

DAVENPORT · 圆法 · 高中详解版

奇异级数（续）

The singular series continued

本章要解决什么 上一章我们把"奇异级数 $\mathfrak{S}(N)$ "这个对象搭起来了，但留下了一个欠条：它到底收不收敛、收敛得快不快，全押在一个估计上——指数和 $S_{a,q}$ （也叫完整 Weyl 和）究竟有多大。本章就专门去把这张欠条还清：证明

$$|S_{a,q}| \ll q^{1-1/k},$$

并由此推出奇异级数在 $s \geq 2k + 1$ 时绝对收敛、而且有正的下界 $\mathfrak{S}(N) \geq C_1 > 0$ 。读完你将掌握三样硬功夫：(1) 用**特征与高斯和**把模素数 p 的指数和拆开估计；(2) 用**"提升一位数字"**的技巧把模 p^ν 的和层层降阶；(3) 用 $S_{a,q}$ 的**乘性**把所有素数的估计拼成对一般 q 的整体估计。最后还会看到 Hardy-Littlewood 关于 $\mathfrak{S}(N)$ 随 N 怎样波动的一段精彩史料。

本页为译文 + 高中详解：灰底"原文"框是对 Davenport 原书该段的忠实翻译；其余彩色框、配图、分步推演是面向高中生的逐句解读，凡原文写"显然""容易看出"的地方，这里都补成一步一步的完整推导。

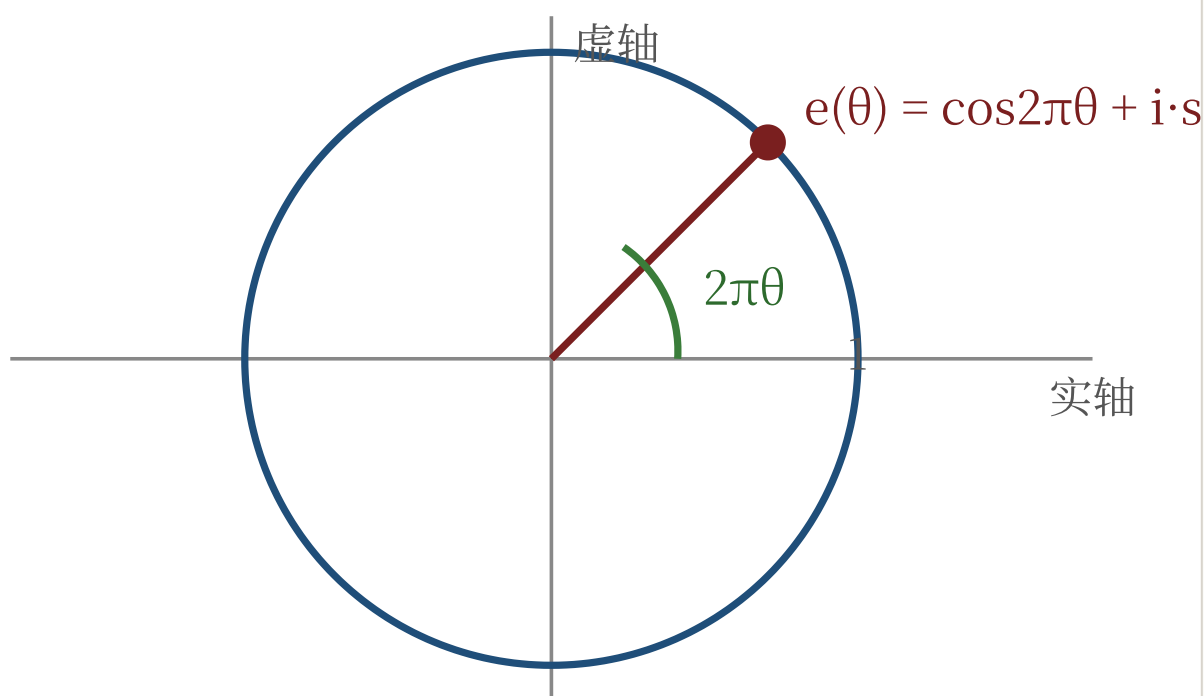
预备：先把本章要用到的符号一次讲清

本章公式里反复出现几个"超纲"记号。高中没学过没关系，下面从零讲，往后遇到就回来查。

记号 $e(\theta)$ (复指数) 约定 $e(\theta) = e^{2\pi i \theta}$, 读作" e 括号 θ ". 这里 i 是虚数单位 ($i^2 = -1$), $e^{i\varphi} = \cos \varphi + i \sin \varphi$ (欧拉公式)。所以

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

它是复平面单位圆上的一个点: 把整个圆周长看成 1, θ 就是"转了几圈"。关键性质: (i) $|e(\theta)| = 1$ (永远在单位圆上, 模长为 1); (ii) $e(\theta)$ 只依赖 θ 的小数部分, 即 $e(\theta + m) = e(\theta)$ 对任意整数 m 成立 (转整数圈等于没转); (iii) $e(\theta_1)e(\theta_2) = e(\theta_1 + \theta_2)$ (指数相加)。



把 $e(\theta)$ 想成"单位圆上转 θ 圈到达的点": 模长恒为 1, 只有方向 (相位角 $2\pi\theta$) 在变。圆法的全部技巧都是在追踪这些点相加时的相互抵消。

记号 $\sum$ (求和号) 与 $\sum_x$ $\sum$ 读作"sigma", 是"把一串数加起来"的缩写。例如 $\sum_{x=0}^{p-1} f(x) = f(0) + f(1) + \cdots + f(p-1)$ 。当下标只写 $\sum_x$ 没标范围时, 按上下文它表示" x 跑遍模 p (或模 q) 的一组完全剩余系", 也就是 $x = 0, 1, 2, \dots, p-1$ 各取一次。

记号 $a \equiv b \pmod{m}$ (同余) 读作" a 与 b 模 m 同余", 意思是 $m \mid (a - b)$, 即 a, b 除以 m 余数相同。例如 $17 \equiv 2 \pmod{5}$ 。 $a \not\equiv 0 \pmod{p}$ 就是" p 不整除 a "。 (a, q) 表示 a 与 q 的最大公约数; $(a, q) = 1$ 即"互素"。 $p \nmid k$ 读作" p 不整除 k "。

记号 $\ll$ 与 $O(\cdot)$ (数量级) $f \ll g$ (等价写法 $f = O(g)$) 读作" f 的量级被 g 控制", 确切含义是: 存在一个与变量无关的常数 C , 使得 $|f| \leq Cg$ 永远成立。它只关心"大约多大", 不关心常数。例如 $3n + 5 \ll n$, $100 \ll 1$ 。本章里 $\ll$ 的常数允许依赖 k, s , 但不依赖 q 或 N ——这正是后面"界与 q 无关"那句话的要害。

记号 $S_{a,q}$ (完整指数和 / Weyl 和) 本书第 5 章的定义:

$$S_{a,q} = \sum_{x=0}^{q-1} e\left(\frac{a}{q}x^k\right).$$

读法: 让 x 跑遍 0 到 $q-1$, 对每个 x 算出单位圆上的点 $e(ax^k/q)$, 再把这 q 个点 (复数) 加起来。 k 是 Waring 问题里的"次数" (要把 N 写成 s 个 k 次方之和)。粗暴地看, q 个模长为 1 的数相加, 最大也就 q (三角不等式); 但这些点方向乱转、彼此抵消, 真实大小要小得多。本章的核心任务就是把"小得多"精确成 $\ll q^{1-1/k}$ 。

记号 $A(q)$ 与奇异级数 $\mathfrak{S}(N)$ 第 5 章定义了

$$A(q) = \sum_{\substack{a=1 \\ (a,q)=1}}^q \left(\frac{S_{a,q}}{q}\right)^s e\left(\frac{-Na}{q}\right), \quad \mathfrak{S}(N) = \sum_{q=1}^{\infty} A(q).$$

$\mathfrak{S}(N)$ (花体 S, "奇异级数"singular series) 是圆法主项里那个只跟"模各种数的同余解多不多"有关的因子; 它 > 0 是" N 能表成 s 个 k 次方和"在**每个素数局部**都没有障碍的体现。 $A(q)$ 是它的第 q 项。要谈 $\mathfrak{S}(N)$ 收不收敛, 先得知道

$A(q)$ 随 q 增大衰减得有多快——而这又取决于 $S_{a,q}$ 有多大。一切都归到 $S_{a,q}$ 的估计上，这就是本章的来由。

记号 φ (欧拉函数)、特征 χ 、原根 欧拉函数 $\varphi(m)$ = "1 到 m 中与 m 互素的个数", 如 $\varphi(p) = p - 1$ 。

原根: 模素数 p 时, 存在一个 g (叫原根) 使 $g^1, g^2, \dots, g^{p-1}$ 恰好跑遍所有非零剩余 $1, 2, \dots, p-1$ (各一次)。这说明"非零剩余在乘法下"像一个长度 $p-1$ 的循环跑道。

(乘法) **特征 χ** : 把每个非零剩余 t 对应到一个模长为 1 的复数 $\chi(t)$, 满足 $\chi(t_1 t_2) = \chi(t_1) \chi(t_2)$ (乘法保持), 并约定 $\chi(0) = 0$ 。用原根写就是: 取定一个 $p-1$ 次单位根 ζ , 令 $\chi(g^y) = \zeta^y$ 。"主特征"是处处取 1 (除 0 外) 的那个 χ_0 ; 非主特征就是别的。特征的灵魂性质 (求和正交):

$$\sum_t \chi(t) = 0 \quad (\chi \text{ 非主}), \quad \sum_t \chi_0(t) = p - 1.$$

本章只需把 χ 当成"一种带乘法的、会自动抵消的权重", 下文用到时再细讲。

一、开场：把欠条点清

原文 我们现在来证明在第 3.1 引理的推论中提到的那个结果, 即

$$|S_{a,q}| \ll q^{1-1/k}. \quad (6.1)$$

这一段交代本章的总目标。"引理 3.1 的推论"里曾经用过这个估计 (在 Weyl 不等式那一章), 但当时是对一般的不完整和顺手提了一句; 现在要对完整和 $S_{a,q}$ 给出干净、独立的证明。先体会一下 (6.1) 说了什么:

- 平凡上界 (什么都不想, 直接三角不等式) 只能给 $|S_{a,q}| \leq q$ 。

- (6.1) 把它压到 $q^{1-1/k}$ 。注意 $1 - 1/k < 1$ ，所以这是实打实地省下了一个 $q^{1/k}$ 因子。 k 越大省得越少，但永远 \(\

数量感 取 $k = 3$, $q = 10^6$ 。平凡界 10^6 ；(6.1) 给

$\ll (10^6)^{1-1/3} = (10^6)^{2/3} = 10^4$ 。从一百万压到一万，差了 100 倍。这就是"抵消"的威力——一百万个单位圆上的点加起来，方向乱转，结果只有一万量级。

