

DAVENPORT · 圆法 · 高中详解版

更高次的齐次方程

Homogeneous equations of higher degree

本章要解决什么 · 读完能掌握什么

前面几章我们已经用「圆法」(Hardy-Littlewood circle method) 成功处理了单个三次齐次方程 (一个三次型 $= 0$ 的整数解)。本章问: 能不能把同样的方法推广到任意高次 (k 次)、而且多个方程联立 (R 个方程) 的情形? 这就是 B. J. Birch 1962 年那篇著名论文的内容。

读完本章你会明白三件事:

1. 会遇到两个新的拦路虎: 高次时"同余条件"不再自动满足; 而且即使局部 (实数 + 每个素数) 都有解, 整体 (整数) 也可能无解——本章给出一个石破天惊的反例 (19.1)。
2. Birch 的新武器是"维数": 他用一个叫奇异轨迹 V^* (singular locus) 的几何对象的维数来度量"方程组有多坏", 并据此给出一个可解性判据。
3. 最终定理 19.1: 只要变量个数 n 比奇异轨迹维数 s 大得足够多 ($n - s > R(R + 1)(k - 1)2^{k-1}$), 再加上局部可解性, 就保证有非零整数解。

本章在原书里是一篇"综述式"的章节——Davenport 只勾勒思路、不给全部证明。所以我们的讲解重点是: 把每一个被略过的符号、每一步"由此推出"背后的逻辑, 统统补成高中生能跟上的完整链条。

阅读方式: 黑色引用块是译文原话 (逐段搬来), 其后的彩色框是面向高中生的逐句详解。凡原文写"显然""容易看出""直接推广"的地方, 我们都补成一步一步的推导。

零、先补三块高中没学过的"地基"

本章满纸都是 $e(\alpha)$ 、 $\int_0^1$ 、 $\sum$ 、 $\ll$ 、"同余"、" p -进"、"维数"这些记号。它们超出高中课本，但每一个都能从零讲清。先把最常用的几块地基打好，后面读正文就顺了。

记号 $e(\alpha)$ ，读作"**e 括号 alpha**"。

定义 $e(\alpha) = e^{2\pi i \alpha}$ 。这里 i 是虚数单位 ($i^2 = -1$)， π 是圆周率， α 是一个实数。由欧拉公式 $e^{i\theta} = \cos \theta + i \sin \theta$ ，所以

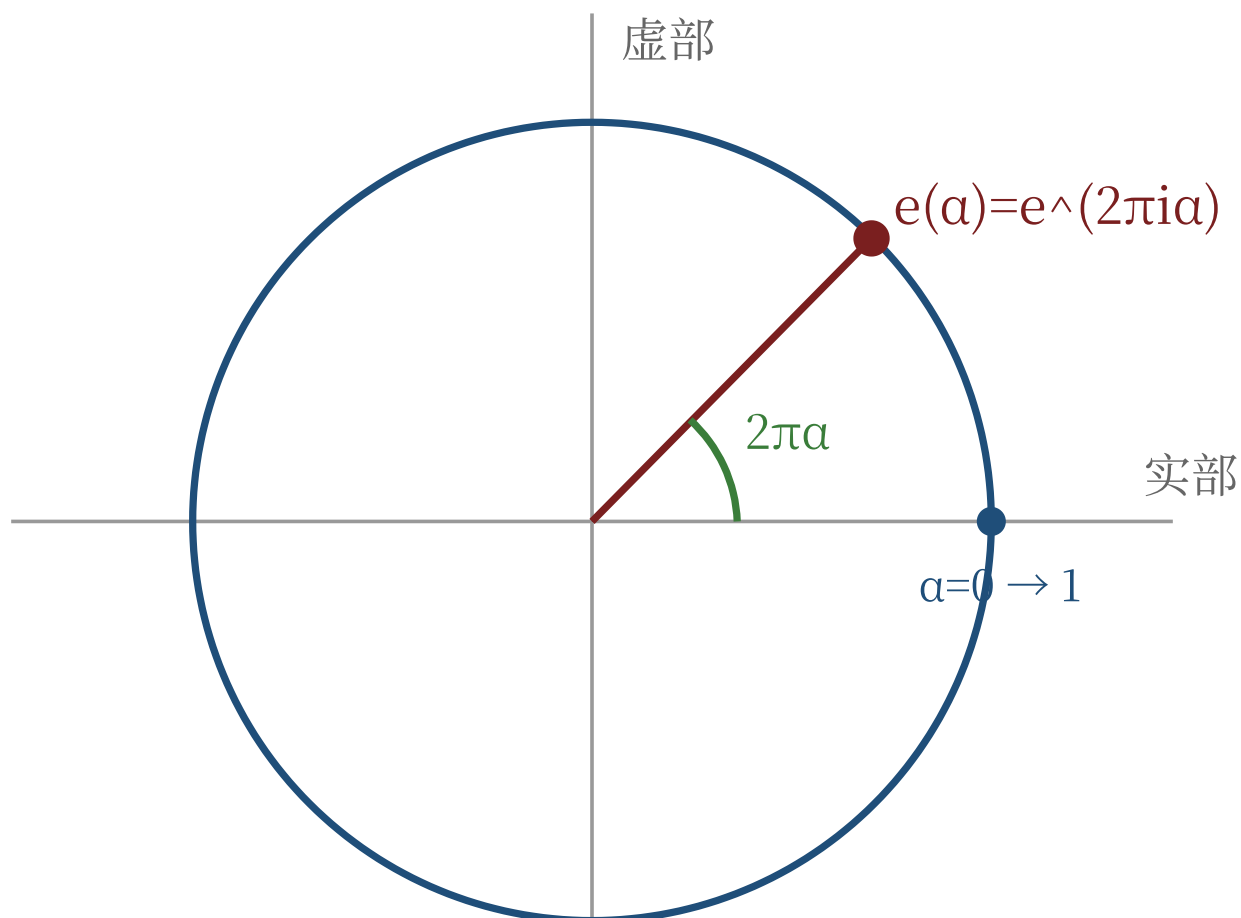
$$e(\alpha) = \cos(2\pi\alpha) + i \sin(2\pi\alpha).$$

它的几何意义：把它画在复平面（横轴实部、纵轴虚部）上， $e(\alpha)$ 永远落在**单位圆**（半径 1 的圆）上，对应的角度是 $2\pi\alpha$ 弧度。 α 每增加 1，角度就转满一圈 2π ，转回原点。所以 $e(\alpha)$ 只关心 α 的**小数部分**： $e(\alpha + \text{整数}) = e(\alpha)$ 。

为什么圆法要用它？关键性质有两条：（一）**模长恒为 1**，即 $|e(\alpha)| = 1$ （点在单位圆上，到原点距离永远是 1）；（二）**正交性**：对整数 m ，

$$\int_0^1 e(m\alpha) d\alpha = \begin{cases} 1, & m = 0, \\ 0, & m \neq 0. \end{cases}$$

正是这条"是 0 就积出 1，不是 0 就积出 0"的性质，让积分能精确地"数出"方程的解的个数。下面会反复用到。



点永远在单位圆上， $|e(\alpha)|=1$ ； α 增加 1 转满一圈

$e(\alpha) = e^{2\pi i \alpha}$ 把实数 α "卷"到单位圆上：角度 $= 2\pi\alpha$ ，模长恒为 1。 $\alpha = 0, 1, 2, \dots$ 都回到点 1。

求和号 $\sum$ 与定积分 $\int_0^1$ 。

$\sum_x g(x)$ 表示把变量 x 跑遍指定范围时所有 $g(x)$ 的值加起来。例如 $\sum_{x=1}^3 x^2 = 1 + 4 + 9 = 14$ 。本章里 x 常常是一个向量（即一组坐标 $\mathbf{x} = (x_1, \dots, x_n)$ ），" $\sum_{\mathbf{x} \in P\mathfrak{B}}$ "就是让这组整点跑遍某个盒子里所有格点。

$\int_0^1 F(\alpha) d\alpha$ 是定积分，几何上是函数 F 在区间 $[0, 1]$ 上"曲线下方的面积"（被积函数取复数值时，就是实部、虚部各自积分）。可以把它想成"把 $[0, 1]$ 切成无穷多小段、每段取值乘以段宽再相加"的极限。圆法之所以叫"圆法"，正是因为 α 跑遍 $[0, 1]$ 等价于 $e(\alpha)$ 跑遍整个单位圆一圈——积分就是沿这个圆"绕一圈求平均"。

记号 $O(\cdot)$ 与 $\ll$ (数量级 / 不等式的"宽松版")。

这两个记号是解析数论的家常便饭，意思一样，只是写法不同。写 $A \ll B$ (读作" A 远小于等于 B "，等价于 $A = O(B)$) 的意思是：存在一个与"主变量"无关的常数 $C > 0$ ，使得 $|A| \leq CB$ 恒成立。它不在乎"具体的常数倍是多少"，只关心"增长的快慢量级"。例如 $3n + 5 \ll n$ (取 $C = 8$ 即可，对 $n \geq 1$)。本章里主变量是 P (盒子的边长，最后要让 $P \rightarrow \infty$)，所以诸如 $q \ll P^{R(k-1)\theta}$ 的意思是" q 至多和 P 的那个幂同一量级"。

与之配套的还有 $\gg$ (远大于等于， $A \gg B$ 即 $B \ll A$) 和 ε (读作"epsilon"，表示一个可以任意小的正数，常用来吸收掉无关紧要的微小幂次)。

记号 $\|\beta\|$ ：到最近整数的距离。

对实数 β ，记 $\|\beta\|$ 为 β 与离它最近的整数之间的距离。例如 $\|3.2\| = 0.2$ ， $\|3.8\| = 0.2$ ， $\|5\| = 0$ 。它永远在 $[0, \frac{1}{2}]$ 之间。这个量衡量" β 离整数有多近"——在圆法里， $\|\beta\|$ 很小，意味着 $e(\beta)$ 离 1 很近，指数和才可能很大。

