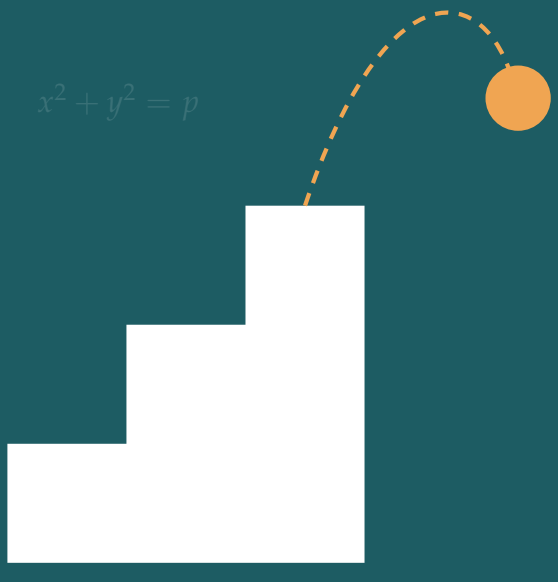
STEP 备试 · 大学数学

Beyond STEP

第 1 章

数论与证明



KEITH McNULTY

Beyond STEP

以历届 STEP 试题为门
走进大学数学

第 1 章

数论与证明

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

第 1 章：数论与证明

本册是 *Beyond STEP* 全书 13 章中的第 1 章。全书、各章的单行本、每一节的可列印工作纸，以及每一道 STEP 试题的详细解答，均可在 `beyondstep.keithmcnulty.org` 免费阅读及下载。

2026 年 11 月首次于网上发表。

本 PDF 内凡属作者本人创作的内容，包括正文、解答、插图、练习及答案，均按照 Creative Commons CC0 1.0 Universal 授权条款贡献予公有领域：你可以为任何目的复制、改编及分享，毋须征求许可。详见 `creativecommons.org/publicdomain/zero/1.0`。

本书讨论的 STEP 试题，版权属 Cambridge University Press & Assessment 所有。本书并无转载这些试题：每一题均附连结，通往官方来源的题目，即 STEP Support Programme 的试题资料库或 OCR 的历届试卷。

如有错漏，欢迎到 `github.com/keithmcnulty/step_book_mcnulty` 提交 issue 指正。

章

1

数论与证明

整数是数学中最古老的对象，却至今仍藏着数学中一些最深奥的秘密。

本章内容

1.1	质数工厂	STEP II 2002
1.2	质数分布得有多稀疏?	STEP III 2016
1.3	最小反例	STEP III 2015
1.4	e 是无理数，e^2 也是	STEP III 2024
1.5	无穷递降法	STEP III 2022
1.6	两个平方数之和	STEP II 2023
1.7	因数分解失效之时	STEP I 2016
1.8	计数有理数	STEP III 2020
1.9	三次曲线的切线	STEP I 2009
A	练习答案	第 67 页

为甚么是整数?

> 数学是科学的皇后，而数论（*number theory*）是数学的皇后。
>
> — 高斯（Carl Friedrich Gauss）

整数是每个人最先接触到的数学对象。你还未识字，就已经学会数数。然而，关于整数的最简单的问题，却可以难得出奇。每一个大于 2 的偶数，是否都是两个质数（prime）之和？没有人知道：哥德巴赫（Christian Goldbach）在 1742 年提出这个问题，电脑已经验证到 4×10^{18} 以外，但至今仍然没有人能够证明。像 11 和 13，或 101 和 103 这样相差 2 的质数对，是否有无穷多对？这同样没有人知道。不过在 2013 年，一位几乎寂寂无名的数学家张益唐（Yitang Zhang）震惊了全世界：他证明了有无穷多对质数，彼此相差不超过 70 百万，而这个数字后来已被降低至 246。费马（Pierre de Fermat）在 1637 年于书页边的空白处随手写下：当 $n \geq 3$ 时，$x^n + y^n = z^n$ 没有正整数解。人们花了超过 350 年，并动用了有史以来最精深的一些数学，才终于由怀尔斯（Andrew Wiles）在 1994 年证明他是对的。

问题连小孩子都听得懂，答案却可能要花上几个世纪——正是这种组合令数论叫人欲罢不能。这也是为甚么数论一直是证明（proof）的训练场。整数有无穷多个，你不可能逐一检查每个情况，于是就必须找出一个能一举涵盖所有情况的论证。你在往后的数学生涯中会用到的证明技巧，大部分最初都是在有关整数的问题上磨练出来的。

纯粹、无用，却支撑着互联网

1940 年，剑桥数学家哈代（G. H. Hardy）颇为自豪地写道：数论是所有数学中最纯粹的，与日常生活相距太远，不可能有任何实际用途，因而也不可能有任何有害的用途。他错得不能再错了。每当你在网上购物、发送加密信息或为手机解锁，你都在依靠数论。RSA 系统（1977 年）所依赖的事实是：把两个大质数相乘很容易，但要从它们的乘积还原出这两个质数，却似乎难得无从入手。现代的系统所用的，正是你会在本章末遇到的那些曲线上的算术。十八、十九世纪的“无用”数学，成了二十一世纪的安全保障。

你在大学会遇到甚么

如果你在大学修读数学，数论会是你最先到达的地方之一，也会是你最后到达的地方之一。

- **在第一个学期**，你几乎一定会修读一门名为*数与集合*（Numbers and Sets）或*数学基础*（Foundations）之类的课程。它会教你怎样正式地书写证明，而其中大部分素材都来自数论：归纳法（induction）、欧几里得算法（Euclidean algorithm）、模算术（modular arithmetic）、费马小定理（Fermat's Little Theorem）、中国余数定理（Chinese Remainder Theorem），以及一个令人惊讶的事实——有些无穷集合比另一些无穷集合更大。
- **在群论（group theory）中**（通常在第一年修读），你会发现许多数论事实，原来都是同一条关于对称的定理——拉格朗日定理（Lagrange's Theorem）——投下的影子。
- **在第二年**，一门*数论*课程通常会涵盖二次互反律（quadratic reciprocity，即高斯的"黄金定理"）、连分数（continued fractions）和佩尔方程（Pell's equation）。
- **再往后**，你还可以走得更远：*代数数论*（algebraic number theory），其中高斯整数（Gaussian integers）等新的数系挽救了唯一分解（unique factorisation）；*解析数论*（analytic number theory），它运用微积分和复分析（complex analysis）来数算质数；以及*椭圆曲线*（elliptic curves）的理论，它既是费马最后定理（Fermat's Last Theorem）证明背后的支柱，也是七道悬赏一百万美元的千禧年大奖难题（Millennium Prize Problems）之一的根源。

贯穿本章的问题

本章的九道 STEP 题目，围绕着三个大问题打转。做题时请把它们记在心上；我们会在本章结尾回到这些问题。

1. **质数有多少个？它们是怎样分布的？**人人都知道质数有无穷多个。但它们越往后变得有多稀疏？甚么样的论证才能告诉我们答案？
2. **怎样证明一个数不是分数？**要说明一个数是有理数（rational），只需举出一个例子。要说明它是无理数（irrational），却要排除无穷多种可能。这到底是怎样做到的？
3. **哪些方程有整数解？当我们改变量系时，又会发生甚么事？**有时我们能证明某个方程完全没有解。有时要了解整数，最好的方法却是走出整数的范围，进入复数（complex numbers），或走到一条曲线上。

关于证明的一章

在这三个问题底下，还贯穿着第四个主题：*怎样证明*。每一节之所以入选，部分原因是它展示了一种不同的证明方法的实际运用；合起来，它们就是一套相当适合大学一年级生

的工具。

§	课题	思考方式
1.1	费马数（Fermat numbers）	一个需要无穷多个质数的无穷族
1.2	质数的乘积	强归纳法（strong induction）；隐藏在二项式系数（binomial coefficients）中的信息
1.3	再谈 $\sqrt{2}$	“取最小的那一个”：良序原理（well-ordering principle）
1.4	e 与 e^2	把一个整数严格地困在 0 与 1 之间
1.5	不可能的方程	无穷递降法（infinite descent）；模 n 运算
1.6	两个平方数之和	转换场景：复数
1.7	一个奇怪的数系	质疑你视为理所当然的事
1.8	把分数逐一列出	构造双射（bijection）；反向的递降
1.9	立方数与曲线	用几何解决算术问题

怎样阅读本章

每一节都以一道 STEP 题目的连结开始，题目与当年考试所出的完全一样。请先自己尝试，至少做半小时，然后才往下读。一道你曾经苦苦思索过的题目，会比一道你只是读过的题目令你学到多得多。当你往下读时，问问自己方框是邀请你在答案出现之前先停下来想一想，而停一停，试一试方框则标示着关键思路已在你触手可及之处的时刻。之后，每一节的背后的数学部分会把目光越过考试，望向你在大学将会遇到的数学。

各节的深度大致上逐步递增，但它们基本上互相独立。如果某一节未能吸引你，就先跳过，稍后再回来。

1.1 质数工厂

题目

STEP II 2002，第 3 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/02-s2-q3`

引子

你几乎肯定早已知道质数（prime）有无穷多个。欧几里得（Euclid）约在公元前 300 年证明了这件事，这也是每个人最早接触到的证明之一。那么，STEP 的题目为甚么要你再证明一次？

因为这个证明的运作方式完全不同。欧几里得的论证是说：*给我任何一张有限的质数清单，我都能指出它并不完整。*而这个证明则说：*这里有一个预先写好、明确给出的数列（sequence），当中每一个数都含有一个其他各数都没有的质数。*它是一部制造新质数的机器，而它之所以行得通，全靠一条简洁优美的恒等式（identity），题目会一步一步引领你走向它。

途中你还会遇上一些连费马（Fermat）本人也被骗倒的数。

动笔之前

先把题目从头到尾读一遍，留意它的结构。它分为四个阶段：

1. 计算一些例子；
2. 找出（其实是题目告诉你）一个规律，并用这些例子验证；
3. 证明这个规律在一般情况下成立；
4. 两次运用这个规律：先推导出关于公因数（common factor）的结论，再推导出关于质数的结论。

这种“例子 → 猜想 → 证明 → 推论”的结构，正是许多数学研究的实际做法，STEP 的

题目也经常依循这个模式。遇到这种结构时，开头的部分通常不只是热身：你日后需要的直觉，正是在那里建立的。

> **想一想**　最后一行问的是*无穷多个质数*，中间问的是*公因数*。“任何两个费马数（Fermat number）都没有公因数”和“质数有无穷多个”之间，究竟有甚么关系？

请记住这个问题，我们稍后会回来讨论。

逐步解题

阶段 1　熟习这些数

要小心记号：2^{2^n} 是指 $2^{(2^n)}$，而不是 $(2^2)^n$。指数（exponent）本身就是 2 的幂，所以这些数增长得*非常*快。

$$F_0 = 2^1 + 1 = 3, \qquad F_2 = 2^4 + 1 = 17,$$
$$F_1 = 2^2 + 1 = 5, \qquad F_3 = 2^8 + 1 = 257.$$

> **想一想**　这四个数有甚么值得留意的地方？

它们全是质数。（验证 257：它不能被 $2, 3, 5, 7, 11$ 或 13 整除，而 $17^2 = 289 > 257$，所以可以就此停下。）下一个数 $F_4 = 2^{16} + 1 = 65537$ 也是质数。皮埃尔．德．费马在 1640 年正正注意到这一点，并猜想（conjecture）*所有*这种形式的数都是质数。请把这件事记在心里，稍后我们会看看他是否猜对了。

阶段 2　验证恒等式

现在就较小的 k 验证 $(*)$：

- $k = 1$：左边只有 $F_0 = 3$，而 $F_1 - 2 = 3$。✓
- $k = 2$：$F_0 F_1 = 3 \times 5 = 15$，而 $F_2 - 2 = 15$。✓
- $k = 3$：$F_0 F_1 F_2 = 15 \times 17 = 255$，而 $F_3 - 2 = 255$。✓

> **想一想**　看看由 $k = 2$ 到 $k = 3$ 的那一步。我们把 15 乘以 17，得到 255。用费马数来表达，就是计算了 $(F_2 - 2) \times F_2$，结果得到 $F_3 - 2$。这是巧合吗？

这个观察就是整个证明的缩影。让我们看看它为甚么应该成立。

阶段 3　证明恒等式

停一停，试一试 继续读下去之前，试试直接证明

$$(F_k - 2)\,F_k = F_{k+1} - 2.$$

提示：把所有东西都用 2 的幂来表示。$F_k - 2$ 是甚么样子的?

我们有 $F_k - 2 = 2^{2^k} - 1$ 和 $F_k = 2^{2^k} + 1$，所以

$$(F_k - 2)\,F_k = \left(2^{2^k} - 1\right)\left(2^{2^k} + 1\right).$$

想一想 这是一个形如 $(a-1)(a+1)$ 的乘积。它等于甚么?

它是平方差（difference of two squares）$a^2 - 1$。取 $a = 2^{2^k}$，则

$$a^2 = \left(2^{2^k}\right)^2 = 2^{2\cdot 2^k} = 2^{2^{k+1}},$$

所以

$$(F_k - 2)\,F_k = 2^{2^{k+1}} - 1 = F_{k+1} - 2. \tag{†}$$

这就是整部机器的引擎。留意平方的作用：它把指数加倍，而 2^k 加倍后就是 2^{k+1}。费马数的定义与平方差的因式分解（factorisation）配合得天衣无缝。

这样一来，归纳法（induction）几乎是水到渠成。

想一想 在归纳步骤中，我们究竟要证明甚么?

假设 $(*)$ 对某个 $k \geq 1$ 成立，即

$$F_0 F_1 \cdots F_{k-1} = F_k - 2,$$

而我们想证明它对 $k+1$ 也成立：

$$F_0 F_1 \cdots F_{k-1} F_k = F_{k+1} - 2.$$

把假设的等式两边同乘以 F_k，再运用 (†)：

$$F_0 F_1 \cdots F_{k-1} F_k = (F_k - 2)\,F_k = F_{k+1} - 2.$$

我们已经验证了 $k = 1$ 的情况，所以根据归纳法，$(*)$ 对所有 $k \geq 1$ 都成立。 ■

想一想 我们能否完全不用归纳法?

可以，而且值得看看怎样做。由 $F_0 - 2 = 1$ 出发，反复运用 (†)：

$$F_k - 2 = (F_{k-1} - 2)F_{k-1} = (F_{k-2} - 2)F_{k-2}F_{k-1} = \cdots = (F_0 - 2)\,F_0 F_1 \cdots F_{k-1}.$$

由于 $F_0 - 2 = 1$，证明便完成了。这种“逐层展开”的做法，其实就是把同一个归纳法完整地写出来，不过有些人会觉得这样更容易看明白。

阶段 4　没有公因数

我们想证明：若 $m \neq n$，则 F_m 与 F_n 没有大于 1 的公因数。

> **想一想**　假设某个数 d 同时整除 F_m 和 F_n，其中 $m < n$。F_m 在 $(*)$ 的哪个位置出现？

在 $(*)$ 中取 $k = n$：

$$F_0 F_1 \cdots F_m \cdots F_{n-1} = F_n - 2.$$

由于 $m < n$，F_m 是左边其中一个因数。所以若 d 整除 F_m，则 d 整除整个左边；也就是说，d 整除 $F_n - 2$。

> **想一想**　现在 d 既整除 F_n，又整除 $F_n - 2$。它还必定整除甚么？

它们的差：d 整除 $F_n - (F_n - 2) = 2$。所以 $d = 1$ 或 $d = 2$。

> **想一想**　d 可以是 2 吗？

不可以：每个费马数都是 2 的幂加 1，所以是奇数。因此 $d = 1$，即任何两个不同的费马数都没有大于 1 的公因数。 ■

数学家会说费马数*两两互质*（pairwise coprime）。

阶段 5　无穷多个质数

现在回到我们一开始提出的问题。

> **想一想**　每个大于 1 的整数都至少有一个质因数（prime factor）。从每个费马数中各选一个质因数。对于你选出的这些质数，你能说些甚么？

设 p_n 为任意一个整除 F_n 的质数。假如其中两个相等，例如 $p_m = p_n$ 而 $m \neq n$，那么这个质数就是 F_m 与 F_n 的一个大于 1 的公因数，但我们刚刚证明了这是不可能的。所以质数

$$p_0,\ p_1,\ p_2,\ p_3,\ \ldots$$

全都*互不相同*。费马数有无穷多个，所以不同的质数也有无穷多个。 ■

留意我们从来不需要知道每个费马数是被*哪一个*质数整除，甚至不需要知道费马数本身是不是质数。我们只需要知道它们永远没有公因数。每个费马数都必须“用掉”至少一个其他费马数不能用的质数。

n	$F_n = 2^{2^n} + 1$	质因数分解
0	3	质数
1	5	质数
2	17	质数
3	257	质数
4	65 537	质数
5	4 294 967 297	$641 \times 6\,700\,417$
6	18 446 744 073 709 551 617	$274\,177 \times 67\,280\,421\,310\,721$

