第 1 章　素　数　(1)

1.1　整　除　性

数

$$\cdots, -3, -2, -1, 0, 1, 2, \cdots$$

称为有理整数(rational integer), 或简称为整数(integer). 数

$$0, 1, 2, 3, \cdots$$

称为非负整数(non-negative integer). 数

$$1, 2, 3, \cdots$$

称为正整数(positive integer). 正整数构成算术的主要对象, 但它基本上常被视为整数或者某个更大范围内的数的一个子集.

以后我们用字母

$$a, b, \cdots, n, p, \cdots, x, y, \cdots$$

表示整数, 它们有时 (但并不总是如此) 会服从某些进一步的限制条件, 比如正数或非负数这样的限制. 我们也常用 “数” 来指代 “整数”(或表示 “正整数” 等), 在正文中的含义明确无误时, 我们考虑的就仅仅是这种特殊类型的数.

称一个整数 a 能被另一个整数 $b(b \neq 0)$整除(divisible), 假设存在第 3 个整数 c 使得

$$a = bc.$$

如果 a 和 b 都是正数, 则 c 必为正数. 用记号

$$b|a$$

来表示 a 被 b 整除, 或 b 是 a 的一个因子(divisor). 于是有

$$1|a, \quad a|a,$$

且对每个不为零的数 b 均有 $b|0$. 有时也用

$$b \nmid a$$

来表示与 $b|a$ 相反的含义. 显然有

$$b|a, c|b \to c|a,$$

$$b|a \to bc|ac \text{ (假设 } c \neq 0),$$

以及

$$c|a, c|b \to c|(ma + nb) \text{ (对任何整数 } m \text{ 和 } n).$$

1.2 素 数

在 1.2 节到 2.9 节中, 我们考虑的数一般都是正整数.[①] 正整数中有一个特别重要的子集, 即素数集合. 数 p 称为*素数*(prime), 如果

(i) $p>1$,

(ii) p 没有除了 1 和 p 以外的正因子.

例如, 37 是一个素数. 要特别注意 1 不算作素数, 这一点很重要. 在第 1 章以及第 2 章里, 我们始终用字母 p 表示素数.[②]

大于 1 且不是素数的数称为*合数*(composite).

下面, 我们引入第一个定理:

定理 1 *除了 1 以外的每个正整数都是素数的乘积.*

n 要么是素数 (此时不需要证明了), 要么 n 有大于 1 且小于 n 的因子. 设 m 是这些因子中最小的一个, 那么 m 必为素数, 否则,

$$\exists l, 1<l<m, \quad l|m,$$

则

$$l|m \to l|n,$$

这与 m 的定义矛盾.

因此, n 要么是素数, 要么可以被一个小于 n 的素数 (比方说 p_1) 整除. 在后一种情形中, 有

$$n=p_1n_1, \quad 1<n_1<n.$$

这里 n_1 要么是素数 (此种情形的证明已经完成), 要么可以被一个小于 n_1 的素数 p_2 整除, 此时有

$$n=p_1n_1=p_1p_2n_2, \quad 1<n_2<n_1<n.$$

重复这个方法, 得到一列递减的数 $n, n_1, \cdots, n_{k-1}, \cdots$, 它们全都大于 1, 对其中每个数, 都同样有以上两种可能性成立. 但迟早我们必定会接受第一种可能性, 此时得到的 n_{k-1} 已经是一个素数, 比如记之为 p_k, 这样就得到

$$n=p_1p_2\cdots p_k. \tag{1.2.1}$$

例如

$$666=2\times 3\times 3\times 37.$$

如果 $ab=n$, 那么 a 和 b 不可能都大于 $\sqrt{n}$. 于是任何合数 n 必可被一个不超过 $\sqrt{n}$ 的素数 p 整除.

---

① 偶尔也有例外, 如在 1.7 节中, e^x 是分析中的指数函数.

② 需要注意的是, 如果本书自始至终都严格遵守这个约定会很不方便, 因而有时也不坚持用它表示素数. 例如第 9 章用 p/q 表示典型的有理分数, 其中的 p 并不总是表示素数. 不过 p 是表示素数的 "自然的" 字母, 因此只要方便的话, 我们总用这个字母来表示素数.

(1.2.1) 式中的素数不一定是互不相同的, 也不一定非要按照某个特定的次序排列. 如果把它们按照递增的顺序排列, 把相同的素数合写成单一的因子, 并适当改变记号, 就得到

$$n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k} \ (a_1 > 0, a_2 > 0, \cdots, p_1 < p_2 < \cdots). \tag{1.2.2}$$

称 n 被表示成了标准型(standard form).

1.3 算术基本定理的表述

在定理 1 的证明中没有证明 (1.2.2) 式是 n 的唯一表示, 换句话说, 除了因子可以重新排列外, (1.2.1) 式是唯一的. 然而考虑几个特殊情形可以立即看出这是正确的.

定理 2(算术基本定理) *n 的标准型是唯一的. 也就是说, 除了因子可以重新排列以外, n 只能用唯一一种方式表示成素数的乘积.*

定理 2 是算术理论体系的基础, 但本章不会用到它, 关于它的证明将在 2.10 节给出. 但是, 证明它是下面较为简单的定理的一个推论还是很方便的.

定理 3(Euclid 第一定理) *如果 p 是素数, 且 $p|ab$, 那么 $p|a$ 或者 $p|b$.*

眼下先将此定理视为已经成立, 由它来推导出定理 2. 这样一来, 定理 2 的证明就简化为证明定理 3, 而定理 3 的证明在 2.10 节中给出.

显然,

$$p|abc\cdots l \to p|a \text{ 或者 } p|b \text{ 或者 } p|c\cdots \text{ 或者 } p|l$$

是定理 3 的一个推论. 特别地, 如果 $a, b, \cdots, l$ 都是素数, 那么 p 是 $a, b, \cdots, l$ 中的一个. 现在假设

$$n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k} = q_1^{b_1} q_2^{b_2} \cdots q_j^{b_j},$$

其中每个乘积都是标准型中的素数乘积. 从而对每个 i 都有 $p_i | q_1^{b_1} \cdots q_j^{b_j}$, 于是每个 p 都是某个 q. 类似地, 每个 q 都是某个 p. 所以有 $k = j$, 又由于这两个素数集合都是按照递增次序排列, 因此对每个 i 有 $p_i = q_i$.

如果 $a_i > b_i$, 用 $p_i^{b_i}$ 来除即得

$$p_1^{a_1} \cdots p_i^{a_i - b_i} \cdots p_k^{a_k} = p_1^{b_1} \cdots p_{i-1}^{b_{i-1}} p_{i+1}^{b_{i+1}} \cdots p_k^{b_k}.$$

左边可以被 p_i 整除, 然而右边则不能: 这是一对矛盾. 类似地, $b_i > a_i$ 也同样推出矛盾. 由此得出有 $a_i = b_i$. 这就完成了定理 2 的证明.

现在就会清楚为什么不把 1 作为素数. 因为如果把 1 作为素数的话, 定理 2 就不能成立, 这是因为此时可以插入任意多个 1 作为乘积因子.

1.4 素 数 序 列

最前面的几个素数是

$$2,3,5,7,11,13,17,19,23,29,31,37,41,43,47,53,\cdots.$$

通过 “Eratosthenes 筛法” 程序, 不难构造出某个界限 N 内的素数表来. 我们已经看到, 如果 $n \leqslant N$, 且 n 不是素数, 那么 n 必定被一个不大于 $\sqrt{N}$ 的素数整除. 记下数

$$2,\ 3,\ 4,\ 5,\ 6,\ \cdots,\ N,$$

相继划掉以下的数:

(i) 4, 6, 8, 10, $\cdots$, 即划掉 2^2 及其后的每个偶数;

(ii) 9, 15, 21, 27, $\cdots$, 即划掉 3^2 及其后的每个未被划掉的 3 的倍数;

(iii) 25, 35, 55, 65, $\cdots$, 即划掉 5^2(3 后面剩下的那个数的平方), 及其后的每个未被划掉的 5 的倍数; 继续此程序直到下一个剩下的数 (在它的倍数最终被删除之后) 大于 $\sqrt{N}$ 为止. 这样剩下的数均为素数. 目前所有的素数表都是通过对这个程序加以修改得到的.

素数表表明: 素数数列是无限的. 人们已经做出了 100 000 000 以内的素数表. 10 000 000 以内的素数共有 664 579 个, 介于 9 900 000 和 10 000 000 之间的素数有 6 134 个. 1 000 000 000 以内的素数总共有 50 847 478 个, 但是所有这些素数并不是每一个都知道. 已知一些很大的素数, 它们大多数是形如 2^p-1 的数 (参见 2.5 节). 迄今已发现的最大的素数超过了 6 500 位.

这些数据使人联想到下面的定理.

定理 4(Euclid 第二定理) *素数无限.*

2.1 节将证明这个定理.

素数的 “平均” 分布是很有规则的: 它的密度显现出稳定而缓慢地减少. 如果每 1 000 个数一组, 则前 5 组所含的素数个数分别为

$$168,\ 135,\ 127,\ 120,\ 119.$$

10 000 000 以内的最后 5 组所含的素数个数分别为

$$62,\ 58,\ 67,\ 64,\ 53.$$

把最后的 1 000 个数等分成 10 组, 则最后的 53 个素数也被分到了这 10 组中, 每组中分别含有

$$5,4,7,4,6,3,6,4,5,9$$

个素数.

另一方面, 素数分布从细节上来说仍是极不规则的.

首先, 素数表显示在区间里有很长的由合数组成的片段. 像素数 370 261 的后面就接连有 111 个合数. 容易看出, 这种由一长串合数组成的片段是一定会出现的. 假设

$$2,\ 3,\ 5,\ \cdots,\ p$$

是不超过 p 的所有素数, 那么所有不超过 p 的数都可以被这些素数中的某一个数整除, 这样一来, 如果

$$2 \times 3 \times 5 \times \cdots \times p = q,$$

则所有 $p-1$ 个数

$$q+2,\ q+3,\ q+4,\ \cdots,\ q+p$$

都是合数. 如果定理 4 为真, 那么 p 可以任意大, 因为如若不然, 则从某处开始往后所有的数均为合数.

定理 5 对任意给定的数 N, 都存在长度超过 N 的仅由连续合数组成的片段.

另一方面, 素数表指出存在像 3, 5 或者 101, 103 这样的始终相差 2 的、不确定的然而持续的素数对. 这样的素数对 $(p, p+2)$ 在 100 000 以下有 1 224 对, 而在 1 000 000 以下有 8 169 对. 如果仔细检查的话, 似乎有证据支持如下的猜想:

素数对 $(p, p+2)$有无穷多个.

的确还可以合理地给出更多的猜想. 数 $p, p+2, p+4$ 不可能全都是素数, 因为它们中必有一个能被 3 整除. 然而却没有显而易见的理由说明 $p, p+2, p+6$ 不能全是素数, 有证据表明这样的三元素数组也是可能持续出现的. 类似地, 三元数组 $(p, p+4, p+6)$ 似乎也可能持续出现. 于是可以猜想:

形如 $(p, p+2, p+6)$ 以及形如 $(p, p+4, p+6)$ 的三元素数组有无穷多个.

这种关于多个素数的集合的猜想还可以举出许多来, 但迄今为止, 无论是证明这些猜想还是否定它们都超出当今数学力所能及的范围之外.

1.5 关于素数的某些问题

对于像素数这样的数列, 提出什么问题是比较自然的呢? 我们已经给出过一些问题, 现在要再来问几个问题.

(1) 对于第 n 个素数 p_n, 是否有一般性的简单公式(这里的公式指的是, 可以对任何给定的 n 用它来计算 p_n 的值, 其计算量要小于用 Eratosthenes 筛法所需的计算量)? 现在还不知道有这样的公式, 且看起来不像有这样的公式存在.

另一方面, 有可能对 p_n 设计出若干个 “公式”. 这些公式中有一些不过是奇特的小玩意而已, 因为这些公式是用 p_n 来定义 p_n 自己, 而前面未知的 p_n 是不能用这些公式计算出来的. 在定理 419 中我们将给出一个例子. 其他一些公式在理论上能保证我们计算出 p_n, 但其代价是所用的计算量比用 Eratosthenes 筛法的计算量要多得多. 还有另外一些公式本质上与 Eratosthenes 筛法等价. 我们将在 2.7 节以及附录 1 和附录 2 中回答这些问题.

类似的注解对于另一个同类问题也一样适用, 此问题即

(2) 从一个给定的素数得到下一个素数是否存在一般性的简单公式 (即像 $p_{n+1} = p_n^2 + 2$ 这样的递推公式)?

另一个自然的问题是:

(3) 有没有这样一个法则存在, 使得对于任何给定的素数 p, 都可以得到一个更大的素数 q?

当然这个问题预先假设了素数个数无穷 (即定理 4). 如果已知有一个简单的函数 $f(n)$, 它能对所有的整数值 n 均取素数值, 那么这个问题就可以给出肯定的回答. 除了已经提到的那种没什么意思的奇特的小玩意而外, 并不知道有这样的函数存在. 关于这种函数的形式, 仅有的合乎情理的猜想是由 Fermat 给出的,[①] 然而 Fermat 的猜想是错误的.

下一个问题是:

(4) 小于一个给定的数 x 的素数有多少个?

这个问题是一个有用得多的问题, 不过需要加以仔细的解释. 像通常那样, 假设我们定义 $\pi(x)$ 是不超过 x 的素数个数, 于是有 $\pi(1)=0, \pi(2)=1, \pi(20)=8$. 如果 p_n 表示第 n 个素数, 那么就有 $\pi(p_n)=n$, 从而 $\pi(x)$(作为 x 的函数) 和 p_n(作为 n 的函数) 是一对反函数. 为寻求 $\pi(x)$ 的任何形式简单的精确公式, 实际上就是重复提出问题 (1).

这样一来, 我们必须换一种方式来解释这个问题. 我们要问 "大约有多少个素数 ……?" 究竟是大多数的数都是素数呢, 还是只有一小部分数是素数呢? 是否存在一个简单的函数 $f(x)$, 它是 $\pi(x)$ 的 "一个好的度量" 呢?

1.8 节以及第 22 章将回答这些问题.

1.6 若 干 记 号

我们将经常使用符号

$$O,\ o,\ \sim, \tag{1.6.1}$$

偶尔会使用符号

$$\prec,\ \succ,\ \asymp. \tag{1.6.2}$$

这些符号定义如下.

设 n 是一个趋向于无穷的整数变量, x 是一个趋向于无穷或趋向于零或趋向于某个另外的极限值的连续变量. $\phi(n)$ 和 $\phi(x)$ 是 n 和 x 的正值函数, $f(n)$ 和 $f(x)$ 是 n 和 x 的任何其他的函数. 那么

(i) $f=O(\phi)$ 表示[②]

$$|f|<A\phi,$$

其中 A 与 n 或者 x 无关 (对问题中涉及的 n 或者 x 的所有的值而言);

(ii) $f=o(\phi)$ 表示 $f/\phi\to 0$;

(iii) $f\sim\phi$ 表示 $f/\phi\to 1$.

于是当 $x\to\infty$ 时有

① 见 2.5 节.

② 如通常在分析中那样, $|f|$ 表示 f 的模或者绝对值.

$$10x = O(x), \quad \sin x = O(1), \quad x = O(x^2),$$
$$x = o(x^2), \quad \sin x = o(x), \quad x + 1 \sim x.$$

而当 $x \to \infty$ 时有

$$x^2 = O(x), \quad x^2 = o(x), \quad \sin x \sim x, \ 1 + x \sim 1.$$

要注意的是 $f = o(\phi)$ 蕴含且强于 $f = O(\phi)$.

关于符号 (1.6.2), 有

(iv) $f \prec \phi$ 表示 $f/\phi \to 0$, 它等价于 $f = o(\phi)$;

(v) $f \succ \phi$ 表示 $f/\phi \to \infty$;

(vi) $f \asymp \phi$ 表示 $A\phi < f < A\phi$,

其中的两个 A (它们自然不相同) 都是正的且与 n 或者 x 无关. 于是 $f \asymp \phi$ 断言 "f 与 ϕ 的大小同阶".

我们常会像在 (vi) 中那样用 A 作为*未明确给出的正的常数*. 不同的 A 通常有不同的值, 即便是当它们出现在同一个公式中时亦如此. 此外, 即便是可以给它们指定确定的值, 这些数值也与讨论无关.

到目前为止, 已经定义了如 "$f = O(1)$", 但没有单独定义 "$O(1)$". 而让记号更为灵活是非常方便的. 约定 "$O(\phi)$" 表示*一个未指定的函数*f, 它满足 $f = O(\phi)$. 例如, 可以写出

$$O(1) + O(1) = O(1) = o(x) \quad (x \to \infty),$$

它的含义是: "如果 $f = O(1)$ 且 $g = O(1)$, 那么就有 $f + g = O(1)$, 当然更有 $f + g = o(x)$". 或者我们还可以写出

$$\sum_{\nu=1}^{n} O(1) = O(n),$$

它的含义是: 每项都小于一个常数的 n 个项的和也小于 n 的一个常数倍.

注意到介于符号 O 和 o 之间的关系 "=" 通常并不是对称的. 比如 $o(1) = O(1)$ 总是正确的, 然而 $O(1) = o(1)$ 通常是错误的. 还要注意 $f \sim \phi$ 等价于 $f = \phi + o(\phi)$, 或者等价于

$$f = \phi\{1 + o(1)\}.$$

此时就说 f 和 ϕ*是渐近等价的*(asymptotically equivalent), 或者说成 f*渐近于*ϕ.

还有另外一个术语在此定义比较方便. 假设 P 是正整数的一个可能具有的性质, 而 $P(x)$ 是小于 x 的数中有此性质的数的个数. 如果当 $x \to \infty$ 时有

$$P(x) \sim x,$$

也就是说, 如果小于 x 的数中不具有此性质的数的个数是 $o(x)$, 那么就说*几乎所有的数*(almost all numbers) 都具有这个性质. 于是, 将有[①] $\pi(x) = o(x)$, 从而几乎所有的数都是合数.

① 可由定理 7 立即得出.

1.7 对 数 函 数

素数分布的理论要求了解对数函数 $\ln x$ 的性质. 假定读者了解对数和指数的通常解析理论, 但这里要着重强调 $\ln x$ 的一个性质[①].

由于当 $x \to \infty$ 时,

$$\mathrm{e}^x = 1 + x + \cdots + \frac{x^n}{n!} + \frac{x^{n+1}}{(n+1)!} + \cdots,$$

从而

$$x^{-n}\mathrm{e}^x > \frac{x}{(n+1)!} \to \infty \quad (x \to \infty).$$

于是 e^x 与 x 的任何幂次相比, 前者趋向于无穷大的速度要快得多. 由此推出, 其反函数$\ln x$ 与 x 的任何正的幂次相比, 前者趋向于无穷大的速度要慢得多. 此时虽然 $\ln x \to \infty$, 然而对每个正数 δ 有

$$\frac{\ln x}{x^\delta} \to 0, \tag{1.7.1}$$

或者说 $\ln x = o(x^\delta)$. 类似地, $\ln\ln x$ 与 $\ln x$ 的任何幂次相比, 前者趋向于无穷大的速度要慢得多.

可以对 $\ln x$ 增长的缓慢性给出一个数值的例证. 如果 $x = 10^9 = 1\ 000\ 000\ 000$, 则有

$$\ln x = 20.72\cdots.$$

由于 $\mathrm{e}^3 = 20.08\cdots$, 故而 $\ln\ln x$ 比 3 稍大一点, 而 $\ln\ln\ln x$ 比 1 略大一点. 如果 $x = 10^{1\ 000}$, 则 $\ln\ln\ln x$ 比 2 要大一点. 尽管如此, $\ln\ln\ln x$ 的无穷大的阶还是在素数论中有它的作用.

函数

$$\frac{x}{\ln x}$$

在素数论中特别重要. 它比 x 趋向于无穷要慢得多. 但鉴于 (1.7.1), 它比 $x^{1-\delta}$ 趋向于无穷要快得多, 也就是说, 它比 x 的任何小于 1 次的幂趋向于无穷要快得多. 而且它是具有这个性质的最简单的函数.

1.8 素数定理的表述

在前面的绪论之后, 本节来叙述一个定理, 它回答了 1.5 节中的问题 (4).

定理 6(素数定理) 不超过 x 的素数个数渐近于 $\dfrac{x}{\ln x}$, 即 $\pi(x) \sim \dfrac{x}{\ln x}$.

这个定理是素数分布理论的核心定理. 第 22 章将给出它的证明. 这个证明并不容易, 不过在同一章里会对下面的较弱的结果给出一个简单得多的证明:

① 当然了, $\ln x$ 是指以 e 为底的自然对数, "一般的" 对数并没有什么数学意义.

定理 7(Tchebychef 定理)

$$\pi(x) \text{ 的阶是 } \frac{x}{\ln x}, \quad \text{即} \pi(x) \asymp \frac{x}{\ln x}.$$

将定理 6 和素数表中的数值进行比较是件很有趣的事情. 对于 $x = 10^3, x = 10^6$ 以及 $x = 10^9$, $\pi(x)$ 的值分别是

$$168, \quad 78\,498, \quad 50\,847\,478;$$

而 $\frac{x}{\ln x}$ 的值 (取离它最接近的整数) 分别是

$$145, \quad 72\,382, \quad 48\,254\,942.$$

它们对应的比值分别是

$$1.159\cdots, \quad 1.084\cdots, \quad 1.053\cdots.$$

尽管这些比值并不是非常快地逼近 1, 但这些数值给出了某种近似. 实际值多于估计值, 可用一般理论给出解释.

如果

$$y = \frac{x}{\ln x},$$

那么

$$\ln y = \ln x - \ln\ln x,$$

由于

$$\ln\ln x = o(\ln x),$$

故有

$$\ln y \sim \ln x, \quad x = y\ln x \sim y\ln y.$$

于是 $\frac{x}{\ln x}$ 的反函数渐近于 $x\ln x$.

由此可以推知, 定理 6 等价于

定理 8 $p_n \sim n\ln n.$

类似地, 定理 7 等价于

定理 9 $p_n \asymp n\ln n.$

第 664 999 个素数是 10 006 721, 读者可以将这些数字与定理 8 进行比较.

我们把要讲的有关素数及其分布的内容安排在第 1 章、第 2 章以及第 22 章这三章里. 本章作为导引, 除了定义和初步的说明之外, 几乎没有什么内容. 除了较容易证明的定理 1(但它也很重要) 以外, 我们没有证明其他什么结论. 第 2 章要证明得更多一些: 特别是 Euclid 的定理 3 和定理 4. 其中定理 3 可以推导出被称为 "基本定理" 的定理 2(见 1.3 节), 我们以后几乎所有的工作都依赖于这个基本定理, 2.10 节和 2.11 节将对它给出两个证明. 2.1 节、2.4 节和 2.6 节要用几种方法来证明定理 4, 其中有的方法可以使这个定理略加扩展. 第 22 章将再次回到素数分布理论, 并尽可能地用初等方法展开这个理论, 在我们要讨论的结果中, 包含证明定理 7, 最后还要证明定理 6.

本 章 附 注

1.3 节. 定理 3 是 Euclid《几何原本》第 7 卷命题 30. 定理 2 看起来在 Gauss 之前还没有被人明确地叙述过 (见 Gauss *D.A.*, 第 16 章). 当然, 早期的数学家是知道这个结果的, 不过 Gauss 是将算术发展成为一门系统科学的第一人. 参见本书 12.5 节.

1.4 节. 最好的因子表是 D. N. Lehmer 的 *Factor Table for the first ten millions* [Carnegie Institution, Washington 105(1909)], 它给出了不超过 10 017 000 且不能被 2, 3, 5 以及 7 整除的所有的数的最小因子. 同一作者的 *List of prime numbers from* 1 *to* 10 006 721(Carnegie Institution, Washington 165(1914)) 被 Baker 和 Gruenberger 扩展到了 10^8(*The first six million prime numbers*, Rand Corp., Microcard Found., Madison 1959). 有关更早期的表的信息可以在 Lehmer 的两卷本著作的引言以及 Dickson 的数论史第 1 卷第 13 章中找到. 我们给出的素数个数比 Lehmer 给出的要少一个, 是因为他把 1 当作素数来处理. Mapes(*Math. Computation* 17(1963), 184–185) 给出 $\pi(x)$ 的一张表, 其中的 x 取值从 10 000 000 的任何倍数直到 1 000 000 000.

在 D. H. Lehmer 的 *Guide to tables in the theory of numbers*(Washington, 1941) 中给出了一张带有客观表述性注记的素数表.

定理 4 是 Euclid《几何原本》第 9 卷命题 20.

关于定理 5, 请参见 Lucas 的 *Théorie des nombres*, i(1891), 359–361.

Kratchick [*Sphinx*, 6(1936), 166 以及 8(1938), 86] 列出了 $10^{12}-10^4$ 和 $10^{12}+10^4$ 之间的所有素数; 而 Jones, Lal 和 Blundon(*Math. Comp.* 21(1967), 103-107) 则列出了从 10^k 到 $10^k+150\,000$ 之间的所有素数 (对于从 8 到 15 的整数 k). 已知最大的素数对 $p, p+2$ 是

$$2\,003\,663\,613\times 2^{195\,000}\pm 1,$$

它是由 Vautier 于 2007 年发现的. 这些素数有 58 711 位数字.

在 22.20 节中我们对不超过 x 的素数对 $(p, p+2)$ 的个数所猜想的公式给出一个简单的讨论. 它和已知的事实很吻合. 这个方法可以用来寻求关于素数对、三素数组以及更大的素数组的许多其他猜想的定理.

1.5 节. 我们这里的问题列表是对 Carmichael 在 *Theory of numbers*, 29 中给出的问题表经过修改得到的. 当然, 在这里我们没有 (也不可能) 定义 "简单的公式" 其含义是什么. 寻求计算第 n 个素数的算法可能更有裨益. 显然有一个算法, 即由 Eratosthenes 给出的筛法. 于是, 一个有意思的问题就是: 这样一个算法能计算多快? 一种以 Lagarias 以及 Odlyzko 的工作为基础的方法 [*J. Algorithms* **8**(1987), 173–191] 在 $O(n^{3/5})$ 的时间内算出 p_n(如果有大量的存储器可用, 或许的确还可以再略微快一些). 对于问题 (2) 和 (3), 可以类似地问: 给定 p_n, 可以以多快的速度算出 p_{n+1}? 或者更一般地, 可以以多快的速度求出大于一个给定素数 p 的任何素数? 目前看来, 最好的方法还仅仅是从 p_n 开始往上检验每个数的素性. 有人或许会猜想这个过程会是极其有效的, 在大致 $O((\lg n)^C)$($C>0$ 是一个常数) 这样的时间内即可找到下一个素数. 我们有一种属于 Agrawal, Kayal 以及 Saxena[*Ann. of Math.* (2) **160**(2004), 781–793] 的非常快速的素性判别法, 但是关于差 $p_{n+1}-p_n$ 已知最著名的上界是 $O(p_n^{0.525})$ (见 Baker, Harman 以及 Pintz, *Proc. London Math. Soc.* (3) **83**(2001), 532–562). 因此, 目前我们只能说: 在给定 p_n 之后, p_{n+1} 可以在 $O(p_n^{\theta})$ 时间内确定出来 (对任意常数 $\theta>0.525$).

