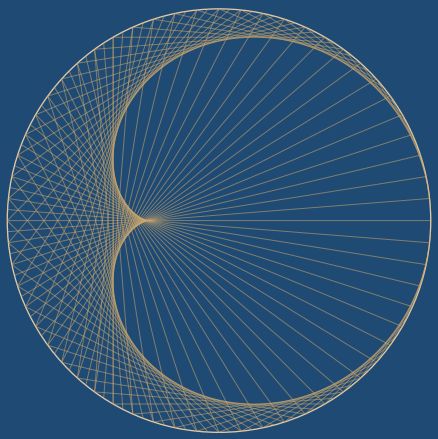
MIT 18.781 THEORY OF NUMBERS SPRING 2012

初等数论

从整除、同余到连分数与有理点



$x \mapsto 2x \pmod{120}$

原课程主讲：Abhinav Kumar

独立中文整理本

MIT OPENCOURSEWARE · 18.781

初等数论逐讲讲义

从整除、同余到连分数与有理点

原课程主讲：Abhinav Kumar

原课程：MIT 18.781　2012 年春季学期

数学基础自学讲义项目　·　2026 年 11 月修订

依据 MIT OpenCourseWare 材料独立整理与扩充

来源、版本与使用说明

原课程：MIT OpenCourseWare 18.781 Theory of Numbers，2012 年春季学期
原课程主讲：Abhinav Kumar
原课程课堂笔记整理：Joseph Lee
课程页面：https://ocw.mit.edu/courses/18-781-theory-of-numbers-spring-2012/

中文整理与扩充：数学基础自学讲义项目
整理本版本：2026 年 11 月修订版　**修订日期**：2026-11-04

本书依据原课程的课堂笔记、习题与练习考试材料独立整理，补充中文讲解、证明步骤、例题与练习。它是学习辅助材料，不代表 MIT 或原课程教师发布、审定或背书；各讲的具体依据见「对应材料」。

参考读物：

- 原课程推荐读物：Ivan Niven, Herbert S. Zuckerman, Hugh L. Montgomery, *An Introduction to the Theory of Numbers*, 5th ed., Wiley。
- 延伸阅读：华罗庚，《数论导引》，科学出版社。

许可与使用：原课程材料的署名与许可按其来源页面标注。本整理本的改写与新增内容采用署名–非商业性使用–相同方式共享 4.0 国际许可协议（CC BY-NC-SA 4.0）。转载或改编时，请保留来源、署名与许可说明，并标明所作修改；第三方材料仍适用各自的许可。课程网站的说明见MIT OCW 使用条款。

排版：XeLaTeX 与 ctex；图形主要由 TikZ 绘制。

前言：从整数问题出发

数论从整数的性质出发：怎样判断两个数是否互素？同余方程有没有解？一个整数可以写成两个平方数之和吗？这些问题容易用小数字试验，但对所有整数成立的结论需要证明。

本书的路线

第一部分讨论整除、欧几里得算法、素数分解和同余，并引入中国剩余定理、素性检验、RSA 与原根。第二部分研究二次剩余、二次互反律、Jacobi 符号、算术函数和生成函数。第三部分从两平方和与连分数出发，讨论有理逼近、Pell 方程、四平方和、勾股数组及二次曲线上的有理点。

本书以初等方法为主，也会在级数、连分数和有理逼近等内容中使用实数、极限或复数。不同工具的作用各不相同：同余用来限制整数解，算术函数组织因数信息，连分数则给出研究有理逼近和 Pell 方程的方法。具体结论及其条件在各讲中建立。

课程材料与本整理本

本书依据 MIT OpenCourseWare 18.781 的课程材料整理，共 24 讲和三次复习课；原课程课堂笔记由 Joseph Lee 整理。中文讲义补写了动机、证明、计算例题和练习，并在各讲注明对应材料。

本书是一份独立的学习辅助材料。各讲保留「对应材料」供读者核对来源；学习时可结合原课程讲义、习题与考试材料。先修要求和各部分路线列在后面的「如何使用本讲义」中。

如何使用本讲义

一讲的阅读顺序

每讲开头的「对应材料」说明来源；讲次与原课程课次的具体关系以该说明为准。

(1) **看目标与动机**：先明确本讲要解决的问题，再进入定义和定理。

(2) **读证明与例题**：先尝试说出思路，再逐步核对条件和计算；遇到「想一想」时，先自己作答。

(3) **做练习再看答案**：答案用于核对步骤。做错的题应回到相应定义、定理或例题查找原因。

(4) **用小结与复习课自测**：合上正文复述主要结论；完成一部分后，再做相应的 R1、R2、…… 复习课。

练习难度按本讲内容标记：★ 直接应用定义或定理；★★ 组合结论或完成较长计算；★★★ 需要额外构造或推广。星级用于安排学习顺序。

版面元素与编号

标记	读法
定义、公理	青色色条给出精确表述。检查对象、条件和量词；青色粗体标记首次引入的术语。
定理、命题、引理	蓝色色条给出结论。先读条件，再读结论；蓝色粗体提示需要记住的内容。
例　解	橙色方块标记例题，解答以青色 ♦ 结束；可以遮住解答先算一遍。
思路　证明	思路说明证明计划，证明给出论证；证明以蓝色 ■ 结束。
想一想	暂停阅读，检查自己对刚出现的概念或结论的理解，再核对「答」。
直观理解	图像或估算用于帮助理解；正式结论仍以定义、定理及证明为准。
易错提醒	核对常见错误及其原因，尝试说明错误方法在哪一步失效。
核心思想	提炼一节或一讲的方法；复习时结合本讲小结阅读。

「定理 3.2」表示第 3 讲的第 2 条定理类结论；「定理 R1.3」属于复习课 1。复习课放在每部分末尾，包含核心结论、题型例题、常见错误与自测。

本课程的准备与路线

开始学习需要整数运算和基本的证明方法（数学归纳法、反证法）。读完后半部还需逐步补充：第 12 讲的复数与单位根，第 14 讲的级数和绝对收敛，第 16 讲的一般线性方程组与行列式，第 18–21 讲的实数与极限，以及第 22–24 讲的矩阵乘法和可逆变换。第 16–17 讲的形式幂级数先按系数运算；代入具体数值求和时另查收敛。各讲在使用位置给出准备和暂缓阅读的提示。

建议的读法：例题和想一想里的数字都亲手算一遍，必要时写几行程序检验。每讲配合「对应材料」中列出的 MIT 习题集练习，并按原课程页面提供的资料核对。复习课 R1、R2、R3 对应两次期中考试和期末考试。

各部分的学习目标

部分	讲次	主要内容	学完应能做到
第一部分	1–8，R1	整除、素数分解、同余、中国剩余定理、素性检验、RSA、Hensel 引理与原根。	解一次同余及同余式组；检查提升与原根的条件，说明素性检验的适用范围。
第二部分	9–16，R2	二次剩余、二次互反律、Jacobi 符号、模素数平方根、分圆多项式、算术函数与生成函数。	区分平方可解性与符号计算；用因数和、反演或生成函数组织计算与证明。
第三部分	17–24，R3	两平方和、连分数与有理逼近；Pell 方程、四平方和、勾股数组与二次曲线有理点。	用渐近分数研究逼近与 Pell 解；核对整数参数、递降条件及有理点的构造。

常用符号与记号约定

本表列出常用记号；模数、素数及互素条件须随定义一起阅读。同一字母或符号在不同讲次可能有不同含义，以所在讲次的定义为准。

记号	含义与约定	查阅位置
$\mathbb{Z}$, $\mathbb{Z}^+$, $\mathbb{N}$	整数与正整数；本表 $\mathbb{Z}^+ = \mathbb{N} = \{1, 2, \ldots\}$。	第 1 讲
$a \mid b$, $a \nmid b$	整除与不整除；$a \mid b$ 指存在整数 k 使 $b = ka$。	第 1 讲
$\gcd(a, b)$, (a, b)	最大公因数的两种记法；本书对 a, b 不全为 0 定义，取正值。	第 1–2 讲
$\operatorname{lcm}(a, b)$, $[a, b]$	最小公倍数；用于正整数时取最小的正公倍数。区间与连分数中的方括号按上下文区分。	第 2、5 讲
$\pi(x)$, $f(x) \sim g(x)$	素数计数函数；$f \sim g$ 表示在指定极限过程中 $f/g \to 1$。	第 2 讲
$a \equiv b \pmod{m}$	模 m 同余，即 $m \mid (a - b)$；模数取正整数。	第 3 讲
$\mathbb{Z}/m\mathbb{Z}$	模 m 的剩余类环；与 m 互素的剩余类构成乘法群 $(\mathbb{Z}/m\mathbb{Z})^\times$。	第 3–7 讲
$a^{-1} \pmod{m}$	模逆元，存在当且仅当 $\gcd(a, m) = 1$；逆元按模 m 的剩余类理解。	第 4–5 讲
$\operatorname{ord}_m(a)$, g	当 $\gcd(a, m) = 1$ 时，阶是使 $a^k \equiv 1 \pmod{m}$ 的最小正整数；原根 g 的阶为 $\varphi(m)$。	第 7–8 讲
$\left(\frac{a}{p}\right)$	Legendre 符号，p 为奇素数；$p \mid a$ 时为 0，非零平方剩余为 1，非剩余为 -1。	第 9 讲
$\left(\frac{a}{n}\right)$	Jacobi 符号，n 为正奇数；按素因子分解定义为 Legendre 符号的乘积。值为 1 不保证平方可解。	第 10 讲
$\varphi(n)$, $\mu(n)$	Euler 函数与 Möbius 函数；$\mu(1) = 1$，可被素数平方整除时为 0，否则为 $(-1)^r$，r 为不同素因子数。	第 4、13 讲
$d(n)$, $\sigma(n)$	正因数个数与正因数和；具体求和公式见正文。	第 13 讲
$\zeta(s)$	在 $\operatorname{Re}(s) > 1$ 时，$\zeta(s) = \sum_{n \geq 1} n^{-s}$ 且有 Euler 乘积；范围外不能直接用此级数定义。	第 14 讲
$\lfloor x \rfloor$, $\{x\}$	不超过 x 的最大整数与小数部分；$\{x\} = x - \lfloor x \rfloor \in [0, 1)$。	第 2、18 讲
$[a_0, a_1, \ldots]$, p_k/q_k	简单连分数与渐近分数；$a_0 \in \mathbb{Z}$，其后系数为正整数。正文也采用分号分隔首项的写法。	第 18–19 讲
$\bar{x}$, $N_d(x + y\sqrt{d})$	二次共轭与范数；$\overline{x + y\sqrt{d}} = x - y\sqrt{d}$，$N_d(x + y\sqrt{d}) = x^2 - dy^2$，参数条件见正文。	第 20–22 讲
$\bar{q}$, $N(q)$	四元数的共轭与平方范数；$N(q)$ 是四个实坐标平方之和。与二次范数按对象区分。	第 22 讲

目录

前言：从整数问题出发 **ii**

如何使用本讲义 **iii**

常用符号与记号约定 **v**

第一部分　整除、同余与原根 **1**

第 1 讲　引言、丢番图方程、整除与最大公因数 **2**

1.1　数论研究什么？ 2

1.2　丢番图方程：只要整数解 3

1.3　自然数的两条基本原理 4

1.4　整除 6

1.5　带余除法 8

1.6　最大公因数与 Bézout 定理 10

1.7　互素及其三个推论 13

练习 16

第 2 讲　欧几里得算法与素数 **18**

2.1　欧几里得算法 18

2.2　扩展欧几里得算法与线性丢番图方程 20

2.3　欧几里得算法有多快 23

2.4　素数与算术基本定理 25

2.5　素数有无穷多个 27

2.6　素数的分布与著名猜想 29

练习 33

第 3 讲　二项式系数与同余 **35**

3.1　从计数到二项式系数 35

3.2　二项式定理与 Pascal 三角 38

3.3　$n!$ 中素数的方幂：Legendre 公式 41

3.4　Kummer 定理：二项式系数与进位 43

3.5 同余 . . . 45
3.6 二项式系数模 p：Lucas 定理 . . . 47
练习 . . . 49

第 4 讲 Fermat、Euler、Wilson 定理与一次同余式 51
4.1 剩余系与 Euler 函数 . . . 52
4.2 Euler 定理与 Fermat 小定理 . . . 54
4.3 Fermat 小定理的另两个证明 . . . 57
4.4 模 m 的逆元 . . . 58
4.5 Wilson 定理 . . . 60
4.6 $x^2 \equiv -1$ 与 $4k+1$ 型素数 . . . 62
4.7 同余方程与一次同余式 . . . 63
练习 . . . 65

第 5 讲 一次同余式、中国剩余定理与算法 68
5.1 一次同余式的解法 . . . 69
5.2 同余式组：从「物不知数」说起 . . . 71
5.3 中国剩余定理 . . . 73
5.4 模数不互素的同余式组 . . . 76
5.5 结构的观点：一一对应与 Euler 函数 . . . 77
5.6 数值计算与算法 . . . 79
练习 . . . 82

第 6 讲 素性检验、因数分解、RSA 与 Hensel 引理 85
6.1 中国剩余定理与同余式的解数 . . . 85
6.2 素性检验 . . . 88
6.3 因数分解与 Pollard ρ 算法 . . . 93
6.4 RSA 公钥密码 . . . 95
6.5 Hensel 引理 . . . 97
练习 . . . 100

第 7 讲 Hensel 引理续、素数模同余式与原根 102
7.1 Hensel 引理：证明与迭代 . . . 102
7.2 奇异根：全部提升或全部消失 . . . 106
7.3 素数模的同余式与 Lagrange 定理 . . . 107
7.4 恰有 n 个根的判别法 . . . 109
7.5 阶 . . . 112

7.6 原根 . . . 115
练习 . . . 116

第 8 讲 原根：个数、指标与存在性 118
8.1 原根的幂与原根的个数 . . . 118
8.2 指标：模 p 的离散对数 . . . 122
8.3 用指标解 $x^d \equiv a \pmod{p}$ . . . 124
8.4 模 p^2 的原根 . . . 127
8.5 模 p^e 与 $2p^e$ 的原根 . . . 129
8.6 没有原根的模 . . . 131
练习 . . . 135

复习课 1 整除、同余与原根的综合复习 138
R1.1 知识地图与定义检查 . . . 138
R1.2 核心结论回顾 . . . 139
R1.3 基础计算：把步骤写到能够回代 . . . 142
R1.4 局部求根、整体合并与证明题 . . . 143
R1.5 准确的阶与指标方程 . . . 145
R1.6 常见错误与交卷前检查 . . . 146
练习 . . . 147

第二部分 二次剩余与算术函数 149

第 9 讲 二次剩余与二次互反律 150
9.1 从二次同余式到 $x^2 \equiv a$ . . . 150
9.2 二次剩余与 Euler 判别法 . . . 152
9.3 Legendre 符号 . . . 157
9.4 Gauss 引理与 $\left(\frac{2}{p}\right)$ . . . 159
9.5 Eisenstein 引理：把符号变成取整和 . . . 161
9.6 二次互反律 . . . 162
练习 . . . 166

第 10 讲 Jacobi 符号与 Zolotareff 定理 169
10.1 回顾：Legendre 符号怎么算 . . . 169
10.2 Jacobi 符号 . . . 171
10.3 Jacobi 符号的互反律 . . . 173
10.4 不做分解的计算方法 . . . 175

10.5 把 Jacobi 符号看成一个置换的符号 . . . 178
练习 . . . 181

第 11 讲 模 p 平方根：Tonelli 算法与连续平方剩余 183
11.1 回顾：用 Jacobi 符号计算二次特征 . . . 183
11.2 固定 a：哪些素数让 a 成为平方 . . . 185
11.3 从「有没有」到「是多少」：两个直接公式 . . . 187
11.4 Tonelli 算法 . . . 189
11.5 算法有多快，难在哪里 . . . 193
11.6 连续的两个平方剩余 . . . 195
练习 . . . 198

第 12 讲 分圆多项式与模 n 余 1 的素数 200
12.1 单位根与本原单位根 . . . 200
12.2 分圆多项式与 $x^n - 1$ 的分解 . . . 203
12.3 整系数与回文性 . . . 206
12.4 模 p 的分圆多项式 . . . 208
12.5 分圆多项式值的素因数 . . . 210
12.6 模 n 余 1 的素数有无穷多个 . . . 211
练习 . . . 213

第 13 讲 算术函数与 Möbius 函数 215
13.1 算术函数 . . . 215
13.2 积性函数 . . . 217
13.3 Dirichlet 卷积 . . . 220
13.4 积性函数的卷积 . . . 222
13.5 偶完全数：Euclid–Euler 定理 . . . 226
13.6 Möbius 函数 . . . 227
练习 . . . 231

第 14 讲 Möbius 反演与 zeta 函数 233
14.1 如何撤销「对所有因数求和」 . . . 233
14.2 素数分解给出 Euler 乘积 . . . 234
14.3 反演的算术应用 . . . 237
14.4 Möbius 函数的抵消与应用 . . . 239
14.5 Dirichlet 级数的乘法 . . . 241
14.6 两个 zeta 值 . . . 242

14.7 从因数和得到密度与概率 . . . 243
练习 . . . 245

第 15 讲 组合计数与线性递推 247
15.1 同一个集合可以怎样数？ . . . 247
15.2 Fibonacci 数与特征多项式的入口 . . . 249
15.3 递推从哪里来？从拆分和字符串计数开始 . . . 252
15.4 容斥原理：把重复计数减回来 . . . 255
15.5 抽屉原理：碰撞会迫使方程有解 . . . 257
15.6 在十个连续整数中找一个「好数」 . . . 259
练习 . . . 260

第 16 讲 生成函数与线性递推 262
16.1 常系数线性递推与特征多项式 . . . 262
16.2 生成函数：把递推折叠成一个分式 . . . 265
16.3 重根与非齐次递推 . . . 268
16.4 斐波那契数的数论性质 . . . 272
16.5 斐波那契数模素数 . . . 274
16.6 共轭根与整除性 . . . 276
练习 . . . 277

复习课 2 平方剩余、算术函数与递推 279
R2.1 知识地图与选用工具 . . . 279
R2.2 核心结论与适用条件 . . . 280
R2.3 题型与完整例题 . . . 283
R2.4 常见错误与快速自检 . . . 287
练习 . . . 288

第三部分 连分数与丢番图方程 290

第 17 讲 生成函数的筛选与两平方和定理 291
17.1 生成函数的四种运算 . . . 291
17.2 单位根筛选 . . . 294
17.3 交错 zeta 和 . . . 296
17.4 两平方和：必要性与最小倍数 . . . 297
17.5 两平方和：下降法 . . . 299
练习 . . . 302

第 18 讲　连分数（一）：欧几里得算法与渐近分数 **305**

18.1 从欧几里得算法到连分数 . . . 305
18.2 无限连分数初探 . . . 310
18.3 渐近分数与递推公式 . . . 312
18.4 行列式恒等式及其推论 . . . 316
18.5 无限连分数的收敛与误差 . . . 318
练习 . . . 320

第 19 讲　连分数（二）：有理逼近的不等式 **323**

19.1 两个相邻渐近分数之间 . . . 323
19.2 每两项中有一个好逼近 . . . 325
19.3 每三项中有一个更好的逼近 . . . 328
19.4 无限多个好有理数与 Hurwitz 常数 . . . 330
19.5 黄金分割比与锐利常数 . . . 331
19.6 周期性与二次方程 . . . 334
练习 . . . 335

第 20 讲　循环连分数与二次无理数 **338**

20.1 二次无理数与共轭 . . . 338
20.2 回顾：完全商与渐近分数 . . . 341
20.3 从循环连分数到二次无理数 . . . 342
20.4 从二次无理数到循环连分数 . . . 343
20.5 纯循环的共轭判别 . . . 346
20.6 Pell 方程与有理逼近 . . . 348
练习 . . . 349

第 21 讲　平方根的周期与 Pell 方程 **351**

21.1 把平方根放进一个纯循环 . . . 351
21.2 完全商的分母如何记录范数 . . . 353
21.3 小范数解为什么一定是渐近分数 . . . 355
21.4 周期奇偶性决定正、负 Pell 方程 . . . 357
21.5 把定理变成可检查的整数算法 . . . 359
21.6 从第一组解到有序的解列 . . . 361
练习 . . . 363

第 22 讲　Pell 方程的全部解与四平方和定理 **365**

22.1 Pell 方程的全部正解 . . . 365

22.2 把 Pell 解用于其他整数方程 . . . 368
22.3 四元数与四平方恒等式 . . . 370
22.4 从一个素数到它的较小倍数 . . . 373
22.5 最小乘数的无穷递降 . . . 375
22.6 把存在性证明转成可核对的计算 . . . 377
练习 . . . 378

第 23 讲 勾股数组与费马无穷递降法 380
23.1 从本原数组到两个互素平方 . . . 380
23.2 单位圆上的有理点与斜率参数 . . . 383
23.3 四次方之和与两次参数化 . . . 386
23.4 同余障碍与三元立方递降 . . . 388
23.5 平方加一的素因子与一个三次方程 . . . 390
练习 . . . 392

第 24 讲 圆锥曲线上的有理点 394
24.1 从一个有理点出发 . . . 394
24.2 齐次化：把分母交给第三个变量 . . . 396
24.3 对角化与系数的规范化 . . . 398
24.4 Legendre 定理：先检查符号与剩余 . . . 399
24.5 充分性的证明：从同余找到真正的零点 . . . 401
24.6 把存在性判别接回参数化 . . . 403
练习 . . . 405

复习课 3 平方和、连分数与丢番图方程：期末复习 408
R3.1 知识地图：从表示到参数 . . . 408
R3.2 核心结论：记条件，也记用途 . . . 409
R3.3 八类题的完整计算 . . . 411
R3.4 六个检查点：避免把条件丢在计算之外 . . . 415
R3.5 全课程：三部分怎样接到一起 . . . 416
练习 . . . 417

第一部分

整除、同余与原根

第 1 讲

引言、丢番图方程、整除与最大公因数

Introduction, Diophantine Equations, Divisibility, GCD

对应材料：MIT 18.781 Lecture 1 讲义；习题集 1。

本讲目标

(1) 知道数论研究什么，什么是丢番图方程，体会「有没有整数解」为什么是一个与「有没有实数解」完全不同的问题。

(2) 掌握数学归纳法与良序原理，理解二者等价，并会用「最小反例」进行论证。

(3) 熟练使用整除的定义和基本性质，特别是「公因数整除任意整系数线性组合」。

(4) 理解带余除法定理的陈述与证明，会对负数做带余除法。

(5) 理解 Bézout 定理：$\gcd(a,b)$ 是 a,b 的最小正整系数线性组合；会用它判断线性丢番图方程是否有解，并推出关于互素的三个推论。

1.1 数论研究什么？

数论最朴素的研究对象是整数 $\mathbb{Z}=\{\dots,-2,-1,0,1,2,\dots\}$ 和自然数 $\mathbb{N}=\{0,1,2,\dots\}$。本讲义沿用课程的约定：$0\in\mathbb{N}$。

整数看上去太简单了，还有什么值得研究？其实恰恰相反。整数是全部数学的地基：从自然数出发，每次「补上缺的东西」，就能一步步造出所有常用的数系（图 1.1）。在 $\mathbb{N}$ 里减法不总能做，于是添上负数得到 $\mathbb{Z}$；在 $\mathbb{Z}$ 里除法不总能做，于是添上分数得到 $\mathbb{Q}$；$\mathbb{Q}$ 里有「洞」（比如 $\sqrt{2}$ 不在其中），用 Dedekind 分割把洞补上得到 $\mathbb{R}$；最后添上 $\sqrt{-1}$ 得到 $\mathbb{C}$。有了 $\mathbb{R}$ 和 $\mathbb{C}$，才有微积分、拓扑和其余的一切。

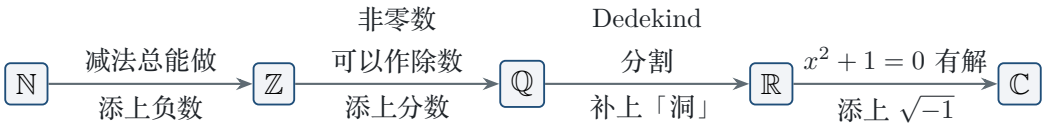


图 1.1　从自然数出发逐步扩充数系，依次补足减法、非零除法、Dedekind 分割和 $x^2+1=0$ 的解。

数学家 Kronecker（1823–1891）有一句名言：「上帝创造了整数，其余都是人的工作。」Gauss（1777–1855）则说：「数学是科学的女王，而数论是数学的女王。」学数论至少有这样几个理由：

- **它是最基础的数学**。上面的图说明，其他数系都建立在整数之上。
- **它有最优美的结果**，而且数论问题常常推动代数、分析、几何、逻辑与计算机科学的发展。

- **它有重要的应用**：RSA 公钥密码（第 6 讲）、纠错码、扩张图的构造等。
- **它是学习读写证明的理想场所**：对象具体，可以随手用小数字检验，但论证必须严格。
- **它盛产「容易陈述、极难证明」的猜想**，下一节就会见到几个。

1.2　丢番图方程：只要整数解

定义 1.1（丢番图方程）　系数为整数的多项式方程，如果只关心它的**整数解**（有时是有理数解），就称为**丢番图方程**（Diophantine equation）。名字来自公元 3 世纪亚历山大的数学家 Diophantus。

同一个方程，在 $\mathbb{R}$ 里求解和在 $\mathbb{Z}$ 里求解，是两个完全不同的问题。先看一个最简单的例子。

■ **例 1.1**　方程 $x^2+y^2=3$ 有没有实数解？有没有整数解？

解　实数解有无穷多个：它们组成以原点为圆心、半径为 $\sqrt{3}$ 的圆。

整数解一个也没有。论证只需要「平方增长得很快」这一点：若 $|x|\geq 2$，则 $x^2\geq 4>3$，而 $y^2\geq 0$，所以 $x^2+y^2>3$。因此整数解必须满足 $|x|\leq 1$，同理 $|y|\leq 1$。于是 $x^2,y^2\in\{0,1\}$，从而 $x^2+y^2\in\{0,1,2\}$，永远等于不了 3。 ♦

图 1.2 把这件事画了出来：半径 $\sqrt{3}$ 的圆从格点之间「穿过」，一个格点也没碰到；而半径 $\sqrt{5}$ 的圆恰好经过 8 个格点 $(\pm1,\pm2),(\pm2,\pm1)$。实数解看的是「圆在哪里」，整数解看的是「圆是否恰好碰到格点」，后者要敏感得多。

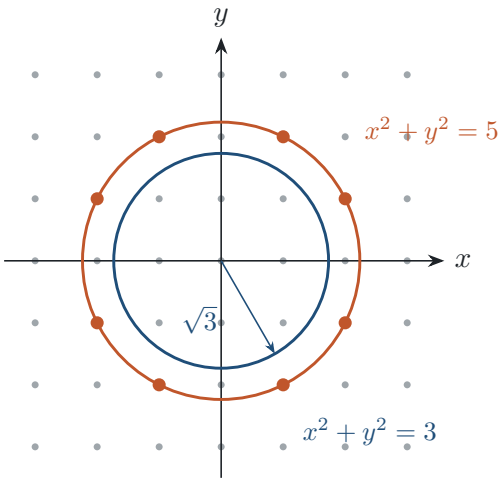


图 1.2　格点与两个圆。蓝色圆 $x^2+y^2=3$ 不经过任何格点；橙色圆 $x^2+y^2=5$ 经过 8 个格点。

■ **例 1.2　勾股数**　方程 $a^2+b^2=c^2$ 的正整数解叫**勾股数**（Pythagorean triple）。试验证 $(3,4,5)$、$(5,12,13)$、$(8,15,17)$、$(7,24,25)$ 都是勾股数。

解　逐个计算：$9+16=25$；$25+144=169=13^2$；$64+225=289=17^2$；$49+576=625=25^2$。 ♦

勾股数有无穷多组，第 23 讲会给出全部勾股数的公式。把指数 2 换成 $n \geq 3$，就得到著名的**费马大定理**（Fermat's Last Theorem）：$a^n + b^n = c^n$ 没有正整数解。Fermat 在 1637 年写下这个断言，直到 1994 年才由 Wiles（与 Taylor 合作）证明，中间相隔 357 年。

还有至今未解决的问题。**完美长方体**（perfect cuboid）问题问：是否存在正整数 a, b, c，使三条面对角线 $\sqrt{a^2+b^2}$、$\sqrt{b^2+c^2}$、$\sqrt{a^2+c^2}$ 和体对角线 $\sqrt{a^2+b^2+c^2}$ 全是整数？只差最后一条的例子是有的：$(a,b,c) = (44, 117, 240)$ 的三条面对角线分别是 $125, 244, 267$，都是整数（例如 $44^2+117^2 = 1936+13689 = 15625 = 125^2$）；但体对角线的平方是 $1936+13689+57600 = 73225$，而 $270^2 = 72900$、$271^2 = 73441$，73225 夹在两者之间，不是完全平方数。这样的问题很好懂，却至今没有人会做。

前面几个方程都是二次以上的。下面这个一次方程会引出本讲的核心。

■ **例 1.3**　方程 $6x + 9y = 10$ 有整数解吗？$6x + 9y = 15$ 呢？

解　对任意整数 x, y，$6x+9y = 3(2x+3y)$ 都是 3 的倍数，而 10 不是 3 的倍数，所以 $6x+9y = 10$ **没有整数解**。

15 是 3 的倍数，试几个小数字：$x = 1$, $y = 1$ 给出 $6+9 = 15$，所以 $6x + 9y = 15$ **有整数解**。 ♦

想一想　对哪些整数 c，方程 $6x + 9y = c$ 有整数解？先试 $c = 3$ 和 $c = 12$。
答　$c = 3$：$x = -1$, $y = 1$。$c = 12$：$x = 2$, $y = 0$。猜想：**有解当且仅当 $3 \mid c$**，而 3 恰好是 6 和 9 的最大公因数。「只有当」的方向上面已经证明了；「当」的方向是 §1.6 的 Bézout 定理。

1.3　自然数的两条基本原理

自然数的一切性质，最终都建立在**后继**（successor）运算 $s(n) = n + 1$ 上：加法是重复取后继，乘法是重复做加法。要对「所有自然数」证明结论，靠的是下面两条原理。

公理 1.2（数学归纳法 PMI）　设 $P(n)$ 是关于自然数 n 的命题。如果

(1)（奠基）$P(0)$ 成立；

(2)（归纳步）对每个 $n \in \mathbb{N}$，由 $P(n)$ 成立可推出 $P(n+1)$ 成立，

那么 $P(n)$ 对所有 $n \in \mathbb{N}$ 成立。

奠基不一定从 0 开始：若 $P(n_0)$ 成立且对 $n \geq n_0$ 有 $P(n) \Rightarrow P(n+1)$，则 $P(n)$ 对一切 $n \geq n_0$ 成立。

公理 1.3（良序原理 WOP）　$\mathbb{N}$ 的每个非空子集都有最小元。

直观理解　数学归纳法像推倒一排多米诺骨牌：第一块倒下，并且每一块倒下都会推倒下一块，于是全部倒下。良序原理说的是：在自然数里不能无限下降。从任何一个自然数出发往下走，走有限步就到底了。这两种说法看起来不一样，下面证明它们其实等价。

定理 1.4　数学归纳法与良序原理等价：以其中任一条为公理，可以证明另一条。

思路　两个方向都用反证法，而且用的是同一个想法：如果结论失败，就去看「第一个失败的地方」。用 WOP 证 PMI 时，「第一个失败的 n」由 WOP 直接给出；用 PMI 证 WOP 时，对「$0, 1, \dots, n$ 都不在集合里」做归纳。

证明　**WOP $\Rightarrow$ PMI**。设 $P(0)$ 成立，且 $P(n) \Rightarrow P(n+1)$ 对一切 n 成立。令

$$S = \{\, n \in \mathbb{N} \mid P(n) \text{ 不成立} \,\}.$$

假设 $S \neq \varnothing$。由 WOP，S 有最小元 m。因为 $P(0)$ 成立，$m \neq 0$，所以 $m-1 \in \mathbb{N}$。又 m 是 S 的最小元，$m-1 \notin S$，即 $P(m-1)$ 成立。由归纳步，$P(m)$ 成立，这与 $m \in S$ 矛盾。所以 $S = \varnothing$，即 $P(n)$ 对一切 n 成立。

PMI $\Rightarrow$ WOP。设 $A \subseteq \mathbb{N}$ 没有最小元，我们证明 $A = \varnothing$。对 $n \in \mathbb{N}$，令 $P(n)$ 为命题「$0, 1, \dots, n$ 都不属于 A」。

奠基：若 $0 \in A$，则 0 就是 A 的最小元（没有比 0 更小的自然数），矛盾。所以 $P(0)$ 成立。

归纳步：设 $P(n)$ 成立，即 $0, \dots, n \notin A$。若 $n+1 \in A$，则 A 中没有比 $n+1$ 更小的元素，$n+1$ 就是最小元，矛盾。所以 $n+1 \notin A$，$P(n+1)$ 成立。

由 PMI，$P(n)$ 对一切 n 成立，特别地每个 n 都不属于 A，即 $A = \varnothing$。这正是 WOP 的逆否命题：非空子集必有最小元。 ■

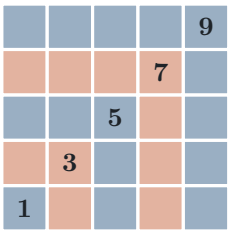
■ **例 1.4**　**归纳法**　证明：对一切 $n \geq 1$，$1 + 3 + 5 + \cdots + (2n-1) = n^2$。

解　记 $P(n)$ 为该等式。奠基：$n = 1$ 时左边 $= 1 = 1^2$。归纳步：设 $P(n)$ 成立，则

$$1 + 3 + \cdots + (2n-1) + (2n+1) = n^2 + (2n+1) = (n+1)^2,$$

这就是 $P(n+1)$。由数学归纳法，等式对一切 $n \geq 1$ 成立。 ♦

归纳法只验证公式，不解释公式从何而来。图 1.3 给出原因：把 $n \times n$ 的方阵一层层剥开，每一层是一个 L 形，第 k 层恰好有 $2k-1$ 个方格。归纳步「$n^2 + (2n+1) = (n+1)^2$」就是「给 $n \times n$ 的方阵再套一层 L 形」。



第 k 层 L 形有 $2k-1$ 个方格（数字标在拐角处）。

$1+3+5+7+9=25=5^2$。

图 1.3　奇数和等于平方数的几何解释。

良序原理最常见的用法是「最小反例」。下面是一个经典例子。

例 1.5　良序原理　用良序原理证明 $\sqrt{2}$ 是无理数。

解　假设 $\sqrt{2}=p/q$，其中 p,q 是正整数，则 $q\sqrt{2}=p$ 是整数。于是集合

$$S=\{\,b\in\mathbb{N}\mid b\geq 1,\ b\sqrt{2}\in\mathbb{Z}\,\}$$

非空（$q\in S$）。由良序原理，S 有最小元 b。

现在用 b 造一个更小的元素。令 $b'=b\sqrt{2}-b$。

- b' 是整数：它是整数 $b\sqrt{2}$ 与整数 b 之差。
- $0<b'<b$：因为 $1<\sqrt{2}<2$，所以 $0<\sqrt{2}-1<1$，而 $b'=b(\sqrt{2}-1)$。
- $b'\sqrt{2}$ 是整数：$b'\sqrt{2}=(b\sqrt{2}-b)\sqrt{2}=2b-b\sqrt{2}$，是两个整数之差。

所以 $b'\in S$ 且 $b'<b$，与 b 是最小元矛盾。因此 $\sqrt{2}$ 是无理数。 ♦

方法：最小反例　要证明「不存在满足某性质的正整数」，可以假设存在，用良序原理取最小的一个，再由它造出一个更小的，得到矛盾。Fermat 把这种论证称为无穷递降法。第 23 讲会用它证明 $x^4+y^4=z^2$ 没有正整数解。

易错提醒　良序原理只对 $\mathbb{N}$ 的子集（更一般地，对 $\mathbb{Z}$ 中有下界的子集）成立。$\mathbb{Z}$ 本身没有最小元；正有理数集 $\{\,x\in\mathbb{Q}\mid x>0\,\}$ 也没有最小元，因为对任何正有理数 x，$x/2$ 更小。「取最小元」之前，先确认集合是 $\mathbb{N}$ 的非空子集。上例中正是 $q\in S$ 保证了非空。

1.4　整除

定义 1.5（整除）　设 $a,b\in\mathbb{Z}$，$a\neq 0$。如果存在整数 x 使 $b=ax$，就说 a 整除（divides）b，记作 $a\mid b$；也说 a 是 b 的因数（divisor），b 是 a 的倍数（multiple）。否则记作 $a\nmid b$。

例如 $3\mid 12$（$12=3\cdot 4$），$-3\mid 12$（$12=(-3)\cdot(-4)$），$5\nmid 12$，$7\mid 0$（$0=7\cdot 0$）。

易错提醒　$a \mid b$ 是一个**命题**（非真即假），不是一个数；a/b 才是数。$3 \mid 12$ 为真，而 $12 \mid 3$ 为假，所以「$\mid$」的左右不能交换。另外，按本讲义的定义 $0 \mid b$ 无意义（要求 $a \neq 0$），常见的另一种约定是「$0 \mid b$ 当且仅当 $b = 0$」。两种约定都不影响后面的结论。

命题 1.6（整除的基本性质）　设 $a, b, c \in \mathbb{Z}$，且下面出现的除数都不为 0。

(1) 对任意 $a \neq 0$，$a \mid 0$；对任意 b，$1 \mid b$。

(2)（传递性）若 $a \mid b$ 且 $b \mid c$，则 $a \mid c$。

(3)（线性组合）若 $a \mid b$ 且 $a \mid c$，则对任意整数 x, y 有 $a \mid bx + cy$。

(4) 若 $a \mid b$ 且 $b \neq 0$，则 $|a| \leq |b|$。

(5) 若 $a \mid b$ 且 $b \mid a$，则 $a = \pm b$。

证明　(1) $0 = a \cdot 0$，$b = 1 \cdot b$。

(2) 设 $b = ax$，$c = by$，则 $c = a(xy)$，而 $xy \in \mathbb{Z}$。

(3) 设 $b = ak$，$c = al$，则 $bx + cy = akx + aly = a(kx + ly)$。

(4) 设 $b = ax$。因为 $b \neq 0$，$x \neq 0$，所以 $|x| \geq 1$，从而 $|b| = |a||x| \geq |a|$。

(5) 由 (4) 得 $|a| \leq |b|$ 且 $|b| \leq |a|$（这里 a, b 都作为除数出现，故都不为 0），所以 $|a| = |b|$。 ■

性质 (3) 是本讲最常用的工具：**一个数若整除 b 和 c，就整除它们的任何整系数组合**。用它的诀窍是：选一个好的组合，把 n 消掉。

■ **例 1.6**　设 n 是整数，d 是正整数，且 $d \mid 5n + 3$、$d \mid 3n + 2$。求 d。

解　想办法消去 n：$5n + 3$ 乘 3，$3n + 2$ 乘 5，再相减：

$$3(5n + 3) - 5(3n + 2) = 15n + 9 - 15n - 10 = -1.$$

由性质 (3)，$d \mid -1$，而 d 是正整数，所以 $d = 1$。这说明对任何整数 n，分数 $\dfrac{5n + 3}{3n + 2}$ 都已是最简分数；分母模 3 余 2，所以不为 0。 ♦

■ **例 1.7**　求所有整数 n，使 $n + 3 \mid n^2 + 2$。

解　首先 $n \neq -3$，否则除数为 0。思路仍是用线性组合消去 n。因为 $n - 3$ 为整数，$n + 3 \mid (n + 3)(n - 3) = n^2 - 9$。若又有 $n + 3 \mid n^2 + 2$，则由性质 (3)，

$$n + 3 \mid (n^2 + 2) - (n^2 - 9) = 11.$$

11 的因数只有 $\pm 1, \pm 11$，所以 $n + 3 \in \{1, -1, 11, -11\}$，即 $n \in \{-2, -4, 8, -14\}$。

反过来要逐个验证，因为上面只推出了必要条件：

n	$n+3$	n^2+2	商
-2	1	6	6
-4	-1	18	-18
8	11	66	6
-14	-11	198	-18

四个值都满足条件，所以答案是 $n \in \{-14, -4, -2, 8\}$。 ♦

想一想　上例为什么要「反过来验证」？在这道题里能省略吗？

答　推理「$n+3 \mid n^2+2 \Rightarrow n+3 \mid 11$」只给出必要条件。这里其实是充要的：$n^2+2 = (n+3)(n-3)+11$，所以 $n+3 \mid n^2+2$ 当且仅当 $n+3 \mid 11$。但养成验证的习惯更安全，很多题目只能推出必要条件。

把良序原理和传递性结合起来，可以证明每个大于 1 的整数都有素因数。先给出素数的定义（第 2 讲会详细讨论）。

定义 1.7（素数）　大于 1 的整数 p，如果它的正因数只有 1 和 p，就称为**素数**（prime）；大于 1 而不是素数的整数称为**合数**（composite）。

■ **例 1.8　最小反例**　证明：每个整数 $n \geq 2$ 都至少有一个素因数。

解　假设结论不对，则集合 $S = \{n \geq 2 \mid n \text{ 没有素因数}\}$ 非空，由良序原理有最小元 m。$m \mid m$，所以 m 本身不是素数（否则 m 就是自己的素因数），即 m 是合数：$m = ab$，其中 $1 < a < m$。因为 $a < m$ 且 $a \geq 2$，$a \notin S$，所以 a 有素因数 p。由 $p \mid a$、$a \mid m$ 和传递性，$p \mid m$，与 $m \in S$ 矛盾。 ♦

1.5　带余除法

小学就学过「17 除以 5 商 3 余 2」，即 $17 = 5 \cdot 3 + 2$，余数 2 满足 $0 \leq 2 < 5$。下面的定理说明：对任何整数都能这样做，而且结果唯一。这是整个初等数论的出发点。

定理 1.8（带余除法）　设 $a, b \in \mathbb{Z}$，$a > 0$。则存在唯一的一对整数 q, r，使

$$b = aq + r, \qquad 0 \leq r < a.$$

q 叫**商**（quotient），r 叫**余数**（remainder）。

直观理解　在数轴上标出 a 的所有倍数 $\ldots, -a, 0, a, 2a, \ldots$，它们把数轴切成长度为 a 的小段。任何整数 b 都落在某一段 $[qa, (q+1)a)$ 里（左闭右开），q 就是商，b 到左端点 qa 的距离 $r = b - qa$ 就是余数，自然有 $0 \leq r < a$。左闭右开保证了 b 只属于一段，这就是唯一性。

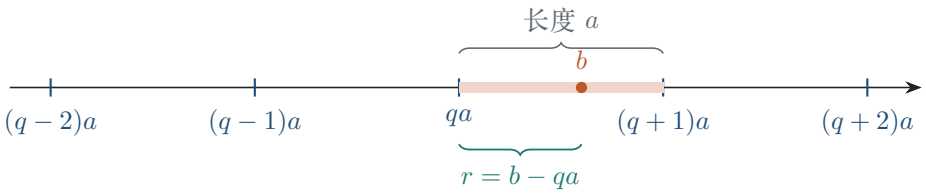


图 1.4　带余除法的几何意义：b 落在区间 $[qa, (q+1)a)$ 中，余数是它到左端点的距离。

思路　存在性要找到「最靠右且不超过 b 的 a 的倍数」，等价于找「b 减去 a 的倍数后最小的非负结果」。「最小的非负」提示用良序原理。唯一性则用「两个余数之差的绝对值小于 a，却又是 a 的倍数」。

证明　存在性。令

$$S = \{\, b + ka \mid k \in \mathbb{Z},\ b + ka \geq 0 \,\} \subseteq \mathbb{N}.$$

先说明 S 非空。若 $b \geq 0$，取 $k = 0$ 得 $b \in S$。若 $b < 0$，取 $k = -b > 0$，则 $b + ka = b - ba = b(1-a)$；因为 $b < 0$ 且 $1 - a \leq 0$，乘积 ≥ 0，所以 $b(1-a) \in S$。这也给出习题集 1 第 1 题所需的严格正数：若得到的值是 0，把 k 加 1 就得到 $a > 0$。

由良序原理，S 有最小元，记为 $r = b + ka$。令 $q = -k$，则 $b = aq + r$，且 $r \geq 0$（因为 $r \in S$）。若 $r \geq a$，则 $r - a = b + (k-1)a \geq 0$ 也属于 S，却比 r 小，矛盾。所以 $0 \leq r < a$。

唯一性。设 $b = aq + r = aq' + r'$，$0 \leq r, r' < a$。则 $a(q - q') = r' - r$，所以 $a \mid r' - r$。而 $-a < r' - r < a$，在这个范围里 a 的倍数只有 0，故 $r = r'$，进而 $aq = aq'$，$q = q'$。 ■

注 1.1　$a \mid b$ 当且仅当 b 除以 a 的余数 $r = 0$。所以「判断整除」总可以化为「计算余数」。商有显式公式 $q = \lfloor b/a \rfloor$，余数 $r = b - a\lfloor b/a \rfloor$。

■ **例 1.9**　对下列 (a, b) 求商 q 和余数 r：(1) $(5, 17)$；(2) $(5, -17)$；(3) $(7, -1)$；(4) $(3, 0)$。

解　关键是始终要求 $0 \leq r < a$。

(1) $17 = 5 \cdot 3 + 2$，$q = 3$，$r = 2$。

(2) $-17 = 5 \cdot (-4) + 3$，$q = -4$，$r = 3$。注意 $-17 = 5 \cdot (-3) - 2$ 不行，因为余数 $-2 < 0$。按图 1.4 的几何意义，-17 落在区间 $[-20, -15)$ 里，左端点是 $5 \cdot (-4)$，到左端点的距离是 3。

(3) $-1 = 7 \cdot (-1) + 6$，$q = -1$，$r = 6$。

(4) $0 = 3 \cdot 0 + 0$，$q = 0$，$r = 0$。

用公式核对 (2)：$\lfloor -17/5 \rfloor = \lfloor -3.4 \rfloor = -4$，$r = -17 - 5 \cdot (-4) = 3$。 ♦

易错提醒　很多编程语言对负数取余的结果和数学约定不同。C、Java 中 `-17 % 5` 得 -2（商向 0 取整）；Python 中 `-17 % 5` 得 3（商向下取整），与定理 1.8 一致。写程序验证数论结论时要留意。

带余除法的一个直接用途是**分类讨论**：除以 a 的余数只有 $0, 1, \ldots, a-1$ 共 a 种可能，所以

任何整数都可以写成 aq，$aq+1$，$\ldots$，$aq+(a-1)$ 之一。

■ **例 1.10**　证明：任何整数的平方除以 4 的余数只能是 0 或 1。由此说明 2023 不能写成两个整数的平方和。

解　按 n 除以 2 的余数分类。若 $n=2k$，则 $n^2=4k^2$，余数 0。若 $n=2k+1$，则 $n^2=4k^2+4k+1=4(k^2+k)+1$，余数 1。

于是 x^2+y^2 除以 4 的余数只能是 $0+0$、$0+1$、$1+1$，即 $0,1,2$ 之一，不可能是 3。而 $2023=4\cdot 505+3$，余数为 3，所以 2023 不是两个平方数之和。 ♦

前后联系　这是「用余数排除整数解」的第一个例子，和本讲开头 $x^2+y^2=3$ 的例子是同一类问题。第 3 讲引入同余记号后，这类论证可以写得更简洁；第 17 讲会给出哪些整数能写成两个平方和的完整答案。

■ **例 1.11**　证明：对任意整数 n，$6\mid n(n+1)(n+2)$。

解　记 $N=n(n+1)(n+2)$。

第一步：$3\mid N$。按 n 除以 3 的余数分类：若 $n=3k$，则 $3\mid n$；若 $n=3k+1$，则 $n+2=3k+3=3(k+1)$；若 $n=3k+2$，则 $n+1=3(k+1)$。三种情况下三个因子中总有一个是 3 的倍数，所以 $3\mid N$。

第二步：$2\mid N$。n 与 $n+1$ 中恰有一个是偶数，所以 $2\mid N$。

第三步：合起来。由前两步，$N=3s$ 且 $N=2t$。注意 $N=3N-2N=3\cdot 2t-2\cdot 3s=6(t-s)$，所以 $6\mid N$。 ♦

想一想　第三步为什么不能直接说「$2\mid N$ 且 $3\mid N$，所以 $6\mid N$」？「$4\mid N$ 且 $6\mid N$，所以 $24\mid N$」对吗？

答　这一步需要理由。第二个说法是错的：$N=12$ 满足 $4\mid 12$、$6\mid 12$，但 $24\nmid 12$。第三步之所以成立，是因为用到了 $1=3\cdot 1-2\cdot 1$，即 2 和 3 的某个整系数组合等于 1；而 4 和 6 的整系数组合 $4x+6y$ 总是偶数，不可能等于 1。这正是下一节 Bézout 定理和推论 1.18 要讲的内容。

1.6　最大公因数与 Bézout 定理

定义 1.9（最大公因数）　设整数 a,b 不全为 0。同时整除 a 和 b 的整数称为它们的**公因数**（common divisor）；公因数中最大的那个称为**最大公因数**（greatest common divisor），记作 $\gcd(a,b)$，常简写为 (a,b)。

定义里的「最大」是有意义的：a,b 至少有一个非零，取其中一个非零数 u。由命题 1.6(4)，每个公因数 d 都满足 $|d|\le|u|$，所以公因数只有有限个；而 1 总是公因数，所以其中有最大的正整数。

■ **例 1.12**　求 $\gcd(12,18)$、$\gcd(0,7)$、$\gcd(-8,12)$。

解　分别列出正因数：

$$12：\ 1,2,3,4,6,12;\qquad 18：\ 1,2,3,6,9,18.$$

公共的有 $1,2,3,6$，所以 $\gcd(12,18)=6$。

每个非零整数都整除 0，所以 0 与 7 的公因数就是 7 的因数，$\gcd(0,7)=7$。

$d\mid -8$ 当且仅当 $d\mid 8$，所以符号不影响最大公因数：$\gcd(-8,12)=\gcd(8,12)=4$。 ♦

列出全部因数的做法对大数不可行。我们换一个角度：考察 a 与 b 的所有整系数线性组合 $ax+by$。下表列出 $a=12$，$b=18$ 时 $12x+18y$ 的一部分取值。

$12x+18y$	$y=-2$	$y=-1$	$y=0$	$y=1$	$y=2$
$x=-2$	-60	-42	-24	-6	12
$x=-1$	-48	-30	-12	$\mathbf{6}$	24
$x=0$	-36	-18	0	18	36
$x=1$	-24	-6	12	30	48
$x=2$	-12	$\mathbf{6}$	24	42	60

表 1.1　$12x+18y$ 的取值。所有值都是 6 的倍数，最小的正值恰好是 $6=\gcd(12,18)$。

表中有两个现象。第一，每个值都是 6 的倍数，原因是 $6\mid 12$、$6\mid 18$，由命题 1.6(3)，6 整除任意组合。第二，6 本身能取到，例如 $12\cdot(-1)+18\cdot 1=6$。下面的定理说明第二个现象不是巧合。

定理 1.10（Bézout 定理）　设整数 a,b 不全为 0。则 $\gcd(a,b)$ 等于集合 $\{\,ax+by\mid x,y\in\mathbb{Z}\,\}$ 中最小的正整数。特别地，存在整数 x,y 使 $ax+by=\gcd(a,b)$。

思路　设最小的正组合为 g。要证 $g=\gcd(a,b)$，需要两件事：(i) g 是公因数；(ii) 任何公因数 d 都不超过 g。(ii) 很容易，因为 d 整除任意组合。(i) 用带余除法：把 a 除以 g，余数仍是一个组合，又比 g 小，只能是 0。

证明　令 $T=\{\,ax+by\mid x,y\in\mathbb{Z},\ ax+by>0\,\}$。$T$ 非空：若 $a\neq 0$，则 $a\cdot a+b\cdot 0=a^2>0$ 属于 T；若 $a=0$，则 $b\neq 0$，$b^2\in T$。由良序原理，T 有最小元 $g=ax_0+by_0$。

g 整除 a。由带余除法，$a=gq+r$，$0\le r<g$。于是

$$r=a-gq=a-(ax_0+by_0)q=a(1-qx_0)+b(-qy_0),$$

也是 a,b 的整系数组合。若 $r>0$，则 $r\in T$ 且 $r<g$，与 g 最小矛盾。所以 $r=0$，$g\mid a$。同理 $g\mid b$。于是 g 是公因数，$g\le\gcd(a,b)$。

任何公因数都不超过 g。设 $d\mid a$，$d\mid b$。由命题 1.6(3)，$d\mid ax_0+by_0=g$，又 $g>0$，由命题 1.6(4)，$d\le|d|\le g$。取 $d=\gcd(a,b)$ 得 $\gcd(a,b)\le g$。

两方面合起来，$g = \gcd(a,b)$。 ■

证明的第二部分其实给出了更强的结论，值得单独记下来。

推论 1.11　设整数 a,b 不全为 0。a,b 的每个公因数都整除 $\gcd(a,b)$。

证明　设 $d \mid a$ 且 $d \mid b$。由 Bézout 定理，存在整数 x_0, y_0 使 $\gcd(a,b) = ax_0 + by_0$，所以 $d \mid \gcd(a,b)$。 ■

所以「最大」公因数不仅在大小上最大，在整除关系下也「最大」：所有公因数都是它的因数。

推论 1.12　设整数 a,b 不全为 0，记 $g = \gcd(a,b)$。则 a,b 的全部整系数线性组合恰好是 g 的全部倍数：

$$\{\, ax + by \mid x, y \in \mathbb{Z} \,\} = g\mathbb{Z} = \{\dots, -2g, -g, 0, g, 2g, \dots\}.$$

证明　$\subseteq$：$g \mid a, g \mid b$，所以 $g \mid ax+by$。$\supseteq$：由 Bézout 定理，$g = ax_0+by_0$，于是 $kg = a(kx_0)+b(ky_0)$ 是组合。 ■

核心思想　Bézout 定理把「整除」这一乘法性质，和「线性组合」这一加法性质联系起来：

$$\gcd(a,b) \text{ 是} ax + by \text{ 的最小正值.}$$

本讲之后几乎每一个关于最大公因数和互素的结论都用这句话证明。

现在可以完整回答本讲开头 $6x + 9y = c$ 的问题了。

推论 1.13（线性丢番图方程的可解性）　设整数 a,b,c 中 a,b 不全为 0。方程 $ax + by = c$ 有整数解，当且仅当 $\gcd(a,b) \mid c$。

证明　由推论 1.12，$ax + by$ 能取到的值恰好是 $\gcd(a,b)$ 的倍数。 ■

■ **例 1.13**　判断 $6x + 9y = c$ 在 $c = 10, 15, 3$ 时是否有整数解，有解时求出一组解。

解　$\gcd(6,9) = 3$。

- $c = 10$：$3 \nmid 10$，无解。这与本讲开头的观察一致：左边总是 3 的倍数。
- $c = 3$：试几个小值，$6 \cdot (-1) + 9 \cdot 1 = 3$，得 $(x,y) = (-1,1)$。
- $c = 15$：$15 = 5 \cdot 3$，把上面的解乘 5：$6 \cdot (-5) + 9 \cdot 5 = -30 + 45 = 15$，得 $(x,y) = (-5,5)$。

解不唯一。对任意整数 t，$6 \cdot 3t + 9 \cdot (-2t) = 0$，所以 $(x,y) = (-5+3t,\ 5-2t)$ 都是 $c = 15$ 的解。$t = 1$ 给出 $(-2,3)$：$-12 + 27 = 15$；$t = 2$ 给出 $(1,1)$：$6 + 9 = 15$。第 2 讲会证明这就是全

部解。 ♦

图 1.5 对比可解与不可解的两条直线：区别在于右端是否被最大公因数整除。

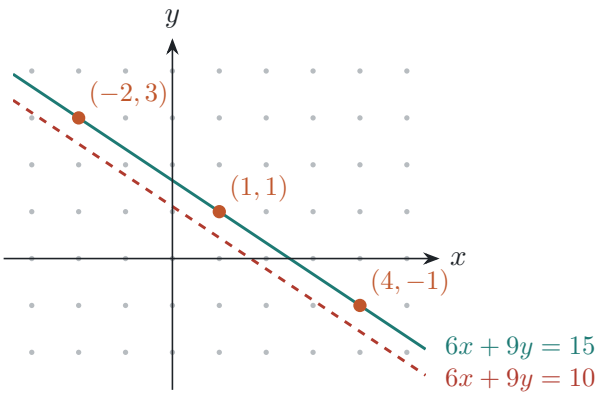


图 1.5　两条平行直线。实线 $6x+9y=15$ 每隔 $(3,-2)$ 穿过一个格点（橙色）；虚线 $6x+9y=10$ 一个格点也不经过。

直观理解　在平面上看，$6x+9y=c$ 是一族斜率为 $-\frac{2}{3}$ 的平行直线。格点 (x,y) 代入后 $6x+9y$ 只能是 3 的倍数，所以只有 $c\in\{\dots,-3,0,3,6,\dots\}$ 对应的那些直线才经过格点，其余直线（例如 $c=10$）夹在中间，碰不到任何格点。

注 1.2　上面的证明用良序原理说明了 x,y 的存在性，但没有告诉我们怎样求出它们。对 $6,9$ 可以凑，对 $\gcd(1071,462)$ 这样的数就不行了。第 2 讲的欧几里得算法（Euclidean algorithm）既能快速算出 $\gcd(a,b)$，又能倒推出 x,y。先看一眼：$18=12\cdot1+6$，$12=6\cdot2+0$，所以 $\gcd(12,18)=6$，且由第一式 $6=12\cdot(-1)+18\cdot1$，正是表 1.1 中 $x=-1,y=1$ 那一格（表中两个加粗的 6 对应两组不同的系数）。

1.7　互素及其三个推论

定义 1.14（互素）　设整数 a,b 不全为 0。若 $\gcd(a,b)=1$，则称 a 与 b 互素（relatively prime, coprime）。

例如 $\gcd(8,15)=1$，所以 8 与 15 互素，尽管它们都不是素数。互素描述的是两个数之间的关系，不是某一个数的性质。

命题 1.15　设整数 a,b 不全为 0。$\gcd(a,b)=1$ 当且仅当存在整数 x,y 使 $ax+by=1$。

证明　「⇒」就是 Bézout 定理。「⇐」：若 $ax+by=1$，则 $\gcd(a,b)\mid 1$，而 $\gcd(a,b)\geq 1$，只能等于 1。 ■

易错提醒　一般地，由 $ax+by=g$ 推不出 $g=\gcd(a,b)$。例如 $12\cdot 1+18\cdot 1=30$，但 $\gcd(12,18)=6\neq 30$。能推出的只是 $\gcd(a,b)\mid g$。命题 1.15 之所以成立，是因为 1 的正因数只有它自己。

命题 1.15 把「互素」变成了一个可以做代数运算的等式 $ax+by=1$。下面三个推论都是「写出这个等式，乘上点什么」。

推论 1.16　设整数 a,b,m 满足 $\gcd(a,m)=\gcd(b,m)=1$。则 $\gcd(ab,m)=1$。

证明　由命题 1.15，有 $ax+my=1$，$bu+mv=1$。两式相乘：

$$1=(ax+my)(bu+mv)=ab\cdot(xu)+m\cdot(axv+ybu+myv).$$

这是 ab 与 m 的整系数组合等于 1，再用一次命题 1.15。 ■

推论 1.17（Euclid 引理的一般形式）　设整数 a,b,c 且 $c\neq 0$。若 $c\mid ab$ 且 $\gcd(c,a)=1$，则 $c\mid b$。

证明　由 $\gcd(c,a)=1$，有 $cx+ay=1$。两边乘 b：

$$b=c(bx)+(ab)y.$$

右边第一项是 c 的倍数；第二项中 $c\mid ab$，所以也是 c 的倍数。由命题 1.6(3)，$c\mid b$。 ■

条件 $\gcd(c,a)=1$ 不能去掉：$6\mid 36=4\cdot 9$，但 $6\nmid 4$，$6\nmid 9$。当 $c=p$ 是素数时，$\gcd(p,a)$ 只能是 1 或 p，于是得到经典的 **Euclid 引理**：若 $p\mid ab$，则 $p\mid a$ 或 $p\mid b$。第 2 讲将用它证明算术基本定理（唯一分解定理）。

推论 1.18　设整数 a,b,n 且 $a,b\neq 0$。若 $a\mid n$，$b\mid n$，且 $\gcd(a,b)=1$，则 $ab\mid n$。

证明　设 $n=as=bt$，$ax+by=1$。两边乘 n：

$$n=anx+bny=a(bt)x+b(as)y=ab(tx+sy).$$

所以 $ab\mid n$。 ■

这正是例子「$6\mid n(n+1)(n+2)$」第三步背后的一般原理，那里用的 $3\cdot 1-2\cdot 1=1$ 就是 $2,3$ 的 Bézout 等式。互素条件同样不能去掉：$4\mid 12$，$6\mid 12$，但 $24\nmid 12$。

图 1.6 汇总 Bézout 定理怎样推出互素条件下的整除性质。

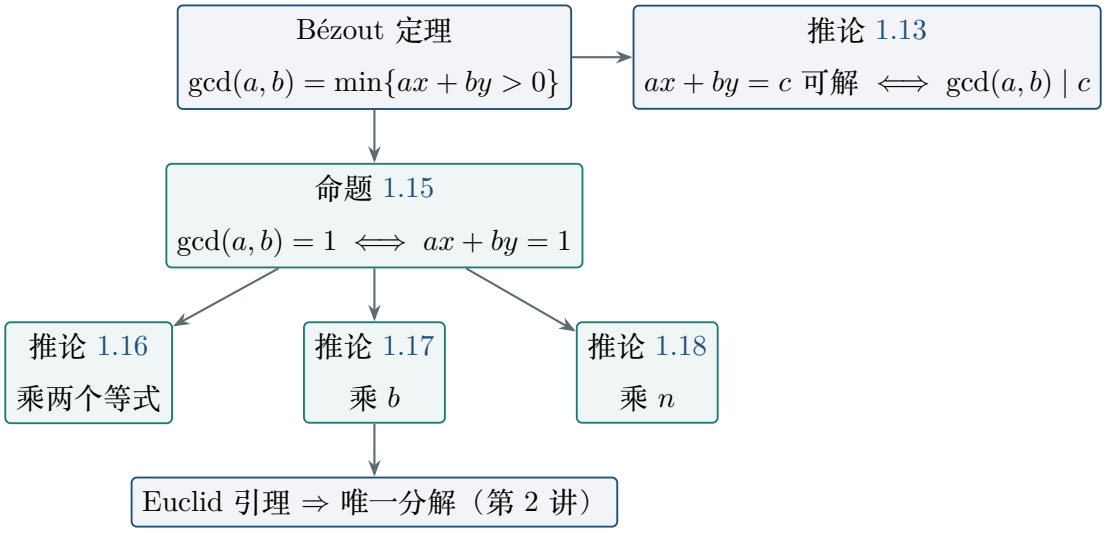


图 1.6　本讲后半部分的逻辑关系：一切都从 Bézout 定理出发。

■ **例 1.14**　证明：对任意整数 n，$\gcd(2n+1,\,3n+1)=1$。

解　找一个消去 n 的组合：$3(2n+1)-2(3n+1)=6n+3-6n-2=1$。由命题 1.15，二者互素。也可以直接说：公因数 d 整除这个组合，即 $d\mid 1$。 ♦

■ **例 1.15　IMO 1959 第 1 题**　证明：对每个自然数 n，分数 $\dfrac{21n+4}{14n+3}$ 都是既约分数。

解　既约就是分子分母互素。要消去 n，取 21 与 14 的最小公倍数 42：

$$3(14n+3)-2(21n+4)=42n+9-42n-8=1.$$

所以 $\gcd(21n+4,14n+3)=1$。一道国际奥赛题，用 Bézout 的观点只需一行。 ♦

■ **例 1.16**　设 $\gcd(a,b)=1$。证明 $\gcd(a+b,\,a-b)$ 等于 1 或 2，并说明两种情况都可能出现。

解　记 $d=\gcd(a+b,a-b)$。d 整除和与差：$d\mid (a+b)+(a-b)=2a$，$d\mid (a+b)-(a-b)=2b$。由 $ax+by=1$ 得 $2ax+2by=2$，而 d 整除左边，所以 $d\mid 2$，$d\in\{1,2\}$。

$a=2,b=1$：$\gcd(3,1)=1$。$a=3,b=1$：$\gcd(4,2)=2$。进一步，$d=2$ 当且仅当 $a+b$ 为偶数，即 a,b 同为奇数（它们互素，不能同为偶数）。 ♦

历史与背景　带余除法与求最大公因数的辗转相除法出现在欧几里得《几何原本》（约公元前 300 年）第 VII 卷。整数情形的 Bézout 等式由法国数学家 Bachet de Méziriac 在其趣味数学书（1624 年第二版）中给出；Étienne Bézout 在 1779 年证明了多项式的类似结论。现在人们习惯把整数版本也称为 Bézout 定理。

本讲小结

- 丢番图方程只问整数解。最常用的排除手段是看余数或看公因数。
- 数学归纳法 $\iff$ 良序原理。用良序原理的标准写法是「取最小反例，再造一个更小的」。
- $a \mid b$ 是命题。公因数整除任意整系数线性组合。
- 对 $a, b \in \mathbb{Z}$ 且 $a > 0$，带余除法 $b = aq + r$ 的商余数存在且唯一，$0 \le r < a$，$q = \lfloor b/a \rfloor$。
- 对不全为 0 的整数 a, b，Bézout：$\gcd(a,b)$ 是 $ax+by$ 的最小正值；组合的全体是 $\gcd(a,b)\mathbb{Z}$；$ax + by = c$ 可解 $\iff$ $\gcd(a,b) \mid c$。
- 对不全为 0 的整数 a, b，互素 $\iff$ $ax + by = 1$。由此得到三个推论：乘法封闭、Euclid 引理；若 $a, b \neq 0$、$\gcd(a,b) = 1$ 且 $a \mid n$, $b \mid n$，则 $ab \mid n$。

练习

1.1 ★ 判断真假，假的给出反例：(a) 若 $a \mid b$ 且 $a \mid c$，则 $a \mid b + c$；(b) 若 $a \mid bc$，则 $a \mid b$ 或 $a \mid c$；(c) 若 $a \mid b$ 且 $b \mid a$，则 $a = b$；(d) 若 $a \mid b$，则 $a^2 \mid b^2$。

1.2 ★ 对 $(a, b) = (7, -30)$，$(4, -1)$，$(13, -13)$ 分别求带余除法 $b = aq + r$ 中的 q, r。

1.3 ★ 证明：任何奇数的平方除以 8 余 1。由此说明方程 $x^2 = 8y + 5$ 没有整数解。

1.4 ★★ 求所有整数 $n \neq -1$，使 $n + 1 \mid n^2 + 1$。

1.5 ★ 设整数 a, b 不全为 0。证明：对任意整数 k，$\gcd(a, b) = \gcd(a,\, b + ka)$。（提示：两对数的公因数完全相同。这是第 2 讲欧几里得算法的基础。）

1.6 ★★ 设整数 a, b 不全为 0，且 $g = \gcd(a, b)$。证明 $\gcd\left(\dfrac{a}{g}, \dfrac{b}{g}\right) = 1$。

1.7 ★★ 仿照本讲 $\sqrt{2}$ 的证明，用良序原理证明 $\sqrt{3}$ 是无理数。

1.8 ★★ 证明：对每个正整数 n，$\gcd(n! + 1, (n + 1)! + 1) = 1$。

1.9 ★★ 斐波那契数列定义为 $F_1 = F_2 = 1$，$F_{n+1} = F_n + F_{n-1}$。用数学归纳法证明相邻两项互素：$\gcd(F_n, F_{n+1}) = 1$。

1.10 ★★★ 只用 3 分和 5 分的邮票（每种数量不限），哪些正整数邮资凑不出来？证明最大的凑不出邮资是 $3 \cdot 5 - 3 - 5$。推广：对互素正整数 a, b，若其中一个等于 1，说明为什么没有凑不出的正整数；若 $a, b > 1$，证明最大的凑不出邮资是 $ab - a - b$。

练习答案与提示

1.1 (a) 真，命题 1.6(3) 取 $x = y = 1$。(b) 假：$6 \mid 2 \cdot 3$，但 $6 \nmid 2$，$6 \nmid 3$；当 a 为素数时为真（Euclid 引理）。(c) 假：$2 \mid -2$，$-2 \mid 2$，但 $2 \neq -2$；正确结论是 $a = \pm b$。(d) 真：$b = ak$ 推出 $b^2 = a^2k^2$。

1.2 $-30 = 7 \cdot (-5) + 5$，$q = -5$，$r = 5$。$-1 = 4 \cdot (-1) + 3$，$q = -1$，$r = 3$。$-13 = 13 \cdot (-1) + 0$，$q = -1$，$r = 0$。每一个都可以用 $q = \lfloor b/a \rfloor$ 核对。

1.3 奇数 $n = 2k + 1$，$n^2 - 1 = (n - 1)(n + 1) = 4k(k + 1)$。$k(k + 1)$ 是相邻两数之积，是偶数，所以 $8 \mid n^2 - 1$。若 $x^2 = 8y + 5$，则 x^2 为奇数，x 为奇数，x^2 除以 8 余 1，不能余 5，矛盾。

1.4 $n^2+1=(n+1)(n-1)+2$，所以 $n+1 \mid n^2+1 \iff n+1 \mid 2 \iff n+1 \in \{\pm 1, \pm 2\}$，即 $n \in \{0,-2,1,-3\}$。验证：$1 \mid 1$，$-1 \mid 5$，$2 \mid 2$，$-2 \mid 10$，都成立；题面排除的 $n=-1$ 会使除数为 0，不在整除定义的范围内。

1.5 若 $d \mid a$，$d \mid b$，则 $d \mid b+ka$。反之若 $d \mid a$，$d \mid b+ka$，则 $d \mid (b+ka)-ka=b$。两对数的公因数集合相同，最大者也相同。

1.6 由 Bézout，$ax+by=g$，两边除以 g：$\frac{a}{g}x+\frac{b}{g}y=1$。由命题 1.15，二者互素。注意 $\frac{a}{g}, \frac{b}{g}$ 是整数，因为 g 是公因数。

1.7 设 $S=\{b \geq 1 \mid b\sqrt{3} \in \mathbb{Z}\}$ 非空，取最小元 b。令 $b'=b\sqrt{3}-b=b(\sqrt{3}-1)$。因为 $1<\sqrt{3}<2$，有 $0<b'<b$；b' 是整数；$b'\sqrt{3}=3b-b\sqrt{3}$ 也是整数。于是 $b' \in S$ 且比 b 小，矛盾。

1.8 $(n+1)!+1=(n+1)\cdot n!+1=(n+1)(n!+1)-n$。公因数 d 整除 $n!+1$ 和 $(n+1)!+1$，因此 $d \mid n$。而 $n \mid n!$，所以 $d \mid n!$，又 $d \mid n!+1$，故 $d \mid 1$。

1.9 $n=1$：$\gcd(F_1,F_2)=1$。设 $\gcd(F_n,F_{n+1})=1$。由练习 5（取 $k=1$），

$$\gcd(F_{n+1},F_{n+2})=\gcd(F_{n+1},F_n+F_{n+1})=\gcd(F_{n+1},F_n)=1.$$

1.10 先证明推广。设 $a,b>1$ 且 $\gcd(a,b)=1$，令 $F=ab-a-b$。若 $F=ax+by$，其中 $x,y \geq 0$，则

$$a(b-1-x)=b(y+1).$$

因为 $\gcd(a,b)=1$，Euclid 引理给出 $a \mid y+1$，故 $y \geq a-1$。于是 $ax+by \geq b(a-1)=ab-b=F+a>F$，矛盾，所以 F 凑不出。

现在取任意 $n>F$。由 Bézout 定理，存在整数 x,y 使 $n=ax+by$。用带余除法选 k，使 $x-bk=x_0$ 且 $0 \leq x_0<b$，并令 $y_0=y+ak$。则 $n=ax_0+by_0$，而

$$by_0=n-ax_0 \geq F+1-a(b-1)=1-b.$$

因此 $y_0 \geq -(b-1)/b>-1$；y_0 是整数，所以 $y_0 \geq 0$。故每个 $n>F$ 都是非负整系数线性组合。若 $a=1$ 或 $b=1$，每个非负整数本身就是其中一个系数的倍数，所以没有凑不出的正整数。

对 $3,5$，推广给出 $F=7$，且所有 $n \geq 8$ 都能凑出；$1,2,4$ 直接凑不出，而 $3,5,6$ 分别是 3、5、$3+3$。所以正整数中凑不出的恰为 $1,2,4,7$，最大值 $7=3\cdot 5-3-5$。

第 2 讲

欧几里得算法与素数

Euclidean Algorithm, Primes

对应材料：MIT 18.781 Lecture 2 讲义；习题集 1 第 3、4、5、6、7、8 题及官方解答。

本讲目标

(1) 会用欧几里得算法（辗转相除法）求最大公因数，理解它为什么正确、为什么一定停止。

(2) 会用扩展欧几里得算法（表格法）求 Bézout 系数，写出线性丢番图方程 $ax+by=c$ 的全部整数解。

(3) 知道欧几里得算法的步数只有对数量级，理解为什么相邻斐波那契数是最坏情形（Lamé 定理）。

(4) 理解算术基本定理的陈述，看清唯一性的证明在哪里用到了 Euclid 引理；会用素因数分解求最大公因数与最小公倍数。

(5) 掌握素数有无穷多个的 Euclid 证明和 Euler 证明，了解素数分布的基本事实与几个著名猜想。

2.1 欧几里得算法

第 1 讲的 Bézout 定理告诉我们：$\gcd(a,b)$ 是 $ax+by$ 的最小正值。但那个证明用的是良序原理，只说明「最小值存在」，没有告诉我们怎样把它算出来。

最直接的办法是列出 a 和 b 的全部因数再比较，或者先把两数分解成素数。对小数可以这样做；对无特殊结构的大整数，一般因数分解可能很困难，位数本身却不能保证困难，例如 10^{299} 的分解可以直接写出。第 6 讲的 RSA 密码体系会利用适当选择的大整数的分解困难性。好在求最大公因数**不需要分解**，只需要反复做带余除法。依据是第 1 讲练习 5 的结论 $\gcd(a,b)=\gcd(a,\,b+ka)$，取 $k=-q$ 就能把一对大数换成一对小数。

> **引理 2.1** 若 $b=aq+r$（a,b,q,r 都是整数）且 a,b 不全为零，则 $\gcd(a,b)=\gcd(r,a)$。

证明 记 $d=\gcd(r,a)$，$e=\gcd(a,b)$。因为 $d\mid a$，$d\mid r$，所以 d 整除线性组合 $aq+r=b$。于是 d 是 a,b 的公因数，由第 1 讲的推论（公因数整除最大公因数），$d\mid e$。

反过来，$e\mid a$，$e\mid b$，所以 e 整除 $b-aq=r$。于是 e 是 r,a 的公因数，$e\mid d$。两个正整数互相整除，只能相等：$d=e$。 ∎

这个证明的套路值得记住：**要证两个正整数相等，就证它们互相整除。**这比直接比较大小好用得多，后面会反复出现。

算法 2.2（欧几里得算法）　输入不全为零的整数 a,b，输出 $\gcd(a,b)$。

(1) 把 a,b 换成 $|a|,|b|$。

(2) 若 $a>b$，交换 a,b，使 $a\le b$。

(3) 若 $a=0$，输出 b，停止。

(4) 做带余除法 $b=aq+r$，$0\le r<a$。把 (a,b) 换成 (r,a)，回到第 3 步。

定理 2.3　对任意不全为零的整数 a,b，欧几里得算法在有限步后停止，并且输出的就是 $\gcd(a,b)$。

思路　正确性要说明每一步都不改变「当前这对数的最大公因数」，这样的量叫作**不变量**（invariant）。终止性要找一个每步都严格变小、又不能无限变小的量。

证明　*正确性。*第 1 步不改变最大公因数，因为 $d\mid a \iff d\mid -a$。第 2 步只是交换次序。第 4 步由引理 2.1，$\gcd(a,b)=\gcd(r,a)$。所以在整个过程中，当前一对数的最大公因数始终等于最初的 $\gcd(a,b)$。算法停止时这对数是 $(0,b)$，并且 $b>0$：开始时两数不全为零，而每次执行第 4 步后，新的 b 就是旧的 a，它在第 3 步刚被检验过不等于 0。0 和 b 的最大公因数是 b，所以输出正确。

*终止性。*每执行一次第 4 步，较小的数从 a 变成 r，而 $0\le r<a$。于是较小的数构成一列严格递减的非负整数。由良序原理，这样的数列不能无限长，所以第 4 步只能执行有限次。 ■

终止性论证只给出「步数不超过 a」。实际步数要少得多，大约是 a 的位数的常数倍（§2.3）。

■ **例 2.1**　用欧几里得算法求 $\gcd(1071,462)$。

解　*第一步：反复做带余除法。*

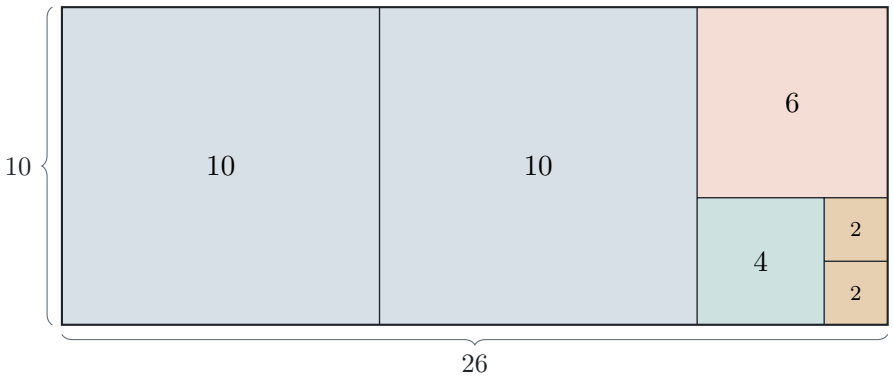
$$\begin{aligned} 1071 &= 462\cdot 2+147, && (462,1071)\to(147,462),\\ 462 &= 147\cdot 3+21, && (147,462)\to(21,147),\\ 147 &= 21\cdot 7+0, && (21,147)\to(0,21). \end{aligned}$$

*第二步：读出答案。*较小的数变成 0 时停止，$\gcd(1071,462)=21$，只做了三次除法。每一行的余数都是下一行的除数，「辗转相除」由此得名；最后一个非零余数就是最大公因数。

第三步：检验。$1071=21\cdot 51$，$462=21\cdot 22$，而 $51=3\cdot 17$ 与 $22=2\cdot 11$ 没有公共素因数。 ♦

欧几里得算法有一个漂亮的几何解释。在 $a\times b$ 的长方形里，沿短边尽量多地切出正方形；剩下的小长方形再这样切，直到恰好切完。切出来的正方形个数依次是各步的商 q，而最后那种正方

形的边长就是 $\gcd(a,b)$。它是能把原长方形无缝铺满的最大正方形的边长。图 2.1 画的是 26×10 的情形。



$$26=10\cdot 2+6 \qquad 10=6\cdot 1+4 \qquad 6=4\cdot 1+2 \qquad 4=2\cdot 2$$

图 2.1　用正方形铺 26×10 的长方形。每一种颜色对应欧几里得算法的一步：先切出 2 个边长 10 的正方形，再切 1 个边长 6 的，1 个边长 4 的，最后 2 个边长 2 的恰好铺满。所以 $\gcd(26,10)=2$。

直观理解　倒着看图 2.1：最小的正方形能铺满上一级的长方形，进而铺满整个长方形，所以 g 是公因数。反过来，任何公因数 d 的方格网都对齐每一刀切口，所以 $d\mid g$。这是引理 2.1 的几何版本。

想一想　如果把第 4 步的带余除法换成「大数减去小数」，算法还正确吗？每轮先把两数排成 $u\ge v\ge 0$，再用 $u-v$ 替换 u；用这种办法求 $\gcd(1071,462)$ 要做多少次减法？

答　仍然正确，因为 $\gcd(u,v)=\gcd(u-v,v)$。1071 减两次 462 得 147；重新排序后，462 减三次 147 得 21；再把 147 减七次 21 得 0，共 $2+3+7=12$ 次，恰好是各步商的和。商很大时它很慢：求 $\gcd(1,10^6)$ 要做 10^6 次减法，带余除法一步即可。这就是《九章算术》的「更相减损术」。

2.2　扩展欧几里得算法与线性丢番图方程

欧几里得算法算出了 $g=\gcd(a,b)$，但 Bézout 定理还承诺了整数 x,y 使 $ax+by=g$。这对 x,y 也能从算法中读出来。每一行除法 $b=aq+r$ 都可以改写成 $r=b-aq$，即「余数是被除数和除数的整系数组合」。从最后一个非零余数出发倒着代回去，就把 g 写成了 a,b 的组合。

例 2.2　回代法　求整数 x,y，使 $1071x+462y=21$。

解　第一步：改写除式。把例 2.1 中的除法写成「余数 = 被除数 − 商 × 除数」。

$$21=462-147\cdot 3, \qquad 147=1071-462\cdot 2.$$

第二步：倒着代入。从最后一个非零余数 21 开始，把 147 用上一行代入。

$$21=462-3\,(1071-462\cdot 2)=462-3\cdot 1071+6\cdot 462=7\cdot 462-3\cdot 1071.$$

第三步：检验。$7\cdot 462 = 3234$，$3\cdot 1071 = 3213$，差为 21。所以 $x=-3$，$y=7$。 ♦

回代法要先存下所有除式再倒着整理，步数一多就容易出错。更好的办法是**边做除法边记录系数**：表格每一行记一个数 r 和把它写成 a,b 组合的系数。

■ **例 2.3　表格法**　求 $\gcd(43,27)$，并把它写成 $43x+27y$ 的形式。

解　第一步：写出前两行。每一行 (r,x,y) 都满足 $r = 43x+27y$。前两行是平凡的：$43 = 43\cdot 1+27\cdot 0$，$27 = 43\cdot 0+27\cdot 1$。

第二步：逐行递推。用上一行的 r 去除上上一行的 r，得商 q；新的一行等于「上上一行 $-q\times$ 上一行」，三列都这样算。

q	r	x	y	检验 $43x+27y$
	43	1	0	43
	27	0	1	27
1	16	1	−1	$43-27=16$
1	11	−1	2	$-43+54=11$
1	5	2	−3	$86-81=5$
2	1	−5	8	$-215+216=1$
5	0	27	−43	$1161-1161=0$

例如第四行：$27 = 16\cdot 1+11$，所以 $q=1$，新行 $=(27,0,1)-1\cdot(16,1,-1)=(11,-1,2)$。

第三步：读出答案。最后一个非零的 r 是 1，所以 $\gcd(43,27)=1$，并且 $1 = 43\cdot(-5)+27\cdot 8$。 ♦

表格最后一行 $(0,27,-43)$ 也有用：它说明 $43\cdot 27+27\cdot(-43)=0$。把这一行的任意倍数加到 $(-5,8)$ 上，组合的值不变，所以 $43x+27y=1$ 有无穷多组解。下面的定理说明这样就得到了全部解。

算法 2.4（扩展欧几里得算法）　输入正整数 $a\le b$，输出 $g=\gcd(a,b)$ 以及整数 x,y，使 $bx+ay=g$。

(1) 令 $(r_{-1},x_{-1},y_{-1})=(b,1,0)$，$(r_0,x_0,y_0)=(a,0,1)$，$k=1$。

(2) 若 $r_{k-1}=0$，输出 $g=r_{k-2}$，$x=x_{k-2}$，$y=y_{k-2}$，停止。

(3) 令 $q_k=\lfloor r_{k-2}/r_{k-1}\rfloor$，三列同时递推：$r_k=r_{k-2}-q_kr_{k-1}$，$x_k=x_{k-2}-q_kx_{k-1}$，$y_k=y_{k-2}-q_ky_{k-1}$。把 k 加 1，回到第 2 步。

正确性　r 这一列就是欧几里得算法本身，最后一个非零的 r 等于 $\gcd(a,b)$。对 k 归纳证明 $r_k=bx_k+ay_k$：$k=-1,0$ 直接验证；若第 $k-2,k-1$ 行成立，则

$$bx_k+ay_k=(bx_{k-2}+ay_{k-2})-q_k(bx_{k-1}+ay_{k-1})=r_{k-2}-q_kr_{k-1}=r_k.$$ ■

这个算法同时给出了 Bézout 定理的一个**构造性证明**：它不但说明系数存在，而且把系数算了出来。

定理 2.5（线性丢番图方程的通解）　设 a,b 不全为零，$g=\gcd(a,b)$，$g\mid c$，(x_0,y_0) 是 $ax+by=c$ 的一组整数解。则全部整数解为

$$x=x_0+\frac{b}{g}t,\qquad y=y_0-\frac{a}{g}t,\qquad t\in\mathbb{Z}.$$

证明　它们都是解。$a\big(x_0+\frac{b}{g}t\big)+b\big(y_0-\frac{a}{g}t\big)=ax_0+by_0+\frac{ab}{g}t-\frac{ab}{g}t=c$。

没有别的解。设 (x,y) 是任意一组解。两式相减得 $a(x-x_0)=-b(y-y_0)$。若 $b=0$，则 $g=|a|$，方程化为 $ax=c$，$x=x_0$ 唯一，y 任意，与公式相符（此时 $b/g=0$，$a/g=\pm1$）。下设 $b\neq0$。记 $a'=a/g$，$b'=b/g$，两边除以 g：

$$a'(x-x_0)=-b'(y-y_0).$$

由第 1 讲练习 6，$\gcd(a',b')=1$。b' 整除右边，从而整除 $a'(x-x_0)$，由 Euclid 引理的一般形式，$b'\mid x-x_0$。记 $x-x_0=b't$，代回得 $a'b't=-b'(y-y_0)$，约去 $b'\neq0$ 得 $y-y_0=-a't$。 ■

证明的结构是：两组解之差满足齐次方程 $au+bv=0$，所以「通解 = 特解 + 齐次通解」，和线性微分方程一样。

易错提醒　步长是 $\big(\frac{b}{g},-\frac{a}{g}\big)$，不是 $(b,-a)$。后者也总给出解，但会漏掉大部分解。第 1 讲的方程 $6x+9y=15$ 有特解 $(-5,5)$，$g=3$，通解是 $(-5+3t,\ 5-2t)$。若写成 $(-5+9t,\ 5-6t)$，就会漏掉 $(-2,3)$、$(1,1)$ 等解。几何上，直线 $6x+9y=15$ 上相邻两个格点之差是 $(3,-2)$，而不是 $(9,-6)$。

■ **例 2.4　改编自习题集 1 第 4 题**　求 $89x+43y=1$ 的全部整数解。

解　第一步：表格法。商依次为 $2,14,3$，表格各行 (r,x,y) 为

$$(89,1,0),\ (43,0,1),\ (3,1,-2),\ (1,-14,29),\ (0,43,-89).$$

第二步：特解。$\gcd(89,43)=1$，特解 $(x_0,y_0)=(-14,29)$。检验：$-1246+1247=1$。

第三步：通解。由定理 2.5，全部解为 $x=-14+43t$，$y=29-89t$，$t\in\mathbb{Z}$。例如 $t=1$ 给出 $(29,-60)$：$2581-2580=1$。表格最后一行 $(43,-89)$ 恰好就是步长。 ♦

■ **例 2.5　百鸡问题**　《张丘建算经》（约公元 5 世纪）：「今有鸡翁一，值钱五；鸡母一，值钱三；鸡雏三，值钱一。凡百钱买鸡百只，问鸡翁、母、雏各几何？」

解　第一步：列方程。设公鸡、母鸡、小鸡分别买 x,y,z 只，$x,y,z\geq0$，且 $3\mid z$。则 $x+y+z=100$，$5x+3y+\frac{1}{3}z=100$。第二式乘 3 减去第一式得 $14x+8y=200$，即 $7x+4y=100$。

第二步：解二元方程。$\gcd(7,4)=1$，且 $1=4\cdot 2-7\cdot 1$，乘 100 得特解 $(x_0,y_0)=(-100,200)$。通解 $x=-100+4t$，$y=200-7t$。令 $s=t-25$，化简为 $x=4s$，$y=25-7s$。

第三步：加上非负约束。$x\ge 0$ 要求 $s\ge 0$，$y\ge 0$ 要求 $s\le 3$。所以 $s=0,1,2,3$，对应 $z=100-x-y=75+3s$：

$$(x,y,z)=(0,25,75),\ (4,18,78),\ (8,11,81),\ (12,4,84).$$

z 都是 3 的倍数。检验 $(4,18,78)$：$20+54+26=100$ 钱。原书只给后三组，因为它要求每种鸡都买。 ♦

前后联系　扩展欧几里得算法在后面会反复出现。第 3、4 讲引入同余之后，「解 $ax+by=1$」就变成「求 a 除以 b 的逆元」；第 6 讲 RSA 算法求私钥用的正是这个表格。各步的商 $q_1,q_2,\dots$ 还有另一层含义：$\frac{1071}{462}=2+1/(3+\frac{1}{7})$，商 $2,3,7$ 恰好是连分数展开的各项。

2.3　欧几里得算法有多快

定理 2.3 的证明只说明步数不超过 a。如果 a 有 100 位，这个上界是 10^{100}，毫无用处。实际上，例 2.1 中 $a=462$ 只用了 3 步。下面证明步数总是 $\log a$ 量级。记欧几里得算法依次出现的较小数为 $r_0=a>r_1>r_2>\cdots$，其中 r_{k+1} 是 r_{k-1} 除以 r_k 的余数（$r_{-1}=b$）。

引理 2.6　若 $r_k>0$，则 $r_{k+1}<r_{k-1}/2$。也就是说，每做两步，较小的数至少减半。

证明　记 $a=r_{k-1}$，$r=r_k$，$s=r_{k+1}$，这里 $0<r\le a$（等号只在 $k=0$、$a=b$ 时出现），s 是 a 除以 r 的余数。分两种情况。若 $r\le a/2$，则 $s<r\le a/2$。若 $r>a/2$，则 $a<2r$，商为 1，$s=a-r<a-a/2=a/2$。 ■

推论 2.7（习题集 1 第 5 题）　设 $0<a\le b$。欧几里得算法求 $\gcd(a,b)$ 的除法次数 n 满足 $n\le 2\log_2 a+2$。

证明　设算法做了 n 次除法，则 $r_0,r_1,\dots,r_{n-1}$ 都是正的，$r_n=0$。若 $n\ge 2j+1$，则 $r_1,\dots,r_{2j}$ 都是正的，对 $k=1,3,\dots,2j-1$ 反复用引理 2.6 得 $1\le r_{2j}\le a/2^j$，所以 $2^j\le a$。取 $j=\lfloor (n-1)/2\rfloor$，有 $2j+1\le n\le 2j+2$，于是 $j\le\log_2 a$，$n\le 2j+2\le 2\log_2 a+2$。 ■

对 $a=462$，这个上界是 $2\log_2 462+2\approx 19.7$。对 100 位的 a，$\log_2 a<100\log_2 10\approx 332.2$，步数不超过 666。从 10^{100} 降到几百，这就是「快」的含义。

■ **例 2.6　最慢的情形**　用欧几里得算法求 $\gcd(55,34)$，数一数要做几次除法。

解　第一步：逐步除。

$$55=34\cdot 1+21,\quad 34=21\cdot 1+13,\quad 21=13\cdot 1+8,\quad 13=8\cdot 1+5,$$
$$8=5\cdot 1+3,\qquad 5=3\cdot 1+2,\qquad 3=2\cdot 1+1,\qquad 2=1\cdot 2+0.$$

第二步：结论。$\gcd(55,34)=1$，共 8 次除法，比 $\gcd(1071,462)$ 多得多，虽然这两个数小得多。

第三步：找原因。除最后一步外，每一步的商都是 1，余数只是「减去一次」，下降得最慢。出现的数 $55,34,21,13,8,5,3,2,1$ 正是倒过来的斐波那契数列 $F_1=F_2=1$，$F_{k+1}=F_k+F_{k-1}$。♦

例 2.6 是一般规律：要让欧几里得算法走 n 步，两个数至少要像斐波那契数那么大。

定理 2.8（Lamé 定理） 设 $0<a<b$，欧几里得算法求 $\gcd(a,b)$ 恰好做了 n 次除法，则 $a\geq F_{n+1}$，$b\geq F_{n+2}$。

思路　从最后一步倒推。最后一个非零余数至少是 1；它前面的数严格比它大，至少是 2；再往前，每个数至少是后面两个数之和（因为商至少是 1）。这正是斐波那契数的递推。

证明　沿用记号 $r_{-1}=b$，$r_0=a$，第 k 次除法是 $r_{k-2}=r_{k-1}q_k+r_k$，$k=1,\dots,n$，$r_n=0$，$r_{n-1}=g\geq 1$。我们对 j 归纳证明 $r_{n-j}\geq F_{j+1}$，$j=1,\dots,n+1$。

$j=1$：$r_{n-1}\geq 1=F_2$。$j=2$：$r_{n-2}>r_{n-1}\geq 1$，所以 $r_{n-2}\geq 2=F_3$。这里用到 $r_{n-2}>r_{n-1}$：当 $n\geq 2$ 时它是余数的性质；当 $n=1$ 时 $r_{-1}=b>a=r_0$ 是假设。

$j\geq 3$：第 $n-j+2$ 次除法是 $r_{n-j}=r_{n-j+1}q+r_{n-j+2}$，商 $q\geq 1$（因为 $r_{n-j}>r_{n-j+1}$）。由归纳假设，

$$r_{n-j}\geq r_{n-j+1}+r_{n-j+2}\geq F_j+F_{j-1}=F_{j+1}.$$

取 $j=n$ 和 $j=n+1$，得 $a=r_0\geq F_{n+1}$，$b=r_{-1}\geq F_{n+2}$。■

反过来，$(a,b)=(F_{n+1},F_{n+2})$ 时欧几里得算法恰好走 n 步（练习 2.5），所以 Lamé 定理的界不能再改进。斐波那契数增长得像几何级数，公比是黄金分割比 $\varphi=\frac{1+\sqrt{5}}{2}\approx 1.618$，它满足 $\varphi^2=\varphi+1$。由此可以把 Lamé 定理换成对数形式。

推论 2.9 设 $0<a<b$，欧几里得算法做了 n 次除法，则 $n\leq 1+\log_\varphi a$。若 a 是 d 位十进制数，则 $n\leq 5d$。

证明　先对 k 归纳证明 $F_k\geq\varphi^{k-2}$。$F_1=1>\varphi^{-1}$，$F_2=1=\varphi^0$。若对 $k-1$ 和 k 成立，则

$$F_{k+1}=F_k+F_{k-1}\geq\varphi^{k-2}+\varphi^{k-3}=\varphi^{k-3}(\varphi+1)=\varphi^{k-3}\varphi^2=\varphi^{k-1}.$$

由 Lamé 定理，$a\geq F_{n+1}\geq\varphi^{n-1}$，取对数得 $n\leq 1+\log_\varphi a$。若 $a<10^d$，则 $n<1+d\log_\varphi 10<1+5d$（$\log_\varphi 10\approx 4.785$），而 n 是整数，所以 $n\leq 5d$。■

例如 $a=34$：$1+\log_\varphi 34\approx 8.33$，所以 $n\leq 8$，而例 2.6 恰好走了 8 步。两个 100 位数，$n<1+100\log_\varphi 10\approx 479.5$，最多 479 次除法；而试除法找因数要试到 $\sqrt{a}\approx 10^{50}$。

历史与背景　欧几里得算法出现在《几何原本》第 VII 卷开头，Knuth 称它为「所有算法的老祖宗」。Lamé 在 1844 年证明了推论 2.9，这常被看作计算复杂度分析的开端。

2.4　素数与算术基本定理

定义 2.10　整数 $p>1$ 称为**素数**（prime），如果它不能写成 $p=ab$，其中 a,b 都是大于 1 的整数。大于 1 而不是素数的整数称为**合数**（composite）。

这和第 1 讲的定义（正因数只有 1 和 p）等价。第 1 讲还用最小反例证明了每个 $n\geq 2$ 都有素因数；把这个论证推进一步，就得到整数乘法的基本结构定理。

定理 2.11（算术基本定理）　每个正整数 n 都可以写成有限个素数（允许重复）的乘积，并且除了因子的次序以外，写法是唯一的。约定 1 是「零个素数的乘积」（空积）。

例如 $360=2\cdot 2\cdot 2\cdot 3\cdot 3\cdot 5$。无论怎样分解 360，最后一定恰好得到三个 2、两个 3 和一个 5。定理包含两部分：**存在性**和**唯一性**。存在性很容易。

存在性的证明　设不能写成素数乘积的正整数构成集合 S。若 S 非空，由良序原理取最小元 n。$n\neq 1$（空积），n 也不是素数（一个素数本身就是乘积）。所以 n 是合数，$n=ab$，$1<a,b<n$。由 n 的最小性，$a,b\notin S$，都能写成素数乘积，把两个乘积拼起来就是 n 的分解，矛盾。 ■

唯一性看起来理所当然，其实不然。下面的例子说明，在一个只比 $\mathbb{Z}$ 稍微「小」一点的乘法系统里，唯一性就会失败。

■ **例 2.7　Hilbert 数**　设 $H=\{1,5,9,13,17,21,\dots\}$ 是除以 4 余 1 的正整数。两个这样的数之积仍在 H 中：$(4a+1)(4b+1)=4(4ab+a+b)+1$。称 $h\in H$，$h>1$ 为 H-*素数*，如果它不能写成两个大于 1 的 H 中元素之积。证明分解成 H-素数的方式不唯一。

解　*第一步：找几个 H-素数。*9 的大于 1 的真因数只有 3，而 $3\notin H$，所以 9 是 H-素数。同理 49（真因数只有 $7\notin H$）和 21（真因数 $3,7\notin H$）都是 H-素数。

第二步：同一个数的两种分解。

$$441=9\cdot 49=21\cdot 21.$$

两边都是 H-素数的乘积，而 $\{9,49\}$ 与 $\{21,21\}$ 不同。在 H 中，存在性仍然成立（证明同上），唯一性却失败了。

*第三步：找原因。*21 整除 $9\cdot 49$，却既不整除 9 也不整除 49。整数中保证唯一性的正是「素数整除乘积就整除某个因子」，而这一点来自 Bézout 定理，需要用到 $ax+by$ 这样含减法的组合。H 对减法不封闭（$9-5=4\notin H$），这条路在 H 中走不通。 ♦

引理 2.12（Euclid 引理）　若 p 是素数，$p\mid ab$，则 $p\mid a$ 或 $p\mid b$。

证明　设 $p\nmid a$。$\gcd(p,a)$ 是 p 的正因数，只能是 1 或 p；它不能是 p，否则 $p\mid a$。所以 $\gcd(p,a)=1$。由第 1 讲的 Euclid 引理的一般形式（$c\mid ab$ 且 $\gcd(c,a)=1$ 推出 $c\mid b$），$p\mid b$。 ■

对因子个数归纳（把 $a_1\cdots a_{n+1}$ 看成 $(a_1\cdots a_n)\cdot a_{n+1}$），立刻得到：

推论 2.13　若 p 是素数，$p \mid a_1a_2\cdots a_n$，则存在某个 i 使 $p \mid a_i$。

唯一性的证明　设有正整数具有两种不同的分解，取其中最小的 n：

$$n = p_1p_2\cdots p_r = q_1q_2\cdots q_s,$$

所有 p_i, q_j 都是素数，两边作为可重集不相同。$n > 1$，所以 $r, s \geq 1$。因为 $p_1 \mid q_1\cdots q_s$，由推论 2.13，$p_1 \mid q_j$ 对某个 j 成立。q_j 是素数，它大于 1 的因数只有它自己，所以 $p_1 = q_j$。两边约去这个素数，得到 n/p_1 的两种分解 $p_2\cdots p_r = q_1\cdots \widehat{q_j}\cdots q_s$（$\widehat{q_j}$ 表示去掉 q_j）。$n/p_1 < n$，由最小性，这两种分解只差次序。再把 $p_1 = q_j$ 放回去，n 的两种分解也只差次序，矛盾。 ■

整个证明中，素数的特殊性只用在一处：p_1 **整除乘积，就整除某个因子**。例 2.7 恰好在这一处失败。

想一想　为什么不把 1 算作素数？

答　若 1 是素数，唯一性立刻失败：$6 = 2\cdot 3 = 1\cdot 2\cdot 3 = 1\cdot 1\cdot 2\cdot 3 = \cdots$。把 1 排除在外，定理才能写得干净。此外 1 在乘法下有逆元（$1\cdot 1 = 1$），它和 -1 一样是「单位」，性质与素数完全不同。

把相同的素数合并，每个 $n > 1$ 都可以唯一地写成**标准分解式**

$$n = p_1^{\alpha_1}p_2^{\alpha_2}\cdots p_k^{\alpha_k}, \qquad p_1 < p_2 < \cdots < p_k, \quad \alpha_i \geq 1.$$

比较两个数时，常把它们写在同一组素数上，允许指数为 0，例如 $12 = 2^2\cdot 3^1\cdot 5^0$，$45 = 2^0\cdot 3^2\cdot 5^1$。

命题 2.14　设 $a = \prod_i p_i^{\alpha_i}$，$b = \prod_i p_i^{\beta_i}$（$\alpha_i, \beta_i \geq 0$）。则

(1) $d \mid a$（$d > 0$）当且仅当 $d = \prod_i p_i^{\delta_i}$，$0 \leq \delta_i \leq \alpha_i$；

(2) $\gcd(a,b) = \prod_i p_i^{\min(\alpha_i,\beta_i)}$，$\operatorname{lcm}(a,b) = \prod_i p_i^{\max(\alpha_i,\beta_i)}$；

(3) $\gcd(a,b)\cdot\operatorname{lcm}(a,b) = ab$。

证明　(1) 若 $a = dm$，把 d 和 m 的分解乘起来就是 a 的一个分解。由唯一性，它就是 a 的标准分解，所以 d 中只出现 p_i，且每个指数不超过 α_i。反过来，指数满足条件时 $a = d\cdot\prod_i p_i^{\alpha_i-\delta_i}$。(2) 由 (1)，公因数的指数满足 $\delta_i \leq \min(\alpha_i,\beta_i)$，取等号得到最大的一个，它也被每个公因数整除；公倍数同理。(3) 由 $\min(\alpha,\beta)+\max(\alpha,\beta) = \alpha+\beta$。 ■

■ **例 2.8**　用素因数分解求 $\gcd(1071, 462)$ 和 $\operatorname{lcm}(1071, 462)$。

解　第一步：分解。$1071 = 3^2\cdot 7\cdot 17$，$462 = 2\cdot 3\cdot 7\cdot 11$。

第二步：列指数表。逐个素数取最小值和最大值。

素数	2	3	7	11	17
1071 的指数	0	2	1	0	1
462 的指数	1	1	1	1	0
min	0	1	1	0	0
max	1	2	1	1	1

第三步：读出结果。$\gcd = 3 \cdot 7 = 21$，与例 2.1 一致；$\operatorname{lcm} = 2 \cdot 3^2 \cdot 7 \cdot 11 \cdot 17 = 23562$。检验：$21 \cdot 23562 = 494802 = 1071 \cdot 462$。 ♦

易错提醒　这个方法只适用于已经分解好的小数。对大数，分解本身比求最大公因数难得多，所以实际计算 gcd 总是用欧几里得算法，再由 $\operatorname{lcm}(a,b) = ab/\gcd(a,b)$ 求最小公倍数。

2.5　素数有无穷多个

算术基本定理说素数是整数的「原子」。原子有多少种？两千多年前欧几里得就给出了答案。

定理 2.15（Euclid）　素数有无穷多个。

证明　设只有有限个素数 $p_1, p_2, \ldots, p_n$。令 $N = p_1 p_2 \cdots p_n + 1 \geq 3$。由第 1 讲，$N$ 有素因数 p，它必是某个 p_i。但由引理 2.1，$\gcd(p_i, N) = \gcd(1, p_i) = 1$，因为 $N = p_i \cdot (\text{其余} p_j \text{ 之积}) + 1$。所以 $p_i \nmid N$，矛盾。 ■

这个证明实际上是一个构造：任给有限个素数，N 的任何一个素因数都是新的素数。从 2 开始，依次取前 n 个素数之积加 1：

n	$p_1 \cdots p_n + 1$	分解	n	$p_1 \cdots p_n + 1$	分解
1	3	素数	5	2311	素数
2	7	素数	6	30031	$59 \cdot 509$
3	31	素数	7	510511	$19 \cdot 97 \cdot 277$
4	211	素数	8	9699691	$347 \cdot 27953$

易错提醒　证明没有说 $N = p_1 \cdots p_n + 1$ 是素数，只说它的素因数是新的。上表第六个 $30031 = 59 \cdot 509$ 就不是素数，但 59 和 509 确实不在 $\{2, 3, 5, 7, 11, 13\}$ 中。

同样的想法可以用来证明某些特殊形式的素数也有无穷多个。

■ **例 2.9**　证明：形如 $4k + 3$ 的素数有无穷多个。

解　第一步：一个观察。除以 4 余 1 的数之积仍除以 4 余 1（例 2.7 已验证两个的情形，多个的情形对个数归纳）。所以**除以 4 余 3 的数，至少有一个素因数除以 4 余 3**：它是奇数，素因数都

是奇数，即除以 4 余 1 或 3，若全余 1，乘积也余 1。

*第二步：构造。*设形如 $4k+3$ 的素数只有 $p_1=3,p_2,\dots,p_n$。令 $N=4p_1p_2\cdots p_n-1$，则 $N=4(p_1\cdots p_n-1)+3$ 除以 4 余 3，且 $N\geq 11$。

*第三步：矛盾。*由第一步，N 有一个形如 $4k+3$ 的素因数，它是某个 p_i。但 $p_i\mid 4p_1\cdots p_n$，于是 $p_i\mid 4p_1\cdots p_n-N=1$，不可能。 ♦

想一想　把例 2.9 中的 N 换成 $4p_1\cdots p_n+1$，能证明形如 $4k+1$ 的素数有无穷多个吗？
答　不能。余 1 的数可以全由 $4k+3$ 型素数组成，如 $21=3\cdot 7$，所以无法断定 N 有 $4k+1$ 型素因数。结论是对的，第 4 讲会用「n^2+1 的奇素因数都是 $4k+1$ 型」证明它。更一般的 Dirichlet 定理说：$\gcd(a,d)=1$ 时，等差数列 $a,a+d,a+2d,\dots$ 中有无穷多个素数。

Euler 在 1737 年给出了一个完全不同的证明。它不构造新素数，而是把算术基本定理改写成一个分析等式。形式上，把每个素数 p 的几何级数相乘：

$$\prod_{p}\Big(1+\frac{1}{p}+\frac{1}{p^2}+\cdots\Big)=\prod_{p}\frac{1}{1-1/p}.$$

展开左边，每一项是 $\frac{1}{p_1^{e_1}\cdots p_k^{e_k}}$，由唯一分解，每个 $\frac{1}{n}$ 恰好出现一次，所以左边「等于」$\sum_n \frac{1}{n}=\infty$。若素数只有有限个，右边是一个有限数，矛盾。下面的有限形式把这个想法变成严格的证明。

定理 2.16（Euler）　对任意正整数 N，

$$\prod_{p\leq N}\frac{1}{1-1/p}\;\geq\;\sum_{n=1}^{N}\frac{1}{n}\;>\;\ln(N+1).$$

因此素数有无穷多个。

证明　*第二个不等式。*$\frac{1}{n}>\int_n^{n+1}\frac{dx}{x}$，对 $n=1,\dots,N$ 求和得 $\sum_{n\leq N}\frac{1}{n}>\int_1^{N+1}\frac{dx}{x}=\ln(N+1)$。

*第一个不等式。*设 $p_1,\dots,p_m$ 是不超过 N 的全部素数，取 K 使 $2^K>N$。每个 $n\leq N$ 的素因数都不超过 N，指数都不超过 K。考虑有限乘积

$$\prod_{i=1}^{m}\Big(1+\frac{1}{p_i}+\cdots+\frac{1}{p_i^{K}}\Big)=\sum_{0\leq e_1,\dots,e_m\leq K}\frac{1}{p_1^{e_1}\cdots p_m^{e_m}}.$$

由存在性，每个 $n\leq N$ 都可以写成 $p_1^{e_1}\cdots p_m^{e_m}$，$e_i\leq K$；由唯一性，不同的 n 对应不同的项。所以右边包含 $\frac{1}{1},\dots,\frac{1}{N}$ 各一次，还有别的正项，右边 $\geq\sum_{n\leq N}\frac{1}{n}$。左边每个因子都小于 $\frac{1}{1-1/p_i}$。

*结论。*若素数只有有限个，左边对所有 N 有一个固定的上界，而 $\ln(N+1)\to\infty$，矛盾。 ■

以 $N=10$ 检验：左边 $=2\cdot\frac{3}{2}\cdot\frac{5}{4}\cdot\frac{7}{6}=4.375$，中间 $=1+\frac{1}{2}+\cdots+\frac{1}{10}\approx 2.929$，右边 $=\ln 11\approx 2.398$。

注 2.1　Euler 的证明能得到更多。对定理 2.16 取对数，用 $-\ln(1-x)\leq x+x^2$（$0\leq x\leq\frac{1}{2}$）和 $\sum_{n\geq 2}\frac{1}{n^2}<1$，得 $\ln\ln(N+1)<\sum_{p\leq N}\frac{1}{p}+1$。所以 $\sum_p\frac{1}{p}$ **发散**，在这个意义下素数比平方数「多」。不过它发散得极慢：$\sum_{p\leq 10^6}\frac{1}{p}\approx 2.887$。

2.6　素数的分布与著名猜想

2.6.1　试除法与筛法

判断 n 是否为素数，最朴素的办法是逐个试除。下面的命题把试除范围从 n 降到 $\sqrt{n}$。

命题 2.17　若 $n>1$ 是合数，则 n 有一个素因数 $p \le \sqrt{n}$。

证明　$n=ab$，$1<a\le b<n$。若 $a>\sqrt{n}$，则 $ab>n$，矛盾，所以 $a\le\sqrt{n}$。由第 1 讲，a 有素因数 p，$p\le a\le\sqrt{n}$，且 $p\mid n$。 ■

例如判断 397：$19^2=361<397<400=20^2$，只需试除 $2,3,5,7,11,13,17,19$。余数依次为 $1,1,2,5,1,7,6,17$，都不为 0，所以 397 是素数。而 391 在 17 处就整除了：$391=17\cdot 23$。

若要找出不超过 N 的全部素数，逐个试除就太浪费了。古希腊的 Eratosthenes 想出了反过来做的办法：不是检验每个数，而是**划掉每个素数的倍数**。

算法 2.18（Eratosthenes 筛法）　列出 $2,3,\dots,N$。取表中最小的未划掉的数 p，它是素数；划掉 $p^2, p^2+p, p^2+2p,\dots$ 中不超过 N 的数。对下一个未划掉的数重复，直到 $p^2>N$。剩下没被划掉的数就是不超过 N 的全部素数。

从 p^2 开始划，是因为比 p^2 小的 p 的倍数 kp（$k<p$）有更小的素因数，已经被划过了；到 $p^2>N$ 停止，正是命题 2.17。对 $N=100$，只需用 $2,3,5,7$ 筛（$11^2>100$），结果见图 2.2。

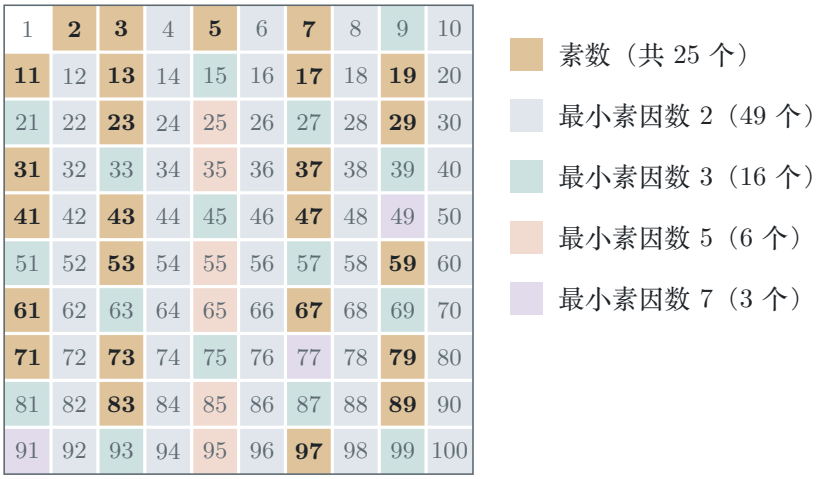


图 2.2　Eratosthenes 筛法，$N=100$。颜色表示每个合数被哪一个素数第一次划掉。只用 $2,3,5,7$ 四个素数就筛完了：7 划掉的第一个数是 $7^2=49$，而 $11^2=121>100$。

2.6.2　素数有多少：$\pi(x)$ 与素数定理

图 2.2 里，1 到 10 中有 4 个素数，91 到 100 中只有 97 一个。素数越往后越稀。稀到什么程度？

定义 2.19（素数计数函数）　对实数 x，记 $\pi(x)$ 为不超过 x 的素数的个数。这里的 π 只是记号，与圆周率无关。

例如 $\pi(10)=4$（$2,3,5,7$），$\pi(100)=25$，$\pi(2.5)=1$，$\pi(1)=0$。Euclid 定理就是 $\pi(x)\to\infty$。$\pi(x)$ 是一个台阶函数：每经过一个素数就上升一格，在两个素数之间保持不变（图 2.3）。

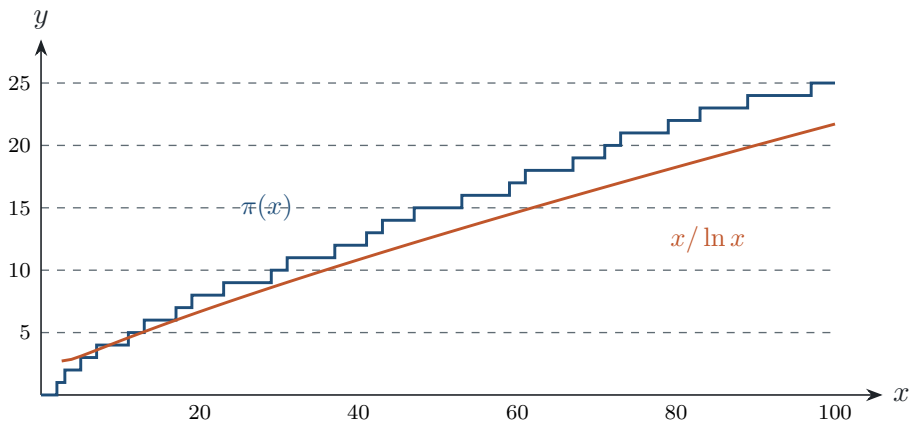


图 2.3　$\pi(x)$（蓝色台阶）与 $x/\ln x$（橙色曲线），$0\le x\le 100$。台阶每遇到一个素数升高一格；两者的变化趋势相近，但不能据图断言曲线始终低于台阶。

图中的橙色曲线 $x/\ln x$ 和台阶贴得相当近。下表把比较推到 10^6，最后一列的 $\mathrm{Li}(x)=\int_2^x \frac{dt}{\ln t}$ 稍后解释。

x	$\pi(x)$	$x/\ln x$	$\pi(x)\ln x/x$	$\mathrm{Li}(x)$
10	4	4.3	0.921	5
10^2	25	21.7	1.151	29
10^3	168	144.8	1.161	177
10^4	1229	1085.7	1.132	1245
10^5	9592	8685.9	1.104	9629
10^6	78498	72382.4	1.084	78627

比值先升后降，从 10^3 起缓慢地向 1 靠拢。这正是下面这个定理所说的。

定理 2.20（素数定理，本课程不证）

$$\lim_{x\to\infty}\frac{\pi(x)}{x/\ln x}=1,\qquad \text{记作}\quad \pi(x)\sim\frac{x}{\ln x}.$$

注意「$\sim$」只说比值趋于 1，并不说差趋于 0：上表中 $\pi(x)-x/\ln x$ 从 23 涨到了 6116。

直观理解　把素数定理读成「密度」：在 x 附近随便取一个整数，它是素数的「概率」约为 $1/\ln x$。例如 $\ln 10^6 \approx 13.8$，所以 10^6 附近大约每 14 个数里有一个素数；实际数一数，$[10^6, 10^6+1000)$ 中有 75 个素数，而 $1000/13.8 \approx 72$。把密度 $1/\ln t$ 从 2 积到 x，就得到表中最后一列 $\mathrm{Li}(x) = \int_2^x \frac{dt}{\ln t}$。它比 $x/\ln x$ 准得多：$x = 10^6$ 时误差只有 129。

这个解释有实际用途。第 6 讲的 RSA 需要随机的 300 位素数。$\ln 10^{300} \approx 690.8$，只在奇数里挑，平均约 345 个里有一个素数；配上快速的素性检验，生成大素数并不难。

历史与背景　Gauss 十五六岁时（约 1792 年）从素数表猜出素数密度约为 $1/\ln x$，Legendre 在 1798 年独立提出类似猜想。Riemann 在 1859 年把问题和 zeta 函数的零点联系起来（第 14 讲会见到 zeta 函数），1896 年 Hadamard 与 de la Vallée Poussin 各自独立完成了证明。

■ **例 2.10**　**改编自习题集 1 第 6 题**　统计 $x = 10000, 20000, \ldots, 100000$ 以内的素数，按除以 4 余 1 和余 3 分成两类，再与 $x/\ln x$ 比较。你观察到什么？

解　*第一步：计算。*用筛法列出 10^5 以内的素数，逐个看除以 4 的余数。结果如下（素数 2 两类都不属于，所以两类之和比 $\pi(x)$ 少 1）。

x	$\pi(x)$	余 1 的	余 3 的	余 3 减余 1	$x/\ln x$
10000	1229	609	619	10	1085.7
20000	2262	1125	1136	11	2019.5
30000	3245	1611	1633	22	2910.1
40000	4203	2085	2117	32	3774.8
50000	5133	2549	2583	34	4621.2
60000	6057	3018	3038	20	5453.5
70000	6935	3449	3485	36	6274.5
80000	7837	3903	3933	30	7086.1
90000	8713	4348	4364	16	7889.5
100000	9592	4783	4808	25	8685.9

第二步：观察。(1) 两类各占约一半，Dirichlet 定理的定量形式说极限下确实如此。(2) 余 3 的一类每行都略多，这叫 **Chebyshev 偏差**（Chebyshev's bias）。但它不是定理：到 $p = 26861$ 时余 1 的一类第一次领先（1473 比 1472），Littlewood 在 1914 年证明领先者会无穷多次易手。(3) $x/\ln x$ 偏小约 10%，比值从 1.120 降到 1.104，收敛很慢。 ♦

2.6.3　素数间隔与几个著名猜想

素数平均间隔约为 $\ln x$，会慢慢变大。但单个间隔可以任意大，构造也很简单。

■ **例 2.11**　对任意正整数 n，找出 n 个连续的合数。

解　取 $(n+1)!+2,\ (n+1)!+3,\ \ldots,\ (n+1)!+(n+1)$，共 n 个连续整数。对 $2 \le j \le n+1$，j 整除 $(n+1)!$，也整除 j，所以 $j \mid (n+1)!+j$；又 $(n+1)!+j > j$，故它是合数。

$n=5$ 时得到 $722,\ldots,726$。这个构造远不是最省的：23 与 29 之间的 $24,\ldots,28$ 已经是 5 个连续合数。 ♦

间隔也不能太大。**Bertrand 公设**（Bertrand's postulate）说：对每个 $n \ge 1$ 都有素数 p 满足 $n < p \le 2n$。Chebyshev 在 1852 年证明了它。下面三个猜想陈述简单，至今没有解决。

猜想 2.21

(1)（Goldbach，1742）每个大于 2 的偶数都是两个素数之和。

(2)（孪生素数）有无穷多对相差 2 的素数 $p,\ p+2$。

(3)（Mersenne 素数）有无穷多个形如 2^n-1 的素数。

Goldbach 猜想。$10=3+7=5+5$，$100=3+97=11+89=\cdots=47+53$，共 6 种。已验证到 4×10^{18}。陈景润在 1973 年证明：每个充分大的偶数是一个素数与一个「至多两个素数之积」的和（俗称「1+2」）。Helfgott 在 2013 年证明了弱形式：每个大于 5 的奇数是三个素数之和。

孪生素数猜想。100 以内有 8 对：$(3,5),(5,7),\ldots,(71,73)$。张益唐在 2013 年证明有无穷多对相邻素数间隔不超过 7×10^7，随后 Maynard 与 Polymath 项目把界降到 246。

Mersenne 素数。先证明一个简单的必要条件。

命题 2.22　若 2^n-1 是素数，则 n 是素数。

证明　设 $n=ab$，$1<a<n$。由 $x^b-1=(x-1)(x^{b-1}+\cdots+x+1)$，取 $x=2^a$，得 $2^a-1 \mid 2^{ab}-1=2^n-1$。而 $1<2^a-1<2^n-1$，所以 2^n-1 是合数。 ■

逆命题不成立：$2^{11}-1=2047=23\cdot89$。Mersenne 数有快速的素性检验，已知最大的素数几乎都是 Mersenne 素数。截至 2024 年 10 月，纪录是 $2^{136279841}-1$，共 41 024 320 位。

易错提醒　不要从几个例子归纳素数的规律。n^2+n+41 对 $n=0,1,\ldots,39$ 都是素数，但 $n=40$ 时等于 $1681=41^2$。Fermat 猜测 $2^{2^k}+1$ 都是素数，Euler 在 1732 年发现 $2^{32}+1=641\cdot6700417$。

本讲小结

- 欧几里得算法：$\gcd(a,b)=\gcd(b \bmod a, a)$，最后一个非零余数就是最大公因数；表格法同时给出 Bézout 系数。$g \mid c$ 时 $ax+by=c$ 的全部解是 $x=x_0+\frac{b}{g}t$，$y=y_0-\frac{a}{g}t$。
- 每两步余数至少减半；Lamé 定理：最坏情形是相邻斐波那契数。
- 算术基本定理：唯一性来自 Euclid 引理；gcd、lcm 取指数的最小、最大值。
- 素数有无穷多个：Euclid 看 $p_1\cdots p_n+1$，Euler 证明 $\sum 1/p$ 发散。
- 试除到 $\sqrt{n}$ 即可判断素数，筛法一次求出全部小素数。素数定理 $\pi(x)\sim x/\ln x$；Goldbach、孪生素数、Mersenne 素数无穷多仍是猜想。

练习

2.1 ★ 用欧几里得算法求 $g=\gcd(2024,1001)$，并求整数 x,y 使 $2024x+1001y=g$。

2.2 ★★ 求方程 $7x+11y=100$ 的全部非负整数解。

2.3 ★★ 设 $a\neq 0$、b,c 为整数，且 $a \mid bc$。证明 $a/\gcd(a,b)$ 整除 c。

2.4 ★★ 设正整数 a,b 互素，ab 是平方数。证明 a,b 都是平方数。不互素呢？

2.5 ★★ 证明：对 $n\geq 1$，欧几里得算法求 $\gcd(F_{n+1},F_{n+2})$ 恰好做 n 次除法；$n\geq 2$ 时前 $n-1$ 个商都是 1，最后一个商是 2。

2.6 ★★★ （习题集 1 第 3 题）设整数 $a\geq 2$，$m,n\geq 1$。证明 $a^m-1 \mid a^n-1 \iff m \mid n$，并证明 $\gcd(a^m-1,a^n-1)=a^{\gcd(m,n)}-1$。用 $\gcd(4095,262143)$ 验证。

2.7 ★★ (a) 设 $a,n\geq 2$。证明：若 a^n-1 是素数，则 $a=2$ 且 n 是素数。(b) 证明：若 2^m+1 是素数（$m\geq 1$），则 m 是 2 的幂。

2.8 ★★★ 记 $T_k=2^{2^k}+1$（Fermat 数）。证明 $T_0T_1\cdots T_{n-1}=T_n-2$，由此证明不同的 Fermat 数两两互素，并再次推出素数有无穷多个。

2.9 ★★★ （习题集 1 第 7 题）甲、乙轮流从 $1,2,\ldots,n$（$n\geq 2$）中取数，甲先。新取的数不能整除已取的任何数（因而不能重复），无数可取者输。证明甲有必胜策略。

练习答案与提示

2.1 $2024=1001\cdot 2+22$，$1001=22\cdot 45+11$，$22=11\cdot 2$，所以 $\gcd=11$。倒推：$11=1001-22\cdot 45=1001-(2024-1001\cdot 2)\cdot 45=1001\cdot 91-2024\cdot 45$，即 $x=-45$，$y=91$。

2.2 由 $1=2\cdot 11-3\cdot 7$ 得特解 $(-300,200)$，全部解为 $x=-300+11t$，$y=200-7t$。$x,y\geq 0$ 要求 $27.3\approx\frac{300}{11}\leq t\leq\frac{200}{7}\approx 28.6$，只有 $t=28$，得 $(8,4)$。

2.3 记 $g=\gcd(a,b)$，$a=ga'$，$b=gb'$，则 $\gcd(a',b')=1$（第 1 讲练习 6）。由 $a \mid bc$ 得 $ga'k=gb'c$，即 $a' \mid b'c$。a' 与 b' 互素，由 Euclid 引理的一般形式，$a' \mid c$。

2.4 设 $ab=n^2$。若素数 p 在 a 中的指数为 $\alpha\geq 1$，则 $p\nmid b$（互素），所以 p 在 n^2 中的指数就是 α，故 α 是偶数。于是 a 是完全平方数，b 同理。去掉互素不成立：$2\cdot 8=16$。

2.5 $2\leq k\leq n$ 时 $F_{k+2}=F_{k+1}\cdot 1+F_k$，$0<F_k<F_{k+1}$，这是商为 1 的带余除法。于是 $(F_{n+1},F_{n+2})\to\cdots\to(F_2,F_3)=(1,2)$ 共 $n-1$ 步，再加最后一步 $2=1\cdot 2+0$（商为 2），共 n 步。

2.6 设 $n = mq + r$，$0 \le r < m$。$a^n - 1 = a^r(a^{mq} - 1) + (a^r - 1)$，且 $a^m - 1 \mid a^{mq} - 1$，故 $\gcd(a^m - 1, a^n - 1) = \gcd(a^m - 1, a^r - 1)$。指数 $(m, n) \to (r, m)$ 正是欧几里得算法的一步，对 m 归纳得 $\gcd = a^{\gcd(m,n)} - 1$，从而 $a^m - 1 \mid a^n - 1 \iff m \mid n$。验证：$\gcd(2^{12} - 1, 2^{18} - 1) = 2^6 - 1 = 63$。

2.7 (a) $a - 1 \mid a^n - 1$ 且 $a - 1 < a^n - 1$，故 $a - 1 = 1$；再由命题 2.22，n 是素数。(b) 若 $m = cb$，$b > 1$ 为奇数，由 $x^b + 1 = (x + 1)(x^{b-1} - \cdots + 1)$ 取 $x = 2^c$，得 $2^c + 1$ 是 $2^m + 1$ 的真因数，矛盾。

2.8 $T_n - 2 = 2^{2^n} - 1 = (2^{2^{n-1}} - 1)(2^{2^{n-1}} + 1) = (T_{n-1} - 2)\,T_{n-1}$，对 n 归纳即得 $T_0 \cdots T_{n-1} = T_n - 2$。$m < n$ 时 $T_m \mid T_n - 2$，故 $d = \gcd(T_m, T_n)$ 整除 2；T_n 是奇数，所以 $d = 1$。各 T_n 的素因数两两不同，素数有无穷多个（这个证明属于 Goldbach）。

2.9 游戏在有限步内结束，没有平局，所以恰有一方有必胜策略。假设是乙。甲先取 1，乙回应 $k > 1$。此后「已取 $1, k$」与「只取 k」的合法后续步相同，故甲可以先取 k，再照搬乙的策略，矛盾；甲有必胜策略。

第 3 讲

二项式系数与同余

Binomial Coefficients, Congruences

对应材料：MIT 18.781 Lecture 3 讲义；习题集 2 第 1、2、3、7 题（部分改编为例题与练习），并借用第 4 题取 $a=-1$ 的二项式判别特例。

本讲目标

(1) 用计数解释 $n!$ 和二项式系数 $\binom{n}{k}$，并由此证明「任意 k 个连续整数之积被 $k!$ 整除」。

(2) 掌握二项式定理、Pascal 递推关系，知道 $\binom{\alpha}{k}$ 对任意复数 α 都有定义，以及广义二项式定理在 $|z|<1$ 时成立。

(3) 会用 Legendre 公式求 $n!$ 中素数 p 的方幂，并会用 p 进制数字和改写它。

(4) 理解 Kummer 定理：$\binom{n}{k}$ 中 p 的方幂等于 p 进制加法 $k+(n-k)$ 的进位次数；会用 Lucas 定理求 $\binom{n}{k}$ 除以 p 的余数。

(5) 掌握同余的定义和它与加、减、乘的相容性；知道指数不能随便换成同余的数，也知道消去律只在 $\gcd(c,m)=1$ 时成立。

3.1 从计数到二项式系数

先看第 1 讲证过的一个结论：对任意整数 n，$6\mid n(n+1)(n+2)$。那里按余数分类，分三步才完成。换一个角度：当 $n\ge 1$ 时，

$$\frac{n(n+1)(n+2)}{6}=\frac{(n+2)(n+1)n}{3!}$$

恰好是「从 $n+2$ 个人里选 3 个人」的方法数。方法数当然是整数，所以 6 整除分子。这一节把这个想法推广：**一个量如果在数某种东西，它就一定是整数**。

3.1.1 排列与阶乘

把 n 个不同的物体排成一排，第一个位置有 n 种选法，第二个位置有 $n-1$ 种，依此类推，最后一个位置只剩 1 种。排法总数是

$$n!=n(n-1)(n-2)\cdots 2\cdot 1,$$

读作 n 的阶乘（factorial）。约定 $0!=1$：把零个物体排成一排只有一种排法，就是什么都不放。这个约定让下面的公式在边界情形也成立。

如果只从 n 个物体中取出 k 个排成一排，同样逐个位置计数，得到

$$n(n-1)\cdots(n-k+1)=\frac{n!}{(n-k)!}\qquad(0\le k\le n),$$

这是一个 k 项连乘积。如果取出的 k 个物体不计顺序，每一组都被重复数了 $k!$ 次，因为同一组 k 个物体有 $k!$ 种排法。

■ **例 3.1**　从 A、B、C、D、E 五人中选出主席、副主席、秘书各一人，有多少种选法？如果只选一个三人委员会、不分职务，又有多少种？

解　有职务时，主席 5 种选法，副主席 4 种，秘书 3 种，共 $5\cdot4\cdot3=60$ 种。

不分职务时，同一个三人组（比如 $\{A,B,C\}$）在上面的 60 种里出现了 $3!=6$ 次：ABC、ACB、BAC、BCA、CAB、CBA。所以委员会个数是 $60/6=10$。逐一列出来核对：

$$ABC,\ ABD,\ ABE,\ ACD,\ ACE,\ ADE,\ BCD,\ BCE,\ BDE,\ CDE,$$

恰好 10 个。 ♦

定义 3.1（二项式系数）　设整数 $0\le k\le n$。从 n 元集合中选出 k 元子集的方法数记为

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

称为**二项式系数**（binomial coefficient），读作「n 选 k」。当 $k>n$ 时约定 $\binom{n}{k}=0$（选不出来）。

由计数含义立即得到对称性 $\binom{n}{k}=\binom{n}{n-k}$：选出要的 k 个，等于选出不要的 $n-k$ 个。

3.1.2 连续整数之积

命题 3.2　对任意整数 n 和正整数 k，有 $k!\mid n(n-1)\cdots(n-k+1)$。换句话说，**任意 k 个连续整数的乘积被 $k!$ 整除**。

思路　当 $n\ge k$ 时，这个乘积等于 $k!\binom{n}{k}$，而 $\binom{n}{k}$ 在计数，是整数。其余情形设法化归到这一种。

证明　分三种情形。

(1) $0<k\le n$。乘积等于 $\dfrac{n!}{(n-k)!}=k!\binom{n}{k}$，被 $k!$ 整除。

(2) $0\le n<k$。乘积 $n(n-1)\cdots(n-k+1)$ 从 n 一直乘到 $n-k+1\le 0$，中间一定经过因子 0，乘积为 0，被任何数整除。

(3) $n<0$。记 $m=-n\ge 1$。每个因子提出一个负号：

$$n(n-1)\cdots(n-k+1)=(-1)^k\,m(m+1)\cdots(m+k-1).$$

右边的 $m(m+1)\cdots(m+k-1)$ 是从 $m+k-1$ 往下数的 k 个正整数之积，由情形 (1)，它等于 $k!\binom{m+k-1}{k}$，被 $k!$ 整除。 ■

■ **例 3.2**　(a) 验证 $4!\mid 7\cdot8\cdot9\cdot10$ 和 $3!\mid(-5)(-4)(-3)$。(b) 证明对一切整数 n，$120\mid n^5-5n^3+4n$。

解　(a) $7\cdot8\cdot9\cdot10=5040=24\cdot210$，商 $210=\binom{10}{4}$。$(-5)(-4)(-3)=-60=6\cdot(-10)$，商 $-10=-\binom{5}{3}$，与情形 (3) 一致。

(b) 先因式分解：

$$n^5-5n^3+4n=n(n^4-5n^2+4)=n(n^2-1)(n^2-4)=(n-2)(n-1)\,n\,(n+1)(n+2).$$

这是 5 个连续整数之积，由命题 3.2，被 $5!=120$ 整除。代入检验：$n=3$ 时值为 $243-135+12=120$；$n=4$ 时值为 $1024-320+16=720=6\cdot120$。 ♦

核心思想　如果要证 $120\mid N$，分别验证 $8\mid N$、$3\mid N$、$5\mid N$ 要讨论很多余数。把 N 写成「$k!$ 乘一个计数」，一步就得到结论。这是组合计数给数论提供的一类证明方法。

3.1.3 推广到任意上标

公式 $\frac{n(n-1)\cdots(n-k+1)}{k!}$ 的分子只是 k 个数相乘，上标 n 不必是自然数。

定义 3.3（广义二项式系数）　对任意复数 α 和自然数 k，定义

$$\binom{\alpha}{k}=\frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!},\qquad \binom{\alpha}{0}=1.$$

当 $\alpha=n\in\mathbb{N}$ 时，它与前面的定义一致：$k\le n$ 时给出 $\binom{n}{k}$；$k>n$ 时分子含因子 $n-n=0$，值为 0，与「选不出来」的约定吻合。固定 k，$\binom{x}{k}$ 是 x 的 k 次多项式。

■ **例 3.3**　计算 $\binom{1/2}{2}$，$\binom{1/2}{3}$，$\binom{-1}{k}$，$\binom{-2}{k}$ 和 $\binom{3}{5}$。

解　按定义逐项代入：

$$\binom{1/2}{2}=\frac{\frac12\cdot\left(-\frac12\right)}{2}=-\frac18,$$
$$\binom{1/2}{3}=\frac{\frac12\cdot\left(-\frac12\right)\cdot\left(-\frac32\right)}{6}=\frac{3/8}{6}=\frac{1}{16}.$$

上标为 -1 时，分子 $(-1)(-2)\cdots(-k)=(-1)^k k!$，所以 $\binom{-1}{k}=(-1)^k$。上标为 -2 时，分子 $(-2)(-3)\cdots(-k-1)=(-1)^k(k+1)!$，所以 $\binom{-2}{k}=(-1)^k(k+1)$。最后，$\binom{3}{5}=\frac{3\cdot2\cdot1\cdot0\cdot(-1)}{120}=0$。

一般地，命题 3.2 证明中情形 (3) 的计算就是

$$\boxed{\binom{-m}{k}=(-1)^k\binom{m+k-1}{k}\qquad(m\ge1).}$$

例如 $\binom{-3}{2}=\frac{(-3)(-4)}{2}=6=\binom{4}{2}$。 ♦

想一想　$\binom{x}{2}=\frac{1}{2}x^2-\frac{1}{2}x$ 的系数不是整数。它在整数 x 处的值一定是整数吗？$\binom{x}{3}$ 呢？

答　一定是。$\binom{x}{k}$ 在整数 x 处的值等于 k 个连续整数之积除以 $k!$，由命题 3.2 是整数。例如 $\binom{-4}{3}=\frac{(-4)(-5)(-6)}{6}=-20$。系数不是整数、但在所有整数处取整数值的多项式叫**整值多项式**（integer-valued polynomial），练习 3.10 会证明它们恰好是 $\binom{x}{k}$ 的整系数组合。

3.2　二项式定理与 Pascal 三角

3.2.1　二项式定理

定理 3.4（二项式定理）　对自然数 n 和任意 x,y（实数、复数，或任何满足交换律的数），

$$(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}.$$

思路　不去归纳，而是直接数：把 n 个括号乘开，看 $x^k y^{n-k}$ 出现了多少次。

证明　$(x+y)^n=(x+y)(x+y)\cdots(x+y)$ 共 n 个括号。用分配律全部乘开，每一项都是从每个括号里各挑一个字母（x 或 y）再相乘，共 2^n 项。一项等于 $x^k y^{n-k}$，当且仅当恰好有 k 个括号挑了 x。挑 x 的那 k 个括号可以是 n 个括号中的任意 k 个，共 $\binom{n}{k}$ 种。合并同类项，$x^k y^{n-k}$ 的系数就是 $\binom{n}{k}$。 ■

■ **例 3.4**　(a) 按上面的证明展开 $(x+y)^3$。(b) 证明行和公式 $\sum_{k=0}^{n}\binom{n}{k}=2^n$；再证明当 $n\geq 1$ 时交错和 $\sum_{k=0}^{n}(-1)^k\binom{n}{k}$ 等于 0。用 $n=4$ 检验两个结论。

解　(a) 从三个括号中各挑一个字母，得到 8 个「单词」：

$$yyy;\quad xyy,\ yxy,\ yyx;\quad xxy,\ xyx,\ yxx;\quad xxx.$$

按 x 的个数分组，组的大小是 $1,3,3,1=\binom{3}{0},\binom{3}{1},\binom{3}{2},\binom{3}{3}$，所以 $(x+y)^3=y^3+3xy^2+3x^2y+x^3$。

(b) 在定理中令 $x=y=1$，得 $2^n=\sum_k\binom{n}{k}$。它的计数含义是：n 元集合的子集总数是 2^n（每个元素选或不选），按子集大小分类就是右边。令 $x=-1$，$y=1$，得 $0=(1-1)^n=\sum_k(-1)^k\binom{n}{k}$ $(n\geq 1)$。$n=4$ 时，

$$1+4+6+4+1=16=2^4,\qquad 1-4+6-4+1=0.$$

♦

3.2.2　Pascal 三角

把二项式系数按行排列，第 n 行是 $\binom{n}{0},\binom{n}{1},\ldots,\binom{n}{n}$，就得到 **Pascal 三角**（Pascal's triangle），中国称为**杨辉三角**（图 3.1）。表中每个数等于它左上方和右上方两数之和。

命题 3.5（Pascal 递推）　对自然数 n 和 $1 \le k \le n$，

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}.$$

证明　数一数从 $n+1$ 个人中选 k 人的方法。固定其中一人 X。选法分成两类：含 X 的，需要从其余 n 人中再选 $k-1$ 人，有 $\binom{n}{k-1}$ 种；不含 X 的，需要从其余 n 人中选 k 人，有 $\binom{n}{k}$ 种。两类不重不漏，相加即得。 ■

两边都是 n 的多项式（固定 k），在无穷多个自然数 n 处相等，所以它们作为多项式相等：$\binom{\alpha+1}{k} = \binom{\alpha}{k-1} + \binom{\alpha}{k}$ 对任意复数 α 成立。

								行和
$n=0$				1				1
$n=1$			1		1			2
$n=2$		1		2		1		4
$n=3$		1		3		3	1	8
$n=4$	1		4		6		4	1	16
$n=5$	1		5		10		10	5	1	32
$n=6$	1		6		15		20	15	6	1	64

图 3.1　Pascal 三角的第 0 到 6 行。橙色标出 $\binom{4}{1} + \binom{4}{2} = \binom{5}{2}$，即 $4+6=10$；右列是行和 2^n。

■ **例 3.5　Vandermonde 恒等式**　证明 $\displaystyle\sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i} = \binom{m+n}{k}$，并由此求 $\displaystyle\sum_{i=0}^{n} \binom{n}{i}^2$。

解　**计数证明**：一个班有 m 名男生、n 名女生，从中选 k 人。按选中的男生人数 i 分类：选 i 名男生有 $\binom{m}{i}$ 种，再选 $k-i$ 名女生有 $\binom{n}{k-i}$ 种。对 i 求和就是全部选法 $\binom{m+n}{k}$。

比较系数的证明：$(1+x)^{m+n} = (1+x)^m (1+x)^n$。左边 x^k 的系数是 $\binom{m+n}{k}$。右边两个多项式相乘，x^k 来自第一个因子的 x^i 乘第二个因子的 x^{k-i}，系数是 $\sum_i \binom{m}{i}\binom{n}{k-i}$。

取 $m = n = k$，并用对称性 $\binom{n}{n-i} = \binom{n}{i}$：

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

$n=4$ 时，$1+16+36+16+1 = 70 = \binom{8}{4}$。 ♦

3.2.3　广义二项式定理

指数不是自然数时，$(1+z)^\alpha$ 展开成无穷级数，系数正是广义二项式系数。

定理 3.6（广义二项式定理）　设 $\alpha \in \mathbb{C}$，复数 z 满足 $|z|<1$。则

$$(1+z)^{\alpha}=\sum_{k=0}^{\infty}\binom{\alpha}{k}z^{k}.$$

这里 $(1+z)^{\alpha}$ 取 $\exp(\alpha\log(1+z))$ 的主值；当 $z\in(-1,1)$、α 为实数时，它就是通常的正实数的幂。课程不证明这个定理，我们只说明系数从哪里来。设 $f(z)=(1+z)^{\alpha}$，每求一次导，指数降 1，原来的指数乘到前面：

$$f^{(k)}(z)=\alpha(\alpha-1)\cdots(\alpha-k+1)\,(1+z)^{\alpha-k}.$$

所以 Taylor 系数 $f^{(k)}(0)/k!$ 等于 $\binom{\alpha}{k}$。当 $\alpha\in\mathbb{N}$ 时，$k>\alpha$ 的系数全为 0，级数退化为定理 3.4（取 $x=z$，$y=1$）。当 $\alpha\notin\mathbb{N}$ 时，系数都不为零，相邻两项之比

$$\left|\frac{\binom{\alpha}{k+1}z^{k+1}}{\binom{\alpha}{k}z^{k}}\right|=\left|\frac{\alpha-k}{k+1}\right|\cdot|z|\ \longrightarrow\ |z|\qquad(k\to\infty).$$

由比值判别法，级数在 $|z|<1$ 时收敛，在 $|z|>1$ 时发散。级数的和确实等于 $(1+z)^{\alpha}$，这需要估计 Taylor 余项，这里承认。

■ **例 3.6**　用广义二项式定理展开 $\dfrac{1}{1+z}$，$\dfrac{1}{(1-z)^2}$ 和 $\sqrt{1+z}$，并用最后一个展开式近似计算 $\sqrt{1.1}$。

解　取 $\alpha=-1$，由 $\binom{-1}{k}=(-1)^k$ 得等比级数

$$\frac{1}{1+z}=1-z+z^2-z^3+\cdots.$$

取 $\alpha=-2$，把 z 换成 $-z$。由 $\binom{-2}{k}=(-1)^k(k+1)$，第 k 项是 $(-1)^k(k+1)(-z)^k=(k+1)z^k$：

$$\frac{1}{(1-z)^2}=1+2z+3z^2+4z^3+\cdots.$$

对等比级数 $\frac{1}{1-z}=\sum z^k$ 逐项求导也得到同样的结果。取 $\alpha=\frac12$，系数依次是 $1,\frac12,-\frac18,\frac{1}{16},-\frac{5}{128}$（其中 $\binom{1/2}{4}=\frac{\frac12\cdot(-\frac12)(-\frac32)(-\frac52)}{24}=-\frac{5}{128}$）：

$$\sqrt{1+z}=1+\frac12 z-\frac18 z^2+\frac{1}{16}z^3-\frac{5}{128}z^4+\cdots.$$

令 $z=0.1$，部分和见表 3.1，真值 $\sqrt{1.1}=1.0488088\cdots$。 ♦

截止到	z^0	z^1	z^2	z^3	z^4
部分和	1	1.05	1.04875	1.0488125	1.0488086
误差	-4.9×10^{-2}	$+1.2\times10^{-3}$	-5.9×10^{-5}	$+3.7\times10^{-6}$	-2.5×10^{-7}

表 3.1　用 $\sqrt{1+z}$ 的级数在 $z=0.1$ 处计算 $\sqrt{1.1}$（误差 = 部分和 $-\sqrt{1.1}$）。误差正负交替，随截断阶数提高而减小；$k\ge1$ 时相邻项绝对值之比为 $|\frac{1/2-k}{k+1}|\cdot0.1<0.1$。

想一想　在 $\frac{1}{1+z}=1-z+z^2-\cdots$ 中代入 $z=3$，左边是 $\frac14$，右边是 $1-3+9-27+\cdots$。哪里出了问题？

答　右边的部分和 $1,-2,7,-20,61,\ldots$ 绝对值越来越大，级数发散。定理只在 $|z|<1$ 时成立；左边在 $z=3$ 处有意义，不能保证右边收敛。

前后联系　取 $\alpha=-n$，把 z 换成 $-z$，再用 $\binom{-n}{k}=(-1)^k\binom{n+k-1}{k}$，得到

$$\frac{1}{(1-z)^n}=\sum_{k=0}^{\infty}\binom{n+k-1}{k}z^k.$$

系数 $\binom{n+k-1}{k}$ 是把 k 个相同的球放进 n 个盒子的方法数。讲生成函数时会把这类展开当作形式幂级数，不再关心收敛。

3.3　$n!$ 中素数的方幂：Legendre 公式

$10!=3628800$ 末尾有两个零。原因可以从分解式看出：$10!=2^8\cdot3^4\cdot5^2\cdot7$，每个末尾零需要一对 2×5，而 5 只有两个。对 100! 这样有 158 位的数，直接分解行不通。这一节给出一个公式，不用算出 $n!$ 就能读出其中每个素数的方幂。

定义 3.7（恰好整除）　设 p 为素数，$N\neq0$ 为整数。若 $p^e\mid N$ 但 $p^{e+1}\nmid N$，就说 p^e **恰好整除**（exactly divides）N，记作 $p^e\parallel N$，并记 $v_p(N)=e$，称为 N 的 **p 进赋值**（p-adic valuation）。

例如 $40=2^3\cdot5$，所以 $2^3\parallel40$，$v_2(40)=3$，$v_5(40)=1$，$v_3(40)=0$。由第 2 讲的算术基本定理，$v_p(N)$ 就是 N 的素因数分解中 p 的指数，因此

$$v_p(ab)=v_p(a)+v_p(b),\qquad v_p(n!)=\sum_{m=1}^{n}v_p(m).$$

引理 3.8　设 d,n 为正整数。$1,2,\ldots,n$ 中 d 的倍数恰有 $\lfloor n/d\rfloor$ 个。

证明　d 的倍数 $d,2d,\ldots,jd$ 不超过 n，当且仅当 $jd\le n$，即 $j\le n/d$。满足条件的正整数 j 共 $\lfloor n/d\rfloor$ 个。 ■

定理 3.9（Legendre 公式）　设 p 为素数，n 为正整数。则

$$v_p(n!)=\lfloor\frac{n}{p}\rfloor+\lfloor\frac{n}{p^2}\rfloor+\lfloor\frac{n}{p^3}\rfloor+\cdots=\sum_{i\ge1}\lfloor\frac{n}{p^i}\rfloor.$$

当 $p^i>n$ 时 $\lfloor n/p^i\rfloor=0$，所以右边只有有限项非零。

思路　在每个 $m\le n$ 上方叠放 $v_p(m)$ 个方块。$v_p(n!)$ 是方块总数。按列数是 $\sum_m v_p(m)$；按行数，

第 i 行的方块恰好落在 p^i 的倍数上方。两种数法结果相同（图 3.2）。

证明　注意 $v_p(m)\ge i$ 当且仅当 $p^i\mid m$。所以 $v_p(m)=\sum_{i\ge 1}[p^i\mid m]$，其中 $[P]$ 在命题 P 成立时取 1，否则取 0。交换求和次序，再用引理 3.8：

$$v_p(n!)=\sum_{m=1}^{n}\sum_{i\ge 1}[p^i\mid m]=\sum_{i\ge 1}\#\{\,m\le n\mid p^i\mid m\,\}=\sum_{i\ge 1}\Big\lfloor\frac{n}{p^i}\Big\rfloor.$$ ■

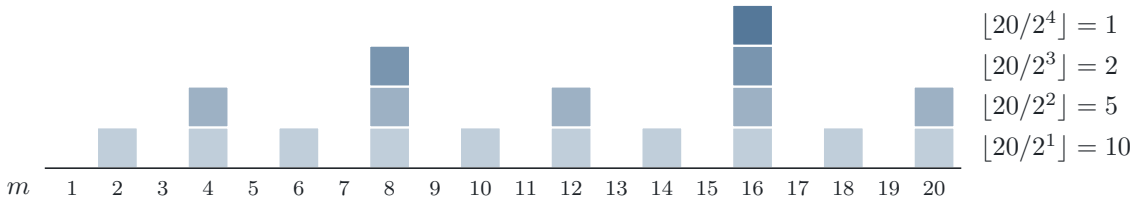


图 3.2　$m=1,\dots,20$ 上方各叠 $v_2(m)$ 个方块。按列数得 $v_2(20!)$；按行数得 $10+5+2+1=18$。

■ **例 3.7**　**100! 的末尾零与 12 的方幂**　(a) 100! 末尾有多少个零？(b) 使 $12^k\mid 100!$ 的最大整数 k 是多少？

解　(a) 末尾零的个数是使 $10^z\mid 100!$ 的最大 z。因为 $10^z=2^z5^z$，所以 $z=\min\{v_2(100!),\,v_5(100!)\}$。由 Legendre 公式，

$$v_5(100!)=\Big\lfloor\tfrac{100}{5}\Big\rfloor+\Big\lfloor\tfrac{100}{25}\Big\rfloor=20+4=24,\qquad v_2(100!)=50+25+12+6+3+1=97.$$

5 远比 2 少，所以 100! 末尾恰有 24 个零。

(b) $12^k=2^{2k}3^k$。需要 $2k\le v_2(100!)=97$ 且 $k\le v_3(100!)$。又

$$v_3(100!)=\Big\lfloor\tfrac{100}{3}\Big\rfloor+\Big\lfloor\tfrac{100}{9}\Big\rfloor+\Big\lfloor\tfrac{100}{27}\Big\rfloor+\Big\lfloor\tfrac{100}{81}\Big\rfloor=33+11+3+1=48.$$

两个条件分别给出 $k\le 48$ 和 $k\le 48$，所以最大的 $k=48$。这次两个素数恰好同时卡住。 ♦

3.3.1　用 p 进制数字和改写

去掉 Legendre 公式中的取整，再把等比级数加到无穷，得到一个上界：

$$v_p(n!)<\frac{n}{p}+\frac{n}{p^2}+\frac{n}{p^3}+\cdots=\frac{n}{p-1}.$$

不等号是严格的，因为 $p^i>n$ 以后左边的项都是 0，右边的项仍为正。例如 $v_2(100!)=97<100$。差额 $100-97=3$ 从哪来？答案藏在 n 的 p 进制表示里。

把 n 写成 p 进制：$n=(a_ra_{r-1}\cdots a_1a_0)_p=a_0+a_1p+\cdots+a_rp^r$，$0\le a_i\le p-1$。记**数字和** (digit sum) $s_p(n)=a_0+a_1+\cdots+a_r$。例如 $100=(1100100)_2=(10201)_3=(400)_5=(202)_7$，所以 $s_2(100)=3$，$s_3(100)=4$，$s_5(100)=4$，$s_7(100)=4$。

命题 3.10　设 p 为素数，$n \geq 1$。则

$$\boxed{v_p(n!) = \frac{n - s_p(n)}{p-1}.}$$

证明　设 $n = a_0 + a_1 p + \cdots + a_r p^r$。除以 p^i 时，低 i 位 $a_0 + \cdots + a_{i-1}p^{i-1} \leq p^i - 1$ 落在小数部分，所以

$$\lfloor \frac{n}{p^i} \rfloor = a_i + a_{i+1} p + \cdots + a_r p^{r-i} \qquad (0 \leq i \leq r).$$

对 $i = 1, \ldots, r$ 求和，并按 a_j 合并：a_j 出现在 $i = 1, \ldots, j$ 这 j 项中，系数为 $p^{j-1} + p^{j-2} + \cdots + 1 = \dfrac{p^j - 1}{p-1}$。于是

$$v_p(n!) = \sum_{j=0}^{r} a_j \cdot \frac{p^j - 1}{p-1} = \frac{\sum_j a_j p^j - \sum_j a_j}{p-1} = \frac{n - s_p(n)}{p-1}.$$

■

核对 $n = 100$：$p = 2, 3, 5, 7$ 分别给出 $\frac{100-3}{1} = 97$，$\frac{100-4}{2} = 48$，$\frac{100-4}{4} = 24$，$\frac{100-4}{6} = 16$，与 Legendre 公式一致。上界 $\frac{n}{p-1}$ 与真值恰好差 $\frac{s_p(n)}{p-1}$。

想一想　是否存在正整数 n 使 $2^n \mid n!$？使 $2^{n-1} \mid n!$ 的 n 又是哪些？

答　由命题 3.10，$v_2(n!) = n - s_2(n)$。$n \geq 1$ 的二进制表示里至少有一个 1，所以 $s_2(n) \geq 1$，$v_2(n!) \leq n - 1$，$2^n \nmid n!$ 永远成立。$2^{n-1} \mid n!$ 当且仅当 $s_2(n) = 1$，即二进制只有一个 1，也就是 n 是 2 的方幂。例如 $v_2(8!) = 7$，$v_2(6!) = 4 = 6 - 2$。

3.4　Kummer 定理：二项式系数与进位

由 $\binom{n}{k} = \dfrac{n!}{k!\,(n-k)!}$ 和 v_p 把乘法变成加法，

$$v_p\binom{n}{k} = v_p(n!) - v_p(k!) - v_p((n-k)!).$$

代入命题 3.10，三个分子中的 n、k、$n-k$ 相互抵消，只剩数字和：

$$\boxed{v_p\binom{n}{k} = \frac{s_p(k) + s_p(n-k) - s_p(n)}{p-1}.}$$

先看一个例子。$\binom{10}{4} = 210 = 2 \cdot 3 \cdot 5 \cdot 7$。二进制加法 $100 + 110 = 1010$ 只在第 2 位进位一次；三进制 $(11)_3 + (20)_3 = (101)_3$ 只在第 1 位进位一次。v_2 和 v_3 都是 1，与进位次数相同。这不是巧合。

引理 3.11　在 p 进制下计算 $a + b$，设共发生 c 次进位。则

$$s_p(a) + s_p(b) - s_p(a+b) = (p-1)\,c.$$

证明　设 $a, b, a+b$ 的第 i 位数字分别是 a_i, b_i, r_i，第 i 位向第 $i+1$ 位的进位是 $c_i \in \{0,1\}$，并令 $c_{-1}=0$。竖式加法的第 i 列说的是

$$a_i + b_i + c_{i-1} = r_i + p\,c_i.$$

（进位最多是 1，因为 $a_i+b_i+c_{i-1} \le 2(p-1)+1 < 2p$。）对所有 i 求和。最高位不再进位，所以 $\sum_i c_{i-1} = \sum_i c_i = c$。于是 $s_p(a)+s_p(b)+c = s_p(a+b)+pc$，移项即得。 ■

直观地说：每进位一次，本位数字少了 p，高一位多了 1，数字和净减少 $p-1$。

定理 3.12（Kummer 定理）　设 p 为素数，$0 \le k \le n$。则 $v_p\binom{n}{k}$ 等于在 p 进制下计算 $k+(n-k)$ 时的进位次数。

证明　在引理 3.11 中取 $a=k$，$b=n-k$，再与上面的方框公式比较。 ■

	1	1	1	1	
		1	2	1	2
+		1	2	1	2
	1	0	2	0	1

三进制：4 次进位

	1	1			1		
		1	1	0	0	1	0
+		1	1	0	0	1	0
	1	1	0	0	1	0	0

二进制：3 次进位

图 3.3　用竖式计算 $50+50=100$。橙色小字是进位，阴影标出接收进位的列。由 Kummer 定理，$3^4 \parallel \binom{100}{50}$，$2^3 \parallel \binom{100}{50}$。

■ **例 3.8　$\binom{100}{50}$ 的素因数分解**　不展开这个 30 位数，求出 $\binom{100}{50}$ 的素因数分解。

解　$\binom{100}{50}$ 的素因数都不超过 100，因为分子 100! 的素因数都不超过 100。对每个 $p \le 100$ 数 $50+50$ 在 p 进制下的进位次数。

小素数。$50 = (110010)_2 = (1212)_3 = (200)_5 = (101)_7$。图 3.3 显示二进制进位 3 次、三进制进位 4 次。五进制 $200+200=400$、七进制 $101+101=202$，各位相加都不到 p，不进位。所以 $v_2=3$，$v_3=4$，$v_5=v_7=0$。

$p \ge 11$。写 $50 = qp+r$，$0 \le r < p$，则 $q \le 4$。个位相加得 $2r$，进位当且仅当 $2r \ge p$。十位最多是 $4+4+1=9<p$，不会再进位。所以 $v_p\binom{100}{50}=1$ 当且仅当 $2r \ge p$，否则为 0。逐个检查：$p = 11, 13, 17, 19, 29, 31$ 时 $r = 6, 11, 16, 12, 21, 19$，都有 $2r \ge p$；$p = 23, 37, 41, 43, 47$ 时 $r = 4, 13, 9, 7, 3$，$2r < p$；$53 \le p \le 97$ 时 $r=50$，$2r \ge p$。于是

$$\binom{100}{50} = 2^3 \cdot 3^4 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 29 \cdot 31 \cdot 53 \cdot 59 \cdot 61 \cdot 67 \cdot 71 \cdot 73 \cdot 79 \cdot 83 \cdot 89 \cdot 97.$$

把右边乘出来确实得到 100891344545564193334812497256。 ♦

Kummer 定理把「p 是否整除 $\binom{n}{k}$」变成「p 进制加法是否进位」，后者一眼就能看出来。

推论 3.13　设 p 为素数，$0<k<p^e$。则 $v_p\binom{p^e}{k} = e - v_p(k) \ge 1$。特别地，$0<k<p$ 时 $p \mid \binom{p}{k}$。

证明　记 $j = v_p(k) < e$，则 k 的最低 j 位是 0，第 j 位非零。$p^e = (10\cdots0)_p$ 的低 e 位全是 0。计算 $k + (p^e - k)$ 时，最低 j 位两数都是 0，不进位；第 j 位两数之和非零而结果是 0，只能是恰好等于 p，进位；此后第 $j+1, \dots, e-1$ 位都带着进位 1，结果仍是 0，所以每位的和都等于 p，各进位一次。共进位 $e - j$ 次。 ■

例如 $\binom{8}{4} = 70$，$v_2(70) = 1 = 3 - v_2(4)$；$\binom{9}{3} = 84$，$v_3(84) = 1 = 2 - 1$；$\binom{9}{1} = 9$，$v_3(9) = 2$。

■ **例 3.9　素数的一个刻画**　证明：$n \geq 2$ 整除 $\binom{n}{1}, \binom{n}{2}, \dots, \binom{n}{n-1}$ 全体，当且仅当 n 是素数。

解　若 $n = p$ 是素数，由推论 3.13 即得。

反之设 n 是合数，取 n 的一个素因数 p，则 $1 < p < n$。我们证明 $n \nmid \binom{n}{p}$。由

$$\binom{n}{p} = \frac{n}{p} \cdot \frac{(n-1)(n-2)\cdots(n-p+1)}{(p-1)!} = \frac{n}{p}\binom{n-1}{p-1},$$

若 $n \mid \binom{n}{p}$，则 $\frac{n}{p}\binom{n-1}{p-1} = nt$，两边约去 $\frac{n}{p}$ 得 $p \mid \binom{n-1}{p-1}$。但 $\binom{n-1}{p-1}$ 的分子 $(n-1)\cdots(n-p+1)$ 中各项与 n 相差 $1, \dots, p-1$，而 $p \mid n$，所以没有一项是 p 的倍数。由 Euclid 引理，p 不整除分子，更不整除 $\binom{n-1}{p-1}$。矛盾。

验证：$\binom{6}{2} = 15$，$6 \nmid 15$；$\binom{9}{3} = 84$，$9 \nmid 84$。并不是每个 $\binom{n}{k}$ 都失败（$9 \mid \binom{9}{1}$），证明只需找到一个 k。 ♦

3.5　同余

第 1 讲多次用到「看余数」：平方数除以 4 余 0 或 1，奇数的平方除以 8 余 1。每次都要写「设 $n = 4k + r$」再展开。Gauss 引入的同余记号把这类论证变成可以像等式一样运算的式子。

定义 3.14（同余）　设 m 为正整数。若 $m \mid a - b$，就说 a 与 b **模 m 同余**（congruent modulo m），记作 $a \equiv b \pmod{m}$；否则记作 $a \not\equiv b \pmod{m}$。m 叫**模**（modulus）。

例如 $17 \equiv 2 \pmod 5$，$-1 \equiv 6 \pmod 7$，$100 \equiv 0 \pmod 4$，$3 \not\equiv 8 \pmod{10}$。$m = 1$ 时任意两数都同余，所以通常取 $m \geq 2$。

$a \equiv b \pmod m$ 当且仅当 a, b 除以 m 的余数相同。证明：写 $a = mq_1 + r_1$，$b = mq_2 + r_2$，$0 \leq r_1, r_2 < m$，则 $m \mid a - b$ 当且仅当 $m \mid r_1 - r_2$；而 $|r_1 - r_2| < m$，m 的倍数中只有 0 满足这一点，所以这等价于 $r_1 = r_2$。

由此，模 m 同余是一个等价关系：$a \equiv a$；$a \equiv b \Rightarrow b \equiv a$；$a \equiv b$，$b \equiv c \Rightarrow a \equiv c$。整数按除以 m 的余数分成 m 个**剩余类**（residue class）$\{r + mk \mid k \in \mathbb{Z}\}$，$r = 0, 1, \dots, m-1$。可以把它想成一个有 m 个刻度的钟面：$-3, 2, 7, 12$ 在模 5 的钟面上指向同一个刻度 2。

定理 3.15（同余与运算相容）　若 $a \equiv b \pmod m$，$c \equiv d \pmod m$，则

$$a + c \equiv b + d, \qquad a - c \equiv b - d, \qquad ac \equiv bd \pmod m.$$

证明　写 $a=b+mk$，$c=d+ml$。则 $(a\pm c)-(b\pm d)=m(k\pm l)$，并且

$$ac-bd=(b+mk)(d+ml)-bd=m\,(bl+dk+mkl).$$ ■

> **推论 3.16**　若 $a\equiv b\pmod m$，则对任意正整数 k，$a^k\equiv b^k\pmod m$；对任意整系数多项式 f，$f(a)\equiv f(b)\pmod m$。

证明　对 k 归纳：$a^{k+1}=a^k\cdot a\equiv b^k\cdot b=b^{k+1}$，用的是定理 3.15 的乘法。$f(a)=\sum c_ia^i$ 的每一项满足 $c_ia^i\equiv c_ib^i$，再用加法。 ■

核心思想　只做加、减、乘时，可以随时把任何一个数换成与它同余的数（通常换成余数），最后结果的余数不变。这让我们在计算 2^{100} 的余数时，中间过程始终只和小数打交道。

第 1 讲「2023 不是两个平方数之和」现在一行就能写完：$x\equiv 0,1,2,3\pmod 4$ 给出 $x^2\equiv 0,1,0,1$，所以 $x^2+y^2\equiv 0,1,2\pmod 4$，而 $2023\equiv 3\pmod 4$。

■ 例 3.10　大数幂的余数

(a) 求 2^{100} 除以 7 的余数。(b) 求 7^{2012} 的个位数字。

解　(a) 先找一个余数为 1 的小幂：$2^3=8\equiv 1\pmod 7$。由于 $100=3\cdot 33+1$，

$$2^{100}=(2^3)^{33}\cdot 2\equiv 1^{33}\cdot 2=2\pmod 7.$$

这里用了推论 3.16：把底数 2^3 换成同余的 1。

(b) 个位数字就是模 10 的余数。$7^2=49\equiv -1\pmod{10}$，所以 $7^4\equiv 1$，个位以 $7,9,3,1$ 循环。$2012=4\cdot 503$，$7^{2012}=(7^4)^{503}\equiv 1$，个位是 1。 ♦

■ 例 3.11　弃九法与模 11 检验

设 $N=(d_rd_{r-1}\cdots d_0)_{10}$。证明 $N\equiv d_0+d_1+\cdots+d_r\pmod 9$，$N\equiv d_0-d_1+d_2-\cdots\pmod{11}$。用它们求 918082 除以 9 和 11 的余数。

解　令 $f(x)=d_rx^r+\cdots+d_1x+d_0$，则 $N=f(10)$。因为 $10\equiv 1\pmod 9$，由推论 3.16，$N=f(10)\equiv f(1)=\sum d_i\pmod 9$。因为 $10\equiv -1\pmod{11}$，$N\equiv f(-1)=d_0-d_1+d_2-\cdots\pmod{11}$。

对 $N=918082$：数字和 $9+1+8+0+8+2=28\equiv 1\pmod 9$，所以 N 除以 9 余 1。从个位起交错求和：$2-8+0-8+1-9=-22\equiv 0\pmod{11}$，所以 $11\mid N$。确实 $918082=11\cdot 83462$。手算乘法 $ab=c$ 时，可以检查 a,b,c 的数字和是否满足同样的关系模 9。 ♦

易错提醒　同余只与加、减、乘相容。下面两种「自然」的操作都是错的。

- **指数不能换成同余的数。**$1\equiv 4\pmod 3$，但 $2^1=2\not\equiv 16=2^4\pmod 3$。推论 3.16 换的是底数，指数必须原样保留。指数应该按什么模化简，要等第 4 讲的 Euler 定理。
- **不能随便约去公因数。**$2\cdot 3=6\equiv 16=2\cdot 8\pmod{10}$，但 $3\not\equiv 8\pmod{10}$。同样，$2\cdot 5\equiv 0\pmod{10}$ 不能推出 $2\equiv 0$ 或 $5\equiv 0$。

约分出错的原因是被约去的 2 与模 10 有公因数。去掉这个公因数，约分就合法了，代价是模也要缩小。

命题 3.17（消去律）　设 $g=\gcd(x,m)$。则 $ax\equiv bx \pmod m$ 当且仅当 $a\equiv b \pmod{m/g}$。特别地，若 $\gcd(x,m)=1$，则由 $ax\equiv bx \pmod m$ 可得 $a\equiv b \pmod m$。

证明　$m\mid (a-b)x$ 等价于 $\frac{m}{g}\mid (a-b)\frac{x}{g}$（两边同除以 g）。由第 1 讲练习 1.6，$\gcd(\frac{m}{g},\frac{x}{g})=1$，再由第 1 讲 Euclid 引理的一般形式，这等价于 $\frac{m}{g}\mid a-b$。（反方向：若 $\frac{m}{g}\mid a-b$，乘以整数 $\frac{x}{g}$ 即可。）∎

回到警告中的例子：$x=2$，$m=10$，$g=2$，所以 $2\cdot 3\equiv 2\cdot 8 \pmod{10}$ 只能推出 $3\equiv 8 \pmod 5$，这是对的。第 4 讲会把「$\gcd(x,m)=1$」解释为 x 模 m 有逆元。

想一想　由 $a^2\equiv b^2 \pmod m$ 能否推出 $a\equiv \pm b \pmod m$？分别考虑 $m=p$ 为素数和 $m=8$。
答　$m=p$ 时可以：$p\mid a^2-b^2=(a-b)(a+b)$，由 Euclid 引理 $p\mid a-b$ 或 $p\mid a+b$。$m=8$ 时不行：$1^2,3^2,5^2,7^2$ 都 $\equiv 1 \pmod 8$，但 $3\not\equiv \pm 1 \pmod 8$。原因是 $8\mid (3-1)(3+1)=2\cdot 4$，而两个因子各分到一部分 2，谁都不被 8 整除。

3.6　二项式系数模 p：Lucas 定理

Kummer 定理回答了 p 是否整除 $\binom{n}{k}$。若不整除，余数是多少？答案同样由 p 进制数字决定。为此要把同余推广到多项式。

两个整系数多项式 f,g 满足 $f\equiv g \pmod m$，是指 $f-g$ 的每个系数都被 m 整除。例如 $(1+x)^2=1+2x+x^2\equiv 1+x^2 \pmod 2$。定理 3.15 的证明原样适用于多项式：多项式同余也可以相加、相乘。

易错提醒　多项式同余比「代入每个整数都同余」强。$x^3-x=(x-1)x(x+1)$ 是三个连续整数之积，所以对每个整数 x 都有 $x^3\equiv x \pmod 3$。但作为多项式 $x^3\not\equiv x \pmod 3$，因为 x^3-x 的系数 $1,-1$ 不是 3 的倍数。下面的证明比较的是系数，所以必须用多项式同余。

引理 3.18　设 p 为素数，$j\ge 0$。则 $(1+x)^{p^j}\equiv 1+x^{p^j} \pmod p$。

证明　$j=0$ 时两边均为 $1+x$。$j=1$：$(1+x)^p=\sum_k \binom{p}{k}x^k$，由推论 3.13，$0<k<p$ 时 $p\mid \binom{p}{k}$，只剩首尾两项。设结论对 j 成立，两边取 p 次方，再对 $y=x^{p^j}$ 用 $j=1$ 的情形：

$$(1+x)^{p^{j+1}}=\left((1+x)^{p^j}\right)^p\equiv (1+x^{p^j})^p\equiv 1+x^{p^{j+1}} \pmod p.$$ ∎

定理 3.19（Lucas 定理）　设 p 为素数，$n=(n_r\cdots n_1n_0)_p$，$k=(k_r\cdots k_1k_0)_p$（位数不足时在前面补 0）。则

$$\boxed{\binom{n}{k}\equiv\binom{n_0}{k_0}\binom{n_1}{k_1}\cdots\binom{n_r}{k_r}\pmod p,}$$

其中约定 $k_i>n_i$ 时 $\binom{n_i}{k_i}=0$。

思路　算两遍 $(1+x)^n$ 中 x^k 的系数。一遍直接用二项式定理，得 $\binom{n}{k}$；另一遍先按 n 的 p 进制把 $(1+x)^n$ 拆成若干因子，再用引理 3.18。

证明　由 $n=\sum_i n_ip^i$ 和引理 3.18，

$$(1+x)^n=\prod_{i=0}^{r}\left((1+x)^{p^i}\right)^{n_i}\equiv\prod_{i=0}^{r}(1+x^{p^i})^{n_i}=\prod_{i=0}^{r}\sum_{j_i=0}^{n_i}\binom{n_i}{j_i}x^{j_ip^i}\pmod p.$$

展开右边，一般项是 $\binom{n_0}{j_0}\cdots\binom{n_r}{j_r}x^{j_0+j_1p+\cdots+j_rp^r}$，其中 $0\le j_i\le n_i\le p-1$。由于 p 进制表示唯一，指数等于 k 当且仅当每个 $j_i=k_i$。若所有 $k_i\le n_i$，x^k 的系数恰为 $\prod\binom{n_i}{k_i}$；若某个 $k_i>n_i$，没有这样的项，系数为 0，与约定一致。比较两边 x^k 的系数即得。 ■

■ **例 3.12　用 Lucas 定理求余数**　求 $\binom{1000}{500}$ 和 $\binom{100}{30}$ 除以 7 的余数。

解　先化成七进制。$1000=2\cdot343+6\cdot49+2\cdot7+6=(2626)_7$，$500=1\cdot343+3\cdot49+1\cdot7+3=(1313)_7$。逐位配对：

$$\binom{1000}{500}\equiv\binom{2}{1}\binom{6}{3}\binom{2}{1}\binom{6}{3}=2\cdot20\cdot2\cdot20=1600\pmod 7.$$

中途就可以约化：$20\equiv-1\pmod 7$，所以乘积 $\equiv2\cdot(-1)\cdot2\cdot(-1)=4$。余数是 4（$1600=7\cdot228+4$）。

$100=(202)_7$，$30=(042)_7$。中间一位是 $\binom{0}{4}=0$，所以 $\binom{100}{30}\equiv0\pmod 7$，即 $7\mid\binom{100}{30}$。Kummer 定理核对：七进制 $(42)_7+(130)_7$ 只在第 1 位进位（$4+3=7$），故 $7\parallel\binom{100}{30}$。 ♦

两个定理在同一处相遇：存在某一位 $k_i>n_i$，当且仅当减法 $n-k$ 至少发生一次借位，也当且仅当加法 $k+(n-k)$ 至少发生一次进位。这里说的是「是否发生」，并非逐位一一对应；低位的借位还可能沿着数字为 0 的位置继续传递。Kummer 数出进位总次数，从而得到 p 的方幂；Lucas 则计算模 p 的余数。

3.6.1　$p=2$：Pascal 三角的奇偶

$p=2$ 时每个 $\binom{n_i}{k_i}$ 只能是 $\binom{0}{0}=\binom{1}{0}=\binom{1}{1}=1$ 或 $\binom{0}{1}=0$。所以

$$\boxed{\binom{n}{k}\text{ 为奇数}\iff k\text{ 为 1 的二进制位上，}n\text{ 也是 1.}}$$

特别地，第 n 行全是奇数当且仅当 $n=2^m-1$（二进制全是 1）；第 2^m 行只有两端是奇数。第 n 行的奇数个数是 $2^{s_2(n)}$：n 的每个 1 位上 k 可取 0 或 1，其余位上只能取 0。例如 $10=(1010)_2$，

奇数只在 $k=0,2,8,10$ 处；第 10 行 $1,10,45,120,210,252,\ldots$ 中奇数确实只有 1 和 45。

把 Pascal 三角中的奇数涂成实心，就得到图 3.4。前 2^m 行由前 2^{m-1} 行的三个副本拼成，中间夹一个全为偶数的倒三角。这个自相似图形叫 Sierpiński 三角。

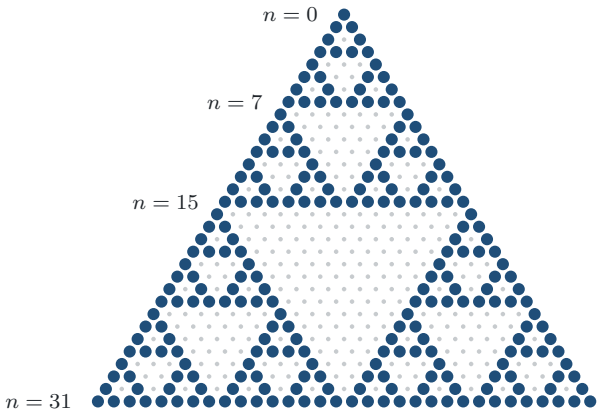


图 3.4　Pascal 三角前 32 行的奇偶：实心点为奇数，小灰点为偶数。第 $7,15,31$ 行全是奇数，第 $8,16$ 行只有两端是奇数。

前后联系　引理 3.18 的 $j=1$ 情形换成两个变量，就是「新生之梦」（freshman's dream）$(x+y)^p \equiv x^p + y^p \pmod p$。第 4 讲会用它对 a 归纳证明 Fermat 小定理 $a^p \equiv a \pmod p$。另一方面，例 3.9 等价于说：$n \geq 2$ 是素数当且仅当 $(1+x)^n \equiv 1 + x^n \pmod n$。2002 年的 AKS 算法从这个判别法出发，给出了第一个确定性多项式时间的素性检验。

历史与背景　同余记号 $\equiv$ 由 Gauss 在《算术研究》（1801）开篇引入，取其与等号相似。$n!$ 中素数方幂的公式见于 Legendre《数论》1808 年第二版；Kummer 的进位定理发表于 1852 年，Lucas 定理发表于 1878 年。杨辉《详解九章算法》（1261）载有 Pascal 三角，注明出自北宋贾宪，比 Pascal 的《论算术三角形》（1654）早约四百年。

本讲小结

- 在数某种东西的量一定是整数。由此 $k!$ 整除任意 k 个连续整数之积。
- 二项式定理与 Pascal 递推。$\binom{\alpha}{k}$ 对任意复数 α 有定义，$|z|<1$ 时 $(1+z)^\alpha = \sum_k \binom{\alpha}{k} z^k$。
- Legendre：$v_p(n!) = \sum_{i\geq 1} \lfloor n/p^i \rfloor = \dfrac{n - s_p(n)}{p-1}$。
- Kummer：$v_p\binom{n}{k}$ 等于 p 进制加法 $k+(n-k)$ 的进位次数；$0<k<p$ 时 $p \mid \binom{p}{k}$。
- $a \equiv b \pmod m$ 即 $m \mid a-b$，即余数相同。同余与加、减、乘相容，但**指数不能换成同余的数**；约去 x 要求 $\gcd(x,m)=1$。
- $(1+x)^{p^j} \equiv 1 + x^{p^j} \pmod p$。Lucas：$\binom{n}{k} \equiv \prod_i \binom{n_i}{k_i} \pmod p$；第 n 行有 $2^{s_2(n)}$ 个奇数。

练习

3.1 ★　用 Legendre 公式求 $v_7(1000!)$，再用命题 3.10 核对。1000! 末尾有多少个零？

3.2 ★★ 证明：没有 n 使 $n!$ 末尾恰好有 5 个零。求所有使 $n!$ 末尾恰好有 100 个零的 n。

3.3 ★ 用两种方法数「从 n 人中选 k 人委员会并指定其中一人为主席」，证明 $k\binom{n}{k}=n\binom{n-1}{k-1}$。由此求 $\sum_{k=0}^{n}k\binom{n}{k}$。

3.4 ★★ 用 $(1-x)^{2n}(1+x)^{2n}=(1-x^2)^{2n}$ 中 x^{2n} 的系数证明 $\sum_k(-1)^k\binom{2n}{k}^2=(-1)^n\binom{2n}{n}$。

3.5 ★★ 证明 $v_2\binom{2n}{n}=s_2(n)$。因此 $n\ge 1$ 时 $\binom{2n}{n}$ 为偶数，且 $4\mid\binom{2n}{n}\iff n$ 不是 2 的方幂。