图 1.1.1 最初几个费马数。费马猜想它们全是质数，但欧拉（Euler）在 1732 年找到了 F_5 的因数 641，而 F_4 之后至今仍未发现任何费马数是质数。即使如此，恒等式 $(*)$ 表明不同费马数的因数永不重叠：每一行都带来新的质数。

回顾

值得停下来看看这道题成功的关键何在，因为同样的想法会在本书中多次重现。

- **例子并不是装饰。**看出 $15 \times 17 = 255$ 其实是“$(F_2 - 2) \times F_2 = F_3 - 2$”，正是关键思路。当题目给你一些简单情况时，要问每一步在做甚么，而不只是看数字是否吻合。
- **一条恒等式完成了所有工作。**一切都源自 $(a-1)(a+1) = a^2 - 1$。看似艰深的 STEP 题目，往往只是由一小段简单的代数推动。
- **最后的论证着眼于结构，而非大小。**我们证明了质数有无穷多个，却没有找出任何一个新的质数。我们证明的是：要构成这个无穷的数族，必须用上无穷多个质数。

背后的数学

费马猜对了吗？

没有。1732 年，莱昂哈德．欧拉证明了

$$F_5 = 2^{32} + 1 = 4\,294\,967\,297 = 641 \times 6\,700\,417.$$

欧拉并不是靠运气找到 641 的。他证明了 F_n 的任何质因数都必定具有一种非常特殊的形式（当 $n = 5$ 时，它必定是 64 的倍数加一），从而大幅减少了需要考虑的可能。你可以在练习 4 中亲自证明这一点。

自此以后，凡是有人成功检验过的费马数，由 F_5 到 F_{32}，结果全都是合成数（composite）。就目前所知，F_0 至 F_4 可能是仅有的费马质数，但没有人能够证明这一点。这是一个货

真价实的未解难题。

妙就妙在我们的证明完全不受影响。不论是质数还是合成数，费马数依然能产生无穷多个质数。

费马质数与正多边形作图

费马质数会在一个相当意想不到的地方出现。古希腊人懂得用直尺和圆规作出等边三角形和正五边形，而请留意 $3 = F_0$ 和 $5 = F_1$。其后两千年间，没有人能作出其他边数为质数的正多边形（regular polygon）。到了 1796 年，年仅 19 岁的卡尔．弗里德里希．高斯（Gauss）示范了怎样作出*正十七边形*，而 $17 = F_2$。

高斯更进一步：正 N 边形可以用直尺和圆规作出，当且仅当 N 是 2 的幂乘以若干个*互不相同的费马质数*。（Pierre Wantzel 在 1837 年补全了这个证明。）所以哪些正多边形可以作图，取决于费马质数有多少个。你会在 Section 6.2（STEP II 1991，第 4 题）遇到高斯正十七边形背后的关键计算，而完整的来龙去脉则属于*伽罗瓦理论*（Galois theory），那是大学本科数学课程的精华之一。

一个定理，多种证明

数学家为甚么要不厌其烦地把同一个定理证明多于一次？因为不同的证明会揭示不同的东西。这个证明一般归功于克里斯蒂安．哥德巴赫（Goldbach）在 1730 年写给欧拉的一封信，它表明：*任何一个由大于 1 的整数组成、各项两两互质的无穷数列，都必然带来无穷多个质数*。费马数只是其中一个方便的例子。（练习 3 给出另一个例子。）

它给出的结论也比欧几里得的证明更多。质数 $p_0, p_1, \ldots, p_{n-1}$ 是 n *个不同的质数*，每个都不超过 F_{n-1}。所以不大于 $2^{2^{n-1}} + 1$ 的质数至少有 n 个。这是一个关于质数如何分布的*定量陈述*。它非常弱（实际情况要好得多），但它是迈向十九世纪数学其中一项伟大成果的第一步，那就是*质数定理*（Prime Number Theorem），它精确地描述了质数怎样变得越来越稀疏。Section 1.2 会踏出下一步。

思维技巧

从结构性质出发建立论证

要证明 A 类对象有无穷多个，可以找出无穷多个 B 类对象，而每个 B 类对象都需要一个专属于自己的 A 类对象。

这里 A = 质数，B = 费马数，而“需要一个专属于自己的”是指“有一个不与

其他任何一个共用的质因数”。同样的模式也出现在计数论证、线性代数（linear algebra）（无穷多个线性独立的向量必然导致无穷维）和分析（analysis）之中。

你也练习了*归纳法*，尤其是养成追问“*由 k 到 $k+1$ 的那一步实际上做了甚么？*”的习惯。这个问题往往比形式上的范本更有用。

延伸练习

1 **欧几里得的版本。**写出欧几里得对“质数有无穷多个”的证明。它在甚么意义上是“反证法”（by contradiction），而费马数的证明又在甚么意义上是“构造性”（constructive）的？

2 **末位数字。**证明当 $n \geq 2$ 时，费马数 F_n 的末位数字是 7。

3 **西尔维斯特（Sylvester）数列。**定义 $a_1 = 2$ 及 $a_{n+1} = a_n^2 - a_n + 1$，于是数列的开头几项是 $2, 3, 7, 43, 1807, \ldots$

(a) 证明 $a_{n+1} - 1 = a_1 a_2 \cdots a_n$。

(b) 由此推导出各项两两互质，从而给出“质数有无穷多个”的另一个证明。

(c) 较难 证明 $\dfrac{1}{a_1} + \dfrac{1}{a_2} + \cdots + \dfrac{1}{a_n} = 1 - \dfrac{1}{a_{n+1} - 1}$。

4 **欧拉的妙着。***本题需要用到费马小定理（Fermat's Little Theorem）：若 p 是质数且 $p \nmid a$，则 $a^{p-1} \equiv 1 \pmod{p}$。*假设质数 p 整除 F_n。

(a) 证明 $2^{2^n} \equiv -1 \pmod{p}$，从而证明 $2^{2^{n+1}} \equiv 1 \pmod{p}$。

(b) 设 d 为满足 $2^d \equiv 1 \pmod{p}$ 的最小正整数。证明 d 整除 2^{n+1} 但不整除 2^n，并由此得出 $d = 2^{n+1}$。

(c) 利用费马小定理，证明 d 整除 $p - 1$。由此推导出 $p \equiv 1 \pmod{2^{n+1}}$。

(d) 当 $n = 5$ 时，这表示 F_5 的任何质因数都具有 $64k + 1$ 的形式。验证 641 确实如此。小于 641 而又具有这种形式的质数有多少个？

5 **质数的一个上界。**利用*一个定理，多种证明*中的论证，证明第 n 个质数不超过 $2^{2^{n-1}} + 1$。当 $n = 5$ 时，这个上界有多准确？

延伸阅读

▶ **Martin Aigner & Günter M. Ziegler, *Proofs from THE BOOK***（Springer），第 1 章“Six proofs of the infinity of primes”（质数无穷性的六个证明）。书中除了这个证明外，还

收录了欧几里得的证明和另外四个证明，其中一个用上了拓扑学（topology）。这是入读大学前最值得一读的书之一。

- **Harold Davenport, *The Higher Arithmetic*** （Cambridge University Press），第 1–2 章。一本文笔优美的入门书，介绍质数与因数分解，最适合在入读大学前的暑假阅读。
- **Michal Křížek, Florian Luca & Lawrence Somer, *17 Lectures on Fermat Numbers*** （Springer，2001）。关于费马数的一切，由欧拉的分解到高斯的正多边形，应有尽有。不妨翻翻开首几章。
- **Ian Stewart, *Why Beauty Is Truth*** （Basic Books），关于高斯的一章。以一般读者为对象，讲述正十七边形的故事。

1.2 质数分布得有多稀疏?

题目

STEP III 2016，第 5 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/16-s3-q5`

引子

在 Section 1.1 中，我们已看到质数（prime）有无穷多个。但“无穷多个”是个很粗略的说法。质数确实愈来愈稀疏：100 以下有 25 个质数，但 900 与 1000 之间只有 14 个。它们究竟会稀疏到甚么程度?

这道题要证明的结论听来相当平凡：把不超过 n 的所有质数相乘，所得的积小于 4^n。然而，这个看似平平无奇的不等式（inequality），正是回答“有多稀疏?”的第一步实质进展，也是数学史上最著名的简短证明之一的关键引理（lemma）；该证明出自年轻的保罗·艾狄胥（Paul Erdős）之手，时为 1932 年。

动笔之前

记号 $P_{r,s}$ 看来有点繁琐，但背后的意思很简单：把某个范围内的质数相乘。务必先弄懂题中的例子。$P_{3,7}$ 用的是大于 3 而不超过 7 的质数，即 5 和 7，所以它等于 35。

? **想一想** 第 (iv) 部分要求你对所有 $n \geq 2$ 证明一个命题。第 (iii) 部分则假设该命题对不超过 $2m$ 的每一个 k 都成立。这是哪一种归纳法（induction）?

这是强归纳法（strong induction）：要证明 n 的情况，你可以假设前面所有情况都成立，而不仅是紧接在前的那一个。请把这一点记在心里；题目之所以这样编排，原因正在于此。

逐步解题

阶段 1　二项式系数不会太大

想一想　在 $(1+x)^{2m+1}$ 的二项展开式（binomial expansion）中代入 $x=1$，会得到甚么？

你会得到

$$2^{2m+1} = \binom{2m+1}{0} + \binom{2m+1}{1} + \cdots + \binom{2m+1}{2m+1}.$$

各项都是正数，所以每一项都小于 2^{2m+1}。这还不够好；我们想要的是 2^{2m}，即它的一半。

想一想　帕斯卡三角形（Pascal's triangle）是对称的。和式中还有哪一项等于 $\binom{2m+1}{m}$？

由对称性，$\binom{2m+1}{m} = \binom{2m+1}{m+1}$，而这是和式中两个不同的项（因为 $m \neq m+1$）。所以

$$2\binom{2m+1}{m} = \binom{2m+1}{m} + \binom{2m+1}{m+1} < 2^{2m+1},$$

这里是严格不等式，因为和式中还有其他正项，例如 $\binom{2m+1}{0} = 1$。两边除以 2，即得 $\binom{2m+1}{m} < 2^{2m}$。 ■

阶段 2　哪些质数整除这个二项式系数？

把这个二项式系数（binomial coefficient）写出来：

$$\binom{2m+1}{m} = \frac{(2m+1)!}{m!\,(m+1)!} = \frac{(m+2)(m+3)\cdots(2m+1)}{m!}.$$

想一想　设 p 为质数，且 $m+1 < p \le 2m+1$。p 整除分子吗？它有可能整除分母吗？

分子 $(m+2)(m+3)\cdots(2m+1)$ 的因子当中就有 p 本身，所以 p 整除分子。分母 $m!$ 是若干个数的积，而这些数全都小于 p；由于 p 是质数，如果它不能整除其中任何一个数，就不能整除它们的积。因此 p 整除分子而不整除分母，于是 p 整除整个二项式系数（别忘记，它是一个整数）。

停一停，试一试　我们现在知道，这个范围内的每一个质数都整除 $\binom{2m+1}{m}$。为甚么由此可知它们的积 $P_{m+1,2m+1}$ 也整除它？（小心：即使 4 和 6 都整除某个数，24 也不一定整除它。）

分别在于，我们的质数是*相异质数*。如果相异质数 $p_1, p_2, \dots$ 各自整除 N，那么它们的积也整除 N：写 $N = p_1 N_1$；这时 p_2 整除 $p_1 N_1$ 但不整除 p_1，所以它整除 N_1；如此类推。（这里用到一个事实：如果一个质数整除某个积，它必整除其中一个因子。这个事实值得我们稍加敬重；Section 1.7 会展示一个世界，在那里与它相应的命题并不成立。）

所以 $P_{m+1,2m+1}$ 整除 $\binom{2m+1}{m}$，而正数的因数不会大于该数本身，故

$$P_{m+1,\,2m+1} \le \binom{2m+1}{m} < 2^{2m}. \qquad \blacksquare$$

阶段 3 把积拆开

想一想 $P_{1,2m+1}$ 是不超过 $2m+1$ 的所有质数的积，怎样用我们已经掌握的积把它拼出来?

把不超过 $2m+1$ 的质数分成两批：不超过 $m+1$ 的，以及大于 $m+1$ 的：

$$P_{1,2m+1} = P_{1,m+1} \times P_{m+1,2m+1}.$$

由第 2 阶段，第二个因子小于 $2^{2m} = 4^m$。至于第一个因子，我们可以假设 $P_{1,k} < 4^k$ 对由 2 至 $2m$ 的每一个 k 都成立，而 $k = m+1$ 正在这个范围之内（因为由 $m \ge 1$ 可得 $2 \le m+1 \le 2m$）。所以

$$P_{1,2m+1} < 4^{m+1} \times 4^m = 4^{2m+1}. \qquad \blacksquare$$

强归纳法正是在这里大派用场：要处理 $2m+1$，我们需要的是 $m+1$ 的情况，它的大小约为一半，而不是紧接在前的那个情况。

阶段 4 所有 n

想一想 第 (iii) 部分处理了奇数。对于偶数 $n = 2m+2$，情况又如何？它是质数吗?

先验证起始情况：$P_{1,2} = 2 < 16$ 及 $P_{1,3} = 6 < 64$。现在假设 $P_{1,k} < 4^k$ 对由 2 至 $n-1$ 的所有 k 都成立，其中 $n \ge 4$。

- 若 n 是偶数，则 n 不是质数（因为它大于 2），所以不超过 n 的质数与不超过 $n-1$ 的质数完全相同。因此 $P_{1,n} = P_{1,n-1} < 4^{n-1} < 4^n$。
- 若 n 是奇数，写 $n = 2m+1$，其中 $m \ge 2$。这时由第 (iii) 部分直接得出 $P_{1,n} < 4^n$。

由强归纳法，$P_{1,n} < 4^n$ 对每一个 $n \ge 2$ 都成立。 $\blacksquare$

回顾

- **二项式系数蕴藏算术信息。**$\binom{2m+1}{m}$ 是单独一个数，它不大（小于 4^m），却必须被许多质数整除。把许多质数挤进一个不大的数里，就说明这些质数不可能太多。
- **当问题会把自己减半时，强归纳法是最自然的工具。**$2m+1$ 的情况取决于 $m+1$ 的情况。
- **小的情况须另行验证。**归纳要到 $n \geq 4$ 才能开始运作，所以 $n = 2$ 和 3 必须亲手计算。

背后的数学

从 4^n 到质数定理

对结果取对数：$\ln P_{1,n} < n \ln 4$，即

$$\sum_{p \leq n} \ln p < (\ln 4)\, n \approx 1.39\, n.$$

在 1850 年代，切比雪夫（Pafnuty Chebyshev）证明了正是这一类的界（bound），而且上界和下界俱备，从而说明 $\sum_{p \leq n} \ln p$ 的增长与 n 成正比。高斯（Gauss）在大约 15 岁时，已从数表中猜到不超过 n 的质数个数大约是 $n / \ln n$。这后来成为*质数定理*（Prime Number Theorem），最终在 1896 年由阿达马（Jacques Hadamard）和德拉瓦莱普桑（Charles de la Vallée Poussin）运用复分析（complex analysis）证明。它的对数形式是

$$\sum_{p \leq n} \ln p \sim n,$$

所以不超过 n 的质数之积大约是 e^n。我们的界 4^n 其实相差不远：图 1.2.1 把三者一并画出。

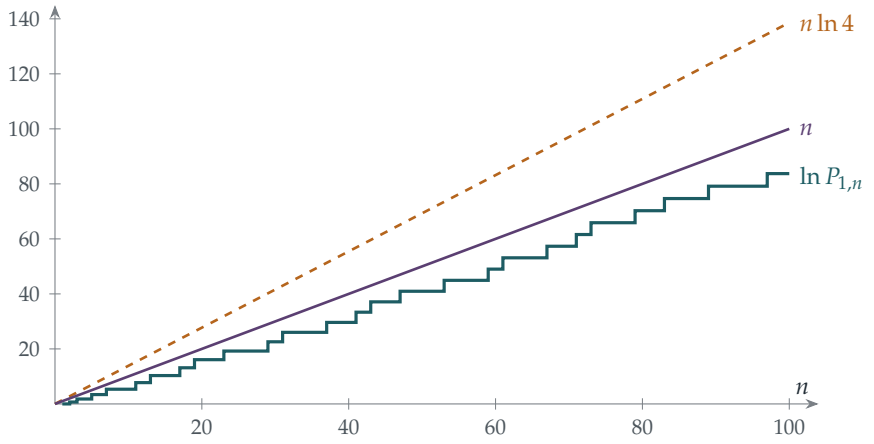


图 1.2.1 不超过 n 的所有质数之积的对数（蓝绿色），在每个质数 p 处跳升 $\ln p$。本节证明它始终低于 $n \ln 4$（琥珀色虚线）；质数定理则指出它的增长与 n 相若（紫红色）。

伯特兰公设

1845 年，伯特兰（Joseph Bertrand）猜想：对每一个 $n \geq 1$，在 n 与 $2n$ 之间总有一个质数。这就是伯特兰公设（Bertrand's postulate）。切比雪夫在 1852 年证明了它。八十年后，年仅 19 岁的艾狄胥找到了一个优美的初等证明。它的关键材料，就是本题的引理，再加上仔细考察哪些质数整除中间的二项式系数 $\binom{2n}{n}$。Martin Aigner 和 Günter Ziegler 把它选为 *Proofs from THE BOOK* 的第 2 章，该书汇集了数学中最优雅的证明。

