arXiv:2611.28614v1 [math.NA] 21 Nov 2026

确定 性非对称特征值问题的正则化

Matthew J. Colbrook*

摘要

我们给出一种在精确算术下实现伪谱分裂的确定性算法，解决了非对称特征值问题复杂度方面的一个重要公开问题。给定一个阶数为 $n \geq 2$ 且满足 $0 < \delta < 1/2$ 的复收缩矩阵 A，我们计算一个满足$\|E\| < \delta$ 的扰动 E，使得 $A + E$ 具有单谱。其特征基条件数与最小特征值间隔之比至多为 $C(n/\delta)^{20/3}$。该算法还计算出 $A + E$ 的精确特征系统。它使用 $O(n^3\log^d(n/\delta))$ 次实域运算和比较，其中 C和 d 为绝对常数。我们依据判别式模的对数的圆周均值下界来选择秩一扰动。二次惩罚用于控制累积更新对矩阵造成的偏移。通过在多次更新中使用同一个参考特征基，我们可以估计所有当前投影算子范数，同时只恢复下一次扰动所需的方向。

1 引言

分离非对称矩阵的特征值时，有两个量至关重要。即使特征值彼此间隔很大，几乎平行的特征向量也会使其对扰动十分敏感。反之，正规矩阵具有正交归一特征基，但其特征值可能任意接近。正则化扰动必须同时控制特征基的条件数和最小特征值间隔。有关谱敏感性的背景介绍，请参见 Trefethen 和 Embree [16]。

Davies [8] 建立了定量的近似对角化界，其指数取决于维度。在高斯扰动下对特征基条件数进行多项式正则化的结果已在 [4] 中证明。Banks 等人 [3] 还控制了特征值间隔，并利用伪谱分裂，在几乎与矩阵乘法相同的时间内获得了一种数值稳定的随机化对角化算法。对于范数有多项式界的矩阵，Shah、Srivastava 和 Zeng [13, 定理 1.1] 证明了：对于期望支集大小为 $n^{1+\alpha}$、其中 $0 < \alpha \leq 1/2$ 为固定值的伯努利–高斯扰动，以高概率可得到特征基条件数和逆间隙的多项式界。

这些结果留下了一个悬而未决的问题：如何确定性地找到一个合适的扰动，使计算量达到立方级（忽略对数因子）。Banks 等人 [3, 第 6] 节提出了这一问题，而 Amsel 等人 [2, 问题 3.1 和公式 (6)] 给出了定量表述。我们解决了这一开放问题的精确算术版本，该问题在 Townsend 的汇编 [15]中列为问题 IE-07。输入可能存在缺陷，即使其特征值已经分离，在每次扰动后计算完整特征基的代价也过高。对于阶数为 $n \geq 2$ 且扰动半径为 $0 < \delta < 1/2$的压缩算子，我们的算法将特征基条件数与最小间隔之比界定为 $C(n/\delta)^{20/3}$，并在 $O(n^3\log^d(n/\delta))$ 次实域运算和比较内返回精确特征系统。

*通讯作者。剑桥大学应用数学与理论物理系，剑桥 CB3 0WA，英国。电子邮箱：m.colbrook@damtp.cam.ac.uk。

下文中， $\|\cdot\|$ 表示向量的欧几里得范数或其诱导矩阵范数， $\|\cdot\|_F$表示弗罗贝尼乌斯范数。符号 M^*表示 M的共轭转置， I表示由表达式所需维度确定的单位矩阵。未带下标的所有对数均为自然对数。压缩算子是算子范数至多为一的矩阵。在有限搜索中出现并列时，按照索引的固定顺序处理。对于非奇异矩阵 X，定义 $\kappa_2(X) = \|X\|\,\|X^{-1}\|$。对于阶数为 $n \geq 2$ 且具有互异特征值 $\lambda_1, \dots, \lambda_n$的矩阵 B，定义

$$\operatorname{gap}(B) = \min_{i \neq j} |\lambda_i - \lambda_j|, \qquad \kappa_V(B) = \inf_{B = X \Lambda X^{-1}} \|X\|\,\|X^{-1}\|,$$

其中下确界取遍对角 Λ和可逆 X。研讨会表述使用单位右特征向量。若 V 是任意特征基，且 D 是其列长度构成的对角矩阵，则$\|V D^{-1}\| \leq \sqrt{n}$ 且 $\|DV^{-1}\| \leq \|V\|\,\|V^{-1}\|$。因此，采用上述约定得到的多项式界，也能给出单位列基的多项式界。对于标量矩阵，同一定义给出 $\kappa_V(B) = 1$。最小间隔仅用于阶数至少为二的矩阵。比值$\kappa_V(B)$/间隙(B)衡量谱能够分离为稳定的独立分量的尺度。事实上，对角化给出

$$\|(zI - B)^{-1}\| \leq \frac{\kappa_V(B)}{\min_i |z - \lambda_i|} \qquad (z \notin \{\lambda_1, \dots, \lambda_n\}).$$

因此， ε-伪谱包含在以特征值为中心、半径为 $\varepsilon\kappa_V(B)$的圆盘并集中。若 $0 < \varepsilon <$ 间隙$(B)/(2\kappa_V(B))$，这些圆盘互不相交。该比值的多项式界给出了伪谱分裂所需的逆多项式尺度。

定理 1.1 （确定性谱正则化）。 存在一种确定性算法，可执行以下操作。对于每个 $n \geq 2$, $A \in \mathbb{C}^{n \times n}$ 满足 $\|A\| \leq 1$，且 $0 < \delta < 1/2$，它计算出满足$\|E\| < \delta$ 的 E ，使得 $B = A + E$ 具有单谱，且

$$\frac{\kappa_V(B)}{\operatorname{gap}(B)} \leq C \left(\frac{n}{\delta}\right)^{20/3}.$$

它还计算 B 的特征值、一个特征基及其逆矩阵。对于某些绝对常数 C, d，该算法使用 $O\left(n^3 \log^d\left(n/\delta\right)\right)$ 次实域运算和实数比较。

基本运算包括加法、减法、乘法、除以非零标量以及比较。每种运算的成本均为单位成本，包括对任意长度整数进行的运算。在对整数进行二分搜索时，我们可以将搜索区间扩展为二进制幂次长度，并在每次比较后将该长度减半。我们采用第 6 节中的单位成本约定，其中有界整数以单个域元素中的独立数位存储。每当算法需要根的数值时，就通过域运算对其进行近似。进一步的扰动可使计算得到的特征值和特征向量变为精确值。我们将此校正计入 $\|E\|$ 的界中。

1.1 选择扰动

为选择扰动，我们使用判别式，即成对特征值差的平方乘积。如果秩一扰动中的两个方向与某个敏感特征值的左、右特征向量有足够大的重叠，那么在某个邻近参数处，矩阵铅笔就会出现重特征值。詹森公式将判别式的这些零点转化为其对数模在圆上的均值下界。我们减去一个矩阵平方弗罗贝尼乌斯范数的倍数，该矩阵用于追踪累积更新。此平方范数中的交叉项在圆上的均值为零，因此我们仍可选择一个相位，使得到的势增大。

我们需要多久更换一次特征基？我们根据每个特征值的投影算子范数及其与最近的其他特征值之间的距离，为其赋予对数权重。当权重之和较大时，一次更新带来的势增量足以抵消更换基的代价。当权重之和较小时，它可以界定参考特征向量所需的总精度。此时，我们便可以在多次更新中使用同一组基。每次更新时，我们都在原始欧几里得坐标中估计所有投影算子范数，并且只求出选择下一个方向所需的向量。当超过局部界的指标数量足够少时，我们保留由其余右特征向量张成的不变子空间，并在尾部对角块上递归处理。一个较小的复数平移可以将两个块谱分开。每个保留特征值所需的分离程度取决于其自身的最近间隙。

1.2 概述

第2节确立判别式增长和确定性方向估计。第3节介绍标量多项式算法，第4节构造初始分离谱。第5节和第6节阐述如何在低秩批次期间更新特征系统并估计投影算子范数。第7节综合这些估计并证明定理 1.1。

1.3 AI 声明

在本项工作开展过程中，作者使用 ChatGPT 6（OpenAI）作为交互式辅助工具。AI 协助探索了引理 5.3 中的预解向量精度估计、命题 5.4 中的精确对角化校正与结构化求逆、引理 6.4 中具有单独精度的三角分解，以及命题6.5和6.6中的范数估计与方向恢复构造。在这些交流中，作者提出了所需的问题、假设和估计，并将建议作为进一步数学研究的素材。作者确定了最终的陈述和结论，并对所有内容承担全部责任。

2　秩一扰动下的判别式增长

对于每个单特征值，我们比较其谱投影算子的范数与它到其余谱的距离。设 $B \in \mathbb{C}^{n\times n}$、$n \geq 2$ 具有单谱，令 P_i 表示对应于 λ_i 的谱投影算子，并定义

$$s_i = \min_{j\neq i} |\lambda_i - \lambda_j|, \qquad K_i = \frac{\|P_i\|}{s_i}, \qquad D(B) = \prod_{i<j} (\lambda_i - \lambda_j)^2.$$

此乘积中有 $d = n$ 个一次因子（按重数计）。该公式也定义了重谱情形下的判别式，此时判别式为零。对于 $x \geq 0$，定义 $\log_2^+ x = \log_2 \max\{1, x\}$。若 $B = X\Lambda X^{-1}$ 且 X 的列 x_i 的范数为 1，则对偶行 w_i^* 的范数为 $\|P_i\|$，因为 $P_i = x_i w_i^*$。用弗罗贝尼乌斯范数界定特征基及其逆矩阵的范数，可得

$$\kappa_V(B) \leq n \max_i \|P_i\|. \tag{2.1}$$

每个投影算子都满足 $\|P_i\| \geq 1$；若 $\|B\| < 2$，则 $s_i < 4$。

该算法只需要近似归一化。对于 $0 < \epsilon \leq 1$，可以将非零列缩放为其范数属于 $[(1+\epsilon)^{-1},\ 1]$。首先除以最大绝对实部或虚部。所得平方范数属于 $[1, 2n]$。通过平方比较进行二分法可得到

长度的上界估计，其相对误差为 ϵ，总成本为 $O(n+\log(n/\epsilon))$。将对偶行乘以列缩放的逆矩阵可精确保持投影算子不变。

引理 2.1 (秩一铅笔的碰撞)。设 $B(t) = B + tuv^*$ 满足 B 无重根的，并令 $a_i = |v^* P_i u|$。对于每个 $r > 0$，多项式 $g(t) = D(B(t))$ 至少有 $\frac{1}{2}\#\{i : ra_i > 4s_i\}$ 个零点位于 $|t| \le 2r$ 中，按重数计。因此，对于每个 $h_0 > 0$，

$$\log_2 \frac{\max_{|t|=h_0} |g(t)|}{|g(0)|} \;\ge\; \frac{1}{2\pi}\int_0^{2\pi} \log_2 \frac{|g(h_0 e^{i\theta})|}{|g(0)|}\, d\theta \ge \frac{1}{2}\sum_i \log_2^+ \frac{h_0 a_i}{8 s_i}. \tag{2.2}$$

证明。令 $p - tq$ 表示 $B(t)$ 的特征多项式。对于给定的 r，称满足 $ra_i > 4s_i$ 的 i 为标记的。若 $q = 0$，则 g 为常数，且每个 $v^* P_i u$ 都消失，因此两个结论均成立。否则，令 $h =$ 为 p, q 的最大公因式，$f = p/h$ 和 $v_0 = q/h$。h 的每个根都给出一个驻定特征值分支，其导数为零。选取 $R \in (r, 2r)$，使得 $|t| = R$ 不含 f/v_0 的临界值。由于该有理映射将无穷远映到无穷远，$|t| < R$ 的原像的每个连通分支都是一个平面域，并且是到某个次数为 j 的圆盘上的真分支覆盖。次数为 j 的每个连通分支恰好包含 f 的 j 个根，因为零是正则值。在次数为 $j \ge 2$ 的连通分支中，分歧至少为 $j - 1$。事实上，去掉分支值后会得到一个连通覆盖，其单值化群由绕这些分支值的回路生成，并且在 j 个覆盖片上具有传递性。长度为 e 的循环至多能将 $e - 1$ 个此前互不相交的覆盖片类合并。在局部次数为 e 的点处，该映射在局部坐标下具有形式 $\zeta \mapsto \zeta^e$。它的 e 个逆分支两两之差的阶为 $1/e$，因此它们的平方差对判别式的贡献阶为 $2\binom{e}{2}/e = e - 1$。由于 $j - 1 \ge j/2$，这些连通分支中的分歧至少为其中所含标记根数的一半。

对于一个次数为一的分支，通过 λ_i 的逆分支是单叶的。对简单特征值方程求导，可得其在零点处的导数 $v^* P_i u$。Koebe 四分之一定理表明，以 λ_i 为中心、半径为 $Ra_i/4$ 的圆盘包含于逆 [9, 第 2] 章的像中。对于一个标记的 i，该圆盘包含 B 的最近的其他特征值。零点的第二个原像不可能属于该分支，因此这个其他特征值是 h 的根。在其经过 f/v_0 的像处，移动分支与驻定分支相交。这给出了 $\mathrm{Res}(h,\ f - tv_0)^2$ 的一个零点。不同的一次逆分支的像彼此不相交。因此，每个标记的一次分支都对应一个不同的驻定根。如果多个碰撞发生在同一个参数处，它们的重数会相加。判别式等于移动因子的判别式、这个结式的平方以及 h 的非零判别式之积。这就证明了零点计数。

令 $N(x)$ 表示 g 中模不超过 x 的零点个数。根据詹森恒等式，中的平均值等于 $\int_0^{h_0} N(x)$(2.2) $dx/(x \log 2)$。一次因子的对数平均公式也适用于圆上的零点。将 $r = x/2$ 代入计数式，并对每个 i 的示性函数积分，即得 (2.2)。含有 $a_i = 0$ 的项为零。

□

左、右特征向量的重叠度以乘积的形式进入这一估计。我们需要这两种重叠度在相同的指标上都很大。我们先为左特征向量选择符号，再利用保留下来的指标为右特征向量选择符号。

引理 2.2 （两个方向的同时选择）。给定一个已存储的特征基、它的逆矩阵和非负权重 γ_i，可以在 $O(n^2)$ 次运算内找到两个向量 u, v，其元素为 $\pm 1/n$，使得在承载至少 $1/25$ 总权重的指标上满足 $a_i \ge \|P_i\|/(4n^2)$。对于任意 个指定指标，假设其单位左、右特征向量在各自相位意义下已知，且欧几里得误差至多为 $(100n^2)$。令 s 为满足 $s^2 \ge n$ 的最小 2 的幂。可以在 $O(nm)$ 次运算内选取 $\bar{u}, \bar{v}$，其元素为 $\pm 1/(2s)$，使得对于至少 $m/25$ 个指标，每个向量与对应真实单位方向的重叠模至少为 $1/(16s)$。

证明。对于独立均匀符号 σ 和一个非零复列向量 x，定义 $Z = |x^{*}\sigma|^{2}/\|x\|^{2}$。由独立性和对称性可得 $\mathbb{E}Z = 1$ 和 $\mathbb{E}Z^{2} \le 3$。多项式 $F(z) = z - z^{2}/6$ 的最大值为 $3/2$，且在 $[0, 1/4)$ 上小于 $1/4$。通过对 $\sum_{i} \gamma_{i} F(Z_{i})$ 取条件期望来确定符号。最终总和至少为 $\sum_{i} \gamma_{i}/2$，因此满足 $Z_{i} \ge 1/4$ 的指标承载至少五分之一的权重。若 m_{0} 是某个形式的部分带符号和，且 c_{j} 是其剩余系数，则其四阶矩为

$$|m_{0}|^{4} + 4|m_{0}|^{2} S + 2\Re(\bar{m}_{0}^{2} T) + 2S^{2} + |T|^{2} - 2U, \qquad S = \sum |c_{j}|^{2}, \quad T = \sum c_{j}^{2}, \quad U = \sum |c_{j}|^{4}.$$

二阶矩为 $|m_{0}|^{2} + S$。更新部分和以及关于剩余系数的三个和，对每个形式和符号只需常数次运算。先将条件期望构造应用于左向量，再应用于保留下来的右向量，并将每个符号向量除以 n。第一个断言由 $P_{i} = x_{i} w_{i}^{*}$ 得出。

对于第二个断言，将相同的两种符号选择应用于权重相等的 m 个观测对。对于每个保留下来的指标及其每个观测向量，未缩放的带符号和的模至少是该向量范数的一半。此范数至少为 $1 - 1/(100n^{2})$，而替换为真实向量后，模的变化至多为 $\sqrt{n}/(100n^{2})$。除以 $2s$ 后即可得到所述下界。每次符号选择至多使用 m 个形式。 $\square$

3 矩阵铅笔的判别式与标量根的近似

每次秩一更新都需要在一个圆上选择参数。我们根据判别式多项式作出这一选择，并考虑扰动平方弗罗贝尼乌斯范数中的交叉项。为在更新后恢复特征值，我们还需要将所得特征多项式的根计算到指定精度。

所有运算次数均按实数域计算，其中复标量用两个实标量表示。在纯代数运算次数中，$O(m)$ 表示绝对 c 下的 $O(m \log^{c}(m+2))$。对精度的额外依赖在各数值子程序中具体说明。对于 $x \ge 1$，我们用 $\mathrm{polylog}(x)$ 表示 $1 + \log x$ 的固定次幂。记号 $\|p\|_{1}$ 表示 p 的系数模之和。在系数域上，准线性多项式乘法 [7] 可用。随后，乘积树和余式树可实现不同节点处的准线性多点求值与插值。

牛顿迭代每进行一步，截断倒数的精度就翻一倍。对 F'/F 积分 F，可得 $F(0) = 1$ 的 $\log$；对 $\log F = g$ 进行牛顿迭代，则可得 $g(0) = 0$ [6, 的指数函数，见第 5] 节。经过阶乘缩放后，平移可化为卷积。用半最大公因式算法计算扩展多项式最大公因式的代价为 $O(M(m) \log(m+2))$，其中 $M(m)$ 为乘法代价 [10,，见第 9–11] 章。每种构造都只使用域运算。

3.1 秩一线上的判别式

秩一扰动的特征多项式关于参数是仿射的。这使我们能够通过乘法算子的迹，求得其判别式中的结式。

引理 3.1 (含一个标量参数的结式)。设 $W \neq 0$ 的次数至多为 $2n$，且 f、v 的次数至多为 n，其中至少一个非零。相差一个固定非零标量因子，关于不定元 t 的结式 $\mathrm{Res}_{z}(W, f - tv)$ 可在 $\widetilde{O}(n^{(1+\omega)/2})$ 次运算内计算，其中 $O(n^{\omega})$ 次运算足以完成方阵乘法，且 $2 \le \omega \le 3$。

