ 时，说明有 $\sum_{i=0}^{n-1} z^{\mathrm{i}}=0$

第四部分

初等数论

第十六章　整数的可除性

§16.1　整除的基本概念

我们知道，整数集 $\mathbb{Z}$ 对于加法、减法和乘法运算是封闭的，也就是说对于任意两个整数之和、差、积仍旧是整数，而整数对于乘法的逆运算———除法并不封闭，因此我们引入整除的概念。

定义

设 a, b $(b \neq 0)$ 是任意两个整数，若存在一个整数 q，使得下列等式

$$a = bq \tag{1}$$

成立，我们说"b 整除 a"或"a 被 b 整除"，记为 $b|a$，此时把 b 称为 a 的**因数**，把 a 称为 b 的倍数。

如果 1 式不成立，那么称"b 整除 a"或"a 不被 b 整除"，记为 $b \nmid a$。

由整除的定义出发，我们可以得到整除的一些基本性质（定理）。

定理

整除的传递性：若 a 是 b 的倍数，若 b 是 c 的倍数，那么 a 也是 c 的倍数，即

$$b|a, c|b \Rightarrow c|a$$

这个定理是显然的，下面我们用代数来证明一下。

证明：由整除的定义可知

$$a = q_1 b, b = q_2 c$$

其中 $q_1, q_2 \in \mathbb{Z}$，那么有

$$a = q_1 q_2 c$$

由于两个整数之积仍为整数，即 $q_1 q_2$ 为整数，故由定义可知 $c|a$。

□

定理

若 a, b 都是 m 的倍数，则 $a \pm b$ 也是 m 的倍数。

证明：由整除的定义有

$$a = q_1 c, b = q_2 c$$

对 $a \pm b$ 有

$$a \pm b = (q_1 \pm q_2)c$$

由于两个整数之和或差仍为整数，即 $q_1 \pm q_2$ 为整数，故由定义可知 $(a \pm b)|c$。

□

我们可以按照相同的方法将上面的结论推广

推论

若 $a_1, a_2, \ldots, a_n$ 都是 m 的倍数，$b_1, b_2, \ldots, b_n$ 是任意的 n 个整数，则

$$a_1b_1 + a_2b_2 + \cdots + a_nb_n = \sum_{k=1}^{n} a_kb_k$$

上式也是 m 的倍数。

◇

证明：由整除的定义知

$$a_k = q_km,\ k = 1, 2, \ldots, n$$

那么将等式两边乘以 b_k 可得

$$a_kb_k = b_kq_km,\ k = 1, 2, \ldots, n$$

显然任意一组 $b_k \cdot q_k$ 都为整数，那么它们的累加和也为整数，即

$$\sum_{k=1}^{n} a_kb_k = \sum_{k=1}^{n} b_kq_km = m\sum_{k=1}^{n} b_kq_k \in \mathbb{Z}$$

根据定义即知

$$\left(\sum_{k=1}^{n} a_kb_k\right) \mid m$$

□

在实际问题中，我们常常会遇到不可以整除的情况，这时我们可以用表示为 $a = bq + r$，其中 $a, b, q, r \in \mathbb{Z}$，且 $0 < r < b$，可以将这个式子化为更一般的情况，也就是可以表示整除的情况。

定理

若 a, b 是任意的两个整数，其中 $b > 0$，则存在两个整数 q 以及 r 使得

$$a = bq + r,\ 0 \leqslant r < b \quad ②$$

成立，q, r 是唯一的。这就是**带余除法**。

♠

证明：作整数序列

$$\cdots, -2b, -b, 0, b, 2b, \cdots$$

则整数 a 必在上述序列的某相邻两项之间，也就是有

$$qb \leqslant a < (q+1)b,\ q \in \mathbb{Z}$$

记 $r = a - qb$，由于 a, b, q 均为整数，那么 r 也为整数，且根据上式，若均减去 qb 可得

$$0 \leqslant r < b$$

这说明存在整数 q, r 使得 ② 成立。

下面来证明 q, r 的唯一性：设 q', r' 是满足 ② 式的两个整数，即

$$a = bq' + r',\ 0 \leqslant r' < b$$

则有

$$\begin{aligned} bq + r &= bq' + r' \\ b(q - q') &= r - r' \\ b|q - q'| &= |r - r'| \end{aligned}$$

由于上式右边中 r, r' 均是小于 b 的正整数，所有上式右边是小于 b 的，如果 $q \neq q'$，那么上式左边将大于 b，这使得等式两边出现相矛盾的结论，因此 $q = q', r = r'$，即 q, r 是唯一的。

□

由带余除法这个定理，我们可以引出余数的定义。

定义

② 式中的 q 叫做 a 被 b 除所得的**不完全商**，r 叫做 a 被 b 除所得到的**余数**。

在小学我们就已经学过有关余数的定义了，不过在小学阶段没能把除数、被除数、商以及余数用抽象（广义）的表示出来，都是用一些具体的数字来辅助理解，*e.g.*

$$417 \div 15 = 27 \cdots\cdots 12$$

如果用现在的定义可以表示为

$$\begin{gathered} a = bq + r,\ 0 \leqslant r < b \\ 417 = 15 \times 27 + 12,\ 0 < 12 < 15 \end{gathered}$$

§16.2 最大公因数以及辗转相除法

定义

设 $a_1, a_2, \ldots, a_n$ 是 $n (n \geqslant 2)$ 个整数，若 d 是它们之中每一个的因数，那么 d 叫做 $a_1, a_2, \ldots, a_n$ 的**公因数**。

整数 $a_1, a_2, \ldots, a_n$ 的公因数中最大的一个称为**最大公因数**，记作 $(a_1, a_2, \ldots, a_n)$，若

$$(a_1, a_2, \ldots, a_n) = 1$$

那么称 $a_1, a_2, \dots, a_n$ **互素**，若 $a_1, a_2, \dots, a_n$ 每两个整数互素，我们就说它们**两两互素**。

显然，若 $a_1, a_2, \dots, a_n$ 两两互素，那么 $(a_1, a_2, \dots, a_n) = 1$，但反过来就不一定。

注意到 0 的任意整数都是 0 的因数，因此当 $a_1, a_2, \dots, a_n$ 全为 0 时，不存在 $(a_1, a_2, \dots, a_n)$；反过来说，若 $a_1, a_2, \dots, a_n$ 不全为 0，那么 $(a_1, a_2, \dots, a_n)$ 一定存在。

e.g. 已知一组整数为 $-3, -2, 0, 13, 2$，另一组整数为 $3, 2, 0, 13, 2$，它们的最大公因数为

$$(-3, -2, 0, 13, 2) = (3, 2, 0, 13, 2) = 1$$

e.g. 已知一组整数为 $-8, 0, 4, 12, 16$，另一组整数为 $8, 0, -4, -12, 16$，它们公因数均为 $-4, -2, -1, 1, 2, 4$，最大公因数为

$$(-8, 0, 4, 12, 16) = (8, 0, -4, -12, 16) = 4$$

这两个例子中，若对这两组数的每一个整数都加上绝对值符号，可以发现两组数是相同的，且它们的最大公因数相等。不妨猜测对于任意的 n 个整数 $a_1, a_2, \dots, a_n$，其 $a_1, a_2, \dots, a_n$ 与 $|a_1|, |a_2|, \dots, |a_n|$ 有相同的公因数，且最大公因数相等，即

$$(a_1, a_2, \dots, a_n) = (|a_1|, |a_2|, \dots, |a_n|)$$

证明：设 d 为 $a_1, a_2, \dots, a_n$ 的任意一个公因数，根据定义有

$$d \mid a_k \iff a_k = q_k d,\ k = 1, 2, \dots, n$$

将等式两边加上绝对值符号有

$$|a_k| = |q_k d| = q_k d \vee -q_k d,\ k = 1, 2, \dots, n$$

故 $d \mid |a_k|,\ k = 1, 2, \dots, n$，可知 $a_1, a_2, \dots, a_n$ 和 $|a_1|, |a_2|, \dots, |a_n|$ 有相同的公因数。

由于 $a_1, a_2, \dots, a_n$ 和 $|a_1|, |a_2|, \dots, |a_n|$ 有相同的公因数，那么立马可以得到

$$(a_1, a_2, \dots, a_n) = (|a_1|, |a_2|, \dots, |a_n|)$$

□

以上内容可以写成定理形式

定理

对于任意的 n 个整数 $a_1, a_2, \dots, a_n$，有：

1. $a_1, a_2, \dots, a_n$ 与 $|a_1|, |a_2|, \dots, |a_n|$ 有相同的公因数
2. $(a_1, a_2, \dots, a_n) = (|a_1|, |a_2|, \dots, |a_n|)$

♠

这个定理在我们研究最大公约数时，可以避免区别正负数而引起的多余的步骤。

由 $(a_1, a_2, \dots, a_n) = (|a_1|, |a_2|, \dots, |a_n|)$ 可知，我们可以仅就讨论非负数的最大公因数。

定理

若 b 是一正整数，则

1. 0 与 b 的因数就是 b 的公因数，反之 b 的因数是 0 与 b 的公因数。
2. $(0,b)=b$。

这个定理的证明是显然的，在此不多赘述。

由这个定理可以得到，对于任意非零整数 b，都有 $(0,b)=|b|$。

定理

设 a,b,c 是任意三个不全为 0 的整数，且

$$a=qb+c$$

其中 q 是非零整数，则 a,b 与 b,c 有相同的公因数，那么有 $(a,b)=(b,c)$。

证明：设 d 为 a,b 的一个公因数，由定义 $d\mid a$，$d\mid b$，c 可以写为

$$c=a-qb$$

根据之前得到的定理可知，$a-qb$ 为 d 的倍数，不妨设为 $a-qb=pd$，显然

$$c=pd\Rightarrow d\mid c$$

a,b 与 b,c 有相同的公因数，反过来也可以证明 b,c 的任一公因数也是 a 的因数，即 a,b 与 b,c 有相同的公因数；由此立得 $(a,b)=(b,c)$。

如果设 a,c 的任一公因数为 d'，那么有 $d'\mid a,d'\mid c$，且

$$b=\frac{a+c}{q}$$

可发现，不能像前两种情况一样，得到 d' 是 b 的因数，因此没有$(a,b)=(b,c)=(a,c)$。

§16.2.1 辗转相除法

辗转相除法可以用来求任意两个正整数的最大公因数。

设任意正整数 a, b，由带余除法，可以得到一系列等式：

$$\begin{aligned}
a &= bq_1 + r_1,\ 0 < r_1 < b \\
b &= r_1 q_2 + r_2,\ 0 < r_2 < r_1 \\
r_1 &= r_2 q_3 + r_3,\ 0 < r_3 < r_2 \\
&\cdots \\
r_{n-2} &= r_{n-1} q_n + r_n,\ 0 < r_n < r_{n-1} \\
r_{n-1} &= r_n q_{n+1} + r_{n+1},\ r_{n+1} = 0
\end{aligned} \tag{3}$$

每进行一次带余除法，余数至少减少 1，因为不完全商“b”是有限的，因此最多进行 b 次带余除法，总可以得到一个余数为零的等式，上式所指出的计算方法称为**辗转相除法**[1]。

定理

设 a, b 为任意两个正整数，则 (a, b) 是 ③ 式中最后一个不为零的余数，即 $(a, b) = r_n$。♠

证明：由之前得到的定理可以知道

$$(a, b) = (b, r_1) = (r_1, r_2) = \cdots = (r_n, r_{n+1})$$

又 $r_{n+1} = 0$，所有等式的最后一项为 $(r_n, 0) = r_n$，即证 $(a, b) = r_n$。

□

推论

a, b 的公因数与 (a, b) 的因数相同。◇

受到上个定理的启发，我们可以证明以上推论。

证明：设 d 为任意两个正整数 a, b 的任一公因数，注意到 ③ 式中的第一个式子

$$r_1 = a - bq_1$$

故 d 也是 r_1 的因数，以此类推 r_k 均为 d 的倍数，亦即

$$d \mid r_n = d \mid (a, b)$$

□

下面再来证明两个最大公因数的性质。

[1]在西方称为**欧几里得除法**，在《九章算术》中叫“更相减损术”。

定理

设 a, b 是两个不全为零的整数。

1. 若 m 是任一正整数，则
$$(am, bm) = (a, b)m$$
2. 若 δ 是 a, b 的一个公因数，则 $\left(\frac{a}{\delta}, \frac{b}{\delta}\right) = \frac{(a,b)}{|\delta|}$，特别的 $\left(\frac{a}{(a,b)}, \frac{b}{(a,b)}\right) = 1$

证明：对于第一条，由于

$$(am, bm) = (|am|, |bm|) = (|a|m, |b|m),\ (a, b)m = (|a|, |b|)m$$

因此不妨设 a, b 均为非负整数，显然当 a, b 任意一个为零时，等式显然成立，当 $m = 1$ 时，等式也显然成立。

我们将 ③ 式的等式两边乘以 m 得到

$$\begin{aligned} am &= bmq_1 + mr_1,\ 0 < r_1 < b \\ mb &= mr_1q_2 + mr_2,\ 0 < r_2 < r_1 \\ mr_1 &= mr_2q_3 + mr_3,\ 0 < r_3 < r_2 \\ &\cdots \\ mr_{n-2} &= mr_{n-1}q_n + mr_n,\ 0 < r_n < r_{n-1} \\ mr_{n-1} &= mr_nq_{n+1} + mr_{n+1},\ r_{n+1} = 0 \end{aligned} \quad ③$$

由上式可知
$$(am, bm) = (bm, mr_1) = \cdots = (mr_n, mr_{n+1}) = (mr_n, 0) = mr_n = m(a, b)$$

因此第一条得证。

对于第二条，由第一条可知

$$\begin{aligned} &\left(\frac{a}{\delta}, \frac{b}{\delta}\right)|\delta| = \left(\frac{|a|}{|\delta|}|\delta|, \frac{|b|}{|\delta|}|\delta|\right) = (a, b) \\ \Rightarrow &\left(\frac{a}{\delta}, \frac{b}{\delta}\right) = \frac{(a, b)}{|\delta|} \end{aligned}$$

当 $\delta = (a, b)$ 时，显然成立 $\left(\frac{a}{(a,b)}, \frac{b}{(a,b)}\right) = 1$。

□

下面我们来讨论多个整数的最大公因数，设 $a_1, a_2, \ldots, a_n$ 是 n 个整数，令

$$
\begin{aligned}
(a_1, a_2) &= d_2 \\
(d_2, a_3) &= d_3 \\
(d_3, a_4) &= d_4 \\
&\cdots\cdots \\
(d_{n-1}, a_n) &= d_n
\end{aligned}
$$

于是有

定理

设 $a_1, a_2, \ldots, a_n$ 是 n 个整数，则 $(a_1, a_2, \ldots, a_n) = d_n$

证明：由设得

$$d_k \mid d_{n-1}, d_k \mid a_k,\ k = 2, 3, \ldots, n$$

且 $d_2 \mid a_2, d_2 \mid a_1$，故根据整除的传递性

$$d_n \mid a_k,\ k = 1, 2, 3, \ldots, n$$

即 d_n 为 $a_1, a_2, \ldots, a_n$ 的一个公因数；再来设 d 是 $a_1, a_2, \ldots, a_n$ 的任一公因数，那么有 $d \mid a_1, d \mid a_2$，推出 $d \mid d_2$，以此类推可得 $d \mid d_n$，因此

$$d_n \geqslant d$$

即证 d_n 为 $a_1, a_2, \ldots, a_n$ 的最大公因数。

□

§16.3　整数的最小公倍数

设 a, b 是任意两个正整数，则可以得到

$$
\begin{aligned}
a &= bq_1 + r_1,\ 0 < r_1 < b \\
b &= r_1q_2 + r_2,\ 0 < r_2 < r_1 \\
r_1 &= r_2q_3 + r_3,\ 0 < r_3 < r_2 \\
&\cdots \\
r_{n-2} &= r_{n-1}q_n + r_n,\ 0 < r_n < r_{n-1} \\
r_{n-1} &= r_nq_{n+1} + r_{n+1},\ r_{n+1} = 0
\end{aligned}
\tag{1}
$$

则可以得到

定理

若 a, b 是任意两个正整数，则

$$Q_k a - P_k b = (-1)^{k-1} r_k,\ k = 1, 2, \ldots, n \quad ②$$

其中

$$\begin{cases} P_0 = 1, & P_1 = q_1, & P_k = q_k P_{k-1} + P_{k-2} \\ Q_0 = 0, & Q_1 = 1, & Q_k = q_k Q_{k-1} + Q_{k-2} \end{cases} \quad k = 2, 3, \ldots, n \quad ③$$

证明：当 $k = 1$ 时，② 式显然成立，当 $k = 2$ 时，有

$$\begin{aligned} r_2 &= b - r_1 q_2 \\ r_2 &= (q_2 q_1 + 1)b - q_2 q_1 b - r_1 q_2 \\ r_2 &= (q_2 q_1 + 1)b - q_2 a \\ (-1)^1 r_2 &= Q_2 - a P_2 b \end{aligned}$$

故当 $k = 2$ 时上式成立；假设 ②,③ 式在 $n \geqslant 3$ 时成立，则

$$\begin{aligned} (-1)^n r_{n+1} &= (-1)^n (r_{n-1} - r_n q_{n+1}) \\ &= Q_{n-1} a - P_{n-1} b + q_{n+1} Q_n a - q_{n+1} P_n b \\ &= (Q_{n-1} + q_{n+1} Q_n) a - (P_{n-1} + q_{n+1} P_n) b \\ &= Q_{n+1} a - P_{n+1} b \end{aligned}$$

根据数学归纳法，定理为真。

□

由此定理我们可以推出一个重要的推论。

推论

若 a, b 是两个不全为零的整数，则存在整数 s, t 使得

$$as + tb = (a, b)$$

上面这个推论又叫做**裴蜀恒等式**，由定理证明非常简单

$$(a, b) = r_n = (-1)^{1-n} Q_k a + (-1)^{2-n} P_k b$$

□

裴蜀恒等式能推广到 n 个整数的情况，即

推论

若 $a_1, a_2, \ldots, a_n$ 是 n 个不全为零的整数，则存在整数 $x_1, x_2, \ldots, x_n$ 使得

$$\sum_{k=1}^{n} a_k x_k = (a_1, a_2, \ldots, a_n)$$

◇

证明：可以用数学归纳法证明，当 $n = 2$ 时，上式成立，假设当 $n - 1$ 时上式也成立，那么有

$$\sum_{k=1}^{n-1} a_k x_k = (a_1, a_2, \ldots, a_{n-1})$$

记 $d_{n-1} = (a_1, a_2, \ldots, a_{n-1})$，则

$$(d_{n-1}, a_n) = d_n = (a_1, a_2, \ldots, a_n)$$

由裴蜀恒等式可知

$$(d_{n-1}, a_n) = sd_{n-1} + ta_n = s\left(\sum_{k=1}^{n-1} a_k x_k\right) + ta_n$$

亦即存在整数 $x'_1, x'_2, \ldots, x'_n$ 满足

$$(a_1, a_2, \ldots, a_n) = s\left(\sum_{k=1}^{n-1} a_k x_k\right) + ta_n = \sum_{k=1}^{n} a_k x'_k$$

由归纳法即证，对于 $n \geqslant 2$ 上式都成立，顺便一提，当 $n = 1$ 时，定理是显然的。

□

下面来看三个整数间的公因数的性质。

定理

设 a, b, c 是三个整数，且 $(a, c) = 1$，则

1. ab, c 与 b, c 有相同多个因数；

2. $(ab, c) = (b, c)$

上面假定 b, c 至少有一个不为零。

♠

证明：由裴蜀恒等式

$$as + tc = 1$$

其中 s, t 为两个整数，等式两边乘以 b 得

$$(ab)s + c(bt) = b$$

设 d 为 ab, c 的任一公因数，则 $d \mid ab, d \mid c$，结合上式可知 $d \mid b$。[②]。这说明了 ab, c 的任一公因数都是 b, c 的公因数，反之亦然，则第一条得证。

[②]这里用到了之前的定理：若 $m \mid a_k$，则 $m \mid \sum q_k a_k$

再来，由于 b, c 不全为零，那么 (b, c) 是存在的，由第一条可知

$$(ab, c) = (b, c)$$

□

推论

若 $(a, c) = 1, c \mid ab$ 那么 $c \mid b$。

证明：易知

$$|c| = (ab, c) = (b, c)$$

即证 $c \mid b$。

□

推论

设两组整数为 $a_1, a_2, \ldots, a_n$ 以及 $b_1, b_2, \ldots, b_m$，若前一组的任一整数都与后一组的任一整数互素[a]，则 $a_1 a_2 \cdots a_n$ 与 $b_1 b_2 \cdots b_m$ 互素。

[a]若 a, b 互素，则意味着 $(a, b) = 1$

证明：由条件可知

$$(a_i, b_j) = 1,\ i = 1, 2, \ldots, n, j = 1, 2, \ldots, m$$

那么有

$$(a_1 a_2 \cdots a_n, b_j) = (a_2 a_3 \cdots a_n, b_j) = \cdots = (a_n, b_j) = 1,\ j = 1, 2, \ldots, m$$

再次用定理得

$$(a_1 a_2 \cdots a_n, b_1 b_2 \cdots b_m) = (a_1 a_2 \cdots a_n, b_2 b_3 \cdots b_m) = \cdots = (a_1 a_2 \cdots a_n, b_m) = 1$$

即证 $a_1 a_2 \cdots a_n$ 与 $b_1 b_2 \cdots b_m$ 互素。

□

下面我们来看最小公倍数的定义。

定义

设 $a_1, a_2, \ldots, a_n$ 是 $n(n \geqslant 2)$ 个整数，若 d 是这 n 个整数的倍数，则 d 就叫这 n 个整数的一个**公倍数**，在 $a_1, a_2, \ldots, a_n$ 的所有公倍数中最小的正数叫做**最小公倍数**，记作 $[a_1, a_2, \ldots, a_n]$。

注意到 0 不是任何整数的倍数，因此在下面的讨论中只讨论不为零的情况。

设 d 为 $a_k, k = 1, 2, \ldots, n$ 的倍数，则有 $a_k \mid d$，令

$$d = q_k a_k,\ k = 1, 2, \ldots, n$$

将等式两边加上绝对值符号得

$$|d| = |q_k a_k| \Rightarrow d = q_k|a_k| \vee -q_k|a_k|$$

故 $|a_k| \mid d$，这意味着当我们研究一组数的公倍数时，可以直接讨论它们均为正数的情况，研究最小公倍数时也是一样的，即

$$[a_1, a_2, \ldots, a_n] = [|a_1|, |a_2|, \ldots, |a_n|]$$

首先我们研究两个整数的最小公倍数的一些性质。

定理

设 a, b 是任意两个正整数，则

1. a, b 的所有公倍数就是 $[a, b]$ 的所有倍数。
2. a, b 的最小公倍数等于以它们的最大公因数除它们的乘积所得的商，即

$$[a, b] = \frac{ab}{(a, b)}$$

特别的，若 $(a, b) = 1$ 那么 $[a, b] = ab$。

♠

证明：记 m 为 a, b 的任一公倍数，有 $a \mid m, b \mid m$，即

$$m = q_1 a = q_2 b$$

再设

$$a = p_1(a, b), b = p_2(a, b)$$

则有

$$\begin{aligned} p_1 &= \frac{a}{(a, b)}, p_2 = \frac{b}{(a, b)} \\ \Rightarrow (p_1, p_2) &= \left(\frac{a}{(a, b)}, \frac{b}{(a, b)}\right) = 1 \end{aligned}$$

由于 p_1, p_2 互素，又 $p_1 q_1 = p_2 q_2$，那么有 $p_2 \mid p_1 q_1 \Rightarrow p_2 \mid q_1$，设

$$q_1 = p_2 t$$

综上得

$$\begin{aligned} m &= q_1 a \\ m &= t\frac{ab}{(a, b)} \end{aligned}$$

当 $t = 1$ 时得到最小正数，故

$$[a, b] = \frac{ab}{(a, b)}$$

故第二条得证，再由

$$m = q_1 a = q_2 b = [a, b]t$$

即证第一条。

□

现在再来讨论两个以上的整数的最小公倍数，与讨论多个整数的最大公因数时类似，设 $a_1, a_2, \ldots, a_n$ 是 $n(n \geqslant 2)$ 个正整数，令

$$\begin{aligned}
[a_1, a_2] &= m_2 \\
[m_2, a_3] &= m_3 \\
[m_3, a_4] &= m_4 \\
&\cdots\cdots \\
[m_{n-1}, a_n] &= m_n
\end{aligned}$$

于是有

定理

若 $a_1, a_2, \ldots, a_n$ 是 $n(n \geqslant 2)$ 个正整数，则

$$[a_1, a_2, \ldots, a_n] = m_n$$

证明：由条件可知 $m_{k-1} \mid m_k$，$k = 3, 4, \ldots, n$，由整除的传递性可知：$a_k \mid m_n$，$k = 1, 2, \ldots, n$。这意味着 m_n 为 $a_1, a_2, \ldots, a_n$ 的一个公倍数，再来设 m 为 $a_1, a_2, \ldots, a_n$ 的任一公倍数，那么有 $a_k \mid m$，可以推出 $m_k \mid m$，$k = 2, 3, \ldots, n$，因此 $m_n \mid m$，亦即 $m_n \leqslant |m|$，故 m_n 为 $a_1, a_2, \ldots, a_n$ 的最小公倍数。

□

§16.4 算术基本定理

我们知道 1 的正因数只有它本身，任何大于 1 的整数都至少有两个正因数，即 1 与它本身，我们对它们进行分类，有如下定义

定义

一个大于 1 的整数，如果它的正因数只有 1 和它本身，那么称为**素数**（素数又叫质数）；否则就称为**合数**。

素数在数论研究中占很重要的地位，下面我们来讨论一些素数的简单性质。

定理

设 a 是任一大于 1 的整数，则 a 除 1 外的最小正因数 q 为一素数，并且当 a 为合数时，$q \leqslant \sqrt{a}$。

证明：假设 q 不是素数，则 q 存在正因数 $q_1 > 1$ 使得 $q_1 \mid q$，又 $q \mid a$，故根据整除的传递性有 $q_1 \mid a$，也即 q 不是 a 的最小正因数，与假设矛盾，故 q 为素数。

当 a 为合数时，令 $a = qa_1$，因为 a 不是素数，则 $a_1 > 1$，又 q 是最小得到正因数，因此有

$$\begin{aligned} & a_1 \geqslant q > 1 \\ \Rightarrow & a_1 q \geqslant q^2 \\ \Rightarrow & a \geqslant \sqrt{q} \end{aligned}$$

□

定理

若 q 为一素数，a 为一整数，那么 a 能被 q 整除或 a 与 q 互素。♠

证明：由条件可知

$$(q,a) \mid q \Rightarrow q = t(q,a)$$

又 $(q,a) > 0$，因此 $(q,a) = 1$ 即 a, q 互素，又或 $(q,a) = q$，即 $q \mid a$。

□

推论

设 $a_1, a_2, \ldots, a_n$ 为 n 个整数，q 为一素数，且 $q \mid a_1 a_2 \cdots a_n$ 那么 q 一定能整除某一个 a_k。◇

证明：假定 $a_1, a_2, \ldots, a_n$ 不能被 q 整除，则由定理可知

$$(q, a_k) = 1,\ k = 1, 2, \ldots, n$$

即 q 为 $a_1, a_2, \ldots, a_n$ 的一个公因数，可以知道

$$1 = (q, a_1) = (q, a_1 a_2) = \cdots = (q, a_1 a_2 \cdots a_n)$$

这与条件 $q \mid a_1 a_2 \cdots a_n$ 矛盾，故推论得证。

□

算术基本定理

任意大于 1 的整数都能表达成素数的乘积，即任意大于 1 的整数

$$a = p_1 p_2 \cdots p_n,\ p_1 \geqslant p_2 \geqslant \cdots \geqslant p_n \tag{1}$$

其中 $p_1, p_2, \ldots, p_n$ 为素数，并且若

$$a = q_1 q_2 \cdots q_m,\ q_1 \geqslant q_2 \geqslant \cdots \geqslant q_m$$

其中 $q_1, q_2, \ldots, q_m$ 为素数，则 $m = n, q_i = p_i, i = 1, 2, \ldots, n$。

这个定理又叫做**质因数分解定理**。♠

证明：先用数学归纳法证明 ① 式成立，当 $a = 2$ 时，显然 ① 式成立，假定对一切小于 a 的正整数都满足 ①

式，当 a 为素数时，① 式显然成立，当 a 为合数时，令 $a=bc$，其中 b,c 为两正整数，且满足 $1<b<a,1<c<a$，由假定

$$b=p'_1p'_2\dots p'_l,\ c=p'_{l+1}p'_{l+2}\dots p'_n$$

其中 $p'_1\leqslant p'_2\leqslant\dots\leqslant p'_l,p'_{l+1}\leqslant p'_{l+2}\leqslant\dots\leqslant p'_n$，于是

$$a=p'_1p'_2\cdots p'_n$$

即得 ① 式，故 ① 式对 a 成立。综上，对于任意大于 1 的正整数 ① 式成立。

若 $a=q_1q_2\cdots q_m$，$q_1\geqslant q_2\geqslant\cdots\geqslant q_m$

$$p_1p_2\cdots p_n=q_1q_2\cdots q_m$$

那么有 $p_1\mid q_1q_2\cdots q_m,q_1\mid p_1p_2\cdots p_n$，由推论即知存在 $p_1\mid q_k,q_1\mid p_s$，注意到 q_k,p_s 均为素数，不能被 1 或它本身整除，故 $p_1=q_k,q_1=p_s$，又 $p_1\leqslant p_s,q_1\leqslant q_k$，故 $q_k=p_1\leqslant p_s=q_1$，因此 $p_1=q_k=q_1$，以此类推，可以推出 $p_2=q_2$，最后即得 $m=n,p_i=q_i,i=1,2,\dots,n$。

□

推论

任一大于 1 的整数 a 都可以唯一的写成

$$a=q_1^{\alpha_1}q_2^{\alpha_2}\cdots q_k^{\alpha_k},\ \alpha_i>0,i=1,2,\dots,k \qquad ②$$

其中 $q_i<q_j,i<j$。

◇

② 式被称为 a 的标准分解式，由算术基本定理，这个推论是显然的，然而在实际应用中，常常会加入一些次数为零的素数，写为下列形式

$$a=q_1^{\alpha_1}q_2^{\alpha_2}\cdots q_l^{\alpha_l},\ \alpha_i\geqslant 0,i=1,2,\dots,l$$

推论

设 a 是一个大于 1 的整数，且

$$a=q_1^{\alpha_1}q_2^{\alpha_2}\cdots q_k^{\alpha_k},\ \alpha_i\geqslant 0,i=1,2,\dots,k$$

则 a 的正因数 d 可以表示成

$$d=q_1^{\beta_1}q_2^{\beta_2}\cdots q_k^{\beta_k},\ \alpha_i\geqslant\beta_i\geqslant 0,i=1,2,\dots,k$$

的形式，而且当 d 可以表示为上述形式时，d 是 a 的正因数。

◇

证明：若 $d\mid a$，则有 $a=qd$，由于 a 的标准分解式是唯一的，则 d 的标准分解式中出现的素数都在 q_j，$j=1,2,\dots,k$ 中出现，且 q_j 在 d 的标准分解式中出现的指数满足 $\beta_j\not>\alpha_j$，即 $\beta_j\leqslant\alpha_j$。反之，当 $\beta_j\leqslant\alpha_j$，显然 d 整除 a。

应用以上推论，可以得到以下推论，这是中学教科书中求最大公因数和最小公倍数的根据。

推论

设 a, b 是任意两个正整数，且

$$a = q_1^{\alpha_1} q_2^{\alpha_2} \cdots q_k^{\alpha_k},\ \alpha_i \geqslant 0, i = 1, 2, \ldots, k$$
$$b = q_1^{\beta_1} q_2^{\beta_2} \cdots q_k^{\beta_k},\ \beta_i \geqslant 0, i = 1, 2, \ldots, k$$

则它们的最大公因数以及最小公倍数可以表示为

$$(a, b) = q_1^{\gamma_1} q_2^{\gamma_2} \cdots q_k^{\gamma_k}$$
$$[a, b] = q_1^{\delta_1} q_2^{\delta_2} \cdots q_k^{\delta_k}$$

其中 $\gamma_i = \min(\alpha_i, \beta_i), \delta_i = \max(\alpha_i, \beta_i), i = 1, 2, \ldots, k$。 ◇

证明：由于 $(a, b) \mid a, (a, b) \mid b$，因此记

$$(a, b) = q_1^{\gamma_1} q_2^{\gamma_2} \cdots q_k^{\gamma_k}$$

其中 $\gamma_j \leqslant \alpha_j, \gamma_j \leqslant \beta_j, j = 1, 2, \ldots, k$，当 $\alpha_j \geqslant \beta_j$ 时 $\gamma_j = \beta_j$，反之当 $\alpha_j < \beta_j$ 时，$\gamma_j = \alpha_j$，因此 $\gamma_j = \min(\alpha_j, \beta_j), j = 1, 2, \ldots, k$。

再来，由于 $a \mid [a, b], b \mid [a, b]$，因此记

$$[a, b] = q_1^{\delta_1} q_2^{\delta_2} \cdots q_k^{\delta_k}$$

其中 $\delta_j \geqslant \alpha_j, \delta_j \geqslant \beta_j, j = 1, 2, \ldots, k$，显然 $\delta_j = \max(\alpha_j, \beta_j), j = 1, 2, \ldots, k$。

这个定理可以验证两个正整数最大公因数和最小公倍数的关系

$$\gamma_j + \delta_j = \alpha_j + \beta_j,\ j = 1, 2, \ldots, k$$
$$\prod_{j=1}^{k} q_j^{\gamma_j + \delta_j} = \prod_{j=1}^{k} q_j^{\alpha_j + \beta_j}$$
$$(a, b)[a, b] = ab$$

□

实际上这个推论可以推广到 n 个正整数的情况，即

推论

设 $a_1, a_2, \ldots, a_n$ 是 n 个任意正整数，且

$$a_s = q_1^{\theta_{s,1}} q_2^{\theta_{s,2}} \cdots q_k^{\theta_{s,k}} = \prod_{j=1}^{k} q_j^{\theta_{s,j}},\ \theta_{s,i} \geqslant 0, i = 1, 2, \ldots k, s = 1, 2, \ldots, n$$

则它们的最大公因数以及最小公倍数可以表示为

$$(a_1, a_2, \ldots, a_n) = q_1^{\gamma_1} q_2^{\gamma_2} \cdots q_k^{\gamma_k}$$
$$[a_1, a_2, \ldots, a_n] = q_1^{\delta_1} q_2^{\delta_2} \cdots q_k^{\delta_k}$$

其中 $\gamma_i = \min(\theta_{1,i}, \theta_{2,i}, \ldots, \theta_{n,i}), \delta_i = \max(\theta_{1,i}, \theta_{2,i}, \ldots, \theta_{n,i}), i = 1, 2, \ldots, k$。

证明：由于 $(a_1, a_2, \ldots, a_n) \mid a_k, k = 1, 2, \ldots, n$，因此记

$$(a_1, a_2, \ldots, a_n) = q_1^{\gamma_1} q_2^{\gamma_2} \cdots q_k^{\gamma_k}$$

其中 $\gamma_j \leqslant \theta_{s,j}, s = 1, 2, \ldots, n$，显然 $\gamma_j = \min(\theta_{1,j}, \theta_{2,j}, \ldots, \theta_{n,j})$，再来由于 $a_k \mid [a_1, a_2, \ldots, a_n], k = 1, 2, \ldots, n$，因此记

$$[a_1, a_2, \ldots, a_n] = q_1^{\delta_1} q_2^{\delta_2} \cdots q_k^{\delta_k}$$

其中 $\delta_j \geqslant \theta_{s,j}, s = 1, 2, \ldots, n$，则 $\delta_j = \max(\theta_{1,j}, \theta_{2,j}, \ldots, \theta_{n,j})$。

□

上面的理论证明了任意一个正整数都有唯一的标准分解式，但在实际的计算中没有简易的方法来判断哪些正整数是素数；任意给一个正整数 N，可以按照以下方法找出小于 N 的素数：首先把不超过 N 的正整数由小到大排列出来

$$1, 2, 3, 4, 5, \ldots, N$$

首先划去 1，第一个得到的数是 2，它是素数

$$\not{1}, 2, 3, 4, 5, \ldots, N$$

接下来，从 2 开始每个一位划去一个数，这样就把 2 的倍数给划去了

$$\not{1}, 2, 3, \not{4}, 5, \not{6}, 7, \not{8}, 9, \not{10} \ldots, N$$

第一个留下未划去的数是 3，它不是 2 的倍数，是一个素数，接下来从 3 开始每隔 $3 - 1 = 2$ 位划掉一个数，所划去的数就是 $3 + 3m, m = 1, 2, \ldots$，也就是除 3 外 3 的倍数。

$$\not{1}, 2, 3, \not{4}, 5, \not{6}, 7, \not{8}, \not{9}, \not{10} \ldots, N$$

第一个留下未划去的数是 5，它不是 2,3 的倍数，是一个素数，接下来从 5 开始每隔 $5 - 1 = 4$ 位划掉一个数，所划去的数就是 $5 + 5m, m = 1, 2, \ldots$，划去的是 5 的倍数（除 5 之外）。

如此进行，所有划去的数都是合数，第一个留下的是都不是比它小的素数的倍数，因此也是一个素数，用这种方法可以将素数唯一的求出来，这种方法称为**埃拉托色尼筛法**。

上述的方法并不能在有限步骤以内把正整数的一切上述都找出来，因为有定理

定理

素数的个数无穷的。 ♠

证明：采用反证法来证明该定理，假设正整数中只有有限个素数，记为 $p_1, p_2, \dots, p_k$，令

$$p_1 p_2 \cdots p_k + 1 = N$$

则 $N > 1$，根据定理，N 一定有一个素数因数 p，这里 $p \neq p_i, i = 1, 2, \dots, k$，否则 $p \mid p_1 p_2 \cdots p_k, p \mid N$，因此 $p \mid 1$，而且与 p 是素数矛盾，故 p 是上面 k 个素数以外的素数，因此定理得证。

□

§16.5 高斯函数在数论的应用

定义

函数 $[x]$ 与 $\{x\}$ 是对于一切实数都有定义的函数，函数 $[x]$ 的值等于不大于 x 的最大整数；函数 $\{x\}$ 的值是 $x - [x]$，我们把 $[x]$ 叫做 x 的**整数部分**，把 $\{x\}$ 叫做 x 的**小数部分**。*e.g.* $[\pi] = 3, [\mathrm{e}] = 2, \{1.2\} = 0.2, [4] = 4$。其中 $y = [x]$ 又被叫做**高斯函数**。

由定义立马可以得到这两个函数的一些简单性质

1. $x = [x] + \{x\}$。
2. $[x] \leqslant x < [x+1],\ x - 1 < [x] \leqslant x,\ 0 \leqslant \{x\} < 1$。
3. $[n + x] = [x] + n,\ n \in \mathbb{Z}$。
4. $[x + y] \geqslant [x] + [y], \{x + y\} \leqslant \{x\} + \{y\}$。
5. $$[-x] = \begin{cases} -[x] - 1, & x \notin \mathbb{Z} \\ -[x], & x \in \mathbb{Z} \end{cases}$$
6. （带余除法）如果 a, b 是两个整数，且 $b > 0$，则
$$a = b\left[\frac{a}{b}\right] + b\left\{\frac{a}{b}\right\},\ 0 \leqslant b\left\{\frac{a}{b}\right\} \leqslant b - 1$$
7. 若 a, b 是任意两正整数，则不大于 a 而为 b 的倍数的正整数的个数是 $\left[\frac{a}{b}\right]$

证明：（只证 6、7）显然
$$\frac{a}{b} = \left[\frac{a}{b}\right] + \left\{\frac{a}{b}\right\} \Rightarrow a = b\left[\frac{a}{b}\right] + b\left\{\frac{a}{b}\right\}$$
再来，由于 $0 \leqslant \{x\} < 1$，因此
$$0 \leqslant b\left\{\frac{a}{b}\right\} < b$$
注意到 $b\left\{\frac{a}{b}\right\}$ 可以表示为 $a - b\left[\frac{a}{b}\right]$，且 a, b 均为整数，因此 $b\left\{\frac{a}{b}\right\}$ 为整数，也即

$$0 \leqslant b\left\{\frac{a}{b}\right\} \leqslant b - 1 < b$$

第六条得证。

□

设 $b \mid d, d \leqslant a$，显然当 $b > a$ 时 d 是不存在的，故令 $0 < b \leqslant d \leqslant a$，则

$$0 < d = d'b \leqslant a \Rightarrow d' \leqslant \frac{a}{b}$$

因此满足以上条件的 d 有 d' 个，因而等于 $\left[\frac{a}{b}\right]$。

□

定理

在 $n!$ 的标准分解式中素因数 $p \leqslant n$ 的指数

$$h = \left[\frac{n}{p}\right] + \left[\frac{n}{p^2}\right] + \cdots = \sum_{k=1}^{\infty} \left[\frac{n}{p^k}\right]$$

注意：若 $p^s > n$，则 $\left[\frac{n}{p^s}\right] = 0$，故上式只有有限项不为零，因而有意义。

♠

证明：设想 $2, 3, \ldots, n$ 都分解成标准分解式，则有算术基本定理，h 就是这 $n-1$ 个分解式中 p 的指数之和，设其中 p 的指数是 r 的有 n_r 个 $r \geqslant 1$，则

$$\begin{aligned}
h =& n_1 + 2n_2 + 3n_3 + \cdots \\
=& n_1 + n_2 + n_3 + \cdots + \\
& n_2 + n_3 + \cdots + \\
& n_3 + \cdots + \\
& \cdots \\
=& N_1 + N_2 + N_3 + \cdots
\end{aligned}$$

其中 $N_r = n_r + n_{r+1} + \cdots$ 恰好是 $2, 3, \ldots, n$ 这 $n-1$ 个数中能被 p^r 除尽的个数，但由第七条性质，$N_r = \left[\frac{n}{p^r}\right]$，故

$$h = \left[\frac{n}{p}\right] + \left[\frac{n}{p^2}\right] + \cdots = \sum_{k=1}^{\infty} \left[\frac{n}{p^k}\right]$$

□

推论

$$n! = \prod_{p \leqslant n} p^{\sum_{k=1}^{\infty} \left[\frac{n}{p^k}\right]}$$

◇

这个有很多的应用，在这里举几个简单的例子。

设 $f(x)$ 是一个 n 次整系数多项式，$f^{(k)}(x)$ 是它的 $k(k \leqslant n)$ 阶导数，则 $\frac{f^{(k)}(x)}{k!}$ 是 $n-k$ 次整系数多项式。

证明：显然 $\frac{f^{(k)}(x)}{k!}$ 是 $n-k$ 次多项式，不妨设

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$$

则 $\frac{f^{(k)}(x)}{k!}$ 中 x^i 的系数为

$$b_i = a_{k+i}\frac{(k+i)(k+i-1)\cdots(i+1)}{k!} = a_{k+i}\frac{(k+i)!}{k!i!}$$

其中 $\frac{(k+i)!}{k!i!}$ 实际上是一组合数，即 $\frac{(k+i)!}{k!i!} = \binom{k+i}{k}$，不妨设 $n' = k+i, m' = k$，我们来证明一般的组合数是一整数：

$$\begin{gathered}
n' = (n'-m') + m' \\
\left[\frac{n'}{p^r}\right] \geqslant \left[\frac{n'-m'}{p^r}\right] + \left[\frac{m'}{p^r}\right] \\
\sum_{r=1}^{\infty}\left[\frac{n'}{p^r}\right] \geqslant \sum_{r=1}^{\infty}\left[\frac{n'-m'}{p^r}\right] + \sum_{r=1}^{\infty}\left[\frac{m'}{p^r}\right]
\end{gathered}$$

因此，有

$$\begin{aligned}
&\prod_{p\leqslant n'} p^{\sum_{r=1}^{\infty}\left[\frac{n'}{p^r}\right]} \mid \prod_{p\leqslant n'} p^{\sum_{r=1}^{\infty}\left[\frac{n'-m'}{p^r}\right]+\sum_{r=1}^{\infty}\left[\frac{m'}{p^r}\right]} \\
&m'!(n'-m')! \mid n'!
\end{aligned}$$

因此 b_i 为整数。

□

第十七章　不定方程

§17.1　二元一次不定方程

解不定方程就是求出一个方程的非负整数解，例如

$$ax + by = c$$

其中 a, b, c 是整数，这种形式称为**二元一次不定方程**的一般形式。

定理

设二元一次不定方程

$$ax + by = c \quad ①$$

（其中 a, b, c 是整数，且 a, b 都不是零）有一整数解 $x = x_0, y = y_0$，则上式的一切解可以表示为

$$x = x_0 - b_1 t, \quad y = y_0 + a_1 t$$

其中 $(a, b) = d, a = a_1 d, b = b_1 d, t = 0, \pm 1, \pm 2, \ldots$。

证明：由于 x_0, y_0 是 ① 式的解，那么有

$$ax_0 + by_0 = c$$

那么显然

$$\begin{aligned} a(x_0 + b_1 t) + b(y_0 - a_1 t) &= ax_0 + by_0 + t(ab_1 - ba_1) \\ &= c + t(a_1 b_1 d - a_1 b_1 d) = c \end{aligned}$$

这表明对于任意整数 t 上式均成立。

反之，设 x', y' 是 ① 式的任一解，则有 $ax' + by' = c$，那么有

$$a(x' - x_0) + b(y' - y_0) = 0 \Rightarrow a(x' - x_0) = b(y_0 - y')$$

又 $d = (a, b)$，故 $(a_1, b_1) = \left(\frac{a}{(a,b)}, \frac{b}{(a,b)}\right) = 1$，因此

$$a_1 \mid (y_0 - y')$$

可设 $y' = y_0 + a_1 t$，带入 $a_1(x' - x_0) = b_1(y_0 - y')$ 可得

$$\begin{aligned} a_1(x' - x_0) &= -ta_1b_1 \\ x' &= x_0 - b_1 t \end{aligned}$$

因此 x', y' 只能表示为

$$\begin{cases} x = x_0 - \frac{b}{(a,b)}t \\ y = y_0 + \frac{a}{(a,b)}t \end{cases} \quad t \in \mathbb{Z}$$

的形式，故上式为二元一次不定方程的所有解。

□

下面，我们来证明什么时候（什么条件下）一个一般的二元一次不定方程有解。

定理

① 式有整数解的充分必要条件为 $(a, b) \mid c$.

♠

证明：设 x_0, y_0 为 ① 式的一组整数解，则

$$ax_0 + by_0 = c$$

因为 $(a, b) \mid a, (a, b) \mid b$，所以

$$\begin{aligned} & a = a_1(a, b), b = b_1(a, b) \\ \Rightarrow & (a, b)(a_1x_0 + b_1y_0) = c \\ \Rightarrow & (a, b) \mid c \end{aligned}$$

即证定理的必要性。

再来证明充分性，根据裴蜀恒等式可知，存在整数 s, t 使得

$$(a, b) = as + bt$$

因而 $(as + bt) \mid c$，不妨令 $c = a(ps) + b(pt)$，亦即 ① 式有整数解，其中

$$\begin{cases} x_0 = ps \\ y_0 = pt \end{cases}$$

□

现在来介绍解二元一次不定方程的一般流程：

1. 判断这个方程是否有整数解。
2. 求出这个方程的一个特解 x_0, y_0。
3. 通过特解得到这个不定方程的通解。

通过以上两个定理，我们可以判断以及通过特解求出通解。

有些不定方程的特解很容易看出，*e.g.*

$$10x+7y=17$$

容易看出该方程有一特解 $x_0=1,y_0=1$，又 $(10,7)=1$，因此该方程的通解为

$$\begin{cases} x=1-7t \\ y=1+10t \end{cases} \quad t\in\mathbb{Z}$$

但更一般的不定方程的特解是不容易看出的，下面来介绍求一特解的一般流程。

在证明“$(a,b)\mid c$ 为不定方程有解的充要条件”这个定理的过程中，我们知道了不定方程的一个特解可以表示为

$$\begin{cases} x_0=ps \\ y_0=pt \end{cases}$$

其中，$p=\frac{c}{(a,b)}$，因此主要要解的为裴蜀恒等式，即

$$as+bt=(a,b)$$

并且只需要 s,t 的一个解即可，为了简单起见，令 $a>0,b>0$，并令 $a'=\frac{a}{(a,b)},b'=\frac{b}{(a,b)}$，那么上式可改写为

$$a's+b't=1$$

且 $(a',b')=1$，应用辗转相除法得

$$\begin{aligned} a'&=b'q_1+r_1,\ 0<r_1<b \\ b'&=r_1q_2+r_2,\ 0<r_2<r_1 \\ r_1&=r_2q_3+r_3,\ 0<r_3<r_2 \\ &\cdots \\ r_{n-2}&=r_{n-1}q_n+r_n,\ 0<r_n<r_{n-1} \\ r_{n-1}&=r_nq_{n+1}+r_{n+1},\ r_{n+1}=0 \end{aligned} \tag{1}$$

其中有

$$Q_ka'-P_kb'=(-1)^{k-1}r_k,\ k=1,2,\ldots,n \tag{2}$$

注意到 $(a',b')=r_n=1$，即

$$a'\left[(-1)^{n-1}Q_n\right]+b'\left[(-1)^nP_n\right]=1$$

对比可知有一特殊解为

$$\begin{cases} s=(-1)^{n-1}Q_n \\ t=(-1)^nP_n \end{cases}$$

可以根据以下递推式

$$\begin{cases} P_0=1, & P_1=q_1, & P_k=q_kP_{k-1}+P_{k-2} \\ Q_0=0, & Q_1=1, & Q_k=q_kQ_{k-1}+Q_{k-2} \end{cases} \quad k=2,3,\dots,n$$

即可求出 s,t 的一特殊解，进而得到 x_0,y_0 的特解。

下面用一个例子来演示：*e.g.* 求不定方程的通解

$$114514x+1919y=810$$

通过辗转相除法可知 x,y 两系数互素，也即 $(114514,1919)=1$，那么这个不定方程一定有解，再来求出该方程的一个通解，令

$$114514s+1919y=1$$

根据辗转相除法得下表

n	0	1	2	3	4	5	6	7	8
q	–	59	1	2	15	3	1	2	1
P	1	59	60	179	2745	8414	11159	30732	41891
Q	0	1	1	3	46	141	187	515	702

综上可知 s,t 的一个特殊解为

$$\begin{cases} s=-702 \\ t=41891 \end{cases}$$

又有 $p=\frac{810}{1}=810$，因此该不定方程的一个特解为

$$\begin{cases} x_0=ps=-568620 \\ y_0=pt=33931710 \end{cases}$$

故该不定方程的通解为

$$\begin{cases} x=-568620-1919t \\ y=33931710+114514t \end{cases} \quad t\in\mathbb{Z}$$

从这可以看出，如果一开始读者注意力惊人，可以发现一组较为简单的特解：

$$\begin{cases} x_0=-596 \\ y_0=35566 \end{cases}$$

由这个特解得到的通解与之前的通解是等价的。

我们发现，以上的做法计算量非常大，如果不使用计算器，很可能会算错，下面再提供一种方法，它是许多中学教科书中的方法，我们来讨论一下它的理论依据。

设给定一个适合下列条件的二元一次不定方程

$$ax+by=c,\ a>b>0,\ (a,b)=1 \tag{①}$$

用带余除法表示 a,c

$$a = q_1 b + r_1,\ 0 \leqslant r_1 < b$$
$$c = q_1' b + r_1',\ 0 \leqslant r_1' < b$$

由于 $(a,b)=1$，因此 $(b,r_1)=(a,b)=1$，故方程

$$by' + r_1 x' = r_1' \qquad ②$$

有整数解。设 $x=x_0, y=y_0$ 为 ① 式的任一整数解，则

$$y_0 = \frac{c-ax_0}{b} = q_1' - q_1 x_0 + \frac{r_1' - r_1 x_0}{b}$$

由于 $y_0, q_1' - q_1 x_0$ 均为整数，则 $\frac{r_1'-r_1x_0}{b}$ 为整数，令 $y_0' = \frac{r_1'-r_1x_0}{b}$，则 $x'=x_0, y'=y_0'$ 是 ② 式的一个整数解，即 ① 的任一解都可表示为：

$$\begin{cases} x = x' \\ y = q_1' - q_1 x' + y' \end{cases} \qquad ③$$

其中 x', y' 是 ② 式的某一整数解，反之，若 x', y' 是 ② 式的任一整数解，则由上式求得 x, y 是 ① 式的一个整数解，由 ②,③ 式可以得出

$$\begin{aligned} y &= q_1' - q_1 x' + y' \\ &= q_1' - q_1 x' + \frac{r_1' - r_1 x_0}{b} \\ &= \frac{c - ax}{b} \end{aligned}$$

故得

定理

对于二元一次不定方程

$$ax + by = c,\ a > b > 0,\ (a,b) = 1 \qquad ①$$

的一切整数解可由

$$\begin{cases} x = x' \\ y = q_1' - q_1 x' + y' \end{cases}$$

得出，只要上式中 x', y' 取

$$by' + r_1 x' = r_1'$$

的一切解。

这个定理可以帮助我们化简不定方程，化简到可以看出它的特解的程度即可，再次拿上个例子演示

$$114514x + 1919y = 810$$

注意到 $(114514, 1919) = 1, 114514 > 1919$，因此

$$y = \frac{810 - 114514x}{1919} = -59x + \frac{810 - 1293x}{1919} = -59x + y'$$

注意到 y' 为整数，因此可列不定方程

$$1919y' + 1293x = 810$$

其中 $(1919, 1293) = 1$，因此

$$x = -y' + \frac{810 - 626y'}{1293} = -y' + x'$$

注意到 x' 为整数，因此继续列不定方程

$$1293x' + 626y' = 810$$

其中 $(626, 1293) = 1$，因此

$$y' = 1 - 2x' + \frac{184 - 41x'}{626} = 1 - 2x' + y''$$

其中 y'' 为整数，继续列不定方程

$$626y'' + 41x' = 184$$

系数互素，故可得

$$x' = 4 - 15y'' + \frac{20 - 11y''}{41} = 4 - 15y'' + x''$$

得到新的不定方程

$$11y'' + 41x'' = 20$$

其中系数互素，因此得

$$y'' = 1 - 3x'' + \frac{9 - 8x''}{11} = 1 - 3x'' + y'''$$

因此可得不定方程

$$11y''' + 8x'' = 9$$

其中系数互素，因此得

$$x'' = 1 - y''' + \frac{1 - 3y'''}{8} = 1 - y''' + x'''$$

其中 x''' 是整数，得到

$$8x''' + 3y''' = 1$$

这时就非常容易看出特解 $x_0''' = -1, y''' = 3$，其通解为

$$\begin{cases} x''' = -1 - 3t \\ y''' = 3 + 8t \end{cases}$$

因此我们得到原不定方程的通解

$$\begin{cases} x=-(1-2(4-15(1-3(1-y'''+x''')+y''')+(1-y'''+x'''))+(1-3(1-y'''+x''')+y''')) \\ +4-15(1-3(1-y'''+x''')+y''')+1-y'''+x''' \\ y=-59(-(1-2(4-15(1-3(1-y'''+x''')+y''')+(1-y'''+x'''))+(1-3(1-y'''+x''')+y''')) \\ +4-15(1-3(1-y'''+x''')+y''')+1-y'''+x''')+1-2(4-15(1-3(1-y'''+x''')+y''') \\ +(1-y'''+x'''))+1-3(1-y'''+x''')+y''' \end{cases}$$

