初等数学笔记

zhcosin

东汉永寿

序

这份笔记开工于 2016 年 4 月 6 日，其实它在这之前几年就应该动笔的，高考后的十年间，除了囫囵吞枣的学了一些高等数学内容外，数学能力没有与年龄同步增长，常引为恨事，所以打算以写作这份笔记为契机，给自己一个继续学习的机会，以期常有所思，常有所得。虽然平时也有一些数学上的思考，但终究没有积累下来，没有形成自己的知识体系，这份笔记算是在这方面的一个尝试。

其实我不能算是这份笔记的作者，只能算是整理者，因为其中属于我自己原创的内容其实很少，就这份笔记所涉及的内容而言，那些定理和结论都是前人几百年前甚至早在公元前就研究透彻了，我辈若是能在没接触前人成果的情况下独立发现一些结论就已经是非常不容易了，何况我是在早已经接触过这些结论甚至还大体记得推导证明过程的基础上进行了重新推导而已。这份笔记主要内容的来源，基本上是参考文献所列出的那几本书籍，这基本上也是我这几年所读过的书，实际上没有哪一部是认真读完过的，还有极少数的内容是我自己在没接触过前人结论的情况下自己推导所得（例如伯努利信封问题，我一直称为错位排列问题），这些内容主要以例子的形式出现。

这份笔记的写作受到了前苏联数学家菲赫金哥尔茨所著《微积分学教程》的影响，这是我最推崇的一部巨著，书中取材之广泛，讨论之深度和广度超乎我的想象能力，我也从这部书中受益良多，这份笔记也在取材的广度和深度方面甚至内容的组织方面都受到它的影响。

这份笔记目前没有成书之日，也没有什么计划之类，受限于自己的数学能力和工作闲暇，不定期的更新而已。

我最佩服的几何学家是古希腊的 Apollonius，中文译作阿波罗尼奥斯，他生活在公元前的古希腊，他所著的《圆锥曲线论》将圆锥曲线的性质几乎一网打尽，以致于后人在长达两千年间没能在这个领域有所建树，直到笛卡尔坐标几何的创立，他所采用的还是纯几何理论，当然他也是在一些前人的研究成果上结合自己的研究写出了这部巨著，有时间会认真读一读这部书的部分内容。让我惊叹的是这竟然是在公元前的希腊完成的，古希腊的数学到底是有多发达，何以古希腊没能在我们今天所称四大文明古国之列呢。

致谢也是不能缺少的，首先要感谢的是本书参考文献的作者们，是他们让我接触到了这么多精彩的数学内容，当然还有一些书没有在参考文献之列（其实是我也想不起来是从哪儿看到的了)，也一并感谢。还要感谢的是《计算机程序设计艺术》一书的作者 Donald E. Knuth，他开发的 TEX 排版系统以及由之发展而来的 LATEX 系统，使得我排版自己的书籍成为可能，还有编辑器 Emacs 的作者 Richard Stallman，这个编辑器

所带来的强大的功能和编辑体验对我完成这份笔记功不可没。最后还要感谢我的夫人和女儿，女儿的降生给了我们这个家庭前所未有的欢乐，我对她的信心是在她的学生生涯，数学学科不至于成为她的拦路虎。夫人在照顾女儿上的付出才让我得以有精力来完成这份笔记。还有我的父母和哥哥。

最后需要说明的是，这份笔记作为一份知识总结，没有在问题的引入和背景介绍方面下多少功夫 (这远不是目前的篇幅所能完成的，而且也受作者经验和水平限制)，因此并不适合于初学者用于学习新知识。

这份笔记的最新文档可以在以下链接找到，并在这个网页上不定期更新:

https://coding.net/u/zhcosin/p/math-notes-publish/git/blob/master/elementary-math-notes.pdf

有任何问题，欢迎邮件讨论: zhcosin@163.com.

zhcosin

2017-03-20 于成都华阳

目录

第一章　绪论

本章讲述一些与其它各章节没有太大直接关联的主题。

1.1　一些说明

1.1.1　求和符号与求积符号

今后会经常遇到类似于 $x_1 + x_2 + \cdots + x_n$ 这样的求和，每次都写成这样长不免浪费篇幅，也显得累赘，因为这求和中实际上只有两个要素，一是求和通项 x_i，二是求和的下标范围 1 到 n，因此把这个求和简记为 $\sum\limits_{i=1}^{n} x_i$，即

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

求和符号中的 x_i 就是求和的通项，通项中的 i 称为求和指标，而上下标 $i = 1$ 和 n 分别给出了求和指标的上下限，即求和范围。

例 1.1.1　二项式定理利用求和符号就可以写成

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

而等比数列的求和公式则可以写成 (公比不为 1)

$$\sum_{i=0}^{n-1} q^i = \frac{1-q^n}{1-q}$$

■

在有求和符号参与的运算过程中，常常对求和指标进行变换，比如把上式的求和顺序倒过来，写成

$$x_n + x_{n-1} + \cdots + x_1 + x_0$$

这时利用求和符号可以写成 $\sum\limits_{i=n}^{1} x_i$，意义不变，也可以写成 $\sum\limits_{i=1}^{n} x_{n+1-i}$，仍然没有本质上的不同，因为当 i 从 1 到 n 取值时，$n+1-i$ 便依次依次取值 $n, n-1, \ldots, 2, 1$。因此，在接触并使用求和符号的初始阶段，将它展开来验证相关等式是有益的，以后熟悉了就可以省去展开操作。

有时会从求和符号中拆分出若干项来，也会将一些项合并到其中去，例如

$$
\begin{aligned}
\sum_{i=0}^{n} x_i &= x_0 + \sum_{i=1}^{n} x_i \\
\sum_{i=1}^{n} x_i + x_{n+1} &= \sum_{i=1}^{n+1} x_i
\end{aligned}
$$

在运算时，有时也会变换求和范围，比如对于

$$\sum_{i=1}^{n} x_i = x_1 + x_1 + \cdots x_n$$

可将其变换为 $\sum\limits_{i=0}^{n-1} x_{i+1}$，同样也是因为当 i 从 0 到 $n-1$ 遍历取值时，$i+1$ 便依次取值 $1, 2, \ldots, n$。在这种变换下，便有如下的运算

$$
\begin{aligned}
& (1+t)\sum_{i=0}^{n-1} t^i - t\sum_{i=0}^{n-2} t^i \\
= & \sum_{i=0}^{n-1} t^i + t\sum_{i=0}^{n-1} t^i - t\sum_{i=0}^{n-2} t^i \\
= & \sum_{i=0}^{n-1} t^i + \sum_{i=0}^{n-1} t^{i+1} - \sum_{i=0}^{n-2} t^{i+1} \\
= & \sum_{i=0}^{n-1} t^i + \sum_{i=1}^{n} t^i - \sum_{i=1}^{n-1} t^i \\
= & \sum_{i=0}^{n} t^i
\end{aligned}
$$

在此基础上，还会遇到多重求和，比如对于一个 n 行 m 列的数阵 (称为矩阵) 的所有元素进行求和，这个数阵的形式为

$$
\begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1m} \\
a_{21} & a_{22} & \cdots & a_{2m} \\
\vdots & \vdots & \cdots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nm}
\end{pmatrix}
$$

要对它的所有元素进行求和，可以先把每一行的元素加起来，得到每一行的和，再把每一行的和加起来，便得到所有元素的和。第 i 行元素的和是 $\sum_{j=0}^{m} a_{ij}$，于是所有元素的和便是

$$\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}$$

同样，也可以先将每一列的元素加起来，再把每一列的和进行相加，显然所得出的结果

应与前面一样，所以

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{m}a_{ij} &= \sum_{i=1}^{n}(a_{i1}+a_{i2}+\cdots+a_{im})\\ &= (a_{11}+\cdots+a_{1m})+(a_{21}+\cdots+a_{2m})+\cdots+(a_{n1}+\cdots+a_{nm})\\ &= (a_{11}+\cdots+a_{n1})+(a_{12}+\cdots+a_{n2})+\cdots+(a_{1m}+\cdots+a_{nm})\\ &= \sum_{j=1}^{m}(a_{1j}+a_{2j}+\cdots+a_{nj})\\ &= \sum_{j=1}^{m}\sum_{i=1}^{n}a_{ij}\end{aligned}$$

于是得到二重求和的一个重要性质：可以交换两个求和指标的求和顺序，但这需要注意的是这里的两个求和指标的范围是互相独立的，如果不是这样则不能交换求和顺序，例如下面这样的求和

$$\sum_{i=0}^{n}\sum_{j=0}^{i}x_ix_j$$

类比于求和符号，也可以用连乘积符号 $\prod$ 来表示多个元素的连乘积，例如

$$\prod_{i=1}^{n}\frac{1}{2^i}=\frac{1}{2}\cdot\frac{1}{4}\cdots\frac{1}{2^n}$$

同样也有多重求积，求积符号与求和符号有着类似的运算性质，不再重复说明了。

有些情况下的求和，将各项从左到右按下标列出比较繁琐，反而不如直接指明求和指标所应满足的限制条件列出，例如，$(a+b+c)^n$ 的展开式中，每一项都具有形式 $t_{i,j,k}a^ib^jc^k$，其中 $i,j,k\geqslant 0$ 并且 $i+j+k=n$，如若按二重求和的写法应写为

$$(a+b+c)^n=\sum_{i=0}^{n}\sum_{j=0}^{n-i}t_{i,j,n-i-j}a^ib^jc^{n-i-j}$$

其意义就是，首先 i 可以从 0 取到 n，但是当 i 确定后，j 就只能取 0 到 $n-i$ 之间的值，因为要保证 $i+j\leqslant n$，在 i 和 j 都取定后，k 就只能取 $n-i-j$ 了，然后按这样的取法进行求和。

显然这样做的本质是，将所有满足 $i+j+k=n$ 的三元非负数对 (i,j,k) 进行排序，以适应求和符号对指标范围的硬性要求，但是这样写出来的结果反而不如 $i+j+k=n$ 这个限定条件来的直观，因此我们直接使用简单的记号

$$\sum_{i,j,k\geqslant 0,\ i+j+k=n}t_{i,j,k}a^ib^jc^k$$

来表示这种形式的求和，同样，在其它类似情形下，只要将求和指标所满足的限定条件列在求和符号下方就行了，例如对不超过 n 的素数进行求和

$$\sum_{p\leqslant n}p$$

这里没有在条件中指明 p 为素数，这需要在上下文中进行指明。

1.2 几个原理

1.2.1 容斥原理

原理 1.2.1 (容斥原理). 用 $|A|$ 表示集合 A 的元素个数，则多个集合的并集的元素个数是:

$$\begin{aligned}\left|\bigcup_{i=1}^{n} A_i\right| &= \sum_{i=1}^{n}|A_i| - \sum_{1\leqslant i<j\leqslant n}|A_i\cap A_j| + \sum_{1\leqslant i<j<k\leqslant n}|A_i\cap A_j\cap A_k| \\ &\quad - \cdots + (-1)^{n+1}\left|\bigcap_{i=1}^{n} A_i\right| \end{aligned} \tag{1.2.1}$$

容斥原理的含义借助韦恩图是显而易见的，其证明用数学归纳法即可，这里从略。

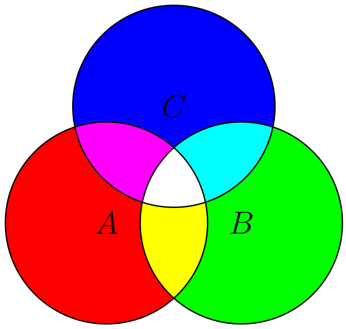

图 1.1 容斥原理示意图

例 1.2.1 伯努利 (Bernoulli) 信封问题 作为容斥原理的一个直接应用，这里讨论一下伯努利信封问题: 有相同数目的信封和信件若干，将这些信件装进这些信封，使得没有任何一封信件与信封搭配正确，问题是有多少种装法。

记信件数目是 n，并把信件和信封依次编号为 $1,2,\ldots,n$，信件与所属的信封编号相同。那么要计算所有信件都搭配错误的组装数，可以考虑其反面即至少有一封信件搭配正确的组装数目，再从总数 $n!$ 中减去它即可，根据容斥原理，构造出集合 A_i 表示第 i 封信件搭配正确的组装方案，则 $|A_i|=(n-1)!$，那么至少有一封信件搭配正确的方案集合就是这些 A_i 的并集，其数目是

$$\begin{aligned} & \sum_{i=1}^{n}|A_i| - \sum_{1\leqslant i<j\leqslant n}|A_i\cap A_j| + \sum_{1\leqslant i<j<k\leqslant n}|A_i\cap A_j\cap A_k| \\ & - \cdots + (-1)^{n+1}|\cap_{i=1}^{n} A_i| \\ = \ & n(n-1)! - C_n^2(n-2)! + C_n^3(n-3)! - \cdots + (-1)^{n+1}C_n^n 0! \\ = \ & n!\sum_{i=1}^{n}(-1)^{i+1}\frac{1}{i!} \end{aligned}$$

所以从总数 $n!$ 中减去它就得到最后的结果，在 $0!=1$ 的约定下，这结果可以写为:

$$\sum_{i=0}^{n}(-1)^{i}\frac{1}{i!} \tag{1.2.2}$$

于是问题得以解决，以后还将介绍利用数列递推方法来解决此问题 (见例 3.7.8)。■

1.2.2 数学归纳法

原理 1.2.2 (第一数学归纳法). *设 A 是一个元素都是整数的集合，如果:*

1. $1 \in A$.

2. 如果 $n \in A$，则必有 $n+1 \in A$.

那么对于任何正整数 n，都有 $n \in A$.

这个原理是显而易见的，它可以用来证明与整数有关的命题 $P(n)$，把 A 视为使得命题成立的那些整数的集合，那么如果命题对 $n=1$ 成立，即 $P(1)$ 为真，又如果由 $P(n)$ 成立能推得 $P(n+1)$ 也成立，则命题对于一切正整数就都成立。

除了常用的第一数学归纳法以外，我们还有以下的:

原理 1.2.3 (第二数学归纳法). *如果与正整数有关的命题 $P(n)$ 满足:*

1. $P(1)$ 成立;

2. 由 $P(1), P(2), \ldots, P(k)$ 成立能够推证出 $P(k+1)$ 成立;

那么该命题对于一切正整数成立.

有些命题在递推过程中，依赖的是前面的所有结论而非仅仅依赖前一个结论，此时第二数学归纳法就非常适用，例如下面这个例子:

例 1.2.2 由平方差公式和立方差

$$\begin{aligned} a^2 - b^2 &= (a-b)(a+b) \\ a^3 - b^3 &= (a-b)(a^2+ab+b^2) \end{aligned}$$

出发，可以利用数学归纳法得出一般的 n 次方差公式

$$a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \cdots + ab^{n-2} + b^{n-1})$$

或者利用求和符号写成

$$a^n - b^n = (a-b)\sum_{i=0}^{n-1} a^{n-1-i}b^i$$

证明. 先用数学归纳法证明公式的特殊情况，即 $a=1, b=t$ 的情况

$$1 - t^n = (1-t)\sum_{i=0}^{n-1} t^i$$

当 $n=2$ 时，即 $1-t^2 = (1-t)(1+t)$，公式是成立的，假定上式对于小于等于 n 的正

整数都成立，那么对于正整数 $n+1$ 的情形，就有

$$
\begin{aligned}
1-t^{n+1} &= (1+t)(1-t^n)-t(1-t^{n-1}) \\
&= (1-t)(1+t)\sum_{i=0}^{n-1}t^i - t(1-t)\sum_{i=0}^{n-2}t^i \\
&= (1-t)\left(\sum_{i=0}^{n-1}t^i + t\sum_{i=0}^{n-1}t^i - t\sum_{i=0}^{n-2}t^i\right) \\
&= (1-t)\left(\sum_{i=0}^{n-1}t^i + \sum_{i=0}^{n-1}t^{i+1} - \sum_{i=0}^{n-2}t^{i+1}\right) \\
&= (1-t)\left(\sum_{i=0}^{n-1}t^i + \sum_{i=1}^{n}t^i - \sum_{i=1}^{n-1}t^i\right) \\
&= (1-t)\sum_{i=0}^{n}t^n
\end{aligned}
$$

于是上式对于正整数 $n+1$ 也成立，在上式中取 $t=\dfrac{a}{b}$，即可得出要证的等式。 □

从证明过程中所可以看出，奠基不是从 1 开始的，而是从 2 开始的，而且为了证明 $P(n+1)$ 成立，利用了假设 $P(n)$ 和 $P(n-1)$ 成立。 ■

原理 1.2.4 (倒推归纳法). *如果与正整数有关的命题 $P(n)$ 满足:*

1. *有无穷多个正整数 n 使命题成立;*

2. *由 $P(k)$ 成立能够推证 $P(k-1)$ 成立;*

那么该命题对于一切正整数成立。

倒推归纳法的原理也是显而易见的，对于任何一个给定的正整数 n，由于不超过 n 的正整数只有有限个，所以必然存在一个大于 n 的正整数 N，使得命题 $P(N)$ 成立，再倒推回来，知 $P(n)$ 成立。

例 1.2.3 均值不等式 作为一个例子，我们用倒推归纳法来证明均值不等式，这比用通常的第一数学归纳法来得更加容易:

均值不等式的内容是：对任意 $n(n\geqslant 2)$ 个正实数 a_i，有下面不等式成立:

$$\frac{1}{n}\sum_{i=1}^{n}a_i \geqslant \left(\prod_{i=1}^{n}a_i\right)^{\frac{1}{n}}$$

证明.[1] 对 $n=2$ 的情形，有 $\frac{1}{2}(a_1+a_2)-\sqrt{a_1a_2}=\frac{1}{2}(\sqrt{a_1}-\sqrt{a_2})^2\geqslant 0$ 知不等式成立。

反复使用 $n=2$ 的结论，我们就可以得到当 n 是 2 的幂的时候不等式是成立的，然而 2 的幂是无穷多的，所以只要证明，不等式如果对 $n+1$ 个正实数成立就必然对 n 个正实数也成立就可以了。

[1]这个倒推归纳法的证明来自于参考文献 [8].

对于任意 n 个正实数，我们再添加一个正实数 $a_{n+1}=\frac{1}{n}\sum_{i=1}^{n}a_i$ 构成 $n+1$ 个正实数，由假设，不等式对 $n+1$ 个正实数是成立的，所以有

$$\frac{1}{n+1}\sum_{i=1}^{n+1}a_i \geqslant \left(\prod_{i=1}^{n+1}a_i\right)^{\frac{1}{n+1}}$$

而由于 a_{n+1} 正好等于其它 n 个实数的平均数，所以

$$\frac{1}{n+1}\sum_{i=1}^{n+1}a_i = \frac{1}{n}\sum_{i=1}^{n}a_i$$

因此前一式即为:

$$\frac{1}{n}\sum_{i=1}^{n}a_i \geqslant \left(\prod_{i=1}^{n}a_i\right)^{\frac{1}{n+1}} \cdot \left(\frac{1}{n}\sum_{i=1}^{n}a_i\right)^{\frac{1}{n+1}}$$

化简即得

$$\frac{1}{n}\sum_{i=1}^{n}a_i \geqslant \left(\prod_{i=1}^{n}a_i\right)^{\frac{1}{n}}$$

即得证。 □

■

原理 1.2.5 (跳跃数学归纳法). *如果与正整数有关的命题 $P(n)$ 满足:*

1. *$P(1)$,$P(2)$,...,$P(m)$ 成立;*
2. *由 $P(k)$ 成立能够推证 $P(k+m)$ 成立;*

那么该命题对于一切正整数成立。

数学归纳法还有其他形式，比如解决对偶性问题（如正余弦）的螺旋归纳法，解决同时与两个正整数相关的命题的二重归纳法等，但它们的道理都是类似的。

下面举几个应用数学归纳法的例子。

例 1.2.4 对于任意两个正整数 $m<n$，证明 $\frac{n!}{m!(n-m)!}$ 是整数。

证明. 这就是要证明组合数 C_n^m 是整数，从组合数的实际意义来看它应该是整数，但我们感兴趣的是数学证明，这个结论在初等数论中应用整数的素数分解定理可以很容易的证明它，这里我们使用数学归纳法来证明。

因为

$$C_n^m=\frac{n!}{m!(n-m)!}=\frac{n(n-1)\cdots(n-m+1)}{m!}$$

分子是连续 m 个正整数的乘积，所以我们只要证明辅助结论：任意 m 个连续正整数的乘积能被 $m!$ 整除。

实际上 $m!$ 自身也是 m 个连续正整数的乘积，而且是最小的乘积，我们对连续正整数的个数 m 施行数学归纳法。

显然 $m=1$ 时结论是成立的，今假定任意 m 个连续正整数能被 $m!$ 整除，来看连续 $m+1$ 个正整数的情况。

连续 $m+1$ 个正整数可以写成 $r(r+1)\cdots(r+m)$，显然 r 从 1 开始逐个取遍所有正整数，这个连乘积就能取遍所有 $m+1$ 个连续正整数的乘积，所以我们这里再嵌套一个对 r 的数学归纳法。

当 $r=1$ 时这连乘积即为 $(m+1)!$ 显然能被 $(m+1)!$ 整除，假定连乘积 $r(r+1)\cdots(r+m)$ 能被 $(m+1)!$ 所整除，那么我们来看当 r 增加 1 时的连续乘积 $(r+1)(r+2)\cdots(r+m+1)$，由

$$(r+1)(r+2)\cdots(r+m+1)-r(r+1)\cdots(r+m)=(m+1)\cdot(r+1)(r+2)\cdots(r+m)$$

由对 m 的归纳假设, $(r+1)(r+2)\cdots(r+m)$ 是 $m!$ 的倍数，所以这个差便是 $(m+1)!$ 的倍数，再由对 r 的归纳假设，左边的减数也是 $(m+1)!$ 的倍数，所以左边的被减数也能被 $(m+1)!$ 所整除，按照对 r 的归纳法，这就证得了任意 $m+1$ 个连续正整数的乘积都能被 $(m+1)!$ 所整除，再按对 m 的数学归纳法，就证得了我们最终的结论。 □

■

例 1.2.5　求证多项式系数 $\frac{n!}{n_1!n_2!\cdots n_m!}$ 是整数，其中 $n=n_1+n_2+\cdots+n_m(n_i\geqslant 0, i=1,2,\ldots,n)$。

证明. 由上一个例子中所证得的，二项式系数是整数，而多项式系数

$$\frac{n!}{n_1!n_2!\cdots n_m!}=\frac{n!}{n_1!(n-n_1)!}\cdot\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\cdots\frac{(n-n_1-n_2-\cdots-n_{m-1})!}{n_m!0!}$$

于是这多项式系数就转化成了若干个二项式系数的乘积，所以它亦是整数。 □

■

1.2.3　抽屉原理

1.3　坐标变换

在高级中学数学教材中，三角函数的伸缩变换已经为人所熟知，而本文要讨论的是坐标变换的一般性理论。

这些变换包括平移、旋转、伸缩、对称，通常所说的位似变换可以通过伸缩变换得到，其中，平移、旋转、对称都是恒等变换，即不改变平面图形的形状和大小，或者说变换前后的图形是全等的，而位似变换前后的图象则是相似的，形状不变，大小可以改变。

1.3.1　平移

平移变换由平移向量 $\vec{s}=(a,b)$ 唯一确定，平面上任意一点 P 与它在变换下的像 P' 满足 $\overrightarrow{PP'}=\vec{s}$ 为常向量，因此如果点 P 的坐标是 (x,y)，点 $P'(x',y')$ 的坐标将是 $(x+a,y+b)$，即

$$\begin{cases} x' &= x+a \\ y' &= y+b \end{cases}\tag{1.3.1}$$

写成矩阵形式则是:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{1.3.2}$$

我们把平移向量为 $\vec{s} = (a, b)$ 的平移变换记作 $S(a, b)$。

例 1.3.1 本目中来建立曲线 $f(x, y) = 0$ 在平移向量为 $\vec{s} = (a, b)$ 的平移变换下的新方程，假使点 $P(x, y)$ 为新曲线上任一点，则它在变换前的坐标则是 $(x - a, y - b)$，而变换前的点是满足原曲线方程的，所以得到新坐标所满足的方程 $f(x - a, y - b) = 0$，此即原曲线在平移变换下的新的曲线的方程。比如说，以原点为圆心的圆 $x^2 + y^2 = r^2$ 在此平移变换下的新方程即为 $(x - a)^2 + (y - b)^2 = r^2$。对于一般函数 $y = g(x)$ 而言，把它写成 $f(x, y) = g(x) - y$ 即可应用此结论。 ■

1.3.2 旋转

假如在平面直角坐标系中，有一个点的坐标是 $P(x, y)$，现在我们把它绕着原点逆时针方向转动一个角度 θ，我们来寻求这个点的新坐标 $P' = (x', y')$ 与原坐标之间的关系。

设向量 $\overrightarrow{OP}$ 与坐标系 x 轴正向成角 α，则向量 $\overrightarrow{OP'}$ 与坐标系 x 轴正向成角 $\alpha + \theta$，记向量 $\overrightarrow{OP}$ 长度为 r，则

$$x = r\cos\alpha, y = r\sin\alpha$$

同样有

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

于是就有

$$\begin{aligned} x' &= x\cos\theta - y\sin\theta \\ y' &= x\sin\theta + y\cos\theta \end{aligned} \tag{1.3.3}$$

写成矩阵形式就是:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

或者写成这样:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

我们把这个旋转变换记作 $R(\theta)$。

例 1.3.2 反比例函数 $xy = 1$ 的图象位于一三象限并且关于这两个象限的角平分线对称，我们尝试把它的图象绕原点顺时针旋转 45 度，看看新的曲线方程的模样。为此目的，记反比例曲线为 C，顺时针旋转 45 度后得到的新曲线记为 C'，我们反过来把 C 看成是由 C' 绕原点逆时针旋转 45 度得到的，就有关系式 $x = \frac{\sqrt{2}}{2}(x' - y'), y = \frac{\sqrt{2}}{2}(x' + y')$，因此新曲线 C' 上的点 $P(x', y')$ 都满足方程 $\frac{\sqrt{2}}{2}(x' - y') \cdot \frac{\sqrt{2}}{2}(x' + y') = 1$，即 $x'^2 - y'^2 = 2$，这是一个离心率为 $\sqrt{2}$ 的双曲线，所以反比例函数的图象是双曲线。 ■

1.3.3 伸缩

伸缩变换是以原点为基准，把平面沿着坐标轴的两个方向进行拉伸或压缩，两个方向上各有一个 *伸缩因子*，分别用 $\lambda(>0)$ 和 $\mu(>0)$ 代表 x 和 y 方向的伸缩因子，则显然新旧坐标之间的关系是:

$$\begin{cases} x' &= \lambda x \\ y' &= \mu y \end{cases} \tag{1.3.4}$$

矩阵形式为

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

或者写成这样:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

如果两个方向上的伸缩因子相等，则称它为以原点为中心的 *位似变换*.

我们把伸缩因子分别为 λ 和 μ 的伸缩变换记作 $L(\lambda, \mu)$。

如果以空间的视角来看伸缩变换，它也可以看作把平面倾斜一定角度后在原平面上的投影，所以关于投影有许多跟伸缩变换完全类似的结论，比如说投影下的面积。

例 1.3.3 二次曲线的离心率决定其形状 抛物线的标准方程 $y^2 = 2px(p > 0)$，对它进行位似变换，记原抛物线为 C，新的曲线为 C'，那么 C 也可以由新曲线进行位似变换得到，假定这个变换的伸缩因子为 λ，就有 $x = \lambda x', y = \lambda y'$，于是新的曲线方程是 $\lambda^2 y'^2 = 2p\lambda x'$，取 $\lambda = p$，则得 $y'^2 = 2x'$，这表明任何一个抛物线 $y^2 = 2px(p > 0)$ 的图象的形状都是一样的，不同的只是大小。事实上，同样的方法应用到圆、椭圆、双曲线上，得出的结论是: 率心率相同的二次曲线都是相似的，换句话说，离心率是确定二次曲线形状的参数，离心率相同，则形状相同，不同的只是大小。 ■

例 1.3.4 伸缩变换下的面积 椭圆的面积 讨论一下伸缩变换对平面图形面积的影响，根据祖暅原理，如果两个几何体在任一水平面上的截面积都相同，那么它俩体积相等。同理也有，如果两个平面图形被某一方向上的任一直线所截得的线段长度都相等，那么它俩面积也相等，进一步，如果它俩被某一方向的任意直线所截得的线段长度都有相同的比例，那么它们面积之比也为此固定比例。而伸缩变换 $L(\lambda, \mu)$ 可以通过两次单方向的伸缩变换 $L(\lambda, 1)$ 和 $L(1, \mu)$ 来实现，易见前者把面积放大 λ 倍，后者放大 μ 倍，所以如果一个平面图形在变换前后的面积分别为 S 和 S'，那么就有

$$S' = \lambda\mu S \tag{1.3.5}$$

这是一个有趣的结论，比如说我们可以通过它得到椭圆的面积，因为标准椭圆 $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1(a > b > 0)$ 在伸缩变换 $L(\frac{1}{a}, \frac{1}{b})$ 下将变为单位圆，单位圆面积是 π，所以椭圆面积就是 $S = \pi ab$。 ■

例 1.3.5 伸缩变换下直线的斜率 椭圆的切线 在伸缩变换 $L(\lambda, \mu)$ 下，假定一条直线 l 被变换成了直线 l'(直线在伸缩变换下仍旧是直线这一事实，可以通过直线方程

在变换之后仍旧是一个二元一次方程这一点上看出)，记原直线斜率为 k，新直线斜率为 k'，那么对直线上两点有

$$k' = \frac{y_1' - y_2'}{x_1' - x_2'} = \frac{\mu}{\lambda} \cdot \frac{y_1 - y_2}{x_1 - x_2} = \frac{\mu}{\lambda} k \tag{1.3.6}$$

因此新旧直线斜率之比是与伸缩因子有关的固定比例。有了这个结论，我们可以利用它求出椭圆曲线在任一点处的切线方程来，因为椭圆 $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1(a > b > 0)$ 在伸缩变换 $L(\frac{b}{a}, 1)$ 下成为圆 $x^2 + y^2 = b^2$，因为圆的切线与过切点处的半径互相垂直，也就是斜率之积为 -1，因此椭圆上任一点处的切线的斜率，与该点与椭圆中心连线的斜率之积为 $-\frac{b^2}{a^2}$，这是一个椭圆中非常有用的结论，在椭圆上任取一点 $P(x_0, y_0)$，那么可以写出过该点的切线方程

$$y - y_0 = -\frac{b^2 x_0}{a^2 y_0}(x - x_0)$$

利用 $\frac{x_0^2}{a^2} + \frac{y_0^2}{b^2} = 1$ 即可将它化简为

$$\frac{x_0 x}{a^2} + \frac{y_0 y}{b^2} = 1$$

■

例 1.3.6 如果允许伸缩因子是虚数，那么在伸缩变换 $L(1, i)$ 下 (i 是虚数单位)，双曲线将被变换为椭圆! 然而这对现阶段的我们似乎也没有什么用处。 ■

还有一种变换称为 *反射变换*，也称为对称变换，即把平面上所有的点关于某一直线作对称，如果该对称轴与坐标轴成倾斜角度，可以先做旋转变换，使得对称轴与坐标轴平行或者垂直，再通过平移使对称轴与坐标轴重合，这时只需要关于坐标轴作对称之后再反平移和反旋转，即可实现一般的反射变换，所以只要讨论关于坐标轴的反射变换就可以了，然而这一点如果允许伸缩变换中的伸缩因子为负就可以得到，所以反射变换本质上是平移、旋转和伸缩的叠加，就不单独讨论了。

1.4 组合与二项式定理

1.4.1 分类计数原理与分步计数原理

1.4.2 排列、线排列与圆排列

1.4.3 组合、若干组合模型

1.4.4 二项式定理

1.4.5 组合恒等式

1.4.6 多项式定理

从中学数学教材中熟知有如下的二项式定理

定理 1.4.1 (二项式定理)**.** 对于任意两个实数 x 和 y，以及任意的正整数 n，有如下等式:

$$(x+y)^n = \sum_{i=0}^{n} C_n^i x^i y^{n-i} \tag{1.4.1}$$

其中每一项的系数 $C_n^i = \frac{n!}{i!(n-i)!}$ 称为 二项式系数。

利用数学归纳法，证明是很容易的，此处略去。

把这定理推广到多个数相加的情况，就有如下的多项式乘幂定理

定理 1.4.2 (多项式乘幂定理)**.** 对于任意 m 个实数 $x_i(i=1,2,\ldots,m)$，以及任意的正整数 n，有如下等式:

$$\left(\sum_{i=1}^{m} x_i\right)^n = \sum_{r_i \geqslant 0, \sum_{i=1}^{m} r_i = n} \frac{n!}{r_1!r_2!\cdots r_m!} x_1^{r_1} x_2^{r_2} \cdots x_m^{r_m} \tag{1.4.2}$$

其中的指数组合 $(r_1, r_2, \ldots, r_m)$ 要遍及方程 $\sum_{i=1}^{m} r_i = n$ 的所有非负整数解，每一项的系数 $\frac{n!}{r_1!r_2!\cdots r_m!}$ 称为 多项式系数。

对加数的个数 m 使用数学归纳法，并利用二项式定理，便可以证明此定理，此处同样略去。

这公式随着次数和加数个数的增加会迅速变长，它的项数就是方程 $\sum_{i=1}^{m} r_i = n$ 的非负整数解的个数，这个值是 C_{n+m-}^{m-1}，比如说，对于 $n=3, m=3$ 的情况，把这公式写出来就是:

$$(a+b+c)^3 = a^3+b^3+c^3+3a^2b+3a^2c+3b^2a+3b^2c+3c^2a+3c^2b+6abc$$

为了以后书写方便，我们引入所谓轮换求和和对称求和的符号，对于三个实数 a, b, c 而言，轮换求和是

$$\sum_{cyc} f(a,b,c) = f(a,b,c) + f(b,c,a) + f(c,a,b)$$

而对称求和是

$$\sum_{sym} f(a,b,c) = f(a,b,c) + f(a,c,b) + f(b,c,a) + f(b,a,c) + f(c,a,b) + f(c,b,a)$$

即轮换求和是将序列 a, b, c 首尾相接后轮换进行求和，而对称求和是要对所有可能的排列进行求和。

在这种符号下，$(a+b+c)^3$ 的展开可以写成如下这样:

$$(a+b+c)^3 = \sum_{sym} a^3 + 3\sum_{sym} a^2b + 6abc$$

而 $(a+b+c+d)^4$ 的展开则是

$$(a+b+c+d)^4 = \sum_{sym} a^4 + 4\sum_{sym} a^3b + 6\sum_{sym} a^2b^2 + 12\sum_{sym} a^2bc + 24abcd$$

现在利用这些简化的符号来表达多项式乘幂定理，可以看出，定理公式中因子次数组合相同的项都具有相同的系数，把系数提出来就是对称求和，所以这定理可以改写为

$$\left(\sum_{i=1}^{m} x_i\right)^n = \sum_{r_i\geqslant 0,\sum_{k=1}^{m} r_k=n} \frac{n!}{r_1!r_2!\cdots r_n!}\sum_{sym} x_1^{r_1}x_2^{r_2}\cdots x_n^{r_n}$$

例 1.4.1 均值不等式 作为多项式乘幂定理的一个应用，我们来证明均值不等式: 对任意 n 个正实数 $x_i(i=1,2,\ldots,n)$，有下面不等式成立:

$$\sqrt[n]{x_1x_2\cdots x_n} \leqslant \frac{x_1+x_2+\cdots x_n}{n}$$

由于证明过程写起来比较晦涩，先写出 $n=3$ 的过程示例，以帮助理解，以 a,b,c 标记这三个数，由 $(a^k+b^k)-(a^{k-1}b+ab^{k-1})=(a-b)(a^{k-1}-b^{k-1})\geqslant 0$ 得 $a^k+b^k\geqslant a^{k-1}b+ab^{k-1}$，于是

$$\sum_{sym} a^2b = \frac{1}{2}\sum_{sym}(a^2b+bc^2) = \frac{1}{2}\sum_{sym} b(a^2+c^2) \geqslant \frac{1}{2}\sum_{sym} 2bac = 6abc$$

同样

$$\sum_{sym} a^3 = \sum_{sym}\frac{a^3+b^3}{2} \geqslant \sum_{sym}\frac{a^2b+ab^2}{2} = \frac{1}{2}\sum_{sym} a^2b \geqslant 3abc$$

所以最终

$$(a+b+c)^3 = \sum_{sym} a^3 + 3\sum_{sym} a^2b + 6abc \geqslant 3abc+18abc+6abc = 27abc$$

于是三元均值不等式得证。

可以看出，这就是一个不断平衡各个因子次数的过程，它基于 $a^k+b^k\geqslant a^{k-1}b+ab^{k-1}$ 这个基本的不等式，下面我们把这个过程一般化。

只需要证明如下的不等式

$$(x_1+x_2+\cdots+x_n)^n \geqslant n^n x_1x_2\cdots x_n$$

我们考虑左边按照多项式乘幂定理展开后的对称求和的通项

$$\frac{n!}{r_1!r_2!\cdots r_n!}\sum_{sym} x_1^{r_1}x_2^{r_2}\cdots x_n^{r_n}$$

这个对称求和是针对 $r_1,r_2,\ldots,r_n$ 这个组合的各种可能的排列的，把右边的对称求和（不要系数）记作 $\sigma(r_1,r_2,\ldots,r_n)$，我们来平衡这些次数。

理想的次数平衡是每个因子的次数 r_i 都是 1，如果不是如此的话，因为这 $\sum_{i=1}^{n} r_i = n$，所以必然存在某两个 r_i 和 r_j 使得 $r_i-1\geqslant r_j+1$ 即 $r_i-r_j\geqslant 2$，我们把这两个次数平衡一次，变成 r_i-1 和 r_j+1，基本不等式是

$$(x_1^{r_i}y_2^{r_j}+x_2^{r_i}x_1^{r_j})-(x_1^{r_i-1}x_2^{r_j+1}+x_2^{r_i-1}x_1^{r_j+1}) = x_1^{r_j}x_2^{r_j}(x_1-x_2)(x_1^{r_i-r_j-1}-x_2^{r_i-r_j-1})\geqslant 0$$

所以

$$\begin{aligned}
\sigma(r_1,\ldots,r_i,\ldots,r_j,\ldots,r_n) &= \sum_{sym} x_i^{r_i} x_j^{r_j} \cdots \\
&= \frac{1}{n-1}\sum_{sym}(x_i^{r_i}x_j^{r_j} + x_i^{r_j}x_j^{r_i})\cdots \\
&\geqslant \frac{1}{n-1}\sum_{sym}(x_1^{r_i-1}x_2^{r_j+1} + x_2^{r_i-1}x_1^{r_j+1})\cdots \\
&= \frac{1}{2(n-1)}\sum_{sym} x_1^{r_i-1}x_2^{r_j+1}\cdots \\
&= \frac{1}{2(n-1)}\sigma(r_1,\ldots,r_i-1,\ldots,r_j+1,\ldots,r_n)
\end{aligned}$$

■

1.5 无理指数幂的定义问题

在中学数学的课本中，没有定义指数为无理数的幂，然而却提出了定义在全实数域上的指数函数，这是因为无理指数幂的定义要用到极限，本文的目的是为了在初等数学范围内给无理指数幂作一个解释，以解答中学生对此问题可能的疑惑。

在高等数学中，对无理指数幂的定义是，对于一个无理数 r 和一个实数 $a>0$，用任意一个以 r 为极限的有理数的序列 $r_i(i=1,2,\ldots)$ 去逼近它，无理指数幂 a^r 的定义为 $a^r=\lim_{n\to\infty}a^{r_n}$.

这个极限值是与序列 r_n 无关的，用任何一个以 r 为极限的有理数序列，所得的那个极限都是相同的。这就是高等数学中无理指数幂的定义。

接下来我们尝试在初等数学范围内解释一下无理指数幂定义。

先回顾一下在中学数学范围内是如何定义出一个实数的。

对于字面上能够写出来的数，比如整数 2 与小数 3.2，我们清楚它的每一个数位上的数字是多少，我们就认为我们定出了一个数。比如 3.2 的个位是 3，十分位为 2，其余为零。

有些数我们是不可能把它的每一位都写出来的，比如说无限循环小数 $\frac{1}{3}$，但是我们清楚它的小数部分每一位都是 3，这样我们也认为我们定出了一个数。

圆周率 π 是一个自然界中存在的常数，这个无限不循环小数我们也不能把它的每一位都写出来，甚至我们为了确定它的小数部分第 100 位数是几都得经过一番计算，但无论如何，它的每一位数我们也是能够确定的，只是这需要一些计算而已[2]。

还有一些数，比如 $\ln 2$，我们是用它所满足的一些性质来刻画它的，如果要问 $\ln 2$ 的某一数位上的数字是几，我们也需要通过一些计算步骤才能得出[3]。

所以我们可以说，我们认为我们定出了一个数，当且仅当我们能够回答出来这个数的每一个数位上的数字是多少，无论这个回答是立刻就能作出的，还是需要经过一系列的运算。

[2]这计算需要利用高等数学中的级数展开式

[3]这同样需要利用高等数学中的级数展开式

有了这点认识，我们要定义如像 $2^{\sqrt{3}}$ 这样的无理指数幂，我们只要能够按照某种规则确定出它的每一个数位上的数值就行了。

把所有实数都用小数的形式来表示，即

$$x = \cdots x_2 x_1 x_0 . x_{-1} x_{-2} \cdots (0 \leqslant x_i \leqslant 9) \tag{1.5.1}$$

在这种表示下，对于两个实数 x 和 y，如果从数从高到低的顺序（双端无穷序列），第一个不相同的数位上，x 的该数位大于 y 的该数位，则有 $x \geqslant y$，如果所有数位都相同，则必有 $x = y$。

上面为何是大于等于而不是大于呢，考察一下 1 与 $0.\dot{9}$ 就知道了，这两个数是相等的，因为它们的差可以小于任何一个正实数，从而这个差值就只能是零，所以有 $1 = 0.\dot{9}$。为了能在上段结论中去掉这个等号，我们约定，如果一个无限循环小数的循环部分是 9，我们就把它收上来，也就是说 $1.12\dot{9} = 1.13$，在这样的约定下，刚才比较大小时如果第一个不相同的数位上谁大，谁的值就大。

提一下一个实数的不足近似值和过剩近似值的概念[4]，对于任何一个实数，它的 n 位 *不足近似值*是

$$\underline{x_n} = \cdots x_2 x_1 x_0 . x_{-1} x_{-2} \cdots x_{-n} \tag{1.5.2}$$

它的 n 位 *过剩近似值*是

$$\overline{x_n} = \cdots x_2 x_1 x_0 . x_{-1} x_{-2} \cdots x_{-n} + 10^{-n} \tag{1.5.3}$$

通俗的说法就是，n 位不足近似值是舍掉小数点后第 n 位以后的数位，而 n 位过剩近似值则是在把这以后的数位收上来。易知序列 $\underline{x_n}$ 单调不减而序列 $\overline{x_n}$ 单调不增，而且有 $\overline{x_n} - \underline{x_n} = 10^{-n}$。

从定义可以看出，对于任何一个实数 x，不等式 $\underline{x_n} \leqslant x \leqslant \overline{x_n}$ 永远成立，但是两边的等号不能同时成立。更进一步，对于任意两个正整数 m 和 n，实际上有 $\underline{x_m} < \overline{x_n}$ 成立，这个很容易说明，假定 $m < n$，就有 $\underline{x_m} \leqslant \underline{x_n} \leqslant x \leqslant \overline{x_n}$，而式中的后两个等号又不能同时取到，所以有此结论。

现在来考虑无理指数幂，对于一个实数 $a > 0$ 和一个无理数 r，我们作出 r 的不足近似值序列 $\underline{r_n}$ 和过剩近似值序列 $\overline{r_n}$，因为 $\underline{r_n}$ 和 $\overline{r_n}$ 都是有理数，我们作两个序列 $L_n = a^{\underline{r_n}}$ 和 $M_n = a^{\overline{r_n}}$。

现在我们来证明，存在唯一一个实数 $P = P(a, r)$，它能够介于所有的 L_n 和所有的 M_n 之间，具体来说，就是在 $a > 1$ 的情况下不等式 $L_n < P < M_n$ 对于任意正整数都成立，在 $0 < a < 1$ 的情况下是 $M_n < P < L_n$ 对一切正整数恒成立。我们将会把这个实数 P 定义为 a^r，这样定义将会使指数函数在整个实数集上保持它在有理数集上同样的单调性。

只就 $a > 1$ 的情况进行论证，至于 $0 < a < 1$ 的情况，有 $\frac{1}{a} > 1$，于是可以验证 $\frac{1}{P(\frac{1}{a}, r)}$ 就符合要求。

在 $a > 1$ 的假定下对任意两个正整数 m 和 n 就有 $L_m < M_n$。这就是说，任何一个 L_m 都小于所有的 M_n，任何一个 M_n 都大于所有的 L_m。

[4]本文这个利用不足近似值和过剩近似值进行夹逼的思路来自于参考文献 [5].

因为

$$M_m - L_m = a^{\overline{x_m}} - a^{\underline{x_m}} = a^{\underline{x_m}}(a^{\overline{x_m}-\underline{x_m}} - 1) < a^{\overline{x_0}}(a^{\overline{x_m}-\underline{x_m}} - 1) < a^{\overline{x_0}}(a^{10^{-m}} - 1)$$

为了定出实数 P 在 10^{-n} 数位上的数值，我们让 $M_m - L_m < 10^{-n}$，这只要下式成立就可以了

$$a^{10^{-m}} < 1 + \frac{1}{10^n a^{\overline{x_0}}} \tag{1.5.4}$$

但是要注意的是我们现在还没有无理指数幂，所以对数的定义也就不完整，暂时也就不要尝试使用对数来从上式中解出 m 了。

现在我们需要一个引理，这个引理的内容是，对于一个实数 $a > 1$ 和任何一个大于 1 的实数 T，存在一个正有理数 t，使得对任何一个满足 $0 < x < t$ 的有理数 x 恒有 $a^x < T$。

这个引理也是显而易见的，只是我们现在需要把指数函数限制在有理数上，这时指数函数的单调性仍然同我们所熟知的，其它性质也与我们所熟知的在实数集上的指数函数一样。首先我们可以找到一对正整数 u 和 v，使得 $a^u < T^v$，也就是 $a^{\frac{u}{v}} < T$，于是根据定义在有理数集上的指数函数单调性，在区间 $(0, \frac{u}{v})$ 上的有理数 x 将恒有 $a^x < T$，所以引理成立。

现在回到前面的问题上来，式1.5.4的右端即是一个大于 1 的与 m 无关的正实数 T_n，根据引理，存在与 n 有关的有理数 t_n，使得对区间 $(0, t_n)$ 上的所有有理数 x 都成立 $a^x < T_n$，于是只要 $10^{-m} < t_n$ 即 $10^m > \frac{1}{t_n}$ 时式1.5.4就能成立，而 $\frac{1}{t_n}$ 显然也是一个与 m 无关的常数，所以也可以找到一个与 n 有关的有理数 s_n，使得 $m > s_n$ 时便有 $10^m > \frac{1}{t_n}$。

所以我们得到的结论就是，存在与正整数 n 有关的有理数 s_n，使得当 $m > s_n$ 时便恒有 $M_m - L_m < 10^{-n}$，于是我们的要找的实数 P 就这样定义，它在 10^{-n} 这一位上的数值，与 $m > s_n$ 时 L_m 在这一位上的数值相同（容易知道 m 继续增大时，这个数位上的值不会变）。通俗的说就是，在 L_m 随着 m 增长（单调不减）的过程中，我们取它某一位上稳定下来的数值（稳定是指随着 m 无限增大都不会改变）。这样我们确实定出了一个实数，接下来需要做的事情就是，证明对于任意正整数 n 都有 $L_n < P < M_n$。

从实数 P 的作法即可知 $L_n \leqslant P$，但是由于 r 是一个无理数，它不可能从某一数位开始后面全是零，所以 $\underline{r_m}$ 随着 m 增大，虽然中途可能有所停留，但必将持续增大，因而实数 P 不可能与某个 L_m 相等，否则就与 $L_n \leqslant P$ 恒成立相矛盾了。所以对于所有正整数 n，恒有 $L_n < P$ 成立，不等式左边得证。

为了证明不等式 $P < M_n$ 对所有正整数成立，先证明实数 P 与序列 L_m 是无限接近的，也即是说，$P - L_n$ 随着 n 的增大可以小于任何正实数，这个是明显的，根据实数 P 的作法即知，存在有理数 s_n，使得 $m > s_n$ 时，$P - L_m < 10^{-n}$，所以有此结论。有了这个，利用反证法就可以说明 $P < M_n$ 对所有正整数 n 成立，若不然，假定 P 能够大于某个 M_{n_0}，则有 $P - L_n \geqslant P - M_{n_0}$，右端为一常数，这与 $P - L_n$ 可以小于任何正实数相矛盾，所以实数 P 确实是介于所有的 L_n 和所有的 M_n 之间的一个实数，并且由于 $M_n - L_n$ 可以任意小，位于这两个序列之间的实数也是唯一的。

现在，我们将把这个实数 P 定义为 a^r，因为 $L_n < P < M_n$，这个定义使得指数函数在实数集上仍然是单调性的。

需要说明的是，以上有一些不甚严格的地方，这严格性需要建立在实数公理化体系之上。另外这基本上已经接触到确界的概念和戴德金的实数分划理论了。

1.6 行列式

1.6.1 奇排列与偶排列

为了讨论行列式的每一项的符号，这一小节来讨论下奇排列与偶排列。

我们来考虑正整数 $1, 2, \ldots, n$ 的任一排列，称其为一个 n 级排列，交换其中两个元素的位置的操作称为一个 对换, 我们先指出如下定理

定理 1.6.1. 正整数 $1, 2, \ldots, n$ 的任一排列 $a_1, a_2, \ldots, a_n$ 均可由原始排列 $1, 2, \ldots, n$ 经过有限次对换得到，同样它也可以经过有限次对换得出原始排列 $1, 2, \ldots, n$。

证明. 我们给出一个实际的对换方案，如果 $a_1 \neq 1$，则必然有某个 $a_i = 1 (i > 1)$，于是我们先对换 a_1 和 a_i，对换之后就可以使得新的 $a_1 = 1$ 了，然后再看 a_2 是否等于 2，如果不是则做类似处理，依次这样下去，最多对换 $n - 1$ 次，便可以得出原始排列 $1, 2, \ldots, n$，这是一个对换方案，然后把这个对换方案做一个逆对换，便可以由原始排列 $1, 2, \ldots, n$ 得出排列 $a_1, a_2, \ldots, a_n$. □

显然由这定理可知，任意两个不同的 n 级排列都可以通过对换操作互相转换。

现在我们准备引入排列的奇偶性，为此先给出如下结论

定理 1.6.2. 如果一个排列是经由原始排列 $1, 2, \ldots, n$ 通过奇数次对换得到，则它也只能经由奇数次对换得到，同样，如果它是由原始排列经过偶数次对换得到，则它也只能通过偶数次对换得到。

证明. 我们对排列的级数 n 作数学归纳法。显然 $n = 1, 2$ 时结论均成立，假定结论对小于等于 n 级的排列都成立，我们来考虑 $n + 1$ 级排列的情况。

假定一个 $n + 1$ 级排列 $a_1, a_2, \ldots, a_n, a_{n+1}$ 是由原始排列 $1, 2, \ldots, n, n + 1$ 经由奇数次对换得到的， □

由这定理可知对于一个确定的 n 级排列，由原始排列得到它的对换次数的奇偶数是确定的，根具体的对换方案无关，仅由排列本身确定，为此我们把只能由原始排列通过奇数次对换而得到的排列称为奇排列，而把只能由偶数次对换而得到的排列称为偶排列，根据上面两个定理可知，一个排列，必然要么是奇排列要么是偶数列，不能两者都是或者两者都不是。

由上述定理立即可得:

定理 1.6.3. 对排列作一次对换将改变其奇偶性。

虽然我们定义了排列的奇偶性，但对于一个具体的排列，要指出它是奇排列还是偶排列却还是不容易的，最简单的办法就是按照定理 1.6.1的证明过程中的步骤构成出一个对换方案来，并由其对换次数来确定排列的奇偶性，但我们还有更可行的方法，为此提出 正序与 逆序的概念。

定义 1.6.1. 对于正整数 $1, 2, \ldots, n$ 的任一排列 $a_1, a_2, \ldots, a_n$，如果其中两个元素 a_i 和 a_j 满足 $i < j$ 且 $a_i < a_j$，则称它们是这排列中的一个 正序，如果它们满足 $i < j$ 而 $a_i > a_j$，则称其为这排列中的一个 逆序，一个排列中所有正序的个数称为 正序数，所有逆序的个数称为 逆序数。

显然，一个排列中的正序数与逆序数之和是 $C_n^2 = n(n-1)/2$.

我们指出

定理 1.6.4. 一个排列的逆序数，与排列本身具备相同的奇偶性。

为了证明这个定理，我们先证如下引理

引理 1.6.1. 正整数 $1, 2, \ldots, n$ 的任一排列都可以由原始排列 $1, 2, \ldots, n$ 通过有限次相邻元素的对换得到。

证明. 我们仍然实际构造出一个方案，假设这排列是 $a_1, a_2, \ldots, n$，如果 $a_1 \neq 1$，则在原始排列中，把数 a_i 依次跟它前一个元素对换，直到把它移动第一个元素的位置，同样的方法逐个应用到 $a_2, a_3, \ldots, a_n$，即可在有限步内将原始排列 $1, 2, \ldots, n$ 变成 $a_1, a_2, \ldots, a_n$。 □

有了这引理，我们回到定理 1.6.4的证明上

证明. 由引理 1.6.1，排列可以由原始排列通过相邻元素的对换而得来，而这对换的次数的奇偶性又是排列自身确定的，而显然相邻元素的对换必然改变逆序数的奇偶性，所以经过多少次相邻元素的对换，逆序数的奇偶性就改变多少次，而原始排列 $1, 2, \ldots, n$ 的逆序数是零，所以逆序数的奇偶性，与对换次数的奇偶性相同，即与排列的奇偶性相同。 □

显然，利用逆序数是容易求得一个具体排列的奇偶性的。

1.6.2 行列式的概念

我们把二级行列式和三级行列式的概念推广到任意 $n \times n$ 的数阵上，对于一个 $n \times n$ 的数阵

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

它的行列式是所有位于不同行且不同列的元素的乘积的代数和，也就是说，求和中的每一项如果不考虑符号都具备形式

$$a_{i_1j_1}a_{i_2j_2}\cdots a_{i_nj_n}$$

其中 $i_1, i_2, \ldots, i_n$ 和 $j_1, j_2, \ldots, j_n$ 都是 $1, 2, \ldots, n$ 的一个排列，而这求和中将会遍及所有可能的排列。我们把上式按 i_k 的顺序重新排列成

$$a_{1j_1}a_{2j_2}\cdots a_{nj_n}$$

这时 $j_1, j_2, \ldots, j_n$ 已经不再是原来的排列了，我们只是仍使用相同的符号，然后这一项的符号是这样定义的：当 $j_1, j_2, \ldots, j_n$ 是奇排列是带负号，为偶排列时带正号，这样所有这样的项的和便是由数阵所决定的行列式的值。

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \sum_{j_1, j_2, \ldots, j_n} (-1)^{\tau(j_1, j_2, \ldots, j_n)} a_{1j_1}a_{2j_2}\cdots a_{nj_n}$$

1.6.3 行列式的性质与计算

1.6.4 行列式按一行 (列) 展开

1.6.5 Cramer 法则

1.6.6 Laplace 定理

第二章　初等数论

本章是参考文献 [6] 一书的学习笔记。

2.1　整除理论

2.1.1　整除的概念与带余除法

定义 2.1.1. 对于两个整数 a 和 b，如果存在某个整数 c 使得 $a = bc$ 成立，就称 b 能整除 a，或者 a 能被 b 所 整除，记为 $b \mid a$，这时，b 称为 a 的 因数，a 称为 b 的 倍数.

