

DAVENPORT · 圆法 · 高中详解版

一般齐次方程与 Birch 定理

General homogeneous equations: Birch's theorem

本章要解决什么 / 读完能掌握什么

前面几章讲的都是**加性方程**，比如 $c_1x_1^k + c_2x_2^k + \cdots + c_nx_n^k = 0$ ，每一项只用到一个变量的 k 次方。本章要把目标提升到**任意的 k 次齐次方程** $f(x_1, \dots, x_n) = 0$ （项里可以出现 $x_1x_2x_3$ 、 $x_1^2x_2$ 这样混合变量的乘积）。

核心结论是 **Birch 定理**：只要每个方程的次数都是**奇数**，并且变量个数 n 足够大（大到某个只与次数有关的界），那么这一组方程一定有非平凡的有理数解；而且解还能填满一整个高维"解空间"。

读完你将掌握：① 什么叫"型"（齐次多项式）和"齐次"；② 圆法之外的另一种思想——**用线性代数把"一般型"逐步化简成"加性型"**，再借用前面圆法的成果收尾；③ 多重线性函数、线性无关、向量空间维数、对次数做数学归纳法等工具；④ Birch 定理 11.1 的完整证明，每一步都补全、不跳步。

阅读方式：黑色叙述是对译文每一段的忠实转述与串讲；彩色框（目标 / 符号 / 分步推演 / 定理 / 证明）和插图是为高中生补充的零省略详解。译文里凡是写"显然""容易看出"的地方，本页都补成一步一句的推导。

一、从加性方程走向一般齐次方程

译文段落（第1段）。"现在我们转向一般的齐次方程……设 $f(x_1, \dots, x_n)$ 是一个整系数的、次数为 k 的齐次多项式，我们称之为一个**型**（form）。我们关心的是 $f(x_1, \dots, x_n) = 0$ 的可解性。"

这一段在交代**研究对象**。我们先把里面每个词讲透。

什么是"多项式 $f(x_1, \dots, x_n)$ "? 就是把若干个变量 $x_1, x_2, \dots, x_n$ 用加、减、乘搭起来的式子，例如 $x_1^2 + 3x_1x_2 - x_3^2$ 。下标只是给变量编号， n 是变量的总个数。 f 是"function（函数）"的首字母，读作"f"。

什么是"次数 k "? 一项里所有变量指数之和叫这一项的**次数**。例如 $x_1^2x_2$ 的次数是 $2 + 1 = 3$ ； $x_1x_2x_3$ 的次数是 $1 + 1 + 1 = 3$ 。整个多项式的次数，是各项次数里最大的那个。 k 读作"k"，是 degree（次数）的代号（德语 *Grad*）。

什么是"齐次"（homogeneous）? 如果一个多项式里**每一项的次数都相等**（都等于 k ），就说它是 k 次**齐次**的。例如 $x_1^2 + x_2^2 - x_3^2$ 是 2 次齐次； $x_1^2 + x_2$ 不是齐次（一项 2 次、一项 1 次）。齐次多项式有个漂亮性质：把所有变量同时放大 t 倍，整个值会放大 t^k 倍，即

$$f(tx_1, \dots, tx_n) = t^k f(x_1, \dots, x_n).$$

这条性质后面会反复用到。"型"（form）是齐次多项式的专门叫法。

译文段落（第2段）。"在整数 $x_1, \dots, x_n$ （不全为 0）中的可解性。由于齐次性，我们可以允许系数和变量取有理数而不取整数……一个显然的必要条件是：该方程必须在实数中可解，且解不全为 0。"

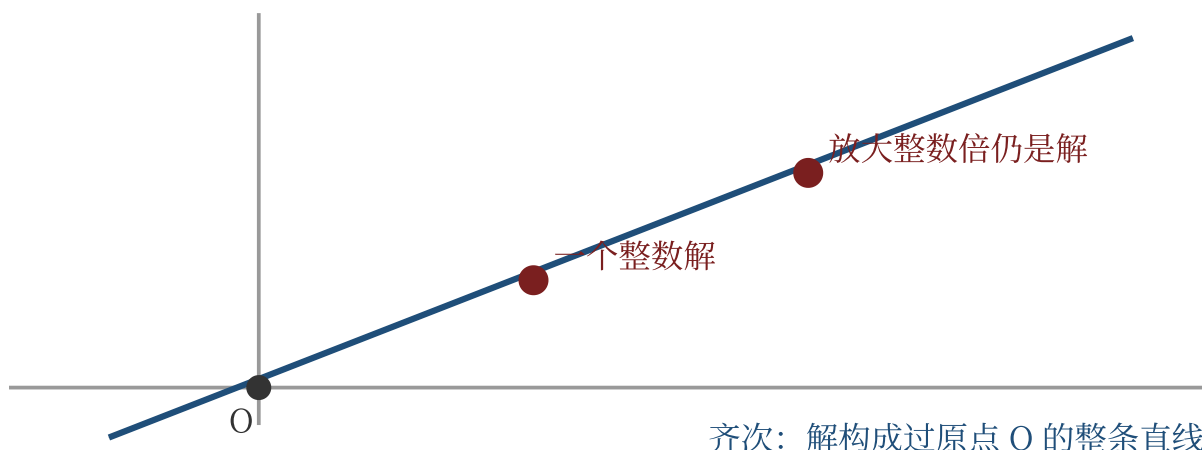
这一段讲三件事，逐条拆解。

逐句补全

1 "不全为 0"为什么要强调? 因为 $x_1 = x_2 = \cdots = x_n = 0$ 必然让 $f = 0$ (每一项都含变量, 代 0 就全没了)。这个解谁都有, 毫无信息, 叫平凡解。我们真正要找的是非平凡解——至少有一个变量不为 0。

2 为什么"取整数"和"取有理数"是同一回事? 设我们找到一组有理数解 $x_1, \dots, x_n$ 。把它们写成分数, 取所有分母的公倍数 D , 令 $x'_j = Dx_j$, 则 x'_j 全是整数。由齐次性, $f(x'_1, \dots, x'_n) = f(Dx_1, \dots, Dx_n) = D^k f(x_1, \dots, x_n) = D^k \cdot 0 = 0$ 。所以有理解能放大成整数解, 二者可解性完全等价。这就是"齐次"带来的便利: 只要找到一束方向, 整条直线上的整点都是解。

3 "在实数中可解"为什么是必要条件? 整数、有理数都是实数。如果连放宽到实数都无解 (只有平凡解), 那在更小的整数范围里更不可能有解。所以"实数中有非平凡解"是"整数中有非平凡解"的必要前提。例如 $x_1^2 + x_2^2 = 0$ 在实数里只有 $x_1 = x_2 = 0$, 所以它在整数里也没有非平凡解。"必要条件"的意思是: A 成立才可能 B 成立; 缺了 A, B 一定不成立。



齐次方程的非平凡解一旦存在, 把它乘以任何倍数仍是解——解集是"过原点的方向", 所以有理数解和整数解一回事。

二、二次型 (k=2): 配方法与 Meyer 定理

译文段落 (第 3 段)。"当 $k = 2$, 即 f 是一个二次型时, 可以借助'配方法'把一般的型表示为加性型……当 $n \geq 5$ 时这些同余条件总能满足, 相应地我们便有 Meyer (1883) 的著名定理: 五个或更多变量中的不定二次型总能非平凡地表示零。"

这一段是背景回顾: 在最简单的 $k = 2$ 情形, 问题早已被彻底解决。它告诉我们"一般型化成加性型"这个思路并不新鲜, 二次时就管用。把里面的概念逐个讲清。

"配方法" (completing the square)。就是高中学过的配方。例如二次型

$x^2 + 2xy + 3y^2$ 含有混合项 $2xy$, 配方得 $(x + y)^2 + 2y^2$ 。令

$u = x + y, v = y$, 它就变成 $u^2 + 2v^2$ ——只剩纯平方项、没有混合项了, 这正是加性型。一般的二次型都能这样消掉所有混合项, 化成 $a_1u_1^2 + a_2u_2^2 + \dots$ 的加性形状。这说明: 对二次型, "一般"和"加性"本质上没区别。本章后面要做的, 正是把这个想法艰难地推广到更高的次数。

"同余" $f \equiv 0 \pmod{p^\nu}$ 。记号 $a \equiv b \pmod{N}$ 读作" a 与 b 模 N 同余", 意思是 $a - b$ 能被 N 整除, 也就是除以 N 的余数相同。例如

$17 \equiv 2 \pmod{5}$ (都余 2)。这里 p 是素数, p^ν 是它的某个幂 (ν 读作"纽", 是希腊字母 nu, 表示指数)。条件 $f \equiv 0 \pmod{p^\nu}$ 就是要求方程在"对 p^ν 取余数"的世界里也有非平凡解。为什么这是必要的? 因为如果真有整数解使 $f = 0$, 那对它取余数当然得 0; 所以"取余后无解"就反证了"整数无解"。

"不定" (indefinite) 二次型。指这个二次型既能取到正值、又能取到负值

(不像 $x^2 + y^2$ 永远 ≥ 0 , 那叫正定)。要想等于 0, 自然必须能"由正变负"地穿过 0, 所以不定是必要的。 $x^2 + y^2 = 0$ 只有零解, 正是因为它正定。

Meyer 定理 (1883) 是什么

这是数论里的一块里程碑：只要变量个数 $n \geq 5$ ，又是不定的整系数二次型，就一定能非平凡地表示 0（即 $f = 0$ 有不全为零的整数解），无须再单独检验那些同余条件——因为 $n \geq 5$ 时它们自动满足。

历史意义：Meyer 在 1883 年证明了这一点，它是“变量足够多就一定可解”这类定理的祖先。Birch 定理（本章主角）正是把这种精神从 2 次推广到所有奇数次。“只对有限个素数 p 才需要检验”是因为：一旦 p 大于系数里出现的因子，同余条件自动成立，所以要查的 p 只有有限个。