1.7 节. Littlewood 有关 $\pi(x)$ 有时比 “对数积分”$\operatorname{li} x$ 稍大的证明依赖于当 x 相当大时 $\log\log\log x$ 的大小. 见 Ingham 的书的第 5 章, 或者参看 Landau, *Vorlesungen*, ii, 123–156.

1.8 节. 定理 7 是由 Tchebychef 在大约 1850 年证明的, 而定理 6 是由 Hadamard 和 de la Vallée Poussin 在 1896 年证明的. 见 Ingham 的书, 4-5; Landau, *Handbuch*, 3-55; 以及本书第 22 章, 特别是 22.14 节至 22.16 节的附记.

$\pi(x)$ 的一个更好的近似值由 “对数积分”

$$\operatorname{Li} x = \int_2^x \frac{\mathrm{d}t}{\ln t}$$

给出. 例如, 对于 $x = 10^9$, $\pi(x)$ 和 $x/\ln x$ 相差大于 2 500 000, 而 $\pi(x)$ 与 $\operatorname{Li} x$ 仅相差大约 1 700.

第 2 章　素　数　(2)

2.1　Euclid 第二定理的第一个证明

Euclid 自己对定理 4 给出的证明如下.

设 $2,3,5,\cdots,p$ 是不大于 p 的所有素数组成的集合, 并令

$$q=2\times 3\times 5\times\ \cdots\ \times p+1, \tag{2.1.1}$$

则 q 不能被 $2,3,5,\cdots,p$ 中任何一个数整除. 于是 q 要么是一个素数, 要么可以被介于 p 和 q 之间的某个素数整除. 无论哪一种情形都会有一个大于 p 的素数存在, 这就证明了该定理.

该定理等价于

$$\pi(x)\to\infty. \tag{2.1.2}$$

2.2　Euclid 方法的更进一步的推论

如果 p 是第 n 个素数 p_n, q 的定义与 (2.1.1) 式中的相同, 那么显然, 对 $n>1$[①]有

$$q<p_n^n+1,$$

从而有

$$p_{n+1}<p_n^n+1.$$

这个不等式使我们能对 p_n 的增长速率给出一个上限, 并对 $\pi(x)$ 的增长速率给出一个下限.

然而, 我们可以得到如下更好的界限. 假设对 $n=1,2,\cdots,N$ 有

$$p_n<2^{2^n}, \tag{2.2.1}$$

那么 Euclid 方法就给出

$$p_{N+1}\leqslant p_1p_2\cdots p_N+1<2^{2+4+\cdots+2^N}+1<2^{2^{N+1}}. \tag{2.2.2}$$

由于 (2.2.1) 对 $n=1$ 为真, 从而它对所有 n 也为真.

现在假设 $n\geqslant 4$, 且

$$\mathrm{e}^{\mathrm{e}^{n-1}}<x\leqslant\mathrm{e}^{\mathrm{e}^n},$$

那么就有[②]

$$\mathrm{e}^{n-1}>2^n,\quad \mathrm{e}^{\mathrm{e}^{n-1}}>2^{2^n}.$$

于是, 根据 (2.2.1) 式就有

$$\pi(x)\geqslant\pi(\mathrm{e}^{\mathrm{e}^{n-1}})\geqslant\pi(2^{2^n})\geqslant n.$$

---

① 当 $n=1,p=2,q=3$ 时, 左右两式相等.

② 它对 $n=3$ 并不成立.

由 $\ln\ln x \leqslant n$ 可推出：对 $x > \mathrm{e}^{\mathrm{e}^3}$ 有

$$\pi(x) \geqslant \ln\ln x.$$

显然此不等式对 $2 \leqslant x \leqslant \mathrm{e}^{\mathrm{e}^3}$ 也成立, 于是就证明了：

定理 10 $\pi(x) \geqslant \ln\ln x \quad (x \geqslant 2)$.

这样就超越了定理 4, 得到了 $\pi(x)$ 的阶的一个下限. 当然这个下限太小, 因而不大合理. 例如, 根据此不等式它对 $x = 10^9$ 才给出 $\pi(x) \geqslant 3$, 而此时实际上 $\pi(x)$ 的值已超过 50 000 000 了.

2.3 某种算术级数中的素数

Euclid 方法还可以沿另外的方向发展.

定理 11 *存在无穷多个形如 $4n+3$ 的素数.*

我们不用 (2.1.1) 式, 而改用

$$q = 2^2 \times 3 \times 5 \times \cdots \times p - 1$$

来定义数 q, 那么 q 就是形如 $4n+3$ 的数, 且它不能被不超过 p 的任何素数整除. 它也不可能仅仅是形如 $4n+1$ 这样的素数的乘积, 这是因为两个形如 $4n+1$ 的数的乘积仍然是一个形如 $4n+1$ 的数. 于是它一定能被一个大于 p 且形如 $4n+3$ 的素数整除.

定理 12 *存在无穷多个形如 $6n+5$ 的素数.*

证明是类似的. 用

$$q = 2 \times 3 \times 5 \times \cdots \times p - 1$$

来定义 q, 并且注意到, 除了 2 和 3 以外的任何素数都形如 $6n+1$ 或者形如 $6n+5$, 且两个形如 $6n+1$ 的数的乘积仍是一个形如 $6n+1$ 的数.

证明形如 $4n+1$ 的素数的无穷性要更困难一些. 我们需要假设后面 (20.3 节) 要证明的一个定理的真实性.

定理 13 *如果 a 和 b 没有公约数, 那么 a^2+b^2 的任何奇素因子都必定形如 $4n+1$.*

如果事先假设这个定理成立, 就能证明存在无穷多个形如 $4n+1$ 的素数. 事实上可以证明

定理 14 *存在无穷多个形如 $8n+5$ 的素数.*

取

$$q = 3^2 \times 5^2 \times 7^2 \times \cdots \times p^2 + 2^2,$$

这是两个没有公约数的平方数之和. 奇数 $2m+1$ 的平方是

$$4m(m+1)+1,$$

这是一个形如 $8n+1$ 的数, 故而 q 是一个形如 $8n+5$ 的数. 根据定理 13, q 的任何素因子均形如 $4n+1$, 也即均形如 $8n+1$ 或者 $8n+5$, 而形如 $8n+1$ 的两个数的乘积仍然是一个形如 $8n+1$ 的数, 这样就可以和以前一样完成证明了.

所有这些定理都是著名的 Dirichlet 定理的特殊情形.

定理 15*(Dirichlet 定理)[①] *如果 a 是一个正数, 且 a 和 b 没有除了 1 以外的公约数, 那么就有无穷多个形如 $an+b$ 的素数存在.*

这个定理的证明过于困难, 不适合放在本书中. 而当 b 等于 1 或 -1 时则有较为简单的证明.

2.4 Euclid 定理的第二个证明

定理 4 的第二个证明 (该证明由 Pólya 给出) 依赖于所谓的 "Fermat 数" 的一个性质.

Fermat 数定义为

$$F_n = 2^{2^n}+1,$$

于是有 $F_1=5$, $F_2=17$, $F_3=257$, $F_4=65\ 537$.

Fermat 数在很多方面都令人感兴趣: 比方说, Gauss 曾经证明过[②], 如果 F_n 是一个素数 p, 那么边数为 p 的正多边形可以用 Euclid 的方法内切到一个圆的内部[③].

与这里的问题有关的 Fermat 数的性质如下.

定理 16 *任何两个 Fermat 数都没有大于 1 的公约数.*

假设 F_n 和 $F_{n+k}(k>0)$ 是两个 Fermat 数, 且

$$m|F_n, \quad m|F_{n+k}$$

如果 $x=2^{2^n}$, 就有

$$\frac{F_{n+k}-2}{F_n}=\frac{2^{2^{n+k}}-1}{2^{2^n}+1}=\frac{x^{2^k}-1}{x+1}=x^{2^k-1}-x^{2^k-2}+\cdots-1,$$

从而有 $F_n|F_{n+k}-2$. 于是就有

$$m|F_{n+k}, \quad m|(F_{n+k}-2),$$

这就给出 $m|2$. 但由于 F_n 是奇数, 从而 $m=1$, 这就证明了定理.

由此推出, $F_1, F_2, \cdots, F_n$ 中的每一个数都能被一个奇素数整除, 且整除其中某一个数的奇素数必不能整除这组数中其他任何一个数. 这样就至少有 n 个不超过 F_n 的奇素数存在, 而这也就证明了 Euclid 的定理. 我们还有

① 定理序号附有一个星号表示本书并不给出这个定理的证明.

② 见 5.8 节.

③ 这个结果可以等价表述为, 如果 F_n 是一个素数 p, 那么边数为 p 的正多边形可以用圆规与直尺作出. —— 译者注

$$p_{n+1} \leqslant F_n = 2^{2^n} + 1,$$

显然, 由这个不等式 [它比 (2.2.1) 式要稍强一点] 可以导出定理 10 的一个证明.

2.5 Fermat 数和 Mersenne 数

前 4 个 Fermat 数都是素数, Fermat 曾猜想所有的 Fermat 数都是素数. 然而, Euler 在 1732 年发现

$$F_5 = 2^{2^5} + 1 = 641 \times 6\ 700\ 417$$

是合数. 因为 $641 = 5^4 + 2^4 = 5 \times 2^7 + 1$ 既整除 $5^4 \times 2^{28} + 2^{32}$ 又整除 $5^4 \times 2^{28} - 1$, 从而它也整除这两个数的差 F_5.

1880 年 Landry 证明了

$$F_6 = 2^{2^6} + 1 = 274\ 177 \times 67\ 280\ 421\ 310\ 721.$$

最近有数学工作者证明了对于

$$7 \leqslant n \leqslant 16, \quad n = 18, 19, 21, 23, 36, 38, 39, 55, 63, 73$$

以及 n 的许多更大的值, F_n 都是合数. F_{14} 尚无已知的因子, 而对于所有其余已证明了是合数的 Fermat 数都有一个因子是已知的.

在 F_4 之后没有发现过取素数值的 F_n, 于是 Fermat 猜想一直未能被证明是一个成功的猜想. 很有可能取素数值的 F_n 的个数是有限的.[①]如果事实确实如此, 那么取素数值的 $2^n + 1$ 就是有限的, 这是因为容易证明下面的定理.

定理 17 *如果 $a \geqslant 2$ 且 $a^n + 1$ 是素数, 那么 a 必为偶数且 $n = 2^m$.*

因为如果 a 是奇数的话, $a^n + 1$ 就是偶数. 又如果 n 有一个奇数因子 k, 且 $n = kl$, 那么 $a^n + 1$ 可以被 $a^l + 1$ 整除:

$$\frac{a^{kl} + 1}{a^l + 1} = a^{(k-1)l} - a^{(k-2)l} + \cdots + 1.$$

将 Fermat 猜想和另一个著名猜想的命运加以比较是很有意思的, 这个猜想说的是形如 $2^n - 1$ 的素数. 我们首先给出另一个与定理 17 几乎同一类型的平凡定理.

定理 18 *如果 $n > 1$ 且 $a^n - 1$ 是素数, 那么 $a = 2$ 且 n 为素数.*

① 这是由概率的考虑提供的结果. 假设定理 7 成立, 可以粗略地讨论如下: 一个数 n 为素数的概率至多是

$$\frac{A}{\ln n},$$

于是 Fermat 素数的总的期望值至多为

$$A \sum \left\{ \frac{1}{\ln(2^{2^n} + 1)} \right\} < A \sum 2^{-n} < A.$$

这个讨论 (除了缺乏严格性以外) 假设了不存在特殊的理由使得某个 Fermat 数像是素数, 而定理 16 和定理 17 使我们想到这种数中有一些是素数.

因为如果 $a>2$, 那么就有 $(a-1)|(a^n-1)$. 又如果 $a=2$ 且 $n=kl$, 那么就有 $(2^k-1)|(2^n-1)$.

这样一来, 判断 a^n-1 是否素数的问题就归结为判断 2^p-1 是否素数. 1644 年 Mersenne 曾断言: 对

$$p=2,3,5,7,13,17,19,31,67,127,257,$$

$M_p=2^p-1$ 都是素数, 且对另外的 44 个小于 257 的 p 的值, M_p 都是合数. Mersenne 结论中的第一个错误是在大约 1886 年被发现的[①], 那一年 Pervusin 和 Seelhoff 发现了 M_{61} 是素数. 其后在 Mersenne 的结论中又发现了 4 个错误, 因而对他的结论不再需要认真对待了. 1876 年 Lucas 发现了一个方法来测试 M_p 是否素数, 并用此方法证明了 M_{127} 是素数. 这个数直到 1951 年都仍然是已知最大的素数, 而在 1951 年 Ferrier 用不同的方法发现了一个更大的素数 (仅用到一台台式计算机), 而 Miller 和 Wheeler(他们用到剑桥的电子计算机 EDSAC 1) 则发现了若干个大素数, 其中最大的一个是

$$180M_{127}^2+1,$$

这个数大于 Ferrier 得到的那个数. 但是 Lucas 的判别法特别适用于在二进制的数值计算机上使用. 后来又在 (Lehmer 和 Robinson, Hurwitz 和 Selfridge, Riesel, Gillies, Tuckerman 以及最后是 Nickel 和 Noll 等人的) 一系列的研究中得到了应用. 现在已知 M_p 对

$$\begin{aligned}p=&2,3,5,7,13,17,19,31,61,89,107,127,521,607,1\ 279,2\ 203,2\ 281,\\&3\ 217,4\ 253,4\ 423,9\ 689,9\ 941,11\ 213,19\ 937,21\ 701\end{aligned}$$

皆为素数, 而对 $p<21\ 700$ 中所有其余的 p, M_p 均为合数. 最大已知的素数是 $M_{21\ 701}$, 它是一个 6 533 位的数.

15.5 节将描述 Lucas 的判别法, 并给出一个 Miller 和 Wheeler 在定理 101 中所用的判别法.

Mersenne 数的问题与 "完全数" 问题有关, 16.8 节中会考虑完全数问题.

我们还会在 6.15 节和 15.5 节中再次回到这个论题.

2.6 Euclid 定理的第三个证明

假设 $2,3,\cdots,p_j$ 是前 j 个素数, 令 $N(x)$ 是不超过 x 且不能被任何素数 $p>p_j$ 整除的数 n 的个数. 如果把这样的 n 表成形式

$$n=n_1^2m,$$

其中 m 是 "无平方因子数", 即它不能被任何素数的平方整除, 这样就有

$$m=2^{b_1}3^{b_2}\cdots p_j^{b_j},$$

① 1732 年 Euler 说过 M_{41} 和 M_{47} 都是素数, 但这是错误的.

其中每个 b 的取值或者为 0 或者为 1. m 的指数恰有 2^j 种可能的选择, 于是 m 有不多于 2^j 个不同的值. 此外, $n_1 \leqslant \sqrt{n} \leqslant \sqrt{x}$, 从而 n_1 有不多于 $\sqrt{x}$ 个不同的值. 故有

$$N(x) \leqslant 2^j\sqrt{x}. \tag{2.6.1}$$

如果定理 4 不真, 那么素数个数就是有限的, 设所有素数为 $2, 3, \cdots, p_j$. 此时对每个 x 有 $N(x) = x$, 因此

$$x \leqslant 2^j\sqrt{x}, \quad x \leqslant 2^{2j},$$

而这对 $x \geqslant 2^{2j} + 1$ 是错误的.

可以用这个方法来证明两个进一步的结果.

定理 19 级数

$$\sum \frac{1}{p} = \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \cdots \tag{2.6.2}$$

是发散的.

如果该级数收敛, 可以选取 j 使得第 j 项以后的余项小于 $\frac{1}{2}$, 也就是说

$$\frac{1}{p_{j+1}} + \frac{1}{p_{j+2}} + \cdots < \frac{1}{2}.$$

满足 $n \leqslant x$ 且能被 p 整除的数 n 的个数至多为 x/p. 因此 $x - N(x)$(它是满足 $n \leqslant x$ 且能被 $p_{j+1}, p_{j+2}, \cdots$ 中一个或多个数整除的数 n 的个数) 不多于

$$\frac{x}{p_{j+1}} + \frac{x}{p_{j+2}} + \cdots < \frac{1}{2}x.$$

于是, 根据 (2.6.1) 式就有

$$\frac{1}{2}x < N(x) \leqslant 2^j\sqrt{x}, \quad x < 2^{2j+2},$$

这对 $x \geqslant 2^{2j+2}$ 是错误的. 从而该级数发散.

定理 20 $\pi(x) \geqslant \dfrac{\ln x}{2\ln 2}(x \geqslant 1)$, $p_n \leqslant 4^n$.

取 $j = \pi(x)$, 于是 $p_{j+1} > x$, $N(x) = x$. 从而

$$x = N(x) \leqslant 2^{\pi(x)}\sqrt{x}, \quad 2^{\pi(x)} \geqslant \sqrt{x},$$

取对数就得到定理 20 的第一部分. 如果令 $x = p_n$, 则有 $\pi(x) = n$, 定理第二部分结论立即得出.

根据定理 20 有 $\pi(10^9) \geqslant 15$, 这仍然是一个远低于实际结果的数.

2.7 关于素数公式的进一步结果

暂时回到 1.5 节中提出的问题. 可以寻求各种意义下的 “素数公式”.

(i) 可以寻找一个简单函数 $f(n)$, 使它取所有的素数值且仅取素数值. 也就是说, 当 n 取值为 $1, 2, \cdots$ 时, 该函数连续取素数值 $p_1, p_2, \cdots$. 这是 1.5 节中讨论过

的问题.

(ii) 可以寻找 n 的一个简单函数, 它只取素数值. Fermat 的猜想如果正确的话, 那就会给出此问题的一个答案[①]. 而现在的情况是还不知道是否会有令人满意的答案. 但是有可能构造出一个 (多个正整数变量的) 多项式, 尽管这个多项式所取的负值是合数, 但它所取的正值全都是素数且包含了所有的素数. 见附录 2.

(iii) 可以适当降低要求, 仅仅来求 n 的一个简单函数, 它取无穷多个素数值. 由 Euclid 定理得知, $f(n)=n$ 就是这样一个函数, 关于这个问题的不太显然的答案由定理 11 至定理 15 给出. 除了平凡的解之外, Dirichlet 定理 15 是已知的仅有解答. 迄今尚未能证明 n^2+1 或者 n 的任何一个另外的二次式能表示出无穷多个素数, 所有这样的问题看起来都极其困难.

有一些简单否定的定理, 它们包含了对于问题 (ii) 的很不完全的回答.

定理 21 *不存在任何非常数的整系数多项式 $f(n)$, 它能对所有 n, 或者对所有充分大的 n 都取素数值.*

可以假设 $f(n)$ 的首项系数是正的, 于是当 $n\to\infty$ 时就有 $f(n)\to\infty$, 且比方说对 $n>N$ 还有 $f(n)>1$ 成立. 如果 $x>N$ 且

$$f(x)=a_0x^k+\cdots=y>1,$$

那么, 对每个整数 r,

$$f(ry+x)=a_0(ry+x)^k+\cdots$$

都能被 y 整除, 并且当 r 趋向于无穷时 $f(ry+x)$ 也趋于无穷. 从而 $f(n)$ 可以取到无穷多个合数值.

有这样的二次式存在, 它对 n 的一列相当长的值都取素数值. 例如 n^2-n+41 对于 $0\leqslant n\leqslant 40$ 都取素数值, 而

$$n^2-79n+1\,601=(n-40)^2+(n-40)+41$$

则对 $0\leqslant n\leqslant 79$ 都取素数值.

一个更为一般的定理 (6.4 节中将证明它) 是

定理 22 *如果*

$$f(n)=P(n,2^n,3^n,\cdots,k^n)$$

是它的变量的一个整系数多项式, 且当 $n\to\infty$ 时有 $f(n)\to\infty$,[②]那么对无穷多个 n 的值, $f(n)$ 都取合数值.

---

① 有人建议用下面的数列来代替 Fermat 数列:

$$2+1,\quad 2^2+1,\quad 2^{2^2}+1,\quad 2^{2^{2^2}}+1,\cdots$$

它的前 4 个数是素数, 但这个数列的第 5 个数, 即 F_{16}, 现在已知是一个合数. 另一个建议是限制 p 取 Mersenne 素数, 认为这样的话数列 M_p 就会只包含素数了. 然而 $M_{13}=8\,191$ 和 $M_{8\,191}$ 都是合数.

② 对此定理的陈述要小心一些, 以避免 $f(n)$ 取成像 $2^n3^n-6^n+5$ 这样的显然对所有 n 均取素数值的情况.

2.8 关于素数的未解决的问题

1.4 节陈述了两个猜想式的命题, 没有人知道它们的证明, 尽管数值证据表明它们很可能是正确的. 还有许多其他的同类猜想.

存在无穷多个形如 n^2+1 的素数. 更一般地, 如果 a,b,c 是没有公约数的整数, a 是正数, $a+b$ 和 c 不全是偶数, 且 b^2-4ac 不是完全平方数, 那么就有无穷多个形如 an^2+bn+c 的素数存在.

2.7 节 (iii) 已经讨论过 n^2+1. 如果 a,b,c 有公约数, 显然在规定形式的数中最多只有一个素数存在. 如果 $a+b$ 和 c 两者均为偶数, 那么 $N=an^2+bn+c$ 始终是偶数. 如果 $b^2-4ac=k^2$, 那么

$$4aN=(2an+b)^2-k^2.$$

这样一来, 如果 N 是素数, 那么要么 $2an+b+k$, 要么 $2an+b-k$ 整除 $4a$, 而这只能对 n 的至多有限多个值为真. 因此猜想中所说的限制条件是至关重要的.

n^2 和 $(n+1)^2$ 之间总有素数存在.

如果 $n>4$ 是偶数, 那么 n 是 2 个奇素数之和.

这就是 “Goldbach 定理”.

如果 $n\geqslant 9$ 是奇数, 那么 n 是 3 个奇素数之和.

从某个数开始往后的所有 n 要么是一个平方数, 要么是一个素数和一个平方数之和.

这个结论并不是对所有的 n 都为真, 比如 34 和 58 就是例外.

一个更加值得怀疑的猜想 (2.5 节中曾经谈到过它) 是:

Fermat 素数的个数是有限的.

2.9 整 数 模

现在给出在 1.3 节中未给出的定理 3 和定理 2 的证明. 另一个证明在 2.11 节中给出, 第三个证明在 12.4 节中给出. 在本节中, 整数指的是正的或者负的有理整数.

这个证明与数的 “模” 这个概念有关. 模指的是一个数系 S, *S 中任何两个数的和与差也是 S 中的元素*, 也就是说,

$$m\in S,\quad n\in S\to (m\pm n)\in S. \tag{2.9.1}$$

一个模里面的数不一定是正数或有理数 (它们也可以是复数, 或者四元数), 不过这里我们只关心整数的模.

单独一个数 0 构成一个模 [零模(null modulus)].

由 S 的定义推出

$$a\in S\to 0=a-a\in S,\quad 2a=a+a\in S.$$

重复这个方法, 我们看出, 对任何 (正的或负的) 整数 n 有 $na \in S$. 更一般地, 对任何整数 x, y 有

$$a \in S, \quad b \in S \to xa + yb \in S. \tag{2.9.2}$$

另一方面容易看出, 如果给定 a 和 b, $xa + yb$ 的值组成的集合就作成一个模.

显然, 除了零模以外, 任何模 S 都含有正数. 假设 d 是 S 中的最小正数, 如果 n 是 S 中任何一个正数, 那么对所有的 z, $n - zd \in S$. 如果 c 是 n 被 d 除得到的余数, 且

$$n = zd + c,$$

则有 $c \in S$ 且 $0 \leqslant c < d$. 既然 d 是 S 中的最小正数, 故有 $c = 0$ 以及 $n = zd$. 于是就得到

定理 23 除了零模以外, 任何模都是某个正数 d 的整倍数组成的集合.

定义两个不全为零的整数 a 和 b 的**最大公约数**(highest common divisor)d: 如果 d 是能同时整除 a 和 b 的最大正整数. 记为

$$d = (a, b).$$

于是有 $(0, a) = |a|$. 可以用同样的方法定义任意一组正整数 $a, b, c, \cdots, k$ 的最大公约数

$$(a, b, c, \cdots, k).$$

对整数 x, y, 形如

$$xa + yb$$

的数组成的集合是一个模, 根据定理 23, 它是某个正数 c 的倍数 zc 组成的集合. 由于 c 整除 S 中的每一个数, 所以它必整除 a 和 b, 于是

$$c \leqslant d.$$

另一方面,

$$d|a, \ d|b \to d|(xa + yb),$$

所以 d 整除 S 中的每一个数, 特别有 d 整除 c. 由此推得

$$c = d,$$

于是 S 就是由 d 的倍数组成的集合.

定理 24 模 $xa + yb$ 是由 $d = (a, b)$ 的倍数组成的集合.

显然我们还附带证明了

定理 25 方程

$$ax + by = n$$

有整数解 x, y, 当且仅当 $d|n$. 特别地,

$$ax + by = d$$

可解.

定理 26 a 和 b 的任何公约数都整除 d.

2.10 算术基本定理的证明

现在可以来证明 Euclid 的定理 3, 从而也就可以证明定理 2 了.

假设 p 是素数且 $p|ab$. 如果 $p \nmid a$, 那么 $(a,p)=1$, 于是根据定理 24 知, 存在一个 x 和一个 y 使 $xa+yp=1$, 也就是

$$xab+ypb=b.$$

但是 $p|ab$ 且 $p|pb$, 故有 $p|b$.

实际上同样的讨论可以证明

定理 27 $(a,b)=d,\ c>0 \to (ac,bc)=dc.$

因为存在一个 x 和一个 y 使有 $xa+yb=d$, 也就是

$$xac+ybc=dc,$$

从而就有 $(ac,bc)|dc$. 反过来, 我们有 $d|a \to dc|ac$ 以及 $d|b \to dc|bc$, 故由定理 26 有 $dc|(ac,bc)$, 从而有 $(ac,bc)=dc$.

2.11 基本定理的另一个证明

称能以多于一种方式分解成素数乘积的数为非正规数(abnormal). 设 n 是最小的非正规数. 同一个素数 P 不可能在 n 的两个不同的因子分解中出现, 因为如果不然, n/P 就是一个非正规数, 且 $n/P<n$. 那样就有

$$n=p_1p_2p_3\cdots=q_1q_2\cdots,$$

其中 p 和 q 都是素数, 且没有一个 p 等于某个 q, 也没有一个 q 等于任何一个 p.

不妨令 p_1 是最小的 p. 由于 n 是合数, 故 $p_1^2 \leqslant n$. 类似地, 如果 q_1 是最小的 q, 则有 $q_1^2 \leqslant n$. 又由于 $p_1 \neq q_1$, 由此推出 $p_1q_1<n$. 因此, 如果 $N=n-p_1q_1$, 则有 $0<N<n$, 且 N 不是非正规数. 现在有 $p_1|n$, 于是 $p_1|N$. 类似地有 $q_1|N$. 于是 p_1 和 q_1 两者都在 N 的唯一分解式中出现, 且 $(p_1q_1)|N$. 由此推出 $(p_1q_1)|n$, 于是 $q_1|(n/p_1)$. 但是 n/p_1 小于 n, 从而有唯一素数分解 $p_2p_3\cdots$. 由于 q_1 不是任何一个 p, 这是不可能的. 于是不可能有任何非正规数, 这正是基本定理的结论.

本 章 附 注

2.2 节. Ingham 先生告诉我们, 这里所用的方法属于 Bohr 和 Littlewood：见 Ingham, **2**.

