

DAVENPORT · 圆法 · 高中详解版

引言：圆法是用来干什么的

Introduction

本章要解决什么 / 读完能掌握什么 这一章是全书的「开场白」，本身没有任何定理证明，作者 Davenport 只是在告诉你：圆法（circle method，又叫 Hardy-Littlewood 解析方法）是为了回答哪一类问题而被发明出来的，它从哪儿来、解决过哪些著名难题。读完这份详解，你应当能用高中语言回答下面这些问题：

1. 什么叫「加性问题」？华林问题、哥德巴赫问题各是什么意思？
2. 「把一个大数写成若干个 k 次方之和」这句话，用数学符号怎么精确地写出来（式 (1.1)）？
3. 圆法的「圆」是哪个圆？为什么后面会冒出复数 $e^{2\pi i\alpha}$ 、积分 $\int_0^1$ 、求和号 $\sum$ ？（本章只埋下伏笔，我们把这些超纲符号**从零讲清**。）
4. 分拆数 $p(n)$ 是什么？为什么 1917 年 Hardy 和 Ramanujan 关于它的一篇文章，被看作整个圆法的「起点」？
5. $O(\cdot)$ （大 O 记号）、 $\ll$ 这些记号是什么意思？ $e^{A\sqrt{n}}$ 这种增长有多快？

本章是「认路」，不是「赶路」。把这些名词和符号先认熟，后面几章的硬推导才不会一上来就把你劝退。

阅读方式：每一段先给出**原文译文**（灰条标注「原文第 N 段」之后的黑色正文就是忠实翻译），再用彩色框做**高中视角的逐句扩展**。凡是原文一笔带过、或用到超纲符号的地方，我们都补成一步一步的完整说明。

原文第1段·在讲什么 给「圆法」定位：它是为「加性问题」造的工具；并举出两个最有名的加性问题——华林问题、哥德巴赫三素数问题。

哈代 (Hardy) 与利特尔伍德 (Littlewood) 的解析方法 (有时称为「圆法」, circle method) 是为处理数论中的**加性问题** (additive problems) 而发展起来的。这类问题关心的是把一个大数表示为某些指定类型的数之和。被加数 (summand) 的个数既可以是固定的, 也可以是不受限制的; 在后一种情形下我们谈论**分拆问题** (partition problems)。最著名的加性问题是华林问题 (Waring's problem), 其中指定的数是 k 次幂, 于是问题就是把一个大数 N 表示为

$$N = x_1^k + x_2^k + \cdots + x_s^k, \quad (1.1)$$

其中 s 与 k 是给定的, 而 $x_1, \dots, x_s$ 是正整数。几乎同样著名的是哥德巴赫三素数问题 (Goldbach's ternary problem), 其中指定的数是素数, 问题是把一个大数 N 表示为

$$N = p_1 + p_2 + p_3.$$

先把「加性问题」一句话说透 「加性」就是「加法」。所谓**加性问题**, 指的是这样一类问题: 给定一堆「允许使用的数」(比如所有平方数, 或所有素数), 问一个大数能不能被写成这堆数里若干个相加? 有几种写法? 关键词是「相加」——我们只用加号去拼出目标数, 不用乘除。这和你熟悉的「质因数分解 (用乘法拼)」是相反方向的问题。

「被加数个数固定 vs 不受限」是什么区别

个数固定 (fixed)

规定「只许用 s 个数相加», s 是事先定好的常数。华林问题就是这种: 式 (1.1) 右边**恰好** s 项。这是本书的主战场。

个数不受限 (unrestricted)

不限制用几个数，能用多少个就用多少个。这时我们问的是「一共有多少种拼法」，这类问题叫**分拆问题** (partition problems)。例如把 4 写成若干正整数之和：4、3+1、2+2、2+1+1、1+1+1+1，共 5 种，所以分拆数 $p(4) = 5$ (这个 p 在第 3 段会重点出现)。

逐符号拆解式 (1.1): $N = x_1^k + x_2^k + \cdots + x_s^k$ 这是全书第一个、也是最核心的式子，每个符号都要认清：

N (读作「大 N 」)

我们要表示的那个**大数**，是一个给定的正整数。比如 $N = 100$ 。「大」是因为圆法只在 N 充分大时才保证有效。

k (次数 / 指数)

幂的**次数**，是事先给定的正整数。 $k = 2$ 就是平方， $k = 3$ 就是立方，依此类推。 x^k 表示 x 自乘 k 次： $x^k = \underbrace{x \cdot x \cdots x}_{k \text{ 个}}$ 。

s (项数 / 被加数个数)