三、高次型为什么远比加性型复杂

译文段落（第 4 段）。“对于 $k > 2$ ，一般的型比加性型一般得多。例如，若 $k = 3$ ，则一般型中有 $\frac{1}{6}n(n+1)(n+2)$ 个独立系数，而由于对 n 个变量作的一个线性变换只有 n^2 个系数，显然几乎不可能作什么简化。”

这一段用数系数的方式，说明“高次时配方法行不通”。我们把这两个数字算清楚。

为什么三次型有 $\frac{1}{6}n(n+1)(n+2)$ 个系数

- 1 三次齐次多项式的每一项形如 $x_i x_j x_k$ ，系数记为 c_{ijk} 。这里 i, j, k 各从 1 到 n ，但交换顺序得到的是同一项（ $x_1 x_2 x_3$ 与 $x_2 x_1 x_3$ 一样），所以独立的系数个数 = 从 n 个变量里可重复地取 3 个的组合数。

- 2 “可重复取 3 个”的组合数公式是 $\binom{n+3-1}{3} = \binom{n+2}{3}$ 。代入 $\binom{n+2}{3} = \frac{(n+2)(n+1)n}{3!} = \frac{n(n+1)(n+2)}{6}$ 。这就是

$\frac{1}{6}n(n+1)(n+2)$ 。(注：原文 PDF 排版把 $\frac{1}{6}$ 错印成 "16"，正确是 $1/6$ 。)

3 它随 n 像 n^3 那样增长 (约 $n^3/6$)。

"线性变换只有 n^2 个系数"是什么意思？ 所谓"线性变换"，就是用一组新变量 $y_1, \dots, y_n$ 去替换旧变量：

$$x_i = a_{i1}y_1 + a_{i2}y_2 + \dots + a_{in}y_n \quad (i = 1, \dots, n).$$

这里的"系数表" (a_{ij}) 一共有 $n \times n = n^2$ 个数，这是我们做"换元化简"时能自由调节的全部旋钮。配方法本质就是找一个合适的线性变换。

这段的关键洞察

三次型有约 $n^3/6$ 个系数要"对付"，而线性变换只提供 n^2 个旋钮。当 n 大时， $n^3/6$ 远多于 n^2 (旋钮远不够用)，所以不可能像二次那样靠一次线性换元就把一般型化成加性型。这正是高次问题困难的根源，也解释了为什么后面 Birch 的办法要"浪费"大量变量、绕一大圈才能成功。

四、退化性 (degeneration)

译文段落 (第 5 段)。“与一般型相关的一个重要问题是退化性……称 n 个变量中的一个型是退化的，如果存在一个从 $x_1, \dots, x_n$ 到 $y_1, \dots, y_n$ 的非奇异线性变换，使得变换后的型只涉及 $y_1, \dots, y_{n-1}$ ；也就是说，所有含 y_n 的项的系数都消失。”

"退化"是一个判断这个型有没有"多余变量"的概念。下面把定义讲透。

"非奇异" (non-singular) 线性变换。指这个换元是可逆的——能从新变量唯一解回旧变量，不丢失信息（用矩阵的话说，系数矩阵行列式不为 0）。只有可逆换元才不改变问题实质。"奇异"则是不可逆、会压缩信息的坏换元。

退化到底长什么样

举例： $f = (x_1 + x_2)^2$ ，表面上有两个变量。令 $y_1 = x_1 + x_2$, $y_2 = x_2$ （可逆），则 $f = y_1^2$ ，只剩 y_1 ， y_2 完全消失。这就叫**退化**——它实质上是"一个变量的型"，第二个变量是摆设。反过来， $x_1^2 + x_2^2$ 无论怎样可逆换元都甩不掉一个变量，它是**非退化**的。

译文（续）。"退化性是一个绝对的性质……用我们的记号，必须有 $\partial f / \partial y_n = 0$ 恒成立；而若 C_j 是 y_n 在 x_j 中的系数，这就等价于

$$C_1 \frac{\partial f}{\partial x_1} + \cdots + C_n \frac{\partial f}{\partial x_n} = 0$$

恒成立。"

偏导数 $\partial f / \partial y_n$ 是什么。 符号 ∂ 读作"偏" (partial)。 $\partial f / \partial x_j$ 的意思是：把除 x_j 以外的变量都当常数，只对 x_j 求导。例如 $f = x_1^2 x_2$ ，则 $\partial f / \partial x_1 = 2x_1 x_2$ （对 x_1 求导）， $\partial f / \partial x_2 = x_1^2$ （对 x_2 求导）。它衡量" f 随该变量变化的快慢"。高中只学过单变量求导，偏导无非是"锁住别的变量、只动一个"，求导规则完全一样。

为什么"不含 y_n "等价于" $\partial f / \partial y_n \equiv 0$ "，再等价于那条链式法则等式

- 1 "型里不出现 y_n " $\iff$ "无论 y_n 怎么变 f 都不变" $\iff$ " f 对 y_n 的偏导恒为 0"，即 $\partial f / \partial y_n \equiv 0$ 。（恒为 0 是因为齐次多项式若含 y_n ，其对 y_n

的偏导必非零。)

- 2 现在 f 本来是用 $x_1, \dots, x_n$ 写的, 而 x_j 又通过换元依赖 y_n 。设 x_j 中 y_n 的系数是 C_j , 即 $\partial x_j / \partial y_n = C_j$ 。

- 3 用多元链式法则 (把每条路径的偏导乘起来再相加):

$$\frac{\partial f}{\partial y_n} = \sum_{j=1}^n \frac{\partial f}{\partial x_j} \cdot \frac{\partial x_j}{\partial y_n} = \sum_{j=1}^n C_j \frac{\partial f}{\partial x_j}.$$

令它恒为 0, 就得到译文那条等式 $C_1 \partial f / \partial x_1 + \dots + C_n \partial f / \partial x_n = 0$ 。这是退化的代数判据。

"绝对性质"是什么意思。就是"是否退化"不依赖于你在哪个数域里换元: 如果在有理数范围内无法把它退化, 那即使允许用更大的数 (无理数、复数) 做换元, 照样退化不了。译文最后一句解释了原因: 上面那条等式拆开来, 是关于 $C_1, \dots, C_n$ 的有限多个线性方程; 线性方程组"有没有非零解"只看系数是否有理, 能在原数域里解就一定能在原数域里解, 扩大数域帮不上忙。

译文段落 (第 6 段)。"然而, 从当前问题——即表示零——的角度看, 我们总可以假设 f 是非退化的; 因为若 f 如上所述退化, 那么对应于解 $y_1 = \dots = y_{n-1} = 0, y_n = 1$, 就有一个 $x_1, \dots, x_n$ 不全为 0 的解。"

这是一个"无伤大雅"的简化, 解释为什么

- 1 如果 f 退化, 换元后 f 不含 y_n 。那么取 $y_1 = \dots = y_{n-1} = 0, y_n = 1$, 代入后 $f = 0$ (因为 f 根本不依赖 y_n , 而其余 y 都为 0, 所有项都消失)。

2 由于换元可逆且 $y_n = 1 \neq 0$ ，回代得到的 $x_1, \dots, x_n$ 不全为 0（否则可逆变换会把非零的 y 也压成零，矛盾）。

3 所以退化的型自动有非平凡解，这种情形已经"白送"解决了。于是研究 $f = 0$ 时，我们只需操心非退化的型即可——这就是"不失一般性地假设 f 非退化"的来由。

五、历史脉络：Brauer、Lewis、Birch

译文段落（第 7 段）。"朝着求解一般齐次方程方向迈出第一步实质性进展的是 R. Brauer [9]……Brauer 定理在 p 进数域中适用，并确立了一个函数 $n_1(k, h)$ 的存在性……只要 $n \geq n_1(k, h)$ ，便在 p 进数中可解。"

这一段是历史，讲清三代人的接力。先讲 Brauer。

Richard Brauer 的贡献与它的局限

是谁、做了什么：Richard Brauer（理查德·布饶尔，德裔美国代数学家）最先证明了一个"约化定理"：要解一个 k 次齐次方程，只要能解所有次数 $\leq k$ 的加性方程就够了——把"一般型问题"归约成"加性型问题"。这正是本章方法的鼻祖思想。

代价：为了让最终得到的加性方程里还剩"够用"的变量，原方程的变量个数 n 必须取得极其庞大。这就是后面反复出现的"变量浪费"。

为什么在有理数里不灵：归约出来的加性方程里会混入偶数次的（比如 4 次）。偶次方永远 ≥ 0 ，若系数同号（全正或全负），方程 $c_1 x_1^4 + \dots = 0$ 除了零解别无可能——在实数（从而有理数）里根本无解。所以 Brauer 定理在有理数域里会卡死在偶次方程上。

" **p 进数域**" (p-adic numbers) 是什么? 这是一套与实数平行的"数系"。实数关心"两个数差多小", p 进数关心"两个数的差能被素数 p 的多高次幂整除"——本质上是把"模 $p, p^2, p^3, \dots$ 同余"的信息全部打包成一种数。在 p 进世界里没有"正负号"的障碍, 所以 Brauer 定理在这里畅通无阻, 给出了函数 $n_1(k, h)$: 只要 $n \geq n_1(k, h)$, n 个变量里任意 h 个次数 $\leq k$ 的方程, 在 p 进数中必有非平凡解。 h 是方程个数。高中阶段把它理解为"加强版的同余可解性"即可。

译文段落 (第 8 段)。"这一领域中最简单的问题是单个三次方程的问题。Lewis 教授 [56] 最先证明了存在一个绝对常数 n_0 , 使得任何 n 个变量中带有理系数的三次方程, 只要 $n \geq n_0$, 便在有理数中可解。"