证明。若 W 为常数，则结式为非零常数。若 W 非常数，则可通过最大公因式判断 W, f, v 是否有公共零点。若有，则结式恒等于零。若没有公共零点且 $v = 0$，则结式为非零常数。在剩下的情形中，将 W 归一化为首一，并定义

$$M = W/\gcd(W, v^{\deg W+1} \bmod W)$$

使用首一最大公因式。被移除的根及其重数只贡献一个非零常数。若 $M = 1$，则返回 1。在 $\mathcal{R} = \mathbb{C}[z]/(M)$ 中，定义 $s = fv^{-1} \bmod M$。对于 M 的每个根 ζ， s 在局部代数上的乘法算子等于标量 $f(\zeta)/v(\zeta)$ 加上一个幂零算子。因此，取遍所有局部因子后，其特征多项式作为关于 t 的多项式，与结式成比例。

所需的迹可以通过按 Brent 和 Kung 的复合算法（ [6, 第2节.1]）对幂进行分组来计算。令 $e = \deg M$，并选取满足 $k^2 > e$ 的最小整数 k。通过连续乘法，在 $\widetilde{O}(ek)$ 次运算内计算模 M 的 s^a 和 s^{jk}，其中 $0 \le a,\ j < k$。对于每个 j，计算以下表达式在无穷远处的前 e 个洛朗系数:

$$\frac{(M' s^{jk}) \bmod M}{M}.$$

z^{-i-1} 的系数为 $\operatorname{tr}_{\mathcal{R}}(z^i s^{jk})$。事实上， M'/M 在各个不同的根处有单极点，其留数为相应根的重数。乘以 s^{jk} 的一个代表元会按其取值缩放留数，而减去多项式部分不会改变这些洛朗系数。计算~这些系数，每个 j 需要 $O(e)$ 次运算。将所得的 $k \times e$ 矩阵与一个 $e \times k$ 矩阵相乘，该矩阵的列由各幂 s^a 的系数组成。其各项包含了每个 $1 \le i \le e$ 的迹 $\operatorname{tr}(s^i)$。将内维度分成长度为 k 的块，代价为 $O((1 + e/k)k^\omega)$。反转首一特征多项式为

$$\exp\left(-\sum_{i=1}^{e} \frac{\operatorname{tr}(s^i)}{i} u^i\right) \bmod u^{e+1}.$$

该恒等式可通过展开按重数计数的特征值乘积的对数得到。

□

命题 *3.2*（判别式与参数选择）。设 p 是具有单根的 $n \ge 2$ 次首一多项式，设 $\deg q < n$，并定义 $g(t) = \operatorname{disc}(p - tq)$。对于 $h_0 > 0$，可以计算出 g（精确到非零比例因子），并找到 $|t_0| = h_0$，使得 $|g(t_0)| \ge 0.98 \max_{|t|=h_0} |g(t)|$。给定至多 n 个进一步的节点 ξ，满足 $p(\xi) \ne 0$，可以选择 $t = t_0(1 + j/(1000n^4))$，$0 \le j \le 2n$，使得

$$|g(t)| \ge 0.9 \max_{|t|=h_0} |g(t)|, \qquad |p(\xi) - tq(\xi)| \ge |p(\xi)|/(10000n^4) \quad \textit{for every such } \xi.$$

代价为 $\widetilde{O}\left(n^{(1+\omega)/2}\right)$。*Strassen* 界 *[14]* 给出 $\omega =$ 对数$_2$ 7 以及 $(1+\omega)/2 < 1.904$。

证明。若 $q = 0$，则判别式为非零常数，圆上的任意 t_0 都是可容许的。否则，设 $h = \gcd(p, q)$ 为首一多项式，并定义 $f = p/h$ 和 $v = q/h$。乘积公式给出

$$g(t) = \operatorname{disc}(h) \operatorname{disc}(f - tv) \operatorname{Res}(h, f - tv)^2,$$

其中常数判别式或线性判别式为 1。由于 p 是无重根的，第一个因子非零。对于 $W = f'v - fv'$，严格次数不等式 $\deg f > \deg v$ 给出 $W \ne 0$。由于 $W = (f - tv)'v - (f - tv)v'$，对首一的 $f - tv$ 进行模约化可得

$$\operatorname{Res}(f - tv, W) = (-1)^{\deg f(\deg f - 1)/2} \operatorname{disc}(f - tv) \operatorname{Res}(f, v).$$

最终结果式是一个非零常数。因此，应用两次引理 3.1 即可计算出完整的参数多项式，精确到一个非零标量，并包含与驻根的所有交点。如有需要，可通过在 $t = 0$ 处计算普通判别式来确定其尺度。

判别式关于系数的齐次性给出 $D = \deg g \le 2n - 2$。参数化曲线 $(1 + is)/(1 - is)$、 $0 \le s \le 1$ 及其四分之一周旋转，以至多为二的速度覆盖单位圆。以有理步长 $1/(100(D + 1))$ 对这些弧取样，并将样本乘以 h_0。对于次数至多为 $D \ge 1$ 的多项式 P，令 $M_P = \max_{|z|=1} |P(z)|$。若 $|a| > M_P$，则由鲁歇定理， $az^D - P(z)$ 的所有 D 根都位于开单位圆盘内。根据高斯–卢卡斯定理，其导数在单位圆上不为零。因此，在边界点取 $a = P'(z)/(Dz^{D-1})$，可得 $|P'(z)| \le DM_P$。对圆进行缩放，即可得到任意正半径下的伯恩斯坦导数不等式。该估计保证存在一个样本，其模与最大模相差不超过 $0.02M_0$ ，其中 $M_0 = \max_{|t|=h_0} |g(t)|$。通过多点求值和模平方比较，选取 t_0 。对于 $D = 0$ ，取 $t_0 = h_0$。

设 $u = 2n/(1000n^4)$。根据应用于反转多项式的最大值原理，半径为 $h_0(1 + u)$ 的外部多项式界为 $(1 + u)^D M_0$。伯恩斯坦不等式进一步给出了扩展圆盘上导数的界 $D(1 + u)^D M_0/h_0$。对于每个试探 j，模的减少量至多为 $uD(1 + u)^D M_0 < 0.08M_0$。对于每个旧节点，令 $a = p(\xi)$、 $b = t_0 q(\xi)$。如果 $|b| \le |a|/2$，所有 j都满足所需的界。否则， $a - b(1 + j/(1000n^4))$ 的连续取值彼此相距超过 $|a|/(2000n^4)$。至多有一个 j 不满足该界。只需测试 $[0, 2n]$ 到$1000n^4 \Re(a/b - 1)$ 中距离最近的整数，该整数可通过比较找到。至多排除 n 个索引。在 $\widetilde{}$构造 g 之后的所有工作成本为 $O(n)$。

□

累积扰动的平方弗罗贝尼乌斯范数为圆上的对数目标函数贡献了一个线性项。取指数后即可得到下一条命题中的权重。

命题 *3.3*（带指数权重的相位选择）。设 $p(x) \in \mathbb{C}[x]$ 是次数为 $n \ge 2$ 且具有单根的首一多项式，设 $q(x) \in \mathbb{C}[x]$ 的次数小于 n，并且 $r > 0$。定义$g(z) = \mathrm{disc}_x(p(x) - rzq(x))$. 设 $c \in \mathbb{C}$，并假设域值 c_0 满足 $|c_0 - c| \le 1/100$。设 A 是 $[1 + |\Re c_0| + |\Im c_0|, 2(1 + |\Re c_0| + |\Im c_0|)]$，中的一个二的整数次幂，且 $J = 2n + 64A$。给定至多 n 个满足 $p(\xi) \ne 0$ 的节点 ξ，可以找到一个域值 z，使得

$$1 \le |z| \le 1 + \frac{1}{500J}, \qquad |g(z)e^{-cz}| \ge 0.9 \max_{|w|=1} |g(w)e^{-cw}|, \qquad |p(\xi) - rzq(\xi)| \ge \frac{|p(\xi)|}{2^{16} nJ}$$

在每个给定节点处。代价为 $\widetilde{O}(n^{(1+\omega)/2} + J)$。

证明。 利用系数递推关系构造 $e^{-c_0 z}$ 的次数为 $64A$ 的泰勒多项式 T。对于 $|z| \le 2$ 和 $l = 64A + 1$，根据指数级数余项和 $l! \ge (l/e)^l$，其相对误差至多为 $e^{4A}(2eA/l)^l < 1/100$。因此，在整个圆盘内， gT 的模介于 $0.99e^{-0.02}$ 倍和 $1.01e^{0.02}$ 倍的 ge^{-cz} 的模之间。命题 3.2 可将 g 计算至一个比例因子。乘积 gT 的次数至多为 J。该命题证明中的有理圆周采样和多点求值给出一个点 z_0 ，使其处的值至少达到其在圆周上的最大模 M 的 0.98 倍。

测试这些点 $z_j = z_0(1 + j/(1000nJ))$、 $0 \le j \le 2n$。在半径为 $1 + 1/(500J)$ 的扩展圆盘上，多项式的模至多为 $(1 + 1/(500J))^J M$。由边界上的伯恩斯坦不等式及导数的最大值原理可得 $|(gT)(z_j) - (gT)(z_0)| < 0.003M$。由于 $0.977 < (0.99/1.01)e^{-0.04} > 0.9$，每次试探的加权模都满足要求。对于给定节点，令$b = rz_0 q(\xi)$。若 $|b| \le |p(\xi)|/2$ ，则所有试探均通过。否则， $p(\xi) - b(1 + j/(1000nJ))$ 的连续取值之差都大于 $|p(\xi)|/(2000nJ)$，因此至多有一次试探不通过。当 $b \ne 0$ 时，可通过二分搜索找到 $\{0, \ldots, 2n\}$ 中最接近 $1000nJ\Re(p(\xi)/b - 1)$ 的整数。测试该整数，且仅当它未达到所需下界时才将其排除。每个节点至多排除一个索引，因此至少还有一次试探$\widetilde{}$。已知 g 后，多项式乘法、采样和节点测试的代价为 $O(J)$。

□

3.2 实域模型中的根近似

根半径估计和递归分裂圆是 Neff 和 Reif [11] 以及 Pan [12, 第 2.2–2.5] 节算法的核心。我们需要给出实域运算次数，并明确说明其对所要求精度的依赖关系。我们控制系数范数中的误差，并使用以下引理来匹配各根，同时计入重数。

引理 3.4 (从系数残差到匹配). 若 p、q 是次数为 $m \geq 1$ 的首一多项式，p 的根位于单位圆盘内，且 $\|q-p\|_1 \leq e < 2^{-m}$，则 q 的根可以与 p 的根进行匹配，使得距离至多为 $4me^{1/m}$，其中包括重数。

证明。在 $|z|=2$ 上，我们有 $|p(z)| \geq 1$ 且 $|q(z)-p(z)| \leq 2^m e < 1$。Rouch´e 定理保证，当 $0 \leq t \leq 1$ 时，$p_t = p + t(q-p)$ 的零点仍位于这个圆内。在任意这样的零点处，$|p(z)| \leq 2^m e$，因此它到 p 最近根的距离至多为 $2e^{1/m}$。将这个半径扩大任意的 $\varepsilon > 0$。所得圆盘并集的每个连通分支的直径至多为 $2m(2e^{1/m}+\varepsilon)$。根作为多重集连续移动，因此每个分支中的根数保持不变。在每个分支内进行匹配，并沿着匹配固定的子序列令 ε 趋于零。 □

定理 3.5 (标量根近似)。对于一个次数为 $m \geq 1$ 、根位于闭单位圆盘内的首一多项式 p 以及一个整数 $b \geq 2$，存在一种确定性算法，可生成 m 个域元素 z_i ，使得

$$\left\| p - \prod_{i=1}^{m} (z - z_i) \right\|_1 \leq 2^{-b}.$$

其代价为 $O(m^{1.9}\mathrm{polylog}(m(b+2)))$ 次实域运算和比较。特别地，$b \geq m(\mu + \lceil \log_2(4m) \rceil)$ 可为 $\mu \geq 0$ 提供 $2^{-\mu}$ 的匹配根的绝对精度。

我们递归地分解多项式。根的距离的包围区间引导我们搜索分离圆或小簇。在每种情况下，我们都会恢复两个近似因子。在每次递归调用开始时，将 b 增加到至少 $2m\lceil \log_2(m+2) \rceil + 10$，并设置 $B = b + 20m + 20, \eta = 2^{-B}/8$。整数指数、二进幂以及下面的精度界，分别通过对整数进行加倍、平方和二分搜索获得。

接下来的四个引理涉及一个次数为 $m \geq 2$ 的首一多项式 p，其根位于闭单位圆盘内，其中 B 和 η 已在上文定义。

引理3.6 （计算根半径的包围区间）。设 C 是一个域元素，满足 $|C| \leq 16$，并定义 $\rho = 2^{-35}/m^2$。将 p的根按其到 C 的距离排序，并按重数计数，记为 ζ_1、……、ζ_m。通过 $m\,\mathrm{polylog}(m(B+2))$ 次运算，可以将真实根对应的有序值 $|\zeta_i - C|$ 包围在区间 $[L_i, U_i]$中。初始部分的包围区间为 $[0, (1+\rho)\eta]$。每个其余的包围区间均具有 $L_i > 0$ 和 $U_i/L_i \leq 1 + 4\rho$。对于低于 $\eta/(1+\rho)$ 的任意半径，返回的包围区间均从零开始。两个端点列表均可按非递减顺序排列。

证明。将 p 平移 C，并利用偶部和奇部进行 Graeffe 平方，直到根成为平移后的根的 N 次幂。取 N 不小于 $2^{38}m^3$ 的最小 2 的幂。通过精确除法移除零根，并在下面的凸包构造中仅保留非零系数。如果多项式是单项式，则所有半径均为零。否则，将剩余多项式记为p_1，其系数为 a_j，非零根为 r_i，并按模的非降序排列。考虑

$$H(t) = \max_{a_j \neq 0} (\log|a_j| + jt), \qquad J(t) = \log|a_{\mathrm{lead}}| + \sum_i \max(t, \log|r_i|).$$

根据每个一次因子的对数平均公式，$\log |p_1(e^t e^{i\theta})|$ 的圆周均值等于 $J(t)$。因此 $J \le H + \log(m+1)$。初等对称系数公式给出 $H \le J + m\log 2$。在两个函数的第 i 个断点之间，将斜率跃变按其重数计数，前一个函数的斜率至少为 i，而后一个函数的斜率至多为 $i-1$。它们在 t 中的间距至多为$m\log 2 + \log(m+1)$。J的第 i 个断点为 $\log |r_i|$。取 N 次方根后，断点估计表明所有原始正半径均处于乘法因子 $1+\rho$ 以内。事实上，对数误差至多为 $(m\log 2 + \log(m+1))/N \le 2m\log 2/N$，其指数小于 $1+\rho$。

断点可以通过比较整数幂来定位。在已排序的横坐标上，通过单调扫描非零系数处的 $(j, \log |a_j|^2)$，构造上凸包。每次方向测试都要比较整数幂，这些幂通过重复平方计算。共有 $O(m)$ 次测试，总计需要 $O(m\log(m+2))$ 次运算。从 i 到 $i+k$ 的一条边给出有序估计值 $(|a_i|^2/|a_{i+k}|^2)^{1/(2kN)}$，其重数为 k。通过求幂将其与 η 比较。如果该估计值不大于η，或者真实半径为零，则返回从零开始的所示区间。否则，对指数进行二分搜索，找到一个包含该估计值、且位于 η 与固定上界 64 之间的相邻二的幂区间。将此区间二分，直到相对宽度至多为 $\rho/4$。将其下端点除以 $1+\rho$，并将上端点乘以 $1+\rho$。由此得到至多为 $1+4\rho$的宽度比。在每个这样的区间内，可以采用相对网格至多为 $\rho/4$ 的固定细分，并在共享端点处进行单调夹逼。若下端点为正，则该估计值的上端夹逼至少为 η，因此扩大的上端点至少为 $(1+\rho)\eta$。因此，两个端点列表都是单调的。此搜索的成本关于 $\log(m(B+2))$ 为多项式级，其中包括比较时使用的求幂运算。

□

Neff 和 Reif [11] 以及 Pan [12,第2节4]利用高阶导数的根作为分割圆的中心。对于每个包含足够多根的圆盘，利用一个初等估计可得一个邻近的导数根。将该估计应用于半径最小的圆盘，就能为所有这类圆盘选取同一个导数根。

在标量根构造中，$D(c,\ r)$表示以 c 为中心、半径为 r的闭圆盘。

引理 3.7（公共导数中心）。设 p 的次数为 m，设 l 为满足 $m/2 \le l < m$的整数，并定义 $d = m - l$。对于每个至少包含 $l+1$ 个 p的零点（按重数计）的闭圆盘 $D(Y, R)$，$p^{(l)}$ 的某个零点都位于 $D(Y,\ (2d+2)R)$ 中。

证明。先固定一个这样的正半径圆盘。经过平移和缩放，可将其化为单位圆盘。选取其中 $l+1$ 个根 $\alpha_0, \ldots\ldots, \alpha_l$（按重数计），并将 $p(z)$表示为 $= \sum a_j z^j$。除以 $\prod_{i=0}^{l}(z-\alpha_i)$所得的商是一个多项式，因此它在 z^{-1} 处的洛朗系数为零。因此

$$0 = a_l + \sum_{r=1}^{d} a_{l+r} h_r, \qquad h_r = [u^r]\prod_{i=0}^{l}(1-\alpha_i u)^{-1}, \qquad |h_r| \le \binom{l+r}{r}.$$

如果 $p^{(l)}$ 的每个根的模都大于 $2d$，那么 $a_l \ne 0$。将 $p^{(l)}(z)/p^{(l)}(0)$ 分解为 d个一次项系数的模小于 $1/(2d)$ 的因子，可得

$$|a_{l+r}|\binom{l+r}{r} \le |a_l|\binom{d}{r}(2d)^{-r}.$$

那么，$r \ge 1$ 的和的模至多为 $|a_l|((1+1/(2d))^d - 1) < |a_l|$，这就产生了矛盾。半径为零时，同样的局部断言可由中心处的重数推出。

现在选择一个半径最小的圆盘 R_*，其中至少包含 $l+1$ 个根。由于根多重集有限且具有紧致性，这样的圆盘必定存在。在其 $2d$ 倍放大圆盘内固定一个导数根。其他每个

题述中的圆盘都与该圆盘至少有一个共同的根，因为 $2(l+1) > m$。它们中心之间的距离至多为 $R + R_*$，且 $R_* \le R$，因此这个导数根到每个这样的中心的距离都在 $(2d+2)R$ 以内。

□

引理 3.8（寻找分离环域）。对于 $m > 200$，定义 $g = \lceil m/50 \rceil$、$h = m - g$、$l = \lfloor 17m/20 \rfloor$ 和 $s = m+3$。给定与 $p^{(l)}$ 的根相匹配且误差为 $a = \eta/(100s)$ 的近似值，可以在 $m\mathrm{polylog}(m(B+2))$ 次运算内找到以下两者之一。