"同余"与" p -进可解" (congruence / p -adic)。

同余：说 $a \equiv b \pmod{m}$ (读作" a 与 b 模 m 同余")，意思是 $a - b$ 能被 m 整除，也就是 a, b 除以 m 余数相同。例如 $17 \equiv 2 \pmod{5}$ 。

同余条件 / p -进可解：要让方程 $f(\mathbf{x}) = 0$ 有整数解，一个必要前提是：对每个素数 p 和每个 p 的幂 p^t ，同余方程 $f(\mathbf{x}) \equiv 0 \pmod{p^t}$ 都要有解 (而且我们想要的是"不退化、能往上一层 p^{t+1} 继续抬"的解，这叫非奇异解)。"对一切 p^t 都可解"这件事，用更高级的语言说就是方程在 p -进数域 $\mathbb{Q}_p$ 中可解。 p -进数你可以暂时理解为"把同余信息对一切 p^t 同时打包"的一种数。直觉：整数解必须"在实数里说得通" (实可解)，还要"对每个素数说得通" (p -进可解) —— 这些叫局部条件；而真正要的"整数解"叫整体。本章的大主题正是：局部都行，整体却未必行。

"齐次型"与"维数"。

k 次齐次型 (form of degree k): 就是每一项的次数都恰好是 k 的多项式。例如 $x^2 + xy + y^2$ 是二次型, $x^3 - 2y^3 + z^3$ 是三次型。"齐次"保证: 若 $\mathbf{x}$ 是解, 则 $\lambda\mathbf{x}$ 也是解 (因为 $f(\lambda\mathbf{x}) = \lambda^k f(\mathbf{x})$), 所以解总是成"射线"出现, 我们关心的是**非零解**。

代数簇 (variety): 一组多项式方程的公共零点集合, 是几何里"曲线、曲面"的高维推广。**维数** (dimension): 粗略说就是"这个解集里能自由变动的参数个数"——一条曲线维数 1, 一张曲面维数 2, 整个 n 维空间维数 n 。一般地, n 元空间里由 R 个"独立"方程切出来的簇, 维数是 $n - R$ (每加一个真正起作用的方程, 自由度就少 1)。这个"减 1"的直觉, 是本章几何论证的核心, 后面会反复用。

一、引言：把三次方程的方法推广到高次

在文献 [6] 中, Birch 给出了对我们处理齐次三次方程所用方法的一个意义深远的推广, 但这一推广涉及一些重要的修改。他考虑求解一个齐次方程, 或一组联立的齐次方程 (它们的次数全都相同) 的问题。这里人们面临两个严重的困难。首先, 即便对于单个次数为 $k > 3$ 的方程, 我们一般也不知道任何关于 k 的合理函数 $n_0(k)$, 使得当 $n \geq n_0(k)$ 时, 同余条件对每个素数都能被满足。(由 Brauer 的工作 (已在第 11 章中引用), 我们知道存在某个 k 的函数, 使得方程在 p -进域中可解。此外, 仅仅在 p -进域中有解还不够; 我们需要一个非奇异解, 以便确保能够满足同余条件。) 因此, 我们必须假定对每个素数 p , 同余条件都被满足。我们还必须假定该方程, 或方程组, 在实数域中可解, 并且具有一个非奇异解。

这一段在讲什么: 交代背景与"第一个困难"。前面几章 (第 15–17 章) 处理的是"一个三次型"。Birch 想一口气推广到"任意 k 次、任意 R 个方程联立"。但一上来就撞到两堵墙, 这段讲第一堵。

文献 [6]、Birch

B. J. Birch (伯奇), 英国数论家。这里的 [6] 指他 1962 年发表的论文 *Forms in many variables* (《多变量的型》)。Davenport 这本讲义的最后一章, 就是把 Birch 这篇论文的思路做一个综述。

k

方程的**次数** (degree)。三次就是 $k = 3$ ，本章要处理 $k > 3$ 乃至一般的 k 。

n

变量个数 (即 $\mathbf{x} = (x_1, \dots, x_n)$ 的坐标数)。圆法的灵魂是"变量足够多时方程一定有解"，所以 n 越大越好办。

$n_0(k)$

一个"门槛函数"：希望找到这样一个只依赖 k 的数，使得"只要变量数 $n \geq n_0(k)$ ，同余条件就自动对每个素数满足"。困难在于：高次时**没人知道**这样一个合理（不大得离谱）的 $n_0(k)$ 。

Brauer 的工作

R. Brauer (布饶尔) 证明了：确实**存在**某个仅依赖 k 的门槛，使方程 p -进可解。但他给出的数值"astronomical (天文数字般大)"，大到没有实用价值。所以 Davenport 说"我们知道存在，但用不上"。

为什么三次时没这个烦恼，高次却有？

对三次方程，前面章节能证明：只要变量足够多，对每个素数 p 的同余条件可以自动验证（这背后用到 Chevalley-Warning 一类定理：变量数大于次数时，模 p 同余方程必有非零解）。但当 k 增大，要保证"对一切 p 、一切幂次 p^t 都有**非奇异解**"所需的变量数会失控，谁也写不出一个好用的公式 $n_0(k)$ 。

于是 Birch 干脆**把它当作假设直接列出来**：不去费力证明同余条件能自动满足，而是把"对每个 p 同余条件满足"和"实数域中有非奇异解"作为**定理的前提**写进去。这是一种诚实的"取舍"：把暂时证不出来的部分明确地假设掉，让方法的主体先跑起来。

"非奇异解"是什么、为什么非要它不可？

一个解 $\mathbf{a}$ (满足 $f(\mathbf{a}) = 0$) 叫**非奇异** (non-singular)，是指在这个点，梯度（所有偏导数 $\partial f / \partial x_1, \dots, \partial f / \partial x_n$ ）**不全为零**。几何上：该点处曲面是"光滑"的、有正常的切平面，不是尖点或自交点。

为什么需要非奇异？ 因为我们要把"模 p 的解"一层层抬到"模 $p^2, p^3, \dots$ 的解"（这套抬升机制叫 *Hensel 引理*，本质就是高维的牛顿迭代）。抬升能进行下去的条件，正是这一点处导数不全为零（分母不为零）。若解是奇异的（导数全 0），抬升就卡住，"对一切 p^t 可解"就无法保证。所以这里反复强调：要的不只是有解，而是**非奇异的解**。

其次——而这一点更为重要——即便这些假定也不总是足以保证方程在整数（或有理数）中可解。下面这个例子是 Swinnerton-Dyer 给我看的：

这一段在讲什么：抛出"第二个、也是更要命的困难"——**局部—整体原则会失效**。这是整章思想转折的关键，下面用一个具体反例砸实。

Swinnerton-Dyer

H. P. F. Swinnerton-Dyer（斯温纳顿-戴尔），英国数论家（也是著名的 Birch-Swinnerton-Dyer 猜想里的那位）。他给 Davenport 看了一个精心构造的反例。

"局部—整体原则"（Hasse 原则）

一个美好的**愿望**：如果方程在实数里有解（"在无穷远处局部可解"），且对每个素数 p 都 p -进可解（"每个素数处局部可解"），那么它就应该在**有理数 / 整数**里有解（整体可解）。对二次型这是对的（Hasse-Minkowski 定理）。但下面这个例子说明：对**三次及更高次**，这个美好愿望会**破产**。

为什么这个反例如此重要 它直接说明：光有前一段那些局部假设（实可解 + 每个 p 可解）**还不够**，必须再加新条件。Birch 论文的全部技术，正是为了找出那个"额外条件"——而这就引出了本章的主角"奇异轨迹的维数"。

$$3(x_1^2 + \cdots + x_r^2)^3 + 4(x_{r+1}^2 + \cdots + x_s^2)^3 = 5(x_{s+1}^2 + \cdots + x_n^2)^3, \quad (19.1)$$

其中 $r < s < n$ 。由 Selmer [78] 的工作可知，方程

$$3X^3 + 4Y^3 = 5Z^3$$

除了 $X = Y = Z = 0$ 之外无解，由此可推出 (19.1) 除了 $x_1 = \cdots = x_n = 0$ 之外无解。另一方面，可以证明 (19.1) 对每个 p 满足同余条件，并且它在实数域中当然也是非奇异可解的。

这一段在讲什么：把上面那个"反例"完整摆出来，并解释它为什么是反例。我们逐块拆解。

先读懂方程 (19.1) 的结构。

它是一个 n 元的方程。把 n 个变量分成三组：第一组 $x_1, \dots, x_r$ ，第二组 $x_{r+1}, \dots, x_s$ ，第三组 $x_{s+1}, \dots, x_n$ （条件 r 非负的（实数的平方非负）。每一组平方和立方后，整体是关于 $(x_1, \dots, x_n)$ 的 6 次齐次型（平方是 2 次，再立方乘 3， $2 \times 3 = 6$ ），所以这其实是个 $k = 6$ 次的方程，刚好属于"高次"的范畴。

记号 [78]、Selmer，以及那个三元三次方程。

[78] 指 E. S. Selmer（塞尔默）的工作。他证明了一个惊人的事实：三元三次方程

$$3X^3 + 4Y^3 = 5Z^3$$

只有平凡解 $X = Y = Z = 0$ （在整数、乃至有理数范围内）。（英文 PDF 此处把右端误印成 $5Y^3$ ，正确应为 $5Z^3$ ，译文已订正。）而妙在：这个方程对每个素数 p 都 p -进可解、也实可解——它正是 Hasse 原则失效的经典例子之一（与更有名的 Selmer 例子 $3X^3 + 4Y^3 + 5Z^3 = 0$ 同类）。Selmer 例子本身就告诉我们：三次时局部—整体已经会崩。