Lewis 的突破

Brauer 在**有理数域**卡住了。Lewis 教授 (D. J. Lewis) 针对最简单的**单个三次方程**, 用**代数数论**的论证去补 Brauer 的不足, 第一次证明了: **存在一个绝对常数 n_0** (不依赖于具体方程, 只是一个固定的大数), 使得变量数 $n \geq n_0$ 的任何有理系数三次方程在有理数里都可解。"绝对常数"是关键——它意味着"变量够多就一定有解"这件事, 对所有三次方程一视同仁。

译文段落 (第 9 段)。"不久之后, Birch [5] 证明了一个更一般的定理, 即: 每个齐次方程组, 只要其中各方程都是奇次的, 并且 n 超过依赖于这些次数的某个函数, 便可解。这个定理就是本章的主题。"

Birch 定理一句话

B. J. Birch 把 Lewis 的"单个三次"一举推广到"任意一组方程", 条件只有一个: **每个方程的次数都是奇数**。只要变量数 n 超过一个仅由这些次数决定的界, 方程组就有非平凡有理解。"奇次"这个条件正是为了避开 Brauer 卡住的

偶次陷阱（奇次方 x^r 既能正又能负，没有符号障碍）。这就是本章要完整证明的定理 11.1。

六、方法预览：先讲单个三次方程

译文段落（第 10 段）。"在以其全部一般性证明 Birch 定理之前，我打算先就最简单的情形——单个三次方程——来说明方法……后者的可解性由定理 8.1 保证，只要变量个数至少为 $k^2(2k - 1) + 1 = 46$ 。（实际上 8 个变量就够了，但这需要特殊的论证。）"

作者的教学策略：先在三次的小场景里把核心招式演示一遍，再上升到一般证明。这一段先把要借用的"外援"说清楚。

定理 8.1 是什么（外援工具）。这是前面章节用圆法证明过的结论：加性方程 $c_1 u_1^k + \cdots + c_s u_s^k = 0$ ，只要变量个数 s 足够大，就有非平凡整数解。对三次（ $k = 3$ ），这个"足够大"的界由公式 $k^2(2k - 1) + 1$ 给出。代入 $k = 3$ ：

$$3^2 \cdot (2 \cdot 3 - 1) + 1 = 9 \times 5 + 1 = 45 + 1 = 46.$$

所以只要有 46 个变量，加性三次方程就保证可解。括号里说"实际 8 个就够"，是因为有更精细但更费力的论证；这里用宽松的 46，是为了证明走得轻松。这正是"取参数时优先取够用而好证的，而非最优的"的体现。

整体策略（极重要）

圆法只会解加性方程（每项一个变量的方幂）。一般三次型有混合项，圆法直接用不上。于是 Birch 的办法是：

① 用纯线性代数，把一般三次型 f 一步步"化简 / 表示"成一个加性三次型 $b_1 u_1^3 + \cdots + b_s u_s^3$ （代价是要很多变量）；② 再对这个加性型调用定理 8.1

求解。只要起初的 n 足够大，化简后还能剩下 $s \geq 46$ 个变量，整条链就走了。下面就来落实第①步。

七、"表示" (represents) 的精确含义

译文段落（第 11 段）。"设 $f(x_1, \dots, x_n)$ 是一个三次型。我们说： f 表示 $g(y_1, \dots, y_m)$ （其中 $m \leq n$ ），如果存在 n 个关于 $y_1, \dots, y_m$ 的线性型，其秩为 m ，使得当用这些线性型替换 $x_1, \dots, x_n$ 时，得到 $f(x_1, \dots, x_n) = g(y_1, \dots, y_m)$ 关于 $y_1, \dots, y_m$ 恒成立。"

" f 表示 g " 是什么。就是：把 f 的 n 个变量 x_j 各自换成 $y_1, \dots, y_m$ 的线性组合后， f 恰好变成 g 。换言之， g 是 f 在某个 m 维"切片"上看到的样子。因为变量从 n 个减少到 m 个（ $m \leq n$ ）， g 是更"瘦"的型。

"线性型"和"秩为 m "。"线性型"就是 y 的一次齐次式，如 $x_j = a_{j1}y_1 + \dots + a_{jm}y_m$ 。我们一共要给 n 个 x_j 各配一个这样的式子。"秩为 m "是说这 n 个线性型里有 m 个是真正独立的（不能互相线性表出），从而 $y_1, \dots, y_m$ 都是"有用的、独立的"新变量，没有偷偷退化成更少的变量。

为什么"表示"对我们有用

关键观察：如果 g 有非平凡零点，那么 f 也有。因为找到使 $g(y_1, \dots, y_m) = 0$ 的非零 y ，回代 $x_j =$ （关于 y 的线性型）就得到使 $f = 0$ 的 x （秩为 m 保证回代出的 x 不全为 0）。所以只要能让 f "表示"一个我们会解的简单型 g （最终是加性型）， f 的解就到手了。

译文段落（第 12 段，核心思路）。"证明的基本思想是：证明 f 表示一个如下类型的

$$a_0 t_0^3 + g_1(t_1, \dots, t_m) \quad (11.1)$$

只要 n 超过某个仅依赖于 m 的数。"

(11.1) 这一步要干什么、为什么这样设计

式 (11.1) 的形状是：一个孤立的纯立方项 $a_0 t_0^3$ ，加上一个完全不含 t_0 的三次型 $g_1(t_1, \dots, t_m)$ 。也就是说，我们成功地把变量 t_0 "剥离"出来——它只以 t_0^3 的纯净形式出现，不和别的变量纠缠。

动机：这正是"把一般型一点点逼近加性型"的单步操作。每做一次，就多剥出一个像 $a_0 t_0^3$ 这样的纯立方项；剩下的 g_1 是变量更少的新三次型，对它再做一次，又剥出一项……反复剥，最后 f 就被表示成纯立方项之和 $b_1 u_1^3 + \dots + b_s u_s^3$ ，即加性型。 t_0 读作"t-零"，是新坐标系里被挑出来的特殊方向。

八、三线性函数 T 与"剥离一个立方项"的构造

译文段落（第 13 段）。"写 $f(\mathbf{x}) = \sum_i \sum_j \sum_k c_{ijk} x_i x_j x_k$ ，其中各求和从 1 到 n ，且 c_{ijk} 是三个下标的对称函数。对三个点 $\mathbf{x}, \mathbf{y}, \mathbf{z}$ 定义三线性函数 $T(\mathbf{x} | \mathbf{y} | \mathbf{z}) = \sum_i \sum_j \sum_k c_{ijk} x_i y_j z_k$ 。"

求和号 $\sum$ （西格玛）。 $\sum_i$ 读作"对 i 求和"，意思是让下标 i 跑遍它的取值范围，把每一项加起来。这里 $\sum_i \sum_j \sum_k$ 是三重求和：让 i, j, k 各自独立地从 1 跑到 n ，把所有 $c_{ijk} x_i x_j x_k$ 加在一起。这就是三次型最一般的写法。

黑体 $\mathbf{x}$ (向量 / 点)。 $\mathbf{x}$ 是把 n 个坐标打包成的一个"点" ($x_1, \dots, x_n$)。 x_i 是它的第 i 个坐标。用一个黑体字母代替一长串坐标，是为了书写简洁。把 $\mathbf{x}$ 想成 n 维空间里的一个箭头/位置即可。

" c_{ijk} 是对称函数"。意思是无论怎样交换三个下标，系数都不变：

$c_{ijk} = c_{jik} = c_{kji} = \dots$ 。为什么能这样规定？因为 $x_i x_j x_k$ 本身就和下标顺序无关（乘法可交换），所以我们总可以把同类项的系数**平均分摊**，做成对称的，这只是记号上的方便，不改变 f 。

三线性函数 $T(\mathbf{x} \mid \mathbf{y} \mid \mathbf{z})$ 是什么、为什么要造它。把 f 里三个相乘的变量分别放到三个不同的点上：第一个因子取 $\mathbf{x}$ 的坐标，第二个取 $\mathbf{y}$ 的，第三个取 $\mathbf{z}$ 的，得到

$$T(\mathbf{x} \mid \mathbf{y} \mid \mathbf{z}) = \sum_i \sum_j \sum_k c_{ijk} x_i y_j z_k.$$

它有两个关键性质：① **在每个位置上都是线性的**（固定另外两个点，对第一个点的坐标而言是一次的）；② **三个点取相同时还原成 f** ： $T(\mathbf{x} \mid \mathbf{x} \mid \mathbf{x}) = f(\mathbf{x})$ 。 T 取"trilinear（三线性）"的首字母。造它的**动机**： f 本身是三次的（非线性，难处理），但 T 在每个槽里都是线性的，**线性方程我们会解**。把一个三次问题拆成"在某些槽里线性"的问题，就能动用线性代数这台机器。这是整段构造的灵魂。

三点相同时 $T(\mathbf{x} \mid \mathbf{x} \mid \mathbf{x}) = f(\mathbf{x})$

$f(\mathbf{x})$: 三因子同点



$T(\mathbf{x} \mid \mathbf{y} \mid \mathbf{z})$: 三因子分到三点

$\mathbf{x} \cdot \mathbf{x} \cdot \mathbf{x}$ (三次, 难)

每个槽都是一次的 (线性, 可解)

把三次型 f 的三个相乘因子"拆"到三个独立的点上，得到对每个点都线性的 T 。这是后面所有构造的工具。

译文段落（第 14 段）。"我们任取 m 个线性无关的点 $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(m)}$ ，它们具有整（或有理）坐标，并考虑关于一个未知点 $\mathbf{y}$ 的方程

$$T(\mathbf{y} \mid \mathbf{y}^{(p)} \mid \mathbf{y}^{(q)}) = 0, \quad 1 \leq p \leq q \leq m.$$

"