2.3 节. 关于定理 11, 12 和 14, 见 Lucas, *Théorie des nombres*, i (1891) 353–354; 关于定理 15, 见 Landau, *Handbuch*, 422–446 以及 *Vorlesungen*, i, 79–96.

定理 15 的一个有趣的推广是由 Shiu 得到的 (*J. London Math. Soc.* (2) **61**(2000), 359–373). 它是说: 对于定理 15 中那样的 a 和 b, 素数序列包含任意长相邻的元素串, 它们全都有 $an+b$ 的形状. 取 $a=1000$ 以及 $b=777$ 为例, 这就意味着我们可以找到如我们所想要的那么多个相邻的素数, 其中每一个素数末尾三位数字都是 777.

2.4 节. 见 Pólya 和 Szegő No. **94**.

2.5 节. 见 Dickson, *History*, 第 1 卷第 1, 5, 16 章, Rouse Ball Mathematical recreations and essays 第 2 章, 有关较早的数值结果, 见 Kraitchik, *Théorie des nombres*, i (Paris,1922), 22, 218 以及 D. H. Lehmer, *Bulletin Amer. Math. Soc.* **38** (1932), 383–384. Miller 和 Wheeler [*Nature* **168** (1951) 838] 给出了他们的大素数, 而 Tuckerman [*Proc. Nat. Acad. Sci. U.S.A.* **68** (1971), 2319–2320] 对 $p=19\,937$ 给出了 Mersenne 素数 M_p, 并给出了用电子计算机发现的其他较小的 Mersenne 素数的参考文献. M_p 对于 $p=21\,701$ 是素数的发现被刊登在了 1978 年 11 月第 17 期 *Times* 上. 关于合数 F_m 的因子, 请见 Hallyburdon 和 Brillhart, *Math. Comp.* **29**(1975), 109-112, 关于 F_8 的因子, 见 Brent, *American Math. Soc. Abstracts*, **1**(1980), 565.

到 2007 年, 对于范围在 $5\leqslant n\leqslant 11$ 中的值, F_n 被公认为都是合数且对它们作了完全的因子分解, 同时对于更大的 n, 也有许多因子被发现. 已知对 $4\leqslant n\leqslant 32$, F_n 都是合数. 没有被发现的 F_n 的因子中最小的 n 是 $n=14$.

类似地, 到 2007 年, 总共发现了 44 个 Mersenne 素数, 最大的一个是 $M_{32\,582\,657}$. 第 39 个 Mersenne 素数已查明是 $M_{13\,466\,917}$, 但尚未对位于这两个数之间的所有 Mersenne 数全部查验完毕.

Ferrier 的素数是 $(2^{148}+1)/7$, 这是不用电子计算机所发现的最大的素数 (很可能这个记录会保持下去).

新的大型计算机使得大数分解以及大数的素性检测成为十分有意思且绝非浅显的研究对象. Guy (*Proc.* 5^{th} *Manitoba Conf. Numerical Math.* 1975, 49–89) 给出了有关因子分解方法的一个完全的说明, 关于素性检测的一些评论以及关于这两个问题的相当丰富的参考文献. 有关素性检测, 也可参见比如说 Brillhart, Lehmer 以及 Selfridge, *Math. Comp.* **29** (1975), 620–647 以及 Selfridge 和 Wunderlich, *Proc.* 4^{th} *Manitoba Conf. Numerical Math.* 1974, 109–120.

根据 Kraitchik 和 Bennett 的说法, 我们给出的 $641|F_5$ 的证明属于 Coxeter (*Introduction to Geometry*, New York, Wiley, 1969).

Ribenboim, *The new book of prime number records* (Springer, New York, 1996) 一书对上面所有的工作以及其他很多研究成果给出了详尽的介绍.

2.6 节. 参见 Erdős, *Mathematica*, **B 7** (1938), 1–2. 定理 19 是 Euler 在 1737 年证明的.

2.7 节. 定理 21 属于 Goldbach (1752), 而定理 22 属于 Morgan Ward, *Journal London Math. Soc.* **5** (1930), 106–107.

2.8 节. 见附录第 3 节.

2.9 节至 2.10 节. 这里的讨论遵循了 Hecke 第 1 章的路线. 模的定义是很自然的, 但有点多余. 只要假设

$$m\in S.\quad n\in S\to m-n\in S$$

就足够了. 因为那样就有

$$0 = n - n \in S,\ -n = 0 - n \in S,\ m + n = m - (-n) \in S.$$

2.11 节. F. A. Lindemann, *Quart. J. of Math.* (Oxford), **4** (1933), 319–320, 以及 Davenport, *Higher arithmetic*, 20. 关于有点类似的证明, 见 Zermelo, *Göttinger Nachrichten* (new series), **i** (1934), 43–44 以及 Hasse, *Journal für Math.* **159** (1928), 3–6.

第 3 章　Farey 数列和 Minkowski 定理

3.1　Farey 数列的定义和最简单的性质

本章主要关注像 1/2 和 7/11 这样的 “正有理数” 或者 “普通分数” 的某些性质. 这样的一个分数可以看成两个正整数之间的一个关系, 因而我们证明的定理也体现了正整数的性质.

n 阶 Farey 数列 $\mathfrak{F}_n$ 是介于 0 和 1 之间且分母不超过 n 的递增的不可约分数序列. 如果

$$0 \leqslant h \leqslant k \leqslant n, \quad (h,k)=1, \tag{3.1.1}$$

那么 h/k 就属于 $\mathfrak{F}_n$. 数 0 和 1 包含在形式 $\frac{0}{1}$ 和 $\frac{1}{1}$ 之中. 例如 $\mathfrak{F}_5$ 是

$$\frac{0}{1},\ \frac{1}{5},\ \frac{1}{4},\ \frac{1}{3},\ \frac{2}{5},\ \frac{1}{2},\ \frac{3}{5},\ \frac{2}{3},\ \frac{3}{4},\ \frac{4}{5},\ \frac{1}{1}.$$

Farey 数列的特征性质由下面几个定理表出.

定理 28　如果 h/k 和 h'/k' 是 $\mathfrak{F}_n$ 中两个相连的项, 那么

$$kh'-hk'=1. \tag{3.1.2}$$

定理 29　如果 h/k、h''/k'' 和 h'/k' 是 $\mathfrak{F}_n$ 中 3 个相连的项, 那么

$$\frac{h''}{k''}=\frac{h+h'}{k+k'}. \tag{3.1.3}$$

我们将在 3.2 节中证明这两个定理是等价的, 然后在 3.3 节、3.4 节和 3.7 节中分别给出这两个定理的 3 个不同的证明. 我们将通过证明 $\mathfrak{F}_n$ 的两个较简单的性质来结束本节.

定理 30　如果 h/k 和 h'/k' 是 $\mathfrak{F}_n$ 中两个相连的项, 那么

$$k+k'>n. \tag{3.1.4}$$

h/k 和 h'/k' 的 “中位数”

$$\frac{h+h'}{k+k'}^{①}$$

落在区间

$$\left(\frac{h}{k},\frac{h'}{k'}\right)$$

中. 因此, 除非 (3.1.4) 式为真, 否则在 $\mathfrak{F}_n$ 中就会有另外一项位于 h/k 和 h'/k' 之间.

定理 31　如果 $n>1$, 则 $\mathfrak{F}_n$ 中不存在两个相连的项能有相同的分母.

① 或这个分数的既约分数.

如果在 $\mathfrak{F}_n$ 中, $k>1$ 且 h'/k 紧跟在 h/k 的后面, 则有 $h+1\leqslant h'<k$, 而

$$\frac{h}{k}<\frac{h}{k-1}<\frac{h+1}{k}\leqslant\frac{h'}{k},$$

从而 $h/(k-1)$[①]在 $\mathfrak{F}_n$ 中就位于 h/k 和 h'/k 之间, 这是一对矛盾.

3.2 两个特征性质的等价性

现在来证明定理 28 和定理 29 分别蕴含另外一个.

(1) 定理 28 蕴含定理 29.

如果假设定理 28 成立, 对 h'' 和 k'' 来解方程

$$kh''-hk''=1,\quad k''h'-h''k'=1, \tag{3.2.1}$$

则得到

$$h''(kh'-hk')=h+h',\quad k''(kh'-hk')=k+k',$$

这就得到 (3.1.3) 式.

(2) 定理 29 蕴含定理 28.

假设定理 29 成立, 并假设定理 28 对 $\mathfrak{F}_{n-1}$ 成立, 要推出定理 28 对 $\mathfrak{F}_n$ 也成立. 显然只要证明: 当 h''/k'' 属于 $\mathfrak{F}_n$ 但不属于 $\mathfrak{F}_{n-1}$(即有 $k''=n$) 时 (3.2.1) 式成立. 此时, 根据定理 31 可知, k 和 k' 两者都小于 k'', 于是 h/k 和 h'/k' 是 $\mathfrak{F}_{n-1}$ 中相连的两项.

由于根据假设有 (3.1.3) 式为真, 且 h''/k'' 是不可约的, 于是就有

$$h+h'=\lambda h'',\quad k+k'=\lambda k'',$$

其中 λ 是一个整数. 既然 k 和 k' 两者都小于 k'', λ 必定等于 1. 从而

$$h''=h+h',\quad k''=k+k',$$
$$kh''-hk''=kh'-hk'=1.$$

类似地, 有

$$k''h'-h''k'=1.$$

3.3 定理 28 和定理 29 的第一个证明

我们的第一个证明是 3.2 节中所用的思想的一个自然展开.

这两个定理对 $n=1$ 均为真. 假设它们对 $\mathfrak{F}_{n-1}$ 成立, 要证它们对 $\mathfrak{F}_n$ 也成立.

设 h/k 和 h'/k' 是 $\mathfrak{F}_{n-1}$ 中两个相连的项, 但它们在 $\mathfrak{F}_n$ 中被 h''/k'' 隔开.[②] 令

$$kh''-hk''=r>0,\quad k''h'-h''k'=s>0. \tag{3.3.1}$$

对 h'' 和 k'' 解这些方程, 记住有

$$kh'-hk'=1,$$

① 或这个分数的既约分数.

② 根据定理 31, h''/k'' 是 $\mathfrak{F}_n$ 中位于 h/k 和 h'/k' 之间仅有的一项, 但证明中并没有假设这一点.

于是得到

$$h'' = sh + rh', \quad k'' = sk + rk'. \tag{3.3.2}$$

这里有 $(r, s) = 1$, 这是因为 $(h'', k'') = 1$.

现在考虑所有分数

$$\frac{H}{K} = \frac{\mu h + \lambda h'}{\mu k + \lambda k'} \tag{3.3.3}$$

的集合 S, 其中 λ 和 μ 都是正整数, 且 $(\lambda, \mu) = 1$. 于是 h''/k'' 属于 S. S 的每个分数都在 h/k 和 h'/k' 之间, 且都是既约分数, 这是因为 H 和 K 的任何公约数都能整除

$$k(\mu h + \lambda h') - h(\mu k + \lambda k') = \lambda$$

和

$$h'(\mu k + \lambda k') - k'(\mu h + \lambda h') = \mu.$$

从而 S 的每个分数或迟或早都会出现在某个 $\mathfrak{F}_q$ 中, 且显然首次出现的那个分数即是使得 K 取最小值者, 也即是使 $\lambda = 1, \mu = 1$ 者. 这个分数必为 h''/k'', 所以

$$h'' = h + h', \quad k'' = k + k'. \tag{3.3.4}$$

如果用这些值来代替 (3.3.1) 式中的 h'' 和 k'', 则可得 $r = s = 1$. 这就对 $\mathfrak{F}_n$ 证明了定理 28. 对于 $\mathfrak{F}_n$ 的 3 个连续的分数来说, (3.3.4) 式一般来说并不为真, 然而 (如前面已经指出的) 当中间那个分数在 $\mathfrak{F}_n$ 中第一次出现时, 这些方程是成立的.

3.4 定理 28 和定理 29 的第二个证明

这个证明不是归纳证明, 它给出 $\mathfrak{F}_n$ 中紧跟在 h/k 之后的那一项的构造法则. 由于 $(h, k) = 1$, 故方程

$$kx - hy = 1 \tag{3.4.1}$$

有整数解 (定理 25). 如果 x_0, y_0 是一组解, 那么对于任何正的或者负的整数 r

$$x_0 + rh, \quad y_0 + rk$$

仍然是该方程的解. 可以选择 r 的值, 使得有 $n - k < y_0 + rk \leqslant n$. 这样一来 (3.4.1) 式就有一组解 (x, y) 使得

$$(x, y) = 1, \quad 0 \leqslant n - k < y \leqslant n. \tag{3.4.2}$$

由于 x/y 已经约分, 且 $y \leqslant n$, 故而 x/y 是 $\mathfrak{F}_n$ 中的一个分数. 同样有

$$\frac{x}{y} = \frac{h}{k} + \frac{1}{ky} > \frac{h}{k},$$

于是在 $\mathfrak{F}_n$ 中 x/y 位于 h/k 的后面. 如果它不是 h'/k', 它就位于 h'/k' 的后面, 且

$$\frac{x}{y} - \frac{h'}{k'} = \frac{k'x - h'y}{k'y} \geqslant \frac{1}{k'y},$$

然而

$$\frac{h'}{k'}-\frac{h}{k}=\frac{kh'-hk'}{kk'}\geqslant\frac{1}{kk'},$$

从而根据 (3.4.2) 就有

$$\frac{1}{ky}=\frac{kx-hy}{ky}=\frac{x}{y}-\frac{h}{k}\geqslant\frac{1}{k'y}+\frac{1}{kk'}=\frac{k+y}{kk'y}>\frac{n}{kk'y}\geqslant\frac{1}{ky},$$

这是一对矛盾. 于是 x/y 必定等于 h'/k', 且有 $kh'-hk'=1$.

比方说, 要在 $\mathfrak{F}_{13}$ 中求 $4/9$ 的后继分数, 我们先要求 $9x-4y=1$ 的某一组解 (x_0,y_0), 例如解 $x_0=1,y_0=2$. 然后来选择 r 使得 $2+9r$ 在 $13-9=4$ 和 13 之间. 这给出 $r=1$, $x=1+4r=5$, $y=2+9r=11$, 于是所求的分数就是 $5/11$.

3.5 整 数 格 点

第三个也是最后一个证明有赖于一个简要的几何思想.

假设在平面上给定了原点 O 以及两个与 O 不共线的点 P,Q. 作出平行四边形 $OPQR$[①], 让它的边不确定, 画出两组等距的平行线, 其中 OP,QR 以及 OQ,PR 是这两组平行线中相邻的两条平行线, 这样它们就把平面分成无穷多个相等的平行四边形. 这样一个图形就称为一个格(lattice). 德语称为 Gitter.

一个格是由线作成的一个图形, 它定义了一个由点构成的图形, 也就是说由线的交点系 (或称为格点) 构成的图形. 我们称这样的一个系统为一个点格(point-lattice).

两个不同的格有可能确定同样的点格. 例如在图 1 中, 基于 OP,OQ 的格和基于 OP,OR 的格所确定的是同一个格点系. 决定同样点格的两个格称为等价的.

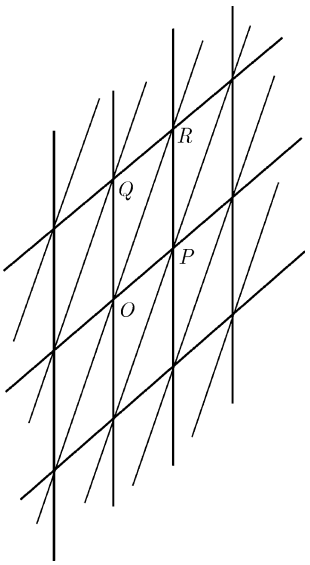

图 1

显然, 一个格的任何格点都可以看成是原点 O, 而且格的性质与原点的选取无关, 且格是关于任意的原点为对称的.

这里有一种类型的格特别重要. 这就是 (当给定直角坐标系时) 由平行于坐标轴且相距单位距离的平行线作成的格, 这些平行线把平面划分成单位正方形. 我们把这样的格称为基本格(fundamental lattice)L, 它所确定的点格 [也就是由整数坐标的点 (x,y) 作成的系统] 称为基本点格(fundamental point-lattice) Λ.

任何点格都可以看作是一个由数或者向量组成的系统, 其中格点的复数坐标为 $x+\mathrm{i}y$, 而向量是从

① 原书如此. 我国的平行四边形表示法与此不同, 应为$\square OPRQ$. —— 编者注

原点出发到格点的向量. 这样的一个系统显然作成在 2.9 节意义下的一个模. 如果 P 和 Q 是点 (x_1, y_1) 和 (x_2, y_2), 则基于 OP 和 OQ 的格中的任何一点 S 的坐标是

$$x = mx_1 + nx_2, \quad y = my_1 + ny_2,$$

其中 m 和 n 是整数. 换言之, 如果 z_1 和 z_2 是 P 和 Q 的复坐标, 那么 S 的复坐标就是

$$z = mz_1 + nz_2.$$

3.6 基本格的某些简单性质

(1) 现在来考虑由

$$x' = ax + by, \quad y' = cx + dy \tag{3.6.1}$$

定义的变换, 其中 a, b, c, d 是给定的正的或者负的整数. 显然, Λ 的每个点 (x, y) 都会变成 Λ 的另一个点 (x', y').

对 x 和 y 求解 (3.6.1) 式, 得到

$$x = \frac{dx' - by'}{ad - bc}, \quad y = -\frac{cx' - ay'}{ad - bc}. \tag{3.6.2}$$

如果

$$\Delta = ad - bc = \pm 1, \tag{3.6.3}$$

那么 x' 和 y' 的任何一组整数值都给出 x 和 y 的一组整数值, 且每个格点 (x', y') 对应于一个格点 (x, y). 此时, Λ 被变换成自己.

反过来, 如果 Λ 被变换成自己, 每一个整数点 (x', y') 必定给出一个整数点 (x, y). 特别地, 取 (x', y') 为 $(1, 0)$ 和 $(0, 1)$, 可以看出

$$\Delta | d, \quad \Delta | b, \quad \Delta | c, \quad \Delta | a,$$

于是

$$\Delta^2 | (ad - bc), \quad \Delta^2 | \Delta.$$

从而有 $\Delta = \pm 1$.

这样就证明了:

定理 32 变换 (3.6.1) 把 Λ 变成自己的充分必要条件是 $\Delta = \pm 1$.

称这样一个变换为幺模变换(unimodular).

(2) 现在假设 P 和 Q 是 Λ 的格点 (a, c) 和 (b, d). 由 OP 和 OQ 所定义的平行四边形的面积是

$$\delta = \pm(ad - bc) = |ad - bc|,$$

其中符号的选取是使 δ 取正数. 基于 OP 和 OQ 的格 Λ' 中的点 (x', y') 由

$$x' = xa + yb, \quad y' = xc + yd$$

给出, 其中 x 和 y 是任意整数. 根据定理 32, Λ' 与 Λ 完全相同的充分必要条件是 $\delta = 1$.

定理 33 基于 OP 和 OQ 的格 L' 等价于格 L 的充分必要条件是由 OP 和 OQ 所定义的平行四边形的面积为 1.

(3) 称格 Λ 的一个点 P 是可视的 (即从原点看去为可视的), 如果在 OP 上没有 Λ 中的介于 O 和 P 之间的点存在. 为使得点 (x, y) 是可视的, 其充分必要条件是 x/y 不可约, 即 $(x, y) = 1$.

定理 34 设 P 和 Q 是 Λ 中的可视点, 且 δ 是由 OP 和 OQ 所定义的平行四边形 J 的面积. 那么

(i) 如果 $\delta = 1$, 则在 J 的内部没有 Λ 的点;

(ii) 如果 $\delta > 1$, 那么 Λ 至少有一个点在 J 的内部, 且除非该点是 J 的对角线的交点, 否则 Λ 至少有两个点在 J 的内部, 每个点都在 J 被 PQ 所分成的两个三角形的一个之中.

当且仅当基于 OP 和 OQ 的格 L' 与格 L 等价时, 也就是当且仅当 $\delta = 1$ 时, 在 J 的内部没有 Λ 的点. 如果 $\delta > 1$, 就至少有一个这样的点 S. 如果 R 是平行四边形 J 的第四个顶点, 且 RT 与 OS 平行且相等, 但其方向相反, 那么 (由于格的性质是对称的, 且与选取哪个特定的点作为原点无关)T 也是 Λ 的一个点, 这样在 J 中就至少有 Λ 的两个点, 除非 T 与 S 重合. 这就是情形 (ii) 中的特例.

不同的情形由图 2a, 2b, 2c 给出.

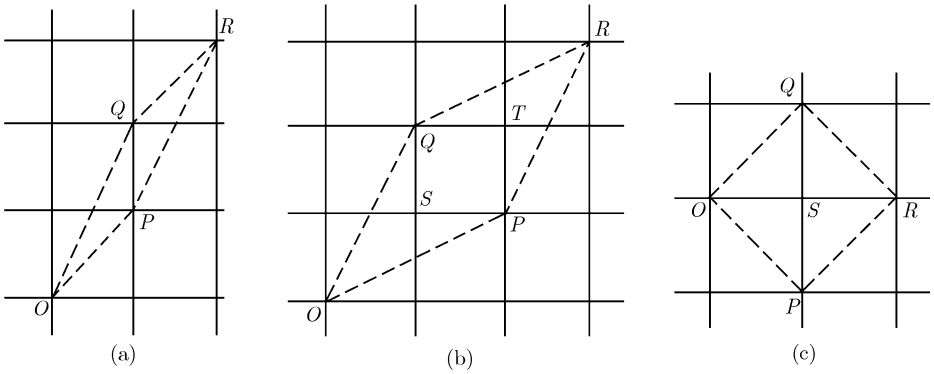

图 2

3.7 定理 28 和定理 29 的第三个证明

满足条件

$$0 \leqslant h \leqslant k \leqslant n, \quad (h, k) = 1$$

的分数 h/k 都是 $\mathfrak{F}_n$ 中的分数, 且对应 Λ 中的可视点 (k, h), 该点在由直线 $y = 0$, $y = x$, $x = n$ 所定义的三角形的内部或边界上.

如果画出一条经过 O 的射线, 并将它绕原点从起始位置 x 轴开始沿逆时针方向旋转, 它就依次经过 Farey 分数所代表的每个点 (k,h). 如果 P 和 P' 是代表相连分数的两个点 (k,h) 和 (k',h'), 那么在三角形 OPP' 的内部以及在连线 PP' 上就没有所表示的点存在, 于是由定理 34 就有

$$kh' - hk' = 1.$$

3.8 连续统的 Farey 分割

在一个圆上表示实数而不是像通常那样在一条直线上表示实数, 常常更加方便, 圆周所表示的实数去掉了整数部分. 取一个由单位圆周作成的圆 C, 取圆周上任意一个点 O 表示数 0, 用点 P_x 来表示 x, 该点在圆周上沿逆时针方向度量的离点 O 的距离就是 x. 显然所有的整数都由同一个点 O 来表示, 且相差一个整数的数有同样的表示点.

有时把 C 的圆周按照下述方式加以划分是有用的. 取 Farey 数列 $\mathfrak{F}_n$, 对相连的分数对 h/k 和 h'/k' 作出所有的中位数

$$\mu = \frac{h+h'}{k+k'}.$$

其中第一个以及最后一个中位数是

$$\frac{0+1}{1+n} = \frac{1}{n+1}, \quad \frac{n-1+1}{n+1} = \frac{n}{n+1}.$$

当然, 这些中位数本身并不属于 $\mathfrak{F}_n$.

现在用点 P_μ 来表示每一个中位数 μ. 圆就被分成了若干弧段 [称为Farey 弧(Farey arc)], 每一段弧都介于两个点 P_μ 之间, 且包含一个Farey 点 (Farey point), 此即 $\mathfrak{F}_n$ 中一项的表示. 于是

$$\left(\frac{n}{n+1}, \frac{1}{n+1}\right)$$

就是包含一个 Farey 点 O 的一段 Farey 弧. 把 Farey 弧的集合称为圆的一个Farey 分割(Farey dissection).

下面假设 $n > 1$. 如果 $P_{h/k}$ 是一个 Farey 点, 且 $h_1/k_1, h_2/k_2$ 是 $\mathfrak{F}_n$ 中的紧接在 h/k 的前面以及紧跟在它后面的项, 那么环绕 $P_{h/k}$ 的 Farey 弧由两部分组成, 这两部分的长度分别为

$$\frac{h}{k} - \frac{h+h_1}{k+k_1} = \frac{1}{k(k+k_1)}, \quad \frac{h+h_2}{k+k_2} - \frac{h}{k} = \frac{1}{k(k+k_2)}.$$

由于 k 和 k_1 不相等 (定理 31) 且二者都不超过 n, 故有 $k+k_1 < 2n$. 又由定理 30 有 $k+k_1 > n$. 于是得到

定理 35 在 n 阶 Farey 分割中 $(n>1)$, 包含 h/k 的表示点的弧的每一部分长度都介于 $\frac{1}{k(2n-1)}$ 和 $\frac{1}{k(n+1)}$ 之间.

事实上, 这种分割有某种 “一致性”, 这种性质显示出它的重要性.

这里要用 Farey 分割来证明用有理数逼近任意实数的一个简单的定理, 我们将在第 11 章中再回到这个问题.

定理 36 如果 ξ 是任意一个实数, n 是一个正整数, 那么必存在一个不可约分数 h/k 使得

$$0 < k \leqslant n, \quad \left|\xi - \frac{h}{k}\right| \leqslant \frac{1}{k(n+1)}. \tag{3.8.1}$$

可以假设 $0 < \xi < 1$. 则 ξ 落在由 $\mathfrak{F}_n$ 中两个相连的分数, 比方说就是 h/k 和 h'/k' 所界限的区间之中, 从而它也就落在区间

$$\left(\frac{h}{k}, \frac{h+h'}{k+k'}\right), \quad \left(\frac{h+h'}{k+k'}, \frac{h'}{k'}\right)$$

中的某一个里. 这样一来, 根据定理 35 知, 要么是 h/k 要么是 h'/k' 满足定理中的条件：如果 ξ 落在第一个区间中, 则有 h/k 满足条件; 如果 ξ 落在第二个区间中, 则有 h'/k' 满足条件.

3.9 Minkowski 的一个定理

如果 P 和 Q 是 Λ 的点, P' 和 Q' 是 P 和 Q 关于原点对称的点, 除了定理 34 中所给的平行四边形 J 外, 我们再给出基于 OQ, OP', 基于 OP', OQ' 以及基于 OQ', OP 的三个平行四边形, 我们得到一个平行四边形 K, 其中心是原点, 其面积 4δ 是 J 的面积的四倍. 如果 δ 的值为 1(这是它最小可能的值), 那么在 K 的边界上就有 Λ 的点, 但在其内部除了 O 以外, 没有 Λ 的点. 如果 $\delta > 1$, 则在 K 的内部除了 O 以外还有 Λ 的点. 这是 Minkowski 的一个著名定理的一个非常特别的情况, Minkowski 定理断言：不仅仅关于原点对称的任何平行四边形 (无论它们是否由 Λ 的点所生成的) 具有同样的性质, 而且关于原点对称的任何 “凸区域” 也有同样性质成立.