为什么 " $3X^3 + 4Y^3 = 5Z^3$ 无解"能推出 (19.1) 无解？把略过的逻辑补全：

1 设 (19.1) 有一组整数解 $(x_1, \dots, x_n)$ 。记 $A = x_1^2 + \dots + x_r^2$ ， $B = x_{r+1}^2 + \dots + x_s^2$ ， $C = x_{s+1}^2 + \dots + x_n^2$ 。这三个都是非负整数。

2 代入 (19.1) 得 $3A^3 + 4B^3 = 5C^3$ 。也就是说 $(X, Y, Z) = (A, B, C)$ 是三元三次方程 $3X^3 + 4Y^3 = 5Z^3$ 的一组整数解。

3 但 Selmer 已证该三次方程只有平凡解，故必须 $A = B = C = 0$ 。

4 而"若干个实数的平方和 $= 0$ ", 当且仅当每个数都为 0 (平方非负, 要加起来为零只能各自为零)。所以 $A = 0 \Rightarrow x_1 = \cdots = x_r = 0$; 同理 $B = 0, C = 0$ 逼出其余变量也全为 0。

5 于是 $x_1 = \cdots = x_n = 0$, 即 (19.1) 除平凡解外无解。◆

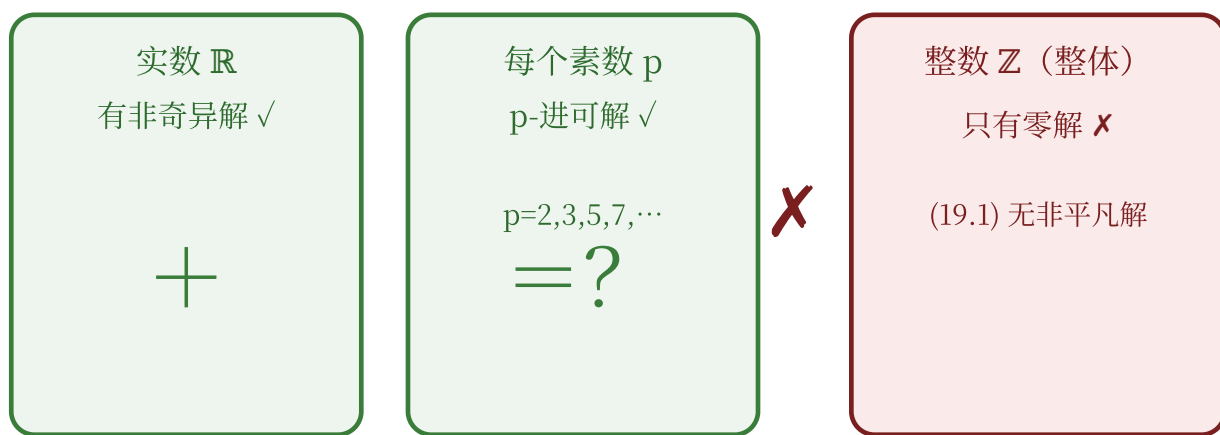
这一步设计极其巧妙: 用"平方和"这个**非负**构件, 把 n 个变量"压缩"成三个非负整数 A, B, C , 从而把一个 n 元方程的无解性, 归结到 Selmer 那个三元方程的无解性。

那为什么它又"局部处处可解"?

原文只说"可以证明它对每个 p 满足同余条件, 实数域中也非奇异可解", 把验证略去了。直觉如下:

- **实可解**: 当变量个数 n 足够大、三组都非空时, 三个平方和 A, B, C 可以取到任意非负实数, 于是 $3A^3 + 4B^3 = 5C^3$ 在实数里随便就能配平 (例如先取 A, B , 解出 $C = \sqrt[3]{(3A^3 + 4B^3)/5}$), 而且能取在非奇异点上。
- **每个 p 可解**: 当每组的变量数足够多时, "平方和"在模 p^t 下能取遍足够多的值 (平方和的取值很"丰富"), 于是总能凑出满足同余的解。变量越多, 这种自由度越大——这正是"圆法相信变量多就好办"的局部版本。

结论: (19.1) **局部处处有 (非奇异) 解, 整体却只有零解**。这就是一个干净利落的"Hasse 原则反例", 把"光靠局部条件"的幻想彻底打碎。



局部处处可解，整体却无解 —— Hasse 原则失效

反例 (19.1)：左、中两块"局部条件"全部满足，却推不出右边的"整体有解"。这正是必须追加新条件的

原因。

因此，如果我们要建立整数可解性，就必须施加某个进一步的条件。Birch 被引导施加的那类条件，是用与方程组相关联的一个'奇异轨迹'的维数来表述的。

这一段在讲什么：点题——既然局部条件不够，就得加一条**新条件**；而 Birch 的新条件是用"奇异轨迹的维数"来写的。这句话是全章的"路标"。

预告：什么是"奇异轨迹" (singular locus) V^* ?

这里先给个直觉，正文后面会给出精确定义。回忆"非奇异点 = 梯度不全为零的点"。反过来，**奇异点**就是梯度（对方程组而言是那个偏导数矩阵）"退化"的点。把方程组所有这样的坏点收集起来，组成的集合就叫**奇异轨迹 V^*** 。它越"大"（维数越高），方程组就越"病态"、越难处理。Birch 的洞见是：**只要 V^* 足够小（维数 s 足够低），圆法就能突破，整体可解性就能成立**。反例 (19.1) 的毛病，恰恰就出在它的奇异轨迹太大。

我们将概述他的论文的总体规划，在可能有帮助的地方与单个三次方程的问题作比较。由于记号不可避免地复杂，细节相当繁难。

这一段在讲什么：是作者的"免责声明 + 阅读指引"。Davenport 坦言：下面只给**提纲 (outline)**，不给全部证明，因为记号太繁、细节太硬。他会不时拿"单个三次方程"（即

$R = 1, k = 3$ 的旧情形) 来对照, 帮你定位。给读者的建议: 下面每出现一个新符号, 先盯住"它退化到 $R = 1, k = 3$ 时是什么", 就能把陌生公式接回前几章学过的东西。

二、搭台：型、指数和、与"用积分数解"

设有 R 个 n 元 k 次齐次型, 其中 $R < n$ 。我们可以把它们写成

$$\begin{aligned} f^{(1)}(\mathbf{x}) &= \sum_{j_0, \dots, j_{k-1}} C_{j_0, \dots, j_{k-1}}^{(1)} x_{j_0} \cdots x_{j_{k-1}}, \\ &\vdots \\ f^{(R)}(\mathbf{x}) &= \sum_{j_0, \dots, j_{k-1}} C_{j_0, \dots, j_{k-1}}^{(R)} x_{j_0} \cdots x_{j_{k-1}}, \end{aligned}$$

其中求和的变量从 1 到 n 。

这一段在讲什么: 正式把"研究对象"写下来—— R 个 k 次型。这串带上标下标的公式看着吓人, 其实只是"把一个 k 次齐次多项式展开成单项式之和"的标准写法。

R

方程 (型) 的个数。 $R = 1$ 就回到"单个方程"。条件 \((R

$f^{(1)}, \dots, f^{(R)}$

这 R 个齐次型本身。上标 (i) 是"第几个型"的编号, 不是幂次。

$\mathbf{x} = (x_1, \dots, x_n)$

变量向量, 加粗的 $\mathbf{x}$ 表示"一整组坐标"。

$C_{j_0, \dots, j_{k-1}}^{(i)}$

第 i 个型里、单项式 $x_{j_0} x_{j_1} \cdots x_{j_{k-1}}$ 前面的系数 (已知的整数)。下标 $j_0, \dots, j_{k-1}$ 各自从 1 跑到 n , 指出这一项乘了哪几个变量。

把这个吓人的展开式"翻译"成人话

取最简单的 $k = 2$ (二次型)、 $n = 2$ 来体会。这时单项式是 $x_{j_0} x_{j_1}$, 下标 $j_0, j_1 \in \{1, 2\}$, 于是

$$f = \sum_{j_0, j_1} C_{j_0 j_1} x_{j_0} x_{j_1} = C_{11} x_1 x_1 + C_{12} x_1 x_2 + C_{21} x_2 x_1 + C_{22} x_2 x_2.$$

合并后就是普通的二次型 $C_{11}x_1^2 + (C_{12} + C_{21})x_1x_2 + C_{22}x_2^2$ 。可见这套" k 个下标连乘"的写法，无非是把一个 k 次齐次多项式的**每一个单项式**都拆成" k 个变量按下标相乘"的形式，再把所有组合加起来。它写起来啰嗦，但好处是：求偏导、做"多重线性化"时特别整齐——这正是后面要用的。

设 $\mathfrak{B}$ 为 n 维空间中的一个盒子，定义指数和

$$S(\alpha_1, \dots, \alpha_R) = \sum_{\mathbf{x} \in P\mathfrak{B}} e\left(\alpha_1 f^{(1)}(\mathbf{x}) + \dots + \alpha_R f^{(R)}(\mathbf{x})\right).$$

这一段在讲什么：引入圆法的核心工具——**指数和 S** （生成函数）。这是整套方法的"发动机"。

$\mathfrak{B}$ （花体 B）

n 维空间里一个固定的**盒子**（box），就是各坐标都被限制在某区间里的"长方体"，比如 $[-1, 1]^n$ 。它规定解的"形状范围"。

P

一个大的正参数（伸缩倍数）。 $P\mathfrak{B}$ 就是把盒子 $\mathfrak{B}$ **放大 P 倍**得到的大盒子。我们最后要让 $P \rightarrow \infty$ ，看大盒子里解的个数怎么增长。

$\sum_{\mathbf{x} \in P\mathfrak{B}}$

让 $\mathbf{x}$ 跑遍大盒子 $P\mathfrak{B}$ 里所有**整点**（坐标都是整数的格点）。盒子边长 $\sim P$ ，所以这样的整点约有 P^n 个。