1. 一个中心 C 和半径为 r 的圆，圆内恰有 $k \in [g, h]$ 个根；相对于 C，内侧模至多为 $(1-\sigma)r$，外侧模至少为 $(1+\sigma)r$，其中 $\sigma \ge c/m$，且 $c > 0$ 是一个绝对有理数。

2. 一个中心 C，已知半径小于 4η 的圆盘内有 $k > h$ 个根。

在这两种情况下，$|C| \le 16$；在第二种情况下，r 表示圆盘半径。

证明。设定 $\psi = 1 + 2^{-20}/m$。在所考虑的每个中心处应用引理3.6。如果至少有 g 个包围区间从零开始，则考虑内半径和外半径分别为 $2\eta\psi^{3j}$ 和 $2\eta\psi^{3j+1}$ 的 $m+1$ 个闭环带，$0 \le j \le m$。这些包围区间均位于它们之下。由于分离比为 $\psi^2 > 1 + 4\rho$，其余每个包围区间至多与一个环带相交。因此，必有某个环带不与任何包围区间相交。通过对端点排序并扫描，或使用二分搜索确定每个包围区间相交的环带，可以找到这样的环带，并计算其内部根的数量。如果根的数量至多为 h，则使用算术中点圆。否则，内半径小于 4η，从而得到第二种结果。如果少于 g 个包围区间从零开始，则在 $g \le i \le h$ 处存在间隙 $L_{i+1} \ge \psi U_i$，同样可得到第一种结果，此时使用半径为 $(U_i + L_{i+1})/2$ 的圆。在这两种中点构造中，分离比均可取为至少 $(\psi-1)/(\psi+1)$。

在剩余的情形下，壳层 $[L_g, U_{h+1}]$ 至少包含 $m - 2g + 2$ 个根，且其外半径与内半径之比至多为

$$(\psi(1+4\rho))^{m+1} \le 1 + 4 \cdot 2^{-20}.$$

令 r_0 表示其外半径，C 表示其中心。在 $C + 2r_0$ 和 $C + 2ir_0$ 处重复半径包围计算；如果其中任一计算得到所述结果之一，立即返回。否则，这三个壳层至少有 $m - 6g + 6 \ge 22m/25 > l$ 个公共根。分别将它们的平方半径区间表示为 $[R_j - \Delta_j, R_j + \Delta_j] = [a_j^2, b_j^2]$，其中 $j = 0, 1, 2$。沿实部和虚部方向对平方距离不等式作差，可将它们的公共交集限制在一个中心为、半径为的圆盘内

$$Y = C + \frac{4r_0^2 + R_0 - R_1}{4r_0} + i\frac{4r_0^2 + R_0 - R_2}{4r_0}, \qquad R = \frac{2\Delta_0 + \Delta_1 + \Delta_2}{4r_0}.$$

两个坐标误差分别以 $(\Delta_0+\Delta_1)/(4r_0)$ 和 $(\Delta_0+\Delta_2)/(4r_0)$ 为界。两者之和给出欧几里得误差的上界，并得到所示半径。某个公共点到任一偏移中心的距离至多为 $3r_0$，因此根据相对宽度可得 $b_j \le 4r_0$。这些宽度给出 $R \le r_0/100$。由于一个公共根距 C 介于 $r_0/(1 + 4 \cdot 2^{-20})$ 和 r_0 之间，我们也有 $r_0/2 \le |Y - C| \le 2r_0$。

将 Y 用作下一个中心。如果三个新的壳层再次确定一个圆盘，其初始壳层包含第 g 近的半径，该半径至多为 R。壳层宽度与该半径可能存在的包围误差 $1 + 4\rho$ 相加，得到的外半径至多为 $r_0/50$。如果 $r_0 \le 50\eta$，那么 $R \le \eta/2$，至少 g 个在 Y 处的包围区间从零开始，而该过程会在下一次计算时返回。

为选择一个起始中心，取剩余导数根近似值的按坐标中位数 C_0，并按重数计数。引理 3.7 表明，每个 s 倍扩张圆盘中都有一个导数根，因为 $2(m-l) + 2 \le s$。在整个

构造过程中，我们始终保留一个与该根匹配的近似值。从 C_0 开始，生成至多 $J = 2 + \lceil \log_2(100s) \rceil$ 个由三个壳层确定的连续圆盘；若引理的两种结论中任一种成立，就立即停止。若生成了全部 J 个圆盘，令Y， R 为最终的中心和半径，并令 r_0 表示第一个壳层的外半径。于是$r_0 > 50\eta$。第一次位移的大小至少为 $r_0/2$ ，之后各次位移之和至多为 $2r_0/49$，因此 $|Y - C_0| > r_0/3$。由几何衰减可得 $sR \le r_0/100$。圆盘 $D(Y,\ sR + a)$ 包含指定的匹配逼近元，且其半径小于 $r_0/50$。它严格位于实部或虚部中位坐标的一侧。移除该圆盘之外的所有近似值后，剩余集合非空，且大小至多为原集合的一半，即使坐标存在并列也如此。经过$O(\log(m+2))$条这样的链后，两种结论中必有一种成立。

还需界定应用引理3.6时各中心的位置。根据高斯–卢卡斯定理，所有导数根的模都不超过一，其匹配近似值的模则小于二。在按坐标中位数 $|C_0| \le 3$处。该中心处初始壳层的外半径至多为 $(1 + 4\rho)(|C_0| + 1)$，因为其上端点界定的半径所对应的包围区间下端点为正。因此，两个偏移中心的模都小于12。由三个壳层得到的后续中心与某个单位圆盘内根的距离小于 $r_0/100$ 。沿此序列，壳层半径逐渐减小，因此这些界限始终成立。当返回一个中点圆时，其半径不超过 $(1 + 4\rho)(1 + |C|)$ 或 4η，因此小于 16。所有比较均使用模平方。半径包围计算的次数为 $O(\log^2(m+2))$。

□

分离环域使我们能够近似计算内部根的幂和。对于小簇，我们可以用其中心代替每个内部根。

引理 3.9 (恢复这两个因子)。 引理 *3.8* 的第一种情形给出的圆产生首一因子 f_o, g_o，其次数分别为 $k, m-k$，且满足 $\|p - f_o g_o\|_1 < 2^{-b-3}$。在第二种情形中，可以取$f_o = (z - C)^k$。计算得到的因子的每个根都在 $|z| < 2$ 中。计算成本为 $m\mathrm{polylog}(m(B+2))$。

证明。 对于分离圆，将 p 变换为首一多项式 $S(y) = r^{-m} p(C + ry)$。令 y_i表示其根。对于 2 的幂 $N \ge 2m$ ，计算

$$h = (1 + y^N)^{-1} \bmod S, \qquad F_j = [y^{-j-1}] \frac{(S'h) \bmod S}{S}, \quad 1 \le j \le k.$$

分离环域保证可逆性。重复模平方、扩展最大公因数算法以及~倒序除法的计算成本为 $O(m \log N)$。与引理 3.1 相同的部分分式计算给出 $F_j = \sum_i y_i^j/(1 + y_i^N)$，其中包含重数。对于 $|y_i| \le 1 - \sigma$，与y_i^j 的绝对差至多为 $2(1 - \sigma)^N$。对于 $|y_i| \ge 1 + \sigma$ ，其贡献至多为 $2(1 + \sigma)^{-(N-m)}$，只要 $(1 + \sigma)^{-N}$， $(1 - \sigma)^N \le 1/2$。取

$$b' = b + 40m + 40 + \lceil \log_2(m+1) \rceil$$

并选择 N 作为一个 2 的幂，其值大于 $m(m + b' + \log(m+2))/c$ 的一个足够大的通用倍数，其中 c 是分离常数。这样，所有 k 幂和都能以 $e = 2^{-b'}$ 为绝对误差近似真实的内层幂和。

估计得到的反转的首一内因子为

$$\exp\left(-\sum_{j=1}^{k} F_j u^j / j \right) \bmod u^{k+1}.$$

精确的反转内因子的系数范数至多为 2^m。指数的扰动范数至多为 me，因此因子误差至多为 $2^m (\exp(me) - 1) \le 2^{m+1} me$。反转

系数列表后，通过映射 $f \mapsto r^k f((z-C)/r)$ 返回到 z，其系数范数放大至多为 33^m，因为 r, $|C| \le 16$。所得的原变量中的因子 f_o 是首一的，且与真实首一因子 F 的差至多为 $e_0 = 2^{-b-20m-20}$。

令 G 表示真实的首一外因子，并令 g_o 为 p 除以 f_o所得的欧几里得除法的商。我们有 $\|G\|_1 \le 2^m$。通过在单位圆盘内根处展开倒数一次因子，反转后的 F 模 u^{m+1} 的逆矩阵范数至多为 $\binom{2m}{m} \le 4^m$。根据诺伊曼级数，反转后的 f_o 对应的逆矩阵范数至多为 $2 \cdot 4^m$。因此，对 $p = f_o G + (F - f_o)G$ 进行倒序除法可得 $\|g_o - G\|_1 \le 2\,8^m e_0$。由三角不等式及 $\|f_o\|_1 \le 2^m + e_0$可得 $\|p - f_o g_o\|_1 < 2^{-b-3}$。

在第二种情形下，令 $w = z - C$，并将 $p(C+w)$除以 w^k。其商 Q_w去掉最低的 k个系数，并将次数下移 k。平移后的真实内因子 F 满足 $\|F - w^k\|_1 \le (1+r)^k - 1 \le 2mr$，其中 $r < 4\eta < 2^{-B}$。中心距某个单位圆盘内根不超过 r，因此 $|C| < 2$。因此，平移后的外因子 G 的范数至多为 4^m，从而得到 $\|Q_w - G\|_1 \le 2mr4^m$。平移回去之前的残差至多为 $4mr4^m$。取 $g_o(z) = Q_w(z - C)$，可得残差至多为 $4mr\,12^m < 2^{-b-3}$。最后，在这两种情形下，Rouch´e 在引理 3.4 中的估计都表明，$f_o g_o$ 的所有零点均位于半径为二的圆内。 □

定理 *3.5* 的证明。 对于 $m > 200$，先递归地近似 $p^{(l)}$ 的首一化多项式的根，其次数为 $d = m - l \le 0.155m$。根据高斯–卢卡斯定理，这些根都在单位圆盘内。引理 3.4 提供所需的匹配精度 $a = \eta/(100s)$，前提是导数调用请求的系数指数为 $d(\lceil \log_2(4d/a) \rceil + 1)$。应用引理 3.8 和 3.9。

对于 $2 \le m \le 200$，使用 $g = 1$、$h = m-1$，并以根据次首项系数计算得到的根的算术平均值作为第一个中心。执行相同的三次半径包围计算。假设这些计算既未得到分离圆，也未在较小圆盘中得到簇。那么每个壳层都包含所有根。它们共有的圆盘 $D(Y,\ R)$ 包含所有根，因此也包含它们的平均值 C。这与 $|Y - C| \ge r_0/2$ 和 $R \le r_0/100$ 矛盾。因此无需调用导数。次数为一的多项式可通过一次域运算得到其根，而空因子则给出空列表。

对于每个尚未分解、次数为 j 的因子，将其缩放 $z = 2u$ 并除以 2^j，使其根位于单位圆盘内。要求系数指数至少为 $b + 20m + 20$，并在每次调用开始时施加该下界。对于较小圆盘中的一个簇，直接输出 k 个 C 的副本，并仅对另一个因子进行递归。消除缩放所付出的代价是在系数范数中至多引入 2^m 的因子。由于每个原始因子的根都位于 $|z| < 2$ 内部，其系数范数至多为 3^m。因此，所返回的一次因子的乘积所引入的额外误差至多为 $2 \cdot 3^m \varepsilon + \varepsilon^2$，其中 $\varepsilon = 2^m 2^{-b-20m-20}$。与初始残差合并后，即得到所断言的 2^{-b}。

每次递归调用所需的精度指数至多为 $C_2 m(b + m + \log(m+2))$，其中 C_2 为绝对常数。对于次数超过 200 的情形，每次递归调用中的每个多项式次数至多为 $0.98m$，并且这些调用中多项式次数的 1.9 次方之和至多为

$$m^{1.9}(0.155^{1.9} + 0.98^{1.9} + 0.02^{1.9}) < 0.993 m^{1.9}.$$

事实上，任一因子中的根所占比例都介于 0.02 和 0.98 之间，而根据凸性，和在两个端点之一处取得最大值。在第二种情形中，递归调用只涉及剩余的次数为 $< 0.02m$ 的因子以及导数。达到有界次数之前，任一路径上的精度指数的对数相对于其初始对数增长 $O(\log^2(m+2))$。对有限个剩余次数进行递归，只会使深度增加一个固定常数，并使该精度指数乘以一个固定因子。因此，局部多对数因子均可由 $\log(m(b+2))$ 的一个固定多项式一致地上界。将递归树上的运算成本按几何级数求和，即可证明运算上界。匹配根断言由引理 3.4 得出。

□

4 谱分离下的初始化

判别式迭代从一个具有分离谱且已知特征系统的矩阵开始。秩一修正使特征多项式的近似根成为精确特征值。随后，我们通过在三角形式中平移对角元素来增大间隙。

命题 4.1。 对于 $n \ge 2$、 $\|A\| \le 1$和 $0 < \delta < 1/2$，可以计算出矩阵 B_0 的精确对角化，使其满足

$$\|B_0 - A\| < \delta/4, \qquad \operatorname{gap}(B_0) \ge \delta/(8n),$$

需要 $O(n^3 \operatorname{polylog}(n/\delta))$ 次算术运算。

证明。 通过对相邻坐标进行共轭变换，将 A 化为上 Hessenberg 形式。要消去非零数对 (a, b) 的第二个元素，在左侧使用矩阵

$$\frac{1}{s}\begin{pmatrix} \bar{a} & \bar{b} \\ -b & a \end{pmatrix}, \qquad \sqrt{|a|^2+|b|^2} \le s \le (1 + 1/(100n^4))\sqrt{|a|^2+|b|^2},$$

并在右侧使用其精确逆矩阵。按照第2节的方法，通过分量缩放和二分法计算 s，跳过零数对。共轭变换次数少于 n^2，每次的条件数至多为 $1+1/(100n^4)$，其乘积的条件数小于 1.02。令 $\eta = \delta/128$。将每个模小于 η 的次对角线元素替换为 η。校正仅支撑在一条次对角线上，且其范数至多为 2η。将所得矩阵记为 H。于是 $\|H\| < 2$，且 e_1 对 H 是循环的。

确实， $v_j = H^j e_1$ 在坐标 $j+1$ 以下为零，其第（$j+1$）个元素等于前 j 个次对角元素之积（对于 $0 \le j < n$）。因此，矩阵 $V = [v_0, \dots, v_{n-1}]$ 的行列式的模至少为 $\eta^{n(n-1)/2}$。计算 v_0， $\dots$ ， v_n 并在此基下求解 v_n。所得递推关系为首一特征多项式 p。对于列 r （ z ） $=\operatorname{adj}$ （ $zI - H$）e_1，将$\sum_{j \ge 0} v_j z^{-j-1}$ 乘以 p（z），并取其多项式部分。 r按降幂排列的系数矩阵，是 V 乘以一个单位上三角 Toeplitz 矩阵。其行列式满足相同的下界。根据 p的根所给出的系数界，其范数至多为 $n^2 6^n$ 乘以 $\|v_j\| \le 2^j$。因此，余子式公式给出了其逆范数的对数上界 O（$n^2 \log$（$2n/\delta$））。

将定理 3.5 应用于 $2^{-n}p(2z)$，并采用匹配精度 $e/4$，从而在原变量中对目标点 $0 < e \le 1$达到精度 $e/2$。按实部对所得节点排序，并将 $je/(2n)$ 加到第 j 个节点 $0 \le j < n$ 上。新节点 μ_j 互不相同，且与 p 的根仍在 e范围内匹配。令$f(z) = \prod_j (z - \mu_j)$。对于 $e \le 1$，通过对一次因子乘积作望远镜求和，可得出 $\|p - f\|_1 \le n 5^n e$的系数一范数界。求解非奇异系数方程组 $\ell^* r = p - f$。根据行列式引理， $H_1 = H + e_1 \ell^*$的特征多项式为 f。前述逆矩阵界表明，只要 $\log_2(1/e)$ 大于某个足够大的常数乘以 $n^2 \log(2n/\delta)$， $\|\ell\| \le \delta/32$ 就成立。可以将此指数加倍，直到平方范数满足界限。由于其系数矩阵可逆，所有 $r(\mu_j)$均非零，而 $(zI - H_1)r(z) = f(z)e_1$ 给出了精确的特征向量。

按节点实部递增的顺序，对这些向量进行格拉姆–施密特正交化。精确的投影系数和近似的长度缩放，得到不变旗标的正交基，其条件数至多为 $1 + 1/(100n^4)$。在此基下， H_1 为三角形。将第 j 个对角元素增大 $j\delta/(8n)$， $0 \le j < n$。现在，相邻特征值的实部之差至少为 $\delta/(8n)$。通过三角代入可求得所有特征向量，Hessenberg坐标变换与正交坐标变换的条件数乘积小于 1。03。原坐标中的扰动有界于

$$1.03(2\eta + \delta/32 + \delta/8) < \delta/4.$$

Hessenberg约化、Krylov序列、系数求解和基构造的计算量至多为立方级，并带有对数因子。根近似的代价仅以精度指数的多对数方式增长，指数加倍的次数则为对数级。

□

5 低秩扰动的预解式

多个连续的秩一扰动可以共用一个参考特征基。在该基下，预解式可将特征多项式归结为一个小型行列式，并给出当前矩阵的左、右特征向量。再进行一次小幅校正，就能将计算节点作为精确特征值。

设 $n \ge 2$ 和 $B_0 = X_0 \Lambda W_0$，其中 $W_0 = X_0^{-1}$ 和 $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ 的元素互异。我们用至多 k 个扰动列表示 $B = B_0 + UKV^*$，并保留它们在参考基中的坐标 W_0U 和 V^*X_0。在本节中，我们始终假设 $\|B_0\|,\ \|B\| \le 2$。$\widetilde{}$记号 O 省略了代数子程序中固定次数的 $1 + \log n$。

复杂度估计采用 Bostan、Jeannerod、Mouilleron 和 Schost 的确定性乘法定理 [5, 定理 1.1]。求逆时，我们使用正定Schur补。设对角矩阵 D_a，D_b 的节点列表长度至多为 n，每个列表中的节点互异，且列表之间没有共同节点。若 $D_aC - CD_b = UV^*$ 由宽度为 $w \ge 1$的生成元给出，则将 C 与 w 个稠密列相乘的代价为 $\widetilde{O}(nw^{\omega-1})$ 次域运算，其中 $w \le n$。应用该定理时，取各节点对应的一次因子。这些因子均为首一因子，且在各自族内两两互素；各族互不相交，因此Sylvester算子可逆。系数域具有特征零，一次因子的伴随矩阵就是相应的标量节点。转置后，在左侧进行乘法也有相同的上界。因此，对 b 个右端项进行填充或分块，所得代价为 $\widetilde{O}(n\max(b, w)w^{\omega-2})$。对于维度至多为 w 的块，采用方形分块进行常规稠密乘法，也可得到相同的上界。