原文 这蕴含了

$$|A(q)| \ll q^{1-s/k},$$

由此可知，若 $s \geq 2k + 1$ ，则奇异级数绝对收敛。第 5 章的 (5.8) 以及引理 5.2 的推论，在同一条件下也都成立。

这一段说明为什么 (6.1) 值得费这么大劲去证：它一步推出整条奇异级数的收敛。我们把这个"蕴含"补成完整推导（原文跳过了）。

1 回到 $A(q) = \sum_{\substack{a=1 \\ (a,q)=1}}^q (S_{a,q}/q)^s e(-Na/q)$ 。先用三角不等式把求和拆

掉、把模长为 1 的 $e(-Na/q)$ 丢掉：

$$|A(q)| \leq \sum_{\substack{a=1 \\ (a,q)=1}}^q \left| \frac{S_{a,q}}{q} \right|^s \cdot \underbrace{|e(-Na/q)|}_{=1}.$$

"丢掉 e 因子"为什么合法？因为 $|e(\theta)| = 1$ ，乘上它不改变模长（预备里的性质 (i)）。

2 对每个满足 $(a, q) = 1$ 的 a ，由 (6.1) 有 $|S_{a,q}| \ll q^{1-1/k}$ ，于是

$$\left| \frac{S_{a,q}}{q} \right|^s \ll \left(\frac{q^{1-1/k}}{q} \right)^s = (q^{-1/k})^s = q^{-s/k}.$$

这里用了 $q^{1-1/k}/q = q^{(1-1/k)-1} = q^{-1/k}$ (同底数幂相除, 指数相减), 再整体 s 次方 (幂的幂, 指数相乘)。

- 3 求和的项数: 满足 $(a, q) = 1$ 、 $1 \leq a \leq q$ 的 a 共有 $\varphi(q)$ 个, 而 $\varphi(q) \leq q$ (互素的不可能比总数多)。把每项的界 $q^{-s/k}$ 乘以项数上界 q :

$$|A(q)| \ll q \cdot q^{-s/k} = q^{1-s/k}.$$

这正是原文那个 $|A(q)| \ll q^{1-s/k}$ 。

- 4 现在看收敛。"绝对收敛"指 $\sum_q |A(q)|$ 是有限数。由上一步,

$$\sum_{q=1}^{\infty} |A(q)| \ll \sum_{q=1}^{\infty} q^{1-s/k} = \sum_{q=1}^{\infty} \frac{1}{q^{s/k-1}}.$$

这是一个 p -级数 ($\sum 1/q^\sigma$ 型), 它收敛当且仅当指数 $\sigma = s/k - 1 > 1$, 即 $s/k > 2$, 即 $s > 2k$ 。因为 s 是整数, " $s > 2k$ "就是" $s \geq 2k + 1$ "。这就解释了条件 $s \geq 2k + 1$ 是从哪冒出来的: 它恰好让 p -级数的指数迈过 1 这道收敛门槛。

补: 为什么 $\sum 1/q^\sigma$ 在 $\sigma > 1$ 才收敛 高中没正式讲过。直觉证法 (积分比较): $\frac{1}{q^\sigma}$ 是减函数, 把 $\sum_{q \geq 2} \frac{1}{q^\sigma}$ 与 $\int_1^\infty \frac{dx}{x^\sigma}$ 比较, 二者同生死。而 $\int_1^\infty x^{-\sigma} dx = \left[\frac{x^{1-\sigma}}{1-\sigma} \right]_1^\infty$ 在 $\sigma > 1$ 时 $x^{1-\sigma} \rightarrow 0$, 积分有限; $\sigma \leq 1$ 时发散。所以分界点正是 $\sigma = 1$ 。这里 $\sigma = s/k - 1$, 要 > 1 。

最后一句"(5.8) 与引理 5.2 推论也在同条件下成立", 意思是: 第 5 章里凡是临时假设了 $S_{a,q}$ 的衰减才得到的结论, 现在 (6.1) 一旦证成, 那些结论就无条件兑现

了。它们都共用同一张欠条，本章一并还清。

二、引理 6.1：模素数 p 的和——用特征和高斯和拿下

原文·引理 6.1 若 $a \not\equiv 0 \pmod{p}$ 且 $\delta = (k, p-1)$ ，则

$$|S_{a,p}| \leq (\delta - 1)p^{1/2}. \quad (6.2)$$

这是整章的基石：先把模数是素数 p 的最简单情形彻底搞定，后面模 p^ν 、模一般 q 都靠它。先读懂记号。

记号 $\delta = (k, p-1)$ δ 是 k 与 $p-1$ 的最大公约数。为什么偏偏是它？因为模 p 的非零剩余在乘法下是个"长度 $p-1$ 的循环跑道"（原根），在跑道上做" k 次方"等价于"在跑道上走 k 步的倍数"，而这种步法实际能落到的格点数恰由 $\gcd(k, p-1) = \delta$ 决定。所以 k 次方映射的"真实复杂度"是 δ 而非 k 。下面第一步就把这件事说透。

先看一眼 (6.2) 多有用 它把 $|S_{a,p}|$ 压到 $\leq (\delta - 1)\sqrt{p}$ 。因为 $\delta \leq k$ ，所以 $|S_{a,p}| \leq (k - 1)\sqrt{p} \ll \sqrt{p}$ ，比平凡界 p 小了个 $\sqrt{p}$ 。这正是 $\sqrt{p}$ 这种"平方根抵消"，是高斯和的招牌。

原文·证明（开头） 由于 $x^k \equiv m \pmod{p}$ 与 $x^\delta \equiv m \pmod{p}$ 具有相同个数的解，我们有

$$S_{a,p} = \sum_x e\left(\frac{a}{p}x^\delta\right).$$

第一步：把次数从 k 换成 δ 。这是关键的化简，原文一句"具有相同个数的解"带过，我们补全。

1 先把 $S_{a,p}$ 按 " x^k 取到的值 m " 分组。设

$N_k(m) = \#\{x \bmod p : x^k \equiv m\}$ (即 k 次方等于 m 的 x 的个数)。
则

$$S_{a,p} = \sum_{x=0}^{p-1} e\left(\frac{a}{p} x^k\right) = \sum_{m=0}^{p-1} N_k(m) e\left(\frac{a}{p} m\right).$$

道理：把求和按 x^k 的值 m 归类，凡 $x^k \equiv m$ 的 x 都贡献同一个 $e(am/p)$ ，这样的 x 有 $N_k(m)$ 个。

2 断言 $N_k(m) = N_\delta(m)$ 对每个 m 成立 (这就是"相同个数的解")。取模 p 的原根 g ，非零的 x 写成 $x = g^y$ ($y = 0, \dots, p-2$)。则 $x^k = g^{ky}$ 。所以 x^k 跑遍的集合 $= \{g^{ky}\} =$ "跑道上以步长 k 走"落到的点。由群论基本事实： $\{g^{ky} \bmod (p-1) \text{ 的指数}\} = \{ky \bmod (p-1)\}$ 这个子群恰等于 $\{\delta y \bmod (p-1)\}$ ，因为 $\gcd(k, p-1) = \gcd(\delta, p-1) = \delta$ (δ 本就整除 $p-1$)。换言之，" k 次方"和" δ 次方"取到的值集完全一样，而且每个值被取到的次数也一样。再补上 $x = 0$ (两边都只贡献 $m = 0$ 一次)。故 $N_k(m) = N_\delta(m)$ 。

3 把 $N_k(m)$ 换成 $N_\delta(m)$ 代回第 1 步，再把分组还原成对 x 的求和：

$$S_{a,p} = \sum_m N_\delta(m) e\left(\frac{a}{p} m\right) = \sum_x e\left(\frac{a}{p} x^\delta\right).$$

于是把次数 k 合法地换成了它的"有效次数" δ 。好处： $\delta \mid (p-1)$ ，下面能直接用阶为 δ 的特征。

非零剩余的乘法跑道（长度 $p-1$ ），原根 g 把它排成一圈：



" k 次方" = 步长 k 走 $\rightarrow$ 实际落点由 $\gcd(k, p-1)=\delta$ 决定

所以把 k 换成 δ ，指数和一字不差： $x^k \equiv m$ 与 $x^\delta \equiv m$ 解数相同

" k 次方"在乘法循环群上的效果，只由 $\gcd(k, p-1) = \delta$ 决定。这就是为什么可以把次数从 k 降到 δ 。

原文·证明（续） 设 χ 是一个阶为 δ 的原特征（模 p ）。那么 $x^\delta \equiv t \pmod{p}$ 的解的个数为

$$1 + \chi(t) + \cdots + \chi^{\delta-1}(t).$$

因此

$$S_{a,p} = \sum_t \{1 + \chi(t) + \cdots + \chi^{\delta-1}(t)\} e\left(\frac{a}{p}t\right), \quad (6.3)$$

其中此处（以及本证明中其他地方）求和都是对模 p 的一个完全剩余系进行的。来自方括号中项 1 的那个和为 0，因为 $a \not\equiv 0 \pmod{p}$ 。

第二步：用特征把"解的个数 $N_\delta(t)$ "写成一串可计算的项。这是整个证明最巧的一招。原文又把两件事说成"显然"：① 解数公式；② 项 1 的和为 0。逐一补全。

"阶为 δ 的特征"是什么 取原根 g 和一个 δ 次单位根 $\omega = e(1/\delta)$ ，定义特征 $\chi(g^y) = \omega^y$ 。它的"阶"是使 $\chi^d = \chi_0$ （主特征）的最小正整数 d ，这里恰为 δ 。于是 $\chi, \chi^2, \dots, \chi^{\delta-1}$ 是 $\delta-1$ 个互不相同的非主特征，而 $\chi^\delta = \chi_0$ （主特征）。"原特征"在这里就指它真正的阶是 δ （不是更小）。

1 为什么解数 $= 1 + \chi(t) + \cdots + \chi^{\delta-1}(t)$ 。分两种情况看右边这串。考虑几何和 $\sum_{j=0}^{\delta-1} \chi^j(t) = \sum_{j=0}^{\delta-1} (\chi(t))^j$ 。

◦ 若 t 是某个 x 的 δ 次方 (即 $t = x^\delta$)，写 $t = g^{\delta m}$ ，则

$\chi(t) = \omega^{\delta m} = 1$ (因 $\omega^\delta = 1$)。于是每项都是 1，和 $= \delta$ 。而此时 $x^\delta \equiv t$ 恰好有 δ 个解 (δ 次方映射是 δ -对一)。两边都 $= \delta$ 。✓

◦ 若 $t \neq 0$ 但不是 δ 次方，则 $\chi(t) \neq 1$ ，用几何级数公式

$$\sum_{j=0}^{\delta-1} \chi(t)^j = \frac{\chi(t)^\delta - 1}{\chi(t) - 1} = \frac{1 - 1}{\chi(t) - 1} = 0 \quad (\text{因}$$

$\chi(t)^\delta = \chi^\delta(t) = \chi_0(t) = 1$)。而此时方程 $x^\delta \equiv t$ 无解，解数 $= 0$ 。两边都 $= 0$ 。✓