接下来的方向

质数的精确分布，由黎曼 ζ 函数（Riemann zeta function）

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots,$$

的零点（zero）所支配。你会在 Section 2.4 中遇见这个函数朴素得多的一面，那里有 $\zeta(2) = \pi^2/6$。黎曼猜想（Riemann Hypothesis）预测了这些零点的确切位置；它若成立，便能把质数定理中的误差限定得紧无可紧。它也许是数学中最著名的未解难题，也是千禧年大奖难题（Millennium Prize Problems）之一。

思维技巧

强归纳法，以及从巧妙的对象中提取信息

如果 n 的情况取决于远小于 $n-1$ 的情况，就一次过假设前面所有情况都成立。

这里的另一个想法较为微妙，也极具数论（number theory）的特色：要证明某个范围内所有质数的某种性质，就找出一个对象（这里是一个二项式系数），使得这些质数每一个都必须整除它，而你又能证明它不会太大。

延伸练习

1 **感受一下数字。**对 $n = 10, 20, 30$ 计算 $P_{1,n}$，并把它与 4^n 及 e^n 比较。（这里可以使用计算机。）

2 **中间的二项式系数。**证明 $\binom{2m}{m}$ 是 $(1+1)^{2m}$ 展开式的 $2m+1$ 项中最大的一项，并由此推出 $\binom{2m}{m} \geq \dfrac{4^m}{2m+1}$。

3 **质数不会太多。**设 $\pi(n)$ 为不超过 n 的质数个数。介乎 $\sqrt{n}$ 与 n 之间的每个质数都大于 $\sqrt{n}$，而它们的积至多为 $P_{1,n} < 4^n$。由此推出

$$\pi(n) - \pi(\sqrt{n}) < \frac{2\ln 4}{\ln n}\, n,$$

从而得出 $\pi(n) < 2.8\, n/\ln n + \sqrt{n}$。把这个结果与质数定理的 $\pi(n) \approx n/\ln n$ 比较。

4 **验证伯特兰公设。**对 $n = 1, 2, \ldots, 50$ 验证伯特兰公设，即在每个情况下找出一个介乎 n 与 $2n$ 之间的质数。你能否利用一张简短的质数表有效率地完成，使表中每个质数都小于前一个的两倍？

延伸阅读

- **Aigner & Ziegler, *Proofs from THE BOOK***，第 2 章“Bertrand's postulate”（伯特兰公设）。
- **Marcus du Sautoy, *The Music of the Primes***（Fourth Estate）。一部讲述质数定理与黎曼猜想历史的普及读物。
- **G. J. O. Jameson, *The Prime Number Theorem***（Cambridge University Press）。严谨而平易近人，为大学本科生而写；开首几章只需要高级程度（A-level）的微积分。

1.3 最小反例

题目

STEP III 2015，第 5 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/15-s3-q5`

引子

你大概见过 $\sqrt{2}$ 是无理数（irrational）的证明：假设 $\sqrt{2} = a/b$ 为最简分数（lowest terms），两边平方，推出 a 是偶数，继而推出 b 也是偶数，于是得出矛盾（contradiction）。这是个漂亮的证明，但它依赖一个关于*偶数*的技巧，而这个技巧看不出可以怎样推广。换成 $\sqrt[3]{2}$ 又如何？$\sqrt{7}$ 呢？

本题给出另一个证明，它的核心想法简单得近乎荒谬：*看看最小的那一个*。在本章里，同一个想法还会改头换面地出现好几次。

动笔之前

这是一道不寻常的 STEP 题目：它给你一个证明的骨架，要你补上当中欠缺的理据。大学生正是这样学习阅读数学的。课本里的证明常常说“显然”或“由此可得”，而你的部分工作，就是查证事实是否真的如此。

想一想 第 4 步说“定义 k 为 S 中最小的正整数”。我们凭甚么可以这样做？假如 S 是一个由正*实*数组成的集合，这样做还可以吗？

任何由正整数组成的非空集合都有最小元素。这称为*良序原理*（well-ordering principle），它与数学归纳法（mathematical induction）原理等价。对于由正实数组成的集合，这个原理*并不成立*：集合 $\{x : x > 0\}$ 就没有最小元素。所以第 4 步既依赖第 3 步（集合必须非空），也依赖 S 由整数组成这一点。

逐步解题

阶段 1　第 3 步：S 非空

想一想　若 $\sqrt{2} = a/b$，其中 a 和 b 是正整数，你能写出 S 的一个元素吗？

若 $\sqrt{2} = a/b$，则 $b\sqrt{2} = a$，是一个整数。所以 b 属于 S，S 非空。

阶段 2　第 5 步：一个更小的元素

现在 k 是 S 的最小元素，所以 $k\sqrt{2}$ 是整数，记为 j。

停一停，试一试　要证明 $(\sqrt{2}-1)k$ 属于 S，你必须检查三件事。是哪三件？

我们需要 $(\sqrt{2}-1)k$ 满足：(a) 是整数；(b) 是正数；(c) 乘以 $\sqrt{2}$ 后得到整数。

- (a) $(\sqrt{2}-1)k = k\sqrt{2} - k = j - k$，是两个整数之差。
- (b) $\sqrt{2} > 1$，所以 $\sqrt{2} - 1 > 0$，因此 $(\sqrt{2}-1)k > 0$。
- (c) $(\sqrt{2}-1)k \cdot \sqrt{2} = 2k - k\sqrt{2} = 2k - j$，是整数。

所以 $(\sqrt{2}-1)k \in S$。

阶段 3　第 6 步：矛盾

想一想　$\sqrt{2} - 1$ 有多大？

由于 $1 < \sqrt{2} < 2$，我们有 $0 < \sqrt{2} - 1 < 1$，所以 $(\sqrt{2}-1)k < k$。我们找到了 S 的一个元素，它比 S 的最小元素 k 还要小。这是不可能的，所以第 1 步的假设不成立：$\sqrt{2}$ 是无理数。 ■

留意我们从头到尾都不需要“最简分数”，也不需要任何关于奇偶数的事实。整个证明只建基于三个事实：$\sqrt{2} \cdot \sqrt{2}$ 是整数；$\sqrt{2}$ 严格地介乎两个连续整数之间；整数具有良序性。

阶段 4　立方根：把两个数联系起来

设 $c = 2^{1/3}$，则 $c^2 = 2^{2/3}$ 且 $c^3 = 2$。

想一想　若 c 是有理数（rational），c^2 是否也是？反过来，若 c^2 是有理数，你能否只用乘法和除法回到 c？

若 c 是有理数，则 $c^2 = c \times c$ 也是有理数。反之，若 c^2 是有理数，则 $2/c^2 = c^3/c^2 = c$

也是有理数。所以 $2^{1/3}$ 是有理数，当且仅当 $2^{2/3}$ 是有理数。 ■

阶段 5 改造论证

想一想 在第 (i) 部分，关键在于 $(\sqrt{2})^2 = 2$ 是整数，所以把 S 的元素乘以 $\sqrt{2}$ 两次，就会回到整数。对于 $c = 2^{1/3}$，你需要乘三次。集合 S 应该怎样修改?

假设 c 是有理数，则 c^2 也是。定义

$$S = \{n \text{ 为正整数} : nc \text{ 和} nc^2 \text{ 都是整数}\}.$$

- S **非空。**若 $c = a/b$ 且 $c^2 = a'/b'$，则 bb' 属于 S。
- **设 k 为它的最小元素**，并考虑 $(c-1)k$。我们有 $1 < c < 2$（因为 $1^3 < 2 < 2^3$），所以 $0 < (c-1)k < k$。
- **它是整数：** $(c-1)k = kc - k$。
- **乘以 c：** $(c-1)k \cdot c = kc^2 - kc$，是整数。
- **乘以 c^2：** $(c-1)k \cdot c^2 = kc^3 - kc^2 = 2k - kc^2$，是整数。

所以 $(c-1)k \in S$，而且比 k 小：矛盾。因此 $2^{1/3}$ 是无理数，而根据第 4 阶段，$2^{2/3}$ 也是无理数。 ■

回顾

- **“取最小的那一个”是一种证明技巧。**假设反例存在，取其中最小的一个，再由它造出一个更小的。这称为*极端原理*（extremal principle），或称*最小反例*（minimal counterexample）论证。
- **集合 S 替我们记帐。**我们不必周旋于分子和分母之间，而是把所有“好”的乘数收进同一个集合。
- **推广，就是追问证明真正行得通的原因。**对于 $\sqrt{2}$，需要乘以 $\sqrt{2}$ 一次；对于 $\sqrt[3]{2}$，则需要乘两次。看出这一点，我们就清楚知道该怎样重新定义 S。

背后的数学

一个画得出来的证明

第 (i) 部分背后有一幅优美的图，出自 Stanley Tennenbaum。若 $\sqrt{2} = a/b$，则 $a^2 = 2b^2$：一个边长为 a 的正方形，面积等于两个边长为 b 的正方形面积之和。如图 1.3.1 所示，把两个边长为 b 的正方形放在边长为 a 的正方形的两个对角。它们重叠之处是一个小正方形，另外还有两个小正方形没有被盖住。由于面积相等，重叠部分必定等于两个未被盖住的正方形之和，于是得到一对更小而具有同样性质的整数。整数对不可能无止境地愈变愈小，所以这样的正方形并不存在。这对较小的整数正是 $(2b - a,\ a - b)$，也就是我们那一步 $(\sqrt{2} - 1)k$ 的化身。

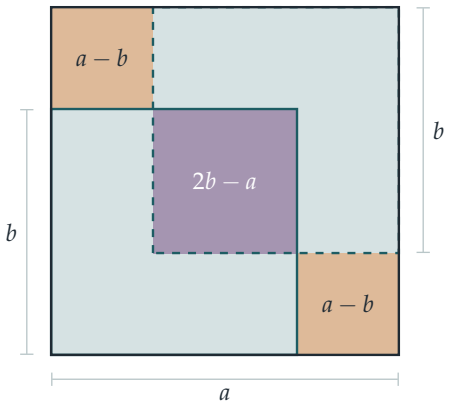


图 1.3.1　Tennenbaum 的证明。两个边长为 b 的正方形（蓝绿色）放在边长为 a 的正方形的两个对角。若 $a^2 = 2b^2$，则被计算了两次的面积（紫红色的重叠部分）必定等于没有计算到的面积（两个琥珀色的角落），所以 $(2b - a)^2 = 2(a - b)^2$：一个更小的解。

良序、归纳与递降

良序原理与归纳法原理在逻辑上等价：两者可以互相推导。到了大学，你会看到两者都被用作自然数的基础。在 Section 1.5 中，我们会以另一种形式运用同一个原理，那就是费马（Fermat）的*无穷递降法*（method of infinite descent）：不存在由正整数组成的无穷严格递减数列。

它能走多远？

这个论证可以证明，只要 N 不是完全平方数（perfect square），$\sqrt{N}$ 就是无理数（见练习）；更一般地，一个整数的 n 次方根要么是整数，要么是无理数。到了大学，这会变成一个关于*代数整数*（algebraic integers）的命题：若一个有理数满足某个整数系数的首一多项式（monic polynomial）方程，例如 $x^3 - 2 = 0$，它就必定是整数。你会在 Section 6.3 再次遇到这个想法。

相传毕达哥拉斯学派（Pythagoreans）因为希帕索斯（Hippasus）泄露了 $\sqrt{2}$ 是无理数而把他淹死，这个传说多半并不属实。属实的是，这个发现迫使希腊数学家把对数的研究与对长度的研究分开；这道裂痕要等到十九世纪，待戴德金（Dedekind）构造出实数，才得以弥合。

思维技巧

极端原理

如果你想证明某件事永远不会发生，就假设它会发生，选出最小（或最大、或最简单）的例子，再用它造出一个更小的例子。

决定“最小”是甚么意思，正是需要创意的一步。在这里，它是使 $n\sqrt{2}$ 成为整数的最小的 n。

延伸练习

1 **其他平方根。**改造这个论证，证明 $\sqrt{3}$ 是无理数。哪一个数担当了 $\sqrt{2}-1$ 的角色？

2 **所有非平方数。**设 N 为一个不是完全平方数的正整数，并设 m 为满足 $m < \sqrt{N} < m+1$ 的整数。考虑 $(\sqrt{N} - m)k$，从而证明 $\sqrt{N}$ 是无理数。

3 **找出漏洞。**试把这个论证套用于 $\sqrt{4}$。究竟是哪一步行不通？

4 **Tennenbaum 的图。**利用图 1.3.1 证明：若 $a^2 = 2b^2$，则 $(2b-a)^2 = 2(a-b)^2$。再用代数方法验证这个恒等式。

延伸阅读

- **Daniel Velleman, *How to Prove It***（Cambridge University Press）。一本全面的证明入门书，内容包括归纳法和良序原理。

- **R. P. Burn, *A Pathway into Number Theory***（Cambridge University Press）。全书由精心编排次序的问题组成，是修读大学一年级课程前的理想准备。
- **John Conway & Joseph Shipman, "Extreme proofs I: the irrationality of $\sqrt{2}$"**（*Mathematical Intelligencer*，2013）。同一个事实的多个证明，比较得妙趣横生。

1.4 e 是无理数，e^2 也是

本节会用到两道题目。第一道出自 1997 年，证明 e 是无理数（irrational number）。第二道出自 2024 年，把同一个想法用得更巧妙，证明 e^2 是无理数，而这原来确实困难得多。

热身题

STEP III 1997，第 7 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/97-s3-q7`

主题目

STEP III 2024，第 7 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2024 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

`www.ocr.org.uk/Images/725753-2024-past-papers.zip`

引子

在 Section 1.3 中，我们证明了 $\sqrt{2}$ 是无理数，用的是 $\sqrt{2}$ 满足一条简单方程 $x^2 = 2$ 这个事实。但就我们所知，e 并不满足任何这样的方程。我们手上只有它的级数（series）：

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots .$$

一个级数怎能告诉你某个数不是分数?

答案是数论（number theory）中最有用的想法之一：*整数（integer）不可能严格介乎 0 与 1 之间*。如果你能证明某个量是整数，同时又严格介乎 0 与 1 之间，就得出矛盾（contradiction）。本节的一切，都是这个主题的变奏。

动笔之前

> **想一想**　假设 $e = p/q$。把 e 乘以 q 会得到整数。那么乘以 $q!$ 又会对 e 的级数的每一项有甚么影响?

只要 $k \le q$，把 $\dfrac{1}{k!}$ 乘以 $q!$ 都会得到整数。因此 $q!\,e$ 可分成整数部分（首 $q+1$ 项）和“尾部”（tail，即其后的所有项）。若 $e = p/q$，则 $q!\,e$ 本身是整数，所以尾部也必定是整数。我们的计划，就是证明尾部严格介乎 0 与 1 之间。

逐步解题

阶段 1　热身题：e 是无理数

级数 b_n 是等比级数（geometric series），首项和公比都是 $\frac{1}{n+1}$，所以

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

> **想一想**　逐项比较 a_n 和 b_n。哪一个较大?

两者的首项相等。其后，$(n+1)(n+2) > (n+1)^2$，而一般来说，当 $k \ge 2$ 时 $(n+1)(n+2)\cdots(n+k) > (n+1)^k$，所以 a_n 其后的每一项都小于 b_n 的对应项。因此 $0 < a_n < b_n = 1/n$。

现在看看 a_n 究竟是甚么。把 e 的级数乘以 $n!$：

$$n!\,e = \underbrace{\frac{n!}{0!} + \frac{n!}{1!} + \cdots + \frac{n!}{n!}}_{\text{整数}} + \underbrace{\frac{n!}{(n+1)!} + \frac{n!}{(n+2)!} + \cdots}_{= a_n}.$$

由于 $0 < a_n < 1$，$n!\,e$ 的整数部分就是第一个括号，而 a_n 就是小数部分（fractional part）：$a_n = n!\,e - [n!\,e]$。

> **停一停，试一试**　现在完成证明。假设 $e = p/q$。你应该选取哪个 n 值?

取 $n = q$。那么 $q!\,e = (q-1)!\,p$ 是整数，所以它的小数部分 a_q 是 0。但 $a_q > 0$。矛盾：e 是无理数。 ∎

这个证明出自傅里叶（Joseph Fourier），约于 1815 年提出。

阶段 2　为何 e^2 较难

想一想 试把同一个想法用于 e^2。它的级数是 $\sum 2^k/k!$。当你乘以 $n!$ 再看尾部时，哪里出了问题?

