

A rigorous interval for the norm in Trefethen's third challenge problem

September 10, 2026

Abstract

The infinite matrix obtained by arranging the reciprocals of the positive integers along successive antidiagonals defines a compact operator on ℓ^2 . We give a computer-assisted enclosure of its norm with width $1.97338654216 \cdot 10^{-871}$, fixing 870 decimal places. The calculation uses positive Gaussian quadrature for counting measure, a finite-rank Laplace approximation of the sampled matrix, and explicit bounds for the discretization error and the infinite tail. The exact interval, numerical data, and verification programs accompany this paper.

1 The problem and the interval

Let

$$a(s, t) = \frac{2}{(s+t-1)^2 + s - t + 1}, \quad A_{jk} = a(j, k), \quad j, k \in \mathbb{N}, \quad (1)$$

and let $T = \|A\|_{\ell^2 \rightarrow \ell^2}$. Thus the first two rows of A are $(1, 1/2, 1/4, \dots)$ and $(1/3, 1/5, 1/8, \dots)$. This is the matrix in Problem 3 of Trefethen's 2002 challenge [1]. Each positive integer occurs exactly once as the denominator of an entry. Consequently

$$\|A\|_{\text{HS}}^2 = \sum_{j,k \geq 1} a(j, k)^2 = \zeta(2) = \frac{\pi^2}{6}. \quad (2)$$

In particular, A is compact. Its positive entries imply that its largest singular value is simple and has positive left and right singular vectors. The same conclusions hold for the finite sections A_N and the sampled kernel matrices B defined below.

Theorem 1. *Let L and U be the terminating decimals specified in Appendix A. Then*

$$L \leq T \leq U, \quad U - L = 1.97338654216 \cdot 10^{-871}.$$

The endpoints have the same first 870 digits after the decimal point.

The norm begins $T = 1.27422415282122818821 \dots$

The enclosure is obtained in three stages. Section 2 bounds the norm of a finite weighted quadrature matrix B . Section 3 compares this norm with that of the integer section $A_N = (a(j, k))_{j,k \leq N}$. Section 4 bounds $T - \|A_N\|$ by a residual expansion. We give the estimates for a family of quadrature rules and truncations. Section 5 discusses parameter selection and cost; Section 6 specifies the calculation proving Theorem 1.

Waldvogel's chapter in the book by Bornemann, Laurie, Wagon, and Waldvogel [2, Chapter 3] develops finite-section, summation, and contour methods for this operator. Lemma 3.1 proves

monotone convergence of finite-section norms and the bound $\pi/\sqrt{6}$. Section 3.4 applies a largest-singular-value version of Stewart’s perturbation expansion [3]; inverse-square vector decay is observed numerically, and an improved fourth-order remainder is conjectured. In §3.5, equation (3.18), Waldvogel introduces the sampled matrix $W^{1/2}(a(c_i, c_j))W^{1/2}$ and evaluates it with a summation formula due to Rolf Strebels, based on Euler–Maclaurin summation. Our matrix B has the same form, with positive Gaussian rules on a finite section. We bound the difference between $\|B\|$ and $\|A_N\|$ and the contribution of the omitted infinite tail. The tail correction has the same perturbative starting point as Waldvogel’s correction; we prove the needed vector decay and bound the remainders explicitly.

Section 3.6 of Waldvogel’s chapter reports a month-long contour calculation with a predicted accuracy of 273 digits. The prediction uses observed convergence and an empirical accuracy law; the long computation serves as the reference for the shorter runs. Trefethen’s later account also records the 273-digit computation [4, p. 16 of the author’s manuscript].

For Gaussian quadrature, Hermite interpolation, and the Hahn-polynomial recurrence used for counting measure, we use [5, §§3.3, 3.5, 18.19, 18.22]. The divided-difference product rule and integral representations are given by de Boor [7]. The spectral estimates are applications of the Rayleigh principle and the Kato–Temple inequality [10]. In Section 3 we extend the sampled singular pair to the integer section and evaluate its Rayleigh quotient, deriving the quadrature and residual estimates for counting measure. Arb supplies the real and complex interval arithmetic [16].