◦ 若 $t = 0$ ：左边解数 $= 1$ (只有 $x = 0$)；右边

$1 + \chi(0) + \cdots = 1 + 0 + \cdots = 1$ (约定 $\chi(0) = 0$)。✓

三种情况都对上，所以这串就是解的个数 $N_\delta(t)$ ，妙处是它把"几何意义的解数"翻译成"可以逐项求和的代数式"。

2 把这串解数公式代入 $S_{a,p} = \sum_t N_\delta(t) e(at/p)$ ，就得到 (6.3)。

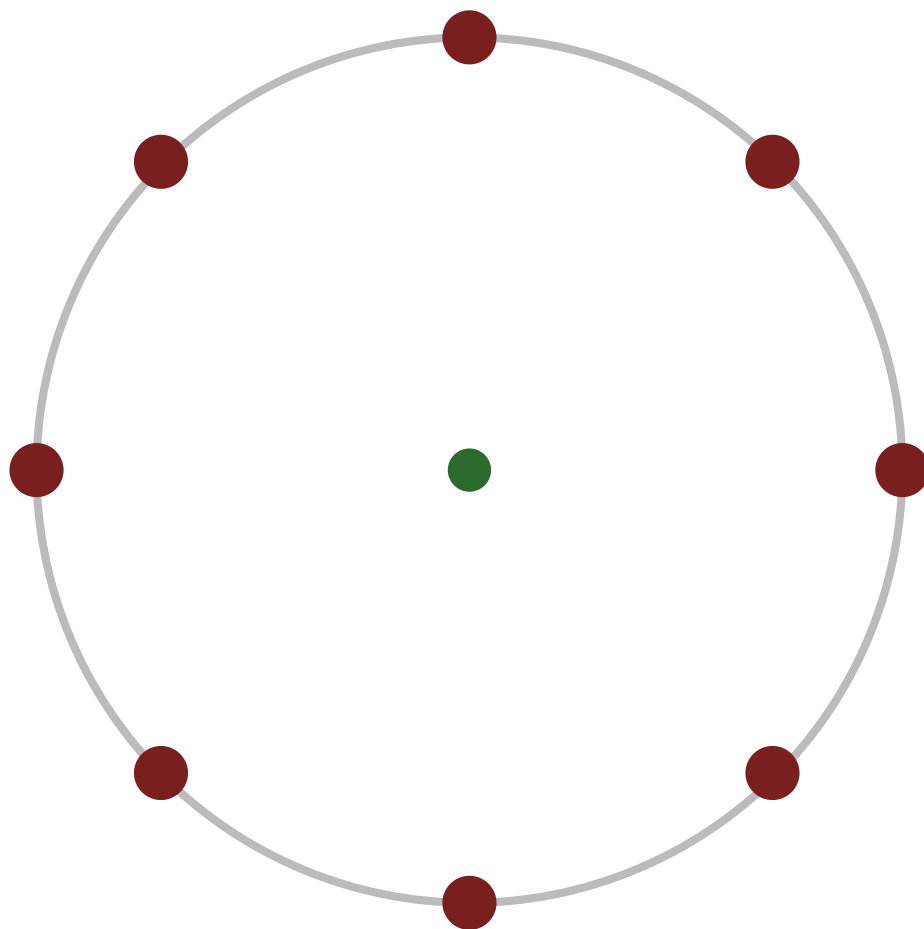
3 项 1 的和为 0。方括号里的"1"单独拎出来是

$$\sum_{t=0}^{p-1} 1 \cdot e\left(\frac{a}{p}t\right) = \sum_{t=0}^{p-1} \left(e\left(\frac{a}{p}\right)\right)^t.$$

记 $z = e(a/p)$ 。因为 $a \not\equiv 0 \pmod{p}$ ， a/p 不是整数，故 $z \neq 1$ 。这是公比 $z \neq 1$ 的几何级数：

$$\sum_{t=0}^{p-1} z^t = \frac{z^p - 1}{z - 1} = \frac{e(a) - 1}{z - 1} = \frac{1 - 1}{z - 1} = 0,$$

用了 $z^p = e(a/p)^p = e(a) = 1$ (a 是整数，转整数圈回原点)。所以方括号里的 1 贡献为 0，可以从 (6.3) 里划掉。**几何意义**： p 个均匀分布在单位圆上的点 (p 次单位根) 相加，正好抵消成 0。



合力 = 0（重心在圆心）

项 1 的和： p 个等间隔的单位根，对称分布，向量相加正好回到原点 0。这就是 $\sum_t e(at/p) = 0$ 的几何图像。

原文 · 证明（高斯和） 若 ψ 是任何一个非主特征（模 p ），则称和

$$T(\psi) = \sum_t \psi(t) e\left(\frac{at}{p}\right)$$

为高斯和（Gauss sum），以纪念这类和在高斯的分圆论工作中所起的重要作用。我们可以很容易地证明 $|T(\psi)| = p^{1/2}$ ，如下所示。我们有

$$|T(\psi)|^2 = \sum_t \sum_u \psi(t) \bar{\psi}(u) e\left(\frac{a}{p}(t-u)\right).$$

这里我们可以略去 $u = 0$ ，因为 $\psi(0) = 0$ 。把变量从 t 换成 v ，其中 $t \equiv vu \pmod{p}$ ，我们得到

$$|T(\psi)|^2 = \sum_v \sum_{u \neq 0} \psi(v) e\left(\frac{au}{p}(v-1)\right).$$

内层和当 $v = 1$ 时为 $p - 1$ ，否则为 $-\psi(v)$ 。因此

$$|T(\psi)|^2 = p\psi(1) - \sum_v \psi(v) = p.$$

这就是前面所述的结果。在 (6.3) 中对 $\psi = \chi, \dots, \chi^{\delta-1}$ 应用此结果，我们便得到 (6.2)。

第三步：高斯和的模长恰好是 $\sqrt{p}$ 。这一段是引理 6.1 的"发动机"。先说动机：(6.3) 划掉项 1 后剩下的，正是 $\delta - 1$ 个高斯和

$$S_{a,p} = T(\chi) + T(\chi^2) + \cdots + T(\chi^{\delta-1}), \quad T(\psi) = \sum_t \psi(t) e(at/p).$$

每个 $T(\psi)$ 的模长若能精确算出 $= \sqrt{p}$ ，则三角不等式立刻给 $|S_{a,p}| \leq (\delta - 1)\sqrt{p}$ ，正是 (6.2)。所以全部问题归为：**证明 $|T(\psi)| = \sqrt{p}$ 。**下面逐句补全。

记号 $\bar{\psi}$ 与 $|z|^2 = z\bar{z}$ 对复数 $z = x + iy$ ，共轭 $\bar{z} = x - iy$ ，且 $|z|^2 = z\bar{z}$ (实数)。对特征， $\bar{\psi}(u) = \overline{\psi(u)}$ 是 $\psi(u)$ 的共轭。关键：对 $u \neq 0$ ， $|\psi(u)| = 1$ ，故 $\psi(u)\bar{\psi}(u) = |\psi(u)|^2 = 1$ 。还会用到 $\overline{e(\theta)} = e(-\theta)$ 。

1 展开 $|T|^2 = T\bar{T}$ 。

$$|T(\psi)|^2 = \left(\sum_t \psi(t) e\left(\frac{at}{p}\right) \right) \overline{\left(\sum_u \psi(u) e\left(\frac{au}{p}\right) \right)} = \sum_t \sum_u \psi(t) \bar{\psi}(u) e\left(\frac{a}{p}(t-u)\right)$$

用了共轭把乘积拆成双重和： $\overline{\psi(u)e(au/p)} = \overline{\psi(u)}e(-au/p)$ ，再与 $e(at/p)$ 合并指数得 $e(a(t-u)/p)$ 。

- 2 略去 $u = 0$ 。当 $u = 0$ ， $\psi(0) = 0$ ，整列项都是 0，删之无害。于是 u 只跑非零剩余。

- 3 换元 $t \equiv vu \pmod{p}$ 。固定一个非零 u ，当 t 跑遍完全剩余系时， v （由 $t = vu$ 决定， $v \equiv tu^{-1}$ ）也跑遍完全剩余系（因为 u 可逆，乘 u 是个一一对应）。代入： $\psi(t) = \psi(vu) = \psi(v)\psi(u)$ ，于是

$$\psi(t)\overline{\psi(u)} = \psi(v)\psi(u)\overline{\psi(u)} = \psi(v) \cdot 1 = \psi(v),$$

而 $t - u = vu - u = u(v - 1)$ ，故指数变成 $e(au(v - 1)/p)$ 。得到

$$|T(\psi)|^2 = \sum_v \sum_{u \neq 0} \psi(v) e\left(\frac{au}{p}(v - 1)\right).$$

注意 $\psi(v)$ 与 u 无关，可提到内层和外面。

- 4 算内层和 $\sum_{u \neq 0} e(au(v - 1)/p)$ 。

- 若 $v = 1$ ：指数里 $(v - 1) = 0$ ，每项都是 $e(0) = 1$ ，而 u 跑 $p - 1$ 个非零值，故内层和 $= p - 1$ 。该 v 的贡献为 $\psi(1)(p - 1) = p - 1$ ($\psi(1) = 1$)。
- 若 $v \neq 1$ ： $a(v - 1) \not\equiv 0 \pmod{p}$ ，所以 $\sum_{u=0}^{p-1} e(au(v - 1)/p) = 0$ （又是 p 个单位根抵消）。去掉 $u = 0$ 那一项（值为 1），得 $\sum_{u \neq 0} = 0 - 1 = -1$ 。该 v 的贡献为 $\psi(v) \cdot (-1) = -\psi(v)$ 。

这正是原文“内层和 $v = 1$ 为 $p - 1$ 、否则为 $-\psi(v)$ ”。

- 5 合计。把所有 v 的贡献加起来：

$$|T(\psi)|^2 = \underbrace{(p-1)}_{v=1} + \sum_{v \neq 1} (-\psi(v)) = (p-1) - \left(\sum_v \psi(v) - \psi(1) \right)$$

因为 ψ 非主, $\sum_v \psi(v) = 0$ (特征正交, 预备里讲过), 又 $\psi(1) = 1$, 所以

$$|T(\psi)|^2 = (p-1) - (0-1) = p.$$

(这与原文写的 $p\psi(1) - \sum_v \psi(v) = p \cdot 1 - 0 = p$ 是同一个结果, 只是合并次序不同。) 于是 $|T(\psi)| = \sqrt{p}$ 。

6 收尾得 (6.2)。把 $\psi = \chi, \chi^2, \dots, \chi^{\delta-1}$ (共 $\delta-1$ 个非主特征, 每个模长都 $= \sqrt{p}$) 代回 $S_{a,p} = \sum_{j=1}^{\delta-1} T(\chi^j)$, 用三角不等式:

$$|S_{a,p}| \leq \sum_{j=1}^{\delta-1} |T(\chi^j)| = (\delta-1)\sqrt{p}.$$

正是 (6.2)。■

为什么是"平方根"而不是别的 $|T|^2 = p$ 这一步是灵魂: 双重和里有 p^2 项, 但因为相位的精确抵消, 只剩下"对角贡献"的 p 量级, 开方就是 $\sqrt{p}$ 。这种 " N 项随机相位相加, 模长约 $\sqrt{N}$ " 是解析数论里无处不在的"平方根抵消"原理, 高斯和是它最干净的范例。