再带入

$$\begin{cases} x'''=-1-3t \\ y'''=3+8t \end{cases}$$

即可求出 x,y 的通解。

上面的式子看起来非常繁琐，在一般情况下，更多会通过 x''',y''' 的通解先写出一个较为简单的特解，再来通过关系式写出 x,y 的一个特解，这样会方便许多。

有时二元一次不定方程有限制条件，这里我们讨论最简单的限制条件，一个二元一次不定方程

$$ax+by=N,\ a>0,b>0,(a,b)=1$$

它的非负整数解的个数为 $\left[\frac{N}{ab}\right]$ 或 $\left[\frac{N}{ab}\right]+1$。

证明：设 x_0,y_0 为该方程的任一整数解，由我们之前的讨论，x,y 可以表示为

$$\begin{cases} x=x_0-bt \\ y=y_0+at \end{cases}$$

由 $x,y\geqslant 0$ 得

$$\begin{cases} \frac{x_0}{b}\geqslant t \\ \frac{y_0}{a}\geqslant -t \end{cases} \Rightarrow \frac{x_0}{b}\geqslant t\geqslant -\frac{y_0}{a}$$

不等式每一项加 $\frac{y_0}{a}$ 得

$$0\leqslant t\leqslant \frac{ax_0+by_0}{ab}$$

因此

$$t=\left[\frac{ax_0+by_0}{ab}\right]=\left[\frac{N}{ab}\right]$$

当 $ab\mid N$ 时，$t=\left[\frac{N}{ab}\right]+1$

□

§17.2 多元一次不定方程

一般地，多元一次不定方程可以写成

$$a_1x_1+a_2x_2+\cdots+a_nx_n=N$$

其中 $a_1,a_2,\ldots,a_n,N$ 均为整数，$n\geqslant 2$，我们可以假定 $a_1,a_2,\ldots,a_n,N$ 都不为零。

首先我们来证明这个不定方程有解的充要条件。

定理

多元一次不定方程

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = N$$

有解的充要条件为 $(a_1, a_2, \ldots, a_n) \mid N$

证明：首先证明必要性，令 $(a_1, a_2, \ldots, a_n) = d$；若该不定方程有解，那么存在 n 个整数 $x'_1, x'_2, \ldots, x'_n$ 满足

$$a_1x'_1 + a_2x'_2 + \cdots + a_nx'_n = N$$

由于 $d \mid a_k, k = 1, 2, \ldots, n$ 因此 $d \mid N$，故充分性得证。

再来，若 $d \mid N$，不妨设 $N = pd$，有裴蜀恒等式的推广可知，存在 n 个整数 $x''_k, k = 1, 2, \ldots, n$ 满足

$$N = dp = (a_1, a_2, \ldots, a_n)p = \sum_{k=1}^{n} pa_kx''_k$$

令 $y'_k = px''_k, k = 1, 2, \ldots, n$，那么 $y'_k, k = 1, 2, \ldots, n$ 就是该方程的整数解。

□

求解一个多元一次不定方程通常是把它化成 $n-1$ 个二元一次不定方程，然后逐个求通解。

先顺次求出 $(a_1, a_2) = d_2, (d_2, a_3) = d_3, \ldots, (d_{n-1}, a_n) = d_n$，若 $d_n \mid N$ 则不定方程有解，反之若 $d_n \nmid N$ 则方程无解。若有解，那么可列方程

$$\begin{aligned} a_1x_1 + a_2x_2 &= d_2t_2 \\ d_2t_2 + a_3x_3 &= d_3t_3 \\ &\cdots\cdots\cdots \\ d_{n-2}t_{n-2} + a_{n-1}x_{n-1} &= d_{n-1}t_{n-1} \\ d_{n-1}t_{n-1} + a_nx_n &= N \end{aligned}$$

首先解出最后一个方程的一切解，然后把 t_{n-1} 的值带入倒数第二个中，以此类推最后得到逐个多元一次不定方程的一切解。

在实际求解多元一次不定方程时，常把 t_i 看作常数，求出上面一系列方程中第 $i-1$ 个方程的整数解的一般形式，再从结果中消去 $t_2, t_3, \ldots, t_n$。但这个方法在解较多元的不定方程时也会比较麻烦。

e.g. 解不定方程 $11x + 45y + 14z + 19a + 19b = 810$。

显然 $(11, 45, 14, 19, 19) = 1$，因此该不定方程一定有解，注意到

$$(11, 45) = 1, (1, 14) = 1, (1, 19) = 1$$

则可列

$$
\begin{aligned}
11x+45y&=t_2\\
t_2+14z&=t_3\\
t_3+19a&=t_4\\
t_4+19b&=810
\end{aligned}
$$

我们依次求解，显然最后一式有特解 $t_4=12,b=42$，且系数互素，则其通解为

$$
\begin{cases}
t_4=12-19u\\
b=42+u
\end{cases}
$$

其中 $u=0,\pm1,\pm2,\dots$，对于倒数第二个式子，将 t_4 看作常数，则显然有一特解 $t_3=t_4,a=0$，即通解为

$$
\begin{cases}
t_3=t_4-19v\\
a=v
\end{cases}
$$

其中 $v=0,\pm1,\pm2,\dots$，对于正数第二个式子，同法将 t_3 看作常数，则显然有特解 $t_2=t_3,z=0$，其通解为

$$
\begin{cases}
t_2=t_3-14w\\
z=w
\end{cases}
$$

$w=0,\pm1,\pm2,\dots$，最后来看第一个式子，显然有一特解 $x=-4t_2,y=t_2$，则通解为

$$
\begin{cases}
x=-4t_2-45s\\
y=t_2+11s
\end{cases}
$$

$s=0,\pm1,\pm2,\dots$，综上我们带入消去 t_2,t_3,t_4 得到不定方程

$$11x+45y+14z+19a+19b=810$$

的解为

$$
\begin{cases}
x=-48+76u+76v+56w-45s\\
y=12-19u-19v-14w+11s\\
z=w & \qquad u,v,w,s=0,\pm1,\pm2,\dots\\
a=v\\
b=42+u
\end{cases}
$$

第十八章　同余

§18.1　同余的基本概念以及性质

定义

给定一个正整数 m，把它叫做**模**，如果用 m 去除任意两个整数 a, b 所得的余数相同，我们就说 a, b 对模 m **同余**，记作 $a \equiv b(\bmod m)$，如果余数不同，我们就说 a, b 对模 m 不同余，记作 $a \not\equiv b(\bmod m)$。

由定义立即可以得到三个基本性质

1.（反身性）$a \equiv a(\bmod m)$；

2.（对称性）若 $a \equiv b(\bmod m)$，则 $b \equiv a(\bmod m)$；

3.（传递性）若 $a \equiv b(\bmod m), b \equiv c(\bmod m)$，则有 $a \equiv c(\bmod m)$。

这三个性质是显然的。

定理

整数 a, b 对模 m 同余的充分必要条件是 $m \mid (a-b)$，即 $a = b + mt$，其中 t 是整数。

证明：用带余除法表示 a, b，即

$$a = a_1 m + r_1, b = b_1 m + r_2, \quad 0 \leqslant r_1 < m, 0 \leqslant r_2 < m$$

若 $a \equiv b(\bmod m)$，则 $r_1 = r_2$，故两式相减为

$$a - b = (a_1 - b_1)m$$

即证 $m \mid (a-b)$。反之若 $m \mid (a-b)$，则有

$$a - b = (a_1 - b_1)m + r_1 - r_2 = qm \Rightarrow r_1 - r_2 = (q - a_1 + b_1)m$$

即 $m \mid (r_1 - r_2)$，又 $|r_1 - r_2| < m$，因此 $r_1 - r_2 = 0$。

□

由这个定理很容易得出同余的其他基本性质。如

1. 若 $a_1 \equiv b_1(\bmod m), a_2 \equiv b_2(\bmod m)$ 则

$$a_1 + a_2 \equiv b_1 + b_2(\bmod m)$$

2. 若 $a + b \equiv c(\bmod m)$，则 $a \equiv c - b(\bmod m)$。

3. 若 $a_1 \equiv b_1(\bmod m), a_2 \equiv b_2(\bmod m)$ 则

$$a_1a_2 \equiv b_1b_2(\bmod m)$$

特别的，若 $a \equiv b(\bmod m)$，则 $ak \equiv bk(\bmod m)$。

证明：由定理得

$$a_1 + a_2 = b_1 + b_2 + m(t_1 + t_2)$$

即证第一条。再来若 $a + b \equiv c(\bmod m)$，则

$$c - b \equiv a + b - b \equiv a(\bmod m)$$

即证第二条。

最后来证明第三条，由定理设 $a_1 = b_1 + mt_1, a_2 = b_2 + mt_2$，则

$$a_1a_2 = b_1b_2 + m(b_1t_2 + b_2t_1 + mt_1t_2)$$

故

$$a_1a_2 \equiv b_1b_2(\bmod m)$$

□

特别的，当 $a \equiv b(\bmod m)$，则

$$ak \equiv bk(\bmod m)$$

定理

若整数 $A_{\alpha_1,\alpha_2,\ldots,\alpha_n}, B_{\alpha_1,\alpha_2,\ldots,\alpha_n}, x_i, y_i, i = 1, 2, \ldots, n$ 满足

$$A_{\alpha_1,\alpha_2,\ldots,\alpha_n} \equiv B_{\alpha_1,\alpha_2,\ldots,\alpha_n}(\bmod m)$$
$$x_i \equiv y_i(\bmod m),\ i = 1, 2, \ldots, n$$

则有

$$\sum_{\alpha_1,\alpha_2,\ldots,\alpha_n} A_{\alpha_1,\alpha_2,\ldots,\alpha_n} x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \equiv \sum_{\alpha_1,\alpha_2,\ldots,\alpha_n} B_{\alpha_1,\alpha_2,\ldots,\alpha_n} y_1^{\alpha_1} y_2^{\alpha_2} \cdots y_n^{\alpha_n}(\bmod m)$$

特别的，若 $a_i \equiv b_i(\bmod m), i = 0, 1, 2, \ldots, n$，则存在两个整数

$$\sum_{k=0}^{n} a_k x_k \equiv \sum_{k=0}^{n} b_k y_k(\bmod m)$$

♠

证明：通过同余的性质对于任意 $x_i, y_i, i = 1, 2, \ldots, n$ 都有

$$x_i \equiv y_i(\bmod m)$$
$$x_i^{\alpha_i} \equiv y_i^{\alpha_i}(\bmod m)$$

因此有

$$x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \equiv y_1^{\alpha_1} y_2^{\alpha_2} \cdots y_n^{\alpha_n}(\bmod m)$$

结合 $A_{\alpha_1,\alpha_2,\ldots,\alpha_n} \equiv B_{\alpha_1,\alpha_2,\ldots,\alpha_n}(\mathrm{mod} m)$ 得

$$A_{\alpha_1,\alpha_2,\ldots,\alpha_n} x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \equiv B_{\alpha_1,\alpha_2,\ldots,\alpha_n} y_1^{\alpha_1} y_2^{\alpha_2} \cdots y_n^{\alpha_n}(\mathrm{mod} m)$$

最后用同余的加法性质可证

$$\sum_{\alpha_1,\alpha_2,\ldots,\alpha_n} A_{\alpha_1,\alpha_2,\ldots,\alpha_n} x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \equiv \sum_{\alpha_1,\alpha_2,\ldots,\alpha_n} B_{\alpha_1,\alpha_2,\ldots,\alpha_n} y_1^{\alpha_1} y_2^{\alpha_2} \cdots y_n^{\alpha_n}(\mathrm{mod} m)$$

特别的，当 $a_i \equiv b_i(\mathrm{mod} m), i = 1, 2, \ldots, n$，有

$$a_0 + \sum_n a_n x^n \equiv b_0 + \sum_n b_n x^n (\mathrm{mod} m)$$
$$\sum_{i=0}^n a_i x^i \equiv \sum_{i=0}^n b_i x^i (\mathrm{mod} m)$$

□

下面继续证明同余的基本性质

1. 若 $a \equiv b(\mathrm{mod} m)$，且 $a = a_1 d, b = b_1 d, (d, m) = 1$，则

$$a_1 \equiv b_1(\mathrm{mod} m)$$

2. 若 $a \equiv b(\mathrm{mod} m), k > 0$，则 $ak \equiv bk(\mathrm{mod} mk)$。
3. 若 $a \equiv b(\mathrm{mod} m)$，$d$ 是 a, b 以及 m 的任一正公因数，则有

$$\frac{a}{d} \equiv \frac{b}{d}(\mathrm{mod} \frac{m}{d})$$

4. 若 $a \equiv b(\mathrm{mod} m_i), i = 1, 2, \ldots, n$，则有

$$a \equiv b(\mathrm{mod}[m_1, m_2, \ldots, m_n])$$

5. 若 $a \equiv b(\mathrm{mod} m), d \mid m, d > 0$，则 $a \equiv b(\mathrm{mod} d)$。
6. 若 $a \equiv b(\mathrm{mod} m)$，则 $(a, m) = (b, m)$，因而若 d 能整除 m 以及 a, b 二数之一，则 d 并能整除 a, b 中的另一个。

证明：对于第一条，由定理可知 $m \mid (a - b)$，因此

$$m \mid d(a_1 - b_1) = a - b$$

又 d, m 互素，根据整除的推论可证

$$m \mid (a_1 - b_1)$$

即证 $a_1 \equiv b_1(\mathrm{mod} m)$。

□

这一条实际上是一个乘法的性质，即若 $a_1d \equiv b_1d(\bmod m)$，且 $(d,m)=1$，故

$$a_1 \equiv b_1(\bmod m)$$

对于第二条，由于 $m \mid (a-b)$，显然 $mk \mid (a-b)k$，因此 $ak \equiv bk(\bmod mk)$。

□

对于第三条，由于 d 是 a,b 以及 m 的任一正公因数，则有 $d \mid a, d \mid b, d \mid m$，不妨设

$$a = a_1d, b = b_1d, m = m_1d,$$

由于 $a \equiv b(\bmod m)$，可知 $m \mid (a-b)$，带入上式得 $m_1d \mid (a_1-b_1)d$，即 $m_1 \mid (a_1-b_1)$，故

$$a_1 \equiv b_1(\bmod m_1) \Rightarrow \frac{a}{d} \equiv \frac{b}{d}(\bmod \frac{m}{d})$$

□

对于第四条，由于 $a \equiv b(\bmod m_i), i=1,2,\ldots,n$，易知 $m_i \mid (a-b), i=1,2,\ldots,n$，即 $a-b$ 是 $m_1,m_2,\ldots,m_n$ 的一个公倍数，故有 $[m_1,m_2,\ldots,m_n] \mid (a-b)$，即证

$$a \equiv b(\bmod [m_1,m_2,\ldots,m_n])$$

□

对于第五条，由于 $a \equiv b(\bmod m)$，则根据定理有 $m \mid (a-b)$，又 $d \mid m$，由整除传递性可得 $d \mid (a-b)$，即证 $a \equiv b(\bmod d)$。

□

对于第六条，由定理可知 $m \mid (a-b)$，不妨设 $a = mt+b$，根据整除的性质可知

$$(a,m) = (b,m)$$

□

下面来介绍同余的两个应用

检查因数的一些方法

引理

一个整数能被 3（或 9）整除的充分必要条件是它的十进制数码的和能被 3（或 9）整除。 ♣

证明：设任一正整数为 a，把 a 写成十进位数的形式，即

$$a = a_0 + 10a_1 + \cdots + 10^{n-1}a_{n-1} + 10^n a_n,\ 0 \leqslant a_i < 10$$

由于 $10 \equiv 1(\bmod 3)$，根据定理可知 $a_k10^k \equiv a_k1^k(\bmod 3)$，故

$$a \equiv a_0 + a_1 + \cdots + a_n(\bmod 3)$$

根据同余的性质，若 $3\mid a$，则必有 $3\mid\sum a_k$ 成立。

同理由于 $10\equiv 1(\mathrm{mod}9)$，根据定理可知 $a_k10^k\equiv a_k1^k(\mathrm{mod}9)$，故

$$a\equiv a_0+a_1+\cdots+a_n(\mathrm{mod}9)$$

若 $9\mid a$，当且仅当 $9\mid\sum a_k$。

□

e.g. 若 $a=1145141919$，则

$$1+1+4+5+1+4+1+9+1+9=36$$

由引理即知，1145141919 能被 3、9 整除。

引理

设正整数

$$a=a_n1000^n+a_{n-1}1000^{n-1}+\cdots+a_0,\ 0\leqslant a_i<1000$$

则 7（或 11，或 13）整除 a 的充分必要条件是 7（或 11，或 13）整除

$$(a_0+a_2+\ldots)-(a_1+a_3+\ldots)=\sum_{k=1}^{n}(-1)^ka_k$$

♣

证明：注意到 $1000\equiv -1(\mathrm{mod}7)$，因此 $a_k1000^k\equiv a_k-1^k$，也即当 $7\mid a$ 时，7 也整除下式

$$(a_0+a_2+\ldots)-(a_1+a_3+\ldots)=\sum_{k=1}^{n}(-1)^ka_k$$

同法可证明 11、13 整除 a 时，同时也整除上式。

□

特别的，注意到 $10\equiv -1(\mathrm{mod}11)$，$a$ 为任意正整数

$$a=a_0+10a_1+\cdots+10^{n-1}a_{n-1}+10^na_n,\ 0\leqslant a_i<10$$

因此当 $11\mid a$，一定有

$$11\mid\sum_{k=1}^{n}(-1)^ka_k$$

e.g. 整数 31424588338，根据引理

$$(338+424)-(588+31)=143$$

因为 $11\mid 143, 13\mid 143, 7\nmid 143$，因此该数被 11 以及 13 整除；另一方面可以用

$$(8+3+8+4+4+3)-(3+8+5+2+1)=11$$

来判断该数被 11 整除。

弃九法

弃九法是用来验算整数计算结果的方法，但会纯在一些误差，这一点在后面提到。

假设由普通乘法的运算方法求出的整数 a, b 的乘积为 P，不妨令

$$a = a_0 + 10a_1 + \cdots + 10^{n-1}a_{n-1} + 10^n a_n,\ 0 \leqslant a_i < 10$$
$$b = b_0 + 10b_1 + \cdots + 10^{m-1}b_{m-1} + 10^m b_m,\ 0 \leqslant b_i < 10$$
$$P = p_0 + 10p_1 + \cdots + 10^{l-1}p_{l-1} + 10^l p_l,\ 0 \leqslant p_i < 10$$

显然有

$$ab \equiv \left(\sum_{i=0}^{n} a_i\right)\left(\sum_{j=0}^{m} b_j\right)(\bmod 9)$$
$$P \equiv \sum_{k=0}^{l} p_k (\bmod 9)$$

因此如若（根据同余的传递性）

$$\left(\sum_{i=0}^{n} a_i\right)\left(\sum_{j=0}^{m} b_j\right) \not\equiv \sum_{k=0}^{l} p_k (\bmod 9)$$

那么所求得的乘积是错误的。

以上就是弃九法的原理，当上式中 a_i, b_j, p_k 中出现 9 时，还可以去掉（因为 $9 \equiv 0(\bmod 9)$），不仅如此凡是数码之和为 9 的通通可以去掉。

e.g.　记 $a = 114514, b = 1919810$，如果用普通方法得到 a, b 的乘积为 $P = 219845922340$，按照上述方法得

$$a \equiv 7(\bmod 9), b \equiv 2(\bmod 9), P \equiv 13(\bmod 9)$$

显然 $7 \times 2 = 14 \not\equiv 13(\bmod 9)$，故上式计算得到的积是错误的，准确的应为 219845122340（自己重新算）。

不过若计算结果为 $P = 218945122340$，其数码和正确答案一样，因此有 $P \equiv 5(\bmod 9)$，故

$$14 \equiv 5(\bmod 9)$$

因此弃九法在碰到这种情况时会有差错。

§18.2　剩余类以及完全剩余系

命题：任意三个连续整数中，必有一个被 3 整除。

证明：我们可以把任意正整数分为三类：$3q, 3q+1, 3q+2, q \in \mathbb{Z}$，注意到它是整数的一个划分，因此任意连续的三个整数都必在这三个集合中，且每个整数仅在上述的一个集合里，即证任意连续的三个整数中必有 $3q$，其为 3 的倍数。

□

通过以上的讨论，其实不难证任意 m（m 是正整数）个连续的整数中，其中必有一个（且只有一个）被 m

整除。因此有

定理

若 m 是一个给定的正整数，则全部整数可以分成 m 个集合，记作 $K_0, K_1, K_2, \ldots, K_{m-1}$，其中 $K_r(r = 0, 1, 2, \ldots, m)$ 为一切形如 $qm + r(r = 0, \pm 1, \pm 2, \ldots)$ 的整数所组成的，这些集合具有以下性质。

1. 每一个整数必包含在而且仅在上述的一个集合中。
2. 两个整数同在一个集合的充分必要条件是这两个数对模 m 同余。

证明：对于第一条，设 a 为任意整数，将它用带余除法展开

$$a = a_1 m + r_a,\ 0 \leqslant r_a < m$$

故 a 在集合 K_{r_a} 中，又 r_a 是唯一确定的，因此 a 只能在 K_{r_a} 内。

对于第二条，设 a, b 为两个整数，并且都在 K_r 内，则

$$a = a_1 m + r,\ b = b_1 m + r$$

显然 $a \equiv b(\bmod m)$，反之当 $a \equiv b(\bmod m)$ 时，a, b 都在某一 K_r 内。

□

由这个定理可知能将整数集划分，我们引入以下定义

定义

定理中的 $K_0, K_1, K_2, \ldots, K_{m-1}$ 叫做模 m 的**剩余类**，一个剩余类中任一数叫做它的**同类的数的剩余**。若 $a_0, a_1, \ldots, a_{m-1}$ 是 m 个整数，并且其中任何两数都在不同在一个剩余类里，则 $a_0, a_1, \ldots, a_{m-1}$ 叫做模 m 的一个**完全剩余系**。

e.g. 模 4 的剩余类 K_0, K_1, K_2, K_3 的通项可以分别表示为 k_0, k_1, k_2, k_3 其中

$$\begin{aligned}
k_0 &= 4q \\
k_1 &= 4q + 1 \\
k_2 &= 4q + 2 \\
k_3 &= 4q + 3
\end{aligned}$$

其中 $q \in \mathbb{Z}$，因此模 4 的剩余类为

$$\begin{aligned}
K_0 &= \{\cdots, -8, -4, 0, 4, 8, \cdots\} \\
K_1 &= \{\cdots, -7, -3, 1, 5, 9, \cdots\} \\
K_2 &= \{\cdots, -6, -2, 2, 6, 10, \cdots\} \\
K_3 &= \{\cdots, -5, -1, 3, 7, 11, \cdots\}
\end{aligned}$$

由定义可知

$$
\begin{aligned}
K_0 &= \{\cdots, -8, \underline{-4}, 0, 4, 8, \cdots\} \\
K_1 &= \{\cdots, -7, -3, \underline{1}, 5, 9, \cdots\} \\
K_2 &= \{\cdots, \underline{-6}, -2, 2, 6, 10, \cdots\} \\
K_3 &= \{\cdots, -5, -1, 3, 7, \underline{11}, \cdots\}
\end{aligned}
$$

如上，用圆圈圈起来的为一个模 4 的完全剩余类，亦或下划线划出的，也就是说只要选出来的数任意两个都不在一个剩余类即可。

显然连续的 m 个整数一定是模 m 的一个完全剩余系（如上画圈的）。

下面给出一个判断 m 个整数是否在模 m 的一个完全剩余系的方法。

推论

m 个整数作成的模 m 的一个完全剩余系的充分必要条件是这 m 个整数两两对模 m 不同余。 ◇

证明：设这 m 个整数中任意两个整数为 a, b，当这两数对模 m 同余时显然有 $a, b \in K_r$，亦即 a, b 在同一个剩余类中。反之当 a, b 在模 m 的一个完全剩余系中，可知 $a, b \notin K_r$，亦即 a, b 对模 m 不同余。

□

利用这个推论我们能判断一个序列是否是模 m 的完全剩余系。

e.g. 已知序列 $0, m+1, \ldots, am+a, \ldots, (m-1)m+(m-1)$。

这个序列中任意两数可以写为 $am+a, bm+b (a \neq b)$，由于

$$m \nmid [(a-b)m+(a-b)]$$

因此该序列为模 m 的一个完全剩余系。

□

e.g. 已知序列 $0, -m+1, \ldots, (-1)^a m+a, \ldots, (-1)^{m-1} m+m-1$。

这个序列中任意两数可以写为 $(-1)^a m+a, (-1)^b m+b, a \neq b$，由于

$$m \nmid [((-1)^a-(-1)^b)m+a-b]$$

因此该序列为模 m 的一个完全剩余系。

□

定理

设 m 为正整数，$(a, m) = 1$，b 是任意整数，若 x 通过模 m 的一个完全剩余系，则 $ax+b$ 也通过模 m 的完全剩余系，也就是说，若 $a_0, a_1, \ldots, a_{m-1}$ 是模 m 的完全剩余系，则 $aa_0+b, aa_1+b, \ldots, aa_{m-1}+b$ 也是模 m 的完全剩余系。 ♠

证明：要证明这个定理，即证 $aa_0+b, aa_1+b, \ldots, aa_{m-1}+b$ 两两不同余即可。下面采用反证法来证明这一点。

假定 $aa_i+b\equiv aa_j+b(\bmod m), i\neq j$，根据同余的性质这个式子等价于

$$aa_i\equiv aa_j(\bmod m)$$

由于 $(a,m)=1$，根据同余的性质即知

$$a_i\equiv a_j(\bmod m)$$

显然这与 $a_0,a_1,\ldots,a_{m-1}$ 是模 m 的完全剩余系矛盾，故定理得证。

□

定理

若 m_1,m_2 是两个互素的正整数，而 x_1,x_2 分别通过模 m_1,m_2 的完全剩余系，则 $m_2x_1+m_1x_2$ 通过模 m_1m_2 的完全剩余系。

♠

也就是说，若 $x_{1,0},x_{1,1},x_{1,2},\ldots,x_{1,m_1-1}$ 是模 m_1 的一个完全剩余系，$x_{2,0},x_{2,1},x_{2,2},\ldots,x_{2,m_2-1}$ 是模 m_2 的一个完全剩余系，那么

$$m_1x_{2,i}+m_2x_{1,j}, i=0,1,2,\ldots,m_2-1, j=0,1,2,\ldots,m_1-1$$

是模 m_1m_2 的完全剩余系。

证明：由条件可知 x_1,x_2 分别通过 m_1,m_2 个整数，那么 $m_2x_1+m_1x_2$ 通过 m_1m_2 个整数。只需证明这 m_1m_2 个整数对模 m_1m_2 两两不同余即可。假定

$$m_2x_1'+m_1x_2'\equiv m_2x_1''+m_1x_2''(\bmod m_1m_2)$$

其中 x_1',x_1'' 是 x_1 所通过的完全剩余系的整数，而 x_2',x_2'' 是 x_2 通过完全剩余系的整数。由同余的性质可知

$$m_1x_2'\equiv m_1x_2''(\bmod m_1m_2),\quad m_2x_1'\equiv m_2x_1''(\bmod m_1m_2)$$

再来由于 m_1,m_2 为正整数，因此根据同余的性质

$$m_1x_2'\equiv m_1x_2''(\bmod m_2),\quad m_2x_1'\equiv m_2x_1''(\bmod m_1)$$

又由于 $(m_1,m_2)=1$，即得

$$x_2'\equiv x_2''(\bmod m_2),\quad x_1'\equiv x_1''(\bmod m_1)$$

由我们一开始得到的定理，得 $x_1'=x_1'',x_2'=x_2''$，故当 x_1',x_1'' 与 x_2',x_2'' 不全相同，则假定不成立，因此定理得证。

□

下面给出几个最简单也是最常用的完全剩余系。

定义

$0,1,\ldots,m-1$ 这 m 个整数叫做模 m 的**最小非负完全剩余系**；

当 m 是偶数时，$-\frac{m}{2},\ldots,-1,0,1,\ldots\frac{m}{2}-1$ 或 $-\frac{m}{2}+1,\ldots,-1,0,1,\ldots,\frac{m}{2}$ 叫做**绝对最小完全剩余**

系；

当 m 是奇数时，$-\frac{m-1}{2},\ldots,-1,0,1,\ldots\frac{m}{2-1}$ 叫做模 m 的**绝对最小完全剩余系**。

§18.3　既约剩余系与欧拉函数

定义

欧拉函数 $\varphi(a)$ 是定义在正整数上的函数，它在正整数 a 上的值等于序列 $0,1,2,\ldots,a-1$ 中与 a 互素的数的个数。

下面举一些简单的例子：

$$\varphi(3)=2,\quad \varphi(10)=4,\quad \varphi(37)=36$$

实际上我们发现，当 $\varphi(a)$ 中的 a 为素数时，$\varphi(a)=a-1, a\geqslant 2$，更一般的结论我们在之后介绍。

定义

如果一个模 m 的剩余类里面的数与 m 互素，就把它叫做一个**模 m 互素的剩余类**。

在与模 m 互素的剩余类中，从每一类各任取一数所作成的数的集合，叫做模 m 的一个**既约剩余系**。

下面通过一个例子来体会这个定义。

e.g.　模 4 的剩余类为

$$\begin{aligned}
K_0&=\{\cdots,-8,-4,0,4,8,\cdots\}\\
K_1&=\{\cdots,-7,-3,1,5,9,\cdots\}\\
K_2&=\{\cdots,-6,-2,2,6,10,\cdots\}\\
K_3&=\{\cdots,-5,-1,3,7,11,\cdots\}
\end{aligned}$$

它们的通项可写为

$$\begin{aligned}
k_0&=4q\\
k_1&=4q+1\\
k_2&=4q+2\\
k_3&=4q+3
\end{aligned}$$

其中有

$$\begin{aligned}
(4,k_0)&=4\\
(4,k_1)&=1\\
(4,k_2)&=2\\
(4,k_3)&=1
\end{aligned}$$

因此模 4 互素的剩余类为 K_1, K_3，在这两个集合中分别取一个数，这两个数就组成了模 4 的一个既约剩余系。

定理

1. 模 m 的剩余类与模 m 互素的充分必要条件是此类中有一数与 m 互素。
2. 因此与模 m 互素的剩余类的个数是 $\varphi(m)$。
3. 模 m 的每一个既约剩余系是由与 m 互素的 $\varphi(m)$ 个对模 m 不同余的整数组成的。

这个定理简化了判断模 m 的互素的剩余类的难度，例如上例中我们通过模 4 的剩余类的通项与 4 的最大公因数来判断哪些为模 4 的互素的剩余类，有了这个定理以后只需每个集合中取一数来判断即可。

证明：设 $K_0, K_1, \ldots, K_{m-1}$ 是模 m 的全部剩余类。若 K_r 是一个与模 m 互素的剩余类，则显然 $(r, m) = 1$。

反之，若 $k_r \in K_r, (k_r, m) = 1$，根据剩余系的定理可知，任意两个整数在同一集合的充要条件是它们模 m 同余。即设 $k_r' \in K_r$，那么有

$$k_r \equiv k_r' (\mathrm{mod} m)$$

根据同余的性质可知

$$(k_r, m) = (k_r', m) = 1$$

即证任一 K_r 中的数都与模 m 互素，故 K_r 是一个与模 m 互素的剩余类。

由上即证模 m 的剩余类 $K_r, r = 0, 1, \ldots, m-1$ 与模 m 互素的有 $\varphi(m)$ 个。

□

由这个定理很容易得到

定理

若 $a_1, a_2, \ldots, a_{\varphi(m)}$ 是 $\varphi(m)$ 个与 m 互素的整数，并且两两对模 m 不同余，则 $a_1, a_2, \ldots, a_{\varphi(m)}$ 是模 m 的一个既约剩余系。

证明：由定理可知，与模 m 互素的剩余类的个数是 $\varphi(m)$ 个。

将 $a_1, a_2, \ldots, a_{\varphi(m)}$ 用带余除法可表示为

$$a_i = mq_i + r_i, \quad 0 \leqslant r_i < m, i = 1, 2, \ldots, \varphi(m)$$

存在集合 K_{r_i} 构成的模 m 的剩余类，其中 $a_i \in K_{r_i}$；由已知 $a_1, a_2, \ldots, a_{\varphi(m)}$ 都与 m 互素

$$(a_i, m) = 1$$

则由定理即知，模 m 的剩余类与模 m 互素的，再由已知

$$a_i \not\equiv a_j (\mathrm{mod} m)$$

故对于任意的 i, j 都有 a_i, a_j 不在一个剩余类中。综上即证 $a_1, a_2, \ldots, a_{\varphi(m)}$ 为模 m 的一个既约剩余系。

□

与上一节中完全剩余系类似的有如下性质。

定理

若 $(a,m)=1$，x 通过模 m 的既约剩余系，则 ax 通过模 m 的既约剩余系。

证明：由条件 $x_1,x_2,\ldots,x_{\varphi(m)}$ 是一个模 m 的既约剩余系，那么有

$$(x_i,m)=1,\quad i=1,2,\ldots,\varphi(m)$$

由于 $(a,m)=1$，由最大公因数的性质可知

$$(ax_i,m)=1$$

如若有

$$ax_i\equiv ax_j(\mathrm{mod}m),i\neq j,i,j=1,2,\ldots,\varphi(m)$$

由同余的性质可知有 $x_i\equiv x_j(\mathrm{mod}m)$，这与条件矛盾，故 ax 通过模 m 的既约剩余系。

□

定理

若 m_1,m_2 是两个互素的正整数，x_1,x_2 分别通过模 m_1,m_2 的既约剩余系，则 $m_2x_1+m_1x_2$ 通过模 m_1m_2 的既约剩余系。

证明：易知既约剩余系是完全剩余系的一个子集，因此我们只需要证明：若 $(x_1,m_1)=(x_2,m_2)=1$，则由 $(m_1,m_2)=1$，以及最大公因数的性质可知

$$(m_2x_1,m_1)=(m_1x_2,m_2)=1$$

再来注意到

$$m_2x_1+m_1x_2=m_2(x_1)+m_1x_2\Rightarrow(m_2x_1+m_1x_2,m_2)=(m_1x_2,m_2)$$

同理得 $(m_2x_1,m_1)=1$，故有

$$(m_2x_1+m_1x_2,m_1m_2)=1$$

反之，若 $(m_2x_1+m_1x_2,m_1m_2)=1$，则

$$(m_1x_2,m_2)=(m_2x_1,m_1)=1$$

即得 $(m_2x_1,m_1)=(m_1x_2,m_2)=1$，由于 $(m_1,m_2)=1$，故 $(x_1,m_1)=(x_2,m_2)=1$。

□

推论

若 m_1,m_2 是两个互素的正整数，则 $\varphi(m_1m_2)=\varphi(m_1)\varphi(m_2)$

证明:由定理可知,若 m_1,m_2 是两个互素的正整数,x_1,x_2 分别通过模 m_1,m_2 的既约剩余系,则 $m_2x_1+m_1x_2$ 通过模 m_1m_2 的既约剩余系。即 $m_2x_1+m_1x_2$ 通过 $\varphi(m_1m_2)$ 个整数，而另一方面由于 x_1 通过 $\varphi(m_1)$ 个整

数，x_2 通过 $\varphi(m_2)$ 个整数，因此

$$\varphi(m_1m_2)=\varphi(m_1)\varphi(m_2)$$

□

定理

设 $a=p_1^{\alpha_1}p_2^{\alpha_2}\cdots p_k^{\alpha_k}$，则

$$\varphi(a)=a\left(1-\frac{1}{p_1}\right)\left(1-\frac{1}{p_2}\right)\cdots\left(1-\frac{1}{p_k}\right)$$

证明：由推论即知

$$\varphi(a)=\varphi(p_1^{\alpha_1})\varphi(p_2^{\alpha_2})\cdots\varphi(p_k^{\alpha_k})$$

再来我们证明，若 p 为素数，一般的有

$$\varphi(p^n)=p^n-p^{n-1}$$

注意到序列

$$0,1,2,\ldots,p\ldots,2p,\ldots,p^2,\ldots,p^2+p,\ldots,2p^2,\ldots,p^n-1,p^n$$

由于 p 为素数，因此不与 p^n 互素的数一定满足 $(p^n,sp)=p^t$，即是 p 的倍数，上述序列中，不超过 p^n，且为 p 的倍数的正整数的个数是

$$\left[\frac{p^n}{p}\right]=p^{n-1}$$

那么这 p^n 个数中，与 p^n 互素的数为

$$p^n-p^{n-1}=\varphi(p^n)$$

综上即知

$$\begin{aligned}\varphi(a)&=\left(p_1^{\alpha_1}-p_1^{\alpha_1-1}\right)\left(p_2^{\alpha_2}-p_2^{\alpha_2-1}\right)\cdots\left(p_k^{\alpha_k}-p_k^{\alpha_k-1}\right)\\&=\varphi(p_1^{\alpha_1})\varphi(p_2^{\alpha_2})\cdots\varphi(p_k^{\alpha_k})\left(1-\frac{1}{p_1}\right)\left(1-\frac{1}{p_2}\right)\cdots\left(1-\frac{1}{p_k}\right)\\&=a\left(1-\frac{1}{p_1}\right)\left(1-\frac{1}{p_2}\right)\cdots\left(1-\frac{1}{p_k}\right)\end{aligned}$$

□

§18.4 欧拉定理、费马定理及其对循环小数的应用

定理

欧拉定理：设 m 是大于 1 的整数，$(a,m)=1$，则

$$a^{\varphi(m)}\equiv 1(\text{mod}m)$$

定理

费马定理：若 p 为素数，则

$$a^p \equiv a(\mathrm{mod} p)$$

第五部分

概率论与数理统计

第十九章　概率

§19.1　随机试验

定义

随机试验：对随机现象的实现和对它的观察，简称试验用 E 表示。

特点：

1. 可重复进行
2. 所有结果明确可知，不止一个
3. 每次试验得到可能结果的一个，但事先无法确定

性质：

1. 有限性 (有限样本空间)
2. 不唯一

样本点：试验中每个可能的基本结果。
样本空间：全体样本点组成的集合；一般用 Ω 表示样本空间，ω 表示样本点。
有限样本空间：一个试验有 n 个结果，是有限的。
随机事件 (简称事件)：样本空间的子集。
基本事件：只包含一个样本点的事件。
特别的：
Ω 作为本身的子集，Ω 必然发生，我们称 Ω 为必然事件。
ϕ 不包含任何的样本点，称之为不可能事件。

§19.2　事件关系

研究概率论我们需要用到集合的工具来描述。
与集合类似的有包含被包含等：
事件 A 包含事件 B，$B \subseteq A$
事件 A 和事件 B 相等，$B = A$
交事件或积事件，$A \cap B, AB$
并事件或和事件，$A \cup B, A + B$
如果 $A \cap B = \phi$，则称事件 A 和事件 B 互斥 (或互不相容)
如果 $A \cap B = \phi, A \cup B = \Omega$ 则称事件 A 和事件 B 互相对立；A 的对立事件可以记为 $\overline{A}$

§19.3　古典概型

定义

概率：随机事件发生的可能性大小的度量，A 的概率用 $P(A)$ 表示。

古典概率模型：简称古典概型，有以下几个共同特征：

1. 有限性：样本空间的样本点只有有限个
2. 等可能性：每个样本点发生的可能性相等

一般地，设试验 E 是古典概型，样本空间 Ω 包含 n 个样本点，事件 A 包含 k 个样本点，则定义事件 A 的概率为

$$P(A)=\frac{k}{n}=\frac{n(A)}{n(\Omega)}$$

$\frac{n(A)}{n(\Omega)}$ 是它的古典定义，$n(A), n(\Omega)$ 表示 A 和 Ω 包含的样本点个数。

基本性质

(1)，对任意事件 A，都有

$$P(A)\geqslant 0$$

(2)，必然事件概率为 1. 不可能事件概率为 0

$$P(\Omega)=1, P(\phi)=0$$

(3)，如果事件 A 与事件 B 互斥，那么 $P(A\cup B)=P(A)+P(B)$，推广：如果事件 $A_i(i=1,2,\ldots,n)$ 互斥则有

$$P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P\left(A_i\right)$$

(4)，如果事件 A 与事件 B 互为对立事件，则

$$P(B)=1-P(A), P(A)=1-P(B)$$

(5)，如果 $A\subseteq B$，那么 $P(A)\leqslant P(B)$

(6)，设 A, B 是一个随机试验中的两个事件，我们有

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

对于任意事件 A, B，如果 $P(AB)=P(A)P(B)$，则称 A, B 相互独立，简称独立。

频率

频率的稳定性：随着试验次数 n 增加，频率偏离概率的幅度会缩小，即 $f_n(A)$ 逐渐稳定于 $P(A)$ 概率。

第二十章　随机变量

这一章没啥好说的，高考试卷每次固定出一道概统大题，新教材加了全概率公式和贝叶斯公式，通常会在概统题里给出公式,22 年新一卷可以看出，这道大题会有比较大量的计算，拉高做题的时间成本。这一章内容的顺序我不是按教材书上的编排来写的，我把随机变量的数字特征放到了后面，方便复习时查看它们的性质。

这一章属于概率论的基本概念，因此也会有比较严格的定义，对于全概率公式和贝叶斯公式等想进一步了解的可以看“概率论与数理统计”的书。

§20.1　条件概率与独立事件

概率论的研究常常将集合论作为工具进行研究，用集合论的视角来理解这些定义定理会比记文字牢靠得多，特别是用 Venn 图“数形结合”的记忆，会比被概念牢靠（个人经验）

条件概率定义

条件概率：在事件 A 发生的条件下，事件 B 发生的条件概率，用 $P(B|A)$ 表示，有

$$P(B|A) = \frac{P(AB)}{P(A)}$$

请理解 $P(AB)$ 与 $P(B|A)$ 的不同。

$$P(B|A) = \frac{n(AB)}{n(A)}, P(AB) = \frac{n(AB)}{n(\Omega)}$$

全集 Ω

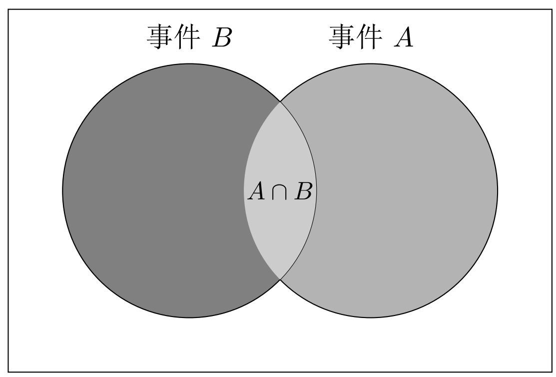


图 20.1: 条件概率

显然 $P(A) > 0$；结合图像，$P(B|A), P(AB)$ 的不同在于它们的“基底”不同，前者是以所有发生了事件 A 的样本点的集合为基底，而后者是以全集 Ω 为基底。
不能验证，条件概率 $P(\cdot|A)$ 满足三个条件：

1. **非负性**：对于每一个事件 B，有 $P(B|A) \geqslant 0$
2. **规范性**：对于必然事件 S，有 $P(S|A) = 1$
3. **可列可加性**：设 $B_i, i = 1, 2, \ldots, n$ 是两两互不相容的事件，则有：

$$P(\bigcup_{i=1}^{\infty} B_i|A) = \sum_{i=1}^{\infty} P(B_i|A)$$

定理

若等式两边同乘 $P(A)$，我们可以得到**乘法定理**：

设 $P(A) > 0$，则有

$$P(B|A) \cdot P(A) = P(AB) \tag{1}$$

推论

(1) 式可推广到多个事件积的情况，设 $A_i(i = 1, 2, \ldots, n)$ 为 n 个事件，$n \geqslant 2$，且 $P(A_1A_2\cdots A_n) > 0$，则有：

$$P(A_1A_2\cdots A_n) = P(A_n|A_1\cdots A_{n-1})P(A_{n-1}|A_1\cdots A_{n-2})\cdots P(A_2|A_1)P(A_1)$$

$$P(\prod_{i=1}^{n} A_i) = P(A) \cdot \prod_{j=2}^{n} P(A_j|\prod_{i=1}^{j-1} A_i) \tag{2}$$

(2) 式是我自己化简的，感觉不太巧妙……

如果 A, B 相互独立，由独立的定义可知 $P(AB) = P(A)P(B)$，则：

$$P(B|A) = \frac{P(A) \cdot P(B)}{P(A)}$$

$$P(B|A) = P(B)$$

这告诉我们如果事件 A, B 相互独立，互不干涉，则事件 B 发生在事件 A 发生的条件下的概率就是事件 B 发生的概率，这句话读起来拗口，读者理解即可。

如果事件 B，C 互斥，则有：

$$P(B \cup C|A) = P(B|A) + P(C|A)$$

结合 Venn 图更好理解，特别的，当事件 B, C 互为对立事件时，即 $PB \cup C = 1$，由条件概率的规范性可知：$P(B \cup C|A) = 1$，读者可自行证明。

定义

设 S 为试验 E 的样本空间，$B_1, B_2, \ldots, B_n$ 为 E 的一组事件，若：

i. $B_iB_j = \varnothing, i \neq j, i, j = 1, 2, \ldots, n$；

ii. $B_1 \cup B_2 \cup \ldots \cup B_n = S$

则称 $B_1, B_2, \ldots, B_n$ 为样本空间 S 的一个**划分**。

在高中教材没有严格定义“划分”，但这个概念有助于我们理解接下来的全概率公式和贝叶斯公式。

定理

设试验 E 的样本空间为 S，B 为 E 的事件，$A_1, A_2, \ldots, A_n$ 为 S 的一个划分，且 $P(A_i) > 0 (i = 1, 2, \ldots, n)$，则

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B|A_i) \tag{17.1}$$

(17.1) 式被称为**全概率公式**。

很多实际问题中的 $P(B)$ 不容易直接求得，但容易找到 S 的一个划分 $A_1, A_2, \ldots, A_n$，且 $P(A_i)$ 和 $P(B|A_i)$ 容易求得，那么可根据全概率公式求出 $P(B)$。

证明：因为

$$\begin{aligned} B = BS &= B(A_1 \cup A_2 \cup \cdots \cup A_n) \\ &= BA_1 \cup BA_2 \cup \cdots \cup BA_n \end{aligned}$$

由假设 $P(A_i) > 0 (i = 1, 2, \ldots, n)$，且 $(BA_i)(BA_j) = \varnothing, i \neq j, i, j = 1, 2, \ldots, n$ 得到

$$\begin{aligned} P(B) &= P(BA_1) + P(BA_2) + \cdots + P(BA_n) \\ &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n) \\ &= \sum_{i=1}^{n} P(A_i) P(B|A_i) \end{aligned}$$

□

定理

设试验 E 的样本空间为 S，B 为 E 的事件，$A_1, A_2, \ldots, A_n$ 为 S 的一个划分，且 $P(B) > 0, P(A_i) > 0 (i = 1, 2, \ldots, n)$，则

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{k=1}^{n} P(A_k)P(B|A_k)}, i = 1, 2, \ldots, n \tag{17.2}$$

(17.2) 式称为**贝叶斯公式**。

贝叶斯公式刻画了 $P(B|A_i)$ 与 $P(A_i|B)$ 的关系，下面来证明贝叶斯公式。

证明：由条件概率的定义以及全概率公式即可得：

$$\begin{aligned} P(A_i|B) &= \frac{P(A_iB)}{P(B)} \\ &= \frac{P(A_i)P(B|A_i)}{\sum_{k=1}^{n} P(A_k)P(B|A_k)}, i = 1, 2, \ldots, n \end{aligned}$$

□

§20.2　离散型随机变量及其分布列

对于随机试验样本空间 Ω 中每个样本点 ω，都有唯一实数 $X(\omega)$ 与之对应，称 X 为“**随机变量**”。

离散型随机变量：可取值为有限个或可列无限多个的随机变量，但在高中一般只讨论有限个取值的情况。

离散型随机变量 X 可能取不同的值为 $x_i,(i=1,2,\ldots)$，X 取各个可能的值的概率，即事件 $X=x_i$ 的概率为：

$$P(X=x_i)=p_i, i=1,2,\ldots \tag{17.3}$$

由概率的定义，p_i 满足下列两个条件：

1. $p_i \geqslant 0, i=1,2,\ldots,n$
2. $\sum_{i=1}^{\infty} p_i = 1$

在高中一般不考虑取值为可列无限多个的情况，因为没有学过级数的东西，因此高中的定义只需取到有限值就行，如 (17.3) 式改成 $P(X=x_i)=p_i, i=1,2,\ldots,n$；另外看到 ∞ 符号就换成 n 即可。

(17.3) 式称为 X 的概率分布列，简称为**分布列**[①]，也可以用表格表示：

X	x_1	x_2	$\ldots$	x_n	$\ldots$
P	p_1	p_2	$\ldots$	p_n	$\ldots$

可知

$$\sum_{i=1}^{\infty} p_i = 1$$

两点分布

设随机变量 X 只可能取 0 与 1 两个值，它们的分布列为

$$P(X=k)=p^k(1-p)^{1-k}, k=0,1 \quad (0<p<1)$$

则称 X 服从以 p 为参数的 $(0-1)$ 分布或两点分布。

两点分布的分布列也可以写成：

X	0	1
p_k	$1-p$	p

§20.2.1　伯努利试验、二项分布

设试验 E 只有两个可能的结果：A 以及 $\overline{A}$，则称 E 为**伯努利试验**，设 $P(A)=p\ (0<p<1)$，此时 $P(\overline{A})=1-p$, 将 E 独立重复地进行 n 次，则称这一串重复的独立试验为n **重伯努利试验**。其中：

1. “重复”是指每一次试验中 $P(A)=p$ (概率相同)；
2. “独立”是指各次试验结果相互独立，即

$$P(\prod_{i=1}^{n} A_i)=\prod_{i=1}^{n} P(A_i)$$

[①]有些书中称为分布律

一般地，在 n 重伯努利试验中，设每次试验中事件 A 发生的概率为 $p(0<p<1)$，用 X 表示 A 发生的次数，则 X 的分布列为

$$P\left(X=k\right)=C_n^k p^k(1-p)^{n-k}, k=0,1,2,\ldots,n$$

如果有以上形式的，则称 X 服从二项分布，记作 $X\sim B(n,p)$。
显然

$$\sum_{k=0}^{n} P\left(X=k\right)=\sum_{k=0}^{n} C_n^k p^k(1-p)^{n-k}=\left[p+(1-p)\right]^n=1$$

特别的，当 $n=1$ 时，就是 $(0-1)$ 分布

§20.2.2 超几何分布

一般地，设一批产品有 N 件，其中有 M 件是次品，从 N 件中随机抽取 n 件 (不放回)，用 X 表示抽取 n 件的次品数，X 分布列为 $P(X=k)=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, k=m,m+1,m+2,\ldots,r$ ，其中

$$n,N,M\in\mathbb{N}^+, M\leqslant N, n\leqslant N$$
$$m=\{0,n-N+M\}_{\max}, r=\{n,M\}_{\min}$$

如过有以上形式，则称 X 服从超几何分布，记 $X\sim H(N,n,M)$。

令 $p=\frac{M}{N}$，则 p 是次品率，$\frac{X}{n}$ 是抽取 n 件中的次品率。

§20.3 随机变量的分布函数

分布函数是高中教材中没有明确定义的，但这和正态密度函数的概念息息相关。

对于非离散型随机变量 X，由于其可能的取值不能一一列举，因而就不能像研究离散型随机变量一样用分布列来研究；

另外，我们通常遇到的非离散型随机变量取任一实数值的概率都为 0，因此我们转而研究随机变量所取的值落在一个区间 $(x_1,x_2]$ 的概率：$P\{x_1<X\leqslant x_2\}$，但由于

$$P\{x_1<X\leqslant x_2\}=P\{X\leqslant x_2\}-P\{X\leqslant x_1\}$$

所以我们只需要知道 $P\{X\leqslant x_2\}$ 和 $P\{X\leqslant x_1\}$ 就可以了，下面引入随机变量的分布函数的概念。

分布函数的定义

设 X 是一个随机变量，x 是任意实数，函数

$$F(x)=P\{X\leqslant x\}, -\infty<x<\infty$$

$F(x)$ 称为 X 的**分布函数**

对于任意实数 $x_1,x_2(x_1<x_2)$，有

$$\begin{aligned}P\{x_1<X\leqslant x_2\}&=P\{X\leqslant x_2\}-P\{X\leqslant x_1\}\\&=F(x_2)-F(x_1)\end{aligned}$$

因此，若已知随机变量 X 的分布函数，那么就可以知道 X 落在区间 $(x_1, x_2]$ 的概率；分布函数是一个普通的函数，通过它，我们将能用数学分析的方法来研究随机变量。

若将 X 看成数轴上的随机点坐标，那么，分布函数 $F(x)$ 在 x 处的函数值就表示 X 落在区间 $(-\infty, x]$ 上的概率；

分布函数具有以下的基本性质：

1. $F(x)$ 是一个不减函数

   对于任意实数 $x_1, x_2(x_1 < x_2)$ 有

$$F(x_2) - F(x_1) = P\{x_1 < X \leqslant x_2\} \geqslant 0$$

2. $0 \leqslant F(x) \leqslant 1$，且

$$F(-\infty) = \lim_{x \to -\infty} = 0, F(\infty) = \lim_{x \to \infty} = 1$$

3. $F(x+0) = F(x)$，即 $F(x)$ 是右连续的。

反之，可证具备以上性质的函数 $F(x)$ 必是某个随机变量的分布函数。

§20.4 连续型随机变量

定义

一般，如果对于随机变量 X 的分布函数 $F(x)$，存在非负可积函数 $f(x)$，使对于任意实数 x 有

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$$

则称 X 为**连续型随机变量**，$f(x)$ 称为 X 的**概率密度函数**，简称**概率密度**。

由定义可知，概率密度 $f(x)$ 具有以下性质：

1. $f(x) \geqslant 0$

2. $\int_{-\infty}^{\infty} x\mathrm{d}x = 1$

3. 对于任意函数实数 $x_1, x_2(x_1 \leqslant x_2)$，

$$P\{x_1 < X \leqslant x_2\} = F(x_2) - F(x_1) = \int_{x_2}^{x_1} f(x)\mathrm{d}x$$