显然，任何整数都能被 1 整除，也能被它自己整除。

易知整除具有下列性质:

性质 2.1.1. 设 a、b 都是整数，若 $a \mid b$ 且 $b \mid c$，则 $a \mid c$，即整除的传递性。

性质 2.1.2. 设 a、b、k 都是整数，若 $k \mid a$ 且 $k \mid b$，则 $k \mid (a \pm b)$

性质 2.1.3. 设 k、$a_i (i = 1, 2, \ldots, n)$ 都是整数，若 $k \mid a_i (i = 1, 2, \ldots, n)$，则 $k \mid \sum_{i=1}^{n} \lambda_i a_i$，这里 $\lambda_i (i = 1, 2, \ldots, n)$ 都是整数.

性质 2.1.4. 设 a 和 b 是两个正整数，若 $b \mid a$，则 $b \leqslant a$.

证明. 因为存在整数 c 使得 $a = bc$，又 a、b 都是正的，所以 c 也必是正整数，所以 $c \geqslant 1$，从而 $a = bc \geqslant b$. □

下面的结论是本小节最重要的结论:

定理 2.1.1 (带余除法)**.** 设 a、b 是任意两个整数，其中 b 是正的，则存在唯一的整数 q 和唯一的非负整数 r 使得 $a = qb + r (0 \leqslant r < b)$ 成立，q 称为 a 除以 b 所得的 商，r 称为 a 除以 b 所得的 余数.

证明. 由 b 是正整数，作无穷序列 $\cdots, -2b, -b, 0, b, 2b, \cdots$，利用这些节点把全体整数划分在左闭右开区间序列 $[tb, (t+1)b)$ 上，任何一个整数必定落在其中某个区间上，而且不能同时两个区间上，对于整数 a，假设这区间是 $[qb, (q+1)b)$，这时就有 $qb \leqslant a \leqslant (q+1)b$，令 $r = a - qb$ 就有 $0 \leqslant r < b$，得证。 □

2.1.2 最大公因数与辗转相除法

定义 2.1.2. 设 a 和 b 都是整数，如果整数 c 能同时整除 a 和 b，就称 c 是 a 和 b 的一个 公因数.

显见公因数与 a 和 b 的符号无关，今后我们都将只讨论正整数的正的公因数。

显然 1 是任何两个整数的公因数，于是公因数的存在性得以解决，对于两个正整数 a 和 b 而言，若 c 是它俩的公因数，则由性质 2.1.4，有 c 不超过 a 和 b 中的较小者，从而公因数必定只有有限个，其中最大的那一个就称为 a 和 b 的 *最大公因数*，并把它记为 (a,b)，公因数和最大公因数的概念不难推广到多个数的情况，符号仍采用相同的小括号。

如果两个正整数 a 和 b 的最大公因数是 1，即 $(a,b)=1$，则称这两个正整数 *互素* 或者 *互质*，多个数的情形仍类似，但需要区别的是多个数互素与它们之间两两互素是不同的，例如三个数 2,3,4 互素，显然并非两两互素。

显然，对于任意非零整数，它自己是它与零的公因式，其绝对值便是最大公因数。

接下来我们讨论 *辗转相除法*，也称 *欧几里德除法*。

定理 2.1.2. 对于等式 $a=qb+r$，如果其中所涉及的四个数都是整数，则 a 与 b 的公因数集合，跟 b 与 r 的公因数集合相等，从而也有相同的最大公因数.

证明很简单，能同时整除 a 和 b 的，按等式也能整除 r，反之，能同时整除 b 和 r 的，也必定能整除 a，所以定理是显然的。

利用这个定理，我们可以通过如下的辗转相除法来求出两个正整数的最大公因数，对于任意两个正整数 a 和 b，不妨假定 $a \geqslant b$，否则就交换 a 和 b 这两个符号，所以这是合理的，根据带余除法，我们有一连等式:

$$\begin{aligned} a &= q_0 b + r_0 \\ b &= q_1 r_0 + r_1 \\ r_0 &= q_2 r_1 + r_2 \\ &\cdots \\ r_{n-1} &= q_{n+1} r_n + r_{n+1} \end{aligned}$$

因为余数总是比除数小，所以有 $b > r_0 > r_1 > \cdots > r_n > r_{n+1} \geqslant 0$，所以这个步骤进行有限步之后，必定得出 $r_{n+1}=0$，根据刚才所得的定理，就有 $(a,b)=(b,r_0)=(r_0,r_1)=\cdots=(r_n,r_{n+1})=(r_n,0)=r_n$，这就是说，这个除法序列进行到余数为零时，最后一个除数便是正整数 a 和 b 的最大公因数，这个反复做除法的方法就叫做辗转相除法，也叫欧几里德除法。

从辗转相除法，我们还得出一个重要结论:

定理 2.1.3. 设正整数 d 是正整数 a 和 b 的最大公因数，则存在整数 s 和 t 使得 $d=sa+tb$.

证明. 根据辗转相除法，设 $a \geqslant b$，设 $r_0 = a, r_1 = b$，则有下面一连串等式

$$
\begin{aligned}
r_0 &= q_1 r_1 + r_2 \\
r_1 &= q_2 r_2 + r_3 \\
&\cdots \\
r_{k-1} &= q_k r_k + r_{k+1} \\
&\cdots \\
r_{n-1} &= q_n r_n + 0
\end{aligned}
$$

其中 r_n 便是 a 和 b 的最大公因数，上式中第 k 个等式是

$$r_{k-1} = q_k r_k + r_{k+1}$$

所以得到

$$r_k = r_{k-2} - q_{k-1} r_{k-1}$$

这说明，每一个 r_k 都可以通过比它下标较小的相邻两个 r_m 表示表示成形状 $u_k r_{k-1} + v_k r_{k-2}$，而 r_{k-1} 又可以进一步利用用 r_{k-2} 和 r_{k-3} 来表达，反复继续下去，最终 r_n 就可以由 r_1 和 r_0(也就是 a 和 b) 通过类似的形状来表达，我们把这过程写成数学归纳法的形式。

我们来证明，最大公因数 r_n 可以表成下面的形式

$$r_n = P_k r_{n-k+1} + Q_k r_{n-k}$$

其中 P_k 和 Q_k 都是整数。

我们对 k 施行数学归纳法，当 $k = 1$ 时，由 $r_n = r_{n-2} - q_{n-1} r_{n-1}$ 知 $k = 2$ 时成立，并且 $P_2 = -q_{n-1}$，$Q_k = 1$，假定 $r_n = P_k r_{n-k+1} + Q_k r_{n-k}$，那么由将等式 $r_{n-k+1} = r_{n-k-1} - q_{n-k} r_{n-k}$ 代入其中，则得

$$r_n = (Q_k - q_{n-k} P_k) r_{n-k} + P_k r_{n-k-1}$$

于是令 $P_{k+1} = Q_k - q_{n-k} P_k, Q_{k+1} = P_k$，就有 $r_n = P_{k+1} r_{n-k} + Q_{k+1} r_{n-k-1}$，于是这个表达式总是成立的，还顺便得到了系数 P_k 和 Q_k 的递推公式。

在上述表达式中令 $k = n$ 便得 $r_n = P_n r_1 + Q_n r_0 = P_n a + Q_n b$，所以定理结论得证，但是要指出的是，定理中的 s 和 t 并不是唯一的，因为假若 s_0 和 t_0 是这样的一对整数，则对任意整数 t，易知 $s_0 + bt$ 和 $t_0 - at$ 也能满足定理中的等式。 □

推论 2.1.1. *两个整数 a 和 b 互素的充分必要条件是，存在整数 s 和 t 使得 $sa + tb = 1$ 成立。*

证明. 如若这两个整数互素，则 1 作为它俩的最大公因数，自然是存在满足这等式的整数 s 和 t 的。反之，如果存在两个整数满足这等式，而 d 是 a 和 b 的最大公因数，按这等式必有 $d \mid 1$，所以 $d = 1$，于是 a 和 b 互素。 □

推论 2.1.2. 正整数 a 和 b 的任一公因数，都是它俩的最大公因数的因数，反之，最大公因数的因数，也都是公因数。

定理的前半部分这由等式 $(a,b)=sa+tb$ 便可直接得出，后半部分由整除的传递性乃显然。

关于这定理中的等式，我们还有以下的更详细的结论:

定理 2.1.4. 设 a 和 b 是两个非零整数，那么集合 $\{m|m=xa+yb,x,y\in Z\}$ 中的所有数按从小到大排列成一个正负对称的 *(双向)* 无穷序列，这个序列将是一个等差数列，其公差便是 a 和 b 的最大公因数，同时因为零也在这序列中，这序列中的最小正数也就正是这个最大公因数。

证明. 首先这集合是正负对称的，并且零也列于其中，作为一个无限的整数集合，其中必有一个最小的正的整数，假定便是 $x_0a+y_0b>0$，我们来证明，它能整除这个集合中的全部数。

由带余数法，存在整数 q 及 $r(0\leqslant r<x_0a+y_0b)$，使得

$$xa+yb=q(x_0a+y_0b)+r$$

这里的余数 r 必定是零，这是因为 $r=(x-qx_0)a+(y-qy_0)b$ 也从属于这集合，如果它是正的，即 $0<r<x_0a+y_0b$，这便与 x_0a+y_0b 是这集合中的最小正整数相矛盾，所以 $r=0$，即这集合中的最小正整数 x_0a+y_0b 能够整除集合中的任意一个数。

其次证明这个 x_0a+y_0b 就是 a 和 b 的最大公因数，因为 a 和 b 本来也在这集合中，所以 x_0a+y_0b 能同时整除它俩，即为它俩的公因数，如果这还不是最大公因数的话，假定最大公因数是 d，而由于存在整数 s 和 t，使得 $d=sa+tb$，所以 d 也属于这集合，从而 $(x_0a+y_0b)\mid d$，而又显然 $d\mid(x_0a+y_0b)$，所以 $x_0a+y_0b=d$。

最后来证明这集合从小到大构成一个等差数列，且公差就是 a 和 b 的最大公因数，显然 $nd=n(x_0a+y_0b)=(nx_0)a+(ny_0)b$，这就是说，最大公因数 d 的任意倍数都在这集合中，而前面已经证得集合中的任意数都是最大公因数的倍数，所以这集合就等于 $\{m|m=nd,n\in Z\}$，定理得证。 □

由这定理即知，除了最大公因数，别的公因数无法表示成 $sa+tb$ 的形式，也就是说，在 a 和 b 的公因数中，只有最大公因数能表示成这种形式，这就是如下的结果:

定理 2.1.5. 设正整数 d 是整数 a 和 b 的一个公因数，如果又存在整数 s 和 t 使得 $d=sa+tb$，则 d 是 a 和 b 的最大公因数。

证明. 假定 a 和 b 的最大公因数是 k，因为公因数都是最大公因数的因数，所以 $d\mid k$，又显然 $k\mid(sa+tb)$，即 $k\mid d$，所以 $d=k$. □

定理 2.1.6. 设整数 a 与 b 互素，c 是一个整数，则 a,c 的公因数和 a,bc 的公因数相同，从而也有相同的最大公因数。

证明. 显然 a,c 的公因数也都是 a 和 bc 的公因数，只要再证明 a,bc 的公因数也都是 a,c 的公因数即可，由条件 a,b 互素，存在整数 s 和 t 使得 $sa+tb=1$，两边乘以 c 得

$(cs)\cdot a+t\cdot bc=c$，所以 a,bc 的公因数必然整除 c，从而它也是 a,c 的公因数，既然两对数有着相同的公因数集合，自然也有相同的最大公因数。 □

推论 2.1.3. *如果 a 与 b 互素，并且 $a \mid bc$，则有 $a \mid c$.*

证明. 因为 $a \mid bc$ 知 a 是 a,bc 的最大公因数，由定理 2.1.6知它也是 a,c 的最大公因数，即 $a \mid c$. □

推论 2.1.4. *如果 $(a,b)=1$，且 $a \mid c$, $b \mid c$，则 $ab \mid c$.*

证明. 设 $c=aa_1$，则 $b \mid aa_1$，由上一推论，有 $b \mid a_1$，从而 $ab \mid aa_1=c$. □

证明二. 由条件，存在整数 s 和 t 使 $sa+tb=1$，又设 $c=ak$，则 $sak+tbk=k$，即 $sc+tbk=k$，显然 b 能够整除等式的左端，因而也能整除右端，即 $b \mid k$，从而 $ab \mid ak=c$. □

推论 2.1.5. *若 $(a,b)=1$, 且 $(a,c)=1$，则 $(a,bc)=1$.*

证明. 设 $d=(a,bc)$，由 $(a,b)=1$，显然有 $(d,b)=1$，这是因为 d 与 b 的公因式必然也是 a 与 b 的公因式，所以 $(d,b)=1$. 又 $d \mid bc$，由推论 2.1.3，有 $d \mid c$，于是 d 成为 a 与 c 的公因式，但 $(a,c)=1$，所以 $d=1$，即 $(a,bc)=1$. □

证明二. 把等式 $s_1a+t_1b=1$ 和 $s_2a+t_2c=1$ 相乘，得 $(s_1s_2a+s_1c+s_2b)a+t_1t_2bc=1$，即知 $(a,bc)=1$. □

这个推论还可以推广为下述的定理

定理 2.1.7. *若 $a=p_1p_2\cdots p_n$，$b=q_1q_2\cdots q_m$，且对于任意 i 和 j 都有 $(p_i,q_j)=1$，那么 $(a,b)=1$.*

证明. 由 $(p_i,q_j)=1$，根据推论 2.1.5便知 $(p_i,b)=1(i=1,2,\ldots,n)$，进一步便有 $(a,b)=1$. □

定理 2.1.8. *如果 $(a,b)=1$，n 是一个正整数，则 $(a,b^n)=1$.*

证明. 由 a、b 互素，存在整数 s 和 t，使得 $sa+tb=1$，于是 $(1-sa)^n=t^nb^n$，左边按二项式定理展开，并把含有 a 的项归并到一起，得 $1+au=t^nb^n$，即 $-ua+t^nb^n=1$，故 $(a,b^n)=1$. □

再补充一个关于最大公因数的结论:

定理 2.1.9. *设有整数 a 和 b，最大公因数是 (a,b)，那么*

1. *对于任意正整数 m，有 $(ma,mb)=m(a,b)$.*
2. *设 δ 是最大公因数 (a,b) 的一个因子，则有 $(\frac{a}{\delta},\frac{b}{\delta})=\frac{(a,b)}{\delta}$.*

证明. (1). 因为存在整数 s 和 t，使得 $(a,b)=sa+tb$，所以 $m(a,b)=s(ma)+t(mb)$，即 $m(a,b)$ 能表示成 ma 和 mb 的组合形式，又显然 $m(a,b)$ 是 ma 和 mb 的公因数，所以它是 ma 和 mb 的最大公因数。(2). 类似的可以证明，略去。 □

对于多个整数的情形，我们也有类似的结论:

定理 2.1.10. 如果 d 是若干个整数 $a_i(i=1,2,\ldots,n)$ 的最大公因数，那么存在整数 $s_i(i=1,2,\ldots,n)$ 使得 $d=\sum_{i=1}^{n}s_ia_i$.

推论 2.1.6. 整数 $a_i(i=1,2,\ldots,n)$ 的公因数，与 $(a_1,a_2,\ldots,a_n)$ 的因数相同。

定理 2.1.11. 设整数 d 是若干个整数 $a_i(i=1,2,\ldots,n)$ 的一个公因数，如果存在 n 个整数 $s_i(i=1,2,\ldots,n)$ 使得 $d=\sum_{i=1}^{n}s_ia_i$，那么 d 便是这 n 个整数的最大公因数。

最后给出多个整数的最大公因数的一个实际的求法:

定理 2.1.12. 对于若干个正整数 $a_i(i=1,2,\ldots,n)$，若记 $d_1=a_1,d_2=(d_1,a_2),d_3=(d_2,a_3),\ldots,d_n=(d_{n-1},a_n)$,则 d_n 就是这 n 个整数的最大公因数,即 $d_n=(a_1,a_2,\ldots,a_n)$.

证明. 因为 $d_k=(d_{k-1},a_k)$，所以 $d_k\mid a_k$，$d_k\mid d_{k-1}$，由数学归纳法即可容易得到 $d_n\mid a_i(i=1,2,\ldots,n)$，即 d_n 是这 n 个整数的一个公因数.

记这 n 个整数的最大公因数为 d，那么由前一推论，有 $d_n\mid d$，又有 $d\mid a_1=d_1$，又 $d\mid a_2$，即 d 同时整除 d_1 和 a_2，于是就有 $d\mid d_2$，依次下去，有 $d\mid d_i(i=1,2,\ldots,n)$，所以 $d\mid d_n$，即 d 跟 d_n 能互相整除，所以 $d=d_n$。

要证明它是最大公因数也可以通过这个 d_n 可以表示为 $\sum_{i=1}^{n}s_ia_i$ 的形式，此处略去。 □

2.1.3 最小公倍数

定义 2.1.3. 如果整数 c 既是 a 的倍数，也是 b 的倍数，即它能同时被 a 和 b 所整除，则称它是 a 和 b 的一个 公倍数.

显然整数 a 和 b 的乘积 ab 便是它俩的一个公倍数。

只考虑正整数的情形，显然两个正整数的公倍数同时大于等于 a 和 b，于是在它俩的所有公倍数中，必定有一个最小的，称之为这两个正整数的 最小公倍数，记为 $[a,b]$，多个数的情形仍有类似的定义及符号。

定理 2.1.13. 对于两个整数 a 和 b,

1. 它俩所有的公倍数，都是最小公倍数的倍数.
2. 它俩的最小公倍数与最大公因数的乘积，等于 a 和 b 的乘积.

证明. 设 m 是 a,b 的最小公倍数，而 M 是它俩的任一公倍数，由带余带法，存在整数 q 和小于 m 的非负整数 r，使得 $M=qm+r(0\leqslant r<m)$，显然按这式子，r 也必然是 a,b 的公倍数，所以只能 $r=0$，否则与 m 是最小公倍数矛盾，所以 $m\mid M$。

再来证明最小公倍数与最大公因数的乘积等于 ab，分别用 d 和 m 表示 a,b 的最大公因数和最小公倍数，则存在整数 a_1,b_1,a_2,b_2 使得

$$a=a_1d,\, b=b_1d$$

$$m=a_2a,\, m=b_2b$$

这时必然有 $(a_1,b_1)=1$ 和 $(a_2,b_2)=1$，若不然，a,b 就会有比 d 更大的公因数和比 m 更小的公倍数。记 $M=a_1b_1d$，显然 M 是 a,b 的一个公倍数，因而它一定是 m 的倍数，所以 $m\mid M$，即 $m=a_1a_2d\mid a_1b_1d=M$，所以 $a_2\mid b_1$，同样由 $m=b_1b_2d\mid a_1b_1d=M$，得 $b_2\mid a_1$.

另一方面，因为 $m=a_2a=b_2b$，所以 $a_1a_2d=b_1b_2d$，即 $a_1a_2=b_1b_2$，所以 $a_1\mid b_1b_2$，但 $(a_1,b_1)=1$，所以 $a_1\mid b_2$，同样由 $b_1\mid a_1a_2$ 并且 $(a_1,b_1)=1$ 可得 $b_1\mid a_2$.

于是 a_1 与 b_2 互相整除，a_2 与 b_1 互相整除，所以 $a_1=b_2,a_2=b_1$，从而 $m=M$，即 a_1b_1d 便是 a,b 的最小公倍数，所以 $dm=ab$. □

但要注意的是，对于多个整数，最大公因数与最小公倍数的乘积就不一定等于这些整数的乘积了，例如 2,4,8 的最大公因数是 2，最小公倍数是 8，显然不满足，只有在这组整数两两互素的情况下才有这结论，这时最大公因数是 1，最小公倍数便是它们的乘积。

对于多个整数的公倍数，我们有以下结论:

定理 2.1.14. *设 $a_i(i=1,2,\ldots,n)$ 是 n 个整数，记 $m_1=a_1,m_2=[m_1,a_2],m_3=[m_2,a_3],\ldots,m_n=[m_{n-1},a_n]$，那么 m_n 便是这 n 个整数的最小公倍数。*

证明. m_n 是这 n 个整数的公倍数是明显的，记 M 是这 n 个整数的任一公倍数，则 $m_1=a_1\mid M$，又 $a_2\mid M$，所以 $m_2\mid M$，如此继续下去，便有 $m_n\mid M$，所以 m_n 便是最小公倍数。 □

2.1.4 素数与算术基本定理

定义 2.1.4. *如果一个整数除 1 与自身以外没有其它因数，则称之为一个* 素数*.*

素数是数论中极其重要的一个概念。

定理 2.1.15. *设 p 为一个素数，而 a 为任一整数，则要么 $(p,a)=1$，要么 $p\mid a$.*

证明. 设 $d=(p,a)$，d 作为 p 的一个因数，由素数定义，或者 $d=1$，或者 $d=p$，若 $d=1$，即 $(p,a)=1$，若 $d=p$，即 $p\mid a$. □

推论 2.1.7. *设 p 为一素数，a 和 b 为两个整数，若 $p\mid ab$，则必有 $p\mid a$ 或者 $p\mid b$.*

证明. 若 $p\mid a$ 则结论成立，而 $p\nmid a$ 时，由定理 2.1.15知 $(p,a)=1$，此时由推论 2.1.3知必有 $p\mid b$，得证。 □

利用归纳法，这结论可以推广到多个数的情形:

推论 2.1.8. *设 p 为一素数，如果 $p\mid a_1a_2\cdots a_n$，则 p 必能整除诸 a_i 中至少一个。*

定义 2.1.5. *如果一个整数不是素数，则称它为一个* 合数*.*

由定义，设 a 为一合数，则其必存在因子 b，使得 $1<b<a$，即 $a=bb_1$，换句话说，合数必能被分解为比它小的两个数之乘积，进一步，如果分解出来的这两个数中还有合数，则继续进行分解，直到乘积中全部是素数为止，这样我们就得到以下著名的定理:

定理 2.1.16 (算术基本定理). *任一大于 1 的正整数 a 都能表成一些素数的乘积，即*

$$a = p_1 p_2 \cdots p_n$$

诸 p_i 均为素数并可以相等，并且这分解式在不考虑各因子的顺序的情况下是唯一的。

证明. 先证明可分解性，使用数学归纳法，对于 $a=2$ 时，已经表成一个素数之积了，假定对于小于 a 的正整数都能分解为素数之积，则对于 a，若它本身是素数，则勿需再分解，若不是素数，则必然可表示为两个比它小的两个因数之乘积，根据归纳假设，这两个因数均能分解为素数之积，把这两个因式的分解式合起来，就是 a 的分解式，这就是证明了 a 也能表为素数之积。

再证明分解式的唯一性，设 a 还有另一种素数分解

$$a = q_1 q_2 \cdots q_m$$

其中诸 q_i 也是素数，则有

$$p_1 p_2 \cdots p_n = q_1 q_2 \cdots q_m$$

则每一个 p_i 都能整除诸 q_i 之积，由素数性质，它必然能整除 q_i 中的某一个，比如说是 q_s，但 q_i 本身是素数，除 1 及自身之外没有其它因数，所以只能 $p_i = q_s$，这就是说，左边每一个 p 都必然等于右边的某个 q，同理，右边的每个 q 也必然等于左边的某个 p，但此时还不能断定左右两个分解式是相同的，因为分解式中允许有相同的素数重复出现的情况，我们还必须证明相同素数出现的重复次数也相同，设某个素数 r 在两个分解式中分解出现 s 次和 t 次，则 r^s 能够整除诸 q_i 之积，由定理 2.1.7，有 r^s 与 q_i 之积中除去 r^t 之外的各项乘积是互素的，所以只能 $r^s \mid r^t$，同理有 $r^t \mid r^s$，所以 $s=t$，这就证明了结论。 □

对于任意正整数 a，在它的分解式中把相同的素数写成幂的形式，就得出如下的 *标准分解式*:

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_m^{\alpha_m},\ \alpha_i > 0 (i = 1, 2, \ldots m)$$

在讨论两个乃至多个正整数时，可以把只出现在其中某个正整数的分解式中的素数也合并到别的数的分解式中去并标以零次，这通常对于讨论问题是有用的，这样对于两个数 a 和 b，就有分解式

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_m^{\alpha_n},\ \alpha_i \geqslant 0 (i = 1, 2, \ldots n)$$

和

$$b = p_1^{\beta_1} p_2^{\beta_2} \cdots p_n^{\beta_n},\ \beta_i \geqslant 0 (i = 1, 2, \ldots n)$$

在此分解式基础上，有如下的重磅结论:

定理 2.1.17. *设两个正整数 a 和 b 有分解式*

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_m^{\alpha_n},\ \alpha_i \geqslant 0 (i = 1, 2, \ldots n)$$

和

$$b = p_1^{\beta_1} p_2^{\beta_2} \cdots p_n^{\beta_n},\ \beta_i \geqslant 0 (i = 1, 2, \ldots n)$$

则 $a \mid b$ 的充分必要条件是 $\alpha_i \leqslant \beta_i (i = 1, 2, \ldots, n)$.

证明. 充分性是显然的，只证必要性，由 $a \mid b$ 知 $p_i^{\alpha_i} \mid b$，但根据定理 2.1.7，$p_i^{\alpha_i}$ 与 b 的分解式中除 $p_i^{\beta_i}$ 以外的那部分子乘积是互素的，所以必然 $p_i^{\alpha_i} \mid p_i^{\beta_i}$，因此 $\alpha_i \leqslant \beta_i (i = 1, 2, \ldots, n)$. □

由此得

推论 2.1.9. 设 a 的标准分解式为

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_m^{\alpha_m},\ \alpha_i > 0 (i = 1, 2, \ldots m)$$

则 a 的所有因数都具有形式

$$d = p_1^{\beta_1} p_2^{\beta_2} \cdots p_m^{\beta_m},\ 0 \leqslant \beta_i \leqslant \alpha_i (i = 1, 2, \ldots m)$$

进一步可得

定理 2.1.18. 仍设两个正整数 a 和 b 有分解式

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_m^{\alpha_n},\ \alpha_i \geqslant 0 (i = 1, 2, \ldots n)$$

和

$$b = p_1^{\beta_1} p_2^{\beta_2} \cdots p_n^{\beta_n},\ \beta_i \geqslant 0 (i = 1, 2, \ldots n)$$

则它俩的最大公因数和最小公倍数分别是

$$(a, b) = p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_m^{\gamma_n},\ \gamma_i = \min\{\alpha_i, \beta_i\} (i = 1, 2, \ldots n)$$

和

$$[a, b] = p_1^{\delta_1} p_2^{\delta_2} \cdots p_m^{\delta_n},\ \delta_i = \max\{\alpha_i, \beta_i\} (i = 1, 2, \ldots n)$$

这几个定理和推论，利用素数分解把整除理论推上了巅峰，提示了整除性的本质，而这一切的基础，正是整数的素数分解，正是基于此理由，素数分解定理才被称为算术基本定理。

2.1.5 高斯函数

2.2 不定方程

2.2.1 二元一次不定方程

定理 2.2.1. 二元一次不定方程 $ax + by = c$ 有整数解的充分必要条件是 a 和 b 的最大公因数 d 能够整除 c.

证明. 先证必要证,如果这方程有解,比如说 $x = x_0, y = y_0$ 是它的一个解,则 $ax_0+by_0 = c$,显然 a 和 b 的最大公因数 d 能够整除左边,自然也就能整除 c,必要性成立。

再证充分性,如果 $d = (a, b) \mid c$,记 $c = c_1 d$,由最大公因数性质,存在整数 s 和 t 使得 $sa + tb = d$,于是 $(c_1 s)a + (c_1 t)b = c$,从而 $x = c_1 s, y = c_1 t$ 是方程的一个解,充分性成立。 □

定理 2.2.2. 如果二元一次不定方程 $ax + by = c$ 有一个解 $x = x_0, y = y_0$,则它的全部解是

$$x = x_0 + b_1 t,\ y = y_0 - a_1 t\ (t = 0, \pm 1, \pm 2, \ldots)$$

其中 a_1 和 b_1 分别是从 a 和 b 中约去它们的最大公因数后剩下的因子。

证明. 显然 $x = x_0 + b_1 t, y = y_0 - a_1 t, (t = 0, \pm 1, \pm 2, \ldots)$ 都是这方程的解,只要证明它给出了方程的全部解即可,设 $x = x_1, y = y_1$ 是方程的任意一个解,即 $ax_1 + by_1 = c$,则

$$a(x_1 - x_0) + b(y_1 - y_0) = 0$$

两边约去 a 和 b 的最大公因数 d 后得

$$a_1(x_1 - x_0) + b_1(y_1 - y_0) = 0$$

这表明 $a_1 \mid b_1(y_1 - y_0)$ 且 $b_1 \mid a_1(x_1 - x_0)$,但 $(a_1, b_1) = 1$,所以必有 $a_1 \mid (y_1 - y_0)$, $b_1 \mid (x_1 - x_0)$,记 $y_1 - y_0 = a_1 t_1$,$x_1 - x_0 = b_1 t_2$,则由上式得 $t_1 + t_2 = 0$,所以记 $t = t_2$,则 $x_1 = x_0 + b_1 t, y_1 = y_0 - a_1 t$,这就表明定理中的解给出了方程的全部解。 □

2.2.2 勾股方程与费马问题

2.3 同余

2.3.1 同余及其性质

2.3.2 剩余类与完全剩余系

2.3.3 欧拉函数、欧拉定理与费马小定理

2.3.4 RSA 加密与解密

2.3.5 一次同余式与中国剩余定理

第三章　初等代数

3.1　函数概要

3.1.1　函数概念

函数是两个数集之间的一个映射，根据对应法则，数集 A 中每一个数在数集 B 中都有唯一一个数与之对应。函数通常写为 $y = f(x)$，但这并不是说，所有函数都能表示成自变量的式子，比如黎曼函数就没有解析式，而隐函数 $f(x, y) = 0$ 甚至不能将 y 解成关于 x 的式子。

在讨论一些函数时，为了方便，将它的对应法则分解成嵌套的多个法则，于是得到复合函数的概念，但它并不是一类新的函数，只是认识函数对应法则的一个视角而已。

对于某些函数，由于它的对应法则的逆法则也正好满足函数定义（只要原法则下不存在多对一，则逆法则就不存在一对多，从而符合函数定义），因此自变量也就可以看成因变量的函数，这就是反函数，反函数与其原来函数是同一对应法则的两种表示方法，图象也是完全重合的，只有在互换 x 和 y 后，两者图象关于一三象限角平分线对称。

3.1.2　函数的性质

比较通用的性质是单调性，对称性（含奇偶性），周期性，凸凹性等。

单调性

单调性反应了两个变量的变化趋势，如果变化趋势一致，则为增函数，变化趋势相反则为减函数。但函数在某一区间上并不必然有某种单调性，有些函数无论你把区间划分得多么小，都没有单调性，比如狄利克雷函数和黎曼函数。

这里讨论下函数 $f(x) = x + \frac{a}{x}$ 的单调性，这里 a 是任何固定的正实数。

因为它是奇函数，奇函数在关于原点对称的区间上单调性情况相同，所以只要讨论 $x > 0$ 的情况即可，此时由于

$$f(x) = \left(\sqrt{x} - \frac{\sqrt{a}}{\sqrt{x}}\right)^2 + 2$$

括号中部分是关于 x 的增函数，但是外面有平方，还得考虑它的符号，在 $x = \sqrt{a}$ 左侧为负右侧为正，所以 $f(x)$ 在 $(0, \sqrt{a}]$ 上单调减少，在 $[\sqrt{a}, +\infty)$ 上单调增加，在 $x = \sqrt{a}$ 处有极小值 $f(\sqrt{a}) = 2\sqrt{a}$，在 x 趋近于 0 和正无穷大时，函数值亦趋向于正无穷大，

而且在这两个情形下，它分别与反比例函数 $y=\frac{1}{x}$ 和正比例函数 $y=x$ 无限逼近，因此它的图象如图所示。

对称性

奇偶性是对称性的特殊情况，更一般的情况是，若函数 $f(x)$ 的图象关于直线 $x=a$ 对称，则 $f(a+x)=f(a-x)$，若它的图象关于点 (a,b) 中心对称，则 $f(a+x)+f(a-x)=2b$。

周期性

周期性反映了函数值重复取值的规律，三角函数的周期性人尽皆知。此处需要说明的是周期函数并不一定存在最小正周期，除了最为特殊的常量函数以外，狄利克雷函数（在任何有理数处函数值为 1，而任何无理数处函数值为零）也可以说明这一点，任何有理数都是它的周期，而最小的正有理数是不存在的。

凸性

凸凹性反映了函数图像的拱形特征，这个性质是一大批不等式的本源，比如说，由对数函数的上凸性即得均值不等式，再由琴生（Jensen）不等式可推得多元均值不等式，更为宽泛的加权均值不等式仍然从对数函数的上凸性获得。本章有专门讨论这一性质的小节。

3.2 方程的求解

3.2.1 Cardano 公式与 Vieta 替换

对于一般的复系数一元三次方程

$$ax^3+bx^2+cx+d=0$$

首先可以先把它的三次方系数化为 1，成为

$$x^3+ax^2+bx+c=0$$

这里的系数不同于原来的系数，只是为了方便仍然以 a,b,c 标识。然后仿照解二次方程的配方法，将前两项凑配三次方，因为 $(x+1)^3$ 展开后三次方与二次方的系数比是 $1:3$，所以作代换 $y=x+a/3$，可以消掉二次项，成为如下形式

$$x^3+px+q=0$$

假定 $x=u+v$ 是这方程的根，那么变形得

$$(u^3+v^3)+(u+v)(3uv+p)+q=0$$

因为是引入了两个自由变量 u 和 v，可以加上一些额外的限制，这里就限定 $3uv+p=0$，于是便得 $u^3+v^3=-q$，如果令 $U=u^3,V=v^3$，就有

$$U+V=-q,\ UV=-\frac{p^3}{27}$$

所以 U 和 V 是下面的一元二次方程的根

$$r^2+qr-\frac{p^3}{27}=0$$

由二次方程的求根公式，有

$$U=\frac{-q+\sqrt{q^2+\frac{4p^3}{27}}}{2},\ V=\frac{-q-\sqrt{q^2+\frac{4p^3}{27}}}{2}$$

再分别将 U 和 V 开三次方，便得出 u 和 v，但有个问题是开三次方就能得出三个 u 和三个 v，$u+v$ 的组合就能有九种，是不是都是三次方程的根呢，不是的，因为有 $3uv=-p$，假定 U 的一个立方根是 u_0，那么它的另外两个立方根就是 ωu_0 和 ω^2u_0，这里 $\omega=\cos\frac{2\pi}{3}+i\sin\frac{2\pi}{3}=-\frac{1}{2}+\frac{\sqrt{3}}{2}i$ 是 1 的一个立方根，同样有 V 的三个立方根是 v_0、ωv_0 和 ω^2v_0，合理选取 u_0 和 v_0 使得 $3u_0v_0=-p$，那么由条件 $3uv=-p$ 的限制，U 和 V 各自的三个立方根只能这样搭配:

$$\begin{aligned}x_1 &= u_0+v_0\\ x_2 &= \omega u_0+\omega^2v_0\\ x_3 &= \omega^2u_0+\omega v_0\end{aligned}$$

这便是三次方程 $x^3+px+q=0$ 的全部三个解，也可以把它写成下面这样

$$x_{1,2,3}=\sqrt[3]{\frac{-q+\sqrt{q^2+\frac{4p^3}{27}}}{2}}+\sqrt[3]{\frac{-q-\sqrt{q^2+\frac{4p^3}{27}}}{2}} \tag{3.2.1}$$

这公式称为 *Cardano* 公式。

注意到在解方程 $x^3+px+q=0$ 时所作的代换 $x=u+v$ 并且后面限定了 $3uv=-p$，所以实际上我们是作了代换

$$x=u-\frac{p}{3u} \tag{3.2.2}$$

这替换称为 *Vieta* 替换，在这替换下，方程 $x^3+px+q=0$ 化为

$$u^3-\frac{p^3}{27}\cdot\frac{1}{u^3}+q=0 \tag{3.2.3}$$

这可以化为一个关于 u^3 的二次方程，于是 u^3 可解出 (两个复数值)，再开立方便得出 u 的值，然后 $x=u-\frac{p}{3u}$ 就是方程 $x^3+px+q=0$ 的解，这样就算是解出了三次方程。但是有个问题是，u^3 是从二次方程解出来的，所以有两个复数值，而对每一个进行开立方都会产生三个复数根，所以最后 u 会有 6 个值，是不是这三次方程就会有 6 个解呢？当然不是，首先可以求出 u^3 的两个值为

$$u^3=\frac{-q\pm\sqrt{q^2+\frac{4p^3}{27}}}{2}$$

假定 u_0 满足

$$u_0^3=\frac{-q+\sqrt{q^2+\frac{4p^3}{27}}}{2}$$

不难得到

$$\left(-\frac{p}{3u_0}\right)^3=\frac{-q-\sqrt{q^2+\frac{4p^3}{27}}}{2}$$

这就是说，如果 u_0 满足方程3.2.3，那么 $-p/(3u_0)$ 也满足它，所以方程3.2.3的六个解是可以搭配为三对的，而对于 Vieta 替换 $x=u-p/(3u)$，如果记 $h(u)=u-p/(3u)$，则计算便知

$$h(u)=h(-\frac{p}{3u})$$

所以刚才的三对 u，每一对中的 u 经 Vieta 替换都得出相同的 x，所以最终方程 $x^3+px+q=0$ 只有三个复数解。

3.2.2 三次方程的三角解法

仍然只考虑方程 $x^3+px+q=0$，作代换 $x=u\cos\theta$，得方程

$$u^3\cos^3\theta+pu\cos\theta+q=0$$

熟知 $\cos 3\theta=4\cos^3\theta-3\cos\theta$，所以为了将上述方程中的余弦归到同一项，令

$$\frac{u^3}{pu}=\frac{4}{-3}$$

得到 $u=2\sqrt{-p/3}$，而原方程则化为

$$-\frac{2p}{3}\sqrt{-\frac{p}{3}}\cos 3\theta+q=0$$

于是就得出了 $\cos 3\theta$ 的值

$$\cos 3\theta=-\frac{3q}{2p}\sqrt{-\frac{3}{p}}$$

这里需要说明的一点是，我们的三角函数目前为止都是限制在实数范围内的，复数的余弦还是一个没有定义的概念，对此这里有两种方案，一种方案是认可方程中的系数 p 和 q 都是实数，也就是说我们只讨论实系数三次方程的三角形式的解，但是要注意的是上面的 $\cos 3\theta$ 的值中有开方，所以即便是实系数的三次方程，上式右端也可能是一个虚数，所以这种方案并不理想，另一方案是我们认为三角函数已经推广到复数，并认为三角函数在实数范围内的运算规则对复数仍然成立，这样就不用关注到底是实数还是复数了，反正运算法则是一样的。

得到了 $\cos 3\theta$ 的值，我们利用反三角求出角度 3θ，并进一步得出 θ

$$\theta=\pm\frac{1}{3}\arccos\left(-\frac{3q}{2p}\sqrt{-\frac{3}{p}}\right)+\frac{2}{3}k\pi,\ (k=0,1,\cdots)$$

于是原三次方程的解 $x=u\cos\theta$ 便可以表为

$$x=2\sqrt{-\frac{p}{3}}\cos\left(\pm\frac{1}{3}\arccos\left(-\frac{3q}{2p}\sqrt{-\frac{3}{p}}\right)+\frac{2}{3}k\pi\right),\ (k=0,1,2)$$

注意到由余弦的周期性，上面的整数 k 只需要取遍 $0,1,2$ 便可以得到三次方程的三个解，这就是三次方程的三角形式的解。

3.2.3 四次方程的解法

3.2.4 牛顿迭代法

3.2.5 题选

3.3 多项式

这一节是参考文献 [4] 的学习笔记，讨论了多项式的整除理论与及因式分解相关内容。

3.3.1 数域

域本来是抽象代数中的一个概念，但早点接触它还是有好处的。

定义 3.3.1. *如果一个数集包含了 0 和 1，并且对集合中任意两数 (可以相同) 进行加减乘除运算所得的结果都仍然在这集合中 (称为对这四种运算具有封闭性)，则称该数集为一个* 数域.

易知任何数域都包含有理数集作为它的一个子集，常见的有理数集、实数集、复数集都是数域，但数域是一个更宽泛的概念，例如下面这个例子。

例 3.3.1 设 π 是一个无理数，则定义数集 $A_\pi = \{x|x = a + b\pi, a, b \in Q\}$，这里 Q 是有理数集，易证这是一个数域，其中除了有理数外，还包含一些无理数，这些无理数都跟一个确定的无理数 π 有关，这个数集可以称为是由无理数 π 生成的 *最小数域*，这类数域在构造某些特例时是有用的。 ■

定义 3.3.2. *系数都在某个数域 P 中的多项式 $f(x)$ 称为数域 P 上的多项式。*

本节所讨论的多项式都是针对某个数域 P 上的多项式而言的，在无特别说明时，这个数域 P 可以是任何数域，当然把它单单理解为实数集和复数集，对结论也没有什么影响。

3.3.2 一元多项式

3.3.3 整除的概念与带余除法

多项式的整除理论与整数的整除理论是两套平行的理论，诸多定理和性质基本上是完全相同的，用抽象代数的语言来说，就是具有相同的代数结构，因而这两套理论可以对照着学习。

定理 3.3.1 (带余除法)**.** *对数域 P 上的任意两个非零多项式 $f(x)$ 和 $g(x)$，存在数域 P 上的另外两个多项式 $q(x)$ 和 $r(x)$，其中 $r(x)$ 次数低于 $g(x)$ 或者是零多项式，使得下*

式成立

$$f(x) = q(x)g(x) + r(x)$$

并且 $q(x)$ 及 $r(x)$ 是唯一的。

证明. 先证明存在性。

如果 $f(x)$ 的次数低于 $g(x)$，则取 $u(x) = 0, r(x) = f(x)$ 就可以了，以下证明 $f(x)$ 比 $g(x)$ 次数高的情况，设

$$f(x) = \sum_{i=0}^{n} a_i x^i,\ g(x) = \sum_{i=0}^{n} b_i x^i$$

其中 $n > m > 0$，则首先用

$$u_0(x) = \frac{a_n}{b_m} x^{n-m}$$

与 $g(x)$ 相乘得出一个 n 次多项式，于是这个多项式与 $f(x)$ 的差就是一个次数低于 n 的多项式，即

$$f(x) = u_0(x)g(x) + f_1(x)$$

如果 $f_1(x)$ 的次数仍然高于或者等于 $g(x)$，则我们再对 $f_1(x)$ 施以同样的手法可得出第二个等式

$$f_1(x) = u_1(x)g(x) + f_2(x)$$

依次下去，必定能够在有限步之内 (最多 $n-m+1$ 步) 使得某个 $f_r(x)$ 的次数低于 $g(x)$ 或者成为零多项式，这时便有

$$f(x) = g(x)\sum_{i=0}^{r-1} u_i(x) + f_r(x)$$

于是取

$$q(x) = \sum_{i=0}^{r-1} u_i(x),\ r(x) = f_r(x)$$

就可以满足定理要求。

再证明唯一性，设有两组符合定理中等式的 $q(x)$ 和 $r(x)$，即有

$$f(x) = q_1(x)g(x) + r_1(x) = q_2(x)g(x) + r_2(x)$$

于是便有

$$(q_1(x) - q_2(x))g(x) = r_1(x) - r_2(x)$$

上式右端的次数低于 $g(x)$ 的次数，因此左边必须有 $q_1(x) = q_2(x)$，这时也就必然有 $r_1(x) = r_2(x)$，唯一性得证。 □

这个证明过程实际上就是多项式除法的运算过程，定理中的 $q(x)$ 称为 $f(x)$ 除以 $g(x)$ 的商，而 $r(x)$ 则为余式.

定义 3.3.3. 对于多项式 $f(x)$ 和 $g(x)$，如果存在多项式 $h(x)$，使得 $f(x)=g(x)h(x)$，则称 $g(x)$ 能够整除$f(x)$，记为 $g(x)\mid f(x)$，此时称 $g(x)$ 是 $f(x)$ 的因式，而 $f(x)$ 则称为 $g(x)$ 的倍式.

多项式整除有以下性质.

性质 3.3.1. 如果多项式 $f(x)$ 和 $g(x)$ 能够互相整除，则它俩只相差一个常数因子，即 $f(x)=cg(x)$，c 为常数。

证明. 因为存在多项式 $h_1(x)$ 和 $h_2(x)$，使得 $f(x)=g(x)h_1(x)$ 和 $g(x)=f(x)h_2(x)$，所以 $h_1(x)h_2(x)=1$，于是 $h_1(x)$ 和 $h_2(x)$ 都只能是常数，得证。 □

还有以下两个性质，证明略。

性质 3.3.2. 如果 $f(x)\mid g(x)$ 且 $g(x)\mid h(x)$，则 $f(x)\mid h(x)$.

性质 3.3.3. 如果 $f(x)\mid g(x)(i=1,2,\ldots,n)$，则有

$$f(x)\mid\sum_{i=1}^{n}u_i(x)g_i(x)$$

式中 $u_i(x)(i=1,2,\ldots,n)$ 是任意多项式。

定理 3.3.2. 多项式 $g(x)$ 能够整除多项式 $f(x)$ 的充分必要条件是，在带余除法等式

$$f(x)=q(x)g(x)+r(x)$$

中的余式 $r(x)$ 是零多项式。

证明. 充分性显然，只证必要性，由 $g(x)\mid f(x)$，知 $g(x)$ 能够整除带余除法等式中三个项的其中两个项，因而也能整除另外一项，即 $g(x)\mid r(x)$，但 $r(x)$ 的次数低于 $g(x)$，所以只能是零多项式，必要性成立。 □

3.3.4 最大公因式与辗转相除法

定义 3.3.4. 如果 $d(x)\mid f(x)$ 且 $d(x)\mid g(x)$，则称 $d(x)$ 是 $f(x)$ 与 $g(x)$ 的一个 公因式.

显然常数因子是任何两个多项式的公因式，因而公因式是存在的。

定义 3.3.5. 多项式 $f(x)$ 与 $g(x)$ 的诸公因式中，次数最高的那一个称为这两个多项式的最大公因式，记为 $(f(x),g(x))$.

这个定义的问题是，有没有可能同时存在多个次数都最高的公因式 (只相差一个常数因子的视为同一个)，也就是最大公因式的唯一性问题。

而流行的定义是:

定义 3.3.6. 设 $d(x)$ 是 $f(x)$ 与 $g(x)$ 的一个公因式，如果 $f(x)$ 与 $g(x)$ 的所有公因式都是 $d(x)$ 的因式，则称 $d(x)$ 是 $f(x)$ 与 $g(x)$ 的最大公因式，记为 $(f(x),g(x))$.

这个定义也有个问题，在这诸公因式中，次数最高的那个，是否能被别的所有公因式都整除呢，所以无论采用哪一种定义，都有些问题要留待对最大公因式有一定程度的讨论后才能解决，这就是最大公因式的性质定理:

定理 3.3.3 (最大公因式性质定理). *设 $d(x)$ 是 $f(x)$ 与 $g(x)$ 的最大公因式，则存在多项式 $u(x)$ 和 $v(x)$，使得*

$$d(x) = u(x)f(x) + v(x)g(x)$$

为证明这个定理，先提出如下引理

引理 3.3.1. *如果有等式*

$$f(x) = q(x)g(x) + r(x)$$

则 $f(x)$ 与 $g(x)$ 的公因式，也必然是 $g(x)$ 与 $r(x)$ 的公因式.

由整除性质，引理是显然的。

现在证明定理 3.3.3:

证明. 根据带余除法，存在多项式 $q_1(x)$ 及 $r_1(x)$(次数低于 $g(x)$ 或者是零多项式)，使

$$f(x) = q_1(x)g(x) + r_1(x)$$

如果 $r_1(x)$ 不是零多项式，则再继续拿 $g(x)$ 除以 $r_1(x)$，得出

$$g(x) = q_2 r_1(x) + r_2(x)$$

如此反复下去，因为每进行一次，$r_i(x)$ 的次数至少减少一，因此必然在有限步之后，$r_i(x)$ 成为零次多项式即为常数，再进行一次之后，$r_i(x)$ 便成为零多项式，把这过程写成等式序列，并记 $r_{-1}(x) = f(x)$,$r_0(x) = g(x)$，便是

$$\begin{array}{rcl}
r_{-1}(x) & = & q_1(x)r_0(x) + r_1(x) \\
r_0(x) & = & q_2(x)r_1(x) + r_2(x) \\
\cdots & & \\
r_{k-1}(x) & = & q_{k+1}(x)r_k(x) + r_{k+1}(x) \\
\cdots & & \\
r_{s-2}(x) & = & q_s(x)r_{s-1}(x) + r_s(x) \\
r_{s-1}(x) & = & q_{s+1}(x)r_s(x) + 0
\end{array}$$

按照上述引理，$f(x)$ 与 $g(x)$ 的公因式，必然也是 $g(x)$ 与 $r_1(x)$ 的公因式，也就必然是 $r_1(x)$ 和 $r_2(x)$ 的公因式，依次推下去，最后必然也是 $r_s(x)$ 与 0 的公因式，从而就必然是 $r_s(x)$ 的因式，于是 $r_s(x)$ 本身便是最大公因式（无论按照前面的两种定义的那一种)。

从上面倒数第二个等式可以看出，$r_s(x)$ 可以用 $r_{s-1}(x)$ 与 $r_{s-2}(x)$ 表示成各自与某个多项式相乘后再相加的形式，而 $r_{s-1}(x)$ 又可以用 $r_{s-2}(x)$ 和 $r_{s-3}(x)$ 用相同的形

式表示出来，依次倒着推回去，最后 $r_s(x)$ 便必定可以用 $f(x)$ 与 $g(x)$ 各自与某个多项式相乘后相加的形式表示出来，当然这也可以用数学归纳法的形式进行严格叙述，略去。 □

定理证明过程中的这个反复做带余除法的过程称为 *辗转相除法*，可以用它来求两个多项式的最大公因式。

从定理证明过程中还可以看出，两个多项式的公因式，都是它俩最大公因式的因式，从而最大公因式在不考虑常数因式的意义下是唯一的。

要说明的是，定理中的 $u(x)$ 及 $v(x)$ 不是唯一的，这从下式中可以看出

$$d(x) = (u(x) + h(x)g(x))f(x) + (v(x) - h(x)f(x))g(x)$$

即如果 $u(x)$、$v(x)$ 符合定理，则 $u(x) + h(x)g(x)$、$v(x) - h(x)f(x)$ 也符合定理，其中 $h(x)$ 是任意多项式。

但是要强调的是，只有最大公因式才能表示成这种形式，其它公因式不具有这种表示，这就是下面的定理

定理 3.3.4. *设 $d(x)$ 是 $f(x)$ 和 $g(x)$ 的一个公因式，如果存在多项式 $u(x)$ 和 $v(x)$ 使得 $d(x) = u(x)f(x) + v(x)g(x)$，则 $d(x)$ 必然是最大公因式。*

证明. 显然根据那等式，$f(x)$ 与 $g(x)$ 的任一公因式也能够整除 $d(x)$，所以 $d(x)$ 是最大公因式。 □

定义 3.3.7. *如果两个多项式的最大公因式为 1，即 $(f(x), g(x)) = 1$，则称这两个多项式* 互素.

定理 3.3.5. *两个多项式 $f(x)$ 与 $g(x)$ 互素的充分必要条件是，存在多项式 $u(x)$ 和 $v(x)$ 使得 $u(x)f(x) + v(x)g(x) = 1$.*

证明. 如果有这等式，则对于 $f(x)$ 与 $g(x)$ 的最大公因式 $d(x)$，必然有 $d(x) \mid 1$，所以两个多项式互素，充分性得证，而必要性是显然的。 □

定理 3.3.6. *如果 $(f(x), g(x)) = 1$，且 $f(x) \mid g(x)h(x)$，则 $f(x) \mid h(x)$.*

证明. 由互素，存在两个多项式 $u(x)$、$v(x)$ 使得 $u(x)f(x) + v(x)g(x) = 1$，两边同乘 $h(x)$ 得

$$u(x)f(x)h(x) + v(x)g(x)h(x) = h(x)$$

显然 $f(x)$ 能够整除等式左边的两项，因而也能整除右边，得证。 □

推论 3.3.1. *如果 $(f(x), g(x) = 1)$，且 $f(x) \mid h(x)$，$g(x) \mid h(x)$，则 $f(x)g(x) \mid h(x)$.*

证明. 设 $h(x) = f(x)r(x)$，则 $g(x) \mid f(x)r(x)$，按前述定理，这时必有 $g(x) \mid r(x)$，于是 $f(x)g(x) \mid f(x)r(x) = h(x)$. □

证明二. 由条件，存在多项式 $u(x)$ 和 $v(x)$ 使 $u(x)f(x)+v(x)g(x)=1$，又设 $h(x)=f(x)f_1(x)$，在前等式两边同乘 $f_1(x)$ 可得 $u(x)f(x)f_1(x)+v(x)g(x)f_1(x)=f_1(x)$，即 $u(x)h(x)+v(x)g(x)f_1(x)=f_1(x)$，显然 $g(x)$ 能整除等式的左边，因而也能整除右边，即 $g(x)\mid f_1(x)$，从而 $f(x)g(x)\mid f(x)f_1(x)=h(x)$. □

推论 3.3.2. 如果 $(f(x),g(x))=1$ 且 $(f(x),h(x))=1$，则 $(f(x),g(x)h(x))=1$.

