

DAVENPORT · 圆法 · 高中详解版

一个丢番图不等式

A Diophantine inequality

本章要解决什么

给五个实数 $\lambda_1, \dots, \lambda_5$ (不全同号, 且它们之间至少有一个比值是无理数), 我们要找不全为零的整数 $x_1, \dots, x_5$, 使得

$$|\lambda_1 x_1^2 + \dots + \lambda_5 x_5^2| < \varepsilon$$

对任意预先给定的小正数 ε 都能办到。这就是 1945 年 Davenport 和 Heilbronn 的著名定理。

读完本章你能掌握: ① 圆法 (Hardy–Littlewood circle method) 处理不等式而非方程的完整套路; ② 怎样用一个"指示函数"把不等式翻译成积分; ③ 怎样把积分拆成"主区间 + 中段 + 尾巴"分别估计; ④ Weyl 不等式、连分数渐近分数、Dirichlet 逼近这些工具如何在关键一步发挥作用。我们对原文每一段、每一个符号、每一处跳步都从零讲透。

读前必备的几样"新工具"

本章用到几个超出高中课本的记号。下面先一次性讲清, 后面正文遇到时还会再点。

复指数 $e(\theta) = e^{2\pi i \theta}$

这里 i 是虚数单位 ($i^2 = -1$)。欧拉公式说 $e^{i\phi} = \cos \phi + i \sin \phi$ 。所以

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

它是复平面单位圆上的一个点, 幅角为 $2\pi\theta$ (即转了 θ 圈)。关键性质: $e(\theta)$ 的周期是 1 (θ 增加 1 就转回原处), 而且 $|e(\theta)| = 1$ (模长恒为 1), 以及

$$e(\theta_1)e(\theta_2) = e(\theta_1 + \theta_2)。$$

复数的模 $|z|$

若 $z = a + bi$, 则 $|z| = \sqrt{a^2 + b^2}$, 是 z 到原点的距离。它满足 $|z_1 z_2| = |z_1| |z_2|$ 和三角不等式 $|z_1 + z_2| \leq |z_1| + |z_2|$ 。

求和号 $\sum_{x=1}^P f(x)$

就是 $f(1) + f(2) + \cdots + f(P)$ 的简写, 把括号里随 x 变化的东西从 $x = 1$ 一直加到 $x = P$ 。

定积分 $\int_a^b f(x) dx$

把区间 $[a, b]$ 切成无数小段, 每段宽 dx 、高 $f(x)$, 把这些细长条面积加起来取极限。可理解为"连续版的求和"。 $\int_{-\infty}^{\infty}$ 表示从负无穷积到正无穷。

大 O 与 $\ll$ 记号

写 $A = O(B)$ 或 $A \ll B$ (两者同义), 意思是"存在一个与变量无关的常数 C , 使得 $|A| \leq C B$ "。它只关心"增长量级", 不关心精确常数。例如 $3n + 5 = O(n)$, $7n^2 = O(n^2)$ 。

写 $A \gg B$ 就是 $B \ll A$, 即 $A \geq cB$ (某正常数 c)。

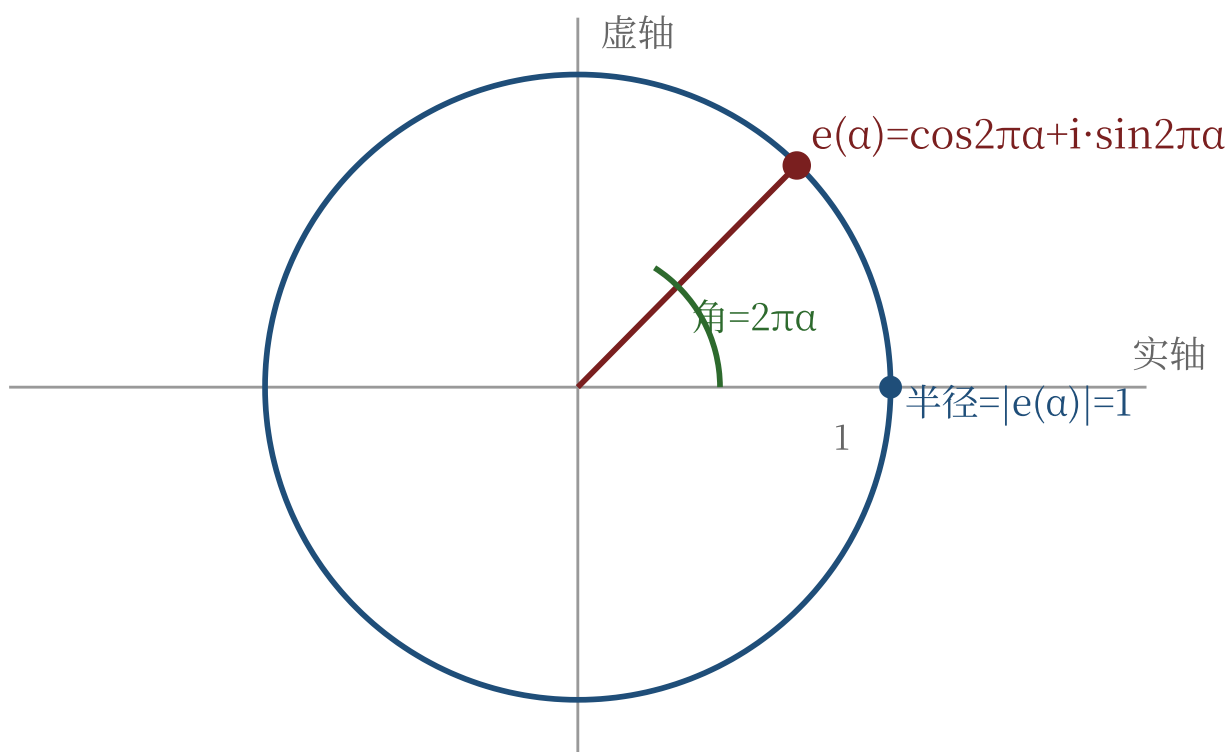
写 $A = o(B)$ (小 o) 更强, 意思是"当变量趋于极限时, 比值 $A/B \rightarrow 0$ ", 即 A 比 B 增长慢得多。例如 $P^2 = o(P^3)$ 。

整除与互素 $(a, q) = 1$

(a, q) 表示整数 a, q 的**最大公约数**。 $(a, q) = 1$ 就是说它们互素 (没有大于 1 的公因数), 即分数 a/q 已是最简形式。

$\mathbb{Q}$ 、 $\mathbb{Z}$

$\mathbb{Q}$ 是全体有理数 (分数) 的集合; $\mathbb{Z}$ 是全体整数的集合。 $x \notin \mathbb{Q}$ 表示 x 不是有理数, 即**无理数**。



复指数 $e(\alpha) = e^{2\pi i \alpha}$ 永远落在单位圆上，模长是 1，幅角是 $2\pi\alpha$ 。 α 每增加 1，点就绕圆整整一圈回到原位——这就是"周期为 1"。圆法之所以叫"圆法"，正因为核心对象都活在这个单位圆上。

一、引子：什么是丢番图不等式

在丢番图不等式这一课题中，我们的目标是用整数解出某种给定类型的不等式，通常涉及具有任意实系数的多项式或形式。数的几何为研究线性不等式的可解性提供了有用的方法，并给出了关于高次多项式不等式的某些信息，但对于后者它的力量是有限的。

这段在讲什么

"丢番图（Diophantus）问题"指的是只允许用**整数**来求解的问题。中学里你解方程允许任意实数解；丢番图问题却规定未知数只能取整数。

本章研究的不是方程 " $= 0$ "，而是不等式 " $|\dots| < \epsilon$ "：我们不要求那个式子精确等于零，只要求它能**任意接近**零。这两件事差别很大：如果系数里有无理数，往往根本无法让式子精确为 0，但"无限接近 0"却可能办到。

"系数是任意**实数**"是关键难点。如果系数都是整数，这就是普通的丢番图方程问题；一旦系数可以是无理数（比如 $\sqrt{2}$ ），整数代进去算出来的几乎都是无理数，要让它接近 0 需要全新的方法。

"数的几何"（geometry of numbers，Minkowski 创立的把整点看成格点、用凸体体积估计格点的理论）能很好地处理**线性**不等式（一次式），但面对**二次及更高次**的式子就力不从心了。所以本章要换一套武器——**圆法**。

次数高于一次的最简单的丢番图不等式是

$$|\lambda_1 x_1^2 + \cdots + \lambda_n x_n^2| < C.$$

逐符号读这条式子

$\lambda_1, \dots, \lambda_n$ （读作"拉姆达 1 到 n"）

给定的 n 个**实数系数**，可以是无理数。它们是题目给的、固定不变的。

$x_1, \dots, x_n$

我们要寻找的**整数**未知数。

x_j^2

每个未知数都**平方**了，所以这是个二次（次数为 2）的表达式。这种"各变量平方加权求和"的式子叫**对角二次型**（或加性二次型）。

$|\cdots| < C$

整个加权和的**绝对值**小于给定正数 C ，也就是要它落在 $(-C, C)$ 这个小区间里。

为什么从"平方和"开始？因为它是"高于一次"里最简单的：次数恰好是 2，而且各变量分开（没有 $x_1 x_2$ 这类交叉项）。先把最简单的情形彻底拿下，再去想更复杂的形式。

二、历史：从 Oppenheim 猜想到 Davenport–Heilbronn 定理

基于与第 11 章中遇到的 Meyer 定理的类比，Oppenheim 在 1929 年猜想：只要 $\lambda_1, \dots, \lambda_n$ 是不全同号的实数，那么当 $n \geq 5$ 时，对所有 $C > 0$ ，该不等式都应可解。当

然，如果 $\lambda_1, \dots, \lambda_n$ 之间的比值都是有理的，我们就能令左端为零，因此问题关涉的是这些比值不全为有理数的情形。

这段的来龙去脉

人物与年代。 Alexander Oppenheim 在 1929 年提出这个猜想。他的灵感来自 **Meyer 定理**（本书第 11 章）：Meyer 1884 年证明，一个**整系数**的不定二次型，只要变量个数 ≥ 5 ，就一定能取到 0（即方程 $Q(\mathbf{x}) = 0$ 有非零整数解）。Oppenheim 大胆类比：把“整系数 + 等于 0”换成“实系数 + 任意接近 0”，门槛是不是也是 5 个变量？

“不全同号”是什么意思、为什么必须。 如果 $\lambda_1, \dots, \lambda_n$ 全是正数，那么 $\lambda_1 x_1^2 + \dots + \lambda_n x_n^2$ 是一堆非负数相加，最小值在 x_j 全为 0 时取到 0；但题目要求 x_j 不全为零，于是这个和至少是 $\min \lambda_j$ （一个固定的正数），永远没法 $< \varepsilon$ 。所以必须有正有负，这样正项和负项才能**相互抵消**，把总和压到接近 0。这种有正有负的二次型叫**不定二次型**（indefinite form）。

“比值都有理就能精确取 0”这句怎么理解。 假设所有比值 λ_i/λ_j 都是有理数。那么可以把所有 λ_j 同时写成“某个公共实数 t 乘以有理数”： $\lambda_j = t \cdot r_j$ ， $r_j \in \mathbb{Q}$ 。再通分让 r_j 都成整数 m_j （同乘一个公分母不改变式子是否为 0）。于是问题变成整系数不定二次型能否取 0——由 Meyer 定理（ $n \geq 5$ ）**能**，而且是精确等于 0，自然也 $< \varepsilon$ 。所以“比值全有理”是平凡情形，已经解决。真正有挑战的是**比值不全为有理数**的情形——这时候无法精确取 0，只能逼近。

1934 年 Chowla [13] 证明了当 $n \geq 9$ 时该结果成立；他是从 Jarník 和 Walfisz [51] 关于大椭圆内整点个数的结果推出来的。1945 年，Davenport 和 Heilbronn [26] 证明了它对 $n \geq 5$ 成立，本章主要致力于阐述这一证明。应当指出，尽管变量个数的界 5 在某种意义上是尽可能好的，但在更深层的意义上它很可能并非如此。如果我们假定比值 λ_i/λ_j 不全为有理数，那么（就我们所知）该结果对 $n \geq 3$ 也可能成立。

进展时间线与“界 5”的微妙之处

- **1934 年 Chowla** ($n \geq 9$)。他借用 Jarník–Walfisz 关于“大椭圆内有多少整点”的计数结果，间接推出 9 个变量够用。这步还比较粗糙。

- 1945 年 Davenport–Heilbronn ($n \geq 5$)。他们用圆法直接攻克，把门槛降到 5，恰好匹配 Oppenheim 的猜想。这正是本章要详细讲的证明。
- "界 5 是 best possible 的两种含义。"说它"在某种意义上最好"：因为存在反例表明 4 个变量时结论会失败（所以不能再降到 4 的一般情形）。但又说"更深层意义上很可能不是最好"：那个 4 变量反例所有比值都是有理的。如果额外假设"比值不全有理"（恰是本章的假设），人们猜测 $n \geq 3$ 就够了——只是至今没人能证。这告诉你：数学结论的"最优性"依赖于你加了哪些前提。