右边一共有**多少项**相加，也是事先给定的正整数。比如 $s = 4$ 就是「用 4 个 k 次方相加」。

$x_1, x_2, \dots, x_s$ (下标变量)

s 个**正整数**，是我们要去寻找的「未知数」。下标 $1, 2, \dots, s$ 只是给它们编号，区分是第几个，并不代表它们要互不相等（允许重复，比如 $x_1 = x_2$ 是允许的）。

$\cdots$ (省略号)

表示「中间一路同样地加下去」，从第 3 项一直到第 $s - 1$ 项都按同样规律，省着不写。

所以式 (1.1) 用大白话念出来是：「能不能找到 s 个正整数，使它们各自的 k 次方加起来正好等于 N ？」

例：把华林问题具体化 ($k = 2, s = 4$) 取 $k = 2$ (平方)、 $s = 4$ (用四项)，式 (1.1) 变成 $N = x_1^2 + x_2^2 + x_3^2 + x_4^2$ 。

- $N = 7$: $7 = 2^2 + 1^2 + 1^2 + 1^2$ (即 $4 + 1 + 1 + 1$)。成立。
- $N = 30$: $30 = 4^2 + 3^2 + 2^2 + 1^2$ ($16 + 9 + 4 + 1$)。四项都是正整数，成立。(注：若像 Lagrange 定理那样允许出现 0，还可写 $30 = 5^2 + 2^2 + 1^2 + 0^2$ ；但式 (1.1) 的华林问题要求 x_i 是正整数，故这里取全为正的那组。)

事实上「每个正整数都能写成 4 个平方之和」是著名的**拉格朗日四平方和定理** (Lagrange, 1770)。华林 (Waring) 当年大胆猜测：对每个次数 k ，都存在一个**固定的项数 s** (只依赖 k 、不依赖 N)，使得**所有**足够大的 N 都能写成 s 个 k 次方之和。这就是「华林问题」的核心——它问的不仅是「能不能写」，更是「最少需要几项 s 」。

「素数 p 」与哥德巴赫三素数问题 素数 (prime) 就是大于 1、且只能被 1 和自身整除的正整数：2, 3, 5, 7, 11, 13, ...。哥德巴赫三素数问题里「指定的数」换成了素数，要把大数 N 写成三个素数之和 $N = p_1 + p_2 + p_3$ 。

- 例： $21 = 3 + 5 + 13$ 、 $21 = 7 + 7 + 7$ 、 $21 = 3 + 7 + 11$ 。

这里 p_1, p_2, p_3 三个下标同样只是编号，允许相等。这就是「三素数」(ternary，意为「三元」) 的含义。

为什么是三个而不是两个？著名的**哥德巴赫猜想**说「每个大于 2 的偶数都是两个素数之和」(如 $8 = 3 + 5$)，但「两素数」版本至今 (指 Davenport 写书的年代乃至现在) **仍未被证明**；而「三素数」版本 (对充分大的奇数 N) 已被维诺格拉多夫用圆法**证明了**。这正说明圆法的威力：它攻得下三素数，却暂时攻不下两素数——后面会解释为什么「多一项」会让问题从「做不出」变成「做得出」。

想表示的大数

$$N = x_1^k + x_2^k + \dots + x_s^k$$

恰好 s 项，每项都是某个正整数的 k 次方（华林问题）

若把「 k 次方」换成「素数」，就是哥德巴赫问题

加性问题的统一模样：把大数 N 拆成「指定类型的数」之和。改变「指定类型」（ k 次方 / 素数）和「项数」（固定 s / 不限），就得到不同的著名问题。

原文第 2 段 · 在讲什么 交代方法的「冠名权」：Hardy–Littlewood 之后，Vinogradov 又把它大大推进，所以也叫「Hardy–Littlewood–Vinogradov 方法」。

哈代与利特尔伍德伟大成就，后来又被维诺格拉多夫（Vinogradov）所取得的卓越进展所继承；因此，我们的俄国同行如今称之为「哈代–利特尔伍德–维诺格拉多夫方法」（Hardy–Littlewood–Vinogradov method），这并非没有道理。

人物与来龙去脉

G. H. Hardy（哈代，1877–1947）与 J. E. Littlewood（利特尔伍德，1885–1977）

20 世纪上半叶英国剑桥的两位顶尖数学家，长期合作。1920 年前后，他们在一系列题为《Some problems of "Partitio Numerorum"》（数的分拆诸问题）的论文里，把这套解析方法系统化，用来攻华林问题。这套方法就是本书的主角。

I. M. Vinogradov（维诺格拉多夫，1891–1983）

苏联数论大家。1937 年，他用经过自己改造的圆法证明了：**每个充分大的奇数都是三个素数之和**（即上一段的哥德巴赫三素数问题）。他还发明

了估计指数和的强力技术，极大改进了华林问题中所需项数 s 的上界。正因他的贡献分量极重，苏联（「俄国同行」）习惯把这套方法连他的名字一起称呼。

为什么这段重要：它告诉你这套方法不是一锤定音、而是几代人接力打磨出来的。本书后面要讲的，正是这条接力线上的核心技术。Davenport 用「这并非没有道理」（not without justice）这种含蓄说法，是在认可 Vinogradov 的贡献确实配得上把名字并列。