引理 5.1（特征铅笔）。给定扰动因子的已存储参考坐标，包括针对提议更新 suv^* 的 W_0u 和 v^*X_0，则 B 的特征多项式以及该更新对应的矩阵铅笔可通过 $\widetilde{O}(n(k+1)^3)$ 次域运算构造。

证明。令 $p_0(z) = \prod_i (z - \lambda_i)$。根据行列式引理，

$$p_K(z) = p_0(z)\det\bigl(I - KV^*X_0(zI - \Lambda)^{-1}W_0U\bigr).$$

我们在 $z = 4 + j/(n+1)$、$0 \le j \le n$ 处计算该恒等式。小矩阵的 k^2 个条目是标量 Cauchy 和。它们在公共分母 p_0 上的分子通过乘积树计算，并利用余式树在外部节点处求值。这需要 $\widetilde{O}(nk^2)$。精确选主元在 $O(nk^3)$ 次运算中计算 $n+1$ 个行列式，并通过插值恢复 p_K。最后，特征多项式关于附加的秩一更新 suv^* 是仿射的。若 p 和 p_1 分别是当前多项式以及在 $s = 1$ 处的多项式，则 $q = p - p_1$ 给出矩阵铅笔 $p - sq$。

□

相位搜索还需要这个矩阵铅笔的判别式，包括与驻根的碰撞。令 $h_0 = \gcd(p, q)$ 并分解 $p = h_0p_1$，$q = h_0q_1$。由于当前谱是无重根的，h_0无平方因子，且与 p_1 互素。$p - sq$ 的判别式为

$$\mathrm{disc}(h_0)\,\mathrm{disc}(p_1 - sq_1)\,\mathrm{Res}(h_0, p_1 - sq_1)^2.$$

若 $q \ne 0$，第3节计算第二个因子。在结式中，最大公因式(h_0, q_1)的根贡献一个非零常数。模 h_0/最大公因式(h_0, q_1)取商后，元素 p_1/q_1 定义良好。其

特征多项式利用第3节的幂投影计算给出剩余因子，精确到一个非零常数。若需要，可通过在 $s = 0$ 处求值来确定该常数。若 $q = 0$，则完整判别式为常数。

如果当前节点与参考特征值重合，结构化乘积中的分母就会消失。初始的秩一扰动会将两个谱分离。在后续步骤中，对试探半径进行微小调整即可使二者保持不相交。

引理 5.2（与参考谱的分离）。给定 $0 < h < 1$ 和一个已知的 $0 < G_0 \leq |D(B_0)|$，范数至多为 h/n^{10} 的初始秩一扰动会使新谱与参考谱不相交，并保留至少 $1 - 1/(500n^3)$ 的判别式模。对于后续矩阵铅笔 $p - sq$，假设在每个参考节点处都有 $p(\lambda_i) \neq 0$。如果在半径为 h 的圆上的一次试探达到了最大判别式模 M 的至少一半，则可以选择相对半径至多为 $2/(1000n^3)$ 的向外调整，使其模至多减少 $M/32$，并保持在每个参考特征值处特征多项式的模具有正下界。

证明。给定 $0 < G_0 \leq |D(B_0)|$ 和 $0 < h < 1$，设

$$\alpha = \frac{1}{n(1 + \|X_0\|_F^2)}, \qquad \beta = \frac{1}{n(1 + \|W_0\|_F^2)}, \qquad u_0 = \alpha X_0 \mathbf{1}, \qquad v_0^* = \beta \mathbf{1}^T W_0.$$

这些向量的范数至多为一，并且可以通过 $O(n^2)$ 次运算计算得到。在每个参考特征值处的留数为 $\alpha\beta$。对于

$$\sigma = \min\left\{h/n^{10},\, G_0/(2^{10} n^3\, 8^{n(n-1)})\right\}$$

添加 $\sigma u_0 v_0^*$ 后，特征多项式满足 $p_\sigma(\lambda_i) = -\sigma\alpha\beta p_0'(\lambda_i) \neq 0$。在单位参数圆上，矩阵范数至多为三，因此其判别式模至多为 $8^{n(n-1)}$。柯西系数界表明，在 σ处的变化至多为 $2\sigma 8^{n(n-1)}$，这就证明了第一个断言。

对于后续的试探 s_0，对 $0 \leq j \leq 2n$使用 $s_j = s_0(1 + j/(1000n^4))$。判别式多项式的次数至多为 $D = 2n - 2$。根据外部最大值原理，其在 $|s| \leq h(1 + 1/D)$上的模至多为 eM。以向外线段为中心、半径为 $h/(2D)$的圆盘包含于这个更大的圆盘中，从而得到导数界 $6DM/h$，且损失低于 $M/32$。对于每个参考特征值，令 $a = p(\lambda_i) \neq 0, b = s_0 q(\lambda_i)$。我们要求

$$|a - b(1 + j/(1000n^4))| \geq |a|/(10000n^4).$$

若 $|b| \leq |a|/2$，则每个试探都符合条件。否则，相邻值之间至少相隔$|a|/(2000n^4)$，因此至多有一个不符合条件。对 $1000n^4\Re(a/b - 1)$进行二分搜索，可找出其最近的允许整数，并可精确检验这些整数的模平方。对每个参考特征值排除不符合条件的试探，然后从剩余试探中选一个。多点求值之后，此搜索的开销为 $O(n\log\, n)$。一次被接受的步骤至多会使先前的下界缩小 $10000n^4$倍。当被接受的矩阵的范数也不超过二时，其首一特征多项式在 λ_i 处的模是 n 个距离的乘积，每个距离都不超过四。将其模的上界除以4^{n-1}，即可得到参考特征值到当前谱中各点的距离下界。

□

由预解式得到左右向量。参考预解式给出估计谱投影算子所需的右向量、左向量和标量配对。将扰动表示在参考坐标中，得到

$$W_0 B X_0 = \Lambda + QFS^*.$$

我们选择 Q，S 使其具有长度属于 $[1/2, 1]$ 的正交列。为此，在参考特征基中对两组扰动因子应用精确的 Gram–Schmidt 正交化，并省略零残差

并使用第2节中的域归一化。每个基中的坐标是内积除以列长度的平方。将这些坐标映射吸收到 F 中，可得 $\|F\| \le 4\|W_0(B-B_0)X_0\|$，即使原始列线性相关也是如此。令 r，$s \le$ 表示 Q，S 的列数的最小值（k，n）。因此 $Q \in \mathbb{C}^{n\times r}$、$S \in \mathbb{C}^{n\times s}$ 和 $F \in \mathbb{C}^{r\times s}$。下文假设当前谱是无重根的，且与参考谱不相交。特别地，两个秩均为正。

对于精度估计，取一个 2 的幂 H，使其大于 100，n^4、X_0，W_0，F 的范数，以及当前间隙和到参考谱距离的倒数。第7节给出了这样的界，其对数是关于 n 和 $\log(n/\delta)$ 的多项式。对于 $R(z) = (zI-\Lambda)^{-1}$，定义

$$J(z) = S^* R(z) Q, \qquad L_R(z) = I_r - F J(z), \qquad L_L(z) = I_s - J(z) F.$$

对于任意方阵 A，令 $\chi_A(z) = \det(zI - A)$。远离参考谱时，两个行列式都等于 $\chi_B(z)/\chi_{B_0}(z)$。在匹配的根近似值 z_j 处，选择 $\operatorname{adj} L_R(z_j)$ 中包含模最大的元素的一列 c。独立地，按照相同规则，从 $\operatorname{adj} L_L(z_j)$ 中选择一行 d。在原坐标中，对应的列 x、行 y 及标量配对为

$$x = X_0 R(z_j) Q c, \qquad y = d S^* R(z_j) W_0, \qquad yx = -d J'(z_j) c. \tag{5.1}$$

这里 y 是一个行向量，因此 yx 是通常的行向量与列向量乘法。

引理 5.3 (预解向量的精度)。设 $|z_j - \mu_j| \le H^{-60n}$ 对应一个真实特征值μ_j。在 μ_j 处取相同的列和行指标，可得到非零的右特征向量和左特征向量。它们的范数以及 (5.1) 中的配对的相对误差均小于 $1/(100n^2)$。欧几里得向量误差也满足相同的相对误差界。

证明。我们首先在真实特征值处对一个余子式给出下界。在 $|z-\mu_j| \le 1/(2H)$ 上，预解式 R、两个小矩阵及其导数分别有界于 $2H$，H^3，H^4。在 μ_j 处，行列式导数满足

$$|\chi_B'(\mu_j)/\chi_{B_0}(\mu_j)| \ge H^{-2n}$$

这里用到了当前间隙的下界，以及到参考谱距离的上界 4。因此，根据雅可比公式，每个小矩阵中都存在一个模至少为 H^{-5n} 的伴随矩阵元素。在整个圆盘内，伴随矩阵及其导数的范数分别至多为 H^{4n} 和 H^{6n}。为说明这个导数界，将余子式导数展开后，至多有 $n!n$ 个乘积，每个乘积包含至多 $n-2$ 个有界于 H^3 的元素，以及一个有界于 H^4 的导数。标量伴随矩阵为常数。因此，在 z_j 处选取的列和行，在 μ_j 处各自仍保有一个模至少为 $H^{-5n}/2$ 的元素，即使最大值由多个元素同时达到也如此。

在真实根处，伴随矩阵恒等式给出 $L_R c = 0$ 和 $d L_L = 0$。由此可得 (5.1) 中向量满足的特征向量方程。Q, S 在其列空间上的最小奇异值至少为 $1/2$，且 $\sigma_{\min}(R(\mu_j)) \ge 1/4$。因此，每个真实向量的范数至少为 H^{-8n}。在圆盘上，选定指标对应的向量函数范数至多为 H^{6n}，导数至多为 H^{8n}。对于真实的单特征值，根据其多项式乘积公式，谱投影算子的范数至多为$(4H)^{n-1}$。单位左向量与右向量的配对之模是其倒数，且至少为 H^{-2n}，因此真实的未归一化配对的模至少为 H^{-18n}。乘积的导数幅度至多为 H^{17n}。结合这些估计与假定的根精度，即可证明所述结论。如果之后对 c或 d 进行重新缩放，则对其对应的真实比较向量和配对中的相应因子也应用相同的缩放。

□

在所有 n查询点处的 J、J'值需要进行 $O(k^2)$次有理多点求值。对于 J'，我们使用商法则，对分子和分母分别求导。直接计算余子式行列式

采用精确选主元的代价为 $O(nk^5)$，包括查询点z_j 为精确根的情况。因此，薄分解以及对小矩阵进行的所有运算的代价为 $\widetilde{O}(nk^2)+O(nk^5)$。我们以预解向量所指定的精度，对（5.1）中的配对进行求值。

估算最近间隙。我们用最近邻 ℓ^1 距离估算欧几里得间隙。对于四种反射方向，先考虑查询点东北象限中的点。按实部降序排序，并在查询前插入每个完整的并列组。对已排序虚部构建的线段树会返回满足 v 不小于查询点虚部的点（u，v）中，位于不同索引处的两个最小 $u+v$ 值。排除查询点的索引并减去其坐标和，即可得到该象限内的最近距离。对四种方向取最小值，即得 $O(n\log n)$中的精确最近 ℓ^1 距离。该距离介于欧几里得间隙与该间隙的$\sqrt{2}$ 倍之间。将每个节点移动 H^{-60n}，其 ℓ^1 间隙的变化至多为 $2\sqrt{2}H^{-60n}$。

计算所得特征向量矩阵一旦可逆，其残差便给出一个校正，使对角化变得精确。我们在参考基变更过程中构造逆矩阵。

命题 5.4（小校正后的显式对角化）。在关于互不相交的参考谱和当前无重根谱的前述假设下，设 $b\in\mathbb{Z}_{\ge 0}$。可以计算出一个弗罗贝尼乌斯范数至多为 2^{-b} 的校正，并在匹配的域节点处得到显式对角化。只需将节点计算到精度 $2^{-b}H^{-Cn}$，其中 C为充分大的绝对常数。完成标量根近似后，代价为 $\widetilde{O}\left(n^2(k+1)^{\omega-2}+nk^5\right)$。

证明。令 $Z=\operatorname{diag}(z_j)$。我们首先构造候选基。未缩放的坐标数组 $Y_{:j}=R(z_j)Qc_j$ 的位移为 $\Lambda Y-YZ=-Q[c_j]$，因此在左侧进行结构化乘法即可得到原坐标中的列 X_0Y。对于每一列，分量重缩放和平方二分法给出一个上范数估计，其误差因子为 $9/8$。用该估计值分别除以原坐标中的列、参考特征基中的对应列以及其生成元列。从现在起，Y，$X=X_0Y$表示重缩放后的数组。X的各列范数属于 $[8/9,1]$。根据引理 5.3，用相同除数缩放后的真实列范数属于 $[1/2,2]$。缩放之前，残差恰好为

$$(B-z_jI)x_j=-X_0Q\det L_R(z_j)e_{\ell(j)},$$

其中 $\ell(j)$是选定的列索引。残差采用与该列相同的除数进行缩放。因此，$\Delta=(XZ-BX)X^{-1}$ 具有薄因子，其较小的格拉姆矩阵可用于计算其平方弗罗贝尼乌斯范数。事实上，对于 $\Delta=UV^*$，该范数的平方为 $\operatorname{tr}((U^*U)(V^*V))$。通过分块计算 X_0Q 以及稠密矩阵与薄因子的其他乘积，开销为 $O(n^2(k+1)^{\omega-2})$。

为界定校正，考虑真实缩放基 X_*。其第 j个对偶行的范数至多为$2\|P_j\|$，因此$\|X_*^{-1}\|\le 2\sqrt{n}(4H)^{n-1}$。缩放前的导数估计和范数下界给出 $\|X-X_*\|_F\le H^{O(n)}\max_j|z_j-\mu_j|$。利用 $BX_*=X_*\operatorname{diag}(\mu_j)$可对$\|XZ-BX\|_F$得到相同的估计。接着由 Neumann 不等式将 $\|X^{-1}\|$ 界定为 $2\|X_*^{-1}\|$，从而证明所需的校正界。可通过域比较检验互异性和校正范数。一旦这些检验通过，$B+\Delta=XZX^{-1}$ 即可精确成立。

还需要在给定的计算成本内构造逆矩阵。我们首先改用实辅助节点列表 $a_i=8+i/n$, $b_i=12+i/n$。令 $(H_1)_{ij}=1/(a_i-\lambda_j)$，$(H_2)_{ij}=1/(z_i-b_j)$。根据 Cauchy 行列式公式，这两个矩阵都是可逆的。对于 $Y'=H_1YH_2$，恒等式

$$D_aY'-Y'D_b=(D_aH_1-H_1\Lambda)YH_2+H_1(\Lambda Y-YZ)H_2+H_1Y(ZH_2-H_2D_b)$$

生成宽度为 $k+2$ 的生成元，因为两个外部位移的秩均为一。薄因子通过结构化乘积获得。通过分块生成元乘积计算 Y' 的所有元素

并对 $a_i - b_j$ 逐项除法。使用不相交节点列表进行一次结构化乘法，即可得到 $G = Y'^{*}Y'$，包括其对角线。选择实数 a，b 可确保通过伴随运算和乘积法则得到宽度为 $w = 2$ 的等节点位移 $D_bG - GD_b$（$k + 2$）。

要将具有已知等节点位移和对角线的 q 行矩阵相乘，可将其节点分成两半。非对角块具有不相交节点列表，而对角块则继承其对角线。重复进行此细分。与 q 个稠密右端项相乘的代价为 $\widetilde{O}(q^2w^{\omega-2})$：每一层的块行数之和为 q，当某个块首次达到不大于 w 的尺寸时，细分便停止。在该层，方形分块利用生成元构造非对角元素，在等节点块上使用已知对角线，并将所得稠密块与右端项相乘。

要对 G 求逆，将当前块划分为 $\left(\begin{smallmatrix} M & B' \\ C' & D' \end{smallmatrix}\right)$，并将其两个位移生成元划分为 $U = (U_1; U_2)$ 和 $V^* = (V_1^*,\ V_2^*)$。我们必须保留舒尔补的对角线及其生成元，因为等节点位移无法确定对角元素。根据 $D' - C'M^{-1}B'$ 计算该对角线。一旦获得 $C'M^{-1}$，对角线校正便需要计算行列内积，运算量为 $O(q^2)$。舒尔补的生成元因子为

$$U_2 - C'M^{-1}U_1, \qquad V_2^* - V_1^*M^{-1}B'.$$

要验证此公式，在左侧用 $[-C'M^{-1}, I]$ 乘以完整位移矩阵，并在右侧用 $\left(\begin{smallmatrix} -M^{-1}B' \\ I \end{smallmatrix}\right)$相乘。因此，传递给每次递归求逆的生成元宽度仍为 w。显式求出一个块的逆矩阵后，将其旧生成元在两侧分别乘以该逆矩阵，并将其中一个因子取负，即可得到逆矩阵的生成元。根据乘积法则，非对角乘积 $C'M^{-1}$和 $M^{-1}B'$的端点节点列表互不相交，宽度至多为 $2w$。将结构化非对角因子与稠密逆矩阵相乘，计算这些乘积的元素。这些表示给出了舒尔补和四个标准逆矩阵分块，包括 $M^{-1} + (M^{-1}B')(D' - C'M^{-1}B')^{-1}(C'M^{-1})$。根据乘积法则，中间生成元的宽度至多为 $3w$。保留显式构造的逆矩阵分块的对角线，以便后续计算相同节点的乘积。当块阶数不超过 w时，停止递归求逆，并在此处使用稠密求逆。在最后一次分割中，两个子块的阶数与 w相当，因此可通过方形分块，利用生成元和对角线还原其元素。对于更大的阶数，按块分解时，将稠密矩阵乘以瘦生成元的每次运算开销为 $O(q^2w^{\omega-2})$。根据前述细分，递归节点处其余乘积的开销为 $\widetilde{O}(q^2w^{\omega-2})$。各块大小的平方构成一个几何级数。由于格拉姆矩阵正定，所有所需的主元均存在。如果初始阶数不超过 w，格拉姆矩阵已显式给出，稠密求逆的开销为 $O(n^{\omega})$。通过递归舒尔补分解和方形矩阵乘法对正定块进行稠密求逆，具有这一开销。格拉姆矩阵递归中的主元选取失败，表明暂定矩阵奇异。