一个**开区域**(open region)R 是具有下述性质的点的集合：(i) 如果 P 属于 R, 那么平面上充分接近 P 的所有的点也都属于 R; (ii) R 的任何两点都可以用一条完全位于 R 内部的连续曲线连接起来. 我们还可以将 (i) 表示成 “R 的任何点都是 R 的**内点**(interior point)”. 于是一个圆或者一个平行四边形的内部都是开区域. R 的**边界**(boundary)C 是由本身并不属于 R 的、R 的极限点组成的集合. 从而一个圆的边界就是它的圆周. 一个**闭区域**(closed region)R^* 是一个开区域 R 加上它的边界所得的集合. 我们仅考虑有界区域.

凸(convex) 区域有两个自然的定义, 可以证明它们是等价的. 第一个定义可以说成：R(或者 R^*) 是凸的, 如果 R 中任何一条弦上的每一点 (即连接 R 的任何两点的线段上的每一点) 都属于 R. 第二个定义可以说成：R(或者 R^*) 是凸的, 如果

经过 C 的每一点 P 都可以画出至少一条直线 l, 使得 R 中所有的点都在 l 的某一侧. 于是, 圆和平行四边形都是凸的. 对于圆来讲, l 就是在点 P 的切线; 而对于平行四边形来讲, 每条直线 l 都是它的一条边 (除了在顶点处以外), 而在顶点处它有无穷多条符合要求的直线.

容易证明这两个条件的等价性. 首先假设根据第二个定义 R 是凸的, 又设 P 和 Q 属于 R, 而 PQ 上有一个点 S 不属于 R. 那么 C 上就有一点 T(也可能就是 S 自己) 在 PS 上, 且有一条经过 T 的直线 l 使得 R 整个位于 l 的一侧, 但因为所有充分靠近 P 或者 Q 的点都属于 R, 这是一对矛盾.

其次, 假设根据第一个定义 R 是凸的, P 是 C 的一个点. 考虑将 P 和 R 的点联接作出的直线的集合 L. 如果 Y_1 和 Y_2 是 R 中的点, Y 是 Y_1Y_2 上的一个点, 那么 Y 就是 R 的一个点且 PY 是 L 中的一条线. 于是就有一个角度 $\angle APB$, 它使得从 P 出发的每一条限于 $\angle APB$ 内部的直线均属于 L, 且没有一条从 P 出发但在 APB 外部的直线是属于 L 的. 如果 $\angle APB > \pi$, 则存在 R 的点 D, E 使得 DE 通过 P, 此时情形 P 属于 R, 但不属于 C, 这是一对矛盾. 从而有 $\angle APB \leqslant \pi$. 如果 $\angle APB = \pi$, 则 AB 就是一条直线 l; 如果 $\angle APB < \pi$, 则任何一条位于这个角的外边且经过点 P 的直线都是直线 l.

显然, 凸性是关于平移以及关于点 O 的伸缩变换的不变量.

凸区域 R 有面积(area) 存在 (例如它的面积可以定义为顶点在 R 内部的小正方形网格总面积的上界).

定理 37 (Minkowski 定理) 任何关于点 O 对称且面积大于 4 的凸区域, 其内部都至少含有 Λ 中异于 O 的一个点.

3.10 Minkowski 定理的证明

先来证明一个简单的定理, 这个定理的真实性是 “直观的”.

定理 38 设 R_O 是包含点 O 的一个开区域, R_P 是与之全等且关于 Λ 中任一点 P 位置类似的一个区域, 且诸区域 R_P 中没有两个是重叠的. 那么 R_O 的面积不超过 1.

如果考虑的 R_O 是由直线 $x = \pm 1/2, y = \pm 1/2$ 界限的正方形, 定理就变成 “显然的”, 此时 R_O 的面积就等于 1, 而区域 R_P 加上它们的边界将会覆盖住整个平面. 可以给出该定理的确切证明如下.

假设 Δ 是 R_O 的面积, A 是 C_O[①]的点离点 O 的最大距离. 考虑与 Λ 的其坐标在数值上都不大于 n 的点所对应的 $(2n+1)^2$ 个区域 R_P. 所有这些区域都位于一个正方形的内部, 这个正方形的边与坐标轴平行且到点 O 的距离为 $n + A$. 从而

① 我们经常用 C 来表示与 R 对应的边界.

(由于诸区域不相重叠)

$$(2n+1)^2\Delta \leqslant (2n+2A)^2, \quad \Delta \leqslant \left(1+\frac{A-\frac{1}{2}}{n+\frac{1}{2}}\right)^2,$$

令 n 趋向无穷就得到所要的结果.

值得注意的是, 在定理 38 中并没有用到对称性或者凸性.

现在容易证明 Minkowski 定理了. Minkowski 本人给出过两个证明, 这两个证明基于凸性的两个定义.

(1) 取第一个定义, 并假设 R_O 是将 R 关于点 O 收缩到它的线性维数一半所得到的结果. 那么 R_O 的面积大于 1, 于是定理 38 中的诸区域 R_P 中有两个是重叠的, 从而有一个格点 P 存在, 使得 R_O 与 R_P 重叠. 设 Q 是 R_O 和 R_P 的一个公共点 (图 3a). 如果 OQ' 与 PQ 相等且平行, Q'' 是 Q' 关于 O 的映像, 于是 Q', Q'' 都在 R_O 中. 这样一来, 根据凸性的定义, QQ'' 的中点在 R_O 中. 但这一点是 OP 的中点, 于是 P 在 R 中.

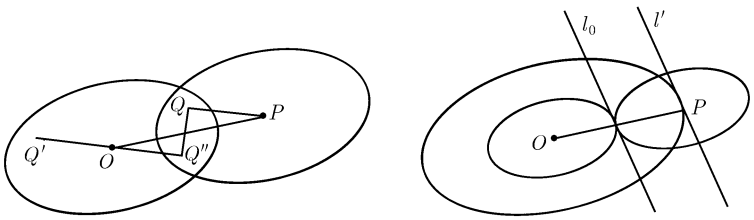

图 3

(2) 取第二个定义, 假设除了点 O 以外没有格点在 R 中. 环绕 O 扩大 R^*(与 R'^* 一样), 直到它首次包含一个格点 P 为止. 那么 P 是 C' 的一个点, 且有经过点 P 的一条线 l, 比方说就是 l'(图 3b). 如果 R_O 是由 R' 环绕点 O 将其线性维数收缩到原来的一半得到的结果, 又 l_O 经过 OP 的中点且与 l 平行, 于是 l_O 对 R_O 来说就是一条直线 l. 它显然也是对 R_P 来说的一条直线 l, 且使得 R_O 和 R_P 各在它相反的两侧, 从而 R_O 和 R_P 不会互相重叠. 进而 R_O 也不和任何其他的 R_P 重叠. 但由于 R_O 的面积大于 1, 这与定理 38 矛盾.

还有若干个可供选择且有意思的证明, 其中最简单的一个证明由 Mordell 给出.

如果 R 是凸的且关于点 O 对称, 且 P_1 和 P_2 是 R 中的坐标为 (x_1, y_1) 和 (x_2, y_2) 的点, 那么 $(-x_2, -y_2)$, 从而坐标为 $\frac{1}{2}(x_1-x_2)$ 和 $\frac{1}{2}(y_1-y_2)$ 的点 M 也是 R 的点.

直线 $x = 2p/t, y = 2q/t$(其中 t 是一个固定的正整数, 而 p 和 q 是任意的整数) 把平面分成面积为 $4/t^2$ 的正方形, 它的角点是 $(2p/t, 2q/t)$. 如果 $N(t)$ 是 R 中角点的个数, 而 A 是 R 的面积, 那么显然当 $t \to \infty$ 时有 $4t^{-2}N(t) \to A$. 如果 $A > 4$,

则对大的 t 有 $N(t) > t^2$. 但是当 p 和 q 被 t 除的时候, 数对 (p, q) 至多给出 t^2 个不同的余数对. 这样一来, R 中就有两个点 P_1 和 P_2, 其坐标为 $2p_1/t, 2q_1/t$ 以及 $2p_2/t, 2q_2/t$, 使得 $p_1 - p_2$ 和 $q_1 - q_2$ 两者都能被 t 整除. 因此点 M(它属于 R) 是 Λ 的一个点.

3.11 定理 37 的进一步拓展

定理 37 的一些进一步推广是第 24 章中所希望得到的, 在这里, 我们很自然地证明这些结果. 我们首先给出一个一般性的说明, 这个说明对于 3.6 节以及 3.9 节和 3.10 节中的所有定理都适用.

我们一直主要对 "基本的" 格 L(或者 Λ) 感兴趣, 但是我们能以各种方式看到, 基本格的性质是如何作为格的一般性质再次被陈述的. 现在用 L 或者 Λ 来表示由直线或者由点构成的格. 如果像 3.5 节中那样, 格是以点 O, P, Q 为基础构建的, 那么就称平行四边形 $OPRQ$ 为 L 或者 Λ 的*基本平行四边形*(fundamental parallelogram).

(i) 可以建立一个以 OP, OQ 为坐标轴的笛卡儿斜坐标系, 并约定 P 和 Q 是点 $(1, 0)$ 和 $(0, 1)$. 那么基本平行四边形的面积就是

$$\delta = OP \cdot OQ \cdot \sin\omega,$$

其中 ω 是 OP, OQ 之间的夹角. 在这个坐标系中对 3.6 节中的论证加以解释就证明了下面的定理.

定理 39 *变换 (3.6.1) 把 Λ 变成自身的充分必要条件是 $\Delta = \pm 1$.*

定理 40 *如果 P 和 Q 是 Λ 的任意两点, 那么, 基于 OP 和 OQ 的格 L' 与格 L 等价的充分必要条件是: 由 OP 和 OQ 所定义的平行四边形的面积等于 Λ 的基本平行四边形的面积.*

(ii) 变换

$$x' = \alpha x + \beta y, \quad y' = \gamma x + \delta y$$

(现在这里的 $\alpha, \beta, \gamma, \delta$ 是任意实数)[①] 把 3.5 节中的基本格变换成由原点以及点 $(\alpha, \gamma), (\beta, \delta)$ 所确定的格. 它把直线变成直线, 把三角形变成三角形. 如果三角形 $P_1P_2P_3$[其中 P_i 是点 (x_i, y_i)] 被变换成三角形 $Q_1Q_2Q_3$, 则这两个三角形的面积为

$$\pm\frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$$

和

① 本节中的 δ 与 (i) 中的 δ 无关, 它在下面还会重复出现.

$$\pm\frac{1}{2}\begin{vmatrix} \alpha x_1+\beta y_1 & \gamma x_1+\delta y_1 & 1 \\ \alpha x_2+\beta y_2 & \gamma x_2+\delta y_2 & 1 \\ \alpha x_3+\beta y_3 & \gamma x_3+\delta y_3 & 1 \end{vmatrix} = \pm\frac{1}{2}(\alpha\delta-\beta\gamma)\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}.$$

于是这两个三角形的面积相差一个常数因子 $|\alpha\delta-\beta\gamma|$. 同样的结论对一般情形的面积也仍然为真, 这是因为在一般情形下它们要么是三角形的面积之和, 或者是三角形面积之和的极限.

这样一来, 可以把一个基本格在适当的线性变换之下的任何性质加以推广. 定理 38 的推广是:

定理 41 假设 Λ 是含有原点 O 的一个格, 且 R_O(关于 Λ) 满足定理 38 中的条件. 那么 R_O 的面积不超过 Λ 的基本平行四边形的面积.

既然要在下一个定理的证明中用到类似的思想, 所以在这里将这个定理的证明从头到尾详尽地给出是恰如其分的. 这个证明依照上面 (i) 的路线, 实际上和 3.10 节中的方法相同.

直线

$$x=\pm n, \quad y=\pm n$$

定义了一个面积为 $4n^2\delta$ 的平行四边形 Π, 有 Λ 的 $(2n+1)^2$ 个点 P 在 Π 的内部或者在它的边界上. 来考虑与这些点所对应的 $(2n+1)^2$ 个区域 R_P. 如果 A 是 $|x|$ 和 $|y|$ 在 C_O 上的最大值, 那么所有这些区域都在一个面积为 $4(n+A)^2\delta$ 的平行四边形 Π' 的内部, 该平行四边形以直线

$$x=\pm(n+A), \quad y=\pm(n+A)$$

为其边界, 且有

$$(2n+1)^2\Delta \leqslant 4(n+A)^2\delta.$$

于是, 令 $n\to\infty$ 就得到

$$\Delta \leqslant \delta.$$

我们还需要一个关于极限情形 $\Delta=\delta$ 的定理. 假设 R_O 是一个平行四边形, 在此假设下我们所证明的结果对于第 24 章中的目的来说是足够的了.

称两个点 (x,y) 和 (x',y') 是关于 L 等价的(equivalent with respect to L), 如果它们在 L 的两个平行四边形中有相似的位置 (因此, 如果一个平行四边形被平行移动到与另一个平行四边形重合时, 这两点就会重合). 如果 L 基于 OP 和 OQ, 且 P 和 Q 是 (x_1,y_1) 和 (x_2,y_2), 那么点 (x_1,y_1) 和 (x_2,y_2) 等价的条件就是

$$x'-x=rx_1+sx_2, \quad y'-y=ry_1+sy_2,$$

其中 r 和 s 是整数.

定理 42 如果 R_O 是一个平行四边形, 其面积与 L 的基本平行四边形的面积相等, 且在 R_O 的内部没有两个点是等价的, 那么在 R_O 的内部或边界上就存在一个点, 它与平面上任何给定的点均等价.

使用 R_P^* 来记与 R_P 对应的闭区域.

假设 "R_O 不包含两个等价的点" 等价于假设 "任意两个 R_P 皆不重叠". 而结论 "R_O^* 中有一个点与平面的任意一点等价" 等价于结论 "R_P^* 覆盖整个平面". 从而要证明的就是: 如果 $\Delta = \delta$ 且 R_P 均不重叠, 那么 R_P^* 就覆盖整个平面.

假设相反的情形出现, 则在所有 R_P^* 的外部就存在一个点 Q. 这个点 Q 在 L 中的某个平行四边形的内部或者边界上, 且在这个平行四边形中有一个区域 D, 它有正的面积 η 且在所有 R_P 的外部, 又在 L 的每一个平行四边形中有一个对应的区域. 因此, 在面积为 $4(n+A)^2\delta$ 的平行四边形 Π' 的内部, 所有的 R_P 的面积不超过

$$4(\delta-\eta)(n+A+1)^2,$$

由此得出

$$(2n+1)^2\delta \leqslant 4(\delta-\eta)(n+A+1)^2.$$

这样一来, 令 $n\to\infty$ 就有

$$\delta \leqslant \delta-\eta.$$

这是一对矛盾, 由此就证明了定理.

最后要说明的是, 所有这些定理都可以推广到任意维数的空间中去. 比如说, 如果 Λ 是三维空间中的基本点格, 即形如 (x,y,z) 且坐标为整数的点的集合, R 是一个关于原点对称的凸区域, 且其体积大于 8, 那么在 R 中就存在 Λ 的异于 O 的点. 在 n 维空间中 8 应代之以 2^n. 第 24 章还要继续讨论一下这个推广, 但并不需要新的思想.

本 章 附 注

3.1 节. "Farey 数列" 的历史非常有趣. 定理 28 和定理 29 似乎是在 1802 年由 Haros 首先提出并予以证明的, 见 Dickson, *History*, i, 156. 直到 1816 年 Farey 才在 *Philosophical Magazine* 的一篇注记中陈述了定理 29. 他没有给出证明, 且这个定理也不像是他所发现的, 因为他似乎至多是一个平凡的数学家.

然而, Cauchy 看到了 Farey 的陈述并补充了证明 (*Exercices de mathématique*, i, 114–116). 通常数学家们都依照 Cauchy 的说法把这个结果归功于 Farey, 于是这个数列就一直冠以他的名字.

有关 Farey 数列的更完整的说明, 见 Rademacher, *Lectures in elementary number theory* (New York, Blaisdell, 1964). 更详细的内容参见 Huxley, *Acta Arith.* **18**(1971), 281–287 以及 Hall, *J. London Math. Soc.* (2) **2** (1970), 139–148.

3.3 节. Hurwitz, *Math. Annalen*, **44**(1894), 417–436. H. G. Diamond 教授使我们注意到在较早的版本中这处证明的不完整性.

3.4 节. Landau, *Vorlesungen*, i. 98–100.

3.5 节至 3.7 节. 这里我们采用了 Pólya 教授的讲稿中的路线.

3.8 节. 定理 36 见 Landau, *Vorlesungen*, i. 100.

3.9 节. 如果读者不乐意的话, 他们不必对这一节里给出的 “区域”、“边界” 等定义给予太多的关注; 他们可以通过用像平行四边形、多边形或者椭圆这样的初等区域的术语来进行思考而不会失去什么. 凸区域是不包含 “拓扑” 困难的简单区域. 凸区域有面积这一结论是由 Minkowski 首先证明的 (*Geometrie der Zahlen*, 第 2 章).

3.10 节. Minkowski 的第一个证明可以在 *Geometrie der Zahlen*, 73-76 中找到, 他的第二个证明给出在 *Diophantische Approximationen*, 28–30 中. Mordell 的证明是在 *Compositio Math.* **1**(1934), 248–253 中给出的. 另外一个有趣的证明是由 Hajós, *Acta Univ. Hungaricae* (Szeged), **6**(1934), 224–225 给出的, 这在本书第 1 版中作了详尽的阐述.

第 4 章　无　理　数

4.1　概　　论

如同在分析教科书中解释的那样, “无理数” 的理论被划分在算术范围之外. 数论首先是研究整数, 接下来是研究有理数 (它可以看成是整数之比), 然后才是特殊形式的无理数、实数或者复数, 比如

$$r+s\sqrt{2}, \quad r+s\sqrt{-5},$$

其中 r 和 s 是有理数. 数论一般并不研究全体无理数或者无理性的一般判别法 (尽管这是一个我们并不很重视的限制).

然而, 还有许多无理性的问题可以看成是算术的一部分. 关于有理数的定理可以被重新表述成关于整数的定理. 因此, 定理

$$\text{“}r^3+s^3=3 \text{ 没有有理数解”}$$

可以被重新表述成下述形式:

$$\text{“}a^3d^3+b^3c^3=3b^3d^3 \text{ 没有整数解”}.$$

对于涉及 “无理性” 的许多定理而言, 同样也可以进行重新表述. 比如说

$$\text{“}\sqrt{2} \text{ 是无理数”} \tag{P}$$

的含义是

$$\text{“}a^2=2b^2 \text{ 没有整数解”}, \tag{Q}$$

这样它就作为一个真正的算术定理出现了. 我们用不着超出算术的正常范围就可以问: “$\sqrt{2}$ 是无理数吗?”, 而且也不必问 “$\sqrt{2}$ 有什么意义?”. 我们不需要对单个符号 $\sqrt{2}$ 作任何解释, 因为 (P) 的意义是作为一个整体定义的, 且与 (Q) 的含义相同[①].

本章将研究问题

$$\text{“}x \text{ 是有理数还是无理数?”},$$

这里 x 是一个像 $\sqrt{2}$, e 或者 π 这样的数, 这些数很自然地出现在分析中.

4.2　已知的无理数

我们考虑的问题一般来说是很困难的, 只对少数不同类型的数 x 找到了问题的解答. 在本章里, 我们仅把注意力集中在几个最简单的情形, 不过, 首先对这方面

① 简言之, 这里 $\sqrt{2}$ 可以在 *Principia Mathematica* 的意义下作为 “不完全的符号” 来处理.

已知的结果给出一个概述也许更加方便. 这个陈述必定是粗略的, 因为任何精确的陈述都需要我们在此基础上进行定义.

广义地说, 在分析中出现的各种数之中, 有两种类型的数的无理性已经得到确认.

(a) *代数无理数.* $\sqrt{2}$ 的无理性是由 Pythagoras 或者他的学生证明的, 后来希腊数学家把这个结论推广到了 $\sqrt{3}$ 及其他的平方根. 现在容易证明: 一般来说, 对整数 m 和 N, $\sqrt[m]{N}$ 都是无理数. 更一般地, 由整系数代数方程所定义的数, 除了"明显" 是有理数的以外, 可以用 Gauss 的一个定理证明它们都是无理数. 我们要在 4.3 节中来证明这个定理 (定理 45).

(b) *数 e 和 π 以及由它们得出的数.* 容易证明 e 是无理数 (见 4.7 节). 证明很简单, 且只涉及该定理后来的推广中所含的最基本的思想. π 是无理数, 但对此并没有真正简单的证明. e 和 π 的所有幂以及 e 和 π 的有理系数多项式都是无理数. 像

$$\mathrm{e}^{\sqrt{2}}, \quad \mathrm{e}^{\sqrt{5}}, \quad \sqrt{7}\mathrm{e}^{\sqrt[3]{2}}, \quad \ln 2$$

这样的数都是无理数. 我们将在第 11 章中 (11.13 节至 11.14 节) 回过头来讨论这个问题.

直到 1929 年才发现了一些定理, 它们在所有重要的方面都超越了 11.13 节至 11.14 节中的那些结果. 最近又有人证明了还有某些种类的数也是无理数, 诸如数

$$\mathrm{e}^{\pi}, \quad 2^{\sqrt{2}}, \quad \mathrm{e}\pi$$

就位列其中. 而像

$$2^{\mathrm{e}}, \quad \pi^{\mathrm{e}}, \quad \pi^{\sqrt{2}}$$

以及 "Euler 常数"γ[①] 这样的数的无理性仍未得到证明.

4.3 Pythagoras 定理及其推广

首先要来证明

定理 43(Pythagoras 定理) $\sqrt{2}$ *是无理数.*

我们将对此定理给出两个证明. 这个定理及其最简单的推广 (虽然其价值微不足道) 仍值得深入研究. 古希腊关于比例的理论以同种的量一定是可公度的这一假设作为基础, 是 Pythagoras 的发现揭示了这个理论的缺陷, 从而为 Eudoxus 建立更为深入的理论 (见《几何原本》第 5 卷) 打通了道路.

(i) *第一个证明.* 如果 $\sqrt{2}$ 是有理数, 那么方程

$$a^2 = 2b^2 \tag{4.3.1}$$

就有整数解 $a, b, (a, b) = 1$. 故有 $b|a^2$, 于是对 b 的任何素因子 p 都有 $p|a^2$. 由此推出有 $p|a$. 既然有 $(a, b) = 1$, 这是不可能的. 从而有 $b = 1$, 而这显然也是错误的.

① $\gamma = \lim\limits_{n\to\infty}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \ln n\right)$.

(ii) 第二个证明. Pythagoras 的传统证明叙述如下. 由 (4.3.1) 可以看出 a^2 是偶数, 于是 a 也是偶数, 也即 $a = 2c$. 从而 $b^2 = 2c^2$, 且 b 也是偶数, 这与假设 $(a, b) = 1$ 矛盾.

这两个证明非常相似, 不过有一个重大的区别. (ii) 中考虑的是被一个给定的数 2 整除的性质. 显然, 如果 $2|a^2$, 则有 $2|a$, 这是因为奇数的平方必为奇数. 另一方面, (i) 中考虑的是被未知的素数 p 整除的性质, 且事实上假设了定理 3 成立. 所以从逻辑上讲 (ii) 是更为简单的证明. 然而, 下面就会看到, (i) 更有助于进行推广.

现在来证明更一般的定理.

定理 44 $\sqrt[m]{N}$ 是无理数, 除非 N 是一个整数 n 的 m 次幂.

(iii) 假设

$$a^m = Nb^m, \tag{4.3.2}$$

其中 $(a, b) = 1$. 则有 $b|a^m$, 于是对 b 的任何素因子 p 都有 $p|a^m$. 因此有 $p|a$, 由此与前一样得出有 $b = 1$. 可以看出这个证明与定理 43 的第一个证明几乎完全一样.

(iv) 为了不用定理 3 来对 $m = 2$ 证明定理 44, 假设

$$\sqrt{N} = a + \frac{b}{c},$$

其中 a, b, c 是整数, $0 < b < c$ 且 b/c 是使此式为真的具有最小分子的分数. 因此有

$$c^2N = (ca + b)^2 = a^2c^2 + 2abc + b^2,$$

故而 $c|b^2$, 也即有 $b^2 = cd$. 从而有

$$\sqrt{N} = a + \frac{b}{c} = a + \frac{d}{b}$$

以及 $0 < d < b$, 这是一对矛盾. 由此推得 $\sqrt{N}$ 是整数或者是无理数.

一个更为一般的定理是:

定理 45 如果 x 是首项系数为 1 的整系数方程

$$x^m + c_1x^{m-1} + \cdots + c_m = 0$$

的一个根, 那么 x 要么是整数, 要么是无理数.

特别地, 如果方程为

$$x^m - N = 0,$$

则定理 45 转化为定理 44.

显然可以假设 $c_m \neq 0$. 我们如上面的 (iii) 那样来进行讨论. 如果 $x = a/b$, 这里 $(a, b) = 1$, 那么

$$a^m + c_1a^{m-1}b + \cdots + c_mb^m = 0.$$

于是 $b|a^m$, 于是与前面一样推出 $b = 1$.

有可能对一般的 m 来证明定理 44, 而且不用定理 3 也可以证明定理 45, 不过这样的论证要稍微冗长一点.

4.4 基本定理在定理 43～45 证明中的应用

鉴于 4.5 节中关于历史的讨论, 所以应该特别注意在 4.3 节的证明、算术基本定理的证明或者 "等价的" 定理 3 的证明中所用到的东西.

定理 44 的证明 (iii) 中的关键推理是

$$"p|a^m \to p|a".$$

这里用到了定理 3. 同样的说明对定理 43 的第一个证明也适用, 唯一的简化是对 $m=2$ 的情形. 在这些证明中定理 3 起着至关重要的作用.

在定理 43 的第二个证明中情形有所不同, 因为这里考虑的是被特殊的数 2 整除的性质. 我们需要 "$2|a^2 \to 2|a$", 这可以用枚举法加以证明, 而不必求助于定理 3. 由于

$$(2s+1)^2 = 4s^2 + 4s + 1,$$

如我们已经说明过的, 奇数的平方是奇数, 由此即得结论.

对于任何特殊的 m 和 N, 可以用类似的枚举法来证明定理 44. 比方说, 假设 $m=2$, $N=5$. 我们需要 "$5|a^2 \to 5|a$". 现在任何不是 5 的倍数的数都有下列形式之一: $5m+1$, $5m+2$, $5m+3$, $5m+4$, 这些数的平方被 5 除的余数是 1, 4, 4, 1.

如果 $m=2$, $N=6$, 我们对 6 的最小素因子 2 来进行讨论, 其证明与定理 43 的第二个证明几乎完全一样. 对于 $m=2$ 和

$$N = 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 18,$$

用因子

$$d = 2, 3, 5, 2, 7, 4, 2, 11, 3, 13, 2, 3, 17, 2$$

来加以讨论: 在 N 是一个奇数倍数的情形, d 是 N 的最小素因子; 而在 $N=8$ 的情形, d 是这个素因子的一个适当的幂. 对于其中的某些情形实地进行证明是有益的, 仅仅是在 N 为素数时, 其证明才完全按照原来的格式进行, 而如果 N 的值很大, 证明会变得繁琐冗长.

可以类似地处理像 $m=3$, $N=2,3$ 或者 5 这样的情形, 但我们仅限于讨论 4.5 节至 4.6 节中所涉及的那些情形.

4.5 历 史 杂 谈

我们并不清楚是在什么时候以及由谁发现了 "Pythagoras 定理". Heath[①]说: "这个发现很难说是由 Pythagoras 本人做出的. 但这个发现肯定是在他的学派中做出的." Pythagoras 生活在大约公元前 570 至前 490 年. 诞生在大约 470 年[②]的 Dem-

① Thomas Heath, *A manual of Greek mathematics*, 54-55. 引号中所引用的内容, 除非特别指出是其他作者所作, 否则均取自这本书或者取自同一作者的*A history of Greek mathematics* 一书.

