

DAVENPORT · 圆法 · 高中详解版

Weyl 不等式与华罗庚不等式

Weyl's inequality and Hua's inequality

本页是对本章译文的**逐段高中详解**：先按原文顺序复述每一段在讲什么，再把其中所有“显然”“容易看出”的跳步补成一步一句的完整推导，每个超出高中范围的符号都从零讲起。黑字是原文的意思，彩色框（目标 / 符号 / 分步推演 / 图）是为高中生补充的讲解。

本章要解决什么

这一章是整本《圆法》的**技术心脏**。圆法要研究“一个数能否写成若干个 k 次方之和”（华林问题），核心是去估计一种叫做**指数和**的东西

$$T(\alpha) = \sum_{x=1}^P e(\alpha x^k) \quad (e(\theta) := e^{2\pi i \theta}).$$

这一章给出两件最重要的“武器”：

- Weyl 不等式**（引理 3.1）：当 α 不太接近“分母很小的分数”时，上面这种和会比它的平凡上界 P **小很多**。这把“和很大”的坏情形局限在很窄的范围里。
- 华罗庚不等式**（引理 3.2）： $|T(\alpha)|$ 的高次方在 $[0, 1]$ 上的**平均**（积分）很小。这是后面建立渐近公式时控制“次弧”的关键。

读完你应当能：理解 $e(\alpha)$ 、积分 $\int_0^1$ 、求和号、 $\ll/O$ 、 $\|\lambda\|$ 、除数函数等记号；并能**独立复述**这两个不等式的每一步证明——为什么对和取平方、为什么

用差分降次、为什么用 Cauchy 不等式、为什么把项按公因子合并、为什么积分恰好是“方程解的个数”。

第 0 节 先把工具备齐：高中生需要的全部预备知识

原文默认读者已经熟悉下面这些记号，直接拿来用。但高中课本里没有它们，所以我们先单独花一节把它们讲透。看不懂证明时，多半是卡在这一节的某个符号上，回来查即可。

符号 $e(\theta)$ ：复指数

定义 $e(\theta) := e^{2\pi i \theta}$ ，读作“e 括号 theta”。这里 i 是虚数单位 ($i^2 = -1$)， e 是自然常数。由欧拉公式

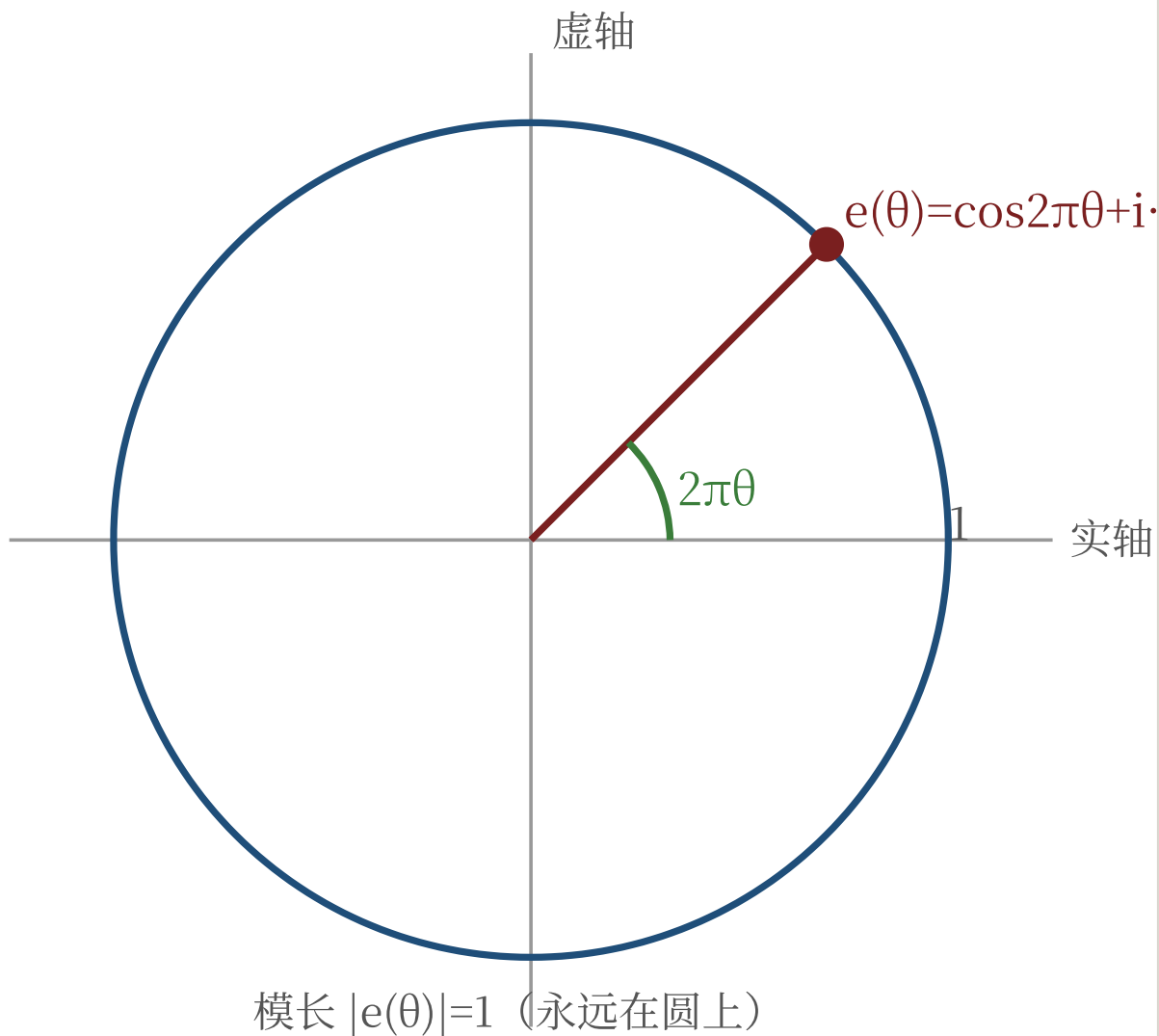
$$e^{i\varphi} = \cos \varphi + i \sin \varphi$$

可得

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

它是复平面单位圆上的一个点，对应的角度是 $2\pi\theta$ （即把整整一圈 2π 分成 θ 份）。两个关键性质：

- **模长恒为 1**： $|e(\theta)| = \sqrt{\cos^2 + \sin^2} = 1$ 。它永远在单位圆上跑，绝不会变长或变短。
- **把加法变成乘法**： $e(\theta_1) e(\theta_2) = e(\theta_1 + \theta_2)$ ，因为指数相加。于是 $e(-\theta) = 1/e(\theta) = \overline{e(\theta)}$ （共轭）。
- **以 1 为周期**： $e(\theta + 1) = e(\theta)$ ，因为转一整圈 (2π) 回到原地。所以 $e(\theta)$ 只关心 θ 的小数部分。



$e(\theta)$ 是单位圆上转了角 $2\pi\theta$ 的点。 θ 每增加 1 就转满一圈回到原处，所以 e 只看 θ 的小数部分。

符号 $\sum$ 求和、 $|z|$ 复数模长

$\sum_{x=1}^P g(x) = g(1) + g(2) + \cdots + g(P)$ ，把方括号内变量从下界取到上界逐项相加。本章里被加的常是复数 $e(\cdots)$ ，所以和也是复数。复数 $z = a + bi$ 的模长

$|z| = \sqrt{a^2 + b^2}$ ，表示它到原点的距离； $|z|^2 = z\bar{z}$ （自己乘以自己的共轭）。我们要估计的就是这种复数和的模长 $\left| \sum e(\cdots) \right|$ 。

符号 $\Re z$ 取实部

$\Re z$ 表示复数 $z = a + bi$ 的实部 a 。一个有用事实：对任意复数 z ， $z + \bar{z} = 2\Re z$ 。本章把“一对 (x_1, x_2) 和它交换后的 (x_2, x_1) ”合并时会用到它——这两项互为共轭，加起来正好是 $2\Re$ 。

符号 $\int_0^1$ 定积分与“正交关系”

$\int_0^1 g(\alpha) d\alpha$ 是函数 g 在区间 $[0, 1]$ 上的定积分，几何上是曲线下的“带符号面积”。本章只需要一个具体结果，叫正交关系：对整数 n ，

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

为什么？当 $n = 0$ 时被积函数恒等于 $e^0 = 1$ ，积分就是区间长度 1。当 $n \neq 0$ 时，

$$\int_0^1 e^{2\pi i n \alpha} d\alpha = \left[\frac{e^{2\pi i n \alpha}}{2\pi i n} \right]_0^1 = \frac{e^{2\pi i n} - e^0}{2\pi i n} = \frac{1 - 1}{2\pi i n} = 0,$$

因为 $e^{2\pi i n} = 1$ (n 是整数，转了整数圈)。这条性质极其重要：它把“积分”变成了“数方程的解”——见引理 3.2。

符号 $\ll$ 与 $O(\cdot)$: Vinogradov / 大 O 记号

$A \ll B$ (读作“ A 远小于等于 B ”，与 $A = O(B)$ 同义) 意思是：存在一个常数 $C > 0$ ，使得 $|A| \leq C B$ 恒成立。这个常数 C 与我们关心的“大变量” (这里是 P) 无关，但允许依赖固定的参数 (这里是次数 k 和 ε)。例如 $3n + 5 \ll n$ ，因为 $3n + 5 \leq 8n$ (取 $C = 8$ ，对 $n \geq 1$)。它让我们只盯量级、忽略不重要的常数，这正是解析数论的工作方式。脚注里说明：本章 $\ll$ 隐含的常数只依赖 k 与 ε 。

符号 ε : 任意小的正数

$\varepsilon > 0$ 表示一个可以取得任意小的固定正数。“对任意 $\varepsilon > 0$ 成立 $A \ll P^\varepsilon$ ”意思是：无论你把 ε 取多小（比如 0.001），都有一个对应的常数把 A 压在 P^ε 之下。直观上 P^ε 增长得比任何 P 的正数次幂都慢（比 $\log P$ 还略大一点），是数论里“几乎可以忽略”的因子。注意：常数随 ε 变小而变大，但只要 ε 固定就没问题。

符号 $(a, q) = 1$ 、同余 $\equiv \pmod{q}$ 、剩余系

(a, q) 是 a 与 q 的最大公约数， $(a, q) = 1$ 表示二者互素（没有公共素因子）。 $am \equiv r \pmod{q}$ 表示 am 与 r 除以 q 余数相同，即 $q \mid (am - r)$ 。一个事实（本章要用）：若 $(a, q) = 1$ ，当 m 取遍 $0, 1, \dots, q-1$ 时， am 除以 q 的余数也恰好取遍 $0, 1, \dots, q-1$ 各一次——这叫“ am 跑遍模 q 的完全剩余系”。原因：若 $am_1 \equiv am_2$ ，则 $q \mid a(m_1 - m_2)$ ，因 $(a, q) = 1$ 得 $q \mid (m_1 - m_2)$ ，在 $0 \leq m$

符号 $\|\lambda\|$ ：到最近整数的距离

$\|\lambda\|$ 表示实数 λ 到离它最近的整数的距离，例如 $\|3.2\| = 0.2$ ， $\|3.8\| = 0.2$ ， $\|5\| = 0$ 。它总满足 $0 \leq \|\lambda\| \leq \frac{1}{2}$ 。为什么本章离不开它？因为 $e(\lambda x)$ 只看 λ 的小数部分，当 λ 接近整数（ $\|\lambda\|$ 小）时 $e(\lambda) \approx 1$ ，求和不会互相抵消、和会很大； $\|\lambda\|$ 正好度量“离危险有多近”。

符号 $d(m)$ ：除数函数

$d(m) = \sum_{d \mid m} 1$ 是 m 的正因子个数（“ $d \mid m$ ”读作“ d 整除 m ”）。例如 $d(6) = 4$ （因子 1, 2, 3, 6）， $d(12) = 6$ 。本章会证明它增长极慢： $d(m) \ll m^\varepsilon$ 。

符号 差分算子 $\Delta_y f$

$\Delta_y f(x) := f(x+y) - f(x)$, 叫“以步长 y 的一阶（前向）差分”。它是导数的离散版本。多次差分写成 $\Delta_{y_1, \dots, y_\nu} f = \Delta_{y_\nu}(\dots \Delta_{y_1} f)$, 即依次以步长 $y_1, \dots, y_\nu$ 做差分。**核心性质**：差分使多项式降低一次。下面证明里会反复用。