"线性无关" (linearly independent)。一组点 $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(m)}$ 线性无关，是指没有一个能被其余的线性组合表示出来；等价地， $c_1 \mathbf{y}^{(1)} + \dots + c_m \mathbf{y}^{(m)} = \mathbf{0}$ 只有全零系数这一种可能。直观上它们指向 m 个"真正不同的方向"，张成一个 m 维空间。上标 (p) 是给这 m 个固定点编号，别和坐标下标混淆。

这组方程在求什么。固定后两个槽为已知点 $\mathbf{y}^{(p)}, \mathbf{y}^{(q)}$ ，让第一个槽是未知点 $\mathbf{y}$ 。由于 T 对第一个槽线性， $T(\mathbf{y} \mid \mathbf{y}^{(p)} \mid \mathbf{y}^{(q)}) = 0$ 是关于 $\mathbf{y}$ 的 n 个坐标的一条线性方程。让 (p, q) 跑遍 $1 \leq p \leq q \leq m$ 的所有组合，就得到一组齐次线性方程组。我们想找一个非零的 $\mathbf{y}$ 同时满足它们。

九、计数方程、找出特殊点 $\mathbf{y}^{(0)}$

译文段落（第 15 段）。"这是关于 $\mathbf{y}$ 的 n 个坐标的 $\frac{1}{2}m(m+1)$ 个线性方程。如果 $n > \frac{1}{2}m(m+1)$ ，则存在一个满足它们的、异于原点的点 $\mathbf{y}$ 。称这样的点为 $\mathbf{y}^{(0)}$ 。"

为什么方程个数是 $\frac{1}{2}m(m+1)$ ，为什么 n 大就有非零解

- 1 数方程个数。下标对 (p, q) 满足 $1 \leq p \leq q \leq m$ ，即从 $\{1, \dots, m\}$ 里可重复地选两个（允许 $p = q$ ）。这样的对数 =

$$\binom{m}{2} + m = \frac{m(m-1)}{2} + m = \frac{m(m+1)}{2}。 \text{（也可用可重组合} \binom{m+1}{2} = \frac{1}{2}m(m+1) \text{。）}$$

所以共 $\frac{1}{2}m(m+1)$ 条线性方程。

- 2 线性代数事实。一个齐次线性方程组，若未知数个数 n 严格大于方程个数，则必有非零解（方程不够多，约束不住所有自由度，总会留下非零的解方向）。

- 3 这里未知数是 $\mathbf{y}$ 的 n 个坐标，方程有 $\frac{1}{2}m(m+1)$ 条。只要 $n > \frac{1}{2}m(m+1)$ ，就保证存在非零解 $\mathbf{y}$ 。把任选的一个这样的非零解命名为 $\mathbf{y}^{(0)}$ （上标 0 表示"新挑出来的特殊点"）。于是

$$T(\mathbf{y}^{(0)} \mid \mathbf{y}^{(p)} \mid \mathbf{y}^{(q)}) = 0, \quad 1 \leq p \leq q \leq m.$$

未知数 n 个

>

方程 $\frac{1}{2}m(m+1)$ 条

$\Rightarrow$ 齐次线性方程组必有非零解 $\mathbf{y}^{(0)}$

"变量比方程多"是本章反复使用的免费午餐：齐次线性方程组一定留下非零解方向。

译文段落（第 16 段，退化情形先处理掉）。"如果点 $\mathbf{y}^{(0)}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(m)}$ 线性相关，那么……我们应当有 $\mathbf{y}^{(0)} = c_1 \mathbf{y}^{(1)} + \dots + c_m \mathbf{y}^{(m)}$ ……但那时

$$T(\mathbf{y}^{(0)} \mid \mathbf{y}^{(0)} \mid \mathbf{y}^{(0)}) = \sum_{p=1}^m \sum_{q=1}^m c_p c_q T(\mathbf{y}^{(0)} \mid \mathbf{y}^{(p)} \mid \mathbf{y}^{(q)}) = 0,$$

由此 $f(\mathbf{y}^{(0)}) = 0$ ，从而给出了所求的解。"

补全这段"显然"的推导

1 为什么线性相关就能这样写。已知 $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(m)}$ 线性无关，而现在 $m+1$ 个点 $\mathbf{y}^{(0)}, \dots, \mathbf{y}^{(m)}$ 线性相关，那"多出来的"那个 $\mathbf{y}^{(0)}$ 必能由其余 m 个表示： $\mathbf{y}^{(0)} = c_1 \mathbf{y}^{(1)} + \dots + c_m \mathbf{y}^{(m)}$ ，系数 c_p 是有理数。

2 把这个代入 T 的第二、第三个槽。因为 T 对每个槽线性，可以把和号拆出来：

$$T(\mathbf{y}^{(0)} \mid \mathbf{y}^{(0)} \mid \mathbf{y}^{(0)}) = T\left(\mathbf{y}^{(0)} \mid \sum_p c_p \mathbf{y}^{(p)} \mid \sum_q c_q \mathbf{y}^{(q)}\right) = \sum_p \sum_q c_p c_q$$

(这里第一个槽保持 $\mathbf{y}^{(0)}$ 不动；线性性允许把第二、三槽里的线性组合"提"成系数乘求和。)

3 每一项都为 0。上一段已证 $T(\mathbf{y}^{(0)} \mid \mathbf{y}^{(p)} \mid \mathbf{y}^{(q)}) = 0$ 对所有 $p \leq q$ 成立；由对称性 $p > q$ 也一样。所以整个双重和是 0。

4 得到解。而 $T(\mathbf{y}^{(0)} \mid \mathbf{y}^{(0)} \mid \mathbf{y}^{(0)}) = f(\mathbf{y}^{(0)})$ (三槽同点还原成 f)。于是 $f(\mathbf{y}^{(0)}) = 0$ ，且 $\mathbf{y}^{(0)} \neq \mathbf{0}$ ，正是要找的非平凡解，大功告成 (这一支直接结束)。

十、独立情形：换元后型的形状

译文段落 (第 17 段)。"于是我们可以假设这 $m+1$ 个点线性无关。变换

$$x_j = t_0 y_j^{(0)} + t_1 y_j^{(1)} + \dots + t_m y_j^{(m)}, \quad (1 \leq j \leq n)$$

的秩为 $m+1$ ，并把 f 表示为 $t_0, t_1, \dots, t_m$ 的一个型。"

这个换元在做什么

1 上一支若线性相关已经出解，所以现在只需讨论 $\mathbf{y}^{(0)}, \dots, \mathbf{y}^{(m)}$ 线性无关的情形。

2 用这 $m + 1$ 个无关方向作"新坐标轴"：任何点 $\mathbf{x}$ 写成它们的线性组合，组合系数就是新变量 $t_0, t_1, \dots, t_m$ 。具体地，第 j 个坐标 $x_j = t_0 y_j^{(0)} + t_1 y_j^{(1)} + \dots + t_m y_j^{(m)}$ ，其中 $y_j^{(p)}$ 是第 p 个点的第 j 个坐标。

3 因为这 $m + 1$ 个点线性无关，这个换元的秩是 $m + 1$ ($t_0, \dots, t_m$ 都是真变量)，符合"表示"的定义。代入后 f 变成只含 $m + 1$ 个变量 $t_0, \dots, t_m$ 的三次型——这正是" f 表示一个 $m + 1$ 元型"。

译文段落（第 18 段，算系数）。"在这个型中 $t_0 t_p t_q$ 的系数是 $T(\mathbf{y}^{(0)} \mid \mathbf{y}^{(p)} \mid \mathbf{y}^{(q)})$ ，故当 $1 \leq p \leq q \leq m$ 时它为 0。因此新的型具有下列形状

$$a_0 t_0^3 + t_0^2(b_1 t_1 + \dots + b_m t_m) + g_1(t_1, \dots, t_m).$$

"

为什么混合项 $t_0 t_p t_q$ 的系数恰是那个 T ，又为什么它是 0

1 把换元 $\mathbf{x} = t_0 \mathbf{y}^{(0)} + t_1 \mathbf{y}^{(1)} + \dots + t_m \mathbf{y}^{(m)}$ 代入 $f(\mathbf{x}) = T(\mathbf{x} \mid \mathbf{x} \mid \mathbf{x})$ 。利用三线性，把三个槽各自展开成 $\sum t_p \mathbf{y}^{(p)}$ ，乘开后每一项形如 $t_a t_b t_c T(\mathbf{y}^{(a)} \mid \mathbf{y}^{(b)} \mid \mathbf{y}^{(c)})$ 。

2 要凑出含 $t_0 t_p t_q$ （其中 $p, q \geq 1$ ）的项，就是三槽里恰好一个取 $\mathbf{y}^{(0)}$ 、其余取 $\mathbf{y}^{(p)}$ 和 $\mathbf{y}^{(q)}$ 。合并这些贡献， $t_0 t_p t_q$ 的系数正比于 $T(\mathbf{y}^{(0)} \mid \mathbf{y}^{(p)} \mid \mathbf{y}^{(q)})$ 。

- 3 而第九节已确保 $T(\mathbf{y}^{(0)} \mid \mathbf{y}^{(p)} \mid \mathbf{y}^{(q)}) = 0$ (当 p, q 都在 1 到 m 之间, 即 $p, q \geq 1$)。所以所有" t_0 出现一次、另两个因子都来自 $t_1, \dots, t_m$ "的项全部消失。

- 4 剩下能含 t_0 的, 只有 t_0 出现三次 (即 t_0^3) 或出现两次 (即 $t_0^2 \cdot t_p$) 的项。于是型必为

$$a_0 t_0^3 + t_0^2(b_1 t_1 + \dots + b_m t_m) + g_1(t_1, \dots, t_m),$$

其中 a_0 是 t_0^3 的系数, b_i 是 $t_0^2 t_i$ 的系数, g_1 是完全不含 t_0 的那部分 (t_0 出现零次)。注意没有 t_0 的一次项 $\times (t_p t_q)$ ——它们正是被我们消掉的。