证明一. 设 $d(x)=(f(x),g(x)h(x))$，则由 $(f(x),g(x))=1$ 知 $(d(x),g(x))=1$，而由 $d(x)\mid g(x)h(x)$ 便得 $d(x)\mid h(x)$，于是 $d(x)$ 是 $f(x)$ 与 $h(x)$ 的一个公因式，但是 $(f(x),h(x))=1$，所以 $d(x)=1$. □

证明二. 将等式 $u_1(x)f(x)+v_1(x)g(x)=1$ 与 $u_2(x)f(x)+v_2(x)h(x)=1$ 相乘得

$$[u_1(x)u_2(x)f(x)+u_2(x)v_1(x)g(x)+u_1(x)v_2(x)h(x)]f(x)+v_1(x)v_2(x)g(x)h(x)=1$$

即得结论. □

这个结论还可以推广到更一般的情况:

定理 3.3.7. 设 $f(x)=f_1(x)f_2(x)\cdots f_n(x)$，$g(x)=g_1(x)g_2(x)\cdots g_m(x)$，并且对于任意的 $i(1\leqslant i\leqslant n)$ 和 $j(1\leqslant j\leqslant m)$ 都有 $(f_i(x),g_j(x))=1$，则 $(f(x),g(x))=1$.

证明. 因为每一个 $f_i(x)$ 都与所有的 $g_j(x)$ 互素，所以 $(f_i(x),g(x))=1(i=1,2,\ldots,n)$，进一步就有 $(f(x),g(x))=1$. □

最大公因式可以推广到多个多项式的情形。

定义 3.3.8. 如果 $d(x)$ 是一组多项式 $f_i(x)(i=1,2,\ldots,n)$ 的公因式，并且这组多项式的任一公因式都是 $d(x)$ 的因式，则称 $d(x)$ 是这组多项式的公因式，也记为 $d(x)=(f_1(x),f_2(x),\ldots,f_n(x))$.

定理 3.3.8. 对一组多项式 $f_i(x)(i=1,2,\ldots,n)$，定义递推序列

$$d_1(x)=f_1(x),\ d_k(x)=(d_{k-1}(x),f_k(x))$$

则 $d_n(x)$ 便是这组多项式的最大公因式。

依据这定理，最大公因式可以逐次求，先求出 $f_1(x)$ 与 $f_2(x)$ 的最大公因式 $d_2(x)$，再求 $d_2(x)$ 与 $f_3(x)$ 的最大公因式，逐次下去，最后求得 $d_{n-1}(x)$ 与 $f_n(x)$ 的最大公因式 $d_n(x)$ 就是这一组多项式的最大公因式。

证明. 记 $r_k(x)=(f_1(x),f_2(x),\ldots,f_k(x))$，易知 $d_k(x)\mid f_i(x)(i=1,2,\ldots,k)$，从而 $d_k(x)\mid r_k(x)$，另一方面，$r_k(x)$ 作为 $f_1(x),f_2(x),\ldots,f_{k-1}(x)$ 的一个公因式，也必然是 $d_{k-1}(x)$ 的因式，所以 $r_k(x)\mid d_{k-1}(x)$，又 $r_k(x)\mid f_k(x)$，所以 $r_k(x)\mid d_k(x)$，因此 $d_k(x)=r_k(x)(i=1,2,\ldots,n)$（不计常数因子）。 □

同样有性质定理

定理 3.3.9. *设一组多项式 $f_i(x)(i=1,2,\dots,n)$ 的最大公因式是 $d(x)$，则存在多项式 $u_i(x)(i=1,2,\dots,n)$ 使得*

$$d(x)=\sum_{i=1}^{n}u_i(x)f_i(x)$$

证明. 对多项式的个数使用数学归纳法，$n=2$ 时显然成立，假若对于 $n-1$ 个多项式也成立，则对于 n 个多项式的情形，由前述定理，$d_n(x)=(d_{n-1}(x),f_n(x))$，因而存在多项式 $p(x)$ 和 $u_n(x)$ 使得

$$d_n(x)=p(x)d_{n-1}(x)+u_n(x)f_n(x)$$

而根据归纳假设，存在 $p_i(x)(i=1,2,\dots,n-1)$，使得

$$d_{n-1}(x)=\sum_{i=1}^{n-1}p_i(x)f_i(x)$$

代入前式即得结论。 □

同样的，有多个项式互素的定义

定义 3.3.9. *如果一组多项式的最大公因式是 1，则称它们互素。*

一组多项式互素与它们两两互素是不同的，两两互素是更强的约束。

定理 3.3.10. *一组多项式 $f_i(x)(i=1,2,\dots,n)$ 互素的充分必要条件是存在一组多项式 $u_i(x)(i=1,2,\dots,n)$，使得*

$$\sum_{i=1}^{n}u_i(x)f_i(x)=1$$

证明略。

3.3.5 因式分解定理

把一个多项式写成它的因式的乘积叫做 *因式分解*.

定义 3.3.10. *如果数域 P 上一个多项式不能被分解成同一数域上的两个比它次数低的多项式的乘积，则称它是一个* 不可约多项式，*或者说该多项式在数域 P 上*不可约.

按定义，不可约多项式的因式只有 1 与它本身 (不考虑常数因子)。

一个多项式是否是不可约多项式与所讨论的数论有关，比如多项式 x^4-4 在有理数域上可以被分解为 $(x^2-2)(x^2+2)$，这两个因子都是有理数域上的不可约多项式，但在实数域上可以进一步分解为

$$(x-\sqrt{2})(x+\sqrt{2})(x^2+2)$$

这些因子都是实数域上的不可约多项式，而在复数域上则还能被继续分解为

$$(x-\sqrt{2})(x+\sqrt{2})(x-\sqrt{2}i)(x+\sqrt{2}i)$$

所以，关于一个多项式还能不能被继续分解，只有在指定了数域后才有意义。

不可约多项式在多项式整除理论中的作用，与整数的整除理论中素数的作用类似。

定理 3.3.11. 设 $f(x)$ 是数域 P 上的不可约多项式，$g(x)$ 是该数域上的任一多项式，则要么 $f(x) \mid g(x)$，要么 $(f(x), g(x)) = 1$.

证明. 设 $d(x) = (f(x), g(x))$，则由于 $f(x)$ 不可约，$d(x)$ 作为 $f(x)$ 的因式，要么是 1，要么便是 $f(x)$ 自身，如果是前者，则 $(f(x), g(x)) = 1$，如果是后者，则 $f(x) \mid g(x)$. □

推论 3.3.3. 设 $f(x)$ 为不可约多项式，如果 $f(x) \mid g(x)h(x)$，则 $f(x) \mid g(x)$ 或者 $f(x) \mid h(x)$.

证明. 如果 $f(x) \mid g(x)$，自然结论成立，若是 $f(x)$ 不能整除 $g(x)$，由 $f(x)$ 不可约，则只能 $(f(x), g(x)) = 1$，于是按照定理 3.3.6就有 $f(x) \mid h(x)$，即得证。 □

利用数学归纳法，这结论可以推广到多个数的情形

推论 3.3.4. 如果 $f(x)$ 是不可约多项式，且 $f(x) \mid g_1(x)g_2(x)\cdots g_n(x)$，则 $f(x)$ 必定整除诸 $g_i(x)$ 中至少一个。

如果一个多项式不是不可约多项式，则它可以表为两个次数较低的多项式之积，若这两个多项式还有不可约的，则再继续进行分解，由于每一次分解都会降低某一因式的次数，所以这个过程不可能无限的进行下去，直到所有因式都成为不可约多项式为止，于是我们得到以下重要的结果

定理 3.3.12 (多项式因式分解定理). 数域 P 上的任一非零多项式 $f(x)$ 都可以被分解为同一数域上的一些不可约多项式的乘积

$$f(x) = p_1(x)p_2(x)\cdots p_n(x)$$

并且此分解在不考虑常数因式及各因式的顺序的情况下是唯一的。

证明. 先证分解式的存在性，对多项式 $f(x)$ 的次数作归纳，如果 $f(x)$ 的次数为一，结论是显然的，因为一次多项式在任何数域上都是不可约的，假定任何次数小于 n 的多项式都能被分解为一些次数更低的不可约多项式之积，则对于任一次数为 n 的多项式 $f(x)$，如果它本身就是不可约的，则结论显然是成立的，如果它不是不可约的，则它能被分解为两个次数低于 n 的因式之积，按照归纳假设，这两个因式都能被分解，于是把这两个因式的分解式合并起来，就得到 $f(x)$ 的分解式。

再证分解式的唯一性，设 $f(x)$ 同时还有如下的分解

$$f(x) = q_1(x)q_2(x)\cdots q_m(x)$$

其中 $q_i(x)$ 都是数域 P 上的不可约多项式，则有

$$p_1(x)p_2(x)\cdots p_n(x) = q_1(x)q_2(x)\cdots q_m(x)$$

由于左边每一个 $p_i(x)$ 都能整除右边，而 $p_i(x)$ 不可约，于是必定能整除某个 $q_j(x)$，但是 $q_j(x)$ 也是不可约的，所以只能 $p_i(x) = cq_j(x)$，即在允许相差一个常数因子的意义下是相等的，于是左边每一个 $p_i(x)$ 都能与右边的某个 $q_j(x)$ 对应上，同理右边每一个

$q_i(x)$ 也能与左边某一个 $p_j(x)$ 对应上，但这还不能说明左右两边是相同的，还必须证明在等式的两边，每一个不可约多项式 (相差常数因子的视为相同) 重复出现的次数出是一样的。假定某个不可约多项式 $h(x)$ 在左右两边重复出现的次数分别是 s 和 t，则 $h^s(x)$ 能整除右边，根据定理 3.3.7知，$h^s(x)$ 与右边除去 $h^t(x)$ 之外的部分积是互素的，于是必然有 $h^s(x) \mid h^t(x)$，同理也有 $h^t(x) \mid h^s(x)$，因而只能 $s=t$，所以这两个分解式是一样的。 □

在定理的分解式中，把每一个因式的最高次项系数都提出来，然后把相同的因式写成幂的形式，就得到如下的 *标准分解式*:

$$f(x) = cp_1^{\alpha_1}(x)p_2^{\alpha_2}(x)\cdots p_n^{\alpha_n}(x),\ \alpha_i > 0(i=1,2,\ldots,n)$$

对于两个多项式，把它们的分解式中出现的因式都并起来，原来并不出现的因式标之以零次幂，则得到如下的表示

$$f(x) = cp_1^{\alpha_1}(x)p_2^{\alpha_2}(x)\cdots p_n^{\alpha_n}(x),\ \alpha_i \geqslant 0(i=1,2,\ldots,n)$$

和

$$g(x) = dp_1^{\beta_1}(x)p_2^{\beta_2}(x)\cdots p_n^{\beta_n}(x),\ \beta_i \geqslant 0(i=1,2,\ldots,n)$$

在此基础上，可导出如下一些重要结论:

定理 3.3.13. *设两个多项式 $f(x)$ 与 $g(x)$ 有如上分解式，则 $f(x) \mid g(x)$ 的充分必要条件是 $\alpha_i \leqslant \beta_i(i=1,2,\ldots,n)$.*

证明. 充分性显然，只证必要性，如果 $f(x) \mid g(x)$，则 $p_i^{\alpha_i}(x) \mid g(x)$，而根据定理 3.3.7，有 $p_i^{\alpha_i}(x)$ 与 $g(x)$ 分解式中除 $p_i^{\beta_i}(x)$ 以外剩余的部分是互素的，所以 $p_i^{\alpha_i}(x) \mid p_i^{\beta_i}(x)$，所以 $\alpha_i \leqslant \beta_i$. □

因为 $f(x)$ 能被它的因式整除，所以由上述定理便得到

定理 3.3.14. *设多项式 $f(x)$ 的标准分解式为*

$$f(x) = cp_1^{\alpha_1}(x)p_2^{\alpha_2}(x)\cdots p_n^{\alpha_n}(x),\ \alpha_i > 0(i=1,2,\ldots,n)$$

则它的所有因式都具有形式

$$d(x) = dp_1^{\beta_1}(x)p_2^{\beta_2}(x)\cdots p_n^{\beta_n}(x),\ 0 \leqslant \beta_i \leqslant \alpha_i(i=1,2,\ldots,n)$$

进一步便可得出

定理 3.3.15. *设 $f(x)$ 和 $g(x)$ 具有标准分解式*

$$f(x) = cp_1^{\alpha_1}(x)p_2^{\alpha_2}(x)\cdots p_n^{\alpha_n}(x),\ \alpha_i \geqslant 0(i=1,2,\ldots,n)$$

和

$$g(x) = dp_1^{\beta_1}(x)p_2^{\beta_2}(x)\cdots p_n^{\beta_n}(x),\ \beta_i \geqslant 0(i=1,2,\ldots,n)$$

则它俩的最大公因式是

$$d(x) = p_1^{\gamma_1}(x)p_2^{\gamma_2}(x)\cdots p_n^{\gamma_n}(x),\ \gamma_i = \min\{\alpha_i,\beta_i\}(i=1,2,\ldots,n)$$

3.3.6　重因式

3.3.7　多项式函数

在此讨论一下有理系数多项式的有理根的问题，因为有理系数多项式方程总可以化为一个整系数多项式方程，所以只要讨论整系数多项式的根就行了，这时我们有以下定理

定理 3.3.16. *整系数 n 次多项式*

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

如果有一个有理根 r/s(r、s 互素)，则 $r \mid a_0$，$s \mid a_n$，在最高次项系数 $a_n = 1$ 的特殊情况下，它的有理数都只能是整数根，而且这些根都是常数项 a_0 的因数。

证明. 由条件得

$$a_n \left(\frac{r}{s}\right)^n + a_{n-1} \left(\frac{r}{s}\right)^{n-1} + \cdots + a_1 \frac{r}{s} + a_0 = 0$$

整理得

$$a_n r^n + a_{n-1} r^{n-1} s + \cdots + a_1 r s^{n-1} + a_0 s^n = 0$$

左边除最后一项外都能被 r 整除，所以 $r \mid a_0 s^{n-1}$，而 r 与 s 互素，所以 $r \mid a_0$，同样，左边除第一项外都能被 s 整除，所以 $s \mid a_n r^n$，从而 $s \mid a_n$. □

3.3.8　对称多项式

3.3.9　插值多项式

3.3.10　题选

题 3.3.1　(台湾台南吴政哲) 已知 a、b、c 是三次方程 $x^3 + 2x^2 + 3x + 4 = 0$ 的三个复根，求表达式 $\dfrac{a}{b} + \dfrac{b}{c} + \dfrac{c}{a}$ 的值. ■

解答. 一眼看上去容易想到对称多项式和根与系数的关系，但是

$$\frac{a}{b} + \frac{b}{c} + \frac{c}{a} = \frac{ab^2 + bc^2 + ca^2}{abc}$$

分子并不是对称多项式，仅仅是轮换多项式，记 $M = ab^2 + bc^2 + ca^2$，再记其对偶式 $N = a^2 b + b^2 c + c^2 a$，可以验证 $M + N$、MN 都是对称多项式

$$\begin{aligned} M + N &= (ab^2 + bc^2 + ca^2) + (a^2 b + b^2 c + c^2 a) \\ &= \frac{1}{3}(a + b + c)^3 - \frac{1}{3}(a^3 + b^3 + c^3) - 2abc \end{aligned}$$

$$\begin{aligned} MN &= (ab^2 + bc^2 + ca^2)(a^2 b + b^2 c + c^2 a) \\ &= abc(a^3 + b^3 + c^3) + (a^3 b^3 + b^3 c^3 + c^3 a^3) + 3a^2 b^2 c^2 \end{aligned}$$

因此，$M + N$ 和 MN 都可以利用根与系数的关系求出，从而 M 可求出，但结果应该是两个值。 □

题 3.3.2 已知整系数四次方程 $x^4+ax^3+bx^2+cx+d=0$ 的一个根是 $1+\sqrt{2}+\sqrt{3}$，求常数项 d 的值. ■

解答. 记 $\alpha=1+\sqrt{2}+\sqrt{3}$，则

$$(\alpha-1)^2=(\sqrt{2}+\sqrt{3})^2=5+2\sqrt{6}$$

所以 $((\alpha-1)^2-5)^2=24$，即

$$\alpha^4-4\alpha^3-4\alpha^2+16\alpha-8=0$$

这就是 α 所满足的一个四次方程，接下来就是讨论下这个方程与题目中的四次方程有什么样的关系，因为 α 有两个根号，所以猜测这两个方程的系数对应成比例，也就是说是同一个方程，为了说明这一点，先来证明 α 不可能是一个次数低于四并具有有理系数的方程的根，反证法，假如 α 满足下面的有理系数方程

$$ax^3+bx^2+cx+d=0$$

那么作平移替换 $x\to x+\frac{b}{3a}$ 后即可消去其中的二次项，成为

$$ax^3+cx+d=0$$

注意现在的系数已经不再是之前的系数了，但仍然是有理数，且在这平移下，α 被变换为 $p+q\sqrt{2}+r\sqrt{3}$ 的形式 (其中 p，q，r 都是有理数)，并不影响接下来的过程。

将 $\alpha=p+q\sqrt{2}+r\sqrt{3}$ 代入上述方程，可知左边三次方展开后会出现 $\sqrt{6}$ 的无理项，且这一项无法在后面得到抵消，所以只能让这一项的系数等于零，从而就有 $a=0$，即变为一次方程，但是有理系数的一次方程是不可能有无理根的，所以这就证明了，α 不可能是一个次数低于四次的有理系数方程的根。

有了这一点就可以证明一开头构造的四次方程与题目中方程必然是系数成比例的了，因为若不然的话，将两个方程的最高次项系数配成相同后相减，则得出一个次数低于四次的有理系数方程，而且 α 必然是这个方程的解，这与前面已经证明的结论相矛盾。

于是最终的结果是，两个方程系数成比例，所以 $d=-8$. □

注：1. 这个题目的背景就是根式的最小多项式问题，即对于一个给定的根式组合 $\alpha=a_1\sqrt[m_1]{r_1}+a_2\sqrt[m_2]{r_2}+\cdots+a_n\sqrt[m_n]{r_n}$，其中 r_i 是互不相同的整数且都没有平方因子，求一个次数最小的有理系数多项式，使它以 α 为其一根，这多项式称为这个根式组合的最小多项式，它的次数也称为这个根式组合的次数，上面这个解答过程其实就是证明了，对于含有两个二次根号的情况 $\alpha=a_1\sqrt{r_1}+a_2\sqrt{r_2}$，它的最小多项式是四次的。

2. 解答过程中用了一个结论，即如果表达式 $a_1\sqrt{r_1}+a_2\sqrt{r_2}+\cdots+a_n\sqrt{r_n}$(约定同上，但全部限定为二次根式) 为有理数，则根式的系数全部为零，换句话说，这表达式的系数若不全为零，则它一定是无理数，这个结论的本质是说，一个最简底数的二次根式是无法用别的最简二次根式的线性组合来表达的。这结论的证明暂没想到，可能需要用到抽象代数的内容。

3.4 函数方程

函数方程，顾名思义是关于函数的方程，它是以函数为未知量的方程，例如方程 $f(x+y)=f(x)+f(y)$，这个方程中的未知量是函数 $f(x)$. 如果某个函数能够满足某一函数方程，则称它是这函数方程的一个解，例如任意奇函数都是函数方程 $f(x)+f(-x)=0$ 的解，求出函数方程的全部解的过程称为解函数方程。

在解函数方程时，有时我们会加上连续性的假设，也就是考虑函数方程的连续解，因为非连续解往往没有通式，而且讨论的意义也不大。

3.4.1 柯西方程

这一节介绍柯西研究过的一个函数方程: 定义在实数集 R 上的 $f(x)$ 对于任意两个实数 x 和 y（可以相等）都成立 $f(x+y)=f(x)+f(y)$，我们要讨论的问题不外乎几点：这样的函数存在吗？如果存在，是否唯一？函数表达式是否有通式？能否求出？

显然任意正比例函数 $f(x)=ax$ 都满足这函数方程，所以问题是还有没有其它形式的解。

首先取 $x=y=0$ 得出 $f(0)=0$，于是 $0=f(0)=f(x+(-x))=f(x)+f(-x)$，所以 $f(x)$ 是奇函数。

再令 $x=y$ 得 $f(2x)=2f(x)$，不难根据数学归纳法得到，对任意正整数 n 和任意实数 x，有 $f(nx)=nf(x)$，再由 $f(x)$ 是奇函数，知这等式对于负整数也同样成立。

在等式 $f(nx)=nf(x)$ 中取 $x=1$，便知对于任意整数 n，有 $f(n)=nf(1)$，也就是说，整个整数集上的函数值，全由 $x=1$ 处的函数值决定。

接着转而讨论有理数的情形，对于整数 p，有

$$f(1)=f\left(p\cdot\frac{1}{p}\right)=pf\left(\frac{1}{p}\right)$$

所以

$$f\left(\frac{1}{p}\right)=\frac{1}{p}f(1)$$

进一步，对于任意有理数 $\dfrac{q}{p}$，有

$$f\left(\frac{q}{p}\right)=f\left(q\cdot\frac{1}{p}\right)=qf\left(\frac{1}{p}\right)=\frac{q}{p}f(1)$$

所以，对于一切有理数 x，都有

$$f(x)=xf(1)$$

于是全体有理数集上的函数值，都由 $f(1)$ 所唯一确定。

到目前为止，事情都很顺利，但是当讨论到无理数时，情况就不同了。因为前面之所以顺利，是因为由 0 和 1，经过有限次加减，可得出全体整数，再经过乘除，可以得出全体有理数，但是无理数，却不能这么简单的得出，讨论的函数方程对于开方都无能为力，更不用说对于圆周率 π 这样的无理数了。

任意待定一个无理数 π（为了突出它的无理性，选用了圆周率符号，但并不表示圆周率），定义集合 $Q_\pi = x|x = a + b\pi, a, b \in Q$，称为由无理数 π 生成的无理数集，显然如果 $x_1, x_2 \in Q_\pi$，则 $x_1 + x_2$ 以及 $\lambda x_1(\lambda \in Q)$ 也都属于这集合，设 $x = a + b\pi \in Q_\pi$，称 a 为它的有理部分，b 为它的无理部分（的系数）。

在此定义下，我们将看到，原来的函数方程对于有理数上的取值和无理数上的取值是互相独立的，比如说规定当 $x \in Q_{pi}$ 时，它的函数值等于它的有理部分的函数值，即 $f(a + b\pi) = f(a)$，现在来验证，这样定义的函数符合原来的函数方程，这只要验证 x 和 y 中有无理数就行了。

设 x 为有理数，而 $y = a+b\pi$ 为无理数，则 $f(x+y) = f((x+a)+b\pi) = f(x+a) = f(x) + f(a)$，而 $f(x) + f(y) = f(x) + f(a)$，因此 $f(x + y) = f(x) + f(y)$. 由交换律，x 为无理数而 y 为有理数的情形也是成立的。

再设 $x_1 = a_1 + b_1\pi$，$x_2 = a_2 + b_2\pi$，则 $f(x_1 + x_2) = f((a_1 + a_2) + (b_1 + b_2)\pi) = f(a_1 + a_2) = f(x_1) + f(x_2)$，这就证明了当 x_1 与 x_2 都是无理数且都从属于同一个 Q_π 时，函数方程是成立的。

如果 x_1 与 x_2 分属于不同的 Q_π.

3.4.2 指数方程与对数方程

指数方程是由指数函数所抽象出的方程: $f(x + y) = f(x)f(y)$，显然零函数和任意指数函数都是它的解，现在讨论的一般解。

直观上来看，如果函数能恒保持正号，则可以在方程两端取对数得

$$\ln f(x + y) = \ln f(x) + \ln f(y)$$

再令 $g(x) = \ln f(x)$ 就有 $g(x + y) = g(x) + g(y)$，从而转化为已经研究过的柯西方程，如果 $f(x)$ 又是连续的，那么 $g(x)$ 也是连续的，于是就有 $g(x) = mx(m = g(1))$，从而 $f(x) = e^{mx}$，令 $a = e^m = e^{g(1)} = e^{\ln f(1)} = f(1)$，则 $f(x) = a^x$.

但这个过程要求函数恒为正，现在从这函数方程出发讨论一下它的符号。令 $x = y = 0$ 得 $f(0) = f^2(0)$，因此 $f(0) = 0$ 或 $f(0) = 1$，如果 $f(0) = 0$，则对任意实数 x 都有 $f(x) = f(x + 0) = f(x)f(0) = 0$，即为零函数。在 $f(0) = 1$ 的情况下，对于任意实数 x，有

$$f(x) = f\left(\frac{x}{2} + \frac{x}{2}\right) = f^2\left(\frac{x}{2}\right) \geqslant 0$$

但是我们将指出，上式只能取正，若不然，假定存在某个 x_0 使得 $f(x_0) = 0$，则对任意实数 x 有 $f(x) = f(x_0 + (x - x_0)) = f(x_0)f(x - x_0) = 0$，于是也有 $f(0) = 0$，与刚才假定的 $f(0) = 1$ 矛盾，所以若 $f(0) = 1$，则函数就恒保持为正。

所以最终的结论是，指数方程的连续解为零函数和全体指数函数。

与指数方程类型，对数方程是由对数函数所抽象出来的函数方程: $f(xy) = f(x) + f(y)$.

还是先考虑一下 $f(0)$，虽然对数函数的定义域是正实数集，但单从这个函数方程来说，还是可以考虑一下的。命 $x = y = 0$ 即得 $f(0) = 0$，再取 x 为任意实数而 $y = 0$，

则 $f(0)=f(x)+f(0)$ 对一切实数成立，从而 $f(x)$ 只能为零函数。这就是说，如果要考虑此方程的非零解，则必须将 $x=0$ 从定义域排除掉。

现在假定 $f(x)$ 是连续函数,先考虑函数在正实数上的情形,令 $g(x)=f(e^x)$,则 $g(x)$ 也是连续函数，并且 $g(x+y)=g(x)+g(y)$，于是 $g(x)$ 是柯西方程的解，$g(x)=g(1)x$，从而 $f(x)=g(\ln x)=\log_a x(a=f(e))$.

最后来考虑负数的情形，取 $x=y=1$ 可得 $f(1)=0$，于是再取 $x=y=-1$ 则有 $f(1)=2f(-1)$，从而 $f(-1)=0$，进一步对任意实数 x 就有 $f(-x)=f(-1)f(x)=0$，所以一旦考虑到负数区间上，就只能得出零函数的结果。

所以最终的结论是，该函数方程在 R 上的连续解只有零函数，在限定定义域为正实数集上时，对数函数也是它的解。

3.4.3　三角方程

三角方程是根据三角函数所满足的关系式所提出来的，根据正余函数和余弦函数的和角公式，有

$$\begin{aligned}\cos(x+y) &= \cos x\cos y-\sin x\sin y\\ \sin(x+y) &= \sin x\cos y+\cos x\sin y\end{aligned}$$

所以很自然的提出如下函数方程：两个函数 $C(x)$ 和 $S(x)$，对于任何两个实数 x 和 y 都成立

$$\begin{aligned}C(x+y) &= C(x)C(y)-S(x)S(y)\\ S(x+y) &= S(x)C(y)+C(x)S(y)\end{aligned}$$

显然正弦函数和余弦函数是其一解，所以问题是，是否是唯一解。

3.4.4　题选

题 3.4.1　(贵州周奎) 设连续的单调函数 f:$(-1,+\infty)\to(-1,+\infty)$，满足 $f(0)=0$，且 $f(x+f(y)+xf(y))\geqslant y+f(x)+yf(x)$，$\forall x,y\in(-1,+\infty)$，求函数 $f(x)$ 的解析式. ∎

解答. 这题不好搞，在题目不等式中取 $x=0$ 得到下式恒成立

$$f(f(y))\geqslant y$$

如果函数 $f(x)$ 是增函数的话，就必有 $f(x)\geqslant x$，因为如果有某个 x_0 使得 $f(x_0)<x_0$，那由单调增就有 $f(f(x_0))<f(x_0)<x_0$ 与上式矛盾，所以在增函数的条件下一定有 $f(x)\geqslant x$. 但对于减函数的情形似乎难以得到类似的结论。

题目之所以难是因为条件中是不等式，那先降低难度，考虑一下如果是等式会怎样，也就是

$$f(x+f(y)+xf(y))=y+f(x)+yf(x)$$

因为 $f(x)$ 单调，所以必有反函数，再假设 $f(x)$ 为满射（似乎题意也要求如此），于是在上式中将 y 替换为 $f^{-1}(y)$，就成为

$$f(x+y+xy)=f(x)+f^{-1}(y)+f(x)f^{-1}(y)$$

互换 x 和 y，又得

$$f(x+y+xy)=f^{-1}(x)+f(y)+f^{-1}(x)f(y)$$

于是

$$f(x)+f^{-1}(y)+f(x)f^{-1}(y)=f^{-1}(x)+f(y)+f^{-1}(x)f(y)$$

整理即得

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

因为 x 和 y 是任意的，所以上式意味着

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

右端 C 是常数，但 $f(0)=f^{-1}(0)=0$，所以 $C=0$，于是得

$$f(x)=f^{-1}(x)$$

这表明函数 $f(x)$ 的图象是关于直线 $y=x$ 对称的. 将它代回前面的等式就得到

$$f(x+y+xy)=f(x)+f(y)+f(x)f(y)$$

这个方程很漂亮，可是，似乎也难以转化为柯西方程，到这又没戏了。。。。。

既然还是难，那就进一步加条件，假设函数 $f(x)$ 不但连续，还是一阶可导的，那么有

$$\frac{f(x+y+xy)-f(x)}{y+xy}=\frac{1+f(x)}{(1+x)}\cdot\frac{f(y)}{y}$$

两边令 $y\to 0$，就有

$$f'(x)=\frac{1+f(x)}{(1+x)}\cdot f'(0)$$

这是一个微分方程，至于其中的常数 $f'(0)$，因为 $f(x)=f^{-1}(x)$，所以必然有 $f'(0)=1$ 或者 $f'(0)=-1$，方程变形为

$$\frac{y'}{1+y}=C\cdot\frac{1}{1+x}$$

其中 $C=1$ 或者 $C=-1$，两边积分得

$$\ln(1+y)=C\ln(1+x)+C'$$

这里 C' 是积分常数，所以得到

$$y=e^{C'}\cdot(1+x)^C-1$$

由 $f(0)=0$ 得 $C'=0$，所以

$$f(x)=(1+x)^C-1$$

其中 $C=1$ 或者 $C=-1$，分别对应着 $f(x)=x$ 与 $f(x)=-\dfrac{x}{1+x}$，但是要注意，这是在加强了条件（不等式改等式并且要求函数一阶可导）得出的结果. □

3.5 三角函数

3.5.1 第一类切比雪夫多项式

我们知道余弦的二倍角公式和三倍角公式

$$\begin{aligned}\cos 2\theta &= 2\cos^2\theta - 1\\ \cos 3\theta &= 4\cos^3\theta - 3\cos\theta\end{aligned}$$

于是 $\cos 3\theta$ 和 $\cos 4\theta$ 分别表成了 $\cos\theta$ 的二次多项式和三次多项式，那么对于一般形式的 $\cos n\theta$，情况又如何呢?

由和差化积公式可得

$$\cos(n+1)\theta + \cos(n-1)\theta = 2\cos n\theta\cos\theta$$

所以

$$\cos(n+1)\theta = 2\cos n\theta\cos\theta - \cos(n-1)\theta$$

假如 $\cos n\theta$ 和 $\cos(n-1)\theta$ 分别是 $\cos\theta$ 的 n 次多项式和 $n-1$ 次多项式，那么 $\cos(n+1)\theta$ 显然也是 $\cos\theta$ 的一个次数不超过 $n+1$ 的多项式，而易见最高次系数 a_n 满足递归关系 $a_{n+1} = 2a_n$，而 $a_2 = 2, a_3 = 4$，所以 $a_n = 2^{n-1}$，最高次项系数恒正，从而 $\cos(n+1)\theta$ 表成了 $\cos\theta$ 的 $n+1$ 次多项式。这结果就是如下的定理:

定理 3.5.1. *对于任意正整数 n，$\cos n\theta$ 可表为 $\cos\theta$ 的 n 次多项式，这多项式的最高次项系数是 2^{n-1}，这多项式称为第一类切比雪夫 (Chebyshev) 多项式。*

第一类切比雪零头多项式以 $T_n(x)$ 标记，按定义有 $\cos n\theta = T_n(\cos\theta)$。

容易得 $T_0(x) = 1, T_1(x) = x$ 并有如下递推式

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \tag{3.5.1}$$

当然也可以用这个递归式来定义第一类切比雪夫多项式，只是少了三角函数的背景，成了为了定义而定义了。

前几个多项式是如下的表:

$$\begin{aligned}T_0(x) &= 1\\ T_1(x) &= x\\ T_2(x) &= 2x^2 - 1\\ T_3(x) &= 4x^3 - 3x\\ T_4(x) &= 8x^4 - 4x^2 + 1\\ T_5(x) &= 16x^5 - 20x^3 + 5x\\ T_6(x) &= 32x^6 - 48x^4 + 18x^2 - 1\end{aligned}$$

在讨论这些多项式的性质之前，我们再从另外一个角度来引导出第一类切比雪夫多项式。由复数的乘幂公式，复数 $z=r(\cos\theta+i\sin\theta)$ 的乘方是

$$z^n=[r(\cos\theta+i\sin\theta)]^n=r^n(\cos n\theta+i\sin n\theta)$$

在上式中将 θ 换成 $-\theta$ 得

$$[r(\cos\theta-i\sin\theta)]^n=r^n(\cos n\theta-i\sin n\theta)$$

两式相加，并按二项式定理将左边展开，所有求和指标为奇数的项都被消去，得

$$\begin{aligned}\cos n\theta &= \sum_{0\leqslant i\leqslant n/2} C_n^{2i}(-1)^i\cos^{n-2i}\theta\sin^{2i}\theta\\ &= \sum_{0\leqslant i\leqslant n/2} C_n^{2i}(-1)^i\cos^{n-2i}\theta(1-\cos^2\theta)^i\end{aligned}$$

上式中右边每一项都是 $\cos\theta$ 的 n 次多项式,所有最高次项的系数和是 $C_n^0+C_n^2+C_n^4+\cdots$，这是对上标为偶数的二次项系数求和，由组合数性质，这个和是 2^{n-1}，于是 $\cos n\theta$ 就是 $\cos\theta$ 的 n 次多项式，最高次项系数是 2^{n-1}，这样也引出了第一类的切比雪夫多项式，还得出了它的具体表达式:

$$T_n(x)=\sum_{0\leqslant i\leqslant n/2} C_n^{2i}(-1)^i x^{n-2i}(1-x^2)^i$$

以下我们来推证这些多项式的一些性质，我们先由公式 $\cos n\theta=T_n(\cos\theta)$ 得出结论，再尝试利用递推关系来加以证明。

性质 3.5.1. 对于任意满足 $|x|\leqslant 1$ 的实数 x，都有 $|T_n(x)|\leqslant 1$.

3.5.2 第二类切比雪夫多项式

第一类切比雪夫多项式刻画的是 $\cos n\theta$ 展开为 $\cos\theta$ 的多项式，那么正弦的情况又如何呢，还是先观察一下二倍角公式和三倍角公式

$$\begin{aligned}\sin 2\theta &= 2\sin\theta\cos\theta\\ \sin 3\theta &= 3\sin\theta-4\sin^3\theta\end{aligned}$$

不难发现，$\sin n\theta$ 无法表示为 $\sin\theta$ 的多项式，但是 $\sin n\theta$ 似乎可以表达为 $\sin\theta$ 与一个关于 $\cos\theta$ 的多项式的乘积，也就是说，$\sin n\theta/\sin\theta$ 可表为 $\cos\theta$ 的多项式。那么对于一般情况，仍然由和差化积

$$\sin(n+1)\theta=2\sin n\theta\cos\theta-\sin(n-1)\theta$$

可见，$\sin(n+1)\theta$ 也能表为 $\sin\theta$ 与一个关于 $\cos\theta$ 的 n 次多项式之积，于是得到

定理 3.5.2. 对于任意正整数 n，$\sin(n+1)\theta/\sin\theta$ 可表为一个关于 $\cos\theta$ 的 n 次多项式，其最高次项系数是 2^n，这多项式称为第二类切比雪夫多项式。

仍然有递推公式 $U_0(x)=1, U_1(x)=2x$ 和

$$U_{n+1}(x)=2xU_n(x)-U_{n-1}(x)$$

可以发现，两类切比雪夫多项的递推公式是相同的，不同的是两个初始值。

第二类切比雪夫多项式以 $U_n(x)$ 标记，有

$$\frac{\sin{(n+1)\theta}}{\sin\theta}=U_n(\cos\theta)$$

同样利用复数的乘方公式，可以得到第二类切比雪夫多项式的具体表达:

$$U_n(x)=\sum_{0\leqslant 2i+1\leqslant n+1}(-1)^i C_{n+1}^{2i+1}x^{n-2i}(1-x^2)^i$$

其最高次项系数是 $C_{n+1}^1+C_{n+1}^3+C_{n+1}^5+\cdots=2^n$。

前几个多项式是

$$\begin{aligned}
U_0(x) &= 1\\
U_1(x) &= 2x\\
U_2(x) &= 4x^2-1\\
U_3(x) &= 8x^3-4x\\
U_4(x) &= 16x^4-12x^2+1\\
U_5(x) &= 32x^5-32x^3+6x\\
U_6(x) &= 64x^6-80x^4+24x^2-1
\end{aligned}$$

3.5.3 两类切比雪夫多项式的关系与它们的根

由三角关系，不难推得两类切比雪夫多项式的一些关系式

性质 3.5.2. *两类切比雪夫多项式间满足如下一些关系式:*

$$\begin{aligned}
T_n^2(x)+(1-x^2)U_{n-1}^2(x) &= 1\\
T_n(x)+\sqrt{x^2-1}U_{n-1}(x) &= (x+\sqrt{x^2-1})^n
\end{aligned}$$

两类多项式还有以下的双重递归:

$$\begin{aligned}
T_{n+1}(x) &= xT_n(x)-(1-x^2)U_{n-1}(x)\\
U_n(x) &= xU_{n-1}(x)+T_n(x)
\end{aligned}$$

初始值 $T_0(x)=1, U_{-1}(x)=1$，这由余弦和正弦的和角公式便可得出。

在高等数学中，切比雪夫多项是一类正交多项，可以利用它们来逼近函数和曲线，如同泰勒多项式中 $1, x, x^2\cdots$ 的地位一样，不过这不在本书讨论范围。

下面来讨论切比雪夫多项式的根，对于第一类，因为 $T_n(\cos\theta)=\cos n\theta$，所以 $T_n(x)$ 的全部根是

$$x_i=\cos\frac{2i-1}{2n}\pi,\ i=1,2,\ldots,n$$

同样，$U_n(x)$ 的全部根是

$$x_i = \cos\frac{i}{n+1}\pi,\ i = 1, 2, \ldots, n$$

显然这些根全部都在区间 $[-1, 1]$ 上。

3.6 函数的凸性

本文讨论上凸函数与下凸函数以及与之有关的不等式。

绝大多数函数的图像都不是一条笔直的直线，它们在某个区间上经常向上凸起或者向下凹陷。现在，我们要用数学语言来给函数的凸性下个定义：

定义 3.6.1. 如果定义在区间 D 上的 连续 函数 $f(x)$ 满足：对区间 D 上任意两个不相等的实数 a、b 都成立

$$f(\frac{a+b}{2}) \geqslant \frac{f(a)+f(b)}{2} \tag{3.6.1}$$

则称函数 $f(x)$ 在区间 D 上是上凸的。把这不等式中的不等号反向，则得到下凸函数的定义。如果这不等式中的等号不成立，则称为严格上凸或者严格下凸。

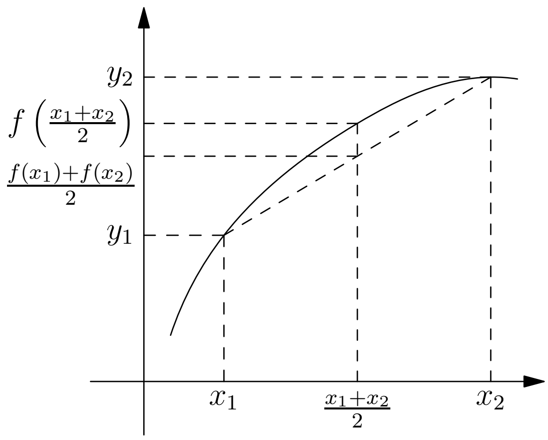

图 3.1

这定义的几何意义如图3.1，本文主要讨论上凸函数，因为下凸函数只要乘以 -1，即可变为上凸函数，但本文关于上凸函数的结论都可以类似的得到下凸函数的结论。

需要特别强调的是上述定义中对函数连续性的要求，这并不是说不连续的函数就没有凸性，而是因为这是一个不完善的定义，只有连续函数的上凸才能用上述不等式刻画，对于非连续函数而言，满足上述不等式还不足以得出它的凸性，这从下面这个函数就可以看出:

$$f(x) = \begin{cases} 0 & x \in Q \\ \sqrt{x} & x \notin Q \end{cases}$$

可以验证，这个函数满足上述定义中的不等式，然而它显然不是上凸的。关于凸函数完善的定义将在本节后文给出。

下面是一些例子。

例 3.6.1 讨论下面这些函数的凸性。

1. $f(x)=x^n$(n 为正整数, $x>0$)

2. $f(x)=\ln x$

3. $f(x)=\sin x, 0\leqslant x\leqslant \pi$

■

解. 因为对于整数 $k(0<k<n)$ 有 $(a^n+b^n)-(a^{n-k}b^k+a^kb^{n-k})=(a^{n-k}-b^{n-k})(a^k-b^k)>0$，因此

$$\begin{aligned}\left(\frac{a+b}{2}\right)^n+\left(\frac{b+a}{2}\right)^n&=\frac{1}{2^n}\sum_{k=0}^{n}C_n^k a^{n-k}b^k+\frac{1}{2^n}\sum_{k=0}^{n}C_n^k b^{n-k}a^k\\&=\frac{1}{2^n}\sum_{k=0}^{n}C_n^k\left(a^{n-k}b^k+a^kb^{n-k}\right)\\&<\frac{1}{2^n}\sum_{k=0}^{n}C_n^k\left(a^n+b^n\right)\\&=a^n+b^n\end{aligned}\tag{3.6.2}$$

所以 $\frac{a^n+b^n}{2}>\left(\frac{a+b}{2}\right)^n$，函数 $f(x)=x^n$ 在正实数区间上是下凸函数。

因为 $\frac{a+b}{2}-\sqrt{ab}=\frac{1}{2}(\sqrt{a}-\sqrt{b})^2>0$，所以 $\ln\frac{a+b}{2}-\frac{\ln a+\ln b}{2}=\ln\frac{a+b}{2}-\ln\sqrt{ab}>0$，所以对数函数在正实数区间上是上凸函数。

因为 $\frac{1}{2}(\sin a+\sin b)=\sin\frac{a+b}{2}\cos\frac{a-b}{2}<\sin\frac{a+b}{2}$，所以正弦函数在区间 $(0,\pi)$ 上是上凸函数。 □

利用数学归纳法，可以很容易的将定义中的两个数推广到任意个数的情形，这就是著名的琴生 *(Jensen)* 不等式:

定理 3.6.1 (琴生 (Jensen) 不等式)**.** *如果函数 $f(x)$ 是区间 D 上的上凸函数，则对于此区间上的任意 n 个实数 $x_i(i=1,2,\ldots,n)$，成立着不等式:*

$$f(\frac{1}{n}\sum_{i=1}^{n}x_i)\geqslant\frac{1}{n}\sum_{i=1}^{n}f(x_i)\tag{3.6.3}$$

证明. 事实上，与第一数学归纳法相比，倒推归纳法更容易证明本定理，读者不妨一试，本文仍使用第一数学归纳法。

由定义知 $n=2$ 的情形成立。现在假定对 n 的情形成立，则对于 $n+1$ 个数的情

形，记 $X_{n+1}=\frac{1}{n+1}\sum_{i=1}^{n+1}x_i$，则

$$\begin{aligned}
f(X_{n+1}) &= f\left(\frac{x_1+x_2+\cdots+x_{n+1}}{n+1}\right)\\
&= f\left(\frac{(x_1+\cdots+x_n)+(x_{n+1}+(n-1)X_{n+1})}{2n}\right)\\
&= f\left(\frac{\frac{x_1+\cdots+x_n}{n}+\frac{x_{n+1}+(n-1)X_{n+1}}{n}}{2}\right)\\
&\geqslant \frac{f\left(\frac{x_1+\cdots+x_n}{n}\right)+f\left(\frac{x_{n+1}+(n-1)X_{n+1}}{n}\right)}{2}\\
&\geqslant \frac{\frac{f(x_1)+\cdots+f(x_n)}{n}+\frac{f(x_{n+1})+(n-1)f(X_{n+1})}{n}}{2}
\end{aligned} \tag{3.6.4}$$

化简即得定理结论。 □

琴生不等式是一个非常重要的不等式，它是一系列不等式的重要来源，例如把它应用到对数函数身上，就得出多元均值不等式：

$$\frac{1}{n}\sum_{i=1}^{n}x_i \geqslant \sqrt[n]{\prod_{i=1}^{n}x_i}$$

另一个例子，因为正弦函数在区间 $(0,\pi)$ 上是上凸函数，所以对于一个三角形的三个内角，由琴生不等式就有：

$$\sin A+\sin B+\sin C \leqslant 3\sin\frac{A+B+C}{3}=\frac{3\sqrt{3}}{2}$$

最后我们来证明：圆的内接凸多边形中，以正多边形的面积为最大。只要记每一条边所对应的圆心角为 $\theta_i(i=1,2,\ldots,n)$，凸多边形的面积

$$\begin{aligned}
S &= \frac{1}{2}R^2\sum_{i=1}^{n}\sin\theta_i\\
&\leqslant \frac{n}{2}R^2\sin\left(\frac{1}{n}\sum_{i=1}^{n}\theta_i\right)\\
&= \frac{n}{2}R^2\sin\frac{2\pi}{n}
\end{aligned}$$

而等号在 $\theta_i=\frac{2\pi}{n}(i=1,2,\ldots,n)$ 时成立，即为正多边形。

从函数图象的几何意义上看，前面的定义仅仅刻画了区间中点处的函数值与区间两个端点函数值的关系，但从直觉上看，作为上凸函数，它在区间上任意一点的函数值，都大于连接函数曲线段 (由区间所截得) 两端点连线所代表的一次函数的函数值，换句话说，定义中的 $\frac{1}{2}$ 应该可以推广到 0 与 1 之间的任意实数。但是下面的定理表明，仅仅由前面的定义，只能将这个 $\frac{1}{2}$ 推广到 0 与 1 之间的有理数，而无理数的情形要依赖于函数的连续性才能得出。也正是由于这个原因，通常把下面定理中这个更强的不等式来作为上凸函数的定义。

定理 3.6.2. 如果函数 $f(x)$ 是定义在区间 D 上的上凸函数，并且在此区间上连续，则对于区间上任意两个实数 a、b，以及任意两个满足 $\alpha+\beta=1(\alpha\geqslant 0,\beta\geqslant 0)$ 的实数 α、β，都成立不等式[1]

$$f(\alpha a+\beta b)\geqslant \alpha f(a)+\beta f(b) \tag{3.6.5}$$

证明. [2] 先证明 α 是有理数的情形 (β 也就同时为有理数)，对于满足 $\alpha+\beta=1$ 的非负有理数 α 和 β，必定存在不同时为零的非负整数 m 和 n，使得

$$\alpha=\frac{m}{m+n},\beta=\frac{n}{m+n}$$

因此，在下面中将 $ma+nb$ 视为 m 个 a 与 n 个 b 分别相加之后求和，再利用琴生不等式，得到

$$\begin{aligned}f(\alpha a+\beta b)&=f\left(\frac{ma+nb}{m+n}\right)\\&\geqslant\frac{mf(a)+nf(b)}{m+n}\\&=\alpha f(a)+\beta f(b)\end{aligned}$$

因此有理数的情形得证。

对于 α 为无理数的情形，存在有理数的序列 α_n 和 β_n，使得 $\lim_{n\to\infty}\alpha_n=\alpha$ 和 $\lim_{n\to\infty}\beta_n=\beta$，由

$$f(\alpha_n a+\beta_n b)\geqslant\alpha_n f(a)+\beta_n f(b)$$

不等式两端令 $n\to\infty$，并利用函数 f 的连续性即得无理数时不等式仍成立。 □

现在来看一些例子，前面已经证明过对数函数在其定义域上是上凸函数，因此套用刚证明过的定理，就得到，对于任意正实数 a,b 以及满足 $\alpha+\beta=1(\alpha>0,\beta>0)$ 的正实数 α,β，都成立不等式

$$\alpha a+\beta b\geqslant a^{\alpha}b^{\beta} \tag{3.6.6}$$

这称为加权均值不等式（二元），而通常情况下的均值不等式 $\frac{a+b}{2}\geqslant\sqrt{ab}$，则不过是其中 $\alpha=\beta=\frac{1}{2}$ 的特殊情况而已。

这里再为这个重要定理提供另外一个证明，它不使用琴生不等式，而是直接从定义出发，不过，篇幅稍长些。

先提出如下两个引理:

引理 3.6.1. 区间上的凸函数，加上或者减去同一区间上的一次函数，不改变凸性。

这个结论是明显的，因为对于一次函数，式3.6.1中的等号总是成立的，因此在一个凸函数的式3.6.1两端同时减去一次函数的该式两端，不等式仍然成立。

引理 3.6.2. 如果函数 $f(x)$ 是区间 I 上的上凸函数，a 和 b 是区间 I 上两个不相等的固定实数，则函数 $h(t)=f((1-t)a+tb)$ 是关于 t 的定义在区间 $[0,1]$ 上的上凸函数。

---

[1] 也有的书上是用这个定理来作为凸函数定义的。

[2] 这个证明来自于参考文献 [8].

证明. 记 $g(t)=(1-t)a+tb$，则 $h(t)=f(g(t))$, 而 $g(t)$ 作为一次函数，显然对于任意 t_1 和 t_2 成立着不等式:

$$g\left(\frac{t_1+t_2}{2}\right)=\frac{g(t_1)+g(t_2)}{2}$$

因此有

$$\begin{aligned}h\left(\frac{t_1+t_2}{2}\right)&=f\left(g\left(\frac{t_1+t_2}{2}\right)\right)\\&=f\left(\frac{g(t_1)+g(t_2)}{2}\right)\\&\geqslant\frac{f(g(t_1))+f(g(t_2))}{2}\\&=\frac{h(t_1)+h(t_2)}{2}\end{aligned}$$

即得证。 □

从这证明过程可以看出，实际上我们可以得到更为通用的关于复合函数凸性的结论：

定理 3.6.3. *若函数 $g(x)$ 是区间 I 上的上凸函数，并且在该区间上的值域为 D，而函数 $f(t)$ 是区间 D 上的单调增加的上凸函数，则复合函数 $h(x)=f(g(x))$ 是区间 I 上的上凸函数，其它条件组合也有类似的结论。*

它的证明仿照引理 3.6.2即可，这里从略。

现在来证明定理 3.6.2

证明. 将不等式左边减去右边的差记为 $g(\alpha)$，则只要证明 $g(\alpha)\geqslant 0$ 对于一切 $0\leqslant\alpha\leqslant 1$ 恒成立即可。

由于完整的证明需要用到高等数学中连续性的定义，所以这里仅就 α 是有理数的情形进行证明 (当然 β 也就同时为有理数，而且函数连续的条件也就用不上)。

现在我们证明：对于任意正整数 n，不等式 $g(\frac{i}{n})\geqslant 0(i=0,1,\ldots,n)$ 都成立。

首先 $g(0)=g(1)=0$，并且利用上面两个引理可以得知 $g(\alpha)$ 仍是一个关于 α 的上凸函数，因此

$$g(\frac{i}{n})\geqslant\frac{1}{2}\left(g(\frac{i-1}{n})+g(\frac{i+1}{n})\right)$$

也即是

$$g(\frac{i+1}{n})-g(\frac{i}{n})\leqslant g(\frac{i}{n})-g(\frac{i-1}{n})$$

令 $\lambda_i=g(\frac{i+1}{n})-g(\frac{i}{n})$，则有 $\lambda_i\leqslant\lambda_{i-1}(i=1,2,\ldots,n-1)$，而

$$0=g(1)-g(0)=\sum_{i=0}^{n-1}\left(g(\frac{i+1}{n})-g(\frac{i}{n})\right)$$

即是 $\sum_{i=0}^{n-1}\lambda_i=0$，因此存在正整数 $k(0<k<n)$，使得

$$\lambda_0\geqslant\cdots\geqslant\lambda_{k-1}\geqslant 0\geqslant\lambda_k\geqslant\ldots\geqslant\lambda_{n-1}$$

也即是

$$0 = g(0) \leqslant g(\frac{1}{n}) \ldots \leqslant g(\frac{k}{n}) \geqslant \ldots \geqslant g(\frac{n-1}{n}) \geqslant g(1) = 0$$

因此，$g(\frac{i}{n}) \geqslant 0 (i = 0, 1, \ldots, n)$ 成立，有理数的情形得证。而无理数的情形仍同前面的证明。 □

同样的，我们还可以把定理3.6.2从两个数的情况推广到多个数的情形:

定理 3.6.4. *设函数 $f(x)$ 在区间 D 上是上凸函数，那么对于该区间上任意 n 个实数 $x_i (i = 1, 2, \ldots, n)$ 以及相应的权重 $\alpha_i (\sum_{i=1}^{n} \alpha_i = 1, \alpha_i \geqslant 0)$，有下面的不等式成立:*

$$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \geqslant \sum_{i=1}^{n} \alpha_i f(x_i) \tag{3.6.7}$$

其证明与定理3.6.1类似，此处从略。

我们还是把这定理套用到对数函数身上，得到如下的加权均值不等式：对于一组实数 $x_i (i = 1, 2, \ldots, n)$ 及一组权值 $\alpha_i > 0 (i = 1, 2, \ldots, n), \sum_{i=1}^{n} \alpha_i = 1$，成立着不等式:

$$\sum_{i=1}^{n} \alpha_i x_i \geqslant \prod_{i=1}^{n} x_i^{\alpha_i} \tag{3.6.8}$$

等号成立的条件是 x_i 全部都相等。它也可被视为均值不等式的推广。

3.7 数列

3.7.1 数列概念

数列就是一个数的序列，可以是有限序列也可以是无限序列，按下标记法可以写成 $a_1, a_2, \ldots$，有时为了方便也可以让下标从零开始 $a_0, a_1, a_2, \ldots$，甚至作为无穷序列时，还可以把下标扩展到负整数:$\cdots, a_{-2}, a_{-1}, a_0, a_1, a_2, \cdots$，成为一个双向无穷数列，如果有一个公式可以把数列的每一项表成它的下标的函数，就称此公式为该数列的*通项公式*。数列本质上是定义在整数集或者正整数集上的函数。

例 3.7.1 任一实数在十进制下可以表为

$$a_n a_{n-1} \cdots a_1 a_0 . a_{-1} a_{-2} \cdots$$

其中 a_i 取值范围是 $0, 1, \ldots, 9$，这便可以视为一个数列，而且是单侧无穷数列，而这个实数的值是

$$a_n \cdot 10^n + a_{n-1} \cdot 10^{n-1} + \cdots + a_1 \cdot 10 + a_0 \cdot 10^0 + a_{-1} \cdot 10^{-1} + a_{-2} \cdot 10^{-2} + \cdots$$

■

除了通项公式之外，*递推公式*也是研究数列的一个重要手段，所谓递推公式是指一个联系着数列中相邻若干项的一个公式，例如可以这样定义一个数列，约定 $a_1 = 1$，$a_2 = 1$，以后的每一项都由它前面的两项之和来定义，即 $a_{n+2} = a_{n+1} + a_n (n \geqslant 1)$，显然这公式唯一的确定了数列的每一项，也就确定了数列本身，它就是有名的 *斐波那契(Fibonacci) 数列*，后文还会讨论它。

3.7.2 等差数列与等比数列

等差数列和等比数列是最常见的两种数列。

定义 3.7.1. 如果一个数列 $\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots$ 中，任一项减去它前一项所得差值都相同，即存在常数 d，对任意整数 n 都有 $a_n - a_{n-1} = d$，则称此数列为 等差数列，而常数 d 称为此等差数列的公差.