$\alpha_1, \dots, \alpha_R$

R 个实参数，每个跑遍 $[0, 1]$ 。它们是圆法里要"积分掉"的变量——每个方程配一个 α 。

$S(\alpha_1, \dots, \alpha_R)$

指数和：把大盒子里每个整点 $\mathbf{x}$ ，算出 $e(\alpha_1 f^{(1)}(\mathbf{x}) + \dots + \alpha_R f^{(R)}(\mathbf{x}))$ （一个单位圆上的复数），再全部加起来。

为什么要造这个 S ？动机是什么？

核心动机是把"数方程的解"这个离散的、看不出规律的问题，变成"算一个积分"这个可以用分析工具（估计、逼近）下手的问题。桥梁正是第 0 节讲的正交性：对整数 m ， $\int_0^1 e(m\alpha) d\alpha$ 在 $m = 0$ 时为 1、在 $m \neq 0$ 时为 0。把它推广到 R 个变量，就能让积分像"探测器"一样，只把那些让所有 $f^{(i)}(\mathbf{x}) = 0$ 的整点挑出来计 1，其余计 0。下一段公式就是这个魔法。

那么， $P\mathfrak{B}$ 中满足联立方程 $f^{(1)}(\mathbf{x}) = 0, \dots, f^{(R)}(\mathbf{x}) = 0$ 的整点 $\mathbf{x}$ 的个数由下式给出

$$\mathcal{N}(P) = \int_0^1 \cdots \int_0^1 S(\alpha_1, \dots, \alpha_R) d\alpha_1 \cdots d\alpha_R.$$

这一段在讲什么：亮出圆法的"基本恒等式"——解的个数 = 指数和的多重积分。这是后面一切工作的出发点。原文没给推导（认为是"直接的"），我们补全。

$\mathcal{N}(P)$

大盒子 $P\mathfrak{B}$ 里、同时满足全部 R 个方程 $f^{(i)}(\mathbf{x}) = 0$ 的整点个数。这正是我们最终想知道大于 0 的量（有解 $\iff \mathcal{N}(P) > 0$ ）。

补全：为什么这个积分恰好等于解的个数

1 先看一维正交性。对任意整数 m ，

$\int_0^1 e(m\alpha) d\alpha = \int_0^1 \cos(2\pi m\alpha) + i \sin(2\pi m\alpha) d\alpha$ 。当 $m \neq 0$ ， $\cos, \sin$ 在整数个周期上积分为 0；当 $m = 0$ ，被积函数恒为 $e(0) = 1$ ，积分得 1。所以 $\int_0^1 e(m\alpha) d\alpha = [m = 0]$ （方括号表示"成立取 1，否则取 0"）。

2 推广到 R 维。因为各 α_i 相互独立，多重积分可拆成连乘：

$$\int_0^1 \cdots \int_0^1 e(\alpha_1 m_1 + \cdots + \alpha_R m_R) d\alpha_1 \cdots d\alpha_R = \prod_{i=1}^R \int_0^1 e(\alpha_i m_i) d\alpha_i.$$

由上一步，这个乘积只有当每个 $m_i = 0$ 时才为 1，否则为 0。即它等于 " $m_1 = \cdots = m_R = 0$ " 这件事的指示器。

- 3 代入 S 的定义并交换求和与积分。把 $S = \sum_{\mathbf{x}} e(\alpha_1 f^{(1)}(\mathbf{x}) + \dots)$ 代入，由于是有限和，可与积分交换次序：

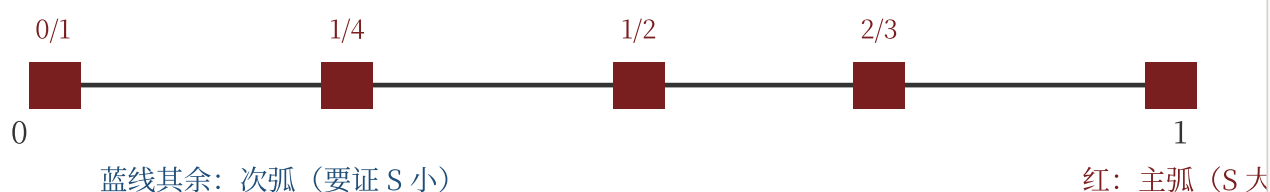
$$\mathcal{N}(P) = \sum_{\mathbf{x} \in P\mathfrak{B}} \int_0^1 \cdots \int_0^1 e(\alpha_1 f^{(1)}(\mathbf{x}) + \cdots + \alpha_R f^{(R)}(\mathbf{x})) d\alpha_1 \cdots d\alpha_R.$$

- 4 用第 2 步的指示器。这里 $m_i = f^{(i)}(\mathbf{x})$ 都是整数（整点代入整系数型）。于是每个整点 $\mathbf{x}$ 的那块积分，等于 " $f^{(1)}(\mathbf{x}) = \cdots = f^{(R)}(\mathbf{x}) = 0$ 吗？是则 1，否则 0"。

- 5 求和。于是 $\mathcal{N}(P) = \sum_{\mathbf{x} \in P\mathfrak{B}} [\mathbf{x} \text{ 是公共解}] = (\text{公共解的个数})$ 。证毕。 ◆

这就是圆法的精髓："积分 = 数解"。剩下的全部工作，就是想办法估计这个积分有多大，从而证明 $\mathcal{N}(P) > 0$ 。估计的标准套路是把积分区间 $[0, 1]^R$ 切成两类区域：主弧（major arcs, α 接近"分母小的有理点", S 很大, 贡献主项）和次弧（minor arcs, 其余地方, 要证 S 很小、贡献可忽略）。本章的全部技术困难，都在"怎么在次弧上把 S 估小"。

把 $[0, 1]$ 切成主弧 + 次弧 ($R=1$ 的示意)



圆法把积分区间 $[0, 1]^R$ 划成"主弧"（围绕分母小的有理点，贡献主项）与"次弧"（剩余部分，需证可忽略）。本章的硬骨头全在次弧上把 $|S|$ 估小。

三、核心估计：从" S 大"挤出"对 α 的有理逼近"

由引理 13.1 的一个直接推广，我们发现，如果

$$|S(\alpha_1, \dots, \alpha_R)| \geq P^{n-K}, \quad (K > 0),$$

那么

$$\sum_{\mathbf{x}^{(1)}} \cdots \sum_{\mathbf{x}^{(k-1)}} \prod_{J=1}^n \min \left\{ P, \left\| \alpha_1 M_J^{(1)} + \cdots + \alpha_R M_J^{(R)} \right\|^{-1} \right\} \gg P^{nk-2^{k-1}K},$$

这一段在讲什么：这是全章技术主线的**第一步**。它说的是一个"二择"逻辑链的起点：*假如*指数和 S 还不算小 ($\geq P^{n-K}$)，那么必然能推出右边一个关于"线性型靠近整数"的大下界。这是 Weyl (外尔) 差分法的标准产物。

P^{n-K} 与 K

S 的"自然大小上限"约是 P^n (因为是 P^n 个模长为 1 的复数相加)。 $K > 0$ 衡量" S 比满额 P^n 小了多少个幂次"。" $|S| \geq P^{n-K}$ "就是说" S 还没小到 K 那么多", 即**仍然偏大**。 K 越大, 这个假设越苛刻。

$\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k-1)}$

$k-1$ 个**独立的整点向量** (每个都是 n 维)。它们来自"**Weyl 差分**": 把一个 k 次型反复做差分 $k-1$ 次, 每差一次次数降 1, 最后降成关于这 $k-1$ 个辅助向量的**多重线性**对象。

$M_J^{(i)}$

差分后得到的**多重线性型** (下段会精确定义)。上标 (i) 对应第 i 个方程, 下标 J (从 1 到 n) 对应第 J 个坐标方向。

$\min\{P, \|\cdots\|^{-1}\}$

取" P "和"那个线性型到最近整数距离的倒数"中的**较小者**。若 $\|\cdots\|$ 很小 (线性型很贴近整数), 倒数很大, 但被 P 截住, 所以这一项最大就是 P 。它度量"线性型有多接近整数", 越接近、这项越大。

2^{k-1}

这个出现在指数上的 2^{k-1} , 正是"差分 $k-1$ 次、每次平方放大一次损耗"的代价: 每做一次 Weyl 差分 (用一次 Cauchy-Schwarz), 损耗就翻倍, 做 $k-1$ 次就乘出 2^{k-1} 。当 $k=3$ 时 $2^{k-1}=4$, 正好对上前面三次方程那一章里的 4。

"引理 13.1 的直接推广"是什么意思?

引理 13.1 是 Davenport 前面处理单个三次方程 ($R = 1, k = 3$) 时证的核心引理：若那时的指数和不少，就能得到类似的右边下界。Birch 的"推广"是把它从 " $R = 1, k = 3$ " 原样升级到"任意 R 、任意 k "——证明思路一模一样（反复用 Weyl 差分 + Cauchy-Schwarz），只是要把"一个方程"换成" R 个方程的线性组合 $\alpha_1 f^{(1)} + \dots + \alpha_R f^{(R)}$ "，把"差分 2 次"换成"差分 $k - 1$ 次"。所以原文说"直接推广"，意思是不需要新想法、照搬即可，故不再重写证明。

这一步的"剧情作用" 把抽象的"指数和 S 偏大"这个解析信息，**翻译**成一个算术信息："有非常多组辅助整点 $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k-1)}$ ，使得那些线性型 $\alpha_1 M_J^{(1)} + \dots$ 同时很贴近整数"。后面就靠"很多组 + 很贴近"这两件事，反推出 α_i 本身必须接近一个分母不大的有理数——也就是说，" S 大"只可能发生在主弧上。这正是次弧估计的标准武器。

其中 $M_J^{(1)}, \dots, M_J^{(R)}$ 是 $k - 1$ 个点 $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k-1)}$ 的多重线性型，由下式定义

$$M_J^{(i)}(\mathbf{x}^{(1)} \mid \dots \mid \mathbf{x}^{(k-1)}) = \sum_{j_1, \dots, j_{k-1}} c_{J, j_1, \dots, j_{k-1}}^{(i)} x_{j_1}^{(1)} \dots x_{j_{k-1}}^{(k-1)},$$