原文·注 (6.2) 在 $p=2$ (于是 $\delta=1$) 时仍然成立, 但此时它是平凡的, 因为 $a=1$ 且 $S_{1,2} = 1 + e^{i\pi} = 0$ 。

这段是**边界情形的核对**。当 $p=2$: $p-1=1$, 故 $\delta=(k,1)=1$, 于是 (6.2) 右边 $(\delta-1)\sqrt{p}=0$ 。这要求 $|S_{a,2}| \leq 0$, 即 $S_{a,2}=0$ 。验证: 模 2 时 $(a,2)=1$ 只有 $a=1$, x 取 0, 1, x^k 取 0, 1,

$$S_{1,2} = e\left(\frac{1}{2} \cdot 0\right) + e\left(\frac{1}{2} \cdot 1\right) = e(0) + e\left(\frac{1}{2}\right) = 1 + e^{i\pi} = 1 + (-1) = 0.$$

(用 $e(\frac{1}{2}) = e^{2\pi i \cdot 1/2} = e^{i\pi} = \cos \pi + i \sin \pi = -1$ 。) 确实成立，且两边都是 0，所以叫"平凡"——成立但没给新信息。这种把公式拿到极端值上对一遍的习惯，是确保推理没有漏洞的好办法。

三、引理 6.2：模 p^ν ($p \nmid k$) ——"提升一位数字"降阶

原文·引理 6.2 设 $a \not\equiv 0 \pmod{p}$ 且 $p \nmid k$ 。那么，对于 $1 < \nu \leq k$,

$$S_{a,p^\nu} = p^{\nu-1}, \quad (6.4)$$

而对于 $\nu > k$,

$$S_{a,p^\nu} = p^{k-1} S_{a,p^{\nu-k}}. \quad (6.5)$$

引理 6.1 解决了"指数 $\nu = 1$ "。现在要处理素数幂 p^ν ，先假设 $p \nmid k$ (p 不整除 k ，是"好"素数，大多数素数都这样)。结论有两条：

- (6.4)：当指数 ν 在 2 到 k 之间，和被精确算出来，等于 $p^{\nu-1}$ (一个干净的整数，没有抵消之外的零头)；
- (6.5)：当 $\nu > k$ ，和满足一个**降阶递推**——把指数从 ν 降到 $\nu - k$ ，每降一次掏出一个因子 p^{k-1} 。这正是后面证 $|S_{a,q}| \ll q^{1-1/k}$ 的命门：递推让"高次幂"问题不断回落到"低次幂"。

原文·证明 在定义

$$S_{a,p^\nu} = \sum_{x=0}^{p^\nu-1} e\left(\frac{a}{p^\nu} x^k\right)$$

中，我们令 $x = p^{\nu-1}y + z$ ，其中 $0 \leq y$

第一步：把 x 拆成"高位 y + 低位 z "。这是全章最核心的技巧，叫"提升变量"或" p -进位分解"。动机：我们想看看"最高那一位数字 y 的变化"对 x^k 的影响——结果会发现它在指数里变成一个干净的线性项，于是对 y 求和能直接抵消。

为什么这样拆 $x = p^{\nu-1}y + z$ 是合法且唯一的 任何 $0 \leq x$ 唯一写成 $x = p^{\nu-1}y + z$ ，其中 y 是" x 写成 p 进制时的最高位" ($0 \leq y$)

1 展开 $x^k = (z + p^{\nu-1}y)^k$ (二项式定理)。

$$(z + p^{\nu-1}y)^k = z^k + \binom{k}{1} z^{k-1} (p^{\nu-1}y) + \binom{k}{2} z^{k-2} (p^{\nu-1}y)^2 +$$

$$\text{即} = z^k + k z^{k-1} p^{\nu-1} y + (\text{含 } p^{2(\nu-1)} \text{ 及更高}).$$

2 高次项在模 p^ν 下消失。从第三项起每项都带因子

$$(p^{\nu-1})^2 = p^{2(\nu-1)} \text{ 或更高幂。而 } \nu \geq 2 \text{ (因 } 1 < \nu) \text{ 时}$$

$2(\nu-1) = 2\nu-2 \geq \nu$ (等价于 $\nu \geq 2$)。所以 $p^{2(\nu-1)}$ 是 p^ν 的倍数，这些项 $\equiv 0 \pmod{p^\nu}$ ，可以扔掉。剩

$$x^k \equiv z^k + k p^{\nu-1} z^{k-1} y \pmod{p^\nu}.$$

这就是原文那句"因为 $2(\nu-1) \geq \nu$ "的完整含义。注意此时还没用到 $p \nmid k$ ，这一步对一般 p 都对。

x (一个 0 到 $p^{\nu-1}$ 的数, p 进制 ν 位):

高位 y

低位 z ($z < p^{(\nu-1)}$)

$$x = p^{(\nu-1)} \cdot y + z \rightarrow x^k \equiv z^k + k \cdot p^{(\nu-1)} \cdot z^{(k-1)} \cdot y \pmod{p^\nu}$$

把 x 按 p 进制拆成"最高位 y "和"低位 z "。最高位 y 在 x^k 里只留下一个线性项，正好用来对 y 做抵消。

原文 · 证明 (续) 因此

$$S_{a,p^\nu} = \sum_{z=0}^{p^{\nu-1}-1} \sum_{y=0}^{p-1} e\left(\frac{az^k}{p^\nu} + \frac{akz^{k-1}y}{p}\right).$$

由于 $ak \not\equiv 0 \pmod{p}$ ，内层和为 0，除非 $z \equiv 0 \pmod{p}$ ，此时它等于 p 。因此，若 $z = pw$ ，

$$S_{a,p^\nu} = p \sum_{w=0}^{p^{\nu-2}-1} e\left(\frac{aw^k}{p^{\nu-k}}\right).$$

第二步：对高位 y 求和，制造抵消。 这里第一次用上 $p \nmid k$ 。

1 把上一步的同余式代入指数。注意

$$\frac{a}{p^\nu} x^k \equiv \frac{a}{p^\nu} (z^k + kp^{\nu-1}z^{k-1}y) = \frac{az^k}{p^\nu} + \frac{akz^{k-1}y}{p} \quad (\text{第二项里 } p^{\nu-1}/p^\nu = 1/p).$$

于是

$$S_{a,p^\nu} = \sum_z \sum_y e\left(\frac{az^k}{p^\nu} + \frac{akz^{k-1}y}{p}\right) = \sum_z e\left(\frac{az^k}{p^\nu}\right) \underbrace{\sum_{y=0}^{p-1} e\left(\frac{akz^{k-1}y}{p}\right)}_{\text{内层}}.$$

含 z 但不含 y 的因子 $e(az^k/p^\nu)$ 提到内层和外面。

2 算内层和。它是几何级数 $\sum_{y=0}^{p-1} (e(akz^{k-1}/p))^y$ ，公比

$$w_0 = e(akz^{k-1}/p).$$

○ 若指数 $akz^{k-1} \not\equiv 0 \pmod{p}$ ，则 $w_0 \neq 1$ ，和

$$= \frac{w_0^p - 1}{w_0 - 1} = 0 \quad (w_0^p = e(akz^{k-1}) = 1).$$

○ 若 $akz^{k-1} \equiv 0 \pmod{p}$, 则每项 = 1, 和 = p 。

3 判断何时内层和非零。 $akz^{k-1} \equiv 0 \pmod{p}$ 何时成立? 因为 $a \not\equiv 0$ 且 $p \nmid k$, 所以 $ak \not\equiv 0 \pmod{p}$ (两个非零相乘还是非零, 模素数下没有零因子)。于是只能是 $z^{k-1} \equiv 0 \pmod{p}$, 即 $p \mid z$, 即 $z \equiv 0 \pmod{p}$ 。这正是 $p \nmid k$ 派上用场的地方: 它保证只有 z 是 p 倍数时内层和才不死。

4 只留 $z = pw$ 。满足 $z \equiv 0 \pmod{p}$ 且 $0 \leq z$

原文·证明(收尾) 若 $\nu \leq k$, 则最后这个和中所有项都为 1, 于是得到 $S_{a,p^\nu} = p^{\nu-1}$ 。若 $\nu > k$, 一般项是 w 的周期函数, 周期为 $p^{\nu-k}$, 因而

$$S_{a,p^\nu} = p p^{k-2} S_{a,p^{\nu-k}}.$$

第三步: 分 $\nu \leq k$ 和 $\nu > k$ 收尾。

1 情形 $\nu \leq k$ (得 (6.4))。此时指数 $\nu - k \leq 0$, 记 $\nu - k = -m$ ($m \geq 0$), 则 $\frac{aw^k}{p^{\nu-k}} = aw^k p^{k-\nu} = aw^k p^m$ 是整数。整数 θ 满足 $e(\theta) = 1$ 。所以和里每项都是 1, 共 $p^{\nu-2}$ 项 (w 从 0 到 $p^{\nu-2} - 1$)。于是

$$S_{a,p^\nu} = p \cdot p^{\nu-2} = p^{\nu-1}.$$

得 (6.4)。(这里需要 $\nu \geq 2$ 让 $p^{\nu-2}$ 有意义, 正是引理条件 $1 < \nu$ 。)

2 情形 $\nu > k$ (得 (6.5))。此时 $\nu - k \geq 1$, 和里的项 $e(aw^k/p^{\nu-k})$ 是 w 的周期函数, 周期为 $p^{\nu-k}$: 因为把 w 换成

$w + p^{\nu-k}$ 时, w^k 模 $p^{\nu-k}$ 不变 (更准确地, 指数 $aw^k/p^{\nu-k}$ 的小数部分不变, 而 e 只看小数部分)。一个完整周期 $w = 0, \dots, p^{\nu-k} - 1$ 上的和正是定义里的 $S_{a,p^{\nu-k}}$ 。求和范围 $w = 0, \dots, p^{\nu-2} - 1$ 共 $p^{\nu-2}$ 项, 含

$$\frac{p^{\nu-2}}{p^{\nu-k}} = p^{(\nu-2)-(\nu-k)} = p^{k-2}$$

个完整周期。所以

$$\sum_{w=0}^{p^{\nu-2}-1} e\left(\frac{aw^k}{p^{\nu-k}}\right) = p^{k-2} S_{a,p^{\nu-k}},$$

从而 $S_{a,p^\nu} = p \cdot p^{k-2} S_{a,p^{\nu-k}} = p^{k-1} S_{a,p^{\nu-k}}$, 即 (6.5)。■