尾部现在是 $\dfrac{2^{n+1}}{n+1}+\dfrac{2^{n+2}}{(n+1)(n+2)}+\cdots$，而 2 的幂把一切都破坏了：尾部不再细小。乘以 $n!$ 对分子中的 2^k 毫无帮助。

2024 年这道题的妙着，是完全避开 e^2。若 $e^2=p/q$，则

$$qe=\frac{p}{e}.$$

现在我们有两个数 e 和 $1/e$，各自都有一个阶乘（factorial）级数，而且没有麻烦的 2 的幂：

$$e=\sum_{k=0}^{\infty}\frac{1}{k!},\qquad \frac{1}{e}=\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}.$$

阶段 3 两个尾部

第 (i) 部分正正就是热身题：$0<f(n)<1/n$。

至于第 (ii) 部分，$g(n)$ 是一个各项的大小递减的交错级数（alternating series）。

想一想 对于交错级数 $t_1-t_2+t_3-\cdots$，其中 $t_1>t_2>t_3>\cdots>0$，为何它的和介乎 0 与 t_1 之间?

把各项两两配对：$(t_1-t_2)+(t_3-t_4)+\cdots$ 是若干个正括号之和，所以是正数。换另一种方式配对：$t_1-(t_2-t_3)-(t_4-t_5)-\cdots$ 是 t_1 减去若干个正括号，所以小于 t_1。这里 $t_1=\frac{1}{n+1}$,而各项递减,因为每一项都是前一项除以一个大于 1 的数。所以 $0<g(n)<\frac{1}{n+1}$。

阶段 4 又见整数

把 e 乘以 $(2n)!$，然后分出尾部，做法与热身题完全一样：

$$(2n)!\,e=(\text{整数})+f(2n).$$

所以 $(2n)!\,e-f(2n)$ 是整数。现在对 $1/e$ 做同样的事：

$$\frac{(2n)!}{e}=\sum_{k=0}^{2n}(-1)^k\frac{(2n)!}{k!}+\sum_{k=2n+1}^{\infty}(-1)^k\frac{(2n)!}{k!}.$$

想一想 尾部的首项（即 $k=2n+1$ 的一项）是甚么符号？尾部与 $g(2n)$ 相比又如何?

尾部首项的 $k = 2n+1$ 是奇数，所以该项是负数。尾部是

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

所以 $\dfrac{(2n)!}{e} + g(2n)$ 是整数。 ■

这就是题目用 $2n$ 而不用 n 的原因：乘数是偶数的阶乘时，$1/e$ 的尾部恰好是 $-g$，符号可以预知。

阶段 5　矛盾

假设 $qe = p/e$。把等式两边乘以 $(2n)!$，并利用第 4 步：

$$q\big(\text{整数} + f(2n)\big) = p\big(\text{整数} - g(2n)\big).$$

移项整理后，$qf(2n) + pg(2n)$ 等于若干整数的组合，所以它是整数。这就是第 (iv) 部分。

> **停一停，试一试**　现在利用第 3 步的上下界。当 n 非常大时，$qf(2n) + pg(2n)$ 会怎样？

$f(2n)$ 和 $g(2n)$ 都是正数，所以 $qf(2n) + pg(2n) > 0$。而且

$$qf(2n) + pg(2n) < \frac{q}{2n} + \frac{p}{2n+1} < \frac{p+q}{2n}.$$

选取 $n > (p+q)/2$。那么 $qf(2n) + pg(2n)$ 就是一个严格介乎 0 与 1 之间的整数。这是不可能的，所以 e^2 是无理数。 ■

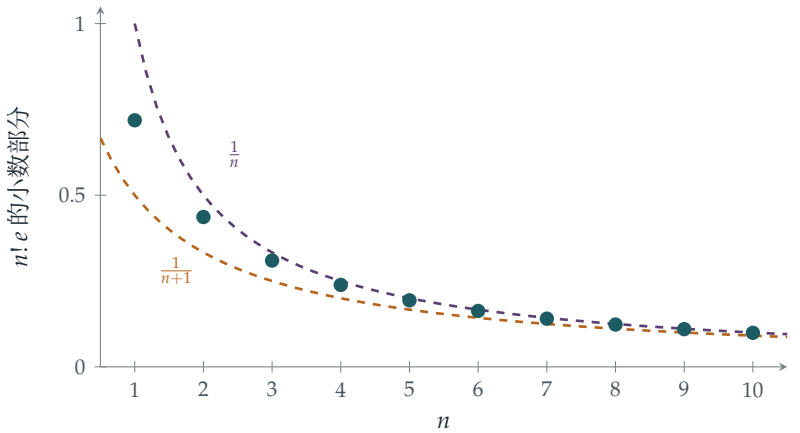


图 1.4.1　尾部 $a_n = n!\,e - [n!\,e]$（青绿色），即 $n!\,e$ 的小数部分。它总是严格介乎 $\frac{1}{n+1}$ 与 $\frac{1}{n}$ 之间：永不为零，但小于 1。假如 e 是分数 $\frac{p}{q}$，那么对每个 $n \geq q$，$n!\,e$ 都会是整数，青绿色的点就会是 0。

回顾

- **推动整个证明的是“没有整数严格介乎 0 与 1 之间”。**两个证明都构造出一个必定是整数的量，然后把它困住。
- **阶乘令尾部变小。**乘以 $n!$ 把级数的开首部分变成整数，剩下一个像 $1/n$ 那样缩小的尾部。
- **正面进攻失败时，就重新整理。**我们没有正面处理 $e^2 = p/q$，而是把它改写成 $qe = p/e$，把困难分摊到两个较易处理的数上。
- **细节很重要。**选用 $2n$ 而非 n，固定了尾部的符号。许多证明的成败，都取决于这样的一个小细节。

背后的数学

e 有多“无理”？

刘维尔（Joseph Liouville）在 1840 年证明了 e^2 是无理数，而且事实上它不满足任何整数系数的二次方程。1873 年，埃尔米特（Charles Hermite）走得远得多：他证明了 e 是*超越数*（transcendental number），意思是它不是任何整数系数多项式的根。他的证明正是上述论证的精密版本：构造一个整数，再证明它严格介乎 0 与 1 之间。九年后，林德曼（Ferdinand von Lindemann）改造了埃尔米特的方法，证明 π 是超越数，终于证实以尺规“化圆为方”这个古老问题是不可能做到的。

我们仍未知道的事

尽管这些方法威力强大，一些简单得令人尴尬的问题却仍未解决。$e + \pi$ 是无理数吗？$e\pi$ 呢？没有人知道，虽然两者之中至少有一个必定是超越数（练习 4）。$\zeta(3) = 1 + \frac{1}{8} + \frac{1}{27} + \cdots$ 是无理数吗？是的：阿培里（Roger Apéry）在 1978 年证明了这一点，震惊了整个数学界。$\zeta(5)$ 呢？没有人知道。

逼近

这些证明背后有一个更深刻的想法：如果一个数能被分数逼近得*太好*，它就是无理数。e 的级数的部分和（partial sum）都是分数，而且很快就非常接近 e；一个分数不可能被其他分数逼近得那么好。这就是*丢番图逼近*（Diophantine approximation）的起点。它是数论的一个分支，你可以在 Section 2.9 进一步探索，那里会用另一族积分得出更好的分数，步步逼近 e。

思维技巧

困住一个整数

要证明一个数是无理数，先假设它是分数，由此构造一个必定是整数的量，再证明这个量严格介乎 0 与 1 之间。

个中的艺术在于选择这个量。在这里，乘以阶乘把级数的大部分变成整数，剩下一个细小但为正数的尾部。

延伸练习

1 **只用 $1/e$。**假设 $1/e = p/q$，并只利用 $1/e$ 的交错级数，证明 e 是无理数。

2 $\sin 1$ **是无理数。**利用 $\sin 1 = 1 - \dfrac{1}{3!} + \dfrac{1}{5!} - \cdots$ 证明 $\sin 1$ 是无理数。（乘以一个合适的奇数的阶乘。）

3 **为何 π 不行？**解释为何这个方法不能直接用于 π。然后查阅 Ivan Niven 在 1947 年发表、只有一页的 π 是无理数的证明，当中用的是积分而不是级数。能力强的高级程度（A-level）学生应可读懂。

4 **其中一个是超越数。**较难 证明 e 和 π 是 $x^2 - (e+\pi)x + e\pi = 0$ 的根。你可以假设：若一个多项式的系数都是代数数（algebraic number），则它的根本身也是代数数；以及 π 不是代数数（林德曼）。由此推出 $e+\pi$ 和 $e\pi$ 之中至少有一个是超越数。

延伸阅读

- **Aigner & Ziegler, *Proofs from THE BOOK***，第 8 章“Some irrational numbers”（一些无理数）。正是用这些想法，证明 e、e^2、e^4 和 π^2 都是无理数。
- **Ivan Niven, *Irrational Numbers***（Mathematical Association of America）。一本篇幅不长的经典，内容由 $\sqrt{2}$ 一直谈到超越性。
- **Julian Havil, *The Irrationals***（Princeton University Press）。一本浅白易读的无理数历史。

1.5 无穷递降法

题目

STEP III 2022，第 2 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2022 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

`www.ocr.org.uk/Images/696587-2022-past-papers.zip`

引子

怎样才能证明一条方程*没有*整数解？你不可能逐一试遍所有可能，因为可能性有无穷多个。费马（Pierre de Fermat）在 1630 年代找到了一个方法：证明任何一个解都会产生另一个数字*更小*的解，然后又产生一个更小的，如此无休止地继续下去。可是正整数不可能永无止境地愈变愈小。所以，从一开始就根本没有解。

费马把这个方法称为*无穷递降法*（method of infinite descent），并且深以为傲。它是 Section 1.3 中良序原理（well-ordering principle）的另一面，至今仍是证明方程无解的主要工具之一。

动笔之前

四个部分的格局都一样：假设有解，把解缩小，得出矛盾（contradiction）。问题在于*怎样*缩小。

> **想一想** 在第 (i) 部分，看看 $a^3 + 2b^3 + 4c^3 = 0$。三项之中有两项明显是偶数。这告诉你第三项有甚么性质？

每一次递降都是由这类观察开始的：某种整除性（divisibility）是你无法回避的。

逐步解题

阶段 1　第一次递降

由于 $a^3 = -2b^3 - 4c^3$，a^3 这个数是偶数，所以 a 是偶数（奇数的立方是奇数）。设 $a = 2p$。那么 $|p| = |a|/2 < |a|$，代入后得

$$8p^3 + 2b^3 + 4c^3 = 0 \quad \Longrightarrow \quad 4p^3 + b^3 + 2c^3 = 0.$$

想一想 新的方程是 $4p^3 + b^3 + 2c^3 = 0$。它的系数和原来一样，只是轮换了位置。你可以再用同一招吗？

现在 $b^3 = -4p^3 - 2c^3$ 是偶数，所以 $b = 2q$，其中 $|q| < |b|$，再除以 2 便得 $2p^3 + 4q^3 + c^3 = 0$。同样地，c 必定是偶数：$c = 2r$，其中 $|r| < |c|$，再除以 2，

$$p^3 + 2q^3 + 4r^3 = 0.$$

这正是原来的方程，只是以 p、q、r 取代了 a、b、c。每个数的大小都减半，而且没有一个是零，因为把非零的数减半不可能得到零。

停一停，试一试 把以上论证写成一个证明（proof），说明非零解并不存在。你究竟应该假设甚么是“最小”的？

在所有非零整数解之中，选取 $|a|$ 最小的一个。（只要有解存在，良序原理就容许我们这样做。）上述论证会产生另一个非零解，并且 $|p| < |a|$，与最小性矛盾。所以没有非零解。 ■

另一个等价的说法是：由一个解出发，我们可以构造出无穷多个解，它们满足 $|a| > |p| > \cdots$，即一个严格递减的无穷正整数数列，而这是不可能的。

阶段 2　换一个质数

想一想 在 $9a^3 + 10b^3 + 6c^3 = 0$ 中，奇偶性（parity）看来帮不上忙。看看系数 9、10、6。哪一个质数（prime）整除其中两个？

质数 3 整除 9 和 6，所以 3 整除 $10b^3$，因此 3 整除 b。设 $b = 3q$ 并除以 3：

$$3a^3 + 90q^3 + 2c^3 = 0.$$

现在 3 整除 $2c^3$，所以 $c = 3r$；再除以 3 得 $a^3 + 30q^3 + 18r^3 = 0$。于是 3 整除 a：$a = 3p$，再除以 3 一次，

$$9p^3 + 10q^3 + 6r^3 = 0.$$

这与原来的方程完全相同，只是换成了较小的非零数。根据同样的递降论证，方程没有非零解。 ■

阶段 3　模 3 下的平方数

想一想 每个整数都可写成 $3n$、$3n+1$ 或 $3n-1$ 的形式。它们的平方模（modulo）3 是甚么？

$(3n\pm1)^2 = 9n^2 \pm 6n + 1 = 3(3n^2 \pm 2n) + 1$，比 3 的倍数多一。而 $(3n)^2$ 是 3 的倍数。所以平方数除以 3 的余数（remainder）是 0 或 1，永远不会是 2。

现在如果 $a^2 + b^2$ 是 3 的倍数，余数的可能组合是 $0+0$、$0+1$、$1+0$ 和 $1+1$，分别得出 0、1、1、2。只有第一种是 3 的倍数，所以 a 和 b 都必定是 3 的倍数。

想一想 如果 $a^2 + b^2 = 3c^2$，而 $a = 3p$、$b = 3q$，你对 c 有甚么结论？

我们得到 $9p^2 + 9q^2 = 3c^2$，所以 $c^2 = 3(p^2 + q^2)$ 是 3 的倍数，因此 $c = 3r$。于是 $9p^2 + 9q^2 = 27r^2$，即 $p^2 + q^2 = 3r^2$：同一条方程，较小的数。又是递降。 ■

阶段 4 变奏：方程变了

想一想 在 $a^2 + b^2 + c^2 = 4abc$ 中，右边是 4 的倍数。平方数模 4 是甚么？

偶数的平方是 4 的倍数，奇数的平方则比 4 的倍数多一（$(2n+1)^2 = 4n^2 + 4n + 1$）。所以 $a^2 + b^2 + c^2$ 除以 4 的余数是 0、1、2 或 3，视乎 a、b、c 之中有多少个是奇数。由于右边是 4 的倍数，它们没有一个可以是奇数。设 $a = 2p$、$b = 2q$、$c = 2r$：

$$4(p^2 + q^2 + r^2) = 32\,pqr \quad\Longrightarrow\quad p^2 + q^2 + r^2 = 8\,pqr.$$

想一想 这并不是同一条方程：4 变成了 8。递降法还行得通吗？

窍门在这里：*证明一个更强的命题*。考虑所有以下形式的方程

$$x^2 + y^2 + z^2 = 2^m\,xyz \qquad \text{其中} m \geq 2.$$

对任何这样的 m，右边都是 4 的倍数，所以完全相同的论证显示 x、y、z 全是偶数。把它们减半，便得到含 2^{m+1} 的方程的一个解，而这条方程仍属这种形式。因此，由 m 方程的一个非零解，我们得到 $m+1$ 方程的一个较小的非零解。在这个方程族中任何一条方程的所有解之中，取 $|x|$ 最小的一个，便得出矛盾。特别地，$a^2 + b^2 + c^2 = 4abc$ 没有非零解。 ■

有时候，你想证明的命题太弱，不足以支撑它自身的归纳（induction），补救的方法是证明一个更一般、*能够自我支撑*的命题。这个技巧在大学里出奇地常见。

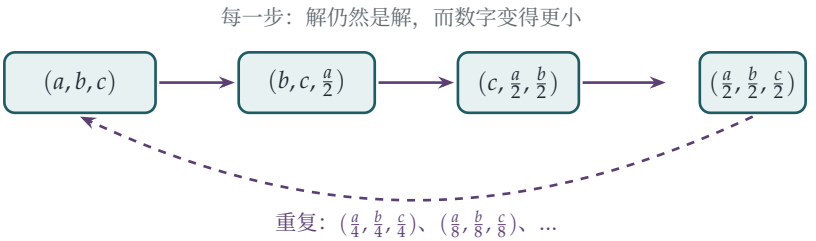


图 1.5.1　第 (i) 部分的递降。由 $a^3+2b^3+4c^3=0$ 的一个解出发，三步之后得到另一个解，每个分量都减半。无休止地重复下去，便会得出一个没有尽头的递减正整数数列 $|a|>|\frac{a}{2}|>|\frac{a}{4}|>\cdots$，而这是不可能的。所以一开始就没有解。

回顾

- **递降需要整除性的来源。**每一次，我们都找到一个质数 p（这里是 2 或 3），它必然整除所有变量，然后把它除去。
- **模 n 运算正是找出它的方法。**考察模 2、3 或 4 的余数，把“有没有任何可能行得通?”变成一个简短的有限验证。
- **如果方程变了，就推广。**第 (iv) 部分需要一整族方程，而这一族方程在递降步骤下是封闭的。

背后的数学

费马、递降法与他的最后定理

费马用无穷递降法证明了 $x^4+y^4=z^2$ 没有正整数解。在他著名的“最后定理”之中，这是他唯一留下证明的情况，它解决了 $x^n+y^n=z^n$ 在 $n=4$ 时的情形。欧拉（Euler）其后用递降法处理 $n=3$ 的情形，当中有一个漏洞，后来才被补上。递降法至今仍是现代理论的起点：关于 Section 1.9 中那类曲线的莫德尔–韦伊定理（Mordell–Weil theorem），其证明就是一个精密的递降。