最后，使用结构化乘子构造出 $Y^{-1} = H_2G^{-1}Y'^{*}H_1$。在 $\Lambda Y - YZ$ 两侧乘以 Y^{-1} 并取负，可得到 $ZY^{-1} - Y^{-1}\Lambda$ 的生成元。最后进行一次不相交节点乘法，得到 $X^{-1} = Y^{-1}W_0$。

□

6 原坐标中的范数

参考特征向量不必正交，因此坐标向量的长度不必等于其所表示的特征向量的范数。在每次更新后将参考基应用于所有向量，会损失低秩带来的节省。我们估计它们的所有范数，使每个参考列具有不同的精度，并且仅在选定目标处将向量恢复到原坐标中。

设参考矩阵 $B_0 = X\Lambda X^{-1}$ 具有 $|\lambda_i| < 2$、$1/2 \le \|x_i\| \le 2$，并令 w_i^* 表示 X的对偶行。这里，X 是参考基，在第5节中记为 X_0，且 $W_0 = X^{-1}$。令$s_i^{(0)} = \min_{j\neq i} |\lambda_i - \lambda_j|$。所需的精度将由满足以下条件的二的幂 b_i 控制

$$b_i \ge 20\log_2(2n) + \log_2^+ \|w_i\| + \log_2^+(1/s_i^{(0)}).$$

通过连续平方和平方比较夹逼，可在常数因子范围内得到这些指数的右侧值。在本节中，始终假设 $\sum_i b_i \le CnL$，其中 $L \ge \log_2(2n/\delta)$，且 C 为绝对常数。将 b_i 增加一个绝对因子后，归一化左列也满足相同的假设：无论取哪一侧，对偶范数都与同一个投影算子范数可比。我们只对参考特征系统要求这个求和界。引理 7.1 证明，只要局部算法从一个参考基开始处理批次，这个假设就成立。

在固定精度下，我们可以将矩阵和向量的舍入后元素缩放为有界整数。在单位成本约定下，我们可以将矩阵的每一列编码为一个整数。对这些整数作线性组合后，其各位数字便给出乘积的元素，至多相差已知的加性修正项。

引理 6.1 (整数打包)。设 $H \ge 1$ 为一个整数，M 为一个固定的 $a\times b$ 整数矩阵，其 a、$b \le n$，且元素属于 $[-H, H]$。对其各列进行预处理后，对于任意 *2* 的幂 $D > 4nH^2$，与整数向量 $v \in [-H, H]^b$ 的乘积可通过 $O((a+b)(1+\log D))$ 次域运算和比较计算得到。

证明。将行从零开始编号，并保留 $C_j = \sum_i (M_{ij} + H)D^i$ 以及行和。$\sum_j C_j(v_j + H)$ 的以 D 为底的展开中，位置 i 上的数字为 $\sum_j (M_{ij} + H)(v_j + H)$，因为该和位于 $[0, D)$ 中，且不会产生进位。从最高位开始，通过二分搜索找出每个数字，并减去其 D^i 的倍数。行和与 $\sum_j v_j$ 可以消除偏置。预处理需要 $O(a)$ 次运算来制表列出二的幂，并需要 $O(ab)$ 次运算来插入已量化的项。

对于幅度为 $O(p)$ 且具有精度位的实数定点输入，先按二的幂缩放，再使用 $O(p)$ 的平方或普通比较来量化每个新的标量。如果 $p \ge \log_2(2n)$，预处理后每次乘积的运算成本为 $O((a+b)p)$。复数对只需四次实数乘积。 □

引理 6.2（参考基中的符号方向）。在上述关于 X 和 b_i 的假设下，设 $\epsilon = 2^{-40}n^{-6}(1+L)^{-2}$。经过 $\widetilde{O}(n^2L)$ 预处理后，以下构造需要 $\widetilde{O}(nm + nL)$ 次运算。给定观测精度达到引理 *2.2* 要求的 m 对当前单位特征向量方向，可得到符号向量 $\bar{u}$、$\bar{v}$，其条目为 $\pm 1/(2s)$，如该引理所述；还可得到参考坐标中的域向量$\mathbf{a}$、$\mathbf{b}$。方向 $u = X\mathbf{a}$ 和 $v^* = \mathbf{b}^*X^{-1}$ 满足

$$\|u - \bar{u}\|,\ \|v - \bar{v}\| \le \epsilon, \qquad \|u\|,\ \|v\| \le 1.$$

这里，P_i 是当前矩阵在某个选定标签处的谱投影算子。在至少 $m/25$ 个这样的标签上，$|v^* P_i u| \ge \|P_i\|/(2^{13}n)$。

证明。设 $W = X^{-1}$，并将其各行记为 w_i^*。选择一个二进制幂次 $e \in [\epsilon/(32n^2),\ \epsilon/(16n^2)]$。在 W 的第 i 行上，以网格 e 量化实部和虚部；在 X 的第 i 列上，以网格 $e\,2^{-b_i}$ 量化。将所得数组记为 $\widehat{W}$，$\widehat{X}$。由于 $\|w_i\| \le 2^{b_i}$ 和 $\|x_i\| \le 2$，缩放后组 $b_i = b$ 中的整数条目需要 $O(b + \log(1/e))$ 位。逐项二分搜索和列打包因此需要 $O(n\sum_i (b_i + \log(1/e))) = \widetilde{O}(n^2L)$。

将引理 2.2 应用于观测向量 $\widehat{W}\bar{u}$ 以及 $\mathbf{b}^* = \bar{v}^*\widehat{X}$。为得到其中任一乘积，按 b_i 对输出项分组，并在除以 $2s$之前计算符号乘积。对每个实数整数数组组，使用引理 6.1 的打包列 C_j。对于未缩放的输入符号

σ_j，在 O 中用 $\sum_j C_j(1+\sigma_j)$ 次运算构造。减去已知的行和与偏置后，各位数字确定带符号行乘积。对于绝对值不超过 H 的项，基数大于 $4n(H+1)^2$ 即足够。恢复一位数字需要 $O(b+\log(1/e))$ 次比较。由于 b_i 是二的幂，因此共有 $O(\log(2+nL))\tilde{}$ 组。选定符号后的总查询成本为 $O(nL)$。

在位置 i，$\mathbf{a}$ 和行向量 $\mathbf{b}^*$ 中的误差分别至多为 $2ne$ 和 $2ne\,2^{-b_i}$。将第一组误差分别与 X 的各列相乘求和，并将第二组误差分别与 W 的各行相乘求和，即可得到所述欧几里得误差。名义方向的范数至多为 $1/2$。对于成功的指标，两种实际重叠与单位特征向量的模都至少为 $1/(16s)-\epsilon \ge 1/(32s)$。它们的乘积乘以 $\|P_i\|$ 为 $|v^*P_iu|$，且 $s<2\sqrt{n}$。两个参考因子 $\mathbf{a}$，$\mathbf{b}^*$ 精确地定义了秩一更新，因此无需将任一因子与稠密基相乘即可构造其矩阵铅笔。 □

引理 6.3 (批次中的 Frobenius 配对)。 设 F_0 是一个满足 $\|F_0\|_F \le \delta$的显式矩阵。在一个批次中，设 $F = F_0 + \sum_{j=1}^{l} t_j \bar{u}_j \bar{v}_j^*$、$l \le k \le n$具有引理 *6.2* 中的符号方向。对于新的符号方向 $\bar{u}$、$\bar{v}$ 以及满足 $|c|\delta \le 2^{10}n^2(1+L)$的实数 c，数值 $c\,\mathrm{tr}(F^*\bar{u}\bar{v}^*)$ 可在 $\widetilde{O}(nL+nk)$ 次运算内计算得到，绝对误差不超过 $1/100$，预处理需 $\widetilde{O}(n^2L)$ 次运算。批次结束时，显式矩阵 F 需要 $O(n^2k^{\omega-2})$ 次运算。

证明。 将 F_0/δ 的实部和虚部量化到误差不超过 2^{-p}，其中 $p = 40+10\lceil \log_2(2n(1+L))\rceil$。按照引理 6.1 对其各列进行打包。将该矩阵作用于 $\bar{v}$ 的未缩放符号，除以 $2s$，再与 $\bar{u}$ 收缩。共轭并乘以 $c\delta$ 后，误差至多为 $2|c|\delta n 2^{-p} < 1/100$。每个存储项的贡献恰好为 $ct_j(\bar{u}_j^*\bar{u})(\bar{v}^*\bar{v}_j)$。最后，将两个 $n\times l$ 的方向数组相乘，并将标量 t_j 吸收到第一个数组中。将两个外部维度按长度为 l 的区间分块，即可得到所述计算成本。当 $l=0$ 时，初始矩阵本身已能给出结果。

□

对于三角分解，我们按所需精度从高到低依次选取各列。两遍正交化既能控制正交性损失，也能控制重构每一列时的误差。

引理 6.4 （具有单独精度的三角分解）。 对参考列和对偶行进行置换，使得 $b_1 \ge \cdots \ge b_n$。对参考特征值以及这些坐标下每个分子因子的行应用相同的置换。目标标签保持原有顺序。右列构造和共轭左列构造各自对其参考数据采用一致的置换。通过 $O\left(n\sum_i b_i\right)$ 次运算（忽略固定的对数因子），可以计算出Q、$R\in\mathbb{C}^{n\times n}$，其中 R 为上三角矩阵，使得

$$|(Q^*Q-I)_{ij}| \le 2^{-2b_j} \qquad (i\le j), \tag{6.1}$$

$$\|x_j - QR_{:j}\| \le 2^{-8b_j}. \tag{6.2}$$

特别地，对于每个 y，

$$\|QRy - Xy\| \le \Big(\sum_i 2^{-8b_i}\|w_i\|\Big)\|Xy\|, \qquad \|Q^*Q-I\| \le n2^{-2\min_i b_i}. \tag{6.3}$$

证明。 令 q_j 为 Q 和 $Q_{<j}=[q_1,\dots,q_{j-1}]$的列 j。对于列 j，对 x_j 相对于 $Q_{<j}$进行两遍正交化。在当前残差 u上，每一遍都会计算 $a \simeq Q_{<j}^*u$，并减去 $Q_{<j}a$的打包近似。对于有界输入，分量精度 2^{-30b_j} 可确保每次乘积的欧几里得误差小于 2^{-20b_j}，其中包括维度因子。计算得到的两个向量 a 之和构成当前列 R的严格上三角部分。

我们首先证明残差始终有一个严格大于零的下界。归纳地，令 T 将 x_i映射为 $QR_{:i}$ ，其中 $i<j$ ，并固定其他基向量列。重构界和对偶性给出$\|T-I\| \le \sum_{i<j} 2^{-8b_i}\|w_i\| \le n2^{-7b_j}$。由于 w_j^*消去前面的列以及 $w_j^* x_j = 1$，x_j 到它们张成空间的距离至少为 $1/\|w_j\|$。由于三角对角元非零，它们变换后的张成空间等于 $Q_{<j}$ 的张成空间。由于 T固定了 x_j，它到该张成空间的距离至少为 $(1-n2^{-7b_j})/\|w_j\|$。两次向量减法产生的误差范数之和小于 2^{-18b_j}，因此最终残差的范数至少为 $0.8\,2^{-b_j}$。

设 $e = \|Q_{<j}^* Q_{<j} - I\| \le n2^{-2b_j}$。每次精确遍历都会将内积向量乘以$I - Q_{<j}^* Q_{<j}$，且 $\|I - Q_{<j}Q_{<j}^*\| \le 1$。由于 $\|x_j\| \le 2$，这些界使残差和系数范数保持在 4 以下，满足打包估计的要求。经过两次遍历后， $\|Q_{<j}^* u\|$ 至多为$4e^2 + O\left(2^{-20b_j}\right)$ 。平方比较通过 $d_j > 0$ 计算 $\|u\|$ 的近似值，其相对误差为 2^{-10b_j}，需要进行 $O\left(b_j\right)$ 次运算。令 $q_j = u/d_j$、 $R_{jj} = d_j$。除以 d_j 后，内积误差小于 2^{-2b_j}，且平方列范数具有相同的误差界。由于 R 使用计算得到的系数 a，重构误差仅受两次向量相减所产生的误差影响。这就证明了（6.1）和（6.2）。对列误差求和，即可由 $y_i = w_i^* X y$ 推出（6.3）中的第一个界。

对于成本界，在达到 b_j之前的每个所需二进制精度下，为 $Q_{<j}$ 和 $Q_{<j}^*$ 保留打包列，并为每种精度使用固定基数。追加 q_j 会在第一个表中增加一列，并在第二个表的 n 列中各增加一位。活动二进制精度之和为 $O(b_j)$。因此，每一步（包括量化）的成本至多为 $O(nb_j)$，忽略对数因子。 □

我们将这一因子分解应用于 n 个目标点 $|z| \le 2$，这些目标点彼此不同，且与$\{\lambda_i\}$不相交。在每个目标点处，原坐标中的坐标列及其范数具有如下形式

$$y_i(z) = \frac{(Ac)_i}{z - \lambda_i}, \qquad P = \|Xy(z)\|. \tag{6.4}$$

这里， A 有 $1 \le r \le n$ 列，而 c 是一个依赖于 z的域向量。对于（5.1）中的右向量， $A = Q$ 是预解因子。然而，在本节其余部分， Q， R 表示引理 6.4 的三角因子。

对于（5.1）中的左行向量，将该构造应用于其共轭转置。随后，目标点和参考节点均取共轭，选定的余子式行则替换为其共轭转置。设 D 为正对角矩阵，用于将 W_0^*的列归一化到规定的常数因子范围内。它在原坐标中的基为 $W_0^* D$，而该基下的分子因子为 $D^{-1}S$。在这两种情况下，（6.4）中的所有数据（包括计算得到的目标点 z）都是精确域值。

我们精确计算高精度行向量的三角坐标。对于其余行向量，我们根据格拉姆矩阵的矩估算它们对平方范数的贡献。

命题 6.5 （估算所有欧几里得范数）。 令 $\tau \ge 1$ 为 *2* 的幂，并令 $d = \#\{i:\ b_i > \tau\}$。在对参考特征系统进行 $\widetilde{O}\left(n\sum_i b_i + n^2\left(\tau + \log n\right)\right)$ 预处理之后，(6.4) 中的 n 平方范数可以估算至相对误差 $1/20$ 以内

$$\widetilde{O}\big(n\max(d,r)r^{\omega-2} + n(r+1)^2(\tau+\log n)^5\big)$$

次域运算。此处及下文中，对数因子均包含 $1 + \log(n/\delta)$ 的幂，只要 L在该应用中出现。

证明。 求和界给出 $d = O(nL/\tau)$。我们首先在每个目标点处准确计算 $Ry(z)$ 的初始 d 坐标。其列为 $y(z)$ 的矩阵具有西尔维斯特位移 $-A[c]$

相对于参考节点和目标节点。第5节中的转置结构化乘法给出了成本中的第一项。根据上三角性，带有 $i \le d$的源列在前 d 个输出坐标以下不产生任何贡献。

对于其余坐标，令 R_t 为行、列索引均大于 d的子矩阵。引理 6.4 给出 $\|R_{:i}\| \le (\|x_i\| + 2^{-8b_i})/\sigma_{\min}(Q) < 3$。因此， $R_t^* R_t$ 的元素的模至多为九，且$\|R_t\| \le 3\sqrt{n}$。使用 $p + 20\lceil \log_2(2n) \rceil$ 个小数位量化 R_t，并将其打包伴随矩阵作用于每一列。这样可以用 $O(n^2 p)$ 次运算，将 $G = R_t^* R_t$ 近似到绝对元素误差 2^{-p}，其中 $p = C_1(\tau + \lceil \log_2(2n) \rceil + 64)$，绝对常数将在下文选定。该格拉姆矩阵仅取决于参考基。

我们还需要低位源行 $i > d$上分子的有界表示。令$r_0 = \operatorname{rank}(A_{\mathrm{low}}) \le r$。精确进行格拉姆–施密特正交化并省略线性相关列，可得 $A_{\mathrm{low}} = U H_A$，且 $U \in \mathbb{C}^{(n-d)\times r_0}$ 和 $H_A \in \mathbb{C}^{r_0 \times r}$。保留 U各行原有的源标签 $i > d$。在对非零残差的范数进行夹逼之前，先将其缩放，使其最大实部或虚部的模为一。因此， U的列可以在 $O(nr^2 + r \log n)$ 次运算内缩放至长度属于 $[1/2, 2]$，而不受 A_{low} 奇异值大小的影响。对于 $c' = H_A c$，我们得到

$$|y_i| \le \|w_i\| P \le 2^{\tau} P, \qquad \|c'\| \le 2\|A_{\mathrm{low}} c\| \le 8\sqrt{n}\, 2^{\tau} P \quad (i > d), \tag{6.5}$$

使用 $|z - \lambda_i| \le 4$。空坐标集和秩为零的分子贡献为零。

现在，我们对源节点分组，使其倒数分母可以进行短展开。用一个固定的正方形覆盖节点圆盘。将每个非空的低层源单元细分为四个全等正方形，直到边长 w介于 $2^{-\tau}/4000$和 $2^{-\tau}/1000$之间。所得的树有 $T = O(\tau + \log n)$ 层。按照固定的半开约定，将边界点分配给唯一的子单元（包括初始正方形的外侧边缘）。通过中点比较，在 $O(T)$ 次运算内确定每个源节点所属的路径。每一层中占用的单元都用其整数行列索引存储；这些索引通过将父单元的索引加倍并加上子单元的偏移量得到。对这些列表排序后，即可通过二分搜索进行邻近查询，所需比较次数为 $O(\log(2n))$。对于一个目标点，若单元 I 的边长为 w_I 且中心为 a_I，并且满足 $|z - a_I| > 8w_I$，则接受该单元；否则将其细分。在每一层中，靠近目标点的单元数量有界，因此下一层中可接受的子单元数量也有界。在最细层级，所有剩余的邻近点都在目标点的 $9w$ 范围内。由于这些点之间至少相隔 $2^{-\tau}$，因此至多只有一个点。使用其单位坐标特征，并精确计算其标量 y_i。

在一个接受单元上，将倒数分母近似为

$$\sum_{j=0}^{s-1} \left(\frac{\lambda_i - a_I}{w_I} \right)^j \frac{(w_I/(z - a_I))^j}{z - a_I}, \qquad i \in I. \tag{6.6}$$

相对误差为 $|(\lambda_i - a_I)/(z - a_I)|^s \le 8^{-s}$。因此，剩余坐标向量中的误差至多为 $8^{-s}\sqrt{n}\, 2^{\tau} P$。取 $s = 128(\tau + \lceil \log_2(2n) \rceil + 20)$ 后，乘以 R_t 后的尾部误差小于 $P/1000$。当消去使真实尾部为零时，该估计同样成立。

这些展开给出一小组向量，其 Gram 矩阵决定了近似尾部范数。对于单元 I，使用在 I 上、对于 $1 \le h \le r_0$ 和 $0 \le j < s$ 具有条目 $U_{ih}((\lambda_i - a_I)/w_I)^j$、其他位置为零的特征。加入最细层单例单位特征。每个目标点使用 $O((r+1)(s+1)(T+1))$ 个特征。我们只需要计算可能在同一目标点处出现的成对特征的矩 $f_J^* G f_I$。