译文段落 (第 19 段, 消去 b 项)。**"如果 $b_1, \dots, b_m$ 全为 0, 我们便得到一个所求类型 (11.1) 的型。如果不, 设 $b_m \neq 0$, 我们令 $t_m = -\frac{1}{b_m}(b_1 t_1 + \dots + b_{m-1} t_{m-1})$, 从而得到一个类型 (11.1) 的型, 但其中 m 被 $m-1$ 代替。"**

如何干掉碍事的 $t_0^2(b_1 t_1 + \dots)$ 项

- 1 **幸运情形**: 若 $b_1 = \dots = b_m = 0$, 那个中间项整体消失, 型已经是 $a_0 t_0^3 + g_1(t_1, \dots, t_m)$, 正是 (11.1), 收工。

- 2 **一般情形**: 至少一个 b 非零, 不妨设 $b_m \neq 0$ 。我们想让 $b_1 t_1 + \dots + b_m t_m = 0$ 永远成立, 从而中间项 $t_0^2 \cdot 0 = 0$ 自动消失。为此把 t_m 不再当作自由变量, 而是定义成

$$t_m = -\frac{1}{b_m}(b_1 t_1 + \dots + b_{m-1} t_{m-1}).$$

代入后确实有 $b_1 t_1 + \cdots + b_m t_m = 0$ (直接验证:
 $b_m t_m = -(b_1 t_1 + \cdots + b_{m-1} t_{m-1})$, 移项即得)。

- 3 这等于把变量个数从 $\{t_1, \dots, t_m\}$ 减少到 $\{t_1, \dots, t_{m-1}\}$ (t_m 被前面的线性表出, 不再独立)。于是型变成 $a_0 t_0^3 + g'_1(t_1, \dots, t_{m-1})$ ——还是 (11.1) 的形状, 只是 m 变成了 $m - 1$ 。代价就是少了一个独立变量, 这就是"变量浪费"的具体来源。

译文段落 (第 20 段, $a_0 \neq 0$)。"我们可以假设 $a_0 \neq 0$, 因为若 $a_0 = 0$, 则对应于 $t_0 = 1, t_1 = \cdots = t_m = 0$, 存在一个 $x_1, \dots, x_n$ 的解。"

为什么 $a_0 = 0$ 又是"白送解"

- 1 若 $a_0 = 0$, 型成了 $g_1(t_1, \dots, t_m)$, 根本不含 t_0 。
- 2 取 $t_0 = 1, t_1 = \cdots = t_m = 0$, 则型值 $= 0$ (g_1 在 t 全零处为 0, 且没有 t_0 项)。
- 3 回代换元, $\mathbf{x} = 1 \cdot \mathbf{y}^{(0)} = \mathbf{y}^{(0)} \neq \mathbf{0}$, 得到 $f(\mathbf{y}^{(0)}) = 0$ 的非平凡解, 又直接出解。所以"麻烦"的情形只剩 $a_0 \neq 0$, 我们就假设它。

十一、反复剥离 $\rightarrow$ 加性型 $\rightarrow$ 调用定理 8.1

译文段落 (第 21 段)。"我们看到, 只要取 n 足够大, 就可以让 m 任意大。同样, 我们也可以对新的型 $g(t_1, \dots, t_m)$ 重复这一过程……存在一个函数 $n_0(s)$, 使得若 $n \geq n_0(s)$, 则型 f 表示一个如下类型的型 $b_1 u_1^3 + \cdots + b_s u_s^3$ 。"

把"剥一项"滚雪球成"剥成纯立方和"

1 第七到第十节证明了：只要 $n > \frac{1}{2}m(m+1)$ ， f 就能表示成 $a_0 t_0^3 + g_1(t_1, \dots, t_m)$ ，其中 g_1 是变量更少的新三次型，且可假设它不表示零（否则又白送解）。

2 对 g_1 重复同样的操作：又剥出一个纯立方项，记 $b_0 u_0^3$ ，剩下更小的三次型。如此一层层剥下去。

3 每剥一次得到一个纯立方项 $b_i u_i^3$ 。要剥出 s 个这样的项，需要起初的变量数 n 足够大；把"需要多大"记作函数 $n_0(s)$ 。于是只要 $n \geq n_0(s)$ ， f 就表示加性型 $b_1 u_1^3 + \dots + b_s u_s^3$ 。这正是把"一般三次型"成功化成了"加性三次型"。

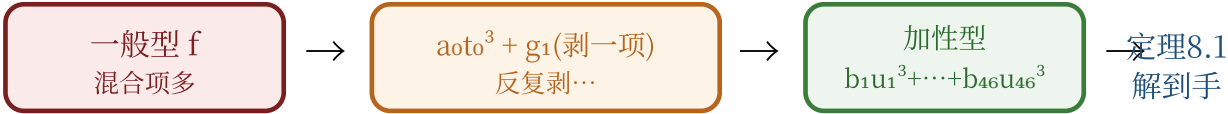
译文段落（第 22 段，收尾）。"由定理 8.1，只要 $s \geq s_0 (= 46)$ ，比方说，相应的齐次方程便有非平凡解，从而当 $n \geq n_0(s_0)$ 时原方程可解。"

最后一步合龙

1 取 $s = 46$ （即 s_0 ）。由第六节的定理 8.1，加性三次方程 $b_1 u_1^3 + \dots + b_{46} u_{46}^3 = 0$ 一定有非平凡整数解。

2 由第七节"表示"的好处： g 的非平凡零点回代成 f 的非平凡零点。

3 所以只要 $n \geq n_0(46)$ ，原一般三次方程 $f = 0$ 就有非平凡有理解。单个三次方程的 Birch 型结论证毕。整条逻辑链是： n 大 $\implies$ 能剥出 46 个立方项 $\implies$ 加性型可解 $\implies f$ 可解。



代价：每剥一层都浪费变量，所以起初的 n 必须很大

核心流水线：用线性代数把含混合项的一般型逐层剥成纯立方和，再交给圆法（定理 8.1）收尾。

十二、Birch 一般定理（定理 11.1）的陈述

译文段落（第 23 段）。引入定理 11.1。

定理 11.1 (Birch)

设 h, m 是正整数，并设 $r_1, \dots, r_h$ 是任意奇正整数。那么存在一个数 $\Psi(r_1, \dots, r_h; m)$ 具有下列性质。设 $f_{r_1}(\mathbf{x}), \dots, f_{r_h}(\mathbf{x})$ 是 $\mathbf{x} = (x_1, \dots, x_n)$ 中分别具有有理系数、次数为 $r_1, \dots, r_h$ 的型。那么，只要 $n \geq \Psi(r_1, \dots, r_h; m)$ ，就存在一个 m 维有理向量空间，其中所有的点都满足

$$f_{r_1}(\mathbf{x}) = 0, \dots, f_{r_h}(\mathbf{x}) = 0.$$

h （方程个数）

要同时满足的方程有几条。读作“ h ”。

$r_1, \dots, r_h$ （各方程的次数）

第 j 个方程的次数 r_j ，**必须都是奇数**。这是定理的核心限制（避开偶次方的符号陷阱）。

m (解空间维数)

结论保证的不只是"有一个解", 而是有一整个 m 维的解空间。 m 你想多大都行, 只要 n 跟着够大。

$\Psi(r_1, \dots, r_h; m)$ (变量门槛)

读作"普西" (希腊字母 Psi)。它是一个只由各次数和 m 决定的数 (门槛): 变量数 n 一旦达到它, 结论就成立。定理只断言"这个门槛存在", 不追求它具体多大。

m 维有理向量空间

指由 m 个线性无关的有理点张成的所有线性组合构成的集合 (像过原点的" m 维平面")。"其中所有点都满足方程"是非常强的结论。

译文段落 (第 24 段)。"注意我们断言的比'存在无穷多个解'更强; 这是为了在归纳证明中方便。"

为什么要"加强"成"整个 m 维空间"

只要解空间维数 $m \geq 1$, 里面就已经有无穷多个点 (一条过原点的有理直线上有无穷多个有理点), 所以" m 维解空间"比"无穷多解"更强。为什么不嫌麻烦要证更强的命题? 因为证明用数学归纳法: 归纳的下一步需要"在前一步得到的整个解空间上"继续操作。如果只知道"有无穷多个孤立解", 没法在上面再做线性代数; 而"有一整个 m 维空间"才能当作新的舞台递归下去。这是"为了归纳走得动, 故意把结论提强"的经典手法。

十三、引理 11.1: 加性方程的解含 m 维空间

译文段落 (第 25 段)。引理 11.1 及其证明。

引理 11.1

存在一个数 $\Phi(r, m)$, 对正整数 m, r (其中 r 为奇数) 有定义, 使得若 $s \geq \Phi(r, m)$, 则任何如下形式的方程

$$c_1 x_1^r + \cdots + c_s x_s^r = 0, \quad (c_j \text{ 有理})$$

都被某个 m 维有理线性空间中的所有点满足。

$\Phi(r, m)$ (读作"斐", 希腊字母 Phi)。是专为加性方程设的门槛: 变量数 s 达到它, 就能保证解里含一个 m 维空间。它和上面的 Ψ (一般型门槛) 配套, 是定理 11.1 证明的"地基石"。

证 (引理 11.1) .