4. 若 $f(x)$ 在点 x 处连续，则有 $F'(x) = f(x)$

反之，若 $f(x)$ 具备性质 1、2，引入 $G(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$，它是某一随机变量 X 的的分布函数，$f(x)$ 是 X 概率密度。

需要指出的是，对于连续型随机变量 X 来说，它取任一指定的实数值 a 的概率均为 0，即 $P\{X = a\} = 0$，事实上，设 X 的分布函数为 $F(x), \delta x > 0$，则由 $\{X = a\} \subset \{a - \delta x < X \leqslant a\}$ 得

$$0 \leqslant P\{X = a\} \leqslant P\{a - \delta x < X \leqslant a\} = F(a) - F(a - \delta x)$$

在上述不等式中令 $\delta x \to 0$，并注意到 X 为连续型随机变量，其分布函数 $F(x)$ 是连续的，即得

$$P\{X = a\} = 0$$

据此，在计算连续型随机变量落在某一区间的概率时，可以不必区分该区间的开闭或半开，例如有

$$P\{a < X \leqslant b\} = P\{a \leqslant X \leqslant b\} = P\{a < X < b\}$$

在这里，事件 $\{X = a\}$ 并非不可能事件，但有 $P\{X = a\} = 0$，就是说，若 A 是不可能事件，则有 $P(A) = 0$；反之，若 $P(A) = 0$，并不一定意为着 A 是不可能事件。

以后当我们提到一个随机变量 X 的"概率分布"时，指的是它的分布函数；或者，当 X 是连续型随机变量时，指的是它的概率密度，当 X 是离散型随机变量时，指得是它的分布列。

下面我们介绍高中唯一涉及的连续型随机变量————**正态分布**

定义

若连续型随机变量 X 的概率密度为

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

其中 $\mu \in \mathbb{R}, \sigma > 0$ 为常数，则称 X 服从参数为 μ, σ **正态分布或高斯分布**，记为 $X \sim N(\mu, \sigma^2)$。♡

显然 $f(x) \geqslant 0$，下面来证明 $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$，令 $\frac{x-\mu}{\sigma} = t$，得到

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{\frac{-t^2}{2}}\mathrm{d}t$$

记

$$I = \int_{-\infty}^{\infty}\mathrm{e}^{\frac{-t^2}{2}}\mathrm{d}t$$

则有

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{e}^{\frac{-(t^2+u^2)}{2}}\mathrm{d}t\mathrm{d}u$$

利用极坐标将它们化成累次积分，得到

$$I^2 = \int_0^{2\pi}\int_0^{\infty} r\mathrm{e}^{\frac{-r^2}{2}}\mathrm{d}r\mathrm{d}\theta = 2\pi$$

而 $I > 0$，故 $I = \sqrt{2\pi}$，既有

$$\because \int_{-\infty}^{\infty}\mathrm{e}^{\frac{-t^2}{2}}\mathrm{d}t = \sqrt{2\pi}$$

$$\therefore \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{\frac{-t^2}{2}}\mathrm{d}t = 1$$

参数 μ,σ 的意义在下一节说明，$f(x)$ 的图像具有以下性质：

1. 曲线关于 $x=\mu$ 对称，这表明对于任意 $h>0$ 有

$$P\{\mu-h<X\leqslant\mu\}=P\{\mu<X\leqslant\mu+h\}$$

2. 当 $x=\mu$ 时取得最大值

$$f(\mu)=\frac{1}{\sigma\sqrt{2\pi}}$$

x 离 μ 越远，$f(x)$ 的值越小。

在 $x=\mu\pm\sigma$ 处曲线有拐点，曲线以 Ox 轴为渐近线；$f(x)$ 的最值以及拐点可以通过求函数的一、二阶导来证明，这里作为读者的课后习题。

另外如果固定 σ，改变 μ 值，图像只会水平移动，而不改变形状，因此 μ 也被称作**位置参数**。如果固定 μ，改变 σ，由 $f(\mu)=\frac{1}{\sigma\sqrt{2\pi}}$ 可知，σ 越小时图像变得越尖，因而 X 落在 μ 附近的概率越大。

不难知道，X 的分布函数为

$$F(x)=\frac{1}{\sqrt{\sigma 2\pi}}\int_{-\infty}^{x}\mathrm{e}^{\frac{-(t-\mu)^2}{2\sigma^2}}\mathrm{d}t$$

特别的, 当 $\mu=0,\sigma=1$ 时，称随机变量 X 服从**标准正态分布**，其概率密度和分布函数分别用 $\varphi(x),\Phi(x)$ 表示，即有

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$$
$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{\frac{-t^2}{2}}\mathrm{d}t$$

易知，$\Phi(-x)=1-\Phi(x)$。

一般，若 $X\sim N(\mu,\sigma^2)$，我们只需通过一个线性变换就能将它化成标准正态分布

引理

若 $X\sim N(\mu,\sigma^2)$，则 $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$

证明：$Z=\frac{X-\mu}{\sigma}$ 的分布函数为

$$\begin{aligned}P\{Z\leqslant x\}&=P\{\frac{X-\mu}{\sigma}\leqslant x\}=P\{X\leqslant\mu+\sigma x\}\\&=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mu+\sigma x}\mathrm{e}^{\frac{-(t-\mu)^2}{2\sigma^2}}\mathrm{d}t\end{aligned}$$

令 $\frac{t-\mu}{\sigma}=u$，得到

$$P\{Z\leqslant x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{\frac{-u^2}{2}}\mathrm{d}u=\Phi(x)$$

由此知 $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$

□

设 $X\sim N\left(\mu,\sigma^2\right)$，有 $\Phi(x)$ 的函数表[2]能得到

$$\begin{aligned}P(\mu-\sigma\leqslant X\leqslant\mu+\sigma)&=\Phi(1)-\Phi(-1)\\&=2\Phi(1)-1\approx0.6827\\P(\mu-2\sigma\leqslant X\leqslant\mu+2\sigma)&=\Phi(2)-\Phi(-2)\approx0.9545\\P(\mu-3\sigma\leqslant X\leqslant\mu+3\sigma)&=\Phi(3)-\Phi(-3)\approx0.9973\end{aligned}$$

可以看出，尽管正态变量的取值是 $(-\infty,\infty)$，但它落在 $(\mu-3\sigma\leqslant X\leqslant\mu+3\sigma)$ 内几乎是肯定的事，这被称为“3σ 原则”或“3σ 法则”。

§20.5 随机变量的数字特征

§20.5.1 数学期望

定义

设离散型随机变量 X 的分布列为

$$P\{X=x_k\}=p_k,k=1,2,\ldots$$

若级数

$$\sum_{k=1}^{\infty}x_kp_k$$

绝对收敛，则称级数 $\sum_{k=1}^{\infty}x_kp_k$ 的和为随机变量 X 的**数学期望**，记为 $E(X)$，即

$$E(X)=\sum_{k=1}^{\infty}x_kp_k$$

设连续型随机变量 X 的概率密度为 $f(x)$，若积分

$$\int_{-\infty}^{\infty}xf(x)\mathrm{d}x$$

绝对收敛，则称积分 $\int_{-\infty}^{\infty}xf(x)\mathrm{d}x$ 的值为随机变量 X 的**数学期望**，记为 $E(X)$，即

$$E(X)=\int_{-\infty}^{\infty}xf(x)\mathrm{d}x$$

数学期望简称**期望**，又称**均值**。

♡

[2]人们编制了 $\Phi(x)$ 的函数表，可供查用

在高中，只学习离散型随机变量的数字特征，高中对于数学期望的定义是：

$$E(X)=\sum_{i=1}^{n}x_ip_i$$

定理

设 Y 是随机变量 X 的函数：$Y=g(X)$（g 是连续函数）

1. 如果 X 是离散型随机变量，它的分布列为 $P\{X=x_k\}=p_k,k=1,2,\dots$ 若

$$\sum_{k=1}^{\infty}g(x_k)=p_k$$

绝对收敛，则有

$$E(Y)=E[g(X)]=\sum_{k=1}^{\infty}g(x_k)=p_k$$

2. 如果 X 是连续型随机变量，它的概率密度为 $f(x)$，若

$$\int_{-\infty}^{\infty}g(x)f(x)\mathrm{d}x$$

绝对收敛，则有

$$E(Y)=E[g(X)]=\int_{-\infty}^{\infty}g(x)f(x)\mathrm{d}x$$

♠

这个定理的重要意义在于当我们求 $E(Y)$ 时，不必算出 Y 的分布列或概率密度，而只需要利用 X 的分布列或概率密度就可以了。

推论

上述定理还可以推广到两个或两个以上随机变量的函数情况，

设 Z 是随机变量 X,Y 的函数 $Z=g(X,Y)$（g 是连续函数），那么 Z 是一个一维随机变量，若二维随机变量 (X,Y) 的概率密度为 $f(x,y)$，则有：

$$E(Z)=E[g(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$$

这里设上式右边的积分绝对收敛，又若 (X,Y) 为离散型随机变量，其分布列为 $P\{X=x_i,Y=y_j\}=p_{ij}(i,j=1,2,\dots)$ 则有

$$E(Z)=E[g(X,Y)]=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}g(x_i,y_i)p_{ij}$$

这里也设上式右边的级数绝对收敛。

◇

现在来证明数学期望的几个重要性质，以下设所遇到的随机变量的数学期望存在。

1. 设 C 是常数，则有 $E(C)=C$

2. 设 X 是一个随机变量，C 是常数，则有 $E(CX) = CE(X)$

3. 设 X, Y 是两个随机变量，则有

$$E(X+Y) = E(X) + E(Y)$$

这一性质可以推广到任意有限个随机变量之和的情况。

4. 设 X, Y 是相互独立的随机变量，则有

$$E(XY) = E(X)E(Y)$$

这一性质可以推广到任意有限个相互独立的随机变量之积的情况。

这里只对离散型随机变量的情况加以证明，性质 1、2 显然成立；
设二维随机变量 (X, Y) 的分布列为 $P\{X = x_i, Y = y_j\} = p_{ij}(i, j = 1, 2, \ldots)$

$$\begin{aligned} E(X+Y) &= \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}(x_i + y_j)p_{ij} \\ &= \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}x_i p_{ij} + \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}y_i p_{ij} \\ &= E(X) + E(Y) \end{aligned}$$

□

性质 3 得证；

又若 X 和 Y 相互独立

$$\begin{aligned} E(XY) &= \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}x_i y_i p_{ij} \\ &= \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}x_i p_i y_i p_j \\ &= \sum_{i=1}^{\infty}x_i p_i \sum_{j=1}^{\infty}y_j p_j \\ &= E(X)E(Y) \end{aligned}$$

□

性质 4 得证。

在高中阶段，只需掌握

$$E(aX+b) = aE(X) + b$$

§20.5.2 方差

定义

设 X 是一个随机变量，若 $E\{[X-E(X)]^2\}$ 存在，则称 $E\{[X-E(X)]^2\}$ 为 X 的**方差**，记为 $D(X)$ 或 $\mathrm{Var}(X)$，即

$$D(X)=\mathrm{Var}(X)=E\{[X-E(X)]^2\}$$

在应用上还引入量 $\sqrt{D(X)}$，记为 $\sigma(X)$，称为**标准差**或**均方差**

按照定义，随机变量 X 的方差表达了 X 的取值与其数学期望的偏离程度，因此 $D(X)$ 是刻画 X 取值分散程度的一个量，$D(X)$ 越小意为着 X 的取值比较集中在 $E(X)$ 的附近。

对于离散型随机变量有

$$D(X)=\sum_{k=1}^{\infty}[x_k-E(X)]^2 p_k$$

其中 $P\{X=x_k\}=p_k, k=1,2,\ldots$ 是 X 的分布列；

对于连续型随机变量有

$$D(X)=\int_{-\infty}^{\infty}[x-E(X)]^2 f(x)\mathrm{d}x$$

其中 $f(x)$ 就是 X 的概率密度。

随机变量 X 的方差可以按下列公式计算

定理

$$D(X)=E(X^2)-[E(X)]^2$$

证明：由数学期望的性质可知

$$\begin{aligned}D(X)&=E[X-E(X)]^2=E\{X^2-2XE(X)+[E(X)]^2\}\\&=E(X^2)-2E(X)E(X)+[E(X)]^2\\&=E(X^2)-[E(X)]^2\end{aligned}$$

□

现在来证明方差的几个重要性质

1. 设 C 是常数，则 $D(X)=0$

2. 设 X 是随机变量，C 是常数，则有

$$D(CX)=C^2D(X),\ D(X+C)=D(X)$$

3. 设 X, Y 是两个随机变量，则有

$$D(X+Y) = D(X) + D(Y) + 2E\{(X - E(X))(Y - E(Y))\}$$

特别的，若 X, Y 相互独立，则有

$$D(X+Y) = D(X) + D(Y)$$

这一性质可以推广到任意有限个相互独立的随机变量之积的情况。

4. $D(X) = 0$ 的充要条件是 X 以概率 1 取常数 $E(X)$，即

$$P\{X = E(X)\} = 1$$

证明：

对于性质 1

$$D(C) = E\{[C - E(C)]^2\} = 0$$

□

对于性质 2

$$\begin{aligned} D(CX) &= E\{[CX - E(CX)]^2\} \\ &= C^2 E\{[X - E(X)]^2\} \\ &= C^2 D(X) \end{aligned}$$

以及

$$\begin{aligned} D(X+C) &= E\{[X + C - E(X+C)]^2\} \\ &= E\{[X - E(X)]^2\} \\ &= D(X) \end{aligned}$$

□

对于性质 3

$$\begin{aligned} D(X+Y) &= E\{[X + Y - E(X+Y)]^2\} \\ &= E\{[(X - E(X)) + (Y - E(Y))]^2\} \\ &= E\{[X - E(X)]^2\} + E\{[Y - E(Y)]^2\} \\ &\quad + 2E\{[X - E(X)][Y - E(Y)]\} \\ &= D(X) + D(Y) + 2E\{[X - E(X)][Y - E(Y)]\} \end{aligned}$$

注意到

$$\begin{aligned}&2E\{[X-E(X)][Y-E(Y)]\}\\&=2E\{XY-XE(Y)-YE(X)+E(X)E(Y)\}\\&=2\{E(XY)-E(X)E(Y)-E(Y)E(X)+E(X)E(Y)\}\\&=2\{E(XY)-E(X)E(Y)\}\end{aligned}$$

若 X,Y 相互独立，可知 $2\{E(XY)-E(X)E(Y)\}=0$，于是

$$D(X+Y)=D(X)+D(Y)$$

□

在高中阶段，只需掌握

$$\begin{aligned}D(X)&=E(X^2)-[E(X)]^2\\D(aX+b)&=a^2D(X)\end{aligned}$$

设随机变量 X 具有两点分布，其分布列为

$$P\{X=0\}=1-p,\quad P\{X=1\}=p$$

则有

$$E(X)=p,\quad D(X)=p(1-p)$$

证明：

$$\begin{aligned}E(X)&=0\cdot(1-p)+1\cdot p=p\\E(X^2)&=0^2\cdot(1-p)+1^2\cdot p=p\\D(X)&=E(X^2)-[E(X)]^2=p(1-p)\end{aligned}$$

□

知道数学期望以及方差的概念后，现在来求**二项分布的数学期望与方差**；

定理

设随机变量 $X\sim B(n,p)$，$E(X),D(X)$ 分别为数学期望与方差，有：

$$\begin{aligned}E(X)&=np\\D(X)&=np(1-p)\end{aligned}$$

♠

证明：

由二项分布的定义可知，随机变量 X 是 n 重伯努利试验中事件 A 发生的次数，且在每次试验中 A 发生

的概率为 p，引入变量

$$X_k = \begin{cases} 1, & A\text{在第 } k \text{ 次试验发生}, \\ 0, & A\text{在第 } k \text{ 次试验不发生}, \end{cases} \quad k = 1, 2, \ldots, n$$

易知

$$X = X_1 + X_2 + \cdots + X_n \tag{1}$$

由于 X_k 只依赖于第 k 次试验，而各次试验相互独立，于是 $X_1, X_2, \ldots, X_n$ 相互独立，又知 $X_k, k = 1, 2, \ldots, n$ 服从同一两点分布

X_k	0	1
p_k	$1-p$	p

(1) 式表明，以 n, p 为参数的二项分布变量，可分解成为 n 个相互独立且都服从以 p 为参数的两点分布的随机变量之和；所以有 $E(X_k) = p,\ D(X_k) = p(1-p), k = 1, 2, \ldots, n$ 故

$$\begin{aligned} E(X) &= E(\sum_{k=1}^{n} X_k) \\ &= \sum_{k=1}^{n} E(X_k) = np \end{aligned}$$

又由于 $X_1, X_2, \ldots, X_n$ 相互独立，则

$$\begin{aligned} D(X) &= D(\sum_{k=1}^{n} X_k) \\ &= \sum_{k=1}^{n} D(X_k) = np(1-p) \end{aligned}$$

综上，

$$E(X) = np, \quad D(X) = np(1-p)$$

□

下面来求**超几何分布**的**数学期望**与**方差**；

定理

设随机变量 $X \sim H(N, n, M)$，$P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$，$k = 0, 1, 2, \ldots, n$，$E(X), D(X)$ 分别是数学期望和方差，有

$$\begin{aligned} E(X) &= \frac{nM}{N} \\ D(X) &= \frac{nM(N-n)(N-M)}{N^2(N-1)} \end{aligned}$$

♠

证明：

引理

$$k\binom{M}{k} = M\binom{M-1}{k-1}$$
$$kC_M^k = MC_{M-1}^{k-1}$$

组合数的意义展开即证。

先证明 $E(X)$：

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} \frac{kC_M^k C_{N-M}^{n-k}}{C_N^n} \\ &= \frac{M}{C_N^n}\sum_{k=1}^{n}(C_{M-1}^{k-1}C_{N-M}^{n-k}) \\ &= \frac{MC_{N-1}^{n-1}}{C_N^n} \\ &= \frac{nM}{N} \end{aligned}$$

再来，对于 $E(X^2)$：

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{n} \frac{k^2C_M^k C_{N-M}^{n-k}}{C_N^n} \\ &= \frac{M}{C_N^n}\sum_{k=1}^{n}[kC_{M-1}^{k-1}C_{N-M}^{n-k}] \\ &= \frac{M}{C_N^n}\left(\sum_{k=1}^{n}\left[(k-1)C_{M-1}^{k-1}C_{N-M}^{n-k}\right] + \sum_{k=1}^{n}\left[C_{M-1}^{k-1}C_{N-M}^{n-k}\right]\right) \\ &= \frac{M}{C_N^n}\left(\sum_{k=1}^{n}\left[(M-1)C_{M-2}^{k-2}C_{N-M}^{n-k}\right] + \sum_{k=1}^{n}\left[C_{M-1}^{k-1}C_{N-M}^{n-k}\right]\right) \\ &= \frac{M}{C_N^n}\left[(M-1)C_{N-2}^{n-2} + C_{N-1}^{n-1}\right] \\ &= \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{nM}{N} \end{aligned}$$

因此可以求得 $D(X)$：

$$\begin{aligned} D(X) &= E(X^2) - [E(X)]^2 \\ &= \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{nM}{N} - \frac{n^2M^2}{N^2} \\ &= \frac{nM(N-n)(N-M)}{N^2(N-1)} \end{aligned}$$

□

§20.5.3 协方差

对于二维随机变量 (X,Y)，除了讨论 X 与 Y 的数学期望与方差以外，还需要讨论 X 与 Y 之间相互关系的数字特征；由方差性质的证明，我们已经知道了，如果两个随机变量 X 和 Y 是相互独立的，则

$$E\{[X-E(X)][Y-E(Y)]\}=0$$

这意为这当 $E\{[X-E(X)][Y-E(Y)]\}\neq 0$ 时，X 与 Y 不相互独立，而是存在一定关系的。

定义

$E\{[X-E(X)][Y-E(Y)]\}$ 称为随机变量 X 与 Y 的**协方差**，记为 $\mathrm{Cov}(X,Y)$，即

$$\mathrm{Cov}(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$$

由定义不难知道：

$$\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X),\qquad \mathrm{Cov}(X,X)=D(X)$$

在方差性质的证明中已经证明了

$$D(X+Y)=D(X)+D(Y)+2\mathrm{Cov}(X,Y)$$

若将 $\mathrm{Cov}(X,Y)$ 的定义式展开，易得

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)$$

协方差具有以下性质：

1. $\mathrm{Cov}(aX,bY)=ab\mathrm{Cov}(X,Y)$，$a,b$ 是常数；

2. $\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$。

§20.6 马尔科夫过程

鉴于 23 年杭二模和新一卷中都出现了有关“马尔可夫链”背景的题，在此简单介绍一些马尔可夫过程的知识，虽然随机过程是数理统计相对独立的，但在这版中我还是放到概率论里；同时，本文不会详细说明随机过程的内容，否则对于高中生会很冗长。

先引入**马尔可夫性或无后效性**：过程（或系统）在时刻 t_0 所处的状态为已知的条件下，过程在时刻 $t>t_0$ 所处状态的条件分布与过程在时刻 t_0 之前所处的状态无关，通俗的说，就是在已经知道过程“现在”的条件下，其“将来”不依赖于“过去”。

现在用分布函数来表述马尔可夫性，设随机过程 $\{X(t),t\in T\}$ 的状态空间为 I；如果对时间 t 的任意 n 个值 $t_1<t_2<\cdots<t_n,n\geqslant 3,t_i\in T$，在条件 $X(t_i)=x_i,x_i\in I,i=1,2,\ldots,n-1$ 下，$X(t_n)$ 的条件分布函数恰

等于在条件 $X(t_{n-1})=x_{n-1}$ 下 $X(t_n)$ 的条件分布函数，即：

$$\begin{aligned}&P\{X(t_n)\leqslant x_n|X(t_1)=x_1,X(t_2)=x_2,\dots,X(t_{n-1})=x_{n-1}\}\\&=P\{X(t_n)\leqslant x_n|X(t_{n-1})=x_{n-1}\},x_n\in\mathbb{R}\end{aligned}$$

则过程 $\{X(t),t\in T\}$ 具有马尔可夫性或无后效性，并且称此过程为**马尔可夫过程**。

时间和状态都是离散的马尔可夫过程称为**马尔科夫链**，简称**马氏链**，记为

$$\{X_n=X(n),n=0,1,\dots\}$$

它可以看作在时间集 $T_1=\{0,1,2,\dots\}$ 上对离散状态的马尔可夫过程相继观察的结果，我们约定记链的状态空间为 $I=\{a_1,a_2,\dots\},a_i\in\mathbb{R}$，在链的情形，马尔可夫性通常用条件分布列来表示，即对任意的正整数 n,r 和 $0\leqslant t_1<t_2<\cdots<t_r<m;t_i,m,n+m\in T_1$，有：

$$\begin{aligned}&P\{X_{m+n}=a_j|X_{t_1}=a_{i_1},X_{t_2}=a_{i_2},\dots,X_{t_r}=a_{i_r},X_m=a_i\}\\&=P\{X_{m+n}=a_j|X_m=a_i\}\end{aligned}$$

其中 $a\in I$ 记上式右端为 $P_{ij}(m,m+n)$，我们称条件概率

$$P_{ij}(m,m+n)=P\{X_{m+n}=a_j|X_m=a_i\} \qquad ①$$

为马氏链在时刻 m 处于状态 a_i 条件下，在时刻 $m+n$ 转移到状态 a_j 的**转移概率**。

由于链在时刻 m 从任何一个状态 a_i 出发，到另一时刻 $m+n$，必然转移到 $a_1,a_2,\cdots$ 诸状态中的某一个，所以

$$\sum_{j=1}^{+\infty}P_{ij}(m,m+n)=1,i=1,2,\cdots$$

由转移概率组成的矩阵 $\boldsymbol{P}(m,m+n)=(P_{ij}(m,m+n))$ 称为马氏链的**转移概率矩阵**。由上式可知，此矩阵的每一行元之和等于 1。

当转移概率 $P_{ij}(m,m+n)$ 只与 i,j 及时间间距 n 有关时，把它记为 $P_{ij}(n)$，即

$$P_{ij}(m,m+n)=P_{ij}(n)$$

并称此转移概率具有**平稳性**，同时也称此链是**齐次的或时齐的**，以下我们限于讨论齐次马氏链。

在马氏链为齐次的情形下，由 ① 式定义的转移概率

$$P_{ij}(n)=P\{X_{m+n}=a_j|X_m=a_i\}$$

称为马氏链的n **步转移概率**，$\boldsymbol{P}(n)=(P_{ij}(n))$ 为n **步转移概率矩阵**，在以下的讨论中特别重要的是**一步转移概率**

$$p_{ij}=P_{ij}(1)=P\{X_{m+1}=a_j|X_m=a_i\}$$

或由它们组成的**一步转移概率矩阵**。

$$X_m\text{的状态}\begin{array}{cc} & \begin{matrix} a_1 & a_2 & \cdots & a_j & \cdots \end{matrix} \\ \begin{matrix} a_1 \\ a_2 \\ \vdots \\ a_i \\ \vdots \end{matrix} & \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1j} & \cdots \\ p_{21} & p_{22} & \cdots & p_{2j} & \cdots \\ \vdots & \vdots & & \vdots & \\ p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots \\ \vdots & \vdots & & \vdots & \end{bmatrix} \end{array} = \boldsymbol{P}(1) \xlongequal{\text{记成}} \boldsymbol{P}$$

在上述矩阵的左侧和上边标上状态 $a_1, a_2, \cdots$ 是为了显示 p_{ij} 的由状态 a_i 经一步转移到状态 a_j 的概率。

§20.6.1 一维随机游走

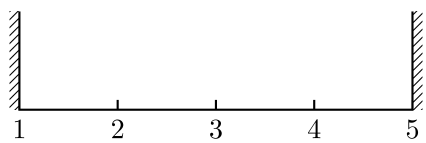


图 20.2: 一维随机游走

设一醉汉 Q（或看作一随机游走的质点），在如图所示的直线的点集 $I = \{1, 2, 3, 4, 5\}$ 上随机游走，且仅在 1 秒、2 秒等时刻发生游走，游走的概率规则为：如果 Q 现在位于点 i $(1 < i < 5)$，则下一时刻以 $\frac{1}{3}$ 的概率向左或向右移动一格，$\frac{1}{3}$ 的概率留在原处；如果 Q 现在位于 1（或 5）这点上，则下一时刻就以概率 1 移动到 2（或 4）这一点上，1 和 5 这两个点称为**反射壁**，上面这种游走称为**带有两个反射壁的随机游走**。

若以 X_n 表示时刻 n 时 Q 的位置，不同的位置就是 X_n 的不同状态，那么 $\{X_n, n = 0, 1, 2, \cdots\}$ 是一随机过程，状态空间就是 I，而且当 $X_n = i, i \in I$ 为已知时，X_{n+1} 所处的状态的概率分布只与 $X_n = i$ 有关，而与 Q 在时刻 n 以前如何到达 i 是完全无关的，所以 $\{X_n, n = 0, 1, 2, \cdots\}$ 是一马氏链，而且还是齐次的，它的一步转移概率和一步转移概率矩阵分别为：

$$p_{ij} = P\{X_{n+1} = j | X_n = i\} = \begin{cases} \frac{1}{3},\ j = i-1, i, i+1, 1 < i < 5, \\ 1,\ i = 1, j = 2\text{或}i = 5, j = 4 \\ 0,\ |j - i| \geqslant 2 \end{cases}$$

和

$$\boldsymbol{P} = \begin{array}{cc} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

如果把 1 这一点改为**吸收壁**，就是说 Q 一旦到达 1 这一点，则就永远留在点 1 上，此时，相应链的转移概率矩阵只需把 $\boldsymbol{P}$ 中第一横行改为 $(1, 0, 0, 0, 0)$；总之，改变游走的概率，就可得到不同方式随机游走和相应的马氏链。

第二十一章　统计

§21.1　统计中一些基本概念

总体：包含所研究的全部个体的集合。跟据其包含的单位数目是否可数可以分为**有限总体**和**无限总体**，区分有限和无限主要是为了判别在抽样中每次抽取是否独立。

个体：组成总体的每一个调查对象。

样本：从总体抽出的那部分个体。从总体中抽取的一部分元素的集合，构成样本的元素的数目称为**样本量**。

全面调查/普查：对每一个调查对象都进行调查。

由于普查的成本较大，因此通常会抽取一部分进行调查即：

抽样调查：从总体中抽取一部分个体进行调查，并以此为依据对总体的情况做出估计和推断。

§21.1.1　几种抽样方式

概率抽样，也称**随机抽样**，是指遵循随机原则进行的抽样，总体中每个个体都有一定的机会被选入样本。需要注意的是概率抽样和等概率抽样是两个不同的概念。

下面介绍几种概率抽样方式。

一、简单随机抽样

一个总体含 N 个个体，从中随机地、一个个地抽取 $n\,(n\in[1,N))$ 个个体做为样本，每个个体入样概率是相等的。(除特殊说明，一般简单随机抽样指不放回简单随机抽样)，常用以下方法：

1. 抽签法
2. 随机数法

二、分层随机抽样

将抽样单位按某种特征或某种规则划分为不同的层，然后从不同的层中独立、随机地抽取样本。

如果每层样本的量与层的大小成比例，那么称这种样本量的分配方式为“比例分配”。

三、系统抽样

将总体中所有个体按一定顺序排列，在规定的范围内随机地抽取一个单位作为初始单位，然后按事先规定好的规则确定其他样本单位。

典型的系统抽样是从数字 1 k 之间随机抽取一个单位作为初始单位，以后依次取 $r+k, r+2k, \cdots$。

§21.2　数字特征

总体均值/总体平均值：总体有 N 个个体，变量值分别为 $Y_i\,(i=1,2,\ldots,N)$，则称 $\overline{Y}=\frac{1}{N}\sum_{i=1}^{N}Y_i$ 为总体均值。

如果总体的 N 个变量中，不同的值共有 $k\,(k\leqslant N)$ 个，不妨记为 $Y_i\,(i=1,2,\ldots,k)$ 其中 Y_i 出现的频数为 f_i，则可以写成加权平均数 $\overline{Y}=\frac{1}{N}\sum_{i=1}^{k}f_iY_i$

如果从总体中抽取一个容量为 n 的样本，它们的变量值为 $y_i\,(i=1,2,\ldots,n)$，则称 $\overline{Y}=\frac{1}{n}\sum_{i=1}^{n}y_i$ 为样本均值。

中位数：记有 n 个数据，当数据量是偶数时，中位数为第 $\frac{n}{2}$ 和 $\frac{n}{2}+1$ 个数的平均数；当 n 为奇数时，中位数等于第 $\frac{n+1}{2}$ 个数。

众数：出现频率最大的数。

平均数：$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$

方差：$s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2=\frac{1}{n}\left(\sum_{i=1}^{n}x_i^2\right)-\overline{x}$

标准差：$\sqrt{s^2}=s$

关于 $\sum$ 的计算我不过多赘述，很多都是在概率论中推到过的结论。

下面讨论两组数据 x_i,y_i，它们的平均数以及方差的关系：

$$\sum_{i=1}^{n}(ax_i+by_i)=a\sum_{i=1}^{n}x_i+b\sum_{i=1}^{n}y_i$$

如果 $y_i=ax_i+b$，容易知道：

$$\begin{aligned}\overline{y}&=a\overline{x}+b\\S_y&=|a|S_x\end{aligned}$$

再来看看更加一般的情况：

设有两组数据 $x_i,y_j(i=1,2,\ldots,n,j=1,2,\ldots,m)$ 它们平均数分别为

$$\begin{cases}\overline{x}= & \frac{1}{n}\sum_{i=1}^{n}x_i\\\overline{y}= & \frac{1}{m}\sum_{i=1}^{m}y_j\end{cases}$$

方差分别为

$$\begin{cases}S_x^2= & \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\\S_y^2= & \frac{1}{m}\sum_{j=1}^{m}(y_j-\overline{y})^2\end{cases}$$

设总体平均数为 $\overline{a}$，总体方差为 b^2，推出：

$$\begin{aligned}\overline{a}&=\frac{n\overline{x}+m\overline{y}}{m+n}\\b^2&=\frac{1}{m+n}\left[nS_x^2+mS_y^2+n(\overline{x}-\overline{a})^2+m(\overline{y}-\overline{a})^2\right]\end{aligned}$$

方差比较难记，但不会常考到，可以分为两部分来理解：

1，$nS_x^2+mS_y^2$ 是它们的层内方差

2，$n(\overline{x}-\overline{a})^2+m(\overline{y}-\overline{a})^2$ 是它们的层外方差

这个可以自己推一下，对记忆有帮助，但就是比较麻烦。

§21.3 数据的图表展示

频率分布直方图 使用“频率分布表”和“频率分布直方图”来整理和表示数据（频率 = 频数/样本数）

e.g.

如图所示，每个小长方形的面积即是改段数据出现的频率，首先介绍对于一组数据如何画频率分布直方图

画图步骤：

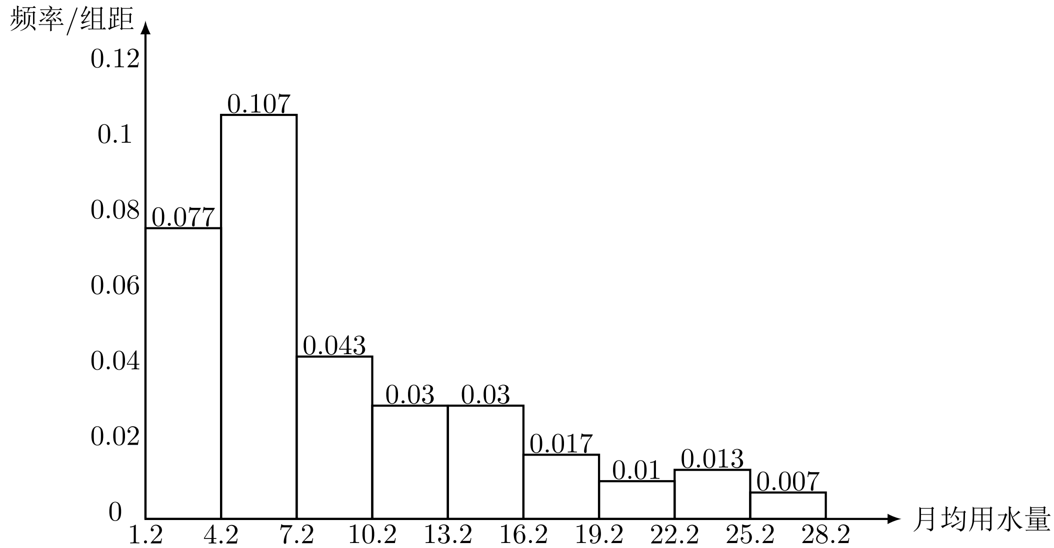


图 21.1: 频率分布直方图

1. 求极差：$X_{\max} - X_{\min}$。
2. 决定组数与组距：分组可以是等距的或不等距的，一般组距相等，且力求取整；组数 = 极差/组距。
3. 将数据分组。
4. 列频率分布表：如第 n 组频数/样本容量。
5. 画频率分布直方图：y 轴为频率/组距，x 轴为数据的意义，小长方形面积 = 组距 · 频率/组距 = 频率。

频率分布直方图可以帮助我们观察数据的集中趋势，也可以估计一些数字特征，如上例，通过图像易知 $(4.2, 7.2)$ 中数据出现的频数最大，因此估计众数为 $\frac{4.2+7.2}{2} = 5.7$。

为了展示更多的数字特征，我们引入**第 p 百分位数**。

一组数据的第 p 百分位数是这样一个值，它使得这组数据中至少有 $p\%$ 的数据小于或等于这个值，且至少有 $(100-p)\%$ 的数据大于或等于这个值.

我们可以通过以下步骤计算一组数据的第 p 百分位数。

1. 将数据从小到大排列。
2. 计算 $i = n \times p\%$，其中 n 为总体，$p\%$ 为百分位数。
3. (a) 当 i 是整数：第 i 项与 $(i+1)$ 项之和的平均数。
   (b) 当 i 不是整数：取比 i 大的第一个整数 j，第 p 百分位数为 j

显然中位数就是第 50 百分位数。

当只有表格没有具体数值时用以下方法求：设样本数据的 $p\%$ 分位数为 i

1. 先从低到高相加频率，直到第 n 组包含 $p\%$。
2. 设第 n 组分组为 $[x, y)$，x 对应数据 a，y 对应数据 b，i 对应数据 c，则

$$x + \frac{c-a}{b-a} \cdot (y-x) = i$$

下面就用上例来演示

e.g. 已知频率分布直方图，求数据的中位数，以及第 67 百分位数。

先求中位数，第一步求每组频率 逐个相加直到包含 0.5，易知 $(1.2, 7.2)$ 的频率为 $0.231 + 0.321 = 0.552$，

第 n 组	1	2	3	4	5	6	7	8	9
频率	0.231	0.321	0.129	0.09	0.09	0.051	0.03	0.039	0.021

图 21.2: 数据频率

而 $(1.2, 4.2)$ 的频率为 0.231，因此中位数在第二组中，即 $(4.2, 7.2)$ 中。

随后，中位数在第二组中的具体位置（所占部分）应为：

$$\frac{0.5 - 0.231}{0.552 - 0.231} = 0.838$$

故中位数约为

$$3 \times 0.838 + 4.2 = 6.714$$

再来我们算第 67 百分位数，记为 p，易知 $0.552 + 0.129 = 0.681$，故 p 在第三组中，则可得

$$\begin{aligned} p &= \frac{0.67 - 0.552}{0.681 - 0.552} \times 3 + 7.2 \\ p &= 9.944 \end{aligned}$$

茎叶图

茎叶图是反应原始数据分布的图形。它由茎和叶两部分组成，其图形由数字构成。

e.g. 有两组数据，甲组：12,23,24,25,32,33,37,41,50；乙组：11,23,34,38,39,40,42,52,56。通过茎叶图可以表示为

甲				乙		
		2	1	1		
5	4	3	2	3		
3	2	7	3	4	8	9
		1	4	0	2	
		0	5	1	6	

图 21.3: 茎叶图

其中中间为茎，两边为叶，茎代表十位数，叶代表个位数。

第二十二章　成对数据统计分析

这一章在高考中并不算难，主要是计算量大，但这一章往往容易被轻视，我个人觉得从样本相关系数引出 n 维向量的角度来理解，以及一元线性回归方程中对于整体接近程度取最小值，从而得到最小二乘估计，教课书上用到的方法都不超过高中范围，非常巧妙。

此外“成对数据”其实就是二维随机变量，其中相关系数这一节就涉及到二维随机变量的数字特征“协方差”，在教科书中没有，我会补充一点。

§22.1　成对数据的统计相关性

变量间的关系常见的有两类：一类是确定关系，即函数关系；另一类是变量与变量间虽然确定存在关系，但又没有确切到可有其中一个去精确的决定另一个的程度，这种关系叫相关关系。

如果从整体看，一个变量增加时，另一个变量的相应值也呈现增加的趋势，则称这两个变量为**正相关**；如果一个变量增加时另一个变量减小则称这两个变量为**负相关**。

散点图：成对数据可以用表格表示，也可以用点的形式表示为 (x_i, y_i)，这一串点的图叫做散点图。

线性相关：如果两个变量的取值呈正/负相关，而且散点落在一条直线附近，则称这两个变量线性相关。

设

$$s_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}, s_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2}$$

$$x_i' = \frac{x_i-\overline{x}}{s_x}, y_i' = \frac{y_i-\overline{y}}{s_y}$$

则有

$$r = \frac{1}{n}\sum_{i=1}^{n}x_i'y_i' = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}}$$

其中 r 称为 x 和 y 的**样本相关系数**，这是高中的定义，如果引入协方差的概念，可以这样定义相关系数：

对于随机变量 X, Y，我们将

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

称为 X 与 Y 的相关系数，这两种定义是等价的。

当 $r > 0$ 时，称成对数据正相关，这是一个数据变小，另一个数据通常也变小；一个数据变大，另一个数据通常也变大；反之当 $r < 0$ 时一个变小另一个通常变大。

将 $x_i', y_i', (i = 1, 2, \ldots, n)$ 分别看成 $\overrightarrow{x'}, \overrightarrow{y'}$ 在 n 维上的坐标；则 $\overrightarrow{x'}, \overrightarrow{y'}$ 的内积公式可以推出，$r = \cos\theta$，故 $r \in [-1, 1]$。

当 $|r|$ 越接近 1 时，线性相关程度越强；$|r|$ 越接近 0 时，线性相关程度越弱。

当 $|r|=1$ 时 $\theta=0$，故 $\overrightarrow{y'}=\lambda\overrightarrow{x'}$。即

$$\frac{y_i-\overline{y}}{s_y}=\lambda\frac{x_i-\overline{x}}{s_x}, i=1,2,\ldots,n$$

表面 (x_i,y_i) 都落在直线 $y-\overline{y}=\frac{\lambda s_y}{s_x}(x-\overline{x})$ 上

§22.2　一元线性回归模型及其应用

设随机变量 Y 与普通变量 x 之间存在某种相关关系，为了近似的掌握 Y 与 x 的关系，我们考察 Y 的数学期望，若 Y 的数学期望 $E(Y)$ 存在，则其值随 x 的取值而定，它是关于 x 的函数，记为 $\mu(x)$，称为 Y 关于 x 的**回归函数**。

在实际问题中，回归函数 $\mu(x)$ 一般是未知的，回归分析的任务是根据试验数据取估计回归函数。

我们对于 x 取定一组不完全相同的值 $x_1,x_2,\ldots,x_n$，设 $Y_1,Y_2,\ldots,Y_n$ 分别是在 $x_1,x_2,\ldots,x_n$ 除对 Y**独立**观察的结果，称

$$(x_1,Y_1),(x_2,Y_2),\ldots,(x_n,Y_n)$$

是一个样本，对应的样本值为

$$(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$$

通过 (x_i,y_i) 确定的散点图，我们可以粗略的看出 $\mu(x)$ 的形式；

若 $\mu(x)$ 为线性函数：$\mu(x)=a+bx$，此时估计 $\mu(x)$ 的问题称为求一元线性回归问题，这节我们讨论这个问题。

假设对于 x（在某个区间内）的每一个值有

$$Y\sim N(a+bx,\sigma^2)$$

其中 a,b 以及 σ^2 都是不依赖于 x 的未知参数，记 $\varepsilon=Y-(a+bx)$，对 Y 进行这样的正态假设，相当于假设：

$$Y=a+bx+\varepsilon,\varepsilon\sim N(0,\sigma^2) \tag{2.1}$$

其中 a,b,σ^2 都不依赖于 x，上式称为**一元线性回归模型**，b 为回归系数或斜率参数、a 为截距参数，ε 是 Y 与 $bx+a$ 间的随机误差。

(2.1) 式说明因变量 Y 由两部分组成，一部分是 x 的线性函数 $a+bx$，另一部分 $\varepsilon\sim N(0,\sigma^2)$ 是随机误差，是人们不可控的。

(二)，a,b 的估计

取 x 的 n 个不完全相同的值 $x_1,x_2,\ldots,x_n$ 作独立试验，得到样本 $(x_1,Y_1),(x_2,Y_2),\ldots,(x_n,Y_n)$，由 (2.1) 式可知

$$Y_i=a+bx_i+\varepsilon_i,\varepsilon_i\sim N(0,\sigma^2)$$

各个 ε_i 相互独立；于是 $Y_i\sim N(a+bx_i,\sigma^2),i=1,2,\ldots,n$，由 $Y_1,Y_2,\ldots,Y_n$ 的独立性可知，$Y_1,Y_2,\ldots,Y_n$

的联合密度为

$$
\begin{aligned}
L &= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y_i - a - bx_i)^2\right] \\
&= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - a - bx_i)^2\right] \qquad (2.2)
\end{aligned}
$$

现在用最大似然估计法来估计未知参数 a, b 对于任意一组观测值 $y_1, y_2, \ldots, y_n$，(2.2) 式就是样本的似然函数。显然，要使 L 取最大值，只要 (2.2) 式中

$$
Q(a,b) = \sum_{i=1}^{n}(y_i - a - bx_i)^2 \qquad (2.3)
$$

取最小值。

求 $Q(a,b)$ 最小值的方法我提供两种，一种是高中教课书上的方法，一种是更一般、应用更广泛的方法，不过需要一定的数学基础；这里我想特别说一下，高中的一元线性回归模型并没提到 $Y \sim N(a+bx, \sigma^2)$，而是用下列式子

$$
\begin{cases}
Y = bx + a + \varepsilon \\
E(\varepsilon) = 0, D(\varepsilon) = \sigma^2
\end{cases}
$$

来表示一元线性回归模型，我认为是因为高中内容太少了，如果要使 $Y \sim N(a+bx, \sigma^2)$，那么还得知道边缘分布、相互独立的随机分布的联合密度以及最大似然法，这知识量还是蛮大的，而且还涉及到数学分析的知识，因此高中的一元线性回归没有 $Y \sim N(a+bx, \sigma^2)$；

若 Y 表示正态变量，则直接使用 (2.3) 式估计 a, b 使 Y 的观察值 y_i 与 $a + bx_i$ 偏差的平方和 $Q(a,b)$ 为最小，这种方法叫**最小二乘法**，若 Y 是正态变量，则最小二乘法与最大似然估计法给出的结果相同。

首先我用高中的方法来估计 a, b，这个方法很巧妙，理解过程，方可对 $\sum$ 运算更熟悉：

记

$$
\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i
$$

因为

$$
\begin{aligned}
Q(a,b) &= \sum_{i=1}^{n}(y_i - a - bx_i)^2 \\
&= \sum_{i=1}^{n}[y_i - bx_i - (\overline{y} - b\overline{x}) + (\overline{y} - b\overline{x}) - a]^2 \\
&= \sum_{i=1}^{n}[(y_i - \overline{y}) - b(x_i - \overline{x}) + (\overline{y} - b\overline{x}) - a]^2 \\
&= \sum_{i=1}^{n}[(y_i - \overline{y}) - b(x_i - \overline{x})]^2 + 2\sum_{i=1}^{n}\{[(y_i - \overline{y}) - b(x_i - \overline{x})] \times [(\overline{y} - b\overline{x}) - a]\} \\
&\quad + n[(\overline{y} - b\overline{x}) - a]^2
\end{aligned}
$$

注意到

$$
\begin{aligned}
&\sum_{i=1}^{n}\{[(y_i-\overline{y})-b(x_i-\overline{x})]\times[(\overline{y}-b\overline{x})-a]\}\\
&=(\overline{y}-b\overline{x}-a)\sum_{i=1}^{n}[(y_i-\overline{y})-b(x_i-\overline{x})]\\
&=(\overline{y}-b\overline{x}-a)[\sum_{i=1}^{n}(y_i-\overline{y})-b\sum_{i=1}^{n}(x_i-\overline{x})]\\
&=(\overline{y}-b\overline{x}-a)[(n\overline{y}-n\overline{y})-b(n\overline{x}-n\overline{x})]\\
&=0
\end{aligned}
$$

所以

$$
Q(a,b)=\sum_{i=1}^{n}[(y_i-\overline{y})-b(x_i-\overline{x})]^2+n(\overline{y}-b\overline{x}-a)^2
$$

上式右边均为非负数，且前 n 项与 a 无关，所以，要是 Q 取到最小值，后一项的值应为 0，即 $a=\overline{y}-b\overline{x}$，此时

$$
\begin{aligned}
Q(a,b)&=\sum_{i=1}^{n}[(y_i-\overline{y})-b(x_i-\overline{x})]^2\\
&=b^2\sum_{i=1}^{n}(x_i-\overline{x})^2-2b\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})+\sum_{i=1}^{n}(y_i-\overline{y})^2
\end{aligned}
$$

上式是关于 b 的二次函数，因此要使 Q 取最小值，当且仅当 b 为

$$
\hat{b}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}
$$

综上，Q 取最小值时有：

$$
\begin{cases}
\hat{b}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\\
\hat{a}=\overline{y}-\hat{b}\overline{x}
\end{cases}
$$

上面的推导我觉得很“巧”，配凑、消元后变成二次函数求最值问题，接下来我用更容易想到也更一般的方法：由于 $Q(a,b)$ 是二元函数，对于这种二元函数无限制条件的求最值问题我们一般如下求解：

取 Q 分别关于 a,b 的偏导数，并令它们为为零：

$$
\begin{cases}
\frac{\partial Q}{\partial a}=-2\sum_{i=1}^{n}(y_i-a-bx_i)=0\\
\frac{\partial Q}{\partial b}=-2\sum_{i=1}^{n}(y_i-a-bx_i)x_i=0
\end{cases}
$$

得到方程组

$$
\begin{cases}
na+(\sum_{i=1}^{n}x_i)b=\sum_{i=1}^{n}y_i\\
(\sum_{i=1}^{n}x_i)a+(\sum_{i=1}^{n}x_i^2)b=\sum_{i=1}^{n}x_iy_i
\end{cases}
\tag{2.4}
$$

(2.4) 式被称为**正规方程组**，得到驻点

$$\begin{cases} \hat{a} = \frac{n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \\ \\ \hat{b} = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \end{cases}$$

求二阶偏导数，并记

$$\frac{\partial^2 Q}{\partial a^2} = 2n = A$$
$$\frac{\partial^2 Q}{\partial a \partial b} = 2\sum_{i=1}^n x_i = B$$
$$\frac{\partial^2 Q}{\partial b^2} = 2\sum_{i=1}^n x_i^2 = C$$

令

$$H = \begin{vmatrix} A & B \\ B & C \end{vmatrix} = 4\left(n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2\right)$$

由柯西不等式可知

$$\left(\sum_{i=1}^n x_i \cdot 1\right)^2 \geqslant \sum_{i=1}^n x_i^2 \sum_{i=1}^n 1^2$$
$$\left(\sum_{i=1}^n x_i\right)^2 \geqslant n\sum_{i=1}^n x_i^2$$

又因 x_i 不完全相同，因此 $\frac{x_i}{1}$ 不为定值，即不能取等

$$\left(\sum_{i=1}^n x_i\right)^2 > n\sum_{i=1}^n x_i^2$$

综上，$H > 0, A > 0$，故 $Q(a,b)$ 有最小值点 $(\hat{a}, \hat{b})$，为了方便计算，一般记

$$\begin{cases} \hat{b} = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} \\ \\ \hat{a} = \overline{y} - \hat{b}\overline{x} \end{cases} \tag{2.5}$$

其中 $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i, \overline{y} = \frac{1}{n}\sum_{i=1}^n y_i$

得到 a, b 的估计值 $\hat{a}, \hat{b}$，对于给定的 x，我们取 $\hat{a} + \hat{b}x$ 作为回归函数 $\mu(x) = a + bx$ 的估计，即 $\hat{\mu(x)} = \hat{a} + \hat{b}x$，称为 Y 关于 x 的经验回归函数

$$\hat{y} = \hat{a} + \hat{b}x \tag{2.6}$$

称为 Y 关于 x 的经验回归方程，简称回归方程，其图形称为回归直线。

将 (2.5) 式中 $\hat{a}$ 的表达式代入 (2.6) 式中则回归方程可写为：

$$\hat{y} = \overline{y} + \hat{b}(x - \overline{x})$$

上式表明，对于样本值 $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$，回归直线通过散点图的几何中心 $(\overline{x}, \overline{y})$。

为了方便计算 $\hat{a}, \hat{b}$ 还可以写成：

$$\begin{cases} \hat{b} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} \\ \hat{a} = \overline{y} - \hat{b}\overline{x} \end{cases}$$

这里读者可自行推导 $\hat{b}$。

记 $\hat{y}_i = \hat{a} + \hat{b}x_i$，称 $y_i - \hat{y}_i$ 为 x_i 处的残差，平方和

$$Q_e = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - \hat{a} - \hat{b}x_i\right)^2$$

称为**残差平方和**，为了方便计算，Q_e 也写成

$$Q_e = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2 - \hat{b}\left(\sum_{i=1}^{n} x_i y_i - n\overline{x}\overline{y}\right)$$

可以用 R^2 (决定系数) 为比较两个模型的拟合效果。

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

§22.2.1 列联表与独立性检验表

分类变量：讨论如“吸烟是否会增加患肺癌风险”的问题，常用一种特殊的随机变量，以区别不同的现象或性质。

2×2 列联表：将数据分类统计，做成表格的数据统计表，例如：

	Y=0	Y=1	合
X=0	a	b	a+b
X=1	c	d	c+d
合	a+c	b+d	a+b+c+d

其中有

$$\chi^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

当 χ^2 越大，X，Y 相关性就越强。(旧教材中 χ^2 用 K^2 表示)

$\chi^2 = k$，k 为观测值，有一套临界值表：

犯错概率	$P(\chi^2 \geqslant k_0)$	0.5	0.40	...
观测值	k_0	0.455	0.708	...