3.6 ★★ 不展开 $\binom{20}{10}$，用 Kummer 定理求它的素因数分解。

3.7 ★ 求 3^{100} 除以 7 的余数、3^{2012} 的个位数字和 2^{100} 的末两位数字。

3.8 ★★ 证明立方数模 9 只能是 $0,1,8$。由此证明方程 $x^3+y^3+z^3=2012$ 没有整数解。

3.9 ★★ (a) 用 Lucas 定理求 $\binom{100}{10}$ 除以 3 的余数。(b) 哪些 $\binom{100}{k}$ 是奇数？

3.10 ★★★ 设 $p(x)$ 是次数 $\le n$ 的复系数多项式。(a) 证明存在唯一的 $c_0,\dots,c_n$ 使 $p(x)=\sum_{k=0}^{n}c_k\binom{x}{k}$。(b) 记 $\Delta f(x)=f(x+1)-f(x)$。证明 $\Delta\binom{x}{k}=\binom{x}{k-1}$（$k\ge 1$），并推出 $c_k=(\Delta^k p)(0)$。(c) 证明：p 在每个整数处取整数值，当且仅当所有 $c_k\in\mathbb{Z}$。(d) 把 $p(x)=\frac{x^3}{3}+\frac{x^2}{2}+\frac{x}{6}$ 写成这种形式。

练习答案与提示

3.1 $v_7(1000!)=142+20+2=164$。$1000=(2626)_7$，$s_7=16$，$(1000-16)/6=164$。末尾零个数为 $v_5(1000!)=200+40+8+1=249$；$v_2(1000!)=1000-s_2(1000)=994$ 更多。

3.2 记 $Z(n)=v_5(n!)$，它只在 5 的倍数处增加。$Z(24)=4$，$Z(25)=6$，跳过了 5。由 $Z(n)<n/4$ 得 $n>400$；$Z(400)=99$，$Z(405)=81+16+3=100$，$Z(410)=101$，所以 $n=405,\dots,409$。

3.3 先选 k 人再从中选主席：$k\binom{n}{k}$；先选主席再从其余 $n-1$ 人中选 $k-1$ 人：$n\binom{n-1}{k-1}$。求和得 $\sum_k k\binom{n}{k}=n\sum_k\binom{n-1}{k-1}=n2^{n-1}$。例如 $n=4$：$4+12+12+4=32$。

3.4 左边 x^{2n} 的系数是 $\sum_k(-1)^k\binom{2n}{k}\binom{2n}{2n-k}=\sum_k(-1)^k\binom{2n}{k}^2$。右边 $(1-x^2)^{2n}=\sum_j(-1)^j\binom{2n}{j}x^{2j}$，取 $j=n$。检验：$n=1$ 得 $1-4+1=-2$，$n=2$ 得 6。

3.5 由 Kummer 定理，$v_2\binom{2n}{n}=2s_2(n)-s_2(2n)=s_2(n)$，因为 $2n$ 的二进制是 n 末尾添 0。$n\ge1$ 时 $s_2(n)\ge1$；$s_2(n)\ge2$ 当且仅当 n 不是 2 的方幂。例：$\binom{6}{3}=20$，$\binom{8}{4}=70$。

3.6 二进制 $1010+1010$ 进位 2 次；三、五、七进制 $101+101$、$20+20$、$13+13$ 不进位；$p=11,\dots,19$ 时个位 $10+10\ge p$，进位一次。所以 $\binom{20}{10}=2^2\cdot11\cdot13\cdot17\cdot19=184756$。

3.7 $3^6\equiv1\pmod 7$，$3^{100}\equiv3^4=81\equiv4$。$3^k$ 的个位以 $3,9,7,1$ 循环，$4\mid2012$，个位是 1。$2^{20}\equiv24^2\equiv76\pmod{100}$，$76^2\equiv76$，所以 $2^{100}=(2^{20})^5\equiv76$。

3.8 写 $x=3t+r$，$r\in\{0,\pm1\}$，则 $x^3=27t^3+27t^2r+9tr^2+r^3\equiv r^3\in\{0,\pm1\}\pmod 9$。三个这样的数之和在 -3 与 3 之间，模 9 只能是 $0,1,2,3,6,7,8$，不会是 4 或 5。而 $2012=9\cdot223+5$。

3.9 (a) $100=(10201)_3$，$10=(00101)_3$，$\binom{100}{10}\equiv\binom{1}{1}\binom{0}{0}\binom{2}{1}\binom{0}{0}\binom{1}{1}=2\pmod 3$。(b) $100=(1100100)_2$，$s_2=3$，共 8 个奇数，位于 $k=0,4,32,36,64,68,96,100$（$\{64,32,4\}$ 的各子集之和）。

3.10 (a) 因 $\binom{x}{k}$ 次数为 k、首项系数为 $1/k!$，逐次消去最高次项可唯一展开。(b) Pascal 恒等式给出 $\Delta\binom{x}{k}=\binom{x}{k-1}$，故 $c_j=\Delta^j p(0)$，表示也唯一。(c) $c_k\in\mathbb{Z}$ 时 $p(m)\in\mathbb{Z}$；反之

$$c_k=\Delta^k p(0)=\sum_{i=0}^{k}(-1)^{k-i}\binom{k}{i}p(i)\in\mathbb{Z},$$

因为各 $p(i)$ 都是整数。(d) 前向差分为 $0,1,5,14$；$1,4,9$；$3,5$；2，故系数为 $1,3,2$，即 $p(x)=\binom{x}{1}+3\binom{x}{2}+2\binom{x}{3}$。

第 4 讲

Fermat、Euler、Wilson 定理与一次同余式

Fermat, Euler, Wilson, Linear Congruences

对应材料：MIT 18.781 Lecture 4 讲义；习题集 2 第 4、5、6 题，习题集 3 第 2 题的 3^{100} 部分（及官方解答）。

本讲目标

(1) 理解完全剩余系与简化剩余系，知道 Euler 函数 $\varphi(m)$ 数的是什么，会判断一组数是不是剩余系。

(2) 掌握 Euler 定理和 Fermat 小定理的「乘积比较」证明，以及 Fermat 小定理基于「新生之梦」$(x+y)^p \equiv x^p + y^p$ 的归纳证明和项链计数证明。

(3) 会用 Euler 定理给指数降次，求大数幂的余数和末位数字；知道 Fermat 小定理的逆命题不成立。

(4) 理解模 m 的逆元何时存在、为何唯一，会用扩展 Euclid 算法和 Euler 定理求逆元。

(5) 掌握 Wilson 定理的配对证明及其逆命题，并用它证明：$x^2 \equiv -1 \pmod p$ 有解当且仅当 $p = 2$ 或 $p \equiv 1 \pmod 4$；由此得到 $4k+1$ 型素数有无穷多个。

(6) 理解同余方程「解」的含义和解数；掌握一次同余式 $ax \equiv b \pmod m$ 有解的充要条件和解的个数。

第 3 讲引入了同余：$a \equiv b \pmod m$ 指 $m \mid a-b$。同余与加法、乘法相容，所以可以「先取余再运算」；但消去律要小心，由 $ax \equiv bx \pmod m$ 推出 $a \equiv b \pmod m$ 需要 $\gcd(x, m) = 1$。本讲用这些工具回答几个具体的问题：

- 3^{100} 的末两位是多少？
- 对素数 p，$(p-1)!$ 除以 p 余几？
- 哪些素数能整除形如 x^2+1 的数？
- 同余式 $6x \equiv 9 \pmod{15}$ 有几个解？

前三个问题的答案都来自同一个想法：**把一整组剩余乘起来，用两种方式计算这个乘积。**

4.1　剩余系与 Euler 函数

4.1.1　完全剩余系

模 m 的余数只有 $0,1,\dots,m-1$ 这 m 种。同余的两个数在加法和乘法中可以互相替换，所以研究模 m 的问题时，每个剩余类只需取一个代表。

定义 4.1（完全剩余系）　设 $m \geq 1$。若整数 $a_1,\dots,a_m$ 模 m 两两不同余，并且每个整数都与其中某个 a_i 模 m 同余，则称 $\{a_1,\dots,a_m\}$ 是模 m 的一个**完全剩余系**（complete residue system）。

换句话说，完全剩余系从每个剩余类里**恰好**取一个数，不重不漏。最常用的是**最小非负剩余系** $\{0,1,\dots,m-1\}$。m 为奇数时也常用**绝对最小剩余系** $\{-\frac{m-1}{2},\dots,-1,0,1,\dots,\frac{m-1}{2}\}$，它让每个代表的绝对值都不超过 $m/2$。例如模 7：

$$\{0,1,2,3,4,5,6\},\qquad \{-3,-2,-1,0,1,2,3\},\qquad \{7,8,9,10,11,12,13\},\qquad \{0,3,6,9,12,15,18\}$$

都是完全剩余系。最后一个是 $\{3k : 0 \leq k \leq 6\}$，模 7 分别余 $0,3,6,2,5,1,4$，恰好每种余数出现一次。

4.1.2　简化剩余系与 Euler 函数

后面要做乘法和消去，而消去只对与 m 互素的数可靠。「与 m 互素」是整个剩余类共有的性质：若 $a \equiv b \pmod m$，写 $a = b + km$，则 a 与 m 的公因数也整除 $b = a - km$，反之亦然，所以 $\gcd(a,m) = \gcd(b,m)$。因此可以只挑出与 m 互素的那些剩余类。

定义 4.2（简化剩余系与 Euler 函数）　若整数 $r_1,\dots,r_k$ 都与 m 互素、模 m 两两不同余，并且每个与 m 互素的整数都与某个 r_i 模 m 同余，则称 $\{r_1,\dots,r_k\}$ 是模 m 的一个**简化剩余系**（reduced residue system）。简化剩余系的元素个数记作 $\varphi(m)$，称为 **Euler 函数**（Euler's totient function）。等价地，

$$\varphi(m) = \#\{\, n \mid 1 \leq n \leq m,\ \gcd(n,m) = 1 \,\}.$$

从任意一个完全剩余系中删去与 m 不互素的数，剩下的就是简化剩余系。以 $m = 12$ 为例，从 $\{1,2,\dots,12\}$ 中只留下与 12 互素的数，得到 $\{1,5,7,11\}$，所以 $\varphi(12) = 4$（见图 4.1）。

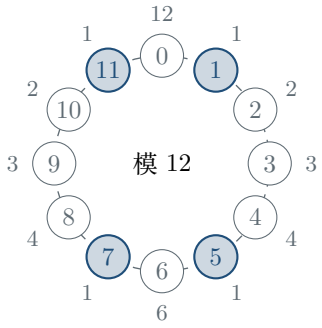


图 4.1　模 12 的钟面，外圈小字是 $\gcd(k,12)$。蓝色的 $1,5,7,11$ 组成简化剩余系，$\varphi(12)=4$。

下表列出 $\varphi(m)$ 的前几个值。对素数 p，$1,2,\dots,p-1$ 都与 p 互素，所以$\varphi(p)=p-1$；而 1 与自身互素，故 $\varphi(1)=1$。

表 4.1 列出直接计数的结果，后面的积性公式将解释这些数值。

m	1	2	3	4	5	6	7	8	9	10	11	12
$\varphi(m)$	1	1	2	2	4	2	6	4	6	4	10	4

表 4.1　Euler 函数的前 12 个值。

对较大的 m 也可以直接数。例如 $1,\dots,100$ 中，2 的倍数有 50 个，5 的倍数有 20 个，10 的倍数（被重复去掉的）有 10 个，所以

$$\varphi(100)=100-50-20+10=40.$$

同理 $\varphi(25)=25-5=20$。φ 的一般公式和「积性」留到第 5 讲。

判断一组数是不是剩余系时，「不漏」往往不好直接验证。下面的命题说，只要个数对，「不重」就够了。

命题 4.3　设 $m\ge 1$。

(1) 任意 m 个模 m 两两不同余的整数构成完全剩余系。

(2) 任意 $\varphi(m)$ 个与 m 互素、模 m 两两不同余的整数构成简化剩余系。

(3) 若 $\{a_1,\dots,a_m\}$ 是完全剩余系，$\gcd(a,m)=1$，b 为任意整数，则 $\{aa_1+b,\dots,aa_m+b\}$ 也是完全剩余系。

证明　(1) 这 m 个数的余数两两不同，而余数只有 m 种，所以每种余数恰好出现一次（鸽笼原理）。

(2) 与 m 互素的剩余类恰有 $\varphi(m)$ 个。这 $\varphi(m)$ 个数落在其中且两两不同，同样每类恰好出现一次。

(3) 由 (1)，只需证两两不同余。若 $aa_i+b\equiv aa_j+b\pmod m$，则 $aa_i\equiv aa_j\pmod m$。因 $\gcd(a,m)=1$，可以消去 a，得 $a_i\equiv a_j$，于是 $i=j$。 ■

(3) 中 $\gcd(a,m)=1$ 不能去掉。取 $m=12$，$a=2$，$b=0$：$\{0,2,4,\dots,22\}$ 模 12 只有偶数余数，每个出现两次。

■ **例 4.1　判断剩余系**　下列各组数中，哪些是模 12 的完全剩余系或简化剩余系？(a) $\{5k : 0 \le k \le 11\}$；　(b) $\{5,25,35,55\}$；　(c) $\{1,3,5,7\}$；　(d) $\{1,5,13,17\}$；　(e) $\{-5,-1,1,5\}$。

解　(a) $\gcd(5,12)=1$，由命题 4.3(3)（取 $a_k=k$，$b=0$），这是完全剩余系。直接算，余数依次是 $0,5,10,3,8,1,6,11,4,9,2,7$，确实不重不漏。

(b) 四个数都与 12 互素，模 12 分别余 $5,1,11,7$，两两不同，个数等于 $\varphi(12)=4$。由命题 4.3(2)，这是简化剩余系。它正好是 $5\cdot\{1,5,7,11\}$。

(c) $\gcd(3,12)=3$，不是简化剩余系；个数不是 12，也不是完全剩余系。

(d) 都与 12 互素，但 $13\equiv 1$，$17\equiv 5$，有重复，漏掉了 7 和 11 两类。不是简化剩余系。

(e) 模 12 分别余 $7,11,1,5$，是简化剩余系。　♦

想一想　定义中说简化剩余系的元素个数「记作 $\varphi(m)$」。这里默认了两个不同的简化剩余系元素个数相同。为什么？

答　每个简化剩余系从「与 m 互素的剩余类」中各取恰好一个数，所以它的元素与这些剩余类一一对应，个数都等于这类剩余类的个数。完全剩余系的个数总是 m，道理相同。

4.2　Euler 定理与 Fermat 小定理

4.2.1　一张幂次表

先看模 7 时 a^k 的余数。

a	a^1	a^2	a^3	a^4	a^5	a^6
1	1	1	1	1	1	**1**
2	2	4	1	2	4	**1**
3	3	2	6	4	5	**1**
4	4	2	1	4	2	**1**
5	5	4	6	2	3	**1**
6	6	1	6	1	6	**1**

表 4.2　模 7 的幂次表。最后一列全是 1，而 $6=\varphi(7)$。

最后一列全是 1。再看模 10：$3,9,7,1$ 是 $3^1,\dots,3^4$ 的余数，$3^4\equiv 1$，而 $\varphi(10)=4$。但 2^k 模 10 依次是 $2,4,8,6,2,\dots$，永远不会是 1，因为 2^k-1 是奇数，不可能被 10 整除。这提示我们：**当 $\gcd(a,m)=1$ 时，$a^{\varphi(m)}\equiv 1 \pmod m$。**

4.2.2 乘以 a 只是重新排列

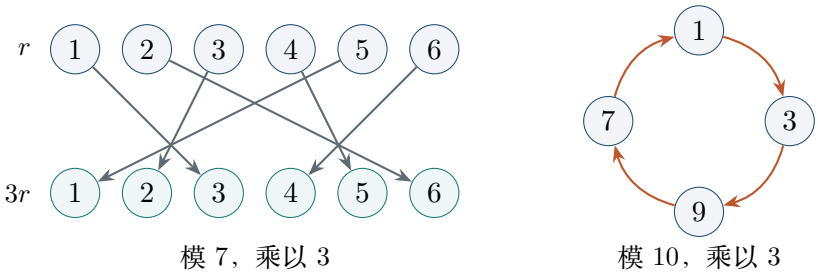
引理 4.4　设 $\gcd(a,m)=1$，$\{r_1,\dots,r_k\}$ 是模 m 的简化剩余系，$k=\varphi(m)$。则 $\{ar_1,\dots,ar_k\}$ 也是模 m 的简化剩余系。

证明　由命题 4.3(2)，只需证明这 k 个数都与 m 互素，且两两不同余。

互素：若素数 q 同时整除 ar_i 和 m，则由 Euclid 引理 $q\mid a$ 或 $q\mid r_i$，与 $\gcd(a,m)=1$ 或 $\gcd(r_i,m)=1$ 矛盾。所以 $\gcd(ar_i,m)=1$。

不同余：若 $ar_i\equiv ar_j\pmod m$，则 $m\mid a(r_i-r_j)$。由 $\gcd(a,m)=1$ 得 $m\mid r_i-r_j$，即 $r_i\equiv r_j$，所以 $i=j$。 ■

于是存在 $\{1,\dots,k\}$ 的一个排列 σ，使 $ar_i\equiv r_{\sigma(i)}\pmod m$：**乘以 a 只是把简化剩余系重新排列了一遍**。图 4.2 画出了两个例子。



模 7，乘以 3　　模 10，乘以 3

图 4.2　左：模 7 时 $r\mapsto 3r$ 把 $1,2,\dots,6$ 变成 $3,6,2,5,1,4$，只是换了顺序。右：模 10 的简化剩余系 $\{1,3,7,9\}$ 在乘 3 下转成一个圈 $1\to 3\to 9\to 7\to 1$。两种情形下，所有数的乘积都不变。

定理 4.5（Euler 定理）　若 $\gcd(a,m)=1$，则

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

思路　把简化剩余系中所有数的乘积算两遍：一遍直接乘；另一遍先把每个数乘以 a 再乘。由引理 4.4，两遍乘的是同一组剩余，只是顺序不同。

证明　取简化剩余系 $r_1,\dots,r_k$，$k=\varphi(m)$，记 $R=r_1r_2\cdots r_k$。由引理 4.4，$ar_i\equiv r_{\sigma(i)}\pmod m$，其中 σ 是一个排列，于是

$$a^kR=(ar_1)(ar_2)\cdots(ar_k)\equiv r_{\sigma(1)}r_{\sigma(2)}\cdots r_{\sigma(k)}=r_1r_2\cdots r_k=R\pmod m.$$

每个 r_i 都与 m 互素，所以 $\gcd(R,m)=1$，可以消去 R，得 $a^k\equiv 1\pmod m$。 ■

注 4.1　为什么不用完全剩余系？它的乘积含因子 0，两边都 $\equiv 0$，得不到任何信息。最后一步消去 R 要求每个因子都与 m 互素，这正是引入简化剩余系的原因。

取 $m=p$ 为素数，就得到数论中最常用的定理之一。

推论 4.6（Fermat 小定理）　设 p 是素数。

(1) 若 $p \nmid a$，则 $a^{p-1} \equiv 1 \pmod p$。

(2) 对一切整数 a，$a^p \equiv a \pmod p$。

证明　(1) $\varphi(p) = p-1$，且 $p \nmid a$ 等价于 $\gcd(a,p)=1$，由 Euler 定理即得。(2) 若 $p \nmid a$，把 (1) 两边乘以 a。若 $p \mid a$，两边都 $\equiv 0 \pmod p$。 ■

历史与背景　Fermat 在 1640 年给 Frénicle 的信中叙述了这个定理，没有给出证明，说是怕信写得太长。第一个发表的证明出自 Euler（1736 年提交，1741 年发表），思路与 §4.3 的归纳法相同。1763 年 Euler 引入了 φ 这个函数（当时还没有专门的记号），把定理推广到一般的模 m；记号 $\varphi(n)$ 来自 Gauss 的《算术研究》(1801)。本节的乘积比较证明出现得更晚，通常归于 Ivory（1806）。

4.2.3　用 Euler 定理降次

若 $\gcd(a,m)=1$，令 $n = q\varphi(m)+s$，$0 \le s < \varphi(m)$，则 $a^n = \left(a^{\varphi(m)}\right)^q a^s \equiv a^s \pmod m$。所以**指数可以模 $\varphi(m)$ 来算**。

■ **例 4.2　大数幂的余数**　(a) 求 3^{100} 除以 7 的余数。(b) 求 3^{100} 的末两位数字。(c) 证明 $13 \mid 2^{70}+3^{70}$。

解　(a) $\varphi(7)=6$，$100 = 6\cdot 16+4$，所以 $3^{100} \equiv 3^4 = 81 \equiv 4 \pmod 7$。这与表 4.2 中 3^4 一列一致。

(b) 末两位就是模 100 的余数。$\gcd(3,100)=1$，$\varphi(100)=40$，$100 = 40\cdot 2+20$，所以 $3^{100} \equiv 3^{20} \pmod{100}$。再算 3^{20}：

$$3^5 = 243 \equiv 43, \qquad 3^{10} \equiv 43^2 = 1849 \equiv 49, \qquad 3^{20} \equiv 49^2 = 2401 \equiv 1 \pmod{100}.$$

所以 $3^{100} \equiv 1 \pmod{100}$，末两位是 01。

(c) 13 是素数，由 Fermat 小定理 $2^{12} \equiv 3^{12} \equiv 1 \pmod{13}$。$70 = 12\cdot 5+10$，所以

$$2^{70}+3^{70} \equiv 2^{10}+3^{10} \pmod{13}.$$

$2^{10} = 1024 = 13\cdot 78+10 \equiv 10$；$3^3 = 27 \equiv 1$，故 $3^{10} = (3^3)^3\cdot 3 \equiv 3$。于是 $2^{70}+3^{70} \equiv 10+3 = 13 \equiv 0 \pmod{13}$。 ♦

(b) 中 3^{20} 已经 $\equiv 1 \pmod{100}$，比 Euler 定理保证的 40 次更早，见本节末的思考题。

换成 2^{100}，$\gcd(2,100)=2$，不能对模 100 用 Euler 定理。可以把 100 拆成互素的 $4\cdot 25$：$2^{100} = 4\cdot 2^{98} \equiv 0 \pmod 4$，模 25 则可用 Euler 定理。把两个余数合成模 100 的余数是第 5 讲中国剩余定理的内容，答案是 76。

■ **例 4.3　习题集 2 第 5 题**　证明：对每个整数 n，$\dfrac{n^7}{7}+\dfrac{n^{11}}{11}+\dfrac{59n}{77}$ 都是整数。

解　通分，原式 $=\dfrac{11n^7+7n^{11}+59n}{77}$。$77=7\cdot 11$，且 $7,11$ 互素，所以只需证分子 N 同时被 7 和 11 整除。

模 7：由 Fermat 小定理 $n^7\equiv n$，又 $7n^{11}\equiv 0$，所以 $N\equiv 11n+59n=70n\equiv 0 \pmod 7$。

模 11：$n^{11}\equiv n$，$11n^7\equiv 0$，所以 $N\equiv 7n+59n=66n\equiv 0 \pmod{11}$。

于是 $77\mid N$。这里用 $a^p\equiv a$ 的形式，它对 $p\mid n$ 也成立，不必分情况。 ♦

易错提醒　Fermat 小定理的**逆命题不成立**：$a^{n-1}\equiv 1 \pmod n$ 不能推出 n 是素数。例如 $341=11\cdot 31$，而 $2^{10}=1024=3\cdot 341+1$，所以 $2^{340}=(2^{10})^{34}\equiv 1 \pmod{341}$。逆否命题须保留条件：设 $n>1$ 且 $\gcd(a,n)=1$，若 $a^{n-1}\not\equiv 1 \pmod n$，则 n 是合数。例如第 5 讲用反复平方法会算出 $3^{340}\equiv 56 \pmod{341}$，换底数 3 就揭穿了 341。互素条件不能省：取素数 $n=2$、底数 $a=2$，就有 $2^1\not\equiv 1 \pmod 2$。第 6 讲会讨论这种素性检验，以及对所有互素底数都能「骗过」它的 Carmichael 数。

想一想　$\varphi(m)$ 是使 $a^n\equiv 1 \pmod m$ 的最小正整数 n 吗？在表 4.2 中，每一行第一次出现 1 的位置是多少？它们与 6 有什么关系？

答　不一定是最小的。各行第一次出现 1 的位置依次是 $1,3,6,3,6,2$，都是 6 的因数。一般地，设 d 是使 $a^d\equiv 1$ 的最小正整数，把 $\varphi(m)$ 写成 $qd+s$，$0\le s<d$，则 $1\equiv a^{\varphi(m)}=(a^d)^q a^s\equiv a^s$。由 d 的最小性只能 $s=0$，即 $d\mid\varphi(m)$。这个 d 叫做 a 模 m 的阶，后面几讲会系统研究。

4.3　Fermat 小定理的另两个证明

初学代数的人常犯一个错误：把 $(x+y)^2$ 写成 x^2+y^2。这被戏称为**新生之梦**（freshman's dream）。模素数 p 时，这个「错误」对 p 次方竟然是对的。

引理 4.7（新生之梦）　设 p 是素数，x,y 是整数，则 $(x+y)^p\equiv x^p+y^p \pmod p$。

证明　由二项式定理，$(x+y)^p=\sum_{k=0}^{p}\binom{p}{k}x^k y^{p-k}$。第 3 讲证明过：$0<k<p$ 时 $p\mid\binom{p}{k}$，因为 $\binom{p}{k}=\dfrac{p!}{k!\,(p-k)!}$ 的分子含因子 p，而分母是小于 p 的数之积，不含因子 p。所以中间各项都 $\equiv 0 \pmod p$，只剩 $k=0$ 和 $k=p$ 两项。 ■

$n=6$	1	6	15	20	15	6	1
$n=7$	1	**7**	**21**	**35**	**35**	**21**	**7**	1

表 4.3　Pascal 三角的第 6、7 行。$n=7$ 是素数，中间各项（粗体）都是 7 的倍数；$n=6$ 时 15 和 20 都不是 6 的倍数。

对 k 归纳，引理推广到多个数：$(x_1+\cdots+x_k)^p\equiv x_1^p+\cdots+x_k^p \pmod p$。

Fermat 小定理 (2) 的第二个证明　先对 $a\ge 0$ 归纳。$a=0$ 时 $0^p=0$。设 $a^p\equiv a \pmod p$，由

引理 4.7，

$$(a+1)^p \equiv a^p + 1^p \equiv a+1 \pmod{p}.$$

所以对一切 $a \geq 0$ 成立。对负整数 a，取 $a' \geq 0$ 使 $a' \equiv a \pmod{p}$，则 $a^p \equiv a'^p \equiv a' \equiv a \pmod{p}$。 ■

直观理解 两个证明的出发点不同。第一个证明看**乘法**：乘以 a 只是重排简化剩余系。第二个证明看**加法**：「p 次方」这个运算模 p 与加法交换，而每个整数都是若干个 1 的和，$1^p = 1$，所以 p 次方把每个整数都映回自己。

还有第三个证明，只用计数，对 $a \geq 1$ 证明 $p \mid a^p - a$。

■ **例 4.4** 项链计数　设 p 是素数，$a \geq 1$。用 a 种颜色给排成一圈的 p 颗珠子染色，旋转后相同的算同一条项链。证明：不是单色的项链恰有 $\frac{a^p-a}{p}$ 条，从而 $p \mid a^p - a$。

解 把珠子编号 $0, 1, \ldots, p-1$，一种染色就是一个颜色序列 $c = (c_0, \ldots, c_{p-1})$，共 a^p 个，其中单色的 a 个。记 $\sigma c = (c_{p-1}, c_0, \ldots, c_{p-2})$ 为转一格，则 $\sigma^p c = c$，一条项链对应的序列就是 $c, \sigma c, \ldots, \sigma^{p-1} c$。

断言：c 不是单色时，这 p 个序列两两不同。若 $\sigma^i c = \sigma^j c$，$0 \leq i < j \leq p-1$，两边再转 $p-i$ 格得 $\sigma^k c = c$，$k = j - i$，$0 < k < p$。即每颗珠子与往后第 k 颗同色（下标模 p），所以 $c_0 = c_k = c_{2k} = \cdots = c_{(p-1)k}$。又若 $jk \equiv j'k \pmod{p}$，$0 \leq j, j' < p$，则 $p \mid (j-j')k$，而 $p \nmid k$，由 Euclid 引理 $j = j'$。所以 $0, k, \ldots, (p-1)k$ 取遍所有珠子，c 是单色的，矛盾。

两条项链若有公共序列，就互相转得到，是同一条。所以 $a^p - a$ 个非单色序列恰好分成若干组，每组 p 个，组数就是非单色项链的条数，$p \mid a^p - a$。

检验 $a = 2$，$p = 3$：$001, 010, 100$ 是一条项链，$011, 110, 101$ 是另一条，$2^3 - 2 = 6 = 2 \cdot 3$。断言用到了 p 是素数：4 颗珠子时 0101 转两格回到自己，只对应 $0101, 1010$ 两个序列；事实上 $2^4 - 2 = 14$ 不被 4 整除。 ♦

4.4 模 m 的逆元

在 $\mathbb{Q}$ 中可以除以任何非零数。模 m 时，「除以 a」应该理解为「乘以一个 b，使 $ab \equiv 1$」。

命题 4.8（逆元） 存在整数 b 使 $ab \equiv 1 \pmod{m}$，当且仅当 $\gcd(a, m) = 1$。此时 b 模 m 唯一，称为 a 模 m 的**逆元**（inverse），记作 $a^{-1} \bmod m$。

证明 必要性：若 $ab = 1 + km$，则 $\gcd(a, m)$ 整除 $ab - km = 1$。

充分性：由 Bézout 定理（第 1 讲），有整数 u, v 使 $au + mv = 1$，所以 $au \equiv 1 \pmod{m}$，取 $b = u$。

唯一性：若 $ab \equiv ab' \equiv 1$，则 $b \equiv b(ab') = (ab)b' \equiv b' \pmod{m}$。 ■

求逆元有三种办法：m 很小时直接试；一般情形用扩展 Euclid 算法求出 Bézout 系数；或者用 Euler 定理，$a \cdot a^{\varphi(m)-1} = a^{\varphi(m)} \equiv 1$，所以

$$a^{-1} \equiv a^{\varphi(m)-1} \pmod{m} \qquad (\gcd(a,m)=1).$$

■ **例 4.5　两种方法求逆元**　求 11 模 26 的逆元，并解同余式 $11x \equiv 5 \pmod{26}$。

解　*方法一：扩展 Euclid 算法。*先做辗转相除，再从下往上回代：

$$\begin{aligned} 26 &= 2\cdot 11 + 4, & 1 &= 4 - 3 \\ 11 &= 2\cdot 4 + 3, & &= 4 - (11 - 2\cdot 4) = 3\cdot 4 - 11 \\ 4 &= 1\cdot 3 + 1, & &= 3\cdot(26 - 2\cdot 11) - 11 = 3\cdot 26 - 7\cdot 11. \end{aligned}$$

所以 $11\cdot(-7) \equiv 1 \pmod{26}$，$11^{-1} \equiv -7 \equiv 19$。验证：$11\cdot 19 = 209 = 8\cdot 26 + 1$。

方法二：Euler 定理。$1,\dots,26$ 中 2 的倍数 13 个，13 的倍数 2 个，26 的倍数 1 个，$\varphi(26) = 26 - 13 - 2 + 1 = 12$，所以 $11^{-1} \equiv 11^{11}$。逐次平方：

$$11^2 = 121 \equiv 17, \qquad 11^4 \equiv 17^2 = 289 \equiv 3, \qquad 11^8 \equiv 9 \pmod{26},$$

于是 $11^{11} = 11^8\cdot 11^2\cdot 11 \equiv 9\cdot 17\cdot 11 = 9\cdot 187 \equiv 9\cdot 5 = 45 \equiv 19 \pmod{26}$，与方法一相同。

*解同余式。*两边乘以 $11^{-1} \equiv 19$：$x \equiv 19\cdot 5 = 95 \equiv 17 \pmod{26}$。验证：$11\cdot 17 = 187 = 7\cdot 26 + 5$。 ♦

m 不大时，可以把所有逆元列成表。下表是模 13 的逆元，下一节要用。

a	1	2	3	4	5	6	7	8	9	10	11	12
a^{-1}	**1**	7	9	10	8	11	2	5	3	4	6	**12**

表 4.4　模 13 的逆元。只有 1 和 12 是自己的逆元。

易错提醒　$a^{-1} \bmod m$ 是一个剩余类，不是分数。$2^{-1} \bmod 7$ 是 4（因为 $2\cdot 4 = 8 \equiv 1$），而 $2^{-1} \bmod 5$ 是 3：同一个 a 换了模，逆元就变了。$\gcd(a,m) > 1$ 时逆元根本不存在。例如模 12 时 $3\cdot 4 \equiv 0$；若 $4b \equiv 1$，则 $3 \equiv 3\cdot 4b \equiv 0\cdot b = 0 \pmod{12}$，矛盾。

想一想　模 12 时哪些数有逆元？各是什么？与表 4.4 比较，有什么不同？

答　有逆元的恰是 $1, 5, 7, 11$。$5^2 = 25$，$7^2 = 49$，$11^2 = 121$ 模 12 都是 1，所以每个都是自己的逆元。模素数 13 时却只有 1 和 12 是自己的逆元。下一节会看到这个差别正是 Wilson 定理的关键。

4.5　Wilson 定理

4.5.1　$(n-1)!$ 模 n

先算几个值。

表 4.5 对比素数与合数情形，提示了 Wilson 定理的结论。

n	**2**	**3**	4	**5**	6	**7**	8	9	10	**11**	12
$(n-1)! \bmod n$	**1**	**2**	2	**4**	0	**6**	0	0	0	**10**	0

表 4.5　$(n-1)!$ 除以 n 的余数。n 为素数（粗体）时余数都是 $n-1$，即 $(n-1)! \equiv -1$；n 为合数时，除 $n=4$ 外余数都是 0。

要解释素数这一行，先看哪些数是自己的逆元。$a \equiv a^{-1}$ 即 $a^2 \equiv 1$。

引理 4.9　设 p 是素数。若 $x^2 \equiv 1 \pmod p$，则 $x \equiv 1$ 或 $x \equiv -1 \pmod p$。

证明　$p \mid x^2 - 1 = (x-1)(x+1)$。由 Euclid 引理，$p \mid x-1$ 或 $p \mid x+1$。 ■

模合数时引理不成立：模 8 时 $1, 3, 5, 7$ 的平方都 $\equiv 1$；模 12 的情形见上一节的思考题。原因是 $m \mid (x-1)(x+1)$ 时，m 的因子可以分摊到两个因式上。

4.5.2　配对证明

定理 4.10（Wilson 定理）　若 p 是素数，则

$$\boxed{(p-1)! \equiv -1 \pmod p.}$$

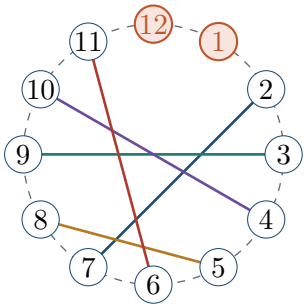
思路　$1, 2, \ldots, p-1$ 每个数都有逆元，而且逆元仍在这些数中。把每个数和它的逆元配成一对，每对的乘积 $\equiv 1$。由引理 4.9，配不上对的只有自己的逆元 1 和 $p-1$。

证明　$p=2$ 时 $1! = 1 \equiv -1 \pmod 2$；$p=3$ 时 $2! = 2 \equiv -1 \pmod 3$。下设 $p \geq 5$。

对 $a \in \{1, \ldots, p-1\}$，由命题 4.8，a 模 p 有唯一的逆元，取它在 $\{1, \ldots, p-1\}$ 中的代表，记作 a'。由唯一性，$(a')' = a$，所以「$a \leftrightarrow a'$」把这 $p-1$ 个数分成若干对，只有满足 $a' = a$ 的数自成一对。由引理 4.9，$a' = a$ 当且仅当 $a = 1$ 或 $a = p-1$。于是 $2, 3, \ldots, p-2$ 恰好分成 $\frac{p-3}{2}$ 对，每对乘积 $\equiv 1$，从而

$$(p-1)! = 1 \cdot \underbrace{\big(2 \cdot 3 \cdots (p-2)\big)}_{\equiv 1} \cdot (p-1) \equiv p-1 \equiv -1 \pmod p.$$ ■

图 4.3 将模 13 的非零数按逆元配对，保留的两个自逆元素决定乘积。



$2 \cdot 7 = 14 \equiv 1$
$3 \cdot 9 = 27 \equiv 1$
$4 \cdot 10 = 40 \equiv 1$
$5 \cdot 8 = 40 \equiv 1$
$6 \cdot 11 = 66 \equiv 1$
$1 \cdot 1 \equiv 1, \quad 12 \cdot 12 \equiv 1$

图 4.3　$p = 13$ 时的配对。每条弦连接互逆的两个数；橙色的 1 和 12 自成一对，所以 $12! \equiv 1 \cdot 12 \equiv -1 \pmod{13}$。

Wilson 定理的逆命题也成立，因此它给出了素数的一个刻画。

命题 4.11（Wilson 定理的逆）　设 $n \geq 2$。若 $(n-1)! \equiv -1 \pmod{n}$，则 n 是素数。

证明　若 n 是合数，取因数 d，$1 < d < n$。则 d 是 $(n-1)!$ 的一个因子，$d \mid (n-1)!$。又 $d \mid n \mid (n-1)! + 1$，所以 $d \mid 1$，矛盾。 ■

于是「n 是素数 $\iff (n-1)! \equiv -1 \pmod{n}$」。但算 $(n-1)! \bmod n$ 要做约 n 次乘法，比试除到 $\sqrt{n}$ 还慢，所以这个判别法没有实用价值。

■ **例 4.6　用 Wilson 定理化简阶乘**　求 16! 除以 19 的余数。

解　由 Wilson 定理 $18! \equiv -1 \pmod{19}$。而 $18! = 16! \cdot 17 \cdot 18$，且 $17 \equiv -2$，$18 \equiv -1$，所以

$$-1 \equiv 16! \cdot (-2)(-1) = 2 \cdot 16! \pmod{19}.$$

$2^{-1} \equiv 10 \pmod{19}$（因 $2 \cdot 10 = 20 \equiv 1$），两边乘以 10：$16! \equiv -10 \equiv 9 \pmod{19}$。验证：$2 \cdot 9 = 18 \equiv -1$。 ♦

想一想　Wilson 定理的证明在哪两处用到了 p 是素数？对 $m = 15$，把简化剩余系 $\{1, 2, 4, 7, 8, 11, 13, 14\}$ 全部乘起来，模 15 得多少？

答　第一处：$1, \ldots, p-1$ 每个数都有逆元（模合数时，与 m 不互素的数没有逆元）。第二处：只有 ± 1 是自己的逆元（引理 4.9）。模 15 时只考虑简化剩余系，第一处没有问题，但 $x^2 \equiv 1$ 有 $1, 4, 11, 14$ 四个解。配对 $2 \leftrightarrow 8$、$7 \leftrightarrow 13$ 后，乘积 $\equiv 1 \cdot 4 \cdot 11 \cdot 14$。$4 \cdot 11 = 44 \equiv 14$，$14 \cdot 14 = 196 \equiv 1$，所以乘积 $\equiv 1$，不是 -1。

历史与背景　这个结果最早可能由 Ibn al-Haytham（约 1000 年）发现。1770 年 Waring 在书中发表了它，归功于他的学生 John Wilson，但两人都没有给出证明。第一个证明由 Lagrange 在 1771 年给出，他同时证明了逆命题。

4.6　$x^2 \equiv -1$ 与 $4k+1$ 型素数

把 x^2+1 对 $x = 1, \dots, 10$ 分解：

$$2,\quad 5,\quad 2\cdot 5,\quad 17,\quad 2\cdot 13,\quad 37,\quad 2\cdot 5^2,\quad 5\cdot 13,\quad 2\cdot 41,\quad 101.$$

出现的素因子 $2, 5, 13, 17, 37, 41, 101$ 除 2 外都 $\equiv 1 \pmod 4$，而 $3, 7, 11, 19, 23, \dots$ 一个也没有出现。素数 p 整除某个 x^2+1，就是说同余式 $x^2 \equiv -1 \pmod p$ 有解。

> **定理 4.12**　设 p 是素数。$x^2 \equiv -1 \pmod p$ 有解当且仅当 $p = 2$ 或 $p \equiv 1 \pmod 4$。$p \equiv 1 \pmod 4$ 时，$x = \left(\frac{p-1}{2}\right)!$ 就是一个解。

思路　无解的方向用 Fermat 小定理：$x^{p-1} = (x^2)^{(p-1)/2}$，把 $x^2 \equiv -1$ 代进去，看 $(-1)^{(p-1)/2}$ 的符号。有解的方向用 Wilson 定理：把 $(p-1)!$ 的后一半 $p-1, p-2, \dots$ 写成 $-1, -2, \dots$，它就变成前一半的平方乘以一个符号。

证明　$p = 2$ 时 $x = 1$ 是解。下设 p 为奇素数，$h = \frac{p-1}{2}$。

$p \equiv 3 \pmod 4$ 时无解。设 $x^2 \equiv -1$，则 $p \nmid x$。$h = \frac{p-1}{2}$ 是奇数，由 Fermat 小定理

$$1 \equiv x^{p-1} = (x^2)^h \equiv (-1)^h = -1 \pmod p,$$

于是 $p \mid 2$，与 p 为奇素数矛盾。

$p \equiv 1 \pmod 4$ 时有解。此时 h 是偶数。对 $1 \le j \le h$，$p - j \equiv -j \pmod p$，所以

$$(p-1)! = \underbrace{1\cdot 2\cdots h}_{\text{前一半}} \cdot \underbrace{(p-h)\cdots(p-2)(p-1)}_{\text{后一半}} \equiv h!\cdot(-1)^h\, h! = (h!)^2 \pmod p.$$

由 Wilson 定理，$(h!)^2 \equiv -1 \pmod p$。 ■

例如 $p = 13$，$h = 6$，$6! = 720 = 55\cdot 13 + 5$，而 $5^2 = 25 \equiv -1 \pmod{13}$。

■ **例 4.7　模 29 的 -1 的平方根**　求 $x^2 \equiv -1 \pmod{29}$ 的全部解。$x^2 \equiv -1 \pmod{43}$ 呢？

解　$29 = 4\cdot 7 + 1$，$h = 14$。逐步乘、随时取余：

$$\begin{aligned} &4! = 24, && 5! \equiv 24\cdot 5 = 120 \equiv 4, && 6! \equiv 24, && 7! \equiv 168 \equiv 23,\\ &8! \equiv 184 \equiv 10, && 9! \equiv 90 \equiv 3, && 10! \equiv 30 \equiv 1, && 11! \equiv 11,\\ &12! \equiv 132 \equiv 16, && 13! \equiv 208 \equiv 5, && 14! \equiv 70 \equiv 12 && \pmod{29}. \end{aligned}$$

验证：$12^2 = 144 = 5\cdot 29 - 1 \equiv -1$。若 $x^2 \equiv -1 \equiv 12^2$，则 $29 \mid (x-12)(x+12)$，所以 $x \equiv \pm 12$。全部解是 $x \equiv 12, 17 \pmod{29}$。

$43 = 4\cdot 10 + 3$，由定理 4.12，$x^2 \equiv -1 \pmod{43}$ 无解。 ♦

第 2 讲用 $N = 4p_1\cdots p_n - 1$ 证明了 $4k+3$ 型素数有无穷多个，但同样的办法对 $4k+1$ 型行不通：$4k+1$ 型数的素因子不一定是 $4k+1$ 型（例如 $21 = 3\cdot 7$）。定理 4.12 补上了这个缺口：形如 x^2+1 的数，奇素因子一定是 $4k+1$ 型。

定理 4.13　形如 $4k+1$ 的素数有无穷多个。

证明　设 $p_1,\dots,p_n$ 是任意有限个 $4k+1$ 型素数，令 $N=(2p_1p_2\cdots p_n)^2+1$。$N>1$ 是奇数，取它的一个素因子 q，则 q 是奇数，且 $x=2p_1\cdots p_n$ 满足 $x^2\equiv -1 \pmod q$。由定理 4.12，$q\equiv 1 \pmod 4$。若 $q=p_i$，则 $q\mid x$，又 $q\mid x^2+1$，得 $q\mid 1$，矛盾。所以 q 是一个新的 $4k+1$ 型素数。任何有限个 $4k+1$ 型素数都不是全部，所以有无穷多个。 ■

例如从 $\{13,17\}$ 出发，$x=2\cdot 13\cdot 17=442$，$N=442^2+1=195365=5\cdot 41\cdot 953$，三个素因子都 $\equiv 1 \pmod 4$，也都是新的。

前后联系　定理 4.12 是第 9 讲二次剩余理论的第一个结论：用 Legendre 符号写，就是 $\left(\frac{-1}{p}\right)=(-1)^{(p-1)/2}$。它也是 Fermat 二平方定理「$p\equiv 1 \pmod 4$ 的素数是两个平方数之和」的出发点。更一般地，Dirichlet 定理断言：$\gcd(a,m)=1$ 时，形如 $mk+a$ 的素数有无穷多个，但一般情形的证明需要分析工具。

4.7　同余方程与一次同余式

4.7.1　同余方程的解

定义 4.14（同余方程）　设 $f(x)$ 是整系数多项式。若整数 x_0 满足 $f(x_0)\equiv 0 \pmod m$，则称 x_0 是同余方程 $f(x)\equiv 0 \pmod m$ 的**解**（solution）。

若 x_0 是解，$x_1\equiv x_0 \pmod m$，则 $f(x_1)\equiv f(x_0)\equiv 0$，所以 x_1 也是解。因此**解总是整个剩余类**。今后说「有几个解」，都指模 m 有几个不同的剩余类，也就是在一个完全剩余系中有几个解。

■ **例 4.8　几个同余方程的解数**　求下列同余方程的解数：(a) $x^p-x\equiv 0 \pmod p$（p 为素数）；(b) $x^2\equiv -1 \pmod 5$；(c) $x^2\equiv -1 \pmod{43}$；(d) $x^2\equiv 1 \pmod{15}$。

解　(a) 由 Fermat 小定理，每个整数都是解，共 p 个解，等于次数。

(b) 在 $\{0,1,2,3,4\}$ 中试：平方依次是 $0,1,4,9\equiv 4,16\equiv 1$，$-1\equiv 4$，所以解是 $x\equiv 2,3$，共 2 个。

(c) $43\equiv 3 \pmod 4$，由定理 4.12，没有解。

(d) 在 $0,\dots,14$ 中试，解是 $x\equiv 1,4,11,14$，即 $\pm 1,\pm 4$：$4^2=16=15+1$，$11^2=121=8\cdot 15+1$，$14^2\equiv(-1)^2$。共 4 个解，比次数 2 还多。 ♦

(d) 说明模合数时，多项式同余方程的解数可以超过次数。原因与引理 4.9 后面的说明相同：$15\mid(x-1)(x+1)$ 时，可以 $3\mid x-1$、$5\mid x+1$（得 $x=4$），或反过来（得 $x=11$）。模素数时，第 7 讲的 Lagrange 定理会证明：非零多项式模 p 的解数不超过次数。

4.7.2　一次同余式

现在完全解决最简单的同余方程 $ax \equiv b \pmod m$。它等价于线性丢番图方程 $ax - my = b$，所以第 1 讲的 Bézout 定理是主要工具。

定理 4.15（一次同余式）　设 $m \geq 1$，$g = \gcd(a, m)$。同余式 $ax \equiv b \pmod m$ 有解当且仅当 $g \mid b$。若有解，则模 m 恰有 g 个解：设 x_1 是一个解，全部解为

$$x \equiv x_1,\ x_1 + \frac{m}{g},\ x_1 + \frac{2m}{g},\ \ldots,\ x_1 + \frac{(g-1)m}{g} \pmod m.$$

证明　*必要性*。若 $ax - b = my$，则 $g \mid a$，$g \mid m$，所以 $g \mid ax - my = b$。

充分性。由 Bézout 定理，有 u, v 使 $au + mv = g$。若 $b = gc$，则 $a(uc) + m(vc) = b$，所以 $x_1 = uc$ 是解。

全部的解。设 x 是任意一个解，则 $a(x - x_1) \equiv b - b = 0 \pmod m$，即 $m \mid a(x - x_1)$。两边除以 g：$\frac{m}{g} \mid \frac{a}{g}(x - x_1)$。而 $\gcd\left(\frac{a}{g}, \frac{m}{g}\right) = 1$（第 1 讲练习），由 Euclid 引理的一般形式，$\frac{m}{g} \mid x - x_1$。反过来，$x = x_1 + t\frac{m}{g}$ 时，$a(x - x_1) = \frac{a}{g}\, t\, m \equiv 0$，也是解。所以解集恰好是 $x \equiv x_1 \pmod{m/g}$。

数个数。模 m/g 的一个剩余类，在模 m 下分裂成 $x_1 + t\frac{m}{g}$（$t = 0, 1, \ldots, g-1$）这 g 个类。它们两两不同余：$t \neq t'$ 时 $0 < |t - t'| < g$，差 $|t - t'|\frac{m}{g}$ 介于 0 和 m 之间，不被 m 整除。任何 $x_1 + t\frac{m}{g}$ 都与其中之一同余（把 t 换成模 g 的余数）。 ■

证明中「除以 g」的一步值得单独记下：有解时，

$$ax \equiv b \pmod m \iff \frac{a}{g}x \equiv \frac{b}{g} \pmod{\frac{m}{g}}.$$

右边 $\gcd\left(\frac{a}{g}, \frac{m}{g}\right) = 1$，用逆元就能解出唯一的 $x \bmod \frac{m}{g}$。

■ **例 4.9　有解、无解与解数**　解下列同余式：(a) $6x \equiv 9 \pmod{15}$；(b) $6x \equiv 7 \pmod{15}$；(c) $12x \equiv 18 \pmod{30}$。

解　(a) $g = \gcd(6, 15) = 3$，$3 \mid 9$，有 3 个解。除以 3：$2x \equiv 3 \pmod 5$。$2^{-1} \equiv 3 \pmod 5$，所以 $x \equiv 3 \cdot 3 = 9 \equiv 4 \pmod 5$。模 15 的解是 $x \equiv 4, 9, 14$。验证：$6 \cdot 4 = 24 \equiv 9$，$6 \cdot 9 = 54 = 3 \cdot 15 + 9$，$6 \cdot 14 = 84 = 5 \cdot 15 + 9$。

(b) $g = 3 \nmid 7$，无解。直观地看，$6x - 15y$ 总是 3 的倍数，不可能等于 7。

(c) $g = \gcd(12, 30) = 6$，$6 \mid 18$，有 6 个解。除以 6：$2x \equiv 3 \pmod 5$，同 (a)，$x \equiv 4 \pmod 5$。模 30 的解是 $x \equiv 4, 9, 14, 19, 24, 29$。 ♦

图 4.4 将第一问的三个解画在模 15 的圆周上。

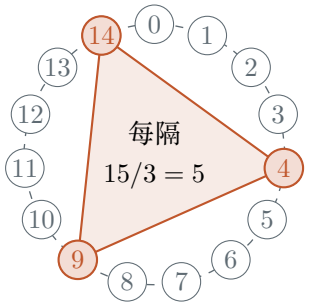


图 4.4　$6x \equiv 9 \pmod{15}$ 的三个解 $4, 9, 14$ 在圆周上等距分布，相邻两解相差 $m/g = 5$。

易错提醒　说「几个解」时要看清是模多少。(a) 和 (c) 化简后都是 $x \equiv 4 \pmod 5$，模 5 只有一个解，但原题问的是模 15 和模 30，分别有 3 个和 6 个解。不要把化简后的模当成原来的模。

想一想　固定 a 和 m，记 $g = \gcd(a, m)$。b 跑遍模 m 的完全剩余系时，有多少个 b 使 $ax \equiv b \pmod m$ 有解？所有方程的解数加起来是多少？为什么一定是这个数？

答　有解当且仅当 $g \mid b$，模 m 的 g 的倍数有 $\frac{m}{g}$ 个，所以有 $\frac{m}{g}$ 个 b，每个恰有 g 个解，共 $\frac{m}{g} \cdot g = m$ 个解。这是必然的：每个 $x \bmod m$ 恰好是一个方程（即 $b \equiv ax$ 的那个）的解，所以总解数等于 x 的个数 m。

$\gcd(a, m) = 1$ 时解唯一，就是 $x \equiv a^{-1}b$（如例 4.5）。第 5 讲会用中国剩余定理处理多个模的同余式组。

本讲小结

- 完全剩余系有 m 个数，简化剩余系有 $\varphi(m)$ 个数；个数对时，只需检查两两不同余。
- 乘以与 m 互素的 a 只是重排简化剩余系；比较乘积得 Euler 定理，特例是 Fermat 小定理。
- $\gcd(a, m) = 1$ 时，a^n 的指数可以模 $\varphi(m)$ 来算；底数与模不互素时，把模拆成互素的部分分别处理，再用第 5 讲的中国剩余定理合并。Fermat 小定理的逆命题不成立（341）。
- a 模 m 有逆元 $\iff \gcd(a, m) = 1$，逆元唯一，可用扩展 Euclid 算法或 $a^{\varphi(m)-1}$ 求出。
- 模素数时 $x^2 \equiv 1$ 只有 ± 1。由此配对得 Wilson 定理 $(p-1)! \equiv -1$，其逆也成立。
- $x^2 \equiv -1 \pmod p$ 有解 $\iff p = 2$ 或 $p \equiv 1 \pmod 4$；$p = 2$ 时解为 1，奇素数有解时解为 $\pm\left(\frac{p-1}{2}\right)!$；由此 $4k+1$ 型素数有无穷多个。
- $ax \equiv b \pmod m$ 有解 $\iff g = \gcd(a, m) \mid b$，此时模 m 恰有 g 个解，相邻两解相差 m/g。模合数时，高次同余式的解数可以超过次数。

练习

4.1 ★　写出模 14 和模 16 的简化剩余系，求 $\varphi(14)$、$\varphi(16)$，并直接验证 $3^{\varphi(14)} \equiv 1 \pmod{14}$

和 $3^{\varphi(16)} \equiv 1 \pmod{16}$。

4.2 ★ 求 2^{1000} 除以 13 的余数，以及 7^{2024} 的末两位数字。

4.3 ★ 写出模 11 的逆元表，把 $2,3,\dots,9$ 配成乘积 $\equiv 1$ 的四对，由此验证 $10! \equiv -1 \pmod{11}$。

4.4 ★★ 解同余式：(a) $9x \equiv 21 \pmod{30}$；(b) $10x \equiv 4 \pmod{25}$；(c) $17x \equiv 3 \pmod{43}$（用扩展 Euclid 算法求 17^{-1}）。

4.5 ★★ 设 p 是奇素数。证明 $(p-2)! \equiv 1 \pmod{p}$，并求 $97!$ 除以 101 的余数。

4.6 ★★ 证明：对每个整数 n，$2730 \mid n^{13} - n$。（提示：$2730 = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 13$。）

4.7 ★★ 设 $n > 4$ 是合数。证明 $(n-1)! \equiv 0 \pmod{n}$。$n = 4$ 时为什么不成立？

4.8 ★★ （习题集 2 第 6 题）设 p 是素数，$e \geq 1$，求 $x^2 \equiv 1 \pmod{p^e}$ 的全部解。特别地，求模 16 的解，并说明奇素数情形的证明在 $p = 2$ 时哪里失效。

4.9 ★★★ 设 p 是奇素数。证明 $1^2 \cdot 3^2 \cdot 5^2 \cdots (p-2)^2 \equiv (-1)^{(p+1)/2} \pmod{p}$。

4.10 ★★ （习题集 2 第 4 题）(a) 设 p 是素数。证明对每个整数 a，$(x-a)^p \equiv x^p - a \pmod{p}$ 作为多项式成立。(b) 设 $n > 1$ 是合数。证明 $(x-a)^n \equiv x^n - a \pmod{n}$ 不能对每个整数 a 都成立。对 $n = 6$、$a = -1$，指出哪些系数不是 6 的倍数。

练习答案与提示

4.1 模 14 的简化剩余系有 6 个元，且 $3^3 \equiv -1$，故 $3^6 \equiv 1$；模 16 有 8 个奇数，且 $3^4 = 81 \equiv 1$，故 $3^8 \equiv 1$。

4.2 $1000 \equiv 4 \pmod{12}$，故 $2^{1000} \equiv 2^4 \equiv 3 \pmod{13}$；又 $7^4 \equiv 1 \pmod{100}$、$4 \mid 2024$，所以末两位为 01。

4.3 $2 \cdot 6 \equiv 3 \cdot 4 \equiv 5 \cdot 9 \equiv 7 \cdot 8 \equiv 1 \pmod{11}$；固定元 $1, 10$ 的乘积为 -1，故 $10! \equiv -1 \pmod{11}$。

4.4 (a) $g = 3$，化为 $3x \equiv 7 \pmod{10}$，得 $x \equiv 9, 19, 29 \pmod{30}$。(b) 无解，因为 $5 \nmid 4$。(c) $17^{-1} \equiv 38 \pmod{43}$，故 $x \equiv 28$；$17 \cdot 28 = 11 \cdot 43 + 3$。

4.5 由 Wilson 定理，$(p-1)! = (p-1)(p-2)! \equiv -1$，故 $(p-2)! \equiv 1$。取 $p = 101$，$99! \equiv 1$ 且 $99! = 97! \cdot 98 \cdot 99 \equiv 6 \cdot 97!$，所以 $97! \equiv 17 \pmod{101}$。

4.6 设 $p \in \{2, 3, 5, 7, 13\}$，则 $p - 1 \in \{1, 2, 4, 6, 12\}$ 整除 12，记 $12 = (p-1)t$。若 $p \mid n$，则 $p \mid n^{13} - n = n(n^{12} - 1)$。若 $p \nmid n$，由 Fermat 小定理 $n^{12} = (n^{p-1})^t \equiv 1 \pmod{p}$，同样 $p \mid n(n^{12} - 1)$。五个素数两两互素，由第 1 讲「$a \mid N$，$b \mid N$，$\gcd(a, b) = 1 \Rightarrow ab \mid N$」逐个合并，得 $2730 \mid n^{13} - n$。例如 $n = 2$：$2^{13} - 2 = 8190 = 3 \cdot 2730$。

4.7 若 $n = ab$，$1 < a < b < n$，则 a, b 是 $1, \dots, n-1$ 中两个不同的数，$ab = n$ 整除 $(n-1)!$。否则只能 $n = q^2$，q 为素数；$n > 4$ 时 $q \geq 3$，$q < 2q < q^2 = n$，所以 q 与 $2q$ 是 $(n-1)!$ 中两个不同的因子，$2q^2 \mid (n-1)!$，更有 $n \mid (n-1)!$。$n = 4$ 时 $2q = 4 = n$，而 $3! = 6 \equiv 2 \pmod{4}$。

4.8 先设 p 为奇素数。由 $p^e \mid (x-1)(x+1)$，若 p 同时整除两个因子便有 $p \mid 2$，矛盾。因此 p^e 整除其中一个因子，解恰为 $x \equiv \pm 1 \pmod{p^e}$。当 $p = 2$：$e = 1$ 时解为 1；$e = 2$ 时解为 $1, 3$。若 $e \geq 3$，解必为奇数，写 $x = 2y + 1$。条件变为 $2^{e-2} \mid y(y+1)$。因 $\gcd(y, y+1) = 1$，所以 $y \equiv 0$ 或 $-1 \pmod{2^{e-2}}$。按模 2^{e-1} 计，这给出 $y \equiv 0, 2^{e-2}, -1, 2^{e-2} - 1$ 四类；对应

$$x \equiv 1, \quad 1 + 2^{e-1}, \quad -1, \quad 2^{e-1} - 1 \pmod{2^e}.$$

直接代入可验四类都是解，且彼此不同。取 $e = 4$，模 16 的解为 $1, 7, 9, 15$。奇素数证明失效在「p

不能同时整除 $x-1$ 与 $x+1$」这一步：$p=2$ 时两因子可以同时为偶数，且 2^e 可分摊在两因子中，例如 $x=7$ 时 $6\cdot 8=48$。

4.9 把 $(p-1)!$ 中的偶数 j 写成 $j=p-k$，$k=p-j$ 是奇数，且 j 跑遍 $2,4,\dots,p-1$ 时 k 跑遍 $p-2,\dots,3,1$。于是

$$(p-1)!=\prod_{k\text{ 奇},\,k\le p-2} k\cdot(p-k)\equiv\prod_{k\text{ 奇}} k\cdot(-k)=(-1)^{(p-1)/2}\cdot 1^2\cdot 3^2\cdots(p-2)^2 \pmod p.$$

由 Wilson 定理左边 $\equiv -1$，所以 $1^2\cdot 3^2\cdots(p-2)^2\equiv -(-1)^{(p-1)/2}=(-1)^{(p+1)/2}$。验证 $p=7$：$1\cdot 9\cdot 25=225=32\cdot 7+1$，而 $(-1)^4=1$。

4.10 (a) 由引理 4.7 的证明，$(x-a)^p\equiv x^p+(-a)^p$ 作为多项式成立。p 为奇数时 $(-a)^p=-a^p\equiv -a$；$p=2$ 时 $a^2\equiv a\equiv -a \pmod 2$。所以 $(x-a)^p\equiv x^p-a$。(b) 取 $a=-1$。若 $(x+1)^n\equiv x^n+1 \pmod n$，则 n 整除 $\binom{n}{1},\dots,\binom{n}{n-1}$ 全体，由第 3 讲「素数的一个刻画」那个例子，n 是素数，矛盾。$n=6$ 时 $(x+1)^6=x^6+6x^5+15x^4+20x^3+15x^2+6x+1$，$x^4,x^3,x^2$ 的系数 $15,20,15$ 不是 6 的倍数（参看表 4.3）。

第 5 讲

一次同余式、中国剩余定理与算法

Linear Congruences, Chinese Remainder Theorem, Algorithms

对应材料：MIT 18.781 Lecture 5 讲义（Linear Congruences, Chinese Remainder Theorem, Algorithms）；习题集 3 第 2、3、5 题（末两位数字、$x^2 \equiv x$ 的解数、连续的非平方自由整数）。

本讲目标

(1) 会用扩展欧几里得算法解一次同余式 $ax \equiv b \pmod m$，包括 $\gcd(a,m) > 1$ 时写出模 m 的全部 $\gcd(a,m)$ 个解。

(2) 理解中国剩余定理的陈述与构造性证明，会用「公式法」和「逐次代入法」两种办法解两两互素模数的同余式组。

(3) 会判断模数不互素的同余式组是否有解：$x \equiv a_1 \pmod{m_1}$，$x \equiv a_2 \pmod{m_2}$ 有解当且仅当 $\gcd(m_1,m_2) \mid a_1 - a_2$，且解模 $\operatorname{lcm}(m_1,m_2)$ 唯一。

(4) 把中国剩余定理看成模 M 与各个模 m_i 之间的一一对应，由此证明 Euler 函数的积性，并会用公式 $\varphi(n) = n\prod_{p\mid n}(1-1/p)$ 计算。

(5) 知道衡量数论算法快慢的标准（位数的多项式），掌握 Karatsuba 乘法的想法和计算 $a^b \bmod N$ 的反复平方法。

第 4 讲证明了一次同余式的可解性定理：$ax \equiv b \pmod m$ 有解当且仅当 $\gcd(a,m) \mid b$，有解时模 m 恰有 $\gcd(a,m)$ 个解。那个证明告诉我们解**存在**，本讲先把它变成一套能一步步执行的**求解步骤**。

接下来处理一个更有意思的问题：一个未知数**同时**满足几个模数不同的同余式。一千多年前的《孙子算经》就问过：一个数除以 3 余 2，除以 5 余 3，除以 7 余 2，这个数是多少？回答这类问题的定理叫**中国剩余定理**（Chinese Remainder Theorem），简称 CRT。它是本课程用得最多的工具之一：本讲用它算 Euler 函数，第 6 讲用它把模 n 的多项式同余式拆成模素数幂来解，后面讲原根、二次剩余时也一再用到。

最后一节换一个角度：这些计算在计算机上要花多少时间？我们会看到，求逆元、解同余式组、计算 $a^b \bmod N$ 都能在「位数的多项式」步内完成，即使 N 有几百位也不在话下。

本讲用 $\gcd(a,m)$ 表示最大公因数，不用简写 (a,m)。原因是后面要大量使用有序数组，例如余数组 $(a_1,\dots,a_k)$，两种记号容易混淆。

5.1　一次同余式的解法

5.1.1　回顾：可解性与逆元

> **定理 5.1（一次同余式的可解性，第 4 讲）**　设 $m \geq 1$，$g = \gcd(a, m)$。同余式 $ax \equiv b \pmod{m}$ 有解当且仅当 $g \mid b$。有解时，它模 m 恰有 g 个解。

第 4 讲的证明是这样造出解的：由 Bézout 定理，有整数 x_0, y_0 使 $ax_0 + my_0 = g$；若 $b = b'g$，两边乘 b' 得 $a(b'x_0) + m(b'y_0) = b$，于是 $a(b'x_0) \equiv b \pmod{m}$，$x = b'x_0$ 就是一个解。所以**求解的全部工作在于求出 Bézout 系数 x_0**，而这正是第 2 讲扩展欧几里得算法做的事。

特别地，当 $\gcd(a, m) = 1$ 时，同余式 $ax \equiv 1 \pmod{m}$ 模 m 有唯一解，记作 a^{-1}，称为 a 模 m 的**逆元**（inverse）（见第 4 讲）。例如 $3 \cdot 5 = 15 \equiv 1 \pmod{7}$，所以 $3^{-1} \equiv 5 \pmod{7}$。逆元让我们能在模 m 下「做除法」。

5.1.2　情形一：$\gcd(a, m) = 1$

这时解唯一，就是 $x \equiv a^{-1}b \pmod{m}$。求逆元只需对 (a, m) 做一次扩展欧几里得算法。

例 5.1　求逆元解同余式　解 $27x \equiv 5 \pmod{43}$。

解　43 是素数，$\gcd(27, 43) = 1$，所以解唯一。先求 $27^{-1} \bmod 43$。左栏做欧几里得算法，右栏从倒数第二行起把余数 1 逐步回代：

$$
\begin{aligned}
43 &= 1 \cdot 27 + 16, & 1 &= 11 - 2 \cdot 5 \\
27 &= 1 \cdot 16 + 11, & &= 11 - 2(16 - 11) = 3 \cdot 11 - 2 \cdot 16 \\
16 &= 1 \cdot 11 + 5, & &= 3(27 - 16) - 2 \cdot 16 = 3 \cdot 27 - 5 \cdot 16 \\
11 &= 2 \cdot 5 + 1, & &= 3 \cdot 27 - 5(43 - 27) = 8 \cdot 27 - 5 \cdot 43.
\end{aligned}
$$

于是 $8 \cdot 27 - 5 \cdot 43 = 1$，两边模 43 得 $27 \cdot 8 \equiv 1 \pmod{43}$，即 $27^{-1} \equiv 8$。这和第 2 讲用表格法得到的 $43 \cdot (-5) + 27 \cdot 8 = 1$ 一致。原同余式两边乘 8：

$$
x \equiv 8 \cdot 27x \equiv 8 \cdot 5 = 40 \pmod{43}.
$$

检验：$27 \cdot 40 = 1080 = 25 \cdot 43 + 5$，余数确实是 5。 ♦

注意解题时用的是「两边同乘 a^{-1}」，而不是「两边同除以 a」。模 43 下 $\frac{5}{27}$ 没有意义，$5 \cdot 27^{-1}$ 才有意义。

5.1.3　情形二：$\gcd(a, m) = g > 1$

若 $g \nmid b$，无解。若 $g \mid b$，写 $a = ga'$，$b = gb'$，$m = gm'$。关键的一步是：

$$
\begin{aligned}
ax \equiv b \pmod{m} &\iff gm' \mid g(a'x - b') \\
&\iff m' \mid a'x - b' \iff a'x \equiv b' \pmod{m'}.
\end{aligned}
$$

也就是说，a、b、m 三者同时除以 g，得到一个等价的同余式。由第 1 讲的习题，$\gcd(a',m') = \gcd(a/g,\, m/g) = 1$，新同余式落入情形一，模 m' 有唯一解 $x \equiv x_1 \pmod{m'}$。

回到模 m 看，一个模 m' 的剩余类分裂成 g 个模 m 的剩余类：

$$x_1,\quad x_1 + m',\quad x_1 + 2m',\quad \ldots,\quad x_1 + (g-1)m'.$$

它们两两模 m 不同余，因为任意两个之差的绝对值在 m' 和 $(g-1)m' < m$ 之间；而任何满足 $x \equiv x_1 \pmod{m'}$ 的整数都模 m 同余于其中之一。这就是「恰有 g 个解」的来历。

算法 5.2（解一次同余式）　输入整数 a, b 和模 $m \geq 2$，输出 $ax \equiv b \pmod{m}$ 模 m 的全部解，或判定无解。

(1) 用欧几里得算法求 $g = \gcd(a, m)$。若 $g \nmid b$，输出「无解」，停止。

(2) 约化：$a' = a/g$，$b' = b/g$，$m' = m/g$。此时 $\gcd(a', m') = 1$。若 $m' = 1$，原同余式对全部整数成立，直接输出 $0, 1, \ldots, m-1$ 并停止。

(3) 用扩展欧几里得算法求 a' 模 m' 的逆元 u，令 $x_1 \equiv ub' \pmod{m'}$，$0 \leq x_1 < m'$。

(4) 输出 $x_1 + jm'$，$j = 0, 1, \ldots, g-1$。

■ **例 5.2　解不唯一的情形**　解 $35x \equiv 14 \pmod{28}$。

解　第 1 步：$35 = 1 \cdot 28 + 7$，$28 = 4 \cdot 7$，所以 $g = 7$。$7 \mid 14$，有解，且模 28 恰有 7 个解。

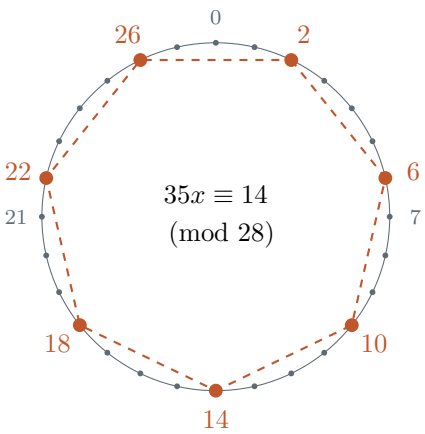
第 2 步：三者同除以 7，得 $5x \equiv 2 \pmod{4}$。

第 3 步：$5 \equiv 1 \pmod{4}$，逆元就是 1，所以 $x_1 = 2$。

第 4 步：$m' = 4$，全部解为

$$x \equiv 2,\ 6,\ 10,\ 14,\ 18,\ 22,\ 26 \pmod{28}.$$

抽查一个：$35 \cdot 6 = 210 = 7 \cdot 28 + 14$。图 5.1 把这 7 个解画在模 28 的「钟面」上。 ♦



约化为 $5x \equiv 2 \pmod{4}$，
即 $x \equiv 2 \pmod{4}$。

模 4 的一个剩余类
= 模 28 的 7 个剩余类，
相邻两个解相差 $m' = 28/7 = 4$。

图 5.1　$35x \equiv 14 \pmod{28}$ 的 7 个解在模 28 的钟面上均匀分布，构成正七边形。

■ **例 5.3　两轮欧几里得算法**　解 $60x \equiv 36 \pmod{228}$；再判断 $60x \equiv 30 \pmod{228}$ 是否有解。

解　求 $\gcd(60,228)$：$228=3\cdot60+48$，$60=1\cdot48+12$，$48=4\cdot12$，所以 $g=12$。$12\mid36$，有解，模 228 恰有 12 个解。

约化：$a'=5$，$b'=3$，$m'=19$，得 $5x\equiv3\pmod{19}$。再求 $5^{-1}\bmod19$：$19=3\cdot5+4$，$5=1\cdot4+1$，回代得

$$1=5-4=5-(19-3\cdot5)=4\cdot5-19,$$

所以 $5^{-1}\equiv4\pmod{19}$，$x_1\equiv4\cdot3=12\pmod{19}$。全部解为

$$x\equiv12+19j\pmod{228},\qquad j=0,1,\dots,11,$$

即 $12,31,50,69,88,107,126,145,164,183,202,221$。检验：$60\cdot12=720=3\cdot228+36$。

至于 $60x\equiv30\pmod{228}$：$12\nmid30$，无解。原因是 $60x-228y$ 总是 12 的倍数，不可能等于 30。 ♦

易错提醒　约化时 a、b、m 必须**同时**除以 g。只约两边而不约模数是常见错误。例如 $6x\equiv4\pmod{10}$ 的解是 $x\equiv4,9\pmod{10}$（$g=2$，约化为 $3x\equiv2\pmod5$，$x\equiv4\pmod5$）。若只把两边除以 2 写成 $3x\equiv2\pmod{10}$，就只剩 $x\equiv4$，丢掉了 $x=9$：$6\cdot9=54\equiv4\pmod{10}$，但 $3\cdot9=27\equiv7\pmod{10}$。第 3 讲说过，消去律 $ax\equiv ay\Rightarrow x\equiv y$ 要求 $\gcd(a,m)=1$，这里 $\gcd(2,10)=2$，不能直接消去。

想一想　为什么解的个数恰好是 g？换个角度：列出 $x=0,1,\dots,27$ 时 $35x \bmod 28$ 的值，它们有哪些、各出现几次？

答　因为 $35\equiv7\pmod{28}$，$35x \bmod 28=7x \bmod 28$，依次是 $0,7,14,21,0,7,14,21,\dots$，只取 $0,7,14,21$ 四个值，每个恰好出现 7 次。所以 $35x\equiv b\pmod{28}$ 当 $b\in\{0,7,14,21\}$（即 $7\mid b$）时恰有 7 个解，否则无解。一般地，映射 $x\mapsto ax \bmod m$ 把 m 个剩余「g 对 1」地映到 g 的倍数 $0,g,2g,\dots,m-g$ 这 m/g 个值上。可解性条件和解数都从这里读出来。

5.2　同余式组：从「物不知数」说起

《孙子算经》（成书于公元 3 至 5 世纪之间）卷下第 26 题：

> 今有物不知其数，三三数之剩二，五五数之剩三，七七数之剩二。问物几何？

用同余的语言，就是求同时满足

$$x\equiv2\pmod3,\qquad x\equiv3\pmod5,\qquad x\equiv2\pmod7$$

的整数 x。这样一组同余式叫**同余式组**（system of congruences）。与上一节不同，这里每个同余式都很简单，难点在于**模数各不相同**。

5.2.1　筛选法

最朴素的办法是从模数最大的同余式出发逐个筛选。满足 $x\equiv2\pmod7$ 的非负整数是

$$2,\ 9,\ 16,\ 23,\ 30,\ 37,\ 44,\ 51,\ 58,\ 65,\ 72,\ 79,\ 86,\ 93,\ 100,\ \dots$$

其中除以 5 余 3 的是 $23, 58, 93, \ldots$（每隔 35 出现一个）；再看除以 3 的余数：$23 \equiv 2, 58 \equiv 1, 93 \equiv 0$。所以 23 满足全部三个条件，而且在 $0 \le x < 105$ 中只有它一个。下一个解是 $23 + 105 = 128$，因为 $105 = 3 \cdot 5 \cdot 7$ 是三个模数的公倍数，加上它不改变任何一个余数。这个答案与《孙子算经》的「答曰：二十三」相同。

筛选法容易理解，但要逐个检查的数多达 $m_2 m_3$ 量级。模数有几十位时这个办法毫无用处。

5.2.2 并非总有解

■ **例 5.4　矛盾的同余式组**　同余式组 $x \equiv 3 \pmod 8$，$x \equiv 1 \pmod{12}$ 是否有解？

解　无解。4 同时整除 8 和 12。由 $x \equiv 3 \pmod 8$ 得 $x \equiv 3 \pmod 4$，由 $x \equiv 1 \pmod{12}$ 得 $x \equiv 1 \pmod 4$，两者矛盾。 ♦

两个模数有公因数 d 时，两个同余式各自决定了 $x \bmod d$，这两个「答案」必须相同，否则无解。若模数两两互素，就不会有这种冲突，这正是中国剩余定理的条件。

5.2.3 逐次代入法

更系统的办法是把同余式一个一个地代入，每一步只解一个一次同余式。

■ **例 5.5　逐次代入解「物不知数」**　用逐次代入法解 $x \equiv 2 \pmod 3$，$x \equiv 3 \pmod 5$，$x \equiv 2 \pmod 7$。

解　第一个同余式等价于 $x = 2 + 3t$，$t \in \mathbb{Z}$。代入第二个：

$$2 + 3t \equiv 3 \pmod 5 \iff 3t \equiv 1 \pmod 5 \iff t \equiv 2 \pmod 5,$$

因为 $3 \cdot 2 = 6 \equiv 1$。所以 $t = 2 + 5s$，$x = 2 + 3(2 + 5s) = 8 + 15s$。代入第三个：

$$8 + 15s \equiv 2 \pmod 7 \iff 1 + s \equiv 2 \pmod 7 \iff s \equiv 1 \pmod 7,$$

这里用了 $8 \equiv 1$、$15 \equiv 1 \pmod 7$。所以 $s = 1 + 7r$，$x = 8 + 15(1 + 7r) = 23 + 105r$，即 $x \equiv 23 \pmod{105}$。 ♦

每一步都形如 $(\text{前面模数之积}) \cdot t \equiv c \pmod{m_i}$。如果 m_i 与前面的模数都互素，这个系数就可逆，于是每一步都恰有一个解。这已经包含了下一节定理的证明思路。

5.3　中国剩余定理

定理 5.3（中国剩余定理）　设 $m_1,\dots,m_k$ 是两两互素的正整数，$M=m_1m_2\cdots m_k$。则对任意整数 $a_1,\dots,a_k$，同余式组

$$x\equiv a_1 \pmod{m_1},\quad x\equiv a_2 \pmod{m_2},\quad \dots,\quad x\equiv a_k \pmod{m_k}$$

有解，且解模 M 唯一：若 x,y 都是解，则 $x\equiv y \pmod M$；反过来，与某个解模 M 同余的整数都是解。

思路　唯一性靠整除：两个解之差被每个 m_i 整除，互素保证它被乘积整除。存在性用「叠加」：先对每个 i 造一个数 e_i，使它模 m_i 余 1、模其余各 m_j 余 0；那么 $a_1e_1+\cdots+a_ke_k$ 就同时满足所有同余式。

证明　*唯一性*。若 x,y 都是解，则对每个 i 有 $x\equiv a_i\equiv y \pmod{m_i}$，即 $m_i\mid x-y$。第 1 讲证明过：若 $\gcd(a,b)=1$，$a\mid n$，$b\mid n$，则 $ab\mid n$。由 $m_1\mid x-y$、$m_2\mid x-y$ 得 $m_1m_2\mid x-y$。m_3 与 m_1、m_2 都互素，所以与 m_1m_2 互素（第 1 讲的乘法封闭性），于是 $m_1m_2m_3\mid x-y$。依此归纳，$M\mid x-y$。反过来，若 $y\equiv x \pmod M$，则 $y\equiv x \pmod{m_i}$ 对每个 i 成立，y 也是解。

存在性。令 $N_i=M/m_i$，它是除 m_i 以外所有模数的乘积。因为每个 m_j（$j\neq i$）都与 m_i 互素，所以 $\gcd(N_i,m_i)=1$，N_i 模 m_i 有逆元，记 $H_i\equiv N_i^{-1} \pmod{m_i}$。令

$$e_i=H_iN_i,\qquad x=\sum_{i=1}^{k}a_ie_i=a_1H_1N_1+a_2H_2N_2+\cdots+a_kH_kN_k.$$

固定 j。对 $i\neq j$，$m_j\mid N_i$，所以 $e_i\equiv 0 \pmod{m_j}$；而 $e_j=H_jN_j\equiv 1 \pmod{m_j}$。于是

$$x\equiv a_j\cdot 1+\sum_{i\neq j}a_i\cdot 0=a_j \pmod{m_j},$$

对每个 j 成立。 ■

由证明可以提炼出一个公式。

算法 5.4（中国剩余定理的公式解法）　输入两两互素的模 $m_1,\dots,m_k$ 和余数 $a_1,\dots,a_k$，输出模 $M=\prod m_i$ 的唯一解。

(1) 对每个 i，计算 $N_i=M/m_i$，再用扩展欧几里得算法求 $H_i\equiv N_i^{-1} \pmod{m_i}$。（为节省计算，可先把 N_i 模 m_i 约化。）

(2) 输出 $x\equiv\sum_i a_iH_iN_i \pmod M$ 在 $0,1,\dots,M-1$ 中的代表。

$$\boxed{x\equiv a_1H_1N_1+a_2H_2N_2+\cdots+a_kH_kN_k \pmod M,\qquad N_i=\frac{M}{m_i},\quad H_iN_i\equiv 1 \pmod{m_i}.}$$

注意 $e_i=H_iN_i$ 只依赖于模数，不依赖于余数 a_i。模数固定时，算好一次 $e_1,\dots,e_k$，以后对任何余数都只需做一次加权求和。

■ **例 5.6　公式法**　解 $x \equiv 2 \pmod 3$，$x \equiv 3 \pmod 5$，$x \equiv 5 \pmod 7$。

解　$M = 105$。逐个计算 N_i、$N_i \bmod m_i$ 和逆元 H_i：

i	m_i	$N_i = M/m_i$	$N_i \bmod m_i$	$H_i = N_i^{-1} \bmod m_i$	$e_i = H_i N_i$
1	3	35	2	2	70
2	5	21	1	1	21
3	7	15	1	1	15

例如 H_1：$35 \equiv 2 \pmod 3$，而 $2 \cdot 2 = 4 \equiv 1 \pmod 3$，所以 $H_1 = 2$。于是

$$x \equiv 2 \cdot 70 + 3 \cdot 21 + 5 \cdot 15 = 140 + 63 + 75 = 278 \equiv 68 \pmod{105}.$$

检验：$68 = 22 \cdot 3 + 2 = 13 \cdot 5 + 3 = 9 \cdot 7 + 5$。

同一组模数下，「物不知数」的余数是 $(2,3,2)$，答案是 $2 \cdot 70 + 3 \cdot 21 + 2 \cdot 15 = 233 \equiv 23 \pmod{105}$，与上一节一致。 ♦

数 $70, 21, 15$ 有清楚的含义：70 模 3 余 1、被 5 和 7 整除；21 模 5 余 1、被 3 和 7 整除；15 模 7 余 1、被 3 和 5 整除。它们分别「只负责」一个模数。

直观理解　把模 105 的整数 x 看成一个「坐标向量」$(x \bmod 3,\ x \bmod 5,\ x \bmod 7)$。那么

$$70 \leftrightarrow (1,0,0), \qquad 21 \leftrightarrow (0,1,0), \qquad 15 \leftrightarrow (0,0,1),$$

它们就像线性代数中的标准基 e_1, e_2, e_3。要得到坐标为 (a_1, a_2, a_3) 的数，只需取线性组合 $a_1 \cdot 70 + a_2 \cdot 21 + a_3 \cdot 15$。中国剩余定理的公式就是「用基向量展开」。

■ **例 5.7　连续三个数都有平方因子**　求最小的正整数 n，使 $4 \mid n$，$9 \mid n+1$，$25 \mid n+2$。

解　条件即 $n \equiv 0 \pmod 4$，$n \equiv -1 \pmod 9$，$n \equiv -2 \pmod{25}$。$4, 9, 25$ 两两互素，$M = 900$。

- $N_1 = 225 \equiv 1 \pmod 4$，$H_1 = 1$。
- $N_2 = 100 \equiv 1 \pmod 9$，$H_2 = 1$。
- $N_3 = 36 \equiv 11 \pmod{25}$。求 $11^{-1} \bmod 25$：$25 = 2 \cdot 11 + 3$，$11 = 3 \cdot 3 + 2$，$3 = 1 \cdot 2 + 1$，回代得 $1 = 3 - 2 = 3 - (11 - 3 \cdot 3) = 4 \cdot 3 - 11 = 4(25 - 2 \cdot 11) - 11 = 4 \cdot 25 - 9 \cdot 11$，所以 $11^{-1} \equiv -9 \equiv 16 \pmod{25}$，$H_3 = 16$。

于是

$$n \equiv 0 \cdot 225 + (-1) \cdot 100 \cdot 1 + (-2) \cdot 36 \cdot 16 = -100 - 1152 = -1252 \equiv 548 \pmod{900}.$$

检验：$548 = 4 \cdot 137$，$549 = 9 \cdot 61$，$550 = 25 \cdot 22$。所以 $548, 549, 550$ 三个连续整数都含有大于 1 的平方因子。 ♦

把 $4, 9, 25$ 换成任意多个不同素数的平方 $p_1^2, \dots, p_N^2$，同样的论证给出任意长的一串连续整数，其中每一个都有平方因子（习题 9）。中国剩余定理常常这样用来**同时安排一个数对多个模数的余数**。

5.3.1 两个模数的情形

只有两个模数时有一个更紧凑的写法。设 $\gcd(m_1, m_2) = 1$，由扩展欧几里得算法得 $m_1 u + m_2 v = 1$。那么 $m_2 v \equiv 1 \pmod{m_1}$、$m_2 v \equiv 0 \pmod{m_2}$，$m_1 u$ 则正好相反。所以

$$x \equiv a_1\, m_2 v + a_2\, m_1 u \pmod{m_1 m_2}$$

满足 $x \equiv a_1 \pmod{m_1}$，$x \equiv a_2 \pmod{m_2}$。这里 $e_1 = m_2 v$，$e_2 = m_1 u$，一次扩展欧几里得算法同时给出两个「基向量」。

■ **例 5.8　末两位数字**　求 2^{100} 的末两位数字。

解　末两位就是模 100 的余数。$\gcd(2, 100) \neq 1$，不能直接对模 100 用 Euler 定理。把 100 拆成互素的 $4 \cdot 25$，分别计算：

- 模 4：$2^{100} = 4 \cdot 2^{98} \equiv 0 \pmod 4$。
- 模 25：$\gcd(2, 25) = 1$，$\varphi(25) = 20$，由 Euler 定理 $2^{20} \equiv 1$，所以 $2^{100} = (2^{20})^5 \equiv 1 \pmod{25}$。

再合并。$4 \cdot (-6) + 25 \cdot 1 = 1$，即 $u = -6$，$v = 1$，所以 $e_1 = 25$，$e_2 = -24$，

$$x \equiv 25 a_1 - 24 a_2 \pmod{100}.$$

代入 $a_1 = 0$，$a_2 = 1$：$x \equiv -24 \equiv 76 \pmod{100}$。所以 2^{100} 的末两位是 76。

同样的公式也能处理 3^{100}：$3^{100} \equiv (-1)^{100} = 1 \pmod 4$，$3^{100} = (3^{20})^5 \equiv 1 \pmod{25}$，$x \equiv 25 - 24 = 1$，末两位是 01。第 4 讲对 3^{100} 直接用了模 100 的 Euler 定理；中国剩余定理的好处是底数与模数不互素时同样适用。 ♦

易错提醒　定理要求模数**两两互素**，只要求 $\gcd(m_1, \dots, m_k) = 1$ 是不够的。例如 $\gcd(2, 3, 4) = 1$，但 2 与 4 不互素：$x \equiv 0 \pmod 2$，$x \equiv 0 \pmod 3$，$x \equiv 1 \pmod 4$ 无解，因为第三个同余式迫使 x 为奇数。即使有解，解也只是模 $\operatorname{lcm}(2, 3, 4) = 12$ 唯一，而不是模 $2 \cdot 3 \cdot 4 = 24$ 唯一。下一节专门处理模数不互素的情形。

历史与背景　《孙子算经》只给了答数和一个具体算法：「凡三三数之剩一，则置七十；五五数之剩一，则置二十一；七七数之剩一，则置十五。一百六以上，以一百五减之，即得。」其中 $70, 21, 15$ 正是上面的 e_1, e_2, e_3。明代程大位《算法统宗》（1592）把它编成口诀：「三人同行七十稀，五树梅花廿一枝，七子团圆正半月，除百零五便得知。」南宋秦九韶《数书九章》（1247）的「大衍求一术」给出了求 H_i 的一般算法，本质上就是扩展欧几里得算法，并处理了模数不互素的情形。在欧洲，Gauss 在《算术研究》（1801）中系统叙述了这一结果。1852 年，英国传教士伟烈亚力（Alexander Wylie）在上海发表文章，首次向西方介绍孙子的问题和秦九韶的大衍术，「中国剩余定理」这个名称由此而来。

5.4　模数不互素的同余式组

例 5.4 说明模数不互素时可能无解。下面的定理给出准确的判别条件。

定理 5.5（两个同余式的一般情形）　设 $g=\gcd(m_1,m_2)$，$L=\operatorname{lcm}(m_1,m_2)$。同余式组 $x\equiv a_1 \pmod{m_1}$，$x\equiv a_2 \pmod{m_2}$ 有解当且仅当 $g\mid a_1-a_2$。有解时，解模 L 唯一。

证明　*必要性*。若 x 是解，则 $m_1\mid x-a_1$，$m_2\mid x-a_2$，而 g 整除 m_1 和 m_2，所以 g 整除 $(x-a_2)-(x-a_1)=a_1-a_2$。

充分性与唯一性。第一个同余式的解是 $x=a_1+m_1t$，$t\in\mathbb{Z}$。代入第二个，得到关于 t 的一次同余式

$$m_1t\equiv a_2-a_1 \pmod{m_2}.$$

由定理 5.1，它有解当且仅当 $g\mid a_2-a_1$；有解时，按 §5.1 的约化，它的解构成一个模 m_2/g 的剩余类 $t=t_0+\frac{m_2}{g}s$。于是

$$x=a_1+m_1t_0+\frac{m_1m_2}{g}\,s=x_0+Ls,\qquad s\in\mathbb{Z},$$

这里用了第 2 讲的 $\gcd(m_1,m_2)\operatorname{lcm}(m_1,m_2)=m_1m_2$。所以全部解恰好是一个模 L 的剩余类。 ■

当 $g=1$ 时 $L=m_1m_2$，条件 $1\mid a_1-a_2$ 自动成立，这就是 $k=2$ 时的中国剩余定理。

■ **例 5.9　有公因数但相容**　解 $x\equiv 3 \pmod 8$，$x\equiv 7 \pmod{12}$。

解　$g=\gcd(8,12)=4$，$a_1-a_2=-4$，$4\mid -4$，有解，且解模 $L=24$ 唯一。

方法一：代入。$x=3+8t$，代入得 $8t\equiv 4 \pmod{12}$。三者同除以 4：$2t\equiv 1 \pmod 3$，$t\equiv 2 \pmod 3$。所以 $x=3+8\cdot 2=19$，$x\equiv 19 \pmod{24}$。

方法二：拆成素数幂。由中国剩余定理，$x\equiv 7 \pmod{12}$ 等价于 $x\equiv 7\equiv 3 \pmod 4$ 且 $x\equiv 1 \pmod 3$。而 $x\equiv 3 \pmod 8$ 已经蕴含 $x\equiv 3 \pmod 4$，所以条件 $x\equiv 3 \pmod 4$ 是多余的。原方程组等价于 $x\equiv 3 \pmod 8$，$x\equiv 1 \pmod 3$，这两个模数互素，解得 $x\equiv 19 \pmod{24}$。

检验：$19=2\cdot 8+3=12+7$。 ♦

方法二说明了一般原理：把每个模数分解成素数幂，对每个素数 p，只需检查涉及 p 的各个条件是否相容，然后保留其中 p 的幂次最高的一个。若出现矛盾（如例 5.4 中的 $x\equiv 3$ 与 $x\equiv 1 \pmod 4$），就无解。

■ **例 5.10　一篮鸡蛋**　一篮鸡蛋，2 个 2 个地数、3 个 3 个地数、……、6 个 6 个地数，都剩 1 个；7 个 7 个地数，正好数完。篮中至少有几个鸡蛋？

解　条件是 $x\equiv 1 \pmod k$（$k=2,3,4,5,6$）且 $x\equiv 0 \pmod 7$。前五个模数两两不互素，但余数都是 1，相容条件自然成立。它们合起来等价于 $k\mid x-1$ 对 $k=2,\dots,6$ 都成立，即

$$\operatorname{lcm}(2,3,4,5,6)=60\mid x-1,\qquad x\equiv 1 \pmod{60}.$$

再与 $x \equiv 0 \pmod 7$ 联立。$x = 1+60t$，代入得 $60t \equiv -1 \pmod 7$。$60 \equiv 4$，$-1 \equiv 6 \pmod 7$，所以 $4t \equiv 6 \pmod 7$。$4^{-1} \equiv 2 \pmod 7$（$4\cdot 2 = 8 \equiv 1$），$t \equiv 12 \equiv 5 \pmod 7$。于是 $x = 1+60\cdot 5 = 301$，全部解为 $x \equiv 301 \pmod{420}$。

检验：$301 = 7\cdot 43$，$300 = 60\cdot 5$。篮中至少有 301 个鸡蛋。 ♦

注 5.1　对 k 个同余式 $x \equiv a_i \pmod{m_i}$，结论是：方程组有解当且仅当两两相容，即对所有 $i<j$ 都有 $\gcd(m_i,m_j) \mid a_i - a_j$；有解时，解模 $\operatorname{lcm}(m_1,\dots,m_k)$ 唯一。必要性同定理 5.5。充分性可以用定理 5.5 逐个合并，也可以用上面「拆成素数幂」的办法证明（习题 4 给出一个具体例子）。注意每一对都要检查：对警告中的 $x \equiv 0 \pmod 2$，$x \equiv 0 \pmod 3$，$x \equiv 1 \pmod 4$，前两个、后两个都相容，矛盾出在第一个与第三个之间。

5.5　结构的观点：一一对应与 Euler 函数

5.5.1　模 M 的数就是一组余数

中国剩余定理可以换一种说法。设 $m_1,\dots,m_k$ 两两互素，$M = m_1\cdots m_k$，考虑映射

$$\Phi:\ \mathbb{Z}/M \longrightarrow \mathbb{Z}/m_1 \times \mathbb{Z}/m_2 \times \cdots \times \mathbb{Z}/m_k, \qquad x \bmod M \longmapsto (x \bmod m_1,\ \dots,\ x \bmod m_k).$$

这里 $\mathbb{Z}/m$ 表示模 m 的全部剩余类 $\{0,1,\dots,m-1\}$。这个映射是良定义的：若 $x \equiv y \pmod M$，则对每个 i 有 $x \equiv y \pmod{m_i}$。

定理 5.6（中国剩余定理的结构形式）　Φ 是一一对应。并且它保持加法和乘法：$\Phi(x+y) = \Phi(x)+\Phi(y)$，$\Phi(xy) = \Phi(x)\Phi(y)$，其中右边的加法、乘法按分量进行。

证明　唯一性部分说的是 Φ 是单射；存在性部分说的是 Φ 是满射。其实只需单射：两边都恰有 $M = m_1m_2\cdots m_k$ 个元素，有限集合之间的单射一定是满射。保持运算是因为同余与加法、乘法相容（第 3 讲）。 ■

把 $\mathbb{Z}/M$ 的元素填进一张 $m_1 \times m_2$ 的表格：x 填在第 $x \bmod m_1$ 行、第 $x \bmod m_2$ 列。定理 5.6 说，模数互素时每个格子恰好填一个数。图 5.2 左边是 $15 = 3\cdot 5$ 的表；右边是 4 与 6 不互素时的对照，那里一半格子是空的，另一半各填了两个数。

$\mathbb{Z}/15$：每格恰一个数

$x \bmod 3$ \ $x \bmod 5$	0	1	2	3	4
0	0	6	12	3	9
1	10	1	7	13	4
2	5	11	2	8	14

$\mathbb{Z}/24$：4 与 6 不互素

$x \bmod 4$ \ $x \bmod 6$	0	1	2	3	4	5
0	0,12		8,20		4,16	
1		1,13		9,21		5,17
2	6,18		2,14		10,22	
3		7,19		3,15		11,23

图 5.2　左：模 15 的数与（模 3，模 5）的余数对一一对应，例如 $8 \leftrightarrow (2,3)$。右：模 4 与模 6 的余数奇偶性必须相同，灰色格子是不相容的组合，没有数；其余格子各有两个数，相差 $\operatorname{lcm}(4,6) = 12$。

直观理解　定理 5.6 的「保持运算」意味着：模 15 的计算可以拆成模 3 和模 5 两个小计算，各算各的，最后再拼回来。例如 $8\cdot 13=104\equiv 14 \pmod{15}$；分量上 $8\leftrightarrow(2,3)$，$13\leftrightarrow(1,3)$，乘积 $(2\cdot 1,\ 3\cdot 3)=(2,4)$，而 $14\leftrightarrow(2,4)$，结果一致。计算机代数中常用这个办法：把一个大模数下的运算拆成许多个小素数模下的运算，并行计算后用中国剩余定理合并。

5.5.2 Euler 函数的积性

第 4 讲定义了 $\varphi(m)$：$1,\dots,m$ 中与 m 互素的数的个数。在一一对应 Φ 下，「与 M 互素」这个性质也按分量拆开。

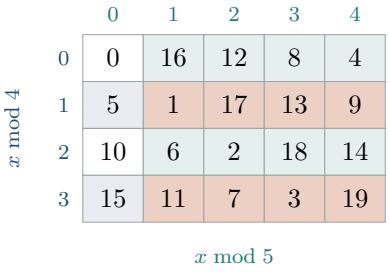
引理 5.7　设 $m_1,\dots,m_k$ 两两互素，$M=m_1\cdots m_k$。则 $\gcd(x,M)=1$ 当且仅当对每个 i 都有 $\gcd(x,m_i)=1$。又因 $\gcd(x,m_i)=\gcd(x \bmod m_i,\ m_i)$，这等价于 $\Phi(x)$ 的每个分量都与相应的模互素。

证明　若素数 p 整除 x 和某个 m_i，则 p 也整除 M；反过来，若素数 p 整除 x 和 $M=m_1\cdots m_k$，由 Euclid 引理 p 整除某个 m_i。所以 x 与 M 有公共素因数，当且仅当 x 与某个 m_i 有公共素因数。最后一句是因为 x 与 $x \bmod m_i$ 相差 m_i 的倍数，由第 2 讲欧几里得算法的关键引理，二者与 m_i 的最大公因数相同。 ■

定理 5.8（Euler 函数是积性函数）　若 $\gcd(m,n)=1$，则 $\varphi(mn)=\varphi(m)\varphi(n)$。

证明　由定理 5.6，$\Phi:\mathbb{Z}/mn\to\mathbb{Z}/m\times\mathbb{Z}/n$ 是一一对应。由引理 5.7，它把「与 mn 互素的剩余类」恰好映到「第一分量与 m 互素、第二分量与 n 互素的数对」。前者有 $\varphi(mn)$ 个，后者有 $\varphi(m)\cdot\varphi(n)$ 个。 ■

图 5.3 用模 20 的剩余表解释互素剩余类的计数为何相乘。



行 1,3：x 与 4 互素，$\varphi(4)=2$ 行；
列 1–4：x 与 5 互素，$\varphi(5)=4$ 列；
交叉处（橙色）：x 与 20 互素，
共 $2\times 4=8$ 个：
$1,3,7,9,11,13,17,19$。

图 5.3　$\varphi(20)=\varphi(4)\varphi(5)$。按 $(x \bmod 4,\ x \bmod 5)$ 排列模 20 的剩余，与 20 互素的数恰好是「好行」与「好列」的交叉。

易错提醒　互素条件不能省。$\varphi(4)=2$（与 4 互素的是 1,3），但 $\varphi(2)\varphi(2)=1$；$\varphi(12)=4$，但 $\varphi(2)\varphi(6)=1\cdot 2=2$。原因正是图 5.2 右边那样：模数不互素时 Φ 不是一一对应。

5.5.3 $\varphi(n)$ 的公式

积性把问题归结为素数幂。$1, 2, \dots, p^e$ 中与 p^e 不互素的数恰好是 p 的倍数 $p, 2p, \dots, p^{e-1}\cdot p$，共 p^{e-1} 个，所以

$$\varphi(p^e) = p^e - p^{e-1} = p^{e-1}(p-1) = p^e\Big(1-\frac{1}{p}\Big).$$

设 $n = p_1^{e_1}\cdots p_r^{e_r}$。不同素数的幂两两互素，对定理 5.8 反复使用，得到

$$\boxed{\varphi(n) = \prod_{i=1}^{r} p_i^{e_i-1}(p_i-1) = n\prod_{p\mid n}\Big(1-\frac{1}{p}\Big).}$$

■ **例 5.11**　计算 φ　计算 $\varphi(100)$，$\varphi(360)$，$\varphi(1001)$，$\varphi(1024)$。

解

- $100 = 2^2\cdot 5^2$：$\varphi(100) = 2\cdot 1\cdot 5\cdot 4 = 40$，或 $100\cdot\frac{1}{2}\cdot\frac{4}{5} = 40$。
- $360 = 2^3\cdot 3^2\cdot 5$：$\varphi(360) = 4\cdot 1\cdot 3\cdot 2\cdot 4 = 96$。
- $1001 = 7\cdot 11\cdot 13$：$\varphi(1001) = 6\cdot 10\cdot 12 = 720$。
- $1024 = 2^{10}$：$\varphi(1024) = 2^9 = 512$，恰好是一半，即全部奇数。

♦

想一想　$\varphi(1) = \varphi(2) = 1$，而 $\varphi(3) = 2$，$\varphi(4) = 2$，$\varphi(5) = 4$，$\varphi(6) = 2$，……都是偶数。为什么 $n \ge 3$ 时 $\varphi(n)$ 总是偶数？试给出两种解释。

答　用公式：若 n 有奇素因子 p，则 $p-1$ 是偶数，作为因子出现在 $\varphi(n)$ 中；否则 $n = 2^e$，$e \ge 2$，$\varphi(n) = 2^{e-1}$ 是偶数。不用公式：当 $\gcd(x,n) = 1$ 时 $\gcd(n-x, n) = 1$，所以与 n 互素的数两两配对 $x \leftrightarrow n-x$。$n \ge 3$ 时 $x \ne n-x$，否则 $n = 2x$，$\gcd(x,n) = x = 1$，$n = 2$。所以这些数个数为偶数。

5.6 数值计算与算法

前面的方法在纸上算小数没有问题。密码学中的模数有几百位，这时要问：计算要多少步？

5.6.1 用位数衡量快慢

一个正整数 N 有大约 $\log_{10} N$ 位十进制数字、$\log_2 N$ 位二进制数字。输入 N 只需写下这么多位，所以衡量算法快慢的尺度是 $n \approx \log N$，而不是 N 本身。若一个算法的步数不超过 $(\log N)^c$（c 为常数），就说它是多项式时间（polynomial time）的，即「有效的」。

对比两个求整除信息的办法。试除法找 N 的因数要试到 $\sqrt{N} = 10^{n/2}$，步数是位数的指数；欧几里得算法的除法次数不超过 5 倍的十进制位数（第 2 讲的 Lamé 定理），每次除法的代价是位数的平方量级，总代价是位数的多项式。N 有 100 位时，前者要 10^{50} 步，后者只要几百次除法。

运算（N 为数的大小，$n \approx \log N$）	方法	位运算次数的量级
加法、减法	竖式	n
乘法	竖式	n^2
乘法	Karatsuba	$n^{\log_2 3} \approx n^{1.585}$
带余除法、取模	竖式	n^2
gcd、逆元、中国剩余定理	扩展欧几里得算法	n^3 以内
$a^b \bmod N$（$b < N$）	反复平方	n^3
找出 N 的一个因数	试除法	$10^{n/2}$

5.6.2 Karatsuba 乘法

竖式乘法把两个 n 位数的每一位两两相乘，要做 n^2 次一位数乘法。1960 年 Kolmogorov 猜想这是最优的，不久 Karatsuba 就找到了更快的方法。把两个 $2k$ 位数各分成高低两半：$x = 10^k a + b$，$y = 10^k c + d$。则

$$xy = 10^{2k} ac + 10^k (ad + bc) + bd.$$

直接算要 4 次 k 位乘法。关键在于中间项可以由一次乘法得到：

$$ad + bc = (a + b)(c + d) - ac - bd.$$

所以只需 ac、bd、$(a+b)(c+d)$ 三次乘法，外加几次加减法。

例 5.12　Karatsuba 乘法　用 Karatsuba 方法计算 1234×5678。

解　$k = 2$，$a = 12$，$b = 34$，$c = 56$，$d = 78$。三次乘法：

$$ac = 12 \cdot 56 = 672, \qquad bd = 34 \cdot 78 = 2652, \qquad (a+b)(c+d) = 46 \cdot 134 = 6164.$$

中间项 $ad + bc = 6164 - 672 - 2652 = 2840$（直接算：$12 \cdot 78 + 34 \cdot 56 = 936 + 1904 = 2840$）。于是

$$1234 \times 5678 = 672 \cdot 10^4 + 2840 \cdot 10^2 + 2652 = 6720000 + 284000 + 2652 = 7006652.$$

♦

递归时还须计入加减法和进位。为避免 $a + b$ 多出一位，可以把中间项改写成 $ad + bc = ac + bd - (a - b)(c - d)$，分别记录差的符号，再乘其绝对值；差的绝对值仍不到 10^k。因此总位运算次数满足

$$T(n) \le 3T(\lceil n/2 \rceil) + Cn,$$

其中 C 为常数，线性项包含加减法、符号处理和移位。展开递归得到 $T(n) = O(n^{\log_2 3})$，其中 $\log_2 3 \approx 1.585$。当 $n = 10^6$ 时，$n^2 / n^{\log_2 3}$ 约为 300；这只比较增长量级，不表示实际运行时间必然相差这个倍数。

5.6.3 反复平方法

计算 $a^b \bmod N$ 是数论算法中最常用的操作：Euler 定理、Fermat 素性检验、RSA 都要用。朴素地连乘 $b-1$ 次，b 有几百位时根本做不完，而且 a^b 本身的位数大约是 $b\log a$，写都写不下。两个想法解决这两个问题：

- **每乘一次就取模**，所有中间结果都小于 N；
- **反复平方**：$a, a^2, a^4, a^8, \dots$ 每一项是前一项的平方，a^{2^j} 只需 j 次乘法。再把 b 写成二进制，$b=\sum_j \beta_j 2^j$，则 $a^b=\prod_{\beta_j=1} a^{2^j}$。

算法 5.9（反复平方法）　输入 a、$b\ge 1$、$N\ge 2$，输出 $a^b \bmod N$。

(1) 写出 b 的二进制展开 $b=(\beta_t\beta_{t-1}\cdots\beta_0)_2$，$t=\lfloor \log_2 b\rfloor$。

(2) 令 $s_0=a \bmod N$，依次计算 $s_j=s_{j-1}^2 \bmod N$，$j=1,\dots,t$。这时 $s_j\equiv a^{2^j}$。

(3) 把 $\beta_j=1$ 的那些 s_j 乘起来，每乘一次就模 N，输出结果。

第 2 步做 t 次平方，第 3 步至多做 t 次乘法，共至多 $2\log_2 b$ 次模 N 乘法。每次的代价是 $(\log N)^2$ 量级，所以总代价约为 $\log b\cdot(\log N)^2$。$b<N$ 时这是 $(\log N)^3$，即位数的三次方。朴素连乘的代价是 $b(\log N)^2$，对 b 的位数是指数级的。

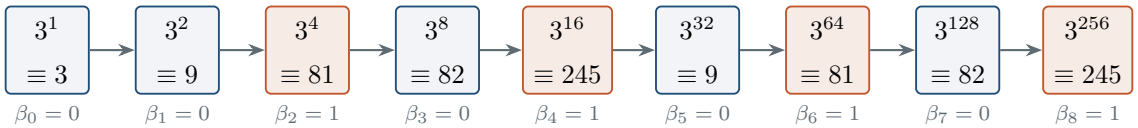


图 5.4　计算 $3^{340} \bmod 341$ 的平方链，每个箭头表示「平方后模 341」。$340=(101010100)_2=2^8+2^6+2^4+2^2$，橙色方框是 $\beta_j=1$ 的项，把它们乘起来就得到 3^{340}。

例 5.13　反复平方　计算 $3^{340} \bmod 341$，并用中国剩余定理检验。

解　$340=256+64+16+4=(101010100)_2$，$t=8$。平方链（图 5.4）：

$$s_0=3,\quad s_1=9,\quad s_2=81,\quad s_3=81^2=6561\equiv 82,\quad s_4=82^2=6724\equiv 245,$$
$$s_5=245^2=60025\equiv 9,\quad s_6=81,\quad s_7=82,\quad s_8=245.$$

（$s_5=9=s_1$，之后循环出现，但算法并不需要发现这一点。）取 $\beta_j=1$ 的项 s_2,s_4,s_6,s_8 相乘：

$$81\cdot 245=19845\equiv 67,\qquad 67\cdot 81=5427\equiv 312,\qquad 312\cdot 245=76440\equiv 56 \pmod{341}.$$

所以 $3^{340}\equiv 56 \pmod{341}$，共用 8 次平方和 3 次乘法，而朴素连乘要 339 次。

检验：$341=11\cdot 31$。模 11：由 Fermat 小定理 $3^{10}\equiv 1$，$3^{340}=(3^{10})^{34}\equiv 1$。模 31：$3^{30}\equiv 1$，$340=11\cdot 30+10$，所以 $3^{340}\equiv 3^{10}$；$3^5=243\equiv -5$，$3^{10}\equiv 25$。解 $x\equiv 1 \pmod{11}$，$x\equiv 25 \pmod{31}$：$x=25+31t$，$25+31t\equiv 3+9t\equiv 1 \pmod{11}$，$9t\equiv -2\equiv 9$，$t\equiv 1$，$x=56$。与上面一致。 ♦

若 341 是素数，Fermat 小定理预言 $3^{340} \equiv 1$。现在 $3^{340} \equiv 56 \neq 1$，说明 341 不是素数，而这个判断完全没有用到 341 的分解。第 4 讲看到，换成底数 2 却会被骗过：$2^{340} \equiv 1 \pmod{341}$。

前后联系 本讲的算法（gcd、逆元、解同余式组、计算 $a^b \bmod N$）都是输入总位数的多项式时间。一般整数分解则没有已知的经典多项式时间算法；某些几百位合数仍可凭特殊结构轻易分解，不能只凭位数断言困难。第 6 讲将讨论 Fermat 素性检验与 RSA：加密、解密可用反复平方，求私钥用扩展欧几里得算法。分解 RSA 模数 N 足以恢复私钥；这里没有证明任意破解 RSA 的办法都必须先分解 N。

历史与背景 1960 年，Kolmogorov 在莫斯科大学的讨论班上猜想 n 位数的乘法需要 n^2 量级的运算，当时 23 岁的学生 Karatsuba 在一周内给出了反例，结果于 1962 年以 Karatsuba 与 Ofman 的名义发表。此后 Schönhage 和 Strassen（1971）用快速 Fourier 变换把乘法降到 $n \log n \log\log n$，Harvey 和 van der Hoeven（2019）达到了 $n \log n$，人们猜想这已是最优。反复平方法则古老得多：据 Knuth《计算机程序设计艺术》第 2 卷的考证，计算 x^n 的二进制方法在公元前 200 年以前就出现在印度 Pingala 的著作中。

本讲小结

- 解 $ax \equiv b \pmod{m}$：$g = \gcd(a, m)$，$g \nmid b$ 则无解；否则 a, b, m 同时除以 g，用扩展欧几里得算法求逆元得 $x \equiv x_1 \pmod{m/g}$，模 m 共 g 个解 $x_1 + j\, m/g$。
- 中国剩余定理：模数两两互素时，同余式组有解且模 $M = \prod m_i$ 唯一，$x \equiv \sum a_i H_i N_i$，$N_i = M/m_i$，$H_i \equiv N_i^{-1} \pmod{m_i}$。也可以逐次代入，每步解一个一次同余式。
- 模数不互素：$x \equiv a_1 \pmod{m_1}$，$x \equiv a_2 \pmod{m_2}$ 有解 $\iff \gcd(m_1, m_2) \mid a_1 - a_2$，解模 $\operatorname{lcm}(m_1, m_2)$ 唯一；多个同余式要求两两相容。
- 结构形式：$\mathbb{Z}/M \to \prod \mathbb{Z}/m_i$ 是保持加法、乘法的一一对应，且与 M 互素恰好对应各分量都与 m_i 互素。由此 φ 是积性的，$\varphi(n) = n \prod_{p \mid n} (1 - 1/p)$。
- 有效算法 = 步数是位数 $\log N$ 的多项式。Karatsuba 乘法 $n^{1.585}$；反复平方法计算 $a^b \bmod N$ 只需至多 $2 \log_2 b$ 次模乘法。

练习

5.1 ★ 求 $12x \equiv 9 \pmod{15}$ 模 15 的全部解；$12x \equiv 10 \pmod{15}$ 有解吗？

5.2 ★ 用扩展欧几里得算法求 $37^{-1} \bmod 101$，并解 $37x \equiv 5 \pmod{101}$。

5.3 ★ 用公式法和逐次代入法两种方法解 $x \equiv 1 \pmod 3$，$x \equiv 2 \pmod 4$，$x \equiv 3 \pmod 5$。你能看出一个更快的办法吗？

5.4 ★★ 同余式组 $x \equiv 5 \pmod 6$，$x \equiv 7 \pmod{10}$，$x \equiv 3 \pmod{15}$ 是否有解？若把第三个同余式改为 $x \equiv 2 \pmod{15}$ 呢？有解时求出全部解。

5.5 ★ 计算 $\varphi(720)$，$\varphi(1155)$，$\varphi(243)$，$\varphi(600)$。

5.6 ★★ 求所有满足 $\varphi(n)=12$ 的正整数 n。（提示：若素数 $p\mid n$，则 $p-1\mid\varphi(n)$。）

5.7 ★★ 求 2^{2026} 的末三位数字。（提示：$1000=8\cdot125$。）

5.8 ★★ 用反复平方法计算 $3^{117}\bmod 101$，写出平方链和乘法过程，再用 Fermat 小定理检验。

5.9 ★★★ 证明：对任意正整数 N，存在 N 个连续的正整数，其中每一个都能被某个大于 1 的完全平方数整除。

5.10 ★★★ 设 m 有 r 个不同的素因子。证明 $x^2\equiv x\pmod m$ 模 m 恰有 2^r 个解，并写出 $m=30$ 时的全部解。

练习答案与提示

5.1 $g=\gcd(12,15)=3\mid 9$，约化为 $4x\equiv 3\pmod 5$。$4^{-1}\equiv 4$（$16\equiv 1$），$x\equiv 12\equiv 2\pmod 5$，所以 $x\equiv 2,7,12\pmod{15}$。$12x\equiv 10$：$3\nmid 10$，无解。

5.2 $101=2\cdot37+27$，$37=1\cdot27+10$，$27=2\cdot10+7$，$10=1\cdot7+3$，$7=2\cdot3+1$。回代：$1=3\cdot7-2\cdot10=3\cdot27-8\cdot10=11\cdot27-8\cdot37=11\cdot101-30\cdot37$。所以 $37^{-1}\equiv-30\equiv71$；检验 $37\cdot71=2627=26\cdot101+1$。$x\equiv71\cdot5=355\equiv52\pmod{101}$。

5.3 公式法：$M=60$；$N_1=20\equiv2\pmod3$，$H_1=2$，$e_1=40$；$N_2=15\equiv3\pmod4$，$H_2=3$，$e_2=45$；$N_3=12\equiv2\pmod5$，$H_3=3$，$e_3=36$。$x\equiv40+2\cdot45+3\cdot36=238\equiv58\pmod{60}$。代入法同样得到 58。更快的办法：三个条件都是 $x\equiv-2$，所以 $x\equiv-2\equiv58\pmod{60}$。

5.4 原方程组无解：$\gcd(6,15)=3$，而 $5-3=2$ 不被 3 整除（模 3 下第一个给出 $x\equiv2$，第三个给出 $x\equiv0$）。改为 $x\equiv2\pmod{15}$ 后，三对都相容：$2\mid5-7$，$3\mid5-2$，$5\mid7-2$。拆成素数幂，三个同余式都归结为 $x\equiv1\pmod2$，$x\equiv2\pmod3$，$x\equiv2\pmod5$，解得 $x\equiv17\pmod{30}$，$30=\operatorname{lcm}(6,10,15)$。检验：$17=2\cdot6+5=10+7=15+2$。

5.5 $720=2^4\cdot3^2\cdot5$，$\varphi=8\cdot6\cdot4=192$。$1155=3\cdot5\cdot7\cdot11$，$\varphi=2\cdot4\cdot6\cdot10=480$。$243=3^5$，$\varphi=3^4\cdot2=162$。$600=2^3\cdot3\cdot5^2$，$\varphi=4\cdot2\cdot20=160$。

5.6 n 的素因子 p 满足 $p-1\mid12$，所以 $p\in\{2,3,5,7,13\}$；又 $p^{e-1}\mid12$，故 $2^4\nmid n$，$3^3\nmid n$，5,7,13 的指数为 1。设 $\varphi(m)$ 为相应剩余因子的值。由前面的结论，$m\ge3$ 时 $\varphi(m)$ 为偶数，而 $\varphi(1)=\varphi(2)=1$。按最大的奇素因子分类：含 13 时其余部分 m 满足 $\varphi(m)=1$，故 $m=1,2$，得 13, 26。含 7 时 $\varphi(m)=2$。此时 m 只可能含 2, 3, 5；若 $5\mid m$，则 $4\mid\varphi(m)$，不可能。写 $m=2^u3^v$：若 $u=0,v\ge1$，由 $2\cdot3^{v-1}=2$ 得 $v=1$；若 $v=0,u\ge1$，由 $2^{u-1}=2$ 得 $u=2$；若 $u,v\ge1$，由 $2^u3^{v-1}=2$ 得 $u=v=1$；$u=v=0$ 时 $\varphi(m)=1$。所以 $m=3,4,6$，得 21, 28, 42。含 5 时需 $\varphi(m)=3$，但 $m=1,2$ 时函数值为 1，$m\ge3$ 时为偶数，故不可能。

若 n 只含 2, 3，写 $n=2^a3^b$。$a=b=0$ 时 $n=1$；若 $a=0,b\ge1$，则 $\varphi(n)=2\cdot3^{b-1}$，不可能等于 12，因为 6 不是 3 的幂；若 $b=0,a\ge1$，则 $\varphi(n)=2^{a-1}$，也不可能等于 12，因为 12 不是 2 的幂。当 $a,b\ge1$ 时，$\varphi(n)=2^a3^{b-1}=12=2^2\cdot3$，故 $a=b=2$、$n=36$。答案：13, 21, 26, 28, 36, 42。

5.7 模 8：$2^{2026}\equiv0$。模 125：$\varphi(125)=100$，$2^{2026}\equiv2^{26}$；$2^7=128\equiv3$，$2^{26}=(2^7)^3\cdot2^5\equiv27\cdot32=864\equiv114$。合并：$x=114+125t\equiv2+5t\equiv0\pmod8$，$5t\equiv6$，$t\equiv6\cdot5=30\equiv6$，$x=114+750=864$。末三位是 864。

5.8 $117=(1110101)_2=64+32+16+4+1$。平方链：$3,\ 9,\ 81,\ 81^2\equiv97,\ 97^2\equiv(-4)^2=16,\ 16^2\equiv54,\ 54^2\equiv88$，分别是 3^{2^j}，$j=0,\dots,6$。取 $j=0,2,4,5,6$：$3\cdot81\equiv41$，$41\cdot16\equiv50$，$50\cdot54\equiv74$，$74\cdot88\equiv48$。所以 $3^{117}\equiv48\pmod{101}$，用了 6 次平方、4 次乘法。检验：$3^{100}\equiv1$，$3^{117}\equiv3^{17}=3^{16}\cdot3\equiv16\cdot3=48$。

5.9 取 N 个不同的素数 $p_1,\dots,p_N$。$p_1^2,\dots,p_N^2$ 两两互素，由中国剩余定理，存在正整数 x 使 $x\equiv-i\pmod{p_i^2}$，$i=1,\dots,N$。于是 $p_i^2\mid x+i$，$x+1,\dots,x+N$ 每个都有平方因子。$N=3$、$p=2,3,5$

时就是例 5.7：$x = 547$，得到 $548, 549, 550$。

5.10 因 $\gcd(x, x-1) = 1$，模每个 $p_i^{e_i}$ 只能有 $x \equiv 0$ 或 1 两解；中国剩余定理使 r 个选择独立，故模 m 有 2^r 解。$m = 30$ 时为 $0, 1, 6, 10, 15, 16, 21, 25$。

第 6 讲

素性检验、因数分解、RSA 与 Hensel 引理

Primality, Factoring, RSA, Hensel's Lemma

对应材料：MIT 18.781 Lecture 6 讲义（Primality, Factoring, RSA, Hensel's Lemma）；习题集 3 第 1 题、第 4(a)、(b)、(c) 题；习题集 4 第 1 部分第 2、3 题。

本讲目标

(1) 会用中国剩余定理把模 n 的多项式同余式拆成模素数幂的同余式，并按「解数相乘」计数。

(2) 知道「快」的标准是位数 $\log n$ 的多项式；理解为什么试除法不算快，而反复平方法算快。

(3) 掌握 Fermat 检验和 Miller–Rabin 强伪素数检验，知道 Carmichael 数能通过所有互素底数的 Fermat 检验，但仍可能通过某些底数的强检验；会说明独立随机多轮检验如何降低误判概率。

(4) 会手算 Pollard ρ 算法，理解「生日悖论」给出的约 $\sqrt{p}$ 步的估计。

(5) 掌握 RSA 的密钥生成、加密与解密，能证明解密的正确性，并说明分解模数为何足以恢复私钥。

(6) 理解 Hensel 引理的陈述：模 p 的非奇异根可以唯一地提升到模 $p^2, p^3, \dots$。

这一讲的几个主题看起来各不相干，其实有一条主线：**判断素数容易，分解整数难**。第 5 讲的中国剩余定理告诉我们，只要知道 n 的素因数分解，很多关于模 n 的问题都可以拆成模素数幂的小问题；本讲的第一节先把这件事推广到任意多项式同余式。可是「知道分解」这个前提本身就很难满足。第二、三节比较两个问题的难度：判断 n 是不是素数有很快的算法，分解 n 却没有。第四节的 RSA 密码正是利用了这个差距。最后一节回到解同余式，用 Hensel 引理处理模素数幂的情形。

记号。第 5 讲为了和有序数组区分，一直把最大公因数写成 $\gcd(a,b)$。从本讲起，在不会与有序数组混淆的地方，按数论书的习惯简写为 (a,b)，例如 $(a,n)=1$ 表示 a 与 n 互素。

6.1 中国剩余定理与同余式的解数

6.1.1 一个二次方程有四个根

在实数里，$x^2=1$ 只有 $x=\pm 1$ 两个根。模 15 呢？把 $x=0,1,\dots,14$ 逐个代入，$x^2\equiv 1 \pmod{15}$ 的解是

$$x \equiv 1,\ 4,\ 11,\ 14 \pmod{15}.$$

比如 $4^2 = 16 = 15+1$，$11^2 = 121 = 8\cdot 15 + 1$。一个二次同余式有四个根，这在实数或复数里不可能发生。

原因要从 $15 = 3\cdot 5$ 看。若 $x^2 \equiv 1 \pmod{15}$，则 $x^2 \equiv 1 \pmod 3$ 且 $x^2 \equiv 1 \pmod 5$。模 3 的解是 $x \equiv \pm 1$，即 $1, 2$；模 5 的解也是 $x \equiv \pm 1$，即 $1, 4$。按中国剩余定理，「模 3 的余数」与「模 5 的余数」可以任意搭配，每种搭配恰好对应一个模 15 的余数：$1 \leftrightarrow (1,1)$，$4 \leftrightarrow (1,4)$，$11 \leftrightarrow (2,1)$，$14 \leftrightarrow (2,4)$。两边各有 2 种选择，于是共有 $2\times 2 = 4$ 个解。其中 $x \equiv \pm 1$ 是两边取同号的搭配，4 和 11 是「一边取 $+1$、另一边取 -1」的搭配。图 6.1 把模 15 的全部余数排成 3×5 的方格，四个解恰好落在两行与两列的交叉处。

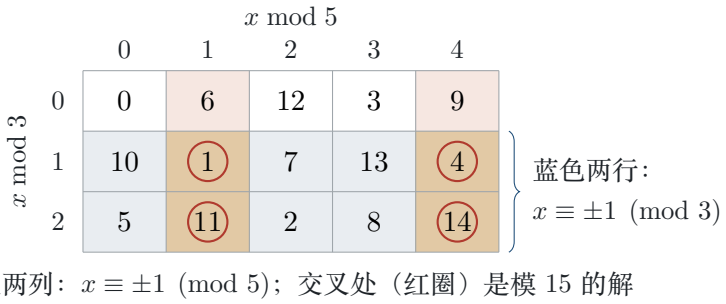


图 6.1　模 15 的余数按 $(x \bmod 3,\ x \bmod 5)$ 排成方格，每个余数恰好出现一次（中国剩余定理）。$x^2 \equiv 1 \pmod{15}$ 的解就是「行满足模 3 的条件、列满足模 5 的条件」的格子，共 $2\times 2 = 4$ 个。

6.1.2 解数的乘法原理

同样的道理对任何整系数多项式都成立。

定理 6.1（同余式解数的乘法原理）　设 f 是整系数多项式，$n = m_1 m_2 \cdots m_r$，其中 $m_1, \dots, m_r$ 两两互素（最常用的是 $n = p_1^{e_1}\cdots p_r^{e_r}$，$m_i = p_i^{e_i}$）。对模 n 的余数 x，令

$$\Phi(x) = (x \bmod m_1,\ \dots,\ x \bmod m_r).$$

则 Φ 是从「$f(x) \equiv 0 \pmod n$ 的解」到「各个 $f(x_i) \equiv 0 \pmod{m_i}$ 的解所组成的 r 元组」的一一对应。特别地，

$$\#\{\text{模} n \text{ 的解}\} = \prod_{i=1}^{r} \#\{\text{模} m_i \text{ 的解}\}.$$

证明　解映到解。若 $n \mid f(x)$，因为 $m_i \mid n$，所以 $m_i \mid f(x)$，即 $x \bmod m_i$ 是模 m_i 的解。

反过来。任给模 m_i 的解 x_i $(i = 1, \dots, r)$。由中国剩余定理，存在唯一的 $x \bmod n$ 使 $x \equiv x_i \pmod{m_i}$ 对所有 i 成立。同余可以代入多项式（第 4 讲：同余式可以相加、相乘），所以

$$f(x) \equiv f(x_i) \equiv 0 \pmod{m_i}, \qquad i = 1, \dots, r.$$

$m_1, \dots, m_r$ 两两互素且都整除 $f(x)$，所以它们的乘积 n 整除 $f(x)$。于是 x 是模 n 的解，而且由中国剩余定理的唯一性，它是唯一一个映到 $(x_1, \dots, x_r)$ 的余数。 ■

■ **例 6.1**　**习题集 3 第 1 题：一个三次同余式**　不用计算机，解 $x^3-9x^2+23x-15\equiv 0 \pmod{143}$。

解　第一步：分解模数。$143=11\cdot 13$。由定理 6.1，只需分别解模 11 和模 13 的同余式，再拼起来。

第二步：解小模数。试 $x=1$：$1-9+23-15=0$，所以 $x-1$ 是因式。做多项式除法得

$$x^3-9x^2+23x-15=(x-1)(x^2-8x+15)=(x-1)(x-3)(x-5).$$

11 和 13 都是素数，乘积被素数整除当且仅当某个因子被它整除（Euclid 引理），所以模 11 的解恰好是 $x\equiv 1,3,5$，模 13 的解也恰好是 $x\equiv 1,3,5$。

第三步：拼接。先用扩展欧几里得算法求 $13u+11v=1$：$13=11+2$，$11=5\cdot 2+1$，倒推得 $1=11-5\cdot 2=11-5(13-11)=6\cdot 11-5\cdot 13$。于是 $-65\equiv 1 \pmod{11}$、$-65\equiv 0 \pmod{13}$，$66\equiv 0 \pmod{11}$、$66\equiv 1 \pmod{13}$。若 $x\equiv a \pmod{11}$，$x\equiv b \pmod{13}$，则

$$x\equiv -65a+66b \pmod{143}.$$

让 a,b 各取 $1,3,5$，得到 $3\times 3=9$ 个解：

	$b=1$	$b=3$	$b=5$
$a=1$	1	133	122
$a=3$	14	3	135
$a=5$	27	16	5

例如 $a=3$，$b=1$：$-195+66=-129\equiv 14 \pmod{143}$。验算：$x=14$ 时三个因子是 $13,11,9$，乘积 $1287=9\cdot 143$。所以全部解是 $x\equiv 1,3,5,14,16,27,122,133,135 \pmod{143}$。 ♦

想一想　$x=14$ 时三个因子 $13,11,9$ 没有一个是 143 的倍数，乘积却是。这种「乘积为 0、因子都不为 0」的现象为什么模素数不会发生？它和「三次方程有 9 个根」有什么关系？

答　模素数 p 时，Euclid 引理说 $p\mid ab$ 推出 $p\mid a$ 或 $p\mid b$，所以 $(x-1)(x-3)(x-5)\equiv 0$ 只能来自某个因子为 0，根至多 3 个。模 143 时，11 可以整除一个因子而 13 整除另一个因子，于是出现「混搭」的根。第 7 讲的 Lagrange 定理会证明：模素数时，k 次多项式至多有 k 个根。

■ **例 6.2**　**$x^2\equiv 1$ 的解数**　设 n 是奇数，有 r 个不同的素因子。证明 $x^2\equiv 1 \pmod{n}$ 恰有 2^r 个解，并写出 $n=105$ 时的全部解。

解　先看一个奇素数幂 p^e。$x^2\equiv 1 \pmod{p^e}$ 即 $p^e\mid (x-1)(x+1)$。两个因子相差 2，奇素数 p 不能同时整除它们（否则 $p\mid 2$），所以 p^e 的全部 e 个因子 p 都在同一个因子里：$p^e\mid x-1$ 或 $p^e\mid x+1$。于是解恰好是 $x\equiv \pm 1 \pmod{p^e}$，这两个数不同余（因为 $p^e>2$），共 2 个解。由定理 6.1，模 n 的解数是 $2\cdot 2\cdots 2=2^r$。

$n=105=3\cdot 5\cdot 7$，$r=3$，共 8 个解。每个解由三个符号 $(\pm 1,\pm 1,\pm 1)$ 决定，逐个用中国剩余定理拼出来，得到

$$x\equiv 1,\ 29,\ 34,\ 41,\ 64,\ 71,\ 76,\ 104 \pmod{105}.$$

例如 $29 \equiv -1 \pmod 3$，$29 \equiv -1 \pmod 5$，$29 \equiv 1 \pmod 7$。除了 ± 1（即 1 和 104）之外的 6 个解，称为 1 的**非平凡平方根**（nontrivial square root of 1）。下一节会看到，它们是揭穿合数的利器。 ♦

解同余式的三层结构　解 $f(x) \equiv 0 \pmod n$ 分三层：

(1) 分解 $n = p_1^{e_1} \cdots p_r^{e_r}$，用中国剩余定理把问题拆成模 $p_i^{e_i}$ 的问题（本节）；

(2) 从模 p 的根出发，用 Hensel 引理逐层提升到模 p^e（第 6.5 节和第 7 讲）；

(3) 模素数 p 的问题是最底层，第 7 讲以后专门研究。

第一层隐含一个前提：**要先知道 n 的分解**。对大的 n，这恰恰是最难的一步。

6.2　素性检验

6.2.1　什么叫「快」

给定正整数 n，判断它是素数还是合数。衡量算法快慢，要看步数随**输入的长度**怎样增长。写下 n 需要约 $\log_{10} n$ 个十进制位（或 $\log_2 n$ 个二进制位），所以输入长度是 $\log n$ 量级。如果步数不超过 $(\log n)^c$（c 是常数），就说算法是**多项式时间**（polynomial time）的，简称「快」。

最直接的办法是**试除法**（trial division）：用 $2, 3, \ldots, \lfloor \sqrt{n} \rfloor$ 逐个去除 n。若 $n = ab$ 且 $1 < a \le b$，则 $a \le \sqrt{n}$，所以试到 $\sqrt{n}$ 就够了。但步数 $\sqrt{n} = e^{\frac{1}{2}\ln n}$ 是输入长度的**指数**函数：对 100 位的 n 要约 10^{50} 次除法，每秒做 10^9 次也要约 3×10^{33} 年。对比第 5 讲的反复平方法：计算 $a^{n-1} \bmod n$ 只需约 $\log_2 n$ 次平方和至多同样多次乘法，100 位的 n 只要约 332 次平方。所以要找的是建立在反复平方之上的检验方法。

■ **例 6.3**　**用反复平方检验 91**　计算 $2^{90} \bmod 91$，并据此判断 91 是否为素数。

解　把指数写成二进制：$90 = 64 + 16 + 8 + 2 = 1011010_2$。反复平方（每一项是前一项的平方再模 91），$2^{2^i} \bmod 91$ 在 $i = 0, 1, \ldots, 6$ 时依次为 $2, 4, 16, 74, 16, 74, 16$，例如 $16^2 = 256 = 2 \cdot 91 + 74$，$74^2 = 5476 = 60 \cdot 91 + 16$。于是

$$2^{90} = 2^{64} \cdot 2^{16} \cdot 2^8 \cdot 2^2 \equiv 16 \cdot 16 \cdot 74 \cdot 4 \pmod{91}.$$

逐步约化：$16 \cdot 16 = 256 \equiv 74$，$74 \cdot 74 \equiv 16$，$16 \cdot 4 = 64$。所以 $2^{90} \equiv 64 \pmod{91}$。

如果 91 是素数，Fermat 小定理要求 $2^{90} \equiv 1 \pmod{91}$。现在 $64 \not\equiv 1$，所以 91 **是合数**。整个过程只做了十次左右的模乘法，而且**没有找到 91 的任何因子**（$91 = 7 \cdot 13$）。 ♦

6.2.2　Fermat 检验与伪素数

上例的做法可以写成一个算法。

算法 6.2（Fermat 检验）　输入 $n>2$。

(1) 随机取 $a\in\{2,3,\dots,n-1\}$。

(2) 用欧几里得算法算 (a,n)。若 $(a,n)>1$，输出「合数」（还顺便得到一个因子）。

(3) 用反复平方法算 $a^{n-1}\bmod n$。若 $a^{n-1}\not\equiv 1\pmod n$，输出「合数」；否则输出「不能确定」，可以换一个 a 再试。

「合数」的结论总是对的：它直接与 Fermat 小定理矛盾。问题出在「不能确定」。

定义 6.3（伪素数）　若合数 n 和与它互素的整数 a 满足 $a^{n-1}\equiv 1\pmod n$，则称 n 是**以 a 为底的伪素数**（pseudoprime to base a），此时也称 a 是 n 的一个 **Fermat 说谎者**（Fermat liar）。

■ **例 6.4　同一个合数，有的底数说谎，有的不说谎**　(1) 证明 91 是以 3 为底的伪素数。(2) 证明 $341=11\cdot 31$ 是以 2 为底的伪素数，但不是以 3 为底的伪素数。

解　(1) $3^6=729=8\cdot 91+1$，所以 $3^6\equiv 1\pmod{91}$，而 $90=6\cdot 15$，于是 $3^{90}=(3^6)^{15}\equiv 1\pmod{91}$。底数 3 看不出 91 是合数，而上例中底数 2 可以。

(2) $2^{10}=1024=3\cdot 341+1$，所以 $2^{340}=(2^{10})^{34}\equiv 1\pmod{341}$，341 是以 2 为底的伪素数（它是最小的一个）。而 $3^{340}\equiv 56\pmod{341}$，底数 3 揭穿了 341。只看模 31 就可得到矛盾：$3^{10}\equiv 25$，$3^{30}\equiv 1$，$3^{340}=(3^{30})^{11}\cdot 3^{10}\equiv 25\not\equiv 1\pmod{31}$。 ♦

伪素数说明一个底数不够，换几个底数常常就能揭穿合数。可惜存在一类合数，**所有**与它互素的底数都说谎。

定义 6.4（Carmichael 数）　若合数 n 对**每个**与 n 互素的 a 都满足 $a^{n-1}\equiv 1\pmod n$，则称 n 是 **Carmichael 数**（Carmichael number）。

定理 6.5（Korselt 判别法）　若合数 n 无平方因子，并且对 n 的每个素因子 p 都有 $p-1\mid n-1$，则 n 是 Carmichael 数。

证明　设 $(a,n)=1$，p 是 n 的任一素因子，写 $n-1=k(p-1)$。$p\nmid a$，由 Fermat 小定理 $a^{p-1}\equiv 1\pmod p$，所以 $a^{n-1}=(a^{p-1})^k\equiv 1\pmod p$。于是 n 的每个素因子都整除 $a^{n-1}-1$。n 无平方因子，n 就是这些互不相同的素数之积，所以 $n\mid a^{n-1}-1$。 ■

逆命题也成立：每个 Carmichael 数都无平方因子，且满足 $p-1\mid n-1$。证明「$p-1\mid n-1$」要用第 7 讲的原根，这里不证。最小的 Carmichael 数是 561，小于 10^4 的只有 561、1105、1729、2465、2821、6601、8911 七个。习题集 3 第 4(c) 题要求给出十个形如 $n=pqr$ 的例子；由 Korselt 判别法，若 p,q,r 是互不相同的素数且 $p-1,q-1,r-1$ 都整除 $pqr-1$，则 n 满足性质 P。例

如：$561 = 3\cdot 11\cdot 17$、$1105 = 5\cdot 13\cdot 17$、$1729 = 7\cdot 13\cdot 19$、$2465 = 5\cdot 17\cdot 29$、$2821 = 7\cdot 13\cdot 31$、$6601 = 7\cdot 23\cdot 41$、$8911 = 7\cdot 19\cdot 67$、$10585 = 5\cdot 29\cdot 73$、$15841 = 7\cdot 31\cdot 73$、$29341 = 13\cdot 37\cdot 61$。1994 年 Alford、Granville、Pomerance 证明了 Carmichael 数有无穷多个。

■ **例 6.5**　**习题集 3 第 4(a) 题：561**　证明 561 是 Carmichael 数。

解　$561 = 3\cdot 11\cdot 17$，无平方因子。$n-1 = 560 = 2^4\cdot 5\cdot 7$。逐个检查 $p-1$：

$$3-1 = 2 \mid 560, \qquad 11-1 = 10 \mid 560, \qquad 17-1 = 16 \mid 560.$$

由 Korselt 判别法，561 是 Carmichael 数。这意味着在 $1,\dots,560$ 中，与 561 互素的 $\varphi(561) = 2\cdot 10\cdot 16 = 320$ 个底数全部是 Fermat 说谎者；Fermat 检验只能靠第 2 步碰巧取到 $3, 11, 17$ 的倍数来揭穿它。561 的素因子很小，这样的底数还不少。但若 $n = pqr$，p, q, r 都是约 100 位的素数，随机的 a 与 n 不互素的概率不超过 $\frac{1}{p}+\frac{1}{q}+\frac{1}{r} \approx 3\cdot 10^{-99}$，这一步几乎从不起作用。　♦

6.2.3　Miller–Rabin 强伪素数检验

改进的出发点是一个只对素数成立的性质。

引理 6.6　若 p 是素数，$x^2 \equiv 1 \pmod p$，则 $x \equiv \pm 1 \pmod p$。

证明　$p \mid x^2-1 = (x-1)(x+1)$，由 Euclid 引理，$p \mid x-1$ 或 $p \mid x+1$。　■

对合数这个性质可能不成立，上一节就看到 $x^2 \equiv 1 \pmod{105}$ 有 8 个解。而一旦找到了 1 的非平凡平方根，就能直接分解 n。

命题 6.7（非平凡平方根给出因子）　若 $x^2 \equiv 1 \pmod n$ 但 $x \not\equiv \pm 1 \pmod n$，则 $(x-1, n)$ 是 n 的非平凡因子（既不是 1 也不是 n）。

证明　$n \mid (x-1)(x+1)$。若 $(x-1,n) = n$，则 $x \equiv 1$；若 $(x-1,n) = 1$，则由 $n \mid (x-1)(x+1)$ 得 $n \mid x+1$（第 1 讲 Euclid 引理的一般形式），即 $x \equiv -1$。两者都被假设排除了。　■

例如模 15 取 $x = 4$：$(3,15) = 3$，$(5,15) = 5$，立刻得到 $15 = 3\cdot 5$。

现在设 n 是奇数，把 $n-1$ 中的因子 2 全部提出来：$n-1 = 2^s d$，d 为奇数。Fermat 检验只看链条的最后一项，强检验则看整条链

$$a^d,\quad a^{2d},\quad a^{4d},\quad \dots,\quad a^{2^s d} = a^{n-1} \pmod n,$$

其中每一项是前一项的平方。如果 n 是素数，最后一项是 1，所以链条中一定有 1。设 1 第一次出现在第 i 项。若 i 是第一项，就是 $a^d \equiv 1$。否则第 $i-1$ 项不等于 1，而它的平方是 1，由引理 6.6，它只能是 -1。此后每一项都是 1 的平方，仍是 1。所以素数的链条只有两种样子：

$$(1, 1, \dots, 1) \qquad \text{或} \qquad (*, \dots, *, -1, 1, \dots, 1).$$

定义 6.8（强伪素数）　设奇合数 n 满足 $n-1=2^s d$（d 为奇数）。若 $a^d \equiv 1 \pmod n$，或存在 $0 \le i < s$ 使 $a^{2^i d} \equiv -1 \pmod n$，则称 n 是**以 a 为底的强伪素数**（strong pseudoprime to base a），a 称为 n 的**强说谎者**（strong liar）。

算法 6.9（Miller–Rabin 检验）　输入奇数 $n>3$，写 $n-1=2^s d$，d 为奇数。

(1) 随机取 $a \in \{2,\dots,n-2\}$，算 $x \equiv a^d \pmod n$。若 $x \equiv \pm 1$，输出「可能是素数」。

(2) 重复至多 $s-1$ 次：$x \leftarrow x^2 \bmod n$。若 $x \equiv -1$，输出「可能是素数」；若 $x \equiv 1$，输出「合数」（上一项是 1 的非平凡平方根）。

(3) 若循环结束仍未输出，输出「合数」。

■ **例 6.6　强检验揭穿 561**　用底数 2 对 561 做 Miller–Rabin 检验。

解　$560 = 2^4 \cdot 35$，$s=4$，$d=35$。反复平方：

指数	35	70	140	280	560
$2^{(\cdot)} \bmod 561$	263	166	67	1	1

最后一项是 1，所以 561 骗过了 Fermat 检验。但 1 第一次出现在指数 280 处，它前面的 67 不是 $-1 \equiv 560$。$67^2 = 4489 = 8 \cdot 561 + 1$，67 是 1 的非平凡平方根，所以 561 是合数。

更进一步，由命题 6.7，$(66,561)=33$，$(68,561)=17$，一下子得到 $561 = 33 \cdot 17$。从中国剩余定理的角度看：$67 \equiv 1 \pmod 3$，$67 \equiv 1 \pmod{11}$，$67 \equiv -1 \pmod{17}$，它是「符号不一致」的平方根。 ♦

■ **例 6.7　同一个 Carmichael 数也能通过强检验**　把底数换成 50，对 561 再做一轮 Miller–Rabin 检验。

解　仍写 $560 = 2^4 \cdot 35$。计算得 $50^{35} \equiv 560 \equiv -1 \pmod{561}$，此后的平方都为 1，所以这一轮输出「可能是素数」。但 $561 = 3 \cdot 11 \cdot 17$ 是合数，故 50 是它的一个强说谎底数。底数 2 能揭穿它，不能推出每个底数都能揭穿它；可靠性要由下面的随机多轮概率界保证。 ♦

同样，对例 6.4 中骗过底数 3 的 $91 : 90 = 2 \cdot 45$，$3^{45} \equiv 27 \pmod{91}$，不是 ± 1，而 $27^2 = 729 \equiv 1$，强检验立即判定 91 是合数，并且 $(26,91)=13$，$(28,91)=7$。

强说谎者比 Fermat 说谎者少得多，而且下面的定理对**所有**奇合数成立，包括 Carmichael 数。

定理 6.10（Rabin–Monier）　若 $n>9$ 是奇合数，则在 $1,\dots,n-1$ 中，n 的强说谎者不超过 $\varphi(n)/4$ 个，从而不超过 $(n-1)/4$ 个。

本课程不证明此定理；证明要用到模 n 乘法群的结构。例如 561 的 $\varphi(561)=320$ 个互素剩余全是 Fermat 说谎者，强说谎者却只有 10 个（连同总会说谎的 1 和 560），远小于上界 80。上

界也可能取到：91 恰有 $\varphi(91)/4 = 18$ 个强说谎者，其中包括 $9, 10, 12$。

于是，对固定的奇合数 $n > 9$，在 $2, \dots, n-2$ 中均匀随机取底数，通过一轮的概率至多 $\frac{1}{4}$：去掉必定通过的 $1, n-1$ 后，说谎者所占比例只会降低。$n = 9$ 可直接检查，也满足此界。独立取 k 个底数，全部通过的概率至多 4^{-k}。取 $k = 50$，这个数约为 8×10^{-31}。每轮只做约 $\log_2 n$ 次模乘法，所以 Miller–Rabin 是多项式时间的随机算法。

易错提醒

- 输出「合数」永远正确；只有「可能是素数」才可能出错。
- 4^{-k} 是**固定合数 n 连续通过 k 轮**的条件概率上界。通过之后 n 是合数的概率还取决于候选数的分布，不能直接等同于这个上界。
- 使用 4^{-k} 界时，底数须**独立、均匀随机**抽取。固定用 $a = 2$ 会被 $2047 = 23 \cdot 89$ 骗过（练习 6）；有限范围内的固定底数判定须另行证明覆盖范围。

6.2.4 AKS：确定性的多项式时间算法

Miller–Rabin 用到了随机性。有没有**不用随机**、又是多项式时间的算法？这个问题直到 2002 年才被解决，出发点是下面这个漂亮的刻画。

定理 6.11　设整数 $n \geq 2$。则 n 是素数，当且仅当作为多项式有

$$(x-1)^n \equiv x^n - 1 \pmod{n},$$

即左边展开后，x^n 与常数项以外的每个系数都被 n 整除，且常数项 $(-1)^n \equiv -1 \pmod{n}$。

证明　$(x-1)^n = \sum_{k=0}^{n} \binom{n}{k} x^k (-1)^{n-k}$。

n 是素数 p。对 $0 < k < p$，$\binom{p}{k} = \frac{p!}{k!\,(p-k)!}$ 的分子含因子 p 而分母不含，所以 $p \mid \binom{p}{k}$（第 3 讲）。常数项 $(-1)^p$：p 奇时为 -1；$p = 2$ 时为 $1 \equiv -1 \pmod 2$。

n 是合数。取 n 的一个素因子 p，设 $p^e \parallel n$（即 $p^e \mid n$，$p^{e+1} \nmid n$），看 x^p 的系数 $\pm\binom{n}{p}$，这里 $0 < p < n$。

$$\binom{n}{p} = \frac{n(n-1)\cdots(n-p+1)}{p!}.$$

分子中 n 含 p^e，其余因子 $n-1, \dots, n-p+1$ 都不被 p 整除（它们模 p 是 $-1, \dots, -(p-1)$）；分母 $p!$ 恰含一个 p。所以 $\binom{n}{p}$ 恰含 p^{e-1}，不被 p^e 整除，更不被 n 整除。这个系数不为 0，同余式不成立。 ∎

例如 $n = 6$：$(x-1)^6$ 中 x^2, x^3 的系数是 $\binom{6}{2} = 15$、$-\binom{6}{3} = -20$，分别不被 2、3 整除，所以 6 是合数。

直接用这个定理并不快：$(x-1)^n$ 有 $n+1$ 个系数。2002 年，Agrawal 和他的两名本科生 Kayal、Saxena 在论文《PRIMES is in P》中想到，只在同时模 n 和模一个低次多项式 $x^r - 1$ 的意义下比较，即对若干个小的 a 检验 $(x-a)^n \equiv x^n - a \pmod{x^r - 1,\ n}$，其中 r 只有 $(\log n)^{O(1)}$ 大小。每个多项式至多 r 项，用反复平方计算，总步数是 $\log n$ 的多项式；难点在于证明截短之

后合数仍一定会露出马脚。这是第一个无条件的确定性多项式时间算法。实际计算中仍然用快得多的 Miller–Rabin。

6.3　因数分解与 Pollard ρ 算法

上一节的检验能告诉我们 n 是合数，却一般不给出因子（除非碰巧遇到 1 的非平凡平方根）。分解是另一个问题：给定合数 n，找到它的一个非平凡因子。试除法仍要约 $\sqrt{n}$ 步。下面介绍一个简单而聪明的改进。

6.3.1　序列模 p 会先进入循环

取 $f(x)=x^2+1$，$x_0=1$，令 $x_{i+1}=f(x_i) \bmod n$。设 p 是 n 的一个（未知的）素因子。因为 $p \mid n$，有 $x_{i+1}\equiv x_i^2+1 \pmod p$，所以序列 $x_i \bmod p$ 是同一个规则在模 p 下的迭代。它只能取 p 个值，最多 $p+1$ 步之后必定出现重复；一旦 $x_i \equiv x_j \pmod p$，此后 $x_{i+1}\equiv x_{j+1}$，序列就进入循环。把它画出来，是一条「尾巴」接一个「圈」，形如希腊字母 ρ（图 6.2）。

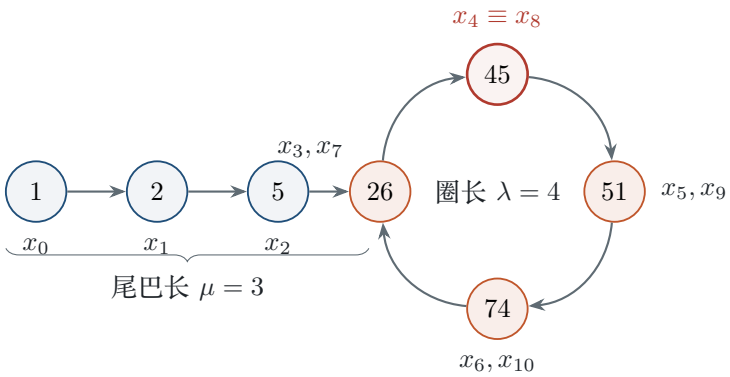


图 6.2　$x_{i+1}=x_i^2+1$，$x_0=1$ 模 79 的轨迹：$1,2,5$ 之后进入圈 $26\to45\to51\to74\to26$。Floyd 判圈法比较 x_k 与 x_{2k}，在 $k=4$ 时第一次相等（红色）。

我们并不知道 p，所以无法直接观察模 p 的序列。但是若 $x_i \equiv x_j \pmod p$ 而 $x_i \not\equiv x_j \pmod n$，则

$$p \mid (x_i - x_j,\ n), \qquad (x_i - x_j,\ n) \neq n,$$

最大公因数就是 n 的一个非平凡因子。要比较的对子太多，Floyd 的判圈法只比较 x_k 与 x_{2k}。

引理 6.12（Floyd 判圈）　设序列 $x_i \bmod p$ 的尾巴长 μ、圈长 λ（即 $x_0,\dots,x_{\mu+\lambda-1}$ 两两不同余，且 $x_{\mu+\lambda}\equiv x_\mu$）。对正整数 k，$x_k \equiv x_{2k} \pmod p$ 当且仅当 $k \geq \mu$ 且 $\lambda \mid k$。特别地，存在 $1 \leq k \leq \mu+\lambda$ 使 $x_k \equiv x_{2k} \pmod p$。

证明　从下标 μ 起序列以 λ 为周期，而尾巴上的项（下标 $<\mu$）只出现一次。若 $k<\mu$，x_k 在尾巴上，不与任何别的项同余。若 $k \geq \mu$，k 与 $2k$ 都在圈上，$x_k \equiv x_{2k}$ 当且仅当 $\lambda \mid 2k-k=k$。在 $\mu+1,\mu+2,\dots,\mu+\lambda$ 这 λ 个连续整数中必有一个是 λ 的倍数，它就是所求的 k。 ■

图 6.2 中 $\mu=3$，$\lambda=4$，第一个满足条件的是 $k=4$。

算法 6.13（Pollard ρ）　输入合数 n。令 $x=y=1$，$f(t)=t^2+1$。重复：

(1) $x \leftarrow f(x) \bmod n$，$y \leftarrow f(f(y)) \bmod n$（于是第 k 轮时 $x=x_k$，$y=x_{2k}$）；

(2) 算 $g=(|x-y|,n)$。若 $1<g<n$，输出因子 g；若 $g=n$，本次失败，换 $f(t)=t^2+c$ 重来；若 $g=1$，继续。

整个过程只需存两个数 x,y，每轮做三次平方和一次欧几里得算法。

■ **例 6.8　分解 8137**　用 Pollard ρ 算法分解 $n=8137$。

解　序列 $x_i \bmod 8137$ 的前几项是 $1,2,5,26,677,2658,2049,7847,2731$。例如 $x_4=26^2+1=677$，$x_5=677^2+1=458330=56\cdot 8137+2658$。逐轮计算：

k	x_k	x_{2k}	$x_{2k}-x_k$	$(x_{2k}-x_k,\ 8137)$
1	2	5	3	1
2	5	677	672	1
3	26	2049	2023	1
4	677	2731	2054	79

第 4 轮得到因子 79（$2054=26\cdot 79$），于是 $8137=79\cdot 103$。

模 79 时 $x_4\equiv x_8\equiv 45$，正是图 6.2 中的红色结点；模 103 时 $\mu=4$，$\lambda=14$，要到 $k=14$ 才相遇，所以最大公因数是 79 而不是 n。 ♦

6.3.2 为什么约 $\sqrt{p}$ 步

如果把 $x_i \bmod p$ 看成从 p 个值中独立随机地抽取，那么前 k 个值两两不同的概率是

$$\left(1-\frac{1}{p}\right)\left(1-\frac{2}{p}\right)\cdots\left(1-\frac{k-1}{p}\right)\approx e^{-(1+2+\cdots+(k-1))/p}\approx e^{-k^2/(2p)},$$

当 k 是 $\sqrt{p}$ 的几倍时，它已经很接近 0。这就是**生日悖论**（birthday paradox）：一年 365 天，只要 23 个人，就有超过一半的概率出现同一天生日（精确值约 50.7%）。所以可以期待 $\mu+\lambda$ 是 $\sqrt{p}$ 量级，Pollard ρ 约 $\sqrt{p}$ 步找到素因子 p。x^2+1 并不真的随机，这只是一个**启发式**的估计，但实际表现与它相符。

习题集 4 第 1 部分第 2 题用这个算法分解 $2^{1231}-1$（371 位），在第 838 轮得到素因子 531793，与 $\sqrt{531793}\approx 729$ 是同一量级。

想一想　为什么不用更简单的 $f(x)=x+1$？

答　$x_i=1+i$，序列模 p 没有尾巴，圈长恰好是 p，要 $k=p$ 才有 $x_k\equiv x_{2k}$。此时 $x_{2k}-x_k=k$，算法相当于依次检验 $(1,n),(2,n),\ldots$，就是试除法。一次函数把模 p 的余数一一打乱，不会「提前撞车」；x^2+c 不是一一的（x 与 $-x$ 的像相同），表现得更像随机函数，才有生日悖论的效果。

更好的分解算法确实存在。椭圆曲线方法（ECM）的启发式步数主要取决于待找素因子 p，量级为 $\exp\big(c\sqrt{\log p \log\log p}\big)$；数域筛法（NFS）的启发式量级为 $\exp\big(c(\log n)^{1/3}(\log\log n)^{2/3}\big)$，两式中的 c 为各自的常数。这些估计优于一般试除法，但没有给出对任意输入的经典多项式时间分解算法。量子计算已有多项式时间分解算法，因此讨论分解的困难时要明确计算模型。

6.4　RSA 公钥密码

6.4.1　共享密钥的麻烦

Alice 想给 Bob 发一条消息，Carol 在窃听。消息转成二进制，就是一个整数 m。一种简单方法是双方预先共享秘密 k，发送 $m+k$（或按位异或 $m\oplus k$），Bob 减去 k（或再异或一次）即可还原。直接重复使用同一掩码会泄露信息：两个密文相减得到 m_1-m_2，异或时得到 $m_1\oplus m_2$。这个缺陷针对上述方法，不能推广到所有共享密钥（shared key）加密方案。密钥还须预先安全分发；若 n 人每对各用一把独立密钥，便需 $\binom{n}{2}$ 把，1000 人需 499500 把。

公钥密码（public key cryptography）的想法是：Bob 公开一把「锁」，谁都可以用它把消息锁上，但只有 Bob 有「钥匙」能打开。这个概念由 Diffie 和 Hellman 在 1976 年提出（第 8 讲），1977 年 Rivest、Shamir、Adleman 用数论实现了它。

6.4.2　RSA 算法

算法 6.14（RSA）　*Bob 生成密钥：*

(1) 随机取两个大小相近且不同的大素数 p,q（用 Miller–Rabin 检验），令 $N=pq$，$\varphi(N)=(p-1)(q-1)$。

(2) 取正整数 $e>1$ 且 $(e,\varphi(N))=1$，用扩展欧几里得算法求正整数 f，使 $ef\equiv 1 \pmod{\varphi(N)}$。

(3) 公开 **公钥** (N,e)；保密 **私钥** f，以及 $p,q,\varphi(N)$。

*Alice 加密：*消息 $0\le m<N$（太长就分段），发送 $c\equiv m^e \pmod N$。
*Bob 解密：*计算 $c^f \bmod N$，它等于 m。

定理 6.15（RSA 的正确性）　在算法 6.14 中，若 $(m,N)=1$，则 $(m^e)^f\equiv m \pmod N$。

证明　$ef\equiv 1 \pmod{\varphi(N)}$，写 $ef=1+k\varphi(N)$，$k\ge 0$。由 Euler 定理 $m^{\varphi(N)}\equiv 1 \pmod N$，所以

$$(m^e)^f = m^{1+k\varphi(N)} = m\cdot\big(m^{\varphi(N)}\big)^k \equiv m\cdot 1^k = m \pmod N.$$ ■

条件 $(m,N)=1$ 其实可以去掉（练习 9）。在 $0\le m<N$ 中，不互素的比例为 $\frac{1}{p}+\frac{1}{q}-\frac{1}{N}$；若 $0<m<N$ 且不互素，计算 (m,N) 就能得到非平凡因子。$m=0$ 单独处理，不能用它分解 N。

图 6.3 标出公开参数、秘密参数和密文的传递方向。

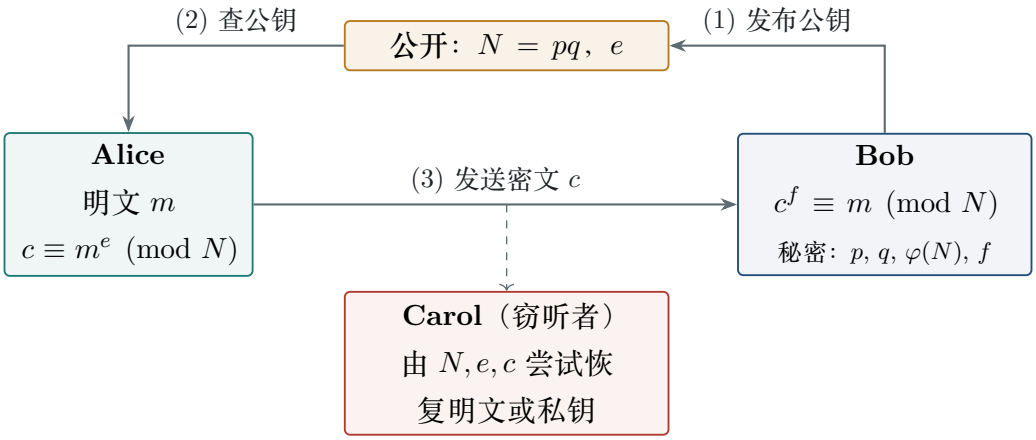


图 6.3　RSA 的流程。已知 p, q 可算 $\varphi(N)$ 和私钥 f；反过来，已知 $N, \varphi(N)$ 可分解 N。这没有证明任意恢复明文的方法都必须分解 N。

■ 例 6.9　**一组完整的小参数**　取 $p = 61$，$q = 53$，$e = 17$。求私钥 f，加密 $m = 65$，再解密验证。

解　生成密钥。$N = 61 \cdot 53 = 3233$，$\varphi(N) = 60 \cdot 52 = 3120$。用第 2 讲的扩展欧几里得算法：$3120 = 183 \cdot 17 + 9$，$17 = 1 \cdot 9 + 8$，$9 = 1 \cdot 8 + 1$，倒推得 $1 = 9 - 8 = 2 \cdot 9 - 17 = 2 \cdot 3120 - 367 \cdot 17$，所以 $f \equiv -367 \equiv 2753 \pmod{3120}$。验算：$17 \cdot 2753 = 46801 = 15 \cdot 3120 + 1$。公钥 $(3233, 17)$，私钥 2753。

加密。$17 = 16 + 1$，只需反复平方四次：

$$65^2 \equiv 992, \quad 65^4 \equiv 992^2 \equiv 1232, \quad 65^8 \equiv 1232^2 \equiv 1547, \quad 65^{16} \equiv 1547^2 \equiv 789 \pmod{3233},$$

$c \equiv 65^{16} \cdot 65 \equiv 789 \cdot 65 = 51285 \equiv 2790 \pmod{3233}$。Alice 发送 $c = 2790$。

解密。Bob 要算 $2790^{2753} \bmod 3233$。他知道 p, q，可以用中国剩余定理分开算。模 61：$c \equiv 45$，由 Fermat 小定理指数可以模 60 约化，$2753 \equiv 53 \pmod{60}$，得 $45^{53} \equiv 4 \pmod{61}$。模 53：$c \equiv 34$，指数模 52 约化，$2753 \equiv 49 \pmod{52}$，得 $34^{49} \equiv 12 \pmod{53}$。再解 $x \equiv 4 \pmod{61}$，$x \equiv 12 \pmod{53}$：$x = 4 + 61t$，$4 + 8t \equiv 12 \pmod{53}$，$t = 1$，$x = 65$。解密得到 $m = 65$。这样算的数只有原来一半长，实际的 RSA 实现都这样做。 ♦

6.4.3 安全性：知道 $\varphi(N)$ 就能分解 N

Carol 看得到 N, e, c。如果她能分解 N，就能算 $\varphi(N) = (p-1)(q-1)$，再像 Bob 一样求出 f。反过来，下面的例子说明，求 $\varphi(N)$ 并不比分解 N 容易。

■ 例 6.10　**习题集 4 第 1 部分第 3 题**　设 $N = pq$，p, q 是不同素数。已知 N 和 $M = \varphi(N)$，怎样求 p, q？对 $N = 3233$，$M = 3120$ 实际算一遍。

解　$M = (p-1)(q-1) = pq - (p+q) + 1 = N - (p+q) + 1$，所以

$$p + q = N - M + 1, \qquad pq = N.$$

两数之和与积都已知，它们就是二次方程 $X^2 - (N - M + 1)X + N = 0$ 的两个根：

$$p,\ q=\frac{s\pm\sqrt{s^2-4N}}{2},\qquad s=N-M+1.$$

对 $N=3233$，$M=3120$：$s=114$，$s^2-4N=12996-12932=64$，$\sqrt{64}=8$，所以 $p,q=\frac{114\pm8}{2}=61,53$。

原题给的是一个 71 位的 N 和它的 $\varphi(N)$，同样只需一次开平方，就得到 $p=2^{127}-1$，$q=2^{107}-1$。♦

想一想 RSA 为什么要求 $(e,\varphi(N))=1$？对上例的 $N=3233$ 改用 $e=3$ 会出什么问题？

答 若 $(e,\varphi(N))>1$，e 模 $\varphi(N)$ 没有逆元，私钥 f 根本不存在。更具体地，$3\mid 60=p-1$，模 61 有非平凡的立方根 13（$13^3=2197=36\cdot61+1$）。用中国剩余定理取 $m'\equiv13\pmod{61}$，$m'\equiv1\pmod{53}$，得 $m'=1538$，它满足 $1538^3\equiv1=1^3\pmod{3233}$。两条不同的消息 1 和 1538 加密成同一个密文，谁也无法解密。

易错提醒

- 私钥是 e 模 $\varphi(N)$ 的逆元，不是模 N 的逆元。指数按 $\varphi(N)$ 约化，底数才按 N 约化。
- 上面的 RSA 只是「教科书版本」，直接使用并不安全。例如 $e=3$ 而 $m<N^{1/3}$ 时，$m^3<N$，密文就是整数 m^3 本身，开立方即得 m；同一条消息每次加密结果都相同，也会泄露信息。实际使用时要先对 m 做随机填充。

6.5 Hensel 引理

第 6.1 节把解 $f(x)\equiv0\pmod m$ 归结为模素数幂 p^e 的问题。逐个试要试 p^e 次。Hensel 引理先解模 p，再一层层「提升」到模 $p^2,p^3,\dots$，每层只需解一个模 p 的一次同余式。

6.5.1 从模 7 到模 343

■ **例 6.11** **把 $x^2\equiv2$ 的根一层层提升**　解 $x^2\equiv2\pmod{49}$ 和 $x^2\equiv2\pmod{343}$。

解 记 $f(x)=x^2-2$。模 7：逐个试 $0,\dots,6$，平方模 7 为 $0,1,4,2,2,4,1$，所以根是 $x\equiv3,4\pmod7$。

模 49。若 $x^2\equiv2\pmod{49}$，则更有 $x^2\equiv2\pmod7$，所以 x 模 7 必是 3 或 4。先看 $x\equiv3\pmod7$，即 $x=3+7t$，$t=0,1,\dots,6$：

t	0	1	2	3	4	5	6
$x=3+7t$	3	10	17	24	31	38	45
$x^2-2 \bmod 49$	7	**0**	42	35	28	21	14

只有 $t=1$ 成立，$10^2-2=98=2\cdot49$。表中第三行是公差为 -7 的等差数列，这不是巧合：

$$f(3+7t)=9+42t+49t^2-2\equiv7+42t=7(1+6t)\pmod{49}.$$

t^2 项含 49，直接消失，剩下的是 t 的一次式。要它被 49 整除，只需 $1+6t \equiv 0 \pmod 7$，即 $6t \equiv -1$，$t \equiv 1 \pmod 7$。不用列表也能得到 $t = 1$。对另一个根，$x \equiv 4 \pmod 7$ 提升为 $x \equiv 39 = 49 - 10$。所以模 49 的解是 $x \equiv 10, 39$。

模 343。把 10 提升：$x = 10 + 49t$，

$$f(10 + 49t) = 98 + 980t + 49^2 t^2 \equiv 98 + 980t = 49(2 + 20t) \pmod{343},$$

需要 $2 + 20t \equiv 0 \pmod 7$，即 $6t \equiv 5$，$t \equiv 2 \pmod 7$，$x = 10 + 98 = 108$。验算：$108^2 - 2 = 11662 = 34 \cdot 343$。同理 39 提升为 $343 - 108 = 235$。所以模 343 的解是 $x \equiv 108, 235$。 ♦

6.5.2 引理的陈述

上例每一层都发生了同样的事：代入 $a + tp^j$ 后，含 t^2 的项消失，t 的系数是 p^j 乘 $f'(a)$（$f'(3) = 6$，$f'(10) = 20$）。这里 f' 是 f 的形式导数（formal derivative），即逐项求导，系数仍是整数。一般情形如下，下文 f 总是整系数多项式，p 是素数。

定理 6.16（Hensel 引理）　设 $j \geq 1$，$f(a) \equiv 0 \pmod{p^j}$，且 $f'(a) \not\equiv 0 \pmod p$。则存在唯一的 $t \in \{0, 1, \ldots, p-1\}$，使

$$f(a + tp^j) \equiv 0 \pmod{p^{j+1}}.$$

也就是说，模 p^{j+1} 恰有一个根 b 满足 $b \equiv a \pmod{p^j}$，称 b 是 a 的提升（lift）。

Hensel 引理的完整证明见第 7 讲；这里先给出提升公式的推导。

引理 6.17　对整数 a, t 和 $j \geq 1$，有 $f(a + tp^j) \equiv f(a) + tp^j f'(a) \pmod{p^{j+1}}$。

这个「截断的 Taylor 展开」只需对单项式 $f(x) = x^n$ 验证，再按系数相加。由二项式定理，

$$(a + tp^j)^n = a^n + na^{n-1} \cdot tp^j + \binom{n}{2} a^{n-2} \cdot t^2 p^{2j} + \cdots,$$

从第三项起每一项都含 p^{2j}，而 $j \geq 1$ 时 $2j \geq j + 1$，这些项都被 p^{j+1} 整除；前两项正是 $a^n + tp^j \cdot na^{n-1}$。

有了引理，定理的思路就清楚了。写 $f(a) = p^j s$，则 $f(a + tp^j) \equiv p^j \big(s + tf'(a)\big) \pmod{p^{j+1}}$。它被 p^{j+1} 整除当且仅当

$$\boxed{f'(a)\, t \equiv -\frac{f(a)}{p^j} \pmod p.}$$

这是 t 的一次同余式，$f'(a)$ 模 p 可逆，所以模 p 恰有一个解。完整的证明和由此得到的迭代公式放在第 7 讲。

想一想　引理为什么要求 $j \geq 1$？取 $j = 0$ 会怎样？

答　$j = 0$ 时 $p^{2j} = 1$，二次项不再被 p 整除，不能丢掉。例如 $f(x) = x^2$，$a = 0$，$t = 1$：$f(1) = 1$，而 $f(0) + f'(0) = 0$。从模 $p^0 = 1$ 到模 p 这一步只能直接解。

推论 6.18　若 $f(a) \equiv 0 \pmod{p}$ 且 $f'(a) \not\equiv 0 \pmod{p}$，则对每个 $e \geq 1$，模 p^e 恰有一个根 b 满足 $b \equiv a \pmod{p}$。特别地，若 f 模 p 的根都满足 $f'(a) \not\equiv 0$，则 f 模 p^e 的根数与模 p 的根数相同。

证明　对 e 归纳。设 b_e 是模 p^e 的根且 $b_e \equiv a \pmod{p}$，则 $f'(b_e) \equiv f'(a) \not\equiv 0 \pmod{p}$，定理 6.16 给出唯一的提升 b_{e+1}，它模 p 仍与 a 同余。反过来，模 p^{e+1} 的根模 p^e 也是根，所以每个根都是这样逐层得到的。 ■

例 6.11 中 $f'(3) = 6$、$f'(4) = 8$ 模 7 都不为 0，由推论，$x^2 \equiv 2 \pmod{7^e}$ 对每个 e 都恰有 2 个解。把推论与定理 6.1 合起来，求解 $f(x) \equiv 0 \pmod{m}$ 就基本归结为模素数 p 的问题，这正是第 6.1 节「三层结构」的最后一层。

易错提醒　条件 $f'(a) \not\equiv 0 \pmod{p}$ 不能去掉。若 $f'(a) \equiv 0 \pmod{p}$（称 a 为**奇异根**（singular root）），由引理 6.17，$f(a + tp^j) \equiv f(a) \pmod{p^{j+1}}$ 与 t 无关：要么 p 个候选全是根，要么一个都不是。例如 $p = 2$，$a = 1$：

- $f(x) = x^2 + 1$，$f(1) = 2$，$f'(1) = 2 \equiv 0$。2 不被 4 整除，所以 1, 3 都不是模 4 的根，提升**不存在**。
- $f(x) = x^2 - 1$，$f(1) = 0$，$f'(1) \equiv 0$。1, 3 都是模 4 的根，提升**不唯一**。

奇异根的完整处理见第 7 讲。

前后联系　第 7 讲补完 Hensel 引理的证明，把提升写成迭代公式 $a \mapsto a - f(a)\,\overline{f'(a)}$（$\overline{f'(a)}$ 是 $f'(a)$ 模 p 的逆元），它在形式上与 Newton 法 $x \mapsto x - f(x)/f'(x)$ 完全一样；然后转向最底层的问题：模素数的同余式最多有几个根。

本讲小结

- 中国剩余定理：模 $m=\prod p_i^{e_i}$ 的解与各模 $p_i^{e_i}$ 的解组一一对应，解数相乘；奇数 n 有 r 个不同素因子时，$x^2\equiv 1 \pmod n$ 有 2^r 个解。
- 快慢要和输入长度 $\log n$ 比：试除法约 $\sqrt{n}$ 步，是指数的；反复平方约 $\log_2 n$ 步，是多项式的。
- Fermat 检验会被 Carmichael 数骗过（Korselt：无平方因子且 $p-1\mid n-1$）。Miller–Rabin 检查链条 $a^d, a^{2d}, \dots, a^{n-1}$，利用「素数模下 1 的平方根只有 ± 1」，每轮出错概率至多 $\frac{1}{4}$。AKS 基于 $(x-1)^n\equiv x^n-1 \pmod n$，是确定性多项式时间的。
- 非平凡平方根 x 给出因子 $(x-1,n)$。Pollard ρ 比较 x_k 与 x_{2k}，找到素因子 p 的启发式步数约为 $\sqrt{p}$。
- RSA：公钥 (N,e)，私钥 $f\equiv e^{-1} \pmod{\varphi(N)}$，加密 m^e，解密 c^f。知道 $\varphi(N)$ 等价于能分解 N。
- Hensel 引理：$f(a)\equiv 0 \pmod{p^j}$，$f'(a)\not\equiv 0 \pmod p$ 时，a 恰有一个模 p^{j+1} 的提升；关键是 $f(a+tp^j)\equiv f(a)+tp^j f'(a) \pmod{p^{j+1}}$。

练习

6.1 ★ 求 $x^2\equiv -1 \pmod{65}$ 的全部解。（先分别解模 5 和模 13。）

6.2 ★★ $x^2\equiv 1 \pmod{360}$ 有多少个解？为什么模 8 的那一部分不符合「每个奇素数幂贡献 2 个解」的规律？用 Hensel 引理的条件解释。

6.3 ★ 用 Korselt 判别法证明 1105 和 1729 都是 Carmichael 数。

6.4 ★★ （习题集 3 第 4(b) 题）利用 Korselt 判别法的逆命题（Carmichael 数无平方因子，且每个素因子 p 满足 $p-1\mid n-1$），证明：Carmichael 数都是奇数，且至少有三个不同的素因子。（提示：试 $a=-1$；若 $n=pq$，$p<q$，看 $q-1\mid n-1$。）

6.5 ★★ 设 n 是合数，且至少有一个 a 满足 $(a,n)=1$、$a^{n-1}\not\equiv 1 \pmod n$。证明：在与 n 互素的 $\varphi(n)$ 个剩余中，Fermat 说谎者至多占一半。

6.6 ★★ 证明 $2047=23\cdot 89$ 是以 2 为底的强伪素数（提示：$2^{11}=2048$），再用底数 3 做一轮 Miller–Rabin 检验。

6.7 ★ 用 Pollard ρ（$f(x)=x^2+1$，$x_0=1$）分解 $n=221$，并说明为什么在第 4 轮找到的是因子 13 而不是 17。

6.8 ★ (a) RSA 中取 $p=5$，$q=11$，$e=3$。求私钥 f，加密 $m=7$，再解密验证。(b) 已知 $N=10403$ 是两个素数之积，且 $\varphi(N)=10200$，分解 N。

6.9 ★★★ 证明 RSA 对所有 $0\le m<N$ 都正确，包括 $(m,N)>1$ 的情形：$m^{ef}\equiv m \pmod N$。（提示：分别证明模 p 和模 q 成立，再用中国剩余定理。）

6.10 ★★ 求 $x^2\equiv 7 \pmod{27}$ 的全部解，并证明对每个 $e\ge 1$，$x^2\equiv 7 \pmod{3^e}$ 恰有 2 个解。

练习答案与提示

6.1 模 5：$x^2 \equiv 4$，$x \equiv 2, 3$；模 13：$x^2 \equiv 12$，$x \equiv 5, 8$（$5^2 = 26 - 1$，$8^2 = 65 - 1$）。四种组合用中国剩余定理合成：$(2,5) \to 57$，$(2,8) \to 47$，$(3,5) \to 18$，$(3,8) \to 8$。解为 $x \equiv 8, 18, 47, 57 \pmod{65}$。验算：$8^2 = 64 \equiv -1$。

6.2 $360 = 8 \cdot 9 \cdot 5$。模 8 有 $1, 3, 5, 7$ 共 4 个解，模 9 有 ± 1，模 5 有 ± 1，共 $4 \cdot 2 \cdot 2 = 16$ 个。$f'(x) = 2x$：奇素数 p 时 $f'(\pm 1) \not\equiv 0 \pmod{p}$，由推论 6.18 模 p^e 恰有 2 个解；$p = 2$ 时根都是奇异根，推论不适用。

6.3 $1105 = 5 \cdot 13 \cdot 17$，$1104 = 2^4 \cdot 3 \cdot 23$，$4, 12, 16$ 都整除 1104。$1729 = 7 \cdot 13 \cdot 19$，$1728 = 2^6 \cdot 3^3$，$6, 12, 18$ 都整除 1728。两数都无平方因子，由 Korselt 判别法得证。

6.4 若 n 是偶合数，则 $n \geq 4$，取 $a = -1$，$(a, n) = 1$，而 $(-1)^{n-1} = -1 \not\equiv 1 \pmod{n}$，矛盾，所以 n 是奇数。n 是合数且无平方因子，所以至少有两个不同的素因子。若恰有两个，$n = pq$，$p < q$，则 $q - 1 \mid n - 1 = p(q-1) + (p-1)$，于是 $q - 1 \mid p - 1$，但 $0 < p - 1 < q - 1$，矛盾。素数不是合数，所以至少三个素因子。

6.5 设 $b_1, \ldots, b_k$ 是全部 Fermat 说谎者（与 n 互素）。则 $ab_1, \ldots, ab_k$ 模 n 互不相同（a 可逆），都与 n 互素，且 $(ab_i)^{n-1} \equiv a^{n-1} \not\equiv 1$，都不是说谎者。所以非说谎者至少 k 个，$2k \leq \varphi(n)$。

6.6 $2046 = 2 \cdot 1023$，$s = 1$，$d = 1023 = 11 \cdot 93$。$2^{11} = 2048 \equiv 1$，所以 $2^{1023} = (2^{11})^{93} \equiv 1 \pmod{2047}$，2 是强说谎者。底数 3：$3^{1023} \equiv 1565 \pmod{2047}$，不是 ± 1，而 $s - 1 = 0$，没有可平方的步骤，输出「合数」。事实上 $3^{2046} \equiv 1013$，连 Fermat 检验都通不过。