显然，正整数数列 $\ldots, -2, -1, 0, 1, 2, \ldots$ 是公差为 1 的等差数列，任何项都取相同值的常数数列则是公差为零的等差数列。

等差数列的递推公式也可以用 $a_{n+1} + a_{n-1} = 2a_n$ 来刻画，而且其中不出现公差，仅仅通过项的关系来就定义出等差数列，这也表明，数列的递推公式可以不唯一。

定理 3.7.1. 对于等差数列中的任意两项（无论顺序）有

$$a_n - a_m = (n - m)d$$

证明. 证明是容易的，只要证明 $n \geqslant m$ 的情况即可（否则可以反过来相减)，首先可以验证 n 与 m 相等时该等式成立，而后有

$$a_n - a_m = \sum_{k=m}^{n-1} (a_{k+1} - a_k) = (n - m)d$$

当然，对差值 $n - m$ 施行归纳法也是可行的。 □

假定我们知道 a_0 和公差 d，在上式中取 $m = 0$，我们就得到等差数列的通项公式

定理 3.7.2. 等差数列的每一项可以由 a_0 及公差 d 表为

$$a_n = a_0 + nd$$

这便是等差数列的通项公式.

当然通项并不一定非得使用 a_0 来表示，$a_n = a_1 + (n-1)d$，$a_n = a_{100} + (n-100)d$ 也都是通项。

易见这通项是关于 n 的一次函数，实际上容易证明，如果数列的通项可表成关于下标的一次函数，则它必然是等差数列。

现在考虑对等差数列进行某种求和，如果从 a_0 开始，向正下标或者负下标进行累加，n 是任意一个整数（无论正负），记 $S_n = \sum_{i=0}^{n} a_n$，则

$$S_n = \sum_{i=0}^{n} a_i = \sum_{i=0}^{n} (a_0 + nd) = (n+1)a_0 + d\sum_{i=0}^{n} i$$

于是这个求和就归结为对自然数序列 $0, 1, 2, \ldots, n$ 进行求和，高斯曾经口算出 $1 + 2 + 3 + \cdots + 100 = 5050$，他的计算方法是，将 1 与 100 相加得 101，2 与 99 相加得 101，如此这般，最后 50 与 51 相加得 101，于是 $1 + 2 + \cdots + 100 = 50 \times 101 = 5050$，把这个方法略加改造便得出如下的计算 $1 + 2 + \cdots + n$ 的方法，为了避免 n 的奇偶性带来

分组分不完从而还需要分奇偶讨论的问题，再拿一个式子 $n+(n-1)+\cdots+2+1$ 与 $1+2+\cdots+(n-1)+n$ 进行相加，则这两个式子对应位置上的两数之和都是 $n+1$，所以最终得公式

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

不必怀疑右端会有出现小数的可能，因为作为连续的两个自然数 n 和 $n+1$ 中必定一奇一偶，所以右端永不会为小数。

把这公式应用到上面的式子便得

定理 3.7.3. *在公差为 d 的等差数列 a_n 中，有*

$$S_n=\sum_{i=0}^{n}a_i=(n+1)a_0+\frac{1}{2}n(n+1)d$$

其中 n 可以是任意整数 (包括负整数)。

上面求 $1+2+\cdots+n$ 的方法称为*倒序相加法*，当然也可以直接把这方法应用到 $S_n=a_0+a_1+\cdots+a_n$ 上，同样可得出上面的结果.

例 3.7.2 一类与组合数有关的和式 在本例中，我们来考虑求和 $C(n,m)=\sum\limits_{i=0}^{n}i^mC_n^i$，其中 m 是一个确定的非负整数。

在 $m=0$ 的情形，由二项式定理即知这个和是 $C(n,0)=2^n$. 而在 $m=1$ 的情况下，和式即 $\sum\limits_{i=0}^{n}iC_n^i$，此时有两种方法可以求出这个和式，其一是利用倒序相加，我们有

$$\begin{aligned}\sum_{i=0}^{n}iC_n^i &= \frac{1}{2}\left(\sum_{i=0}^{n}iC_n^i+\sum_{i=0}^{n}iC_n^{n-i}\right)\\ &= \frac{1}{2}\left(\sum_{i=0}^{n}iC_n^i+\sum_{i=0}^{n}(n-i)C_n^i\right)\\ &= \frac{n}{2}\sum_{i=0}^{n}C_n^i=n2^{n-1}\end{aligned}$$

即 $C(n,1)=n2^{n-1}$.

另一个方法是利用了组合恒等式 $iC_n^i=nC_{n-1}^{i-1}(i>0)$，这很容易验证，利用它我们有

$$\begin{aligned}C(n,1) &= \sum_{i=0}^{n}iC_n^i\\ &= n\sum_{i=1}^{n}C_{n-1}^{i-1}\\ &= n2^{n-1}\end{aligned}$$

受此启发，对于 $m=2$ 我们也有类似方法，在 $n>1$ 时有

$$\begin{aligned}i^2C_n^i &= niC_{n-1}^{i-1}\\ &= n(1+(i-1))C_{n-1}^{i-1}\\ &= nC_{n-1}^{i-1}+n(i-1)C_{n-1}^{i-1}\\ &= nC_{n-1}^{i-1}+n(n-1)C_{n-2}^{i-2}\end{aligned}$$

因此，

$$\begin{aligned} C(n,2) &= \sum_{i=0}^{n} i^2 C_n^i \\ &= C_n^1 + n\sum_{i=2}^{n} C_{n-1}^{i-1} + n(n-1)\sum_{i=2}^{n} C_{n-2}^{i-2} \\ &= n + n(2^{n-1}-1) + n(n-1)2^{n-2} \\ &= n2^{n-1} + n(n-1)2^{n-2} \end{aligned}$$

在一般情形下，对于 $C(n,m)=\sum\limits_{i=0}^{n} i^m C_n^i$，可以利用 $C_{n+1}^i = C_n^i + C_n^{i-1}(1\leqslant i\leqslant n)$，有

$$\begin{aligned} i^m C_{n+1}^i &= i^m C_n^i + i^m C_n^{i-1} \\ &= i^m C_n^i + (1+(i-1))^m C_n^{i-1} \\ &= i^m C_n^i + \sum_{k=0}^{m} C_m^k (i-1)^k C_n^{i-1} \end{aligned}$$

显然对于上式最后一行的前一项求和 (从 1 到 n) 就是 $C(n,m)$，而对后一项从 1 到 n 求和则是

$$\sum_{i=1}^{n}\sum_{k=0}^{m} C_m^k (i-1)^k C_n^{i-1} = \sum_{k=0}^{m} C_m^k \sum_{i=1}^{n} (i-1)^k C_n^{i-1} = \sum_{k=0}^{m} C_m^k (C(n,k)-n^k)$$

所以最终

$$C(n+1,m) = (n+1)^m + C(n,m) + \sum_{k=0}^{m} C_m^k (C(n,k)-n^k)$$

■

定义 3.7.2. 如果一数列 $\ldots,a_{-2},a_{-1},a_0,a_1,a_2,\ldots$ 中，任意一项与前一项的比值都是同一非零常数，即存在非零常数 q，使得对于任意整数 n 都有 $a_n/a_{n-1}=q$ 成立，则称此数列是 等比数列，而比值常数 q 称为它的 公比.

对于等比数列，如果它的各项都是正的，取对数即可变身为一个等差数列。对于所有的等比数列，同样可以得出类似于等差数列的结论来，即是

定理 3.7.4. 对于公比为 q 的等比数列 $\cdots,a_{-2},a_{-1},a_0,a_1,a_2,\cdots$，对任意整数 n 和任意整数 m 都成立

$$\frac{a_n}{a_m} = q^{n-m}$$

并由此得出通项

定理 3.7.5. 公比为 q 的等比数列 $\ldots,a_{-2},a_{-1},a_0,a_1,a_2,\ldots$ 的通项可表为

$$a_n = a_0 q^n$$

关于等比数列的求和

$$S_n = a_0 + a_1 + \cdots + a_n = a_0(1 + q + \cdots + q^n)$$

这就归结为对 $1 + q + q^2 + \cdots + q^n$ 求值，

由数学归纳法可得出下面的乘法公式，它也是平方差公式和立方差公式的自然推广 (见例 1.2.2)

$$a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \cdots + ab^{n-2} + b^{n-1}) \tag{3.7.1}$$

式中 n 是任意正整数，

在式中把 n 换成 $n+1$ 并取 $a = 1, b = q$ 便可得出，对于任意正整数 n，成立

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

若 n 为负整数，则上式中以 $1/q$ 代 q，便知上式仍然成立，于是对于等比数列就有

定理 3.7.6. *在以 $q(q \neq 1)$ 为公比的等比数列 $\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots$ 中，对于任意整数 n 都有*

$$S_n = \sum_{i=0}^{n} a_i = \frac{a_0(1 - q^{n+1})}{1 - q}$$

而在 $q = 1$ 时，$S_n = (n+1)a_0$.

还可以以另外一种方式得出这公式，因为

$$S_n = a_0 + a_1 + \cdots + a_n$$

两端同乘以公比 q，便可得出

$$qS_n = q(a_0 + a_1 + \cdots + a_n) = a_1 + a_2 + \cdots + a_n + a_{n+1}$$

于是

$$(1-q)S_n = a_0 - a_{n+1} = a_0(1 - q^{n+1})$$

所以得出

$$S_n = \frac{a_0 - a_{n+1}}{1 - q} = \frac{a_0(1 - q^{n+1})}{1 - q}$$

当然这过程中要求 $q \neq 1$，而对于 $q = 1$ 的情况，$S_n = (n+1)a_0$.

这个方法称为 *错位相减法*，下面这个例子提示了这个方法的一些用途。

例 3.7.3 错位相减法的一个用途 考虑对数列 $a_n = nq^n (n = 1, 2, 3, \ldots)$ 进行求和，这里要求 $q \neq 1$，否则就变成早已解决过的问题了。同样有

$$S_n = a_1 + a_2 + \cdots + a_n = q + 2q^2 + 3q^3 + \cdots + (n-1)q^{n-1} + nq^n$$

两端同乘以 q 得

$$qS_n = q^2 + 2q^3 + 3q^4 + \cdots + (n-1)q^n + nq^{n+1}$$

把以上两式相减，得

$$(1-q)S_n = (q + q^2 + \cdots + q^n) - nq^{n+1}$$

显然上式右端括号中的部分是等比数列的求和，于是 S_n 便可以求出。

接下来再考虑数列 $a_n = n^m q^n (n = 1, 2, 3, \ldots)$，其中 m 是个固定的正整数, 记

$$S_m(n) = \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} n^m q^n = q + 2^m q^2 + \cdots + (n-1)^m q^{n-1} + n^m q^m$$

进行同样的过程有

$$qS_m(n) = q^2 + 2^m q^3 + \cdots + (n-1)^m q^n + n^m q^{n+1}$$

两式相减

$$(1-q)S_m(n) = q + (2^m - 1)q^2 + (3^m - 2^m)q^3 + \cdots + (n^m - (n-1)^m)q^m - n^m q^{n+1}$$

而利用下面的公式 (它可以由二项式定理得出)

$$(1+k)^n - k^n = 1 + nk + C_n^2 k^2 + \cdots + C_n^{n-1} k^{n-1}$$

上式可化为

$$\begin{aligned}
(1-q)S_m(n) &= \sum_{i=1}^{n} (i^m - (i-1)^m) q^i - n^m q^{n+1} \\
&= \sum_{i=0}^{n-1} ((1+i)^m - i^m) q^{i+1} - n^m q^{n+1} \\
&= \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} C_m^j i^j q^{i+1} - n^m q^{n+1} \\
&= \sum_{j=0}^{m-1} C_m^j \sum_{i=0}^{n-1} i^j q^{i+1} - n^m q^{n+1} \\
&= q \sum_{j=0}^{m-1} C_m^j S_j(n-1) - n^m q^{n+1}
\end{aligned}$$

这意味着 $S_m(n)$ 可以用 $S_0(n-1), S_1(n-1), \ldots, S_{m-1}(n-1)$ 来表示，即具有递推性，而 $S_0(n-1)$ 显然就是等比数列求和，于是 $S_m(n)$ 便可以求出来。

进一步，便可以求形如 $a_n = p(n)q^n$ 这样的数列的和，其中 $p(n)$ 是关于 n 的多项式，只是这过程将越来越繁琐，如此机械化的计算方法，由计算机程序来进行是再合适不过了。 ■

例 3.7.4 自然数的幂和 在上面得出了公式

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

现在就来讨论下一般的 $S_m(n) = \sum_{i=1}^{n} i^m$ 的求和公式，仍然由

$$(1+k)^m - k^m = 1 + mk + C_m^2 k^2 + \cdots + C_m^{m-1} k^{m-1}$$

对 $k=1,2,\ldots,n$ 进行累加，便得

$$(1+n)^m-1=S_0(n)+C_m^1S_1(n)+C_m^2S_2(n)+\cdots C_m^{m-1}S_{m-1}(n)$$

显然这便是联系着诸 $S_m(n)$ 的递推关系，把它写成更美观的形式

$$\sum_{i=0}^{m}C_m^iS_i(n)=(1+n)^{m+1}-1$$

且有初始公式

$$S_0(n)=n$$

由此出发便能求出任何 $S_m(n)$ 的表达式来，比如说取 $m=2$ 便得

$$S_0(n)+2S_1(n)+S_2(n)=(1+n)^3-1$$

于是得出

$$S_2(n)=1^2+2^2+\cdots+n^2=\frac{1}{6}n(n+1)(2n+1)$$

同样不必怀疑右端的整数性，因为 n 与 $n+1$ 作为相邻两个正整数，必然有一个偶数，故 $2\mid n(n+1)$，还需证明 $3\mid n(n+1)(2n+1)$，若是 n 与 $n+1$ 中有 3 的倍数，则自然不消说，若是 n 与 $n+1$ 都不是 3 的倍数，则它俩被 3 除所得余数必然一个是 1，另一个是 2，于是 $2n+1=n+(n+1)$ 便必然是 3 的倍数，于是 2 和 3 都能整除 $n(n+1)(2n+1)$ 且 2 与 3 互素，所以 6 也能整除它。

同样再取 $m=3$，便可得出

$$S_3(n)=1^3+2^3+\cdots+n^3=\frac{1}{4}n^2(n+1)^2$$

一般的可以知道，$S_m(n)$ 是关于 n 的 $m+1$ 次多项式。 ■

3.7.3 差分与高阶等差数列

定义 3.7.3. 对于无穷数列 $\ldots,a_{-2},a_{-1},a_0,a_1,a_2,\ldots$，定义它的 一阶差分数列 为

$$\Delta a_n=a_n-a_{n-1}$$

进一步递归的定义它的 m 阶差分数列是

$$\Delta^m a_n=\Delta^{m-1}a_n-\Delta^{m-1}a_{n-1}$$

即 m 阶差分是 $m-1$ 阶差分的差分，这称为 高阶差分，为了方便，可以把数列本身看作它自己的 零阶差分数列.

差分将一个数列变换为另一个数列，这与平方运算把一个数变成另一个数，求导运算把一个函数变成另一个函数没什么差别，所以差分是在数列这种数学对象上的一种运算。

为了讨论的方便，用 $a_n^{(m)}$ 代表数列 a_n 的 m 阶差分数列，并规定 $a_n^{(0)} = a_n$，上标加了括号，以与幂相区别，在这定义就有

$$a_n^{(m)} = a_n^{(m-1)} - a_{n-1}^{(m-1)}$$

接下来有高阶等差数列的定义

定义 3.7.4. 如果一个数列 a_n 的 m 阶差分数列是等差数列，则该数列称为 m 阶等差数列.

显然等差数列本身即是零阶等差数列，同时，如果一个数列是 m 阶等差数列，则它也必然是 $m+1$ 阶等差数列，$m+2$ 阶等差数列，等等。

易见

定理 3.7.7. 数列 a_n 是 m 阶等差数列的充分必要条件是 Δa_n 是 $m-1$ 阶等差数列.

假定 a_n 是一个 m 阶等差数列，那么存在常数 d，使得对任意整数 n 成立下式

$$a_n^{(m)} - a_{n-1}^{(m)} = d$$

把式中的 m 阶差分数列用 $m-1$ 阶差分的差分来替代，就得到

$$a_n^{(m-1)} - 2a_{n-1}^{(m-1)} + a_{n-2}^{(m-1)} = d$$

同样的手段再施行两次，得

$$\begin{aligned} a_n^{(m-2)} - 3a_{n-1}^{(m-2)} + 3a_{n-2}^{(m-2)} - a_{n-3}^{(m-2)} &= d \\ a_n^{(m-3)} - 4a_{n-1}^{(m-3)} + 6a_{n-2}^{(m-3)} - 4a_{n-3}^{(m-3)} + a_{n-4}^{(m-3)} &= d \end{aligned}$$

易见各项式的系数正好便是二项式系数并交错的带上正负号，并且 $m-k$ 阶差分数列的公式中便含有相邻 $k+2$ 项，照此规律，当差分的阶数降为零时，公式成为

$$C_{m+1}^0 a_n - C_{m+1}^1 a_{n-1} + \cdots + (-1)^{m+1} C_{m+1}^{m+1} a_{n-m-1} = d$$

这便是 m 阶等差数列的一个递推公式，但是公式中带有未知常数 d，现在想办法去掉它，因为对于 m 阶等差数列而言，它的 m 阶差分数列成为等差数列，那么它的 $m+1$ 阶差分数列就成为一个常数数列，也就是公差为零的等差数列，因此按照上面的公式，原来数列的递推公式可以表为

$$C_{m+2}^0 a_n - C_{m+2}^1 a_{n-1} + \cdots + (-1)^{m+2} C_{m+2}^{m+2} a_{n-m-2} = 0$$

于是得到以下重要结果

定理 3.7.8. $m-2$ 阶等差数列的递推公式是

$$C_m^0 a_n - C_m^1 a_{n-1} + \cdots + (-1)^m C_m^m a_{n-m} = 0$$

这只要把上面的推导，改用数学归纳法写出来就可以证明它。

这结果对于等差数列即零阶等差数列而言就成为 $a_n-2a_{n-1}+a_{n-2}=0$.

现在讨论一下高阶等差数列的通项问题，先考虑一阶等差数列 a_n，按定义，它的一阶差分数列是等差数列，于是有

$$a_n-a_{n-1}=rn+s$$

于是对于整数 n(包括负整数)，就有

$$\begin{aligned} a_n &= a_0+\sum_{i=1}^{n}(a_i-a_{i-1}) \\ &= a_0+\sum_{i=1}^{n}(ri+s) \\ &= a_0+ns+r\sum_{i=1}^{n}i \\ &= a_0+ns+\frac{1}{2}rn(n+1) \end{aligned}$$

这表明它的通项是关于 n 的二次多项式，再重复同样的过程并利用自然数的幂和公式 (例 3.7.4) 便可得出二阶等差数列的通项是关于 n 的三次多项式，如此等等，于是得出如下重要结果

定理 3.7.9. *无穷数列 $\ldots,a_{-2},a_{-1},a_0,a_1$ 是 m 阶等差数列的充分必要条件是，它的通项能表成关于下标的 $m+1$ 次多项式.*

这结果表明，高阶等差数列本质上就是多项式数列。

证明. 对 m 施行数学归纳法，$m=0$ 的情况是显然的，假设必要性对于小于 m 的正整数都成立，那么对于一个 m 阶等差数列 a_n，它的一阶差分数列是 $m-1$ 阶等差数列，因而通项为关于下标的 m 次多项式，于是

$$a_n-a_{n-1}=\sum_{j=0}^{m}b_jn^j$$

累加得

$$\begin{aligned} a_n &= a_0+\sum_{i=0}^{n}(a_i-a_{i-1}) \\ &= a_0+\sum_{i=0}^{n}\sum_{j=0}^{m}b_ji^j \\ &= a_0+\sum_{j=0}^{m}b_j\sum_{i=0}^{n}i^j \end{aligned}$$

在例 3.7.4中，我们就已经知道 $\sum_{i=1}^{n}i^j$ 是 $j+1$ 次多项式，于是 a_n 便是关于 n 的 $m+1$ 次多项式，所以必要性成立.

再证充分性，同样施行归纳法，如果数列的通项是关于下标的一次多项式，显然它必然是 (零阶) 等差数列，即 $m=0$ 时充分性成立，假若充分性对于小于 m 的正整数都成立，现在设数列 a_n 的通项是关于 n 的 $m+1$ 次多项式，即

$$a_n=\sum_{i=0}^{m+1} b_i n^i$$

则它的差分数列

$$a_{n+1}-a_n=\sum_{i=0}^{m+1} b_i((1+n)^i-n^i)$$

显然，n^{m+1} 被抵销，这成为一个次数低于 $m+1$ 的多项式，按归纳假设，这差分数列必然是某一阶的等差数列，且阶数小于 m，于是原来的数列 a_n 也是某一阶的等差数列，且阶数小于等于 m，自然也是 m 阶等差数列. □

3.7.4 线性递推数列

本节讨论常系数线性递推数列的通项求法问题[3]，这个都是有固定结论的内容，本文只是粗略转叙一番而已。

所谓常系数线性递推数列，是指它的递推公式形如 $\sum_{k=1}^{n}\lambda_k a_k=c$ 的数列，例如斐波那契数列 $a_{n+2}=a_{n+1}+a_n$。

先看最简单的一种，递推式为 $a_{n+1}=pa_n+q$ 的数列，当 $p=1$ 时它成为等差数列，当 $q=0$ 则成为等比数列，所以此处限定 $p\neq 1, q\neq 0$。

要求它的通项，只要在它两端同时除以 p^{n+1}，就有

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

因此有

$$\begin{aligned}\frac{a_n}{p^n} &= \frac{a_1}{p}+\sum_{k=2}^{n}\left(\frac{a_k}{p^k}-\frac{a_{k-1}}{p^{k-1}}\right)\\ &= \frac{a_1}{p}+\sum_{k=2}^{n}\frac{q}{p^k}\end{aligned}$$

剩下的就是对一个等比数列进行求和了。

另外一种方法比较巧妙，假定存在一个实数 λ，使得 $a_{n+1}+\lambda=p(a_n+\lambda)$，展开与原递推式比较即得 $\lambda=\frac{q}{p-1}$，于是数列 $a_n-\lambda$ 就成为一个等比数列了。

现在看二阶的情形，每一项需要它前面两项才能确定：$a_{n+2}=pa_{n+1}+qa_n$，假想有两个实数 r 和 s 能够使得

$$a_{n+2}-ra_{n+1}=s(a_{n+1}-ra_n) \tag{3.7.2}$$

展开后与原递推式比较可得

$$r+s=p$$

$$rs=-q$$

[3]线性递推数列的通项求法这部分内容，主要参考了文献 [2].

因此 r 和 s 是方程 $x^2=px+q$ 的两个根，在复数范围内，它必有两个解 (可以相等，重根按重数计算)，于是数列 $a_{n+1}-ra_n$ 成为等比数列，求出它的通项 $a_{n+1}-ra_n=f(n)$ 后，只要两端同时除以 r^{n+1} 即可求出 a_n 的通项，其结果如下

如果方程有两个等根 $x=r$，那么

$$a_n=\left(\frac{a_2-ra_1}{r^2}-\frac{a_2-2ra_1}{r^2}\right)r^n$$

如果方程有两个不相等的根 $x_1=r,x_2=s$，那么

$$a_n=\frac{a_2-sa_1}{r(r-s)}r^n-\frac{a_2-ra_1}{s(r-s)}s^n$$

如果这两个根不相等，则还有另一种求法，因为 r 和 s 都是方程 $x^2=px+q$ 的根，因此既然有3.7.2成立，也就必然有

$$a_{n+2}-sa_{n+1}=r(a_{n+1}-sa_n) \tag{3.7.3}$$

成立，递推下去，便有

$$a_n-ra_{n-1}=s^{n-2}(a_2-ra_1)$$
$$a_n-sa_{n-1}=r^{n-2}(a_2-sa_1)$$

从中解出 a_n 来:

$$a_n=\frac{(a_2-sa_1)r^{n-1}-(a_2-ra_1)s^{n-1}}{r-s}$$

易见这公式具有如下形式

$$a_n=c_1r^n+c_2s^n$$

其中 c_1 和 c_2 是常数，这常数跟 a_1、a_2 以及 r 和 s 有关，跟 r 和 s 有关实际上就是跟 p 与 q 有关，可以直接将 a_1 和 a_2 的值代入上式中定出 c_1 和 c_2，这只需求解一个二元一次方程组就可以了。

而对于有两个等根的情形，通项具有形式

$$a_n=(c_1+c_2n)r^n$$

的情形，同样可以利用待定系数法求出 c_1,c_2.

更一般的情形是

定理 3.7.10. 对于线性递推数列 $\sum_{k=1}^n\lambda_ka_{p+k}=c$，称方程 $\sum_{k=1}^n\lambda_kx^k=c$ 为它的特征方程，在复数范围内这个特征方程必有 n 个根（重根按重数计算），假定这些根是 $x_i(i=1,2,\cdots,m,m\le n)$，相应根的重数是 $r_i(i=1,2,\cdots,m,\sum_{i=1}^m r_i=n)$，则它的通项是:

$$a_n=\sum_{i=1}^m P_{r_i-1}(n)x_i^n$$

上式中 $P_{r_i-1}(n)$ 表示一个关于 n 的次数是 r_i-1 的多项式，如果哪个根是单重根，则它的系数是常数。

此定理便是说，线性递推数列，其实就是多项式作系数的指数的组合，以后在微分方程中还会看到，常系数线性微分方程的解也是指数函数的组合，这两个结果表明，差分与微分存在某种类似，差分可以视为微分的离散化。

例 3.7.5 作为一个例子，现在来求斐波那契数列的通项，递推公式为 $a_{n+2}=a_{n+1}+a_n$，特征方程是 $x^2=x+1$，其两个根是 $x_{1,2}=\frac{1}{2}(1\pm\sqrt{5})$，于是通项应为：

$$a_n=\alpha_1x_1^n+\alpha_2x_2^n$$

将 $a_1=1$ 和 $a_2=1$ 带入求出两个系数，最后得：

$$a_n=\frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n-\left(\frac{1-\sqrt{5}}{2}\right)^n\right]$$

令人惊讶的是一个所有项都是正整数的数列，其通项居然出现了无理数，事实上，利用二项定理可以证明，这个表达式将永远是正整数。 ■

例 3.7.6 讨论数列 $a_1=1,a_2=1$，而以后的项由 $a_{n+2}=a_{n+1}-a_n$ 决定，其特征方程是 $x^2=x-1$，方程在复数范围内有两个根

$$\alpha=\frac{1+\sqrt{3}i}{2},\ \beta=\frac{-1+\sqrt{3}i}{2}$$

通项则有形式 $a_n=c_1\alpha^n+c_2\beta^n$ 的形式，将 a_1 和 a_2 的值代入通项定出 c_1 和 c_2 后得

$$a_n=\frac{1}{\sqrt{3}i}\left(\frac{1+\sqrt{3}i}{2}\right)^n-\frac{1}{\sqrt{3}i}\left(\frac{-1+\sqrt{3}i}{2}\right)^n$$

这时的通项，就不得不借助复数来表达了，实际上这还是一个周期数列。 ■

例 3.7.7 我们已经知道，一个 $m-2$ 阶等差数列的递推公式是

$$C_m^0a_n-C_m^1a_{n-1}+\cdots+(-1)^mC_m^ma_{n-m}=0$$

这显然是一个线性递推数列，其特征方程是

$$\sum_{i=0}^{m}(-1)^iC_m^ix^i=0$$

也就是 $(1-x)^m=0$，这方程只有一个 m 重根 $x=1$，因而通项可表为 $a_n=p(n)\cdot 1^n$，其中 $p(n)$ 是一个 $m-1$ 次多项式，这样就再次证明了定理 3.7.9. ■

3.7.5 分式型递推数列

3.7.6 递推方法的应用

有些问题本身并不是数列问题，但因为问题与自然数有关，递推方法往往能发挥作用，这一节举几个例子。

例 3.7.8 伯努利信封问题 伯努利信封问题早在介绍容斥原理时就解决过 (见例 1.2.1)，此处介绍运用数列递推方法的解决方式，有两个方法，其一是大数学家欧拉所提供的，思维精巧，其二是我自己的解法。

简单复述一下此问题，有相同数目的一一配对的信件和信封，把这些信件装进这些信封中，问没有任何一封信件装进正确的信封的装法总数是多少。

欧拉对此问题有一个借助数列递推的解法，记信封数目为 n 时的结果为 a_n，那么 $a_1=0$，来看下这个数列的递推情况，在有 $n+1$ 个信封时，考虑编号为 1 的那封信件，除 1 号信封外它有 n 个信封可以装入，假定它装入的信封编号是 r，那再考虑编号为 r 的信件，它此时有两个选择，一是它可以装入 1 号信封，这时其它的 $n-1$ 个信件的装法是 a_{n-1}，它的另一个选择是不装入 1 号信封，这时由于 1 号信封等同于 r 号信封，所以除 1 号信件以外的 n 封信件有 a_n 种装法，于是得到该数列的递推公式为:

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

两边同除以 $(n+1)!$ 并记 $b_n=\frac{a_n}{n!}$ 可得

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

即

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

所以

$$b_n-b_{n-1}=\sum_{i=2}^{n}(b_i-b_{i-1})=\sum_{i=2}^{n}(-1)^i\frac{1}{i!}$$

于是

$$a_n=n!b_n=n!(b_1+\sum_{i=2}^{n}(b_i-b_{i-1}))=n!\sum_{i=2}^{n}(-1)^i\frac{1}{i!}$$

同样在 $0!=1$ 的约定下即有

$$a_n=n!\sum_{i=0}^{n}(-1)^i\frac{1}{i!}$$

以下介绍另一个解法，这个解法是我在高中时期搞出来的，那时我还没听说过伯努利信封问题，当时考虑的问题是：有编号为 $1,2,\ldots,n$ 的 n 个人和同样编号的 n 个座位，如果每个人都不允许坐与自身编号相同的座位，有多少种坐法？我当时称之为错位排列问题，原稿早已遗失（也太啰嗦，长达三页，当时花费了好多天，写出来后自己不会 word，花了十元钱请打印店打出来，成天自我欣赏，这大概代表了我高中时期在数学上的最高成就)，此处是根据思路重新写的。

考虑编号为 1 的人，假如他坐在编号为 r_1 的座位上，又看编号为 r_1 的人，他可以坐在 1 号座位上，也可以坐在 r_2 号座位上，如果是坐在 1 号座位上，那么 1 号和 r_1 号这两个人由于交叉坐，与其他人无涉，如果是坐在 r_2 号座位上，则再继续看 r_2 号人，依此顺着链条 $1\to r_1\to r_2\to\cdots$，这样下去的结果是，从 1 号人开始，能够找到一个环，使得该环内的人和座位与环外的人和座位无涉，可以独立开来，如果找不到这样的环，那无非是所有人一起构成了一个整体环，即是无法分割。这样的环可能不止一个，但我们仅考虑包含 1 号人的那个环。

假定 n 个人的错位排列数为 a_n，设包含 1 号人的环共含有 $m(2\leqslant m\leqslant n-2)$ 个人，那么只要从除 1 号之外的人中选出 $m-1$ 个人排在这个座位链上，因此座法是

$(m-1)!C_{n-1}^{m-1} = \dfrac{(n-1)!}{(n-m)!}$，而剩下的 $n-m$ 个人又组成了一个较小的错位排列，因此其坐法数是 a_{n-m}，此外，当 $m=n$ 时，所有人构成一个环，只要把他们按照座位链排成一排即可，所以此时的坐法是 $(n-1)!$，于是总共的坐法是：

$$a_n = \sum_{m=2}^{n-2} \frac{(n-1)!}{(n-m)!} a_{n-m} + (n-1)! = (n-1)!(\sum_{m=2}^{n-2} \frac{a_{n-m}}{(n-m)!} + 1)$$

将式中的求和顺序倒过来，就有

$$a_n = (n-1)!(\sum_{m=2}^{n-2} \frac{a_m}{m!} + 1) \tag{3.7.4}$$

此即其作为数列的递推公式，自然的，$a_1 = 0, a_2 = 1$。

下面来求它的通项公式，上式可以变形为：

$$\frac{a_n}{n!} = \frac{1}{n}(1 + \sum_{m=2}^{n-2} \frac{a_m}{m!}) \tag{3.7.5}$$

只要记 $b_n = \dfrac{a_n}{n!}$，则有 $b_1 = 0$ 和

$$b_n = \frac{1}{n}(1 + \sum_{m=2}^{n-2} b_m) \tag{3.7.6}$$

于是有

$$nb_n - (n-1)b_{n-1} = b_{n-2} \tag{3.7.7}$$

也就是

$$b_n - b_{n-1} = -\frac{1}{n}(b_{n-1} - b_{n-2}) \tag{3.7.8}$$

而 $b_2 - b_1 = \dfrac{1}{2}$，所以

$$b_n - b_{n-1} = (-1)^{n-2} \frac{1}{n!} = (-1)^n \frac{1}{n!} \tag{3.7.9}$$

于是

$$b_n = b_1 + \sum_{m=2}^{n} (b_m - b_{m-1}) = \sum_{m=2}^{n} (-1)^m \frac{1}{m!} \tag{3.7.10}$$

所以最终

$$a_n = n!b_n = n! \sum_{m=2}^{n} (-1)^m \frac{1}{m!} \tag{3.7.11}$$

在 $0! = 1$ 的约定下，也可以把求和指标从零开始

$$a_n = n! \sum_{m=0}^{n} (-1)^m \frac{1}{m!} \tag{3.7.12}$$

这就是最终的结果。

■

3.7.7 题选

题 3.7.1 已知各项都是正实数的数列 x_n 对一切正整数 n 都成立 $x_n+\frac{1}{x_{n+1}}<2$，求证该数列所有项都满足 $x_n<1$. ■

解答. 如果用上极限理论，则可以很容易的得出它单调增加并以 1 为极限，结论不证自明，所以这里主要讨论的是初等证明。

因为

$$x_n+\frac{1}{x_{n+1}}<2\leqslant x_{n+1}+\frac{1}{x_{n+1}}$$

所以 $x_n<x_{n+1}$，即该数列单调增加。

又显然 $x_n<2$，所以

$$2>x_n+\frac{1}{x_{n+1}}>x_n+\frac{1}{2}$$

于是 $x_n<2-\frac{1}{2}$，我们得到一个更加好的上限，重复这个过程，我们由 $x_n<y_m$ 就可以得到

$$x_n<2-\frac{1}{y_m}$$

所以我们作数列 y_m，它由 $y_1=2$ 和

$$y_{m+1}=2-\frac{1}{y_m}$$

来确定。

数列 y_m 的每一项都大过数列 x_n 的全部项，所以它的下标特意用 m 而不是 n 来表示，以示不相关。

现在来求 y_m 的通项公式，由于

$$\frac{1}{y_{m+1}-1}=1+\frac{1}{y_m-1}$$

因此数列 $\frac{1}{y_m-1}$ 是等差数列，它的通项为 $y_m=1+\frac{1}{m}$[4]，于是 $x_n<y_m$ 对一切正整数 n 和 m 都成立，所以必定有 $x_n\leqslant 1$（用反证法），而 x_n 的单调性则保证了等号是不能取的。

下面关于 y_m 再给个不求通项的玩法[5]，$y_m>1$ 这一点根据数学归纳法是明显成立的。下面证明它可以任意接近 1，也就是要证明，对于无论多么小的正实数 δ，总存在 y_m 中的某一项 y_M，使得 $y_M<1+\delta$。采用反证法，假定存在某个正实数 δ，使得 y_m 中的所有项都满足 $y_m\geqslant 1+\delta$，则

$$y_{m+1}-1=\frac{1}{y_m}(y_m-1)\leqslant\frac{1}{1+\delta}(y_m-1)$$

于是

$$y_m-1\leqslant\frac{1}{(1+\delta)^m}$$

显然与假设矛盾，故得证。 □

[4]这是分式型递推数列求通项的不动点解法。

[5]其实这是高数的玩法，就差提到确界二字了。

题 3.7.2 数列 a_n 满足：$a_1=\frac{1}{2}$，$a_{n+1}=\frac{na_n+a_n^2}{n+1}$，

1. 求证该数列是递减的。
2. 求证 $a_n<\frac{7}{4n}$

■

解答. 根据数列归纳法易知 $0<a_n<1$，所以

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

所以数列递减，第二问，只要证明 $n>3$ 时有如下更强的不等式即可（使用数学归纳法，过程略去）

$$a_n\leqslant\frac{7}{4n}-\frac{4}{n^2}$$

□

题 3.7.3 记 $I_n=1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots+\frac{1}{2n-1}-\frac{1}{2n}$，求证，在正整数 $n\geqslant 100$ 时，有 $0.68<I_n<0.7$. ■

证明. 记 $J_n=\sum_{k=1}^{n}(\frac{1}{2k}-\frac{1}{2k+1})$，由于 $z_n=\frac{1}{n-1}-\frac{1}{n}=\frac{1}{n(n-1)}$ 是递减的，并且相邻两项也相差越来越小，所以有不等式 $z_{2k}<\frac{1}{2}(z_{2k-1}+z_{2k+1})$, 也就是如下的:

$$\frac{1}{2k-1}-\frac{1}{2k}<\frac{1}{2}\left[\left(\frac{1}{2k-2}-\frac{1}{2k-1}\right)+\left(\frac{1}{2k}-\frac{1}{2k+1}\right)\right]$$

对上式左边进行累加，但从 $k=3$ 到 $k=n$ 使用右边放缩，得

$$I_n<\frac{1}{2}+\frac{1}{12}+\frac{1}{2}\left[\left(J_n-\frac{1}{6}-\frac{1}{2n(2n+1)}\right)+\left(J_n-\frac{1}{6}-\frac{1}{20}\right)\right]$$

化简

$$I_n<J_n+\frac{47}{120}-\frac{1}{4n(2n+1)}$$

利用 $I_n+J_n=1-\frac{1}{2n+1}$ 从上式中换掉 J_n 得

$$I_n<\frac{167}{240}-\frac{1}{2(2n+1)}-\frac{1}{8n(2n+1)}<\frac{167}{240}<\frac{168}{240}=0.7 \tag{3.7.13}$$

于是不等式的右边得证，接下来考虑左边不等式，同样因为 z_n 是递减的，有不等式 $z_{2k}>\frac{1}{2}(z_{2k}+z_{2k+1})$，也就是

$$\frac{1}{2k-1}-\frac{1}{2k}>\frac{1}{2}\left[\left(\frac{1}{2k-1}-\frac{1}{2k}\right)+\left(\frac{1}{2k}-\frac{1}{2k+1}\right)\right]$$

对左边进行累加，在 $k\geqslant 3$ 时使用右边放缩，得到

$$I_n>\frac{1}{2}+\frac{1}{12}+\frac{1}{2}\left(\frac{1}{5}-\frac{1}{2n+1}\right)$$

也就是

$$I_n > \frac{41}{60} - \frac{1}{2(2n+1)} \tag{3.7.14}$$

在 $n \geqslant 100$ 时，有

$$I_n \geqslant I_{100} > \frac{41}{60} - \frac{1}{402} = 0.680845771... > 0.68$$

所以不等式左边得证.

其实证明左边所用的放缩是比较松的，实际上因为 $\lim_{n\to\infty} I_n = \ln 2 = 0.693147...$，所以左边不等式的放缩余地较大，所以这样的放缩也能达到要求，现在来尝试使用更强的放缩，看看能得到一个什么样的结果。

对于 z_n，不等式 $z_n > 2z_{n+1} - z_{n+2}$ 将是一个更强的放缩，所以我们有 $z_{2k} > 2z_{2k+1} - z_{2k+2}$，也就是下面的不等式

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

对上式左边进行累加，但只在 $k \geqslant 3$ 时使用右边放缩，即得

$$I_n > \frac{1}{2} + \frac{1}{12} + 2\left(J_n - \frac{1}{6} - \frac{1}{20}\right) - \left(I_n - \frac{1}{2} - \frac{1}{12} - \frac{1}{30} + \frac{1}{(2n+1)(2n+2)}\right)$$

将其中的 J_n 用 $1 - I_n - \frac{1}{2n+1}$ 替换掉，即得

$$I_n > \frac{83}{120} - \frac{1}{2(2n+1)} - \frac{1}{4(2n+1)(2n+2)}$$

因此在 $n \geqslant 100$ 时，便有

$$I_n \geqslant I_{100} > \frac{83}{120} - \frac{1}{2(2\times 100+1)} - \frac{1}{4(2\times 100+1)(2\times 100+2)} = 0.6891729471.....$$

这个值已经非常接近 0.69 了。 □

题 3.7.4 已知数列 $\{a_n\}$ 满足: $a_1 = 1$，$a_{n+1} = a_n + \frac{1}{a_n^2}$，求证 $a_{2015} > 18$. ■

证明. 易见这是一个递增的正项数列，在递推式两边同时三次方:

$$a_{n+1}^3 = \left(a_n + \frac{1}{a_n^2}\right)^3 = a_n^3 + 3 + \frac{3}{a_n^3} + \frac{1}{a_n^6} > a_n^3 + 3$$

所以 $a_{2015} > a_1^3 + 3 \times 2014 = 6043 > 5832 = 18^3$.

遗留问题，如果要证明的是 $18.2 < a_{2015} < 18.3$ 呢（编程计算知这是成立的）？ □

题 3.7.5 数列 a_n 满足 $a_1 = 2$，$(n+1)a_{n+1}^2 = na_n^2 + a_n$，求证

$$\sum_{i=2}^{n} \frac{a_i^2}{i^2} < \frac{9}{5}$$

■

证明一. 由数列归纳法易证 $a_n > 1$，所以

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

于是数列递减，所以当 $n > 1$ 时，$a_n < a_1 = 2$

$$(n+1)a_{n+1}^2 = na_n^2 + a_n < na_n^2 + 2$$

于是累加下去，就有

$$na_n^2 < a_1^2 + 2(n-1) = 2(n+1)$$

所以

$$a_n^2 < 2\left(1 + \frac{1}{n}\right)$$

于是

$$\sum_{i=2}^{n} \frac{a_i^2}{i^2} < 2\left(\sum_{i=2}^{n} \frac{1}{i^2} + \sum_{i=2}^{n} \frac{1}{i^3}\right)$$

借用放缩

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

和

$$\frac{1}{i^3} < \frac{1}{(i-1)i(i+1)} = \frac{1}{2}\left(\frac{1}{(i-1)i} - \frac{1}{i(i+1)}\right)$$

从 $i \geqslant 4$ 开始放缩，累加即得

$$\begin{aligned}
\sum_{i=2}^{n} \frac{a_i^2}{i^2} &< 2\left(\frac{1}{4} + \frac{1}{9} + (\frac{1}{3} - \frac{1}{n}) + \frac{1}{8} + \frac{1}{27} + \frac{1}{2}(\frac{1}{12} - \frac{n}{n+1})\right) \\
&< 2\left(\frac{1}{4} + \frac{1}{9} + \frac{1}{3} + \frac{1}{8} + \frac{1}{27} + \frac{1}{24}\right) \\
&= 2\left(\frac{3}{4} + \frac{4}{27}\right) < \frac{9}{5}
\end{aligned}$$

□

证明二. 不以要证的不等式为目标，研究下这个数列的性态，因为

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

显然 $\frac{na_n+1}{n+1}$ 是 a_n 和 1 的加权平均，因为 $a_n > 1$ 有 $\frac{na_n+1}{n+1} < a_n$，所以有

$$\begin{aligned}
a_{n+1} &= \sqrt{a_n \cdot \frac{na_n + 1}{n+1}} \\
&< \frac{1}{2}\left(a_n + \frac{na_n + 1}{n+1}\right) \\
&= \frac{2n+1}{2n+2}a_n + \frac{1}{2n+2}
\end{aligned}$$

另一方面，由 $\frac{na_n+1}{n+1} < a_n$，所以

$$a_{n+1}^2 = a_n \cdot \frac{na_n+1}{n+1} > \left(\frac{na_n+1}{n+1}\right)^2$$

所以

$$a_{n+1} > \frac{n}{n+1}a_n + \frac{1}{n+1}$$

综合这两个估计，得到

$$\frac{n}{n+1}a_n + \frac{1}{n+1} < a_{n+1} < \frac{2n+1}{2n+2}a_n + \frac{1}{2n+2}$$

左右都是 a_n 和 1 的加权平均，只是权重不同，上式改写为

$$\frac{n}{n+1}(a_n-1) < a_{n+1}-1 < \frac{2n+1}{2n+2}(a_n-1)$$

所以最后就有估计式

$$1+\frac{1}{n} < a_n < 1+\frac{1}{2}\cdot\frac{(2n-1)!!}{(2n)!!}$$

对于后面的双阶乘，由熟知的放缩

$$\begin{aligned}
&\left(\frac{1}{2}\cdot\frac{3}{4}\cdots\frac{2n-1}{2n}\right)^2\\
=&\left(\frac{1}{2}\cdot\frac{1}{2}\right)\left(\frac{3}{4}\cdot\frac{3}{4}\right)\cdots\left(\frac{2n-1}{2n}\cdot\frac{2n-1}{2n}\right)\\
<&\left(\frac{1}{2}\cdot\frac{2}{3}\right)\left(\frac{3}{4}\cdot\frac{4}{5}\right)\cdots\left(\frac{2n-1}{2n}\cdot\frac{2n}{2n+1}\right)\\
=&\frac{1}{2n+1}
\end{aligned}$$

所以 a_n 的估计式两端都以 1 为极限，由夹逼定理，a_n 极限为 1.

而仍由那估计式，可以得出

$$a_n^2 < \left(1+\frac{1}{2\sqrt{2n+1}}\right)^2 < 2+\frac{1}{4}\frac{1}{2n+1} < 2+\frac{1}{8n}$$

由这不等式，仍同证明一中的放缩，同样可证得题目中的不等式。 □

题 3.7.6 [6] 数列 a_n 满足，$a_1=1$，$na_na_{n+1}=1$，求证:

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

2.
$$2(\sqrt{n+1}-1) \leqslant \frac{1}{2a_3}+\frac{1}{3a_4}+\cdots+\frac{1}{(n+1)a_{n+2}} \leqslant n$$

■

[6] 题目来自悠闲数学娱乐论坛.

证明. 首先在 $a_{n+1}a_n = \frac{1}{n}$ 中将 n 替换为 $n+1$ 得 $a_{n+2}a_n = \frac{1}{n+1}$，这两式相除即得

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

第一问得证。第二问，通项

$$\frac{1}{(k+1)a_{k+2}} = \frac{1}{ka_k} = a_{k+1}$$

所以只是要证明下式

$$2(\sqrt{n+1}-1) \leqslant a_2 + a_3 + \cdots + a_{n+1} \leqslant n$$

这只要能证明下面这个估计就能办到 $(n > 1)$

$$2(\sqrt{n}-\sqrt{n-1}) \leqslant a_n \leqslant 1$$

用数学归纳法试了一下，递推存在一点困难，倒是证明更强的放缩容易些 $(n > 1)$

$$\frac{1}{\sqrt{n-1}} \leqslant a_n \leqslant 1$$

以下就来数学归纳法吧，计算得 $a_2 = 1$，$a_3 = \frac{1}{2}$，都符合这不等式，于是假定 a_n 满足这不等式，来看 a_{n+2} 的情况:

$$a_{n+2} = \frac{n}{n+1}a_n < a_n \leqslant 1$$

同时

$$a_{n+2} = \frac{n}{n+1}a_n \geqslant \frac{n}{n+1} \cdot \frac{1}{\sqrt{n-1}} > \frac{1}{\sqrt{n+1}}$$

于是得证。说明：这里数归是从 n 到 $n+2$ 而不是 $n+1$ 是因为，试算了前面几项，发现它是一个锯齿数列，也就是偶数项都向 1 靠近，奇数项都向 0 靠近，所以就分别考虑两个子列了。 □

题 3.7.7 [7] 数列 a_n 满足: $a_1 = 1/3$，递推公式为 $a_{n+1} = a_n + \frac{a_n^2}{n^2}$，求证:

$$\frac{n}{2n+1} \leqslant a_n \leqslant \frac{2n-1}{2n+1}$$

■

证明. 这个递推公式比较有意思，在证出这题目后还可以进一步研究 a_1 的取值对数列的收敛性的影响，因为 $a_n = n$ 是符号这递推式的，而题目的结论表明当 a_1 取 1/3 时数列有上界，显然数列又是递增的，所以是收敛的。

左边不等式是很松的，因为数列是递增的，而

$$a_4 = \frac{30760}{59049} = 0.520923\cdots > \frac{1}{2} > \frac{n}{2n+1}$$

右端尚未有思路，但是根据程序显示的结果，数列在 a_4 就已经超过 0.57，但是直到 a_{100000} 都还没超过 0.61，所以此数列增长极其缓慢，从要证的结论来看它有上界，因而有极限，猜测这极限应该小于 1，若能证明这一点，则结论便能得到证明。 □

[7]来自 kuing 的悠闲数学娱乐论坛。

题 3.7.8 两个数列 a_n 和 b_n 满足: $a_1=0$, $b_1=\frac{1}{2}$，并有递推关系

$$a_{n+1}=\frac{a_n+b_n}{2},\ b_{n+1}=\sqrt{a_{n+1}b_n}$$

求证这两个数列有共同的极限，并求出这个极限。 ■

题目来源: 悠闲数学娱乐论坛,原帖链接: http://kuing.orzweb.net/viewthread.php?tid=4569&extra=page%3D1，这题目命题人 (网友" 其妙") 给出了它的几何背景，是周长为 2 的正 2^n 边形的内切圆半径和外接圆半径.

证明一. (解答于 2017-05-02) 因为

$$\frac{a_{n+1}}{b_{n+1}}=\frac{\frac{a_n+b_n}{2}}{\sqrt{a_{n+1}b_n}}=\frac{\frac{a_n+b_n}{2}}{\sqrt{\frac{a_n+b_n}{2}b_n}}=\sqrt{\frac{1+\frac{a_n}{b_n}}{2}}$$

所以令 $c_n=a_n/b_n$，就有

$$c_{n+1}=\sqrt{\frac{1+c_n}{2}}$$

受半角公式

$$\cos\frac{\theta}{2}=\sqrt{\frac{1+\cos\theta}{2}}$$

启发，令 $c_n=\cos\theta_n$，则可取 $\theta_{n+1}=\theta_n/2$，结合 $\cos\theta_1=c_1=a_1/b_1=0$ 可取 $\theta_1=\pi/2$，于是 $\theta_n=\pi/2^n$，所以得到

$$a_n=\cos\frac{\pi}{2^n}b_n$$

再回到 b_n 的递推式，有

$$b_{n+1}^2=a_{n+1}b_n=\cos\frac{\pi}{2^{n+1}}b_{n+1}b_n$$

所以

$$b_{n+1}=\cos\frac{\pi}{2^{n+1}}b_n$$

于是便不难求得

$$b_n=b_1\cos\frac{\pi}{2^2}\cos\frac{\pi}{2^3}\cdots\cos\frac{\pi}{2^n}$$

对这个余弦的连乘积，将它乘以 $\sin\frac{\pi}{2^n}$ 再利用正弦的二倍角公式便会发生连锁反应，反应的结果便是这连乘积等于 $\left(2^{n-1}\sin\frac{\pi}{2^n}\right)^{-1}$，所以

$$b_n=\frac{1}{2^n\sin\frac{\pi}{2^n}}$$

而

$$a_n=\cos\frac{\pi}{2^n}b_n=\frac{1}{2^n}\cot\frac{\pi}{2^n}$$

由熟知的极限 $\lim_{x\to 0}\frac{\sin x}{x}=1$ 便知 a_n 和 b_n 有共同的极限 $\frac{1}{\pi}$.