通过 χ^2 的取值推断 X 和 Y 是否独立的方法称为 χ^2 独立性检验 (卡方独立性检验)，简称独立性检验。
判断 $X = 1$ 和 $Y = 1$ 间是否有关联，需判断假定关系是否成立

$$H_0 : P(Y = 1|X = 0) = P(Y = 1|X = 1)$$

称 H_0 为零假设或原假设。

第六部分

几何

第二十三章　平面几何基础

这一章主要介绍平面几何的基础知识，主要是纯几何的内容，由于高中考察不多，所以不会有大量篇幅。

§23.1　初中拾遗

这一节主要回忆一下初中学的基本的知识，此外还有一些不复杂的、初中知识即可证明的我都会放在这一节。

§23.1.1　射影定理

在解三角形中常常用到 $A+B+C=\pi$ 这个性质，这告诉我们三角形中的一个角可以用另外两个角来表示，例如：

$$A=\pi-B-C$$

因此有：

$$\begin{aligned}\sin A&=\sin(\pi-B-C)\\ \sin A&=\sin(B+C)\\ \sin A&=\sin B\cos C+\sin C\cos B\end{aligned}$$

通过正弦定理，将上式中齐次项可角化边为：

$$a=b\cos C+c\cos B$$

用同样的方法可以证明：

定理

在 $\triangle ABC$ 中，角 A,B,C 所对的边分别为 a,b,c，有：

$$\begin{aligned}a&=b\cos C+c\cos B\\ b&=a\cos C+c\cos A\\ c&=b\cos A+a\cos B\end{aligned}$$

这个定理称为**射影定理**。

在直角三角形中讨论的射影定理与以上的射影定理不相干。

定理

在直角三角形 $\triangle ABC, \angle ACB = \frac{\pi}{2}$ 中，过 C 作 $CD \perp AB$ 垂足为 D，有

$$|AC|^2 = |AD| \cdot |AB|$$
$$|BC|^2 = |BD| \cdot |BA|$$
$$|CD|^2 = |AD| \cdot |BD|$$

我认为广义的来说，任何利用这个模型得出的相似比关系都可以算在射影定理的范畴中。♠

证明：

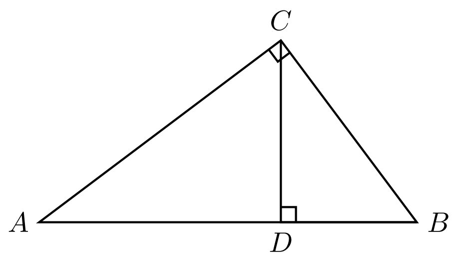


图 23.1: 直角三角形射影定理

证明：由图易知

$$\triangle ABC \sim \triangle CBD \sim \triangle ACD$$

即证

$$\frac{|AC|}{|AB|} = \frac{|AD|}{|AC|} \iff |AC|^2 = |AD| \cdot |AB|$$
$$\frac{|BC|}{|BD|} = \frac{|AB|}{|BC|} \iff |BC|^2 = |BD| \cdot |BA|$$
$$\frac{|CD|}{|AD|} = \frac{|BD|}{|CD|} \iff |CD|^2 = |AD| \cdot |BD|$$

□

由这个射影定理立即可以推出

推论

在直角三角形 $\triangle ABC, \angle ACB = \frac{\pi}{2}$ 中，过 C 作 $CD \perp AB$ 垂足为 D，有

$$\frac{|AD|}{|BD|} = \frac{|AC|^2}{|BC|^2}$$

◇

§23.1.2 一些分线定理

在三角形中，过一顶点到其对边的一点，顶点到这一点的连线有一些比较常用的定理，当这连线为角平分线、中线等时又有比较特殊的意义，这一节详细介绍。

§23.1.3 角平分线定理

定理

在 $\triangle ABC$ 中，$\angle BAC$ 的角平分线与 BC 交于 D，则有：

$$\frac{|AB|}{|AC|} = \frac{|BD|}{|DC|}$$

该定理称为**角平分线定理**，以及逆定理：若 $\triangle ABC$ 中 D 是 BC 上的点且满足：

$$\frac{|AB|}{|AC|} = \frac{|BD|}{|DC|}$$

那么 AD 是 $\angle BAC$ 的角平分线。

在初中阶段，会用到等体积法或者构造的方法来证明这个定理，不过我们以及学过正弦定理了，能更轻松的证明这个定理。

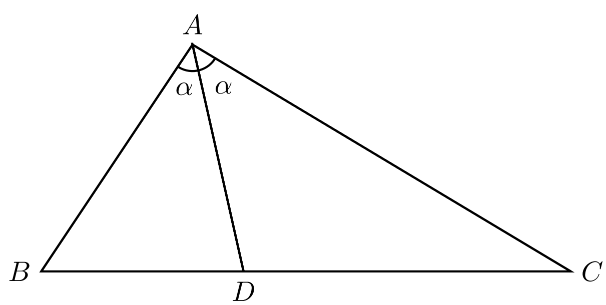


图 23.2: 角平分线定理

证明：设 $\angle BAD = \alpha, \angle BDA = \beta$，通过正弦定理有：

$$\frac{|BD|}{\sin\alpha} = \frac{|AB|}{\sin\beta}$$

另一方面在 $\triangle ADC$ 中有

$$\frac{|DC|}{\sin\alpha} = \frac{|AC|}{\sin(\pi-\beta)} = \frac{|AC|}{\sin\beta}$$

两式对照，即证

$$\frac{|AB|}{|AC|} = \frac{|BD|}{|DC|}$$

同时，其逆定理也成立，证明是显然的。

§23.1.4　张角定理

定理

在 $\triangle ABC$ 中，点 D 为 BC 上一点，则有

$$\frac{\sin\angle BAC}{|AD|}=\frac{\sin\angle BAD}{|AC|}+\frac{\sin\angle CAD}{|AB|}$$

上式被称为**张角定理**。

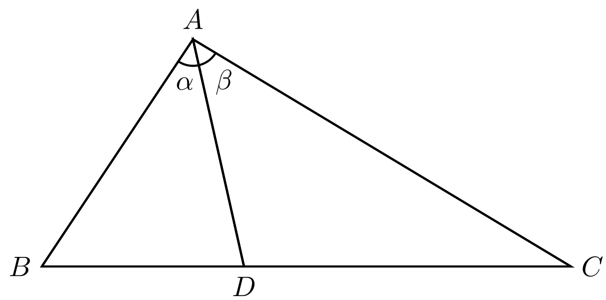


图 23.3: 张角定理

证明：记 $\triangle ABD,\triangle ACD,\triangle ABC$ 的面积为 S_{ABD},S_{ACD},S_{ABC}，$\angle BAD=\alpha,\angle CAD=\beta$ 则有

$$S_{ABD}=\frac{1}{2}\sin\alpha|AD||AB|,\quad S_{ACD}=\frac{1}{2}\sin\beta|AD||AC|,\quad S_{ABC}=\frac{1}{2}\sin(\alpha+\beta)|AB||AC|$$

且 $2S_{ABD}+2S_{ACD}=2S_{ABC}$，则

$$\begin{gathered}\sin\alpha|AD||AB|+\sin\beta|AD||AC|=\sin(\alpha+\beta)|AB||AC|\\ \frac{\sin\alpha}{|AC|}+\frac{\sin\beta}{|AB|}=\frac{\sin(\alpha+\beta)}{|AD|}\qquad ①\end{gathered}$$

□

其逆定理为，如果 A,B,C,D 满足上式，则 B,C,D 三点共线；张角定理的逆定理是证明三点共线的一方法。

当 $\alpha=\beta$ 时，即 AD 为 $\angle BAC$ 的角平分线，① 式为

$$\begin{aligned}\frac{\sin\alpha}{|AC|}+\frac{\sin\alpha}{|AB|}&=\frac{\sin(2\alpha)}{|AD|}\\ \frac{1}{|AC|}+\frac{1}{|AB|}&=\frac{2\cos\alpha}{|AD|}\end{aligned}$$

由均值不等式可知：

$$\begin{aligned}\frac{|AD|}{\cos\alpha}=\frac{2}{\frac{1}{|AC|}+\frac{1}{|AB|}}&\leqslant\sqrt{|AB||AC|}\\ \frac{|AD|^2}{\cos^2\alpha}&\leqslant|AB||AC|\end{aligned}$$

当且仅当 $|AB|=|AC|$ 时等号成立，从几何的角度 $|AB|=|AC|$ 时 $\triangle ABC$ 为等腰三角形，三线合一致使

$$|AD|=\cos\alpha|AB|=\cos\alpha|AC|$$

§23.1.5 分角线定理

定理

在 $\triangle ABC$ 中，点 D 为 BC 上一点，则有

$$\frac{|BD|}{|CD|}=\frac{\sin\angle BAD|AB|}{\sin\angle CAD|AC|}$$

上式被称为**分角线定理**，其证明过程也是用到了面积法。

证明：沿用张角定理证明的图以及设，另设过 A 点的高为 h，则有

$$\begin{aligned}\frac{S_{ABD}}{S_{ADC}}&=\frac{S_{ABD}}{S_{ADC}}\\ \frac{|BD|h}{|DC|h}&=\frac{\sin\alpha|AB||AD|}{\sin\beta|AC||AD|}\\ \frac{|BD|}{|CD|}&=\frac{\sin\alpha|AB|}{\sin\beta|AC|}\end{aligned}$$

□

当 AD 为中线时，$|BD|=|DC|$，也即

$$\frac{\sin\alpha}{\sin\beta}=\frac{|AC|}{|AB|}$$

或

$$|AB|\sin\alpha=|AC|\sin\beta$$

上式被称为**中线定理**。

§23.2 全等与相似

第二十四章　立体几何

新版的这一章节内容是老版必修二的第一章空间几何体与第二章点线面关系组成，内容多，东西杂，二级结论也很多，但是只要记住那五个公理和九个定理的用法，做起题来就不会复杂。

§24.1　基础立体几何图形

多面体：由若干个平面多边形围成的几何体叫多面体，各个多边形叫做多面体的面，相邻两个面的公共边叫做多面体的棱，棱与棱的公共点叫做多面体的顶点。

旋转体：由一个平面图形绕它所在平面内的一条定直线旋转所形成的封闭几何体，这条定直线叫旋转体的轴。

多面体

一、**棱柱**：一般地，有两个面互相平行，其余各面都是四边形，并且每相邻两个四边形的公共边都互相平行，由这些面所围成的多面体叫做棱柱。

1. 两个互相平行的面叫做底面
2. 除底边外的其余各面叫做棱柱的侧面
3. 相邻侧面的公共边叫做棱柱的侧棱
4. 侧棱与地面的公共顶点叫做棱柱的顶点

底边是 n 边形的棱柱是 n 棱柱 $(n \geqslant 3, n \in \mathbb{Z})$
侧棱不垂直于底面的棱柱叫做**斜棱柱**。
侧棱垂直于底面的棱柱叫**直棱柱**。
底面是正多边形的直棱柱叫做**正直棱柱**。

二，**棱锥**：一般地，有一个面是多边形，其余各面都是有一个公共顶点的三角形，由这些面所围成的多面体叫做棱锥。

1. 多边形面叫做底面；
2. 有公共顶点的各三角形面叫做侧面；
3. 各侧面的公共顶点也叫棱锥的顶点；
4. 相邻侧面的公共边叫做棱锥的侧棱。

底边是 n 边形的棱锥是 n 棱锥 $(n \geqslant 3, n \in \mathbb{Z})$
用顶点和底面各顶点字母表示棱锥，如 $S - ABCD$

正棱锥：如果一个棱锥的底面是正多面体，并且顶点在底面的射影是底面的中心，这样的棱锥叫正棱锥。各棱长均相等的四面体叫<u>**正四面体**</u>。

三，**棱台**用一个平行于棱锥底面的平面去截棱锥，底面与截面之间的部分，这样的多面体叫做棱台。原棱锥的底面和截面分别叫做棱台的下底面和上底面，棱台也有侧面、侧棱、顶点。

由 n 棱锥截得的棱台是 n 棱台 $(n \geqslant 3, n \in \mathbb{Z})$

正棱台：由正棱锥截得的棱台叫做正棱台，正棱台各侧棱相等，各侧面都是全等的等腰梯形。

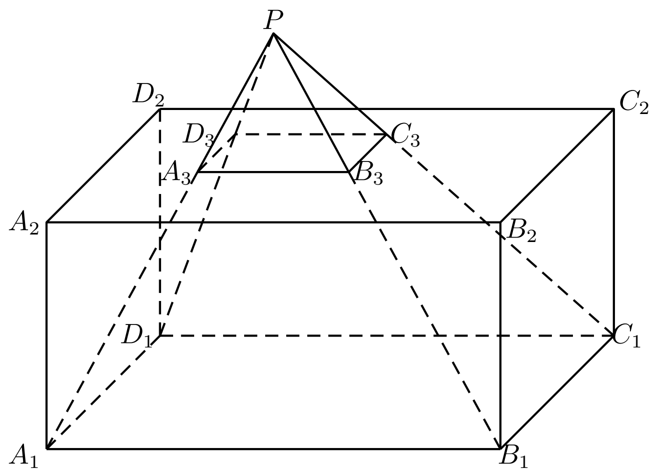


图 24.1: 三合一

如图，有直棱柱 $A_1B_1C_1D_1 - A_2B_2C_2D_2$，棱锥 $P - A_1B_1C_1D_1$ 以及棱台 $A_1B_1C_1D_1 - A_3B_3C_3D_3$。

旋转体

四，**圆柱**：以矩形的一边所在直线为旋转轴，其余三边旋转形成的面所围成的旋转体叫做圆柱。

1. 旋转轴叫做圆柱的轴；
2. 垂直于轴的边旋转而成的圆面叫做圆柱的底面；
3. 平行于轴的边旋转而成的曲面叫做圆柱的侧面；
4. 无论旋转到什么位置，不垂直于轴的边都叫做圆柱侧面的母线。

圆柱和棱柱统称<u>**柱体**</u>，圆柱侧面展开是矩形。

五，**圆锥**：以直角三角形的一条直角边所在直线为旋转轴，其余两边旋转形成的面所围成的旋转体叫做圆锥。

1. 旋转轴叫做圆锥的轴；
2. 垂直于轴的边旋转而成的圆面叫做圆锥的底面；
3. 不垂直于轴的边旋转而成的曲面叫做圆锥侧面；

4. 无论旋转到什么位置，不垂直于轴的边都叫做圆锥的母线。

圆锥和棱锥统称为**锥体**。

六，**圆台**：用平行于圆锥底面的平面去截圆锥，底面于截面之间的部分叫圆台。与圆柱圆锥一样，圆台也有轴、底面、侧面、母线。

七，**球**：以半圆的直径所在直线为旋转轴，半圆面旋转一周形成的旋转体叫做球体，简称球。
半圆的圆心叫做球的球心，半圆的半径也是球的半径。
用一个平面去截球，球心到截面的距离 d 与球半径 R 以及截面圆的半径 r 满足：$r=\sqrt{R^2-d^2}$。

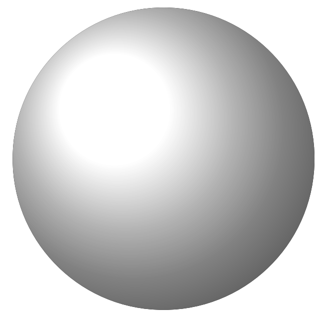


图 24.2: 球

§24.2　直观图

画空间几何时，常常用斜二测画法，斜二测画法是将三维物体平行投影到二维平面的一种画法。

1，将 x 轴和 y 轴交于一点 O，使 $\angle xOy=45°$。

2，原平面中平行 x 轴的线段，长度不变，平行于 y 轴的线段为原平面中的一半。

§24.3　体积公式

体积公式：

棱柱体积：$V=Sh$

棱锥体积：$V=\frac{1}{3}Sh$

棱台体积：$V=\frac{1}{3}h\left(S+\sqrt{SS_1}+S_1\right)$

圆柱体积：$V=\pi r^2h$

圆锥体积：$V=\frac{1}{3}\pi r^2h$

圆台体积：$V=\frac{1}{3}\pi h\left(r^2+rr_1+r_1^2\right)$

球体积公式：$V=\frac{4}{3}\pi R^3$

下面仅证明棱台与圆台的体积公式：

先证明棱台的体积

如图为一棱锥，底面积为 S，平行于底面截得面积为 S'，记得到的棱台体积为 V 高为 h，棱锥顶点为 P，大棱锥的高为 $|OP|=h+h_1$，小棱锥的高为 $|O'P|=h_1$，因为两棱锥相似，因此有

$$\frac{h+h_1}{h_1}=\frac{\sqrt{S}}{\sqrt{S'}}$$

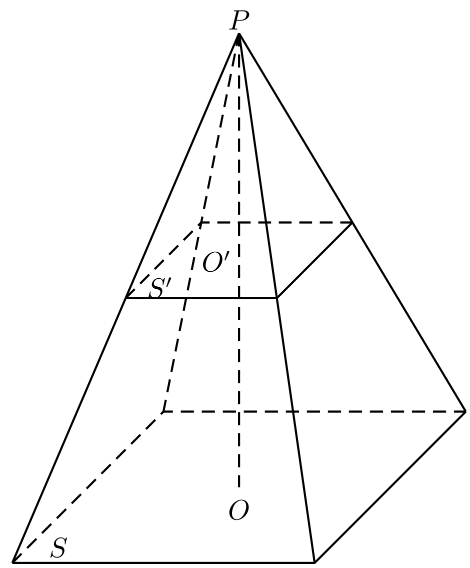


图 24.3: 棱台体积证明

记大棱锥体积为 V_1，小棱锥体积为 V_2，则有：

$$\begin{aligned}V &= V_1 - V_2\\ &= \frac{1}{3}\left(S(h+h_1) - S'h_1\right)\\ &= \frac{h}{3}\frac{\sqrt{S^3}-\sqrt{S'^3}}{\sqrt{S}-\sqrt{S'}}\\ &= \frac{h(S+\sqrt{SS'}+S')}{3}\end{aligned}$$

再来证明圆台的体积

如图为一圆锥，顶点为 P，底面圆心为 O 半径为 R，平行于底面截半径为 r，圆心为 O' 的圆，记截得的圆台体积为 V 高为 h，大圆锥高为 $|OP| = h + h_1$，体积为 V_1，小圆锥高为 $|O'P|$，体积为 V_2。

由于两圆锥相似，则有

$$\frac{R}{r} = \frac{h+h_1}{h_1}$$

另一方面

$$\begin{aligned}V &= V_1 - V_2\\ &= \frac{\pi}{3}\left(R^2(h+h_1) - r^2h_1\right)\\ &= \frac{\pi}{3}\frac{hR^3-hr^3}{R-r}\\ &= \frac{h\pi}{3}\left(R^2+Rr+r^2\right)\end{aligned}$$

□

表面积公式：

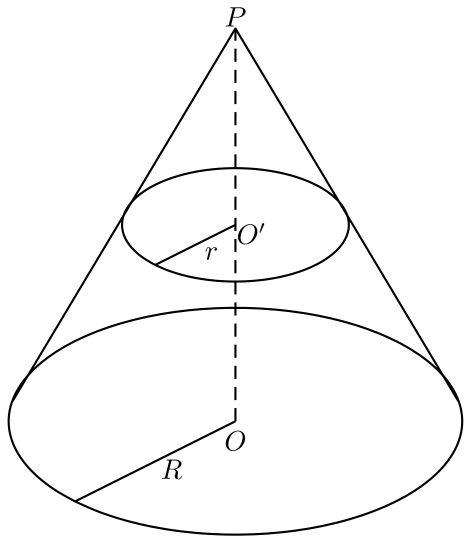


图 24.4: 圆台体积证明

圆柱表面积：$S = 2\pi r\,(r + l)$

圆锥表面积：$S = \pi r\,(r + l)$，圆锥侧面积：$S = \pi rl$

圆台表面积：$S = \pi\,(r_1^2 + r^2 + rl + r_1 l)$

球表面积：$S = 4\pi R^2$

§24.4 点线面关系与公理

线与线之间：分为异面直线和共面直线，共面直线又分为相交直线和平行直线。

线与面之间：分为直线在平面内和直线在平面外，直线在平面外又分为平行于相交关系。

面与面之间：平行或相交。

接下来是四个公理 (新教材叫基本事实)

公理 1 ： 过不在一条直线的三个点，有且只有一个平面。(不共线的三点确定一个平面)

公理 2 ： 如果一条直线上两个点在一个平面内，那么这条直线在这个平面内。

公理 3 ： 如果两个不重合的平面，有一公共点，那么它们有且只有一条过该点的公共直线。

公理 4 ： 平行于同一直线的两条直线平行。

这些公理乍看之下是废话，但却很有用；公理 1234 可以以点线面平行来记忆。另外还有一个等角定理

定理 1

定理 1： 如果空间中两个角的邻边分别平行，那么这两个角相等或互补。 ♠

§24.5　平行的定理

四条关于平行的定理：

平行定理

定理 2：如果平面外一直线与平面内任一直线平行，则该直线与该平面平行

定理 3：一条直线与一个平面平行。如果过该直线的平面与该平面相交，则交线与该直线平行。

定理 4：如果一个平面内两条相交的直线与平面平行，那么这两个平面平行。

定理 5：两个平面平行，如果另一个平面与这两个平面相交，那么两条交线平行。

可以看出来其实就是平面与直线的平行互相转换，我画了一个转换图便于理解。

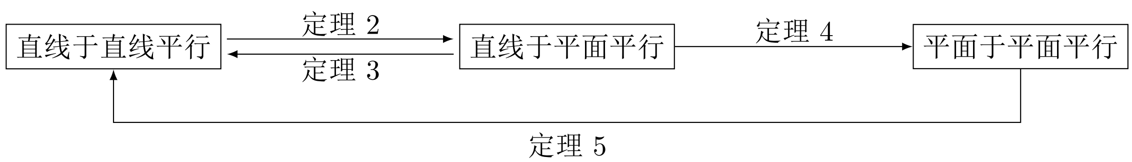


图 24.5: 直线与平面平行的转换

§24.6　垂直的定理

垂直定理

定理 6：如果一条直线与一个平面内的两条相交的直线垂直，则这条直线与该平面垂直。

定理 7：垂直于同一平面的两条直线平行。

定理 8：如果一个平面过另一个平面的垂线，则这两个平面垂直。

定理 9：两个平面垂直，如果一个平面内有一条直线垂直于这两个平面的交线，那么这条直线与另一个平面垂直。

其中直观的关系如下

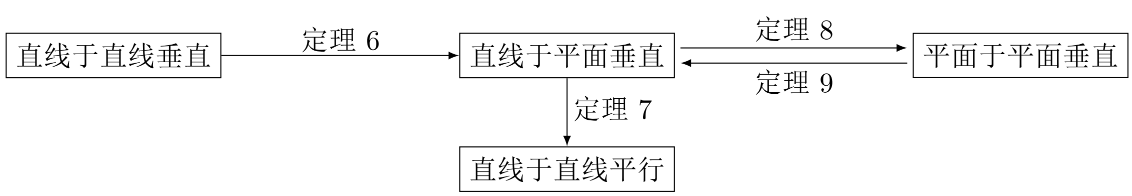


图 24.6: 直线与平面垂直的转换

§24.7　二面角

从一条直线出发的两个半平面，所组成的图形叫做**二面角**，这条直线叫二面角的**棱**，这两个半平面叫做**二面角的面**。棱为 AB。面为 α,β 的二面角记作“二面角 $\alpha-AB-\beta$”，二面角的取值范围是 $[0,\pi]$。

二面角的平面角：二面角 $\alpha - l - \beta$ 中 l 上的一点 P 为垂足，分别在 α, β 平面上作垂直于 l 的 PA, PB 射线, 二面角的平面角取值范围 $[0, \pi)$

一般地，两个面相交，如果它们所成的角是直二面角，就说两个平面互相垂直 $\alpha \perp \beta$。

§24.8 补充

§24.8.1 祖暅原理

定理

祖暅原理：“缘幂势既同，则积不容异。”其中“幂”代表截面积，“势”代表高；

用现代的汉语来解释就是：“介于两个平行平面间的两个几何体，被任一平行于这两平面的平面所截，若这两截面的面积总相同，则这两个几何体的体积相同。”

祖暅原理并不难理解，如果把一叠 A4 纸落在一起放在桌上组成一个几何体，若使它倾斜一个角度，则几何体形状发生了改变，变成一个新几何体，但由于它们的高和截面积相同所以体积也相同。

依据这个原理可以证明球的体积公式：

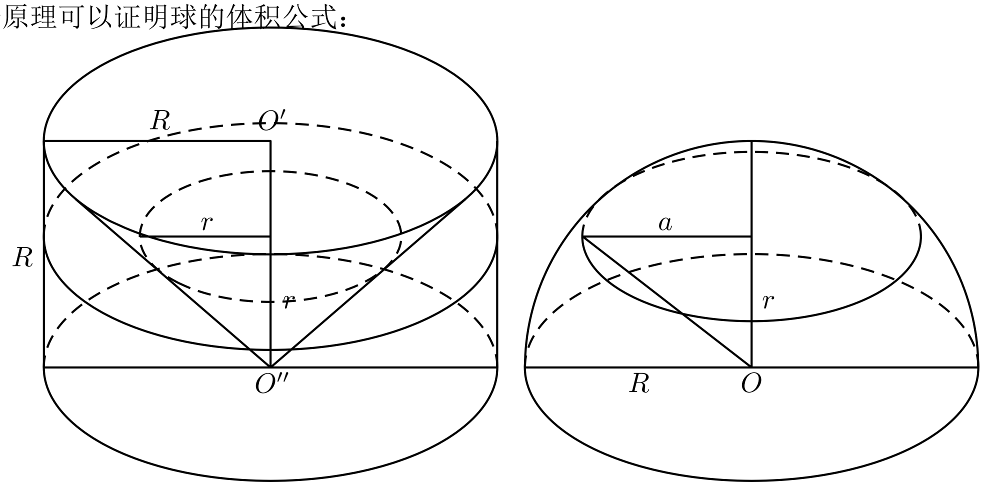


图 24.7: 祖暅原理

如图,构造一个半径为 R 的半球,球心为 O,对于距离球心 O 为 r 的截面圆,设其半径为 a,易知 $a^2 = R^2 - r^2$,面积为 πa^2;

这时，新构造一个底面半径为 R，高为 R 的圆柱，其上顶面圆的圆心为 O'，以上顶面为底面作一个高为 R 的圆锥，其顶点位于圆柱底面圆的圆心 O''。

现在我们构造了这样一个局面，对于距离 O'' 为 r 且平行于底面的圆锥截面的面积，不难知道为 πr^2，又可以知道此时这个截面截圆柱的面积为 πR^2，不难看出 $\pi R^2 - \pi r^2$ 的面积就是圆环的面积；

综上，这圆柱（内嵌圆锥）和半球，介于两个平行平面之间，且被任意平行于这两平面的平面所截的两截面面积都相同，即

$$\pi a^2 = \pi(R^2 + r^2) = \pi R^2 - \pi r^2$$

根据圆柱和圆锥的体积公式即可求这个半球的体积

$$V_{\text{半球}} = \pi R^3 - \frac{\pi R^3}{3} = \frac{2}{3}\pi R^3$$

最后可得球的体积公式：

$$V = 2V_{\text{半球}} = \frac{4}{3}\pi R^3$$

§24.8.2　三垂线定理

定理

三垂线定理：平面内的一条直线若和这个平面的一条斜线的射影垂直，则它也和这条斜线垂直。

三垂线定理的逆定理：平面内一条直线若和这个平面的一条斜线垂直，则他也和这条斜线的射影垂直。

证明：

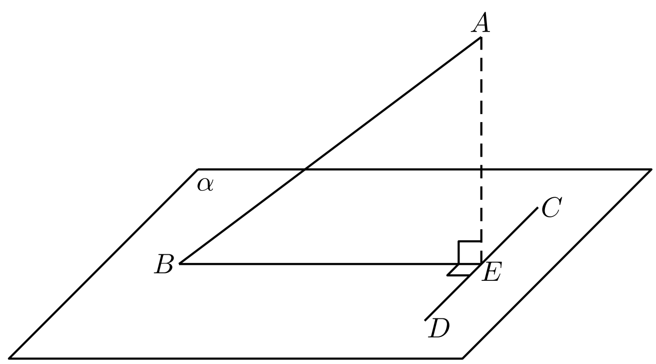


图 24.8: 三垂线定理

如图，直线 AB 在平面上的射影为 BE，显然 $AE \perp \alpha$，设 $DC \perp BE$，平面 ABE 记作 β

$$\begin{aligned}
&\because DC \perp BE, AE \perp \alpha, DC \subset \alpha \\
&\therefore AE \perp \alpha, AE \cap BE = E \\
&\therefore DC \perp \beta \\
&\therefore DC \perp AB
\end{aligned}$$

即证。对于逆定理，证明是类似的，如下

$$\begin{aligned}
&\because DC \perp AB, AE \perp \alpha, DC \subset \alpha \\
&\therefore AE \perp \alpha, AE \cap AB = A \\
&\therefore DC \perp \beta \\
&\therefore DC \perp BE
\end{aligned}$$

□

在高考中，三垂线定理不能直接用，需要证明才可以使用。

§24.8.3 三余弦定理

定理

直线 AB 与平面 π 交于点 B，作 AB 到平面 π 的射影 BC（BA 沿方向 CA 下的外射影），过点 B 在平面 α 上作一条直线 l，过点 C 作 $CD \perp l$，垂足为 D，记 $\angle ABD = \theta, \angle ABC = \alpha, \angle CBD = \beta$，则有：

$$\cos\theta = \cos\alpha \cdot \cos\beta$$

上式被称为**三余弦定理**

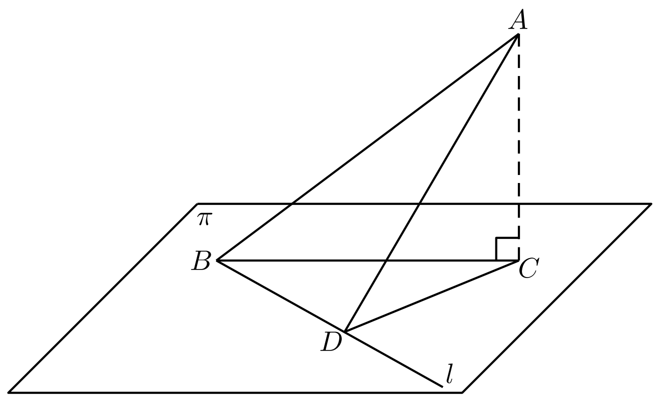


图 24.9: 三余弦定理

证明：如图，已知 $DC \perp BD$，而 DC 可以看作 AB 在平面 π 上的射影，根据三垂线定理的逆定理，可知：$BD \perp AD$，因此 $\triangle BDC, \triangle ABD, \triangle ABC$ 均为直角三角形，即有

$$\begin{cases} \cos\alpha & = \frac{|BC|}{|AB|} \\ \cos\beta & = \frac{|BD|}{|BC|} \\ \cos\theta & = \frac{|BD|}{|AB|} \end{cases}$$

即证：

$$\cos\theta = \frac{|BD|}{|AB|} = \frac{|BD| \cdot |BC|}{|AB| \cdot |BC|} = \cos\alpha \cdot \cos\beta$$

□

推论

斜线与平面所成的线面角是斜线与平面内直线所成的角的最小角，这被称为**最小角定理**。

证明：

通过三余弦定理可以快速证得，设斜线 AB 与平面 π 交于点 B，C 为 A 在平面 π 上的投影，平面内任意一条过 B 点的直线 l，作 $CD \perp l$ 垂足为 D，令 $\angle ABD = \theta, \angle ABC = \alpha, \angle CBD = \beta$，可知斜线 AB 与 π 的线面角为 α，易知：

$$\cos\alpha = \frac{\cos\theta}{\cos\beta} > \cos\theta, \beta \in (0, \frac{\pi}{2})$$

其中 θ 就是斜线与平面内直线所成的角。

□

§24.8.4　射影面积法

定理

平面 $A'DE$ 与平面 ADE 相交于直线 DE，其中 A 是 A' 在平面 ADE 上的投影，设二面角 $A-DE-A'$ 的平面角为 θ，则二面角的余弦值为：

$$\cos\theta = \frac{S_{ADE}}{S_{A'DE}}$$

♠

证明：

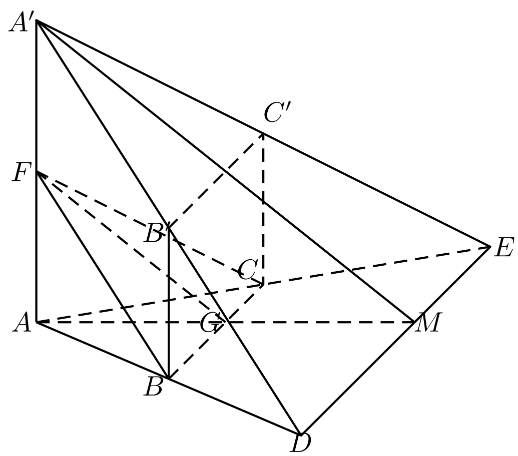


图 24.10: 射影面积法

如图，过 A' 作 $A'M \perp DE$，易知 $\triangle A'DE$ 在平面 ADE 上的射影为 $\triangle ADE$，因此 AM 是 $A'M$ 在平面 ADE 上的射影，由三垂线定理的逆定理可知

$$DE \perp AM$$

即有

$$\begin{aligned}\cos\theta &= \frac{|AM|}{|A'M|}\\ &= \frac{2|AM||DE|}{2|A'M||DE|}\\ &= \frac{S_{ADE}}{S_{A'DE}}\end{aligned}$$

□

推论

$\triangle A'B'C'$ 在平面 ADE 上的射影为 $\triangle ABC$，设平面 $A'B'C'$ 与平面 ABC 二面角的平面角为 θ，则

有

$$\cos\theta = \frac{S_{ABC}}{S_{A'B'C'}}$$

这个推论是显然的，如果将 $\triangle A'B'C'$ 平移至与平面 ABC 相交，得到 $\triangle FBC$，则由定理可知推论成立。

射影面积法的重要意义在于，我们在求一个二面角的余弦值时，有时不知道它们的相交直线，这时用射影面积法仍可以快速求得。

但**需要注意的是**，由于射影的性质，射影面积法只能求平面角为锐角的情况，对于平面角是钝角的情况，只需用射影面积法求它的补角即可。

§24.8.5 外接球

外接球的内容我会先讲一些通用的方法，之后再去说一些特殊的模型以及方法。

对于一个四面体（三棱锥），它的外接球球心 O 到四个顶点的距离相同，例如对于四面体 $ABCD$ 有：

$$|OA| = |OB| = |OC| = |OD| = R$$

其中 R 是外接球球半径，我们先只关注前三个等式：

$$|OA| = |OB| = |OC|$$

由平面几何的知识我们知道，对于平面上一个 $\triangle ABC$ 其外心 O' 到三角形顶点距离相同，这一点告诉我们球心 O 在平面 ABC 的射影为 O'，也就是说球心 O 在过三角形外心 O' 且垂直于平面 ABC 的直线 l 上，如果我们作 D 到平面 ABC 的射影为 D' 如图可列：

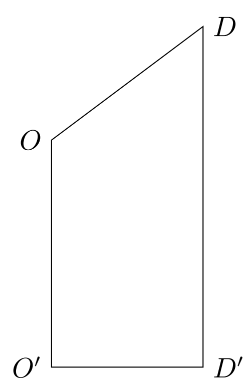


图 24.11: 外接球

$$|OD| = R = |O'D'|^2 + (|DD'| - |OO'|)^2$$

由于 $|O'D'|, |DD'|$ 是确定的，所以只需确定 $|OO'|$ 即可，注意到

$$|OO'|^2 = |OA|^2 - |O'A|^2 = R^2 - |O'A|^2$$

即可求 R 的值。

另外用解析几何建系来算 O 点也是可以的，已知四面体四点，通过

$$|OA| = |OB| = |OC| = |OD| = R$$

即可解出 O 的坐标以及 R。

§24.8.6 内切球

定理

四面体 $ABCD$ 的内切球半径：

$$R = \frac{3V}{S}$$

其中 R 是内切球半径，V 是四面体体积，S 是四面体表面积。

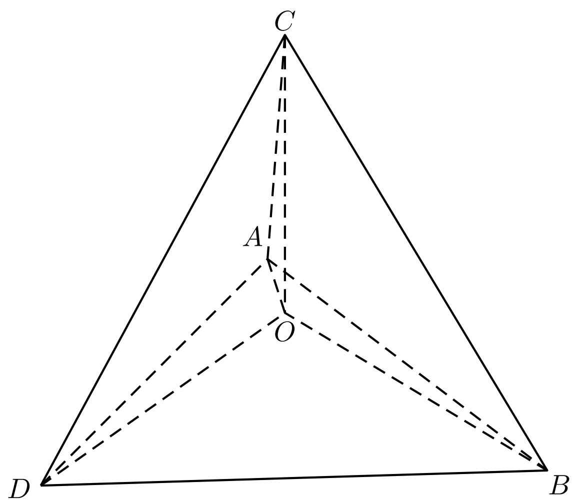


图 24.12: 内切球

证明过程与三角形内切圆半径的证明过程相同，设内切球球心为 O，则 O 到平面 ABC、平面 BCD、平面 ACD、平面 ABD 的距离都为 R，

则有：

$$V_{OABC} = \frac{R}{3}S_{PABC}$$
$$V_{OABD} = \frac{R}{3}S_{PABD}$$
$$V_{OBCD} = \frac{R}{3}S_{PBCD}$$
$$V_{OADC} = \frac{R}{3}S_{PADC}$$

即证：

$$R = \frac{3V}{S}$$

□

仔细观察会发现，这和三角形内切圆半径的公式非常相似，即：

$$R=\frac{S}{p}=\frac{2S}{C}$$

其中 R 是内切圆半径，S 是三角形面积，p,C 分别是半周长和周长。

§24.9 一些几何体

下面介绍一些几何体

阳马、堑堵、鳖臑

九章算术第五卷商功中介绍了一些特别的几何体：“斜解立方，得两堑堵。斜解堑堵，其一为阳马，一为鳖臑。阳马居二，鳖臑居一，不易之率也。合两鳖臑三而一，验之以棊，其形露矣。”

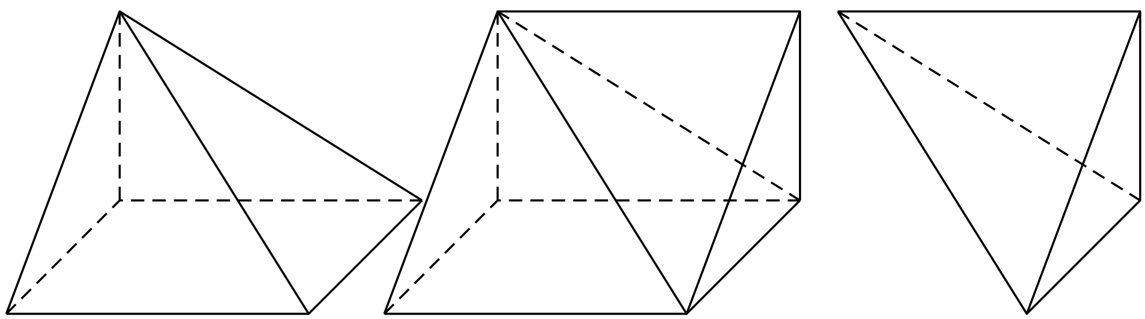


图 24.13: 阳马、堑堵、鳖臑

显然它们是共外接球的，如若遇到这种类型，可以构造长方体来求得外接球半径等。

其中我认为最有意思的是鳖臑模型，鳖臑可以看作是底面为直角三角形，高为不在直角这个端点（在其余两锐角其中一个的端点上）的三棱锥，由于鳖臑在长方体中，因此易证明其四个面都为直角三角形。

对棱模型

当四面体对棱相等时，我们可以构造一个长方体，使得四面体的三组对棱分别是这个长方体的三组面对角线。

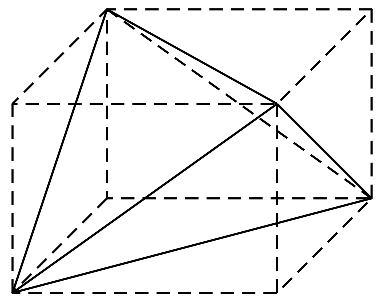


图 24.14: 对棱模型

如图，不妨设长方体三个边分别为 a, b, c，四面体三组对棱长度分别为 d, e, f，有

$$a^2 + b^2 = d^2$$
$$a^2 + c^2 = e^2$$
$$c^2 + b^2 = f^2$$

这个四面体的外接球半径 R 为

$$R = \sqrt{a^2 + b^2 + c^2} = \sqrt{2(d^2 + e^2 + f^2)}$$

第二十五章　向量

高中把平面向量和空间向量分为两块，实际上是没必要的（对于完整学习），所以我现在把他们捏在一起，在这一章中会简单介绍向量的运算以及性质，然后用向量研究一些平面几何或立体几何中的问题。

§25.1　向量的定义以及线性运算

向量的定义

既有大小又有方向的量称为**向量**（也称**矢量**），用 $\boldsymbol{a}, \boldsymbol{b}$ 或 $\vec{a}$ 表示；我们可以用一段有向线段来表示向量，将起点是 A 终点是 B 的向量用 $\overrightarrow{AB}$ 表示。

向量的大小也称向量的长度[①]，记作 $|\boldsymbol{a}|, \left|\overrightarrow{AB}\right|$，规定长度相等方向相同的有向线段为同一向量。

1. 长度为零的向量称为**零向量**，记作 $\mathbf{0}$，方向不确定，或者说方向任意。
2. 长度为一的向量称为**单位向量**，与 $\boldsymbol{a}$ 方向相同的单位向量记为 $\boldsymbol{a}^0$[②]。
3. 长度相等方向相反的向量为**反向量**，如 $\boldsymbol{a}, -\boldsymbol{a}$，$\overrightarrow{AB} = -\overrightarrow{BA}$。

此外我们还有如下定义

方向向量：直线 l 上有一向量 $\boldsymbol{a}$，我们把与 $\boldsymbol{a}$ 平行的非零向量称为直线 l 的方向向量。

共面向量：平行于同一个平面的向量。

向量的加法

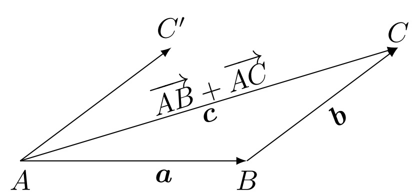


图 25.1: 向量的三角形法则

向量的加法

对于 $\boldsymbol{a}, \boldsymbol{b}$ 做 $\overrightarrow{AB} = \boldsymbol{a}, \overrightarrow{BC} = \boldsymbol{b}, \overrightarrow{AC}$ 表示为 $\boldsymbol{c}$ 称为 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的和，记作 $\boldsymbol{c} = \boldsymbol{a} + \boldsymbol{b}, \overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$，这个公式表示向量的加法规则，称为**三角形法则**。

此外，还有**平行四边形法则**，不过平行四边形法则就是三角形法则换一种说法，如图中 $\overrightarrow{AC'} = \overrightarrow{BC}$ 所以有 $\overrightarrow{AB} + \overrightarrow{AC'} = \overrightarrow{AC}$ 所以只用知道三角形法则就行了。

[①]向量的长度也叫模长

[②]这个记号在解析几何中常用

对于向量的减法，也可以用三角形法则来定义，将

$$\boldsymbol{a}-\boldsymbol{b}$$

看作 $\boldsymbol{a}+(-\boldsymbol{b})$。

对于向量加法有以下规律：

定理

1. 结合律 $(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}=\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c})$
2. 交换律 $\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{b}+\boldsymbol{a}$
3. $\boldsymbol{a}+\boldsymbol{0}=\boldsymbol{a}$
4. $\boldsymbol{a}+(-\boldsymbol{a})=\boldsymbol{0}$

♠

向量的数量乘法

实数 λ 与 $\boldsymbol{a}$ 的乘积 $\lambda\boldsymbol{a}$ 是一个向量，其长度为

$$|\lambda\boldsymbol{a}|=\lambda\,|\boldsymbol{a}|$$

它的方向当 $\lambda>0$ 时与 $\boldsymbol{a}$ 的方向相同，当 $\lambda<0$ 时与 $\boldsymbol{a}$ 的方向相反。

设 $\boldsymbol{a}\neq 0$ 因为 $|\boldsymbol{a}|^{-1}\boldsymbol{a}$ 与 $\boldsymbol{a}$ 同向，并且

$$||\boldsymbol{a}|^{-1}\boldsymbol{a}|=|\boldsymbol{a}|^{-1}|\boldsymbol{a}|$$

所以 $\boldsymbol{a}^0=\frac{\boldsymbol{a}}{|\boldsymbol{a}|}$，得到一个与 $\boldsymbol{a}$ 同向的单位向量，这称为把 $\vec{a}$ **单位化**。

向量的数量积满足以下规律：以下 $\lambda,\mu\in\mathbb{R}$

推论

1. $1\boldsymbol{a}=\boldsymbol{a},-\boldsymbol{a}=(-1)\boldsymbol{a}$
2. $\lambda\left(\mu\boldsymbol{a}\right)=\left(\lambda\mu\right)\boldsymbol{a}$
3. $\left(\lambda+\mu\right)\boldsymbol{a}=\lambda\boldsymbol{a}+\mu\boldsymbol{a}$
4. $\lambda\left(\boldsymbol{a}+\boldsymbol{b}\right)=\lambda\boldsymbol{a}+\lambda\boldsymbol{b}$

◇

§25.1.1 共线以及共面的向量组

以上介绍了向量的加法以及数量乘法，它们统称为向量的线性运算。

设 $\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_n$ 是一组向量，$k_1,k_2,\ldots,k_n$ 是一组实数，则

$$\sum_{i=1}^{n}k_i\boldsymbol{a}_i=k_1\boldsymbol{a}_1+k_2\boldsymbol{a}_2+\cdots+k_n\boldsymbol{a}_n$$

是一个向量，称它为 $\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_n$ 的一个**线性组合**，称 $k_1,k_2,\ldots,k_n$ 为这个组合的**系数**。

定义

向量组若用同一起点得到有向线段表示后，它们在一条直线（一个平面）上，则称这个向量组是**共线的（共面的）**。

若向量 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 满足

$$\boldsymbol{c}=\lambda\boldsymbol{a}+\mu\boldsymbol{b}$$

其中 λ,μ 是实数，那么 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 在同一平面。

证明：

当 $\boldsymbol{a},\boldsymbol{b}$ 共线，那么显然 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 共面；当 $\boldsymbol{a},\boldsymbol{b}$ 不共线时，即当 $\lambda,\mu\neq 0$ 时，根据三角形法则

$$\lambda\boldsymbol{a}+\mu\boldsymbol{b}$$

是与 $\boldsymbol{a},\boldsymbol{b}$ 共面的一向量，即 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 在同一平面。

□

对于任意三点 A,B,C 它们共线的条件是

$$\overrightarrow{OA}=\lambda\overrightarrow{OB}+\mu\overrightarrow{OC}$$

且 $\lambda+\mu=1$，O 是任取的一点。

以上这个命题在解决很多共线问题时是非常有用的，它的证明以及推论我放在后面与定比分点一起讨论。

下面看到对于以上这个命题的“升维”命题。

如果四点 A,B,C,D 共面则：

$$\overrightarrow{OA}=x\overrightarrow{OB}+y\overrightarrow{OC}+z\overrightarrow{OD}$$

其中 $x+y+z=1$，O 是任取的一点。

证明：

设 A 在平面 BCD 上则 $\overrightarrow{OA}$ 可以表示为

$$\begin{aligned}\overrightarrow{OA}&=x\overrightarrow{OB}+y\overrightarrow{OC}+z\overrightarrow{OD}\\ \overrightarrow{OA}&=(x+y+z)\overrightarrow{OA}+x\overrightarrow{AB}+y\overrightarrow{AC}+z\overrightarrow{AD}\\ -(x+y+z-1)\overrightarrow{OA}&=x\overrightarrow{AB}+y\overrightarrow{AC}+z\overrightarrow{AD}\end{aligned}$$

记右边的向量和为 $\overrightarrow{AA'}$，显然 $\overrightarrow{AA'}$ 在平面 BCD 上，而向量 $\overrightarrow{OA}$ 不在该平面上，即两向量不共线故

$$x+y+z-1=0\Rightarrow x+y+z=1$$

□

§25.2 几何空间的线性结构

几何空间 V 是空间中所有点组成的集合。取一点 O，以 O 为起点的向量称为**定位向量**，所有定位向量组成的集合与 V 一一对应，故 V 也可以看作由所有向量组成的集合。在之前的内容中，已经介绍了 V 中向量的

加法以及数量乘法运算，这为线性结构打下了良好的基础。

在几何空间中，非常直观的任意向量都可以表示为几个向量的线性组合，因此给出以下定理。

定理

几何空间 V 中，任意给定三个不共面的向量 $\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3$，则任意向量 $\boldsymbol{m}$ 可以被唯一的表示为 $\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3$ 的线性组合。

证明：

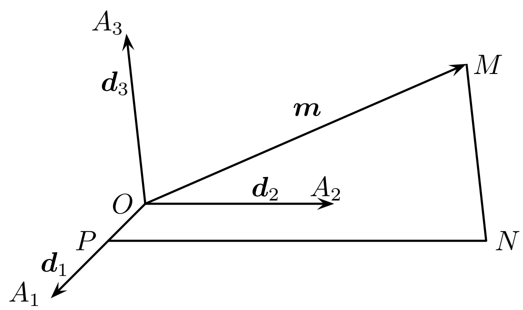


图 25.2: 向量可表性和唯一性的证明

可表性：取一点 O，作 $\overrightarrow{OA_1},\overrightarrow{OA_2},\overrightarrow{OA_3},\overrightarrow{OM}$ 分别表示 $\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3,\boldsymbol{m}$。

过点 M 作直线 MN 平行于 $\overrightarrow{OA_3}$，并且于平面 OA_1A_2 交于点 N；过点 N 作直线 NP 平行于 $\overrightarrow{OA_2}$，且与直线 OA_1 相交于点 P。

因为 $\overrightarrow{OP}$ 与 $\boldsymbol{d}_1$ 共线，$\overrightarrow{PN}$ 与 $\boldsymbol{d}_2$ 共线，$\overrightarrow{NM}$ 与 $\boldsymbol{d}_3$ 共线，因此存在实数 x,y,z 使得

$$\overrightarrow{OP}=x\boldsymbol{d}_1,\quad \overrightarrow{PN}=y\boldsymbol{d}_2,\quad \overrightarrow{NM}=z\boldsymbol{d}_3$$

从而

$$\boldsymbol{m}=\overrightarrow{OM}=\overrightarrow{OP}+\overrightarrow{PN}+\overrightarrow{NM}=x\boldsymbol{d}_1+y\boldsymbol{d}_2+z\boldsymbol{d}_3$$

唯一性：若

$$\boldsymbol{m}=x\boldsymbol{d}_1+y\boldsymbol{d}_2+z\boldsymbol{d}_3=x'\boldsymbol{d}_1+y'\boldsymbol{d}_2+z'\boldsymbol{d}_3$$

则有：

$$(x-x')\boldsymbol{d}_1+(y-y')\boldsymbol{d}_2+(z-z')\boldsymbol{d}_3=\boldsymbol{0}$$

由于 $\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3$ 不共面，则

$$\begin{cases} x &= x' \\ y &= y' \\ z &= z' \end{cases}$$

□

这个定理告诉我们，只要给定 V 中三个不共面的向量，那么 V 中的任意向量都可以被它们表示。

定义

几何空间 V 中任意三个有次序的不共面的向量 $\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3$ 称为 V 的一个**基**，对于几何空间中任一向量 $\boldsymbol{m}$，若

$$\boldsymbol{m}=x\boldsymbol{d}_1+y\boldsymbol{d}_2+z\boldsymbol{d}_3$$

则把三元有序实数组 (x,y,z) 称为 $\boldsymbol{m}$ 在基 $\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3$ 下的坐标，记作

$$(x,y,z)^T$$

以上是向量坐标的定义，下面我们引入空间中点的坐标。

定义

几何空间中的一个点 O 和一个基 $\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3$ 合在一起称为几何空间中的一个**仿射标架**或**仿射坐标系**，记作

$$[O;\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3]$$

其中，O 称为**原点**，对于几何空间中的任意一点 M，把它的定位向量 $\overrightarrow{OM}$ 在基 $\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3$ 下的坐标称为点 M 在仿射标架 $[O;\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3]$ 中的坐标。

通过定位向量，几何空间中的点和三元有序数组的集合之间建立了一个一一对应。

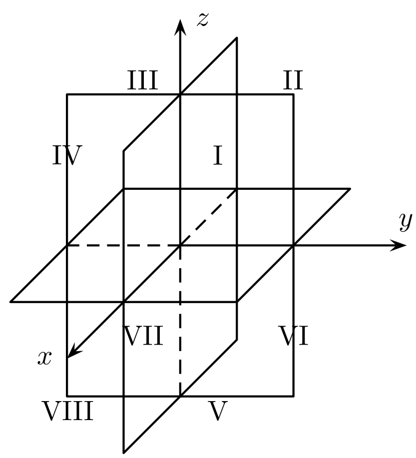


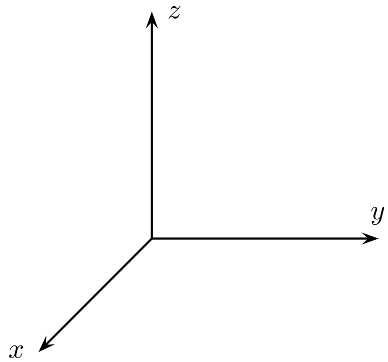
图 25.3: 八个卦限

设 $[O;\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3]$ 为几何空间的一个仿射标架，过原点 O 且分别以 $\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3$ 为方向的有向直线分别称为 x 轴、y 轴、z 轴，统称**坐标轴**，由每两根坐标轴决定的平面称为**坐标平面**，它们分别为 Oxy 平面、Oyz 平面、Oxz 平面；坐标平面把空间分成八个部分，称为八个**卦限**。

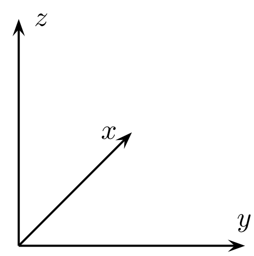
卦限	I	II	III	IV	V	VI	VII	VIII
x	+	−	−	+	+	−	−	+
y	+	+	−	−	+	+	−	−
z	+	+	+	+	−	−	−	−

将右手四指（除拇指）从 x 轴方向弯向 y 轴方向（转角小于 π），如果拇指所指的方向与 z 轴方向在 Oxy

平面的同侧，那么称此坐标系为**右手坐标系**，简称右手系；否则，称为**左手坐标系**，简称左手系。



如上为右手系，在高中的课本以及课标等所有涉及建系的内容，都统一使用右手直角坐标系，因此无需特别说明。



如上的左手系是相对较少见的，但在一些软件中仍会使用。

定义

如果 $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ 两两垂直，并且它们都是单位向量，则 $[O; \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3]$ 称为一个**直角标架**或**直角坐标系**。♡

可以看出，直角标架是仿射标架的一种特殊情况。

需要注意的是，高中的课表规定，向量的坐标（直角标架）应表示为

$$\boldsymbol{a} = (a_1, a_2, a_3)$$

而点的坐标表示为

$$A(a_1, a_2, a_3)$$

即向量的坐标须用等号连接，而点的坐标不能用等号连接。

§25.3 用坐标做向量的线性运算

取仿射标架 $[O; \boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3]$，设 $\boldsymbol{a}$ 的坐标是 $(a_1, a_2, a_3)^T$，$\boldsymbol{b}$ 的坐标是 $(b_1, b_2, b_3)^T$，则

$$\begin{aligned} \boldsymbol{a} + \boldsymbol{b} &= a_1\boldsymbol{d}_1 + a_2\boldsymbol{d}_2 + a_3\boldsymbol{d}_3 + b_1\boldsymbol{d}_1 + b_2\boldsymbol{d}_2 + b_3\boldsymbol{d}_3 \\ &= (a_1 + b_1)\boldsymbol{d}_1 + (a_2 + b_2)\boldsymbol{d}_2 + (a_3 + b_3)\boldsymbol{d}_3 \end{aligned}$$

因此 $\boldsymbol{a} + \boldsymbol{b}$ 的坐标为 $(a_1 + b_1, a_2 + b_2, a_3 + b_3)^T$。再来看数量乘法，设 $\lambda \in \mathbb{R}$，则

$$\lambda\boldsymbol{a} = \lambda a_1\boldsymbol{d}_1 + \lambda a_2\boldsymbol{d}_2 + \lambda a_3\boldsymbol{d}_3$$