原文第 3 段·在讲什么 追溯方法的「起源」：1917 年 Hardy–Ramanujan 研究分拆数 $p(n)$ 的那篇论文，是圆法的胚胎；并提到平方和问题能给出精确公式，而更高次幂则不太可能。

值得回顾的是，哈代–利特尔伍德方法的起源可以追溯到哈代与拉马努金（Ramanujan）于 1917 年发表的一篇论文 [69]，其内容是关于 $p(n)$ 的渐近行为，即 n 的分拆总数。函数 $p(n)$ 的增长大致像 $e^{A\sqrt{n}}$ ，其中 A 是某个正常数；哈代与拉马努金为它求得了一个渐近级数，如果在最小项处截断，便给出 $p(n)$ ，误差为 $O(n^{-1/4})$ 。这种高度精确性的内在解释，被哈代描述为「不可思议的」（uncanny），它由拉德马赫（Rademacher）于 1937 年给出 [68]：存在一个收敛级数，它精确地表示 $p(n)$ ，并且这个级数起初几乎与那个渐近级数完全相同。还有另外一组问题，哈代–利特尔伍德方法对它们能给出精确的公式；这些问题关心的是把一个数表示为给定个数的平方之和。对于更高次幂，似乎不大可能存在这样的公式。

谁是 Ramanujan，这篇 1917 年论文为何是「起点」 S. Ramanujan（拉马努金，1887–1920）是印度传奇数学天才，几乎自学成才，1914 年被 Hardy 邀请到剑桥合作。他们 1917 年那篇论文研究的是分拆数 $p(n)$ 。为了求出 $p(n)$ 的精确近似，他们发明了一种「把信息装进一个函数、再用积分

把它取出来」的技巧——这个技巧的内核，正是后来圆法的雏形。所以 Davenport 说方法的「起源 (genesis, 意为创世、起源)」在这篇文章里。

分拆数 $p(n)$ 到底是什么 (读作「p of n」) $p(n)$ 表示：把正整数 n 写成若干个正整数之和的方法总数，其中不计较加数的次序 ($3 + 1$ 和 $1 + 3$ 算同一种)，加数个数不限。

$$p(1) = 1$$

只有 1。

$$p(2) = 2$$

2、 $1 + 1$ 。

$$p(3) = 3$$

3、 $2 + 1$ 、 $1 + 1 + 1$ 。

$$p(4) = 5$$

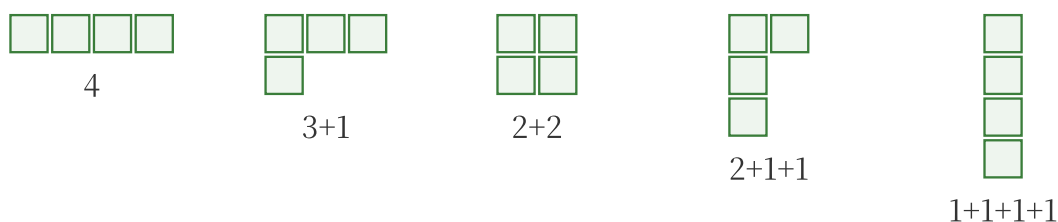
4、 $3 + 1$ 、 $2 + 2$ 、 $2 + 1 + 1$ 、 $1 + 1 + 1 + 1$ 。

$$p(5) = 7$$

5、 $4 + 1$ 、 $3 + 2$ 、 $3 + 1 + 1$ 、 $2 + 2 + 1$ 、 $2 + 1 + 1 + 1$ 、 $1 + 1 + 1 + 1 + 1$ 。

这正是第 1 段说的「被加数个数不受限」的分拆问题。注意 $p(n)$ 增长得飞快： $p(10) = 42$ ， $p(100) = 190569292$ （接近两亿）！手数是数不过来的，所以才需要一个公式去逼近它。

$n = 4$ 的全部 5 种分拆（每种用一摞方块表示， $p(4)=5$ ）



把每个加数画成一行方块、由大到小从上往下码放，就得到分拆的「Ferrers 图」。

$n = 4$ 共有 5 张这样的图，故 $p(4) = 5$ 。这是「不计次序」的直观表示。

$e^{A\sqrt{n}}$ 是什么量级的增长（先认识 $\sqrt{\cdot}$ 、指数函数） e 是自然常数 ≈ 2.718 ， e^x 是指数函数，随 x 增大而**爆炸式**变大。 $\sqrt{n}$ （根号 n ）是 n 的平方根。所以 $e^{A\sqrt{n}}$ （ A 是某个正的常数）意思是： $p(n)$ 大约随「 e 的『 A 乘根号 n 』次方」增长。