2 Gaussian discretization and the sampled singular value

Choose integers $e > r_0 \geq 6$, and put

$$N = 2^e - 1, \quad n_0 = 2^{r_0} - 1.$$

Retain the integers $1, \dots, n_0$ as nodes of weight one. For each $a = 2^r$, $r = r_0, \dots, e - 1$, choose an integer $1 \leq p_a < a$ and use the p_a -point Gaussian rule for counting measure on $[a, 2a - 1] \cap \mathbb{N}$. We call these $\ell = e - r_0$ cells active. The restriction $r_0 \geq 6$ ensures the domain needed for the two-variable error estimate in Section 3. Write all nodes and weights as c_i, w_i . Their number and the sampled matrix are

$$n = n_0 + \sum_{\text{active } a} p_a, \quad B_{ij} = \sqrt{w_i} a(c_i, c_j) \sqrt{w_j}, \quad 1 \leq i, j \leq n. \quad (3)$$

All weights are positive, and

$$\sum_i w_i = N.$$

The exact prefix can be enlarged when higher quadrature orders are needed.

For a cell containing a equally spaced points, use the coordinate $z = 2(t - a)/(a - 1) - 1$. The monic orthogonal polynomials obey

$$P_0 = 1, \quad P_1 = z, \quad P_{k+1} = zP_k - \frac{k^2(a^2 - k^2)}{(4k^2 - 1)(a - 1)^2} P_{k-1}. \quad (4)$$

These are the monic, rescaled Hahn polynomials $Q_k(x; 0, 0, a - 1)$; the recurrence is the specialization of DLMF (18.22.2)–(18.22.3) [5]. The zeros of P_{p_a} are the Gaussian nodes in this coordinate. If $p_0, \dots, p_{p_a-1}$ are orthonormal for the uniform probability measure on the transformed cell, the weight at a zero z is $a / \sum_k p_k(z)^2$. The verifier isolates all zeros in disjoint intervals with opposite

polynomial signs at their endpoints, then encloses the weights by this formula. There are p_a such intervals; degree counting proves that no zero has been omitted. The rule is exact on polynomials of degree at most $2p_a - 1$.

Let $s = \|B\|$, and let x, y be its positive unit singular vectors: $By = sx$, $B^T x = sy$. Choose positive trial vectors $\hat{x}, \hat{y}$ and normalize each to unit norm. Interval matrix products enclose

$$\mu = \hat{x}^T B \hat{y}, \quad R_B = \frac{\|B\hat{y} - \mu\hat{x}\|^2 + \|B^T \hat{x} - \mu\hat{y}\|^2}{2}. \quad (5)$$

Appendix D describes how those products are evaluated without storing B . Their working precision and expansion order are parameters of the computation.

We use the following elementary form of the Kato–Temple inequality. Let H be a compact self-adjoint operator with simple largest eigenvalue s , and let q_1 be a corresponding unit eigenvector. Suppose the spectrum of H on $q_1^\perp$ is contained in $(-\infty, \beta]$. If q is a unit vector and $\mu = \langle q, Hq \rangle > \beta$, then

$$\mu \leq s \leq \mu + \frac{\|Hq - \mu q\|^2}{\mu - \beta}. \quad (6)$$

Indeed, $(s - H)(H - \beta)$ is nonpositive on the orthogonal complement of the leading eigenvector and vanishes on it. Taking its expectation gives $(s - \mu)(\mu - \beta) \leq \|Hq - \mu q\|^2$. Spectral projection also gives

$$\sin \angle(q, q_1) \leq \frac{\|Hq - \mu q\|}{\mu - \beta}. \quad (7)$$

More generally, if P_1 projects onto the leading eigenspace, then for any shift $t > \beta$ and any unit vector q ,

$$\|(I - P_1)q\| \leq \frac{\|(H - t)q\|}{t - \beta}. \quad (8)$$

This follows by expanding in a spectral basis on $q_1^\perp$, where $|t - \lambda_i| \geq t - \beta$. We use this form when the residual is centered at a known approximate eigenvalue rather than its Rayleigh quotient.

For $H_B = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$ and $q = (\hat{x}, \hat{y})/\sqrt{2}$, the quantities in (5) are the Rayleigh quotient and squared residual. Appendix B gives an upper bound F for $\|B\|_{\text{HS}}^2$. Choose $\beta_0 \geq 0$ such that $F - \mu^2 \leq \beta_0^2$ and $\mu > \beta_0$, checking both inequalities with outward bounds. Then $\sigma_2(B) \leq \beta_0$, and (6) applies.