其中 $i = 1, \dots, R$ 。引理 13.1 本身就是 $R = 1, k = 3$ 的情形。

这一段在讲什么：精确写出上一步里出现的"多重线性型 $M_J^{(i)}$ "是怎么从原来的型 $f^{(i)}$ 来的。

"多重线性型" (multilinear form) 是什么、为什么会冒出来？

"多重线性"意思是：把 $M_J^{(i)}$ 看成 $k - 1$ 个向量 $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k-1)}$ 的函数时，**固定其余、只动其中任意一个，它都是线性的**（满足 $f(\dots, a\mathbf{u} + b\mathbf{v}, \dots) = af(\dots, \mathbf{u}, \dots) + bf(\dots, \mathbf{v}, \dots)$ ）。公式里每一项 $x_{j_1}^{(1)} x_{j_2}^{(2)} \dots x_{j_{k-1}}^{(k-1)}$ 都是"从每个向量各取一个坐标相乘"，所以对每个向量都是一次的——这就是多重线性。

它怎么来的？ 把一个 k 次型 $f^{(i)}$ 反复做"差分"（把变量换成 $\mathbf{x} + \mathbf{x}^{(j)}$ 再相减，每次消掉一层），做 $k - 1$ 次之后，原来 k 次的型就被"剥"成关于 $k - 1$ 个增量向量的多重线性

型。系数 $c_{J,j_1,\dots,j_{k-1}}^{(i)}$ 由原系数 $C^{(i)}$ 经差分组合而来（差一个常数倍 $k!$ 之类，对量级估计无影响）。下标 J ：差分一次会"留下一个自由坐标方向"， J 就标这个方向，从 1 到 n 共 n 个，所以才有 $\prod_{J=1}^n$ 那 n 个因子。

用最熟悉的二次型把"差分降次"看明白

取 $k=2$ 、单变量 $f(x)=x^2$ 。做一次差分： $f(x+h)-f(x)=(x+h)^2-x^2=2xh+h^2$ 。其中关于 h 的"主部"是 $2xh$ ——它对 x 和 h 各是一次，这就是一个双线性型，对应 $k-1=1$ 个增量。再看 $k=3$ 、 $f(x)=x^3$ ：差两次会得到 $\sim 6h_1h_2x$ 这种"对 h_1, h_2 各一次"的项——这正是引理 13.1 ($R=1, k=3$) 里 M_J 的雏形。一般 k 次型差 $k-1$ 次，就降成对 $k-1$ 个增量各一次的多重线性型。这就是为什么辅助向量恰好有 $k-1$ 个。

正如在引理 13.2 中那样，由此推出，满足

$$|\mathbf{x}^{(1)}| < P, \dots, |\mathbf{x}^{(k-1)}| < P,$$

$$\|\alpha_1 M_J^{(1)} + \dots + \alpha_R M_J^{(R)}\| < P^{-1}, \quad (1 \leq J \leq n),$$

的 $k-1$ 个整点的组的个数

$$\gg P^{(k-1)n-2^{k-1}K-\varepsilon}.$$

这一段在讲什么：把上一步那个"乘积下界"翻译成更直观的**计数**结论：满足两类条件（向量不太长 + 线性型很贴近整数）的辅助整点组，**个数很多**。

$|\mathbf{x}^{(j)}|$

每个辅助向量的长度（各坐标绝对值）都

$\|\alpha_1 M_J^{(1)} + \dots + \alpha_R M_J^{(R)}\|$

那 R 个 α 加权出的线性型，到最近整数的距离小于 $1/P$ （非常贴近整数）。这对每个方向 $J=1, \dots, n$ 都要成立。

补全：从"乘积下界"到"计数下界"的逻辑（引理 13.2 型论证）

- 1 回到上一步的乘积 $\prod_{j=1}^n \min\{P, \|\cdots_j\|^{-1}\}$ 。每个因子最大是 P ，所以整个乘积最大是 P^n 。
- 2 要让乘积达到大下界 $\gg P^{nk-2^{k-1}K}$ ，就必须有很多组 $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k-1)})$ 让乘积接近满额 P^n ；而乘积接近 P^n ，要求每个因子都接近 P ，也就是每个 $\|\cdots_j\|^{-1} \gtrsim P$ ，即 $\|\cdots_j\| \lesssim P^{-1}$ 。这就是第二个条件的由来。
- 3 把"贡献大的那些组"数一数：总下界 $P^{nk-2^{k-1}K}$ 除以"每组至多贡献 P^n "，得到合格组数 $\gg P^{nk-2^{k-1}K}/P^n = P^{(k-1)n-2^{k-1}K}$ 。（这一步是粗略的"平均化"，严格处理会引入可忽略的 P^ε 因子，故下界写成 $P^{(k-1)n-2^{k-1}K-\varepsilon}$ 。）
- 4 ε 的出现：把诸如 $\log P$ 之类的"小毛刺"统一吸收进 P^ε （因为对任意固定 $\varepsilon > 0$ ，当 P 大时 $\log P \ll P^\varepsilon$ ）。这是解析数论里"把不重要因子扫进 ε "的惯用手法。 ◆

把引理 12.6 使用 $k-1$ 次（而不是像引理 13.3 的证明中那样使用两次），我们推出，满足

$$|\mathbf{x}^{(1)}| < P^\theta, \dots, |\mathbf{x}^{(k-1)}| < P^\theta,$$

$$\|\alpha_1 M_J^{(1)} + \cdots + \alpha_R M_J^{(R)}\| < P^{-k+(k-1)\theta}$$

的 $k-1$ 个整点的组的个数

$$\gg P^{(k-1)n\theta-2^{k-1}K-\varepsilon}.$$

这一段在讲什么：把上一步的结论"压缩"——把辅助向量的搜索盒子从边长 P 缩小到 P^θ （ θ 是 0 到 1 之间的小数），代价是线性型贴近整数的要求也相应收紧。这是为后面"提取有理逼近"做准备。

θ (读作"theta")

一个待定的参数， $0 < \theta < 1$ 。它控制"把辅助向量限制得多小"： P^θ 比 P 小（因为 $\theta < 1$ ）。选好 θ 是后面平衡各项幂次的关键调节旋钮。

$$P^{-k+(k-1)\theta}$$

收紧后的"贴近整数"门槛。把盒子缩小，能换来线性型更贴近整数（指数更负，门槛更苛刻），这正是"缩小搜索范围以提纯逼近质量"的交易。

引理 12.6

前面的一条工具引理（一种"格点缩放/转换"原理）：它能把"在大盒子里有很多贴近整数的组"转化为"在小盒子里也有相当多（贴得更近）的组"。这里要把它连用 $k - 1$ 次。

为什么是"用 $k - 1$ 次"，而不是三次方程那章的"两次"？

这正是原文特意点明的"取舍/对照"。在处理三次方程（ $k = 3$ ）的引理 13.3 时，辅助向量只有 $k - 1 = 2$ 个，所以引理 12.6 只需用**两次**（每个辅助向量缩一次）。现在 k 一般，辅助向量有 $k - 1$ 个，就得对每个都缩一次，故用 $k - 1$ 次。这是把"三次专用"升级成"一般 k 通用"时唯一需要改的地方——次数 k 决定了辅助向量个数，进而决定缩放次数。指数里的 θ 也因此从"2 次的形态"变成" $(k - 1)$ 次的形态"（如下界里出现 $(k - 1)n\theta$ ）。

四、二择：要么 S 小，要么得到对 α 的良好联立逼近

如果在这些 $k - 1$ 个点的组中存在某一个，使得矩阵

$$\begin{pmatrix} M_1^{(1)} & \cdots & M_1^{(R)} \\ \vdots & & \vdots \\ M_n^{(1)} & \cdots & M_n^{(R)} \end{pmatrix}$$

的秩为 R ，那么我们就得到对 $\alpha_1, \dots, \alpha_R$ 的良好有理逼近，它们都具有相同的分母 q 。这个分母来自上述矩阵中某个 R 阶行列式（非零）的值。事实上我们得到

$$|q\alpha_i - a_i| \ll P^{-k+R(k-1)\theta}$$

以及

$$q \ll P^{R(k-1)\theta}.$$

这一段在讲什么：到了"二择"逻辑的分叉口的第一支。上面已得到"很多组辅助点让 R 个线性型同时贴近整数"。现在问：这些组里有没有一组，使得那个 $n \times R$ 矩阵满秩（秩 $= R$ ）？若有，就能反解出 $\alpha_1, \dots, \alpha_R$ 都是"分母相同且不大"的近似有理数。

"秩" (rank) 是什么？

一个矩阵的秩，就是它的列（或行）里"真正相互独立、互不能由其它线性表出"的最大个数；等价地，是它含有的非零行列式的最大阶数。这里矩阵有 n 行、 R 列（ $n \geq R$ 线性无关，于是矩阵里能挑出一个 $R \times R$ 的非零子行列式。这个非零行列式的值，就用来当公共分母 q 。

q

公共分母：取自那个 R 阶非零行列式的值。因为辅助向量都 $\leq P^{-k}$

a_i

逼近 α_i 的那个有理数的分子（整数），即 $\alpha_i \approx a_i/q$ 。

$$|q\alpha_i - a_i| \ll P^{-k+R(k-1)\theta}$$

逼近的精度： α_i 与 a_i/q 的偏差小到这个量级。指数 $-k$ 来自上一步线性型那个很苛刻的贴近门槛 $P^{-k+(k-1)\theta}$ ，再经 Cramer 法则用分母 q 放大，凑出 $R(k-1)\theta$ 。

补全：满秩为什么能"解出"公共分母的逼近（Cramer 法则）

1 把" $\alpha_1 M_J^{(1)} + \dots + \alpha_R M_J^{(R)}$ 贴近整数 b_J "写成近似等式：对每个 J ，
 $\sum_{i=1}^R \alpha_i M_J^{(i)} \approx b_J$ (b_J 是最近的整数)，误差 $\leq P^{-k}$