6.7 序列模 221：$1, 2, 5, 26, 14, 197, 135, 104, 209, \ldots$。$k = 1, 2, 3$ 时最大公因数都是 1；$k = 4$：$(209 - 14, 221) = (195, 221) = 13$，$221 = 13 \cdot 17$。模 13 序列为 $1, 2, 5, 0, 1, \ldots$，$\mu = 0$，$\lambda = 4$，由引理 6.12 $k = 4$ 时相遇；模 17 为 $1, 2, 5, 9, 14, 10, 16, 2, \ldots$，$\mu = 1$，$\lambda = 6$，要到 $k = 6$。

6.8 (a) $\varphi(55) = 40$，$3 \cdot 27 = 81 = 2 \cdot 40 + 1$，$f = 27$。$c = 7^3 = 343 = 6 \cdot 55 + 13 \equiv 13$。解密：模 5，$13^{27} \equiv 3^{27} \equiv 3^3 \equiv 2$；模 11，$13^{27} \equiv 2^{27} \equiv 2^7 = 128 \equiv 7$。$x \equiv 2 \pmod{5}$，$x \equiv 7 \pmod{11}$ 给出 $x \equiv 7$。(b) $s = N - \varphi(N) + 1 = 204$，$s^2 - 4N = 41616 - 41612 = 4$，$p, q = \frac{204 \pm 2}{2} = 103, 101$。验算 $103 \cdot 101 = 10403$。

6.9 $ef = 1 + k(p-1)(q-1)$，$k \geq 0$。模 p：若 $p \nmid m$，由 Fermat 小定理 $m^{ef} = m \cdot \left(m^{p-1}\right)^{k(q-1)} \equiv m$；若 $p \mid m$，两边都 $\equiv 0$。所以总有 $m^{ef} \equiv m \pmod{p}$，同理模 q 也成立。$p \neq q$ 互素，由中国剩余定理（或第 1 讲的推论「$a \mid n$，$b \mid n$，$(a, b) = 1 \Rightarrow ab \mid n$」），$m^{ef} \equiv m \pmod{N}$。

6.10 $f(x) = x^2 - 7$ 的模 3 根为 $1, 2$，且 $f'(1) = 2$、$f'(2) = 4$ 均不被 3 整除。由推论 6.18，每个模 3 根在模 3^e 下恰有一个提升，故每个 $e \geq 1$ 恰有两个解。求模 27 的根：令 $x = 1 + 3t$，模 9 的条件化为 $2t - 2 \equiv 0 \pmod{3}$，得 $t = 1$、$x = 4$。再令 $x = 4 + 9t$，模 27 的条件化为 $1 + 8t \equiv 0 \pmod{3}$，也得 $t = 1$、$x = 13$。因 $f(-x) = f(x)$，另一个根为 $-13 \equiv 14 \pmod{27}$，所以全部解为 $x \equiv 13, 14 \pmod{27}$。

第 7 讲

Hensel 引理续、素数模同余式与原根

Hensel's Lemma (cont.), Congruences mod Primes, Primitive Roots

对应材料：MIT 18.781 Lecture 7 讲义（Hensel's Lemma, Congruences mod Primes, Order, Primitive Roots）；习题集 4 第 1 部分第 1、4、5 题，第 2 部分第 2、3、4、5 题。

本讲目标

(1) 补完 Hensel 引理的证明，并把它改写成可以反复迭代的公式；会把模 p 的根一步步提升到模 p^e，理解它与 Newton 迭代的关系。

(2) 会处理 $f'(a) \equiv 0 \pmod{p}$ 的「奇异根」：判断它是全部提升还是全部消失，并画出根的提升树。

(3) 会把模素数的同余式化简为首一、次数小于 p 的形式；掌握 Lagrange 定理（n 次同余式模 p 至多 n 个根），知道它在合数模下为什么失效。

(4) 掌握判别法「f 恰有 n 个根 $\iff f \mid x^p - x$」，并由它推出 $x^d \equiv 1$ 的根数、Wilson 定理和初等对称多项式的同余性质。

(5) 理解阶 $\mathrm{ord}_m(a)$ 的定义与三条基本引理；理解原根的定义，并证明每个素数都有原根。

第 6 讲把求解 $f(x) \equiv 0 \pmod{m}$ 的问题拆成了两步：先用中国剩余定理把 m 分解成素数幂 p^e，再用 Hensel 引理把模 p 的根「提升」到模 p^e。那里已推导提升公式，本讲补全引理证明，并把公式写成可重复迭代的算法。之后问题归结到模素数 p：同余式最多有几个根，何时能取到上界？最后引入阶和原根，描述模 p 乘法的结构。

记号。同第 6 讲，(a, m) 是 $\gcd(a, m)$ 的简写。

7.1 Hensel 引理：证明与迭代

先回顾第 6 讲的陈述。下文 f 总是整系数多项式，p 是素数，f' 是 f 的形式导数（逐项求导，系数仍是整数）。

定理 7.1（Hensel 引理） 设 $j \geq 1$，$f(a) \equiv 0 \pmod{p^j}$，且 $f'(a) \not\equiv 0 \pmod{p}$。则存在唯一的 $t \in \{0, 1, \ldots, p-1\}$，使

$$f(a + tp^j) \equiv 0 \pmod{p^{j+1}}.$$

换句话说，模 p^j 的根 a 恰好有一个模 p^{j+1} 的「提升」，它模 p^j 仍与 a 同余。

证明见下面的 Taylor 引理及其后的定理证明。关键是展开后从二次项起均被目标模数整除。

引理 7.2　对整系数多项式 f 和整数 a, t，以及 $j \geq 1$，有

$$f(a + tp^j) \equiv f(a) + tp^j f'(a) \pmod{p^{j+1}}.$$

证明　由线性，只需对单项式 $f(x) = x^n$ 验证。二项式定理给出

$$(a + tp^j)^n = a^n + na^{n-1}\, tp^j + \binom{n}{2} a^{n-2}\, t^2 p^{2j} + \cdots.$$

从第三项起，每一项都含因子 p^{2j}，而 $j \geq 1$ 时 $2j \geq j + 1$，所以它们都被 p^{j+1} 整除。前两项恰好是 $a^n + tp^j \cdot (x^n)'|_{x=a}$。二项式系数都是整数，这一点保证了「丢掉的项」确实是 p^{j+1} 的整数倍。 ■

定理 7.1 的证明　因为 $p^j \mid f(a)$，可以写 $f(a) = p^j s$，$s \in \mathbb{Z}$。由引理 7.2，

$$f(a + tp^j) \equiv p^j \big(s + t f'(a)\big) \pmod{p^{j+1}}.$$

右边被 p^{j+1} 整除，当且仅当 $p \mid s + tf'(a)$，即

$$f'(a)\, t \equiv -s \pmod{p}.$$

这是关于 t 的一次同余式。$f'(a) \not\equiv 0 \pmod{p}$ 意味着 $(f'(a), p) = 1$，由第 5 讲关于一次同余式的结论，它模 p 恰有一个解：

$$\boxed{\,t \equiv -\frac{f(a)}{p^j} \cdot \overline{f'(a)} \pmod{p},\,}$$

其中 $\overline{f'(a)}$ 表示 $f'(a)$ 模 p 的逆元。这个 t 在 $0, 1, \ldots, p-1$ 中唯一。 ■

证明说明了「提升」在做什么：模 p^j 的根 a 已经确定了答案的前 j 位 p 进制数字，提升就是确定下一位数字 t。这一位只依赖 $f(a)/p^j$ 和 $f'(a)$ 的模 p 值，所以只需解一个模 p 的一次方程。

与 Newton 法的类比　实数上求 $f(x) = 0$ 的 Newton 法是 $x_{\text{新}} = x - \dfrac{f(x)}{f'(x)}$：用切线代替曲线，把误差 $f(x)$ 除以斜率 $f'(x)$ 修正掉。Hensel 引理是同一件事的 p 进版本。把 t 代回，新的根是

$$a + tp^j \equiv a - f(a) \cdot \overline{f'(a)} \pmod{p^{j+1}},$$

形式和 Newton 公式完全一样。实数上「误差小」指 $|f(x)|$ 小；这里「误差小」指 $f(a)$ 被 p 的高次幂整除。条件 $f'(a) \not\equiv 0$ 对应 Newton 法中「切线不能水平」。

上面的公式里，逆元 $\overline{f'(a)}$ 只需要模 p。下一步提升时，按道理应该换成 $f'(a + tp^j)$ 的逆元。但 $a + tp^j \equiv a \pmod{p}$，所以 $f'(a + tp^j) \equiv f'(a) \pmod{p}$，逆元可以一直沿用。于是得到一个非常省事的算法。

算法 7.3（Hensel 迭代）　设 $f(a_1)\equiv 0 \pmod p$，$f'(a_1)\not\equiv 0 \pmod p$。取定整数 u 使 $uf'(a_1)\equiv 1 \pmod p$，令

$$a_{j+1}\equiv a_j - f(a_j)\,u \pmod{p^{j+1}},\qquad j=1,2,\dots$$

则 $f(a_j)\equiv 0 \pmod{p^j}$，$a_{j+1}\equiv a_j \pmod{p^j}$，且 a_j 是模 p^j 下唯一满足 $a_j\equiv a_1 \pmod p$ 的根。

证明　对 j 归纳。设 $f(a_j)=p^j s_j$。所有 a_j 模 p 都与 a_1 同余，因此 $f'(a_j)\equiv f'(a_1)$，u 也是 $f'(a_j)$ 模 p 的逆元。由引理 7.2（取 $t=-s_j u$），

$$f(a_{j+1})\equiv f(a_j)-s_j u p^j f'(a_j)=p^j s_j\bigl(1-uf'(a_j)\bigr) \pmod{p^{j+1}},$$

括号里被 p 整除，所以 $p^{j+1}\mid f(a_{j+1})$。唯一性就是定理 7.1 中 t 的唯一性，逐层使用即可。 ■

■ **例 7.1**　$x^2\equiv -1 \pmod{125}$　求 $x^2+1\equiv 0 \pmod{125}$ 的全部解，并进一步求模 625 的解。

解　令 $f(x)=x^2+1$，$f'(x)=2x$。先解模 5：把 $x=0,1,2,3,4$ 代入，$f(x) \bmod 5$ 依次为 $1,2,0,0,2$，根是 2 和 3。$f'(2)=4$，$f'(3)=6\equiv 1$，模 5 都不为 0，所以两个根都可以用算法 7.3 提升。

从 $a_1=2$ 出发。$f'(2)=4\equiv -1 \pmod 5$，取 $u=-1$，迭代公式是 $a_{j+1}\equiv a_j+f(a_j)$：

- $a_2\equiv 2+f(2)=2+5=7 \pmod{25}$。验证：$7^2+1=50=2\cdot 25$。
- $a_3\equiv 7+f(7)=7+50=57 \pmod{125}$。验证：$57^2+1=3250=26\cdot 125$。

从 $a_1=3$ 出发。$f'(3)\equiv 1$，取 $u=1$，迭代公式是 $a_{j+1}\equiv a_j-f(a_j)$：

- $a_2\equiv 3-10=-7\equiv 18 \pmod{25}$。验证：$18^2+1=325=13\cdot 25$。
- $a_3\equiv 18-325=-307\equiv 68 \pmod{125}$。验证：$68^2+1=4625=37\cdot 125$。

模 125 的任何根模 5 都是 2 或 3，而每个模 5 的根只有唯一的提升，所以模 125 恰有两个解：$x\equiv 57$，$x\equiv 68 \pmod{125}$。二者之和是 125，这符合预期：x 是根时 $-x$ 也是根。

再走一步：$a_4\equiv 57+3250=3307\equiv 182 \pmod{625}$，$a_4\equiv 68-4625=-4557\equiv 443 \pmod{625}$。同样 $182+443=625$。 ♦

把每一步的 $t=(a_{j+1}-a_j)/p^j$ 记下来，就得到根的 5 进制数字。从 2 出发的链上 t 依次是 $1,2,1$，所以 $182=2+1\cdot 5+2\cdot 5^2+1\cdot 5^3=(1212)_5$。图 7.1 把两条链画在一起：每提升一次，就在 5 进制展开的左边补上一位数字，已有的数字不再改变。无限提升下去，得到的是 $x^2=-1$ 在 5 进数中的两个解；本讲不需要这个概念，只把它当作一幅图景。

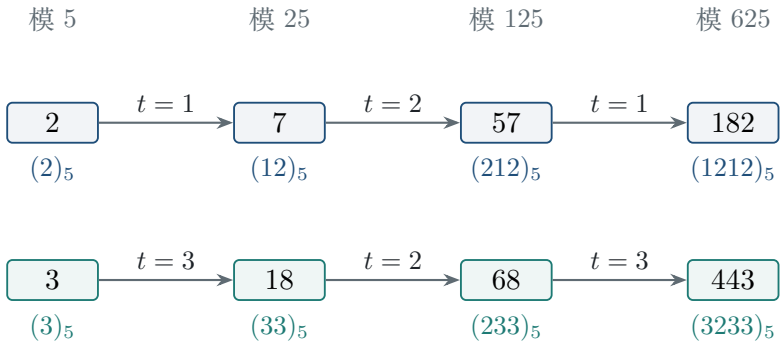


图 7.1　$x^2 \equiv -1$ 的根从模 5 提升到模 625。每个箭头上的 t 是新增的 5 进制数字，下方是根的 5 进制写法。两条链在每一层都互为相反数。

■ **例 7.2　−3 模 1024 的立方根，改编自习题集 4 第 1 部分**　求 $x^3 \equiv -3 \pmod{1024}$ 的全部解。

解　令 $f(x) = x^3 + 3$，$1024 = 2^{10}$。模 2 只有一个根 $x \equiv 1$（$f(0) = 3$ 是奇数，$f(1) = 4$ 是偶数）。$f'(x) = 3x^2$，$f'(1) = 3$ 是奇数，所以这是非奇异根，模 2 的逆元取 $u = 1$，迭代公式为

$$a_{j+1} \equiv a_j - f(a_j) \pmod{2^{j+1}}.$$

表 7.1 记录了全部十步。第 3 列写出 $f(a_j)$ 恰好被 2 的几次幂整除。

表 7.1　$x^3 + 3 \equiv 0$ 的根从模 2 提升到模 2^{10}

j	$a_j \pmod{2^j}$	$f(a_j)$	下一步 $a_{j+1} \equiv a_j - f(a_j) \pmod{2^{j+1}}$
1	1	$4 = 2^2$	$1 - 4 = -3 \equiv 1 \pmod 4$
2	1	$4 = 2^2$	$1 - 4 = -3 \equiv 5 \pmod 8$
3–6	5	$128 = 2^7$	$128 \equiv 0 \pmod{2^{j+1}}$，不变
7	5	$128 = 2^7$	$5 - 128 = -123 \equiv 133 \pmod{256}$
8	133	$2352640 = 2^9 \cdot 4595$	$f(133) \equiv 0 \pmod{512}$，不变
9	133	$2352640 \equiv 512 \pmod{1024}$	$133 - 512 = -379 \equiv 645 \pmod{1024}$
10	645	$268336128 = 2^{10} \cdot 262047$	完成

所以 $x \equiv 645 \pmod{1024}$。模 2 的根唯一，每次提升又唯一，因此 645 是唯一解。 ♦

表中 a_j 有好几步原地不动。原因是 $f(5) = 128$ 碰巧比需要的更「小」：$5^3 = 125 = 128 - 3$，所以 5 直接就是模 128 的根。习题集原题的提示正是这个意思：从一个「离立方数差 3」的 2 的方幂（这里是 128）开始，可以省掉前面几步。

同样的论证给出习题集 4 第 1 部分第 1(a) 题：$x^3 - 1 \equiv 0 \pmod{1024}$ 模 2 只有根 1，且 $(x^3 - 1)' = 3x^2$ 在 $x = 1$ 处是奇数，所以模 1024 只有一个解。代入 $x = 1$ 有 $1^3 - 1 = 0$，故唯一解为 $x \equiv 1 \pmod{1024}$。

7.2　奇异根：全部提升或全部消失

Hensel 引理要求 $f'(a)\not\equiv 0 \pmod p$。如果 $f'(a)\equiv 0 \pmod p$，称 a 是奇异根 (singular root)。这时一次同余式 $f'(a)t\equiv -s \pmod p$ 的系数是 0，变成 $0\equiv -s$，要么对所有 t 成立，要么对所有 t 都不成立。

> **命题 7.4（奇异根的提升）**　设 $j\geq 1$，$f(a)\equiv 0 \pmod{p^j}$，$f'(a)\equiv 0 \pmod p$。
>
> (1) 若 $f(a)\equiv 0 \pmod{p^{j+1}}$，则 a 的全部 p 个提升 $a+tp^j$（$t=0,1,\dots,p-1$）都是模 p^{j+1} 的根。
>
> (2) 若 $f(a)\not\equiv 0 \pmod{p^{j+1}}$，则 a 的提升没有一个是模 p^{j+1} 的根。

证明　由引理 7.2，$f(a+tp^j)\equiv f(a)+tp^jf'(a) \pmod{p^{j+1}}$。因为 $p\mid f'(a)$，有 $p^{j+1}\mid tp^jf'(a)$，所以 $f(a+tp^j)\equiv f(a) \pmod{p^{j+1}}$，与 t 无关。 ■

于是模 p^j 的一个根在模 p^{j+1} 下的命运只有三种：非奇异时恰有 1 个提升；奇异时有 p 个或 0 个提升。

■ **例 7.3　习题集 4 第 1 部分第 1(c) 题**　解 $x^5+x^4+1\equiv 0 \pmod{81}$。

解　令 $f(x)=x^5+x^4+1$。模 3：$f(0)=1$，$f(1)=3\equiv 0$，$f(2)=49\equiv 1$，只有根 $x\equiv 1$。$f'(x)=5x^4+4x^3$，$f'(1)=9\equiv 0 \pmod 3$，所以 1 是奇异根。$f(1)=3$ 不被 9 整除，由命题 7.4(2)，1 的三个提升 $1,4,7$ 都不是模 9 的根。直接验证：$f(1)=3$，$f(4)=1281$，$f(7)=19209$，模 9 都余 3。

模 9 无解，模 81 就更不可能有解：模 81 的根自动是模 9 的根。所以原同余式无解。 ♦

一个奇异根可以连续几层都「全部提升」，然后在某一层突然全部消失。下面的例子把这个过程完整画成一棵树。

■ **例 7.4　根的提升树**　对 $k=1,2,3,4$，求 $x^2+x+7\equiv 0 \pmod{3^k}$ 的全部解。

解　令 $f(x)=x^2+x+7$，$f'(x)=2x+1$。模 3：$f(0)=7$，$f(1)=9$，$f(2)=13$，只有根 1。凡是模 3 余 1 的 x，都有 $f'(x)=2x+1\equiv 0 \pmod 3$，所以这棵树上每个根都是奇异根，每一层都用命题 7.4 判断。

- 模 9：$f(1)=9\equiv 0 \pmod 9$，全部提升，得到 $1,4,7$。
- 模 27：$f(1)=9$，$f(4)=27$，$f(7)=63$。只有 $27\mid f(4)$，所以 1 和 7 消失，4 全部提升为 $4,13,22$。
- 模 81：$f(4)=27$，$f(13)=189$，$f(22)=513$，模 81 都余 27，三个根全部消失。

所以解的个数依次是 $1,3,3,0$：模 3 为 $\{1\}$，模 9 为 $\{1,4,7\}$，模 27 为 $\{4,13,22\}$，模 81 无解，从而模 3^k（$k\geq 4$）都无解。整个过程见图 7.2。 ♦

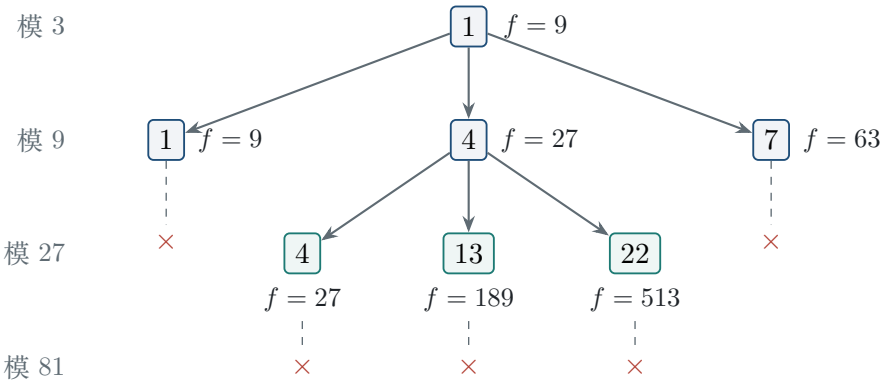


图 7.2　$x^2 + x + 7$ 模 3^k 的根。每个根都是奇异根：若 $3^{k+1} \mid f$ 就分出 3 个提升，否则这一支断掉（红叉）。到模 81 时全部断掉。

这棵树为什么在模 81 处断掉？配方可以看得更清楚：$4f(x) = (2x+1)^2 + 27$。若模 81 有解，则 $(2x+1)^2 \equiv -27 \pmod{81}$。设 $3^v \parallel 2x+1$（即 $3^v \mid 2x+1$ 但 $3^{v+1} \nmid 2x+1$）。若 $v \geq 2$，左边被 81 整除，而 -27 不被 81 整除；若 $v \leq 1$，左边恰好被 3^{2v} 整除，$2v \in \{0, 2\}$，而 -27 恰好被 3^3 整除，两边模 81 不可能相等。归根结底，**平方数含 3 的偶数次幂，而 27 含 3 的奇数次幂**。

易错提醒

(1) 判断奇异与否看的是 $f'(a)$ 模 p，不是模 p^j。例 7.2 中 $f'(5) = 75$ 模 8 不为 0，模 2 也不为 0，是非奇异根；例 7.4 中 $f'(4) = 9$ 模 27 不为 0，但模 3 为 0，是奇异根。

(2) 奇异根不一定提升失败，也不一定提升成功，必须检查 $f(a)$ 是否被 p^{j+1} 整除。

(3) 定理 7.1 的「唯一」指一个非奇异根只有一个提升。同余式模 p^{j+1} 的解数，等于模 p^j 的每个根的提升个数之和，其中奇异根贡献 0 个或 p 个。

注 7.1　$p = 2$ 时奇异根很常见。$x^2 - 1$ 的导数 $2x$ 永远是偶数，所以每个根都是奇异根。逐层计算：模 $2, 4, 8, 16, 32$ 的解分别是 $\{1\}$，$\{1, 3\}$，$\{1, 3, 5, 7\}$，$\{1, 7, 9, 15\}$，$\{1, 15, 17, 31\}$。从模 8 到模 16，1 和 7 各分出两个根（$1^2 - 1 = 0$，$7^2 - 1 = 48$ 被 16 整除），3 和 5 消失（$3^2 - 1 = 8$，$5^2 - 1 = 24$，模 16 都余 8）。所以从 $k = 3$ 起，$x^2 \equiv 1 \pmod{2^k}$ 恰有 4 个解（练习 7.3）。一个二次同余式有 4 个解，下一节会解释为什么这种事在素数模下不会发生。

7.3　素数模的同余式与 Lagrange 定理

有了中国剩余定理和 Hensel 引理，解 $f(x) \equiv 0 \pmod{m}$ 只剩一个核心问题：模素数 p 怎么解？这一节先把同余式化简，再证明模 p 的根数有一个和实数情形一样的上界。

7.3.1　三种化简

设 $f(x) = a_n x^n + \cdots + a_1 x + a_0$，下面的操作都不改变 $f(x) \equiv 0 \pmod{p}$ 的解集。

(1) **系数取模**：把每个 a_i 换成它模 p 的余数。系数被 p 整除的项可以直接删掉。

(2) **化为首一**：若 $p \nmid a_n$，把 f 乘以 a_n 模 p 的逆元，首项系数就变成 1。乘一个模 p 可逆的数不改变「$\equiv 0$」。

(3) 降次：由 Fermat 小定理，对每个整数 x 都有 $x^p \equiv x \pmod p$（$p \mid x$ 时两边都是 0）。所以当 $k \geq p$ 时，可以把 $x^k = x^{k-p} \cdot x^p$ 换成 $x^{k-p} \cdot x = x^{k-(p-1)}$。反复使用，次数就降到 $p-1$ 以下。等价的说法是：用 $x^p - x$ 去除 f，余式 r 与 f 在每个整数处的值模 p 相同。

若系数模 p 化简后不是零多项式，可用 $x^p \equiv x$ 把次数降到 $p-1$ 以下，再乘以首项系数的逆元化为首一多项式。若化简后是零多项式，则所有 p 个剩余类都是根，不能把它化成首一且次数小于 p 的多项式。这两种情形覆盖所有同余式，原则上都可直接检查 $0, 1, \ldots, p-1$。

■ **例 7.5　化简后求解**　解 $2x^{13} + 5x^9 + 6x^2 + 1 \equiv 0 \pmod 7$。

解　$p - 1 = 6$。降次：$x^{13} \to x^7 \to x$，$x^9 \to x^3$。同余式化为

$$2x + 5x^3 + 6x^2 + 1 \equiv 0, \quad \text{即} \quad 5x^3 + 6x^2 + 2x + 1 \equiv 0 \pmod 7.$$

首项系数 5 的逆元是 3（$5 \cdot 3 = 15 \equiv 1$），两边乘 3：

$$x^3 + 18x^2 + 6x + 3 \equiv x^3 + 4x^2 + 6x + 3 \equiv 0 \pmod 7.$$

代入 $x = 0, 1, \ldots, 6$，左边模 7 依次为 $3, 0, 4, 0, 1, 6, 0$，解是 $x \equiv 1, 3, 6 \pmod 7$。实际上

$$(x-1)(x-3)(x-6) = x^3 - 10x^2 + 27x - 18 \equiv x^3 + 4x^2 + 6x + 3 \pmod 7,$$

三次式恰好分解成三个一次因子。代回原式也可以核对，例如 $x = 1$ 时 $2+5+6+1 = 14 \equiv 0$。 ♦

7.3.2 Lagrange 定理

实系数 n 次多项式至多有 n 个实根。证明是：每找到一个根 α，就能提出因子 $x - \alpha$，次数降低 1。模素数时这个论证照样成立。

> **定理 7.5（Lagrange）**　设 $f(x) = a_n x^n + \cdots + a_0$ 是整系数多项式，p 是素数，$p \nmid a_n$。则 $f(x) \equiv 0 \pmod p$ 至多有 n 个模 p 互不同余的解。

思路　对次数 n 归纳。若 f 有一个根 α，就把 f 写成 $(x-\alpha)g(x)$ 加上一个被 p 整除的常数。f 的其他根 β 满足 $(\beta - \alpha)g(\beta) \equiv 0$。因为 p 是素数，由 Euclid 引理，β 必须是 g 的根，而 g 的次数是 $n-1$。

证明　$n = 0$ 时 $f = a_0$，$p \nmid a_0$，没有根，结论成立。设 $n \geq 1$，且结论对 $n-1$ 次多项式成立。若 f 模 p 无根，已经成立。否则取一个根 α。做多项式带余除法（除式 $x - \alpha$ 首一，商和余数都是整系数）：

$$f(x) = (x-\alpha)g(x) + r, \qquad g(x) = a_n x^{n-1} + \cdots, \quad r \in \mathbb{Z}.$$

代入 $x = \alpha$ 得 $r = f(\alpha) \equiv 0 \pmod p$。若 β 是 f 的另一个根，$\beta \not\equiv \alpha$，则

$$0 \equiv f(\beta) \equiv (\beta - \alpha)\, g(\beta) \pmod p.$$

p 是素数且 $p \nmid \beta - \alpha$，由 Euclid 引理 $p \mid g(\beta)$。所以 f 除 α 以外的根都是 g 的根。g 的次数为 $n-1$，首项系数 a_n 不被 p 整除，由归纳假设它至多有 $n-1$ 个根。于是 f 至多有 $1 + (n-1) = n$ 个根。 ■

条件 $p \nmid a_n$ 不能去掉。例如 $f(x) = 7x^2 + x$ 模 7 实际上就是 x，只有根 0，这无妨；但 $f(x) = 7x^2 - 7x$ 的每个系数都被 7 整除，模 7 的每个数都是根。下面的推论说明，「根太多」只可能是这种情形。

推论 7.6　设 $f(x) = a_n x^n + \cdots + a_0$ 是整系数多项式。若 $f(x) \equiv 0 \pmod p$ 有多于 n 个模 p 互不同余的解，则 f 的每个系数都被 p 整除。

证明　反设有系数不被 p 整除，取其中下标最大的 a_k，$k \le n$。删去 $a_n x^n, \ldots, a_{k+1}x^{k+1}$ 这些系数被 p 整除的项，不改变解集，剩下一个首项系数 $a_k \not\equiv 0$ 的 k 次多项式。由定理 7.5，它至多有 $k \le n$ 个根，矛盾。 ■

推论 7.6 常用来证明多项式同余式：要说明两个次数 $\le n$ 的多项式 F, G 的系数逐个模 p 同余，只要找到 $n+1$ 个模 p 互不同余的 x 使 $F(x) \equiv G(x)$。下一节的证明全靠这个技巧。

模数不是素数时，定理失效。注 7.1 中，二次同余式 $x^2 \equiv 1 \pmod 8$ 有 4 个解 $1, 3, 5, 7$。模 15 也一样：$x^2 \equiv 1 \pmod{15}$ 的解是 $1, 4, 11, 14$。

想一想　Lagrange 定理的证明用在 $x^2 - 1 \pmod 8$ 上，究竟哪一步出了问题？

答　取 $\alpha = 1$，$x^2 - 1 = (x-1)(x+1)$，$g(x) = x + 1$。对 $\beta = 3$，$(\beta - \alpha)g(\beta) = 2 \cdot 4 = 8 \equiv 0 \pmod 8$，但 $2 \not\equiv 0$，$4 \not\equiv 0$。「乘积为 0 则某个因子为 0」在模 8 下不成立，Euclid 引理只对素数成立。这个现象也有用处：若找到 $b \not\equiv \pm 1$ 而 $b^2 \equiv 1 \pmod n$，则 $n \mid (b-1)(b+1)$ 但 n 不整除任何一个因子，于是 $(b-1, n)$ 是 n 的非平凡因子。例如模 15 取 $b = 4$，$(3, 15) = 3$。

7.4　恰有 n 个根的判别法

Lagrange 定理只给出上界。有一个多项式取到了上界：由 Fermat 小定理，$x^p - x$ 在 $0, 1, \ldots, p-1$ 处都 $\equiv 0$，这个 p 次多项式恰有 p 个根。本节证明，任何取到上界的首一多项式都是 $x^p - x$ 的因式。先说清楚「模 p 整除」的意思。

定义 7.7　设 f, g 是整系数多项式。若存在整系数多项式 h，使 $g - fh$ 的每个系数都被 p 整除，就说 f **模 p 整除**（divides mod p）g，记作 $f \mid g \pmod p$，并写 $g \equiv fh \pmod p$。

这是系数的同余，比「在每个整数处取值同余」更强。例如 x^p 与 x 在每个整数处取值模 p 同余，但作为多项式并不模 p 同余。

定理 7.8　设 $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ 是首一整系数多项式，p 是素数。则 $f(x) \equiv 0 \pmod p$ 恰有 n 个模 p 互不同余的解，当且仅当 $f(x) \mid x^p - x \pmod p$。

思路　两个方向都是数根。「⇒」：用 f 去除 $x^p - x$，余式次数小于 n 却有 n 个根，只能模 p 为零。「⇐」：$x^p - x \equiv fg$ 的 p 个根分给 f 和 g，g 至多分到 $p - n$ 个，f 至少分到 n 个。

证明　「⇒」：模 p 只有 p 个剩余类，所以 $n \le p$。f 首一，带余除法给出整系数多项式 g, r 使

$$x^p - x = f(x)g(x) + r(x), \qquad \deg r < n.$$

设 α 是 f 的根。代入得 $\alpha^p - \alpha = f(\alpha)g(\alpha) + r(\alpha)$。左边由 Fermat 小定理 $\equiv 0$，$f(\alpha) \equiv 0$，所以 $r(\alpha) \equiv 0$。于是 r 有 n 个互不同余的根，而 $\deg r < n$。由推论 7.6，r 的系数全被 p 整除，即 $x^p - x \equiv f(x)g(x) \pmod p$。

「⇐」：设 $x^p - x \equiv f(x)g(x) \pmod p$。比较首项可知，$g$ 模 p 化简后是首一的 $p-n$ 次多项式，由定理 7.5，它至多有 $p-n$ 个根。所以至少有 $p - (p-n) = n$ 个 $\alpha \in \{0, 1, \dots, p-1\}$ 使 $g(\alpha) \not\equiv 0$。对这些 α，

$$f(\alpha)g(\alpha) \equiv \alpha^p - \alpha \equiv 0 \pmod p,$$

由 Euclid 引理 $f(\alpha) \equiv 0$。f 至少有 n 个根，由 Lagrange 定理又至多有 n 个，所以恰有 n 个。 ■

■ **例 7.6**　**两个模 7 的检验**　(1) 例 7.5 中的 $f(x) = x^3 + 4x^2 + 6x + 3$ 有 3 个根。验证 $f \mid x^7 - x \pmod 7$。(2) 用定理 7.8 判断 $x^2 + 1 \equiv 0 \pmod 7$ 有没有解。

解　(1) $x^7 - x$ 的根是 $0, 1, \dots, 6$，f 的根是 $1, 3, 6$，剩下的是 $0, 2, 4, 5$。所以商式应当是

$$g(x) = x(x-2)(x-4)(x-5) = x^4 - 11x^3 + 38x^2 - 40x \equiv x^4 + 3x^3 + 3x^2 + 2x \pmod 7.$$

直接相乘：$f(x)g(x) = x^7 + 7x^6 + 21x^5 + 35x^4 + 35x^3 + 21x^2 + 6x$，除首项外各系数依次是 $7, 21, 35, 35, 21, 6$，模 7 为 $0, 0, 0, 0, 0, 6$。所以 $fg \equiv x^7 + 6x \equiv x^7 - x \pmod 7$。

(2) 用 $x^2 + 1$ 去除 $x^7 - x$。因为 $x^2 \equiv -1$ 模 $x^2 + 1$ 成立，可以把每个 x^2 换成 -1：$x^7 = x \cdot (x^2)^3 \to -x$，余式是 $-x - x = -2x$。实际上

$$x^7 - x = (x^2 + 1)(x^5 - x^3 + x) - 2x.$$

余式 $-2x$ 的系数不被 7 整除，所以 $x^2 + 1 \nmid x^7 - x \pmod 7$，解的个数小于 2。而解总是成对出现（$x$ 是解则 $-x$ 也是，$x \equiv -x$ 只在 $x \equiv 0$ 时发生，0 不是解），所以**没有解**。直接验证：$x^2 \bmod 7$ 只取 $0, 1, 4, 2$，取不到 $6 \equiv -1$。这与第 4 讲的结论一致：$x^2 \equiv -1 \pmod p$ 有解当且仅当 $p = 2$ 或 $p \equiv 1 \pmod 4$。 ♦

定理 7.8 最常用的情形是 $f = x^d - 1$。

> **推论 7.9**　若 $d \mid p - 1$，则 $x^d \equiv 1 \pmod p$ 恰有 d 个模 p 互不同余的解。

证明　设 $p - 1 = dk$。对 $y = x^d$ 用因式分解 $y^k - 1 = (y-1)(y^{k-1} + \cdots + y + 1)$，得到整系数多项式的恒等式

$$x^p - x = x\big(x^{p-1} - 1\big) = (x^d - 1) \cdot x\big(x^{d(k-1)} + \cdots + x^d + 1\big),$$

所以 $x^d - 1 \mid x^p - x$，由定理 7.8 即得。 ■

■ **例 7.7**　**模 11 的五次单位根**　求 $x^5 \equiv 1 \pmod{11}$ 的全部解，并与 $x^3 \equiv 1 \pmod{11}$ 比较。

解　$5 \mid 10$，由推论 7.9，恰有 5 个解。不必逐个计算五次幂：对任何 $a \not\equiv 0$，$(a^2)^5 = a^{10} \equiv 1$，所以**每个非零平方数都是解**。$1^2, 2^2, 3^2, 4^2, 5^2$ 模 11 是 $1, 4, 9, 5, 3$，已经有 5 个，所以解恰好是 $x \equiv 1, 3, 4, 5, 9 \pmod{11}$。

$x^3 \equiv 1$ 的情形不同，$3 \nmid 10$。设 $a^3 \equiv 1$。又 $a^{10} \equiv 1$，而 $21 = 3 \cdot 7 = 10 \cdot 2 + 1$，所以 $a \equiv a \cdot (a^{10})^2 = a^{21} = (a^3)^7 \equiv 1$。只有一个解 $x \equiv 1$。同样的论证说明，$x^d \equiv 1$ 的解和 $x^{(d, p-1)} \equiv 1$ 的解完全相同，由推论 7.9，**解数是** $(d, p-1)$。 ♦

7.4.1　Wilson 定理与对称多项式

第 4 讲用「配对」证明了 Wilson 定理。定理 7.8 给出另一个证明，而且得到的信息多得多。

> **推论 7.10（Wilson 定理）**　对素数 p，有多项式同余式
>
> $$\boxed{x(x-1)(x-2)\cdots\big(x-(p-1)\big) \equiv x^p - x \pmod{p},}$$
>
> 特别地 $(p-1)! \equiv -1 \pmod{p}$。

证明　左边 $F(x)$ 是首一 p 次多项式，恰有 p 个根 $0, 1, \ldots, p-1$。由定理 7.8，$x^p - x \equiv F(x)h(x)$。两边都是首一 p 次，所以 $h \equiv 1$，$F \equiv x^p - x$。比较 x 的系数：左边是 $(-1)(-2)\cdots(-(p-1)) = (-1)^{p-1}(p-1)!$，右边是 -1。p 为奇数时 $(-1)^{p-1} = 1$，得 $(p-1)! \equiv -1$；$p = 2$ 时 $1! = 1 \equiv -1 \pmod 2$ 直接成立。 ■

两边约去 x，对奇素数 p 得 $(x-1)(x-2)\cdots(x-(p-1)) \equiv x^{p-1} - 1$。左边展开后，$x^{p-1-k}$ 的系数是 $(-1)^k \sigma_k$，其中

$$\sigma_k = \sum_{1 \le i_1 < i_2 < \cdots < i_k \le p-1} i_1 i_2 \cdots i_k$$

是 $1, 2, \ldots, p-1$ 的第 k 个**初等对称多项式**（elementary symmetric polynomial），即从中任取 k 个数相乘、再对所有取法求和。例如 $\sigma_1 = 1 + 2 + \cdots + (p-1)$，$\sigma_{p-1} = (p-1)!$。右边中间的系数全是 0，于是得到：

> **命题 7.11（习题集 4 第 1 部分第 5(a) 题）**　设 p 是奇素数，则 $\sigma_1 \equiv \sigma_2 \equiv \cdots \equiv \sigma_{p-2} \equiv 0 \pmod{p}$，$\sigma_{p-1} \equiv -1 \pmod{p}$。

证明　由前面的多项式恒等式，将 $\prod_{j=1}^{p-1}(x-j)$ 与 $x^{p-1} - 1$ 的各次项系数模 p 比较。中间项系数为 $(-1)^k \sigma_k$，故 $1 \le k \le p-2$ 时 $\sigma_k \equiv 0$；常数项因 $p-1$ 为偶数等于 σ_{p-1}，故它同余于 -1。 ■

$k = 1$ 的情形可以直接验证：$\sigma_1 = \frac{p(p-1)}{2}$，$p$ 为奇数时 $\frac{p-1}{2}$ 是整数，所以 $p \mid \sigma_1$。$k = 2$ 时，$\sigma_2 = \sum_{i<j} ij$ 被 p 整除就不那么明显了。当 $p = 3$ 时 $k = 2 = p - 1$，所以命题对 σ_2 的断言从 $p \ge 5$ 开始；例如 $p = 5$ 时 $\sigma_2 = 1\cdot 2 + 1\cdot 3 + 1\cdot 4 + 2\cdot 3 + 2\cdot 4 + 3\cdot 4 = 35$。

■ **例 7.8**　$p = 7$ **的全部系数**　展开 $(x-1)(x-2)\cdots(x-6)$，检验命题 7.11。

解　把首尾配对可以简化计算：$(x-1)(x-6)=y+6$，$(x-2)(x-5)=y+10$，$(x-3)(x-4)=y+12$，其中 $y=x^2-7x$。逐步相乘得到

$$(x-1)(x-2)\cdots(x-6)=x^6-21x^5+175x^4-735x^3+1624x^2-1764x+720.$$

于是 $\sigma_1,\dots,\sigma_6$ 依次是 $21,175,735,1624,1764,720$。前五个都是 7 的倍数：$21=7\cdot3$，$175=7\cdot25$，$735=7\cdot105$，$1624=7\cdot232$，$1764=7\cdot252$。最后 $720=7\cdot103-1\equiv-1$，这就是 Wilson 定理。 ♦

例 7.8 中 $\sigma_5=1764=36\cdot49$ 不仅被 7 整除，还被 7^2 整除。这不是巧合，练习 7.6 证明：$p\geq5$ 时 $\sigma_{p-2}\equiv0\pmod{p^2}$。

历史与背景　Waring 在 1770 年的《代数沉思录》中发表了这个定理，归功于他的学生 John Wilson，但没有证明；第一个证明由 Lagrange 在 1771 年给出。$\sigma_{p-2}\equiv0\pmod{p^2}$ 等价于调和和 $1+\frac12+\cdots+\frac{1}{p-1}$ 的分子被 p^2 整除（因为这个和等于 $\sigma_{p-2}/(p-1)!$），例如 $1+\frac12+\cdots+\frac16=\frac{49}{20}$。Wolstenholme 在 1862 年证明了这一结论。

7.5　阶

Euler 定理说：$(n,35)=1$ 时 $n^{24}\equiv1\pmod{35}$，因为 $\varphi(35)=24$。24 能不能换成更小的数？也就是说，使 $n^N\equiv1\pmod{35}$ 对所有 $(n,35)=1$ 都成立的最小正整数 N 是多少？要回答它，需要对单个的 n 问同样的问题。

定义 7.12　设 $(a,m)=1$。使 $a^h\equiv1\pmod m$ 成立的最小正整数 h 叫做 a 模 m 的阶（order），记作 $\operatorname{ord}_m(a)$，在不致混淆时简写为 $\operatorname{ord}a$。

由 Euler 定理，这样的 h 存在且 $h\leq\varphi(m)$。条件 $(a,m)=1$ 是必要的：若 $a^h\equiv1$，则 $a\cdot a^{h-1}\equiv1$，a 模 m 可逆。第 4 讲模 7 的幂次表中各行第一次出现 1 的位置就是阶：$1,2,\dots,6$ 模 7 的阶依次是 $1,3,6,3,6,2$，都是 6 的因数。下面的引理说明这不是巧合。

引理 7.13　设 $h=\operatorname{ord}_m(a)$，$k\geq0$ 是整数。则 $a^k\equiv1\pmod m$ 当且仅当 $h\mid k$。

证明　若 $k=hq$，则 $a^k=(a^h)^q\equiv1$。反之，设 $a^k\equiv1$，作带余除法 $k=hq+r$，$0\leq r<h$。则 $1\equiv a^k=(a^h)^qa^r\equiv a^r$。若 $r>0$，这与 h 的最小性矛盾，所以 $r=0$，$h\mid k$。 ■

于是 a^k 模 m 以 h 为周期循环：$a^i\equiv a^j$ 当且仅当 $i\equiv j\pmod h$（两边乘 a^j 的逆元即可化为引理 7.13）。取 $k=\varphi(m)$ 还得到：$\operatorname{ord}_m(a)\mid\varphi(m)$，特别地 $\operatorname{ord}_p(a)\mid p-1$。所以求阶时只需检查 $\varphi(m)$ 的因数。

■ **例 7.9　回答模 35 的问题**　求 $\operatorname{ord}_{35}(2)$，并求使 $n^N\equiv1\pmod{35}$ 对所有 $(n,35)=1$ 都成立的最小正整数 N。

解　$35 = 5 \cdot 7$，所以 $2^N \equiv 1 \pmod{35}$ 当且仅当 $2^N \equiv 1 \pmod 5$ 且 $2^N \equiv 1 \pmod 7$。模 5：$2, 4, 3, 1$，$\operatorname{ord}_5(2) = 4$。模 7：$2, 4, 1$，$\operatorname{ord}_7(2) = 3$。由引理 7.13，条件变成 $4 \mid N$ 且 $3 \mid N$，即 $12 \mid N$。所以 $\operatorname{ord}_{35}(2) = 12$。验证：$2^{12} = 4096 = 117 \cdot 35 + 1$。

再看一般的 n。$(n, 35) = 1$ 时，由 Fermat 小定理 $n^4 \equiv 1 \pmod 5$，$n^6 \equiv 1 \pmod 7$。12 是 4 和 6 的倍数，所以 $n^{12} \equiv 1$ 模 5 和模 7 都成立，从而模 35 成立。另一方面 N 必须对 $n = 2$ 成立，所以 $12 \mid N$。答案是 $N = 12 = \operatorname{lcm}(4, 6)$，只有 Euler 给出的 24 的一半。

由此可见，模 35 的每个元素的阶都整除 12，没有一个元素的阶等于 $\varphi(35) = 24$。 ♦

已知 a 的阶，a 的各次幂的阶由下面的公式给出。

引理 7.14　设 $h = \operatorname{ord}_m(a)$，$k \geq 1$。则

$$\boxed{\operatorname{ord}_m(a^k) = \frac{h}{(h, k)}.}$$

证明　由引理 7.13，$(a^k)^j \equiv 1$ 当且仅当 $h \mid kj$。记 $d = (h, k)$，$h = dh'$，$k = dk'$，则 $(h', k') = 1$。$h \mid kj$ 等价于 $h' \mid k'j$，又等价于 $h' \mid j$。所以使 $(a^k)^j \equiv 1$ 的最小正整数 j 是 $h' = h/d$。 ■

■ **例 7.10　模 11 的各元素的阶**　先求 $\operatorname{ord}_{11}(2)$，再用引理 7.14 求 $1, 2, \ldots, 10$ 模 11 的阶。

解　阶整除 10，只需检查 $1, 2, 5$：$2^1 = 2$，$2^2 = 4$，$2^5 = 32 \equiv 10$，都不是 1，所以 $\operatorname{ord}_{11}(2) = 10$。于是 $2^0, 2^1, \ldots, 2^9$ 两两不同余，依次是

$$1,\ 2,\ 4,\ 8,\ 5,\ 10,\ 9,\ 7,\ 3,\ 6,$$

恰好是全部非零剩余。每个 a 都可写成 2^s，$\operatorname{ord}(2^s) = 10/(10, s)$：

- $s = 1, 3, 7, 9$：阶 10，即 $a = 2, 8, 7, 6$；
- $s = 2, 4, 6, 8$：阶 5，即 $a = 4, 5, 9, 3$；
- $s = 5$：阶 2，即 $a = 10$；$s = 0$：阶 1，即 $a = 1$。

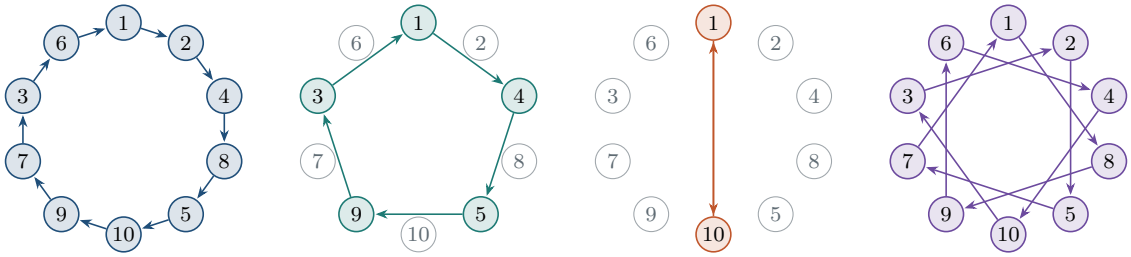
阶为 1 或 5 的元素合起来是 $\{1, 3, 4, 5, 9\}$，与例 7.7 中 $x^5 \equiv 1$ 的解一致。 ♦

把 $2^0, \ldots, 2^9$ 按顺序排在圆周上，乘以 2^s 就是沿圆周一次走 s 格。从 1 出发，要走回 1，需要走的次数正是 $10/(10, s)$，见图 7.3。

要构造阶很大的元素，最自然的办法是把两个元素乘起来。阶互素时这样做是有效的。

引理 7.15　设 $\operatorname{ord}_m(a) = h$，$\operatorname{ord}_m(b) = k$，且 $(h, k) = 1$。则 $\operatorname{ord}_m(ab) = hk$。

证明　$(ab)^{hk} = (a^h)^k (b^k)^h \equiv 1$，所以 $r = \operatorname{ord}_m(ab)$ 整除 hk。另一方面，$1 \equiv (ab)^{rh} = (a^h)^r\, b^{rh} \equiv b^{rh}$，由引理 7.13，$k \mid rh$；又 $(k, h) = 1$，所以 $k \mid r$。对称地 $h \mid r$。h, k 互素，所以 $hk \mid r$。于是 $r = hk$。 ■



乘 2：每次走 1 格　　乘 4：每次走 2 格　　乘 10：每次走 5 格　　乘 8：每次走 3 格

$\mathrm{ord}_{11}(2)=10$　　$\mathrm{ord}_{11}(4)=5$　　$\mathrm{ord}_{11}(10)=2$　　$\mathrm{ord}_{11}(8)=10$

图 7.3　模 11 的非零剩余按 $2^0,2^1,\dots,2^9$ 排成一圈（1 在正上方，顺时针）。从 1 出发反复乘 $a=2^s$，经过的点着色，未经过的点为灰色。走回 1 所需的步数就是阶 $10/(10,s)$。

易错提醒

(1) 引理 7.15 的条件 $(h,k)=1$ 不能去掉。模 7 时 3 和 5 的阶都是 6，但 $3\cdot5=15\equiv1$，阶为 1；3 和 3 的阶都是 6，$3\cdot3\equiv2$，阶为 3。阶不互素时，乘积的阶既不是 hk，也不一定是 $\mathrm{lcm}(h,k)$。

(2) 引理 7.14 的公式是 $h/(h,k)$，不是 (h,k)。检验一下 $k=h$：$a^h\equiv1$，阶为 $1=h/h$。

(3) 阶依赖于模数。$\mathrm{ord}_5(2)=4$，$\mathrm{ord}_7(2)=3$，$\mathrm{ord}_{35}(2)=12$。写 $\mathrm{ord}\,a$ 时要清楚模数是什么。

想一想　已知 $a^{12}\equiv1$ 且 $a^{18}\equiv1\pmod m$。关于 $\mathrm{ord}_m(a)$ 能得出什么？这些可能性都会出现吗？

答　由引理 7.13，阶同时整除 12 和 18，所以整除 $(12,18)=6$，只能是 $1,2,3,6$。都会出现：模 7 时 $1,6,2,3$ 的阶分别是 $1,2,3,6$。一般地 $a^i\equiv a^j\equiv1$ 推出 $a^{(i,j)}\equiv1$，例 7.7 中的 $x^3\equiv1\Rightarrow x\equiv1$ 就是 $(3,10)=1$ 的情形。

■ **例 7.11　Fermat 数的素因子，习题集 4 第 2 部分第 3 题**　(1) 证明：若 $a>1$ 且 a^k+1 是素数，则 k 是 2 的方幂。(2) 证明：若奇素数 $p\mid a^{2^n}+1$，则 $p\equiv1\pmod{2^{n+1}}$。(3) 求 $2^{32}+1$ 的一个素因子。

解　(1) 若 k 有奇因子 $r>1$，记 $b=a^{k/r}$，则 $a^k+1=b^r+1=(b+1)(b^{r-1}-b^{r-2}+\cdots+1)$，且 $1<b+1<a^k+1$，与素数矛盾。

(2) $a^{2^n}\equiv-1\pmod p$（于是 $p\nmid a$），平方得 $a^{2^{n+1}}\equiv1$，所以 $h=\mathrm{ord}_p(a)$ 整除 2^{n+1}，是 2 的方幂。若 $h\mid2^n$，则 $a^{2^n}\equiv1$，与 $a^{2^n}\equiv-1\not\equiv1$（$p$ 是奇数）矛盾。所以 $h=2^{n+1}$。又 $h\mid p-1$，得 $p\equiv1\pmod{2^{n+1}}$。

(3) 由 (2)，$2^{32}+1$ 的素因子 $\equiv1\pmod{64}$。依次检查这类素数 $193,257,449,577,641$：$2^{32}+1$ 除以它们的余数是 $109,2,325,288,0$。所以 $641\mid2^{32}+1$，事实上 $2^{32}+1=641\cdot6700417$。Fermat 猜想 $2^{2^n}+1$ 都是素数，Euler 在 1732 年用这个分解推翻了它。不用 (2) 就要从 3 开始逐个试除，(2) 把候选者减少到大约 $1/64$。　♦

7.6　原根

阶整除 $\varphi(m)$，所以 $\varphi(m)$ 是阶可能取到的最大值。取到最大值的元素最有用。

定义 7.16　设 $(g,m)=1$。若 $\operatorname{ord}_m(g)=\varphi(m)$，就称 g 是模 m 的**原根**（primitive root）。

若 g 是原根，则 $g^0,g^1,\dots,g^{\varphi(m)-1}$ 两两不同余（引理 7.13 后的说明），又都与 m 互素，共 $\varphi(m)$ 个，所以**恰好构成模 m 的一个简化剩余系**：每个与 m 互素的数都同余于 g 的某个幂。乘法于是变成指数的加法，这是第 8 讲「指标」的出发点。

- 模 7：3 和 5 的阶是 6，是原根；2, 4 的阶是 3，6 的阶是 2，都不是。
- 模 11：由例 7.10，原根是 2, 6, 7, 8。图 7.3 第一幅和第四幅中，从 1 出发走遍了全部 10 个点。
- 模 8：$1^2,3^2,5^2,7^2$ 都 $\equiv 1$，阶至多为 $2<\varphi(8)=4$，**没有原根**。
- 模 35：由例 7.9，每个阶都整除 $12<24$，也没有原根。

哪些模数有原根，留到第 8 讲回答。本讲证明素数模的情形，这是全部理论的基础。

定理 7.17　每个素数 p 都有原根。

思路　把 $p-1$ 分解为 $q_1^{e_1}q_2^{e_2}\cdots q_r^{e_r}$。只要对每个 i 找到阶恰为 $q_i^{e_i}$ 的元素 g_i，由引理 7.15，乘积 $g_1g_2\cdots g_r$ 的阶就是 $p-1$。要找到 g_i，只需数根：$x^{q^e}\equiv 1$ 的根比 $x^{q^{e-1}}\equiv 1$ 的根多。完整证明见下面的引理及其后的定理证明。

引理 7.18　设 p 是素数，q 是素数，$q^e\mid p-1$，$e\ge 1$。则存在模 p 阶恰为 q^e 的元素。

证明　由推论 7.9，$x^{q^e}\equiv 1$ 恰有 q^e 个根，$x^{q^{e-1}}\equiv 1$ 恰有 q^{e-1} 个根。设 α 是前者的根，由引理 7.13，$\operatorname{ord}\alpha\mid q^e$，所以 $\operatorname{ord}\alpha$ 是 q^e 或 q^{e-1} 的因数。后一种情形 α 是 $x^{q^{e-1}}\equiv 1$ 的根，这样的 α 只有 q^{e-1} 个。$q^e-q^{e-1}>0$，所以至少有一个根的阶恰为 q^e。 ■

定理 7.17 的证明　$p=2$ 时 1 是原根。设 p 为奇素数，$p-1=q_1^{e_1}\cdots q_r^{e_r}$。由引理 7.18，对每个 i 取 g_i 使 $\operatorname{ord}_p(g_i)=q_i^{e_i}$。这些阶两两互素，反复使用引理 7.15，$g=g_1g_2\cdots g_r$ 的阶是 $q_1^{e_1}\cdots q_r^{e_r}=p-1=\varphi(p)$。 ■

引理 7.18 只说明这样的元素存在，没有说怎么找。可以这样找：任取 $a\not\equiv 0$，令 $b=a^{(p-1)/q^e}$，则 $b^{q^e}=a^{p-1}\equiv 1$，$\operatorname{ord}b\mid q^e$。所以 $\operatorname{ord}b=q^e$ 当且仅当 $b^{q^{e-1}}=a^{(p-1)/q}\not\equiv 1$。失败的 a 恰好是 $x^{(p-1)/q}\equiv 1$ 的根，由推论 7.9，共 $(p-1)/q$ 个，所以**随便取一个 a，成功的机会是 $1-1/q$**。

■ **例 7.12　按证明构造模 37 的原根**　按定理 7.17 的证明，构造模 37 的一个原根。

解　$p-1=36=2^2\cdot 3^2$，需要一个阶为 9 的元素和一个阶为 4 的元素。

阶为 9（$q=3, e=2$）：取 $a=3, b=3^{36/9}=3^4=81\equiv 7$。检验 $b^3=343=9\cdot 37+10\equiv 10\not\equiv 1$，所以 $\operatorname{ord}_{37}(7)=9$。

阶为 4（$q=2, e=2$）：先试 $a=3$，$b=3^{36/4}=3^9$。由 $3^4\equiv 7$，$3^9=(3^4)^2\cdot 3\equiv 49\cdot 3\equiv 12\cdot 3=36\equiv -1$，于是 $b^2\equiv 1$，b 的阶只有 2，**失败**。换 $a=2$：$2^9=512=13\cdot 37+31$，$b=31\equiv -6$，$b^2\equiv 36\equiv -1\not\equiv 1$，所以 $\operatorname{ord}_{37}(31)=4$。

相乘：$(4,9)=1$，由引理 7.15，$g=31\cdot 7=217=5\cdot 37+32$，即 $g\equiv 32\equiv -5$ 的阶是 36。直接核对：36 的真因数都整除 12 或 18，所以只需检查 $g^{12}, g^{18}\not\equiv 1$。由 $(-5)^2=25\equiv -12$，$(-5)^3\equiv 60\equiv 23$，$(-5)^6\equiv 529\equiv 11$，得 $(-5)^{12}\equiv 121\equiv 10$，$(-5)^{18}\equiv 10\cdot 11=110\equiv -1$，都不是 1。

$a=3$ 失败是因为 $3^{18}=(3^9)^2\equiv 1$，这样的 a 占一半。同样的检查给出 $2^{12}\equiv 26$，$2^{18}\equiv -1$，所以 2 本身就是原根。实际中通常从小到大直接试，第 8 讲把这个检查整理成判别法。 ♦

历史与背景 Lambert 在 1769 年断言每个素数都有原根，没有证明；Euler 在 1773 年命名了原根，给出的证明有漏洞；第一个完整的证明见 Gauss 1801 年的《算术研究》。

第 8 讲将证明：模 p 恰有 $\varphi(p-1)$ 个原根；有原根的模数恰是 $2, 4, p^e, 2p^e$（p 为奇素数）。

本讲小结

- Hensel 迭代：$a_{j+1}\equiv a_j-f(a_j)u \pmod{p^{j+1}}$，$uf'(a_1)\equiv 1 \pmod p$。奇异根（$p\mid f'(a)$）：$p^{j+1}\mid f(a)$ 时 p 个提升全是根，否则全不是。
- 模 p 的同余式可化为首一、次数 $\le p-1$。Lagrange：$p\nmid a_n$ 时至多 n 个根。
- 首一 f 恰有 $\deg f$ 个根 $\iff f\mid x^p-x \pmod p$。由此得 $x^d\equiv 1$（$d\mid p-1$）恰有 d 个根、Wilson 定理和 $p\mid \sigma_k$（$k\le p-2$）。
- 阶：$a^k\equiv 1 \iff \operatorname{ord} a\mid k$；$\operatorname{ord}(a^k)=\operatorname{ord} a/(\operatorname{ord} a, k)$；阶互素时 $\operatorname{ord}(ab)=\operatorname{ord} a\cdot\operatorname{ord} b$。
- 原根的幂走遍简化剩余系。素数 p 有原根：对每个 $q^e\parallel p-1$ 取阶为 q^e 的元素，再相乘。

练习

7.1 ★ 用算法 7.3 求 $x^2\equiv 2 \pmod{343}$ 的全部解。

7.2 ★★ 证明：$k\ge 2$ 时 $x^3\equiv 8 \pmod{3^k}$ 恰有 3 个解，它们是 $x\equiv 2 \pmod{3^{k-1}}$。（提示：令 $x=2y$，化为 $y^3\equiv 1$。）

7.3 ★★ 证明：$k\ge 3$ 时 $x^2\equiv 1 \pmod{2^k}$ 恰有 4 个解 ± 1，$2^{k-1}\pm 1$。这和 Lagrange 定理矛盾吗？

7.4 ★★ （习题集 4 第 1 部分第 4 题）设 $f(a)\equiv 0 \pmod{p^j}$，$f'(a)\not\equiv 0 \pmod p$，$v$ 满足 $vf'(a)\equiv 1 \pmod{p^{2j}}$，令 $b=a-f(a)v$。证明 $f(b)\equiv 0 \pmod{p^{2j}}$。用它从 x^2+1 模 25 的根 7 一步求出模 625 的根。

7.5 ★ 用定理 7.8 说明 $x^4+1\equiv 0 \pmod{17}$ 恰有 4 个解，并求出它们。

7.6 ★★★ （习题集 4 第 1 部分第 5(b) 题）证明：素数 $p \geq 5$ 时 $\sigma_{p-2} \equiv 0 \pmod{p^2}$。（提示：在 $(x-1)\cdots(x-(p-1)) = x^{p-1} - \sigma_1 x^{p-2} + \cdots + \sigma_{p-1}$ 中令 $x = p$。）

7.7 ★ 求 $\operatorname{ord}_{19}(2)$，再用引理 7.14 求 $\operatorname{ord}_{19}(7)$，并写出模 19 的全部原根。

7.8 ★★ （习题集 4 第 2 部分第 2 题）设 m 是奇数，$n \geq 1$。证明 $(2^m - 1, 2^n + 1) = 1$。m 为偶数时结论还成立吗？

7.9 ★★ （习题集 4 第 2 部分第 4 题）设 $n > 1$，$a^{n-1} \equiv 1 \pmod{n}$，且对 $n-1$ 的每个真因数 d 都有 $a^d \not\equiv 1 \pmod{n}$。证明 n 是素数。用 $a = 5$ 说明 97 是素数。

7.10 ★★★ （习题集 4 第 2 部分第 5 题）证明：数列 $1^1, 2^2, 3^3, \ldots$ 模素数 p 是周期的，最小正周期是 $p(p-1)$。

练习答案与提示

7.1 模 7 的根是 $3, 4$。从 3 出发：$f'(3) = 6 \equiv -1$，取 $u = -1$，$a_{j+1} \equiv a_j + f(a_j)$。$a_2 = 3 + 7 = 10$，$a_3 = 10 + 98 = 108$。验证 $108^2 - 2 = 11662 = 34 \cdot 343$。另一个根是 $-108 \equiv 235$。所以 $x \equiv 108, 235 \pmod{343}$。

7.2 8 模 3^k 可逆，$x = 2y$ 化为 $8(y^3 - 1) \equiv 0$，即 $y^3 \equiv 1$。由 Fermat 小定理 $y \equiv y^3 \equiv 1 \pmod{3}$，写 $y = 1 + 3t$，$y^3 - 1 = 9t(1 + 3t + 3t^2)$，括号与 3 互素，所以 $3^k \mid y^3 - 1 \iff 3^{k-2} \mid t \iff y \equiv 1 \pmod{3^{k-1}}$，模 3^k 有 3 个，即 $x \equiv 2 \pmod{3^{k-1}}$。2 是奇异根（$f'(2) = 12$）。

7.3 x 必为奇数，$x-1$ 和 $x+1$ 是相差 2 的偶数，不能都被 4 整除。$2^k \mid (x-1)(x+1)$，所以其中一个被 2^{k-1} 整除，即 $x \equiv \pm 1 \pmod{2^{k-1}}$，模 2^k 给出 ± 1，$2^{k-1} \pm 1$。$k \geq 3$ 时它们互不同余，且 $(2^{k-1} \pm 1)^2 = 2^{2k-2} \pm 2^k + 1 \equiv 1$。不矛盾：Lagrange 定理只对素数模成立。

7.4 记 $h = -f(a)v$，$p^j \mid h$。由 $f(a+h) = f(a) + hf'(a) + h^2 g$（$g$ 整系数），$f(b) \equiv f(a)\big(1 - vf'(a)\big) \equiv 0 \pmod{p^{2j}}$，因为两个因子分别被 p^j 整除。$f(7) = 50$，$f'(7) = 14$，$14 \cdot 134 = 3 \cdot 625 + 1$，$b = 7 - 50 \cdot 134 \equiv 182 \pmod{625}$，与例 7.1 一致。Hensel 迭代每步多一位，这里每步位数加倍。

7.5 $x^{16} - 1 = (x^8 - 1)(x^8 + 1)$，$x^8 - 1 = (x^4 - 1)(x^4 + 1)$，所以 $x^4 + 1 \mid x^{16} - 1 \mid x^{17} - x$，恰有 4 个解。$2^4 = 16 \equiv -1$，所以 2 是解。$x^4 \equiv 1$ 的根是 $\pm 1, \pm 4$，乘上它们得 $2, 15, 8, 9$，都是解。所以 $x \equiv 2, 8, 9, 15 \pmod{17}$。

7.6 左边为 $(p-1)! = \sigma_{p-1}$，与右边常数项消去，再除以 p（x 的系数是 $-\sigma_{p-2}$），得 $\sigma_{p-2} = p^{p-2} - \sigma_1 p^{p-3} + \cdots + \sigma_{p-3} p$。$p - 2 \geq 3$，除最后一项外 p 的指数都 ≥ 2；最后一项 $\sigma_{p-3} p$ 由命题 7.11 也被 p^2 整除。例：$p = 5$ 时 $\sigma_3 = 50$；$p = 3$ 时 $\sigma_1 = 3$，结论不成立。

7.7 只需检查 9 和 6：$2^9 \equiv 18$，$2^6 \equiv 7$，都不是 1，所以 $\operatorname{ord}_{19}(2) = 18$，$\operatorname{ord}(7) = \operatorname{ord}(2^6) = 18/6 = 3$。原根是 2^s，$(s, 18) = 1$：$s = 1, 5, 7, 11, 13, 17$ 给出 $2, 13, 14, 15, 3, 10$。

7.8 两数都是奇数。若素数 p 整除两者，令 $h = \operatorname{ord}_p(2)$。由 $2^n \equiv -1$ 得 $h \mid 2n$ 但 $h \nmid n$，故 h 为偶数；而 $2^m \equiv 1$ 给出 $h \mid m$，与 m 奇矛盾。m 偶时反例为 $(2^2 - 1, 2^1 + 1) = 3$。

7.9 若 $d = (a, n) > 1$，由 $a^{n-1} \equiv 1 \pmod{n}$ 得 $d \mid a^{n-1}$、$d \mid n$，进而 $d \mid 1$，矛盾。因此 $h = \operatorname{ord}_n(a)$ 存在。$h \mid n - 1$，条件排除其为真因数，故 $h = n - 1$；Euler 定理给出 $h \mid \varphi(n) \leq n - 1$，于是 $\varphi(n) = n - 1$，n 为素数。对 97，$5^{32} \equiv 35$、$5^{48} \equiv -1$、$5^{96} \equiv 1$；96 的每个真因数整除 32 或 48，对应的幂都不为 1，故判据适用。

7.10 令 $P = p(p-1)$。$n^n \bmod p$ 只由 n 模 p 与 $p-1$ 决定（$p \mid n$ 时值为 0），所以 P 是周期。设 T 为周期；取 $n = p$ 得 $p \mid T$。再取 CRT 解 $n \equiv g \pmod{p}$、$n \equiv 0 \pmod{p-1}$，其中 g 为模 p 原根，则 $n^n \equiv 1$，且 $(n+T)^{n+T} \equiv g^T$。周期性给出 $g^T \equiv 1$，故 $p - 1 \mid T$，于是 $P \mid T$。

第 8 讲

原根：个数、指标与存在性

Primitive Roots: Counting, Index Calculus, Existence

对应材料：MIT 18.781 Lecture 8 讲义（Primitive Roots (Prime Powers), Index Calculus）；习题集 4 第 1 部分第 1、6 题，第 2 部分第 1 题；习题集 5 第 2、5 题。

原题与题解的对应。习题集 4 第 2 部分第 1(a) 的题面要求求模 23^3 的原根，所附官方题解的提升计算止于模 23^2。本讲保留平方情形作为中间步骤，在证明模 p^e 的原根定理后补齐三次幂情形。

本讲目标

(1) 理解原根 g 的幂 $1, g, \dots, g^{\varphi(m)-1}$ 恰好走遍模 m 的全部简化剩余类，并由此数出原根的个数 $\varphi(\varphi(m))$。

(2) 会用「对 $\varphi(m)$ 的每个素因数 q 检验 $g^{\varphi(m)/q} \not\equiv 1$」来判断原根。

(3) 掌握指标（离散对数）：它把模 p 的乘法变成模 $p-1$ 的加法，会用指标表解一次同余式和 $x^d \equiv a \pmod{p}$。

(4) 会判断 $x^d \equiv a \pmod{p}$ 是否有解、有几个解，并理解判别式 $a^{(p-1)/(d,p-1)} \equiv 1$。

(5) 掌握原根存在定理：模 m 有原根当且仅当 $m = 1, 2, 4, p^e, 2p^e$（p 为奇素数），并能写出每一步证明。

第 7 讲证明了：对每个素数 p，存在模 p 的原根。本讲回答三个接着冒出来的问题。

- 有多少个？模 p 的原根一共几个？怎样从一个原根找出全部？
- 有什么用？原根让模 p 的每个非零剩余都能写成 g^k，于是可以像对数那样，把乘法问题化成加法问题。
- 别的模呢？模 8、模 15 有没有原根？模 25、模 50 呢？完整答案是一张很短的清单。

8.1 原根的幂与原根的个数

8.1.1 回顾：阶与原根

设 $(a, m) = 1$。第 7 讲定义了 a 模 m 的阶（order）$\operatorname{ord}_m(a)$：使 $a^h \equiv 1 \pmod{m}$ 的最小正整数 h。若 $\operatorname{ord}_m(a) = \varphi(m)$，就称 a 是模 m 的原根（primitive root）。下面两条引理来自第 7 讲，本讲反复使用。

引理 8.1（阶的两条基本性质）　设 $h=\operatorname{ord}_m(a)$。

(1) $a^k\equiv 1 \pmod m$ 当且仅当 $h\mid k$。特别地 $h\mid \varphi(m)$。

(2) 对任意正整数 k，$\operatorname{ord}_m(a^k)=\dfrac{h}{(h,k)}$。

第 7 讲的存在性证明是这样的：把 $p-1$ 分解为 $q_1^{e_1}\cdots q_r^{e_r}$，对每个素数幂 $q_i^{e_i}$，比较 $x^{q_i^{e_i}}-1$ 与 $x^{q_i^{e_i-1}}-1$ 的根数（分别是 $q_i^{e_i}$ 和 $q_i^{e_i-1}$ 个），得到一个阶恰为 $q_i^{e_i}$ 的元素 g_i；各 g_i 的阶两两互素，所以乘积 $g=g_1\cdots g_r$ 的阶是 $q_1^{e_1}\cdots q_r^{e_r}=p-1$。

那个证明只告诉我们原根存在，并没有说原根长什么样。下面从一个具体的原根出发，看它的幂能告诉我们什么。

例 8.1　模 13 的原根 2　验证 2 是模 13 的原根，并列出 $2^0,2^1,\dots,2^{11}$ 模 13 的值。

解　每一步把上一个结果乘 2 再模 13：

k	0	1	2	3	4	5	6	7	8	9	10	11
$2^k \bmod 13$	1	2	4	8	3	6	12	11	9	5	10	7

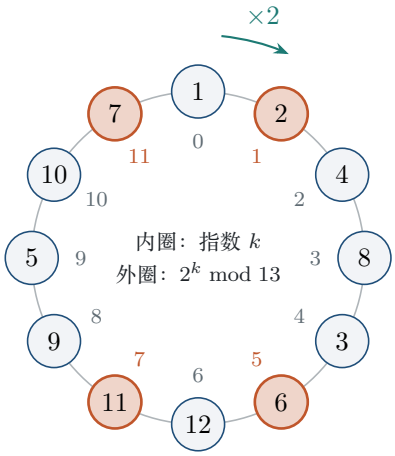
例如 $2^4=16\equiv 3$，$2^5\equiv 3\cdot 2=6$，$2^7\equiv 12\cdot 2=24\equiv 11$。表中第一次回到 1 是在 $k=12$（$2^{12}\equiv 7\cdot 2=14\equiv 1$），所以 $\operatorname{ord}_{13}(2)=12=\varphi(13)$，2 是原根。第二行恰好是 $1,2,\dots,12$ 的一个重排，每个非零剩余出现一次。 ♦

「每个非零剩余恰好出现一次」不是巧合，对任意有原根的模都成立。

命题 8.2　设 g 是模 m 的原根。则 $1,g,g^2,\dots,g^{\varphi(m)-1}$ 两两模 m 不同余，并且恰好构成模 m 的一个简化剩余系。

证明　每个 g^i 都与 m 互素，因为 $(g,m)=1$。假设 $g^i\equiv g^j \pmod m$，$0\le i<j\le \varphi(m)-1$。两边乘 g^{-i}（g 模 m 可逆）得 $g^{j-i}\equiv 1$，而 $0<j-i<\varphi(m)$，与 $\operatorname{ord}_m(g)=\varphi(m)$ 矛盾。所以这 $\varphi(m)$ 个数两两不同余。模 m 的简化剩余类一共只有 $\varphi(m)$ 个，所以它们恰好各出现一次。 ■

把这件事画成一个钟面（图 8.1）：从 1 出发，每次乘 g 就顺时针走一格，走 $\varphi(m)$ 步回到起点，途中每个简化剩余恰好经过一次。**在原根的眼里，模 m 的乘法就是在钟面上转圈。**



橙色是原根 $2, 6, 7, 11$，对应的指数 $k = 1, 5, 7, 11$ 恰好是与 12 互素的那些。

2^k 的阶是 $12/(12, k)$，只有 $(k, 12) = 1$ 时才等于 12。

图 8.1　模 13 的「原根钟面」。每次乘 2 顺时针走一格，12 步走遍全部非零剩余。

8.1.2 原根的个数

钟面上的每个元素都是 g^k，而引理 8.1(2) 给出了 g^k 的阶。于是「哪些 g^k 是原根」就变成一个关于指数 k 的问题。

定理 8.3（原根的个数）　若模 m 有原根 g，则模 m 的全部原根是

$$g^k, \qquad 1 \le k \le \varphi(m),\ (k, \varphi(m)) = 1,$$

一共 $\varphi(\varphi(m))$ 个。特别地，模素数 p 恰有 $\varphi(p-1)$ 个原根。

证明　设 a 与 m 互素。由命题 8.2，a 与 $g^0, g^1, \dots, g^{\varphi(m)-1}$ 中恰好一个同余；因为 $g^{\varphi(m)} \equiv 1 = g^0$，也可以说存在唯一的 $k \in \{1, 2, \dots, \varphi(m)\}$ 使 $a \equiv g^k$。由引理 8.1(2)，

$$\operatorname{ord}_m(a) = \operatorname{ord}_m(g^k) = \frac{\varphi(m)}{(k, \varphi(m))}.$$

它等于 $\varphi(m)$ 当且仅当 $(k, \varphi(m)) = 1$。在 $1 \le k \le \varphi(m)$ 中这样的 k 恰有 $\varphi(\varphi(m))$ 个，不同的 k 给出不同的剩余类，所以原根恰有 $\varphi(\varphi(m))$ 个。 ∎

证明里真正起作用的一步是：**原根把「求阶」化成了「求最大公因数」**。同样的推理还告诉我们每种阶各有多少个元素。

例 8.2　模 13 各元素的阶　求模 13 每个非零剩余的阶，并统计每种阶的元素个数。

解　用原根 $g = 2$ 和例中的幂表。2^k 的阶是 $12/(k, 12)$：

阶 d	对应的指数 k（$(k,12)=12/d$）	元素 $2^k \bmod 13$	个数
1	0	1	1
2	6	12	1
3	4, 8	3, 9	2
4	3, 9	8, 5	2
6	2, 10	4, 10	2
12	1, 5, 7, 11	2, 6, 11, 7	4

例如 $k=9$：$(9,12)=3$，阶为 $12/3=4$，验证 $5^2=25\equiv 12$，$5^4\equiv 12^2=144\equiv 1$，而 $5^2\not\equiv 1$，阶确实是 4。最后一列是 $1,1,2,2,2,4$，恰好是 $\varphi(1),\varphi(2),\varphi(3),\varphi(4),\varphi(6),\varphi(12)$，加起来是 12。 ♦

一般地，对 $p-1$ 的每个因数 d，模 p 中阶为 d 的元素恰有 $\varphi(d)$ 个（对应 $(k,p-1)=(p-1)/d$ 的那些 k）。把所有 d 加起来就得到恒等式 $\sum_{d\mid n}\varphi(d)=n$，这个恒等式在第 13 讲会再次出现。

想一想　$\varphi(\varphi(8))=\varphi(4)=2$。能否由此说「模 8 有 2 个原根」？

答　不能。定理 8.3 的前提是「模 m 有原根」。模 8 的简化剩余是 $1,3,5,7$，而 $3^2=9$、$5^2=25$、$7^2=49$ 都 $\equiv 1 \pmod 8$，最大的阶是 $2<\varphi(8)=4$，所以模 8 一个原根也没有。公式 $\varphi(\varphi(m))$ 只在「至少有一个原根」时才数得对。哪些 m 有原根，是第 8.6 节的主题。

8.1.3　怎样判断一个数是不是原根

按定义检验 g 是不是原根，要算 $g,g^2,\dots,g^{\varphi(m)-1}$，确认都不是 1。下面的判别法只需算少数几个幂。

命题 8.4（原根判别法）　设 $(g,m)=1$，$\varphi(m)$ 的不同素因数为 $q_1,\dots,q_r$。则 g 是模 m 的原根当且仅当

$$g^{\varphi(m)/q_i}\not\equiv 1 \pmod m, \qquad i=1,\dots,r.$$

证明　若 g 是原根，$\varphi(m)/q_i<\varphi(m)$，当然 $g^{\varphi(m)/q_i}\not\equiv 1$。反过来，设 $h=\operatorname{ord}_m(g)$，则 $h\mid\varphi(m)$。若 $h<\varphi(m)$，则 $\varphi(m)/h>1$ 有某个素因数 q_i，于是 $h\mid\varphi(m)/q_i$，由引理 8.1(1) 得 $g^{\varphi(m)/q_i}\equiv 1$，与假设矛盾。所以 $h=\varphi(m)$。 ■

例 8.3　模 23 的原根，改编自习题集 4 第 2 部分第 1(a)　找出模 23 的一个原根，并写出模 23 的全部原根。

解　$\varphi(23)=22=2\cdot 11$，所以只需检验 g^2 和 g^{11} 都不是 1。

试 $g=2$：$2^5=32\equiv 9$，$2^{10}\equiv 81\equiv 12$，$2^{11}\equiv 24\equiv 1$。不是原根。

试 $g=3$：$3^3=27\equiv 4$，$3^9\equiv 4^3=64\equiv 18$，$3^{11}=3^9\cdot 3^2\equiv 18\cdot 9=162\equiv 1$。不是原根。

试 $g=5$：$5^2=25\equiv 2\not\equiv 1$；$5^{11}=(5^2)^5\cdot 5\equiv 2^5\cdot 5=160\equiv 160-161=-1\not\equiv 1$。所以 5 是

原根。

由定理 8.3，全部原根是 5^k，k 取 $1 \le k \le 22$ 中与 22 互素的数 $1,3,5,7,9,13,15,17,19,21$，共 $\varphi(22)=10$ 个。逐个计算得

$$5,\ 10,\ 20,\ 17,\ 11,\ 21,\ 19,\ 15,\ 7,\ 14,$$

即模 23 的原根为 $5,7,10,11,14,15,17,19,20,21$。

♦

易错提醒　判别法要对 $\varphi(m)$ 的每一个素因数检验。模 23 时 $2^{22/11}=2^2=4\not\equiv 1$，若只检验 $q=11$ 就会误判 2 为原根，实际上 $2^{22/2}=2^{11}\equiv 1$。另一个常见错误是把指数 $\varphi(m)/q_i$ 写成 q_i：模 13 时 $\varphi(13)=12$ 的素因数是 $2,3$，要检验的是 g^6 和 g^4，而不是 g^2 和 g^3。

历史与背景　模 p 的原根有 $\varphi(p-1)$ 个。构造时通常先分解 $p-1$，再用命题 8.4 检验候选数。一个固定的整数 a 是不是对无穷多个 p 都是原根？Artin 在 1927 年提出猜想：只要 $a\neq -1$ 且 a 不是完全平方数，答案就是肯定的，而且这样的 p 在素数中占正比例。对 $a=2$，猜想的比例是 Artin 常数 $\prod_p\left(1-\frac{1}{p(p-1)}\right)\approx 0.3740$；在 10^5 以内的 9591 个奇素数中，2 是原根的有 3603 个，约 37.6%。Hooley（1967）在广义 Riemann 猜想成立的前提下证明了猜想；Heath-Brown（1986）无条件地证明：至多有两个素数 a 使猜想失败，但没有指出它们。按本讲原始课程材料（2012 春）的研究状态，当时尚无具体的 a 已被无条件证明满足猜想。

想一想　为什么 Artin 猜想要排除完全平方数？

答　设 $a=b^2$，$p\nmid a$ 是奇素数。由 Fermat 小定理，$a^{(p-1)/2}=b^{p-1}\equiv 1 \pmod p$，所以 $\operatorname{ord}_p(a)\le (p-1)/2$，$a$ 对任何奇素数都不是原根。

8.2　指标：模 p 的离散对数

8.2.1　定义与指标表

实数里，对数把乘法变成加法：$\log(xy)=\log x+\log y$。有了原根，模 p 也能这样做。

定义 8.5（指标）　设 p 是素数，g 是模 p 的原根，$p\nmid a$。由命题 8.2，存在唯一的 $k\in\{0,1,\dots,p-2\}$ 使 $g^k\equiv a \pmod p$。这个 k 称为 a 以 g 为底的**指标**（index），或**离散对数**（discrete logarithm），记作 $\operatorname{ind}_g a$，底数明确时简写为 $\operatorname{ind} a$。

指标只对与 p 互素的 a 有定义（0 不是 g 的任何幂），并且依赖于底 g 的选取。对一般的有原根的模 m，可以同样定义，指标取值于 $0,1,\dots,\varphi(m)-1$。

把例 8.1 的幂表「倒过来查」，就得到模 13、底为 2 的指标表：

a	1	2	3	4	5	6	7	8	9	10	11	12
$\operatorname{ind}_2 a$	0	1	4	2	9	5	11	3	8	10	7	6

表 8.1　模 13 的指标表（底 $g=2$）。本讲后面的模 13 计算都查这张表。

命题 8.6（指标的运算律）　设 g 是模 p 的原根，a, b 都与 p 互素，$n \geq 0$。

(1) $g^k \equiv a \pmod p$ 当且仅当 $k \equiv \operatorname{ind} a \pmod{p-1}$。

(2) $\operatorname{ind}(ab) \equiv \operatorname{ind} a + \operatorname{ind} b \pmod{p-1}$。

(3) $\operatorname{ind}(a^n) \equiv n \operatorname{ind} a \pmod{p-1}$。

(4) $\operatorname{ind} 1 = 0$；当 p 为奇素数时 $\operatorname{ind}(-1) = \dfrac{p-1}{2}$。

证明　(1) $g^k \equiv a \equiv g^{\operatorname{ind} a}$ 当且仅当 $g^{k-\operatorname{ind} a} \equiv 1$（$k < \operatorname{ind} a$ 时两边同乘 $g^{\operatorname{ind} a - k}$ 也一样），由引理 8.1(1)，当且仅当 $p-1 \mid k - \operatorname{ind} a$。

(2) $g^{\operatorname{ind} a + \operatorname{ind} b} = g^{\operatorname{ind} a} g^{\operatorname{ind} b} \equiv ab$，再用 (1)。(3) 同理，$g^{n \operatorname{ind} a} = (g^{\operatorname{ind} a})^n \equiv a^n$。

(4) $g^0 = 1$。令 $c = g^{(p-1)/2}$，则 $c^2 = g^{p-1} \equiv 1$，所以 $p \mid (c-1)(c+1)$，$c \equiv \pm 1$。又 $(p-1)/2 < p-1$，$c \neq 1$，只能 $c \equiv -1$。 ■

对照表 8.1：$\operatorname{ind} 12 = \operatorname{ind}(-1) = 6 = (13-1)/2$，与 (4) 一致。

直观理解　指标就是模 p 的对数，它把两种运算对应起来（图 8.2）：

模 p 的乘法（$p-1$ 个非零剩余）　$\longleftrightarrow$　模 $p-1$ 的加法（指数 $0, \ldots, p-2$）。

乘积对应指标之和，幂对应指标的倍数。在钟面图 8.1 上，$\operatorname{ind} a$ 就是从 1 顺时针走到 a 所需的格数。

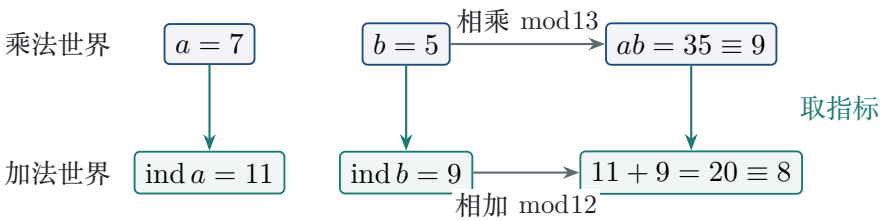


图 8.2　指标把模 13 的乘法变成模 12 的加法。查表 8.1：$\operatorname{ind} 9 = 8$，与 $11 + 9 \equiv 8$ 一致。

指标的运算在模 $p-1$ 下进行　同余式 $a \equiv b \pmod p$ 取指标后变成 $\operatorname{ind} a \equiv \operatorname{ind} b \pmod{p-1}$，模数从 p 变成了 $p-1$。模 13 时 $\operatorname{ind} 7 + \operatorname{ind} 5 = 20$，要化成 $20 - 12 = 8$，不能化成 $20 - 13 = 7$（$\operatorname{ind} 7 = 11$ 对应的数是 7，而 $7 \cdot 5 = 35 \equiv 9 \neq 7$）。同理，解出的指标方程 $dk \equiv \ell$ 是模 $p-1$ 的一次同余式，要用第 5 讲的方法在模 $p-1$ 下解。

8.2.2 用指标解同余式

■ **例 8.4　用指标解一次同余式**　用表 8.1 解 $7x \equiv 5 \pmod{13}$。

解　$x \not\equiv 0$（否则左边为 0），所以可以取指标。两边取指标：

$$\operatorname{ind} 7 + \operatorname{ind} x \equiv \operatorname{ind} 5 \pmod{12}, \qquad 11 + \operatorname{ind} x \equiv 9 \pmod{12},$$

得 $\operatorname{ind} x \equiv -2 \equiv 10 \pmod{12}$。查表，指标为 10 的数是 10，所以 $x \equiv 10$。验证：$7 \cdot 10 = 70 = 5 \cdot 13 + 5 \equiv 5$。

当然，第 5 讲的方法（求 7 的逆元）也能解这道题。指标的好处是：一旦有了表，所有乘除法都变成查表与加减，对下面的高次同余式尤其有用。 ♦

离散对数与密码学　已知 g, k 时，用第 6 讲的反复平方法计算 $g^k \bmod p$ 很快；给定 g 和 $y \equiv g^k \pmod{p}$，反过来求 k 称为离散对数问题（DLP）。对适当选取的密码学参数，经典计算模型下尚无已知的多项式时间算法。Diffie–Hellman 密钥交换直接涉及的是计算 Diffie–Hellman 问题（CDH）：给定 $g, g^a, g^b \pmod{p}$，求 $g^{ab} \pmod{p}$。两者相关但不是同一个问题：若能求离散对数，就能恢复 a, b 并算出 g^{ab}；CDH 的任务则只要求算出这个共享值。参数是否安全还要看具体群和可用攻击，不能只凭 p 的位数判断。

小例子只展示运算关系：公开 $p = 23$、$g = 5$；甲秘密选 $a = 6$，公开 $5^6 \equiv 8$；乙秘密选 $b = 15$，公开 $5^{15} \equiv 19$。甲算 $19^6 \equiv 2$，乙算 $8^{15} \equiv 2$，双方得到 $5^{ab} \equiv 2$，旁听者只看到 8 和 19。第 6 讲的 RSA 以因数分解困难作为安全性假设；Diffie–Hellman 通常以所用群上的 CDH 难以计算作为安全性假设。

8.3 用指标解 $x^d \equiv a \pmod{p}$

8.3.1 化为一次同余式

现在来看本讲原讲义的核心应用。考虑同余式 $x^d \equiv a \pmod{p}$，其中 $p \nmid a$。任何解 x 都与 p 互素，所以可以写成 $x \equiv g^k$。代入并取指标：

$$x^d \equiv a \iff g^{dk} \equiv g^{\operatorname{ind} a} \iff dk \equiv \operatorname{ind} a \pmod{p-1}.$$

未知数从 x 换成了指数 k，d 次同余式变成了一次同余式。第 5 讲讲过一次同余式 $dk \equiv \ell \pmod{n}$ 的解法：有解当且仅当 $(d, n) \mid \ell$，有解时恰有 (d, n) 个模 n 的解。取 $n = p - 1$ 就得到下面的定理。

定理 8.7（d 次剩余）　设 p 是素数，$p \nmid a$，$d \geq 1$，记 $\delta = (d, p-1)$。

(1) $x^d \equiv 1 \pmod{p}$ 恰有 δ 个解，它们是 g^k，k 取 $0 \leq k < p-1$ 中 $\frac{p-1}{\delta}$ 的倍数。

(2) $x^d \equiv a \pmod{p}$ 有解 $\iff \delta \mid \operatorname{ind} a \iff a^{(p-1)/\delta} \equiv 1 \pmod{p}$。有解时恰有 δ 个解。

证明　(2) 前面已化为 $dk \equiv \operatorname{ind} a \pmod{p-1}$，第一个等价和解数由一次同余式的理论得到，并

且不同的 $k \bmod (p-1)$ 给出不同的 $x \bmod p$（命题 8.2）。第二个等价：

$$\delta \mid \operatorname{ind} a \iff p-1 \,\Big|\, \frac{p-1}{\delta}\operatorname{ind} a \iff g^{\frac{p-1}{\delta}\operatorname{ind} a} \equiv 1 \iff a^{(p-1)/\delta} \equiv 1 \pmod p,$$

中间一步用了引理 8.1(1)，最后一步因为 $g^{\operatorname{ind} a} \equiv a$。

(1) 是 $a=1$ 的情形：$\operatorname{ind} 1 = 0$，方程 $dk \equiv 0 \pmod{p-1}$ 等价于 $\frac{p-1}{\delta} \mid \frac{d}{\delta}k$，而 $\frac{p-1}{\delta}$ 与 $\frac{d}{\delta}$ 互素，所以等价于 $\frac{p-1}{\delta} \mid k$。在 $0 \le k < p-1$ 中这样的 k 有 δ 个。 ■

判别式 $a^{(p-1)/\delta} \equiv 1$ 的好处是**不需要知道原根**：只要用反复平方法算一个幂。取 $d=2$，$\delta = 2$，它就是「a 是模 p 的平方数当且仅当 $a^{(p-1)/2} \equiv 1$」，即第 9 讲开头的 Euler 判别法。

8.3.2 例题

例 8.5　四次同余式　解 $x^4 \equiv 3 \pmod{13}$ 和 $x^4 \equiv 5 \pmod{13}$。

解　$d=4$，$p-1=12$，$\delta = (4,12) = 4$。

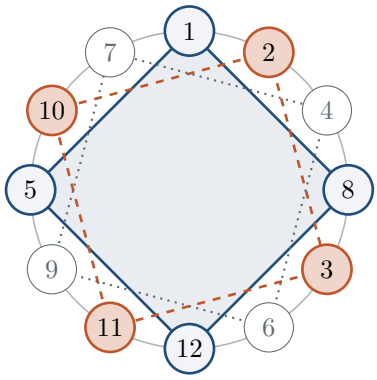
$x^4 \equiv 3$。查表 8.1，$\operatorname{ind} 3 = 4$，$4 \mid 4$，有 4 个解。设 $x \equiv 2^k$，指标方程为

$$4k \equiv 4 \pmod{12} \iff k \equiv 1 \pmod 3 \iff k \in \{1,4,7,10\}.$$

查幂表得 $x \equiv 2^1, 2^4, 2^7, 2^{10} \equiv 2, 3, 11, 10$。验证 $x=3$：$3^4 = 81 = 6\cdot 13 + 3$；$x = 10 \equiv -3$，$x = 11 \equiv -2$ 的四次方分别与 3^4、2^4 相同；$2^4 = 16 \equiv 3$。四个解是 $x \equiv 2, 3, 10, 11 \pmod{13}$。

$x^4 \equiv 5$。$\operatorname{ind} 5 = 9$，而 $4 \nmid 9$，无解。不查表也能判断：$5^{12/4} = 5^3 = 125 = 9\cdot 13 + 8 \equiv 8 \not\equiv 1$，由定理 8.7(2) 无解。 ♦

图 8.3 画出了解的位置。$x^4 \equiv 1$ 的四个解在钟面上构成一个正方形（指数 $0,3,6,9$）；$x^4 \equiv 3$ 的解是把这个正方形转一格（乘一个特解 2）得到的。一般地，**$x^d \equiv a$ 的全部解等于一个特解乘上 $x^d \equiv 1$ 的全部解**，后者在钟面上是一个正 δ 边形。



蓝色实线：$x^4 \equiv 1$ 的解 $1, 8, 12, 5$（指数 $0, 3, 6, 9$）。

橙色虚线：$x^4 \equiv 3$ 的解 $2, 3, 11, 10$（指数 $1, 4, 7, 10$）。

灰色点线：$x^4 \equiv 9$ 的解 $4, 6, 9, 7$（指数 $2, 5, 8, 11$）。

三个正方形恰好铺满 12 个非零剩余，所以四次剩余只有 $1, 3, 9$。

图 8.3　在钟面上看 $x^4 \equiv 1$ 与 $x^4 \equiv 3 \pmod{13}$ 的解：橙色正方形是蓝色正方形转了一格。

例 8.6　$\delta = 1$ 的情形　解 $x^5 \equiv 3 \pmod{13}$。

解　$\delta=(5,12)=1$，所以对任何 a 都恰有一个解。设 $x\equiv 2^k$，$\operatorname{ind}3=4$，指标方程为 $5k\equiv 4 \pmod{12}$。$5\cdot 5=25\equiv 1 \pmod{12}$，所以 5 模 12 的逆是 5，$k\equiv 5\cdot 4=20\equiv 8 \pmod{12}$，$x\equiv 2^8\equiv 9$。

验证：$9^2=81\equiv 3$，$9^4\equiv 3^2=9$，$9^5\equiv 9\cdot 9=81\equiv 3$。

换个角度：$5\cdot 5\equiv 1 \pmod{12}$，所以 $(x^5)^5=x^{25}=x^{24}\cdot x\equiv x$，映射 $x\mapsto x^5$ 的逆还是它自己。于是 $x\equiv 3^5=243=18\cdot 13+9\equiv 9$，结果相同，而且不需要指标表。一般地，$(d,p-1)=1$ 时取 e 使 $de\equiv 1 \pmod{p-1}$，$x^d\equiv a$ 的唯一解就是 $x\equiv a^e$。 ♦

■ **例 8.7**　**模 $p\equiv 2 \pmod 3$ 的立方根，改编自习题集 5**　设素数 $p\equiv 2 \pmod 3$，$p\nmid a$。证明 $x^3\equiv a \pmod p$ 有唯一解 $x\equiv a^{(2p-1)/3}$，并对 $p=11$，$a=2$ 求出它。

解　$p\equiv 2 \pmod 3$ 推出 $p-1\equiv 1 \pmod 3$，所以 $(3,p-1)=1$，由定理 8.7，解唯一。又 $2p-1\equiv 4-1\equiv 0 \pmod 3$，$(2p-1)/3$ 是整数。验证它是解：

$$\left(a^{(2p-1)/3}\right)^3=a^{2p-1}=a\cdot\left(a^{p-1}\right)^2\equiv a\cdot 1=a \pmod p.$$

（用指数的语言说：$3\cdot\frac{2p-1}{3}=2(p-1)+1\equiv 1 \pmod{p-1}$，$\frac{2p-1}{3}$ 就是 3 模 $p-1$ 的逆。）

$p=11$，$a=2$：$(2p-1)/3=7$，$x\equiv 2^7=128=11\cdot 11+7\equiv 7$。验证：$7^3=343=31\cdot 11+2\equiv 2$。 ♦

■ **例 8.8**　**立方剩余**　判断 2 和 3 是否是模 31 的立方数；若是，求出全部立方根。

解　$p-1=30$，$\delta=(3,30)=3$，判别式是 $a^{30/3}=a^{10}\equiv 1$。

$a=2$：$2^5=32\equiv 1$，所以 $2^{10}\equiv 1$，有 3 个解。$a=3$：$3^5=243=7\cdot 31+26\equiv -5$，$3^{10}\equiv 25\not\equiv 1$，无解。

求 2 的立方根。试小数字：$4^3=64=2\cdot 31+2\equiv 2$，所以 $x\equiv 4$ 是一个解。其余解是 4 乘以 $x^3\equiv 1$ 的解。$x^3\equiv 1$ 的解除了 1 还有两个：$5^3=125=4\cdot 31+1$，所以 5 是一个，另一个是 $5^2=25$（$25^3=(5^3)^2\equiv 1$）。于是全部解为

$$x\equiv 4,\quad 4\cdot 5=20,\quad 4\cdot 25=100=3\cdot 31+7\equiv 7 \pmod{31}.$$

验证 $20^3=8000=258\cdot 31+2$，$7^3=343=11\cdot 31+2$，都 $\equiv 2$。 ♦

> **推论 8.8（d 次剩余的个数）**　模 p 的非零剩余中，能写成 x^d 的恰有 $\dfrac{p-1}{(d,p-1)}$ 个。特别地，当 $(d,p-1)=1$ 时 $x\mapsto x^d$ 是非零剩余的一个置换。

证明　$p-1$ 个非零的 x 被映到 d 次方，每个能取到的值恰好被 $\delta=(d,p-1)$ 个 x 取到（定理 8.7），所以能取到的值有 $(p-1)/\delta$ 个。 ■

模 13 的四次剩余有 $12/4=3$ 个，正是图 8.3 中的 $1,3,9$；模 31 的立方剩余有 $30/3=10$ 个。

原根除了用来解方程，也能直接用来求和。下面这个结果出自习题集 4。

■ **例 8.9　幂和，习题集 4 第 1 部分第 6 题**　设 p 为素数，$k \geq 1$。证明

$$S_k = 1^k + 2^k + \cdots + (p-1)^k \equiv \begin{cases} -1 \pmod p, & (p-1) \mid k, \\ 0 \pmod p, & (p-1) \nmid k. \end{cases}$$

解　取原根 g。$1, 2, \ldots, p-1$ 模 p 恰好是 $g^0, g^1, \ldots, g^{p-2}$ 的重新排列（命题 8.2），所以

$$S_k \equiv \sum_{j=0}^{p-2} \left(g^j\right)^k = \sum_{j=0}^{p-2} r^j, \qquad r = g^k.$$

若 $(p-1) \mid k$，则 $r \equiv 1$，和式是 $p-1$ 个 1，$S_k \equiv p-1 \equiv -1$。

若 $(p-1) \nmid k$，则 $r = g^k \neq 1$（g 的阶是 $p-1$），$r-1$ 可逆。用等比数列求和：

$$(r-1)S_k \equiv r^{p-1} - 1 \equiv 0 \pmod p,$$

最后一步用了费马小定理。两边乘以 $(r-1)^{-1}$ 得 $S_k \equiv 0$。

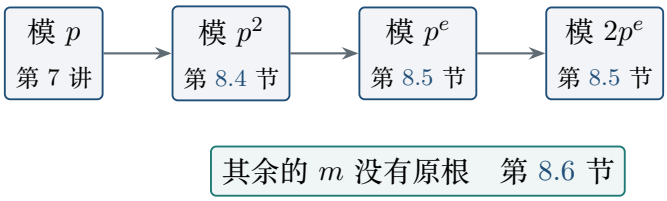
检验 $p=7$：$1^2 + \cdots + 6^2 = 91 = 13 \cdot 7 \equiv 0$，$1^3 + \cdots + 6^3 = 441 = 63 \cdot 7 \equiv 0$。对 $k = 1, \ldots, 12$ 逐个计算，$S_k \bmod 7$ 依次是 $0,0,0,0,0,6,0,0,0,0,0,6$，只有 $k = 6, 12$ 处等于 -1。 ♦

8.4　模 p^2 的原根

现在回答引言中的第三个问题：哪些模数有原根？第 7 讲证明了素数 p 有原根。本讲的主定理是：

> **定理 8.9（原根存在定理）**　模 m 有原根，当且仅当 $m = 1, 2, 4, p^e, 2p^e$，其中 p 是奇素数，$e \geq 1$。

证明分四步，路线如下：



本节完成第一步。思路和 Hensel 引理（第 6、7 讲）一样：已知模 p 的原根 g，在它的 p 个「提升」$g,\ g+p,\ \ldots,\ g+(p-1)p$ 中寻找模 p^2 的原根。

8.4.1　先看一个例子

■ **例 8.10　模 25 的提升**　2 和 3 是模 5 的原根。它们模 25 的提升中，哪些是模 25 的原根？（$\varphi(25) = 20$。）

解　逐个计算阶。以 7 为例：$7^2 = 49 \equiv -1 \pmod{25}$，所以 $7^4 \equiv 1$，$\operatorname{ord}_{25} 7 = 4$。对其余的数，只需检验 $x^4 \neq 1$ 和 $x^{10} \neq 1$（$20 = 2^2 \cdot 5$，命题 8.4）。结果是

2 的提升	2	7	12	17	22
模 25 的阶	20	**4**	20	20	20
3 的提升	3	8	13	18	23
模 25 的阶	20	20	20	**4**	20

每组 5 个提升里恰好坏了一个。好的共 8 个：$2,3,8,12,13,17,22,23$，与 $\varphi(\varphi(25))=\varphi(20)=8$ 一致。 ♦

两个现象值得解释：坏的提升的阶是 $4=p-1$，不是别的数；每组恰好坏一个。坏的那个也有规律：$2^5=32\equiv 7$，$3^5=243\equiv 18 \pmod{25}$，都是 g^p。

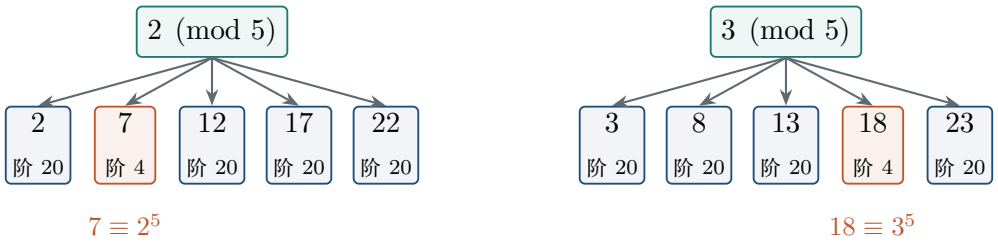


图 8.4　模 5 的原根 2,3 各有 5 个模 25 的提升。每组恰有一个提升（橙色）的阶掉到 $p-1=4$，它就是 $g^p \bmod p^2$；其余的都是模 25 的原根。

8.4.2 恰有一个坏的提升

引理 8.10　设 g 是模 p 的原根，$g'\equiv g \pmod p$。则 $\operatorname{ord}_{p^2} g'$ 只能是 $p-1$ 或 $p(p-1)$。因此 g' 是模 p^2 的原根，当且仅当 $g'^{p-1}\not\equiv 1 \pmod{p^2}$。

证明　记 $h=\operatorname{ord}_{p^2} g'$。一方面 $h\mid\varphi(p^2)=p(p-1)$。另一方面 $g'^h\equiv 1 \pmod{p^2}$ 推出 $g^h\equiv g'^h\equiv 1 \pmod p$，而 $\operatorname{ord}_p g=p-1$，所以 $(p-1)\mid h$。$p(p-1)$ 的因数中是 $p-1$ 倍数的只有 $p-1$ 和 $p(p-1)$（因为 p 是素数）。 ■

定理 8.11（模 p^2 的原根）　设 g 是模 p 的原根。在 g 的 p 个提升 $g+tp$（$t=0,1,\dots,p-1$）中，恰有一个不是模 p^2 的原根，它就是 $g^p \bmod p^2$。特别地，模 p^2 有原根。

证明　由引理，只需数出有多少个 t 使 $(g+tp)^{p-1}\equiv 1 \pmod{p^2}$。用二项式定理展开，含 p^2 的项全部丢掉：

$$(g+tp)^{p-1}\equiv g^{p-1}+(p-1)g^{p-2}\cdot tp\equiv g^{p-1}-g^{p-2}tp \pmod{p^2}.$$

由费马小定理，$g^{p-1}=1+cp$，c 是某个整数。代入得

$$(g+tp)^{p-1}\equiv 1+p\big(c-g^{p-2}t\big) \pmod{p^2}.$$

所以 $(g+tp)^{p-1}\equiv 1 \pmod{p^2}$ 当且仅当 $g^{p-2}t\equiv c \pmod p$。$p\nmid g^{p-2}$，这个关于 t 的一次同余式模 p 恰有一个解。于是恰有一个坏的提升，其余 $p-1$ 个都是原根。

坏的提升是哪个？$g^p \equiv g \pmod p$（费马），所以 $g^p \bmod p^2$ 是 g 的一个提升；又 $(g^p)^{p-1} = g^{p(p-1)} = g^{\varphi(p^2)} \equiv 1 \pmod{p^2}$（欧拉定理），所以它就是坏的那个。 ■

直观理解　证明的关键一步，是 $(g+tp)^{p-1}$ 的「第二位数字」（p 进制下 p^1 那一位）关于 t 是一次的：$c - g^{p-2}t$。一次函数模 p 恰好在一个点为零，所以恰好坏一个。这和 Hensel 引理里「修正量满足一次同余式」是同一个现象。

■ **例 8.11**　**原题的中间步骤：模 $23^2 = 529$　5 是模 23 的原根（例 8.3）。它是不是模 529 的原根？哪个提升是坏的？模 529 共有几个原根？**

解　由引理 8.10，只需算 $5^{22} \bmod 529$。反复平方：$5^4 = 625 \equiv 96$，$5^8 \equiv 96^2 = 9216 \equiv 223$，$5^{16} \equiv 223^2 = 49729 \equiv 3$，于是

$$5^{22} = 5^{16} \cdot 5^4 \cdot 5^2 \equiv 3 \cdot 96 \cdot 25 = 7200 = 13 \cdot 529 + 323 \equiv 323 \not\equiv 1.$$

所以 5 是模 529 的原根。坏的提升是 $5^{23} \equiv 5 \cdot 323 = 1615 = 3 \cdot 529 + 28 \equiv 28$，确实 $28 = 5 + 23$。模 529 的原根个数是 $\varphi(\varphi(529)) = \varphi(506) = \varphi(2)\varphi(11)\varphi(23) = 220$；与定理一致：模 23 有 $\varphi(22) = 10$ 个原根，每个有 22 个好的提升，$10 \cdot 22 = 220$。 ♦

模 p 的原根不一定是模 p^2 的原根　很多人默认「g 是模 p 的原根，g 本身就是模 p^2 的原根」。小例子里常常如此，但反例存在：14 是模 29 的原根，而 $14^{28} \equiv 1 \pmod{841}$，所以 14 不是模 29^2 的原根。这是最小的反例，下两个是 $p = 37$，$g = 18$ 和 $p = 43$，$g = 19$。补救很简单：换成 $g + p$。由定理 8.11，g 与 $g + p$ 不会同时是坏的；这里 43 模 841 的阶是 $812 = \varphi(841)$。

8.5　模 p^e 与 $2p^e$ 的原根

从 p^2 到 p^e 不再需要挑选：模 p^2 的原根自动是所有更高次幂的原根。关键是下面这个关于「离 1 有多近」的引理。

引理 8.12（升幂）　设 p 是奇素数，$j \geq 1$，$p \nmid b$。则

$$(1 + bp^j)^p = 1 + b'p^{j+1}, \qquad b' \equiv b \pmod p, \text{ 特别地} p \nmid b'.$$

证明　二项式展开：

$$(1 + bp^j)^p = 1 + \binom{p}{1} bp^j + \binom{p}{2} b^2 p^{2j} + \sum_{i \geq 3} \binom{p}{i} b^i p^{ij}.$$

第二项是 bp^{j+1}。第三项：$\binom{p}{2} = \frac{p(p-1)}{2}$，因为 p 是奇数，$\frac{p-1}{2}$ 是整数，所以 $\binom{p}{2}$ 被 p 整除，第三项被 p^{2j+1} 整除，而 $2j + 1 \geq j + 2$。$i \geq 3$ 的项被 p^{3j} 整除，$3j \geq j + 2$。于是 $(1 + bp^j)^p \equiv 1 + bp^{j+1} \pmod{p^{j+2}}$，即为所求。 ■

$p=2$ 时引理不成立　证明用到了 $\binom{p}{2}$ 被 p 整除，$p=2$ 时 $\binom{2}{2}=1$，这一步失效。反例：$j=1$，$b=1$，$(1+2)^2=9=1+2\cdot 2^2$，得到的 $b'=2$ 是偶数，结论 $p\nmid b'$ 不成立；实际上 $9=1+2^3$，一步就升了两次幂。所以 2 的方幂要单独处理（第 8.6 节）。

定理 8.13（模 p^e 的原根）　设 p 是奇素数，g 是模 p^2 的原根。则对所有 $e\geq 2$，g 是模 p^e 的原根。

证明　g 是模 p^2 的原根，所以 $g^{p-1}\equiv 1 \pmod p$ 但 $g^{p-1}\not\equiv 1 \pmod{p^2}$，即 $g^{p-1}=1+bp$，$p\nmid b$。反复用引理 8.12：

$$g^{(p-1)p^{e-2}}=1+b_e p^{e-1},\qquad p\nmid b_e.$$

所以 $g^{(p-1)p^{e-2}}\not\equiv 1 \pmod{p^e}$。

记 $h=\operatorname{ord}_{p^e} g$。$g^h\equiv 1 \pmod{p^e}$ 推出 $g^h\equiv 1 \pmod p$，所以 $(p-1)\mid h$；又 $h\mid \varphi(p^e)=(p-1)p^{e-1}$。于是 $h=(p-1)p^i$，$0\leq i\leq e-1$。若 $i\leq e-2$，则 $h\mid (p-1)p^{e-2}$，与上式矛盾。所以 $h=(p-1)p^{e-1}=\varphi(p^e)$。 ■

核心思想　三步合起来：取模 p 的原根 g；若 $g^{p-1}\not\equiv 1 \pmod{p^2}$ 就用 g，否则用 $g+p$；得到的数是所有 p^e 的原根。检验只需要在模 p^2 做一次。

■ **例 8.12　完成习题集 4 第 2 部分第 1(a)：模 23^3**　求模 $23^3=12167$ 的一个原根，并写出它的阶。

解　先用已经核对过的平方情形，再应用提升定理。例 8.11 中 $5^{22}\equiv 323\not\equiv 1 \pmod{529}$，已经说明 5 是模 23^2 的原根。23 是奇素数，由定理 8.13，同一个整数 5 也是模 23^3 的原根，其阶为 $\varphi(23^3)=23^2(23-1)=529\cdot 22=11638$。这一步将官方题解中的平方计算延伸到了原题要求的三次幂；不需要从头列出模 12167 的全部幂。 ♦

■ **例 8.13　2 是所有 5^e 的原根**　证明 2 是模 5^e 的原根（$e\geq 1$），并观察 $2^{\varphi(5^{e-1})}$ 离 1 有多近。

解　2 是模 5 的原根，且 $2^4=16=1+3\cdot 5\not\equiv 1 \pmod{25}$，由引理 8.10，2 是模 25 的原根（这和图 8.4 一致），再由定理 8.13，它是所有 5^e 的原根。

定理证明中的 b_e 可以直接看到。$2^4=16=1+3\cdot 5$，然后

$$\begin{aligned}2^{20}&\equiv 76=1+3\cdot 5^2 \pmod{125},\\ 2^{100}&\equiv 376=1+3\cdot 5^3 \pmod{625},\\ 2^{500}&\equiv 1876=1+3\cdot 5^4 \pmod{3125}.\end{aligned}$$

每做一次 p 次方，「1+ 后面那一项」恰好多一个因子 5，而系数模 5 保持为 3，和引理 8.12 说的一样。阶依次是 $4,20,100,500,2500$，每次乘 5。 ♦

最后处理 $2p^e$。

推论 8.14（模 $2p^e$ 的原根）　设 p 是奇素数，g 是模 p^e 的原根。则 g 与 $g+p^e$ 中的奇数是模 $2p^e$ 的原根。

证明　记 G 为其中的奇数，则 $G\equiv g \pmod{p^e}$，$(G,2p^e)=1$。由中国剩余定理，$G^k\equiv 1 \pmod{2p^e}$ 当且仅当 $G^k\equiv 1 \pmod 2$（对奇数 G 总成立）且 $G^k\equiv 1 \pmod{p^e}$。所以 $\operatorname{ord}_{2p^e} G=\operatorname{ord}_{p^e} G=\varphi(p^e)$。而 $\varphi(2p^e)=\varphi(2)\varphi(p^e)=\varphi(p^e)$，所以 G 是原根。 ■

■ **例 8.14　模 50 与模 54**　求模 50 和模 54 的一个原根。

解　模 $50=2\cdot 5^2$：2 是模 25 的原根，但它是偶数，换成 $2+25=27$。$\varphi(50)=20$，检验 $27^4\not\equiv 1$，$27^{10}\not\equiv 1 \pmod{50}$ 即可确认 $\operatorname{ord}_{50} 27=20$。（3 也是模 25 的原根，而且是奇数，所以 3 同样是模 50 的原根。）

模 $54=2\cdot 3^3$：2 是模 3 的原根，$2^2=4\not\equiv 1 \pmod 9$，所以 2 是模 9、从而是模 27 的原根，阶 18。它是偶数，换成 $2+27=29$，$\operatorname{ord}_{54} 29=18=\varphi(54)$。 ♦

想一想　为什么推论 8.14 对 $4p^e$ 失效？例如模 $20=4\cdot 5$ 有没有原根？

答　没有。$\varphi(20)=\varphi(4)\varphi(5)=8$，但对 $(a,20)=1$，a 模 4 的阶整除 2，模 5 的阶整除 4，所以 $a^{\operatorname{lcm}(2,4)}=a^4\equiv 1 \pmod{20}$，阶不超过 $4<8$。$2p^e$ 能成功，是因为模 2 那部分 $\varphi(2)=1$，不占任何位置；模 4 那部分 $\varphi(4)=2$ 是偶数，和 $\varphi(p^e)$ 的偶数「撞车」了。下一节把这个想法一般化。

8.6　没有原根的模

前两节证明了 p^e 和 $2p^e$ 有原根。主定理还说：**其余的 m 都没有原根**。先看数据。对 $m\le 40$，逐个求出简化剩余的最大阶，与 $\varphi(m)$ 比较，结果如图 8.5。

1	2	3	4	5	6	7	**8**	9	10
11	**12**	13	14	**15**	**16**	17	18	19	**20**
21	22	23	**24**	25	26	27	**28**	29	**30**
31	**32**	**33**	34	**35**	**36**	37	38	**39**	**40**

蓝色：有原根。
橙色：没有原根，最大阶是 $\varphi(m)$ 的真因数，例如 $m=15$：4 对 8；$m=24$：2 对 8；$m=35$：12 对 24。

图 8.5　$1\le m\le 40$ 中哪些模有原根。橙色格子要么是 2^e（$e\ge 3$），要么能拆成两个互素且都大于 2 的因数之积。

橙色的数分两类。一类是 $8,16,32$，即 2^e（$e\ge 3$）。另一类是 $12=3\cdot 4$，$15=3\cdot 5$，$20=4\cdot 5$，$21=3\cdot 7$，$24=3\cdot 8$，$28=4\cdot 7$，$30=5\cdot 6$，$33=3\cdot 11$，$35=5\cdot 7$，$36=4\cdot 9$，$39=3\cdot 13$，$40=5\cdot 8$：每个都能写成**两个互素且都大于 2** 的数之积。而 $10=2\cdot 5$，$18=2\cdot 9$ 这类只能拆出因数 2，它们是蓝色的。下面分别处理这两类。

8.6.1 两个互素的因数都大于 2

■ **例 8.15**　模 15　求模 15 每个简化剩余的阶，说明模 15 没有原根。

解　$\varphi(15)=8$，简化剩余是 $1,2,4,7,8,11,13,14$。逐个求阶：

a	1	2	4	7	8	11	13	14
$\operatorname{ord}_{15} a$	1	4	2	4	4	2	4	2

例如 $2^2=4$，$2^4=16\equiv 1$；$11^2=121=8\cdot 15+1$。最大的阶是 $4<8$，所以没有原根。

原因可以从中国剩余定理看出。a 与 15 互素时，模 3 有 $a^2\equiv 1$（$\varphi(3)=2$），模 5 有 $a^4\equiv 1$（$\varphi(5)=4$）。4 同时是 2 和 4 的倍数，所以 $a^4\equiv 1$ 对模 3 和模 5 都成立，从而 $a^4\equiv 1 \pmod{15}$。两个分量的「周期」都是偶数，它们的最小公倍数 4 比乘积 8 小。 ♦

例 8.15 的论证只用到一件事：两个分量的 φ 都是偶数。

> **引理 8.15**　当 $n\geq 3$ 时，$\varphi(n)$ 是偶数。

证明　把与 n 互素的 $a\in\{1,\dots,n-1\}$ 和 $n-a$ 配成一对；$(n-a,n)=(a,n)=1$，所以 $n-a$ 也在其中。若 $a=n-a$，则 $n=2a$，$a=(a,2a)=(a,n)=1$，$n=2$，与 $n\geq 3$ 矛盾。所以每一对都由两个不同的数组成，$\varphi(n)$ 是偶数。 ■

> **命题 8.16**　设 $n=mm'$，$(m,m')=1$，$m>2$，$m'>2$。则对每个与 n 互素的 a，
>
> $$a^{\varphi(n)/2}\equiv 1 \pmod{n},$$
>
> 所以模 n 没有原根。

思路　模 m 和模 m' 分别有 Euler 定理。找一个同时是 $\varphi(m)$ 与 $\varphi(m')$ 倍数、却比 $\varphi(n)=\varphi(m)\varphi(m')$ 小的指数。两个数都是偶数，所以乘积的一半就行。

证明　φ 是积性函数，$\varphi(n)=\varphi(m)\varphi(m')$。由引理 8.15，$\varphi(m)$、$\varphi(m')$ 都是偶数，所以

$$N=\frac{\varphi(n)}{2}=\varphi(m)\cdot\frac{\varphi(m')}{2}=\frac{\varphi(m)}{2}\cdot\varphi(m')$$

既是 $\varphi(m)$ 的整数倍，也是 $\varphi(m')$ 的整数倍。a 与 m、m' 都互素，由 Euler 定理，

$$a^N=\left(a^{\varphi(m)}\right)^{\varphi(m')/2}\equiv 1 \pmod{m},\qquad a^N=\left(a^{\varphi(m')}\right)^{\varphi(m)/2}\equiv 1 \pmod{m'}.$$

m,m' 互素，所以 $a^N\equiv 1 \pmod{mm'}$。于是 $\operatorname{ord}_n a\leq N<\varphi(n)$，没有 a 的阶等于 $\varphi(n)$。 ■

关键一步是「两个偶数的最小公倍数至多是乘积的一半」。条件 $m,m'>2$ 不能去掉：$10=2\cdot 5$ 中 $\varphi(2)=1$ 是奇数，$N=\varphi(10)/2=2$ 不是 $\varphi(5)=4$ 的倍数，论证失效，而 10 确实有原根 3。

注 8.1　[回到第 7 讲的问题] 第 7 讲问过：Euler 定理说 $(n,35)=1$ 时 $n^{24}\equiv 1 \pmod{35}$，24 能否换成更小的数？同样的推理给出 $n^{\operatorname{lcm}(\varphi(5),\varphi(7))}=n^{\operatorname{lcm}(4,6)}=n^{12}\equiv 1 \pmod{35}$。12 不能再小：$\operatorname{ord}_{35}2=12$（2 模 5 的阶是 4，模 7 的阶是 3）。使 $a^N\equiv 1 \pmod n$ 对所有 $(a,n)=1$ 成立的最小 N 叫做 Carmichael 函数 $\lambda(n)$。模 n 有原根当且仅当 $\lambda(n)=\varphi(n)$。

8.6.2　2 的方幂

模 8 的简化剩余是 $1,3,5,7$，它们的平方都是 1：

$$3^2=9=8+1,\qquad 5^2=25=3\cdot 8+1,\qquad 7^2=49=6\cdot 8+1.$$

最大阶是 2，而 $\varphi(8)=4$。原因是任何奇数 $a=2k+1$ 满足 $a^2-1=4k(k+1)$，而 $k(k+1)$ 是偶数，所以 $8\mid a^2-1$。这个现象会一直传上去。

命题 8.17　设 $e\ge 3$，a 是奇数。则 $a^{2^{e-2}}\equiv 1 \pmod{2^e}$。因此模 2^e（$e\ge 3$）没有原根。

证明　对 e 归纳。$e=3$ 就是上面的 $a^2\equiv 1 \pmod 8$。设 $a^{2^{e-2}}=1+t\cdot 2^e$，t 是整数。两边平方：

$$a^{2^{e-1}}=\left(1+t\cdot 2^e\right)^2=1+t\cdot 2^{e+1}+t^2\cdot 2^{2e}\equiv 1 \pmod{2^{e+1}},$$

因为 $2e\ge e+1$。这就是 $e+1$ 时的结论。于是每个奇数模 2^e 的阶都整除 2^{e-2}，小于 $\varphi(2^e)=2^{e-1}$。∎

为什么要 $e\ge 3$　命题对 $e=2$ 不成立：$3^{2^0}=3\not\equiv 1 \pmod 4$。模 4 的简化剩余只有 $1,3$，3 的阶是 $2=\varphi(4)$，所以 3 是模 4 的原根。归纳的起点 $a^2\equiv 1 \pmod 8$ 用到了 $k(k+1)$ 是偶数带来的额外一个因子 2，模 4 时还没有这个因子可用。这就是清单里出现 4 却没有 8 的原因。

模 2^e 虽然没有原根，但离「有原根」只差一个符号。最大阶 2^{e-2} 是取得到的，而且由 5 取到。

定理 8.18（模 2^e 的结构）　设 $e\ge 3$。则 $\operatorname{ord}_{2^e}5=2^{e-2}$，并且

$$\pm 5^k,\qquad 0\le k<2^{e-2}$$

恰好构成模 2^e 的一个简化剩余系。

证明　先证：对 $j\ge 0$，$5^{2^j}=1+2^{j+2}u_j$，u_j 是奇数。$j=0$ 时 $5=1+4\cdot 1$。若结论对 j 成立，则

$$5^{2^{j+1}}=\left(1+2^{j+2}u_j\right)^2=1+2^{j+3}u_j+2^{2j+4}u_j^2=1+2^{j+3}\left(u_j+2^{j+1}u_j^2\right),$$

括号里是奇数加偶数，仍是奇数。取 $j=e-3$：$5^{2^{e-3}}=1+2^{e-1}u$，u 为奇数，所以 $5^{2^{e-3}}\not\equiv 1 \pmod{2^e}$。由命题 8.17，$\operatorname{ord}_{2^e}5$ 整除 2^{e-2}；它又不整除 2^{e-3}，所以等于 2^{e-2}。

于是 $5^0,5^1,\dots,5^{2^{e-2}-1}$ 模 2^e 两两不同余（与命题 8.2 的证明相同）。$5\equiv 1 \pmod 4$，所以它们都 $\equiv 1 \pmod 4$，而 $-5^k\equiv -1\equiv 3 \pmod 4$。两组没有公共元素，一共 $2\cdot 2^{e-2}=2^{e-1}=\varphi(2^e)$ 个两两不同余的奇数，恰好是全部简化剩余。∎

以 $e=5$ 为例，$\mathrm{ord}_{32}(5)=8$，模 32 的 16 个奇剩余排成两个环（图 8.6）。和图 8.1 的单个钟面相比，这里是两个钟面：5^k 负责在环内转圈，-1 负责在两个环之间跳。

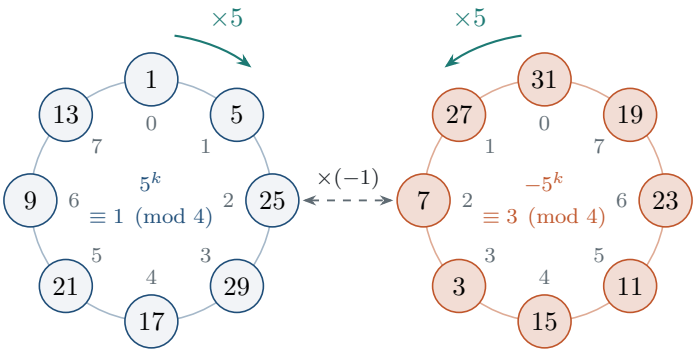


图 8.6　模 32 的奇剩余分成两个环，环内的小数字是指数 k。乘 5 在环内转一格，乘 -1 在两环之间跳（例如 $25 \leftrightarrow 7$）。每个元素的阶至多是 8，没有一个数能走遍全部 16 个剩余。

■ **例 8.16　模 32 的平方根**　用定理 8.18 解 $x^2 \equiv 17 \pmod{32}$，并说明模 32 的奇数中哪些是平方数。

解　查图 8.6：$17=5^4$。设 $x \equiv \varepsilon 5^k$，$\varepsilon=\pm1$，$0 \le k < 8$。因为 $\varepsilon^2=1$，

$$x^2 \equiv 5^{2k} \equiv 5^4 \pmod{32} \iff 2k \equiv 4 \pmod 8 \iff k \equiv 2 \pmod 4 \iff k \in \{2,6\}.$$

符号 ε 随便取，所以有 $2\times 2=4$ 个解：$x \equiv \pm 5^2, \pm 5^6 \equiv \pm 25, \pm 9$，即

$$x \equiv 7,\ 9,\ 23,\ 25 \pmod{32}.$$

验证：$9^2=81=2\cdot 32+17$，$7^2=49=32+17$，$23^2=529=16\cdot 32+17$，$25^2=625=19\cdot 32+17$。

同样的推理说明：奇数 $a \equiv \varepsilon 5^j$ 是平方数当且仅当 $\varepsilon=1$ 且 j 是偶数，即 $a \in \{5^0, 5^2, 5^4, 5^6\} = \{1, 25, 17, 9\}$，也就是 $a \equiv 1 \pmod 8$。每个这样的 a 恰有 4 个平方根。对比模奇素数 p：那里的平方剩余恰有 2 个平方根 $\pm b$。 ♦

8.6.3　原根存在定理的完整证明

现在把各节的结论拼起来。

定理 8.9 的证明　*充分性*。$m=1,2$ 时只有一个简化剩余，1 就是原根；$m=4$ 时 3 是原根。奇素数 p 的情形是第 7 讲的定理；p^2 由定理 8.11，p^e（$e \ge 2$）由定理 8.13，$2p^e$ 由推论 8.14。

必要性。设 m 不在清单中，写 $m=2^a u$，u 为奇数。

- 若 $u=1$，则 $m=2^a$ 且 $a \ge 3$（$a \le 2$ 在清单里）。由命题 8.17，没有原根。
- 若 $u>1$ 且 u 至少有两个不同的素因数，取其中一个 p，设 p^b 是 m 中 p 的最高次幂，令 $m_1=p^b$，$m_2=m/p^b$。m_2 含有另一个奇素数，所以 $m_1, m_2 \ge 3$，且互素。
- 若 $u=p^b$ 是奇素数的方幂，则 $a \ge 2$（$a=0,1$ 在清单里）。令 $m_1=2^a \ge 4$，$m_2=p^b \ge 3$。

后两种情形都把 m 拆成了两个互素且都大于 2 的因数，由命题 8.16，没有原根。 ■

表 8.2 将存在性结论与排除情形合并，求原根前可先按模数类型判断。

m	原根？	理由	例子
$1, 2, 4$	有	直接验证	4：原根 3
p^e（p 奇素数）	有	第 7 讲；定理 8.11、8.13	25：原根 2
$2p^e$	有	推论 8.14	50：原根 27
2^e，$e \geq 3$	无	命题 8.17：$a^{2^{e-2}} \equiv 1$	8：最大阶 2
其余	无	命题 8.16：$a^{\varphi(m)/2} \equiv 1$	$12 = 4 \cdot 3$，$15 = 3 \cdot 5$

表 8.2　模 m 的原根：完整的分类。

循环群　用群的语言说，模 m 的简化剩余在乘法下组成一个阶为 $\varphi(m)$ 的有限交换群 $(\mathbb{Z}/m\mathbb{Z})^\times$，原根就是它的生成元。所以定理 8.9 回答的是「这个群什么时候是循环群」。定理 8.18 说 $(\mathbb{Z}/2^e\mathbb{Z})^\times \cong \{\pm 1\} \times \langle 5 \rangle$，是两个循环群的直积，阶分别为 2 和 2^{e-2}。命题 8.16 则说两个偶数阶循环群的直积不是循环群。

本讲小结

- 若 g 是模 m 的原根，则 $g^0, g^1, \ldots, g^{\varphi(m)-1}$ 恰好走遍简化剩余。$\operatorname{ord}_m(g^k) = \varphi(m)/(k, \varphi(m))$，所以原根恰有 $\varphi(\varphi(m))$ 个，即 g^k，$(k, \varphi(m)) = 1$。
- 判别法：g 是原根 $\iff$ 对 $\varphi(m)$ 的每个素因数 q，$g^{\varphi(m)/q} \not\equiv 1$。
- 指标 ind_g 把模 p 的乘法变成模 $p-1$ 的加法：$\operatorname{ind}(ab) \equiv \operatorname{ind} a + \operatorname{ind} b$，$\operatorname{ind}(a^k) \equiv k \operatorname{ind} a \pmod{p-1}$。
- 记 $\delta = (d, p-1)$。$x^d \equiv a \pmod p$ 有解 $\iff \delta \mid \operatorname{ind} a \iff a^{(p-1)/\delta} \equiv 1$，有解时恰有 δ 个解。
- 模 p 的原根 g 的 p 个提升中，只有 $g^p \bmod p^2$ 不是模 p^2 的原根。模 p^2 的原根是所有模 p^e 的原根，其中的奇数是模 $2p^e$ 的原根。
- 模 m 有原根 $\iff m = 1, 2, 4, p^e, 2p^e$。其余的 m 要么有两个都大于 2 的互素因数，此时 $a^{\varphi(m)/2} \equiv 1$；要么是 2^e（$e \geq 3$），此时 $a^{2^{e-2}} \equiv 1$，简化剩余是 $\pm 5^k$。

练习

8.1 ★　54，98，100，162，250，1000 中哪些模有原根？有的话一共几个？并写出模 98 的一个原根。

8.2 ★　证明 $3^8 \equiv -1 \pmod{17}$，由此说明 3 是模 17 的原根，再写出模 17 的全部原根。它们和模 17 的平方非剩余有什么关系？

8.3 ★　用表 8.1 解下列同余式：(a) $6^x \equiv 5 \pmod{13}$；(b) $x^3 \equiv 5 \pmod{13}$；(c) $x^3 \equiv 2 \pmod{13}$。

8.4 ★★　求模 121 的一个原根，并说明它对所有 11^e 都是原根。再求模 242 的一个原根。

8.5 ★★ 设 p 是奇素数，g 是模 p 的原根。证明：$-g$ 是模 p 的原根当且仅当 $p \equiv 1 \pmod 4$。

8.6 ★★ 设 $p \geq 5$ 是素数。证明模 p 的全部原根之积 $\equiv 1 \pmod p$。$p = 3$ 时结果是什么？

8.7 ★★ 证明：$x^3 \equiv 1 \pmod{1024}$ 只有解 $x \equiv 1$。再解 $x^3 \equiv -3 \pmod{1024}$。（提示：$\varphi(1024) = 512$ 与 3 互素。）

8.8 ★★★ 证明：对每个素数 p，$x^8 \equiv 16 \pmod p$ 都有解。（提示：用定理 8.7 化成关于 $\operatorname{ind} 2$ 的条件；$p \equiv 1 \pmod 8$ 时，取阶为 8 的 ζ，计算 $(\zeta + \zeta^{-1})^2$。）

8.9 ★★★ 求所有正整数 m，使得对每个与 m 互素的 a 都有 $a^2 \equiv 1 \pmod m$。由此说明：对每个素数 $p \geq 5$，$24 \mid p^2 - 1$。

8.10 ★ （习题集 4 第 2 部分第 1(a)）先求模 $23^2 = 529$ 的一个原根，再求原题要求的模 $23^3 = 12167$ 的一个原根，并写出后者的阶。

练习答案与提示

8.1 $54 = 2 \cdot 3^3$，$98 = 2 \cdot 7^2$，$162 = 2 \cdot 3^4$，$250 = 2 \cdot 5^3$ 都是 $2p^e$ 型，有原根。个数为 $\varphi(\varphi(m))$：$\varphi(18) = 6$，$\varphi(42) = 12$，$\varphi(54) = 18$，$\varphi(100) = 40$。$100 = 4 \cdot 25$ 和 $1000 = 8 \cdot 125$ 都能拆成两个大于 2 的互素因数，没有原根（最大阶分别是 20 和 100）。模 98：$3^2 \equiv 2$，$3^3 \equiv 6 \equiv -1 \pmod 7$，所以 3 是模 7 的原根；$3^6 = 729 = 14 \cdot 49 + 43 \not\equiv 1 \pmod{49}$，由引理 8.10，3 是模 49 的原根；3 是奇数，由推论 8.14，它是模 98 的原根。

8.2 $3^2 = 9$，$3^4 = 81 \equiv 13 \equiv -4$，$3^8 \equiv 16 \equiv -1 \pmod{17}$。$\varphi(17) = 16$ 的素因数只有 2，而 $3^8 \not\equiv 1$，由命题 8.4，3 是原根。全部原根是 3^k，k 为奇数：$3, 10, 5, 11, 14, 7, 12, 6$，即 $\{3, 5, 6, 7, 10, 11, 12, 14\}$。模 17 的平方剩余是 $\{1, 2, 4, 8, 9, 13, 15, 16\}$，所以原根恰好是全部平方非剩余。原因是 16 是 2 的方幂，「$(k, 16) = 1$」等价于「k 为奇数」，而奇数指标的元素正是平方非剩余。

8.3 查表：$\operatorname{ind} 6 = 5$，$\operatorname{ind} 5 = 9$，$\operatorname{ind} 2 = 1$。(a) $5x \equiv 9 \pmod{12}$，$5^{-1} \equiv 5$，$x \equiv 45 \equiv 9 \pmod{12}$。验证 $6^9 \equiv 5$。(b) $3 \operatorname{ind} x \equiv 9 \pmod{12}$，$\operatorname{ind} x \equiv 3 \pmod 4$，$\operatorname{ind} x \in \{3, 7, 11\}$，$x \in \{8, 11, 7\}$。验证 $8^3 = 512 = 39 \cdot 13 + 5$。(c) $3 \operatorname{ind} x \equiv 1 \pmod{12}$ 无解，因为 $(3, 12) = 3 \nmid 1$。也可以用定理 8.7(2) 的判别法：$2^{12/3} = 16 \equiv 3 \not\equiv 1$。

8.4 $2^5 = 32 \equiv -1 \pmod{11}$，所以 $\operatorname{ord}_{11}(2) = 10$，2 是模 11 的原根。$2^{10} = 1024 = 8 \cdot 121 + 56 \not\equiv 1 \pmod{121}$，由引理 8.10，2 是模 121 的原根；由定理 8.13，它是所有模 11^e 的原根。2 是偶数，与 242 不互素，由推论 8.14 取 $2 + 121 = 123$。

8.5 由第 7 讲，$x^2 \equiv 1 \pmod p$ 只有解 ± 1。$g^{(p-1)/2}$ 的平方是 1 而它本身不是 1，所以 $g^{(p-1)/2} \equiv -1$，$-g \equiv g^{1+(p-1)/2}$。由定理 8.3，$-g$ 是原根 $\iff \left(1 + \frac{p-1}{2}, p - 1\right) = 1$。若 $\frac{p-1}{2}$ 是奇数，$1 + \frac{p-1}{2}$ 与 $p - 1$ 都是偶数，不互素。若 $\frac{p-1}{2}$ 是偶数，设 d 是公因数，则 d 整除 $2\left(1 + \frac{p-1}{2}\right) - (p - 1) = 2$，而 $1 + \frac{p-1}{2}$ 是奇数，所以 $d = 1$。故 $-g$ 是原根 $\iff$ $\frac{p-1}{2}$ 为偶数 $\iff p \equiv 1 \pmod 4$。例：$p = 5$，$-2 \equiv 3$ 是原根；$p = 7$，$-3 \equiv 4$ 的阶是 3。

8.6 若 g 是原根，则 g^{-1} 也是（它们的阶相同）。把原根按 $\{g, g^{-1}\}$ 配对，每对之积为 1。只有 $g \equiv g^{-1}$，即 $g^2 \equiv 1$、$g \equiv \pm 1$ 时配不成对；但 ± 1 的阶至多为 $2 < p - 1$，不是原根。所以乘积 $\equiv 1$。$p = 3$ 时唯一的原根是 $2 \equiv -1$，乘积为 $2 \equiv -1$。

8.7 解 x 必为奇数。$(3, 512) = 1$，且 $3 \cdot 171 = 513 \equiv 1 \pmod{512}$。对奇数 x，$x^3 \equiv a \Rightarrow x \equiv x^{513} = (x^3)^{171} \equiv a^{171}$（Euler 定理）；反之 $(a^{171})^3 = a^{513} \equiv a$。所以解唯一，$x \equiv a^{171}$。$a = 1$ 给出 $x \equiv 1$。$a = -3$ 给出 $x \equiv -3^{171}$。反复平方：$3^8 \equiv 417$，$3^{32} \equiv 641$，$3^{128} \equiv 513$；$171 = 128 + 32 + 8 + 2 + 1$，依次相乘得 $513 \cdot 641 \equiv 129$，$129 \cdot 417 \equiv 545$，$545 \cdot 9 \equiv 809$，$809 \cdot 3 \equiv 379$。于是 $x \equiv -379 \equiv 645$。

验证：$645^2 \equiv 281$，$281 \cdot 645 \equiv 1021 \equiv -3 \pmod{1024}$。

8.8 $p = 2$ 时 $x = 0$。设 p 为奇素数，g 为原根，$\operatorname{ind} 16 = 4 \operatorname{ind} 2$。由定理 8.7，有解 $\iff \delta = (8, p-1)$ 整除 $4 \operatorname{ind} 2$。若 $8 \nmid p-1$，则 $\delta \mid 4$，自动成立。若 $8 \mid p-1$，需要 $\operatorname{ind} 2$ 为偶数，即 2 是平方数。令 $\zeta = g^{(p-1)/8}$，它的阶是 8，$\zeta^4 \equiv -1$，于是 $\zeta^{-2} \equiv -\zeta^2$，

$$(\zeta + \zeta^{-1})^2 = \zeta^2 + 2 + \zeta^{-2} \equiv \zeta^2 + 2 - \zeta^2 = 2 \pmod{p}.$$

所以 2 是平方数。例：$p = 17$，$\zeta = 3^2 = 9$，$9^{-1} \equiv 2$，$11^2 = 121 \equiv 2$。

8.9 由中国剩余定理，条件对 m 成立当且仅当对 m 的每个素数方幂因子 p^b 成立。若 p 是奇素数，p^b 有原根 g，$g^2 \equiv 1$ 要求 $\varphi(p^b) = p^{b-1}(p-1) \mid 2$，只有 $p^b = 3$。对 2^a：$a \le 3$ 时成立（奇数平方 $\equiv 1 \pmod 8$）；$a \ge 4$ 时 $3^2 = 9 \not\equiv 1 \pmod{16}$，不成立。所以 $m \mid 24$，即 $m \in \{1, 2, 3, 4, 6, 8, 12, 24\}$，这些都满足条件。素数 $p \ge 5$ 与 24 互素，取 $m = 24$，$a = p$，得 $p^2 \equiv 1 \pmod{24}$。

8.10 由例 8.3，5 是模 23 的原根。例 8.11 算得 $5^{22} \equiv 323 \not\equiv 1 \pmod{529}$，由引理 8.10，5 是模 529 的原根。再由定理 8.13，23 是奇素数，故 5 也是模 12167 的原根，阶为 $23^2 \cdot 22 = 11638$。

复习课 1

整除、同余与原根的综合复习

Review of Divisibility, Congruences, and Primitive Roots

对应材料：MIT 18.781 第 1–8 讲；Guidelines for Midterm 1；Practice problems for Midterm 1 第 1–10、13–16 题及解答（第 11、12 题融入定义表与 CRT 回顾）；Problem Set 4 第一部分第 6 题。

本讲目标

(1) 能准确说出第 1–8 讲的主要定义与定理条件，并按依赖关系选择适当方法。

(2) 会检查约分、降幂、中国剩余定理、Hensel 提升和原根判别的条件。

(3) 能完成带完整性说明的计算：全部解、全部根、元素的准确阶，而不仅是找到一个候选。

(4) 能辨认原练习题与答案中的条件遗漏、抄写错误，并用回代或反例核对。

R1.1 知识地图与定义检查

给出 $ax+by=c$、$a^k \bmod m$、$f(x)\equiv 0 \pmod{p^e}$ 三类问题，第一步应当不同：整数方程先看最大公因数，幂余数先看底数与模数是否互素，素数幂同余先从模 p 的根开始。第 1–8 讲积累的结论可以按这几种入口串联。图 R1.1 中实线表示直接使用前一个结论，橙色虚线表示另一条主题中的工具也参与计算。

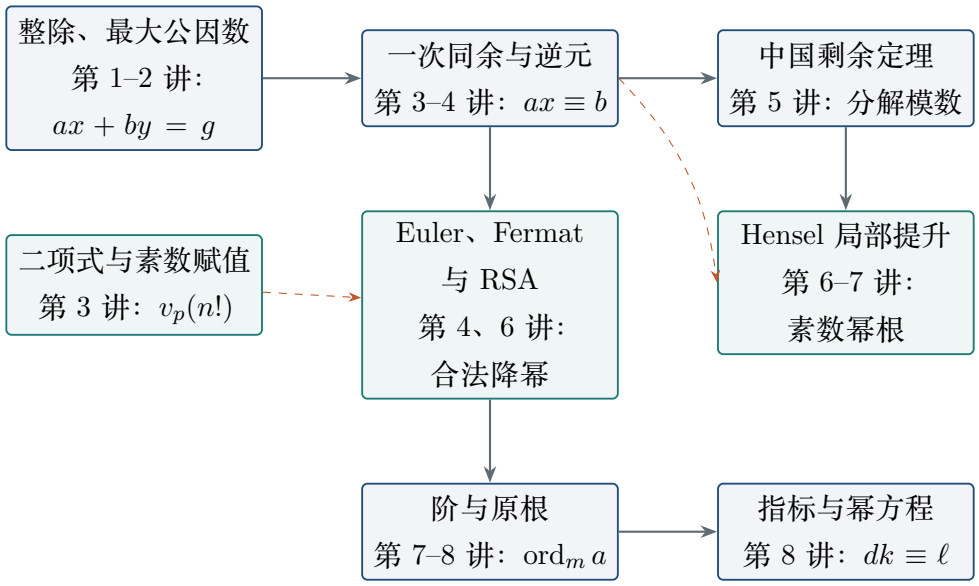


图 R1.1　整除支持逆元与一次同余，分解模数支持局部求根；两条橙色虚线分别提示二项式支持 Fermat 证明、逆元支持非奇异根的提升。

读图时要区分两个「模」：数值运算通常在模 m 下进行；若底数与模数互素，指数可按该底数的阶约化，也可按它的某个已知倍数约化。指标题中，这个区别尤其明确：原方程模 p，指数方程模 $p-1$。原课程期中指引强调定义、标准定理及其用法；只会背公式，漏掉取值范围，计算即使短也可能整题失效。

遮住表 R1.1 的中间一列，检查能否说出定义中的每个条件。最右列提供了检验定义是否真正会用的小任务。

概念	定义要点	典型用法
整除、最大公因数与互素	$a \mid b$ 表示 $b = ak$、$k \in \mathbb{Z}$、$a \neq 0$；$a, b \in \mathbb{Z}$ 不全为 0 时，$\gcd(a, b)$ 是正最大公因数，互素指它为 1。	把公因数变成整系数组合的因数。
素数、合数与唯一分解	对 $n > 1$，素数只有正因数 $1, n$，合数则有其他正因数；素因数分解除次序外唯一。	确定试除范围，逐素数比较指数。
同余与剩余系	$a \equiv b \pmod m$ 表示 $m \mid a - b$；完全剩余系对 m 个类各取一个代表，简化剩余系对每个与 m 互素的类各取一个，共 $\varphi(m)$ 个。	同余与整系数多项式代入相容；模 12 的简化剩余系可取 $\{1, 5, 7, 11\}$。
二项式系数	$n, k \in \mathbb{Z}$ 且 $0 \le k \le n$ 时，$\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$；它计数 n 元集合的 k 元子集。	使用二项式定理和素数赋值判断整除性。
阶、原根与指标	$m \ge 1$ 且 $\gcd(a, m) = 1$ 时，$\operatorname{ord}_m(a)$ 是使 $a^h \equiv 1 \pmod m$ 的最小正整数；原根还要求阶为 $\varphi(m)$。若 p 为素数、g 为模 p 的原根且 $p \nmid a$，指标是唯一的 $k \bmod (p-1)$ 使 $g^k \equiv a \pmod p$。	把模 p 的幂方程化为模 $p-1$ 的一次同余。

表 R1.1　定义中的非零、正数、互素与最小性，决定后续公式何时可以使用。

R1.2　核心结论回顾

R1.2.1　整除、算法与唯一分解（第 1–2 讲）

> **定理（Bézout 与整数方程）**　设整数 a, b 不全为零，$g = \gcd(a, b) > 0$。存在整数 u, v 使 $au + bv = g$；$ax + by = c$ 有整数解当且仅当 $g \mid c$。若 (x_0, y_0) 是一组解，则全部解为 $x = x_0 + (b/g)t$、$y = y_0 - (a/g)t$，$t \in \mathbb{Z}$。

关键　第 1 讲用最小正整系数组合得到 g；第 2 讲用带余除法保持公因数集合，欧几里得算法算出 g，回代算出系数。通解由 $a(x - x_0) = -b(y - y_0)$ 再约去最大公因数得到。完整证明见第 1、2 讲。

用法　先求 g，再判断可解性；需要「所有解」时还要写步长。代入通解只能证明它们都是解，需补上从任意解倒推出整数参数的理由。求逆元则是取 $c = 1$，其存在条件也来自同一个定理。

定理（唯一分解与素数赋值）　正整数的素因数分解除次序外唯一。若 $a=\prod p^{\alpha_p}$、$b=\prod p^{\beta_p}$，则最大公因数取指数的最小值，最小公倍数取最大值；特别地 $\gcd(a,b)\operatorname{lcm}(a,b)=ab$。

关键　素数整除乘积时必整除一个因子，逐个消去得到唯一性。欧几里得证明素数无穷时，取有限列表 $p_1,\dots,p_r$，考察 $p_1\cdots p_r+1$。完整证明见第 2 讲；Euclid 引理的一般形式在第 1 讲已由 Bézout 等式推出。用良序原理取最小反例时，须先确认集合非空且有下界。

用法　分解已经给出时，逐素数比较指数最直接；对较大且未分解的整数求最大公因数，应先用欧几里得算法。素数试除只需试到平方根，但「能终止」与「对于输入位数高效」是不同要求，第 5、6 讲专门比较过它们。

R1.2.2　二项式、赋值与同余（第 3 讲）

定理（二项式与阶乘中的素数指数）　对 $n\ge 0$，$(1+x)^n=\sum_{k=0}^{n}\binom{n}{k}x^k$。对素数 p，$v_p(n!)=\sum_{j\ge 1}\lfloor n/p^j\rfloor$，且 $v_p\binom{n}{k}=v_p(n!)-v_p(k!)-v_p((n-k)!)$。

关键　二项式系数计数选择集合的方法；阶乘赋值逐层数 $p,p^2,\dots$ 的倍数。完整证明见第 3 讲。Kummer 将同一个指数解释成 p 进制的进位数，Lucas 则把模 p 的二项式系数拆成逐位乘积。

用法　加权和 $\sum k\binom{n}{k}$ 可对二项式恒等式求导，或把权重解释成「再指定一个被选中的元素」。判断整除性先比较 v_p；判断二项式系数的余数则考虑 Lucas，注意它要求底数 p 是素数。

同余的基础是 $m\mid a-b$。它与加、减、乘和代入整系数多项式相容；消去同一个因子 c 时，准确的等价式为 $ac\equiv bc\pmod m\iff a\equiv b\pmod{m/\gcd(c,m)}$。所以「消去」后模数可能缩小。后面的每一种解法都建立在这条整除关系上，而非实数除法的习惯上。

R1.2.3　降幂、一次同余与分解模数（第 4–5 讲）

定理（Euler、Fermat 与 Wilson）　若 $\gcd(a,m)=1$，则 $a^{\varphi(m)}\equiv 1\pmod m$。对素数 p，有 $a^p\equiv a\pmod p$；$p\nmid a$ 时可写成 $a^{p-1}\equiv 1\pmod p$。另有 $(p-1)!\equiv -1\pmod p$，且对 $n\ge 2$，$(n-1)!\equiv -1\pmod n$ 也能反推 n 为素数。

关键　Euler 的证明用乘以互素数来重排简化剩余系，再合法地消去所有单位；Wilson 把非零剩余中的逆元配对，留下 ±1。完整证明见第 4 讲。两种配对都要求相关逆元存在。

用法　求大幂先检查互素，再决定降指数还是反复平方。Fermat 检验失败能判合数，通过却不能判素数；Wilson 是充要判别，但计算阶乘通常昂贵。若指数靠近 $p-1$，可把问题改为求一个小数的逆元。

Wilson 定理还推出 $x^2\equiv -1\pmod p$ 有解当且仅当 $p=2$ 或 $p\equiv 1\pmod 4$，并由此证明有无穷多个 $4k+1$ 型素数（证明见第 4 讲）。

定理（一次同余与中国剩余定理）　设 $a,b\in\mathbb{Z}$、$m\geq 1$ 且 $g=\gcd(a,m)$。$ax\equiv b\pmod m$ 有解当且仅当 $g\mid b$，有解时模 m 恰有 g 个解。若 m_i 是两两互素的正整数，则 $x\equiv a_i\pmod{m_i}$ 有且仅有一个模 $M=\prod m_i$ 的解类。对正整数模数 m_1,m_2，有解当且仅当 $\gcd(m_1,m_2)\mid a_1-a_2$，解模最小公倍数唯一。

关键　一次同余式等价于 $ax-my=b$，所以直接使用整数方程的可解性。CRT 构造在一个分量为 1、其他分量为 0 的整数，再相加；唯一性来自互素模数的乘积整除解之差。完整证明见第 4、5 讲。

用法　先把模数拆成互素的素数幂，再在各分量求解。每个局部解组合都对应一个整体解，因此局部解数相乘；这个计数能检查「求所有解」有没有漏项。第 5 讲也由 CRT 得到 φ 的积性及 $\varphi(n)=n\prod_{p\mid n}(1-1/p)$。

R1.2.4　素数幂、素性检验与 RSA（第 6 讲）

定理（非奇异根的 Hensel 提升）　设 $f\in\mathbb{Z}[x]$、$a\in\mathbb{Z}$，p 为素数，$j\geq 1$。若 $f(a)\equiv 0\pmod{p^j}$ 且 $p\nmid f'(a)$，则 $a+tp^j$（$0\leq t<p$）中恰有一个是模 p^{j+1} 的根。

关键　展开到一次项：$f(a+tp^j)\equiv f(a)+tp^jf'(a)\pmod{p^{j+1}}$。除去 p^j 后是关于 t 的模 p 一次同余，导数条件保证其系数可逆。若 $p\mid f'(a)$，同层 p 个候选全为根当且仅当 $p^{j+1}\mid f(a)$；否则全不为根。见第 7 讲。

用法　根要先在模 p 下找全，再对每个根检查导数。提升只负责同一素数的幂次；不同素数幂之间的合并仍用 CRT。根数保持不变的前提是每个起始根均非奇异。

Fermat 检验有合数通过；Miller–Rabin 检查 $n-1=2^sd$ 的平方链，素数必通过。Rabin–Monier 定理：固定奇合数 $n>9$ 时，强说谎底数至多有 $\varphi(n)/4$ 个；在 $\{2,\dots,n-2\}$ 独立均匀抽底 k 次，误判概率至多 4^{-k}。Pollard ρ 用 $\gcd(x_{2k}-x_k,n)$ 找因子，约 $\sqrt{p}$ 步是启发式估计。

RSA 取不同素数 p,q，令 $N=pq$、$ef\equiv 1\pmod{\varphi(N)}$；对与 N 互素的消息，$(m^e)^f\equiv m\pmod N$，私钥模数是 $\varphi(N)$。若已知 $N,\varphi(N)$，由 $p+q=N-\varphi(N)+1$ 解二次方程即可恢复 p,q。

R1.2.5　根数、阶、原根与指标（第 7–8 讲）

定理（素数模根数与元素的阶）　设 $f(x)=a_nx^n+\cdots+a_0\in\mathbb{Z}[x]$、$p$ 为素数且 $p\nmid a_n$，则 $f(x)\equiv 0\pmod p$ 至多有 n 个根。若 $m\geq 1$、$\gcd(a,m)=1$、$h=\operatorname{ord}_m(a)$、$k\in\mathbb{Z}_{\geq 0}$，则 $a^k\equiv 1\pmod m\iff h\mid k$；对 $k\geq 1$，$\operatorname{ord}_m(a^k)=h/\gcd(h,k)$。

关键　根数上界对每个已知根提出一个一次因子，利用素数模下的消去律归纳。阶的整除性质对 k 除以最小正指数 h，排除非零余数。完整证明见第 7 讲。两元素阶互素时，乘积的阶为两阶乘积；阶不互素时不能照用。

用法　根数上界可在找到足够多根后证明完整性；合数模下须改用局部根数与 CRT。求准确阶时，Euler 只提供一个倍数，要继续排除较小约数。模 p^j 的奇异根有 0 或 p 个提升，也是第 7 讲保留的另一条完整性规则。

> **定理（原根判别与幂同余）**　若 $m \geq 1$、$\gcd(g, m) = 1$，则 g 为原根当且仅当对 $\varphi(m)$ 的每个不同素因子 q 均有 $g^{\varphi(m)/q} \not\equiv 1 \pmod{m}$。若 p 为素数、g 是原根、$p \nmid a$、$d \geq 1$，则 $x^d \equiv a \pmod{p}$ 等价于 $dk \equiv \operatorname{ind}_g a \pmod{p-1}$；令 $\delta = \gcd(d, p-1)$，有解时恰有 δ 个解。

关键　原根判别排除阶为 $\varphi(m)$ 的真因子：任何真因子都整除某个 $\varphi(m)/q$。指标则把 g^k 的乘法变成指数加法。完整证明见第 8 讲；指标方程本身使用第 4、5 讲的一次同余理论。

用法　先确认原根，再制幂表，最后取指标；不能先假定某个数是原根。若模数根本没有原根，这条路线就不可用。第 8 讲的存在性清单是 $1, 2, 4, p^e, 2p^e$，其中 p 为奇素数；有原根时其个数为 $\varphi(\varphi(m))$。

R1.3　基础计算：把步骤写到能够回代

三个短计算检验三种不同能力：例 R1.1 检验整除与回代，例 R1.2 比较两条求幂路线，例 R1.3 检查唯一分解与互素条件。即使最后数字正确，也要能说明为什么所用的公式在这里合法。

■ **例 R1.1**　**期中练习第 1、2 题**　求 $\gcd(621, 483)$，找出一组 Bézout 系数，并写出 $621x + 483y = 69$ 的全部整数解。

解　先用欧几里得算法求公因数，再逆着余数等式回代：

$$621 = 483 + 138, \quad 483 = 3 \cdot 138 + 69, \quad 138 = 2 \cdot 69.$$

故最大公因数为 69，而

$$69 = 483 - 3 \cdot 138 = 483 - 3(621 - 483) = -3 \cdot 621 + 4 \cdot 483.$$

得到特解 $(-3, 4)$。约化系数为 $621/69 = 9$、$483/69 = 7$，因此全部解为

$$x = -3 + 7t, \qquad y = 4 - 9t, \qquad t \in \mathbb{Z}.$$

代回时参数项为 $621 \cdot 7t - 483 \cdot 9t = 0$，常数项为 69。反过来，任意解与特解相减，得 $9(x+3) = -7(y-4)$；由 $\gcd(9, 7) = 1$，有 $7 \mid x + 3$，即 $x + 3 = 7t$，继而 $y - 4 = -9t$，所以没有遗漏。 ♦

点评　看到一对未分解的大整数，先做带余除法，不必先试素因子。本题把第 1 讲的 Bézout 与第 2 讲的算法、通解结合起来。原解答的一行把 483 抄成了 638；回代题面中的数即可发现该笔误。若右端不被 69 整除，回代前就应判无解。

■ **例 R1.2**　**期中练习第 3、4 题**　分别用反复平方与 Fermat 小定理计算 $3^{64} \bmod 67$。

解　第一种方法逐次平方，每次立刻取模，避免出现不必要的大整数：

k	2	4	8	16	32	64
$3^k \bmod 67$	9	14	62	25	22	15

例如 $3^8 \equiv 14^2 = 196 \equiv -5$，下一次平方得 25，再平方得 $625 \equiv 22$，最后 $484 \equiv 15$。

第二种方法利用 67 为素数，且 $67 \nmid 3$，所以 $3^{66} \equiv 1 \pmod{67}$。将指数 64 看成 $66-2$，得到 $9 \cdot 3^{64} \equiv 1$。因此只需求 9 的逆元；由 $15 \cdot 9 = 135 = 2 \cdot 67 + 1$，可得 $3^{64} \equiv 15 \pmod{67}$。两种路线相互核对。 ♦

点评　指数恰为二的幂时反复平方很直接；指数靠近 $p-1$ 时，Fermat 加逆元往往更短。第一种方法不依赖模数为素数，第二种依赖素数与互素条件；若换成合数模，必须先改用合法的 Euler 指数或元素的阶，不能仍把 $m-1$ 当作通用周期。

■ **例 R1.3　期中练习第 5、9、10 题**　计算 $\varphi(576)$，并解释 $\varphi(mn) = \varphi(m)\varphi(n)$ 中的互素条件能否删去。

解　先分解模数：$576 = 2^6 \cdot 3^2$。两个素数幂互素，所以

$$\varphi(576) = \varphi(2^6)\varphi(3^2) = (2^6 - 2^5)(3^2 - 3) = 32 \cdot 6 = 192.$$

互素条件不能删去。例如 $m = n = 2$，有 $\varphi(m)\varphi(n) = 1$，但 $\varphi(mn) = \varphi(4) = 2$。

事实上，原题还要求判断逆命题。设同时整除 m, n 的素数全体为 S，由乘积公式逐素数比较，得到

$$\frac{\varphi(m)\varphi(n)}{\varphi(mn)} = \prod_{p \in S} \left(1 - \frac{1}{p}\right).$$

只出现于一个数的素因子，其因子在分子分母中抵消；每个公共素因子则多留一个 $1 - 1/p$。若 S 非空，右边严格小于 1；若 S 为空，右边为空积 1。所以等式成立当且仅当 $\gcd(m, n) = 1$，逆命题也成立。 ♦

点评　先写分解式再套 φ 公式，可以避免把 $\varphi(p^e)$ 错记成 $p^e - 1$。第 2 讲的唯一分解为逐素数比较提供依据，第 5 讲的 CRT 解释互素时为什么积性成立。上面的逆命题是已有公式的直接应用，没有把积性定义擅自推广到所有整数对。

R1.4　局部求根、整体合并与证明题

模 25 与模 35 看起来接近，处理顺序却不同：前者沿着同一个素数提升，后者把两个素数模分别求解再合并。例 R1.4 同时展示这两个流程；例 R1.5 检查证明题中最容易漏掉的特殊因子结构；例 R1.6 与例 R1.7 分别训练构造和组合解释。

■ **例 R1.4　期中练习第 6、7 题**　令 $f(x) = x^3 - x + 1$，分别求 $f(x) \equiv 0 \pmod{25}$ 与 $f(x) \equiv 0 \pmod{35}$ 的全部解。

解　先把模素数的根找全。逐个代入得到余数表：

$x \bmod 5$	0	1	2	3	4
$f(x) \bmod 5$	1	1	2	0	1

$x \bmod 7$	0	1	2	3	4	5	6
$f(x) \bmod 7$	1	1	0	4	5	2	1

因此模 5 唯一根为 $3 \equiv -2$，模 7 唯一根为 2。

对模 25，写候选 $x = -2 + 5t$、$t \in \{0,1,2,3,4\}$。$f(-2) = -5$，$f'(-2) = 11 \equiv 1 \pmod 5$，导数可逆。利用一次展开，

$$f(-2+5t) \equiv -5 + 5t \cdot 11 \equiv 5(-1+t) \pmod{25}.$$

所以 $t \equiv 1 \pmod 5$，得到 $x \equiv 3 \pmod{25}$。每个模 25 的根必须约化为唯一的模 5 根，而这根只有一个提升，因此结果已穷尽；回代 $f(3) = 25$。

对模 35，必须同时满足 $x \equiv 3 \pmod 5$、$x \equiv 2 \pmod 7$。令 $x = 3 + 5t$，得到 $5t \equiv -1 \equiv 6 \pmod 7$，$5^{-1} \equiv 3 \pmod 7$，故 $t \equiv 4 \pmod 7$，即 $x \equiv 23 \pmod{35}$。验算 $23 \equiv 3 \pmod 5$、$23 \equiv 2 \pmod 7$，或直接算 $f(23) = 12145 = 35 \cdot 347$。两个局部根各只有一个，CRT 保证模 35 也只有一个解。 ♦

点评　这是第 5–7 讲的综合题：分解模数决定是否用 Hensel 或 CRT，导数决定提升是否唯一，局部计数决定整体结果有没有遗漏。若模 p 有多个根，必须全部逐根处理；若某个根奇异，也不能仅凭导数不可逆就判它无提升。

■ **例 R1.5　期中练习第 13 题，补足边界条件**　证明对合数 $n > 4$ 有 $(n-1)! \equiv 0 \pmod n$，并检查 $n = 4$。

解　先将合数写成两个较小因子的乘积，再检查它们能否同时在阶乘中使用。取 p 为 n 的最小素因子。若 $n \neq p^2$，则 p 与 n/p 不相等，且都在 $2, \dots, n-1$ 内，是阶乘中两个不同的因子；其乘积为 n，所以 $n \mid (n-1)!$。

若 $n = p^2$，不能把同一个 p 使用两次。由于 $n > 4$，有 $p \geq 3$，因此 p 与 $2p$ 是两个不同且小于 p^2 的正整数。阶乘包含它们，因而被 $2p^2 = 2n$ 整除，也就被 n 整除。最后 $n = 4$ 时 $3! = 6 \equiv 2 \pmod 4$，结论失败。原练习题面漏写 $n > 4$，解答中已经补上；检验最小合数即可发现边界。 ♦

点评　看到阶乘整除题，可以寻找互不相同的因子，把它们的乘积放进阶乘。平方情形最容易把同一个位置数两遍；第 2 讲的唯一分解或第 3 讲的赋值也能核对这一点。条件 $n > 4$ 正是排除唯一的合数例外，不能为了统一写法将它省掉。

■ **例 R1.6　期中练习第 14 题**　求同余式组 $x \equiv 1 \pmod 3$、$x \equiv 2 \pmod 5$、$x \equiv 3 \pmod 7$ 的全部整数解。

解　先检查模数两两互素，再构造只在一个分量为 1 的基元。总模数为 $M = 3 \cdot 5 \cdot 7 = 105$。对应 $3, 5, 7$ 的补积分别为 $35, 21, 15$，它们模相应素数的逆元分别为 $-1, 1, 1$。于是可取基元

$$E_3 = -35, \quad E_5 = 21, \quad E_7 = 15.$$

例如 $E_3 \equiv 1 \pmod 3$，而模 5 与模 7 都为 0；其余两个基元同样逐项满足要求。将目标余数乘到相应基元并求和：

$$x \equiv 1 \cdot (-35) + 2 \cdot 21 + 3 \cdot 15 = 52 \pmod{105}.$$

验算 $52 = 3 \cdot 17 + 1 = 5 \cdot 10 + 2 = 7 \cdot 7 + 3$。若 x' 也满足三式，则 $3, 5, 7$ 都整除 $x' - 52$，它们两两互素，故乘积 105 整除差，得到全部解 $x = 52 + 105t$、$t \in \mathbb{Z}$。 ♦

点评　基元构造同时解释存在性与可计算性；解之差的整除性解释唯一性。若只写出 52，还没有回答「所有整数解」。若将模数改成不互素的数，需要先检查余数相容，再把最后的模数换成最小公倍数。

■ **例 R1.7　期中练习第 15 题**　对正整数 n，证明 $\sum_{i=1}^{n} i\binom{n}{i} = n2^{n-1}$。

解　将权重 i 解释成一个额外选择：从 n 个元素中选一个非空子集，并在其中指定一个元素。若先选子集大小 i，有 $\binom{n}{i}$ 个子集，每个子集有 i 种指定方式，合计是左边。

若先选被指定的元素，有 n 种方式。其余 $n-1$ 个元素各自决定是否进入子集，彼此独立，故有 2^{n-1} 种选择，合计是右边。两种顺序数的是同一批对象，等式成立。

也可以逐项使用 $i\binom{n}{i} = n\binom{n-1}{i-1}$，得到

$$\sum_{i=1}^{n} i\binom{n}{i} = n\sum_{j=0}^{n-1}\binom{n-1}{j} = n(1+1)^{n-1}.$$

这里换指标 $j = i - 1$ 后，上下限分别变成 $0, n-1$。端点 $n = 1$ 给出 $1 = 1$，也符合计数解释。 ♦

点评　二项式系数乘一个指标，常表示「从已选元素中再指定一个」。这是第 3 讲的组合意义与二项式定理两种语言的对应。若题目把 i 换成 $i(i-1)$，就尝试指定两个有顺序且不同的元素；不能在改指标时保留原来的求和边界。

R1.5　准确的阶与指标方程

$a^{\varphi(m)} \equiv 1$ 只给出指数周期的一个上界，实际阶可能更小。例 R1.8 先在互素分量求准确阶，再合并；例 R1.9 将模 p 的幂方程转为模 $p-1$ 的一次同余，训练模数改变时的检查。

■ **例 R1.8　期中练习第 16 题**　求 3 模 301 的阶。

解　先分解 $301 = 7 \cdot 43$，两个因子互素。模 7 时，阶整除 6；检验 $3^2 \equiv 2 \not\equiv 1$、$3^3 \equiv -1 \not\equiv 1$，排除 6 的两个最大真因子 $2, 3$，因此 $\operatorname{ord}_7(3) = 6$。

模 43 时，阶整除 $42 = 2 \cdot 3 \cdot 7$，只需排除 $42/2, 42/3, 42/7$：

$$3^{21} \equiv 42, \qquad 3^{14} \equiv 36, \qquad 3^{6} \equiv 41 \pmod{43}.$$

三者均不等于 1。若阶是 42 的真因子，它必整除三个被检验指数中的某一个，与对应计算矛盾，所以 $\operatorname{ord}_{43}(3) = 42$。

最后，$3^k \equiv 1 \pmod{301}$ 当且仅当它模 7 与模 43 均为 1，亦即 $6 \mid k$ 且 $42 \mid k$。满足条件的最小正 k 是 $\operatorname{lcm}(6,42) = 42$，故所求阶为 42。这也核对了原解答最终结论，不能把两分量的阶直接相乘成 252。 ♦

点评　这题综合第 5 讲的 CRT 与第 7、8 讲的阶判别。分量阶合并时取最小公倍数；同一模数下两个元素相乘时，只有阶互素才可直接把阶相乘。这两个操作作用于不同对象，不能仅因都出现「乘法」而混用。

■ **例 R1.9　第 8 讲方法的综合练习**　证明 2 是模 13 的原根，再求 $x^4 \equiv 3 \pmod{13}$ 的全部解。

解　先验证原根条件，再把未知数写成原根的幂。$\varphi(13) = 12$ 的不同素因数为 $2, 3$，需要检验 $2^6 \equiv 12$、$2^4 \equiv 3 \pmod{13}$，二者均非 1，所以 2 的阶为 12。幂表为

k	0	1	2	3	4	5	6	7	8	9	10	11
$2^k \bmod 13$	1	2	4	8	3	6	12	11	9	5	10	7

因右端 3 非零，所有解都非零，可以唯一写成 $x \equiv 2^k$、$k \bmod 12$。又 $3 \equiv 2^4$，故原式等价于

$$4k \equiv 4 \pmod{12} \iff k \equiv 1 \pmod{3}.$$

模 12 的四个指标是 $1, 4, 7, 10$，对应 $x \equiv 2, 3, 11, 10 \pmod{13}$。所以全部解类为 $2, 3, 10, 11$。每个指标满足一次同余，反向代入原根幂即满足原方程；原根遍历全部非零剩余，所以也没有遗漏。根数 4 与 $\gcd(4, 12)$ 一致。 ♦

点评　这是第 8 讲原根与第 4、5 讲一次同余的综合。指标方程的模数是 12，不是 13；除以 4 后模数变成 3，但原问题需要模 12 的四个指标。若右端为 0，指标没有定义，必须先单独处理 $x \equiv 0$。

R1.6　常见错误与交卷前检查

证明题里最危险的遗漏往往只少几个字：互素、素数、最小、全部。下面五个错误分别对应消去、降幂、合并模数、根数上界与准确阶。发现同类错误时，要回到定理条件修正整条推理，不能仅改最后一个余数。

约去公因子却保留原模数　由 $8x \equiv 4 \pmod{12}$ 直接写成 $2x \equiv 1 \pmod{12}$ 会漏解。正确约化是 $2x \equiv 1 \pmod{3}$，即 $x \equiv 2 \pmod{3}$，在模 12 下给出 $2, 5, 8, 11$ 四个解。约分后先记住新模数，再返回题目要求的原模数列全解类。

不互素时照搬 Euler 降幂　不能把任意指数都直接按 $\varphi(m)$ 约化。模 8 时，$\varphi(8) = 4$，但 $2^4 \equiv 0$，而 $2^0 \equiv 1$。底数与模数不互素时，先分素数幂处理整除性；只有互素的分量才能按 Euler 指数合法降幂。

不检查相容就合并同余式　$x \equiv 1 \pmod 4$、$x \equiv 2 \pmod 6$ 无解：前式要求奇数，后式要求偶数。一般情形先检查 $\gcd(4,6) = 2$ 是否整除余数差 $1-2$。只有通过相容性检查，才谈最小公倍数模下的唯一解；模数乘积唯一性还要求两两互素。

把次数当作所有模数的根数上界　$x^2 \equiv 1 \pmod{15}$ 有 $1, 4, 11, 14$ 四个根，超过次数 2。Lagrange 根数上界要求素数模且首项系数不被该素数整除。合数模的根可来自多个局部分量组合，应先分解再用 CRT 计数，不能在找到两个根后宣布结束。

得到一个可行指数就当作准确阶　$2^{12} \equiv 1 \pmod 7$，但 2 模 7 的阶是 3，并非 12。等式只说明准确阶整除 12。若已有 $a^h \equiv 1$，要证明阶为 h，还需对 h 的每个不同素因子 q 检查 $a^{h/q} \not\equiv 1$；漏检任意一个素因子都可能误判。

交卷前按问题类型做最后一次核对：整数通解代回并检查参数步长；一次同余回到原模数数解；CRT 在每个局部模下代入；Hensel 同时检查原根与导数；阶判别检查所有不同素因子。每一步核对都应针对原题条件，避免用同一条有误的中间式再验证自身。

本讲小结

- 对不全为 0 的 a, b，整数线性方程先求 $g = \gcd(a,b) > 0$：可解性为 $g \mid c$，通解步长为 $(b/g, -a/g)$。
- 阶乘、二项式的整除性看素数赋值；带权二项式和先尝试计数解释或求导。
- 大幂先检查互素条件，再用 Euler、Fermat 或准确阶降指数；反复平方可在任意正模数下使用。
- $ax \equiv b \pmod m$ 约化时系数、右端、模数同时除以 $g = \gcd(a,m)$；若 $g \mid b$，模 m 恰有 g 个解，否则无解。
- 素数幂同余先在模 p 下找根：非奇异根由 Hensel 唯一提升；奇异根的 p 个同层候选全为根当且仅当 $p^{j+1} \mid f(a)$；不同素数幂再由 CRT 合并。
- 求准确阶须排除真因子；指标把模 p 的幂方程化成模 $p-1$ 的一次同余，模数随之改变。

练习

R1.1 ★　求 $\gcd(18, 30)$，写一组 Bézout 系数，进而求 $18x + 30y = 6$ 的全部整数解，说明为什么没有遗漏。（第 1–2 讲）

R1.2 ★　用 Legendre 公式求 $v_2(100!)$、$v_5(100!)$，再求 $v_2\binom{20}{10}$、$v_5\binom{20}{10}$。（第 3 讲）

R1.3 ★　求最小正整数 N，使每个 $n \geq N$ 都满足 $\varphi(n) \geq 5$；证明你的 N 确实最小。（第 5 讲；期中练习第 8 题）

R1.4 ★　用 Wilson 定理求 $9! \bmod 11$，不得逐项计算整个阶乘。（第 4 讲）

R1.5 ★★ 求 $x \equiv 1 \pmod 4$、$x \equiv 2 \pmod 9$ 的全部整数解，验算并给出唯一性所在的模数。（第 5 讲）

R1.6 ★★ 在教科书 RSA 算术练习中取 $p = 5, q = 11, e = 3$。求 $N, \varphi(N)$、私钥指数 f，再对消息 $m = 12$ 加密、解密，核对结果。（第 6 讲）

R1.7 ★★ 求 2 与 3 模 7 的准确阶，再求 2 模 35 的阶。只给出 Euler 上界不算完成。（第 7 讲）

R1.8 ★★ (a) 求模 13 的全部原根，并核对个数。（第 8 讲）(b) 设 p 为素数、g 为模 p 的原根，说明 $1, g, \dots, g^{p-2}$ 遍历全部非零剩余，再求 $S_k = \sum_{a=1}^{p-1} a^k$ 的模 p 闭式，其中 $k \geq 1$。（第 8 讲；习题集 4 第一部分第 6 题）

R1.9 ★★★ 求 $x^2 \equiv 1 \pmod{75}$ 的全部解。先求模 5 的根，再提升到模 25，最后与模 3 的根合并，解释为何四个解全找到了。（第 5–7 讲综合）

R1.10 ★★★ 利用例 R1.9 的原根表，求 $x^6 \equiv 12 \pmod{13}$ 的全部解。写出指标方程，说明它有多少解，再变回剩余类。（第 4、5、8 讲综合）

练习答案与提示

R1.1 $30 = 18 + 12$、$18 = 12 + 6$、$12 = 2 \cdot 6$，所以最大公因数为 6。回代 $6 = 18 - (30 - 18) = 2 \cdot 18 - 30$，取特解 $(2, -1)$。全部解为 $x = 2 + 5t$、$y = -1 - 3t$、$t \in \mathbb{Z}$。代回参数项抵消；任意解与特解相减得 $3(x - 2) = -5(y + 1)$，故 $5 \mid x - 2$，进而得到上述整数参数，保证不漏解。

R1.2 $v_2(100!) = 50 + 25 + 12 + 6 + 3 + 1 = 97$，$v_5(100!) = 20 + 4 = 24$。$v_2(20!) = 10 + 5 + 2 + 1 = 18$，$v_2(10!) = 5 + 2 + 1 = 8$，故 $v_2\binom{20}{10} = 18 - 2 \cdot 8 = 2$。另有 $v_5(20!) = 4$、$v_5(10!) = 2$，所以 $v_5\binom{20}{10} = 4 - 2 \cdot 2 = 0$。用阶乘赋值作差，不能把二项式系数误当作 $20!/10!$。

R1.3 对奇数 m，素数情形有 $\varphi(m) = m - 1$；若 m 合数，则含 p^2 时 $\varphi(m) \geq p(p-1) \geq 6$，含两个不同奇素因子 p, q 时 $\varphi(m) \geq (p-1)(q-1) \geq 8$。因此 $m \geq 3, 5, 7$ 时分别有 $\varphi(m) \geq 2, 4, 6$。写 $n = 2^e m$、m 为奇数。$n \geq 13$ 时，$e = 0, 1$ 给出 $m \geq 13, 7$，故 $\varphi(n) \geq 6$；$e = 2, 3$ 给出 $m \geq 5, 3$，故 $\varphi(n) \geq 8$；$e \geq 4$ 时 $\varphi(n) = 2^{e-1}\varphi(m) \geq 8$。所以 $N = 13$ 可行，而 $\varphi(12) = 4$ 证明它最小。

R1.4 Wilson 给出 $10! \equiv -1 \pmod{11}$。又 $10! \equiv -9! \pmod{11}$，因此 $9! \equiv 1 \pmod{11}$。将 10 换成 -1 后仅乘以 -1，无需计算 9! 的十进制值。

R1.5 令 $x = 1 + 4t$，有 $4t \equiv 1 \pmod 9$。因 $4^{-1} \equiv 7$，得 $t \equiv 7 \pmod 9$，从而 $x \equiv 29 \pmod{36}$。回代 $29 \equiv 1 \pmod 4$、$29 \equiv 2 \pmod 9$。两模互素，所以解模 36 唯一；全部整数解为 $29 + 36s$、$s \in \mathbb{Z}$。

R1.6 $N = 55$、$\varphi(N) = (5-1)(11-1) = 40$。由 $3 \cdot 27 = 81 \equiv 1 \pmod{40}$，取 $f = 27$。消息 12 与 55 互素，加密得 $c \equiv 12^3 = 1728 \equiv 23 \pmod{55}$。解密时 $23^2 \equiv 34$、$23^4 \equiv 1$，故 $23^{27} \equiv 23^3 \equiv 23 \cdot 34 \equiv 12 \pmod{55}$，回到原消息。这也与 Euler 证明 $12^{81} \equiv 12$ 一致。