第1节 引言：Weyl 不等式的来历

这一段在讲什么：交代 Weyl 不等式是谁、什么时候、为什么提出的——它是研究华林问题乃至整个解析数论“最重要的单件工具”。

研究华林问题 (Waring's problem), 乃至研究数论解析理论中的许多其他问题时, 最重要的单一工具就是 Weyl 不等式。这一不等式由 Weyl 在其 1916 年关于数列模 1 均匀分布的伟大论文中以一种不太显式的形式给出。针对多项式、并以其最高次系数表述的显式形式, 则由 Hardy 与 Littlewood 给出。

历史背景：人物、年代、难点

华林问题 (1770)：英国数学家 Waring 猜想——对每个固定的 $k \geq 2$, 存在一个数 s , 使得每个正整数都能写成至多 s 个 k 次方数之和 ($k=2$ 是四平方和定理, $k=3$ 是若干立方数之和……)。这要数“某数有多少种写法”, 自然导出指数和 $T(\alpha) = \sum e(\alpha x^k)$ 。

Hermann Weyl (1885–1955) 是 20 世纪最伟大的数学家之一。他 1916 年的论文研究“数列 $\{\alpha n^k\}$ 的小数部分是否在 $[0, 1]$ 上均匀分布”。要判断均匀分布, 按 Weyl 判别法就得估计 $\sum e(\alpha x^k)$ 的大小——若这个和相对 P 很小, 就说明转动充分“打散”、分布均匀。为此他发明了“反复平方降次”的技巧, 这就是 Weyl 不等式的雏形。

Hardy 与 Littlewood 在他们 1920 年代关于华林问题的系列论文（书中简称 P.N. I, 即 “Partitio Numerorum I”）中，把 Weyl 的思想整理成针对一般 k 次多项式、并用“最高次系数 α 的有理逼近”来表述的**显式**不等式——也就是下面引理 3.1 的形式。为什么“显式”重要？因为圆法要把 $[0, 1]$ 切成“主弧”（ α 接近小分母分数）和“次弧”（其余），在次弧上必须有一个能**具体算出量级**的上界，Weyl 不等式正好提供它。

第 2 节 引理 3.1: Weyl 不等式的陈述

这一段在讲什么：正式写出 Weyl 不等式。它说：只要 α 有一个“好的有理逼近” a/q ，那么指数和 $\sum e(f(x))$ 就有一个明确的上界。

引理 3.1. (Weyl 不等式) 设 $f(x)$ 是一个次数为 k 、最高次系数为 α 的实系数多项式：

$$f(x) = \alpha x^k + \alpha_1 x^{k-1} + \cdots + \alpha_k.$$

假设 α 有一个满足下述条件的有理逼近 a/q ：

$$(a, q) = 1, \quad q > 0, \quad \left| \alpha - \frac{a}{q} \right| \leq \frac{1}{q^2}.$$

那么，对任意 $\varepsilon > 0$ ，有

$$\left| \sum_{x=1}^P e(f(x)) \right| \ll P^{1+\varepsilon} \left(P^{-\frac{1}{K}} + q^{-\frac{1}{K}} + \left(\frac{P^k}{q} \right)^{-\frac{1}{K}} \right), \quad (3.1')$$

其中 $K = 2^{k-1}$ ，所含的隐性常数只依赖于 k 与 ε 。

$f(x)$

一个 k 次实系数多项式。注意——只有**最高次系数** α 进入最终的上界，低次系数 $\alpha_1, \dots, \alpha_k$ 完全不影响结果。后面会看到原因：证明用差分反复降次，最后只剩下与 α 有关的线性项，其余系数都被“差掉了”。

P

求和的项数（自变量 x 从 1 取到 P ）。它是“大变量”，我们关心 $P \rightarrow \infty$ 时和的量级。**平凡上界**是 P ：因为每项模长为 1， P 项相加模长至多 P （三角不等式 $|\sum| \leq \sum |\cdot| = P$ ）。引理的价值在于给出比 P 更小的上界。

a/q

α 的有理逼近， $q > 0$ 是分母， $(a, q) = 1$ 保证它是“最简分数”， $|\alpha - a/q| \leq 1/q^2$ 保证逼近“足够好”。**为什么这种逼近一定存在？** 这是 Dirichlet 逼近定理：对任意实数 α 与任意 $Q \geq 1$ ，总能找到 $1 \leq q \leq Q$ 与整数 a ，使 $|\alpha - a/q| \leq 1/(qQ) \leq 1/q^2$ 。所以这个假设对每个 α 都能满足，引理是普遍适用的。

$K = 2^{k-1}$

“ K ”是后面反复平方所产生的指数。每平方一次幂指数翻倍，做 $k - 1$ 次后变成 2^{k-1} 。所以 $k = 2$ 时 $K = 2$ ， $k = 3$ 时 $K = 4$ ， $k = 4$ 时 $K = 8 \cdots \cdots$ 它随 k 指数级增大，这也是后来 Vinogradov 改进的动机（见后注）。

第 3 节 引理后的注：这个上界好在哪、什么时候有用

这一段在讲什么：解读 Weyl 不等式“有用”的范围—— q 不能太小也不能太大，并解释为什么 q 太小时它必然失效（这是事物本身的性质，不是证明的缺陷）。

注. 当 $P^\delta \leq q \leq P^{k-\delta}$ （对某个固定的 $\delta > 0$ ）时，该不等式对平凡上界 P 给出了某种改进。若 $P \leq q \leq P^{k-1}$ ，则我们得到估计 $P^{1-1/K+\varepsilon}$ ；正是在这些条件下，Weyl 不等式得到最为常见的应用。显然不可能从中提取出比这更好的估计。注意，若 q 很小，则 Weyl 不等式不能给出任何有用的信息；这

是很自然的，因为若 $f(x) = \alpha x^k$ 且 α 非常接近于一个分母很小的有理数，则该和确实具有趋近于 P 的量级。

把这段注一句句拆开

1 括号里三项谁说了算。上界括号里是 $P^{-1/K} + q^{-1/K} + (P^k/q)^{-1/K}$ 。括号外有个 $P^{1+\varepsilon}$ 。三项之和由**最大的那一项**决定量级。第一项 $P^{-1/K}$ 是固定的“底线改进”；第二项随 q 增大而减小；第三项 $(P^k/q)^{-1/K} = q^{1/K}/P^{k/K}$ 随 q 增大而增大。所以 q 太小时第三项也许还行但第二项 $q^{-1/K}$ 很大（接近 1）， q 太大时第三项很大。两头都坏，中间才好。

2 为什么取 $P^\delta \leq q \leq P^{k-\delta}$ 。要让上界真正小于 P ，需要括号 $\ll P^{-\eta}$ （某个 $\eta > 0$ ）。第二项 $q^{-1/K} \leq P^{-\delta/K}$ 要求 $q \geq P^\delta$ ；第三项 $(P^k/q)^{-1/K} \leq P^{-\delta/K}$ 要求 $P^k/q \geq P^\delta$ 即 $q \leq P^{k-\delta}$ 。两条合起来就是注里的范围。这就是“为什么这样取参数”的理由：恰好让两端的坏项都被压下去。

3 “最常用”的 $P \leq q \leq P^{k-1}$ 。此时 $q^{-1/K} \leq P^{-1/K}$ 且 $(P^k/q)^{-1/K} \leq P^{-1/K}$ ，括号三项都 $\leq P^{-1/K}$ ，于是整体 $\ll P^{1+\varepsilon} \cdot P^{-1/K} = P^{1-1/K+\varepsilon}$ 。这就是那个著名的省力上界：比平凡的 P 小了一个因子 $P^{1/K}$ （再差一点点 P^ε ）。圆法在次弧上用的正是这个版本。

4 为什么 q 很小时必然没用——而且这是“正常”的。取最简单的 $f(x) = \alpha x^k$ 且 α 极接近一个小分母分数 a/q （ q 很小）。那么 $\alpha x^k \approx (a/q)x^k$ ，而 $e((a/q)x^k)$ 关于 x 是周期为 q 的，许多项指向同一个方向、互相叠加而非抵消，整个和的模长真的会接近 P 。既然和本身就大，任何上界都不可能小于它——所以不是 Weyl 不等式“弱”，而是**事实**

如此。这正是圆法把这种 α （小分母附近）单独划成“主弧”去精细处理的原因。

第 4 节 引理 3.1 的证明（上）：平方降次的基本操作

这一段在讲什么：说明整个证明的核心套路——把指数和的模长取平方，就能把一个 k 次多项式的和，换成许多 $(k-1)$ 次多项式的和的平均。这是 Weyl 的发明，叫“Weyl 平方”或“Weyl differencing”。

证明中的基本操作是对一个指数和的绝对值取平方，从而把该和与一些次数更低的多项式的同类和的平均值联系起来。令

$$S_k(f) = \sum_{x=P_1+1}^{P_2} e(f(x)),$$

其中 $0 \leq P_2 - P_1 \leq P$ ，下标 k 用于标记 $f(x)$ 的次数。

为什么引入 $S_k(f)$ 这个一般记号

原来的和是 x 从 1 到 P 。但证明里反复降次后，会出现“ x 取遍某个长度 $\leq P$ 的区间”的和。为了让归纳能进行，作者把和写成更一般的 $\sum_{x=P_1+1}^{P_2}$ ，区间长度 $P_2 - P_1$ 介于 0 和 P 之间。下标 k 提醒我们里面的多项式 f 是 k 次的。动机：降一次次数，下标就从 k 变 $k-1$ ，区间也会变（依赖于差分步长 y ），用统一记号才好递推。

则

$$\begin{aligned} |S_k(f)|^2 &= \sum_{x_1} \sum_{x_2} e(f(x_2) - f(x_1)) \\ &= P_2 - P_1 + 2\Re \sum_{\substack{x_1, x_2 \\ x_2 > x_1}} e(f(x_2) - f(x_1)). \end{aligned}$$

补全：模长平方为什么等于双重和

- 1 对复数 z , $|z|^2 = z\bar{z}$ 。这里 $z = S_k(f) = \sum_{x_2} e(f(x_2))$, 它的共轭 $\bar{z} = \sum_{x_1} \overline{e(f(x_1))} = \sum_{x_1} e(-f(x_1))$ (因为 $\overline{e(\theta)} = e(-\theta)$, 用了 $\overline{e^{i\varphi}} = e^{-i\varphi}$)。

- 2 两个和相乘, 按分配律把每个 x_2 项乘每个 x_1 项:

$$|S_k|^2 = \left(\sum_{x_2} e(f(x_2)) \right) \left(\sum_{x_1} e(-f(x_1)) \right) = \sum_{x_1} \sum_{x_2} e(f(x_2) - f(x_1))$$

这里又用了 $e(A)e(-B) = e(A - B)$ 。两个变量都跑遍同一个区间 $(P_1, P_2]$ 。

- 3 分出对角项与非对角项。当 $x_1 = x_2$ 时, 指数 $f(x_2) - f(x_1) = 0$, $e(0) = 1$ 。这样的对角项有“区间内整数个数”那么多, 即 $P_2 - P_1$ 个, 贡献 $\sum 1 = P_2 - P_1$ 。

- 4 剩下 $x_1 \neq x_2$ 的项可两两配对: 项 (x_1, x_2) 与项 (x_2, x_1) 的指数互为相反数, 于是这两项 $e(\theta) + e(-\theta) = 2\Re e(\theta)$ (用了 $z + \bar{z} = 2\Re z$)。把每对只数一次 (约定 $x_2 > x_1$), 就得到 $2\Re \sum_{x_2 > x_1} e(f(x_2) - f(x_1))$ 。

- 5 合并即得所示等式。关键收获: 平方后, 被研究的对象从 $f(x)$ 变成了差 $f(x_2) - f(x_1)$ 。下一步会看到这个差的次数比 f 低一次。