这条递推的妙处 (6.5) 说 S_{a,p^ν} 与 $S_{a,p^{\nu-k}}$ 只差一个明确的整数因子 p^{k-1} 。所以无论 ν 多大, 反复用 (6.5) 就能把它压回到 $\nu \leq k$ 的情形, 那里已被 (6.4) 或引理 6.1 算清。换句话说: 无穷多个素数幂的和, 本质上只有有限种"花样"。后面引理 6.4 正是吃这口饭。

四、引理 6.3: 补上 $p \mid k$ 的"坏"素数

原文 · 引理 6.3 引理 6.2 的第二个结果在 $p \mid k$ 时也成立。

引理 6.2 假设了 $p \nmid k$ 。但 k 的素因子 (如 $k = 4$ 时的 $p = 2$) 也得照顾。这一段证明递推 (6.5) 对 $p \mid k$ 同样成立 (注意: 只是第二个结果 (6.5), 不是 (6.4))。为什么单要 (6.5)? 因为引理 6.4 的降阶只用 (6.5); 坏素数的 " $\nu \leq k$ 部分" 会用别的方式裹进常数里。

原文·证明（开头） 像之前一样令 $k = p^\tau k_0$ ，并注意由于 $\nu > k$ ，我们有

$$\nu > p^\tau k_0 \geq 2^\tau \geq \tau + 1,$$

因而 $\nu \geq \tau + 2$ 。事实上， $k \geq \tau + 2$ ，因为当 $\tau = 1$ 时 $k \geq 6$ 。

记号 $k = p^\tau k_0$ 把 k 中所有 p 的因子提出来： $\tau \geq 1$ 是" p 在 k 里出现的次数" (p -adic 赋值)， k_0 是剩下与 p 互素的部分 ($p \nmid k_0$)。例如 $k = 12, p = 2$: $12 = 2^2 \cdot 3$ ，故 $\tau = 2, k_0 = 3$ 。引理 6.2 是 $\tau = 0$ 的情形；现在 $\tau \geq 1$ 。

1 先确立一些大小关系（后面会反复用）。 因为 $\nu > k = p^\tau k_0$ 。又 $p \geq 2, k_0 \geq 1$ ，故 $p^\tau k_0 \geq 2^\tau$ 。再由 $2^\tau \geq \tau + 1$ （这对 $\tau \geq 1$ 都成立： $2^1 = 2 \geq 2, 2^2 = 4 \geq 3, \dots$ 指数增长甩开线性）。串起来 $\nu > 2^\tau \geq \tau + 1$ ，因 ν 是整数得 $\nu \geq \tau + 2$ 。

2 更强的： $k \geq \tau + 2$ 。 若 $\tau \geq 2$ ，则 $k \geq p^\tau \geq 2^\tau \geq \tau + 2$ 。若 $\tau = 1$ ：当 p 为奇素数时 $k \geq p \geq 3 = \tau + 2$ 直接成立；最需要小心的是 $p = 2$ （后面归纳里 $p = 2$ 的条件最苛刻），此时 k 为偶数但 $4 \nmid k$ ，即 $k = 2k_0$ (k_0 为奇数)，在 Waring 背景 $k \geq 3$ 的前提下 $k_0 \geq 3$ ，故 $k \geq 6$ ——这正是原文" $k \geq 6$ 若 $\tau = 1$ "所针对的（最紧的 $p = 2$ ）情形。无论哪种情况都有 $k \geq \tau + 2$ 。这个不等式保证后面出现的指数 $k - \tau - 2 \geq 0$ 不会变成负数。

原文·证明（修改的拆分） 我们修改前面的证明，令 $[x = p^{\nu-\tau-1}y + z, \text{qqquad } 0 \leq y$

第一步：为什么要换一种拆法。 引理 6.2 里拆 $x = p^{\nu-1}y + z$ （只取最高 1 位作 y ），靠的是" y 的线性项系数 ak 与 p 互素"才有抵消。可现

在 $p \mid k$, 那个系数会被 p 整除, 抵消会失灵。对策: 让 y 多吃几位——取 y 占最高 $\tau + 1$ 位 ($0 \leq y$

- 1 承认 (6.6) 先往下走。把 (6.6) 代入 $S_{a,p^\nu} = \sum_x e(ax^k/p^\nu)$ 。指数里

$$\frac{a}{p^\nu} x^k \equiv \frac{a}{p^\nu} (z^k + kp^{\nu-\tau-1} z^{k-1} y) = \frac{az^k}{p^\nu} + \frac{akz^{k-1}y}{p^{\tau+1}}.$$

再把 $k = p^\tau k_0$ 代入第二项: $\frac{ak}{p^{\tau+1}} = \frac{ap^\tau k_0}{p^{\tau+1}} = \frac{ak_0}{p}$ 。于是第二项 = $\frac{ak_0 z^{k-1} y}{p}$ 。(原文里那个负号来自 (6.6) 推导的符号约定, 不影响后面"是否为零"的判断, 因为只关心指数是否为整数。) 得到原文那个双重和。

- 2 内层对 y 求和 (y 跑 $p^{\tau+1}$ 个值)。内层和
 $= \sum_{y=0}^{p^{\tau+1}-1} e\left(\frac{ak_0 z^{k-1}}{p} y\right)$ 。关键: 现在系数里是 ak_0 而不是 ak , 而 $p \nmid a$, $p \nmid k_0$, 故 $ak_0 \not\equiv 0 \pmod{p}$! 抵消恢复了。该几何和为 0 除非 $p \mid z^{k-1}$ 即 $z \equiv 0 \pmod{p}$, 否则 ($z \equiv 0$) 每项为 1、和 = $p^{\tau+1}$ 。

- 3 只留 $z = pw$, 降阶。满足 $(z \equiv 0, 0 \leq z$

原文·证明 (归约到二项同余) 还需证明同余式 (6.6)。只需证明

$$(z + p^{\nu-\tau-1} y)^{p^\tau} \equiv z^{p^\tau} + p^{\nu-1} z^{p^\tau-1} y \pmod{p^\nu}$$

即可, 因为接下来把两边升到 k_0 次方不会带来任何困难。令 $\nu - \tau - 1 = \lambda$, 我们要证明

$$(z + p^\lambda y)^{p^\tau} \equiv z^{p^\tau} + p^{\lambda+\tau} z^{p^\tau-1} y \pmod{p^{\lambda+\tau+1}}. \quad (6.7)$$

这并不像看上去那样显然，因为并非 $(A + B)^{p^\tau}$ 展开式中的所有二项式系数都能被 p^τ 整除。

第二步：把要证的 (6.6) 归约成纯粹的二项式同余 (6.7)。这里出现了证明里"最技术"的一段，原文承认"并不像看上去那样显然"，我们要把它讲透。

1 为什么"先证 p^τ 次方就够"。因为 $k = p^\tau k_0$ ，所以

$x^k = (x^{p^\tau})^{k_0}$ 。如果能证明

$$(z + p^{\nu-\tau-1}y)^{p^\tau} \equiv z^{p^\tau} + p^{\nu-1}z^{p^\tau-1}y \pmod{p^\nu}, \quad (*)$$

那么把 (*) 两边再升到 k_0 次方：右边 $(z^{p^\tau} + p^{\nu-1}(\dots))^{k_0}$ ，

用二项式，除首项 $z^{p^\tau k_0} = z^k$ 外，其余每项都含

$(p^{\nu-1})^j, j \geq 1$ ；其中 $j = 1$ 项给

$k_0 z^{(p^\tau)(k_0-1)} \cdot p^{\nu-1} z^{p^\tau-1} y = k_0 p^{\nu-1} z^{k-1} y, j \geq 2$ 项含

$p^{2(\nu-1)} \equiv 0 \pmod{p^\nu}$ (因 $\nu \geq 2$)。于是

$$x^k \equiv z^k + k_0 p^{\nu-1} z^{k-1} y \pmod{p^\nu}.$$

咦，这里系数是 k_0 而原文 (6.6) 写的是 k (在 $\text{mod } p$ 意义下) ——其实 (6.6) 是 $\text{mod } p$ 的弱断言，而我们这里得到更强的 $\text{mod } p^\nu$ ；代入指数 $\frac{a}{p^\nu} \cdot k_0 p^{\nu-1} z^{k-1} y = \frac{ak_0}{p} z^{k-1} y$ ，与第一步算出的第二项一致。所以"升到 k_0 次方不带来困难"。问题归到证 (*)。

2 换记号 $\lambda = \nu - \tau - 1$ ，得 (6.7)。注意 (*) 右边第二项指数

$\nu - 1 = \lambda + \tau$ ，模数 $p^\nu = p^{\lambda+\tau+1}$ ，于是 (*) 就写成

(6.7):

$$(z + p^\lambda y)^{p^\tau} \equiv z^{p^\tau} + p^{\lambda+\tau} z^{p^\tau-1} y \pmod{p^{\lambda+\tau+1}}.$$

- 3 为什么不能一步硬展开。若直接对 $(z + p^\lambda y)^{p^\tau}$ 用二项式定理，第 j 项是 $\binom{p^\tau}{j} z^{p^\tau-j} (p^\lambda y)^j$ 。我们想说 $j \geq 2$ 的项都 $\equiv 0 \pmod{p^{\lambda+\tau+1}}$ 。但 $\binom{p^\tau}{j}$ 不一定被 p^τ 整除（例如 $\binom{p^\tau}{p^{\tau-1}}$ 只含较低的 p 幂），单靠二项式系数提供的 p 幂不够，得不到 $p^{\lambda+\tau+1}$ 。这就是原文说“不像看上去那样显然”的精确含义。对策：不一次展开，而是一次只升一个 p 次方，分 τ 个阶段（对 τ 归纳）。

原文·证明（一步一个 p 次方）然而，我们可以分阶段证明此结果（换句话说，对 τ 作归纳）。我们首先证明

$$(z + p^\lambda y)^p \equiv z^p + p^{\lambda+1} z^{p-1} y \pmod{p^{\lambda+2}} \quad (6.8)$$

其中要求 $\lambda \geq 1$ （当 $p > 2$ 时）或 $\lambda \geq 2$ （当 $p = 2$ 时）。二项式展开中唯一需要检验的项是最后一项；对此我们需要 $\lambda p \geq \lambda + 2$ ，而当 $p > 2$ 且 $\lambda \geq 1$ 时，或当 $p = 2$ 且 $\lambda \geq 2$ 时，这是成立的。

第三步：基础引理 (6.8)——升一个 p 次方。

- 1 对 $(z + p^\lambda y)^p$ 用二项式定理：

$$(z + p^\lambda y)^p = z^p + \binom{p}{1} z^{p-1} (p^\lambda y) + \sum_{j=2}^{p-1} \binom{p}{j} z^{p-j} (p^\lambda y)^j$$

- 2 第 $j = 1$ 项： $\binom{p}{1} = p$ ，给 $p \cdot z^{p-1} \cdot p^\lambda y = p^{\lambda+1} z^{p-1} y$ ，正是目标里要保留的项。