R1.7 模 7 时 $2^1 \equiv 2$、$2^2 \equiv 4$、$2^3 \equiv 1$，准确阶为 3。$3^2 \equiv 2$、$3^3 \equiv -1$，排除 6 的真因子中最大的 2, 3，所以准确阶为 6。模 5 时 2 的阶为 4，模 7 时为 3；因 5, 7 互素，$2^k \equiv 1 \pmod{35}$ 等价于 $4 \mid k$ 且 $3 \mid k$，因此准确阶为 12。

R1.8 (a) 2 是模 13 的原根，故指数 $k = 1, 5, 7, 11$ 给出全部原根 $2, 6, 7, 11$，共 $\varphi(12) = 4$ 个。(b) 原根的幂遍历非零剩余，所以 $S_k \equiv -1 \pmod p$ 若 $p - 1 \mid k$，否则 $S_k \equiv 0 \pmod p$；后一情形中等比和的分母可逆、分子为 0。

R1.9 模 25 的局部根为 $1, 24$，模 3 的根为 $1, 2$；CRT 给出 $x \equiv 1, 26, 49, 74 \pmod{75}$。四个局部组合一一对应这四类，故无遗漏，不能只取整体的 ± 1。

R1.10 $12 \equiv 2^6 \pmod{13}$，令 $x \equiv 2^k$ 得 $6k \equiv 6 \pmod{12}$，故 k 为奇数，根为 $2, 5, 6, 7, 8, 11$。原根遍历全部非零剩余且 0 不是根，所以答案完备。

第二部分

二次剩余与算术函数

第 9 讲

二次剩余与二次互反律

Quadratic Residues, Quadratic Reciprocity

对应材料：MIT 18.781 Lecture 9 讲义（Quadratic Residues, Quadratic Reciprocity）；习题集 5 第 3–11 题。

本讲目标

(1) 会用配方把模 p 的二次同余式化成 $y^2 \equiv D \pmod p$，知道它的解数只能是 0、1 或 2。

(2) 理解二次剩余的定义，知道模 p 的非零剩余类恰有一半是二次剩余，掌握 Euler 判别法 $a^{(p-1)/2} \equiv \pm 1$。

(3) 会用 Legendre 符号的乘法性，把一般的 $\left(\frac{a}{p}\right)$ 化归为 $\left(\frac{-1}{p}\right)$、$\left(\frac{2}{p}\right)$ 和 $\left(\frac{q}{p}\right)$（q 为奇素数）。

(4) 理解 Gauss 引理的证明，并由它推出 $\left(\frac{2}{p}\right)$ 的公式和 Eisenstein 的取整和公式。

(5) 理解二次互反律的格点证明，会用它计算 Legendre 符号，并用同余条件刻画「a 是模 p 的二次剩余」的素数 p。

9.1 从二次同余式到 $x^2 \equiv a$

第 4、5 讲彻底解决了一次同余式：$ax \equiv b \pmod m$ 有解当且仅当 $\gcd(a, m) \mid b$。对更高次的多项式同余式 $f(x) \equiv 0 \pmod m$，前几讲也搭好了框架：中国剩余定理（第 5 讲）把模 m 拆成模素数幂，Hensel 引理（第 6、7 讲）又把模 p 的解提升到模 p^e。所以真正要回答的是**模素数 p 的同余式**。本讲研究其中最简单的非线性情形，即二次同余式

$$ax^2 + bx + c \equiv 0 \pmod p, \qquad p \nmid a.$$

$p = 2$ 时只有两个剩余类，把 $x = 0, 1$ 代进去看一看就行。以下总设 p 是**奇素数**。

9.1.1 配方

中学里解 $ax^2 + bx + c = 0$ 靠的是配方，得到求根公式 $x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$。模 p 时同样可以配方，只要把「除以 $2a$」理解为「乘 $2a$ 的逆」。因为 p 是奇素数且 $p \nmid a$，所以 $p \nmid 4a$，$4a$ 模 p 可逆。原同余式两边乘 $4a$：

$$4a^2x^2 + 4abx + 4ac = (2ax + b)^2 - (b^2 - 4ac) \equiv 0 \pmod p.$$

记**判别式**（discriminant）$D = b^2 - 4ac$，令 $y = 2ax + b$，就得到 $y^2 \equiv D \pmod p$。

命题 9.1（配方）　设 p 是奇素数，$p\nmid a$，$D=b^2-4ac$。则对应 $x\mapsto y=2ax+b$ 是同余式

$$ax^2+bx+c\equiv 0 \pmod p \qquad 与 \qquad y^2\equiv D \pmod p$$

的解（作为模 p 的剩余类）之间的一一对应，逆对应是 $x\equiv (2a)^{-1}(y-b)$。

证明　由上面的恒等式，$4a\,f(x)=(2ax+b)^2-D$。因为 $4a$ 模 p 可逆，$f(x)\equiv 0$ 当且仅当 $(2ax+b)^2\equiv D$。又 $2a$ 模 p 可逆，$x\mapsto 2ax+b$ 是模 p 剩余类集合到自身的双射。 ■

判别式也说明根是否为重根。恒等式 $4af(x)=(2ax+b)^2-D$ 给出 $D=0$ 时唯一根满足 $2ax+b=0$，即 $f'(x)=0$；若 $D\neq 0$ 且有根，则 $2ax+b=\pm\sqrt{D}\neq 0$，两根互异且导数不为 0。这也补足习题集 5 第 3 题中的判别式边界。

这里用到了两次「p 是奇数」：2 和 a 都要可逆。于是**一切模 p 的二次同余式都归结为 $x^2\equiv a \pmod p$**。下一个问题是：这种同余式有几个解？

9.1.2　解的个数

命题 9.2（解数只有三种可能）　设 p 是奇素数。

(1) 若 $p\mid a$，则 $x^2\equiv a \pmod p$ 恰有一个解 $x\equiv 0$。

(2) 若 $p\nmid a$，且 $x^2\equiv a$ 有一个解 $x\equiv b$，则它恰有两个解 $x\equiv \pm b$，并且 $b\not\equiv -b$。

所以 $x^2\equiv a \pmod p$ 的解数只能是 0、1 或 2。

证明　(1) $x^2\equiv 0$ 即 $p\mid x\cdot x$，由 Euclid 引理 $p\mid x$。

(2) $x^2\equiv a\equiv b^2$ 即 $p\mid x^2-b^2=(x-b)(x+b)$，由 Euclid 引理 $p\mid x-b$ 或 $p\mid x+b$，所以解只有 $\pm b$。若 $b\equiv -b$，则 $p\mid 2b$；但 p 是奇数且 $p\nmid b$（因为 $p\nmid b^2=a$），矛盾。 ■

这也是第 7 讲 Lagrange 定理（模素数的 n 次同余式至多 n 个解）在 $n=2$ 时的特例，上面的证明只是把它写得更具体。

■ **例 9.1　判别式决定一切**　在模 13 下讨论下列三个同余式：(a) $2x^2+5x+1\equiv 0$；(b) $x^2+3x+5\equiv 0$；(c) $4x^2+4x+1\equiv 0$。

解　(a) $D=5^2-4\cdot 2\cdot 1=17\equiv 4 \pmod{13}$。$y^2\equiv 4$ 的解是 $y\equiv\pm 2$。由 $y=2ax+b=4x+5$ 得 $x\equiv 4^{-1}(y-5)$。因为 $4\cdot 10=40\equiv 1$，$4^{-1}\equiv 10$，于是

$$y=2:\ x\equiv 10\cdot(-3)=-30\equiv 9, \qquad y=-2:\ x\equiv 10\cdot(-7)=-70\equiv 8 \pmod{13}.$$

检验：$2\cdot 81+45+1=208=16\cdot 13$，$2\cdot 64+40+1=169=13^2$。两个解 $x\equiv 8,9$。

(b) $D=9-20=-11\equiv 2 \pmod{13}$。把 $1,\dots,6$ 平方得 $1,4,9,3,12,10$，再加上 $(13-x)^2\equiv x^2$，模 13 的非零平方只有 $\{1,3,4,9,10,12\}$，其中没有 2。所以 $y^2\equiv 2$ 无解，原同余式无解。

(c) $D=16-16=0$，$y^2\equiv 0$ 只有 $y\equiv 0$。这里 $y=2ax+b=8x+4$，而 $8\cdot 5=40\equiv 1$，$8^{-1}\equiv 5$，所以 $x\equiv 5\cdot(-4)=-20\equiv 6 \pmod{13}$，是唯一解（「二重根」）。直接看也一样：$4x^2+4x+1=(2x+1)^2$，而 $2\cdot 6+1=13$。 ♦

易错提醒　「至多两个平方根」依赖于模数是素数。在合数模下，它一般不成立。例如模 15 时 $x^2\equiv 4$ 有四个解 $x\equiv 2,7,8,13$（$7^2=49=3\cdot 15+4$）；模 8 时 $x^2\equiv 1$ 也有四个解 $1,3,5,7$。证明里的「$p\mid(x-b)(x+b)\Rightarrow p\mid x-b$ 或 $p\mid x+b$」对合数不再成立：$15\mid(7-2)(7+2)=45$，但 15 既不整除 5 也不整除 9。

前后联系　模素数幂的情形可以由 Hensel 引理（第 6 讲）补上。取 $f(x)=x^2-a$，$p\nmid a$。模 p 的根 b 满足 $f'(b)=2b\not\equiv 0$，所以它唯一地提升为模 p^e 的根。因此对奇素数 p 和 $p\nmid a$，$x^2\equiv a \pmod{p^e}$ 的解数与模 p 时相同（0 或 2）。再用中国剩余定理，模任意奇数 m 的解数就是各素数幂因子解数之积。

9.2　二次剩余与 Euler 判别法

9.2.1　定义与计数

定义 9.3（二次剩余）　设 p 是奇素数，$p\nmid a$。若 $x^2\equiv a \pmod{p}$ 有解，称 a 是模 p 的**二次剩余**（quadratic residue, QR）；否则称 a 是模 p 的**二次非剩余**（quadratic nonresidue, QNR）。

由于 $(p-x)^2\equiv x^2$，列出模 p 的全部平方只需平方 $x=1,2,\dots,\frac{p-1}{2}$。表 9.1 是几个小素数的结果。

表 9.1　小素数的二次剩余（平方 $x=1,\dots,\frac{p-1}{2}$ 得到）

p	$x^2 \bmod p$（$x=1,\dots,\frac{p-1}{2}$）	二次剩余	二次非剩余
5	$1,4$	$1,4$	$2,3$
7	$1,4,2$	$1,2,4$	$3,5,6$
11	$1,4,9,5,3$	$1,3,4,5,9$	$2,6,7,8,10$
13	$1,4,9,3,12,10$	$1,3,4,9,10,12$	$2,5,6,7,8,11$

从表中可以看出几件事：每一行二次剩余与非剩余各占一半；$-1\equiv p-1$ 在 $p=5,13$ 时是剩余，在 $p=7,11$ 时不是；2 只在 $p=7$ 时是剩余。第一件事由下述平方根配对论证得到，后两件要到本讲后半部分才能完全解释。

命题 9.4　设 p 是奇素数。模 p 的二次剩余恰有 $\frac{p-1}{2}$ 个，它们是 $1^2,2^2,\dots,\left(\frac{p-1}{2}\right)^2$ 模 p 的值；二次非剩余也恰有 $\frac{p-1}{2}$ 个。

证明　每个非零剩余类都可以写成 $\pm k$，$1 \le k \le \frac{p-1}{2}$，而 $(\pm k)^2 = k^2$，所以二次剩余就是这 $\frac{p-1}{2}$ 个平方。它们两两不同余：若 $k^2 \equiv l^2$，则由命题 9.2，$k \equiv \pm l$；但 $1 \le k, l \le \frac{p-1}{2}$ 时 $k+l$ 介于 2 和 $p-1$ 之间，不能被 p 整除，所以 $k \equiv l$，即 $k = l$。剩下的 $(p-1) - \frac{p-1}{2} = \frac{p-1}{2}$ 个非零剩余类是非剩余。 ■

另一种看法：平方映射 $x \mapsto x^2$ 把 $p-1$ 个非零剩余类「二对一」地映到它的像上（$\pm x$ 对应同一个像），所以像恰好有 $\frac{p-1}{2}$ 个元素。

注 9.1　[前半区间中的剩余偏差] 令 $h = (p-1)/2$，$R = \#\{1 \le a \le h : \left(\frac{a}{p}\right) = 1\}$，$N = \#\{1 \le a \le h : \left(\frac{a}{p}\right) = -1\}$。习题集 5 第 10 题要求统计前 100 个奇素数，数据呈现 $p \equiv 1 \pmod 4$ 时 $R = N$、$p \equiv 3 \pmod 4$ 时 $R > N$。第一种情形可证明：-1 是平方，a 与 $p-a$ 的 Legendre 符号相同，故全体 h 个二次剩余组成 $h/2$ 对；每对在 $1, \dots, h$ 中恰有一个代表，所以 $R = N = h/2$。第二种不等式与类数公式有关，本讲不证明；课程原解答指出其证明用到 L 函数和 Dirichlet 类数公式。

表 9.2 列出这 100 个奇素数的计数。$p \equiv 1 \pmod 4$ 的行可由上面的证明直接推出；$p \equiv 3 \pmod 4$ 的计数来自逐个计算 Legendre 符号。

p	R	N	p	R	N	p	R	N	p	R	N
3	1	0	5	1	1	7	2	1	11	4	1
13	3	3	17	4	4	19	6	3	23	7	4
29	7	7	31	9	6	37	9	9	41	10	10
43	12	9	47	14	9	53	13	13	59	19	10
61	15	15	67	18	15	71	21	14	73	18	18
79	22	17	83	25	16	89	22	22	97	24	24
101	25	25	103	28	23	107	31	22	109	27	27
113	28	28	127	34	29	131	40	25	137	34	34
139	39	30	149	37	37	151	41	34	157	39	39
163	42	39	167	47	36	173	43	43	179	52	37
181	45	45	191	54	41	193	48	48	197	49	49
199	54	45	211	57	48	223	59	52	227	64	49
229	57	57	233	58	58	239	67	52	241	60	60
251	73	52	257	64	64	263	72	59	269	67	67
271	73	62	277	69	69	281	70	70	283	75	66
293	73	73	307	81	72	311	87	68	313	78	78
317	79	79	331	87	78	337	84	84	347	94	79
349	87	87	353	88	88	359	99	80	367	96	87
373	93	93	379	99	90	383	104	87	389	97	97
397	99	99	401	100	100	409	102	102	419	118	91
421	105	105	431	118	97	433	108	108	439	117	102
443	118	103	449	112	112	457	114	114	461	115	115
463	119	112	467	127	106	479	132	107	487	125	118
491	136	109	499	129	120	503	136	115	509	127	127
521	130	130	523	138	123	541	135	135	547	141	132

表 9.2　习题集 5 第 10 题：前 100 个奇素数的半区间二次剩余数 R 与非剩余数 N。

想一想　定义里为什么要求 $p \nmid a$？$0 = 0^2$ 明明是一个平方。

答　0 的性质和非零元截然不同：它只有一个平方根（命题 9.2(1)），而且 0 乘任何数都是 0，放进去会破坏后面「剩余 × 非剩余 = 非剩余」之类的乘法规律。所以把 0 单独处理，下一节的 Legendre 符号给它一个专门的值 0。

9.2.2　用原根看二次剩余

第 7、8 讲证明了模 p 存在原根 g，非零剩余类恰好是 $g^0, g^1, \dots, g^{p-2}$。在这个「坐标」下，平方就是把指数乘 2：$(g^l)^2 = g^{2l}$。所以平方数应当是**偶数次幂**。

命题 9.5　设 g 是模 p 的原根，$a \equiv g^k$，$0 \le k \le p-2$。则 a 是二次剩余当且仅当 k 是偶数。

证明　若 $k=2l$，则 $a\equiv (g^l)^2$。反之，若 $a\equiv b^2$，写 $b\equiv g^l$，则 $g^k\equiv g^{2l}$。原根的阶是 $p-1$，所以 $k\equiv 2l \pmod{p-1}$。$p-1$ 是偶数，于是 k 与 $2l$ 同奇偶，k 是偶数。 ■

偶数次幂 $g^0, g^2, \dots, g^{p-3}$ 恰有 $\frac{p-1}{2}$ 个，这又一次证明了命题 9.4。图 9.1 画出了 $p=13$、$g=2$ 的情形。

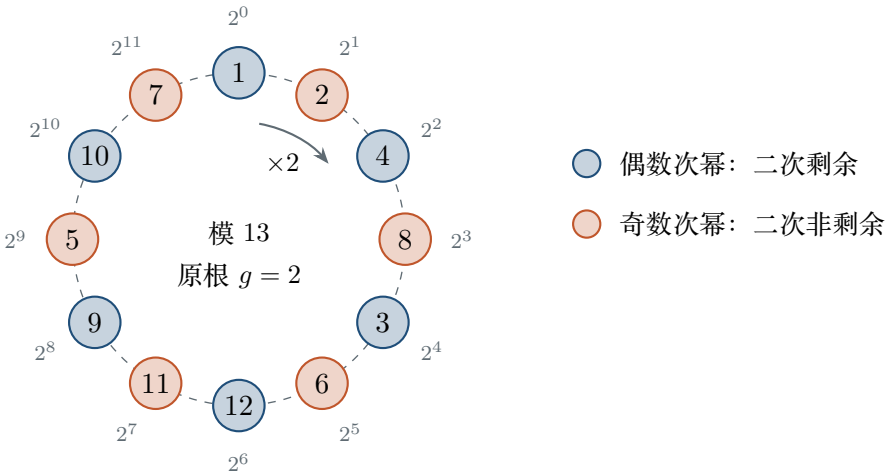


图 9.1　模 13 的非零剩余类按 2^k 排成一圈。每乘一次 2 走一格，剩余与非剩余交替出现；蓝色的 $1,4,3,12,9,10$ 正是表 9.1 中的二次剩余。

9.2.3 Euler 判别法

列表和找原根都只适合小素数。我们想要一个可以直接计算的判别法。由 Fermat 小定理，$p\nmid a$ 时 $\left(a^{(p-1)/2}\right)^2 = a^{p-1}\equiv 1$，所以 $a^{(p-1)/2}$ 是 $y^2\equiv 1$ 的根。由命题 9.2，这个同余式只有 $y\equiv \pm 1$ 两个根。于是$a^{(p-1)/2}$ 只能是 1 或 -1。Euler 发现这个符号恰好区分了剩余与非剩余。

定理 9.6（Euler 判别法）　设 p 是奇素数，$p\nmid a$。则

$$a^{(p-1)/2}\equiv \begin{cases} 1 \pmod{p}, & a \text{ 是二次剩余},\\ -1 \pmod{p}, & a \text{ 是二次非剩余}. \end{cases}$$

思路　有两条路。一是用原根把 a 写成 g^k，问题变成「$g^{k(p-1)/2}\equiv 1$ 当且仅当 k 是偶数」，这只涉及阶。二是数根：所有二次剩余都是 $x^{(p-1)/2}-1$ 的根，而 Lagrange 定理说这个多项式至多有 $\frac{p-1}{2}$ 个根，正好被二次剩余占满，非剩余就只能给出 -1。

证明一（原根）　设 g 是原根，$a\equiv g^k$。则 $a^{(p-1)/2}\equiv g^{k(p-1)/2}$。由于 $\operatorname{ord}_p(g)=p-1$，$g^{k(p-1)/2}\equiv 1$ 当且仅当 $(p-1)\mid k\cdot\frac{p-1}{2}$，即 $2\mid k$。由命题 9.5，这恰好是 a 为二次剩余的条件。不为 1 时，前面已说明它只能是 -1。 ■

证明二（数根）　若 $a\equiv b^2$，则 $a^{(p-1)/2}\equiv b^{p-1}\equiv 1$。所以 $\frac{p-1}{2}$ 个二次剩余都是 $x^{(p-1)/2}-1\equiv 0 \pmod{p}$ 的根。由 Lagrange 定理（第 7 讲），这个 $\frac{p-1}{2}$ 次同余式至多有 $\frac{p-1}{2}$ 个根，所以它的根恰好是二次剩余。于是非剩余 a 满足 $a^{(p-1)/2}\not\equiv 1$，只能 $a^{(p-1)/2}\equiv -1$。 ■

■ **例 9.2**　**用 Euler 判别法检验**　判断 5 和 3 是不是模 31 的二次剩余。

解　$\frac{p-1}{2}=15$。

对 $a=5$：$5^2=25\equiv -6$，$5^3\equiv -30\equiv 1 \pmod{31}$。所以 $5^{15}=(5^3)^5\equiv 1$，5 是二次剩余。确实 $6^2=36\equiv 5$，两个平方根是 ± 6，即 6 和 25。

对 $a=3$：$3^3=27\equiv -4$，$3^5=3^3\cdot 9\equiv -36\equiv -5 \pmod{31}$。所以 $3^{15}=(3^5)^3\equiv (-5)^3=-125=-4\cdot 31-1\equiv -1$，3 是二次非剩余。 ♦

Euler 判别法用反复平方法（第 6 讲）只需 $O(\log p)$ 次乘法，所以判断很快。但它只回答「有没有」，不给出平方根本身；求平方根的算法留到第 11 讲。

推论 9.7（-1 何时是平方）　设 p 是奇素数。-1 是模 p 的二次剩余当且仅当 $p\equiv 1 \pmod 4$。

证明　由 Euler 判别法，-1 是二次剩余当且仅当 $(-1)^{(p-1)/2}\equiv 1$。两边都是 ± 1 而 $p\geq 3$，所以同余就是相等，即 $\frac{p-1}{2}$ 是偶数，$p\equiv 1 \pmod 4$。 ■

第 4 讲已经用 Wilson 定理证明过这个结论，而且还给出了一个具体的平方根，下面重新推一遍。

■ **例 9.3**　**用阶乘写出 $\sqrt{-1}$**　证明：$\left(\left(\frac{p-1}{2}\right)!\right)^2\equiv (-1)^{(p+1)/2} \pmod p$。由此对 $p=13$ 和 $p=17$ 求出 $x^2\equiv -1$ 的一个解。

解　记 $h=\frac{p-1}{2}$。Wilson 定理说 $(p-1)!\equiv -1$。把 $(p-1)!$ 的后一半因子 $p-k$ 换成 $-k (k=1,\dots,h)$：

$$-1\equiv (p-1)!=\prod_{k=1}^{h} k\cdot\prod_{k=1}^{h}(p-k)\equiv h!\cdot(-1)^h h!=(-1)^h(h!)^2 \pmod p.$$

两边乘 $(-1)^h$ 得 $(h!)^2\equiv (-1)^{h+1}=(-1)^{(p+1)/2}$。当 $p\equiv 1 \pmod 4$ 时 $\frac{p+1}{2}$ 是奇数，所以 $(h!)^2\equiv -1$，$h!$ 就是 -1 的一个平方根。

$p=13$：$6!=720=55\cdot 13+5\equiv 5$，检验 $5^2=25=26-1\equiv -1$。

$p=17$：$8!=40320=2371\cdot 17+13\equiv 13$，检验 $13^2=169=170-1\equiv -1$。 ♦

■ **例 9.4**　**形如 $4k+1$ 的素数有无穷多个**　证明形如 $4k+1$ 的素数有无穷多个。

解　5 就是一个。假设这样的素数只有有限个：$p_1,\dots,p_n$。令

$$N=(2p_1p_2\cdots p_n)^2+1.$$

$N>1$ 是奇数，取它的一个素因子 q，则 q 是奇素数，且 $(2p_1\cdots p_n)^2\equiv -1 \pmod q$。所以 -1 是模 q 的二次剩余，由推论 9.7，$q\equiv 1 \pmod 4$，于是 q 是某个 p_i。但 $p_i\mid N-1$，又 $p_i\mid N$，推出 $p_i\mid 1$，矛盾。

（具体看：从 $\{5\}$ 出发，$N=10^2+1=101$，是素数且 $101\equiv 1 \pmod 4$；从 $\{5,13\}$ 出发，$N=130^2+1=16901$，也是 $4k+1$ 型素数。） ♦

9.3　Legendre 符号

9.3.1　定义与基本性质

把「是不是二次剩余」编码成 ± 1，Euler 判别法就变成一个等式，乘法规律也会变得一目了然。

定义 9.8（Legendre 符号）　设 p 是奇素数，$a \in \mathbb{Z}$。定义**Legendre 符号**（Legendre symbol）

$$\left(\frac{a}{p}\right) = \begin{cases} 1, & p \nmid a \text{ 且 } a \text{ 是模 } p \text{ 的二次剩余}, \\ -1, & p \nmid a \text{ 且 } a \text{ 是模 } p \text{ 的二次非剩余}, \\ 0, & p \mid a. \end{cases}$$

读作「a 对 p 的 Legendre 符号」。

易错提醒　$\left(\frac{a}{p}\right)$ 不是分数 a/p，只是借用了分数的写法。下面的公式里，分母位置上永远是奇素数，$\left(\frac{a}{2}\right)$、$\left(\frac{a}{15}\right)$ 都没有定义（第 10 讲会把分母推广到奇合数，但那时的含义有所不同）。

有了这个记号，两个已有结论可以写得很紧凑。Euler 判别法变成

$$\boxed{\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p}}$$

（$p \mid a$ 时两边都是 0，也成立）；命题 9.2 变成「$x^2 \equiv a \pmod{p}$ 恰有 $1 + \left(\frac{a}{p}\right)$ 个解」，三种情形 $-1, 0, 1$ 分别对应 $0, 1, 2$ 个解。

定理 9.9（Legendre 符号的基本性质）　设 p 是奇素数，$a, b \in \mathbb{Z}$。

(1)（只看剩余类）若 $a \equiv b \pmod{p}$，则 $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$。

(2)（完全积性）$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$。

(3)（平方因子可去）若 $p \nmid b$，则 $\left(\frac{ab^2}{p}\right) = \left(\frac{a}{p}\right)$；特别地 $\left(\frac{b^2}{p}\right) = 1$。

(4) $\left(\frac{1}{p}\right) = 1$，$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$。

证明　(1) 由定义，$x^2 \equiv a$ 与 $x^2 \equiv b$ 是同一个同余式。

(2) 由 Euler 判别法，

$$\left(\frac{ab}{p}\right) \equiv (ab)^{(p-1)/2} = a^{(p-1)/2} b^{(p-1)/2} \equiv \left(\frac{a}{p}\right)\left(\frac{b}{p}\right) \pmod{p}.$$

两边都属于 $\{-1, 0, 1\}$，差的绝对值至多为 $2 < p$，而它又是 p 的倍数，所以差为 0。

(3) 由 (2)，$\left(\frac{ab^2}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)^2 = \left(\frac{a}{p}\right)$，因为 $\left(\frac{b}{p}\right) = \pm 1$。

(4) 前者因为 $1 = 1^2$，后者就是推论 9.7（同样用「± 1 同余即相等」）。 ■

性质 (2) 的证明是本讲反复用到的技巧：**两个属于 $\{-1,0,1\}$ 的数模 $p\geq 3$ 同余，则相等。**

9.3.2 剩余与非剩余的乘法

把 (2) 翻译回文字，就是表 9.3。

表 9.3　模奇素数 p 的乘法规律

×	剩余（+1）	非剩余（−1）
剩余（+1）	剩余	非剩余
非剩余（−1）	非剩余	剩余

直观理解　在原根坐标下 $a=g^k$，剩余是偶数次幂，非剩余是奇数次幂。乘法就是指数相加，表 9.3 只是在说指数的奇偶性：$\left(\frac{a}{p}\right)=(-1)^{\operatorname{ind}_g a}$，奇 + 奇 = 偶，所以两个非平方之积是平方。模 13 时 2 和 5 都是非剩余，$2\cdot 5=10\equiv 6^2$。

易错提醒

- **合数模下乘法规律失效**。模 8 时 $3,5,7$ 都不是平方，但 $3\cdot 5\equiv 7$ 也不是平方。表 9.3 依赖原根存在、剩余恰占一半，这是素数模的性质。
- **Legendre 符号没有加法性**：$\left(\frac{1+1}{5}\right)=\left(\frac{2}{5}\right)=-1$，而 $\left(\frac{1}{5}\right)+\left(\frac{1}{5}\right)=2$。

想一想　设模奇素数 p 的非零剩余类被分成两个非空集合 S_1,S_2，满足：同一集合中两数之积在 S_1 中，不同集合中两数之积在 S_2 中。证明 S_1 恰是二次剩余，S_2 恰是非剩余。

答　任一非零 x 与自己同属一个集合，所以 $x^2\in S_1$，全部二次剩余都在 S_1。若 S_1 还含有一个非剩余 n，则每个二次剩余 r 与 n 同在 S_1，故 $rn\in S_1$。当 r 遍历全部二次剩余时，rn 遍历全部非剩余：乘以固定非剩余 n 是一个双射，并把平方类映到非平方类。因此 S_1 包含所有非零剩余，和 S_2 非空矛盾。所以 S_1 恰为二次剩余，S_2 恰为非剩余。

9.3.3 化归到素数

■ **例 9.5**　**用性质计算**　计算 $\left(\frac{-4}{79}\right)$、$\left(\frac{18}{31}\right)$、$\left(\frac{15}{31}\right)$。

解　$-4=(-1)\cdot 2^2$，平方因子可去：$\left(\frac{-4}{79}\right)=\left(\frac{-1}{79}\right)=(-1)^{39}=-1$。所以 $x^2\equiv -4 \pmod{79}$ 无解。

$18=2\cdot 3^2$，所以 $\left(\frac{18}{31}\right)=\left(\frac{2}{31}\right)$。由 Euler 判别法，$2^5=32\equiv 1$，$2^{15}=(2^5)^3\equiv 1$，所以 $\left(\frac{18}{31}\right)=1$。确实 $7^2=49=31+18$。

$15=3\cdot 5$，由例 9.2，$\left(\frac{15}{31}\right)=\left(\frac{3}{31}\right)\left(\frac{5}{31}\right)=(-1)\cdot 1=-1$。 ♦

一般地，把 a 分解为 $a=\pm 2^e q_1^{e_1}\cdots q_r^{e_r}$（$q_i$ 为奇素数），由完全积性

$$\left(\frac{a}{p}\right)=\left(\frac{\pm 1}{p}\right)\left(\frac{2}{p}\right)^e\left(\frac{q_1}{p}\right)^{e_1}\cdots\left(\frac{q_r}{p}\right)^{e_r},$$

而偶数次幂的因子都等于 1。所以只需会算三类符号：$\left(\frac{-1}{p}\right)$（已解决）、$\left(\frac{2}{p}\right)$、以及 $\left(\frac{q}{p}\right)$（$q \neq p$ 为奇素数）。后两类是下面两节的主题。

9.4 Gauss 引理与 $\left(\frac{2}{p}\right)$

9.4.1 绝对最小剩余与 Gauss 引理

Euler 判别法要算 $a^{(p-1)/2}$，看不出 $\left(\frac{a}{p}\right)$ 与 p 之间有什么规律。Gauss 换了一个角度：不去算大幂，而是看 a 的前 $\frac{p-1}{2}$ 个倍数落在哪里。

定义 9.10（绝对最小剩余）　对奇数 p 和整数 x，记 $(x)_p$ 为与 x 模 p 同余、且落在开区间 $\left(-\frac{p}{2}, \frac{p}{2}\right)$ 中的唯一整数，称为 x 模 p 的**绝对最小剩余**（least absolute residue）。

例如 $(10)_{13} = -3$，$(20)_{13} = -6$，$(15)_{13} = 2$。因为 p 是奇数，$\pm\frac{p}{2}$ 不是整数，这个区间恰好含 p 个整数 $-\frac{p-1}{2}, \dots, \frac{p-1}{2}$，所以 $(x)_p$ 定义明确。

引理 9.11（Gauss 引理）　设 p 是奇素数，$p \nmid a$。在 $(a)_p, (2a)_p, \dots, \left(\frac{p-1}{2}a\right)_p$ 中，设有 n 个是负数。则

$$\left(\frac{a}{p}\right) = (-1)^n.$$

先看一个例子。取 $p = 13$，$a = 5$：

$$5,\ 10,\ 15,\ 20,\ 25,\ 30 \longrightarrow (5k)_{13} = 5,\ -3,\ 2,\ -6,\ -1,\ 4.$$

有 $n = 3$ 个负数，引理给出 $\left(\frac{5}{13}\right) = -1$，与表 9.1 一致。这六个数的绝对值 $5, 3, 2, 6, 1, 4$ 恰好是 $1, 2, \dots, 6$ 的一个排列（图 9.2）。证明就建立在这件事上。

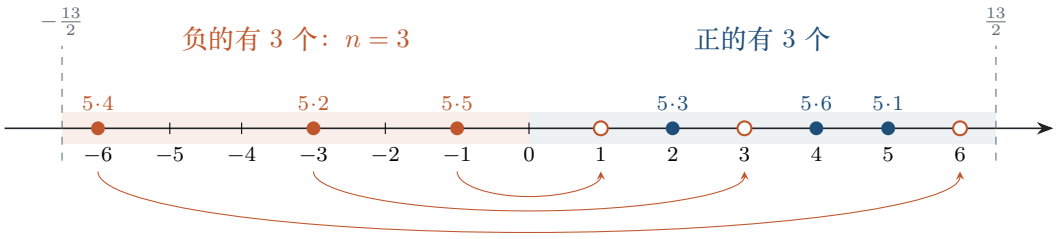


图 9.2　$p = 13$、$a = 5$ 时 $(5k)_{13}$ 的位置。蓝点为正，橙点为负。把负的关于 0 翻折（下方弧线）后得到空心点 $1, 3, 6$，与蓝点 $2, 4, 5$ 合起来恰好每个 $1, \dots, 6$ 出现一次。

证明　记 $h = \frac{p-1}{2}$，$r_k = (ka)_p$（$k = 1, \dots, h$）。

第一步：$|r_1|, \dots, |r_h|$ 两两不同。若 $|r_k| = |r_l|$，则 $ka \equiv \pm la$，消去 a 得 $p \mid k \mp l$。但 $1 \le k, l \le h$ 时 $0 \le |k - l| < p$、$2 \le k + l \le p - 1$，所以只能 $k = l$。又 $p \nmid ka$，故 $r_k \neq 0$，$1 \le |r_k| \le h$。h 个两两不同的数落在 $\{1, \dots, h\}$ 中，所以 $\{|r_1|, \dots, |r_h|\} = \{1, \dots, h\}$。

第二步：把 $r_1 r_2 \cdots r_h$ 算两遍。一方面 $r_k \equiv ka$，乘积 $\equiv a^h\, h!$；另一方面其中恰有 n 个负数，乘积 $= (-1)^n |r_1| \cdots |r_h| = (-1)^n h!$。所以

$$a^h\, h! \equiv (-1)^n h! \pmod{p}.$$

$p \nmid h!$，消去得 $a^{(p-1)/2} \equiv (-1)^n$。由 Euler 判别法 $\left(\frac{a}{p}\right) \equiv (-1)^n$，两边都是 ± 1，所以相等。 ■

证明的关键是只取前一半倍数。若取全部 $a, 2a, \ldots, (p-1)a$，则 ka 与 $(p-k)a$ 互为相反数，绝对值成对重复，就得不到排列。这与第 4 讲证明 Fermat 小定理时「把 $a, 2a, \ldots, (p-1)a$ 乘起来与 $(p-1)!$ 比较」是同一个想法，只是多记录了符号。

想一想　对 $a = -1$ 用 Gauss 引理，能得到什么？

答　$-1, -2, \ldots, -h$ 本身都在 $\left(-\frac{p}{2}, 0\right)$ 中，所以 $n = h = \frac{p-1}{2}$，$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$，这正是推论 9.7。

9.4.2 第二补充律：$\left(\frac{2}{p}\right)$

定理 9.12（2 何时是平方）　设 p 是奇素数，则

$$\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8} = \begin{cases} 1, & p \equiv \pm 1 \pmod 8, \\ -1, & p \equiv \pm 3 \pmod 8. \end{cases}$$

证明　取 $a = 2$。倍数 $2, 4, \ldots, 2h = p-1$ 都在 $(0, p)$ 中，$(2k)_p$ 为负当且仅当 $2k > \frac{p}{2}$，即 $k > \frac{p}{4}$。所以 $n = h - \lfloor p/4 \rfloor$。对 p 模 8 的四种情形逐一计算（写 $p = 8m + r$）：

p	$h = \frac{p-1}{2}$	$\lfloor p/4 \rfloor$	n	n 的奇偶
$8m+1$	$4m$	$2m$	$2m$	偶
$8m+3$	$4m+1$	$2m$	$2m+1$	奇
$8m+5$	$4m+2$	$2m+1$	$2m+1$	奇
$8m+7$	$4m+3$	$2m+1$	$2m+2$	偶

另一方面 $\frac{p^2-1}{8} = \frac{(p-1)(p+1)}{8}$ 在这四种情形下分别为 $m(8m+2)$、$(4m+1)(2m+1)$、$(2m+1)(4m+3)$、$(4m+3)(2m+2)$，奇偶性依次为偶、奇、奇、偶，与 n 一致。 ■

检验：$p = 7, 17, 23, 31$ 都 $\equiv \pm 1 \pmod 8$，确实 $3^2 \equiv 2 \pmod 7$，$6^2 = 36 \equiv 2 \pmod{17}$，$5^2 = 25 \equiv 2 \pmod{23}$，$8^2 = 64 \equiv 2 \pmod{31}$；而 $p = 3, 5, 11, 13$ 都 $\equiv \pm 3 \pmod 8$，表 9.1 中 2 都不是剩余。

■ **例 9.6 Mersenne 数的因子**　设 p 是素数，$p \equiv 3 \pmod 4$，且 $q = 2p + 1$ 也是素数。证明 $q \mid 2^p - 1$。由此说明 $2^{11} - 1$ 和 $2^{23} - 1$ 都是合数。

解　写 $p=4m+3$，则 $q=8m+7\equiv -1 \pmod 8$，由定理 9.12，$\left(\frac{2}{q}\right)=1$。Euler 判别法给出

$$2^{(q-1)/2}=2^p\equiv\left(\frac{2}{q}\right)=1 \pmod q,$$

即 $q\mid 2^p-1$。

$p=11$：$q=23$，$2^{11}-1=2047=23\cdot 89$。$p=23$：$q=47$，$2^{23}-1=8388607=47\cdot 178481$。当 $p>3$ 时 $2^p-1>2p+1=q$，所以 2^p-1 有真因子 q，是合数，尽管指数 p 是素数。（$p=3$ 时 $q=7=2^3-1$ 本身就是素数。）♦

9.5　Eisenstein 引理：把符号变成取整和

Gauss 引理里的 n 不容易直接计算。Eisenstein 发现，当 a 是奇数时，n 的奇偶性等于一个取整和的奇偶性，而取整和可以解释为数格点，这正是证明二次互反律所需要的。

引理 9.13（Eisenstein）　设 p 是奇素数，$p\nmid a$，$h=\frac{p-1}{2}$，$t=\sum_{k=1}^{h}\left\lfloor\frac{ka}{p}\right\rfloor$。设 n 如 Gauss 引理，则

$$n\equiv t+(a-1)\frac{p^2-1}{8} \pmod 2.$$

特别地，当 a 是奇数时 $\left(\frac{a}{p}\right)=(-1)^t$。

证明　做带余除法 $ka=p\lfloor ka/p\rfloor+\rho_k$，$0<\rho_k<p$。$\rho_k<\frac{p}{2}$ 时 $(ka)_p=\rho_k$；$\rho_k>\frac{p}{2}$ 时 $(ka)_p=\rho_k-p$，后者恰有 n 个，此时 $|(ka)_p|=p-\rho_k\equiv 1+\rho_k \pmod 2$（$p$ 是奇数）。由 Gauss 引理证明的第一步，$|(ka)_p|$ 跑遍 $1,\dots,h$，所以

$$\sum_{k=1}^{h}k=\sum_{k=1}^{h}|(ka)_p|\equiv\sum_{k=1}^{h}\rho_k+n \pmod 2.$$

另一方面，对 $ka=p\lfloor ka/p\rfloor+\rho_k$ 求和，并用 $p\equiv 1 \pmod 2$：

$$a\sum_{k=1}^{h}k=pt+\sum_{k=1}^{h}\rho_k\equiv t+\sum_{k=1}^{h}\rho_k \pmod 2.$$

两式相减得 $(a-1)\sum_{k=1}^{h}k\equiv t-n\equiv t+n \pmod 2$。而 $\sum_{k=1}^{h}k=\frac{h(h+1)}{2}=\frac{p^2-1}{8}$，移项即得结论。$a$ 为奇数时 $a-1$ 是偶数，$n\equiv t$，再用 Gauss 引理。■

例 9.7　再算一次 $\left(\frac{5}{13}\right)$　用 Eisenstein 引理计算 $\left(\frac{5}{13}\right)$，并用 $a=2$ 重新得到定理 9.12。

解　$h=6$，$\lfloor 5k/13\rfloor$（$k=1,\dots,6$）依次为 $\lfloor 5/13\rfloor,\lfloor 10/13\rfloor,\dots,\lfloor 30/13\rfloor=0,0,1,1,1,2$，所以 $t=5$ 是奇数，$\left(\frac{5}{13}\right)=(-1)^5=-1$。注意这里 $t=5$ 而 Gauss 引理中 $n=3$，两者不相等，只是奇偶相同。

取 $a=2$：$2k\le p-1<p$，所以每一项 $\lfloor 2k/p\rfloor=0$，$t=0$，引理给出 $n\equiv\frac{p^2-1}{8} \pmod 2$，于是 $\left(\frac{2}{p}\right)=(-1)^{(p^2-1)/8}$。这是定理 9.12 的第二个证明，不用分情形。♦

易错提醒　「$\left(\frac{a}{p}\right)=(-1)^t$」只对奇数 a 成立。偶数 a 时必须带上修正项 $(a-1)\frac{p^2-1}{8}$。例如 $a=2$、$p=5$：$t=\lfloor 2/5\rfloor+\lfloor 4/5\rfloor=0$，但 $\left(\frac{2}{5}\right)=-1$。计算 $\left(\frac{a}{p}\right)$ 时，若 a 是偶数，先用积性把 2 的幂分出去。

取整和就是数格点　设 a 为正数。对固定的 k，$\lfloor ka/p\rfloor$ 是竖线 $x=k$ 上满足 $1\le y<\frac{a}{p}k$ 的整数 y 的个数，也就是这条竖线上严格位于直线 $y=\frac{a}{p}x$ 下方、x 轴上方的格点数。所以 t 是三角形 $0<x\le h$、$0<y<\frac{a}{p}x$ 内的格点数。下一节对两个奇素数 p,q 同时做这件事，两个三角形恰好拼成一个矩形。

9.6　二次互反律

9.6.1　从数据中看出规律

到此为止，$\left(\frac{a}{p}\right)$ 的计算已化归到 $\left(\frac{q}{p}\right)$，其中 $q\neq p$ 都是奇素数。把小素数的结果列成表 9.4，第 p 行第 q 列是 $\left(\frac{p}{q}\right)$，即「p 是不是模 q 的平方」。

表 9.4　$\left(\frac{p}{q}\right)$ 的符号（行为 p，列为 q）。橙色格子是 $p\equiv q\equiv 3\pmod 4$ 的位置

$p\backslash q$	3	5	7	11	13	17	19
3	·	−	−	+	+	−	−
5	−	·	−	+	−	−	+
7	+	−	·	−	−	−	+
11	−	+	+	·	−	−	+
13	+	−	−	−	·	+	−
17	−	−	−	−	+	·	+
19	+	+	−	−	−	+	·

表中看不出某一行有什么规律，但把表和它的转置对比就有了：白色格子关于对角线对称，$\left(\frac{p}{q}\right)=\left(\frac{q}{p}\right)$；橙色格子反对称，$\left(\frac{p}{q}\right)=-\left(\frac{q}{p}\right)$。例如 $\left(\frac{5}{11}\right)=\left(\frac{11}{5}\right)=1$，而 $\left(\frac{3}{7}\right)=-1$、$\left(\frac{7}{3}\right)=1$。橙色恰是两个素数都 $\equiv 3\pmod 4$ 的情形。

定理 9.14（二次互反律）　设 $p\neq q$ 是奇素数，则

$$\boxed{\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)=(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}}$$

即：若 p,q 中至少一个 $\equiv 1\pmod 4$，则 $\left(\frac{p}{q}\right)=\left(\frac{q}{p}\right)$；若 $p\equiv q\equiv 3\pmod 4$，则 $\left(\frac{p}{q}\right)=-\left(\frac{q}{p}\right)$。连同两个补充律 $\left(\frac{-1}{p}\right)=(-1)^{(p-1)/2}$、$\left(\frac{2}{p}\right)=(-1)^{(p^2-1)/8}$，它们决定了所有 Legendre 符号。

直观理解　「q 是不是模 p 的平方」与「p 是不是模 q 的平方」看上去是两个毫不相干的问题，一个在 $\mathbb{Z}/p$ 里，一个在 $\mathbb{Z}/q$ 里。互反律说它们的答案几乎相同，至多差一个只取决于 p,q 模 4 的符号。下面用格点计数证明这个关系；Gauss 把它称为「金定理」（theorema aureum）。

9.6.2　格点证明

思路　由 Eisenstein 引理（p,q 都是奇数），$\left(\frac{q}{p}\right)=(-1)^{t_1}$、$\left(\frac{p}{q}\right)=(-1)^{t_2}$，其中 t_1,t_2 是两个取整和。只需证明 $t_1+t_2=\frac{p-1}{2}\cdot\frac{q-1}{2}$。把 t_1,t_2 看成两个三角形中的格点数，它们恰好拼成一个 $\frac{p-1}{2}\times\frac{q-1}{2}$ 的格点矩形，而对角线上没有格点。

证明　记 $h=\frac{p-1}{2}$，$h'=\frac{q-1}{2}$，

$$t_1=\sum_{x=1}^{h}\lfloor\frac{qx}{p}\rfloor,\qquad t_2=\sum_{y=1}^{h'}\lfloor\frac{py}{q}\rfloor.$$

由 Eisenstein 引理，$\left(\frac{q}{p}\right)=(-1)^{t_1}$，$\left(\frac{p}{q}\right)=(-1)^{t_2}$。考虑格点矩形 $R=\{(x,y)\in\mathbb{Z}^2\mid 1\le x\le h,\ 1\le y\le h'\}$，共 hh' 个点，以及过原点的直线 $L:\ y=\frac{q}{p}x$。

(i) L 上没有 R 中的点：若 $py=qx$，则 $p\mid qx$，而 $p\nmid q$，故 $p\mid x$，与 $1\le x\le h<p$ 矛盾。

(ii) L 下方的点数是 t_1：竖线 x（$1\le x\le h$）上满足 $1\le y<\frac{qx}{p}$ 的整数 y 有 $\lfloor qx/p\rfloor$ 个。这些 y 都不超过 h'，因为 $\frac{qx}{p}\le\frac{qh}{p}<\frac{q}{2}$，所以 $y\le\frac{q-1}{2}$，确实落在 R 中。

(iii) L 上方的点数是 t_2：同理，横线 y（$1\le y\le h'$）上满足 $1\le x<\frac{py}{q}$ 的点有 $\lfloor py/q\rfloor$ 个，且都满足 $x\le h$。

由 (i)，R 的每个点恰在 L 的一侧，所以 $t_1+t_2=hh'$，

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)=(-1)^{t_1+t_2}=(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}.$$

■

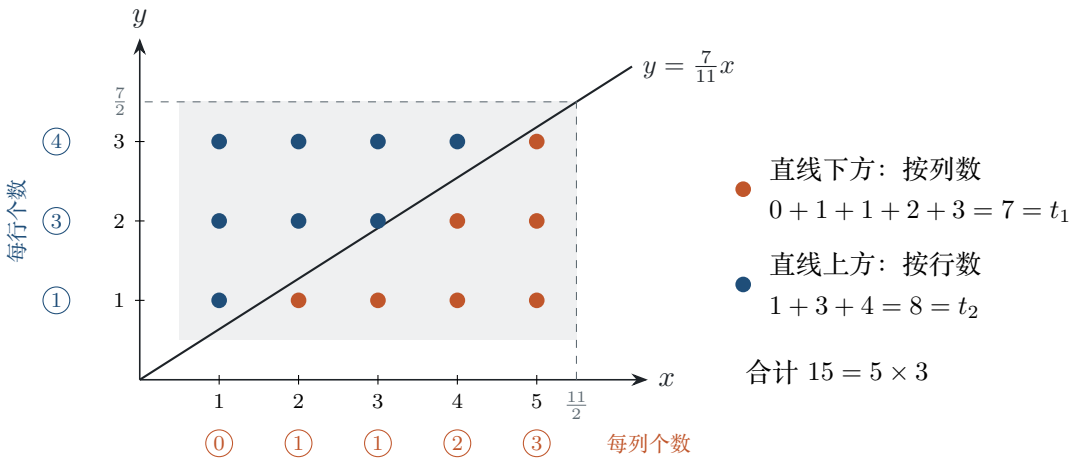


图 9.3　$p=11$、$q=7$ 时的格点矩形 $1\le x\le 5$、$1\le y\le 3$。直线 $y=\frac{7}{11}x$ 不经过任何格点，恰好穿过矩形的角 $\left(\frac{11}{2},\frac{7}{2}\right)$，把 15 个点分成下方 7 个和上方 8 个。

对照图 9.3：$t_1=7$ 是奇数，$\left(\frac{7}{11}\right)=-1$；$t_2=8$ 是偶数，$\left(\frac{11}{7}\right)=\left(\frac{4}{7}\right)=1$。两者之积 $-1=(-1)^{5\cdot3}$，符合互反律。与表 9.4 的第 7 行第 11 列、第 11 行第 7 列也一致。

9.6.3 用互反律计算

现在任何 Legendre 符号都能算了。算法和第 2 讲的欧几里得算法很像：**约化**、**拆分**、**翻转**，反复进行，每一轮分子分母都变小。

(1) 约化：用 $a \equiv b \Rightarrow \left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$，把分子换成模 p 的较小代表（可以取负的）。

(2) 拆分：把分子分解，分出 -1 和 2 的幂，用两个补充律处理；平方因子直接去掉。

(3) 翻转：对剩下的奇素数 q，用互反律把 $\left(\frac{q}{p}\right)$ 换成 $\pm\left(\frac{p}{q}\right)$，再回到第 1 步。

■ **例 9.8　两种算法**　(a) $x^2 \equiv 7 \pmod{227}$ 有没有解？（227 是素数。改编自习题集 5。）(b) 计算 $\left(\frac{-42}{61}\right)$。

解　(a) $7 \equiv 227 \equiv 3 \pmod 4$，翻转时要变号：

$$\left(\frac{7}{227}\right) = -\left(\frac{227}{7}\right) = -\left(\frac{3}{7}\right) = -\left(-\left(\frac{7}{3}\right)\right) = \left(\frac{1}{3}\right) = 1.$$

第二步用了 $227 = 32 \cdot 7 + 3$；第三步 $3 \equiv 7 \equiv 3 \pmod 4$，又变一次号；第四步 $7 \equiv 1 \pmod 3$。所以有两个解。实际上 $37^2 = 1369 = 6 \cdot 227 + 7$，两个解是 $x \equiv \pm 37$。

(b) 先拆分：$-42 = (-1) \cdot 2 \cdot 3 \cdot 7$，$61 \equiv 1 \pmod 4$，$61 \equiv 5 \pmod 8$。

$$\left(\frac{-1}{61}\right) = 1, \qquad \left(\frac{2}{61}\right) = -1, \qquad \left(\frac{3}{61}\right) = \left(\frac{61}{3}\right) = \left(\frac{1}{3}\right) = 1, \qquad \left(\frac{7}{61}\right) = \left(\frac{61}{7}\right) = \left(\frac{5}{7}\right).$$

后两个用了互反律，因为 $61 \equiv 1 \pmod 4$，不变号。最后 $\left(\frac{5}{7}\right) = \left(\frac{7}{5}\right) = \left(\frac{2}{5}\right) = -1$。所以

$$\left(\frac{-42}{61}\right) = 1 \cdot (-1) \cdot 1 \cdot (-1) = 1.$$

也可以先约化：$-42 \equiv 19 \pmod{61}$，而 19 是素数，$\left(\frac{19}{61}\right) = \left(\frac{61}{19}\right) = \left(\frac{4}{19}\right) = 1$，一步到位。验证：$18^2 = 324 = 5 \cdot 61 + 19$。先约化往往省事，但分子是合数时**必须先分解**再翻转。 ♦

翻转之前先检查　互反律只适用于两个**不同的奇素数**。常见的错误有三种：

- 对合数分子直接翻转，例如把 $\left(\frac{15}{p}\right)$ 写成 $\pm\left(\frac{p}{15}\right)$。$\left(\frac{p}{15}\right)$ 作为 Legendre 符号没有定义，要先拆成 $\left(\frac{3}{p}\right)\left(\frac{5}{p}\right)$。（第 10 讲的 Jacobi 符号会让这种写法合法，但那是另一个定理。）
- 对 2 用互反律。2 不是奇素数，$\left(\frac{2}{p}\right)$ 只能用第二补充律。
- 忘记变号。只有在两个素数**都** $\equiv 3 \pmod 4$ 时才变号，每次翻转前都要看一眼。

9.6.4 用同余条件刻画「a 是平方」的素数

互反律更深刻的用途是**固定分子、让分母变化**。问「3 是模哪些素数的平方」，原本要对每个 p 单独检验；翻转之后，分母变成固定的 3，答案只取决于 p 除以某个固定数的余数。

■ **例 9.9　3 何时是平方**　求所有使 $\left(\frac{3}{p}\right) = 1$ 的素数 $p \neq 2, 3$。

解　由互反律，$\left(\frac{3}{p}\right)=(-1)^{\frac{p-1}{2}\cdot\frac{3-1}{2}}\left(\frac{p}{3}\right)=(-1)^{\frac{p-1}{2}}\left(\frac{p}{3}\right)$。两个因子分别取决于 p 模 4 和模 3：

$$(-1)^{\frac{p-1}{2}}=\begin{cases}1, & p\equiv 1 \pmod 4,\\ -1, & p\equiv 3 \pmod 4,\end{cases}\qquad \left(\frac{p}{3}\right)=\begin{cases}1, & p\equiv 1 \pmod 3,\\ -1, & p\equiv 2 \pmod 3.\end{cases}$$

由中国剩余定理，p 模 4 与模 3 的余数合起来就是 p 模 12 的余数。与 12 互素的余数只有 $1,5,7,11$ 四个：

$p \bmod 12$	$p \bmod 4$	$p \bmod 3$	$(-1)^{(p-1)/2}\left(\frac{p}{3}\right)$	$\left(\frac{3}{p}\right)$
1	1	1	$(+1)(+1)$	$+1$
5	1	2	$(+1)(-1)$	-1
7	3	1	$(-1)(+1)$	-1
11	3	2	$(-1)(-1)$	$+1$

所以 **3 是模 p 的二次剩余当且仅当 $p\equiv\pm1 \pmod{12}$**。检验：$5^2\equiv 3 \pmod{11}$，$4^2\equiv 3 \pmod{13}$，$7^2\equiv 3 \pmod{23}$，$15^2=6\cdot 37+3$；而 $p=5,7$ 时表 9.1 中 3 不是剩余。 ♦

想一想　同样的方法求使 $\left(\frac{5}{p}\right)=1$ 的素数 $p\neq 2,5$。为什么这次只需要看 p 模 5，不用看模 4？

答　$5\equiv 1 \pmod 4$，互反律不带符号：$\left(\frac{5}{p}\right)=\left(\frac{p}{5}\right)$。模 5 的剩余是 $1,4$，所以 $\left(\frac{5}{p}\right)=1$ 当且仅当 $p\equiv\pm1 \pmod 5$（例如 $4^2\equiv 5 \pmod{11}$，$9^2\equiv 5 \pmod{19}$）。而 $3\equiv 3 \pmod 4$，翻转时的符号依赖 p 模 4，所以上例要看模 12。

核心思想　固定整数 $a\neq 0$。对不整除 $2a$ 的素数 p，$\left(\frac{a}{p}\right)$ 只取决于 p 模 $4|a|$ 的余数。原因是：把 a 分解后，$\left(\frac{-1}{p}\right)$ 只看 $p \bmod 4$，$\left(\frac{2}{p}\right)$ 只看 $p \bmod 8$，每个奇素因子 q 翻转后 $\left(\frac{q}{p}\right)=\pm\left(\frac{p}{q}\right)$ 只看 $p \bmod 4$ 和 $p \bmod q$。看起来毫无规律的条件变成了一组同余条件。

「a 是模 p 的平方」还可以反过来用：若 $p\mid x^2-a$，则 $\left(\frac{a}{p}\right)=1$，从而 p 落在特定的同余类中。例 9.4 就是这样证明 $4k+1$ 型素数有无穷多的，练习里还有 $8k+7$ 型和 $3k+1$ 型的版本。

历史与背景　Euler 在 1740 年代从数值例子中发现了互反律，但没能证明。Legendre 1785 年的证明假设了尚未证明的「算术级数中有无穷多素数」，并在 1798 年引入了符号 $\left(\frac{a}{p}\right)$。第一个完整证明由 18 岁的 Gauss 在 1796 年给出（1801 年发表）。本讲的证明基于 Gauss 的第三个证明（1808 年），格点写法来自 Eisenstein（1844 年）。如今已知的证明数以百计。

前后联系　第 10 讲把分母推广到奇合数（Jacobi 符号），并介绍 Zolotareff 的置换定义。第 11 讲给出求平方根的 Tonelli 算法。把「二次」换成高次，就通向三次、四次互反律和类域论。

本讲小结

- 配方把二次同余式化为 $y^2 \equiv D$；$x^2 \equiv a \pmod p$ 恰有 $1+\left(\frac{a}{p}\right)$ 个解。剩余与非剩余各占一半，剩余就是原根的偶数次幂。
- Euler 判别法 $\left(\frac{a}{p}\right) \equiv a^{(p-1)/2}$ 给出完全积性和 $\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$。
- Gauss 引理 $\left(\frac{a}{p}\right) = (-1)^n$ 给出 $\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8}$；Eisenstein 引理把 n 的奇偶换成格点数 $t = \sum \lfloor ka/p \rfloor$。
- 二次互反律 $\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}$：两个三角形拼成矩形。计算时「约化、拆分、翻转」；$\left(\frac{a}{p}\right)$ 只取决于 $p \bmod 4|a|$。

练习

9.1 ★ 列出模 17 和模 19 的全部二次剩余，并用 Euler 判别法验证：3 是模 17 的非剩余，5 是模 19 的剩余。

9.2 ★ 用 Gauss 引理计算 $\left(\frac{3}{11}\right)$，用 Eisenstein 引理计算 $\left(\frac{7}{13}\right)$，再与平方表核对。

9.3 ★ 计算 $\left(\frac{-2}{59}\right)$、$\left(\frac{30}{37}\right)$、$\left(\frac{-105}{229}\right)$（$59, 37, 229$ 都是素数）。

9.4 ★★ 下列同余式各有几个解？(a) $x^2 \equiv -2 \pmod{118}$；(b) $x^2 \equiv -1 \pmod{244}$；(c) $x^2 \equiv -1 \pmod{365}$；(d) $x^2 \equiv 267 \pmod{789}$。（改编自习题集 5。提示：先分解模数。）

9.5 ★★ 用同余条件刻画使 $\left(\frac{10}{p}\right) = 1$ 的素数 p。（习题集 5）

9.6 ★★ (a) 证明：对素数 $p > 3$，-3 是模 p 的二次剩余当且仅当 $p \equiv 1 \pmod 3$。(b) 由此证明形如 $3k+1$ 的素数有无穷多个。(c) 证明形如 $3k-1$ 的素数也有无穷多个。（习题集 5 第 8 题）

9.7 ★★ 证明形如 $8k+7$ 的素数有无穷多个。（习题集 5。提示：$N = (p_1 \cdots p_n)^2 - 2$。）

9.8 ★★★ 证明：对每个素数 p，$x^8 \equiv 16 \pmod p$ 都有解。（习题集 5。提示：$-1, 2, -2$ 中至少有一个是模 p 的平方。）

9.9 ★★ 对奇素数 p，证明 $\sum_{a=1}^{p-1}\left(\frac{a}{p}\right) = 0$，且模 p 的全部二次剩余之积 $\equiv (-1)^{(p+1)/2}$。

9.10 ★★★ 设 p 是奇素数，$p \nmid k$。(a) 证明 $y^2 \equiv x^2 + k \pmod p$ 恰有 $p-1$ 组解 (x, y)。(b) 由此证明 $\sum_{x=0}^{p-1}\left(\frac{x^2+k}{p}\right) = -1$。(c) 设 $p \nmid ab$，证明 $ax^2 + by^2 \equiv 1 \pmod p$ 恰有 $p - \left(\frac{-ab}{p}\right)$ 组有序解 (x, y)。（习题集 5 第 9 题）

练习答案与提示

9.1 模 17 的剩余：$\{1, 2, 4, 8, 9, 13, 15, 16\}$；$3^4 = 81 \equiv -4$，$3^8 \equiv 16 \equiv -1$，故 3 是非剩余。模 19 的剩余：$\{1, 4, 5, 6, 7, 9, 11, 16, 17\}$；$5^4 \equiv 36 \equiv -2$，$5^8 \equiv 4$，$5^9 \equiv 20 \equiv 1$，故 5 是剩余（$9^2 = 81 \equiv 5$）。

9.2 $(3k)_{11}$（$k = 1, \ldots, 5$）为 $3, -5, -2, 1, 4$，$n = 2$，$\left(\frac{3}{11}\right) = 1$；确实 $5^2 \equiv 3$。$\lfloor 7k/13 \rfloor$（$k = 1, \ldots, 6$）为 $0, 1, 1, 2, 2, 3$，$t = 9$ 是奇数，$\left(\frac{7}{13}\right) = -1$；表 9.1 中 7 不是模 13 的剩余。

9.3 $59 \equiv 3 \pmod 8$，$\left(\frac{-2}{59}\right) = \left(\frac{-1}{59}\right)\left(\frac{2}{59}\right) = (-1)(-1) = 1$。$37 \equiv 5 \pmod 8$ 且 $37 \equiv 1 \pmod{12}$，$\left(\frac{30}{37}\right) =$

$\left(\frac{2}{37}\right)\left(\frac{3}{37}\right)\left(\frac{5}{37}\right)=(-1)\cdot 1\cdot\left(\frac{2}{5}\right)=1$。$229\equiv 1 \pmod 4$，$\left(\frac{-1}{229}\right)=1$，翻转不变号：$\left(\frac{3}{229}\right)=\left(\frac{1}{3}\right)=1$，$\left(\frac{5}{229}\right)=\left(\frac{4}{5}\right)=1$，$\left(\frac{7}{229}\right)=\left(\frac{5}{7}\right)=\left(\frac{2}{5}\right)=-1$，乘积为 -1。

9.4 (a) $118=2\cdot 59$：模 2 有 1 个解，模 59 有 2 个（练习 3），共 2 个。(b) $244=4\cdot 61$，而 $x^2\equiv -1\equiv 3 \pmod 4$ 无解，共 0 个。(c) $365=5\cdot 73$，$5\equiv 73\equiv 1 \pmod 4$，各 2 个，共 4 个。(d) $789=3\cdot 263$，$267\equiv 0 \pmod 3$ 有 1 个解，$267\equiv 4=2^2 \pmod{263}$ 有 2 个，共 2 个。

9.5 $p\neq 2,5$。$\left(\frac{10}{p}\right)=\left(\frac{2}{p}\right)\left(\frac{p}{5}\right)$，前者看 $p \bmod 8$，后者为 1 当且仅当 $p\equiv\pm 1 \pmod 5$。逐一合并两个条件，乘积为 $+1$ 的类是

$$p\equiv \pm 1,\ \pm 3,\ \pm 9,\ \pm 13 \pmod{40},$$

即 $1,3,9,13,27,31,37,39$；其余允许类 $\pm 7,\pm 11,\pm 17,\pm 19$ 给出 -1。这一步不能只保留「两个因子同号」的四个类，因为乘积为 -1 的两种异号情形也会合成为 $+1$。

9.6 (a) $\left(\frac{-3}{p}\right)=\left(\frac{-1}{p}\right)\left(\frac{3}{p}\right)=(-1)^{\frac{p-1}{2}}\cdot(-1)^{\frac{p-1}{2}}\left(\frac{p}{3}\right)=\left(\frac{p}{3}\right)$，为 1 当且仅当 $p\equiv 1 \pmod 3$。(b) 设只有 $p_1,\dots,p_n$，令 $N=(2p_1\cdots p_n)^2+3$。N 是奇数且 $N\equiv 1 \pmod 3$，所以素因子 $q\neq 2,3$。由 $(2p_1\cdots p_n)^2\equiv -3 \pmod q$ 和 (a)，$q\equiv 1 \pmod 3$，于是 $q=p_i$；但 $p_i\mid N$ 推出 $p_i\mid 3$，矛盾。

设形如 $3k-1$ 的素数只有 $p_1=2,p_2,\dots,p_n$。令 $P=p_1\cdots p_n$，$N=3P-1$。因为 P 是偶数，$N>1$ 且为奇数；又 $N\equiv -1 \pmod 3$。素因子都不是 3，按重数计的素因子在模 3 下各余 ± 1，而它们的乘积为 -1，故至少有一个素因子 $q\equiv -1 \pmod 3$。它是奇素数，因而在假设列表中，$q\mid P$。但 $q\mid N=3P-1$ 也推出 $q\mid 1$，矛盾。具体地，列表 $\{2\}$ 给出 $N=5$，列表 $\{2,5\}$ 给出 $N=29$，两者都是新的 $3k-1$ 型素数。

9.7 设形如 $8k+7$ 的素数只有 $p_1=7,\dots,p_n$，令 $P=p_1\cdots p_n$、$N=P^2-2$。P 为奇数，故 $N>1$ 且 $N\equiv 7 \pmod 8$。若素数 $q\mid N$，则 q 为奇数且 $q\nmid P$，因为同时整除 P 与 N 会推出 $q\mid 2$。于是 $P^2\equiv 2 \pmod q$，故 $\left(\frac{2}{q}\right)=1$，由 2 的判别式 $q\equiv\pm 1 \pmod 8$。把 N 完全分解为 $N=\prod_{j=1}^{r}q_j^{e_j}$，其中 q_j 互异。模 8 有

$$-1\equiv N\equiv\prod_{j=1}^{r}q_j^{e_j}\equiv(-1)^{\sum_{q_j\equiv 7\ (8)}e_j} \pmod 8.$$

因此指数和为奇数，存在某个 $q_j\equiv 7 \pmod 8$ 且 e_j 为奇数。这样的 q_j 必是列表中的某个 p_i，于是 $q_j\mid P$ 且 $q_j\mid N=P^2-2$，推出 $q_j\mid 2$，与 q_j 为奇素数矛盾。

9.8 $p=2$ 时 $16\equiv 0 \pmod 2$，取 $x=0$ 即可。设 p 为奇素数，$\left(\frac{-1}{p}\right)\left(\frac{2}{p}\right)\left(\frac{-2}{p}\right)=\left(\left(\frac{-1}{p}\right)\left(\frac{2}{p}\right)\right)^2=1$，三者不能全为 -1。分解 $x^8-16=(x^2-2)(x^2+2)\big((x+1)^2+1\big)\big((x-1)^2+1\big)$（用 $x^4+4=(x^2+2)^2-(2x)^2$）。若 $\pm 2\equiv c^2$，取 $x=c$；若 $-1\equiv c^2$，取 $x=c-1$。

9.9 剩余与非剩余各 $\frac{p-1}{2}$ 个，所以和为 0。二次剩余恰是互不同余的 $1^2,\dots,h^2$（$h=\frac{p-1}{2}$），积为 $(h!)^2\equiv(-1)^{(p+1)/2}$（例 9.3）。例如 $p=7$：$1\cdot 2\cdot 4=8\equiv 1$。

9.10 (a) 令 $z=y-x$，$w=y+x$。2 可逆，所以 $(x,y)\mapsto(z,w)$ 是双射，方程变为 $zw\equiv k$：每个非零 z 对应唯一的 $w\equiv kz^{-1}$。(b) 固定 x，y 的个数是 $1+\left(\frac{x^2+k}{p}\right)$。对 x 求和得 $p-1=p+\sum_x\left(\frac{x^2+k}{p}\right)$。例如 $p=7,k=1$ 时，x^2+1 的剩余类依次是 $1,2,5,3,3,5,2$，Legendre 符号为 $1,1,-1,-1,-1,-1,1$，和为 -1。

(c) 记解对数为 M。固定 x，关于 y 的同余式等价于 $y^2\equiv(1-ax^2)b^{-1} \pmod p$，故有 $1+\left(\frac{(1-ax^2)b^{-1}}{p}\right)$ 个解。对 $x=0,\dots,p-1$ 求和，并用 $\left(\frac{b^{-1}}{p}\right)=\left(\frac{b}{p}\right)$ 及 $1-ax^2\equiv(-a)(x^2-a^{-1})$，得到

$$M=p+\left(\frac{-ab}{p}\right)\sum_{x=0}^{p-1}\left(\frac{x^2-a^{-1}}{p}\right).$$

由 (b)（取 $k=-a^{-1}$），和为 -1，所以 $M=p-\left(\frac{-ab}{p}\right)$。例如 $p=7,a=b=1$ 时 $\left(\frac{-1}{7}\right)=-1$，公

式给出 8 组；逐类看，$x=0$ 有 $y=\pm 1$，$x=\pm 1$ 有 $y=0$，$x=\pm 2$ 各有 $y=\pm 2$，$x=\pm 3$ 无解，共 $2+2+4=8$ 组。

第 10 讲

Jacobi 符号与 Zolotareff 定理

Jacobi Symbol, Computation, Zolotareff's Definition

对应材料：MIT 18.781 Lecture 10 讲义（Jacobi Symbol, Computation, Zolotareff's Definition）；习题集 5 第 11、12 题；期中 2 练习题第 4、5 题。

本讲目标

(1) 熟练运用 Legendre 符号的运算规则，并看清它的瓶颈：每次用互反律翻转之前，都要先分解分子。

(2) 掌握 Jacobi 符号的定义与乘法性；知道 $\left(\frac{P}{Q}\right)=-1$ 说明 P 不是模 Q 的平方，而 $\left(\frac{P}{Q}\right)=1$ 单凭这一点不能确定平方性。

(3) 理解 Jacobi 符号的互反律和两个补充律为什么能从 Legendre 符号「继承」下来，并用它们在对数步数内算出任意 Legendre 符号，全程不做因数分解。

(4) 会写置换的轮换分解，掌握置换的符号：逆序数、乘法性、k 轮换的符号是 $(-1)^{k-1}$。

(5) 理解 Zolotareff 定理：$\left(\frac{P}{Q}\right)$ 等于「乘以 P」这个置换的符号；并用它重新得到补充律和互反律。

10.1 回顾：Legendre 符号怎么算

10.1.1 小素数的平方表

先看几个具体的数。把 $x=1,2,\dots,\frac{p-1}{2}$ 平方，再取绝对最小剩余（介于 $-\frac{p}{2}$ 与 $\frac{p}{2}$ 之间的代表），就得到模 p 的全部二次剩余：

$$
\begin{aligned}
p=5:&\quad 1,\ \ 4\equiv -1;\\
p=7:&\quad 1,\ \ 4\equiv -3,\ \ 9\equiv 2;\\
p=11:&\quad 1,\ \ 4,\ \ 9\equiv -2,\ \ 16\equiv 5,\ \ 25\equiv 3;\\
p=13:&\quad 1,\ \ 4,\ \ 9\equiv -4,\ \ 16\equiv 3,\ \ 25\equiv -1,\ \ 36\equiv -3.
\end{aligned}
$$

想一想 模 13 的剩余表关于正负对称：±1、±3、±4 成对出现。模 11 却不对称：-2 是剩余，2 不是；$3,4,5$ 是剩余，$-3,-4,-5$ 都不是。为什么？

答 由乘法性，$\left(\frac{-a}{p}\right)=\left(\frac{-1}{p}\right)\left(\frac{a}{p}\right)$。$13\equiv 1 \pmod 4$，$\left(\frac{-1}{13}\right)=1$，所以 a 与 $-a$ 同为剩余或同为非剩余，表是对称的。$11\equiv 3 \pmod 4$，$\left(\frac{-1}{11}\right)=-1$，$a$ 与 $-a$ 恰有一个是剩余：剩余 $\{1,-2,3,4,5\}$ 取负号，正好得到全部非剩余。

10.1.2 运算规则

第 9 讲对奇素数 p 定义了 **Legendre 符号**（Legendre symbol）：$p \mid a$ 时 $\left(\frac{a}{p}\right)=0$；$a$ 是模 p 的二次剩余时为 1；是非剩余时为 -1。计算它只需要下面五条。

命题（第 9 讲）　设 p, q 是不同的奇素数。

(1) 约化：$a \equiv b \pmod p \Rightarrow \left(\frac{a}{p}\right)=\left(\frac{b}{p}\right)$。

(2) 完全积性：$\left(\frac{ab}{p}\right)=\left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$。特别地，与 p 互素的平方因子可以直接去掉。

(3) 第一补充律：$\left(\frac{-1}{p}\right)=(-1)^{(p-1)/2}$，即 $p \equiv 1 \pmod 4$ 时为 1，$p \equiv 3 \pmod 4$ 时为 -1。

(4) 第二补充律：$\left(\frac{2}{p}\right)=(-1)^{(p^2-1)/8}$，即 $p \equiv \pm 1 \pmod 8$ 时为 1，$p \equiv \pm 3 \pmod 8$ 时为 -1。

(5) 互反律：$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)=(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}$，即只有 $p \equiv q \equiv 3 \pmod 4$ 时翻转要变号。

乘法性在原根的坐标下最透明。取 $p=7$，$g=3$ 是原根，它的幂依次是

$$3^1=3,\quad 3^2\equiv 2,\quad 3^3\equiv 6,\quad 3^4\equiv 4,\quad 3^5\equiv 5,\quad 3^6\equiv 1 \pmod 7.$$

偶数次幂 $2,4,1$ 正好是上表中的剩余 $1,2,4$（即 $1,-3,2$）。一般地，若 $a \equiv g^k$，则 $\left(\frac{a}{p}\right)=(-1)^k$（第 9 讲）。两个数相乘时指数相加，指数的奇偶按「偶 + 偶 = 偶，偶 + 奇 = 奇，奇 + 奇 = 偶」相加，这就是「剩余乘剩余是剩余，剩余乘非剩余是非剩余，非剩余乘非剩余是剩余」。

■ **例 10.1　两种算法**　计算 $\left(\frac{17}{31}\right)$。（17、31 都是素数。）

解　方法一：先翻转。$17 \equiv 1 \pmod 4$，翻转不变号：

$$\left(\frac{17}{31}\right)=\left(\frac{31}{17}\right)=\left(\frac{-3}{17}\right)=\left(\frac{-1}{17}\right)\left(\frac{3}{17}\right)=\left(\frac{3}{17}\right)=\left(\frac{17}{3}\right)=\left(\frac{2}{3}\right)=-1.$$

这里用了 $31=17+14\equiv -3 \pmod{17}$、$\left(\frac{-1}{17}\right)=1$（$17\equiv 1 \pmod 4$），以及 $17\equiv 2 \pmod 3$、$3\equiv 3 \pmod 8$。

方法二：先约化。$17 \equiv 17+31=48=4^2\cdot 3 \pmod{31}$，平方因子 4^2 去掉：

$$\left(\frac{17}{31}\right)=\left(\frac{3}{31}\right)=-\left(\frac{31}{3}\right)=-\left(\frac{1}{3}\right)=-1.$$

这次 $3\equiv 31\equiv 3 \pmod 4$，翻转要变号。两种方法结果一致，17 不是模 31 的平方。检验：模 31 的平方是 $1,2,4,5,7,8,9,10,14,16,18,19,20,25,28$，其中没有 17。 ♦

注 10.1　MIT 讲义对方法一的记录是「$\left(\frac{17}{31}\right)=\left(\frac{31}{17}\right)=\left(\frac{3}{17}\right)=-\left(\frac{17}{3}\right)=-\left(\frac{1}{3}\right)=-1$」。第二个等号省略了经过 -3 的一步；后两个等号各有一处笔误：$17\equiv 1 \pmod 4$，翻转不应变号；$17\equiv 2 \pmod 3$，不能换成 1。两处错误恰好抵消，最后的 -1 是对的。这类计算每一步都要核对模 4、模 8 的余数。

10.1.3 瓶颈：翻转之前必须分解

上面的计算之所以顺利，是因为每次要翻转的分子都是素数。再看一个分子会变成合数的例子。

■ 例 10.2　只用 Legendre 符号　计算 $\left(\frac{127}{233}\right)$。（127、233 都是素数，$127 \equiv 3$，$233 \equiv 1 \pmod 4$。）

解　$233 \equiv 1 \pmod 4$，翻转不变号；再约化，$233 = 127 + 106$，$106 \equiv -21 \pmod{127}$：

$$\left(\frac{127}{233}\right) = \left(\frac{233}{127}\right) = \left(\frac{-21}{127}\right) = \left(\frac{-1}{127}\right)\left(\frac{21}{127}\right) = -\left(\frac{21}{127}\right).$$

21 不是素数，Legendre 符号的互反律不能直接用，必须先分解 $21 = 3 \cdot 7$，再分别翻转。$3 \equiv 7 \equiv 127 \equiv 3 \pmod 4$，两次都变号：

$$\left(\frac{3}{127}\right) = -\left(\frac{127}{3}\right) = -\left(\frac{1}{3}\right) = -1, \qquad \left(\frac{7}{127}\right) = -\left(\frac{127}{7}\right) = -\left(\frac{1}{7}\right) = -1.$$

这里 $127 = 42 \cdot 3 + 1 = 18 \cdot 7 + 1$。所以 $\left(\frac{21}{127}\right) = (-1)(-1) = 1$，$\left(\frac{127}{233}\right) = -1$，127 不是模 233 的平方。 ♦

分解 21 不费力，但若每次翻转都依赖因数分解，就没有得到高效算法：某些由两个大素数相乘得到的两百位合数，用现有经典通用算法分解可能很困难（第 6 讲）。可是「a 是不是模 p 的平方」这个问题本身并不难：用 Euler 判别法 $\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod p$ 和反复平方，几百次乘法就能回答。所以只靠 Legendre 符号的「约化、拆分、翻转」在算法上是有缺陷的。我们希望对合数直接翻转，像欧几里得算法那样一路做带余除法，永远不分解。下一节的 Jacobi 符号正好做到这一点。

10.2 Jacobi 符号

10.2.1 定义

定义 10.1（Jacobi 符号）　设 Q 是正奇数，$Q = q_1 q_2 \cdots q_k$ 是它的素因数分解（q_i 允许重复），P 是任意整数。定义 **Jacobi 符号**（Jacobi symbol）

$$\left(\frac{P}{Q}\right) = \left(\frac{P}{q_1}\right)\left(\frac{P}{q_2}\right)\cdots\left(\frac{P}{q_k}\right),$$

右边每个因子都是 Legendre 符号。约定 $\left(\frac{P}{1}\right) = 1$（空乘积）。

Q 是奇素数时，Jacobi 符号就是原来的 Legendre 符号，所以沿用同一个记号不会产生歧义。只要 $\gcd(P, Q) > 1$，就有某个 $q_i \mid P$，对应因子为 0，整个符号为 0；反之 $\gcd(P, Q) = 1$ 时每个因子都是 ± 1，符号也是 ± 1。

■ 例 10.3　按定义计算　计算 $\left(\frac{-1}{77}\right)$、$\left(\frac{2}{15}\right)$、$\left(\frac{7}{45}\right)$。

解　$77 = 7 \cdot 11$，$7 \equiv 11 \equiv 3 \pmod 4$，所以 $\left(\frac{-1}{77}\right) = \left(\frac{-1}{7}\right)\left(\frac{-1}{11}\right) = (-1)(-1) = 1$。

$15 = 3\cdot 5$，$3 \equiv 5 \equiv \pm 3 \pmod 8$，所以 $\left(\frac{2}{15}\right) = \left(\frac{2}{3}\right)\left(\frac{2}{5}\right) = (-1)(-1) = 1$。

$45 = 3\cdot 3\cdot 5$，素因子 3 出现两次：$\left(\frac{7}{45}\right) = \left(\frac{7}{3}\right)^2\left(\frac{7}{5}\right) = \left(\frac{7}{5}\right) = \left(\frac{2}{5}\right) = -1$。 ♦

前两个结果都是 1，但 -1 不是模 77 的平方（它连模 7 的平方都不是），2 也不是模 15 的平方（模 15 的平方只有 $0,1,4,6,9,10$）。第三个结果是 -1，而 7 确实不是模 45 的平方。这两种情况的区别可以说清楚。

命题 10.2　设 Q 是正奇数，$\gcd(P,Q) = 1$。若 P 是模 Q 的平方，则 $\left(\frac{P}{Q}\right) = 1$。等价地，$\left(\frac{P}{Q}\right) = -1$ 时 P 一定不是模 Q 的平方。

证明　若 $P \equiv x^2 \pmod Q$，则对每个素因子 $q_i \mid Q$ 都有 $P \equiv x^2 \pmod{q_i}$。又 $q_i \nmid P$，所以 $\left(\frac{P}{q_i}\right) = 1$，乘积也是 1。 ■

反方向不成立，因为两个 -1 相乘也是 1。由中国剩余定理（第 5 讲）和 Hensel 引理（第 6、7 讲），真正的判别法是：$\gcd(P,Q) = 1$ 时，P 是模 Q 的平方当且仅当对 Q 的每一个素因子 q 都有 $\left(\frac{P}{q}\right) = 1$。Jacobi 符号只记录了这些 ± 1 的乘积，丢掉了每个因子各自的值。图 10.1 把模 15 的情形全部列了出来。

a	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\left(\frac{a}{3}\right)$	+1	−1	0	+1	−1	0	+1	−1	0	+1	−1	0	+1	−1
$\left(\frac{a}{5}\right)$	+1	−1	−1	+1	0	+1	−1	−1	+1	0	+1	−1	−1	+1
$\left(\frac{a}{15}\right)$	+1	+1	0	+1	0	0	−1	+1	0	0	−1	0	−1	−1
是平方?	是	否	否	是	否	是	否	否	是	是	否	否	否	否

图 10.1　模 15 的 Jacobi 符号。与 15 互素的 8 个 a 中，$\left(\frac{a}{15}\right) = 1$ 的有 $1,2,4,8$ 四个，真正的平方只有 $1,4$（它们在模 3、模 5 下都是剩余）。红框中的 $2,8$ 模 3、模 5 都是非剩余，两个 -1 相乘得到 1。

Jacobi 符号不是「是否为平方」

- $\left(\frac{P}{Q}\right) = 1$ 推不出 P 是模 Q 的平方，例如 $\left(\frac{2}{15}\right) = 1$。只有 $\left(\frac{P}{Q}\right) = -1$ 是确定的结论（命题 10.2）。
- Euler 判别法对合数分母不成立：$\left(\frac{2}{15}\right) = 1$，但 $2^{(15-1)/2} = 2^7 = 128 \equiv 8 \pmod{15}$，连 ± 1 都不是。这个「缺陷」在 §10.4 的素性检验中反而成了工具。

10.2.2　乘法性

Jacobi 符号由 Legendre 符号相乘得到，Legendre 符号的乘法规则就逐个因子地传了过来。

命题 10.3　设 Q, Q' 是正奇数，P, P' 是整数。

(1) $P \equiv P' \pmod{Q} \Rightarrow \left(\frac{P}{Q}\right) = \left(\frac{P'}{Q}\right)$。

(2) 对分子积性：$\left(\frac{PP'}{Q}\right) = \left(\frac{P}{Q}\right)\left(\frac{P'}{Q}\right)$。

(3) 对分母积性：$\left(\frac{P}{QQ'}\right) = \left(\frac{P}{Q}\right)\left(\frac{P}{Q'}\right)$。

(4) $\gcd(P, Q) = 1$ 时 $\left(\frac{P^2}{Q}\right) = \left(\frac{P}{Q^2}\right) = 1$，所以分子、分母中的平方因子都可以去掉。

证明　设 $Q = q_1 \cdots q_k$。(1) $P \equiv P' \pmod{Q}$ 推出 $P \equiv P' \pmod{q_i}$，每个因子不变。(2) 每个因子 $\left(\frac{PP'}{q_i}\right) = \left(\frac{P}{q_i}\right)\left(\frac{P'}{q_i}\right)$，再相乘。(3) QQ' 的素因数分解就是把 Q 和 Q' 的分解拼在一起，这正是定义。(4) 由 (2)(3)，$\left(\frac{P^2}{Q}\right) = \left(\frac{P}{Q}\right)^2$，$\left(\frac{P}{Q^2}\right) = \left(\frac{P}{Q}\right)^2$，而 $\left(\frac{P}{Q}\right) = \pm 1$。 ■

想一想　取 $Q = 9$。对哪些 P 有 $\left(\frac{P}{9}\right) = 1$？它们都是模 9 的平方吗？

答　由定义 $\left(\frac{P}{9}\right) = \left(\frac{P}{3}\right)^2$，与 3 互素时恒为 1。可是模 9 的单位 $1, 2, 4, 5, 7, 8$ 平方后只得到 $1, 4, 7$，一半不是平方。分母是完全平方数时，Jacobi 符号对互素的分子恒等于 1，不提供任何信息。分母不是完全平方数时情况不同：与 Q 互素的剩余类中恰有一半使符号为 1（可选一个 Jacobi 值为 -1 的单位相乘，将两类一一配对），图 10.1 中是 8 个里的 4 个。

表 10.1 对比两种符号。

表 10.1　Legendre 符号与 Jacobi 符号

	Legendre 符号 $\left(\frac{a}{p}\right)$	Jacobi 符号 $\left(\frac{P}{Q}\right)$
分母	奇素数 p	正奇数 Q
值为 0	$p \mid a$	$\gcd(P, Q) > 1$
值为 -1	a 不是平方	P 不是平方
值为 1	a 是非零平方	不能下结论
Euler 判别法	$\left(\frac{a}{p}\right) \equiv a^{(p-1)/2}$	一般不成立
两个补充律、互反律	成立	成立（定理 10.4）
翻转前要分解分子吗	要	不要

10.3　Jacobi 符号的互反律

10.3.1　定理与证明

表 10.1 中最重要的一行是：Legendre 符号的三条「翻转规则」对 Jacobi 符号原样成立。

定理 10.4（Jacobi 符号的互反律）　设 P, Q 是正奇数。

(1) $\left(\frac{-1}{Q}\right) = (-1)^{(Q-1)/2}$。

(2) $\left(\frac{2}{Q}\right) = (-1)^{(Q^2-1)/8}$。

(3) 若 $\gcd(P,Q) = 1$，则 $\left(\frac{P}{Q}\right)\left(\frac{Q}{P}\right) = (-1)^{\frac{P-1}{2}\cdot\frac{Q-1}{2}}$。也就是说，$P \equiv Q \equiv 3 \pmod 4$ 时 $\left(\frac{P}{Q}\right) = -\left(\frac{Q}{P}\right)$，否则 $\left(\frac{P}{Q}\right) = \left(\frac{Q}{P}\right)$。

思路　两边都拆成 Legendre 符号。左边按定义是若干个 Legendre 符号之积，每个因子可以用第 9 讲的结论算出，得到 $(-1)^{\text{某个和}}$。剩下的问题是：这个指数和与右边的指数 $\frac{Q-1}{2}$（或 $\frac{Q^2-1}{8}$）奇偶性相同吗？这归结为下面关于奇数的一个算术引理。

引理 10.5　对奇数 u, v，

$$\frac{uv-1}{2} \equiv \frac{u-1}{2} + \frac{v-1}{2} \pmod 2, \qquad \frac{u^2v^2-1}{8} \equiv \frac{u^2-1}{8} + \frac{v^2-1}{8} \pmod 2.$$

证明　直接相减：$\frac{uv-1}{2} - \frac{u-1}{2} - \frac{v-1}{2} = \frac{(u-1)(v-1)}{2}$。$u-1$、$v-1$ 都是偶数，乘积是 4 的倍数，除以 2 后仍是偶数。同理 $\frac{u^2v^2-1}{8} - \frac{u^2-1}{8} - \frac{v^2-1}{8} = \frac{(u^2-1)(v^2-1)}{8}$，而奇数的平方减 1 是 8 的倍数，乘积是 64 的倍数，除以 8 后是偶数。 ■

定理 10.4 的证明　记 $\varepsilon(u) = \frac{u-1}{2}$，$\omega(u) = \frac{u^2-1}{8}$。对因子个数归纳，引理 10.5 推广为：对奇数 $u_1, \dots, u_k$，

$$\varepsilon(u_1 \cdots u_k) \equiv \varepsilon(u_1) + \cdots + \varepsilon(u_k), \qquad \omega(u_1 \cdots u_k) \equiv \omega(u_1) + \cdots + \omega(u_k) \pmod 2.$$

设 $Q = q_1 \cdots q_k$。(1) 由第一补充律，$\left(\frac{-1}{Q}\right) = \prod_j \left(\frac{-1}{q_j}\right) = (-1)^{\sum_j \varepsilon(q_j)} = (-1)^{\varepsilon(Q)}$。(2) 同理，用第二补充律和 ω。

(3) 再设 $P = p_1 \cdots p_l$。$\gcd(P,Q) = 1$，所以每个 p_i 都不等于任何 q_j。由定义和 Legendre 符号的积性，

$$\begin{aligned}\left(\frac{P}{Q}\right)\left(\frac{Q}{P}\right) &= \prod_j \prod_i \left(\frac{p_i}{q_j}\right) \cdot \prod_i \prod_j \left(\frac{q_j}{p_i}\right) \\ &= \prod_{i,j} \left(\frac{p_i}{q_j}\right)\left(\frac{q_j}{p_i}\right) = \prod_{i,j} (-1)^{\varepsilon(p_i)\varepsilon(q_j)}.\end{aligned}$$

最后一步对每一对不同的奇素数用了互反律。由引理 10.5，右边的指数满足

$$\sum_{i,j} \varepsilon(p_i)\varepsilon(q_j) = \Big(\sum_i \varepsilon(p_i)\Big)\Big(\sum_j \varepsilon(q_j)\Big) \equiv \varepsilon(P)\,\varepsilon(Q) \pmod 2.$$

■

核心思想　证明只用到两件事：两个符号都对分子、分母分别积性；$(-1)^{\varepsilon(u)}$ 和 $(-1)^{\omega(u)}$ 作为奇数 u 的函数也是积性的（它们只看 u 模 4、模 8 的余数）。两边都是积性的，在素数上相等，在所有正奇数上就相等。互反律由此从 Legendre 符号「继承」给了 Jacobi 符号。

10.3.2 用 Jacobi 互反律计算

■ **例 10.4　不分解地翻转**　用 Jacobi 符号重新计算例 10.2 的 $\left(\frac{127}{233}\right)$。

解　前两步与例 10.2 相同。到 $\left(\frac{21}{127}\right)$ 时，$21 \equiv 1 \pmod 4$，由定理 10.4(3) 直接翻转，不变号，也不用分解 21：

$$\left(\frac{127}{233}\right) = \left(\frac{233}{127}\right) = \left(\frac{-21}{127}\right) = -\left(\frac{21}{127}\right) = -\left(\frac{127}{21}\right) = -\left(\frac{1}{21}\right) = -1,$$

其中 $127 = 6 \cdot 21 + 1$。这就是 MIT 讲义中的计算链。中间的 $\left(\frac{127}{21}\right)$ 是 Jacobi 符号，它本身不回答「127 是不是模 21 的平方」（$127 \equiv 1$ 当然是），只是计算过程中的一个中间量。 ♦

■ **例 10.5　期中 2 练习题第 5 题**　设 $n \geq 3$ 是奇数，求 $\left(\frac{n^3}{n-2}\right)$。

解　$n-2$ 是正奇数，$\gcd(n, n-2) = \gcd(n, 2) = 1$。由积性，$\left(\frac{n^3}{n-2}\right) = \left(\frac{n}{n-2}\right)^3 = \left(\frac{n}{n-2}\right)$。

方法一（约化）：$n \equiv 2 \pmod{n-2}$，由第二补充律，$\left(\frac{n}{n-2}\right) = \left(\frac{2}{n-2}\right) = (-1)^{\omega(n-2)}$。它为 1 当且仅当 $n-2 \equiv \pm 1 \pmod 8$，即 $n \equiv 1, 3 \pmod 8$。

方法二（翻转）：$\varepsilon(n)\varepsilon(n-2) = \frac{n-1}{2} \cdot \frac{n-3}{2}$ 是两个相邻整数之积，为偶数，翻转不变号：

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

按 $n \bmod 8 = 1, 3, 5, 7$，两个因子分别为 $(1)(1)$、$(-1)(-1)$、$(1)(-1)$、$(-1)(1)$。两种方法一致：

$$\boxed{\left(\frac{n^3}{n-2}\right) = \begin{cases} 1, & n \equiv 1, 3 \pmod 8, \\ -1, & n \equiv 5, 7 \pmod 8. \end{cases}}$$

检验：$n = 5$ 时 $\left(\frac{125}{3}\right) = \left(\frac{2}{3}\right) = -1$；$n = 9$ 时 $729 = 104 \cdot 7 + 1$，$\left(\frac{729}{7}\right) = 1$；$n = 3$ 时分母为 1，值为 1。 ♦

10.4 不做分解的计算方法

10.4.1 算法

有了定理 10.4，第 9 讲的「约化、拆分、翻转」里只剩下拆出 2 的幂这一种拆分。得到的算法与欧几里得算法（第 2 讲）几乎一样。

算法 10.6（计算 Jacobi 符号）　输入整数 P 和正奇数 Q；维护符号 s，初值 $s = 1$。

1. 约化：$P \leftarrow P \bmod Q$（取 $0 \leq P < Q$）。若 $P = 0$：$Q = 1$ 时输出 s，否则输出 0。
2. 去 2：当 P 为偶数时，$P \leftarrow P/2$；若 $Q \equiv \pm 3 \pmod 8$，$s \leftarrow -s$。
3. 翻转：交换 P 与 Q；若两者都 $\equiv 3 \pmod 4$，$s \leftarrow -s$。回到第 1 步。

正确性：每一步都不改变 $s\cdot\left(\frac{P}{Q}\right)$ 的值。第 1 步用命题 10.3(1)，第 2 步用定理 10.4(2) 和积性，第 3 步用定理 10.4(3)。第 3 步交换时 P 是正奇数，可以做分母。若 $\gcd(P,Q)>1$，交换前后两个符号都是 0，等式也成立。

步数：记第 t 次翻转后的分母为 Q_t。下一轮的 P 是 Q_{t-1} 除以 Q_t 的余数再去掉 2，所以 $Q_{t+1}\le Q_{t-1}\bmod Q_t$。与第 2 讲证明欧几里得算法步数时的论证一样，$Q_{t-1}\bmod Q_t<Q_{t-1}/2$，于是分母每两轮至少减半，翻转不超过 $2\log_2 Q+2$ 次。每做一次「去 2」，P 减半，所以去 2 的总次数也不超过 $\log_2 Q$ 量级。**整个计算只要 $O(\log Q)$ 次带余除法**，每一步需要查看的只是模 4、模 8 的余数。

■ **例 10.6　一个较大的例子**　计算 $\left(\frac{2023}{7919}\right)$（7919 是素数），并判断 $x^2\equiv 2023\pmod{7919}$ 是否有解。

解　按算法 10.6 逐步记录，「依据」一列说明这一步为什么变号或不变号。

当前形式	操作	依据
$\left(\frac{2023}{7919}\right)=-\left(\frac{7919}{2023}\right)$	翻转，变号	$2023\equiv 7919\equiv 3\pmod 4$
$-\left(\frac{7919}{2023}\right)=-\left(\frac{1850}{2023}\right)$	约化	$7919=3\cdot 2023+1850$
$-\left(\frac{1850}{2023}\right)=-\left(\frac{925}{2023}\right)$	去 2，不变号	$2023\equiv 7\pmod 8$
$-\left(\frac{925}{2023}\right)=-\left(\frac{2023}{925}\right)=-\left(\frac{173}{925}\right)$	翻转、约化	$925\equiv 1\pmod 4$；$2023=2\cdot 925+173$
$-\left(\frac{173}{925}\right)=-\left(\frac{925}{173}\right)=-\left(\frac{60}{173}\right)$	翻转、约化	$173\equiv 1\pmod 4$；$925=5\cdot 173+60$
$-\left(\frac{60}{173}\right)=-\left(\frac{15}{173}\right)$	去 2^2	平方因子直接去掉
$-\left(\frac{15}{173}\right)=-\left(\frac{173}{15}\right)=-\left(\frac{8}{15}\right)$	翻转、约化	$173\equiv 1\pmod 4$；$173=11\cdot 15+8$
$-\left(\frac{8}{15}\right)=-\left(\frac{2}{15}\right)^3=-1$	去 2^3	$15\equiv 7\pmod 8$

所以 $\left(\frac{2023}{7919}\right)=-1$，同余式无解。中间出现的分母 $2023=7\cdot 17^2$、$925=5^2\cdot 37$、$15=3\cdot 5$ 都是合数，在 Jacobi 符号的框架里这毫无影响。

一共翻转 5 次。作为对比，欧几里得算法求 $\gcd(2023,7919)$ 要做 10 次除法，拆出 2 的幂让数变小得更快。

用分解来核对：$\left(\frac{2023}{7919}\right)=\left(\frac{7}{7919}\right)\left(\frac{17}{7919}\right)^2=\left(\frac{7}{7919}\right)$。$7\equiv 7919\equiv 3\pmod 4$，$7919=1131\cdot 7+2$，所以 $\left(\frac{7}{7919}\right)=-\left(\frac{2}{7}\right)=-1$。用 Euler 判别法也可以核对：$2023^{3959}\equiv -1\pmod{7919}$。 ♦

图 10.2 把上例的计算画成六个状态。每个框记录当前的 (P,Q,s)，框间的箭头合并一次翻转、约化及去掉 2 的因子。

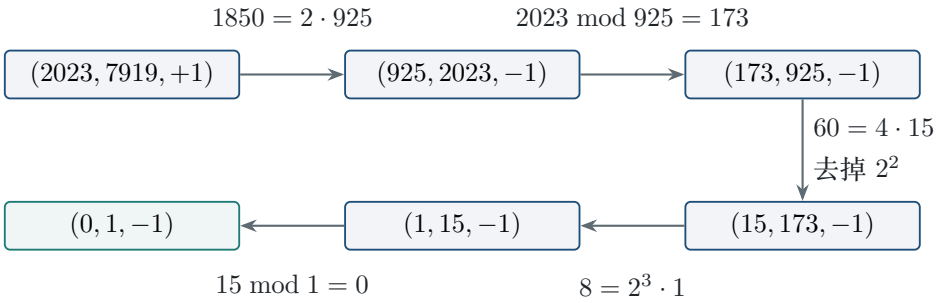


图 10.2　Jacobi 算法的状态变化。第一个箭头使 s 变为 -1，后续符号保持不变；终点 $Q=1$，输出 $s=-1$。

从左上框沿箭头走到左下框，每个状态都满足 $s\left(\frac{P}{Q}\right)=-1$。因此数变小以后，原问题的答案仍被完整保留。若终点的 $Q>1$，则说明原来的分子、分母不互素，应输出 0；不能只读取 s。

注 10.2　算法结束于 $P=0$ 时，最后的 Q 恰好是 $\gcd(P_0,Q_0)$：约化和翻转不改变最大公因数，Q 始终是奇数，去 2 也不改变。所以输出 0 的同时还顺带求出了公因数。

10.4.2 应用：Solovay–Strassen 素性检验

第 6 讲的 Fermat 检验看 a^{n-1} 是否为 1，Miller–Rabin 再看这个幂的平方根链条。Jacobi 符号给出另一种检验。若 n 是奇素数，Euler 判别法说对每个与 n 互素的 a 都有

$$a^{(n-1)/2}\equiv\left(\frac{a}{n}\right)\pmod n.$$

左边用反复平方计算，右边用算法 10.6 计算，都只要 $O(\log n)$ 步，而且都不需要知道 n 是否为素数、怎样分解。若某个 a 使两边不同余，n 必是合数，称 a 为 n 的 **Euler 证据**（Euler witness）。若 n 是合数而两边同余，称 a 为 **Euler 说谎者**（Euler liar）。

■ **例 10.7　揭穿 341**　第 6 讲说过，$341=11\cdot31$ 是以 2 为底的伪素数。用 $a=2$ 做上面的检验。

解　$2^{10}=1024=3\cdot341+1$，所以 $2^{170}=(2^{10})^{17}\equiv1\pmod{341}$。另一边，$341=42\cdot8+5$，由补充律 $\left(\frac{2}{341}\right)=-1$。两边不同余，2 是 Euler 证据，341 是合数。

两边为什么对不上：$\left(\frac{2}{341}\right)=\left(\frac{2}{11}\right)\left(\frac{2}{31}\right)=(-1)\cdot1$，2 模 11 不是平方。可是 2 模 11 的阶是 10，模 31 的阶是 5，都整除 170，所以 2^{170} 模两个素因子都是 1。n 是合数时，指数 $(n-1)/2$ 通常不同于各素因子 p 对应的 $(p-1)/2$，因此不能直接用模 p 的 Euler 判别法把它换成平方性的符号。 ♦

定理 10.7（Solovay–Strassen）　设 n 是奇合数。在 $1,\dots,n-1$ 中与 n 互素的 $\varphi(n)$ 个数里，Euler 说谎者至多占一半。

证明　记说谎者的集合为 L。设 b 与 n 互素且 $b\notin L$。若 $a\in L$ 且 $ab\in L$，则 $a^{(n-1)/2}b^{(n-1)/2}\equiv\left(\frac{a}{n}\right)\left(\frac{b}{n}\right)$，而 $a^{(n-1)/2}\equiv\left(\frac{a}{n}\right)=\pm1$ 可以约去，得到 $b\in L$，矛盾。所以 $a\mapsto ab \bmod n$ 把 L 一对一地映到 L 之外，L 至多占一半。剩下只需找出一个这样的 b。

n 有平方因子。写 $n=p^km$，$k\ge2$，$p\nmid m$。取模 p^k 的原根 g（第 8 讲），由中国剩余定

理取 $b \equiv g \pmod{p^k}$、$b \equiv 1 \pmod{m}$。若 $b \in L$，则 $b^{n-1} \equiv \left(\frac{b}{n}\right)^2 = 1 \pmod{p^k}$，于是 g 的阶 $p^{k-1}(p-1)$ 整除 $n-1$，得 $p \mid n-1$，与 $p \mid n$ 矛盾。

*n 无平方因子。*写 $n = pm$，p 是素数，$m > 1$。取模 p 的非剩余 c，再取 $b \equiv c \pmod{p}$、$b \equiv 1 \pmod{m}$。由命题 10.3，$\left(\frac{b}{n}\right) = \left(\frac{c}{p}\right)\left(\frac{1}{m}\right) = -1$。若 $b \in L$，则 $b^{(n-1)/2} \equiv -1 \pmod{n}$，从而模 m 也 $\equiv -1$；但 $b \equiv 1 \pmod{m}$ 给出 $b^{(n-1)/2} \equiv 1 \pmod{m}$，而 $m \geq 3$，矛盾。 ■

于是对奇合数，在互素底数中均匀随机取样，一轮检验被骗的概率至多 $\frac{1}{2}$，独立做 k 轮至多 2^{-k}。若在 $1, \ldots, n-1$ 中均匀取样，先算最大公因数，发现非互素底数就判合数，同样满足这个上界。这个算法是 Solovay 和 Strassen 在 1977 年发表的，是最早的随机素性检验之一。比较第 6 讲的数据：Carmichael 数 561 的 320 个底数全是 Fermat 说谎者，Euler 说谎者只有 80 个，强说谎者只有 10 个。可以证明强说谎者一定是 Euler 说谎者，所以实际计算仍用 Miller–Rabin。1729 恰有 $648 = \varphi(1729)/2$ 个 Euler 说谎者（可枚举互素底数核对），定理中的「一半」不能改进。Solovay–Strassen 检验能做到多项式时间，前提是 Jacobi 符号不分解 n 就能算，这正是本讲前半部分的内容。

10.5　把 Jacobi 符号看成一个置换的符号

前面的算法把 $\left(\frac{P}{Q}\right)$ 变成一串模 4、模 8 的判断。Zolotareff 给出另一个完全不同的视角：把 $0, 1, \ldots, Q-1$ 中的每个数乘以 P，看这个重新排列的符号。

定义 10.8（置换与符号）　集合 $\{0, 1, \ldots, Q-1\}$ 的一个**置换**（permutation）是它到自身的双射。若 σ 是置换，记

$$\operatorname{inv}(\sigma) = \#\{(i, j) : i < j,\ \sigma(i) > \sigma(j)\}$$

为逆序数，并定义置换的**符号**（sign）

$$\operatorname{sgn}(\sigma) = (-1)^{\operatorname{inv}(\sigma)} \in \{1, -1\}.$$

当 $\gcd(P, Q) = 1$ 时，映射

$$M_{P,Q} : x \longmapsto Px \pmod{Q}$$

确实是置换：若 $Px \equiv Py \pmod{Q}$，则 $Q \mid P(x-y)$；因为 P 与 Q 互素，得到 $Q \mid x-y$。反过来，有限集合上的单射就是满射。

命题 10.9（轮换的符号）　一个长度为 k 的轮换的符号是 $(-1)^{k-1}$。因此，把置换写成不相交轮换后，置换符号等于

$$(-1)^{\text{偶长度轮换的个数}}.$$

证明　轮换 $(a_1\ a_2\ \cdots\ a_k)$ 可以写成

$$(a_1\ a_k)(a_1\ a_{k-1}) \cdots (a_1\ a_2),$$

即 $k-1$ 个换位的复合。相邻换位恰好改变一个逆序；任意换位都能写成奇数个相邻换位，所以每个换位都会翻转逆序数的奇偶性。于是轮换符号为 $(-1)^{k-1}$。不相交轮换彼此独立，符号相乘；奇长度轮换贡献 1，偶长度轮换贡献 -1。 ■

■ **例 10.8**　$P=5, Q=7$ **的轮换分解**　计算 $M_{5,7}$ 的轮换分解，并用它求 $\left(\frac{5}{7}\right)$。

解　逐点计算 $5x \bmod 7$：

$$0\mapsto 0,\quad 1\mapsto 5,\quad 2\mapsto 3,\quad 3\mapsto 1,\quad 4\mapsto 6,\quad 5\mapsto 4,\quad 6\mapsto 2$$

所以

$$M_{5,7}=(0)(1\ 5\ 4\ 6\ 2\ 3)$$

第二个轮换长度为 6，是偶长度，故 $\operatorname{sgn}(M_{5,7})=-1$。另一方面，5 是模 7 的非剩余（平方表只有 $1,2,4$），所以 $\left(\frac{5}{7}\right)=-1$。两种计算吻合。 ♦

图 10.3 把这个置换画成轮换图；每条箭头表示一次乘以 5，可以沿箭头连续追踪同一个元素。

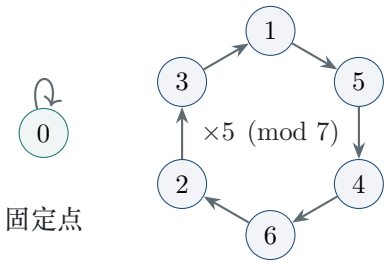


图 10.3　乘以 5 的置换。沿箭头反复走，得到轮换 $(0)(1\ 5\ 4\ 6\ 2\ 3)$。

定理 10.10（Zolotareff 定理）　设 Q 是正奇数，$\gcd(P,Q)=1$。则

$$\operatorname{sgn}(M_{P,Q})=\left(\frac{P}{Q}\right)$$

右边是 Jacobi 符号；当 Q 是素数时就是 Legendre 符号。

思路　先处理奇素数 p：把非零剩余按 $\{\pm 1\},\dots,\{\pm(p-1)/2\}$ 成对排列，乘以 P 时，每出现一个负的绝对最小剩余，就把这一对的方向翻转一次，这正是 Gauss 引理中的计数。再把结论从 p 推到 p^e，最后用中国剩余定理把互素的素数幂拼起来。

证明　**第一步：素数分母。**设 p 是奇素数，$h=(p-1)/2$。对 $k=1,\dots,h$，把 Pk 化成绝对最小剩余 r_k，即 $-h\le r_k\le h$。这些 $|r_k|$ 两两不同，恰好是 $1,2,\dots,h$ 的一个排列；否则 $Pk\equiv\pm P\ell \pmod p$ 会推出 $k\equiv\pm\ell \pmod p$，与 $1\le k,\ell\le h$ 矛盾。

在非零剩余中按成对的 $\{j,-j\}$ 排列。乘以 P 先置换这些 h 对，再在第 k 对中根据 r_k 的正负决定是否交换两个方向。置换这些「对」的符号在完整的 $2h$ 个元素上出现两次，贡献 1；因此

$$\operatorname{sgn}(M_{P,p})=(-1)^{\#\{k:r_k<0\}}$$

右边由 Gauss 引理等于 $\left(\frac{P}{p}\right)$。0 是固定点，不改变符号。

第二步：素数幂分母。用归纳证明 $\operatorname{sgn}(M_{P,p^e}) = \left(\frac{P}{p}\right)^e$。把 $\mathbb{Z}/p^e\mathbb{Z}$ 按模 p 的余数分成 p 个纤维，每个纤维有 p^{e-1} 个元素。纤维之间的置换是 $M_{P,p}$；每个位置在纤维内有 p^{e-1} 个副本，故贡献 $\left(\frac{P}{p}\right)^{p^{e-1}} = \left(\frac{P}{p}\right)$。固定一个纤维后，内部映射形如 $t \mapsto Pt + c$。平移在 p^{e-1} 个元素上的轮换长度都是奇数，符号为 1；乘法部分按归纳假设贡献 $\left(\frac{P}{p}\right)^{e-1}$。共有 p 个纤维，内部贡献的乘积为 $\left(\frac{P}{p}\right)^{p(e-1)} = \left(\frac{P}{p}\right)^{e-1}$。合并两部分，得到 $\left(\frac{P}{p}\right)^e$。

第三步：互素分母相乘。若 $\gcd(m,n) = 1$，中国剩余定理把 $\mathbb{Z}/mn\mathbb{Z}$ 识别为 $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$，并把 $M_{P,mn}$ 识别为两个置换的直积。若 α 作用在 r 个元素上、β 作用在 s 个元素上，则直积置换的符号是

$$\operatorname{sgn}(\alpha \times \beta) = \operatorname{sgn}(\alpha)^s \operatorname{sgn}(\beta)^r$$

这里 m, n 都是奇数，故指数仍为奇数。于是

$$\operatorname{sgn}(M_{P,mn}) = \operatorname{sgn}(M_{P,m}) \operatorname{sgn}(M_{P,n}) = \left(\frac{P}{m}\right)\left(\frac{P}{n}\right) = \left(\frac{P}{mn}\right)$$

把 Q 分解成互素的素数幂，结论随之成立。 ■

■ **例 10.9　合数分母也能直接看出符号**　用轮换分解计算 $\left(\frac{2}{15}\right)$。

解　乘以 2 的置换分解为

$$(0)(1\ 2\ 4\ 8)(3\ 6\ 12\ 9)(5\ 10)(7\ 14\ 13\ 11)$$

共有四个偶长度轮换，所以符号为 $(-1)^4 = 1$。按 Jacobi 符号直接算也得到

$$\left(\frac{2}{15}\right) = \left(\frac{2}{3}\right)\left(\frac{2}{5}\right) = (-1)(-1) = 1$$

这里的 1 仍不能推出 2 是模 15 的平方；模 3 和模 5 的两个 Legendre 符号都为 -1，乘积才变成 1。 ♦

10.5.1　从置换符号重看补充律与互反律

令 $Q = 2h + 1$。映射 $M_{-1,Q}$ 固定 0，并把 k 与 $Q - k$ 两两交换（$1 \le k \le h$），所以它由 h 个换位组成。定理 10.10 立刻给出

$$\left(\frac{-1}{Q}\right) = (-1)^h = (-1)^{(Q-1)/2}$$

这就是 Jacobi 符号的第一补充律。

乘以 2 后，$0, 1, \ldots, Q-1$ 的像依次为

$$0, 2, 4, \ldots, 2h, 1, 3, \ldots, 2h-1$$

每一段内部递增，逆序只出现在两段之间。对每个奇数 $2j+1$，前一段里比它大的偶数有 $h - j$ 个，因此逆序数为

$$h + (h-1) + \cdots + 1 = \frac{h(h+1)}{2} = \frac{Q^2 - 1}{8}$$

于是

$$\left(\frac{2}{Q}\right) = (-1)^{(Q^2-1)/8}$$

得到第二补充律。

前面已经证明了 Jacobi 互反律。借助 Zolotareff 定理，它也可以读成两个置换符号的乘积公式：对互素正奇数 P, Q，

$$\operatorname{sgn}(M_{P,Q})\operatorname{sgn}(M_{Q,P}) = (-1)^{\frac{P-1}{2}\cdot\frac{Q-1}{2}}$$

当 $P \equiv Q \equiv 3 \pmod 4$ 时乘积为 -1，其余情形为 1。这一步把已证的互反律换成了置换语言。

前后联系　第 9 讲的 Gauss 引理给出了 Legendre 符号的「负号计数」；本讲的 Zolotareff 定理把这个计数解释成一个置换的符号。两种语言都在记录同一个奇偶性。第 11 讲转向 Tonelli 算法时，模素数的平方性仍可用 Jacobi 算法计算 Legendre 符号来判断；真正求平方根需要新的构造。

本讲小结

- Jacobi 符号的值为 -1 时，分子一定不是模分母的平方；值为 1 时不能直接下结论。
- Jacobi 互反律把「翻转分子和分母」变成欧几里得式的快速计算，整个算法只需对数级别的带余除法。
- Solovay–Strassen 检验用 $a^{(n-1)/2}$ 与 $\left(\frac{a}{n}\right)$ 的不一致性排除合数；奇合数的 Euler 说谎者至多占互素底数的一半。
- 置换符号定义为 $(-1)^{\operatorname{inv}(\sigma)}$；长度为 k 的轮换符号是 $(-1)^{k-1}$。
- Zolotareff 定理：对奇数 Q 与 $\gcd(P,Q)=1$，乘以 P 的置换满足 $\operatorname{sgn}(M_{P,Q}) = \left(\frac{P}{Q}\right)$。

练习

10.1 ★　计算 $\left(\frac{21}{55}\right)$、$\left(\frac{-1}{99}\right)$ 和 $\left(\frac{2}{45}\right)$。

10.2 ★　用算法 10.6 计算 $\left(\frac{1001}{9907}\right)$，并记录每次翻转后的分母。

10.3 ★　写出 $M_{4,7}$ 的轮换分解，并由轮换长度求 $\left(\frac{4}{7}\right)$。

10.4 ★★　验证 $M_{2,9} = (0)(1\ 2\ 4\ 8\ 7\ 5)(3\ 6)$，用它求 $\left(\frac{2}{9}\right)$。

10.5 ★★　证明置换直积的符号公式：$\operatorname{sgn}(\alpha \times \beta) = \operatorname{sgn}(\alpha)^{|B|}\operatorname{sgn}(\beta)^{|A|}$。

10.6 ★★　设 p 是奇素数，非零剩余类分成两个非空集合 S_1, S_2，满足「同一集合中两元素的乘积属于 S_1，而 S_1 与 S_2 中各取一元素的乘积属于 S_2」。证明 S_1 恰为二次剩余，S_2 恰为非剩余。

10.7 ★★　求所有 $a \in \{1, 2, \ldots, 14\}$ 使 $\left(\frac{a}{15}\right) = 1$，并指出其中哪些是模 15 的平方。

10.8 ★★★　证明对奇素数 p、满足 $p \nmid a$ 的整数 a 与任意 $e \geq 1$，$\operatorname{sgn}(M_{a,p^e}) = \left(\frac{a}{p}\right)^e$。

10.9 ★★★ 证明 $\left(\frac{15}{p}\right)=1$ 的素数 $p>5$ 恰落在模 60 的剩余类 $\pm1,\pm7,\pm11,\pm17$ 中。

10.10 ★★★ 设 $n>1$ 是奇数，证明若某个与 n 互素的底数满足 $a^{(n-1)/2}\not\equiv\left(\frac{a}{n}\right)\pmod n$，则 n 必为合数，并说明为什么重复独立检验 k 次后错误接受概率至多 2^{-k}。

练习答案与提示

10.1 $\left(\frac{21}{55}\right)=\left(\frac{21}{5}\right)\left(\frac{21}{11}\right)=1\cdot(-1)=-1$；$\left(\frac{-1}{99}\right)=\left(\frac{-1}{3}\right)^2\left(\frac{-1}{11}\right)=1\cdot(-1)=-1$；$\left(\frac{2}{45}\right)=\left(\frac{2}{3}\right)^2\left(\frac{2}{5}\right)=1\cdot(-1)=-1$。

10.2 $\left(\frac{1001}{9907}\right)=-1$。约化、去掉 2 的幂并翻转后，分母依次为 $1001,449,103,37,29,1$；最后得到 $-\left(\frac{1}{1}\right)=-1$。

10.3 $M_{4,7}=(0)(1\ 4\ 2)(3\ 5\ 6)$，没有偶长度轮换，符号为 1，所以 $\left(\frac{4}{7}\right)=1$。

10.4 直接计算 $2x \bmod 9$ 得 $M_{2,9}=(0)(1\ 2\ 4\ 8\ 7\ 5)(3\ 6)$。六轮换和二轮换各贡献 -1，总符号为 1；这与 $\left(\frac{2}{9}\right)=\left(\frac{2}{3}\right)^2=1$ 一致。

10.5 把 $A\times B$ 中的元素按 B 分成块。α 在每一列复制 $|B|$ 次，贡献 $\operatorname{sgn}(\alpha)^{|B|}$；固定 A 坐标时，β 在每一行复制 $|A|$ 次，贡献 $\operatorname{sgn}(\beta)^{|A|}$。两步复合得到公式。

10.6 每个非零平方 x^2 都属于 S_1，因为 x 必在 S_1 或 S_2，而同一集合中两元素的乘积属于 S_1。取 $u\in S_2$；它不能是平方。对每个平方 $r\in S_1$，$ur\in S_2$，而这些 ur 是互不相同的非剩余，共有 $(p-1)/2$ 个，恰好穷尽全部非剩余。因此 S_1 只有平方，S_2 只有非平方。

10.7 Jacobi 值为 1 的是 $a=1,2,4,8$；模 15 的平方只有 $1,4$，所以 $2,8$ 是 Jacobi 值为 1 但并非平方的例子。

10.8 按证明中的纤维分解，纤维置换贡献 $\left(\frac{a}{p}\right)$，纤维内部的乘法按归纳贡献 $\left(\frac{a}{p}\right)^{e-1}$；平移由奇长度轮换组成，符号为 1，故总符号为 $\left(\frac{a}{p}\right)^{e}$。

10.9 对 $p>5$，互反律给出 $\left(\frac{15}{p}\right)=(-1)^{(p-1)/2}\left(\frac{p}{3}\right)\left(\frac{p}{5}\right)$。模 12 下，$\left(\frac{3}{p}\right)=1$ 对应 $p\equiv\pm1$，为 -1 对应 $p\equiv\pm5$；模 5 下，平方类是 ±1，非平方类是 ±2。两个符号相同的 CRT 组合为 $p\equiv1,7,11,17,43,49,53,59\pmod{60}$，即 $\pm1,\pm7,\pm11,\pm17$。

10.10 若 n 为素数，Euler 判别法保证同余成立，所以不一致直接排除素数。Solovay–Strassen 定理说明奇合数的说谎者至多占互素底数的一半；独立抽取 k 个底数同时被骗的概率至多 $(1/2)^k$。

第 11 讲

模 p 平方根：Tonelli 算法与连续平方剩余

Square Roots mod p, Tonelli's Algorithm

对应材料：MIT 18.781 Lecture 11 讲义（Square Roots, Tonelli's Algorithm, Number of Consecutive Pairs of Squares mod p）；习题集 5 第 1、2、9 题；期中模拟题 2 第 3、4、7 题。

本讲目标

(1) 会用「约化、拆分、翻转」快速计算 Jacobi 符号，理解它和欧几里得算法的相似之处；知道 Jacobi 符号等于 1 不保证是平方。

(2) 会证明「$\left(\frac{p}{q}\right)$ 只依赖于 $q \bmod 4p$」，并用它列出使固定的数 a 成为平方剩余的全部素数类。

(3) 掌握两个直接求平方根的公式：$p \equiv 3 \pmod 4$ 时 $x = a^{(p+1)/4}$，$p \equiv 5 \pmod 8$ 时先算 $a^{(p+3)/8}$ 再修正。

(4) 理解 Tonelli 算法：先找一个非剩余 n，再用 $c = n^t$ 一层一层地把 A 的阶中 2 的幂次降到零；会手算完整的例子。

(5) 知道这个算法为什么快、哪一步依赖随机性，以及模合数开平方与分解因数的关系。

(6) 用 Legendre 符号求和，精确数出模 p 有多少个 x 使 x 与 $x+1$ 都是平方。

11.1 回顾：用 Jacobi 符号计算二次特征

第 9 讲证明了二次互反律，第 10 讲把 Legendre 符号推广成 Jacobi 符号 $\left(\frac{P}{Q}\right)$（Q 为正奇数）。本讲反复要用的是下面这些性质，先集中列出。

命题 11.1（Jacobi 符号的计算规则） 设 Q, Q' 为正奇数，P, P' 为整数。

(1) 约化：$P \equiv P' \pmod Q$ 时 $\left(\frac{P}{Q}\right) = \left(\frac{P'}{Q}\right)$。

(2) 乘法性：$\left(\frac{PP'}{Q}\right) = \left(\frac{P}{Q}\right)\left(\frac{P'}{Q}\right)$，$\left(\frac{P}{QQ'}\right) = \left(\frac{P}{Q}\right)\left(\frac{P}{Q'}\right)$。

(3) 补充律：$\left(\frac{-1}{Q}\right) = (-1)^{(Q-1)/2}$，$\left(\frac{2}{Q}\right) = (-1)^{(Q^2-1)/8}$，即 $Q \equiv \pm 1 \pmod 8$ 时为 1，$Q \equiv \pm 3 \pmod 8$ 时为 -1。

(4) 互反律：P, Q 是互素的正奇数时，$\left(\frac{P}{Q}\right)\left(\frac{Q}{P}\right) = (-1)^{\frac{P-1}{2}\cdot\frac{Q-1}{2}}$，即只有 $P \equiv Q \equiv 3 \pmod 4$ 时翻转要变号。

当 Q 是素数时，Jacobi 符号就是 Legendre 符号。

有了这些规则，计算 $\left(\frac{a}{p}\right)$ 时**不必分解** a：只要反复做三件事，约化（把上面的数对下面的数取余）、拆分（拿出因子 2）、翻转（用互反律交换上下）。

例 11.1　MIT 讲义的例子　223 是素数。判断 1729 是不是模 223 的二次剩余。

解　逐步计算，每一步写明用到的规则。

$$
\begin{aligned}
\left(\frac{1729}{223}\right) &= \left(\frac{168}{223}\right) && 1729 = 7\cdot 223 + 168 \text{（约化）}\\
&= \left(\frac{4}{223}\right)\left(\frac{42}{223}\right) = \left(\frac{42}{223}\right) && 168 = 4\cdot 42,\ \left(\frac{4}{223}\right) = 1\\
&= \left(\frac{2}{223}\right)\left(\frac{21}{223}\right) = \left(\frac{21}{223}\right) && 223 \equiv 7 \pmod 8,\ \left(\frac{2}{223}\right) = 1\\
&= \left(\frac{223}{21}\right) && 21 \equiv 1 \pmod 4 \text{（翻转不变号）}\\
&= \left(\frac{13}{21}\right) = \left(\frac{21}{13}\right) && 223 = 10\cdot 21 + 13,\ 13 \equiv 1 \pmod 4\\
&= \left(\frac{8}{13}\right) = \left(\frac{2}{13}\right)^3 = \left(\frac{2}{13}\right) && 21 = 13 + 8\\
&= -1 && 13 \equiv 5 \pmod 8.
\end{aligned}
$$

所以 1729 不是模 223 的二次剩余。也可以用 Euler 判别法核对：用反复平方算出 $1729^{111} \equiv -1 \pmod{223}$，结论一致。

如果只许用 Legendre 符号，到 $\left(\frac{21}{223}\right)$ 这一步就得把 21 分解成 $3\cdot 7$，再分别计算：$\left(\frac{3}{223}\right) = -\left(\frac{223}{3}\right) = -\left(\frac{1}{3}\right) = -1$（3 与 223 都 $\equiv 3 \pmod 4$），$\left(\frac{7}{223}\right) = -\left(\frac{223}{7}\right) = -\left(\frac{6}{7}\right) = -\left(\frac{-1}{7}\right) = 1$，乘积还是 -1。数小时分解不难，数大时分解很慢，Jacobi 符号的好处就在于完全绕开分解。 ♦

这个计算过程和求最大公因数的欧几里得算法（第 2 讲）几乎一样：每一步都把一对数换成一对更小的数，而且换法本质上是做一次带余除法。表 11.1 把二者并排对照。

	欧几里得算法	Jacobi 符号算法
处理的对象	数对 (a, b)	数对 (P, Q)，Q 为正奇数
每一步	$(a, b) \to (b \bmod a,\ a)$	约化 $P \bmod Q$，拆出 2 的幂，再翻转
依据	$\gcd(a, b) = \gcd(b \bmod a, a)$	命题 11.1 的四条规则
1729, 223 的过程	$1729 \to 168 \to 55 \to 3 \to 1$	$1729 \to 168 \to 21 \to 13 \to 8 \to 2$
步数	$O(\log b)$	$O(\log Q)$
要不要分解因数	不要	不要
停止条件	余数为 0	上面的数变成 1（或 $\gcd > 1$，值为 0）

表 11.1　欧几里得算法与 Jacobi 符号算法的对照。两者都反复使用带余除法与交换，步数的上界随输入位数线性增长；这不表示每组输入都需要同样多的步骤。

Jacobi 符号等于 1 不代表是平方　Q 不是素数时，$\left(\frac{P}{Q}\right)=1$ 推不出 P 是模 Q 的平方。例如 $\left(\frac{2}{15}\right)=\left(\frac{2}{3}\right)\left(\frac{2}{5}\right)=(-1)(-1)=1$，可是模 15 的平方只有 $0,1,4,6,9,10$，没有 2（模 3 时 2 已经不是平方）。反过来，$\left(\frac{P}{Q}\right)=-1$ 时 Q 必有一个素因子 q 使 $\left(\frac{P}{q}\right)=-1$，所以 P 一定不是模 Q 的平方。上面例题里出现了 $\left(\frac{223}{21}\right)$ 这样分母是合数的中间量，这没有问题：中间量只是计算工具，最后的答案是关于素数 223 的。

11.2　固定 a：哪些素数让 a 成为平方

第 9 讲用互反律证明了：3 是模奇素数 $q\neq 3$ 的平方，当且仅当 $q\equiv\pm1 \pmod{12}$。这类结论有一个共同的模式：**把 a 固定、让素数 q 变**，$\left(\frac{a}{q}\right)$ 只取决于 q 除以某个固定的数的余数。下面先对 a 是奇素数的情形给出准确的陈述。

> **引理 11.2**　设 p,q,r 是互不相同的奇素数，且 $q\equiv r \pmod{4p}$，则 $\left(\frac{p}{q}\right)=\left(\frac{p}{r}\right)$。如果 $p\equiv 1 \pmod 4$，只要 $q\equiv r \pmod p$ 就够了。

思路　$\left(\frac{p}{q}\right)$ 的「上面」是固定的 p，「下面」在变，没法直接约化。用互反律把它翻转成 $\left(\frac{q}{p}\right)$，这时 q 在上面，可以按模 p 约化。翻转多出来的符号只取决于 $q \bmod 4$。模 $4p$ 同余正好同时管住了 $q \bmod p$ 和 $q \bmod 4$。

证明　由 $4p\mid q-r$ 得 $p\mid q-r$，所以 $\left(\frac{q}{p}\right)=\left(\frac{r}{p}\right)$。又 $4\mid q-r$，所以 q,r 同为 $1 \bmod 4$ 或同为 $3 \bmod 4$，于是 $\frac{q-1}{2}$ 与 $\frac{r-1}{2}$ 奇偶性相同，

$$(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}=(-1)^{\frac{p-1}{2}\cdot\frac{r-1}{2}}.$$

由互反律，

$$\left(\frac{p}{q}\right)=\left(\frac{q}{p}\right)(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}=\left(\frac{r}{p}\right)(-1)^{\frac{p-1}{2}\cdot\frac{r-1}{2}}=\left(\frac{p}{r}\right).$$

若 $p\equiv 1 \pmod 4$，则 $\frac{p-1}{2}$ 是偶数，两个符号因子都是 1，证明中只用到了 $q\equiv r \pmod p$。 ■

证明的关键一步是**翻转之后再约化**。翻转把「下面在变」的问题变成「上面在变」的问题，而上面的数可以约化。

注 11.1　用同样的论证（或直接用 Jacobi 符号的互反律）可以证明更一般的结论：对任意非零整数 a，$\left(\frac{a}{q}\right)$ 作为奇素数 $q\nmid a$ 的函数只依赖于 $q \bmod 4|a|$。这正是第 9 讲末尾的核心思想。在模 $4p$ 的 $\varphi(4p)=2(p-1)$ 个与 $4p$ 互素的类中，恰有一半，即 $p-1$ 个类，使 $\left(\frac{p}{q}\right)=1$：当 $p\equiv 1 \pmod 4$ 时，这些类就是 $q \bmod p$ 为平方的类（每个模 p 的类对应模 $4p$ 的两个类）；当 $p\equiv 3 \pmod 4$ 时，$q\equiv 1 \pmod 4$ 的类要求 q 是模 p 的平方，$q\equiv 3 \pmod 4$ 的类要求 q 不是平方，各贡献 $\frac{p-1}{2}$ 个。

■ **例 11.2　MIT 讲义的例子：17**　刻画所有使 17 成为模 q 的平方的素数 q。

解　先看两个特殊的素数。$q=2$：模 2 下 $17\equiv 1=1^2$，是平方。$q=17$：$17\equiv 0=0^2$，也是平方。

设 q 是奇素数且 $q \neq 17$。因为 $17 \equiv 1 \pmod 4$，引理 11.2 说明只需看 $q \bmod 17$，并且翻转不变号：$\left(\frac{17}{q}\right) = \left(\frac{q}{17}\right)$。所以 17 是模 q 的平方当且仅当 q 是模 17 的平方。把 $1, \dots, 8$ 平方后模 17：

$$1,\ 4,\ 9,\ 16,\ 25 \equiv 8,\ 36 \equiv 2,\ 49 \equiv 15,\ 64 \equiv 13,$$

即平方类为 $\{1, 2, 4, 8, 9, 13, 15, 16\} = \{\pm 1, \pm 2, \pm 4, \pm 8\}$。

结论：17 是模 q 的平方当且仅当 $q = 2$，$q = 17$，或 $q \equiv \pm 1, \pm 2, \pm 4, \pm 8 \pmod{17}$。在 110 以内的奇素数中，满足条件的是 $13, 17, 19, 43, 47, 53, 59, 67, 83, 89, 101, 103$。抽查两个：$17 \equiv 4 = 2^2 \pmod{13}$；$17 \equiv -2 \pmod{19}$，而 $6^2 = 36 = 19 + 17$，所以 $17 \equiv 6^2 \pmod{19}$。 ♦

■ **例 11.3　MIT 讲义的例子：19**　刻画所有使 19 成为模 q 的平方的奇素数 $q \neq 19$。

解　这一次 $19 \equiv 3 \pmod 4$，翻转是否变号取决于 $q \bmod 4$，所以要看 $q \bmod 76$。分两种情况。

- $q \equiv 1 \pmod 4$：$\left(\frac{19}{q}\right) = \left(\frac{q}{19}\right)$，要求 q 是模 19 的平方。
- $q \equiv 3 \pmod 4$：$\left(\frac{19}{q}\right) = -\left(\frac{q}{19}\right)$，要求 q 不是模 19 的平方。

模 19 的平方类是 $\{1, 4, 5, 6, 7, 9, 11, 16, 17\}$，非平方类是 $\{2, 3, 8, 10, 12, 13, 14, 15, 18\}$。对每个模 19 的类与模 4 的类，用中国剩余定理合成一个模 76 的类，得到表 11.2。

$q \equiv 1 \pmod 4$，$q \bmod 19$ 为平方	1	4	5	6	7	9	11	16	17
对应的 $q \bmod 76$	1	61	5	25	45	9	49	73	17
$q \equiv 3 \pmod 4$，$q \bmod 19$ 非平方	2	3	8	10	12	13	14	15	18
对应的 $q \bmod 76$	59	3	27	67	31	51	71	15	75

表 11.2　使 $\left(\frac{19}{q}\right) = 1$ 的模 76 剩余类。两行合起来共 $18 = 19 - 1$ 个类。

把这 18 个类按大小排好并配对：19 是模 q 的平方当且仅当

$$q \equiv \pm 1,\ \pm 3,\ \pm 5,\ \pm 9,\ \pm 15,\ \pm 17,\ \pm 25,\ \pm 27,\ \pm 31 \pmod{76}.$$

结果关于 $q \mapsto -q$ 对称，这不是巧合：把 q 换成 $-q$ 时，$q \bmod 4$ 从 1 变 3（翻转多一个负号），同时 $\left(\frac{-q}{19}\right) = \left(\frac{-1}{19}\right)\left(\frac{q}{19}\right) = -\left(\frac{q}{19}\right)$ 也多一个负号，两个负号抵消。抽查：$q = 3$ 时 $19 \equiv 1 = 1^2 \pmod 3$；$q = 5$ 时 $19 \equiv 4 = 2^2 \pmod 5$。 ♦

想一想　5 和 43 都 $\equiv 5 \pmod{19}$。19 是不是模 5 的平方？是不是模 43 的平方？这说明了什么？

答　$19 \equiv 4 = 2^2 \pmod 5$，是平方。$\left(\frac{19}{43}\right) = -\left(\frac{43}{19}\right) = -\left(\frac{5}{19}\right) = -\left(\frac{19}{5}\right) = -\left(\frac{4}{5}\right) = -1$，不是平方（19 和 43 都 $\equiv 3 \pmod 4$，翻转变号）。两个素数模 19 同余，答案却相反，因为 $5 \equiv 1$，$43 \equiv 3 \pmod 4$。所以 $p \equiv 3 \pmod 4$ 时引理 11.2 中的模 $4p$ 不能换成模 p。

前后联系　例 11.2 的答案是一组同余类。每个与模数互素的同余类里都有无穷多个素数，这是 Dirichlet 定理，所以上面的每个类都不是空的。第 12 讲会用分圆多项式证明其中的特殊情形：$q \equiv 1 \pmod{n}$ 的素数有无穷多个。

11.3　从「有没有」到「是多少」：两个直接公式

Euler 判别法回答了 $x^2 \equiv a \pmod{p}$ 有没有解，但没有给出解。逐个试 $x = 1, 2, \ldots, \frac{p-1}{2}$ 要做约 $p/2$ 次乘法，p 有几百位时完全不可行。我们要的是步数为 $\log p$ 的多项式的算法。先看两种容易的情形。

命题 11.3（$p \equiv 3 \pmod 4$）　设 $p \equiv 3 \pmod 4$，$\left(\frac{a}{p}\right) = 1$，则 $x = a^{(p+1)/4}$ 满足 $x^2 \equiv a \pmod p$，两个解是 $\pm a^{(p+1)/4}$。

证明　$p \equiv 3 \pmod 4$ 保证 $\frac{p+1}{4}$ 是整数。由 Euler 判别法 $a^{(p-1)/2} \equiv \left(\frac{a}{p}\right) = 1$，所以

$$x^2 = a^{(p+1)/2} = a^{(p-1)/2} \cdot a \equiv 1 \cdot a = a \pmod p.$$ ■

这个证明的想法是：$a^{(p-1)/2} \equiv 1$ 说明 $a^{(p+1)/2} \equiv a$，而 $\frac{p+1}{2}$ 是偶数时就能「除以 2」。$\frac{p+1}{2}$ 是偶数恰好就是 $p \equiv 3 \pmod 4$。

■ **例 11.4**　解 $x^2 \equiv 3 \pmod{83}$。

解　先判断有没有解。83 是素数，$83 \equiv 3 \pmod 4$，$3 \equiv 3 \pmod 4$，翻转变号：$\left(\frac{3}{83}\right) = -\left(\frac{83}{3}\right) = -\left(\frac{2}{3}\right) = -(-1) = 1$。有解。

由命题 11.3，$x \equiv 3^{(83+1)/4} = 3^{21}$。用反复平方（第 6 讲）：

$$3^1 = 3, \quad 3^2 = 9, \quad 3^4 = 81 \equiv -2, \quad 3^8 \equiv 4, \quad 3^{16} \equiv 16 \pmod{83}.$$

$21 = 16 + 4 + 1$，所以 $3^{21} = 3^{16} \cdot 3^4 \cdot 3 \equiv 16 \cdot (-2) \cdot 3 = -96 \equiv -13 \pmod{83}$。检验：$13^2 = 169 = 2 \cdot 83 + 3$。两个解是 $x \equiv \pm 13$，即 $x \equiv 13$ 或 $70 \pmod{83}$。整个计算只用了 4 次平方和 2 次乘法。 ♦

想一想　设 $p \equiv 3 \pmod 4$。如果忘了先判断，对一个非剩余 a 照样算 $x = a^{(p+1)/4}$，得到的 x^2 是什么？用 $a = 2$，$p = 83$ 试一试。

答　$x^2 = a^{(p-1)/2} \cdot a \equiv -a$，得到的是 $-a$ 的平方根。$p \equiv 3 \pmod 4$ 时 $\left(\frac{-1}{p}\right) = -1$，所以 a 与 $-a$ 恰有一个是剩余，公式总给出「是剩余的那一个」的平方根。$83 \equiv 3 \pmod 8$，2 是非剩余；$2^{21} \equiv 74 \equiv -9$，而 $(-9)^2 = 81 \equiv -2 \pmod{83}$。所以算完一定要平方检验。

■ **例 11.5　改编自期中模拟题 2**　解同余式 $3x^2 + 4x - 2 \equiv 0 \pmod{31}$。

解　按第 9 讲的配方法，两边乘 $4a=12$（与 31 互素，不改变解集）：

$$36x^2+48x-24=(6x+4)^2-16-24=(6x+4)^2-40.$$

于是同余式等价于 $(6x+4)^2\equiv 40\equiv 9 \pmod{31}$，判别式 $D=4^2-4\cdot 3\cdot(-2)=40$。这里 $9=3^2$，平方根一眼看出：$6x+4\equiv\pm 3$。

如果判别式不是一眼能看出的平方，就用命题 11.3（$31\equiv 3 \pmod 4$）。为了演示，换一种配方：先乘 $3^{-1}\equiv 21 \pmod{31}$（$3\cdot 21=63=2\cdot 31+1$）化成首一，$x^2+84x-42\equiv x^2+22x+20$，再配方 $(x+11)^2\equiv 121-20=101\equiv 8$。$\left(\frac{8}{31}\right)=\left(\frac{2}{31}\right)^3=1$（$31\equiv 7 \pmod 8$），所以 $\sqrt{8}\equiv\pm 8^{(31+1)/4}=\pm 8^8$。$8^2=64\equiv 2$，$8^4\equiv 4$，$8^8\equiv 16$；检验 $16^2=256=8\cdot 31+8$。

两种做法给出同样的答案。用第二种：$x+11\equiv\pm 16$，$x\equiv 5$ 或 $x\equiv -27\equiv 4 \pmod{31}$。代回检验：$3\cdot 25+20-2=93=3\cdot 31$；$3\cdot 16+16-2=62=2\cdot 31$。解是 $x\equiv 4,5 \pmod{31}$。 ♦

下一个情形是 $p\equiv 5 \pmod 8$。这时 $\frac{p+1}{2}$ 是奇数，上面的「除以 2」行不通，但可以「除以 4」，代价是多一个修正因子。

> **命题 11.4（$p\equiv 5 \pmod 8$）**　设 $p\equiv 5 \pmod 8$，$\left(\frac{a}{p}\right)=1$，$x_0=a^{(p+3)/8}$。则 $a^{(p-1)/4}\equiv\pm 1$，并且
>
> - 若 $a^{(p-1)/4}\equiv 1$，则 $x_0^2\equiv a$；
> - 若 $a^{(p-1)/4}\equiv -1$，令 $i=2^{(p-1)/4}$，则 $i^2\equiv -1$，且 $(ix_0)^2\equiv a$。

证明　写 $p=8k+5$，则 $\frac{p+3}{8}=k+1$，$\frac{p-1}{4}=2k+1$ 都是整数。$y=a^{(p-1)/4}$ 满足 $y^2=a^{(p-1)/2}\equiv 1$，而模素数 p 时 $y^2\equiv 1$ 只有 $y\equiv\pm 1$（第 9 讲）。又 $x_0^2=a^{(p+3)/4}=a^{(p-1)/4}\cdot a=ya$。$y\equiv 1$ 时 $x_0^2\equiv a$。$y\equiv -1$ 时：$p\equiv 5 \pmod 8$ 使 $\left(\frac{2}{p}\right)=-1$，由 Euler 判别法 $i^2=2^{(p-1)/2}\equiv -1$，于是 $(ix_0)^2\equiv(-1)(-a)=a$。 ■

■ **例 11.6**　$p=29\equiv 5 \pmod 8$。解 $x^2\equiv 5$ 与 $x^2\equiv 7 \pmod{29}$。

解　两个都有解：$\left(\frac{5}{29}\right)=\left(\frac{29}{5}\right)=\left(\frac{4}{5}\right)=1$，$\left(\frac{7}{29}\right)=\left(\frac{29}{7}\right)=\left(\frac{1}{7}\right)=1$（$29\equiv 1 \pmod 4$，翻转不变号）。这里 $\frac{p+3}{8}=4$，$\frac{p-1}{4}=7$。

$a=5$：$5^2=25\equiv -4$，$5^4\equiv 16$，$5^7=5^4\cdot 5^2\cdot 5\equiv 16\cdot(-4)\cdot 5=-320=-11\cdot 29-1\equiv -1$。属于第二种情形。$x_0=5^4\equiv 16$，$x_0^2=256\equiv 24\equiv -5$，确实差一个负号。修正因子 $i=2^7=128\equiv 12$，$12^2=144=5\cdot 29-1\equiv -1$。$x=ix_0\equiv 12\cdot 16=192\equiv 18$，检验 $18^2=324=11\cdot 29+5$。解为 $x\equiv\pm 18$，即 $11,18 \pmod{29}$。

$a=7$：$7^2=49\equiv -9$，$7^4\equiv 81\equiv -6$，$7^7\equiv(-6)(-9)\cdot 7=378=13\cdot 29+1\equiv 1$。属于第一种情形，$x=7^4\equiv -6$，检验 $6^2=36\equiv 7$。解为 $x\equiv\pm 6$，即 $6,23 \pmod{29}$。 ♦

剩下 $p\equiv 1 \pmod 8$。照搬上面的思路会卡在两个地方。第一，$a^{(p-1)/8}$ 的平方是 $a^{(p-1)/4}$，可能是 ± 1，也可能是 -1 的平方根，情形越分越多。第二，修正因子 $i=2^{(p-1)/4}$ 失效了：$p\equiv 1 \pmod 8$ 时 2 是剩余，$2^{(p-1)/2}\equiv 1$，$i^2=1$ 而不是 -1。我们需要一个已知的非剩余来造修正因

子，而没有一个固定的整数对所有 $p \equiv 1 \pmod 8$ 都是非剩余：对固定的奇素数 n，凡是 $p \equiv 1 \pmod{8n}$ 的素数都有 $\left(\frac{n}{p}\right) = \left(\frac{p}{n}\right) = \left(\frac{1}{n}\right) = 1$，这样的素数有无穷多个（第 12 讲）。图 11.1 总结了到目前为止的局面。

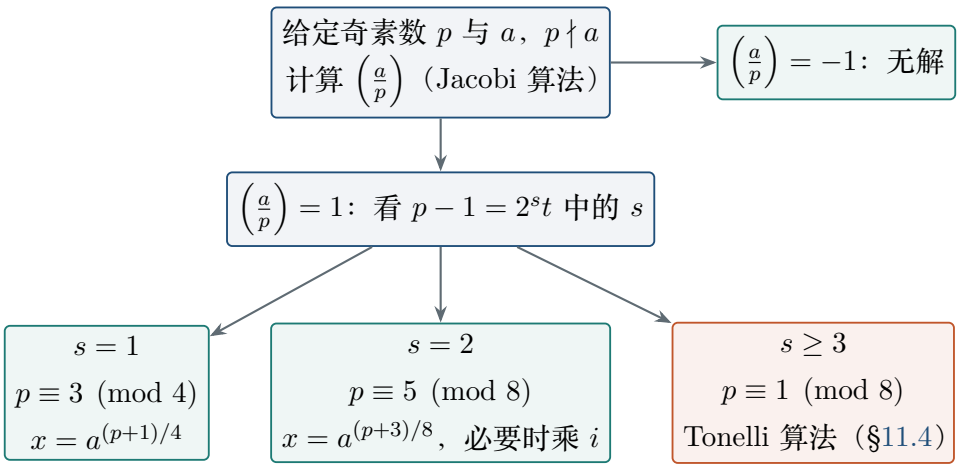


图 11.1　求 $x^2 \equiv a \pmod p$ 的路线图。s 越大，2 在 $p-1$ 中的幂次越高，问题越难；Tonelli 算法统一处理所有情形。

11.4　Tonelli 算法

11.4.1　准备：阶里 2 的幂次

从现在起固定奇素数 p，写

$$p - 1 = 2^s\, t, \qquad t \text{ 为奇数}, \ s \geq 1.$$

模 p 的每个非零元 x 的阶 $\operatorname{ord} x$ 整除 $p-1$，因此 $\operatorname{ord} x = 2^j \cdot (t \text{ 的某个因数})$，$0 \leq j \leq s$。整个算法只关心其中 2 的幂次 j。本讲把它记作 $\ell(x) = j$，叫作 x 的**层数**：$\operatorname{ord} x$ 恰好被 $2^{\ell(x)}$ 整除，不被 $2^{\ell(x)+1}$ 整除。层数可以用「看 ± 1」来测量。

> **引理 11.5（层数测试）**　对 $0 \leq k < s$，$x^{2^k t} \equiv 1$ 当且仅当 $\ell(x) \leq k$；$\ell(x) = k+1$ 时 $x^{2^k t} \equiv -1$。

证明　$x^{2^k t} \equiv 1$ 当且仅当 $\operatorname{ord} x \mid 2^k t$。$\operatorname{ord} x = 2^{\ell(x)} u$，$u \mid t$ 为奇数，所以这又等价于 $2^{\ell(x)} \mid 2^k$，即 $\ell(x) \leq k$。若 $\ell(x) = k+1$，则 $y = x^{2^k t}$ 满足 $y^2 = x^{2^{k+1} t} \equiv 1$ 而 $y \not\equiv 1$，只能 $y \equiv -1$。 ■

取 $k = s-1$：$x^{(p-1)/2} = x^{2^{s-1} t} \equiv 1$ 当且仅当 $\ell(x) \leq s-1$。和 Euler 判别法对照，得到**非剩余恰好是层数为 s 的元素**，剩余的层数都 $\leq s-1$。用原根 g 看得更清楚：$\operatorname{ord}(g^e) = \frac{p-1}{\gcd(e, p-1)}$，所以 e 中 2 的幂次每多一个，层数就少一层；e 为奇数（非剩余）时层数是 s。层数为 j 的元素共有 $2^{j-1} t$ 个（$j \geq 1$），层数为 0 的有 t 个。

> **引理 11.6（MIT 讲义引理 44 的推广）**　若 $\ell(a) = \ell(b) = j \geq 1$，则 $\ell(ab) < j$。

证明　由引理 11.5（取 $k = j-1$），$a^{2^{j-1} t} \equiv -1$，$b^{2^{j-1} t} \equiv -1$，相乘得 $(ab)^{2^{j-1} t} \equiv 1$，再由同一

引理，$\ell(ab) \le j-1$。 ■

这条引理是整个算法的发动机：**两个同层的元素相乘，层数一定下降**。降多少不确定，可能一次降好几层。

命题 11.7　设 n 是模 p 的非剩余，$c = n^t$，则 $\operatorname{ord} c = 2^s$。因此对 $0 \le i \le s$，c^{2^i} 与 c^{-2^i} 的层数都是 $s-i$。

证明　$c^{2^s} = n^{2^s t} = n^{p-1} \equiv 1$，所以 $\operatorname{ord} c$ 整除 2^s，是 2 的幂。由 Euler 判别法，$c^{2^{s-1}} = n^{(p-1)/2} \equiv \left(\frac{n}{p}\right) = -1 \not\equiv 1$，所以 $\operatorname{ord} c$ 不整除 2^{s-1}，只能是 2^s。于是 $\operatorname{ord}(c^{2^i}) = 2^{s-i}$；$c^{-2^i}$ 是它的逆元，阶相同。 ■

MIT 讲义用原根证明：$n = g^k$，k 必为奇数（否则 n 是剩余），$c^{2^{s-1}} = (g^{(p-1)/2})^k \equiv (-1)^k = -1$。两种写法是一回事。

11.4.2　算法与正确性

想法是维持一个不变式

$$a \equiv A\, b^2 \pmod p,$$

开始时 $A = a$，$b = 1$。如果最后 A 的层数降到 0，则 $A^t \equiv 1$，$A^{(t+1)/2}$ 就是 A 的平方根，从而 $x = bA^{(t+1)/2}$ 是 a 的平方根。降层用引理 11.6：当 A 的层数恰好是 $s-i$ 时，乘上同层的 c^{-2^i}。因为 $c^{-2^i} = (c^{-2^{i-1}})^2$ 是平方，把 b 同时乘上 $c^{2^{i-1}}$，不变式就保持了。

算法 11.8（Tonelli 算法）　输入奇素数 p 与 a，$p \nmid a$。输出 $x^2 \equiv a \pmod p$ 的一个解，或报告无解。

(1) 若 $a^{(p-1)/2} \not\equiv 1$，报告无解，停止。

(2) 写 $p-1 = 2^s t$，t 为奇数。找一个非剩余 n，令 $c = n^t$，$A = a$，$b = 1$。

(3) 对 $i = 1, 2, \dots, s-1$：若 $A^{2^{s-1-i}t} \equiv -1$，则令 $A \leftarrow A\,c^{-2^i}$，$b \leftarrow b\,c^{2^{i-1}}$；否则不动。

(4) 输出 $x = b\,A^{(t+1)/2}$。

定理 11.9　算法 11.8 的输出满足 $x^2 \equiv a \pmod p$。而且第 i 步之后 $\ell(A) \le s-1-i$，第 i 步的检验值 $A^{2^{s-1-i}t}$ 只可能是 ± 1。

证明　*不变式*。开始时 $Ab^2 = a$。修正一步后 $(Ac^{-2^i})(bc^{2^{i-1}})^2 = Ab^2 c^{-2^i} c^{2^i} = Ab^2$，不变。

层数。第 1 步之前 $A = a$ 是剩余，$\ell(A) \le s-1$。设第 i 步之前 $\ell(A) \le s-i$。由引理 11.5（取 $k = s-1-i$），检验值只能是 ± 1：若为 1，则 $\ell(A) \le s-1-i$，不动即可；若为 -1，则 $\ell(A) = s-i$，而由命题 11.7，$\ell(c^{-2^i}) = s-i$，由引理 11.6，新的 A 层数 $< s-i$。两种情况下，第 i 步之后都有 $\ell(A) \le s-1-i$。

结论。第 $s-1$ 步之后 $\ell(A)=0$，即 $\operatorname{ord} A$ 是奇数，整除 t，所以 $A^t\equiv 1$。（$s=1$ 时没有循环，$\ell(A)\le 0$ 已在开始时成立。）于是

$$x^2=b^2A^{t+1}=(Ab^2)\cdot A^t\equiv a\cdot 1=a \pmod p.$$ ■

证明里最关键的是「检验值只能是 ±1，而且 -1 恰好意味着 A 和修正因子同层」。算法本身不需要知道层数，层数只出现在证明里。

注 11.2　$s=1$ 时循环为空，输出 $x=a^{(t+1)/2}=a^{(p+1)/4}$，就是命题 11.3，而且根本用不到 n。$s=2$ 时只有一步，检验值 $A^t=a^{(p-1)/4}$；若取 $n=2$，则 $c=2^{(p-1)/4}=i$，输出就是命题 11.4 的 x_0 或 $\pm ix_0$（修正时 $c^{-2}=i^{-2}=-1$，A 变成 $-a$）。所以前面两个公式都是 Tonelli 算法的特例。

11.4.3 两个完整的例子

■ 例 11.7　习题集 5 第 1 题　用 Tonelli 算法解 $x^2\equiv 21 \pmod{41}$。

解　*第 1 步，判断有解。*$\left(\frac{21}{41}\right)=\left(\frac{41}{21}\right)=\left(\frac{20}{21}\right)=\left(\frac{-1}{21}\right)=1$（$41\equiv 1 \pmod 4$，翻转不变号；$21\equiv 1 \pmod 4$）。有解。

第 2 步，准备。$p-1=40=2^3\cdot 5$，$s=3$，$t=5$。$41\equiv 1 \pmod 8$，2 是剩余，不能用。$\left(\frac{3}{41}\right)=\left(\frac{41}{3}\right)=\left(\frac{2}{3}\right)=-1$，取 $n=3$，$c=3^5=243=6\cdot 41-3\equiv -3$。循环有 $i=1,2$ 两步，检验指数分别是 $2^1\cdot 5=10$ 和 5。要用到 $c^{-2}=9^{-1}$：$9\cdot 32=288=7\cdot 41+1$，所以 $c^{-2}\equiv 32\equiv -9$。

第 3 步，循环。$i=1$，$A=21$：$21^2=441\equiv -10$，$21^4\equiv 100\equiv 18$，$21^8\equiv 324\equiv -4$，$21^{10}=21^8\cdot 21^2\equiv(-4)(-10)=40\equiv -1$。要修正：

$$A\leftarrow 21\cdot(-9)=-189\equiv 16,\qquad b\leftarrow 1\cdot c=-3.$$

核对不变式：$16\cdot(-3)^2=144\equiv 21$。$i=2$，$A=16$：$16^2=256\equiv 10$，$16^4\equiv 100\equiv 18$，$16^5\equiv 18\cdot 16=288\equiv 1$。不动。

第 4 步，输出。$(t+1)/2=3$，$16^3=16\cdot 10=160\equiv -4$，$x=b\cdot A^3\equiv(-3)(-4)=12$。检验 $12^2=144=3\cdot 41+21$。解为 $x\equiv\pm 12$，即 $12,29 \pmod{41}$。 ♦

这个例子里 21 的层数是 2（$21^{10}\equiv -1$），修正一次就落到层数 0（$16^5\equiv 1$），第 2 步无事可做。下一个例子 $s=5$，能看到层数一级一级往下走，也能看到一次跳过两级。

■ 例 11.8　用 Tonelli 算法解 $x^2\equiv 2 \pmod{97}$。

解　$97\equiv 1 \pmod 8$，所以 $\left(\frac{2}{97}\right)=1$，有解，而且 $n=2$ 不能用。$\left(\frac{3}{97}\right)=\left(\frac{97}{3}\right)=\left(\frac{1}{3}\right)=1$，也不行；$\left(\frac{5}{97}\right)=\left(\frac{97}{5}\right)=\left(\frac{2}{5}\right)=-1$，取 $n=5$。$p-1=96=2^5\cdot 3$，$s=5$，$t=3$，$c=5^3=125\equiv 28$。先把要用的 c 的幂与逆元算好：

$$c^2\equiv 8,\quad c^4\equiv 64,\quad c^8\equiv 22,\quad c^{16}\equiv -1;\qquad c^{-2}\equiv 85,\quad c^{-4}\equiv 47,\quad c^{-16}\equiv -1.$$

（例如 $8\cdot 85=680=7\cdot 97+1$，$64\cdot 47=3008=31\cdot 97+1$。$c^{16}\equiv -1$ 印证了 $\operatorname{ord} c=32$。）

循环 $i=1,2,3,4$，检验指数依次为 $24,12,6,3$。逐步记录在表 11.3 中。几个检验值的算法：$2^8=256\equiv 62$，$2^{16}\equiv 62^2\equiv 61$，$2^{24}\equiv 61\cdot 62\equiv -1$；$73^2\equiv -6$，$73^4\equiv 36$，$73^8\equiv 35$，$73^{12}\equiv 35\cdot 36\equiv -1$；$36^2\equiv 35$，$36^4\equiv 61$，$36^6\equiv 61\cdot 35\equiv 1$，$36^3\equiv 36\cdot 35\equiv -1$。

i	当前 A	检验指数	检验值	动作	新 A	新 b	ℓ(新A)
开始				$A=2$，$b=1$	2	1	4
1	2	24	-1	$A\cdot c^{-2}$，$b\cdot c$	73	28	3
2	73	12	-1	$A\cdot c^{-4}$，$b\cdot c^{2}$	36	30	1
3	36	6	1	不动	36	30	1
4	36	3	-1	$A\cdot c^{-16}$，$b\cdot c^{8}$	61	78	0

表 11.3　$x^2\equiv 2 \pmod{97}$ 的 Tonelli 过程（$n=5$，$c=28$）。最后一列是为了理解而附上的，算法本身不需要它。

最后 $A=61$，$A^{(t+1)/2}=61^2=3721\equiv 35$，$x=b\cdot 35=78\cdot 35=2730\equiv 14$。检验 $14^2=196=2\cdot 97+2$。解为 $x\equiv\pm 14$，即 $14,83 \pmod{97}$。也可以核对不变式：$61\cdot 78^2\equiv 61\cdot 70=4270\equiv 2$。 ♦

图 11.2 把表 11.3 画成「下梯子」。模 97 的 96 个非零元按层数分成 6 层，越往下人越少。A 从第 4 层出发，每一步之后都必须落在红色虚线（第 i 步之后 $\ell\le 4-i$）之下；第 2 步一次跳过了两层，第 3 步因此无事可做。

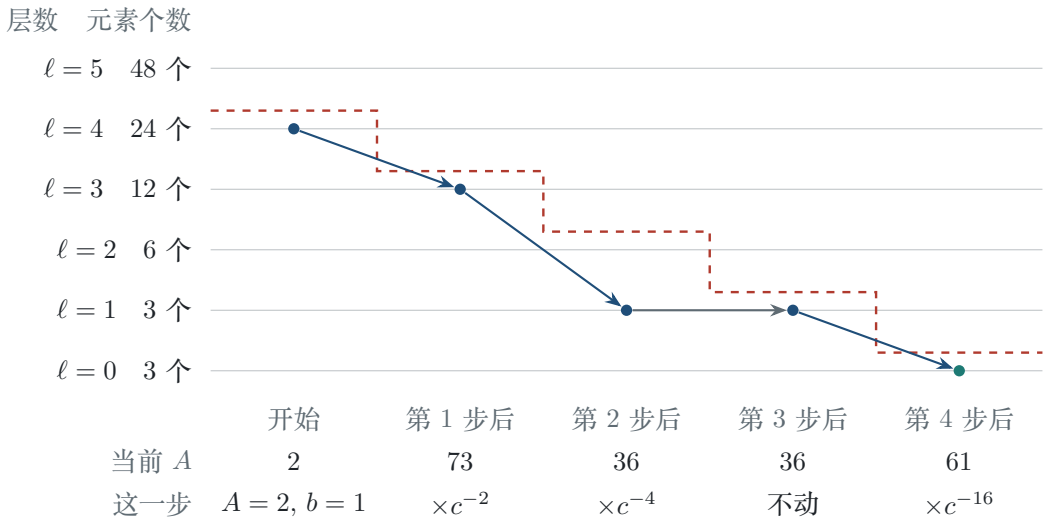


图 11.2　$x^2\equiv 2 \pmod{97}$ 时 A 的层数变化。左侧是每层的元素个数（层数 5 的 48 个恰是全部非剩余）。红色虚线是定理 11.9 保证的上界，蓝点是实际的层数；下面两行是每一步之后的 A 和这一步做的乘法。第 2 步一次下降两层。

第 1 步不能省　如果 a 是非剩余，第 3 步的检验值可能根本不是 ±1，输出也毫无意义。例如对 $a=3$，$p=41$（$3^{20}\equiv-1$，非剩余）硬跑算法：两步的检验值 $3^{10}\equiv9$ 和 $3^5\equiv38$ 都不是 ±1，于是都「不动」，输出 $3^3=27$，而 $27^2\equiv32\not\equiv3$。所以先判断 $\left(\frac{a}{p}\right)=1$，最后再平方检验一次，检验只需一次乘法。

想一想　在例 11.7 中，如果误取 $n=2$（$c=2^5=32$），会发生什么？

答　$41\equiv1\pmod 8$，2 是剩余，$c=32$ 的阶只有 4（$32^2\equiv-1$），达不到 2^3。第 1 步修正因子 $c^{-2}\equiv-1$ 的层数是 1，与 $A=21$（层数 2）不同层，乘完 $A=-21\equiv20$ 层数没降。第 2 步检验值 $20^5\equiv32$ 不是 ±1，算法「不动」，最后输出 $x=32\cdot20^3\equiv37$，而 $37^2\equiv16\not\equiv21$。修正因子要和 A 同层，这要求 $\operatorname{ord}c=2^s$，也就是 n 必须是非剩余。

直观理解　用原根写 $A=g^e$，则 $\ell(A)=s-\min\{s,v_2(e)\}$；当 $e=0$ 时约定 $v_2(0)=+\infty$，所以 $\ell(1)=0$。再写 $c=g^u$，由 $\operatorname{ord}c=2^s$ 可知 u 为奇数。第 i 步之前，e 已被 2^i 整除；检验值为 -1 恰好表示 $e/2^i$ 为奇数。乘上 c^{-2^i} 后，指数变成 $e-u2^i$，除以 2^i 后是两个奇数之差，因而多出一个因子 2。算法无需计算离散对数 e,u，只需计算检验值和修正因子。

11.5　算法有多快，难在哪里

11.5.1　运算量

算法 11.8 的每次乘方都用反复平方，代价是 $O(\log p)$ 次模 p 乘法。第 1 步一次乘方；循环 $s-1\le\log_2 p$ 步，每步一次乘方加两次乘法；c 的幂与逆元可以预先算好。合计 $O(\log^2 p)$ 次模乘法，每次模乘法对 $\log p$ 位的数是 $O(\log^2 p)$ 次位运算，总共 $O(\log^4 p)$ 次位运算，是 p 的位数的多项式。对 300 位十进制的素数，$\log_2 p\approx1000$，乘法次数约为 10^6 量级，仍是可实际处理的规模；而逐个试 x 要 10^{299} 次。

11.5.2　怎样找非剩余

唯一没说清的是第 2 步「找一个非剩余 n」。模 p 的非零类恰有一半是非剩余（第 9 讲），所以**随机取一个数，有一半的机会是非剩余**，用 Jacobi 算法检验只要 $O(\log p)$ 步。连续 k 次都取到剩余的概率是 2^{-k}；$k=\log_2 p$ 时这个概率是 $1/p$。平均只要试 2 次。这样得到的是一个**随机化算法**（randomized algorithm）：答案总是对的（最后可以检验），只有运行时间是随机的。如果不想要随机性，最自然的办法是按 $2,3,4,\dots$ 的顺序逐个检验，碰到第一个非剩余就停。记它为 n_p，称为模 p 的**最小非剩余**（least quadratic nonresidue）。问题变成：n_p 最大能有多大？

命题 11.10（最小非剩余的两个初等性质）　设 p 为奇素数，n_p 是模 p 的最小正非剩余。

(1) n_p 是素数。

(2) $n_p<\sqrt{p}+1$。

证明　(1) 若 $n_p=ab$，$1<a,b<n_p$，则 a,b 比 n_p 小，都是剩余，乘积 n_p 也是剩余，矛盾。

(2) 记 $n=n_p$，取 $m=\lfloor p/n\rfloor+1$，即使 $mn>p$ 的最小正整数。于是 $0<mn-p\le n$；又 p 是素数，$n\nmid p$，所以 $0<mn-p<n$。$mn-p$ 比 n 小，是剩余；而 $mn\equiv mn-p\pmod p$，所以 $\left(\frac{m}{p}\right)\left(\frac{n}{p}\right)=1$，$\left(\frac{n}{p}\right)=-1$ 迫使 $\left(\frac{m}{p}\right)=-1$。$m$ 也是非剩余，由最小性 $m\ge n$。于是 $n\le m<p/n+1$，即 $n^2-n<p$，得 $n<\frac12+\sqrt{p+\frac14}<\sqrt p+1$。 ■

$\sqrt p+1$ 对算法毫无用处：p 有 300 位时它有 150 位。实际情况要好得多。$p\equiv\pm3\pmod 8$ 时 $n_p=2$，这已经是一半的素数；表 11.4 列出 $p<10^6$ 范围内 n_p 创新高的位置。

p	3	7	23	71	311	479	1559	5711	10559	18191	31391	366791
n_p	2	3	5	7	11	13	17	19	23	29	31	43
$\sqrt p+1$	2.7	3.6	5.8	9.4	18.6	22.9	40.5	76.6	103.8	135.9	178.2	606.6
$2(\ln p)^2$	2.4	7.6	19.7	36.3	65.9	76.2	108.1	149.7	171.7	192.4	214.4	328.3

表 11.4　$p<10^6$ 时最小非剩余 n_p 的历次新高。每个 n_p 都是素数，增长远比 $\sqrt p$ 慢。

在**广义 Riemann 假设**（generalized Riemann hypothesis, GRH）之下，Ankeny（1952）证明 $n_p=O(\log^2 p)$，Bach（1990）给出了具体常数：$p>3$ 时 $n_p<2(\ln p)^2$。如果 GRH 成立，逐个检验 $n=2,3,\dots$ 就是一个确定性的多项式时间算法。但不假设 GRH 时，目前最好的无条件上界（Burgess，1957）大约是 $p^{0.152}$，仍是 p 的位数的指数。所以至今**没有人能无条件地给出求模 p 平方根的确定性多项式时间算法**，尽管随机化算法在实践中毫无困难。平均来看，n_p 很小：Erdős 证明 n_p 的平均值趋于 $\sum_k q_k/2^k\approx3.67$（$q_k$ 是第 k 个素数）；$p<10^6$ 时实算的平均值是 3.65。

想一想　哪些素数满足 $n_p\ge5$？为什么表 11.4 里 $n_p=5$ 第一次出现在 $p=23$？

答　$n_p\ge5$ 当且仅当 $2,3,4$ 都是剩余。4 总是剩余；2 是剩余当且仅当 $p\equiv\pm1\pmod 8$；3 是剩余当且仅当 $p\equiv\pm1\pmod{12}$（第 9 讲）。两者同时成立当且仅当 $p\equiv\pm1\pmod{24}$。这样的最小素数是 23，而 $\left(\frac{5}{23}\right)=\left(\frac{23}{5}\right)=\left(\frac35\right)=-1$，所以 $n_{23}=5$。

11.5.3　模合数开平方等价于分解

模合数 $n=pq$（$p\ne q$ 为奇素数）时，先分别模 p、模 q 开平方，再用中国剩余定理（第 5 讲）拼起来。$\gcd(a,n)=1$ 时，$x^2\equiv a\pmod n$ 的解数是 $\left(1+\left(\frac{a}{p}\right)\right)\left(1+\left(\frac{a}{q}\right)\right)$：模 p 与模 q 都有解时恰有 4 个解。

■ **例 11.9**　解 $x^2\equiv386\pmod{437}$，已知 $437=19\cdot23$。

解　模 19。$386=20\cdot19+6$，$a\equiv6$。$19\equiv3\pmod4$，用命题 11.3：$6^2\equiv-2$，$6^4\equiv4$，$x\equiv6^5\equiv24\equiv5$。检验 $5^2=25\equiv6$。

模 23。$386=16\cdot23+18$，$a\equiv18\equiv-5$。$23\equiv3\pmod4$：$(-5)^2\equiv2$，$x\equiv(-5)^6\equiv2^3=8$。检验 $8^2=64=2\cdot23+18$。

拼接。找 $e_1\equiv1\pmod{19}$，$e_1\equiv0\pmod{23}$：$23\equiv4\pmod{19}$，$4\cdot5=20\equiv1$，取

$e_1 = 23 \cdot 5 = 115$。同理 $19 \cdot 17 = 323 = 14 \cdot 23 + 1$，取 $e_2 = 323$。于是

$$x \equiv (\pm 5) \cdot 115 + (\pm 8) \cdot 323 \pmod{437}.$$

两个都取正号：$575 + 2584 = 3159 = 7 \cdot 437 + 100$，得 100；取 $+5$、-8：$575 - 2584 = -2009 \equiv 176$。四个解是 ± 100，± 176，即 $100, 176, 261, 337$。检验 $100^2 = 10000 = 22 \cdot 437 + 386$，$176^2 = 30976 = 70 \cdot 437 + 386$。 ♦

反过来看这个例子：$100^2 \equiv 176^2 \pmod{437}$，但 $176 \not\equiv \pm 100$。于是 $437 \mid (176 - 100)(176 + 100) = 76 \cdot 276$，却不整除任何一个因子，因此 $\gcd(76, 437) = 19$ 和 $\gcd(276, 437) = 23$ 都是真因子。这说明**会开平方就会分解**。

命题 11.11　设 $n = pq$，$p \neq q$ 为奇素数。若有一个算法 $\mathcal{A}$，对每个与 n 互素的平方 a 输出它模 n 的某个平方根，则可以用期望至多 2 轮试验分解 n；每轮调用至多一次 $\mathcal{A}$，并计算至多两次最大公因数。

证明　随机取 $x \in [1, n-1]$。若 $\gcd(x, n) > 1$，已经得到因子。否则令 $a \equiv x^2 \pmod{n}$，$y = \mathcal{A}(a)$。a 的四个平方根中，$\mathcal{A}$ 只看到 a，不知道我们取的是哪一个；而 x 在这四个根中是等可能的，所以 $y \equiv \pm x$ 的概率是 $1/2$。若 $y \not\equiv \pm x \pmod{n}$，由 $y^2 \equiv x^2$ 得 $y \equiv \pm x$ 分别模 p、模 q 成立，且两个素数上的符号不同，比如 $y \equiv x \pmod{p}$，$y \equiv -x \pmod{q}$。这时 $p \mid x - y$；而 $q \nmid x - y$，否则 $q \mid 2x$。所以 $\gcd(x - y, n) = p$。在互素分支中成功的概率为 $1/2$；非互素分支直接成功，所以每轮成功概率至少为 $1/2$，期望轮数至多为 2。 ■

另一个方向更简单：知道 p, q 就能用 Tonelli 算法和中国剩余定理开平方。所以

$$\boxed{n = pq \text{ 时，模} n \text{ 开平方} \iff n \text{ 分解}}$$

Rabin（1979）据此设计了一种公钥密码：明文 m 加密为 $m^2 \bmod n$，解密就是开平方。破解它可以证明和分解 n 一样难；对 RSA（第 6 讲），这样的等价性至今没有证明。

11.6　连续的两个平方剩余

把 0 也算作平方。模 11 的平方类是 $\{0, 1, 3, 4, 5, 9\}$。逐个看这六个数的下一个数，只有 $0 \to 1$、$3 \to 4$、$4 \to 5$ 仍在集合中，因此答案是 3。这个结果不是简单的「约四分之一」：相邻两项的平方性之间有一个可精确计算的关系。

记 $\chi(x) = \left(\frac{x}{p}\right)$，并沿用 $\chi(0) = 0$。先算一项关键的字符和。

引理 11.12　对任意奇素数 p，

$$\sum_{x \bmod p} \chi\big(x(x+1)\big) = -1.$$

证明　数方程 $y^2 = x(x+1)$ 的解对 (x,y)。令 $u = 2x+1$、$v = 2y$，则 $u^2 - v^2 = 1$，等价于 $(u-v)(u+v) = 1$。任取 $r = u - v \in \mathbb{F}_p^\times$，必须有 $u+v = r^{-1}$；因 2 可逆，这两个等式唯一确定 u, v，也唯一确定 x, y。所以解对恰有 $p-1$ 个。

固定 x 时，$y^2 = x(x+1)$ 有 $1 + \chi(x(x+1))$ 个解。按 x 求和便得

$$p + \sum_{x \bmod p} \chi\big(x(x+1)\big) = p - 1,$$

结论成立。 ∎

> **定理 11.13（连续平方剩余的个数）**　设 p 是奇素数。模 p 的剩余类 x 中，使 x 与 $x+1$ 都是平方（允许为 0）的恰有
>
> $$\frac{p + 2 + \left(\frac{-1}{p}\right)}{4}$$
>
> 个。

思路　对 $x \neq 0$，x 是平方的指示值为 $\frac{1+\chi(x)}{2}$。因此把两个指示值相乘后求和即可；但 $x = 0$ 和 $x = -1$ 各有一个零因子，必须单独计入，不能直接把乘积公式用在所有 x 上。

证明　记 $e = \chi(-1)$。$x = 0$ 一定计入，贡献 1；$x = -1$ 的第二项为 0，第一项为平方的指示值，贡献 $(1+e)/2$。对其余 $p-2$ 个 x，两个指示值的乘积为

$$\frac{1}{4}\big(1 + \chi(x) + \chi(x+1) + \chi(x)\chi(x+1)\big).$$

逐项求和：常数项给 $p-2$；由于 $\sum_x \chi(x) = 0$，去掉 $x = -1$ 后得到 $-e$；同理去掉 $x = 0, -1$ 后，$\sum \chi(x+1) = -1$。最后

$$\chi(x)\chi(x+1) = \chi\big(x(x+1)\big),$$

零点本来就不贡献，故由引理 11.12 得到 -1。于是

$$N = 1 + \frac{1+e}{2} + \frac{(p-2) - e - 1 - 1}{4} = \frac{p+2+e}{4}.$$

∎

■ **例 11.10　模 11 的直接计数**　列出模 11 中使 x 与 $x+1$ 都为平方的所有 x，并与定理 11.13 对照。

解　平方类为 $\{0, 1, 3, 4, 5, 9\}$。逐项检查：$0 \to 1$、$3 \to 4$、$4 \to 5$ 合格；$1 \to 2$、$5 \to 6$、$9 \to 10$ 不合格，其余 x 的第一项已经不是平方。因此答案是 $x \equiv 0, 3, 4 \pmod{11}$，共有 3 个。

定理给出的数也是

$$\frac{11 + 2 + \left(\frac{-1}{11}\right)}{4} = \frac{11 + 2 - 1}{4} = 3,$$

两种方法一致。图 11.3 将这三个「成功跳一步」画出来：蓝色节点是平方类，青绿色箭头表示 $x \mapsto x+1$ 仍留在平方类中。 ♦



蓝色节点：平方类；青绿色箭头：连续两个都是平方

图 11.3　模 11 的连续平方。恰有 $0\to1$、$3\to4$、$4\to5$ 三条箭头同时落在蓝色节点上。

从这个小例子还能看出两个边界点。$x=0$ 必须计入，因为 0 是平方；$x=-1$ 是否计入取决于 -1 本身是不是平方，这正是公式中 $\left(\frac{-1}{p}\right)$ 出现的原因。

■ 例 11.11　三个小素数的比较　分别求 $p=3,5,7$ 时连续平方的个数。

解　直接代入公式：

p	$\left(\frac{-1}{p}\right)$	$\dfrac{p+2+\left(\frac{-1}{p}\right)}{4}$	合格的x
3	-1	1	0
5	1	2	$0,4$
7	-1	2	$0,1$

例如模 5 的平方类为 $0,1,4$，只有 $0\to1$ 和 $4\to0$；模 7 的平方类为 $0,1,2,4$，只有 $0\to1$ 和 $1\to2$。公式在最小的奇素数 3 上也成立。 ♦

想一想　如果把「平方」改成「非零平方」，公式会不会不变？

答　不会。原公式把 $x=0$ 和 $x+1=0$ 都视为平方，所以必须分别处理。若只允许非零平方，就要从原来的计数中删去 $x=0$，再在 $x=-1$ 为平方时删去 $x=-1$；因此个数为

$$\frac{p+2+\left(\frac{-1}{p}\right)}{4}-1-\frac{1+\left(\frac{-1}{p}\right)}{2}=\frac{p-4-\left(\frac{-1}{p}\right)}{4}.$$

例如 $p=11$ 时只剩 $x=3,4$ 两个。

前后联系　习题集 5 第 9 题对 $k\not\equiv0\pmod p$ 证明 $y^2\equiv x^2+k\pmod p$ 的解数为 $p-1$，再求出 $\sum_x\left(\frac{x^2+k}{p}\right)=-1$。本节取 $k=1$，把同一个字符和用于两个相邻平方的计数；若要求 $x-1,x,x+1$ 全是平方，就会出现三次乘积，计算明显更复杂。

本讲小结

- Jacobi 符号的值为 -1 时，分子一定不是模分母的平方；值为 1 时不能直接下结论。
- 对固定的奇素数 p，若 $p-1=2^s t$ 且 t 为奇数，Tonelli 算法用一个非剩余构造阶为 2^s 的修正因子，逐步降低 A 的层数。
- Tonelli 算法维持 $a \equiv Ab^2 \pmod p$；当 $A^t \equiv 1$ 时，$bA^{(t+1)/2}$ 就是 a 的平方根。
- 模 $n=pq$ 开平方可以借助 Tonelli 和中国剩余定理完成；反过来，拿到同一个平方的两个非相反平方根就能用最大公因数分解 n。
- 对 $\chi(x)=\left(\frac{x}{p}\right)$，有 $\sum_{x \bmod p} \chi(x(x+1)) = -1$。
- 连续两个数都为平方的剩余类恰有 $\dfrac{p+2+\left(\frac{-1}{p}\right)}{4}$ 个；其中 $x=0$ 总计入，$x=-1$ 是否计入由 $p \bmod 4$ 决定。

练习

11.1 ★ 用定理 11.13 求模 17 中连续两个平方剩余的个数，并列出所有 x。

11.2 ★★ 用 Tonelli 算法求 $x^2 \equiv 10 \pmod{41}$ 的全部解。取非剩余 $n=3$，逐步记录检验值、A 与 b，并检验输出。

11.3 ★ 证明：对任意奇素数 p，$\sum_{x \bmod p}\left(\frac{x}{p}\right)=0$。

11.4 ★★ 仿照引理 11.12，证明对 $k \not\equiv 0 \pmod p$，$\sum_{x \bmod p}\left(\frac{x^2+k}{p}\right)=-1$。