令 $x_2 = x_1 + y$ 。则 $1 \leq y < P_2 - P_1$, 且

$$f(x_2) - f(x_1) = f(x_1 + y) - f(x_1) = \Delta_y f(x_1),$$

此处采用了显然的记号。于是

$$|S_k(f)|^2 = P_2 - P_1 + 2\Re \sum_{y=1}^P \sum_x e(\Delta_y f(x)),$$

其中对 x 的求和取遍一个依赖于 y 但包含于 $P_1 < x \leq P_2$ 之中的区间。
对某些 y 的取值而言，该区间可能为空。

补全：换元 $x_2 = x_1 + y$ 与差分降次

1 换元的动机：非对角和里出现 $f(x_2) - f(x_1)$ ，我们想固定“间距”
 $y = x_2 - x_1$ 来求和。因为 $x_2 > x_1$ 且都在 $(P_1, P_2]$ 内，所以 y 至少是 1，至多是区间长度减 1，即 $1 \leq y < P_2 - P_1 \leq P$ 。把外层换成对 y 求和，内层对 $x := x_1$ 求和。

2 记 $\Delta_y f(x) = f(x + y) - f(x)$ （前面定义过的差分算子）。则
 $f(x_2) - f(x_1) = \Delta_y f(x_1)$ 。

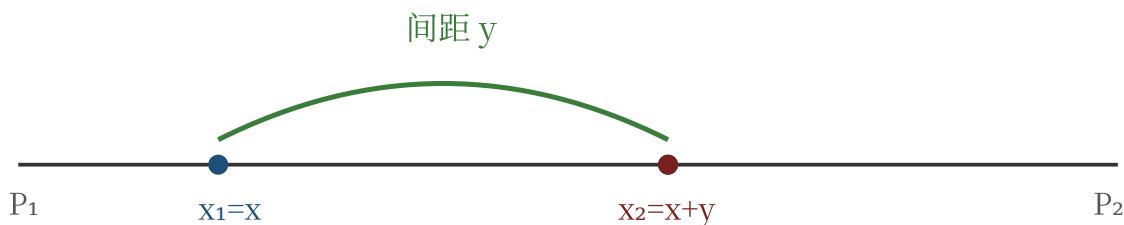
3 为什么 $\Delta_y f$ 比 f 低一次（这是全章发动机）。设 $f(x) = \alpha x^k + (\text{低次})$ 。
看最高次项：

$$(x + y)^k - x^k = \left(x^k + kyx^{k-1} + \dots \right) - x^k = kyx^{k-1} + (\text{更低次}),$$

其中用了二项式定理 $(x + y)^k = \sum_j \binom{k}{j} x^{k-j} y^j$ ， $j = 0$ 的 x^k 被减掉，
剩下最高是 $j = 1$ 的 kyx^{k-1} 。于是 $\Delta_y f(x) = k\alpha y x^{k-1} + \dots$ ，确是
 $k - 1$ 次，最高次系数变成 $k\alpha y$ 。

4 对 x 求和的范围：要求 $x = x_1$ 和 $x_2 = x_1 + y$ 都落在 $(P_1, P_2]$ 内，即 \

固定间距 y ，让左端点 x 在区间里滑动，
 每一对 $(x, x+y)$ 贡献 $e(\Delta_y f(x))$ ；
 $\Delta_y f$ 的次数比 f 低 1。



平方 + 换元的几何图像：把“所有数对”按它们的间距 y 分类，对每个 y 再让左端点 x 滑过区间。被研究的多项式从 k 次降到 $k-1$ 次。

特别地，

$$|S_k(f)|^2 \leq P + 2 \sum_{y=1}^P |S_{k-1}(\Delta_y f)|,$$

其中 S_{k-1} 所对应的区间正是刚才描述的那种。

补全：从等式到不等式

1 对角贡献 $P_2 - P_1 \leq P$ ，放大成 P 。

2 对 $2\Re \sum_y \sum_x e(\Delta_y f(x))$ ：先用 $\Re w \leq |w|$ ，再用三角不等式 $|\sum_y (\dots)| \leq \sum_y |\sum_x e(\Delta_y f(x))|$ 。注意内层 $\sum_x e(\Delta_y f(x))$ 正是一个 $(k-1)$ 次多项式 $\Delta_y f$ 的指数和，记作 $S_{k-1}(\Delta_y f)$ 。

3 于是 $|S_k(f)|^2 \leq P + 2 \sum_{y=1}^P |S_{k-1}(\Delta_y f)|$ 。意义：一个 k 次和的平方，被一堆 $(k-1)$ 次和的模长之和控制住了。降次成功一次。

重复上述论证，我们得到

$$|S_{k-1}(\Delta_y f)|^2 \leq P + 2 \sum_{z=1}^P |S_{k-2}(\Delta_{y,z} f)|,$$

其中 S_{k-2} 的求和区间同时依赖于 y 与 z 。利用 Cauchy 不等式，便可把第二个不等式中的 S_{k-1} 代入第一个不等式：

$$\begin{aligned} |S_k(f)|^4 &\ll P^2 + P \sum_{y=1}^P |S_{k-1}(\Delta_y f)|^2 \\ &\ll P^3 + P \sum_{y=1}^P \sum_{z=1}^P |S_{k-2}(\Delta_{y,z} f)|. \end{aligned}$$

工具：Cauchy–Schwarz 不等式

柯西不等式（高中可理解的求和形式）：对任意实数 $a_1, \dots, a_n$,

$$\left(\sum_{i=1}^n a_i \right)^2 \leq n \sum_{i=1}^n a_i^2.$$

（这是 $(\sum a_i \cdot 1)^2 \leq (\sum a_i^2)(\sum 1^2)$ 的特例。）直观：把“和的平方”换成“项数乘以平方和”，代价是多一个因子 n 。**为什么这里要用它？** 因为我们手上有 $\sum_y |S_{k-1}|$ （一次方之和），而降次公式给的是 $|S_{k-1}|^2$ （平方）的上界。柯西不等式正好把“一次方之和”转成“平方之和”，于是能代入。

补全：怎么用 Cauchy 把两层不等式接起来

- 1 对第一层 $|S_k|^2 \leq P + 2 \sum_{y=1}^P |S_{k-1}(\Delta_y f)|$ 两边平方。用 $(A + B)^2 \leq 2A^2 + 2B^2$ （这是 $(A - B)^2 \geq 0$ 的推论）：

$$|S_k|^4 \leq 2P^2 + 8 \left(\sum_{y=1}^P |S_{k-1}| \right)^2.$$

2 对括号里的和用柯西不等式 ($n = P$ 项):

$(\sum_{y=1}^P |S_{k-1}|)^2 \leq P \sum_{y=1}^P |S_{k-1}|^2$ 。于是

$$|S_k|^4 \leq 2P^2 + 8P \sum_{y=1}^P |S_{k-1}(\Delta_y f)|^2 \ll P^2 + P \sum_{y=1}^P |S_{k-1}|^2.$$

(常数 2、8 都吸进 $\ll$ 。) 这正是第一行。

3 把第二层 $|S_{k-1}(\Delta_y f)|^2 \leq P + 2 \sum_{z=1}^P |S_{k-2}(\Delta_{y,z} f)|$ 代入:

$$P \sum_y |S_{k-1}|^2 \leq P \sum_y \left(P + 2 \sum_z |S_{k-2}| \right) = P \cdot P \cdot P + 2P \sum_y \sum_z |$$

第一块 $P \cdot P \cdot P = P^3$ (外层 $\sum_y$ 有 P 项, 每项里那个 P , 再乘前面的 P)。

4 合起来 $|S_k|^4 \ll P^3 + P \sum_{y=1}^P \sum_{z=1}^P |S_{k-2}(\Delta_{y,z} f)|$ 。降次成功第二次, 代价是和的层数多了一层、外面的 P 幂也升高了。

此过程可以继续进行, 由此建立的一般不等式为

$$|S_k(f)|^{2^\nu} \ll P^{2^\nu-1} + P^{2^\nu-\nu-1} \sum_{y_1=1}^P \cdots \sum_{y_\nu=1}^P |S_{k-\nu}(\Delta_{y_1, \dots, y_\nu} f)|. \quad (3.1)$$

这只需对 ν 作归纳即可证明。须知 (3.1) 中对 x 的求和范围是一个依赖于 $y_1, \dots, y_\nu$ 的区间, 但包含于 $P_1 < x \leq P_2$ 。

补全: 对 ν 的归纳证明 (3.1)

1 基础 $\nu = 1$: 就是上面证过的 $|S_k|^2 \leq P + 2 \sum_y |S_{k-1}|$, 对照 (3.1):

$2^1 - 1 = 1$ (前一项 P^1), $2^1 - 1 - 1 = 0$ (后一项前缀 $P^0 = 1$), 完全吻合。

- 2 归纳假设：设 (3.1) 对 ν 成立。两边平方并用 $(A+B)^2 \leq 2A^2 + 2B^2$ ：

$$|S_k|^{2^{\nu+1}} \ll P^{2^{\nu+1}-2} + P^{2^{\nu+1}-2\nu-2} \left(\sum_{y_1, \dots, y_\nu} |S_{k-\nu}| \right)^2.$$

- 3 对那个多重和用柯西不等式，项数是 P^ν （共 ν 个变量各跑 P ）：

$$\left(\sum_{y_1, \dots, y_\nu} |S_{k-\nu}| \right)^2 \leq P^\nu \sum_{y_1, \dots, y_\nu} |S_{k-\nu}|^2.$$

- 4 对每个 $|S_{k-\nu}(\Delta_{y_1, \dots, y_\nu} f)|^2$ 再用一次“平方降次”基本操作（引入新变量 $y_{\nu+1}$ ）： $|S_{k-\nu}|^2 \leq P + 2 \sum_{y_{\nu+1}} |S_{k-\nu-1}|$ 。

- 5 把这些都串起来，整理 P 的幂次：第一项变 $P^{2^{\nu+1}-1}$ ，第二项前缀变 $P^{2^{\nu+1}-(\nu+1)-1}$ ，多重和变成对 $y_1, \dots, y_{\nu+1}$ 的 $\nu+1$ 重和，被加的是 $|S_{k-\nu-1}|$ 。这正是把 (3.1) 里的 ν 换成 $\nu+1$ 。归纳完成。

- 6 幂次记忆法：每平方一次，左边指数翻倍 ($2^\nu \rightarrow 2^{\nu+1}$)；第一项 $P^{2^\nu-1}$ 是“全程用平凡上界”留下的；第二项前缀 $P^{2^\nu-\nu-1}$ 里的 $-\nu$ 来自 ν 次柯西不等式各贡献的一个 P （被分摊）。

在此我们插入一点说明，它在引理 3.2 的证明中会有用。这就是：若在导出 (3.1) 的最后一步中，我们把基本操作以其原始形式应用，则得到

$$|S_k(f)|^{2^\nu} \ll P^{2^\nu-1} + P^{2^\nu-\nu-1} \sum_{y_1=1}^P \cdots \sum_{y_\nu=1}^P \Re S_{k-\nu}(\Delta_{y_1, \dots, y_\nu} f). \quad (3.2)$$

此处同样， $S_{k-\nu}$ 中 x 的取值范围依赖于 $y_1, \dots, y_\nu$ ，并且有时可能为空。

补全：(3.2) 与 (3.1) 差在哪

1 回忆最初的等式 $|S_k|^2 = (P_2 - P_1) + 2\Re \sum_y \sum_x e(\Delta_y f(x))$ 是带 $\Re$ 的等式。在推 (3.1) 时，最后一步我们用 $\Re(\cdot) \leq |\cdot|$ 放成了绝对值。

2 但如果最后一步不放绝对值，保留 $\Re$ 与“求和后再取实部”的原始形式，就得到 (3.2)：被加的是 $\Re S_{k-\nu}$ 而不是 $|S_{k-\nu}|$ 。

3 为什么要专门留这个版本？因为引理 3.2 要把 $S_{k-\nu}$ （不取绝对值的、带正负号的复数）整体拿去乘 $|T|^{2\nu}$ 再积分——积分会用到正交关系，把它变成“某方程解的个数”。若提前取了绝对值就破坏了这个结构。这是一个“为后文埋的伏笔”，体现了好的证明会保留尽量多的信息。

第 5 节 引理 3.1 的证明（中）：降到一次、化为几何级数

这一段在讲什么：取 $\nu = k - 1$ ，把多项式一路差分到一次（线性）。一次多项式的指数和是等比数列求和，能精确算出，于是出现了 $\|\lambda\|$ 。