因此 $\lambda\boldsymbol{a}$ 的坐标为 $(\lambda a_1, \lambda a_2, \lambda a_3)^T$。

对于减法，由向量的加法以及数量乘法立即得到

$$\boldsymbol{a}-\boldsymbol{b} \text{ 的坐标为 } (a_1-b_1, a_2-b_2, a_3-b_3)^T$$

§25.3.1 共线的条件

设平面上的两个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的坐标分别为 $(a_1, a_2)^T, (b_1, b_2)^T$，则它们共线的充要条件为

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = 0$$

即 $a_1 b_2 = a_2 b_1$。

证明：（只证必要性）设 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 共线，则存在实数 λ 使得 $\boldsymbol{b} = \lambda\boldsymbol{a}$，从而有

$$b_i = \lambda a_i,\ i = 1, 2$$

即

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1 b_2 - a_2 b_1 = \lambda a_1 a_2 - \lambda a_1 a_2 = 0$$

□

对于平面上三个点 A, B, C，它们的坐标为 $(a_1, a_2)^T, (b_1, b_2)^T, (c_1, c_2)^T$，倘若三点共线，即有

$$\overrightarrow{AB} = \lambda\overrightarrow{BC}$$

更加以上的结论可知

$$0 = \begin{vmatrix} b_1 - a_1 & c_1 - b_1 \\ b_2 - a_2 & c_2 - b_2 \end{vmatrix} = \begin{vmatrix} a_1 - b_1 & c_1 - b_1 & b_1 \\ a_2 - b_2 & c_2 - b_2 & b_2 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} a_1 & c_1 & b_1 \\ a_2 & c_2 & b_2 \\ 1 & 1 & 1 \end{vmatrix}$$

下面用向量方法证明一些之前用纯几何证明过的三角形中的定理：

梅涅劳斯定理

塞瓦定理

§25.3.2 三点共线的条件与定比分点

平面上三点 A, B, C，取仿射标架 $[O; \overrightarrow{OA}, \overrightarrow{OB}]$，则 $\overrightarrow{OC}$ 可表示为

$$\overrightarrow{OC} = x\overrightarrow{OA} + y\overrightarrow{OB}$$

其中，A, B, C 共线 $\iff$ $x + y = 1$。

证明：

先证必要性：已知 $x+y=1$，则 $\overrightarrow{OC}$ 可表示为

$$\overrightarrow{OC}=x\overrightarrow{OA}+(1-x)\overrightarrow{OB}$$

展开即得

$$\overrightarrow{OC}+\overrightarrow{BO}=x\left(\overrightarrow{OA}+\overrightarrow{BO}\right)$$
$$\overrightarrow{BC}=x\overrightarrow{BA}$$

即证 A,B,C 共线。

再证明充分性，我一并引入定比分点：

定义

对于线段 $AB(A\neq B)$，如果点 C 满足：

$$\overrightarrow{AC}=\lambda\overrightarrow{CB}$$

则称**点 C 分线段 AB 成定比** λ。

1. 当 $\lambda>0$ 时，$\overrightarrow{AC}$ 与 $\overrightarrow{CB}$ 同向，点 C 是线段 AB 内部的点，称 C 为**内分点**；
2. 当 $\lambda<0$ 时，$\overrightarrow{AC}$ 与 $\overrightarrow{CB}$ 反向，点 C 是线段 AB 外部的点，称 C 为**外分点**；
3. 当 $\lambda=0$ 时，点 C 与点 A 重合，假如 $\lambda=-1$，则 $\overrightarrow{AC}=-\overrightarrow{CB}$，即 $\overrightarrow{AB}=\overrightarrow{0}$，矛盾，所以 $\lambda\neq-1$。♡

引理

在线段 AB 上有内分点 C，设 $|AC|=m,|CB|=n,n,m\neq0$，则有

$$\overrightarrow{OC}=\frac{n}{m+n}\overrightarrow{OA}+\frac{m}{m+n}\overrightarrow{OB}$$

其中 O 是平面上异于这三点的一点 ♣

证明：O 在直线 AB 上时的证明留给读者，现在讨论 O 不在直线 AB 上，O,A,B 构成一个三角形如下图

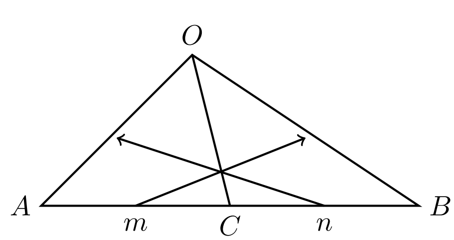


图 25.4: 定比分点

其中 $|AC| = m, |BC| = n$，则有

$$\begin{aligned}\overrightarrow{OC} &= \overrightarrow{OA} + \overrightarrow{AC} \\ &= \overrightarrow{OA} + \frac{m}{m+n}\overrightarrow{AB} \\ &= \frac{n}{m+n}\overrightarrow{OA} + \frac{m}{m+n}\overrightarrow{OB}\end{aligned}$$

□

显然 $\frac{m}{m+n} + \frac{n}{m+n} = 1$，当 C 是外分点时命题也成立，证明留给读者。

特别的，当 $m = n$ 时，即内分点 C 为 AB 中点时，有：

$$2\overrightarrow{OC} = \overrightarrow{OA} + \overrightarrow{OB}$$

在初中阶段被称为“倍长中线”。

定理

线段 AB 上有一点 P 分线段 AB 成定比 λ，设 $A(x_1, y_1), B(x_2, y_2), P(x, y)$，则有：

$$\begin{cases} x &= \frac{x_1+\lambda x_2}{1+\lambda} \\ y &= \frac{y_1+\lambda y_2}{1+\lambda} \end{cases}$$

这被称为**定比分点坐标公式**。

证明：因为 $\frac{|AP|}{|PB|} = \lambda$，由引理知

$$\begin{aligned}\overrightarrow{OP} &= \frac{|BP|}{|AP|+|BP|}\overrightarrow{OA} + \frac{|AP|}{|AP|+|BP|}\overrightarrow{OB} \\ &= \frac{1}{\lambda+1}\overrightarrow{OA} + \frac{\lambda}{\lambda+1}\overrightarrow{OB}\end{aligned}$$

带入它们的坐标即证：

$$(x, y) = \left(\frac{x_1+\lambda x_2}{\lambda+1}, \frac{y_1+\lambda y_2}{\lambda+1}\right)$$

□

当然这个定理通过 $\overrightarrow{AP} = \lambda\overrightarrow{PB}$ 也可以容易地证明，只需带入坐标即可。

§25.4 向量的内积

§25.4.1 射影和分量

几何空间 V 中，给了一个单位向量 $\boldsymbol{e}$，过点 O 作直线 l，其方向向量为 $\boldsymbol{e}$；过点 O 作平面 π 与 l 垂直，在平面 π 上取两个互相垂直的单位向量 $\boldsymbol{e_1}, \boldsymbol{e_2}$，如图所示，则 $[O; \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}]$ 是几何空间 V 的一个直角坐标系，于是，任意向量 $\boldsymbol{a}$，它可以唯一的分解成

$$\boldsymbol{a} = x\boldsymbol{e}_1 + y\boldsymbol{e}_2 + z\boldsymbol{e} = \boldsymbol{a}_2 + \boldsymbol{a}_1$$

其中 $\boldsymbol{a}_1 = x\boldsymbol{e}_1 + y\boldsymbol{e}_2, \boldsymbol{a}_2 = z\boldsymbol{e}$。

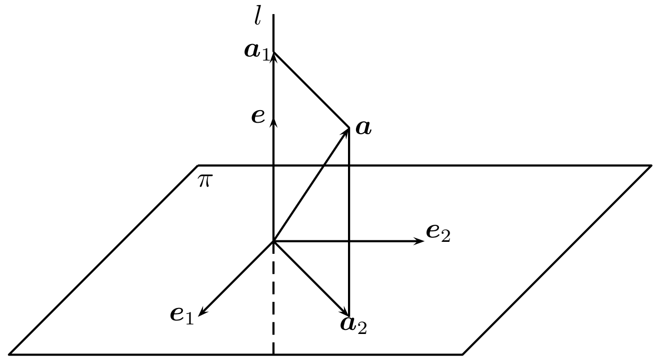


图 25.5: 射影

可见 $\boldsymbol{a_2} \perp \boldsymbol{e}$，$\boldsymbol{a}_1$ 与 $\boldsymbol{e}$ 共线，我们把 $\boldsymbol{a}_1$ 称为 $\boldsymbol{a}$ 在方向 $\boldsymbol{e}$ 上的**内射影**（也称 $\boldsymbol{a}_1$ 是 $\boldsymbol{a}$ 在方向向量为 $\boldsymbol{e}$ 的轴 l 上的正投影），记作 $\mathbf{P}_{\boldsymbol{e}}(\boldsymbol{a})$；把 $\boldsymbol{a}_2$ 称为 $\boldsymbol{a}$ 沿方向 $\boldsymbol{e}$ 的**外射影**。

内射影的一些基本性质：对于几何空间中任意向量 $\boldsymbol{a}, \boldsymbol{b}$，任意实数 λ，有

$$\mathbf{P}_{\boldsymbol{e}}(\boldsymbol{a}+\boldsymbol{b}) = \mathbf{P}_{\boldsymbol{e}}(\boldsymbol{a}) + \mathbf{P}_{\boldsymbol{e}}(\boldsymbol{b}) \quad ①$$
$$\mathbf{P}_{\boldsymbol{e}}(\lambda\boldsymbol{a}) = \lambda\mathbf{P}_{\boldsymbol{e}}(\boldsymbol{a}) \quad ②$$

由于 $\boldsymbol{a}$ 在方向 $\boldsymbol{e}$ 上的内射影 $\boldsymbol{a}_1$ 与 $\boldsymbol{e}$ 共线，因此存在在一个唯一的实数 μ，使得 $\boldsymbol{a}_1 = \mu\boldsymbol{e}$，把这个实数 μ 称为 $\boldsymbol{a}$ 在方向 $\boldsymbol{e}$ 上的**分量**，记作 $\Pi_{\boldsymbol{e}}(\boldsymbol{a})$。

几何空间中任意向量 $\boldsymbol{a}$ 在方向 $\boldsymbol{e}$ 上的分量为

$$\Pi_{\boldsymbol{e}} = |\boldsymbol{a}| \cos\langle\boldsymbol{a}, \boldsymbol{e}\rangle$$

其中 $\langle\boldsymbol{a}, \boldsymbol{e}\rangle$ 表示向量 $\boldsymbol{a}$ 与 $\boldsymbol{e}$ 之间的夹角。

§25.4.2　内积的定义与性质

定义

$\boldsymbol{a}, \boldsymbol{b}$ 的内积记作 $\boldsymbol{a} \cdot \boldsymbol{b}$ 规定为一个实数：

$$\boldsymbol{a} \cdot \boldsymbol{b} = |\boldsymbol{a}|\,|\boldsymbol{b}| \cos\langle\boldsymbol{a}, \boldsymbol{b}\rangle$$

由分量也可以的定义为也定义为:

$$\boldsymbol{a} \cdot \boldsymbol{b} = \left(\Pi_{\boldsymbol{b}^0}(\boldsymbol{a})\right)|\boldsymbol{b}|$$

其中 $\Pi_{\boldsymbol{b}^0}(\boldsymbol{a})$ 是 $\boldsymbol{a}$ 在方向 $\boldsymbol{b}^0$ 的分量

可以得出一些简单的推导：

1. $|\boldsymbol{a}| = \sqrt{\boldsymbol{a} \cdot \boldsymbol{a}}$
2. $\cos\langle\boldsymbol{a}, \boldsymbol{b}\rangle = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|}$

向量的内积满足以下规律：

1. 对称性 $\boldsymbol{a}\cdot\boldsymbol{b}=\boldsymbol{b}\cdot\boldsymbol{a}$

2. 线性性质其一 $(\lambda\boldsymbol{a})\cdot\boldsymbol{b}=\lambda(\boldsymbol{a}\cdot\boldsymbol{b})$

3. 线性性质其二 $(\boldsymbol{a}+\boldsymbol{c})\cdot\boldsymbol{b}=\boldsymbol{a}\cdot\boldsymbol{b}+\boldsymbol{c}\cdot\boldsymbol{b}$

4. 正定性 $\boldsymbol{a}\cdot\boldsymbol{a}\geqslant 0$，等号当且仅当 $\boldsymbol{a}=\boldsymbol{0}$ 时成立。

§25.4.3 用坐标计算向量的内积

取一仿射标架 $[O;\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3]$，设 $\boldsymbol{a},\boldsymbol{b}$ 的坐标分别为 $(a_1,a_2,a_3)^T,(b_1,b_2,b_3)^T$，则根据内积的定义

$$\begin{aligned}\boldsymbol{a}\cdot\boldsymbol{b}&=(a_1\boldsymbol{d}_1+a_2\boldsymbol{d}_2+a_3\boldsymbol{d}_3)\cdot(a_1\boldsymbol{d}_1+a_2\boldsymbol{d}_2+a_3\boldsymbol{d}_3)\\&=a_1b_1\boldsymbol{d}_1\cdot\boldsymbol{d}_1+a_1b_2\boldsymbol{d}_1\cdot\boldsymbol{d}_2+a_1b_3\boldsymbol{d}_1\cdot\boldsymbol{d}_3\\&+a_2b_1\boldsymbol{d}_2\cdot\boldsymbol{d}_1+a_2b_2\boldsymbol{d}_2\cdot\boldsymbol{d}_2+a_2b_3\boldsymbol{d}_2\cdot\boldsymbol{d}_3\\&+a_3b_1\boldsymbol{d}_3\cdot\boldsymbol{d}_1+a_3b_2\boldsymbol{d}_3\cdot\boldsymbol{d}_2+a_3b_3\boldsymbol{d}_3\cdot\boldsymbol{d}_3\end{aligned}$$

由上可知，当知道 $\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3$ 两两之间的内积，即可知道任意两个向量的内积。上式中两两之间的内积共有 9 个，它们被称为仿射标架 $[O;\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3]$ 的**度量参数**。

现在设 $[O;\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_3]$ 为直角标架，注意到

$$\boldsymbol{e}_i\cdot\boldsymbol{e}_i=1,\quad \boldsymbol{e}_i\cdot\boldsymbol{e}_j=0,i\neq j$$

于是得到

$$\boldsymbol{a}\cdot\boldsymbol{b}=a_1b_1+a_2b_2+a_3b_3$$

也即

定理

在直角坐标系中，两个向量的内积等于它们对应坐标的乘积之和。♠

类似的，可以证明在平面直角坐标系中，两个向量的内积等于它们对应坐标之和，设 $\boldsymbol{a},\boldsymbol{b}$ 的坐标分别为 $(a_1,a_2),(b_1,b_2)$ 即有

$$\boldsymbol{a}\cdot\boldsymbol{b}=a_1b_1+a_2b_2$$

对于 $\boldsymbol{a}=(a_1,a_2,a_3)$，可知其长度为

$$|\boldsymbol{a}|=\sqrt{\boldsymbol{a}^2}=\sqrt{a_1^2+a_2^2+a_3^2}$$

在空间中的两点 A,B 的距离可以看作 $|AB|=|\overrightarrow{AB}|$，因此设 $A(x_1,y_1,z_1),B(x_2,y_2,z_2)$，那么有

$$|AB|=\sqrt{x_2-x_1{}^2+y_2-y_1{}^2+z_2-z_1{}^2}$$

垂直的条件

当两个向量 $\boldsymbol{a},\boldsymbol{b}$ 垂直时，有 $\cos\langle\boldsymbol{a},\boldsymbol{b}\rangle=0$，易知

$$\boldsymbol{a}\cdot\boldsymbol{b}=0$$

若 $\boldsymbol{a},\boldsymbol{b}$ 的坐标分别为 $(a_1,a_2,a_3),(b_1,b_2,b_3)$，那么用坐标表示为

$$a_1b_1+a_2b_2+a_3b_3=0$$

同理，在平面上设 $\boldsymbol{a},\boldsymbol{b}$ 的坐标分别为 $(a_1,a_2),(b_1,b_2)$ 有

$$a_1b_1+a_2b_2=0$$

§25.4.4 方向角与方向余弦

在直角坐标系中，还可以用向量 $\boldsymbol{a}$ 与基向量的内积来计算 $\boldsymbol{a}$ 的坐标，设直角标架 $[O;\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_3]$，以及 $\boldsymbol{a}=(a_1,a_2,a_3)$ 则

$$\boldsymbol{a}=a_1\boldsymbol{e}_1+a_2\boldsymbol{e}_2+a_3\boldsymbol{e}_3$$

$\boldsymbol{a}$ 与基向量的内积为

$$\boldsymbol{a}\cdot\boldsymbol{e}_i=a_i,\ i=1,2,3$$

特别的，单位向量 $\boldsymbol{a}^0$ 的坐标为

$$\begin{cases} x &= \boldsymbol{a}^0\cdot\boldsymbol{e}_1=\cos\langle\boldsymbol{a},\boldsymbol{e}_1\rangle\\ y &= \boldsymbol{a}^0\cdot\boldsymbol{e}_2=\cos\langle\boldsymbol{a},\boldsymbol{e}_2\rangle\\ z &= \boldsymbol{a}^0\cdot\boldsymbol{e}_3=\cos\langle\boldsymbol{a},\boldsymbol{e}_3\rangle \end{cases}$$

我们把一个向量 $\boldsymbol{a}$ 与直角标架中的基向量 $\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_3$ 所成的角 α,β,γ 称为方向 $\boldsymbol{a}$ 的**方向角**，把方向角的余弦

$$\cos\alpha,\cos\beta,\cos\gamma$$

称为方向 $\boldsymbol{a}$ 的**方向余弦**，有上面的推导可知，$\boldsymbol{a}$ 的方向余弦就是 $\boldsymbol{a}^0$ 的直角坐标，且有

$$\cos^2\alpha+\cos^2\beta+\cos^2\gamma=1$$

§25.5 外积

定义

在空间中给定平面 α，以及直线 l，若有 $l\perp\alpha$，取 l 的方向向量 $\boldsymbol{a}$，我们称 $\boldsymbol{a}$ 为平面 α 的一个**法向量**，方向向量的长度随意但不能为 0。

不难知道，任意 $\boldsymbol{b}/\!/\alpha$ 都有：

$$\boldsymbol{b}\cdot\boldsymbol{a}=0$$

通过以上的结论，可以通过找到两个平行于平面 α 的向量（或者在平面 α 上找三点）即可联立方程解出

平面 α 的一个法向量。

定义

外积：两个向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的外积记作：

$$\boldsymbol{a}\times\boldsymbol{b}$$

仍是一个向量，它的长度规定为

$$|\boldsymbol{a}\times\boldsymbol{b}|:=|\boldsymbol{a}||\boldsymbol{b}|\sin\langle\boldsymbol{a},\boldsymbol{b}\rangle$$

且当 $|\boldsymbol{a}\times\boldsymbol{b}|\neq 0$ 时，它的方向规定为：与 $\boldsymbol{a},\boldsymbol{b}$ 均垂直，并且使 $(\boldsymbol{a},\boldsymbol{b},\boldsymbol{a}\times\boldsymbol{b})$ 成右手系，即当右手四指从 $\boldsymbol{a}$ 弯向 $\boldsymbol{b}$（转角小于 π）时，拇指指的就是 $\boldsymbol{a}\times\boldsymbol{b}$ 的方向。

易知，当 $\boldsymbol{a},\boldsymbol{b}$ 中有一个为 $\boldsymbol{0}$，则 $\boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{0}$，由定义可知 $\boldsymbol{a}\times\boldsymbol{b}$ 的充要条件是 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 共线，因此需要注意，若 $\boldsymbol{a}\times\boldsymbol{b}=0$，不能因此判断 $\boldsymbol{a},\boldsymbol{b}$ 中必有一个为 $\boldsymbol{0}$。

§25.5.1 外积的几何意义

当 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 不共线时，$|\boldsymbol{a}\times\boldsymbol{b}|$ 表示以 $\boldsymbol{a},\boldsymbol{b}$ 为邻边的平行四边形的面积。而 $\boldsymbol{a}\times\boldsymbol{b}$ 方向的几何意义是给平面定向。

平面的定向，就是平面上的旋转方向，在平面几何中，我们只需要用“逆时针方向”或“顺时针方向”即可描述平面上两个方向的旋转，但是在三维空间中的平面，这种说法不足以描述平面上的旋转方向，因为在平面一侧看是顺时针旋转，而在另一侧看就是逆时针旋转，因此通常用另一种方法来描述。

给定平面 π_0 上的一对不共线的向量，如果规定了它们先后顺序，则从第一个向量到第二个向量的转角小于 π 的旋转方向就称为平面 π_0 的一个定向。

平面的两个定向，也可以用平面的两侧来代表：如果右手四指沿平面上取定的旋转方向弯曲，拇指必指向平面的一侧，这样，平面的两个定向就对应于平面的两侧，而平面两侧又可以用垂直于该平面的两个方向来刻画，因此通常也用垂直于平面的方向来表示平面的定向。

现在来看外积 $\boldsymbol{a}\times\boldsymbol{b}$ 的方向的几何意义，$\boldsymbol{a}\times\boldsymbol{b}$ 的方向给出了以 $\boldsymbol{a},\boldsymbol{b}$ 为邻边的平行四边形的边界的一个环形方向，即让右手的拇指指向 $\boldsymbol{a}\times\boldsymbol{b}$ 的方向，右手其余四指的弯向（转角小于 π）就是以 $\boldsymbol{a},\boldsymbol{b}$ 为邻边的平行四边形的边界环形方向，则称它是**定向平行四边形**。因此，$\boldsymbol{a}\times\boldsymbol{b}$ 的方向的几何意义就是以 $\boldsymbol{a},\boldsymbol{b}$ 为邻边的平行四边形确定了一个方向。

外积的运算规律

对于任意向量 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 和任意实数 λ，有

1. $\boldsymbol{a}\times\boldsymbol{b}=-\boldsymbol{b}\times\boldsymbol{a}$（反交换律）；

2. $(\lambda\boldsymbol{a})\times\boldsymbol{b}=\lambda(\boldsymbol{a}\times\boldsymbol{b})=\boldsymbol{a}\times(\lambda\boldsymbol{b})$；

3. $\boldsymbol{a}\times(\boldsymbol{b}+\boldsymbol{c})=\boldsymbol{a}\times\boldsymbol{b}+\boldsymbol{a}\times\boldsymbol{c}$（左分配律）；

   $(\boldsymbol{b}+\boldsymbol{c})\times\boldsymbol{a}=\boldsymbol{b}\times\boldsymbol{a}+\boldsymbol{c}\times\boldsymbol{a}$（右分配律）。

§25.5.2 坐标计算外积

在右手系直角标架 $[O;\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3]$，设向量 $\boldsymbol{a},\boldsymbol{b}$ 坐标分别为 $(a_1,a_2,a_3)^T,(b_1,b_2,b_3)^T$，则有

$$
\begin{aligned}
\boldsymbol{a}\times\boldsymbol{b} &= (a_1\boldsymbol{d}_1+a_2\boldsymbol{d}_2+a_3\boldsymbol{d}_3)\times(b_1\boldsymbol{d}_1+b_2\boldsymbol{d}_2+b_3\boldsymbol{d}_3)\\
&= (a_1b_2-a_2b_1)\boldsymbol{d}_1\times\boldsymbol{d}_2+(a_3b_1-a_1b_3)\boldsymbol{d}_3\times\boldsymbol{d}_1\\
&+(a_2b_3-a_3b_2)\boldsymbol{d}_2\times\boldsymbol{d}_3
\end{aligned}
$$

现设 $[O;\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_3]$ 是右手直角标架，根据

$$\boldsymbol{e}_1\times\boldsymbol{e}_2=\boldsymbol{e}_3,\quad \boldsymbol{e}_2\times\boldsymbol{e}_3=\boldsymbol{e}_1,\quad \boldsymbol{e}_3\times\boldsymbol{e}_1=\boldsymbol{e}_2$$

可以得到

$$
\begin{aligned}
\boldsymbol{a}\times\boldsymbol{b} &= (a_2b_3-a_3b_2)\boldsymbol{e}_1+(a_3b_1-a_1b_3)\boldsymbol{e}_2+(a_1b_2-a_2b_1)\boldsymbol{e}_3\\
&= \begin{vmatrix} a_2 & b_2\\ a_3 & b_3\end{vmatrix}\boldsymbol{e}_1-\begin{vmatrix} a_1 & b_1\\ a_3 & b_3\end{vmatrix}\boldsymbol{e}_2+\begin{vmatrix} a_1 & b_1\\ a_2 & b_2\end{vmatrix}\boldsymbol{e}_3
\end{aligned}
$$

于是我们有

定理

设 $\boldsymbol{a},\boldsymbol{b}$ 在有时直角坐标系中的坐标分别为

$$(a_1,a_2,a_3)^T,\quad (b_1,b_2,b_3)^T$$

则 $\overrightarrow{a}\times\overrightarrow{b}$ 的坐标为

$$\left(\begin{vmatrix} a_2 & b_2\\ a_3 & b_3\end{vmatrix},-\begin{vmatrix} a_1 & b_1\\ a_3 & b_3\end{vmatrix},\begin{vmatrix} a_1 & b_1\\ a_2 & b_2\end{vmatrix}\right)^T$$

从而

$$|\boldsymbol{a}\times\boldsymbol{b}|=\sqrt{(a_2b_3-a_3b_2)^2+(a_3b_1-a_1b_3)^2+(a_1b_2-a_2b_1)^2}$$

由外积的几何意义可知，上式也是以 $\boldsymbol{a},\boldsymbol{b}$ 为邻边的平行四边形的面积公式。

右手直角坐标系中 $\boldsymbol{a}\times\boldsymbol{b}$ 也可以写成

$$\boldsymbol{a}\times\boldsymbol{b}=\begin{vmatrix} \boldsymbol{e}_1 & a_1 & b_1\\ \boldsymbol{e}_2 & a_2 & b_2\\ \boldsymbol{e}_3 & a_3 & b_3\end{vmatrix}$$

§25.5.3 二重外积

我们考虑三个向量间外积的运算

$$\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c})$$

当括号不同时，上式会发生改变吗？也即外积符合结合律吗？

由外积的定义可知，$\boldsymbol{b},\boldsymbol{c}$（设它们不共线）确定一个平面 π，又由于 $\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c})$ 与 $\boldsymbol{b}\times\boldsymbol{c}$ 垂直，因此 $\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c})$ 与 pi 平行，也即在平面内，从而有

$$\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c})=k_1\boldsymbol{b}+k_2\boldsymbol{c}$$

其中 k_1,k_2 待定，但可以用这三个向量表示，下面证明一下：

取一个右手直角坐标系，设

$$\boldsymbol{a}=(a_1,a_2,a_3),\ \boldsymbol{b}=(b_1,b_2,b_3),\ \boldsymbol{c}=(c_1,c_2,c_3)$$

记 $\boldsymbol{b}\times\boldsymbol{c}=(d_1,d_2,d_3)$，以及 $\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c})=(e_1,e_2,e_3)$，则有

$$\begin{aligned}
e_1&=a_2d_3-a_3d_2\\
&=a_2(b_1c_2-b_2c_1)-a_3(b_3c_1-b_1c_3)\\
&=b_1(\boldsymbol{a}\cdot\boldsymbol{c}-a_1c_1)-c_1(\boldsymbol{a}\cdot\boldsymbol{b}-a_1b_1)\\
&=b_1(\boldsymbol{a}\cdot\boldsymbol{c})-c_1(\boldsymbol{a}\cdot\boldsymbol{b})
\end{aligned}$$

同理可得

$$\begin{aligned}
e_2&=b_2(\boldsymbol{a}\cdot\boldsymbol{c})-c_2(\boldsymbol{a}\cdot\boldsymbol{b})\\
e_3&=b_3(\boldsymbol{a}\cdot\boldsymbol{c})-c_3(\boldsymbol{a}\cdot\boldsymbol{b})
\end{aligned}$$

综上可知

$$\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c})=(\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b}-(\boldsymbol{a}\cdot\boldsymbol{b})\boldsymbol{c}$$

上式被称为**二重外积公式**，再来由反交换律可知

$$(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}=-\boldsymbol{c}\times(\boldsymbol{a}\times\boldsymbol{b})=-(\boldsymbol{b}\cdot\boldsymbol{c})\boldsymbol{a}+(\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b}$$

因此，在一般情况下 $\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c})\neq(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}$，即向量的外积不适合结合律。

Jacobi 等式

定理

$$\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c})+\boldsymbol{b}\times(\boldsymbol{c}\times\boldsymbol{a})+\boldsymbol{c}\times(\boldsymbol{a}\times\boldsymbol{b})=\boldsymbol{0}$$

♠

证明：由二重外积公式易知

$$\begin{aligned}
\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c})&=(\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b}-(\boldsymbol{a}\cdot\boldsymbol{b})\boldsymbol{c}\\
\boldsymbol{b}\times(\boldsymbol{c}\times\boldsymbol{a})&=(\boldsymbol{b}\cdot\boldsymbol{a})\boldsymbol{c}-(\boldsymbol{b}\cdot\boldsymbol{c})\boldsymbol{a}\\
\boldsymbol{c}\times(\boldsymbol{a}\times\boldsymbol{b})&=(\boldsymbol{c}\cdot\boldsymbol{b})\boldsymbol{a}-(\boldsymbol{c}\cdot\boldsymbol{a})\boldsymbol{b}
\end{aligned}$$

即证

$$
\begin{aligned}
&\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c})+\boldsymbol{b}\times(\boldsymbol{c}\times\boldsymbol{a})+\boldsymbol{c}\times(\boldsymbol{a}\times\boldsymbol{b})\\
=&(\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b}-(\boldsymbol{a}\cdot\boldsymbol{b})\boldsymbol{c}+(\boldsymbol{b}\cdot\boldsymbol{a})\boldsymbol{c}-(\boldsymbol{b}\cdot\boldsymbol{c})\boldsymbol{a}+(\boldsymbol{c}\cdot\boldsymbol{b})\boldsymbol{a}-(\boldsymbol{c}\cdot\boldsymbol{a})\boldsymbol{b}\\
=&\mathbf{0}
\end{aligned}
$$

§25.6　混合积

§25.7 用向量解决平面几何的问题

这一节中，主要用向量来讨论一些平面几何中的问题，通过向量这个工具，很多问题研究起来就会比较方便。

§25.7.1 极化恒等式

定理

对于任意 $\boldsymbol{a},\boldsymbol{b}$，有恒等式

$$\boldsymbol{a}\cdot\boldsymbol{b}=\frac{1}{4}\left[(\boldsymbol{a}+\boldsymbol{b})^2-(\boldsymbol{a}-\boldsymbol{b})^2\right]$$

这个恒等式被称为**极化恒等式**，它沟通了向量内积与长度的关系，通过构造平行四边形来证明也很直观，在一些夹角不易求得的情况下，这个恒等式常被使用。

下面来证明这个恒等式，显然当 $\boldsymbol{a},\boldsymbol{b}$ 任何一方为零向量或均为零向量时，这个恒等式显然成立，下面我们来讨论它们均不为零向量的情况；

已知 $\boldsymbol{a},\boldsymbol{b}$，构造以它们为邻边的平行四边形 $\square ABCD$

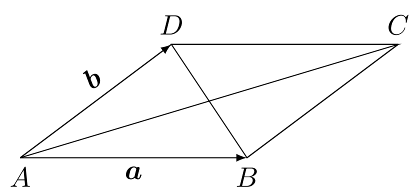


图 25.6: 极化恒等式

如图，$\boldsymbol{a}=\overrightarrow{AB},\boldsymbol{b}=\overrightarrow{AD},\boldsymbol{a}+\boldsymbol{b}=\overrightarrow{AC},\boldsymbol{a}-\boldsymbol{b}=\overrightarrow{DB}$，而 AC 与 DB 分别是 $\boldsymbol{a},\boldsymbol{b}$ 所“夹着”和“连接”的两条对角线，容易得到：

$$\overrightarrow{AC}^2=\boldsymbol{a}^2+\boldsymbol{b}^2+2\boldsymbol{a}\cdot\boldsymbol{b} \tag{1}$$

$$\overrightarrow{DB}^2=\boldsymbol{a}^2+\boldsymbol{b}^2-2\boldsymbol{a}\cdot\boldsymbol{b} \tag{2}$$

我们让 (1) 式减去 (2) 式

$$\begin{aligned}&\overrightarrow{AC}^2-\overrightarrow{DB}^2\\=&\boldsymbol{a}^2+\boldsymbol{b}^2+2\boldsymbol{a}\cdot\boldsymbol{b}-\boldsymbol{a}^2-\boldsymbol{b}^2+2\boldsymbol{a}\cdot\boldsymbol{b}\\=&4\boldsymbol{a}\cdot\boldsymbol{b}\end{aligned}$$

综上

$$\boldsymbol{a}\cdot\boldsymbol{b}=\frac{1}{4}(\overrightarrow{AC}^2-\overrightarrow{DB}^2)$$

$$\boldsymbol{a}\cdot\boldsymbol{b}=\frac{1}{4}\left[(\boldsymbol{a}+\boldsymbol{b})^2-(\boldsymbol{a}-\boldsymbol{b})^2\right]$$

□

可以数形结合，将 $\boldsymbol{a}+\boldsymbol{b}$ 和 $\boldsymbol{a}-\boldsymbol{b}$ 理解为这两向量构成的平行四边形的对角线；

另外通过倍长中线的知识，可以转化到三角形上，即对于三角形 $\triangle ABC$，O 为 BC 上的中点，有

$$\overrightarrow{AB}\cdot\overrightarrow{AC}=\overrightarrow{AO}^2-\frac{1}{4}\overrightarrow{BC}^2$$

这个证明是显然的，证明留给读者。

在这一节，如无特别说明，那么通常设 $\triangle ABC$ 中，顶点 A,B,C 对边分别是 a,b,c

§25.7.2 正弦定理

定理

$$\frac{a}{\sin A}=\frac{b}{\sin B}=\frac{c}{\sin C}=2R$$
$$\sin A=\frac{a}{2R}$$

其中 R 是 $\triangle ABC$ 的外接圆，正弦定理沟通了三角形内对角与对边的关系，常用于两边与两角的情况下。♠

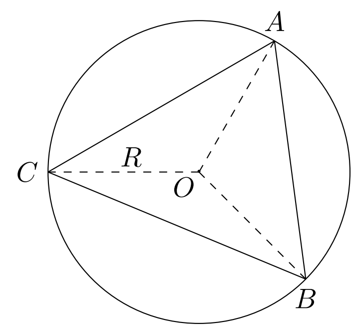


图 25.7: 正弦定理

证明正弦定理有许多方法，篇幅有限只选其中一种，下面来证明：

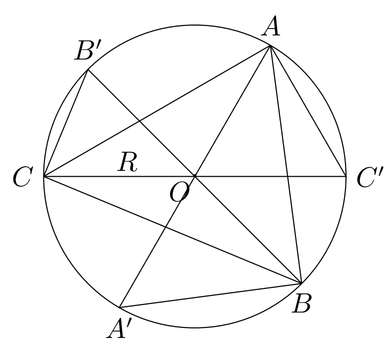


图 25.8: 正弦定理推导

对于 $\triangle ABC$，外心为 O，外接圆半径为 R，如图，分别连接并延长 AO,BO,CO 交圆于 A',B',C'，连接 $A'B,B'C,C'A$，由于

$$AA'=BB'=CC'=2R$$

因此有：

$$\angle ABA'=\angle BCB'=\angle CAC'=90\circ$$

根据圆的性质不难知道

$$\angle AA'B=\angle ACB,\ \angle BB'C=\angle BAC,\ \angle CC'A=\angle CBA$$

因此，可以得到：

$$\sin\angle ACB = \frac{AB}{AA'}$$
$$\sin\angle BAC = \frac{BC}{BB'}$$
$$\sin\angle CBA = \frac{AC}{CC'}$$

令 $\angle A = \angle BAC, \angle B = \angle CBA, \angle C = \angle ACB$，综上可得

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} = 2R$$

□

§25.7.3 余弦定理

定理

对于任意 $\triangle ABC$，有

$$c^2 = b^2 + a^2 - 2ab\cos C$$
$$\cos C = \frac{a^2+b^2-c^2}{2ab}$$

同理可推出 $\cos A, \cos B$，这个定理称为**余弦定理**。

下面只对 $\cos C$ 进行证明（其它两角的证明过程一样）：

对于 $\triangle ABC$，令 $AB = c, AC = b, BC = a$，过点 A 作 BD 的垂线 AD 垂足为 D；不难知道

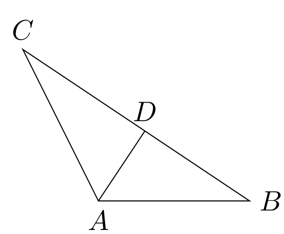


图 25.9: 余弦定理推导

$$\begin{aligned}
c^2 &= |AD|^2 + |BD|^2 \\
c^2 &= b^2\sin^2 C + (a - b\cos C)^2 \\
c^2 &= b^2\sin^2 C + a^2 + b^2\cos^2 C - 2ab\cos C \\
c^2 &= a^2 + b^2 - 2ab\cos C \\
\cos C &= \frac{a^2+b^2-c^2}{2ab}
\end{aligned}$$

□

同样的方法可推出 $\cos A, \cos B$。

此外，在高中常用向量推导，推导过程也更简单与直接：

$$\begin{aligned}\overrightarrow{AB} &= \overrightarrow{AC} + \overrightarrow{CB} \\ \overrightarrow{AB}^2 &= \overrightarrow{CA}^2 + \overrightarrow{CB}^2 - 2\cos C|CA|\cdot|BC| \\ c^2 &= b^2 + a^2 - 2\cos Cab\end{aligned}$$

□

§25.7.4 关于面积的推论

推论

$\triangle ABC$ 面积 $S_{\triangle ABC}$, 有

$$S_{\triangle ABC} = \frac{1}{2}ab\sin C = \frac{1}{2}bc\sin A = \frac{1}{2}ac\sin B$$

◇

下面证明其中一种，另外两种可以类似证明，证明 $S_{\triangle ABC} = \frac{1}{2}ab\sin C$，在 $\triangle ABC$ 中设边 a 上的高为 h，则有

$$h = b\sin C$$

因此三角形的面积可表示为

$$S_{\triangle ABC} = \frac{1}{2}ab\sin C$$

□

此外，关于三角形面积还有一个重要的定理———**海伦公式**

定理

海伦公式，若用 p 表示三角形的半周长，则面积可表示为：

$$S_{\triangle ABC} = \sqrt{p(p-a)(p-b)(p-c)}$$

♠

证明：通过上面的推论、三角恒等式以及余弦定理即证，不过列式比较复杂；

$$\begin{cases} S_{\triangle ABC} & = \frac{1}{2}ab\sin C \\ \sin^2 C + \cos^2 C & = 1 \\ \cos C & = \frac{a^2+b^2-c^2}{2ab} \end{cases}$$

综上，

$$\begin{aligned}
{S_{\triangle ABC}}^2 &= \frac{(2ab)^2}{16}\left(1-\left(\frac{a^2+b^2-c^2}{2ab}\right)^2\right)\\
&= \frac{1}{16}\left((2ab)^2-\left(a^2+b^2-c^2\right)^2\right)\\
&= \frac{1}{16}(2p)(2p-2a)(2p-2b)(2p-2c)\\
&= p(p-a)(p-b)(p-c)
\end{aligned}$$

□

海伦公式以对称性闻名，通过观察它的形式记忆即可

三角形中的齐次化简

当等式两边 a,b,c 和 $\sin A,\sin B,\sin C$ 齐次（指数的次数相同）时，通过正弦定理，即 $\frac{a}{\sin A}=2R$，可将等式两边转换。

例如:

$$\sin Ab=\sin Ca$$

$$ab=ac$$

此外可以根据恒等式构造

$$\begin{aligned}
&2\sin^2\alpha+\sin\alpha\cos\alpha\\
=&\frac{2\sin^2\alpha+\sin\alpha\cos\alpha}{\sin^2\alpha+\cos^2\alpha}\\
=&\frac{2\tan^2\alpha+\tan\alpha}{\tan^2\alpha+1}
\end{aligned}$$

上式说明，齐次不仅用在解三角形中可以应用，引入三角恒等式后对于任意角，齐次后变换，也是一个重要的工具。

三角形五心

三角形的五心分别是：重心、垂心、外心、内心以及旁心，其中前面四个是高中关注的重点，这一小节通过向量的角度来研究三角形的这五心，首先给出定义：

定义

重心：$\triangle ABC$ 的三条中线交于一点，该点被称为“重心”;

垂心：$\triangle ABC$ 的三条边的高（所在直线）交于一点，该点被称为“垂心”;

外心：$\triangle ABC$ 的外接圆的圆心，称为“外心”;

内心：$\triangle ABC$ 内切圆的圆心，称为“内心”;

旁心：$\triangle ABC$ 旁边圆（与三角形的一边和其他两边的延长线相切的圆）的圆心，称为“旁心”。

重心的性质

1. 重心到顶点的距离与重心到对边中点的距离之比为 $2:1$，也就是说重心在中线的三等分点上；
2. 重心与三角形任意两个顶点组成的三个三角形面积相等，即重心到三条边的距离与三条边的长成反比；
3. 重心到三角形三个顶点的距离的平方和最小；
4. 在平面直角坐标系中，重心的坐标是顶点的算术平均数，即设 A,B,C 的坐标分别为 (x_i,y_i)，$i=1,2,3$，则重心 $G=(\frac{x_1+x_2+x_3}{3},\frac{y_1+y_2+y_3}{3})$

下面来证明：

设 $\triangle ABC$，角 A,B,C 的对边分别为 a,b,c，重心为 G，边 a,b,c 上的中点分别为 D,E,F；对于性质 1：

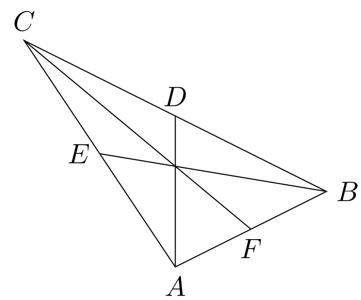


图 25.10: 重心

设 $\boldsymbol{a}=\overrightarrow{AB}$，$\boldsymbol{b}=\overrightarrow{AC}$，可以得到

$$\overrightarrow{AD}=\frac{1}{2}(\boldsymbol{a}+\boldsymbol{b})$$

由于 B,G,E 共线，不难知道

$$\overrightarrow{AG}=\lambda\boldsymbol{a}+\frac{1-\lambda}{2}\boldsymbol{b}$$

又由于 A,G,D 共线，可知

$$\mu\overrightarrow{AG}=\overrightarrow{AD}$$
$$\mu\lambda\boldsymbol{a}+\frac{\mu-\mu\lambda}{2}\boldsymbol{b}=\frac{1}{2}\boldsymbol{a}+\frac{1}{2}\boldsymbol{b}$$

因 $\boldsymbol{a},\boldsymbol{b}$ 不共线，因此得到二元方程组

$$\begin{cases}\mu\lambda &= \frac{1}{2}\\ \frac{\mu-\mu\lambda}{2} &= \frac{1}{2}\end{cases}$$

可以解得

$$\begin{cases}\mu &= \frac{3}{2}\\ \lambda &= \frac{1}{3}\end{cases}$$

□

即证性质 1，类似的可以证明另外两条中线的三等分点为重心。

对于性质 2，由性质一可知 $|AG|h=2|DG|h$，因此有

$$S_{\triangle GAB}=S_{\triangle GAC}=2S_{\triangle DGB}=S_{\triangle GBC}$$

□

先证性质 4

引理

若 G 是 $\triangle ABC$ 的重心，则满足：

$$\overrightarrow{GA}+\overrightarrow{GB}+\overrightarrow{GC}=\mathbf{0}$$

♣

证明：由性质 1，与倍长中线的推论可知：

$$\overrightarrow{GA}+\overrightarrow{GB}=2\overrightarrow{GF}=\overrightarrow{CG}$$

即证引理，这时引入平面直角坐标系，设 A,B,C 的坐标分别为 (x_i,y_i)，$i=1,2,3$，重心 $G(x_0,y_0)$，根据引理有：

$$\begin{aligned}&(x_1+x_2+x_3-3x_0,y_1+y_2+y_3-3y_0)=(0,0)\\ \Rightarrow &x_0=\frac{x_1+x_2+x_3}{3}\\ \Rightarrow &y_0=\frac{y_1+y_2+y_3}{3}\end{aligned}$$

综上，$G(\frac{x_1+x_2+x_3}{3},\frac{y_1+y_2+y_3}{3})$。

□

再来证明性质 3：任意一点 P 到三个顶点 A,B,C 的距离分别为 d_i，$i=1,2,3$，有

$$\begin{aligned}d_1^2+d_2^2+d_3^2&=\overrightarrow{PA}^2+\overrightarrow{PB}^2+\overrightarrow{PA}^2\\ &=(\overrightarrow{PG}+\overrightarrow{GA})^2+(\overrightarrow{PG}+\overrightarrow{GB})^2+(\overrightarrow{PG}+\overrightarrow{GC})^2\\ &=3\overrightarrow{PG}^2+2\overrightarrow{PG}(\overrightarrow{GA}+\overrightarrow{GB}+\overrightarrow{GC})+\overrightarrow{GA}^2+\overrightarrow{GB}^2+\overrightarrow{GC}^2\end{aligned}$$

根据引理可知

$$d_1^2+d_2^2+d_3^2=3\overrightarrow{PG}^2+\overrightarrow{GA}^2+\overrightarrow{GB}^2+\overrightarrow{GC}^2$$

其中当 $\overrightarrow{PG}=\mathbf{0}$ 时取得最小值，即 $P=G$ 时取到最小值。

□

垂心的性质

1. 三角形的外心 O，重心 G，垂心 H 三点共线，且 $OG:GH=1:2$ 此线被称为三角形的欧拉线（Euler line）。
2. 垂心到三角形任一顶点的距离等于此三角形外心到该顶点对边的距离的两倍。

垂心的性质在高考中基本不会涉及，最多是用在构造三角形内向量的关系。

下面来证明[3]：

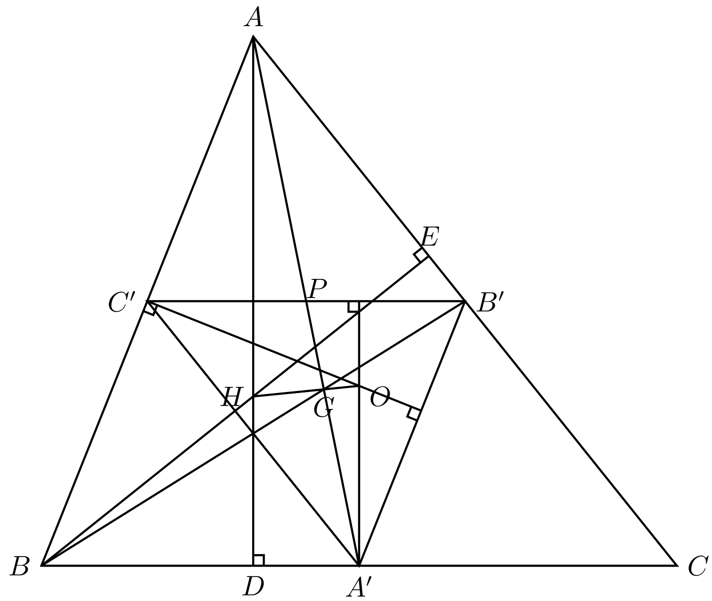


图 25.11: Medial triangle and Euler line

作三角形 $\triangle ABC$，顶点 A, B, C 对边的中点分别为 A', B', C'，连接中点可以得到中点三角形（Medial triangle)，作 BC, AC 的高，垂足为 D, E；可得垂心 H，中线的交点可得重心 G，作 $\triangle A'B'C'$ 的垂心 O，不难发现 O 即是 $\triangle ABC$ 的外心（因为满足中垂线这个条件）。

由中位线的知识可以推出：$\triangle ABC \sim \triangle A'B'C'$ 且 $AB : A'B' = 2 : 1$，因此可以看出 $AB'A'C'$ 为平行四边形（对边平行且相等)，得到 P 为 AA' 的中点；不仅如此，还可以推出 $\square BA'B'C'$，得到 $\triangle A'B'C'$ 的两中点、中线，进而我们推出 $\triangle A'B'C'$ 的重心也是 G；通过重心的性质可以得到 $AG : A'G = 2 : 1$；

又由于 $\triangle ABC, \triangle A'B'C'$ 相似，所以有 $AH : A'O = 2 : 1$；

因为 $AD \,//\, A'O$，因此 $\angle HAG = \angle OA'G$；

综上，$\triangle HAG \sim \triangle OA'G$，得到 $GH : OG = 2 : 1$。

□

性质 1 得证，其实在这过程中也证明了性质 2，故性质 1，2 得证。

内心的性质

1. 三角形的三个角平分线的交点交于一点，那一点即是“内心”；
2. 三角形的内心到边的距离（内切圆半径 r）与三边和三角形面积的关系：$r = \frac{2S_{\triangle ABC}}{a+b+c}$。
3. 直角三角形的内心到边的距离等于两直角边的和减去斜边的差的二分之一；

证明：

由于角平分线到两边距离相等，因此内心在角平分线上显然成立，性质 1 得证；

[3]欧拉线的证明引用 Geometry Revisited[?] 的证明方法

对于性质 2，作内心 O 到三角形三边的垂线 h（三条垂线长度相等），设内切圆半径为 r 易得：

$$\begin{aligned} S_{\triangle ABC} &= S_{\triangle OAB} + S_{\triangle OAC} + S_{\triangle OBC} \\ &= \frac{r}{2}(a+b+c) \end{aligned}$$

整理即得：

$$r = \frac{2S_{\triangle ABC}}{a+b+c}$$

即证性质 2，如果用 p 表示半周长，可以表示为 $r = \frac{S_{\triangle ABC}}{p}$，根据海伦公式：

$$r = \sqrt{\frac{(p-a)(p-b)(p-c)}{p}}$$

□

对于性质 3，不妨令 a, b 为直角边，则 $S_{\triangle ABC} = \frac{1}{2}ab$，根据性质 2：

$$r = \frac{ab}{a+b+c}$$

由勾股定理知：

$$\begin{aligned} 2ab &= 2ab + a^2 + b^2 - c^2 \\ 2ab &= (a+b)^2 - c^2 \\ 2ab &= (a+b+c)(a+b-c) \\ \frac{ab}{a+b+c} &= \frac{a+b-c}{2} \\ r &= \frac{a+b-c}{2} \end{aligned}$$

□

性质 3 得证。

外心的性质

1. 三角形的三条边的垂直平分线的交点为外心；
2. 若 O 是 $\triangle ABC$ 的外心，则 $\angle BOC = 2\angle A$（$\angle A$ 为锐角或直角），或 $\angle BOC = 2\pi - 2\angle A$（$\angle A$ 为钝角时）；
3. 外心到三顶点的距离相等。

外心的性质是显然的，证明留给读者，这里提供思路：

性质 1 可以由垂直平分线的性质证明；由于外心是外接圆圆心，显然性质 3 成立；性质 2 可以通过圆的性质证明。

旁心的性质

旁切圆[④]的性质

1. 三角形一内角平分线和另外两顶点处的外角平分线交于一点，该点即为三角形的旁心；

2. 每个三角形都有三个旁心；

3. 旁心到三角形的三边距离相等；

4. 旁心与半周长 p 形影不离，如图，$AG = AF = p$。

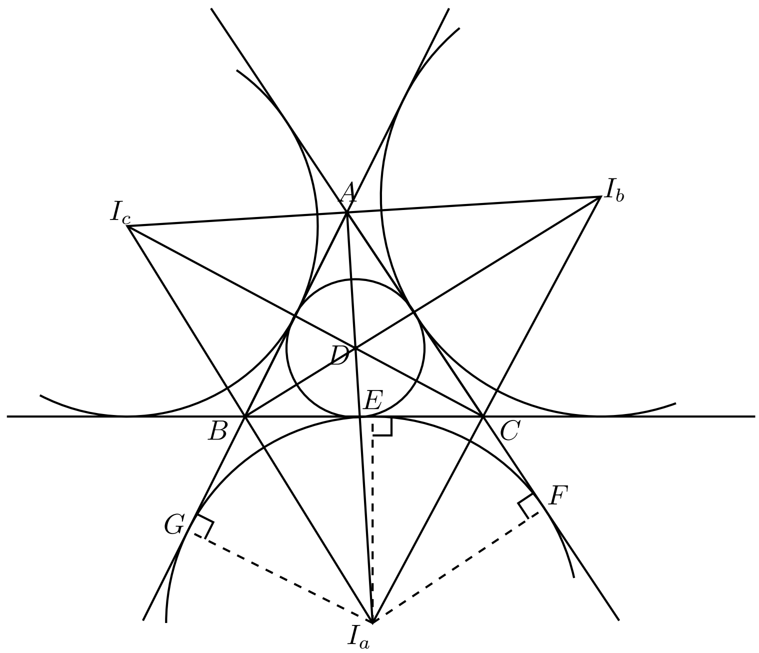


图 25.12: 旁切圆与内切圆

如图，$\triangle ABC$ 内角平分线 AI_a, BI_b, CI_c 交于一点 D，I_a, I_b, I_c 分别是三个旁切圆的圆心，E, F, G 是旁切圆 O_a 与三边的相切点；通过图中不难看出，三角形的内切圆分别于三个旁切圆外切，下面来证明性质：

通过角平分线的性质易证，旁心到三条边的距离相等，例如图中角平分线 AI_a, BI_a 使得它们的交点到三角形三边距离相等，因此性质 1，3 得证；而性质 2 显然成立；

对于性质 4：如图，由于 $\angle AGI_a = \angle AFI_a = \frac{\pi}{2}, GI_a = FI_a, AI_a = AI_a$，因此有

$$\triangle AGI_a \sim \triangle AFI_a$$

则 $AF = AG$，同理可得

$$AG = AB + BE, AF = AC + CE$$

[④]小蓝本中叫旁边圆、日本那边也有叫旁接圆，但我查到的资料更多的是叫旁切圆，所以下面统一用“旁切圆”称呼。

因此有：

$$\begin{aligned} AG+AF &= AC+CE+AB+BE \\ &= AB+BC+CA \\ &= 2p \end{aligned}$$

即证 $AG=AF=p$。

□

§25.8 线面关系证明

直线与直线：
平行:$\exists\lambda\in\mathbb{R}:\overrightarrow{a}=\lambda\overrightarrow{b}$
垂直：$\overrightarrow{a}\cdot\overrightarrow{b}=0$
夹角：$\cos\theta=\left|\frac{\overrightarrow{a}\cdot\overrightarrow{b}}{|\overrightarrow{a}|\cdot|\overrightarrow{b}|}\right|$

直线与平面：设 $\overrightarrow{m}$ 是平面 α 的法向量
平行：$\overrightarrow{a}\cdot\overrightarrow{m}=0$
垂直：$\exists\lambda\in\mathbb{R}:\overrightarrow{a}=\lambda\overrightarrow{m}$
夹角：$\gamma=\frac{\pi}{2}-\theta$