- 它比任何多项式 n^d 都快（因为指数最终碾压幂函数）；
- 但又比纯指数 e^{cn} 慢（因为指数里是 $\sqrt{n}$ 而非 n ）。

这解释了上面看到的现象： $p(n)$ 涨得很猛（百以内就近两亿），但还没猛到「翻倍式」纯指数那么夸张。准确的公式是 $A = \pi\sqrt{2/3}$ ，但本章不需要这个细节，认识「这是一种介于多项式和纯指数之间、相当快的增长」就够了。

大 O 记号 $O(\cdot)$ ：误差「不超过某量级」的速记（高中没学，从零讲） 写「误差为 $O(n^{-1/4})$ 」是什么意思？这是数论里描述**误差大小上界**的标准记号，读作「大 O 的 n 的负四分之一次方」。

定义（直观版）

说「某个量 $= O(g(n))$ 」，意思是：当 n 足够大时，这个量的**绝对值不会超过**「某个固定常数 C 」乘以 $g(n)$ 。用式子写就是「存在常数 C ，使

$|\text{误差}| \leq C g(n)$ 」。它只关心量级（涨落的快慢），不关心那个常数 C 具体是多少。

$n^{-1/4}$ 是多大

$n^{-1/4} = \frac{1}{\sqrt[4]{n}}$ (n 的四次方根的倒数)。 n 越大它越接近 0。例如

$n = 10000$ 时 $\sqrt[4]{n} = 10$, 故 $n^{-1/4} = 0.1$ 。

顺带认识 $\ll$ (维诺格拉多夫记号)

本章正文没用到它，但后面章节会频繁出现，先认个脸：「 $A \ll B$ 」与「 $A = O(B)$ 」意思完全相同，都表示「当变量足够大时， $|A| \leq C B$ 对某个固定常数 C 成立」。Davenport 后文常用 $\ll$ 来代替 $O(\cdot)$ ，二者可以互换着读。

所以「在最小项处截断、误差 $O(n^{-1/4})$ 」是说：他们的近似公式算出来的值，与真实的 $p(n)$ （一个整数）之间的差距，会随 n 增大而趋于 0。差距小于 0.5 时，把近似值四舍五入到最近的整数就恰好得到 $p(n)$ ！这就是为什么 Hardy 觉得「不可思议 (uncanny)」——一个本是「近似」的级数，竟精确到能算出整数答案。

「渐近级数」与「收敛级数」的区别，以及 Rademacher 补上的最后一块

渐近级数 (asymptotic series)

一种特殊的无穷级数：它整体加起来未必收敛（项数取太多反而会发散变糟），但只取前面有限几项、在「最小的那一项」处停下来，就能给出极好的近似。Hardy-Ramanujan 求得的就是这种。「在最小项处截断」正是它的标准用法。

收敛级数 (convergent series)

真正的无穷级数：一直加下去，和会越来越接近一个确定的值。

Rademacher (拉德马赫) 1937 年把 Hardy-Ramanujan 的渐近级数改造成了一个真正收敛的级数，它精确等于 $p(n)$ （不再是「近似」），而且这个收敛级数的开头几项几乎和原来的渐近级数一模一样。这就从根本上解释了那个「不可思议的精确」：因为它背后本来就藏着一个精确的收敛级数。

这段对全书的意义：它示范了圆法的「最高境界」——不仅能估计、有时还能给出**精确公式**。但作者紧接着提醒：这种「精确」是**奢侈品**，只在两类问题上出现：① 分拆数 $p(n)$ ；② 把数表示成「平方和」($k = 2$)。对 $k \geq 3$ 的更高次幂，「精确公式」基本无望，只能退而求「渐近公式 + 误差估计」——这正是本书第 4 章以后要做的事。先记住这个「精确是例外、估计是常态」的基调。

关于「平方和」精确公式的背景补充 原文说「把数表示成给定个数的平方之和」也能得到精确公式。例如：把 n 写成**4 个平方之和**的方法数，有 Jacobi 的漂亮公式（与 n 的因数之和有关）；写成 2 个、3 个平方之和也都有经典结果。这些属于 19 世纪就发展起来的「平方和理论」。圆法能把它们统一地重新推出来，是它的一个「加分项」。但同样地，一旦次数升到立方 ($k = 3$) 及以上，这种古典的精确公式就不复存在了。