回到 (3.1)，我们取 $\nu = k - 1$ ，并在原始的 S_k 中取 $P_1 = 0, P_2 = P$ 。我们注意到

$$\Delta_{y_1, \dots, y_{k-1}} f(x) = k! \alpha y_1 \cdots y_{k-1} x + \beta,$$

不妨这样记，其中 β 是一些与 x 无关的项的集合。

补全：为什么 $k - 1$ 次差分后系数是 $k! \alpha y_1 \cdots y_{k-1}$

1 取 $\nu = k - 1$ ，从 k 次降 $k - 1$ 次到 1 次。每差分一次，次数减 1、最高次系数乘上（当前次数） $\times$ （步长），前面在 $(x + y)^k - x^k = ky x^{k-1} + \cdots$ 已看到这个规律。

2 逐次追踪最高次系数：起点 α (k 次)。第一次差分（步长 y_1 ）：系数 $\rightarrow k\alpha y_1$ ($k-1$ 次)。第二次（步长 y_2 ）： $\rightarrow (k-1) \cdot k\alpha y_1 y_2$ ($k-2$ 次)。……第 $k-1$ 次：系数 $\rightarrow k(k-1)(k-2)\cdots 2 \cdot \alpha y_1 \cdots y_{k-1}$ (1 次)。

3 而 $k(k-1)\cdots 2 = k!/1! = k!$ (这是从 k 一直乘到 2，正好是 k 的阶乘)。所以一次项系数恰为 $k!\alpha y_1 \cdots y_{k-1}$ 。

4 常数项 β 是什么？差分得到的一次多项式 = (一次系数) x + (常数)。那个常数依赖于所有 y_i 和 f 的低次系数 $\alpha_1, \dots, \alpha_k$ ，但不依赖 x 。我们把它统称 β 。下一步会看到， $e(\beta)$ 是一个模长为 1 的常数，提到求和号外面不影响 $\left| \sum \right|$ ——这就是为什么低次系数最终全部消失、只有最高次系数 α 进入答案。

于是

$$|S_1(\Delta_{y_1, \dots, y_{k-1}} f)| = \left| \sum_x e(k! \alpha y_1 \cdots y_{k-1} x) \right|.$$

右边的和取遍任一长度至多为 P 的 x 区间，形如

$$\left| \sum_{x=x_1}^{x_2-1} e(\lambda x) \right| \leq \frac{2}{|1 - e(\lambda)|} = \frac{1}{|\sin \pi \lambda|} \ll \frac{1}{\|\lambda\|},$$

其中 $\|\lambda\|$ 表示 λ 到最近整数的距离。当 λ 是整数时这一估计失效，但我们可以用显然的上界 P 来补充。

补全：等比数列求和与 $\|\lambda\|$ 估计（每一步）

1 把 $e(\beta)$ 提到外面：

$$\left| \sum_x e(\lambda x + \beta) \right| = |e(\beta)| \cdot \left| \sum_x e(\lambda x) \right| = \left| \sum_x e(\lambda x) \right|, \text{ 其中}$$

$\lambda = k! \alpha y_1 \cdots y_{k-1}$ 。这一步用 $|e(\beta)| = 1$ ，确认了 β 不影响大小。

- 2 等比求和。记 $w = e(\lambda)$ ，则 $\sum_{x=x_1}^{x_2-1} e(\lambda x) = \sum_x w^x$ 是公比 w 的等比数列。用求和公式 $\sum_{x=x_1}^{x_2-1} w^x = \frac{w^{x_1} - w^{x_2}}{1 - w}$ ($w \neq 1$)。取模长，分子 $|w^{x_1} - w^{x_2}| \leq |w^{x_1}| + |w^{x_2}| = 2$ (三角不等式, $|w| = 1$)，所以

$$\left| \sum_x e(\lambda x) \right| \leq \frac{2}{|1 - w|} = \frac{2}{|1 - e(\lambda)|}.$$

- 3 把 $|1 - e(\lambda)|$ 算成 $\sin$ 。用 $1 - e^{i\varphi} = 1 - \cos \varphi - i \sin \varphi$ ，其模长平方 $(1 - \cos \varphi)^2 + \sin^2 \varphi = 2 - 2 \cos \varphi = 4 \sin^2 \frac{\varphi}{2}$ ，故 $|1 - e^{i\varphi}| = 2 \left| \sin \frac{\varphi}{2} \right|$ 。代 $\varphi = 2\pi\lambda$ 得 $|1 - e(\lambda)| = 2 \left| \sin \pi\lambda \right|$ 。于是 $\frac{2}{|1 - e(\lambda)|} = \frac{1}{|\sin \pi\lambda|}$ 。

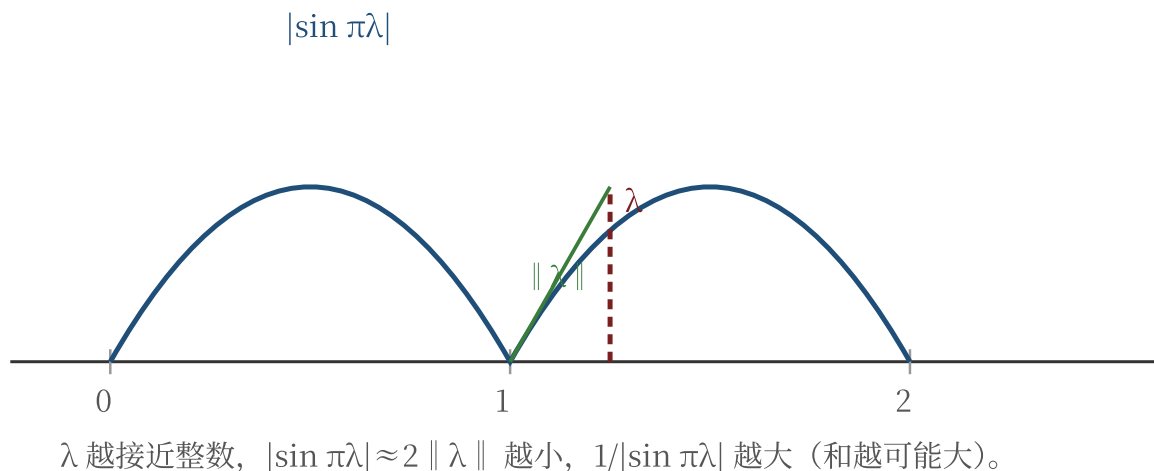
- 4 用 $\|\lambda\|$ 替换 $\sin$ 。 $\sin \pi\lambda$ 只看 λ 小数部分，且当 λ 离整数距离为 $\|\lambda\| \leq \frac{1}{2}$ 时， $|\sin \pi\lambda| = \sin(\pi\|\lambda\|)$ 。利用初等不等式 $\sin(\pi t) \geq 2t$ (对 $0 \leq t \leq \frac{1}{2}$ ，因为正弦在这段是凹的、端点连线在曲线下方)，得 $|\sin \pi\lambda| \geq 2\|\lambda\|$ 。所以

$$\frac{1}{|\sin \pi\lambda|} \leq \frac{1}{2\|\lambda\|} \ll \frac{1}{\|\lambda\|}.$$

- 5 整数情形的补丁。若 λ 恰是整数， $\|\lambda\| = 0$ 、上面分母为 0、估计失效；其实那时每项 $e(\lambda x) = 1$ 、和等于项数，模长至多区间长 $\leq P$ 。所以无论如何都有

$$\left| \sum_x e(\lambda x) \right| \leq \min\left(P, \frac{1}{\|\lambda\|}\right) \quad (\text{约定 } \frac{1}{\|\lambda\|} = \infty \text{ 当 } \|\lambda\| = 0).$$

取“ P 与 $1/\|\lambda\|$ 中较小者”，两种情形就统一了——这正是后面 $\min(P, \dots)$ 的来历。



线性指数和的大小由 λ 到最近整数的距离 $\|\lambda\|$ 决定：离整数远则和小，离整数近（甚至等于整数）则需用平凡上界 P 兜底。

于是 (3.1) 给出

$$|S_k(f)|^K \ll P^{K-1} + P^{K-k} \sum_{y_1=1}^P \cdots \sum_{y_{k-1}=1}^P \min(P, \|k! \alpha y_1 \cdots y_{k-1}\|^{-1})$$

补全：代入 $\nu = k - 1$ 的幂次核对

① 在 (3.1) 里取 $\nu = k - 1$ ：左边指数 $2^\nu = 2^{k-1} = K$ ，第一项 $P^{2^\nu-1} = P^{K-1}$ ，第二项前缀 $P^{2^\nu-\nu-1} = P^{K-(k-1)-1} = P^{K-k}$ 。完全对上。

② 多重和里被加项 $|S_1(\Delta_{y_1, \dots, y_{k-1}} f)| \leq \min(P, \|k! \alpha y_1 \cdots y_{k-1}\|^{-1})$ ，由上一步。代入即得本式。这就把“高次指数和”彻底化简为“关于 $\|k! \alpha \cdot (\text{乘积})\|$ 的算术和”。

第6节 引理 3.1 的证明（下）：除数函数 + 分块求和

这一段在讲什么：多重和里出现的是乘积 $y_1 \cdots y_{k-1}$ ，很多不同的 $(y_1, \dots, y_{k-1})$ 给出同一个乘积值。把它们按乘积值 m 合并，需要“一个数有多少种因子分解”的估计——这就是除数函数 $d(m) \ll m^\varepsilon$ 。

现在我们诉诸初等数论中的一个结果，它使我们能把和中所有使 $k! y_1 \cdots y_{k-1}$ 取某个给定值（记为 m ）的项合并到一起。这样的项的个数是 $\ll m^\varepsilon$ 。为证明此点，只需证明

$$d(m) \ll m^\varepsilon, \quad (3.3)$$

对任意整数 m 成立。事实上，对于每个 $y_1, \dots, y_{k-1}$ ，至多有 $d(m)$ 种可能。

补全：为什么“合并项数”被 $d(m)$ 控制

1 固定乘积值 $m = k! y_1 \cdots y_{k-1}$ ，问：有多少组 $(y_1, \dots, y_{k-1})$ 给出它？等价于问 $y_1 \cdots y_{k-1} = m/k!$ 有多少有序解。

2 先看 y_1 ：它必须是 $m/k!$ 的一个因子，所以 y_1 的取法 $\leq d(m)$ 种。定下 y_1 后， y_2 是 $(m/k!)/y_1$ 的因子，取法又 $\leq d(m)$ ……每个变量都至多 $d(m)$ 种。所以总组数 $\leq d(m)^{k-1}$ 。

3 因为 $k-1$ 固定， $d(m)^{k-1} \ll (m^\varepsilon)^{k-1} = m^{(k-1)\varepsilon}$ ，把 $(k-1)\varepsilon$ 重新记作 ε （ ε 任意小，乘个常数还是任意小），得到“合并项数 $\ll m^\varepsilon$ ”。所以只需证核心事实 (3.3)： $d(m) \ll m^\varepsilon$ 。

为建立 (3.3)，设 $m = p_1^{\lambda_1} p_2^{\lambda_2} \cdots$ ，并注意到

$$\frac{d(m)}{m^\varepsilon} = \prod_i \frac{\lambda_i + 1}{p_i^{\varepsilon \lambda_i}} \leq \prod_{p_i \leq 2^{1/\varepsilon}} \frac{\lambda_i + 1}{2^{\varepsilon \lambda_i}} \leq C(\varepsilon),$$

因为 $2^{-\varepsilon\lambda}(\lambda + 1)$ 对 $\lambda > 0$ 是有上界的。

补全: $d(m) \ll m^\varepsilon$ 的完整证明

- 1 除数函数的乘积公式。设 m 的素因子分解 $m = p_1^{\lambda_1} p_2^{\lambda_2} \cdots$ 。 m 的每个因子形如 $p_1^{a_1} p_2^{a_2} \cdots$ ，其中 $0 \leq a_i \leq \lambda_i$ ，每个 a_i 有 $\lambda_i + 1$ 种选择，互相独立，所以 $d(m) = \prod_i (\lambda_i + 1)$ 。同时 $m^\varepsilon = \prod_i p_i^{\varepsilon\lambda_i}$ 。相除得

$$\frac{d(m)}{m^\varepsilon} = \prod_i \frac{\lambda_i + 1}{p_i^{\varepsilon\lambda_i}}.$$