平面与平面：设平面 π_1,π_2 的法向量分别为 $\overrightarrow{n_1},\overrightarrow{n_2}$
平行：$\exists\lambda\in\mathbb{R}:\overrightarrow{n_1}=\lambda\overrightarrow{n_2}$
垂直：$\overrightarrow{n_1}\cdot\overrightarrow{n_2}=0$
二面角：$\theta=\langle\overrightarrow{n_1},\overrightarrow{n_2}\rangle$, 二面角为 α, 有 $\theta=\pi-\alpha$

§25.9 补充

二面角是算其中比较复杂的，判断二面角的正负比较麻烦，通常只用通过观察来判断正负，但如果没有图则需要在两法向量与各自所垂直的平面的交点，连接这两交点，令其为一个向量 $\overrightarrow{a}$，如果两法向量与 $\overrightarrow{a}$ 的内积同号，则 α 与 θ 互补，如果内积异号则 α 与 θ 相等。具体查看A.4。

这算一个偏冷门的考点，一般来说都可以直接看图判断，很少用到这个。

空间中点线面距离

其实就研究 1，点到面；2，线到面；3，点到线；都可以转到点与线面的关系一，**点到面**

设 $M\in\pi,N\in\pi,M\neq N,P\notin\pi,\overrightarrow{PM}\perp\pi$ $\overrightarrow{n}\perp\pi$，已知 $P,N,\overrightarrow{n}$

PM 即为所求

$$\left|\overrightarrow{PM}\right|=\left|\cos\left\langle\overrightarrow{PN},\overrightarrow{n}\right\rangle\right|\cdot\left|\overrightarrow{PN}\right|=\frac{\left|\overrightarrow{n}\cdot\overrightarrow{PN}\right|}{|\overrightarrow{n}|}$$

主要是记思路。

二，**线到面**

找到线上一点转换成点到面的问题即可。

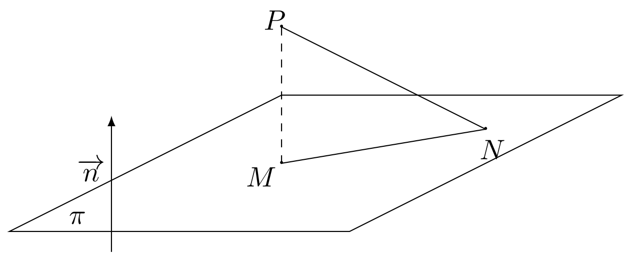


图 25.13: 点到面

三，**点到线** 设 $P \notin l, M \in l, N \in l, Q \in l, M \neq N \neq Q, \overrightarrow{PM} \perp l$，已知 P, N, Q，PM 即为所求

$$\left|\overrightarrow{PM}\right| = \sqrt{\left|\overrightarrow{PN}\right|^2 - \left(\cos\left\langle\overrightarrow{PN}, \overrightarrow{QN}\right\rangle \cdot \left|\overrightarrow{PN}\right|\right)^2}$$

教科书上用方向向量其实是一样的，这个更直观，其实就是勾股定理。

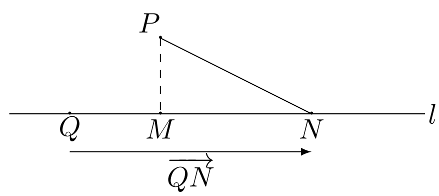


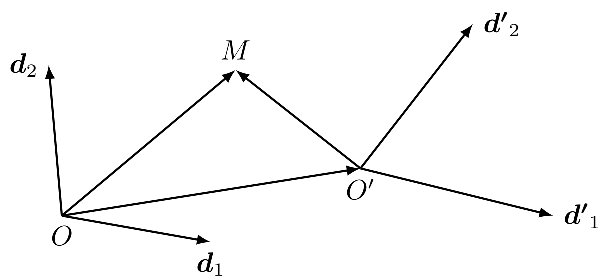
图 25.14: 点到直线

第二十六章　坐标变换

在研究实际问题时，常常有变换坐标系的需求，这样能使问题简单化。研究同一个点在两个坐标系中的坐标之间的关系，这样的关系式称为**坐标变换公式**。本章就来讨论坐标变换的问题。

§26.1　平面的仿射坐标变换

平面上有两仿射坐标系：$[O;\boldsymbol{d}_1,\boldsymbol{d}_2],[O';\boldsymbol{d}'_1,\boldsymbol{d}'_2]$，为方便，称前者为旧坐标系，记为 I；后者为新坐标系，记为 II。点 M 在 I 上的坐标称为旧坐标，在 II 上的坐标称为新坐标。为了研究同一个点 M 的 I 坐标和 II 坐标的关系，首先得确定 I 与 II 的相对位置。



如图，设 II 的原点 O' 的旧坐标为 $(x_0,y_0)^T$，II 的基向量 $\boldsymbol{d}'_1,\boldsymbol{d}'_2$ 的旧坐标分别是 $(a_{11},a_{21})^T,(a_{12},a_{22})^T$。现在我们来求点 M 的旧坐标 $(x,y)^T$ 与它的新坐标 $(x',y')^T$ 之间的关系，注意到

$$\overrightarrow{OM}=\overrightarrow{OO'}+\overrightarrow{O'M}=(x_0\boldsymbol{d}_1+y_0\boldsymbol{d}_2)+(x'\boldsymbol{d}'_1+y'\boldsymbol{d}'_2)$$

展开得到

$$\begin{aligned}\overrightarrow{OM}&=(x_0\boldsymbol{d}_1+y_0\boldsymbol{d}_2)+x'(a_{11}\boldsymbol{d}_1+a_{21}\boldsymbol{d}_2)+y'(a_{12}\boldsymbol{d}_1+a_{22}\boldsymbol{d}_2)\\&=(a_{11}x'+a_{12}y'+x_0)\boldsymbol{d}_1+(a_{21}x'+a_{22}y'+y_0)\boldsymbol{d}_2\end{aligned}$$

所以

$$\begin{cases}x&=a_{11}x'+a_{12}y'+x_0\\y&=a_{21}x'+a_{22}y'+y_0\end{cases}\qquad ①$$

上式被称为平面上坐标系 I 到 II 的**点的仿射坐标变换公式**。

注意到新坐标系的基向量不共线，因此 ① 式中的系数行列式不等于零。即

$$\begin{vmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\end{vmatrix}\neq 0$$

不妨记系数行列式为 D，① 式中的 x', y' 有唯一解

$$\begin{cases} x' &= \frac{1}{D}\left(a_{22}x - a_{12}y - a_{22}x_0 + a_{12}y_0\right) \\ y' &= \frac{1}{D}\left(-a_{21}x + a_{11}y + a_{21}x_0 - a_{11}y_0\right) \end{cases}$$

上式是把平面上任意一点 M 的新坐标 x', y' 表示成它旧坐标 x, y 的一次多项式，称上式为平面上坐标系 II 到 I 的**点的仿射坐标变换公式**。

§26.1.1 向量的仿射坐标变换公式

平面上一向量 $\boldsymbol{m}$ 的旧坐标为 $(u, v)^T$，新坐标为 $(u, v)^T$，不妨设 $\boldsymbol{m} = \overrightarrow{M_1M_2}$，其中

$$M_i\text{的旧坐标}(x_i, y_i),\quad M_i\text{的新坐标}(x_i', y_i'),\quad i = 1, 2$$

那么根据点的仿射坐标变换公式即得

$$\begin{cases} u &= x_2 - x_1 = a_{11}u' + a_{12}v' \\ v &= y_2 - y_1 = a_{21}u' + a_{22}v' \end{cases}$$

上式被称为平面上坐标系 I 到 II 的**向量的仿射坐标变换公式**，与 ① 式不同的是，上式没有常数项，这是因为新坐标系的原点位置并不影响向量的坐标。（向量平移的性质），同样的是上式的系数行列式不为零，因此可以反解出

$$\begin{cases} u' &= \frac{1}{D}(a_{22}u - a_{12}v) \\ v' &= \frac{1}{D}(-a_{21}u + a_{11}v) \end{cases}$$

这就得到了平面上坐标系 II 到 I 的**向量的仿射坐标变换公式**。

第二十七章　直线和圆的方程

在初中我们已经简单了解了直线与圆，这一节主要是以代数的视角来解释直线与圆。

§27.1　直线

§27.1.1　直线的定义

两个点可以确定一条直线，据此可以写出直线的解析式，下面通过引入一条直线来解释不同角度下直线的解析式

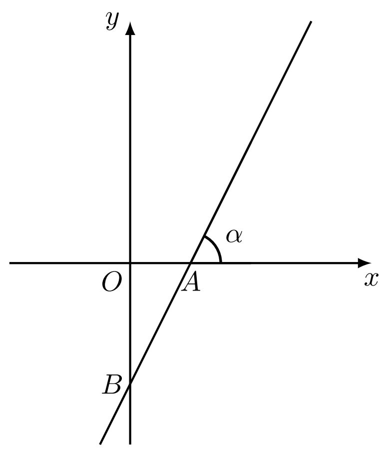


图 27.1: 直线的一些定义

如图，一直线 l 过两点 $A(x_1,y_1),B(x_2,y_2)$，这两点分别在 x 轴和 y 轴上，直线 l 与 x 轴正半轴的夹角记为 α，可以发现，直线也可以通过一个角和一个点确定，因此我们引入定义：

定义

倾斜角：直线 l 与 x 轴正方向上，向上的方向之间的角 $\alpha,(0^\circ \leqslant \alpha < 180^\circ)$

斜率：直线 l 的倾斜角 α 的正切值，用 k 表示；即

$$\tan\alpha = k$$

截距：直线与 x 轴的截距为 OA 的长度，直线与 y 轴的截距为 OB 的长度，但要注意截距有正负，因此截距不只是长度，例如上例直线 l 的 x 轴的截距就是正的，而 y 轴的截距就是负的。

当 α 不同取值时可以得到

$$\begin{cases} \alpha \in (0^\circ, 90^\circ) & k > 0 \\ \alpha = 90^\circ & k \text{ 不存在} \\ \alpha \in (90^\circ, 180^\circ) & k < 0 \end{cases}$$

如果 $A(x_1,y_1),B(x_2,y_2)$ 确定，则可以通过下式求斜率

$$k=\frac{y_2-y_1}{x_2-x_1}$$

对于两条直线 l_1,l_2，当 $l_1 /\!/ l_2$ 时，有 $k_1=k_2$，这是显然的。

当 $l_1 \perp l_2$ 时，有 $k_1\cdot k_2=-1$，这个证明也比较显然：

记直线 l_1,l_2 的倾斜角分别是 θ_1,θ_2，因为两直线垂直，所以

$$\begin{aligned}\theta_2+\pi-\theta_1&=\frac{\pi}{2}\\ \theta_2+\frac{\pi}{2}&=\theta_1\\ -\cot\theta_2&=\tan\theta_1\end{aligned}$$

因此

$$k_1\cdot k_2=-1$$

□

同一个点和斜率来确定直线的解析式，称为**点斜式**，由于斜率不变因此，对于直线 l 上的任意一点 $P(x,y)$ 都有

$$k=\frac{y-y_1}{x-x_1}=\frac{y_1-y_2}{x_1-x_2} \tag{1}$$

移项即得

$$y-y_1=k(x-x_1) \tag{2}$$

点 $A(x_1,y_1)$ 可以替换成任意属于直线 l 的点。

若将 ② 式展开并移项可得

$$y=kx-kx_1+y_1$$

若令 $x=0$，可得

$$y_0=y_1-kx_1$$

其中 y_0 是直线 y 轴上的截距，不妨记为 b，这时 ② 式可以改写为

$$y=kx+b$$

这就是常用的**斜截式**。

以上两种解析式都是建立在 k 存在的情况下，当 k 不存在时不适用，即直线垂直于 x 轴时不能使用。

若记 a,b 分别为直线 l 在 x 轴和 y 轴的截距，可知

$$k=-\frac{b}{a}$$

由斜截式得

$$
\begin{aligned}
ay &= -bx + ab \\
\frac{x}{a} + \frac{y}{b} &= 1
\end{aligned}
$$

上式被称为**截距式**，显然 $ab \neq 0$，该式不适用于与坐标轴垂直的直线或过原点的直线。

倘若我们把 ② 式（点斜式）的斜率展开，即

$$
\begin{aligned}
y - y_1 &= \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) \\
\frac{y - y_1}{y_2 - y_1} &= \frac{x - x_1}{x_2 - x_1}
\end{aligned}
$$

上式也被称为**两点式**，该式不适用于与坐标轴垂直的直线。

上面介绍了这么多种解析式，但每一种都有缺陷，那么有没有一种可以适用于所有直线的解析式呢？

通过定义，直线可以被两个点确定，在平面上求这个直线的解析式也就是解二元一次方程组，对于二元一次方程的一般形式可以写为

$$Ax + By + C = 0$$

其中 A, B 不都为 0，若直线 l 过点 $A(x_1, y_1), B(x_2, y_2)$，带入上式联立可解出 A, B, C。因此对于任意直线都可以用上式来表示，上式被称为**直线的一般式**

换为斜截式就是

$$y = -\frac{A}{B}x - \frac{C}{B}$$

即知斜率为 $k = -\frac{A}{B}$，y 轴上的截距为 $-\frac{C}{B}$。

对于两直线 $l_1 : A_1x + B_1y + C_1 = 0, l_2 : A_2x + B_2y + C_2 = 0$，若 $l_1 \,/\!/\, l_2$，那么

$$A_1B_2 = A_2B_1$$

需要注意两条直线不能重合。

若 $l_1 \perp l_2$，那么

$$A_1A_2 = B_1B_2$$

这样，我们就将之前的结论推广到了对于任意直线都成立的情况。

§27.1.2 直线的一些结论

定理

平面上的两点 $A(x_1, y_1), B(x_2, y_2)$，它们的距离为 $|AB|$，可以表示为

$$|AB| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

上式被称为**两点间距离公式**（平面上的），通过勾股定理即证。 ♠

若点 $P(x_0, y_0)$ 不在直线 l 上，那么它到直线的距离 d 可以表示为

$$d = \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}}$$

证明：

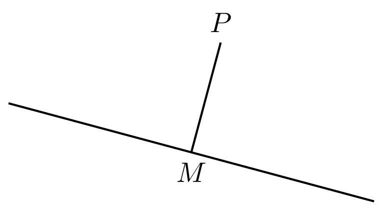


图 27.2: 点到直线的距离

设直线 $l : Ax + By + C = 0$ 以及线外一点 $P(x_0, y_0)$，记 PM 垂直 l，垂足为点 $M(x_1, y_1)$，点 P 到直线的距离即为 $d = |PM|$，易知 l_{PM} 的方程为

$$Ay - Ay_0 = Bx - Bx_0$$

将两直线方程联立可得

$$\begin{cases} x_1 = & \frac{B^2x_0 - ABy_0 - AC}{A^2+B^2} \\ y_1 = & \frac{A^2y_0 - ABx_0 - BC}{A^2+B^2} \end{cases}$$

根据两点间距离公式可得

$$\begin{aligned} d = |PM| &= \sqrt{\left(\frac{B^2x_0 - ABy_0 - AC}{A^2 + B^2} - x_0\right)^2 + \left(\frac{A^2y_0 - ABx_0 - BC}{A^2 + B^2} - y_0\right)^2} \\ &= \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}} \end{aligned}$$

□

推论

若两直线 l_1, l_2 平行，即 $l_1 /\!/ l_2$，则两直线距离为

$$d = \frac{|C_1 - C_2|}{\sqrt{A^2 + B^2}}$$

上式称为**平行线间的距离公式**（平面上的），上式的参数具体会在下面推到中说明。 ◇

证明：由于 $l_1 /\!/ l_2$，即两直线斜率相同，不妨设

$$\begin{cases} l_1 : Ax + By + C_1 = 0 \\ l_2 : Ax + By + C_2 = 0 \end{cases}$$

直线 l_1 上有一点 $P(x_0,y_0)$，它到直线 l_2 的距离即是两平行线间的距离，故

$$d=\frac{|Ax_0+By_0+C_2|}{\sqrt{A^2+B^2}}$$

又因点 P 在 l_1 上，则有

$$Ax_0+By_0+C_1=0$$

综上

$$d=\frac{|C_1-C_2|}{\sqrt{A^2+B^2}}$$

□

设 $l: Ax+By+C=0$，则 $\overrightarrow{a}=(A,B)$，有 $\overrightarrow{a}\perp l$。（在空间解析几何中也有类似结论，设平面 $\pi=Ax+By+Cz+D=0$，则 $\overrightarrow{a}(A,B,C)^T$ 是平面 π 的一个法向量）

§27.1.3 倒角公式

定理

倒角公式：

$$\tan\theta=\frac{k_2-k_1}{1+k_2k_1}$$

或

$$\tan\theta=\frac{k_1-k_2}{1+k_2k_1}$$

其中 θ 是直线 l_1,l_2 的其中一个夹角，这两直线的斜率分别为 k_1,k_2。

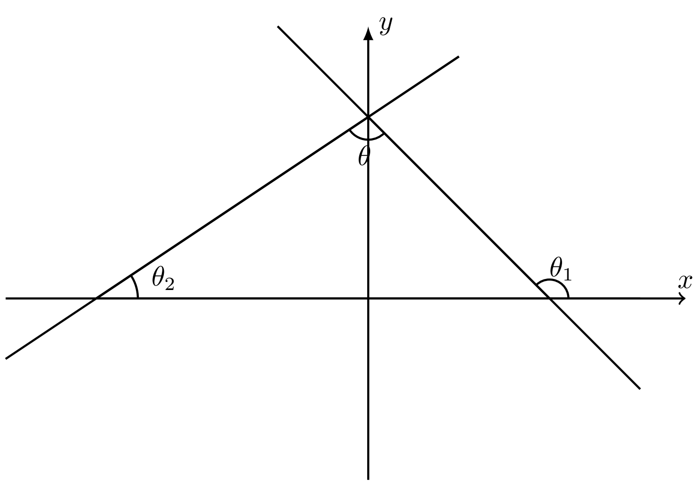


图 27.3: 倒角公式

证明：如图直线 l_1,l_2 的倾斜角分别是 θ_1,θ_2，两直线夹角（其中一个）为 θ；

通过观察，直线 l_1,l_2 与 x 轴构成一个三角形，不难知道 $\theta=\theta_2-\theta_1$，则有：

$$\begin{aligned}\tan\theta &= \tan(\theta_2-\theta_1) \quad (27.1)\\ &= \frac{\tan\theta_2-\tan\theta_1}{1+\tan\theta_2\tan\theta_1} \quad (27.2)\\ &= \frac{k_2-k_1}{1+k_2k_1} \quad (27.3)\end{aligned}$$

□

注意到两直线的另一个夹角是 θ 的补角，设为 θ_3，有

$$\tan\theta_3=\tan(-\theta)=\frac{k_1-k_2}{1+k_2k_1}$$

这个定理联系斜率的几何意义，沟通了三角函数中正切的差角公式，主要是数形结合，不用死记硬背。

§27.2 圆

圆可以看作定点到定长的点组成的集合，因此由定点 O，和定长 r 就可以得到圆的方程。

不妨设 $O(a,b)$，对于圆上任意一点 $P(x,y)$ 都有 $|OP|=r$，根据两点间距离公式得

$$(x-a)^2+(y-b)^2=r^2$$

上式被称为**圆的标准方程**。

如果将圆的标准方程展开可得：

$$(x-a)^2+(y-b)^2=r^2$$
$$x^2+y^2-2ax-2by+a^2+b^2-r^2=0$$

记 $D=-2a,E=-2b,F=a^2+b^2-r^2$ 则有

$$x^2+y^2+Dx+Ey+F=0$$

上式被称为**圆的一般方程**，对它配方可得

$$(x+\frac{D}{2})^2+(y+\frac{E}{2})^2=\frac{D^2+E^2-4F}{4}$$

观察可知，当 $D^2+E^2-4F>0$ 则该图像是一个圆，$D^2+E^2-4F=0$ 则是一个点，$D^2+E^2-4F<0$ 则是一个虚圆。

§27.3 阿氏圆

圆的第二定义

已知平面上两点 A,B，则所有满足

$$\frac{|PA|}{|PB|}=\lambda,\ \lambda\neq 1$$

的 P 的轨迹是一个以定比 λ 内分和外分定线段 AB 的两个分点的连线为直径的圆，这个轨迹最早由阿波罗尼斯发现，因此称为**阿波罗尼斯圆**或**阿氏圆**。

证明：设点 $P(x,y),A(x_1,y_1),B(x_2,y_2)$，由定义知有

$$\begin{aligned}\sqrt{(x-x_1)^2+(y-y_1)^2}&=\lambda\sqrt{(x-x_2)^2+(y-y_2)^2}\\ x^2+x_1^2-2xx_1+y^2+y_1^2-2yy_1&=\lambda^2(x^2+x_2^2-2xx_2+y^2+y_2^2-2yy_2)\\ x_1^2-\lambda^2x_2^2+y_1^2-\lambda^2y_2^2&=(\lambda^2-1)(x^2+y^2)+2(x_1-\lambda^2x_2)x+2(y_1-\lambda^2y_2)y\end{aligned}$$

令

$$\begin{aligned}D&=\frac{2(x_1-\lambda^2x_2)}{\lambda^2-1}\\ E&=\frac{2(y_1-\lambda^2y_2)}{\lambda^2-1}\\ F&=\frac{\lambda^2(x_2^2+y_2^2)-(x_1^2+y_1^2)}{\lambda^2-1}\end{aligned}$$

则式子可化为：

$$x^2+y^2+Dx+Ey+F=0$$

易知 P 的轨迹是圆心为

$$\left(\frac{(x_1-\lambda^2x_2)}{1-\lambda^2},\frac{2(y_1-\lambda^2y_2)}{1-\lambda^2}\right)$$

半径为

$$r=\frac{\lambda\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}}{|\lambda^2-1|} \tag{1}$$

的圆。

现在我们用几何直观的来分析，因为关于 AB 对称的两点 P,P' 有

$$\frac{|PA|}{|PB|}=\lambda,\frac{|P'A|}{|P'B|}=\lambda$$

圆 O 关于 AB 对称，因此圆心 O 在 AB 这条直线上，圆 O 与直线 AB 交于两点 C,D，易知 $|CD|$ 即为直径，且 C,D 都满足

$$\frac{|CA|}{|CB|}=\frac{|DA|}{|DB|}=\lambda$$

因此我们知道 C,D 一个是 AB 的内分点，另一个是外分点，不妨设 C 为内分点 D 为外分点；如图：

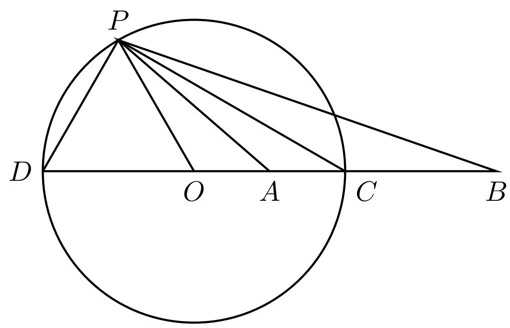


图 27.4: 阿氏圆

对于 P 不在直线 AB 上时，因为有：

$$\frac{|CA|}{|CB|} = \frac{|DA|}{|DB|} = \frac{|PA|}{|PB|} = \lambda$$

因此根据角平分线定理有：

$$\angle APC = \angle CPB,\ \pi - \angle APB = 2\angle APD$$

可以知道

$$\angle DPC = \angle DPA + \angle APC = \frac{\pi}{2}$$

即证对于 P 不在直线 AB 上时，它的轨迹是一个圆，且直径为 $|CD|$。

注意到任意

$$\frac{a}{b} = \frac{c}{d}, b \neq d$$

都有

$$\frac{a}{b} = \frac{c}{d} = \frac{a+c}{b+d} = \frac{a-c}{b-d}$$

可以用相似或代数理解上式，因此有

$$\begin{aligned}\frac{|CA| + |DA|}{|CB| + |DB|} &= \frac{|CA| - |DA|}{|CB| - |DB|}\\ \frac{2r}{2|OB|} &= \frac{-2|OA|}{-2r}\\ \frac{|OP|}{|OB|} &= \frac{|OA|}{|OP|}\end{aligned}$$

又 $\angle POA = \angle BOA$，因此 $\triangle POA \sim \triangle BOA$。

并且它们的比值为 λ，即

$$\frac{r}{|OB|} = \frac{|OA|}{r} = \lambda$$

注意到 $|AB| = ||OB| - |OA||$，联立即可得到

$$r = \frac{\lambda|AB|}{|\lambda^2 - 1|}$$

这与我们用代数推导出来的 ① 式结果一样，综上：

$$\begin{cases} r &= \frac{\lambda|AB|}{|\lambda^2-1|} \\ |OA| &= \lambda r = \frac{\lambda^2|AB|}{|\lambda^2-1|} \\ |OB| &= \frac{r}{\lambda} = \frac{|AB|}{|\lambda^2-1|} \end{cases}$$

§27.4 位置关系

§27.4.1 直线与圆的位置关系

直线与圆的位置关系分为：相离，相切，相交。

可以通过圆心与直线的距离 d，和圆的半径 r 来判断他们的位置关系。

$$\begin{cases} r < d & \text{相离} \\ r = d & \text{相切} \\ r < d & \text{相交} \end{cases}$$

也可以通过联立它们两的方程，消去一个未知数，得到一个一元二次方程，判断它的零点个数，它的零点个数就表示直线与圆的交点个数。

当圆与直线相交，交点 A, B，弦长为 $|AB| = 2\sqrt{r^2 - d^2}$。也是用勾股定理推的。

§27.4.2 圆与圆的位置关系

圆与圆关系分为：外离，相切，相交，内切，内含。

主要是通过两个圆的半径与两圆心距离 d 的关系来判断的

$$\begin{cases} d > r_1 + r_2 & \text{外离} \\ d = r_1 + r_2 & \text{外切} \\ |r_1 - r_2| < d < r_1 + r_2 & \text{相交} \\ d = |r_1 - r_2| & \text{内切} \\ d < |r_1 - r_2| & \text{内含} \end{cases}$$

可以配合图像理解

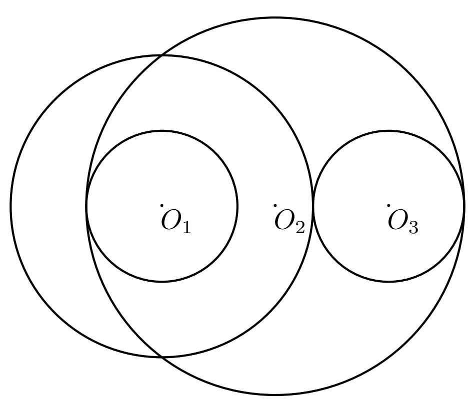


图 27.5: 圆与圆的位置关系

如图，O_1, O_2, O_3 共线，圆 C_1, C_2, C_3, C_4 的半径分别为 2、1、2.5、1，圆心分别为 O_1, O_1, O_2, O_3，其中 C_1, C_2 是同心圆，位置关系为内含，C_1, C_3 相交，C_1, C_4 外切，C_2, C_3 内切，C_2, C_4 外离。

两圆相交时，公共弦所在的直线的方程可以用两圆的方程作差获得。

§27.5 补充

米勒定理

第二十八章　圆锥曲线

这一章介绍圆锥曲线的基本性质，以及高考中常用到的方法；同时这一章是比较具象、直观的，会由最朴素的思想开始入手研究圆锥曲线，初步引入解析几何的思想。

§28.1　椭圆

§28.1.1　第一定义

椭圆

第一定义：平面内到两个定点 F_1, F_2 的距离的和等于常数 (大于 $|F_1F_2|$) 的点的轨迹叫做**椭圆**。
这两个定点叫做椭圆的焦点，$|F_1F_2| = 2c$ 叫做焦距，其中 c 叫做半焦距。
若点 M 在这个椭圆上，则满足 $|MF_1| + |MF_2| = 2a(2a > |F_1F_2|)$。

在高中，我们经常研究以下两种椭圆，它们的焦点都在坐标轴上，且关于原点对称。

椭圆的标准方程 (1)：对于焦点在 x 轴上，分别是 $F_1(-c, 0), F_2(c, 0)$ 的椭圆，在椭圆上的点的轨迹方程是

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \ (a > b > 0)$$

其中 $a^2 = b^2 + c^2$。

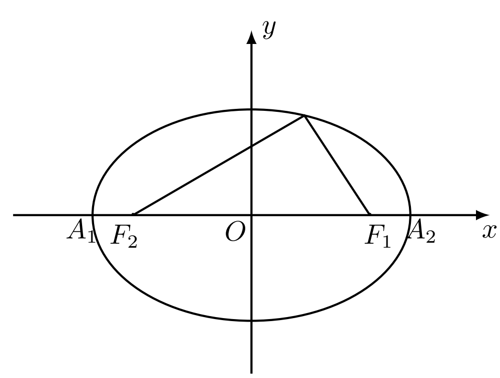


图 28.1: 交点在 x 轴上的椭圆

证明：设 $M(x, y)$ 在椭圆 C 上，C 的焦点分别为 $F_1(-c, 0), F_2(c, 0)$，根据定义有

$$|MF_1| + |MF_2| = 2a$$

因此，可以写出：

$$\sqrt{(x+c)^2+y^2}+\sqrt{(x-c)^2+y^2}=2a$$
$$(x+c)^2+y^2=4a^2+(x-c)^2+y^2-4a\sqrt{(x-c)^2+y^2}$$
$$4xc-4a^2=-4a\sqrt{(x-c)^2+y^2}$$
$$x^2c^2+a^4-2a^2xc=a^2(x-c)^2+a^2y^2$$
$$(a^2-c^2)x^2+a^2y^2=a^2(a^2-c^2)$$
$$\frac{x^2}{a^2}+\frac{y^2}{a^2-c^2}=1$$

若令 $a^2-c^2=b^2$，即可得到椭圆的标准方程，显然有 $a^2>b^2$ 即 $a>b$。

□

a,b 的几何意义：a 是椭圆的半长轴，b 是椭圆的半短轴，通过图像，这一性质是显然的。

椭圆的标准方程 (1)：对于焦点在 y 轴上，分别是 $F_1(0,c),F_2(0,-c)$ 的椭圆，在椭圆上的点的轨迹方程是

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1\ (a>b>0)$$

其中 $a^2=b^2+c^2$（这保证了 a,b 的几何意义）。
与证明焦点在 x 轴上的椭圆标准方程类似的步骤，可以证明在 y 轴上的情况，这里不再赘述。

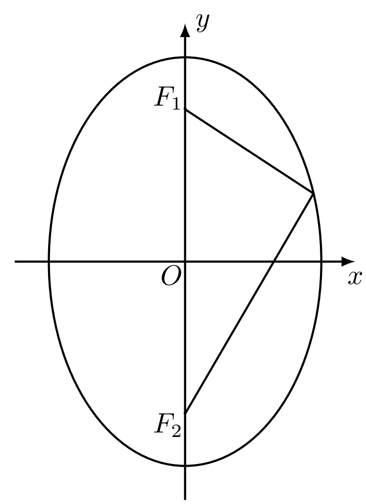


图 28.2: 交点在 y 轴上的椭圆

§28.1.2 第二定义

椭圆

第二定义：到一定点 F_1 的距离与到一定直线 l 的距离之比为定值（小于 1）的点所形成的轨迹称为**椭圆**，该定直线 l 称为**准线**。

由第一定义，不难知道

$$\sqrt{(x+c)^2+y^2}+\sqrt{(x-c)^2+y^2}=2a \quad ①$$

写出它的共轭根式

$$\sqrt{(x+c)^2+y^2}-\sqrt{(x-c)^2+y^2}=K \quad ②$$

注意到 ①×②

$$\begin{aligned}(x+c)^2+y^2-(x-c)^2-y^2&=2aK\\ K&=\frac{2xc}{a}\end{aligned}$$

再将两式相加 ①+②

$$\sqrt{(x+c)^2+y^2}=a+\frac{xc}{a}$$

注意到左边是非负的，且它的几何意义是点 (x,y) 到焦点 $F(-c,0)$ 的距离，而右边也应是非负的，即

$$\begin{aligned}\sqrt{(x+c)^2+y^2}&=\frac{c}{a}\left|\frac{a^2}{c}+x\right|\\ \frac{\sqrt{(x+c)^2+y^2}}{\left|\frac{a^2}{c}+x\right|}&=\frac{c}{a}\end{aligned}$$

这里，$|\frac{a^2}{c}+x|$ 可以看作点 (x,y) 到直线 $l:x=-\frac{a^2}{c}$ 的距离，综上我们证明了第二定义，其中 F_1 为左焦点，l 称为左准线，类似的有右焦点以及右准线，可证明：

$$\frac{\sqrt{(x-c)^2+y^2}}{\left|x-\frac{a^2}{c}\right|}=\frac{c}{a}$$

到定点 F_1 的距离与到一定直线 l 的距离之比的这个定值为 $\frac{c}{a}$，这个定值称为**离心率**，用 e 表示，有

$$e=\frac{c}{a}=\sqrt{\frac{c^2}{a^2}}=\sqrt{\frac{a^2-b^2}{a^2}}=\sqrt{1-\frac{b^2}{a^2}}$$

可以看出 $e\in(0,1),c^2=a^2-b^2$，当椭圆的离心率越接近 1 时，椭圆越“扁”。

椭圆简单的几何性质是明显的，有两个对称轴，它们相交于椭圆的对称中心，对于焦点在 x 轴上的椭圆，上顶点 $(0,b)$、下顶点 $(0,-b)$、左顶点 $(-a,0)$、右顶点 $(a,0)$，通过顶点可以得到 x,y 的取值范围。

由第二定义，可以直观的得出，若点 $A(x_0,y_0)$ 在椭圆 $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$ 上，焦点为 $F_1(-c,0),F_2(c,0)$ 则它到两焦点的距离为：

$$\begin{aligned}|AF_1|&=a+ex_0\\ |AF_2|&=a-ex_0\end{aligned}$$

显然 $|AF_1|+|AF_2|=2a$ 是符合第一定义的。

§28.1.3 第三定义

椭圆

第三定义：动点 P 到两定点 A,B 的斜率的乘积为定值（大于 -1 且小于 0），则动点轨迹为**椭圆**[a]。

$$k_{AP}\cdot k_{BP}=-\frac{b^2}{a^2}=e^2-1$$

[a]这里讨论焦点在 x 轴上的情况。

推论

第三定义的推论：椭圆上关于椭圆中心对称的两点（设为 A,B）与异于这两点且在椭圆上的点（设为 P）的斜率的乘积为定值，即

$$k_{AP}\cdot k_{BP}=-\frac{b^2}{a^2}$$

证明：

由题设 $A(x_0,y_0),B(-x_0,-y_0),P(x_1,y_1)$，椭圆 $C:\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$。

$$\because A,P\text{ 在椭圆 }C\text{ 上}$$
$$\therefore\begin{cases}\frac{x_0^2}{a^2}+\frac{y_0^2}{b^2}=1\\ \frac{x_1^2}{a^2}+\frac{y_1^2}{b^2}=1\end{cases}$$

两式相减得到：

$$\begin{aligned}\frac{x_0^2-x_1^2}{a^2}+\frac{y_0^2-y_1^2}{b^2}&=0\\ \frac{(y_0-y_1)(y_0+y_1)}{b^2}&=-\frac{(x_0-x_1)(x_0+x_1)}{a^2}\\ \frac{y_0-y_1}{x_0-x_1}\cdot\frac{y_0+y_1}{x_0+x_1}&=-\frac{b^2}{a^2}\end{aligned}$$

注意到 $k_{AP}=\frac{y_0-y_1}{x_0-x_1},k_{BP}=\frac{y_1+y_0}{x_1+x_0}$，带入方程

$$k_{AP}\cdot k_{BP}=-\frac{b^2}{a^2}$$

□

在考试中，如果在大题用到椭圆的第三定义及其推论，需要对其进行证明，显然第三定义是此推论的特殊情况，即证第三定义。

推论

焦点在 x 轴上的椭圆 C 上，任意取互异两点 P,Q，它们的中点为 M，原点为 O，都有

$$k_{OM}\cdot k_{PQ}=-\frac{b^2}{a^2}$$

通过几何直观来证明。

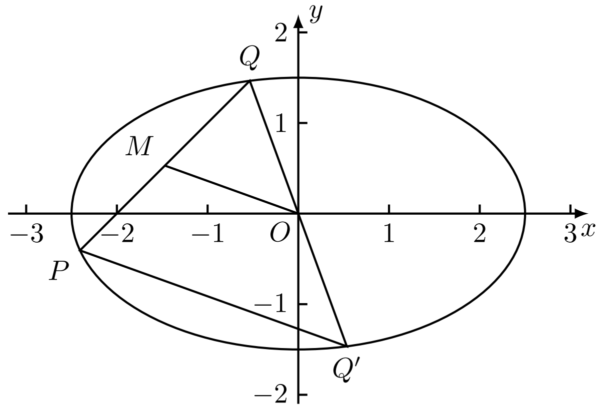


图 28.3: 椭圆中点弦

证明：取 Q 关于原点对称的点 Q'，由第三定义的推论可知

$$k_{PQ} \cdot k_{PQ'} = -\frac{b^2}{a^2}$$

由于 O 是线段 QQ' 的中点、M 是线段 PQ 的中点，所以在 $\triangle PQQ'$ 中，OM 是一条中位线，即 $k_{OM} \cdot k_{PQ'}$

$$k_{PQ} \cdot k_{OM} = -\frac{b^2}{a^2}$$

□

通过代数来证明：设 $Q(x_1, y_1), Q'(-x_1, -y_1), P(x_2, y_2)$，椭圆 $C : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$。

$$\because Q, P \text{ 在椭圆 } C \text{ 上}$$
$$\therefore \begin{cases} \frac{x_1^2}{a^2} + \frac{y_1^2}{b^2} = 1 \\ \frac{x_2^2}{a^2} + \frac{y_2^2}{b^2} = 1 \end{cases}$$

两式相减得到：

$$\begin{aligned} \frac{x_1^2 - x_2^2}{a^2} + \frac{y_1^2 - y_2^2}{b^2} &= 0 \\ \frac{(y_1 - y_2)(y_1 + y_2)}{b^2} &= -\frac{(x_1 - x_2)(x_1 + x_2)}{a^2} \\ \frac{y_1 - y_2}{x_1 - x_2} \cdot \frac{y_2 + y_1}{x_1 + x_2} &= -\frac{b^2}{a^2} \end{aligned}$$

注意到 $k_{PQ} = \frac{y_1 - y_2}{x_1 - x_2}, k_{OM} = \frac{y_2 + y_1}{x_1 + x_2}$，即证

$$k_{PQ} \cdot k_{OM} = -\frac{b^2}{a^2}$$

□

这个证明方法和证明第三定义推论的过程如出一辙，这个方法称为**点差法**。

如图，垂直于椭圆长轴的焦点弦叫做椭圆的**通径**，即图中 $|PQ|$，通过椭圆的第二定义可以较快求得通径的长度：

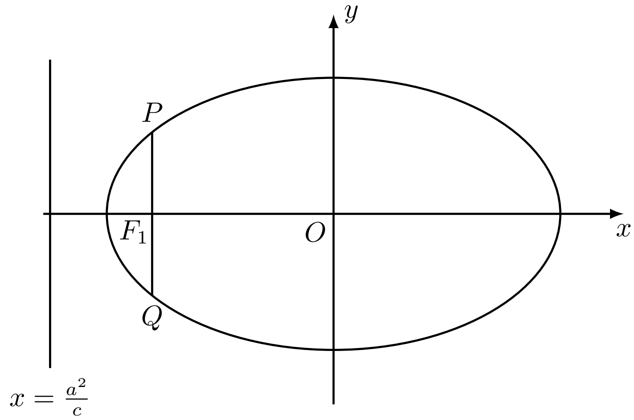


图 28.4: 椭圆通径

$$
\begin{aligned}
|PQ| &= |PF_1| + |QF_1| \\
&= 2a + 2ex_0 \\
&= 2\frac{a^2-c^2}{a} \\
&= 2\frac{b^2}{a}
\end{aligned}
$$

当然，通过解三角形以及第一定义，可以得到同样的结论。

§28.1.4　椭圆与直线的关系

研究圆锥曲线与直线的关系，通常是将它们联立，通过联立后的一元二次方程来研究它们间的关系；

$$
\begin{cases} C: & \frac{x^2}{a^2}+\frac{y^2}{b^2}=1, \quad a>b>0 \\ l: & y=kx+m \end{cases}
$$

联立可得

$$(a^2k^2+b^2)x^2+2kma^2x+a^2m^2-a^2b^2=0$$

这时可以得到判别式

$$
\begin{aligned}
\varDelta &= 4k^2m^2a^4-4(a^2m^2-a^2b^2)(a^2k^2+b^2) \\
&= 4a^2b^2(b^2+a^2k^2-m^2)
\end{aligned}
$$

注意到 $4a^2b^2>0$，因此只需判断 $b^2+a^2k^2-m^2$ 的大小即可判直线与椭圆的交点个数，因此可以得到下表：

在求离心率的问题中，常常会用到齐次式，以下是一些经常用到的齐次思路：

焦点三角形背景下求离心率：

$$e=\frac{c}{a}=\frac{2c}{2a}=\frac{|F_1F_2|}{|PF_1|+|PF_2|}=\frac{\sin\angle F_1PF_2}{\sin\angle PF_1F_2+\sin\angle PF_2F_1}$$

判别式	位置关系	交点数
$b^2+a^2k^2-m^2>0$	相交	2
$b^2+a^2k^2-m^2=0$	相切	1
$b^2+a^2k^2-m^2<0$	相离	0

表 28.1: 直线与椭圆的关系

§28.1.5 椭圆的焦点三角形

§28.2 双曲线

§28.2.1 第一定义

双曲线

第一定义：把平面内两个定点 F_1,F_2 的距离之差的**绝对值**为非零常数 (小于 $|F_1F_2|$) 的点的轨迹叫做**双曲线**。

设这个动点为 P，有

$$||PF_1|-|PF_2||=2a,\ 2a<|F_1F_2|$$

这两个定点叫做双曲线的焦点，焦点间的距离叫做双曲线的焦距（$|F_1F_2|=2c$），故在双曲线中 $c>a$。

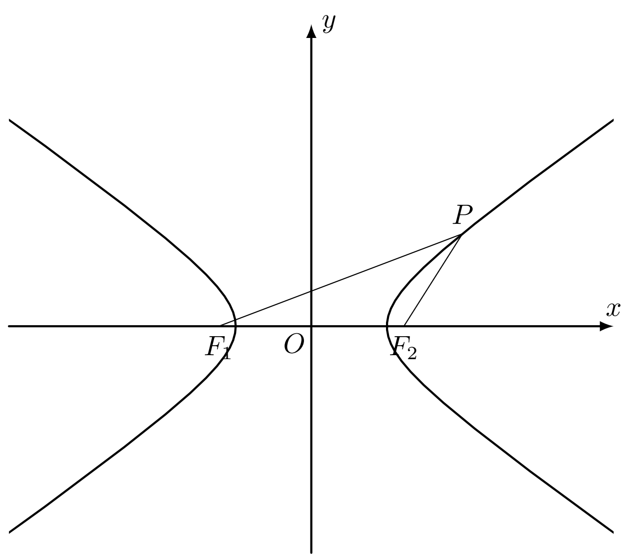


图 28.5: 双曲线

在高中，主要研究以下两种双曲线，它们的焦点都在坐标轴上，且关于原点对称。

双曲线的标准方程 (1)：

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1(a>0,b>0)$$

它表示焦点在 x 轴上，分别是 $F_1(-c,0),F_2(c,0)$ 的双曲线，这里 $c^2=a^2+b^2$。

证明：设 $F_1(-c,0),F_2(c,0),P(x,y)$，$P$ 在双曲线 C 上，根据定义有

$$\left|\sqrt{(x+c)^2+y^2}-\sqrt{(x-c)^2+y^2}\right|=2a$$

去掉绝对值后，右边可能为 $2a$ 或 $-2a$，我们先讨论右边为 $2a$ 的情况，即

$$\sqrt{(x+c)^2+y^2}-\sqrt{(x-c)^2+y^2}=2a \quad ①$$

它的共轭根式为

$$\sqrt{(x+c)^2+y^2}+\sqrt{(x-c)^2+y^2}=K \quad ②$$

易知 ①×② 为

$$\begin{aligned}(x+c)^2+y^2-(x-c)^2-y^2&=2aK\\4xc&=2aK\\K&=2\frac{xc}{a}\end{aligned}$$

又 ①+②

$$\begin{aligned}\sqrt{(x+c)^2+y^2}&=a+\frac{xc}{a} \quad ③\\x^2+c^2+2xc+y^2&=a^2+\frac{x^2c^2}{a^2}+2xc\\x^2\frac{a^2-c^2}{a^2}+y^2&=a^2-c^2\\\frac{x^2}{a^2}-\frac{y^2}{c^2-a^2}&=1\end{aligned}$$

若令 $b^2=c^2-a^2$，即得

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1(a>0,b>0)$$

□

对于去绝对值后右边为 $-2a$ 的情况，最后得到的结果一样，故不证明。

双曲线的标准方程 (2)：

$$\frac{y^2}{a^2}-\frac{x^2}{b^2}=1(a>0,b>0)$$

它表示焦点在 y 轴上，焦点分别是 $F_1(0,-c),F_2(0,c)$ 的双曲线，这里 $c^2=a^2+b^2$。证明过程仿照上例。

§28.2.2 第二定义

双曲线

第二定义：到一定点 F_1 的距离与到一定直线 l_1 的距离之比为定值（大于 1）的点所形成的轨迹称为**双曲线**，该定直线 l 称为**准线**。

在证明双曲线的标准方程 (1) 时，注意到 ③ 式的几何意义，即

$$\frac{\sqrt{(x+c)^2+y^2}}{\left|x+\frac{a^2}{c}\right|}=\frac{c}{a}$$

同样有

$$\frac{\sqrt{(x-c)^2+y^2}}{\left|x-\frac{a^2}{c}\right|}=\frac{c}{a}$$

即证第二定义。以上两个式子说明了，双曲线与椭圆一样，两个焦点分别有两个准线。

同样的，定义比值 $\frac{c}{a}$ 为离心率，用 e 表示，因为 $c>a$，能推出 $e\in(1,+\infty)$。

由第二定义，可以直观的得出，若点 $P(x_0,y_0)$ 在双曲线 $C:\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$ 上，C 的左右焦点分别为 $F_1(-c,0),F_2(c,0)$，若点 P 在双曲线 C 的右支，有

$$\begin{aligned}|F_1P|&=x_0e+a\\|F_2P|&=x_0e-a\end{aligned}$$

若点 P 在双曲线 C 的左支，有

$$\begin{aligned}|F_1P|&=-x_0e-a\\|F_2P|&=a-x_0e\end{aligned}$$

显然，能验证 $||F_1P|-|F_2P||=2a$。

双曲线的通径：垂直于实轴的焦点弦叫做双曲线的通径，通过第二定义能快速求得通径长度，设通径交双曲线于 P,Q 两点，有：

$$\begin{aligned}|PQ|&=|PF_1|+|QF_1|\\&=2ec-2a\\&=2\frac{b^2}{a}\end{aligned}$$

关于 a,b 的几何意义，a 被称为双曲线的**实半轴长**，即是双曲线与坐标轴相交点与圆心的距离。b 被称为**虚半轴长**。

双曲线的渐近线：双曲线无限延伸，逐渐接近的一条直线。

双曲线 $C:\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$，的渐近线有两条，即

$$y=\pm\frac{b}{a}x$$

证明：下面的证明只证明其中一种，其余情况同理可证；设双曲线的渐近线为 $y=mx+n$，将双曲线显化得到

$$f(x)=y=\frac{b}{a}\sqrt{x^2-a^2}$$

显然有

$$\lim_{x\to\infty}\frac{f(x)}{x}=\frac{b}{a}\lim_{x\to\infty}\frac{\sqrt{x^2-a^2}}{x}=\frac{b}{a}=m$$

另一面，有

$$\begin{aligned}n=&\lim_{x\to\infty}[f(x)-mx]\\=&\lim_{x\to\infty}[\frac{b}{a}\sqrt{x^2-a^2}-\frac{b}{a}x]\\=&\frac{b}{a}\lim_{x\to\infty}\frac{-a^2}{\sqrt{x^2-a^2}+x^2}=0\end{aligned}$$

即证，双曲线的一条渐近线为

$$y=\frac{b}{a}x$$

□

双曲线的两个渐近线分别为：

$$y=\frac{b}{a}x,\quad y=-\frac{b}{a}x$$

可以写成

$$\frac{y}{b}-\frac{x}{a}=0 \qquad ①$$
$$\frac{y}{b}+\frac{x}{a}=0 \qquad ②$$

若两式相乘，即 ① × ②，得到

$$\frac{y^2}{b^2}-\frac{x^2}{a^2}=0$$

可以看出，双曲线的两条渐近线就是上式的两个分解，观察可知，上式与双曲线的标准方程相似，只是右边的 1 变成 0；所以可以用上面这个式子推导双曲线的渐近线方程。

§28.2.3 第三定义

双曲线

第三定义：平面内动点到两定点 $A_1(-a,0),A_2(a,0)$ 的斜率乘积为定值（大于 0），则动点轨迹为**双曲线**[a]。

$$k_{A_1P}\cdot k_{A_2P}=\frac{b^2}{a^2}=e^2-1$$

[a] 焦点在 x 轴的情况

♡

推论

第三定义的推论：双曲线上关于原点中心对称的两点（设为 A,B）与异于这两点且在双曲线上的点（设为 P）的斜率的乘积为定值，即

$$k_{AP}\cdot k_{BP}=-\frac{b^2}{a^2}$$

◇

证明（点差法）：由题设 $A(x_0, y_0), B(-x_0, -y_0), P(x_1, y_1)$，双曲线 $C: \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$。

$$\because A, P \text{ 在双曲线 } C \text{ 上}$$
$$\therefore \begin{cases} \frac{x_0^2}{a^2} - \frac{y_0^2}{b^2} = 1 \\ \frac{x_1^2}{a^2} - \frac{y_1^2}{b^2} = 1 \end{cases}$$

两式相减得到：

$$\frac{x_0^2 - x_1^2}{a^2} + \frac{y_1^2 - y_0^2}{b^2} = 0$$
$$\frac{(y_1 - y_0)(y_1 + y_0)}{b^2} = -\frac{(x_0 - x_1)(x_0 + x_1)}{a^2}$$
$$\frac{y_1 - y_0}{x_1 - x_0} \cdot \frac{y_1 + y_0}{x_1 + x_0} = \frac{b^2}{a^2}$$

注意到 $k_{AP} = \frac{y_1 - y_0}{x_1 - x_0}, k_{BP} = \frac{y_1 + y_0}{x_1 + x_0}$，带入方程

$$k_{AP} \cdot k_{BP} = \frac{b^2}{a^2}$$

□

显然第三定义是推论的特殊情况，因此可以证明第三定义。

推论

焦点在 x 轴上的双曲线 C，任意取两点 P, Q，P, Q 中点为 M，原点为 O，有

$$k_{PQ} \cdot k_{OM} = \frac{b^2}{a^2}$$

◇

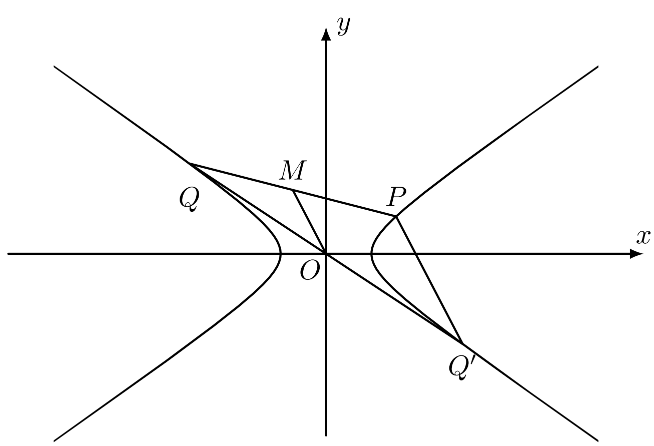


图 28.6: 双曲线中点弦

有第三定义推论结合三角形中位线的知识，这个推论是显然的，证明过程与椭圆中点弦推论证明类似，故不再次证明。

下面介绍一些特殊的双曲线

等轴双曲线: 实轴和虚轴等长的双曲线叫做等轴双曲线,即 $a=b$,标准方程为: $x^2-y^2=a^2$ 或者 $y^2-x^2=a^2$,可以推出 $e=\sqrt{2}$,渐近线为 $y=\pm x$。

共轭双曲线: 设双曲线 $C:\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$,则它的共轭双曲线为

$$C':\frac{y^2}{a^2}-\frac{x^2}{b^2}=1$$

易知,它们有共同的渐近线。

§28.2.4 双曲线与直线的关系

与研究椭圆一样联立直线与双曲线的方程:

$$\begin{cases} C: & \frac{x^2}{a^2}-\frac{y^2}{b^2}=1, \quad a,b>0 \\ l: & y=kx+m \end{cases}$$

可以得到

$$\begin{aligned} \frac{x^2}{a^2}-\frac{(kx+m)^2}{b^2}&=1 \\ (b^2-a^2k^2)x^2-2kma^2x-a^2m^2-a^2b^2&=0 \end{aligned} \quad ①$$

注意到二次项系数 $b^2-a^2k^2$ 有可能为 0,因此我们分类讨论

当 $b^2-a^2k^2=0$ 时,① 为一个一次函数,可以根据一次项系数判断直线与双曲线的关系。

1. 当一次项系数 $-2kma^2=0$ 时,直线与双曲线无交点;

2. 当一次项系数 $-2kma^2\neq 0$ 时,直线与双曲线有且仅有一个交点。

实际上 $k,a\neq 0$,因此只需判断 m 即可,即 $m=0$ 时,直线与双曲线无交点;$m\neq 0$ 时有一个交点。

注意到 $b^2-a^2k^2=0$ 时,直线斜率 k 与双曲线的一条渐近线斜率相同,当 $m=0$ 时

$$l:y=\pm\frac{a}{b}x$$

就是其中一条渐近线,显然无交点。

当 $b^2-a^2k^2\neq 0$ 时,我们可以得到 ① 的判别式为

$$\Delta=4a^2b^2(m^2+b^2-a^2k^2)$$

注意到 $4a^2b^2$ 恒大于 0,因此只需判断 $m^2+b^2-a^2k^2$ 即可,这里与椭圆的判断过程一样,不在赘述;直接给出下边表:

<table>
<tr><td>斜率</td><td>判别式</td><td>位置关系</td><td>交点数</td></tr>
<tr><td rowspan="2">$k^2 = \frac{b^2}{a^2}$</td><td>$m \neq 0$</td><td>相交</td><td>1</td></tr>
<tr><td>$m = 0$</td><td>渐近线</td><td>0</td></tr>
<tr><td rowspan="3">$k^2 \neq \frac{b^2}{a^2}$</td><td>$m^2 + b^2 - a^2k^2 > 0$</td><td>相交</td><td>2</td></tr>
<tr><td>$m^2 + b^2 - a^2k^2 = 0$</td><td>相切</td><td>1</td></tr>
<tr><td>$m^2 + b^2 - a^2k^2 < 0$</td><td>相离</td><td>0</td></tr>
</table>

表 28.2: 直线与双曲线的关系

§28.3 抛物线

抛物线

第一定义：把平面内与一个定点 F 和一个条定直线 l (l 不经过点 F) 距离相等的点的轨迹叫做抛物线，点 F 叫做抛物线的焦点，直线 l 叫做抛物线的准线。

当抛物线焦点在 x 轴上，准线垂直于 x 轴时，其标准方程：

$$y^2 = 2px(p > 0), \text{ 焦点 } F(\tfrac{p}{2}, 0), \text{准线方程} x = -\frac{p}{2}$$

$$y^2 = -2px(p > 0), \text{ 焦点 } F(-\tfrac{p}{2}, 0), \text{准线方程} x = \frac{p}{2}$$

当抛物线焦点在 y 轴上，准线垂直于 y 轴时，其标准方程：

$$x^2 = 2px(p > 0), \text{ 焦点 } F(0, \tfrac{p}{2}), \text{准线方程} x = -\frac{p}{2}$$

$$x^2 = -2px(p > 0), \text{ 焦点 } F(0, -\tfrac{p}{2}), \text{准线方程} x = \frac{p}{2}$$

其中 p 被称为**焦准距**。

对于抛物线开口方向判断，简单来说只要记得我们最熟悉的 $y = x^2$ 开口向上，$y = -x^2$ 开口向下，它们的反函数与它们的函数图像关于 $y = x$ 对称，就可以判断其他抛物线开口方向了。

§28.3.1 抛物线焦点弦

过抛物线焦点的直线交抛物线于 A, B 两点，这时线段 AB 被称为**焦点弦**。

如图，AB 为抛物线的焦点弦，分别过 A, B 作垂直于 y 轴的直线，分别交抛物线的准线为 A', B'，设 $A(x_a, y_a), B(x_b, y_b)$。

关于焦点弦 AB 有如下性质：

1. $|AB| = x_a + x_b + p$；
2. $\angle A'FB' = 90°$；
3. $|A'C| \cdot |B'C| = p^2$；
4. $\angle ACF = \angle BCF$

对于性质 1：

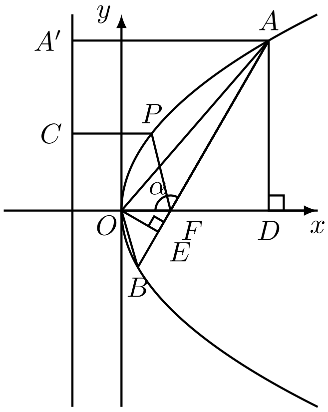


图 28.7: 抛物线角版

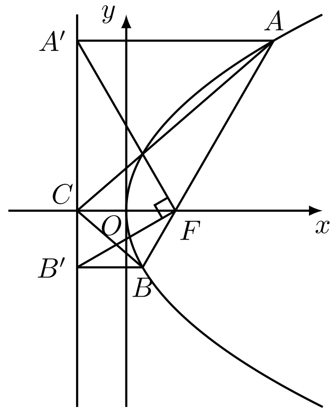


图 28.8: 抛物线焦点弦

引理

抛物线 $C: y^2 = 2px, p > 0$ 上一点 $P(x_0, y_0)$ 到焦点 F 的距离为 $x_0 + \frac{p}{2}$。

证明：由抛物线的性质可知 P 到 F 的距离等于 P 到准线 l 的距离，显然 P 到 l 的距离为 $x_0 + \frac{p}{2}$。

□

由引理可以直接得到性质 1。

推论

抛物线 $C: y^2 = 2px, p > 0$ 上，对于焦点弦 AB 其中 $A(x_a, y_a), B(x_b, y_b)$ 有：

$$|AB| = x_a + x_b + p$$

证明：由引理可知 $|AF| = x_a + \frac{p}{2}, |BF| = x_b + \frac{p}{2}$，因此

$$|AB| = |AF| + |BF| = x_a + x_b + p$$

□

对于性质 2，设 $\angle A'AF = \alpha$，由抛物线定义可知

$$|AF| = |AA'|,\ |BF| = |BB'|$$

即

$$\angle AFA' = \frac{\pi - \alpha}{2}$$
$$\angle BFB' = \frac{\pi - (\pi - \alpha)}{2} = \frac{\alpha}{2}$$

因此

$$
\begin{aligned}
\angle A'FB' &= \pi - \angle AFA' - \angle BFB' \\
&= \frac{2\pi - \pi - \alpha + \alpha}{2} = \frac{\pi}{2}
\end{aligned}
$$

□

对于性质 3，在 $\triangle A'FB'$ 中，由性质 $2\angle A'FB' = \angle A'CF = \frac{\pi}{2}$，根据射影定理（三角形相似）可知

$$
\begin{gathered}
\frac{|A'C|}{|CF|} = \frac{|CF|}{|B'C|} \\
|A'C| \cdot |B'C| = |CF|^2 = p^2
\end{gathered}
$$

□

对于性质 4，由于 $AA' \mathbin{/\!/} BB' \mathbin{/\!/} CF$，可知，若要证 $\angle ACF = \angle BCF$，即证 $\angle CAA' = \angle CBB'$，又因 $\angle CA'A = \angle CB'B = \frac{\pi}{2}$，所以只需证

$$\triangle AA'C \sim \triangle BB'C$$

即证明

$$\frac{|A'C|}{|AA'|} = \frac{|B'C|}{|BB'|}$$

由抛物线的定义可知 $|AA'| = |AF|, |BB'| = |BF|$，通过平行线的性质可知

$$\frac{|A'C|}{|AF|} = \frac{|B'C|}{|BF|}$$

显然成立，因此 $\angle ACF = \angle BCF$。

□

在抛物线角版的图中令 $\angle AFO = \alpha$，通过研究 α 得出有关焦点弦结论：

1. $|AF| = \frac{p}{1+\cos\alpha}, |BF| = \frac{p}{1-\cos\alpha}$；
2. $|AB| = \frac{2p}{\sin^2\alpha}$；
3. $S_{\triangle AOB} = \frac{p^2}{2\sin\alpha}$。

在这里我还是说一下，性质不是最重要的，理解推导过程会更重要一点，通过理解证明加深对性质本身的理解。

对于性质 1，不难看出在 $\triangle AFD$ 中有：

$$
\begin{aligned}
\cos(\pi - \alpha) &= \frac{|FD|}{|AF|} \\
-\cos\alpha &= \frac{|AA'| - p}{|AF|} \\
|AF| + |AF|\cos\alpha &= p \\
|AF| &= \frac{p}{1+\cos\alpha}
\end{aligned}
$$

类似的可以证明 $|BF| = \frac{p}{1-\cos\alpha}$

□

对于性质 2，由性质 1 可得

$$\begin{aligned}|AB| &= |AF| + |BF| \\ &= \frac{p}{1+\cos\alpha} + \frac{p}{1-\cos\alpha} \\ &= \frac{p(1-\cos\alpha+1+\cos\alpha)}{1-\cos^2\alpha} \\ &= \frac{2p}{\sin^2\alpha}\end{aligned}$$

□

对于性质 3，作 OE 垂直 AB 垂足为 E，显然 OE 就是 $\triangle OAB$ 边 AB 的高，因此

$$\begin{aligned}S_{\triangle OAB} &= \frac{1}{2}|AB|\cdot|OE| \\ &= \frac{1}{2}|AB|\cdot\sin\alpha|OF|\end{aligned}$$

结合性质 2，得到

$$\begin{aligned}S_{\triangle OAB} &= \frac{1}{2}\cdot\frac{2p}{\sin^2\alpha}\cdot\frac{p\sin\alpha}{2} \\ &= \frac{p^2}{2\sin\alpha}\end{aligned}$$

□

与椭圆、双曲线一样，抛物线也有通径，平行于准线的焦点弦叫做抛物线的通径，由焦点弦的知识可以快速求得通径大小，设抛物线 $C: y^2 = 2px$ 的通径为 PQ，交 C 于 P, Q，则

$$\begin{aligned}|PQ| &= |PF| + |QF| \\ &= 2x_0 + p = 2p\end{aligned}$$

§28.3.2 阿基米德三角形

定义

过圆锥曲线上任意两点 A, B 作切线，两切线交于点 P，$\triangle PAB$ 称为**阿基米德三角形**，当 AB 过圆锥曲线焦点时，称为**阿基米德焦点三角形**，称 $\angle P$ 为顶角，AB 为底边。

在高考中主要研究抛物线上的阿基米德三角形的性质。

设抛物线 $C: y^2 = 2px, p > 0$，焦点为 F，抛物线上两点 A, B，过这两点作抛物线的切线，两切线交于一点 P，A, B 的中点为 M，P, M 中点为 N，有如下性质：

1. MP 垂直于准线，或着说 MP 平行于轴。

2. 设 $|AB| = a$，则 $\triangle PAB$ 面积最大值为 $\frac{a^3}{8p}$。

3. $\angle PFA = \angle PFB$。

4. $|AF| \cdot |BF| = |PF|^2$。

5. PM 的中点 Q 在抛物线上，且过 Q 作抛物线的切线 l，有 $l // l_{AB}$。

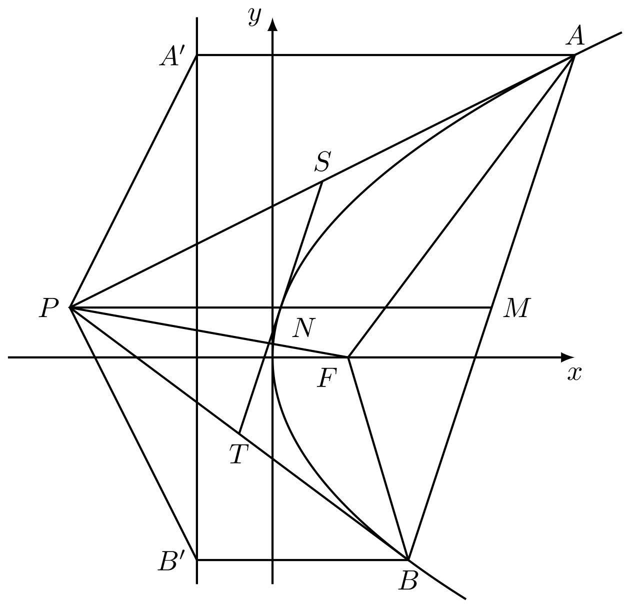


图 28.9: 抛物线阿基米德三角形

证明以上性质：

引理

抛物线 $C: y^2 = 2px, p > 0$，其中 $A(x_1, y_1), B(x_2, y_2)$ 在抛物线 C 上，过 A, B 作抛物线的切线，两切线交于点 $P(x_0, y_0)$，则有：

$$\begin{cases} x_0 &= \frac{y_1 y_2}{2p} \\ y_0 &= \frac{y_1 + y_2}{2} \end{cases}$$

♣

证明：对 $y^2 = 2px$ 两边微分，易知：

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{p}{y}$$

则过 A, B 的两切线 l_1, l_2 方程分别为

$$l_1 : y - y_1 = \frac{p}{y_1}(x - x_1)$$
$$l_2 : y - y_2 = \frac{p}{y_2}(x - x_2)$$

由于 A, B 都在 C 上，则考虑用 y_i 代替 $x_i, i = 1, 2$，即

$$A\left(\frac{y_1^2}{2p}, y_1\right), \quad B\left(\frac{y_2^2}{2p}, y_2\right)$$

联立两个直线方程，即得交点 P 的坐标

$$\begin{aligned}\frac{p}{y_1}(x_0-x_1)+y_1&=\frac{p}{y_2}(x_0-x_2)+y_2\\ \frac{px_0}{y_1}+\frac{y_1}{2}&=\frac{px_0}{y_2}+\frac{y_2}{2}\\ 2px_0(y_2-y_1)&=y_1y_2(y_2-y_1)\\ x_0&=\frac{y_1y_2}{2p}\end{aligned}$$

再将 x_0 带入 l_1 中即可得：

$$\begin{aligned}y_0-y_1&=\frac{p}{y_1}\left(\frac{y_1(y_2-y_1)}{2p}\right)\\ y_0&=\frac{y_1+y_2}{2}\end{aligned}$$

□

对于性质 1，A,B 的中点 M 坐标为

$$M\left(\frac{x_1+x_2}{2},\frac{y_1+y_2}{2}\right)$$

通过引理即有

$$\overrightarrow{MP}=\left(\frac{y_1y_2-p(x_1+x_2)}{2p},0\right)$$

易知 $\overrightarrow{MP}$ 与准线垂直，即准线的方向向量 $\overrightarrow{v}=(0,1)$，显然：

$$\overrightarrow{MP}\cdot\overrightarrow{v}=0$$

□

对于性质 2，设 $|AB|=a$，直线 AB 的方程为 $x=my+n$，点 P 到直线 AB 的距离为 d，易知：

$$\begin{aligned}d\geqslant|PM|&=|\overrightarrow{PM}|\\ &=\left|\frac{y_1y_2-p(x_1+x_2)}{2p}\right|\\ &=\frac{(y_1-y_2)^2}{4p}\end{aligned}$$

联立抛物线与直线 AB 的方程可以将 a 表示出来：

$$\begin{aligned}a&=\sqrt{(1+m^2)(y_1-y_2)^2}\\ a^2&=(1+m^2)(y_1-y_2)^2\end{aligned}$$

注意到 $1+m^2$ 是恒大于等于 1 的，因此

$$\frac{a^2}{4p}\geqslant\frac{(y_1-y_2)^2}{4p}=d$$

最后根据 $S_{\triangle PAB}=\frac{1}{2}|AB|d$ 可以得到

$$S_{\triangle PAB}\leqslant\frac{a}{2}\cdot\frac{a^2}{4p}=\frac{a^3}{8p}$$

当且仅当 $m=0$ 时取得最大值。

□

对于性质 3，作 $AA'\perp l$ 垂足为 A'，l 为准线，作 $BB'\perp l$ 垂足为 B'，连接 $A'P,AA',PB,BB'$，容易得到：

$$k_{A'F}=-\frac{y_1}{p}$$

不难得出：

$$k_{A'F}\cdot k_{PA}=-\frac{y_1}{p}\cdot\frac{p}{y_1}=-1$$

即 $A'F\perp PA$，再据 $|AA'|=|AF|$，设 PA 与 $A'F$ 相交于 Q，$|QA|=|QA|$，则有

$$\triangle QAA'\sim\triangle QAF$$

可得 $\angle A'AP=\angle FAP$，因此

$$\triangle PAA'\cong\triangle PAF$$

有 $|PA'|=|PF|,\angle PA'A=\angle PFA$，同理可证 $|PB'|=|PF|,\angle PB'B=\angle PFB$。因为 $|PA'|=|PB'|$，所以 $\angle PA'B'=\angle PB'A'$，则：

$$\begin{aligned}\angle PA'B'+\pi&=\angle PB'A'+\pi\\\angle PA'A&=\angle PB'B\\\angle PFA&=\angle PFB\end{aligned}$$

□

对于性质 4，一方面

$$\begin{aligned}|AF|\cdot|BF|&=(x_1+\frac{p}{2})(x_2+\frac{p}{2})\\&=x_1x_2+\frac{p}{2}(x_1+x_2)+\frac{p^2}{4}\\&=\frac{y_1^2y_2^2}{4p^2}+\frac{y_1^2+y_2^2+p^2}{4}\end{aligned}$$

另一方面

$$\begin{aligned}|PF|^2&=\frac{(y_1y_2-p^2)^2}{4p^2}+\frac{(y_1+y_2)^2}{4}\\&=\frac{y_1^2y_2^2}{4p^2}+\frac{y_1^2+y_2^2+p^2}{4}\end{aligned}$$

即证

$$|AF|\cdot|BF|=|PF|^2$$

□

对于性质 5，由于 $N(x_N, y_N)$ 是 P, M 的中点，易知

$$\begin{aligned} x_N &= \frac{x_P + x_M}{2} \\ &= \frac{y_1 y_2}{2p} + \frac{x_1 + x_2}{2} \\ &= \frac{(y_1 + y_2)^2}{4p} \end{aligned}$$

可得

$$\begin{cases} x_N &= \frac{(y_1+y_2)^2}{8p} \\ y_N &= \frac{y_1+y_2}{2} \end{cases}$$

显然 N 在抛物线上，因为有：

$$2px_N = y_N$$

且过 N 作抛物线 C 的切线 l'，斜率为 k'，有

$$k' = \frac{p}{y_N} = \frac{2p}{y_1 + y_2}$$

直线 AB 的斜率 k_{AB} 为

$$k_{AB} = \frac{y_1 - y_2}{x_1 - x_2} = \frac{2p}{y_1 + y_2}$$

□

特别的，若直线 AB 过焦点 F，即**阿基米德焦点三角形**，除了以上的性质外还有一些特殊的性质：

1. $\angle PFA = \angle PFB = \frac{\pi}{2}$。
2. $|AF| \cdot |BF| = |PF|^2$。

这两条性质的证明是显然的，当 F 在直线 AB 上时，P 在准线上，这一点我会在曲线轨迹那一节中给出推导，可以证明 $\angle APB$ 为直角，性质 2 也就可以看成直角三角形中的射影定理（相似）。

抛物线与直线的关系

与研究椭圆、双曲线一样，我们联立圆锥曲线和直线的方程

$$\begin{cases} C: & y^2 = 2px, \quad p > 0 \\ l: & y = kx + m \end{cases}$$

可得：

$$(kx + m)^2 = 2px$$
$$k^2x^2 + (2km - 2p)x + m^2 = 0$$

与双曲线一样，需要讨论二次项系数为零的情况；

当 $k^2 = 0$，显然这时直线平行于 x 轴，于抛物线有且只有一个交点；

当 $k^2 \neq 0$，可以得到二次方程的判别式：

$$\begin{aligned} \varDelta &= (2km-2p)^2 - 4k^2m^2 \\ &= 4p(p-2km) \end{aligned}$$

注意到 $4p > 0$，因此只需判断 $p-2km$ 即可；可得下表：

斜率	判别式	位置关系	交点数
$k=0$	–	相交	1
$k \neq 0$	$p-2km>0$	相交	2
	$p-2km=0$	相切	1
	$p-2km<0$	相离	0

表 28.3: 直线与抛物线的关系

焦点三角形的面积问题，设 $\angle F_1PF_2 = \theta$，有：

1. 在椭圆中，$S_{\triangle PF_1F_2} = b^2 \tan\frac{\theta}{2}$；
2. 在双曲线中，$S_{\triangle PF_1F_2} = \frac{b^2}{\tan\frac{\theta}{2}}$。

§28.4 极坐标系

平面上的一点，在直角坐标系中可以用坐标，即有序数对 (x,y) 来表示；在复数的学习过程中，我们也知道了一个向量（复数）可以用其模长以及辐角主值来唯一确定，而向量和点的坐标可以做到一一对应，因而想到用一个长度参数和一个角度参数来确定平面上一个点，由此建立的坐标系称为**极坐标系**。

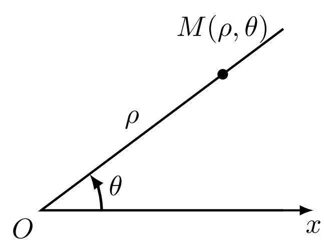


图 28.10: 极坐标系

如图，在平面上取一点 O，称为**极点**；自极点引一条射线 Ox，称为**极轴**；在选定一个长度单位、一个角度单位及其正方向，这样就建立了一个**极坐标系**。

通常情况下，角度单位选取弧度制，并且以逆时针为正方向。

设 M 是平面内一点，极点 O 到 M 的距离 $|OM|$ 叫做点 M 的极径，记为 ρ；以极轴 Ox 为始边，射线 OM 为终边的角 xOM 叫做点 M 的**极角**，记为 θ，有序数对 (ρ,θ) 叫做点 M 的**极坐标**，记为 $M(\rho,\theta)$。

注意到当 (ρ,θ) 表示的极点和 $(-\rho,\theta+\pi)$ 表示的极点相同，以及 (ρ,θ) 表示的极点与 $(\rho,\theta+2k\pi)$ 表示的极点相同，因此为了使极坐标和平面上的点是双射（一一对应），可以规定

$$\rho \geqslant 0, \quad \theta \in [0, 2\pi)$$

特别的当 $\rho = 0$ 时，θ 取何值都是极点。

极坐标和直角坐标转换

极坐标和直角坐标可以互相转换，延上节设，点 M 的极坐标为 (ρ, θ)，直角坐标为 (x, y)，那么易知

$$\rho^2 = x^2 + y^2,\ \ \tan\theta = \frac{y}{x}, x \neq 0$$

由上式可推出

$$y\cos\theta = x\sin\theta$$

因此有

$$\begin{cases} x &= \rho\cos\theta \\ y &= \rho\sin\theta \end{cases}$$

§28.4.1　极坐标方程

一般地，在极坐标系中，如果平面曲线 C 上任取一点至少有一个满足方程

$$f(\rho, \theta) = 0$$

并且坐标适合方程 $f(\rho, \theta) = 0$ 的点都在曲线 C 上，那么方程 $f(\rho, \theta) = 0$ 称为曲线 C 的**极坐标方程**。

我们先从简单的情况开始讨论。

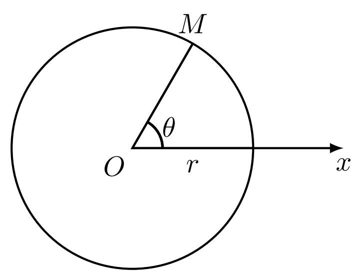


图 28.11: 圆心在极点的圆

如图为一圆心在极点的圆，由于圆到圆心的距离恒为 r，因此该曲线的极坐标方程为

$$\rho = r$$

式中没有 θ，就像在直角坐标系中平行于坐标轴的直线方程没有 x 或 y 一样，θ 的取任意值时，点 $M(\rho, \theta)$ 都在圆上。

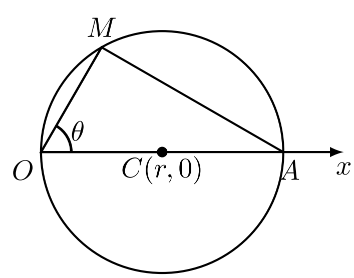


图 28.12: 极点在圆上的圆

如图，圆 C 的圆心在极径上，且极点为圆上一点，记圆上一点为 $M(\rho,\theta)$，注意到 $\angle OMA=\frac{\pi}{2}$，因此有

$$\rho=|OM|=2r\cos\theta$$

当 $\theta\in(-\frac{\pi}{2},\frac{\pi}{2})$ 时，上式就表示该圆上异于极点的点的极坐标方程。

下面我们讨论更一般的情况，也就是在平面直角坐标系中，圆心为 $C(x_0,y_0)$，半径为 r 的圆的极坐标方程。

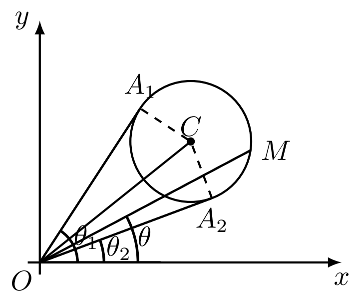


图 28.13: 一般情况的圆

我们知道，这个圆在直角坐标系下的方程为

$$(x-x_0)^2+(y-y_0)^2=r^2$$

由于

$$\begin{cases} x &= \rho\cos\theta \\ y &= \rho\sin\theta \end{cases}$$

因此带入可知

$$(\rho\cos\theta-x_0)^2+(\rho\sin\theta-y_0)^2=r^2$$

需要注意 $\rho=|OM|$，上述方程就是该圆的极坐标方程，不过从上述方程很难看出 ρ,θ 的取值范围，因此我们从几何上分析。

对于 ρ，不难看出是直线 OC 与圆的两个交点，其一为最小值，另一为最大值，即有

$$\min\{\rho\}=|OC|-r=\sqrt{x_0^2+y_0^2}-r$$
$$\max\{\rho\}=|OC|+r=\sqrt{x_0^2+y_0^2}+r$$

因此得到 ρ 的取值范围：

$$\sqrt{x_0^2+y_0^2}-r\leqslant\rho\leqslant\sqrt{x_0^2+y_0^2}+r$$

再来分析 θ 的取值，过原点做圆 C 的两条切线 OA_1,OA_2，记其斜率分别为 θ_1,θ_2，显然

$$\theta_2\leqslant\theta\leqslant\theta_1$$

下面再来考虑如何表示 θ_1,θ_2，由点 A_i 在圆上，以及勾股定理易知：

$$\begin{cases} (x_a-x_0)^2+(y_a-y_0)^2 &= r^2 \\ x_a^2+y_a^2+r^2=x_0^2+y_0^2 \end{cases}$$

由上式解出 A_1, A_2 的坐标后，即知 θ_1, θ_2 的取值，用反三角函数来表示，需要注意的是当在不同象限时，两个点的计算方式不一样。如图所示有

$$\arctan\left(\frac{y_{a_2}}{x_{a_2}}\right) \leqslant \theta \leqslant \arctan\left(\frac{y_{a_1}}{x_{a_1}}\right)$$

最后在来讨论一种特殊的情况，也即圆心在极轴上，但极点不在圆上。有以上的推导可知，其极坐标方程为

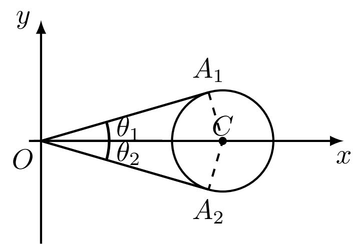


图 28.14: 极径上的圆

$$(\rho\cos\theta - x_0)^2 + \rho^2\sin^2\theta = r^2$$

这时 ρ, θ 的范围就很好确定了：

$$x_0 - r \leqslant \rho \leqslant x_0 + r$$

而 θ

$$\arcsin\left(-\frac{r}{x_0}\right) \leqslant \theta \leqslant \arcsin\left(\frac{r}{x_0}\right)$$

直线方程

经过极点的直线可以简单的表示为

$$\theta = \theta_0$$

其中，θ_0 是该直线的倾斜角。

下面我们来讨论一般情况：

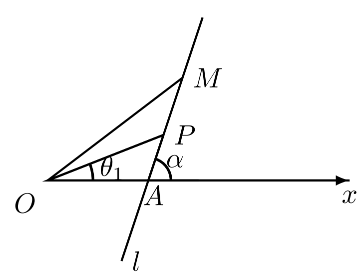


图 28.15: 一般的直线

如图，设点 $M(\rho, \theta)$ 为直线 l 上除 P 外的一点，连接 OM，可知

$$|OM| = \rho, \quad \angle MOA = \theta$$

不妨记点 $P(\rho_1, \theta_1)$，则

$$|OP| = \rho_1, \quad \angle POA = \theta_1$$

设直线的倾斜角为 α，于是

$$\angle OMP = \alpha - \theta, \quad \angle OPM = \pi - \alpha + \theta_1$$

由正弦定理可知

$$\frac{|OM|}{\sin\angle OPM} = \frac{|OP|}{\sin\angle OMP}$$

即

$$\rho\sin(\alpha - \theta) = \rho_1\sin(\alpha - \theta_1)$$

上式就是直线 l 的极坐标方程。

我们也可以用直角坐标系和极坐标系的转换来求，直线的一般方程为

$$Ax + By + C = 0$$

因此

$$A\rho\cos\theta + B\rho\sin\theta + C = 0$$

即得，由此易知

$$\rho = -\frac{C}{A\cos\theta + B\sin\theta}$$

上式不可以描述 $A\cos\theta + B\sin\theta = 0$ 的情况，即直线过极点的情况。

§28.4.2 圆锥曲线统一定义

圆锥曲线的极坐标方程可以像上面讨论的一样，用直角坐标和极坐标的变换来得到，下面我们讨论一种特殊的情况，即当极点在焦点的情况。

前面的学习中，我们了解了圆锥曲线的一、二、三定义，这些定义中唯独第二定义能囊括这三类曲线。

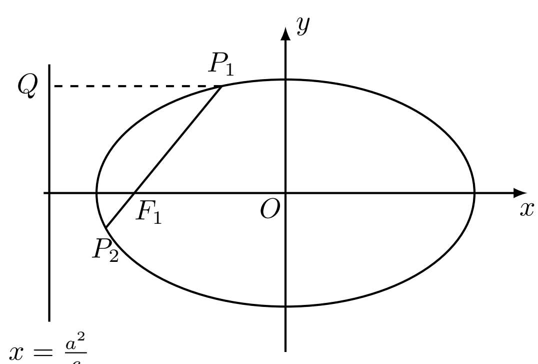


图 28.16: 圆锥曲线统一定义

一圆锥曲线 C，以一焦点为极点，沿对称轴方向建立极坐标系，记曲线上一点 $P_1(\rho_1, \theta_1)$，那么有

$$|P_1F_1| = \rho, \quad |P_1Q| = |P_1F_1|\cos\theta + p$$

其中 p 是焦点到准线的距离，称为**焦准距**，故离心率可以表述为

$$e=\frac{|P_1F_1|}{|P_1Q|}=\frac{\rho}{\rho\cos\theta+p}$$

整理上式，可得圆锥曲线的极坐标方程为

$$\rho=\frac{pe}{1-e\cos\theta} \quad ①$$

上式就是圆锥曲线的统一方程。

进一步，对于椭圆，其焦准距可以表示为 $\frac{a^2}{c}-c=\frac{b^2}{c}$，因此 ① 式可表示为

$$\rho=\frac{b^2}{a-c\cos\theta}$$

对于双曲线，其焦准距可表示为 $c-\frac{a^2}{c}=\frac{b^2}{c}$，因此 ① 式也是表示为

$$\rho=\frac{b^2}{a-c\cos\theta}$$

图中 $|P_1P_2|$ 为焦点弦，由统一方程可得

$$\begin{aligned}|P_1P_2|&=|F_1P_1|+|F_1P_2|\\&=\rho_1+\rho_2\\&=\frac{pe}{1-e\cos\theta}+\frac{pe}{1-e\cos(\pi-\theta)}\\&=\frac{2pe}{1-e^2\cos^2\theta}\end{aligned} \quad ②$$

这其实就是之前我们在抛物线中研究的“角板”内容。

通过 ① 式，可以发现 ρ_1,ρ_2 及其对称的形式，由此可以推出一些特别的性质。

例如，不难发现 ρ_1,ρ_2 的倒数和为定值。

$$\frac{1}{\rho_1}+\frac{1}{\rho_2}=\frac{1-e\cos\theta}{pe}+\frac{1+e\cos\theta}{pe}=\frac{2}{pe}$$

由算术平均和调和平均可知

$$\begin{aligned}\frac{\rho_1+\rho_2}{2}&\geqslant\frac{2}{\frac{1}{\rho_1}+\frac{1}{\rho_2}}\\\frac{|P_1P_2|}{2}&\geqslant pe\\|P_1P_2|&\geqslant 2pe\end{aligned}$$

当且仅当 $\rho_1=\rho_2$ 时上式取等，当曲线为椭圆时，上式右边为 $2\frac{b^2}{a}$，即通径长度，进而证明了在椭圆中最短的焦点弦为通径，当曲线为双曲线或抛物线时同理。

再来，当 ρ_1,ρ_2 的比值为定值 λ 时，即

$$\frac{\rho_1}{\rho_2}=\lambda=\frac{1+e\cos\theta}{1-e\cos\theta}$$

整理可知

$$e\cos\theta = \frac{\lambda - 1}{\lambda + 1}$$

当 θ 为钝角时

$$e\cos\theta = -\frac{\lambda - 1}{\lambda + 1}$$

为简洁可表示为

$$|e\cos\theta| = \frac{|\lambda - 1|}{|\lambda + 1|}$$

带入 ② 式可知，焦点弦的弦长为一定值。

这意味着，过焦点 F 的直线交圆锥曲线与 A, B 两点，若点 F 分线段 AB 成定比 λ，那么焦点弦的弦长 $|AB|$ 为定值

$$|AB| = \frac{(1+\lambda)^2 pe}{2\lambda}$$

值得注意的是 F 为内分点，因此 $\lambda > 0$，上式为定值。

§28.5 一些结论

圆锥曲线有一些共同的结论，我汇总放在一起说。

§28.5.1 焦点三角形面积

椭圆的焦点三角形面积为 $S_{\triangle PF_1F_2} = b^2 \tan\frac{\theta}{2}$。

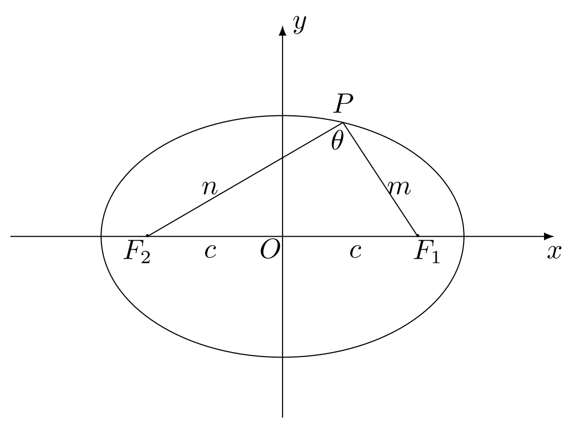


图 28.17: 焦点三角形

证明：

设 $PF_1 = m, PF_2 = n, \angle F_1PF_2 = \theta$

$$\begin{aligned} |F_1F_2|^2 &= m^2 + n^2 - 2mn\cos\theta \\ 4c^2 &= (m+n)^2 - 2mn(\cos\theta + 1) \\ 2b^2 &= mn(\cos\theta + 1) \\ mn &= \frac{2b^2}{\cos\theta + 1} \end{aligned}$$

由正弦定理可知：

$$\begin{aligned}
\because S_{PF_1F_2} &= \frac{1}{2}mn\sin\theta \\
\therefore S_{PF_1F_2} &= \frac{1}{2}\cdot\frac{2b^2}{\cos\theta+1}\cdot\sin\theta \\
\Longrightarrow S_{PF_1F_2} &= b^2\cdot\frac{\sin\theta}{1+\cos\theta} \\
\Longrightarrow S_{PF_1F_2} &= b^2\cdot\frac{2\sin\frac{\theta}{2}\cos\frac{\theta}{2}}{1+2\cos^2\frac{\theta}{2}-1} \\
\Longrightarrow S_{PF_1F_2} &= b^2\frac{\sin\frac{\theta}{2}}{\cos\frac{\theta}{2}} \\
\Longrightarrow S_{PF_1F_2} &= b^2\tan\frac{\theta}{2}
\end{aligned}$$

□

这个推论的证明用到了余弦定理、正弦定理推论、倍角公式，值得去推一推。用类似的方法可以推出双曲线的焦点三角形面积为 $S_{\triangle PF_1F_2}=\frac{b^2}{\tan\frac{\theta}{2}}$。

§28.5.2 弦长公式

当直线和圆锥曲线 $f(x,y)=0$ 相交于 A,B 两点时，求弦长 $|AB|$。有：

1. 设直线方程为：$y=kx+t$，联立后得到：$|AB|=\sqrt{1+k^2}\cdot|x_1-x_2|=\sqrt{1+k^2}\cdot\frac{\sqrt{\Delta}}{|a|}$

2. 设直线方程为：$x=my+n$，联立后得到：$|AB|=\sqrt{1+m^2}\cdot|y_1-y_2|=\sqrt{1+m^2}\cdot\frac{\sqrt{\Delta}}{|a|}$

这个结论并不难证明，在这我仅对直线方程为 $y=kx+t$ 是作证明。

联立两个方程

$$\begin{cases} f(x,y)=0 \\ y=kx+t \end{cases}$$

消去 y 会得到一个一元二次方程

$$ax^2+bx+c=0$$

由于直线交圆锥曲线于 A,B 两点，因此对于这个一元二次方程一定有两解，不妨设 $A(x_1,y_1),B(x_2,y_2)$ 其中 x_1,x_2 就是这个一元二次方程的两个解，可以列出

$$\begin{aligned}
|AB| &= \sqrt{(x_1-x_2)^2+(y_1-y_2)^2} \\
&= \sqrt{(x_1-x_2)^2+k^2(x_1-x_2)^2} \\
&= \sqrt{1+k^2}|x_1-x_2|
\end{aligned}$$

□

在实际应用中，可以不需要把 x_1, x_2 解出来，利用韦达定理，可以作如下转换

$$
\begin{aligned}
|x_1 - x_2| &= \sqrt{(x_1 + x_2)^2 - 4x_1x_2} \\
&= \sqrt{\frac{b^2}{a^2} - \frac{4ac}{a^2}} \\
&= \frac{\sqrt{\Delta}}{|a|}
\end{aligned}
$$

§28.6 补充

$$
\begin{cases}
e \in (0,1), & \text{椭圆} \\
e = 1, & \text{抛物线} \\
e \in (1,+\infty), & \text{双曲线}
\end{cases}
$$

§28.6.1 曲线轨迹

曲线的**轨迹**与**轨迹方程**是两个不同的概念，*e.g.*：

动点 P 的轨迹为椭圆，其轨迹方程为

$$\frac{x^2}{5^2} + \frac{y^2}{3^2} = 1$$

求轨迹主要分为两个方法：**直接法**与**间接法**

直接法

定义法、直译法都属于直接法的范畴，通过已知条件，结合所学过的曲线轨迹的性质，得到动点轨迹；

高中我们已经学过了椭圆、双曲线、抛物线，这些曲线的几何意义，如到两定点距离之和为常数，或到定点和定直线的距离相等的轨迹；但题目中往往还会联合其他的知识，例如平面几何的知识、立体几何的知识等等。

e.g. 正方体 $ABCD - A'B'C'D'$ 中，P 是侧面 $BB'C'C$ 内一动点，若 P 到直线 BC 与到直线 $C'D'$ 的距离相等，求 P 的轨迹。

由立体几何的知识易知，直线 $D'C' \perp$ 平面 $BB'C'C$，因此 P 到 $D'C'$ 的距离即为 $|PC'|$；因此动点 P 到定点 C' 和到定直线 BC 的距离相等，其轨迹是以 C' 为焦点，直线 BC 为准线的抛物线。

间接法

有些动点轨迹无法直接求出，动点的运动是依靠其他点的运动而运动的，这时通过确定动点所**依靠运动**的点的轨迹方程，在通过动点与其依赖点的关系即可转换为动点轨迹方程。

e.g. O 是滑槽 AB 的中点，短杆 ON 可绕 O 转动，长杆 MN 通过 N 处铰链与 ON 连接，MN 的栓子 D 可沿滑槽 AB 滑动，且 $|DN| = |ON| = 1|MN| = 3$，当栓子 D 在 AB 作往复运动时，带动 N 绕 O 转动一周（D 不动时，N 也不动），M 处的笔尖画出的曲线记为 C，以 O 为原点，AB 为 x 轴，建立直角坐标系，求 C 的轨迹方程。

通过题目已知动点 M 依靠 N 运动，而 N 的运动轨迹为圆，且易列：

$$x_N^2 + y_N^2 = 1 \tag{①}$$

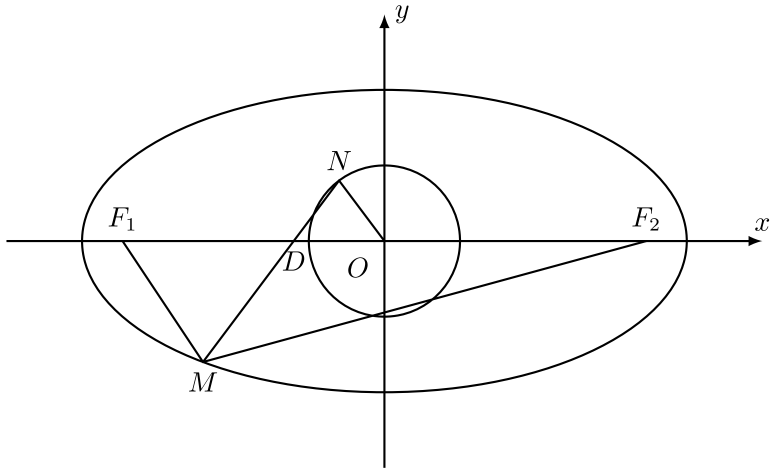


图 28.18: 轨迹方程间接法

由于 $|DN| = |ON|$ 可知点 D 的坐标

$$\begin{cases} x_D &= 2x_N \\ y_D &= 0 \end{cases}$$

注意到 $\overrightarrow{DM} = 2\overrightarrow{ND}$，即有

$$\begin{cases} x_M - x_D &= 2x_D - 2x_N &= 4x_N \\ y_M - y_D &= 2y_D - 2y_N &= -2y_N \end{cases}$$

综上，带入 ① 得到

$$\frac{x_M^2}{4^2} + \frac{y_M^2}{2^2} = 1$$

即 C 是焦点为 $(-2\sqrt{3}, 0), (2\sqrt{3}, 0)$，长轴为 8 的椭圆，其轨迹方程就是上式。

e.g. 抛物线 $C: y^2 = 2px, p > 0$，A, B 为抛物线上的两点，过 A, B 作抛物线的切线，两切线交于点 P，若 AB 过抛物线内一定点 C，求 P 的轨迹方程。

记 $A(x_1, y_1), B(x_2, y_2), C(x_0, y_0)$，易知 $\triangle PAB$ 是阿基米德三角形，因此

$$P\left(\frac{y_1y_2}{2p}, \frac{y_1+y_2}{2}\right)$$

又 A, B, C 三点共线，因此有

$$\begin{aligned} y_0 - y_1 &= \frac{y_2 - y_1}{x_2 - x_1}(x_0 - x_1) \\ (y_0 - y_1)(y_2^2 - y_1^2) &= (y_2 - y_1)(2px_0 - y_1^2) \\ y_0y_2 + y_0y_1 - y_1y_2 &= 2px_0 \\ y_0(y_1 + y_2) - y_1y_2 &= 2px_0 \qquad ① \end{aligned}$$

当 $C(x_0, y_0)$ 为定值时，可以通过上式确定 $y_1 + y_2$ 与 y_1y_2 的关系；如若把其中一个看为参数 t，那么即可确定

P 所在直线，特别的，当 C 为抛物线焦点 F 时，有

$$\begin{cases} x_0 & = \frac{p}{2} \\ y_0 & = 0 \end{cases}$$

带入 ① 式可得

$$y_1 y_2 = -p^2$$

P 的坐标为

$$P\left(-\frac{p}{2}, \frac{y_1+y_2}{2}\right)$$

即 P 在准线上。

通过 ① 式也可以知道到，如果 P 在定直线上运动，那么 A,B 过定点 C。

§28.6.2 定比点差法

前文中，已经在椭圆、双曲线的第三定义以及中点弦的推导里运用了**点差法**；点差法主要用于有关中点的问题中，现在我们类比点差法来介绍它“推广”的形式（事实上并不能说是推广，这一点我会在后面说明），也即**定比点差法**。

已知两点 $A(x_1,y_1),B(x_2,y_2)$ 在二次曲线 $\frac{x^2}{a^2}\pm\frac{y^2}{b^2}=1$（椭圆或双曲线上）,在直线 AB 上有两点 $N(x_n,y_n),M(x_m,y_m)$，满足

$$\frac{|AN|}{|BN|}=\frac{|AM|}{|BM|}=\lambda,\lambda\neq\pm1$$

不难看出，N,M 是 AB 的内、外分点，分线段定比 λ，不妨设 $\lambda>0$，N 为内分点，M 为外分点；则有：

$$\overrightarrow{AN}=\lambda\overrightarrow{NB},\quad \overrightarrow{AM}=-\lambda\overrightarrow{MB}$$

也即

$$\overrightarrow{ON}=\frac{\overrightarrow{OA}+\lambda\overrightarrow{OB}}{\lambda+1}$$
$$(x_n,y_n)=\left(\frac{x_1+\lambda x_2}{\lambda+1},\frac{y_1+\lambda y_2}{\lambda+1}\right)$$

以及

$$\overrightarrow{OM}=\frac{\overrightarrow{OA}-\lambda\overrightarrow{OB}}{1-\lambda}$$
$$(x_m,y_m)=\left(\frac{x_1-\lambda x_2}{1-\lambda},\frac{y_1-\lambda y_2}{1-\lambda}\right)$$

由于 A,B 在二次曲线 $C:\frac{x^2}{a^2}\pm\frac{y^2}{b^2}=1$ 上，带入有

$$\frac{x_1^2}{a^2}\pm\frac{y_1^2}{b^2}=1 \quad ①$$
$$\frac{x_2^2}{a^2}\pm\frac{y_2^2}{b^2}=1$$
$$\frac{x_2^2\lambda^2}{a^2}\pm\frac{y_2^2\lambda^2}{b^2}=\lambda^2 \quad ②$$

① – ② 两式做差得到

$$\frac{x_1^2-x_2^2\lambda^2}{a^2}+\frac{\pm y_1^2\mp y_2^2\lambda^2}{b^2}=1-\lambda^2$$
$$\frac{(x_1-x_2\lambda)(x_1+x_2\lambda)}{a^2(1-\lambda)(1+\lambda)}\pm\frac{(y_1-y_2\lambda)(y_1+y_2\lambda)}{b^2(1-\lambda)(1+\lambda)}=1$$
$$\frac{x_mx_n}{a^2}\pm\frac{y_my_n}{b^2}=1$$

可以看出，当 N,M 任意一个被确定，那么另一点一定在一个定直线上运动。

综上，$\lambda\neq\pm1$，这意味着 N,M 不能为中点，所以不能算是点差法的推论，但能它在更广泛的情况下使用，以及与点差法有着相同的思路。

§28.7　定点问题

很多过定点问题都可以用“先猜后证”的方法解决，下面展示一般流程：

一，先猜定点，常用方法有：

1. 取特殊直线，两直线交于一点，确定定点。
2. 利用圆锥曲线的对称性，找到两条满足参数条件且“对称”的直线，确定定点。

二，证明直线这个定点

当定点猜出来以后，只需要通过三点共线来证明即可，例如定点为 A，直线上两点为 P,Q（可以表示出来的两点），可以用向量证明：

$$\overrightarrow{AP}=\lambda\overrightarrow{AQ}$$

或者用斜率来证明：

$$k_{AP}=k_{AQ}$$

此外，还有猜不出定点的情况，这时候转换为恒成立问题来解决：

通过已知条件整理过定点的直线方程，如：

$$(a_1k+b_1)x+(a_2k+b_2)y+a_3k+b_3=0$$

其中 $a_i,b_i,i=1,2,3$ 是已知常数，k 是参数（往往是截距或斜率），将 k 设为主元整理：

$$(a_1x+a_2y+a_3)k+b_1x+b_2y+b_3=0$$

上式是关于 k 的一元一次方程，由于对于任意 k 直线都过定点 (x_0,y_0)，即将 (x_0,y_0) 带入上式对于任意 k 恒

成立，所以有：

$$\begin{cases} a_1x_0+a_2y_0+a_3 &= 0 \\ b_1x_0+b_2y_0+b_3 &= 0 \end{cases}$$

解这个方程即可得到定点 (x_0,y_0)，如果不理解为什么这个方程组成立，可以想象一下如令：

$$a_1x_0+a_2y_0+a_3=A$$
$$b_1x_0+b_2y_0+b_3=B$$

则关于 k 的一元一次方程为 $Ak+B=0$，若 $A\neq 0, B\neq 0$，则该一元一次方程是一个直线，不能满足对于所有 k 都成立。

常见的定点模型

一、过椭圆 $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$ 上一定点 $P(x_0,y_0)$ 引两条直线，交椭圆于 A,B 两点；若 $k_{AP}\cdot k_{BP}=\lambda$ 其中 λ 为定值，且 $\lambda\neq\frac{b^2}{a^2}$ 则直线 AB 过定点

$$\left(\frac{b^2+\lambda a^2}{\lambda a^2-b^2}x_0, \frac{b^2+\lambda a^2}{b^2-\lambda a^2}y_0\right)$$

二、过椭圆 $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$ 上一定点 $P(x_0,y_0)$ 引两条直线，交椭圆于 A,B 两点；若 $k_{AP}+k_{BP}=\mu$ 其中 μ 为定值，且 $\mu\neq 0$ 则直线 AB 过定点

$$\left(x_0+\frac{2y_0}{\mu}, -y_0-\frac{2b^2x_0}{a^2\mu}\right)$$

对上面的模型进行记忆是无意义的，下面介绍一个处理方法来解决这类问题。

§28.7.1 齐次化

已知 $\tan\theta=-2$，求解 $\sin 2\theta-\cos^2\theta$，我们可以将他们转为齐二次方程

$$\begin{aligned}\sin 2\theta-\cos^2\theta &= 2\sin\theta\cos\theta-\cos^2\theta \\ &= \frac{2\sin\theta\cos\theta-\cos^2\theta}{\sin^2\theta+\cos^2\theta} \\ &= \frac{2\tan\theta-1}{\tan^2\theta+1}=-1\end{aligned}$$

注意，齐二次后，通过上下同时除以 $\cos^2\theta$ 将二元的方程转为一元，简化了计算的过程。

延续上例的思路，关于 x,y 的齐二次方程同时除以 x^2，就可转化为有关斜率 k 的一元二次方程。下面我们讨论在二次曲线中的齐次处理。

我们注意到直线的一般式：

$$Ax+By+C=0$$

如果将常数项 C 消去可得

$$-\frac{A}{C}x-\frac{B}{C}y=1$$
$$mx+ny=1$$

由于默认了 $C \neq 0$，因此上式可以表达不过原点的所有直线。

通过上式我们可以将非齐次二次方程转为齐二次的方程，这个过程称为**齐次化**，而齐次化的意义在于它能帮助我们化简计算，下面通过齐次化来研究定点问题。

一般的，平面直角坐标系中的二次曲线上有一定点 A，以及两动点 P, Q，若定点与两动点的斜率之积或之和为定值，即

$$k_{AP} \cdot k_{AQ} = C, \text{ 或} k_{AP} + k_{AQ} = C$$

其中 C 为常数，那么直线 PQ 过定点。

常规思路是设点作标 $A(x_0, y_0), P(x_1, y_1), Q(x_2, y_2)$，设二次曲线 C 以及直线 PQ 为

$$C : F(x, y) = 0, \quad l : mx + ny = 1$$

联立它们得到关于 x 或 y 的一元二次方程，通过斜率之积或之和为定值的条件，推出定点的坐标，不过这种方法实际操作非常繁琐，因此我们采用齐次化来简化这类计算。下面举一例来看两种方法的区别：

e.g. 已知直线 AB 交抛物线 $y^2 = 2px, p > 0$ 于 A, B 两点，且 $k_{OA} \cdot k_{OB} = -\frac{1}{2}$，证明：直线 A, B 过定点。

首先是常规思路：设 $A(x_1, y_1), B(x_2, y_2)$，直线 AB 的方程为

$$x = ky + b$$

将其与 $y^2 = 2px$ 联立可得

$$y^2 - 2pky - 2pb = 0$$

根据条件可知

$$k_{OA} \cdot k_{OB} = -\frac{1}{2} \Rightarrow \frac{y_1}{x_1} \cdot \frac{y_2}{x_2} = -\frac{1}{2}$$

根据韦达定理即知：

$$\begin{aligned} \frac{y_1}{x_1} \cdot \frac{y_2}{x_2} &= -\frac{2pb}{(ky_1 + b)(ky_2 + b)} \\ &= -\frac{2pb}{-2k^2pb + 2pk^2b + b^2} \\ &= -\frac{2p}{-2k^2p + 2pk^2 + b} \end{aligned}$$

也即

$$\begin{aligned} \frac{2p}{-2k^2p + 2pk^2 + b} &= \frac{1}{2} \\ 4p &= b \end{aligned}$$

故当 k 取任意值时，都有直线 AB 过定点 $(4p, 0)$。

下面用齐次化来处理：

由题设直线 AB 为

$$mx + ny = 1$$

将它与抛物线方程联立：

$$\begin{aligned} y^2 &= 2px(mx+ny) \\ y^2 - 2pmx^2 - 2pnxy &= 0 \\ k^2 - 2pnk - 2pm &= 0 \end{aligned}$$

由条件得

$$k_1 \cdot k_2 = -2pm = -\frac{1}{2} \Rightarrow m = \frac{1}{4p}$$

因此无论 n 取何值，直线都过定点 $(4p, 0)$。

通过上例可以发现齐次化稍微简化了我们的计算，不过以上是最简单的类型，即定点为原点。更一般的情况是定点不在原点，而对它们的处理一种是将坐标系平移，使得定点为原点，而另一种做法是不平移，下面通过一例来感受：

e.g. 已知点 $A(2,1)$ 和双曲线 $E : \frac{x^2}{2} - y^2 = 1$，$P, Q$ 为双曲线上两动点，它们满足

$$k_{AP} + k_{AQ} = 0$$

求直线 PQ 的斜率。

首先我们看平移的做法，将坐标系平移，使得 $A(2,1)$ 为原点，那么双曲线的方程改写为

$$\frac{(x+2)^2}{2} - (y+1)^2 = 1$$

由于直线 PQ 不过 A，因此设直线方程为

$$mx + ny = 1$$

联立得

$$\begin{aligned} \frac{(x+2)^2}{2} - (y+1)^2 &= 1 \\ x^2 + 4 + 4x - 2y^2 - 2 - 4y &= 2 \\ x^2 - 2y^2 + 4(x-y)(ny+mx) &= 0 \end{aligned}$$

由条件可知

$$k_{AP} + k_{AQ} = 0$$

因此我们只需提取 xy 项与 y^2 的系数比即可

$$\frac{4m - 4n}{2 + 4n} = 0 \Rightarrow m = n$$

即得 PQ 斜率（平移不改变斜率）

$$k=-\frac{m}{n}=-1$$

下面我们采用不平移的做法：

由于直线 PQ 不过点 A，因此设

$$m(x-2)+n(y-1)=1$$

双曲线方程可以写为

$$\begin{aligned}\frac{x^2}{2}-y^2=1\\ \frac{(x-2+2)^2}{2}-(y-1+1)^2=1\\ (x-2)^2+4-4(x-2)-2(y-1)^2-2-4(y-1)=2\\ (x-2)^2-2(y-1)^2-4[(x-2)+(y-1)][m(x-2)+n(y-1)]=0\end{aligned}$$

由条件可知

$$k_1+k_2=\frac{y_1-1}{x_1-2}+\frac{y_2-1}{x_2-2}=0$$

也即 $(x-2)(y-1)$ 的系数与 $(y-1)^2$ 的系数比为 0，立得

$$\frac{4m-4n}{2+4n}=0\Rightarrow m=n$$

即得直线 PQ 的斜率为-1。

不难发现，实际上平移和不平移的方法本质上**毫无区别**，后者不过是没对平移进行说明罢了，光是体现在代数上，不采用新坐标系（平移后的坐标系）的写法；这是高考中规避平移变换而被扣分的做法，在实际研究中，没什么实际意义。