② 另一说为公元前 460 年. —— 译者注

ocritus 曾经写过 “在无理线以及立体上” 这样的话, 并且还说过 “很难拒绝 $\sqrt{2}$ 的无理性在 Democritus 的时代之前就已经被人发现的结论”.

看起来在超过 50 年的时间里没有对此定理作出推广. Plato 的*Theaetetus*这篇对话中有一段很著名的论述, 其中提到 Theodorus(Plato 的老师) 证明了

$$\sqrt{3}, \sqrt{5}, \cdots$$

的无理性, “(他) 取所有个别的情形一直做到 17 平方英尺的平方根, 就在这儿, 由于某种原因, 他止步不前, 停了下来”. 对此我们缺乏确切的信息, 而且我们对 Theodorus 的其他发现也一无所知, 但是 Plato 生活在公元前 429 至前 348 年, 因而这项发现的合理日期应该是在公元前大约 410 至前 400 年.

至于 Theodorus 如何证明他的定理, 这个问题使每个历史学家都绞尽了脑汁. 自然会猜想他是用了如同在 4.4 节里讨论过的 Pythagoras“传统” 方法的某种修改. 在那种情形中, 由于他不可能已经知道基本定理,[①] 而且他也不可能知道 Euclid 的定理 3 , 因而他可能像我们在 4.4 节末尾讨论的那样来进行论证. 对此的反对意见是 (反对意见系由 Zeuthen 和 Heath 这样的历史学家给出)：(i) 这个证明非常明显地采用了对于 $\sqrt{2}$ 的证明, 因而不应该被看成新东西; (ii) 早在证明 $\sqrt{17}$ 之前就明显可以看出, 这个证法是通用的. 然而, 对于这种观点, 应该注意到 Theodorus 不得不重新考虑每个不同的 d, 且处理 $\sqrt{11}$、$\sqrt{13}$ 以及 $\sqrt{17}$(在 $\sqrt{17}$ 之后还潜藏有 $\sqrt{19}$ 和 $\sqrt{23}$) 时的工作量非常大, 这才是公正的.

然而, 有关 Theodorus 的证明方法还有另外两个猜想. 这些方法非常复杂, 一个是在 $\sqrt{17}$, 另一个是在 $\sqrt{19}$. 它们中的哪一个与希腊词汇 $\mu\varepsilon\chi\rho\tau$[它被 Heath 翻译成 “直到”, 它的含义是指 “直到且不包含” 还是 “直到且包含”(“through” 一词的美国用法) 呢?] 的精确含义关系更密切呢? 正统的学者们告诉我前者更有可能, 如果是这样, 下面的由 McCabe 提出的方法就是一个很有可能的证法. 它的优点是本质上依赖于奇数和偶数之间的区别, 这在古希腊数学中是很重要的.

对 N 的连续值考虑 $\sqrt{N}$, 由于 Theodorus 已经处理了 $\sqrt{n}$, 所以他可能会忽略 $N = 4n$ 的情形. N 的其他偶数值形如 $2(2n+1)$, 而 $\sqrt{2}$ 的证明可以立即推广到这种情形. 这样一来, 我们只需要考虑 N 为奇数的情形. 对这样的 N, 如果 $\sqrt{N} = a/b$ 且 $(a, b) = 1$, 我们就有 $Nb^2 = a^2$, 且 a 和 b 两者必定均为奇数. 记 $a = 2A + 1$, $b = 2B + 1$, 于是就得到

$$N(2A + 1)^2 = (2B + 1)^2.$$

数 N 必定有下述形式之一：

$$4n + 3, \quad 8n + 5, \quad 8n + 1.$$

如果 $N = 4n + 3$, 将该等式乘开并除以 2 就得到

$$8nA(A + 1) + 6A(A + 1) + 2n + 1 = 2B(B + 1),$$

这是不可能的, 因为它的一边是奇数, 而另一边却是偶数. 如果 $N = 8n + 5$, 再次将

① 有关这一点的进一步讨论, 见 12.5 节.

该等式乘开并除以 4 就有

$$8nA(A+1)+5A(A+1)+2n+1=B(B+1),$$

这仍然是不可能的, 因为 $A(A+1)$ 和 $B(B+1)$ 都是偶数.

剩下的是形如 $8n+1$ 的数, 也即 $1,9,17,\cdots$, 其中 1 和 9 是平凡的, 困难首先出现在 $N=17$ 上. 如前面一样进行讨论, 得到方程

$$17(B^2+B)+4=A^2+A,$$

它的两边都是偶数. 这样就必须考虑多种可能性, 因而问题就变得复杂多了. (读者不妨动手尝试一下.) 因此, 如果这就是 Theodorus 的方法, 他会很自然地恰好在 $\sqrt{17}$ 之前止步.

Zeuthen 提出一个有意思的方法, 这个方法涉及经过几个变换后开始无限循环的比值, 这就引导出一个反证法. 这项工作一直延伸到 17 并包含 17, 而 18 当然是平凡的, 但是 19 在达到无限循环的链之前需要 8 个比值. 我们在 4.6 节中要给出他对 $\sqrt{5}$ 的证明. 但是, 即使 $\mu\varepsilon\chi\rho\tau$ 在这段文字中的含义是 "直到且包含", Plato 或许更有理由说过 "直到且包含 18". 总而言之, McCabe 的猜想看起来是最合理的.

4.6 $\sqrt{5}$ 无理性的几何证明

Zeuthen 提出的证法随着数的变化而变化. 其变化本质上依赖于表示 $\sqrt{N}$ 的周期连分数[①]的形式, 我们取最简单的情形 ($N=5$) 作为一个有代表性的例子.

用

$$x=\frac{1}{2}(\sqrt{5}-1)$$

来进行讨论. 这样就有

$$x^2=1-x.$$

从几何上说, 如果 $AB=1, AC=x$, 那么

$$AC^2=AB\cdot CB$$

且 AB 被 C 点划分成 "黄金分割比例". 这些关系在圆内接正五边形的构造中是基本的 (Euclid《几何原本》第 4 卷, 命题 11).

如果用 x 来除 1, 取最大可能的整数商, 也就是 1[②], 余数是 $1-x=x^2$. 如果用 x^2 来除 x, 商再次为 1, 而余数是 $x-x^2=x^3$. 接下去再用 x^3 来除 x^2, 并无限继续这个过程. 在每一步, 被除数、除数以及余数的比值都是同样的. 从几何上说, 如果取 CC_1 与 CB 相等且方向相反, CA 在 C_1 被分成的比例与 AB 在 C 被分成的比例相同, 也即黄金分割比. 如果取 C_1C_2 与 C_1A 相等且方向相反, 那么 C_1C 在 C_2

① 见 10.12 节.
② 因为 $1/2<x<1$.

被分成黄金分割比. 如此下去 (见图 4)[①]. 由于每一步我们都在处理被分成同样比例的线段, 故而这个过程是不可能终结的.

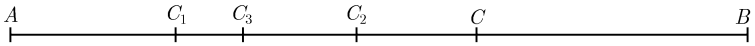

图 4

容易看出, 这与 x 的有理性的假设矛盾. 如果 x 是有理数, 那么 AB 和 AC 都是同一个长度 δ 的整倍数, 同样的结论对

$$C_1C = CB = AB - AC, \quad C_1C_2 = AC_1 = AC - C_1C, \cdots$$

也为真, 也就是说, 所有这些线段都在该图中. 因此可以构造一个由 δ 的整倍数组成的递减的无穷序列, 而这显然是不可能的.

4.7 更多的无理数

根据定理 44 可以知道, $\sqrt{7}, \sqrt[3]{2}, \sqrt[4]{11}\cdots$. 都是无理数. 根据定理 45, $x = \sqrt{2} + \sqrt{3}$ 是无理数, 这是因为它不是整数且满足方程

$$x^4 - 10x^2 + 1 = 0.$$

如同我们将在第 9 章和第 10 章中看到的那样, 可以利用十进制小数或者连分数任意地构造出无理数. 但是, 如果没有我们在 11.13 节和 11.14 节要证明的那些定理, 要想把在分析中自然出现的许多数添加到我们的无理数行列中来, 可不是一件容易的事.

定理 46 $\lg 2$ 是无理数.

这个结论是平凡的, 因为

$$\lg 2 = \frac{a}{b}$$

就蕴含 $2^b = 10^a$, 而这是不可能的. 更一般地, $\log_n m$ 是无理数, 如果 m 和 n 是整数, 且二者中的一个数有一个另一个数所没有的素因子.

定理 47 e 是无理数.

假设 e 是有理数, 比方说 $\mathrm{e} = a/b$, 其中 a 和 b 是整数. 如果 $k \geqslant b$ 且

$$\alpha = k!\left(\mathrm{e} - 1 - \frac{1}{1!} - \frac{1}{2!} - \cdots - \frac{1}{k!}\right),$$

那么 $b|k!$ 和 α 是一个整数. 但是

$$0 < \alpha = \frac{1}{k+1} + \frac{1}{(k+1)(k+2)} + \cdots < \frac{1}{k+1} + \frac{1}{(k+1)^2} + \cdots = \frac{1}{k},$$

① C_2C_3 与 C_2C 相等且方向相反, C_3C_4 与 C_3C_1 相等且方向相反, $\cdots\cdots$. 所定义的新线段交替地向左边和右边进行度量.

而这是一对矛盾.

在这个证明中, 假设定理不真, 从而推导出 α(i) 是整数, (ii) 是正数, (iii) 小于 1, 这就得到一个明显的矛盾. 通过对同样思想的更加复杂的应用再来证明两个进一步的定理.

对任意的正整数 n, 记

$$f = f(x) = \frac{x^n(1-x)^n}{n!} = \frac{1}{n!}\sum_{m=n}^{2n} c_m x^m,$$

其中 c_m 为整数. 对 $0 < x < 1$ 我们有

$$0 < f(x) < \frac{1}{n!}. \tag{4.7.1}$$

又有 $f(0) = 0$ 以及 $f^{(m)}(0) = 0$(如果 $m < n$ 或者 $m > 2n$). 但是, 如果 $n \leqslant m \leqslant 2n$, 那么

$$f^{(m)}(0) = \frac{m!}{n!}c_m$$

是一个整数. 因此 $f(x)$ 和它的所有导数在 $x = 0$ 时都取整数值. 由于 $f(1-x) = f(x)$, 故同样的结论对 $x = 1$ 也为真.

定理 48 对每个有理数 $y \neq 0$, e^y 都是无理数.

如果 $y = h/k$ 且 e^y 是有理数, 则 $\mathrm{e}^{ky} = \mathrm{e}^h$ 亦然. 再次, 如果 e^{-h} 是有理数, 则 e^h 亦然. 于是只要证明 "如果 h 是正整数, 则 e^h 不可能是有理数" 就够了. 假设此结论不真, 则有 $\mathrm{e}^h = a/b$, 其中 a 和 b 都是正整数. 记

$$F(x) = h^{2n}f(x) - h^{2n-1}f'(x) + \cdots - hf^{(2n-1)}(x) + f^{(2n)}(x),$$

从而 $F(0)$ 和 $F(1)$ 都是整数. 我们有

$$\frac{\mathrm{d}}{\mathrm{d}x}\left\{\mathrm{e}^{hx}F(x)\right\} = \mathrm{e}^{hx}\left\{hF(x) + F'(x)\right\} = h^{2n+1}\mathrm{e}^{hx}f(x).$$

于是

$$b\int_0^1 h^{2n+1}\mathrm{e}^{hx}f(x)\mathrm{d}x = b\left[\mathrm{e}^{hx}F(x)\right]\Big|_0^1 = aF(1) - bF(0)$$

是一个整数. 但由 (4.7.1) 知, 对足够大的 n 有

$$0 < b\int_0^1 h^{2n+1}\mathrm{e}^{hx}f(x)\mathrm{d}x < \frac{bh^{2n}\mathrm{e}^h}{n!} < 1,$$

这是一对矛盾.

定理 49 π 和 π^2 是无理数.

设 π^2 是有理数, 则有 $\pi^2 = a/b$, 其中 a 和 b 都是正整数. 记

$$G(x) = b^n\left\{\pi^{2n}f(x) - \pi^{2n-2}f''(x) + \pi^{2n-4}f^{(4)}(x) - \cdots + (-1)^n f^{(2n)}(x)\right\},$$

从而 $G(0)$ 和 $G(1)$ 都是整数. 有

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}x}\{G'(x)\sin\pi x-\pi G(x)\cos\pi x\}\\&=\{G''(x)+\pi^2G(x)\}\sin\pi x=b^n\pi^{2n+2}f(x)\sin\pi x\\&=\pi^2a^n\sin\pi xf(x).\end{aligned}$$

于是

$$\pi\int_0^1 a^n\sin\pi xf(x)\mathrm{d}x=\left[\frac{G'(x)\sin\pi x}{\pi}-G(x)\cos\pi x\right]_0^1=G(0)+G(1)$$

是一个整数. 但是由 (4.7.1) 知, 对足够大的 n 有

$$0<\pi\int_0^1 a^n\sin\pi xf(x)\mathrm{d}x<\frac{\pi a^n}{n!}<1,$$

这是一对矛盾.

本 章 附 注

4.2 节. e 和 π 的无理性是由 Lambert 在 1761 年证明的; 而 e^π 的无理性是由 Gelfond 在 1929 年证明的. 见第 11 章的 “本章附注”.

4.3 节至 4.6 节. 对希腊数学感兴趣的读者请参看 4.5 节中提到的 Heath 的书, 也见 van der Waerden, *Science Awakening*(Gronnigen, Nordhoff, 1954) 以及 Knorr, *Evolution of the Euclidean Elements*(Boston, Reidel, 1975). 有关 McCabe 关于 Theodorus 的证明方法的猜想, 请见 McCabe, *Math. Mag.* **49**(1976), 201–203.

我们并未给出专门的参考文献, 也不打算对希腊定理指定它们真正的发现者. 所以我们是在用 “Pythagoras” 来代表 “Pythagoras 学派的某些数学家”.

4.3 节. Alexander Oppenheim 爵士发现了定理 44 的证明 (iv)(由 R. Rado 教授作了改进), 而定理 45 的对应的证明参见 4.3 节的末尾. 在 Gauss, *D.A.* 一书第 42 章中对定理 45 以更一般的形式给出了证明.

4.7 节. 我们给出的定理 48 的证明基于 Hermite 的证明 (*Œuvres*, **3**, 154), 而我们给出的定理 49 的证明基于 Niven 的证明 (*Bulletin Amer. Math. Soc.* **53**(1947), 509).

根据定理 49,

$$\zeta(2)=\sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{\pi^2}{6}$$

是无理数, 又根据定理 205, $\zeta(4)=\dfrac{\pi^4}{90}$ 也是无理数, 且对所有正的偶数 m, $\zeta(m)$ 之值亦然. 然而, 当 m 取奇数值时却知之甚少. Apéry(1978) 证明了 $\zeta(3)$ 是无理数; 作为一个短小精悍的证明, 见 Beukers(*Bull. London Math. Soc.* **11**(1979), 268–272). 现在, 人们仍不知道 $\zeta(5)$ 是否是无理数. 不过, Ball 与 Rivaoal(*Inventiones Math.* **146**(2001), 193–207) 证明了: 序列 $\zeta(3),\zeta(5),\zeta(7),\zeta(9),\cdots\cdots$ 中含有无穷多个无理数.

第 5 章　同余和剩余

5.1　最大公约数和最小公倍数

我们已经定义了两个数 a 和 b 的最大公约数 (a,b). 关于这个数有一个简单的公式.

分别用 $\min(x,y)$ 和 $\max(x,y)$ 表示 x 和 y 中较小的和较大的那个数. 例如,

$$\min(1,2)=1,\quad \max(1,1)=1.$$

定理 50　*如果*

$$a=\prod_p p^{\alpha}\ (\alpha\geqslant 0),^{①}\quad b=\prod_p p^{\beta}\ (\beta\geqslant 0),$$

那么

$$(a,b)=\prod_p p^{\min(\alpha,\beta)}.$$

本定理是定理 2 以及最大公约数 (a,b) 的定义的直接推论.

两个整数 a 和 b 的最小公倍数(least common multiple) 是同时能被 a 和 b 整除的最小正数. 用 $\{a,b\}$ 来表示, 于是有

$$a|\{a,b\},\quad b|\{a,b\},$$

并且 $\{a,b\}$ 是有此性质的最小的数.

定理 51　*在定理 50 的记号下, 有*

$$\{a,b\}=\prod_p p^{\max(\alpha,\beta)}.$$

由定理 50 和定理 51 可以推出

定理 52　$\{a,b\}=\frac{ab}{(a,b)}$.

---

① 符号

$$\prod_p f(p)$$

表示取遍 p 的所有素数值的乘积. 而符号

$$\prod_{p|m} f(p)$$

则表示取遍所有整除 m 的素数的乘积. 在定理 50 的第一个公式中, 除非有 $p|a$, 否则相应的 α 等于 0 (从而该乘积中实际上只有有限项). 也可以将它写成

$$a=\prod_{p|a} p^{\alpha},$$

此时的每个 α 都是正数.

如果 $(a,b)=1$, 则称 a 与 b互素(coprime). 诸数 $a,b,c,\cdots,k$ 称为互素的, 如果其中任意两个数都互素. 说这些数是互素的要强于说

$$(a,\ b,\ c,\ \cdots,\ k)=1,$$

后者仅表示除了 1 以外, 不存在其他的数能同时整除 $a,b,c,\cdots,k$ 中所有的数.

有时候我们说 "a 和 b 没有公约数" 是指它们没有大于 1 的公约数, 也即它们互素.

5.2 同余和剩余类

如果 m 是 $x-a$ 的一个因子, 就说 x 和 a 关于模 m 同余, 并记为

$$x\equiv a\ (\mathrm{mod}\ m).$$

这个定义并没有引进任何新的思想, 因为 "$x\equiv a\ (\mathrm{mod}\ m)$" 和 "$m|(x-a)$" 有同样的含义, 但是每一种记号都有它自己的优点. 我们已经在 2.9 节中使用 "模" 这个词表示另外的意义, 但是这种多义性不会产生任何混淆[①].

用 $x\not\equiv a\ (\mathrm{mod}\ m)$ 表示 x 和 a 不同余.

如果 $x\equiv a\ (\mathrm{mod}\ m)$, 那么 a 就叫做 x 模 m 的一个剩余(residue). 若 $0\leqslant a\leqslant m-1$, 那么 a 称作是 x 模 m 的最小剩余(least residue)[②]. 因此, 关于模 m 同余的两个数 a 和 b 就有相同的剩余 $(\mathrm{mod}\ m)$. 模 m 的一个剩余类(class of residue)是由与某个给定的剩余 $(\mathrm{mod}\ m)$ 同余的所有数所组成的一个类, 这个类的每一个成员都叫做这个类的一个代表(representative). 显然, 总共有 m 个剩余类, 它们分别由

$$0,\ 1,\ 2,\ \cdots,\ m-1$$

作为代表. 这 m 个数组成的集合, 或者任何 m 个分别属于这 m 个剩余类的数组成的一个集合, 都称为模 m 的一个完全剩余系(complete system of incongruent residues to modulus m), 或简称为模 m 的一个完系(complete system).

同余在日常生活中具有极为重要的实用性. 比如, "今天是星期六" 就是从某个确定的日期开始所经过的天数关于模 7 的一个同余性质, 这个性质通常要比从某个时间点 (例如创世伊始) 开始所经过的天数重要得多. 课程表和列车时刻表同样也是同余表, 课程表中涉及的模是 365、7 和 24.

想知道发生了某个特定事件的某一天究竟是星期几, 实际上就是对模 7 解一个算术问题. 在这样的算术中, 同余的数是等价的, 因此这种算术完全是有限的系统, 其中所有的问题都可以通过尝试来获得解答. 例如, 一个讲座每两天举办一次 (包括星期天) 且第一次讲座在星期一举行, 那么第几次讲座首次在星期二举办呢? 如果这次讲座是第 $x+1$ 次, 那么有

① 一词双用是有意的, 这是因为 "关于一个数作成的模的同余" 这一概念要在这个理论的后面阶段中才会出现, 虽然我们在本书中不会用到这个概念.

② 严格地说, 应该是指最小非负剩余.

$$2x \equiv 1 \pmod 7,$$

通过尝试可以求得最小的正数解是

$$x = 4.$$

从而第五次讲座将会在星期二开讲, 而且这也将是第一次在星期二举办的讲座.

类似地, 可以用尝试法求得同余式

$$x^2 \equiv 1 \pmod 8$$

有 4 个解, 即为

$$x \equiv 1, 3, 5, 7 \pmod 8.$$

有时候, 即使出现的变量不是整数, 我们也使用同余符号, 这样做有时是很方便的. 比方说, 只要 $x - y$ 是 z 的整数倍, 就可以写成

$$x \equiv y \pmod z,$$

例如, 这样就有

$$\frac{3}{2} \equiv \frac{1}{2} \pmod 1, \quad -\pi \equiv \pi \pmod{2\pi}.$$

5.3 同余式的初等性质

显然, 对于给定的模 m, 同余式有如下性质:

(i) $a \equiv b \to b \equiv a$;

(ii) $a \equiv b, b \equiv c \to a \equiv c$;

(iii) $a \equiv a', b \equiv b' \to a + b \equiv a' + b'$.

又如果 $a \equiv a', b \equiv b', \cdots$, 就有

(iv) $ka + lb + \cdots \equiv ka' + lb' + \cdots$;

(v) $a^2 \equiv a'^2, a^3 \equiv a'^3$.

如此类推. 最后, 如果 $\phi(a, b, \cdots)$ 为任意的整数系数多项式, 就有

(vi) $\phi(a, b, \cdots) \equiv \phi(a', b', \cdots)$.

定理 53 如果 $a \equiv b \pmod m$ 以及 $a \equiv b \pmod n$, 那么 $a \equiv b \pmod{\{m, n\}}$. 特别, 如果 $(m, n) = 1$, 那么 $a \equiv b \pmod{mn}$.

这可以由定理 50 推出. 如果 p^c 是能够整除 $\{m, n\}$ 的 p 的最高幂, 那么 $p^c | m$ 或者 $p^c | n$, 于是有 $p^c | (a - b)$. 这对于 $\{m, n\}$ 的每个素因子来说都成立, 故而

$$a \equiv b \pmod{\{m, n\}}.$$

这条定理很容易推广到任意多个同余式的情形.

5.4 线性同余式

5.3 节介绍的性质 (i) 至性质 (vi) 与普通的代数方程的性质相像, 但是我们很快就会遇到它们之间的一个差别. 下面的性质在同余式中就未必成立:

$$ka \equiv ka' \to a \equiv a'.$$

比如

$$2 \times 2 \equiv 2 \times 4 \pmod 4,$$

但是

$$2 \not\equiv 4 \pmod 4.$$

接下来我们要研究在这个方向上有什么结果是成立的.

定理 54 如果 $(k, m) = d$, 那么

$$ka \equiv ka' \pmod m \to a \equiv a' \left(\mod \frac{m}{d}\right).$$

反过来也成立.

因为 $(k, m) = d$, 则有

$$k = k_1 d, \quad m = m_1 d, \quad (k_1, m_1) = 1.$$

那么

$$\frac{ka - ka'}{m} = \frac{k_1(a - a')}{m_1},$$

又因为 $(k_1, m_1) = 1$, 故有

$$m | ka - ka' \equiv m_1 | a - a'.^{①}$$

这就证明了定理. 特别地, 有

定理 55 如果 $(k, m) = 1$, 那么

$$ka \equiv ka' \pmod m \to a \equiv a' \pmod m.$$

反过来也成立.

定理 56 如果 $a_1, a_2, \cdots, a_m$ 是模 m 的一个完全剩余系, 且有 $(k, m) = 1$, 那么 $ka_1, ka_2, \cdots, ka_m$ 也是模 m 的一个完全剩余系.

根据定理 55, 由 $ka_i - ka_j \equiv 0 \pmod m$ 可以推导出 $a_i - a_j \equiv 0 \pmod m$, 这只有当 $i = j$ 时才可能成立. 更一般地, 如果 $(k, m) = 1$, 那么

$$ka_r + l \ (r = 1, 2, 3, \cdots, m)$$

也是模 m 的完全剩余系.

定理 57 如果 $(k, m) = d$, 那么

$$kx \equiv l \pmod m \tag{5.4.1}$$

有解, 当且仅当 $d | l$, 且有解时它恰有 d 个解. 特别地, 如果 $(k, m) = 1$, 那么该同余式只有一个解.

定理 57 中的同余式等价于

$$kx - my = l,$$

① 这里 '≡' 是逻辑等价的符号: 如果 P 和 Q 都是命题, 那么 $P \equiv Q$ 成立当且仅当 $P \to Q$ 以及 $Q \to P$ 成立.

因此, 这个结果部分地包含在定理 25 之中. 当我们说到同余式 "恰有 d 个" 解的时候, 自然理解成把同余的解看成是同样的解.

如果 $d=1$, 定理 57 就是定理 56 的推论. 如果 $d>1$, 则同余式 (5.4.1) 显然是不可解的, 除非有 $d|l$. 如果 $d|l$, 那么

$$m=dm', \quad k=dk', \quad l=dl',$$

故而该同余式等价于

$$k'x \equiv l' \pmod{m'}. \tag{5.4.2}$$

由于 $(k',m')=1$, 所以 (5.4.2) 恰有一个解. 如果这个解是

$$x \equiv t \pmod{m'},$$

那么

$$x=t+ym',$$

而 (5.4.1) 的完全解集就可以通过给 y 取所有的值来求得, 这里 y 的取值要使得诸 $t+ym'$ 关于模 m 互不同余.

由于

$$t+ym' \equiv t+zm' \pmod{m} \equiv m|m'(y-z) \equiv d|(y-z),$$

从而恰有 d 个解, 这些解可以表示成

$$t,\ t+m',\ t+2m',\ \cdots,\ t+(d-1)m'.$$

这就证明了定理.

5.5 Euler 函数 $\phi(m)$

用 $\phi(m)$ 来记不大于 m 的正整数中与 m 互素的整数的个数, 也就是说满足

$$0<n\leqslant m,\ (n,m)=1^{①}$$

的整数 n 的个数. 如果 a 与 m 互素, 那么任何一个与 a 同余 $\pmod m$ 的数 x 也与 m 互素. 于是有 $\phi(m)$ 个与 m 互素的剩余类, 从每个这样的剩余类中任取一个数所得到的任何一组 $\phi(m)$ 个剩余作成的集合都称为一个*与 m 互素的完全剩余系*(complete set of residues prime to m)[②] 一个这样的完全系是 $\phi(m)$ 个小于 m 且与 m 互素的数组成的集合.

定理 58 *如果 $a_1,a_2,\cdots,a_{\phi(m)}$ 是一个与 m 互素的完全剩余系, 且 $(k,m)=1$, 那么*

$$ka_1,ka_2,\cdots,ka_{\phi(m)}$$

仍然是一个与 m 互素的完全剩余系.

显然, 第二组数也都与 m 互素, 且如同定理 56 的证明中那样, 它们中没有任何两个数是同余的.

① 仅当 $m=1$ 时 n 才可能等于 m. 此时有 $\phi(1)=1$.