- 2 把因子分两类。对“大素数” $p_i > 2^{1/\varepsilon}$ (即 $p_i^\varepsilon > 2$)，这一项 $\frac{\lambda_i + 1}{p_i^{\varepsilon\lambda_i}} \leq \frac{\lambda_i + 1}{2^{\lambda_i}} \leq 1$ (因为 $2^\lambda \geq \lambda + 1$ 对 $\lambda \geq 0$ 恒成立)。所以大素数的因子不会让乘积变大，可以丢掉 (≤ 1)。于是只剩“小素数” $p_i \leq 2^{1/\varepsilon}$ 的因子。

- 3 每个小素数因子有上界。对小素数项 $\frac{\lambda_i + 1}{p_i^{\varepsilon\lambda_i}} \leq \frac{\lambda_i + 1}{2^{\varepsilon\lambda_i}}$ (因为 $p_i \geq 2$)。考虑函数 $g(\lambda) = 2^{-\varepsilon\lambda}(\lambda + 1)$ ：当 $\lambda \rightarrow \infty$ 指数衰减压过线性增长，所以 g 在 $\lambda \geq 0$ 上有最大值，记为 $M(\varepsilon)$ 。于是每个小素数项 $\leq M(\varepsilon)$ 。

- 4 小素数个数有限。满足 $p \leq 2^{1/\varepsilon}$ 的素数只有有限个 (个数只依赖 ε ，记为 $N(\varepsilon)$)。于是

$$\frac{d(m)}{m^\varepsilon} \leq M(\varepsilon)^{N(\varepsilon)} =: C(\varepsilon),$$

一个只依赖 ε 、与 m 无关的常数。即 $d(m) \leq C(\varepsilon) m^\varepsilon$ ，亦即 $d(m) \ll m^\varepsilon$ 。证毕。

- 5 为什么这个估计“几乎是平凡的却不可缺”：它精确表达了“因子分解的方式不会太多”，把组合爆炸控制在可忽略的 m^ε 内，让我们放心地把多重和压缩成单重和。

如上所述把各项合并，我们得到

$$|S_k(f)|^K \ll P^{K-1} + P^{K-k+\varepsilon} \sum_{m=1}^{k!P^{k-1}} \min(P, \|\alpha m\|^{-1}).$$

补全：多重和 $\rightarrow$ 单重和

- 1 原来的 $(k-1)$ 重和 $\sum_{y_1, \dots, y_{k-1}} \min(P, \|k! \alpha y_1 \cdots y_{k-1}\|^{-1})$ ，被加项只通过乘积 $m' = y_1 \cdots y_{k-1}$ （等价地 $m = k! m'$ ）起作用。

- 2 按 m 合并：每个 m 值对应的项数 $\ll m^\varepsilon \leq (k!P^{k-1})^\varepsilon \ll P^\varepsilon$ （因为 $y_i \leq P$ ，乘积 $\leq P^{k-1}$ ， $m \leq k!P^{k-1}$ ）。把这个 P^ε 提出来。

- 3 m 的取值范围： $m = k! y_1 \cdots y_{k-1}$ 从最小 $k!$ 到最大 $k!P^{k-1}$ ，放宽下界写成 $m = 1$ 到 $k!P^{k-1}$ （多加的项非负，不影响 $\ll$ ）。注意被加项里 $\|k! \alpha y \cdots\| = \|\alpha m\|$ （把 $k!$ 吸进 m ）。于是多重和 $\ll P^\varepsilon \sum_{m=1}^{k!P^{k-1}} \min(P, \|\alpha m\|^{-1})$ 。

- 4 把这个并入 (3.1) 第二项的前缀， $P^{K-k} \cdot P^\varepsilon = P^{K-k+\varepsilon}$ ，即得所示式。现在问题归结为估计单重和 $\sum_m \min(P, \|\alpha m\|^{-1})$ 。

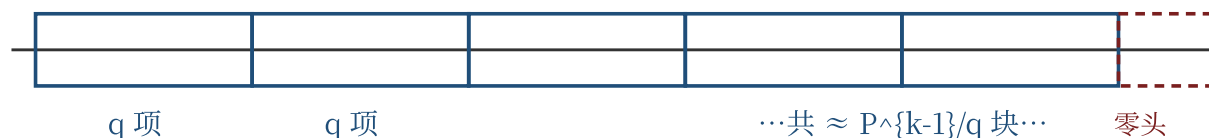
剩下的是用题设中提到的对 α 的有理逼近 a/q 来估计最后这个和。我们把对 m 的求和分成若干个由 q 个连续项构成的块，这样的块的个数是 $\ll \frac{P^{k-1}}{q} + 1$ 。

补全：为什么按 q 个一块分

1 动机：逼近条件是 $|\alpha - a/q| \leq 1/q^2$ 。当 m 在长为 q 的一段里变化时， αm 与 $(a/q)m$ 的差不超过 $q \cdot 1/q^2 = 1/q$ ，是个可控的小扰动。而 $(a/q)m$ 在一段长 q 内会把分母 q 的各个余数跑一遍，结构清楚。所以以 q 为块长最自然。

2 总项数约 $k!P^{k-1}$ ，每块 q 项，块数 $\approx k!P^{k-1}/q$ ，可能多出一个不满 q 的零头块，故块数 $\ll P^{k-1}/q + 1$ ($k!$ 是常数被吸收，“+1”管那个零头)。

$m: 1 \rightarrow k! \cdot P^{k-1}$ ，每 q 个连续整数为一块



把单重和切成长为 q 的块，每块内部 αm 近似按 a/q 的节奏跑遍所有余数，便于逐块估计。

考虑任一个块上的和，它形如

$$\sum_{m=0}^{q-1} \min(P, \|\alpha(m_1 + m)\|^{-1}),$$

其中 m_1 是该块中的第一个数。我们有

$$\alpha(m_1 + m) = \alpha m_1 + \frac{am}{q} + O\left(\frac{1}{q}\right),$$

这是因为 $|\alpha - a/q| \leq q^{-2}$ 且 $0 \leq m < q$ 。

补全：这一行的代数细节

- 1 块内第 m 项的自变量是 $m_1 + m$ ($m = 0, \dots, q-1$)。把 α 写成 $\alpha = \frac{a}{q} + \theta$, 其中 $|\theta| \leq 1/q^2$ 。则

$$\alpha(m_1 + m) = \alpha m_1 + \alpha m = \alpha m_1 + \left(\frac{a}{q} + \theta\right)m = \alpha m_1 + \frac{am}{q} + \theta m$$

- 2 误差项 θm : $|\theta m| \leq \frac{1}{q^2} \cdot (q-1) < \frac{1}{q}$, 所以 $\theta m = O(1/q)$ 。这就是式中的 $O(1/q)$ ——一个绝对值不超过 $1/q$ 的小量。

- 3 意义: 在一整块里, αm_1 是固定常数, 真正变化的是 $\frac{am}{q}$ (规整地跑余数) 加一个小扰动 $O(1/q)$ 。

当 m 从 0 取到 $q-1$ 时, am 遍历模 q 的完全剩余系。令 $am \equiv r \pmod{q}$, 则该和为

$$\sum_{r=0}^{q-1} \min\left(P, \frac{1}{\|(r+b)/q + O(1/q)\|}\right),$$

其中我们取 b 为最接近 $q\alpha m_1$ 的整数。

补全: 换成对余数 r 求和

- 1 因 $(a, q) = 1$, 前面预备知识里讲过: m 跑 $0, \dots, q-1$ 时 $am \bmod q$ 也跑遍 $0, \dots, q-1$ 。记 $r = am \bmod q$, 则求和指标可从 m 换成 r (双射, 不重不漏)。

- 2 此时 $\frac{am}{q} \equiv \frac{r}{q} \pmod{1}$ (差一个整数, $\|\cdot\|$ 与 $e(\cdot)$ 都不在乎整数差)。再把固定的 αm_1 也按整数拆开: 取 $b =$ “最接近 $q\alpha m_1$ 的整数”, 则

$\alpha m_1 \equiv \frac{b}{q} + (\text{小数})$ ，而那点小数已被并进 $O(1/q)$ 。综合得括号里距离参数是 $\frac{r+b}{q} + O(1/q)$ 。

- 3 于是块和 $= \sum_{r=0}^{q-1} \min(P, \|\frac{r+b}{q} + O(1/q)\|^{-1})$ 。问题变成：当 r 跑遍余数时， $\frac{r+b}{q}$ 这些点离整数多近。

在和中有 $O(1)$ 个 r 值使得最小值中的第二个表达式无用，即那些使得 $r+b \pmod{q}$ 的绝对最小剩余很小的 r 。对这些 r ，我们必须取 P 。对于其余的 r 值，若 s 表示 $r+b \pmod{q}$ 的绝对最小剩余，则有

$$\left\| \frac{r+b}{q} + O\left(\frac{1}{q}\right) \right\| \gg \frac{s}{q}.$$

因此上述和为

$$\ll P + \sum_{s=1}^{q/2} \frac{q}{s} \ll P + q \log q.$$

符号 “绝对最小剩余” s

把 $r+b$ 除以 q 的余数，调整到 $-q/2$ 与 $q/2$ 之间（例如 $q=10$ 时余数 7 记为 -3 ），它的绝对值 $s = |\cdot|$ 就是“ $r+b$ 离最近的 q 的倍数有多远”。于是 $\|\frac{r+b}{q}\| = \frac{s}{q}$ （在 $0 \leq s \leq q/2$ 时）。 s 从 0（恰是倍数）到约 $q/2$ 。

补全：逐块求和的最后估计

- 1 坏的 r （少数）。当 s 很小（ $\frac{s}{q}$ 与误差 $O(1/q)$ 同量级，即 $s = O(1)$ ）时， $\|\frac{r+b}{q} + O(1/q)\|$ 可能很小甚至为 0，倒数无意义，只能用平凡上界 P 。这样的 r 只有 $O(1)$ 个（ $s = 0, \pm 1, \dots$ 中有限几个），合计贡献 $\ll P$ 。

2 好的 r (多数)。当 s 不太小, 主项 $\frac{s}{q}$ 压过误差 $O(1/q)$, 故 $\|\frac{r+b}{q} + O(1/q)\| \gg \frac{s}{q}$, 于是该项 $\min(P, \dots) \leq (\frac{s}{q})^{-1} = \frac{q}{s}$ 。

3 对好 r 求和。不同 r 对应不同 s , s 从 1 到约 $q/2$ (每个 s 最多对应正负两个 r , 常数无妨)。于是好 r 部分 $\ll \sum_{s=1}^{q/2} \frac{q}{s} = q \sum_{s=1}^{q/2} \frac{1}{s}$ 。

4 调和级数。 $\sum_{s=1}^N \frac{1}{s} \ll \log N$ (调和级数的标准估计: 用 $\int_1^N \frac{dt}{t} = \log N$ 比较)。这里 $N \approx q/2$, 所以 $\sum_{s=1}^{q/2} \frac{1}{s} \ll \log q$, 好 r 部分 $\ll q \log q$ 。

5 坏 r 与好 r 相加: 每块的和 $\ll P + q \log q$ 。

计入块的数目, 我们得到

$$|S_k(f)|^K \ll P^{K-1} + P^{K-k+\varepsilon} \left(\frac{P^{k-1}}{q} + 1 \right) (P + q \log q).$$

由于可以假设 $q \leq P^k$, 因子 $\log q$ 可被吸收进 P^ε 中。于是右边为

$$\ll P^{K+\varepsilon} (P^{-1} + q^{-1} + P^{-k}q),$$

这就给出了所要的结果。

补全: 从这里到引理结论的全部代数

1 合并块。单重和 = (块数) $\times$ (每块和) $\ll (\frac{P^{k-1}}{q} + 1)(P + q \log q)$ 。代回上一式得展示的式子。

2 处理 $\log q$ 。若 $q > P^k$, 回看引理结论: 括号里 $(P^k/q)^{-1/K} = (q/P^k)^{1/K} > 1$, 于是右边 $> P^{1+\varepsilon}$, 比平凡上界 P 还大, 结论自动成立 (平凡)。所以不妨设 $q \leq P^k$, 从而

$\log q \leq \log(P^k) = k \log P \ll P^\varepsilon$ (对数远小于任何正幂)。把 $\log q$ 并进 P^ε 。

- 3 展开乘积。 $(\frac{P^{k-1}}{q} + 1)(P + q \log q)$, 吸收 $\log q$ 进 P^ε 后逐项乘开 (共四项):

$$\frac{P^{k-1}}{q} \cdot P = \frac{P^k}{q}, \quad \frac{P^{k-1}}{q} \cdot q = P^{k-1}, \quad 1 \cdot P = P, \quad 1 \cdot q = q.$$