11.5 ★★ 求模 13 中 x、$x+1$、$x+2$ 都是平方的所有 x。

11.6 ★★ 设 $p \equiv 3 \pmod 4$。证明连续平方的个数为 $(p+1)/4$，并解释为什么 $x=-1$ 不贡献。

11.7 ★★★ 证明：若 $p>3$，则至少有两个模 p 的剩余类 x 使 x 与 $x+1$ 都是平方。

11.8 ★★ 用 $p \equiv 5 \pmod 8$ 的公式求 $x^2 \equiv 10 \pmod{13}$ 的全部解。说明初值是否需要修正，并把所得根平方检验。

11.9 ★★★ 已知 $13^2 \equiv 20^2 \pmod{77}$。用最大公因数分解 77；再解释若只拿到 13 和 64 这对平方根，为什么同样的做法得不到真因子。

练习答案与提示

11.1 $17 \equiv 1 \pmod 4$，所以个数为 $(17+3)/4=5$。平方类为 $0,1,2,4,8,9,13,15,16$；逐项检查得 $x=0,1,8,15,16$。

11.2 $40=2^3\cdot 5$，故 $s=3, t=5, c=3^5 \equiv 38$。初始 $(A,b)=(10,1)$；两个检验值分别为 $10^{10}\equiv 1$ 与 $10^5 \equiv 1$，两步均不修正，始终有 $Ab^2 \equiv 10$。输出 $x=10^3 \equiv 16$，而 $16^2=256 \equiv 10 \pmod{41}$。另一根为 $-16 \equiv 25$；模奇素数的非零平方恰有两个根，所以全部解为 $16, 25 \pmod{41}$。

11.3 非零平方和非平方各有 $(p-1)/2$ 个，对应的 Legendre 符号分别为 $+1$ 与 -1，恰好抵消；$x=0$ 项为 0。

11.4 数方程 $y^2=x^2+k$。令 $u=y-x$、$v=y+x$，则 $uv=k$。任取非零 u，唯一确定 $v=k/u$，再由 2 可逆唯一恢复 x,y，所以共有 $p-1$ 个解。固定 x 的解数为 $1+\left(\frac{x^2+k}{p}\right)$，求和即得字符和为 -1。

11.5 模 13 的平方类为 $\{0,1,3,4,9,10,12\}$。逐项检查，只有 $x=12$ 合格，因为 $12,0,1$ 都是平方类；其余相邻三项至少有一项不是平方。

11.6 $p\equiv 3 \pmod 4$ 时 $\left(\frac{-1}{p}\right)=-1$，公式化为 $(p+1)/4$；$x=-1$ 的第一项是非平方，第二项为 0，所以不贡献。

11.7 若 $p\equiv 3 \pmod 4$，公式为 $(p+1)/4$；若 $p\equiv 1 \pmod 4$，公式为 $(p+3)/4$。两种情形下，$p>3$ 时个数都至少为 2。

11.8 $13\equiv 5 \pmod 8$，$10\equiv 6^2 \pmod{13}$，确为非零平方。初值 $x_0=10^2\equiv 9$，但 $10^3\equiv 12\equiv -1$，所以需要修正。取 $i=2^3\equiv 8$，$i^2\equiv -1$，得 $x=ix_0\equiv 72\equiv 7$。检验 $7^2\equiv 10$，另一根为 6，故全部解为 $6,7 \pmod{13}$。

11.9 两个平方都模 77 余 15，且 $\gcd(20-13,77)=\gcd(7,77)=7$、$\gcd(20+13,77)=\gcd(33,77)=11$，得到 $77=7\cdot 11$。但 $64\equiv -13 \pmod{77}$，这时 $\gcd(64+13,77)=77$、$\gcd(64-13,77)=1$，只得到平凡因子。这说明需要的是同一个平方的两个非相反根。

第 12 讲

分圆多项式与模 n 余 1 的素数

Cyclotomic Polynomials, Primes Congruent to 1 mod n

对应材料：MIT 18.781 Lecture 12 讲义（Cyclotomic Polynomials, Primes Congruent to 1 mod n；命题 45、引理 46、定理 47）；习题集 6 第 10 题（Bonus，仅作预告）；习题集 5 第 6、8 题（改编为对比表与练习）。

本讲目标

(1) 会把 n 次单位根画成正 n 边形的顶点，知道 $e^{2\pi ij/n}$ 的阶是 $\dfrac{n}{(j,n)}$，本原 n 次单位根恰有 $\varphi(n)$ 个。

(2) 掌握分圆多项式 $\Phi_n(x)$ 的定义和基本恒等式 $x^n-1=\prod_{d\mid n}\Phi_d(x)$，会用它递推地算出 Φ_n。

(3) 理解 Φ_n 为什么是整系数的（首一多项式的除法），为什么 $n\geq 2$ 时它是回文的，并由此得到 $\Phi_n(0)=1$。

(4) 掌握本讲的核心事实：若素数 $p\nmid n$ 且 $p\mid\Phi_n(a)$，则 $\operatorname{ord}_p(a)=n$，从而 $p\equiv 1\pmod n$。会用它分解 $2^{11}-1$、$2^{32}+1$ 这类数。

(5) 理解定理「模 n 余 1 的素数有无穷多个」的 Euclid 式证明，并把它和第 2、4、9 讲中 $4k+3$、$4k+1$、$3k+1$、$8k+7$ 型素数的证明放在一起比较。

12.1 单位根与本原单位根

第 7、8 讲研究的是**模 p 的阶**：使 $a^h\equiv 1\pmod p$ 的最小正整数 h。本讲先离开同余，到复数里问同一个问题：哪些复数 z 满足 $z^n=1$？它们的「阶」是多少？回答这个问题会自然地造出一族整系数多项式 $\Phi_n(x)$。把 Φ_n 再拿回到模 p 的世界，就能证明 Dirichlet 定理的一个特殊情形：**模 n 余 1 的素数有无穷多个**。

准备。本节使用复数的四则运算、模与辐角，以及三角函数的加法公式；若尚不熟悉，应先补学这些概念。以下回顾本讲需要的极形式与乘法规则。

每个非零复数都可以写成 $z=re^{i\theta}$，其中 $r=|z|$ 是长度，θ 是辐角，$e^{i\theta}=\cos\theta+i\sin\theta$。两个复数相乘时**长度相乘、辐角相加**：$re^{i\theta}\cdot se^{i\psi}=rs\,e^{i(\theta+\psi)}$。特别地，乘以单位圆上的一个数 $e^{i\psi}$，就是把平面绕原点旋转角度 ψ。

定义 12.1（单位根）　满足 $z^n = 1$ 的复数 z 叫 n 次单位根（n-th root of unity）。记

$$\zeta_n = e^{2\pi i/n} = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}.$$

若 $z^n = 1$，取长度得 $|z|^n = 1$，所以 $|z| = 1$，$z = e^{i\theta}$。$z^n = e^{in\theta} = 1$ 当且仅当 $n\theta$ 是 2π 的整数倍，即 $\theta = 2\pi j/n$。所以 n 次单位根恰好是

$$1,\ \zeta_n,\ \zeta_n^2,\ \ldots,\ \zeta_n^{n-1},$$

它们两两不同；$n \geq 3$ 时，是单位圆上以 1 为一个顶点的正 n 边形的 n 个顶点。边界 $n = 1$ 时只有 1，$n = 2$ 时是 $1, -1$。$x^n - 1$ 是 n 次首一多项式，恰好有这 n 个不同的根，因此

$$\boxed{x^n - 1 = \prod_{j=0}^{n-1}(x - \zeta_n^j).}$$

指数只需看模 n 的余数：$\zeta_n^j = \zeta_n^{j'}$ 当且仅当 $j \equiv j' \pmod n$。于是 $\zeta_n^j \cdot \zeta_n^k = \zeta_n^{j+k}$，单位根相乘就是指数模 n 相加。

定义 12.2（单位根的阶，本原单位根）　设 z 是某个单位根。使 $z^d = 1$ 的最小正整数 d 叫 z 的阶（order）。阶恰好为 n 的单位根叫本原 n 次单位根（primitive n-th root of unity）。

例如 -1 的阶是 2，i 的阶是 4（$i^2 = -1$，$i^3 = -i$，$i^4 = 1$），1 的阶是 1。ζ_n 本身是本原的：$0 < k < n$ 时 ζ_n^k 的辐角 $2\pi k/n$ 不是 2π 的倍数。这与模 p 的阶完全平行，下面的命题和第 7 讲的结论也一模一样。

命题 12.3（单位根的阶）

(1) 设单位根 z 的阶为 d。则 $z^k = 1$ 当且仅当 $d \mid k$。特别地，若 $z^n = 1$，则 $d \mid n$。

(2) ζ_n^j 的阶是 $\dfrac{n}{(j,n)}$。所以 ζ_n^j 是本原 n 次单位根当且仅当 $(j,n) = 1$，本原 n 次单位根恰有 $\varphi(n)$ 个。

证明　(1) 带余除法 $k = qd + r$，$0 \leq r < d$。$z^k = (z^d)^q z^r = z^r$。若 $r > 0$，由 d 的最小性 $z^r \neq 1$；若 $r = 0$，$z^k = 1$。

(2) 记 $g = (j,n)$。$(\zeta_n^j)^k = \zeta_n^{jk} = 1$ 当且仅当 $n \mid jk$，两边除以 g，当且仅当 $\frac{n}{g} \mid \frac{j}{g}k$。$\frac{n}{g}$ 与 $\frac{j}{g}$ 互素，由 Euclid 引理，这又当且仅当 $\frac{n}{g} \mid k$。满足条件的最小正整数 k 是 $\frac{n}{g}$，这就是阶。阶等于 n 当且仅当 $g = 1$；$0 \leq j < n$ 中与 n 互素的 j 有 $\varphi(n)$ 个。 ■

(2) 就是第 7 讲的公式 $\operatorname{ord}_m(g^k) = \dfrac{h}{(h,k)}$，这里 ζ_n 扮演阶为 $h = n$ 的元素 g。

■ **例 12.1　十二次单位根按阶分类**　把 12 个十二次单位根 ζ_{12}^j（$0 \leq j \leq 11$）按阶分类，并验证

阶为 d 的恰有 $\varphi(d)$ 个。

解　由命题 12.3(2)，ζ_{12}^{j} 的阶是 $12/(j,12)$。逐个计算 $(j,12)$ 再分组：

阶 d	1	2	3	4	6	12
指数 j	0	6	4, 8	3, 9	2, 10	1, 5, 7, 11
个数	1	1	2	2	2	4
$\varphi(d)$	1	1	2	2	2	4

两行个数完全一致，合计 $1+1+2+2+2+4=12$。原因是：阶为 d 的 ζ_{12}^{j} 满足 $(\zeta_{12}^{j})^{d}=1$，它们正好是**本原 d 次单位根**。例如 $d=4$ 的两个是 $\zeta_{12}^{3}=e^{\pi i/2}=i$ 和 $\zeta_{12}^{9}=-i$，$d=3$ 的两个是 $\zeta_{12}^{4}=\zeta_3$ 和 $\zeta_{12}^{8}=\zeta_3^{2}$。本原十二次单位根是 $\zeta_{12}^{\pm 1}=\frac{\sqrt{3}}{2}\pm\frac{1}{2}i$ 和 $\zeta_{12}^{\pm 5}=-\frac{\sqrt{3}}{2}\pm\frac{1}{2}i$。 ♦

图 12.1 把这张表画在单位圆上，同一种颜色的点就是同一个阶。阶为 3 的两点加上 1 是正三角形的顶点，阶为 4 的两点加上 ± 1 是正方形的顶点：较小的正多边形嵌在正十二边形里面。

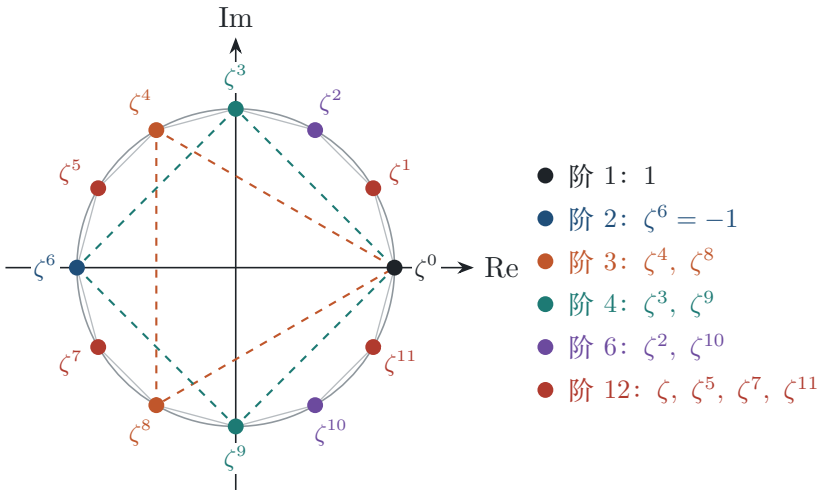


图 12.1　十二次单位根 ζ^{j}（$\zeta=\zeta_{12}$），同色的点阶相同。橙色虚线三角形的顶点是三次单位根，青色虚线正方形的顶点是四次单位根；四个红点是本原十二次单位根。

直观理解　乘以 ζ_n 就是逆时针转 $\frac{1}{n}$ 圈。ζ_n^{j} 的阶，就是「每次跳 j 格，在 n 格的钟面上跳几次第一次回到起点」。每次跳 j 格时，只能落在 (j,n) 的倍数上，这样的格子有 $n/(j,n)$ 个，所以跳 $n/(j,n)$ 次回到起点。$(j,n)=1$ 时每一格都会被踩到，这正是「本原」的含义：**ζ_n^{j} 的幂跑遍全部 n 次单位根**，就像原根的幂跑遍全部非零剩余。

想一想　i 满足 $i^8=1$，它是本原八次单位根吗？一个单位根能同时是本原 d 次和本原 d' 次单位根（$d\neq d'$）吗？

答　不是。$i=\zeta_8^{2}$，阶是 $8/(2,8)=4$，它是本原四次单位根。「$z^8=1$」只说明阶整除 8。每个单位根的阶是唯一确定的一个数，所以它恰好对一个 d 是本原的。于是全部 n 次单位根按阶分成互不相交的若干组，每组对应 n 的一个因数 d，这就是下一节分解 x^n-1 的出发点。

12.2　分圆多项式与 x^n-1 的分解

定义 12.4（分圆多项式）　以全部本原 n 次单位根为根的首一多项式

$$\Phi_n(x)=\prod_{\substack{1\le j\le n\\(j,n)=1}}\left(x-\zeta_n^{\,j}\right)$$

叫第 n 个**分圆多项式**（cyclotomic polynomial）。由命题 12.3，它的次数是 $\varphi(n)$。

「分圆」的意思是把圆周等分：x^n-1 的根把单位圆分成 n 等份，Φ_n 挑出其中「恰好需要 n 份才能回到起点」的那些分点。

■ **例 12.2**　**直接从根算 Φ_n**　由定义求 $\Phi_1,\Phi_2,\Phi_3,\Phi_4,\Phi_6$。

解　$n=1,2$：本原一次单位根只有 1，本原二次单位根只有 -1，所以 $\Phi_1(x)=x-1$，$\Phi_2(x)=x+1$。

$n=4$：本原四次单位根是 $\zeta_4=i$ 和 $\zeta_4^3=-i$，$\Phi_4(x)=(x-i)(x+i)=x^2+1$。

$n=3,6$ 要用一个小公式。单位圆上一对共轭的点 $e^{\pm i\theta}$ 之和是 $2\cos\theta$，之积是 1，所以

$$(x-e^{i\theta})(x-e^{-i\theta})=x^2-2\cos\theta\cdot x+1.$$

本原三次单位根是 $\zeta_3=e^{2\pi i/3}$ 和 $\zeta_3^2=e^{-2\pi i/3}$，$\cos\frac{2\pi}{3}=-\frac12$，得 $\Phi_3(x)=x^2+x+1$。本原六次单位根是 $e^{\pm\pi i/3}$，$\cos\frac{\pi}{3}=\frac12$，得 $\Phi_6(x)=x^2-x+1$。 ♦

素数的情形可以一步到位。若 p 是素数，$1\le j\le p-1$ 都与 p 互素，除 1 以外的 p 次单位根全是本原的。于是

$$\boxed{\Phi_p(x)=\frac{x^p-1}{x-1}=x^{p-1}+x^{p-2}+\cdots+x+1.}$$

例如 $\Phi_5(x)=x^4+x^3+x^2+x+1$。$\Phi_3=x^2+x+1$ 也符合这个公式。

一般的 Φ_n 很难这样直接算，因为 $\cos\frac{2\pi}{n}$ 一般没有简单的表达式。下面的恒等式给出一个只用多项式除法的递推办法，它是本讲第一个核心结论。

命题 12.5（x^n-1 的分解）　对每个正整数 n，

$$\boxed{x^n-1=\prod_{d\mid n}\Phi_d(x).}$$

证明　$x^n-1=\prod_z(x-z)$，z 跑遍全部 n 次单位根。由命题 12.3(1)，每个 n 次单位根的阶 d 整除 n；反过来，对 n 的每个因数 d，本原 d 次单位根 w 满足 $w^n=(w^d)^{n/d}=1$，也是 n 次单位根。所以全部 n 次单位根按阶分成互不相交的组，每个 $d\mid n$ 对应一组，组里恰好是全部本原 d

次单位根。按组把乘积重排：

$$x^n-1=\prod_{d\mid n}\ \prod_{z\ \text{的阶}=d}(x-z)=\prod_{d\mid n}\Phi_d(x).$$ ■

比较两边的次数，马上得到第 8 讲用原根数出来的恒等式的另一个证明：

$$\boxed{\sum_{d\mid n}\varphi(d)=n.}$$

例 12.1 的表就是 $n=12$ 的情形：$x^{12}-1=\Phi_1\Phi_2\Phi_3\Phi_4\Phi_6\Phi_{12}$，次数 $1+1+2+2+2+4=12$。同一个恒等式在第 8 讲里数的是模 p 中各阶元素的个数，这里数的是各阶单位根的个数，两边的计数方式完全一样。

把命题 12.5 中 $d=n$ 的一项单独拿出来，得到递推公式

$$\boxed{\Phi_n(x)=\frac{x^n-1}{\displaystyle\prod_{d\mid n,\ d<n}\Phi_d(x)}.}$$

算 Φ_n 时，较小的 Φ_d 都已经算好了，只需做一次多项式除法。

■ **例 12.3　递推计算**　用递推公式求 $\Phi_6,\Phi_8,\Phi_9,\Phi_{12}$。

解　一个省力的技巧：若 $m\mid n$，由命题 12.5，$\prod_{d\mid m}\Phi_d(x)=x^m-1$，分母里的这一组因子可以整组换成 x^m-1。

$n=6$：分母 $\Phi_1\Phi_2\Phi_3$ 中 $\Phi_1\Phi_3=x^3-1$，所以

$$\Phi_6(x)=\frac{x^6-1}{(x^3-1)(x+1)}=\frac{x^3+1}{x+1}=x^2-x+1,$$

与例 12.2 一致。

$n=8$：$\Phi_1\Phi_2\Phi_4=x^4-1$，$\Phi_8(x)=\dfrac{x^8-1}{x^4-1}=x^4+1$。

$n=9$：$\Phi_1\Phi_3=x^3-1$，$\Phi_9(x)=\dfrac{x^9-1}{x^3-1}=x^6+x^3+1$。

$n=12$：真因数是 $1,2,3,4,6$，其中 $\Phi_1\Phi_2\Phi_3\Phi_6=x^6-1$，剩下 $\Phi_4=x^2+1$：

$$\Phi_{12}(x)=\frac{x^{12}-1}{(x^6-1)(x^2+1)}=\frac{x^6+1}{x^2+1}=x^4-x^2+1.$$

用根直接验算：本原十二次单位根是 $e^{\pm\pi i/6}$ 和 $e^{\pm5\pi i/6}$，$\cos\frac{\pi}{6}=\frac{\sqrt3}{2}$，$\cos\frac{5\pi}{6}=-\frac{\sqrt3}{2}$，

$$\Phi_{12}(x)=(x^2-\sqrt3x+1)(x^2+\sqrt3x+1)=(x^2+1)^2-3x^2=x^4-x^2+1.$$

两种算法一致，而递推法完全不需要知道 $\cos\frac{\pi}{6}$。 ♦

不能直接除以 x^d-1　递推公式的分母是 Φ_d，不是 x^d-1。若误写成 $\dfrac{x^6-1}{(x-1)(x^2-1)(x^3-1)}$，分母里 Φ_1 出现了三次（$x-1$、x^2-1、x^3-1 各含一个），结果是 $\dfrac{\Phi_6}{\Phi_1^2}$，连多项式都不是。只用 x^d-1 写出 Φ_n 也是可以的，但必须有乘有除，那要用第 13、14 讲的 Möbius 函数。

表 12.1 列出了 Φ_1 到 Φ_{18}，其中合数下标的都可以用上面的办法算出来。

n	$\Phi_n(x)$	n	$\Phi_n(x)$
1	$x-1$	10	$x^4-x^3+x^2-x+1$
2	$x+1$	11	$x^{10}+x^9+\cdots+x+1$
3	x^2+x+1	12	x^4-x^2+1
4	x^2+1	13	$x^{12}+x^{11}+\cdots+x+1$
5	$x^4+x^3+x^2+x+1$	14	$x^6-x^5+x^4-x^3+x^2-x+1$
6	x^2-x+1	15	$x^8-x^7+x^5-x^4+x^3-x+1$
7	$x^6+x^5+\cdots+x+1$	16	x^8+1
8	x^4+1	17	$x^{16}+x^{15}+\cdots+x+1$
9	x^6+x^3+1	18	x^6-x^3+1

表 12.1　前 18 个分圆多项式。Φ_n 的次数是 $\varphi(n)$。

从表中可以读出几条规律，后面会逐一说明或证明：

- 系数**都是整数**。定义里出现的是 $\cos\frac{2\pi}{15}$ 这样的无理数，乘出来却全是整数（§12.3）。
- $n\ge 2$ 时常数项都是 1，系数从两头读**完全对称**。例如

$$\Phi_{15}(x)=x^8-x^7+x^5-x^4+x^3-x+1$$

（§12.3）。

- $\Phi_8=x^4+1=\Phi_2(x^4)$，$\Phi_9=\Phi_3(x^3)$，$\Phi_{16}=\Phi_2(x^8)$：素数幂下标可以化成素数下标（命题 12.10）。
- $\Phi_6(x)=\Phi_3(-x)$，$\Phi_{10}(x)=\Phi_5(-x)$，$\Phi_{14}(x)=\Phi_7(-x)$，$\Phi_{18}(x)=\Phi_9(-x)$；这来自本原根取负号后的阶变化，下面给出理由。
- 系数只有 $0,\pm1$。这一条**并不总是对的**，见下面的注。

注 12.1　第一个反例是 $n=105=3\cdot5\cdot7$：Φ_{105} 的次数是 $\varphi(105)=48$，其中 x^7 和 x^{41} 的系数都是 -2；对所有 $n<105$，Φ_n 的系数都在 $\{-1,0,1\}$ 中。系数的绝对值第一次达到 3 是在 $n=385=5\cdot7\cdot11$。配套核验脚本按递推公式逐个生成 $n\le385$ 的多项式，检查了这些有限范围结论；这不证明系数在一般情况下有统一上界。

上面的取负号规律适用于奇数 $m\ge3$：若 z 的阶是 m，则 $(-z)^2=z^2$ 的阶仍为 m，而 $(-z)^m=-1$，所以 $-z$ 的阶恰为 $2m$。取负号将两组本原根一一对应；又 $\varphi(m)$ 为偶数，故 $\Phi_{2m}(x)=\Phi_m(-x)$。

$m=1$ 要单独处理：$\Phi_1(-x)=-x-1=-\Phi_2(x)$，不能漏掉负号。

想一想　Φ_n 的根大多不是实数，为什么它的系数都是实数？由此说明 $n\ge 3$ 时 $\varphi(n)$ 是偶数。
答　复共轭 $\overline{\zeta_n^{j}}=\zeta_n^{-j}=\zeta_n^{n-j}$，而 $(n-j,n)=(j,n)$，所以本原 n 次单位根的共轭还是本原 n 次单位根。$n\ge 3$ 时它们都不是实数（实的单位根只有 ±1，阶为 $1,2$），于是两两配成共轭对 $e^{\pm i\theta}$，每对贡献一个实系数因子 $x^2-2\cos\theta\cdot x+1$。根的个数 $\varphi(n)$ 是对数的两倍，是偶数。

12.3　整系数与回文性

§12.2 的递推公式给出 Φ_n 是两个整系数多项式的商。商一般不是整系数的，例如 $x=(2x)\cdot\frac{1}{2}$。关键在于除式是首一的。

引理 12.6（首一多项式的除法）　设 f,g 是整系数多项式，g 首一，且 $f=gh$，其中 h 是复系数多项式。则 h 也是整系数多项式。

证明　用 g 对 f 做多项式长除法。每一步用余式的首项除以 g 的首项 x^m，商的这一项的系数就是余式首项的系数，是整数；减去它乘 g 后余式仍是整系数的。所以 $f=gq+r$，q,r 是整系数多项式，$\deg r<\deg g$。另一方面 $f=gh+0$。复系数多项式的带余除法是唯一的，所以 $h=q$，$r=0$。 ∎

命题 12.7（整系数）　对每个 $n\ge 1$，$\Phi_n(x)$ 是首一的整系数多项式。

证明　对 n 用强归纳法。$\Phi_1=x-1$。设 $d<n$ 时 Φ_d 都是首一整系数的，则 $g(x)=\prod_{d\mid n,\,d<n}\Phi_d(x)$ 是首一整系数的，而 $x^n-1=g(x)\Phi_n(x)$。由引理 12.6，Φ_n 是整系数的；又由定义，它是首一的。 ∎

例 12.4　用长除法求 Φ_{15}　求 Φ_{15}，并观察长除法的每一步为什么不会出现分数。

解　15 的因数是 $1,3,5,15$，$\Phi_1\Phi_5=x^5-1$，所以

$$\Phi_{15}(x)=\frac{x^{15}-1}{(x^5-1)\,\Phi_3(x)}=\frac{x^{10}+x^5+1}{x^2+x+1}.$$

用 $g=x^2+x+1$ 做长除法。每一步取「余式首项 $\div x^2$」作为商的下一项，减去它乘 g：

商的项	减去	得到的余式
		$x^{10}+x^5+1$
x^8	$x^{10}+x^9+x^8$	$-x^9-x^8+x^5+1$
$-x^7$	$-x^9-x^8-x^7$	x^7+x^5+1
x^5	$x^7+x^6+x^5$	$-x^6+1$
$-x^4$	$-x^6-x^5-x^4$	x^5+x^4+1
x^3	$x^5+x^4+x^3$	$-x^3+1$
$-x$	$-x^3-x^2-x$	x^2+x+1
1	x^2+x+1	0

所以 $\Phi_{15}(x)=x^8-x^7+x^5-x^4+x^3-x+1$。因为除式的首项系数是 1，商的每一项都保留当前余式的首项系数，只降低相应的次数，不需要把系数除以别的数。这正是引理 12.6 的证明。如果除式是 $2x^2+x+1$，第一步商就是 $\frac{1}{2}x^8$ 了。 ♦

再看回文性。对 d 次多项式 f，令 $f^*(x)=x^d f(1/x)$，它就是把 f 的系数倒过来写：$f=a_dx^d+\cdots+a_1x+a_0$ 时 $f^*=a_0x^d+a_1x^{d-1}+\cdots+a_d$。若 $f^*=f$，称 f 是回文多项式（palindromic polynomial）。两条性质直接由定义得到：

$$(fg)^*=f^*g^*,\qquad (x^m-1)^*=x^m\bigl(x^{-m}-1\bigr)=-(x^m-1).$$

命题 12.8（回文性）　对 $n\ge 2$，$x^{\varphi(n)}\,\Phi_n(1/x)=\Phi_n(x)$，即 Φ_n 是回文多项式。

证明　对 $n\ge 2$ 用强归纳法，设 $2\le d<n$ 时 $\Phi_d^*=\Phi_d$。把 x^n-1 写成

$$x^n-1=\Phi_1(x)\cdot P(x)\cdot\Phi_n(x),\qquad P(x)=\prod_{d\mid n,\ 1<d<n}\Phi_d(x).$$

两边取 *：左边是 $-(x^n-1)$；右边是 $\Phi_1^*P^*\Phi_n^*=-\Phi_1\,P\,\Phi_n^*$，这里用了 $\Phi_1^*=(x-1)^*=-\Phi_1$ 和归纳假设 $P^*=P$。于是 $\Phi_1P\Phi_n=\Phi_1P\Phi_n^*$，约去非零多项式 Φ_1P 得 $\Phi_n^*=\Phi_n$。$n=2$ 时 $P=1$，论证同样成立。 ■

直观理解　f^* 的根是 f 的根的倒数。本原 n 次单位根的倒数 ζ_n^{-j} 还是本原 n 次单位根，所以 Φ_n^* 和 Φ_n 的根完全相同。两者的首项系数分别是 $\Phi_n(0)$ 和 1；下面的推论说明它们也相等，于是两个多项式相同。

推论 12.9　对 $n\ge 2$，$\Phi_n(0)=1$；而 $\Phi_1(0)=-1$。

证明　$\Phi_n(0)$ 是 Φ_n 的常数项，也就是 Φ_n^* 的首项系数。由命题 12.8，它等于 Φ_n 的首项系数 1。 ■

命题 12.10（素数幂）　对素数 p 和正整数 k，

$$\Phi_{p^k}(x)=\Phi_p(x^{p^{k-1}})=\frac{x^{p^k}-1}{x^{p^{k-1}}-1}=x^{(p-1)p^{k-1}}+x^{(p-2)p^{k-1}}+\cdots+x^{p^{k-1}}+1.$$

特别地，对任意素数 p 与 $k\geq 1$，$\Phi_{p^k}(1)=p$（练习 12.7）。

证明　p^k 的小于它的因数恰好是 p^{k-1} 的因数，所以

$$x^{p^k}-1=\Phi_{p^k}(x)\prod_{d\mid p^{k-1}}\Phi_d(x)=\Phi_{p^k}(x)\cdot(x^{p^{k-1}}-1).$$

另一方面，由命题 12.5，$x^{p^k}-1$ 的根是 $\zeta_{p^k}^{j}$，$x^{p^{k-1}}-1$ 的根是 $\zeta_{p^k}^{pj}$；而 $(pj,p^k)=p(j,p^{k-1})$，所以后者的阶都整除 p^{k-1}，不等于 p^k，它们恰好组成 Φ_{p^k} 根的补集。将公式中的 p 项几何级数在 $x=1$ 处求值，得 $\Phi_{p^k}(1)=1+1+\cdots+1=p$，这里包括 $p=2$。 ■

■ **例 12.5　素数幂的情形**　用命题 12.10 计算 $\Phi_8,\Phi_9,\Phi_{16},\Phi_{25},\Phi_{27}$，并把二次幂下标和 Fermat 数联系起来。

解　代入素数幂公式：$\Phi_8=\Phi_2(x^4)=x^4+1$，$\Phi_{16}=\Phi_2(x^8)=x^8+1$；$\Phi_9=\Phi_3(x^3)=x^6+x^3+1$，$\Phi_{27}=\Phi_3(x^9)=x^{18}+x^9+1$；$\Phi_{25}=\Phi_5(x^5)=x^{20}+x^{15}+x^{10}+x^5+1$。

特别地，$\Phi_{2^{k+1}}(x)=x^{2^k}+1$，所以 $\Phi_{2^{k+1}}(2)=2^{2^k}+1=F_k$，即第 k 个 Fermat 数。$F_5=\Phi_{64}(2)=2^{32}+1$。 ♦

前后联系　Φ_n 在有理数域 $\mathbb{Q}$ 上不可约，即不能写成两个次数 ≥ 1 的有理系数多项式的乘积。原课程陈述了这个结论，本讲不证明它。不可约与整系数并不保证它模每个素数都分解成一次因子。下一节会证明：在 $p\nmid n$ 的前提下，Φ_n 模 p 完全分裂当且仅当 $p\equiv 1\pmod n$。

Gauss 在 1796 年发现正十七边形可以用尺规作图，因为 Φ_{17} 的 16 个复数根可以通过连续开平方算出来（$\varphi(17)=16=2^4$）。Wantzel 1837 年证明了 Gauss–Wantzel 定理：正 n 边形可作当且仅当 n 是 2 的幂与若干个不同 Fermat 素数的乘积。

12.4　模 p 的分圆多项式

从 §12.2 开始已经建立了单位根和分圆多项式在复数域里的理论，现在把 Φ_n 拿到模 p 的世界。核心问题是：$\Phi_n(x)\equiv 0\pmod p$ 有解吗？有的话有几个解？答案和 n 的阶联系在一起，本节给出本讲的关键定理：**若 $p\nmid n$ 且 $p\mid\Phi_n(a)$，则 $\operatorname{ord}_p(a)=n$，从而 $p\equiv 1\pmod n$。**

先解释为什么 $p\nmid n$ 是必要的。回到复数：x^n-1 的正式导数是 nx^{n-1}。多项式 f 有至少二重的根 α 当且仅当 $f(\alpha)=f'(\alpha)=0$（这称为重根判别法）。模 p 时完全一样。

取 $n=4$，有 $\Phi_4(x)=x^2+1$。模 13 时，$5^2\equiv 8^2\equiv -1$，所以 $5,8$ 都是 Φ_4 的根；它们的阶恰好是 4。这不是偶然：在模数不整除 n 时，分圆多项式的根不能同时落在较低阶的单位根里。

先看阻止「两个因子撞在同一个根上」的引理。

引理 12.11（互素模数下没有重根）　设 p 是素数且 $p\nmid n$。若 m 是 n 的真因数，则 $\Phi_n(x)$ 与 x^m-1 在模 p 下没有公共根。

证明　反设 a 是公共根。由 $a^m\equiv 1 \pmod p$，有 $p\nmid a$。分解公式给出

$$x^n-1=\Phi_n(x)\prod_{\substack{d\mid n\\ d<n}}\Phi_d(x),\qquad x^m-1=\prod_{d\mid m}\Phi_d(x).$$

因为 $m<n$ 且 $m\mid n$，第二个乘积的每个因子都出现在第一个乘积里。于是模 p 的 x^n-1 在 a 处有重根：两个相乘的因子都在 a 处为零。

若多项式 F 在 a 处有重根，则 $(x-a)^2\mid F$，从而 $F'(a)=0$。但这里

$$(x^n-1)'(a)=na^{n-1}\not\equiv 0 \pmod p,$$

因为 $p\nmid n$ 且 $p\nmid a$。矛盾。 ■

定理 12.12（分圆根的阶判别）　设 $n\ge 2$，p 是素数，$p\nmid n$，且 a 是模 p 的 Φ_n 的根。则

$$\operatorname{ord}_p(a)=n,\qquad \text{因而} n\mid p-1,\ \text{即} p\equiv 1 \pmod n.$$

思路　分解 a^n-1，先得到 $\operatorname{ord}_p(a)\mid n$。如果阶是某个真因数 $m<n$，那么 a 同时是 x^m-1 和 Φ_n 的根，正好违反引理 12.11。最后用费马小定理把阶与 $p-1$ 联系起来。

证明　由 $x^n-1=\prod_{d\mid n}\Phi_d(x)$，代入 a 得 $a^n\equiv 1 \pmod p$，所以 $d=\operatorname{ord}_p(a)$ 存在且 $d\mid n$。若 $d<n$，则 $a^d\equiv 1$，说明 a 是 x^d-1 的根，又是 Φ_n 的根，与引理矛盾。因此 $d=n$。

另一方面，费马小定理给出 $a^{p-1}\equiv 1 \pmod p$，按阶的定义 $n=\operatorname{ord}_p(a)\mid p-1$。所以 $p\equiv 1 \pmod n$。 ■

推论 12.13（模 p 下完全分裂）　设 $n\ge 2$ 且 p 是不整除 n 的素数。则 $\Phi_n(x)$ 在模 p 下完全分解成一次因子，当且仅当 $p\equiv 1 \pmod n$。

证明　若完全分裂，Φ_n 有 $\varphi(n)$ 个模 p 的根；它们互不相同，因为 x^n-1 在 $p\nmid n$ 时没有重根。定理 12.12 说明每个根的阶为 n，故 $n\mid p-1$。

反过来，若 $n\mid p-1$，循环群 $\mathbb{F}_p^\times$ 中有恰好 $\varphi(n)$ 个阶为 n 的元素。每个这样的元素是 x^n-1 的根；由分解公式，它是某个 Φ_d 的根，其中 $d\mid n$。若 $d<n$，则 $\Phi_d\mid x^d-1$ 会推出它的阶整除 d，矛盾。因此这 $\varphi(n)$ 个元素全是 Φ_n 的根。由于 $\deg\Phi_n=\varphi(n)$，它完全分裂。 ■

条件 $p\nmid n$ 不能删掉。例如 $n=p=3$ 时，$\Phi_3(x)=x^2+x+1\equiv(x-1)^2 \pmod 3$，$a=1$ 是根，但 $\operatorname{ord}_3(1)=1\neq 3$。这里 x^3-1 在模 3 下有重根，正是导数 $3x^2$ 恒为零造成的。

■ **例 12.6　模 13 的 Φ_4 根**　求模 13 的 $\Phi_4(x)=x^2+1$ 的所有根，并验证它们的阶。

解　$x^2 \equiv -1 \equiv 12 \pmod{13}$ 的两个根是 $x \equiv 5, 8$，因为 $5^2 = 25 \equiv -1$，而 $8 \equiv -5$。它们都不是 1，平方又是 -1，所以四次幂为 1，阶恰为 4。定理也预言了 $4 \mid 13 - 1$。 ♦

想一想　如果 $p \mid \Phi_n(a)$，但 $p \mid n$，还能断言 $n \mid p - 1$ 吗?
答　不能。上面的 $p = n = 3, a = 1$ 就是反例：$3 \mid \Phi_3(1) = 3$，但 $3 \nmid 2$。使用阶判别定理前，必须先检查 $p \nmid n$。

12.5　分圆多项式值的素因数

定理 12.12 的输入是「模 p 的根」，而整数计算里更常见的输入是一个整除关系 $p \mid \Phi_n(a)$。二者只差一步：把整数等式看成模 p 的等式。这里 $n \geq 2$，所以 $\Phi_n(0) = 1$；若 $p \mid \Phi_n(a)$，就不可能有 $p \mid a$。

推论 12.14（分圆值的素因数）　设 $n \geq 2$，q 是素数，$q \mid \Phi_n(a)$，且 $q \nmid n$。则

$$\operatorname{ord}_q(a) = n, \qquad q \equiv 1 \pmod{n}.$$

证明　由 $\Phi_n(0) = 1$，$q \mid a$ 会推出 $\Phi_n(a) \equiv 1 \pmod{q}$，与假设矛盾。因此 a 是模 q 的非零根。把 $q \mid \Phi_n(a)$ 写成 $\Phi_n(a) \equiv 0 \pmod{q}$，应用定理 12.12 即得结论。 ■

图 12.2 把这条推理链画出来。最上方同时列出整除关系和额外假设 $q \nmid n$；不能由整除关系推出这个假设。右边的「低阶都不成立」则是引理 12.11 对每个真因数的结论。

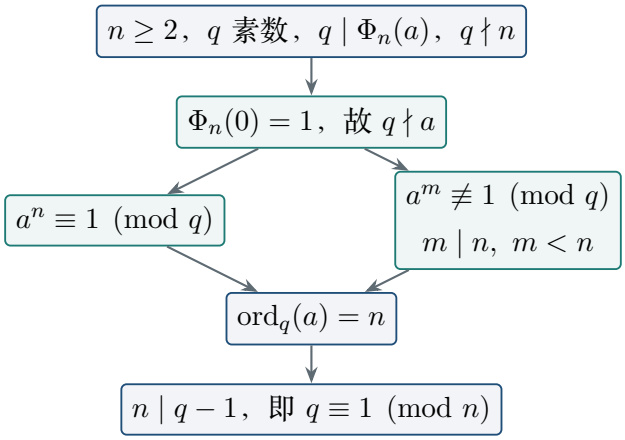


图 12.2　在 $n \geq 2$、q 为素数且 $q \nmid n$ 的条件下，$q \mid \Phi_n(a)$ 给出一个阶恰为 n 的模 q 单位根，因而 $q \equiv 1 \pmod{n}$。

■ **例 12.7**　**分解 $2^{11} - 1$**　说明 $2^{11} - 1$ 的素因数为什么都落在 $1 \pmod{11}$ 的剩余类中，并完成分解。

解　因为 11 是素数，

$$2^{11} - 1 = \Phi_{11}(2).$$

直接试除得到

$$2047 = 23 \cdot 89, \qquad 23 = 2 \cdot 11 + 1, \quad 89 = 8 \cdot 11 + 1.$$

试除可知 23 和 89 都是素数，且都不等于 11，所以满足推论 12.14 的条件。也可以核对阶：$2^{11} \equiv 1 \pmod{23}$ 且 $2 \not\equiv 1 \pmod{23}$，因 11 是素数，故 $\operatorname{ord}_{23}(2) = 11$；对 89 同理。于是 $11 \mid 23 - 1$ 和 $11 \mid 89 - 1$。

这说明分圆多项式给出的不是「任意因数都可能出现」的分解：它先把候选素数限制在一个同余类里，再做有限的试除。 ♦

■ **例 12.8　Fermat 数 F_5 的两个因数**　验证

$$2^{32} + 1 = 641 \cdot 6700417$$

并解释为什么两个素因数都满足 $p \equiv 1 \pmod{64}$。

解　由命题 12.10，$\Phi_{64}(x) = x^{32} + 1$，所以左边就是 $\Phi_{64}(2)$。验证 641 可用恒等式 $641 = 5 \cdot 2^7 + 1 = 5^4 + 2^4$。模 641 下，$(5 \cdot 2^7)^4 \equiv 1$ 且 $5^4 \equiv -2^4$，消去可逆的 2^4 得 $2^{32} \equiv -1$。数值上

$$641 \cdot 6700417 = 4294967297 = 2^{32} + 1.$$

第一个因数的出现可以用 $641 = 5 \cdot 2^7 + 1$ 追踪；直接取模也可验证 $2^{32} \equiv -1 \pmod{641}$。两个因数都不整除 64，而

$$641 - 1 = 10 \cdot 64, \qquad 6700417 - 1 = 104694 \cdot 64.$$

两因数经试除到各自平方根可核对为素数，因此它们都符合推论 12.14。这里先认出 $2^{32} + 1$ 是一个分圆值，便能限制其素因数的同余类。 ♦

使用阶判别前先检查素数幂条件　从 $p \mid \Phi_n(a)$ 直接写出 $p \equiv 1 \pmod{n}$，少了一步检查。必须确认 $p \nmid n$；例如 $\Phi_3(1) = 3$，但 $3 \not\equiv 1 \pmod{3}$。若 $p \mid n$，$x^n - 1$ 可能出现重根，低阶根会和 Φ_n 的根合并。

12.6　模 n 余 1 的素数有无穷多个

现在把上面的推论反过来使用。若我们能构造一个数 A，使 $\Phi_n(A) > 1$，那么它的某个素因数 q 会满足 $q \equiv 1 \pmod{n}$，只要同时保证 $q \nmid n$。把已经列出的所有候选素数乘进 A，就能保证新得到的 q 不在旧列表中。

定理 12.15（模 n 余 1 的素数）　对每个正整数 n，都有无穷多个素数 q 满足 $q \equiv 1 \pmod{n}$。

思路　若只有有限个这样的素数 $q_1, \dots, q_N$，把它们和 n 的所有素因子一起乘进 A。分圆值 $\Phi_n(A)$ 的每个素因数都不会落在这个乘积里：代入模 q_i 得到常数项 1。但阶判别又说它必须是 $1 \pmod{n}$ 的素数，矛盾。

证明　$n = 1$ 时结论就是素数有无穷多个。以下设 $n \geq 2$，反设满足 $q \equiv 1 \pmod{n}$ 的素数只有

$$q_1, q_2, \dots, q_N.$$

允许 $N=0$，此时约定空乘积为 1。取正整数 ℓ 足够大，令

$$A=\ell n q_1 q_2\cdots q_N.$$

因为 Φ_n 是首一多项式，A 足够大时 $M=\Phi_n(A)>1$。取 p 为 M 的一个素因数。

先排除旧素数。对每个 i，有 $q_i\mid A$，而 $\Phi_n(0)=1$，所以

$$\Phi_n(A)\equiv\Phi_n(0)=1 \pmod{q_i};$$

因此 $p\neq q_i$。同样的计算说明 $p\nmid A$：若 $p\mid A$，就会得到 $M\equiv 1 \pmod p$，与 $p\mid M$ 矛盾。特别地，因为 $n\mid A$，所以 $p\nmid n$。

现在 $p\mid\Phi_n(A)$ 且 $p\nmid n$。推论 12.14 给出 $p\equiv 1 \pmod n$，所以 p 应该出现在列表中；但上一段已经证明 p 不等于列表中的任何 q_i，矛盾。因此这样的素数不可能只有有限个。 ■

证明中没有假设列表非空。若 $N=0$，直接取 $A=\ell n$；例如 $n=3$、$\ell=1$ 时

$$\Phi_3(3)=3^2+3+1=13,\qquad 13\equiv 1 \pmod 3.$$

这已经从「没有已知候选」的空列表里造出一个候选。若 $n=4$ 取 $A=4$，则

$$\Phi_4(4)=4^2+1=17,\qquad 17\equiv 1 \pmod 4.$$

改变 ℓ 就会得到无穷多次同样的构造；定理保证每次得到的素因数不能永远停留在同一个有限列表里。

想一想　为什么构造 $A=\ell q_1\cdots q_N$ 时还要把 n 乘进去？

答　需要排除 $p\mid n$ 的重根情形。若只把旧的 q_i 乘进 A，新得到的素因数有可能整除 n，此时阶判别定理不能使用。加入因子 n 后，任何 $p\mid n$ 都同时满足 $p\mid A$；但 $p\mid\Phi_n(A)$ 又会给出 $\Phi_n(A)\equiv\Phi_n(0)=1 \pmod p$，矛盾。因此新素因数自动与 n 互素。

前后联系　这是一种 Euclid 式构造：先假设候选已经列完，再用一个精心选择的整数制造列表之外的新素数。与「$4k+3$ 型素数无穷多」的证明相比，变化在于这里用 $\Phi_n(A)$ 代替了一个固定的二次式；分圆多项式把「低阶因子已经被排除」编码进了它的根结构。

习题集 6 的 Bonus 10(c) 还要求计算全部本原 n 次单位根之和，答案是 Möbius 函数 $\mu(n)$。把 $x^n-1=\prod_{d\mid n}\Phi_d(x)$ 与各因子根的和结合，再用第 14 讲的 Möbius 反演，可以得到这个结论。

本讲小结

- ζ_n^j 的阶是 $n/(j,n)$；本原 n 次单位根有 $\varphi(n)$ 个。
- 全部单位根按阶分类，给出 $x^n-1=\prod_{d\mid n}\Phi_d(x)$ 和 $\sum_{d\mid n}\varphi(d)=n$。
- Φ_n 是首一整系数多项式；$n\geq 2$ 时 $x^{\varphi(n)}\Phi_n(1/x)=\Phi_n(x)$ 且 $\Phi_n(0)=1$。
- $\Phi_{p^k}(x)=\dfrac{x^{p^k}-1}{x^{p^{k-1}}-1}$，因此 $\Phi_{p^k}(1)=p$，包括 $p=2$。
- 若 $n\geq 2$、$p\nmid n$ 且 $p\mid\Phi_n(a)$，则 $\operatorname{ord}_p(a)=n$，所以 $p\equiv 1\pmod n$。
- 对每个正整数 n，都有无穷多个素数 $p\equiv 1\pmod n$；$n=1$ 是 Euclid 定理，其余情形用 $\Phi_n(\ell n q_1\cdots q_N)$ 构造新素因数。

练习

12.1 ★ 求 ζ_{20}^6 的阶，并求本原 20 次单位根的个数。

12.2 ★ 用 $x^n-1=\prod_{d\mid n}\Phi_d(x)$ 计算 $\Phi_{10}(x)$ 和 $\Phi_{14}(x)$。

12.3 ★ 计算 $\Phi_{15}(2)$，并说明它的素因数满足什么同余条件。

12.4 ★★ 分解 $\Phi_{64}(2)$，核对两个因数都是素数，并验证它们模 64 的余数。

12.5 ★★ 求模 11 的 $\Phi_5(x)$ 的全部根，逐个核对它们的阶。

12.6 ★★ 判断「若 $p\mid\Phi_n(a)$，则 $p\equiv 1\pmod n$」是否总成立；若不成立，给出 $p\mid n$ 的反例。

12.7 ★★ 由素数幂公式证明 $\Phi_{p^k}(1)=p$，并单独核对 $p=2,k=1$。

12.8 ★★ 在定理 12.15 的构造中，分别取 $n=3$、$n=4$，用空列表造出一个符合结论的素数。

12.9 ★★★ 证明对每个 $n\geq 2$，都有无穷多个素数 $p\equiv 1\pmod n$。证明必须覆盖候选列表为空的情形。

练习答案与提示

12.1 $\operatorname{ord}(\zeta_{20}^6)=20/(6,20)=10$；$\varphi(20)=20(1-1/2)(1-1/5)=8$，所以本原二十次单位根有 8 个。

12.2 $\Phi_{10}(x)=\Phi_5(-x)=x^4-x^3+x^2-x+1$，$\Phi_{14}(x)=\Phi_7(-x)=x^6-x^5+x^4-x^3+x^2-x+1$；代回分解恒等式即可核对。

12.3 $\Phi_{15}(x)=x^8-x^7+x^5-x^4+x^3-x+1$，故 $\Phi_{15}(2)=151$。试除不超过 $\sqrt{151}<13$ 的素数可知 151 为素数，且 $151=10\cdot 15+1$。

12.4 $\Phi_{64}(2)=2^{32}+1=4294967297=641\cdot 6700417$。试除到各自平方根可核对两因数为素数；$641=10\cdot 64+1$，$6700417=104694\cdot 64+1$。

12.5 $\Phi_5(x)=x^4+x^3+x^2+x+1$ 在模 11 下的根是 $3,4,5,9$。它们的五次幂都为 1，且都不等于 1；因 5 是素数，四个根的阶全为 5。

12.6 不总成立。例如 $p=n=5,a=1$ 时，$\Phi_5(1)=5$，所以 $5\mid\Phi_5(1)$，但 $5\not\equiv 1\pmod 5$。阶判别需要 $p\nmid n$。

12.7 $\Phi_{p^k}(x)=1+x^{p^{k-1}}+\cdots+x^{(p-1)p^{k-1}}$，令 $x=1$ 得 p 项之和为 p。当 $p=2,k=1$ 时 $\Phi_2(x)=x+1$，故 $\Phi_2(1)=2$。

12.8 $n=3$ 取 $A=3$，$\Phi_3(A)=13\equiv 1\pmod 3$；$n=4$ 取 $A=4$，$\Phi_4(A)=17\equiv 1\pmod 4$。

12.9 反设候选只有 $q_1,\dots,q_N$，允许 $N=0$ 且空乘积为 1。取 ℓ 足够大，使 $A=\ell n\prod_i q_i$ 满足 $M=$

$\Phi_n(A) > 1$；取素数 $p \mid M$。若 $p \mid A$，则 $M \equiv \Phi_n(0) = 1 \pmod{p}$，矛盾。因此 $p \nmid n$ 且 $p \neq q_i$；推论 12.14 却给出 $p \equiv 1 \pmod{n}$，与列表穷尽所有候选矛盾。$N = 0$ 时 $A = \ell n$，证明不变。

第 13 讲

算术函数与 Möbius 函数

Arithmetic Functions, the Möbius Function

对应材料：MIT 18.781 Lecture 13 讲义（定理 48–53，命题 51–52）；习题集 6 第 1、2、3、4、5、6、8(b)、10(c) 题。

本讲目标

(1) 认识最常用的算术函数 $d(n)$、$\sigma(n)$、$\sigma_k(n)$、$\omega(n)$、$\Omega(n)$、$\varphi(n)$，会对具体的 n 计算它们。

(2) 区分积性与完全积性；理解积性函数由它在素数幂上的值完全决定。

(3) 掌握 Dirichlet 卷积 $(f*g)(n)=\sum_{d\mid n} f(d)g(n/d)$ 的定义，会用对称写法 $\sum_{ab=n}$ 证明它满足交换律、结合律，知道单位元是 $\mathbf{1}$。

(4) 证明「两个积性函数的卷积仍是积性函数」，并由此推出 $d(n)$、$\sigma_k(n)$ 的公式和恒等式 $\sum_{d\mid n}\varphi(d)=n$。

(5) 证明 Euclid–Euler 定理：偶完全数恰好是 $2^{m-1}(2^m-1)$，其中 2^m-1 是素数。

(6) 定义 Möbius 函数 μ，证明 $\mu * U=\mathbf{1}$，即 $\sum_{d\mid n}\mu(d)$ 在 $n=1$ 时为 1、在 $n>1$ 时为 0。

13.1 算术函数

拿一个具体的数，比如 12，可以问它很多问题：它有几个正因数？这些因数加起来是多少？它有几个不同的素因数？$1,\dots,12$ 里有几个与它互素？不超过它的素数有几个？答案分别是 6、28、2、4、5。把 12 换成一般的 n，每个问题就给出一个「输入正整数、输出一个数」的函数。本讲研究这类函数的共同结构。

定义 13.1（算术函数） 定义在正整数集 $\mathbb{Z}_{>0}=\{1,2,3,\dots\}$ 上、取复数值的函数 $f:\mathbb{Z}_{>0}\to\mathbb{C}$，称为**算术函数**（arithmetic function）。

本课程约定 $0\in\mathbb{N}$，而算术函数在 0 处没有定义，所以这里写 $\mathbb{Z}_{>0}$。定义里没有任何限制，任意一个数列 $f(1),f(2),\dots$ 都是算术函数。有意思的是那些**与整除结构相关**的函数。下面是本讲要用到的几个。

- $\pi(n)$：不超过 n 的素数个数。例如 $\pi(12)=5$（$2,3,5,7,11$）。
- $d(n)$：n 的正因数个数。
- $\sigma(n)$：n 的正因数之和。

- $\sigma_k(n)=\sum_{d\mid n} d^k$：$n$ 的正因数的 k 次方之和。于是 $\sigma_1=\sigma$，$\sigma_0=d$。
- $\omega(n)$：n 的不同素因数的个数。
- $\Omega(n)$：n 的素因数个数，按重数计算。若 $n=p_1^{e_1}\cdots p_r^{e_r}$，则 $\omega(n)=r$，$\Omega(n)=e_1+\cdots+e_r$。
- $\varphi(n)$：$1,\dots,n$ 中与 n 互素的数的个数（第 4 讲）。

记号 $\sum_{d\mid n}$ 表示对 n 的全部正因数 d 求和。有些书把 $d(n)$ 记作 $\tau(n)$；本讲把 τ 留给 §13.4 末尾的 Ramanujan τ 函数。

表 13.1 列出 $n\le 12$ 时的值。最后一行 $\mu(n)$ 是 §13.6 才定义的 Möbius 函数，先放在这里，读到那里时再回来看。

n	1	2	3	4	5	6	7	8	9	10	11	12
$d(n)$	1	2	2	3	2	4	2	4	3	4	2	6
$\sigma(n)$	1	3	4	7	6	12	8	15	13	18	12	28
$\omega(n)$	0	1	1	1	1	2	1	1	1	2	1	2
$\Omega(n)$	0	1	1	2	1	2	1	3	2	2	1	3
$\varphi(n)$	1	1	2	2	4	2	6	4	6	4	10	4
$\pi(n)$	0	1	2	2	3	3	4	4	4	4	5	5
$\mu(n)$	1	−1	−1	0	−1	1	−1	0	0	1	−1	0

表 13.1　常用算术函数在 $n\le 12$ 时的值。注意 $n=1$ 那一列：1 没有素因数，所以 $\omega(1)=\Omega(1)=0$。

例 13.1　对 $n=12$ 计算 $d(n)$、$\sigma(n)$、$\sigma_2(n)$、$\omega(n)$、$\Omega(n)$、$\varphi(n)$。

解　$12=2^2\cdot 3$，正因数是 $1,2,3,4,6,12$。于是

$$d(12)=6,\qquad \sigma(12)=1+2+3+4+6+12=28,$$

$$\sigma_2(12)=1+4+9+16+36+144=210.$$

不同的素因数是 $2,3$，所以 $\omega(12)=2$；按重数数，$12=2\cdot 2\cdot 3$，所以 $\Omega(12)=3$。与 12 互素的是 $1,5,7,11$，所以 $\varphi(12)=4$。 ♦

这种「逐个列因数」的办法对大数不可行。§13.4 会给出只看素因数分解就能算的公式，例如 $\sigma(12)=(1+2+4)(1+3)=7\cdot 4=28$。

13.1.1　完全数

古希腊人特别关心一个数与它的真因数（除自身以外的正因数）之和的关系。6 的真因数是 $1,2,3$，加起来恰好是 6；28 的真因数 $1,2,4,7,14$ 加起来也恰好是 28。真因数之和等于 $\sigma(n)-n$，所以「真因数之和等于 n」就是 $\sigma(n)=2n$。

定义 13.2（完全数） 若 $\sigma(n)=2n$，则称正整数 n 为**完全数**（perfect number）。若 $\sigma(n)<2n$，称 n 为**亏数**（deficient）；若 $\sigma(n)>2n$，称 n 为**盈数**（abundant）。

由表 13.1，12 是盈数（$\sigma(12)=28>24$），8 是亏数（$\sigma(8)=15<16$），每个素数 p 都是亏数（$\sigma(p)=p+1<2p$）。

■ **例 13.2**　验证 496 是完全数，并观察 6、28、496 的素因数分解有什么共同点。

解　$496=16\cdot 31=2^4\cdot 31$，而 31 是素数。正因数分成两行：不含 31 的和含 31 的，

$$\begin{aligned} &1,\ 2,\ 4,\ 8,\ 16 && \text{和为}31,\\ &31,\ 62,\ 124,\ 248,\ 496 && \text{和为}31\cdot(1+2+4+8+16)=31\cdot 31=961. \end{aligned}$$

所以 $\sigma(496)=31+961=992=2\cdot 496$，496 是完全数。

三个数的分解是

$$6=2\cdot 3,\qquad 28=2^2\cdot 7,\qquad 496=2^4\cdot 31.$$

每个都形如 $2^{m-1}(2^m-1)$：$m=2,3,5$ 时 $2^m-1=3,7,31$，都是素数。上面「分两行」的算法其实就是 $\sigma(2^4\cdot 31)=\sigma(2^4)\,\sigma(31)$。§13.5 将证明：偶完全数全都是这个形状。 ♦

历史与背景　Euclid《几何原本》第 IX 卷命题 36 证明了：若 2^m-1 是素数，则 $2^{m-1}(2^m-1)$ 是完全数。两千年后 Euler 证明了反过来也对：每个偶完全数都是这种形状。形如 2^m-1 的素数叫 **Mersenne 素数**（Mersenne prime），到 2024 年已知 52 个，所以已知的完全数也恰好是 52 个，全是偶数。是否存在奇完全数，至今没有答案；如果存在，它必须大于 10^{1500}（Ochem 与 Rao，2012）。「每个完全数都是偶数」是数论中最古老的未解猜想之一。

想一想　$\pi(n)$ 与乘法有没有关系？比较 $\pi(6)$ 与 $\pi(2)\pi(3)$，$\sigma(6)$ 与 $\sigma(2)\sigma(3)$。

答　$\pi(6)=3$，$\pi(2)\pi(3)=1\cdot 2=2$，不相等。$\pi(n)$ 数的是区间 $[1,n]$ 里的素数，它关心 n 的大小，与 n 的因数分解无关。$\sigma(6)=12=3\cdot 4=\sigma(2)\sigma(3)$，相等。下一节说明这不是巧合。

13.2　积性函数

从表 13.1 里多找几个例子：

$$d(6)=4=d(2)d(3),\qquad \sigma(10)=18=3\cdot 6=\sigma(2)\sigma(5),\qquad \varphi(12)=4=2\cdot 2=\varphi(4)\varphi(3).$$

但是 $d(4)=3$，而 $d(2)d(2)=4$；$\sigma(4)=7$，而 $\sigma(2)\sigma(2)=9$。规律似乎是：两个数互素时，函数值可以相乘。

定义 13.3（积性与完全积性）　设 f 是不恒为零的算术函数。

(1) 若对所有满足 $\gcd(m,n)=1$ 的正整数 m,n 都有 $f(mn)=f(m)f(n)$，称 f 为**积性函数**（multiplicative function）。

(2) 若对**所有**正整数 m,n（不要求互素）都有 $f(mn)=f(m)f(n)$，称 f 为**完全积性函数**（completely multiplicative function）。

完全积性一定是积性。积性函数总有 $f(1)=1$：取 $m=1$，则 $\gcd(1,n)=1$，所以 $f(n)=f(1)f(n)$ 对一切 n 成立；f 不恒为零，取 $f(n)\neq 0$ 的那个 n 约去，得 $f(1)=1$。因此「$f(1)=1$」是检验积性的第一道关卡。

13.2.1　例子与反例

下面几个函数会在本讲反复出现，记号沿用原讲义。

- 函数 $\mathbf{1}$：$\mathbf{1}(1)=1$，$n>1$ 时 $\mathbf{1}(n)=0$。它是完全积性的：若 m,n 都是 1，两边都是 1；否则 $mn>1$，左边是 0，右边至少有一个因子是 0。§13.3 会说明它为什么叫 $\mathbf{1}$。
- 常值函数 U：对一切 n，$U(n)=1$，故 $U(mn)=1=U(m)U(n)$，所以它完全积性。
- 幂函数 $r_k(n)=n^k$（k 为固定的非负整数）：$(mn)^k=m^kn^k$，完全积性。$r_0=U$，$r_1(n)=n$。
- φ：积性（第 5 讲用中国剩余定理证明），但不是完全积性，因为 $\varphi(4)=2\neq 1=\varphi(2)^2$。
- 固定整数 D，令 $f(n)=\left(\frac{D}{n}\right)$（$n$ 为奇数时取 Jacobi 符号，n 为偶数时令 $f(n)=0$）。由第 10 讲，Jacobi 符号对分母满足 $\left(\frac{D}{mn}\right)=\left(\frac{D}{m}\right)\left(\frac{D}{n}\right)$，所以 f 是完全积性的。

再看两个与素因数个数有关的函数。ω **不是**积性的：$\omega(1)=0\neq 1$，第一道关卡就通不过。但互素的 m,n 没有公共素因数，mn 的不同素因数恰好是 m 的和 n 的合在一起，所以

$$\omega(mn)=\omega(m)+\omega(n)\qquad(\gcd(m,n)=1).$$

这样的函数叫**加性函数**（additive function）。把加法变成乘法，取指数即可：$2^{\omega(mn)}=2^{\omega(m)}\cdot 2^{\omega(n)}$，所以 $2^{\omega(n)}$ 是积性函数。它不是完全积性的：$2^{\omega(4)}=2$，而 $2^{\omega(2)}\cdot 2^{\omega(2)}=4$。

Ω 更好：对一切 m,n 都有 $\Omega(mn)=\Omega(m)+\Omega(n)$（素因数按重数数，乘起来重数相加），称为**完全加性**。于是 **Liouville 函数**（Liouville function）

$$\lambda(n)=(-1)^{\Omega(n)}$$

是完全积性的。例如 $\lambda(12)=(-1)^3=-1$，$\lambda(2)\lambda(6)=(-1)\cdot(+1)=-1$。

■ **例 13.3**　固定正整数 k，令 $f(n)=\gcd(n,k)$。证明 f 是积性函数；取 $k=12$，说明它不是完全积性的。

解　设 $\gcd(m,n)=1$。对每个素数 p，比较两边 p 的指数。记 $v_p(x)$ 为 x 的分解中 p 的指数。最大公因数的 p-指数是两数 p-指数的较小者，所以

$$v_p\big(\gcd(mn,k)\big)=\min\big(v_p(m)+v_p(n),\ v_p(k)\big).$$

因为 m,n 互素，$v_p(m)$ 与 $v_p(n)$ 至少有一个是 0。若 $v_p(n)=0$，上式等于 $\min(v_p(m),v_p(k))+0=v_p(\gcd(m,k))+v_p(\gcd(n,k))$，因为此时 $v_p(\gcd(n,k))=\min(0,v_p(k))=0$。$v_p(m)=0$ 的情形同理。每个素数的指数都对上，所以 $\gcd(mn,k)=\gcd(m,k)\gcd(n,k)$。又 $f(1)=\gcd(1,k)=1$。

取 $k=12$：$\gcd(4\cdot3,12)=12=4\cdot3=\gcd(4,12)\gcd(3,12)$，符合积性；但 $\gcd(2\cdot4,12)=4$，而 $\gcd(2,12)\gcd(4,12)=2\cdot4=8$。2 与 4 不互素，乘法公式失效，所以 f 不是完全积性的。边界 $k=1$ 时，$f(n)=1$ 对所有 n 成立，此时却是完全积性的。 ♦

13.2.2 积性函数由素数幂上的值决定

定理 13.4　设 f 是积性函数，$n=p_1^{e_1}p_2^{e_2}\cdots p_r^{e_r}$ 是 n 的标准分解（p_i 是互不相同的素数，$e_i\geq1$）。则

$$f(n)=f(p_1^{e_1})\,f(p_2^{e_2})\cdots f(p_r^{e_r}).$$

若 f 是完全积性的，进一步有 $f(n)=f(p_1)^{e_1}\cdots f(p_r)^{e_r}$。反过来，任意指定每个素数幂 p^e 处的值，恰有一个积性函数取这些值。

证明　对 r 归纳。$r=0$ 时 $n=1$，$f(1)=1$ 等于空积。设 $r\geq1$，令 $m=p_2^{e_2}\cdots p_r^{e_r}$。$p_1^{e_1}$ 与 m 没有公共素因数，所以互素，于是 $f(n)=f(p_1^{e_1})f(m)$；由归纳假设 $f(m)=\prod_{i\geq2}f(p_i^{e_i})$。完全积性时再用 $f(p^e)=f(p\cdots p)=f(p)^e$。

反过来，给定素数幂上的值，用上面的乘积公式定义 $f(n)$（$f(1)=1$）。由算术基本定理，标准分解唯一，所以 f 定义得好。若 $\gcd(m,n)=1$，m 与 n 的分解用到的素数互不相同，mn 的标准分解就是两者合在一起，乘积公式自然拆成 $f(m)f(n)$。唯一性由第一部分得到。 ■

核心思想　积性函数与「每个素数幂上的一个值」一一对应。要证明关于积性函数的恒等式，或算它在某个 n 处的值，只需处理 $n=p^e$ 这一种情形。第 5 讲的 $\varphi(n)=\prod p_i^{e_i-1}(p_i-1)$ 就是这样得到的：先算 $\varphi(p^e)=p^{e-1}(p-1)$，再相乘。

图 13.1 区分完全积性、积性与一般算术函数，反例说明包含关系严格。

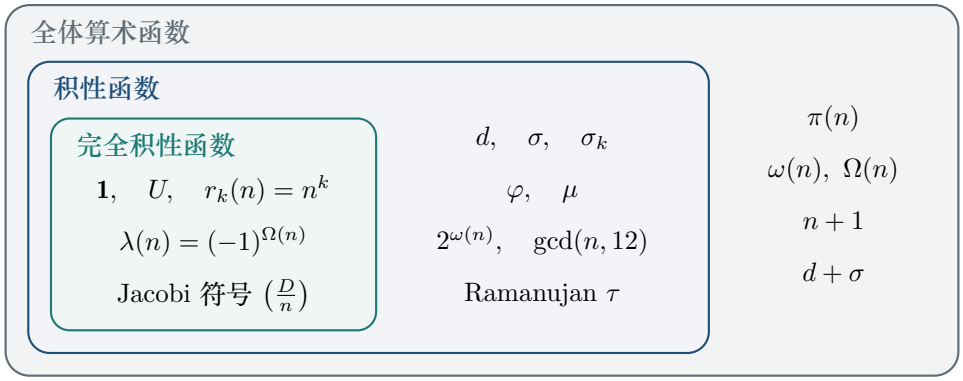


图 13.1　三类算术函数的包含关系。外层右栏的函数均在 $n=1$ 处不等于 1，因此不积性；中层选用 $\gcd(n,12)$，避免把 $k=1$ 时完全积性的特例归错类。

易错提醒

- 积性只保证互素时 $f(mn)=f(m)f(n)$。用积性计算时，要把 n 拆成不同素数的幂，不能随便拆成两个因数。例如 $d(12)\neq d(2)d(6)=2\cdot 4=8$，正确的是 $d(12)=d(4)d(3)=3\cdot 2=6$。
- 两个积性函数的逐点乘积 $f(n)g(n)$ 仍是积性的，但和 $f+g$ 一般不是：$(f+g)(1)=2\neq 1$。例如 $d+\sigma$ 不是积性函数。

想一想　设 f 是积性函数，并且对每个素数 p 都有 $f(p)=p$。f 一定是 $r_1(n)=n$ 吗？如果 f 是完全积性的呢？

答　不一定。由定理 13.4，f 在 $p^2,p^3,\dots$ 处的值可以随意指定。例如令 $f(p^e)=p$（对一切 $e\geq 1$），得到 $f(n)=\prod_{p\mid n}p$，它满足 $f(p)=p$，但 $f(4)=2\neq 4$。若 f 是完全积性的，则 $f(p^e)=f(p)^e=p^e$，此时 f 与 r_1 在每个素数幂上相等，所以 $f=r_1$。

13.3　Dirichlet 卷积

把 d 和 σ 的定义改写一下：

$$d(n)=\sum_{d\mid n}1\cdot 1=\sum_{d\mid n}U(d)\,U\!\left(\frac{n}{d}\right),\qquad \sigma(n)=\sum_{d\mid n}d\cdot 1=\sum_{d\mid n}r_1(d)\,U\!\left(\frac{n}{d}\right).$$

两者都是「对 n 的每个因数 d，把一个函数在 d 处的值乘以另一个函数在 n/d 处的值，再求和」。这个运算值得单独起名字。

定义 13.5（Dirichlet 卷积）　算术函数 f,g 的 **Dirichlet 卷积**（Dirichlet convolution）是算术函数

$$(f*g)(n)=\sum_{d\mid n}f(d)\,g\!\left(\frac{n}{d}\right)=\sum_{ab=n}f(a)\,g(b),$$

其中第二种写法对满足 $ab=n$ 的全部有序正整数对 (a,b) 求和。

两种写法相同：d 跑遍 n 的正因数时，$(a,b)=(d,n/d)$ 恰好跑遍 $ab=n$ 的全部有序对。第二种写法把 a,b 放在平等的位置上，证明性质时更方便。

图 13.2 把卷积求和的各项放到因子对组成的双曲线上。

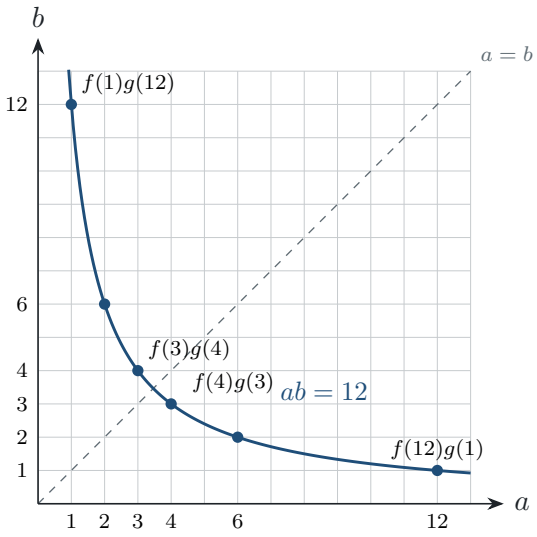


图 13.2　$(f*g)(12)$ 就是对双曲线 $ab=12$ 上的 6 个格点求 $f(a)g(b)$ 之和。关于直线 $a=b$ 翻转，把 (a,b) 换成 (b,a)，得到同一组格点，这就是交换律。

前后联系　微积分里的卷积是 $(f*g)(x)=\int f(t)g(x-t)\,dt$，两个自变量相加等于 x；多项式乘法中 x^n 的系数是 $\sum_{i+j=n}a_ib_j$，下标也是相加。Dirichlet 卷积把「相加」换成了「相乘」：$ab=n$。它对应的是 Dirichlet 级数 $\sum f(n)n^{-s}$ 的乘法，因为 $a^{-s}b^{-s}=(ab)^{-s}$。第 14 讲会沿这条线走到 ζ 函数。

■ **例 13.4**　计算 $(U*U)(12)$、$(r_1*U)(12)$、$(U*r_1)(12)$、$(\varphi*U)(12)$、$(r_1*r_1)(12)$。

解　12 的因数对 (a,b) 是 $(1,12),(2,6),(3,4),(4,3),(6,2),(12,1)$，共 6 对。逐个代入：

- $(U*U)(12)=\sum_{ab=12}1\cdot1=6$，就是有序对的个数，即 $d(12)$。
- $(r_1*U)(12)=\sum_{ab=12}a=1+2+3+4+6+12=28=\sigma(12)$。
- $(U*r_1)(12)=\sum_{ab=12}b=12+6+4+3+2+1=28$。同样的数换了顺序相加，与上一个相等。
- $(\varphi*U)(12)=\varphi(1)+\varphi(2)+\varphi(3)+\varphi(4)+\varphi(6)+\varphi(12)=1+1+2+2+2+4=12$。
- $(r_1*r_1)(12)=\sum_{ab=12}ab=6\cdot12=72$，因为每一项都等于 $ab=12$。

同样的推理对一般的 n 都成立，得到

$$U*U=d,\qquad r_k*U=\sigma_k,\qquad (r_k*r_k)(n)=n^k\,d(n).$$

第四个结果 $(\varphi*U)(12)=12$ 不是巧合，§13.4 会证明 $\varphi*U=r_1$。 ♦

命题 13.6（卷积的运算律）　对任意算术函数 f, g, h：

(1) 单位元：$f * \mathbf{1} = \mathbf{1} * f = f$；

(2) 交换律：$f * g = g * f$；

(3) 结合律：$(f * g) * h = f * (g * h)$；

(4) 分配律：$f * (g + h) = f * g + f * h$。

证明　(1) $(f * \mathbf{1})(n) = \sum_{d \mid n} f(d)\,\mathbf{1}(n/d)$。只有 $n/d = 1$ 即 $d = n$ 的那一项不为零，所以和等于 $f(n)$。再由 (2) 得 $\mathbf{1} * f = f$。

(2) $\sum_{ab=n} f(a)g(b)$ 与 $\sum_{ba=n} g(b)f(a)$ 是对同一组有序对求和，只是换了字母。

(3) 两边都展开成对有序三元组求和：

$$\big((f * g) * h\big)(n) = \sum_{mc=n} \Big(\sum_{ab=m} f(a)g(b) \Big) h(c) = \sum_{abc=n} f(a)g(b)h(c),$$

$$\big(f * (g * h)\big)(n) = \sum_{am=n} f(a) \Big(\sum_{bc=m} g(b)h(c) \Big) = \sum_{abc=n} f(a)g(b)h(c).$$

右边相同，所以左边相等。

(4) 由和式的线性立即得到。 ■

于是全体算术函数在加法和 Dirichlet 卷积下成为一个有单位元的交换环，单位元就是 $\mathbf{1}$，这正是它名字的来历。结合律保证 $f*g*h$ 不必加括号，它等于 $\sum_{abc=n} f(a)g(b)h(c)$。例如 $(U*U*U)(n)$ 就是满足 $abc = n$ 的有序三元组个数。

易错提醒

- 不要把 U 与 $\mathbf{1}$ 弄混。$\mathbf{1}$ 只在 $n = 1$ 处为 1，它是卷积的单位元；U 处处为 1，与它卷积是「对因数求和」：$(f * U)(n) = \sum_{d \mid n} f(d)$，一般不等于 $f(n)$。
- 卷积不是逐点乘积。$U \cdot U = U$，而 $U * U = d$；$r_1 \cdot r_1 = r_2$，而 $(r_1 * r_1)(12) = 72 \neq 144 = r_2(12)$。

13.4　积性函数的卷积

上一节得到 $d = U * U$，$\sigma = r_1 * U$，而 U、r_1 都是积性的，d、σ 也是积性的。这提示了一个一般规律。证明它要用到互素的数的因数怎样组合。

引理 13.7　设 $\gcd(m, n) = 1$。对 m 的每个正因数 d_1 和 n 的每个正因数 d_2，乘积 $d_1 d_2$ 是 mn 的正因数；反过来，mn 的每个正因数 d 都恰好以一种方式写成这样的乘积，即 $d_1 = \gcd(d, m)$，$d_2 = \gcd(d, n)$。

证明　由 $m = d_1 u, n = d_2 v$，有 $mn = d_1 d_2 uv$，故 $d_1 d_2 \mid mn$。设 $d \mid mn$。m, n 没有公共素因数，对 m 的素因数 p，d 中 p 的指数不超过它在 mn 中的指数，也就是在 m 中的指数；对 n 的素

因数同理。所以把 d 的分解按「属于 m 的素数」和「属于 n 的素数」分成两组，得到 $d=d_1d_2$，$d_1\mid m$，$d_2\mid n$。分组方式由 d 的素因数分解唯一决定，这就是 $d_1=\gcd(d,m)$，$d_2=\gcd(d,n)$。 ■

图 13.3 用 $72=8\cdot 9$ 展示这个双射：每个因数由一个 8 的因数和一个 9 的因数相乘得到。

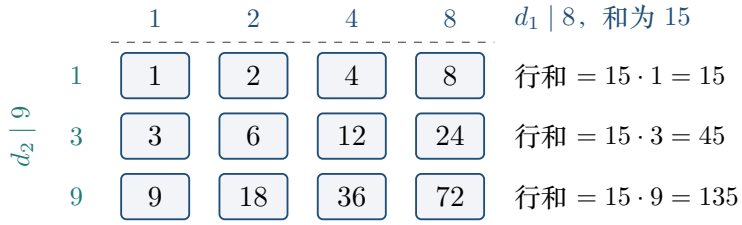


图 13.3　$72=8\cdot 9$，$\gcd(8,9)=1$。8 的因数排成列，9 的因数排成行，交叉处的乘积恰好把 72 的 12 个因数各列出一次。所以 $d(72)=4\cdot 3=d(8)d(9)$，$\sigma(72)=15+45+135=195=15\cdot 13=\sigma(8)\sigma(9)$。

读一行时，固定来自 9 的因数，再让来自 8 的因数跑遍四列；因此行和分别是 15、45、135。互素条件保证每个因数只占一个格子：若换成 $4\cdot 6$，因数 2 就会同时出现在 $1\cdot 2$ 与 $2\cdot 1$ 两格。

定理 13.8　若 f,g 都是积性函数，则 $f*g$ 也是积性函数。

证明　先看 $n=1$：$(f*g)(1)=f(1)g(1)=1$，所以 $f*g$ 不恒为零。设 $\gcd(m,n)=1$。由引理 13.7，对 mn 的因数 d 求和，等于对 $d_1\mid m$、$d_2\mid n$ 的全部组合求和，其中 $d=d_1d_2$。这时 $\frac{mn}{d}=\frac{m}{d_1}\cdot\frac{n}{d_2}$，而 d_1 与 d_2 互素，$\frac{m}{d_1}$ 与 $\frac{n}{d_2}$ 也互素。于是

$$\begin{aligned}(f*g)(mn)&=\sum_{d_1\mid m}\sum_{d_2\mid n}f(d_1d_2)\,g\!\left(\frac{m}{d_1}\cdot\frac{n}{d_2}\right)=\sum_{d_1\mid m}\sum_{d_2\mid n}f(d_1)f(d_2)\,g\!\left(\frac{m}{d_1}\right)g\!\left(\frac{n}{d_2}\right)\\&=\Big(\sum_{d_1\mid m}f(d_1)\,g\!\left(\frac{m}{d_1}\right)\Big)\Big(\sum_{d_2\mid n}f(d_2)\,g\!\left(\frac{n}{d_2}\right)\Big)=(f*g)(m)\,(f*g)(n).\end{aligned}$$

第二个等号用了 f、g 的积性。 ■

推论 13.9　若 f 是积性函数，则 $F(n)=\sum_{d\mid n}f(d)$ 也是积性函数。特别地，d、σ、σ_k 都是积性函数，并且对 $n=p_1^{e_1}\cdots p_r^{e_r}$ 有

$$d(n)=\prod_{i=1}^{r}(e_i+1),\qquad \sigma_k(n)=\prod_{i=1}^{r}\big(1+p_i^{k}+p_i^{2k}+\cdots+p_i^{e_ik}\big)=\prod_{i=1}^{r}\frac{p_i^{(e_i+1)k}-1}{p_i^{k}-1}\quad(k\ge 1).$$

证明　$F=f*U$，U 是积性的，由定理 13.8 得 F 积性。$d=U*U$，$\sigma_k=r_k*U$ 都属于这种情形。由定理 13.4，只需算素数幂：p^e 的因数是 $1,p,\dots,p^e$，共 $e+1$ 个，所以 $d(p^e)=e+1$；它们的 k 次方之和是公比为 p^k 的等比数列，所以 $\sigma_k(p^e)=1+p^k+\cdots+p^{ek}=\frac{p^{(e+1)k}-1}{p^k-1}$。 ■

■ **例 13.5**　求 $d(720)$、$\sigma(720)$、$\sigma_2(720)$，并判断 720 是亏数、完全数还是盈数。

解　$720 = 2^4 \cdot 3^2 \cdot 5$。由推论 13.9，

$$d(720) = (4+1)(2+1)(1+1) = 30,$$

$$\sigma(720) = (1+2+4+8+16)(1+3+9)(1+5) = 31 \cdot 13 \cdot 6 = 2418,$$

$$\sigma_2(720) = (1+4+16+64+256)(1+9+81)(1+25) = 341 \cdot 91 \cdot 26 = 806806.$$

$2418 > 1440 = 2 \cdot 720$，所以 720 是盈数。逐个列出 30 个因数再相加也能得到 2418，但公式只需要三次小加法和两次乘法。 ♦

直观理解　展开乘积 $(1+2+4+8+16)(1+3+9)(1+5)$，每一项都是从三个括号里各挑一个数相乘，形如 $2^a 3^b 5^c$，其中 $0 \le a \le 4$，$0 \le b \le 2$，$0 \le c \le 1$。这恰好是 720 的全部因数，每个出现一次。所以展开后的和就是 $\sigma(720)$，项数 $5 \cdot 3 \cdot 2 = 30$ 就是 $d(720)$。图 13.3 是两个括号时的同一件事。

13.4.1　φ 的因数和

例 13.4 中算出 $(\varphi * U)(12) = 12$。这对一切 n 都成立。

定理 13.10　对一切正整数 n，$\displaystyle\sum_{d \mid n} \varphi(d) = n$，即 $\varphi * U = r_1$。

证明　φ 积性，由推论 13.9，$\varphi * U$ 积性；r_1 也积性。由定理 13.4，两个积性函数只要在每个素数幂上相等就处处相等。对 $n = p^e$，$1, \dots, p^j$ 中不被 p 整除的数有 $p^j - p^{j-1}$ 个，所以 $\varphi(p^j) = p^j - p^{j-1}$ $(j \ge 1)$。于是

$$\sum_{d \mid p^e} \varphi(d) = 1 + (p-1) + (p^2 - p) + \cdots + (p^e - p^{e-1}) = p^e,$$

中间各项正负相消。 ■

这个证明很短，但没有说明为什么因数的 φ 值加起来恰好是 n。Gauss 给过一个直接数数的证明。

■ **例 13.6**　把 $\frac{1}{12}, \frac{2}{12}, \dots, \frac{12}{12}$ 都约成最简分数，按分母分组，数每组有几个。

解　约分后的结果见图 13.4。分母只能是 12 的因数。分母为 12 的是 $\frac{1}{12}, \frac{5}{12}, \frac{7}{12}, \frac{11}{12}$，分子恰好是 $1, \dots, 12$ 中与 12 互素的数，共 $\varphi(12) = 4$ 个。分母为 6 的是 $\frac{1}{6}, \frac{5}{6}$，分子是与 6 互素的 $1, 5$，共 $\varphi(6) = 2$ 个。依此类推，分母为 d 的分数恰好是 $\frac{a}{d}$，$1 \le a \le d$，$\gcd(a, d) = 1$，共 $\varphi(d)$ 个。12 个分数分完，得到 $\varphi(1) + \varphi(2) + \varphi(3) + \varphi(4) + \varphi(6) + \varphi(12) = 1 + 1 + 2 + 2 + 2 + 4 = 12$。 ♦

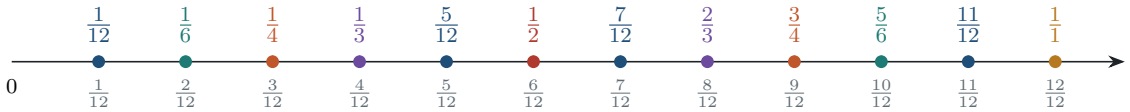


图 13.4　$\frac{k}{12}$ 约分后按分母着色。分母为 d 的分数恰有 $\varphi(d)$ 个，6 组合起来正好是全部 12 个分数。

对一般的 n 论证完全一样：$\frac{k}{n}$（$1 \le k \le n$）约分成 $\frac{a}{d}$，其中 $d \mid n$，$1 \le a \le d$，$\gcd(a,d)=1$；反过来每个这样的 $\frac{a}{d}$ 都来自唯一的 $k = a \cdot \frac{n}{d}$。所以 $n = \sum_{d \mid n} \varphi(d)$。这给出了定理 13.10 的第二个证明，它不用积性。

前后联系　第 8 讲用原根数过同一件事：模素数 p 的 $p-1$ 个非零剩余按阶分组，阶为 d 的恰有 $\varphi(d)$ 个（$d \mid p-1$），所以 $\sum_{d \mid p-1} \varphi(d) = p-1$。那里是在循环群里按阶分组，这里是按约分后的分母分组。两种分组都把 n 个对象分成以 n 的因数为标号的若干组，第 d 组恰有 $\varphi(d)$ 个。

13.4.2　一个深刻的例子：Ramanujan τ 函数

到目前为止，每个积性函数的积性都能从定义直接看出来，或者用一次卷积定理得到。下面这个函数的积性完全看不出来。把无穷乘积展开成 q 的幂级数：

$$q\prod_{i=1}^{\infty}(1-q^i)^{24} = \sum_{n=1}^{\infty}\tau(n)\,q^n = q - 24q^2 + 252q^3 - 1472q^4 + 4830q^5 - 6048q^6 + \cdots,$$

q^n 的系数 $\tau(n)$ 叫 **Ramanujan τ 函数**（Ramanujan tau function）。

■ **例 13.7**　从定义算出 $\tau(2)$ 和 $\tau(3)$，并用上面列出的值检查 $\tau(6)$ 与 $\tau(2)\tau(3)$、$\tau(4)$ 与 $\tau(2)^2$。

解　$\tau(n)$ 是 $\prod(1-q^i)^{24}$ 中 q^{n-1} 的系数，所以只需把乘积展开到 q^2。$i \ge 3$ 的因子 $(1-q^i)^{24} = 1 - 24q^i + \cdots$ 对 q^2 及更低次没有贡献，只看前两个：

$$(1-q)^{24} = 1 - 24q + \binom{24}{2}q^2 - \cdots = 1 - 24q + 276q^2 - \cdots, \qquad (1-q^2)^{24} = 1 - 24q^2 + \cdots.$$

相乘：q^1 的系数是 -24；q^2 的系数是 $276 - 24 = 252$。所以 $\tau(2) = -24$，$\tau(3) = 252$。

于是 $\tau(2)\tau(3) = -24 \cdot 252 = -6048 = \tau(6)$，与积性一致；而 $\tau(2)^2 = 576 \ne -1472 = \tau(4)$，所以 τ 不是完全积性的。实际上 $\tau(4) = \tau(2)^2 - 2^{11} = 576 - 2048$。 ♦

定理 13.11（Mordell，1917）　τ 是积性函数。

本课程不证明此定理；证明要用模形式（$\sum \tau(n)q^n$ 是一个权为 12 的模形式，积性来自 Hecke 算子），超出本课程的范围。它说明「积性」不总是来自因数的组合，有时反映了更深的对称性。

历史与背景 Ramanujan 在 1916 年的论文中算出了 $\tau(n)$ 的前 30 个值，猜想了三件事：τ 是积性的；$\tau(p^{e+1}) = \tau(p)\tau(p^e) - p^{11}\tau(p^{e-1})$；$|\tau(p)| \le 2p^{11/2}$。前两条在第二年由 Mordell 证明。第三条难得多，1974 年 Deligne 证明 Weil 猜想时才作为推论得到。至今未解决的是 Lehmer 在 1947 年提出的问题：是否对一切 n 都有 $\tau(n) \neq 0$？

13.5　偶完全数：Euclid–Euler 定理

例 13.2 中观察到 $6, 28, 496$ 都形如 $2^{m-1}(2^m-1)$，其中 2^m-1 是素数。有了 σ 的积性，可以把这个观察变成定理，而且两个方向都能证明。

定理 13.12（Euclid–Euler）

(1)（Euclid）若 2^m-1 是素数，则 $2^{m-1}(2^m-1)$ 是完全数。

(2)（Euler）每个偶完全数都形如 $2^{m-1}(2^m-1)$，其中 2^m-1 是素数。

思路 把偶数写成 $n = 2^r s, r \ge 1, s$ 为奇数。2^r 与 s 互素，所以 $\sigma(n) = \sigma(2^r)\sigma(s) = (2^{r+1}-1)\sigma(s)$。(1) 直接代入。(2) 把 $\sigma(n) = 2n$ 写成 $(2^{r+1}-1)\sigma(s) = 2^{r+1}s$，左边的奇数因子 $2^{r+1}-1$ 必须整除 s；再比较 $\sigma(s)$ 与 s 的大小，逼出 s 只有两个因数。

证明 (1) 设 $p = 2^m-1$ 是素数，$n = 2^{m-1}p$。2^{m-1} 与奇数 p 互素，所以

$$\sigma(n) = \sigma(2^{m-1})\,\sigma(p) = (1+2+\cdots+2^{m-1})(1+p) = (2^m-1)\cdot 2^m = 2\cdot 2^{m-1}(2^m-1) = 2n.$$

(2) 设 n 是偶完全数，$n = 2^r s$，$r \ge 1$，s 为奇数。记 $M = 2^{r+1}-1$，则 $\sigma(2^r) = M$，条件 $\sigma(n) = 2n$ 变成

$$M\,\sigma(s) = 2^{r+1}s. \tag{$*$}$$

M 是奇数，与 2^{r+1} 互素，而 M 整除右边，由 Euclid 引理 $M \mid s$。写 $s = Mt$，代回 $(*)$ 并约去 M：

$$\sigma(s) = 2^{r+1}t = (M+1)t = s + t.$$

因为 $r \ge 1$，有 $M \ge 3$，所以 $t < s$，t 和 s 是 s 的两个不同的因数，它们的和已经等于 $\sigma(s)$。若 $t > 1$，则 1 是第三个因数，$\sigma(s) \ge 1 + t + s > s + t$，矛盾。所以 $t = 1$，并且 s 的因数只有 1 和 s，即 $s = M = 2^{r+1}-1$ 是素数。令 $m = r+1$，得 $n = 2^{m-1}(2^m-1)$。 ■

关键一步是 $\sigma(s) = s + t$：它说 s 的全部因数之和只比「s 和 t 两个因数之和」不多，于是 s 没有别的因数。这里用到了 $r \ge 1$，也就是 n 是偶数。对奇数 n，$r = 0$，$M = 1$，$(*)$ 退化成 $\sigma(s) = 2s$，什么也没得到。这就是奇完全数问题难以下手的原因之一。

■ **例 13.8**　(1) 用定理 13.12 找出第四个偶完全数，并验证。(2) $m = 11$ 时，$2^{10}(2^{11}-1)$ 是完全数吗？

解 (1) 依次试 $m = 2, 3, 4, 5, 6, 7$：$2^m - 1 = 3, 7, 15, 31, 63, 127$。其中 $15 = 3\cdot 5$，$63 = 3^2\cdot 7$ 不是素数。127 是素数：$\sqrt{127} < 12$，而 127 不被 $2, 3, 5, 7, 11$ 整除。所以前四个偶完全数对应

$m=2,3,5,7$，第四个是

$$2^6\cdot 127=64\cdot 127=8128.$$

验证：$\sigma(8128)=\sigma(2^6)\sigma(127)=127\cdot 128=16256=2\cdot 8128$。

(2) $2^{11}-1=2047=23\cdot 89$，不是素数，由定理的 (2)，$n=2^{10}\cdot 2047=2096128$ 不是完全数。直接算也可以：$2^{10},23,89$ 两两互素，

$$\sigma(n)=\sigma(2^{10})\,\sigma(23)\,\sigma(89)=2047\cdot 24\cdot 90=4421520>4192256=2n,$$

n 是盈数。2047 的因数 $23,89$ 让 σ 多出了一截。 ♦

易错提醒

- 2^m-1 是素数时 m 必须是素数：若 $m=ab$，$a,b>1$，在 $x^b-1=(x-1)(x^{b-1}+\cdots+x+1)$ 中取 $x=2^a$，得 $2^a-1\mid 2^m-1$，而 $1<2^a-1<2^m-1$。例如 $2^3-1=7$ 整除 $2^6-1=63$。
- 反过来不对：$m=11$ 是素数，$2^{11}-1=23\cdot 89$ 却不是。「m 是素数」只是必要条件。
- 定理只说偶完全数。它对奇完全数什么也没说：既没有证明它们不存在，也没有给出它们的形状。

想一想　素数幂 p^e 能是完全数吗？

答　不能。$\sigma(p^e)-p^e=1+p+\cdots+p^{e-1}=\dfrac{p^e-1}{p-1}\le p^e-1<p^e$，所以真因数之和小于 p^e，每个素数幂都是亏数。练习 10 沿这个思路证明：奇完全数至少有三个不同的素因数。

13.6　Möbius 函数

由 §13.3，算术函数在加法和卷积下构成交换环，单位元是 $\mathbf{1}$。环里自然要问：哪些元素可逆？最值得问的是 U，因为与 U 卷积就是「对因数求和」：$F=f*U$ 即 $F(n)=\sum_{d\mid n}f(d)$。若 U 有逆 g，则 $F*g=f*(U*g)=f$，就能从因数和 F 反解出 f。

■ **例 13.9**　求算术函数 g，使 $g*U=\mathbf{1}$，即对一切 n 都有 $\sum_{d\mid n}g(d)=\mathbf{1}(n)$。先算出 $g(1),\dots,g(12)$。

解　把 $d=n$ 那一项单独写出来，条件变成

$$g(n)=\mathbf{1}(n)-\sum_{d\mid n,\ d<n}g(d).$$

右边只用到比 n 小的 d 处的值，所以可以从 $n=1$ 开始逐个解出 g，而且每一步的解唯一。

- $n=1$：$g(1)=1$。
- $n=p$ 为素数：$g(p)=0-g(1)=-1$。所以 $g(2)=g(3)=g(5)=g(7)=g(11)=-1$。
- $n=4$：$g(4)=-(g(1)+g(2))=-(1-1)=0$。同理 $g(9)=-(g(1)+g(3))=0$。
- $n=8$：$g(8)=-(g(1)+g(2)+g(4))=-(1-1+0)=0$。
- $n=6$：$g(6)=-(g(1)+g(2)+g(3))=-(1-1-1)=1$。同理 $g(10)=1$。

▶ $n=12$：$g(12)=-(g(1)+g(2)+g(3)+g(4)+g(6))=-(1-1-1+0+1)=0$。

结果依次是 $1,-1,-1,0$，$-1,1,-1,0$，$0,1,-1,0$，恰好是表 13.1 的最后一行。规律：n 含平方因子（$4,8,9,12$）时 $g(n)=0$；n 是 r 个不同素数之积时 $g(n)=(-1)^r$。 ♦

若一个正整数不被任何素数的平方整除，称它**无平方因子**（squarefree）。$1,2,3,5,6,7,10,11$ 都无平方因子，$4,8,9,12$ 不是。

定义 13.13（Möbius 函数）　Möbius 函数（Möbius function）μ 定义为

$$\mu(n)=\begin{cases}(-1)^{\omega(n)}, & n \text{ 无平方因子},\\ 0, & \text{否则}.\end{cases}$$

等价地：$\mu(1)=1$；若 $n=p_1p_2\cdots p_r$ 是 r 个不同素数之积，$\mu(n)=(-1)^r$；若有素数 p 使 $p^2\mid n$，$\mu(n)=0$。

例如 $30=2\cdot3\cdot5$，$\mu(30)=-1$；$210=2\cdot3\cdot5\cdot7$，$\mu(210)=1$；$1001=7\cdot11\cdot13$，$\mu(1001)=-1$；$60=2^2\cdot15$，$\mu(60)=0$。

μ 是积性的。设 $\gcd(m,n)=1$。若 m 或 n 含平方因子，mn 也含，两边都是 0。若 m,n 都无平方因子，它们没有公共素因数，mn 也无平方因子，而 $\omega(mn)=\omega(m)+\omega(n)$，所以 $(-1)^{\omega(mn)}=(-1)^{\omega(m)}(-1)^{\omega(n)}$。它在素数幂上的值是

$$\mu(1)=1,\qquad \mu(p)=-1,\qquad \mu(p^e)=0\quad(e\ge2).$$

μ 不是完全积性的：$\mu(4)=0\neq1=\mu(2)^2$。

定理 13.14　$\mu*U=\mathbf{1}$，即

$$\sum_{d\mid n}\mu(d)=\begin{cases}1, & n=1,\\ 0, & n>1.\end{cases}$$

证明一（积性）　μ、U 积性，由定理 13.8，$\mu*U$ 积性；$\mathbf{1}$ 也积性。由定理 13.4，只需在 $n=1$ 和素数幂 p^e（$e\ge1$）上比较。$n=1$ 时 $(\mu*U)(1)=\mu(1)=1$。$n=p^e$ 时，因数是 $1,p,\dots,p^e$，从 p^2 起 μ 都是 0：

$$(\mu*U)(p^e)=\mu(1)+\mu(p)+\mu(p^2)+\cdots+\mu(p^e)=1+(-1)+0+\cdots+0=0=\mathbf{1}(p^e).$$ ■

证明二（直接数）　设 $n>1$，$p_1,\dots,p_r$ 是 n 的全部不同素因数，$r\ge1$。含平方因子的因数贡献 0，只剩无平方因子的因数，它们是从 $\{p_1,\dots,p_r\}$ 中选一个子集相乘。选 k 个素数的子集有 $\binom{r}{k}$ 个，每个贡献 $(-1)^k$，所以

$$\sum_{d\mid n}\mu(d)=\sum_{k=0}^{r}\binom{r}{k}(-1)^k=(1-1)^r=0.$$ ■

图 13.5 用因数格展示含平方因子的项如何在 Möbius 求和中消失。

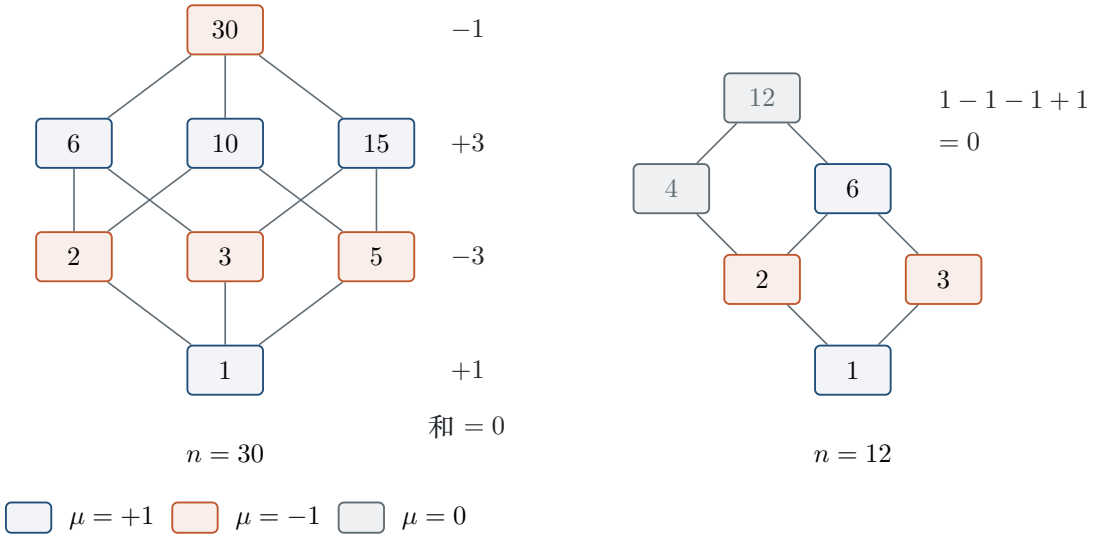


图 13.5　因数格，连线表示乘一个素数。左：30 的 8 个因数排成立方体，按层（素因数个数 k）的 μ 值之和是 $\binom{3}{k}(-1)^k$，合计 $1-3+3-1=0$。右：12 的因数中 $4,12$ 含平方因子，$\mu=0$，剩下的 $1,2,3,6$ 与 $n=6$ 的情形相同。

直观理解　μ 是**容斥原理**的符号。数 $1,\dots,30$ 中不被 $2,3,5$ 任何一个整除的数：先全数，减去 $2,3,5$ 的倍数，加回被两个整除的，再减去被三个都整除的，

$$30-(15+10+6)+(5+3+2)-1=8.$$

每一项是 $\frac{30}{d}$，前面的符号恰好是 $\mu(d)$，d 跑遍 30 的因数。一个与 30 有公因数的数 k 在这个式子里被数了 $\sum_{d\mid\gcd(k,30)}\mu(d)=0$ 次，与 30 互素的 k 被数了 $\mu(1)=1$ 次。所以结果是 $\varphi(30)=8$。第 14 讲会把它写成一般公式 $\varphi(n)=\sum_{d\mid n}\mu(d)\frac{n}{d}$。

13.6.1　μ 是 U 的卷积逆

例 13.9 的递推对任何 $f(1)\neq 0$ 的函数都能做，这给出可逆元的完整刻画（习题集 6 第 8(b) 题）。

定理 13.15　算术函数 f 有卷积逆 f^{-1}（即 $f*f^{-1}=f^{-1}*f=\mathbf{1}$）当且仅当 $f(1)\neq 0$，并且逆是唯一的。

证明　若 $f*g=\mathbf{1}$，在 $n=1$ 处取值得 $f(1)g(1)=1$，所以 $f(1)\neq 0$。反过来设 $f(1)\neq 0$，令 $g(1)=1/f(1)$，并对 $n>1$ 依次令

$$g(n)=-\frac{1}{f(1)}\sum_{d\mid n,\ d>1} f(d)\,g\!\left(\frac{n}{d}\right).$$

右边只用到 g 在小于 n 处的值，所以 g 定义得好，而且这个式子正是 $(f*g)(n)=0$ 移项的结果。于是 $f*g=\mathbf{1}$，由交换律 $g*f=\mathbf{1}$。若另有 $h*f=\mathbf{1}$，则 $h=h*(f*g)=(h*f)*g=g$，所以逆唯一。 ∎

$U(1)=1\neq 0$，所以 U 有唯一的逆；定理 13.14 说 $\mu * U=\mathbf{1}$，所以这个逆就是 μ，与例 13.9 逐项算出的结果一致。积性函数都满足 $f(1)=1$，因此都可逆。

前后联系　有了 $U^{-1}=\mu$，「对因数求和」就能撤销：若 $F=f*U$，两边卷积 μ 得 $F*\mu=f*(U*\mu)=f$，即 $f(n)=\sum_{d\mid n}\mu(d)F(n/d)$。这就是 **Möbius 反演公式**（Möbius inversion formula），第 14 讲的主题。例如由 $\varphi * U=r_1$ 立刻得到 $\varphi=\mu * r_1$。

■ **例 13.10**　设 f 是积性函数，$n>1$，$p_1,\dots,p_r$ 是 n 的不同素因数。证明 $\sum_{d\mid n}\mu(d)f(d)=\prod_{i=1}^{r}\bigl(1-f(p_i)\bigr)$，并对 $f=d$、$f=\sigma$、$f(n)=\frac{1}{n}$ 写出结果，在 $n=12$ 处核对。

解　逐点乘积 μf 是积性的，所以 $\sum_{d\mid n}\mu(d)f(d)=(\mu f*U)(n)$ 也是积性的，只需算素数幂。p^e（$e\geq 1$）的因数中只有 1 和 p 使 $\mu\neq 0$，所以和等于 $f(1)-f(p)=1-f(p)$。再把各素数幂的结果相乘即得。

- $f=d$：$1-d(p)=1-2=-1$，所以 $\sum_{d\mid n}\mu(d)d(d)=(-1)^{\omega(n)}$。$n=12$：$d(1)-d(2)-d(3)+d(6)=1-2-2+4=1=(-1)^2$。
- $f=\sigma$：$1-\sigma(p)=-p$，所以和为 $(-1)^{\omega(n)}p_1\cdots p_r$。$n=12$：$1-3-4+12=6=(-2)(-3)$。
- $f(n)=\frac{1}{n}$：和为 $\prod(1-\frac{1}{p_i})$。$n=12$：$1-\frac{1}{2}-\frac{1}{3}+\frac{1}{6}=\frac{1}{3}=(1-\frac{1}{2})(1-\frac{1}{3})$。乘以 n 得 $12\cdot\frac{1}{3}=4=\varphi(12)$，这正是第 5 讲的公式 $\varphi(n)=n\prod_{p\mid n}(1-\frac{1}{p})$。

取 $f=U$ 时 $1-U(p)=0$，又回到定理 13.14。 ♦

唯一性本身也是一个工具：要证某个函数 c 等于 μ，只要证 $c*U=\mathbf{1}$。

■ **例 13.11　改编自习题集 6 第 10 题**　记 $c(n)=\sum_{1\leq a\leq n,\ \gcd(a,n)=1}e^{2\pi i a/n}$，即全部本原 n 次单位根之和。证明 $c(n)=\mu(n)$。

解　x^n-1 的 n 个根是 $\zeta^k=e^{2\pi ik/n}$，$k=1,\dots,n$。像例 13.6 那样把 $\frac{k}{n}$ 约成最简分数 $\frac{a}{d}$，则 $\zeta^k=e^{2\pi ia/d}$ 是本原 d 次单位根，$d\mid n$；每个本原 d 次单位根（$d\mid n$）恰好出现一次。所以按阶分组，

$$\sum_{d\mid n}c(d)=\sum_{k=1}^{n}e^{2\pi ik/n}=\begin{cases}1, & n=1,\\ 0, & n>1,\end{cases}$$

后一种情形是因为 $\zeta\neq 1$，等比数列求和 $\zeta\frac{\zeta^n-1}{\zeta-1}=0$。这说明 $c*U=\mathbf{1}$，由逆的唯一性，$c=\mu$。

核对：$n=4$，本原根是 $i,-i$，和为 $0=\mu(4)$；$n=6$，本原根是 $e^{\pm\pi i/3}=\frac{1}{2}\pm\frac{\sqrt{3}}{2}i$，和为 $1=\mu(6)$。用第 12 讲的语言，本原 n 次单位根是 $\Phi_n(x)$ 的全部根，所以 $\mu(n)$ 等于 Φ_n 次高项系数的相反数，例如 $\Phi_6(x)=x^2-x+1$。 ♦

易错提醒

- $\sum_{d\mid n}\mu(d)=0$ 只对 $n>1$ 成立，$n=1$ 时和是 1。用这个结论时要把 $n=1$ 单独处理。
- μ 是 U 的逆，不是 $\mathbf{1}$ 的逆（$\mathbf{1}$ 的逆是它自己）。也不要把卷积逆和倒数 $1/f(n)$ 混淆：U 的倒数仍是 U，卷积逆却是 μ。

想一想　ω、π、φ、λ 中哪些有卷积逆？

答　看 $f(1)$。$\omega(1)=0$，$\pi(1)=0$，没有逆。$\varphi(1)=1$，$\lambda(1)=1$，都有逆。由 $\varphi=\mu*r_1$ 得 $\varphi^{-1}=U*r_1^{-1}$；λ^{-1} 是 $|\mu|$，见练习 9。

本讲小结

- 算术函数是 $\mathbb{Z}_{>0}\to\mathbb{C}$ 的函数。积性：互素时 $f(mn)=f(m)f(n)$；完全积性：对一切 m,n 成立。积性函数有 $f(1)=1$，并由素数幂上的值完全决定。
- Dirichlet 卷积 $(f*g)(n)=\sum_{ab=n}f(a)g(b)$ 满足交换律、结合律，单位元是 $\mathbf{1}$。$f*U$ 就是对因数求和。
- 积性函数的卷积仍积性。由 $d=U*U$、$\sigma_k=r_k*U$ 得 $d(n)=\prod(e_i+1)$，$\sigma_k(n)=\prod\frac{p_i^{(e_i+1)k}-1}{p_i^k-1}$（$k\geq 1$；$k=0$ 时用 $\sigma_0=d$）；由 $\varphi*U=r_1$ 得 $\sum_{d\mid n}\varphi(d)=n$。
- 偶完全数恰好是 $2^{m-1}(2^m-1)$，2^m-1 为素数。奇完全数是否存在，至今未知。
- $\mu(n)$：含平方因子时为 0，否则为 $(-1)^{\omega(n)}$。$\mu*U=\mathbf{1}$，即 $\mu=U^{-1}$；f 可逆当且仅当 $f(1)\neq 0$。

练习

13.1 ★　对 $n=360$ 和 $n=1001$ 求 $d(n)$、$\sigma(n)$、$\varphi(n)$、$\mu(n)$、$\omega(n)$、$\Omega(n)$，并判断它们是亏数还是盈数。

13.2 ★　下列函数是完全积性、积性但非完全积性，还是非积性？(a) $(-1)^{n+1}$；(b) $2^{\Omega(n)}$；(c) $\mu(n)^2$；(d) $d(n)^2$；(e) $\sigma(n)-n$。

13.3 ★　求恰有 12 个正因数的最小正整数。

13.4 ★★　证明 n 的全部正因数之积等于 $n^{d(n)/2}$。（习题集 6 第 1 题）

13.5 ★★　设 k 为正整数。证明 $\sigma_k(n)$ 为奇数当且仅当 n 是平方数或平方数的 2 倍。（习题集 6 第 2 题）

13.6 ★★　证明：若 $\gcd(a,b)>1$，则 $d(ab)<d(a)d(b)$，$\sigma_k(ab)<\sigma_k(a)\sigma_k(b)$。这说明积性定义里的互素条件不能去掉。（习题集 6 第 3 题）

13.7 ★★　证明 Liouville 函数 $\lambda(n)=(-1)^{\Omega(n)}$ 完全积性，并且 n 为平方数时 $\sum_{d\mid n}\lambda(d)=1$，否则为 0。（习题集 6 第 5 题）

13.8 ★★　证明 $\sum_{d\mid n}d(d)^3=\left(\sum_{d\mid n}d(d)\right)^2$，并在 $n=12$ 处验证。（习题集 6 第 6 题）

13.9 ★★ 设 f 完全积性。证明 $f^{-1}=\mu f$（逐点乘积）。由此求 r_1^{-1}，并证明 $\lambda^{-1}=|\mu|$。再计算 $(d*\mu d)(4)$，说明「完全」不能去掉。

13.10 ★★★ 证明奇完全数至少有三个不同的素因数。（提示：$\sigma(n)/n$ 是积性的，且 $\sigma(p^a)/p^a<\frac{p}{p-1}$。）

练习答案与提示

13.1 $360=2^3\cdot3^2\cdot5$：$d=4\cdot3\cdot2=24$，$\sigma=15\cdot13\cdot6=1170$，$\varphi=4\cdot6\cdot4=96$，$\mu=0$，$\omega=3$，$\Omega=6$；$1170>720$，盈数。$1001=7\cdot11\cdot13$：$d=8$，$\sigma=8\cdot12\cdot14=1344$，$\varphi=6\cdot10\cdot12=720$，$\mu=-1$，$\omega=\Omega=3$；$1344<2002$，亏数。

13.2 (a) 积性：m,n 互素时不能都是偶数；若都奇，两边都是 1；若恰一个偶，两边都是 -1。但 $f(4)=-1\neq f(2)^2=1$，不是完全积性。(b) 完全积性，因为 Ω 完全加性。(c) 积性（μ 积性），但 $\mu(4)^2=0\neq\mu(2)^4=1$。(d) 积性，但 $d(4)^2=9\neq d(2)^4=16$。(e) 非积性：在 1 处取值 0；又 $\sigma(6)-6=6$，而 $(\sigma(2)-2)(\sigma(3)-3)=1$。

13.3 $d(n)=\prod(e_i+1)=12$，指数组（降序）只能是 (11)、$(5,1)$、$(3,2)$、$(2,1,1)$。大指数给小素数，候选为 $2^{11}=2048$，$2^5\cdot3=96$，$2^3\cdot3^2=72$，$2^2\cdot3\cdot5=60$。最小的是 60。

13.4 记积为 P。$d\mapsto n/d$ 是因数集合上的双射，所以 $P=\prod_{d\mid n}\frac{n}{d}$。两式相乘，$P^2=\prod_{d\mid n}d\cdot\frac{n}{d}=n^{d(n)}$，开方得 $P=n^{d(n)/2}$。例：12 的因数之积 $1\cdot2\cdot3\cdot4\cdot6\cdot12=1728=12^3$。

13.5 σ_k 积性，所以 $\sigma_k(n)$ 为奇当且仅当每个 $\sigma_k(p^e)$ 为奇。$p=2$ 时 $\sigma_k(2^e)=1+(\text{偶数})$，总是奇数。$p$ 为奇素数时，$\sigma_k(p^e)$ 是 $e+1$ 个奇数之和，为奇当且仅当 e 为偶。于是条件是：$n=2^am^2$，m 为奇数。a 偶时 n 是平方数，a 奇时 $n=2\cdot(2^{(a-1)/2}m)^2$。

13.6 映射 $(d_1,d_2)\mapsto d_1d_2$ 把 $\{d_1\mid a\}\times\{d_2\mid b\}$ 映到 ab 的因数。它是满射：对 $d\mid ab$，取 $d_1=\gcd(d,a)$，则 $\frac{d}{d_1}$ 与 $\frac{a}{d_1}$ 互素且整除 $\frac{a}{d_1}b$，所以 $\frac{d}{d_1}\mid b$。设素数 $p\mid\gcd(a,b)$，则 $(p,1)$ 与 $(1,p)$ 都映到 p，不是单射。所以 $d(ab)<d(a)d(b)$。同理 $\sigma_k(a)\sigma_k(b)=\sum(d_1d_2)^k$ 中每个 e^k（$e\mid ab$）至少出现一次，p^k 至少出现两次，故严格大于 $\sigma_k(ab)$。例：$d(24)=8<12$，$\sigma(24)=60<84$。

13.7 $\Omega(mn)=\Omega(m)+\Omega(n)$ 对一切 m,n 成立，所以 $\lambda(mn)=\lambda(m)\lambda(n)$。$F=\lambda*U$ 是积性的，$F(p^e)=1-1+1-\cdots+(-1)^e$，$e$ 偶时为 1，e 奇时为 0。因此 $F(n)=1$ 当且仅当所有指数都是偶数，即 n 为平方数；否则 $F(n)=0$。

13.8 两边都是积性函数（d^3*U 积性；$d*U$ 积性，其平方也积性），只需在 p^e 处比较。p^e 的因数 p^j 有 $d(p^j)=j+1$，于是左边 $=1^3+2^3+\cdots+(e+1)^3$，右边 $=(1+2+\cdots+(e+1))^2$。由经典恒等式 $\sum_{i=1}^m i^3=\left(\frac{m(m+1)}{2}\right)^2$，二者相等。$n=12$：$d$ 在 $1,2,3,4,6,12$ 处为 $1,2,2,3,4,6$，左边 $1+8+8+27+64+216=324$，右边 $18^2=324$。

13.9 f 完全积性，所以对每个 $d\mid n$ 有 $f(d)f(n/d)=f(n)$（不需要互素）。于是 $(f*\mu f)(n)=\sum_{d\mid n}f(d)\mu(\frac{n}{d})f(\frac{n}{d})=f(n)\sum_{d\mid n}\mu(\frac{n}{d})$，由定理 13.14 这在 $n=1$ 时为 1，否则为 0。故 $r_1^{-1}(n)=\mu(n)n$。$\mu(n)\neq0$ 时 n 无平方因子，$\lambda(n)=\mu(n)$，所以 $\mu\lambda=\mu^2=|\mu|$。d 不是完全积性：$(d*\mu d)(4)=\sum_{a\mid4}d(a)\mu(\frac{4}{a})d(\frac{4}{a})=1\cdot0\cdot3+2\cdot(-1)\cdot2+3\cdot1\cdot1=-1\neq0$。

13.10 设 $h(n)=\sigma(n)/n$，它是两个积性函数之商，积性。$h(p^a)=1+\frac{1}{p}+\cdots+\frac{1}{p^a}<\frac{p}{p-1}$，且 $\frac{p}{p-1}$ 随 p 增大而减小。奇数 n 的素因数都 ≥3。若 $n=p^a$，$h(n)<\frac{3}{2}$；若 $n=p^aq^b$，$3\leq p<q$，则 $q\geq5$，$h(n)<\frac{3}{2}\cdot\frac{5}{4}=\frac{15}{8}$。两种情形都小于 2，$n=1$ 时 $h=1$，所以奇完全数至少有三个不同的素因数。已知的结果远强于此：Nielsen（2015）证明至少要有 10 个。

第 14 讲

Möbius 反演与 zeta 函数

Mobius Inversion and the Zeta Function

对应材料：MIT 18.781 Lecture 14 讲义（定理 54–56）；习题集 6 第 7(a)、8(a)(b)、10(a) 题。

本讲目标

(1) 熟练使用 Möbius 反演：已知 $F(n)=\sum_{d\mid n} f(d)$ 时，能恢复 $f(n)$。

(2) 会把 φ、平方自由数等具体问题写成带 μ 的因数和，并逐项核对。

(3) 理解 Dirichlet 级数 $Z(f,s)$ 把 Dirichlet 卷积变成乘法的原因。

(4) 会从积性函数的局部数据推导 Euler 乘积，写出 $\zeta(s)$ 与 μ 的乘法逆关系。

(5) 能用 $\zeta(2)$ 与 Möbius 反演计算平方自由数密度和两个随机正整数互素的概率。

14.1 如何撤销「对所有因数求和」

第 13 讲中出现了函数

$$F(n)=\sum_{d\mid n} f(d)=(f*U)(n)$$

它把 f 的值混合在一起。例如若 $f=\varphi$，则 $F(n)=n$；若 $f=\mathbf{1}$，则 $F(n)=U(n)=1$。已知 F 时，能否把 f 拆回来？先对小数做一次计算：

$$F(1)=f(1),\qquad F(2)=f(1)+f(2),\qquad F(6)=f(1)+f(2)+f(3)+f(6)$$

每一步都可以减去已经知道的较小下标。Möbius 函数把这个递推过程写成一个一次完成的公式。

定理 14.1（Möbius 反演公式） 设 f,F 是算术函数。若对每个正整数 n 有

$$F(n)=\sum_{d\mid n} f(d),$$

则

$$f(n)=\sum_{d\mid n}\mu(d)F\left(\frac{n}{d}\right)=\sum_{d\mid n}\mu\left(\frac{n}{d}\right)F(d)$$

思路　把第一式写成 $F=f*U$，然后在两边卷积 U 的逆 μ。关键只在于结合律和 $\mu*U=\mathbf{1}$；最后把卷积展开成因数和。

证明　由第 13 讲的恒等式 $\mu * U = \mathbf{1}$，以及 Dirichlet 卷积的结合律，

$$F * \mu = (f * U) * \mu = f * (U * \mu) = f * \mathbf{1} = f$$

在 n 处取值便得

$$(F * \mu)(n) = \sum_{d \mid n} F(d)\mu\left(\frac{n}{d}\right) = \sum_{d \mid n} \mu(d) F\left(\frac{n}{d}\right)$$

第二个等号只把因数 d 换成 n/d。这两个表达式都等于 $f(n)$。 ∎

当 $n = 6$ 时，反演如何把混合项消掉？图 14.1 把 $F(6) - F(3) - F(2) + F(1)$ 逐项展开。每一列对应同一个 $f(d)$，沿列相加后，只有 $f(6)$ 留下。

	$f(1)$	$f(2)$	$f(3)$	$f(6)$
$F(6)$	$+1$	$+1$	$+1$	$+1$
$-F(3)$	-1		-1	
$-F(2)$	-1	-1		
$+F(1)$	$+1$			
总系数	0	0	0	1

图 14.1　反演在每个较小因数的列中相消，留下 $f(6)$ 的系数为 1。空白位置表示该项未出现，系数为 0。

例如 $f(1)$ 的系数是 $1 - 1 - 1 + 1 = 0$，$f(2)$ 的系数是 $1 - 1 = 0$。一般证明中的 $\mu * U = \mathbf{1}$ 正是在每个因数处实现同样的抵消，因而反演可以恢复任意算术函数，而不只适用于 φ。

14.2　素数分解给出 Euler 乘积

*准备与阅读范围。*以下使用级数的收敛、绝对收敛及绝对收敛级数的重排和乘法。尚未学过级数时，可以先完成上一节的有限因数和与反演，补学后再读本节及后面的密度论证。主要计算可先取实数 s；复变量情形约定 $n^{-s} = e^{-s \log n}$，这里 $\log n$ 是正整数 n 的实对数。

普通整数的乘法由素数幂唯一决定。Dirichlet 级数中，素数 p 对应一个局部因子，里面收集 $1, p, p^2, \ldots$ 的贡献。

定义 14.2（Dirichlet 级数）　算术函数 f 的 **Dirichlet 级数**（Dirichlet series）定义为

$$Z(f, s) = \sum_{n \geq 1} \frac{f(n)}{n^s}.$$

其中 s 是复变量。若 $\sum_{n \geq 1} |f(n)| n^{-\sigma}$ 对某个实数 σ 收敛，则在 $\mathrm{Re}(s) \geq \sigma$ 上绝对收敛。

先看收敛区域如何由系数增长决定。对常值函数 $U(n) = 1$，级数是 $\sum n^{-s}$，当 $\mathrm{Re}(s) > 1$ 时由 p 级数判别法绝对收敛；在 $s = 1$ 时它变成调和级数而发散。对 $r_1(n) = n$，则

$$Z(r_1, s) = \sum_{n \geq 1} n^{1-s} = \zeta(s - 1), \qquad \mathrm{Re}(s) > 2.$$

这两个例子说明：把 $f(n)$ 的增长多乘一个 n，绝对收敛边界就向右移动 1。后面的乘积与换序都在绝对收敛区域内进行；边界上的条件收敛不能直接套用同一套重排理由。

更一般地，若存在常数 $C, m > 0$ 使 $|f(n)| \le Cn^m$，则

$$\left|\frac{f(n)}{n^s}\right| \le \frac{C}{n^{\mathrm{Re}(s)-m}}.$$

比较 p 级数可知 $\mathrm{Re}(s) > m+1$ 是一个绝对收敛区域。它是由增长率给出的充分条件，不一定是最好的边界；例如 $\mu(n)$ 有界，$\varphi(n) \le n$，分别可在 $\mathrm{Re}(s) > 1$、$\mathrm{Re}(s) > 2$ 使用这个估计。

定理 14.3（积性函数的 Euler 乘积）　设 f 是积性函数，$f(1) = 1$，且在 $\mathrm{Re}(s) = \sigma$ 上

$$\sum_{n\ge 1} |f(n)| n^{-\sigma} < \infty.$$

则

$$Z(f,s) = \prod_{p \text{ prime}} \left(\sum_{e\ge 0} \frac{f(p^e)}{p^{es}} \right), \qquad \mathrm{Re}(s) = \sigma$$

条件中的绝对收敛有实际作用：它不只保证左边的级数有意义，还保证把素数幂的局部级数相乘、再按整数重新分组时不改变和。对 U 来说，结论在 $\mathrm{Re}(s) > 1$ 使用；例如 $s = 2$ 时每个局部因子都是收敛几何级数，而 $s = 1$ 时调和级数已发散，不能从有限乘积直接跳到一个有限极限。

思路　先只取有限多个素数。展开有限乘积时，每一项选择各素数的一个幂；唯一分解定理保证这些选择与所有素因子来自所选素数集的整数一一对应。再用绝对收敛让素数集合增大到全部素数。

证明　对任意有限素数集合 P，乘开局部级数得到

$$\prod_{p\in P} \sum_{e\ge 0} \frac{f(p^e)}{p^{es}} = \sum_{\substack{n\ge 1 \\ \text{所有素因数属于} P}} \frac{f(n)}{n^s}$$

这里每个 n 对应唯一分解 $n = \prod_{p\in P} p^{e_p}$，且积性给出 $f(n) = \prod_{p\in P} f(p^{e_p})$。令 P 包含不超过 y 的所有素数，$y \to \infty$ 时右边逐项覆盖全部正整数。由于绝对收敛，遗漏项的绝对值和趋于 0，故极限就是 $Z(f,s)$。 ∎

对常值函数 $U(n) = 1$，它的 Dirichlet 级数是

$$\zeta(s) = Z(U,s) = \sum_{n\ge 1} \frac{1}{n^s} \qquad (\mathrm{Re}(s) > 1)$$

这个函数称为 Riemann **zeta 函数**（zeta function）。

取 $f = U$，每个局部因子都是几何级数：

$$\zeta(s) = \prod_p \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots\right) = \prod_p \frac{1}{1-p^{-s}} \qquad (\mathrm{Re}(s) > 1)$$

■ **例 14.1　只保留素数 2、3 时，乘积包含哪些整数？**　在 $s=2$ 时展开 $(1-2^{-2})^{-1}(1-3^{-2})^{-1}$，说明它为什么小于 $\zeta(2)$。

解　先分别展开两个几何级数，再相乘。一个因子选出 2^{-2e}，另一个选出 3^{-2f}，这一对选择就贡献 $1/(2^e3^f)^2$。唯一分解保证不同的 (e,f) 不会得到同一个整数。于是

$$\sum_{e,f\ge 0}\frac{1}{(2^e3^f)^2}=\frac{1}{1-1/4}\,\frac{1}{1-1/9}=\frac{3}{2}.$$

这些项没有包含 $1/5^2$、$1/7^2$ 等正数，所以总和严格小于 $\zeta(2)$。再加入素数 5 的局部因子，乘积增为 $(3/2)(25/24)=25/16$，仍未包含素因子为 7 的项。 ♦

图 14.2 将两种指数选择排成网格。格子中的 n 表示级数项 $1/n^2$；橙格说明 $n=12$ 只能由 $e=2,f=1$ 这一种选择产生。

	2^0	2^1	2^2	2^3
3^0	1	2	4	8
3^1	3	6	12	24
3^2	9	18	36	72
3^3	27	54	108	216

每个格子对应 $n=2^e3^f$。

继续向右、向下延伸，覆盖所有只含素因子 $2,3$ 的正整数。

$5,7,10$ 都不在这个网格中。

图 14.2　有限素数乘积已包含无限多个整数项，但只包含素因子来自所选集合的整数；图中显示 $0\le e,f\le 3$ 的部分。

网格区分了两种截断：只保留有限个素数，仍可使用每个素数的任意非负次幂；只画有限个格子，则又限制了指数。Euler 乘积的证明先做前一种截断，最后让所选素数覆盖全部素数。

取 $f=\mu$，由于 $\mu(p)=-1$ 且 $\mu(p^e)=0$（$e\ge 2$），局部因子变成 $1-p^{-s}$。所以

$$Z(\mu,s)=\prod_p(1-p^{-s})=\frac{1}{\zeta(s)}$$

这正是函数恒等式 $\mu * U=\mathbf{1}$ 在 Dirichlet 级数中的像。

局部因子只依赖 f 在素数幂上的值。令 $x=p^{-s}$，下表列出几个已学函数的局部级数：

函数 f	素数幂上的值	对应的局部因子
U	$U(p^e)=1$	$\displaystyle\sum_{e\ge 0}x^e=(1-x)^{-1}$
μ	$1,-1,0,0,\dots$	$1-x$
d	$d(p^e)=e+1$	$\displaystyle\sum_{e\ge 0}(e+1)x^e=(1-x)^{-2}$
φ	$\varphi(p^e)=p^{e-1}(p-1),\ e\ge 1$	$\dfrac{1-p^{-s}}{1-p^{1-s}}$

表 14.1　四个乘法函数的素数幂值及其局部 Dirichlet 因子。

表 14.1 的第三行也可由 $d = U * U$ 得到：两个几何级数相乘后，x^e 有 $e+1$ 种幂次拆分。因此

$$Z(d,s) = \prod_p (1-p^{-s})^{-2} = \zeta(s)^2$$

最后一行从 $\varphi(1)=1$ 开始。记 $y = p^{1-s}$，则

$$1 + \sum_{e\ge 1} \frac{p^{e-1}(p-1)}{p^{es}} = 1 + \frac{1-p^{-1}}{1-y}y = \frac{1-p^{-s}}{1-p^{1-s}}$$

当 $\mathrm{Re}(s) > 2$ 时对所有素数相乘，得到

$$Z(\varphi,s) = \prod_p \frac{1-p^{-s}}{1-p^{1-s}} = \frac{\zeta(s-1)}{\zeta(s)}$$

这条公式把本讲的反演结果 $\varphi = \mu * r_1$（$r_1(n)=n$）变成级数恒等式：$Z(\varphi,s) = Z(\mu,s)Z(r_1,s) = \zeta(s-1)/\zeta(s)$。

若 f 是**完全积性函数**（completely multiplicative function），则 $f(p^e) = f(p)^e$，每个局部因子是几何级数：

$$\sum_{e\ge 0} \frac{f(p^e)}{p^{es}} = \sum_{e\ge 0} \left(\frac{f(p)}{p^s}\right)^e = \frac{1}{1-f(p)p^{-s}}$$

因此完全积性函数的 Euler 乘积化为

$$Z(f,s) = \prod_p \frac{1}{1-f(p)p^{-s}}$$

例如 Liouville 函数 $\lambda(n) = (-1)^{\Omega(n)}$ 完全积性，因为 $\Omega(mn) = \Omega(m)+\Omega(n)$。它在素数幂上满足 $\lambda(p^e) = (-1)^e$，所以

$$Z(\lambda,s) = \prod_p \frac{1}{1+p^{-s}}$$

这里需要区分 $\Omega(n)$（素因数按重数计数）与 $\omega(n)$（不同素因数个数）；第 13 讲的 Möbius 函数用的是后者。

14.3　反演的算术应用

■ **例 14.2**　**从因数和恢复 φ**　已知 $\sum_{d\mid n}\varphi(d) = n$，用 Möbius 反演求 $\varphi(12)$，并写出一般公式。

解　这里 $F(n)=n$，所以

$$\varphi(12) = \sum_{d\mid 12} \mu(d)\frac{12}{d} = 12 - 6 - 4 + 2 = 4$$

只有 $d = 1,2,3,6$ 的 $\mu(d)$ 非零；含平方因子的 $d = 4, 12$ 自动消失。一般地，若

$$n = p_1^{e_1}\cdots p_r^{e_r},$$

则非零项对应于从 $p_1,\dots,p_r$ 中选取一个子集，因而

$$\varphi(n) = n\prod_{p\mid n}\left(1-\frac{1}{p}\right)$$

这说明反演公式不仅恢复单个数值，还会自动产生积性函数的乘法公式。 ♦

为什么乘积公式只看不同的素数，而不看指数 e_i？在反演和式中，$\mu(d)$ 只允许 d 取平方自由因数；因此对 $n=p_1^{e_1}\cdots p_r^{e_r}$，每个素数只有「选入一次」或「不选入」两种状态。选入的素数集合 $S\subseteq\{1,\dots,r\}$ 对应

$$d=\prod_{i\in S}p_i,\qquad \mu(d)=(-1)^{|S|}.$$

于是反演和是一个包含 2^r 项的容斥展开。把它重新分组，就得到

$$\sum_{S}(-1)^{|S|}\frac{n}{\prod_{i\in S}p_i}=n\prod_{i=1}^{r}\left(1-\frac{1}{p_i}\right).$$

指数只影响前面的 n，不影响每个素数是否需要排除。

一个正整数若不被任何素数的平方整除，就称为平方自由。例如 $30=2\cdot3\cdot5$ 是平方自由，而 $12=2^2\cdot3$ 不是。如何把这个条件变成一个可以求和的公式？答案仍然是 Möbius 函数。

定义 14.4（平方自由指标）　定义

$$\mathrm{sf}(n)=\begin{cases}1, & n\text{ 平方自由},\\ 0, & n\text{ 含有某个素数的平方因子}\end{cases}$$

命题 14.5（平方自由指标的反演公式）　对每个正整数 n，

$$\mathrm{sf}(n)=\sum_{d^2\mid n}\mu(d)$$

证明　设 $n=\prod p^{e_p}$。因数 d 满足 $d^2\mid n$，所以在每个素数 p 上只能取指数 $0\le v_p(d)\le\lfloor e_p/2\rfloor$。由于 $\mu(d)$ 在 d 含平方因子时为 0，每个素数实际上只有选择「不取」或「取一次」两种贡献。因此

$$\sum_{d^2\mid n}\mu(d)=\prod_{p^2\mid n}(1+\mu(p))=\prod_{p^2\mid n}(1-1)$$

若没有 $p^2\mid n$，空乘积为 1；若至少有一个这样的素数，乘积为 0，正好得到 $\mathrm{sf}(n)$。 ■

■ **例 14.3　数出 30 以内的平方自由数**　用上式计算 $Q(30)=\#\{n\le30:n\text{ 平方自由}\}$。

解　交换有限和：

$$Q(30)=\sum_{n\le30}\sum_{d^2\mid n}\mu(d)=\sum_{d\le\sqrt{30}}\mu(d)\left\lfloor\frac{30}{d^2}\right\rfloor$$

只有 $d=1,2,3,5$ 的 $\mu(d)$ 非零，故

$$Q(30)=30-7-3-1=19$$

直接列出 $1,2,3,5,6,7,10,11,13,14,15,17,19,21,22,23,26,29,30$ 也得到 19 个。这一列举是公式的数值核对，不是证明的替代。 ♦

注 14.1　例如 $\varphi(30)=30(1-\frac12)(1-\frac13)(1-\frac15)=8$。若 $n=p^e$，乘积中只有一个素数因子，于是 $\varphi(p^e)=p^e-p^{e-1}$，与第 13 讲逐项数互素剩余类的结果相同。

14.4　Möbius 函数的抵消与应用

图 14.3 把 Möbius 反演的一种直观来源画了出来：编号为 m 的门会被它的每个正因数翻转一次。

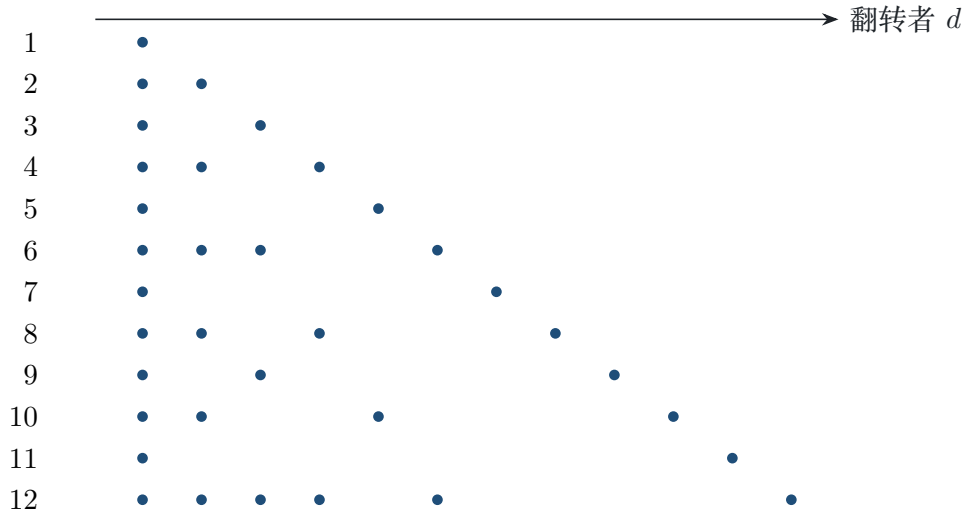


图 14.3　门 n 被翻转 $d(n)$ 次，最终开着当且仅当 n 是平方数。

门 n 被翻转的次数是 $d(n)$。若 $n=p_1^{e_1}\cdots p_r^{e_r}$，则 $d(n)=\prod_i(e_i+1)$ 为奇数当且仅当每个 e_i 都是偶数，也就是 n 为平方数。因此 $1,4,9,16,\dots$ 号门最后开着。这个例子把「因数的个数」连到了「是否为平方」；下一步会看到 μ 如何对任意因数和做符号消去。

想一想　反演公式里保持 $F(n/d)$ 不变，只把权重 $\mu(d)$ 换成 $\mu(n/d)$，结果还相同吗？

答　一般不同。取 $n=2$、$F(n)=n$，原式为 $F(2)-F(1)=1$；只换权重后为 $-F(2)+F(1)=-1$。正确的换元要同时改变两个因子：令 $e=n/d$，则 $\sum_{d\mid n}\mu(d)F(n/d)=\sum_{e\mid n}\mu(n/e)F(e)$。因数互换保持的是整个乘积，不能只换其中一项。

命题 14.6（反演公式的逆命题）　若对每个正整数 n 有

$$f(n)=\sum_{d\mid n}\mu(d)F\left(\frac{n}{d}\right),$$

则 $F(n)=\sum_{d\mid n}f(d)$。

证明　假设 $f=\mu*F$。对两边卷积 U，由交换律、结合律和 $\mu*U=\mathbf{1}$，

$$f*U=(\mu*F)*U=F*(\mu*U)=F*\mathbf{1}=F$$

展开左边即为 $\sum_{d\mid n}f(d)$。 ∎

对 $n=6$，因数 $1,2,3,6$ 的 Möbius 值是 $1,-1,-1,1$，所以它们相加为 0。一般地，若 $n>1$，其不同素因数给出一个非空集合；将因数中是否选入每个素数配对，正负项逐一抵消。这正是 $\mu * U = \mathbf{1}$ 的素因数解释。

■ **例 14.4**　**$\mu*\mu$ 在素数幂上的值**　计算 $(\mu*\mu)(p^e)$，再求 $(\mu*\mu)(12)$。

解　对素数幂，卷积和只有 $e+1$ 项：

$$(\mu*\mu)(p^e)=\sum_{i=0}^{e}\mu(p^i)\mu(p^{e-i})$$

当 $e=0$ 时结果为 1；当 $e=1$ 时两项各为 -1，和为 -2；当 $e=2$ 时只有中间项非零，结果为 1。若 $e\ge 3$，每一项至少有一个指数不小于 2，该项就为 0。因此局部值依次是 $1,-2,1,0,0,\dots$。

函数 $\mu*\mu$ 积性。因为 $12=2^2\cdot 3$，所以

$$(\mu*\mu)(12)=(\mu*\mu)(2^2)(\mu*\mu)(3)=1\cdot(-2)=-2$$

这题说明：卷积的乘法结构允许我们先处理素数幂，再相乘得到一般整数的值。 ♦

若 $n=\prod p^{e_p}$，上面的素数幂计算给出

$$(\mu*\mu)(n)=\prod_{p\mid n}\begin{cases}1, & e_p=2,\\ -2, & e_p=1,\\ 0, & e_p\ge 3\end{cases}$$

所以只要 n 含有某个素数的三次方因子，$(\mu*\mu)(n)=0$。不被任何素数的三次方整除的正整数称为**立方自由数** (cube-free integer)；上式说明 $(\mu*\mu)(n)\neq 0$ 当且仅当 n 立方自由。例如 $12=2^2\cdot 3$ 立方自由，卷积值为 -2；$24=2^3\cdot 3$ 不立方自由，卷积值为 0。若 n 平方自由且有 r 个素因数，则 $(\mu*\mu)(n)=(-2)^r$。

易错提醒　$\mu(n)=0$ 只说明 n 含有平方因子，不能推出 $(\mu*\mu)(n)=0$。例如 $n=4$ 时 $\mu(4)=0$，但 $(\mu*\mu)(4)=1$。卷积会把多个因数位置的值相加，必须先算卷积，不能把函数值直接相乘。

■ **例 14.5**　**用开关门找平方数**　有 12 扇关着的门。第 k 个人翻转所有编号为 k 的倍数的门。最后哪些门开着？

解　门 n 被翻转 $d(n)$ 次，所以它最终开着当且仅当 $d(n)$ 为奇数。若 $n=\prod p_i^{e_i}$，则

$$d(n)=\prod_i(e_i+1)$$

为奇数当且仅当所有 e_i 都为偶数，即 n 是完全平方数。$1\le n\le 12$ 的平方数是 $1,4,9$，所以只有这三扇门开着。逐一核对：$d(1)=1$、$d(4)=3$、$d(9)=3$ 为奇数，而 $d(12)=6$ 为偶数。 ♦

想一想　部分和 $M(x)=\sum_{n\le x}\mu(n)$ 会一直保持在 $\sqrt{x}$ 的量级以内吗?

答　Mertens 猜想断言 $|M(x)|<\sqrt{x}$，但该猜想已被反例推翻。较弱的界 $M(x)=O_\varepsilon(x^{1/2+\varepsilon})$ 对每个 $\varepsilon>0$ 成立，当且仅当 Riemann 猜想成立。这里的结论说明 μ 的正负抵消和素数分布相连；本讲不证明这个深层等价。

先拆素数幂　要计算积性函数，先求它在 p^e 上的值；要计算卷积，先按 p^e 展开因数和，再利用乘法性组合。

14.5　Dirichlet 级数的乘法

上一节定义了 $Z(f,s)$，并在积性函数的情形得到 Euler 乘积。现在直接证明它与 Dirichlet 卷积相容。普通生成函数按加法下标收集项；这里的下标关系是 $ab=n$。

定理 14.7（卷积的 Dirichlet 乘法定理）　设 $s\in\mathbb{C}$，且 $Z(f,s)$、$Z(g,s)$ 都绝对收敛，则

$$Z(f*g,s)=Z(f,s)Z(g,s)$$

思路　把右边展开成有序对 (a,b) 的双重和，再按乘积 $n=ab$ 分组。固定的 n 只有有限多个因数对，所以 n^{-s} 的系数正好是 $(f*g)(n)$。

证明　在绝对收敛区域内可以交换求和次序:

$$\begin{aligned}Z(f,s)Z(g,s)&=\sum_{a\ge1}\sum_{b\ge1}\frac{f(a)g(b)}{(ab)^s}\\&=\sum_{n\ge1}\frac{1}{n^s}\sum_{ab=n}f(a)g(b)=\sum_{n\ge1}\frac{(f*g)(n)}{n^s}.\end{aligned}$$

右端就是 $Z(f*g,s)$。 ■

■ **例 14.6　卷积的系数核对**　设 $f=U$，$g=U$。写出 $Z(U,s)^2$ 中 $n=12$ 的系数。

解　乘积 $ab=12$ 的有序因数对是

$$(1,12),(2,6),(3,4),(4,3),(6,2),(12,1)$$

一共有 $6=d(12)$ 对。因此 n^{-s} 的系数是 6，这正是

$$(U*U)(12)=\sum_{d\mid12}1=d(12)=6$$

这也解释了 $Z(d,s)=\zeta(s)^2$。 ♦

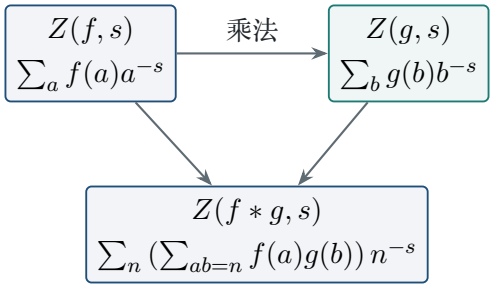


图 14.4　Dirichlet 级数的乘法按 $ab=n$ 收集项，得到卷积后的系数。

图 14.4 中的箭头表示同一件事的两种读法：先把两个级数相乘，再按 n 收集；或者先做 Dirichlet 卷积，再取级数。乘法下标 $ab=n$ 是本讲与普通幂级数的关键区别。

14.6　两个 zeta 值

Euler 乘积揭示了 $\zeta(s)$ 与素数的关系。另一方面，$\zeta(2)$ 和 $\zeta(4)$ 还有漂亮的闭式：

$$\zeta(2)=\sum_{n\ge 1}\frac{1}{n^2}=\frac{\pi^2}{6},\qquad \zeta(4)=\sum_{n\ge 1}\frac{1}{n^4}=\frac{\pi^4}{90}$$

先看第一项的数值近似：$1+1/4+1/9=49/36\approx 1.3611$，加入更多正项会从下方趋近 $\pi^2/6\approx 1.6449$。

命题 14.8（Basel 和与四次方和）

$$\zeta(2)=\frac{\pi^2}{6},\qquad \zeta(4)=\frac{\pi^4}{90}$$

证明　正弦乘积公式是经典分析结论；本讲只用它比较幂级数系数，不展开其证明。使用经典的正弦乘积

$$\frac{\sin x}{x}=\prod_{n\ge 1}\left(1-\frac{x^2}{n^2\pi^2}\right)$$

左边的 Taylor 展开是 $1-x^2/6+x^4/120-\cdots$。右边的 x^2 系数为 $-\pi^{-2}\sum_{n\ge 1}n^{-2}$，比较得到 $\zeta(2)=\pi^2/6$。

右边的 x^4 系数为

$$\frac{1}{\pi^4}\sum_{i<j}\frac{1}{i^2j^2}$$

因此

$$\sum_{i<j}\frac{1}{i^2j^2}=\frac{\pi^4}{120}$$

又因为

$$\zeta(2)^2=\zeta(4)+2\sum_{i<j}\frac{1}{i^2j^2},$$

代入 $\zeta(2)=\pi^2/6$ 后得

$$\zeta(4)=\frac{\pi^4}{36}-\frac{\pi^4}{60}=\frac{\pi^4}{90}$$

■

■ **例 14.7**　**截断和的误差方向**　比较 $\sum_{n=1}^{3} n^{-2}$ 与 $\zeta(2)$，并给出一个简单的尾项上界。

解　部分和为

$$1+\frac{1}{4}+\frac{1}{9}=\frac{49}{36}<\frac{\pi^2}{6}$$

因为 $x\mapsto x^{-2}$ 单调递减，

$$0<\sum_{n\ge 4}\frac{1}{n^2}<\int_{3}^{\infty}\frac{dx}{x^2}=\frac{1}{3}$$

所以截断值与 $\zeta(2)$ 的差小于 $1/3$。这说明正项级数的部分和从下方逼近极限。 ♦

14.7　从因数和得到密度与概率

记 $Q(X)$ 为不超过 X 的平方自由数个数。由命题 14.5，

$$Q(X)=\sum_{n\le X}\mathrm{sf}(n)=\sum_{d\le\sqrt{X}}\mu(d)\Big\lfloor\frac{X}{d^2}\Big\rfloor$$

对每个 $d\le\sqrt{X}$，地板函数与 X/d^2 的差小于 1，所以这些误差合计至多 $\sqrt{X}$。于是

$$Q(X)=X\sum_{d\le\sqrt{X}}\frac{\mu(d)}{d^2}+O(\sqrt{X})$$

令 $m=\lfloor\sqrt{X}\rfloor$。补回被截掉的尾和时，积分比较给出

$$X\sum_{d>\sqrt{X}}\frac{1}{d^2}\le X\int_{m}^{\infty}\frac{dt}{t^2}=\frac{X}{m}=O(\sqrt{X}).$$

所以

$$Q(X)=X\sum_{d\ge 1}\frac{\mu(d)}{d^2}+O(\sqrt{X})=\frac{6}{\pi^2}X+O(\sqrt{X})$$

自然密度是计数函数除以 X 后的极限；因此平方自由数的自然密度是 $6/\pi^2\approx 0.6079$。也可以把这个常数写成局部因子的乘积 $\prod_p(1-p^{-2})$。逐素数相乘很像「每个 p^2 独立排除」的概率模型，但严谨依据是上面的有限求和和尾项估计，不需要假设无限多个事件相互独立。这里的主项来自

$$\sum_{d\ge 1}\frac{\mu(d)}{d^2}=Z(\mu,2)=\frac{1}{\zeta(2)}$$

■ **例 14.8**　**互素数对计数**　在 $1\le a,b\le 6$ 的 36 个有序数对中，有多少对满足 $\gcd(a,b)=1$？

解　按最大值分类：最大值为 1 时有 $(1,1)$；最大值为 $m\ge 2$ 时，边界上的互素点有 $2\varphi(m)$ 个。因此总数为

$$1+2\sum_{m=2}^{6}\varphi(m)=2\sum_{m=1}^{6}\varphi(m)-1=2(1+1+2+2+4+2)-1=23$$

比例是 $23/36\approx 0.639$，已经接近极限 $6/\pi^2\approx 0.608$，但有限范围仍有明显波动。 ♦

定理 14.9（两个随机正整数互素的概率）　令 C_N 为正方形 $\{1,\dots,N\}^2$ 中满足 $\gcd(a,b)=1$ 的有序对数。则

$$\lim_{N\to\infty}\frac{C_N}{N^2}=\frac{6}{\pi^2}$$

思路　用 $\sum_{d\mid\gcd(a,b)}\mu(d)$ 表示「互素」这个指标，交换有限求和后得到 $\sum\mu(d)\lfloor N/d\rfloor^2$。除以 N^2，逐项极限是 $\mu(d)/d^2$，尾项由 $\sum d^{-2}$ 控制。

证明

$$C_N=\sum_{a,b\le N}\sum_{d\mid\gcd(a,b)}\mu(d)=\sum_{d\le N}\mu(d)\Big\lfloor\frac{N}{d}\Big\rfloor^2$$

给定 d，有 $\lfloor N/d\rfloor^2/N^2\to d^{-2}$。由于

$$\left|\mu(d)\frac{\lfloor N/d\rfloor^2}{N^2}\right|\le\frac{1}{d^2}$$

且 $\sum d^{-2}$ 收敛，可以先固定有限个 d 再控制尾和，得到

$$\lim_{N\to\infty}\frac{C_N}{N^2}=\sum_{d\ge1}\frac{\mu(d)}{d^2}=\frac{1}{\zeta(2)}=\frac{6}{\pi^2}$$

■

固定正整数 k。满足 $\gcd(a,b)=k$ 的数对恰好写成 $(a,b)=(ka',kb')$，其中 $\gcd(a',b')=1$。因此

$$\lim_{N\to\infty}\Pr\big(\gcd(a,b)=k\big)=\frac{6}{\pi^2k^2}=\frac{1}{\zeta(2)k^2}$$

这里还需说明：不能只凭每个固定 k 的极限，就交换极限与无限求和。若 $\gcd(a,b)>K$，则存在某个 $d>K$ 同时整除 a,b，故并集估计给出

$$\Pr\big(\gcd(a,b)>K\big)\le\sum_{d=K+1}^{N}\frac{\lfloor N/d\rfloor^2}{N^2}\le\sum_{d>K}\frac{1}{d^2}.$$

右端随 $K\to\infty$ 趋于 0，且与 N 无关。因此可以先对有限个 k 取极限，再用这个一致尾界扩大范围，得到全部极限概率之和 $\sum_{k\ge1}1/(\zeta(2)k^2)=1$。

■ **例 14.9　两组最大公因数相等**　取四个相互独立、在 $1,\dots,N$ 上均匀分布的正整数 a,b,c,d。当 $N\to\infty$ 时，$\gcd(a,b)=\gcd(c,d)$ 的概率是多少？

解　记 $G_1=\gcd(a,b)$、$G_2=\gcd(c,d)$。两组数独立。先限于 $k\le K$，各项的概率乘积收敛到 $1/(\zeta(2)^2k^4)$。超过 K 的相等事件概率不超过 $\Pr(G_1>K)$，而上面的统一尾界令它在 $K\to\infty$ 时趋于 0。所以可将有限和的极限放大为

$$\lim_{N\to\infty}\Pr(G_1=G_2)=\frac{1}{\zeta(2)^2}\sum_{k\ge1}\frac{1}{k^4}=\frac{\zeta(4)}{\zeta(2)^2}=\frac{\pi^4/90}{(\pi^2/6)^2}=\frac{2}{5}$$

也就是 40%。这里先按 gcd 的可能值分情况，再把互斥情形的概率相加。 ♦

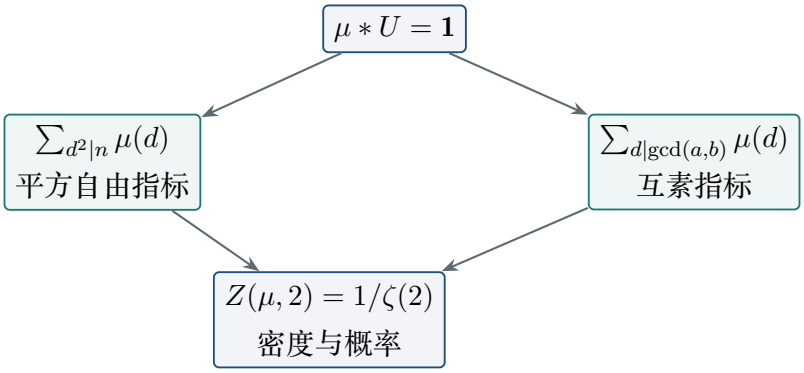


图 14.5　Möbius 函数把平方自由与互素条件写成指标和，取平均后出现 $1/\zeta(2)$。

图 14.5 总结了本讲的计数路线：μ 的抵消先把算术条件变成有限和；把这些和对整数取平均，就得到 zeta 级数。两种条件都导向同一个常数 $1/\zeta(2)=6/\pi^2$。

历史与背景　Möbius 函数以德国数学家 August Ferdinand Möbius 命名；它在 1832 年的工作中出现。Riemann 在 1859 年研究素数分布时引入了现在的 zeta 函数视角。Euler 乘积则把「所有整数的乘法」和「素数的局部因子」直接联系起来。

本讲小结

- 若 $F(n)=\sum_{d\mid n}f(d)$，则 **Möbius 反演**给出 $f(n)=\sum_{d\mid n}\mu(d)F(n/d)$。
- $\varphi(n)=\sum_{d\mid n}\mu(d)n/d=n\prod_{p\mid n}(1-1/p)$。
- 积性函数满足 Euler 乘积 $Z(f,s)=\prod_p\sum_{e\ge0}f(p^e)p^{-es}$；特别地 $\zeta(s)=\prod_p(1-p^{-s})^{-1}$。
- Dirichlet 卷积对应级数乘法：$Z(f*g,s)=Z(f,s)Z(g,s)$，所以 $Z(\mu,s)=1/\zeta(s)$。
- 平方自由指标为 $\sum_{d^2\mid n}\mu(d)$，平方自由数的密度是 $\sum\mu(d)/d^2=1/\zeta(2)=6/\pi^2$。
- 两个随机正整数互素的概率也是 $6/\pi^2$；其 gcd 等于 k 的概率为 $6/(\pi^2k^2)$，两组 gcd 相等的概率为 $\zeta(4)/\zeta(2)^2=2/5$。

练习

14.1 ★　计算 $\mu(1),\mu(2),\mu(6),\mu(12),\mu(30)$，并说明哪些数平方自由。

14.2 ★　已知 $F(n)=n^2=\sum_{d\mid n}f(d)$，用反演求 $f(12)$。

14.3 ★　证明：若 $f(n)=\sum_{d\mid n}\mu(d)F(n/d)$，则 $F(n)=\sum_{d\mid n}f(d)$。

14.4 ★★　用反演或乘积公式求 $\varphi(36)$，并列出 $\varphi(36)$ 的非零 Möbius 项。

14.5 ★★　求 $(\mu*\mu)(60)$，其中 $60=2^2\cdot3\cdot5$。

14.6 ★★　在 $Z(f,s)$ 与 $Z(g,s)$ 都绝对收敛的区域，证明 $Z(f*g,s)=Z(f,s)Z(g,s)$，并写出 $Z(d,s)$ 与 $\zeta(s)$ 的关系。

14.7 ★★　(a) 设 $f(1)\neq0$，构造卷积逆 g，并说明 $f(1)=0$ 时为何不可逆。(b) 设 $f(n)\neq0$ 且

$F(n)=\prod_{d\mid n} f(d)$，证明 $f(n)=\prod_{d\mid n} F(d)^{\mu(n/d)}$。

14.8 ★★ 用 $\sum_{d^2\mid n}\mu(d)$ 计算 $Q(50)$，即 50 以内平方自由数的个数。

14.9 ★★ 证明两个随机正整数互素的概率为 $6/\pi^2$，并计算 $N=6$ 时的有限比例。

14.10 ★★★ (a) 在 $\{1,\dots,N\}$ 中独立均匀抽取四个整数并分成两组，证明两组最大公因数相等的概率在 $N\to\infty$ 时趋于 $2/5$。(b) 利用第 12 讲的 $x^n-1=\prod_{d\mid n}\Phi_d(x)$，证明 $\Phi_n(x)=\prod_{d\mid n}(x^d-1)^{\mu(n/d)}$。

练习答案与提示

14.1 $\mu(1)=1$，$\mu(2)=-1$，$\mu(6)=1$，$\mu(12)=0$，$\mu(30)=-1$。其中 $1,2,6,30$ 平方自由，12 不是。

14.2 反演中只有 $d=1,2,3,6$ 的项非零，故 $f(12)=144-36-16+4=96$。

14.3 写成 $f=\mu*F$，则 $f*U=(\mu*F)*U=F*(\mu*U)=F*\mathbf{1}=F$，展开即得 $F(n)=\sum_{d\mid n}f(d)$。

14.4 $36=2^2\cdot 3^2$，所以 $\varphi(36)=36(1-1/2)(1-1/3)=12$。非零 Möbius 项对应 $d=1,2,3,6$，其和为 $36-18-12+6=12$。

14.5 $\mu*\mu$ 在 2^2、3、5 上的局部值分别为 $1,-2,-2$，故 $(\mu*\mu)(60)=1\cdot(-2)\cdot(-2)=4$。

14.6 展开双重和并按 $n=ab$ 分组：

$$Z(f,s)Z(g,s)=\sum_{a,b\ge 1}\frac{f(a)g(b)}{(ab)^s}=\sum_{n\ge 1}\frac{\sum_{ab=n}f(a)g(b)}{n^s}=Z(f*g,s)$$

因为 $d=U*U$，所以 $Z(d,s)=Z(U,s)^2=\zeta(s)^2$。

14.7 在 $n=1$ 处必须有 $f(1)g(1)=1$，故 $f(1)=0$ 不可逆。若 $f(1)\neq 0$，令 $g(1)=1/f(1)$，并递推定义

$$g(n)=-\frac{1}{f(1)}\sum_{\substack{d\mid n\\ d>1}} f(d)g(n/d)$$

右侧只用到 g 的较小下标，故定义良好，并使 $(f*g)(n)=0$（$n>1$）。对 (b) 直接交换有限乘积的次序，避开复数对数的分支问题：

$$\prod_{d\mid n}F(d)^{\mu(n/d)}=\prod_{e\mid n}f(e)^{\sum_{e\mid d\mid n}\mu(n/d)}=f(n)$$

最后一项用到 $\sum_{r\mid n/e}\mu(r)=\mathbf{1}(n/e)$。

14.8

$$Q(50)=\sum_{d\le\sqrt{50}}\mu(d)\Big\lfloor\frac{50}{d^2}\Big\rfloor=50-12-5-2+1-1=31$$

非零项来自 $d=1,2,3,5,6,7$。

14.9 由 Möbius 恒等式，

$$C_N=\sum_{d\le N}\mu(d)\Big\lfloor\frac{N}{d}\Big\rfloor^2,\qquad \frac{C_N}{N^2}\longrightarrow\sum_{d\ge1}\frac{\mu(d)}{d^2}=\frac{1}{\zeta(2)}=\frac{6}{\pi^2}$$

当 $N=6$ 时，按最大值分类得到 $C_6=23$，有限比例为 $23/36$。

14.10 设 G_1,G_2 是两组的 gcd。利用正文中固定 gcd 的极限分布及统一尾界，$\lim_{N\to\infty}\Pr(G_1=G_2)=\sum_{k\ge1}36/(\pi^4k^4)=36\zeta(4)/\pi^4=2/5$。(b) 在 $\mathbb{Q}(x)$ 中，对第 12 讲的分解应用练习 **14.7**(b)；约分后右端仍为多项式：

$$\Phi_n(x)=\prod_{d\mid n}(x^d-1)^{\mu(n/d)}.$$

第 15 讲

组合计数与线性递推

Combinatorial Counting and Linear Recurrences

对应材料：MIT 18.781 Lecture 15 讲义；习题集 6、7 相关题目。

本讲目标

(1) 会用两种方式数同一个有限集合，并用它证明 $\sum_{d\mid n}\varphi(d)=n$ 及二项式恒等式。

(2) 会正确使用抽屉原理，证明奇素数模下一个三元二次型存在非平凡零点。

(3) 会写出容斥公式，理解交集的符号为何交替，并用它计算错排数和 $\varphi(n)$。

(4) 能从有序拆分、受限字符串等计数问题建立递推，并用初值和归纳法求出答案。

(5) 理解 Fibonacci 数的 Binet 公式是线性递推的特征根方法的结果，能解释它为何给出最近整数估计。

15.1 同一个集合可以怎样数？

先看一个熟悉的数：12。把 $1,2,\dots,12$ 按它们与 12 的最大公因数分组，得到的各组大小应当加起来等于 12。另一方面，每一组的大小又能用 Euler 函数表示。这种「先分组、再求和」的方法是本节的核心。

原讲义开头回顾了两个随机正整数的 gcd 分布。第 14 讲已从 Möbius 反演证明：极限中 $\Pr(\gcd(a,b)=k)=6/(\pi^2k^2)$，特别地互素概率为 $6/\pi^2$。原讲义随后给出「每个素数独立排除」的直觉；那种解释适合猜常数，但依赖性和极限交换需要额外论证。本讲把注意力转向可直接核查的有限计数恒等式。

定理 15.1（按最大公因数分组的恒等式） 对每个正整数 n，有

$$\sum_{d\mid n}\varphi(d)=n.$$

思路 把集合 $\{1,\dots,n\}$ 按 $\gcd(x,n)$ 的值分成互不相交的组。固定一组 $\gcd(x,n)=d$，除以 d 后就得到 $1,\dots,n/d$ 中与 n/d 互素的数，组的大小是 $\varphi(n/d)$。最后把因数 d 换成 n/d。

证明 对每个 $d\mid n$，令

$$S_d=\{x\in\{1,\dots,n\}:\gcd(x,n)=d\}.$$

每个 x 恰好属于一个 S_d，所以这些集合不交且并为 $\{1,\dots,n\}$。若 $x\in S_d$，写 $x=dy$，则

$1 \le y \le n/d$，并且

$$\gcd(x,n)=d \quad\Longleftrightarrow\quad \gcd(y,n/d)=1.$$

映射 $y \mapsto dy$ 在这两个集合之间是双射，因此 $|S_d|=\varphi(n/d)$。于是

$$n=\sum_{d\mid n}|S_d|=\sum_{d\mid n}\varphi(n/d)=\sum_{d\mid n}\varphi(d).$$

■

■ **例 15.1　把 12 按 gcd 分组**　列出 S_d（$d \mid 12$）的大小，并用它核对定理 15.1。

解　直接按最大公因数计算：

d	1	2	3	4	6	12
$\lvert S_d\rvert$	4	2	2	2	1	1
$\varphi(12/d)$	4	2	2	2	1	1

例如 $S_4=\{4,8\}$，除以 4 后得到 $\{1,2\}$，两个数都与 3 互素。这里 S_4 对应的是 $\varphi(12/4)=\varphi(3)=2$。各组大小之和为 $4+2+2+2+1+1=12$。这说明分组的关键是先固定 gcd，再把两个数同时除以它。 ♦

图 15.1 把 $1 \le x \le 12$ 的分组画成一条带颜色的数轴；每种颜色代表一个 $\gcd(x,12)$。

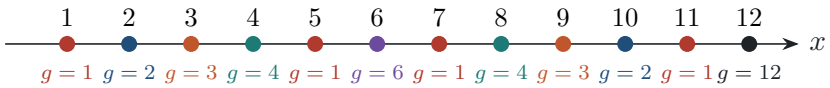


图 15.1　$1,\dots,12$ 按 $\gcd(x,12)$ 分组；每组大小是相应 Euler 函数值。

从图中看出，$g=1$ 的红色点正好有 $\varphi(12)=4$ 个；其他颜色的点对应

$$\varphi(6),\quad \varphi(4),\quad \varphi(3),\quad \varphi(2),\quad \varphi(1).$$

定理不是把这些数值碰巧相加，而是来自一个逐组的双射。

同一个思想也能处理组合数。设有两排各 n 个位置，要从 $2n$ 个位置中选 n 个。直接选择得到 $\binom{2n}{n}$；按第一排选了几个位置分类，则得到另一种计数。

命题 15.2（二项式卷积恒等式）　对每个 $n \ge 0$，

$$\binom{2n}{n}=\sum_{k=0}^{n}\binom{n}{k}^2.$$

证明　设 X 是从两排各 n 个位置组成的集合中选出 n 个位置的所有方案。直接数 X 得 $|X|=\binom{2n}{n}$。若第一排选了 k 个位置，则第二排必须选 $n-k$ 个，有 $\binom{n}{k}\binom{n}{n-k}=\binom{n}{k}^2$ 种。$k=0,1,\dots,n$ 互不相交且覆盖全部方案，求和即得结论。 ■

■ **例 15.2　核对 $n=4$ 的二项式恒等式**　计算等式两边，确认 $\binom{8}{4}=\sum_{k=0}^{4}\binom{4}{k}^2$。

解　左边为 $\binom{8}{4}=70$。右边是

$$\binom{4}{0}^2+\binom{4}{1}^2+\binom{4}{2}^2+\binom{4}{3}^2+\binom{4}{4}^2=1+16+36+16+1=70.$$

两边相等。平方来自两排各自的选择，不能把第二个因子误写成 $\binom{n}{k}$ 之外的数量。 ♦

想一想　为什么定理 15.1 中的分组必须按 $\gcd(x,n)$，而不能按「x 是否为偶数」分成两组？
答　偶奇分组只有两个组，不能直接得到 $\varphi(d)$ 这样的局部计数；更关键的是，x 与 n 的互素性由 n 的全部素因数共同决定。例如 $n=12$ 时，$x=5$ 和 $x=7$ 都是奇数，但 $\gcd(5,12)=\gcd(7,12)=1$；$x=3$ 也是奇数，却有 $\gcd(3,12)=3$。gcd 分组恰好记录了除去公共因子后剩下的互素条件。

15.2　Fibonacci 数与特征多项式的入口

定义 Fibonacci 数列

$$F_0=0,\qquad F_1=1,\qquad F_n=F_{n-1}+F_{n-2}\quad(n\ge 2).$$

前几项为 $0,1,1,2,3,5,8,13,21,34,55,\dots$。后面奇数拆分的计数会说明，同一个递推可以来自完全不同的组合问题。

定义 15.3（特征多项式）　对二阶常系数线性递推

$$u_n+au_{n-1}+bu_{n-2}=0\qquad(n\ge 2),$$

把

$$\chi(T)=T^2+aT+b$$

称为它的**特征多项式**（characteristic polynomial）。若 $\lambda\neq 0$ 是 χ 的根，则 $u_n=\lambda^n$ 是递推的一个解。

验证最后一句只需代入：

$$\lambda^n+a\lambda^{n-1}+b\lambda^{n-2}=\lambda^{n-2}(\lambda^2+a\lambda+b)=0.$$

若特征根为 0，则 $b=0$；取 $u_0=1$、$u_n=0$（$n\ge 1$）也满足递推。这样单独写出零根对应的数列，就不必使用 0^0 的记号。Fibonacci 递推写成 $F_n-F_{n-1}-F_{n-2}=0$，所以特征多项式为 T^2-T-1。它的两个根是

$$\alpha=\frac{1+\sqrt{5}}{2},\qquad \beta=\frac{1-\sqrt{5}}{2}.$$

命题 15.4（Binet 公式）　对每个 $n\ge 0$，

$$F_n=\frac{\alpha^n-\beta^n}{\alpha-\beta}=\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n-\frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n.$$

证明　由于 $\alpha^2=\alpha+1$、$\beta^2=\beta+1$，数列 α^n 和 β^n 都满足 Fibonacci 递推，故它们的线性组合也满足。令

$$G_n=\frac{\alpha^n-\beta^n}{\alpha-\beta}.$$

有 $G_0=0$、$G_1=1$，与 F_0,F_1 相同。递推关系和初值唯一决定整个数列，所以 $G_n=F_n$。 ■

■ **例 15.3**　**用 Binet 公式核对 F_8**　计算 F_8，并验证公式给出的值是整数 21。

解　递推得到 $F_2=1,F_3=2,F_4=3,F_5=5,F_6=8,F_7=13,F_8=21$。另一方面，Binet 公式给出

$$F_8=\frac{\alpha^8-\beta^8}{\sqrt{5}}.$$

由 $\alpha^2=\alpha+1$ 逐步得到 $\alpha^3=2\alpha+1$、$\alpha^4=3\alpha+2$，进而 $\alpha^8=21\alpha+13$；对 β 作同样计算得 $\beta^8=21\beta+13$。两式相减为 $21(\alpha-\beta)=21\sqrt{5}$，所以 $F_8=21$。这个例子提醒我们：公式中的两个无理数项必须一起计算，单独取整会失去抵消。 ♦

因为 $|\beta|=(\sqrt{5}-1)/2<1$，Binet 公式还给出一个很有用的估计：

$$\left|F_n-\frac{\alpha^n}{\sqrt{5}}\right|=\frac{|\beta|^n}{\sqrt{5}}<\frac{1}{2}\qquad(n\ge 1).$$

因此 F_n 是 $\alpha^n/\sqrt{5}$ 的最近整数。

同一个误差估计还说明相邻 Fibonacci 数的比值趋于 α。因为

$$\frac{F_{n+1}}{F_n}=\frac{\alpha^{n+1}-\beta^{n+1}}{\alpha^n-\beta^n}=\alpha\,\frac{1-(\beta/\alpha)^{n+1}}{1-(\beta/\alpha)^n},$$

而 $|\beta/\alpha|<1$，分子分母中的幂都趋于 0，所以

$$\lim_{n\to\infty}\frac{F_{n+1}}{F_n}=\alpha=\frac{1+\sqrt{5}}{2}.$$

这里的极限论证也解释了为什么递推的增长率由绝对值最大的特征根决定；若两个根的绝对值相同，则需要保留振荡项，不能直接套这一句。

图 15.2 将前几项比值画成折线：奇数下标的比值从下方靠近 α，偶数下标的比值从上方靠近。

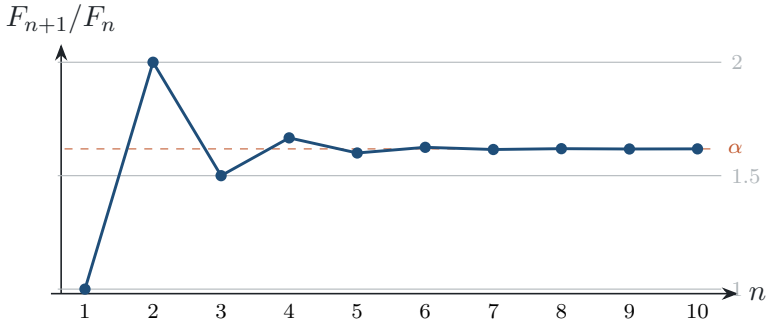


图 15.2　连续 Fibonacci 数之比在极限 α 两侧交替靠近；虚线为 $\alpha\approx 1.618034$。

由 Binet 公式相减可得 $F_{n+1}-\alpha F_n=\beta^n$，所以比值与极限的差恰为 β^n/F_n（$n\ge 1$）。因为 $F_n>0$、$\beta<0$，奇数项在下、偶数项在上；又 $|\beta|<1$ 且 F_n 不减，误差绝对值逐步缩小。

命题 15.5（Fibonacci 加法恒等式）　对 $m, n \geq 1$，有

$$F_{m+n} = F_{m-1}F_n + F_m F_{n+1}.$$

证明　固定 m 对 n 归纳。$n = 1$ 时右边为 $F_{m-1} + F_m = F_{m+1}$；$n = 2$ 时由一次 Fibonacci 递推也成立。设结论对 n 和 $n-1$ 成立，则

$$\begin{aligned} F_{m+n+1} &= F_{m+n} + F_{m+n-1} \\ &= F_{m-1}(F_n + F_{n-1}) + F_m(F_{n+1} + F_n) \\ &= F_{m-1}F_{n+1} + F_m F_{n+2}. \end{aligned}$$

这就是下一个 n 的结论。 ■

推论 15.6（整除性）　若 $m \mid n$，则 $F_m \mid F_n$。

证明　写 $n = km$。$k = 1$ 时 $F_{km} = F_m$。若 $F_m \mid F_{km}$，由命题 15.5

$$F_{(k+1)m} = F_{m-1}F_{km} + F_m F_{km+1}$$

也是 F_m 的倍数，归纳即得。 ■

命题 15.7（素数模下的 Fibonacci 检验）　设 $p > 5$ 为素数，则

$$F_p \equiv \left(\frac{5}{p}\right) \pmod p, \qquad F_{p+1} \equiv \frac{1 + \left(\frac{5}{p}\right)}{2} \pmod p.$$

因此 $p \equiv \pm 1 \pmod 5$ 时 $F_{p+1} \equiv 1$，$p \equiv \pm 2 \pmod 5$ 时 $F_{p+1} \equiv 0$。若再有 $p \equiv \pm 1 \pmod 5$，则 $F_{n+p-1} \equiv F_n \pmod p$ 对所有 $n \geq 0$ 成立。

证明　由 Binet 公式和二项式定理，

$$F_j = 2^{1-j} \sum_r \binom{j}{2r+1} 5^r.$$

当 $j = p$ 时，中间二项式系数被 p 整除，费马小定理给出

$$F_p \equiv 5^{(p-1)/2} \equiv \left(\frac{5}{p}\right) \pmod p;$$

当 $j = p + 1$ 时，只剩首项和末项，故

$$F_{p+1} \equiv \frac{1 + 5^{(p-1)/2}}{2} \equiv \frac{1 + \left(\frac{5}{p}\right)}{2} \pmod p.$$

二次互反律把 $\left(\frac{5}{p}\right)$ 化为 $\left(\frac{p}{5}\right)$，于是得到按 p 模 5 的两种情形。若 $p \equiv \pm 1 \pmod 5$，则 $F_p \equiv 1$、$F_{p+1} \equiv 1$，从而 $F_{p-1} = F_{p+1} - F_p \equiv 0$。命题 15.5 取 $m = p - 1$，得到

$$F_{n+p-1} = F_{p-2}F_n + F_{p-1}F_{n+1} \equiv F_n \pmod p.$$

■

例如 $p=11\equiv 1 \pmod 5$，所以 $F_{11}=89\equiv 1$、$F_{12}=144\equiv 1 \pmod{11}$，并且周期 $p-1=10$ 给出 $F_{n+10}\equiv F_n \pmod{11}$。

■ **例 15.4　最近整数估计**　证明 F_8 是 $\alpha^8/\sqrt{5}$ 的最近整数，并给出这个实数的近似值。

解　由上面的误差界，$|F_8-\alpha^8/\sqrt{5}|<1/2$，所以它必为最近整数。数值上 $\alpha\approx 1.618034$，从而 $\alpha^8/\sqrt{5}\approx 21.0095$，最近整数是 $21=F_8$。这里的「最近整数」来自共轭根 β 的绝对值小于 1，不是因为幂本身恰好是整数。 ♦

可以把这个估计用于快速检查较大的项。例如 $F_{10}=55$，而 $\alpha^{10}/\sqrt{5}$ 与 55 的距离小于 $1/2$；若用递推逐项算，$F_9=34$、$F_{10}=34+21=55$。两种方法的角色不同：递推保证整数值和精确计算，Binet 公式揭示增长率并给出近似。

直观理解　特征根把递推看成「每一步乘以一个固定比例」。Fibonacci 数列同时含有增长根 $\alpha>1$ 和衰减根 $|\beta|<1$；前者决定长期规模，后者负责把两个初值调到整数序列。第 16 讲会用生成函数和部分分式系统地解释这一点，并处理重根、非齐次项及更高阶递推。

从计数到公式　先按首项、末项或某个结构把对象分成互不相交的几类，得到递推；再写出初值。若递推是常系数线性的，就把 T^n 代入得到特征多项式，根给出通项的候选形式。

15.3　递推从哪里来？从拆分和字符串计数开始

很多计数问题不适合直接写出一个乘法公式，却能按「最后一步」或「第一个符号」分类。分类后，规模为 n 的答案由规模更小的答案组成，这就是递推。

定义 15.8（递推关系）　若数列 (u_n) 的每一项由前面若干项和固定的规则确定，就称这个规则为数列的**递推关系**（recurrence）。例如 $u_0=1$、$u_n=nu_{n-1}$ 生成阶乘数列；若递推中各项的系数与 n 无关且只对 u 做线性组合，则称为**常系数线性递推**（linear recurrence with constant coefficients）。

定义本身不要求递推是线性的：$u_n=nu_{n-1}$ 是递推但系数随 n 变化，$u_n=u_{n-1}^2+1$ 是递推但不是线性。后面只在需要时讨论线性情形。

例如 $a_0=0$、$a_n=a_{n-1}+n$ 给出 $0,1,3,6,10,15,\dots$，即逐项累加正整数；而 $b_0=1$、$b_n=nb_{n-1}$ 给出 $1,1,2,6,24,\cdots=n!$。两个规则都能生成数列，但只有第一个在把 n 当作变量时是「加上一个新末项」；第二个的倍数 n 随位置改变。描述递推时要同时注明起始下标、初值和适用范围。

命题 15.9（初值与递推唯一确定数列）　设递推式能把 u_n 唯一写成 $u_{n-1},\dots,u_{n-k}$ 的函数。给定 $u_0,\dots,u_{k-1}$ 后，至多有一个数列满足递推；若递推中的运算在给定数系内总能进行，则恰有一个。

证明　从 $n=k$ 开始，递推式将 u_k 唯一确定；再将已知的 $u_1,\dots,u_k$ 代入，确定 u_{k+1}。逐项归纳便能构造整个数列。若另一个数列具有相同初值并满足同一递推，那么它的第 k 项与第一个数列相同，归纳下去每一项都相同。 ■

■ **例 15.5　逐项求和得到三角数**　设 $a_0=0$，且 $a_n=a_{n-1}+n$（$n\ge 1$）。证明 $a_n=n(n+1)/2$。

解　展开递推得到 $a_n=1+2+\cdots+n$。也可以归纳：$n=0$ 时公式成立；若 $a_{n-1}=(n-1)n/2$，则

$$a_n=a_{n-1}+n=\frac{(n-1)n}{2}+n=\frac{n(n+1)}{2}.$$

所以每个 $n\ge 0$ 都满足公式。递推给出序列，归纳则验证闭式与递推、初值一致。 ♦

■ **例 15.6　正整数的有序拆分**　令 $r(n)$ 表示把 n 写成有序正整数之和的方法数，例如 $4=4=1+3=3+1=2+2=1+1+2=\cdots$。证明 $r(n)=2^{n-1}$。