补充说明：在令 $c_n=\cos\theta_n$ 时尚需证明 $|c_n|\leqslant 1$，这利用 c_n 的递推式和初始值 $c_1=0$，由数学归纳法便知 $0\leqslant c_n\leqslant 1$，如果初始 $c_1>1$，便不能使用余弦了，但却可以使用双曲函数 $\cosh x=(e^x+e^{-x})/2$，双曲函数有着相同的倍半公式。 □

证明二. (解答于 2012-05-18) 首先在 a_n 的递推式两边取极限即知它俩如果收敛就必定收敛到相同的极限，将 $b_n=2a_{n+1}-a_n$ 代入 b_n 的递推式中换掉 b_n 和 b_{n+1} 得到

$$(2a_{n+2}-a_{n+1})^2=a_{n+1}(2a_{n+1}-a_n)$$

展开整理得

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

逐次把下标推下去，就有

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

把它视为关于 a_{n+1} 的二次方程，解之得

$$a_{n+1}=\frac{1}{2}\left(a_n+\sqrt{a_n^2+\frac{1}{4^n}}\right)$$

令 $a_n=1/(2^n\tan\theta_n)$，代入上式得 $\tan\theta_{n+1}=\tan\frac{\theta_n}{2}$，所以可以取 $\theta_{n+1}=\theta_n/2$，进而得 $\theta_n=\pi/2^n$，所以 $n>1$ 时有

$$a_n=\frac{1}{2^n\tan\frac{\pi}{2^n}}$$

比较 a_n 的递推式和定义式可得

$$b_n=\sqrt{a_n^2+\frac{1}{4^n}}=\frac{1}{2^n\sin\frac{\pi}{2^n}}$$

由熟知的极限 $\lim_{x\to 0}\frac{\sin x}{x}=1$ 即知 a_n 和 b_n 有共同的极限 $1/\pi$. □

题 3.7.9 (isee)[8] 如果无穷数列 $a_1,a_2,a_3,\dots$ 对任意不同的两个正整数 $i,j(i<j)$ 都有 $|a_i-a_j|\geqslant\dfrac{1}{j}$，那么就称其为“难亲数列”，问题是难亲数列是否一定无界？ ■

解答. 初步想了下，好像是可以的，而且所有点都可以被限制在一个长度为 2 的闭区间上，当前 n 个点都确定后，a_{n+1} 就不能落在前 n 个点以每一个点为中心半径为 $\dfrac{1}{n+1}$ 的开邻域内，这些邻域的总长不超过 $\dfrac{2n}{n+1}<2$，所以如果前 n 个点都被限制在一个长度为 2 的闭区间上的话，则这闭区间仍然有空隙可以容得下 a_{n+1}，因此结论应该是可以的。

再考虑构造出一个方法，让所有点都落在区间 $[0,2]$ 上，首先让 $a_1=0$，$a_2=2$，以后的点的确定，采用反复分割区间取中点的方法:

第一次分割前，区间数目为 1，每个区间长度为 2，按区间中点分割 $a_3=1$，分割后区间数目为 2，每个区间长度为 1.

第二次分割前，区间数目为 2，每个区间长度为 1，取各个区间中点为分割 $a_4=1/2$，$a_5=3/2$，分割后区间数目为 2^2，每个区间长度为 $\dfrac{1}{2}$.

第三次分割前，区间数目为 2^2，每个区间长度为 $\dfrac{1}{2}$，各个区间中点 $a_6=1/4$，$a_7=3/4$,$a_8=5/4$,$a_9=7/4$，分割后区间数目为 2^3，每个区间长度为 $\dfrac{1}{2^2}$.

[8]原帖见 http://kuing.orzweb.net/viewthread.php?tid=4853

……

第 $n+1$ 次分割前，区间数目为 2^n，每个区间长度为 $\dfrac{1}{2^{n-1}}$，各个区间中点 $a_{2^n+2+i}=\dfrac{2i+1}{2^n}(i=0,1,\cdots,2^n-1)$，分割后区间数目为 2^{n+1}，每个区间长度为 $\dfrac{1}{2^n}$.

按此方法可以确定 a_n，把 n 改写为 $n=2^m+2+i(i=0,1,\cdots,2^m-1)$，于是 $m=[\log_2(n-2)]$，$i=n-2^{[\log_2(n-2)]}$, 这里中括号是向下取整，在这种表示下，就有

$$a_{2^m+2+i}=\frac{2i+1}{2^m}$$

或者写成

$$a_n=\frac{1+2(n-2-2^{[\log_2(n-2)]})}{2^{[\log_2(n-2)]}}$$

后者其实还不如前一种好看。

在这种分法下，$a_{2^m+2+i}(i=0,1,\cdots,2^m-1)$ 是第 $m+1$ 次分割产生的分点，而在这一次分割后，每个区间的长度为 $\dfrac{1}{2^m}$，所以到这一次分割后为止，所得到的 a_i 中的最大下标是 $2^{m+1}+1$，即 $i=2^m-1$ 的那个分点, 此时任意两个 a_i 之差都 $\geqslant\dfrac{1}{2^m}>\dfrac{1}{2^m+2+i}$，因此，由这分割所确定的数列符合要求。 □

3.8 不等式

本节叙述一些在初等数学领域非常重要的不等式[9]，这些不等式应用广泛，刻画的不等式关系非常深刻。

3.8.1 绝对值不等式

定理 3.8.1. 对于任意两个实数 a 和 b，成立 $|a+b|\leqslant|a|+|b|$，等号成立的充分必要条件是 $ab\geqslant 0$.

证明很简单，略去。

推论 3.8.1. 对于任意两个实数 a 和 b，成立不等式 $||a|-|b||\leqslant|a+b|\leqslant|a|+|b|$，左边取等的条件是 $ab\leqslant 0$，右边取等条件是 $ab\geqslant 0$。

证明. 只需证明左边，因为 $|a|=|(a+b)-b|\leqslant|a+b|+|b|$，所以 $|a|-|b|\leqslant|a+b|$，类似还可得到 $|b|-|a|\leqslant|a+b|$，所以 $||a|-|b||\leqslant|a+b|$。 □

例 3.8.1 对于任意非零实数 x，成立不等式 $\left|x+\frac{1}{x}\right|\geqslant 2$.

最简单的证明就是由两项同号，有 $\left|x+\frac{1}{x}\right|=|x|+\frac{1}{|x|}\geqslant 2$，这里利用了二元均值不等式。 ■

推论 3.8.2. 对于任意 n 个实数 $a_i(i=1,2,\ldots,n)$，成立不等式 $|a_1+a_2+\cdots+a_n|\leqslant|a_1|+|a_2|+\cdots+|a_n|$。

[9]这一节主要参考了文献 [3] 和 [1] 以及 [2].

3.8.2 排序不等式

定理 3.8.2 (排序不等式). 对于任何两组按照相同的大小顺序排列好（均从小到大或者均从大到小）的两组实数 $x_i(i=1,2,\ldots,n)$ 和 $y_i(i=1,2,\ldots,n)$，有下面的不等式成立

$$\begin{aligned} & x_1y_n + x_2y_{n-1} + \cdots + x_ny_1 \\ \leqslant\ & x_1y_{r_1} + x_2y_{r_2} + \cdots + x_ny_{r_n} \\ \leqslant\ & x_1y_1 + x_2y_2 + \cdots + x_ny_n \end{aligned} \tag{3.8.1}$$

其中 $r_i(i=1,2,\ldots,n)$ 是 $1,2,\ldots,n$ 的任意一个排列。

不等式的通俗说法就是，反序和最小，同序和最大，乱序和居中。

证明. 对于中间的乱序和，如果 $r_1 \neq 1$，那么 y_1 就必定跟某个 x_m 搭配着，也就是 $x_1y_{r_1}$ 和 x_my_1 这两项，我们尝试交换一下这两个搭配，看看和会怎么变化

$$(x_1y_1 + x_my_{r_1}) - (x_1y_{r_1} + x_my_1) = (x_1 - x_m)(y_1 - y_{r_1}) \geqslant 0$$

因此这样做的结果是使得和式变大了，照此逐步调整下去，所得和式将会继续变大，直到所有 $r_i = i(i=1,2,\ldots,n)$ 都成立时得到同序和为最大，这就证明了不等式的左边，这个方法称为 *逐步调整法*。

至于不等式的右边，只要把第二组数 $y_1, y_2, \ldots, y_n$ 换为 $-y_n, -y_{n-1}, \ldots, -y_1$，则它与 x_i 单调性相同，把左边不等式应用到这两组数上即可得到右边不等式。 □

例 3.8.2 Nesbitt 不等式[10] 对于任意三个非负实数 a, b, c，有下面不等式成立

$$\frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} \geqslant \frac{3}{2} \tag{3.8.2}$$

证明. 由对称性，不失一般性，设 $a \geqslant b \geqslant c$，则也有 $a+b \geqslant c+a \geqslant b+c$，因而由排序不等式，有

$$\begin{aligned} \frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} &\geqslant \frac{a}{a+b} + \frac{b}{b+c} + \frac{c}{c+a} \\ \frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} &\geqslant \frac{b}{a+b} + \frac{c}{b+c} + \frac{a}{c+a} \end{aligned}$$

两式相加即得证。 □

■

[10] 来自于参考文献 [8]。

3.8.3 均值不等式

定理 3.8.3 (均值不等式). 对于任意 n 个正实数 $x_i(i=1,2,\ldots,n)$，记

$$H_n = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} \tag{3.8.3}$$

$$G_n = (\prod_{i=1}^{n} x_i)^{\frac{1}{n}} \tag{3.8.4}$$

$$A_n = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.8.5}$$

$$Q_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{3.8.6}$$

称 H_n 为这 n 个正实数的 调和平均数，而 A_n 为 算术平均数，G_n 为 几何平均数，Q_n 为 平方平均数，则有下面不等式成立

$$H_n \leqslant G_n \leqslant A_n \leqslant Q_n \tag{3.8.7}$$

其中的等号只有在 x_i 全部都相等时才成立。

我们在节 1.4中曾经利用多项式乘幂定理证明过中间的 $A_n \geqslant G_n$，在小节 1.2.2中也曾经使用倒推形式的数学归纳法证明过这个不等式，所以在这一节我们使用另外的方法。

证明. 先来证明 $A_n \geqslant G_n$.

因为 $\frac{x_1+x_2}{2} - \sqrt{x_1x_2} = \frac{1}{2}(\sqrt{x_1} - \sqrt{x_2})^2 \geqslant 0$，所以不等式 $A_2 \geqslant G_2$ 成立。

假定 $A_n \geqslant G_n$，来尝试证明 $A_{n+1} \geqslant G_{n+1}$，因为 A_{n+1} 是一个算术平均数，对一组数而言，如果新添加的数等于原来这些数的算术平均数，则算术平均数保持不变，于是有如下的策略

$$\begin{aligned}
A_{n+1} &= \frac{1}{n+1}\sum_{i=1}^{n+1} x_i \\
&= \frac{\sum_{i=1}^{n} x_i + [x_{n+1} + (n-1)A_{n+1}]}{2n} \\
&= \frac{1}{2}\left(\frac{x_1+\cdots+x_n}{n} + \frac{x_{n+1}+(n-1)A_{n+1}}{n}\right) \\
&\geqslant \frac{1}{2}\left(\sqrt[n]{x_1\cdots x_n} + \sqrt[n]{x_{n+1}A_{n+1}^{n-1}}\right) \\
&\geqslant \sqrt{\sqrt[n]{x_1\cdots x_n}\cdot\sqrt[n]{x_{n+1}A_{n+1}^{n-1}}} \\
&= \sqrt[2n]{x_1\cdots x_{n+1}A_{n+1}^{n-1}}
\end{aligned}$$

也就是 $A_{n+1}^{2n} \geqslant G_{n+1}^{n+1}A_{n+1}^{n-1}$，整理即得 $A_{n+1} \geqslant G_{n+1}$，得证，等号成立的条件也是显而易见的。

在 $A_n \geqslant G_n$ 中，把每个数 x_i 都换成 $\frac{1}{x_i}$，立即便得到 $\frac{1}{H_n} \geqslant \frac{1}{G_n}$，所以 $H_n \leqslant G_n$，至于 $A_n \leqslant Q_n$，利用数学归纳法和基本不等式 $2ab \leqslant a^2+b^2$ 就可以轻松获证，这里就不啰嗦了。

下面再提供 $A_n \geqslant G_n$ 的另一证法.

这个不等式可以改写为如下形式

$$\frac{x_1}{G_n}+\frac{x_2}{G_n}+\cdots+\frac{x_n}{G_n} \geqslant n$$

因为左边各项之积为 1，所以我们只要能够证明：如果 n 个正实数 $x_i(i=1,2,\ldots,n)$ 满足 $x_1x_2\cdots x_n=1$，则 $x_1+x_2+\cdots+x_n \geqslant n$ 就可以了。

对于 $n=2$ 的情况是很容易验证的，假定对于 n 个数也成立，现在来看 $n+1$ 个数的情况，先把 x_nx_{n+1} 绑在一起视为一个数，这样 $n+1$ 个数就成了 n 个数，根据归纳假设可得 $x_1+\cdots+x_{n-1}+x_nx_{n+1} \geqslant n$，于是我们只要证明 $x_n+x_{n+1} \geqslant x_nx_{n+1}+1$ 就可以了，也就是要证明 $(x_n-1)(x_{n+1}-1) \leqslant 0$，这个似乎不一定成立，但是乘积为 1 的 $n+1$ 个正实数，必定至少有一个大于等于 1，同时至少有一个小于等于 1，我们可以重新排序这些正实数，使得这两个数放在最后（或者干脆一开始绑定两个数的时候就绑这两个数），于是不等式 $(x_n-1)(x_{n+1}-1) \leqslant 0$ 就成立了，所以得证。 □

证明三. 仍同证明二，只要证明辅助命题: 如果 n 个正实数 $x_i(i=1,2,\ldots,n)$ 乘积为 1，则它们的和不小于 n 就可以了。

利用排序不等式，这 n 个正实数乘积为 1，则存在另外 n 个正实数 $y_i(i=1,2,\ldots,n)$ 使得 $x_1=y_1/y_2, x_2=y_2/y_3,\ldots,x_n=y_n/y_1$，于是对于序列 y_i 和 $1/y_i$，就有 (利用了乱序和大于等于反序和)

$$\begin{aligned} x_1+x_2+\cdots+x_n &= \frac{y_1}{y_2}+\frac{y_2}{y_3}+\cdots+\frac{y_n}{y_1} \\ &\geqslant \frac{y_1}{y_1}+\frac{y_2}{y_2}+\cdots+\frac{y_n}{y_n}=n \end{aligned}$$

所以辅助命题成立，从而得证。 □

例 3.8.3 均值不等式中的几个项都有相似的构成形式: $f^{-1}\left(\frac{1}{n}\sum_{i=1}^n f(x_i)\right)$，对于 A_n，$f(x)=x$，对于 H_n，$f(x)=\frac{1}{x}$，对于 G_n，唔...$f(x)=\ln x$，对于 Q_n，$f(x)=x^2$，所以这些平均数的大小关系完全取决于对应函数的性质，从几何意义上看，比算术平均数小的平均数所对应函数都是向左凸起的，而比算术平均数大的平均数所对应函数则是向右凸起的。 ■

例 3.8.4 匀变速直线运动中间时刻瞬时速度和中间位移瞬时速度的大小 我们来考察一下匀变速直线运动中，中间时刻的瞬时速度和经过中间位移处瞬时速度的大小关系。因为加速度是恒定的，速度是时间的一次函数，所以对于起始时刻 t_1 和终止时刻 t_2 来说，中间时刻的瞬时速度是起始速度 v_1 和终止速度 v_2 的算术平均 $\frac{1}{2}(v_1+v_2)$，根据速度公式 $v(t)=v_1+at$ 和位移公式 $s(t)=v_1t+\frac{1}{2}at^2$(其中 a 是加速度) 可得出经过中间位移处的瞬时速度是 $\sqrt{\frac{v_1^2+v_2^2}{2}}$，由均值不等式即知中间位移处的瞬时速度要大一些。

■

例 3.8.5　我们在例 3.8.2中已经证过 Nesbitt 不等式，那时利用的是排序不等式，现在我们用均值不等式加以证明，由 AM-GM 不等式得

$$\frac{a+b}{b+c}+\frac{b+c}{c+a}+\frac{c+a}{a+b}\geqslant 3$$

以及

$$\frac{a+c}{b+c}+\frac{b+a}{c+a}+\frac{c+b}{a+b}\geqslant 3$$

两式相加即得证。 ■

例 3.8.6　在这个例子中,我们来建立一个有用的不等式。设 $x_i>-1(i=1,2,\ldots,n)$，记 $s=\sum_{i=1}^n x_i$，则有

$$\prod_{i=1}^n(1+x_i)<1+s+\frac{s^2}{2!}+\cdots+\frac{s^n}{n!}$$

并且，如果所有 x_i 的符号都相同，则还有

$$\prod_{i=1}^n(1+x_i)\geqslant 1+s$$

证明. 因 $1+x_i>0$，由 AM-GM 不等式得

$$\prod_{i=1}^n(1+x_i)\leqslant\left(1+\frac{s}{n}\right)^n=\sum_{k=0}^n\frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{s^n}{k!}<\sum_{k=0}^n\frac{s^k}{k!}$$

第一个不等式得证，再证第二个，如果 x_i 的符号都相同（允许有零），则按归纳法，$n=1$ 时显然是成立的，假定对 n 个数时也是成立的，那么对于 $n+1$ 个数，则有

$$\prod_{i=1}^n(1+x_i)\geqslant(1+x_{n+1})(1+\sum_{i=1}^n x_i)=1+\sum_{i=1}^{n+1}x_i+x_{n+1}\sum_{i=1}^n x_i\geqslant 1+\sum_{i=1}^{n+1}x_i$$

□

■

对于均值不等式中最基本的 $G_n\leqslant A_n$，将每个数的权重一般化，得如下的加权形式

定理 3.8.4 (加权平均值不等式). 对于任意 n 个正实数 $x_i(i=1,2,\ldots,n)$ 和一组权值 $\alpha_i\geqslant 0(i=1,2,\ldots,n),\sum_{i=1}^n\alpha_i=1$，有如下的不等式

$$\sum_{i=1}^n\alpha_i x_i\geqslant\prod_{i=1}^n x_i^{\alpha_i}\tag{3.8.8}$$

等号也只有在 x_i 全部都相等时才成立.

证明. 先证 $\alpha_i(i=1,2,\ldots,n)$ 都是有理数的情况，此时存在非负整数 $p_i(i=1,2,\ldots,n)$ 使得 $\alpha_i=\frac{p_i}{\sum_{k=1}^n p_k}$，只要把 p_ix_i 看成 p_i 个 x_i 相加，就有

$$\begin{aligned}\sum_{i=1}^n \alpha_i x_i &= \sum_{i=1}^n \frac{p_i}{\sum_{k=1}^n p_k} x_i \\ &= \frac{1}{\sum_{i=1}^n p_i}\cdot\sum_{i=1}^n p_i x_i \\ &\geqslant \left(\prod_{i=1}^n x_i^{p_i}\right)^{\frac{1}{\sum_{i=1}^n p_i}} \\ &= \prod_{i=1}^n x_i^{\alpha_i}\end{aligned}$$

所以不等式得证。

对于某个 α_i 为无理数的情况，使用有理数序列去逼近它，再两边取极限即得证。 □

3.8.4 柯西 (Cauchy) 不等式

定理 3.8.5 (柯西不等式). 对于任何两组实数 $x_i(i=1,2,\ldots,n)$ 和 $y_i(i=1,2,\ldots,n)$，有下面不等式成立

$$\left(\sum_{i=1}^n x_i y_i\right)^2 \leqslant \left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n y_i^2\right) \tag{3.8.9}$$

其中等号成立的唯一场景是：存在一个共同的实数 λ 使得 $x_i=\lambda y_i(i=1,2,\ldots,n)$ 都成立。

等号成立的条件，通俗的说就是两组实数成比例，如果我们约定分数中的分母可以为零，并且此时分子也必须为零，则可以把取等条件改写为

$$\frac{x_1}{y_1}=\frac{x_2}{y_2}=\cdots=\frac{x_n}{y_n}$$

如果记向量 $\vec{a}=(x_1,x_2,\ldots,x_n), \vec{b}=(y_1,y_2,\ldots,y_n)$，则柯西不等式即表明 $|\vec{a}\cdot\vec{b}|\leqslant|\vec{a}|\cdot|\vec{b}|$.

证明很简单:

$$\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n x_i^2\right)-\left(\sum_{i=1}^n x_i y_i\right)^2=\sum_{1\leqslant i<j\leqslant n}(x_iy_j-x_jy_i)^2\geqslant 0$$

但更为流行的是下面的证明:

证明. 构造二次函数

$$f(t)=t^2\sum_{i=1}^n x_i^2+2t\sum_{i=1}^n x_iy_i+\sum_{i=1}^n y_i=\sum_{i=1}^n(x_it+y_i)^2\geqslant 0$$

既然是恒为非负的二次函数，其判别式必小于等于零，于是即得柯西不等式，等号成立的条件就是二次函数有零点 t_0，使得 $x_it_0+y_i=0(i=1,2,\ldots,n)$，也就是两组数成比例。 □

例 3.8.7 高中数学教材上有一类习题，是说点 $P(x,y)$ 是椭圆 $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1(a>0,b>0)$ 上任一点，λ 和 μ 是两个固定的实数，求 $\lambda x+\mu y$ 的最值。

这题目可以使用判别式来求解，也可以使用椭圆参数方程来求，最直接的就是使用柯西不等式:

$$(\lambda x+\mu y)^2=(a\lambda\cdot\frac{x}{a}+b\mu\cdot\frac{y}{b})^2\leqslant(a^2\lambda^2+b^2\mu^2)(\frac{x^2}{a^2}+\frac{y^2}{b^2})=a^2\lambda^2+b^2\mu^2$$

可以验证等号是可以取到的。 ■

本书倾向于认为在均值不等式、柯西不等式、排序不等式中，排序不等式是最基本的不等式，因为从1.4节利用多项式乘幂定理证明均值不等式的过程中可以看出，均值不等式赖以成立的基础是不等式 $x^n+y^n\geqslant x^{n-1}y+xy^{n-1}(n\in N_+)$，而柯西不等式则是得益于不等式 $x^2+y^2\geqslant 2xy$，而这两个简单不等式都是排序不等式的直接推论。

3.8.5 切比雪夫 (Chebyshev) 不等式

定理 3.8.6 (切比雪夫不等式). 对于任意两组按照相同的大小顺序排列好的实数 $x_i(i=1,2,\ldots,n)$ 和 $y_i(i=1,2,\ldots,n)$，成立着下面不等式

$$\frac{1}{n}\sum_{i=1}^{n}x_iy_{n+1-i}\leqslant\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right)\leqslant\frac{1}{n}\sum_{i=1}^{n}x_iy_i \tag{3.8.10}$$

其中的等号成立的唯一场景是：两组数中至少有一组数的值都相同。

切比雪夫不等式是排序不等式的直接推论，这从如下的证明过程即可看出。

证明. 先证明右边不等式，由排序不等式，可得如下的 n 个等式或不等式:

$$\begin{aligned}
x_1y_1+x_2y_2+\cdots+x_ny_n &= x_1y_1+x_2y_2+\cdots+x_ny_n\\
x_1y_2+x_2y_3+\cdots+x_ny_1 &\leqslant x_1y_1+x_2y_2+\cdots+x_ny_n\\
x_1y_3+x_2y_4+\cdots+x_ny_2 &\leqslant x_1y_1+x_2y_2+\cdots+x_ny_n\\
&\cdots\\
x_1y_n+x_2y_1+\cdots+x_ny_{n-1} &\leqslant x_1y_1+x_2y_2+\cdots+x_ny_n
\end{aligned}$$

将这些等式或不等式全部相加，即得切比雪夫不等式的右边。

至于不等式的左边，只要在把序列 $y_1,y_2,\ldots,y_n$ 替换为序列 $-y_n,-y_{n-1},\ldots,-y_1$ 后应用不等式的右边即可得出 (当然也可以使用排序不等式的左边也像刚才那样累加得出)。 □

3.8.6 琴生不等式

琴生不等式是一个与函数凸性有关的不等式。

定理 3.8.7 (琴生不等式). 设函数 $f(x)$ 在区间 D 上是上凸函数，则对于该区间上任意 n 个实数 $x_i(i=1,2,\ldots,n)$，有如下不等式成立:

$$f\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)\geqslant\frac{1}{n}\sum_{i=1}^{n}f(x_i) \tag{3.8.11}$$

同样有加权形式的琴生不等式

定理 3.8.8 (加权形式的琴生不等式). 设函数 $f(x)$ 在区间 D 上是上凸函数,则对于该区间上任意 n 个实数 $x_i(i=1,2,\ldots,n)$ 和一组权值 $\alpha_i \geqslant 0(i=1,2,\ldots,n), \sum_{i=1}^n \alpha_i = 1$,有如下不等式成立:

$$f\left(\sum_{i=1}^n \alpha_i x_i\right) \geqslant \sum_{i=1}^n \alpha_i f(x_i) \tag{3.8.12}$$

我们在3.6节已经证明过琴生不等式，所以这里就省去了。

3.8.7 幂平均值不等式

定理 3.8.9 (幂平均值不等式). 对于任意 n 个正实数 $x_i(i=1,2,\ldots,n)$ 和两个正实数 $0<p<q$，成立着不等式

$$\left(\frac{1}{n}\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}} \leqslant \left(\frac{1}{n}\sum_{i=1}^n x_i^q\right)^{\frac{1}{q}} \tag{3.8.13}$$

等号当且仅当 $x_i(i=1,2,\ldots,n)$ 全部相等时成立。

取 $p=1,q=2$，便可得算术平均数小于等于平方平均数。

将每个数的权重 $\frac{1}{n}$ 一般化，得如下的加权形式

定理 3.8.10 (加权幂平均值不等式). 对任意 n 个正实数 $x_i(i=1,2,\ldots,n)$ 和任意两个正实数 $0<p<q$，以及一组权值 $\alpha_i \geqslant 0(i=1,2,\ldots,n), \sum_{i=1}^n \alpha_i = 1$，成立着下面不等式

$$\left(\sum_{i=1}^n \alpha_i x_i^p\right)^{\frac{1}{p}} \leqslant \left(\sum_{i=1}^n \alpha_i x_i^q\right)^{\frac{1}{q}} \tag{3.8.14}$$

3.8.8 赫尔德 (Holder) 不等式

定理 3.8.11 (赫尔德不等式). 设 $x_i>0, y_i>0(i=1,2,\ldots,n)$ 且 $\frac{1}{p}+\frac{1}{q}=1$，那么在 $p>1$ 时成立下面的不等式

$$\sum_{i=1}^n x_i y_i \leqslant \left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^n x_i^q\right)^{\frac{1}{q}} \tag{3.8.15}$$

在 $p<1, p\neq 0$ 不等式反向，等号当且仅当向量 $(x_1^p,\ldots,x_n^p)$ 和 $(y_1^q,\ldots,y_n^q)$ 共线时成立。

在赫尔德不等式中，取 $p=q=2$，即得柯西不等式。

3.8.9 闵可夫斯基 (Minkowski) 不等式

定理 3.8.12 (闵可夫斯基不等式). 设 $x_i>0, y_i>0(i=1,2,\ldots,n)$，那么在 $p>1$ 时成立下面的不等式

$$\left(\sum_{i=1}^n (x_i+y_i)^p\right)^{\frac{1}{p}} \leqslant \left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^n y_i^p\right)^{\frac{1}{p}} \tag{3.8.16}$$

在 $p < 1$ 时不等式反向，等号当且仅当存在一个共同的实数 λ 使得 $x_i = \lambda y_i (i = 1, 2, \ldots, n)$ 时成立。

在闵可夫斯基不等式中取 $n = 2, p = 2$，便是三角形两边之和大于第三边的坐标表达。

3.8.10 舒尔不等式

3.8.11 嵌入不等式

定理 3.8.13 (嵌入不等式)**.** 对 $\triangle ABC$ 和任意的实数 x, y, z 均有下面不等式成立

$$x^2 + y^2 + z^2 \geqslant 2yz\cos A + 2zx\cos B + 2xy\cos C \tag{3.8.17}$$

其中，当且仅当 $x : y : z = \sin A : \sin B : \sin C$ 时等号成立。

3.8.12 一些例子

例 3.8.8 Nesbitt 不等式的其它证法 我们在例 3.8.2和例 3.8.5中分别利用排序不等式和均值不等式证明了 Nesbitt 不等式，此处再给几个证明。

证明一. 由齐次性，设 $a + b + c = 3$，则只需证

$$\frac{a}{3-a} + \frac{b}{3-b} + \frac{c}{3-c} \geqslant \frac{3}{2}$$

由取等条件 $a = b = c = 1$ 出发，上切线法，当 $0 < a \leqslant 3$ 时，有

$$\frac{a}{3-a} \geqslant \frac{1}{2} + \frac{3}{4}(a-1)$$

于是

$$\sum_{cyc} \frac{a}{3-a} \geqslant \frac{3}{2} + \frac{3}{4}\sum_{cyc}(a-1) = \frac{3}{2}$$

□

证明二. 看着三个分母，想用等比定理想得痒痒的，试了一下，还真的可以:

$$\sum_{cyc} \frac{a}{b+c} = \sum_{cyc} \frac{a + \frac{a^2}{b+c}}{a+b+c} = 1 + \frac{1}{a+b+c}\sum_{cyc} \frac{a^2}{b+c}$$

所以只要证明

$$\sum_{cyc} \frac{a^2}{b+c} \geqslant \frac{1}{2}(a+b+c)$$

待定系数 λ，由 AM-GM 有

$$\lambda^2(b+c) + \frac{a^2}{b+c} \geqslant 2\lambda a$$

所以

$$2\lambda^2(a+b+c)+\sum_{cyc}\frac{a^2}{b+c}\geqslant 2\lambda(a+b+c)$$

即

$$\sum_{cyc}\frac{a^2}{b+c}\geqslant 2\lambda(1-\lambda)(a+b+c)$$

为了照顾取等条件，取 $\lambda=\frac{1}{2}$，即得

$$\sum_{cyc}\frac{a^2}{b+c}\geqslant\frac{1}{2}(a+b+c)$$

□

证明三.（导数法）由齐次性，假设 $a+b+c=1$，则只需证

$$\frac{a}{1-a}+\frac{b}{1-b}+\frac{c}{1-c}\geqslant\frac{3}{2}$$

固定 a，让 b 和 c 变化，考虑此时的最小值，因为有 $b+c=1-a$ 的限制，所以左端实际上是 b 的一元函数

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

求导得

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

易见当 $b=\frac{1-a}{2}$ 时达到最小值，所以

$$f(b)\geqslant f\left(\frac{1-a}{2}\right)=\frac{a}{1-a}+\frac{2(1-a)}{1+a}$$

记上式右边为 $g(a)$，再次求导得

$$g'(a)=-\frac{(3-a)(1-3a)}{(1-a)^2(1+a)^2}$$

显见 $g(a)$ 在 $a=\frac{1}{3}$ 时取得最小值，所以

$$g(a)\geqslant g\left(\frac{1}{3}\right)=\frac{3}{2}$$

□

■

3.8.13 题选

题 3.8.1 已知三个正实数 a,b,c 满足

$$\frac{1}{1+a}+\frac{1}{1+b}+\frac{1}{1+c}=1$$

求证

$$(a-1)(b-1)(c-1)\leqslant 1$$

■

证明. 作代换

$$\frac{1}{1+a}=\frac{x}{x+y+z},\ \frac{1}{1+b}=\frac{y}{x+y+z},\ \frac{1}{1+c}=\frac{z}{x+y+z}$$

其中 x,y,z 是三个正实数，于是只要证明

$$(y+z-x)(z+x-y)(x+y-z)\leqslant xyz$$

如果左边为负，不等式显然成立，否则可再作代换

$$r=y+z-x,\ s=z+x-y,\ t=x+y-z$$

于是只要证明

$$rst\leqslant\frac{s+t}{2}\frac{t+r}{2}\frac{r+s}{2}$$

由均值，这显然。 □

题 3.8.2 [11] 设 a,b,c 是三个互不相等的实数，求证

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

■

证明. 作代换

$$x=\frac{a}{a-b},\ y=\frac{b}{b-c}\ z=\frac{c}{c-a}$$

则只需证 $x^2+y^2+z^2\geqslant 1$，而在这代换下，易知

$$\left(1-\frac{1}{x}\right)\left(1-\frac{1}{y}\right)\left(1-\frac{1}{z}\right)=1$$

整理即为

$$x+y+z=(xy+yz+zx)+1$$

记等式左右两边的公共值为 t，则

$$x^2+y^2+z^2=(x+y+z)^2-2(xy+yz+zx)=t^2-2(t-1)=(t-1)^2+1\geqslant 1$$

得证。 □

题 3.8.3 [12] 已知非负实数 x、y、z 满足 $x+y+z=1$，求证:

$$1\leqslant\frac{x}{1-yz}+\frac{y}{1-zx}+\frac{z}{1-xy}\leqslant\frac{9}{8}$$

■

[11] 2008 年 IMO 试题.

[12] 2017 年德国数学奥林匹克试题.

证明. 左边有最直接的证明，因为每个分母都小于等于 1，所以有

$$\frac{x}{1-yz}+\frac{y}{1-zx}+\frac{z}{1-xy}\geqslant x+y+z=1$$

此外我们还有以下的换元法证明: 显见 $0\leqslant x,y,z\leqslant 1$，作代换

$$a=\frac{x}{1-yz},\ b=\frac{y}{1-zx},\ c=\frac{z}{1-xy}$$

只需证明 $1\leqslant a+b+c\leqslant 9/8$，在这代换下有 $a,b,c\geqslant 0$，并且

$$a=x+ayz$$
$$b=y+bzx$$
$$c=z+cxy$$

三式相加并利用 $x+y+z=1$ 得

$$a+b+c=1+ayz+bzx+cxy$$

显然右边大于等于 1，所以 $a+b+c\geqslant 1$，原不等式左边得证。右边待证。 □

题 3.8.4 (Sqing55) 已知实数 $a,b,c>0$，且 $a+b+c=1$，求证: 当 $\lambda\geqslant\frac{1}{5}$ 时有下面不等式成立:

$$\frac{a+\lambda}{a+bc}+\frac{b+\lambda}{b+ca}+\frac{c+\lambda}{c+ab}\geqslant\frac{9}{4}(1+3\lambda)$$

证明. (在 Kuing 的提示下完成)，先分离出 λ

$$f(\lambda)=\left(\frac{1}{a+bc}+\frac{1}{b+ca}+\frac{1}{c+ab}-\frac{27}{4}\right)\lambda+\frac{a}{a+bc}+\frac{b}{b+ca}+\frac{c}{c+ab}\geqslant\frac{9}{4}$$

我们先证明 λ 的系数是非负的，这样就有 $f(\lambda)\geqslant f\left(\frac{1}{5}\right)$，从而只要证明 $f\left(\frac{1}{5}\right)\geqslant\frac{9}{4}$ 就可以了。

因为 $a+b+c=1$，所以

$$\frac{1}{a+bc}\geqslant\frac{1}{a+\left(\frac{b+c}{2}\right)^2}=\frac{1}{a+\left(\frac{1-a}{2}\right)^2}=\frac{4}{(a+1)^2}$$

由切线法，易证

$$\frac{4}{(a+1)^2}\geqslant\frac{9}{4}-\frac{27}{8}\left(a-\frac{1}{3}\right)$$

于是

$$\sum\frac{1}{a+bc}\geqslant\frac{27}{4}$$

于是 $f(\lambda)$ 是单调不减的，所以接下来证明 $f\left(\frac{1}{5}\right)\geqslant\frac{9}{4}$，以下过程中 $\lambda=1/5$，因为

$$\sum\frac{a+\lambda}{a+bc}=\sum\frac{a+\lambda}{(a+b)(a+c)}$$

所以只需证明

$$\sum(a+\lambda)(b+c)\geqslant\frac{9}{4}(a+b)(b+c)(c+a)$$

为齐次化，左边作处理

$$(a+\lambda)(b+c)=(a+\lambda(a+b+c))(b+c)(a+b+c)$$

于是只要证明

$$(a+b+c)\sum(b+c)((1+\lambda)a+b+c)\geqslant\frac{9}{4}(a+b)(b+c)(c+a)$$

暴力展开，整理后，即是要证

$$a^3+b^3+c^3+3abc\geqslant a^2(b+c)+b^2(c+a)+c^2(a+b)$$

这等价于熟知的

$$abc\geqslant(b+c-a)(c+a-b)(a+b-c)$$

□

■

题 3.8.5　设正实数 a、b、c 满足 $a+b+c=1$，求证

$$\frac{1+a}{1-a}+\frac{1+b}{1-b}+\frac{1+c}{1-c}\leqslant 2\left(\frac{a}{b}+\frac{b}{c}+\frac{c}{a}\right)$$

■

3.9　题集

第四章　初等几何

4.1　向量

4.1.1　概念及运算

物理学上有一类物理量，不但有大小，还有方向，例如力、位移、动量，它们有一些共同的运算法则，在数学上解决一些几何问题时，引入带正负符号的线段，有时更是引入有向线段也会使问题更易于表达，在此基础上抽象出一个数学模型便称为向量。

定义 4.1.1. *既有大小也有方向的量称为* 向量，*也称为* 矢量，*向量的大小也称为它的*模。

向量可以用有向线段来表示，此时有向线段的长度就代表它的大小，它的方向就代表向量的方向，向量用粗体符号 $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \ldots$ 来表示，手写体可以用带箭头的字母来表示 $\vec{a}, \vec{b}, \vec{c}, \ldots$，在用有向线段表示时，也用它的起点和终点表示成 $\overrightarrow{AB}$，其中 A 是有向线段的起点，而 B 是有向线段的终点。

向量没有位置的概念，把一个有向线段的起点和终点移动到任何别的地方，只要新的有向线段的长度与原来相等，而方向与原来相同，则两个有向量线段代表的是同一个向量。今后对有向线段和向量不进行严格区分。

向量的模借用绝对值记号，如 $|\boldsymbol{\alpha}|, |\vec{a}|, |\overrightarrow{AB}|$.

定义 4.1.2. *大小为 1 的向量称为单位向量，单位向量用 $\boldsymbol{e}$ 表示，大小为零的向量称为*零向量，*记作 $\mathbf{0}$，零向量的方向没有实际意义，在具体问题中可以作灵活约定。*

定义 4.1.3. *方向相同或相反的向量称为* 共线向量，*约定零向量与任意向量共线，两个向量 $\boldsymbol{\alpha}$ 及 $\boldsymbol{\beta}$ 共线记作 $\boldsymbol{\alpha} \parallel \boldsymbol{\beta}$. 如果两个向量的大小相等，方向相同，则称这两个向量*相等，*记作 $\boldsymbol{\alpha} = \boldsymbol{\beta}$。大小相同，方向相反的两个向量互为*相反向量.

注意共线向量只是就两个向量的方向而言，并不是说代表两个向量的两个有向线段在同一直线上，当然，同一直线上的两个有向线段代表的两个向量必然共线。

定义 4.1.4. *对于空间中的多个向量，如果通过平移使得它们的起点相同时，起点与所有的终在同一平面内，则称这些向量是*共面向量。*否则便称它们不共面。*

对于两个向量，还可以引入夹角概念:

定义 4.1.5. 通过平移使得两个向量 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 起点相同，此时两个向量所在两条射线的夹角便称为这两个向量的 夹角，记作 $<\boldsymbol{\alpha},\boldsymbol{\beta}>$，其取值范围是 $[0,\pi]$，夹角为直角的两个向量称为互相 垂直，记作 $\boldsymbol{\alpha}\perp\boldsymbol{\beta}$.

显然，两个共线向量的夹角是 0 或者 π(不考虑零向量)。

定义 4.1.6. 两个向量可以进行相加，通过平移使得第二个向量的起点与第一个向量的终点重合，这时由第一个向量的起点为起点并以第二个向量的终点为终点的向量，便称为这两个向量的和，即 $\overrightarrow{AB}+\overrightarrow{BC}=\overrightarrow{AC}$.

将这法则用有向线段画出来，便是一个三角形，所以这个定义称为向量加法的三角形法则，向量加法也可以使用平行四边形法则，即让两个向量的起点重合，以这两个有向线段为邻边作平行四边形 (若两个向量是共线的，则这是一个压扁退化的平行四边形)，则由共同的起点指向平行四边形另一对角顶点所得向量即代表它俩的和，显然这与三角形法则是等价的。

不难验证，向量加法符合交换律和结合律，即

$$\begin{aligned}\boldsymbol{\alpha}+\boldsymbol{\beta} &= \boldsymbol{\beta}+\boldsymbol{\alpha}\\ (\boldsymbol{\alpha}+\boldsymbol{\beta})+\boldsymbol{\gamma} &= \alpha+(\boldsymbol{\beta}+\boldsymbol{\gamma})\end{aligned}$$

类比于数，将 $\boldsymbol{\alpha}+\boldsymbol{\alpha}$ 记为 $2\boldsymbol{\alpha}$，由此引出实数与向量的乘法运算，这运算定义如下

定义 4.1.7. 实数 λ 与向量 $\boldsymbol{\alpha}$ 的乘积也是一个向量，记为 $\lambda\boldsymbol{\alpha}$，它的大小是 $|\lambda\boldsymbol{\alpha}|=|\lambda||\boldsymbol{\alpha}|$，它的方向在 $\lambda>0$ 的情况下与 $\boldsymbol{\alpha}$ 相同，在 $\lambda<0$ 的情况下与 $\boldsymbol{\alpha}$ 相反，在 $\lambda=0$ 时，$0\boldsymbol{\alpha}=\mathbf{0}$ 是零向量。

由定义，一个向量 $\boldsymbol{\alpha}$ 的相反向量是 $(-1)\boldsymbol{\alpha}$，简记为 $-\boldsymbol{\alpha}$，即 $\boldsymbol{\alpha}+(-\boldsymbol{\alpha})=\mathbf{0}$.

不难得到，数与向量的乘法运算满足两种形式的分配律以及对实数的结合律

$$\begin{aligned}(\lambda+\mu)\boldsymbol{\alpha} &= \lambda\boldsymbol{\alpha}+\mu\boldsymbol{\alpha}\\ \lambda(\boldsymbol{\alpha}+\boldsymbol{\beta}) &= \lambda\boldsymbol{\alpha}+\lambda\boldsymbol{\beta}\\ \lambda(\mu)\boldsymbol{\alpha} &= (\lambda\mu)\boldsymbol{\alpha}\end{aligned}$$

在此基础上可以定义向量的减法

定义 4.1.8. 向量 $\boldsymbol{\alpha}$ 与向量 $\boldsymbol{\beta}$ 的差定义为向量 $\boldsymbol{\alpha}$ 与向量 $\boldsymbol{\beta}$ 的相反向量相加，即 $\boldsymbol{\alpha}-\boldsymbol{\beta}=\boldsymbol{\alpha}+(-\boldsymbol{\beta})$.

向量减法也有明确的几何意义，因为 $\overrightarrow{AB}-\overrightarrow{AC}=\overrightarrow{AB}+(-\overrightarrow{AC})=\overrightarrow{AB}+\overrightarrow{CA}=\overrightarrow{CA}+\overrightarrow{AB}=\overrightarrow{CB}$，因此，只要让两个向量的起点重合，由减向量的终点指向被减向量的终点所得向量即是它们的差，这就是向量减法的三角形法则，显然，任意向量减去它自身，所得差为零向量。

不难看出，对于两个向量构成的平行四边形中，两条对角形所代表的向量，便分别是它们的和与差所代表的向量，后文将由此得出一个平行四边形的性质，显然，两个向量的和向量与差向量，与原来的两个向量共面。

而对于空间中的三个向量相加，很容易把平行四边形法则推广为平行六面体法则，即让三个向量的起点相同，以这三个向量为邻边作平行六面体，则以共同起点到对角顶点的向量就代表它们的和向量。

关于向量共线的一条重要结果是

定理 4.1.1 (向量共线定理). *向量 $\boldsymbol{\alpha}$ 与非零向量 $\boldsymbol{\beta}$ 共线的充分必要条件是存在唯一实数 λ，使得 $\boldsymbol{\alpha}=\lambda\boldsymbol{\beta}$.*

证明. 充分性可由实数与向量乘法的定义得到，下证必要性，若向量 $\boldsymbol{\alpha}$ 与非零向量 $\boldsymbol{\beta}$ 共线，如果 $\boldsymbol{\alpha}$ 为零向量，则 $\lambda=0$ 就符合要求，在 $\boldsymbol{\alpha}$ 为非零向量的情况下，根据共线的定义，$\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的方向相同或相反，取实数 λ 满足 $|\lambda|=\dfrac{|\alpha|}{|\beta|}$，并根据方向相反还是相反来取定其符号，则显然有 $\boldsymbol{\alpha}=\lambda\boldsymbol{\beta}$，又若还存在另一实数 μ 使得 $\boldsymbol{\alpha}=\mu\boldsymbol{\beta}$，则有 $(\lambda-\mu)\boldsymbol{\beta}=\mathbf{0}$，而 $\boldsymbol{\beta}$ 是非零向量，根据数与向量的乘法定义，这必须 $\lambda=\mu$。 □

4.1.2 分解定理

对于三个共面向量 $\boldsymbol{\alpha}$,$\boldsymbol{\beta}$,$\boldsymbol{\gamma}$，通过平移使它们的起点重合，并记共同起点为 O 以及 $\overrightarrow{OA}=\boldsymbol{\alpha}$,$\overrightarrow{OB}=\boldsymbol{\beta}$,$\overrightarrow{OC}=\boldsymbol{\gamma}$，若 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 不共线，则过 $\boldsymbol{\gamma}$ 的终点 C 分别作直线 OB 与 OA 的平行线，分别与直线 OA 及 OB 相交于 M 和 N，则显然 $\overrightarrow{OM}$ 与 $\boldsymbol{\alpha}$ 共线，$\overrightarrow{ON}$ 与 $\boldsymbol{\beta}$ 共线，于是由定理 4.1.1，存在唯一的一对实数 (λ,μ)，使得 $\overrightarrow{OM}=\lambda\boldsymbol{\alpha}$ 以及 $\overrightarrow{ON}=\mu\boldsymbol{\beta}$，于是 $\boldsymbol{\gamma}=\overrightarrow{OM}+\overrightarrow{ON}=\lambda\boldsymbol{\alpha}+\mu\boldsymbol{\beta}$，这便得出如下的共面向量分解定理

定理 4.1.2 (共面向量分解定理). *设向量 $\boldsymbol{\alpha}$ 与向量 $\boldsymbol{\beta}$ 不共线，则对于空间中任一向量 γ，它能与 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$ 共面的充分必要条件是，存在唯一的一对实数 λ 和 μ，使得 $\boldsymbol{\gamma}=\lambda\boldsymbol{\alpha}+\mu\boldsymbol{\beta}$.*

证明. 对于必要性，其中实数对的存在性已经由上面的推导得出，假若还有另一对实数 u 和 v 使得 $\boldsymbol{\gamma}=u\boldsymbol{\alpha}+v\boldsymbol{\beta}$，便有 $(\lambda-u)\boldsymbol{\alpha}+(\mu-v)\boldsymbol{\beta}=\mathbf{0}$，即 $(\lambda-u)\boldsymbol{\alpha}=-(\mu-v)\boldsymbol{\beta}$，若是 $\lambda\neq u$，那么便能得出 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 共线，定理条件矛盾，因此 $\lambda=u$，同时由这等式便有 $\mu=v$，因此实数对是唯一的。

再证充分性，如若 $\boldsymbol{\gamma}=\lambda\boldsymbol{\alpha}+\mu\boldsymbol{\beta}$，则显然向量 $\lambda\boldsymbol{\alpha}$ 与 $\boldsymbol{\alpha}$ 共线，向量 $\mu\boldsymbol{\beta}$ 与 $\boldsymbol{\beta}$ 共线，于是按平行四边形法则，$\lambda\boldsymbol{\alpha}+\mu\boldsymbol{\beta}$ 在 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 所在平面内。 □

与之相仿的，还有空间向量分解定理

定理 4.1.3 (空间向量分解定理). *设 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$、γ 是空间中三个不共面的向量，则对于空间中任一向量 $\boldsymbol{\delta}$，存在唯一的实数三元组 (λ,μ,ν)，使得 $\boldsymbol{\delta}=\lambda\boldsymbol{\alpha}+\mu\boldsymbol{\beta}+\nu\gamma$.*

证明. 仍然通过平移使这些向量的起点都是点 O，记 $\boldsymbol{\alpha}=\overrightarrow{OA}$，$\boldsymbol{\beta}=\overrightarrow{OB}$，$\boldsymbol{\gamma}=\overrightarrow{OC}$，$\boldsymbol{\delta}=\overrightarrow{OD}$，过点 D 作平面 BOC 的平行平面并与直线 OA 相交于 M，过 D 作平面 COA 的平行平面与 OB 交于 N，再过 D 作平面 AOB 的平行平面交 OC 于 P，再以 OM 和 ON 为边作平行四边形 $OMQN$，则 $\overrightarrow{OD}=\overrightarrow{OQ}+\overrightarrow{QD}$，由定理 4.1.2，存在唯一实数对 λ 和 μ，使得 $\overrightarrow{OQ}=\lambda\overrightarrow{OA}+\mu\overrightarrow{OB}$，而由辅助线作法知新作的三个平面与三个向量两两决定的三个平面决定了一个平行六面体，而 $\overrightarrow{QD}$ 与 $\overrightarrow{OP}$ 正好是一组

对棱，所以 $\overrightarrow{QD}=\overrightarrow{OP}$，而由定理 4.1.1，存在唯一实数 ν，使得 $\overrightarrow{OP}=\nu\overrightarrow{OC}$，于是 $\boldsymbol{\delta}=\overrightarrow{OD}=\overrightarrow{OM}+\overrightarrow{ON}+\overrightarrow{OD}=\lambda\boldsymbol{\alpha}+\mu\boldsymbol{\beta}+\nu\boldsymbol{\gamma}$，存在性得证。

若是还存在另一组实数 u、v、w 使得 $\boldsymbol{\delta}=u\boldsymbol{\alpha}+v\boldsymbol{\beta}+w\boldsymbol{\gamma}$，则有 $(\lambda-u)\boldsymbol{\alpha}+(\mu-v)\boldsymbol{\beta}+(\nu-w)\boldsymbol{\gamma}=0$，若是 $\lambda\neq u$，便得出 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$、$\boldsymbol{\gamma}$ 共面，这与定理条件矛盾，所以 $\lambda=u$，进一步类似的方法得出 $\mu=v$ 和 $\nu=w$. □

例 4.1.1 对于一个三角形 ABC 及其所在平面内的任一点 P，显然向量 $\overrightarrow{PA}$、$\overrightarrow{PB}$、$\overrightarrow{PC}$ 不可能全部共线，假定 PB 与 PC 不共线，那么存在唯一分解式 $\overrightarrow{PA}=p\overrightarrow{PB}+q\overrightarrow{PC}$，或者写成

$$\overrightarrow{PA}-p\overrightarrow{PB}-q\overrightarrow{PC}=\mathbf{0}$$

反之，若存在不全为零的三个实数 α、β、γ 使得

$$\alpha\overrightarrow{PA}+\beta\overrightarrow{PB}+\gamma\overrightarrow{PC}=\mathbf{0}$$

因为三个系数不全为零，假定 α 不为零，于是 $\overrightarrow{PA}$ 便可以写成

$$\overrightarrow{PA}=-\frac{\beta}{\alpha}\overrightarrow{PB}-\frac{\gamma}{\alpha}\overrightarrow{PC}$$

因此，$\overrightarrow{PA}$ 与 $\overrightarrow{PB}$ 和 $\overrightarrow{PC}$ 共面，因此得到结论：

定理 4.1.4. *点 P 位于三角形 ABC 所在平面内的充分必要条件是：存在不全为零的三个实数 α、β、γ 使得下式成立，*

$$\alpha\overrightarrow{PA}+\beta\overrightarrow{PB}+\gamma\overrightarrow{PC}=\mathbf{0}$$

并且当点 P 在此平面内时，这个方程中的三个系数在允许相差一个常数因子的意义下是唯一的。

前面已经推证了充分性和必要性，这里只就系数在允许相差一个常数因子的意义下是唯一的加以说明，事实上，将定理等式中所涉及的向量全部改写为以 A 为起点的向量，即

$$\alpha(-\overrightarrow{AP})+\beta(\overrightarrow{AB}-\overrightarrow{AP})+\gamma(\overrightarrow{AC}-\overrightarrow{AP})=\mathbf{0}$$

即

$$\overrightarrow{AP}=\frac{\beta}{\alpha+\beta+\gamma}\overrightarrow{AB}+\frac{\gamma}{\alpha+\beta+\gamma}\overrightarrow{AC}$$

由向量的分解定理，这系数 $\dfrac{\beta}{\alpha+\beta+\gamma}$ 跟 $\dfrac{\gamma}{\alpha+\beta+\gamma}$ 必然是唯一的，不妨设前者为 p，后者为 q，而记 $t=\alpha+\beta+\gamma$，则

$$\beta=pt,\ \gamma=qt,\ \alpha=(1-p-q)t$$

于是 $\alpha:\beta:\gamma=p:q:(1-p-q)$，而 t 可以任取，于是这三个系数在允许相差一个常量因子的情况下是唯一的。

以后我们会利用向量的外积给出这三个系数的具体表达，见例 4.1.6. ■

4.1.3 坐标表示

取定从同一点 O(称为原点) 出发，沿三个不共面方向的单位向量 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$、$\boldsymbol{e}_3$，称为一组基底，简称基，则空间中任一向量 $\boldsymbol{\alpha}$ 都有唯一分解式 $\boldsymbol{\alpha}=x\boldsymbol{e}_1+y\boldsymbol{e}_2+z\boldsymbol{e}_3$，将三元数组 (x,y,z) 称为向量 $\boldsymbol{\alpha}$ 在在这组基下的坐标，记作 $\boldsymbol{\alpha}=(x,y,z)$，于是建立起坐标系统。显然，如果将向量 $\boldsymbol{\alpha}$ 的起点移至原点，则它的终点在此坐标系中的坐标也就正是 (x,y,z)，于是向量便与此坐标系下的点建立起一一映射关系，于是坐标原点就代表零向量，零向量的坐标是 $(0,0,0)$，如果坐标系中的三个单位向量两两垂直，则称为*直角坐标系*，直角坐标系的好处是使得距离与角这类几何量的公式简单化，例如在直角坐标系下，向量 $\boldsymbol{\alpha}$ 的长度就是终点到原点的距离 $|\boldsymbol{\alpha}|=\sqrt{x^2+y^2+z^2}$，今后如无特殊说明，均是在直角坐标系中进行讨论。

显然，若两个向量相等，则它俩坐标相等。

设 $\boldsymbol{\alpha}=(x_1,y_1,z_1)$，$\boldsymbol{\beta}=(x_2,y_2,z_2)$，则

$$\begin{aligned}\boldsymbol{\alpha}\pm\boldsymbol{\beta} &= (x_1\pm x_2)\boldsymbol{e}_1+(y_1\pm y_2)\boldsymbol{e}_2+(z_1\pm z_2)\boldsymbol{e}_3\\ \lambda\boldsymbol{\alpha} &= \lambda x_1\boldsymbol{e}_1+\lambda y_1\boldsymbol{e}_2+\lambda z_1\boldsymbol{e}_3\end{aligned}$$

即是说

$$\begin{aligned}\boldsymbol{\alpha}\pm\boldsymbol{\beta} &= (x_1\pm x_2,y_1\pm y_2,z_1\pm z_2)\\ \lambda\boldsymbol{\alpha} &= (\lambda x_1,\lambda y_1,\lambda z_1)\end{aligned}$$

于是向量 $\boldsymbol{\alpha}=(x,y,z)$ 的相反向量是 $-\boldsymbol{\alpha}=(-x,-y,-z)$.