原文第 4 段 · 在讲什么 说明圆法不止能做华林问题：把它推广到「更一般的加性方程」原则上不难，并写出两个推广形式。

华林问题关心的是特殊的丢番图方程 (1.1)。把哈代–利特尔伍德方法推广到处理更一般的加性型方程¹，原则上并无困难，例如

$$N = f(x_1) + f(x_2) + \cdots + f(x_s),$$

其中 $f(x)$ 是取整数值的多项式；特别地，推广到方程

$$N = a_1 x_1^k + a_2 x_2^k + \cdots + a_s x_s^k. \quad (1.2)$$

什么是「丢番图方程」(Diophantine equation) 得名于古希腊数学家丢番图 (Diophantus)。它指的是只允许在**整数（或正整数）范围内求解的方程**。同一个方程，若允许取小数解可能轻而易举，但限定「解必须是整数」往往就变得极难。式 (1.1) 就是一个丢番图方程： $N = x_1^k + \cdots + x_s^k$ ，要求 x_i 全是**正整数**。圆法本质上就是一套数「**整数解有多少组**」的解析工具。

「取整数值的多项式 $f(x)$ 」与第一个推广式

多项式 $f(x)$

就是形如 $f(x) = c_d x^d + \cdots + c_1 x + c_0$ 的式子（高中熟悉的「关于 x 的多项式」）。

「取整数值」（taking integer values）

要求：只要 x 代入整数，算出来的 $f(x)$ 也一定是整数。注意系数本身不一定是整数也能满足这点！例如 $f(x) = \frac{x(x+1)}{2}$ （第 x 个三角形数）系数是分数 $\frac{1}{2}$ ，但 x 取整数时 $\frac{x(x+1)}{2}$ 永远是整数。这类多项式叫「整值多项式」。

第一个推广式 $N = f(x_1) + \cdots + f(x_s)$ 的意思是：「指定类型的数」从「 k 次幂 x^k 」换成「任意整值多项式的值 $f(x)$ 」。显然取 $f(x) = x^k$ 就退回华林问题 (1.1)，所以它确实是 (1.1) 的推广。

逐符号看推广式 (1.2)：带系数的华林方程

$$N = a_1 x_1^k + a_2 x_2^k + \cdots + a_s x_s^k.$$

与 (1.1) 相比，唯一的新东西是每一项前多了**系数** $a_1, a_2, \dots, a_s$ （一组给定的整数，例如 $a_1 = 2, a_2 = 3, \dots$ ）。

- 当所有 $a_i = 1$ 时，(1.2) 就退回 (1.1)。所以 (1.2) 把 (1.1) 包含在内，是更一般的形式。
- 例 ($k = 2, s = 3$): $N = 2x_1^2 + 3x_2^2 + 5x_3^2$ ，问哪些 N 能这样表示、有几种。

为什么作者要把式子编号成 (1.2) 并强调它？因为本书前半部分的主攻对象就是 (1.2) 这类「带系数的加性方程」（下一段会明说）。它比 (1.1) 一般，又仍保持「各项分开、互不相乘」的**加性**结构，所以圆法照样适用——这正是「原则上并无困难」的含义：方法的骨架不变，只是计算时得多带着系数 a_i 走。

为什么说「原则上并无困难」却仍值得专门讲？ 「原则上 (of principle) 并无困难」是数学家的常用说法，意思是大思路照搬即可、不需要全新的想法；但「原则上不难」不等于「实际操作不费事」——把系数 a_i 一路带进所有估计里，许多细节会变复杂。这就是为什么后面要花好几章专门处理 (1.2)。先建立这个预期：骨架通用，血肉因题而异。

原文第 5 段·在讲什么 区分「加性方程」和「一般（非加性）方程」：后者直到近年才有大进展；本书先讲加性的，因为一般情形都建立在加性情形之上。

然而，只是在最近这些年里，才在把该方法改造以适用于一般的（即非加性的）丢番图方程方面取得了长足的进展。关于这些发展的叙述将在后面的讲座中给出，但我们起初将主要关注华林问题以及形如 (1.2) 的加性方程。所有关于一般丢番图方程的工作都严重依赖于加性方程方面工作的方法或结果。

「加性」与「非加性」方程的本质差别

加性方程 (additive)

各个未知数彼此分离、从不相乘。式 (1.1)、(1.2) 都是：每一项只含一个变量 (x_1 归 x_1 , x_2 归 x_2)，变量之间没有 x_1x_2 这种「交叉乘积」。这种「各变量各管一摊」的结构，是圆法最初能施展的关键。

一般（非加性）方程 (general / non-additive)

出现了变量之间的乘积/交叉项，例如二次型 $x_1x_2 + x_3x_4 = N$ 或一般的齐次多项式方程。这时变量「纠缠在一起」，圆法的原始版本不能直接套用，需要更精巧的改造——这正是「只是在最近这些年里」才取得进展的部分（本书后半 13 章以后讲的「三次型」等就属于此类）。