② 现代的数论著作中不再用这个术语, 而改称它是一个模 m 的缩剩余系(或简化剩余系). —— 译者注

定理 59 假设 $(m,m')=1$, 且 a 取遍模 m 的一个完全剩余系, a' 取遍模 m' 的一个完全剩余系. 那么 $a'm+am'$ 取遍模 mm' 的一个完全剩余系.

这里有 mm' 个数 $a'm+am'$. 如果

$$a'_1m+a_1m'\equiv a'_2m+a_2m'\ (\mathrm{mod}\ mm'),$$

那么

$$a_1m'\equiv a_2m'\ (\mathrm{mod}\ m),$$

所以

$$a_1\equiv a_2\ (\mathrm{mod}\ m).$$

类似地有

$$a'_1\equiv a'_2\ (\mathrm{mod}\ m').$$

从而这 mm' 个数都是互不同余的, 于是它们构成了模 mm' 的一个完全剩余系.

一个函数 $f(m)$ 称为是积性的, 如果 $(m,m')=1$ 就蕴含

$$f(mm')=f(m)f(m').$$

定理 60 $\phi(n)$ 是积性的.

如果 $(m,m')=1$, 那么根据定理 59 可知, 当 a 和 a' 分别取遍模 m 和模 m' 的完全剩余系时, $a'm+am'$ 取遍模 mm' 的一个完全剩余系. 又有

$$\begin{aligned}(a'm+am',mm')=1&\equiv(a'm+am',m)=1,\ (a'm+am',m')=1\\&\equiv(am',m)=1,\ (a'm,m')=1\\&\equiv(a,m)=1,\ (a',m')=1.\end{aligned}$$

从而这 $\phi(mm')$ 个小于 mm' 且与 mm' 互素的数是这 $\phi(m)\phi(m')$ 个数 $a'm+am'$ 的最小正剩余, 其中 a 与 m 互素, 而 a' 与 m' 互素, 从而有

$$\phi(mm')=\phi(m)\phi(m').$$

附带我们还证明了

定理 61 如果 $(m,m')=1$, a 取遍一个与 m 互素的完全剩余系, 而 a' 则取遍一个与 m' 互素的完全剩余系, 那么 $am'+a'm$ 取遍一个与 mm' 互素的完全剩余系.

现在可以对 m 的任意的值求出 $\phi(m)$ 的值. 根据定理 60 可见, 只要对 m 为素数幂的情形来计算 $\phi(m)$ 就行了. 小于 p^c 的正数一共有 p^c-1 个, 其中有 $p^{c-1}-1$ 个是 p 的倍数, 剩下的数均与 p 互素, 从而有

$$\phi\left(p^c\right)=p^c-1-\left(p^{c-1}-1\right)=p^c\left(1-\frac{1}{p}\right),$$

故而 $\phi(m)$ 的一般的值可由定理 60 得出.

定理 62 如果 $m=\prod p^c$, 那么

$$\phi(m) = m\prod_{p|m}\left(1-\frac{1}{p}\right).$$

我们将需要下面的结果.

定理 63 $\sum\limits_{d|m}\phi(d) = m.$

如果 $m = \prod p^c$, 那么 m 的因子就是诸数 $d = \prod p^{c'}$, 其中对每个 p 都有 $0 \leqslant c' \leqslant c$, 且根据 $\phi(m)$ 的积性性质有

$$\Phi(m) = \sum_{d|m}\phi(d) = \sum_{p,c'}\prod\phi\left(p^{c'}\right) = \prod_{p}\left\{1+\phi\left(p\right)+\phi\left(p^2\right)+\cdots+\phi\left(p^c\right)\right\}.$$

但是

$$1+\phi\left(p\right)+\cdots+\phi\left(p^c\right) = 1+(p-1)+p\left(p-1\right)+\cdots+p^{c-1}\left(p-1\right) = p^c,$$

从而有

$$\Phi(m) = \prod_{p}p^c = m.$$

5.6 定理 59 和定理 61 对三角和的应用

在数论中有某种重要的三角和, 它们要么是在 5.5 节的意义下是 “积性的”, 要么具有十分类似的性质.

记[①]

$$\mathrm{e}(\tau) = \mathrm{e}^{2\pi\mathrm{i}\tau},$$

我们只关心 τ 的有理值. 显然, 当 $m \equiv m' \pmod{n}$ 时有

$$\mathrm{e}\left(\frac{m}{n}\right) = \mathrm{e}\left(\frac{m'}{n}\right).$$

正是这个性质给出了三角和的算术重要性.

(1) Gauss 和的积性性质. Gauss 和定义为

$$S(m,n) = \sum_{h=0}^{n-1}\mathrm{e}^{2\pi\mathrm{i}h^2m/n} = \sum_{h=0}^{n-1}\mathrm{e}\left(\frac{h^2m}{n}\right),$$

它在二次剩余的理论中特别重要. 由于对任何 r 有

$$\mathrm{e}\left(\frac{(h+rn)^2m}{n}\right) = \mathrm{e}\left(\frac{h^2m}{n}\right),$$

故而只要 $h_1 \equiv h_2 \pmod{n}$, 就有

$$\mathrm{e}\left(\frac{h_1^2m}{n}\right) = \mathrm{e}\left(\frac{h_2^2m}{n}\right).$$

于是可以记

① 在本节里, e^{ζ} 都是复变量 ζ 的指数函数 $\mathrm{e}^{\zeta} = 1+\zeta+\cdots$. 假设读者了解指数函数的初等性质.

$$S(m,n)=\sum_{h(n)}\mathrm{e}\left(\frac{h^2m}{n}\right),$$

这个记号表示 h 取遍模 n 的任意一个完全剩余系. 当不致产生混淆时, 用 h 来代替 $h(n)$.

定理 64 如果 $(n,n')=1$, 那么

$$S(m,nn')=S(mn',n)S(mn,n').$$

设 h,h' 分别取遍模 n,n' 的完全剩余系. 那么, 根据定理 59 可知,

$$H=hn'+h'n$$

取遍模 nn' 的一个完全剩余系. 我们还有

$$mH^2=m\left(hn'+h'n\right)^2\equiv mh^2n'^2+mh'^2n^2\pmod{nn'}.$$

于是

$$\begin{aligned}S(mn',n)S(mn,n')&=\left\{\sum_{h}\mathrm{e}\left(\frac{h^2mn'}{n}\right)\right\}\left\{\sum_{h'}\mathrm{e}\left(\frac{h'^2mn}{n'}\right)\right\}\\&=\sum_{h,h'}\mathrm{e}\left(\frac{h^2mn'}{n}+\frac{h'^2mn}{n'}\right)=\sum_{h,h'}\mathrm{e}\left(\frac{m\left(h^2n'^2+h'^2n^2\right)}{nn'}\right)\\&=\sum_{H}\mathrm{e}\left(\frac{mH^2}{nn'}\right)=S(m,nn').\end{aligned}$$

(2) Ramanujan 和的积性性质. Ramanujan 和是

$$c_q(m)=\sum_{h^*(q)}\mathrm{e}\left(\frac{hm}{q}\right),$$

这里的记号表示 h 仅取遍与 q 互素的剩余类. 当不致产生混淆时, 我们有时用 h 来代替 $h^*(q)$.

可以将 $c_q(m)$ 表示成另外的形式, 其中引进了一个有更一般的重要性的记号. 称 ρ 是一个本原 q 次单位根(primitive q-th root of unity), 如果 $\rho^q=1$, 但是对 r 的任何小于 q 的正值, ρ^r 都不等于 1.

假设 $\rho^q=1$, 且 r 是使得 $\rho^r=1$ 成立的最小正整数, 那么 $q=kr+s$, 其中 $0\leqslant s<r$. 从而

$$\rho^s=\rho^{q-kr}=1,$$

所以有 $s=0$ 以及 $r\,|q$. 从而有

定理 65 任何 q 次单位根都是对 q 的某个因子 r 而言的一个本原 r 次单位根.

定理 66 q 次单位根是下列诸数

$$\mathrm{e}\left(\frac{h}{q}\right)\ (h=0,1,\cdots,q-1),$$

一个根是本原单位根的一个充分必要条件是 h 与 q 互素.

现在可以将 Ramanujan 和表成形式

$$c_q(m) = \sum \rho^m,$$

其中 ρ 取遍本原 q 次单位根.

定理 67 如果 $(q, q') = 1$, 那么

$$c_{qq'}(m) = c_q(m)c_{q'}(m).$$

因为根据定理 61 有

$$c_q(m)c_{q'}(m) = \sum_{h,h'} \mathrm{e}\left\{m\left(\frac{h}{q} + \frac{h'}{q'}\right)\right\} = \sum_{h,h'} \mathrm{e}\left\{\frac{m\,(hq' + h'q)}{qq'}\right\} = c_{qq'}(m).$$

(3) Kloosterman 和的积性性质. Kloosterman 和 (它要更困难一些) 是

$$S(u, v, n) = \sum_h \mathrm{e}\left(\frac{uh + v\hbar}{n}\right),$$

其中 h 取遍与 n 互素的一个完全剩余系, 而 $\hbar$ 定义为

$$h\hbar \equiv 1 \pmod{n}.$$

定理 57 表明: 给定任何 h, 则存在唯一的 $\hbar \pmod{n}$ 满足这个条件. 我们用不到
邮 Kloosterman 和, 但是对它的积性性质的证明过程极好地解释了前面几节中的思想.
电 **定理 68** 如果 $(n, n') = 1$, 那么

$$S(u, v, n)S(u, v', n') = S(u, V, nn'),$$

其中

$$V = vn'^2 + v'n^2.$$

如果

$$h\hbar \equiv 1 \pmod{n}, \quad h'\hbar' \equiv 1 \pmod{n'},$$

那么

$$\begin{aligned} S(u, v, n)S(u, v', n') &= \sum_{h,h'} \mathrm{e}\left(\frac{uh + v\hbar}{n} + \frac{uh' + v'\hbar'}{n'}\right) \\ &= \sum_{h,h'} \mathrm{e}\left\{u\left(\frac{hn' + h'n}{nn'}\right) + \frac{v\hbar n' + v'\hbar' n}{nn'}\right\} \\ &= \sum_{h,h'} \mathrm{e}\left(\frac{uH + K}{nn'}\right), \end{aligned} \tag{5.6.1}$$

其中

$$H = hn' + h'n, \quad K = v\hbar n' + v'\hbar' n.$$

根据定理 61, H 取遍与 nn' 互素的一个完全剩余系. 于是, 如果能证明

$$K \equiv V\overline{H} \pmod{nn'}, \tag{5.6.2}$$

其中 $\overline{H}$ 定义为

$$H\overline{H} \equiv 1 \pmod{nn'},$$

那么 (5.6.1) 将被化简成

$$S(u,v,n)S(u,v',n') = \sum_H \mathrm{e}\left(\frac{uH+V\overline{H}}{nn'}\right) = S(u,V,nn').$$

现在有

$$(hn'+h'n)\overline{H} = H\overline{H} \equiv 1 \pmod{nn'}.$$

从而

$$hn'\overline{H} \equiv 1 \pmod{n}, \quad n'\overline{H} \equiv \hbar hn'\overline{H} \equiv \hbar \pmod{n},$$

所以有

$$n'^2\overline{H} \equiv n'\hbar \pmod{nn'}. \tag{5.6.3}$$

类似地, 可以看出

$$n^2\overline{H} \equiv n\hbar' \pmod{nn'}, \tag{5.6.4}$$

而由 (5.6.3) 和 (5.6.4) 可以推出

$$V\overline{H} = (vn'^2+v'n^2)\overline{H} \equiv vn'\hbar + v'n\hbar' \equiv K \pmod{nn'}.$$

这就是 (5.6.2), 由此得出定理.

5.7 一个一般性的原理

我们暂时回到证明定理 65 时用到的论证方法. 如果我们用更一般的方式来对这个定理及其证明重新加以表述, 以后就会避免掉许多重复. 用 $P(a)$ 来记断言非负整数 a 所具有的某个性质的任意命题.

定理 69 如果

(i) 对每个 a 和 b(只要在第二种情形下有 $b \leqslant a$ 即可), $P(a)$ 和 $P(b)$ 就蕴含 $P(a+b)$ 和 $P(a-b)$;

(ii) r 是使得 $P(r)$ 成立的最小的正整数;

那么

(a) 对每个非负整数 k, $P(kr)$ 也为真;

(b) 任何使得 $P(q)$ 成立的 q 都是 r 的倍数.

首先 (a) 是显然的.

为证明 (b), 注意到, 根据 r 的定义有 $0 < r \leqslant q$. 从而可以记

$$q = kr + s, \quad s = q - kr,$$

其中 $k \geqslant 1$ 且 $0 \leqslant s < r$. 但是根据 (a) 有 $P(r) \to P(kr)$, 从而根据 (i) 就有

$$P(q), \quad P(kr) \to P(s).$$

故而再次利用 r 的定义知, s 必须为 0, 且有 $q = kr$.

我们还能从定理 23 推导出定理 69. 在定理 65 中, $P(a)$ 是 $\rho^a = 1$.

5.8 正十七边形的构造

我们将简要地补充介绍初等几何的一个著名问题, 也就是正 n 边形 (或者说内角为 $\alpha = 2\pi/n$ 的正多边形) 的构造问题, 以此来结束本章.

假设 $(n_1, n_2) = 1$, 并假设该问题对 $n = n_1$ 以及 $n = n_2$ 均可解. 则存在整数 r_1 和 r_2 使得

$$r_1 n_1 + r_2 n_2 = 1,$$

或者

$$r_1\alpha_2 + r_2\alpha_1 = r_1\frac{2\pi}{n_2} + r_2\frac{2\pi}{n_1} = \frac{2\pi}{n_1 n_2}.$$

因此, 如果该问题对 $n = n_1$ 以及 $n = n_2$ 均可解, 它就对 $n = n_1 n_2$ 也可解. 由此可知, 只需要考虑 n 是一个素数幂的情况即可. 下面假设 $n = p$ 是素数.

如果能够构造出 $\cos\alpha$(或者 $\sin\alpha$), 就能构造出 α. 诸数

$$\cos k\alpha + i\sin k\alpha \ (k = 1, 2, \cdots, n-1)$$

是

$$\frac{x^n - 1}{x - 1} = x^{n-1} + x^{n-2} + \cdots + 1 = 0 \tag{5.8.1}$$

的根. 所以, 如果能作出 (5.8.1) 的根, 那么就能作出 α 了.

从分析上来说, "Euclid" 作图法 (即用直尺和圆规作图) 等价于求解一系列的线性方程或者二次方程.[①] 因此, 如果能将 (5.8.1) 的求解问题转化成一系列这样的方程, 那么相应的作图就是可能的.

这个问题被 Gauss 解决, 他证明了 (详见 2.4 节): 这种转化是可能的, 当且仅当 n 是一个 "Fermat 素数"[②]

$$n = p = 2^{2^h} + 1 = F_h.$$

h 的前面 5 个值, 也即 0, 1, 2, 3, 4, 给出

$$n = 3, 5, 17, 257, 65\,537,$$

它们全都是素数, 因而在这些情形下, 该问题是可解的.

对于 $n = 3$ 和 $n = 5$, 相应的正三边形和正五边形的构造法是熟知的. 这里给出 $n = 17$ 的构造法. 我们不打算对 Gauss 的理论给出系统的说明, 但是这个特定的构造法对于他的方法步骤给出了一个恰当的例子, 读者应当明白 (从一开始这就是合情合理的): 当 $n = p$ 且 $p - 1$ 不含有除了 2 以外的任何其他素数因子时, Euclid 作图法是能够完成这一构造的. 这就要求 p 是形如 $2^m + 1$ 的素数, 而仅有的这种特征的素数就是 Fermat 素数.[③]

然后假设 $n = 17$. 对应的方程是

---

① 见 11.5 节.
② 见 2.5 节.
③ 见 2.5 节定理 17.

$$\frac{x^{17}-1}{x-1}=x^{16}+x^{15}+\cdots+1=0. \tag{5.8.2}$$

记

$$\alpha=\frac{2\pi}{17},\quad \varepsilon_k=\mathrm{e}\left(\frac{k}{17}\right)=\cos k\alpha+\mathrm{i}\sin k\alpha,$$

所以 (5.8.2) 的根是

$$x=\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_{16}. \tag{5.8.3}$$

由这些根可以形成一定的和, 它们称为周期(period), 它们皆为二次方程的根.

诸数

$$3^m\ (0\leqslant m\leqslant 15)$$

按照某种次序分别与 $k=1,2,\cdots,16$ 同余 $(\mathrm{mod}\ 17)$,[①]对应关系如下:

$m=0,$	1,	2,	3,	4,	5,	6,	7,	8,	9,	10,	11,	12,	13,	14,	15,
$k=1,$	3,	9,	10,	13,	5,	15,	11,	16,	14,	8,	7,	4,	12,	2,	6.

用

$$x_1=\sum_{2\mid m}\varepsilon_k=\varepsilon_1+\varepsilon_9+\varepsilon_{13}+\varepsilon_{15}+\varepsilon_{16}+\varepsilon_8+\varepsilon_4+\varepsilon_2,$$

$$x_2=\sum_{2\nmid m}\varepsilon_k=\varepsilon_3+\varepsilon_{10}+\varepsilon_5+\varepsilon_{11}+\varepsilon_{14}+\varepsilon_7+\varepsilon_{12}+\varepsilon_6.$$

来定义 x_1 和 x_2, 而用

$$y_1=\sum_{m\equiv 0\ (\mathrm{mod}\ 4)}\varepsilon_k=\varepsilon_1+\varepsilon_{13}+\varepsilon_{16}+\varepsilon_4,$$

$$y_2=\sum_{m\equiv 2\ (\mathrm{mod}\ 4)}\varepsilon_k=\varepsilon_9+\varepsilon_{15}+\varepsilon_8+\varepsilon_2,$$

$$y_3=\sum_{m\equiv 1\ (\mathrm{mod}\ 4)}\varepsilon_k=\varepsilon_3+\varepsilon_5+\varepsilon_{14}+\varepsilon_{12},$$

$$y_4=\sum_{m\equiv 3\ (\mathrm{mod}\ 4)}\varepsilon_k=\varepsilon_{10}+\varepsilon_{11}+\varepsilon_7+\varepsilon_6.$$

来定义 y_1,y_2,y_3,y_4. 由于

$$\varepsilon_k+\varepsilon_{17-k}=2\cos k\alpha,$$

故而有

$$\begin{aligned}
&x_1=2\left(\cos\alpha+\cos 8\alpha+\cos 4\alpha+\cos 2\alpha\right),\\
&x_2=2\left(\cos 3\alpha+\cos 7\alpha+\cos 5\alpha+\cos 6\alpha\right),\\
&y_1=2\left(\cos\alpha+\cos 4\alpha\right),\quad y_2=2\left(\cos 8\alpha+\cos 2\alpha\right),\\
&y_3=2\left(\cos 3\alpha+\cos 5\alpha\right),\quad y_4=2\left(\cos 7\alpha+\cos 6\alpha\right).
\end{aligned}$$

首先证明 x_1 和 x_2 是一个有有理系数的二次方程的根. 由于 (5.8.2) 的根是 (5.8.3) 中诸数, 所以有

① 事实上, 在即将在 6.8 节中解释的那种意义下, 3 是 “17 的一个原根”.

$$x_1+x_2=2\sum_{k=1}^{8}\cos k\alpha=\sum_{k=1}^{16}\varepsilon_k=-1.$$

又有

$$x_1x_2=4\left(\cos\alpha+\cos 8\alpha+\cos 4\alpha+\cos 2\alpha\right)\times\left(\cos 3\alpha+\cos 7\alpha+\cos 5\alpha+\cos 6\alpha\right).$$

如果把它的右边乘开来, 并利用恒等式

$$2\cos m\alpha\cos n\alpha=\cos(m+n)\alpha+\cos(m-n)\alpha, \tag{5.8.4}$$

就可以得到

$$x_1x_2=4(x_1+x_2)=-4.$$

因此, x_1 和 x_2 是

$$x^2+x-4=0 \tag{5.8.5}$$

的根. 又有

$$\cos\alpha+\cos 2\alpha>2\cos\frac{1}{4}\pi=\sqrt{2}>-\cos 8\alpha,\quad \cos 4\alpha>0.$$

从而 $x_1>0$ 且

$$x_1>x_2. \tag{5.8.6}$$

接下来, 证明 y_1,y_2 和 y_3,y_4 都是关于 x_1 和 x_2 的、系数为有理数的二次方程的根. 因为

$$y_1+y_2=x_1,$$

再次利用 (5.8.4) 有

$$\begin{aligned}y_1y_2&=4\left(\cos\alpha+\cos 4\alpha\right)\left(\cos 8\alpha+\cos 2\alpha\right)\\&=2\sum_{k=1}^{8}\cos k\alpha=-1.\end{aligned}$$

于是, y_1,y_2 是方程

$$y^2-x_1y-1=0 \tag{5.8.7}$$

的根, 而且显然有

$$y_1>y_2. \tag{5.8.8}$$

类似地, 有

$$y_3+y_4=x_2,\quad y_3y_4=-1,$$

所以 y_3,y_4 是方程

$$y^2-x_2y-1=0 \tag{5.8.9}$$

的根, 且有

$$y_3>y_4. \tag{5.8.10}$$

最后有

$$2\cos\alpha+2\cos 4\alpha=y_1,$$

$$4\cos\alpha\cos 4\alpha=2\left(\cos 5\alpha+\cos 3\alpha\right)=y_3.$$

又有 $\cos\alpha>\cos 4\alpha$. 因此 $z_1=2\cos\alpha$ 和 $z_2=2\cos 4\alpha$ 就是二次方程

$$z^2-y_1z+y_3=0 \tag{5.8.11}$$

的根, 且有

$$z_1 > z_2. \tag{5.8.12}$$

现在可以通过求解 4 个二次方程 (5.8.5)、(5.8.7)、(5.8.9) 以及 (5.8.11), 并记住相关的不等式, 从而可以定出 $z_1 = 2\cos\alpha$ 的值. 得到

$$\begin{aligned} 2\cos\alpha = {} & \frac{1}{8}\left\{-1+\sqrt{17}+\sqrt{34-2\sqrt{17}}\right\} \\ & +\frac{1}{8}\sqrt{68+12\sqrt{17}-16\sqrt{34+2\sqrt{17}}-2\left(1-\sqrt{17}\right)\sqrt{34-2\sqrt{17}}}, \end{aligned}$$

这是一个仅包含有理数以及平方根的表达式. 现在这个数就可以仅用直尺和圆规构造出来, 从而 α 可以 (用直尺和圆规) 构造出来.

还有一个更简单的几何作图法. 设 $\angle C$ 是使得 $\tan 4C = 4$ 成立的最小的正的锐角, 从而 $\angle C$、$2\angle C$ 以及 $4\angle C$ 全都是锐角. 这样一来, (5.8.5) 就可以写成

$$x^2 + 4x\cot 4C - 4 = 0.$$

这个方程的根是 $2\tan 2C$ 和 $-2\cot 2C$. 由于 $x_1 > x_2$, 这就给出 $x_1 = 2\tan 2C$ 以及 $x_2 = -2\mathrm{cot}2C$. 代入 (5.8.7) 和 (5.8.9) 中, 并求解方程即得

$$y_1 = \tan\left(C+\frac{1}{4}\pi\right), \quad y_3 = \tan C,$$

$$y_2 = \tan\left(C-\frac{1}{4}\pi\right), \quad y_4 = -\mathrm{cot}C.$$

于是有

$$\begin{cases} 2\cos 3\alpha + 2\cos 5\alpha = y_3 = \tan C, \\ 2\cos 3\alpha \cdot 2\cos 5\alpha = 2\cos 2\alpha + 2\cos 8\alpha = y_2 = \tan\left(C-\frac{1}{4}\pi\right). \end{cases} \tag{5.8.13}$$

现在设 OA, OB(图 5) 是一个圆中两条互相垂直的半径. 取 OI 是 OB 的 $\frac{1}{4}$, 且 $\angle OIE$(E 在 OA 上) 是 $\angle OIA$ 的 $\frac{1}{4}$. 在 AO 上求一个点 F 使得 $\angle EIF = \frac{1}{4}\pi$. 设以 AF 为直径的圆与 OB 相交于 K, 并设中心在 E、半径为 EK 的圆与 OA 相交于 N_3 和 N_5(N_3 在 OA 上, N_5 在 AO 上). 画出与 OA 垂直的 N_3P_3, N_5P_5, 它们与原来的圆的圆周交于 P_3 和 P_5.

这样就有 $\angle OIA = 4\angle C$ 以及 $\angle OIE = \angle C$. 又有

$$2\cos\angle AOP_3 + 2\cos\angle AOP_5 = 2\frac{ON_3 - ON_5}{OA} = \frac{4OE}{OA} = \frac{OE}{OI} = \tan C,$$

$$\begin{aligned} 2\cos\angle AOP_3 \cdot 2\cos\angle AOP_5 &= -4\frac{ON_3 \cdot ON_5}{OA^2} = -4\frac{OK^2}{OA^2} \\ &= -4\frac{OF}{OA} = -\frac{OF}{OI} = \tan\left(C-\frac{1}{4}\pi\right). \end{aligned}$$

将这些方程与 (5.8.13) 比较, 可以看出 $\angle AOP_3 = 3\alpha$ 以及 $\angle AOP_5 = 5\alpha$. 由此推出, A, P_3, P_5 是圆内接正十七边形的第一、第四和第六个顶点, 于是怎样构造出这个正

多边形就是显而易见的事了.

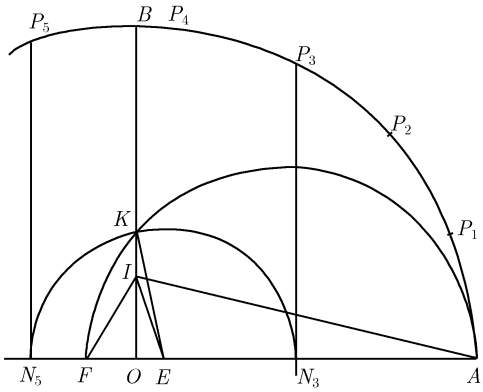

图 5

本 章 附 注

5.1 节. 这一章的内容全部都是 "经典的"(除了在 5.6 节中证明的 Ramanujan 和以及 Kloosterman 和的性质之外), 它们均可在教科书中找到. 同余式的理论首先是由 Gauss, *D.A.* 系统所发展出来的, 虽然其中主要的结果已经为像 Fermat 和 Euler 这样的更早期的数学家所知. 我们偶尔给出一些参考资料, 特别是当某个有名的函数或者定理习惯上与一个特殊的数学家的名字相连在一起时, 但是不打算给出系统的阐述.

5.5 节. Euler, *Novi Comm. Acad. Petrop.* **8** (1760–1761), 74–104 [*Opera* (1), ii. 531–544].

看起来更为自然的是称 $f(m)$ 为积性的, 如果对所有 m, m' 都有

$$f(mm') = f(m)f(m').$$

但这个定义限制性太强, 正文中给出的较为宽松的定义要有用得多.

5.6 节. 这一节里的和出现在 Gauss, "*Summatio quarumdam serierum singularium*" (1808), *Werke*, ii. 11–45; Ramanujan, *Trans. Camb. Phil. Soc.* **22** (1918), 259–276 (*Collected Papers*, 179–199); Kloosterman, *Acta Math.* **49** (1926), 407–464 之中; "Ramanujan 和" 可以在更早期的论著中找到; 例如, 见 Jensen, *Beretning d. tredje Skand. Matematikercongres* (1913), 145 以及 Landau, *Handbuch*, 572. 但是 Ramanujan 是看到它的重要性并系统地应用它的第一位数学家. 这种和在用平方和来表示数的理论中是特别重要的. 有关 Gauss 和的计算、它们的应用以及历史, 见 Davenport, *Multiplicative number theory*, (Markham, Chicago, 1967), 有关 Kloosterman 和的信息以及参考文献, 见 Weil, *Proc. Nat. Acad. Sci. U.S.A.* **34** (1948), 204–207.