局部障碍

本节的每一个证明，都是靠考察模某个数的余数来完成的。如果一条方程对某个 n 而言模 n 无解，或者没有实数解，它当然没有整数解。这类理由称为*局部障碍*（local obstructions）；到了大学，你会遇到*p 进数*（p-adic numbers），那是一种同时研究模 $p, p^2, p^3, \ldots$ 所有余数的方法。

哈塞（Hasse）与闵可夫斯基（Minkowski）有一个精彩的定理：对于像 $a^2+b^2=3c^2$ 这样的二次方程，局部障碍就是仅有的障碍——如果没有任何“局部”的理由令它无解，它就确实有有理数解。对三次方程而言，这个结论却彻底失效。Ernst Selmer 在 1951 年证明了

$$3x^3+4y^3+5z^3=0$$

模任何整数都有解，在实数中也有解，但没有非零整数解。解释“局部到整体”原理何时失效、为何失效，是现代数论的一大主题。

一条确实有解的方程

第 (iv) 部分有一个著名的近亲：$x^2+y^2+z^2=3xyz$，即*马尔可夫方程*（Markov equation）。它有无穷多个正整数解，由 $(1,1,1)$、$(1,1,2)$、$(1,2,5)$、$(1,5,13)$ 和 $(2,5,29)$ 开始，而这些解与无理数能被分数逼近到甚么程度有关。见练习 4。

思维技巧

无穷递降法

要证明一条方程无解，可证明任何一个解都会产生一个严格更小的解。这样，最小的解便不可能存在。

把它与模 n 运算配合使用，找出推动递降的整除性；如果命题在递降步骤下不能保持，就要准备把命题推广。

延伸练习

1 **STEP I 热身题。**（*STEP I 2010 第 8 题。*）证明 $a^3+3b^3=9c^3$ 的唯一整数解是 $a=b=c=0$。然后，藉考虑末位数字，证明 $p^4+2q^4=5r^4$ 也有同样的结论。

2 **只用模运算，毋须递降。**只用模 4 运算，证明 $x^2+y^2=4z+3$ 完全没有整数解。

3 **有理点。**利用第 (iii) 部分，证明圆 $x^2+y^2=3$ 上没有坐标为*有理数*的点。（试与 $x^2+y^2=2$ 对比，后者包含 $(1,1)$。）

4 **马尔可夫三元组。**证明若 (x,y,z) 满足 $x^2+y^2+z^2=3xyz$，则 $(x,y,3xy-z)$ 亦满足。由 $(1,1,1)$ 出发，生成上文列出的三元组。（这个“跳跃”技巧正好与递降相反。）

延伸阅读

- **Harold Davenport, *The Higher Arithmetic***，第 2 及第 5 章。
- **Joseph Silverman, *A Friendly Introduction to Number Theory***（Pearson）。非常易读；包括费马对 $x^4 + y^4 = z^2$ 的递降证明。
- **Simon Singh, *Fermat's Last Theorem***（Fourth Estate）。为一般读者而写的故事，由费马的页边笔记讲到怀尔斯（Wiles）的证明。

1.6 两个平方数之和

题目

STEP II 2023，第 7 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2023 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

`www.ocr.org.uk/Images/707102-2023-past-papers.zip`

引子

哪些整数可以写成两个平方数之和？$5 = 1^2 + 2^2$ 和 $13 = 2^2 + 3^2$ 可以，但 3、7 和 21 却不行。这里有一个奇怪的事实：$65 = 5 \times 13$ 可以用*两种*不同的方法写成两个平方数之和，即 $1^2 + 8^2$ 和 $4^2 + 7^2$。为甚么两个这样的数相乘，会得出另一个这样的数？又为甚么会有两种写法？

答案来自一个意想不到的方向。这道题目要你放下整数，改用*复数*（complex number）来思考。在复数的世界里，两平方和变得很简单，一些看似令人摸不着头脑的算术也变得近乎机械化。

动笔之前

? **想一想** 若 $z = a + \mathrm{i}b$，其中 a 和 b 为整数，那么 $|z|^2$ 是甚么？

它是 $a^2 + b^2$，即两个平方数之和。所以“N 是两个平方数之和”的意思正正就是“存在某个实部（real part）和虚部（imaginary part）皆为整数的复数 z，使得 $N = |z|^2$”。这样的复数称为*高斯整数*（Gaussian integer）。请牢记这个转译；整道题目都建基于此。

逐步解题

阶段 1 模相乘

想一想　以 a、b、c、d 写出 zw。关于 $|zw|^2$，你需要证明甚么？

我们有 $zw = (ac - bd) + \mathrm{i}(ad + bc)$，所以

$$|zw|^2 = (ac - bd)^2 + (ad + bc)^2 = a^2c^2 + b^2d^2 + a^2d^2 + b^2c^2,$$

因为交叉项 $-2abcd$ 和 $+2abcd$ 互相抵消。而这个式子可以因式分解（factorise）：

$$a^2c^2 + b^2d^2 + a^2d^2 + b^2c^2 = (a^2 + b^2)(c^2 + d^2) = |z|^2|w|^2.$$

取平方根，得 $|zw| = |z||w|$。 ■

现在把它看成一个关于整数的命题：

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2. \qquad (\star)$$

两个两平方和的乘积，仍是一个两平方和，而复数更确切地告诉你那两个平方数是甚么。

阶段 2　学以致用

第 (ii) 部分：$5 = |2 + \mathrm{i}|^2$，而 $221 = 10^2 + 11^2 = |10 + 11\mathrm{i}|^2$。于是相乘：

$$(2 + \mathrm{i})(10 + 11\mathrm{i}) = 20 + 22\mathrm{i} + 10\mathrm{i} - 11 = 9 + 32\mathrm{i},$$

得出 $5 \times 221 = 9^2 + 32^2$。（验算：$81 + 1024 = 1105$。）

想一想　如果改为把 $10 + 11\mathrm{i}$ 乘以 $2 - \mathrm{i}$，会怎么样？它的模（modulus）与 $2 + \mathrm{i}$ 的相同。

$(2 - \mathrm{i})(10 + 11\mathrm{i}) = 20 + 22\mathrm{i} - 10\mathrm{i} + 11 = 31 + 12\mathrm{i}$，所以同样有 $1105 = 31^2 + 12^2$。选用 $2 + \mathrm{i}$ 还是它的共轭（conjugate），会得出不同的表示法。这就解释了引子中 65 的两种写法（图 1.6.1）：$65 = |(2 + \mathrm{i})(3 + 2\mathrm{i})|^2 = |4 + 7\mathrm{i}|^2$ 以及 $65 = |(2 - \mathrm{i})(3 + 2\mathrm{i})|^2 = |8 + \mathrm{i}|^2$。

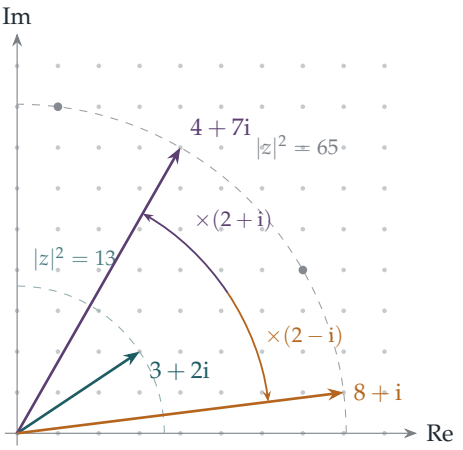


图 1.6.1 乘以 $2 \pm \mathrm{i}$ 会把长度放大 $\sqrt{5}$ 倍，并旋转约 26.6°：$2+\mathrm{i}$ 为逆时针方向，$2-\mathrm{i}$ 为顺时针方向。由圆 $|z|^2 = 13$ 上的 $3+2\mathrm{i}$ 出发，两种选择会落在圆 $|z|^2 = 65$ 上不同的格点（lattice point）。灰色的点是它们对直线 $y = x$ 的反射。

阶段 3 先作因式分解

? **想一想** 8045 的末位数字是 5。这提示了甚么?

$8045 = 5 \times 1609$。我们需要把 1609 写成两个平方数之和；试试略小于 1609 的平方数，很快便找到 $1609 = 40^2 + 3^2$。于是

$$(2+\mathrm{i})(40+3\mathrm{i}) = 77 + 46\mathrm{i}, \qquad \text{所以} \qquad 8045 = 77^2 + 46^2,$$

而共轭亦给出 $8045 = 83^2 + 34^2$。

第 (iv) 部分更快：$36 = 6^2 + 0^2$，而乘以 6 只是按比例放大：$36 \times 50805 = 612^2 + 1206^2$。

阶段 4 三种写法

? **想一想** 25 可以用多于一种方法写成两个平方数之和。若不计正负号和次序对调，有多少个高斯整数的模为 5?

$25 = 5^2 + 0^2 = 3^2 + 4^2$，分别对应 5、$3+4\mathrm{i}$ 和 $3-4\mathrm{i}$。接着看 1002082：由于 $1001^2 = 1002001$，我们有 $1002082 = 1001^2 + 9^2$。把 $1001 + 9\mathrm{i}$ 分别乘以这三个数，得：

$$\begin{aligned} 5\,(1001+9\mathrm{i}) &= 5005 + 45\mathrm{i}, \\ (3+4\mathrm{i})(1001+9\mathrm{i}) &= 2967 + 4031\mathrm{i}, \\ (3-4\mathrm{i})(1001+9\mathrm{i}) &= 3039 - 3977\mathrm{i}. \end{aligned}$$

所以三对 (r,s) 是 $(45,5005)$、$(2967,4031)$ 和 $(3039,3977)$。

阶段 5　除法

> **停一停，试一试**　第 (vi) 部分把这部机器倒过来运转。我们知道 $109 = 10^2 + 3^2$ 和 $1002037 = 1001^2 + 6^2$。怎样才能不靠搜索，就把 9193 写成两个平方数之和？

既然高斯整数相乘时模会相乘，那么相除时模就应该相除。试试

$$\frac{1001 + 6\mathrm{i}}{10 + 3\mathrm{i}} = \frac{(1001 + 6\mathrm{i})(10 - 3\mathrm{i})}{109} = \frac{10028 - 2943\mathrm{i}}{109} = 92 - 27\mathrm{i}.$$

刚好整除，所以 $9193 = 92^2 + 27^2$。（验算：$8464 + 729 = 9193$。）

> **想一想**　如果改为除以 $10 - 3\mathrm{i}$，会怎么样？

你会得到 $(1001 + 6\mathrm{i})(10 + 3\mathrm{i})/109 = (9992 + 3063\mathrm{i})/109$，这不是高斯整数，因为 9992/109 不是整数。在 109 的两个“因数”之中，只有一个能整除 $1001 + 6\mathrm{i}$，正如只有某些质数（prime）能整除一个给定的整数。高斯整数有它们自己的质数，也有它们自己的因式分解。

回顾

- **转译、求解、再译回来。**“两个平方数之和”变成了“高斯整数的模的平方”。用这种语言来说，问题就变得容易。
- **先作因式分解。**每一部分都是先把一个大数拆成较小的数，而这些较小的数的表示法很容易找到。
- **共轭带来新答案。**表示法的数目，取决于有多少种方法选择一个因数或它的共轭。

背后的数学

费马二平方和定理

哪些质数是两个平方数之和？$2 = 1^2 + 1^2$、5、13、17、29、37、41 是；3、7、11、19、23、31 则不是。规律是：除 2 以外，一个质数是两个平方数之和，当且仅当它除以 4 的余数为 1。费马（Fermat）于 1640 年提出这个结果；欧拉（Euler）于 1749 年给出证明。用高斯整数的语言来说：质数 $p \equiv 1 \pmod 4$ 会分裂（split）成 $p = (a + b\mathrm{i})(a - b\mathrm{i})$，而质数 $p \equiv 3 \pmod 4$ 则仍然是质数。例如 $5 = (2 + \mathrm{i})(2 - \mathrm{i})$，但 3 完全无法分解。

1990 年，Don Zagier 发表了二平方和定理的一个证明，它以全长只有一句而闻名。这

个证明连同解说收录于 *Proofs from THE BOOK*。

一个拥有自己质数的新数系

高斯（Gauss）发现，高斯整数的表现与普通整数极为相似。它们有除法算法（division algorithm）和欧几里得算法（Euclidean algorithm），而且每个高斯整数都可以唯一地分解成高斯质数（Gaussian prime）的乘积（不计乘以 ± 1 或 $\pm i$ 的差别）。结合 $(\star)$，你只需知道一个数的质因数分解，就能准确数出它有多少种方法写成两个平方数之和。

在大学里，这是通往代数数论（algebraic number theory）的大门：研究 $\mathbb{Z}[i]$ 或 $\mathbb{Z}[\sqrt{-5}]$ 这类数系，借此了解普通的整数。Section 1.7 说明了为甚么需要创立这门学科：并非每个这样的数系都像高斯整数那样规矩。

四个平方数及其他

每个正整数都是四个平方数之和（拉格朗日（Lagrange），1770 年）。证明依赖 $(\star)$ 的四平方版本，它源自哈密顿（Hamilton）的四元数（quaternion），情形就如 $(\star)$ 源自复数一样。你会在 Section 7.8 见到它。

思维技巧

> **转换场景**
>
> 一个问题在某个数系中显得棘手，在一个更大的数系中却可能十分自然。
>
> 两平方和在整数中很棘手，在复数中却不过是模而已。数学家经常这样扩大数系：由整数到有理数，由有理数到实数，由实数到复数。每次扩大，都会令一些旧问题变得容易。

延伸练习

1. **一个障碍。**证明没有任何形如 $4k+3$ 的数是两个平方数之和。（平方数模 4 是甚么?）
2. **两种写法。**把 65 和 85 各自用两种不同的方法写成两个平方数之和，并用高斯整数解释你的答案。
3. **四种写法。**$1105 = 5 \times 13 \times 17$。找出四种不同的方法，把 1105 写成两个正平方数之和。为甚么是四种?

4 **高斯质数。** 较难 证明若 $3 = (a + b\mathrm{i})(c + d\mathrm{i})$，其中 a, b, c, d 为整数，则其中一个因数的模为 1。（取模的平方。）这说明 3 是高斯质数。

延伸阅读

- **John Stillwell, *Elements of Number Theory***（Springer）。循序渐进讲到高斯整数及更深入的内容，并附历史背景。
- **Aigner & Ziegler, *Proofs from THE BOOK***，第 4 章，“Representing numbers as sums of two squares”（把数表示为两个平方数之和）。
- **Joseph Silverman, *A Friendly Introduction to Number Theory***，有关平方和与高斯整数的各章。

1.7 因数分解失效之时

题目

STEP I 2016，第 7 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/16-s1-q7`

引子

每个大于 1 的整数都可以写成质数(prime)的乘积,而且写法只有一种:$60 = 2 \times 2 \times 3 \times 5$,再没有别的写法。这个事实你大概已经用过无数次，熟悉得让人觉得显然成立，根本不用证明。

这道题构造了一个小小的数字世界，看上去完全合情合理，但在这个世界里，那个“显然”的事实却是错的。这是一道 STEP I 的题目，所用的数学并不艰深，但当中的教训却十分深刻。1847 年，一个著名的费马最后定理（Fermat's Last Theorem）证明尝试，正是败在这个教训之上，而代数学的一个全新分支亦因此诞生。

动笔之前

想一想 列出 S 和 T 的首几个元素。按照第 (iv) 部分的定义，S 的哪些元素是S-质数?