定点模型的证明

下面来证明之前提到的定点模型：

椭圆 $E:\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$ 上一定点 $P(x_0,y_0)$ 引两条直线，交椭圆于 A,B 两点；若 $k_{AP}\cdot k_{BP}=\lambda,k_{AP}+k_{BP}=\mu$，且 $\lambda\neq\frac{b^2}{a^2},\mu\neq 0$ 则直线 AB 过定点

$$\left(\frac{b^2+\lambda a^2}{\lambda a^2-b^2}x_0,\frac{b^2+\lambda a^2}{b^2-\lambda a^2}y_0\right)$$

将坐标系平移，使 $P(x_0,y_0)$ 为原点，则新坐标系中椭圆的方程应改写为

$$E:\frac{(x+x_0)^2}{a^2}+\frac{(y+y_0)^2}{b^2}=1$$

设新坐标系中直线 AB 的直线方程为

$$mx+ny=1$$

带入得

$$b^2(x+x_0)^2+a^2(y+y_0)^2=a^2b^2$$
$$b^2x^2+a^2y^2+2(b^2x_0x+a^2y_0y)=a^2b^2-b^2x_0^2-a^2y_0^2$$

注意到 (x_0,y_0) 在曲线上，则

$$b^2x^2+a^2y^2+2(b^2x_0x+a^2y_0y)(mx+ny)=0 \tag{1}$$

我们只取 x^2 与 y^2 的系数比

$$\lambda=\frac{b^2+2b^2x_0m}{a^2+2a^2y_0n}$$
$$\lambda(a^2+2a^2y_0n)=b^2+2b^2x_0m$$

带回直线 AB 的方程，消去 m，即

$$\frac{(\lambda(a^2+2a^2y_0n)-b^2)}{2b^2x_0}x+ny=1$$

由于 n 也是参数，因此将 n 也整理出来，得到

$$n\left(\frac{2\lambda a^2y_0nx}{2b^2x_0}+y\right)+\frac{\lambda a^2-b^2}{2b^2x_0}x=1$$

将 n 看作主元，那么要是上式恒成立即有

$$\begin{cases}\frac{2\lambda a^2y_0x'}{2b^2x_0}+y' &=0\\ \frac{\lambda a^2-b^2}{2b^2x_0}x' &=1\end{cases}$$

故直线 AB 过定点

$$\begin{cases}x' &=\frac{2b^2x_0}{\lambda a^2-b^2}\\ y' &=-\frac{2\lambda a^2y_0}{\lambda a^2-b^2}\end{cases}$$

最后平移回原坐标即是

$$\begin{cases}x &=x'+x_0=\frac{b^2+\lambda a^2}{\lambda a^2-b^2}x_0\\ y &=y'+y_0=\frac{b^2+\lambda a^2}{b^2-\lambda a^2}y_0\end{cases}$$

□

再来对于斜率之积为定值，提取 ① 中的 xy 项系数与 y^2 系数比值得

$$\mu=-\frac{2b^2x_0n+2a^2y_0m}{a^2+2a^2y_0n}$$
$$-(a^2+2a^2y_0n)\mu=2b^2x_0n+2a^2y_0m$$

带回直线方程，消去 m 得

$$-\frac{(a^2+2a^2y_0n)\mu+2b^2x_0n}{2a^2y_0}x+ny=1$$

将参数 n 提出来得到

$$n\left(\frac{2a^2y_0\mu+2b^2x_0}{2a^2y_0}x+y\right)+\frac{a^2\mu}{2a^2y_0}x=1$$

联立可得

$$\begin{cases}\frac{2a^2y_0\mu+2b^2x_0}{2a^2y_0}x'+y' &=0\\ \frac{a^2\mu}{2a^2y_0}x' &=1\end{cases}$$

故直线 AB 过定点

$$\begin{cases}x' &=\frac{2y_0}{\mu}\\ y' &=-\frac{2a^2y_0\mu+2b^2x_0}{a^2\mu}\end{cases}$$

最后平移回原坐标即是

$$\begin{cases}x &=x'+x_0=x_0+\frac{2y_0}{\mu}\\ y &=y'+y_0=-y_0-\frac{2b^2x_0}{a^2\mu}\end{cases}$$

□

§28.8　定值问题

定值问题往往和其他问题联动，如定点问题，考察的类型主要有：面积、斜率、内积、面积等；定值问题主要是翻译这些定值的条件，与定点问题不同的是，定值问题往往用一般方法不好猜出来，而且也不需要去猜，将条件翻译好后用韦达定理解决，这就引出下面的问题：

§28.8.1　非对称韦达定理

当面积、斜率、内积、面积等条件翻译完成以后，有时不能直接使用韦达定理，如设，联立圆锥曲线与直线方程得等到的一元二次方程的两个根为 x_1, x_2，使用韦达定理的条件是翻译后得到的式子中出现 x_1x_2 或 x_1+x_2 的对称式，但如果翻译后的式子没有这种对称式就无法直接使用韦达定理了；

这种没法直接使用韦达定理的情况被称为**“非对称韦达定理”**，虽然名字中有定理二字，但却不是定理，是一种题目类型，以下整理一些解决非对称韦达定理的思路，我用一道例题总结一些通用方法：

e.g.　已知椭圆 $E:\frac{x^2}{4}+y^2=1$，左右顶点分别为 A, B，过点 $P(1,0)$ 做直线 l 与椭圆交于 C, D，设直线 AD, BC 的斜率分别为 k_1, k_2，证明 $\frac{k_1}{k_2}$ 为定值。

和积互换

找到对称式 x_1x_2 与 x_1+x_2 的关系，互相转换。

设点 $C(x_1,y_1), D(x_2,y_2)$，易知 $A(-2,0), B(2,0)$ 则有：

$$k_1=\frac{y_1}{x_1+2}$$
$$k_2=\frac{y_2}{x_2-2}$$

因此

$$\frac{k_1}{k_2}=\frac{y_1(x_2-2)}{y_2(x_1+2)} \quad ①$$

设 $l: x = my + 1$，则 ① 可以表示为

$$\frac{k_1}{k_2} = \frac{my_1y_2 - y_1}{my_1y_2 + 3y_2} \qquad ②$$

将直线 l 与椭圆 E 的方程联立即可得到

$$(m^2+4)y^2 + 2my - 3 = 0$$

通过韦达定理知道

$$\begin{aligned} y_1 + y_2 &= -\frac{2m}{m^2+4} \\ y_1y_2 &= -\frac{3}{m^2+4} \end{aligned}$$

显然 ② 式中只有部分能使用韦达定理消去，无法完全消去 y_1, y_2，但注意到 y_1y_2 与 $y_1 + y_2$ 的关系：

$$y_1y_2 = \frac{3(y_1+y_2)}{2m}$$

将这个关系式带入 ② 中，此步称为**和积互换**

$$\begin{aligned} \frac{k_1}{k_2} &= \frac{3(y_1+y_2) - 2y_1}{3(y_1+y_2) + 6y_2} \\ &= \frac{y_1 + 3y_2}{3y_1 + 9y_2} \\ &= \frac{1}{3} \end{aligned}$$

即得定值。

以曲代直

在上一种方法中，① $\to$ ② 的过程中我们用直线方程消去 x_1, x_2，而以曲代直就是用曲线方程转化为对称结构。

延续上例，注意到椭圆 E 的方程

$$\begin{aligned} \frac{x^2}{4} + y^2 &= 1 \\ y^2 &= 4 - x^2 \\ y^2 &= (2-x)(2+x) \end{aligned}$$

因此 ① 可表示为：

$$\begin{aligned}\frac{k_1}{k_2}&=\frac{y_1y_1(x_2-2)}{y_1y_2(x_1+2)}\\&=\frac{(2-x_1)(x_2-2)}{y_1y_2}\\&=\frac{2(x_1+x_2)-x_1x_2-4}{y_1y_2}\end{aligned} \tag{3}$$

③ 对称结构，因此可以用韦达定理。

硬解定理

硬解定理不过是好听的名字，说白了就是爆算，用求根公式把 y_1,y_2 表示出来之后带入 ①

延续上例，可得

$$\begin{aligned}y_1&=\frac{-m+2\sqrt{m^2+3}}{m^2+4}\\y_2&=\frac{-m-2\sqrt{m^2+3}}{m^2+4}\end{aligned}$$

带入 ② 可得：

$$\begin{aligned}\frac{k_1}{k_2}&=\frac{-3m+m-2\sqrt{m^2+3}}{m^2+4}\cdot\frac{m^2+4}{-3m-3m-6\sqrt{m^2+3}}\\&=\frac{-2m-2\sqrt{m^2+3}}{-6m-6\sqrt{m^2+3}}\\&=\frac{1}{3}\end{aligned}$$

综上得到定值。

硬解定理虽然是最为通用、适用面最广的方法，但考虑到计算量巨大，时间成本以及试错成本高，这种方法不推荐在考试中用。

以上为解决非对称韦达定理情况的几个常用的方法，但当我们思考一个问题，“非对称”的结构与“对称”结构同时出现在一个式子中，例如 ②

$$\frac{my_1y_2-y_1}{my_1y_2+3y_2}$$

其中对称结构为 my_1y_2，非对称结构为 $-y_1,3y_2$，它们的比值我们已经知道为 $\frac{1}{3}$；实际上有以下定理：

定理

若 $\frac{A}{B}$ 是 x_1,x_2 的二元对称式，且 x_1,x_2 是一元二次方程的两个互异的实根，且 $\frac{A+Cx_1}{B+Dx_1},\frac{A+Cx_1}{B+Dx_2}$ 均为定值，则有

$$\frac{A+Cx_1}{B+Dx_1}=\frac{C}{D}$$

以及

$$\frac{A+Cx_1}{B+Dx_2}=-\frac{C}{D}$$

这个“定理”我最早是在作者“申申本质数学与通法”那看到的，我认为这个定理极大程度化简了计算的过程，像是以上例题，虽然我用三个方法都写出来了，但在更多题目中，要么是无法直接转换成对称结构，要么是计算量过大，通过此定理可以更深刻的理解非对称韦达定理。

证明：先证明

$$\frac{A+Cx_1}{B+Dx_1}=\frac{C}{D}$$

由于 x_1,x_2 是一元二次方程的两根，因此根据求根公式即可写出：

$$x_1=\frac{-b+\sqrt{\Delta}}{2a}$$
$$x_2=\frac{-b-\sqrt{\Delta}}{2a}$$

注意到 $\sqrt{\Delta}$ 为根式，在多项式中不可约，因此若 $\frac{A+Cx_1}{B+Dx_1}$ 可以约分，那么 $\sqrt{\Delta}$ 的系数比即为那个定值，即

$$\begin{aligned}\frac{A+Cx_1}{B+Dx_1}&=\frac{2aA-bC+C\sqrt{\Delta}}{2aB-bD+D\sqrt{\Delta}}\\&=\frac{C}{D}\end{aligned}$$

再来，对于

$$\frac{A+Cx_1}{B+Dx_2}$$

带入 x_1,x_2 可得：

$$\begin{aligned}\frac{A+Cx_1}{B+Dx_1}&=\frac{2aA-bC+C\sqrt{\Delta}}{2aB-bD-D\sqrt{\Delta}}\\&=-\frac{C}{D}\end{aligned}$$

□

e.g. 已知

$$\frac{kx_1x_2+kx_2+x_1+1}{kx_1x_2-kx_1+x_2-1}$$

为定值，x_1,x_2 为一元二次方程的两个根，求该定值；

首先找到对称结构，这里配凑一下即得：

$$\frac{kx_1x_2+x_1+x_2+1+(k-1)x_2}{kx_1x_2+x_1+x_2-1-(k+1)x_1}$$

注意到对称结构为：

$$\frac{kx_1x_2+x_1+x_2+1}{kx_1x_2+x_1+x_2-1}$$

这里需要注意的是常数项并不影响对称，读者可以通过将 x_1,x_2 的位置互换来检查是否为对称结构（对称式），而非对称结构为：

$$\frac{(k-1)x_2}{-(k+1)x_1}$$

根据定理可知，这个定值为：

$$\frac{k-1}{k+1}$$

第七部分

组合学

第二十九章　排列与组合

计数原理属于组合数学里的东西，组合数学的题就是我们通常所说的“杂题”，在数学竞赛中有些甚至与数论齐名的难度，这种题高考一般不会遇到，不过这个计数原理在我们生活中是非常有用的，我觉得是为数不多对我们生活起直接作用的数学方法；可以将排列组合应用于生物遗传计算、概率论等等，更方便的研究概率的东西。

§29.1　计数的基本原则

分类加法计数原理：完成一件事有两类不同方案，在第一类方案中有 m 种不同的方法，在第二类方案中又 n 种不同的方法，那么完成这件事共有 $N=m+n$ 种不同的方法。

可以推广为:完成一件事有 n 类不同的方案,在第 i 类有 m_i 种不同的方法,则完成这件事共有 $N=\sum_{i=1}^{n}m_i$ 种不同的方法。

分步乘法计数原理：完成一件事需要两个步骤，做第一步有 n 不同的方法，做第二步有 m 种不同的方法，则完成这件事共有 $N=m\cdot n$ 种不同的方法。

同样可以推广：完成一件事需要 n 个步骤，做第 i 步有 m_i 不同的方法，则完成这件事共有 $N=\prod_{i=1}^{n}m_i$ 种不同的方法。

要分清楚“分类”和“分步骤”完成的区别，在它们结合在一起时先分类在明确步骤。

§29.2　排列

排列

排列的定义：从 n 个不同的元素中取出 $m(m\leqslant n)$ 个元素，按照一定的顺序排成一列，叫做从 n 个不同元素中取出 m 个元素的一个排列。

排列数：从 n 个不同元素中取出 $m(m\leqslant n)$ 个元素的所有不同排列，这个的个数叫做从 n 个不同元素中取出 m 个元素的排列数，用符号 A_n^m 表示。

排列数的一些公式：

排列数公式：$A_n^m=n(n-1)(n-2)\dots(n-m+1),n,m\in\mathbb{N}^+$，并且 $m\leqslant n$，排列数的公式可以写成：$A_n^m=\frac{n!}{(n-m)!}$

全排列：将 n 个不同元素每一个都抽出来的一个排列叫全排列，这时全排列的排列数为 $A_n^n=n!$

组合

组合的定义：一般地，从 n 个不同元素中取出 $m(m\leqslant n)$ 个元素合成一组，叫从 n 个不同元素中取出 m 个元素的一个组合。组合问题与元素的顺序无关。

组合数：从 n 个不同元素中取出 $m(m\leqslant n)$ 个元素的所有不同组合，这个的个数叫做从 n 个不同元素中

取出 m 个元素的组合数，用符号 C_n^m 表示，也有用 $\binom{n}{m}$ 表示的，注意上下顺序与前者相反。

组合数公式：

$$C_n^m = \frac{A_n^m}{A_m^m} = \frac{n(n-1)(n-2)\dots(n-m+1)}{m!}$$

还可以写成：$C_n^m = \frac{n!}{m!(n-m)!}$，规定 $C_n^0 = 1$。

组合数具有性质：

1. $C_n^m = C_n^{n-m}$
2. $C_{n+1}^m = C_n^m + C_n^{m-1}$

可以看出 $A_n^m = C_n^m \cdot A_m^m$，对于组合数我们可以简记，如 $C_{100}^3 = \frac{100\cdot99\cdot98}{1\cdot2\cdot3}$，把分式下面部分给去掉就是排列数了。

§29.3 二项式定理

$$(a+b)^n = C_n^0 a^n + C_n^1 a^{n-1} b + \dots + C_n^n b^n, n \in \mathbb{N}^+$$

简记为：

$$\prod_{i=1}^{n}(a+b) = \sum_{i=0}^{n} C_n^i a^{n-i} b^i$$

左边我不用指数用求连乘积是为了更直观的体现，左边有 n 项，而右边有 $n+1$ 项。

其中 C_n^i 叫做二项式系数。

通项：二项展开式中第 $k+1$ 项 $T_{k+1} = C_n^k a^{n-k} b^k, (0 \leqslant k \leqslant n, k \in \mathtt{N}^+)$ 叫做二项展开式的通项。

二项式系数的性质

1. **对称性**：与首末两端“等距离”的两个二项式系数相等，即 $C_n^k = C_n^{n-k}$。
2. 增减性与最大值：当 $k < \frac{n+1}{2}$ 时二项式系数是逐渐增大的，由对称性可知后面逐渐减小，最大值在中间；
   (a) 当 n 是偶数时，中间的一项 $C_n^{\frac{n}{2}}$ 为最大值。
   (b) 当 n 为奇数时，中间的两项 $C_n^{\frac{n-1}{2}}, C_n^{\frac{n+1}{2}}$ 相等且同时为最大值。
3. 各二项式系数的和，推导过程：

$$(1+x)^n = C_n^0 + C_n^1 x + \dots + C_n^n x^n$$

   令 $x=1$ 就得到：

$$\sum_{i=0}^{n} C_n^i = 2^n$$

§29.4 补充

§29.4.1 间接法

间接法相当于“求反”的思想，也叫正难则反，意思是如果直接去计算很麻烦的问题，可以试着求它的反；用集合的思想比较容易理解，如我们要求的情况属于集合 A，又不好直接求 A，可以用全集减去 $\overline{A}$，如下图：

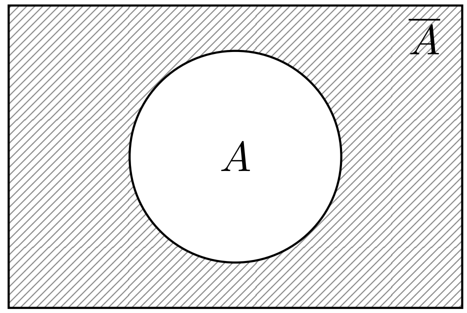


图 29.1: 间接法 1

对于单条件限制的问题，没什么重复的情况，不容易出错；但对于多条件限制的问题，就需要考虑到重复减去的部分，用集合的观点也很好理解，下面举一个例子：六个人排成一排，甲不能站最左边，乙不能站最右边，请问一共有多少种排法？

这个问题有两个条件限制，分别是“甲不能站最左边”、“乙不能站最右边”，我们用间接法来解决此问题，将所有排列设为全集 Ω（无限制条件），不符合第一个限制条件的排列组成集合 A, 不符合第二个限制条件的排列组成集合 B，这里应注意，存在既不符合条件一，也不符合条件二的排列，也就是说 A、B 有交集，即“甲站最左边且乙站最右边”的排列，我们要求的排列种数应为

$$|\Omega| - |A \bigcup B|$$

一开始不熟悉时容易写成

$$|\Omega| - |A| - |B|$$

注意到有被重复减去的部分 $|A \bigcap B|$，所以应加上，即

$$\begin{aligned}&|\Omega| - |A| - |B| + |A \bigcap B|\\=&6! - 5! - 5! + 4!\\=&504\end{aligned}$$

用下图可更好的理解：

全集 Ω

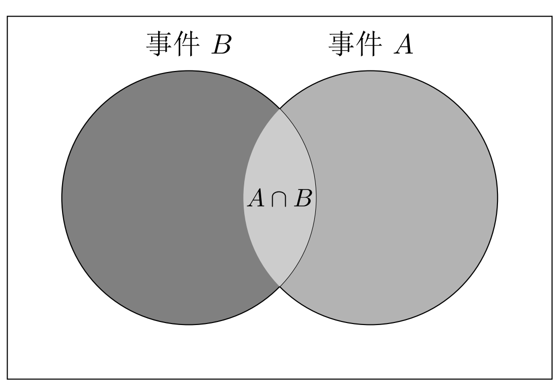


图 29.2: 间接法 2

当有更多条件限制时，应注意交集部分，用集合的思想去理解它，就不容易漏差情况。

§29.4.2 捆绑法、插空法、隔板法

一、捆绑法

当遇到<u>**相邻问题**</u>时，可以把相邻的元素看作一个整体，再进行排列，这里需要注意的是把相邻元素看作整体后，元素“层内”之间也有不同排序，这是容易漏掉的，下面结合一个例题来理解：

e.g. 五个人排成一排，其中甲乙必须相邻，问有多少种排法？

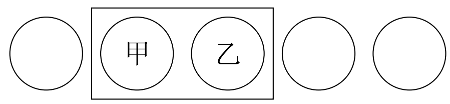


图 29.3: 捆绑法

如果把甲乙看作一个整体则有 4! 种排法，但考虑到甲乙“层内”的排序，结果应是 $2!\cdot 4!$。

由上题可知，要点就是第一步把“层外”的排序完后，第二步考虑“层内”的排列情况。

二、插空法

当遇到<u>**不相邻问题**</u>时，除了用“捆绑 + 间接”以外常用插空法，要点是第一步把无特殊需求的元素进行排列，第二步将不能相邻的元素插入空，如有六个空插入两个人即 A_6^2。

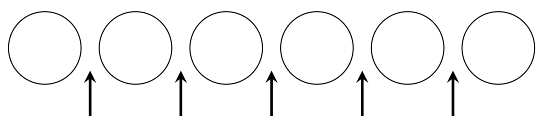


图 29.4: 插空法

当然插空法不止可以解决不相邻问题；当一组元素有一定顺序时，将其他元素插入，需要注意插入元素的“相邻问题”，如果可以插入的元素可以相邻，那么空会越插越多，以及注意第二步中元素是否相同，要用组合亦或排列。

其实不难看出当插入元素不相邻时就是“不相邻问题”，所以这是此问题更一般的情况。

三、隔板法

有一类问题是：将相同的元素放到不同集合；这时大部分情况可用隔板法。

隔板法就是这些相同的元素排成一列，然后用相同的板子隔开，使得板与板之间的元素归于一个集合，如下图：

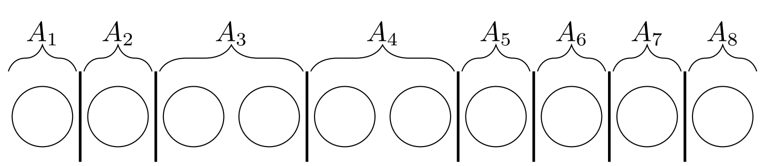


图 29.5: 隔板法

这样，我们就可以解决一类问题“将 m 个相同元素，放到 n 个集合 $x_i, i=1,2,\cdots,n$ 中，其中 $n\leqslant m$，一共有多少种放法”，我们将这类问题再细分一下，分为两种情况：

1. 每个集合中至少有一个元素
2. 允许有集合是空集

先来讨论第一种情况，对于第一种情况，将 m 个元素排成一列，由于它们无异，所以并不存在有无顺序的差别，这 m 个元素除去两端的空，元素之间夹着的空一共有 $m-1$，又因为要分成 n 个集合（也就是 n 个组），所以需要插入 $n-1$ 个板子；
由插空法可知，我们需要在 $m-1$ 个空中选 $n-1$ 个空，且板子也无异，所以一共有 C_{m-1}^{n-1} 种分法[①]。

e.g. “将十个相同的球分到七个班里，问有几种分发？”
利用隔板法不难知道，一共分法为：

$$C_{10-1}^{7-1} = C_9^6 = C_9^3$$

我们可以用这个结论来解决不定方程 $x_1 + x_2 + x_3 + \cdots + x_n = r(r \geqslant n)$ 的正整数解的个数。这是一个隔板法一般化很重要一步，如果把 r 看作 r 个小球排成一排，那么 x_i 可以看成有不定个数小球的集合，例如：

$$5 + 5 + 3 + 2 = 15$$

结合隔板法，对于不定方程 $x_1 + x_2 + x_3 + \cdots + x_n = r(r \geqslant n)$ 的正整数解的个数有: C_{r-1}^{n-1}

现在讨论第二种情况，也就是问题不变，但允许有空集，如果用分类的方法就会麻烦很多，那我们不妨先给每个集合各一个元素（元素相同），这样一来不管怎么分配都不会出现空集的情况，且又回到我们熟悉的情况一中了。

e.g. 求“不定方程 $x_1 + x_2 + x_3 + \cdots + x_n = r(r \geqslant n)$ 的非负整数解的个数”，给每个集合加一个“球”，就会为：

$$x_1 + x_2 + x_3 + \cdots + x_n = r$$
$$(x_1 + 1) + (x_2 + 1) + (x_3 + 1) + \cdots + (x_n + 1) = r + n$$

故一共有 $C_{r+n-1}^{n-1} = C_{r+n-1}^{r}$ 种非负整数解。

我们可以把解决这个不定方程的方法和思想类比为现实问题，就如 m 个小球放入 n 个集合中，允许又空集的情况下问有多少种放法。

§29.4.3 定序问题

像是“六个人排成一排，乙必须在甲左边”，让它更一般化“n 个元素排列，其中 m 个元素顺序确定”，这类问题叫做定序问题；
对于定序问题常用的处理方法叫做“缩倍法”，下面举一个例子：六个人排成一排，其中甲必须在乙左边，问有多少种排法？

我们知道，无视任何规则，六个人一共有 A_6^6 种排法，也就是 6! 种排法，但甲必须在乙的左边，所以进行缩倍处理除以 A_2^2

$$\frac{A_6^6}{A_2^2} = \frac{6!}{2!} = 360$$

也就是一共有 360 种排法。

我认为缩倍法的本质就是“排列数”到“组合数”的过程，想一想，“定序”无非是只取排列的一种情况，就如上例中甲和乙的关系，甲乙两个元素全排列为 A_2^2，无非是“甲，乙”和“乙，甲”两种排列，而我们只需要其中的一种，所以对于六个人的全排列除以 A_2^2，取其中“甲，乙”的排列。

[①]注意看这里是 $\binom{m-1}{n-1}$

如果你理解了上述的例子，那我把一般情况的定序问题改一下，变成“n 个元素排列，其中 n 个元素顺序确定”，这不就只有一种情况吗，如果再改一下，变成“从 N 个元素里取 n 个元素，其中 n 个元素顺序一定”，可以看出上述两个例子写成数学公式就是：

$$\frac{A_n^n}{A_n^n} = 1$$
$$\frac{A_N^n}{A_n^n} = C_N^n$$

这两个例子可以看出定序问题，与排列数推导到组合数之间的联系。

e.g. 如果六个人排队，求下列情况的排列种数：

1. 甲在乙左边，丙在乙左边。
2. 甲在乙左边，丙在甲左边。

首先六个人全排列一共有 A_6^6 种排法，而这两个问题就是甲乙丙三人的定序问题；

对于 1、甲乙丙间一共有 A_3^3 种排法，而符合条件的只有一种，即“甲，乙，丙”，所以一共有 $\frac{A_6^6}{A_3^3}$ 种排法。

对于 2、符合条件的有“甲，乙，丙”、“甲，丙，乙”这两种排法，所以我们需要的是 A_3^3 中的其中两种，也就是一共有 $2\frac{A_6^6}{A_3^3}$ 种排法。

如果有多重定序，比如说“六个人排队，甲在乙左边，丙在丁左边”这种问题，只需分步缩倍就好，即先算甲在乙左边的情况，得到的情况再算丙在丁左边的情况，最后得到 $\frac{A_6^6}{A_2^2 \cdot A_2^2}$

§29.4.4 分组分堆

分组分堆问题主要考察到对于排列组合的理解，当将 n 个不同的元素分到 m 个组中（$m \geqslant n$），当这 m 个组有区别时，即可以表示成 $m_1, m_2, \ldots, m_m$ 个组，那么对每个组分配元素的问题可以直接使用分步乘法计数原理，第一步将 a 个元素分配到 m_1 组中，第二部将 b 个元素分配到 m_2 组中，直到最后将 n 个元素分配到这 m 个组中即可。

当 m 个组无区别时，即“分成 m 组、分成 m 堆”的情况时，要特别注意“平均分组”的问题，平均分组意味着有 a 个组中存在同样的元素数，例如“将 7 本书分成三堆，一堆三本，一堆两本，一堆两本”，由于组与组之间无异，所以需要进行“消序”处理。

例如：将 23 本不同的书分成六堆，其中一堆七本、两堆各五本、三堆各两本，问一共有多少种分法。
为了方便理解，我们先将这六堆分别编上号，即为 $m_1, m_2, \cdots, m_6$，令 m_1 堆中放入 7 本，m_2, m_3 堆中放入都是 5 本，m_4, m_5, m_6 堆中放入的是 2 本。这时只需用分步乘法计数原理即可算出一共有多少种分配方法：

$$C_{23}^7 \cdot C_{16}^5 C_{11}^5 \cdot C_6^2 C_4^2 C_2^2$$

由于原题中这六堆无异[2]，所以进行消序处理：

$$C_{23}^7 \cdot \frac{C_{16}^5 C_{11}^5}{A_2^2} \cdot \frac{C_6^2 C_4^2 C_2^2}{A_3^3}$$

[2]这里说的是堆与堆之间无异，而不是每堆放的书的个数无异

请理解上述例题，其实就是将排列转换为组合的过程，包括此类题型的衍生“先分后派”问题，不过是将组合又进行排列了。

§29.4.5 圆排列

定理

设有 n 个元素圆排列，则一共有 $(n-1)!$ 种排列方法。

直观的来理解，若有 n 个人手拉手围成一个圈，则当某相邻的两个人放手时，就变成了直线排列，同一圆排列相邻两人放手的方式有 n 种（可以结合多边形的欧拉公式来理解），因此 n 个元素圆排列一共有

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

$(n-1)!$ 种排列方法。

$\square$

§29.4.6 杨辉三角

第零行						1						
第一行					1		1					
第二行				1		2		1				
第三行			1		3		3		1			
第四行		1		4		6		4		1		
$\cdots$	$\cdots$	$\cdots$		$\cdots$		$\cdots$		$\cdots$		$\cdots$		
第 n 行	C_n^0	C_n^1		$\cdots$		$\cdots$		$\cdots$		C_n^{n-1}		C_n^n

图 29.6: 杨辉三角

上图为杨辉三角，通过杨辉三角我们可以直观的得到一些组合数的性质；

不难看出，每个数字都是由该数字上一行的相邻两个数加起来得到，如果我们从左至右编号，可以得到第 N 行第 n 项是由第 $N-1$ 行的第 n 项加上 $n-1$ 项得到的，倘若第 N 行第 n 项没有数字，则将该项视为 0，例如第三行第 3 项是由 $1+0$ 得到的。

将行数推广到第 n 行后，我们发现这一行有 $n+1$ 项，第 i 项 $(i=1,2,\ldots,n)$ 可以由 C_n^i 表示，简单的解释是因为 $C_n^m = C_{n-1}^m + C_{n-1}^{m-1}$，满足刚刚讲得“每个数字都是由该数字上一行的相邻两个数加起来得到的”这一条件，所以杨辉三角和组合数是“等价”关系。

§29.4.7 染色问题

染色问题最常用的通法就是推理、演绎、分类讨论，或者说“模拟过程，跳个分类”，下面我结合一道例题来诠释通法的过程：

e.g. 给图中 A,B,C,D,E,F 六块区域染色，每个区域染一种总颜色，且相邻区域不同色，有四种颜色可选，一共有多少种染色方法？

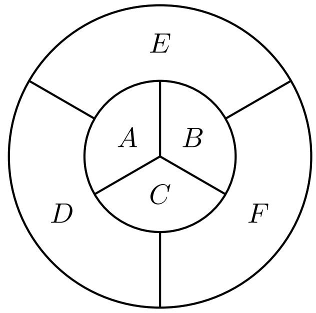


图 29.7: 一例染色问题

首先模拟过程，即演绎，我们先给 E 区域染色，一共有 4 种颜色可以选，与 E 区域相邻的区域有 4 个，都会受到影响，因此我们进行跳格分类；

本题中只有 C 区域不与 E 区域相邻，因此我们对此进行分类：

1. C 区域与 E 区域同色。

2. C 区域与 E 区域不同色。

对于情况 1，其余区域都与 C 区域相邻，因此随意对一个区域染色，它有 3 种染法，例如对 A 区域有 3 种染法，继续分析，与 A 跳格的 F 区域没有染色，因此需要分类：

1. A 区域与 F 区域同色。

2. A 区域与 F 区域不同色。

对于情况 1.1，区域 B 相邻区域都已经确定，故有两种染法，D 区域也有两种染法；因此对于情况 1.1 共有 $4 \times 3 \times 2 \times 2 = 48$ 种；

对于情况 1.2，F 有两种染法，B, D 都区域只有一种染法，所以有 $4 \times 3 \times 2 = 24$ 种；

对于情况 2，对 A 区域有两种染法，与上面的做法一样，对它跳格分类：

1. A 区域与 F 区域同色。

2. A 区域与 F 区域不同色。

对于情况 2.1，B, D 区域都只有一种染法，所以有 $4 \times 3 \times 2 = 24$ 种；

对于情况 2.2，B, D 区域相邻部分都有四种颜色，因此无法染色，所以不可能染色。

综上，一共有

$$48 + 24 + 24 = 96$$

种染法。

以上的通法可以解决高中阶段绝大部分的染色问题，当一个点（区域）模拟过程中不好分类时，换一个点去讨论可能会更方便。

下面讨论一种特殊的染色情况：对 n 个首尾相连的区域染色，每个区域只能染一种颜色，相邻区域必须不同色，有 m 种颜色可以染，那么一共有多少种方案？

实际上，以上命题可以等价为：m 个人传球，第 $n+1$ 次传球时球回到第一次传球的人的脚下。

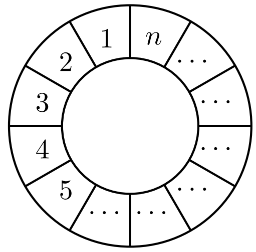


图 29.8: 环形染色

我们构造一个数列使得第 n 个区域的颜色有 a_n 中染法；易知第一个区域有 m 种染法，第二个区域就有 $m-1$ 种染法，因此有：

$$a_n = m(m-1)^{n-1}$$

不过稍加思考我们会发现，以上式子没有考虑到第 n 个区域与第一个区域同色的情况，为了避免同色的情况，我们应该考虑相邻两项的染色方法，也即：$a_n + a_{n-1}$，易知第一个区域与第二个区域有 $m(m-1)$ 种染法，第二个区域与第三个区域有 $m(m-1)^2$ 种染法，易推：

$$a_n + a_{n-1} = m(m-1)^{n-1}$$

这样，当取到第 n 以及 $n+1$ 个区域时就考虑到首尾同色的情况了，下面我们对上面的递推式求通项：

$$\begin{aligned} a_n &= -a_{n-1} + m(m-1)^{n-1} \\ \frac{a_n}{(-1)^n} &= \frac{a_{n-1}}{(-1)^{n-1}} - m(1-m)^{n-1} \end{aligned}$$

这里构造了一个可以用累加法的数列，令 $b_n = \frac{a_n}{(-1)^n}$，则有

$$\begin{aligned} b_n - b_{n-1} &= -m(1-m)^{n-1} \\ \sum_{i=3}^{n}(b_n - b_{n-1}) &= (1-m)^n - (m-1)^2 \\ b_n &= b_2 + (1-m)^n - (m-1)^2 \\ b_n &= (1-m)^n + m - 1 \\ a_n &= (m-1)^n + (m-1)(-1)^n \end{aligned}$$

综上得到染法种数。

§29.4.8 范德蒙德卷积

定理

范德蒙德卷积公式

$$\sum_{i=1}^{k}\binom{n}{i}\binom{m}{k-i} = \binom{n+m}{k}$$

♠

先从直观示例理解这个公式，在一个大小为 $m+n$ 的集合中选 k 个数，等价于把这个集合拆分成两个集

合，大小分别为 m,n，从 n 中取出 i 个数，再从 m 中取 $k-i$ 个数的方案个数。

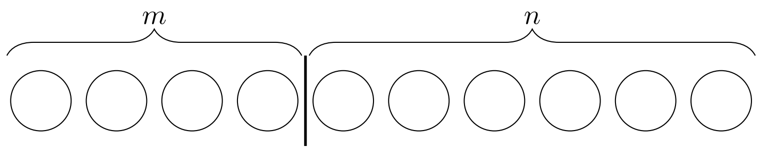


图 29.9: 范德蒙德卷积

当然可以用代数来证明：

$$\begin{aligned}\sum_{k=0}^{m+n}\binom{n+m}{k}x^k &= (1+x)^{n+m}\\ &= (1+x)^n(1+x)^m\\ &= \sum_{r=0}^{n}\binom{n}{r}x^n\sum_{s=0}^{m}\binom{m}{s}x^m\\ &= \sum_{k=0}^{m+n}\sum_{r=0}^{n}\binom{n}{r}\binom{m}{k-r}x^k\end{aligned}$$

即有：

$$\sum_{i=0}^{k}\binom{n}{i}\binom{m}{k-i}=\binom{n+m}{k}$$

$\square$

附录 A　附录

§A.1　行列式

在解方程中常常会遇到一些比较复杂的方程式，在本科会学习到关于行列式的内容，这里简单介绍一下，本书中主要是用于简化式子，定义来自 [5]。

首先了解一下 n 元排列

n 个不同的正整数的一个全排列称为一个**n 元排列**，如，正整数 $1,2,3$ 的三元排列有

$$123,132,213,231,312,321$$

给定 n 个不同正整数，它们形成的全排列有 $n!$ 个。

4 元排列 2341 中，2 与 3 形成的数对 23，小的数在前，大的数在后，此时称这一对数构成严格**顺序**；而 2 与 1 形成的数对 21，大的数在前，小的数在后，此时称这一对数构成一个**逆序**。排列 2341 中，构成逆序的数对有 21,31,41 共三对，此时称排列 2341 的**逆序数**是 3，记作 $\tau(2341)=3$。

在 n 元排列 $a_1a_2\cdots a_n$ 中，从左到右任取一对数 $a_ia_j(i<j)$，如果 $a_i<a_j$，那么称这一对数构成一个**顺序**；如果 $a_i<a_j$，那么称这一对数构成一个**逆序**。一个 n 元排列中逆序的总数称为**逆序数**，记作 $\tau(a_1a_2\cdots a_n)$

4 元排列 2143 中，构成逆序的数对有 21,43，共两对，于是

$$\tau(2143)=2$$

逆序数为奇数的排列称为**奇排列**，逆序数为偶数的排列称为**偶排列**。

把排列 2341 的 3 和 1 互换位置，其余数不动，便得到排列 2143，像这样的变换称为一个**对换**，记作 $(3,1)$，对换的概念也适用于 n 元排列。

定理

对换改变 n 元排列的奇偶性 ♠

2 阶行列式

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22}-a_{12}a_{21}$$

2 阶行列式是 $2(=2!)$ 项的代数和，其中每一项是位于不同行、不同列的两个元素的乘积，把这两个元素按照行指标成自然序排好，其列指标所成排列是偶排列时，该项带正号，成奇排列时该项带负号，于是 2 阶行列式为

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \sum_{j_1j_2} (-1)^{\tau(j_1j_2)} a_{1j_1}a_{2j_2}$$

从 2 阶行列式的定义得到启发，给出 n 阶行列式的定义如下：

定义

n 级矩阵 $\boldsymbol{A}=(a_{ij})$ 的行列式（简称为n **阶行列式**）

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

是 $n!$ 项的代数和，其中每一项的都是位于不同行、不同列的 n 个元素的乘积，把这 n 个元素以行指标为自然顺序排好位置，当列指标构成的排是偶排列时，该项带正号；是奇排列是，该项带正号，即

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \sum_{j_1j_2\cdots j_n} (-1)^{\tau(j_1j_2\cdots j_n)} a_{1j_1}a_{2j_2}\cdots a_{nj_n}$$

其中 $j_1j_2\cdots j_n$ 是 n 元排列，$\sum_{j_1j_2\cdots j_n}$ 表示对所有 n 元排列求和。上式称为 n 阶行列式的**完全展开式**。

n 阶行列式也记作 $|\boldsymbol{A}|$ 或者 $\det \boldsymbol{A}$。

由定义立即得到：

一阶行列式 $|a|=a$

由于三元排列 123,231,312 是偶排列，321,213,132 是奇排列，因此三阶行列式

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

三阶行列式的六项及其所带符号可以采用下图记忆：

其中主对角线上的三个元素的乘积 $a_{11}a_{22}a_{33}$,以及与主对角线平行的线上三个元素的乘积 $a_{31}a_{23}a_{31},a_{13}a_{21}a_{32}$ 都带正号;反对角线上三个元素的乘积 $a_{13}a_{22}a_{31}$,以及与反对角线平行线上的三个元素的乘积 $a_{12}a_{21}a_{33},a_{11}a_{23}a_{32}$ 都带负号。

§A.2　迭代递归

这种一般算比较偏的内容，高考和竞赛都有这方面的题，高中里更多是“类递归”就是间接复合自己的函数，但不会像计算机中有去有回的递归，更多是考察换元法的运用；而竞赛中多是考迭代函数性质，解决一些代数问题，都是很难的东西。

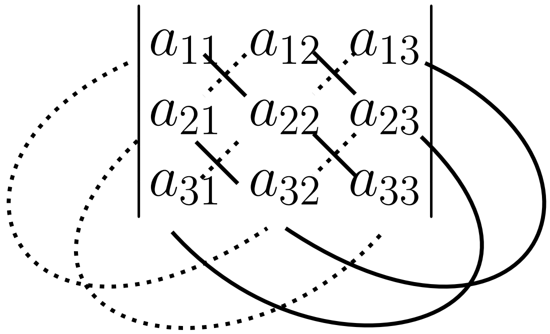


图 A.1: 三阶行列式

递归函数：递归函数更多是计算机方面的东西，它就是在执行一个函数时，会直接或间接调用到自己的函数，着重于“递推”和“回归”，有去有回，递归函数在重复执行直到一个终值时，会结束下一层的执行，然后回归。

1. 直接调用：如 $f(x)=f(x+1)+2$；

2. 间接调用：如 $f(x)=g(x+1)+x,g(x)=f(x+1)+2$。

例如在用计算机解决“拉马努金恒等式”时，对函数抽象出来进行建模得到一个递归函数，在设一个终值执行就可以求出答案。

$$
\begin{aligned}
&= \sqrt{1+2\sqrt{1+3\sqrt{1+4\sqrt{\cdots}}}} \\
&= \sqrt{1+(i+1)\sqrt{1+(i+2)\sqrt{1+(i+3)\sqrt{\cdots}}}},(i=1) \\
f(x) &= \sqrt{1+(x+1)f(x+1)},(x=1)
\end{aligned}
$$

迭代函数：重复的与自身进行复合的函数。

例如一个函数 $f(x)$ 的 $n(n\in\mathbb{N}^+)$ 次迭代，可以表示为 $f^{(n)}(x)$，有：

$$f^{(n)}(x)=\underbrace{f\left(f\left(f\left(\cdots f\left(x\right)\right)\right)\right)}_{\text{迭代 n 次}}$$

不难看出，迭代函数满足 $R_f\subseteq D_f$
对于任意 $q,p\in\mathbb{N}^+$，有：

$$f^{(q)}\circ f^{(p)}(x)=f^{(q+p)}(x)$$

如果 $f(x)$ 有反函数，$f(x)$ 的反函数为 $f^{(-1)}(x)$，对于 $f^{(0)}(x)$ 定义为 $f^{(0)}(x)=x$，这样我们可以将迭代次数 n 推广到任意整数中。

§A.3 参数方程

三角恒等式：

$$\sin^2\alpha+\cos^2\alpha=1$$
$$\sec^2\alpha-\tan^2\alpha=1$$
$$\csc^2\alpha-\cot^2\alpha=1$$

我们观察圆的方程 (圆心在原点)：

$$\begin{aligned} & x^2+y^2=r^2 \\ \Longrightarrow\ & \frac{x^2}{r^2}+\frac{y^2}{r^2}=1 \end{aligned}$$

如果我们令:$\sin^2 t=\frac{x^2}{r^2},\cos^2 t=\frac{y^2}{r^2}$，我们就可以得到关于 x,y 参数为 t 的参数方程：

$$\begin{cases} x=r\sin t \\ y=r\cos t \end{cases}$$

所以平面中点 $P(r\sin t,r\cos t)$ 的轨迹就是以原点为圆心半径为 r 的圆。

§A.4 向量方法判断二面角余弦值正负

设平面 α，平面 β 的法向量分别为 $\overrightarrow{n_1},\overrightarrow{n_2}$；存在两点 A,B 满足 $A\in\alpha,B\in\beta$。
如果通过 $\overrightarrow{n_1},\overrightarrow{n_2}$ 的余弦值来判断平面 α,β 的二面角大小，可能会导致判断错误，这是因为没选好法向量而导致的；在大多数情况下通过图示即可判断二面角是钝角或锐角 (二面角余弦值正负) ，但在一些特殊情况下，无法直接得到二面角是锐角还是钝角，就需要以下方法进行准确判断。

无论二面角是锐角或钝角，通过内积公式都是可以判断出来的，但一个平面的法向量有两种方向，“选取哪一组法向量”会影响我们判断，我们需要引入一个新的向量 $\overrightarrow{AB}$；对于任意一组法向量的选择，我们都可以列出来：

α 的法向量	β 的法向量	它们夹角的余弦值
$\overrightarrow{n_1}$	$\overrightarrow{n_2}$	$\cos\langle\overrightarrow{n_1},\overrightarrow{n_2}\rangle$
$\overrightarrow{n_1}$	$-\overrightarrow{n_2}$	$-\cos\langle\overrightarrow{n_1},\overrightarrow{n_2}\rangle$
$-\overrightarrow{n_1}$	$\overrightarrow{n_2}$	$-\cos\langle\overrightarrow{n_1},\overrightarrow{n_2}\rangle$
$-\overrightarrow{n_1}$	$-\overrightarrow{n_2}$	$\cos\langle\overrightarrow{n_1},\overrightarrow{n_2}\rangle$

这里夹角的余弦值是通过 $\cos\langle\overrightarrow{n_1},\overrightarrow{n_2}\rangle=\frac{\overrightarrow{n_1}\cdot\overrightarrow{n_2}}{|\overrightarrow{n_1}|\cdot|\overrightarrow{n_2}|}$ 算出来的。

α 的法向量	β 的法向量	α 的法向量与 $\overrightarrow{AB}$ 的内积	β 的法向量与 $\overrightarrow{AB}$ 的内积
$\overrightarrow{n_1}$	$\overrightarrow{n_2}$	$+$	$+$
$\overrightarrow{n_1}$	$-\overrightarrow{n_2}$	$+$	$-$
$-\overrightarrow{n_1}$	$\overrightarrow{n_2}$	$-$	$+$
$-\overrightarrow{n_1}$	$-\overrightarrow{n_2}$	$-$	$-$

这只是其中一种情况，但不同的情况只是在正负号上同时变化，通过这个图表可以看出规律，我们称 $\overrightarrow{AB}$ 为判断向量

1. 当两个法向量与“判断向量”内积呈同号时，那么二面角的余弦值就是 $\cos\langle\overrightarrow{n_1},\overrightarrow{n_2}\rangle$；
2. 当两个法向量与“判断向量”内积呈异号时，那么二面角的余弦值就是 $-\cos\langle\overrightarrow{n_1},\overrightarrow{n_2}\rangle$。

§A.5 韦达定理

在初中，我们就学习了一元二次函数中的韦达定理，对于一元二次方程 $ax^2+bx+c=0$ 的两根 x_1,x_2，有

$$\begin{aligned} x_1+x_2&=-\frac{b}{a}\\ x_1\cdot x_2&=\frac{c}{a}\end{aligned}$$

可以通过求根公式来验证韦达定理。

现在将韦达定理推广到高次方程中，对于 n 次方程

$$\sum_{k=0}^{n}a_kx^k=a_0+a_1x+a_2x^2+\cdots+a_nx^n=0$$

有 n 个根 $x_1,x_2,\ldots,x_n$，一般的有

$$\begin{aligned} \sum_{i=1}^{n}x_i&=-\frac{a_{n-1}}{a_n}\\ \sum_{\substack{i,j=1\\ i<j}}^{n}x_ix_j&=\frac{a_{n-2}}{a_n}\\ \sum_{\substack{i,j,k=1\\ i<j<k}}^{n}x_ix_jx_k&=-\frac{a_{n-3}}{a_n}\\ &\cdots\cdots\cdots\\ \prod_{i=1}^{n}x_i&=(-1)^n\frac{a_0}{a_n}\end{aligned}$$

附录 B　常用符号解释

§B.1　累加和与连乘积

§B.1.1　$\sum$ 累加和

有很多人和我反应，即便是学过概统，但还是记不住 Σ 是什么意思，这可能是由于不常用导致的，累加和与连乘积的简记是一种非常方便的记号，每当忘记时，只需要看下此节即可：

累加和运算：对于一长串数据加起来，如：

$$a_0 + a_1 + a_2 + \cdots + a_n$$

我们可以记为：

$$\sum_{i=0}^{n} a_i$$

其中 Σ 符合表示累加和运算，它下面的 $i=0$ 意为这起始项为 0 开始，而它上面的 n 表示它的终止项为 n 结束，而右边的 a_i 中 $i=0,1,2,\dots,n$，$(n \in \mathbb{N})$ 就意味着每当 i 取一个值时，a_i 都会化为一个“项”，当 i 取完所有值时，将这些得到的项全部加起来就得到我们需要的式子了。

可以结合以下例子理解：

$$\begin{aligned}
&\sum_{i=1}^{n} i = 1 + 2 + 3 + \cdots + n \\
&\sum_{i=1}^{n} a = na \\
&\sum_{i=1}^{n} \left(i a_i^i 2^{2i-1}\right) = 2a_1 + 8a_2^2 + 96a_3^3 + \cdots + n a_n^n 2^{2n-1}
\end{aligned}$$

又例如贝叶斯公式：

$$\begin{aligned}
P(A_i|B) &= \frac{P(B|A_i) \cdot P(A_i)}{P(B)} (i = 1, 2, \dots, n) \\
P(A_i|B) &= \frac{P(B|A_i) \cdot P(A_i)}{\sum_{k=1}^{n} P(B|A_k) \cdot P(A_k)} (i = 1, 2, \dots, n)
\end{aligned}$$

这里请注意 k, i 的意义，求累加和的运算中的起始项字母 k 仅代表“求累加和”这个运算内的范围。

以上就是高中范围内可能用到的 $\sum$ 运算的内容，如果你接触到更难的内容，可能会涉及到更难看懂的 $\sum$

运算，就比如以下我举的二项式定理、三项式定理以及 n 项式定理：

$$
\begin{aligned}
(a+b)^n &= \sum_{r=0}^{n} \frac{n!}{(n-r)!r!} a^{n-r} b^r = \sum_{r+q=n} \frac{n!}{q!r!} a^q b^r \\
(a+b+c)^n &= \sum_{i+j+k=n} \frac{n!}{i!j!k!} a^i b^j c^k \\
(x_1+x_2+\cdots+x_t)^n &= \sum_{n_1+n_2+\cdots+n_t=n} \frac{n!}{n_1!n_2!\cdots n_t!} x_1^{n_1} x_2^{n_2} \cdots x_t^{n_t}
\end{aligned}
$$

其中二项式定理中有 $q=n-r$，你会发现有些 $\sum$ 的符号上没有终止项，而且下面也很奇怪；这一类可以理解为：满足 $\sum$ 符号下面的条件的非负整数都算作一项，之后把满足条件的项都加起来。

§B.1.2 $\prod$ 连乘积

$\sum$ 运算和 $\prod$ 运算基本一样，只不过是将“+”换成了“×”，如：

$$
\prod_{i=1}^{n} a_i = a_1 \times a_2 \times a_3 \times \cdots \times a_n
$$

所以只要理解 $\sum$ 运算就可以推出 $\prod$ 运算了，请读者理解以下例子：

$$
\begin{aligned}
&\prod_{i=1}^{n} i = 1 \times 2 \times \cdots \times n = n! \\
&\prod_{i=m}^{n} i = m \times (m+1) \times \cdots \times n = \frac{n!}{(m-1)!}
\end{aligned}
$$

以上便是高中可能涉及到的（不过高中教材似乎没涉及到过，竞赛有些题目上会有，不过都是出于简化式子的目的）

你更深入钻研肯能会遇到复杂的情况，譬如多个函数乘积的导数，我们在高中仅学习过两个函数乘积的导数，现在看看它更一般的形式：

$$
\left[\prod_{i=1}^{n} f_i(x)\right]' = \sum_{j=1}^{n} \left\{ f_j'(x) \prod_{i=1, i\neq j}^{n} f_i(x) \right\}
$$

由于涉及到 i, j 两个变量，我们要一步一步分析，当 j 取一个确定的值时，可以将 j 看作定量，也就是把 j“冻结”住，再来带入到这个大括号中，注意到，由于 j 取的值在 $[1, n]$ 之间，所以 i 必定会取到，又因 $i \neq j$ 所以，只需要将 $i=j$ 这一项排除掉再连乘，这是 j 为一个定值时取得的一个项，最后将所有满足 j 的条件的项累加起来就是我们需要的结果了。

如果你还是觉得晦涩难懂，不妨将 n 设为一个较小的值（如 $n=3$）来推推看，当 $n=3$ 时只需要用换元法就可以简单证明，不过带入较小的值取展开式子，在很多情况都有利于我们理解式子。

§B.2 阶乘

§B.2.1 阶乘

若一个正整数为 $n, n \in \mathbb{N}^+$，则

$$\prod_{i=1}^{n} i$$

上式称为 n 的**阶乘**，用 $n!$ 来表示，即

$$n! = \prod_{i=1}^{n} i$$

特别的，当 $n=0$ 时，有 $0! = 1$。

以上的阶乘定义是高中阶段所需要知道的。

如果对阶乘的定义由自然数扩充到整个实数（除非负整数），那么有

$$z! = \Gamma(z+1) = \int_0^{+\infty} t^z \mathrm{e}^{-t} \mathrm{d}t$$

§B.2.2 双阶乘

正整数的双阶乘，定义为小于等于该数且具有相同奇偶性的正整数的阶乘。即

$$(2n-1)!! = 1 \times 3 \times 5 \times \cdots \times (2n-1)$$
$$(2n)!! = 2 \times 4 \times 6 \times \cdots \times 2n$$

需要注意的是，要区分某数的双阶乘与某数的阶乘的阶乘，它们的定义是不同的。

e.g.

$$4!! = 2 \times 4 = 8, \quad (4!)! = (1 \times 2 \times 3 \times 4)! = 24!$$

参考文献

[1] 盛骤，谢式千，潘承毅. *概率论与数理统计*. 高等教育出版社, 4 edition, 2008.

[2] 陈纪修，於崇华，金路. *数学分析*, volume 1. 高等教育出版社, 2 edition, 2004.

[3] 蔡德锦，李尚泽. *高考数学真题分类狂刷基础 + 中档 2000 题*. 首都师范大学出版社出版发行, 1 edition, 2021.

[4] 闵嗣鹤，严士键. *初等数论*. 高等教育出版社, 4 edition, 2020.

[5] 丘维声. *高等代数*, volume 1. 清华大学出版社, 2 edition, 2019.