- 3 中间项 $2 \leq j \leq p-1$ ：素数 p 的二项式系数 $\binom{p}{j}$ 当 $1 \leq j \leq p-1$ 时都被 p 整除（因 $\binom{p}{j} = \frac{p}{j} \binom{p-1}{j-1}$ ，分子带 p

, 分母 \j

- 4 最后一项 $j = p$ (原文"唯一需要检验的"): $(p^\lambda y)^p = p^{\lambda p} y^p$ 。
要它 $\equiv 0 \pmod{p^{\lambda+2}}$, 需 $\lambda p \geq \lambda + 2$, 即 $\lambda(p-1) \geq 2$ 。
- $p > 2$ ($p \geq 3$) 且 $\lambda \geq 1$: $\lambda(p-1) \geq 1 \cdot 2 = 2 \checkmark$ 。
 - $p = 2$: $p-1 = 1$, 需 $\lambda \geq 2 \checkmark$ ($\lambda = 1$ 时 $\lambda(p-1) = 1 < 2$ 会失败, 所以 $p = 2$ 必须 $\lambda \geq 2$, 这就是为何对 $p = 2$ 要额外条件)。

- 5 于是在相应条件下, 除前两项外都消失, (6.8) 成立。它说: 把 $(z + p^\lambda y)$ 升一个 p 次方, 结果形状不变——还是" z 的 p 次方 + 一个线性余项", 只是 λ 升级为 $\lambda + 1$ 、模数升级一格。这正好可以接力。

原文·证明 (归纳接力) 最后, (6.7) 通过对 (6.8) 的重复应用而得出; 在下一阶段我们得到

$$\begin{aligned}(z + p^\lambda y)^{p^2} &= (z^p + p^{\lambda+1} z^{p-1} y_1)^p \\ &\equiv z^{p^2} + p^{\lambda+2} z^{p^2-1} y_1 \pmod{p^{\lambda+3}} \\ &\equiv z^{p^2} + p^{\lambda+2} z^{p^2-1} y \pmod{p^{\lambda+3}},\end{aligned}$$

其中 $y_1 \equiv y \pmod{p}$, 并且我们在此用 $\lambda + 1$ 替代 λ 应用了 (6.8)。论证继续下去, 便给出 (6.7)。

第四步: 用 (6.8) 当积木, 归纳出 (6.7)。

- 1 先记基底 $B = z + p^\lambda y$ 。由 (6.8),
 $B^p \equiv z^p + p^{\lambda+1} z^{p-1} y \pmod{p^{\lambda+2}}$ 。把它写成" z^p + 高位"的形状: $B^p = z^p + p^{\lambda+1} z^{p-1} y_1$, 其中
 $y_1 \equiv y \pmod{p}$ (同余允许我们把 B^p 真实值的高位部分记

成某个 y_1 , 它模 p 等于 y)。注意: 现在它形如 $z' + p^{\lambda+1}y_1$ (把 $z' = z^p$ 当新"底", $\lambda + 1$ 当新" λ "), 与 (6.8) 的输入同型。

- 2 再升一个 p 次方, 对 $(z^p + p^{\lambda+1}z^{p-1}y_1)^p$ 套 (6.8) (用 $\lambda + 1$ 代 λ 、用 z^p 代 z):

$$(z + p^\lambda y)^{p^2} = B^{p^2} \equiv z^{p^2} + p^{\lambda+2}z^{p^2-1}y_1 \equiv z^{p^2} + p^{\lambda+2}z^{p^2-1}y$$

最后一步把 y_1 换回 y : 因为 $y_1 \equiv y \pmod{p}$, 而它前面乘了 $p^{\lambda+2}$, 差出来的 $p^{\lambda+2}(y_1 - y)$ 含 $p^{\lambda+3}$, 在 $\text{mod } p^{\lambda+3}$ 下可忽略。所以 $\tau = 2$ 的情形成立。

- 3 归纳。假设升 $p^{\tau-1}$ 次方后已成立

$$B^{p^{\tau-1}} \equiv z^{p^{\tau-1}} + p^{\lambda+\tau-1}z^{p^{\tau-1}-1}y \pmod{p^{\lambda+\tau}}, \text{ 再用 (6.8)}$$

(这次以 $\lambda + \tau - 1$ 代 λ) 升最后一个 p 次方, 同样的 " $y_1 \rightarrow y$ 可换"论证, 便得到 (6.7) 的 p^τ 版本。每一步都只升一个 p 次方、模数升一格, 所以二项式系数提供的那一个 p 因子始终够用——这正是"分阶段"绕开 " $\binom{p^\tau}{j}$ 不够整除"困难的办法。

原文·证明 (核对条件, 收尾) 当 $\lambda = \nu - \tau - 1$ 时, 关于 λ 的条件得到满足。我们已经看到 $\nu - \tau - 1 \geq 1$, 而如前所注, 若 $p = 2$ 则有 $\nu \geq k + 1 \geq \tau + 3$ 。因此证明完成。

第五步: 核对归纳用到的 λ 条件确实满足。 (6.8) 在归纳里被反复用, 每次要求 " $\lambda \geq 1$ ($p > 2$)" 或 " $\lambda \geq 2$ ($p = 2$)", 其中起始 $\lambda = \nu - \tau - 1$ (之后只增不减, 所以查起始值最严)。

- 1 $p > 2$: 要 $\lambda = \nu - \tau - 1 \geq 1$ 。由第一步 $\nu \geq \tau + 2$ 得 $\nu - \tau - 1 \geq 1 \checkmark$ 。

2 $p = 2$: 要 $\lambda \geq 2$ 即 $\nu \geq \tau + 3$ 。由前面 $\nu > k$ 即 $\nu \geq k + 1$, 又 $k \geq \tau + 2$, 故 $\nu \geq (\tau + 2) + 1 = \tau + 3$ ✓。这正是为什么前面要费劲确立 $k \geq \tau + 2$: 专门为 $p = 2$ 这个最挑剔的情形兜底。

条件都满足, (6.7) 成立, 从而 (6.6)、从而 (6.5) 对 $p \mid k$ 也成立。■

五、引理 6.4: 拼装——对一般 q 的整体估计 (6.1)

原文·引理 6.4 当 $(a, q) = 1$ 时, $|S_{a,q}| \ll q^{1-1/k}$ 。

这就是本章开头的目标 (6.1)。前面三条引理分别拿下了 $\nu = 1$ (引理 6.1)、 $\nu \leq k$ 且 $p \nmid k$ (引理 6.2 (6.4))、以及所有 $\nu > k$ 的降阶 (引理 6.2/6.3 的 (6.5))。现在用 $S_{a,q}$ 的乘性把不同素数的估计拼成一块。

原文·证明 令 $T(a, q) = q^{-1+1/k} S_{a,q}$ 。我们要证明 $T(a, q)$ 有一个与 q 无关的界。若 $q = p_1^{\nu_1} p_2^{\nu_2} \cdots$, 则由引理 5.1 证明中 $S_{a,q}$ 的乘性性质,

$$T(a, q) = T(a_1, p_1^{\nu_1}) T(a_2, p_2^{\nu_2}) \cdots,$$

其中 $a_1, a_2, \dots$ 取适当的值, 每一个都与相应的 p^ν 互素。由引理 6.2 的第二部分以及引理 6.3, 我们有

$$T(a, p^\nu) = T(a, p^{\nu-k})$$

对 $\nu > k$ 成立, 于是我们可以假定所有 ν_i 都 $\leq k$ 。

为什么定义 $T(a, q) = q^{-1+1/k} S_{a,q}$ 我们想证 $|S_{a,q}| \ll q^{1-1/k}$ ，等价于证归一化后的量 $|T(a, q)| = |S_{a,q}|/q^{1-1/k}$ 被一个与 q 无关的常数控制。引入 T 的好处：(1) 目标变成最干净的" T 有界"；(2) T 仍然乘性，可以逐素数拆开。这里 $q^{-1+1/k} = q^{-(1-1/k)}$ (指数就是 $-1 + 1/k$)。

1 乘性。第 5 章引理 5.1 证过：当 $q = q_1 q_2$ 且 $(q_1, q_2) = 1$ ，存在 $(a_1, q_1) = (a_2, q_2) = 1$ 使 $S_{a,q} = S_{a_1,q_1} S_{a_2,q_2}$ 。直观原因：中国剩余定理把 $x \bmod q$ 一一对应到 $(x \bmod q_1, x \bmod q_2)$ ，求和因此劈成两个独立求和的乘积。把 q 按素因子全分解 $q = \prod p_i^{\nu_i}$ ，得 $S_{a,q} = \prod_i S_{a_i, p_i^{\nu_i}}$ 。

2 T 也乘性。因为 $q^{1-1/k} = \prod_i (p_i^{\nu_i})^{1-1/k}$ (幂对乘积可拆)，把 S 的乘积逐因子除以对应的 $(p_i^{\nu_i})^{1-1/k}$ ：

$$T(a, q) = \frac{\prod_i S_{a_i, p_i^{\nu_i}}}{\prod_i (p_i^{\nu_i})^{1-1/k}} = \prod_i T(a_i, p_i^{\nu_i}).$$

所以只要每个 $|T(a, p^\nu)|$ 有界，乘起来就有界（只要参与相乘的"超过 1 的因子"个数有限——下面正是要说明这点）。

3 用 (6.5) 把 ν 压到 $\leq k$ 。由 $S_{a, p^\nu} = p^{k-1} S_{a, p^{\nu-k}}$ (引理 6.2 第二部分或引理 6.3)，换算成 T ：

$$T(a, p^\nu) = \frac{S_{a, p^\nu}}{(p^\nu)^{1-1/k}} = \frac{p^{k-1} S_{a, p^{\nu-k}}}{p^{\nu(1-1/k)}}.$$

而 $T(a, p^{\nu-k}) = \frac{S_{a, p^{\nu-k}}}{p^{(\nu-k)(1-1/k)}}$ 。两者相除，验证

$$T(a, p^\nu) = T(a, p^{\nu-k}) : \text{需 } \frac{p^{k-1}}{p^{\nu(1-1/k)}} = \frac{1}{p^{(\nu-k)(1-1/k)}},$$

即比较指数

$$(k-1) - \nu(1 - \frac{1}{k}) \stackrel{?}{=} -(\nu-k)(1 - \frac{1}{k}).$$

$$\text{右边} = -\nu(1 - \frac{1}{k}) + k(1 - \frac{1}{k}) = -\nu(1 - \frac{1}{k}) + (k-1)$$

, 与左边相同 ✓ (用 $k(1 - 1/k) = k-1$)。所以

$T(a, p^\nu) = T(a, p^{\nu-k})$ 。反复用, 可把任何 ν 降到 $\leq k$ 。于是不妨设所有 $\nu_i \leq k$ 。

原文 · 证明 (逐因子定界) 由引理 6.1,

$$T(a, p) \leq kp^{1/2}p^{-(1-1/k)} \leq kp^{-1/6},$$

而由引理 6.2 的第一部分,

$$T(a, p^\nu) = p^{\nu-1}p^{-\nu(1-1/k)} \leq 1 \quad \text{当 } 1 < \nu \leq k.$$

因此 $T(a, p^\nu) \leq 1$, 除非可能在 $\nu = 1$ 且 $p \leq k^6$ 的情形。于是

$$T(a, q) \leq \prod_{p \leq k^6} (kp^{-1/6}),$$

而右边的数与 q 无关。

逐个素因子估计 T , 找出"可能大于 1"的只有有限个。

1 $\nu = 1$ 的因子。由引理 6.1, $\|S_{a,p}\| \leq (\delta-1)\sqrt{p} \leq (k-1)\sqrt{p}$

2 这个界何时 ≤ 1 ?