其中 $P^{k-1} \leq P^{k-1}$ 与 P 都被更大的项盖过 (在 $q \leq P^k$ 下 $\frac{P^k}{q} \geq 1$)。主要留下 $\frac{P^k}{q}$ 、 P^{k-1} 、 q 这几类。

- 4 乘上前缀并整理。第二大项整体 $\ll P^{K-k+\varepsilon}(\frac{P^k}{q} + P^{k-1} + q)$ 。逐项乘 P^{K-k} :

$$P^{K-k} \cdot \frac{P^k}{q} = \frac{P^K}{q} = P^K q^{-1}, \quad P^{K-k} \cdot P^{k-1} = P^{K-1}, \quad P^{K-k} \cdot q$$

再加上第一项 $P^{K-1} = P^K \cdot P^{-1}$ 。三类合起来正是

$$|S_k(f)|^K \ll P^{K+\varepsilon}(P^{-1} + q^{-1} + P^{-k}q).$$

- 5 开 K 次方。两边取 $1/K$ 次幂。用

$(A + B + C)^{1/K} \leq A^{1/K} + B^{1/K} + C^{1/K}$ (次可加性, 因 $1/K \leq 1$):

$$|S_k(f)| \ll (P^{K+\varepsilon})^{1/K} (P^{-1} + q^{-1} + P^{-k}q)^{1/K} \ll P^{1+\varepsilon'} (P^{-1/K} + q$$

注意 $(P^{-k}q)^{1/K} = (q/P^k)^{1/K} = (P^k/q)^{-1/K}$ 。把 ε' 重记为 ε 。这正是引理 3.1 的结论 (3.1')。■

第 7 节 注：大 k 时 Vinogradov 的改进

这一段在讲什么：提醒读者——当 k 很大时，Weyl 不等式里那个指数 $K = 2^{k-1}$ 增长太快（指数级），效果变差；Vinogradov 用更巧的方法把它换成增长慢得多的 $4k^2 \log k$ 。

注. 若 k

较大，则 Vinogradov 给出了一个好得多的估计，其中（粗略地说） 2^{k-1} 被 $4k^2 \log k$ 所取代。

为什么这是重大改进

回忆“最常用”的上界是 $P^{1-1/K+\varepsilon}$ ，改进幅度由 $1/K$ 决定—— K 越小，节省越多。Weyl 的 $K = 2^{k-1}$ 随 k 指数爆炸（ $k = 10$ 已是 512），于是 $1/K$ 微乎其微，几乎回到平凡上界。Vinogradov 在 1930 年代发明“中值定理”，把有效指数压到 $\sim 4k^2 \log k$ （多项式级增长）， $1/K$ 大得多，节省显著。这是华林问题中 $G(k)$ （所需 k 次方个数）估计取得突破的关键，也是为什么大 k 要换工具。本书选择先讲 Weyl 不等式，是因为它简单、自足、足以处理小 k ，符合“先易后难”的取舍。

第 8 节 推论：完全和 $S_{a,q}$ 的估计

这一段在讲什么：把 Weyl 不等式用到一个特殊但极重要的对象——“完全指数和” $S_{a,q} = \sum_{z=1}^q e(az^k/q)$ 。它在后面建立奇异级数时反复出现。

推论（引理 3.1 之推论）. 令

$$S_{a,q} = \sum_{z=1}^q e(az^k/q),$$

其中 a, q 是互素的整数且 $q > 0$ 。则

$$S_{a,q} \ll q^{1-1/K+\varepsilon}.$$

补全：为什么它是引理 3.1 的特例

- 1 在引理 3.1 里取 $f(z) = \frac{a}{q}z^k$ ，即最高次系数 $\alpha = \frac{a}{q}$ ，没有低次项；并取和长度 $P = q$ (z 从 1 到 q)。
- 2 找 $\alpha = a/q$ 的有理逼近：它本身就是分数 a/q ，取逼近就用 a/q ，则 $|\alpha - a/q| = 0 \leq 1/q^2$ ，且 $(a, q) = 1$ ，条件全部满足，逼近分母正是 q 。
- 3 代入结论括号 $P^{-1/K} + q^{-1/K} + (P^k/q)^{-1/K}$ ，因 $P = q$ ：第一项 $q^{-1/K}$ ，第二项 $q^{-1/K}$ ，第三项 $(q^k/q)^{-1/K} = q^{-(k-1)/K} \leq q^{-1/K}$ ($k \geq 2$)。三项都 $\leq q^{-1/K}$ 。
- 4 于是 $S_{a,q} \ll P^{1+\varepsilon} \cdot q^{-1/K} = q^{1+\varepsilon} \cdot q^{-1/K} = q^{1-1/K+\varepsilon}$ 。证毕。

这是引理 3.1 的一个特例，对应于 $\alpha = a/q$ 且 $P = q$ 。我们以后（引理 6.4）将证明更精确的估计 $q^{1-1/k}$ 以取代 $q^{1-1/K+\varepsilon}$ ，但目前上述结果已足够使用。

解读：这里坦白指出当前结果不是最优——以后第 6 章会把指数从 $1 - 1/K + \varepsilon$ ($K = 2^{k-1}$) 改进到更好的 $1 - 1/k$ （注意 $1/k > 1/K$ ，所以 $q^{1-1/k}$ 更小、更强），而且没有 ε 。但“够用就先用”，体现了写作上的取舍：不在此处提前展开更难证明。

第 9 节 引理 3.2：华罗庚不等式的陈述

这一段在讲什么：给出本章第二件武器——华罗庚不等式。它不再估计单个 α 处的和，而是估计 $|T(\alpha)|$ 的高次幂在整个 $[0, 1]$ 上的平均（积分）。

引理 3.2. (华罗庚不等式) 若

$$T(\alpha) = \sum_{x=1}^P e(\alpha x^k),$$

则

$$\int_0^1 |T(\alpha)|^{2^k} d\alpha \ll P^{2^k - k + \varepsilon},$$

对任意固定的 $\varepsilon > 0$ 成立。

历史与动机：华罗庚是谁、这个不等式解决什么

华罗庚 (1910–1985) 是中国现代数学的奠基人之一，自学成才，在解析数论、矩阵几何、多复变等领域有奠基性贡献。这条以他命名的不等式发表于 1938 年，是华林问题圆法处理中的标准工具。

它为什么重要？ 圆法要算 $\int_0^1 T(\alpha)^s e(-N\alpha) d\alpha$ （它正好等于 N 写成 s 个 k 次方之和的方法数）。把 $[0, 1]$ 分成主弧（贡献主项）和次弧（希望贡献可忽略）。在次弧上要证明积分很小，关键就是要有 $\int |T|$ 某次幂的好上界——华罗庚不等式正好提供 2^k 次幂的均值估计。

平凡上界对照： $|T(\alpha)| \leq P$ ，所以 $\int_0^1 |T|^{2^k} \leq P^{2^k} \cdot \int_0^1 1 = P^{2^k}$ 。华罗庚不等式把它改进成 $P^{2^k - k + \varepsilon}$ ，省下了 $P^{k - \varepsilon}$ 这么大一个因子。指数里的 $-k$ 正是“每个变量积分一次省一点”累积的结果。

$T(\alpha)$

专指最高次单项 $f(x) = \alpha x^k$ （没有低次项）的指数和。这是华林问题真正要用的和。

$$\int_0^1 |T(\alpha)|^{2^k} d\alpha$$

把 $|T|$ 升到 2^k 次方再在 $[0, 1]$ 上求平均。为什么偏偏是 2^k 次？因为证明用“平方降次”，每平方一次幂翻倍，配合 k 步归纳，自然落到 2^k 。后面的注会讨论其它次幂（如 6 次）能否更好。

第 10 节 引理 3.2 的证明：归纳 + “积分=数解数”

这一段在讲什么：用对 ν 的归纳证明一串递增的估计 $I_\nu \ll P^{2^\nu - \nu + \varepsilon}$ ， $\nu = k$ 时正是要证的。核心新思想：这种积分等于某个方程整数解的个数，于是“分析问题”变成“数数问题”。

记

$$I_\nu = \int_0^1 |T(\alpha)|^{2^\nu} d\alpha.$$

我们对 ν 作归纳，证明

$$I_\nu \ll P^{2^\nu - \nu + \varepsilon}, \quad \nu = 1, \dots, k, \quad (3.4)$$

其中 $\nu = k$ 的情形即为引理所断言的结果。

为什么引入 I_ν 并对 ν 归纳

直接证 $\nu = k$ 太难，但相邻两步 $I_\nu \rightarrow I_{\nu+1}$ 之间有清晰联系（靠 (3.2) 那条“原始形式”不等式）。于是设一个一般的 I_ν ，证明“前一步成立 $\Rightarrow$ 后一步成立”，再加一个简单的起点，就能像多米诺骨牌一样推到 $\nu = k$ 。这就是数学归纳法在解析数论里的典型用法。

对 $\nu = 1$ ，该估计是显然的。我们有

$$I_1 = \int_0^1 \sum_{x_1} e(\alpha x_1^k) \sum_{x_2} e(-\alpha x_2^k) d\alpha = P,$$

因为当 $x_1 = x_2$ 时对 α 的积分为 1, 否则为 0。

补全：基础情形 $\nu = 1$ (把“积分=数解”看明白)

1 $I_1 = \int_0^1 |T|^2 d\alpha = \int_0^1 T \bar{T} d\alpha$ 。把 $T = \sum_{x_1} e(\alpha x_1^k)$ 、
 $\bar{T} = \sum_{x_2} e(-\alpha x_2^k)$ 乘开：

$$I_1 = \int_0^1 \sum_{x_1} \sum_{x_2} e(\alpha(x_1^k - x_2^k)) d\alpha = \sum_{x_1} \sum_{x_2} \int_0^1 e(\alpha(x_1^k - x_2^k)) d\alpha.$$

(先把有限求和与积分交换顺序，合法。)

2 用正交关系 (预备知识那条)：内层积分 $\int_0^1 e(n\alpha) d\alpha$ 在
 $n = x_1^k - x_2^k = 0$ 时为 1, 否则为 0。而 $x_1^k = x_2^k$ (正整数)
 $\iff x_1 = x_2$ 。

3 所以

$$I_1 = \#\{(x_1, x_2) : 1 \leq x_1, x_2 \leq P, x_1 = x_2\} = \#\{x_1 = x_2\} = P。$$

4 对照 (3.4)： $\nu = 1$ 时 $P^{2^1-1+\varepsilon} = P^{1+\varepsilon} \geq P$, 成立 (其实更强，等于 P)。这一步揭示了全证明的灵魂： $\int_0^1 |T|^{2m} d\alpha$ 恰好数 “
 $x_1^k + \cdots + x_m^k = y_1^k + \cdots + y_m^k$ 的解数”。

$$\int_0^1 |T(\alpha)|^{2m} d\alpha$$

(分析：一个积分)

=

$$x_1^k + \cdots + x_m^k = y_1^k + \cdots + y_m^k$$

在 $[1, P]$ 内的整数解个数

正交关系把“求积分”翻译成“数整数解”——这是圆法均值方法的核心。

关键转译： $|T|^{2m}$ 的积分 = 一个对称丢番图方程的解的个数。证明 3.2 全靠把解的个数数清楚。

现在假设 (3.4) 对某个特定的整数 $\nu \leq k-1$ 成立；我们要推出把 ν 换成 $\nu+1$ 时相应的结果。我们回顾前一证明中的不等式 (3.2)；将其中的 $S_k(f)$ 换成 $T(\alpha)$ ，它表明

$$|T(\alpha)|^{2\nu} \ll P^{2\nu-1} + P^{2\nu-\nu-1} \Re \sum_{y_1=1}^P \cdots \sum_{y_\nu=1}^P S_{k-\nu},$$

其中

$$S_{k-\nu} = \sum_x e(\alpha \Delta_{y_1, \dots, y_\nu}(x^k)).$$

补全：为什么这里要用 (3.2) 而不是 (3.1)

1 把引理 3.1 证明里的一般多项式 f 取成 $f(x) = \alpha x^k$ 、 $S_k(f) = T(\alpha)$ 。
(3.2) 给出 $|T|^{2\nu}$ 的上界，关键是它保留了 $\Re S_{k-\nu}$ (带正负号)，**没有取绝对值**。