三、要证明的定理与第一步化简

正式地表述，我们要证明的结果是：

定理 20.1. 设 $\lambda_1, \dots, \lambda_5$ 是实数，没有一个为 0，且既非全正也非全负。设其中至少有一个比值 λ_i/λ_j 是无理的。那么对任意 $\varepsilon > 0$ ，存在不全为 0 的整数 $x_1, \dots, x_5$ ，使得

$$|\lambda_1 x_1^2 + \dots + \lambda_5 x_5^2| < \varepsilon.$$

把定理的每个条件读懂

"没有一个为 0"

若某个 $\lambda_j = 0$ ，则对应的 x_j 在式子里消失，相当于变量更少，会让问题退化。所以排除。

"既非全正也非全负"

即不定，前面解释过：要靠正负抵消才能接近 0。

"至少有一个比值 λ_i/λ_j 无理"

排除"全有理"的平凡情形——那种情形 Meyer 定理已直接给出精确解。

"对任意 $\varepsilon > 0$ "

无论你把目标精度定得多苛刻（ ε 多小），都能找到整数解。注意找到的 x_j 会随 ε 而变， ε 越小往往需要越大的 x_j 。

"不全为 0"

必须排除 $x_1 = \cdots = x_5 = 0$ 这个废解（它让式子等于 0 但毫无意义）。

不失一般性，我们可以假定

$$\lambda_1 > 0, \quad \lambda_5 < 0, \quad \lambda_1/\lambda_2 \notin \mathbb{Q}.$$

"不失一般性"凭什么这么假定

"不失一般性" (without loss of generality) 是说：这些额外假定只是**重新命名/调整顺序**，不会削弱定理，证明了特殊摆放就等于证明了一般情形。逐条解释：

1 $\lambda_1 > 0, \lambda_5 < 0$ 。因为既有正的又有负的，我们就把变量**重新编号**，让第 1 个系数是正的、第 5 个是负的。下标只是名字，随便换。

2 $\lambda_1/\lambda_2 \notin \mathbb{Q}$ 。题目只保证"某一对"比值无理。无理性其实可以"传染"到含 λ_1 的某个比值上：如果 λ_i/λ_j 无理，那么 λ_1/λ_i 与 λ_1/λ_j 至少有一个无理（否则两者都有理，相除得 λ_j/λ_i 有理，矛盾）。于是总能选出一个与 λ_1 有关的无理比值，再把那个变量编号为第 2 个。这样就凑出了 λ_1/λ_2 无理。

这一步的好处：把"无理性"这张唯一的王牌，集中安放在 λ_1, λ_2 这一对上，后面用起来方便。

只需证明下式可解即可

$$|\lambda_1 x_1^2 + \cdots + \lambda_5 x_5^2| < 1, \quad (20.1)$$

因为在最后这个不等式中把 $\lambda_1, \dots, \lambda_5$ 替换为 $\lambda_1/\varepsilon, \dots, \lambda_5/\varepsilon$ ，就可由此推出表面上更一般的不等式的可解性。

为什么把 ε 换成 1 不损失一般性

1 定理要的是 $|\sum \lambda_j x_j^2| < \varepsilon$ 。两边同除以 $\varepsilon > 0$ （除以正数不改变不等号方向）：

$$\left| \frac{\lambda_1}{\varepsilon} x_1^2 + \cdots + \frac{\lambda_5}{\varepsilon} x_5^2 \right| < 1.$$

2 记 $\mu_j = \lambda_j/\varepsilon$ 。这组新系数 μ_j 仍满足定理全部条件：没有 0、不全同号（同除正数不改变符号）、 $\mu_1/\mu_2 = \lambda_1/\lambda_2$ 仍无理。

3 所以"对任意系数证明 < 1 可解"这一件事，自动覆盖了"对任意 ε 证明 $< \varepsilon$ 可解"。于是只需主攻把右端固定为 1 的式子 (20.1)。

这是一个常见手法：把一个"带参数 ε "的命题，通过缩放变量/系数，归约成一个"参数固定"的命题，少操一份心。

四、第一块砖：把不等式翻译成积分（引理 20.1）

第一步是构造一个实变量 Q 的函数，它在 $|Q| < 1$ 时为正，在 $|Q| \geq 1$ 时为零。下面的引理给出了这样一个函数，但也存在各种类似的函数。

为什么要造这样一个函数（动机）

这是整个圆法的灵魂一招。我们想"数"有多少组整数 $(x_1, \dots, x_5)$ 让 $Q = \lambda_1 x_1^2 + \cdots + \lambda_5 x_5^2$ 满足 $|Q| < 1$ 。如果手里有一个函数 $K(Q)$ ，它在 $|Q| < 1$ 时为正、在 $|Q| \geq 1$ 时为 0，那么把 $K(Q)$ 对所有整点求和：

$$\sum_{x_1, \dots, x_5} K(Q)$$

就只剩下满足 $|Q| < 1$ 的那些点有贡献——这等于把"满足不等式的解"自动筛选出来了。若能证明这个和 > 0 ，就说明至少存在一组满足 $|Q| < 1$ 的整数解。"造一个能把不等式翻译成可计算量的函数"——这就是动机。

为什么要"它在 $|Q| \geq 1$ 时严格为 0"？因为我们只关心 $|Q| < 1$ 的解，不想让 $|Q| \geq 1$ 的点掺进来污染计数。为什么要" $|Q| < 1$ 时为正"？因为只有正贡献，才能从"和 > 0 "推出"确有解"。

引理 20.1. 我们有

$$\int_{-\infty}^{\infty} e(\alpha Q) \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha = \begin{cases} 1 - |Q|, & |Q| \leq 1, \\ 0, & |Q| \geq 1. \end{cases}$$

先认识这个被积函数

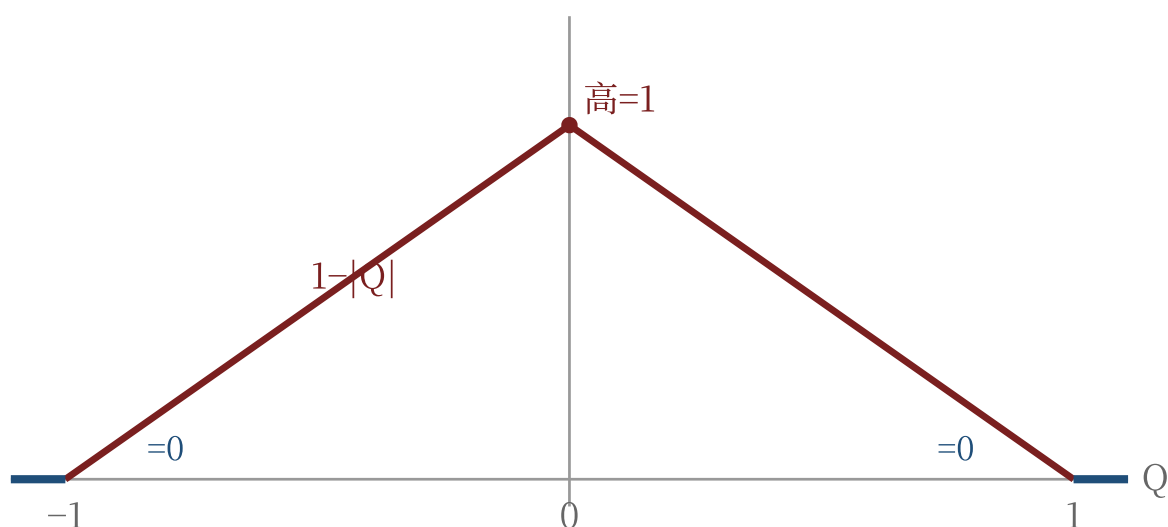
$$\left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2$$

这是函数 sinc 的平方。 $\frac{\sin \pi \alpha}{\pi \alpha}$ 在 $\alpha = 0$ 处理解为极限值 1 (因为 $\lim_{t \rightarrow 0} \frac{\sin t}{t} = 1$)，随着 $|\alpha|$ 增大像波浪一样衰减。平方后恒非负，且衰减得更快 (约像 $1/\alpha^2$)，所以从 $-\infty$ 积到 $+\infty$ 收敛。它是个"集中在原点附近的钟形权重"。

$$e(\alpha Q) = e^{2\pi i \alpha Q}$$

带着参数 Q 的复指数。把 Q 当常数、 α 当积分变量。

整条积分的结果是一个只依赖 Q 的实数：在 $|Q| \leq 1$ 处等于 $1 - |Q|$ (一个三角形帽子， $Q = 0$ 时最高为 1，到 $|Q| = 1$ 时降到 0)，在 $|Q| \geq 1$ 处恒等于 0。这正是我们要的"指示型"函数。



引理 20.1 给出的函数：一个底边 $[-1, 1]$ 、顶点在 $(0, 1)$ 的三角形帽子，区间外恒为 0。它在 $|Q| < 1$ 处为正，正好用来"筛"出满足不等式的解。

证. 众所周知

$$\int_{-\infty}^{\infty} \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha = 1.$$

因此对任意实数 η ,

$$\int_{-\infty}^{\infty} \left(\frac{\sin \pi \eta \alpha}{\pi \alpha} \right)^2 d\alpha = |\eta|.$$

这给出

$$\begin{aligned} \int_{-\infty}^{\infty} e(i\alpha Q) \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha &= \int_{-\infty}^{\infty} \cos 2\pi \alpha Q \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha \\ &= \frac{1}{2} \int_{-\infty}^{\infty} \frac{\sin^2 \pi \alpha (Q+1) + \sin^2 \pi \alpha (Q-1) - 2 \sin^2 \pi \alpha}{(\pi \alpha)^2} d\alpha \\ &= \frac{1}{2} \{|Q+1| + |Q-1| - 2|Q|\}, \end{aligned}$$

由此得到结论。 ■

把这个证明的每一步补全

1 起点公式 $\int_{-\infty}^{\infty} (\sin \pi \alpha / \pi \alpha)^2 d\alpha = 1$ 。这是著名的"sinc 平方积分" (Fejér 核的总质量), 属于分析里的标准结论。它说那个钟形权重的总面积恰好是 1。这里直接当已知用。

2 缩放得到 $\int (\sin \pi \eta \alpha / \pi \alpha)^2 d\alpha = |\eta|$ 。在第 1 步积分里做**换元** $\beta = \eta \alpha$ (设 $\eta \neq 0$)。则 $\alpha = \beta / \eta$, $d\alpha = d\beta / \eta$, 而 $\pi \alpha = \pi \beta / \eta$ 。代入:

$$\int \left(\frac{\sin \pi \eta \alpha}{\pi \alpha} \right)^2 d\alpha = \int \left(\frac{\sin \pi \beta}{\pi \beta / \eta} \right)^2 \frac{d\beta}{\eta} = \eta^2 \cdot \frac{1}{\eta} \int \left(\frac{\sin \pi \beta}{\pi \beta} \right)^2 d\beta = \eta.$$

若 $\eta < 0$, 换元会让积分上下限对调并多出一个负号, 最终得到 $|\eta|$ (积分本身非负, 所以取绝对值)。这条是后面三处 " $\sin^2 / \alpha^2$ 积分" 的公式来源。

3 复指数化成余弦: 丢掉虚部。由欧拉公式 $e(i\alpha Q) = \cos 2\pi \alpha Q + i \sin 2\pi \alpha Q$ 。被积函数另一半 $(\sin \pi \alpha / \pi \alpha)^2$ 是**偶函数** (关于 $\alpha \rightarrow -\alpha$ 不变)。而 $\sin 2\pi \alpha Q$ 是**奇**

函数，奇 $\times$ 偶=奇，奇函数在对称区间 $(-\infty, \infty)$ 上的积分为0。于是虚部消失，只剩余弦项 $\cos 2\pi\alpha Q \cdot (\sin \pi\alpha/\pi\alpha)^2$ 。这解释了第一个等号。

- 4 关键三角恒等式。要把 $\cos 2\pi\alpha Q \cdot \sin^2 \pi\alpha$ 拆成三个 $\sin^2$ 之差。用降幂公式 $\sin^2 \theta = \frac{1}{2}(1 - \cos 2\theta)$:

$$\begin{aligned} & \sin^2 \pi\alpha(Q+1) + \sin^2 \pi\alpha(Q-1) - 2\sin^2 \pi\alpha Q \\ &= \frac{1}{2} [(1 - \cos 2\pi\alpha(Q+1)) + (1 - \cos 2\pi\alpha(Q-1)) - 2(1 - \cos 2\pi\alpha Q)] \\ &= \frac{1}{2} [2\cos 2\pi\alpha Q - \cos 2\pi\alpha(Q+1) - \cos 2\pi\alpha(Q-1)]. \end{aligned}$$

再用和差化积 $\cos(A+B) + \cos(A-B) = 2\cos A \cos B$ ，取 $A = 2\pi\alpha Q$, $B = 2\pi\alpha$:

$$\cos 2\pi\alpha(Q+1) + \cos 2\pi\alpha(Q-1) = 2\cos 2\pi\alpha Q \cos 2\pi\alpha.$$

代回:

$$= \frac{1}{2} [2\cos 2\pi\alpha Q - 2\cos 2\pi\alpha Q \cos 2\pi\alpha] = \cos 2\pi\alpha Q (1 - \cos 2\pi\alpha) = \cos 2\pi\alpha Q \cdot 2\sin^2 \pi\alpha$$