设 I 为较粗单元。父单元处的近邻测试使两个中心都位于目标点的 $O(w_I)$ 范围内。因此， J位于 $N(I)$ 的并集中；该并集由 I 层级上的固定数量的相邻网格单元构成。这些邻域在每个层级上都有有界重叠。保留

将舍入子矩阵 $G_{N(I),\ I}$ 以打包形式保存。由于其元素已经量化，每个参考基只需对整棵树插入一次，耗时为 $O(n^2T)$。

对于 I 上的每个特征，打包乘积给出 $G_{N(I),\ I} f_I$。在每个更细或相同的层级上，对于每种 J 特征类型，扫描 $N(I)$ 中的源，并按单元 J 累积内积。对于每个更粗的层级，对 I 求和得到的扫描长度为 $O(n)$。在每个层级保持非空单元键有序，我们得到计算成本

$$\widetilde{O}\big(n(r+1)^2(s+1)^2(T+1)^2(p+\log n)\big).$$

对于参考基处的每次更新，都会重复使用树和打包的 Gram 子矩阵。更新后，只需重新计算涉及 U 的特征。

接下来需要控制这些矩中的舍入误差。通过特征系数的二次型的实部估计尾部平方范数。取实部不会增大误差。根据 (6.5)，最小叶大小和 $|w_I/(z-a_I)|<1/8$，每个远端系数的模至多为 $2^{2\tau+16}nP$。每个近端系数恰好为 y_i，并且由 (6.5) 给出界。所有特征项的模都至多为二。

使用 $p+20\lceil\log_2(2n)\rceil$ 个小数位对有界特征、Gram 条目和打包乘积输入进行量化，可将每个矩误差限制为 $128n^2 2^{-p}$。每个压缩算子至多包含 n^2 项，额外的 $20\lceil\log_2(2n)\rceil$ 个位用于吸收两个特征因子中的误差。系数至多有$2^{14}n(s+1)(T+1)$ 个。因此，二次型误差至多为

$$2^{4\tau+67}n^6(s+1)^2(T+1)^2 2^{-p}P^2.$$

取 $C_1=4096$ 可使该值小于 $P^2/1000$，即使存在消去也是如此。将其与截断误差、精确的前 d 个坐标，以及（6.3）中的近等距性和重构界结合起来。在原坐标中，平方范数的最终绝对误差至多为 $P^2/20$。计算这 n 个二次型由矩构造界涵盖。如果没有尾部坐标，则省略尾部计算。如果 $d=0$，则没有需要精确计算的初始坐标。

□

选定目标点后，符号构造还需要它们在原坐标中的左右方向。

命题 6.6（恢复选定方向）。对于（5.1）中的未缩放候选值（由（6.4）表示），在满足引理 *5.3* 的精度假设时，任选满足 $P>0$ 的 $1\le m\le n$ 个目标点。完成参考特征系统的预处理后，可在尺度 $P'\in[P/2,2P]$ 下恢复这些目标点在原坐标中的向量，误差小于 n^{-3}，其中 ℓ^2，适用于 Xy/P'。除当前分子的范数估计和 $\widetilde{O}$（$nr^2+r\sum_i b_i$）次运算外，所需成本为

$$\widetilde{O}\left(m\Big(\sum_i b_i + r^2\log(2+\max b_i)+n\log n+\log(2+n\log H)\Big)\right).$$

证明。命题 6.5 通过二进尺度上的平方比较给出 P'。对于未缩放候选值，引理 5.3 将 P 和 $1/P$ 均界定为不超过 $H^{O(n)}$。将范数估计与二进尺度的平方进行比较；若估计值小于 1，则先对其取倒数。通过连续平方夹定二进指数。按降序使用存储的幂次，可在 $O(\log(2+n\log H))$ 次运算内确定该指数。在对数尺度上最接近估计长度的二进尺度与该长度相差不超过因子 $\sqrt{2}$，因此与 P 相差不超过因子 2。几何中点处的比较直接使用平方范数估计。

将具有相同 $b_i = b$ 的索引 J 分组。与低阶行的处理相同，构造秩不超过 $\min(r, |J|)$ 的 $A_J = U_J H_J$，其中秩为 r_J。因子分解的代价为 $O(nr^2 + r\log\, n\#\{J\})$，而瘦基的量化与打包需要 $O(r\sum_i b_i)$ 次运算。

对于每个选定目标点，用 $O(rr_J)$ 次运算计算 $c_J = H_J c$。式 (6.5) 的论证给出$\|c_J\|/P' \le 16\sqrt{n}\,2^b$。扫描组节点即可找到距离 z不超过 $2^{-b}/4$的任何极点。这样的极点至多有一个，因为两个组节点之间至少相距 2^{-b}。根据式 (6.4)，可精确计算该极点处的 y_i/P'。在其他节点处，倒数分母至多为 $4\cdot 2^b$。

因此，在 $O(b)$ 定点精度下计算乘积 $U_J(c_J/P')$，然后进行精确除法，可使每个归一化坐标的绝对误差至多为 n^{-12}。中间幅度和倒数分母各需要 $O(b + \log\, n)$ 位，而 $b \ge 20\log_2(2n)$。此外，真实的归一化坐标的模至多为 2^{b+1}，因为 $|y_i| \le \|w_i\|P$。每组的打包乘积耗费 $O((|J| + r_J)b)$，而近极点处的精确行向量耗费 $O(r)$。组的数量至多为$O(\log(2 + \max_i b_i))$。

令 $d_b = \#\{i : b_i \ge b\}$。$R$ 的组列在第 d_b 行以下均为零，因此在 $O(b)$ 位精度下，通过一次打包矩阵-向量乘积计算它们对 Ry/P' 的贡献。系数有界，且输入的取值范围为 2^{b+1}。当绝对精度常数足够大时，每个块中的总坐标误差（包括输入误差）至多为 n^{-9}。计算量为 $O((|J| + d_b)b)$。每个参考基只需以 $O(n\sum_i b_i)$ 次运算的代价打包这些块一次。由于精度是二进制幂次，$\sum_b d_b b \le 2\sum_i b_i$。

将各块的贡献相加。由 (6.3)，$\|Ry/P'\|$ 有绝对常数上界。因此，只需在 $O(\log n)$ 位精度下用 Q 进行一次最终的打包乘积。增大绝对精度常数后，累积的欧几里得误差（包括 (6.3) 中的误差）将小于 n^{-3}。若要获得指定的固定前置因子，则在进行二进制舍入前，还应将 b_i 的下界增加一个绝对常数。这样可将 (6.3) 中的重构误差控制在相同的前置因子以内，并保持 $\sum_i b_i = O(nL)$。 □

令 P_i 和 s_i 分别表示当前矩阵在第 i 个匹配特征值处的谱投影算子和最近间隙。为估计 $\|P_i\|/s_i$，取该目标处（5.1）中的计算列向量 x_i 和行向量 y_i。令 $N_{R,i}$,$N_{L,i}$ 表示它们的范数平方估计，并令 $d_i > 0$ 表示第 i 个计算节点与其在 ℓ^1 中最近邻之间的距离，该距离按照第5节中的方法获得。在范数估计和全精度配对中使用相同的列向量和行向量尺度，定义

$$\widehat{K}_i^2 = \frac{N_{R,i} N_{L,i}}{|y_i x_i|^2 d_i^2}.$$

根据引理 5.3 和命题 6.5，K_i 与 $\|P_i\|/s_i$ 相差不超过因子 4。实域运算给出配对的模平方，所有比较均可使用 $\widehat{K}_i^2$ 进行。

对于选定目标，命题 6.6 给出的向量在观测误差之前，其范数处于两个固定正数常数之间。使用允许的精度 $10^{-6}n^{-3}$，并通过有界二分法进行归一化。引理 5.3 结合相同的节点精度或更大的绝对指数，以及各自的观测界，给出与真实方向相差 $1/(100n^2)$ 以内的单位方向。随后，引理 6.2 以所述成本为附近的实际方向构造两个名义符号方向和精确的参考因子。

对于迭代，设 $r = O(n^b)$，并选择 τ，使其与 $n^a L$ 相差不超过一个固定因子，其中 $0 < a < 1$ 和 $0 < b \le 1/2$ 为固定值。范数估计的成本为

$$\widetilde{O}\big(n^{2-a+(\omega-2)b} + n^{1+(\omega-1)b} + n^{1+2b+5a}\big)$$

每次更新，在为参考基完成 $\widetilde{O}(n^{2+a})$ 准备工作后。在应用日志中为 H =$(256nL)^8$ log2，因此额外的尺度搜索对每个目标点的成本为 $O(\log(nL))$。由于 $\sum_i b_i = O(nL)$，并且

$r^2 \le O(n)$，恢复任意 m 选定方向的成本为 $\widetilde{O}(nm)$，此外还需为其当前分子付出 $\widetilde{O}(nr^2 + rnL)$ 的成本。不要求 m与 r 之间存在任何关系。我们使用

$$a = \frac{23}{192}, \qquad b = \frac{25}{144}, \qquad \rho = \frac{545}{576}.$$

矩阵乘法界 $\omega < 2.38$ 由 [1] 给出

$$2 - a + \frac{19}{50}b = 1 + 2b + 5a = 2 + a - b = 1 + \rho.$$

因此，范数估计的每次更新成本为 $\widetilde{O}(n^{1+\rho})$，而在 $k = \lceil n^b \rceil$ 次更新中分摊准备工作的成本也具有相同的界。

7 迭代与正则化定理的证明

在每个阶段，我们将每个谱投影算子的范数与其特征值到谱中其余部分的距离进行比较。如果超过局部界的比值足够少，则其他指标对应的右特征向量张成一个我们可以保留的不变子空间。这样，我们只需进行一次递归调用，作用于末尾的对角块。

在每次调用中，输入是一个阶数为 n 的压缩算子，扰动容差为 $0 < \delta < 1/2$。顶层调用的深度为零，每次递归调用都会使深度增加一。当维度为零时，调用返回空作用和基。当维度为一时，调用返回给定的标量和单位列向量及其逆矩阵。对于 $n \ge 2$，如果其深度为零或一且 $n \ge 2^{20}$，则调用会得到加速。在这两个深度下，分别令

$$\alpha = \begin{cases} 947/1000, & \text{at depth zero,} \\ 39/50, & \text{at depth one,} \end{cases} \qquad m = \lfloor n^{\alpha} \rfloor.$$

所有其他调用均为终止调用，然后 $m = 1$。定义

$$\begin{gathered} L = \lceil \log_2(64n/\delta) \rceil, \qquad d = n(n-1), \qquad T = 2^{16} n^2 L, \\ \nu = 2^{24} n^2 L/\delta^2, \qquad \eta = \delta/(2^{30} n^2 L), \\ G_0 = (\delta/(8n))^d/16, \qquad H = 2^{(256nL)^8}. \end{gathered} \tag{7.1}$$

选择最大的二进半径 $r = 2^j$， $j \in \mathbb{Z}$，使得

$$r^2 \le 2^{-40} \frac{\delta^2 m}{n^2 L}, \tag{7.2}$$

并令

$$M = \frac{2^{20} n}{r}, \qquad Q = 32M, \qquad Q_{\mathrm{term}} = \frac{2^{12}}{r}, \qquad k = \lceil n^{25/144} \rceil. \tag{7.3}$$

因此 $0 < r < 1$，并且

$$Q \le 2^{70} \frac{n^2 \sqrt{L}}{\delta \sqrt{m}} \quad \text{in an accelerated invocation,} \qquad Q_{\mathrm{term}} \le 2^{70} \frac{n \sqrt{L}}{\delta} \quad \text{in a terminal invocation.} \tag{7.4}$$

该算法通过 r 平方比较进行选择。平方根仅出现在界中。在一次加速调用中，取满足 $\alpha = u/v$ 的正整数 u， v。在正整数中， m 是

满足 $m^v \le n^u$ 的最大值，而 k 是满足 $k^{144} \ge n^{25}$ 的最小值。对于 τ 的二进尺度，同样可以通过将 τ^{192} 与 $n^{23}L^{192}$ 进行比较来选择。一个批次的初始扰动会额外贡献一列，因此低秩子程序至多使用 $k+1$ 个扰动列。

将命题 4.1 应用于本次调用的输入，其参数为 δ。将其输出记为 B_{s}，并将其右列缩放至长度属于 $[1/2,1]$，同时对相应的对偶行应用逆缩放。它与输入的距离小于 $\delta/4$，其最小间隔至少为 $\delta/(8n)$。除当前矩阵 B 外，保留一个初始为零的矩阵 $\mathcal{E}$，用于近似 $B - B_{\mathrm{s}}$。我们使用以下势

$$\Phi(B,\mathcal{E}) = \log|D(B)| - \nu\|\mathcal{E}\|_F^2 \tag{7.5}$$

来控制位移和更新次数。为了在一个批次中评估依赖相位的惩罚项，我们只需计算一个弗罗贝尼乌斯内积。令 $B_0 = X\Lambda X^{-1}$ 为参考矩阵，其对角元素为 $\Lambda = \mathrm{diag}(\lambda_1,\dots,\lambda_n)$。令 x_i 为 X 的第 i 列，令 w_i^{*} 为 X^{-1} 的第 i 行。初始时 $B_0 = B_{\mathrm{s}}$。符号 B_0, X, Λ 会在每次参考对象变更时更新。在参考对象变更之间，以参考坐标存储每个精确的秩一更新。若要变更参考对象，则应用命题 5.4；该命题要求校正的弗罗贝尼乌斯范数至多为 η，且校正后的判别式模至少为校正前数值的 0.9 倍。矩阵 $\mathcal{E}$ 不受此校正影响。

7.1 局部算法

在单次更新和批次之间进行选择，取决于在参考基中计算出的权重和。较大的权重和可确保在 Φ 中获得足够的收益，以抵偿重新计算特征系统的开销。较小的权重和则限定了继续使用当前基所需的精度。

在一次加速调用中，重复执行以下操作。

1. 将右列归一化为长度为 $[1/2,1]$，并计算

$$K_i^2 = \|x_i\|^2\|w_i\|^2 / s_i^2$$

精确计算，其中 s_i 为参考谱中的最近间隙。如果超过 Q^2的值不多于 m 个，则返回参考矩阵及其特征系统。否则，赋予权重

$$\gamma_i = \begin{cases} \lfloor \log_2(rK_i/(2^{16}n)) \rfloor, & K_i > 2^{20}n/r, \\ 0, & K_i \le 2^{20}n/r, \end{cases} \qquad W = \sum_i \gamma_i .$$

对于正权重，将 $(rK_i/(2^{16}n))^2$ 与 4 的幂进行比较。通过连续平方确定指数的范围，再按降序排列的已存储幂，找出其值不超过平方比值的最大整数指数。

2. 如果 $W \ge n$，执行以下单次更新，更改参考，然后返回第一步。

3. 如果 $W < n$，请按照第 6 节的说明对两个参考基进行预处理，包括引理 6.2 的符号变换表。应用引理 5.2 的初始扰动，并将其半径参数替换为 $\min(r,\ \eta)$，同时保持 $\mathcal{E}$ 不变。其范数至多为 η。在一个批次内最多执行 k 次更新。每次更新前，计算当前特征多项式、匹配根近似值、欧几里得范数估计、传递配对和最近距离估计。使用$\widehat{}$ $\widehat{}$一致的缩放，得到满足 $K_i/4 \le K_i \le 4K_i$的 K_i，其中 K_i 现在指当前$\widehat{}$谱中与之匹配的标签。标记 $K_i > 4M$。如果标记的索引不超过 m 个，则保留所有

未标记标签，将参考更改为满足以下测试的对象，并返回其特征系统。否则，重构任意 m 标记的配对并执行批量更新。经过 k 次更新后，更改参考并返回第一步。

终止调用则使用以下过程，直到每个 $K_i \le Q_{\mathrm{term}}$。最后一个小节将确定是否递归以及如何组装输出。

单次更新的方向。 在引理 2.2 的第一个断言中，使用权重 γ_i选择符号。令 s 为满足 $s^2 \ge n$的最小整数幂，于是 $\sqrt{n} \le s < 2\sqrt{n}$，并将两个符号向量重新缩放，使其各项为 $\pm 1/(2s)$。它们的范数至多为 $\overline{1}/2$，且

$$|v^* P_i u| \ge \|P_i\|/(64n)$$

在承载至少 $W/25$ 总权重的指标上。在$\mathbf{a} = X^{-1}u$ 和$\mathbf{b}^* = v^* X$中，将模小于

$$\theta = \left[100n^3(1 + \|X\|_F^2 + \|X^{-1}\|_F^2)\right]^{-1}$$

的每一项替换为 θ, 并重新构造两个方向。每个方向的欧几里得变化量至多为

$$2\sqrt{n}\,\theta \max(\|X\|_F, \|X^{-1}\|_F) \le \frac{1}{100n^{5/2}}.$$

因此，两个方向的范数都至多为一。每个成功的坐标，其模至少为相应列范数或对偶行范数除以 $4s$，因此大于 θ 且保持不变。所以重叠估计仍然成立。非零坐标还使得对于每个非零且谱为无重根的参数，新谱都与参考谱不相交。

批量更新的方向。 对于 m 选定的配对，引理 6.2 给出名义向量u_0，v_0 其分量为 $\pm 1/(2s)$，以及精确的参考因子 $\mathbf{a}$，$\mathbf{b}^*$。将实际方向定义为 $u = X\mathbf{a}$，$v^* = \mathbf{b}^* X^{-1}$。实际方向和名义方向的范数都不超过一，并且至少在 $m/25$个选定标签上，

$$|v^* P_i u| \ge \|P_i\|/(2^{13}n).$$

原坐标中的误差至多为

$$\epsilon = 2^{-40} n^{-6} (1 + L)^{-2}.$$

批次更新期间只需要精确的参考因子。对于一个接受的参数 t，将B 更新 tuv^*，并将 $\mathcal{E}$ 更新 $tu_0 v_0^*$。下面的参数界给出

$$\|tuv^* - tu_0 v_0^*\|_F \le 3r\epsilon \le \eta.$$

在一个批次期间，$\mathcal{E}$ 由批次开始时的显式值以及待处理的名义秩一项组成。在任一退出时，按照引理 6.3 形成所得的显式矩阵。在提前退出时，还要求在与保留的未标记标签对应的校正列上满足 $K_i^2 \le Q^2$。