「先加性、后一般」的安排逻辑——本书的路线图 这段其实是 Davenport 给出的全书阅读顺序的说明：

1 先讲**华林问题 (1.1)** 和**加性方程 (1.2)**（前半部分）。原因：它们结构最干净，是练基本功的地方。

2 再讲**一般（非加性）方程**（后半部分）。

3 而第 2 步之所以放后面，不只是「由易到难」，更因为**一般方程的处理「严重依赖（depends heavily）」加性方程的方法或结果**——也就是说，做一般方程时，会反复把它归约到、或借用加性方程已经建立好的工具。所以加性部分不是「热身」，而是后面一切的地基。

读到这里你就明白：前面几章哪怕看起来只在折腾「 k 次方之和」这种特殊问题，其实是在为后面更一般的方程攒弹药。**不要跳过加性部分。**

原文第 6 段 · 在讲什么 最后预告一个更弱的主题：丢番图「不等式」。它也有一般结果，但不如「方程」情形完整、精确。

最后，我们将略微谈及丢番图不等式（Diophantine inequalities）这一课题。在这里，同样地，现在已经知道了一些一般性的结果，但它们不及方程情形的结果那样完整和精确。

从「方程」到「不等式」：要求被放松了

丢番图方程

要求某个表达式**精确**等于一个整数 N ，如 $x_1^k + \cdots + x_s^k = N$ 。

丢番图不等式（Diophantine inequality）

把「 $=$ 」放松成「**充分接近**」。典型形式是：要求 $|\lambda_1 x_1^k + \cdots + \lambda_s x_s^k| < \varepsilon$ ，其中系数 λ_i 可以是**无理数**， ε 是一个很小的正数。这时右边不再是整数 N ，问题变成「能否让这个带无理系数的式子**任意接近 0**（或接近某个目标值）」。

为什么要放松成不等式？因为当系数是无理数时，「精确等于某个整数」可能根本不可能或意义不大，转而问「能多接近」才是自然的问题。本书最后一章（第 20 章「A Diophantine inequality」）就讲这个。

为什么不等式情形「不及方程那样完整、精确」 方程情形里，圆法的核心工具是「**精确捕捉整数**」的积分 $\int_0^1$ （下面的预告框会讲它为什么能数整数解）。而不等式情形里没有这么干净的「整数捕捉器」，只能用更柔性的近似手段，所以结论往往是「存在解」而难以给出解的个数的**精确渐近公式**——这就是「不那么完整、不那么精确」的具体含义。作者把它放在「最后、略微谈及（touch on）」，正反映了它在本书中是**较次要、较前沿、较不成熟**的话题。

脚注

1. 原文脚注 1：「参见专著 [63]。」——这里 [63] 是参考文献编号，指向书末文献表中的某本专著（monograph，即专门论述某一主题的学术著作），用来支撑「圆法可推广到更一般加性方程」这一说法。读到这种 [数字] 时，到全书的「参考文献（References）」页按编号查即可。↩ [返回正文](#)

承上启下：本章埋下的伏笔——「圆法」的「圆」在哪里？

本章正文一次都没真正写出圆法的核心机器，但「circle method（圆法）」这个名字、以及第 3 段提到的「用积分把信息取出来」，都在向你预告后面几章会冒出的几个超纲符号。这里先把它**从零认一遍**，等第 4 章正式用到时你就不陌生了。这一节属于「补全原文省略的动机」，请配合下图理解。

① **复指数 $e(\alpha) = e^{2\pi i \alpha}$ ：把「角度」变成「单位圆上的点」** 这是全书最关键的记号，读作「e 的 $2\pi i \alpha$ 次方」，作者用简写 $e(\alpha)$ 代表它。

i (虚数单位)

满足 $i^2 = -1$ 的「数」。复数就是形如 $a + bi$ 的数 (a, b 为实数)。这在高中拓展内容里会见到。

欧拉公式 (Euler's formula)