所以

$\cos 2\pi\alpha Q \cdot \sin^2 \pi\alpha = \frac{1}{2} [\sin^2 \pi\alpha(Q+1) + \sin^2 \pi\alpha(Q-1) - 2\sin^2 \pi\alpha Q]$ 。把它除以 $(\pi\alpha)^2$ 并积分，就得到第二个等号（那个 $\frac{1}{2}$ 和分子的三项）。

- 5 用第2步的公式逐项积分。对每一项用 $\int (\sin \pi\eta\alpha/\pi\alpha)^2 d\alpha = |\eta|$: 第一项 $\eta = Q+1$ 给 $|Q+1|$ ，第二项 $\eta = Q-1$ 给 $|Q-1|$ ，第三项 $\eta = Q$ 给 $2|Q|$ （前面带-2）。整体乘 $\frac{1}{2}$ ，得

$$\frac{1}{2} \{|Q+1| + |Q-1| - 2|Q|\}.$$

- 6 分情况算出最终结果。

- 若 $|Q| \leq 1$ （即 $-1 \leq Q \leq 1$ ）： $Q+1 \geq 0$ 故 $|Q+1| = Q+1$ ； $Q-1 \leq 0$ 故 $|Q-1| = 1-Q$ 。相加 $= 2$ 。于是 $\frac{1}{2}\{2 - 2|Q|\} = 1 - |Q|$ 。
- 若 $Q \geq 1$ ： $|Q+1| = Q+1$ ， $|Q-1| = Q-1$ ，相加 $= 2Q$ ，而 $2|Q| = 2Q$ ，差为0。

◦ 若 $Q \leq -1$: 对称地同样得 0。
合起来正是引理所断言的分段结果。

五、把"求和"与"积分"两条线并排 (生成函数 S, I)

设 P 是一个大的正整数。定义

$$S(\alpha) = \sum_{x=1}^P e(\alpha x^2), \quad I(\alpha) = \int_0^P e(\alpha x^2) dx.$$

认识两位主角 $S(\alpha)$ 与 $I(\alpha)$

P

一个我们自己挑选的**大正整数**，相当于"搜索范围的上限"——我们只在 $1 \leq x_j \leq P$ 里找解。 P 越大，候选越多。最后会让 $P \rightarrow \infty$ 。

$S(\alpha) = \sum_{x=1}^P e(\alpha x^2)$ (指数和 / 生成函数)

把 $x = 1, 2, \dots, P$ 的复指数 $e(\alpha x^2)$ 加起来。它是个随 α 变化的复数。**为什么引入它?** 因为后面 (20.2) 里会出现"对整点求和"，而每个 $e(\alpha \lambda_j x_j^2)$ 都能各自打包成 $S(\lambda_j \alpha)$ ，于是五重求和就分解成五个 S 的乘积——这正是圆法把多重和"因式分解"的威力。**它叫"生成函数"**，因为关于平方数的信息都被编码进这一个 α 的函数里。注意 S 以 1 为周期 (因为 x^2 是整数， e 周期为 1)。

$I(\alpha) = \int_0^P e(\alpha x^2) dx$ (对应的积分)

把 S 里的"对整数求和"换成"对连续变量积分"，是 S 的"光滑连续版本"。**为什么也要它?** 因为 I 容易精确计算和估计 (它本质是 Fresnel 积分)，可以当作 S 的"参照标准"。证明的主线就是说明在关键区域 $S \approx I$ ，从而把难算的求和换成好算的积分。

在引理 20.1 的结果中取 $Q = \lambda_1 x_1^2 + \dots + \lambda_5 x_5^2$ ，并对 $x_1, \dots, x_5$ 求和，我们得到

$$\int_{-\infty}^{\infty} S(\lambda_1 \alpha) \cdots S(\lambda_5 \alpha) \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha = \sum_{\substack{x_1, \dots, x_5 \\ |Q| < 1}} (1 - |Q|), \quad (20.2)$$

其中求和遍历满足 $1 \leq x_j \leq P$ 且服从 (20.1) 的整数。

逐步推出 (20.2)——这是核心恒等式

- 1 对每一组固定的整数 $(x_1, \dots, x_5)$ ，先看引理 20.1。取 $Q = \lambda_1 x_1^2 + \dots + \lambda_5 x_5^2$ ，引理 20.1 说

$$\int_{-\infty}^{\infty} e(\alpha Q) \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha = \begin{cases} 1 - |Q|, & |Q| \leq 1 \\ 0, & |Q| \geq 1. \end{cases}$$

- 2 把指数里的 Q 拆开。因为 e 满足 $e(A + B) = e(A)e(B)$ ：

$$e(\alpha Q) = e(\alpha(\lambda_1 x_1^2 + \dots + \lambda_5 x_5^2)) = e(\lambda_1 \alpha x_1^2) e(\lambda_2 \alpha x_2^2) \cdots e(\lambda_5 \alpha x_5^2).$$

这一步把"五个变量纠缠在一起"的指数，拆成了五个各管一个变量的因子。

- 3 对所有 $(x_1, \dots, x_5)$ （每个从 1 到 P ）求和，并与积分交换次序。左边变成

$$\sum_{x_1=1}^P \cdots \sum_{x_5=1}^P \int_{-\infty}^{\infty} e(\lambda_1 \alpha x_1^2) \cdots e(\lambda_5 \alpha x_5^2) \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha.$$

有限个求和与积分可以交换顺序（项数有限，绝对收敛），把求和移到积分内部：

$$\int_{-\infty}^{\infty} \left(\sum_{x_1=1}^P e(\lambda_1 \alpha x_1^2) \right) \cdots \left(\sum_{x_5=1}^P e(\lambda_5 \alpha x_5^2) \right) \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha.$$

这里用了"求和的乘积 = 乘积的求和"（即 $(\sum_i a_i)(\sum_j b_j) = \sum_{i,j} a_i b_j$ 的多重版本）。

- 4 每个括号就是一个 S 。因为 $\sum_{x=1}^P e(\lambda_j \alpha x^2) = S(\lambda_j \alpha)$ （把 S 定义里的 α 换成 $\lambda_j \alpha$ ）。于是左边正好是 (20.2) 的左端 $\int S(\lambda_1 \alpha) \cdots S(\lambda_5 \alpha) (\sin \pi \alpha / \pi \alpha)^2 d\alpha$ 。

- 5 右边由引理 20.1 直接给出。对每组 $(x_1, \dots, x_5)$ ，内层积分等于 $1 - |Q|$ （当 $|Q| < 1$ ）或 0（当 $|Q| \geq 1$ ）。所以求和后只剩 $|Q| < 1$ 的那些组有贡献，得

$$\sum_{\substack{1 \leq x_j \leq P \\ |Q| < 1}} (1 - |Q|).$$

这正是 (20.2) 的右端。

(20.2) 的意义：右端是一堆**正数**（每项 $1 - |Q|$ 在 $|Q| < 1$ 时为正）的和，而且每出现一组满足 (20.1) 的解就贡献一个正项。所以“右端 > 0 ” $\iff$ “存在满足 (20.1) 的整数解”。我们的全部目标就转化为：**证明左端那个积分是正的（而且很大）**。这就是把“找整数解”翻译成了“估计一个积分”。

类似地，对 $x_1, \dots, x_5$ 积分而非求和，我们得到

$$\int_{-\infty}^{\infty} I(\lambda_1 \alpha) \cdots I(\lambda_5 \alpha) \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha = \int \cdots \int (1 - |Q|) dx_1 \cdots dx_5, \quad (20.3)$$

其中积分遍历满足 $0 \leq x_j \leq P$ 且服从 (20.1) 的实变量。

(20.3) 是 (20.2) 的“连续对照版”

1 把上面推导里所有“对整数 x_j 求和 $\sum_{x_j=1}^P$ ”统统换成“对实数 x_j 积分 $\int_0^P dx_j$ ”。

2 于是每个 $\sum_x e(\lambda_j \alpha x^2)$ 变成 $\int_0^P e(\lambda_j \alpha x^2) dx = I(\lambda_j \alpha)$ 。左端成了五个 I 的乘积。

3 右端则把“对满足 (20.1) 的整点求和”换成“对满足 (20.1) 的实点（一块五维区域）积分”，得到 $\int \cdots \int (1 - |Q|) dx_1 \cdots dx_5$ 。

为什么要造这个“假想的连续版”？因为右端是个普通的多重定积分，可以老老实实算出量级（引理 20.2 会证它 $\gg P^3$ ）；它充当真实问题 (20.2) 的**标尺**。证明策略就是：先量好标尺 (20.3)，再证明 (20.2) 与 (20.3) 只差一点点。

六、证明的总蓝图

证明的总体思路是比较 (20.2) 与 (20.3)。容易证明当 $P \rightarrow \infty$ 时 (20.3) 的右端为 $\gg P^3$ （见下面的引理 20.2）。如果我们能够证明 (20.2) 与 (20.3) 的左端之差当 $P \rightarrow \infty$ 时为

$o(P^3)$ ，那么就能推出 (20.2) 的右端也是 $\gg P^3$ 。这就意味着 (20.1) 有 $\gg P^3$ 组整数解 $(x_1, \dots, x_5)$ ，且满足 $1 \leq x_j \leq P$ 。

把这套逻辑讲成一条清晰的因果链

- 1 (标尺) 引理 20.2 将证: (20.3) 右端 $\gg P^3$ 。由 (20.3)，它的左端积分也 $\gg P^3$ 。
- 2 (核心估计) 若能证: $\left[(20.2) \text{左端} \right] - \left[(20.3) \text{左端} \right] = o(P^3)$ ，即两者之差比 P^3 小一个量级。
- 3 (相减得结论) 那么 (20.2) 左端 = (20.3) 左端 + $o(P^3) \gg P^3$ (一个 $\gg P^3$ 的量加上一个相对小得多的 $o(P^3)$ ，仍 $\gg P^3$)。
- 4 由 (20.2) 这个恒等式，左端 $\gg P^3 \Rightarrow$ 右端 $\sum_{|Q| < 1} (1 - |Q|) \gg P^3$ 。
- 5 右端是正项和、且 $\gg P^3$ ，说明满足 $|Q| < 1$ 的整点不仅存在，而且数量级有 $\gg P^3$ 这么多。定理就证完了 (我们只要存在性，结果还白送了"解很多")。

所以全章的技术难点集中在第 2 步：估计 S -积分与 I -积分之差。这正是接下来一系列引理 20.3–20.8 要干的活。

我们将证明在 $\alpha = 0$ 附近有一个小区间，在其中 $S_j(\alpha)$ 与 $I_j(\alpha)$ 相差很小，并由此推出该区间对两个积分的贡献实际上相同 (引理 20.4)。容易证明对所有其他的 α ，对 (20.3) 左端积分的贡献可忽略不计。困难在于估计这样的 α 对 (20.2) 左端积分的贡献。正是在这里 (而且只在这里)，我们才用到 λ_1/λ_2 为无理数这一假设，并且我们不会对所有大的 P 证明所论结果，而只对某个特定序列证明。

把 α 轴分成三段——圆法的"主弧 / 次弧"思想

整个积分 $\int_{-\infty}^{\infty}$ 沿 α 轴展开。我们按 α 的大小把它切成三段，每段用不同手法对付。这种"主弧 (major arc) + 次弧 (minor arc)"的分割是圆法的标志性套路：

第一段： α 很小， $|\alpha| < (4\lambda P)^{-1}$ ("主弧")

这里 S 与 I 几乎相等 (引理 20.3)，所以 S -积分与 I -积分在这一段几乎贡献相同，且这一段就是产生主项 $\gg P^3$ 的地方 (引理 20.4)。

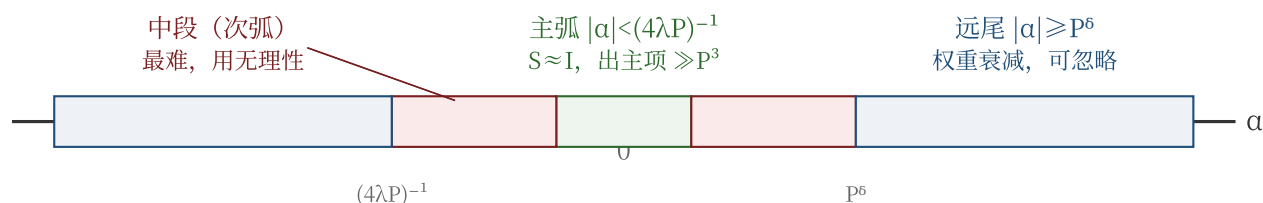
第三段： α 很大， $|\alpha| \geq P^\delta$ ("远尾")

$(\sin \pi \alpha / \pi \alpha)^2$ 这个权重像 $1/\alpha^2$ 衰减，再配合对 $|S|$ 的均方估计，这一段贡献可忽略 (引理 20.6)。

第二段：中段 $(4\lambda P)^{-1} < |\alpha| < P^\delta$ ("次弧"，真正的硬骨头)

这是整个证明最难、最关键的一段。 I 在这里贡献已被证明很小 (引理 20.4 顺带处理)，难的是证明 S 在这里贡献也是 $o(P^3)$ 。恰恰在这一段、也只在这一段，我们必须动用" λ_1/λ_2 无理"这张王牌 (引理 20.8)。