解　按第一个加数分类。若第一个加数是 j，剩下部分是 $n-j$ 的一个有序拆分；当 $j=n$ 时剩下空和，贡献 1。因此

$$r(n)=1+r(1)+r(2)+\cdots+r(n-1)\qquad(n\ge 2).$$

对 $n-1$ 使用同一式，得到 $r(n-1)=1+r(1)+\cdots+r(n-2)$，相减可得 $r(n)=2r(n-1)$。初值 $r(1)=1$，归纳得到 $r(n)=2^{n-1}$。

还有一个直接的双射：把 n 个相邻的小球排成一行，$n-1$ 个间隙中每个都可以放或不放隔板。放隔板的位置唯一决定拆分，例如 $1|11|1$ 表示 $4=1+2+1$，所以共有 2^{n-1} 种。 ♦

■ **例 15.7　奇数加数的拆分**　令 c_n 表示把 n 写成有序奇数正整数之和的方法数。证明 $c_n=F_n$，其中 $F_1=F_2=1$。

解　按第一个加数分类。第一个加数为 1 时，删掉它得到 $n-1$ 的拆分；第一个加数至少为 3 时，把它减去 2，得到 $n-2$ 的拆分。这两个操作互为逆操作，因此

$$c_n=c_{n-1}+c_{n-2}\qquad(n\ge 3).$$

初值为 $c_1=1$（只有 1）和 $c_2=1$（只有 $1+1$），与 F_1,F_2 相同。由递推的唯一性，$c_n=F_n$。 ♦

这里使用的「递推唯一性」很简单，却是证明计数公式的关键：如果两个数列的前两项相同，并且都满足 $u_n=u_{n-1}+u_{n-2}$，那么它们的第三项都等于同一个数，第四项也相同，继续归纳便处处相同。因此只要把计数问题的递推和已知数列的递推、初值逐项对齐，就可以把一个组合问题转化为数列识别问题。

例 15.8　禁止相邻的 0 和 1　令 $f(n)$ 为长度为 n 的字符串 $a_1\cdots a_n$ 的数量，其中每个 $a_i\in\{0,1,2\}$，且不出现 00 或 11。证明

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

解　设 $g(n)$ 为满足条件且首字符为 0 的字符串数。交换 0,1 不改变条件，所以首字符为 1 的数量也是 $g(n)$；首字符为 2 时后面任意，数量为 $f(n-1)$。故

$$f(n)=2g(n)+f(n-1).$$

对 $n\ge 2$，若首字符是 0，第二字符为 2 时后缀任意，有 $f(n-2)$ 种。若第二字符为 1，剩下的长度 $n-2$ 字符串不能以 1 开头；交换 0,1 后，等价于数不以 0 开头的字符串，而在这些字符串前加 0 与长度 $n-1$、首字符为 0 的合法串一一对应，故有 $g(n-1)$ 种。因此 $g(n)=g(n-1)+f(n-2)$。又 $g(n)=(f(n)-f(n-1))/2$，代入前式得到

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

整理即得 $f(n)-2f(n-1)-f(n-2)=0$。特征多项式是 T^2-2T-1，根为 $\lambda=1+\sqrt{2}$ 与 $\mu=1-\sqrt{2}$。通解是 $A\lambda^n+B\mu^n$。由 $f(1)=3$、$f(2)=7$ 可取 $A=\lambda/2$、$B=\mu/2$，因为 $\lambda^2+\mu^2=6$、$\lambda^3+\mu^3=14$。于是得到题中的公式。代入 $n=1,2$ 分别为 $3,7$，与直接列举一致；之后所有项都由同一递推和初值唯一决定。 ♦

图 15.3 把「按首字符分类」画成递推图：0、1 两类要记住上一层的受限状态，2 类则只需要任意后缀。

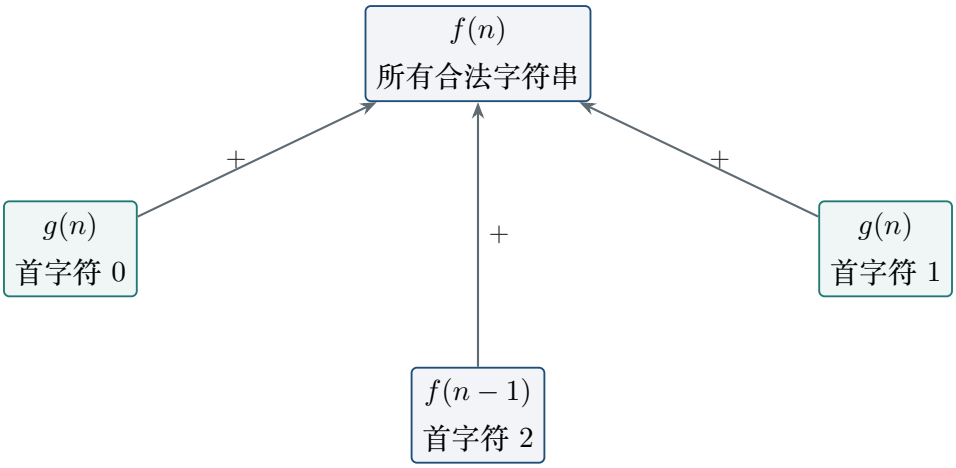


图 15.3　受限字符串按首字符拆分；两个受限状态各贡献 $g(n)$，首字符为 2 的部分贡献 $f(n-1)$。

从图中看出，递推不是凭空猜出的公式，而是把一个集合拆成互不相交的几类，再记录每一类与较小规模集合的双射。特征多项式只是把这个分类得到的线性关系转成代数方程。

易错提醒　不能只凭前几项就断言递推。例如 $1,1,2,3,5$ 既符合 Fibonacci 递推，也可能来自许多别的规则。要证明计数公式，必须先说明分类覆盖全部对象且各类互不相交，再验证初值。

想一想　为什么奇数拆分的递推中没有 c_{n-3}、c_{n-5} 等更多项？

答　按第一个加数分类时，1 对应删去 1，其余奇数 $m \geq 3$ 都先减去 2，统一对应到 $n-2$ 的拆分；不同的较大奇数已经被这个减 2 的双射整体收集。因此所有项合并成 $c_{n-1}+c_{n-2}$，不需要逐项列出。

15.4　容斥原理：把重复计数减回来

两组计数解决的是「把集合拆开」；容斥原理解决相反的困难：几个条件的集合会重叠。若只把各集合大小相加，交集中的元素就被重复计算，必须按交集层数交替加减。

定理 15.10（容斥原理）　设 X 是有限集合，$A_1,\dots,A_r \subseteq X$。则

$$\left|\bigcup_{i=1}^{r} A_i\right| = \sum_{k=1}^{r}(-1)^{k-1} \sum_{1\leq i_1<\cdots<i_k\leq r} |A_{i_1}\cap\cdots\cap A_{i_k}|.$$

约定空交集是 X，则补集形式为

$$\left|X \setminus \bigcup_{i} A_i\right| = \sum_{k=0}^{r}(-1)^{k} \sum_{i_1<\cdots<i_k} |A_{i_1}\cap\cdots\cap A_{i_k}|.$$

思路　不要逐个数交集的几何形状，而是固定一个元素 x。若 x 落在恰好 m 个事件中，它在右边的系数是 $\sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k}=1$；若 $m=0$，系数是 0。逐元素核对就完成证明。

证明　固定 $x \in X$。若 x 属于恰好 m 个集合 A_i，右边只会从这 m 个集合中选交集，因此 x 的总系数为

$$\sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k} = 1-(1-1)^m = 1 \qquad (m\geq 1).$$

这正是 x 属于并集时在左边的系数。若 $m=0$，右边没有贡献，左边也没有贡献。对所有 x 求和即得并集公式；补集公式同样是把空交集的 $k=0$ 项加入后逐元素核对。 ■

图 15.4 显示三集合中一个元素会在单交集、双交集和三重交集中被怎样加减。

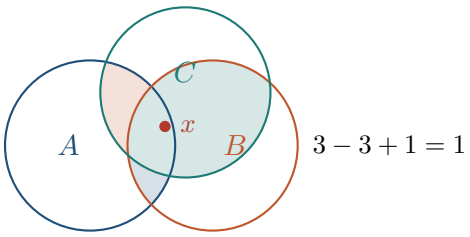


图 15.4　若 x 属于 $A\cap B\cap C$，它在三层中的总系数为 $3-3+1=1$。

■ **例 15.9**　计算 $\varphi(60)$　用容斥原理计算 $1,\dots,60$ 中与 60 互素的数的个数。

解　$60=2^2\cdot 3\cdot 5$。令 A_2,A_3,A_5 分别表示被 $2,3,5$ 整除的数。与 60 互素的数在三个集合的并集

之外，所以

$$\begin{aligned}\varphi(60) &= 60 - \frac{60}{2} - \frac{60}{3} - \frac{60}{5} + \frac{60}{6} + \frac{60}{10} + \frac{60}{15} - \frac{60}{30} \\ &= 60 - 30 - 20 - 12 + 10 + 6 + 4 - 2 = 16.\end{aligned}$$

整理就是 $60(1-1/2)(1-1/3)(1-1/5)=16$。指数 2 没有额外项，因为「被 2 整除」已经包含了「被 4 整除」的数。 ♦

把这个计算写成指标函数会更紧凑。对每个素数 $p \mid n$，令 $I_p(x)=1$ 若 $p \mid x$，否则 $I_p(x)=0$。那么

$$\mathbf{1}_{\gcd(x,n)=1} = \prod_{p \mid n} (1 - I_p(x)).$$

右边在所有 $p \nmid x$ 时等于 1，只要有一个共同素因子就为 0。对 $1 \le x \le n$ 求和，展开乘积后，选中一组素数 S 的交集就是被 $\prod_{p\in S} p$ 整除的数，有 $n/\prod_{p\in S} p$ 个；这正是容斥的每一项。

例如 $n=30$ 的不同素因子为 $2,3,5$，展开有八项：

$$30 - 15 - 10 - 6 + 5 + 3 + 2 - 1 = 8 = \varphi(30).$$

这个指标函数写法也能延伸到多个互素条件，是许多 Möbius 反演恒等式的有限版本。

例如，若要数 $1 \le x \le 30$ 中既不被 2 整除、也不被 3 整除、也不被 5 整除的数，就把三个指标相乘并求和。容斥的八项分别对应选择排除哪些素数；这与第 14 讲中 $\sum_{d^2 \mid n} \mu(d)$ 的写法有相同结构：每个被选中的素数贡献一个负号，交集贡献正号。区别在于，这里是在一个固定有限区间里数点，那里是在每个整数上写出一个平方自由指标。

■ **例 15.10　帽子错排**　有 4 个人和 4 顶分别写着姓名的帽子。帽子随机归还，求没有人拿回自己的帽子的方案数。

解　令 A_i 表示第 i 个人拿回自己的帽子。总方案数是 $4!=24$。指定 k 个人都拿回自己的帽子时，剩下的人任意分配，有 $(4-k)!$ 种；从 4 个人中选这 k 个人有 $\binom{4}{k}$ 种。容斥给出

$$24 - \binom{4}{1}3! + \binom{4}{2}2! - \binom{4}{3}1! + \binom{4}{4}0! = 24 - 24 + 12 - 4 + 1 = 9.$$

所以错排概率是 $9/24 = 3/8$。这题说明容斥的每一项都必须带上「先选出哪些条件同时成立」的组合数。 ♦

对一般的 n 个人，同样的计算给出错排数

$$!n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)! = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

这里第二个等号只是约去 $k!(n-k)!$。因此 $!n/n!$ 是交替和 $1 - 1 + 1/2! - 1/3! + \cdots$ 的前 $n+1$ 项；随着 n 增大，它接近 e^{-1}。本讲不需要用无穷级数证明这个极限，但有限公式已经说明错排概率为何稳定在约 36.8%。

易错提醒　错误做法是直接用 $4!-4\cdot 3!$，得到 0；它把「至少一个人拿对帽子」的方案重复扣除了。正确做法是继续加回两人同时拿对的方案，再减去三人同时拿对的方案，最后加回四人同时拿对的方案。

想一想　容斥公式中的符号为什么必须交替？若一个元素恰好属于三个 A_i，它在第一层被算了几次？

答　它先在单集合层被算 $\binom{3}{1}=3$ 次，在双交集层被减去 $\binom{3}{2}=3$ 次，在三重交集层加回 $\binom{3}{3}=1$ 次，总系数是 $3-3+1=1$。若不交替，重叠元素就不会只留下一个贡献。

15.5　抽屉原理：碰撞会迫使方程有解

把 $p+1$ 件物品放入 p 个盒子，至少有一个盒子装了两件。这句话很短，却能把「存在一个解」变成一个计数问题。

定理 15.11（抽屉原理）　设 $N\ge 1$、$q\ge 0$ 为整数。若 $N+1$ 个对象被放入 N 个盒子，则至少有一个盒子包含两个对象。更一般地，若 $qN+1$ 个对象被放入 N 个盒子，则某个盒子至少包含 $q+1$ 个对象。

证明　反设每个盒子至多一个对象，则 N 个盒子至多容纳 N 个对象，与有 $N+1$ 个对象矛盾。一般情形中，若每个盒子至多 q 个对象，总数至多 qN，同样矛盾。 ■

■ **例 15.11　连续整数中找互素数**　从集合 $\{1,2,\dots,2n+1\}$ 中选出 $n+1$ 个整数，证明其中有两个互素。说明只选 n 个时结论可能失败。

解　若选出的集合含有 1，它与任何整数都互素。否则把剩下的 $2n$ 个整数配成

$$\{2,3\},\ \{4,5\},\ \dots,\ \{2n,2n+1\}.$$

这是 n 个抽屉。选出的 $n+1$ 个数放入这些抽屉，抽屉原理保证某一对相邻整数同时被选出；相邻整数的最大公因数只能整除它们的差 1，所以二者互素。

只选 n 个时，选择全部偶数 $\{2,4,\dots,2n\}$ 就可能失败：任意两个数的 gcd 至少为 2。 ♦

本讲还需要一个「平方剩余抽屉」。设 p 是奇素数，x 只在 $0,1,\dots,(p-1)/2$ 中变化。若 $x\neq y$，则

$$x^2\equiv y^2 \pmod p \quad\Longrightarrow\quad (x-y)(x+y)\equiv 0 \pmod p.$$

由于 $0<x+y<p$，两因子都不被 p 整除，故平方值彼此不同。于是有 $(p+1)/2$ 个不同的平方剩余。

定理 15.12（三元二次型的非平凡零点）　设 p 是奇素数，且 $p\nmid abc$。则同余方程

$$ax^2+by^2+cz^2\equiv 0 \pmod p$$

有一个非平凡解，即 $(x,y,z)\not\equiv(0,0,0)\pmod p$。

思路　构造两组各有 $(p+1)/2$ 个互不相同的剩余类：第一组是 ax^2，第二组是 $-c-by^2$。它们总共有 $p+1$ 个元素，却只有 p 个剩余类，所以两组发生碰撞；碰撞恰好给出令 $z=1$ 的解。

证明　令

$$A=\{ax^2:0\le x\le (p-1)/2\},\qquad B=\{-c-by^2:0\le y\le (p-1)/2\}.$$

因为 $p\nmid a,b$，上面的平方剩余论证说明 A、B 各自都有 $(p+1)/2$ 个不同元素。若 A 与 B 不相交，它们合起来会有 $p+1$ 个不同的模 p 剩余类，这是不可能的；所以存在 x,y 使

$$ax^2\equiv -c-by^2 \pmod p.$$

取 $z=1$ 即得 $ax^2+by^2+cz^2\equiv 0\pmod p$。因为 $z=1$，这个解自动非平凡。 ■

■ **例 15.12　模 7 的具体碰撞**　用定理 15.12 找出

$$x^2+y^2+z^2\equiv 0 \pmod 7$$

的一个非平凡解。

解　取 $(x,y,z)=(1,2,3)$。代入得到

$$1^2+2^2+3^2=1+4+9=14=2\cdot 7,$$

所以三元组模 7 为零，且三个坐标不全为零。按证明中的两组计算，

$$A=\{0,1,2,4\},\qquad B=\{-1,-2,-5,-3\}=\{6,5,2,4\}.$$

交集中的 2 或 4 就给出碰撞。 ♦

易错提醒　抽屉原理只保证「有两个对象落在同一盒子」，不能保证指定的两个对象碰撞。比如把 4 个数放入 3 个盒子，只能推出某一盒至少有两个，不能指定第一个和第二个在一起。使用定理时必须先说明对象如何放入盒子。

想一想　上面的二次型定理为什么要求 p 是奇素数？

答　证明用到了区间 $0\le x\le (p-1)/2$，并用 $0<x+y<p$ 排除 $x+y\equiv 0\pmod p$；当 $p=2$ 时这个区间和排除步骤退化。例如 $x^2+y^2+z^2\equiv 0\pmod 2$ 虽然也有非平凡解 $(1,1,0)$，但不能由同一组「半数平方剩余」计数直接推出。

15.6　在十个连续整数中找一个「好数」

习题集把抽屉和容斥结合起来，问：任意十个连续正整数中，能否找到一个数与其余九个都互素？例如 $20, 21, \dots, 29$ 中，23 与其他九个都互素。下面的论证说明这样的数总能找到。

定理 15.13（十个连续正整数中存在共同互素元） 对任意正整数 m，集合 $\{m, m+1, \dots, m+9\}$ 中存在一个数 x，使 x 与集合内其余每个数都互素。

思路 若 x 与另一个元素 y 不互素，就有素数 p 同时整除它们，因此 $p \mid (y-x)$，而 $1 \le |y-x| \le 9$。所以「坏数」只能被 $2, 3, 5, 7$ 中一个素数整除。逐个估计这些倍数在十个数中的新覆盖量，证明坏数至多有九个。

证明 令 $S = \{m, \dots, m+9\}$，称 $x \in S$ 为坏数，如果它与某个不同的 $y \in S$ 不互素。取素数 $p \mid x$ 和 $p \mid y$，则 $p \mid (y-x)$，而 $y - x$ 的绝对值在 1 到 9 之间。因此 $p \in \{2, 3, 5, 7\}$，每个坏数都属于这些素数倍数的并集。

十个连续整数中恰有五个偶数。奇数的 3 倍数相邻相距 6，所以最多有两个不在偶数集合里的 3 倍数。两个相邻的 5 倍数相距 5，十个连续数中恰有两个；这两个数一奇一偶，因此最多新增一个不在偶数集合里的数。十个连续数中最多有两个 7 的倍数；若有两个，它们相差 7，其中一个为偶数，所以最多也新增一个。

故坏数最多为 $5 + 2 + 1 + 1 = 9$ 个，而 S 有十个元素，至少剩下一个不是坏数。这个元素与其他九个全都互素。 ■

■ **例 15.13　检查 20 到 29**　找出 $20, 21, \dots, 29$ 中一个与其余九个都互素的数。

解 取 23。逐一计算：

$$\gcd(23, 20) = \gcd(23, 21) = \gcd(23, 22) = 1,$$

且 23 是素数，大于 23 的这组数只有 $24, 25, 26, 27, 28, 29$，都不是 23 的倍数，所以 23 与它们也互素。因而 23 是一个好数。一般证明没有试图猜中这个数，而是先界定所有坏数可能落入的集合，再证明该集合装不下十个数。 ♦

历史与背景 抽屉原理通常与 Dirichlet 的名字联系在一起，但它只是一个直接的有限计数事实。容斥原理在古典组合计数中反复出现，错排数的表达式也因此常被称为 derangement formula。Fibonacci 数列以 Fibonacci 在《算盘书》中研究兔子繁殖问题而闻名；本讲只使用它的递推结构，不依赖这个故事的现实解释。

本讲小结

- 按 $\gcd(x,n)$ 分组给出 **Euler 函数恒等式** $\sum_{d\mid n}\varphi(d)=n$；每组与 $1,\dots,n/d$ 中的互素数通过除以 d 建立双射。
- 两排各 n 个位置的双重计数给出 $\binom{2n}{n}=\sum_{k=0}^{n}\binom{n}{k}^2$。
- 抽屉原理把 $N+1$ 个对象放入 N 个盒子时的碰撞转成存在性证明；奇素数模下的三元二次型因此有非平凡零点。
- 容斥原理按交集层数交替加减；特别地 $\varphi(n)=n\prod_{p\mid n}(1-1/p)$，4 人错排数为 9。
- 有序正整数拆分满足 $r(n)=2r(n-1)$；有序奇数拆分满足 $c_n=c_{n-1}+c_{n-2}$，所以 $c_n=F_n$。
- Fibonacci 递推的特征多项式是 T^2-T-1，其根给出 Binet 公式；因为 $|(1-\sqrt{5})/2|<1$，F_n 是 $\alpha^n/\sqrt{5}$ 的最近整数。

练习

15.1 ★ 对 $n=18$ 按 $\gcd(x,18)$ 分组，写出各组大小，并核对 $\sum_{d\mid 18}\varphi(d)=18$。

15.2 ★ 直接计算并核对 $\binom{6}{3}=\sum_{k=0}^{3}\binom{3}{k}^2$。

15.3 ★★ 对递推 $u_n=6u_{n-1}-11u_{n-2}+6u_{n-3}$ 写出特征多项式、特征根和通解的形式；说明还需要哪些初值才能唯一确定数列。

15.4 ★★ 设 $n\ge 1$。证明：从 $\{1,\dots,2n+1\}$ 选出 $n+1$ 个数，其中必有一对互素数。再给出一个只选 n 个数、且所选任意两数都不互素的例子。

15.5 ★★ 用容斥计算 5 个人的错排数，并给出错排概率。

15.6 ★★ 证明有序正整数拆分数 $r(5)=16$，并用隔板双射再算一次。

15.7 ★★ 求 7 拆成有序奇数正整数之和的方式数，并按加数个数分类列出数量。

15.8 ★★ 计算禁止相邻 00、11 的三字母字符串数 $f(4)$，并用 $f(n)=2f(n-1)+f(n-2)$ 核对。

15.9 ★★★ 证明 Fibonacci 加法恒等式

$$F_{m+n}=F_{m-1}F_n+F_mF_{n+1}\qquad (m,n\ge 1),$$

并推出 $F_m\mid F_n$（当 $m\mid n$）。

15.10 ★★★ 设 $p>5$ 为素数。证明 $F_p\equiv\left(\frac{5}{p}\right)\pmod p$，并确定 $F_{p+1}\pmod p$；若 $p\equiv\pm1\pmod 5$，再证明 $F_{n+p-1}\equiv F_n\pmod p$ 对所有 $n\ge 0$ 成立。

练习答案与提示

15.1 因数 $1,2,3,6,9,18$ 对应组大小 $\varphi(18),\varphi(9),\varphi(6),\varphi(3),\varphi(2),\varphi(1)=6,6,2,2,1,1$，总和为 18。

15.2 $\binom{6}{3}=20$；右边 $\binom{3}{0}^2+\binom{3}{1}^2+\binom{3}{2}^2+\binom{3}{3}^2=1+9+9+1=20$。

15.3 特征多项式为 $T^3-6T^2+11T-6=(T-1)(T-2)(T-3)$，根是 $1,2,3$，通解为 $u_n=A+B2^n+C3^n$。需要 u_0,u_1,u_2 三个初值解出 A,B,C。

15.4 把 $\{2,3\},\{4,5\},\dots,\{2n,2n+1\}$ 作为 n 个盒子；若选集含 1，则 1 与任一所选数互素。若不含 1，$n+1$ 个数放入 n 个盒子，抽屉原理保证某盒中的一对相邻数同时被选出，它们互素，因为公因数必整除差 1。只选 n 个时取全部偶数 $\{2,4,\dots,2n\}$；任意两个不同的偶数都有公因数 2，所以没有互素对。

15.5 $!5 = 5! - \binom{5}{1}4! + \binom{5}{2}3! - \binom{5}{3}2! + \binom{5}{4}1! - \binom{5}{5}0! = 120 - 120 + 60 - 20 + 5 - 1 = 44$，概率为 $44/120 = 11/30$。

15.6 递推给出 $r(1)=1, r(2)=2, r(3)=4, r(4)=8, r(5)=16$。隔板法中 5 个球之间的 4 个间隙各有放或不放两种选择，共 $2^4 = 16$。

15.7 数量为 $F_7 = 13$。按加数个数分别为 $1, 6, 5, 1$（1 个、3 个、5 个、7 个加数），总和 13。

15.8 $f(1)=3, f(2)=7$，故 $f(3) = 2\cdot 7 + 3 = 17$，$f(4) = 2\cdot 17 + 7 = 41$。

15.9 固定 m 对 n 归纳。$n = 1, 2$ 时直接代入 Fibonacci 递推即可。对 $n \geq 2$，若恒等式对 n 和 $n-1$ 成立，则

$$F_{m+n+1} = F_{m+n} + F_{m+n-1} = F_{m-1}F_{n+1} + F_m F_{n+2}.$$

若 $n = mk$，利用恒等式 $F_{(k+1)m} = F_{m-1}F_{km} + F_m F_{km+1}$，归纳得 $F_m \mid F_{km}$。

15.10 见命题 15.7 的二项式展开证明：$F_p \equiv \left(\frac{5}{p}\right)$，且 $F_{p+1} \equiv (1 + \left(\frac{5}{p}\right))/2$。当 $p \equiv \pm 1 \pmod 5$ 时，$F_{p-1} \equiv 0$、$F_{p-2} \equiv 1$；代入练习 15.9 的恒等式即得 $F_{n+p-1} \equiv F_n \pmod p$。

第 16 讲

生成函数与线性递推

Generating Functions and Linear Recurrences

对应材料：MIT 18.781 Lecture 16 讲义；习题集 7 第 4、6、7、8 题。

本讲目标

(1) 会写出常系数线性递推的特征多项式；特征根互不相同时，会写出通解 $u_n = \sum c_i \lambda_i^n$，并用初值解范德蒙德方程组定出系数。

(2) 理解生成函数的观点：满足 k 阶递推的数列，生成函数恰好是分母为 $1 + a_1 z + \cdots + a_k z^k$ 的有理函数；部分分式加几何级数就给出通项公式，λ^n 的形式是推出来的，不是猜出来的。

(3) 会处理重根（出现 $n\lambda^n$、$n^2\lambda^n$）和非齐次项（通解 = 特解 + 齐次通解，注意「共振」时要乘 n）。

(4) 会用 Binet 公式证明 Cassini 恒等式，并由此证明：素数 $p > 5$ 恰好整除 F_{p-1} 和 F_{p+1} 中的一个。

(5) 会用「共轭根之和是整数」的技巧，证明 $\lfloor (1+\sqrt{3})^{2n+1} \rfloor$ 能被 2^{n+1} 整除这类问题。

16.1 常系数线性递推与特征多项式

第 15 讲已经见过斐波那契数列 $F_0 = 0$，$F_1 = 1$，$F_n = F_{n-1} + F_{n-2}$，并证明了 Binet 公式

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

本讲回答两个问题：这样的公式是**怎么得到的**；得到之后**能拿来做什么**。先把要研究的对象说清楚。

定义 16.1（常系数线性递推） 设 $k \geq 1$，$a_1, \ldots, a_k$ 是常数（整数、有理数或复数均可），$a_k \neq 0$。数列 $(u_n)_{n \geq 0}$ 若对所有 $n \geq k$ 满足

$$u_n + a_1 u_{n-1} + a_2 u_{n-2} + \cdots + a_k u_{n-k} = 0,$$

就称它满足一个 **k 阶常系数齐次线性递推**（linear recurrence with constant coefficients）。$u_0, \ldots, u_{k-1}$ 称为**初值**（initial values）。

与递推对应的多项式

$$\chi(T) = T^k + a_1 T^{k-1} + \cdots + a_{k-1} T + a_k$$

称为这个递推的**特征多项式**（characteristic polynomial），它的根称为**特征根**。

斐波那契数列写成 $F_n - F_{n-1} - F_{n-2} = 0$，所以 $k = 2$，$a_1 = a_2 = -1$，$\chi(T) = T^2 - T - 1$。等比数列 $u_n = 3u_{n-1}$ 是 $k = 1$ 的情形，$\chi(T) = T - 3$。下面几个**不是**本节意义下的线性递推：

- $u_n = n\,u_{n-1}$（$u_n = n!$）：系数随 n 变化，不是常系数；
- $u_n = u_{n-1}^2 + 1$：不是线性的；
- $u_n = 5u_{n-1} - 6u_{n-2} + n^2$：多了一个不含 u 的项，称为非齐次递推，留到 §16.3 处理。

记号上，二阶递推常写成 $u_n + a u_{n-1} + b u_{n-2} = 0$，$\chi(T) = T^2 + aT + b$。这样斐波那契数列对应 $a = b = -1$。

整套理论建立在下面三个简单事实上。

引理 16.2（三个基本事实）　固定一个 k 阶常系数线性递推，记它的全体解（复数列）为 V。

(1) **初值决定一切**：若两个解的前 k 项相同，则它们完全相同。反之任给 k 个初值，都恰有一个解以它们为初值。

(2) **叠加原理**：若 $(u_n), (v_n) \in V$，则对任意常数 α, β，$(\alpha u_n + \beta v_n) \in V$。

(3) **指数解**：设 $\lambda \neq 0$。数列 $u_n = \lambda^n$ 是解，当且仅当 $\chi(\lambda) = 0$。

证明　(1) 递推把 u_n 写成 $u_{n-1}, \ldots, u_{n-k}$ 的函数：$u_n = -a_1 u_{n-1} - \cdots - a_k u_{n-k}$。对 n 用强归纳法，前 k 项相同推出第 k 项相同，再推出第 $k+1$ 项相同，以此类推。任给初值，也可以用这个式子逐项算出整个数列。

(2) 递推式左边对 u 是线性的：$(\alpha u_n + \beta v_n) + \sum_i a_i(\alpha u_{n-i} + \beta v_{n-i}) = \alpha \cdot 0 + \beta \cdot 0 = 0$。

(3) 把 $u_n = \lambda^n$ 代入左边：$\lambda^n + a_1\lambda^{n-1} + \cdots + a_k\lambda^{n-k} = \lambda^{n-k}\,\chi(\lambda)$。因为 $\lambda \neq 0$，它对所有 $n \geq k$ 为零当且仅当 $\chi(\lambda) = 0$。 ∎

准备。下面的一般 k 阶论证使用向量空间、线性方程组与行列式，特别是范德蒙德行列式。尚未学过时，可先读紧随定理后的二阶消元与具体例题，并按下一节的系数规则学习生成函数；补学这些线性代数知识后，再回读一般 k 阶通解的证明。

(1)(2) 合起来说：V 是一个向量空间，而「取前 k 项」是 $V \to \mathbb{C}^k$ 的线性双射，所以 $\dim V = k$。于是只要找到 k 个「彼此独立」的解，就找到了全部解。(3) 正好提供候选：每个特征根给出一个解。

定理 16.3（特征根互不相同的情形）　若特征多项式 χ 有 k 个互不相同的根 $\lambda_1,\dots,\lambda_k$，则递推的每个解都能唯一地写成

$$u_n = c_1\lambda_1^n + c_2\lambda_2^n + \cdots + c_k\lambda_k^n,$$

系数 $c_1,\dots,c_k$ 由初值通过**范德蒙德方程组**（Vandermonde system）唯一确定：

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_k \\ \vdots & \vdots & & \vdots \\ \lambda_1^{k-1} & \lambda_2^{k-1} & \cdots & \lambda_k^{k-1} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} = \begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_{k-1} \end{pmatrix}.$$

证明　由引理 16.2(2)(3)，右边的表达式是解。系数矩阵的行列式是 $\prod_{i<j}(\lambda_j-\lambda_i)$，根互不相同时它不为零，所以方程组有唯一解 (c_i)，此时表达式与 (u_n) 的前 k 项相同。再由引理 16.2(1)，二者处处相同。 ■

$k=2$ 时，记两个特征根为 $\lambda\neq\mu$，方程组就是 $\alpha+\beta=u_0$，$\alpha\lambda+\beta\mu=u_1$，解得

$$\alpha = \frac{u_1-\mu u_0}{\lambda-\mu}, \qquad \beta = \frac{\lambda u_0 - u_1}{\lambda-\mu}. \tag{16.1}$$

对斐波那契数列，$\lambda,\mu=\frac{1\pm\sqrt5}{2}$，$\lambda-\mu=\sqrt5$，$u_0=0$，$u_1=1$，于是 $\alpha=1/\sqrt5$，$\beta=-1/\sqrt5$，这正是 Binet 公式。图 16.1 画出了三种描述同一个解的方式之间的关系。

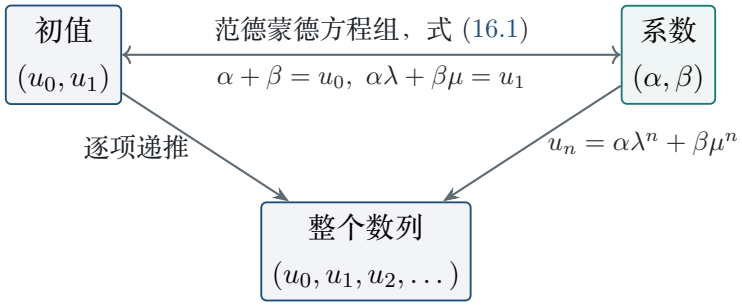


图 16.1　二阶递推（$\lambda\neq\mu$）的解有三种等价描述。初值和系数都是两个数，二者通过一个可逆的 2×2 线性方程组相互确定。

■ **例 16.1**　求满足 $u_n=5u_{n-1}-6u_{n-2}$，$u_0=0$，$u_1=1$ 的数列的通项。

解　递推写成 $u_n-5u_{n-1}+6u_{n-2}=0$，特征多项式 $\chi(T)=T^2-5T+6=(T-2)(T-3)$，特征根 $2,3$ 互不相同。由定理 16.3，$u_n=\alpha\cdot3^n+\beta\cdot2^n$。代入初值：

$$n=0:\ \alpha+\beta=0, \qquad n=1:\ 3\alpha+2\beta=1.$$

第二式减去第一式的 2 倍得 $\alpha=1$，于是 $\beta=-1$。所以

$$u_n = 3^n - 2^n.$$

核对：递推逐项算出 $0,1,5,19,65,211,665,\dots$；公式给出 $3^2-2^2=5$，$3^3-2^3=19$，$3^4-2^4=65$，$3^6-2^6=729-64=665$，一致。 ♦

■ **例 16.2　三阶递推**　求满足 $u_n=6u_{n-1}-11u_{n-2}+6u_{n-3}$，$u_0=1$，$u_1=2$，$u_2=6$ 的数列的通项。

解　特征多项式 $\chi(T)=T^3-6T^2+11T-6$。试根：$\chi(1)=1-6+11-6=0$，所以 $T-1$ 是因式，做除法得 $\chi(T)=(T-1)(T^2-5T+6)=(T-1)(T-2)(T-3)$。三个根 $1,2,3$ 互不相同，所以 $u_n=\alpha+\beta\cdot 2^n+\gamma\cdot 3^n$。范德蒙德方程组是

$$\alpha+\beta+\gamma=1,\qquad \alpha+2\beta+3\gamma=2,\qquad \alpha+4\beta+9\gamma=6.$$

用第二、第三式分别减去第一式：$\beta+2\gamma=1$，$3\beta+8\gamma=5$。把 $\beta=1-2\gamma$ 代入后一式，$3+2\gamma=5$，得 $\gamma=1$，$\beta=-1$，$\alpha=1$。所以

$$u_n=1-2^n+3^n.$$

核对：递推给出 $u_3=6\cdot 6-11\cdot 2+6\cdot 1=20$，$u_4=6\cdot 20-11\cdot 6+6\cdot 2=66$；公式给出 $1-8+27=20$，$1-16+81=66$。数列是 $1,2,6,20,66,212,666,\dots$。 ♦

特征根可以是复数。只要它们互不相同，定理 16.3 照样成立，因为证明只用到了代数运算。

想一想　数列 $u_n=u_{n-1}-u_{n-2}$，$u_0=0$，$u_1=1$ 的前几项是 $0,1,1,0,-1,-1,0,1,1,\dots$，看起来以 6 为周期。它的特征根是复数。请写出通项，并从特征根解释周期为什么是 6。

答　$\chi(T)=T^2-T+1$，根 $\lambda,\mu=\frac{1\pm i\sqrt{3}}{2}=e^{\pm i\pi/3}$，$\lambda-\mu=i\sqrt{3}$。由式 (16.1)，$u_n=\frac{\lambda^n-\mu^n}{i\sqrt{3}}=\frac{2}{\sqrt{3}}\sin\frac{n\pi}{3}$，每一项都是整数。因为 $(T+1)(T^2-T+1)=T^3+1$，有 $\lambda^3=\mu^3=-1$，所以 $\lambda^6=\mu^6=1$，从而 $u_{n+6}=u_n$。实际上还有 $u_{n+3}=-u_n$。

16.2　生成函数：把递推折叠成一个分式

上一节的方法是先猜解的形式 λ^n，再验证。这一节换一个角度：把整个数列装进一个「函数」，让递推自己告诉我们答案。

定义 16.4（生成函数）　数列 $(u_n)_{n\ge 0}$ 的**生成函数**（generating function）是形式幂级数

$$U(z)=\sum_{n\ge 0}u_nz^n=u_0+u_1z+u_2z^2+\cdots.$$

「形式」的意思是：我们只把 $U(z)$ 当作一串系数的记账本，按多项式的规则做加法和乘法（z^n 的系数只依赖有限多项），不讨论收敛。两个形式幂级数相等，当且仅当每个 z^n 的系数相等。最基本的例子是几何级数：因为 $(1-\lambda z)(1+\lambda z+\lambda^2z^2+\cdots)=1$，

$$\frac{1}{1-\lambda z}=\sum_{n\ge 0}\lambda^nz^n. \tag{16.2}$$

这里 $\frac{1}{1-\lambda z}$ 指的就是 $1-\lambda z$ 的乘法逆元，它的系数正是数列 λ^n。

现在设 $u_n + au_{n-1} + bu_{n-2} = 0$（$n \ge 2$）。把 U、azU、bz^2U 按 z 的幂次对齐写出来，再相加（图 16.2）：乘以 z 就是把系数向右平移一格。

	1	z	z^2	z^3	$\cdots$	z^n
U	u_0	u_1	u_2	u_3	$\cdots$	u_n
azU		au_0	au_1	au_2	$\cdots$	au_{n-1}
bz^2U			bu_0	bu_1	$\cdots$	bu_{n-2}
和	u_0	$u_1 + au_0$	0	0	$\cdots$	0

高亮这一列之和 $= u_n + au_{n-1} + bu_{n-2} = 0$

图 16.2　移位相加。从 z^2 起，每一列恰好是递推式的左边，因此为零；只有前两列留下与初值有关的项。

于是

$$U(z)\,(1 + az + bz^2) = u_0 + (u_1 + au_0)\,z, \qquad \text{即} \qquad U(z) = \frac{u_0 + (u_1 + au_0)z}{1 + az + bz^2}.$$

分母 $1 + az + bz^2$ 是特征多项式 $T^2 + aT + b$ 的系数倒过来写。若 $T^2 + aT + b = (T - \lambda)(T - \mu)$，把 $T = 1/z$ 代入再乘以 z^2，得到

$$1 + az + bz^2 = (1 - \lambda z)(1 - \mu z).$$

当 $\lambda \ne \mu$ 时做部分分式，再用几何级数 (16.2) 展开：

$$U(z) = \frac{A}{1 - \lambda z} + \frac{B}{1 - \mu z} = \sum_{n \ge 0} (A\lambda^n + B\mu^n)\, z^n.$$

比较 z^n 的系数，$u_n = A\lambda^n + B\mu^n$。解的形式 λ^n 在这里是**推出来的**：它来自几何级数。

易错提醒　分母 $1 + az + bz^2$ 的根是 $1/\lambda$ 和 $1/\mu$，不是 λ, μ。例如斐波那契数列的分母是 $1 - z - z^2$，它的根是 $\frac{-1 \pm \sqrt{5}}{2}$，恰是 $\frac{1 \pm \sqrt{5}}{2}$ 的倒数。做部分分式时要把分母写成 $(1 - \lambda z)(1 - \mu z)$ 的形式，展开才是 λ^n；若写成 $(z - r_1)(z - r_2)$，展开出来的是 r_1^{-n}，容易弄错。

同样的移位相加对任意阶都成立。

引理 16.5（递推 $\iff$ 有理生成函数）　设 $Q(z) = 1 + a_1 z + \cdots + a_k z^k (a_k \ne 0), U(z) = \sum u_n z^n$。则 (u_n) 满足递推 $u_n + a_1 u_{n-1} + \cdots + a_k u_{n-k} = 0$（$n \ge k$），当且仅当

$$U(z) = \frac{R(z)}{Q(z)}, \qquad R \text{ 是次数} < k \text{ 的多项式.}$$

此时 R 的系数由初值决定：$R(z)$ 就是 $Q(z)U(z)$ 的前 k 项。

证明　按多项式乘法，$Q(z)U(z)$ 中 z^n 的系数是 $u_n + a_1 u_{n-1} + \cdots + a_k u_{n-k}$（约定下标为负的 u 为零）。递推成立，等价于 $n \ge k$ 时这些系数全为零，即 $Q(z)U(z)$ 是次数 $< k$ 的多项式 $R(z)$。Q 的常数项是 1，在形式幂级数中可逆，所以 $U = R/Q$。 ■

生成函数的对应关系

递推这边		生成函数这边
特征多项式 $\chi(T)=T^k+a_1T^{k-1}+\cdots+a_k$	$\longleftrightarrow$	分母 $Q(z)=z^k\chi(1/z)$
特征根 λ_i	$\longleftrightarrow$	分母的因式 $1-\lambda_i z$
k 个初值	$\longleftrightarrow$	分子 $R(z)$ 的 k 个系数
通项 $\sum c_i\lambda_i^n$	$\longleftrightarrow$	部分分式 $\sum \frac{c_i}{1-\lambda_i z}$

■ **例 16.3　用生成函数重做例 16.1**　用生成函数求 $u_n=5u_{n-1}-6u_{n-2}$，$u_0=0$，$u_1=1$ 的通项。

解　这里 $a=-5$，$b=6$。分子是 $u_0+(u_1+au_0)z=0+(1-0)z=z$，分母是 $1-5z+6z^2=(1-2z)(1-3z)$，所以

$$U(z)=\frac{z}{(1-2z)(1-3z)}=\frac{A}{1-3z}+\frac{B}{1-2z}.$$

去分母：$z=A(1-2z)+B(1-3z)$。令 $z=\frac13$，得 $\frac13=A\cdot\frac13$，$A=1$；令 $z=\frac12$，得 $\frac12=B\cdot(-\frac12)$，$B=-1$。于是

$$U(z)=\frac{1}{1-3z}-\frac{1}{1-2z}=\sum_{n\ge0}(3^n-2^n)z^n,\qquad u_n=3^n-2^n.$$

与例 16.1 的结果相同。也可以直接展开核对前几项：$\frac{z}{1-5z+6z^2}=z+5z^2+19z^3+65z^4+\cdots$。 ♦

■ **例 16.4　推导 Binet 公式**　用生成函数推导斐波那契数列的通项公式。

解　$F_n-F_{n-1}-F_{n-2}=0$，即 $a=b=-1$；初值 $F_0=0$，$F_1=1$，分子为 $0+(1-0)z=z$。所以

$$F(z)=\sum_{n\ge0}F_nz^n=\frac{z}{1-z-z^2}=z+z^2+2z^3+3z^4+5z^5+8z^6+\cdots.$$

设 $\lambda,\mu=\frac{1\pm\sqrt5}{2}$ 是 T^2-T-1 的两个根，则 $\lambda+\mu=1$，$\lambda\mu=-1$，$\lambda-\mu=\sqrt5$，并且 $1-z-z^2=(1-\lambda z)(1-\mu z)$。部分分式 $z=A(1-\mu z)+B(1-\lambda z)$：令 $z=1/\lambda$，得 $\frac1\lambda=A\cdot\frac{\lambda-\mu}{\lambda}$，$A=\frac{1}{\lambda-\mu}=\frac{1}{\sqrt5}$；令 $z=1/\mu$，同理 $B=\frac{1}{\mu-\lambda}=-\frac{1}{\sqrt5}$。于是

$$F(z)=\frac{1}{\sqrt5}\left(\frac{1}{1-\lambda z}-\frac{1}{1-\mu z}\right),\qquad \text{从而}$$

$$\boxed{F_n=\frac{\lambda^n-\mu^n}{\sqrt5}=\frac{\lambda^n-\mu^n}{\lambda-\mu}.}$$

因为 $|\mu|\approx0.618<1$，第二项 $|\mu^n/\sqrt5|<\frac12$，所以 F_n 是离 $\lambda^n/\sqrt5$ 最近的整数（第 15 讲用过这一点）。例如 $\lambda^{10}/\sqrt5\approx55.0036$，$F_{10}=55$。 ♦

■ **例 16.5　三进制串计数，习题集 7 第 7 题**　用 $0,1,2$ 组成长度为 n 的串，要求不出现子串 00，也不出现子串 11。设这样的串共有 $f(n)$ 个，求 $f(n)$。

解　第一步：建立递推。约定 $f(0)=1$（空串）。$f(1)=3$；长度 2 的串共 9 个，去掉 $00,11$，$f(2)=7$。按最后一个字符分类。设 $s(n)$ 是以 2 结尾的合格串个数，$t(n)$ 是以 0 或 1 结尾的个数。

- 任何长度 $n-1$ 的合格串后面接 2 都合格，所以 $s(n)=f(n-1)$。
- 以 0 结尾：前面 $n-1$ 位必须以 1 或 2 结尾。由 $0\leftrightarrow 1$ 的对称性，以 1 结尾的有 $t(n-1)/2$ 个，所以以 0 结尾的串有 $t(n-1)/2+s(n-1)$ 个。以 1 结尾的同样多，因此 $t(n)=t(n-1)+2s(n-1)$。

相加：$f(n)=f(n-1)+t(n-1)+2s(n-1)=2f(n-1)+s(n-1)=2f(n-1)+f(n-2)$，对 $n\geq 2$ 成立。核对：$f(2)=2\cdot 3+1=7$。

第二步：生成函数。$a=-2$，$b=-1$，分子 $1+(3-2\cdot 1)z=1+z$，

$$\sum_{n\geq 0} f(n)z^n=\frac{1+z}{1-2z-z^2}=\frac{A}{1-\lambda z}+\frac{B}{1-\mu z},\qquad \lambda,\mu=1\pm\sqrt{2}.$$

令 $z=1/\lambda$：$1+\frac{1}{\lambda}=A\cdot\frac{\lambda-\mu}{\lambda}$，$A=\frac{\lambda+1}{\lambda-\mu}=\frac{2+\sqrt{2}}{2\sqrt{2}}=\frac{1+\sqrt{2}}{2}=\frac{\lambda}{2}$。同理 $B=\frac{\mu+1}{\mu-\lambda}=\frac{\mu}{2}$。所以

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

核对：$n=2$ 时 $(1\pm\sqrt{2})^3=7\pm 5\sqrt{2}$，$f(2)=14/2=7$。前几项是 $1,3,7,17,41,99,239,577,1393$，与直接枚举一致。 ♦

16.3　重根与非齐次递推

16.3.1　重根

若 $\chi(T)=(T-\lambda)^2$，定理 16.3 的表达式 $\alpha\lambda^n+\beta\mu^n$ 退化为 $(\alpha+\beta)\lambda^n$，只剩一个自由参数，配不上两个初值。例如 $u_n=4u_{n-1}-4u_{n-2}$，$\chi(T)=(T-2)^2$；若硬写 $u_n=c\cdot 2^n$，则 $u_0=1$ 要求 $c=1$，$u_1=4$ 又要求 $c=2$，矛盾。缺的那个解是什么？

生成函数直接给出答案。此时分母是 $(1-\lambda z)^2$，而几何级数的平方是

$$\frac{1}{(1-\lambda z)^2}=\Big(\sum_{i\geq 0}\lambda^i z^i\Big)\Big(\sum_{j\geq 0}\lambda^j z^j\Big)=\sum_{n\geq 0}(n+1)\lambda^n z^n,$$

因为 z^n 的系数是 λ^n 乘以满足 $i+j=n$ 的非负整数对 (i,j) 的个数 $n+1$。部分分式写成 $\frac{A}{1-\lambda z}+\frac{B}{(1-\lambda z)^2}$，展开得到 $A\lambda^n+B(n+1)\lambda^n$：多出来的解是 $n\lambda^n$。

> **定理 16.6（二重根）**　设 $\chi(T)=T^2+aT+b=(T-\lambda)^2$，$\lambda\neq 0$。则递推 $u_n+au_{n-1}+bu_{n-2}=0$ 的每个解都能唯一地写成
>
> $$u_n=(\alpha+\beta n)\lambda^n,\qquad \alpha=u_0,\quad \beta=\frac{u_1}{\lambda}-u_0.$$

证明　不用生成函数也能直接验证 $n\lambda^n$ 是解。由 $(T-\lambda)^2=T^2-2\lambda T+\lambda^2$ 得 $a=-2\lambda$，$b=\lambda^2$。代入：

$$n\lambda^n+a(n-1)\lambda^{n-1}+b(n-2)\lambda^{n-2}=n\lambda^{n-2}\big(\lambda^2+a\lambda+b\big)-\lambda^{n-2}\big(a\lambda+2b\big)=0-\lambda^{n-2}(-2\lambda^2+2\lambda^2)=0.$$

由叠加原理，$(\alpha+\beta n)\lambda^n$ 是解。初值条件为 $\alpha=u_0$，$(\alpha+\beta)\lambda=u_1$，因为 $\lambda\neq 0$，它有唯一解。再由引理 16.2(1) 即得。 ■

■ **例 16.6**　求 $u_n=4u_{n-1}-4u_{n-2}$，$u_0=1$，$u_1=4$ 的通项，并用生成函数核对。

解　**方法一（定理 16.6）**。$\lambda=2$，$\alpha=u_0=1$，$\beta=\frac{4}{2}-1=1$，所以 $u_n=(n+1)2^n$。

方法二（生成函数）。$a=-4$，$b=4$，分子 $u_0+(u_1+au_0)z=1+(4-4)z=1$，所以

$$U(z)=\frac{1}{1-4z+4z^2}=\frac{1}{(1-2z)^2}=\sum_{n\geq 0}(n+1)2^nz^n.$$

这次部分分式一步都不用做。两种方法都给出 $1,4,12,32,80,192,\ldots$；递推核对：$u_2=4\cdot 4-4\cdot 1=12$，$u_3=4\cdot 12-4\cdot 4=32$。 ♦

一般情形的证明思路完全相同：重数为 m 的根贡献分母因式 $(1-\lambda z)^m$，而 $\frac{1}{(1-\lambda z)^j}$ 是 j 个几何级数之积，z^n 的系数等于 λ^n 乘以把 n 写成 j 个有序非负整数之和的方法数 $\binom{n+j-1}{j-1}$，这是 n 的 $j-1$ 次多项式。

定理 16.7（一般情形）　设 $\chi(T)=(T-\lambda_1)^{m_1}\cdots(T-\lambda_r)^{m_r}$，$\lambda_i$ 互不相同且非零，$m_1+\cdots+m_r=k$。则递推的解恰好是形如

$$u_n=P_1(n)\lambda_1^n+P_2(n)\lambda_2^n+\cdots+P_r(n)\lambda_r^n,\qquad \deg P_i<m_i$$

的数列，并且多项式 $P_1,\ldots,P_r$ 由 (u_n) 唯一确定。

证明概要　由引理 16.5，$U=R/Q$，$Q(z)=\prod_i(1-\lambda_iz)^{m_i}$，$\deg R<k$。部分分式把 U 写成 $\sum_i\sum_{j=1}^{m_i}\frac{c_{ij}}{(1-\lambda_iz)^j}$，展开后 $u_n=\sum_i\bigl(\sum_j c_{ij}\binom{n+j-1}{j-1}\bigr)\lambda_i^n$，括号里是次数 $<m_i$ 的多项式。反过来，$\binom{n}{0},\binom{n+1}{1},\ldots,\binom{n+m-1}{m-1}$ 构成次数 $<m$ 的多项式的一组基，所以每个这种形式的数列都对应一个分母为 Q 的真分式，由引理 16.5 满足递推。两边都是 k 维空间，唯一性随之成立。 ■

■ **例 16.7**　**三重根**　求 $u_n=3u_{n-1}-3u_{n-2}+u_{n-3}$，$u_0=0$，$u_1=1$，$u_2=3$ 的通项。

解　$\chi(T)=T^3-3T^2+3T-1=(T-1)^3$，$1$ 是三重根。由定理 16.7，$u_n=(\alpha+\beta n+\gamma n^2)\cdot 1^n=\alpha+\beta n+\gamma n^2$。代入初值：$\alpha=0$，$\beta+\gamma=1$，$2\beta+4\gamma=3$。解得 $\gamma=\frac{1}{2}$，$\beta=\frac{1}{2}$，所以

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

即三角数 $0,1,3,6,10,15,\ldots$。核对：$u_3=3\cdot 3-3\cdot 1+0=6$，$u_4=3\cdot 6-3\cdot 3+1=10$。

这个结果有直观解释。递推可以改写成 $(u_n-u_{n-1})-2(u_{n-1}-u_{n-2})+(u_{n-2}-u_{n-3})=0$，也就是说三阶差分为零，而三阶差分为零的数列恰好是 n 的二次多项式，正如三阶导数为零的函数是二次多项式。 ♦

易错提醒　重根时只写 $\alpha\lambda^n+\beta\lambda^n$ 是最常见的错误：这两项可以合并，实际只有一个参数。重数为 m 的根必须配上次数 $<m$ 的多项式 $P(n)$，即 $\lambda^n, n\lambda^n, \dots, n^{m-1}\lambda^n$ 共 m 个解。全部重数加起来恰好是 k，参数个数才与初值个数相等。

16.3.2　非齐次递推

现在考虑右边多一项 $g(n)$ 的递推

$$u_n + a_1 u_{n-1} + \cdots + a_k u_{n-k} = g(n) \qquad (n \ge k). \tag{16.3}$$

命题 16.8（通解 = 特解 + 齐次通解）　设 (p_n) 是 (16.3) 的某一个解（称为**特解**（particular solution））。则 (16.3) 的全部解是 $u_n = p_n + h_n$，其中 (h_n) 取遍对应齐次递推（右边为 0）的解。

证明　若 u, p 都满足 (16.3)，两式相减，$h = u - p$ 满足右边为 $g(n)-g(n)=0$ 的递推。反之 $p+h$ 代入左边得 $g(n)+0=g(n)$。 ∎

剩下的问题是找特解。对多项式、指数这类 $g(n)$，用**待定系数法**：猜一个与 g 同类型的 p_n，代入后比较系数。

■ **例 16.8　多项式非齐次项**　求 $u_n - 5u_{n-1} + 6u_{n-2} = n^2$（$n \ge 2$），$u_0 = u_1 = 0$ 的通项。

解　**特解**。猜 $p_n = c_2 n^2 + c_1 n + c_0$。因为 $(n-1)^2 = n^2 - 2n + 1$，$(n-2)^2 = n^2 - 4n + 4$，

$$p_n - 5p_{n-1} + 6p_{n-2} = 2c_2\, n^2 + (2c_1 - 14c_2)\, n + (2c_0 - 7c_1 + 19c_2).$$

（n^2 的系数是 $c_2(1-5+6)$；n 的系数是 $c_2(10-24)+c_1(1-5+6)$；常数项是 $c_2(-5+24)+c_1(5-12)+c_0(1-5+6)$。）令它等于 n^2：

$$2c_2 = 1, \qquad 2c_1 - 14c_2 = 0, \qquad 2c_0 - 7c_1 + 19c_2 = 0,$$

依次得 $c_2 = \tfrac12$，$c_1 = \tfrac72$，$c_0 = \tfrac12\left(\tfrac{49}{2} - \tfrac{19}{2}\right) = \tfrac{15}{2}$。所以 $p_n = \frac{n^2+7n+15}{2}$。

通解与初值。齐次部分的特征根是 $2, 3$（例 16.1），所以 $u_n = \alpha 2^n + \beta 3^n + \frac{n^2+7n+15}{2}$。代入初值：

$$n=0:\ \alpha+\beta+\tfrac{15}{2}=0, \qquad n=1:\ 2\alpha+3\beta+\tfrac{23}{2}=0.$$

第二式减去第一式的 2 倍：$\beta - \tfrac72 = 0$，$\beta = \tfrac72$，$\alpha = -11$。所以

$$u_n = -11\cdot 2^n + \frac{7}{2}\cdot 3^n + \frac{n^2+7n+15}{2}.$$

核对：递推给出 $u_2 = 0-0+4 = 4$，$u_3 = 5\cdot 4 - 0 + 9 = 29$，$u_4 = 5\cdot 29 - 6\cdot 4 + 16 = 137$。公式给出 $u_2 = -44 + \tfrac{63}{2} + \tfrac{33}{2} = 4$，$u_3 = -88 + \tfrac{189}{2} + \tfrac{45}{2} = 29$。 ♦

■ **例 16.9　共振**　求 $u_n - 5u_{n-1} + 6u_{n-2} = 3^n$（$n \ge 2$），$u_0 = u_1 = 0$ 的通项。

解　先试 $p_n = A\cdot 3^n$。代入左边得 $A\cdot 3^{n-2}(9-15+6) = A\cdot 3^{n-2}\chi(3) = 0$，无论 A 取什么都等于不了 3^n。原因是 3 本身是特征根，$A\cdot 3^n$ 是齐次解。

改试 $p_n = Cn\cdot 3^n$。代入左边：

$$C\,3^{n-2}\big(9n - 15(n-1) + 6(n-2)\big) = C\,3^{n-2}\big(n(9-15+6) + (15-12)\big) = C\cdot 3^{n-1}.$$

令 $C\cdot 3^{n-1} = 3^n$，得 $C = 3$，特解 $p_n = n\cdot 3^{n+1}$。

初值。$u_n = \alpha 2^n + \beta 3^n + n\cdot 3^{n+1}$。$n=0$：$\alpha+\beta = 0$；$n=1$：$2\alpha+3\beta+9 = 0$。解得 $\beta = -9$，$\alpha = 9$，于是

$$u_n = 9\cdot 2^n + (n-3)\,3^{n+1}.$$

核对：递推给出 $u_2 = 9, u_3 = 5\cdot 9+27 = 72, u_4 = 5\cdot 72 - 6\cdot 9 + 81 = 387$；公式给出 $u_2 = 36-27 = 9$，$u_3 = 72+0 = 72$，$u_4 = 144+243 = 387$。 ♦

易错提醒　若非齐次项 $g(n) = c\,\lambda^n$ 中的 λ 是特征根，猜 $A\lambda^n$ 必然失败（代入后得零）。这时要乘以 n：若 λ 是 m 重根，就猜 $An^m\lambda^n$。同样，若 $g(n)$ 是多项式而 1 是 m 重特征根，多项式的猜测要整体乘以 n^m。例 16.8 中 $\chi(1) = 2 \neq 0$，所以二次多项式就够了。

学过 18.03（常微分方程）的读者会发现，本节与常系数线性常微分方程的理论几乎逐条对应（表 16.1）。原因是「乘以 z 平移系数」和「求导」都是线性算子，而 λ^n 与 e^{rt} 分别是平移和求导的「特征向量」。

表 16.1　线性递推与常系数线性常微分方程的对照

	递推 $u_n + au_{n-1} + bu_{n-2} = g(n)$	微分方程 $y'' + ay' + by = g(t)$
特征方程	$\lambda^2 + a\lambda + b = 0$	$r^2 + ar + b = 0$
基本解	λ^n	e^{rt}
二重根	$\lambda^n,\ n\lambda^n$	$e^{rt},\ te^{rt}$
初值	u_0, u_1	$y(0), y'(0)$
通解结构	特解 + 齐次通解	特解 + 齐次通解
共振	$g = \lambda^n$，$\lambda \neq 0$ 是 m 重特征根，猜 $An^m\lambda^n$	$g = e^{rt}$，r 是 m 重特征根，猜 $At^m e^{rt}$
齐次衰减	$g = 0$ 且所有 $\lvert\lambda\rvert < 1$ 时 $u_n \to 0$	$g = 0$ 且所有 $\operatorname{Re} r < 0$ 时 $y \to 0$

表 16.1 的衰减条件只控制齐次部分。非齐次项会持续改变解：例如 $u_n - \frac{1}{2}u_{n-1} - \frac{1}{4}u_{n-2} = 1$ 的两个特征根为 $(1\pm\sqrt{5})/4$，绝对值都小于 1；但取 $u_0 = u_1 = 4$，每一步都是 $4-2-1 = 1$，解恒为 4。有驱动项时，应另外求特解并检查其极限。

看读数列　线性递推会出现在意想不到的地方。从 1 出发，每一项都「读出」上一项：1 读作「一个 1」，得 11；11 读作「两个 1」，得 21；接着是 1211，111221，312211，13112221，…。这叫**看读数列**（look-and-say sequence）。J. H. Conway 在 1986 年证明：任何一项经过足够多步后，都会分裂成 92 种互不干扰的「元素」（他用化学元素给它们命名）。因此各项的长度 $1, 2, 2, 4, 6, 6, 8, 10, 14, 20, 26, \ldots$ 从某项起满足一个 71 阶常系数线性递推，相邻两项长度之比趋于这个递推特征多项式的最大实根，即 Conway 常数 ≈ 1.3036。

想一想　从 1 出发的看读数列中，为什么永远不会出现数字 4?

答　每一项（第一项除外）都是一串「个数数字」对 $c_1 d_1 c_2 d_2 \cdots$，并且相邻的对数字不同：$d_i \neq d_{i+1}$，否则它们在读的时候会被并成一段。若某一项第一次出现 4，它只能是某个个数 c_i（数字 d_i 来自上一项，上一项没有 4），即上一项有四个相同的数字连在一起。但上一项的任意连续四个位置要么是 $c_i d_i c_{i+1} d_{i+1}$，要么是 $d_i c_{i+1} d_{i+1} c_{i+2}$，两种情形都包含 d_i 和 d_{i+1}，而它们不相等。所以 4 永远不出现，只有 1, 2, 3 三个数字。

16.4　斐波那契数的数论性质

Binet 公式不仅能算数，还能把关于 F_n 的恒等式变成关于 λ, μ 的代数运算。本节始终记

$$\lambda = \frac{1+\sqrt{5}}{2}, \qquad \mu = \frac{1-\sqrt{5}}{2}, \qquad \lambda + \mu = 1, \qquad \lambda\mu = -1, \qquad F_n = \frac{\lambda^n - \mu^n}{\sqrt{5}}.$$

定理 16.9（Cassini 恒等式）　对所有 $n \geq 1$，

$$F_{n+1}F_{n-1} - F_n^2 = (-1)^n.$$

证明　**方法一（Binet）**。乘出

$$5F_{n+1}F_{n-1} = \lambda^{2n} + \mu^{2n} - (\lambda\mu)^{n-1}(\lambda^2 + \mu^2),$$

而 $5F_n^2 = \lambda^{2n} + \mu^{2n} - 2(\lambda\mu)^n$。相减，并用 $\lambda\mu = -1$：

$$5\big(F_{n+1}F_{n-1} - F_n^2\big) = (\lambda\mu)^{n-1}\big(2\lambda\mu - \lambda^2 - \mu^2\big) = -(-1)^{n-1}(\lambda-\mu)^2 = (-1)^n \cdot 5.$$

方法二（矩阵）。对 n 归纳可得 $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^n = \begin{pmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{pmatrix}$。两边取行列式，左边是 $(-1)^n$，右边是 $F_{n+1}F_{n-1} - F_n^2$。 ∎

核对：$n = 6$ 时 $F_7F_5 - F_6^2 = 13 \cdot 5 - 64 = 1$；$n = 7$ 时 $F_8F_6 - F_7^2 = 21 \cdot 8 - 169 = -1$。Cassini 恒等式有两个直接推论。其一，$F_n$ 与 F_{n+1} 的公因数整除 $F_{n+1}F_{n-1} - F_n^2 = \pm 1$，所以 $\gcd(F_n, F_{n+1}) = 1$（第 1 讲的练习用归纳法证过）。其二，取 $n = 2k$，得

$$F_{2k+1}F_{2k-1} = F_{2k}^2 + 1,$$

例如 $13 \cdot 5 = 8^2 + 1$，$34 \cdot 13 = 21^2 + 1$。这个「多出来的 1」就是一个著名谜题的谜底。

■ **例 16.10**　$64 = 65$?　把 8×8 的正方形按图 16.3 左边剪成四块（两个直角梯形、两个直角三角形），重新拼成右边 13×5 的长方形。面积从 64 变成了 65，多出的 1 从哪里来？

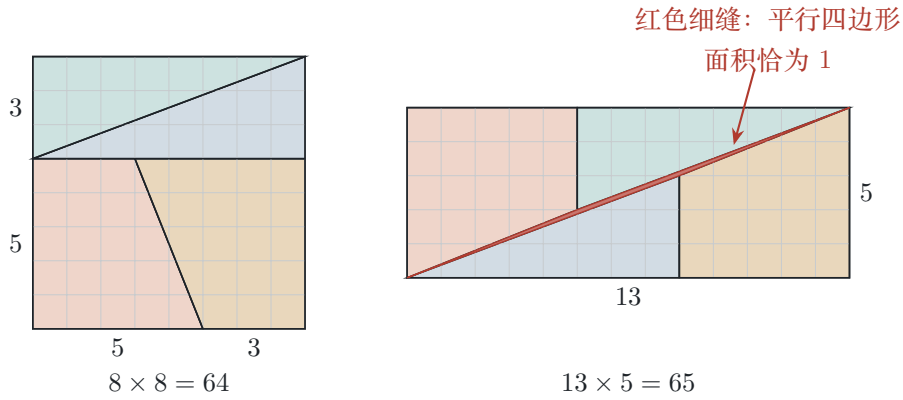


图 16.3　$64 = 65$ 谜题。左边正方形的边长是 $F_6 = 8$，切口用到 $F_5 = 5$ 和 $F_4 = 3$；右边长方形是 $F_7 \times F_5 = 13 \times 5$。长方形的「对角线」其实是两条折线，中间夹着一条面积为 1 的细缝。

解　四块的面积都没变，变的是「拼得严不严」。看右图从 $(0,0)$ 到 $(13,5)$ 的「对角线」：下方两块的斜边依次是三角形的斜边（斜率 $\frac{3}{8} = 0.375$）和梯形的斜边（斜率 $\frac{2}{5} = 0.4$）；上方两块是梯形的斜边（斜率 $\frac{2}{5}$）和三角形的斜边（斜率 $\frac{3}{8}$）。真正的对角线斜率是 $\frac{5}{13} \approx 0.385$。三个斜率不相等，所以两边都是折线：下方折线经过 $(8,3)$，上方折线经过 $(5,2)$，二者之间夹着以 $(0,0),(8,3),(13,5),(5,2)$ 为顶点的平行四边形。它由向量 $(8,3)$ 和 $(5,2)$ 张成，面积是

$$|8 \cdot 2 - 3 \cdot 5| = 1.$$

所以 $65 = 64 + 1$，多出的 1 就是这条细缝。上下边界的竖直间距最大只有 $\frac{1}{8}$ 格：例如在 $x = 5$ 处，上边界高为 2，下边界高为 $\frac{3}{8} \cdot 5 = \frac{15}{8}$，相差 $\frac{1}{8}$。这个间距很小，容易被误看成拼合的边线。

这里的数都是斐波那契数，$13 \cdot 5 - 8^2 = 1$ 正是 Cassini 恒等式。换成 $F_{2k-1}, F_{2k}, F_{2k+1}$ 可以造出更大、更难察觉的版本，细缝面积永远是 1。 ♦

16.4.1　加法公式和整除性

Cassini 恒等式说明相邻 Fibonacci 数互素。本节再推导一个更灵活的公式，它可以把指标相加，也能直接回答「什么时候 F_m 整除 F_n」。

定理 16.10（斐波那契加法公式）　对 $m \geq 1$、$n \geq 0$，

$$F_{m+n} = F_{m-1}F_n + F_m F_{n+1}.$$

证明　固定 m，令 $G_n = F_{m-1}F_n + F_m F_{n+1}$。因为 F_n 和 F_{n+1} 都满足 Fibonacci 递推，G_n 也满足 $G_n = G_{n-1} + G_{n-2}$（$n \geq 2$）。初值为

$$G_0 = F_m, \qquad G_1 = F_{m-1} + F_m = F_{m+1}.$$

而数列 F_{m+n} 的前两项也正是 F_m, F_{m+1}。初值决定递推数列，所以两列相同。 ■

■ **例 16.11**　**一条加法公式的数值核对**　用定理 16.10 计算 F_{10}，并用递推逐项核对。

解　取 $m=6, n=4$，则

$$F_{10} = F_5F_4 + F_6F_5 = 5\cdot 3 + 8\cdot 5 = 55.$$

递推给出 $F_0, F_1, \dots, F_{10} = 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55$，结果一致。公式把相距较远的项写成较小指标 Fibonacci 数的乘积和。 ♦

定理 16.11（指标整除推出 Fibonacci 数整除）　若 m, n 是正整数且 $m \mid n$，则 $F_m \mid F_n$。

证明　写 $n = mk$，对 k 归纳。$k=1$ 时 $n=m$，所以 F_m 整除自身。设 $k \geq 2$ 且 $F_m \mid F_{m(k-1)}$。由加法公式，

$$F_{mk} = F_{m(k-1)+m} = F_{m(k-1)-1}F_m + F_{m(k-1)}F_{m+1}.$$

右边两项都被 F_m 整除，故 $F_m \mid F_{mk}$。 ■

习题集 7 第 6 题把同一递推结构用在计数上。一个正整数的有序拆分（composition）是把它写成若干正整数之和，并保留加数的顺序；例如 $5 = 1+3+1$ 与 $5 = 3+1+1$ 是不同拆分。现在要求每个加数都是奇数。

定理 16.12（奇数拆分数是 Fibonacci 数）　设 c_n 是 n 写成有序奇数正整数之和的方法数。则 $c_1 = c_2 = 1$，且对 $n \geq 3$，

$$c_n = c_{n-1} + c_{n-2} = F_n.$$

证明　按一个拆分的第一个加数分类。若第一个加数是 1，删去它，得到 $n-1$ 的奇数拆分；若第一个加数至少是 3，从它减去 2，得到 $n-2$ 的奇数拆分。两种操作都可逆，故方法数相加。$n = 1, 2$ 时只有 1 和 $1+1$ 两种基例；它们与 $F_1 = F_2 = 1$ 相同，递推遂给出 $c_n = F_n$。 ■

■ **例 16.12**　**拆分 7**　求 7 拆成有序奇数加数的方式数，并列出全部方式。

解　按定理，方式数为 $F_7 = 13$。按加数个数分类：一个加数时是 7；三个加数时有

$$5+1+1,\ 1+5+1,\ 1+1+5,\ 3+3+1,\ 3+1+3,\ 1+3+3;$$

五个加数时，把唯一的 3 放在五个位置之一，有 5 种；七个加数时只有 $1+1+1+1+1+1+1$。合计 $1+6+5+1 = 13$。 ♦

16.5　斐波那契数模素数

前面的恒等式只涉及整数等式；把 Binet 公式放到模素数的环境中，还能读出 F_p、F_{p-1} 和 F_{p+1} 的整除性。这里 p 始终表示大于 5 的素数，因而 2、5 在模 p 下都有逆元。

定理 16.13（素数指标的 Fibonacci 同余）　若 $p>5$ 是素数，则

$$F_p \equiv \left(\frac{5}{p}\right) \pmod p.$$

特别地，$F_p \equiv 1 \pmod p$ 当 $p \equiv \pm 1 \pmod 5$，而 $F_p \equiv -1 \pmod p$ 当 $p \equiv \pm 2 \pmod 5$。

思路　把 $F_p = (\lambda^p - \mu^p)/\sqrt{5}$ 中的两个 p 次幂展开。模 p 时，中间的二项式系数都消失，只留下首项和最高次项；后者经 Euler 判别法变成 Legendre 符号。

证明　将 $\lambda = (1+\sqrt{5})/2$、$\mu = (1-\sqrt{5})/2$ 代入 Binet 公式并展开：

$$F_p = \frac{1}{2^{p-1}}\left(\binom{p}{1} + \binom{p}{3}5 + \cdots + 5^{(p-1)/2}\right).$$

当 $1<j<p$ 时，$p \mid \binom{p}{j}$，所以模 p 只剩最后一项。由费马小定理 $2^{p-1} \equiv 1$，由 Euler 判别法 $5^{(p-1)/2} \equiv \left(\frac{5}{p}\right)$，得到第一式。

因为 $5 \equiv 1 \pmod 4$，二次互反律给出 $\left(\frac{5}{p}\right) = \left(\frac{p}{5}\right)$。模 5 的非零平方类是 $1, 4$，故第二个结论成立。 ■

定理 16.14（相邻两项中恰有一项被 p 整除）　若 $p>5$ 是素数，则 p 恰好整除 F_{p-1} 和 F_{p+1} 中的一个。更具体地，

$p \pmod 5$	F_{p-1}	F_{p+1}
± 1	0	1
± 2	1	0

$(\mathrm{mod} p).$

证明　展开 F_{p+1} 时，模 p 只有二项式展开中的首项和最高次项留下：

$$F_{p+1} \equiv 2^{-p}\left(1 + 5^{(p-1)/2}\right) \equiv \frac{1+\left(\frac{5}{p}\right)}{2} \pmod p.$$

定理 16.13 给出：当 $p \equiv \pm 1 \pmod 5$ 时，$F_p \equiv 1$、$F_{p+1} \equiv 1$；当 $p \equiv \pm 2 \pmod 5$ 时，$F_p \equiv -1$、$F_{p+1} \equiv 0$。由于 $F_{p+1} = F_p + F_{p-1}$，相减即得表格。 ■

小素数需要单独看：$p=2$ 时 $2 \mid F_3$，$p=3$ 时 $3 \mid F_4$；而 $p=5$ 时 $F_4 = 3, F_6 = 8$，两项都不被 5 整除。因此「恰有一项」的整洁结论从 $p>5$ 开始，不能把 $p=5$ 包含进去。

■ **例 16.13　模 11 的周期片段**　验证 11 整除 F_{10}，并检查 $F_{n+10} \equiv F_n \pmod{11}$ 对 $n = 0, 1, 2, 3$ 成立。

解　$11 \equiv 1 \pmod 5$，所以定理 16.14 预言 $11 \mid F_{10}$ 且 $F_{11} \equiv 1$。数值上 $F_9 = 34 \equiv 1$、$F_{10} = 55 \equiv 0$、$F_{11} = 89 \equiv 1$。用加法公式取 $m = 10$ 得

$$F_{n+10} = F_9 F_n + F_{10} F_{n+1} \equiv F_n \pmod{11}.$$

例如 $F_{10}, F_{11}, F_{12}, F_{13}$ 模 11 分别为 $0, 1, 1, 2$，与 F_0, F_1, F_2, F_3 相同。 ♦

推论 16.15（一个 Fibonacci 周期） 若 $p > 5$ 且 $p \equiv \pm 1 \pmod 5$，则 $p - 1$ 是 Fibonacci 数列模 p 的一个周期：

$$F_{n+p-1} \equiv F_n \pmod p \qquad (n \geq 0).$$

证明 在定理 16.14 的第一种情形中，$F_{p-1} \equiv 0$、$F_p \equiv 1$。对加法公式取 $m = n$、第二个指标为 $p - 1$：

$$F_{n+p-1} = F_{n-1}F_{p-1} + F_nF_p \equiv F_n \pmod p.$$

当 $n = 0$ 时，结论直接是 $F_{p-1} \equiv F_0 = 0$；所以所有 $n \geq 0$ 都成立。 ■

易错提醒 Cassini 恒等式与 $F_p \equiv \pm 1$ 先给出 $p \mid F_{p-1}F_{p+1}$，这只能说明「至少一项被 p 整除」。若两项都被 p 整除，递推 $F_p = F_{p+1} - F_{p-1}$ 会推出 $p \mid F_p$，与 $F_p \equiv \pm 1 \pmod p$ 矛盾；这一步才排除了「两项都整除」。

16.6　共轭根与整除性

Binet 公式中 μ 的绝对值小于 1，所以 Fibonacci 数是一个实数幂的最近整数。类似地，若一个代数数和它的共轭之和是整数，就可以把 floor 符号换成这个整数，再研究它的递推和整除性。

■ **例 16.14　一个共轭根的 floor 问题**　证明对每个 $n \geq 0$，

$$2^{n+1} \mid \lfloor (1+\sqrt{3})^{2n+1} \rfloor.$$

解　令 $\alpha = 1 + \sqrt{3}$、$\beta = 1 - \sqrt{3}$，并记

$$S_k = \alpha^k + \beta^k.$$

因为 $\alpha + \beta = 2$、$\alpha\beta = -2$，有

$$S_0 = 2, \qquad S_1 = 2, \qquad S_k = 2S_{k-1} + 2S_{k-2}.$$

所以每个 S_k 都是整数。又 $-1 < \beta < 0$，对奇数 $2n+1$ 有 $-1 < \beta^{2n+1} < 0$，从而

$$\lfloor \alpha^{2n+1} \rfloor = \alpha^{2n+1} + \beta^{2n+1} = S_{2n+1}.$$

令 $T_n = S_{2n+1}$。数 $\alpha^2 = 4 + 2\sqrt{3}$、$\beta^2 = 4 - 2\sqrt{3}$ 的和为 8、积为 4，因此

$$T_0 = 2, \qquad T_1 = S_3 = 20, \qquad T_n = 8T_{n-1} - 4T_{n-2}.$$

归纳证明 $2^{n+1} \mid T_n$：T_0 被 2 整除，T_1 被 4 整除；若前两项分别被 2^n、2^{n-1} 整除，则 $8T_{n-1}$ 和 $4T_{n-2}$ 都被 2^{n+1} 整除。于是结论成立。 ♦

想一想　为什么这里要取奇数次幂？如果改成偶数次幂，$\lfloor \alpha^{2n} \rfloor$ 还能直接等于共轭根之和吗？

答　β 为负数。奇数次幂时 $-1 < \beta^{2n+1} < 0$，所以 $\alpha^{2n+1} = S_{2n+1} - \beta^{2n+1}$ 位于整数 S_{2n+1} 和 $S_{2n+1}+1$ 之间，floor 正好是 S_{2n+1}。偶数次幂且 $n \geq 1$ 时，$0 < \beta^{2n} < 1$，此时 $\alpha^{2n} = S_{2n} - \beta^{2n}$，floor 是 $S_{2n} - 1$。边界 $n = 0$ 也满足 $\lfloor \alpha^0 \rfloor = 1 = S_0 - 1$，但此时 $\beta^0 = 1$，不适用严格不等式。

本讲小结

- 互异特征根 λ_i 给出通项 $u_n = \sum_i c_i \lambda_i^n$；重数为 m 的根贡献 $P(n)\lambda^n$，其中 $\deg P < m$。
- k 阶齐次常系数递推的生成函数为 $U(z) = R(z)/(1 + a_1 z + \cdots + a_k z^k)$，且 $\deg R < k$。
- 非齐次递推的通解是一个特解加齐次通解；共振时试探解乘以 n 的幂。
- $F_{m+n} = F_{m-1}F_n + F_m F_{n+1}$；若 $m \mid n$，则 $F_m \mid F_n$。
- 对素数 $p > 5$，$F_p \equiv \left(\frac{5}{p}\right) \pmod p$；$p \mid F_{p-1}F_{p+1}$，且恰好一项被 p 整除。
- 若 $p > 5$ 且 $p \equiv \pm 1 \pmod 5$，则 $p - 1$ 是 Fibonacci 数列模 p 的一个周期。

练习

16.1 ★　递推 $u_n = 2u_{n-1} + 3u_{n-2}$，$u_0 = 1, u_1 = 2$。求生成函数，并算出 u_0, u_1, u_2, u_3。

16.2 ★　求 $u_n = 4u_{n-1} - 3u_{n-2}$，$u_0 = 0, u_1 = 2$ 的通项。

16.3 ★　将 7 写成有序奇数正整数之和有多少种？按加数个数分类核对。

16.4 ★★　求 $u_n = 6u_{n-1} - 9u_{n-2}$，$u_0 = 2, u_1 = 9$ 的通项，并指出根的重数如何反映在答案中。

16.5 ★★　用指标整除定理说明 $F_4 \mid F_{12}$，并计算商。

16.6 ★★　对 $p = 7$，计算 F_p, F_{p-1}, F_{p+1} 模 p 的余数，并核对定理 16.14。

16.7 ★★　「若 $F_m \mid F_n$，则 $m \mid n$」是否成立？给出反例或证明。

16.8 ★★　令 $A_n = \lfloor (1+\sqrt{3})^{2n+1} \rfloor$。计算 $A_0, \ldots, A_4$ 及 $A_n / 2^{n+1}$，并核对它们都是整数。

16.9 ★★★　证明 Fibonacci 数列模 11 的最小正周期是 10，而不只是一个周期为 10。

练习答案与提示

16.1 $U(z) = \frac{1}{1-2z-3z^2} = \frac{3/4}{1-3z} + \frac{1/4}{1+z}$，故 $u_n = \frac{3}{4}\,3^n + \frac{1}{4}(-1)^n$；前四项为 $1, 2, 7, 20$。

16.2 $\chi(T) = (T-1)(T-3)$，故 $u_n = A + B3^n$；初值给 $A = -1, B = 1$，所以 $u_n = 3^n - 1$。

16.3 方式数为 $F_7 = 13$：一个加数有 1 种；三个加数有 $(5,1,1)$ 和 $(3,3,1)$ 的各 3 种排列；五个加数有 5 种；七个加数有 1 种。合计 $1+3+3+5+1 = 13$。

16.4 $\chi(T) = (T-3)^2$，故 $u_n = (A+Bn)3^n$。初值给 $A = 2$、$3(A+B) = 9$，所以 $B = 1$，即 $u_n = (n+2)3^n$；因子 n 对应二重根。

16.5 $4 \mid 12$，所以 $F_4 \mid F_{12}$。$F_4 = 3, F_{12} = 144$，商为 48。

16.6 $7 \equiv 2 \pmod 5$，所以 $(5/7) = -1$；$F_7 = 13 \equiv 6 = -1$，$F_6 = 8 \equiv 1$，$F_8 = 21 \equiv 0 \pmod 7$，与表格一致。

16.7 不成立：$F_2 = 1 \mid F_3 = 2$，但 $2 \nmid 3$。

16.8 递推给出 $T_2 = 152, T_3 = 1136, T_4 = 8480$，故 $A_0, \ldots, A_4 = 2, 20, 152, 1136, 8480$。除以 $2, 4, 8, 16, 32$ 分别得 $1, 5, 19, 71, 265$。

16.9 若最小周期为 d，写 $10 = qd + r$、$0 \le r < d$；周期相减知 r 也是周期，故 $d \mid 10$。其真因数 $1, 2, 5$ 均不可能，因为 $F_1, F_2, F_5 \not\equiv 0 \pmod{11}$，所以最小周期是 10。

复习课 2

平方剩余、算术函数与递推

Quadratic Residues, Arithmetic Functions, and Recurrences

对应材料：MIT 18.781 第二次期中考试复习指南、模拟题及解答；习题集 5–7；本讲义第 9–16 讲。模拟题编号以题目文件为准，解答文件后半部分的编号有错位。

本讲目标

(1) 能区分「平方是否存在」「求出平方根」「确定根的个数」三种问题，并检查素数与互素条件。

(2) 能把因数求和写成 Dirichlet 卷积，再用素数幂分解或 Möbius 反演计算。

(3) 能从不交分类建立递推，结合初值写出特征根通项与有理生成函数。

(4) 能完成连接二次互反律、分圆多项式、算术函数与 Fibonacci 同余的综合题，并识别来源材料中的下标笔误。

R2.1 知识地图与选用工具

$3x^2+4x-2\equiv 0 \pmod{31}$、$\sum_{d\mid n}\mu(d)\varphi(d)$、$u_n=5u_{n-1}-6u_{n-2}$ 看似分属三类题，但做题的共同步骤很具体：先辨认结构，检查条件，再选能保留该结构的运算。平方问题先配方；因数求和先看卷积；递推先写初值和特征多项式。

图 R2.1 分三行整理这段课程。灰色实线表示同一主题内的推导，橙色虚线表示不同主题之间的应用关系。图上的联系不能代替条件检查，例如从分圆值的素因子推出阶，仍需排除整除分圆指标的素数。

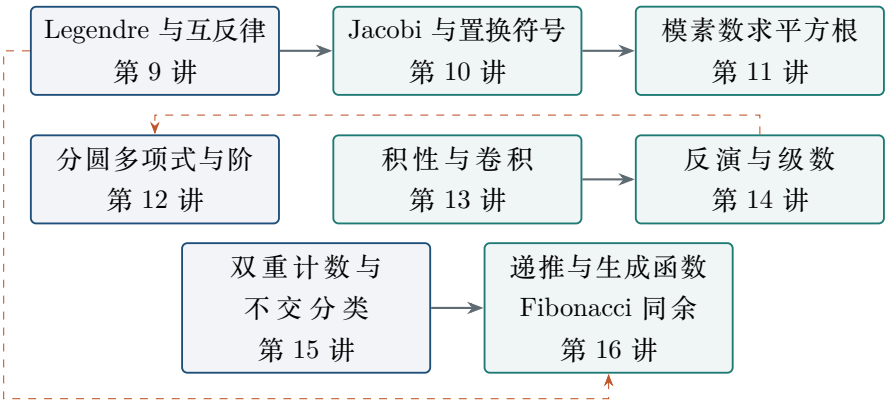


图 R2.1 互反律连接平方判别与 Fibonacci 同余，Möbius 反演也能用于分圆乘积；计数分类产生递推，递推对应生成函数。

复习时可以遮住下表的中间一列，先写出定义和条件，再看最右边的用法。若只能记得公式，不能说清对象所在的数系，计算很容易在第一步就用错工具。

概念	定义或条件	遇到什么问题时使用
Legendre 符号	分母是奇素数；值为 $0, 1, -1$	二次同余的存在性、根数与互反计算
Jacobi 符号	分母为正奇数；按素因子相乘	快速符号计算与乘法置换的奇偶性
模平方根	先判断剩余，再用幂或 Tonelli 修正	从「有解」继续求出全部根
分圆多项式	Φ_n 的根恰为本原 n 次单位根	因式分解、阶与素因子的同余条件
Dirichlet 卷积	$(f*g)(n)=\sum_{d\mid n} f(d)g(n/d)$	因数和、积性与反演
Möbius 反演	$F=f*U$ 当且仅当 $f=F*\mu$	由累计因数和恢复原函数
不交分类	每个对象恰落入一类	建立计数递推，不重复也不漏计
普通生成函数	$A(z)=\sum_{n\ge 0} a_n z^n$	把下标移位与普通卷积变成代数运算

R2.2　核心结论与适用条件

R2.2.1　平方判别、互反律和置换（第 9–10 讲）

定理（素数模的平方判别与互反律）　设 p 是奇素数，$a\in\mathbb{Z}$。$x^2\equiv a \pmod p$ 的解数为 $1+\left(\frac{a}{p}\right)$；$p\nmid a$ 时 Euler 判别法给出 $a^{(p-1)/2}\equiv\left(\frac{a}{p}\right) \pmod p$。若 p,q 是不同奇素数，则

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)=(-1)^{(p-1)(q-1)/4},\quad \left(\frac{-1}{p}\right)=(-1)^{(p-1)/2},\quad \left(\frac{2}{p}\right)=(-1)^{(p^2-1)/8}.$$

关键　平方映射把非零剩余类两两配对；Gauss 引理把 Legendre 符号化为负代表元的个数，格点计数给出互反律。完整证明见第 9 讲。

用法　先约化分子，分出 -1 和 2 的幂，再翻转奇素数因子。只有双方都余 3 模 4 时翻转变号。求一般二次式 ax^2+bx+c 的根之前，先确认 $p\nmid 2a$，再转成 $(2ax+b)^2\equiv b^2-4ac$。

命题（Jacobi 符号与 Zolotareff 定理）　设 n 为正奇数，$\gcd(a,n)=1$。Jacobi 符号按素因数分解定义，并满足与上面相同的两个补充律；对互素的正奇数 m,n，也满足相同形式的奇数互反律。若 a 是模 n 的平方，则 $\left(\frac{a}{n}\right)=1$；反向一般不成立。乘法置换 $M_a: x\mapsto ax$ 在模 n 的全部剩余类上满足

$$\operatorname{sgn}(M_a)=\left(\frac{a}{n}\right).$$

关键与用法　Jacobi 的计算由分子、分母的乘法性归约到 Legendre 符号；置换的符号由不相交轮换相乘，长度 k 的轮换贡献 $(-1)^{k-1}$。完整证明见第 10 讲。用 Jacobi 算出的 1 只代表一个乘积的符号，求合数模平方根仍需分别检查各素数幂。

R2.2.2 求平方根与分圆根的阶（第 11–12 讲）

命题（从存在性走到实际平方根）　设 p 是奇素数，$p\nmid a$ 且 $\left(\frac{a}{p}\right)=1$。若 $p\equiv 3 \pmod 4$，则两个平方根为 $\pm a^{(p+1)/4}$。一般情形写 $p-1=2^s t$、t 奇，Tonelli 算法使用一个非剩余，保持不变式 $a\equiv Ab^2$，最终使 $A^t\equiv 1$；此时

$$x = bA^{(t+1)/2}, \qquad x^2 \equiv a \pmod p.$$

关键与用法　用阶为 2^s 的修正因子逐层消去 A 的二幂部分，最后由 $x^2\equiv Ab^2A^t\equiv a$ 验证。完整算法与证明见第 11 讲。非剩余输入应在开始时报告无解；$a\equiv 0$ 则另行给出唯一根 0。

另一条第 11 讲的联系是**开平方与分解**：对 $n=pq$，$p\neq q$ 为奇素数，若 $r^2\equiv s^2 \pmod n$ 且 $r\not\equiv \pm s$，并且二者都与 n 互素，那么 $\gcd(r-s,n)$ 与 $\gcd(r+s,n)$ 给出两个素因子。原因是每个素数模下只能有两个平方根，而两个符号在 p,q 处未同时一致或同时相反。

定理（分圆分解与阶判别）　对每个正整数 n，

$$x^n-1=\prod_{d\mid n}\Phi_d(x), \qquad \Phi_n(x)\in\mathbb{Z}[x], \qquad \deg\Phi_n=\varphi(n).$$

若 $n\geq 2$、p 为素数、$p\nmid n$，且 $p\mid\Phi_n(a)$，则 $\operatorname{ord}_p(a)=n$，因而 $p\equiv 1 \pmod n$。素数幂还满足 $\Phi_{p^k}(x)=(x^{p^k}-1)/(x^{p^{k-1}}-1)$。

关键　把所有 n 次单位根按它们的实际阶分类，得到分解；模 p 时，条件 $p\nmid n$ 保证 x^n-1 无重根，避免较低阶的根混入。完整证明见第 12 讲。

用法　计算 Φ_n 时递归移去真因子指标对应的 Φ_d，然后核对次数；讨论其值的素因子时，先检查该素因子是否整除 n，再使用阶判别。由阶判别构造新素因子，正是第 12 讲证明余 1 模 n 的素数有无穷多个的办法。

R2.2.3 因数求和、积性与反演（第 13–14 讲）

命题（卷积与积性函数）　记 $U(n)=1$，$\mathbf{1}(1)=1$、$\mathbf{1}(n)=0$（$n>1$）。Dirichlet 卷积定义为 $(f*g)(n)=\sum_{d\mid n}f(d)g(n/d)$，其单位元是 $\mathbf{1}$。积性函数的卷积仍积性，且积性函数由各个素数幂上的值决定。特别地，

$$d=U*U, \quad \sigma=r_1*U, \quad \varphi*U=r_1, \quad \mu*U=\mathbf{1}, \qquad r_1(n)=n.$$

关键与用法　当 $\gcd(m,n)=1$ 时，mn 的每个因数唯一写成 m 的因数乘 n 的因数，所以因数和能够分离。完整证明见第 13 讲。实际计算时先把 n 拆成不同素数的幂，不能把不互素因子随意拆开；先辨认 U 与 $\mathbf{1}$，也不能把卷积当作逐点乘法。

第 13 讲还给出：$\mu(n)=0$ 当 n 含平方因子，否则 $\mu(n)=(-1)^{\omega(n)}$；算术函数存在卷积逆当

且仅当 $f(1)\neq 0$。若积性函数 f 不恒为零，取 m 使 $f(m)\neq 0$，由 $f(m)=f(m\cdot 1)=f(m)f(1)$ 可约去 $f(m)$，所以 $f(1)=1$。偶完全数恰好为 $2^{m-1}(2^m-1)$，其中 2^m-1 是素数；这个结论只讨论偶完全数，不能据此排除奇完全数。

定理（Möbius 反演与 Dirichlet 级数）　对所有正整数 n，有

$$F(n)=\sum_{d\mid n} f(d) \quad\Longleftrightarrow\quad f(n)=\sum_{d\mid n}\mu(d)F(n/d).$$

令 $Z(f,s)=\sum_{n\geq 1} f(n)n^{-s}$。在两边绝对收敛的条件下，$Z(f*g,s)=Z(f,s)Z(g,s)$。若 f 积性、$f(1)=1$ 且级数绝对收敛，则

$$Z(f,s)=\prod_{p}\left(\sum_{e\geq 0} f(p^e)p^{-es}\right).$$

关键与用法　反演就是在 $F=f*U$ 两边卷积 μ；级数乘法则把二重和按乘积 $ab=n$ 分组。完整证明见第 14 讲。用到无穷级数重排时必须先检查绝对收敛；有限的因数和不需要这个分析条件。

特别地，$\varphi(n)=n\prod_{p\mid n}(1-1/p)$，$Z(\mu,s)=1/\zeta(s)$ $(\operatorname{Re} s>1)$。第 14 讲用 $\sum\mu(d)/d^2=6/\pi^2$ 计算平方自由数的密度和两随机正整数互素的极限比例。这里「随机」是先在 $1,\dots,N$ 中均匀取值，再令 $N\to\infty$；不能直接对全部正整数指定一个均匀分布。

R2.2.4 计数、递推与生成函数（第 15–16 讲）

命题（先分类，再使用计数公式）　第 15 讲的两种基本等式是

$$\sum_{d\mid n}\varphi(d)=n, \qquad \binom{2n}{n}=\sum_{k=0}^{n}\binom{n}{k}^2.$$

对有限集合 X 及其子集 $A_1,\dots,A_r$，容斥原理给出

$$\left|X\setminus\bigcup_i A_i\right|=\sum_{S\subseteq\{1,\dots,r\}}(-1)^{|S|}\left|\bigcap_{i\in S}A_i\right|,$$

空集指标对应的交集约定为 X。抽屉原理则说明 $N+1$ 个对象放进 N 个盒子时必有碰撞。

关键与用法　前两式分别按最大公因数和两排位置中的选取个数做不交分类；容斥按「同时落入哪些坏集合」纠正重计。完整证明见第 15 讲。抽屉题先写清对象与盒子，递推题先说明分类为什么不重不漏，不能靠前几项恰好相同来认定两个数列相等。

定理（常系数递推与普通生成函数）　设 k 阶齐次递推为 $u_n + a_1 u_{n-1} + \cdots + a_k u_{n-k} = 0$ $(n \geq k)$，其中 $a_k \neq 0$。给定前 k 项后解唯一，且

$$\sum_{n \geq 0} u_n z^n = \frac{R(z)}{1 + a_1 z + \cdots + a_k z^k}, \qquad \deg R < k.$$

若特征根 λ_i 的重数是 m_i，则通项为 $\sum_i P_i(n)\lambda_i^n$，其中 $\deg P_i < m_i$，各常数由初值唯一确定。

关键与用法　乘上生成函数的分母后，高次项由递推抵消，只留下初值；部分分式把每个根的重数转成 n 的多项式因子。完整证明见第 16 讲。这里可以按形式幂级数运算，不必先确定复变量 z 的收敛域；若要赋予具体数值，则需另行讨论收敛。

Fibonacci 数列取 $F_0 = 0, F_1 = 1$。当 $m \geq 1, n \geq 0$ 时，$F_{m+n} = F_{m-1}F_n + F_m F_{n+1}$；对正整数指标，$m \mid n \Rightarrow F_m \mid F_n$。对素数 $p > 5$，有 $F_p \equiv \left(\frac{5}{p}\right) \pmod p$，且 p 恰好整除 F_{p-1}, F_{p+1} 中的一项：$p \equiv \pm 1 \pmod 5$ 时整除前者，$p \equiv \pm 2 \pmod 5$ 时整除后者。这把二次剩余的符号与递推数列的下标联系起来。

R2.3　题型与完整例题

R2.3.1　题型一：先判别，再求出所有根

■ **例 R2.1　第二次期中模拟题第 3 题**　求同余式 $3x^2 + 4x - 2 \equiv 0 \pmod{31}$ 的全部解。

解　先配方，再求平方根。判别式 $D = 4^2 - 4 \cdot 3 \cdot (-2) = 40 \equiv 9 \pmod{31}$。由于 6 可逆，原同余式等价于

$$(6x + 4)^2 \equiv 9 \pmod{31}.$$

9 的两个根为 ± 3；也可按第 11 讲的 $31 \equiv 3 \pmod 4$ 公式求得 $9^8 \equiv 28 \equiv -3$。因此 $6x + 4 \equiv 3$ 或 -3。由 $6^{-1} \equiv 26$，得到

$$x \equiv 26(3 - 4) \equiv 5, \qquad x \equiv 26(-3 - 4) \equiv 4 \pmod{31}.$$

回代：$x = 4$ 时原式为 $48 + 16 - 2 = 62$，$x = 5$ 时为 $75 + 20 - 2 = 93$，都被 31 整除。配方是一一对应，平方根又只有两个，故全部解恰为 $x \equiv 4, 5 \pmod{31}$。 ♦

点评　例 R2.1 连接第 9 讲的判别式与第 11 讲的求根。看到一般二次式先查首项系数是否可逆；仅算 Legendre 符号只能回答有解与否，不能替代回代与完整性。若模数整除首项系数，应先将原式按一次或常数同余处理。

R2.3.2　题型二：把互反律整理成素数的同余类

■ **例 R2.2　第二次期中模拟题第 4 题**　确定所有使 15 是模 p 的平方的素数 p；允许平方等于 0。

解　先排除会让符号为零的特殊素数。$p = 2$ 时 $15 \equiv 1$；$p = 3, 5$ 时 $15 \equiv 0$，都为平方。以下设 $p > 5$，这时 $\gcd(p, 60) = 1$。

利用乘法性，把 15 分成 3 与 5。互反律给出

$$\left(\frac{15}{p}\right)=\left(\frac{3}{p}\right)\left(\frac{5}{p}\right),\qquad \left(\frac{3}{p}\right)=\begin{cases}1, & p\equiv 1,11 \pmod{12},\\ -1, & p\equiv 5,7 \pmod{12},\end{cases}$$

并且 $\left(\frac{5}{p}\right)=\left(\frac{p}{5}\right)$，故 $p\equiv 1,4 \pmod 5$ 时为 1，$p\equiv 2,3 \pmod 5$ 时为 -1。两个符号的乘积为 1，要求它们同号。枚举模 60 的 16 个可逆类，得到

$$p\equiv 1,7,11,17,43,49,53,59 \pmod{60},$$

也就是 $p\equiv \pm 1,\pm 7,\pm 11,\pm 17 \pmod{60}$。例如 $p\equiv 7$ 时两个符号都为 -1，乘积仍为 1；$p\equiv 13$ 时分别为 $1,-1$，所以不符合。加上特殊素数 $2,3,5$，便得到全部答案。 ♦

点评　例 R2.2 的计算依赖第 9 讲的互反律，整理时使用同余类的有限分类。先把整除分子的素数单独列出，避免把 Legendre 符号为零误写成 ± 1。只写出必要的模 60 条件还不够；这里逐类符号判别是等价计算，因此也保证充分性。

R2.3.3 题型三：分圆因子的次数与模素数的阶

■ **例 R2.3　第二次期中模拟题第 8 题及延伸核对**　计算 $\Phi_{12}(x)$，并由 $\Phi_{12}(2)$ 检查分圆根的阶判别。

解　按根的实际阶移去已有因子。先写

$$x^{12}-1=(x^6-1)(x^6+1),\qquad x^6+1=(x^2+1)(x^4-x^2+1).$$

x^6-1 已包含指标 $1,2,3,6$ 的分圆因子；剩下 x^6+1 对应的阶只有 $4,12$。其中 $x^2+1=\Phi_4(x)$，所以

$$\Phi_{12}(x)=x^4-x^2+1.$$

它首一、次数为 $4=\varphi(12)$，也与第 12 讲的一般次数一致。

代入 $x=2$，得 $\Phi_{12}(2)=16-4+1=13$。因为 $13\nmid 12$，阶判别保证 $\operatorname{ord}_{13}(2)=12$，于是 $13\equiv 1 \pmod{12}$。还可以直接核对：$2^{12}\equiv 1$、$2^6\equiv -1$、$2^4\equiv 3\not\equiv 1$。12 的真因子都整除 6 或 4，所以这些核对已排除了所有更小的可能阶。 ♦

点评　例 R2.3 先把单位根按阶分类，再用次数和实际幂核对结果。不能把每个 x^d-1 都直接除掉，否则共同含有的低阶因子会被反复扣除。若代入所得素因子整除指标 n，则模该素数可能出现重根，不能继续使用阶等于 n 的结论。

R2.3.4 题型四：先算素数幂，再乘回一般整数

■ **例 R2.4　第二次期中模拟题第 10、15 题**　记 $U(n)=1$。计算 $(U*U*U)(n)$，并求所有满足 $\sum_{d\mid n}\mu(d)\varphi(d)=0$ 的正整数 n。

解　两问都先看积性，再计算素数幂。三重卷积可写成

$$(U*U*U)(n)=\sum_{abc=n}1.$$

当 $n=p^e$，每个因子必须是 p 的幂，故它等于 $e_1+e_2+e_3=e$ 的非负整数解数。固定 $e_1=j$ 后，e_2 可从 0 到 $e-j$，共 $e-j+1$ 种，e_3 随之确定。因此

$$(U*U*U)(p^e)=\sum_{j=0}^{e}(e-j+1)=\frac{(e+1)(e+2)}{2}.$$

不同素因子的指数独立分配，于是若 $n=\prod p_i^{e_i}$，

$$(U*U*U)(n)=\prod_i\frac{(e_i+1)(e_i+2)}{2}.$$

例如 $72=2^3\cdot3^2$，答案为 $10\cdot6=60$；$n=1$ 的空乘积为 1，对应唯一三元组 $(1,1,1)$。对第二问，令 $h(n)=\mu(n)\varphi(n)$，这是两个积性函数的逐点乘积，仍积性。所求因数和是 $h*U$，因此也积性。对 $e\ge1$，$\mu(p^j)=0$（$j\ge2$），所以

$$(h*U)(p^e)=\mu(1)\varphi(1)+\mu(p)\varphi(p)=1-(p-1)=2-p.$$

故

$$\sum_{d\mid n}\mu(d)\varphi(d)=\prod_{p\mid n}(2-p).$$

$n=1$ 时和为 1，不为零；$n>1$ 时乘积为零当且仅当某个素因子为 2。答案恰好是全部正偶数。例如 $n=45$ 时为 $(2-3)(2-5)=3$，$n=72$ 时为 0。 ♦

点评　例 R2.4 连接第 13 讲的积性闭包与第 14 讲的因数和运算。每个「仍积性」都对应一个具体运算：逐点乘积或 Dirichlet 卷积，二者不能互换。若未确认积性，即使已经算出全部素数幂处的值，也不能直接把它们相乘作为一般答案。

R2.3.5　题型五：按奇偶拆开 Dirichlet 级数

■ **例 R2.5**　**第二次期中模拟题第 9 题**　设 $f(n)=(-1)^n$，已知 $\zeta(2)=\pi^2/6$，计算 $Z(f,2)=\sum_{n\ge1}(-1)^n/n^2$。

解　先检查能否重排。因为 $\sum|(-1)^n|/n^2=\zeta(2)<\infty$，可以按偶数和奇数分组。记原和为 S，则把 S 与 $\zeta(2)$ 相加后，所有奇数项抵消，偶数项出现两次：

$$S+\zeta(2)=2\sum_{m\ge1}\frac{1}{(2m)^2}=\frac{1}{2}\zeta(2).$$

所以 $S=-\zeta(2)/2=-\pi^2/12$。首项是 -1，偶数项虽为正但绝对值更小，结果为负也符合交错配对的检查。 ♦

点评　例 R2.5 用第 14 讲的级数分组，不必强行套 Euler 乘积。函数 $(-1)^n$ 在 $n=1$ 处等于 -1，已经不满足非零积性函数的单位值条件，所以不能直接当成积性函数。若级数只有条件收敛，按任意子类重排需要另外证明合法性。

R2.3.6 题型六：把重复扣除写成容斥

■ **例 R2.6　习题集 7 第 3 题的四人情形**　四个人各交一件不同的外套，随机发还，每人一件。求没有人拿到自己外套的发还方案数。

解　把每个人拿到自己外套作为一个坏事件。全部方案有 $4!$ 种，记第 i 人拿对的事件为 A_i。指定 k 个人拿对后，剩下 $4-k$ 人任意排列，故相应交集有 $(4-k)!$ 个方案。注意这些交集允许其他人也拿对，正好适合容斥。

$$\begin{aligned} D_4 &= 4! - \binom{4}{1}3! + \binom{4}{2}2! - \binom{4}{3}1! + \binom{4}{4}0! \\ &= 24 - 24 + 12 - 4 + 1 = 9. \end{aligned}$$

因此共有 9 种。最后的 $0! = 1$ 对应四人都被固定后的唯一空排列，不能把这一项误当成零。 ♦

点评　例 R2.6 是第 15 讲容斥的标准结构：题目说「没有一个坏事件发生」，就数坏事件交集。只扣除单个坏事件会重扣，继续加回两两交集才能纠正。若要求「恰好一人拿对」，则需先选拿对的人，再让剩余三人形成错排，不能沿用当前公式而不改事件条件。

R2.3.7 题型七：不交分类同时给出递推与生成函数

■ **例 R2.7　习题集 7 第 6 题**　设 c_n 为把正整数 n 写成若干个奇数正整数之和的方法数，计入顺序。证明 $c_n = F_n$，并求 $\sum_{n\ge 1} c_n z^n$。

解　先按最后一个加数分类，再用变换得到相邻下标。$n=1$ 只有 (1)，$n=2$ 只有 $(1,1)$，所以 $c_1 = c_2 = 1$。对 $n \ge 3$，将所有表示分成两类：

- 最后一项为 1：删去这一项，得到 $n-1$ 的任意有序奇数拆分；反向补一个 1，构成双射，共 c_{n-1} 种。
- 最后一项至少为 3：从这一项减去 2，得到 $n-2$ 的有序奇数拆分；反向给最后一项加 2，构成双射，共 c_{n-2} 种。

两类互斥且覆盖全部表示，故 $c_n = c_{n-1} + c_{n-2}$。初值与 Fibonacci 数列相同，由第 16 讲的递推唯一性得到 $c_n = F_n$。

令 $C(z) = \sum_{n\ge 1} c_n z^n$。按最后一个奇数加数再看一次：一个奇数部件的生成函数为 $B(z) = z + z^3 + \cdots = z/(1-z^2)$。恰有 k 个部件的有序表示对应 $B(z)^k$，所以

$$C(z) = \sum_{k\ge 1} B(z)^k = \frac{B(z)}{1-B(z)} = \frac{z}{1-z-z^2}.$$

这是形式幂级数等式：每个固定次数只由有限个 k 贡献。核对 $n=5$：表示为 (5)、$(3,1,1)$、$(1,3,1)$、$(1,1,3)$、$(1,1,1,1,1)$，共 $5 = F_5$ 种。 ♦

点评　例 R2.7 综合第 15 讲的双射分类与第 16 讲的初值唯一性、生成函数。递推证明不能漏掉 $n=1,2$，生成函数则要说明有序部件相乘的含义。若改为不计顺序的拆分，$(3,1,1)$ 与其重排会合并，当前双射和 $B(z)^k$ 的解释都需要重新考虑。

R2.3.8 题型八：递推通项与素数模整除的综合

■ **例 R2.8　第二次期中模拟题第 13 题，按解答修正下标**　设 $a_0=0, a_1=1$，且 $a_n=5a_{n-1}-6a_{n-2}$（$n\ge 2$）。求通项和普通生成函数，并证明对每个素数 $p>3$，有 $p\mid a_{p-1}$。

解　先解递推，再把指数化为适合 Fermat 定理的 $p-1$。特征多项式为

$$T^2-5T+6=(T-2)(T-3).$$

两根不同，所以 $a_n=A2^n+B3^n$。初值给出 $A+B=0$、$2A+3B=1$，解得 $A=-1, B=1$。因此

$$a_n=3^n-2^n.$$

回代初值为 $0,1$；两个指数序列都满足递推，故由唯一性这就是题目要求的整个数列。对 $p>3$，两底数都与 p 互素，Fermat 定理给出

$$a_{p-1}=3^{p-1}-2^{p-1}\equiv 1-1=0 \pmod p.$$

再设 $A(z)=\sum_{n\ge 0}a_nz^n$，从 $n=2$ 开始求和得 $A(z)-z=5zA(z)-6z^2A(z)$，所以

$$A(z)=\frac{z}{1-5z+6z^2}=\frac{1}{1-3z}-\frac{1}{1-2z}.$$

展开的前几项为 $0,1,5,19,65,211$，与递推相符；部分分式也重新给出同一个通项。 ♦

点评　例 R2.8 连接第 16 讲的特征根、生成函数与前面学过的 Fermat 定理。初值不能只核对一个；在模素数的步骤里，$p>3$ 保证两个底数可逆。原题文件写成 $p\mid a_p$，但解答文件使用 $p\mid a_{p-1}$；前者确实为假，因为 $a_p\equiv 3-2=1 \pmod p$。

R2.4 常见错误与快速自检

每次做完一道题，可以用「对象、条件、边界、回代」检查一次：符号表示的究竟是什么对象；使用的结论要求什么；$n=1$ 或判别式为零是否已处理；最终答案能否代回原题。下面五项对应不同的失误，不能只靠最后数值看起来合理来判断。

把 Jacobi 值一当成存在平方根　错误推断是「$\left(\frac{a}{n}\right)=1$，所以 $x^2\equiv a \pmod n$ 有解」。例如 $\left(\frac{2}{9}\right)=\left(\left(\frac{2}{3}\right)\right)^2=1$，但 2 模 3 已不是平方，模 9 也不可能是平方。正确做法是：素数分母下可直接判别，合数分母下值为 1 时还要检查各素数幂；值为 -1 才能直接排除。

分圆分解反复扣除低阶因子　错误分母是 $\prod_{d\mid n,d<n}(x^d-1)$。例如求 Φ_4 时会写成 $(x^4-1)/((x-1)(x^2-1))=(x^2+1)/(x-1)$，连多项式都不是。正确分母应为 $\Phi_1\Phi_2=x^2-1$，所以 $\Phi_4=x^2+1$；每种阶的单位根只能移去一次。

把卷积逆当作逐点倒数　$U(n)=1$ 的逐点倒数仍是 U，但卷积逆是 μ。同样，$\mu(9)=0$ 不意味着 $(\mu*\mu)(9)=0$：后者为 $\mu(1)\mu(9)+\mu(3)^2+\mu(9)\mu(1)=1$。先写出卷积求和式，再使用积性；不能把同一个 n 处的两个值相乘代替所有因数对。

重根只写指数解导致参数不足　递推 $u_n = 4u_{n-1} - 4u_{n-2}$ 有二重根 2。若只写 $C2^n$，初值 $u_0 = 1, u_1 = 4$ 无法同时满足。正确形式是 $(A + Bn)2^n$，此处 $A = B = 1$。每个重数为 m 的根需保留 m 个参数，才能匹配全部初值。

看到素数下标就直接声称整除　对例 R2.8，取 $p = 5$，有 $a_5 = 211 \not\equiv 0 \pmod 5$，但 $a_4 = 65 \equiv 0 \pmod 5$。正确下标来自 $3^{p-1} \equiv 2^{p-1} \equiv 1$，应是 $p - 1$。核对来源中的结论也要从公式出发；题面与解答冲突时，应找一个小素数验证，而非直接复制其中一句。

本讲小结

- **平方问题**先查模数和首项可逆性，配方后用 Legendre 符号判别；找出根后还要回代，并说明没有遗漏。
- **Jacobi 与互反律**可用于快速符号计算，但合数分母下符号为 1 不保证是平方；Zolotareff 定理将它与乘法置换的符号联系。
- **分圆与阶**使用 $x^n - 1 = \prod_{d|n} \Phi_d(x)$；由 $p \mid \Phi_n(a)$ 推出阶为 n 前，先确认 $p \nmid n$。
- **积性与反演**先在素数幂处计算，再乘回；$F = f * U$ 当且仅当 $f = F * \mu$，U 与卷积单位元 $\mathbf{1}$ 必须分清。
- **两类生成函数**的卷积不同：Dirichlet 级数乘法按 $ab = n$ 分组，普通生成函数乘法按 $i+j = n$ 分组。无穷级数的数值重排需检查绝对收敛。
- **计数与递推**先做不交分类、检查初值，再用特征根或有理生成函数；重根必须保留多项式因子，素数模结论必须检查实际下标。

练习

R2.1 ★　（第 9、11 讲）求 $x^2 \equiv -2 \pmod{17}$ 的全部解，先用补充律判断，再求根并核对。

R2.2 ★★　（第 10 讲；模拟题第 5 题）设 $n \geq 3$ 为奇数，计算 Jacobi 符号 $\left(\frac{n^3}{n-2}\right)$，答案按 $n \pmod 8$ 分类。

R2.3 ★　（第 11 讲）验证 1 与 34 是模 77 的同一个数的两个非相反平方根，用两个最大公因数分解 77。说明为何只拿到 1 与 -1 时这一方法无效。

R2.4 ★　（第 12 讲）求 $\Phi_8(x)$，计算 $\Phi_8(2)$，并使用阶判别确定 2 模所得素数的阶。

R2.5 ★　（第 13 讲；模拟题第 16 题）计算 $d(360)$、$\varphi(360)$、$\sigma(360)$、$\mu(360)$，并证明不恒为零的积性函数 f 满足 $f(1) = 1$。计算项写出素数幂分解或乘积依据。

R2.6 ★★　（第 14 讲）由 $\sum_{d|n} \varphi(d) = n$ 写出 Möbius 反演式，并用它求 $\varphi(30)$；列出全部因数项。

R2.7 ★★　（第 15 讲；习题集 7 第 3 题的五人情形）求五个人都拿不到自己外套的方案数，说明每个容斥交集的大小。

R2.8 ★★　（第 16 讲）设 $u_0=1, u_1=4$、$u_n=4u_{n-1}-4u_{n-2}$，求通项与普通生成函数，解释为何必须考虑二重根。

R2.9 ★★★　（综合第 9、15、16 讲；习题集 7 第 8 题改编）分别判断 11 与 13 应整除相邻 Fibonacci 数 F_{p-1}, F_{p+1} 中哪一个，用二次互反律说明依据，再计算有关项核对。

R2.10 ★★★　（综合第 13、15、16 讲）记 $H = U * U * U * U$。将 $H(p^e)$ 解释为指数分配问题，求 $H(12)$。再写出以 e 为下标的普通生成函数 $\sum_{e\ge 0} H(p^e) z^e$，说明这里的普通卷积与原先的 Dirichlet 卷积为何并不矛盾。

练习答案与提示

R2.1 $\left(\frac{-2}{17}\right)=\left(\frac{-1}{17}\right)\left(\frac{2}{17}\right)=1\cdot 1=1$，因为 $17\equiv 1 \pmod 4$ 且 $17\equiv 1 \pmod 8$。试平方得 $7^2=49\equiv 15\equiv -2$，故根为 7 和 $-7\equiv 10$。奇素数模下的非零平方恰有两个根，所以没有遗漏；$10^2=100\equiv 15$ 也直接核对通过。

R2.2 因 $n, n-2$ 互素且一者余 1 模 4，互反时不变号。由分子乘法性和立方的奇指数，$\left(\frac{n^3}{n-2}\right)=\left(\frac{n}{n-2}\right)=\left(\frac{n-2}{n}\right)=\left(\frac{-2}{n}\right)$。两个补充律给出：$n\equiv 1,3 \pmod 8$ 时为 1，$n\equiv 5,7 \pmod 8$ 时为 -1。当 $n=3$，分母为 1 的 Jacobi 符号按空乘积为 1，也符合结论。

R2.3 $34^2=1156=15\cdot 77+1$，所以两个数的平方同余；$34\not\equiv \pm 1$。计算 $\gcd(34-1,77)=11$、$\gcd(34+1,77)=7$，得到 $77=7\cdot 11$。若两个根为 $1,-1$，两个最大公因数只会给出 $\gcd(2,77)=1$ 与 $\gcd(0,77)=77$，没有非平凡因子。

R2.4 $x^8-1=(x^4-1)(x^4+1)$，所有真因子指标对应的根都包含在 x^4-1 中，所以 $\Phi_8=x^4+1$。$\Phi_8(2)=17$，$17\nmid 8$，故 $\operatorname{ord}_{17}(2)=8$。直接核对 $2^4=16\equiv -1$，也排除了阶为 $1,2,4$。

R2.5 $360=2^3\cdot 3^2\cdot 5$。于是 $d(360)=(3+1)(2+1)(1+1)=24$；$\varphi(360)=360(1-1/2)(1-1/3)(1-1/5)=96$；$\sigma(360)=(1+2+4+8)(1+3+9)(1+5)=15\cdot 13\cdot 6=1170$；因含 2^2 等平方因子，$\mu(360)=0$。证明部分取 m 使 $f(m)\neq 0$。由 $\gcd(m,1)=1$ 和积性，$f(m)=f(m\cdot 1)=f(m)f(1)$；约去非零的 $f(m)$ 得 $f(1)=1$。

R2.6 反演给出 $\varphi(n)=\sum_{d\mid n}\mu(d)n/d$。对 $n=30$，因数依次为 $1,2,3,5,6,10,15,30$，逐项求和为

$$\varphi(30)=30-15-10-6+5+3+2-1=8.$$

这也等于 $30(1-1/2)(1-1/3)(1-1/5)=8$，两种计算互相核对。

R2.7 指定 k 人拿对后，剩余者可任意排列，交集大小为 $(5-k)!$。因此 $D_5=5!-\binom{5}{1}4!+\binom{5}{2}3!-\binom{5}{3}2!+\binom{5}{4}1!-\binom{5}{5}0!=120-120+60-20+5-1=44$。容斥的「指定拿对」不等于「恰好这些人拿对」，所以交集里不应再施加错排条件。

R2.8 特征多项式为 $(T-2)^2$，所以 $u_n=(A+Bn)2^n$。$u_0=1$ 给出 $A=1$，$u_1=4$ 给出 $2(1+B)=4$，故 $B=1$。于是 $u_n=(n+1)2^n$。设 $G(z)=\sum u_n z^n$，移位求和得 $(1-4z+4z^2)G(z)=u_0+(u_1-4u_0)z=1$，故 $G(z)=1/(1-2z)^2$。若漏掉因子 n，便只剩一个常数，无法同时匹配两个初值。

R2.9 $\left(\frac{5}{11}\right)=\left(\frac{11}{5}\right)=\left(\frac{1}{5}\right)=1$，所以 $11\mid F_{10}$；$F_{10}=55$，而 $F_{12}=144\equiv 1 \pmod{11}$。$\left(\frac{5}{13}\right)=\left(\frac{13}{5}\right)=\left(\frac{3}{5}\right)=-1$，所以 $13\mid F_{14}$；$F_{14}=377=29\cdot 13$，而 $F_{12}=144\equiv 1 \pmod{13}$。互反律的分子 $5\equiv 1 \pmod 4$，所以这两次翻转都不变号。

R2.10 $H(p^e)$ 数非负整数四元组 $e_1+e_2+e_3+e_4=e$。将 e 个相同标记与 3 个隔板排成一行，隔板间四段长度就是四个指数，故共有 $\binom{e+3}{3}$ 种，允许空段。由积性，$H(12)=H(2^2)H(3)=\binom{5}{3}\binom{4}{3}=10\cdot 4=40$。又每个指数贡献 $1+z+z^2+\cdots$，所以 $\sum_{e\ge 0}H(p^e)z^e=(1-z)^{-4}$。Dirichlet 卷积原本要求四个整数因子的乘积为 p^e，固定素数后这恰好等价于指数之和为 e，因此在指数下标上转成普通卷积。

第三部分

连分数与丢番图方程

第 17 讲

生成函数的筛选与两平方和定理

Generating Functions and the Two-Square Theorem

对应材料：MIT 18.781 Lecture 17 讲义；习题集 7 第 10 题；习题集 8 第 1、2、4 题，第 3 题 (a)(c) 与第 5 题 (a)（选编改写）。

本讲目标

(1) 会对普通生成函数做线性组合、移位、求导和乘法，并从乘法的系数读出卷积。

(2) 会用 m 次单位根筛选下标满足 $n \equiv r \pmod m$ 的系数和，能处理具体的几何级数。

(3) 能用生成函数计算一个交错 zeta 和，并看懂它与 Euler 乘积在素数 2 处的差别。

(4) 掌握两平方和定理的必要性与充分性证明，尤其是「最小倍数 + 下降」的关键构造。

(5) 会用两平方恒等式构造新的表示，并判断一个整数的素因子指数条件是否允许它成为两个整数平方和。

17.1 生成函数的四种运算

第 16 讲把一个递推数列装进生成函数，得到一个有理函数。本讲先把这个工具本身拆开：每一种代数运算都对应数列上的一个明确操作。设数列 $(a_n)_{n\ge 0}$ 的普通生成函数为

$$A(x)=\sum_{n\ge 0} a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots .$$

这里仍把它看作形式幂级数；因此只比较每个幂次的系数，不需要先讨论 x 的取值范围。

定义 17.1（普通生成函数） 数列 $(a_n)_{n\ge 0}$ 的**普通生成函数**（ordinary generating function）是形式幂级数 $A(x)=\sum_{n\ge 0} a_n x^n$。如果 $A(x)$ 的系数能逐个恢复 a_n，就说 A 编码了这个数列。

最基本的例子是等比数列。由几何级数恒等式，

$$\frac{1}{1-rx}=\sum_{n\ge 0} r^n x^n. \tag{17.1}$$

例如 $r=1/2$ 时，$1/(1-x/2)$ 的系数依次为 $1, 1/2, 1/4, \ldots$；$r=0$ 时只有常数项 1，这也是边界情形。

命题 17.2（线性、移位与加权）　若 $A(x)$、$B(x)$ 分别编码 (a_n)、(b_n)，则：

(1) $\alpha A(x)+\beta B(x)$ 编码 $(\alpha a_n+\beta b_n)$；

(2) $xA(x)$ 编码 $(0,a_0,a_1,\dots)$；

(3) $xA'(x)$ 编码 $(na_n)_{n\ge 0}$。

证明　三条都直接比较系数。第一条把常数提出求和号：

$$\sum_{n\ge 0}(\alpha a_n+\beta b_n)x^n=\alpha\sum_{n\ge 0}a_nx^n+\beta\sum_{n\ge 0}b_nx^n.$$

第二条是

$$xA(x)=a_0x+a_1x^2+a_2x^3+\cdots.$$

第三条先逐项求导：$A'(x)=\sum_{n\ge 1}na_nx^{n-1}$，再乘以 x 即得 $\sum_{n\ge 0}na_nx^n$；$n=0$ 项本来就是 0。 ∎

例 17.1　从几何级数得到 n 与 n^2　求数列 $(n/2^n)$ 与 $(n^2/2^n)$ 的生成函数。

解　取式 (17.1) 中的 $r=1/2$，先得

$$A(x)=\frac{1}{1-x/2}.$$

由加权规则，

$$\sum_{n\ge 0}\frac{n}{2^n}x^n=xA'(x)=\frac{x/2}{(1-x/2)^2}.$$

再加权一次：

$$\sum_{n\ge 0}\frac{n^2}{2^n}x^n=x\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x/2}{(1-x/2)^2}\right)=\frac{x(1+x/2)}{2(1-x/2)^3}.$$

例如前五个系数分别是 $0,1/2,1,9/8,1$，直接代入公式也得到同样结果。 ♦

乘法对应卷积。图 17.1 把 x^3 的系数画成一条对角线：只有下标相加为 3 的项会贡献给它。

$i\backslash j$	0	1	2	3
0	a_0b_0	a_0b_1	a_0b_2	a_0b_3
1	a_1b_0	a_1b_1	a_1b_2	a_1b_3
2	a_2b_0	a_2b_1	a_2b_2	a_2b_3
3	a_3b_0	a_3b_1	a_3b_2	a_3b_3

$$c_3=a_0b_3+a_1b_2+a_2b_1+a_3b_0$$

图 17.1　乘法的 x^3 系数是蓝色四格之和；它们的行、列下标都满足 $i+j=3$。

图中第 i 行、第 j 列代表乘积 $a_ib_jx^{i+j}$。沿蓝色斜线相加就得到 x^3 的系数；若要找 x^2 的系数，则改取 $i+j=2$ 的三个格子。

命题 17.3（乘法与卷积）　若 $A(x)=\sum_{n\ge 0}a_nx^n$、$B(x)=\sum_{n\ge 0}b_nx^n$，则

$$A(x)B(x)=\sum_{n\ge 0}c_nx^n,\qquad c_n=\sum_{k=0}^{n}a_kb_{n-k}.$$

证明　展开乘积后，x^n 的来源必须满足 $k+(n-k)=n$。这样的配对恰好由 $k=0,1,\dots,n$ 给出，所以该系数就是所写的有限和。 ■

■ **例 17.2　卷积的一个具体系数**　求 $(1+x+x^2)(1+2x+x^2)$ 中 $x^0,\dots,x^4$ 的系数，并解释它如何对应两个长度为 3 的数列。

解　令 $(a_0,a_1,a_2)=(1,1,1)$、$(b_0,b_1,b_2)=(1,2,1)$，其余项为 0。卷积给出

$$(c_0,c_1,c_2,c_3,c_4)=(1,3,4,3,1).$$

例如 $c_2=a_0b_2+a_1b_1+a_2b_0=1+2+1=4$。这正是多项式乘法逐列相加的系数版本。 ♦

普通生成函数的乘法给出没有权重的卷积；当对象带有「标号」时，组合数会自然出现。为此定义**指数生成函数**（exponential generating function）

$$\widetilde{A}(x)=\sum_{n\ge 0}a_n\frac{x^n}{n!}.$$

同一个数列换一种编码，移位和乘法的公式也会稍微改变。

命题 17.4（指数生成函数的移位与乘法）　若 $\widetilde{A}(x)$ 编码 (a_n)，则 $\widetilde{A}'(x)$ 编码 (a_{n+1})。若 $\widetilde{A}$、$\widetilde{B}$ 分别编码 (a_n)、(b_n)，且 $\widetilde{A}\widetilde{B}$ 编码 (c_n)，则

$$c_n=\sum_{k=0}^{n}\binom{n}{k}a_kb_{n-k}.$$

证明　逐项求导即可：

$$\frac{\mathrm{d}}{\mathrm{d}x}\sum_{n\ge 0}a_n\frac{x^n}{n!}=\sum_{n\ge 1}a_n\frac{x^{n-1}}{(n-1)!}=\sum_{n\ge 0}a_{n+1}\frac{x^n}{n!}.$$

乘法中 x^n 的系数是

$$\sum_{k=0}^{n}\frac{a_k}{k!}\frac{b_{n-k}}{(n-k)!}.$$

乘以 $n!$ 后，$n!/(k!(n-k)!)=\binom{n}{k}$，得到结论。 ■

■ **例 17.3　指数生成函数与指数函数**　设 $a_n=r^n$。求它的指数生成函数，并解释为什么微分方程能唯一确定答案。

解　记 $E(x)=\sum_{n\ge 0} r^n x^n/n!$。由移位命题，$E'(x)$ 编码 (r^{n+1})，而逐项乘以 r 得到的 $rE(x)$ 也编码同一数列。因此 $E'=rE$。初值 $E(0)=1$，解为

$$E(x)=e^{rx}=\sum_{n\ge 0} r^n\frac{x^n}{n!}.$$

若两个形式幂级数都满足 $E'=rE$ 且常数项为 1，比较递推得到的各阶系数即可知它们相同；所以这里没有额外的「选解」问题。 ♦

上面的对象可以看作两种不同层次的记账。普通生成函数的系数直接等于方案数，所以独立选择两个对象时，方案数相乘后按总大小求和，得到普通卷积。指数生成函数则把大小为 n 的对象除以 $n!$；当若干块对象要分配到一组标号上时，选择哪些标签给第一块会产生二项式系数。两种生成函数都编码同一类数列，但乘法的组合意义不同，不能在中途把 $n!$ 因子丢掉。

从计算角度看，形式幂级数把「按系数计算」与「在某个数值点求和」分开了。例如 $1/(1-x)$ 在形式意义下等于 $1+x+x^2+\cdots$；作为实数级数，它只在 $|x|<1$ 收敛。第 16 讲求线性递推的生成函数时主要使用前一种意义；本讲把 x 代成具体数来求无穷和时，才需要核对收敛范围。

想一想　若 $A(x)$ 编码 (a_n)，$x^2A(x)$ 编码什么？如果要删去原数列前两项，并把原来的 a_2 重新记在下标 0，应如何改写？

答　$x^2A(x)$ 编码 $(0,0,a_0,a_1,\dots)$。若要删去前两项而保持下标从 0 开始，应使用 $(A(x)-a_0-a_1x)/x^2$，它编码 $(a_2,a_3,\dots)$。

17.2　单位根筛选

生成函数不仅能恢复单个系数，还能一次选出某个同余类的所有系数。先看最小的例子：

$$1+(-1)^n=\begin{cases}2, & n\text{ 为偶数},\\ 0, & n\text{ 为奇数}.\end{cases}$$

四次**单位根**（root of unity）$1,-1,i,-i$ 能把「偶数」细分成模 4 的两个余数类。

引理 17.5（单位根的正交性）　设 $\omega=e^{2\pi i/m}$，则对任意整数 n，

$$\sum_{j=0}^{m-1}\omega^{jn}=\begin{cases}m, & m\mid n,\\ 0, & m\nmid n.\end{cases}$$

证明　若 $m\mid n$，每项都是 1，和为 m。否则 $q=\omega^n\neq 1$ 且 $q^m=1$，几何级数给出

$$1+q+\cdots+q^{m-1}=\frac{1-q^m}{1-q}=0.$$

■

定理 17.6（单位根筛选公式）　设 $A(x)=\sum_{n\ge 0}a_nx^n$，$\omega=e^{2\pi i/m}$。对 $0\le r<m$，下标满足 $n\equiv r \pmod m$ 的系数和（在收敛的 x 上）由

$$\sum_{n\equiv r\,(\mathrm{mod}\,m)} a_nx^n=\frac{1}{m}\sum_{j=0}^{m-1}\omega^{-rj}A(\omega^j x)$$

给出。

证明　将右侧展开为幂级数并交换有限的 j 和无限的 n：

$$\frac{1}{m}\sum_{j=0}^{m-1}\omega^{-rj}\sum_{n\ge 0}a_n(\omega^j x)^n=\sum_{n\ge 0}a_nx^n\left(\frac{1}{m}\sum_{j=0}^{m-1}\omega^{j(n-r)}\right).$$

括号由引理等于 1（当 $n-r$ 被 m 整除）或 0（否则），正好留下所需的项。 ■

直观理解　每个 $A(\omega^j x)$ 都给第 n 项乘上相位 ω^{jn}。加权求和后，不符合余数 r 的相位彼此抵消，符合的相位全部同向相加。单位根筛选就是用相消来做一张「同余类过滤器」。

■ **例 17.4**　筛出 $n\equiv 0 \pmod 4$　计算 $\displaystyle\sum_{n\equiv 0\,(\mathrm{mod}\,4)} 2^{-n}$。

解　直接看下标，这是几何级数

$$1+2^{-4}+2^{-8}+\cdots=\frac{1}{1-2^{-4}}=\frac{16}{15}.$$

用单位根筛选也能得到同一个答案。令 $A(x)=1/(1-x/2)$，在 $x=1,-1,i,-i$ 处代入并相加，因

$$1+(-1)^n+i^n+(-i)^n=\begin{cases}4, & 4\mid n,\\ 0, & 4\nmid n,\end{cases}$$

所以四个值的和除以 4 正是所求结果 $16/15$。这个例子说明筛选公式能把「看得出」的几何级数写成统一算法。 ♦

■ **例 17.5**　筛出 $n\equiv 3 \pmod 4$　计算 $\displaystyle\sum_{n\equiv 3\,(\mathrm{mod}\,4)} 2^{-n}$，并说明它与上一个例子的关系。

解　下标为 $3,7,11,\ldots$，所以

$$2^{-3}+2^{-7}+2^{-11}+\cdots=\frac{1/8}{1-1/16}=\frac{2}{15}.$$

在单位根公式中把四个值分别乘 $1^{-3},(-1)^{-3},i^{-3},(-i)^{-3}$，再相加除以 4，就会只保留 $n\equiv 3 \pmod 4$ 的项。四个余数类的和为

$$\frac{16}{15}+\frac{8}{15}+\frac{4}{15}+\frac{2}{15}=2=\sum_{n\ge 0}2^{-n},$$

这也是一次核对。 ♦

■ **例 17.6　三次单位根的筛选**　令 $\omega = e^{2\pi i/3}$，计算

$$\sum_{k\ge 0}\binom{\frac{1}{3}}{3k+1}.$$

要求分别写出直接计算和单位根筛选两种做法。

解　直接按下标看，所求和为

$$\frac{1}{3}+\frac{1}{3^4}+\frac{1}{3^7}+\cdots=\frac{1/3}{1-1/27}=\frac{9}{26}.$$

单位根筛选取 $A(x)=1/(1-x/3)$ 与余数 $r=1$，得到

$$\frac{1}{3}\left(A(x)+\omega^{-1}A(\omega x)+\omega^{-2}A(\omega^2 x)\right).$$

令 $x=1$ 后，第 n 项的权重是 $(1+\omega^{n-1}+\omega^{2(n-1)})/3$，它在 $n\equiv 1 \pmod 3$ 时为 1，其余时候为 0。因此筛选公式确实返回 $9/26$。三个复数值的总和最后仍是实数。 ♦

正文中的四个方向正如图 17.2 所示；它们对应四个相位的相消与保留。

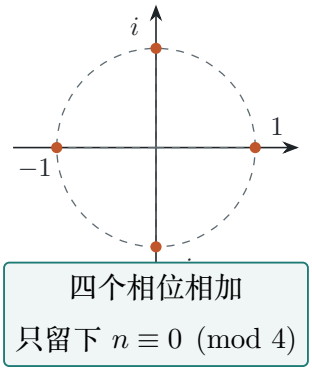


图 17.2　四次单位根位于单位圆的四个方向；相位和对模 4 的非零余数类抵消。

易错提醒　单位根筛选的权重不能漏掉 ω^{-rj}。只把 $A(\omega^j x)$ 相加得到的是 $r=0$ 类；要筛 $r=3$，必须乘上 ω^{-3j}。另一个常见错误是忘记最后除以 m，那会把保留下来的每一项放大 m 倍。

17.3　交错 zeta 和

第 14 讲用 Möbius 函数把整除条件编码进 zeta 函数。这里看一个很短但很有代表性的计算：

$$S=-1+\frac{1}{4}-\frac{1}{9}+\frac{1}{16}-\cdots=\sum_{n\ge 1}\frac{(-1)^n}{n^2}.$$

它既可以按奇偶下标分组，也可以从 Euler 乘积看出为什么素数 2 是特殊的。

■ **例 17.7　按偶数下标分组**　用 $\zeta(2)=\pi^2/6$ 计算 S。

解　偶数项在 S 中为正，奇数项为负。故

$$\zeta(2)+S=2\left(\frac{1}{2^2}+\frac{1}{4^2}+\frac{1}{6^2}+\cdots\right)=2\cdot\frac{1}{2^2}\zeta(2)=\frac{1}{2}\zeta(2).$$

于是

$$S=-\frac{1}{2}\zeta(2)=-\frac{\pi^2}{12}.$$

这里的系数 2 来自「在 $\zeta(2)+S$ 中，奇数项抵消、偶数项加倍」，不能直接把 S 当成 $-\zeta(2)$。 ♦

■ **例 17.8**　**从 Euler 因子检查同一个答案**　用 Euler 乘积说明 $-S$ 与 $\zeta(2)$ 的比值只改变素数 2 的因子。

解　先写

$$-S=1-\frac{1}{2^2}+\frac{1}{3^2}-\frac{1}{4^2}+\cdots=\left(1-\frac{1}{2^2}-\frac{1}{2^4}-\frac{1}{2^6}-\cdots\right)\prod_{p\text{ odd}}\left(1-\frac{1}{p^2}\right)^{-1}.$$

而

$$\zeta(2)=\left(1+\frac{1}{2^2}+\frac{1}{4^2}+\cdots\right)\prod_{p\text{ odd}}\left(1-\frac{1}{p^2}\right)^{-1}.$$

奇素数因子完全相同。两个 2-因子的比值为

$$\frac{1-(1/4+1/16+1/64+\cdots)}{1+1/4+1/16+\cdots}=\frac{1-1/3}{1/(1-1/4)}=\frac{2/3}{4/3}=\frac{1}{2}.$$

所以 $(-S)/\zeta(2)=1/2$，再次得到 $S=-\pi^2/12$。 ♦

这里可以交换乘积与求和，是因为 $\sum_{n\ge1}1/n^2$ 绝对收敛：先把有限个素数对应的几何级数相乘，再逐步加入新的素数；遗漏项的绝对值总和随截断增大而趋于 0。如果把 $1/n^2$ 换成只条件收敛的级数，就不能只凭「素因子分解唯一」任意重排项。

想一想　若把 S 改成 $\sum_{n\ge1}(-1)^{n-1}/n^2$，答案的符号如何变化？若只保留偶数下标，和 $\zeta(2)$ 的比值是多少？

答　前一个级数就是 $-S$，因此等于 $\pi^2/12$。只保留偶数下标得到 $\sum_{k\ge1}(2k)^{-2}=\zeta(2)/4=\pi^2/24$；若在原级数中把偶数项加倍，才会出现上一个例子里的系数 2。

17.4　两平方和：必要性与最小倍数

哪些素数能写成两个整数平方的和？先列出几个小素数：

$$2=1^2+1^2,\quad 5=1^2+2^2,\quad 13=2^2+3^2,\quad 17=1^2+4^2,$$

而 $3,7,11,19$ 都没有这样的表示。模 4 正好解释了这张表。

命题 17.7（素数两平方和的必要条件）　若奇素数 $p=a^2+b^2$，则 $p\equiv1\pmod4$。另外 $2=1^2+1^2$。

证明　先证必要性。$p=2$ 时 $2=1^2+1^2$。设 p 为奇素数且 $p=a^2+b^2$。平方模 4 只有 $0,1$ 两种可能。若 a,b 都偶，p 为偶数；若都奇，$p\equiv 1+1\equiv 2 \pmod 4$；所以一个偶、一个奇，从而 $p\equiv 0+1\equiv 1 \pmod 4$。

■

充分性要解决的是：给定一个 $p\equiv 1 \pmod 4$ 的素数，怎样从同余信息构造出两个平方。下面对 p 做强归纳；关键是先找到一个小于 p 的正整数 m，使 mp 已经是两个平方和，再在 $m>1$ 时构造更小的倍数，矛盾。这样，必要性和充分性在逻辑上各自独立，最后由定理 17.11 合并。

要开始下降，需要模 p 的 -1 平方根。由于 $p\equiv 1 \pmod 4$，Euler 判别法给出

$$\left(\frac{-1}{p}\right)=(-1)^{(p-1)/2}=1,$$

所以存在 x 使 $x^2\equiv -1 \pmod p$。

命题 17.8（两平方恒等式）　对任意整数 a,b,c,d，

$$(a^2+b^2)(c^2+d^2)=(ac-bd)^2+(ad+bc)^2.$$

证明　展开右侧：

$$(ac-bd)^2+(ad+bc)^2=a^2c^2-2abcd+b^2d^2+a^2d^2+2abcd+b^2c^2=(a^2+b^2)(c^2+d^2).$$

两项交叉项一正一负，恰好抵消。 ■

引理 17.9（小倍数）　若 $p\equiv 1 \pmod 4$ 为奇素数，则存在整数 a 与 m 满足

$$0<m<p,\qquad a^2+1=mp.$$

证明　从同余类 $\pm 1,\pm 2,\dots,\pm(p-1)/2$ 中选出满足 $x^2\equiv -1 \pmod p$ 的代表元 x，于是 $|x|\le (p-1)/2$。令 $a=x$，因为 $p\mid x^2+1$，存在正整数 $m=(x^2+1)/p$。又因 $p\ge 5$，

$$0<x^2+1\le \frac{(p-1)^2}{4}+1<p^2,$$

所以 $0<m<p$。这里选择对称代表元的用处正是把 x 控制在 $p/2$ 以内。 ■

■ **例 17.9　小倍数的具体计算**　对 $p=13$，找出引理中的 x、m，并核对 mp 是两个平方和。

解　$5^2=25\equiv -1 \pmod{13}$，并且 $|5|<13/2$。所以

$$5^2+1=26=2\cdot 13,$$

即 $m=2$，且 $26=5^2+1^2$。这个 m 还不是最终答案，但它是下降法的起点。 ♦

这个例子也展示了下降中的「偶数步骤」。因为 1 与 5 都是奇数，

$$\left(\frac{1+5}{2}\right)^2+\left(\frac{1-5}{2}\right)^2=3^2+(-2)^2=13.$$

从 $2\cdot 13=1^2+5^2$ 直接降到了 $13=3^2+(-2)^2$。负号不影响平方；保留它能让两个坐标都来自同一个恒等变换。

17.5　两平方和：下降法

现在完成充分性证明。下降法的对象不是 p 本身，而是所有满足「是 p 的正整数倍且为两个平方和」的倍数系数。

思路　先取最小的 $m>0$，使 $mp=a^2+b^2$。证明 $(a,b)=1$ 且 m 为奇数；再取 m 的任一奇素因子 q，用 $a^2\equiv -b^2 \pmod q$ 说明 $q\equiv 1 \pmod 4$。归纳假设把 q 写成 c^2+d^2，两平方恒等式会把 mp 降成 $(m/q)p$。

> **引理 17.10（下降构造）**　设 $p\equiv 1 \pmod 4$，并假设每个小于 p 且同余 $1 \pmod 4$ 的素数都是两个平方和。若 m 是满足 $mp=a^2+b^2$ 的最小正整数，则 $m=1$。

证明　小倍数引理保证这样的 m 存在且 $m<p$。若 $m=1$，结论成立。以下假设 $m>1$。

先证明 $(a,b)=1$。若 $g=(a,b)>1$，则 $g^2\mid mp$。由 $a^2+b^2=mp<p^2$，得到 $|a|,|b|<p$，所以 $g<p$，从而 $(g,p)=1$。因此 $g^2\mid m$，并且

$$\left(\frac{a}{g}\right)^2+\left(\frac{b}{g}\right)^2=\frac{m}{g^2}p$$

给出一个正整数倍数系数小于 m，与最小性矛盾。

再证 m 为奇数。若 m 偶，则 a^2+b^2 偶，a,b 同奇偶；互素排除了同为偶，所以二者都奇。于是

$$\left(\frac{a+b}{2}\right)^2+\left(\frac{a-b}{2}\right)^2=\frac{a^2+b^2}{2}=\frac{m}{2}p,$$

产生更小的倍数系数，仍矛盾。

取 m 的奇素因子 q，写 $m=qn$。由 $q\mid a^2+b^2$ 且 $(a,b)=1$，q 不可能同时整除 a,b。若 $q\mid b$，则也有 $q\mid a$，矛盾；所以 b 在模 q 下可逆，并且

$$\left(ab^{-1}\right)^2\equiv -1 \pmod q.$$

于是 -1 是模 q 的平方，Euler 判别法给出 $q\equiv 1 \pmod 4$。由于 $q\le m<p$，归纳假设给出 $q=c^2+d^2$。

模 q 下，$a^2\equiv -b^2$、$c^2\equiv -d^2$。相乘得 $(ac)^2\equiv (bd)^2$，所以 $ac\equiv \pm bd \pmod q$。把 c 换成 $-c$ 不改变 c^2+d^2，可设 $q\mid ac-bd$。

使用命题 17.8 的两平方恒等式

$$(a^2+b^2)(c^2+d^2)=(ac-bd)^2+(ad+bc)^2,$$

左边为 pq^2n。因为 $q \mid ac-bd$，恒等式模 q 又给出 $q \mid ad+bc$。因此

$$\left(\frac{ac-bd}{q}\right)^2+\left(\frac{ad+bc}{q}\right)^2=pn.$$

这里 $n=m/q<m$，与 m 的最小性矛盾。所以 $m>1$ 不可能，只能是 $m=1$。 ■

图 17.3 用两组数分别展示偶数步骤和奇素因子步骤。图中的 $m=2,5$ 都是选出的初始乘数；它们尚未最小，下降后才得到 $m=1$。

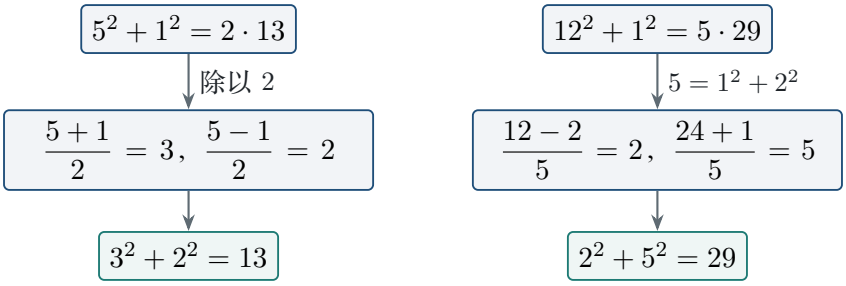


图 17.3　下降改变的是 p 前面的正整数乘数：左侧从 2 降为 1，右侧从 5 降为 1，两条链都保留「两个整数平方和」的形式。

右侧使用 $(a,b,c,d,q)=(12,1,1,2,5)$。两个新坐标分别来自 $(ac-bd)/q$ 与 $(ad+bc)/q$，除法后仍为整数；恒等式再保证它们的平方和为原值的 $1/q$。一般证明先取最小乘数，再说明若它大于 1，同类构造就会给出更小的正整数乘数，形成矛盾。

> **定理 17.11（素数两平方和定理）**　素数 p 能写成两个整数平方的和，当且仅当 $p=2$ 或 $p\equiv 1 \pmod 4$。

证明　必要性由命题 17.7 给出。证明充分性时，对所有 $p\equiv 1 \pmod 4$ 的素数做强归纳。基例 $p=5$ 有 $5=1^2+2^2$；对更大的 p，小倍数引理与下降引理给出最小倍数 $m=1$，于是 $p=a^2+b^2$。素数 2 已由 $2=1^2+1^2$ 覆盖。 ■

■ **例 17.10　用恒等式合成表示**　已知 $5=1^2+2^2$、$13=2^2+3^2$，构造 65 的一个两平方表示；再用 13 和 $17=1^2+4^2$ 构造 221 的表示。

解　取恒等式中的 $(a,b,c,d)=(1,2,2,3)$，得到

$$65=(1\cdot 2-2\cdot 3)^2+(1\cdot 3+2\cdot 2)^2=(-4)^2+7^2.$$

对 $13\cdot 17$ 取 $(a,b,c,d)=(2,3,1,4)$，得到

$$221=(2-12)^2+(8+3)^2=10^2+11^2.$$

符号的另一个选择会给出同一个整数的其他表示；恒等式只负责保证至少有一种表示。 ♦

> **推论 17.12（素因子指数的必要条件）**　若正整数 $N=a^2+b^2$，且素数 $q\equiv 3 \pmod 4$ 整除 N，则 q 的指数为偶数。

证明　由 $a^2 \equiv -b^2 \pmod q$，若 $q \nmid b$ 就会得到 -1 是模 q 的平方，与 $q \equiv 3 \pmod 4$ 矛盾。因此 $q \mid b$，再由 $q \mid a^2$ 得 $q \mid a$。写 $a = qa_1, b = qb_1$，便有 $N = q^2(a_1^2 + b_1^2)$。若 q 的指数是奇数，除掉所有完整的 q^2 因子后，剩余的两平方和仍被 q 整除；同一论证又迫使它被 q^2 整除，矛盾。因此指数只能是偶数。 ■

定理 17.13（整数的两平方和判别）　正整数 N 能写成两个整数平方的和，当且仅当 N 的素因子分解中，每个满足 $q \equiv 3 \pmod 4$ 的素数都以偶数次出现。

证明　必要性就是上面的推论。反过来，写

$$N = 2^e \prod_{p_i \equiv 1(4)} p_i^{\alpha_i} \prod_{q_j \equiv 3(4)} q_j^{2\beta_j}.$$

由两平方和定理，$2 = 1^2 + 1^2$，每个 p_i 是两平方和；而 $q_j^2 = q_j^2 + 0^2$ 直接给出 q_j^2 是两平方和。边界 $N = 1$ 可写成 $1^2 + 0^2$。两个两平方和的乘积由恒等式仍是两平方和。因此把这些因子逐个相乘，就得到 N 的表示。 ■

这个判别把「素数问题」变成了「指数问题」。例如 $105 = 3 \cdot 5 \cdot 7$ 含有两个不同的 $3 \pmod 4$ 素因子，但它们各自的指数都是 1，所以不能表示；而 $90 = 2 \cdot 3^2 \cdot 5$ 中 3 的指数为 2，所以可以表示。

为了把「逐个相乘」写成可执行的步骤，可以先列出基本因子：

因子	两平方表示
2	$1^2 + 1^2$
$q^{2\beta},\ q \equiv 3 \pmod 4$	$(q^\beta)^2 + 0^2$
$p^\alpha,\ p \equiv 1 \pmod 4$	把p 的表示重复相乘α 次

表中第三行每重复一次，就使用一次两平方恒等式；因此这给出一个递归构造算法。比如 $90 = 2 \cdot 3^2 \cdot 5$：先把 2 与 5 合成为 $10 = 1^2 + 3^2$，再和 $3^2 = 3^2 + 0^2$ 相乘，得到

$$90 = (3 \cdot 1 - 0 \cdot 3)^2 + (3 \cdot 3 + 0 \cdot 1)^2 = 3^2 + 9^2.$$

等价的形式是 $(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2$。它来自把复数乘积的实部、虚部换一组符号；做题时可以选择让绝对值较小的那个坐标出现在第一项。

命题 17.14（有理平方和的清分母）　若正整数 N 是两个有理数的平方和，则 N 也是两个整数的平方和。

证明　写成 $N = (a/d)^2 + (b/d)^2$，取 $a, b, d \in \mathbb{Z}$、$d > 0$ 且 $\gcd(a, b, d) = 1$。于是

$$Nd^2 = a^2 + b^2.$$

若 $q \equiv 3 \pmod 4$ 整除 d，则 $q \mid a^2 + b^2$，上面的推论迫使 $q \mid a, b$，这与 $\gcd(a, b, d) = 1$ 矛盾。因此这样的 q 不整除 d。另一方面，q 在 Nd^2 中的指数为偶数；因为 $q \nmid d$，它在 N 中的指数也为偶数。整数两平方和判别遂给出 $N = x^2 + y^2$，其中 $x, y \in \mathbb{Z}$。 ■

■ **例 17.11　判断一个整数能否为两平方和**　判断 65、90、45 是否能写成两个整数平方的和。

解　$65 = 5 \cdot 13$，两个素因子都为 2 或 1 (mod 4)，且上面的恒等式已经给出 $65 = 4^2 + 7^2$。对 90 先把 2 与 5 合成：

$$10 = (1^2+1^2)(1^2+2^2) = (1\cdot 1 - 1\cdot 2)^2 + (1\cdot 2 + 1\cdot 1)^2 = 1^2 + 3^2.$$

再与 $9 = 3^2 + 0^2$ 相乘，得到

$$90 = (3\cdot 1 - 0\cdot 3)^2 + (3\cdot 3 + 0\cdot 1)^2 = 3^2 + 9^2.$$

最后 $45 = 3^2 \cdot 5 = 3^2(1^2+2^2) = 3^2 + 6^2$。三个数都满足指数条件。 ♦

最小倍数下降　要证明一个素数本身是两个平方和，可以先允许一个小倍数 mp，再取最小的 m。两平方恒等式把 m 的一个素因子除掉；「构造出更小的同类对象」正是矛盾的来源。

前后联系　本讲的下降法先利用模 p 的平方根，再把同余关系转成整数平方和。下一讲会回到欧几里得算法，将每一步的商记录为连分数，并用同样的「逐步缩小余数」观点研究有理数的表示与逼近。

本讲小结

- 普通生成函数的线性组合、移位和加权分别对应数列的线性组合、前面补零和乘以下标：$xA'(x) = \sum_{n\ge 0} n a_n x^n$。
- 乘法对应卷积：若 $C(x) = A(x)B(x)$，则 $c_n = \sum_{k=0}^{n} a_k b_{n-k}$。
- 单位根筛选公式为 $\displaystyle\sum_{n\equiv r(m)} a_n x^n = \frac{1}{m}\sum_{j=0}^{m-1} \omega^{-rj} A(\omega^j x)$。
- $\displaystyle\sum_{n\ge 1} (-1)^n / n^2 = -\zeta(2)/2 = -\pi^2/12$。
- 素数两平方和定理：$p = a^2 + b^2$ 当且仅当 $p = 2$ 或 $p \equiv 1 \pmod 4$。
- 两平方恒等式 $(a^2+b^2)(c^2+d^2) = (ac-bd)^2 + (ad+bc)^2$ 是下降证明的核心；若 $q \equiv 3 \pmod 4$ 整除 a^2+b^2，则 q 的指数为偶数。

练习

17.1 ★　用单位根筛选或直接按余数分类，计算 $\displaystyle\sum_{k\ge 0} \binom{n}{3k}$，并按 $n \pmod 6$ 写出结果。

17.2 ★　设 $\widetilde{A}(x) = \sum_{n\ge 0} a_n x^n / n!$。证明 $\widetilde{A}'(x)$ 编码左移数列 $(a_{n+1})_{n\ge 0}$；再证明数列 $a_n = r^n$ 的指数生成函数是 e^{rx}。

17.3 ★　计算 $5, 13, 17, 29, 37$ 的两平方表示，并用模 4 证明 $3, 7, 11, 19$ 没有两平方表示。

17.4 ★★　定义 B_n 为 $x/(e^x - 1) = \sum_{n\ge 0} B_n x^n / n!$ 的系数。求 $B_0, \dots, B_6$，并由乘法恒等式推出 $B_0 + \binom{n}{1} B_1 + \cdots + \binom{n}{n-1} B_{n-1} = 0$ $(n \ge 2)$。

17.5 ★★ 证明两个两平方和的正整数之积仍是两平方和，并据此表示 $2^3\cdot 5\cdot 13$；说明为什么 $3\cdot 5$ 不能是两平方和。

17.6 ★★ 令 $\omega=e^{2\pi i/3}$。求 $\Phi_3(x)$，计算 ω^2，并证明 $N(a+b\omega)=a^2-ab+b^2$ 的乘法性；再说明哪些素数不可能是这种范数。

17.7 ★★ 判断 $45, 90, 105$ 是否为两平方和；若可以，给出表示；若不可以，指出违反了哪一个素因子指数条件。

17.8 ★★★ 完成两平方和定理的下降证明：对最小 m，逐项证明 $(a,b)=1$、m 为奇数，以及取 $q\mid m$ 后如何得到更小倍数。

17.9 ★★★ 设正整数 N 是两个有理数的平方和。证明 N 也是两个整数的平方和；证明中要解释为什么分母不能含 $3 \pmod 4$ 的素因子。

17.10 ★★★ 设 C_n 是 $n+1$ 个变量完全加括号的方式数。由首个乘法位置推导

$$C_{n+1}=\sum_{i=0}^{n} C_i C_{n-i},$$

再求生成函数并写出 C_n 的闭式。

练习答案与提示

17.1 令 $\omega=e^{2\pi i/3}$。把 $(1+x)^n$ 分别代入 $1,\omega,\omega^2$ 后相加，得到 $\sum_k \binom{n}{3k}=[2^n+(1+\omega)^n+(1+\omega^2)^n]/3$。因 $1+\omega=-\omega^2$，按 $n \pmod 6$ 分别为 $(2^n+2)/3$（$n\equiv 0$）、$(2^n-2)/3$（$n\equiv 3$）、$(2^n-1)/3$（$n\equiv 2,4$）、$(2^n+1)/3$（$n\equiv 1,5$）。

17.2 求导得 $\widetilde{A}'(x)=\sum_{n\ge 0} a_{n+1}x^n/n!$。取 $a_n=r^n$，于是 $E'(x)=rE(x)$ 且 $E(0)=1$，解为 $E(x)=e^{rx}=\sum_{n\ge 0} r^n x^n/n!$。

17.3 $5=1^2+2^2$，$13=2^2+3^2$，$17=1^2+4^2$，$29=2^2+5^2$，$37=1^2+6^2$。模 4 时平方只能是 $0,1$；若奇素数是平方和，一个平方为 0、另一个为 1，总和只能是 $1 \pmod 4$。四个待排除素数都为 $3 \pmod 4$。

17.4 展开 $x=(e^x-1)\sum B_n x^n/n!$。比较 x^n（$n\ge 2$）得

$$\sum_{k=0}^{n-1}\binom{n}{k}B_k=0.$$

前七个值为 $1,-1/2,1/6,0,-1/30,0,1/42$。

17.5 恒等式给出乘法性。取

$$65=1^2+8^2,\qquad 8=2^2+2^2,$$

则

$$520=(1^2+8^2)(2^2+2^2)=(-14)^2+18^2=14^2+18^2.$$

而 $15=3\cdot 5$ 不行，因为素数 $3\equiv 3 \pmod 4$ 的指数为 1。

17.6 $\Phi_3(x)=x^2+x+1$，故 $\omega^2=-1-\omega$。若 $z=a+b\omega$，其共轭为 $a+b\omega^2$，乘积为 a^2-ab+b^2。乘法性来自复数乘法：$N(zw)=N(z)N(w)$。若素数 $p\equiv 2 \pmod 3$ 且 $p=N(a+b\omega)$，则模 3 有 $p\equiv (2a-b)^2 \pmod 3$，右边只能是 0 或 1，矛盾。

17.7 $45=3^2+6^2$，$90=3^2+9^2$。$105=3\cdot 5\cdot 7$，素数 $7\equiv 3 \pmod 4$ 的指数为 1，所以不能表示；只要有一个这样的素因子出现奇数次，判别定理就排除它。

17.8 取最小 m 使 $mp = a^2 + b^2$。若 $g = (a,b) > 1$，则 $g^2 \mid mp$；由 $|a|, |b| < p$ 可知 $g < p$，所以 $g \neq p$，从而 $g^2 \mid m$，除以 g^2 就得到更小的系数 m/g^2。若 m 偶，则 a, b 均为奇数，且

$$\left(\frac{a+b}{2}\right)^2 + \left(\frac{a-b}{2}\right)^2 = \frac{m}{2}p,$$

也矛盾。因此 m 为奇数。取奇素因子 $q \mid m$，互素性保证 $q \nmid b$，故 $(ab^{-1})^2 \equiv -1 \pmod q$，于是 $q \equiv 1 \pmod 4$。归纳假设写 $q = c^2 + d^2$；调整 c 的符号使 $q \mid ac - bd$，再由两平方恒等式和模 q 的关系得到 $q \mid ad + bc$。除以 q^2 后得到

$$\left(\frac{ac-bd}{q}\right)^2 + \left(\frac{ad+bc}{q}\right)^2 = \frac{m}{q}p,$$

其系数 $m/q < m$，矛盾。

17.9 写 $N = (a/d)^2 + (b/d)^2$，其中 a, b, d 为整数、$d > 0$，且 $\gcd(a,b,d) = 1$。于是 $Nd^2 = a^2 + b^2$。若素数 $q \equiv 3 \pmod 4$ 整除 d，则 $q \mid a^2 + b^2$，上面的必要性证明迫使 $q \mid a, b$，与互素矛盾。因此 $q \nmid d$。另一方面，q 在 $a^2 + b^2 = Nd^2$ 中的指数为偶数，而 d 不含 q，所以 q 在 N 中的指数也为偶数。对每个 $q \equiv 3 \pmod 4$ 都如此，素因子判别定理于是给出 $N = x^2 + y^2$，其中 x, y 为整数。

17.10 设 $C(x) = \sum_{n \ge 0} C_n x^n$ 且 $C_0 = 1$。按根部最后一次相乘的位置分类，$C_{n+1} = \sum_{i=0}^{n} C_i C_{n-i}$，故 $C(x) = 1 + xC(x)^2$。解二次方程并取常数项为 1 的分支，得

$$C(x) = \frac{1 - \sqrt{1-4x}}{2x}, \qquad C_n = \frac{1}{n+1}\binom{2n}{n}.$$

最后一步可对 $\sqrt{1-4x}$ 用二项式展开：x^n 的系数是 $-\frac{1}{2}\binom{1/2}{n+1}(-4)^{n+1}$。把广义二项式系数的分子 $\frac{1}{2}(\frac{1}{2}-1)\cdots(\frac{1}{2}-n)$ 中的奇数因子配对整理，便得到 $(2n)!/[n!(n+1)!]$，即上式的闭式。

第 18 讲

连分数（一）：欧几里得算法与渐近分数

Continued Fractions I

对应材料：MIT 18.781 Lectures 18–19（Continued Fractions I；收敛与误差部分）；习题集 8 第 7 题；习题集 9 第 1、6、9 题（改编）。

本讲目标

(1) 会把任意有理数写成有限简单连分数，看清它和欧几里得算法记录的是同一份信息；知道每个有理数恰好有两种写法。

(2) 通过 $\varphi=[1,1,1,\dots]$ 与 $\sqrt{2}=[1,2,2,\dots]$ 体会无限连分数，知道有理数对应有限展开，并认识周期展开与二次无理数的关系将在第 20 讲证明。

(3) 掌握渐近分数 p_k/q_k 的递推公式，会用一张三行的表快速算出全部渐近分数。

(4) 掌握行列式恒等式 $p_{k-1}q_k-q_{k-1}p_k=(-1)^k$ 及其矩阵证明，会用它求解 $ax+by=1$、求模逆元，并推出 $\gcd(p_k,q_k)=1$。

(5) 证明无限简单连分数总收敛：偶数项渐近分数递增、奇数项递减、两边夹出极限，误差小于 $1/q_k^2$；并由此推出无限简单连分数的值是无理数。

18.1 从欧几里得算法到连分数

我们习惯用小数表示实数，比如 $\frac{415}{93}=4.462365\dots$。可是小数依赖「十进制」这个人为的选择：换成二进制，同一个数的数字就完全不同。有没有一种只依赖数本身的写法？

连分数给出的答案是：**反复「取整数部分，把剩下的小数部分倒过来」**。下面先用一个具体的分数演示。

例 18.1　$\frac{415}{93}$ 的连分数　把 $\dfrac{415}{93}$ 写成「整数 + 倒数」的嵌套形式。

解　$415=4\cdot 93+43$，所以 $\dfrac{415}{93}=4+\dfrac{43}{93}=4+\dfrac{1}{93/43}$。对 $\frac{93}{43}$ 重复同样的操作：$93=2\cdot 43+7$，$\frac{93}{43}=2+\frac{1}{43/7}$。再来：$43=6\cdot 7+1$，$\frac{43}{7}=6+\frac{1}{7}$。最后 $\frac{7}{1}=7$ 是整数，过程停止。一层层代回去：

$$\frac{415}{93}=4+\cfrac{1}{2+\cfrac{1}{6+\cfrac{1}{7}}}.$$

把每一步的除法并排写出来，就是第 2 讲求 $\gcd(415,93)$ 的欧几里得算法：

$$\begin{aligned} 415 &= \mathbf{4}\cdot 93 + 43,\\ 93 &= \mathbf{2}\cdot 43 + 7,\\ 43 &= \mathbf{6}\cdot 7 + 1,\\ 7 &= \mathbf{7}\cdot 1 + 0. \end{aligned}$$

连分数中的部分商，正是这些除法步骤的商：$4, 2, 6, 7$。

最后的非零余数是 1，所以 $\gcd(415,93) = 1$。 ♦

这种嵌套写法太占地方，我们引入记号。

定义 18.1（连分数）　对实数 a_0 和正实数 $a_1,\dots,a_n$，记

$$[a_0,a_1,\dots,a_n] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}},$$

称为（有限）**连分数**（continued fraction），a_i 称为**部分商**（partial quotient）。若 $a_0\in\mathbb{Z}$，且 $a_1,\dots,a_n$ 都是正整数，则称它为**简单连分数**（simple continued fraction）。

严格地说，记号是递归定义的：$[a_0] = a_0$，而

$$[a_0,a_1,\dots,a_n] = a_0 + \frac{1}{[a_1,\dots,a_n]} \qquad (n\ge 1).$$

对 n 归纳可知 $[a_1,\dots,a_n]\ge a_1 > 0$（每一层都是「正数加正数的倒数」），所以分母永远不为零，记号总有意义。这就是定义里要求 $a_1,\dots,a_n>0$ 的第一个理由。a_0 不在任何分母里，所以没有限制。

记号 $[a_0,\dots,a_n]$ 可以「从前面拆」，也可以「从后面合」。这两条规则后面会反复用到。

引理 18.2　设 $a_0\in\mathbb{R}$，$a_1,\dots,a_n>0$。

(1) 对 $1\le k\le n$，有 $[a_0,\dots,a_n] = [a_0,\dots,a_{k-1},\,[a_k,\dots,a_n]]$。

(2) 当 $n\ge 1$ 时，$[a_0,\dots,a_{n-1},a_n] = [a_0,\dots,a_{n-2},\,a_{n-1}+\frac{1}{a_n}]$。

证明　(1) 对 k 归纳（对所有满足条件的序列同时归纳）。$k=1$ 就是递归定义。设 $k\ge 2$，对序列 $a_1,\dots,a_n$ 用归纳假设（拆分位置为 $k-1$）：

$$[a_0,\dots,a_n] = a_0 + \frac{1}{[a_1,\dots,a_n]} = a_0 + \frac{1}{[a_1,\dots,a_{k-1},[a_k,\dots,a_n]]} = [a_0,\dots,a_{k-1},[a_k,\dots,a_n]].$$

最后一步又是定义，这里用到 $[a_k,\dots,a_n]>0$。(2) 当 $n=1$ 时就是 $[a_0,a_1] = a_0+\frac{1}{a_1}$；当 $n\ge 2$ 时是 (1) 取 $k=n-1$，因为最后一层 $[a_{n-1},a_n] = a_{n-1}+\frac{1}{a_n}$。 ■

几个小例子帮助熟悉记号：

$$[1,2,3]=1+\cfrac{1}{2+\cfrac{1}{3}}=1+\frac{3}{7}=\frac{10}{7},\qquad [0,1,4]=\cfrac{1}{1+\cfrac{1}{4}}=\frac{4}{5},\qquad [-2,3]=-2+\frac{1}{3}=-\frac{5}{3}.$$

它们都是简单连分数。$[1,\frac{1}{2},2]=1+\frac{1}{1/2+1/2}=2$ 是连分数但不是简单连分数，因为 $\frac{1}{2}$ 不是整数。

想一想　$[1,0,3]$ 不满足定义（$a_1=0$），但硬算的话等于多少？它和 $[4]$ 有什么关系？这说明定义里要求 $a_1,\dots,a_n\ge 1$ 还有什么好处？

答　硬算：$[0,3]=0+\frac{1}{3}=\frac{1}{3}$，$[1,0,3]=1+\frac{1}{1/3}=4=[4]$。一般地，$a+\cfrac{1}{0+\cfrac{1}{t}}=a+t$，所以**一个 0 会把左右两个部分商「粘」成一个**：$[\dots,a,0,b,\dots]=[\dots,a+b,\dots]$。如果允许 0，同一个数就会有无穷多种写法，例如 $4=[1,0,3]=[2,0,2]=[1,0,1,0,2]=\cdots$。要求 $a_i\ge 1$ 排除了这种冗余，才有下面的唯一性定理。

现在证明本节的主定理：有理数恰好就是能写成有限简单连分数的数，而且写法可以由欧几里得算法直接读出。

定理 18.3　每个有限简单连分数都是有理数。反之，设 a 为正整数，b 为整数，对 b,a 做欧几里得算法得到的商依次为 $a_0,a_1,\dots,a_n$，则

$$\frac{b}{a}=[a_0,a_1,\dots,a_n],$$

其中 $a_0=\lfloor b/a\rfloor$ 可以是任意整数，$a_1,\dots,a_n\ge 1$，并且当 $n\ge 1$ 时 $a_n\ge 2$。

思路　第一句对长度归纳即可。第二句把欧几里得算法的每一步 $r_{i-1}=a_ir_i+r_{i+1}$ 除以 r_i，得到 $\frac{r_{i-1}}{r_i}=a_i+\frac{1}{r_i/r_{i+1}}$，这正是连分数的一层；再用引理 18.2 把各层接起来。

证明　*连分数是有理数。*对长度归纳。长度为 1 时，$[a_0]=a_0\in\mathbb{Q}$。若 $y=[a_1,\dots,a_n]$ 是正有理数，则 $[a_0,\dots,a_n]=a_0+\frac{1}{y}$ 也是有理数。

*有理数是连分数。*记 $r_{-1}=b$，$r_0=a$，做带余除法

$$r_{i-1}=a_ir_i+r_{i+1},\qquad 0\le r_{i+1}<r_i\qquad (i=0,1,2,\dots),$$

直到某一步余数 $r_{n+1}=0$ 为止。余数 $a=r_0>r_1>r_2>\cdots\ge 0$ 严格递减，由良序原理，过程必在有限步内停止。令 $x_i=r_{i-1}/r_i$（$0\le i\le n$），则 $x_0=\frac{b}{a}$，并且

$$x_i=a_i+\frac{r_{i+1}}{r_i}=a_i+\frac{1}{x_{i+1}}\quad (i<n),\qquad x_n=a_n.$$

从后往前归纳：$x_n=[a_n]$；若 $x_{i+1}=[a_{i+1},\dots,a_n]$，则 $x_i=a_i+\frac{1}{x_{i+1}}=[a_i,\dots,a_n]$。取 $i=0$ 即得 $\frac{b}{a}=[a_0,\dots,a_n]$。

部分商的范围。$a_0=\lfloor b/a\rfloor$ 是任意整数。对 $1\le i\le n$，$r_{i-1}>r_i>0$，所以 $a_i=\lfloor r_{i-1}/r_i\rfloor\ge 1$。若 $n\ge 1$，最后一步 $r_{n-1}=a_nr_n$ 且 $r_{n-1}>r_n$，所以 $a_n\ge 2$。 ■

数 $x_i=[a_i,a_{i+1},\dots,a_n]$ 称为第 i 个**完全商**（complete quotient）：它是「从第 i 层往下的全部尾巴」，部分商 a_i 是它的整数部分。

定理 18.3 有一个几何解释。先考虑整数 $b\ge a>0$，把 $b\times a$ 的长方形不断切掉尽可能大的正方形：先切掉 a_0 个边长为 a 的正方形，剩下 $r_1\times a$ 的长条；再切掉 a_1 个边长为 r_1 的正方形；如此继续。**每种尺寸的正方形个数，恰好是一个部分商**。图 18.1 画的是 $45=2\cdot16+13$，$16=1\cdot13+3$，$13=4\cdot3+1$，$3=3\cdot1$，即 $\frac{45}{16}=[2,1,4,3]$。

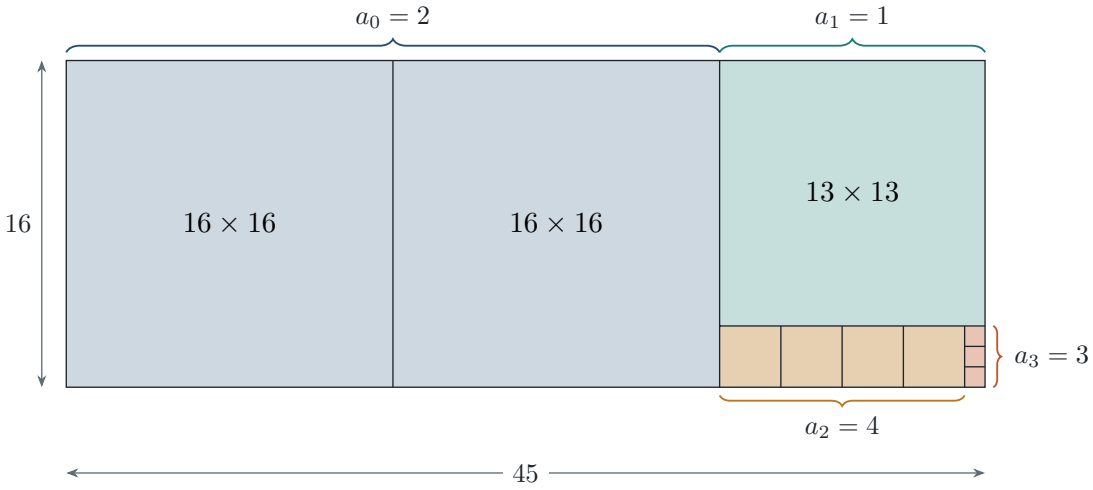


图 18.1　用正方形填满 45×16 的长方形：蓝色有两个 16×16 正方形，青色有一个 13×13 正方形，金色有四个 3×3 正方形，橙色有三个 1×1 正方形。个数 $2,1,4,3$ 给出部分商，最小边长给出 $\gcd(45,16)=1$。

■ **例 18.2**　**习题集 8 第 7 题**　求 $\dfrac{6157}{783}$ 和 $\sqrt{15}$ 的简单连分数；前者从里往外代回验证，后者说明部分商如何重复。

解　做欧几里得算法，把商加粗：

$$6157=\mathbf{7}\cdot783+676,\quad 783=\mathbf{1}\cdot676+107,\quad 676=\mathbf{6}\cdot107+34,\quad 107=\mathbf{3}\cdot34+5,$$
$$34=\mathbf{6}\cdot5+4,\qquad 5=\mathbf{1}\cdot4+1,\qquad 4=\mathbf{4}\cdot1+0.$$

所以 $\dfrac{6157}{783}=[7,1,6,3,6,1,4]$，顺便得到 $\gcd(6157,783)=1$。验证时从最后一层开始，每次算 $a_i+1/(\text{上一步结果})$：

$$4\to1+\tfrac14=\tfrac54\to6+\tfrac45=\tfrac{34}{5}\to3+\tfrac{5}{34}=\tfrac{107}{34},$$
$$\to6+\tfrac{34}{107}=\tfrac{676}{107}\to1+\tfrac{107}{676}=\tfrac{783}{676}\to7+\tfrac{676}{783}=\tfrac{6157}{783}.$$

中间结果正是完全商 $x_i=r_{i-1}/r_i$，分子分母是欧几里得算法里相邻的两个余数。

同题还要求展开 $\sqrt{15}$。有

$$\frac{1}{\sqrt{15}-3}=\frac{\sqrt{15}+3}{6}=1+\frac{\sqrt{15}-3}{6},\qquad \frac{1}{(\sqrt{15}+3)/6-1}=\sqrt{15}+3=6+(\sqrt{15}-3).$$

下一步取倒数又回到 $1/(\sqrt{15}-3)$，所以部分商从开头的 3 之后按 $1,6$ 重复：

$$\sqrt{15}=[3,\overline{1,6}]=[3,1,6,1,6,\dots].$$

这里横线表示 $1,6$ 这一块反复出现。对有理数例题，「从里往外」的算法必须先知道全部部分商；第 18.3 节会给出「从左往右」的算法，每读入一个部分商就更新一次。 ♦

■ **例 18.3　负数与倒数**　求 $-\dfrac{415}{93}$ 和 $\dfrac{93}{415}$ 的简单连分数，与 $\dfrac{415}{93}=[4,2,6,7]$ 比较。

解　$-\frac{415}{93}=-4.46\ldots$，带余除法要求余数非负，所以商是 $\lfloor -4.46\ldots\rfloor=-5$，而不是 -4：

$$-415=(-5)\cdot 93+50,\quad 93=1\cdot 50+43,\quad 50=1\cdot 43+7,\quad 43=6\cdot 7+1,\quad 7=7\cdot 1.$$

于是 $-\frac{415}{93}=[-5,1,1,6,7]$。验证：$[6,7]=\frac{43}{7}$，$[1,6,7]=\frac{50}{43}$，$[1,1,6,7]=\frac{93}{50}$，$-5+\frac{50}{93}=-\frac{415}{93}$。前面变了，尾巴 $6,7$ 没变。

对 $\frac{93}{415}$，第一步是 $93=0\cdot 415+93$，商为 0；之后就是对 $415,93$ 做欧几里得算法。所以 $\frac{93}{415}=[0,4,2,6,7]$：**取倒数只是在前面添一个 0**（或者反过来，去掉开头的 0）。 ♦

一般地，若 $x>1$ 且 $x=[a_0,\ldots,a_n]$，则 $\frac{1}{x}=[0,a_0,\ldots,a_n]$，这直接来自定义 $[0,a_0,\ldots,a_n]=0+\frac{1}{x}$。负数时仍先向下取整，再对正的小数部分继续算法，上面的负数例子展示了这一点。

最后讨论唯一性。欧几里得算法给出的规范展开若 $n\geq 1$，末项满足 $a_n\geq 2$；若 $n=0$，a_0 是任意整数。因而它的末项可以「拆开」：$a_n=(a_n-1)+\frac{1}{1}$，所以

$$[a_0,\ldots,a_{n-1},a_n]=[a_0,\ldots,a_{n-1},a_n-1,1].$$

例如 $\frac{415}{93}=[4,2,6,7]=[4,2,6,6,1]$，$4=[4]=[3,1]$。下面的定理说，除此之外再没有别的写法。

> **定理 18.4（唯一性）**　设 $[a_0,\ldots,a_n]=[b_0,\ldots,b_m]$ 是两个简单连分数，并且长度大于 1 时末项 ≥ 2（即 $n\geq 1\Rightarrow a_n\geq 2$，$m\geq 1\Rightarrow b_m\geq 2$）。则 $m=n$，且 $a_i=b_i$ 对所有 i 成立。因此每个有理数恰好有两种简单连分数写法，一种末项 ≥ 2（或长度为 1），另一种末项为 1。