5.8 节. 一般的理论是由 Gauss, *D. A.*, 第 335 章 –366 章发展起来的. 正十七边形的第一个明显的几何作图是由 Erchinger 给出的 (见 Gauss, *Werke*, ii. 186–187). 正文中给出的作图法属于 Richmond, *Quarterly Journal of Math.* **26** (1893), 206–207 以及 *Math. Annalen*,

67 (1909), 459–461. 我们的图取自 Richmond 的论文.

Gauss (*D.A.*, 第 341 章) 证明了：方程 (5.8.1) 是不可约的, 也就是说, 当 n 为素数时, 它的左边不可能分解成更低次数的有理系数因子之积. 更加一般地, Kronecker 和 Eisenstein 证明了：由 $\phi(n)$ 个本原 n 次单位根所满足的方程是不可约的. 可参见 Mathews, *Theory of numbers* (Cambridge, Deighton Bell, 1892), 186–188. Grandjot 指出了, 该定理可以从 Dirichlet 的定理 15 很容易地推导出来：见 Landau, *Vorlesungen*, iii. 219.

第 6 章　Fermat 定理及其推论

6.1　Fermat 定理

本章要运用第 5 章里的一般性思想来证明主要是属于 Fermat、Euler、Legendre 以及 Gauss 的一系列经典定理.

定理 70　*如果 p 是素数, 那么*

$$a^p \equiv a \pmod{p}. \tag{6.1.1}$$

定理 71(Fermat 定理)　*如果 p 是素数, 且 $p \nmid a$, 那么*

$$a^{p-1} \equiv 1 \pmod{p}. \tag{6.1.2}$$

当 $p \nmid a$ 时, 同余式 (6.1.1) 和 (6.1.2) 是等价的; 而当 $p|a$ 时, (6.1.1) 是平凡的, 这是因为此时有 $a^p \equiv 0 \equiv a$. 因此定理 70 与定理 71 是等价的.

定理 71 是更为一般的定理 72 的特殊情形.

定理 72(Fermat-Euler 定理)　*如果 $(a, m) = 1$, 那么*

$$a^{\phi(m)} \equiv 1 \pmod{m}.$$

如果 x 取遍与 m 互素的完全剩余系[①], 那么, 根据定理 58, ax 也取遍这样一个 (简化) 剩余系. 取每个 (简化) 剩余系中诸数之乘积, 于是就有

$$\prod(ax) \equiv \prod x \pmod{m},$$

也即

$$a^{\phi(m)} \prod x \equiv \prod x \pmod{m}.$$

由于每个数 x 皆与 m 互素, 故而它们的乘积也与 m 互素, 于是根据定理 55, 有

$$a^{\phi(m)} \equiv 1 \pmod{m}.$$

如果 $(a, m) > 1$, 这个结果显然不成立.

6.2　二项系数的某些性质

Euler 是发表了对于 Fermat 定理的证明的第一人. 这个证明 (很容易加以拓展来证明定理 72) 依赖于二项系数的最简单的算术性质.

定理 73　*如果 m 和 n 是正整数, 那么二项系数*

① 用现在的数论语言可以改述成 x 取遍 “模 m 的简化剩余系”, 也就是取遍 “模 m 的缩系”. —— 译者注

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

都是整数.

这是我们需要的这个定理的第一部分, 但是, 由于

$$\binom{-m}{n}=(-1)^n\binom{m+n-1}{n},$$

因此这两部分是等价的. 也就是, 每一部分都可以表述成一种引人瞩目的格式.

定理 74 任何 n 个连续正整数的乘积均可被 $n!$ 整除.

这些定理显然起源于二项式系数, 即 $(1+x)(1+x)\cdots$ 或者

$$(1-x)^{-1}(1-x)^{-1}\cdots=(1+x+x^2+\cdots)(1+x+x^2+\cdots)\cdots$$

中的 x 的幂的系数. 可以如下用归纳法来证明它们. 选取定理 74, 该定理断言

$$(m)_n=m(m+1)\cdots(m+n-1)$$

可以被 $n!$ 整除. 这对 $n=1$ 和所有的 m 显然为真, 对 $m=1$ 和所有 n 也显然为真. 假设 (a) 对 $n=N-1$ 以及所有 m 为真, (b) 对 $n=N$ 以及 $m=M$ 为真. 那么

$$(M+1)_N-M_N=N(M+1)_{N-1},$$

故而 $(M+1)_{N-1}$ 能被 $(N-1)!$ 整除. 从而 $(M+1)_N$ 能被 $N!$ 整除, 从而定理对 $n=N$ 以及 $m=M+1$ 为真. 由此推出定理对 $n=N$ 以及所有 m 为真. 由于它对 $n=N+1$ 以及 $m=1$ 也为真, 我们可以重复这个讨论, 从而该定理结论成立.

定理 75 如果 p 是素数, 那么

$$\binom{p}{1},\ \binom{p}{2},\ \cdots,\ \binom{p}{p-1}$$

均能被 p 整除.

如果 $1\leqslant n\leqslant p-1$, 那么由定理 74 得

$$n!|p(p-1)\cdots(p-n+1).$$

但是 $n!$ 与 p 互素, 故有

$$n!|(p-1)(p-2)\cdots(p-n+1).$$

从而

$$\binom{p}{n}=p\frac{(p-1)(p-2)\cdots(p-n+1)}{n!}$$

能被 p 整除.

定理 76 如果 p 是素数, 那么 $(1-x)^{-p}$ 中除了 $1,x^p,x^{2p},\cdots$ 之外, 其余所有项的系数都能被 p 整除, 而 $1,x^p,x^{2p},\cdots$ 的系数均同余于 1 (mod p).

根据定理 73,

$$(1-x)^{-p}=1+\sum_{n=1}^{\infty}\binom{p+n-1}{n}x^n$$

中的系数全都是整数. 由于

$$(1-x^p)^{-1} = 1 + x^p + x^{2p} + \cdots,$$

故而不得不证明

$$(1-x^p)^{-1} - (1-x)^{-p} = (1-x)^{-p}(1-x^p)^{-1}\{(1-x)^p - 1 + x^p\}$$

的展开式中的每一个系数都能被 p 整除. 由于 $(1-x)^{-p}$ 和 $(1-x^p)^{-1}$ 的展开式中的系数都是整数, 所以只要证明多项式 $(1-x)^p - 1 + x^p$ 中的每个系数都能被 p 整除就足够了. 对于 $p=2$, 这是显然的; 而对于 $p \geqslant 3$, 由于

$$(1-x)^p - 1 + x^p = \sum_{r=1}^{p-1} (-1)^r \binom{p}{r} x^r,$$

故此时的结论可以由定理 75 推出.

第 19 章中将需要这个定理.

定理 77 如果 p 是素数, 那么

$$(x + y + \cdots + w)^p \equiv x^p + y^p + \cdots + w^p \pmod{p}.$$

因为根据定理 75 有

$$(x+y)^p \equiv x^p + y^p \pmod{p},$$

故而一般性的结果可以通过重复使用这个结论而得到.

定理 75 的另外一个有用的推论是:

定理 78 如果 $\alpha > 0$ 且

$$m \equiv 1 \pmod{p^\alpha},$$

那么

$$m^p \equiv 1 \pmod{p^{\alpha+1}}.$$

因为 $m = 1 + kp^\alpha$, 其中 k 是一个整数, 且 $\alpha p \geqslant \alpha + 1$. 这样就有

$$m^p = (1 + kp^\alpha)^p = 1 + lp^{\alpha+1},$$

其中 l 是一个整数.

6.3 定理 72 的第二个证明

现在可以对定理 72 给出 Euler 的证明. 假设 $m = \prod p^\alpha$. 根据定理 53, 只需要证明

$$a^{\phi(m)} \equiv 1 \pmod{p^\alpha}$$

就够了. 但是

$$\phi(m) = \prod \phi(p^\alpha) = \prod p^{\alpha-1}(p-1),$$

所以只需证明当 $p \nmid a$ 时有

$$a^{p^{\alpha-1}(p-1)} \equiv 1 \pmod{p^\alpha}$$

即可.

根据定理 77 有

$$(x+y+\cdots)^p \equiv x^p + y^p + \cdots \pmod{p}.$$

取 $x=y=z=\cdots=1$, 并假设其中有 a 个数, 可以得到

$$a^p \equiv a \pmod{p},$$

这也就是

$$a^{p-1} \equiv 1 \pmod{p}.$$

于是, 根据定理 78 就有

$$a^{p(p-1)} \equiv 1 \pmod{p^2},\ a^{p^2(p-1)} \equiv 1 \pmod{p^3},\ \cdots,\ a^{p^{\alpha-1}(p-1)} \equiv 1 \pmod{p^\alpha}.$$

6.4 定理 22 的证明

在着手 Fermat 定理更为重要的应用之前, 先用它来证明第 2 章中的定理 22.

可以将 $f(n)$ 写成形式

$$f(n) = \sum_{r=1}^{m} Q_r(n) a_r^n = \sum_{r=1}^{m} \left(\sum_{s=0}^{q_r} c_{r,s} n^s \right) a_r^n,$$

其中诸数 a 和 c 皆为整数, 且

$$1 \leqslant a_1 < a_2 < \cdots < a_m.$$

对于很大的 n, $f(n)$ 中的项按照大小递增的次序排列, 当 n 很大时, $f(n)$ 的大小被它的最后一项

$$c_{m,q_m} n^{q_m} a_m^n$$

所控制 (最后那个系数 c 是正数).

如果对所有很大的 n, $f(n)$ 都是素数, 那么就存在一个 n, 使得

$$f(n) = p > a_m,$$

这里 p 是素数. 那么, 对所有整数 k 和 s 有

$$\{n + kp(p-1)\}^s \equiv n^s \pmod{p}.$$

又由 Fermat 定理有

$$a_r^{p-1} \equiv 1 \pmod{p},$$

故而对所有正整数 k 有

$$a_r^{n+kp(p-1)} \equiv a_r^n \pmod{p}.$$

从而

$$\{n + kp(p-1)\}^s a_r^{n+kp(p-1)} \equiv n^s a_r^n \pmod{p},$$

这样一来, 对所有正整数 k 就有

$$f\{n + kp(p-1)\} \equiv f(n) \equiv 0 \pmod{p}.$$

这是一对矛盾.

6.5 二次剩余

假设 p 是一个奇素数, $p \nmid a$, 且 x 是诸数

$$1, 2, 3, \cdots, p-1$$

中的一个. 那么, 根据定理 58, 诸数

$$1 \cdot x, 2 \cdot x, \cdots, (p-1)x$$

中恰有一个与 a 同余 $(\bmod p)$. 于是存在唯一的 x' 使得

$$xx' \equiv a \pmod{p}, \quad 0 < x' < p.$$

称 x' 是 x 的相伴数(associate). 这样就有两种可能性: 或者有至少一个 x 与自己相伴, 从而有 $x' = x$; 或者没有这样的 x 存在.

(1) 假设第一种情形成立, 且 x_1 与自己相伴. 此时, 同余式

$$x^2 \equiv a \pmod{p}$$

有解 $x = x_1$. 此时就说 a 是 p 的一个**二次剩余**(quadratic residue), 或者 (在没有误解的危险时) 简称为 p 的一个**剩余**(residue), 并记为 a R p. 显然

$$x = p - x_1 \equiv -x_1 \pmod{p}$$

是这个同余式的另外一个解. 再者, 如果对 x 的任何一个其他的值 x_2 有 $x' = x$, 我们就有

$$x_1^2 \equiv a, \quad x_2^2 \equiv a, \quad (x_1 - x_2)(x_1 + x_2) = x_1^2 - x_2^2 \equiv 0 \pmod{p}.$$

故而或者有 $x_2 \equiv x_1$, 或者有

$$x_2 \equiv -x_1 \equiv p - x_1.$$

于是该同余式恰好有两个解, 也就是 x_1 和 $p - x_1$.

此时, 诸数

$$1, 2, \cdots, p-1$$

可以分成 $x_1, p - x_1$ 以及 $\frac{1}{2}(p-3)$ 对不相等的相伴数. 现在有

$$x_1(p - x_1) \equiv -x_1^2 \equiv -a \pmod{p},$$

而对任何一对相伴数 x, x' 均有

$$xx' \equiv a \pmod{p}.$$

从而

$$(p-1)! = \prod x \equiv -a \cdot a^{\frac{1}{2}(p-3)} \equiv -a^{\frac{1}{2}(p-1)} \pmod{p}.$$

(2) 如果第二种可能的情形成立, 且没有任何 x 与自己相伴, 此时就说 a 是 p 的一个**二次非剩余**(quadratic non-residue), 或者简称为 p 的一个**非剩余**(non-residue), 并记为 a N p. 此时, 同余式

$$x^2 \equiv a \pmod{p}$$

没有解, 而且诸数

$$1, 2, \cdots, p-1$$

可以分成 $\frac{1}{2}(p-1)$ 对不相等的数组成的相伴数对. 从而

$$(p-1)! = \prod x \equiv a^{\frac{1}{2}(p-1)} \pmod{p}.$$

定义 “Legendre 符号”$\left(\frac{a}{p}\right)$ 如下：

$$\text{如果 } a \text{ R } p, \left(\frac{a}{p}\right) = +1;$$

$$\text{如果 } a \text{ N } p, \left(\frac{a}{p}\right) = -1.$$

其中 p 是一个奇素数, 而 a 是任意一个不被 p 整除的数. 显然, 如果 $a \equiv b \pmod{p}$, 则有

$$\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right).$$

这样就证明了

定理 79 如果 p 是一个奇素数, 且 a 不是 p 的倍数, 那么

$$(p-1)! \equiv -\left(\frac{a}{p}\right) a^{\frac{1}{2}(p-1)} \pmod{p}.$$

我们一直假设 p 是奇数. 显然 $0 = 0^2, 1 = 1^2$, 故而所有的数都是 2 的二次剩余. 当 $p = 2$ 时, 我们不定义 Legendre 符号, 后面将不考虑这种情形. 当 $p = 2$ 时, 定理中有一些是成立的 (不过也是平凡的).

6.6 定理 79 的特例：Wilson 定理

两个最简单的情形是 $a = 1$ 和 $a = -1$ 的情形.

(1) 首先设 $a = 1$, 则

$$x^2 \equiv 1 \pmod{p}$$

有解 $x = \pm 1$. 因此 1 是 p 的一个二次剩余, 且有

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

如果在定理 79 中取 $a = 1$, 它就变成

定理 80(Wilson 定理) $(p-1)! \equiv -1 \pmod{p}$.

从而有 $11 \mid 3\,628\,801$.

同余式

$$(p-1)! + 1 \equiv 0 \pmod{p^2}$$

对于

$$p = 5, \quad p = 13, \quad p = 563$$

为真, 但对于小于 200 000 的其他 p 值都不成立. 关于这个同余式尚无一般性的定理.

如果 m 是合数, 那么

$$m|(m-1)!+1$$

不真, 这是因为存在一个数 d 使得

$$d|m, \quad 1<d<m,$$

而 d 不整除 $(m-1)!+1$. 从而我们得出:

定理 81 如果 $m>1$, 那么 m 是素数的充分必要条件是

$$m|(m-1)!+1.$$

当然, 这个定理作为一个给定的数 m 的素性的实际判别法来说是没有什么用处的.

(2) 其次假设 $a=-1$. 此时定理 79 和定理 80 表明

$$\left(\frac{-1}{p}\right)\equiv -(-1)^{\frac{1}{2}(p-1)}(p-1)!\equiv(-1)^{\frac{1}{2}(p-1)}.$$

定理 82 数 -1 是形如 $4k+1$ 的素数的二次剩余, 而是形如 $4k+3$ 的素数的二次非剩余. 也即有

$$\left(\frac{-1}{p}\right)=(-1)^{\frac{1}{2}(p-1)}.$$

更一般地, 定理 79 和定理 80 合起来给出:

定理 83 $\left(\dfrac{a}{p}\right)\equiv a^{\frac{1}{2}(p-1)} \pmod p$.

6.7 二次剩余和非剩余的初等性质

诸数

$$1^2,2^2,3^2,\cdots,\left\{\frac{1}{2}(p-1)\right\}^2 \tag{6.7.1}$$

均不同余. 这是因为 $r^2\equiv s^2$ 蕴含 $r\equiv s$ 或者 $r\equiv -s \pmod p$, 而第二种情况在这里是不可能的. 再者,

$$r^2\equiv(p-r)^2 \pmod p.$$

由此推出 p 有 $\dfrac{1}{2}(p-1)$ 个剩余和 $\dfrac{1}{2}(p-1)$ 个非剩余.

定理 84 奇素数 p 有 $\dfrac{1}{2}(p-1)$ 个剩余和 $\dfrac{1}{2}(p-1)$ 个非剩余.

接下来证明:

定理 85 两个剩余或者两个非剩余的乘积是一个剩余, 而一个剩余和一个非剩余的乘积是一个非剩余.

(1) 用 $\alpha,\alpha',\alpha_1,\cdots$ 来表示剩余, 用 $\beta,\beta',\beta_1,\cdots$ 来表示非剩余. 那么每个 $\alpha\alpha'$ 都是一个 α, 这是因为

$$x^2 \equiv \alpha,\ y^2 \equiv \alpha' \to (xy)^2 \equiv \alpha\alpha' \pmod{p}.$$

(2) 如果 α_1 是一个固定的剩余, 那么

$$1\cdot\alpha_1, 2\cdot\alpha_1, 3\cdot\alpha_1, \cdots, (p-1)\alpha_1$$

是模 p 的一个完全剩余系. 既然每个 $\alpha\alpha_1$ 都是剩余, 所以每个 $\beta\alpha_1$ 必定都是一个非剩余.

(3) 类似地, 如果 β_1 是一个固定的非剩余, 则每个 $\beta\beta_1$ 都是一个剩余. 这是因为

$$1\cdot\beta_1,\ 2\cdot\beta_1,\ \cdots,\ (p-1)\beta_1$$

是模 p 的一个完全剩余系, 且每个 $\alpha\beta_1$ 都是一个非剩余, 故而每个 $\beta\beta_1$ 都是一个剩余.

定理 85 也是定理 83 的一个推论.

再增加两个在第 20 章里要用到的定理. 第一个定理仅仅是定理 82 一部分的一个重新表述.

定理 86 *如果 p 是一个形如 $4k+1$ 的素数, 那么存在一个 x 使得有*

$$1 + x^2 = mp,$$

其中 $0 < m < p$.

这是因为, 根据定理 82, -1 是 p 的一个剩余, 故而它与 (6.7.1) 中诸数之一 (比方说是 x^2) 同余, 且有

$$0 < 1 + x^2 < 1 + \left(\frac{1}{2}p\right)^2 < p^2.$$

定理 87 *如果 p 是一个奇素数, 那么存在数 x 和 y 使得有*

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

其中 $0 < m < p$.

$\frac{1}{2}(p+1)$ 个数

$$x^2 \quad \left(0 \leqslant x \leqslant \frac{1}{2}(p-1)\right) \tag{6.7.2}$$

都是不同余的, 所以如下 $\frac{1}{2}(p+1)$ 个数

$$-1 - y^2 \quad \left(0 \leqslant y \leqslant \frac{1}{2}(p-1)\right) \tag{6.7.3}$$

也是不同余的. 但是在这两个集合中一共有 $p+1$ 个数, 却只有 p 个剩余类 (mod p), 于是 (6.7.2) 中必有某个数与 (6.7.3) 中某个数同余. 从而有一个 x 和一个 y 存在, 二者都小于 $\frac{1}{2}p$, 使得

$$x^2 \equiv -1 - y^2, \quad 1 + x^2 + y^2 = mp.$$

又有

$$0 < 1 + x^2 + y^2 < 1 + 2\left(\frac{1}{2}p\right)^2 < p^2,$$

所以有 $0 < m < p$.

定理 86 表明, 当 $p = 4k + 1$ 时可以取 $y = 0$.

6.8 $a \pmod m$ 的阶

由定理 72 可以知道, 如果 $(a, m) = 1$, 那么

$$a^{\phi(m)} \equiv 1 \pmod m.$$

用 d 来记使得

$$a^x \equiv 1 \pmod m \tag{6.8.1}$$

成立的 x 的最小正值, 则有 $d \leqslant \phi(m)$.

将同余式 (6.8.1) 称作为命题 $P(x)$, 那么显然 $P(x)$ 和 $P(y)$ 蕴含 $P(x+y)$. 再者, 如果 $y \leqslant x$ 且

$$a^{x-y} \equiv b \pmod m,$$

则有

$$a^x \equiv ba^y \pmod m,$$

从而 $P(x)$ 和 $P(y)$ 蕴含 $P(x-y)$. 于是 $P(x)$ 满足定理 69 的条件, 且

$$d | \phi(m).$$

称 d 是 $a \pmod m$ 的阶(order)[①], 并说成 a 属于(belong to) $d \pmod m$. 由于

$$2 \equiv 2, \quad 2^2 \equiv 4, \quad 2^3 \equiv 1 \pmod 7,$$

故而 2 属于 3 (mod 7). 如果 $d = \phi(m)$, 则称 a 是 m 的一个原根(primitive root). 于是 2 是 5 的一个原根, 这是因为

$$2 \equiv 2, \quad 2^2 \equiv 4, \quad 2^3 \equiv 3, \quad 2^4 \equiv 1 \pmod 5,$$

而 3 是 17 的一个原根. m 的原根这一概念与在 5.6 节中所说的本原单位根这个代数概念有某种相似性. 7.5 节中将证明：每个奇素数 p 均有原根.

现在可以将已经证明的结果总结成：

定理 88 任何与 m 互素的数 a 都属于 $\phi(m)$ 的一个因子 $\pmod m$. 如果 d 是 a 的阶 $\pmod m$, 那么 $d | \phi(m)$. 如果 m 是一个素数 p, 那么 $d | (p-1)$. 同余式 $a^x \equiv 1 \pmod m$ 成立与否, 要根据 x 是否 d 的倍数来确定.

6.9 Fermat 定理的逆定理

Fermat 定理的直接的逆命题是不正确的. 也就是说, "如果 $m \nmid a$ 且

$$a^{m-1} \equiv 1 \pmod m, \tag{6.9.1}$$

① 常称为指数(index), 但是这个词在群论中有完全不同的含义.

那么 m 一定是一个素数" 这个结论是不正确的. 甚至下面的结论也是不正确的: 如果 (6.9.1) 对所有与 m 互素的 a 皆为真, 那么 m 是素数. 例如, 假设 $m = 561 = 3 \times 11 \times 17$. 如果 $3 \nmid a$, $11 \nmid a$, $17 \nmid a$, 则根据定理 71 有

$$a^2 \equiv 1 \pmod 3, \quad a^{10} \equiv 1 \pmod{11}, \quad a^{16} \equiv 1 \pmod{17}.$$

但是 $2 \mid 560$, $10 \mid 560$, $16 \mid 560$, 所以对 3, 7, 11 中的每一个模, 从而对于模 $3 \times 11 \times 17 = 561$, 都有 $a^{560} \equiv 1$ 成立.

如果 (6.9.1) 对一个特别的 a 和一个合数 m 为真, 则称 m 是关于 a 的一个伪素数(pseudo-prime). 如果 m 关于每一个满足 $(a, m) = 1$ 的 a 都是伪素数, 则称 m 为一个Carmichael 数(Carmichael number). 现在还不知道是否有无穷多个 Carmichael 数, 甚至也不知道是否存在无穷多个合数 m, 使得有 $2^m = 2$[①]和 $3^m = 3 \pmod m$ 成立. 但是可以证明

定理 89 *关于每个 $a > 1$ 都有无穷多个伪素数存在.*

设 p 是任意一个不整除 $a(a^2 - 1)$ 的奇素数. 取

$$m = \frac{a^{2p} - 1}{a^2 - 1} = \left(\frac{a^p - 1}{a - 1}\right)\left(\frac{a^p + 1}{a + 1}\right), \tag{6.9.2}$$

故而 m 显然是合数. 现在有

$$(a^2 - 1)(m - 1) = a^{2p} - a^2 = a(a^{p-1} - 1)(a^p + a).$$

由于 a 和 a^p 两者同为奇数或者同为偶数, 故有 $2 \mid (a^p + a)$. 再者, $a^{p-1} - 1$ 能被 p 整除 (根据定理 71), $a^{p-1} - 1$ 还能被 $a^2 - 1$ 整除, 这是因为 $p - 1$ 是偶数. 由于 $p \nmid (a^2 - 1)$, 这就意味着有 $p(a^2 - 1) \mid (a^{p-1} - 1)$. 于是

$$2p(a^2 - 1) \mid (a^2 - 1)(m - 1),$$

所以 $2p \mid (m - 1)$, 且对某个整数 u 有 $m = 1 + 2pu$. 现在, 关于模 m 有

$$a^{2p} = 1 + m(a^2 - 1) \equiv 1, \quad a^{m-1} = a^{2pu} \equiv 1,$$

而这就是 (6.9.1). 由于对每个不整除 $a(a^2 - 1)$ 的奇素数 p, 我们都有 m 的一个不同的值, 这就证明了定理.

定理 71 的一个正确的逆命题是:

定理 90 *如果 $a^{m-1} \equiv 1 \pmod m$ 且对 $m - 1$ 的小于 $m - 1$ 的任何因子 x, 都有 $a^x \not\equiv 1 \pmod m$, 那么 m 是一个素数.*

显然 $(a, m) = 1$. 如果 d 是 $a \pmod m$ 的阶, 则根据定理 88 有 $d \mid (m - 1)$ 以及 $d \mid \phi(m)$. 由于 $a^d \equiv 1$, 我们必定有 $d = m - 1$, 故而 $(m - 1) \mid \phi(m)$. 但是, 如果 m 是合数, 就有

$$\phi(m) = m \prod_{p \mid m} \left(1 - \frac{1}{p}\right) < m - 1,$$

从而 m 必为素数.

① 最好改为 $2^m \equiv 2$. 以下同此, 不再重复说明. —— 译者注

6.10 $2^{p-1}-1$ 能否被 p^2 整除

根据 Fermat 定理, 如果 $p>2$, 就有

$$2^{p-1}-1\equiv 0 \pmod{p}.$$

同余式

$$2^{p-1}-1\equiv 0 \pmod{p^2}$$

是否成立? 这个问题在 "Fermat 大定理" 的理论中有重要意义 (见第 13 章). 这种情况确实有出现, 不过非常稀少.

定理 91 *存在一个素数 p, 使得有*

$$2^{p-1}-1\equiv 0 \pmod{p^2}.$$

事实上, 当 $p=1\,093$ 时它是成立的, 这可以通过直接计算来验证. 我们给出一个更加简短的证明, 其中的所有同余式都是关于模 $p^2=1\,194\,649$ 的.