相位选择。对于拟议的更新，令 $F = uv^*$ 表示其实际方向矩阵。将用于更新 $\mathcal{E}$ 的方向记为 F_0。因此，对于单独更新或终止更新，有 $F_0 = F$；而在批次中，有 $F_0 = u_0 v_0^*$。二次惩罚中的交叉项决定了相位搜索中的系数 c。定义

$$g(z) = D(B + rzF), \qquad c = 2\nu r \ \mathrm{tr}(\mathcal{E}^* F_0).$$

对于单次更新或终止更新，直接计算 c，并设置 $c_0 = c$。对于批次，应用引理 6.3，使用实乘子 $2\nu r$，得到满足 $|c_0 - c| \le 1/100$ 的 c_0。其数据为批次开始时的显式矩阵以及待处理的名义更新。下方的前缀估计验证了该引理的假设。对 g 和 c_0 应用命题 3.3。对于批次，将参考节点 λ_i 作为附加节点提供。初始扰动和先前的相位选择确保当前特征多项式 p 满足 $p(\lambda_i) \neq 0$。对于单次更新和终止更新，使用空的附加节点列表，因为调整后的方向已确保与参考谱不相交。选定相位在单位圆上至少保留 $|g(z) e^{-cz}|$ 最大值的 0.9，半径界如下：

$$1 \le |z| \le 1 + \frac{1}{500 J}, \qquad J = 2n + 64A,$$

其中，A 是不小于 $1 + |\operatorname{Re} c_0| + |\operatorname{Im} c_0|$ 的最小 2 的幂。使用 $t = rz$。因此，$|t| < 2r$。对于单次更新，将 tF 加到 $\mathcal{E}$ 上。命题中的泰勒多项式使算法能够通过域运算和比较来选择相位。

终端更新。在一个参考矩阵处，任选一个满足 $K_i > Q_{\mathrm{term}}$ 的索引（如果存在）。计算其对偶行向量和右列范数的上界估计 c_w，c_x，使估计值与实际范数之比不超过 1.01，然后取

$$u = \frac{w_i}{2c_w}, \qquad v = \frac{x_i}{2c_x}.$$

两个范数都至多为 $1/2$，且 $|v^* P_i u| \ge \|P_i\|/8$。应用与单次更新相同的坐标调整。两个目标坐标都超过 θ，因此重叠得以保留，且调整后的两个范数都至多为一。使用满足 $F_0 = F$ 的加权相位选择，通过实际的秩一变更更新 $\mathcal{E}$，并更改参考矩阵。如果不存在这样的索引，则返回整个特征系统。

7.2 增益、精度与终止

在一个批次开始时，我们已知 $W < n$。这为各个参考特征向量所需精度之和提供了上界。

引理 7.1 (参考特征系统的精度)。在一个批次开始时，第 6 节中的列指数可以选取为 $\sum_i b_i = O(nL)$，使其同时适用于两个参考基。额外的符号变换精度 ϵ 具有相同的总体预处理界。

证明。令 b_i 为大于 $20\lceil \log_2(2n) \rceil$ 的第一个 2 的幂，再加上对对偶范数和间隙倒数这两个非负对数分别取上整后的估计值。对于对偶范数，采用分别缩放右基和左基时较大的要求。它们的原始列长度属于 $[1/2, 1]$，因此其对偶范数介于 $\|P_i\|$ 和 $2\|P_i\|$ 之间。由于 $\|P_i\| \ge 1$ 和 $s_i \le 4$，这两个非负对数之和与 $\log_2^+ K_i$ 相差至多一个绝对常数。根据 r 的定义，$\log_2(1/r) = O(L)$。在权重阈值以下，这给出 $O(L)$；在阈值以上，上界为

$O(L) + \gamma_i$。二进制舍入会使因子增至两倍。因此 $\sum_i b_i = O(nL + W) = O(nL)$。符号表每行或每列只需额外 $O(\log n + \log(1 + L)) = O(L)$ 位，这不会影响该求和界。

□

引理 7.2 (每次更新的增益)。一次单次更新使 $\Phi(B, \mathcal{E})$ 至少增加 $W/100$，其中包括其矩阵校正。每次校正前被接受的批量更新至少获得 $m/25$。每次终止更新至少获得一，包括其校正。当参考判别式模至少为 G_0 时，一个批次的初始扰动和最终校正造成的总损失小于二。

证明。将（2.2）中的均值估计转换为自然对数，并结合 $\int_0^{2\pi} \mathrm{Re}(ce^{it})\, dt = 0$ 可知，$\log |g(z) e^{-cz}| - \log |g(0)|$ 的圆周均值具有与均值判别式增益相同的下界。根据命题 3.3，相对于真实加权最大值，对数模的损失至多为 $-\log(0。9) < 1/4$。此外，

$$\nu|t|^2 \|F_0\|_F^2 \le 2\nu r^2 \le 2^{-15} m.$$

这里在第一个不等式中使用了更精确的界 $|t|/r \le 1 + 1/(500J)$。减去这个二次项后，得到 Φ 的增量。

对于一次单次更新，成功配对的权重至少为 $W/25$ 。在权重为正的配对上，

$$\frac{r|v^* P_i u|}{8 s_i} \ge \frac{r K_i}{512 n} \ge 2^{\gamma_i}.$$

因此，平均对数增益至少为 $(\log 2) W/50$。由于 $m \le n \le W$ 和 $W \ge 2^{20}$，相位损失、二次项以及最终因子 0.9 之后，增益仍至少为 $W/100$。校正之前也有相同的下界。

对于一次批量更新，选定的真实比值超过 M。至少在 $m/25$ 个这样的指标上，

$$(r/16)|v^* P_i u| > 4 s_i.$$

引理 2.1 给出 g 在 $|z| \le 1/8$中的零点至少有 $m/50$ 个。每个零点对圆周均值的贡献至少为 $3 \log 2$。由于加速调用中有 $m \ge 2^{15}$ ，扣除相位损失和二次项损失后，增益仍至少为 $m/25$。

对于终止更新， $K_i > 2^{12}/r$ 和重叠 $\|P_i\|/8$ 给出相同的严格碰撞不等式。由整数性可知， g 在 $|z| \le 1/8$ 中至少有一个零点。其贡献 $3 \log 2$ 扣除相位损失、二次项损失和校正损失后仍大于一。最后，批次的初始扰动保留其判别式模的至少 $1 - 1/(500 n^3)$，而校正保留 0.9。二者均不改变 $\mathcal{E}$，因此总损失小于二。

□

命题 7.3 （局部阶段的终止）。局部算法会返回一个与输入的距离不超过 $\delta/3$ 的无重根矩阵。在加速调用中，至多有 m 个指标具有 $K_i > Q$。在终止调用中，每个 $K_i \le Q_{\mathrm{term}}$。设 $\rho = 545/576$。加速情况下，局部算法的代价为 $\widetilde{O}\left(n^3 + n^{3+\rho-\alpha}\right)$ ；终止情况下为 $\widetilde{O}\left(n^4\right)$ ，其中省略了固定次幂的 $1 + \log\left(n/\delta\right)$ 。

证明。考虑一个至多包含 T 次矩阵变化的前缀，并分别计入初始扰动、接受的秩一更新和校正。我们首先界定该前缀中各状态与输入的距离，以及所需的精度。在参考对象变更时，顺序至关重要：试验矩阵的范数界和判别式界必须足以提供计算其校正所需的精度。随后，我们统计更新次数，并证明算法会在发生 T 次变化之前停止。

根据构造，

$$\|B - B_{\mathrm{s}} - \mathcal{E}\|_F \le T\eta = \delta/2^{14}. \tag{7.6}$$

在批量更新时，引理 6.2 将实际更新与名义更新之间的差异界定为 η，其中包括因子 $|t|$。在单独更新或终止更新时，对 $\mathcal{E}$ 的改动是精确的。每个初始扰动和每项校正均从 $\mathcal{E}$ 中省略，且其弗罗贝尼乌斯范数至多为 η。

种子满足 $|D(B_{\mathrm{s}})| \geq 16G_0$。增益引理给出

$$\Phi(B, \mathcal{E}) \geq \log |D(B_{\mathrm{s}})| - 2$$

在计算根或基的每个状态下：完整更新和批次的增益均为正，而唯一未完成批次的损失小于二。只要当前范数小于二，$|D(B)| \leq 4^d$，因此

$$\nu \|\mathcal{E}\|_F^2 \leq \log \frac{4^d}{|D(B_{\mathrm{s}})|} + 2 < 2n^2 L + 2.$$

ν 的定义因此给出 $\|\mathcal{E}\|_F < \delta/2^{10}$。结合 (7.6) 和初始的 $\delta/4$，可将与输入的距离界定为

$$\delta/4 + \delta/2^{10} + \delta/2^{14} < \delta/3.$$

每次试探变更的范数至多为 $2r < \delta/2^{18}$，批次启动项和校正的范数至多为 η。为证明范数始终小于二，从距离输入不超过 $\delta/3$ 的状态开始。接下来的试探范数至多为 $1 + \delta/3 + 2r + \eta < 2$。在对其根进行近似之前，解析阶段估计适用。校正测试通过后，校正的弗罗贝尼乌斯范数至多为η，并且至少保留判别式模的 0.9。再次应用势函数估计，即可得出下一状态的位移界。严格的范数裕量也确保匹配近似值在下文所用的精度下位于 $|z| < 2$。

由于二次项非负，同一个势函数下界意味着 $|D(B)| \geq e^{-2}|D(B_{\mathrm{s}})| > G_0$。对于选定阶段的矩阵，使用校正前的增益。因此，其范数和判别式给出了逼近其根所需的下界。校正通过判别式检验后，我们在势函数估计中计入其额外损失，至多为 $-\log(0.9)$。

令 $\mu_1, \ldots, \mu_n$ 为当前特征值。每个当前间隙至少为 $g = G_0/4^d$。投影算子公式

$$P_i = \prod_{j \neq i} (B - \mu_j I)/(\mu_i - \mu_j)$$

给出 $\|P_i\| \leq (4/g)^{n-1}$。对于归一化参考列，

$$\log_2(1/g) \leq 4n^2 L, \qquad \log_2(2 + \|X\| + \|X^{-1}\|) \leq 8n^3 L.$$

在初始扰动下，每个参考节点处的特征多项式的模为$|\sigma\alpha\beta p_0'(\lambda_i)|$，其中 $|p_0'(\lambda_i)| \geq g^{n-1}$，采用引理 5.2中的记号。将其半径参数替换为 $\min(r, \eta)$，其倒数对数仅增加 $O(L + \log(1 + L))$。在单独更新或终止更新时，调整后的坐标给出下界 $r\theta^2 |p_0'(\lambda_i)|$。在一个批次期间，每次径向调整至多将前一个下界除以 $2^{16} nJ$。

在每个批次开始时，$\|\mathcal{E}\|_F < \delta$，并且

$$2\nu r\delta \leq 2^5 n\sqrt{mL} \leq 2^{10} n^2 (1 + L).$$

因此，实乘子 $2\nu r$ 满足引理 6.3 的假设。对 $\|\mathcal{E}\|_F$ 的界还给出$|c| \leq 2\nu r \|\mathcal{E}\|_F \|F_0\|_F = O(n\sqrt{mL})$，因此 $J = O(n\sqrt{mL} + n)$。 consequently，多项式倒数下界的对数至多为

$$O\big(n^3 L + T \log(n(1 + L))\big).$$

将这个下界除以 4^{n-1}，即可得到当前特征值与参考特征值之间每个距离的下界。对于当前参考矩阵 B_0，薄中间因子的范数至多为 $4\|X^{-1}\|\|X\|\|B-B_0\|$。为了与 H进行显式比较，上面对 g的两个界以及基范数给出：$160n^3L$ 对于 $|\sigma\alpha\beta p_0'(\lambda_i)|$ 和 $r\theta^2|p_0'(\lambda_i)|$两者，其倒数对数均至多为 $160n^3L$。事实上$|p_0'(\lambda_i)| \ge g^{n-1}$，分母中的每个平方弗罗贝尼乌斯范数至多为平方算子范数的 n 倍，而 $\log_2(1/\sigma)$受初始量定义中两个指数的较大者所界定。此外，$J \le 2^{16}n^2L$，因此每次径向选择在以二为底的对数中产生的代价至多为 $10L$。因此，倒数距离对数由 $160n^3L + 10TL + 2n < (256nL)^8$所界定。所显示的薄中间因子界也由同一个指数涵盖。

半径为二的圆盘中的匹配近似值，可通过将标量算法应用于 $2^{-n}p(2z)$ 获得，所需绝对误差减半。对于匹配误差 e，根据定理 3.5，这个缩放后的多项式所需的一个充分系数指数为 $n(\lceil\log_2(2/e)\rceil + \lceil\log_2(4n)\rceil)$。要获得弗罗贝尼乌斯范数至多为 η 的校正，只需令 $\lceil\log_2(1/e)\rceil$ 为 $n\log H + \log(1/\eta)$ 的一个足够大的绝对常数倍。因此，系数指数的对数为 $O(\log(nL))$。所需的二的幂通过 $O(\log(nL))$ 次重复平方构造。固定参考表使用的各个指数由引理 7.1 给出。弗罗贝尼乌斯配对的表格只需额外 $O(\log n + \log(1+L))$ 位，即可将相位系数近似到绝对误差至多为 $1/100$。

首先考虑参考对象变更时的判别式检验。对于间隙至少为 H^{-1} 的情形，将每个节点移动至多 e，会使每个两两距离的变化至多为 $2e$，相对误差至多为 $2He$。不等式 $(1-2He)^d \ge 1-2dHe$ 在满足 $2dHe < 1/10$ 后给出所需的 0.9 判别式检验。提前退出时还需进行第二项检验。校正前，每个保留标签的真实比值至多为 $16M$。保持这些标签不变。将每个匹配的真实列向量乘以其计算得到的除数，使其长度处于 $[1/2, 2]$。若其误差至多为 e'，则矩阵误差至多为 $\sqrt{n}e'$，而真实的缩放矩阵 X_t 的逆范数至多为 $2\sqrt{n}(4H)^{n-1}$。由 Neumann 不等式可得

$$\|X^{-1} - X_t^{-1}\| \le 2\|X_t^{-1}\|^2\sqrt{n}e'$$

前提是 $\|X_t^{-1}\|\sqrt{n}e' < 1/2$。每个真实列向量和对偶行向量的范数都至少为 $1/2$。根据所示的逆矩阵估计和引理 5.3，在根精度中取足够大的绝对指数，可使计算列向量和对偶行向量的范数至多为其真实范数的 1.1 倍。同时取 $2He \le 1/11$，可将最近间隙的倒数限制在原值的 1.1 倍以内。由于 $1.1^3 < 3/2$，校正后每个保留比值都小于 $24M < Q$。如果需要细化，则可从第一批传递近似值周围互不相交的半径为 $1/(4H)$ 的圆盘中恢复这些标签。在参考对象变更时，以 $O(n^2)$ 的工作量检验所有节点对。在形成柯西分母之前，按字典序对两个节点列表进行排序，使用 $O(n\log n)$ 次比较，即可同时检验试验节点是否互异，以及它们是否与参考节点互不相同。校正范数根据其薄因子进行检验。旧判别式可由当前的首一特征多项式通过标量结式求得，新判别式则可由试验节点的两两差求得。比较时只需使用模平方。在参考对象变更时，这些检验的开销至多为 $\widetilde{O}(n^2(k+1)^{\omega-2} + nk^5)$。上述固定精度足以应对首次尝试。或者，从传递精度开始，将请求的指数加倍，直到检验通过；这样开销至多增加 $O(\log(nL))$ 倍。

最后，令 q 统计前缀中的单次更新、批次更新和终止更新。完整批次在经历两次损失后，每次秩一更新至少增加一。未完成批次最多损失两。若单次更新或终止更新被截断在相位选择与校正之间，则仍至少增加一。因此

$$\Phi(B, \mathcal{E}) - \Phi(B_{\mathrm{s}}, 0) \ge q - 2.$$

另一方面，这一增量小于 $2n^2L$。因此 $q < 2n^2L + 2$。每个按规定长度完成的批次都包含 k 次更新。只有最终批次可能提前返回，且不包含任何更新。将其初始扰动和最终校正分别计算，矩阵变化总数至多为 $4q + 4$，严格小于 T。因此，没有任何前缀能达到 T 次变化。由于每个内部过程都会终止，局部阶段必定返回。在任一返回测试中，这些比率都满足给定的截断值，且矩阵与输入相差不超过 $\delta/3$。

对于运算次数，根据 [1]、$a = 23/192$、$b = 25/144$ 和 $\rho = 545/576$，使用 $\omega < 2.38$。令$m = \lfloor n^{\alpha} \rfloor$ 和 $k = \lceil n^{b} \rceil$，其中 α 是当前深度对应的指数。一个批次的准备工作和参考对象变更的开销为

$$\widetilde{O}\big(n^{2+a} + n^2 k^{\omega-2} + nk^5\big).$$

其中包括打包参考变换以及用于弗罗贝尼乌斯配对的表格。按照引理 6.3，通过将已存储符号方向数组相乘，在批次结束时更新累积矩阵 $\mathcal{E}$ 的开销为 $\widetilde{O}(n^2k^{\omega-2})$。该表给出了批次内单次更新的开销。

计算	忽略固定的对数因子后的开销
判别式、加权相位和根	$n^{(1+\omega)/2} + n^{1.9} + n\sqrt{m}$
转移余子式	nk^5
高精度坐标	$n^{2-a}k^{\omega-2} + nk^{\omega-1}$
其他范数估计	nk^2n^{5a}
选定向量及符号	nm
参考与惩罚查询	$n + nk$

未选定分子的准备工作耗费 $\widetilde{O}(nk^2 + nk)$，已包含在所示界限中。恒等式

$$2 - a + \frac{19}{50}b = 1 + 2b + 5a = 2 + a - b = 1 + \rho$$

表明，除选定向量项外，每个行向量均为 $\widetilde{O}(n^{1+\rho})$；批次准备工作除以 k 后也如此。对于其余项，使用 $5b < \rho$, $2 + (\omega - 3)b < 1 + \rho$, $1 + (\omega - 1)b < 1 + \rho$ 和 $1 + \alpha/2 \le 3/2 < 1 + \rho$。因此，一个完整批次的代价为 $\widetilde{O}(k(n^{1+\rho} + nm))$，并带来 $\Omega(km)$ 的增益。势的总可用增量为 $O(n^2L)$，因此这些批次的总成本为 $\widetilde{O}(n^{3+\rho-\alpha} + n^3)$。单次更新的代价为 $\widetilde{O}(n^2)$，增益为 $\Omega(n)$，因此还需 $\widetilde{O}(n^3)$。

至多有一个未完成批次。其准备工作和参考对象变更的成本为 $\widetilde{O}(n^{2+a})$，因为 $2 + (\omega - 2)b < 2 + a$ 和 $1 + 5b < 2 + a$。其已接受的更新所带来的每单位增益成本，与完整批次中的相同。最终扫描的成本为 $\widetilde{O}(n^{1+\rho} + nm)$，低于三次复杂度上限。初始扰动和精确参考测试的成本为每批 $O(n^2)$，并计入其准备工作中。生成返回的对角化结果的成本计入参考对象变更成本。由对角化结果生成局部输出矩阵的成本为 $O(n^3)$，且只需计算一次。因此，局部加速成本为

$$\widetilde{O}\big(n^3 + n^{3+\rho-\alpha} + n^{2+a}\big).$$

一次终止更新的代价为 $\widetilde{O}(n^2)$：两个方向和弗罗贝尼乌斯配对使用存储基与显式累积矩阵，矩阵铅笔的宽度为一，且相位搜索的代价为 $m = 1$。此类更新共有 $O(n^2L)$ 次，总工作量为 $\widetilde{O}(n^4)$。这些界包含初始特征系统和最终比值检验的开销。