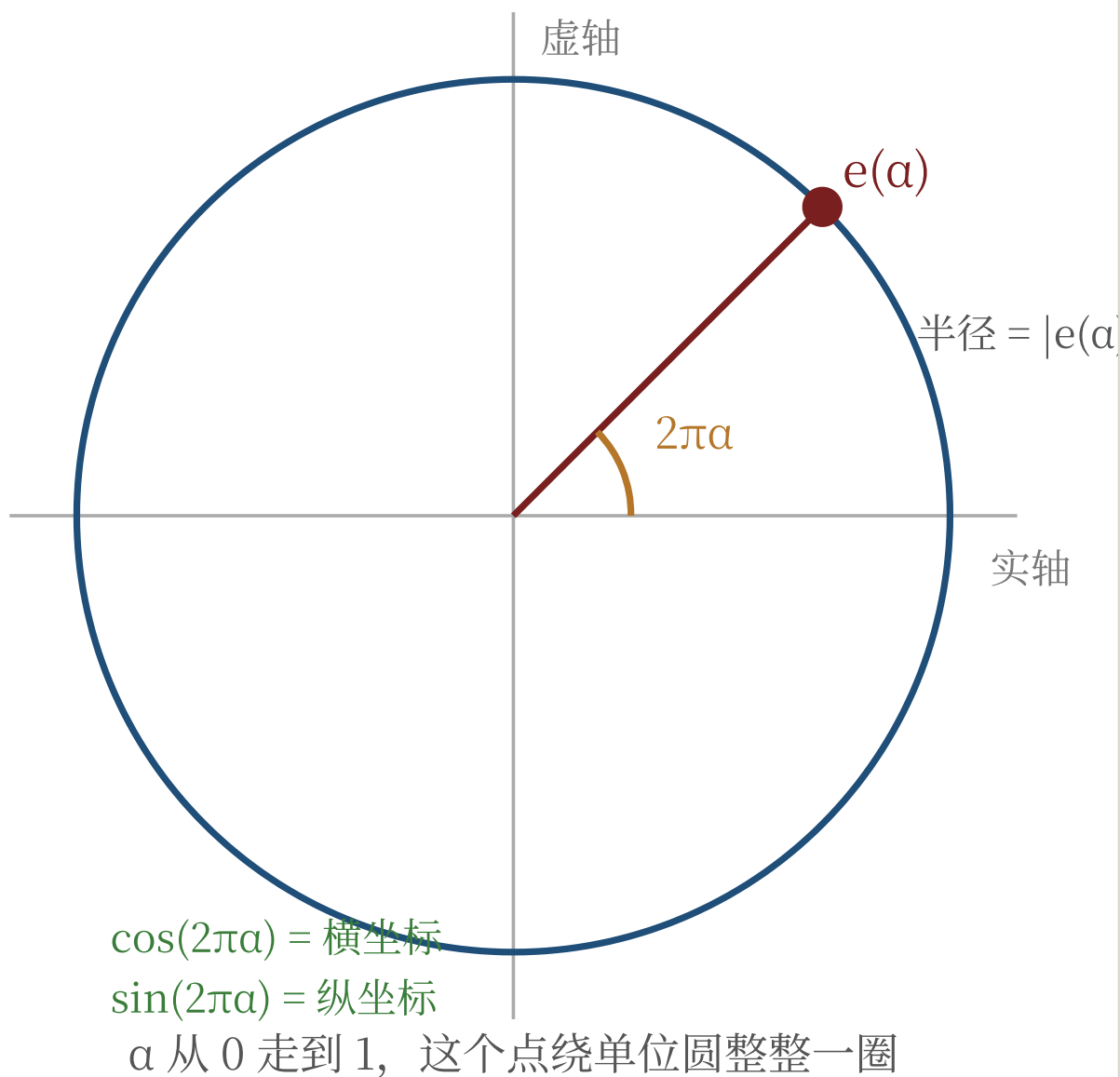
有一条桥梁公式: $e^{i\theta} = \cos \theta + i \sin \theta$ 。它把「指数」和「三角函数」连了起来。代入 $\theta = 2\pi\alpha$ 就得到

$$e(\alpha) = e^{2\pi i \alpha} = \cos(2\pi\alpha) + i \sin(2\pi\alpha).$$

模长 (绝对值) $|e(\alpha)|$

复数 $a + bi$ 的「模长」是 $\sqrt{a^2 + b^2}$, 表示它在复平面上到原点的距离。由 $\cos^2 + \sin^2 = 1$ 得 $|e(\alpha)| = \sqrt{\cos^2(2\pi\alpha) + \sin^2(2\pi\alpha)} = 1$ 。所以 $e(\alpha)$ 永远落在「以原点为心、半径为 1 的圆」(单位圆) 上!

当 α 从 0 增加到 1, 角度 $2\pi\alpha$ 从 0 转到 2π (整整一圈), 点 $e(\alpha)$ 沿单位圆绕行一周。「圆法」的「圆」就是这个单位圆——这就是名字的来历。



$e(\alpha) = \cos(2\pi\alpha) + i \sin(2\pi\alpha)$ 始终在单位圆上。圆法的全部把戏，都发生在 α 跑遍 $[0, 1]$ （即点绕圆一周）的过程中。

② 求和号 $\sum$ 与生成函数 $T(\alpha)$ ：把所有解「打包」进一个函数

$\sum$ （西格玛，求和号）

$\sum_{x=1}^P a_x$ 读作「对 x 从 1 到 P 求和」，就是 $a_1 + a_2 + \cdots + a_P$ 的速记。它

只是「把一串数加起来」的简写。

生成函数 $T(\alpha)$ （generating function，后面第 4 章的主角）

把所有 k 次方都塞进复指数再加起来：

$$T(\alpha) = \sum_{x=1}^P e(\alpha x^k) = e(\alpha \cdot 1^k) + e(\alpha \cdot 2^k) + \cdots + e(\alpha \cdot P^k).$$