再设 $\boldsymbol{\alpha}=\overrightarrow{OA}=(x_1,y_1,z_1)$，$\boldsymbol{\beta}=\overrightarrow{OB}=(x_2,y_2,z_2)$，在 $\triangle OAB$ 中应用余弦定理可以求出

$$\cos<\boldsymbol{\alpha},\boldsymbol{\beta}>=\frac{x_1x_2+y_1y_2+z_1z_2}{\sqrt{x_1^2+y_1^2+z_1^2}\sqrt{x_2^2+y_2^2+z_2^2}}$$

这便是两个向量的夹角公式.

以上所有讨论只要去除竖坐标 z，便是平面向量的对应表达，因此不再单独写出来了。

关于共线向量的结果则是

定理 4.1.5. *向量 $\boldsymbol{\alpha}=(x_1,y_1,z_1)$ 与向量 $\boldsymbol{\beta}=(x_2,y_2,z_2)$ 共线的充分必要条件是 $\dfrac{x_1}{x_2}=\dfrac{y_1}{y_2}=\dfrac{z_1}{z_2}$(分母为零时，约定分子也为零，此时也视为等式成立).*

证明. 若 $\boldsymbol{\beta}$ 是零向量，结论显然成立，在 $\boldsymbol{\beta}$ 非零时，根据定理 4.1.1，两个向量共线的充分必要条件是存在唯一实数 λ，使 $\boldsymbol{\alpha}=\lambda\boldsymbol{\beta}$，于是 $x_1=\lambda x_2$,$y_1=\lambda y_2$,$z_2=\lambda z_2$，所以定理中等式成立。反之，若有定理中等式成立，记比值为 λ，则显然就有 $\boldsymbol{\alpha}=\lambda\boldsymbol{\beta}$，因此充分性也成立。 □

而对于向量共面，坐标形式的结论是

定理 4.1.6. 空间中的三个向量 $\boldsymbol{\alpha}=(x_1,y_1,z_1)$，$\boldsymbol{\beta}=(x_2,y_2,z_2)$，$\boldsymbol{\gamma}=(x_3,y_3,z_3)$ 共面的充分必要条件是

$$\begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}=0$$

证明. 假定 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 不共线，则 $\boldsymbol{\gamma}$ 能与前述两个向量共面的充分必要条件是存在唯一一对实数 λ、μ，使得 $\gamma=\lambda\boldsymbol{\alpha}+\mu\boldsymbol{\beta}$，于是关于 u、v、2 的三元一次齐次方程组

$$\begin{cases} ux_1+vx_2+wx_3=0 \\ uy_1+vy_2+wy_3=0 \\ uz_1+vz_2+wz_3=0 \end{cases}$$

有非零解 $(u,v,w)=(\lambda,\mu,-1)$，因而

$$\begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}=0$$

反之，由定理中的行列式为零，知关于 λ、μ、ν 的方程组

$$\begin{cases} \lambda x_1+\mu x_2+\nu x_3=0 \\ \lambda y_1+\mu y_2+\nu y_3=0 \\ \lambda z_1+\mu z_2+\nu z_3=0 \end{cases}$$

有非零解，不妨设 $\nu\neq 0$，于是 $\boldsymbol{\gamma}=-\dfrac{\lambda}{\nu}\boldsymbol{\alpha}-\dfrac{\mu}{\nu}\boldsymbol{\beta}$，于是三个向量共面。 □

由此可见，向量的共线、共面等特性与线性方程组理论有着千丝万缕的联系，这在线性代数中有更一般性的结论。

4.1.4 定比分点

设有一直线 AB，点 P 是直线上任一点，当点 P 不与 A、B 重合时就存在唯一实数 λ 使得 $\overrightarrow{AP}=\lambda\overrightarrow{PB}$，若设 O 是平面上任一点，那么容易推得

$$\overrightarrow{OP}=\frac{1}{1+\lambda}\overrightarrow{OA}+\frac{\lambda}{1+\lambda}\overrightarrow{OB} \tag{4.1.1}$$

当点 P 位于线段 AP 上时 $\lambda>0$，位于 AB 延长线上 B 一侧时 $\lambda<-1$，位于 A 一侧时 $-1<\lambda<0$，上式称为向量的 定比分点公式。

注意到上式中的两个系数和为 1，还有另外一种表示，如果记 $t=\lambda/(1+\lambda)$，那么有

$$\overrightarrow{OP}=(1-t)\overrightarrow{OA}+t\overrightarrow{OB} \tag{4.1.2}$$

式中 t 的意义是 $\overrightarrow{AP}=t\overrightarrow{AB}$，而且这式子对于点 P 与线段 AB 端点重合时也有效，因而它能将直线上的点与实数建立起一一对应。

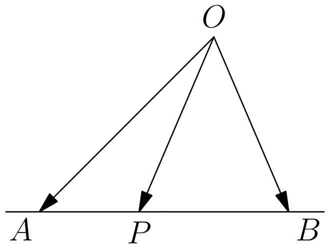

图 4.1

利用复数则可以去掉点 O 的细枝末节，以各点的字母代指该点在复平面上所代表的复数，就有

$$P=(1-t)A+tB \tag{4.1.3}$$

例 4.1.2 在例 4.1.1中已经知道，点 P 位于三角形 ABC 所在平面上的充分必要条件是，存在三个不全为零的实数 α、β、γ 使得下式成立

$$\alpha\overrightarrow{PA}+\beta\overrightarrow{PB}+\gamma\overrightarrow{PC}=\mathbf{0}$$

现在将这式中的向量都用从空间中任取的一点 O 出发的向量来表示，即用 $\overrightarrow{OA}-\overrightarrow{OP}$ 来替换 $\overrightarrow{PA}$，其余类推，则上式便变化为

$$\overrightarrow{OP}=\frac{\alpha\overrightarrow{OA}+\beta\overrightarrow{OB}+\gamma\overrightarrow{OC}}{\alpha+\beta+\gamma}$$

显见右边的三个系数之和为 1，于是例 4.1.1中的结论也可以叙述为，点 P 位于平面 ABC 上的充分必要条件是，存在三个不全为零且满足 $\alpha+\beta+\gamma=1$ 的实数 α、β、γ，使得

$$\overrightarrow{OP}=\alpha\overrightarrow{OA}+\beta\overrightarrow{OB}+\gamma\overrightarrow{OC}$$

或者写成

$$\alpha\overrightarrow{PA}+\beta\overrightarrow{PB}+\gamma\overrightarrow{PC}=\mathbf{0}$$

这两个等式是等价的，但是后一个等式是不需要三个系数之和为 1 的限制的，因为它的两端可以同时乘上任意一个非零实数。

这与定比分点的公式非常类似，在某种意义上，它就是定比分点的一种推广，即由两个点变成了三个点，两个点组成的线段推广成了三个点组成的三角形。

本例中所主要讨论的结果，如果三个系数都为非负，那么点 P 位于三角形的内部或者三条边上，其中位于内部时三个系数都为正而和为 1，位于三条边上时有一个系数为零而另外两个相加为 1，当有两个系数为零另外一个为 1 时正好是顶点之一。

只对三个系数都为正的情形加以说明，在这时有

$$\begin{aligned}\overrightarrow{PA}&=-\left(\frac{\beta}{\alpha}\overrightarrow{PB}+\frac{\gamma}{\alpha}\overrightarrow{PC}\right)\\&=-\frac{\beta+\gamma}{\alpha}\left(\frac{\beta}{\beta+\gamma}\overrightarrow{PB}+\frac{\gamma}{\beta+\gamma}\overrightarrow{PC}\right)\end{aligned}$$

括号中的部分两个系数和为 1，而且都是正数，因而在线段 BC 上存在唯一的一个点 D，使得括号中的部分正是 $\overrightarrow{PD}$，于是有 $\overrightarrow{PA}=-\dfrac{\beta+\gamma}{\alpha}\overrightarrow{PD}$，因此点 P 在线段 AD 上，因而在三角形内。

还可以从另一个可能更加直观的观点来看待这一点，把式中的向量全部改写为以 A 为起点的向量，即 $\alpha(-\overrightarrow{AP})+\beta(\overrightarrow{AB}-\overrightarrow{AP})+\gamma(\overrightarrow{AC}-\overrightarrow{AP})=\mathbf{0}$，即

$$\overrightarrow{AP}=\beta\overrightarrow{AB}+\gamma\overrightarrow{AC}$$

易见右端两个系数都为正，但其和小于 1，于是将它再改写为

$$\overrightarrow{AP}=\frac{\beta}{\beta+\gamma}\left((\beta+\gamma)\overrightarrow{AB}\right)+\frac{\gamma}{\beta+\gamma}\left((\beta+\gamma)\overrightarrow{AC}\right)$$

在边 AB 和 AC 上分别取点 E 和 F，使得 $\overrightarrow{AE}=(\beta+\gamma)\overrightarrow{AB}$ 以及 $\overrightarrow{AF}=(\beta+\gamma)\overrightarrow{AF}$，则有

$$\overrightarrow{AP}=\frac{\beta}{\beta+\gamma}\overrightarrow{AE}+\frac{\gamma}{\beta+\gamma}\overrightarrow{AF}$$

因而点 P 在线段 EF 上，而显然 EF 与 BC 边是平行的，并且除去端点以外都在三角形内。

这结论还可以推广到任意凸多边形的情况，设有凸多边形 $A_1A_2\cdots A_n$，由向量 $\overrightarrow{OP}=\sum_{i=1}^{n}\alpha_i\overrightarrow{OA_i}$ 其中 $\alpha_i\geqslant 0$ 并且 $\sum_{i=1}^{n}\alpha_i=1$ 所决定的点 P 位于凸多边形的内部或者边界上，如果有某个 α_i 为 1 而其余的 α_i 全为零，则它是凸多边形的顶点，如果有两个相邻的 α_i 之和为 1 而其余 α_i 全为零，则它在凸多边形的边上，其余情况下这点位于凸多边形的内部。 ■

例 4.1.3 三角形的重心　作为向量方法的一个展示，我们来证明：三角形的三条中线交于一点，这点称为三角形的重心。

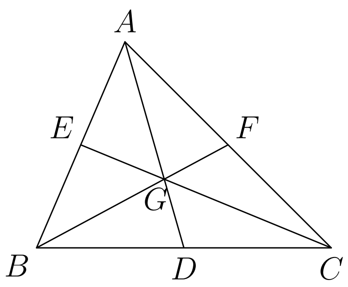

图 4.2

先由 AB 边上的中线 CE 和 AC 边上的中线 BF 相交于点 G，只要证明 BC 边上的中线 AD 通过此点就可以了，这只要证明 $\overrightarrow{AG}$ 与 $\overrightarrow{AD}$ 共线就行了。

因为点 G 同时位于 BF 和 CE 上，所以存在两个实数 t 和 s，使得

$$\overrightarrow{AG}=(1-t)\overrightarrow{AB}+t\overrightarrow{AF}=(1-t)\overrightarrow{AB}+\frac{1}{2}t\overrightarrow{AC}$$

$$\overrightarrow{AG}=s\overrightarrow{AE}+(1-s)\overrightarrow{AC}=\frac{1}{2}s\overrightarrow{AB}+(1-s)\overrightarrow{AC}$$

于是得方程组

$$
\begin{aligned}
1-t &= \frac{1}{2}s \\
1-s &= \frac{1}{2}t
\end{aligned}
$$

解得 $t=s=2/3$，所以 $AG=\frac{1}{3}(\overrightarrow{AB}+\overrightarrow{AC})$，而 $\overrightarrow{AD}=\frac{1}{2}(\overrightarrow{AB}+\overrightarrow{AC})$，所以 $\overrightarrow{AG}=\frac{2}{3}\overrightarrow{AD}$，因而共线，得证，还顺便得到结论：重心把每一条中线都分为 $2:1$ 的两段。

由向量式 $AG=\frac{1}{3}(\overrightarrow{AB}+\overrightarrow{AC})$ 中的向量都按定点 O 来表出，就得 $\overrightarrow{OG}=\frac{1}{3}(\overrightarrow{OA}+\overrightarrow{OB}+\overrightarrow{OC})$，这就是重心的向量表达，由此还得重心的坐标:

$$\left(\frac{x_A+x_B+x_C}{3},\frac{y_A+y_B+y_C}{3}\right)$$

■

例 4.1.4 贝塞尔 (Bezier) 曲线 贝塞尔曲线的初衷是为了将一些离散的点用光滑的曲线连接起来，它的基本原理很简单，如图4.3，Z_0 和 Z_1 是两个点，点 C 是不与 Z_0 和 Z_1 共线的一个指定点（称为控制点），对于一个实数 $\lambda>0$，可以在线段 Z_0C 和 CZ_1 上分别定出点 A 和 B，使得 $Z_0A:AC=\lambda$ 和 $CB:BZ_1=\lambda$，然后在线段 AB 上再定出一个点 Z 使得 $AZ:ZB=\lambda$，当 λ 在正实数范围内变动时，点 Z 的轨迹就是一条曲线，这就是贝塞尔曲线。

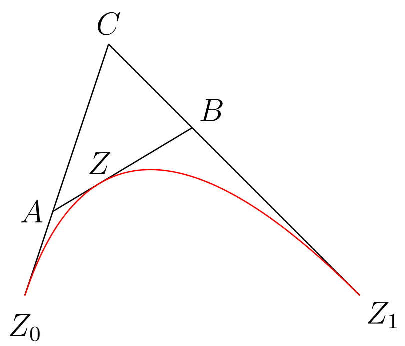

图 4.3 贝塞尔曲线 (一个控制点的情形)

由这原理知贝塞尔曲线是以 λ 为参数的曲线，我们来推导它的参数式，记 $t=\lambda/(1+\lambda)$，并以各点的字母也同时代指该点所对应的复数，就有

$$
\begin{aligned}
Z &= (1-t)A+tB \\
&= (1-t)[(1-t)Z_0+tC]+t[(1-t)C+tB] \\
&= (1-t)^2Z_0+2t(1-t)C+t^2Z_1
\end{aligned}
$$

这是只有一个控制点的情况，如果有两个控制点 C_1 和 C_2，先在线段 Z_0C_1、C_1C_2、C_2Z_1 上各取点 A_1、A_2、A_3 使得 $\frac{Z_0A_1}{A_1C_1}=\frac{C_1A_2}{A_2C_2}=\frac{C_2A_3}{A_3Z_1}=\lambda$，然后再在线段 A_1A_2 和 A_2A_3 上分别取点 B_1 和 B_2 使得 $\frac{A_1B_1}{B_1A_2}=\frac{A_2B_2}{B_2A_3}=\lambda$，最后在线段 B_1B_2 上取点 Z 使得 $\frac{B_1Z}{ZB_2}=\lambda$，点 Z 随着 λ 变动的轨迹就是由控制点 C_1 和 C_2 所确定的贝塞尔曲线，如图4.4所示。

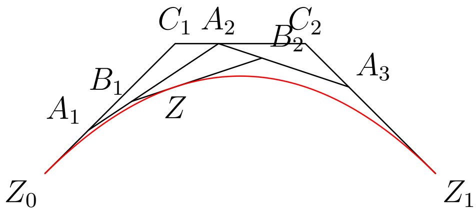

图 4.4 贝塞尔曲线 (两个控制点的情形)

这时的贝塞尔曲线的参数方程是

$$Z = (1-t)^3 Z_0 + 3(1-t)^2 t C_1 + 3t(1-t)^2 C_2 + t^3 Z_1$$

其中 $t = \lambda/(1+\lambda)$，类似的可以得到由端点 Z_0、Z_1 和 $n-1$ 个控制点 $C_k(k = 1, 2, \dots, n-1)$ 所决定的贝塞尔曲线，它的参数方程是

$$Z = \sum_{i=0}^{n} C_n^i (1-t)^{n-i} t^i C_i$$

其中 C_0 和 C_n 分别代指端点 Z_0 和 Z_1，容易发现各个控制点的系数就是 $((1-t)+t)^n$ 按照二项式定理展开后的各项。

现在来证明贝塞尔曲线的一个性质，如图，三角形 ABC 中，以 B、C 为端点以 A 为控制点的贝塞尔曲线段，将全位于由 BC 边上的中位线 EF 所分割出来的梯形 $EBCF$ 中，并且中位线 EF 的中点是这中位线与贝塞尔曲线的唯一公共点。

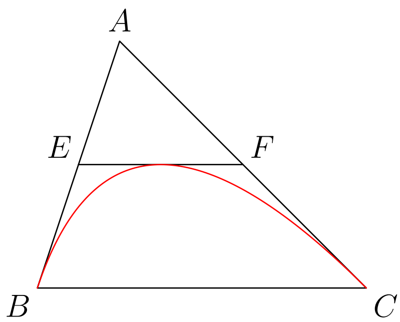

图 4.5

证明很简单，只要将参数方程 $Z = (1-t)^2 B + 2t(1-t)A + t^2 C$ 改写为以 A 为起点的向量形式

$$\overrightarrow{AZ} = (1-t)^2 \overrightarrow{AB} + t^2 \overrightarrow{AC}$$

根据均值不等式知系数和大于等于 1/2，并且两个系数都为非负，所以这曲线全被包含于梯形 $EBCF$ 中，并且当 $t = 1/2$ 时系数和刚好等于 1/2，所以中位线的中点是这曲线与中位线的唯一公共点。

实际上，图4.3中的线段 AB 都是贝塞尔曲线的切线，刚才所证明的性质不过是 $t = 1/2$ 的特殊情况罢了。 ∎

4.1.5 内积

在物理学中，力所做的功是这样一个物理量，当物体在力 F 作用下 (可能还有其它力)，沿着力的方向发生了一段位移 s，则力对物体所做的功是 Fs，这里的位移是沿力的方向所产生的位移，若是位移与力的方向有一定夹角，则将位移投影到力的方向上来，称为在力的方向上的分位移。由此抽象出两个向量的内积概念。

先来讨论下一个向量在另一个向量方向上的投影，设 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 是两个非零向量，将它们移动到共同起点 O，并设终点分别是 A 和 B，则过 A 向直线 OB 引垂线并设垂直是 A_1，则向量 $\overrightarrow{OA_1}$ 称为向量 $\boldsymbol{a}$ 在向量 $\boldsymbol{b}$ 的方向上的投影向量，设向量 $\boldsymbol{a}$ 与向量 $\boldsymbol{b}$ 的夹角是 θ，那么显然 $\overrightarrow{OA_1} = \dfrac{|\boldsymbol{a}|\cos\theta}{|\boldsymbol{b}|}\boldsymbol{b}$.

由此引入两个向量的内积概念.

定义 4.1.9. *两个向量的* 内积 *是两个向量 $\boldsymbol{a}$ 与向量 $\boldsymbol{b}$ 的一种二元运算，记为 $\boldsymbol{a}\cdot\boldsymbol{b}$，或者简记为 $\boldsymbol{ab}$，其结果是一个实数，其值规定为：如果 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 共线，则其绝对值为两个向量之长度乘积，其符号则由两个向量的方向来决定，方向相同时为正，方向相反时为负，若两者中有零向量，则内积为零；若 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 不共线，则两者之内积为其中之一在另一向量上的投影向量与另一向量之内积。*

对于向量与自己的内积，借用乘方的符号写成 $\boldsymbol{a}^2$，显然 $\boldsymbol{a}^2 = |\boldsymbol{a}|^2$，但是需要注意 $\boldsymbol{a}^3$ 有所不同，它表示 $(\boldsymbol{a}\cdot\boldsymbol{a})\boldsymbol{a}$，括号中的内积是一个实数，因而再将其与向量 $\boldsymbol{a}$ 作数与向量的乘法，其结果是一个与 $\boldsymbol{a}$ 同向的向量。

需要说明几点：1. 两个向量的内积不再是一个向量，而是一个实数，这与外积不同（稍后会讲到，两个向量的外积仍是一个向量）. 2. 在两个向量不共线时，使用哪一个向量的投影无关紧要，换句话说，向量的内积满足交换律，即

定理 4.1.7. *两个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的内积是实数 $|\boldsymbol{a}|\cdot|\boldsymbol{b}|\cdot\cos\theta$，其中 θ 是两个向量的夹角，即 $\boldsymbol{a}\cdot\boldsymbol{b} = |\boldsymbol{a}|\cdot|\boldsymbol{b}|\cdot\cos\theta$，从而两个向量互相垂直的充分必要条件是它们的内积为零。*

证明. 如果两个向量共线，则等式显然成立，在不共线的情形下，向量 $\boldsymbol{a}$ 在向量 $\boldsymbol{b}$ 上的投影是 $\dfrac{|\boldsymbol{a}|\cos\theta}{|\boldsymbol{b}|}\boldsymbol{b}$，于是便得等式成立，由这等式便得向量垂直的充分必要条件。 □

由此定理即知向量内积满足交换律，即

$$\boldsymbol{a}\cdot\boldsymbol{b} = \boldsymbol{b}\cdot\boldsymbol{a}$$

显然还有

$$(\lambda\boldsymbol{a})\cdot\boldsymbol{b} = \lambda(\boldsymbol{a}\cdot\boldsymbol{b})$$

从图 4.6还可以验证，向量内积还满足对加法的分配律，即

$$(\boldsymbol{a}+\boldsymbol{b})\cdot\boldsymbol{c} = \boldsymbol{a}\cdot\boldsymbol{c} + \boldsymbol{b}\cdot\boldsymbol{c}$$

这是因为，向量 $\overrightarrow{OA}$ 与向量 $\boldsymbol{OB}$ 在向量 $\overrightarrow{OC}$ 上的两个投影向量之和，正好便是 $\overrightarrow{OD}$ 在向量 $\overrightarrow{OC}$ 上的投影向量。

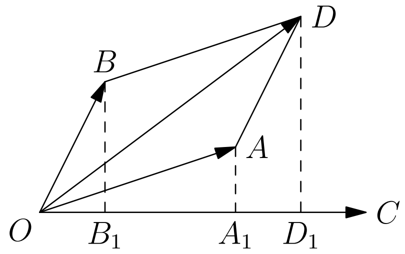

图 4.6

接下来讨论向量内积的坐标表示，设在空间直角坐标系中，两个向量的坐标为 $\boldsymbol{a}=(x_1,y_1,z_1)$ 和 $\boldsymbol{b}=(x_2,y_2,z_2)$，设沿 x 轴、y 轴、z 轴正方向的单位向量分别是 $\boldsymbol{i}$、$\boldsymbol{j}$、$\boldsymbol{k}$，那么向量的坐标表达等价于

$$\begin{aligned}\boldsymbol{a} &= x_1\boldsymbol{i}+y_1\boldsymbol{j}+z_1\boldsymbol{k}\\ \boldsymbol{b} &= x_2\boldsymbol{i}+y_2\boldsymbol{j}+z_2\boldsymbol{k}\end{aligned}$$

利用向量的交换律和对加法的结合律，以及

$$\boldsymbol{i}\cdot\boldsymbol{j}=\boldsymbol{j}\cdot\boldsymbol{k}=\boldsymbol{k}\cdot\boldsymbol{i}=0,\boldsymbol{i}^2=\boldsymbol{j}^2=\boldsymbol{k}^2=1$$

便得

$$\begin{aligned}\boldsymbol{a}\cdot\boldsymbol{b} &= (x_1\boldsymbol{i}+y_1\boldsymbol{j}+z_1\boldsymbol{k})\cdot(x_2\boldsymbol{i}+y_2\boldsymbol{j}+z_2\boldsymbol{k})\\ &= x_1x_2\boldsymbol{i}^2+y_1y_2\boldsymbol{j}^2+z_1z_2\boldsymbol{k}^2+(x_1y_2+x_2y_1)(\boldsymbol{i}\cdot\boldsymbol{j})+(y_1z_2+z_1y_2)(\boldsymbol{j}\cdot\boldsymbol{k})+(z_1x_2+x_1z_2)(\boldsymbol{k}\cdot\boldsymbol{i})\\ &= x_1x_2+y_1y_2+z_1z_2\end{aligned}$$

这就是向量内积的坐标表达式，对于平面向量，也有完全类似的结论，只是没有 z 坐标，由此可知

定理 4.1.8. 空间向量 $\boldsymbol{a}=(x_1,y_1,z_1)$ 与 $\boldsymbol{b}=(x_2,y_2,z_2)$ 互相垂直的充分必要条件是 $x_1x_2+y_1y_2+z_1z_2=0$.

根据向量内积的表达式 $\boldsymbol{a}\cdot\boldsymbol{b}=|\boldsymbol{a}||\boldsymbol{b}|\cos\theta$，可知有不等式

$$(\boldsymbol{a}\cdot\boldsymbol{b})^2\leqslant|\boldsymbol{a}|^2|\boldsymbol{b}|^2$$

设平面向量 $\boldsymbol{a}=(x_1,y_1)$，$\boldsymbol{b}=(x_2,y_2)$，则上述不等式即为

$$(x_1x_2+y_1y_2)^2\leqslant(x_1^2+y_1^2)(x_2^2+y_2^2)$$

同理，如果是空间向量，则相应的得出

$$(x_1x_2+y_1y_2+z_1z_2)^2\leqslant(x_1^2+y_1^2+z_1^2)(x_2^2+y_2^2+z_2^2)$$

这两个不等式分别就是柯西不等式的二元情形和三元情形，柯西不等式的详细介绍见小节 3.8.4.

例 4.1.5 平行四边形的一个性质、三角形中线长公式 根据向量内积的交换律和分配律，可知有如下恒等式

$$(\boldsymbol{a}+\boldsymbol{b})^2+(\boldsymbol{a}-\boldsymbol{b})^2=2(\boldsymbol{a}^2+\boldsymbol{b}^2)$$

而向量加减法的平行四边形法则表明，$\boldsymbol{a}+\boldsymbol{b}$ 与 $\boldsymbol{a}-\boldsymbol{b}$ 分别是平行四边形的两条对角线所代表的向量，于是由此恒等式得到结论：平行四边形两条对角线的平方和，等于四条边的平方和。

进一步，在 $\triangle ABC$ 中，设 BC 边上的中线是 AD，那么 $2\overrightarrow{AD}=\overrightarrow{AB}+\overrightarrow{AC}$，$\overrightarrow{CB}=\overrightarrow{AB}-\overrightarrow{AC}$，于是按上式，便有

$$4AD^2+BC^2=2(AB^2+AC^2)$$

即

$$AD^2=\frac{1}{4}(2AB^2+2AC^2-BC^2)$$

或者简写为

$$l_a^2=\frac{1}{4}(2b^2+2c^2-a^2)$$

这便是三角形中线长公式。 ■

4.1.6 外积

从物理学上的力矩等物理量的概念，可以抽象出两个向量的另一种形式的乘积，即外积，也叫矢量积，叉积，与内积不同，两个向量的外积仍然是一个向量。

定义 4.1.10. 两个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的外积仍然是一个向量，用 $\boldsymbol{a}\times\boldsymbol{b}$ 表示，它的大小是 $|\boldsymbol{a}||\boldsymbol{b}|\sin\theta$，其中 θ 是两个向量的夹角，而它的方向是这样定义的，如果 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 不共线，则 $\boldsymbol{a}\times\boldsymbol{b}$ 同时垂直于 $\boldsymbol{a}$ 和 $\boldsymbol{b}$，并且按右手定则，握住右手，并让四个小手指沿从 $\boldsymbol{a}$ 按两个向量的夹角旋转到 $\boldsymbol{b}$ 的方向，此时大姆指所指的方向便是 $\boldsymbol{a}\times\boldsymbol{b}$ 的方向。当 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 共线时，由于 $\boldsymbol{a}\times\boldsymbol{b}$ 的大小为零，所以此时的外积为零向量。

由定义，零向量与任何向量之外积皆为零向量，且若两个向量共线，则它们的外积亦是零向量。而且，对于三角形 ABC，有 $S_{\triangle ABC}=\frac{1}{2}|\overrightarrow{AB}|\times|\overrightarrow{AC}|$，这是向量外积的一个几何意义（仅限于其大小，在许多情况下，通常用它的方向作为有向面积的符号依据）。

向量的外积与原来两个向量同时垂直，这意味着它实际上成为原来两个向量所共平面的法向量。

由定义，可知外积不满足交换律，交换两个向量的位置，所得外积与原来的外积是相反向量，亦即

$$\boldsymbol{b}\times\boldsymbol{a}=-\boldsymbol{a}\times\boldsymbol{a}$$

显然也有

$$(\lambda\boldsymbol{a})\times(\mu\boldsymbol{b})=\lambda\mu(\boldsymbol{a}\times\boldsymbol{b})$$

仍然从图 4.6可以验证向量外积的 (左) 分配律

$$(\boldsymbol{a}+\boldsymbol{b})\times\boldsymbol{c}=\boldsymbol{a}\times\boldsymbol{c}+\boldsymbol{b}\times\boldsymbol{c}$$

这是因为，对于图中的情形，显然有点 D 到直线 OC 的距离等于 A、B 两点到直线 OC 的距离之和，于是 $S_{\triangle DOC}=S_{\triangle AOC}+S_{\triangle BOC}$，于是便得上式，这里虽然都是按正的距

离和面积来考虑的，实际上如果把距离和面积都看成带符号的代数距离和代数面积，便得出完整的验证。

有了左分配律，便可得出右分配律

$$\begin{aligned}\boldsymbol{a}\times(\boldsymbol{b}+\boldsymbol{c}) &= -(\boldsymbol{b}+\boldsymbol{c})\times\boldsymbol{a}\\ &= -(\boldsymbol{b}\times\boldsymbol{a}+\boldsymbol{c}\times\boldsymbol{a})\\ &= (-\boldsymbol{b}\times\boldsymbol{a})+(-\boldsymbol{c}\times\boldsymbol{a})\\ &= \boldsymbol{a}\times\boldsymbol{b}+\boldsymbol{a}\times\boldsymbol{c}\end{aligned}$$

因为两个向量的外积仍然是一个向量，因此所得结果还可以继续参与外积运算，即式子 $(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}$ 有意义，我们来讨论一下这个式子，假定三个向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 不共面 (从而其中任何两者也就不共线)，于是 $\boldsymbol{a}\times\boldsymbol{b}$ 同时垂直于 $\boldsymbol{a}$ 和 $\boldsymbol{b}$，即它成为 $\boldsymbol{a}$、$\boldsymbol{b}$ 两个向量所共面的法向量。而 $(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}$ 显然又垂直于 $\boldsymbol{a}\times\boldsymbol{b}$，于是 $(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}$ 又与 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 共面！这引起了我们的兴趣，于是多做一些讨论。

既然 $(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}$ 与 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 共面，按向量分解定理，存在唯一一对实数 λ 和 μ，使得

$$(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}=\lambda\boldsymbol{a}+\mu\boldsymbol{b}$$

成立，我们来尝试具体求出 λ 和 μ。

因为 $(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}$ 与 $\boldsymbol{c}$ 垂直，因此它与 $\boldsymbol{c}$ 的内积应为零，即

$$\lambda(\boldsymbol{a}\cdot\boldsymbol{c})+\mu(\boldsymbol{b}\cdot\boldsymbol{c})=0$$

定理 4.1.9. 对于三个向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$，则向量 $(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}$ 与向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 共面，并且

$$(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}=-(\boldsymbol{b}\cdot\boldsymbol{c})\boldsymbol{a}+(\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b}$$

由此定理可知向量外积不满足结合律，因为 $(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}$ 与向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 共面，而向量 $\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c})$ 与 $\boldsymbol{b}$、$\boldsymbol{c}$ 共面，一般情况下两者并不相等。

例 4.1.6 在例 4.1.1中，已经知道对于三角形 ABC 平面内的任一点 P，有向量方程

$$\alpha\overrightarrow{PA}+\beta\overrightarrow{PB}+\gamma\overrightarrow{PC}=\mathbf{0} \tag{4.1.4}$$

在这个例子中，我们将利用向量的外积来具体的找出系数 α、β、γ 的具体表达式。

首先将上面的向量方程的两端分别与 $\overrightarrow{PA}$、$\overrightarrow{PB}$、$\overrightarrow{PC}$ 作外积，得到向量方程组

$$\begin{aligned}\beta(\overrightarrow{PB}\times\overrightarrow{PA})+\gamma(\overrightarrow{PC}\times\overrightarrow{PA}) &= \mathbf{0}\\ \alpha(\overrightarrow{PA}\times\overrightarrow{PB})+\gamma(\overrightarrow{PC}\times\overrightarrow{PB}) &= \mathbf{0}\\ \alpha(\overrightarrow{PA}\times\overrightarrow{PC})+\beta(\overrightarrow{PB}\times\overrightarrow{PC}) &= \mathbf{0}\end{aligned}$$

显然 $\overrightarrow{PA}\times\overrightarrow{PB}$、$\overrightarrow{PB}\times\overrightarrow{PC}$、$\overrightarrow{PC}\times\overrightarrow{PA}$ 这三个向量是共线的，因为都在过点 P 且与平面 ABC 垂直的直线上，这意味着上述方程有可能转化为关于实数的方程组，任取平面 ABC 的单位法向量为 $\boldsymbol{n}$，记

$$\overrightarrow{PB}\times\overrightarrow{PC}=u\boldsymbol{n},\ \overrightarrow{PC}\times\overrightarrow{PA}=v\boldsymbol{n},\ \overrightarrow{PA}\times\overrightarrow{PB}=w\boldsymbol{n}$$

显然 $|u| = |\overrightarrow{PB} \times \overrightarrow{PC}|$，$|v| = |\overrightarrow{PC} \times \overrightarrow{PA}|$，$|w| = |\overrightarrow{PA} \times \overrightarrow{PB}|$，而 u、v、w 的符号则视这三个外积向量的方向而定，与 $\boldsymbol{n}$ 同向的为正，反向的为负，于是上述方程组转化为下面关于 u、v、w 的方程组。

$$\begin{cases} -\beta w + \gamma v &= 0 \\ \alpha w - \gamma u &= 0 \\ -\alpha v + \beta u &= 0 \end{cases}$$

把它视为关于 α、β、γ 的三元一次齐次方程组，容易发现由前两个方程消去 γ 即得第三个方程，因此它等价于如下的方程组

$$\begin{cases} -\beta w + \gamma v &= 0 \\ \alpha w - \gamma u &= 0 \end{cases}$$

未知数的个数多于方程的个数，因此有自由未知量，解之得 (γ 是自由未知量)

$$\begin{cases} \alpha &= \frac{u}{w}\gamma \\ \beta &= \frac{v}{w}\gamma \end{cases}$$

这便是原方程组的全部解，它有无穷多组解 (任意取定 γ 便得出一组解)，但这所有的解都满足 $\alpha : \beta : \gamma = u : v : w$，而且在式 4.1.4两端同乘上一个因子并无多大意义，于是便得

$$u\overrightarrow{PA} + v\overrightarrow{PB} + w\overrightarrow{PC} = \mathbf{0}$$

其中 $\overrightarrow{PB} \times \overrightarrow{PC} = u\boldsymbol{n}$, $\overrightarrow{PC} \times \overrightarrow{PA} = v\boldsymbol{n}$, $\overrightarrow{PA} \times \overrightarrow{PB} = w\boldsymbol{n}$，这里 $\boldsymbol{n}$ 是平面 ABC 的单位法向量。这就是最终的结果。

注意到外积 $\overrightarrow{PA} \times \overrightarrow{PB}$ 与三角形 PAB 面积的关系，这结果还可以写成

$$S_{\triangle PBC}\overrightarrow{PA} + S_{\triangle PCA}\overrightarrow{PB} + S_{\triangle PAB}\overrightarrow{PC} = \mathbf{0}$$

但式中的 S 均是代数面积，即其绝对值是对应三角形的面积，但其符号由外积向量的方向决定，特别情况是，当点 P 位于三角形 ABC 内时，这三个外积向量同向，于是这里的三个面积就是同号的，从而可以都取为正值。 ■

4.2 曲线、曲面与方程

在笛卡尔坐标几何创立之前，几何学与代数学是两个研究对象和研究方法都各自独立的学科，彼此之间并无多少实质性的联系，几何学历来以难著称，层出不穷的几何证明技巧，惊为神来之笔的各种几何辅助线给人以深刻的印象，但是一直以来，缺乏一种通用的或者说万能的方法来处理几何问题。

古希腊数学家阿波罗尼奥斯的名著《圆锥曲线论》将古典几何推向了一个巅峰，该书利用几何方法将圆锥曲线的性质几乎一网打尽，以致于后人在长达两千余年间没能在这个领域有多少重大发现。但笛卡尔的坐标几何改变了这一点。在笛卡尔的坐标思想中，利用坐标来刻画点的位置，于是位置、距离、角度等几何量统统被坐标量化，于是

产生了一种新的研究几何学的方法，就是利用纯粹的代数运算来证明几何性质，这就是解析几何，利用这种新的方法，人们又发现了圆锥曲线的一些新的性质，对圆锥曲线的研究才又有了新的突破性发展。

4.2.1 曲线与曲面的方程、方程的曲线或曲面

今后所称曲线，是通指线型的几何图形，包括直线，只是按一般情形进行通称，并不意味曲线是弯曲的。

定义 4.2.1. *在建立了直角坐标系的平面中，对于曲线 C 和方程 $f(x,y)=0$，如果*

1. *曲线上任一点 $P(x_P,y_P)$，其坐标都满足方程 $f(x,y)=0$，即 $f(x_P,y_P)=0$.*

2. *任一个坐标满足方程 $f(x,y)=0$ 的点 $P(x_p,y_p)$ 都在曲线 C 上.*

则称方程 $f(x,y)=0$ 是曲线 C 的方程，而曲线 C 是方程 $f(x,y)=0$ 的曲线。

类似的有空间曲面的方程，与方程的曲面的概念，这里的曲面同样也包括平面.

定义 4.2.2. *在空间直角坐标系中，对于曲面 C 和方程 $f(x,y,z)=0$，如果*

1. *曲面上任一点 $P(x_P,y_P,z_P)$，其坐标都满足方程 $f(x,y,z)=0$，即 $f(x_P,y_P,z_p)=0$.*

2. *任一个坐标满足方程 $f(x,y,z)=0$ 的点 $P(x_p,y_p,z_p)$ 都在曲面 C 上.*

则称方程 $f(x,y,z)=0$ 是曲面 C 的方程，而曲面 C 是方程 $f(x,y,z)=0$ 的曲面。

方程 $f(x,y)=0$ 或者 $f(x,y,z)=0$ 称为曲线或曲面的 *普通方程*，实际上很多曲线难以用普通方程表达出来，而更常用的是 *参数方程*.

定义 4.2.3. *如果曲线 C 上任一点的坐标，都是某个数 t 的函数，即*

$$\begin{cases} x &= x(t) \\ y &= y(t) \\ z &= z(t) \end{cases}$$

则称该方程 (组) 是曲线 C 的参数方程，类似的，如果曲面上任一点的坐标，都是两个参数 t 和 s 的二元函数，即

$$\begin{cases} x &= x(t,s) \\ y &= y(t,s) \\ z &= z(t,s) \end{cases}$$

则称该方程 (组) 是曲面的 参数方程.

4.2.2 直线和平面的方程

先来考虑平面上直线的方程，为此先给出直线的方向向量和法向量的定义.

定义 4.2.4. 能够与直线上两点所确定的向量共线的非零向量，称为该直线的 方向向量，而能与直线的方向向量垂直的向量，称为该直线的 法向量。

定义 4.2.5. 对于空间中的一张平面，凡能与平面内两点所确定的向量共线的非零向量称为该平面的方向向量，即由方向向量决定的直线族与平面只能是平行或包含的关系。

定义 4.2.6. 空间中，如果直线与平面垂直，则直线称为平面的法线，而平面称为直线的法平面，与平面的法线共线的向量称为平面的法向量。

方向向量和法向量均唯一的确定了直线的走向。

为了确定直线，除了直线的走向，还需要确定直线的位置，于是再给出直线上的一个点就行了，所以先来考虑，给定直线 l 的方向向量 $\boldsymbol{v}=(a,b)$ 及直线上一点 $P(x_0,y_0)$ 的情形下，直线的方程。

设直线上任一点的坐标是 $Q(x,y)$，于是有 $\overrightarrow{PQ}$ 与方向向量 $\boldsymbol{v}$ 共线，于是按照定理 4.1.5在平面上的结论，就有

$$\frac{x-x_0}{a}=\frac{y-y_0}{b}$$

显然，凡满足此方程的任一对坐标所对应的点，其与点 P 所构成的向量也与方向向量 $\boldsymbol{v}$ 共线，从而也必然在直线 l 上，因此这个方程就是所要求的，它称为直线的 *点向式方程*，在分母为零时约定分子也为零。

同理，对于空间中由方向向量 $\boldsymbol{v}=(a,b,c)$ 及直线上一点 $P(x_0,y_0,z_0)$ 所确定的直线的方程是

$$\frac{x-x_0}{a}=\frac{y-y_0}{b}=\frac{z-z_0}{c}$$

显然这并不是一个方程，而是一个方程组，这就是说，空间中的直线，需要两个三元一次方程来确定。

因为直线只需要两个点就可以唯一确定，所以考虑在平面上由两个点 $A(x_1,y_1)$ 和 $B(x_2,y_2)$ 所确定的直线的方程，因为这时方向向量是 $\overrightarrow{AB}=(x_1-x_2,y_1-y_2)$，所以应用刚才得出的点向式方程 (同时还过点 A)，得出直线方程是

$$\frac{x-x_1}{x_1-x_2}=\frac{y-y_1}{y_1-y_2}$$

这便是直线的*两点式方程*，一个特殊情况是，经过 $A(a,0)$ 和 $B(0,b)$ 两点的直线方程是

$$\frac{x}{a}+\frac{y}{b}=1$$

这里的 A 和 B 两点分别是直线与两个坐标轴的交点，而数 a 和 b 分别称为直线在两个坐标轴上的*截距* (即截点到原点的代数距离)，所以这方程称为直线的*截距式方程*.

同理可得空间中，经过两点 $A(x_1,y_1,z_1)$ 和 $B(x_2,y_2,z_2)$ 的直线方程是

$$\frac{x-x_1}{x_1-x_2}=\frac{y-y_1}{y_1-y_2}=\frac{z-z_1}{z_1-z_2}$$

接着再来考虑由直线上一点 $P(x_0, y_0)$ 及它的法向量 $\boldsymbol{n} = (a, b)$ 所确定的直线的方程，设直线上任一点 $Q(x, y)$，则有 $\boldsymbol{n} \perp \overrightarrow{PQ}$，所以 $\boldsymbol{n} \cdot \overrightarrow{PQ} = 0$，即

$$a(x - x_0) + b(y - y_0) = 0$$

这就是所要求的，它称为直线的 *点法式方程*.

对于空间中的直线，显然一条法向量并不足以确定它的走向，需要两个不共线的法向量 (均与法平面垂直，法平面是指与直线垂直的平面)，设两个法向量分别是 $\boldsymbol{n_1} = (a_1, b_1, c_1)$ 和 $\boldsymbol{n_2} = (a_2, b_2, c_2)$，同样求内积为零可得出

$$\begin{cases} a_1(x - x_0) + b_1(y - y_0) + c_1(z - z_0) & = & 0 \\ a_2(x - x_0) + b_2(y - y_0) + c_2(z - z_0) & = & 0 \end{cases}$$

这再一次验证了空间直线需要两个三元一次方程才能确定，实际上接下来会看到，一个三元一次方程代表空间中的一张平面，而直线则是两个平面的交线。

还是回头考虑由直线上一点 $P(x_0, y_0)$ 和直线的方向向量 $\boldsymbol{v} = (a, b)$ 所确定的直线，对于直线上任一点 $Q(x, y)$，由 $\boldsymbol{v}$ 与 $\overrightarrow{PQ}$ 共线，存在唯一实数 t，使得 $\overrightarrow{PQ} = t\boldsymbol{v}$，展开坐标就是

$$\begin{cases} x & = & x_0 + at \\ y & = & y_0 + bt \end{cases}$$

可见，直线上任一点均由参量 t 的不同取值决定，t 的取值与直线上的点构成一一对应的关系，所以这个方程组就是直线的 *参数方程*.

同样，对于空间直线上任一点 $P(x_0, y_0, z_0)$ 及其方向向量 $\boldsymbol{v} = (a, b, c)$ 所确定的直线的参数方程是

$$\begin{cases} x & = & x_0 + at \\ y & = & y_0 + bt \\ z & = & z_0 + ct \end{cases}$$

接着讨论空间平面的方程，类似于平面上的直线，平面由其法向量 $\boldsymbol{n} = (a, b, c)$ 及平面上一点 $P(x_0, y_0, z_0)$ 所唯一确定，设 $Q(x, y)$ 是该平面上任一点，则 $\boldsymbol{n} \perp \overrightarrow{PQ}$，于是 $\boldsymbol{n} \cdot \overrightarrow{PQ} = 0$，所以

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

这称为平面的 *点法式方程*。

如果要用方向向量而不是法向量，则平面需要两个不共线的方向向量才能唯一确定平面的倾斜方向，设两个方向向量是 $\boldsymbol{v_1} = (a_1, b_1, c_1)$ 和 $\boldsymbol{v_2} = (a_2, b_2, c_2)$，则对于平面上任一点 $Q(x, y, z)$，有 $\overrightarrow{PQ}$ 与 $\boldsymbol{v_1}$ 及 $\boldsymbol{v_2}$ 共面，由定理 4.1.6可得

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix} = 0$$

或者展开成为

$$\begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}(x-x_0)-\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}(y-y_0)+\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}(z-z_0)=0$$

这两个方程就是平面的 *点向式方程*.

而由 $\overrightarrow{PQ}$ 与 $\boldsymbol{v_1}$ 及 $\boldsymbol{v_2}$ 共面，存在唯一一对实数 u 和 v，使得 $\overrightarrow{PQ}=u\boldsymbol{v_1}+v\boldsymbol{v_2}$，写成坐标形式就是

$$\begin{cases} x &= x_0+ua_1+va_2 \\ y &= y_0+ub_1+vb_2 \\ z &= z_0+uc_1+vc_2 \end{cases}$$

这就是平面的 *参数方程*.

在上面看到，平面上的直线方程都是二元一次方程，空间中的平面方程也都是三元一次方程，那反过来，是否每一个二元一次方程和每一个三元一次方程都分别代表二维平面上的一条直线和空间中的一张平面呢？

二元一次方程的一般形式是 $Ax+By+C=0$，其中 A、B 不同时为零，假定 $A\neq 0$，则点 $P(-\dfrac{C}{A},0)$ 在这方程所表示的图形上，于是方程可改写为

$$A(x+\frac{C}{A})+B(y-0)=0$$

显然这表示通过点 P 并以 $\boldsymbol{n}=(A,B)$ 为法向量的直线，所以它必然表示一条直线，方程 $Ax+By+C=0$ 便称为平面上直线的 *一般方程*。

同样，在空间中，三元一次方程 $Ax+By+Cz+D=0$ 中，A、B、C 不同时为零，假定 $A\neq 0$，于是它便是通过点 $P(-\dfrac{D}{A},0,0)$ 并以 $\boldsymbol{n}=(A,B,C)$ 为法向量的平面，它便称为空间平面的 *一般方程*.

4.2.3 圆和球面的方程

在平面直角坐标系中，设某圆的圆心坐标坐标是 $C(a,b)$，半径为 r，则对于圆上任一点 $P(x,y)$，有 $|PC|=r$，即 $\sqrt{(x-a)^2+(y-b)^2}=r$，这等价于

$$(x-a)^2+(y-b)^2=r^2$$

反过来，如果某个点 Q 的坐标满足上述方程，则两边开方就得出 $|QC|=r$，即点 Q 必在此圆上，所以这个方程就是以点 (a,b) 为圆心，r 为半径的圆的方程，它称为圆的 *标准方程*.

将圆的标准方程展开，得到一个关于 x 和 y 的二元二次方程

$$x^2+y^2-2ax-2ay+a^2+b^2-r^2=0$$

这方程具有几个特点：(I)x^2 与 y^2 项系数相同，(II) 不含 xy 项. 对于满足这两个条件的一般的二元二次方程 (将 x^2 和 y^2 系数化为 1)

$$x^2+y^2+Dx+Ey+F=0$$

它可以配方成为

$$\left(x+\frac{D}{2}\right)^2+\left(y+\frac{E}{2}\right)^2=\frac{1}{4}(D^2+E^2-4F)$$

显然在 $D^2+E^2-4F>0$ 时它表示以 $\left(-\frac{D}{2},-\frac{E}{2}\right)$ 为圆心，以 $\sqrt{D^2+E^2-4F}$ 为半径的圆，于是一般的二元二次方程需要满足三个条件才能成为圆的方程，上面特殊形式的二元二次方程称为圆的 *一般方程*。

类似的可以得到，空间中以点 $C(a,b,c)$ 为球心，r 为半径的球面方程是

$$(x-a)^2+(y-b)^2+(z-c)^2=r^2$$

对于以 $C(a,b)$ 为圆心，r 为半径的圆，由圆心沿 x 轴正方向所确定的点是 $A(a+r,b)$，对圆心上其它任意一点 $Q(x,y)$，可以视为由点 A 绕圆心旋转一定的角度 θ 而得到，于是按三角函数的定义有 $\cos\theta=\dfrac{x-a}{r},\sin\theta=\dfrac{y-b}{r}$，于是得到

$$\begin{cases} x &=& a+r\cos\theta \\ y &=& b+r\sin\theta \end{cases}$$

这就是圆的 *参数方程*.