□

7.3 三角组装与一致界

在局部阶段结束时，满足截断条件的索引所对应的右特征向量张成一个不变子空间。然而，更改尾随块仍可能改变保留特征值对应的投影算子。我们为该块选择一个移位，并利用递归之前谱中的间隙来界定变化。

定理 *1.1* 的证明。 我们使用两个加速阶段，其中

$$\alpha_0 = \frac{947}{1000}, \qquad \alpha_1 = \frac{39}{50}, \qquad \rho = \frac{545}{576},$$

随后执行终止过程。在深度 $j = 0$，**1** 处，例外维数至多为$m = \lfloor n^{\alpha_j} \rfloor$。低于固定维数阈值的调用会立即使用终止过程。我们采用 (7.3) 中的局部截断值 Q 和 Q_{term}，以及它们在 (7.4) 中的界。

不变子空间与尾随块。 在非终止调用中，令 B 为局部过程返回的单矩阵。它与输入的距离小于 $\delta/3$。取所有满足比值条件 $K_i \le Q$的右特征向量的不变张成空间。其余维数 ℓ 至多为$m < n$。若 $\ell = 0$，则返回 B 及其存储的特征系统。否则，使用精确的格拉姆–施密特投影，补入标准坐标向量并按第2节中的方式缩放，可得正交列基 S，满足 $\kappa_2(S) \le 1.02$，并且

$$S^{-1} B S = \begin{pmatrix} G & C \\ 0 & H_b \end{pmatrix}.$$

设 $p = n - \ell$ 是 G 的阶数。两个对角块的范数都不超过 $1.02(1 + \delta/3) < 2$，因此 $H_b/2$ 是一个压缩算子。以容差 $\delta/64$ 对 $H_b/2$ 递归应用该算法，并将返回的作用加倍，得到 H_{rec}。对剩余深度使用归纳法可得 $\|H_{\mathrm{rec}} - H_b\| < \delta/32$。若尾随块为标量，则返回其给定的作用和单位列向量。

适应前导间隙的平移。 设 $\lambda_1, \dots, \lambda_p$ 是 G 中保留的特征值，σ_i 是 B 的完整旧谱中 λ_i 的最近特征值间隙。因此，原始谱投影算子 P_i^{old} 满足

$$\|P_i^{\mathrm{old}}\| \le Q \sigma_i.$$

圆盘 $D(\lambda_i, \sigma_i/2)$ 的内部互不相交。由于 $\|B\| < 2$ 和 $\sigma_i < 4$，它们位于 $D(0,4)$ 中。因此

$$\sum_{i=1}^{p} \sigma_i^2 \le 64, \qquad \sum_{i=1}^{p} \sigma_i \le 8\sqrt{p} \le 8\sqrt{n}. \tag{7.7}$$

设 $\mu_1, \dots, \mu_\ell$ 表示 H_{rec} 的特征值。我们选取一个较小的复数平移 z，使得

$$|\lambda_i - (\mu_j + z)| \ge \frac{\delta}{2^{20}} \max\left\{ \frac{1}{n}, \frac{\sigma_i}{\sqrt{\ell}} \right\} \quad (1 \le i \le p,\ 1 \le j \le \ell). \tag{7.8}$$

通过域运算构造平移：通过比较距离平方，选取距离 λ_i 最近的旧特征值，并令 $\widetilde{\sigma}_i$ 为两者之差的实部和虚部绝对值之和。于是 $\sigma_i \le \widetilde{\sigma}_i \le 2\sigma_i$。通过比较整数平方，找到满足 $\sqrt{\ell} \le t \le 2\sqrt{\ell}$ 的整数 t。定义

$$\gamma = \frac{\delta}{2^{20} n}, \qquad q = 2^8 n, \qquad r_i = \frac{\delta}{2^{20}} \max\left\{ \frac{1}{n}, \frac{2\widetilde{\sigma}_i}{t} \right\}.$$

考虑平移 $z=\gamma\,(a+\mathrm{i}b)$，其中整数 $-q\le a,b\le q$。对于每一对 (i,j)，标记以 $\lambda_i-\mu_j$为中心、最大范数半径为 r_i 的闭正方形内的网格点。对于 $R_i=r_i/\gamma\ge 1$，该正方形至多包含 $25R_i^2$ 个网格点，并且

$$R_i^2\le 1+\frac{16n^2\sigma_i^2}{\ell}.$$

根据 (7.7)，计数重复项后，被标记点的总数至多为

$$25\left(p\ell+16n^2\sum_i\sigma_i^2\right)\le 25(1/4+1024)n^2<2^{15}n^2.$$

共有 $(2q+1)^2>2^{18}n^2$ 个候选平移。因此存在一个未标记点，并且满足 (7.8)。其范数至多为 $\sqrt{2}\,q\gamma=\sqrt{2}\,\delta/2^{12}<\delta/1024$。每个被标记正方形的整数端点可通过在 $[-q,q]$ 内进行比较来找到。枚举被标记点、对其排序并找到缺失点的代价为 $O(n^2\log n)$ 次运算和比较。其中包括确定旧最近间隙所需的 $O(n^2)$ 次比较。

令 $H_b'=H_{\mathrm{rec}}+zI$ 和 $\Delta=H_b'-H_b$。则

$$\|\Delta\|<\frac{33\delta}{1024}. \tag{7.9}$$

特别地，$\|H_b'\|\le 1.02(1+\delta/3)+33\delta/1024<2$。两个对角谱均为无重根的且不相交。定义

$$T_{\mathrm{new}}=\begin{pmatrix}G & C\\ 0 & H_b'\end{pmatrix},\qquad B_{\mathrm{out}}=ST_{\mathrm{new}}S^{-1}.$$

从 B_{out} 到输入的距离小于

$$\delta/3+1.02(\delta/32+\delta/1024)<\delta.$$

前导特征基继承自所选的特征向量，而递归调用提供尾部特征基。在这些基下，通过将每个条目除以一个交叉差来计算 Sylvester 修正。这将在 $O(n^3)$ 次运算内给出完整的精确特征基及其逆矩阵。谱测试和网格选择也使用实数比较。

投影算子范数之和。对于非空的无重根矩阵 A，定义 $\mathcal{S}(A)=\sum_{\lambda\in\mathrm{spec}(A)}\|P_\lambda^A\|$。特别地，标量具有 $\mathcal{S}=1$。我们使用

$$\kappa_V(A)\le\mathcal{S}(A). \tag{7.10}$$

事实上，选取单位右列 x_i 和对偶行 w_i^*。如果 $p_i=\|P_i^A\|=\|w_i\|$，将列i 乘以 $\sqrt{p_i}$，并将其对偶行乘以 $1/\sqrt{p_i}$，就会使两个平方弗罗贝尼乌斯范数都等于 $\sum_i p_i$。算子范数受弗罗贝尼乌斯范数的界定。此重新缩放仅用于条件估计。

对于每个保留特征值，S 坐标下的旧投影算子具有如下形式

$$\Pi_i=S^{-1}P_i^{\mathrm{old}}S=\begin{pmatrix}U_i & R_i\\ 0 & 0\end{pmatrix},\qquad \|\Pi_i\|\le A_0\sigma_i,\qquad A_0=1.02Q,$$

其中， U_i 是 G对应的投影算子。令 $Y=\sum_i R_i\in\mathbb{C}^{p\times \ell}$ 和 $S_g=\sum_i \|\Pi_i\|$。由投影算子的恒等式可得

$$\sum_i U_i = I_p,\qquad U_i Y = R_i,\qquad C = GY - YH_b,\qquad \|Y\|\le S_g \le 8A_0\sqrt{n}.$$

将 Π_i 和 $S^{-1}BS$ 的交换关系求和，并对 $i\neq j$、$\Pi_i^2=\Pi_i$ 使用 $\Pi_i\Pi_j=0$，即可得到这些恒等式。

对于 H_b' 的特征值 μ，令 T_μ 为其谱投影算子，并设 $\mathcal{S}_{\mathrm{ch}}=\sum_{\mu\in\mathrm{spec}(H_b')}\|T_\mu\|$。加倍和平移不会改变投影算子，因此这就是递归调用返回的和。它至少为 1。由于 $U_iC=R_i(\lambda_i I-H_b)$，三角形预解式给出了新的前导投影算子的右上块如下：

$$R_i' = R_i\big[I+\Delta(\lambda_i I - H_b')^{-1}\big].$$

新的尾随投影算子的右上块为

$$(\mu I-G)^{-1}CT_\mu = -YT_\mu + (\mu I-G)^{-1}Y\Delta T_\mu,$$
$$(\mu I-G)^{-1}Y = \sum_i \frac{R_i}{\mu-\lambda_i}.$$

定义 $F(\mu)=\sum_i \sigma_i/|\mu-\lambda_i|$。在投影算子 T_μ 中展开 $(\lambda_i I-H_b')^{-1}$，前导投影算子的校正项范数之和至多为 $\|\Delta\|A_0\sum_\mu F(\mu)\|T_\mu\|$。尾随投影算子中包含 Δ 的项满足相同的界。尾随投影算子的其余右上项至多贡献 $\|Y\|\mathcal{S}_{\mathrm{ch}}$。计入对角块可得

$$\mathcal{S}(T_{\mathrm{new}}) \le S_g + \left(1 + S_g + 2\|\Delta\|A_0\sup_\mu F(\mu)\right)\mathcal{S}_{\mathrm{ch}}.$$

为估计 F，选取距离 μ 最近的首个特征值 λ_*。方程 (7.8) 将其贡献限制在 $2^{20}\sqrt{\ell}/\delta$ 以内。对于其他每个指标，$\sigma_i\le|\lambda_i-\lambda_*|\le 2|\mu-\lambda_i|$。将这些指标分组成壳层$r\le|\mu-\lambda_i|<2r$，其二进制尺度半径从 γ 开始。对于其中一个壳层，不相交的圆盘 $D(\lambda_i,\ \sigma_i/4)$ 位于 $D(\mu,\ 3r)$ 内，因此 $\sum_{\text{壳层}}\sigma_i^2\le 144r^2$。若该壳层包含 h 个指标，则柯西–施瓦茨不等式将其对 F 的贡献限制在 $r^{-1}(h\sum_{\text{壳层}}\sigma_i^2)^{1/2}\le 12\sqrt{h}$ 以内。经过末尾扰动后，两个对角块的范数都小于二。因此，所有这些距离都小于八，至多需要 $L+17\le 3L$ 个壳层。对各壳层应用柯西–施瓦茨不等式可得

$$\sup_\mu F(\mu)\le \frac{2^{20}\sqrt{\ell}}{\delta}+21\sqrt{nL}. \tag{7.11}$$

利用 (7.9)、$\ell\le n$、$\delta<1/2$ 和 $S_g\le 8A_0\sqrt{n}$，再通过 S 共轭变换回去，得到

$$\mathcal{S}(B_{\mathrm{out}})\le 2^{18}Q\sqrt{nL}\,\mathcal{S}_{\mathrm{ch}}. \tag{7.12}$$

该常数包含两个 1.02 因子。若尾随块为空，则改用直接界$\mathcal{S}(B)\le 8Q\sqrt{n}$。空块不赋予间隙或条件数。

一致幂。 对于输入维度 N 和容差 δ_*，定义

$$L_0 = \lceil \log_2(64N/\delta_*) \rceil.$$

令 n_j、δ_j和 L_j 分别表示深度为 j 时的维度、容差和对数参数。因此，$\delta_j = \delta_*/64^j$ 并且，对于 $j \le 2$，$L_j \le L_0 + 12 \le 3L_0$。 successive 维度满足

$$n_0 = N, \qquad n_1 \le N^{947/1000}, \qquad n_2 \le N^{36933/50000}.$$

在终端块中，将 (7.7) 应用于其完整谱和截断 Q_{term}，可得 $\mathcal{S} \le 8Q_{\mathrm{term}}\sqrt{n}$。结合 (7.4)、(7.12) 和 $\lfloor n^\alpha \rfloor \ge n^\alpha/2$，得到

$$\mathcal{S}(B_{\mathrm{out}}) \le C_1 \delta_*^{-3} L_0^{5/2} N^{256633/50000}. \tag{7.13}$$

该指数中的三个维度项为

$$\frac{5-\alpha_0}{2} = \frac{4053}{2000}, \qquad \alpha_0 \frac{5-\alpha_1}{2} = \frac{199817}{100000}, \qquad \frac{3}{2}\alpha_0\alpha_1 = \frac{110799}{100000}.$$

它们的和为 $256633/50000$。每个加速因子 $Q\sqrt{n}L$贡献一个因子 $\delta_j^{-1}L_j$，而终端因子 $Q_{\mathrm{term}}\sqrt{n}$ 贡献 $\delta_j^{-1}L_j^{1/2}$。空尾随块会从估计中移除后续因子。标量贡献一个因子。如果某次调用因维度低于固定阈值而提前使用终止过程，则其依赖维度的因子可并入绝对常数。这些情形同样满足 (7.13)。

对于一个保留的首个特征值，$\|P_i^{\mathrm{old}}\| \ge 1$ 意味着 $\sigma_i \ge Q^{-1}$。丢弃其他特征值只会在存在最近剩余邻点时增大它与该邻点之间的距离。相应的终端界为 $1/$间隙$(B) \le Q_{\mathrm{term}}$，只要该块至少包含两个特征值。单元素块没有内部点对。交叉点对的间隔至少为 $\delta/(2^{20}n)$。在向根节点推进的过程中，每次加倍都会使内部间隙乘以二。后续的平移和基变换会保持这些间隙不变。在两个加速逆截断中，N 的幂次为

$$2 - \frac{\alpha_0}{2} = \frac{3053}{2000}, \qquad \alpha_0\left(2 - \frac{\alpha_1}{2}\right) = \frac{152467}{100000}.$$

终端项的幂次为 $36933/50000$，而交叉间隙的幂次至多为一。第一个显示的幂次最大。因此

$$\mathrm{gap}(B_{\mathrm{out}})^{-1} \le C_2 \delta_*^{-1} L_0^{1/2} N^{3053/2000}. \tag{7.14}$$

结合（7.10）、（7.13）和（7.14），我们得到

$$\frac{\kappa_V(B_{\mathrm{out}})}{\mathrm{gap}(B_{\mathrm{out}})} \le C_3 N^{166479/25000} \delta_*^{-4} L_0^3.$$

$20/3$ 与 $166479/25000$ 之差为 $\vartheta = 563/75000 > 0$。对于 $R = N/\delta_* > 4$，我们有$L_0 \le 7 + \log_2 R$，且 $(7+\log_2 R)^3 \le C_\vartheta R^\vartheta$。这里 C_ϑ 是有限的，因为 $(7+x)^3\, 2^{-\vartheta x}$ 对于 $x \ge 2$有界。由于 $166479/25000 > 4$ 和 $\delta_* < 1$，前面的估计至多为 $CR^{20/3}$，其中 C是一个绝对常数。

算术成本。 每个加速阶段只有一次递归调用。维度为 n 、指数为 α_j 时的局部成本为 $\widetilde{O}(n^{3+\max(0,\rho-\alpha_j)})$。顶层指数为三，因为 $\alpha_0 > \rho$。在深度一时，其作为 N 的幂的指数为

$$\frac{947}{1000}\left(3+\frac{545}{576}-\frac{39}{50}\right)=\frac{43176571}{14400000}<3.$$

四次终端计算的指数为 $4\alpha_0\alpha_1 = 36933/12500 < 3$。一次维度有界的较早终端调用，其成本至多是其三次成本乘以某个固定常数。正交基与分块组装的成本为 $O(n^3)$，而自适应移位的成本为 $O(n^2 \log n)$。所有对数参数均为 $O(1 + \log(N/\delta_*))$，且至多有三次非平凡调用。因此，总成本为 $O(N^3 \log^d(N/\delta_*))$，其中 d 是一个绝对常数。每次从递归返回时，扰动界均成立。从最终矩阵中减去输入矩阵得到 E。 □

参考文献

[1] J. Alman、R. Duan、V. Vassilevska Williams、Y. Xu、Z. Xu 和 R. Zhou，《More asymmetry yields faster matrix multiplication》，Proc. 2025 ACM–SIAM 离散算法年度研讨会（SODA），SIAM，2025。[2] N. Amsel 等人，《线性系统与特征值问题：Simons 研讨会提出的开放问题》，arXiv:2602.05394v3（2026）。[3] J. Banks、J. Garza-Vargas、A. Kulkarni 和 N. Srivastava，《伪谱分裂、符号函数与近似矩阵乘法时间内的对角化》，Found. Comput. Math. 23（2023），1959–2047。[4] J. Banks、A. Kulkarni、S. Mukherjee 和 N. Srivastava，《伪谱的高斯正则化与 Davies 猜想》，Comm. Pure Appl. Math. 74（2021），2114–2131。[5] A. Bostan、C.-P. Jeannerod、C. Mouilleron 和 E. Schost，《关于具有位移结构的矩阵：广义算子与更快的算法》，SIAM J. Matrix Anal. Appl. 38（2017），733–775。[6] R. P. Brent 和 H. T. Kung，《操作形式幂级数的快速算法》，J. ACM 25（1978），581–595。[7] D. G. Cantor 和 E. Kaltofen，《任意代数上多项式的快速乘法》，Acta Inform. 28（1991），693–701。[8] E. B. Davies，《近似对角化》，SIAM J. Matrix Anal. Appl. 29（2008），1051–1064。[9] P. L. Duren，《单叶函数》，《数学科学基础丛书》259，施普林格，1983。[10] J. von zur Gathen 和 J. Gerhard，《现代计算机代数》，第三版，剑桥大学出版社，2013。[11] C. A. Neff 和 J. H. Reif，《复根问题的高效算法》，J. Complexity 12（1996），81–115。[12] V. Y. Pan，《单变量多项式：数值因式分解与求根的近乎最优算法》，J. Symbolic Comput. 33（2002），701–733。[13] R. Shah、N. Srivastava 和 E. Zeng，《稀疏伪谱粉碎》，arXiv:2411.19926v3（2026）。[14] V. Strassen，《高斯消元并非最优》，Numer. Math. 13（1969），354–356。[15] A. Townsend，《数值线性代数中的开放问题》，GitHub 仓库（2026），问题 IE-07，访问日期：2026 年 11 月 19 日。[16] L. N. Trefethen 和 M. Embree，《谱与伪谱：非正规矩阵与算子的行为》，普林斯顿大学出版社，2005。