$S = \{1, 5, 9, 13, 17, 21, 25, 29, 33, \ldots\}$，而 $T = \{3, 7, 11, 15, 19, 23, \ldots\}$。$S$ 中的一个数如果不能写成 S 中较小元素（1 除外）的乘积，它就是 S-质数。留意 9 是 S-质数，尽管在普通整数中 $9 = 3 \times 3$，因为 3 并不在 S 之中。一个数是否“质数”，取决于你获准使用哪些数。

逐步解题

阶段 1 并集、交集与乘积

$S \cup T$ 是所有正奇数组成的集合，即两者的并集（union），因为每个奇数除以 4 的余数不是 1 就是 3。$S \cap T$ 是空集，即两者的交集（intersection）没有任何元素，因为没有一个数会有两个不同的余数。

> **想一想**　S 的一般元素应该怎样写？两个这样的元素相乘，乘积是甚么？

S 的元素是 $4a+1$，其中 $a \geq 0$，而

$$(4a+1)(4b+1) = 16ab + 4a + 4b + 1 = 4(4ab + a + b) + 1 \in S.$$

所以 S 在乘法下封闭（closed）。至于 T：

$$(4a+3)(4b+3) = 16ab + 12a + 12b + 9 = 4(4ab + 3a + 3b + 2) + 1,$$

这个数属于 S，而不属于 T。例如 $3 \times 7 = 21 \in S$。所以 T 在乘法下并不封闭。

阶段 2　属于 T 的质因数

> **想一想**　在 T 中取一个不是质数的数。它的质因数（prime factor）全是奇数。假如这些质因数全都属于 S，会有甚么后果？

这个数是奇数，所以它的质因数全是奇数，各自属于 S 或 T。假如它们全部属于 S，那么由于 S 在乘法下封闭，它们的乘积也会属于 S。但这个乘积就是我们原来的数，而它属于 T。所以至少有一个质因数属于 T。 ■

阶段 3　奇数个因数

> **想一想**　把偶数个 T 的元素乘起来，乘积会落在哪里？奇数个又如何？

每一对 T-数相乘都得到一个 S-数，而 S-数相乘仍然是 S-数，所以偶数个 T-因数的乘积属于 S。奇数个的话，乘积就是（S 中的某个数）$\times$（T 的一个元素），用同类的计算可知它属于 T。

现在取 $t \in T$。如果 t 是 T-质数，那就完成了。否则，把 t 写成 T 的元素的乘积，其中每个元素都大于 1，因而小于 t。当中任何不是 T-质数的因数都可以再分解下去，如此类推。因数每次都变小，所以这个过程必会停止（这里再次用到良序性质（well-ordering）），最后 t 就写成了 T-质数的乘积。由于乘积属于 T，因数的个数必定是奇数。 ■

阶段 4　两种分解

> **停一停，试一试**　在 S 中找一个数，它有两种真正不同的 S-质数分解。提示：利用像 9 这样的数，它们之所以是 S-质数，只因为它们在普通整数中的质因数并

不在 S 之中。

S-质数依次为 $5, 9, 13, 17, 21, 29, 33, 37, 41, 49, \ldots$。$21 = 3 \times 7$、$9 = 3 \times 3$ 和 $49 = 7 \times 7$ 全都是 S-质数，因为 3 和 7 都不在 S 之中。而且

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

这就是 441 的两种不同的 S-质数分解（图 1.7.1）。唯一分解（unique factorisation）在 S 中并不成立。 ■

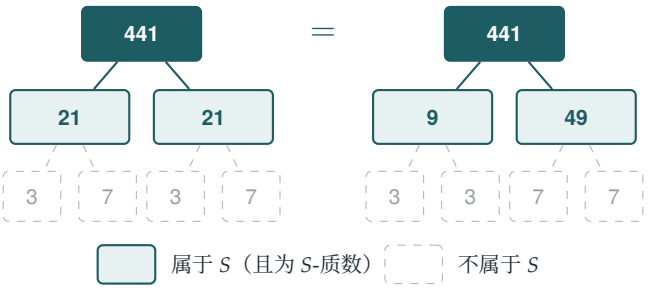


图 1.7.1 441 的两种 S-质数分解。在普通整数中，两棵树都会继续向下分解至 $3 \times 3 \times 7 \times 7$，但 3 和 7 都不在 S 之中，所以在 S 之内，分解到第二行便停止，而两种分解并不相同。

回顾

- **是否“质数”取决于所在的系统。**9 和 21 在整数中是合数（composite），在 S 中却是质数，因为分解它们所需的数并不存在。
- **封闭性很重要。**S 在乘法下封闭；T 则不然。第 (iii) 和第 (iv) 部分的论证，全靠这一点分别。
- **显然的事实也可以是一条定理。**整数的唯一分解是需要证明的。这道题说明了原因：在一个看上去非常相似的系统中，它并不成立。

背后的数学

整数有何特别之处

为甚么分解在普通整数中确实行得通？关键在于欧几里得引理（Euclid's lemma）：如果质数 p 整除乘积 ab，那么 p 整除 a 或 p 整除 b。这条引理是用辗转相除法（Euclidean algorithm）证明的，而它正正是在 S 中失效的地方。在那里，S-质数 9 整除 $21 \times 21 = 441$

（因为 $441 = 9 \times 49$），但 9 在 S 中并不整除 21。到了大学一年级，你会严谨地证明欧几里得引理，并看到唯一分解如何由它推导出来。

费马最后定理的失败证明

1847 年 3 月 1 日，拉梅（Gabriel Lamé）向巴黎科学院宣布，他已证明了费马最后定理。他的想法是利用复数单位根（roots of unity）把 $x^n + y^n$ 分解，

$$x^n + y^n = (x+y)(x+\zeta y)(x+\zeta^2 y)\cdots(x+\zeta^{n-1}y), \qquad \zeta = e^{2\pi i/n}$$

（其中 n 为奇数），然后就每个括号的质因数进行论证。刘维尔（Joseph Liouville）随即提出质疑：由 ζ 构成的数，凭甚么一定有唯一分解？数星期后，人们才知道库默尔（Ernst Kummer）早已证明这类数往往没有唯一分解。最先失效的情况是 $n = 23$。

库默尔尽力补救，发明了“理想数”（ideal numbers），这些额外的对象能够恢复唯一分解，他并借此对一大类指数证明了费马最后定理。戴德金（Richard Dedekind）把库默尔的理想数发展成理想（ideal），成为近世代数的核心概念之一。

一个值得记住的较小例子

在数系 $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$ 中，

$$6 = 2 \times 3 = (1+\sqrt{-5})(1-\sqrt{-5}),$$

而这四个因数在该数系中全都是“质数”（练习 3）。试把这个例子与 Section 1.6 的高斯整数（Gaussian integers）比较，后者的分解确实是唯一的。这类数系之中哪些具有唯一分解，是一个很深刻的问题。高斯（Gauss）猜想，在形如 $\mathbb{Z}[\sqrt{-d}]$ 的数系及与其相近的数系之中，只有九个具有唯一分解。这个猜想最终由黑格纳（Heegner）、贝克（Baker）和斯塔克（Stark）在 1952 年至 1967 年间证明。

思维技巧

> **质疑你视为理所当然的事**
>
> 当一个事实看来显然得不用证明时，问问它依赖甚么，并找出一个令它失效的情境。
>
> 构造反例（counterexample），也就是构造一个令“显然”的事实不成立的世界，是弄清一个证明真正需要甚么的最佳方法之一。大学本科代数有很大部分，就是在探问整数那些熟悉的性质，有哪些能在更一般的系统中保留下来。

延伸练习

1 **另一个例子。**证明 693 有两种不同的 S-质数分解。你能找出一个制造这类例子的通用方法吗?

2 **偶数。**设 $E = \{2, 4, 6, 8, \ldots\}$。哪些数是 E-质数?证明 60 有多于一种 E-质数分解。

3 $\mathbb{Z}[\sqrt{-5}]$**。**定义*范数*（norm）$N(a + b\sqrt{-5}) = a^2 + 5b^2$。证明 $N(zw) = N(z)N(w)$。先证明没有任何元素的范数是 2 或 3，由此推导出 2、3 和 $1 \pm \sqrt{-5}$ 在 $\mathbb{Z}[\sqrt{-5}]$ 中不能再分解。

4 **欧几里得引理的应用。**较难 假设欧几里得引理成立（如果一个质数整除 ab，它就整除 a 或 b），证明每个大于 1 的整数都只有一种质数分解，因数的次序不计。

延伸阅读

- **Ian Stewart & David Tall,** ***Algebraic Number Theory and Fermat's Last Theorem***（CRC Press）。开首几章讲述拉梅、库默尔与理想的故事，程度适合好学的大学一年级学生。
- **John Stillwell,** ***Elements of Number Theory***。把整数的唯一分解与高斯整数的唯一分解并列比较。
- **Harold Edwards,** ***Fermat's Last Theorem: A Genetic Introduction***（Springer）。适合有志挑战的读者：循历史发展解说库默尔的工作。

1.8 计数有理数

题目

STEP III 2020，第 8 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2020 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

`www.ocr.org.uk/Images/696589-2020-past-papers.zip`

引子

你能否把所有正分数逐一列出，使每个分数都恰好出现一次？这看来是不可能的：任意两个分数之间还有无穷多个分数，根本没有“下一个”分数可以写下来。然而，康托尔（Georg Cantor）在 1870 年代证明了这是做得到的。这表示，在某个精确的意义下，分数并不比整数多。

康托尔的列表颇为笨拙：表中既有 $\frac{1}{2}$，也有 $\frac{2}{4}$，重复的项要逐一删去。本题要构造的列表却完全没有重复，而且来自一个规则简单得出奇的数列（sequence）。这个列表由 Neil Calkin 和 Herbert Wilf 于 2000 年发表，不过数列本身早在 1858 年已由 Moritz Stern 研究过。

动笔之前

想一想 在证明任何东西之前，先计算最初十五项左右。你留意到比值 u_n/u_{n+1} 有甚么特别之处？

数列的开首几项是

$$1,\ 1,\ 2,\ 1,\ 3,\ 2,\ 3,\ 1,\ 4,\ 3,\ 5,\ 2,\ 5,\ 3,\ 4,\ 1,\ 5,\ \ldots$$

而比值 u_n/u_{n+1} 依次是

$$\frac{1}{1},\ \frac{1}{2},\ \frac{2}{1},\ \frac{1}{3},\ \frac{3}{2},\ \frac{2}{3},\ \frac{3}{1},\ \frac{1}{4},\ \frac{4}{3},\ \frac{3}{5},\ \frac{5}{2},\ \frac{2}{5},\ \frac{5}{3},\ \frac{3}{4},\ \frac{4}{1},\ \ldots$$

每个分数都以最简分数（lowest terms）的形式出现，而且没有一个重复。题目要你证明这个现象会一直持续下去，并且每个正分数迟早都会出现。

逐步解题

阶段 1　奇数项胜过偶数项

相邻两项不是 (u_{2k}, u_{2k+1})，就是 (u_{2k+1}, u_{2k+2})。

> **想一想**　试用 u_k 和 u_{k+1} 表示这两种数对。哪一项较大？

$(u_{2k}, u_{2k+1}) = (u_k,\ u_k + u_{k+1})$，而 $(u_{2k+1}, u_{2k+2}) = (u_k + u_{k+1},\ u_{k+1})$。由于每一项都是正整数（用归纳法（induction）不难证明），$u_k + u_{k+1}$ 比 u_k 和 u_{k+1} 都大。唯一的例外是第一对 $(u_1, u_2) = (1, 1)$，它不符合上述任何一种形式，因为 u_1 是直接定义的。 ■

阶段 2　相邻两项互质

> **想一想**　假设 $d > 1$ 同时整除 u_{2k} 和 u_{2k+1}，即同时整除 u_k 和 $u_k + u_{k+1}$。那么 d 还整除甚么？

d 整除 $(u_k + u_{k+1}) - u_k = u_{k+1}$，所以 d 整除相邻数对 u_k, u_{k+1}，而这一对在数列中的位置较前。另一种情况完全相同：若 d 整除 $u_k + u_{k+1}$ 和 u_{k+1}，它便整除 u_k。

因此，相邻数对的公因数会传回较前的一对，然后再传回更前的一对，如此类推。位置不可能无止境地减小（又是良序性（well-ordering）），所以 d 最终必须整除第一对 $u_1 = u_2 = 1$。当 $d > 1$ 时这是不可能的。所以每一对相邻项都互质（coprime）。 ■

阶段 3　没有数对出现两次

> **停一停，试一试**　阶段 1 告诉你一件有用的事：看到一个相邻数对 (a, b) 时，你可以判断 a 位于奇数位置还是偶数位置。怎样判断？然后，你能否还原出阶段 2 所指的、在它之前的那个数对？

若 $a < b$，则（由阶段 1）b 位于奇数位置，所以对某个 k，这个数对是 $(u_{2k}, u_{2k+1}) = (u_k, u_k + u_{k+1})$，而它的“父”数对 (u_k, u_{k+1}) 就是 $(a,\ b - a)$。若 $a > b$，这个数对是 (u_{2k+1}, u_{2k+2})，其父数对是 $(a - b,\ b)$。若 $a = b$，阶段 1 表明这个数对必定是最初的一对 $(1, 1)$。

所以除第一对外，每个相邻数对都决定了它的父数对，*以及父数对的位置*（这个结构可以看成一棵树，见图 1.8.1）：当子数对位于 $2k$ 或 $2k + 1$ 时，父数对位于位置 k。现在假设 (a, b) 出现在两个不同的位置。它们的父数对便是出现在两个位置的同一数对，而这些父数对的父数对亦然，如此类推。这样一直追溯上去，我们最终会在 1 以外的位置找到 $(1, 1)$，而阶段 1 已排除了这个可能。 ■

阶段 4 每个互质数对都会出现

想一想 假设某个互质数对 (a,b) 从未以相邻两项的形式出现。你应该考虑哪一个这样的数对？

在所有“缺席”的互质数对之中，取 $a+b$ 的值最小的一个：这就是极小反例（minimal counterexample）。它不可能是 $(1,1)$，因为这一对确实出现了，所以 $a \neq b$（互质而又相等，必有 $a=b=1$）。

- 若 $a>b$：数对 $(a-b,b)$ 互质（$a-b$ 和 b 的任何公因数都整除 a），而它的和 a 小于 $a+b$。由极小性，这个数对确实出现，设它位于位置 $k,k+1$。但这样便有 $u_{2k+1}=u_k+u_{k+1}=a$ 和 $u_{2k+2}=u_{k+1}=b$，所以 (a,b) 终究还是出现了。
- 若 $a<b$：同理，$(a,b-a)$ 出现在某个 $k,k+1$，于是 $(u_{2k},u_{2k+1})=(a,b)$。

无论哪种情况都得出矛盾。所以每一对互质的正整数都会以相邻两项的形式出现。 ■

（题目把第 (iv) 部分写成递降法（descent）：一个缺席的数对会产生一个更小的缺席数对。这其实是同一个论证，只是倒过来读而已。）

阶段 5 所有分数的列表

每个正有理数（rational number）都可以唯一地写成最简分数 a/b，其中 a,b 互质。由阶段 4，(a,b) 会以某个相邻数对 (u_n,u_{n+1}) 的形式出现，所以 $f(n)=a/b$：f 的值域（range）是全体正有理数。由阶段 3，它只出现一次；又由阶段 2，每个值 $f(n)$ 本身已是最简分数，所以两个不同的 n 不可能给出同一个分数。因此 f 是由正整数到正有理数的双射（bijection），并且有逆函数（inverse）。 ■

回顾

- **递降与它的逆向。**公因数和重复的数对都是往回追溯到数列的起点。缺席的数对则用极小反例处理，向前构造。
- **双射要证两半。**“每个分数都出现”（满射，onto）和“没有分数出现两次”（单射，one-to-one）是两个独立的命题，要分开证明。
- **先动手计算。**写出十五项之后，这些命题便显得可信，同时也提示了推动整个证明的父子结构。

背后的数学

Calkin–Wilf 树

把正分数排成一棵树：把 $\frac{1}{1}$ 放在顶端，并给每个分数 $\frac{a}{b}$ 两个子节点，左边是 $\frac{a}{a+b}$，右边是 $\frac{a+b}{b}$。逐行由左至右读这棵树，得到的正是本题的列表 $f(1), f(2), f(3), \ldots$（图 1.8.1）。阶段 3 说明每个分数都有唯一的父节点，阶段 4 说明每个分数都在树中。2003 年，Moshe Newman 发现连这棵树也不需要：列表中每个分数都可以由前一个分数算出，公式是

$$x_{n+1} = \frac{1}{2\lfloor x_n \rfloor - x_n + 1}.$$

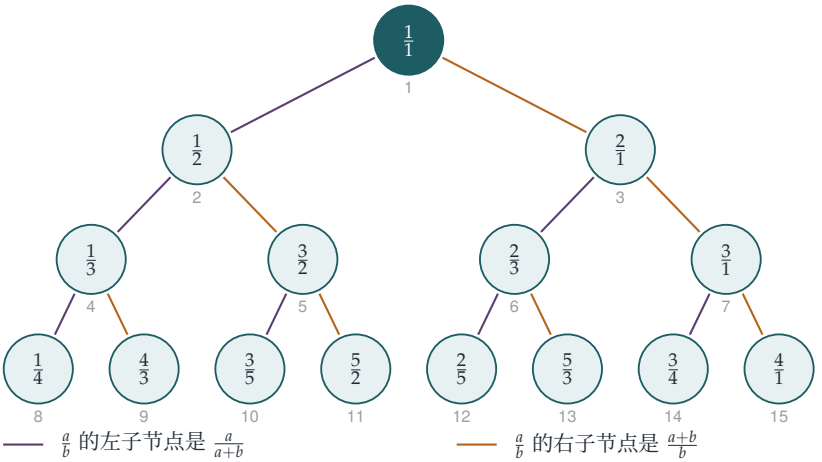


图 1.8.1　Calkin–Wilf 树的首四行。逐行由左至右读，便得到 $f(1), f(2), f(3), \ldots$；灰色小数字是位置 n。每个正分数都以最简分数的形式恰好出现一次。

不同大小的无穷

可以这样逐一列出的集合称为*可数*（countable）集合。分数是可数的，这令人意外，因为分数看来比整数多得多。在大学的第一个学期，你还会遇到康托尔的*对角论证法*（diagonal argument），它证明*实数*（real numbers）是不可数的：任何实数列表都必定有所遗漏。所以无穷确实有不同的大小，而实数所构成的无穷严格地大于有理数所构成的无穷。这是你在大学一年级课程中会遇到的最令人震惊的结果之一。

Stern 数列

数列 u_n 称为 *Stern 双原子数列*（Stern's diatomic sequence），它还有许多其他面貌。例如，u_{n+1} 等于把 n 写成 2 的幂之和、而每个幂最多用两次的方法数目。它与 *Stern–Brocot 树*以及连分数（continued fractions）关系密切，而连分数能给出无理数的最佳有理逼近。