2 因为矩阵秩为 R ，可从 n 个方向 J 里挑出 R 个，使对应的 $R \times R$ 子矩阵 $\mathbf{M}$ 的行列式 $q = \det \mathbf{M} \neq 0$ 。只看这 R 个方程，得到一个关于未知数 $\alpha_1, \dots, \alpha_R$ 的（近似）线性方程组 $\mathbf{M} \boldsymbol{\alpha} \approx \mathbf{b}$ 。

3 用 Cramer 法则（克莱姆法则）解： $\alpha_i \approx \det(\mathbf{M}_i) / \det(\mathbf{M}) = a_i/q$ ，其中 $\mathbf{M}_i$ 是把第 i 列换成整数向量 $\mathbf{b}$ 得到的矩阵， $a_i = \det(\mathbf{M}_i)$ 是整数（整数矩阵的行列式是整数）。于是 α_i 有了一个分母同为 q 的有理逼近。

- 4 估精度与分母大小： $q = \det \mathbf{M}$ 是 R 个量级 $P^{(k-1)\theta}$ 的元素的行列式，故 $q \ll P^{R(k-1)\theta}$ ；把第 1 步的小误差经行列式放大，得 $|q\alpha_i - a_i| \ll P^{-k+R(k-1)\theta}$ 。

一句话：矩阵满秩 $\Rightarrow$ 线性方程组可解 $\Rightarrow$ 反推出 α 的优质有理逼近。"优质"指分母小（ q 小）且逼得准（误差小），这正是" α 落在主弧上"的标志。

这里的指数分别对应于引理 13.4 的选择 B 中的 $-3 + 2\theta$ 与 2θ 。

这一段在讲什么：一句"对照注脚"，帮你把现在的一般公式接回三次方程那章。把 $R = 1, k = 3$ 代入：精度指数 $-k + R(k-1)\theta = -3 + 1 \cdot 2 \cdot \theta = -3 + 2\theta$ ；分母指数 $R(k-1)\theta = 2\theta$ 。完全吻合引理 13.4"选择 B"里的那两个数。这证实了"一般 k, R 的公式确实是三次情形的忠实推广"——退化回去就一字不差。

为什么管这叫"选择 B" 引理 13.4 把局面分成两种"选择" (alternative)：选择 A（矩阵不满秩、退化的坏情形）与选择 B（矩阵满秩、能得到好逼近的好情形）。我们现在做的正是"选择 B"那一支——好情形。下一节就要处理"选择 A"那个真正的拦路虎。

五、真正的困难：矩阵恒不满秩——引入"奇异轨迹 V^* "

真正的困难出现在上述情形失效时，即对所有的 $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k-1)}$ ，上述矩阵的秩都 $\leq R - 1$ 时。在 $R = 1$ 的情形，这将意味着多重线性型 M_J 在所有这些整点组上全都消失。

这一段在讲什么：正式进入"选择 A"——那个让前面方法失效的坏情形：无论怎么取辅助点组，矩阵的秩永远 $\leq R - 1$ （永不满秩）。这时第四节那套"反解出好逼近"的把戏就用不了，必须另想办法。

"秩永远 $\leq R - 1$ "在 $R = 1$ 时意味着什么？

当 $R = 1$ ，矩阵只有 1 列，秩要么是 0、要么是 1。"秩 $\leq R - 1 = 0$ "就等于"这一列全为零"，即所有 $M_J = 0$ （每个方向的多重线性型都恒等于零）。换句话说：差分出来的多重线性型在所有辅助点组上**整个塌掉**。这是最退化的情形——原文用它当作直觉锚点：一般 R 的"秩 $\leq R - 1$ "，就是这种"塌陷"的高维版本（列向量之间永远线性相关，凑不出满秩）。

为什么这是"真正的困难" 第四节靠"满秩 $\Rightarrow$ 好逼近"把坏的 α 关进主弧。若秩恒不满，这条路就断了——我们既得不到对 α 的逼近，也就无法断定 S 在次弧上很小，整个圆法卡死。Birch 的全部新意，就是为了绕开这一关：他不再纠缠于单个点，而是把"秩恒不满"翻译成**一个几何对象的维数问题**。

Birch 论文的主要新思想，是用簇（variety）的维数来表述这种可能性。我们把一组 $k - 1$ 个点看作 $(k - 1)n$ 维空间中的单个点。上述矩阵的秩应 $\leq R - 1$ 这一条件，在该空间中定义了一个代数簇；而由其上整点个数的下界，我们推出这个簇的维数

$$\geq (k - 1)n - 2^{k-1}K/\theta + \varepsilon.$$

这一段在讲什么：亮出 Birch 的**核心新思想**——把"坏情形"几何化为"一个簇的维数下界"。这是全章思想的最高点，要慢讲。

第一步妙手：把"一组点"看成"一个点"。

一组辅助点是 $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k-1)})$ ，每个 $\mathbf{x}^{(j)}$ 有 n 个坐标，合起来共 $(k - 1)n$ 个坐标。Birch 把这一整组坐标**打包成一个单独的点**，住在 $(k - 1)n$ 维的大空间里。于是"很多组辅助点"就变成"大空间里很多个整点"。

第二步妙手：用"很多整点"反推"维数很高"。

"秩 $\leq R - 1$ "这个条件，是若干个行列式 $= 0$ 的多项式方程，它在大空间里切出一个**代数簇 W** （坏点的集合）。现在我们已知（第三节末）满足这些条件的**整点非常多**（ $\gg P^{(k-1)n\theta - 2^{k-1}K - \varepsilon}$ ，在边长 P^θ 的盒子里）。一个一般性原理是：一个簇若在边长 P^θ

的盒里含有 $\gg (P^\theta)^d$ 个整点，则它的维数至少是 d （维数 d 的簇，盒里整点数大约就是（边长） d 量级；整点越多，维数越高）。把整点数 $P^{(k-1)n\theta-2^{k-1}K}$ 写成 $(P^\theta)^{(k-1)n-2^{k-1}K/\theta}$ ，读出指数就得到维数下界

$$\dim W \geq (k-1)n - \frac{2^{k-1}K}{\theta} + \varepsilon.$$

这就是把"解析的计数"翻译成"几何的维数"的关键一跃。

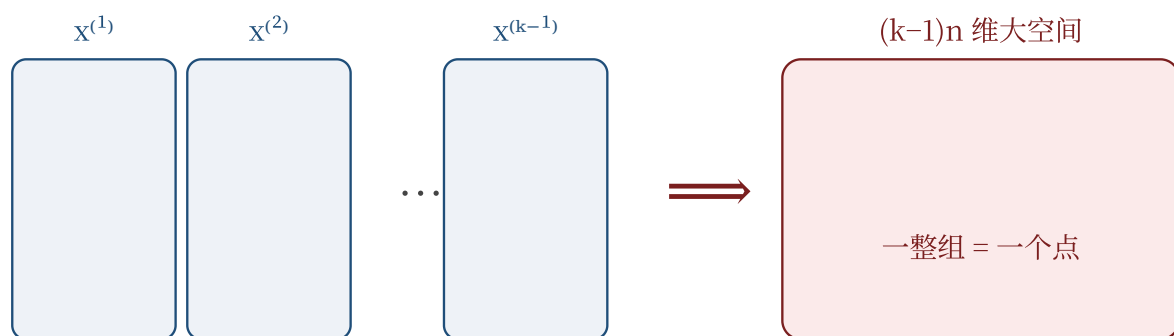
补全：那个"整点多 $\Rightarrow$ 维数高"的指数是怎么对上的

1 已知坏点（整点组）个数 $N \gg P^{(k-1)n\theta-2^{k-1}K-\varepsilon}$ ，且都落在边长 $L = P^\theta$ 的盒子里。

2 把整点数用盒子边长 L 作底来表达： $N \gg L^d$ ，求指数 d 。因 $L = P^\theta$ ，有 $L^d = P^{\theta d}$ 。令 $\theta d = (k-1)n\theta - 2^{k-1}K$ ，解得 $d = (k-1)n - 2^{k-1}K/\theta$ 。

3 由"维数 = 盒内整点数随边长增长的指数"这一原则，簇维数 $\geq d = (k-1)n - 2^{k-1}K/\theta$ （误差 ε 来自前面计数里的 ε ）。 ◆

注意 K 越大（即假设" S 越大"越苛刻），这个维数下界就越小——这层关系马上会被用来"选 K 把坏情形堵死"。



把 $k-1$ 个 n 维点拼成一个 $(k-1)n$ 维点；坏点集成一个代数簇

Birch 的几何化：把 $k-1$ 个 n 维辅助点拼成 $(k-1)n$ 维空间里的单个点；"秩 $\leq R-1$ " 的坏点全体构成一个代数簇，其维数可由整点计数下界估出。

代数几何中一个简单的原理是：若我们用一个由 t 个方程定义的线性空间去与一个簇相交，则簇的维数（即其任一绝对不可约分量的最大维数）至多减少 t 。因此，上述簇与'对角'线性空间

$$\mathbf{x}^{(1)} = \mathbf{x}^{(2)} = \dots = \mathbf{x}^{(k-1)}$$

（它由 $(k-2)n$ 个方程定义）的交，其维数

$$\geq n - 2^{k-1}K/\theta - \varepsilon.$$

这一段在讲什么：用一条"维数相交原理"，把上面那个住在 $(k-1)n$ 维空间的坏簇，切回到原来 n 维的变量空间。手法是和"对角线"相交——强行让 $k-1$ 个辅助点全部相等。

"切一刀维数最多掉1"的原理。

直觉：一个 d 维的簇，用一个方程去切（要求点落在某超平面上），结果维数最多降到 $d-1$ （少一个自由度，有时甚至不降）。用 t 个方程切，维数最多降 t 。注意是"最多减 t "（可能减得更少），所以下界方向是安全的： $\dim(\text{交}) \geq \dim(\text{原簇}) - t$ 。这里括注"绝对不可约分量的最大维数"只是把"维数"定义严谨化——簇可能由几块拼成，取最大那块的维数。