$kp^{-1/6} \leq 1 \iff p^{1/6} \geq k \iff p \geq k^6$ 。所以当 $p \geq k^6$, $\nu = 1$ 的因子已经 ≤ 1 , 乖乖收敛不添乱; 只有 $(p|k)$ 。

3 $2 \leq \nu \leq k$ 的因子。由 (6.4) $S_{a,p^\nu} = p^{\nu-1}$, 故

$$T(a, p^\nu) = \frac{p^{\nu-1}}{p^{\nu(1-1/k)}} = p^{(\nu-1)-\nu/k} = p^{-1+\nu/k}.$$

因 $\nu \leq k$, 指数 $-1 + \nu/k \leq -1 + 1 = 0$, 故

$T(a, p^\nu) \leq p^0 = 1$ 。这些因子永远 ≤ 1 , 从不添乱。(坏素数 $p \mid k$ 的 $2 \leq \nu \leq k$ 因子怎么办? 它们的 $\backslash(p$

4 拼装。把所有素因子的 T 乘起来。除了 $\backslash(p \mid k)$, 其余全 ≤ 1 。所以

$$|T(a, q)| = \prod_i |T(a_i, p_i^{\nu_i})| \leq \prod_{p \leq k^6} (k p^{-1/6}).$$

右边是固定有限个 ($p \leq k^6$ 的素数个数有限) 数的乘积, 它只依赖 k , 与 q 完全无关。记这个常数为 C 。于是

$|T(a, q)| \leq C$, 即

$$|S_{a,q}| = q^{1-1/k} |T(a, q)| \leq C q^{1-1/k} \ll q^{1-1/k}.$$

正是 (6.1)。■

把 q 拆成素数幂, 逐个看归一化和 $T(a, p^\nu)$:

$p < k^6$ 且 $\nu=1$
可能 >1 (有限个)

其余全部 ≤ 1
($p \geq k^6$ 的 $\nu=1$; 及所有 $\nu \geq 2$)

乘积 $|T(a, q)| \leq \prod_{p \leq k^6} (k \cdot p^{-1/6}) = \text{常数 } C$, 与 q 无关

$\Rightarrow |S(a, q)| \leq C \cdot q^{1-1/k}$, 证毕 (6.1)

乘性 + 逐因子定界: 只有有限个"小素数 + 一次幂"的因子可能超过 1, 它们的乘积是个与 q 无关的常数, 于是整体被 $q^{1-1/k}$ 控制。

六、定理 6.1：奇异级数收敛且有正下界

原文·定理 6.1 若 $s \geq 2k + 1$ ，则奇异级数 $\mathfrak{S}(N)$ 以及乘积 $\prod_p \chi(p)$ 绝对收敛，并且

$$\mathfrak{S}(N) \geq C_1(k, s) > 0$$

若 $s \geq 2k + 1$ (k 为奇数) 或 $s \geq 4k$ (k 为偶数)。

这是本章的主定理，把前面所有引理收口。它有两层：

- **绝对收敛** ($s \geq 2k + 1$)：级数形 $\mathfrak{S}(N) = \sum_q A(q)$ 与乘积形 $\prod_p \chi(p)$ 都收敛、且与 N 无关地一致有界。
- **正下界** (k 奇要 $s \geq 2k + 1$, k 偶要 $s \geq 4k$)： $\mathfrak{S}(N)$ 不只是收敛，还离 0 有正的距离。这一步至关重要——圆法主项是 $\mathfrak{S}(N)$ 乘别的正量，只有 $\mathfrak{S}(N) > 0$ 才能断言" N 真的可表为 s 个 k 次方和"。

记号 $\chi(p)$ 与 $C_1(k, s)$ $\chi(p)$ (注意：这是局部密度因子，与前面引理里的乘法特征 χ 同字母但完全不同含义，靠上下文区分) 指"在模 p 的各次幂下，方程 $x_1^k + \cdots + x_s^k \equiv N$ 解的相对密度的极限"，它度量"在素数 p 这个局部， N 可表的容易程度"。奇异级数有欧拉乘积表示 $\mathfrak{S}(N) = \prod_p \chi(p)$ 。 $C_1(k, s)$ 是一个只依赖 k, s (不依赖 N) 的正常数，即统一的正下界。

原文·证明 绝对收敛性如前所述得出，只需用引理 6.4 代替引理 3.1 的推论；而最后的断言由引理 5.6 及引理 5.2 的推论得出。

1 **绝对收敛**。就是本章第一节"把欠条点清"里那段推导：

$$|S_{a,q}| \ll q^{1-1/k} \quad (\text{引理 6.4})$$

$$\Rightarrow |A(q)| \ll q^{1-s/k} \Rightarrow \sum_q |A(q)| \ll \sum_q q^{1-s/k}, \text{ 在}$$

$s \geq 2k + 1$ 时是收敛的 p -级数。第 3 章曾用"引理 3.1 推论"那个较弱的版本，现在换成更干净的引理 6.4，结论照旧成立。

2 正下界。这部分调用第 5 章成果：引理 5.6 把 $\mathfrak{S}(N)$ 与"各局部因子 $\chi(p)$ 之积"挂钩并保证每个 $\chi(p) > 0$ ；引理 5.2 推论则把"小素数部分的乘积"与"大素数部分（接近 1）"分别控制住，得到统一的正下界 $C_1(k, s)$ 。奇偶分类（ k 偶要 $s \geq 4k$ ）来自：偶次幂在模 2 的幂下解更受限（如 $x^4 \equiv 0, 1 \pmod{16}$ ），需要更多变量 s 才能保证局部有解、密度不退化到 0。■

原文 定理 6.1 表明，就奇异级数本身而言（除 $k = 4$ 外），在改进条件 $s \geq 2k + 1$ 使其成立方面不会有任何困难。困难的关键在于次弧（minor arcs），而不在奇异级数。

这段是**战略评论**，点明圆法里"真正难的是哪一块"。回顾全局：圆法把表示数 $r(N) = \int_0^1 T(\alpha)^s e(-N\alpha) d\alpha$ 拆成主弧（贡献主项，主项 = 奇异积分 $\times$ 奇异级数）+ 次弧（要证它小到淹不过主项）。

记号 主弧 $\mathfrak{M}$ 与次弧 $\mathfrak{m}$ 把 $[0, 1]$ 按 α 是否靠近"分母小的分数 a/q "切两块：靠近的小区间并起来是主弧 $\mathfrak{M}$ （ $T(\alpha)$ 在那里大、有渐近公式，给主项）；剩下的全是次弧 $\mathfrak{m}$ （ $T(\alpha)$ 剧烈震荡、没有封闭公式，只能用 Weyl/Hua 不等式压上界）。

这段话的意思是：**奇异级数**（主弧那一侧的算术因子）现在已经被本章**彻底驯服**——只要 $s \geq 2k + 1$ 它就乖乖收敛、还正。所以若想把可表性的门槛 s 降低，瓶颈不在奇异级数，而在**次弧**：怎样对一般 α 把 $|T(\alpha)|$ 压得更狠。"除 $k = 4$ 外"是个有趣的例外： $k = 4$ 时模 16 的局部约束特别强（下文史料会细讲 $\chi(2)$ 的巨大波动），让奇异级数的正下界要更小心。

七、历史注：Hardy-Littlewood 的 $s \geq 4$ 与"最优" $s \geq 2k + 1$

原文·注 哈代（Hardy）与李特尔伍德（Littlewood）证明了奇异级数，以

$$\sum_{q=1}^{\infty} A(q)$$

的形式，当 $s \geq 4$ 时绝对收敛，对乘积形式也同样成立。其要旨在于确保对 a 求和时所发生的抵消，即

$$\sum_{\substack{a=1 \\ (a,q)=1}}^q (q^{-1} S_{a,q})^s e\left(\frac{-Na}{q}\right).$$

然而，这种绝对收敛对 N 不再是一致的。若取 $q^{-1} S_{a,q}$ 的绝对值，则定理 6.1 中的条件 $s \geq 2k + 1$ 是最优的（best possible）。

背景：Hardy 与 Littlewood 是谁 G. H. Hardy（哈代，1877–1947）与 J. E. Littlewood（李特尔伍德，1885–1977）是 20 世纪初英国剑桥的两位巨匠，长期合作。1920 年代他们在一系列题为 "Some problems of 'Partitio Numerorum'"（数的分拆问题，常缩写 P.N. I, II, ...）的论文里系统创立了**圆法**，把 Waring 问题、Goldbach 问题等放进统一的解析框架。本章的奇异级数、主弧次弧的分法，都源出于此。

这段对比了**两种收敛门槛**，要点在于"**是否利用对 a 求和的抵消**"。逐句拆解：

1 本章的做法（保守）：我们证收敛时，第一步就用三角不等式 $|A(q)| \leq \sum_a |q^{-1}S_{a,q}|^s$ ，把 $e(-Na/q)$ 和 a 求和里的相位全丢了。这相当于“取绝对值”。代价是门槛较高：要 $s \geq 2k + 1$ 。好处是：界与 N 完全无关（一致），这对后面证主项很关键。

2 Hardy-Littlewood 的做法（精打细算）：他们保留对 a 的求和 $\sum_a (q^{-1}S_{a,q})^s e(-Na/q)$ 里的相互抵消（就像引理 6.1 里 $\sum_t e(at/p) = 0$ 那种抵消，但这次发生在对 a 求和上）。抵消额外省下次方，于是门槛降到惊人的 $s \geq 4$ （与 k 无关！）。

3 但有代价：不一致。他们的 $s \geq 4$ 收敛依赖 N ——抵消的程度随 N 而变，收敛速度对不同 N 不一样，不能对所有 N 给统一的界。而圆法证主项 $\gg P^{s-k}$ 恰恰需要“对 N 一致”的下界，所以实际证可表性时还是要用本章这种一致版本。

4 “ $s \geq 2k + 1$ 最优”是什么意思。“最优 / best possible”指：如果你坚持取绝对值 $|q^{-1}S_{a,q}|$ （不利用对 a 的抵消），那么 $s \geq 2k + 1$ 这个条件不能再降——存在例子使得 $s = 2k$ 时 $\sum_q \sum_a |q^{-1}S_{a,q}|^s$ 真的发散。这与第一节算的 p -级数指数 $s/k - 1 > 1 \iff s > 2k$ 的分界完全吻合：恰好卡在收敛边缘，多一不可、少一发散。所以本章的 $2k + 1$ 不是偷懒，而是“绝对值框架下的极限”。