思路　关键观察：在末项 ≥ 2 的约定下，a_0 一定是 x 的整数部分 $\lfloor x\rfloor$。于是两种写法的首项相同；去掉首项后两边的「尾巴」也相等，对长度归纳即可。

证明　记 $x=[a_0,\ldots,a_n]$。先证明：若 $n\geq 1$，则尾巴 $t=[a_1,\ldots,a_n]>1$。当 $n=1$ 时 $t=a_1\geq 2$；当 $n\geq 2$ 时 $t=a_1+\frac{1}{[a_2,\ldots,a_n]}>a_1\geq 1$。于是 $0<\frac{1}{t}<1$，$x=a_0+\frac{1}{t}$ 不是整数，并且 $a_0=\lfloor x\rfloor$。当 $n=0$ 时 $x=a_0$ 是整数。总之：**$a_0=\lfloor x\rfloor$；x 是整数当且仅当 $n=0$**。

现在对 $\min(m,n)$ 归纳。若 x 是整数，则 $m=n=0$，$a_0=x=b_0$。否则 $m,n\geq 1$，且 $a_0=\lfloor x\rfloor=b_0$，从而

$$[a_1,\ldots,a_n]=\frac{1}{x-a_0}=\frac{1}{x-b_0}=[b_1,\ldots,b_m].$$

两边仍是满足末项约定的简单连分数（首项 $a_1,b_1\geq 1$），长度更短。由归纳假设，$n-1=m-1$ 且 $a_i=b_i$（$i\geq 1$）。

最后一句：末项 ≥ 2（或长度为 1）的写法由欧几里得算法给出且唯一；末项为 1 的写法 $[c_0,\ldots,c_{k-1},1]$（$k\geq 1$）由引理 18.2(2) 等于 $[c_0,\ldots,c_{k-2},c_{k-1}+1]$（$k=1$ 时就是 $[c_0+1]$），后者末项 ≥ 2 或长度为 1，所以末项为 1 的写法也被唯一确定。 ■

易错提醒　「x 的连分数」通常指末项 ≥ 2 的那一种，也就是欧几里得算法直接给出的那一种。但另一种写法并非没用：两种写法的长度奇偶性不同，第 18.4 节解 $ax+by=1$ 时，可以借此调整解的符号。比较两个连分数是否相等时，要先把末项统一成同一种约定。

18.2　无限连分数初探

对有理数，「取整数部分、倒过来」的过程一定会停。对无理数 x 呢？若过程在第 n 步停下，就有 $x=[a_0,\dots,a_n]$，由定理 18.3，x 是有理数，矛盾。所以无理数的展开永不停止，得到一个无穷序列 $a_0,a_1,a_2,\dots$。这个算法的细节和性质留到第 19 讲；这一节先看几个例子，体会会出现什么现象。

例 18.4　π 的前几个部分商　用 $\pi=3.14159265358979\dots$ 求前五个部分商。

解　每一步取整数部分，再把小数部分倒过来：

$$\begin{aligned}
\pi &= \mathbf{3}+0.14159265\dots,\\
1/0.14159265\dots &= 7.06251330\cdots = \mathbf{7}+0.06251330\dots,\\
1/0.06251330\dots &= \mathbf{15}+0.99659440\dots,\\
1/0.99659440\dots &= 1.00341723\cdots = \mathbf{1}+0.00341723\dots,\\
1/0.00341723\dots &= \mathbf{292}+0.63459\dots.
\end{aligned}$$

所以 $\pi=[3,7,15,1,292,\dots]$。继续算下去（需要更多位的 π）得到

$$\pi=[3,7,15,1,292,1,1,1,2,1,3,1,\dots],$$

列出的前几项没有呈现出像 φ、$\sqrt{2}$ 那样的短周期。手算时要小心：每取一次倒数，小数部分的误差就会被放大，第 k 步的放大倍数大约是 a_k^2。用有限位小数计算时，后续部分商需要更高精度才能可靠确定。 ♦

与 π 相比，下面两个数的展开出奇地整齐。

例 18.5　黄金比 φ　求 $\varphi=\dfrac{1+\sqrt{5}}{2}=1.6180339\dots$ 的连分数。

解　φ 是 $x^2=x+1$ 的正根。两边除以 x，得 $\varphi=1+\frac{1}{\varphi}$，也就是

$$\varphi-1=\frac{1}{\varphi},\qquad \text{即}\qquad \frac{1}{\varphi-1}=\varphi.$$

按算法：$\lfloor\varphi\rfloor=1$，小数部分 $\varphi-1=0.618\dots$，倒过来得 $\frac{1}{\varphi-1}=\varphi$，又回到了 φ 本身。于是下一步的整数部分还是 1，再下一步还是 φ……永远循环：

$$\varphi=1+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\cdots}}}=[1,1,1,1,\dots].$$

♦

■ **例 18.6**　$\sqrt{2}$　求 $\sqrt{2}=1.41421356\ldots$ 的连分数。

解　$\lfloor\sqrt{2}\rfloor=1$，小数部分 $\sqrt{2}-1$。倒过来时分母有理化：

$$\frac{1}{\sqrt{2}-1}=\frac{\sqrt{2}+1}{(\sqrt{2}-1)(\sqrt{2}+1)}=\sqrt{2}+1=2.414\ldots.$$

整数部分为 2，小数部分 $(\sqrt{2}+1)-2=\sqrt{2}-1$，与上一步的小数部分相同。所以接下来每一步都得到 $\sqrt{2}+1$，部分商永远是 2：

$$1+\sqrt{2}=2+\frac{1}{1+\sqrt{2}}=[2,2,2,\ldots],\qquad \sqrt{2}=[1,2,2,2,\ldots].$$

检验一下：截断 $[1,2]=\frac{3}{2}=1.5$，$[1,2,2]=\frac{7}{5}=1.4$，$[1,2,2,2]=\frac{17}{12}=1.4167$，$[1,2,2,2,2]=\frac{41}{29}=1.41379$，确实一大一小地逼近 $\sqrt{2}=1.41421$。♦

直观理解　两个例子的共同点是「自相似」：去掉开头几项后，尾巴的值和原来的某个完全商一模一样，所以部分商只能循环。对 φ，尾巴 $[a_1,a_2,\ldots]$ 就是 φ 自己；对 $\sqrt{2}$，从第二项开始的尾巴都是 $1+\sqrt{2}$。沿用图 18.1 的切正方形方法，另取一个 $\varphi\times 1$ 的黄金长方形：切掉一个单位正方形后，剩余长方形的长宽比为 $1/(\varphi-1)=\varphi$，因此每一步都只能切一个正方形。

e 的展开也有规律：

$$e=2.718281828\cdots=[2,1,2,1,1,4,1,1,6,1,1,8,1,1,10,\ldots],$$

列出的前几项呈现出三项一组的规律；本讲只记录原材料所列的部分商，不据此证明全展开的一般公式。下表汇总了这些例子。

数	简单连分数	形态	一般规律
$415/93$	$[4,2,6,7]$	有限	有限 $\Longleftrightarrow$ 有理数（定理 18.3、18.16）
φ	$[1,1,1,1,\ldots]$	循环	循环 $\Longleftrightarrow$ 二次无理数（第 20 讲）
$\sqrt{2}$	$[1,2,2,2,\ldots]$	从第二项起循环	同上
e	$[2,1,2,1,1,4,1,1,6,\ldots]$	前几项有规律	本讲不证明一般形式
π	$[3,7,15,1,292,1,1,\ldots]$	本节只列前缀	不由有限前缀判定
$\sqrt[3]{2}$	原材料未列具体展开	待研究的例子	本讲不外推一般结论

表 18.1　几类实数的连分数。「二次无理数」指满足某个整系数二次方程的无理数，如 $\varphi,\sqrt{2}$。

有限连分数与有理数的对应本讲就能完全证明。周期连分数与二次无理数的对应将在第 20 讲证明。MIT Lecture 18 把 $\sqrt[3]{2}$ 作为展开规律尚待研究的例子；这里不补写原材料未列出的部分商，也不把这个例子外推到其他代数数。表 18.1 把这些形态放在一起比较。

易错提醒　上面对 φ 和 $\sqrt{2}$ 的推导是形式的：我们还没有定义无限连分数 $[a_0, a_1, a_2, \dots]$ 等于什么。对无穷表达式做形式运算是危险的。例如令 $S = 1+2+4+8+\cdots$，则 $2S = S-1$，「解得」$S = -1$，荒谬。同样，方程 $x = 1 + \frac{1}{x}$ 有两个根 φ 和 $\frac{1-\sqrt{5}}{2} = -0.618\dots$，凭什么 $[1,1,1,\dots]$ 等于前者？第 18.5 节会把无限连分数定义为截断 $[a_0,\dots,a_k]$ 当 $k \to \infty$ 的极限，证明极限存在，再严格地算出这两个值。

想一想　不用任何收敛性理论，你能说明 $[1,1,1,\dots]$ 不可能等于 $-0.618\dots$ 吗？（假设它确实是截断的极限。）

答　每个截断 $[1,1,\dots,1]$ 都是「1 加一个正数的倒数」，所以 ≥ 1。极限保持不等号，所以极限 ≥ 1，不可能是负数。剩下的问题只是极限是否存在，这正是第 18.5 节要证的。

18.3　渐近分数与递推公式

例 18.2 从里往外计算连分数，必须先知道全部部分商；对无限连分数这根本做不到。我们换个方向：依次计算 $[a_0]$，$[a_0,a_1]$，$[a_0,a_1,a_2]$，……。以 $\frac{415}{93} = [4,2,6,7]$ 为例：

$$[4] = \frac{4}{1}, \qquad [4,2] = 4 + \frac{1}{2} = \frac{9}{2}, \qquad [4,2,6] = 4 + \cfrac{1}{2 + \cfrac{1}{6}} = 4 + \frac{6}{13} = \frac{58}{13}, \qquad [4,2,6,7] = \frac{415}{93}.$$

数值上是 4，4.5，$4.4615\dots$，$4.46237\dots$，一小一大地逼近最后的值。

定义 18.5（渐近分数）　对有限或无限的简单连分数 $[a_0,a_1,a_2,\dots]$，截断 $C_k = [a_0,a_1,\dots,a_k]$ 称为它的第 k 个渐近分数（convergent）。

看上面的分子 $4, 9, 58, 415$ 和分母 $1, 2, 13, 93$：

$$58 = \mathbf{6}\cdot 9 + 4, \quad 13 = \mathbf{6}\cdot 2 + 1; \qquad 415 = \mathbf{7}\cdot 58 + 9, \quad 93 = \mathbf{7}\cdot 13 + 2.$$

新分子 = 新部分商 × 上一个分子 + 上上个分子，分母也一样。对 $9 = \mathbf{2}\cdot 4 + 1$，$2 = \mathbf{2}\cdot 1 + 0$ 同样成立；如果在最前面补上两个「虚拟」的分子 $0, 1$ 和分母 $1, 0$，连 $4 = \mathbf{4}\cdot 1 + 0$，$1 = \mathbf{4}\cdot 0 + 1$ 也成立。这就引出下面的定义。

定义 18.6（渐近分数的分子与分母）　对实数 a_0 和正实数 $a_1, a_2, \dots$，定义

$$\begin{aligned} &p_{-2} = 0, \quad p_{-1} = 1, \quad p_k = a_k p_{k-1} + p_{k-2}, \\ &q_{-2} = 1, \quad q_{-1} = 0, \quad q_k = a_k q_{k-1} + q_{k-2} \qquad (k \geq 0). \end{aligned}$$

由归纳法，$q_0 = 1$，且对 $k \geq 1$ 有 $q_k = a_k q_{k-1} + q_{k-2} > 0$，所以所有 q_k（$k \geq 0$）都是正数。

定理 18.7　设 $a_0 \in \mathbb{R}$，$a_1, a_2, \cdots > 0$。对每个 $k \geq 0$，

$$[a_0, a_1, \ldots, a_k] = \frac{p_k}{q_k}.$$

思路　对 k 归纳。困难在于 $[a_0, \ldots, a_{k+1}]$ 比 $[a_0, \ldots, a_k]$ 多一个部分商，结构不同。办法是用引理 18.2(2) 把最后两个部分商**合并成一个实数** $a_k + \frac{1}{a_{k+1}}$，这样又变回长度为 $k+1$ 的连分数，可以用归纳假设。正因为合并后的部分商一般不是整数，定理才要对实数部分商陈述。

证明　对 k 归纳，归纳命题是「对所有满足条件的序列 $a_0, \ldots, a_k$ 成立」。$k = 0$：$[a_0] = a_0$，而 $p_0 = a_0 \cdot 1 + 0 = a_0$，$q_0 = a_0 \cdot 0 + 1 = 1$。

设命题对 k 成立。给定 $a_0, \ldots, a_{k+1}$，令 $a_k' = a_k + \frac{1}{a_{k+1}}$（$k \geq 1$ 时 $a_k' > 0$）。由引理 18.2(2)，$[a_0, \ldots, a_k, a_{k+1}] = [a_0, \ldots, a_{k-1}, a_k']$。对新序列 $a_0, \ldots, a_{k-1}, a_k'$ 用归纳假设。它的前 k 项与原序列相同，所以 $p_{k-1}, p_{k-2}, q_{k-1}, q_{k-2}$ 不变，只有第 k 项变了：

$$\begin{aligned}
[a_0, \ldots, a_{k+1}] &= \frac{a_k' p_{k-1} + p_{k-2}}{a_k' q_{k-1} + q_{k-2}} \\
&= \frac{(a_k p_{k-1} + p_{k-2}) + p_{k-1}/a_{k+1}}{(a_k q_{k-1} + q_{k-2}) + q_{k-1}/a_{k+1}} \\
&= \frac{p_k + p_{k-1}/a_{k+1}}{q_k + q_{k-1}/a_{k+1}}.
\end{aligned}$$

分子分母同乘 a_{k+1}，得 $\dfrac{a_{k+1}p_k + p_{k-1}}{a_{k+1}q_k + q_{k-1}} = \dfrac{p_{k+1}}{q_{k+1}}$。 ■

证明中的计算其实给出了更多：把最后一个部分商换成任意正实数 t，结果是 p_k, p_{k-1} 与 q_k, q_{k-1} 的「加权组合」。

推论 18.8　对 $k \geq 0$ 和任意实数 $t > 0$，$[a_0, \ldots, a_k, t] = \dfrac{tp_k + p_{k-1}}{tq_k + q_{k-1}}$。特别地，若 $x = [a_0, \ldots, a_n]$，$x_{k+1} = [a_{k+1}, \ldots, a_n]$ 是完全商，则 $x = \dfrac{x_{k+1}p_k + p_{k-1}}{x_{k+1}q_k + q_{k-1}}$。

证明　对序列 $a_0, \ldots, a_k, t$ 用定理 18.7，它的第 $k+1$ 项是 t，所以 $p_{k+1} = tp_k + p_{k-1}$，$q_{k+1} = tq_k + q_{k-1}$。第二句由引理 18.2(1) 得 $x = [a_0, \ldots, a_k, x_{k+1}]$。 ■

实际计算时，把部分商写在第一行，分子分母写在下面两行，最左边是两列初值，然后从左往右填表（图 18.2）。

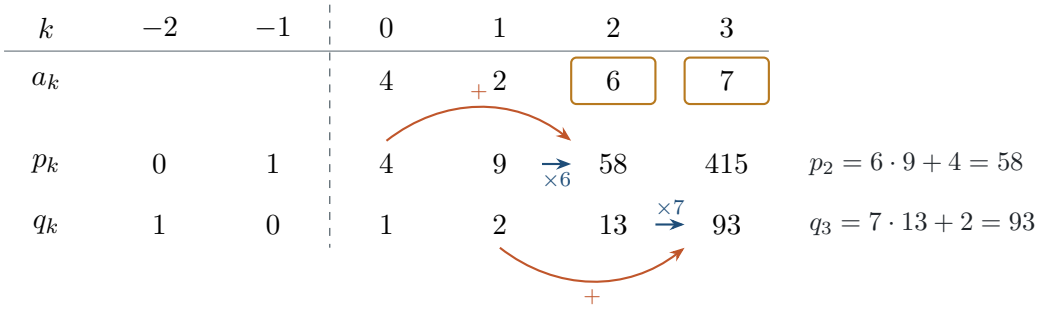


图 18.2　计算渐近分数的三行表（以 $\frac{415}{93} = [4,2,6,7]$ 为例）。虚线左边是初值。每个新数等于「它头上的部分商 × 左边一个数 + 左边第二个数」；图中标出了 p_2 和 q_3 的计算。渐近分数依次为 $\frac{4}{1}, \frac{9}{2}, \frac{58}{13}, \frac{415}{93}$。

■ **例 18.7**　用三行表计算 $\frac{6157}{783} = [7,1,6,3,6,1,4]$ 的全部渐近分数。

解

k	-2	-1	0	1	2	3	4	5	6
a_k			7	1	6	3	6	1	4
p_k	0	1	7	8	55	173	1093	1266	6157
q_k	1	0	1	1	7	22	139	161	783

例如 $p_3 = 3\cdot 55 + 8 = 173$，$q_4 = 6\cdot 22 + 7 = 139$，$p_6 = 4\cdot 1266 + 1093 = 6157$。最后一列回到了 $\frac{6157}{783}$，与定理 18.7 一致。与例 18.2 相比，这里从左往右一遍算完，不需要回头。 ♦

■ **例 18.8　π 的渐近分数**　由 $\pi = [3,7,15,1,292,\dots]$ 求前五个渐近分数。

解

k	-2	-1	0	1	2	3	4
a_k			3	7	15	1	292
p_k	0	1	3	22	333	355	103993
q_k	1	0	1	7	106	113	33102

其中 $p_2 = 15\cdot 22+3 = 333$，$p_3 = 333+22 = 355$，$q_3 = 106+7 = 113$，$p_4 = 292\cdot 355+333 = 103993$，$q_4 = 292\cdot 113 + 106 = 33102$。数值上

$$\frac{22}{7} = 3.142857\dots, \qquad \frac{333}{106} = 3.141509\dots, \qquad \frac{355}{113} = 3.14159292\dots, \qquad \pi = 3.14159265\dots.$$

$\frac{22}{7}$ 和 $\frac{355}{113}$ 就是祖冲之的「约率」和「密率」。$\frac{355}{113}$ 的绝对误差小于 3×10^{-7}，原因要到第 18.5 节才能看清。 ♦

■ **例 18.9　φ 的渐近分数与斐波那契数**　求 $\varphi = [1,1,1,\dots]$ 的渐近分数。

解　所有 $a_k = 1$，递推变成 $p_k = p_{k-1} + p_{k-2}$，$q_k = q_{k-1} + q_{k-2}$，这正是斐波那契递推。初值 $p_{-2} = 0$，$p_{-1} = 1$ 与 $q_{-2} = 1$，$q_{-1} = 0$ 分别是 F_0, F_1 与 F_{-1}, F_0（约定 $F_0 = 0$，$F_1 = 1$，$F_{-1} = 1$），

所以

$$p_k = F_{k+2}, \qquad q_k = F_{k+1}, \qquad C_k = \frac{F_{k+2}}{F_{k+1}} : \quad \frac{1}{1}, \frac{2}{1}, \frac{3}{2}, \frac{5}{3}, \frac{8}{5}, \frac{13}{8}, \frac{21}{13}, \frac{34}{21}, \cdots$$

由第 15、16 讲的 Binet 公式 $F_n = \frac{\varphi^n - \psi^n}{\sqrt{5}}$（$\psi = \frac{1-\sqrt{5}}{2}$，$|\psi| < 1$）可知 $F_{k+2}/F_{k+1} \to \varphi$。所以 $[1,1,1,\ldots]$ 的截断确实收敛到 φ，而不是 $\psi = -0.618\ldots$。第 18.5 节会对所有简单连分数证明收敛性，不需要 Binet 公式。 ♦

■ **例 18.10　$\sqrt{2}$ 的渐近分数**　求 $\sqrt{2} = [1,2,2,2,\ldots]$ 的渐近分数，并计算 $p_k^2 - 2q_k^2$。

解　三行表：

k	-2	-1	0	1	2	3	4	5	6	7
a_k			1	2	2	2	2	2	2	2
p_k	0	1	1	3	7	17	41	99	239	577
q_k	1	0	1	2	5	12	29	70	169	408
$p_k^2 - 2q_k^2$			-1	1	-1	1	-1	1	-1	1

例如 $17^2 - 2\cdot 12^2 = 289 - 288 = 1$，$41^2 - 2\cdot 29^2 = 1681 - 1682 = -1$。看起来 $p_k^2 - 2q_k^2 = (-1)^{k+1}$。

证明：先用归纳法验证 $p_k = p_{k-1} + 2q_{k-1}$，$q_k = p_{k-1} + q_{k-1}$（$k \geq 1$）。$k = 1$：$3 = 1 + 2$，$2 = 1 + 1$。若对 k 成立，则 $p_k + p_{k-1} = 2q_k$，$q_k + q_{k-1} = p_k$，于是

$$p_{k+1} = 2p_k + p_{k-1} = p_k + 2q_k, \qquad q_{k+1} = 2q_k + q_{k-1} = q_k + p_k.$$

代入得 $p_k^2 - 2q_k^2 = (p_{k-1} + 2q_{k-1})^2 - 2(p_{k-1} + q_{k-1})^2 = -(p_{k-1}^2 - 2q_{k-1}^2)$，每走一步变一次号，而 $p_0^2 - 2q_0^2 = -1$。 ♦

前后联系　$p^2 - 2q^2 = \pm 1$ 是 Pell 方程的一个实例；第 21 讲会继续研究这类方程与连分数的关系。这里先用本讲已证的恒等式推出误差公式：因式分解得

$$\frac{p_k}{q_k} - \sqrt{2} = \frac{p_k^2 - 2q_k^2}{q_k(p_k + q_k\sqrt{2})} = \frac{\pm 1}{q_k(p_k + q_k\sqrt{2})},$$

所以误差约为 $\frac{1}{2\sqrt{2}\,q_k^2}$。例如 $\frac{99}{70} - \sqrt{2} = 7.215 \times 10^{-5}$，而 $\frac{1}{70(99+70\sqrt{2})} = 7.215 \times 10^{-5}$。

想一想　$\frac{415}{93}$ 还可以写成 $[4,2,6,6,1]$。用这种写法填三行表，最后一列还是 $\frac{415}{93}$ 吗？倒数第二列是什么？

答　表为 p：$4, 9, 58, 357, 415$；q：$1, 2, 13, 80, 93$，其中 $357 = 6\cdot 58 + 9$，$80 = 6\cdot 13 + 2$，$415 = 1\cdot 357 + 58$，$93 = 1\cdot 80 + 13$。最后一列仍是 $\frac{415}{93}$（两种写法值相同，且下一节会证明 p_k, q_k 互素，所以分子分母也相同），但倒数第二列从 $\frac{58}{13}$ 变成了 $\frac{357}{80}$。所以「倒数第二个渐近分数」依赖于选哪种写法，下一节会用到这一点。

18.4　行列式恒等式及其推论

回到 $\frac{415}{93}$ 的三行表，计算相邻两列组成的 2×2 行列式 $p_{k-1}q_k - q_{k-1}p_k$：

$$\begin{vmatrix}0 & 1\\ 1 & 0\end{vmatrix} = -1,\quad \begin{vmatrix}1 & 4\\ 0 & 1\end{vmatrix} = 1,\quad \begin{vmatrix}4 & 9\\ 1 & 2\end{vmatrix} = -1,\quad \begin{vmatrix}9 & 58\\ 2 & 13\end{vmatrix} = 1,\quad \begin{vmatrix}58 & 415\\ 13 & 93\end{vmatrix} = -1.$$

（最后一个：$58\cdot 93 - 13\cdot 415 = 5394 - 5395 = -1$。）结果只有 ± 1，并且交替出现。

定理 18.9（行列式恒等式）　对任意实数 a_0 和正实数 $a_1, a_2, \ldots$，以及每个 $k \geq -1$，

$$p_{k-1}q_k - q_{k-1}p_k = (-1)^k.$$

证法一（归纳）　$k=-1$：$p_{-2}q_{-1} - q_{-2}p_{-1} = 0\cdot 0 - 1\cdot 1 = -1$。设对 k 成立，用递推式替换 p_{k+1}, q_{k+1}：

$$p_kq_{k+1} - q_kp_{k+1} = p_k(a_{k+1}q_k + q_{k-1}) - q_k(a_{k+1}p_k + p_{k-1}) = p_kq_{k-1} - q_kp_{k-1} = -(-1)^k = (-1)^{k+1}.$$

含 a_{k+1} 的两项正好抵消，这就是结论与部分商无关的原因。 ∎

第二个证明更有启发性：递推式其实是矩阵乘法。

证法二（矩阵）　递推 $p_k = a_kp_{k-1} + p_{k-2}$，$q_k = a_kq_{k-1} + q_{k-2}$ 可以写成

$$\begin{pmatrix}p_k & p_{k-1}\\ q_k & q_{k-1}\end{pmatrix} = \begin{pmatrix}p_{k-1} & p_{k-2}\\ q_{k-1} & q_{k-2}\end{pmatrix}\begin{pmatrix}a_k & 1\\ 1 & 0\end{pmatrix}.$$

初值矩阵 $\begin{pmatrix}p_{-1} & p_{-2}\\ q_{-1} & q_{-2}\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$ 是单位矩阵（这就是初值那样取的原因），所以

$$\begin{pmatrix}p_k & p_{k-1}\\ q_k & q_{k-1}\end{pmatrix} = \begin{pmatrix}a_0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}a_1 & 1\\ 1 & 0\end{pmatrix}\cdots\begin{pmatrix}a_k & 1\\ 1 & 0\end{pmatrix}.$$

每个因子的行列式是 -1，两边取行列式得 $p_kq_{k-1} - p_{k-1}q_k = (-1)^{k+1}$，即 $p_{k-1}q_k - q_{k-1}p_k = (-1)^k$。 ∎

对 $\frac{415}{93}$：

$$\begin{pmatrix}4 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}2 & 1\\ 1 & 0\end{pmatrix} = \begin{pmatrix}9 & 4\\ 2 & 1\end{pmatrix},$$
$$\begin{pmatrix}9 & 4\\ 2 & 1\end{pmatrix}\begin{pmatrix}6 & 1\\ 1 & 0\end{pmatrix} = \begin{pmatrix}58 & 9\\ 13 & 2\end{pmatrix},$$
$$\begin{pmatrix}58 & 9\\ 13 & 2\end{pmatrix}\begin{pmatrix}7 & 1\\ 1 & 0\end{pmatrix} = \begin{pmatrix}415 & 58\\ 93 & 13\end{pmatrix}.$$

三行表的相邻两列，正是这些矩阵乘积的两列。

推论 18.10（倒序读部分商）　对简单连分数和 $k \geq 1$，

$$\frac{q_k}{q_{k-1}} = [a_k, a_{k-1}, \ldots, a_1].$$

若另有 $a_0 > 0$，则 $p_k/p_{k-1} = [a_k, a_{k-1}, \ldots, a_0]$。

证明　分母递推给出 $q_k/q_{k-1} = a_k + 1/(q_{k-1}/q_{k-2})$，逐层向前拆，直到 $q_1/q_0 = a_1$，得到第一式。分子递推完全相同；当 $a_0 > 0$ 时所有相关分子为正，可拆到 $p_0/p_{-1} = a_0$，得到第二式。例如 $93/13 = [7,6,2]$、$415/58 = [7,6,2,4]$，可直接从三行表末两列读出。若 $a_0 = 0$，则 $p_0 = 0$，不能照此除法，这就是第二式另加条件的原因。 ■

核心思想　矩阵 $\begin{pmatrix} a & 1 \\ 1 & 0 \end{pmatrix}$ 对应分式线性变换 $t \mapsto \frac{at+1}{t} = a + \frac{1}{t}$，也就是连分数的「一层」。矩阵相乘对应变换的复合，所以 $k+1$ 层连分数就是 $k+1$ 个矩阵之积。推论 18.8 的公式 $[a_0, \ldots, a_k, t] = \frac{tp_k + p_{k-1}}{tq_k + q_{k-1}}$ 就是乘积矩阵作用在 t 上。

行列式恒等式有一串推论。下面设 $a_0 \in \mathbb{Z}$，$a_1, a_2, \ldots$ 是正整数，此时所有 p_k, q_k 都是整数。

推论 18.11　相邻渐近分数之差为

$$\frac{p_{k-1}}{q_{k-1}} - \frac{p_k}{q_k} = \frac{(-1)^k}{q_{k-1}q_k} \qquad (k \geq 1).$$

证明　通分：左边 $= \dfrac{p_{k-1}q_k - q_{k-1}p_k}{q_{k-1}q_k}$，分子由定理 18.9 等于 $(-1)^k$。 ■

推论 18.12　$\gcd(p_k, q_k) = 1$。所以渐近分数 $\frac{p_k}{q_k}$ 自动是既约分数。

证明　若 $d \mid p_k$ 且 $d \mid q_k$，则 d 整除整系数组合 $p_{k-1}q_k - q_{k-1}p_k = \pm 1$，所以 $d = \pm 1$。 ■

行列式还可以读作面积。对 $[0,1,1]$，相邻渐近分数是 $C_1 = 1/1$、$C_2 = 1/2$，对应两个整数向量 $u = (1,1)$、$v = (1,2)$。图 18.3 画出它们张成的平行四边形；面积为 $|1 \cdot 2 - 1 \cdot 1| = 1$。

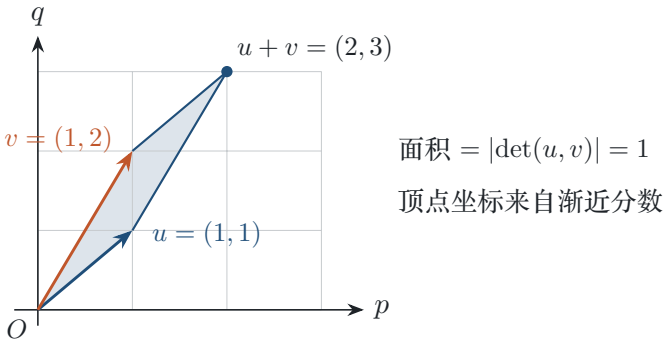


图 18.3　把相邻渐近分数的分子、分母作为向量坐标，得到面积为 1 的平行四边形；行列式的符号记录两列向量的次序。

这里用的是坐标 (p,q)；在这张图里 p/q 等于横坐标除以纵坐标。一般的相邻两列也给出面积 1。互素性则仍按上面的整除证明理解：若一列的两个坐标有公因数 $d>1$，整个整数行列式就会被 d 整除，不可能等于 ±1。

■ **例 18.11　从相邻列直接读出 Bézout 系数**　由 $[4,2,6]=58/13$ 的渐近分数求 $58u+13v=1$ 的一组整数解，并求 58 模 13 的逆元。

解　取最后两列 $(9,2)$ 和 $(58,13)$。行列式给出 $9\cdot13-2\cdot58=1$，所以 $(u,v)=(-2,9)$；回代为 $-116+117=1$。模 13 化简，$58(-2)\equiv1\pmod{13}$，故逆元也可写成 11。这把第 2 讲的 Bézout 等式直接读成了连分数表的相邻行列式。 ♦

18.5　无限连分数的收敛与误差

前面用 $\sqrt{2}$ 的截断看见了「一低一高」地逼近。现在证明这不是数值巧合，而是行列式恒等式的直接结果。

引理 18.13（分母趋于无穷）　若 $a_1,a_2,\dots$ 是正整数，则渐近分数分母 q_k 趋于无穷；并且 $q_{k+1}>q_k$（$k\ge1$）。

证明　$q_0=1$、$q_1=a_1\ge1$，而对 $k\ge2$，递推给出 $q_k=a_kq_{k-1}+q_{k-2}\ge q_{k-1}+q_{k-2}$。因此 $q_2>q_1$，之后严格递增；同时它至少按 Fibonacci 递推增长，故 $q_k\to\infty$。 ■

定理 18.14（奇偶渐近分数夹逼）　设 $a_0\in\mathbb{Z}$，$a_1,a_2,\dots$ 是正整数，令 $C_k=p_k/q_k=[a_0,\dots,a_k]$。则

$$C_0<C_2<C_4<\cdots<C_5<C_3<C_1,$$

更准确地说，偶数下标的 C_{2j} 严格递增，奇数下标的 C_{2j+1} 严格递减，且每个偶数项都小于每个奇数项。两列有相同极限 x，所以 $C_k\to x$。我们把这个极限记作无限连分数 $[a_0,a_1,a_2,\dots]$。

证明　由推论 18.11，

$$C_{k-1} - C_k = \frac{(-1)^k}{q_{k-1}q_k}.$$

右边符号随 k 奇偶交替，且不为 0。还需检查隔两项的差：由递推式和行列式恒等式，

$$C_{k+2} - C_k = \frac{a_{k+2}(p_{k+1}q_k - p_k q_{k+1})}{q_k q_{k+2}} = \frac{(-1)^k a_{k+2}}{q_k q_{k+2}}.$$

因为 $a_{k+2} > 0$，偶数子列严格递增，奇数子列严格递减。相邻差又说明 $C_{2j} < C_{2j+1}$ 以及 $C_{2j+2} < C_{2j+1}$，所以区间 $[C_{2j}, C_{2j+1}]$ 逐个嵌套。两列分别单调有界，故有极限；又

$$0 < C_{2j+1} - C_{2j} = \frac{1}{q_{2j}q_{2j+1}} \longrightarrow 0,$$

所以两列极限相同。交错夹逼遂给出整个 C_k 序列收敛。 ∎

图 18.4 把平方根 $\sqrt{2}$ 的几次夹逼区间依次放大到相同长度。每行的两个端点都是相邻渐近分数；青色竖线标出 $\sqrt{2}$ 的位置。端点越来越近，奇数项从上方、偶数项从下方夹住同一个极限。

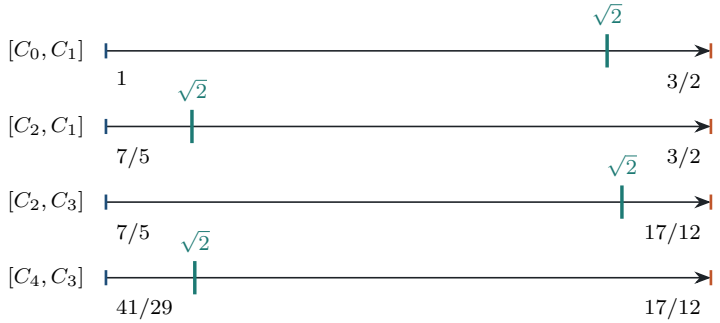


图 18.4　相邻渐近分数给出嵌套夹逼。每行区间都独立缩放到相同长度，竖线在行内的位置按实际数值计算。

各行的绝对长度不同，图中特意单独缩放，不能用不同横行的线段长度互相比较。读者应比较每一行内部的位置：第一个区间中 $\sqrt{2}$ 靠近右端；加入后续部分商后，夹逼区间变窄，端点交替落在极限两侧。

定理 18.15（渐近分数的误差）　令 $x = [a_0, a_1, a_2, \ldots]$。对每个 $k \geq 0$，x 严格位于 C_k 与 C_{k+1} 之间，且

$$\left| x - \frac{p_k}{q_k} \right| < \frac{1}{q_k q_{k+1}} \leq \frac{1}{q_k^2}.$$

特别地，偶数渐近分数从下方逼近 x，奇数渐近分数从上方逼近 x。

证明　固定 k，记尾巴 $t = [a_{k+1}, a_{k+2}, \ldots]$。该尾巴也由上一条定理收敛，且 $t > a_{k+1}$。由推论 18.8（对有限尾巴取极限），

$$x = \frac{tp_k + p_{k-1}}{tq_k + q_{k-1}}.$$

相减并使用行列式恒等式：

$$x - \frac{p_k}{q_k} = \frac{p_{k-1}q_k - q_{k-1}p_k}{q_k(tq_k + q_{k-1})} = \frac{(-1)^k}{q_k(tq_k + q_{k-1})}.$$

因 $t > a_{k+1}$，分母中的第二因子大于 $a_{k+1}q_k + q_{k-1} = q_{k+1}$，得第一条误差界。上式的符号是 $(-1)^k$，而误差绝对值小于 $|C_{k+1} - C_k| = 1/(q_k q_{k+1})$，所以 x 严格位于 C_k 与 C_{k+1} 之间。$q_{k+1} \geq q_k$，所以第二条成立；$k = 0$ 且 $q_1 = q_0$ 时第一条仍为严格不等式。 ■

交错不仅说明从哪一边逼近，也能定量回答「分母为 q_k 时误差有多大」。例如对 $\sqrt{2}$，例 18.10 中的 $99/70$ 是上方渐近分数，而 $239/169$ 从下方逼近；两者分别受 $1/(70 \cdot 169)$ 与 $1/(169 \cdot 408)$ 控制。

定理 18.16（无限简单连分数是无理数）　若简单连分数有无穷多个正整数部分商，则其值 $x = [a_0, a_1, \ldots]$ 是无理数。

证明　反设 $x = A/B$，其中 A, B 互素且 $B > 0$。行列式恒等式说明每个 p_k/q_k 都是既约分数，而此前已经证明 x 严格位于相邻渐近分数之间，所以每个 C_k 都不等于 x。对充分大的 k，两个不同的分数之差至少为 $1/(Bq_k)$；但定理 18.15 给出

$$0 < \left| x - \frac{p_k}{q_k} \right| < \frac{1}{q_k^2}.$$

当 $q_k > B$ 时，$1/(Bq_k) > 1/q_k^2$，矛盾。因此 x 无理。 ■

前后联系　本讲把欧几里得算法的商变成了一个逼近实数的递推过程：行列式恒等式保证相邻渐近分数互素，误差定理保证它们从两侧快速逼近。第 19 讲会继续研究这种逼近的最优性，第 20 讲再证明「循环连分数恰好来自二次无理数」。

本讲小结

- 有理数的简单连分数由欧几里得算法的商唯一确定；末项可以拆成 $a_n - 1, 1$，所以每个有理数有两种写法。
- 渐近分数满足 $p_k = a_k p_{k-1} + p_{k-2}$、$q_k = a_k q_{k-1} + q_{k-2}$，并且 $[a_0, \ldots, a_k] = p_k/q_k$。
- 行列式恒等式为 $p_{k-1}q_k - q_{k-1}p_k = (-1)^k$，因此 $\gcd(p_k, q_k) = 1$，相邻渐近分数之差为 $(-1)^k/(q_{k-1}q_k)$。
- 无限简单连分数的偶数项递增、奇数项递减，两列夹出同一个极限；这一定义了无限连分数的值。
- 若 $x = [a_0, a_1, \ldots]$，则 $|x - p_k/q_k| < 1/(q_k q_{k+1}) \leq 1/q_k^2$，偶数项从下方、奇数项从上方逼近。
- 无穷简单连分数的值必为无理数；φ 与 $\sqrt{2}$ 展示了循环及最终循环的部分商序列。

练习

18.1 ★　用欧几里得算法求 $43/18$ 的简单连分数，并从最后一项开始代回验证。

18.2 ★　对 $[4, 2, 6, 7]$ 填写三行表，列出所有 p_k, q_k，并核对每个相邻行列式的符号。

18.3 ★★ 证明 $[a_0,\dots,a_{n-1},a_n]=[a_0,\dots,a_{n-1},a_n-1,1]$（$a_n\ge 2$），并说明这给出有理数的第二种写法。

18.4 ★★ 用三行表计算 $\pi=[3,7,15,1,292,\dots]$ 的前五个渐近分数，并比较 $22/7$ 与 $355/113$ 的误差。

18.5 ★★ 设 $x=[a_0,a_1,\dots]$ 是无限简单连分数。用行列式恒等式和误差界证明 x 不可能是有理数。

18.6 ★★★ 设 m 为正整数，且 $x=[m,m,m,\dots]$。求 x；写出 p_k,q_k 的递推和闭式，并证明 $p_k/q_k\to x$。

18.7 ★★ 对 $\sqrt{2}=[1,2,2,\dots]$，证明 $p_k^2-2q_k^2=(-1)^{k+1}$，并用它给出 $99/70$ 与 $239/169$ 的误差公式。

18.8 ★★ 验证 $22/7$、$333/106$、$355/113$ 分别是 π 的渐近分数，并检查它们满足定理 18.15 的上界。

18.9 ★★★ 设 13 为素数，$5^2\equiv -1 \pmod{13}$。对 $5/13$ 做连分数，取最大满足 $q_i\le\sqrt{13}$ 的渐近分数，构造 $x^2+y^2=13$。

18.10 ★★ 用行列式恒等式求一组整数解 (u,v)，使 $415u+93v=1$，并说明为什么这同时给出了 415 模 93 的逆元。

练习答案与提示

18.1 $43=2\cdot 18+7$，$18=2\cdot 7+4$，$7=1\cdot 4+3$，$4=1\cdot 3+1$，$3=3\cdot 1$，所以 $43/18=[2,2,1,1,3]$。从末项代回：$3\to 4/3\to 7/4\to 18/7\to 43/18$。

18.2 初值为 $(p_{-2},p_{-1})=(0,1)$、$(q_{-2},q_{-1})=(1,0)$，逐列得到

$$(p_k)=(0,1,4,9,58,415),\qquad (q_k)=(1,0,1,2,13,93).$$

例如 $9\cdot 13-58\cdot 2=1$、$58\cdot 93-415\cdot 13=-1$，正好是 $(-1)^2$、$(-1)^3$ 对应的相邻行列式（按列下标排列）。

18.3 因为 $a_n=(a_n-1)+1/1$，按连分数递归定义有 $[a_0,\dots,a_{n-1},a_n]=[a_0,\dots,a_{n-1},a_n-1,1]$。当 $a_n\ge 2$ 时，右边仍是简单连分数且末项为 1；反过来把末项 1 与前一项合并，就回到欧几里得算法给出的规范写法。

18.4 前五项为 $3/1,22/7,333/106,355/113,103993/33102$。其中 $|\pi-22/7|\approx 1.264\times 10^{-3}$、$|\pi-355/113|\approx 2.668\times 10^{-7}$，且分别小于 $1/(7\cdot 106)$、$1/(113\cdot 33102)$。

18.5 设 $x=A/B$ 为既约有理数。误差定理说明每个渐近分数都与 x 不同，故 $|x-p_k/q_k|\ge 1/(Bq_k)$；但误差定理给出 $|x-p_k/q_k|<1/q_k^2$。当 $q_k>B$ 时两者矛盾，所以 x 无理。

18.6 对 $k\ge 0$，递推为 $p_k=mp_{k-1}+p_{k-2}$、$q_k=mq_{k-1}+q_{k-2}$，初值是 $p_{-2}=0,p_{-1}=1,q_{-2}=1,q_{-1}=0$。方程 $x=m+1/x$ 的正根为 $\alpha=(m+\sqrt{m^2+4})/2$；令 $\beta=(m-\sqrt{m^2+4})/2$，则

$$p_k=\frac{\alpha^{k+2}-\beta^{k+2}}{\alpha-\beta},\qquad q_k=\frac{\alpha^{k+1}-\beta^{k+1}}{\alpha-\beta},$$

α,β 都满足 $t^2=mt+1$，所以闭式满足上述递推与初值。又 $|\beta|<\alpha$，故

$$\frac{p_k}{q_k}=\alpha\frac{1-(\beta/\alpha)^{k+2}}{1-(\beta/\alpha)^{k+1}}\longrightarrow \alpha=x.$$

18.7 $k=1$ 时 $p_1=p_0+2q_0$、$q_1=p_0+q_0$。若 $p_k=p_{k-1}+2q_{k-1}$、$q_k=p_{k-1}+q_{k-1}$，则 $p_k+p_{k-1}=2q_k$、$q_k+q_{k-1}=p_k$，因此由部分商恒为 2 的递推得到下一步同样两式。于是

$$p_k^2-2q_k^2=(p_{k-1}+2q_{k-1})^2-2(p_{k-1}+q_{k-1})^2=-(p_{k-1}^2-2q_{k-1}^2),$$

而 $p_0^2-2q_0^2=-1$，故 $p_k^2-2q_k^2=(-1)^{k+1}$。因此

$$\frac{99}{70}-\sqrt{2}=\frac{1}{70(99+70\sqrt{2})},\qquad \frac{239}{169}-\sqrt{2}=-\frac{1}{169(239+169\sqrt{2})}.$$

18.8 三行表给出 $22/7$、$333/106$、$355/113$，它们的分母下一项分别为 $106,113,33102$。由 $3.1415926535<\pi<3.1415926536$，三项误差分别小于

$$\begin{aligned}\frac{22}{7}-\frac{31415926535}{10^{10}}&<\frac{1}{742},\\ \frac{31415926536}{10^{10}}-\frac{333}{106}&<\frac{1}{106\cdot 113},\\ \frac{355}{113}-\frac{31415926535}{10^{10}}&<\frac{1}{113\cdot 33102}.\end{aligned}$$

这些有理数不等式通分即可验证，右侧正是相应的误差上界。

18.9 $5/13=[0,2,1,1,2]$，渐近分数为 $0/1,1/2,1/3,2/5,5/13$。最大满足 $q_i\le\sqrt{13}$ 的是 $1/3$，取 $x=q_i=3$、$y=p_i\cdot 13-5q_i=13-15=-2$，于是 $x^2+y^2=9+4=13$。

18.10 由相邻列 $58\cdot 93-13\cdot 415=-1$，得 $13\cdot 415-58\cdot 93=1$，所以 $(u,v)=(13,-58)$。模 93 看即 $415\cdot 13\equiv 1\pmod{93}$，故 13 是所求逆元。

第 19 讲

连分数（二）：有理逼近的不等式

Continued Fractions II: Inequalities

对应材料：MIT 18.781 Lecture 19 讲义（Theorems 66–70）；习题集 9 第 1、4、5、9 题（改编）。

本讲目标

(1) 会用渐近分数的行列式恒等式写出相邻误差的精确和，并把它变成误差上界。

(2) 理解并证明：每两个相邻渐近分数中至少有一个满足 $|x-p/q|\le 1/(2q^2)$。

(3) 理解并证明 Hurwitz 型结论：每三个相邻渐近分数中至少有一个满足 $|x-p/q|<1/(\sqrt{5}q^2)$。

(4) 会把结论用于 $\sqrt{2}$、$\sqrt{3}$、π 和黄金分割比，比较不同渐近分数的逼近质量。

(5) 理解 $1/\sqrt{5}$ 为何是不能统一改进的常数，并知道连分数周期性与二次无理数的 Lagrange 定理。

19.1 两个相邻渐近分数之间

第 18 讲中，$\sqrt{2}=[1,2,2,\ldots]$ 的渐近分数从两侧靠近：

$$1,\quad \frac{3}{2},\quad \frac{7}{5},\quad \frac{17}{12},\quad \frac{41}{29},\quad \frac{99}{70},\quad \frac{239}{169},\ldots$$

例如 $41/29<\sqrt{2}<99/70$。相邻两项的距离是

$$\frac{99}{70}-\frac{41}{29}=\frac{1}{70\cdot 29}.$$

这个小数值同时控制两项到 $\sqrt{2}$ 的误差。先把这种观察写成统一记号。

设 x 是无理数，简单连分数的渐近分数记为

$$C_n=\frac{p_n}{q_n}=[a_0,a_1,\ldots,a_n],\qquad q_n>0.$$

第 18 讲的递推为

$$p_n=a_np_{n-1}+p_{n-2},\qquad q_n=a_nq_{n-1}+q_{n-2},$$

其中 $p_{-2}=0,p_{-1}=1,q_{-2}=1,q_{-1}=0$。行列式恒等式给出

$$p_{n-1}q_n-q_{n-1}p_n=(-1)^n. \tag{19.1}$$

命题 19.1（相邻项的精确距离）　对 $n \geq 1$，有

$$|C_n - C_{n-1}| = \frac{1}{q_n q_{n-1}}.$$

并且 C_n 与 C_{n-1} 分居 x 的两侧，所以

$$|x - C_n| + |x - C_{n-1}| = |C_n - C_{n-1}| = \frac{1}{q_n q_{n-1}}.$$

证明　通分并使用式 (19.1)：

$$C_{n-1} - C_n = \frac{p_{n-1}q_n - q_{n-1}p_n}{q_{n-1}q_n} = \frac{(-1)^n}{q_{n-1}q_n}.$$

第 18 讲已经证明渐近分数交错夹住极限 x，所以 x 位于这两个端点之间。线段上的距离满足「两段之和等于整段长度」，得到第二式。 ■

例如对 $\sqrt{2}$ 的 $41/29$ 与 $99/70$，式 (19.1) 给出

$$\frac{99}{70} - \frac{41}{29} = \frac{29 \cdot 99 - 70 \cdot 41}{70 \cdot 29} = \frac{1}{70 \cdot 29}.$$

图 19.1 把这两个相邻渐近分数及 $\sqrt{2}$ 画在数轴上；偶数项从左侧靠近，奇数项从右侧靠近。

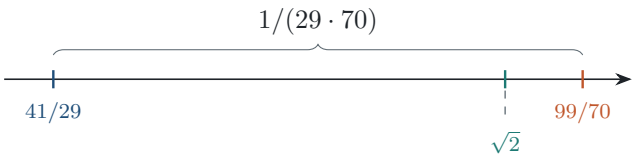


图 19.1　将区间 $[41/29, 99/70]$ 等比例放大后，$\sqrt{2}$ 约在从左端起 85.35% 的位置；端点距离为 $1/(29 \cdot 70)$。

整段宽度由分母乘积确定，但 $\sqrt{2}$ 并不在中点：它到 $99/70$ 的距离更小。两项误差各自多大仍取决于 x 在区间内的位置；下一节用它们的和证明至少一个误差很小。

例 19.1　$\sqrt{2}$ 的相邻误差　比较 $41/29$ 和 $99/70$ 到 $\sqrt{2}$ 的误差，给出不需要小数近似的表达式。

解　利用 $2 \cdot 29^2 - 41^2 = 1$ 和 $99^2 - 2 \cdot 70^2 = 1$，分别有

$$\sqrt{2} - \frac{41}{29} = \frac{2 \cdot 29^2 - 41^2}{29(29\sqrt{2} + 41)} = \frac{1}{29(29\sqrt{2} + 41)},$$

$$\frac{99}{70} - \sqrt{2} = \frac{99^2 - 2 \cdot 70^2}{70(99 + 70\sqrt{2})} = \frac{1}{70(99 + 70\sqrt{2})}.$$

二者之和必须等于 $1/(29 \cdot 70)$，这也可以直接由命题 19.1 得到。保留根式比先取小数更精确，适合比较误差界。 ♦

想一想　为什么相邻渐近分数的距离能写成 $1/(q_n q_{n-1})$，而不是只得到一个近似值？

答　因为分子 $p_{n-1}q_n - q_{n-1}p_n$ 是行列式恒等式给出的 ± 1，没有任何未控制的项。连分数的递推正是把分子压到 ± 1 的结构。

易错提醒　对任意有理数，分母较大不保证误差更小。例如 $7/5$ 到 $\sqrt{2}$ 的误差约为 0.014214，而分母更大的 $10/7$ 误差约为 0.014358，反而更大。本讲比较连续渐近分数，不能把结论套给任意选出的两个分数。

19.2　每两项中有一个好逼近

对一对相邻项 C_n, C_{n+1}，命题 19.1 只告诉我们两种误差加起来是多少。若两项都不够好，误差和就会过大。算术–几何平均不等式恰好给出所需的比较。

定理 19.2（两项逼近定理）　设 x 是无理数，$C_n = p_n/q_n$ 是它的渐近分数。对每个 $n \geq 0$，下列两项中至少有一项满足

$$|x - C_j| \leq \frac{1}{2q_j^2}, \qquad j \in \{n, n+1\}.$$

思路　反设两项都违反目标不等式。两项夹住 x，所以误差之和等于相邻项距离；再用 $uv \leq (u^2 + v^2)/2$，就会得到严格相反的不等式。

证明　记 $e_j = |x - C_j|$。由于 C_n, C_{n+1} 在 x 两侧，命题 19.1 给出

$$e_n + e_{n+1} = \frac{1}{q_n q_{n+1}}.$$

反设 $e_n > 1/(2q_n^2)$ 且 $e_{n+1} > 1/(2q_{n+1}^2)$。于是

$$e_n + e_{n+1} > \frac{1}{2}\left(\frac{1}{q_n^2} + \frac{1}{q_{n+1}^2}\right).$$

对正数 q_n^2, q_{n+1}^2 使用算术–几何平均不等式，

$$\frac{1}{2}\left(\frac{1}{q_n^2} + \frac{1}{q_{n+1}^2}\right) \geq \frac{1}{q_n q_{n+1}}.$$

这与 $e_n + e_{n+1} = 1/(q_n q_{n+1})$ 矛盾。因此至少一项满足结论。 ■

等号什么时候可能出现？算术–几何平均取等要求 $q_n = q_{n+1}$，而两个严格递增的分母通常不满足；即使分母相等，前面的「反设」仍给出严格大于。因此定理中的「至少一项」不需要指定是哪一项。

这里的「相邻」不能随意删掉。证明的误差和等于两端点间距，依赖于 x 被这两个连续渐近分数夹住。若跳过中间项，x 仍可能在端点之间，但端点距离不再是 $1/(q_n q_{n+2})$；行列式分子会包含中间部分商，不再固定为 1。因此两项界是连分数相邻结构的结论，不是对任意两个分数的结论。

■ **例 19.2　用两项界比较 π 的渐近分数**　对 $\pi = [3, 7, 15, 1, 292, \ldots]$，检查 $22/7$ 和 $333/106$ 这一对渐近分数，看看两项界选中了哪一个。

解　它们的分母分别是 7 和 106。用

$$3.14159265358979 < \pi < 3.14159265358980$$

作区间计算：

$$|\pi - 22/7| \approx 1.2645 \times 10^{-3}, \qquad \frac{1}{2 \cdot 7^2} = 1.0204 \times 10^{-2},$$

所以 $22/7$ 已满足两项界。另一项的误差约为 8.32×10^{-5}，而

$$\frac{1}{2 \cdot 106^2} \approx 4.45 \times 10^{-5},$$

它在这一次没有达到更严格的半系数界。定理保证至少有一个好项，并不保证两个都好。 ♦

■ **例 19.3**　**一个边界情形**　设两个相邻分母为 $q_n = 1, q_{n+1} = 2$，相邻项距离为 $1/2$。两项误差能否同时等于各自的 $1/(2q^2)$？

解　若同时取等，误差之和应为

$$\frac{1}{2 \cdot 1^2} + \frac{1}{2 \cdot 2^2} = \frac{5}{8},$$

但相邻项距离是 $1/(1 \cdot 2) = 1/2$，两者不相等。因此不能同时取等。这个计算也提醒我们：算术–几何平均只是证明工具，渐近分数还受到递推关系的额外限制。 ♦

把「至少一个」写成反证　遇到「两个对象中至少有一个满足界」时，先假设两个都不满足，再把两个误差相加。连分数把误差和精确变成 $1/(q_n q_{n+1})$，AM–GM 再把它和两个平方倒数联系起来。

两项定理给出的是「渐近分数之间至少有一个好项」。反过来，一个任意有理数如果好到足够严格，也不能脱离渐近分数序列。

定理 19.3（Legendre 判别法）　设 x 是无理数，$p \in \mathbb{Z}$、$q \in \mathbb{N}$ 且 $q > 0$。如果

$$|x - \frac{p}{q}| < \frac{1}{2q^2},$$

那么 p/q 必是 x 的某个渐近分数。

证明见下面的第二类最佳逼近引理及其后的证明。

引理 19.4（渐近分数的第二类最佳逼近）　设 x 为无理数，$n \geq 0$，a 为整数，b 为正整数，且 $b < q_{n+1}$、$a/b \neq p_n/q_n$。则

$$|bx - a| \geq |q_n x - p_n|.$$

证明　相邻渐近分数的行列式为 ± 1，所以存在整数 u, v 使 $(a, b) = u(p_n, q_n) + v(p_{n+1}, q_{n+1})$。若 $u = 0$，则 b 是 q_{n+1} 的正倍数，与 $b < q_{n+1}$ 矛盾；若 $v = 0$，则 $a/b = p_n/q_n$，也被排除。因此 u, v 都非零。

若 $v>0$，由 $uq_n=b-vq_{n+1}<0$ 得 $u<0$；若 $v<0$，则 $uq_n=b-vq_{n+1}>0$，得 $u>0$。故 u,v 异号。又 C_n,C_{n+1} 分居 x 两侧，所以 q_nx-p_n 与 $q_{n+1}x-p_{n+1}$ 异号。两项相乘后同号，于是

$$|bx-a|=|u|\,|q_nx-p_n|+|v|\,|q_{n+1}x-p_{n+1}|\geq|q_nx-p_n|.$$

■

思路　取最大的 n 使 $q_n\leq q$，于是 $q<q_{n+1}$。这里不把整个分母序列说成从 q_0 开始严格递增，因为 $q_0=q_1=1$ 可能发生；取最大的下标正好避开这个边界。第二类最佳逼近引理控制 C_n 的误差；若 p/q 不是 C_n，分数行列式的整数性又给出距离下界。三角不等式会推出这个距离严格小于同一个下界，形成矛盾。

证明　分母序列最终严格递增且趋于无穷，所以可以取最大的 n 使 $q_n\leq q$，从而 $q<q_{n+1}$。若 $p/q=C_n$，结论成立。否则

$$|pq_n-p_nq|\geq 1,\qquad \left|\frac{p}{q}-C_n\right|\geq\frac{1}{qq_n},$$

因为左边的分子是非零整数。由第二类最佳逼近引理，

$$|x-C_n|=\frac{|q_nx-p_n|}{q_n}\leq\frac{|qx-p|}{q_n}=\frac{q}{q_n}\left|x-\frac{p}{q}\right|.$$

因此三角不等式给出

$$\left|\frac{p}{q}-C_n\right|\leq\left(1+\frac{q}{q_n}\right)\left|x-\frac{p}{q}\right|<\frac{1+q/q_n}{2q^2}.$$

乘以 qq_n，得到

$$|pq_n-p_nq|<\frac{q_n+q}{2q}\leq 1,$$

最后一步用了 $q_n\leq q$。但 pq_n-p_nq 是非零整数，绝对值至少为 1，矛盾。因此只能有 $p/q=C_n$。 ■

■ **例 19.4　判别 π 的一个有理逼近**　说明 355/113 为什么一定是 π 的渐近分数，并解释同样的论证为什么不能选出 333/106。

解　有

$$\left|\pi-\frac{355}{113}\right|\approx 2.6676\times10^{-7}<\frac{1}{2\cdot113^2}\approx 3.9157\times10^{-5}.$$

Legendre 判别法因此保证 355/113 是渐近分数。对 333/106，误差约为 8.3220×10^{-5}，而 $1/(2\cdot106^2)\approx4.4500\times10^{-5}$，严格不等式失败；这不说明 333/106 不是渐近分数，只说明本判别法在它身上没有结论。 ♦

易错提醒　Legendre 判别法是充分条件，不是必要条件。一个渐近分数可能不满足 $1/(2q^2)$，特别是部分商较小的地方；反过来，未通过这个数值检验也不能据此断言某个分数不是渐近分数。

19.3　每三项中有一个更好的逼近

两项定理的常数是 $1/2 = 0.5$。把观察窗口扩大到三项，常数可以改成 $1/\sqrt{5} \approx 0.4472$。关键不是再做一次 AM–GM，而是把「连续两个比值都太小」与连分数分母的递推放在一起。

记

$$\varphi = \frac{1+\sqrt{5}}{2}, \qquad \frac{1}{\varphi} = \varphi - 1 = \frac{\sqrt{5}-1}{2}.$$

它们满足 $\varphi + 1/\varphi = \sqrt{5}$。这个恒等式正是黄金分割比在证明中出现的原因。

定理 19.5（三项逼近定理）　设 x 是无理数，$C_n = p_n/q_n$ 是渐近分数。对每个 $n \geq 0$，三个数

$$C_n, \quad C_{n+1}, \quad C_{n+2}$$

中至少有一个满足

$$|x - C_j| < \frac{1}{\sqrt{5}q_j^2}.$$

思路　反设三项误差都不小于目标值。相邻误差和迫使两个分母比值不超过 φ；比值是有理数，不可能恰等于无理数 φ，所以实际都小于 φ。递推式又要求第二个比值大于 $1 + 1/\varphi = \varphi$，矛盾。

证明　反设三个误差都不小于目标值，即对 $j = n, n+1, n+2$，

$$e_j := |x - C_j| \geq \frac{1}{\sqrt{5}q_j^2}.$$

对相邻的 $n, n+1$，命题 19.1 给出

$$\frac{1}{q_n q_{n+1}} = e_n + e_{n+1} \geq \frac{1}{\sqrt{5}}\left(\frac{1}{q_n^2} + \frac{1}{q_{n+1}^2}\right).$$

令 $r = q_{n+1}/q_n > 0$，乘以 $q_n q_{n+1}$ 得

$$r + \frac{1}{r} \leq \sqrt{5}.$$

函数 $r + 1/r$ 在 $[1, \infty)$ 上递增，而 $r \geq 1$。由 $r + 1/r \leq \sqrt{5}$ 得 $r \leq \varphi$；但 r 是有理数，φ 是无理数，所以 $r < \varphi$。于是

$$\frac{q_n}{q_{n+1}} > \frac{1}{\varphi}.$$

同样对 $n+1, n+2$ 应用相同论证，得到 $q_{n+2}/q_{n+1} < \varphi$。

另一方面，分母递推式给出

$$\frac{q_{n+2}}{q_{n+1}} = a_{n+2} + \frac{q_n}{q_{n+1}} \geq 1 + \frac{q_n}{q_{n+1}} > 1 + \frac{1}{\varphi} = \varphi,$$

这与 $q_{n+2}/q_{n+1} < \varphi$ 矛盾。因此三个误差中至少有一个满足目标界。 ■

严格结论需要留意边界：如果三项误差都等于或大于目标值，分母比值最多等于 φ；整数分母的比值是有理数，不能取到这个等号。因此至少一项严格小于目标值。

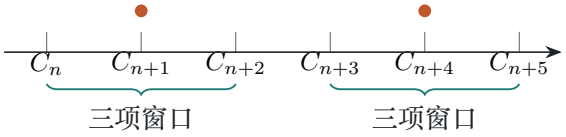


图 19.2　把渐近分数按三个一组分窗：每一组至少有一个分数达到 $1/(\sqrt{5}q^2)$ 的误差界。

图 19.2 中的橙点不表示固定位置；它可能落在窗口的任意一项。定理的力量在于无论部分商怎样变化，窗口都不能连续出现三个「坏项」。

■ **例 19.5**　$\sqrt{2}$ **的三项窗口**　检查 $\sqrt{2}$ 的前三个窗口，找出达到 $1/(\sqrt{5}q^2)$ 界的渐近分数。

解　取

$$C_0 = 1, \quad C_1 = \frac{3}{2}, \quad C_2 = \frac{7}{5}, \quad C_3 = \frac{17}{12}, \quad C_4 = \frac{41}{29}.$$

逐项比较误差和目标上界：

j	C_j	$\lvert\sqrt{2} - C_j\rvert$	$1/(\sqrt{5}q_j^2)$
0	1	0.414214	0.447214
1	3/2	0.085786	0.111803
2	7/5	0.014214	0.017889
3	17/12	0.002453	0.003106
4	41/29	0.000420	0.000532

表中五项都达到界，因此窗口 (C_0, C_1, C_2)、(C_1, C_2, C_3) 和 (C_2, C_3, C_4) 均有好项（本例中三项全是好项）。这些小数用于核对具体例子；对任意无理数，普遍保证仍由三项逼近定理给出。　♦

■ **例 19.6**　**用分母比值复核证明中的矛盾**　若某一段连分数的分母满足 $q_n = 13, q_{n+1} = 21$，证明下面三个误差不可能同时不小于各自的目标界：

$$|x - C_j| \geq \frac{1}{\sqrt{5}q_j^2}, \qquad j = n, n+1, n+2.$$

要求用分母递推式说明矛盾发生在哪里。

解　这里 $q_{n+1}/q_n = 21/13 \approx 1.615 < \varphi$，所以前两个分母本身还没有产生矛盾。若前两个误差都不小于目标界，证明会进一步给出

$$\frac{q_n}{q_{n+1}} > \frac{1}{\varphi}.$$

另一方面，$q_{n+2} = a_{n+2}q_{n+1} + q_n$ 且 $a_{n+2} \geq 1$，故

$$\frac{q_{n+2}}{q_{n+1}} \geq 1 + \frac{13}{21} = \frac{34}{21} \approx 1.619 > \varphi.$$

因此第二个相邻比值必然大于 φ；但反设中后两个误差（$j = n+1, n+2$）都不小于目标界，又要求这个比值小于 φ，矛盾。第三项误差的假设在这一步不可缺少：只知道前两项的误差大，并不能约束后一对的误差和。　♦

想一想　证明三项定理时，为什么不能只把两项定理连续用两次？

答　两项定理只给出 $1/(2q^2)$ 界，再用一次也不会改变常数。它还可能在窗口 (C_n, C_{n+1}) 中选中 C_{n+1}，在窗口 (C_{n+1}, C_{n+2}) 中又选中同一个 C_{n+1}；它没有保证三个不同项里出现两个好项。三项定理需要利用分母递推，正是这一步带来了更小的常数 $1/\sqrt{5}$。

19.4　无限多个好有理数与 Hurwitz 常数

把三项定理应用到无穷长的渐近分数序列，就得到一个对所有无理数都成立的逼近结论。

推论 19.6（Hurwitz 逼近）　对任意无理数 x，存在无穷多个互异的既约分数 p/q（$q>0$），使

$$|x-\frac{p}{q}|<\frac{1}{\sqrt{5}q^2}.$$

证明　把渐近分数分成不重叠的三元组

$$(C_0,C_1,C_2),\quad (C_3,C_4,C_5),\quad (C_6,C_7,C_8),\dots.$$

每个三元组由定理 19.5 选出一个好项。渐近分数的分母最终严格递增并趋于无穷，所以去掉有限个最前面的三元组后，各组所选好项互异；第 18 讲的行列式恒等式又说明每个 p_n/q_n 已经既约。于是得到无穷多个满足不等式的既约分数。 ■

注 19.1　「无穷多个」不能弱化成「至少一个」。任给一个有理数 p/q，总能取更高的渐近分数；它们的分母趋于无穷，构成的是一列越来越细的逼近。另一方面，结论只对无理数成立。若 $x=a/b$ 为既约有理数，而 $p/q\neq a/b$ 也是既约分数，则

$$|x-\frac{p}{q}|=\frac{|aq-bp|}{bq}\geq\frac{1}{bq}.$$

因此对固定的 $K>0$，当 $q\geq Kb$ 时不可能有 $|x-p/q|<K/q^2$；剩下的有限多个分母各自也只允许有限多个分子。故有理数只能有有限多个不同的既约分数满足固定的 K/q^2 误差界。

■ **例 19.7**　**三项分组的实际使用**　用 π 的部分商 $[3,7,15,1,292,1,1,1,2,\dots]$ 把前九个渐近分数分成三组，并指出每组中的一个好项。

解　三行递推表给出

n	0	1	2	3	4	5	6	7	8
p_n	3	22	333	355	103993	104348	208341	312689	833719
q_n	1	7	106	113	33102	33215	66317	99532	265381

前一组 $(n=0,1,2)$ 中 $22/7$ 满足界；第二组中 $355/113$ 的误差约 2.67×10^{-7}，远小于 $1/(\sqrt{5}\cdot113^2)$；第三组 $(n=6,7,8)$ 中，$C_7=312689/99532$ 满足界。实际计算只是在找出具体成员，存在性来自三项逼近定理。 ♦

■ **例 19.8　与 Dirichlet 型结论的比较**　比较普通的 $1/q^2$ 误差界与 Hurwitz 界。对 $q=113$，两个上界分别是多少？

解　普通界为 $1/q^2 = 1/12769 \approx 7.83\times10^{-5}$，Hurwitz 界为

$$\frac{1}{\sqrt{5}q^2} = \frac{1}{\sqrt{5}\cdot 12769} \approx 3.50\times10^{-5}.$$

后者更严格，代价是它只保证在无穷多个渐近分数中反复出现，而不是每一个渐近分数都满足。对 $355/113$，实际误差约 2.67×10^{-7}，两个界都远大于实际误差。 ♦

易错提醒　Hurwitz 结论说的是「存在无穷多个 p/q」，不是「所有足够大的 q 都有一个分数满足界」。渐近分数的分母会跳过许多整数；三项窗口的结论只排除了连续三个误差未达界的渐近分数。

19.5　黄金分割比与锐利常数

前一节给出的常数 $1/\sqrt{5}$ 看起来来自一个证明中的代数恒等式。要判断它是不是偶然的，最自然的测试对象是所有部分商都等于 1 的连分数：

$$\varphi = [1,1,1,\ldots].$$

截断后得到

$$1,\quad 2,\quad \frac{3}{2},\quad \frac{5}{3},\quad \frac{8}{5},\quad \frac{13}{8},\quad \frac{21}{13},\ldots.$$

分子、分母正好是相邻的 Fibonacci 数。记 $F_0=0, F_1=1$，并令 $F_{m+1}=F_m+F_{m-1}$（$m\ge 1$）。

命题 19.7（黄金分割比的渐近分数）　对 $m\ge 1$，φ 的第 m 个非平凡渐近分数为

$$C_m = \frac{F_{m+2}}{F_{m+1}}.$$

并且

$$F_{m+2}^2 - F_{m+2}F_{m+1} - F_{m+1}^2 = (-1)^{m+1}.$$

证明　部分商全为 1 时，渐近分数递推式变为

$$p_m = p_{m-1}+p_{m-2},\qquad q_m = q_{m-1}+q_{m-2},$$

初值为 $(p_{-2},p_{-1})=(0,1)$、$(q_{-2},q_{-1})=(1,0)$。因此 $(p_m,q_m)=(F_{m+2},F_{m+1})$，可用归纳法逐项核对。

令 $P=F_{m+2}$、$Q=F_{m+1}$。因为 $F_{m+2}=F_{m+1}+F_m$，有

$$P^2 - PQ - Q^2 = F_{m+2}F_m - F_{m+1}^2.$$

右边是 Cassini 恒等式的一个移位形式，等于 $(-1)^{m+1}$。这也说明每个分数都已既约。 ■

■ **例 19.9　计算 Fibonacci 渐近分数的误差**　求 $\frac{F_{12}}{F_{11}}$，并比较它到 φ 的误差与 $1/(\sqrt{5}F_{11}^2)$。

解　递推给出 $F_{11}=89, F_{12}=144$，所以渐近分数是 $144/89$。由 $\varphi^2-\varphi-1=0$，

$$\varphi-\frac{144}{89}=\frac{89\varphi-144}{89}\approx 5.6461\times 10^{-5}.$$

目标界为

$$\frac{1}{\sqrt{5}\cdot 89^2}\approx 5.6459\times 10^{-5},$$

所以 $144/89$ 略微不满足严格的 Hurwitz 界；下一项 $233/144$ 则落在界内。这个交替现象正是极限计算中误差项的符号造成的。 ♦

把几项数值列在一起，可以看清「从两侧逼近」的细节：

m	F_{m+1}/F_m	$F_m^2\lvert\varphi-F_{m+1}/F_m\rvert$	$1/\sqrt{5}$
8	$34/21$	0.447011	0.447214
9	$55/34$	0.447291	0.447214
10	$89/55$	0.447184	0.447214
11	$144/89$	0.447225	0.447214
12	$233/144$	0.447209	0.447214

奇偶项从两侧交错靠近 $1/\sqrt{5}$，而不是单调靠近。表中每个分母都来自同一递推，因而可以同时检验渐近分数、误差和极限三件事。

图 19.3 将表中的缩放误差画成折线。纵轴只显示 0.4470 附近的一小段，以便看清各项与 $1/\sqrt{5}$ 的大小关系。

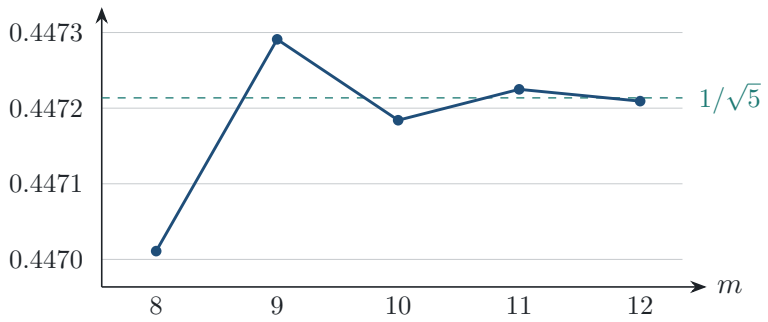


图 19.3　缩放误差 $F_m^2\lvert\varphi-F_{m+1}/F_m\rvert$ 在 $1/\sqrt{5}$ 两侧交替。纵轴局部放大；小于水平虚线的项满足严格 Hurwitz 界。

折线的振幅逐渐减小；表中偶数 m 的点在线下，奇数 m 的点在线上。图只展示这五项，后面的恒等式会证明极限确实是 $1/\sqrt{5}$。

这里还可以看出一个容易忽略的区别。定理保证「每个三项窗口至少有一个好项」，但好项的位置可以变化；在黄金分割比的表中，好项恰好落在偶数 m 上，这是本例的特殊性质。不能把定理改写成「所有偶数项都好」或「所有奇数项都好」。真正稳定的量是窗口长度和误差尺度，而不是某一个固定的下标。

从相邻距离还可以得到一个不带黄金分割比的粗界。因为 $q_{n+1}\geq q_n$，命题 19.1 给出

$$|x-C_n|<|C_{n+1}-C_n|=\frac{1}{q_nq_{n+1}}\leq\frac{1}{q_n^2}.$$

这解释了为什么连分数渐近分数至少总有 $1/q^2$ 量级的逼近；第 19 讲的两个定理进一步把「至少」提升为固定窗口内的 $1/(2q^2)$ 和 $1/(\sqrt{5}q^2)$。常数的改进来自相邻误差不能同时偏大的额外结构，而不是来自把分母简单地换成更大的数。

Hurwitz 推论中的「无穷多个」也有一个具体的分组含义。把渐近分数分成三项一组：

$$(C_0, C_1, C_2), \quad (C_3, C_4, C_5), \quad (C_6, C_7, C_8), \quad \dots.$$

每组至少选出一个好项。

因为分母从某一项起严格递增，来自充分靠后各组的好项不可能是同一个分数（允许最前面出现 $q_0 = q_1 = 1$）。即使相邻两组的好项位置相同，它们的下标也相差至少 3。这一步把局部的三项结论转成全局的无限性，是存在性证明中不能省略的部分。

另一种容易混淆的量词是「任意大的分母」。本推论确实给出任意大的 q，因为好渐近分数有无穷多个且分母趋于无穷；但它没有说每个足够大的整数 q 都能找到这样的 p。例如分母 q 可能跳过很多整数，定理只在渐近分数的分母序列上作保证。把存在的分数限制在渐近分数中，是结论成立的重要信息。

设 $\psi = (1-\sqrt{5})/2 = -1/\varphi$ 是 φ 的共轭。由 Binet 公式

$$F_m = \frac{\varphi^m - \psi^m}{\sqrt{5}}$$

而 $|\psi/\varphi| = 1/\varphi^2 < 1$，所以

$$\frac{F_{m+1}}{F_m} = \varphi \frac{1 - (\psi/\varphi)^{m+1}}{1 - (\psi/\varphi)^m} \longrightarrow \varphi.$$

这一步说明 Fibonacci 比值确实趋于黄金分割比。对 $P = F_{m+1}$、$Q = F_m$，Cassini 恒等式给出 $P^2 - PQ - Q^2 = (-1)^m$，另一方面

$$\left(\varphi - \frac{P}{Q}\right)\left(\psi - \frac{P}{Q}\right) = -1 - \frac{P}{Q} + \frac{P^2}{Q^2} = \frac{P^2 - PQ - Q^2}{Q^2} = \frac{(-1)^m}{Q^2}.$$

因此

$$Q^2 \left|\varphi - \frac{P}{Q}\right| = \frac{1}{|\psi - P/Q|}. \tag{19.2}$$

当 $m \to \infty$ 时，$P/Q \to \varphi$，所以式 (19.2) 的右端趋于

$$\frac{1}{|\psi - \varphi|} = \frac{1}{\sqrt{5}}.$$

定理 19.8（Hurwitz 常数的最优性）　对任意常数 $0 < K < 1/\sqrt{5}$，存在无理数 x（可以取 $x = \varphi$），使得只有有限多个既约分数 p/q（$p \in \mathbb{Z}$、$q > 0$）满足

$$\left|x - \frac{p}{q}\right| < \frac{K}{q^2}.$$

所以 Hurwitz 结论中的系数 $1/\sqrt{5}$ 不能再减小；等价地，分母中的 $\sqrt{5}$ 不能换成更大的常数。

思路　取 $x = \varphi$，利用它的整系数二次方程。任何有理数 p/q 代入多项式 $T^2 - T - 1$ 后，所得分子是非零整数，绝对值至少为 1。若一列分数越来越靠近 φ，把这个整数恒等式除以线性因子，就得到误差的下极限至少是 $1/\sqrt{5}$。

证明　令 $p/q \neq \varphi$ 为既约分数，并设 $N = p^2 - pq - q^2$。N 是整数；若 $N = 0$，则 p/q 是 $T^2 - T - 1$ 的根 φ 或 ψ，二者都无理，所以 $N \neq 0$ 且 $|N| \geq 1$。因式分解

$$\left(\frac{p}{q} - \varphi\right)\left(\frac{p}{q} - \psi\right) = \frac{N}{q^2}$$

给出

$$q^2\left|\varphi - \frac{p}{q}\right| = \frac{|N|}{|p/q - \psi|} \geq \frac{1}{|p/q - \psi|}.$$

若有无穷多个分数满足 $|\varphi - p/q| < K/q^2$，因有界分母只对应有限多个满足该界的分数，可以选出分母趋于无穷的子列；沿该子列有 $p/q \to \varphi$。沿这列分数取下极限，得到

$$\liminf q^2\left|\varphi - \frac{p}{q}\right| \geq \frac{1}{|\varphi - \psi|} = \frac{1}{\sqrt{5}},$$

这与 $K < 1/\sqrt{5}$ 矛盾。因此这样的分数至多有限多个。 ■

定理 19.8 与 Fibonacci 计算合在一起，说明常数的含义：对一般无理数，三项定理保证无穷多次达到 $1/\sqrt{5}$ 的尺度；对黄金分割比，这个尺度沿着一列自然出现的渐近分数逼近，且任何更小的系数最终都会失败。

易错提醒　「常数最优」容易被说反。写成 K/q^2 时，K 越小，要求越严格；因此不能把 $1/\sqrt{5}$ 换成更小的 K。若把界写成 $1/(cq^2)$，则不能把 $c = \sqrt{5}$ 换成更大的 c。

19.6　周期性与二次方程

黄金分割比的周期是 $[\overline{1}]$。周期连分数并不只是一个记号游戏：重复的尾巴可以被当成一个未知数，从而直接得到二次方程。这是下一讲 Lagrange 定理的一个方向。

设

$$\theta = [\overline{a_0, a_1, \ldots, a_{m-1}}]$$

是一个纯周期简单连分数，其中 $m \geq 1$，所有 a_i 均为正整数。令 $p_j/q_j = [a_0, \ldots, a_j]$，并沿用第 18 讲的初值 $p_{-2} = 0, p_{-1} = 1, q_{-2} = 1, q_{-1} = 0$。

命题 19.9（周期连分数满足二次方程）　上述 θ 满足

$$q_{m-1}\theta^2 + (q_{m-2} - p_{m-1})\theta - p_{m-2} = 0.$$

因此每一个周期连分数都是二次无理数。

证明　把一个完整周期后面再次出现的尾巴仍记作 θ。有限连分数的矩阵递推，或直接从后往前合并，给出

$$[a_0, \ldots, a_{m-1}, \theta] = \frac{p_{m-1}\theta + p_{m-2}}{q_{m-1}\theta + q_{m-2}}.$$

左边就是 θ。交叉相乘并移项，得到所列二次方程。系数是整数，且无限简单连分数不是有理数，所以它确实是二次无理数。 ■

■ **例 19.10　两个周期连分数的方程**　分别求 $[\overline{1}]$ 与 $[\overline{1,2,3}]$ 满足的二次方程，并取出正根。

解　对 $[\overline{1}]$，$m=1$，有 $p_0=q_0=1, p_{-1}=1, q_{-1}=0$，代入命题得到

$$\theta^2-\theta-1=0.$$

正根为 $\theta=(1+\sqrt{5})/2=\varphi$。

对周期 $1,2,3$，渐近分数表给出

$$(p_0,p_1,p_2)=(1,3,10),\qquad (q_0,q_1,q_2)=(1,2,7).$$

于是

$$7\theta^2-8\theta-3=0,\qquad \theta=\frac{4+\sqrt{37}}{7},$$

其中另一个根为负数，不可能是正的简单连分数。这里必须把周期重复后的尾巴保留为未知数；只计算有限值 $[1,2,3]=10/7$ 会得到另一个问题。 ♦

想一想　为什么有限连分数 $[1,2,3]$ 与周期连分数 $[\overline{1,2,3}]$ 不能混用？

答　前者在最后一个部分商 3 处停止，值是 $10/7$；后者在 3 后面又接回 $1,2,3,\ldots$，所以尾巴本身是未知数，值为 $(4+\sqrt{37})/7$。周期性改变了方程的边界条件，不能用有限值代替无限尾巴。

前后联系　Lagrange 定理断言：无理数是二次无理数，当且仅当其简单连分数最终周期。本讲证明周期尾巴满足二次方程，有限前缀不改变所属二次域。第 20 讲证明反向：由完全商对应的有界整数状态必重复，得到最终周期。

本讲小结

- 相邻渐近分数满足 $|C_n-C_{n-1}|=1/(q_nq_{n-1})$，且两项到 x 的误差之和等于这个距离。
- 每两个相邻渐近分数中至少有一个满足 $|x-p/q|\le 1/(2q^2)$；证明把两个误差相加，再用 AM–GM。
- 每三个相邻渐近分数中至少有一个满足 $|x-p/q|<1/(\sqrt{5}q^2)$；分母递推与 $\varphi+1/\varphi=\sqrt{5}$ 共同产生矛盾。
- 对任意无理数 x，存在无穷多个既约 p/q 满足 $|x-p/q|<1/(\sqrt{5}q^2)$，这就是 Hurwitz 逼近结论。
- $C_m=F_{m+2}/F_{m+1}$ 且 $F_{m+1}^2|\varphi-C_m|\to 1/\sqrt{5}$，故此常数最优。
- 若 $\theta=[\overline{a_0,\ldots,a_{m-1}}]$，则 θ 满足 $q_{m-1}\theta^2+(q_{m-2}-p_{m-1})\theta-p_{m-2}=0$；周期性必然产生二次无理数。

练习

19.1 ★　对 $\sqrt{3}$ 的相邻渐近分数 $7/4$ 与 $19/11$，验证它们的距离为 $1/(4\cdot 11)$，并写出两项误差之和的精确表达式。

19.2 ★ 对 π 的渐近分数 $22/7$ 与 $333/106$，分别检查 $1/(2q^2)$ 界，并指出定理 19.2 在这一对中至少保证了什么。

19.3 ★★ 设三个连续分母为 $q_n = 13, q_{n+1} = 21, q_{n+2} = 34$。验证它们满足递推，并说明为什么三项反设不能在这三个分母上同时成立。

19.4 ★★ 证明：若 x 的渐近分数中有三个连续项都不满足 $1/(\sqrt{5}q^2)$ 界，那么其中两个相邻分母的比值必小于 φ。

19.5 ★★ 计算 $\sqrt{2}$ 的渐近分数 $41/29$ 与 $99/70$ 的两个误差，并验证至少一个满足 $1/(2q^2)$ 界。

19.6 ★★ 令 $F_0 = 0, F_1 = 1$。证明 $F_{m+1}F_{m-1} - F_m^2 = (-1)^m$，并由此证明 F_{m+1}/F_m 已既约。

19.7 ★★ 对 φ 的渐近分数 $144/89$，计算 $89^2|\varphi - 144/89|$ 的数值，并与 $1/\sqrt{5}$ 比较。

19.8 ★★ 证明：若 p/q 是任意有理数且 $p/q \neq \varphi$，则

$$q^2\left|\varphi - \frac{p}{q}\right| = \frac{|p^2 - pq - q^2|}{|p/q - \psi|}, \qquad \psi = \frac{1-\sqrt{5}}{2}.$$

19.9 ★★★ 求 $[\overline{2}]$ 满足的二次方程和正根，并解释它与 $\sqrt{2}$ 的关系。

19.10 ★★★ 设 $\theta = [\overline{1,2,3}]$。用命题 19.9 求出 θ，再验证正根确实大于 1。

练习答案与提示

19.1 $19/11 - 7/4 = (76-77)/44 = -1/44$，所以距离为 $1/44$。由于两项夹住 $\sqrt{3}$，误差之和为 $1/44$。

19.2 $|\pi - 22/7| \approx 1.2645\times10^{-3} < 1/98$；而 $|\pi - 333/106| \approx 8.3220\times10^{-5} > 1/(2\cdot106^2) \approx 4.4500\times10^{-5}$。因此这对中只有 $22/7$ 达到界，符合定理保证的「至少一项」。

19.3 $34 = 21 + 13$，正是 $a_{n+2} = 1$ 时的递推。第一比值 $21/13 \approx 1.615 < \varphi$，但第二比值 $34/21 \approx 1.619 > \varphi$。若三项都不满足目标界，前一对误差要求第一比值小于 φ，后一对又要求第二比值小于 φ，与实际第二比值矛盾。

19.4 设 $r = q_{n+1}/q_n$。相邻误差和给出 $r + 1/r \le \sqrt{5}$。因 $r \ge 1$ 且该函数在 $[1,\infty)$ 上递增，得 $r \le \varphi$；又 r 有理而 φ 无理，所以 $r < \varphi$。

19.5 $2\cdot29^2 - 41^2 = 1$、$99^2 - 2\cdot70^2 = 1$，所以误差分别为 $1/[29(29\sqrt{2}+41)]$ 与 $1/[70(99+70\sqrt{2})]$。两者之和为 $1/(29\cdot70)$；而第一项约为 0.000420，小于 $1/(2\cdot29^2) \approx 0.000594$。

19.6 记 $D_m = F_{m+1}F_{m-1} - F_m^2$。基例 $D_1 = -1$。利用 $F_{m+2} = F_{m+1} + F_m$ 与 $F_{m+1} - F_m = F_{m-1}$，有 $D_{m+1} = F_{m+2}F_m - F_{m+1}^2 = -D_m$，故 $D_m = (-1)^m$。当 $m \ge 1$ 时，$\gcd(F_{m+1}, F_m)$ 整除 $D_m = \pm1$，所以 F_{m+1}/F_m 已既约。

19.7 $89^2|\varphi - 144/89| \approx 0.447225$，略大于 $1/\sqrt{5} \approx 0.447214$。随着 m 增大，这个数沿 Fibonacci 渐近分数从两侧趋近 0.447214。

19.8 因为 $\varphi^2 - \varphi - 1 = 0$，有 $(p/q - \varphi)(p/q - \psi) = p^2/q^2 - p/q - 1 = (p^2 - pq - q^2)/q^2$。取绝对值并移项即得。

19.9 令 $\theta = [\overline{2}]$。去掉第一个部分商后尾巴仍为 θ，所以 $\theta = 2 + 1/\theta$。乘以正数 θ 得 $\theta^2 - 2\theta - 1 = 0$。求根得 $1 \pm \sqrt{2}$；其中 $1 - \sqrt{2} < 0$，不可能是正的连分数值，故取 $1 + \sqrt{2}$。代回原式也成立：$2 + 1/(1+\sqrt{2}) = 1 + \sqrt{2}$。因此 $[\overline{2}] = 1 + \sqrt{2}$。再用定义，

$$[1,\overline{2}] = 1 + \frac{1}{1+\sqrt{2}} = \sqrt{2},$$

所以 $\sqrt{2}=[1,\overline{2}]$。

19.10 直接把周期尾巴保留为未知数，也有

$$\theta = 1 + \frac{1}{2 + \frac{1}{3+1/\theta}}.$$

令 $t=\theta-1$。由上式，

$$\frac{1}{t} - 2 = \frac{1}{3+1/\theta} \quad\Longrightarrow\quad \frac{3-2\theta}{\theta-1} = \frac{\theta}{3\theta+1}.$$

交叉相乘得到 $7\theta^2-8\theta-3=0$。这与渐近分数递推相符：前三项的分子分母为 $(1,1),(3,2),(10,7)$，代入命题 19.9：

$$7\theta^2 + (2-10)\theta - 3 = 0,$$

即 $7\theta^2-8\theta-3=0$。判别式为 $148=4\cdot 37$，两根为 $(4\pm\sqrt{37})/7$。因为 $\sqrt{37}>4$，负号对应负根；周期简单连分数为正，所以只能取

$$\theta = \frac{4+\sqrt{37}}{7} \approx 1.4404 > 1.$$

代回原方程：

$$7\left(\frac{4+\sqrt{37}}{7}\right)^2 - 8\left(\frac{4+\sqrt{37}}{7}\right) - 3 = 0,$$

展开后根式项抵消，常数项也为 0。这个检验能发现把有限值 $10/7$ 误当周期值的错误。一般周期为 $a_0,\dots,a_{m-1}$ 时，同样的尾巴代换会产生

$$q_{m-1}\theta^2 + (q_{m-2}-p_{m-1})\theta - p_{m-2} = 0,$$

这正是命题 19.9 的统一公式。第 20 讲的任务是说明二次无理数为什么一定能反过来产生这样的周期。

第 20 讲

循环连分数与二次无理数

Periodic Continued Fractions, Quadratic Irrationals

对应材料：MIT 18.781 Lecture 20 讲义（及 Lecture 19 末尾定理 70 的陈述与证明开头）；习题集 8 第 7 题；习题集 9 第 4、7、9 题。

本讲目标

(1) 会把二次无理数写成标准形 $\dfrac{B+\sqrt{d}}{C}$（$C \mid d-B^2$），会求共轭，知道共轭保持四则运算。

(2) 会由循环连分数解二次方程求出它的值（Euler 方向）。

(3) 理解 Lagrange 定理的证明：用一对整数 (B_i, C_i) 记录完全商，证明它们是整数、有界，再用抽屉原理得到循环。会用这套整数算法手算任何二次无理数的连分数。

(4) 掌握 Galois 判别法：连分数纯循环当且仅当 $x>1$ 且共轭落在 $(-1,0)$ 内。

(5) 看到 $\sqrt{d}$ 的连分数与 Pell 方程 $x^2-dy^2=1$ 之间的联系，为下一讲做准备。

20.1 二次无理数与共轭

第 18 讲算过几个连分数：黄金分割比 $\varphi=\frac{1+\sqrt{5}}{2}=[1,1,1,\dots]$，$\sqrt{2}=[1,2,2,2,\dots]$。用同样的方法再多算几个，结果列在表 20.1 中。

数	连分数的前若干项	规律
$\sqrt{2}$	$[1,2,2,2,2,2,\dots]$	从第 2 项起重复 2
$\sqrt{3}$	$[1,1,2,1,2,1,2,\dots]$	从第 2 项起重复 1,2
$\sqrt{7}$	$[2,1,1,1,4,1,1,1,4,\dots]$	从第 2 项起重复 1,1,1,4
$\frac{1+\sqrt{5}}{2}$	$[1,1,1,1,1,1,\dots]$	从头重复 1
$\sqrt[3]{2}$	$[1,3,1,5,1,1,4,1,1,8,1,14,1,10,2,\dots]$	看不出规律
π	$[3,7,15,1,292,1,1,1,2,1,3,\dots]$	看不出规律

表 20.1　几个实数的连分数。前四个是二次无理数，连分数最终循环；后两行只列出有限项，不能据此判断是否循环。

表中的现象不是巧合。回忆小数的情形：一个实数的十进制小数最终循环，当且仅当它是有理数。连分数有一个平行的结论，只是「有理数」换成了「二次无理数」：

展开方式	有限	最终循环
十进制小数	分母只含因子 $2,5$ 的有理数	有理数
连分数	有理数	二次无理数

本讲的任务就是证明这一行的最后一格，并刻画「从头开始循环」的那些数。

定义 20.1（二次无理数） 如果实数 x 是无理数，并且满足某个整系数二次方程 $ux^2+vx+w=0$（$u,v,w\in\mathbb{Z}$，$u\neq 0$），就称 x 为**二次无理数**（quadratic irrational）。

由求根公式，$x=\frac{-v\pm\sqrt{D}}{2u}$，$D=v^2-4uw$。$x$ 是实数且是无理数，等价于 $D>0$ 且 D 不是完全平方数。把 D 写成 f^2t（$t>1$ 无平方因子），就得到 $x=r+s\sqrt{t}$，$r,s\in\mathbb{Q}$，$s\neq 0$。反过来，这样的数满足 $(x-r)^2=s^2t$，去分母就是一个整系数二次方程。所以**二次无理数就是形如 $r+s\sqrt{t}$（$r,s\in\mathbb{Q}$，$s\neq 0$）的数**，这正是第 19 讲定理 70 中的说法。

做连分数时，另一种写法更方便。

命题 20.2（标准形） 每个二次无理数 x 都能写成

$$x=\frac{B+\sqrt{d}}{C},\qquad B,C,d\in\mathbb{Z},\ \ d>0\text{ 不是完全平方数},\ \ C\neq 0,\ \ C\mid d-B^2.$$

证明 由求根公式，$x=\frac{-v+\sqrt{D}}{2u}$ 或 $x=\frac{-v-\sqrt{D}}{2u}=\frac{v+\sqrt{D}}{-2u}$。无论哪种，$x=\frac{P+\sqrt{D}}{Q}$，$P,Q,D$ 为整数，$D>0$ 不是平方数，$Q\neq 0$。分子分母同乘 $|Q|$：

$$x=\frac{P|Q|+\sqrt{DQ^2}}{Q|Q|}.$$

取 $B=P|Q|$，$C=Q|Q|=\pm Q^2$，$d=DQ^2$，则 $d-B^2=Q^2(D-P^2)$ 被 $C=\pm Q^2$ 整除。 ■

例如 $\frac{1+\sqrt{5}}{3}$ 不满足整除条件（$3\nmid 5-1$），乘 3 后变成 $\frac{3+\sqrt{45}}{9}$，此时 $45-9=36$ 被 9 整除。标准形里 d 不必无平方因子，关键在于整除条件，它在 §20.4 的证明中起决定作用。

定义 20.3（共轭） 设 $x=r+s\sqrt{t}$（$r,s\in\mathbb{Q}$，$t>1$ 无平方因子）。称 $\bar{x}=r-s\sqrt{t}$ 为 x 的**共轭**（conjugate）。用标准形写，$x=\frac{B+\sqrt{d}}{C}$ 的共轭是 $\bar{x}=\frac{B-\sqrt{d}}{C}$。

r,s 由 x 唯一确定：若 $r+s\sqrt{t}=r'+s'\sqrt{t}$ 且 $s\neq s'$，则 $\sqrt{t}=\frac{r'-r}{s-s'}$ 是有理数，矛盾。所以共轭是良定义的。另一个看法是：x 与 $\bar{x}$ 恰好是方程 $(X-r)^2=s^2t$ 的两个根，二者之和 $2r$、之积 r^2-s^2t 都是有理数。

把形如 $r+s\sqrt{t}$（$r,s\in\mathbb{Q}$，允许 $s=0$）的数的全体记作 $\mathbb{Q}(\sqrt{t})$。共轭在 $\mathbb{Q}(\sqrt{t})$ 上有定义，在有理数上就是恒等映射。

引理 20.4（共轭保持四则运算）　对 $\alpha,\beta\in\mathbb{Q}(\sqrt{t})$，有 $\overline{\alpha\pm\beta}=\bar\alpha\pm\bar\beta$，$\overline{\alpha\beta}=\bar\alpha\,\bar\beta$；当 $\beta\neq0$ 时 $\overline{\alpha/\beta}=\bar\alpha/\bar\beta$。

证明　设 $\alpha=r+s\sqrt{t}$，$\beta=u+v\sqrt{t}$。加减法直接按有理系数分别相加减。乘法：$\alpha\beta=(ru+svt)+(rv+su)\sqrt{t}$，它的共轭 $(ru+svt)-(rv+su)\sqrt{t}$ 恰好等于 $(r-s\sqrt{t})(u-v\sqrt{t})$。除法：$\beta\neq0$ 时 $\bar\beta\neq0$（$\bar\beta=0$ 推出 $u=v=0$）。令 $\gamma=\alpha/\beta$，由乘法情形 $\bar\alpha=\bar\gamma\bar\beta$，所以 $\bar\gamma=\bar\alpha/\bar\beta$。∎

引理 20.5（分式线性变换）　设 x 是二次无理数，$a,b,c,e\in\mathbb{Z}$，$ae-bc\neq0$。则 $cx+e\neq0$，$y=\dfrac{ax+b}{cx+e}$ 也是二次无理数，并且 $\bar y=\dfrac{a\bar x+b}{c\bar x+e}$。

证明　若 $cx+e=0$，则 $c\neq0$（否则 $e=0$，与 $ae-bc\neq0$ 矛盾），于是 $x=-e/c$ 是有理数，矛盾。所以 y 有定义，且 $y\in\mathbb{Q}(\sqrt{t})$。若 y 是有理数，由 $y(cx+e)=ax+b$ 得 $(cy-a)x=b-ey$。$cy-a=0$ 时 $b-ey=0$，两式消去 y 得 $ae-bc=0$，矛盾；$cy-a\neq0$ 时 $x=\frac{b-ey}{cy-a}$ 是有理数，矛盾。所以 y 是二次无理数。共轭的公式由引理 20.4 得到。∎

连分数的一步 $x\mapsto\frac{1}{x-a}=\frac{0\cdot x+1}{1\cdot x-a}$ 就是这样的变换（$ae-bc=-1$）。所以**二次无理数的每个完全商仍是二次无理数**，而且完全商的共轭可以由共轭一步步算出来。

■ **例 20.1**　把 $x=\dfrac{2+\sqrt{3}}{1+\sqrt{3}}$ 化为标准形，求它的共轭和它满足的整系数二次方程。

解　分子分母同乘 $1-\sqrt{3}$：

$$x=\frac{(2+\sqrt{3})(1-\sqrt{3})}{(1+\sqrt{3})(1-\sqrt{3})}=\frac{2-2\sqrt{3}+\sqrt{3}-3}{1-3}=\frac{-1-\sqrt{3}}{-2}=\frac{1+\sqrt{3}}{2}.$$

这里 $B=1,C=2,d=3,d-B^2=2$ 被 $C=2$ 整除，已经是标准形。共轭是 $\bar x=\frac{1-\sqrt{3}}{2}\approx-0.3660$。

二次方程可以由根与系数的关系写出：$x+\bar x=1$，$x\bar x=\frac{1-3}{4}=-\frac12$，所以 $x,\bar x$ 是 $X^2-X-\frac12=0$ 的两根，即

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

检验：$x\approx1.3660$，$2x^2-2x-1\approx2\times1.8660-2.7321-1\approx0$。♦

> **想一想**　小王求上例的共轭时没有先化简，而是直接把原式中的 $\sqrt{3}$ 都换成 $-\sqrt{3}$，得到 $\frac{2-\sqrt{3}}{1-\sqrt{3}}$。这样做对吗？如果原式里同时出现 $\sqrt{3}$ 和 $\sqrt{12}$ 呢？
>
> 答　对。由引理 20.4，共轭可以逐项、逐个运算地做：$\frac{2-\sqrt{3}}{1-\sqrt{3}}=\frac{(2-\sqrt{3})(1+\sqrt{3})}{-2}=\frac{-1+\sqrt{3}}{-2}=\frac{1-\sqrt{3}}{2}$，与上面一致。$\sqrt{12}=2\sqrt{3}$ 也在 $\mathbb{Q}(\sqrt{3})$ 中，它的共轭是 $-2\sqrt{3}=-\sqrt{12}$，所以照样把每个根号变号即可。前提是所有根号都是同一个 $\sqrt{t}$ 的有理倍数。像 $\sqrt{2}+\sqrt{3}$ 这样的数满足 $x^4-10x^2+1=0$，不在任何 $\mathbb{Q}(\sqrt{t})$ 中，也不是二次无理数。

现在可以陈述本讲的主定理了。它在第 19 讲末尾已经出现。

定理 20.6（Lagrange 定理，课程定理 70）　实数 x 是二次无理数，当且仅当它的简单连分数是**循环的**（定义见 §20.2）。

本讲后面证明这两个方向。

「循环 $\Rightarrow$ 二次无理数」属于 Euler（§20.3），「二次无理数 $\Rightarrow$ 循环」属于 Lagrange（§20.4）。后者难得多，是本讲的核心。

20.2　回顾：完全商与渐近分数

本节集中列出第 18 讲的记号和结论，后面反复使用。设 x 是无理数。令 $x_0 = x$，对 $i \geq 0$ 递归定义

$$a_i = \lfloor x_i \rfloor, \qquad x_{i+1} = \frac{1}{x_i - a_i}.$$

因为 x 是无理数，每个 $x_i - a_i \in (0,1)$，这个过程永不停止。a_i 称为**部分商**（partial quotient），x_i 称为**完全商**（complete quotient）。对 $i \geq 1$，$x_i > 1$，所以 $a_i \geq 1$。完全商就是从第 i 项开始的「尾巴」：

$$x_i = [a_i, a_{i+1}, a_{i+2}, \dots], \qquad x = [a_0, a_1, \dots, a_{n-1}, x_n].$$

渐近分数 $\frac{p_k}{q_k} = [a_0, \dots, a_k]$ 由递推

$$p_k = a_k p_{k-1} + p_{k-2}, \qquad q_k = a_k q_{k-1} + q_{k-2}, \qquad p_{-2} = 0,\ p_{-1} = 1,\ q_{-2} = 1,\ q_{-1} = 0$$

给出，满足 $p_k q_{k-1} - p_{k-1} q_k = (-1)^{k+1}$，并且 $1 = q_0 \leq q_1 < q_2 < \cdots$。把最后一项换成完全商 x_n，同样的递推给出

$$\boxed{x = \frac{x_n\, p_{n-1} + p_{n-2}}{x_n\, q_{n-1} + q_{n-2}} \qquad (n \geq 1).}$$

它的「行列式」$p_{n-1}q_{n-2} - p_{n-2}q_{n-1} = (-1)^n \neq 0$。

定义 20.7（循环与纯循环）　如果存在 $m \geq 0$ 和 $n \geq 1$，使得对所有 $i \geq m$ 都有 $a_{i+n} = a_i$，就称连分数 $x = [a_0, a_1, \dots]$ 是**循环的**（periodic），记作

$$x = [a_0, \dots, a_{m-1}, \overline{a_m, \dots, a_{m+n-1}}].$$

上横线下的部分称为**循环节**（period），满足条件的最小的 n 称为周期长度。如果可以取 $m = 0$，即 $x = [\overline{a_0, \dots, a_{n-1}}]$，就称 x 是**纯循环的**（purely periodic）。

例如 $\sqrt{2} = [1, \overline{2}]$ 是循环的但不是纯循环的，$\varphi = [\overline{1}]$ 是纯循环的。

引理 20.8（循环等价于完全商重复）　x 的连分数从第 m 项起以 n 为周期，当且仅当 $x_m = x_{m+n}$。特别地，x 纯循环当且仅当存在 $n \geq 1$ 使 $x_n = x$。

证明　若 $a_{i+n}=a_i$ 对 $i\geq m$ 成立，则 $x_m=[a_m,a_{m+1},\dots]$ 与 $x_{m+n}=[a_{m+n},a_{m+n+1},\dots]$ 的各项全同，所以相等。反过来，若 $x_m=x_{m+n}$，则 $a_m=\lfloor x_m\rfloor=\lfloor x_{m+n}\rfloor=a_{m+n}$，进而 $x_{m+1}=\frac{1}{x_m-a_m}=x_{m+n+1}$。如此继续，所有 $a_{i+n}=a_i$（$i\geq m$）。 ■

这个引理把「数字序列的循环」变成了「完全商这个数的重复」。下一节和 §20.4 都从这里出发：前者由 $x_m=x_{m+n}$ 列方程，后者设法证明某个完全商必定重复。

20.3　从循环连分数到二次无理数

命题 20.9（Euler 方向）　循环连分数的值是二次无理数。

证明　设 $x=[a_0,\dots,a_{m-1},\overline{a_m,\dots,a_{m+n-1}}]$，令 $\theta=x_m=[\overline{a_m,\dots,a_{m+n-1}}]$。由引理 20.8，$x_{m+n}=\theta$，即

$$\theta=[a_m,\dots,a_{m+n-1},\theta].$$

设 $\frac{P}{Q}$ 和 $\frac{P'}{Q'}$ 是有限连分数 $[a_m,\dots,a_{m+n-1}]$ 的最后两个渐近分数（$n=1$ 时 $\frac{P'}{Q'}=\frac{1}{0}$）。由 §20.2 的公式，

$$\theta=\frac{P\theta+P'}{Q\theta+Q'}\quad\Longrightarrow\quad Q\theta^2+(Q'-P)\theta-P'=0.$$

Q 是渐近分数的分母，$Q\geq 1$，所以这是一个真正的二次方程。θ 的连分数是无限的，由第 18 讲，θ 是无理数。因此 θ 是二次无理数。

若 $m=0$，$x=\theta$，证毕。若 $m\geq 1$，$x=[a_0,\dots,a_{m-1},\theta]=\frac{p_{m-1}\theta+p_{m-2}}{q_{m-1}\theta+q_{m-2}}$，行列式为 ±1。由引理 20.5，x 也是二次无理数。 ■

这个证明同时给出了求值的方法：先对循环节列出 $\theta=[\text{循环节},\theta]$，解出 θ，再把前面的非循环部分「套上去」。两根之积为 $-P'/Q<0$，所以方程一正一负两个根，θ 取正根。

想一想　证明中说「θ 是无理数，因为它的连分数是无限的」。能不能不用这个事实，直接从方程 $Q\theta^2+(Q'-P)\theta-P'=0$ 看出它没有有理根？

答　能。由 $PQ'-P'Q=(-1)^n$，判别式 $\Delta=(Q'-P)^2+4QP'=(P+Q')^2-4(-1)^n$。记 $k=P+Q'$。若 $\Delta=s^2$ $(s\geq 0)$，则 n 为偶数时 $(k-s)(k+s)=4$，n 为奇数时 $(s-k)(s+k)=4$。两个因子奇偶性相同，只能都等于 2，于是 n 偶时 $k=2$，n 奇时 $k=0$。但循环节各项 ≥ 1，所以 $P\geq 1$；$n\geq 2$ 时更有 $P\geq 2$、$Q'\geq 1$，$k\geq 3$。两种情形都不可能，所以 Δ 不是平方数。这条路走得通，但直接引用「无限连分数是无理数」省事得多。

■ **例 20.2**　求 $\theta=[\overline{1,2,3}]$ 的值。

解　有限连分数 $[1,2,3]$ 的渐近分数是

$$\frac{p_0}{q_0}=\frac{1}{1},\qquad \frac{p_1}{q_1}=\frac{2\cdot1+1}{2\cdot1+0}=\frac{3}{2},\qquad \frac{p_2}{q_2}=\frac{3\cdot3+1}{3\cdot2+1}=\frac{10}{7}.$$

由 $\theta=[1,2,3,\theta]$ 得

$$\theta=\frac{10\theta+3}{7\theta+2}\quad\Longrightarrow\quad 7\theta^2+2\theta=10\theta+3\quad\Longrightarrow\quad 7\theta^2-8\theta-3=0.$$

求根：$\theta=\frac{8\pm\sqrt{64+84}}{14}=\frac{8\pm2\sqrt{37}}{14}=\frac{4\pm\sqrt{37}}{7}$。取正根，

$$\theta=\frac{4+\sqrt{37}}{7}\approx 1.4404.$$

检验：$\theta\approx1.4404$，$\lfloor\theta\rfloor=1$；$\theta_1=1/(\theta-1)\approx2.2707$，部分商 2；$\theta_2=1/(\theta_1-2)\approx3.6943$，部分商 3；$1/(\theta_2-3)=\theta$，回到起点。这里始终用精确值递推，小数只用于展示。另一根 $\frac{4-\sqrt{37}}{7}\approx-0.2975$ 落在 $(-1,0)$ 中，这一点 §20.5 会解释。 ♦

■ **例 20.3**　**习题集 9 第 9(a) 题**　设 m 是正整数，求 $x=[\overline{m}]=[m,m,m,\ldots]$ 的值。

解　循环节只有一项，$x=[m,x]=m+\frac{1}{x}$，即 $x^2-mx-1=0$。两根之积为 -1，取正根：

$$[\overline{m}]=\frac{m+\sqrt{m^2+4}}{2}.$$

$m=1$ 时得到 $\frac{1+\sqrt5}{2}=\varphi$；$m=2$ 时得到 $1+\sqrt2$，与 $\sqrt2=[1,\overline{2}]$ 一致；$m=3$ 时得到 $\frac{3+\sqrt{13}}{2}\approx3.3028$。另一根 $\frac{m-\sqrt{m^2+4}}{2}=-\frac{1}{x}$ 也落在 $(-1,0)$ 中。 ♦

■ **例 20.4**　求 $x=[2,1,\overline{3}]$ 的值。

解　先处理循环部分：$\theta=[\overline{3}]=\frac{3+\sqrt{13}}{2}$（上例 $m=3$）。再套上前两项。$x=[2,1,\theta]=2+\dfrac{1}{1+\dfrac{1}{\theta}}=2+\frac{\theta}{\theta+1}$。计算

$$\frac{\theta}{\theta+1}=\frac{3+\sqrt{13}}{5+\sqrt{13}}=\frac{(3+\sqrt{13})(5-\sqrt{13})}{25-13}=\frac{2+2\sqrt{13}}{12}=\frac{1+\sqrt{13}}{6},$$

所以

$$x=2+\frac{1+\sqrt{13}}{6}=\frac{13+\sqrt{13}}{6}\approx2.7676.$$

也可以用渐近分数：$[2,1]$ 的渐近分数为 $\frac21,\frac31$，$x=\frac{3\theta+2}{\theta+1}$，代入得同样结果。

检验：$\lfloor x\rfloor=2$；$\frac{1}{x-2}\approx\frac{1}{0.7676}\approx1.3028$，$\lfloor\cdot\rfloor=1$；$\frac{1}{0.3028}\approx3.3028$，此后一直是 3。共轭 $\bar x=\frac{13-\sqrt{13}}{6}\approx1.5657$ 是正数，不在 $(-1,0)$ 中。 ♦

20.4　从二次无理数到循环连分数

由引理 20.8，要证明二次无理数 x 的连分数循环，只要证明某个完全商重复出现。完全商有无穷多个，怎样说明它们只能取有限多个值？

思路　把每个完全商都写成 $x_i=\frac{B_i+\sqrt{d}}{C_i}$，其中 d 始终不变。然后证明三件事：(1) B_i,C_i 都是整数；(2) 当 i 足够大时，B_i,C_i 落在一个有限范围内；(3) 有限多个「状态」(B_i,C_i) 要出现无穷多次，必有重复（抽屉原理），完全商随之重复。

20.4.1 完全商的整数表示

由命题 20.2，把 $x = x_0$ 写成 $\frac{B_0+\sqrt{d}}{C_0}$，$C_0 \mid d - B_0^2$（课程中称为第 0 步）。设已有 $x_i = \frac{B_i+\sqrt{d}}{C_i}$ 和 $a_i = \lfloor x_i \rfloor$。令 $B_{i+1} = a_i C_i - B_i$，则 $x_i - a_i = \frac{\sqrt{d}-B_{i+1}}{C_i}$，于是

$$x_{i+1} = \frac{C_i}{\sqrt{d} - B_{i+1}} = \frac{C_i(\sqrt{d}+B_{i+1})}{d - B_{i+1}^2} = \frac{B_{i+1}+\sqrt{d}}{C_{i+1}}, \qquad C_{i+1} = \frac{d - B_{i+1}^2}{C_i}.$$

因为 d 不是平方数，$d - B_{i+1}^2 \neq 0$，所以 $C_{i+1} \neq 0$。这就归纳地说明了每个完全商都具有 $\frac{B_i+\sqrt{d}}{C_i}$ 的形式，而且 (B_i, C_i) 由上面的递推唯一确定。

引理 20.10（整数性）　对所有 $i \geq 0$，B_i, C_i 都是整数，并且 $C_i \mid d - B_i^2$。

证明　对 i 归纳。$i = 0$ 正是第 0 步的要求。设结论对 i 成立。$B_{i+1} = a_i C_i - B_i$ 是整数，因为 a_i, C_i, B_i 都是整数。展开

$$C_{i+1} = \frac{d - (a_i C_i - B_i)^2}{C_i} = \frac{d - B_i^2}{C_i} - a_i^2 C_i + 2 a_i B_i,$$

由归纳假设 $\frac{d-B_i^2}{C_i}$ 是整数，所以 C_{i+1} 是整数。最后，$\frac{d-B_{i+1}^2}{C_{i+1}} = C_i$ 是整数，即 $C_{i+1} \mid d - B_{i+1}^2$。 ■

整除条件 $C_i \mid d - B_i^2$ 是归纳能走下去的原因。两个关系式 $B_i + B_{i+1} = a_i C_i$ 和 $C_i C_{i+1} = d - B_{i+1}^2$ 后面会反复用到。

算法 20.11（二次无理数的连分数）　输入标准形 $x = \frac{B_0+\sqrt{d}}{C_0}$（$C_0 \mid d - B_0^2$）。对 $i = 0, 1, 2, \ldots$ 依次计算

$$a_i = \left\lfloor \frac{B_i + \sqrt{d}}{C_i} \right\rfloor, \qquad B_{i+1} = a_i C_i - B_i, \qquad C_{i+1} = \frac{d - B_{i+1}^2}{C_i}.$$

一旦 $(B_j, C_j) = (B_i, C_i)$（$i < j$），就得到 $x = [a_0, \ldots, a_{i-1}, \overline{a_i, \ldots, a_{j-1}}]$。

图 20.1 展示了整数状态如何一步步产生部分商，并在状态重复时读出循环节。

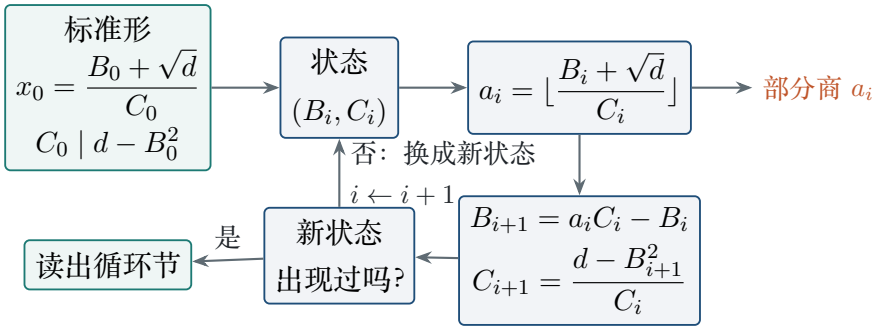


图 20.1　算法 20.11 的流程。每一步只用到当前状态 (B_i, C_i) 和固定的 d，所以下一状态完全由当前状态决定。

注 20.1　[只用整数算 a_i] 记 $r = \lfloor \sqrt{d} \rfloor$。当 $C > 0$ 时，对整数 k 有 $k \leq \frac{B+\sqrt{d}}{C} \iff kC - B \leq r$，

所以 $a=\lfloor\frac{B+r}{C}\rfloor$。当 $C<0$ 时，令 $N=-C>0$。由 $r<\sqrt{d}<r+1$，

$$\frac{-B-r-1}{N}<\frac{B+\sqrt{d}}{C}<\frac{-B-r}{N}.$$

两个端点的分子相差 1，所以它们之间没有整数；若端点本身为整数，中间值分别严格位于左端点上方或右端点下方。因此 $a=\lfloor\frac{-B-r-1}{-C}\rfloor$，取整仍只需整数运算。后面的证明将说明 C_i 只有开头有限多个可能为负。例如 $d=2,B=-4,C=-1$ 满足 $C\mid d-B^2$，且

$$x=\frac{-4+\sqrt{2}}{-1}=4-\sqrt{2}\approx 2.586.$$

此时 $r=1$，负分母公式给出 $a_0=\lfloor(4-1-1)/1\rfloor=2$；递推得到 $(B_1,C_1)=(2,2)$，对应 $x_1=(2+\sqrt{2})/2$。直接按定义计算也得到 $1/(x_0-2)=(2+\sqrt{2})/2$，这说明 $C_i<0$ 时不等号反向后要使用相应的取整公式。

■ **例 20.5**　**习题集 8 第 7 题：$\sqrt{15}$**　用状态算法求 $\sqrt{15}$ 的简单连分数。

解　标准形为 $x_0=(0+\sqrt{15})/1$，从 $(B_0,C_0)=(0,1)$ 开始。取 $r=\lfloor\sqrt{15}\rfloor=3$，依次得到

$$(B_0,C_0)=(0,1)\xrightarrow{a_0=3}(3,6)\xrightarrow{a_1=1}(3,1)\xrightarrow{a_2=6}(3,6).$$

状态 $(3,6)$ 在第 1 步和第 3 步重合，因此 a_1,a_2 循环重复，而首项 $a_0=3$ 留在循环外：

$$\sqrt{15}=[3,\overline{1,6}].$$

♦

■ **例 20.6**　用算法 20.11 求 $x=\dfrac{2+\sqrt{7}}{3}$ 的连分数。

解　$B_0=2$，$C_0=3$，$d=7$，$7-4=3$ 被 3 整除，已是标准形。$r=\lfloor\sqrt{7}\rfloor=2$，所有 $C_i>0$，用 $a_i=\lfloor\frac{B_i+2}{C_i}\rfloor$ 计算。

i	B_i	C_i	$x_i\approx$	a_i	$B_{i+1}=a_iC_i-B_i$	$C_{i+1}=(7-B_{i+1}^2)/C_i$
0	2	3	1.5486	1	$1\cdot3-2=1$	$(7-1)/3=2$
1	1	2	1.8229	1	$1\cdot2-1=1$	$(7-1)/2=3$
2	1	3	1.2153	1	$1\cdot3-1=2$	$(7-4)/3=1$
3	2	1	4.6458	4	$4\cdot1-2=2$	$(7-4)/1=3$
4	2	3	与 $i=0$ 的状态相同，开始重复			

所以 $\dfrac{2+\sqrt{7}}{3}=\overline{[1,1,1,4]}$，纯循环，周期长度 4。

用 Euler 方向验证：$[1,1,1,4]$ 的渐近分数是 $\frac{1}{1},\frac{2}{1},\frac{3}{2},\frac{14}{9}$，由 $\theta=\frac{14\theta+3}{9\theta+2}$ 得 $9\theta^2-12\theta-3=0$，即 $3\theta^2-4\theta-1=0$，正根 $\theta=\frac{4+\sqrt{28}}{6}=\frac{2+\sqrt{7}}{3}$，与出发点一致。♦

图 20.2 把这张表的四个不同状态连成闭合路径。箭头上的数是从该状态读出的部分商。

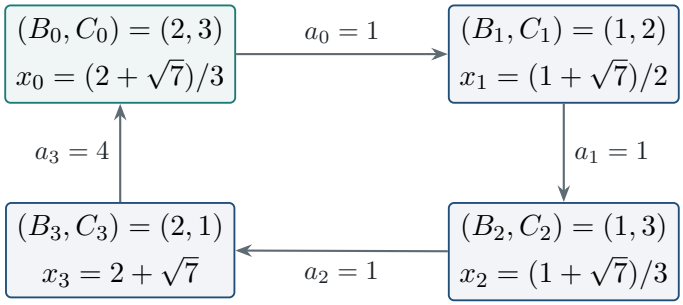


图 20.2　固定 $d = 7$ 后，四步回到完全相同的整数状态；循环节按箭头顺序读为 $1, 1, 1, 4$。

回到 $(2, 3)$ 时，完全商也回到 x_0，所以后续每一步都会重走这条路径。只看到某个部分商再次等于 1 还不够：图中前三个状态的部分商相同，下一状态却各不相同。判断循环必须比较整个状态。

Lagrange 方向的证明　还需证明算法最终一定重复。令 $y_i = \bar{x}_i$，即完全商 x_i 的共轭。把

$$x = \frac{p_{n-1}x_n + p_{n-2}}{q_{n-1}x_n + q_{n-2}}$$

中的 x 换成共轭，解出 y_n：

$$y_n = -\frac{q_{n-2}y_0 - p_{n-2}}{q_{n-1}y_0 - p_{n-1}} = -\frac{q_{n-2}}{q_{n-1}}\frac{y_0 - p_{n-2}/q_{n-2}}{y_0 - p_{n-1}/q_{n-1}}.$$

这里取 $n \geq 2$。渐近分数趋于 x，且 $y_0 \neq x$，所以最后一个分式充分大时为正；分母 q_{n-2}, q_{n-1} 为正，故 $y_n < 0$。

另一方面 $x_n > 1$，并且

$$x_n - y_n = \frac{2\sqrt{d}}{C_n}.$$

因此充分大时 $C_n > 0$。递推关系给出

$$C_nC_{n+1} = d - B_{n+1}^2.$$

右边为正整数，于是 $1 \leq C_n \leq d$，且 $|B_{n+1}| < \sqrt{d}$。所以充分大的下标处，整数对 (B_n, C_n) 只有有限多个可能值。

无限多个状态落在有限集合里，必有 $i < j$ 使 $(B_i, C_i) = (B_j, C_j)$，从而 $x_i = x_j$。由连分数算法的确定性，后续部分商逐项重复：$a_{i+k} = a_{j+k}$ 对所有 $k \geq 0$。因此 x 的连分数最终循环，Lagrange 方向得证。 ■

20.5　纯循环的共轭判别

Lagrange 定理只说明「最终循环」。什么时候非循环的开头也能删掉？答案由共轭的位置决定。

定理 20.12（Galois 判别法）　设 x 是二次无理数，$\bar{x}$ 是它的共轭。则 x 的简单连分数纯循环，当且仅当

$$x > 1, \qquad -1 < \bar{x} < 0.$$

思路　若 $x>1$ 且共轭在 $(-1,0)$，先用 Lagrange 定理得到某个尾巴重复，再用 $a_i=\lfloor x_i\rfloor$ 的递推把重复位置逐项向前推。反向方向则把一个完整周期当成未知尾巴，所得二次方程在 -1 与 0 之间变号。

证明　先设 $x>1$ 且 $-1<\bar{x}<0$。由 Lagrange 定理，存在 m,n 使 $x_m=x_{m+n}$。记 $y_i=\overline{x_i}$。若 $-1<y_i<0$，则 $a_i=\lfloor x_i\rfloor\geq 1$，且

$$y_{i+1}=\frac{1}{y_i-a_i}\in(-1,0).$$

归纳得所有 y_i 都在 $(-1,0)$。

因为 $x_m=x_{m+n}$，共轭也相等，即 $y_m=y_{m+n}$。从这一对相等开始向前归纳：若 $x_i=x_{i+n}$ 且 $y_i=y_{i+n}$，则由

$$a_{i-1}=\left\lfloor\frac{-1}{y_i}\right\rfloor,\qquad y_{i-1}=a_{i-1}+\frac{1}{y_i}$$

可知 $a_{i-1}=a_{i+n-1}$、$y_{i-1}=y_{i+n-1}$，并且 $x_{i-1}=a_{i-1}+1/x_i=x_{i+n-1}$。逐项向前推到 $i=1$，得到 $x_0=x_n$，所以连分数从开头就循环。

反过来，设 $x=[\overline{a_0,\dots,a_{n-1}}]$。周期性给出 $a_0=a_n\geq 1$，再由 $x_1>1$ 得 $x=a_0+1/x_1>1$。令 $p_j/q_j=[a_0,\dots,a_j]$，则

$$f(T)=q_{n-1}T^2+(q_{n-2}-p_{n-1})T-p_{n-2}$$

满足 $f(x)=0$。由 $p_{-1}=1,q_{-1}=0,p_{-2}=0,q_{-2}=1$ 以及递推式，

$$f(0)=-p_{n-2}<0,$$

$$f(-1)=q_{n-1}-q_{n-2}+p_{n-1}-p_{n-2}>0.$$

当 $n=1$ 时，直接用初值得 $f(-1)=a_0>0$；当 $n\geq 2$ 时，把最后一式展开为

$$(a_{n-1}-1)(q_{n-2}+p_{n-2})+(q_{n-3}+p_{n-3}),$$

第一项非负，第二项严格为正（$n=2$ 时为 $q_{-1}+p_{-1}=1$）。介值定理说明 f 在 $(-1,0)$ 有根。另一个根是 x 的共轭，因此 $-1<\bar{x}<0$。 ■

■ **例 20.7**　比较 φ、$\sqrt{2}$ 与 $[2,1,\overline{3}]$　用 Galois 判别法判断 φ、$\sqrt{2}$ 与 $[2,1,\overline{3}]$ 哪些是纯循环。

解　φ 的共轭 $(1-\sqrt{5})/2$ 在 $(-1,0)$，所以 $\varphi=[\overline{1}]$ 纯循环。$\sqrt{2}$ 的共轭是 $-\sqrt{2}<-1$，所以它不是纯循环。上面的 $[2,1,\overline{3}]$ 共轭约为 $1.5657>0$，因此只能最终循环，不能纯循环。 ♦

想一想　为什么判别法中的两个不等式都不能删掉？如果只知道 $\bar{x}\in(-1,0)$，取 $x=\sqrt{2}/2$；如果只知道 $x>1$，取 $x=\sqrt{2}$，它们分别说明缺少哪一个条件？

答　前者的共轭 $-\sqrt{2}/2\in(-1,0)$，但 $0<x<1$，故首项为 0；后续部分商全为正，首项 0 不会再出现，所以不能纯循环。后者满足 $x>1$，而共轭 $-\sqrt{2}<-1$，其连分数只有从第二项开始循环。两个例子分别说明 $x>1$ 和共轭区间条件不可删。

■ **例 20.8　习题集 9 第 7 题**　设 d 是正的非平方整数，求所有使

$$x = \frac{\lfloor \sqrt{d} \rfloor + \sqrt{d}}{c}$$

具有纯循环连分数的正整数 c。

解　令 $r = \lfloor \sqrt{d} \rfloor$。共轭为 $\bar{x} = (r - \sqrt{d})/c$。因为 $0 < \sqrt{d} - r < 1$ 且 $c \geq 1$，所以 $-1 < \bar{x} < 0$ 对每个正整数 c 都成立。Galois 判别法只剩条件 $x > 1$：

$$c < r + \sqrt{d}.$$

由于 $r < \sqrt{d} < r + 1$，右端落在 $(2r, 2r+1)$，故整数 c 必须满足 $1 \leq c \leq 2r$。答案是 $c = 1, 2, \ldots, 2\lfloor \sqrt{d} \rfloor$。 ♦

图 20.3 把两个共轭放到数轴上，显示一个位于判别区间内，一个在左侧之外。

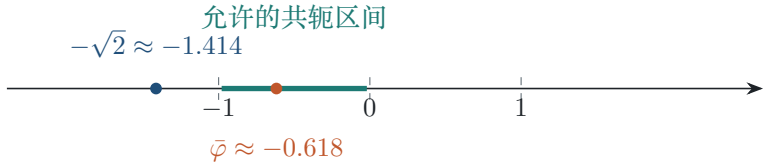


图 20.3　纯循环判别要求共轭严格落在 $(-1, 0)$；$\bar{\varphi}$ 在区间内，而 $\overline{\sqrt{2}} = -\sqrt{2}$ 在区间左侧。

20.6　Pell 方程与有理逼近

求解

$$X^2 - dY^2 = 1, \qquad X, Y \in \mathbb{Z}_{>0},$$

也可以转化为寻找 $\sqrt{d}$ 的好有理逼近。若 X/Y 很接近 $\sqrt{d}$，平方差就可能恰好等于 1；反过来，方程本身会强迫这种逼近极好。

定理 20.13（Pell 解给出渐近分数）　设 $d \geq 2$ 是非平方整数。若正整数 X, Y 满足 $X^2 - dY^2 = 1$，则既约分数 X/Y 是 $\sqrt{d}$ 的一个渐近分数。

证明　若 $g = \gcd(X, Y)$，则 g^2 整除 $X^2 - dY^2 = 1$，故 $g = 1$。又 $X/Y > \sqrt{d}$，并且

$$\left| \sqrt{d} - \frac{X}{Y} \right| = \frac{1}{Y(X + Y\sqrt{d})} = \frac{1}{Y^2(X/Y + \sqrt{d})} < \frac{1}{2Y^2},$$

因为 $X/Y + \sqrt{d} > 2\sqrt{d} \geq 2$。第 19 讲的 Legendre 判别法因此说明 X/Y 是渐近分数。 ■

■ **例 20.9　从 Pell 方程得到 $\sqrt{2}$ 的渐近分数**　验证 $(X, Y) = (3, 2)$ 给出 $\sqrt{2}$ 的渐近分数，并检查它的误差界。

解　$3^2 - 2 \cdot 2^2 = 1$，定理 20.13 保证 $3/2$ 是 $\sqrt{2}$ 的渐近分数；它也确实是 $[1, 2]$。误差为

$$\frac{3}{2} - \sqrt{2} = \frac{3 - 2\sqrt{2}}{2} \approx 0.0858 < \frac{1}{2 \cdot 2^2} = \frac{1}{8}.$$

第 18 讲的渐近分数序列中，满足 $p^2-2q^2=1$ 的项继续出现为 $17/12, 99/70, \ldots$。并非每个渐近分数都满足这个 Pell 方程，例如 $1/1$ 也是渐近分数，却有 $1^2-2\cdot 1^2=-1$。 ♦

图 20.4 中，Pell 解是双曲线上的整数格点；从原点连向这个格点，斜率就是有理数 X/Y。

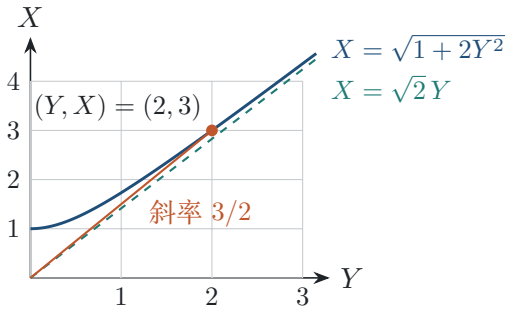


图 20.4　双曲线 $X^2-2Y^2=1$ 位于渐近线 $X=\sqrt{2}Y$ 上方；格点 $(Y,X)=(2,3)$ 给出从上方逼近 $\sqrt{2}$ 的斜率 $3/2$。

曲线上两者的竖直距离是 $X-\sqrt{2}Y = 1/(X+\sqrt{2}Y)$；再除以横坐标 Y，才得到斜率误差 $X/Y-\sqrt{2}$。这解释了证明中分母的两个因子从哪里来。

易错提醒　Legendre 判别法的结论方向不能倒过来：渐近分数未必满足 $1/(2q^2)$。例如 π 的渐近分数 $333/106$ 的误差约为 8.3214×10^{-5}，大于 $1/(2\cdot 106^2)\approx 4.4488\times 10^{-5}$。它不满足判别条件，却仍然是渐近分数。

本讲小结

- 二次无理数可写成 $\dfrac{B+\sqrt{d}}{C}$，并满足 $C\mid d-B^2$；共轭把 $\sqrt{d}$ 换成 $-\sqrt{d}$，保持四则运算。
- 循环连分数的重复尾巴满足一个分式线性方程，因此是二次无理数；例如 $\theta=[\overline{1,2,3}]$ 满足 $7\theta^2-8\theta-3=0$。
- 对完全商 $x_i=(B_i+\sqrt{d})/C_i$，递推为 $B_{i+1}=a_iC_i-B_i$、$C_{i+1}=(d-B_{i+1}^2)/C_i$；整除条件保证每一步仍为整数状态。
- 共轭完全商最终为负，因而 C_i 最终为正；由 $C_iC_{i+1}=d-B_{i+1}^2$ 得状态只有有限多个，抽屉原理给出 Lagrange 定理。
- Galois 判别法：二次无理数 x 纯循环，当且仅当 $x>1$ 且 $-1<\bar{x}<0$。
- Pell 方程 $X^2-dY^2=1$ 的正解满足 $\left|\sqrt{d}-X/Y\right|<1/(2Y^2)$，所以 X/Y 必是 $\sqrt{d}$ 的渐近分数。

练习

20.1 ★　把 $\dfrac{2+\sqrt{3}}{1+\sqrt{3}}$ 化成标准形，写出共轭和一个整系数二次方程。

20.2 ★　求 $[\overline{1,2,3}]$ 的二次方程和正根。

20.3 ★ 设 m 是正整数，求 $[\overline{m}]$ 的值。

20.4 ★★ 对 $x=(2+\sqrt{7})/3$，列出算法 20.11 的四个不同状态，并读出循环节。

20.5 ★★ 证明：若 $C_i \mid d-B_i^2$，则由状态递推得到 $C_{i+1}\in\mathbb{Z}$ 且 $C_{i+1}\mid d-B_{i+1}^2$。

20.6 ★★ 解释 Lagrange 方向证明中为什么 $y_n<0$ 会推出 $C_n>0$，以及为什么状态数有限。

20.7 ★★ 用 Galois 判别法判断 φ、$\sqrt{2}$ 和 $[2,1,\overline{3}]$ 哪些是纯循环。

20.8 ★★ 设 d 是正的非平方整数，求使 $(\lfloor\sqrt{d}\rfloor+\sqrt{d})/c$ 纯循环的所有正整数 c。

20.9 ★★★ 验证 $(X,Y)=(3,2)$ 与 $(17,12)$ 是 $X^2-2Y^2=1$ 的解，并说明它们为什么出现在 $\sqrt{2}$ 的渐近分数中。

20.10 ★★★ 对 $x=[\overline{m}]$，写出 p_n,q_n 的递推和特征根，说明 $p_n/q_n\to x$。

练习答案与提示

20.1 有 $x=(1+\sqrt{3})/2$，共轭为 $(1-\sqrt{3})/2$。根与系数关系给出 $2x^2-2x-1=0$。

20.2 由 $\theta=(10\theta+3)/(7\theta+2)$，得到 $7\theta^2-8\theta-3=0$，正根为 $(4+\sqrt{37})/7$。

20.3 $\theta=m+1/\theta$，所以 $\theta^2-m\theta-1=0$；正根为 $(m+\sqrt{m^2+4})/2$。

20.4 状态依次为 $(2,3),(1,2),(1,3),(2,1)$，下一状态回到 $(2,3)$，所以 $x=[\overline{1,1,1,4}]$。

20.5 $B_{i+1}=a_iC_i-B_i$ 是整数，并且

$$C_{i+1}=\frac{d-B_i^2}{C_i}-a_i^2C_i+2a_iB_i\in\mathbb{Z}.$$

再由 $C_iC_{i+1}=d-B_{i+1}^2$，得 $C_{i+1}\mid d-B_{i+1}^2$。

20.6 $x_n-y_n=2\sqrt{d}/C_n$，而 $x_n>1$、$y_n<0$，所以差为正，故 $C_n>0$。随后 $C_nC_{n+1}=d-B_{n+1}^2$ 是不超过 d 的正整数，给出 $1\le C_n\le d$、$|B_{n+1}|<\sqrt{d}$，状态只有有限多个。

20.7 φ 的共轭为 $(1-\sqrt{5})/2\in(-1,0)$，所以纯循环；$\sqrt{2}$ 的共轭为 $-\sqrt{2}<-1$，不是纯循环；$[2,1,\overline{3}]$ 的共轭约为 1.5657，不是纯循环。

20.8 令 $r=\lfloor\sqrt{d}\rfloor$。共轭自动在 $(-1,0)$，而 $x>1$ 等价于 $c<r+\sqrt{d}$。因为 $2r<r+\sqrt{d}<2r+1$，答案是 $1\le c\le 2r$。

20.9 $3^2-2\cdot2^2=1$、$17^2-2\cdot12^2=1$。对应分数 $3/2$、$17/12$ 的误差分别小于 $1/(2\cdot2^2)$、$1/(2\cdot12^2)$，由 Legendre 判别法它们是渐近分数；直接看 $\sqrt{2}=[1,\overline{2}]$ 也得到这两项。

20.10 $p_n=mp_{n-1}+p_{n-2}$、$q_n=mq_{n-1}+q_{n-2}$，初值为 $p_{-2}=0,p_{-1}=1,q_{-2}=1,q_{-1}=0$。特征方程 $z^2-mz-1=0$ 的根为 $\lambda_\pm=(m\pm\sqrt{m^2+4})/2$，令 $\Delta=\lambda_+-\lambda_-=\sqrt{m^2+4}$，则

$$p_n=\frac{\lambda_+^{n+2}-\lambda_-^{n+2}}{\Delta},\qquad q_n=\frac{\lambda_+^{n+1}-\lambda_-^{n+1}}{\Delta}.$$

这两个式子满足递推和初值。又 $\lambda_+=x$、$|\lambda_-|=1/\lambda_+<1$，所以

$$\frac{p_n}{q_n}=\lambda_+\frac{1-(\lambda_-/\lambda_+)^{n+2}}{1-(\lambda_-/\lambda_+)^{n+1}}\longrightarrow\lambda_+=x.$$

第 21 讲

平方根的周期与 Pell 方程

Square-root Periods and the Brahmagupta–Pell Equation

对应材料：MIT 18.781 Lecture 21 讲义；习题集 9 第 4 题。

本讲目标

(1) 会用整数状态递推求出 $\sqrt{d}$ 的最小周期，并解释周期末项 $2\lfloor\sqrt{d}\rfloor$ 的来源。

(2) 会证明并使用 $p_n^2 - dq_n^2 = (-1)^{n+1}C_{n+1}$，把逼近误差与整数方程联系起来。

(3) 掌握 $|N| < \sqrt{d}$ 时正本原解必来自渐近分数的理由，特别会处理 $N < 0$ 的情形。

(4) 根据最小周期长度的奇偶性，确定正、负 Pell 方程的全部解下标及最小正解。

(5) 理解边界条件与解的排序，能用已有解构造新解，并区分构造与完整性证明。

21.1 把平方根放进一个纯循环

$\sqrt{7} = [2, \overline{1, 1, 1, 4}]$ 的末项 4 是首项 2 的两倍，$\sqrt{13} = [3, \overline{1, 1, 1, 1, 6}]$ 也有相同现象。为什么循环从首项之后开始？为什么每次走到周期末尾都会出现两倍的整数部分？这两件事可以放在同一个纯循环中解释。

从现在起，除非另外说明，d 是**正的非平方整数**，记 $s = \lfloor\sqrt{d}\rfloor$。于是 $d \geq 2$，$s \geq 1$，且

$$s < \sqrt{d} < s + 1.$$

第 20 讲给出的 Galois 判别法说：二次无理数 u 的简单连分数纯循环，当且仅当 $u > 1$ 且 $-1 < \bar{u} < 0$。平方根本身的共轭为 $-\sqrt{d} < -1$，不满足后一个条件；向右平移 s 后，情形发生了变化。

命题 21.1（平方根的平移） 数 $u = s + \sqrt{d}$ 的简单连分数纯循环，且它的首项为 $2s$。因此存在唯一的最小正整数 r，使

$$u = [\overline{2s, a_1, \ldots, a_{r-1}}], \qquad \sqrt{d} = [s, \overline{a_1, \ldots, a_{r-1}, 2s}].$$

当 $r = 1$ 时，后式解释为 $\sqrt{d} = [s, \overline{2s}]$。

证明 $u > 1$，而它的共轭满足 $-1 < s - \sqrt{d} < 0$，所以由 Galois 判别法 u 纯循环。又 $2s < u < 2s + 1$，首项等于 $2s$。减去整数 s 只改变连分数的第一项，之后的完全商保持不变。于是 u 的循环数字从第二项起依次为 $a_1, \ldots, a_{r-1}, 2s$，给出 $\sqrt{d}$ 的展开。无理数的无限简单连分数唯一，因此两种写法各自都不存在更短的周期。 ■

这里的 r 始终指**最小周期长度**（least period length）。例如把一个长为 4 的循环节连写两遍，会得到长为 8 的重复块，但后面的解下标必须依据最小周期。写出一个重复块本身，还没有证明它最短；整数状态首次回到起点才给出最小周期。

■ **例 21.1　先看最短的周期**　求 $\sqrt{2}$ 的连分数，并比较 $\sqrt{2}$ 与 $1+\sqrt{2}$ 的循环起点。

解　有理化第一个余数的倒数：

$$\frac{1}{\sqrt{2}-1}=1+\sqrt{2}=2+(\sqrt{2}-1).$$

因此从第二个完全商开始，每次取走整数部分 2 后，余数都回到 $\sqrt{2}-1$，得到 $\sqrt{2}=[1,\overline{2}]$。而 $1+\sqrt{2}=[\overline{2}]$ 从第一项开始就是循环。最小周期都是 $r=1$，只是首项与周期的相对位置不同。 ♦

> **想一想**　既然 $\sqrt{d}$ 最终循环，为什么还要平移成 $s+\sqrt{d}$?
> **答**　最终循环只保证算法迟早重复，不能直接指出何时返回分母状态 $C=1$。平移后的每个完全商都满足同一个共轭区间条件；当 $C=1$ 时，该区间会迫使分子整数部分恰为 s。下一节就用这个唯一性把周期位置精确定位。平移的价值在于把一个存在性结论变成可以判断下标的结论。

为了看清循环，回忆第 20 讲的整数状态递推。若

$$x_i=\frac{B_i+\sqrt{d}}{C_i},\quad a_i=\lfloor x_i\rfloor,\quad B_{i+1}=a_iC_i-B_i,\quad C_{i+1}=\frac{d-B_{i+1}^2}{C_i},$$

则 $x_{i+1}=1/(x_i-a_i)$。从 $\sqrt{d}$ 出发时 $(B_0,C_0)=(0,1)$；从 $s+\sqrt{d}$ 出发时是 $(s,1)$。两者第一步得到的状态完全相同，因为前者 $a_0=s$，后者 $a_0=2s$，都给出 $B_1=s$、$C_1=d-s^2$。以后按同一个确定规则走下去。

■ **例 21.2　完整记录一个周期**　用整数状态计算 $\sqrt{7}$ 的最小周期。

解　取 $s=2$，初始状态 $(B_0,C_0)=(0,1)$、$a_0=2$。第一步是 $B_1=2$、$C_1=7-2^2=3$。以后的计算为

i	1	2	3	4	5
B_i	2	1	1	2	2
C_i	3	2	3	1	3
a_i	1	1	1	4	1

例如从 $(1,3)$ 出发，$a_3=\lfloor(1+\sqrt{7})/3\rfloor=1$，所以 $B_4=1\cdot3-1=2$、$C_4=(7-2^2)/3=1$。状态 $i=5$ 首次重现 $i=1$ 的状态，前四个状态互不相同，故最小周期为 $r=4$。 ♦

图 21.1 画的是平移后的纯循环：把刚才的第 4 个状态作为起点，四次递推恰好返回它。每个节点写出 (B,C)，箭头写出离开该节点时使用的部分商。

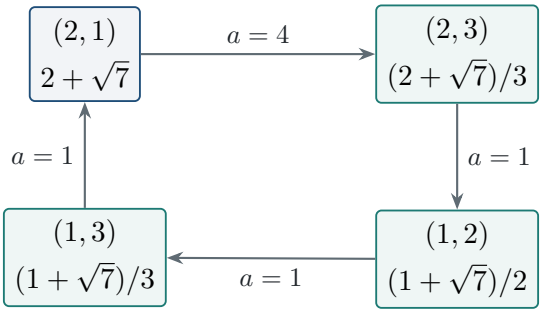


图 21.1　$2+\sqrt{7}$ 的四个整数状态构成一个纯循环，分母状态 $C=1$ 每四步出现一次。

图中每条箭头都同时改变 B 与 C，所以只比较部分商会丢失信息：连续三个 1 并不意味着周期长度是 1。能保证后续数字完全重现的是同一个完整状态再次出现。

21.2　完全商的分母如何记录范数

刚才的 $\sqrt{7}$ 状态中，C_i 按 $3,2,3,1$ 重复。另一方面，渐近分数 $2/1, 3/1, 5/2, 8/3$ 代入 p^2-7q^2，恰好得到 $-3,2,-3,1$。绝对值与状态分母相同，正负号逐项交替。要把这个观察用于求方程，先证明 $C_i=1$ 的出现位置。

引理 21.2（分母为一的返回位置）　设 $\sqrt{d}$ 的最小周期为 r，$x_i=(B_i+\sqrt{d})/C_i$ 是从 $x_0=\sqrt{d}$ 得到的完全商。则对 $i\ge 1$，有

$$C_i>0, \qquad C_i=1 \iff r \mid i.$$

当 $C_i=1$ 时，必有 $B_i=s$、$x_i=s+\sqrt{d}$。

思路　改从 $u=s+\sqrt{d}$ 出发。纯循环中的每个完全商都满足 Galois 的两个不等式；它们先给出 $C>0$，再在 $C=1$ 时确定 B。最后利用最小周期，判断返回 u 的准确时刻。

证明　把 u 的完全商记为 $u_i=(\widetilde{B}_i+\sqrt{d})/\widetilde{C}_i$。因为 u 纯循环，每个 u_i 都是同一个循环节的循环移位，仍纯循环，故

$$u_i>1, \qquad -1<\bar{u}_i<0.$$

于是 $u_i-\bar{u}_i=2\sqrt{d}/\widetilde{C}_i>0$，推出 $\widetilde{C}_i>0$。若 $\widetilde{C}_i=1$，则共轭不等式给出 $-1<\widetilde{B}_i-\sqrt{d}<0$，即 $\widetilde{B}_i<\sqrt{d}<\widetilde{B}_i+1$。整数 $\widetilde{B}_i$ 只能是 s，所以 $u_i=u$。

反过来，若 $u_i=u$，则

$$\widetilde{B}_i-\widetilde{C}_i s=(\widetilde{C}_i-1)\sqrt{d}.$$

左边是整数，$\sqrt{d}$ 无理，所以 $\widetilde{C}_i=1$，且 $\widetilde{B}_i=s$。

还需说明返回下标恰为 r 的倍数。先看 $u_0,\dots,u_{r-1}$：若其中 $u_i=u_j$、$0\le i<j<r$，确定性的递推就给出 $u_{i+k}=u_{j+k}$。取 $k=r-j$，得到 $u_{r-(j-i)}=u_r=u_0$，产生小于 r 的正周期，矛盾。因此这 r 个数互异。将任意 i 写成 $kr+t$、$0\le t<r$，有 $u_i=u_t$，故 $u_i=u_0$ 当且仅当 $t=0$。

上一节已经证明，u 与 $\sqrt{d}$ 从第 1 步起有完全相同的状态。把上述结论搬到 $i\ge 1$ 的 x_i，即得所需结论。特别地，C_i 不可能等于 -1。 ■

易错提醒　「$C_i = 1$ 当且仅当 $r \mid i$」在本讲只对平方根的规定初始状态成立，不能套到任意二次无理数。例如 $(2+\sqrt{7})/3$ 本身纯循环，最小周期也是 4，但从它开始时 $C_0 = 3$。改变初始完全商，会改变返回位置。即使对 $\sqrt{d}$，初值 $C_0 = 1$ 也要与后续的 $i \geq 1$ 分开处理。

定理 21.3（渐近分数的范数恒等式）　令 $p_n/q_n = [a_0, \ldots, a_n]$ 是 $\sqrt{d}$ 的第 n 个渐近分数，编号从 $n = 0$ 开始。则对所有 $n \geq 0$，

$$\boxed{p_n^2 - dq_n^2 = (-1)^{n+1} C_{n+1}.}$$

直观理解　p_n/q_n 交替落在 $\sqrt{d}$ 的两侧，所以 $p_n^2 - dq_n^2$ 的符号也应当交替。公式还指出，绝对值恰好等于下一完全商的整数分母。周期算法保留下来的状态，因此直接控制着一个整数二次式的值。

思路　使用第 18 讲的完全商表达式，把 x_{n+1} 代回 $\sqrt{d}$。乘去分母后，分别比较有理项和 $\sqrt{d}$ 项；再消掉 B_{n+1}，剩余的二阶行列式由渐近分数恒等式确定。

证明　记 $B = B_{n+1}$、$C = C_{n+1}$。由完全商公式，

$$\sqrt{d} = \frac{x_{n+1} p_n + p_{n-1}}{x_{n+1} q_n + q_{n-1}} = \frac{(B + \sqrt{d}) p_n + C p_{n-1}}{(B + \sqrt{d}) q_n + C q_{n-1}}.$$

分母为正，允许交叉相乘。将有理项和无理项分开：

$$dq_n + \sqrt{d}(B q_n + C q_{n-1}) = B p_n + C p_{n-1} + \sqrt{d}\, p_n.$$

由于 $\sqrt{d}$ 无理，有理系数分别相等，因此

$$p_n = B q_n + C q_{n-1}, \qquad dq_n = B p_n + C p_{n-1}.$$

第一式乘 p_n，第二式乘 q_n，两式相减，得到

$$p_n^2 - dq_n^2 = C(p_n q_{n-1} - q_n p_{n-1}) = C(-1)^{n+1}.$$

最后一步使用第 18 讲的行列式恒等式。它对 $n = 0$ 也成立：约定 $p_{-1} = 1$、$q_{-1} = 0$，此时行列式等于 -1，与公式相符。 ■

■ **例 21.3　逐项检查范数**　列出 $\sqrt{7}$ 的前八个渐近分数，并找出其中的单位范数。

解　用 $p_n = a_n p_{n-1} + p_{n-2}$、$q_n = a_n q_{n-1} + q_{n-2}$ 逐项递推。初值为 $p_{-2} = 0, p_{-1} = 1, q_{-2} = 1, q_{-1} = 0$，第一项是 $2/1$；下一项是 $(1 \cdot 2 + 1)/(1 \cdot 1 + 0) = 3/1$。完整结果见表 21.1。 ♦

n	0	1	2	3	4	5	6	7
a_n	2	1	1	1	4	1	1	1
p_n	2	3	5	8	37	45	82	127
q_n	1	1	2	3	14	17	31	48
C_{n+1}	3	2	3	1	3	2	3	1
$p_n^2 - 7q_n^2$	-3	2	-3	**1**	-3	2	-3	**1**

表 21.1　$\sqrt{7}$ 的范数绝对值按 $3,2,3,1$ 重复，单位范数出现在 $n=3,7$，且都为正。

表中的 $8^2-7\cdot 3^2=1$、$127^2-7\cdot 48^2=1$，已经给出两组 Pell 解。这里每隔四项才到达 $C_{n+1}=1$，而四步是偶数，符号没有改变。若周期为奇数，再走一个周期会发生什么？第四节给出完整答案。

方法：把单位范数改写为返回条件　对 $\sqrt{d}$ 的渐近分数，先用范数恒等式把 $p_n^2-dq_n^2=\pm 1$ 改写成 $C_{n+1}=1$，再用引理 21.2 改写成 $r\mid n+1$。求一个二元整数方程，由此变成选择一列确定的下标。

21.3　小范数解为什么一定是渐近分数

$3^2-7\cdot 1^2=2$，所以 $3/1$ 是 $\sqrt{7}$ 的一个有理逼近，也确实出现在表 21.1 中。若把右边换成别的整数 N，仍能保证每组解都来自渐近分数吗？关键在于 $|N|$ 与 $\sqrt{d}$ 的大小，以及分数是否已经约分。

定义 21.4（正解与本原解）　对整数方程 $x^2-dy^2=N$，满足 $x>0$、$y>0$ 的整数解称为**正解**（positive solution）；满足 $\gcd(x,y)=1$ 的整数解称为**本原解**（primitive solution）。

例如 $(3,1)$ 是 $x^2-7y^2=2$ 的正本原解；$(6,2)$ 是 $x^2-7y^2=8$ 的正解，但不是本原解，因为公因数为 2。$(1,0)$ 满足 $x^2-dy^2=1$，却不叫正解；它是边界上的平凡解。对 $N=\pm 1$，每组整数解自动本原，因为公因数的平方必须整除 1。

若 $g=\gcd(x,y)$，令 $x=gx'$、$y=gy'$，则

$$(x')^2-d(y')^2=N/g^2.$$

所以非本原解可以先约去公因数，代价是右边也要除以 g^2。从 ± 1 出发时这个问题自动消失；研究一般 N 时，不能只把分子分母约分，却忘记改变方程右边。

引理 21.5（从方程得到逼近界）　设 $\rho>0$，$\sqrt{\rho}$ 无理，s_1,t_1 是正整数。若

$$s_1^2-\rho t_1^2=\sigma,\qquad 0<\sigma<\sqrt{\rho},$$

则 $0<s_1/t_1-\sqrt{\rho}<1/(2t_1^2)$。

证明　因为 $\sigma > 0$，有 $s_1 > t_1\sqrt{\rho}$。对误差有理化：

$$0 < \frac{s_1}{t_1} - \sqrt{\rho} = \frac{\sigma}{t_1(s_1 + t_1\sqrt{\rho})} < \frac{\sigma}{2t_1^2\sqrt{\rho}} < \frac{1}{2t_1^2}.$$

两个严格不等号分别用到 $s_1 > t_1\sqrt{\rho}$ 与 $\sigma < \sqrt{\rho}$。 ■

> **定理 21.6（小范数判别）**　设 $d > 0$ 为非平方整数，$N \in \mathbb{Z}$、$|N| < \sqrt{d}$。若 (x, y) 是 $x^2 - dy^2 = N$ 的正本原解，则 x/y 是 $\sqrt{d}$ 的一个渐近分数。

思路　当 $N > 0$ 时，x/y 在平方根上方，可以直接使用引理 21.5 与第 19 讲的 Legendre 判别法。当 $N < 0$ 时，改看倒数 y/x：它在 $1/\sqrt{d}$ 的上方，从而回到已经处理的正差情形。

证明　若 $N = 0$，则正整数解会给出 $\sqrt{d} = x/y \in \mathbb{Q}$，与非平方条件矛盾，所以无需处理这一情形。

若 $N > 0$，在引理中取 $\rho = d$、$\sigma = N$、$s_1 = x$、$t_1 = y$，得到

$$\left|\sqrt{d} - \frac{x}{y}\right| < \frac{1}{2y^2}.$$

分数 x/y 已经既约，因此 Legendre 判别法说明它是渐近分数。

若 $N < 0$，将方程改写为

$$y^2 - \frac{1}{d}x^2 = \frac{|N|}{d}, \qquad 0 < \frac{|N|}{d} < \frac{1}{\sqrt{d}} = \sqrt{\frac{1}{d}}.$$

现在在引理中取 $\rho = 1/d$、$\sigma = |N|/d$、$s_1 = y$、$t_1 = x$。注意 $1/\sqrt{d}$ 仍无理，且 y/x 仍既约。于是 y/x 是 $1/\sqrt{d}$ 的渐近分数。

若 $\sqrt{d} = [a_0, a_1, \dots]$，则 $1/\sqrt{d} = [0, a_0, a_1, \dots]$。非零的有限截断满足

$$[0, a_0, \dots, a_n] = \frac{1}{[a_0, \dots, a_n]} = \frac{q_n}{p_n}.$$

因为 $y/x > 0$，它不可能是首项截断 0，故必有 $y/x = q_n/p_n$。取倒数得 $x/y = p_n/q_n$。正负两种情况均已覆盖。 ■

■ **例 21.4　正、负右端各做一次**　说明 $x^2 - 7y^2 = 2$ 的正本原解只能出现在 $\sqrt{7}$ 的渐近分数中，并解释 $(18, 5)$ 为什么是 $\sqrt{13}$ 的渐近分数。

解　先检查小范数界。$2 < \sqrt{7}$，所以第一问的每个正本原解都满足定理 21.6 的条件。具体的 $(3, 1)$ 给出

$$0 < 3 - \sqrt{7} = \frac{2}{3 + \sqrt{7}} < \frac{1}{2}.$$

这足以判定 $3/1$ 是渐近分数，与前面的状态表吻合；这里没有声称它是该方程的唯一解。

第二问先验算 $18^2 - 13 \cdot 5^2 = 324 - 325 = -1$，且 $\gcd(18, 5) = 1$。因为 $1 < \sqrt{13}$，小范数判别给出结论。若具体走一遍倒数证明，得到

$$5^2 - \frac{1}{13}18^2 = \frac{1}{13}, \qquad 0 < \frac{5}{18} - \frac{1}{\sqrt{13}} < \frac{1}{2 \cdot 18^2}.$$

所以 $5/18$ 是 $1/\sqrt{13}$ 的渐近分数，取倒数即可。证明中真正利用的，是误差小于由分母平方决定的阈值，而不只是「看起来很接近」。 ♦

易错提醒　当 $N<0$ 时，不能照搬 $x+y\sqrt{d}>2y\sqrt{d}$。此时 $x<y\sqrt{d}$，该不等号恰好反向。虽然有理化恒等式仍成立，这个简单的分母估计已经失效。定理 21.6 通过取倒数重新得到正差，从而合法地使用同一个误差引理。

想一想　把条件 $|N|<\sqrt{d}$ 删掉，是否仍能只查渐近分数？

答　不能。取 $d=2$、$(x,y)=(2,1)$，则 $x^2-2y^2=2$，是正本原解，但 $2/1$ 不在 $\sqrt{2}=[1,\overline{2}]$ 的渐近分数中。首个分数为 $1/1$；从第二项起分母至少为 2 并继续增长，故不可能再出现分母为 1 的 $2/1$。这里 $|N|=2>\sqrt{2}$，恰好落在判别范围之外。反过来，超出范围也不自动意味着分数不是渐近分数；定理给的是一个足够条件。

本原条件的作用也应说准确：它使方程中的 x,y 就是既约渐近分数的分子、分母。若暂时不要求本原，先除以 $g=\gcd(x,y)$，就得到右端 N/g^2，仍满足小范数界，故约分后的比值仍是渐近分数。原来的整数对则未必等于某个 (p_n,q_n)。例如 $8^2-17\cdot 2^2=-4$，且 $4<\sqrt{17}$；约分后为 $4/1$，它是 $\sqrt{17}$ 的首个渐近分数，但未约分的 $(8,2)$ 不是渐近分数的标准分子分母。

前后联系　第 19 讲的 Legendre 判别法把一个误差不等式转化为连分数结构；本节的方程把这个误差不等式自动提供出来。两步合起来，便能先证明候选解不遗漏，再用上一节的范数恒等式筛选候选。直接算几个渐近分数只能找到解，补上这一步后才能谈「所有解」。

21.4　周期奇偶性决定正、负 Pell 方程

$\sqrt{2}$ 的周期长为 1，前几个渐近分数范数依次为 $-1,1,-1,1$；$\sqrt{7}$ 的周期长为 4，每次单位范数却都是 1。这个差别只依赖于周期长度的奇偶性。

定理 21.7（正负 Pell 方程的全部解下标）　设 $d>0$ 为非平方整数，$\sqrt{d}$ 的最小周期长度为 r，渐近分数从 p_0/q_0 开始编号。所有正解 $x^2-dy^2=\pm 1$ 均来自

$$(x,y)=(p_{kr-1},q_{kr-1}),\qquad k=1,2,\dots,$$

且相应右端为 $(-1)^{kr}$。具体地：若 r 为偶数，负 Pell 无整数解，正 Pell 取全部 $k\ge 1$；若 r 为奇数，负 Pell 取正奇数 k，正 Pell 取正偶数 k。

思路　先由小范数判别保证所有正解都被渐近分数覆盖，再由范数恒等式和返回引理说明只有 $n+1=kr$ 的下标能留下。最后逐一检查留下的下标确实给出解，并读出符号。必要性和充分性都需要写出来。

证明　设 (x,y) 是正解，$x^2-dy^2=\pm 1$。若 $g\mid x,y$，则 $g^2\mid 1$，所以 $\gcd(x,y)=1$。又 $d\ge 2$，故 $1<\sqrt{d}$，可用定理 21.6。因此 $x/y=p_n/q_n$，且双方既约、分子分母都为正，得到 $(x,y)=(p_n,q_n)$。

由定理 21.3，

$$1 = |p_n^2 - dq_n^2| = C_{n+1}.$$

引理 21.2 给出 $r \mid n+1$，也就是 $n = kr - 1$、$k \geq 1$。反过来，每个这样的下标都满足 $C_{kr} = 1$，所以

$$p_{kr-1}^2 - dq_{kr-1}^2 = (-1)^{kr}.$$

p_{kr-1}, q_{kr-1} 为正整数，故这些候选全都给出所声称的正解，没有多余候选。

若 r 偶，则 kr 始终为偶数，右端只能为 1。负 Pell 若有任何整数解，其坐标均非零：$y = 0$ 会给出 $x^2 = -1$；$x = 0$ 会给出 $dy^2 = 1$，与 $d \geq 2$ 矛盾。取坐标绝对值会得到正解，所以负 Pell 实际上没有任何整数解。若 r 奇，则 kr 与 k 奇偶性相同，分类随即得到。 ■

> **推论 21.8（存在性与第一组解）**　对每个正的非平方整数 d，正 Pell 方程 $x^2 - dy^2 = 1$ 有无穷多组正解。其最小正解来自 $n = r - 1$（r 偶）或 $n = 2r - 1$（r 奇）。负 Pell 有整数解当且仅当 r 奇；有解时，其最小正解来自 $n = r - 1$。

证明　在定理 21.7 中，正 Pell 总能取到无穷多个 k。渐近分数互不相同，因而得到无穷多组解。分子递推中的部分商均为正，$p_1 = a_1 p_0 + 1 > p_0$，以后也严格递增；所以满足相应符号的第一个下标给出最小的 x。第五、六节会进一步解释最小性的计算与排序。 ■

■ **例 21.5　奇周期与偶周期的对照**　分别求 $x^2 - 2y^2 = \pm 1$ 与 $x^2 - 3y^2 = \pm 1$ 的前几组正解。

解　先求周期，再按下标筛选。对 $d = 2$，$\sqrt{2} = [1, \overline{2}]$，$r = 1$。渐近分数与范数为

n	0	1	2	3	4	5
p_n/q_n	$1/1$	$3/2$	$7/5$	$17/12$	$41/29$	$99/70$
$p_n^2 - 2q_n^2$	-1	1	-1	1	-1	1

于是负 Pell 的前三组正解为 $(1,1), (7,5), (41,29)$，正 Pell 的前三组为 $(3,2), (17,12), (99,70)$。例如 $41^2 - 2 \cdot 29^2 = 1681 - 1682 = -1$。

对 $d = 3$，整数状态给出 $\sqrt{3} = [1, \overline{1,2}]$、$r = 2$。其前四个渐近分数为

$$1/1, \quad 2/1, \quad 5/3, \quad 7/4,$$

范数依次为 $-2, 1, -2, 1$。所以正 Pell 的前两组正解为 $(2,1), (7,4)$，负 Pell 无解。也可以用模 3 独立核对负 Pell 无解：若 $x^2 - 3y^2 = -1$，就有 $x^2 \equiv 2 \pmod 3$，与平方剩余只有 $0, 1$ 矛盾。局部同余在这个例子里足以排除，周期定理则适用于所有非平方 d。 ♦

图 21.2 将「按周期返回」与「按符号筛选」分开画出。横轴标的是 k，对应渐近分数下标 $n = kr - 1$；图中没有把其他下标也当作解。

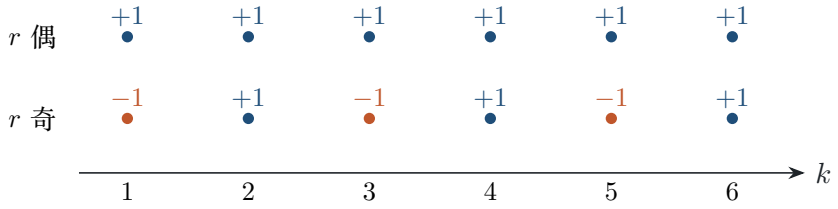


图 21.2　单位范数只在 $n = kr - 1$ 出现；偶周期全给出正范数，奇周期的范数则按 k 的奇偶交替。

图中蓝点对应正 Pell，橙点对应负 Pell。奇周期要走两遍才能首次回到正范数；这解释了为什么把 $r-1$ 一律当作正 Pell 基本解下标会出错。

21.5　把定理变成可检查的整数算法

$x^2 - 13y^2 = 1$ 的最小正解是 $(649, 180)$。逐个试 y 固然能找到它，却很难事先知道试到哪里；$d = 61$ 时这个问题更加突出。周期定理给出了有终止依据的算法：先算一个周期，根据奇偶决定算一遍还是两遍渐近分数。

算法 21.9（求 Pell 方程的最小正解）　输入正的非平方整数 d，令 $s = \lfloor \sqrt{d} \rfloor$，初值为 $(B, C, a) = (0, 1, s)$。

(1) 重复更新 $B \leftarrow aC - B$、$C \leftarrow (d - B^2)/C$、$a \leftarrow \lfloor (s + B)/C \rfloor$，记录部分商，直到第一次更新后 $C = 1$。记录的步数为最小周期 r，末项为 $2s$。

(2) 由 $a_0 = s$ 开始做渐近分数递推；周期后续部分商按刚才的数字循环使用。

(3) 正 Pell 在 r 偶时取 (p_{r-1}, q_{r-1})，r 奇时取 (p_{2r-1}, q_{2r-1})。负 Pell 在 r 偶时无解，r 奇时取 (p_{r-1}, q_{r-1})。

算法第一步的赋值有顺序：更新 B 时使用旧的 a, C, B；更新 C 时使用新的 B 和旧的 C；最后用新的 B, C 计算部分商。若写程序时把旧、新变量混在一起，会破坏恒等式 $C_i C_{i+1} = d - B_{i+1}^2$。用三个临时变量同时写出下一状态，通常更容易核对。

这里用 s 代替 $\sqrt{d}$ 取整没有近似误差。因为 B, C 是整数且 $C > 0$，写 $s + B = qC + t$、$0 \le t < C$，再令 $\delta = \sqrt{d} - s \in (0, 1)$，就有

$$\frac{B + \sqrt{d}}{C} = q + \frac{t + \delta}{C}, \qquad 0 < t + \delta < C.$$

所以其整数部分确为 $q = \lfloor (s + B)/C \rfloor$。整个过程只需要整数平方根、整数加减乘除；除法在状态递推中是整除，在求部分商时则向下取整。

■ **例 21.6　从五步周期算到基本解**　求 $x^2 - 13y^2 = 1$ 与 $x^2 - 13y^2 = -1$ 的最小正解。

解　先做状态表，判定最小周期。$s=3$，从 $(0,1)$ 出发，依次得到

i	1	2	3	4	5
B_i	3	1	2	1	3
C_i	4	3	3	4	1
a_i	1	1	1	1	6

第 5 步首次出现 $C=1$，因此 $r=5$，$\sqrt{13}=[3,\overline{1,1,1,1,6}]$。周期奇，故负 Pell 取 $n=4$，正 Pell 取 $n=9$。表 21.2 将这两次返回之间的所有渐近分数列全，便于检查是否少算或多算一项。 ♦

n	0	1	2	3	4	5	6	7	8	9
a_n	3	1	1	1	1	6	1	1	1	1
p_n	3	4	7	11	18	119	137	256	393	649
q_n	1	1	2	3	5	33	38	71	109	180
范数	-4	3	-3	4	-1	4	-3	3	-4	1

表 21.2　$\sqrt{13}$ 的第一个周期给出范数 -1 的 $(18,5)$，第二个周期才给出范数 1 的 $(649,180)$。

逐项回代端点：$18^2-13\cdot 5^2=-1$，$649^2-13\cdot 180^2=1$。表中所有更早的下标范数均不满足相应要求，再结合定理 21.7 的完整性，才证明了这两组解各自最小。只列出一组解再验算，还不足以证明最小性。

易错提醒　不要把周期末项 $a_r=2s$ 算进 p_{r-1}/q_{r-1}。对 $d=13$，负 Pell 的第一组解是 $[3,1,1,1,1]=18/5$；再添上 6 会变成 $119/33$，范数为 4，不再是单位范数。定位公式始终是 C_{n+1}，所以返回步 r 对应的是渐近分数下标 $r-1$。

■ **例 21.7　小参数也会给出很大的基本解**　用周期法确定 $x^2-61y^2=1$ 的最小正解。

解　先算完整周期，避免无界枚举 y。整数状态算法给出

$$\sqrt{61}=[7,\overline{1,4,3,1,2,2,1,3,4,1,14}],\qquad r=11.$$

周期为奇数，所以第一个正 Pell 解在 $n=21$，此前 $n=10$ 给出负 Pell 的最小正解。递推得到

$$(p_{10},q_{10})=(29718,3805),\qquad (p_{21},q_{21})=(1766319049,226153980).$$

两次代入分别等于 -1 与 1。因此所求最小正解是 $(1766319049,226153980)$。这个例子说明，d 的大小不能直接用来猜基本解的大小；决定算法停止位置的是周期返回，而数值规模由整个部分商序列共同决定。 ♦

上例的两个大数不必仅靠一次长乘法来信任。可以在递推每一步同时检查范数恒等式，或先核对负范数的 $(29718,3805)$，再用下一节的乘法把它平方，核对正范数解。两种独立算路一致，能更快发现抄写下标或数字时的错误。

想一想　算法为什么要求 d 是正的非平方整数？若 d 为平方数、为 0 或为负数，正 Pell 方程怎样处理？

答　若 $d=m^2>0$，方程分解为 $(x-my)(x+my)=1$，两个整数因子同为 1 或同为 -1，从而 $y=0$、$x=\pm1$，没有 $x,y>0$ 的解。此时 $\sqrt{d}$ 是有理数，连分数终止，不能沿用无限周期。若 $d=0$，方程变成 $x^2=1$，y 是任意整数，全部解为 $(\pm1,y)$。若 $d<0$，方程为 $x^2+|d|y^2=1$，有 $|x|\le 1$ 且 $|d|y^2\le 1$；$d=-1$ 时为 $(\pm1,0),(0,\pm1)$，$d\le -2$ 时只有 $(\pm1,0)$。这些情形都应先单独处理。

「非平方」无需加强成「无平方因子」。本讲允许 d 含平方因子，只要不是一个完全平方数，$\sqrt{d}$ 仍无理，所有证明都成立。保留这个范围很有用：若把 $d=f^2d_0$ 化成无平方因子的 d_0，方程会变成 $x^2-d_0(fy)^2=1$；新方程的解还必须满足第二坐标被 f 整除。因此简单地丢掉平方因子，会改变整数解问题。

21.6　从第一组解到有序的解列

已有 $3^2-2\cdot 2^2=1$。把 $3+2\sqrt{2}$ 平方得到 $17+12\sqrt{2}$，这恰好是前面连分数筛出的第二组正解。连分数算法与这种代数乘法为何一致？本讲先证明乘法确实产生解，并说明「最小」的含义；第 22 讲再证明每一组正解都会由最小解的幂产生。

命题 21.10（正解的排序）　固定 $d>0$ 及整数 N。若 (x_1,y_1)、(x_2,y_2) 都是 $x^2-dy^2=N$ 的正解，则

$$x_1<x_2 \iff y_1<y_2.$$

因此对正 Pell 解，最小的第一坐标与最小的第二坐标确定同一组解。

证明　两式相减得 $x_2^2-x_1^2=d(y_2^2-y_1^2)$。$d>0$，故两边的平方差符号相同。四个坐标均为正，所以平方函数在这些坐标上严格递增，于是两组严格不等式等价。等号情形也一致：一坐标相同，另一坐标便相同。 ■

正坐标条件用于从平方比较回到大小比较；若允许负坐标，$(3,2)$ 与 $(-17,12)$ 都满足 $x^2-2y^2=1$，但第一坐标变小、第二坐标变大。称「最小正解」时，正号范围必须先明确，不能把所有带符号整数解混在一起排序。

定义 21.11（基本解）　正 Pell 方程 $x^2-dy^2=1$ 中，第一坐标最小的正解称为**基本解**（fundamental solution），记为 (x_1,y_1)。

由推论 21.8，正解集合非空，良序原理保证第一坐标有最小值；相应 $y_1>0$ 由方程唯一确定。比如 $d=2$ 的基本解为 $(3,2)$，$d=3$ 的基本解为 $(2,1)$。边界解 $(1,0)$ 不在定义范围内，也不能用它生成任何非平凡解。

命题 21.12（范数的乘法）　若 $x^2-dy^2=M$、$u^2-dv^2=N$，则

$$X=xu+dyv,\qquad Y=xv+yu$$

满足 $X^2-dY^2=MN$。特别地，两组正 Pell 正解相乘，仍给出正 Pell 正解。

证明　直接计算，或利用共轭乘法。记 $\alpha=x+y\sqrt{d}$、$\beta=u+v\sqrt{d}$，则 $\alpha\beta=X+Y\sqrt{d}$，而

$$(X+Y\sqrt{d})(X-Y\sqrt{d})=(\alpha\bar{\alpha})(\beta\bar{\beta})=MN.$$

这是所需的范数等式。若 x,y,u,v 都为正且 $d>0$，则 $X,Y>0$，所以正性也保留。 ■

■ **例 21.8　用乘法核对连分数输出**　用 $(3,2)$ 构造 $x^2-2y^2=1$ 的后两组正解，再用 $(18,5)$ 核对 $d=13$ 的正 Pell 基本解。

解　把整数对相乘，避免依赖小数近似。对 $d=2$，

$$(3+2\sqrt{2})^2=(3^2+2\cdot 2^2)+(2\cdot 3\cdot 2)\sqrt{2}=17+12\sqrt{2},$$

再乘一次 $3+2\sqrt{2}$ 得

$$(17+12\sqrt{2})(3+2\sqrt{2})=(51+48)+(34+36)\sqrt{2}=99+70\sqrt{2}.$$

范数乘法保证两组解的范数都为 1，与例 21.5 中的输出一致。对 $d=13$，$18+5\sqrt{13}$ 的范数为 -1，平方的范数为 1，且

$$(18+5\sqrt{13})^2=(324+325)+180\sqrt{13}=649+180\sqrt{13}.$$

这与第 9 个渐近分数一致。负 Pell 的一组解平方总给出正 Pell 的解；本题中它之所以还是基本解，需要前面的周期下标判别来保证。 ♦

反复乘以基本解会产生无穷多组不同的正解：若 $X,Y>0$，则新第一坐标 $x_1X+dy_1Y>X$，因为 $x_1\geq 1$ 且第二项为正。这说明构造不会停在同一组解上。不过，仅证明幂次都是解，还没有排除两个幂次之间藏有其他正解。

前后联系　第 22 讲将证明完整的幂次生成定理：若 (x_1,y_1) 是基本解，则全部正解恰为 (x_n,y_n)，其中 $x_n+y_n\sqrt{d}=(x_1+y_1\sqrt{d})^n$、$n\geq 1$。证明会用一个任意正解除以适当的基本解幂，把它降到基本解之前，再利用最小性排除。这里给出该结论的陈述与证明位置，本讲已经完成的完整性结论则是定理 21.7 的连分数下标分类。

本讲小结

- 设 $s=\lfloor\sqrt{d}\rfloor$，则 $s+\sqrt{d}$ 纯循环，$\sqrt{d}=[s,\overline{a_1,\dots,a_{r-1},2s}]$，其中 r 取最小周期。
- 从 $\sqrt{d}$ 出发的整数状态满足 $C_i>0$，且对 $i\ge 1$，$C_i=1 \iff r\mid i$；返回时 $B_i=s$。
- 范数恒等式为 $p_n^2-dq_n^2=(-1)^{n+1}C_{n+1}$。注意下一状态的下标是 $n+1$。
- 若 $|N|<\sqrt{d}$，则 $x^2-dy^2=N$ 的正本原解给出渐近分数 x/y。负 N 可先对倒数 y/x 应用正差逼近界。
- 全部单位范数解位于 $n=kr-1$，范数为 $(-1)^{kr}$。r 偶时负 Pell 无解；r 奇时奇数 k 对应负 Pell，偶数 k 对应正 Pell。
- 正 Pell 总有无穷多正解，基本解下标为 $r-1$（r 偶）或 $2r-1$（r 奇）。同一方程的正解满足 $x_1<x_2 \iff y_1<y_2$。
- 范数相乘：$(x+y\sqrt{d})(u+v\sqrt{d})$ 的范数为 $(x^2-dy^2)(u^2-dv^2)$。基本解的正整数幂都是正解；穷尽全部解的幂次证明见第 22 讲。

练习

21.1 ★ 判断真假并说明理由：(a) 若 $\sqrt{d}$ 的周期为偶数，则 $x^2-dy^2=-1$ 无整数解；(b) 若周期为奇数，则 $p_{r-1}^2-dq_{r-1}^2=1$；(c) 只要 $x^2-dy^2=N$，x/y 就是渐近分数；(d) $C_0=1$ 足以说明 $\sqrt{d}$ 纯循环。

21.2 ★ 用整数状态求 $\sqrt{5}$ 的最小周期，求正、负 Pell 的最小正解，并分别代入验证。

21.3 ★ 对 $d=6$ 与 $d=11$，分别求最小周期、正 Pell 的基本解，并判断负 Pell 是否有整数解。

21.4 ★ 设 $d=7$，求所有满足 $x,y>0$、$y\le 50$ 的 $x^2-7y^2=1$ 的整数解。必须解释为何没有遗漏。

21.5 ★★ 不借助小数，证明算法中 $\lfloor(B+\sqrt{d})/C\rfloor=\lfloor(B+s)/C\rfloor$，其中 $B\in\mathbb{Z}$、C 为正整数、$s=\lfloor\sqrt{d}\rfloor$。若 $C<0$，原证明的哪一步不能照用？

21.6 ★★ 求 $x^2-10y^2=\pm 1$ 的最小正解，并通过负范数解的平方核对正范数解。

21.7 ★★ 用整数状态求 $\sqrt{14}$ 的周期，再求 $x^2-14y^2=1$ 的基本解，说明负 Pell 是否有解。

21.8 ★★ 证明若 $x^2-dy^2=N$、$g=\gcd(x,y)$，则 $g^2\mid N$。解释当 $|N|<\sqrt{d}$ 时非本原解与渐近分数的准确关系，并用 $(8,2)$、$d=17$ 说明。

21.9 ★★★ 若负 Pell 有正解 (u,v)，证明 $(u^2+dv^2,2uv)$ 是正 Pell 的正解。若 (u,v) 是负 Pell 的最小正解，能否仅凭范数乘法就断定平方后是正 Pell 的基本解？说明你已证明了什么，还缺什么。

21.10 ★★★ 设 $m\ge 1$ 为整数，令 $d=m^2+1$。求 $\sqrt{d}$ 的周期，进而求正、负 Pell 的最小正解，验证正 Pell 的解为 $(2m^2+1,2m)$。

练习答案与提示

21.1 (a) 真，由定理 21.7，所有单位范数下标都为 $kr-1$，偶周期使 kr 始终为偶数。(b) 假；当 r 奇时，第一周期给出范数 $(-1)^r=-1$，例如 $d=2$ 时 $p_0=q_0=1$。(c) 假；$d=2$、$(x,y)=(2,1)$ 给出 $N=2$，但 $2/1$ 不是渐近分数，见第三节的反例。(d) 假；平方根初始状态均有 $C_0=1$，但共轭为 $-\sqrt{d}<-1$，不满足纯循环判别条件。正确结论是平移后的 $s+\sqrt{d}$ 纯循环。

21.2 取 $s=2$。从 $(B_0,C_0)=(0,1)$ 出发，一步得到 $(B_1,C_1)=(2,1)$、$a_1=4$，所以 $\sqrt{5}=[2,\overline{4}]$，$r=1$。负 Pell 取 $n=0$，得到 $(2,1)$，验证 $4-5=-1$。正 Pell 取 $n=1$，由递推得 $p_1=4\cdot2+1=9$、$q_1=4$，得到 $(9,4)$，验证 $81-5\cdot16=1$。最小性来自定理 21.7 的首个允许下标。

21.3 $d=6$ 时 $s=2$，状态为 $(2,2)$、$(2,1)$，部分商为 $2,4$，所以 $\sqrt{6}=[2,\overline{2,4}]$、$r=2$。取 $n=1$，基本解为 $(5,2)$，验算 $25-24=1$。$d=11$ 时 $s=3$，状态为 $(3,2)$、$(3,1)$，部分商为 $3,6$，所以 $\sqrt{11}=[3,\overline{3,6}]$、$r=2$；基本解为 $(10,3)$，验算 $100-99=1$。两个周期均为偶数，故两道负 Pell 都无整数解。

21.4 $r=4$，所以正解仅出现在 $n=4k-1$。表 21.1 给出前两组 $(8,3)$、$(127,48)$。下一组的分母是 $765>50$：从 $p_7/q_7=127/48$ 继续递推，分母依次为 $223,271,494,765$，故 $q_{11}=765$。分母从 $n\ge1$ 起严格递增，后面的允许下标也不可能回到 50 以内。答案恰为这两组，分别满足 $64-63=1$、$16129-16128=1$。

21.5 写 $B+s=qC+t$，其中 $0\le t<C$；令 $\delta=\sqrt{d}-s\in(0,1)$，有 $0<t+\delta<C$。所以 $(B+\sqrt{d})/C=q+(t+\delta)/C$ 的整数部分为 q，等于 $(B+s)/C$ 的整数部分。若 $C<0$，除以 C 会反转不等号，$0<(t+\delta)/C<1$ 这一关键步骤无法照用。例如 $d=2,B=0,C=-1$ 时，左边为 -2，右边为 -1，所以等式确实不再成立。

21.6 $s=3$，一步状态为 $(3,1)$、部分商 6，所以 $\sqrt{10}=[3,\overline{6}]$，$r=1$。负 Pell 的最小正解为 $(3,1)$；正 Pell 的最小正解为 $(19,6)$，因为 $[3,6]=19/6$。验证 $9-10=-1$、$361-10\cdot36=1$。再算 $(3+\sqrt{10})^2=19+6\sqrt{10}$，得到同一组正解。

21.7 $s=3$。依次得到状态 $(3,5)$、$(2,2)$、$(2,5)$、$(3,1)$，对应部分商 $1,2,1,6$；所以 $\sqrt{14}=[3,\overline{1,2,1,6}]$，$r=4$。前四个渐近分数是 $3/1,4/1,11/3,15/4$。周期偶，正 Pell 的基本解为 $(15,4)$，验算 $225-14\cdot16=1$，负 Pell 无解。

21.8 由 $x=gx'$、$y=gy'$，可得 $N=g^2((x')^2-d(y')^2)$，所以 $g^2\mid N$。若原解为正且 $|N|<\sqrt{d}$，则 $x',y'>0$、$\gcd(x',y')=1$，且 $|N/g^2|<\sqrt{d}$，故 x'/y' 是渐近分数。原比值 x/y 与它相等，但原坐标不一定就是标准的 (p_n,q_n)。具体地 $8^2-17\cdot2^2=-4$、$g=2$；约分后 $(4,1)$ 满足范数 -1，且 $4/1$ 是 $\sqrt{17}$ 的首项，未约分的 $(8,2)$ 不能充作既约渐近分数的分子分母。

21.9 由范数乘法，$(u+v\sqrt{d})^2$ 的范数为 $(-1)^2=1$，其整数坐标为 $(u^2+dv^2,2uv)$，两坐标均为正。这证明平方后是正解；最小性还要求排除更小的正解，单凭范数乘法没有完成这一步。周期定理给出负、正 Pell 最小解的下标分别为 $r-1$、$2r-1$，但识别两个渐近分数之间的平方关系仍需额外证明。第 22 讲会补上降解论证，不能把「范数为 1」直接等同于「基本解」。

21.10 由 $m^2<d<(m+1)^2$，有 $s=m$ 且 d 非平方。一步到 $(B_1,C_1)=(m,1)$，下一部分商为 $2m$，故 $\sqrt{d}=[m,\overline{2m}]$、$r=1$。负 Pell 最小正解为 $(m,1)$，其范数为 -1。正 Pell 取 $n=1$，由 $[m,2m]=(2m^2+1)/(2m)$ 得基本解 $(2m^2+1,2m)$，回代核对：

$$(2m^2+1)^2-(m^2+1)(2m)^2=4m^4+4m^2+1-4m^4-4m^2=1.$$

$m\ge1$ 保证第二坐标为正。若 $m=0$，则 $d=1$ 为平方数，周期与正解结论均不能沿用。

第 22 讲

Pell 方程的全部解与四平方和定理

All Solutions to the Pell Equation and the Four-Square Theorem

对应材料：MIT 18.781 Lecture 22 讲义；本讲练习按原讲义的定理与两个丢番图应用改编。

本讲目标

(1) 掌握 Pell 方程 $x^2-dy^2=1$ 全部正整数解的形式：幂次生成定理的证明。

(2) 会由基本解 (x_1,y_1) 生成全部解，会算具体的 (x_n,y_n)；理解共轭与范数降序的关系。

(3) 会由 Pell 单位生成 $x^2-dy^2=N$ 的无穷多组解，并处理 $N=0$ 的例外；会构造连续三整数均为二平方和的例子。

(4) 理解四平方和定理的陈述：每个非负整数都是四个整数平方之和。

(5) 掌握 Hamilton 四元数 $\mathbb{H}$ 的乘法、共轭和平方范数，会展开 Euler 四平方恒等式。

(6) 理解四平方和定理的证明：先证引理 81（mp 是四平方和，$m<p$），再用无穷降递证明 $m=1$。

22.1 Pell 方程的全部正解

第 21 讲已经保证：若 d 是正的非平方整数，方程 $x^2-dy^2=1$ 有正整数解。以 $d=2$ 为例，$(3,2)$、$(17,12)$、$(99,70)$ 都是解，而且

$$(3+2\sqrt{2})^2=17+12\sqrt{2},\qquad (3+2\sqrt{2})^3=99+70\sqrt{2}.$$

继续乘下去当然能得到更多解，但这还没有回答「会不会漏掉某些解」。本节要证明：从最小正解出发取幂，恰好得到全部正整数解。

这里的正解始终指 $x>0$ 且 $y>0$。$(1,0)$ 满足方程，但不计入正解；它在乘法中充当单位元。由于 d 为非平方正整数，有 $d\geq 2$。对正解，

$$x=\sqrt{1+dy^2}>\sqrt{d}\,y,\qquad x-y\sqrt{d}=\frac{1}{x+y\sqrt{d}}>0.$$

因此共轭数虽然小于 1，却没有变成负数；后面乘不等式时就靠这个正性。

定义 22.1（基本解与对应的实数） 设 $d\geq 2$ 为非平方整数。$x^2-dy^2=1$ 中 x 最小的正解记为 (x_1,y_1)，称为基本解（fundamental solution）。记

$$\varepsilon=x_1+y_1\sqrt{d}>1.$$

这个最小值存在，因为正解存在，而正整数的非空子集有最小元。还需要确认：用 x 排序和用 y、$x+y\sqrt{d}$ 排序不会得出不同的「基本解」。对两个正解 (x,y) 与 (z,w)，

$$x<z \iff 1+dy^2<1+dw^2 \iff y<w.$$

若 $x<z$，两个坐标都严格较小，所以 $x+y\sqrt{d}<z+w\sqrt{d}$；反过来，若 $x\geq z$，则 $y\geq w$，实数的大小关系也随之反向。因此三种排序一致。

例如 $d=2$ 时 $y=1$ 给出 $x^2=3$，无整数 x；$y=2$ 给出 $x=3$，所以基本解是 $(3,2)$。$(17,12)$ 虽然也是正解，却不是基本解。边界解 $(1,0)$ 对应实数 1，把它当作基本解只会生成同一个解，不能生成任何正解。

定理 22.2（全部正解的幂次形式；原讲义定理 78）　设 $d\geq 2$ 是非平方整数，(x_1,y_1) 是基本解。则 $x^2-dy^2=1$ 的全部正整数解恰好由

$$x_n+y_n\sqrt{d}=(x_1+y_1\sqrt{d})^n=\varepsilon^n,\qquad n=1,2,3,\dots$$

给出，且每个正解对应唯一的 n。

直观理解　把一个任意正解对应的实数 $A=x+y\sqrt{d}$，放进相邻两项 ε^n 和 ε^{n+1} 之间。除以较小的一项后，结果落在 1 和 ε 之间。若这个结果仍是一个正整数解，就比基本解更小，产生矛盾。完整性证明的任务是核对除法之后两个系数确实都是正整数。

思路　先用范数乘法说明每个幂都给出整数解。再假设存在没有列出的正解，选出夹住它的相邻幂。除以左端点时改为乘以共轭，保持整数系数。最后用一个实数与其倒数之和、之差，分别证明新解的两个坐标为正。

证明　**第一步：幂次确实产生正解。**两个整系数表达式相乘时，有

$$(a+b\sqrt{d})(c+e\sqrt{d})=(ac+dbe)+(ae+bc)\sqrt{d}.$$

因此 $\varepsilon^n=x_n+y_n\sqrt{d}$ 的系数为整数，且从 $n=1$ 起两个系数都为正。共轭保持乘法，于是

$$x_n-y_n\sqrt{d}=(x_1-y_1\sqrt{d})^n=\varepsilon^{-n},\qquad x_n^2-dy_n^2=\varepsilon^n\varepsilon^{-n}=1.$$

由于 $\varepsilon>1$，序列 ε^n 严格递增且趋于无穷。

第二步：夹住一个假想的遗漏解。设 (s,t) 是正解，$A=s+t\sqrt{d}$ 不等于任何 ε^n。基本解的最小性给出 $A\geq\varepsilon$，等号又已排除，故存在 $n\geq 1$ 使

$$\varepsilon^n<A<\varepsilon^{n+1}.$$

乘以正数 $\varepsilon^{-n}=x_n-y_n\sqrt{d}$，得到

$$1<B=A\varepsilon^{-n}<\varepsilon.$$

把乘积展开为 $B=a+b\sqrt{d}$，则

$$a=sx_n-dty_n\in\mathbb{Z},\qquad b=tx_n-sy_n\in\mathbb{Z}.$$

两个展开式中有减法，所以还不能仅凭 $s,t,x_n,y_n>0$ 就断言 $a,b>0$。

第三步：核对范数和正性。范数为

$$(a+b\sqrt{d})(a-b\sqrt{d})=(s^2-dt^2)(x_n^2-dy_n^2)=1.$$

因此 $a-b\sqrt{d}=B^{-1}$，而 $B>1$ 蕴含 $0<B^{-1}<1$。相加、相减得

$$a=\frac{B+B^{-1}}{2}>0,\qquad b=\frac{B-B^{-1}}{2\sqrt{d}}>0.$$

这说明 (a,b) 是正整数解，又有 $1<a+b\sqrt{d}<\varepsilon$，与基本解的最小性矛盾。故不存在遗漏解。最后，若 $\varepsilon^r=\varepsilon^s$，严格递增性给出 $r=s$，所以表示唯一。 ■

图 22.1 用对数坐标画出证明中的两步。横坐标取 $\log_\varepsilon A$；上方省略中间的幂，仅两侧短区间各表示一个对数单位。相邻幂在完整对数轴上等距，在普通实数轴上则不等距。

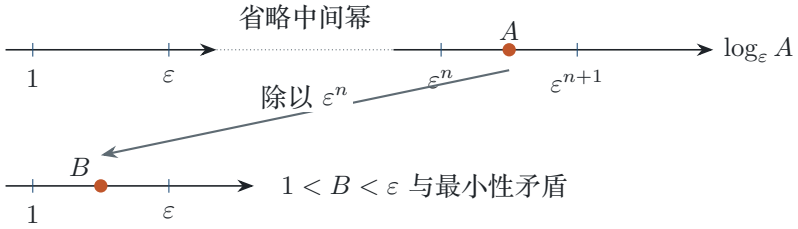


图 22.1　对数轴上省略中间幂；若 $\varepsilon^n<A<\varepsilon^{n+1}$，除以 ε^n 就得到 $1<B<\varepsilon$。

下方区间里没有正整数解。矛盾的依据是 B 同时满足整数系数、范数为 1 和坐标正性。

■ **例 22.1　生成并核对全部正解**　写出 $x^2-2y^2=1$ 的前五组正解，并求出其中满足 $y\le 100$ 的全部正解。

解　从基本解 $(3,2)$ 出发，每次乘 $3+2\sqrt{2}$，坐标递推为

$$x_{n+1}=3x_n+4y_n,\qquad y_{n+1}=2x_n+3y_n.$$

计算结果如下，每行最后一列都重新代回原方程核对。

n	x_n	y_n	$x_n^2-2y_n^2$
1	3	2	1
2	17	12	1
3	99	70	1
4	577	408	1
5	3363	2378	1

由定理 22.2 和 y_n 严格递增，$y\le 100$ 的全部正解为 $(3,2)$、$(17,12)$、$(99,70)$。代回核对只能确认它们是解；排除遗漏依靠全部解定理。 ♦

■ **例 22.2　基本解的选择决定是否漏解**　求 $x^2-3y^2=1$ 的前四组正解；若从 $x^2-2y^2=1$ 的解 $(17,12)$ 出发取幂，会发生什么？

解　先确定真正的基本解。$d=3$ 时 $y=1$ 就给出 $x=2$，所以取 $2+\sqrt{3}$。由

$$(x+y\sqrt{3})(2+\sqrt{3})=(2x+3y)+(x+2y)\sqrt{3},$$

前四组依次是 $(2,1)$、$(7,4)$、$(26,15)$、$(97,56)$。例如 $26^2-3\cdot 15^2=676-675=1$。

对 $d=2$，有 $17+12\sqrt{2}=(3+2\sqrt{2})^2$，所以从它取幂得到 $(3+2\sqrt{2})^{2n}$，只覆盖偶数指数。最初两组是 $(17,12)$ 和 $(577,408)$，漏掉了 $(3,2)$、$(99,70)$ 等。有一个正解足以生成无穷多解；要说生成全部正解，必须从基本解开始。 ♦

想一想　定理只写正解，那么整数解中的负号和 $y=0$ 怎么处理？

答　$y=0$ 时 $x^2=1$，得 $(1,0)$、$(-1,0)$。$y\neq 0$ 时，把 (x,y) 换成 $(|x|,|y|)$ 就得到正解，因而所有整数解是 $(\pm x_n,\pm y_n)$，两个符号独立，再加上 $(\pm 1,0)$。对应的乘法表达式可统一写为 $\pm\varepsilon^n$，其中 $n\in\mathbb{Z}$；负指数用共轭实现，例如 $(3+2\sqrt{2})^{-1}=3-2\sqrt{2}$。

非平方条件保证存在非平凡的基本解。若把 $d=4$ 代入，有 $(x-2y)(x+2y)=1$。两个整数因子只能同时为 1 或同时为 -1，从而 $y=0$；没有正解可供取幂。若把 d 改成负数，例如 $d=-1$，方程变成 $x^2+y^2=1$，整数解只有四个。正且非平方这两个条件共同保证了上一讲的存在性结论。

22.2　把 Pell 解用于其他整数方程

$3^2+4^2=5^2$ 是一组相邻整数的平方和恰好为平方数的例子。能否再找一组，或者一次证明有无穷多组？直接搜索会遇到越来越大的数；把题目改写成范数方程后，可以系统地生成答案。

对整数 a,b 定义二次表达式的范数

$$N_d(a+b\sqrt{d})=(a+b\sqrt{d})(a-b\sqrt{d})=a^2-db^2.$$

共轭保持乘法，所以 $N_d(uv)=N_d(u)N_d(v)$。因此乘一个范数为 1 的 Pell 解，会保留原来的右端项。接下来既要检查范数不变，也要检查所得解彼此不同。

命题 22.3（一般右端项；原讲义命题 79 的非零版本）　设 $d\geq 2$ 是非平方整数，$M\in\mathbb{Z}\setminus\{0\}$。若方程

$$x^2-dy^2=M$$

有一组整数解 (a,b)，则它有无穷多组整数解。具体地，把

$$x_n+y_n\sqrt{d}=(a+b\sqrt{d})\varepsilon^n,\qquad n=0,1,2,\dots$$

展开，即得到一列彼此不同的解。

证明　记 $\beta=a+b\sqrt{d}$。每个乘积的系数都是整数，而

$$N_d(\beta\varepsilon^n)=N_d(\beta)N_d(\varepsilon)^n=M.$$

由于 $M \neq 0$，有 $\beta \neq 0$。若两组系数相同，则 $\beta\varepsilon^r = \beta\varepsilon^s$。除以非零数 β 得 $\varepsilon^r = \varepsilon^s$，所以 $r = s$，证明各组解彼此不同。 ■

原讲义把右端项允许为任意整数，却没有单独说明 $M = 0$。非平方 d 下，$x^2 = dy^2$ 若有 $y \neq 0$，就会使 $\sqrt{d} = |x/y|$ 为有理数，矛盾。因此 $M = 0$ 只有 $(0,0)$ 一组整数解；将它反复相乘仍得到零。命题中的 $M \neq 0$ 正是为了排除这个例外。

易错提醒　「有一组解就有无穷多组解」不表示每个右端项都有解。例如 $x^2 - 2y^2 = 3$ 模 8 不可解：y 偶时右侧要求 $x^2 \equiv 3$，y 奇时要求 $x^2 \equiv 5$，而平方模 8 只可能为 $0, 1, 4$。先找到种子解，才能使用命题 22.3。

■ **例 22.3　连续三个数都是二平方和**　证明：有无穷多组三个连续的正整数，每个数都能写成两个整数平方之和；允许其中一个平方为 0。

解　让中间一个数本身成为平方数。取 $n = x^2$，则

$$n = x^2 + 0^2, \qquad n + 1 = x^2 + 1^2.$$

还需安排 $n - 1$。若 $x^2 - 2y^2 = 1$，就有 $n - 1 = y^2 + y^2$，所以任意正 Pell 解都能完成构造。基本解 $(3, 2)$ 给出

$$8 = 2^2 + 2^2, \qquad 9 = 3^2 + 0^2, \qquad 10 = 3^2 + 1^2.$$

下一解 $(17, 12)$ 给出 $288 = 12^2 + 12^2$、$289 = 17^2 + 0^2$、$290 = 17^2 + 1^2$；再下一解 $(99, 70)$ 给出 $9800, 9801, 9802$。由于 x_n 严格递增，所生成的三元组彼此不同，而且 $x_n \geq 3$ 保证三项全为正。这个构造证明存在无穷多组，并没有声称列出了所有符合题意的三元组。 ♦

■ **例 22.4　相邻整数的平方和还是平方**　证明：有无穷多个正整数 n，使 $n^2 + (n+1)^2$ 为完全平方数，并写出最初三个构造值。

解　先配方，把 n 的一次项并入一个平方。令 $n^2 + (n+1)^2 = m^2$，逐步整理为

$$2n^2 + 2n + 1 = m^2 \iff (2n+1)^2 - 2m^2 = -1.$$

记 $\ell = 2n + 1$。方程 $\ell^2 - 2m^2 = -1$ 有种子解 $(1, 1)$，所以乘 $3 + 2\sqrt{2}$ 生成无穷多组整数解。具体序列为

$$\ell_k + m_k\sqrt{2} = (1 + \sqrt{2})(3 + 2\sqrt{2})^k, \qquad k = 0, 1, 2, \ldots.$$

从 $(1, 1)$ 开始，递推 $\ell_{k+1} = 3\ell_k + 4m_k$、$m_{k+1} = 2\ell_k + 3m_k$，得到

$$(\ell_k, m_k) = (1, 1), (7, 5), (41, 29), (239, 169), \ldots.$$

反向变换也要核对：方程模 2 给出 $\ell^2 \equiv 1 \pmod 2$，所以 ℓ 为奇数，$n = (\ell - 1)/2$ 是整数。代回原式，正好得到 $n^2 + (n+1)^2 = m^2$。$k = 0$ 给出边界值 $n = 0$；从 $k = 1$ 起 n 为正，最初三个构

造值为

$$\begin{aligned} n=3: &\quad 3^2+4^2=25=5^2,\\ n=20: &\quad 20^2+21^2=841=29^2,\\ n=119: &\quad 119^2+120^2=28561=169^2. \end{aligned}$$

ℓ_k 严格递增，故 n 也严格递增。这同时核对了整数性、正性、原方程以及无穷多个不同值，不能只停在「有一个负 Pell 解」这一步。 ♦

想一想　一般右端项的方程也能像 $M=1$ 那样，任取一个种子就生成全部正解吗?