"只对某个特定序列的 P 证明"是什么意思？第二段的估计办不到对所有大 P 都成立——它依赖把 P 取成 λ_1/λ_2 连分数分母的平方 $P = q_0^2$ 。但定理只需要"存在解"，所以哪怕只对一串特殊的 P (这串里 P 仍然要多大有多大) 证明，就已经足够找到任意接近 0 的解了。这是一个聪明的让步：放弃"对所有 P "，换取"对一串 P 能严格证明"。



把 α 轴按 $|\alpha|$ 大小分三段处理：主弧 (绿，产生主项)、中段次弧 (红，最难，用 λ_1/λ_2 无理性)、远尾 (蓝，可忽略)。这是 Hardy-Littlewood 圆法的招牌分割。

七、量好标尺：主项 $\gg P^3$ (引理 20.2)

引理 20.2. 我们有

$$\int_{-\infty}^{\infty} I(\lambda_1 \alpha) \cdots I(\lambda_5 \alpha) \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha \gg P^3.$$

证. 在 (20.3) 的右端令 $|\lambda_i| x_i^2 = y_i$ 。除去一个常数因子外, 该积分变为

$$\int_0^{|\lambda_1|P^2} \cdots \int_0^{|\lambda_5|P^2} \{1 - |y_1 \pm y_2 \pm \cdots - y_5|\} (y_1 \cdots y_5)^{-1/2} dy_1 \cdots dy_5,$$

其中积分遍历满足 $|y_1 \pm y_2 \pm \cdots - y_5| < 1$ 的 $y_1, \dots, y_5$, 符号取 $\lambda_1, \dots, \lambda_5$ 之符号。

我们把变量 y_2, y_3, y_4 限制在区间 $\frac{1}{2}\gamma P^2 < y_j < \gamma P^2$ 内, 把 y_5 限制在区间 $4\gamma P^2 < y_5 < 5\gamma P^2$ 内, 并把 y_1 限制在区间

$$|y_1 \pm y_2 \pm y_3 \pm y_4 - y_5| < \frac{1}{2}.$$

那么所有剩下的点都满足 $0 < y_j < |\lambda_j|P^2$, 只要 $9\gamma < \min |\lambda_j|$ 。因此我们有积分区域的一部分, 其体积为 $\gg (P^2)^4$ 。在此区域中, 被积函数为

$$\gg (y_1 \cdots y_5)^{-1/2} \gg (P^{10})^{-1/2}.$$

因此该积分为 $\gg P^3$ 。 ■

逐步看懂这次换元与体积估计

1 换元 $y_i = |\lambda_i| x_i^2$ 的细节。 因为 $0 \leq x_i \leq P$, 所以 y_i 从 0 跑到 $|\lambda_i|P^2$ 。由 $y_i = |\lambda_i| x_i^2$ 得 $x_i = (y_i/|\lambda_i|)^{1/2}$, 于是

$$dx_i = \frac{1}{2} |\lambda_i|^{-1/2} y_i^{-1/2} dy_i.$$

五个变量都这样换, 雅可比 (各 dx_i 的乘积) 就产生因子 $\frac{1}{2} |\lambda_i|^{-1/2} \cdot y_i^{-1/2}$ 。把 $\prod \frac{1}{2} |\lambda_i|^{-1/2}$ 这堆不依赖 y 的常数提到积分外 (这就是"除去一个常数因子"), 剩下被积函数里出现 $(y_1 \cdots y_5)^{-1/2}$ 。

2 Q 变成 $\pm y$ 的带符号和。原来

$Q = \sum \lambda_i x_i^2 = \sum \operatorname{sgn}(\lambda_i) |\lambda_i| x_i^2 = \sum \operatorname{sgn}(\lambda_i) y_i$ 。因为约定 $\lambda_1 > 0, \lambda_5 < 0$ ，写成 $y_1 \pm y_2 \pm y_3 \pm y_4 - y_5$ （最后一项一定带减号，第一项带正号，中间符号视各 λ_i 而定）。条件 $|Q| < 1$ 就成了 $|y_1 \pm \cdots - y_5| < 1$ ，而 $1 - |Q|$ 就成了 $1 - |y_1 \pm \cdots - y_5|$ 。

3 为什么“只取一块子区域”就够。被积函数 $(1 - |\cdots|)(y_1 \cdots y_5)^{-1/2}$ 在积分域上处处非负（因为 $1 - |\cdots| > 0$ 且 $y_i^{-1/2} > 0$ ）。所以整个积分 $\geq$ 任何一块子区域上的积分。我们只要找到一块“又大、被积函数又不太小”的子区域，给出下界即可。这是估计下界的常用招：缩小积分域。

4 选定的子区域。令 γ 是一个固定的小正常数，满足 $9\gamma < \min_j |\lambda_j|$ 。在子区域里规定： $y_2, y_3, y_4 \in (\frac{1}{2}\gamma P^2, \gamma P^2)$ （各自长度 $\frac{1}{2}\gamma P^2$ ）； $y_5 \in (4\gamma P^2, 5\gamma P^2)$ （长度 γP^2 ）；最后让 y_1 落在使 $|y_1 \pm y_2 \pm y_3 \pm y_4 - y_5| < \frac{1}{2}$ 的范围内。

5 这块区域确实在原积分域内。需要验证两点：(a) 每个 $0 < y_j < |\lambda_j|P^2$ ；(b) $|y_1 \pm \cdots - y_5| < 1$ 。

对 (b)：我们要求的是 $< \frac{1}{2} < 1$ ，比原条件更严，自动满足。

对 (a)： $y_2, y_3, y_4 < \gamma P^2 < |\lambda_j|P^2$ （因 $\gamma < \min |\lambda_j|$ ）； $y_5 < 5\gamma P^2 < |\lambda_5|P^2$ （因 $5\gamma < 9\gamma < |\lambda_5|$ ）；而 y_1 由约束 $|y_1 \pm y_2 \pm y_3 \pm y_4 - y_5| < \frac{1}{2}$ 反解， y_1 约等于 $y_5 \mp y_2 \mp y_3 \mp y_4$ ，其值不超过

$y_5 + y_2 + y_3 + y_4 < 5\gamma P^2 + 3\gamma P^2 = 8\gamma P^2 < 9\gamma P^2 < |\lambda_1|P^2$ ，且因 y_5 较大、其余较小而保证 $y_1 > 0$ 。所以 $9\gamma < \min |\lambda_j|$ 这个条件正是为了让整块区域稳稳落在积分域内。

6 子区域的体积 $\gg (P^2)^4$ 。给定 y_2, y_3, y_4, y_5 后， y_1 被限制在一个长度为常数（约 1，与 P 无关）的小区间里。而 y_2, y_3, y_4, y_5 各自的取值区间长度都 $\propto P^2$ 。所以五维体积 $\approx (\text{常数}) \times (\frac{1}{2}\gamma P^2)^3 \times (\gamma P^2) \times (\text{常数}) \gg (P^2)^4 = P^8$ 。

7 被积函数的下界。在子区域里， $1 - |y_1 \pm \cdots - y_5| > 1 - \frac{1}{2} = \frac{1}{2}$ （常数级， $\gg 1$ ）。而每个 $y_i \leq 5\gamma P^2 \ll P^2$ ，所以

$(y_1 \cdots y_5)^{-1/2} \gg (P^2 \cdots P^2)^{-1/2} = (P^{10})^{-1/2} = P^{-5}$ 。故被积函数 $\gg P^{-5}$ 。

8 合成下界。 积分 $\geq$ (子区域体积) $\times$ (被积函数下界) $\gg P^8 \times P^{-5} = P^3$ 。证毕。

这就量好了标尺：连续模型 (20.3) 给出的主项是 P^3 的量级。它告诉我们后面所有"误差"只要能压到 $o(P^3)$ ，主项就稳住了。

八、主弧上 $S \approx I$ (引理 20.3)

引理 20.3. 若 $|\alpha| < (4\lambda P)^{-1}$ ，其中 $\lambda = \max |\lambda_j|$ ，则

$$S(\lambda_j \alpha) = I(\lambda_j \alpha) + O(1).$$

证. 这是引理 9.1 (van der Corput 引理) 的一个特例，其中 $f(x) = \lambda_j \alpha x^2$ 。我们有

$$|f'(x)| = |2\lambda_j \alpha x| \leq 2|\lambda_j \alpha|P \leq \frac{1}{2},$$

且 $f''(x)$ 符号固定。 ■

讲清这步在做什么、为什么成立

$$\lambda = \max |\lambda_j|$$

五个系数绝对值里最大的那个，固定常数。

$$S(\lambda_j \alpha) = I(\lambda_j \alpha) + O(1)$$

意思是：求和 S 与积分 I 只差一个**有界**的量 ($O(1)$ ：被某个不依赖 P, α 的常数控制)。注意 S, I 本身可以大到 P 的量级，而它们的差却只有 $O(1)$ ——所以在主弧上，相对而言 S 与 I 几乎一模一样。

1 van der Corput 引理 (本书引理 9.1) 说什么。 它是一个"求和 $\approx$ 积分"的定理：若 f 在 $[0, P]$ 上的导数满足 $|f'(x)| \leq \frac{1}{2}$ (变化足够缓和) 且 f'' 不变号 (单调)，则

$$\sum_{x=1}^P e(f(x)) = \int_0^P e(f(x)) dx + O(1).$$

直觉：当 f 变化平缓时，相邻整点上的 $e(f(x))$ 几乎一样，求和就很接近积分，误差只在端点处累积出 $O(1)$ 。这本质是 Euler-Maclaurin / 一阶导数控制下的"梯形逼近"思想。

2 代入 $f(x) = \lambda_j \alpha x^2$ 。这正是 $S(\lambda_j \alpha) = \sum e(\lambda_j \alpha x^2)$ 与 $I(\lambda_j \alpha) = \int e(\lambda_j \alpha x^2) dx$ 的相位函数。

3 验证 $|f'(x)| \leq \frac{1}{2}$ 。求导 $f'(x) = 2\lambda_j \alpha x$ 。在 $[0, P]$ 上 $x \leq P$ ，故 $|f'(x)| \leq 2|\lambda_j \alpha|P$ 。又 $|\lambda_j| \leq \lambda$ ，且条件给 $|\alpha| < (4\lambda P)^{-1}$ ，所以

$$2|\lambda_j \alpha|P \leq 2\lambda|\alpha|P < 2\lambda P \cdot \frac{1}{4\lambda P} = \frac{1}{2}.$$

恰好 $\leq \frac{1}{2}$ ，满足引理 9.1 的条件。这正是主弧半径取成 $(4\lambda P)^{-1}$ 的原因——这个数被精心选出，刚好让导数不超过 $\frac{1}{2}$ 。

4 验证 f'' 不变号。 $f''(x) = 2\lambda_j \alpha$ 是个常数，符号由 $\lambda_j \alpha$ 决定，全程不变号。满足条件。

5 两个条件都满足，van der Corput 引理直接给出 $S(\lambda_j \alpha) = I(\lambda_j \alpha) + O(1)$ 。

九、主弧贡献出主项（引理 20.4）

引理 20.4. 我们有

$$\int_{|\alpha| < (4\lambda P)^{-1}} S(\lambda_1 \alpha) \cdots S(\lambda_5 \alpha) \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha \gg P^3.$$

证. 首先我们注意到，对任意 α ，

$$|I(\lambda_j \alpha)| \ll \min(P, |\alpha|^{-1/2}).$$

估计 P 是显然的, 而估计 $|\alpha|^{-1/2}$ 则由

$$I(\lambda_j \alpha) = \int_0^P e(\lambda_j \alpha x^2) dx = \frac{1}{2} |\lambda_j \alpha|^{-1/2} \int_0^{|\lambda_j \alpha| P^2} t^{-1/2} e(\pm t) dt$$

推出, 因为最后这个积分是有界的。由引理 20.3 可知, 若 $|\alpha| < (4\lambda P)^{-1}$, 则也有

$$|S(\lambda_j \alpha)| \ll \min(P, |\alpha|^{-1/2})$$

成立。因此, 把这两个估计与引理 20.3 联用, 我们在上述区间内有

$$S(\lambda_1 \alpha) \cdots S(\lambda_5 \alpha) - I(\lambda_1 \alpha) \cdots I(\lambda_5 \alpha) \ll \min(P^4, \alpha^{-2}).$$

因此该差在 $|\alpha| < (4\lambda P)^{-1}$ 上的积分为 $O(P^2)$ 。

于是只需证明

$$\int_{|\alpha| < (4\lambda P)^{-1}} I(\lambda_1 \alpha) \cdots I(\lambda_5 \alpha) \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha \gg P^3.$$

我们已知对应于区间 $(-\infty, \infty)$ 的积分确实如此。现在由上面对 $I(\lambda_j \alpha)$ 的估计, 我们有

$$\int_{|\alpha| \geq (4\lambda P)^{-1}} |I(\lambda_1 \alpha) \cdots I(\lambda_5 \alpha)| d\alpha \ll \int_{|\alpha| \geq (4\lambda P)^{-1}} \alpha^{-5/2} d\alpha \ll P^{3/2}.$$