首先有

$$3^7=2\,187=2p+1,\quad 3^{14}=(2p+1)^2\equiv 4p+1. \tag{6.10.1}$$

其次有

$$2^{14}=16\,384=15p-11,\quad 2^{28}\equiv -330p+121,$$
$$3^2\times 2^{28}\equiv -2\,970p+1\,089=-2\,969p-4\equiv -1\,876p-4,$$

所以

$$3^2\times 2^{26}\equiv -469p-1.$$

于是, 根据二项定理, 由 (6.10.1) 就有

$$3^{14}\times 2^{182}\equiv -(469p+1)^7\equiv -3\,283p-1\equiv -4p-1\equiv -3^{14}.$$

由此推出

$$2^{182}\equiv -1,\quad 2^{1\,092}\equiv 1 \pmod{1\,093^2}.$$

同样的结论对 $p=3\,511$ 也为真, 但对其他的 $p<3\times 10^7$ 皆不成立.

6.11 Gauss 引理和 2 的二次特征

如果 p 是一个奇素数, $n \pmod p$ 就恰好有一个剩余[①]位于 $-\frac{1}{2}p$ 和 $\frac{1}{2}p$ 之间. 称这个剩余为 $n \pmod p$ 的最小(minimal) 剩余. 最小剩余是正数还是负数, 要根据 n 的最小非负剩余是位于 0 和 $\frac{1}{2}p$ 之间还是位于 $\frac{1}{2}p$ 和 p 之间而定.

现在假设 m 是一个正的或者负的整数, 它不能被 p 整除, 考虑下面 $\frac{1}{2}(p-1)$ 个数

$$m,\ 2m,\ 3m,\ \cdots,\ \frac{1}{2}(p-1)m. \tag{6.11.1}$$

① 当然, 这里的 "剩余" 有它通常的意义, 而不是 "二次剩余" 的缩写.

的最小剩余. 可以把这些剩余写成形式

$$r_1, r_2, \cdots, r_\lambda, \quad -r'_1, -r'_2, \cdots, -r'_\mu,$$

其中

$$\lambda + \mu = \frac{1}{2}(p-1), \quad 0 < r_i < \frac{1}{2}p, \quad 0 < r'_i < \frac{1}{2}p.$$

由于 (6.11.1) 中的数都不同余, 没有任何两个 r 是相等的, 也没有任何两个 r' 是相等的. 如果有一个 r 和一个 r' 是相等的, 比方说 $r_i = r'_j$, 设 am, bm 是 (6.11.1) 中满足

$$am \equiv r_i, \quad bm \equiv -r'_j \pmod{p}$$

的两个数, 那么

$$am + bm \equiv 0 \pmod{p},$$

所以有

$$a + b \equiv 0 \pmod{p},$$

然而, 由于 $0 < a < \frac{1}{2}p, 0 < b < \frac{1}{2}p$, 故而这是不可能的.

由此推出, 诸数 r_i, r'_j 是诸数

$$1,\ 2,\ \cdots,\ \frac{1}{2}(p-1)$$

的一个重新排列. 于是

$$m \cdot 2m \cdot \cdots \cdot \frac{1}{2}(p-1)m \equiv (-1)^\mu 1 \times 2 \times \cdots \times \frac{1}{2}(p-1) \pmod{p},$$

所以

$$m^{\frac{1}{2}(p-1)} \equiv (-1)^\mu \pmod{p}.$$

但是根据定理 83 有

$$\left(\frac{m}{p}\right) \equiv m^{\frac{1}{2}(p-1)} \pmod{p}.$$

这样就得到:

定理 92(Gauss 引理)

$$\left(\frac{m}{p}\right) = (-1)^\mu,$$

其中 μ 是集合

$$m,\ 2m,\ 3m,\ \cdots,\ \frac{1}{2}(p-1)m$$

中最小正剩余 $\pmod{p}$ 大于 $\frac{1}{2}p$ 的数的个数.

特别地, 取 $m = 2$, 故而 (6.11.1) 中的数就是

$$2,\ 4,\ \cdots,\ p-1.$$

此时, λ 就是其中小于 $\frac{1}{2}p$ 的正偶数的个数.

这里引进一个记号, 以后会频繁地使用它. 用 $[x]$ 来表示 "x 的整数部分", 即不超过 x 的最大整数. 于是

$$x = [x] + f,$$

其中 $0 \leqslant f < 1$. 例如

$$\left[\frac{5}{2}\right] = 2, \quad \left[\frac{1}{2}\right] = 0, \quad \left[-\frac{3}{2}\right] = -2.$$

利用这个记号, 就有

$$\lambda = \left[\frac{1}{4}p\right].$$

但是

$$\lambda + \mu = \frac{1}{2}(p-1),$$

所以

$$\mu = \frac{1}{2}(p-1) - \left[\frac{1}{4}p\right].$$

如果 $p \equiv 1 \pmod 4$, 那么

$$\mu = \frac{1}{2}(p-1) - \frac{1}{4}(p-1) = \frac{1}{4}(p-1) = \left[\frac{1}{4}(p+1)\right],$$

而如果 $p \equiv 3 \pmod 4$, 则有

$$\mu = \frac{1}{2}(p-1) - \frac{1}{4}(p-3) = \frac{1}{4}(p+1) = \left[\frac{1}{4}(p+1)\right].$$

从而

$$\left(\frac{2}{p}\right) \equiv 2^{\frac{1}{2}(p-1)} \equiv (-1)^{\left[\frac{1}{4}(p+1)\right]} \pmod p,$$

这就是说

$$\left(\frac{2}{p}\right) = 1, \text{ 如果 } p = 8n+1 \text{ 或者 } p = 8n-1,$$

$$\left(\frac{2}{p}\right) = -1, \text{ 如果 } p = 8n+3 \text{ 或者 } p = 8n-3.$$

如果 $p = 8n \pm 1$, 那么 $\frac{1}{8}(p^2-1)$ 是偶数; 而当 $p = 8n \pm 3$ 时, 它是奇数. 从而有

$$(-1)^{\left[\frac{1}{4}(p+1)\right]} = (-1)^{\frac{1}{8}(p^2-1)}.$$

总结起来, 有下面的定理.

定理 93 $\left(\frac{2}{p}\right) = (-1)^{\left[\frac{1}{4}(p+1)\right]}$.

定理 94 $\left(\frac{2}{p}\right) = (-1)^{\frac{1}{8}(p^2-1)}$.

定理 95 2 是形如 $8n \pm 1$ 的素数的二次剩余, 是形如 $8n \pm 3$ 的素数的二次非剩余.

Gauss 引理可以用来确定以任何一个给定的整数 m 作为二次剩余的那种素数. 例如, 取 $m = -3$, 并假设 $p > 3$. 则 (6.1.1) 中的数就是

$$-3a \left(1 \leqslant a < \frac{1}{2}p\right),$$

且 μ 是这些数中最小正剩余位于 $\frac{1}{2}p$ 和 p 之间的那些数的个数. 现在有

$$-3a \equiv p - 3a \pmod{p},$$

而当 $1 \leqslant a < \dfrac{1}{6}p$ 时, $p-3a$ 介于 $\dfrac{1}{2}p$ 和 p 之间. 如果 $\dfrac{1}{6}p < a < \dfrac{1}{3}p$, 那么 $p-3a$ 就介于 0 和 $\dfrac{1}{2}p$ 之间. 如果 $\dfrac{1}{3}p < a < \dfrac{1}{2}p$, 则有

$$-3a \equiv 2p - 3a \pmod{p},$$

故而 $2p-3a$ 介于 $\dfrac{1}{2}p$ 和 p 之间. 于是满足条件的 a 的值是

$$1,\ 2,\ \cdots,\ \left[\frac{1}{6}p\right],\ \left[\frac{1}{3}p\right]+1,\ \left[\frac{1}{3}p\right]+2,\ \cdots,\ \left[\frac{1}{2}p\right],$$

从而

$$\mu = \left[\frac{1}{6}p\right] + \left[\frac{1}{2}p\right] - \left[\frac{1}{3}p\right].$$

如果 $p = 6n+1$, 那么 $\mu = n + 3n - 2n$ 是偶数, 而如果 $p = 6n+5$, 那么

$$\mu = n + (3n+2) - (2n+1)$$

是奇数.

定理 96 -3 是形如 $6n+1$ 的素数的二次剩余, 是形如 $6n+5$ 的素数的二次非剩余.

下面的定理是一个进一步的例子, 暂时把它留给读者考虑.[①]

定理 97 5 是形如 $10n \pm 1$ 的素数的二次剩余, 是形如 $10n \pm 3$ 的素数的二次非剩余.

6.12 二次互倒律

这个领域里最著名的一个定理是 Gauss 的 “二次互倒律”.

定理 98 如果 p 和 q 是奇素数, 那么

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{p'q'},$$

其中

$$p' = \frac{1}{2}(p-1), \quad q' = \frac{1}{2}(q-1).$$

如果 p 和 q 中有一个数形如 $4n+1$, 那么 $p'q'$ 是偶数, 而如果 p 和 q 都形如 $4n+3$, 那么 $p'q'$ 是奇数, 所以还可以将该定理表述成

定理 99 如果 p 和 q 都是奇素数, 那么

① 一个依赖于 Gauss 互倒律的证明见 6.13 节.

$$\left(\frac{p}{q}\right)=\left(\frac{q}{p}\right),$$

除非 p 和 q 两者都形如 $4n+3$, 此时有

$$\left(\frac{p}{q}\right)=-\left(\frac{q}{p}\right).$$

我们需要一个引理.

定理 100[①] 如果

$$S(q,p)=\sum_{s=1}^{p'}\left[\frac{sq}{p}\right],$$

那么

$$S(q,p)+S(p,q)=p'q'.$$

它的证明可以表述成几何形式. 在图中 (图 6), AC 和 BC 是 $x=p, y=q$, 而 KM 和 LM 是 $x=p', y=q'$. 如果 (如在图中那样)$p>q$, 那么 $q'/p'<q/p$, 且 M 位于对角线 OC 的下方. 由于

$$q'<\frac{qp'}{p}<q'+1,$$

所以在 $KM=q'$ 和 $KN=qp'/p$ 之间没有整数存在.

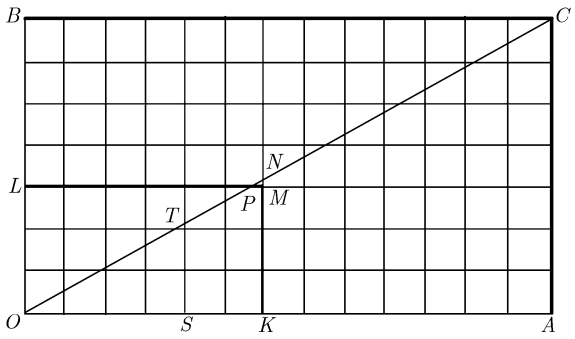

图 6

可以用两种不同的方法计算长方形 $OKML$ 中的格点个数, KM 和 LM 上的点计入在内, 但不计入数轴上的那些点. 首先, 这个数显然是 $p'q'$. 但在 OC 上没有格点 (因为 p 和 q 是素数), 在三角形 PMN 内 (除了在 PM 上可能有格点以外) 没有格点. 因此在 $OKML$ 中的格点个数是在三角形 OKN 和 OLP 中的格点个数之和 (计入在 KN 和 LP 上的格点, 但不计入在数轴上的格点).

ST(直线 $x=s$) 上的格点个数是 $[sq/p]$, 这是因为 T 的坐标是 sq/p. 故而 OKN 中的格点个数是

$$\sum_{s=1}^{p'}\left[\frac{sq}{p}\right]=S(q,p).$$

① 这个记号与 5.6 节中的记号有关.

类似地, OLP 中的格点个数是 $S(p,q)$, 这就得到了结论.

6.13 二次互倒律的证明

可以记

$$kq = p\left[\frac{kq}{p}\right] + u_k, \tag{6.13.1}$$

其中

$$1 \leqslant k \leqslant p', \quad 1 \leqslant u_k \leqslant p-1.$$

这里 u_k 是 $kq \pmod p$ 的最小正剩余. 如果 $u_k = v_k \leqslant p'$, 那么 u_k 是 6.11 节中的诸个最小剩余 r_i 中的一个, 而当 $u_k = w_k > p'$ 时, 则 $u_k - p$ 是诸个最小剩余 $-r'_j$ 中的一个. 于是, 对每个 i,j 以及某个 k 有

$$r_i = v_k, \quad r'_j = p - w_k.$$

诸 r_i 和 r'_j(详见 6.11 节) 是诸数 $1,2,\cdots,p'$ 按照某种次序的一个排列. 因此, 如果

$$R = \sum r_i = \sum v_k, \quad R' = \sum r'_j = \sum (p - w_k) = \mu p - \sum w_k$$

(详见 6.11 节, 其中的 μ 是 r'_j 的个数), 就有

$$R + R' = \sum_{\nu=1}^{p'} \nu = \frac{1}{2}\frac{p-1}{2}\frac{p+1}{2} = \frac{p^2-1}{8},$$

所以

$$\mu p + \sum v_k - \sum w_k = \frac{1}{8}(p^2-1). \tag{6.13.2}$$

另一方面, 对 (6.13.1) 从 $k=1$ 到 $k=p'$ 求和, 就有

$$\frac{1}{8}q(p^2-1) = pS(q,p) + \sum u_k = pS(q,p) + \sum v_k + \sum w_k. \tag{6.13.3}$$

由 (6.13.2) 和 (6.13.3) 可以推出

$$\frac{1}{8}(p^2-1)(q-1) = pS(q,p) + 2\sum w_k - \mu p. \tag{6.13.4}$$

现在 $q-1$ 是偶数, 而 $p^2-1 \equiv 0 \pmod 8$,[①] 所以 (6.13.4) 的左边是偶数, 而且它右边的第二项也是偶数. 于是 (由于 p 是奇数)

$$S(q,p) \equiv \mu \pmod 2,$$

从而根据定理 92 有

$$\left(\frac{q}{p}\right) = (-1)^{\mu} = (-1)^{S(q,p)}.$$

最后由定理 100 得到

$$\left(\frac{q}{p}\right)\left(\frac{p}{q}\right) = (-1)^{S(q,p)+S(p,q)} = (-1)^{p'q'}.$$

① 如果 $p = 2n+1$, 那么 $p^2 - 1 = 4n(n+1) \equiv 0 \pmod 8$.

现在利用互倒律来证明定理 97. 如果

$$p = 10n + k,$$

其中 k 是 1, 3, 7 或者 9, 那么 (因为 5 形如 $4n+1$)

$$\left(\frac{5}{p}\right) = \left(\frac{p}{5}\right) = \left(\frac{10n+k}{5}\right) = \left(\frac{k}{5}\right).$$

5 的二次剩余是 1 和 4. 因此 5 是形如 $5n+1$ 和 $5n+4$ 的素数的二次剩余, 也即是形如 $10n+1$ 和 $10n+9$ 的素数的二次剩余, 而是其他奇素数的二次非剩余.

6.14 素数的判定

现在来证明两个定理, 它们提供了某种特殊类型的数的素性判别法. 这两个定理都与 Fermat 定理密切相关.

定理 101 如果 $p > 2, h < p, n = hp + 1$ 或者 $n = hp^2 + 1$, 且

$$2^h \not\equiv 1, \quad 2^{n-1} \equiv 1 \pmod{n}, \tag{6.14.1}$$

那么 n 是一个素数.

记 $n = hp^b + 1$, 其中 $b = 1$ 或者 2, 并假设 d 是 2 (mod n) 的阶. 根据定理 88, 由 (6.14.1) 推出有 $d \nmid h$ 且有 $d|(n-1)$, 也即有 $d|hp^b$. 从而 $p|d$. 但是, 再次根据定理 88 有 $d|\phi(n)$, 故而 $p|\phi(n)$. 如果

$$n = p_1^{a_1} \cdots p_k^{a_k},$$

就有

$$\phi(n) = p_1^{a_1-1} \cdots p_k^{a_k-1}(p_1 - 1) \cdots (p_k - 1),$$

又因为 $p \nmid n$, 故而 p 至少整除 $p_1 - 1, p_2 - 1, \cdots, p_k - 1$ 中的一个. 于是 n 有一个素因子 $P \equiv 1 \pmod{p}$.

设 $n = Pm$. 由于 $n \equiv 1 \equiv P \pmod{p}$, 则有 $m \equiv 1 \pmod{p}$. 如果 $m > 1$, 则有

$$n = (up+1)(vp+1,) \quad 1 \leqslant u \leqslant v \tag{6.14.2}$$

以及

$$hp^{b-1} = uvp + u + v.$$

如果 $b = 1$, 这就是 $h = uvp + u + v$, 所以

$$p \leqslant uvp < h < p,$$

这是一对矛盾. 如果 $b = 2$, 则有

$$hp = uvp + u + v, \quad p|(u+v), \quad u + v \geqslant p,$$

故而有

$$2v \geqslant u + v \geqslant p, \quad v > \frac{1}{2}p$$

以及

$$uv < h < p, \quad uv \leqslant p - 2, \quad u \leqslant \frac{p-2}{v} < \frac{2(p-2)}{p} < 2.$$

于是 $u=1$, 从而有

$$v \geqslant p-1, \quad uv \geqslant p-1,$$

这是一对矛盾. 于是 (6.14.2) 是不可能的, 从而有 $m=1$ 以及 $n=P$.

定理 102 设 $m \geqslant 2, h < 2^m$, 且设 $n = 2^m h + 1$ 是某个奇素数 p 的二次非剩余 $(\mathrm{mod}\ p)$. 那么 n 是素数的充分必要条件是

$$p^{\frac{1}{2}(n-1)} \equiv -1 \ (\mathrm{mod}\ n). \tag{6.14.3}$$

首先假设 n 是素数. 由于 $n \equiv 1 \ (\mathrm{mod}\ 4)$, 故而根据定理 99 有

$$\left(\frac{p}{n}\right) = \left(\frac{n}{p}\right) = -1.$$

这样 (6.14.3) 就立即由定理 83 推出. 从而条件是必要的.

现在假设 (6.14.3) 为真. 设 P 是 n 的任意一个素因子, 并设 d 是 $p \ (\mathrm{mod}\ P)$ 的阶. 则有

$$p^{\frac{1}{2}(n-1)} \equiv -1, \quad p^{n-1} \equiv 1, \quad p^{P-1} \equiv 1 \ (\mathrm{mod}\ P),$$

故而根据定理 88 有

$$d \nmid \frac{1}{2}(n-1), \quad d \,|\, (n-1), \quad d \,|\, (P-1),$$

这也就是

$$d \nmid 2^{m-1}h, \quad d \,|\, 2^m h, \quad d \,|\, (P-1),$$

故有 $2^m \,|\, d$ 以及 $2^m \,|\, (P-1)$. 于是有 $P = 2^m x + 1$.

由于 $n \equiv 1 \equiv P \ (\mathrm{mod}\ 2^m)$, 我们有 $n/P \equiv 1 \ (\mathrm{mod}\ 2^m)$, 所以

$$n = (2^m x + 1)(2^m y + 1), \quad x \geqslant 1, \quad y \geqslant 0.$$

于是就有

$$2^m xy < 2^m xy + x + y = h < 2^m, \quad y = 0$$

以及 $n = P$. 从而定理的条件也是充分的.

如果令 $h = 1, m = 2^k$, 根据 2.4 节中的记号我们有 $n = F_k$. 由于 $1^2 \equiv 2^2 \equiv 1 \ (\mathrm{mod}\ 3)$, 且有 $F_k \equiv 2 \ (\mathrm{mod}\ 3)$, 故而 F_k 是一个非剩余 $(\mathrm{mod}\ 3)$. 于是, F_k 是素数的一个充分必要条件是 $F_k \,|\, (3^{\frac{1}{2}(F_k-1)} + 1)$.

6.15 Mersenne 数的因子; Euler 的一个定理

暂时回到 2.5 节中提到的 Mersenne 数这个问题. 关于 $M_p = 2^p - 1$ 的可分解性, Euler 给出了一个很简单的判别法.

定理 103 如果 $k > 1$ 且 $p = 4k + 3$ 是素数, 那么 $2p + 1$ 是素数的一个充分必要条件是

$$2^p \equiv 1 \ (\mathrm{mod}\ 2p+1). \tag{6.15.1}$$

这样一来, 如果 $2p+1$ 是素数, 那么就有 $(2p+1) \,|\, M_p$, 从而 M_p 是合数.

首先, 假设 $2p+1=P$ 是素数. 由于 $P\equiv 7 \pmod 8$, 故而根据定理 95 知, 2 是一个二次剩余 $(\mathrm{mod}\ P)$, 而根据定理 83, 有

$$2^p = 2^{\frac{1}{2}(P-1)} \equiv 1 \pmod P.$$

于是条件 (6.15.1) 是必要的, 且有 $P|M_p$. 但是 $k>1$, 故有 $p>3$ 以及

$$M_p = 2^p - 1 > 2p+1 = P.$$

从而 M_p 是合数.

其次, 假设 (6.15.1) 为真. 在定理 101 中取 $h=2, n=2p+1$. 显然有 $h<p$ 以及 $2^h = 4 \not\equiv 1 \pmod n$, 又由 (6.15.1) 有

$$2^{n-1} = 2^{2p} \equiv 1 \pmod n.$$

从而 n 是素数, 且条件 (6.15.1) 是充分的.

定理 103 包含了有关 Mersenne 数的特征的已知最简单的判别法. 对于下面这些数

$$p = 11,\ 23,\ 83,\ 131,\ 179,\ 191,\ 239,\ 251$$

所对应的前 8 个 Mersenne 数 M_p, 这个判别法给出了 M_p 的一个因子.

本 章 附 注

6.1 节. Fermat 于 1640 年陈述了他的定理 (*Œuvres*,ii. 209). Euler 于 1736 年给出他的第一个证明, 1760 年给出了推广的结论. 有关详情见 Dickson, *History*, i, 第 3 章.

6.5 节. Legendre 在 1798 年首次出版的 *Essai sur la théorie des nombres* 一书中引进了 "Legendre 符号". 例如, 参见该书第 2 版 (1808 年) 第 135 章.

6.6 节. Wilson 的定理首先是由 Waring 发表的 [*Meditationes algebraicae* (1770), 288]. 有证据表明 Leibniz 在这之前很早就已经知道这个结果. Goldberg [*Journ. London Math. Soc.* **28**(1953), 252–256] 对于 $p<10\,000$ 给出了 $(p-1)!+1$ 关于模 p^2 的剩余. 有关 $(\mathrm{mod}\, p^2)$ 的同余式的命题, 见 E. H. Pearson[*Math. Computation* **17**(1963), 194–195]. 到 2007 年, 计算已经扩展到了 5×10^8, 但未发现进一步的例子.

6.7 节. 我们可以用定理 85 来求出模 p 的最小正的二次非剩余 q 的一个上界. 设 $m=[p/q]+1$, 则有 $p<mq<p+q$. 由于 $0<mq-p<q$, 我们看到 $mq-p$ 必定是一个二次剩余, 从而 mq 也必为一个二次剩余. 于是 m 是一个二次非剩余, 从而有 $q<m$. 这样就有 $q^2<p+q$, 故而 $q<\sqrt{(p+\frac{1}{4})}+\frac{1}{2}$. Burgess [*Mathematika* **4**(1957), 106–112] 证明了, 对于任何固定的 $a>\frac{1}{4}e^{-1/2}$, 当 $p\to\infty$ 时有 $q=O(p^a)$.

6.9 节. 定理 89 属于 Cipolla, *Annali di Mat.* (3), **9**(1903), 139–160. 下面这些数, 也就是 $3\times 11\times 17$, $5\times 13\times 17$, $5\times 17\times 29$, $5\times 29\times 73$, $7\times 13\times 19$ 都是 Carmichael 数. 除了这些数以外, 小于 2 000 的数中关于 2 的伪素数还有

$$341 = 11\times 31,\quad 645 = 3\times 5\times 43,\quad 1\,387 = 19\times 73,\quad 1\,905 = 3\times 5\times 127.$$

见 Dickson, *History*, i, 91-95, Lehmer, *Amer. Math. Monthly*, **43**(1936), 347–354 以及 Leveque, *Reviews*, **1**, 47-53(从中可查到更多的参考资料).

Alford, Granville 以及 Pomerance 证明了 [*Ann. of Math.* (2) **139**(1994), 703–722]：事实上存在无穷多个 Carmichael 数. 确实, 他们构造出的数与 6 互素, 产生出满足 $2^m \equiv 2$ 以及 $3^m \equiv 3 \pmod{m}$ 的合数 m. 1899 年, Korselt 曾经证明了 (*L'inermédiaire des math.* **6**(1899), 142–143)：n 是一个 Carmichael 数, 当且仅当对每个素数 $p|\,n$ 均有 $p-1|\,n-1$.

定理 90 属于 Lucas, *Amer. Journal of Math.* **1** (1878), 302. D. H. Lehme 以及其他一些人用各种方法对这个结果加以修改, 以期能对给定的大数 m 作为素数或者合数的特征得到有实用价值的判别法. 见 Lehmer 的上面提到的引文和 *Bulletin Amer. Math. Soc.* **33**(1927), 327–340 以及 **34**(1928), 54–56, 还有 Duparc, *Simon Stevin* **29**(1952), 21–24.

6.10 节. 这里的证明是 Landau, *Vorlesungen*, iii, 275 给出的, 由 R.F.Whitehead 作了改进. 定理 91 关于 $p = 3\,511$ 的结果是 Beeger 给出的. 有关该节末尾的数值结果, 可参见 Pearson 上面提到的引文以及 Fröberg [*Computers in Math. Research*, (North Holland, 1968), 84–88]. 现在 (2007) 已知：在不超过 1.25×10^{15} 的范围内没有其他满足所描述性质的素数存在.

6.11 节至 6.13 节. 定理 95 是由 Euler 首先证明的. 定理 98 是由 Euler 和 Legendre 陈述的, 但是该结论的第一批令人满意的证明是由 Gauss 给出的. 关于这个论题的历史以及许多其他的证明, 见 Bachmann, *Niedere Zahlentheorie*, i, 第 6 章.

6.14 节. Miller 和 Wheeler 在定理 101 中取已知的素数 $2^{127}-1$ 作为 p, 求得 $n = 190p^2+1$ 满足该判别法. 见我们关于 2.5 节的附注. 当 $n = hp^3+1$ 时, 定理 101 亦为真, 只要 $h < \sqrt{p}$ 且 h 不是一个立方数即可. 见 Wright, *Math. Gazette*, **37**(1953), 104–106.

Robinson 推广了定理 102 [*Amer. Math. Monthly*, **64**(1957), 703–710], 他还和 Selfridge 用到这个定理关于 $p = 3$ 的情形, 从而得到了一大批形如 $h \cdot 2^m + 1$ 的素数 [*Math. tables and other aids to computation*, **11**(1957), 21–22]. 其中有一些素数是 Fermat 数的因子. 也见 15.5 节的注释.

Lucas[*Théorie des nombres*, i (1891), p. xii] 陈述了 F_k 的素性的判别法. Hurwitz [*Math. Werke*, ii. 747] 给出了一个证明. 人们用这个判别法证明了 F_7 和 F_{10} 是合数, 虽然它们的真实的因子是后来才找到的.

这方面最重要的进展无疑是 Agrawal, Kayal 以及 Saxena[*Ann. of Math.* (2) **160**(2004), 781–793] 的结果, 它给出一个素性判别法, 它最终以 Fermat 定理作为基础, 可在时间 $(\log n)^c$ 之内检验数 n 的素性. 这里 c 是一个数值常数, 根据 Lenstra 与 Pomerance 的工作, 此常数可取为 6.

6.15 节. 定理 103; Euler, *Comm. Acad. Petro.* **6**(1732–1733), 103[*Opera*(1), ii. 3].