1 先解一小撮。由定理 8.1 或定理 10.1, 存在数 $t = t(r)$, 使得任何 t 个变量的加性 r 次方程 $c_1 y_1^r + \cdots + c_t y_t^r = 0$ 都有非零整数解 $(y_1, \dots, y_t)$ 。关键: 因为 r 是奇数, 对系数符号没有任何要求 (奇次方 y^r 可正可负, 正负项能互相抵消); 又若某个 $c_j = 0$, 取对应变量为 1、其余为 0 就是显然解。

2 把变量分组。把 s 个变量切成若干段, 每段 t 个: 第一段 $y_1, \dots, y_t$, 第二段 $y_{t+1}, \dots, y_{2t}$, 依此类推。每一段都按第 1 步各自解出一组非零整数解。

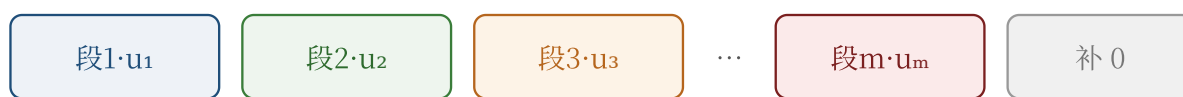
3 拼出 m 维空间。若 $s \geq mt$, 至少能切出 m 段。引入 m 个自由参数 $u_1, \dots, u_m$, 构造点

$$(u_1 y_1, \dots, u_1 y_t, u_2 y_{t+1}, \dots, u_2 y_{2t}, \dots, u_m y_{(m-1)t+1}, \dots, u_m y_{mt})$$

即第 i 段坐标整体乘上参数 u_i , 多出的变量补 0。

4 验证它恒满足方程。把这个点代入 $\sum c_j x_j^r$: 第 i 段贡献 $u_i^r \cdot (\text{第 } i \text{ 段的 } \sum c_j y^r) = u_i^r \cdot 0 = 0$ (每段本就是该段方程的解); 补 0 的部分贡献 0。所以对一切 $u_1, \dots, u_m$, 方程值都是 0。

5 它确实是 m 维空间。当 $u_1, \dots, u_m$ 独立取遍有理数时, 这些点张成一个 m 维有理线性空间 (m 个参数互不牵连)。于是结论成立, 门槛可取 $\Phi(r, m) = mt$ 。 ■



每段各 t 个变量、各自满足方程; m 个参数 $u_1 \cdots u_m$ 张成 m 维解空间

引理 11.1 的构造: 分段独立解出, 再用 m 个自由参数拼成整块 m 维解空间。

十四、定理 11.1 的证明 (一): 归纳框架与多重线性函数

译文段落 (第 26 段)。"设 $R = \max r_j$, 于是 R 是一个奇正整数。当 $\max r_j = 1$ 时结果当然成立。我们对 R (限于奇数值) 作归纳来证明它。我们将首先证明: 若结果对 $\max r_j \leq R - 2$ 的方程组成立, 则它对单个次数为 R 的方程成立。"

$R = \max r_j$ 。 $\max$ 取一组数里最大的那个。这里 R 是所有方程次数里最高的。所有 r_j 都奇, 所以 R 奇。证明对 R 做数学归纳法, 但因为次数只取奇数, 所以是 "1, 3, 5, 7, ..." 地往上走, 相邻奇数差 2, 故归纳是 "从 $\leq R - 2$ 推到 R "。

归纳法的两根支柱

奠基 ($R = 1$): 次数全是 1, 方程是线性的 $f_{r_j}(\mathbf{x}) = 0$ 。一组齐次线性方程, 只要变量够多就有 m 维解空间 ("变量多于方程"那条免费午餐), 所以结论显然成立。

归纳步: 分两小步。① **先证**"若结论对所有 $\max r_j \leq R - 2$ 的方程组成立, 则它对单个次数 R 的方程也成立" (十四—十六节)。② **再把**单个方程推广到一组次数 $\leq R$ 的方程 (十七节)。两步合起来完成归纳。

译文段落 (第 27 段)。定义多重线性函数:

$$f(x_1, \dots, x_n) = \sum c_{i_1, \dots, i_R} x_{i_1} \dots x_{i_R}, \quad M(\mathbf{x}^{(1)} \mid \dots \mid \mathbf{x}^{(R)}) = \sum c_{i_1, \dots, i_R}$$

多重线性函数 M (R 个槽的推广版 T)。这是把第八节的三线性 T 从 3 个槽推广到 R 个槽: f 是 R 次型, 它的每一项是 R 个变量相乘 $x_{i_1} \dots x_{i_R}$; 把这 R 个因子分别放到 R 个不同的点 $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(R)}$ 上, 第 ℓ 个因子取第 ℓ 个点的坐标, 就得到 M 。它对每个槽都线性, 且 R 个槽全取同一点时还原成 f :

$M(\mathbf{x} \mid \dots \mid \mathbf{x}) = f(\mathbf{x})$ 。 M 取 "multilinear (多重线性)" 的首字母。作用和 T 完全一样: 把高次的非线性问题, 转化成 "在若干槽里线性" 的问题, 好用线性方程组求解。

十五、证明 (二): 造出特殊点 $\mathbf{y}^{(0)}$ 与子空间 $\mathcal{Y}_1$

译文段落 (第 28 段)。"我们先选取 h 个线性无关的点 $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(h)}$ 。设 $\mathcal{Y}$ 是由它们生成的 h 维线性空间。考虑方程

$$M(\underbrace{\mathbf{y} \mid \dots \mid \mathbf{y}}_{\rho} \mid \mathbf{y}^{(p_1)} \mid \dots \mid \mathbf{y}^{(p_{R-\rho})}) = 0,$$

其中 ρ 取从 1 到 $R - 2$ 的所有奇数值，而 $p_1, \dots, p_{R-\rho}$ 取从 1 到 h 的所有值。"

$\mathcal{Y}$ (花体 Y) 和"生成的空间"。 $\mathcal{Y}$ 是 $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(h)}$ 的所有线性组合 $u_1 \mathbf{y}^{(1)} + \dots + u_h \mathbf{y}^{(h)}$ 构成的集合，是一个 h 维线性空间 (这 h 个点是它的"坐标轴")。 ρ (读"柔",rho)、 σ (读"西格玛",sigma)、 τ (读"陶",tau) 都是希腊字母，这里当"在某个槽里重复放同一个点的次数"的计数器。下方花括号 $\underbrace{\quad}_{\rho}$ 表示" $\mathbf{y}$ 连续出现 ρ 次"。

这组方程是关于 $\mathbf{y}$ 的、次数 $\leq R - 2$ 的奇次方程组

1 固定后面 $R - \rho$ 个槽为已知点 $\mathbf{y}^{(p_1)}, \dots, \mathbf{y}^{(p_{R-\rho})}$ ，前 ρ 个槽都放未知点 $\mathbf{y}$ 。由于 M 对每个槽线性，前 ρ 个槽放同一个 $\mathbf{y}$ ，使表达式成为 $\mathbf{y}$ 的 ρ 次齐次型。所以每个方程关于 $\mathbf{y}$ 的次数是 ρ 。

2 ρ 只取 1 到 $R - 2$ 的奇数，所以这些方程都是奇次、且次数 $\leq R - 2$ 。这正好落进归纳假设的适用范围 ($\max \leq R - 2$)。

3 数方程总数。 ρ 的取值不超过 R 种；对每个 ρ ，后面 $R - \rho$ 个下标各自从 1 到 h 独立取值，至多 $h^{R-\rho} \leq h^R$ 种组合。所以总数 \(\leq h^R (R-1)!\)。

译文段落 (第 29 段)。"因此，由归纳假设 (对 $m = 1$ 的情形)，只要 $n \geq n_0(R, h)$ ，这些方程在 $\mathbf{y}$ 中就有了一个非零解。记这样的解为 $\mathbf{y}^{(0)}$ 。我们现在有

$$M(\underbrace{\mathbf{y}^{(0)} \mid \dots \mid \mathbf{y}^{(0)}}_{\rho} \mid \underbrace{\mathbf{y} \mid \dots \mid \mathbf{y}}_{R-\rho}) = 0$$

对 $\mathcal{Y}$ 中所有 $\mathbf{y}$ 以及所有奇数 $\rho \leq R - 2$ 成立。"

补全：从"找到一个解 $\mathbf{y}^{(0)}$ "到"对整个 $\mathcal{Y}$ 成立"

1 上面那组方程都是次数 $\leq R - 2$ 的奇次方程，由归纳假设（取 $m = 1$ ），只要变量数 n 够大（记门槛 $n_0(R, h)$ ），它们有公共非零解。挑一个记为 $\mathbf{y}^{(0)}$ 。

2 于是对所有奇数 $\rho \leq R - 2$ 和所有 $1 \leq p_1, \dots, p_{R-\rho} \leq h$ ：
 $M(\underbrace{\mathbf{y}^{(0)} \dots \mathbf{y}^{(0)}}_{\rho} \mid \mathbf{y}^{(p_1)} \mid \dots) = 0$ 。注意这里 $\mathbf{y}^{(0)}$ 是放在前 ρ 个槽（用线性性，把原来" $\mathbf{y}$ 当未知"换成" $\mathbf{y}^{(0)}$ 当已知"，等式照样成立）。

3 升级到任意 $\mathbf{y} \in \mathcal{Y}$ 。 $\mathcal{Y}$ 中任意点 $\mathbf{y} = u_1 \mathbf{y}^{(1)} + \dots + u_h \mathbf{y}^{(h)}$ 。把它代入后 $R - \rho$ 个槽，用 M 的多重线性把每个槽展开成 $\sum u_p \mathbf{y}^{(p)}$ ，整体变成各 $M(\mathbf{y}^{(0)} \dots \mid \mathbf{y}^{(p_1)} \mid \dots)$ 的线性组合——而每一项都已是 0。所以

$$M(\underbrace{\mathbf{y}^{(0)} \mid \dots \mid \mathbf{y}^{(0)}}_{\rho} \mid \underbrace{\mathbf{y} \mid \dots \mid \mathbf{y}}_{R-\rho}) = 0 \quad \text{对所有 } \mathbf{y} \in \mathcal{Y}.$$