为什么要造这么个怪东西？因为它是个「编码器」：把「哪些 k 次方可用」这件事，全部编码进了一个 α 的函数里。**动机**是——只要把 s 个这样的函数相乘，乘积 $T(\alpha)^s$ 展开后的每一项，恰好对应式 (1.1) 的一种拆法。这样「数拆法」就变成了「分析一个函数」，这正是「解析方法」(analytic method) 名称的由来。

③ 定积分 $\int_0^1$ ：一台「只认 N 、不认其它」的过滤器

$\int_0^1 \cdots d\alpha$ (从 0 到 1 的定积分)

高中里积分大致理解为「曲线下的面积 / 连续求和」。这里只需记住它的一条**神奇性质** (可由欧拉公式直接算出)：

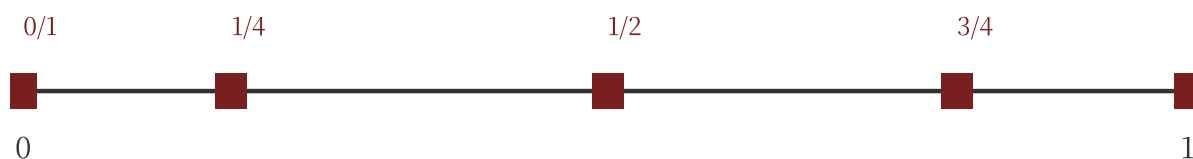
$$\int_0^1 e(m\alpha) d\alpha = \int_0^1 e^{2\pi i m \alpha} d\alpha = \begin{cases} 1, & m = 0, \\ 0, & m \neq 0 \text{ (整数)}. \end{cases} \quad (\star)$$

为什么 (★) 成立？给一个高中能懂的说法： $e(m\alpha)$ 是单位圆上转 m 圈的点；当 $m \neq 0$ 时，它在圆上**转了整数圈、对称地分布**，「正负相消」，平均 (积分) 下来是 0；而 $m = 0$ 时它恒等于 1，积分自然是 1。

这台过滤器有什么用？ 把它和生成函数配合：

$$r(N) = \int_0^1 T(\alpha)^s e(-N\alpha) d\alpha.$$

展开 $T(\alpha)^s e(-N\alpha)$ 后，每一项长得像 $e((x_1^k + \cdots + x_s^k - N)\alpha)$ 。由 (★)：只有当指数里的 $x_1^k + \cdots + x_s^k - N = 0$ (即这组 x 恰好拼出 N) 时，这一项积分才贡献 1；否则贡献 0。于是整个积分 $r(N)$ 正好等于「 N 的表示法总数」！这就是圆法的**心脏**：用一个积分，把「数解的个数」这件离散的、看似要逐个去试的事，变成了一道**可以用分析手段去估计的积分题**。本章第 3 段说的 Hardy-Ramanujan 「用积分把信息取出来」，指的就是这个机制。



红色小段 = 主弧 $\mathfrak{M}$ (靠近简单分数, 贡献主项)
 其余整条线 = 次弧 $\mathfrak{m}$ (贡献是要被压下去的「噪声」)

圆法把积分区间 $[0, 1]$ (等价于单位圆一整圈) 切成两类: 主弧 (major arcs, α 靠近分母小的分数处) 和次弧 (minor arcs, 其余部分)。「主弧出主项、次弧当噪声」是后面所有章节反复出现的总框架——本章只是先让你知道它的存在。

本章小结: 你现在应当能答出

1. **圆法是干嘛的:** 回答「把大数 N 拆成指定类型的数之和, 能不能拆、有几种拆法」这类**加性问题** (华林、哥德巴赫)。
2. **它从哪来:** 胚胎是 1917 年 Hardy-Ramanujan 研究分拆数 $p(n)$ 的工作; 经 Hardy-Littlewood 系统化、Vinogradov 大力推进。
3. **「精确」是奢侈品:** 只有 $p(n)$ 和「平方和」能得精确公式; $k \geq 3$ 一般只能求渐近公式加误差估计。
4. **本书路线:** 先加性方程 (1.1)、(1.2), 再一般 (非加性) 方程, 最后略谈不等式——因为后者都建立在前者之上。
5. **核心机器 (伏笔):** 复指数 $e(\alpha)$ 把角度变成单位圆上的点; 生成函数 $T(\alpha)$ 把所有解打包; 积分 $\int_0^1$ 凭性质 (★) 把「数解的个数」过滤出来; 再把 $[0, 1]$ 分成主弧 / 次弧分别处理。这套机器从第 4 章起会被一步步拆开来看。

带着这五条, 去读下一章吧。

返回 [全书目录](#)