2 **为什么必须保留正负号？** 下一步要把这个不等式乘 $|T|^{2\nu}$ 再积分。 $\Re S_{k-\nu}$ 里的 $e(\alpha \Delta(x^k))$ 与 $|T|^{2\nu} = T^{2\nu-1} \bar{T}^{2\nu-1}$ 里的指数会**合并成一个总指数**，再用正交关系变成“某方程解数”。若提前取绝对值，指数结构被破坏，正交关系就用不上了。这就是第 4 节末特意留下 (3.2) 的原因——伏笔在此回收。

- 3 这里 $S_{k-\nu} = \sum_x e(\alpha \Delta_{y_1, \dots, y_\nu}(x^k))$ 是对 x^k 做 ν 次差分后 ($k - \nu$ 次多项式) 的指数和, x 取值范围依赖各 y_i 但含于 $[1, P]$ 。

把不等式两边都乘以 $|T(\alpha)|^{2^\nu}$, 并从 0 积到 1, 我们得到

$$I_{\nu+1} \ll P^{2^\nu-1} I_\nu + P^{2^\nu-\nu-1} \sum_{y_1, \dots, y_\nu} \Re \int_0^1 S_{k-\nu} |T|^{2^\nu} d\alpha.$$

补全: 两边乘 $|T|^{2^\nu}$ 再积分

- 1 左边: $|T|^{2^\nu} \cdot |T|^{2^\nu} = |T|^{2^{\nu+1}}$, 积分得 $\int_0^1 |T|^{2^{\nu+1}} d\alpha = I_{\nu+1}$ 。这正是我们想要估计的目标。

- 2 右边第一项: $P^{2^\nu-1} \cdot |T|^{2^\nu}$ 积分得 $P^{2^\nu-1} \int_0^1 |T|^{2^\nu} d\alpha = P^{2^\nu-1} I_\nu$ 。这里就要用到归纳假设来控制 I_ν 。

- 3 右边第二项: $P^{2^\nu-\nu-1} \sum_{y_1, \dots, y_\nu} \Re \int_0^1 S_{k-\nu} |T|^{2^\nu} d\alpha$ (积分、求和、取实部都是线性运算, 可交换次序)。剩下的工作就是搞清楚那个积分 $\int_0^1 S_{k-\nu} |T|^{2^\nu} d\alpha$ 到底是什么。

最后这个积分为

$$\int_0^1 \sum_x e(\alpha \Delta_{y_1, \dots, y_\nu}(x^k)) \sum_{\substack{u_1, \dots, u_{2^\nu-1} \\ v_1, \dots, v_{2^\nu-1}}} e(\alpha u_1^k + \dots) e(-\alpha v_1^k - \dots) d\alpha,$$

其中 u_i 与 v_i 从 1 取到 P 。该积分等于

$$\Delta_{y_1, \dots, y_\nu}(x^k) + u_1^k + \dots - v_1^k - \dots = 0 \quad (3.5)$$

的解的个数。

补全: 把 $|T|^{2^\nu}$ 展开, 用正交关系数解

1 展开 $|T|^{2^\nu}$ 。 $|T|^{2^\nu} = T^{2^{\nu-1}} \cdot \bar{T}^{2^{\nu-1}}$ 。 $T^{2^{\nu-1}} = \left(\sum_u e(\alpha u^k)\right)^{2^{\nu-1}}$ 展开成对 $u_1, \dots, u_{2^{\nu-1}}$ 的和，指数是 $\alpha(u_1^k + \dots + u_{2^{\nu-1}}^k)$ ；同理 $\bar{T}^{2^{\nu-1}}$ 给出 $-\alpha(v_1^k + \dots + v_{2^{\nu-1}}^k)$ 。一共 $2^{\nu-1} + 2^{\nu-1} = 2^\nu$ 个变量，正好是 $|T|^{2^\nu}$ 里的 2^ν 次幂。

2 合并所有指数。再乘上 $S_{k-\nu} = \sum_x e(\alpha \Delta(x^k))$ ，所有 $e(\dots)$ 的指数相加，得 $e(\alpha [\Delta_{y_1, \dots, y_\nu}(x^k) + u_1^k + \dots - v_1^k - \dots])$ 。

3 对 α 积分，用正交关系。记方括号里的整数为 n 。 $\int_0^1 e(\alpha n) d\alpha = 1$ 当 $n = 0$ ，否则 0。于是积分把每组 (x, u_i, v_i) “筛选”：只有让 $n = 0$ （即方程 (3.5) 成立）的那些组才计 1，其余计 0。

4 所以这个积分 = 方程 (3.5) 的解的个数（在给定 $y_1, \dots, y_\nu$ 下，对 x, u_i, v_i 数）。再对 $y_1, \dots, y_\nu$ 求和，就得到所有变量上的总解数 N 。注意：解数是非负整数，所以这个量本身是实数， $\Re$ 是多余的、只是形式上保留。

对 $y_1, \dots, y_\nu$ 求和，便给出在所有变量上的解的总数。因此

$$I_{\nu+1} \ll P^{2^\nu-1} I_\nu + P^{2^\nu-\nu-1} N, \quad (3.6)$$

其中 N 表示 (3.5) 在所有变量上的解的个数，这些变量现在都是 $[1, P]$ 中的整数。

解读：到这一步，分析（积分）已经彻底变成组合（数解数）。下一段就专心去估计 N 。

现在重要的是要注意到：由于 $y_1, \dots, y_\nu$ 与 x 都是正的，我们有

$$\Delta_{y_1, \dots, y_\nu}(x^k) > 0.$$

此外，这个数能被 $y_1, \dots, y_\nu$ 中的每一个整除。因此，若我们给定 $u_1, \dots, u_{2^{\nu-1}}$ 与 $v_1, \dots, v_{2^{\nu-1}}$ 任意值，则由 (3.3)， $y_1, \dots, y_\nu$ 中每一个

的可能取法数为 $\ll P^\varepsilon$ 。于是 x 至多有一种可能取法，因为 $\Delta_{y_1, \dots, y_\nu}(x^k)$ 是 x 的严格递增函数。 u_i 与 v_i 的可能取法数为 $\ll P^{2^\nu}$ ，由此可知

$$N \ll P^{2^\nu + \nu\varepsilon}.$$

补全：怎样把解数 N 数清楚（每个论断都验证）

1 论断 A: $\Delta_{y_1, \dots, y_\nu}(x^k) > 0$ 。因为 x^k 是关于 x 严格递增且“高阶差分为正”的函数（对 $\nu \leq k-1$ ，第 ν 阶差分仍是 x 的次数 ≥ 1 的多项式，最高次系数 $k(k-1)\cdots(k-\nu+1)y_1\cdots y_\nu > 0$ ，在正整数处取正值）。直观：步长 $y_i > 0$ 时 $f(x+y) - f(x) > 0$ 对递增的 f 成立，逐次差分保持正。

2 论断 B: $\Delta_{y_1, \dots, y_\nu}(x^k)$ 被每个 y_i 整除。看一次差分 $\Delta_y g(x) = g(x+y) - g(x)$ 。对整系数多项式（如 x^k 及其整数差分）， $g(x+y) \equiv g(x) \pmod{y}$ （因为 $(x+y)^j \equiv x^j \pmod{y}$ ），所以 $\Delta_y g(x)$ 能被 y 整除。对每一步差分用此结论，得 $\Delta_{y_1, \dots, y_\nu}(x^k)$ 被 $y_1, \dots, y_\nu$ 中每一个整除。

3 由 A、B 控制 y_i 的取法。先随便固定 u_i, v_i （共 2^ν 个变量，各 $\leq P$ ，取法 $\ll P^{2^\nu}$ 种）。方程 (3.5) 要求 $\Delta_{y_1, \dots, y_\nu}(x^k) = v_1^k + \cdots - u_1^k - \cdots =: M$ ，一个由 u, v 决定的固定数。由 A 这个 $M > 0$ （在有解时），由 B 每个 y_i 都整除 M 。所以每个 y_i 只能取 M 的因子 $\rightarrow$ 取法 $\leq d(M) \ll M^\varepsilon \ll P^\varepsilon$ （用 (3.3)，且 $M \leq$ 若干个 P^k 之和 $\ll P^k$ ，故 $M^\varepsilon \ll P^{k\varepsilon} \ll P^\varepsilon$ ）。 ν 个 y_i 合计 $\ll P^{\nu\varepsilon}$ 种。

4 x 至多一种。 y_i 与 M 都定了之后， $\Delta_{y_1, \dots, y_\nu}(x^k) = M$ 是关于 x 的方程；因为左边是 x 的严格递增函数（论断 A 的加强：导数恒正，注意

$\nu \leq k-1$ 保证它真的还含 x 、不是常数), 严格递增函数取定值至多一个解。所以 x 至多 1 种。

5 汇总。

$N \leq (u, v \text{ 取法}) \times (y \text{ 取法}) \times (x \text{ 取法}) \ll P^{2^\nu} \cdot P^{\nu\varepsilon} \cdot 1 = P^{2^\nu + \nu\varepsilon}$ 。这就是所要的 N 估计。妙处在于：本来 y_i 和 x 也各能跑 P 个值（朴素地 N 可达 $P^{2^\nu + \nu + 1}$ ），但“被整除”和“严格单调”这两条把它们几乎钉死，省下了 $P^{\nu+1}$ 量级——这正是华罗庚的关键观察。

代入 (3.6) 并利用归纳假设，我们得到

$$I_{\nu+1} \ll P^{2^\nu-1} P^{2^\nu-\nu+\varepsilon} + P^{2^\nu-\nu-1} P^{2^\nu+\nu\varepsilon} \ll P^{2^{\nu+1}-(\nu+1)+\nu\varepsilon}.$$

这正是把 ν 换成 $\nu+1$ 后的 (3.4)，只是 ε 有所改变，而这无关紧要。

补全：最后的幂次合并

1 第一项。归纳假设 $I_\nu \ll P^{2^\nu-\nu+\varepsilon}$ ，乘 $P^{2^\nu-1}$ ：指数相加

$$(2^\nu - 1) + (2^\nu - \nu + \varepsilon) = 2 \cdot 2^\nu - \nu - 1 + \varepsilon = 2^{\nu+1} - (\nu + 1) + \varepsilon$$

。

2 第二项。 $P^{2^\nu-\nu-1} \cdot N \ll P^{2^\nu-\nu-1} \cdot P^{2^\nu+\nu\varepsilon}$ ：指数相加

$$(2^\nu - \nu - 1) + (2^\nu + \nu\varepsilon) = 2^{\nu+1} - \nu - 1 + \nu\varepsilon = 2^{\nu+1} - (\nu + 1) + \nu\varepsilon$$

。

3 两项同阶。两个指数都是 $2^{\nu+1} - (\nu + 1)$ 再加一点 ε 级的项。取较大者（把 ε 与 $\nu\varepsilon$ 统一重记为 ε ，因 $\nu \leq k$ 固定）：

$$I_{\nu+1} \ll P^{2^{\nu+1}-(\nu+1)+\varepsilon}.$$

这恰是 (3.4) 中 $\nu \rightarrow \nu+1$ 的形式。

4 归纳收尾。基础 $\nu = 1$ 已验证，归纳步骤成立，故 (3.4) 对 $\nu = 1, \dots, k$ 都成立。取 $\nu = k$: $I_k = \int_0^1 |T|^{2^k} d\alpha \ll P^{2^k - k + \varepsilon}$ ，正是引理 3.2。■

5 “ ε 改变无关紧要”是什么意思？每一步归纳里 ε 会增大一点点（比如从 ε 变 2ε ）。但因为总共只做 k 步（ k 固定），最终的 ε 仍是某个固定小数。换句话说：先决定最终想要的 ε ，再倒推每步用更小的 ε/k 即可。这是解析数论里处理 ε 的标准说辞。

第 11 节 注： $k = 3$ （立方）时的含义与“八个立方数”悬念

这一段在讲什么：专门考察 $k = 3$ 时华罗庚不等式说了什么。引入一个函数 $\lambda(m)$ 记录各次幂均值的最优指数，证明它是凸的，并指出一个至今未解决的问题——若能在中间改进一点，就能把“九个立方数”改进成“八个立方数”。