"对角"线性空间 $\mathbf{x}^{(1)} = \dots = \mathbf{x}^{(k-1)}$

要求 $k-1$ 个辅助点全相等的那一组点。把它们都设为同一个 $\mathbf{x}$ ，相当于在 $(k-1)n$ 维空间里嵌入一条 n 维的"对角线"子空间。

为什么是 $(k-2)n$ 个方程

"全相等" $\mathbf{x}^{(1)} = \mathbf{x}^{(2)}, \mathbf{x}^{(2)} = \mathbf{x}^{(3)}, \dots, \mathbf{x}^{(k-2)} = \mathbf{x}^{(k-1)}$ 一共 $k-2$ 个"相邻相等"，每个是 n 个坐标方程，合计 $(k-2) \times n = (k-2)n$ 个方程。

补全：相交后的维数下界怎么算出来

1 原坏簇维数 $\geq (k-1)n - 2^{k-1}K/\theta + \varepsilon$ 。

2 用 $t = (k-2)n$ 个方程（对角条件）去切，维数最多减 t ：

$$\dim(\text{交}) \geq [(k-1)n - 2^{k-1}K/\theta] - (k-2)n.$$

3 化简首尾的 n 项：

$$(k-1)n - (k-2)n = [(k-1) - (k-2)]n = 1 \cdot n = n. \text{ 所以}$$

$$\dim(\text{交}) \geq n - 2^{k-1}K/\theta - \varepsilon.$$

意义：经过这一刀，坏簇被搬回到原始的 n 维变量空间里——而落在对角线上的坏点，恰恰就是 " $k-1$ 个点全相等于某个 $\mathbf{x}$ " 的点。下一段就说明：这种点正是方程组的**奇异点**。

如果 $\mathbf{x} = \mathbf{x}^{(1)} = \dots = \mathbf{x}^{(k-1)}$ ，则这个新簇由所有满足下列条件的点 $\mathbf{x}$ 组成：
矩阵

$$\begin{pmatrix} \frac{\partial f^{(1)}}{\partial x_1} & \dots & \frac{\partial f^{(R)}}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial f^{(1)}}{\partial x_n} & \dots & \frac{\partial f^{(R)}}{\partial x_n} \end{pmatrix}$$

的秩 $\leq R-1$ 。我们称之为与给定方程相关联的**奇异轨迹**，并记为 V^* 。于是当前情形导致

$$\dim V^* \geq n - 2^{k-1}K/\theta - \varepsilon.$$

这一段在讲什么：揭晓谜底——切回对角线后得到的那个簇，正是**奇异轨迹** V^* （梯度矩阵退化的点集）。于是上一节的"维数下界"变成了 " $\dim V^*$ 的下界"。本章的中心概念终于精确定义。

为什么"全相等的多重线性型退化"= "梯度矩阵退化"？

关键是**欧拉关系**与多重线性型的对称性。把 $k-1$ 个辅助点都取成同一个 $\mathbf{x}$ ，多重线性型 $M_J^{(i)}(\mathbf{x}|\cdots|\mathbf{x})$ 就退化成只关于 $\mathbf{x}$ 的东西——而经过差分构造，它恰好正比于偏导数 $\partial f^{(i)}/\partial x_J$ 。理由：把 k 次型 $f^{(i)}$ 差分 $k-1$ 次再让所有增量相等，等价于对它求 $k-1$ 阶导的对角部分，而由齐次型的求导规律，这恰好回到一阶偏导（差一个常数因子）。所以 " $M_J^{(i)}$ 构成的矩阵秩 $\leq R-1$ " 在对角线上就变成 "偏导数矩阵 $(\partial f^{(i)}/\partial x_J)$ 秩 $\leq R-1$ "。

奇异轨迹 V^* 的精确定义与几何含义。

那个 $n \times R$ 的矩阵叫**雅可比矩阵** (Jacobian)，第 J 行第 i 列是 $\partial f^{(i)}/\partial x_J$ (第 i 个方程对第 J 个变量的偏导)。

$$V^* = \left\{ \mathbf{x} : \text{rank} \left(\partial f^{(i)}/\partial x_J \right) \leq R-1 \right\}.$$

几何含义：在 V^* 上， R 个曲面 $f^{(i)} = 0$ 的法向量"挤在一起、线性相关"，曲面相交得不"干净"——这些就是方程组的**奇异点**（非光滑、退化的点）。回忆第一节：抬升解、做圆法都需要"非奇异"，所以 V^* 就是"所有麻烦的来源"。 $\dim V^* = s$ 越小，麻烦越少。

把整条逻辑链串起来回看

" S 偏大" $\Rightarrow$ (Weyl 差分) 很多组辅助点让线性型贴近整数 $\Rightarrow$ 若有一组满秩则得 α 的好逼近 (好情形/主弧)；否则 (秩恒 $\leq R-1$ ，坏情形) $\Rightarrow$ 坏点簇维数高 $\Rightarrow$ (切对角线) $\Rightarrow$ **奇异轨迹 V^* 维数高**。于是 Birch 把"坏情形会不会发生"等价转化成 " V^* 到底有多大"。下一段就用这个转化来"主动堵死"坏情形。

六、关键取舍：选好 K ，把坏情形彻底堵死

如果 $\dim V^* = s$ ，我们就可以通过选取

$$K = \frac{\theta}{2^{k-1}}(n - s - 2\varepsilon)$$

来阻止这种情况发生（从而排除现在对应于引理 13.4 选择 A 的局面）。

这一段在讲什么：这是全章最精彩的“取舍”——既然坏情形会逼出“ $\dim V^* \geq n - 2^{k-1}K/\theta - \varepsilon$ ”，那我们就反过来选 K ，让这个下界超过已知的真实值 s ，造成矛盾，从而证明坏情形根本不可能发生。

$s = \dim V^*$

奇异轨迹的**真实维数**（这是方程组本身的固定属性，由 $f^{(i)}$ 决定，是定理里要用到的关键不变量）。

补全：为什么这样选 K 恰好“卡死”坏情形

1 坏情形若发生，必有 $\dim V^* \geq n - 2^{k-1}K/\theta - \varepsilon$ 。但 $\dim V^*$ 的真实值是 s ，所以坏情形发生**必须** $s \geq n - 2^{k-1}K/\theta - \varepsilon$ 。

2 我们想让这个不等式**不可能成立**——也就是反过来让 $n - 2^{k-1}K/\theta - \varepsilon > s$ （右边的“门槛”严格高于真实 s ）。整理： $2^{k-1}K/\theta < n - s - \varepsilon$ 。

3 把它取在临界附近（留出 2ε 余量保证严格）就得到原文的选择：

$$2^{k-1}K/\theta = n - s - 2\varepsilon \implies K = \frac{\theta}{2^{k-1}}(n - s - 2\varepsilon).$$

4 验证：代入后 $n - 2^{k-1}K/\theta - \varepsilon = n - (n - s - 2\varepsilon) - \varepsilon = s + \varepsilon > s$ 。于是“坏情形要求 $s \geq s + \varepsilon$ ”——这是**矛盾**！故坏情形（选择 A）**不可能发生**。◆

为什么这样取、不那样取： K 是“我们假设 S 有多大”的门槛。取得太大（要求 S 极大），坏情形的维数下界就压得很低、容易低于 s ，貌似更容易排除坏情形——但代价是“ $|S| \geq P^{n-K}$ ”这个前提会几乎对所有 α 都不成立，使结论变空洞、主弧太窄而无法拼出主项。取得太小，又排除不掉坏情形。所以 K 必须取在**临界点**：恰好让坏情形不可能、同时让“ S 不大”的范围尽量宽。原文选的 $K = \frac{\theta}{2^{k-1}}(n - s - 2\varepsilon)$ 就是这个最优平衡点。注意它**正比于** $n - s$ ——这第一次让“ $n - s$ ”（变量数减奇异轨迹维数）作为决定成败的关键量登场，后面定理 19.1 的条件正是关于 $n - s$ 的。

做出这一选择后，我们就有了一个类似于选择 B 的局面；也就是说，对每一组 $\alpha_1, \dots, \alpha_R$ ，要么有一个关于 $|S(\alpha_1, \dots, \alpha_R)|$ 的估计，要么有一组对 $\alpha_1, \dots, \alpha_R$ 的良好联立逼近。这在原则上为一种类似于第 15、16、17 章中处理单个三次方程的处理方法奠定了基础。

这一段在讲什么：收束阶段性成果——既然坏情形被堵死，剩下的**永远是好情形**（选择 B）：对任何一组 α ，二者必居其一：要么 $|S|$ 本来就小（次弧，可忽略），要么 α 有好的联立有理逼近（主弧，贡献主项）。这正是圆法需要的"干净二分"，于是可以照搬第 15–17 章处理三次方程的整套流程（划主弧次弧、算主弧主项、证次弧可忽略）。

"良好联立逼近" (simultaneous approximation)。

"联立"是指 $\alpha_1, \dots, \alpha_R$ 被同一个分母 q 同时逼近：存在不大的 q 和整数 a_i ，使每个 α_i 都很接近 a_i/q 。这正是第四节算出来的东西。"对每组 α 要么 S 小、要么有联立逼近"——这个二分，就是把积分区间 $[0, 1]^R$ 干净地切成次弧（ S 小）和主弧（有联立逼近）的**理论依据**。

七、最后一道坎：奇异积分，与定理 19.1

主要的困难在于奇异积分，而这里奇异簇的维数又一次起作用。对该积分的处理过于繁复，无法在此概述。本质上需要假设原始方程组定义一个维数为 $n - R$ 的簇。

这一段在讲什么：提醒还有最后一块硬骨头——**奇异积分**。即便主弧次弧已分好，要真正算出主项、并证明它为正，还得处理"奇异积分"和"奇异级数"，而这里奇异簇的维数再次起决定作用。Davenport 坦言太繁，只点到为止。