因此结论得证。 ■

把引理 20.4 的每一处估计补全

1 估计 $|I(\lambda_j \alpha)| \ll \min(P, |\alpha|^{-1/2})$ 。这是说 I 同时被两个上界压住, 取小的那个。

○ 上界 P :

$|I(\lambda_j \alpha)| = \left| \int_0^P e(\lambda_j \alpha x^2) dx \right| \leq \int_0^P |e(\lambda_j \alpha x^2)| dx = \int_0^P 1 dx = P$ (因为 $|e(\cdot)| = 1$)。这是"显然"的那半。

○ 上界 $|\alpha|^{-1/2}$: 在 I 里换元 $t = |\lambda_j \alpha| x^2$ 。则 $x = (t/|\lambda_j \alpha|)^{1/2}$,

$dx = \frac{1}{2} |\lambda_j \alpha|^{-1/2} t^{-1/2} dt$, x 从 0 到 P 对应 t 从 0 到 $|\lambda_j \alpha| P^2$ 。又

$\lambda_j \alpha x^2 = \pm t$ (符号取 $\lambda_j \alpha$ 的符号)。代入得

$$I(\lambda_j \alpha) = \frac{1}{2} |\lambda_j \alpha|^{-1/2} \int_0^{|\lambda_j \alpha| P^2} t^{-1/2} e(\pm t) dt.$$

最后那个积分 $\int_0^X t^{-1/2} e(\pm t) dt$ 是著名的 **Fresnel 型积分**：尽管被积函数在 0 附近有 $t^{-1/2}$ 的奇性、在远处不衰减，但由于 $e(\pm t)$ 的剧烈振荡相互抵消，这个积分对所有上限 X 都**一致有界**（即 $\leq$ 某常数，与 X 无关）。因此

$$|I(\lambda_j \alpha)| \ll |\lambda_j \alpha|^{-1/2} \ll |\alpha|^{-1/2} \quad (\lambda_j \text{ 是固定常数}).$$

两个上界都成立，所以取较小者： $|I(\lambda_j \alpha)| \ll \min(P, |\alpha|^{-1/2})$ 。直觉： α 很小时 $|\alpha|^{-1/2}$ 很大，是 P 占优； α 较大时 $|\alpha|^{-1/2}$ 小，由它占优。

2 同样的估计搬到 S 上。在主弧 $|\alpha| < (4\lambda P)^{-1}$ 里，引理 20.3 给

$$S(\lambda_j \alpha) = I(\lambda_j \alpha) + O(1), \text{ 所以}$$

$$|S(\lambda_j \alpha)| \leq |I(\lambda_j \alpha)| + O(1) \ll \min(P, |\alpha|^{-1/2}).$$

（那个 $O(1)$ 被 $\min(P, |\alpha|^{-1/2})$ 吸收，因为在主弧上这个 $\min$ 至少是常数级。）

3 估计两个乘积之差 $\ll \min(P^4, \alpha^{-2})$ 。记 $m = \min(P, |\alpha|^{-1/2})$ 。我们有每个 $|S_j|, |I_j| \ll m$ ，且 $S_j - I_j = O(1)$ 。用"逐项替换"的望远镜法：

$$S_1 S_2 S_3 S_4 S_5 - I_1 I_2 I_3 I_4 I_5 = \sum_{j=1}^5 I_1 \cdots I_{j-1} (S_j - I_j) S_{j+1} \cdots S_5.$$

（即一次换一个因子从 I 变 S ，逐项相消展开。）每一项里有 4 个 I 或 S 因子（各 $\ll m$ ）乘上一个差 $S_j - I_j = O(1)$ ，故每项 $\ll m^4 \cdot O(1) = O(m^4)$ 。五项相加仍 $\ll m^4$ 。而 $m^4 = \min(P, |\alpha|^{-1/2})^4 = \min(P^4, |\alpha|^{-2})$ 。这就解释了那条 $\ll \min(P^4, \alpha^{-2})$ 。

4 差的积分是 $O(P^2)$ 。把 $\int_{|\alpha| < (4\lambda P)^{-1}} \min(P^4, \alpha^{-2}) d\alpha$ 分两段算（用对称性只算 $\alpha > 0$ 再乘 2）：

○ 当 $0 < \alpha < P^{-2}$ ：此时 $\alpha^{-2} > P^4$ ， $\min$ 取 P^4 。这段长度 P^{-2} ，贡献 $\leq P^4 \cdot P^{-2} = P^2$ 。

○ 当 $P^{-2} < \alpha < (4\lambda P)^{-1}$ ： $\min$ 取 α^{-2} 。

$$\int_{P^{-2}}^{\cdots} \alpha^{-2} d\alpha = [-\alpha^{-1}] \leq \frac{1}{P^{-2}} = P^2 \quad (\text{下限 } P^{-2} \text{ 处贡献最大}).$$

两段都 $\ll P^2$ ，所以差的积分是 $O(P^2)$ 。注意 $P^2 = o(P^3)$ ：这个误差比主项 P^3 小一个量级，正合我们意。

5 于是把 S -积分换成 I -积分。因为 S -积分 = I -积分 + $O(P^2)$ ，要证 S -积分 $\gg P^3$ ，只需证主弧上的 I -积分 $\gg P^3$ 。

6 主弧上的 I -积分 $\gg P^3$ 。引理 20.2 已证整条直线上的 I -积分 $\gg P^3$ 。我们只要说明"切掉主弧之外那部分"损失不大，剩下的主弧部分仍 $\gg P^3$ 。切掉的部分（远尾 $|\alpha| \geq (4\lambda P)^{-1}$ ）用第 1 步的 $|I(\lambda_j \alpha)| \ll |\alpha|^{-1/2}$ ，五个相乘 $\ll |\alpha|^{-5/2}$ ，再注意 $(\sin \pi \alpha / \pi \alpha)^2 \leq 1$ ：

$$\int_{|\alpha| \geq (4\lambda P)^{-1}} |I_1 \cdots I_5| d\alpha \ll \int_{|\alpha| \geq (4\lambda P)^{-1}} \alpha^{-5/2} d\alpha.$$

算这个尾积分： $\int_c^\infty \alpha^{-5/2} d\alpha = \left[\frac{\alpha^{-3/2}}{-3/2} \right]_c^\infty = \frac{2}{3} c^{-3/2}$ ，取 $c = (4\lambda P)^{-1}$ ，得 $\frac{2}{3} (4\lambda P)^{3/2} \ll P^{3/2}$ 。所以远尾贡献 $\ll P^{3/2} = o(P^3)$ 。

7 相减得结论。主弧上的 I -积分 = 整条直线 I -积分 - 远尾 $\gg P^3 - O(P^{3/2}) \gg P^3$ 。再由第 4-5 步， S -积分 $\gg P^3$ 。证毕。

十、转入核心：估计 S 在主弧外的贡献

现在我们来问题的核心，即对

$$\int_{|\alpha| \geq (4\lambda P)^{-1}} |S(\lambda_1 \alpha) \cdots S(\lambda_5 \alpha)| \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha$$

的估计。

为什么这是核心、难在哪里

到这里，主弧 ($|\alpha| < (4\lambda P)^{-1}$) 已经稳稳给出主项 $\gg P^3$ 。要让 (20.2) 左端 $\gg P^3$ 成立，剩下只需证明主弧之外的 S -积分相对很小，即 $o(P^3)$ ，不至于把主项吃掉。难点在于： I 在主弧外很小（有 $|\alpha|^{-1/2}$ 衰减），但 S 是求和，它在某些"共振"的 α 附

近 (α 接近有理数 a/q) 会变得相当大, 没有那么乖。要控制 S 就需要数论里的**指数和估计**: Weyl 不等式 (引理 20.7)、四次均值 (引理 20.5), 以及最后那张王牌——无理性 (引理 20.8)。接下来三段 (远尾、化简、中段) 就是逐一攻克。

十一、一个均方工具：四次积分 (引理 20.5)

引理 20.5. 对任意 $\varepsilon > 0$, 我们有

$$\int_0^1 |S(\alpha)|^4 d\alpha \ll P^{2+\varepsilon}.$$

证. 由 $S(\alpha)$ 的定义, 该积分等于 $x_1^2 + x_2^2 = y_1^2 + y_2^2$ 在 1 到 P (含) 之间的整数解的个数。其中 $x_2 = y_2$ 的解的个数为 P^2 。在其他解中, x_2 和 y_2 的值决定了 $x_1 - y_1$ 和 $x_1 + y_1$ 的值, 它们是 $x_2^2 - y_2^2$ 的因子, 至多有 P^ε 种可能。由此得结论。 ■

为什么"四次积分=解数", 又为什么 $\ll P^{2+\varepsilon}$

1 $|S(\alpha)|^4$ 展开成四重和。注意 $|S|^4 = S \cdot S \cdot \bar{S} \cdot \bar{S} = S(\alpha)^2 \overline{S(\alpha)}^2$, 其中 $\bar{S}$ 是共轭, $\overline{e(\alpha x^2)} = e(-\alpha x^2)$ 。于是

$$|S(\alpha)|^4 = \sum_{x_1, x_2, y_1, y_2=1}^P e(\alpha(x_1^2 + x_2^2 - y_1^2 - y_2^2)).$$

2 正交关系把积分变成"数解"。关键事实: 对整数 n ,

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

($n = 0$ 时被积函数恒为 1, 积分为 1; $n \neq 0$ 时它在一个周期上积分为 0。)把这个套到上面四重和的每一项: 只有当指数里的整数 $x_1^2 + x_2^2 - y_1^2 - y_2^2 = 0$ 时该项积分为 1, 否则为 0。所以

$$\int_0^1 |S(\alpha)|^4 d\alpha = \#\{(x_1, x_2, y_1, y_2) : 1 \leq x_i, y_i \leq P, x_1^2 + x_2^2 = y_1^2 + y_2^2\}.$$

这就是"积分 = 方程 $x_1^2 + x_2^2 = y_1^2 + y_2^2$ 的解数"。这是圆法里反复用的"积分 $\leftrightarrow$ 计数"对应。

- 3 第一类解： $x_2 = y_2$ 。此时方程变成 $x_1^2 = y_1^2$ ，在正整数范围内即 $x_1 = y_1$ 。 $x_1 (= y_1)$ 有 P 种取法， $x_2 (= y_2)$ 也有 P 种取法，共 $P \times P = P^2$ 组。

- 4 第二类解： $x_2 \neq y_2$ 。把方程改写成 $x_1^2 - y_1^2 = y_2^2 - x_2^2$ ，即

$$(x_1 - y_1)(x_1 + y_1) = y_2^2 - x_2^2.$$

固定 x_2, y_2 (且 $x_2 \neq y_2$)，右边是一个确定的非零整数 $N = y_2^2 - x_2^2$ ，且 $|N| < P^2$ 。现在 $x_1 - y_1$ 与 $x_1 + y_1$ 是 N 的一对因子 (它们相乘等于 N)。一个整数 N 的因子个数记作 $d(N)$ ，有经典的除数界： $d(N) \ll N^\varepsilon \ll P^\varepsilon$ (对任意 $\varepsilon > 0$ ，除数个数增长比任何幂次都慢)。所以 $x_1 - y_1$ 至多有 P^ε 种取值；一旦定了 $x_1 - y_1$ 和 $x_1 + y_1$ (它们的乘积固定为 N ，二者也就一一对应)，就能解出 $x_1 = \frac{1}{2}[(x_1 + y_1) + (x_1 - y_1)]$ 与 y_1 。

- 5 计数第二类。 x_2, y_2 的取法有 $\leq P^2$ 种 (各 $\leq P$)，对每一种， (x_1, y_1) 至多 P^ε 种。所以第二类解 $\ll P^2 \cdot P^\varepsilon = P^{2+\varepsilon}$ 。

- 6 合计。两类相加 $P^2 + P^{2+\varepsilon} \ll P^{2+\varepsilon}$ 。证毕。

这条引理的用处：它是" $|S|$ 的四次平均不会太大"的定量保证。后面只要哪里出现四个 S 相乘的积分，就能用它+Hölder 不等式压住。

十二、远尾可忽略 (引理 20.6)

引理 20.6. 对任意固定的 $\delta > 0$ ，我们有

$$\int_{|\alpha| \geq P^\delta} |S(\lambda_1 \alpha) \cdots S(\lambda_5 \alpha)| \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha \ll P^{3-\delta/2}.$$

证. 鉴于平凡估计 $|S(\lambda_5 \alpha)| \leq P$, 只需证明

$$\int_{|\alpha| \geq P^\delta} |S(\lambda_1 \alpha) \cdots S(\lambda_4 \alpha)| \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha \ll P^{2-\delta/2}.$$

由 Hölder 不等式, 只需证明

$$\int_{|\alpha| \geq P^\delta} |S(\lambda_j \alpha)|^4 \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2 d\alpha \ll P^{2-\delta/2},$$

因而只需证明

$$\int_{|\alpha| \geq P^\delta} |S(\alpha)|^4 \frac{d\alpha}{\alpha^2} \ll P^{2-\delta/2}.$$