注. 考察当 $k = 3$ 时引理 3.2 所给出的信息是颇有意思的。令 $\lambda(m)$ 记使得

$$\int_0^1 |T(\alpha)|^{2^m} d\alpha \ll P^\lambda$$

成立的指数 λ 的下确界。由 Cauchy 不等式可知

$$\lambda\left(\frac{m_1 + m_2}{2}\right) \leq \frac{1}{2}(\lambda(m_1) + \lambda(m_2)),$$

因此 $\lambda(m)$ 作为 m 的函数是凸的。

符号 $\lambda(m)$ ：最优均值指数

对每个实数 $m > 0$ ，问“ $\int_0^1 |T|^{2^m} d\alpha$ 最小能被 P^λ 控制时， λ 取多小”，这个最小的 λ （下确界）记为 $\lambda(m)$ 。它把“各种次幂的均值估计”浓缩成一条曲线，方

便比较好坏。凸是指：图像上任意两点的连线段都在图像上方（像 $y = x^2$ 那样“向上弯”）。

补全：用 Cauchy 不等式证 λ 的凸性

- 1 对积分形式的 Cauchy-Schwarz: $\int fg \leq (\int f^2)^{1/2} (\int g^2)^{1/2}$ 。取 $f = |T|^{m_1}$, $g = |T|^{m_2}$, 则 $fg = |T|^{m_1+m_2}$, $f^2 = |T|^{2m_1}$, $g^2 = |T|^{2m_2}$:

$$\int_0^1 |T|^{m_1+m_2} d\alpha \leq \left(\int_0^1 |T|^{2m_1} d\alpha \right)^{1/2} \left(\int_0^1 |T|^{2m_2} d\alpha \right)^{1/2}.$$

- 2 把右边两个积分用各自的最优界 $\ll P^{\lambda(m_1)}$ 、 $\ll P^{\lambda(m_2)}$ 代入:

$$\int_0^1 |T|^{2 \cdot \frac{m_1+m_2}{2}} d\alpha \ll (P^{\lambda(m_1)})^{1/2} (P^{\lambda(m_2)})^{1/2} = P^{\frac{1}{2}(\lambda(m_1)+\lambda(m_2))}.$$

注意左边正是“次幂参数为 $\frac{m_1+m_2}{2}$ ”的均值。

- 3 按 λ 的定义（最小可行指数），它 $\leq$ 上面这个可行的指数 $\frac{1}{2}(\lambda(m_1) + \lambda(m_2))$ 。即

$$\lambda\left(\frac{m_1 + m_2}{2}\right) \leq \frac{1}{2}(\lambda(m_1) + \lambda(m_2)).$$

这正是“中点处的值不超过两端值的平均”，即凸性。

引理 3.2 告诉我们

$$\lambda(1) \leq 1, \quad \lambda(2) \leq 2, \quad \lambda(4) \leq 5,$$

并且可以证明这三处实际上都取等号。于是该图像落在连接 $(1, 1), (2, 2), (4, 5)$ 的两条线段之上或之下。

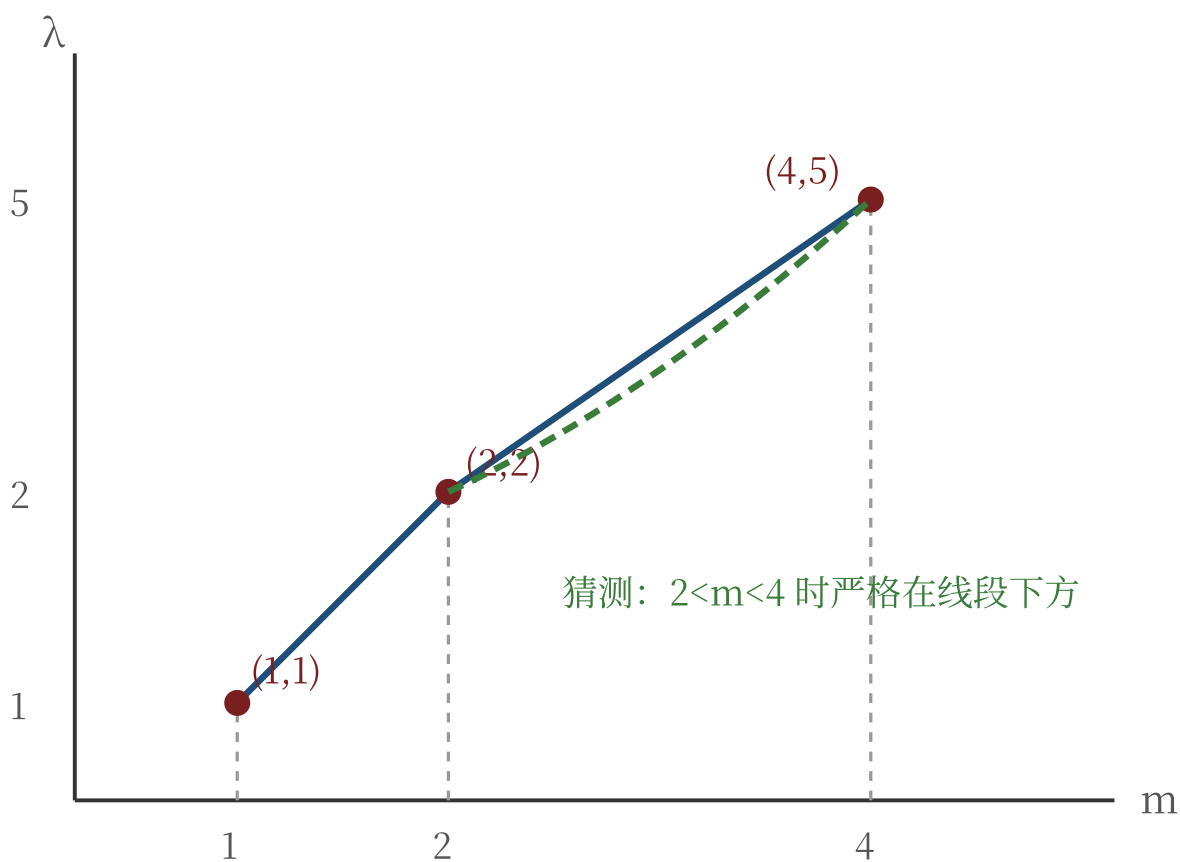
补全：这三个值从哪来

1 $\lambda(1) \leq 1$: $m = 1$ 即 $\int_0^1 |T|^2 d\alpha = I_1 = P = P^1$ (前面算过, 等于 P), 所以 $\lambda(1) \leq 1$ 。

2 $\lambda(2) \leq 2$: $m = 2$ 即 $2m = 4 = 2^2$, 正是引理 3.2 取 $k = 2$? 不——这里 $k = 3$ 固定, $2m = 4 = 2^\nu$ 对应 $\nu = 2$ 。由 (3.4) 的 $\nu = 2$ 情形 $I_2 \ll P^{2^2-2+\varepsilon} = P^{2+\varepsilon}$, 故 $\lambda(2) \leq 2$ 。

3 $\lambda(4) \leq 5$: $2m = 8 = 2^3 = 2^k$ ($k = 3$), 正是引理 3.2 本身: $I_3 \ll P^{2^3-3+\varepsilon} = P^{5+\varepsilon}$, 故 $\lambda(4) \leq 5$ 。

4 取等号 (即这些指数无法再降) 可由“数解数的下界”证明: 例如 $x_1 = y_1$ 这种平凡解就贡献了 $P^{\text{对应幂}}$ 量级, 使指数不能更小。于是三点 $(1, 1), (2, 2), (4, 5)$ 都精确落在曲线上。由凸性, 曲线在 $1 \leq m \leq 4$ 间夹在“两条连线段”与“图像本身”之间。



$k = 3$ 时 $\lambda(m)$ 的已知三点 $(1, 1)$, $(2, 2)$, $(4, 5)$ 与连接它们的折线。人们猜测在 $(2, 2)$

看来当 $2 < m < 4$ 时该图像很可能严格地位于线段之下，但这一点从未被证明。如果能证明它，便能够建立八个立方数（而非九个立方数， $9 = 2^k + 1$ ）的渐近公式。例如，只要能证明

$$\int_0^1 |T(\alpha)|^6 d\alpha \ll P^{7/2-\delta}$$

对某个正数 δ 成立就够了。这等价于断言：

$$x_1^3 + x_2^3 + x_3^3 = y_1^3 + y_2^3 + y_3^3$$

（所有变量介于 0 与 P 之间）的解的总数为 $\ll P^{7/2-\delta}$ 。

补全：为什么 $\int |T|^6 \ll P^{7/2-\delta}$ 能换来“八个立方数”

1 $|T|^6$ 对应 $m = 3$ (在 2 与 4 之间)。折线在 $m = 3$ 处的值是 $(2, 2)$ 与 $(4, 5)$ 的中点纵坐标 $\frac{2+5}{2} = 3.5 = \frac{7}{2}$ 。所以折线给的是 $\lambda(3) \leq \frac{7}{2}$, 即 $\int |T|^6 \ll P^{7/2}$ 。若能改进到 $P^{7/2-\delta}$ (严格低于折线), 就证实了“中间严格在下”的猜测。

2 积分=立方和方程解数。由前面“积分=数解数”的原理, $\int_0^1 |T|^6 d\alpha$ ($= \int_0^1 T^3 \bar{T}^3$) 恰好等于

$$x_1^3 + x_2^3 + x_3^3 = y_1^3 + y_2^3 + y_3^3, \quad 0 \leq x_i, y_i \leq P$$

的解的总数。所以“积分有更好上界” $\iff$ “这个对称立方方程解不太多”。这就是文末那句等价表述。

3 九个 vs 八个立方数。 $9 = 2^k + 1$ ($k = 3$) 个立方数的渐近公式可由现有的 $2^k = 8$ 次幂均值 (即引理 3.2) 直接得到——这是华罗庚不等式的标准应用。要把所需个数减到 8, 需要一个“半个变量的节省”, 恰好就是把 $|T|^6$ 的均值压低一点点 ($\delta > 0$)。难点: 这要求精确地知道立方和方程的解数没有“反常聚集”, 目前数学家既无法证明也无法否定——这是立方华林问题里著名的未决问题, 凸性方法到此为止, 再进一步需要全新的想法。

4 取舍说明: Davenport 在此坦诚地标出“方法的边界”——凸性 (Cauchy 不等式) 只能给到折线, 要突破必须引入更强的工具。这种对“已知/未知边界”的清楚交代, 是好的数学叙述的标志。

第 12 节 脚注: 关于 $\ll$ 记号

我们使用 Vinogradov 记号 $\ll$ 来表示一个带有未指定“常数”因子的不等式。在当前情形中, 所出现的因子实际上与 k 无关, 但我们无需了解这一点。

解读：再次强调 $\ll$ 隐含一个常数 C （见第 0 节）。这里补充一个细节：本章里那个常数其实连 k 都不依赖（只依赖 ε ），但作者说“我们不需要知道这一点”——因为在应用中 k 是固定的，常数依不依赖 k 都无所谓。这体现了 $\ll$ 记号的便利：让我们把注意力集中在量级（ P 的幂次）上，不被常数分心。

本章小结：两件武器各管什么

- **Weyl 不等式（点估计）：**给定一个 α ，若它的有理逼近分母 q 在 $[P, P^{k-1}]$ 间，则 $|\sum e(f(x))| \ll P^{1-1/K+\varepsilon}$ ，比平凡的 P 小。用途：在圆法的**次弧**上逐点压低指数和。证明套路：平方降次（ $k-1$ 次） $\rightarrow$ 几何级数 $\rightarrow$ 除数函数合并 $\rightarrow$ 分块用有理逼近。
- **华罗庚不等式（均值估计）：** $\int_0^1 |T|^{2^k} d\alpha \ll P^{2^k-k+\varepsilon}$ ，比平凡的 P^{2^k} 小 $P^{k-\varepsilon}$ 。用途：在**整段** $[0, 1]$ 上控制高次幂的平均，是次弧贡献可忽略的依据。证明套路：积分=数解数 $\rightarrow$ 用差分的“被整除+单调”狠狠压制解数 $\rightarrow$ 对 ν 归纳。
- **共同内核：**两者都建立在“Weyl 平方降次”这一发动机上，区别只在最后一步取不取绝对值（(3.1) vs (3.2)）。把握住这个内核，整章就连成了一条线。

返回 [全书目录](#)