"奇异积分" (singular integral) 与"奇异级数" (singular series) 是什么、为什么必不可少？

在圆法里，主弧上 S 的贡献最终会分解成两个因子相乘：

$$\mathcal{N}(P) \approx \mathfrak{S} \cdot J \cdot P^{n-Rk} \quad (\text{量级}).$$

- **奇异积分 J** (也叫"实密度"): 度量方程组在**实数**范围内解的"局部密度"——大致是"在实数里 $f^{(i)} \approx 0$ 的概率"。第一节假设的"实可解非奇异点"保证了 $J > 0$ 。
- **奇异级数 $\mathfrak{S}$** : 把所有素数 p 的同余解密度连乘起来——大致是"对每个 p , 模 p^t 有解的概率"的乘积 $\mathfrak{S} = \prod_p \chi_p$ 。第一节假设的"每个 p 同余条件满足且非奇异"保证了 $\mathfrak{S} > 0$ 。

为什么必不可少: $\mathcal{N}(P) > 0$ (有整数解) 需要 $\mathfrak{S} > 0$ 且 $J > 0$, 即**每个局部条件都得贡献正密度**。这正是把第一节那些"局部假设"派上用场的地方——它们不是摆设, 而是保证主项真的为正的命门。而第二节反例 (19.1) 的奇异轨迹太大, 会让这套处理失效 (主项被破坏), 所以才需要 V^* 维数的限制。

为什么"原始方程组要定义维数 $n - R$ 的簇"?

回忆第 0 节: n 元空间里 R 个"真正独立"的方程, 切出的簇 V 维数应是 $n - R$ 。"假设 $\dim V = n - R$ "就是要求这 R 个方程**没有冗余、确实各起一份作用** (不会有两个方程其实是一回事)。若维数比 $n - R$ 还大, 说明方程"没切够", 有效约束不足, 奇异积分的标准量级 P^{n-Rk} 就不对了, 整套渐近分析会崩。所以这是个"非退化"前提。

Birch 论文的结果如下:

定理 19.1. 设 $f_1, \dots, f_R$ 是具有整系数的 n 元 k 次型, 其中 $n > R \geq 1$ 。设 V 表示代数簇

$$f_1(\mathbf{x}) = 0, \dots, f_R(\mathbf{x}) = 0,$$

并假设 V 的维数为 $n - R$ 。设 V^* 为相关联的奇异轨迹, 并令 $s = \dim V^*$ 。假设 V 上有一个非奇异实点, 并且对每个素数 p , V 上有一个非奇异 p -进点。那么只要

$$n - s > R(R + 1)(k - 1)2^{k-1},$$

V 上就存在一个整点 $\mathbf{x} \neq \mathbf{0}$ 。

这一段在讲什么：本章的**终极结论**——Birch 定理。它把前面所有铺垫凝结成一句可检验的判据。我们逐条拆解每个前提的来历，并看清那个核心不等式的含义。

$f_1, \dots, f_R$ (整系数 k 次型, $n > R \geq 1$)

研究对象： R 个 n 元 k 次齐次方程，系数都是整数。 $n > R$ ：变量比方程多（否则自由度不够）。

V (维数 $= n - R$)

方程组的公共零点簇；"维数 $n - R$ "即第七节解释的**非退化条件**—— R 个方程互相独立、无冗余。

V^* , $s = \dim V^*$

奇异轨迹（雅可比矩阵秩 $\leq R - 1$ 的点集）及其维数。 s 小 = 方程组"奇异点少、整体光滑"。

非奇异实点 + 每个 p 的非奇异 p -进点

就是第一节列出的两条**局部假设**，它们分别保证奇异积分 $J > 0$ 、奇异级数 $\mathfrak{S} > 0$ 。缺一不可（反例 (19.1) 满足这些却仍无解，正说明它们**不够**，所以才要补下面那条 $n - s$ 的不等式）。

结论：存在整点 $\mathbf{x} \neq \mathbf{0}$

方程组有**非平凡整数解**（齐次方程总有零解，关键是要非零解）。

核心不等式 $n - s > R(R + 1)(k - 1)2^{k-1}$ 的来历与读法。

这是整套论证"算总账"得出的门槛。把它和我们前面见过的量对照：

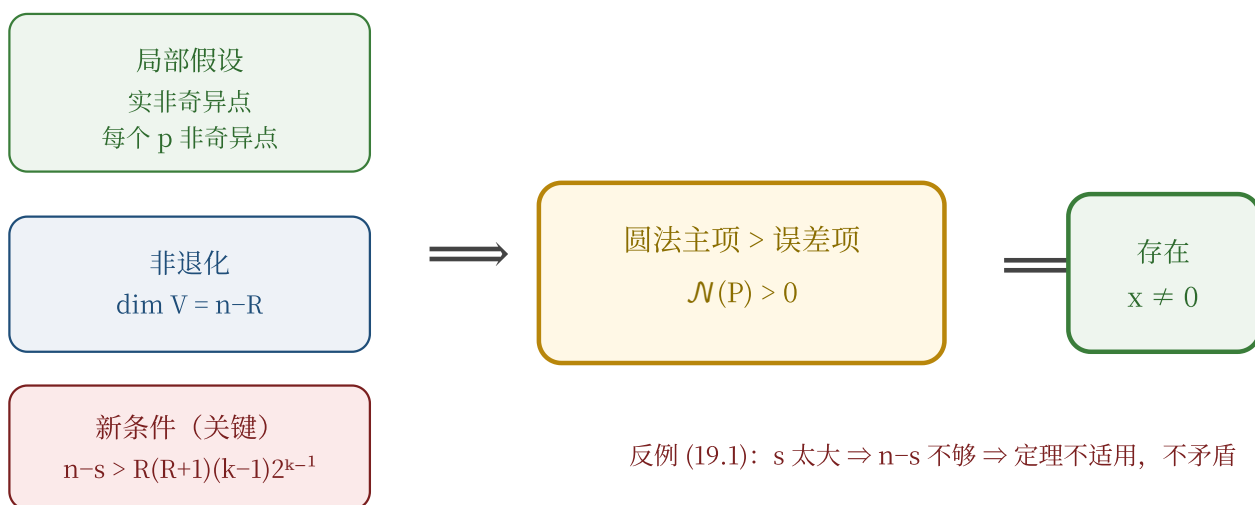
- **左边 $n - s$ ：**变量数减奇异轨迹维数。它度量"方程组离病态有多远"—— n 越大（变量越多、越自由）、 s 越小（奇异点越少）， $n - s$ 越大，越好办。这正是第六节里那个决定 K 的关键组合 $n - s$ 。
- **右边各因子的"出处"：** 2^{k-1} 来自 Weyl 差分 $k - 1$ 次的指数损耗（第三节）； $(k - 1)$ 来自辅助向量有 $k - 1$ 个、做 $k - 1$ 次缩放（第三节）； $R(R + 1)$ 来自处理 R 个方程的联立逼近、以及奇异积分中 R 维的各种求和带来的代价（第四、七节）。

读法：只要变量足够多、奇异轨迹足够小，使 $n - s$ 超过这个由 R, k 决定的常数，圆法的主项就一定压得过误差项，于是 $\mathcal{N}(P) > 0$ ，非零整数解必然存在。**验证退化情形：**

取 $R = 1, k = 3$ (单个三次方程), 右边 $= 1 \cdot 2 \cdot 2 \cdot 4 = 16$, 即条件成 $n - s > 16$ ——与前几章处理单个三次方程时得到的变量数门槛一脉相承, 再次印证本章是那套理论的忠实推广。

回到反例：为什么 (19.1) 不满足定理、因而不矛盾

反例 (19.1) 局部处处可解却整体无解, 会不会和定理 19.1 冲突? **不会**。原因正是它**违反了 $n - s > \dots$ 这条新条件**: (19.1) 由"平方和再立方"造出, 结构高度退化, 它的奇异轨迹 V^* **非常大** (s 很高), 导致 $n - s$ 太小、远够不到右边的门槛。于是定理 19.1 **根本不适用于** (19.1), 自然不保证它有解——这恰恰说明: 那条 $n - s$ 的不等式正是用来排除 (19.1) 这类病例的。第二节抛出的难题, 到这里被定理的新条件干净利落地接住了。整章首尾呼应, 逻辑闭合。



定理 19.1 全景: 三组假设 (局部可解 + 非退化 + $n - s$ 足够大) 共同保证圆法主项压过误差, 得出非零整数解。反例 (19.1) 因奇异轨迹过大而被新条件排除在外。

全章小结：一条逻辑主线

1. **动机**: 把三次方程的圆法推广到高次、多方程, 撞上两个困难——同余条件不再自动满足 (改为假设); 局部—整体失效 (反例 19.1)。

2. **工具**：指数和 S + 基本恒等式"积分 = 数解"；Weyl 差分把 k 次型降成 $k - 1$ 个辅助点的多重线性型 $M_J^{(i)}$ 。
3. **二分**：若 S 偏大，则那个雅可比型矩阵要么满秩（得 α 的好联立逼近，主弧）、要么恒不满秩（坏情形）。
4. **新思想**：把"坏情形"几何化为"奇异轨迹 V^* 维数高"；再通过精选 K ，制造矛盾、堵死坏情形。决定成败的量是 $n - s$ 。
5. **收官**：补上奇异积分/奇异级数（由局部假设保证为正），在 $n - s > R(R + 1)(k - 1)2^{k-1}$ 时主项为正，得非零整数解——即定理 19.1。

一句话记住本章：**变量越多、奇异轨迹越小，高次齐次方程组就越逃不过"有非零整数解"的命运**——而"多到什么程度才够"，由 $n - s > R(R + 1)(k - 1)2^{k-1}$ 精确给出。

返回 [全书目录](#)