思维技巧

构造双射

要证明两个集合大小相同，就要把它们的元素准确地一一配对：证明目标集合的每个元素都被对应到（满射），而且没有元素被对应两次（单射）。

双射是计数的基础，不论有限还是无穷，你在大学里会经常用到。在本题中，两半的证明都是把一个结构追溯回它的起点。

延伸练习

1 **找出分数。**对哪一个 n 有 $f(n) = \frac{3}{5}$？又 $f(n) = \frac{5}{3}$ 呢？利用父节点的规则倒推回 $\frac{1}{1}$。

2 **Newman 公式。**验证由 $x_1 = 1$ 开始，Newman 公式能给出列表中的首十个分数。

3 **康托尔的对角论证法。**假设你有一个列表，以小数形式列出了 0 与 1 之间的所有实数。构造一个小数，使它与列表中第 n 个数在第 n 位数字上不同。为甚么这就说明列表并不完整？（对于像 $0.0999\ldots = 0.1$ 这样的小数，需要留意甚么？）

4 **可数并集。** 较难 给出一个与正整数之间的明确双射，从而证明由全体正整数数对 (m, n) 组成的集合是可数的。（试试 $2^{m-1}(2n-1)$。）

延伸阅读

- **Neil Calkin & Herbert Wilf, "Recounting the rationals"**（*American Mathematical Monthly*，2000 年；〈重新数算有理数〉）。全文四页，完全读得懂。
- **Aigner & Ziegler, *Proofs from THE BOOK***（《数学天书中的证明》），第 17 章 "Sets, functions, and the continuum hypothesis"（集合、函数与连续统假设），当中包括 Calkin–Wilf 树。
- **Ian Stewart & David Tall, *The Foundations of Mathematics***（《数学基础》，Oxford Uni-

versity Press）。可数性、康托尔与实数，内容与大学一年级课程相若。

1.9 三次曲线的切线

题目

STEP I 2009，第 2 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/09-s1-q2`

引子

这题看来只是一道例行的坐标几何（coordinate geometry）题：求导、找切线、解三次方程。可是最后一行却要你找出一个立方数，它是另外三个立方数之和，而几何竟像变魔术一样把它变了出来。

这个窍门是过曲线上一个已知点作直线，从而找出新的点，它可追溯至公元三世纪亚历山大港的丢番图（Diophantus）。到了现代数学家手上，它发展成椭圆曲线（elliptic curves）的理论。这套理论位于当今数论的核心，是费马最后定理（Fermat's Last Theorem）证明的基础，也保障着全球一大部分互联网通讯的安全。

动笔之前

? **想一想** 点 $(-a, b)$ 在曲线上。为甚么？这跟立方和又有甚么关系？

代入：$(-a)^3 + a^3 + b^3 = b^3$，所以确实在曲线上。在 $y^3 = x^3 + a^3 + b^3$ 上坐标为有理数（rational）的点正是我们想要的：若 $x = q/s'$ 而 $y = p/s'$，去分母后便得到一条立方数之间的关系式。我们的计划是从一个容易找到的点出发，利用几何找出一个没那么明显的点。

逐步解题

阶段 1 切线

对 $y^3 = x^3 + a^3 + b^3$ 作隐函数求导（implicit differentiation）：$3y^2\dfrac{dy}{dx} = 3x^2$，所以 $\dfrac{dy}{dx} = \dfrac{x^2}{y^2}$。在 $(-a, b)$ 处斜率（gradient）为 a^2/b^2，所以切线（tangent）是

$$y - b = \frac{a^2}{b^2}(x + a) \quad\Longrightarrow\quad b^2y - b^3 = a^2x + a^3 \quad\Longrightarrow\quad b^2y - a^2x = a^3 + b^3. \qquad \blacksquare$$

阶段 2　切线与曲线再次相交之处

取 $a = 1$、$b = 2$，曲线是 $y^3 = x^3 + 9$，而在 $(-1, 2)$ 的切线是 $4y - x = 9$，即 $y = \dfrac{x+9}{4}$。代入曲线方程：

$$\frac{(x+9)^3}{64} = x^3 + 9 \quad\Longrightarrow\quad x^3 + 27x^2 + 243x + 729 = 64x^3 + 576,$$

整理后得 $63x^3 - 27x^2 - 243x - 153 = 0$，再除以 9 便得 $7x^3 - 3x^2 - 27x - 17 = 0$。 $\blacksquare$

阶段 3　解三次方程

> **想一想**　不用做任何代数运算，你已经知道这个三次方程的一个根。是哪一个？它为甚么应该出现两次？

切线在 $x = -1$ 处与曲线相切（图 1.9.1），而“相切”意味着重根（repeated root），正如抛物线的切线与抛物线相交于某个二次方程的二重根。所以 $(x+1)^2$ 应该是一个因式。验算：$7(-1)^3 - 3(1) - 27(-1) - 17 = -7 - 3 + 27 - 17 = 0$，而导数 $21x^2 - 6x - 27$ 在 $x = -1$ 处也等于零。相除后得

$$7x^3 - 3x^2 - 27x - 17 = (x+1)^2(7x - 17).$$

> **停一停，试一试**　所以切线与曲线再次相交于 $x = \frac{17}{7}$ 之处。求出 y，然后把这个点化为一条立方数之间的关系式。

$y = \dfrac{17/7 + 9}{4} = \dfrac{80/7}{4} = \dfrac{20}{7}$。点 $\left(\frac{17}{7}, \frac{20}{7}\right)$ 在 $y^3 = x^3 + 9$ 上，所以

$$\frac{20^3}{7^3} = \frac{17^3}{7^3} + 9 \quad\Longrightarrow\quad 20^3 = 17^3 + 9 \times 7^3.$$

> **想一想**　我们希望 9×7^3 是两个立方数之和。回想一下那个 9 是怎样来的。

$9 = a^3 + b^3 = 1^3 + 2^3$，所以 $9 \times 7^3 = 7^3 + 14^3$。因此

$$20^3 = 17^3 + 14^3 + 7^3.$$

验算：$8000 = 4913 + 2744 + 343$。 ■

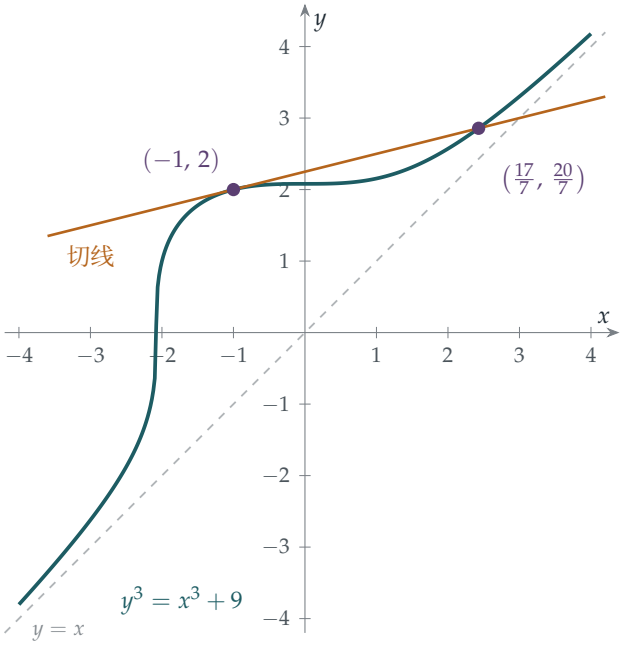


图 1.9.1 曲线 $y^3 = x^3 + 9$（蓝绿色）在远离原点处趋近直线 $y = x$。在有理点 $(-1, 2)$ 的切线（琥珀色）在该处与曲线相切，这算作二重交点，并在另一个有理点 $\left(\frac{17}{7}, \frac{20}{7}\right)$ 再与曲线相交一次。

回顾

- **相切即二重根。**这令一个没有明显根的三次方程，变成可以即时因式分解的方程。
- **新的根自动是有理数。**三个根之中一旦有两个是有理数（这里两个都是 -1），第三个也必定是有理数，因为各根之和 $\frac{3}{7}$ 是有理数。这正是整个方法的精髓。
- **几何带来了算术。**我们从没有刻意搜索立方数；是一条切线把它们找了出来。

背后的数学

弦切法

取一条三次曲线（cubic curve），例如 $y^3 = x^3 + 9$，或者更标准的形式 $y^2 = x^3 + Ax + B$。假如你知道曲线上两个坐标为有理数的点，那么过这两点的直线恰好与曲线再相交于一点，而该点的坐标也是有理数（理由见回顾中的论证）。假如你只知道一个点，就改用

该点的切线。无论哪种情况，旧的有理点都能产生新的有理点。

值得注意的是，这个“弦切”（chord–tangent）作图法可以变成一种把曲线上的点*相加*的方法，于是有理点便构成一个*群*（group），这是你在大学一年级会学到的一种代数结构。莫德尔（Louis Mordell）在 1922 年证明了这个群由有限个点生成：所有有理点，不论有多少个，都可以由一个有限集合出发，利用弦和切线构造出来。他的证明是 Section 1.5 中费马（Fermat）下降法的一个版本。

今日的椭圆曲线

这类曲线称为*椭圆曲线*（并非因为它们是椭圆，而是因为历史上它们与椭圆的弧长有关）。它们在现代数学中无处不在。

- **费马最后定理。**在 1980 年代，Gerhard Frey 留意到 $x^n + y^n = z^n$ 的一个解会产生一条非常古怪的椭圆曲线。怀尔斯（Andrew Wiles）在 1994 年证明了这样古怪的曲线不可能存在，从而证明了费马最后定理。
- **贝赫与斯维讷通－戴尔猜想（Birch and Swinnerton-Dyer conjecture）。**有理点所构成的群需要多少个生成元（generators）？在 1960 年代，有人凭借剑桥一些数学史上最早的电脑实验提出了一个猜想，它用一个函数来预测答案，而这个函数是由计数曲线在模质数之下的点数构造出来的。它是七条千禧年大奖难题（Millennium Prize Problems）之一。
- **密码学（cryptography）。**在有限数系上的椭圆曲线上，把一个点 P 自己相加 n 次是很快的，即使 n 有数百位数字也一样（把 P 反复加倍，就像你用反复平方来计算高次幂那样）。但已知 P 和结果 nP，要找出 n 却似乎极之困难：这就是*椭圆曲线离散对数问题*（elliptic-curve discrete logarithm problem）。这种不对称性保护着今日大部分的网上通讯。

同次幂之和

能写成三个正立方数之和的最小立方数是 $6^3 = 3^3 + 4^3 + 5^3$；切线法找到的是 $20^3 = 7^3 + 14^3 + 17^3$，它并非前者的倍数，而重复使用这个方法会不断产生新的恒等式（虽然有些会涉及负数）。要多少个 k 次幂才能凑成另一个 k 次幂？欧拉（Euler）在 1769 年猜想，至少需要 k 个 k 次幂才能凑成另一个 k 次幂。对于四次幂和五次幂，他错了：1966 年 Leon Lander 和 Thomas Parkin 以电脑搜索找到

$$27^5 + 84^5 + 110^5 + 133^5 = 144^5$$

而在 1988 年，Noam Elkies 利用椭圆曲线找到了四次幂的反例（counterexamples）。

思维技巧

以几何解决算术

要找出一个方程的整数解或有理数解，不妨把它看成一条曲线，然后寻找过已知点的直线。

背后的代数很简单：若一个有理系数多项式（polynomial）除了一个根之外其余的根都是有理数，那么最后一个根也是有理数。这样在图像与方程之间来回转换，是数学家所能拥有的最强大的习惯之一。

延伸练习

1 **圆上的有理点。**过 $(-1,0)$ 而斜率为 t 的直线与圆 $x^2+y^2=1$ 再次相交。证明另一个交点是 $\left(\dfrac{1-t^2}{1+t^2},\dfrac{2t}{1+t^2}\right)$，并由此得出无穷多组毕氏三元数（Pythagorean triples）。

2 **一条落空的弦。**点 $(1,2)$ 和 $(2,1)$ 在 $x^3+y^3=9$ 上。证明过这两点的直线与曲线再没有其他交点。那个“第三点”到哪里去了？（查一查无穷远点（point at infinity）。）

3 **故技重施。**求 $y^3=x^3+9$ 在 $\left(\frac{17}{7},\frac{20}{7}\right)$ 的切线，以及它与曲线相交的第三点。（使用电脑代数系统会有帮助；留意数字增长得有多快。）

4 **的士数（taxicab numbers）。**$1729=1^3+12^3=9^3+10^3$。求过 $(1,12)$ 和 $(9,10)$ 的直线与曲线 $x^3+y^3=1729$ 相交的第三点，并由此推导出一条形如 $p^3=q^3+r^3+s^3$ 的新恒等式。

延伸阅读

- **Joseph Silverman & John Tate, *Rational Points on Elliptic Curves***（Springer）。为大学本科生而写；第 1 章所需的知识不超出英国高考（A-level）程度。
- **Avner Ash & Robert Gross, *Elliptic Tales***（Princeton University Press）。一本普及读物，一路讲到贝赫与斯维讷通－戴尔猜想。
- **Simon Singh, *Fermat's Last Theorem***。讲述椭圆曲线如何解决一个有 350 年历史的难题。

我们学到了甚么

在本章开首，我们就整数提出了三个问题，并预告还有第四个主题贯穿其中：怎样证明（prove）一件事。现在是时候看看，九道 STEP 题目把我们带到了多远。

1. 质数有多少个？它们是怎样分布的？

在 Section 1.1 中，我们证明了质数（prime）有无穷多个，却连一个新的质数也没有找出来。我们构造了一族无穷多个的数，即费马数（Fermat numbers），它们两两之间没有公因数，所以每一个都需要一个属于自己的质数。这个证明谈的是*结构*，而不是大小。

在 Section 1.2 中，我们更进一步，证明了不超过 n 的所有质数之积小于 4^n。这是我们第一个*定量*的结论：它限制了质数可以排得多密。取对数后，它说的是 $\sum_{p \le n} \ln p < 1.39\, n$，而质数定理（Prime Number Theorem）指出真实的值大约是 n。我们这个 STEP 程度的论证，已经相当接近了。

我们*没有*做到的，是确切找出质数怎样分布。这由黎曼 ζ 函数（Riemann zeta function）所支配，而把这件事说得精确，就是黎曼猜想（Riemann Hypothesis），也就是数学中最著名的未解决问题。你现在所知的，已足以明白它在说甚么。

2. 怎样证明一个数不是分数？

我们找到了三种截然不同的方法。

- **取最小的那一个**（Section 1.3）。假如 $\sqrt{2}$ 是分数，就会有一个最小的正整数 k 使 $k\sqrt{2}$ 为整数；但这样一来，$(\sqrt{2}-1)k$ 就是一个更小的。同样的论证也适用于立方根。
- **把一个整数困在 0 与 1 之间**（Section 1.4）。假如 e 或 e^2 是分数，乘以一个阶乘（factorial）后，就会得出一个严格介乎 0 与 1 之间的整数。埃尔米特（Hermite）和林德曼（Lindemann）证明 e 和 π 是超越数（transcendental），背后正是这个想法。
- **找出一个不可能成立的整除关系**（Section 1.5，习题 3）。证明 $x^2 + y^2 = 3$ 上没有有理点（rational point），等同于证明当 x 是有理数时，$\sqrt{3 - x^2}$ 永远不是有理数。

我们也认识到还有多少事情仍然未知：至今没有人能证明 $e + \pi$ 是无理数（irrational）。

3. 哪些方程有整数解?