再来考虑球面的参数方程，设球心是 $C(a,b,c)$，半径为 r，过球心作 xOy 平面的平行平面 α，点 $A(a+r,b,c)$ 是该平面与球面交线上的一个点，设 $Q(x,y,z)$ 是球面上任一点，设射线 OQ 与平面 α 的夹角是 θ，过点 Q 再作平面 xOy 的平行平面 β，与球面交成一个纬线圈，则点 Q 可由点 A 经过两次旋转而得到，第一次旋转由点 A 出发沿着经线圈旋转一个角度到达点 Q 所在的纬线圈上的点 M，显然旋转角度就是 θ，再由 M 出发沿着该纬线圈旋转一个角度 φ 到达点 Q，于是点 Q 所决定的纬线圈的圆心坐标是 $K(a,b,c+r\sin\theta)$，半径是 $r\cos\theta$，于是由圆的参数方程有

$$\begin{cases} x_Q &=& x_k+r\cos\theta\cos\varphi \\ y_Q &=& y_k+r\cos\theta\sin\varphi \\ z_Q &=& z_k \end{cases}$$

于是得出球面的参数方程是

$$\begin{cases} x_Q &=& a+r\cos\theta\cos\varphi \\ y_Q &=& b+r\cos\theta\sin\varphi \\ z_Q &=& c+r\sin\theta \end{cases}$$

4.3 几个重要的定理

本节讲述几个在平面几何中占有重要地位的定理。

4.3.1 张角定理

定理 4.3.1 (张角定理). *在三角形 ABC 中，D 是 BC 边上任意一点，有下式成立:*

$$\frac{\sin\angle BAC}{AD}=\frac{\sin\angle CAD}{AB}+\frac{\sin\angle BAD}{AC} \tag{4.3.1}$$

张角定理的逆也是成立的，即如果平面上四个点满足定理中的等式，就有 B、C、D 三点共线，这也可以用来处理共线问题。

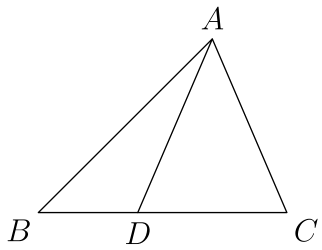

图 4.7 张角定理

证明一. 因为 $S_{\triangle ABC} = S_{\triangle ABD} + S_{\triangle ACD}$，所以

$$\frac{1}{2}AB \cdot AC \cdot \sin\angle BAC = \frac{1}{2}AB \cdot AD \cdot \sin\angle BAD + \frac{1}{2}AC \cdot AD \cdot \sin\angle CAD$$

两边同除以 $\frac{1}{2}AB \cdot AC \cdot AD$ 即得证。 □

证明二. 由正弦定理有 $AB\sin\angle BAC = BC\sin\angle ACB$ 以及 $AD\sin\angle CAD = CD\sin\angle ACD$，两式相除得

$$\frac{\sin\angle CAD}{\sin\angle BAC} \cdot \frac{AD}{AB} = \frac{CD}{BC}$$

同理也有

$$\frac{\sin\angle BAD}{\sin\angle BAC} \cdot \frac{AD}{AC} = \frac{BD}{BC}$$

两式相加再整理即得定理中等式。 □

例 4.3.1 三角形的角平分线长度公式 我们利用张角定理来建立三角形的内角平分线长度与三边的关系，假若在图 4.7中，AD 正好平分 $\angle BAC$，其长度设为 ω_a，并设三个顶点 A、B、C 所对的边长分别为 a、b、c，再设 $\angle BAC = 2\theta$，那么由张角定理，有

$$\frac{\sin 2\theta}{\omega_a} = \frac{\sin\theta}{b} + \frac{\sin\theta}{b}$$

于是

$$\frac{2\cos\theta}{\omega_a} = \frac{1}{b} + \frac{1}{c} \tag{4.3.2}$$

而由半角公式和余弦定理，有

$$\begin{aligned}\cos\theta &= \sqrt{\frac{1+\cos 2\theta}{2}} \\ &= \sqrt{\frac{1+\frac{b^2+c^2-a^2}{2bc}}{2}} \\ &= \sqrt{\frac{(b+c)^2-a^2}{4bc}}\end{aligned}$$

所以最后

$$\omega_a = \sqrt{bc\left(1-\left(\frac{a}{b+c}\right)^2\right)}$$

这就是三角形的内角平分线长度公式。

由式 4.3.2还可得不等式 $\omega_a < H(b,c)$，这里 H 表调和平均数。 ■

4.3.2 梅涅劳斯 (Menelaus) 定理

定理 4.3.2 (梅涅劳斯定理)**.** 设 X、Y、Z 是三角形 ABC 的三边 AB、BC、CA 或其延长线上的点，其中有奇数个点在边的延长线上，则此三点共线的充分必要条件是下式成立:

$$\frac{AX}{XB}\cdot\frac{BY}{YC}\cdot\frac{CZ}{ZA}=1 \tag{4.3.3}$$

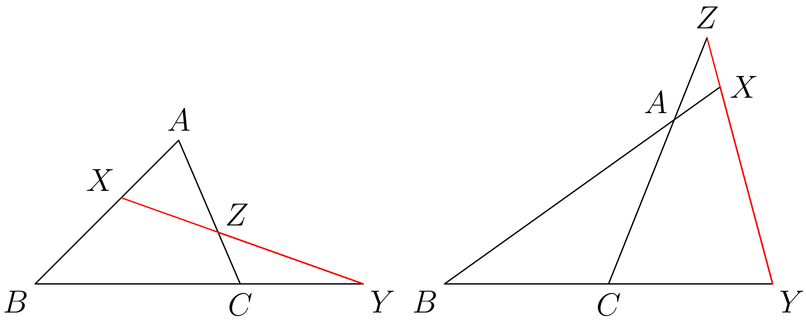

图 4.8 梅涅劳斯定理

只证明必要性，得证之后利用同一法即可轻松得到充分性的证明。最精巧的是下面这个辅助线的证明

证明一. 过 C 作 XYZ 的平行线，与边 AB 相交于点 P，那么有

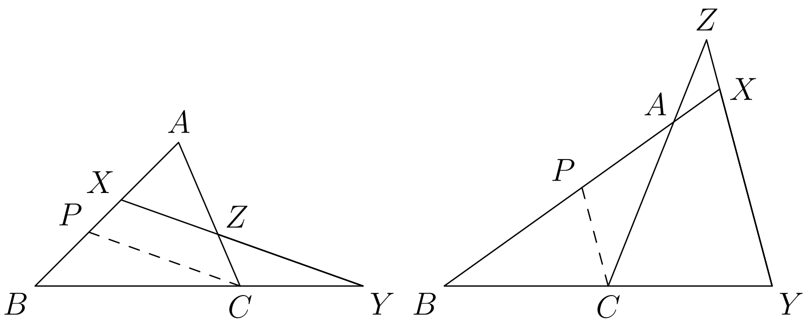

图 4.9 梅涅劳斯定理的证明

$$\frac{AX}{XB}\cdot\frac{BY}{YC}\cdot\frac{CZ}{ZA}=\frac{AX}{XB}\cdot\frac{BX}{XP}\cdot\frac{PX}{XA}=1$$

所以必要性成立，充分性利用同一法即可得证，略去。 □

另外一种辅助线的作法是分别过三角形的三个顶点向直线 XYZ 引垂线，剩下的过程与上面类似，此处就不再详述了。

面积方法仍然是行之有效的手段:

证明二.

$$\frac{AX}{XB}\cdot\frac{BY}{YC}\cdot\frac{CZ}{ZA}=\frac{S_{\triangle AYX}}{S_{\triangle BYX}}\cdot\frac{S_{\triangle BYX}}{S_{\triangle CYX}}\cdot\frac{S_{\triangle CYX}}{S_{\triangle AYX}}=1$$

□

再给一个向量方法的证明:

证明三. 记 $\overrightarrow{AX}=\lambda_1\overrightarrow{XB}$, $\overrightarrow{BY}=\lambda_2\overrightarrow{YC}$, $\overrightarrow{CZ}=\lambda_3\overrightarrow{ZA}$，则有

$$\overrightarrow{XZ}=\overrightarrow{AZ}-\overrightarrow{AX}=\frac{1}{1+\lambda_3}\overrightarrow{AC}-\frac{\lambda_1}{1+\lambda_1}\overrightarrow{AB}$$

同时

$$\begin{aligned}\overrightarrow{XY} &= \overrightarrow{AY}-\overrightarrow{AX}=\left(\frac{1}{1+\lambda_2}\overrightarrow{AB}+\frac{\lambda_2}{1+\lambda_2}\overrightarrow{AC}\right)-\frac{\lambda_1}{1+\lambda_1}\overrightarrow{AB}\\ &= \left(\frac{1}{1+\lambda_2}-\frac{\lambda_1}{1+\lambda_1}\right)\overrightarrow{AB}+\frac{\lambda_2}{1+\lambda_2}\overrightarrow{AC}\end{aligned}$$

这两个向量共线的充分必要条件是上面这两组系数对应成比例，经过计算即为等式 $\lambda_1\lambda_2\lambda_3=-1$，因为是有奇数个点在边的延长线上，所以它等价于定理中的等式。 □

从向量形式中我们得到了梅涅劳斯定理的向量表达: $\lambda_1\lambda_2\lambda_3=-1$，而且向量形式的证明，充分性与必要性是统一的。

例 4.3.2 考虑三个半径两两不同的圆，有三组外公切线，每两个圆的一组外公切线都有一个交点，这样的交点有三个，今将说明这三点共线[1]。

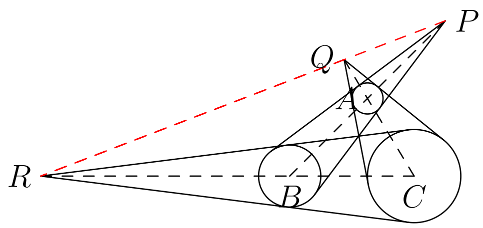

图 4.10 三个半径两两不同的圆的三组外公切线交点共线

如图4.10，易知 P、Q、R 三点在三个圆心形成的三角形的三条边的延长线上，而 $\frac{PA}{PB}=\frac{r_A}{r_B}$，仿此有另外两式，三式相乘即知这满足梅涅劳斯定理中的等式，所以这三点共线。

这结论还有另外一种解释，把这三个圆看成三个球，这三个球有公切面（实际上有两个），这三个点 P、Q、R 将同时位于这公切面和三个球心所决定的平面上，从而位于两个平面的交线上（三个球半径两两不等，所以这两个平面必定相交）。那这三个点

[1]这个例子及证法来自于参考文献 [7]

为什么在公切面上呢，可以这样理解，每两个球有一个公切圆锥，三个公切圆锥的顶点就是这三个点，而这三个圆锥与前述公切面必定都相切于一条圆锥的母线，所以作为圆锥顶点必定在这条母线上，也就在公切面上。 ■

例 4.3.3 梅涅劳斯定理的推广[2]及三角形内接三角形的最大面积 我们先把梅涅劳斯定理推广成如下的结论: 在三角形 ABC 中，点 D、E、F 分别是三边 BC、CA、AB 所在直线上的点，设 $\overrightarrow{BD}=\lambda_1\overrightarrow{DC}$，$\overrightarrow{CE}=\lambda_2\overrightarrow{EA}$，$\overrightarrow{AF}=\lambda_3\overrightarrow{FB}$，则有面积比

$$\frac{S_{\triangle DEF}}{S_{\triangle ABC}}=\frac{1+\lambda_1\lambda_2\lambda_3}{(1+\lambda_1)(1+\lambda_2)(1+\lambda_3)}$$

上式中的 $S_{\triangle DEF}$ 是有向面积，有正负号。

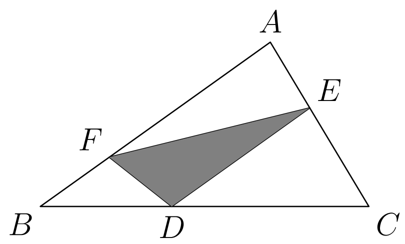

图 4.11 梅涅劳斯定理的推广

证明. 先证明三个点分别在三条边上的情况，这时 $\lambda_i>0(i=1,2,3)$，显然有 $S_{\triangle DEF}=S_{\triangle ABC}-(S_{\triangle AEF}+S_{\triangle BDF}+S_{\triangle CED})$，而

$$\frac{S_{\triangle AEF}}{S_{\triangle ABC}}=\frac{AF}{AB}\cdot\frac{AE}{AC}=\frac{\lambda_1}{1+\lambda_1}\cdot\frac{1}{1+\lambda_3}$$

同理还有

$$\frac{S_{\triangle BDF}}{S_{\triangle ABC}}=\frac{\lambda_2}{(1+\lambda_2)(1+\lambda_1)},\ \frac{S_{\triangle CED}}{S_{\triangle ABC}}=\frac{\lambda_3}{(1+\lambda_3)(1+\lambda_2)}$$

于是便可得结论。

对于有点在边的延长线上的情况，此时需要将面积 $S_{\triangle AEF}$、$S_{\triangle BDF}$、$S_{\triangle CFD}$、$S_{\triangle DEF}$ 理解为有向面积，于是关系式 $S_{\triangle DEF}=S_{\triangle ABC}-(S_{\triangle AEF}+S_{\triangle BDF}+S_{\triangle CED})$ 仍然成立，所以结论也是成立的。 □

在三个点 D、F、F 都在边上而非延长线上的情形，三角形 DEF 称为三角形 ABC 的内接三角形，利用刚证明的结论，我们可以考虑三角形的内接三角形的最大面积问题。

■

4.3.3 塞瓦 (Ceva) 定理

定理 4.3.3 (塞瓦定理). *在三角形 ABC 中，点 X、Y、Z 分别是三边 AB、BC、CA 或其延长线上的点，其中有偶数个点在边的延长线上，则三条线段 AY、BZ、CX 交于一点 P 或者相互平行的充分必要条件是*

$$\frac{AX}{XB}\cdot\frac{BY}{YC}\cdot\frac{CZ}{ZA}=1 \tag{4.3.4}$$

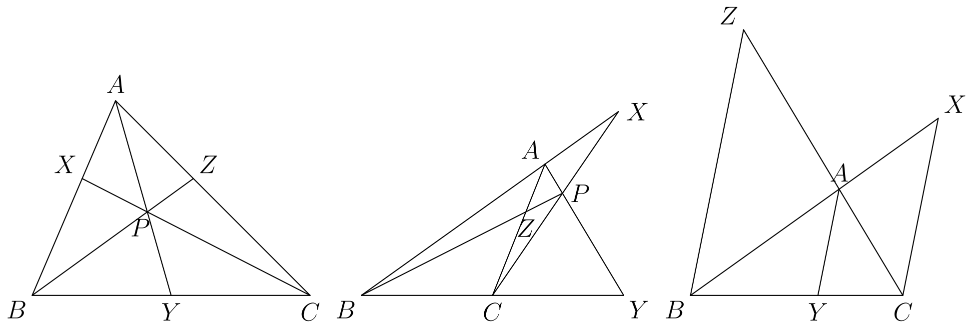

图 4.12 塞瓦定理

三线平行也可以视为相交于无穷远的点。

同样只证明必要性，第一个证明是直接使用了梅涅劳斯定理:

证明一. 对三角形 ABY 和截线 XPC 使用梅涅劳斯定理得

$$\frac{AX}{XB}\cdot\frac{BC}{CY}\cdot\frac{YP}{PA}=1$$

再对三角形 ACY 和截线 BPZ 使用梅涅劳斯定理得

$$\frac{YB}{BC}\cdot\frac{CZ}{ZA}\cdot\frac{AP}{PY}=1$$

两式相乘即得塞瓦定理。 □

面积方法仍然是最直接的:

证明二.

$$\frac{AX}{XB}\cdot\frac{BY}{YC}\cdot\frac{CZ}{ZA}=\frac{S_{\triangle CPA}}{S_{\triangle CPB}}\cdot\frac{S_{\triangle APB}}{S_{\triangle APC}}\cdot\frac{S_{\triangle BPC}}{S_{\triangle BPA}}=1$$

□

同样有向量方法

证明三. 仍然记 $\overrightarrow{AX}=\lambda_1\overrightarrow{XB}$，$\overrightarrow{BY}=\lambda_2\overrightarrow{YC}$，$\overrightarrow{CZ}=\lambda_3\overrightarrow{ZA}$，有

$$\overrightarrow{AX}=\frac{\lambda_1}{1+\lambda_1}\overrightarrow{AB},\ \overrightarrow{AZ}=\frac{1}{1+\lambda_3}\overrightarrow{AC}$$

因为点 P 是 BZ 与 CX 的交点，所以存在两个实数 s 和 t，使得

$$\begin{aligned}
\overrightarrow{AP} &= (1-s)\overrightarrow{AB}+s\overrightarrow{AZ}=(1-s)\overrightarrow{AB}+\frac{s}{1+\lambda_3}\overrightarrow{AC} \\
\overrightarrow{AP} &= (1-t)\overrightarrow{AC}+t\overrightarrow{AX}=\frac{t\lambda_1}{1+\lambda_1}\overrightarrow{AB}+(1-t)\overrightarrow{AC}
\end{aligned}$$

于是得方程组

$$1-s = \frac{t\lambda_1}{1+\lambda_1} \tag{4.3.5}$$

$$\frac{s}{1+\lambda_3} = 1-t \tag{4.3.6}$$

[2]这个推广来自于参考文献 [10].

解之得

$$t = 1 - \frac{1}{(1+\lambda_1)(1+\lambda_3) - \lambda_1}$$

所以

$$\overrightarrow{AP} = \frac{\lambda_1\lambda_3}{1+\lambda_3+\lambda_1\lambda_3}\overrightarrow{AB} + \frac{1}{1+\lambda_3+\lambda_1\lambda_3}\overrightarrow{AC}$$

另一方面有

$$\overrightarrow{AY} = \frac{1}{1+\lambda_2}\overrightarrow{AB} + \frac{\lambda_2}{1+\lambda_2}\overrightarrow{AC}$$

于是 $\overrightarrow{AP}$ 与 $\overrightarrow{AY}$ 共线的充分必要条件是两组系数成比例，而上面两个式子中的系数比分别为 $\lambda_1\lambda_3 : 1$ 和 $1 : \lambda_2$，所以得到 $\lambda_1\lambda_2\lambda_3 = 1$，即得证。 □

例 4.3.4 梅涅劳斯定理与塞瓦定理的统一　根据这证明过程，实际上梅涅劳斯定理和塞瓦定理可以统一起来：假定 X、Y、Z 是三角形 ABC 三边 AB、BC、CA 或它们的延长线上的点，有分比 $\overrightarrow{AX} = \lambda_1\overrightarrow{XB}$，$\overrightarrow{BY} = \lambda_2\overrightarrow{YC}$，$\overrightarrow{CZ} = \lambda_3\overrightarrow{ZA}$，那么这三点共线的充分必要条件是 $\lambda_1\lambda_2\lambda_3 = -1$（梅涅劳斯定理），而 AY、BZ、CX 三线共点的充分必要条件是 $\lambda_1\lambda_2\lambda_3 = 1$（塞瓦定理）。 ■

例 4.3.5 三角形的几个心及其向量表达　利用塞瓦定理，我们知道三角形的三条中线是交于同一点的，该点称为三角形的*重心*。

同样，利用塞瓦定理和三角形内角平分线定理，还可以知道三角形的三条内角平分线也是交于同一点的，这点称为三角形的*内心*，它是三角形内切圆圆心。

对于三角形的一个角的内角平分线和另外两个角的外角平分线，根据三角形的内外角平分线定理，可得这三线也是相交于同一点的，这一点称为三角形的*旁心*，它是旁切圆的圆心。

对于三条高线，假定 BC 边上的高线垂足是 D，则有 $\frac{BD}{CD} = \frac{c\cos B}{b\cos C}$，仿此有另外两式，三式相乘并根据塞瓦定理即知三条高也相交于同一点，称该点为三角形的*垂心*。

这四个心再加上三角形的*外心* (三边的中垂线交点)，合称为三角形的五心。

而且，在塞瓦定理的向量形式的证明过程中已经得出点 P 的向量表达

$$\overrightarrow{AP} = \frac{\lambda_1\lambda_3}{1+\lambda_3+\lambda_1\lambda_3}\overrightarrow{AB} + \frac{1}{1+\lambda_3+\lambda_1\lambda_3}\overrightarrow{AC}$$

或者将它改写为以平面上任一点 O 为起点的向量方程

$$\overrightarrow{OP} = \frac{\lambda_3}{1+\lambda_3+\lambda_1\lambda_3}\overrightarrow{OA} + \frac{\lambda_1\lambda_3}{1+\lambda_3+\lambda_1\lambda_3}\overrightarrow{OB} + \frac{1}{1+\lambda_3+\lambda_1\lambda_3}\overrightarrow{OC}$$

还可以将其中向量全部改写为以 P 为起点的向量方程，即

$$\lambda_3\overrightarrow{PA} + \lambda_1\lambda_3\overrightarrow{PB} + \overrightarrow{PC} = \mathbf{0}$$

这三个向量方程都是由塞瓦定理确定出的点 P 的向量表达，其中最后一种较为简洁，在此利用它来表达三角形的五心。

首先对于重心 G，显然 $\lambda_1 = \lambda_2 = \lambda_3 = 1$，因此

$$\overrightarrow{OG} = \frac{1}{3}(\overrightarrow{OA} + \overrightarrow{OB} + \overrightarrow{OC})$$

或者写成

$$\overrightarrow{GA}+\overrightarrow{GB}+\overrightarrow{GC}=\mathbf{0}$$

而对于内心 I，$\lambda_1=b/a$，$\lambda_2=c/b$，$\lambda_3=a/c$，因而

$$\overrightarrow{OI}=\frac{a}{a+b+c}\overrightarrow{OA}+\frac{b}{a+b+c}\overrightarrow{OB}+\frac{c}{a+b+c}\overrightarrow{OC}$$

或者写成

$$a\overrightarrow{IA}+b\overrightarrow{IB}+c\overrightarrow{IC}=\mathbf{0}$$

对于 BC 边上的旁心 I_A，有 $\lambda_1=-b/a$，$\lambda_2=c/b$，$\lambda_3=-a/c$，所以

$$\overrightarrow{OI_A}=\frac{-a}{-a+b+c}\overrightarrow{OA}+\frac{b}{-a+b+c}\overrightarrow{OB}+\frac{c}{-a+b+c}\overrightarrow{OC}$$

同样可以写成

$$-a\overrightarrow{I_AA}+b\overrightarrow{I_AB}+c\overrightarrow{I_AC}=\mathbf{0}$$

对于垂心 H，$\lambda_1=b\cos A/a\cos B$, $\lambda_2=c\cos B/b\cos C$,$\lambda_3=a\cos C/c\cos A$，于是

$$\overrightarrow{OH}=\frac{a\cos B\cos C\overrightarrow{OA}+b\cos C\cos A\overrightarrow{OB}+c\cos A\cos B\overrightarrow{OC}}{a\cos B\cos C+b\cos C\cos A+c\cos A\cos B}$$

注意到 $a\cos B\cos C:b\cos C\cos A:c\cos A\cos B=\tan A:\tan B:\tan C$，于是这式子还可以写为

$$\tan A\overrightarrow{HA}+\tan B\overrightarrow{HB}+\tan C\overrightarrow{HC}=\mathbf{0}$$

对于外心 O，$\lambda_1=b\cos B/a\cos A$,$\lambda_2=c\cos C/b\cos B$,$\lambda_3=a\cos A/c\cos C$，所以 (这里点 O 是外心，任意点换成了点 M)

$$\overrightarrow{MO}=\frac{a\cos A\overrightarrow{MA}+b\cos B\overrightarrow{MB}+c\cos C\overrightarrow{MC}}{a\cos A+b\cos B+c\cos C}$$

同样注意到 $a\cos A:b\cos B:c\cos C=\sin 2A:\sin 2B:\sin 2C$，所以式子可以写为

$$\sin 2A\overrightarrow{OA}+\sin 2B\overrightarrow{OB}+\sin 2C\overrightarrow{OC}=\mathbf{0}$$

这便是三角形五心的向量表达，进一步，利用向量的坐标，便可以得出五心的坐标公式，此处不再叙述。 ■

例 4.3.6 仿照例4.3.2，把那里的外公切线改为内公切线，那么图4.13中的三线共点。

因为 $\frac{AP}{PB}=\frac{r_A}{r_B}$，同样有另外两式，三式相乘后由塞瓦定理知三线共点。 ■

例 4.3.7 如图4.14，圆周上有顺次六个点 A、B、C、D、E、F，那么三条线段 AD、BE、CF 交于一点的充分必要条件是

$$\frac{AB}{BC}\cdot\frac{CD}{DE}\cdot\frac{EF}{FA}=1$$

证明是很容易的，连接 AC、CE、EA，对三角形 ACE 使用塞瓦定理就有

$$\frac{AX}{XC}\cdot\frac{CY}{YE}\cdot\frac{EZ}{ZA}=1$$

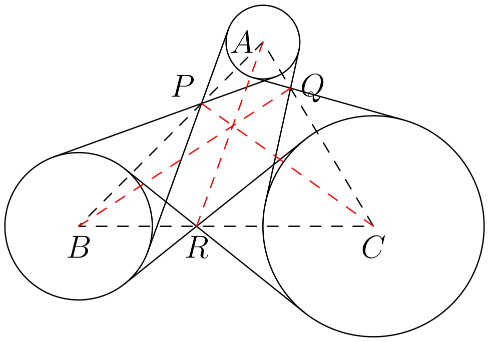

图 4.13 三个半径两两不同的圆的三组内公切线交点形成的塞瓦三线

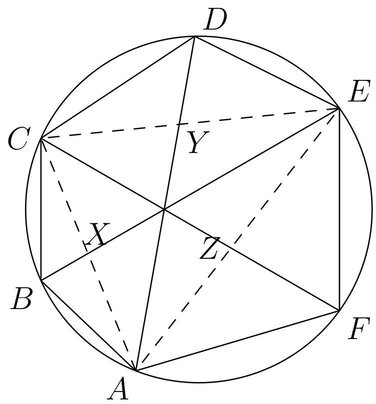

图 4.14

但是又有

$$\frac{AX}{XC}=\frac{AB\sin\angle ABX}{BC\sin\angle CBX}$$

仿此有另外两式，三式相乘并利用 $\angle ABX=\angle EDY$ 等一系列等角，即可得结论的必要性，而充分性仍然由同一法得出，略去。 ■

例 4.3.8 塞瓦定理的一个推广[3] 塞瓦定理可以推广成如下的结论：在三角形 ABC 中，点 D、E、F 分别是三边 BC、CA、AB 上的点，有比例 $\overrightarrow{AF}=\lambda_1\overrightarrow{FB}$，$\overrightarrow{BD}=\lambda_2\overrightarrow{DC}$，$\overrightarrow{CE}=\lambda_3\overrightarrow{EA}$，$BE$ 与 CF 相交于点 X，CF 与 AD 相交于 Y，AD 与 BE 相交于 Z，则有

$$\frac{S_{\triangle XYZ}}{S_{\triangle ABC}}=\frac{(\lambda_1\lambda_2\lambda_3-1)^2}{(1+\lambda_1+\lambda_1\lambda_2)(1+\lambda_2+\lambda_2\lambda_3)(1+\lambda_3+\lambda_3\lambda_1)}$$

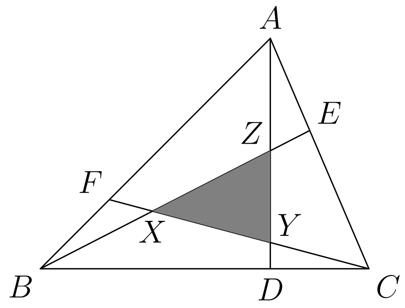

图 4.15 塞瓦定理的一个推广

证明. 如图4.15，显然有 $S_{\triangle XYZ}=S_{\triangle ABC}-(S_{\triangle ABZ}+S_{\triangle BCX}+S_{\triangle CAY})$，于是只要求出此式中被减掉的那三个三角形的面积就行了，因为

$$S_{\triangle ABZ}=\frac{BZ}{BE}S_{\triangle ABE}=\frac{BZ}{BE}\cdot\frac{AE}{AC}S_{\triangle ABC}$$

式中 AE/AC 是容易求的，BZ/BE 可以对三角形 BCE 和截线 AZD 使用梅涅劳斯定理得

$$\frac{BZ}{ZE}\cdot\frac{EA}{AC}\cdot\frac{CD}{DB}=1$$

所以

$$\frac{BZ}{ZE}=\lambda_2(1+\lambda_3)$$

从而

$$\frac{BZ}{BE}=\frac{\lambda_2(1+\lambda_3)}{1+\lambda_2(1+\lambda_3)}$$

所以最终

$$\frac{S_{\triangle ABZ}}{S_{\triangle ABC}}=\frac{BZ}{BE}\cdot\frac{AE}{AC}=\frac{\lambda_2}{(1+\lambda_2+\lambda_2\lambda_3)}$$

同样可得另外两式

$$\begin{aligned}\frac{S_{\triangle BCX}}{S_{\triangle ABC}}&=\frac{\lambda_3}{1+\lambda_3+\lambda_3\lambda_1}\\\frac{S_{\triangle CAY}}{S_{\triangle ABC}}&=\frac{\lambda_1}{1+\lambda_1+\lambda_1\lambda_2}\end{aligned}$$

[3]这个推广来自于参考文献 [10].

由此三式即可得结论。 □

■

例 4.3.9　如图4.16，在 $\triangle ABC$ 中，点 D 是角 BAC 的平分线与 BC 边的交点，过 D 向 AB 和 AC 引垂线，垂足分别是 E、F，连接 BF 和 CE 相交于点 P，今来证明 $AP \perp BC$。

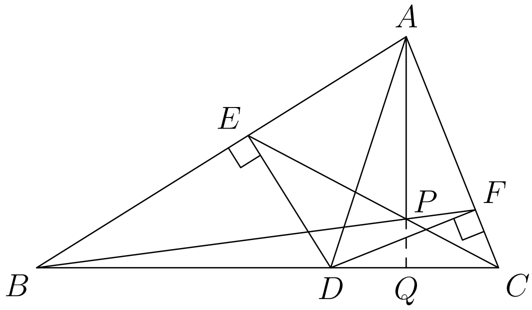

图 4.16

延长 AP 与 BC 边相交于点 Q，那么由塞瓦定理有

$$\frac{AE}{EB}\cdot\frac{BQ}{QC}\cdot\frac{CF}{FA}=1$$

另外显然有 $AE=AF$，所以由上式得 $\frac{BQ}{QC}=\frac{EB}{FC}$，而由几何关系和角平分线定理得

$$\frac{EB}{FC}=\frac{BD\cos B}{CD\cos C}=\frac{AB\cos B}{AC\cos C}$$

所以

$$\frac{BQ}{QC}=\frac{AB\cos B}{AC\cos C}$$

这说明点 Q 与 BC 边上的高线垂足重合，所以 AP 就是高线。 ■

例 4.3.10　如图4.17，在 $\triangle ABC$ 中，AD 是 BC 边上的高线，点 P 是这高线上任一点，直线 CP、FP 分别与边 AB、AC 相交于 E、F，连接 DE 和 DF，求证 $\angle ADE=\angle ADF$。

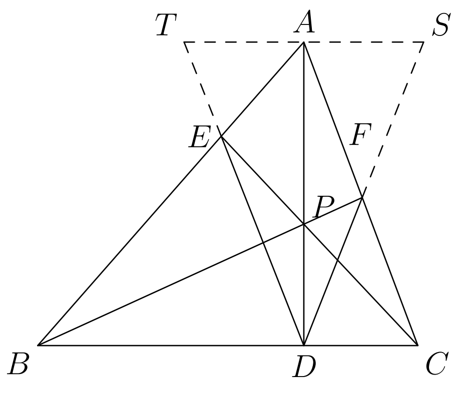

图 4.17

这是某年的全国高中联赛试题，也是一道名题，证法很多，从纯几何方法到坐标方法都有，这里只提供两个利用塞瓦定理的证明。

证明一. 由塞瓦定理有

$$\frac{AE}{EB}\cdot\frac{BD}{DC}\cdot\frac{CF}{FA}=1$$

另一方面有

$$\frac{AE}{EB}=\frac{S_{\triangle AED}}{S_{\triangle BED}}=\frac{AD\sin\angle ADE}{BD\sin\angle BDE}=\frac{AD\sin\angle ADE}{BD\cos\angle ADE}$$

同理有

$$\frac{AF}{FC}=\frac{AD\sin\angle ADF}{CD\cos\angle ADF}$$

这两式相除得

$$\frac{AE}{EB}\cdot\frac{BD}{DC}\cdot\frac{CF}{FA}=\frac{\sin\angle ADE\cos\angle ADF}{\cos\angle ADE\sin\angle ADF}$$

将这式子与由塞瓦定理所得式子相比较得 $\sin\angle ADE\cos\angle ADF=\cos\angle ADE\sin\angle ADF$，于是 $\sin(\angle ADE-\angle ADF)=0$，而这两个都是锐角，所以 $\angle ADE=\angle ADF$。 □

证明二. 如图4.17，过点 A 作 BC 边的平行线，分别与 DE 和 DF 和延长线相交于 T、S，由塞瓦定理

$$\frac{AE}{EB}\cdot\frac{BD}{DC}\cdot\frac{CF}{FA}=1$$

但由于

$$\frac{AE}{EB}=\frac{AT}{BD},\ \frac{CF}{FA}=\frac{CD}{AS}$$

所以得到 $AT=AS$，结合 $TS\perp AD$ 知 $\angle ADE=\angle ADF$。 □

■

4.3.4 斯特瓦尔特 (Stewart) 定理

定理 4.3.4 (斯特瓦尔特定理). *在三角形 ABC 中，点 D 是 BC 边上任一点，有下式成立:*

$$CD\cdot AB^2+BD\cdot AC^2=BC\cdot AD^2+BD\cdot CD\cdot BC \tag{4.3.7}$$

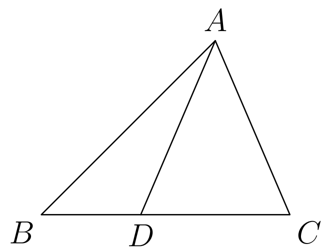

图 4.18 斯特瓦尔特定理

这定理中的等式较长，不太容易记住，写成如下的含比例的式子或许会有所帮助:

$$\frac{CD}{BC}\cdot AB^2+\frac{BD}{BC}\cdot AC^2=AD^2+BD\cdot CD \tag{4.3.8}$$

证明一. 由余弦定理

$$\begin{aligned}\cos\angle ADB &= \frac{AD^2+BD^2-AB^2}{2AD\cdot BD}\\ \cos\angle ADC &= \frac{AD^2+CD^2-AC^2}{2AD\cdot CD}\end{aligned}$$

因为 $\angle ADB$ 与 $\angle ADC$ 互补，所以 $\cos\angle ADB+\cos\angle ADC=0$，将上两式相加化简就得定理中的等式。 □

证明二. 设 $\overrightarrow{BD}=\lambda\overrightarrow{DC}$，则

$$\overrightarrow{AD}=\frac{1}{1+\lambda}\overrightarrow{AB}+\frac{\lambda}{1+\lambda}\overrightarrow{AC}$$

将 λ 换为 $\frac{BD}{DC}$，就是

$$\overrightarrow{AD}=\frac{CD}{BC}\overrightarrow{AB}+\frac{BD}{BC}\overrightarrow{AC}$$

两边平方

$$AD^2=\frac{CD^2}{BC^2}\cdot AB^2+\frac{BD^2}{BC^2}\cdot AC^2+\frac{BD\cdot CD}{BC^2}\cdot 2AB\cdot AC\cdot\cos\angle BAC$$

利用余弦定理将上式中最后一个余弦值换掉，化简就得定理中的等式。 □

例 4.3.11 在这个例子中，我们来建立一个有用的不等式，在三角形 ABC 中，点 D 是 BC 边上的异于端点的任意一点，那么有如下不等式成立:

$$AD<\frac{CD}{BC}\cdot AB+\frac{BD}{BC}\cdot AC \tag{4.3.9}$$

或者写成乘积的形式

$$BC\cdot AD<CD\cdot AB+BD\cdot AC$$

这个不等式与四边形中的托勒密不等式极为相似，实际上它就是四边形退化为三角形的情形，这个例子就是说托勒密不等式在这种情况下仍然成立。

最简单的是利用向量的证明:

证明一. 设 $\overrightarrow{BD}=\lambda\overrightarrow{DC}$，那么

$$\begin{aligned}\overrightarrow{AD} &= \frac{1}{1+\lambda}\overrightarrow{AB}+\frac{\lambda}{1+\lambda}\overrightarrow{AC}\\ &= \frac{CD}{BC}\overrightarrow{AB}+\frac{BD}{BC}\overrightarrow{AC}\end{aligned}$$

由此

$$\begin{aligned}|\overrightarrow{AD}| &= \left|\frac{CD}{BC}\overrightarrow{AB}+\frac{BD}{BC}\overrightarrow{AC}\right|\\ &< \left|\frac{CD}{BC}\overrightarrow{AB}\right|+\left|\frac{BD}{BC}\overrightarrow{AC}\right|\\ &= \frac{CD}{BC}|\overrightarrow{AB}|+\frac{BD}{BC}|\overrightarrow{AC}|\end{aligned}$$

即得证。 □

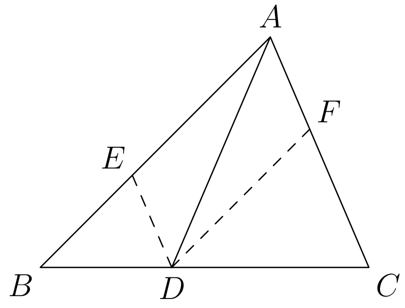

图 4.19

证明二. 如图4.19，过点 D 分别作边 AC、AB 的平行线，分别与 AB、AC 相交于点 E、F，有

$$\frac{DC}{BC}AB+\frac{BD}{BC}AC=\frac{AE}{AB}AB+\frac{AF}{AC}AC=AE+AF=AE+ED>AD$$

所以不等式得证。 □

证明三. 第三个证明就是利用斯特瓦尔特定理，在定理中的等式两边同时除以 BC，改为如下含比例的等式:

$$\begin{aligned}
AD^2 &= \frac{CD}{BC}\cdot AB^2+\frac{BD}{BC}\cdot AC^2-BD\cdot CD\\
&= \frac{CD\cdot BC}{BC^2}\cdot AB^2+\frac{BD\cdot BC}{BC^2}\cdot AC^2-BD\cdot CD\\
&= \frac{CD\cdot(BD+CD)}{BC^2}\cdot AB^2+\frac{BD\cdot(BD+CD)}{BC^2}\cdot AC^2-BD\cdot CD\\
&= \frac{CD^2}{BC^2}\cdot AB^2+\frac{BD^2}{BC^2}\cdot AC^2+\frac{BD\cdot CD}{BC^2}(AB^2+AC^2-BC^2)\\
&= \frac{CD^2}{BC^2}\cdot AB^2+\frac{BD^2}{BC^2}\cdot AC^2+\frac{BD\cdot CD}{BC^2}\cdot 2\cdot AB\cdot AC\cdot\cos\angle BAC\\
&< \frac{CD^2}{BC^2}\cdot AB^2+\frac{BD^2}{BC^2}\cdot AC^2+\frac{BD\cdot CD}{BC^2}\cdot 2\cdot AB\cdot AC\\
&= \left(\frac{CD}{BC}\cdot AB+\frac{BD}{BC}\cdot AC\right)^2
\end{aligned}$$

所以不等式成立。 □

当点 D 在 BC 边的延长线上时，比如说在 C 这一侧，有什么样的结论呢，这时把点 C 看成在三角形 ABD 的边 BD 上，应用刚证明的不等式，就有

$$BD\cdot AC<CD\cdot AB+BC\cdot AD$$

所以

$$AD>-\frac{CD}{BC}AB+\frac{BD}{BC}AC$$

这个结果看起来难以记忆，但实际上那两个系数跟点 D 分 $\overrightarrow{BC}$ 的比例有关，我们把这结果统一成下面这样:

命题 4.3.1. 在三角形 ABC 中，点 D 是 BC 边所在直线上不与 B 和 C 重合的任一点，设 $\overrightarrow{BD}=\lambda\overrightarrow{DC}$，那么在 $\lambda>0$ 时成立不等式

$$AD<\frac{1}{1+\lambda}AB+\frac{\lambda}{1+\lambda}AC \tag{4.3.10}$$

在 $\lambda<0$ 时不等式反向。

■

例 4.3.12 在这个例子中，我们利用斯特瓦尔特定理来建立三角形中的中线长公式、角平分线长公式、高线长公式。

三角形 ABC 的 BC 边上的中线长是

$$l_a^2=\frac{1}{4}(2b^2+2c^2-a^2) \tag{4.3.11}$$

由此可以继续推得一个不等式 $l_a<\frac{1}{2}(b+c)$。

利用内角平分线定理可求得角平分线分对边的比例，于是可以得出角 A 的平分线长度

$$\omega_a^2=bc\left(1-\left(\frac{a}{b+c}\right)^2\right) \tag{4.3.12}$$

由此可得不等式 $\omega<\sqrt{bc}$。

用斯特瓦尔特定理求高线长比较麻烦，倒是直接使用余弦定理方便:

$$\begin{aligned}
h_a &= b\sin C\\
&= b\sqrt{1-\cos^2 C}\\
&= b\sqrt{1-\left(\frac{a^2+b^2-c^2}{2ab}\right)^2}\\
&= \frac{1}{a}\sqrt{a^2b^2-\frac{(a^2+b^2-c^2)^2}{4}}\\
&= \frac{1}{2a}\sqrt{(a+b+c)(a+b-c)(b+c-a)(c+a-b)}
\end{aligned}$$

这就是高线长公式。

由于 $S_{\triangle ABC}=\frac{1}{2}a\cdot h_a$，再记 $p=\frac{1}{2}(a+b+c)$ 为半周长，那么我们实际上得出了著名的 海伦公式:

$$S_{\triangle ABC}=\sqrt{p(p-a)(p-b)(p-c)}$$

4.3.5　托勒密定理与托勒密不等式

4.3.6　西姆松定理

4.3.7　帕普斯 (Pappus) 定理

4.3.8　笛沙格定理

4.4　三角形

三角形是平面几何中最基本的图形之一，简单的图形却蕴含了丰富的内容，三角形的内容可以说是平面几何基础中最重要的内容。

4.4.1　内外角平分线定理

定理 4.4.1 (三角形内外角平分线定理). *三角形 ABC 中，角 A 的内角平分线交 BC 边于 E，同时它的外角平分线交 BC 边延长线于 F，那么有*

$$\frac{BE}{EC} = \frac{AB}{AC} = \frac{BF}{FC} \tag{4.4.1}$$

反之，如果三角形 BC 边上和边的延长线上分别有两个点 E 和 F 能满足这比例式，那么 AE 和 AF 就分别是 $\angle BAC$ 的内角平分线和外角平分线。

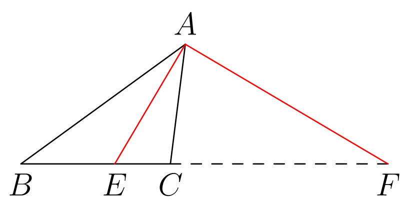

图 4.20 三角形内外角平分线定理

如果外角平分线与 BC 边平行，可以视为点 F 无穷远，此时右边的 $BF : FC$ 按极限理解，等式也是成立的。

只证明定理本身，逆定理根据同一法是容易得到的。

证明一. 过点 C 分别作内角平分线和外角平分线的平行线，分别与 AB 边及其延长线相交于点 P 和 Q，那么易得 $AC = AP$ 和 $AC = AQ$，所以

$$\frac{BE}{EC} = \frac{BA}{AP} = \frac{AB}{AC}$$

和

$$\frac{BF}{FC} = \frac{BA}{AQ} = \frac{AB}{AC}$$

于是定理得证。 □

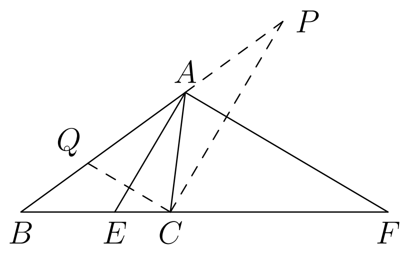

图 4.21 三角形内外角平分线定理的证明

证明二. 辅助线的作法同证明一，对三角形 QBC 和截线 AE 应用梅涅劳斯定理可得 $BE:EC=AB:AC$，对三角形 PBC 和截线 AF 应用梅涅劳斯定理可得 $BF:FC=AB:AC$。 □

例 4.4.1 阿波罗尼奥斯圆 今来考虑平面上到两个定点的距离之比为常数的动点轨迹，记这两个定点为 A 和 B，动点 P 在运动过程中始终满足 $PA:PB=\lambda(\neq 1)$，这里限制常数不等于 1 是因为那是线段 AB 的垂直平分线。

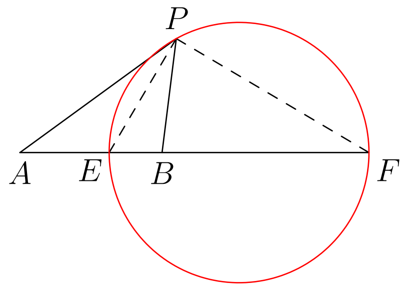

图 4.22 Apollonius 圆

首先在直线 AB 上可以找到两个点 E 和 F,使得 $\overrightarrow{AE}=\lambda\overrightarrow{EB}$ 和 $\overrightarrow{AF}=-\lambda\overrightarrow{FB}$,这里点 E 在线段 AB 上而点 F 在延长线上,动点 P 在运动过程中恒有 $PA:PB=AE:EB$ 和 $PA:PB=AF:FB$，当 P 不在直线 AB 上时，根据三角形的内外角平分线定理的逆定理，就有 PE 和 PF 分别是 $\angle APB$ 的内角平分线和外角平分线，所以 $PE\perp PF$，于是点 P 的轨迹就是以线段 EF 为直径的圆。这个结论最早是由古希腊几何学家阿波罗尼奥斯[4](Apollonius) 发现的，所以称为阿波罗尼奥斯圆。

值得一提的是，点 E 和 F 分线段 AB 的比为 λ，则点 A 和 B 分线段 EF 的比也相同，这个比值是 $|1-\lambda|/(1+\lambda)$，证明是很容易的，只要把线段 AE、EB、BF、AF 的长度都用 AB 的长度来表示就能得到这一点，这个结论是有用的，比如说任给一个圆和一个不等于 1 的正实数，我们可以反过来找出它是关于哪两个点的阿波罗尼奥斯圆，当然这两个点不是唯一的，在每一条直径上都能找出两对。 ■

[4] 参考文献 [9] 的作者。

4.4.2 正弦定理和余弦定理

4.4.3 海伦公式

定理 4.4.2 (海伦 (Helen) 公式)**.** *三角形面积 S 可表为 (式中 s 为半周长)*

$$S=\sqrt{s(s-a)(s-b)(s-c)} \tag{4.4.2}$$

证明. 由面积公式

$$\begin{aligned}
S &= \frac{1}{2}bc\sin A\\
&= \frac{1}{2}bc\sqrt{1-\cos^2 A}\\
&= \frac{1}{2}bc\sqrt{1-\left(\frac{b^2+c^2-a^2}{2bc}\right)^2}\\
&= \frac{1}{4}\sqrt{4b^2c^2-(b^2+c^2-a^2)^2}\\
&= \frac{1}{4}\sqrt{((b+c)^2-a^2)(a^2-(b-c)^2)}\\
&= \frac{1}{4}\sqrt{(a+b+c)(b+c-a)(a+b-c)(a+c-b)}\\
&= \sqrt{s(s-a)(s-b)(s-c)}
\end{aligned}$$

□

这与我国古代数学家秦九韶在《数书九章》中提出的三斜求积术是等价的（见推导过程）：

$$S_{\triangle ABC}=\frac{1}{2}\sqrt{a^2b^2-\left(\frac{a^2+b^2-c^2}{2}\right)} \tag{4.4.3}$$

■

4.4.4 一些三角恒等式

性质 4.4.1. *在三角形 ABC 中，有如下恒等式成立:*

$$\sin A+\sin B+\sin C = 4\cos\frac{A}{2}\cos\frac{B}{2}\cos\frac{C}{2} \tag{4.4.4}$$

$$\cos A+\cos B+\cos C = 1+4\sin\frac{A}{2}\sin\frac{B}{2}\sin\frac{C}{2} \tag{4.4.5}$$

证明. 只证明第一个，第二个也是类似的。

$$\begin{aligned}
\sin A+\sin B+\sin C &= (\sin A+\sin B)+\sin(A+B)\\
&= 2\sin\frac{A+B}{2}\cos\frac{A-B}{2}+2\sin\frac{A+B}{2}\cos\frac{A+B}{2}\\
&= 4\sin\frac{A+B}{2}\cos\frac{A}{2}\cos\frac{B}{2}\\
&= 4\cos\frac{A}{2}\cos\frac{B}{2}\cos\frac{C}{2}
\end{aligned}$$

□

4.4.5 重心

定理 4.4.3. 三角形的三条中线相交于一点。

三条中线的交点称为三角形的 重心。

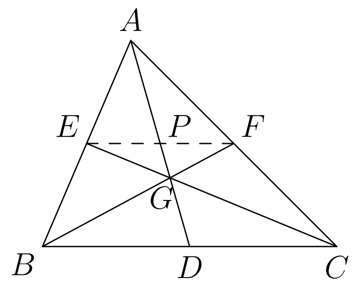

图 4.23 三角形的重心

根据塞瓦定理，这是显然的，在例 4.1.3中也给出了向量方法的证明，这里再给出几个证明:

证明一. 先由 AB 边上的中线 CE 和 AC 边上的中线 BF 相交得交点 G,易得 $S_{\triangle AGC} = S_{\triangle BGC}$ 和 $S_{\triangle AGB} = S_{\triangle CGB}$，从而得到 $S_{\triangle AGB} = S_{\triangle AGC}$，所以 AG 延长后与 BC 的交点 D 必是 BC 边中点，得证。 □

证明二. 仍然先由 AB 边上的中线 CE 和 AC 边上的中线 BF 相交于交点 G，直线 AG 与 BC 边相交于 D，中位线 EF 与直线 AG 相交于 P，那么有 $\triangle EFG \sim \triangle BCG$ 且相似比为 $1/2$，于是 $BG = 2GF$，所以 $\triangle PFG \sim \triangle BDG$ 且相似比也为 $1/2$，所以 $BD = 2PF$，而显然有 $BD = 2EP$，所以点 P 是线段 EF 中点，于是点 D 也是线段 BC 中点，于是得证。 □

性质 4.4.2. 对于平面上任一点 P 和三角形 ABC 的重心 G，有下式成立:

$$PA^2 + PB^2 + PC^2 = 3PG^2 + GA^2 + GB^2 + GC^2 \tag{4.4.6}$$

证明.

$$\overrightarrow{PA}^2 = (\overrightarrow{PG} + \overrightarrow{GA})^2 = \overrightarrow{PG}^2 + \overrightarrow{GA}^2 + 2\overrightarrow{PG} \cdot \overrightarrow{GA}$$

仿此有另外两式，三式相加并利用 $\overrightarrow{GA} + \overrightarrow{GB} + \overrightarrow{GC} = \overrightarrow{0}$ 即得证。 □

推论 4.4.1. 三角形的重心是到三个顶点的距离平方和最小的点。

4.4.6 外接圆与外心

性质 4.4.3. 三角形的外接圆半径 R，面积 S，和三边之间有如下的关系式

$$abc = 4RS \tag{4.4.7}$$

这性质意味着，如果以三角形为底面，外接圆半径为高作三棱柱，其体积与以三边为棱长的长方体的体积之间，有着固定的比例关系，不知这能否有构造法说明这一点?