由于 $S(\alpha)$ 以 1 为周期, 引理 20.5 蕴含

$$\int_m^{m+1} |S(\alpha)|^4 d\alpha \ll P^{2+\varepsilon}.$$

因此, 若 $M = [P^\delta]$, 所论积分至多为

$$\begin{aligned} \sum_{m=M}^{\infty} \int_m^{m+1} |S(\alpha)|^4 \frac{d\alpha}{\alpha^2} &\ll P^{2+\varepsilon} \sum_{m=M}^{\infty} \int_m^{m+1} \frac{d\alpha}{\alpha^2} \\ &\ll P^{2+\varepsilon} M^{-1} \\ &\ll P^{2-\delta+\varepsilon} \\ &\ll P^{2-\delta/2}. \end{aligned}$$

由此得结论。 ■

先认识两个新角色

δ (小正数, 读"德尔塔")

一个我们自己挑的、固定的小正数, 用来标定"远尾从哪里开始" ($|\alpha| \geq P^\delta$)。后面在引理 20.8 还会要求它足够小。

$[P^\delta]$ (取整符号)

$[x]$ 表示不超过 x 的最大整数 (高斯取整)。 $M = [P^\delta]$ 就是把 P^δ 向下取整, 得到一个整数起点。

Hölder 不等式

柯西-施瓦茨不等式的推广。本处用到的形式是: 对四个函数之积的积分,

$$\int |f_1 f_2 f_3 f_4| \leq \left(\int |f_1|^4 \right)^{\frac{1}{4}} \left(\int |f_2|^4 \right)^{\frac{1}{4}} \left(\int |f_3|^4 \right)^{\frac{1}{4}} \left(\int |f_4|^4 \right)^{\frac{1}{4}}.$$

直觉: 几个东西乘起来的积分, 被它们各自"四次方平均"的几何平均控制住。这让我们能把"四个不同的 S 相乘"归约成"同一个 S 的四次方"来估计。

逐步补全引理 20.6

- 1 先扔掉一个因子。用最粗的"平凡估计" $|S(\lambda_5 \alpha)| \leq P$ (求和共 P 项, 每项模长 1, 故和 $\leq P$)。把它提到积分外, 于是只需证剩下四个因子的积分 $\ll P^{2-\delta/2}$ (这样乘上 P 才得 $P^{3-\delta/2}$)。

- 2 用 Hölder 把四个不同 S 统一成四次方。把权重 $w(\alpha) = (\sin \pi \alpha / \pi \alpha)^2$ 平均分到四份 $w = w^{1/4} \cdot w^{1/4} \cdot w^{1/4} \cdot w^{1/4}$, 对 $f_j = S(\lambda_j \alpha) w^{1/4}$ 用 Hölder:

$$\int |S(\lambda_1 \alpha) \cdots S(\lambda_4 \alpha)| w d\alpha \leq \prod_{j=1}^4 \left(\int |S(\lambda_j \alpha)|^4 w d\alpha \right)^{1/4}.$$

所以只要每个 $\int |S(\lambda_j \alpha)|^4 w d\alpha \ll P^{2-\delta/2}$, 几何平均也 $\ll P^{2-\delta/2}$ 。这就把"四个不同 λ_j "的麻烦, 化归为"单个 S 的四次方"。

- 3 换元去掉 λ_j 、并用权重的衰减。在 $\int |S(\lambda_j \alpha)|^4 w d\alpha$ 里令 $\beta = \lambda_j \alpha$ (线性换元, 只改变常数因子和积分区间的常数倍, 不影响量级)。又

$$(\sin \pi \alpha / \pi \alpha)^2 \leq \frac{1}{(\pi \alpha)^2} \ll \frac{1}{\alpha^2} \quad (\text{因 } |\sin| \leq 1). \text{ 于是只需证}$$

$$\int_{|\alpha| \geq P^\delta} |S(\alpha)|^4 \frac{d\alpha}{\alpha^2} \ll P^{2-\delta/2}.$$

- 4 把无穷积分切成整数段、用周期性。 $S(\alpha)$ 以 1 为周期 (因为 $e(\alpha x^2)$ 中 x^2 是整数), 所以在任何长度为 1 的区间上 $|S|^4$ 的积分都一样。由引理 20.5,

$$\int_m^{m+1} |S(\alpha)|^4 d\alpha = \int_0^1 |S(\alpha)|^4 d\alpha \ll P^{2+\varepsilon}.$$

- 5 逐段放大并求和。把远尾 $\int_{\alpha \geq P^\delta}$ 拆成 $\sum_{m=M}^{\infty} \int_m^{m+1}$ ($M = [P^\delta]$)。在 $[m, m+1]$ 上 $\alpha^{-2} \leq m^{-2}$, 提到积分外:

$$\sum_{m=M}^{\infty} \int_m^{m+1} |S|^4 \frac{d\alpha}{\alpha^2} \leq \sum_{m=M}^{\infty} m^{-2} \int_m^{m+1} |S|^4 d\alpha \ll P^{2+\varepsilon} \sum_{m=M}^{\infty} m^{-2}.$$

- 6 估计 $\sum_{m \geq M} m^{-2}$. 这是收敛级数的尾部,

$$\sum_{m=M}^{\infty} m^{-2} \ll \int_{M-1}^{\infty} t^{-2} dt = \frac{1}{M-1} \ll M^{-1}. \text{ 而 } M = [P^\delta] \asymp P^\delta, \text{ 故 } M^{-1} \ll P^{-\delta}. \text{ 代回得}$$

$$\ll P^{2+\varepsilon} \cdot P^{-\delta} = P^{2-\delta+\varepsilon}.$$

- 7 吸收 ε 。因为 ε 可以取得任意小 (特别地取 $\varepsilon \leq \delta/2$), $P^{2-\delta+\varepsilon} \ll P^{2-\delta/2}$ 。证毕 (对称地 $\alpha \leq -P^\delta$ 一侧同理)。

结论 $P^{3-\delta/2} = o(P^3)$ (指数 $3 - \delta/2 < 3$), 所以远尾对 (20.2) 的贡献可忽略。三段里的第三段就此搞定。

十三、把问题逼到中段 (化归到 (20.4))

由 (20.2) 以及引理 20.4 和引理 20.6, 只需证明

$$\int_{(4\lambda P)^{-1} < |\alpha| < P^\delta} |S(\lambda_1 \alpha) \cdots S(\lambda_5 \alpha)| d\alpha = o(P^3).$$

因为这样一来 (20.2) 的右端就为 $\gg P^3$, 这正是我们想证明的。如前所述, 我们只能对某些受限的 P 值证明这一点。

为什么"只剩中段要管"

- 1 把 (20.2) 左端的全直线积分 $\int_{-\infty}^{\infty}$ 按三段拆开: 主弧 $|\alpha| < (4\lambda P)^{-1}$ + 中段 $(4\lambda P)^{-1} < |\alpha| < P^{\delta}$ + 远尾 $|\alpha| \geq P^{\delta}$ 。
- 2 主弧: 引理 20.4 已证 $\gg P^3$ (这是主项)。
- 3 远尾: 引理 20.6 已证 $\ll P^{3-\delta/2} = o(P^3)$ (可忽略)。
- 4 所以只要再证中段的 S -积分也是 $o(P^3)$, 三段合起来 (20.2) 左端 = 主项 + $o(P^3) \gg P^3$, 从而 (20.2) 右端 (解的计数) $\gg P^3$, 定理成立。注意这里连权重 $(\sin \pi \alpha / \pi \alpha)^2 \leq 1$ 都直接放掉了 (用 ≤ 1 上界), 因为中段更看重 S 本身的大小。

由引理 20.5 和 Hölder 不等式可得

$$\int_{(4\lambda P)^{-1} < |\alpha| < P^{\delta}} |S(\lambda_{i_1} \alpha) \cdots S(\lambda_{i_4} \alpha)| d\alpha \ll P^{2+\delta+\varepsilon}$$

这对任意四个不同的下标 $i_1, \dots, i_4$ 都成立。把它分别用于下标 2, 3, 4, 5 和 1, 3, 4, 5, 我们看到只需对积分范围内的每个 α 都有

$$\min(|S(\lambda_1 \alpha)|, |S(\lambda_2 \alpha)|) \ll P^{1-2\delta} \quad (20.4)$$

即可。为此我们必须用到 λ_1/λ_2 的无理性。

把"四因子积分"和 (20.4) 的逻辑讲透

- 1 四因子积分 $\ll P^{2+\delta+\varepsilon}$ 的由来。同样用 Hölder:

$$\int_{\text{中段}} |S(\lambda_{i_1} \alpha) \cdots S(\lambda_{i_4} \alpha)| d\alpha \leq \prod_{k=1}^4 \left(\int_{\text{中段}} |S(\lambda_{i_k} \alpha)|^4 d\alpha \right)^{1/4}.$$

对每个因子换元 $\beta = \lambda_{i_k} \alpha$ 。中段 $|\alpha| < P^{\delta}$ 映成 $|\beta| < |\lambda_{i_k}| P^{\delta} \ll P^{\delta}$ 的区间, 长度 $\ll P^{\delta}$ 。把这区间按长度 1 切片, 每片 $\int |S|^4 \ll P^{2+\varepsilon}$ (引理 20.5+周期性), 共约 P^{δ} 片, 故

$$\int_{\text{中段}} |S(\lambda_{i_k} \alpha)|^4 d\alpha \ll P^\delta \cdot P^{2+\varepsilon} = P^{2+\delta+\varepsilon}.$$

四个因子的几何平均仍是 $P^{2+\delta+\varepsilon}$ 。

2 关键的"取 min"技巧。我们要估计五因子积分 $\int_{\text{中段}} |S(\lambda_1 \alpha) \cdots S(\lambda_5 \alpha)| d\alpha$ 。在每个 α 处，五个因子里 $|S(\lambda_1 \alpha)|$ 与 $|S(\lambda_2 \alpha)|$ 较小的那个记为 $\min(|S(\lambda_1 \alpha)|, |S(\lambda_2 \alpha)|)$ 。把这个最小的因子单独拎出来，剩下的四个因子恰好构成"四个不同下标的积"：

○ 若在该 α 处 $|S(\lambda_1 \alpha)| \leq |S(\lambda_2 \alpha)|$ ，则

$|S(\lambda_1) \cdots S(\lambda_5)| = \min \cdot |S(\lambda_2) S(\lambda_3) S(\lambda_4) S(\lambda_5)|$ ，剩四个下标是 2, 3, 4, 5。

○ 若反之 $|S(\lambda_2 \alpha)| < |S(\lambda_1 \alpha)|$ ，则剩四个下标是 1, 3, 4, 5。

两种情形都把"五因子积"写成了 $\min \times (\text{某四因子积})$ 。

3 若能证 $\min \ll P^{1-2\delta}$ (即 (20.4))，则中段积分 $= o(P^3)$ 。因为

$$\int_{\text{中段}} |S(\lambda_1) \cdots S(\lambda_5)| d\alpha \leq P^{1-2\delta} \int_{\text{中段}} (\text{四因子积}) d\alpha \ll P^{1-2\delta} \cdot P^{2+\delta+\varepsilon} = P^{3-\delta+\varepsilon}$$

(中段按 min 是哪一个分成两块，每块用对应的四因子积，都得同样的界。) 取 $\varepsilon < \delta$ ，则 $P^{3-\delta+\varepsilon} = o(P^3)$ 。这正是我们要的中段估计！

4 所以全部归结为证 (20.4)：在中段每个 α 处，

$\min(|S(\lambda_1 \alpha)|, |S(\lambda_2 \alpha)|) \ll P^{1-2\delta}$ 。直觉： $S(\lambda_1 \alpha)$ 与 $S(\lambda_2 \alpha)$ 不可能同时都很大。这正是要动用 λ_1/λ_2 无理性的地方——下面引理 20.7、20.8 专办此事。

十四、控制单个 S 的大小：Weyl 不等式 (引理 20.7)

引理 20.7. 设 $(a, q) = 1$ 且 $|\alpha - a/q| < q^{-2}$ 。那么

$$|S(\alpha)| \ll P^{1+\varepsilon} \left\{ P^{-1/2} + q^{-1/2} + \left(\frac{P^2}{q} \right)^{-1/2} \right\}.$$

证. 这是引理 3.1 (Weyl 不等式), 其中 $k = 2$, 因而也有 $K = 2$. ■

Weyl 不等式在说什么、为什么这样长相

背景与动机。 $S(\alpha) = \sum_{x \leq P} e(\alpha x^2)$ 是一个"指数和"。如果 α 让相位 αx^2 在转圈时分布很均匀, 那么这些单位向量 $e(\alpha x^2)$ 会相互抵消, 和远小于 P (平凡上界)。Hermann Weyl 在研究"等分布"时发明了一套方法来量化这种抵消, 结论就是 Weyl 不等式。它把 $|S(\alpha)|$ 的大小和" α 离有理数 a/q 有多近、那个分母 q 有多大"挂钩。

条件 $(a, q) = 1, |\alpha - a/q| < q^{-2}$

说 α 被某个**最简分数** a/q 很好地逼近 (误差小于 $1/q^2$)。Dirichlet 逼近定理保证这样的 a/q 总是存在。

三项 $P^{-1/2} + q^{-1/2} + (P^2/q)^{-1/2}$

把 $P^{1+\varepsilon}$ 乘进去看: 第一项给 $P^{1/2+\varepsilon}$; 第二项给 $P^{1+\varepsilon} q^{-1/2}$ (q 越大越小); 第三项 $(P^2/q)^{-1/2} = q^{1/2}/P$, 给 $P^\varepsilon q^{1/2}$ (q 越小越小)。所以当 q 不太小也不太大 (在 $1 \ll q \ll P^2$ 的"中间地带") 时, $|S(\alpha)|$ 明显小于平凡的 P 。这正是我们需要的"抵消"。