答　未必。例如 $x^2-2y^2=17$ 有 $(5,2)$ 和 $(7,4)$ 两组正解。若后者能由前者乘一个整数次的 $3+2\sqrt{2}$ 得到，则二者的商必须有整数系数。但是

$$\frac{7+4\sqrt{2}}{5+2\sqrt{2}}=\frac{19+6\sqrt{2}}{17},$$

系数都不为整数，故它们不在同一列中。幂次定理对右端项 1 的完整性，依靠基本解与范数 1 的逆元；一般右端项的命题只保证一列无穷多解。

在上面的反例中，两组种子继续乘 Pell 单位，分别得到 $(23,16)$ 和 $(37,26)$，范数都仍为 17。把「保持方程的乘法操作」与「所有解的分类」分开，是读这类丢番图论证时需要始终保持的区别。

方法：先找一个范数，再乘范数为一的元素　遇到 $x^2-dy^2=M$，先用同余判断可能性并找一组种子解，再乘 $x_1+y_1\sqrt{d}$ 生成新解。若要证明无穷多组，检查种子不为零和新解彼此不同；若要声称全部解，还必须另给完整性证明。

22.3　四元数与四平方恒等式

两平方和满足

$$(a^2+b^2)(c^2+e^2)=(ac-be)^2+(ae+bc)^2,$$

因为复数乘法保留平方模。想证明四平方和的乘积仍是四平方和，也可以寻找一个有四个实坐标、范数满足乘法性质的代数。四元数恰好提供这个工具；本讲只使用它的加法、乘法、共轭和整数系数，不需要讨论更深的结构。

定义 22.4（Hamilton 四元数）　**四元数**（quaternion）的集合记为

$$\mathbb{H}=\{\,a+bi+cj+dk \mid a,b,c,d\in\mathbb{R}\,\}.$$

两个四元数相等，当且仅当四个系数分别相等；加法逐坐标进行。乘法对加法满足分配律，实数与所有元素交换，并规定

$$i^2=j^2=k^2=-1,\qquad ij=k,\quad jk=i,\quad ki=j,$$

交换后两两相乘的顺序会改变符号：$ji=-k$、$kj=-i$、$ik=-j$。

例如 $2+3i$ 是 $c=d=0$ 的四元数；实数 5 是 $b=c=d=0$ 的情形；0 的四个坐标全为零。形式表达式 $1+2i+3j$ 与 $1+2i+3k$ 不相等，因为 j 坐标不同。这一点和只在复数平面里使用一个虚数单位的习惯不同。

乘法规则可以排成表。行标表示左因子，列标表示右因子，例如第 i 行、第 j 列读作 $ij=k$。

$\cdot$	1	i	j	k
1	1	i	j	k
i	i	-1	k	$-j$
j	j	$-k$	-1	i
k	k	j	$-i$	-1

由此 $ijk=k^2=-1$。但 $(ij)k=-1$ 与 $(ji)k=1$，说明交换次序确实会改变结果。若在展开中擅自把 ji 改写成 ij，不仅个别符号错，后面的范数恒等式也会被破坏。

易错提醒　四元数的乘法不交换：$ij=k$，而 $ji=-k$。因此不能把 $q\bar{r}$ 改成 $\bar{r}q$，也不能随意调整多个因子的顺序。乘法仍然满足结合律，所以可以改变括号；下一段给出保证结合律的一种直接检验。

准备与可暂缓部分。下一段的结合律验证使用复数、矩阵乘法及其结合律。若尚未学过矩阵乘法，可先将结合律作为本段引用的代数事实，继续按乘法表核对后面的共轭与范数计算，补学后再回读这项验证。

结合律可以通过复矩阵验证。把四元数对应到

$$\Phi(a+bi+cj+dk)=\begin{pmatrix} a+bi & c+di \\ -c+di & a-bi \end{pmatrix}.$$

这里矩阵里的 i 是复数单位。四个实坐标可以从矩阵各项恢复，所以 Φ 为单射。将 $1,i,j,k$ 分别代入后相乘，得到的乘法表正好是上表；再利用分配律，就有 $\Phi(qr)=\Phi(q)\Phi(r)$。矩阵乘法满足结合律，故

$$\Phi((qr)s)=\Phi(q)\Phi(r)\Phi(s)=\Phi(q(rs)).$$

由单射性得 $(qr)s=q(rs)$。这只用来给乘法规则一个可靠的实现；后面的计算仍直接用四个实坐标。

定义 22.5（共轭与平方范数）　若 $q=a+bi+cj+dk$，定义其**共轭**（conjugate）与**平方范数**（norm）为

$$\bar{q}=a-bi-cj-dk,\qquad N(q)=q\bar{q}=\bar{q}q=a^2+b^2+c^2+d^2.$$

本讲沿用原讲义把四个平方的和作为范数，因此 $N(q)$ 就是通常欧氏长度的平方，不再开平方。例如 $N(1+2i+3j+4k)=30$，$N(i)=1$，$N(0)=0$。对整数系数的 q，$N(q)$ 是非负整数；$N(q)=0$ 当且仅当 $q=0$。实数 t 乘四元数时有 $N(tq)=t^2N(q)$，所以除以整数 ℓ 会把范数除以 ℓ^2，这一平方在递降中不能遗漏。

命题 22.6（共轭反序与范数乘法）　对任意 $q, r \in \mathbb{H}$，有

$$\overline{qr} = \bar{r}\bar{q}, \qquad N(qr) = N(q)N(r).$$

特别地，两个整数系数四元数相乘仍有整数系数，且其范数为原来两个四平方和的乘积。

思路　共轭公式在基元素 $1, i, j, k$ 上核对后，由分配律推广。注意共轭会把乘法顺序反过来。范数计算随后只移动括号，并利用 $r\bar{r}$ 是实数这一点，避免使用不存在的交换律。

证明　共轭对加法和实数倍都保持运算。若某个基元素为 1，反序公式直接成立。若两个虚单位相同，例如 i, i，则 $\overline{i^2} = -1 = (-i)(-i)$。若 u, v 是两个不同虚单位，乘法表给出 $uv = \pm w$ 和 $vu = -uv$，故

$$\overline{uv} = -uv = vu = (-v)(-u) = \bar{v}\bar{u}.$$

再展开一般的 q, r，每一项都满足这个公式，求和便得到 $\overline{qr} = \bar{r}\bar{q}$。于是

$$N(qr) = (qr)\overline{qr} = qr\bar{r}\bar{q} = qN(r)\bar{q} = N(r)q\bar{q} = N(q)N(r).$$

其中将 $N(r)$ 移到前面合法，因为它是实数。整数系数的乘积仍为整数系数，则由乘法表和分配律直接得到。 ■

■ **例 22.5　把四元数乘法完整展开**　计算 $(1 + 2i + 3j + 4k)(2 - i + j)$，并核对范数乘法。

解　按实部、i、j、k 四个坐标分别收集。设一般乘积为

$$(a + bi + cj + dk)(e + fi + gj + hk) = A + Bi + Cj + Dk.$$

从乘法表得到

$$A = ae - bf - cg - dh, \quad B = af + be + ch - dg,$$
$$C = ag - bh + ce + df, \quad D = ah + bg - cf + de.$$

题中 $(a, b, c, d) = (1, 2, 3, 4)$、$(e, f, g, h) = (2, -1, 1, 0)$，故

$$A = 2 + 2 - 3 = 1, \quad B = -1 + 4 - 4 = -1,$$
$$C = 1 + 6 - 4 = 3, \quad D = 2 + 3 + 8 = 13.$$

乘积为 $1 - i + 3j + 13k$。原来两个范数是 30 与 6，乘积的范数是 $1 + 1 + 9 + 169 = 180 = 30 \cdot 6$。这里 C 中的 df 为负，而 D 中的 $-cf$ 为正，逐坐标收集可以避免只凭记忆套符号。 ♦

上例的一般坐标公式，结合命题 22.6，给出 **Euler 四平方恒等式**(Euler's four-square identity)：

$$\begin{aligned}(a^2 + b^2 + c^2 + d^2)(e^2 + f^2 + g^2 + h^2) \\ = (ae - bf - cg - dh)^2 + (af + be + ch - dg)^2 \\ + (ag - bh + ce + df)^2 + (ah + bg - cf + de)^2.\end{aligned}$$

它对任意实数都成立，尤其对整数成立。因此**四平方和的集合对乘法封闭**。若不想引入四元数，也能逐项展开上式：四次混合项成对抵消。四元数让这些符号的来源更清楚，还保留了递降所需的逐坐标同余信息。

■ **例 22.6　由两个表示乘出一个新表示**　已知 $7=2^2+1^2+1^2+1^2$、$13=2^2+2^2+2^2+1^2$，构造 91 的四平方表示。

解　把两组坐标放进四元数 $q=2+i+j+k$、$r=2+2i+2j+k$。用刚才的坐标公式，

$$qr=-1+5i+7j+4k.$$

于是 $N(qr)=N(q)N(r)=7\cdot 13=91$，具体为

$$91=(-1)^2+5^2+7^2+4^2=1+25+49+16.$$

负坐标对表示没有影响，因为最终取平方。这个例子说明：一旦素数都有四平方表示，合数的表示就可以沿着素因数分解逐次相乘构造。　♦

22.4　从一个素数到它的较小倍数

$7=2^2+1^2+1^2+1^2$ 需要四个非零平方；$13=3^2+2^2+0^2+0^2$ 只需要两个。允许零坐标以后，两种情形可以放进同一个命题中。

定理 22.7（Lagrange 四平方和定理；原讲义定理 80）　每个非负整数都能写成四个整数平方之和，即对每个 $n\in\mathbb{Z}$、$n\geq 0$，存在 $a,b,c,d\in\mathbb{Z}$ 使

$$n=a^2+b^2+c^2+d^2.$$

这里「四个」允许有零项，也不要求各项不同。例如 $0=0^2+0^2+0^2+0^2$，$1=1^2+0^2+0^2+0^2$。若要求四个数全为正，1 就是反例；若允许 $n<0$，平方和非负也立即排除。因此定理的范围和表示的约定都不可省略。

思路　证明分三步。由 Euler 恒等式先把一般整数归约到素数。对奇素数 p，抽屉原理给出某个较小的正倍数 mp 的四平方表示。最后取这样的最小乘数 m，证明若 $m>1$，总能构造更小的正乘数，逼出 $m=1$。本节完成前两步，下一节完成递降。

第一步：归约到奇素数　0 和 1 的表示已经写出，而 $2=1^2+1^2+0^2+0^2$。若每个奇素数都有四平方表示，则每个正整数 $n>1$ 都可分解为有限个素数的乘积。每个素因子的表示相乘后，由命题 22.6 仍得到四平方表示；素因子有重复也不影响。因而只需证明任意奇素数 p 都能表示。证明的剩余两步见引理 22.8 和下一节的递降。　■

只由模 p 的解不能直接得到等于 p 的表示。例如模 7 有 $2^2+3^2+1\equiv 0$，但整数和是 14，不是 7。先接受得到一个倍数，再设法把多出的乘数消去，是本证明的安排。

引理 22.8（较小倍数；原讲义引理 81）　对任意奇素数 p，存在正整数 m 满足 $1\leq m<p$，使得 mp 是四个整数平方之和。

直观理解　模 p 的平方剩余（包括 0）占 $(p+1)/2$ 个同余类。把另一组平方先取负、再减 1，仍占同样多的位置。两组共超过 p 个位置，必有重叠。重叠给出一个三平方和被 p 整除；选择接近零的代表元，把这个平方和的大小控制在 p^2 以下。

证明　先找模 p 的解。考虑 $\mathbb{Z}/p\mathbb{Z}$ 中的两个集合

$$S=\{\,x^2 \bmod p \mid 0\le x\le (p-1)/2\,\},\qquad T=\{\,-1-y^2 \bmod p \mid 0\le y\le (p-1)/2\,\}.$$

在范围 $0\le x,z\le (p-1)/2$ 内，若 $x^2\equiv z^2 \pmod p$，则 $p\mid (x-z)(x+z)$。素性给出 $x\equiv z$ 或 $x\equiv -z$。第一种情形只能是 $x=z$；第二种情形由 $0\le x+z\le p-1$ 得 $x=z=0$。所以 S 有 $(p+1)/2$ 个元素。平移和取负不改变基数，T 也一样大。

若 S,T 不相交，它们的并集就有 $p+1$ 个元素，超过总共 p 个同余类，矛盾。因此存在 x,y 使

$$x^2\equiv -1-y^2 \pmod p,\qquad x^2+y^2+1\equiv 0 \pmod p.$$

再控制整数和的大小。对 x,y 都取绝对值不超过 $(p-1)/2$ 的整数代表元。于是

$$0<x^2+y^2+1\le 2\left(\frac{p-1}{2}\right)^2+1=\frac{p^2-2p+3}{2}<p^2.$$

最后一个严格不等式对 $p\ge 3$ 成立，因为 $p^2+2p-3>0$。左边被 p 整除，写成 mp 后便有 $0<m<p$；再补上 0^2 就是四平方表示。 ■

素数条件在集合计数中有实际作用：模合数时，$x^2\equiv z^2$ 可能没有 $x\equiv \pm z$。例如模 8，$1^2\equiv 3^2$，但 $3\not\equiv \pm 1$。这并不妨碍四平方和定理对合数成立，只是说明刚才为素数准备的计数证明不能原样搬过去。

■ **例 22.7　抽屉原理给出的是哪个倍数？**　分别对 $p=7$ 和 $p=13$，找出引理中的 x,y,m。

解　先在允许的代表元范围内试平方，再检查总和。

(1) $p=7$：取 $x=2,y=3$，二者绝对值均不超过 3。有 $x^2+y^2+1=4+9+1=14=2\cdot 7$，故 $m=2<7$。

(2) $p=13$：取 $x=0,y=5$，绝对值均不超过 6。有 $x^2+y^2+1=0+25+1=26=2\cdot 13$，仍得 $m=2<13$。

这两个构造的乘数都恰好为偶数，下一节可以立即把它减半。引理不保证总能挑到 $m=1$，也不要求我们找出的 m 已经最小。 ♦

想一想　为什么一定要取接近零的代表元？把模 7 的 $2,3$ 改成 $9,10$ 不也满足同余吗？

答　同余仍成立，但 $9^2+10^2+1=182=26\cdot 7$，乘数 26 已不小于 7。选小代表元的作用是同时控制大小；纯粹的同余关系只保证「是一个倍数」，不保证「是较小的倍数」。

22.5　最小乘数的无穷递降

从 $14 = 3^2 + 1^2 + 2^2 + 0^2$ 出发，可以直接得到

$$7 = \left(\frac{3+1}{2}\right)^2 + \left(\frac{3-1}{2}\right)^2 + \left(\frac{2+0}{2}\right)^2 + \left(\frac{2-0}{2}\right)^2 = 2^2 + 1^2 + 1^2 + 1^2.$$

这里先将同奇偶的坐标配成一对，再取和与差的一半，使范数恰好减半。若乘数是奇数，不能直接除以 2；这时需要把坐标先模某个奇素数缩小，再用四元数乘法恢复整数性。

四平方和定理的剩余证明：最小乘数必为一　固定奇素数 p。由引理 22.8，集合

$$\mathcal{M} = \{\, m \in \mathbb{Z} \mid m \geq 1,\ mp \text{ 可写成四个整数平方之和} \,\}$$

非空。取它的最小元 m，则 $m < p$。选择一个表示

$$mp = x^2 + y^2 + z^2 + w^2.$$

若 $m = 1$，p 的表示已经得到。以下假设 $m > 1$，分别处理偶数和奇数，证明都与最小性矛盾。

情形一：m 为偶数。此时 mp 为偶数。平方和模 2 等于奇坐标的个数模 2，故四个坐标中奇数的个数只能是 $0, 2, 4$。无论哪种情况，都能重新排列，使 x, y 同奇偶，z, w 同奇偶。于是

$$\frac{x+y}{2},\quad \frac{x-y}{2},\quad \frac{z+w}{2},\quad \frac{z-w}{2}$$

都是整数，并且

$$\left(\frac{x+y}{2}\right)^2 + \left(\frac{x-y}{2}\right)^2 + \left(\frac{z+w}{2}\right)^2 + \left(\frac{z-w}{2}\right)^2 = \frac{x^2+y^2+z^2+w^2}{2} = \frac{m}{2}\,p.$$

因此 $m/2 \in \mathcal{M}$，又 $0 < m/2 < m$，与最小性矛盾。

情形二：m 为奇数。取任意素因数 $\ell \mid m$，写 $m = g\ell$。由于 m 为奇数，ℓ 为奇素数；又 $\ell \leq m < p$，所以 $\ell \neq p$。把四个坐标分别模 ℓ 约到中心代表元 x_0, y_0, z_0, w_0，满足

$$x_0 \equiv x,\ y_0 \equiv y,\ z_0 \equiv z,\ w_0 \equiv w \pmod{\ell}, \qquad |x_0|, |y_0|, |z_0|, |w_0| \leq \frac{\ell - 1}{2}.$$

接下来必须区分四个余数是否全为零。*分支 A：四个余数全为零。*这时 ℓ 整除 x, y, z, w，所以 $\ell^2 \mid mp$。由于 $\ell \neq p$ 且二者为素数，有 $\gcd(\ell^2, p) = 1$，从而 $\ell^2 \mid m$。因此

$$\left(\frac{x}{\ell}\right)^2 + \left(\frac{y}{\ell}\right)^2 + \left(\frac{z}{\ell}\right)^2 + \left(\frac{w}{\ell}\right)^2 = \frac{m}{\ell^2}\,p$$

给出一个正整数乘数 $m/\ell^2 < m$，与最小性矛盾。这里不能只写「把等式除以 ℓ^2」：还要说明新乘数确实为整数，$\ell \neq p$ 正是在此使用的。

*分支 B：四个余数不全为零。*记

$$\rho = x + yi + zj + wk, \qquad \sigma = x_0 + y_0 i + z_0 j + w_0 k.$$

由于平方和同余，有 $N(\sigma) \equiv N(\rho) = mp \equiv 0 \pmod{\ell}$。另一方面，不全为零保证 $N(\sigma) > 0$，小代表元保证

$$N(\sigma) \leq 4\left(\frac{\ell - 1}{2}\right)^2 = (\ell - 1)^2 < \ell^2.$$

所以存在整数 h，满足 $N(\sigma)=h\ell$ 且 $0<h<\ell$。

现在构造新的整数四元数。所谓 $\rho\equiv\sigma\pmod{\ell}$，指四个坐标分别同余。乘法的每个坐标都是整系数的双线性式，因此同余保持乘法，有

$$\rho\bar{\sigma}\equiv\sigma\bar{\sigma}=N(\sigma)\equiv 0\pmod{\ell}.$$

最右边的 0 是四个坐标全为零的四元数。于是 $\rho\bar{\sigma}$ 的四个坐标全被 ℓ 整除，

$$\beta=\frac{\rho\bar{\sigma}}{\ell}$$

确实具有整数系数。由命题 22.6 和实数倍的范数公式，

$$N(\beta)=\frac{N(\rho)N(\bar{\sigma})}{\ell^2}=\frac{(g\ell p)(h\ell)}{\ell^2}=ghp.$$

共轭只改变坐标符号，所以 $N(\bar{\sigma})=N(\sigma)$。因为 $g>0$ 且 $0<h<\ell$，有 $0<gh<g\ell=m$。四元数 β 的坐标给出 $gh\in\mathcal{M}$，再次与最小性矛盾。

两种情形都排除 $m>1$，故最小元必为 1。于是每个奇素数 p 都是四平方和；结合上一节的素因数乘法归约，四平方和定理对所有非负整数成立。 ■

图 22.2 将递降的三个出口放在一起。三个出口都要给出正整数乘数并证明它严格变小，少检查其中任意一项都不能与最小性矛盾。

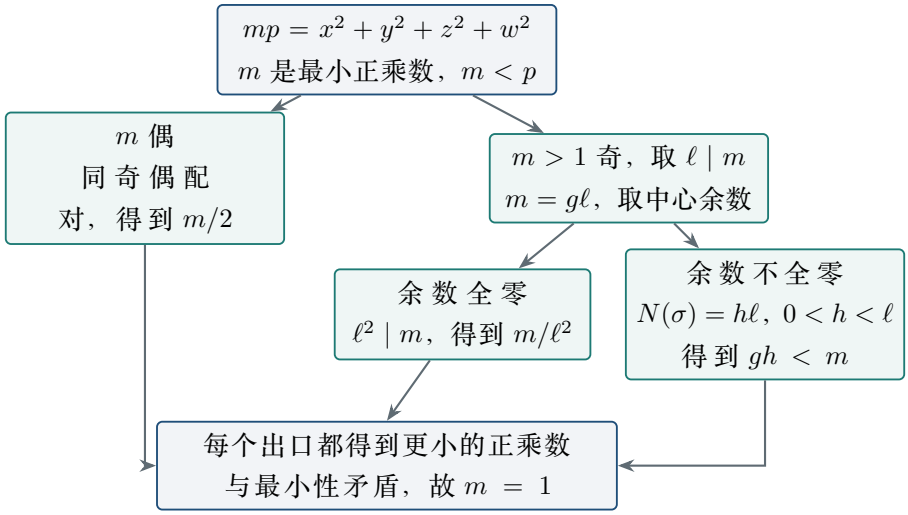


图 22.2　最小乘数的三个递降出口分别为 $m/2$、m/ℓ^2 与 gh，它们都严格小于 m。

这张图也说明：证明只需取 m 的一个奇素因数，完全不要求 m 本身为素数。新乘数 gh 也可能是合数，只要它是一个更小的正整数就足够了。

易错提醒　若直接模偶数 m 取最小绝对值代表元，只能保证 $|r_i|\le m/2$，进而得到 $\sum r_i^2\le m^2$；仅凭这个估计不能写成严格小于。例如 $m=2$、四个余数全为 1 时正好取等号。本证明先用配对处理偶乘数，再模奇素因数 ℓ 取余，使每个 $|r_i|\le(\ell-1)/2$，从根本上保证严格界。

■ **例 22.8　亲手完成奇数乘数的递降**　从 $21=4^2+2^2+1^2+0^2$ 出发，将 $p=7$ 的乘数 $m=3$ 降为 1。

解　取 $\ell=3$、$g=1$。四个中心余数为 $(1,-1,1,0)$，故令

$$\rho=4+2i+j,\qquad \sigma=1-i+j,\qquad N(\sigma)=3=1\cdot\ell.$$

所以 $h=1<\ell$。直接乘出

$$\rho\bar{\sigma}=(4+2i+j)(1+i-j)=3+6i-3j-3k.$$

每个坐标确实被 3 整除，得到 $\beta=1+2i-j-k$。于是

$$N(\beta)=1^2+2^2+(-1)^2+(-1)^2=7=ghp.$$

原乘数 3 降为 $gh=1$。这个例子把两个不同的整除检查分开了：$N(\sigma)$ 被 ℓ 整除用来定义 h；乘积的每个坐标被 ℓ 整除，才允许构造整数系数的 β。 ♦

■ **例 22.9**　**余数全零时不定义正数 h**　从 $117=9^2+6^2+0^2+0^2=9\cdot13$ 出发，用 $\ell=3$ 做递降。

解　四个坐标模 3 的余数全为零，因此 $\sigma=0$、$N(\sigma)=0$。这时不能使用要求 $h>0$ 的分支，直接将四个坐标都除以 3，得到

$$13=3^2+2^2+0^2+0^2.$$

新的乘数是 $m/\ell^2=9/9=1$。题中 $m=9<p=13$ 且 $\ell\neq p$，和证明的条件一致。若错误地把 $N(\sigma)=h\ell$ 的 $h=0$ 带入另一分支，就只会得到零范数，失去「正乘数」这个关键条件。 ♦

22.6　把存在性证明转成可核对的计算

对 $p=19$，试 $x=1,y=6$ 得到 $1^2+6^2+1^2=38$。这已足够完成一条具体的构造：先取一个倍数，再把乘数从 2 降到 1。把证明里的存在性步骤逐项做出来，最终就有整数坐标可回代验证。

■ **例 22.10**　**从头构造 19 的四平方表示**　使用本讲的引理和递降方法，给出 19 的四平方表示，并说明每一步的依据。

解　*第一步：给出较小倍数。*模 19 有 $1^2+6^2+1=38\equiv0$，且 $|1|,|6|\le9$，所以这是引理允许的代表元。于是

$$2\cdot19=1^2+6^2+1^2+0^2,$$

乘数 $m=2$ 满足 $1<m<19$。

*第二步：按同奇偶配对。*两个奇坐标 $1,1$ 配成一对，两个偶坐标 $6,0$ 配成一对。取和、差的一半，分别得到

$$\frac{1+1}{2}=1,\quad \frac{1-1}{2}=0,\quad \frac{6+0}{2}=3,\quad \frac{6-0}{2}=3.$$

所以 $19=1^2+0^2+3^2+3^2$，直接核对为 $1+0+9+9=19$。新乘数已是 1，构造到此停止。 ♦

一般奇素数也可如此构造：有限搜索 x,y 后，按当前乘数的奇偶与余数是否全零选择分支，直到 $m=1$。实际计算不必预知最小乘数；每步严格下降便保证有限步终止。记录每行的 m 和四坐标，逐次核对平方和等于 mp，即可检查整个过程。

两种下降，分别检查什么　Pell 全部解证明中，把实数 A 除以 ε^n，要证明所得两系数是正整数，且对应实数小于基本解。四平方和证明中，把整数乘数 m 降小，要证明新四坐标为整数、新乘数为正且小于 m。范数乘法负责保留方程；大小估计与整除性检查负责完成递降。

本讲小结

- **Pell 的全部正解**：正非平方 d 的基本解对应 $\varepsilon = x_1 + y_1\sqrt{d}$，全部正解为 $x_n + y_n\sqrt{d} = \varepsilon^n$，$n \geq 1$。
- **一般右端项**：$M \neq 0$ 时，任一种子 β 乘 ε^n 给出无穷多个不同解；$M = 0$ 只有零解。一般不保证单个种子覆盖全部解。
- **四元数的共轭与范数**：$\overline{qr} = \bar{r}\bar{q}$，$N(q) = a^2 + b^2 + c^2 + d^2$，$N(qr) = N(q)N(r)$，且 $N(q/\ell) = N(q)/\ell^2$。
- **素数归约**：Euler 恒等式使四平方和对乘法封闭；只需处理 $0, 1, 2$ 和奇素数。
- **较小倍数**：抽屉原理给出 $x^2 + y^2 + 1 \equiv 0 \pmod{p}$，中心代表元保证 $mp = x^2 + y^2 + 1^2 + 0^2$ 且 $0 < m < p$。
- **严格递降**：偶乘数降到 $m/2$；奇素因数 $\ell \mid m$ 下，全零余数降到 m/ℓ^2，非零余数用 $\rho\bar{\sigma}/\ell$ 降到 $gh < m = g\ell$。故最小乘数为 1，每个非负整数都是四平方和。

练习

22.1 ★　判断真假，假的给出反例：(a) 任一 Pell 正解的幂都能生成全部正解；(b) 若 $\ell \mid N(q)$，则 q 的每个整数坐标都被 ℓ 整除；(c) $N(q/\ell) = N(q)/\ell$。

22.2 ★　已知 $x^2 - 7y^2 = 1$ 的基本解为 $(8, 3)$，求前三组正解，并求其中 $y \leq 100$ 的全部正解。

22.3 ★　对 $x^2 - 3y^2 = 1$ 写出由 $(2, 1)$ 生成解的坐标递推。令 $\varepsilon = 2 + \sqrt{3}$，证明 $x_n = (\varepsilon^n + \varepsilon^{-n})/2$、$y_n = (\varepsilon^n - \varepsilon^{-n})/(2\sqrt{3})$。

22.4 ★★　用 $x^2 - 2y^2 = 1$ 的第二组正解，写出连续三整数均为二平方和的一组具体表示。说明为何同样的构造产生无穷多组。

22.5 ★★　本讲相邻平方和的构造给出 $n_0 = 0, n_1 = 3, n_2 = 20, \ldots$。证明 $n_{k+2} = 6n_{k+1} - n_k + 2$。（提示：先对 $\ell_k = 2n_k + 1$ 写齐次递推。）

22.6 ★　计算 $(i + j)^2$、$N(i + j)$ 和 $N((i + j)^2)$，说明混合项为何消失。

22.7 ★★　从 $5 = 2^2 + 1^2$ 和 $13 = 3^2 + 2^2$ 出发，用四元数乘法构造 65 的四平方表示。

22.8 ★★　对 $p = 11$ 找出引理 22.8 中一组 x, y, m；这个例子是否必须继续递降？

22.9 ★★★　从 $33 = 4^2 + 4^2 + 1^2 + 0^2 = 3 \cdot 11$ 出发，取 $\ell = 3$，完整写出中心余数、ρ、σ、h、$\rho\bar{\sigma}$ 和下一组四平方表示。

22.10 ★★★　设 a, b 为非负整数，证明 $4^a(8b + 7)$ 不能写成三个整数平方之和。由 7 的例子说明四平方和定理中的「四」不能统一改成「三」。

练习答案与提示

22.1 三项都假。(a) 对 $d=3$，$(7,4)$ 对应 $(2+\sqrt{3})^2$，其幂漏掉基本解 $(2,1)$；必须从基本解取幂才保证全部。(b) 取 $\ell=3$、$q=1+i+j$，则 $N(q)=3$ 被 3 整除，坐标 1 却不被 3 整除。(c) 取 $q=2$、$\ell=2$，则 $N(q/\ell)=1$，而 $N(q)/\ell=2$；正确分母为 ℓ^2。

22.2 递推为 $x_{n+1}=8x_n+21y_n$、$y_{n+1}=3x_n+8y_n$。从 $(8,3)$ 开始，得到 $(127,48)$ 和 $(2024,765)$。例如 $127^2-7\cdot 48^2=16129-16128=1$。全部解定理与 y_n 递增说明 $y\le 100$ 时只有 $(8,3)$、$(127,48)$。

22.3 乘 $2+\sqrt{3}$ 得 $x_{n+1}=2x_n+3y_n$、$y_{n+1}=x_n+2y_n$。由 $x_n+y_n\sqrt{3}=\varepsilon^n$，取共轭得 $x_n-y_n\sqrt{3}=\varepsilon^{-n}$，相加、相减并除以 2、$2\sqrt{3}$，分别得到题中两式。所除的 $\sqrt{3}$ 非零。

22.4 第二组解为 $(17,12)$，因此 $288=12^2+12^2$、$289=17^2+0^2$、$290=17^2+1^2$。一般地，每组正解 (x_n,y_n) 给出 $x_n^2-1=2y_n^2$、$x_n^2=x_n^2+0^2$、$x_n^2+1=x_n^2+1^2$。因为 x_n 严格递增，这些连续三元组彼此不同。

22.5 记 $u=3+2\sqrt{2}$，则 $u^{-1}=3-2\sqrt{2}$，故 $u+u^{-1}=6$，从而 $u^{k+2}=6u^{k+1}-u^k$。乘 $1+\sqrt{2}$ 并比较有理系数，得 $\ell_{k+2}=6\ell_{k+1}-\ell_k$。代入 $\ell_k=2n_k+1$ 后整理为 $n_{k+2}=6n_{k+1}-n_k+2$。核对 $k=0$：$20=6\cdot 3-0+2$。

22.6 $(i+j)^2=i^2+ij+ji+j^2=-1+k-k-1=-2$。所以 $N(i+j)=2$，$N((i+j)^2)=N(-2)=4=2^2$。混合项消失的依据是 $ij+ji=0$；若误用交换律把它写成 $2ij$，就会得到错误答案。

22.7 取 $q=2+i$、$r=3+2i$，乘得 $qr=6-2+(4+3)i=4+7i$。于是 $65=4^2+7^2+0^2+0^2$，回代为 $16+49=65$。这个例子恰好退化为复数的二平方恒等式。

22.8 取 $x=1,y=3$，均在 $[-5,5]$ 内。$x^2+y^2+1=1+9+1=11$，所以 $m=1$，已经得到 $11=1^2+3^2+1^2+0^2$，无须继续递降。

22.9 模 3 的中心余数为 $(1,1,1,0)$。令 $\rho=4+4i+j$、$\sigma=1+i+j$，则 $N(\sigma)=3$，$h=1$、$g=1$。乘法给出 $\rho\bar{\sigma}=9-3j-3k$，故 $\beta=3-j-k$。于是 $11=3^2+0^2+(-1)^2+(-1)^2$，新乘数是 $gh=1<3$。

22.10 平方模 4 为 0 或 1。若三个平方之和被 4 整除，则奇坐标个数在 $0,1,2,3$ 中只能取 0，所以三个坐标全偶。若 $4^a(8b+7)=x^2+y^2+z^2$，对 $a>0$ 可将三坐标都除以 2、等式除以 4；重复 a 次，得到 $8b+7=X^2+Y^2+Z^2$。平方模 8 为 $0,1,4$，三个这样的余数不可能相加为 7 模 8：零个 4 时总和至多 3；一个 4 时另外两项之和至多 2；两个 4 时总余数为 0 或 1；三个 4 时余 4。矛盾。特别地，7 不能用三个平方表示，却有 $7=2^2+1^2+1^2+1^2$，所以统一保证确实需要四项。

第 23 讲

勾股数组与费马无穷递降法

Pythagorean Triples and Fermat Descent

对应材料：MIT 18.781 Lecture 23 讲义；期末练习卷第 8 题；其余配套练习根据本讲证明编写。

本讲目标

(1) 推出本原勾股数组的全部参数，说明互素和奇偶条件各在哪里使用。

(2) 用过单位圆上固定有理点的直线，找出全部有理点并处理漏掉的基点。

(3) 用两次勾股参数化证明四次方之和不可能是非零平方。

(4) 区分同余矛盾、公共因子递降和乘积素因子矛盾三类论证。

23.1 从本原数组到两个互素平方

$3^2+4^2=5^2$、$5^2+12^2=13^2$ 给出了两组整数直角三角形。把第一组三边同时乘 2，得到 $6^2+8^2=10^2$。后者并没有增加一种新的形状。要找出所有整数直角三角形，合理的第一步是先除掉公共倍数，再研究剩下的互素情形。

第 22 讲用四平方范数的乘法构造整数表示；这一讲把目标换成一个更严格的二次关系。平方和恰好仍为平方，迫使三个数的公因数、奇偶性及因子分解彼此配合。后半讲还要在直角边本身为平方时，再用一次同样的分类。

定义 23.1（勾股数组与本原性） 满足 $x^2+y^2=z^2$ 的正整数三元组 (x,y,z) 称为**勾股数组**（Pythagorean triple）。若 $\gcd(x,y,z)=1$，则称为**本原勾股数组**（primitive Pythagorean triple）。这里直角边可以交换，斜边始终记为 z。

$(3,4,5)$ 是本原的，$(6,8,10)$ 的公因数为 2，因此不是本原的。$(0,1,1)$ 虽满足等式，却有一条边为零，本讲的定义把这类退化情形排除。讨论全体整数解时，再把它们及各坐标的符号补回即可；参数化正整数三角形时不应把退化情形混进去。

令 $d=\gcd(x,y,z)$。由于 d 整除三边，除以 d^2 得

$$(x/d)^2+(y/d)^2=(z/d)^2.$$

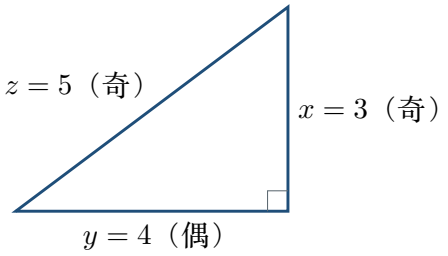
除后的三边仍为正整数，且没有大于 1 的公共因子。于是任意勾股数组都是本原数组的正整数倍。这里「本原」最初只排除了三者共有的因子，但方程还会迫使任意两者都互素。

引理 23.2（本原解的互素与奇偶性）　本原勾股数组的三边两两互素；两条直角边恰有一条为偶数，另一条与斜边均为奇数。

证明　若素数 p 同时整除 x, y，则 $p \mid z^2$，由素因数分解得 $p \mid z$，违背本原性。若 p 同时整除 x, z，由 $y^2 = z^2 - x^2$ 得 $p \mid y$，也矛盾。对 y, z 使用 $x^2 = z^2 - y^2$ 得到相同结论。每个大于 1 的公因数都有素因子，所以三对最大公因数均为 1。

两条直角边不能同偶，否则有公共因子 2；也不能同奇，因为平方模 4 只有 $0, 1$，两奇平方之和模 4 为 2，不可能等于 z^2。故两边异奇偶，平方和为奇数，z 也为奇数。 ■

交换直角边后，约定 x 为奇数、y 为偶数。图 23.1 用最小的一组数组说明这一约定：奇偶区分针对的是直角边的角色，不要求较长直角边一定为偶数。



$3^2 + 4^2 = 5^2$。
本原性排除三边的公共倍数；
模 4 的平方剩余区分两条直角边。

图 23.1　本原勾股数组的一条直角边为奇数，另一条为偶数，斜边为奇数。

图中的奇边短于偶边，但 $(15, 8, 17)$ 中奇边长于偶边。后面的公式统一用奇偶命名，避免在比较长短时反复交换符号。

引理 23.3（互素乘积为平方）　若 A, B 为互素正整数，且 AB 是完全平方数，则 A, B 各自都是完全平方数。

证明　固定一个整除 A 的素数 p。由于 $\gcd(A, B) = 1$，p 在 B 的素因数分解中不出现，因此它在 AB 中的指数等于在 A 中的指数。AB 为平方，所有素指数均为偶数，所以 A 的每个素指数为偶数，A 是平方。对整除 B 的素数作同一分析，得到 B 为平方。 ■

这个结论包含 $A = 1$ 或 $B = 1$，因为 $1 = 1^2$。互素条件不可省：$2 \cdot 8 = 16$ 为平方，两个因子都不是平方。正整数条件也保证平方根为整数且可选为正数；若只讨论整数乘积，则 $(-1)(-1) = 1$ 不能推出两个负因子是整数平方。

定理 23.4（本原勾股数组的完全分类）　正整数三元组 (x, y, z) 是本原勾股数组且 x 为奇数，当且仅当存在唯一的正整数 $r > s$，满足 $\gcd(r, s) = 1$、r, s 异奇偶，并且

$$x = r^2 - s^2, \qquad y = 2rs, \qquad z = r^2 + s^2.$$

直观理解　把 z^2-x^2 分解后，两个因子 $z-x, z+x$ 都是偶数，公共的 2 会妨碍直接使用互素平方引理。各除以 2 后，这个障碍消失；乘积仍是平方，因而可以把两因子分别写成平方。

思路　先从已有三角形导出参数，再反向验证参数确实给出本原三角形。正向证明先核查两个半和、半差互素；反向证明排除公共奇素因子与公共因子 2。

证明　由引理 23.2，x, z 都为奇数，y 为偶数，所以

$$\left(\frac{y}{2}\right)^2 = \frac{z+x}{2}\frac{z-x}{2}$$

是正整数的乘积。设正整数 h 同时整除这两个因子，那么 h 整除它们的和 z 及差 x；$\gcd(z,x)=1$，因此 $h=1$。由引理 23.3，存在正整数 r, s 使

$$\frac{z+x}{2} = r^2, \qquad \frac{z-x}{2} = s^2.$$

$x>0$ 给出 $r>s$，相加、相减给出 $z=r^2+s^2$、$x=r^2-s^2$，再由 $y>0$ 得 $y=2rs$。两个平方互素，故 $\gcd(r,s)=1$；x 为奇数，故 r^2, s^2 异奇偶，r, s 也异奇偶。这两个正平方根由三边唯一确定，所以参数唯一。

反过来，设 $r>s>0$、互素且异奇偶，则三边为正，直接展开有

$$(r^2-s^2)^2+(2rs)^2 = r^4+2r^2s^2+s^4 = (r^2+s^2)^2.$$

x, z 为奇数，三边不能有公共因子 2。若奇素数 p 同时整除三边，则 p 整除 $z+x=2r^2$、$z-x=2s^2$。由于 p 为奇数，可约去 2，再由素数整除平方得 $p \mid r, s$，与互素矛盾。因此数组本原。 ■

■ **例 23.1　斜边小于三十的全部本原数组**　求 $z<30$ 的本原勾股数组，仍按奇直角边、偶直角边、斜边的顺序排列。

解　分类定理把三边搜索变成了两个参数的搜索。$r^2+s^2<30$ 且 $1\le s<r$，所以 $2\le r\le 5$。逐行只保留互素、异奇偶的参数：

r	可用的 s	(x,y,z)
2	1	$(3,4,5)$
3	2	$(5,12,13)$
4	1, 3	$(15,8,17), (7,24,25)$
5	2	$(21,20,29)$

例如 $r=5, s=4$ 虽满足互素和奇偶条件，却给出 $z=41\ge 30$，所以不能保留。上表遍历了参数的全部可能，因而没有漏掉符合范围的本原数组。若要数不区分直角边顺序的三角形，共有五个；交换两条直角边只改变有序表示。 ♦

想一想　为什么公式要求 r, s 异奇偶？只要求它们互素是否已经足够？

答　不够。取互素的 $r=3, s=1$，公式给出 $(8,6,10)$，三个数均为偶数。因为两奇数的平方和、平方差均为偶数，$2rs$ 也是偶数。相反，r, s 异奇偶保证奇边与斜边为奇数，再配合互素条件才得到本原性。若不要求 $r>s$，还可能得到零边或负边。

■ **例 23.2　一条直角边固定为十二**　求含直角边 12 的全部正整数直角三角形，并指出其中的本原解。

解　这里不能预先把三角形当成本原，否则会丢掉一些解。设另一条直角边为 a，斜边为 c，则 $(c-a)(c+a)=144$。两个因子同奇偶，乘积为偶数，所以都为偶数。令 $c-a=2u$、$c+a=2v$，得 $uv=36$、$v>u>0$，并且 $a=v-u, c=v+u$。

因子对只有 $(u,v)=(1,36),(2,18),(3,12),(4,9)$；$(6,6)$ 给出 $a=0$，应排除。因此全部三角形为

$$(a,12,c)=(35,12,37),(16,12,20),(9,12,15),(5,12,13).$$

逐组检查最大公因数，只有第一组与最后一组本原。本原公式与整数因子对方法各有用途：前者适于分类本原解，后者能直接保留固定边问题中的公共倍数。 ♦

■ **例 23.3　与 Pell 方程连接：期末练习第八题**　找一个两条直角边相差 1、每条边都至少为 100 的整数直角三角形。

解　直角边相差 1 使它们互素，所以可用本原参数公式。所求条件为

$$|r^2-s^2-2rs|=1,\quad \text{即}\quad |(r-s)^2-2s^2|=1.$$

令 $A=r-s, B=s$，便得到第 21、22 讲研究过的 Pell 型方程 $A^2-2B^2=\pm1$。取其中一个解 $(A,B)=(7,5)$，有 $49-50=-1$，则 $r=A+B=12, s=B=5$。两参数互素且异奇偶，故本原公式给出

$$(r^2-s^2,2rs,r^2+s^2)=(144-25,120,144+25)=(119,120,169).$$

三边均至少为 100，且两直角边相差 1，满足要求。这道题只要求构造一组，因此找到并回代核查即可；若要求全部这样的三角形，才需要结合正、负 Pell 方程全部解的分类。 ♦

23.2　单位圆上的有理点与斜率参数

整数三角形 $(3,4,5)$ 除以斜边，得到单位圆上的点 $(3/5,4/5)$。另一组三角形 $(5,12,13)$ 给出 $(5/13,12/13)$。这两个分母不同的点，能否由同一种几何作图产生？固定圆上最容易写出的有理点 $P=(1,0)$，把直线的斜率当成参数，就能把这个问题化成一次求根。

> **命题　23.5（整数三角形与有理圆点）**　正整数勾股数组给出单位圆第一象限的有理点 $(x/z,y/z)$。反过来，若 $u,v\in\mathbb{Q}$、$u,v>0$ 且 $u^2+v^2=1$，则清除分母并约掉公因数后，得到唯一的本原勾股数组，其归一化点为 (u,v)。

证明　第一方向把 $x^2+y^2=z^2$ 除以正数 z^2 即可。反过来，选正公分母 d，使 $a=du, b=dv$ 都为正整数，则 $a^2+b^2=d^2$。约去三者最大公因数，得到本原数组。为证明唯一性，若两个本原整数向量沿同一条正有理射线，写其比例为既约分数 $h/k>0$，则 k 整除第一个向量的所有坐标，故 $k=1$；同样 h 整除第二个向量的所有坐标，故 $h=1$。 ■

这个命题仍依赖正坐标限制。单位圆上的 $(-3/5, 4/5)$ 也是有理点，却不是正边长三角形的归一化点；取坐标绝对值后才对应三角形。圆上的四个轴端点则对应零直角边，属于退化情形。几何参数化首先得到整个圆的有理点，再用坐标符号筛选所需的部分。

定理 23.6（单位圆的有理参数化）　除 $P=(1,0)$ 外，单位圆 $u^2+v^2=1$ 的有理点恰好为

$$(u,v)=\left(\frac{m^2-1}{m^2+1},\frac{-2m}{m^2+1}\right),\qquad m\in\mathbb{Q}.$$

每个这样的点对应唯一斜率 $m=v/(u-1)$；基点 P 需要另行补上。

思路　直线 $v=m(u-1)$ 已有一个交点 P，代入圆方程后二次式已有已知因子 $u-1$。除掉它就能求出另一个交点。最后反求斜率，确认任何其他有理点都被覆盖。

证明　对固定 $m\in\mathbb{Q}$，代入圆方程得到

$$u^2+m^2(u-1)^2-1=(u-1)\big((1+m^2)u+1-m^2\big)=0.$$

一个根为 $u=1$；另一根为 $(m^2-1)/(m^2+1)$。由于 m 是实有理数，分母 $m^2+1>0$，并且这一根不等于 1。代回直线方程得 $v=-2m/(m^2+1)$。两个坐标都为有理数，代回所得分解式便知它们位于圆上。

反过来，若 $(u,v)\neq(1,0)$ 位于单位圆，则 $u\neq 1$：因为 $u=1$ 会强制 $v=0$。所以 $m=v/(u-1)$ 有定义，若 u,v 有理，则 m 有理。点满足刚才的二次方程，又不是根 $u=1$，只能是公式给出的另一个根。直线上两点确定斜率，因此参数唯一。 ■

图 23.2 中取 $m=-2$。蓝圆上的点 $Q=(3/5,4/5)$ 与基点 P 连成橙色直线；图里延长线不承担任何新的代数条件，它只让斜率的符号和第二交点的位置更清楚。

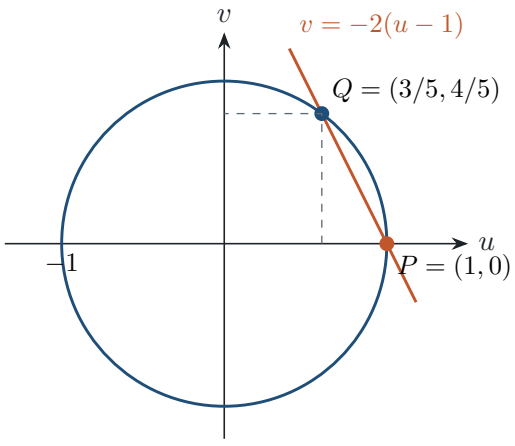


图 23.2　过基点 $(1,0)$ 的有理斜率直线与单位圆的另一个交点仍有有理坐标。

图中 Q 位于第一象限，所以 $u-1<0$、$v>0$，斜率为负。要让公式中的 $u>0$，还须 $m^2>1$；因此第一象限恰对应 $m<-1$。$m=-1$ 时得到 $(0,1)$，是象限边界；$m=0$ 时得到 $(-1,0)$。这些边界都由同一公式处理，只有基点 P 无法由有限斜率得到。

■ **例 23.4　从斜率恢复三角形**　分别取 $m=-2$ 和 $m=-3/2$，求圆上的有理点与对应的本原勾股数组。

解　代入时保留分数，便于看清约分过程。$m=-2$ 给出

$$(u,v)=\left(\frac{4-1}{4+1},\frac{4}{4+1}\right)=\left(\frac{3}{5},\frac{4}{5}\right),$$

清除分母得 $(3,4,5)$。$m=-3/2$ 给出

$$u=\frac{9/4-1}{9/4+1}=\frac{5}{13},\qquad v=\frac{3}{13/4}=\frac{12}{13},$$

所以得到 $(5,12,13)$。两次清除分母后的三边均互素，已经是本原数组。反求斜率：

$$\frac{4/5}{3/5-1}=-2,\qquad \frac{12/13}{5/13-1}=-\frac{3}{2},$$

与所选参数一致。 ♦

一般地，第一象限的斜率可唯一写成 $m=-r/s$，其中 $r>s>0$、$\gcd(r,s)=1$。代入公式得到

$$(u,v)=\left(\frac{r^2-s^2}{r^2+s^2},\frac{2rs}{r^2+s^2}\right).$$

这与定理 23.4 的公式完全相同。但此处 r,s 可能同奇，清除分母后还需要除以公共因子 2。若希望奇直角边总排在第一位，就必须结合前一节的奇偶条件重新排列边长。

■ **例 23.5　清分母后的公因子**　用斜率 $m=-3$ 求单位圆上的有理点。为什么它与代数分类中参数 $(r,s)=(2,1)$ 得到同一个无序直角三角形？

解　有理参数公式给出

$$(u,v)=(8/10,6/10)=(4/5,3/5).$$

直接用未约分的分母清分母得到 $(8,6,10)$，约掉 2 后为 $(4,3,5)$。交换两条直角边就得到 $(3,4,5)$，其奇边在先的唯一参数为 $(2,1)$。

这里并没有违背「有理点对应唯一斜率」：点 $(4/5,3/5)$ 的斜率为 -3，点 $(3/5,4/5)$ 的斜率为 -2，它们是圆上不同的点。无序三角形允许交换直角边，圆点的两个坐标则有固定顺序。分清这两个约定，才能正确理解各自的唯一性。 ♦

易错提醒　「有理斜率给出全部有理点」需要补上基点 $(1,0)$。例如任何有限 m 都满足 $(m^2-1)/(m^2+1)<1$，所以公式不可能给出基点。也不能把 $m=\infty$ 当成有理数代入算式；若使用竖直线，只能另行说明它在 P 处与圆相切。

想一想　若圆上任取一个实数点，都能反求一个实斜率。为何这一办法能精确区分有理点？

答　非基点处，$m=v/(u-1)$ 是坐标的有理运算；所以有理坐标必给有理斜率。反向公式也是 m 的有理运算，且分母不为零；所以有理斜率必给有理坐标。两方向合在一起才是有理性的等价。只画一条实直线并不保证第二个交点有理，必须检查所选参数。

前后联系　单位圆的计算依赖一个已知的有理点以及一个二次方程的已知根。第 24 讲将把这一步推广到适当的圆锥曲线，并讨论何时存在可用的有理基点。若曲线 $u^2+v^2+1=0$ 连实点都没有，当然无从选择有理基点。

23.3　四次方之和与两次参数化

试把勾股公式的两条直角边同时要求为平方。例如 $x=1, y=2$ 时，$x^4+y^4=17$，夹在 $4^2=16$ 和 $5^2=25$ 之间；更大的 x,y 是否可能恰好给出平方？有限枚举只能排除有限范围。本节要证明所有正整数范围内都不可能。

定义 23.7（无穷递降法）　无穷递降法（infinite descent）是利用正整数良序性证明不存在性的方法：假设某类整数解存在，为每个解指定一个正整数大小，并从任意一个解构造同类的、更小的解。取大小最小的解便得到矛盾。

这里「更小」必须是明确的整数不等式。参数的个数减少、公式看起来简短或实数大小趋于零，都不能替代这一步。例如正有理数 $1, 1/2, 1/4, \ldots$ 可以无限严格下降，因为这个集合没有最小元。我们将把 $x^4+y^4=z^2$ 的大小选为正整数 z，并明确得到新的斜边 $z'<z$。

定理 23.8（四次方之和不是正平方）　方程 $x^4+y^4=z^2$ 没有 $x,y,z\in\mathbb{Z}_{>0}$ 的解。等价地，不存在两条直角边同时为正整数平方的整数直角三角形。

直观理解　若 (x^2,y^2,z) 是勾股数组，第一次参数化会让一个乘积 $2rs$ 成为平方；奇偶和互素性迫使 r 与 $s/2$ 分别为平方。与此同时 $x^2+s^2=r^2$ 又是勾股关系，第二次参数化会把新的两个参数也变成平方。这样原来的方程便在更小的参数中重新出现。

思路　反设有解，取 z 最小的一组。先证明该组互素，再交换 x,y 使 x 奇、y 偶。第一次参数化得到 (x^2,y^2,z) 的参数 r,s；第二次参数化处理 (x,s,r)。每次拆平方乘积都先核查互素，最后验证所得新变量正且新斜边严格小于 z。

证明　**第一步：取最小解并约去公因子。**假设正整数解集合非空，取其中 z 最小的一组 (x,y,z)。令 $d=\gcd(x,y)$。因为 $d^4\mid x^4+y^4=z^2$，逐个比较素因子的指数，得到 $d^2\mid z$。若 $d>1$，则

$$\left(\frac{x}{d}\right)^4+\left(\frac{y}{d}\right)^4=\left(\frac{z}{d^2}\right)^2$$

给出更小的正整数解，矛盾。因此 $\gcd(x,y)=1$，勾股数组 (x^2,y^2,z) 本原。其两直角边异奇偶，交换 x,y 后可设 x 奇、y 偶。

第二步：第一次勾股参数化及偶参数的确定。由定理 23.4，存在 $r>s>0$、$\gcd(r,s)=1$、r,s 异奇偶，使

$$x^2=r^2-s^2,\qquad y^2=2rs,\qquad z=r^2+s^2.$$

还需确定谁为偶数。若 r 偶、s 奇，则 $x^2=r^2-s^2\equiv -1\equiv 3 \pmod 4$，不可能。因此 r 奇、s

偶。设 $s=2w$，则

$$\left(\frac{y}{2}\right)^2 = rw.$$

若整数整除 r,w，它也整除 r,s，所以 $\gcd(r,w)=1$。由互素平方乘积引理，存在正整数 a,b 使

$$r=a^2, \qquad w=b^2, \qquad s=2b^2.$$

这里 $s/2$ 为平方。把它误写成 s 为平方，会使下一次参数化的乘积条件发生变化。

第三步：第二次勾股参数化。由 $x^2=r^2-s^2$ 得 $x^2+s^2=r^2$。这是正整数勾股数组，而且本原：若某素数同时整除 x,s，则由 $r^2=x^2+s^2$ 得它也整除 r，与 $\gcd(r,s)=1$ 矛盾。此时奇直角边为 x，偶直角边为 s，所以存在互素、异奇偶的正整数 $t>u$ 使

$$x=t^2-u^2, \qquad s=2tu, \qquad r=t^2+u^2.$$

比较 $s=2b^2=2tu$，得到 $tu=b^2$。由于 $\gcd(t,u)=1$，再用一次引理，得正整数 k,ℓ 使 $t=k^2, u=\ell^2$。于是

$$a^2=r=t^2+u^2=k^4+\ell^4.$$

这正是原方程的另一组正整数解 (k,ℓ,a)。

第四步：核查严格下降。 $a\geq 1$ 且 $s>0$，所以

$$0<a\leq a^2=r\leq r^2<r^2+s^2=z.$$

新解的平方根变量为 a，严格小于最初选取的最小 z，矛盾。故正整数解不存在。 ■

图 23.3 记录了证明中真正传递的信息。每个向下的箭头都使用了一个已核查的整数条件；最底部回到与最上方同形的方程，但正整数大小严格减小。

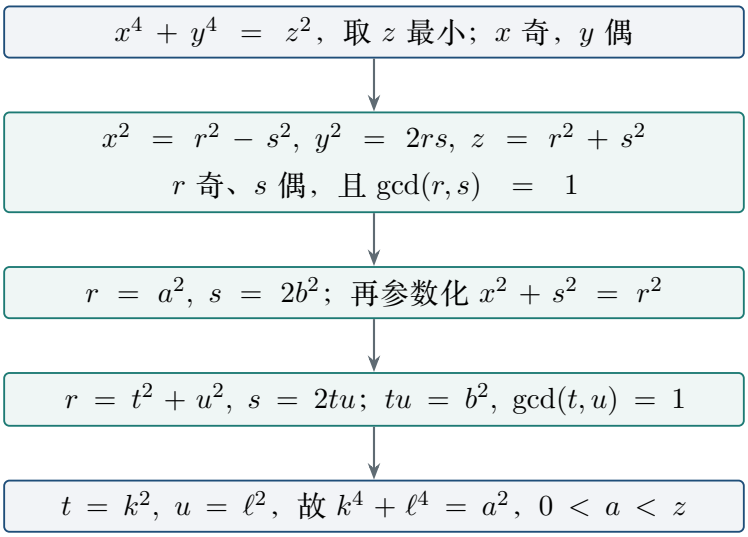


图 23.3　两次勾股参数化与两次互素平方分解产生同类解，并使斜边从 z 严格降到 a。

图中的 $r=a^2$ 出现在第一次拆乘积时，而 $t=k^2, u=\ell^2$ 出现在第二次拆乘积时。把这两次平方性连接起来，才得到新方程；仅仅知道两次勾股公式，并不足以证明递降。

易错提醒　不能由 $y^2 = 2rs$、$\gcd(r,s) = 1$ 直接断言 r, s 都是平方。反例是 $r = 1, s = 2$：$2rs = 4$ 为平方，s 却不是平方。证明中先确认 r 奇、s 偶，再写成 $(y/2)^2 = r(s/2)$，此时两个因子才互素且分别为平方。

■ **例 23.6**　**费马方程指数为四的情形**　证明 $X^4 + Y^4 = Z^4$ 没有非零整数解，并推广到所有正整数指数 n 满足 $4 \mid n$ 的方程 $X^n + Y^n = Z^n$。

解　若 X, Y, Z 均非零，取绝对值不会改变四次方，令 $x = |X|, y = |Y|, z = Z^2$，便得到 $x^4 + y^4 = z^2$ 的正整数解，与定理 23.8 矛盾。

一般地写 $n = 4h$，其中 $h \geq 1$。若原方程有非零整数解，令 $A = |X|^h, B = |Y|^h, C = |Z|^h$，因 n 为偶数，有 $A^4 + B^4 = C^4$，又回到刚排除的情形。因此所有被 4 整除的指数都已排除。这个归约没有证明奇素指数的费马方程；一般指数的结论需要另外的理论。 ♦

想一想　原定理为何要求 x, y 都严格为正？能否把结论写成「没有整数解」？

答　不能。$x = 0$ 时，任意整数 y 都给出 $y^4 = (y^2)^2$，所以 $(0, y, \pm y^2)$ 是解；$y = 0$ 时同样如此。非零的 x, y 可以取绝对值化成正数，z 也可取绝对值。因而全体整数解恰为 $x = 0, z = \pm y^2$ 或 $y = 0, z = \pm x^2$，两类在全零解处重合。

■ **例 23.7**　**只要求斜边为平方会怎样**　「勾股数组中不能出现平方边」这个说法是否成立？说明本节定理实际排除的条件。

解　取参数 $r = 4, s = 3$，得到 $(7, 24, 25)$。它是本原勾股数组，斜边 $25 = 5^2$ 为平方，所以题中说法错误。本节排除的是两条正直角边同时为平方，对斜边是否为平方没有单独的排除结论。

若把一条直角边也要求为零，则 $(0, 1, 1)$ 连直角边都各为平方，却是退化三角形。这说明「正」「两条」「直角边」三个限制都应在引用时保留。把定理压缩成一句模糊的「平方不可能」会丢掉真正使用的条件。 ♦

怎样检查一次递降　构造之后逐项核查：新变量是整数、仍满足原方程、仍在规定的正数范围内、所选正整数大小严格减小。只有这四项都成立，才能用良序原理结束证明。

23.4　同余障碍与三元立方递降

方程 $x^2 = 3 + 4y^3$ 看起来比勾股方程复杂，但模 4 后右边固定为 3，立刻排除全部整数解。另一方面，方程 $x^3 + 2y^3 + 4z^3 = 0$ 模 2 时仍有很多剩余类解，仅靠第一次取余无法结束。沿着偶性继续推下去，会迫使三个坐标都有公共因子 2，这就适合另一种更短的递降。

命题 23.9（同余不可解是整数不可解的充分条件）　设 F 是整系数多项式。若存在正整数 m，使 $F(x_1, \ldots, x_k) \equiv 0 \pmod m$ 没有剩余类解，则 $F(x_1, \ldots, x_k) = 0$ 没有整数解。

证明　整数解逐坐标约化模 m 后，多项式的加法和乘法均与约化相容，必给出一个同余解。因此

若同余解不存在，整数解就不可能存在。这是把「整数解推出同余解」取逆否命题。 ■

模数 $m=1$ 时所有整数均同余，没有排除能力；选择较大的模数也不一定更有效，关键是幂次的剩余类是否足够少。「找到了某个模数的解」只能说该模数没有立即给出矛盾，不能宣布原方程可解。下节会看到还需因式分解配合的例子。

■ **例 23.8　平方模四的障碍**　证明 $x^2=3+4y^3$ 在整数中无解，并解释为何不需要限制 y 正负。

解　先列出平方模 4 的可能值。x 偶时 $x^2\equiv 0 \pmod 4$；x 奇时 $x^2\equiv 1 \pmod 4$。右边对每个整数 y 都有 $3+4y^3\equiv 3 \pmod 4$，两边剩余类不相容。因此无整数解。

这里用的只是 $4\mid 4y^3$，对负整数、零、正整数都成立。用大小估计也许要把 y 分类，模 4 的论证则一次覆盖全体整数。与四次方之和的长证明相比，应先试这个成本很低的必要条件。 ♦

■ **例 23.9　立方模七的障碍**　证明方程 $x^3+y^3-7z^3=3$ 没有整数解。

解　先计算模 7 的立方表：

a	0	1	2	3	4	5	6
$a^3 \bmod 7$	0	1	1	6	1	6	6

因此立方剩余只有 $0,1,-1$。两个立方之和模 7 只能落在

$$\{-2,-1,0,1,2\} \pmod 7 = \{5,6,0,1,2\},$$

其中没有 3。原方程约化模 7 却要求 $x^3+y^3\equiv 3 \pmod 7$，矛盾。变量 z 被模数完全消掉，这是选 7 的直接动机。

也可以用第 4 讲的费马小定理解释这个立方表：若 $7\nmid a$，则 $(a^3)^2=a^6\equiv 1 \pmod 7$；在素数模下 $(a^3-1)(a^3+1)\equiv 0$，故 $a^3\equiv\pm 1$。若 $7\mid a$，立方为 0。这一证明同时说明为什么剩余类这么少。 ♦

定理 23.10（三元立方的公共因子递降）　方程 $x^3+2y^3+4z^3=0$ 的整数解只有 $(0,0,0)$。

思路　先用模 2 迫使 x 偶；代入并除以 2 后再迫使 y 偶；再代入并除以 2 后迫使 z 偶。因为方程齐次，同除以 2 得到同一个方程的新解。为容纳负坐标，大小选 $|x|+|y|+|z|$。

证明　若存在非零整数解，取 $H=|x|+|y|+|z|$ 最小的一组。非零保证 H 为正整数。原方程模 2 得 $x^3\equiv 0 \pmod 2$，因此 x 偶，写 $x=2x_1$。代回并除以 2：

$$4x_1^3+y^3+2z^3=0.$$

再次模 2，得到 y 偶，写 $y=2y_1$。代回该式并除以 2，得到

$$2x_1^3+4y_1^3+z^3=0.$$

所以 z 偶，写 $z=2z_1$。回到最初方程，整体除以 8 后有

$$x_1^3+2y_1^3+4z_1^3=0.$$

新三元组仍非零，因为同时乘 2 才恢复原三元组；其大小为 $H/2<H$，与 H 的最小性矛盾。最后代入确认全零组确实满足方程，故它是唯一解。 ■

这段证明也可以表述为「先取一个本原非零解，随后证明它的三坐标全偶」。因为方程是三次齐次的，任意非零解除以三个坐标最大公因数仍为整数解，故本原化合法。对于非齐次方程却要小心：若右边是 1，除以公因数会改变右端项，不能照抄这一操作。

同余与递降的分工　同余把一个整数解限制在某些剩余类；若这些限制迫使所有坐标含有同一因子，齐次性又允许整体约去该因子，就能把局部的整除信息转成严格下降。三次同余判断合成了一个关于全体整数解的结论。

23.5　平方加一的素因子与一个三次方程

考虑 $y^2=x^3+7$。它没有前面那样一眼可见的全局因式分解。若只枚举 $x=1,5,9,13,17$，右边依次为 $8,132,736,2204,4920$，都不是平方，但这仍是有限范围的观察。关键是先用模 4 限制 x,y，再把常数 7 改写成邻近的立方 8，从而出现两个整数因子。

引理 23.11（平方加一的奇素因子）　若奇素数 p 整除 a^2+1，则 $p\equiv 1 \pmod 4$。特别地，若 a 为偶数，则 a^2+1 的每个素因子都模 4 余 1。

证明　由 $p\mid a^2+1$ 得 $a^2\equiv -1 \pmod p$，所以 $p\nmid a$。若 $p\equiv 3 \pmod 4$，则 $(p-1)/2$ 为奇数，由费马小定理有

$$1\equiv a^{p-1}=(a^2)^{(p-1)/2}\equiv(-1)^{(p-1)/2}=-1 \pmod p,$$

这将迫使 $p\mid 2$，与 p 为奇素数矛盾。奇素数模 4 只可能为 1 或 3，因此只能为 1。若 a 偶，则 a^2+1 奇，不含素因子 2，故所有素因子均适用前半结论。 ■

a 偶的条件只用于排除素因子 2：当 $a=1$ 时，$a^2+1=2$，它的素因子就不是模 4 余 1 的素数。对「奇素因子」的结论，a 不必偶。另一方面，若正整数 $N\equiv 3 \pmod 4$，则其素因数分解中至少出现一个模 4 余 3 的素数；否则它是模 4 余 1 的数的乘积，也会余 1。实际还可推出至少一个此类素数出现奇数次。

定理 23.12（一个无解的三次丢番图方程）　方程 $y^2=x^3+7$ 没有整数解。

思路　先用平方非负得到 $x\ge -1$；模 4 后排除偶 x，并得到 y 偶、$x\equiv 1 \pmod 4$。再把方程改写成 $y^2+1=(x+2)(x^2-2x+4)$。左边的素因子类型受引理限制，右边的正因子 $x+2\equiv 3 \pmod 4$ 却必含相反类型的素因子。

证明　假设存在整数解。若 $x\le -2$，则 $x^3+7\le -8+7=-1<0$，不可能等于 y^2，因此 $x\ge -1$。

若 x 为偶数，则 x^3 被 8 整除，从而 $y^2 \equiv 7 \equiv 3 \pmod 4$，与平方剩余矛盾。因此 x 为奇数，x^3+7 为偶数，故 y 偶。再模 4，有

$$x^3 \equiv y^2 - 7 \equiv -7 \equiv 1 \pmod 4.$$

奇数模 4 余 1 或 3，它的立方与自身同余，因此 $x \equiv 1 \pmod 4$。结合 $x \geq -1$ 得 $x \geq 1$，于是 $x+2 \geq 3$ 且 $x+2 \equiv 3 \pmod 4$。

将原等式两边加 1，利用立方和分解：

$$y^2 + 1 = x^3 + 8 = (x+2)(x^2 - 2x + 4).$$

由于 $x+2$ 是正的且模 4 余 3，它有某个素因子 $p \equiv 3 \pmod 4$。于是 $p \mid y^2+1$；而 y 偶，引理 23.11 说明它的每个素因子均模 4 余 1。两结论矛盾，故不存在整数解。 ■

证明没有假设 x 一开始为正，也没有要求 y 为正。$y=0$ 以及所有负 y 都包含在上述推理中。$x \geq -1$ 这一不等式的作用，是保证随后取得的因子 $x+2$ 为正，从而能直接用它模 4 的余数判断素因子。若对负因子套用同样的话，符号 -1 本身也会改变模 4 的余数。

■ **例 23.10　只过第一层筛选仍不够**　原方程模 4 时剩下哪些 (x,y) 剩余类？取通过筛选的最小正 $x=1$，说明为什么它还不是整数解。

解　按证明可知 $x \equiv 1 \pmod 4$、y 偶，所以剩余类恰为 $(1,0)$、$(1,2)$。代回模 4 的原方程：两类左边均为 0，右边为 $1+7 \equiv 0$，确实都是同余解。

但 $x=1$ 时，整数方程要求 $y^2=8$，而 $2^2 < 8 < 3^2$，因此没有对应的整数 y。若再看模 16，平方剩余只有 $0,1,4,9$，也没有 8。这个例子说明一次成功的剩余类筛选只留下候选，不能保证候选能提升为真正的整数解。 ♦

■ **例 23.11　常数改变后结论会变**　对照方程 $y^2 = x^3 + 1$ 和 $y^2 = x^3 - 2$，各给出一组两个坐标都非零的整数解，并说明本节证明为何依赖常数 7。

解　直接计算 $3^2 = 2^3 + 1$、$5^2 = 3^3 - 2$，所以分别有 $(x,y) = (2,3)$ 与 $(3,5)$。常数 7 的特殊作用是加 1 后成为 $8 = 2^3$，既在左边形成平方加一，又在右边形成整系数的立方和分解。其他常数未必同时产生这两种结构。一个方程的无解证明不能仅凭式子的形状推广到所有 $y^2 = x^3 + D$。 ♦

本讲小结

- 本原勾股数组两两互素；两直角边异奇偶，斜边为奇数。固定奇边在先时，全部本原解为 $(r^2-s^2, 2rs, r^2+s^2)$，其中 $r>s>0$、互素且异奇偶。
- 单位圆除 $(1,0)$ 外的全部有理点由有理斜率参数 m 给出：$u=(m^2-1)/(m^2+1)$、$v=-2m/(m^2+1)$。第一象限对应 $m<-1$。
- 互素正整数乘积为平方时，各因子均为平方。对 $(y/2)^2 = r(s/2)$ 应先核查 s 偶与 $\gcd(r, s/2)=1$。
- 两次勾股参数化把假想解 $x^4+y^4=z^2$ 变成 $k^4+\ell^4=a^2$ 且 $0<a<z$，矛盾于正整数良序性。因此费马方程所有 $4 \mid n$ 的指数情形均无非零解。
- 同余不可解推出整数不可解；立方模 7 为 $0, \pm 1$。齐次方程 $x^3+2y^3+4z^3=0$ 迫使三坐标全偶，再用递降只剩零解。
- 奇素数 $p \mid a^2+1$ 必有 $p \equiv 1 \pmod 4$。在 $y^2=x^3+7$ 中，y 偶而正因子 $x+2 \equiv 3 \pmod 4$，与 $y^2+1=(x+2)(x^2-2x+4)$ 的素因子限制矛盾。

练习

23.1 ★ 判断并说明理由；假的给反例。(a) 两个正整数之积为平方，则两者都是平方；(b) 本原勾股数组的三边两两互素；(c) 互素参数 $r>s>0$ 总给出本原勾股数组；(d) 单位圆的每个有理点均由某个有限有理斜率 m 给出。

23.2 ★ 求斜边为 65 的全部本原勾股数组，按奇直角边在先排列；说明参数搜索为何完整。

23.3 ★★ 求一条直角边为 15 的全部正整数直角三角形，不要求本原。(提示：分解 $c^2-b^2=225$。)

23.4 ★ 对 $m=3/4$ 求单位圆参数公式给出的点，并由该点反求斜率。解释它为何没有对应第一象限的正三角形归一化点。

23.5 ★★ 写出 $x^4+y^4=z^2$ 的全部整数解，包括零坐标与负坐标，并说明为什么没有遗漏。

23.6 ★★ 证明任何整数直角三角形的面积是 6 的倍数。先证明本原情形，再考虑三边的公共倍数。

23.7 ★★ 证明两个奇数的四次方之和不可能是平方。解释为何这一模 16 的论证不能独自替代本讲的无穷递降证明。

23.8 ★★ 直接从模 8 分析证明：若 $x^3+2y^3+4z^3 \equiv 0 \pmod 8$，则 x, y, z 全偶。然后用齐次性重新证明它的整数零点只有原点。

23.9 ★★ 设 a 为偶整数。证明 a^2+1 的每个正因子都模 4 余 1。用这个结论改写 $y^2=x^3+7$ 无解证明的最后一步。

23.10 ★★★ 证明单位圆不存在满足 $u=v$ 的有理点。可把有理点清除分母，给出一个只用偶性和最小正分母的递降证明；比较实数情形。

练习答案与提示

23.1 (a) 假，例如 $2\cdot 8=16$，但两个因子都不是平方；互素条件不能丢。(b) 真：任一素数若同时整除两边，由方程也整除第三边，与本原性矛盾。(c) 假，$(r,s)=(3,1)$ 给出 $(8,6,10)$，还需参数异奇偶。(d) 假，有限斜率公式不包含基点 $(1,0)$，因为 $m^2-1<m^2+1$。

23.2 有 $r^2+s^2=65$、$r>s\geq 1$，所以 $\sqrt{65/2}<r<\sqrt{65}$，整数 r 仅为 $6,7,8$。相应 $s^2=29,16,1$，只有 $(r,s)=(7,4),(8,1)$ 可用，均互素、异奇偶。公式得到 $(33,56,65)$ 与 $(63,16,65)$。区间限制已枚举全部参数，所以没有别的本原数组；交换两条直角边得到另一种排列。

23.3 设另一边 $b>0$、斜边 $c>0$。由 $(c-b)(c+b)=225$，两个正因子均奇，令 $c-b=u, c+b=v$，有 $uv=225, v>u>0$。因子对为 $(1,225),(3,75),(5,45),(9,25)$，而 $(15,15)$ 给零边，应排除。故 $(b,c)=(112,113),(36,39),(20,25),(8,17)$，即三角形为 $(15,112,113),(15,36,39),(15,20,25),(15,8,17)$。每组由因子恒等式自动满足平方和；其中第一组和最后一组本原。

23.4 $m^2=9/16$，所以

$$u=\frac{9/16-1}{9/16+1}=-\frac{7}{25},\qquad v=\frac{-3/2}{25/16}=-\frac{24}{25}.$$

反求 $v/(u-1)=(-24/25)/(-32/25)=3/4$，与原斜率一致。两坐标都为负，点在第三象限；若取绝对值，则得到正三角形 $(7,24,25)$ 的归一化点，但这改变了圆点本身。

23.5 若 x,y 均非零，取绝对值后得到本讲定理排除的正整数解，所以至少有一个为零。$x=0$ 时 $z^2=y^4$，故 $z=\pm y^2$；$y=0$ 时 $z=\pm x^2$。因此全部解为

$$(0,t,\pm t^2)\quad\text{或}\quad(t,0,\pm t^2),\qquad t\in\mathbb{Z}.$$

全零解在两类中重复出现，没有额外影响；每一类直接代入都成立。

23.6 本原参数给出的面积为

$$\frac{xy}{2}=rs(r^2-s^2).$$

r,s 异奇偶，所以 rs 为偶数，面积被 2 整除。若 $3\mid r$ 或 $3\mid s$，面积被 3 整除；否则 $r^2\equiv s^2\equiv 1 \pmod 3$，于是 $3\mid r^2-s^2$，面积仍被 3 整除。由于 $2,3$ 互素，面积被 6 整除。一般数组是本原数组的 d 倍，面积乘 d^2，故结论保持。此处所得面积确实为整数，因为原三角形至少有一条偶直角边。

23.7 奇数平方模 8 为 1，也可写为模 16 的 1 或 9，再平方后均为 1。因此两个奇数的四次方之和模 16 为 2；平方模 16 却只能为 $0,1,4,9$，矛盾。这只排除了 x,y 同奇的情形。若一奇一偶，则四次方和模 16 为 1，没有立即矛盾；本原化后恰好会留下这一情形，所以仍需要递降。

23.8 先模 2，x 偶，故 $x^3\equiv 0 \pmod 8$。原同余化为 $2y^3+4z^3\equiv 0 \pmod 8$，即 $y^3+2z^3\equiv 0 \pmod 4$。模 2 得 y 偶，因此 $y^3\equiv 0 \pmod 4$，从而 $2z^3\equiv 0 \pmod 4$，z 偶。反过来三坐标全偶时三个立方都被 8 整除，同余也成立。若原整数方程有非零解，取 $H=|x|+|y|+|z|$ 最小的一组；全偶允许同除以 2，三次齐次性使方程仍成立，而 H 减半，矛盾。故只剩全零解。

23.9 a 偶使 a^2+1 为奇数，其每个素因子由引理 23.11 都模 4 余 1。任意正因子都是这些素数的非负整数次幂之积，故也模 4 余 1；正因子 1 对应空乘积，同样满足。对原三次方程，已有 y 偶及正整数 $x+2\equiv 3 \pmod 4$，但乘积分解说明 $x+2$ 是 y^2+1 的正因子，按刚证明的结论又应余 1，矛盾。这个写法把寻找特定素因子的最后一步合并成了正因子判别。

23.10 假设有理点满足 $u=v$，则 $2u^2=1$，所以 $u\neq 0$。取既约分数 $u=a/b$，其中 $b>0$，有 $2a^2=b^2$。于是 b 偶，写 $b=2c$，代回得 $a^2=2c^2$，故 a 也偶，与既约性矛盾。

也可不预先要求既约：在全部满足 $2a^2=b^2$ 的整数对中取正分母 b 最小的一组。同样推出 a,b 全偶，除以 2 后得到分母 $b/2$ 的同类表示，违背最小性。实数情形则有两个点 $(1/\sqrt{2},1/\sqrt{2})$ 与 $(-1/\sqrt{2},-1/\sqrt{2})$；不存在有理点，并不等于不存在实点。

第 24 讲

圆锥曲线上的有理点

Rational Points on Conics

对应材料：MIT 18.781 Lecture 24：斜率参数化、齐次化、三元二次型的化简与 Legendre 定理（原讲义定理 84、85）。练习按本讲构造改编。

本讲目标

(1) 会从一个已知有理点出发，用斜率参数化非退化圆锥曲线，并补查竖线和切线。

(2) 理解仿射方程与齐次三元方程的关系，能区分有理点、整数解与无穷远点。

(3) 会用有理可逆变换对角化三元二次型，将系数化成乘积无平方因子的形式。

(4) 掌握 Legendre 定理的必要条件与充分性证明，能用同余条件判断有理点是否存在。

(5) 能把「先判存在，再做参数化」落实为完整计算，并说明结果对整数点的含义。

24.1 从一个有理点出发

第 23 讲用过单位圆 $x^2+y^2=1$：过 $(1,0)$ 作有理斜率直线，就能得到勾股数组。现在换成 $x^2+y^2=5$。它已有整数点 $(1,2)$，能不能借它找出全部有理点？这个问题有两层：知道一个点后怎样找其他点；连第一个点都不知道时怎样判断是否存在。本讲先解决第一层，再用同余解决第二层。

*准备与阅读范围。*圆与双曲线的斜率例题主要使用二次方程和有理数运算；一般非退化条件、对角化及逆向还原还使用矩阵、行列式、秩与可逆线性变换。尚未学过这些线性代数知识时，可先读具体曲线的参数化例题，补学后再读一般情形的证明。

定义 24.1（圆锥曲线与非退化条件） 由有理系数二次方程

$$ax^2+by^2+cxy+dx+ey+f=0$$

定义的平面曲线称为**圆锥曲线**（conic）。将它齐次化后，对称矩阵为

$$A=\begin{pmatrix} a & c/2 & d/2 \\ c/2 & b & e/2 \\ d/2 & e/2 & f \end{pmatrix}.$$

本讲讨论 $\det A\neq 0$ 的**非退化**（nondegenerate）情形；等价地，射影曲线没有奇点。

矩阵条件的作用很具体：非退化二次曲线不能含一整条直线，否则二次多项式分解成两个一次因子，在交点出现奇点。圆、椭圆、双曲线、抛物线都可在这个范围内；两条相交直线 $xy=0$ 和重合直线 $x^2=0$ 不在范围内。$x^2+y^2+1=0$ 虽无实点，矩阵仍可逆，所以「非退化」没有保证点的存在。

■ **例 24.1　为圆 $x^2+y^2=5$ 写出参数式**　以 $P=(1,2)$ 为基点，求全部有理点，并检查斜率 $0,1,-1$ 的结果。

解　先固定经过 P 的非竖直线 $y=2+m(x-1)$，其中 $m\in\mathbb{Q}$。代入圆的方程并利用 $1^2+2^2=5$，可因式分解为

$$(x-1)\bigl[(1+m^2)x+1+4m-m^2\bigr]=0.$$

第一个根 $x=1$ 对应已知点；第二个根给出

$$\boxed{\,x=\frac{m^2-4m-1}{1+m^2},\qquad y=\frac{2-2m-2m^2}{1+m^2}.\,}$$

分母 $1+m^2$ 对有理 m 始终非零。将两个分子平方相加，得到 $5(1+m^2)^2$，所以每个参数确实产生圆上的有理点。三个斜率依次给出 $(-1,2)$、$(-2,-1)$、$(2,1)$，均可直接验算。

接着证明没有漏掉一般的点。若 (u,v) 是圆上有理点且 $u\neq 1$，则连接它与 P 的斜率 $m=(v-2)/(u-1)$ 属于 $\mathbb{Q}$，故 (u,v) 必由参数式得到。若 $u=1$，圆方程给 $v=\pm 2$。基点 $(1,2)$ 对应 $m=-1/2$，此时括号中的第二个根也等于 1，直线是切线。另一个点 $(1,-2)$ 来自竖直线 $x=1$，须单独补上。因此全部有理点恰是参数式的值，加上 $(1,-2)$。 ♦

例 24.1 中，除掉 $x-1$ 是求第二个交点的计算步骤，不能理解为从答案中永远删去基点。切线把第二个交点带回基点，竖线则补回有限斜率无法表示的点。图 24.1 同时标出了这两种需要检查的直线。

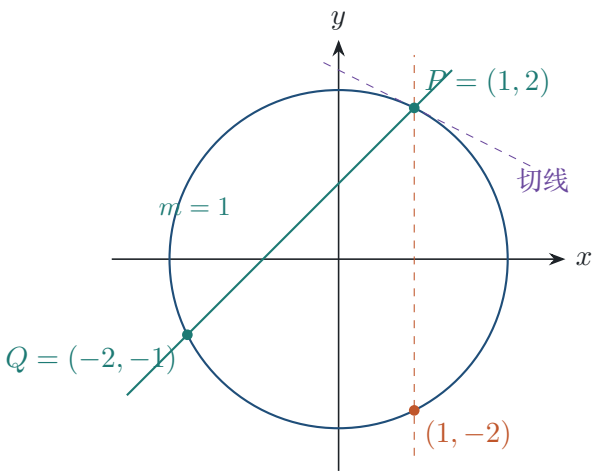


图 24.1　有理斜率直线给出第二个有理交点；切线返回基点，竖线需要单独补查。

图中圆上已画出的三个点都有整数坐标，但方法并未要求第二个点是整数。例如取 $m=3$，得到 $(-2/5,-11/5)$；验算为 $(4+121)/25=5$。参数化解决的是有理数问题，整数点还要额外检查分母能否约掉。

命题 24.2（一个点给出有理参数化）　有理系数的非退化圆锥曲线若有一个有理点 P，则经过 P 的有理直线给出其全部射影有理点的参数化。特别地，曲线有无穷多个仿射有理点。

直观理解　直线与二次曲线相交会得到一个二次方程。已知一个有理根后，另一个根由根与系数关系决定，也属于 $\mathbb{Q}$。让斜率变化，就让另一个交点沿曲线移动。

证明　选一个含 P 的有理仿射坐标图，并记曲线齐次化后的二次式为 H。过 P 的有理直线有无穷多条，例如有限斜率 $m \in \mathbb{Q}$ 给出无穷多条，另有竖直方向。取其中一条，将它写成 $[u:v] \mapsto [vP + uD]$，其中 P, D 是线性无关的有理三元向量。将 H 限制到这条直线，因 $H(P) = 0$，可写成

$$H(vP + uD) = u(Au + Bv), \qquad A, B \in \mathbb{Q}.$$

非退化保证这条直线不整条包含在曲线上，故 A, B 不会同时为零。已知交点对应 $u = 0$；另一个交点满足 $Au + Bv = 0$。若 $A \neq 0$，它是有理参数 $[u:v] = [-B:A]$；若 $A = 0$，则 $B \neq 0$，第二点为 $[u:v] = [1:0]$，也仍是有理点。若 $B = 0$，第二交点就是 P，这正是切线情形。若第二交点在原仿射图的无穷远直线上，上面的齐次计算仍适用。

反过来，任一不同于 P 的有理点都与 P 确定一条有理直线。两个不同方向若给出同一个不等于 P 的第二交点，连接这两点的直线便不唯一，矛盾；因此只有至多一个切线方向会把第二交点送回 P，其余无穷多个方向给出互不相同的点。无穷远直线与非退化二次曲线至多有两个交点，删去它们后仍有无穷多个仿射有理点。 ■

想一想　有一个有理点就有无穷多个有理点，能否把「有理」换成「整数」?

答　不能。圆 $x^2 + y^2 = 5$ 只有 $(\pm 1, \pm 2)$、$(\pm 2, \pm 1)$ 共八个整数点，因为整数 x, y 都满足绝对值不超过 2；但参数式给出无穷多个有理点。斜率可以取任意有理数，分母却不一定被消去。若只求整数点，还要结合有界性、整除或其他结构。

24.2　齐次化：把分母交给第三个变量

点 $(-2/5, -11/5)$ 可以写成整数三元组 $(-2, -11, 5)$。圆方程随之变成 $X^2 + Y^2 = 5Z^2$，回代确实有 $4 + 121 = 125$。第三个变量记录公分母，并使每一项都成为二次项。

定义 24.3（齐次方程与射影坐标）　对仿射方程 $F(x, y) = ax^2 + by^2 + cxy + dx + ey + f = 0$，其齐次化（homogenization）为

$$H(X, Y, Z) = aX^2 + bY^2 + cXY + dXZ + eYZ + fZ^2 = 0.$$

非零有理三元组按共同非零有理倍数视为同一点，记作 $[X:Y:Z]$。当 $Z \neq 0$ 时，它对应仿射点 $(X/Z, Y/Z)$；$Z = 0$ 的点称为无穷远点（point at infinity）。

例如 $[-2:-11:5] = [-4:-22:10]$，两者是同一个射影点。三元组 $(0,0,0)$ 不允许，因为它无法表示任何比值，也无法确定直线方向。「有非平凡解」始终指坐标不全为零，并不要求三个

坐标各自非零。

定理 24.4（齐次化保留有理点的存在性）　设 $F=0$ 是本讲意义下的非退化圆锥曲线，H 是其齐次化。下列三件事等价：

(1) $F=0$ 有仿射有理点；

(2) $H=0$ 有非零有理三元解；

(3) $H=0$ 有非零整数三元解。

思路　有理数与整数之间只需清分母。真正要补的一步是：一个齐次解可能满足 $Z=0$，不能直接除以 Z；须先借它构造一个 $Z\neq 0$ 的有理解。

证明　由仿射有理点 (r,s) 得齐次有理解 $(r,s,1)$；乘以 r,s 的公分母就得非零整数解。非零整数解当然也是有理解。

现设 $[X:Y:Z]$ 是齐次有理点。若 $Z\neq 0$，取 $(x,y)=(X/Z,Y/Z)$ 即可。若 $Z=0$，至少 X,Y 之一非零。若 $X=0$，则 $Y\neq 0$，在 $Y=1$ 的坐标图中作同样论证，得到某个 $w\neq 0$ 的点 $[u:1:w]$，它对应仿射点 $(u/w,1/w)$。以下写 $X\neq 0$ 的情形。除以 X^2，在 $X=1$ 的坐标图中考察

$$a+bv^2+cv+dw+evw+fw^2=0.$$

它已有有理点 $(v,w)=(Y/X,0)$，且仍是非退化圆锥曲线的一张坐标图。命题 24.2 的斜率法在这张图中产生无穷多个有理点。直线 $w=0$ 不是曲线的分量，故至多含两个交点；因此其中有点满足 $w\neq 0$。这个点给原齐次方程的解 $[1:v:w]$，再取 $(x,y)=(1/w,v/w)$，便得到原仿射曲线的有理点。 ■

这里用到非退化，正是为了保证「无穷多点」与「直线至多两个交点」。不能把任意退化二次方程也纳入定理。比如仿射方程 $x^2+1=0$ 无有理点，而其齐次方程 $X^2+Z^2=0$ 有 $[0:1:0]$；齐次曲线退化，这个点不能按上述方法产生仿射点。

■ **例 24.2　一个无穷远点怎样带回仿射点**　曲线 $x^2-y^2=1$ 的齐次方程有无穷远点 $[1:1:0]$。从它出发找仿射有理点。

解　在 $X=1$ 图中写 $v=Y/X, w=Z/X$，方程是 $1-v^2=w^2$，已知点为 $(v,w)=(1,0)$。取过它的直线 $w=t(v-1)$，代入并约去已知因子 $v-1$，得

$$v=\frac{t^2-1}{t^2+1},\qquad w=\frac{-2t}{t^2+1}.$$

只要 $t\in\mathbb{Q}$ 且 $t\neq 0$，便有 $w\neq 0$，于是

$$x=-\frac{t^2+1}{2t},\qquad y=-\frac{t^2-1}{2t}.$$

例如 $t=-1$ 给 $(1,0)$，$t=-2$ 给 $(5/4,3/4)$；后者满足 $25/16-9/16=1$。无穷远点提供了构造的起点，最终仍须选择 $Z\neq 0$ 的点才可回到原平面。 ♦

齐次整数解不等于仿射整数点　由 $X^2+Y^2=5Z^2$ 的整数解 $(-2,-11,5)$ 得到的仿射点是 $(-2/5,-11/5)$，并非整数点。齐次化把「仿射有理点存在」化成「非零整数三元解存在」，没有消除仿射坐标的分母。若 $Z=0$，连直接取比值这一步也不能做。

24.3　对角化与系数的规范化

对 $X^2+Y^2-5Z^2$，三个平方项已经分开。一般二次型还含 XY, XZ, YZ，直接讨论剩余条件会很笨重。先换有理坐标，把它化成 $aU^2+bV^2+cW^2$，才能让每个系数对应一组同余检验。

命题 24.5（有理对角化）　任一有理系数非退化三元二次型，经一个有理可逆线性变换，都可化为

$$aU^2+bV^2+cW^2, \qquad a,b,c\in\mathbb{Q}, \quad abc\neq 0.$$

变换前后存在非零有理解是等价的。

证明　若某个平方项系数非零，不妨写成

$$AX^2+DXY+EXZ+BY^2+FYZ+CZ^2, \qquad A\neq 0.$$

令 $U=X+\frac{D}{2A}Y+\frac{E}{2A}Z$，则二次型等于

$$AU^2+\left(B-\frac{D^2}{4A}\right)Y^2+\left(F-\frac{DE}{2A}\right)YZ+\left(C-\frac{E^2}{4A}\right)Z^2.$$

这个变换为三角形且对角元全为 1，可用有理运算逆回去。剩下二元型若有非零平方项，再配一次方；若只有非零交叉项，用 $V=Y+Z, W=Y-Z$，由 $YZ=(V^2-W^2)/4$ 得平方项。原三元型若起初全无平方项但有非零交叉项，也先对那两个变量作和差变换，就能制造一个非零平方项。这样有限步得到对角式。

可逆变换保持矩阵的秩，非退化的原矩阵秩为 3，所以对角式不可能有零系数。变换及其逆均为有理线性运算，非零向量也不会被送到零，因此非零有理解在两边一一对应。 ■

■ **例 24.3　没有平方项也可以开始配方**　把 $XY=Z^2$ 化成三个平方项的方程，并解释整数解是否原样对应。

解　令 $U=X+Y, V=X-Y, W=2Z$，则

$$XY-Z^2=\frac{U^2-V^2-W^2}{4}.$$

所以原式等价于 $U^2-V^2-W^2=0$。逆变换是 $X=(U+V)/2, Y=(U-V)/2, Z=W/2$，在 $\mathbb{Q}$ 上始终允许。例如 $(X,Y,Z)=(3,12,6)$ 得 $(U,V,W)=(15,-9,12)$，且 $225-81-144=0$。对角式的整数解 $(1,0,1)$ 却逆回 $(1/2,1/2,1/2)$，须整体乘 2 才得到整数三元组 $(1,1,1)$。可逆有理变换首先保证有理解的等价；它不必把每个整数向量原样变成整数向量。 ♦

得到对角式后，先乘公分母把 a,b,c 化成非零整数。若 $a=s^2a_0$，用 $U'=sU$ 就能从系数中删去平方因子；其他系数同理。再除去三系数公因数。仅做到每个系数无平方因子还不够，例如 $6,10,-15$ 各自无平方因子，乘积却含有平方因子。

命题 24.6（把系数化成两两互素）　通过有理可逆的变量缩放和方程的非零倍乘，可将非退化对角方程化成

$$aU^2+bV^2+cW^2=0,\quad a,b,c\in\mathbb{Z}\setminus\{0\},\quad |abc|\text{ 无平方因子.}$$

最后一个条件等价于：三个系数各自无平方因子，并且两两互素。

证明　先完成上段的平方因子删除与三系数公因数消去。若素数 p 同时整除 a,b，它不能再整除 c。写 $a=pa_1,b=pb_1$，令 $W=pW_1$，然后将全式除以 p，得到

$$a_1U^2+b_1V^2+pcW_1^2=0.$$

因为原系数各自无平方因子且 $p\nmid c$，新系数仍各自无平方因子；其乘积绝对值从 $p^2|a_1b_1c|$ 降到 $p|a_1b_1c|$。若还有公共素因子，重复此步。乘积为正整数且严格下降，故过程终止，终止时三系数两两互素。反向的变量替换 $W_1=W/p$ 在 $\mathbb{Q}$ 中合法，故全程保留非零有理解的存在性。 ■

■ **例 24.4　把 $6X^2+10Y^2-15Z^2=0$ 化到规范形**　三个系数各自无平方因子，怎样进一步使乘积无平方因子?

解　依次消去同时出现在两个系数里的素数。先让 $Z=2W$ 并除以 2，再让 $Y=3V$ 并除以 3，最后让 $X=5U$ 并除以 5：

$$\begin{aligned}6X^2+10Y^2-15Z^2=0 &\iff 3X^2+5Y^2-30W^2=0\\ &\iff X^2+15V^2-10W^2=0\\ &\iff 5U^2+3V^2-2W^2=0.\end{aligned}$$

最后乘积为 -30，绝对值无平方因子。三个缩放合起来为 $(X,Y,Z)=(5U,3V,2W)$，在 $\mathbb{Q}$ 上可逆。新方程的整数解 $(U,V,W)=(1,1,2)$ 给原方程解 $(5,3,4)$；验算 $150+90-240=0$。乘积绝对值由 900 依次降为 $450,150,30$，对应除以 $2,3,5$。 ♦

想一想　既然只要求存在非零有理解，为什么还要费力删去平方因子并消去公共素因子?
答　后面的剩余判别要从模一个素数的平方根，拼出模整个系数的平方根。系数无平方因子时，中国剩余定理只涉及不同素数；两两互素又保证另外两个系数在该模数下可逆。比如直接对 $X^2+Y^2-12Z^2=0$ 检查模 12 的条件，会把本来可以吸收到 Z 中的平方因子 4 与真正的模 3 障碍混在一起。先化成 $X^2+Y^2-3W^2=0$，判别的假设和需要检查的模数才完全一致。

24.4　Legendre 定理：先检查符号与剩余

圆 $x^2+y^2=3$ 在实平面上有很多点，却没有有理点。若有，清分母并约去公因数后会得到本原整数解 $X^2+Y^2=3Z^2$。平方模 3 只有 $0,1$，左边要同余于零，必须 $3\mid X,Y$；于是 $9\mid 3Z^2$，又得 $3\mid Z$，与本原性矛盾。这个例子表明：实数上的符号检验并不够，还要看有限模数。

定理 24.7（Legendre 的三元二次型判别）　设 a,b,c 为非零整数，$|abc|$ 无平方因子。方程

$$aX^2+bY^2+cZ^2=0$$

有非零整数解（等价地，非零有理解），当且仅当以下条件同时成立：

(1) a,b,c 不全同号；

(2) $-bc$ 是模 $|a|$ 的平方，$-ac$ 是模 $|b|$ 的平方，$-ab$ 是模 $|c|$ 的平方。

模 1 的条件约定自动满足；「是模 m 的平方」指存在整数 u 使 u^2 同余于该数。

直观理解　同号时，实数平方和无法抵消；某个模数不允许抵消时，整数方程也无解。令人惊讶的是，对已经规范化的三元二次型，这些必要条件合起来已经充分。证明充分性时会先把三个模数上的分解用中国剩余定理拼成一个分解，再用抽屉原理找到足够小的向量。

必要性的证明　把非零有理解清分母，再约去三个坐标的公因数，得到 $\gcd(X,Y,Z)=1$ 的整数解。若系数同号，三个非负平方乘以同号系数，只有全零才能使和为零，矛盾。

先证三个坐标也两两互素。若素数 $p\mid X,Y$，则 $p^2\mid cZ^2$。因为 c 无平方因子，p 在 c 中至多出现一次，故 $p\mid Z$，与本原性矛盾。另两对同理。

现取 $p\mid a$。系数两两互素给 $p\nmid b,c$，而模 p 的方程为 $bY^2+cZ^2\equiv 0$。若 $p\mid Y$，则 $p\mid Z$，违反两两互素；故 Y,Z 均可逆。两边乘 $-cY^{-2}$，得到

$$-bc\equiv (cZ/Y)^2 \pmod p.$$

对 $|a|$ 的每个素因子选择一个平方根，再用中国剩余定理合并。因 $|a|$ 无平方因子，便得模 $|a|$ 的平方根。交换变量可得另外两个条件。这里推出的是「同余于某个平方」，并非「同余于零」。 ■

■ **例 24.5　两个圆，两个不同结论**　比较 $x^2+y^2=3$ 与 $x^2+y^2=5$ 的有理点存在性。

解　齐次化后系数分别为 $(1,1,-3)$ 与 $(1,1,-5)$，乘积绝对值均无平方因子，符号也均不全相同。模 1 的两项检验自动通过，剩下只需判断 -1 是否为模 3 或模 5 的平方。模 3 平方为 $\{0,1\}$，$-1\equiv 2$ 不在其中，故第一个圆无有理点；模 5 平方为 $\{0,1,4\}$，$-1\equiv 4=2^2$，故第二个圆有有理点。实际基点 $(1,2)$ 正好验证了后一个结论。两者都有实点，区别来自同余条件。 ♦

模数取绝对值，并保留素数 2　Legendre 判别写「模 c」时，实际模数为 $|c|$。模 1 没有障碍，模 2 仍须按平方剩余定义理解：平方有 0,1 两类。规范形中若一个系数为偶数，另外两个必为奇数，相应的负乘积模 2 等于 1，所以这一项自动满足；不能为了省事直接将所有偶数因子从系数中删掉，那会改变方程。

24.5　充分性的证明：从同余找到真正的零点

模每个系数都有平方根，怎样把这些有限信息变成整数等式？先把符号统一：重排三个变量，必要时把方程乘以 -1，可设

$$a>0,\qquad b<0,\qquad c<0,\qquad M=abc>0.$$

若 $a=1,b=c=-1$，已有非零解 $(1,1,0)$。下面排除这个最小情形。记 $Q(X,Y,Z)=aX^2+bY^2+cZ^2$。

充分性分成三个可检查的步骤　第一步，把模 $a,|b|,|c|$ 的二次型分别分解成两个一次式，再用中国剩余定理得到 $Q\equiv LR\pmod M$。第二步，在一个小长方体内找两个点，使 L 的值同余；两点之差给非零向量，且 Q 只能等于 0 或 $-M$。第三步，若得到 $-M$，用一个恒等式把它转成真正的零点，并排除转成零向量的情况。

24.5.1　用中国剩余定理合并一次因子

设 $k^2\equiv -bc\pmod a$。因为 $\gcd(a,b)=1$，可以在模 a 下用 b^{-1}，从而

$$Q\equiv bY^2+cZ^2\equiv (Y+kb^{-1}Z)(bY-kZ)\pmod a.$$

展开右边得到 $bY^2-k^2b^{-1}Z^2$，而 $-k^2b^{-1}\equiv c$，所以等式成立。模 $|b|$ 与模 $|c|$ 的分解用同样方法得到。若某个模数为 1，该处的系数任意选择，不必定义逆元；它不会给中国剩余定理增加约束。

将各模数下的两个一次式分别写成 $L_i=\alpha_iX+\beta_iY+\gamma_iZ$、$R_i=\rho_iX+\sigma_iY+\tau_iZ$。对六个系数分别使用中国剩余定理，得到整数系数一次式 L,R，使它们在每个模数下都还原为相应的 L_i,R_i。由于三模数两两互素，乘积为 M，所以

$$Q(X,Y,Z)\equiv L(X,Y,Z)R(X,Y,Z)\pmod M$$

对所有整数三元组成立。我们不必要求 $L=0$，只需找到 $L\equiv 0\pmod M$，便能保证 $M\mid Q$。

24.5.2　小长方体中的抽屉原理

取所有满足

$$0\le X<\sqrt{|bc|},\qquad 0\le Y<\sqrt{|ac|},\qquad 0\le Z<\sqrt{|ab|}$$

的整数点。区间 $0\le n<t$ 含 $\lceil t\rceil$ 个整数：若 t 是整数，含 $0,\dots,t-1$；若非整数，含 $0,\dots,\lfloor t\rfloor$。因此点数至少为

$$\sqrt{|bc|}\sqrt{|ac|}\sqrt{|ab|}=M.$$

事实上严格多于 M。若三个根号都为整数，因 a,b,c 各自无平方因子且两两互素，$|ab|$、$|bc|$、$|ac|$ 都是平方就迫使 $|a|=|b|=|c|=1$，即已单独处理的情形。除此之外，至少一个向上取整严格增大，点数便严格大于 M。

把每个点按 L 模 M 的余数放进 M 个抽屉，必有两个不同点落入同一个抽屉。记它们的差为 (x,y,z)。两点不同，故此向量非零；同余相减则给出

$$L(x,y,z)\equiv 0\pmod M,\qquad M\mid Q(x,y,z),$$

以及严格界

$$|x| < \sqrt{|bc|}, \qquad |y| < \sqrt{|ac|}, \qquad |z| < \sqrt{|ab|}.$$

注意这里取的是两个点的差，差的坐标可为负；三条绝对值界却依旧严格成立。再利用系数符号，

$$Q(x,y,z) \le ax^2 < M, \qquad Q(x,y,z) \ge by^2 + cz^2 > -2M.$$

区间 $(-2M, M)$ 内的 M 的整数倍只有 $-M$ 与 0。若已得到 $Q = 0$，充分性得证。

24.5.3 把负的倍数转成零

若 $Q(x,y,z) = -M$，定义

$$x' = xz - by, \qquad y' = yz + ax, \qquad z' = z^2 + ab.$$

直接展开可见

$$a(xz-by)^2 + b(yz+ax)^2 = (z^2+ab)(ax^2+by^2),$$
$$ax'^2 + by'^2 + cz'^2 = (z^2+ab)\big(Q(x,y,z) + abc\big) = 0.$$

第一行的交叉项 $-2abxyz$ 与 $2abxyz$ 正好抵消，第二行再提出一个 z^2+ab；这就解释了三个新坐标的选择。它们为整数，满足所需方程。但还要确认没有同时为零。

如果新三元组为零，则 $z' = 0$ 给 $z^2 = -ab$。由于 $|a|, |b|$ 无平方因子且互素，$|ab|$ 也无平方因子，一个正整数既无平方因子又是平方，只能为 1。因此 $a = 1, b = -1$。这时原方程是 $X^2 - Y^2 + cZ^2 = 0$，已有非零整数解 $(1,1,0)$。所以不论变换是否退化，都能得到一个非零解，定理 24.7 的充分性完成。

这个证明提供了有限搜索的范围：先算 L，再枚举长方体内的点并找相同余数。它主要证明存在性，并不保证所得解最小，也不保证对很大的系数是最快算法。若只为找一个小基点，通常先检查几个小整数，找不到时才按证明提供的范围搜索。

■ **例 24.6　把充分性证明完整算一遍**　对 $5X^2 - 3Y^2 - 2Z^2 = 0$，写出模 30 的一次因子，并演示 $Q = -30$ 的转换。

解　系数乘积为 30。三项剩余条件分别是 $-6 \equiv 4 \pmod 5$、$10 \equiv 1 \pmod 3$、$15 \equiv 1 \pmod 2$，都成立。可选择

$$L = 5X + 3Y + 2Z, \qquad R = 25X + 9Y + 14Z.$$

相乘后，平方项系数模 30 分别为 $5, 27, 28$，交叉项系数为 $120, 120, 60$，均为 30 的倍数，故 $LR \equiv Q \pmod{30}$。

盒中整数坐标的范围是 $X = 0,1,2$、$Y = 0,1,2,3$、$Z = 0,1,2,3$，共 $48 > 30$ 个点。例如 $(0,2,3)$ 与 $(0,0,0)$ 都在盒内，L 分别为 $12, 0$，它们并不落在同一抽屉；不能随便挑两个点就声称满足抽屉结论。真正可取 $(0,2,0)$ 与 $(0,0,3)$，其 L 均等于 6，差为 $(0,2,-3)$，且 $Q = -12 - 18 = -30$。

用变换得 $x' = 6, y' = -6, z' = -6$，所以 $5\cdot 36 - 3\cdot 36 - 2\cdot 36 = 0$。约去公因数 6 得 $(1,-1,-1)$，再利用平方不受符号影响，也得到更直观的解 $(1,1,1)$。这里枚举盒子、找差、检查界、处理 $-M$ 四步都可逐一回代验证。 ♦

图 24.2 把例 24.6 中的两点放在长方体的 $X = 0$ 截面上。这个截面只含 16 个点；抽屉原理用的是整个长方体的 48 个点。图中的一对碰巧已在同一截面内，因而可以直接看出差向量的后两个坐标。

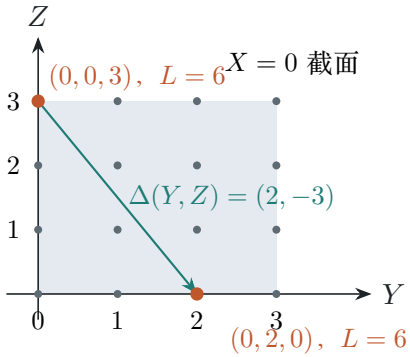


图 24.2　两个盒内点的 L 值相等，它们的非零差满足 $30 \mid Q$；差向量允许出现负坐标。

箭头从 $(0,0,3)$ 指向 $(0,2,0)$，对应相减次序不应反过来。得到的是差向量 $(0,2,-3)$，不是图上任意一个端点；代入 Q 得到 -30，再用恒等式转成方程的零点。这样，图上的「同一余数」才与证明中需要的整数等式接起来。

24.6　把存在性判别接回参数化

现在可以同时处理开头的两层问题。给定一条有理系数非退化圆锥曲线，先齐次化并化成规范对角式，再检查 Legendre 的符号与平方条件。任一条件失败，就已给出无有理点的证明；全部通过时，定理保证至少有一点，此后找一个基点，再用斜率法求全部有理点。图 24.3 把各步依赖画在一起。

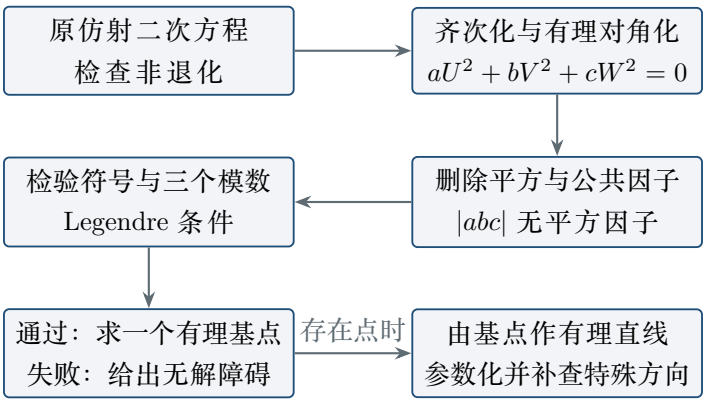


图 24.3　判存在须先化到规范形；存在性通过后，才把一个实际基点用于全部有理点的参数化。

图中「求一个基点」仍包含计算工作。定理的证明给出了一个有限但可能较大的搜索范围；找到对角式的解后，还要按此前保存的坐标变换逆回原曲线。若逆回的点在原无穷远直线上，使用齐次化定理中的换图方法再找仿射点。坐标变换一旦丢失，单独一个规范形的解并不能直接当作原方程的解。

■ **例 24.7　从模数检验到全部有理点**　判断 $2x^2 + 3y^2 = 11$ 是否有有理点；若有，从 $(2,1)$ 出发

参数化。

解　齐次式为 $2X^2+3Y^2-11Z^2=0$。系数乘积绝对值为 66，无平方因子，且符号不同。三项平方条件是

$$33\equiv 1 \pmod 2,\qquad 22\equiv 1 \pmod 3,\qquad -6\equiv 5\equiv 4^2 \pmod{11}.$$

因此有非零有理解。这里正的两个平方系数还保证非零解必有 $Z\neq 0$：若 $Z=0$，$2X^2+3Y^2=0$ 只可能 $X=Y=0$。搜索小整数即可找到 $(X,Y,Z)=(2,1,1)$，故基点为 $(2,1)$，回代有 $8+3=11$。

写 $x=2+t, y=1+mt$。代入得 $t[(2+3m^2)t+8+6m]=0$，所以第二个交点为

$$x=\frac{6m^2-6m-4}{2+3m^2},\qquad y=\frac{2-8m-3m^2}{2+3m^2},\qquad m\in\mathbb{Q}.$$

分母始终为正。斜率 $m=1$ 给 $(-4/5,-9/5)$，检验 $2\cdot 16/25+3\cdot 81/25=11$。切线斜率 $m=-4/3$ 返回 $(2,1)$；竖直线 $x=2$ 补上 $(2,-1)$。任意其他有理点与基点连线的斜率都是有理数，因此上述参数式加上补点就是全部有理点。 ♦

■ **例 24.8**　**有实点也可能被一个素数挡住**　证明 $x^2-2y^2=3$ 无有理点，并说明这是怎样的同余障碍。

解　规范齐次式为 $X^2-2Y^2-3Z^2=0$，系数乘积绝对值 6 无平方因子。它有实点 $(\sqrt{3},0)$，所以符号检验通过。但模 3 所需的 $-ab=2$ 不是平方，Legendre 条件失败。

也可直接给出递降式的必要性论证。假设有本原非零整数解，则模 3 得 $X^2+Y^2\equiv 0$，故 $3\mid X,Y$。再代回 $X^2-2Y^2=3Z^2$，得 $3\mid Z$，与本原性矛盾。后一种论证只用了模 3 的平方类，已经足够排除所有有理点，无须尝试更多分母。这同时说明模一个素数的无解检验，可能比寻找实数参数简便得多。 ♦

判别前不能跳过规范化　把定理机械套在 $(a,b,c)=(1,1,-4)$ 上，会要求 -1 是模 4 的平方；这个条件不成立，但原方程明明有 $(2,0,1)$。原因是 $|abc|=4$ 含平方因子，不符合定理假设。正确做法先令 $W=2Z$，化成 $X^2+Y^2-W^2=0$，规范后的三个模数都是 1，与已有解完全一致。

本讲的存在性判别常被概括为「局部条件决定整体解」。这里的「局部」只指实数符号与有限个模数上的平方条件，「整体」指同一个有理数域中的非零解。这个结论有严格的适用范围：有理系数、非退化、三元二次齐次式，且已经完成系数规范化。第 23 讲的四次方与三次方方程需要专门的递降论证，不能仅凭本讲就断言它们也满足同样的判别。

回看本课程的三条线索：同余把整数方程压缩到有限余数类；连分数从无限展开中抽出特殊的有理逼近；递降与抽屉原理则把「可能一直找不到」的问题转成有限步骤中的矛盾或构造。本讲把其中的中国剩余定理、平方剩余、互素性与抽屉原理接在同一个证明里。读者应能指出每个假设在证明中的落点：非退化用于参数化与齐次化，平方因子删除用于模数拼接及边界判断，符号用于估计 $-2M<Q<M$，有理数域用于合法缩放与清分母。把这些落点分清，遇到新的方程时才能知道已有方法可以用到哪里。

本讲小结

- **有理参数化**：非退化圆锥曲线已有有理点时，过基点的有理直线给出全部射影有理点；仿射公式另查切线、竖线与分母为零的方向。
- **齐次化**：$F(x,y)=0$ 换为 $H(X,Y,Z)=Z^2F(X/Z,Y/Z)=0$。非退化时，有仿射有理点当且仅当有非零整数三元解；$Z=0$ 须先换坐标图。
- **有理可逆变换**：配方与和差变换可消去交叉项；变换保留非零有理解存在性，不保证每个整数解逐个对应整数解。
- **规范形**：删除平方因子、除三系数公因数，再转移两系数共有的素因子，可使 $|abc|$ 无平方因子；必须保存逆变换。
- **Legendre 判别**：a,b,c 不全同号，且 $-bc,-ac,-ab$ 分别为模 $|a|,|b|,|c|$ 的平方，恰好刻画规范三元二次型的非零有理解存在性。
- **充分性构造**：模数上的一次分解经中国剩余定理拼接；小长方体与抽屉原理给 $Q=0$ 或 $-abc$；恒等式将后一种情况转为非零零点。

练习

24.1 ★ 对 $x^2+y^2=5$ 的参数式，计算 $m=0,1,-1,3$ 的点；找出返回基点 $(1,2)$ 的参数，并补齐遗漏的竖线交点。

24.2 ★ 判断下列说法真假并说明理由：齐次整数三元解必对应仿射整数点；非退化二次曲线必有有理点；非退化二次曲线若有一个有理点就有无穷多个有理点。

24.3 ★★ 按期末练习卷第 7 题处理

$$49x^2+5y^2+38z^2-28xy+70xz-28yz=0$$

找出有理可逆变换，把左边化为 $U^2+V^2-3W^2$，并判断是否有非零整数解。

24.4 ★★ 对二次型 $XY-Z^2$ 作有理对角化，写出变换和逆变换；用对角式的解 $(1,0,1)$ 解释为什么要允许清分母。

24.5 ★★ 对曲线 $x^2-3y^2=1$ 参数化全部有理点。以 $(1,0)$ 为基点，给出有理斜率参数式，检查斜率 $0,1,-1$，并说明基点如何处理。

24.6 ★★ 分别判断 $x^2+y^2=3$、$x^2+y^2=5$、$x^2-2y^2=3$ 是否有有理点；每次列出规范系数与实际起作用的模数。

24.7 ★★ 对 $5X^2-3Y^2-2Z^2$，验证例 24.6 的 L,R，写出长方体内两个 L 同余的不同点，计算差与变换后的非零解。

24.8 ★★★ 验证恒等式 $a(xz-by)^2+b(yz+ax)^2+c(z^2+ab)^2=(z^2+ab)(ax^2+by^2+cz^2+abc)$。在充分性证明的系数假设下，证明新三元组为零时原方程仍有非零解。

24.9 ★★★ 对 $2x^2+3y^2=11$，先用平方剩余判别有理点存在，再从 $(2,1)$ 推导全部有理点。列出一个非整数点，并检查切线与竖线。

24.10 ★★★ 解释为何规范化是必要步骤：用 $X^2+Y^2-4Z^2=0$ 给出直接套判别的错误结论，再正确规范化。最后证明：在 $|abc|$ 无平方因子的前提下，充分性证明长方体的点数恰等于 $|abc|$ 只可能发生在 $|a|=|b|=|c|=1$。

练习答案与提示

24.1 将四个参数代入 $x=(m^2-4m-1)/(1+m^2)$、$y=(2-2m-2m^2)/(1+m^2)$，依次得到 $(-1,2)$、$(-2,-1)$、$(2,1)$、$(-2/5,-11/5)$。前面三个点的平方和均为 5，最后一个为 $(4+121)/25=5$。令参数式的 $x=1$，得 $m^2-4m-1=1+m^2$，所以 $m=-1/2$；代回 $y=2$，这就是切线方向。竖线 $x=1$ 与圆相交于 $y=\pm2$，须再补 $(1,-2)$。若某点的横坐标不等于 1，与基点的斜率必为有理数，所以没有其他遗漏。

24.2 第一项假：整数解 $(-2,-11,5)$ 满足 $X^2+Y^2=5Z^2$，但对应 $(-2/5,-11/5)$；若 $Z=0$，甚至不能直接取比值。第二项假：$x^2+y^2+1=0$ 的齐次矩阵为单位阵，行列式非零，但它连实点都没有。第三项真：过已知有理点作有理直线，二次交点方程已有一个有理根，另一个也是有理数；不同方向除有限个特殊情况外给不同点。无穷远直线至多删去两个点，所以仿射有理点仍无穷多。三个命题分别涉及坐标的分母、点的存在与参数的变化，不能互相替代。

24.3 令 $U=7x-2y+5z$、$V=y-4z$、$W=z$，展开得

$$U^2+V^2-3W^2=49x^2+5y^2+38z^2-28xy+70xz-28yz.$$

变换矩阵的行列式为 7，故它在有理数上可逆，逆变换为 $z=W$、$y=V+4W$、$x=(U+2V+3W)/7$。若原式有非零整数解，则 (U,V,W) 是非零整数解；将它约成坐标互素的本原解仍满足同一齐次方程。模 3 有 $U^2+V^2\equiv0$，平方只可能为 $0,1$，所以 $3\mid U,V$。写 $U=3u,V=3v$ 并代回，得到 $W^2=3(u^2+v^2)$，故 $3\mid W$，与本原性矛盾。因此无非零整数解；可逆变换也说明原式无非零有理解。

24.4 令 $U=X+Y,V=X-Y,W=2Z$，则 $4(XY-Z^2)=U^2-V^2-W^2$。逆变换为 $X=(U+V)/2,Y=(U-V)/2,Z=W/2$，均是有理线性式；线性变换的行列式非零，故不会把非零向量变成零。对角式的整数解 $(1,0,1)$ 逆回 $(1/2,1/2,1/2)$，满足 $XY=Z^2=1/4$。整体乘以 2 后得到整数解 $(1,1,1)$，仍非零且仍满足原式，因为原式齐次。若要求逐个整数向量互相对应，还须加奇偶约束；存在性问题则只需允许清分母。

24.5 令 $y=m(x-1)$，代入后约去已知根，解得 $x=(3m^2+1)/(3m^2-1)$、$y=2m/(3m^2-1)$。分母对有理 m 不为零：若写 $m=p/q$ 为既约分数，$3m^2=1$ 会给 $3p^2=q^2$；于是 $3\mid q$，再由等式得 $3\mid p$，矛盾。取 $m=0,1,-1$ 得 $(-1,0),(2,1),(2,-1)$，逐一满足原方程。任一非基点都可由与 $(1,0)$ 的有理斜率得到，故参数式给出所有非基点；再补上 $(1,0)$。竖线 $x=1$ 只交于基点，没有另一个有限点。

24.6 第一式的规范系数为 $(1,1,-3)$，需 -1 为模 3 的平方；但平方类为 $0,1$，所以无解。第二式为 $(1,1,-5)$，所需 $-1\equiv4=2^2\pmod5$，其他模数为 1，故有解，例如 $(1,2)$。第三式为 $(1,-2,-3)$，需 2 为模 3 的平方，失败，所以无解。三式的系数乘积绝对值分别为 $3,5,6$，均无平方因子，且各自符号不全相同，故判断确实发生在定理允许的范围内。前后两个「无解」都指没有有理点，不仅是没找到小整数点。

24.7 展开 LR，平方项为 $125X^2+27Y^2+28Z^2$，交叉项为 $120XY+120XZ+60YZ$。减去 $5X^2-3Y^2-2Z^2$ 后，每项系数都是 30 的倍数，所以同余成立。取盒内两点 $(0,2,0)$ 与 $(0,0,3)$，两者 L 均为 6，差为 $(0,2,-3)$，满足严格坐标界且 $Q=-30$。用 $a=5,b=-3,c=-2$ 的变换得 $(x',y',z')=(6,-6,-6)$。回代为 $180-108-72=0$，约去 6 得 $(1,-1,-1)$。非零性和实际等式都已检查，抽屉原理在这里没有只停留在同余层面。

24.8 展开前两个平方，交叉项抵消后得 $(z^2+ab)(ax^2+by^2)$；第三项写成 $c(z^2+ab)(z^2+ab)$，合并即为题中恒等式。如果原向量满足 $Q=-abc$，右边为零，因此新向量满足 $Q=0$。若新向量同时为零，尤其有 $z^2=-ab$。在 $a>0,b,c<0$ 以及 $|abc|$ 无平方因子时，$|ab|$ 无平方因子，却又为完全平方，故只能是 1；于是 $a=1,b=-1$，原方程 $X^2-Y^2+cZ^2=0$ 有 $(1,1,0)$。这一步不可省，因为「构造出满足等式的三元组」本身未排除零向量。

24.9 规范系数为 $(2,3,-11)$，33 模 2、22 模 3 都同余于 1，-6 模 11 同余于 $5=4^2$，故判别通过。基点 $(2,1)$ 确实满足 $2\cdot4+3=11$。写 $x=2+t,y=1+mt$，消去已知根 $t=0$ 后得 $t=-(8+6m)/(2+3m^2)$，因此参数式为例 24.7 中的两式。取 $m=1$ 得 $(-4/5,-9/5)$，回代为 $(32+243)/25=11$。分子 $8+6m=0$ 时第二根也为零，故 $m=-4/3$ 是切线；横坐标为 2 时原式给 $y=\pm1$，所以竖线额外补点 $(2,-1)$。其他点都由与基点的有理斜率表示，故答案完整。

24.10 原系数 $(1,1,-4)$ 不满足乘积无平方因子，直接要求 -1 为模 4 的平方会错误判无解；实际 $(2,0,1)$ 已是非零整数解。令 $W=2Z$ 得 $X^2+Y^2-W^2=0$，规范系数 $(1,1,-1)$ 的条件均通过。在盒子点数问题中，三个因子分别为 $\lceil\sqrt{|bc|}\rceil,\lceil\sqrt{|ac|}\rceil,\lceil\sqrt{|ab|}\rceil$，各不小于相应根号，根号乘积为 $|abc|$。若乘积恰相等，每个因子都必须等于根号，所以三个成对乘积都是平方。若某个系数有素因子 p，由于系数两两互素且各自无平方因子，p 在含该系数的成对乘积中恰出现一次，与平方矛盾。故三个系数绝对值均为 1。反过来它们都为 1 时，盒内只有原点，点数确实等于 $1=|abc|$，所以最小情形必须另行处理。

复习课 3

平方和、连分数与丢番图方程：期末复习

Sums of Squares, Continued Fractions, and Diophantine Equations

对应材料：MIT 18.781 第 17–24 讲；期末复习指南；期末练习卷第 1、2、6、7、8、9、14 题及解答。其余题为所学内容的自测题。

本讲目标

(1) 把第 17–24 讲的平方和、连分数、Pell 方程与有理圆锥曲线连接起来。

(2) 面对计算、构造、判别与无解题，先选方法，再检查定理条件。

(3) 复查周期下标、范数符号、有理与整数坐标，并能解释全课程三部分的联系。

R3.1 知识地图：从表示到参数

同样是寻找整数，问题可能要求完全不同的答案：325 是否为两平方和，问的是存在性；Pell 方程的所有正解，问的是分类；圆锥曲线的全部有理点，问的是参数化；四次方方程无解，则需要一个能结束全部可能性的论证。复习时先把输出目标说清，再选择计算工具。

图 R3.1 把第 17–24 讲按三条联系排列。连分数的误差估计解释为何 Pell 解必出现在渐近分数中；周期长度决定取哪个下标；基本解的幂次再把个别解变成全部解。平方和与勾股数组则把因子分解、范数乘法和递降连接起来。

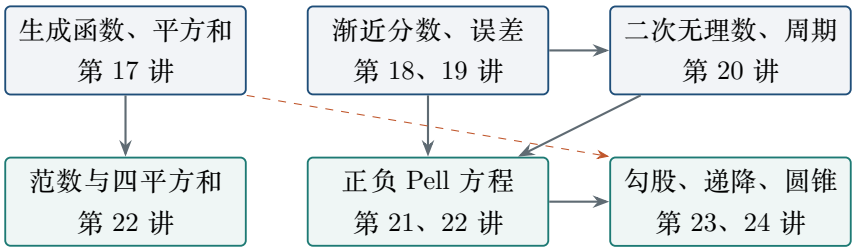


图 R3.1　误差与周期控制 Pell 解，范数控制平方和表示；橙色虚线表示平方和中的互素分解方法也用于勾股递降和圆锥判别。

图中的箭头表示解题时的依赖，不要求所有题都走完整条路径。模 8 已有矛盾时，可以立即结束；已有有理点的圆锥可直接作直线参数化，不必先做最一般的存在性判别。

下表把容易混用的概念分开。建议先遮住中间一列，仅凭左列名称回忆条件，再用右列检查自己是否知道它解决什么任务。

概念	定义或条件	典型用法
平方和判别	两平方和检查 3 (mod 4) 素因子的指数；四平方和允许零项	先判能否表示，再构造坐标
渐近分数	由部分商递推得到既约 p_n/q_n	误差界、定位小范数解
循环与纯循环	前者允许前段；后者从首项开始重复	用二次共轭辨别是否需要平移
Pell 基本解	$x, y > 0$ 中最小的单位范数正解	用基本单位的幂产生全部正解
本原数组	三坐标公因数为 1；勾股关系进一步给出两两互素	排除缩放冗余，合法拆分平方乘积
有理圆锥	二次曲线、非退化，且已知基点时才直接参数化	分清存在性与参数化，补查异常方向

R3.2　核心结论：记条件，也记用途

R3.2.1　筛选、平方和与范数（第 17、22 讲）

> **定理（两平方和与四平方和）**　正整数 N 可写成两个整数平方之和，当且仅当每个 $q \equiv 3 \pmod 4$ 的素因子在 N 中的指数均为偶数。每个非负整数都可写成四个整数平方之和，允许其中若干项为零。

关键　两平方和用复数范数乘法及素因子递降；四平方和用四元数范数乘法及乘数递降。完整证明分别见第 17、22 讲。两种范数都可相乘，但四元数乘法的次序不能任意交换。

用法　问「两平方和」先分解素因数，不能只检查 $N \pmod 4$；问「四平方和」先知存在，再寻找具体四个数。四平方和结论不能直接把项数减少为二，也不要求四个数都非零。

第 17 讲的另一部分是**按下标同余类筛选**。若 $A(t) = \sum_{n\ge 0} a_n t^n$、$\omega = e^{2\pi i/m}$，则

$$\sum_{n \equiv r \,(\mathrm{mod}\, m)} a_n t^n = \frac{1}{m} \sum_{j=0}^{m-1} \omega^{-rj} A(\omega^j t).$$

在形式幂级数中可逐项理解；若把 t 代成具体复数，则应在相关级数收敛处使用。多项式没有收敛问题。权重来自有限几何级数：当 $m \mid n - r$ 时指数和为 m，否则为零。

生成函数还给出交错平方倒数和：$\sum_{n\ge 1} (-1)^n/n^2 = -\zeta(2)/2 = -\pi^2/12$。这项计算把偶数项从 $\zeta(2)$ 中分离出来。

第 22 讲要同时记住**范数乘法**与**Pell 解分类**这两件事。若 $\varepsilon = x_1 + y_1\sqrt{d}$ 对应正 Pell 基本解，则全部正解恰为 ε^n，$n \ge 1$。对一般非零右端项 M，已有解 β 乘 ε^n 给出无穷多个解，却未必覆盖该右端项的所有解。存在种子与分类完备是两个不同的结论。

R3.2.2 渐近分数与逼近的方向（第 18、19 讲）

定理（递推、行列式与误差）　简单连分数的渐近分数满足

$$p_n = a_n p_{n-1} + p_{n-2}, \quad q_n = a_n q_{n-1} + q_{n-2},$$

初值为 $p_{-2} = 0, p_{-1} = 1, q_{-2} = 1, q_{-1} = 0$，且

$$p_{n-1}q_n - q_{n-1}p_n = (-1)^n.$$

对无限简单连分数的值 x，有 $|x - p_n/q_n| < 1/(q_n q_{n+1})$。

关键　递推与二阶行列式同时归纳；相邻渐近分数分居极限两侧，距离为 $1/(q_n q_{n+1})$，所以分别到极限的距离更小。完整证明见第 18 讲。注意行列式中相邻两项的排列决定右边符号。

用法　给定若干部分商时，先列 p_n, q_n 两行；只有需要误差时才引入后一个分母。精确的分数递推通常比分别计算小数更可靠，尤其适合比较两条有公共前缀的连分数。

第 19 讲还给出相邻逼近保证：每两个相邻渐近分数中至少一个满足 $|x - p/q| \le 1/(2q^2)$；每三个连续项中至少一个满足 $|x - p/q| < 1/(\sqrt{5}q^2)$。它们保证的是窗口内至少有一项，不是每项都满足。

第 19 讲的**Legendre 逼近判别**说：若 x 无理、$p \in \mathbb{Z}$、$q > 0$，且 $|x - p/q| < 1/(2q^2)$，则 p/q 一定是渐近分数。这是充分条件，反向不能照写。**Hurwitz 逼近**则说每个无理数有无穷多个互异既约分数 p/q（$p \in \mathbb{Z}$, $q > 0$）满足 $|x - p/q| < 1/(\sqrt{5}q^2)$；该常数不能对所有无理数统一减小。前者辨认一个给定分数，后者保证无穷多个分数存在，不应混淆量词。

R3.2.3 循环连分数定位 Pell 解（第 20、21 讲）

定理（循环性、共轭与单位范数）　实数是二次无理数，当且仅当其简单连分数最终循环。二次无理数 x 的连分数纯循环，当且仅当 $x > 1$ 且 $-1 < \bar{x} < 0$。

设 $d \ge 2$ 为非平方整数，$\sqrt{d}$ 的最小周期长度为 r。所有正整数解 $X^2 - dY^2 = \pm 1$ 均位于渐近分数下标 $n = kr - 1$，其中 k 取正整数；相应范数为 $(-1)^{kr}$。

关键　第 20 讲把完全商写成有限多个整数状态，从状态重复得到循环；共轭条件区分是否从首项就开始重复。第 21 讲用完全商分母的返回位置，证明单位范数只会出现在 $n + 1$ 是最小周期长度 r 的倍数时。完整证明见这两讲。

用法　先算最小周期，再看奇偶：r 偶时负 Pell 无解，正 Pell 基本解在 $r - 1$；r 奇时负 Pell 基本解在 $r - 1$，正 Pell 基本解在 $2r - 1$。这里渐近分数从下标零开始，周期末项不直接等同于基本解的截断位置。

两条常用的辅助结论也值得保留。第一，正 Pell 解满足

$$\left|\sqrt{d} - \frac{X}{Y}\right| = \frac{1}{Y^2(X/Y + \sqrt{d})} < \frac{1}{2Y^2},$$

所以由第 19 讲的判别，它必为渐近分数。第二，第 21 讲的小范数判别还覆盖 $|N|<\sqrt{d}$ 的正本原解 $X^2-dY^2=N$；去掉本原条件，X/Y 约分后可能已经改变原来的范数。遇到负右端项时，不能照搬正 Pell 的分母估计，应按原证明处理符号。

R3.2.4 互素分解、递降与圆锥（第 23、24 讲）

> **定理（两类参数化和一个无解结论）**　本原勾股数组按奇直角边在先排列为
>
> $$(r^2-s^2,2rs,r^2+s^2),\qquad r>s>0,\quad \gcd(r,s)=1,$$
>
> 其中 r,s 异奇偶。方程 $x^4+y^4=z^2$ 没有三个坐标均为正的整数解。非退化有理圆锥曲线若已有有理点，则经过该点的有理直线给出全部射影有理点；仿射公式应另查异常方向。

关键　勾股分类把互素的半和、半差分别写成平方；四次方结论用两次分类产生更小同类解。圆锥参数化则利用二次方程已有一个有理根，另一个根可由有理运算求出。完整证明见第 23、24 讲。

用法　看到平方乘积先查公因数；看到「所有有理点」先找基点并保留分母；看到「无整数解」先取小模数，若迫使全体坐标有公共因子，再考虑递降。三种方法的目标不同，但都要求明确正、零、负坐标的范围。

第 24 讲的存在性判别在参数化之前使用。把非退化三元二次型经有理可逆变换化成 $aU^2+bV^2+cW^2=0$，再规范到 $a,b,c\in\mathbb{Z}\setminus\{0\}$ 且 $|abc|$ 无平方因子。它有非零有理解，当且仅当三系数不全同号，并且

$$-bc \text{ 为模}|a| \text{ 的平方},\quad -ac \text{ 为模}|b| \text{ 的平方},\quad -ab \text{ 为模}|c| \text{ 的平方}.$$

模 1 条件自动成立。这是**Legendre 三元二次型判别**，与第 19 讲同名的逼近判别是不同结论。清除分母把非零有理解转成非零整数三元解，但仿射整数点不随任意有理变换逐个保留。

R3.3 八类题的完整计算

R3.3.1 筛选系数与构造平方和

例 R3.1 练习第 17 讲与生成函数前序内容的连接：下标的同余类不应通过逐项相加来硬算，先把题目转成单位根权重。例 R3.2 则比较两个表示问题的限制强弱。

■ **例 R3.1　第 17 讲单位根筛选的自测**　求 $S=\sum_{0\le k\le 10,\ k\equiv 1\,(\mathrm{mod}\,3)}\binom{10}{k}$。

解　取多项式 $A(t)=(1+t)^{10}$，令 $\omega^3=1,\omega\neq 1$。要筛出余 1 的下标，使用权重 ω^{-j}：

$$S=\frac{1}{3}\left(A(1)+\omega^{-1}A(\omega)+\omega^{-2}A(\omega^2)\right).$$

因为 $1+\omega+\omega^2=0$，有 $1+\omega=-\omega^2$、$1+\omega^2=-\omega$，所以

$$A(\omega)=(-\omega^2)^{10}=\omega^2,\qquad A(\omega^2)=(-\omega)^{10}=\omega.$$

代入得到 $S = (1024 + \omega + \omega^2)/3 = (1024 - 1)/3 = 341$。再用题中只有四项这一事实核对：$\binom{10}{1} + \binom{10}{4} + \binom{10}{7} + \binom{10}{10} = 10 + 210 + 120 + 1 = 341$。两种计算一致。 ♦

点评　特征是「二项系数且下标满足同余」。权重应由余数 r 决定；余 0 时才可不加权。这里是多项式，代入单位根不涉及级数收敛；若换成无限生成函数，就要先交代代入是否合法。

■ **例 R3.2　两平方和与四平方和的比较**　判断 325 能否写成两个整数平方之和，并构造一组表示。7 能否写成两个平方之和？能否写成四个平方之和？

解　先分解 $325 = 5^2 \cdot 13$。其奇素因子均模 4 余 1，所以满足两平方和判别。为了构造，可利用 $25 = 3^2 + 4^2$、$13 = 2^2 + 3^2$，由范数乘法

$$325 = (3 \cdot 2 - 4 \cdot 3)^2 + (3 \cdot 3 + 4 \cdot 2)^2 = (-6)^2 + 17^2 = 6^2 + 17^2.$$

回代得 $36 + 289 = 325$。题目只要求一组表示，因此不必进一步枚举；例如 $1^2 + 18^2$ 和 $10^2 + 15^2$ 也成立。

7 是模 4 余 3 的素数，指数为奇数，所以不是两平方和。四平方和则总存在，并可直接写成 $7 = 2^2 + 1^2 + 1^2 + 1^2$。这四项都非零只是本例碰巧如此，定理本身允许零项。 ♦

点评　先判存在，再构造坐标，避免在无解的情形里盲目搜索。第 17 讲控制两平方表示的素因子，第 22 讲保证四平方表示；表示的项数改变，适用的结论也会改变。

R3.3.2　由共同前缀控制两条连分数

例 R3.3 来自期末练习卷第 2 题。原解答中有一个渐近分数的分母排印错误，复习时应按递推重算，不能把印出的分数与后文的误差界混着使用。

■ **例 R3.3　期末练习卷第 2 题**　设 $\theta_1 = [3, 1, 5, 9, a_1, a_2, \ldots]$、$\theta_2 = [3, 1, 5, 7, b_1, b_2, \ldots]$，其中尾部部分商均为正整数。证明 $|\theta_1 - \theta_2| < 49/7095$。

解　两者共享渐近分数 $[3, 1, 5] = 23/6$。先按递推列出准确的分子、分母：

部分商	3	1	5	9（或 7）
p_n	3	4	23	211（或 165）
q_n	1	1	6	55（或 43）

因此 $[3, 1, 5, 9] = 211/55$，不是 $211/56$。第 18 讲的误差估计分别给出

$$\left|\theta_1 - \frac{23}{6}\right| < \frac{1}{6 \cdot 55}, \qquad \left|\theta_2 - \frac{23}{6}\right| < \frac{1}{6 \cdot 43}.$$

用三角不等式把两个值都连接到共同的 $23/6$：

$$|\theta_1 - \theta_2| < \frac{1}{330} + \frac{1}{258} = \frac{49}{7095}.$$

这里两个尾部是无限的，误差界为严格不等式，最后的结论也严格。 ♦

点评　看到公共前缀，就想到共同渐近分数。若只要求一个上界，无需求未知尾部；后一个部分商已经确定了下一分母。第 18 讲提供递推和误差，第 19 讲强调逼近结论的使用方向。误差估计的 q_nq_{n+1} 不能误写为随意选择的两个分母乘积。

R3.3.3 从周期计算基本解，再用乘法生成

例 R3.4 与例 R3.5 分别展示偶周期、奇周期。先定位正负范数，再谈幂次生成，计算顺序就不容易错。

例 R3.4　期末练习卷第 1 题　先求 $\sqrt{6}$ 的简单连分数，再求 $|X^2-6Y^2|=1$ 的三组正整数解，并说明负号是否可能。

解　取倒数得到

$$\frac{1}{\sqrt{6}-2}=\frac{\sqrt{6}+2}{2}=2+\frac{\sqrt{6}-2}{2},\qquad \frac{2}{\sqrt{6}-2}=\sqrt{6}+2=4+(\sqrt{6}-2).$$

所以 $\sqrt{6}=[2,\overline{2,4}]$，最小周期 $r=2$。正 Pell 基本解的下标为 $r-1=1$，由 $[2,2]=5/2$ 得 $(X_1,Y_1)=(5,2)$；直接检查 $25-6\cdot 4=1$。

周期为偶数，负 Pell 方程无解。其余正解用 $\varepsilon=5+2\sqrt{6}$ 的幂：

$$\varepsilon^2=49+20\sqrt{6},\qquad \varepsilon^3=485+198\sqrt{6}.$$

因此所求三组为 $(5,2),(49,20),(485,198)$。范数乘法保证三组右端均为 1；第 22 讲还说明所有正解都在这条幂次序列中。♦

点评　此题连用第 20 讲的循环结构、第 21 讲的周期下标、第 22 讲的全部解形式。求三组时乘基本单位即可；若要求「最小」或「全部」，必须引用相应定位或分类定理，不能只展示三个算出来的解。

例 R3.5　期末练习卷第 6 题　求 $\alpha=[12,\overline{24}]$ 的值，并求 $X^2-145Y^2=1$ 的最小正整数解。

解　令纯循环尾巴 $t=[\overline{24}]$，则 $t=24+1/t$，解得 $t=12\pm\sqrt{145}$。由于 $t>24$，只能取正根 $t=12+\sqrt{145}$。于是

$$\alpha=12+\frac{1}{12+\sqrt{145}}=12+(\sqrt{145}-12)=\sqrt{145}.$$

这里有理化用到了 $145-144=1$。因此 $\sqrt{145}$ 的最小周期为 1，为奇数。第一截断 $[12]=12/1$ 给出 $12^2-145=-1$，对应负 Pell 基本解，不能作为正 Pell 的答案。

正 Pell 基本解取下标 $2r-1=1$，即

$$[12,24]=12+\frac{1}{24}=\frac{289}{24}.$$

故最小正解为 $(289,24)$，验算 $289^2-145\cdot 24^2=1$。也可先平方范数为 -1 的单位：$(12+\sqrt{145})^2=289+24\sqrt{145}$，得到同一结果。♦

点评　看到长度为奇数的周期，应立刻区分正、负 Pell。尾巴方程有两个根，选根要依赖连分数的正性；不能仅凭「满足同一个二次方程」就宣布相等。若题目改求负 Pell 基本解，答案就是 $(12,1)$。

R3.3.4 配方判无解与直线参数化

例 R3.6 先问是否有点；例 R3.7 已给出基点，直接问全部有理点。两种任务在第 24 讲中应按不同顺序处理。

■ **例 R3.6　期末练习卷第 7 题**　判断 $49x^2+5y^2+38z^2-28xy+70xz-28yz=0$ 是否有非零整数解。

解　先保留可逆变换，配方为

$$(7x-2y+5z)^2+(y-4z)^2-3z^2=0.$$

令 $U=7x-2y+5z, V=y-4z, W=z$。逆变换为

$$z=W,\qquad y=V+4W,\qquad x=(U+2V+3W)/7.$$

这是有理可逆变换，因此原方程有非零有理解当且仅当 $U^2+V^2-3W^2=0$ 有非零有理解。规范系数为 $(1,1,-3)$，乘积绝对值 3 无平方因子，且不同号；但 Legendre 判别还要求 -1 为模 3 的平方。平方模 3 只有 $0,1$，所以条件失败，非零有理解不存在，原方程自然也无非零整数解。

也可直接核查障碍的整除含义：任一非零有理解分母清除并约去坐标公因数后，得到本原非零整数解。由 $U^2+V^2=3W^2$ 模 3 可得 $3\mid U,V$；代入后再得 $3\mid W$，与本原性矛盾。 ♦

点评　交叉项提示先配方。变换含除以 7，保留的是有理可解性，不是整数解之间的逐点双射；判无解只需在更大的有理范围内排除即可。系数不同号只是实数可解的必要条件，还要检查模数条件。

■ **例 R3.7　期末练习卷第 14 题**　参数化双曲线 $x^2-3y^2=1$ 的全部有理点，并各给出一个整数点和一个坐标不全为整数的有理点。

解　取基点 $P=(1,0)$，过它的非竖直有理直线写成 $y=t(x-1)$，$t\in\mathbb{Q}$。代入并因式分解：

$$x^2-3t^2(x-1)^2-1=(x-1)\big(x+1-3t^2(x-1)\big)=0.$$

除去已知交点 $x=1$，另一根满足 $(3t^2-1)x=3t^2+1$，所以

$$x=\frac{3t^2+1}{3t^2-1},\qquad y=\frac{2t}{3t^2-1}.$$

对有理 t，分母不会为零，否则 $3t^2=1$ 将说明 $\sqrt{3}$ 有理。两坐标均有理，回代可核查 $x^2-3y^2=1$。任意非基点有理点的 $x\neq 1$，故可反求 $t=y/(x-1)\in\mathbb{Q}$，因此没有漏掉非基点。最后另行补上 P；竖直线 $x=1$ 在此处切触，只给出基点。

例如 $t=1$ 得整数点 $(2,1)$；$t=2$ 得 $(13/11,4/11)$，其坐标不全为整数；$t=0$ 得 $(-1,0)$。参数公式与基点合起来才是全部有理点。 ♦

点评　已知有理点与二次方程的已知根，使另一个根可以显式计算。要求「全部」时，要写反求参数和异常点，而不只给公式。这里整数正点另由 Pell 基本单位的幂分类，随意取有理斜率只保证有理坐标。

R3.3.5 先用低模数，必要时再递降

例 R3.8 提醒我们：方程出现多个变量，并不意味着需要复杂的消元。有时只看平方的少数剩余类就已足够。

■ **例 R3.8　期末练习卷第 9 题**　证明 $x^2+2y^2=8z+5$ 没有整数解。

解　原方程模 2 强迫 x 为奇数，所以 $x^2\equiv 1 \pmod 8$。再模 8 得

$$2y^2\equiv 5-1\equiv 4 \pmod 8.$$

但 y 偶时 y^2 被 4 整除，$2y^2\equiv 0 \pmod 8$；y 奇时 $y^2\equiv 1 \pmod 8$，$2y^2\equiv 2 \pmod 8$。两类都不能为 4，矛盾。论证覆盖所有正、零、负坐标。 ♦

点评　模数取 8 是因为它消去 z，并将奇平方固定为 1。这里已有剩余类矛盾，证明结束。第 23 讲的四次方无解则在异奇偶情形下没有这种直接矛盾，需要用互素分解构造更小的解；不要把一种无解方法当成对所有方程都必定有效的步骤。

R3.4 六个检查点：避免把条件丢在计算之外

筛选余数时漏掉权重　错误说法：直接平均 $A(\omega^j)$ 就能取得任何余数类。反例：对 $(1+t)^{10}$，不加权的三项平均取得余 0 类，结果为 341；余 2 类的正确和却为 342。正确做法是先写目标余数 r，再放入权重 ω^{-rj}，最后除以模数。

把充分条件当成等价条件　错误说法：渐近分数必满足 $|x-p/q|<1/(2q^2)$。反例：黄金分割比 $\varphi=(1+\sqrt5)/2$ 的首个渐近分数为 $1/1$，误差 $(\sqrt5-1)/2>1/2$。正确说法是满足这个严格误差界的既约分数一定是渐近分数；未通过它，仍需按部分商检查。

奇周期直接取第一周期截断作正解　错误说法：周期长 r 时，正 Pell 基本解总在下标 $r-1$。反例：$\sqrt{145}$ 的周期长 1，首截断 $12/1$ 的范数为 -1。正确做法是先看 r 奇偶，再选 $r-1$ 或 $2r-1$；末尾始终代回范数检查符号。

把任意种子都当成全部解的起点　错误说法：任意一组正解乘基本单位的非负整数幂，都覆盖全部正解。反例：$x^2-2y^2=1$ 的解 $(17,12)$ 对应基本单位 $(3+2\sqrt2)^2$，从它向前乘幂永远漏掉更小的 $(3,2)$。正确说法是正 Pell 的全部正解由基本单位的全部正整数幂给出；一般右端项还可能需要区分不同种子。

未规范化就使用三元二次型判别　错误说法：对 $X^2+Y^2-4Z^2=0$ 直接要求 -1 是模 4 的平方。反例：$(2,0,1)$ 已是非零整数解，所写剩余条件却失败。正确做法先通过有理可逆缩放令 $W=2Z$，化成系数 $(1,1,-1)$，再使用 $|abc|$ 无平方因子的判别。

把齐次整数解读成仿射整数点　错误说法：$X^2+Y^2=5Z^2$ 有整数三元解，所以该组坐标除以 Z 后仍为整数。反例：$(-2,-11,5)$ 给出 $(-2/5,-11/5)$，它有理但非整数。正确说法是清分母保留有理点的存在性；要找仿射整数点，必须另外要求分母整除分子，并处理 $Z=0$ 的情形。

这六个检查点可以用在答题末尾：公式里是否带对权重，结论方向是否说反，Pell 范数是否验过，是否真的证明了「所有」，有理变换的逆式是否保存，坐标范围是否与题目一致。逐项核查比只复算算术更容易找出结构性的错误。

R3.5　全课程：三部分怎样接到一起

图 R3.2 按既定三次复习的范围概括全课程。每个部分都同时有计算工具和证明方法；后面的题常会回到第一部分的公因数、同余与中国剩余定理。

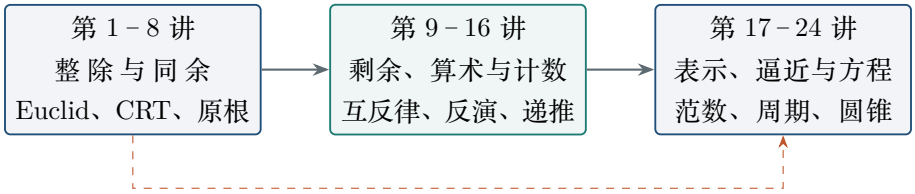


图 R3.2　课程从整除与同余走向算术结构，再处理表示和丢番图方程；橙色虚线表示公因数、同余与 CRT 在最后一部分仍直接参与证明。

第 1–8 讲：把整数条件化成可计算的形式。最大公因数与 Bézout 等式回答一次方程能否求解，也给出互素约分的依据。带余除法发展为同余运算，中国剩余定理把互素模数上的信息拼接起来。幂的周期、阶与原根进一步组织非零剩余类，支撑素性检验和 RSA 等应用。最后一部分里，连分数来自欧几里得算法，互素乘积拆平方依赖唯一分解，圆锥判别的构造仍使用 CRT。

第 9–16 讲：把算术对象按结构分解。二次剩余与互反律把「一个数是不是平方」的问题转成素数之间的同余信息。分圆多项式用阶来组织单位根，算术函数与 Möbius 反演则把整除求和拆回原函数。组合计数、线性递推和生成函数把无限多个系数压缩进一个代数表达式。第 17 讲的单位根筛选接续这些工具，而平方和与圆锥的局部条件重新使用二次剩余。

第 17–24 讲：同时控制大小、整除和几何。平方和表示用范数乘法把素数层面的结果推广到一般整数。连分数一面给出误差界，一面用循环状态把二次无理数与 Pell 整数解联系起来。勾股数组将有理圆点、互素分解及无穷递降放在同一问题中，圆锥曲线再把这种参数化推广。面对一道综合题，先分清要判断存在、求一组、列全部还是证明无解，再选择局部同余、参数公式或最小解方法。

本讲小结

- 下标满足同余类时用单位根筛选，保留 ω^{-rj} 与 $1/m$；无限生成函数还要检查代入的收敛条件。
- 两平方和先检查 3 (mod 4) 素因子的指数；四平方和总存在，但坐标允许为零。范数乘法负责从已有表示构造新表示。
- 连分数先列 p_n, q_n，用行列式检验计算，再用 $1/(q_n q_{n+1})$ 控误差；Legendre 的严格界是充分条件。
- 求 Pell 最小解先算最小周期及奇偶；列全部正解再用基本单位幂次。每个候选都应回代检查范数符号。
- 勾股参数先检查互素、异奇偶与正性；无穷递降还要验证新解为整数、同类、正且严格更小。
- 圆锥题先判断有无有理基点；有点时作直线参数化，无点时配方、规范化并查符号和平方剩余条件，最后分清有理点与整数点。

练习

R3.1 ★ 用单位根筛选求 $\sum_{0\le k\le 8,\ k\equiv 0 \,(\mathrm{mod}\, 4)} \binom{8}{k}$，并直接核对结果。（第 17 讲）

R3.2 ★ 求 $[2,3,1,4]$ 的各个渐近分数，用末两项检验行列式恒等式。（第 18 讲）

R3.3 ★★ 对 $\varphi = (1+\sqrt{5})/2$，判断 $1/1$、$2/1$ 是否为渐近分数；分别检查它们是否满足 Legendre 严格误差界。说明能从检查结果推出什么。（第 19 讲）

R3.4 ★ 用共轭条件判断 $\sqrt{3}$ 与 $1+\sqrt{3}$ 是否纯循环，并求后者的最小循环块。（第 20 讲）

R3.5 ★★ 求 $X^2 - 3Y^2 = 1$ 的基本解和前三组正解，说明负 Pell 方程是否有解。（第 21、22 讲）

R3.6 ★ 判断 31 是否为两平方和，并给出一个四平方表示。说明两个定理为什么不矛盾。（第 17、22 讲）

R3.7 ★★ 在 $x^4 + y^4 = z^2$ 的假想最小正解证明中，已得到 $x^2 = r^2 - s^2, y^2 = 2rs$，其中 r, s 互素、异奇偶。补出 r 奇、s 偶，$r = a^2, s = 2b^2$ 的理由；再说明第二次参数化怎样得到更小同类解。（第 23 讲）

R3.8 ★★ 对 $X^2 + Y^2 - 5Z^2 = 0$ 检查规范化与 Legendre 的所有条件，给出非零整数解，并指出 $(-2,-11,5)$ 对应的仿射点是否为整数点。（第 24 讲）

R3.9 ★★★ 综合题：先将「两条整数直角边相差 1」写成 Pell 型方程，再构造三边均至少为 100 的一组三角形。必须说明本原性、参数条件以及最终回代。（第 21–23 讲；期末练习卷第 8 题）

R3.10 ★★★ 综合题：参数化 $x^2 - 3y^2 = 1$ 的全部有理点，并用 Pell 理论另写其全部正整数点。各给一个整数点与非整数有理点，解释为什么第一种参数化并未直接给出整数点分类。（第 18–24 讲）

练习答案与提示

R3.1 令 $A(t)=(1+t)^8$、$\omega=i$。单位根筛选给出 $\frac{1}{4}(A(1)+A(i)+A(-1)+A(-i))$。其中 $A(1)=256, A(-1)=0$；由 $(1+i)^2=2i$ 得 $A(i)=A(-i)=16$，所以结果为 72。直接核对 $\binom{8}{0}+\binom{8}{4}+\binom{8}{8}=1+70+1=72$。

R3.2 从规定初值开始递推得到 $2/1, 7/3, 9/4, 43/19$。最后两个渐近分数满足 $p_2q_3-q_2p_3=9\cdot19-4\cdot43=-1=(-1)^3$；行列式不为零且为单位，也说明末项分子、分母互素。最终值为 $43/19$。

R3.3 $\varphi=[1,\overline{1}]$，因此 $1/1$、$2/1$ 都是渐近分数。前者误差为 $\varphi-1=(\sqrt{5}-1)/2>1/2$，不满足严格界；后者误差为 $2-\varphi=(3-\sqrt{5})/2<1/2$，满足严格界。第二个检查独立保证 $2/1$ 是渐近分数，第一个检查失败却不能排除 $1/1$，因为判别仅为充分条件。

R3.4 $\sqrt{3}$ 的共轭 $-\sqrt{3}<-1$，故它不是纯循环。令 $x=1+\sqrt{3}$，则 $x>2$、$-1<\bar{x}=1-\sqrt{3}<0$，且 $x^2-2x-2=0$ 给出 $x=2+1/(1+1/x)$，所以纯循环块为 $(2,1)$。两部分商不同，最小周期为 2。

R3.5 $\sqrt{3}=[1,\overline{1,2}]$，周期为偶数 2，正 Pell 基本解取 $[1,1]=2/1$，负 Pell 无解。正单位为 $2+\sqrt{3}$，其平方为 $7+4\sqrt{3}$，立方为 $26+15\sqrt{3}$，所以前三组正解为 $(2,1),(7,4),(26,15)$。第 22 讲的全解定理保证按幂次列出的不只是部分正解。

R3.6 31 为模 4 余 3 的素数，在自己的分解中指数为 1，所以不是两平方和。四平方表示为 $31=5^2+2^2+1^2+1^2$，验算 $25+4+1+1=31$。四个平方之和允许比两个平方之和更多的组合，两项表示不存在并不排除四项表示。

R3.7 若 r 偶、s 奇，则 $x^2\equiv3\pmod4$，矛盾；故 r 奇、s 偶，且 y 为偶数。因此

$$(y/2)^2=r(s/2),\qquad \gcd(r,s/2)=1.$$

互素乘积为平方，故 $r=a^2, s=2b^2$。于是 (x,s,r) 是本原勾股数组：若素数同时整除 x,s，也会整除 r，与 $\gcd(r,s)=1$ 矛盾。第二次参数化给出 $x=t^2-u^2, s=2tu, r=t^2+u^2$，其中 t,u 互素且为正。由 $tu=b^2$ 得 $t=k^2, u=\ell^2$，再由 $r=a^2$ 得 $k^4+\ell^4=a^2$。新解坐标均为正，且 $a\le r<r^2+s^2=z$，与最小正 z 矛盾。

R3.8 系数为 $(1,1,-5)$，均非零且乘积绝对值 5 无平方因子；不全同号。模 1 的两项条件自动满足，模 5 需 -1 为平方，而 $2^2\equiv-1\pmod5$，条件成立。因此有非零整数解，例如 $(1,2,1)$。另一个给定解满足 $4+121=5\cdot25$，但对应仿射点 $(-2/5,-11/5)$，坐标不全为整数。

R3.9 由 $|a-b|=1$ 得 $\gcd(a,b)=1$，所以三角形本原。写 $a=r^2-s^2, b=2rs, c=r^2+s^2$，其中 $r>s>0$、互素且异奇偶；条件化为 $|(r-s)^2-2s^2|=1$。令 $A=r-s, B=s$，负 Pell 解 $(7,5)$ 满足 $49-50=-1$，故 $r=12, s=5$，仍满足参数条件。得到 $(a,b,c)=(119,120,169)$，且 $119^2+120^2=169^2$、$|119-120|=1$、$119\ge100$。

R3.10 由例 R3.7，全部有理点为

$$\left(\frac{3t^2+1}{3t^2-1},\frac{2t}{3t^2-1}\right),\quad t\in\mathbb{Q},$$

再补上 $(1,0)$。参数分母不为零；任一非基点可反求 $t=y/(x-1)$，故没有漏点。正整数点由 $x_n+y_n\sqrt{3}=(2+\sqrt{3})^n$（$n\ge1$）给出，因为周期长 2、基本解为 $(2,1)$，再用第 22 讲的全解定理。$t=1$ 得整数点 $(2,1)$，$t=2$ 得非整数有理点 $(13/11,4/11)$。参数化本身不分类整数点；全体整数解还需包含 $(\pm1,0)$ 并允许两个坐标分别变号。