证明一. 由正弦定理得 $2R=\frac{a}{\sin A}$，再由面积公式得 $S=\frac{1}{2}bc\sin A$，两式相乘即得结论。 □

证明二. 由正弦定理和海伦公式，过程中 $p=\frac{1}{2}(a+b+c)$ 为半周长

$$\begin{aligned}
2R &= \frac{a}{\sin A} \\
&= \frac{a}{\sqrt{1-\cos^2 A}} \\
&= \frac{a}{\sqrt{1-\left(\frac{b^2+c^2-a^2}{2bc}\right)^2}} \\
&= \frac{2abc}{\sqrt{4b^2c^2-(b^2+c^2-a^2)^2}} \\
&= \frac{2abc}{\sqrt{((b+c)^2-a^2)(a^2-(b-c)^2)}} \\
&= \frac{2abc}{\sqrt{(a+b+c)(b+c-a)(a+b-c)(c+a-b)}} \\
&= \frac{abc}{2\sqrt{p(p-a)(p-b)(p-c)}} \\
&= \frac{abc}{2S}
\end{aligned}$$

□

证明二实际上是又把海伦公式推了一遍。

4.4.7 内切圆与内心

性质 4.4.4. 三角形 ABC 中，设项点 A 与内心 I 所在直线与三角形外接圆相交于另一点 K，则 $KB=KI=KC$。

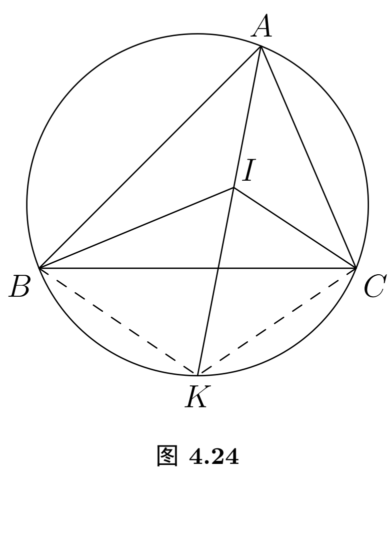

图 4.24

证明. $KB = KC$ 是显然的，只证明 $KI = KB$，易得 $\angle KIB = \angle KAB + \angle IBA = (\angle BAC + \angle ABC)/2$，同时 $\angle KBI = \angle KBC + \angle IBC = \angle KAC + \angle IBC = (\angle BAC + \angle ABC)/2$，所以 $KB = KI$。 □

性质 4.4.5. 三角形的内切圆半径 r 和外接圆半径 R 之间有如下等量关系

$$\frac{r}{R} = 4\sin\frac{A}{2}\sin\frac{B}{2}\sin\frac{C}{2} \tag{4.4.8}$$

证明. 因为 $r = \frac{1}{2}(b+c-a)\tan\frac{A}{2}$，$R = a/(2\sin A)$，所以

$$\begin{aligned}
\frac{r}{R} &= \frac{b+c-a}{a}\cdot\tan\frac{A}{2}\sin A \\
&= \frac{\sin B+\sin C-\sin A}{\sin A}\cdot\tan\frac{A}{2}\sin A \\
&= (\sin B+\sin C-\sin A)\tan\frac{A}{2} \\
&= (\sin B+\sin C-\sin(B+C))\tan\frac{A}{2} \\
&= \left(2\sin\frac{B+C}{2}\cos\frac{B-C}{2} - 2\sin\frac{B+C}{2}\cos\frac{B+C}{2}\right)\tan\frac{A}{2} \\
&= 4\sin\frac{B+C}{2}\sin\frac{B}{2}\sin\frac{C}{2}\tan\frac{A}{2} \\
&= 4\sin\frac{A}{2}\sin\frac{B}{2}\sin\frac{C}{2}
\end{aligned}$$

□

推论 4.4.2. 三角形的外心到三边的距离的代数和是 $R + r$，代数和是指：对于某一边，如果外心和另一个顶点在此边所在直线同侧，则外心到这边的距离取正值，否则取负值，R 和 r 分别指代外接圆半径和内切圆半径。

证明. 由外心性质，容易得这个距离和是 $R(\cos A + \cos B + \cos C)$，结合性质 4.4.1中的三角恒等式和性质 4.4.5，即可得证。 □

性质 4.4.6. 三角形的内切圆半径 r，外接圆半径 R，以及半周长 s，与三边之间有如下关系

$$abc = 4Rrs \tag{4.4.9}$$

证明. 由性质 4.4.3，结合面积 $S = \frac{1}{2}(a+b+c)r = sr$ 即得证。 □

4.4.8 欧拉不等式

定理 4.4.4 (欧拉 (Euler) 不等式)**.** 三角形的外接圆半径 R 和内切圆半径 r 满足不等式 $R \geqslant 2r$，等号仅当三角形为正三角形时取得。

证明一. 由性质 4.4.3和面积公式 $S = sr$，这里 s 为半周长，欲证不等式等价于

$$\frac{abc}{4S} \geqslant \frac{2S}{s}$$

即要证 $8S^2 \leqslant sabc$，由海伦公式，只需证

$$\begin{aligned}
&\iff 8s(s-a)(s-b)(s-c) \leqslant sabc \\
&\iff (a+b-c)(b+c-a)(c+a-b) \leqslant abc
\end{aligned}$$

作代换 $x=a+b-c$，$y=b+c-a$，$z=c+a-b$，则又等价于

$$xyz \leqslant \frac{x+z}{2}\cdot\frac{y+x}{2}\cdot\frac{z+y}{2}$$

由均值不等式，这显然成立，易见取等条件是三角形为正三角形。 □

证明二. 由 $2R=\frac{a}{\sin A}$ 及 $r=\frac{1}{2}(b+c-a)\tan\frac{A}{2}$ 得

$$\begin{aligned}
\frac{r}{R} &= \frac{(b+c-a)}{a}\sin A\tan\frac{A}{2} \\
&= \frac{(b+c-a)}{a}\cdot 2\sin^2\frac{A}{2} \\
&= \frac{(b+c-a)}{a}(1-\cos A) \\
&= \frac{(b+c-a)}{a}\left(1-\frac{b^2+c^2-a^2}{2bc}\right) \\
&= \frac{(b+c-a)[a^2-(b-c)^2]}{2abc} \\
&= \frac{(b+c-a)(c+a-b)(a+b-c)}{2abc}
\end{aligned}$$

因此要证的不等式等价于

$$(b+c-a)(c+a-b)(a+b-c) \leqslant abc$$

以下同证明一。 □

证明三. 由性质 4.4.5，及均值不等式和正弦函数在区间 $(0,\pi)$ 上的上凸性质，有

$$\begin{aligned}
\frac{r}{R} &= 4\sin\frac{A}{2}\sin\frac{B}{2}\sin\frac{C}{2} \\
&\leqslant 4\left(\frac{\sin\frac{A}{2}+\sin\frac{B}{2}+\sin\frac{C}{2}}{3}\right)^3 \\
&\leqslant 4\left(\sin\frac{\frac{A}{2}+\frac{B}{2}+\frac{C}{2}}{3}\right)^3 \\
&= \frac{1}{2}
\end{aligned}$$

□

4.4.9 垂心与垂心组

性质 4.4.7. 在三角形 ABC 中，O 和 H 分别为外心和垂心，D 是 BC 边中点，则有 $OD \parallel AH$ 并且 $OD=\frac{1}{2}AH$。

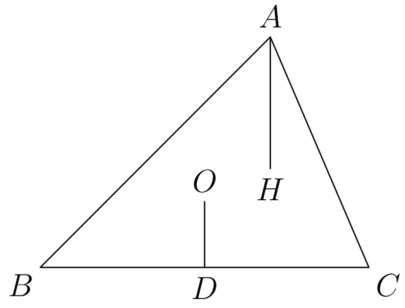

图 4.25 外心和垂心的一个性质

证明. 平行是明显的，只证明 2 倍比例关系，如图 4.26，作出外接圆和过 B 的直径 BP，连接 AP、CP，那么显然有 $OD \parallel CP$ 并且 $OD = \frac{1}{2}CP$，又容易证得四边形 $AHCP$ 为平行四边形，所以 AH 与 CP 平行且相等，故得证。

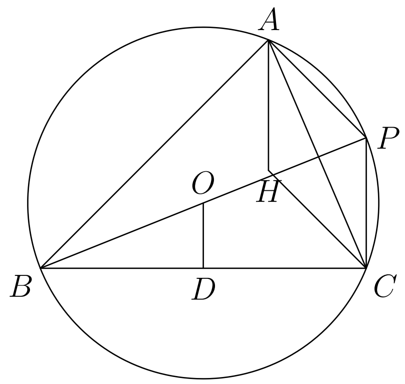

图 4.26

□

易证有如下性质

性质 4.4.8. 三角形的垂心关于三边的对称点，都落在其外接圆上。

性质 4.4.9. 三角形的垂心，是三条高线的垂足所组成的三角形的内心，反之，分别过三角形三顶点作对应顶点处的外角平分线构成一个新三角形，则原三角形的内心是这新三角形的垂心。

我们由下面这个定理引出 垂心组的概念。

定理 4.4.5. 三角形的三个顶点及其垂心构成一个四点组，这四点中每一个点都是以另外三点为顶点的三角形的垂心。

由这定理，这四点的地位是相同的，垂心的关系是相互的，于是就构成一个各点地位平等的四点组，这四点就称为平面上的一个 垂心组。

性质 4.4.10. 垂心组构成的四个三角形的外接圆半径相同，并且这四个外接圆的圆心也构成一个垂心组。

性质 4.4.11. 如果三个半径相同的圆交于一点，则该点与这三个圆的另外三个交点构成一个垂心组。

4.4.10　欧拉 (Euler) 定理与欧拉公式

定理 4.4.6 (欧拉 (Euler) 定理)**.** 三角形 ABC 的外心 O，重心 G，垂心 H 共线且 $OG = \frac{1}{2}GH$。

定理中的这条直线称为三角形的 欧拉线。

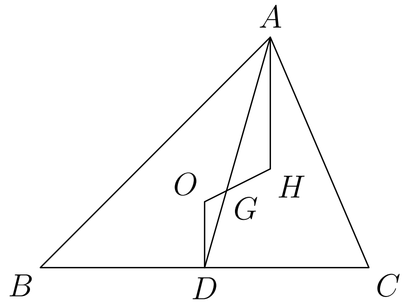

图 4.27

证明. 作出中线 AD，假如直线 OG 与直线 AH 相交于点 H'，那么由重心性质，$AG : GD = 2 : 1$，而 $OD \parallel AH$，所以 $AH' = 2OD$，根据性质 4.4.7，H' 必是垂心，得证。 □

定理 4.4.7. 三角形的外接圆圆心和内切圆圆心的距离是 $d = \sqrt{R(R-2r)}$，其中 R 和 r 分别指代外接圆半径和内切圆半径。

这公式称为欧拉公式。

证明一. 如图 4.28，设外心和内心连线所在直线与外接圆相交于 U、V 两点，角 A 的平分线与外接圆相交于另一点 K，由相交弦定理有 $AI \cdot IK = UI \cdot IV$，下面来计算各条线段长度。

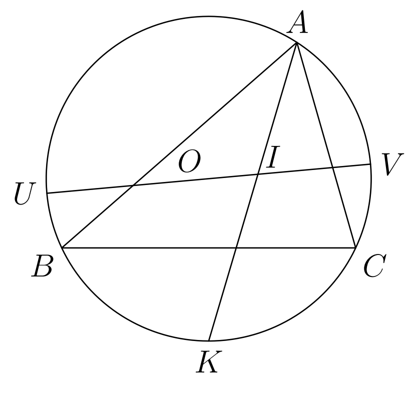

图 4.28

由性质 4.4.4，有 $KI = KB = 2R\sin\frac{A}{2}$，$AI = r/\sin\frac{A}{2}$，记 $OI = d$ 则 $UI\cdot VI = (R+d)(R-d)$，于是 $R^2 - d^2 = 2Rr$，得证。 □

证明二. 在三角形 AOI 中，对 $\angle OAI$ 使用余弦定理，由 $AO = R$，$AI = r/\sin\frac{A}{2}$，$\angle OAI = |\angle OAB - \angle IAB| = \left|\left(\frac{\pi}{2} - B\right) - \frac{A}{2}\right| = \left|\frac{\pi}{2} - \left(B + \frac{A}{2}\right)\right|$，再利用性质 4.4.5，有

$$
\begin{aligned}
d^2 &= R^2 + \left(\frac{r}{\sin\frac{A}{2}}\right)^2 - 2R\cdot\frac{r}{\sin\frac{A}{2}}\cdot\cos\left(\frac{\pi}{2} - \left(B + \frac{A}{2}\right)\right)\\
&= R^2 + \frac{r^2}{\sin^2\frac{A}{2}} - 2Rr\frac{\sin\left(B+\frac{A}{2}\right)}{\sin\frac{A}{2}}\\
&= R^2 + 2Rr\cdot\frac{2\sin\frac{B}{2}\sin\frac{C}{2}}{\sin\frac{A}{2}} - 2Rr\cdot\frac{\sin\left(B+\frac{A}{2}\right)}{\sin\frac{A}{2}}\\
&= R^2 + 2Rr\left(\frac{2\sin\frac{B}{2}\sin\frac{C}{2}}{\sin\frac{A}{2}} - \frac{\sin\left(B+\frac{A}{2}\right)}{\sin\frac{A}{2}}\right)\\
&= R^2 + 2Rr\left(\frac{2\sin\frac{B}{2}\sin\frac{C}{2}}{\sin\frac{A}{2}} - \frac{\cos\left(\frac{B-C}{2}\right)}{\sin\frac{A}{2}}\right)\\
&= R^2 - 2Rr\cdot\frac{\cos\left(\frac{B+C}{2}\right)}{\sin\frac{A}{2}}\\
&= R^2 - 2Rr
\end{aligned}
$$

□

4.4.11　旁切圆与旁心

4.4.12　费马点

4.4.13　内接三角形的最小周长

4.4.14　九点圆定理

4.5　圆

4.5.1　圆幂

早在中学数学中，我们就已经知道切割线定理和相交弦定理，而今我们将以统一的观点来看待这两个定理，这两个定理的共同点是，从一个定点作两条直线与定圆相交（相切可以视为极端情形），在其中每一条直线上，从定点出发到两个交点的两条线段的长度之积，都是相同的，也就是说这个乘积与具体的直线无关，仅与这个点与圆有关，于是我们便用如下的圆幂定理来取代切割线定理和相交弦定理:

定理 4.5.1 (圆幂定理). *过平面上一个定点引直线与一个定圆相交得到两个交点 (相切时两个交点重合)，则从该定点出发分别到达两个交点的两条线段长度之积是与直线位置无关的定值。*

那么接下来就要问，这个定值是多少？对于切割线定理而来，显然这个定点就是该点到圆的切线长的平方，而相交弦定理来说，让通过此点的弦以该点为中点，则显见这个定值是半径的平方与弦心距的平方之差，于是我们引入如下定义：

定义 4.5.1. *对于平面上的一个定点和一个定圆，称定点到圆心距离的平方减去半径的平方所得的差为该点对该圆的* 幂。

由定义，圆外的点对圆的幂为正，圆内的点对圆的幂为负，圆上的点对此圆的幂为零，于是圆幂定理中的定值，便是定点对定圆的幂的绝对值。

4.5.2 反演变换

4.5.3 极点与极线

4.5.4 根轴与根心

4.6 圆锥曲线

椭圆曲线形状优雅，方程优美，性质优酷[5]...... 拜托我不能再吹下去了，不然全世界的汽车车轮都该改成椭圆形了，想象一下坐在这样的车上是什么感受......

本节所谓的标准椭圆是指方程为 $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1(a>b>0)$ 的椭圆。

4.6.1 第一定义与第二定义

椭圆的两个定义是

定义 4.6.1. *平面上到两个定点的距离之和为定值 (大于两定点之间的距离) 的动点轨迹曲线称为* 椭圆.

定义 4.6.2. *平面上到一个定点的距离与一条不通过该定点的定直线的距离之比为一个小于 1 的正常数的动点轨迹曲线称为* 椭圆.

例 4.6.1 设点 P 是某椭圆上一个动点，F_1、F_2 分别是这椭圆的两个焦点，那么三角形 PF_1F_2 的重心、外心、内心、垂心、旁心的轨迹分别是什么? ■

椭圆两个定义是等价的，其实从根据各自的定义得出相同的椭圆方程就可以看出，但这里感兴趣的是纯几何角度的证明，这只要证明下面这个命题就可以了：

命题 4.6.1. *设 a 和 c 是两个正实数，且 $a>c$，F_1 和 F_2 是平面上相距为 $2c(c>0)$ 的两个定点，O 为它们的中点，L 为射线 OF_1 上满足 $OL=\frac{a^2}{c}$ 的点 (由于 $\frac{a^2}{c}>c$，因此 L 比 F_1 距离 O 更远一些)，直线 l 过点 L 且与直线 F_1F_2 垂直，则平面上的动点 P 到 F_1、F_2 的距离之和为 $2a$ 的充分必要条件是 $\frac{PF_1}{d}=\frac{c}{a}$，其中 d 是动点 P 到直线 l 的距离。*

[5] 又美又酷。

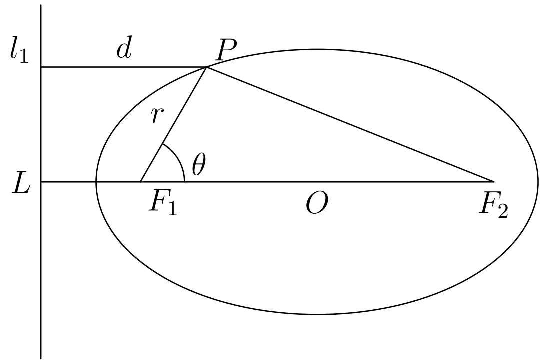

图 4.29 椭圆的第一定义与第二定义

证明. 先证必要性。设直线 PF_1 的倾斜角为 θ，记 $PF_1 = r$，则 $PF_2 = 2a - r$，在三角形 PF_1F_2 中由余弦定理得

$$(2a - r)^2 = r^2 + (2c)^2 - 2r \cdot 2c \cdot \cos\theta$$

记 $b^2 = a^2 - c^2$，则由上式解得

$$r = \frac{b^2}{a - c\cos\theta} \tag{4.6.1}$$

另外，根据几何关系显然有

$$d = LF_1 + PF_1\cos\theta = \frac{a^2}{c} - c + r\cos\theta = \frac{b^2}{c} + r\cos\theta \tag{4.6.2}$$

因此

$$\frac{d}{r} = \frac{1}{r}\frac{b^2}{c} + \cos\theta$$

将式4.6.1代入上式右端，立即得到 $\frac{d}{r} = \frac{a}{c}$，必要性得证。

下证充分性，倾斜角及 r 的定义同前，此时式4.6.2仍然成立，因此

$$\frac{a}{c} = \frac{d}{r} = \frac{b^2}{c} \cdot \frac{1}{r} + \cos\theta$$

解出 r 即得到4.6.1，由此之前的余弦定理式也是成立的，因此必然 $PF_2 = 2a - r$，充分性得证。 □

4.6.2 焦半径公式

椭圆上任一点与焦点的连线称为焦半径，在标准椭圆上任取一点 $P(x_0, y_0)$，那么它到左准线的距离 $d = x_0 + \frac{a^2}{c}$，根据第二定义，它到左焦点 F_1 的距离是 $PF_1 = ed = a + ex_0$，再根据第一定义，它到右焦点的距离是 $PF_2 = 2a - PF_1 = a - ex_0$，于是得坐标形式的 焦半径公式:

$$PF_1 = a + ex_0,\ PF_2 = a - ex_0 \tag{4.6.3}$$

假定向量 $\overrightarrow{F_1P}$ 与 x 轴正方向单位向量成角 θ，记其模 $r = |\overrightarrow{F_1P}|$ 在 $\triangle PF_1F_2$ 中使用余弦定理可得 $(2a - r)^2 = r^2 + 4c^2 - 4cr\cos\theta$，解得

$$r = \frac{b^2}{a - c\cos\theta} \tag{4.6.4}$$

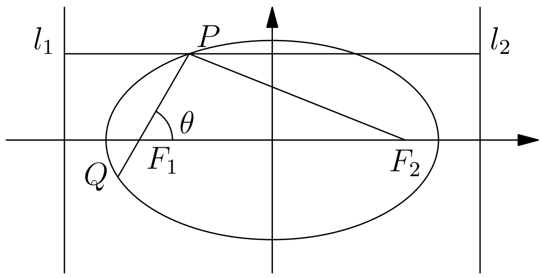

图 4.30 椭圆的焦半径

这是倾斜角形式的 焦半径公式。

例 4.6.2 如果 PF_1 延长线与椭圆相交于另一点 Q，易知向量 $\overrightarrow{F_1Q}$ 与 x 轴正方向单位向量成角 $\theta+\pi$，因此在上式中将 θ 换成 $\theta+\pi$，即得可知

$$F_1P=\frac{b^2}{a-c\cos\theta},\ F_1Q=\frac{b^2}{a+c\cos\theta} \tag{4.6.5}$$

从上式立刻得出，如果过焦点 F 的直线与椭圆相交于 P、Q 两点，那么有

$$\frac{1}{FP}+\frac{1}{FQ}=\frac{2a}{b^2} \tag{4.6.6}$$

■

例 4.6.3 设 A、B 是椭圆上的两个动点，且两个动点在运动过程中始终保持 $AF_1 \parallel BF_2$，我们来确定 AF_2 与 BF_1 的交点 P 的轨迹。

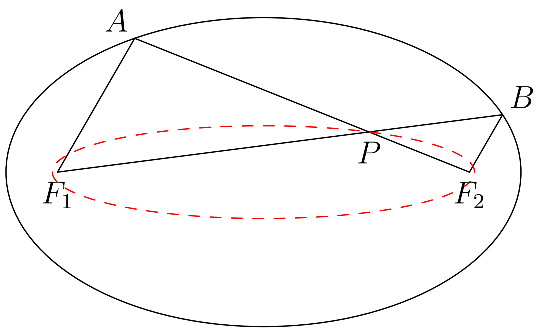

图 4.31

如图 4.31，假定 $AF_1=r_1$，$BF_2=r_2$，由 $\triangle PAF_1\sim\triangle PBF_2$ 且相似比为 r_1/r_2，

有

$$\begin{aligned}PF_1+PF_2 &= \frac{PF_1}{BF_1}BF_1+\frac{PF_2}{AF_2}AF_2\\ &= \frac{r_1}{r_1+r_2}BF_1+\frac{r_2}{r_1+r_2}AF_2\\ &= \frac{r_1}{r_1+r_2}(2a-r_2)+\frac{r_2}{r_1+r_2}(2a-r_1)\\ &= 2a-\frac{2r_1r_2}{r_1+r_2}\\ &= 2a-\frac{2}{\frac{1}{r_1}+\frac{1}{r_2}}\\ &= 2a-\frac{b^2}{a}\end{aligned}$$

最后一个等号处利用了上一个例子中的结论 (BF_2 平行且等于 AF_1 延长后与椭圆相交的另一段线段)，这结果这表明点 P 的轨迹是一个具有相同焦点的椭圆。 ■

4.6.3 切线与光学性质

椭圆曲线将其所在平面分成了三个部分：椭圆内部、椭圆曲线上、椭圆外部，对于椭圆曲线上的点，正如大家所熟知的，它们的坐标都满足椭圆的方程：

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \tag{4.6.7}$$

而对于椭圆内部的点，它们的坐标都满足不等式：

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}<1 \tag{4.6.8}$$

这个很好理解，对于椭圆内部任何一点，总能在椭圆上找到一个对应点，使其横坐标相同而纵坐标的绝对值大于椭圆内的点，而椭圆曲线上的这个对应点的坐标满足方程4.6.7，那么椭圆内的点的坐标满足不等式4.6.8就是显然的事情了。至于椭圆外部的点，那就只有成立不等式：

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}>1 \tag{4.6.9}$$

因为之前椭圆曲线与其方程之间以及椭圆内部的平面区域与其不等式之间都是可以互推的。

接下来讨论椭圆的切线问题，所谓切线，仿照圆的定义，按照直线与椭圆的交点数目，把直线与椭圆的位置关系划分为三类：相切、相交、相离。

在椭圆上任取一个点 $P(x_0,y_0)$，我们在例1.3.5中利用伸缩变换曾经得出过如下的切线方程:

$$\frac{x_0x}{a^2}+\frac{y_0y}{b^2}=1 \tag{4.6.10}$$

在这里，我们换一个角度来阐明它就是切线，在这直线上任取一点 $T(x_T,y_T)$，有：

$$(\frac{x_0^2}{a^2}+\frac{y_0^2}{b^2})+(\frac{x_T^2}{a^2}+\frac{y_T^2}{b^2})\geqslant 2(\frac{x_0x_T}{a^2}+\frac{y_0y_T}{b^2})=2 \tag{4.6.11}$$

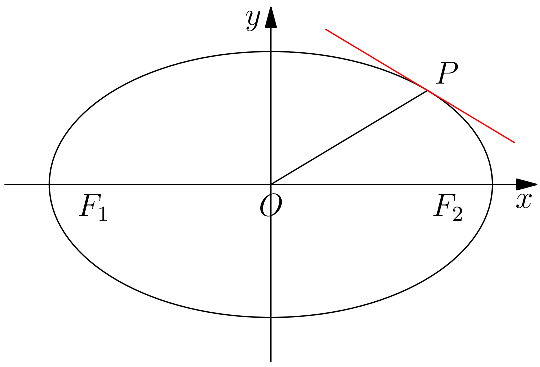

图 4.32 椭圆的切线

所以得到:

$$\frac{x_T^2}{a^2}+\frac{y_T^2}{b^2}\geqslant 1 \tag{4.6.12}$$

这表明直线4.6.10上除点 P 外任何一点都在椭圆外，与椭圆只有 P 一个交点，所以它理所当然就是点 P 处的切线方程。

椭圆具有一个漂亮的光学性质：从椭圆的焦点之一发出的光线，经椭圆反射之后，必经过另一个焦点。需要说明的是光线经椭圆反射，等同于经椭圆上相应点处的切线反射。

为着证明这一点，提出如下定理:

定理 4.6.1 (椭圆切线定理). *椭圆上某点处的切线，是该点与椭圆两个焦点连线所形成的焦点三角形在该点处的外角平分线。*

证明. 分两步证明，先证明定理中所说的外角平分线是椭圆的切线，再证明切线的唯一性。

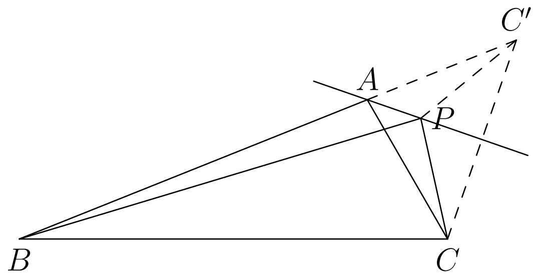

图 4.33 三角形外角平分线的一个性质

要证明此外角平分线与椭圆相切，只需证明它与椭圆只有唯一公共点即可。这实际上来自于平面几何中的一个结论：如图4.33，点 P 是三角形 $\triangle ABC$ 在顶点 A 处的外角平分线上异于点 A 的任意一点，那么有不等式 $PB+PC>AB+AC$ 成立。

证明也很简单，作出点 C 关于此外角平分线的对称点 C'，那么有 $PB+PC=PB+PC'>BC'=AB+AC'=AB+AC$.

这表明点 P 不可能在椭圆上，所以此外角平分线与椭圆除点 A 外无其它公共点，即为切线。

再继续证明切线的唯一性，这只要证明在上述三角形 ABC 中，过点 A 的其它直线上，必存在另一个异于点 A 的点 Q，使得 $QB+QC=AB+AC$ 即可，因为这样点 Q 就是该直线与椭圆的另一个交点，从而直线与椭圆是相交而非相切。

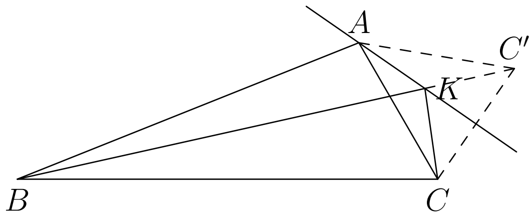

图 4.34 椭圆切线唯一性

如图4.34，同样作点 C 关于这直线的对称点 C'，记 BC' 与直线交点为 K，则 $KB+KC=KB+KC'=BC'<AB+AC'=AB+AC$，这表明点 K 到 B 和 C 两点的距离之和小于 $AB+AC$，显然直线上无穷远处的点到 B 和 C 的距离之和大于 $AB+AC$，所以直线上必存在一点 Q，使得 $QB+QC=AB+AC$，于是得证。 □

由上述定理立即可得椭圆的光学性质，这里我们再利用刚刚到得的切线方程来证明这一事实，假定从焦点发出的光线到达椭圆上一点 $P(x_0,y_0)$，而该点处的切线便是前文所求者，于是得出法线[6]的方程为

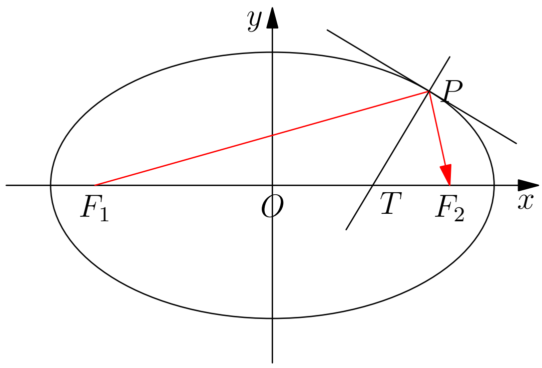

图 4.35 椭圆的光学性质

$$y-y_0=\frac{a^2y_0}{b^2x_0}(x-x_0)$$

容易求得法线与 x 轴交点是 $T(e^2x_0,0)$，于是

$$\frac{TF_1}{TF_2}=\frac{c+e^2x_0}{c-e^2x_0}=\frac{a+ex_0}{a-ex_0}=\frac{PF_1}{PF_2}$$

这意味着法线平分了反射点处的两条焦半径，所以由一个焦点处发出的光线经切线反射后必然经过另一焦点。

[6]法线是经过反射点并与反射面垂直的直线。

4.6.4 极点与极线

现在在椭圆外任选一个点 $P(x_0,y_0)$,并由它向椭圆引两条切线,切点分别为 $T_1(x_1,y_1)$ 和 $T_2(x_2,y_2)$，由于点 P 同时在两条切线上，所以同时成立着:

$$\frac{x_0x_1}{a^2}+\frac{y_0y_1}{b^2}=1,\frac{x_0x_2}{a^2}+\frac{y_0y_2}{b^2}=1 \tag{4.6.13}$$

由此立刻得知切点弦 T_1T_2 的方程是:

$$\frac{x_0x}{a^2}+\frac{y_0y}{b^2}=1 \tag{4.6.14}$$

这一点令人惊讶，要注意这时点 P 是在椭圆外，所以不要与刚才所证的切线方程混淆了。同样的方程，因为点 P 位置的不同，它代表的直线也不同。相信读者此时也会不失时机的提出如下问题：如果点 P 是在椭圆内部，那么方程4.6.14又代表什么样的直线呢？此刻还不好回答，但后文自然而然的会得出结论。

接着讨论弦的中点的问题，在椭圆上任取两点 $T_1(x_1,y_1)$ 和 $T_2(x_2,y_2)$，线段 T_1T_2 中点为 $Q(x_Q,y_Q)$，那么首先成立着等式：

$$\frac{x_1^2}{a^2}+\frac{y_1^2}{b^2}=1,\frac{x_2^2}{a^2}+\frac{y_2^2}{b^2}=1 \tag{4.6.15}$$

两式相减得到:

$$\frac{x_1^2-x_2^2}{a^2}+\frac{y_1^2-y_2^2}{b^2}=0 \tag{4.6.16}$$

进一步变形为:

$$\frac{y_1-y_2}{x_1-x_2}\frac{\frac{y_1+y_2}{2}}{\frac{x_1+x_2}{2}}=-\frac{b^2}{a^2} \tag{4.6.17}$$

所以弦 T_1T_2 及中点与椭圆中心连线 OQ 两者斜率之积为定值[7]:

$$k_{T_1T_2}k_{OQ}=-\frac{b^2}{a^2} \tag{4.6.18}$$

据此可以知道若另作一弦使其所在直线平行于 OQ，那么其中点与椭圆中心的连线必平行于直线 T_1T_2，即两个方向存在着相互性，这称之为椭圆的一组“共轭方向”，而这个结论则可以视为椭圆中的垂径定理。此外还可以知道，对于椭圆内部任何一个点，在经过它的所有弦中，只有唯一一条能使它成为弦的中点，而该弦的方向，就是该点与椭圆中心连线的共轭方向。而如果在这里平行移动弦 T_1T_2，直到它成为椭圆的切线 l，假如切点为 $P(x_0,y_0)$，那么我们应该得出：

$$k_lk_{OP}=-\frac{b^2}{a^2} \tag{4.6.19}$$

这与我们刚才所推导的点 P 处的切线方程正好吻合。

回到刚才过椭圆外一点 $P(x_0,y_0)$ 引两条切线 PT_1 及 PT_2 的场景中，以 Q 标记切点弦 T_1T_2 的中点，则点 Q 的坐标必然满足切点弦 T_1T_2 的方程：

$$\frac{x_0x_Q}{a^2}+\frac{y_0y_Q}{b^2}=1 \tag{4.6.20}$$

[7]本文中都不考虑斜率不存在等特殊情况，只就一般性的情况进行叙述，以得出一般性的结论。

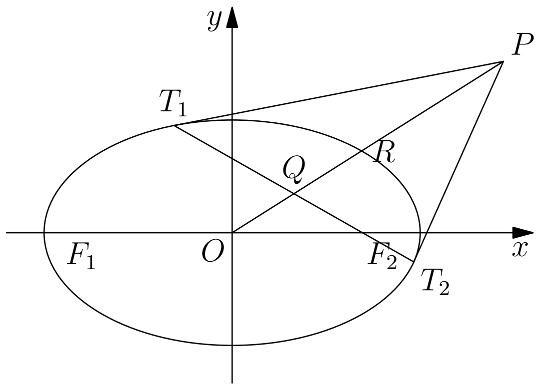

图 4.36 椭圆的极点与极线

这一点就很有意思了，因为这意味着点 P 位于下面的直线上：

$$\frac{x_Q x}{a^2}+\frac{y_Q y}{b^2}=1 \tag{4.6.21}$$

这是一条什么样的直线呢？容易发现，它的方向就是 OQ 的共轭方向！现在回想一下前面所提的悬而未决的问题：对于椭圆内的一点，方程4.6.14代表什么样的直线，现在就可以理直气壮的回答说，先找出以椭圆内的这点为中点的弦，然后过此弦两端点引椭圆切线，两切线有一个交点，这条直线就经过了这个点，那么方向呢？将椭圆内的这点与椭圆中心相连，其共轭方向就是该直线的方向！这段长长的描述显得有些不符合数学美感，但至少我们解决了如下问题：针对椭圆上、椭圆内、椭圆外的点 $P(x_0,y_0)$，方程4.6.14各自分别代表什么样的直线，而且还可以知道，椭圆内 (除去中心) 与椭圆外的点之间，存在着一种对称关系，借此我们可以在这两大平面区域之间，建立一对一的映射关系。

进一步，直线 OQ 与 T_1T_2 是共轭的，而现在直线4.6.21的方向也与 OQ 共轭，这表明直线4.6.21与切点弦 T_1T_2 是平行的！比较一下方程4.6.21与方程 4.6.14就可以得出结论：直线 PQ 是经过椭圆中心的！进而记 R 为线段 PQ 与椭圆的交点，那么显然：

$$\frac{x_R^2}{a^2}+\frac{y_R^2}{b^2}=1 \tag{4.6.22}$$

比较一下方程4.6.20与4.6.22，再加之我们知道 P、Q、R 三点共线，而且还过椭圆中心，所以：

$$x_0x_Q=x_R^2, y_0y_Q=y_R^2 \tag{4.6.23}$$

由此得到：

$$OP\cdot OQ=OR^2 \tag{4.6.24}$$

一切都是如此的完美。

前面曾经考虑过，当点 $P(x_0,y_0)$ 在椭圆内部时，方程4.6.10所代表的直线似乎难以描述，现在我们给它取一个好听的名字：极线。准确的说，是点 P 关于该椭圆的极线，而点 P 则称为直线4.6.10关于椭圆的极点。当点 P 在椭圆曲线上时，它关于椭圆的极线就是椭圆在该点处的切线（相切），当点 P 在椭圆外部时，它关于椭圆的极线就是对

应的切点弦所在直线（相交），而当点 P 在椭圆内时，对应的极线就在椭圆外部了（相离）。而且我们看到，点 P 无论位于哪个位置（椭圆中心除外），它关于椭圆的极线的方向都是 OP 的共轭方向，所以如果让点 P 在一条从椭圆中心出发的射线上由椭圆内到椭圆外缓慢移动，我们将看到它关于椭圆的极线由远及近平行移动，当点 P 移到椭圆上的那一刻，极点刚好位于极线上，而极线也成为过极点的切线，此后，当点 P 移动到椭圆外时，极线则与椭圆椭圆相交，当点 P 向无穷远处移动时，极线则向椭圆中心无限靠近。容易看出，对于平面上任何一点（椭圆中心除外），它关于椭圆的极线都是唯一存在的，反过来，对于平面上任何一条不通过椭圆中心的直线，它关于椭圆的极点也是唯一存在的。而切点与切线，不过是极点刚好位于极线上的一种特殊情况罢了。

4.6.5 纯几何观点下的椭圆

这一节尝试从纯几何的角度对椭圆进行讨论，所谓纯几何，就是不使用坐标几何。

定义 4.6.3. *平面上到两个定点的距离之和为定值 (大于这两个定点之间的线段长) 的动点轨迹称为椭圆。*

这两个点称为椭圆的焦点，以 F_1 和 F_2 来表示，两个焦点之间的距离称为椭圆的焦距，记为 $2c$，这个距离和记为 $2a$。

相应的，如果某个点 Q 满足 $QF_1+QF_2<2a$，称点在椭圆的内部，不等式如果是大于，则称点在椭圆的外部，这里的内部和外部都是形式化的定义，暂时还不能把它们与封闭曲线的内部和外部等同起来。

性质 4.6.1. *椭圆是轴对称图形，两个焦点所在直线以及它们的垂直平分线是两条对称轴。椭圆也是中心对称图形，两个焦点的中点是对称中心。*

这个性质是显而易的，如果点 P 满足 $PF_1+PF_2=2a$，那么该点关于上面所说的两条对称轴的对称点，以及关于两个焦点的中点的对称点，也都能够满足这个距离和的要求。

两个焦点的中点称为椭圆的中心。

性质 4.6.2. *椭圆是有界图形，即存在外含矩形或者外含圆。*

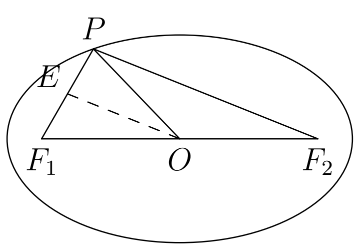

图 4.37 椭圆弦上的点除端点外都在椭圆内部

证明. 如图4.37，点 P 是椭圆上任一点，那么有

$$OP=\frac{1}{2}(OP+OP)<\frac{1}{2}((OF_1+PF_1)+(OF_2+PF_2))=\frac{1}{2}(F_1F_2+PF_1+PF_2)=a+c$$

所以椭圆上的任意点都被包含于以椭圆中心为圆心，以 $a+c$ 为半径的圆内。 □

其实这个外含圆的半径还可以进一步缩小，取 PF_1 中点 E，有

$$OP < PE + OE = \frac{1}{2}PF_1 + \frac{1}{2}PF_2 = a$$

然后因为在直线 F_1F_2 的两侧延长线上各有一点 P_1、P_2，使得 $P_1F_1 = P_2F_2 = a - c$，易知这两个点在椭圆上，所以这个半径 a 不能再减小了。

证明二. 利用三角形的中线长公式 $l^2 = \frac{1}{4}(2a^2 + 2b^2 - c^2)$，这里 a、b、c 分别代指三角形三边而非椭圆中的相关意义。那么有

$$\begin{aligned} l^2 - \left(\frac{a+b}{2}\right)^2 &= \frac{1}{4}(2a^2 + 2b^2 - c^2) - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{1}{4}\left((a-b)^2 - c^2\right) \end{aligned}$$

显然这是负值，所以中线长小于另外两边和的一半，于是应用到性质中就有 $OP < \frac{1}{2}(PF_1 + PF_2) = a$，得证。 □

椭圆上任意两个不同的点之间的连线段称为椭圆的弦。

性质 4.6.3. 椭圆任一条弦上的除端点外的任意点都在椭圆的内部，而延长线上的点都在椭圆的外部。

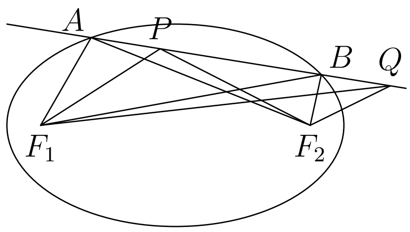

图 4.38 椭圆弦上的点除端点外都在椭圆内部

证明. 如图4.38，AB 是椭圆的一条弦，点 P 是这弦上异于端点的任意一点，而 Q 是 AB 延长线上点 B 一侧上的任一点，根据我们在例4.3.11中所证的不等式，有

$$PF_1 < \frac{PB}{AB}AF_1 + \frac{PA}{AB}BF_1$$
$$PF_2 < \frac{PB}{AB}AF_2 + \frac{PA}{AB}BF_2$$

以及

$$QF_1 > -\frac{QB}{AB}AF_1 + \frac{QA}{AB}BF_1$$
$$QF_2 > -\frac{QB}{AB}AF_2 + \frac{QA}{AB}BF_2$$

两式相加得

$$\begin{aligned} PF_1 + PF_2 &< \frac{PB}{AB}(AF_1 + AF_2) + \frac{PA}{AB}(BF_1 + BF_2) \\ &= \frac{PB}{AB} \cdot 2a + \frac{PA}{AB} \cdot 2a \\ &= 2a \end{aligned}$$

以及

$$\begin{aligned} QF_1 + QF_2 &> -\frac{QB}{AB}(AF_1 + AF_2) + \frac{QA}{AB}(BF_1 + BF_2) \\ &= -\frac{QB}{AB} \cdot 2a + \frac{QA}{AB} \cdot 2a \\ &= 2a \end{aligned}$$

所以结论成立。 □

性质 4.6.4. 设点 P 是椭圆上任一点，点 O 是椭圆的中心，那么线段 OP 上除点 P 外的任一点都在椭圆内部，而 OP 延长线上位于 P 一侧的点都在椭圆的外部。

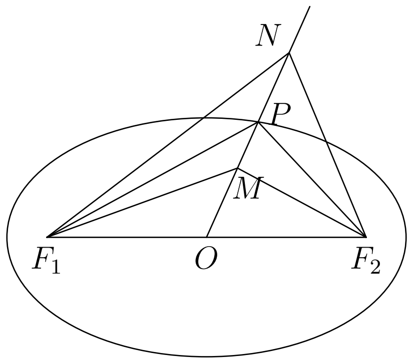

图 4.39

证明. 先证明一个几何不等式，如图4.40，点 P 是三角形 ABC 内任一点，那么有不等式 $PB + PC < AB + AC$ 成立。为这不等式给出两个证法，第一个是作辅助线，延长 CP 与 AB 相交于点 E，那么有

$$PB + PC < (BE + EP) + PC = BE + EC < BE + (EA + AC) = AB + AC$$

这就证明了这个不等式，另外一个证明是利用我们在例4.3.11中所建立的不等式，延长 AP 与 BC 相交于点 D，有

$$\begin{aligned} BP &< \frac{AP}{AD}BD + \frac{DP}{AD}AB \\ CP &< \frac{AP}{AD}CD + \frac{DP}{AD}AC \end{aligned}$$

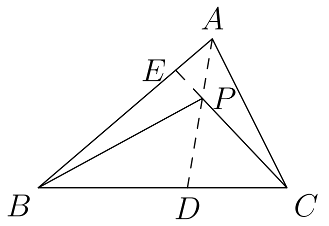

图 4.40

两式相加得

$$\begin{aligned} BP+CP &< \frac{AP}{AD}(BD+CD)+\frac{DP}{AD}(AB+AC) \\ &= \frac{AP}{AD}BC+\frac{DP}{AD}(AB+AC) \\ &< \frac{AP}{AD}(AB+AC)+\frac{DP}{AD}(AB+AC) \\ &= AB+AC \end{aligned}$$

这样也证明了这个不等式。

利用这个不等式，在图4.39，中，就有

$$MF_1+MF_2<PF_1+PF_2<NF_1+NF_2$$

也就是

$$MF_1+MF_2<2a<NF_1+NF_2$$

故得证。 □

性质 4.6.5. 过椭圆的两个焦点向椭圆任一切线引两条垂线，则无论切线位置如何，两个垂足始终位于同一圆上，该圆以椭圆中心为圆心，以椭圆半长轴为半径。

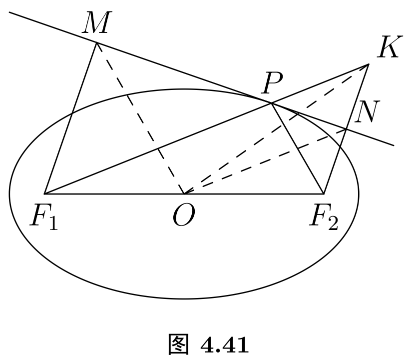

图 4.41

证明. 如图4.41，首先由直角梯形 F_1MNF_2 中位线易知 $OM=ON$，只要继续证明 ON 等于椭圆半长轴即可，由椭圆切线定理知，F_1P 延长后与 F_2N 的延长线相交于点 K，则点 K 是点 F_2 关于直线 MN 的对称点，所以 $F_1K=PF_1+PF_2=2a$，而三角形 F_1KF_2 中位线 $ON=\frac{1}{2}F_1K=a$。 □

4.7 抛物线

性质 4.7.1. 如图 *4.42*所示，过抛物线对称轴上焦点外侧任一点 D，分别向过焦点 F 的弦 MN 的两个端点引直线，分别与准线相交于点 P 和 Q，准线与对称轴的交点是 R，那么*: (1)* 直线 DR、PN、QM 三线共点 *(*设为点 T*)*，*(2)* 对称轴上的四点 R、T、F、D 构成调和点列。

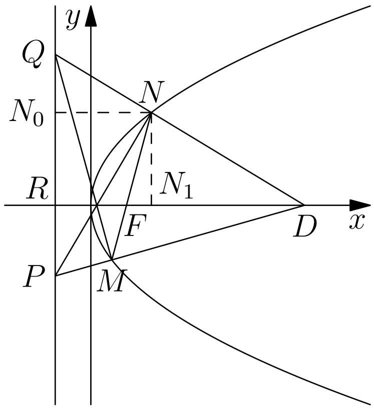

图 4.42 抛物线焦点弦的一个性质

证明. (1) 由塞瓦定理，只需证

$$\frac{DN}{NQ}\cdot\frac{QR}{RP}\cdot\frac{PM}{MD}=1 \tag{4.7.1}$$

设点 N 在准线上的投影是 N_0，则有 $NF=NN_0$，再设点 N 在抛物线对称轴上的投影是 N_1，则

$$\frac{DN}{NQ}=\frac{NN_1}{QN_0}=\frac{NN_1}{NF}\cdot\frac{NN_0}{QN_0}=\sin\angle NFD\cdot\tan\angle DQP$$

同理可得

$$\frac{DM}{MP}=\sin\angle MFR\cdot\tan\angle DPQ$$

而显然 $\angle NFD=\angle MFR$，而且

$$\frac{QR}{RP}=\frac{\frac{DR}{RP}}{\frac{DR}{QR}}=\frac{\tan\angle DPQ}{\tan\angle DQP}$$

因此式4.7.1成立，结论得证。

(2) 对 $\triangle QRD$ 和截线 PTN 使用梅涅劳斯定理得

$$\frac{DN}{NQ}\cdot\frac{QP}{PR}\cdot\frac{RT}{TD}=1$$

再对 $\triangle PRD$ 和截线 QTM 使用梅涅劳斯定理得

$$\frac{DM}{MP}\cdot\frac{PQ}{QR}\cdot\frac{RT}{TD}=1$$

由此二式便得

$$\frac{RT}{TD}\left(\frac{DN}{NQ}+\frac{DM}{MP}\right)=1$$

记 $\angle NFD=\theta$，$\angle DQP=\alpha$，$\angle DPQ=\beta$，则由 (1) 的证明过程中所得结果可得

$$\frac{DN}{NQ}+\frac{DM}{MP}=\sin\theta(\tan\alpha+\tan\beta)$$

记 $DF=d$，有熟知的焦半径

$$NF=\frac{p}{1-\cos\theta},\ MF=\frac{p}{1+\cos\theta}$$

由几何关系得

$$\tan\alpha=\tan\angle DNN_1=\frac{DN_1}{NN_1}=\frac{DF-NF\cos\theta}{NF\sin\theta}=\frac{d(1-\cos\theta)-p\cos\theta}{p\sin\theta}$$

同理可得

$$\tan\beta=\frac{d(1+\cos\theta)+p\cos\theta}{p\sin\theta}$$

由此二式便得

$$\sin\theta(\tan\alpha+\tan\beta)=\frac{2d}{p}$$

即

$$\frac{DN}{NQ}+\frac{DM}{MP}=\frac{2d}{p}$$

所以

$$\frac{RT}{TD}=\frac{p}{2d}$$

由此便知，当抛物线和点 D 的位置确定后，点 T 的位置便固定下来，与焦点弦的具体位置无关，由上式不难求得

$$RT=\frac{p(p+d)}{p+2d},\ TF=\frac{dp}{p+2d}$$

于是

$$\frac{RT}{TF}=\frac{p+d}{d}=\frac{RD}{DF}$$

所以这四点为调和点列. □

4.8 题集

题 4.8.1 如图，在三角形 ABC 中，AO 是 BC 边上的高，D 是 AO 的中点，点 E 是其内切圆 $\odot I$ 与 BC 边的切点，连接 ED 并延长与内切圆相交于另一点 F，过点 C 并与 BC 垂直的直线与 BI 的延长线相交于 G，求证:(1) $EF\perp FG$,(2)EF 平分角 BFC。 ■

题 4.8.2 (姚佳斌供题) 如图 4.44，在 $\triangle ABC$ 中，点 H 是垂心，AD 是 BC 边上的中线，过 H 作 AD 的垂线，分别与 AC、AB 边相交于 E、F，连结 BE、CF 相交于点 K，求证 A、H、K 三点共线。 ■

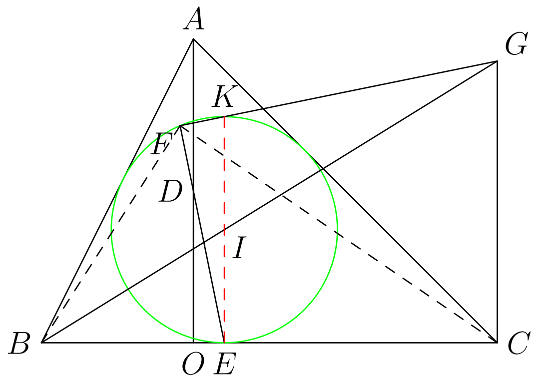

图 4.43

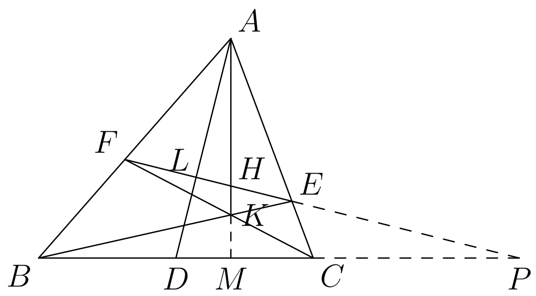

图 4.44

证明. 只要证明 BE、CF、AH 三线共点就可以了，设 BC 边上的高线垂足是 M，由塞瓦定理，只需证

$$\frac{AF}{FB}\cdot\frac{BM}{MC}\cdot\frac{CE}{EA}=1$$

延长 FE 与 BC 边延长线相交于点 P，再设 EF 与 AD 交点为 L，则对 $\triangle ABC$ 和截线 FEP 应用梅涅劳斯定理得

$$\frac{AF}{FB}\cdot\frac{BP}{PC}\cdot\frac{CE}{EA}=1$$

所以只需要证明下式就可以了:

$$\frac{BM}{MC}=\frac{BP}{PC}$$

现在就来证明这一点，易知 L、D、M、H 四点共圆，所以

$$AL\cdot AD=AH\cdot AM$$

又易知 A、L、M、P 四点共圆，所以

$$DL\cdot DA=DM\cdot DP$$

这两式相加得

$$AD^2=AH\cdot AM+DM\cdot DP$$

以下就来从这式中求出 DP，易知

$$\begin{aligned} AD^2 &= \frac{1}{2}(b^2+c^2)-\frac{1}{4}c^2 \\ AH &= \frac{\cos A}{\sin B}b \\ AM &= c\sin B \\ DM &= \frac{1}{2}a-b\cos C \end{aligned}$$

全部代入前式中即得

$$DP=\frac{a^2}{2(a-2b\cos C)}$$

进而 (结合余弦定理)

$$\begin{aligned} BP &= DP+\frac{a}{2}=\frac{a^2+c^2-b^2}{2(a-2b\cos C)} \\ PC &= DP-\frac{a}{2}=\frac{a^2+b^2-c^2}{2(a-2b\cos C)} \end{aligned}$$

于是

$$\frac{BP}{PC}=\frac{a^2+c^2-b^2}{a^2+b^2-c^2}=\frac{2ac\cos B}{2ab\cos C}=\frac{c\cos B}{b\cos C}=\frac{BM}{MC}$$

得证。 □

题 4.8.3 如图4.45，分别以三角形 ABC 的边 AB 和 AC 向形外作正方形，两个正方形的中心分别记为 O_1 和 O_2，X 是 BC 边中点，其余各点的意义如图所示，求证:

1. 三角形 O_1XO_2 是等腰直角三角形.

2. $ST \parallel PQ$.

■

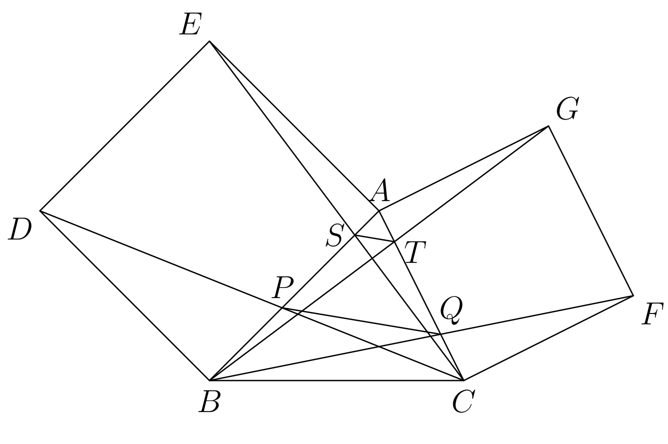

图 4.45

证明. 易见 $\triangle ACE \cong \triangle AGB$，而 $O_1X \parallel EC$ 并且 $|O_1X| = \frac{1}{2}|EC|$，同样 $O_2X \parallel BG$ 并且 $|O_2X| = \frac{1}{2}|BG|$，但是 $EC = BG$，并且 $EC \perp BG$，所以 $O_1X = O_2X$ 并且 $O_1X \perp O_2X$，所以 $\triangle O_1XO_2$ 是等腰直角三角形.

第二问的最简单证法是设 DE 延长线与 CA 延长线交于 U，FG 延长线与 BA 延长线交于 V，则有 $\triangle AEU \sim \triangle AGV$，于是

$$\frac{AS}{SP} = \frac{UE}{ED} = \frac{UE}{EA} = \frac{VG}{GA} = \frac{VG}{GF} = \frac{AT}{TQ}$$

所以 $ST \parallel PQ$.

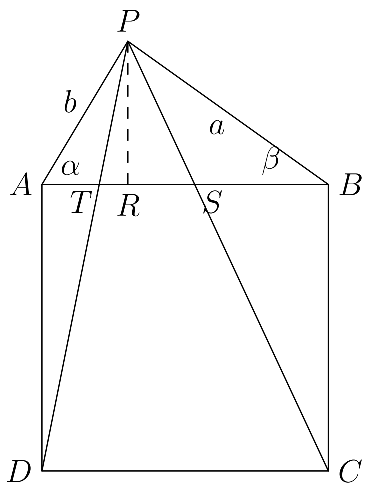

图 4.46

如果注意到图中包含着两个同样的基本图形，则可以有如下的证明，先提一个引理：如图4.46，设点 P 为正方形 $ABCD$ 的边 AB 外侧一点，PC 与 PD 分别与边 AB 相交于 S 和 T，记 $\angle PAB = \alpha, \angle PBA = \beta$，则有

$$AT : TS : SB = \cos\alpha\sin\beta : \sin\alpha\sin\beta : \sin\alpha\cos\beta$$

利用此引理知，$AT : TS$ 仅与角 α 有关，因此在原题中立刻便有 $\frac{AS}{SP} = \frac{AT}{TQ}$，因此两线平行，而这个引理的证明也很简单，记点 P 在边 AB 上的投影为 R，并假定正方形边长为 1，并记 $PB = a, PA = b$，根据正弦定理可得

$$\frac{a}{\sin\alpha} = \frac{1}{\sin\alpha+\beta} = \frac{b}{\sin\beta}$$

于是

$$a = \frac{\sin\alpha}{\sin\alpha+\beta}, b = \frac{\sin\beta}{\sin\alpha+\beta}$$

根据几何关系有

$$AR = b\cos\alpha = \frac{\cos\alpha\sin\beta}{\sin\alpha+\beta}, PR = b\sin\alpha = \frac{\sin\beta}{\sin\alpha+\beta}$$

而由 $\frac{TR}{AT}=\frac{PR}{AD}$ 得

$$\begin{aligned} AT &= \frac{AT}{AT+TR}AR=\frac{1}{1+\frac{TR}{AT}}AR=\frac{1}{1+\frac{\sin\alpha\sin\beta}{\sin\alpha+\beta}}\cdot\frac{\cos\alpha\sin\beta}{\sin\alpha+\beta} \\ &= \frac{\cos\alpha\sin\beta}{\sin\alpha\sin\beta+\sin\alpha+\beta} \end{aligned}$$

根据对称性，互换 α 和 β 可得

$$BS=\frac{\sin\alpha\cos\beta}{\sin\alpha\sin\beta+\sin\alpha+\beta}$$

于是

$$TS=\frac{\sin\alpha\sin\beta}{\sin\alpha\sin\beta+\sin\alpha+\beta}$$

于是引理得证。 □

题 4.8.4 [8] 如图 4.47，三角形 ABC 的内切圆 $\odot I$ 与三边分别相切于 D、E、F，线段 BE 与线段 CF 相交于点 P，直线 DE 与直线 BA 相交于点 M，直线 DF 与直线 CA 相交于点 N，求证: $PI\perp MN$.

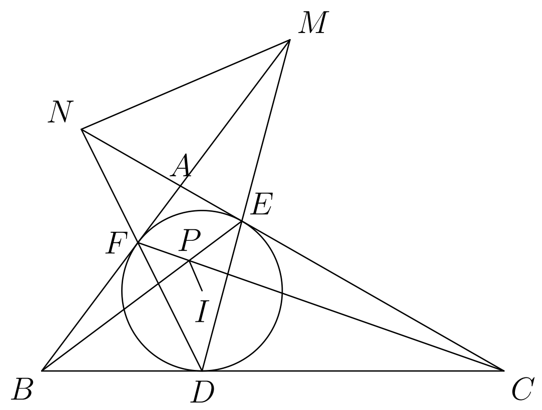

图 4.47

■

题 4.8.5 若 $\triangle ABC$ 中成立 $b+c=2a$，求证 $OI\perp AI$，其中 O 和 I 分别表外心和内心。 ■

证明一. 只需证明 $AO^2=OI^2+AI^2$，记外接圆半径和内切圆半径分别为 R 和 r，由几何关系及欧拉公式有

$$AO=R,\ OI^2=R^2-2Rr,\ AI=\frac{r}{\sin\frac{A}{2}}$$

以上三式代入勾股式后知只需证

$$\frac{r}{R}=2\sin^2\frac{A}{2}$$

[8] 2017 年北京大学中学生数学奖个人能力挑战赛试题.

而

$$R=\frac{a}{2\sin A}$$

和

$$r=\frac{1}{2}(b+c-a)\tan\frac{A}{2}=\frac{1}{2}a\tan\frac{A}{2}$$

由此二式便知要证的等式成立。 □

证明二. (魏子豪)

设角平分线 AI 交 BC 边于 E，交外接圆于另一点 D，则易证 $AI=2IE$，又 $DC:DE=BA:BE=2:1$ 以及 $DC=DI$ 知 $IE=ED$，于是 I 为 AD 中点，于是由垂径定理，$AI\perp AD$. □

参考文献

[1] [前苏联] 菲赫金哥尔茨. 微积分学教程. 高等教育出版社, 2006.

[2] 钱展望, 朱华伟. 奥林匹克数学. 湖北教育出版社, 2002.

[3] 陈传理, 张同君. 竞赛数学教程. 高等教育出版社, 2002.

[4] 北京大学数学系几何与代数教研室前代数小组. 高等代数. 高等教育出版社, 2004.

[5] 华东师范大学数学系, 数学分析. 高等教育出版社, 2004.

[6] 闵嗣鹤, 严士健. 初等数论. 高等教育出版社, 1983. 1983.

[7] kuing. kuing 网络撸题集. 网络电子书, 2015.

[8] [越南] 范建熊 (著), 隋振林 (译). 不等式的秘密 (第一卷). 哈尔滨工业大学出版社, 2012.

[9] [古希腊] 阿波罗尼奥斯. 圆锥曲线论. 陕西科学技术出版社, 2007.

[10] 沈文选, 杨清桃. 几何瑰宝. 哈尔滨工业大学出版社, 2010.