这是用"在 h 个基点上成立 + 线性性 $\implies$ 在整个空间上成立"，下面还会再用。

译文段落（第 30–31 段）。" $\mathcal{Y}$ 中一个任意点 $\mathbf{y}$ 具有形式 $u_1 \mathbf{y}^{(1)} + \dots + u_h \mathbf{y}^{(h)}$ 。现在考虑方程

$$M(\underbrace{\mathbf{y}^{(0)} \mid \dots \mid \mathbf{y}^{(0)}}_{R-\sigma} \mid \underbrace{\mathbf{y} \mid \dots \mid \mathbf{y}}_{\sigma}) = 0,$$

其中 σ 取从 1 到 $R - 2$ 的所有奇数值……这些是关于 $u_1, \dots, u_h$ 的次数 $\leq R - 2$ 的奇次方程，其个数 \(\backslash(\)

第二轮：再压出一个低维好空间 $\mathcal{Y}_1$

1 上一步处理的是" $\mathbf{y}^{(0)}$ 出现 ρ (少) 次"的情形。现在反过来看" $\mathbf{y}^{(0)}$ 出现 $R - \sigma$ (多) 次、 $\mathbf{y}$ 出现 σ (少, 奇数) 次"的等式 $M(\underbrace{\mathbf{y}^{(0)} | \cdots | \mathbf{y}^{(0)}}_{R-\sigma} | \underbrace{\mathbf{y} | \cdots | \mathbf{y}}_{\sigma}) = 0$

。

2 把 $\mathbf{y} = u_1 \mathbf{y}^{(1)} + \cdots + u_h \mathbf{y}^{(h)}$ 代入, 由于 σ 个槽放 $\mathbf{y}$, 等式成为关于系数 $u_1, \dots, u_h$ 的 σ 次齐次方程。 σ 取 1 到 $R - 2$ 的奇数, 所以这些是关于 u 的、次数 $\leq R - 2$ 的奇次方程, 个数 \(\backslash(\)

3 这些方程的变量是 $u_1, \dots, u_h$ (共 h 个)。又落进归纳假设范围。这次我们要的不是单个解, 而是"给定任意目标维数 ℓ , 存在 $\ell + 1$ 维的解空间": 由归纳假设, 只要 $h \geq h_0(R, \ell)$, 这些关于 u 的方程会被 $\mathcal{Y}$ 的某个 $(\ell + 1)$ 维有理子空间里所有点满足。把这个子空间记为 $\mathcal{Y}_1$, 不妨设它由 $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(\ell+1)}$ 生成。

译文段落 (第 32 段)。**"我们便有**

$$M(\underbrace{\mathbf{y}^{(0)} | \cdots | \mathbf{y}^{(0)}}_{\tau} | \underbrace{\mathbf{y} | \cdots | \mathbf{y}}_{R-\tau}) = 0$$

对所有 $\tau = 1, \dots, R - 1$ 以及 $(\ell + 1)$ 维空间 $\mathcal{Y}_1$ 中所有 $\mathbf{y}$ 成立。为保证这一点, 我们只需假设 $n \geq n_1(R, \ell)$ 。"

两轮合并: 所有 τ (从 1 到 $R-1$) 都被搞定

1 第一轮 (第 29 段) 给出: 当 $\mathbf{y}^{(0)}$ 出现奇数次 $\rho \leq R - 2$ 时, 等式对所有 $\mathbf{y} \in \mathcal{Y}$ (从而对子空间 $\mathcal{Y}_1$) 成立。

2 第二轮 (第 31 段) 给出: 当 $\mathbf{y}^{(0)}$ 出现 $R - \sigma$ 次 (即 $\mathbf{y}$ 出现奇数次 $\sigma \leq R - 2$) 时, 等式对 $\mathcal{Y}_1$ 中所有 $\mathbf{y}$ 成立。

- 3 把两轮的 τ ($=\mathbf{y}^{(0)}$ 出现的次数) 合起来。因为 R 是奇数, $R - \tau$ 与 τ 一奇一偶; 当 τ 从 1 跑到 $R - 1$, " τ 奇"和" $R - \tau$ 奇"两类合起来恰好覆盖全部 $\tau = 1, \dots, R - 1$ (第一轮管一半、第二轮管另一半)。于是

$$M(\underbrace{\mathbf{y}^{(0)} \mid \dots \mid \mathbf{y}^{(0)}}_{\tau} \mid \underbrace{\mathbf{y} \mid \dots \mid \mathbf{y}}_{R-\tau}) = 0, \quad \tau = 1, \dots, R - 1, \forall \mathbf{y} \in \mathcal{Y}_1.$$

- 4 为同时满足两轮对变量数的要求, 取 $n \geq n_0(R, h_0(R, \ell)) =: n_1(R, \ell)$ 即可 (先要 h 够大撑起第二轮, 再要 n 够大撑起第一轮, 两个门槛复合)。

十六、证明 (三): 换元、剥出 $a_0 t_0^R$, 再化加性型

译文段落 (第 33 段)。“如果 $\mathbf{y}^{(0)}$ 在由 $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(\ell+1)}$ 生成的子空间 $\mathcal{Y}_1$ 中, 我们可以通过去掉其中一个生成点 (比方说 $\mathbf{y}^{(\ell+1)}$), 得到 $\mathcal{Y}_1$ 的一个 ℓ 维子空间 $\mathcal{Y}_2$, 它不含 $\mathbf{y}^{(0)}$ 。现在点 $\mathbf{y}^{(0)}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(\ell)}$ 是线性无关的。”

为什么能凑出“含 $\mathbf{y}^{(0)}$ 在内、共 $\ell + 1$ 个无关点”

- 1 我们想做一个秩 $\ell + 1$ 的换元, 需要 $\ell + 1$ 个线性无关的方向, 其中一个 是 $\mathbf{y}^{(0)}$ 。
- 2 若 $\mathbf{y}^{(0)} \notin \mathcal{Y}_1$: 那 $\mathbf{y}^{(0)}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(\ell)}$ 自动无关, 直接用。
- 3 若 $\mathbf{y}^{(0)} \in \mathcal{Y}_1$ (译文讨论的情形): $\mathcal{Y}_1$ 是 $\ell + 1$ 维的、由 $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(\ell+1)}$ 生成。去掉一个生成点 (比如 $\mathbf{y}^{(\ell+1)}$), 剩下 $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(\ell)}$ 张成一个 ℓ 维子空间 $\mathcal{Y}_2$, 可安排得使 $\mathbf{y}^{(0)} \notin \mathcal{Y}_2$ 。于是 $\mathbf{y}^{(0)}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(\ell)}$ 这 $\ell + 1$ 个点线性无关。

- 4 关键是上一节的结论对 $\mathcal{Y}_1$ 成立，自然对其中的 $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(\ell)}$ 也成立，下一步换元用得上。

译文段落（第 34 段）。"线性变换 $x_j = t_0 y_j^{(0)} + t_1 y_j^{(1)} + \dots + t_\ell y_j^{(\ell)}$ ($1 \leq j \leq n$) 的秩为 $\ell + 1$ ，并给出 $f(\mathbf{x}) = g(t_0, \dots, t_\ell)$ 。在 g 中 $t_0^\tau t_{p_1} \dots t_{p_{R-\tau}}$ 的系数……是

$$M(\underbrace{\mathbf{y}^{(0)} | \dots | \mathbf{y}^{(0)}}_{\tau} | \mathbf{y}^{(p_1)} | \dots | \mathbf{y}^{(p_{R-\tau})}) = 0,$$

而这对 $\tau = 1, \dots, R - 1$ 以及任意选取的 $p_1, \dots, p_{R-\tau}$ 都成立。因此 $g(t_0, \dots, t_\ell) = a_0 t_0^R + g_1(t_1, \dots, t_\ell)$ 。"

这是第十节"剥立方"在 R 次的翻版

- 1 用 $\ell + 1$ 个无关方向作新坐标轴，换元 $\mathbf{x} = t_0 \mathbf{y}^{(0)} + t_1 \mathbf{y}^{(1)} + \dots + t_\ell \mathbf{y}^{(\ell)}$ ，秩 $\ell + 1$ ，把 f 表示成 $\ell + 1$ 元的 R 次型 $g(t_0, \dots, t_\ell)$ 。

- 2 展开 $f(\mathbf{x}) = M(\mathbf{x} | \dots | \mathbf{x})$ ，含 $t_0^\tau t_{p_1} \dots t_{p_{R-\tau}}$ 的项 (t_0 出现 τ 次，其余 p_j 取 1 到 ℓ) 的系数恰是

$$M(\underbrace{\mathbf{y}^{(0)} | \dots | \mathbf{y}^{(0)}}_{\tau} | \mathbf{y}^{(p_1)} | \dots | \mathbf{y}^{(p_{R-\tau})}).$$

- 3 对 $\tau = 1, \dots, R - 1$ ，由第十五节末这些系数全是 0。也就是说： t_0 只要出现"至少一次、但不到 R 次"的所有混合项，系数都为 0，被一扫而空。

- 4 剩下的项只有两类： t_0 出现 R 次（即 $a_0 t_0^R$ ），和 t_0 出现 0 次（完全不含 t_0 的型 $g_1(t_1, \dots, t_\ell)$ ）。于是

$$g(t_0, \dots, t_\ell) = a_0 t_0^R + g_1(t_1, \dots, t_\ell).$$

注意： R 次情形比三次干净得多——三次时还剩一个 $t_0^2(\dots)$ 中间项要靠“令 t_m 受约束”消掉；这里因为把全部 $\tau = 1, \dots, R-1$ 一次性清零，连中间项都没有了。这正是前面费力造出 $\mathcal{Y}_1$ 、覆盖所有 τ 的回报。

译文段落（第 35 段）。"于是，对任何 ℓ ，只要 $n \geq n_1(R, \ell)$ ， f 就表示一个上述类型的型。重复这一论证可证明 f 表示一个如下类型的型 $\underbrace{a_0 t_0^R + b_0 u_0^R + \dots}_s$ ，