八、 $\mathfrak{S}(N)$ 怎样随 N 波动： $k = 4, s = 21$ 的著名计算

原文 一个有趣的问题是：奇异级数之和 $\mathfrak{S}(N)$ 是如何随 N 波动的。每个因子 $\chi(p)$ 主要取决于 N 所属的剩余类（模 p^γ ）。随 N 波动最大的因子是那些 p 整除 k 的因子，但那些 $p-1$ 与 k 有较大公因子的因子也可能波动得相当可观。

这段提出一个**定量问题**并给出定性回答。 $\mathfrak{S}(N) = \prod_p \chi(p)$ ，每个因子 $\chi(p)$ 是“模 p 各次幂下 N 可表的局部密度”。它怎么随 N 变？

- “ $\chi(p)$ 主要取决于 N 模 p^γ 的剩余类”：局部密度只看 N 在素数 p 处的“低位信息”（模 p 的某个固定次幂 p^γ ），所以 $\chi(p)$ 是 N 的周期函数（周期 p^γ ）。
- “波动最大的是 $p \mid k$ 的因子”：呼应引理 6.3—— k 的素因子让 k 次方在模 p^ν 下取值高度受限（比如 x^4 只能 $\equiv 0, 1 \pmod{16}$ ），于是“ N 落在哪个剩余类”对解的多少影响极大， $\chi(p)$ 随之大起大落。
- “ $p-1$ 与 k 公因子大的也会波动”：呼应引理 6.1 里的 $\delta = (k, p-1)$ —— δ 大意味着 k 次方在模 p 下只能取到 $\frac{p-1}{\delta}$ 个值，约束强、波动大。

原文 在他们早期的论文中，哈代与李特尔伍德主要使用 $\mathfrak{S}(N)$ 的用指数和 $S_{a,q}$ 表示的定义，而不是用同余式（模 p^ν ）的表达式。在 P.N. II [38] 中，他们必须证明 $\mathfrak{S}(N)$ 在 $k=4$ 、 $s=21$ 情形下有正下界。在此情形下波动最大的因子 $\chi(p)$ 是 $\chi(2)$ 和 $\chi(5)$ ；所有其他因子之积与 1 相差不大。他们发现 $\chi(5)$ 在约 0.7 与 1.3 之间变化。但 $\chi(2)$ 的变化幅度约为 200 倍。

这是一段**具体史料**，讲圆法在最经典战例上的真实数字。背景：Waring 问题对 $k=4$ （四次方）要确定“每个充分大的 N 都是多少个四次方之和”，Hardy-Littlewood 在 P.N. II ([38] 是书末参考文献编号) 里处理 $s=21$ （21 个四次方），必须验证 $\mathfrak{S}(N) > 0$ 。

- "用 $S_{a,q}$ 定义而非同余式": 两种等价定义。
 $\chi(p) = \sum_{\nu \geq 0} (\text{模 } p^\nu \text{ 的项})$, 也可写成同余解数的极限。早期他们偏好指数和形式。
- 对 $k = 4, s = 21$: 波动最大的局部因子是 $\chi(2)$ (因 $2 \mid 4$, 对应 $p \mid k$) 和 $\chi(5)$ (因 $p - 1 = 4 = k$, $\delta = (4, 4) = 4$ 很大, 对应 " $p - 1$ 与 k 公因子大")。其余素数的因子之积几乎恒等于 1, 不影响正负。
- 定量: $\chi(5)$ 在 $0.7 \sim 1.3$ 间小幅起伏; 而 $\chi(2)$ 波动约 200 倍——这正是 $k = 4$ 那个"麻烦例外"的根源。

原文 哈代与李特尔伍德证明了 (在所提及的特定情形下)

$$\begin{aligned} \chi(2) = & 1 - 1.3307 \cos \frac{(2N - 5)\pi}{16} + 0.415 \cos \frac{(6N + 1)\pi}{16} \\ & - 0.3793 \cos \frac{(2N + 3)\pi}{8} + \varepsilon(N), \end{aligned}$$

其中 $|\varepsilon(N)| < 0.002$ 。可以验证, 当 $N \equiv 2$ 或 $3 \pmod{16}$ 时, $\chi(2)$ 变得非常小 (但仍为正)。当 $N \equiv 10$ 或 $11 \pmod{16}$ 时它相对较大。这些结果对应于这样一个事实: $x^4 \equiv 0$ 或 $1 \pmod{16}$, 因而在 $x_1^4 + \cdots + x_{21}^4 \equiv N \pmod{16}$ 中对 $x_1, \dots, x_{21}$ 的选择, 在前一种情形下要比后一种情形下受到更多的限制。

这是本章的压轴细节, 把抽象的 $\chi(2)$ 写成 N 的显式三角函数。逐项读懂:

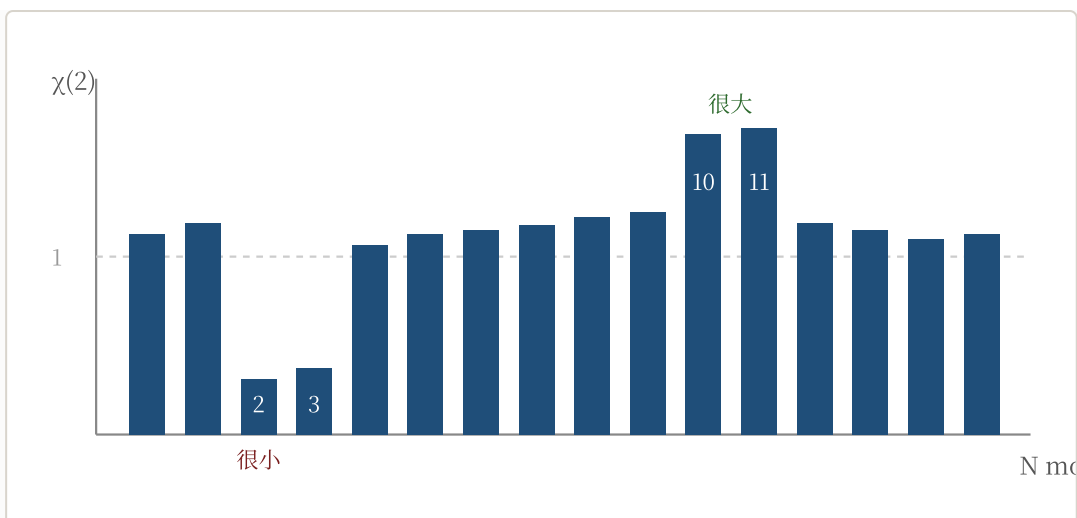
- 1 公式形状: $\chi(2)$ 是 N 的周期函数 (周期 16, 因主要由模 $16 = 2^4$ 的剩余类决定), 写成几个余弦波的叠加, 外加一个极小的误差 $\varepsilon(N)$ ($|\varepsilon| < 0.002$, 可忽略)。cos 的幅角里有 $\frac{(2N-5)\pi}{16}$ 这种, 随 N 整数地跳动, 每 16 个 N 循环一次。

2 为什么是模 16: 因为 $x^4 \bmod 16$ 只能取 0 或 1。验证: 偶数 $x = 2m$, $x^4 = 16m^4 \equiv 0$; 奇数 x , $x^2 \equiv 1 \pmod{8}$ 故 $x^4 \equiv 1 \pmod{16}$ 。所以每个四次方在模 16 下"非 0 即 1", 约束极强, $\chi(2)$ 的全部戏码都在模 16 上演。

3 为什么 $N \equiv 2, 3 \pmod{16}$ 时 $\chi(2)$ 很小: 要凑 $x_1^4 + \cdots + x_{21}^4 \equiv N \pmod{16}$, 每个 x_i^4 只能贡献 0 或 1。于是这 21 项里"取 1 的个数"必须 $\equiv N \pmod{16}$ 。若 $N \equiv 2$ 或 3, 能取 1 的项数很少 (只能是 2 或 3 个, 或 18, 19 个等), 组合方式少、约束紧, 局部解密度低, $\chi(2)$ 小。

4 为什么 $N \equiv 10, 11 \pmod{16}$ 时 $\chi(2)$ 大: 此时"取 1 的项数" $\equiv 10$ 或 11, 处于 $0 \sim 21$ 的中段, 对应的组合数 $\binom{21}{10}, \binom{21}{11}$ 巨大, 选法多、约束松, 局部解密度高, $\chi(2)$ 大。

5 仍然为正: 即便最小的情形 ($N \equiv 2, 3$), $\chi(2)$ 也只是"很小"但严格 > 0 。这正是 Hardy-Littlewood 必须验的——只要每个 $\chi(p) > 0$ 且其积有正下界, $\mathfrak{S}(N) > 0$ 就成立, 从而 21 个四次方足以表示所有大 N 。这就是定理 6.1 的"正下界"在最经典战例里的真实样貌。



示意（非精确值）： $\chi(2)$ 随 $N \bmod 16$ 起伏。 $N \equiv 2, 3$ 处约束最紧、最小（但仍 > 0 ）； $N \equiv 10, 11$ 处选法最多、最大。整体 $\chi(2)$ 极大值与极小值相差约 200 倍——这就是 $k = 4$ 成为"例外"的算术根源。

本章小结

1. 引理 6.1 ($\nu = 1$): 把 k 降成 $\delta = (k, p - 1)$, 用特征拆出 $\delta - 1$ 个高斯和, 每个模长 $\sqrt{p}$, 得 $|S_{a,p}| \leq (\delta - 1)\sqrt{p}$ 。核心是平方根抵消 $|T(\psi)|^2 = p$ 。
2. 引理 6.2 / 6.3 (素数幂): 把 x 按 p 进制拆成"高位 + 低位", 对高位求和制造抵消, 得到 $\nu \leq k$ 的精确值 $p^{\nu-1}$ 与 $\nu > k$ 的降阶递推 $S_{a,p^\nu} = p^{k-1}S_{a,p^{\nu-k}}$; 坏素数 $p \mid k$ 用"多吃 τ 位 + 分阶段二项式归纳"补齐。
3. 引理 6.4 (一般 q): 靠 $S_{a,q}$ 的乘性逐素数拼装, 只有有限个小素数因子可能 > 1 , 乘积是与 q 无关的常数, 得 $|S_{a,q}| \ll q^{1-1/k}$, 即 (6.1)。
4. 定理 6.1: 由 (6.1) 推出 $s \geq 2k + 1$ 时奇异级数绝对收敛、且 $\mathfrak{S}(N) \geq C_1 > 0$ 。瓶颈不在奇异级数, 而在次弧。
5. 史料: Hardy-Littlewood 用抵消把门槛降到 $s \geq 4$ (但对 N 不一致); 取绝对值时 $2k + 1$ 最优。 $k = 4, s = 21$ 时

$\chi(2)$ 随 $N \bmod 16$ 波动达 200 倍，根源是
 $x^4 \equiv 0, 1 \pmod{16}$ 。

返回 [全书目录](#)