" $k = 2$ 故 $K = 2$ "

一般 Weyl 不等式 (引理 3.1) 处理 k 次多项式相位 $\sum e(\alpha x^k)$, 参数 $K = 2^{k-1}$ 。这里我们的相位是 x^2 ($k = 2$), 代入得 $K = 2^{2-1} = 2$, 于是出现上面那种带平方根 ($-1/2$ 次幂) 的具体形状。

十五、把 P 取成连分数分母的平方

我们选取 λ_1/λ_2 的连分数的任一渐近分数 a_0/q_0 , 于是有

$$\left| \frac{\lambda_1}{\lambda_2} - \frac{a_0}{q_0} \right| < \frac{1}{q_0^2}. \quad (20.5)$$

我们取 $P = q_0^2$ ；这将 P 限制为一个无穷序列的取值。

连分数与渐近分数（从零讲）

什么是连分数？任何实数 ξ 都能写成"连续的分数的嵌套"：

$$\xi = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}},$$

其中 a_0 是整数、 $a_1, a_2, \dots$ 是正整数。把这个无穷嵌套在第 n 层"砍断"，得到的有理数 p_n/q_n 叫第 n 个渐近分数（convergent，渐近分数/收敛子）。

渐近分数的黄金性质。渐近分数 p_n/q_n 是"用分母不超过 q_n 的分数逼近 ξ 的最佳选择"，并且满足

$$\left| \xi - \frac{p_n}{q_n} \right| < \frac{1}{q_n^2}.$$

这正是 (20.5)：取 $\xi = \lambda_1/\lambda_2$ ，它的某个渐近分数 a_0/q_0 满足 $|\lambda_1/\lambda_2 - a_0/q_0| < 1/q_0^2$ 。

为什么需要"无理"？如果 λ_1/λ_2 是无理数，它的连分数无限不终止，于是有无穷多个渐近分数，分母 q_0 可以要多大有多大。这保证了 $P = q_0^2$ 能取到任意大的值——这正是定理只需"一串 P "就够的来源。倘若 λ_1/λ_2 有理，连分数会终止， q_0 不能无限增大，这套就垮了。这就是无理性假设头一次显出威力。

为什么偏偏取 $P = q_0^2$ （取舍的理由）

这是本证明最精妙的参数选择。我们需要 P 与 q_0 之间有一个精确的指数关系，好让后面引理 20.8 里两条估计——"逼近误差 $\ll P^{-3/2} = q_0^{-3}$ "与"连分数误差 $< q_0^{-2}$ "——能放在同一把尺子上比较，从而逼出矛盾。取 $P = q_0^2$ 时： $P^{-3/2} = q_0^{-3}$ ，比 q_0^{-2} 还小一个 q_0 因子，这点"富余"正好用来制造矛盾。若取别的关系（如 $P = q_0$ 或 $P = q_0^3$ ），两条误差的量级配不上，矛盾就凑不出来。换言之， $P = q_0^2$ 是为了让两套逼近"刚好打架"而量身定制的。代价是 P 不再能取所有整数，只能取这串平方值——但对"存在解"足够了。

十六、王牌登场：用无理性证 (20.4) (引理 20.8)

引理 20.8. 当 P 受上述限制时, 估计 (20.4) 对范围 $(4\lambda P)^{-1} < \alpha < P^\delta$ 内的每个 α 都成立。

证. 存在 $\lambda_1\alpha$ 的一个有理逼近 a_1/q_1 , 使得

$$(a_1, q_1) = 1, \quad 1 \leq q_1 \leq P^{3/2}, \quad \left| \lambda_1\alpha - \frac{a_1}{q_1} \right| < \frac{1}{q_1 P^{3/2}}. \quad (20.6)$$

我们注意到 $a_1 \neq 0$, 因为若 $a_1 = 0$, 则 $|\lambda_1\alpha| < P^{-3/2}$, 与假设矛盾。类似地, 存在 $\lambda_2\alpha$ 的一个有理逼近 a_2/q_2 , 使得

$$(a_2, q_2) = 1, \quad 1 \leq q_2 \leq P^{3/2}, \quad \left| \lambda_2\alpha - \frac{a_2}{q_2} \right| < \frac{1}{q_2 P^{3/2}}, \quad (20.7)$$

并且同样有 $a_2 \neq 0$ 。

若 $q_1 > P^{5\delta}$, 我们可以对 $S(\lambda_1\alpha)$ 应用引理 20.7, 以 a_1, q_1 代替 a, q , 这给出

$$|S(\lambda_1\alpha)| \ll P^{1+\varepsilon-5\delta/2} \ll P^{1-2\delta}.$$

类似地, 若 $q_2 > P^{5\delta}$, 我们对 $|S(\lambda_2\alpha)|$ 得到相应的结果。在这两种情形的任一种下, (20.4) 都满足。因此我们可以假定

$$q_1 \leq P^{5\delta}, \quad q_2 \leq P^{5\delta}. \quad (20.8)$$

现在我们可以从 (20.6) 和 (20.7) 推出 λ_1/λ_2 被 a_1q_2/a_2q_1 很好地逼近。我们注意到, 由于 a_1/q_1 是 $\lambda_1\alpha$ 的一个逼近, 且 $|\alpha| < P^\delta$, 我们有 $|a_1| \ll P^{6\delta}$, 类似地 $|a_2| \ll P^{6\delta}$ 。因此

$$\frac{\lambda_1}{\lambda_2} = \frac{\lambda_1\alpha}{\lambda_2\alpha} = \frac{\frac{a_1}{q_1}(1 + O(P^{-3/2}))}{\frac{a_2}{q_2}(1 + O(P^{-3/2}))} = \frac{a_1q_2}{a_2q_1} \left(1 + O(P^{-3/2})\right),$$

又由于 $|a_1q_2/a_2q_1|$ 有上界, 这蕴含

$$\left| \frac{\lambda_1}{\lambda_2} - \frac{a_1q_2}{a_2q_1} \right| \ll P^{-3/2}.$$

我们有 $1 \leq |a_2 q_1| \ll P^{11\delta}$ 。

将最后这个结果与 (20.5) 比较, 若 δ 足够小 (且 q_0 足够大), 就得到一个矛盾。因为我们有

$$\left| \frac{\lambda_1}{\lambda_2} - \frac{a_1 q_2}{a_2 q_1} \right| \ll P^{-3/2} + q_0^{-2} \ll q_0^{-2},$$

这里 $P = q_0^2$, 而左端却是

$$\geq \frac{1}{q_0 |a_2| q_1} \gg \frac{1}{q_0 P^{11\delta}} \gg q_0^{-1-6\delta}.$$

这就完成了引理 20.8 的证明; 而由我们前面的论述, 引理 20.8 完成了定理 20.1 的证明。 ■

先补一个工具: Dirichlet 逼近定理

(20.6)、(20.7) 里"存在一个有理逼近"用的是 **Dirichlet 逼近定理**: 对任意实数 θ 和任意正整数 N , 存在既约分数 a/q , 使

$$1 \leq q \leq N, \quad \left| \theta - \frac{a}{q} \right| < \frac{1}{qN}.$$

(证明用抽屉原理: 把 $0, \{\theta\}, \{2\theta\}, \dots, \{N\theta\}$ 共 $N+1$ 个小数部分塞进 N 个长 $1/N$ 的小格, 必有两个落同格, 相减即得。) 这里取 $\theta = \lambda_1 \alpha$ 、 $N = P^{3/2}$, 就得到 (20.6); 取 $\theta = \lambda_2 \alpha$ 得 (20.7)。

把引理 20.8 的整条逻辑链拆开

1 给 $\lambda_1 \alpha, \lambda_2 \alpha$ 各找一个好逼近。由 Dirichlet 定理 ($N = P^{3/2}$) 得 (20.6)、(20.7): 既约分数 $a_1/q_1, a_2/q_2$, 分母 $\leq P^{3/2}$, 误差分别 $< \frac{1}{q_1 P^{3/2}}, \frac{1}{q_2 P^{3/2}}$ 。

2 为什么 $a_1 \neq 0$ 。若 $a_1 = 0$, 则 (20.6) 说 $|\lambda_1 \alpha| < \frac{1}{q_1 P^{3/2}} \leq P^{-3/2}$ 。但中段里 $\alpha > (4\lambda P)^{-1}$, 故 $|\lambda_1 \alpha| \geq |\lambda_1| \cdot \frac{1}{4\lambda P} \gg P^{-1}$, 远大于 $P^{-3/2}$, 矛盾。所以 $a_1 \neq 0$ (同理 $a_2 \neq 0$)。这一步用上了" α 在中段、不太靠近 0"这一区间信息。

3 分情形：只要某个分母大，就直接用 Weyl。若 $q_1 > P^{5\delta}$ ，对 $S(\lambda_1\alpha)$ 用引理 20.7（取 $a = a_1, q = q_1$ ，且 (20.6) 满足 $|\lambda_1\alpha - a_1/q_1| < \frac{1}{q_1 P^{3/2}} \leq \frac{1}{q_1^2}$ 因 $q_1 \leq P^{3/2}$ ）。看 Weyl 三项乘 $P^{1+\varepsilon}$ ：

- $P^{1+\varepsilon} \cdot P^{-1/2} = P^{1/2+\varepsilon}$ ，很小；
- $P^{1+\varepsilon} \cdot q_1^{-1/2} < P^{1+\varepsilon} \cdot P^{-5\delta/2} = P^{1+\varepsilon-5\delta/2}$ ；
- $P^{1+\varepsilon} \cdot (P^2/q_1)^{-1/2} = P^\varepsilon q_1^{1/2} \leq P^\varepsilon P^{3/4} = P^{3/4+\varepsilon}$ ，也小。

三项里主导的是 $P^{1+\varepsilon-5\delta/2}$ ，故 $|S(\lambda_1\alpha)| \ll P^{1+\varepsilon-5\delta/2} \ll P^{1-2\delta}$ （只要 $\varepsilon \leq \delta/2$ ，脚注也提醒“ δ 取小”）。这就直接给出 (20.4)。 $q_2 > P^{5\delta}$ 时同理给出 $|S(\lambda_2\alpha)|$ 小。所以只剩下两个分母都小的情形要对付：(20.8) $q_1, q_2 \leq P^{5\delta}$ 。

4 估计分子 $|a_1|, |a_2| \ll P^{6\delta}$ 。由 (20.6)， $|\frac{a_1}{q_1}| \leq |\lambda_1\alpha| + \frac{1}{q_1 P^{3/2}}$ 。中段里 $|\alpha| < P^\delta$ ，故 $|\lambda_1\alpha| \ll P^\delta$ 。于是 $|a_1| \leq q_1(|\lambda_1\alpha| + P^{-3/2}) \ll P^{5\delta} \cdot P^\delta = P^{6\delta}$ （用了 (20.8) 的 $q_1 \leq P^{5\delta}$ ）。同理 $|a_2| \ll P^{6\delta}$ 。

5 用两个逼近“做除法”逼近 λ_1/λ_2 。关键观察：公共因子 α 在比值里约掉， $\frac{\lambda_1}{\lambda_2} = \frac{\lambda_1\alpha}{\lambda_2\alpha}$ 。把 (20.6) 写成相对误差形式：

$$\lambda_1\alpha = \frac{a_1}{q_1} + \theta_1, |\theta_1| < \frac{1}{q_1 P^{3/2}} \Rightarrow \lambda_1\alpha = \frac{a_1}{q_1} \left(1 + \frac{\theta_1 q_1}{a_1}\right).$$

其中 $|\frac{\theta_1 q_1}{a_1}| < \frac{1}{P^{3/2}|a_1|} \leq P^{-3/2}$ （因 $|a_1| \geq 1$ ）。所以

$\lambda_1\alpha = \frac{a_1}{q_1}(1 + O(P^{-3/2}))$ ，同理 $\lambda_2\alpha = \frac{a_2}{q_2}(1 + O(P^{-3/2}))$ 。两式相除：

$$\frac{\lambda_1}{\lambda_2} = \frac{a_1/q_1}{a_2/q_2} \cdot \frac{1 + O(P^{-3/2})}{1 + O(P^{-3/2})} = \frac{a_1 q_2}{a_2 q_1} (1 + O(P^{-3/2})).$$

（用了 $\frac{1+O(t)}{1+O(t)} = 1 + O(t)$ ，当 $t = P^{-3/2} \rightarrow 0$ 。）

6 转成绝对误差。移项： $\frac{\lambda_1}{\lambda_2} - \frac{a_1 q_2}{a_2 q_1} = \frac{a_1 q_2}{a_2 q_1} \cdot O(P^{-3/2})$ 。而 $|\frac{a_1 q_2}{a_2 q_1}| \approx |\lambda_1/\lambda_2|$ 是有界常数，所以

$$\left| \frac{\lambda_1}{\lambda_2} - \frac{a_1 q_2}{a_2 q_1} \right| \ll P^{-3/2}.$$