有时答案是*没有*，而 Section 1.5 展示了费马（Fermat）证明这一点的方法：任何一个解都会产生一个更小的解，如此无休无止。以一个精心挑选的质数为模（modulo）来运算，就能告诉我们该利用哪一个整除关系。

有时答案是*有，而且这里有一部制造解的机器*。在 Section 1.6 中，复数（complex numbers）把二平方和变成了一条乘法法则。在 Section 1.9 中，三次曲线（cubic curve）的一条切线给出了 $20^3 = 17^3 + 14^3 + 7^3$。

还有些时候，正确的一步是*换一个数系*。高斯整数（Gaussian integers）令二平方和变得容易。但 Section 1.7 提出了一个警告：在一个看来完全合理的数系中，数可以用两种不同的方式分解成“质数”。这个警告在 1847 年被人忽视，结果催生了理想（ideals）和近世代数。

Section 1.8 还另添了一重转折：分数看似远比整数多，却可以逐一列举出来。无穷比乍看之下有更丰富的结构。

4. 证明的工具箱

以下是你在本章用过的方法。每一种都会在大学数学课程中一再出现。

方法	你在哪里用过
归纳法（induction）与强归纳法（strong induction）	费马数（1.1）；质数之积（1.2）
良序性（well-ordering）、最小反例（minimal counterexample）	$\sqrt{2}$ 与 $\sqrt[3]{2}$（1.3）；列举分数（1.8）
无穷递降法（infinite descent）	不可能成立的方程（1.5）；互质（coprime）数对（1.8）
反证法（proof by contradiction）	贯穿全章，尤其是 1.3 和 1.4
模 n 运算	1.5；1.7
估界与夹逼（bounding and squeezing）	1.2；1.4
构造双射（bijection）	1.8
构造反例（counterexample）	非唯一分解（1.7）
转换场景	复数（1.6）；几何（1.9）

回头看看这些题目，你会留意到 STEP 的拟题人往往替你走了第一步：“考虑二项展开式（binomial expansion）…”、“考虑复数 $2 + \mathrm{i}$ 和 $10 + 11\mathrm{i}$…”。到了大学，这种帮助会逐渐消失，而找出正确的第一步就成了真正的本领。最好的准备正是你一直在做的事：每做完一道题，问问自己那个提示*为甚么*奏效，下次你就能自己想到。

下一步

本章有好几条线索会在书中其他地方延续下去。

- 以分数逼近 e 的课题会在 Section 2.9 再次出现，而黎曼 ζ 函数则见于 Section 2.4，那里有 $\zeta(2) = \pi^2/6$。
- 单位根（roots of unity）曾在拉梅（Lamé）那个失败的证明中出现，它是 Chapter 5 的主题；而在 Section 1.1 中预告过的高斯（Gauss）正十七边形作图法，就是 Section 6.2。
- Section 1.6 中那条恒等式的四平方版本，会通过四元数（quaternions）在 Section 7.8 中出现。

如果你想在上大学前走得更远

- **Harold Davenport, *The Higher Arithmetic***：初等数论入门书中最好的一本。
- **Aigner & Ziegler, *Proofs from THE BOOK***：第 1–4、8 和 17 章与本章直接相关。
- **Joseph Silverman, *A Friendly Introduction to Number Theory***：几乎涵盖本章的全部内容，而且还有更多，节奏平缓。
- **Marcus du Sautoy, *The Music of the Primes*** 和 **Simon Singh, *Fermat's Last Theorem***：数学背后的人物故事。

数论总有办法把人吸引进去。它的问题简单得第一天就可以提出，却又深奥得足以耗上一生。如果其中某些章节令你很想知道接下来会怎样，那正是本书的用意所在。

延伸练习答案

以下只是简短的答案和提示，并非完整解答：只列出结果和关键步骤，方便你核对自己的做法。请先认真做过每一道练习，然后才看答案。

1.1 质数工厂

1. 假设 $p_1, \dots, p_k$ 就是全部的质数 (prime)；那么 $p_1 \cdots p_k + 1$ 有一个不在此列之中的质因数 (prime factor)，矛盾 (contradiction)。用费马数 (Fermat number) 的证明则是构造性 (constructive) 的：它具体给出无穷多个两两互质 (pairwise coprime) 的数 $F_0, F_1, \dots$，每一个都提供一个新的质数。

2. 当 $n \geq 2$ 时，2^n 是 4 的倍数，所以 $2^{2^n} = 16^{2^{n-2}}$，而 16 的任何次幂的末位数字都是 6。因此 F_n 的末位数字是 7。

3. (a) 用归纳法 (induction)：$a_{n+1} - 1 = a_n(a_n - 1) = a_n \cdot a_1 \cdots a_{n-1}$。(b) a_m 与 a_n ($m < n$) 的公因数必整除 $a_1 \cdots a_{n-1} = a_n - 1$ 和 a_n，所以只能是 1；每一项都有一个不与其他项共有的质因数。(c) $\frac{1}{a_n} = \frac{1}{a_n - 1} - \frac{1}{a_{n+1} - 1}$；求和时各项裂项相消 (telescope)，得 $1 - \frac{1}{a_{n+1} - 1}$。

4. (a) 由 $p \mid F_n$ 得 $2^{2^n} \equiv -1$；两边平方得 $2^{2^{n+1}} \equiv 1$。(b) 凡满足 $2^e \equiv 1$ 的 e 都能被 d 整除，所以 $d \mid 2^{n+1}$；又因为 $2^{2^n} \equiv -1 \not\equiv 1$ (因 p 是奇数)，所以 $d \nmid 2^n$。故 $d = 2^{n+1}$。(c) $2^{p-1} \equiv 1$，所以 $2^{n+1} \mid p - 1$。(d) $641 = 64 \times 10 + 1$。小于 641 而形如 $64k + 1$ 的质数有 193、257、449 和 577：只需试四个候选数，欧拉 (Euler) 正是这样找到 641 的。

5. 第 n 个质数不会超过 $F_0, \dots, F_{n-1}$ 之中任何一个数的最大质因数，所以 $p_n \leq F_{n-1} = 2^{2^{n-1}} + 1$。当 $n = 5$ 时，这个上界是 65 537，而 $p_5 = 11$：结论正确，但极为粗疏。

1.2 质数分布得有多稀疏?

1. $P_{1,10} = 210$，$P_{1,20} = 9\,699\,690$，$P_{1,30} = 6\,469\,693\,230$。与 $4^{30} \approx 1.15 \times 10^{18}$ 和 $e^{30} \approx 1.07 \times 10^{13}$ 比较：这个乘积远小于 4^n，增长速度大致与 e^n 相若 (事实上 $\ln P_{1,n} \sim n$，这是质数定理 (Prime Number Theorem) 的一种形式)。

2. $\binom{2m}{k+1} / \binom{2m}{k} = \frac{2m-k}{k+1}$，此值大于 1 当且仅当 $k < m$，所以各系数先递增至中间一项，然后递减。$2m + 1$ 项之和为 4^m，其中最大的一项至少等于它们的平均值。

3. 区间 $(\sqrt{n}, n]$ 内的质数每个都大于 $\sqrt{n}$,而它们的乘积小于 4^n,所以 $(\sqrt{n})^{\pi(n)-\pi(\sqrt{n})} < 4^n$；两边取对数（logarithm）。由于 $2\ln 4 \approx 2.77$ 且 $\pi(\sqrt{n}) \le \sqrt{n}$，即得所求的上界。它具有正确的形式 $n/\ln n$，只是常数为 2.8，而非真正的 1。

4. 取质数 $2, 3, 5, 7, 13, 23, 43, 83$，其中每一个都小于前一个的两倍。对任意 $n \le 50$，此列中第一个大于 n 的质数小于前一个质数的两倍，而前一个质数不超过 n，所以它介乎 n 与 $2n$ 之间。

1.3 最小反例

1. 设 k 是使 $k\sqrt{3}$ 为整数的最小正整数，则 $k' = (\sqrt{3}-1)k$ 是一个更小的正整数，而且 $k'\sqrt{3} = 3k - k\sqrt{3}$ 是整数。这里 $\sqrt{3}-1$ 所起的作用，与 $\sqrt{2}-1$ 相同。

2. 由于 $0 < \sqrt{N} - m < 1$，$k' = (\sqrt{N}-m)k$ 是一个小于 k 的正整数，而 $k'\sqrt{N} = kN - mk\sqrt{N}$ 是整数：矛盾。

3. 当 $N = 4$ 时，$m = 2 = \sqrt{4}$，所以 $\sqrt{N} - m = 0$，而 $k' = 0$ 并非正整数。“$0 < \sqrt{N} - m$”这一步不成立。

4. 在图中，边长为 a 的正方形内有两个边长为 b 的正方形互相重叠，它们把中央一个边长为 $2b - a$ 的正方形覆盖了两次，并留下两个边长为 $a - b$ 的正方形未被覆盖。用代数表示，就是 $(2b-a)^2 - 2(a-b)^2 = 2b^2 - a^2 = 0$。

1.4 e 是无理数，e^2 也是

1. 若 $\frac{1}{e} = \frac{p}{q}$，把 $\frac{1}{e} = \sum_{k\ge 0} \frac{(-1)^k}{k!}$ 乘以 $q!$。$k \le q$ 的各项都是整数，而尾部 $\sum_{k>q}(-1)^k q!/k!$ 是各项递减的交错级数（alternating series），所以其绝对值严格介乎 0 与 $\frac{1}{q+1}$ 之间。因此 $q!/e$ 不是整数：矛盾。

2. 若 $\sin 1 = \frac{p}{q}$，取 $2n+1 \ge q$，两边乘以 $(2n+1)!$。左边是整数，级数中直至 $\frac{1}{(2n+1)!}$ 的各项也是整数；尾部是交错级数，首项的大小为 $\frac{1}{(2n+2)(2n+3)} < 1$，所以它不是整数。

3. 这个方法需要一个各项均为有理数（rational）的级数，而且其尾部缩小的速度要比分母增长的速度快，就像 $\frac{1}{k!}$ 那样；对于 π，目前并不知道有这样的级数。Niven 的证明用到 $\int_0^\pi \frac{x^n(\pi-x)^n}{n!}\sin x\,\mathrm{d}x$：假如 π 是有理数，当 n 足够大时，这个积分就会是一个小于 1 的正整数。

4. $x^2 - (e+\pi)x + e\pi$ 的根是 e 和 π。假如 $e+\pi$ 和 $e\pi$ 都是代数数（algebraic number），π 就是一个代数系数多项式的根，因而也是代数数，这与林德曼（Lindemann）定理矛盾。

1.5 无穷递降法

1. $3 \mid a$；设 $a = 3a_1$，得 $9a_1^3 + b^3 = 3c^3$，所以 $3 \mid b$，于是 $3a_1^3 + 9b_1^3 = c^3$，所以 $3 \mid c$，而 (a_1, b_1, c_1) 是一个更小的解：这就是递降（descent）。至于 $p^4 + 2q^4 = 5r^4$：四次方数模（modulo）5 只能是 0 或 1（它们的末位数字是 0、1、5 或 6），所以 $p^4 + 2q^4 \equiv 0$ 迫使 $5 \mid p$ 且 $5 \mid q$；于是 $5^4 \mid 5r^4$，所以 $5 \mid r$，同样可以递降。

2. 平方数模 4 只能是 0 或 1，所以 $x^2 + y^2$ 只能是 0、1 或 2，绝不会是 3。

3. 若有有理点 $(\frac{a}{c}, \frac{b}{c})$，便会得到 $a^2 + b^2 = 3c^2$，其中 $c \neq 0$，而第 (iii) 部分已排除这个可能。

4. 固定 x、y 时，方程是二次方程 $z^2 - 3xyz + (x^2 + y^2) = 0$，其两根之和为 $3xy$。由 $(1,1,1)$ 出发：得 $(1,1,2)$；把 $(1,2)$ 所对的那个 1 换掉，得 $3 \cdot 1 \cdot 2 - 1 = 5$，即 $(1,2,5)$；然后是 $(1,5,13)$ 和 $(2,5,29)$。

1.6 两个平方数之和

1. 平方数模 4 只能是 0 或 1，所以两个平方数之和模 4 绝不会是 3。

2. $65 = 1^2 + 8^2 = 4^2 + 7^2$ 及 $85 = 2^2 + 9^2 = 6^2 + 7^2$。利用 $65 = (2+\mathrm{i})(2-\mathrm{i})(3+2\mathrm{i})(3-2\mathrm{i})$：$(2+\mathrm{i})(3+2\mathrm{i}) = 4 + 7\mathrm{i}$ 及 $(2+\mathrm{i})(3-2\mathrm{i}) = 8 - \mathrm{i}$。同理，对于 85，$(2+\mathrm{i})(4 \pm \mathrm{i})$ 给出 $7 + 6\mathrm{i}$ 和 $9 - 2\mathrm{i}$。

3. $1105 = 4^2 + 33^2 = 9^2 + 32^2 = 12^2 + 31^2 = 23^2 + 24^2$。5、13、17 各自都可分解为一个高斯整数（Gaussian integer）与其共轭（conjugate）之积；从每一对中各选一个因数，共得 8 个乘积，它们两两互为共轭（给出相同的平方和），所以剩下 4 个。

4. 取模的平方，得 $9 = (a^2 + b^2)(c^2 + d^2)$。由于 3 不是两个平方数之和，其中一个因数必为 1。

1.7 因数分解失效之时

1. $693 = 9 \times 77 = 21 \times 33$，四个因数全是 S-质数。方法：取四个形如 $4k+3$ 的质数 p, q, r, s；那么 $(pq)(rs) = (pr)(qs)$ 就给出两种分解为 S-质数的方式。

2. E-质数就是形如 $2 \pmod 4$ 的数（它们不可能是两个偶数之积）。$60 = 2 \times 30 = 6 \times 10$。

3. 展开 $(ac - 5bd)^2 + 5(ad + bc)^2$ 即得 $N(zw) = N(z)N(w)$。$a^2 + 5b^2$ 绝不会等于 2 或 3。由于 $N(2) = 4$、$N(3) = 9$ 及 $N(1 \pm \sqrt{-5}) = 6$，若要分解为非单位（non-unit）之积，就需要一个范数（norm）为 2 或 3 的因数。

4. 设 n 是具有两种不同分解 $p_1 \cdots p_r = q_1 \cdots q_s$ 的最小的数。根据欧几里得（Euclid）

引理，p_1 整除某个 q_j，所以 $p_1 = q_j$；约去后得到一个更小而又有两种不同分解的数，矛盾。

1.8 计数有理数

1. $\frac{3}{5}$：逐级追溯其父节点，$(3,5) \leftarrow (3,2) \leftarrow (1,2) \leftarrow (1,1)$，对应的位置为 $1 \to 2 \to 5 \to 10$，所以 $f(10) = \frac{3}{5}$。$\frac{5}{3}$：$(5,3) \leftarrow (2,3) \leftarrow (2,1) \leftarrow (1,1)$，位置为 $1 \to 3 \to 6 \to 13$，所以 $f(13) = \frac{5}{3}$。

2. Newman 的法则 $x_{n+1} = \frac{1}{2\lfloor x_n \rfloor - x_n + 1}$ 给出 1、$\frac{1}{2}$、2、$\frac{1}{3}$、$\frac{3}{2}$、$\frac{2}{3}$、3、$\frac{1}{4}$、$\frac{4}{3}$、$\frac{3}{5}$，与 $f(1), \ldots, f(10)$ 吻合。

3. 第 n 位数字一律取 5，除非第 n 个数的第 n 位数字是 5，这时则取 4。新的小数与列表中每一个数都不相同，而且由于避开了 0 和 9，它只有一种小数展开式，所以确实是一个不同的数。

4. 每个正整数都可唯一地写成 2 的幂乘以一个奇数，即 $2^{m-1}(2n-1)$，所以 $(m,n) \mapsto 2^{m-1}(2n-1)$ 是一个双射（bijection）。

1.9 三次曲线的切线

1. 把 $y = t(x+1)$ 代入 $x^2 + y^2 = 1$，得 $(1+t^2)x^2 + 2t^2 x + t^2 - 1 = 0$，其根为 -1 和 $\frac{1-t^2}{1+t^2}$。取 $t = \frac{q}{p}$，该点为 $\left(\frac{p^2-q^2}{p^2+q^2}, \frac{2pq}{p^2+q^2}\right)$，由此得到三元数组 $(p^2-q^2, 2pq, p^2+q^2)$。

2. 该直线为 $x + y = 3$，而 $x^3 + (3-x)^3 = 9$ 可化简为 $9x^2 - 27x + 18 = 0$：三次项互相抵消，只剩下 $x = 1, 2$。这条直线与渐近线（asymptote）$y = -x$ 平行，第三个交点在无穷远处。

3. 切线（tangent）的斜率（gradient）为 $\frac{x^2}{y^2} = \frac{289}{400}$，它与曲线再相交于 $\left(-\frac{188479}{90391}, -\frac{36520}{90391}\right)$。去分母后，得 $188479^3 = 36520^3 + 90391^3 + 180782^3$。

4. 该直线与 $x^3 + y^3 = 1729$ 再相交于 $\left(-\frac{37}{3}, \frac{46}{3}\right)$，所以 $46^3 = 37^3 + 27 \times 1729 = 37^3 + 3^3 + 36^3$（这里用到 $27 = 3^3$ 和 $27 \times 1728 = 36^3$）。

Beyond STEP

STEP 是大多数中学生一生中应考过最艰深的数学考试，也是最出色的考试之一：题目简短、陌生而深刻，当中最精彩的，其实是乔装了的大学数学小品。

本书精选最有趣的 STEP 试题，由 1987 年首次开考之前的样本试卷（specimen papers），到 2025 年的试卷，并以每一道题目作为一扇门。每道题目都会仔细解答，沿途提出优秀学生会问的问题，然后带你穿过这扇门：这个想法从何而来，到了大学会演变成甚么，以及接下来该读些甚么。

本书为正在准备 STEP、有天分又有志投身数学的学生而写。他们想要的不止是模拟试卷：他们想开始像数学系本科生那样思考，并一探前方有甚么正等着自己。

- 共 13 章，涵盖纯数学（pure mathematics）、力学（mechanics）、概率（probability）与统计（statistics）
- 136 道 STEP 历届试题，每题均附引导式解答
- 每道题目背后的大学数学
- 附答案的延伸练习，以及延伸阅读

全书共 13 章，本册为第 1 章。 全书各章及所有详细解答均可免费阅读、下载及分享，网址为

beyondstep.keithmcnulty.org

文字内容已贡献予公有领域（CC0 1.0）。STEP 试题 © Cambridge University Press & Assessment。