只要 $n \geq n_2(R, s)$ 。由引理 11.1……因此原方程的解包含一个 m 维线性空间，只要 $n \geq n_2(R, \Phi(R, m))$ 。"

滚雪球 + 调用引理，单个 R 次方程证毕

- 1 上一步把 f 剥成 $a_0 t_0^R + g_1$ ，其中 g_1 是变量更少的新 R 次型。对 g_1 重复同样过程，再剥出一个 R 次纯幂项 $b_0 u_0^R$ ，依此类推。

- 2 要剥出 s 个纯 R 次幂项 $a_0 t_0^R + b_0 u_0^R + \dots$ （共 s 项，这是加性 R 次型），需要起初的 $n \geq n_2(R, s)$ （门槛随剥的层数复合增大）。

- 3 对这个加性型用引理 11.1：只要项数 $s \geq \Phi(R, m)$ ，把它令为 0 的方程的解里就含一个 m 维线性空间。

- 4 由“表示”的传递性， $f = 0$ 的解也含这个 m 维空间。取 $s = \Phi(R, m)$ ，则只要 $n \geq n_2(R, \Phi(R, m))$ ，单个 R 次方程的定理 11.1 成立。归纳步的第①小步完成。

十七、证明（四）：从单个方程推广到一组方程

译文段落（第 36 段）。"现在设有 h 个方程 $f_{r_1} = 0, \dots, f_{r_h} = 0$ ，其中每个 $r_j \leq R$ 。我们对 h 作归纳……给定 m_1 ，由定理对 $h = 1$ 的情形，存在一个 m_1 维有理线性空间，在其上 $f_{r_h} = 0$ ，只要 $n \geq \Psi(r_h; m_1)$ 。"

对方程个数 h 再做一层归纳

1 $h = 1$ 已证（上一节，单个方程）。现在假设结论对 $h - 1$ 个方程成立，要推到 h 个。

2 先用一个方程砍出一个大空间。挑出最后一个方程 $f_{r_h} = 0$ 。由 $h = 1$ 的结论，只要 $n \geq \Psi(r_h; m_1)$ ，存在一个 m_1 维有理空间，在它上面 f_{r_h} 处处为 0。把这个空间的点参数化为 $\mathbf{x} = v_1 \mathbf{x}^{(1)} + \dots + v_{m_1} \mathbf{x}^{(m_1)}$ (v 是新的 m_1 个自由参数)。 m_1 待定，先取得足够大。

译文段落（第 37 段）。"对这些点，型 $f_{r_1}, \dots, f_{r_{h-1}}$ 成为 $v_1, \dots, v_{m_1}$ 中的型。由 $h - 1$ 的情形，存在一个 m 维线性空间，在其上它们全部消失，只要 $m_1 \geq \Psi(r_1, \dots, r_{h-1}; m)$ 。因此……

$$\Psi(r_1, \dots, r_h; m) = \Psi(r_h; \Psi(r_1, \dots, r_{h-1}; m)).$$

"

在那个空间上解剩下的 $h-1$ 个方程

1 把参数化 $\mathbf{x} = v_1 \mathbf{x}^{(1)} + \dots + v_{m_1} \mathbf{x}^{(m_1)}$ 代入其余 $h - 1$ 个型 $f_{r_1}, \dots, f_{r_{h-1}}$ 。因为代入的是线性表达式，每个 f_{r_j} 变成关于新变量 $v_1, \dots, v_{m_1}$ 的型，次数不变（仍 $\leq R$ ，且仍奇）。

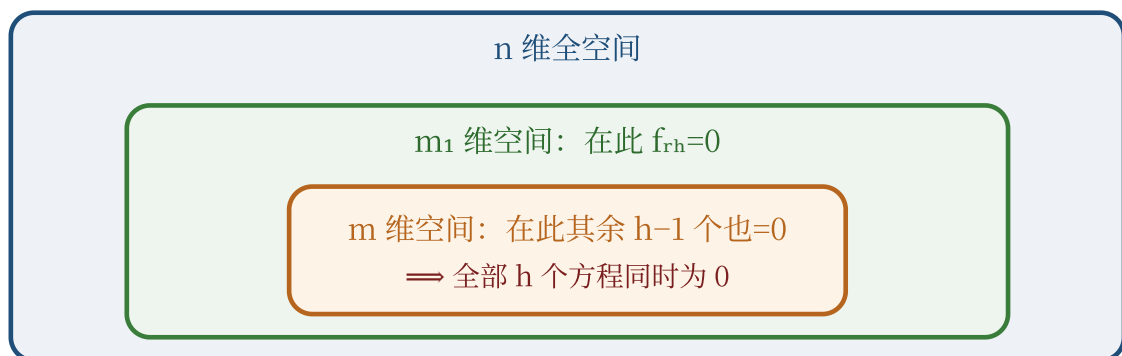
2 现在这是关于 v (共 m_1 个变量) 的 $h - 1$ 个方程的方程组。由归纳假设 ($h - 1$ 个方程的情形), 只要变量数 $m_1 \geq \Psi(r_1, \dots, r_{h-1}; m)$, 存在一个 m 维有理空间, 在其上这 $h - 1$ 个型全部为 0。

3 这个 m 维空间是原 m_1 维空间的子空间, 而在原 m_1 维空间上 f_{r_h} 本就处处为 0。所以在这个 m 维空间上, 全部 h 个型同时为 0。 h 个方程的情形成立。

4 门槛的复合公式。要让上述成立, 需 $m_1 = \Psi(r_1, \dots, r_{h-1}; m)$, 再要 $n \geq \Psi(r_h; m_1)$, 把两层套起来:

$$\Psi(r_1, \dots, r_h; m) = \Psi(r_h; \Psi(r_1, \dots, r_{h-1}; m)).$$

这是一个"门槛套门槛"的递归式——它解释了为什么 Birch 定理要求的变量数会大得惊人 (每多一个方程就把门槛函数嵌套一次)。证明全部完成。 ■



多方程的处理: 层层套娃。先在大空间满足一个方程, 再在其子空间满足下一个, 门槛函数随之嵌套。

十八、推论: 推广到代数数域

译文段落（第 38 段）。"推论. 即使型的系数与 $x_1, \dots, x_n$ 取值于一个代数数域 K 中, 定理 11.1 仍然成立, 但此时 Ψ 依赖于 K 。"

"代数数域 K "。简单说, 是把有理数 $\mathbb{Q}$ 添加上某个代数数 θ (某个整系数多项式的根, 例如 $\sqrt{2}$) 后得到的"扩大的数系"。 ν (次数) 是 K 相对 $\mathbb{Q}$ 的维数, $\{1, \theta, \dots, \theta^{\nu-1}\}$ 是它的一组"基"—— K 里每个数都能唯一写成 $a_0 + a_1\theta + \dots + a_{\nu-1}\theta^{\nu-1}$ (a_i 有理)。高中阶段类比: 就像复数都能写成 $a + bi$, 这里是写成 θ 的幂的有理组合。

译文段落（第 39 段）。"当我们把每个系数与每个变量用 K 的一组基……线性地表示出来时, 每个方程都等价于 ν 个具有相同次数、带有理系数与变量的方程。"

为什么" K 上一个方程 = $\mathbb{Q}$ 上 ν 个方程"

1 把每个变量 x_j 写成基的有理组合 $x_j = x_{j0} + x_{j1}\theta + \dots + x_{j,\nu-1}\theta^{\nu-1}$, 系数也照此展开。代入 $f = 0$ 后, 整理成 θ 的多项式:
 $A_0 + A_1\theta + \dots + A_{\nu-1}\theta^{\nu-1} = 0$, 其中各 A_i 是新有理变量 $x_{j0}, x_{j1}, \dots$ 的型。

2 由于 $1, \theta, \dots, \theta^{\nu-1}$ 是基 (线性无关), 这个等式为 0 当且仅当每个分量 $A_0 = A_1 = \dots = A_{\nu-1} = 0$ 。一个 K 上的方程, 就裂成 ν 个 $\mathbb{Q}$ 上、同样次数 (仍奇) 的方程。

3 于是 K 上的问题完全化归为 $\mathbb{Q}$ 上的方程组问题, 再套用已证的定理 11.1 (方程个数变成 ν 倍, 变量个数也按基展开变多, 所以门槛 Ψ 会依赖 K , 即依赖 ν 和 θ)。推论得证。

十九、承上启下：下一步要更精确

译文段落 (第 40 段)。"我们接下来的目标将是证明一个更精确的结果，它关于使单个齐次三次方程可解所需的变量个数。在工作过程中，我们将需要数的几何中的一些结果，因此我们将以对这一主题某些方面的叙述作为开始。"

本章地位 & 下章预告

本章证明的 Birch 定理只断言"**存在**一个门槛 Ψ "，并没有算出它具体多大——而我们一路看到，门槛函数是层层嵌套的，会大得离谱（"变量浪费"严重）。

所以接下来的目标是**定量精确化**：针对最受关注的"单个三次方程"，问"到底需要多少个变量才够"，给出尽量好的具体界。为此需要一门新工具——**数的几何** (geometry of numbers, 研究格点与凸体的学问)，下一章将从它讲起。本章的定性结论（奇次 + 变量足够多 $\implies$ 可解）是地基，下一章在它上面追求更锋利的定量结果。

全章一句话总结

Birch 定理的证明=**两件武器的组合**：① 多重线性函数 M 把高次型的"剥离一个纯幂项"变成可解的线性方程组（靠"变量多于方程"反复制造特殊点 $\mathbf{y}^{(0)}$ 和好子空间）；② 对次数 R 和方程个数 h 的**双重数学归纳法**，把一般问题层层归约到加性方程，最后由圆法成果（定理 8.1 / 引理 11.1）一锤定音。奇次的条件，自始至终是为了让奇次方"可正可负"、绕开偶次方程的符号死局。

返回 [全书目录](#)