这说明分数 $\frac{a_1 q_2}{a_2 q_1}$ 把 λ_1/λ_2 逼近到 $P^{-3/2} = q_0^{-3}$ 这么准。

- 7 这个新分数的分母很小。 $|a_2 q_1| \leq |a_2| q_1 \ll P^{6\delta} \cdot P^{5\delta} = P^{11\delta}$ ，且 ≥ 1 。即用一个分母只有 $P^{11\delta}$ （很小）的分数，竟把 λ_1/λ_2 逼近到了 q_0^{-3} 这么准——这“好得不真实”，下一步就把它和 a_0/q_0 一比，逼出矛盾。

- 8 制造矛盾（核心）。现在有两个分数都很接近 λ_1/λ_2 ：连分数渐近分数 $\frac{a_0}{q_0}$ （误差 $< q_0^{-2}$ ，见 (20.5)）和 $\frac{a_1 q_2}{a_2 q_1}$ （误差 $\ll P^{-3/2} = q_0^{-3}$ ）。考虑它俩的差

$$\Delta = \frac{a_0}{q_0} - \frac{a_1 q_2}{a_2 q_1}.$$

- 上界：用三角不等式经过 λ_1/λ_2 中转，

$$|\Delta| \leq \left| \frac{a_0}{q_0} - \frac{\lambda_1}{\lambda_2} \right| + \left| \frac{\lambda_1}{\lambda_2} - \frac{a_1 q_2}{a_2 q_1} \right| \ll q_0^{-2} + q_0^{-3} \ll q_0^{-2}.$$

- 下界（若 $\Delta \neq 0$ ）：两个分母分别为 q_0 与 $a_2 q_1$ 的不相等有理数，其差的绝对值 $\geq \frac{1}{q_0 \cdot |a_2 q_1|}$ （因为 $\Delta = \frac{\text{非零整数}}{q_0 a_2 q_1}$ ，分子至少为 1）。由第 7 步 $|a_2 q_1| \ll P^{11\delta} = q_0^{22\delta}$ ，故

$$|\Delta| \geq \frac{1}{q_0 |a_2 q_1|} \gg \frac{1}{q_0 \cdot q_0^{22\delta}} = q_0^{-1-22\delta}.$$

（原书此处写作 $q_0^{-1-6\delta}$ ，但由 $P = q_0^2$ 知 $P^{11\delta} = q_0^{22\delta}$ ，老实算应为 $-1 - 22\delta$ ，原书的 6δ 当是笔误；无论如何这个指数都形如 $-1 - (\text{常数})\delta$ ，下面的矛盾照样成立。）

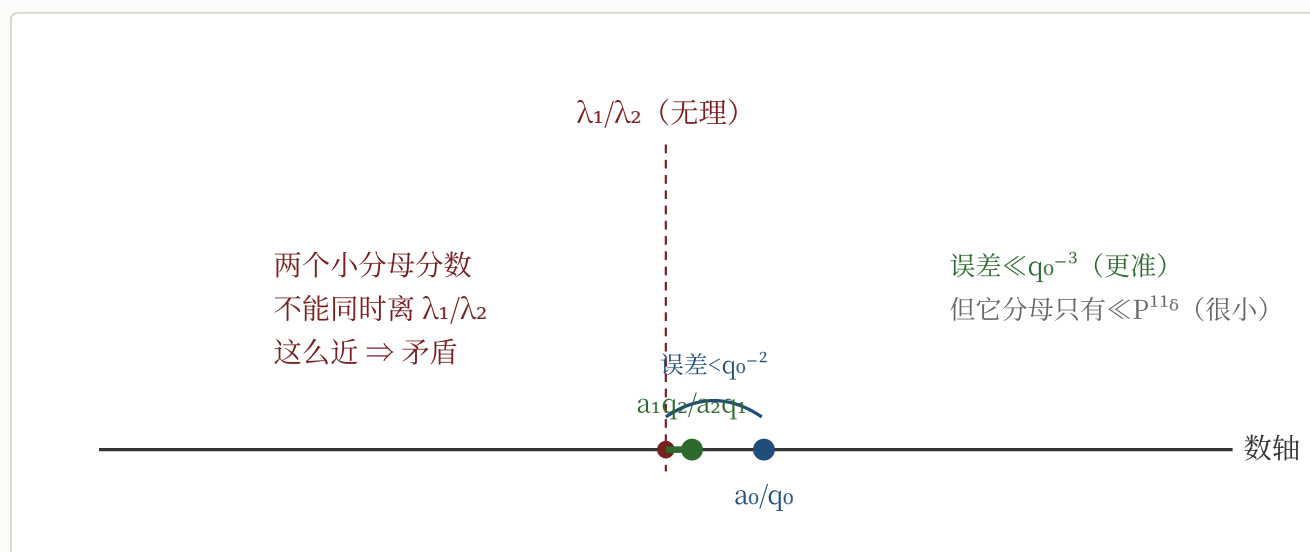
把上下界放一起： $q_0^{-1-22\delta} \leq |\Delta| \ll q_0^{-2}$ 。但只要 δ 取得足够小（使 $1 + 22\delta < 2$ ，即 $\delta < \frac{1}{22}$ ）、并让 q_0 足够大，就有 $q_0^{-1-22\delta} \gg q_0^{-2}$ ，与上界直接冲突。矛盾！

- 9 排除 $\Delta = 0$ 的可能。若 $\Delta = 0$ ，即 $\frac{a_0}{q_0} = \frac{a_1 q_2}{a_2 q_1}$ 。因 a_0/q_0 既约，必有 $q_0 \mid a_2 q_1$ ，从而 $q_0 \leq |a_2 q_1| \ll q_0^{22\delta}$ ，对 $\delta < \frac{1}{22}$ 与大 q_0 同样不可能。所以 $\Delta = 0$ 也被排除。

- 10 收网。无论 Δ 是否为 0 都导致矛盾——而这一切都是在假设 (20.8)（两个分母都小）之下推出的。所以 (20.8) 不可能成立，即必有 $q_1 > P^{5\delta}$ 或 $q_2 > P^{5\delta}$ 。由第

3 步，这恰好给出 (20.4) $\min(|S(\lambda_1\alpha)|, |S(\lambda_2\alpha)|) \ll P^{1-2\delta}$ 。引理 20.8 证毕。

回看整条线。 (20.4) 成立 $\implies$ 中段积分 $= o(P^3)$ (第十三节) $\implies$ 配合主弧主项 $\gg P^3$ (引理 20.4) 与远尾可忽略 (引理 20.6) $\implies$ (20.2) 左端 $\gg P^3 \implies$ (20.2) 右端解数 $\gg P^3 \implies$ 存在满足 (20.1) 的非零整数解 $\implies$ 定理 20.1 成立。无理性恰好、且仅仅在引理 20.8 的"制造矛盾"那一步用到——这就是为什么本章反复强调"here and here only"。



矛盾的几何图景： a_0/q_0 (分母 q_0) 与 a_1q_2/a_2q_1 (分母 $\ll P^{11\delta}$ 很小) 都贴着无理数 λ_1/λ_2 。两个分母不大的分数若都离同一无理数极近，它们彼此也必极近，但分母小的两个不等有理数差不可能太小——挤不下，于是矛盾。这就是无理性发力的瞬间。

十七、推广与边界（注记）

注记. 定理 20.1 的某些推广几乎是显然的。首先，对于不等式

$$|\lambda_1 x_1^2 + \cdots + \lambda_5 x_5^2 - \mu| < \varepsilon,$$

对任意实数 μ (假设比值 λ_i/λ_j 不全为有理数)，我们可以证明同样的结果。其次，我们可以把平方替换为 k 次幂，只要变量个数至少为 $2^k + 1$ 。在这种情形下，我们应当用 Hua 不等式 (引理 3.2) 代替上面的引理 20.7。更精确的结果已由 Davenport 和 Roth [30] 以及 Danicic [16] 证明。

两条推广分别意味着什么

1 挪动靶心 μ 。原定理让式子接近 0；推广成让 $\lambda_1 x_1^2 + \cdots + \lambda_5 x_5^2$ 接近任意实数 μ 。证明几乎照搬：引理 20.1 里把指示函数中心从 0 平移到 μ （即用 $Q - \mu$ 代替 Q ，被积函数乘上 $e(-\alpha\mu)$ ），其余完全平行。它说明：这种二次型的取值在实轴上处处稠密（能逼近任何 μ ）。

2 把平方换成 k 次幂。研究 $|\lambda_1 x_1^k + \cdots + \lambda_n x_n^k| < \varepsilon$ 。这时单个指数和是 $\sum e(\alpha x^k)$ ，控制它需要更强的工具——Hua 不等式（引理 3.2，给出 $\int_0^1 |\sum e(\alpha x^k)|^{2^k} d\alpha$ 的界），它接替本章的引理 20.7/20.5 的角色。变量个数门槛升到 $2^k + 1$ （ $k = 2$ 时正是 $2^2 + 1 = 5$ ，与本章一致）。Davenport-Roth、Danicic 后来把这些结果做得更精细。

注记. 把丢番图不等式推广到涉及具有实系数的一般形式（而非加性形式）的情形，所遇困难也许比丢番图方程的相应推广还要大。

"加性形式"与"一般形式"的区别

本章的 $\lambda_1 x_1^2 + \cdots + \lambda_5 x_5^2$ 是加性（对角）形式：各变量分开、没有交叉项，所以指数和能漂亮地因式分解成 $S(\lambda_1 \alpha) \cdots S(\lambda_5 \alpha)$ 。一旦换成一般二次型（含 $x_i x_j$ 交叉项），这种因式分解失效，圆法处理起来困难大增——甚至比相应的方程问题还棘手。这就是下面 (20.9) 系列结果为何需要"非常复杂的工作"。

迄今所得的全部结果都依赖于关于丢番图方程的结果，而且通常需要其更精确的形式，即对解的大小有一个估计。

这句话的方法论含义

不等式的可解性证明往往"寄生"在方程结果之上：要么直接引用方程的解的存在性，要么需要更精细的"解可以取得多小/多大"的定量信息（如本章里对 P 范围、对逼近误差

的精确控制)。光知道"方程有解"不够，还要知道"解的尺寸"，才能在不等式的逼近论证里用上。

自然首先出现的问题是建立下式的可解性

$$|Q(x_1, \dots, x_n)| < \varepsilon \quad (20.9)$$

对任意 $\varepsilon > 0$ ，其中 Q 是任意不定二次型。通过一些非常复杂的工作，已经证明它当 $n \geq 21$ 时成立，该工作是 Birch、Davenport 和 Ridout [70] 的联合成果。由 Oppenheim [64, 65] 的一个结果可知，如果 Q 不与整系数形式成比例，那么不等式

$$|Q(x_1, \dots, x_n) - \mu| < \varepsilon,$$

对任意实数 μ 都是可解的。于是，任意实不定二次型在 21 个或更多个变量中的取值，要么是离散的（如果该形式与整形式成比例），要么处处稠密。

一般不定二次型的结局

$Q(x_1, \dots, x_n)$ (一般不定二次型)

含交叉项的二次型，如 $x_1^2 - 3x_1x_2 + 2x_2^2 - \dots$ ，系数为任意实数。"不定"即取值有正有负。

$n \geq 21$ (Birch–Davenport–Ridout)

对一般不定二次型，目前能证明 (20.9) 可解的变量门槛是 21——远高于加性情形的 5，正因为没有了因式分解的便利。

"离散 vs 处处稠密"的二分

这是个漂亮的结构性结论：一个 ≥ 21 元实不定二次型的取值集合，只有两种可能——① 若 Q 与某个整系数形式成比例（即 $Q = c \cdot (\text{整形式})$ ），取值落在 $c\mathbb{Z}$ 这样的离散集合上；② 否则（不成比例），取值在实轴上处处稠密，能逼近任何实数 μ （Oppenheim 的结果）。没有中间状态。这正是 Oppenheim 当年猜想的精神被完整实现的样子。

对任意实三次型，(20.9) 的类似结果已由 Pitman [66] 证明，但所需的变量个数相当大。对于五次的形式，证明任何类似结果似乎存在原则性的困难。

更高次的前沿与困难

三次型：Pitman 证明了类似结论，但变量个数要求很大（次数越高，需要的变量越多，圆法的"次弧"越难控制）。**五次型：**存在"原则性困难"——这是奇数次（尤其 5 次）里圆法和已知方法的固有障碍：缺少类似 Hua 不等式那样足够强的均值估计，使得现有手段无法把次弧贡献压下去。这告诉你，本章这套优雅方法虽强，却不是万能的；数论里仍有大片未解的疆域。

全章一句话总览

用引理 20.1 把" $|Q| < 1$ 的整数解计数"写成积分恒等式 (20.2)；用连续模型 (20.3) + 引理 20.2 量出主项 P^3 ；把 α 轴分三段：主弧上 $S \approx I$ 出主项 (20.3、20.4)，远尾靠四次均值 + 权重衰减可忽略 (20.5、20.6)，中段则靠 Weyl 不等式 (20.7) + 把 P 取成连分数分母平方 q_0^2 + 无理性逼出矛盾 (20.5、20.8) 来压成 $o(P^3)$ 。三段一合，解数 $\gg P^3$ ，定理得证。无理性这张牌，全程只在引理 20.8 制造矛盾时打出一次。

返回 [全书目录](#)