The Collatz–Wielandt upper ratio bound [12] gives, for positive trial vectors,

$$u_{\text{rat}} = \left(\max_i \frac{(B\hat{y})_i}{\hat{x}_i} \max_i \frac{(B^T \hat{x})_i}{\hat{y}_i} \right)^{1/2}.$$

Indeed, if the two maxima are a_0, b_0 , then $B^T B \hat{y} \leq a_0 b_0 \hat{y}$; taking its inner product with the positive leading eigenvector of $B^T B$ gives $s^2 \leq a_0 b_0$. We retain the upper bound

$$u_* = \min \left\{ u_{\text{rat}}, \mu + \frac{R_B}{\mu - \beta_0} \right\}.$$

2.1 A compression bound for the remaining singular values

Select disjoint subsets of the sampled coordinates. Each set may be one retained integer, a block of retained integers, or all the nodes on a counting cell. If such a set I has total weight $m_I = \sum_{i \in I} w_i$, define

$$(p_I)_i = \begin{cases} \sqrt{w_i/m_I}, & i \in I, \\ 0, & i \notin I. \end{cases}$$

These vectors are orthonormal. Let P_r and P_c collect any two such families, for rows and columns respectively, and put $\mathcal{C} = P_r^T B P_c$. Its entries are

$$\mathcal{C}_{IJ} = \frac{1}{\sqrt{m_I m_J}} \sum_{i \in I} \sum_{j \in J} w_i w_j a(c_i, c_j). \quad (9)$$

Lemma 2. *Let W be a real matrix with as many rows as $\mathcal{C}$ has columns, and let k be its number of columns. Put*

$$\varepsilon_W = \|W^T W - I\|_{\text{HS}},$$

and let

$$\theta_1 \geq \dots \geq \theta_k \geq 0$$

be the eigenvalues of $(CW)^T CW$. Then

$$\sum_{j \geq 3} \sigma_j(B)^2 \geq \frac{1}{1 + \varepsilon_W} \sum_{j=3}^k \theta_j =: m_{\perp}. \quad (10)$$

An empty sum is zero.

Proof. The singular-value variational principle gives $\sigma_j(\mathcal{C}) \leq \sigma_j(B)$, because P_r and P_c are isometries. Also $\sigma_j(CW) \leq \|W\| \sigma_j(\mathcal{C})$ and $\|W\|^2 \leq 1 + \varepsilon_W$. Squaring these inequalities and summing over $j = 3, \dots, k$ proves the result. $\square$

In the computation, W has exact dyadic entries. Interval arithmetic encloses ε_W and the eigenvalues, whose ordering is checked by disjoint intervals. Substituting nonnegative lower endpoints for θ_j and an upper endpoint for ε_W in (10) gives a certified lower bound for the sum. If $0 < l \leq s = \|B\|$ is already known, the upper bound F for $\|B\|_{\text{HS}}^2$ gives

$$\sigma_2(B) \leq \beta := (F - l^2 - m_{\perp})^{1/2}. \quad (11)$$

The calculation checks that the radicand is nonnegative and that $\mu > \beta$.

We enclose the entries in (9) by Gaussian rules of smaller order on the selected cells. Write I_a for counting measure on $[a, 2a - 1]$, G_a for its sampled rule, and R_a for an auxiliary Gaussian rule. Define

$$e_a = (I_a - G_a)(t - \tfrac{1}{2})^{-2}, \quad f_a = (I_a - R_a)(t - \tfrac{1}{2})^{-2}.$$

Both errors are nonnegative. An exactly represented integer block has $G_a = I_a$ and $e_a = 0$.

For either variable, with the other argument at least one, the factorization of the kernel gives

$$\left| \frac{d^n}{dt^n} a(t, y) \right|, \left| \frac{d^n}{dt^n} a(y, t) \right| \leq 2(n+1)! (t - \tfrac{1}{2})^{-n-2}.$$

Indeed, both kernel orientations are twice a product of two reciprocals whose shifts have real part at least $-1/2$. Leibniz's rule proves the displayed inequality. The positive Gaussian Peano kernel therefore gives, uniformly in the other argument,

$$|(I_a - G_a)a| \leq 2e_a, \quad |(I_a - R_a)a| \leq 2f_a, \quad |(G_a - R_a)a| \leq 2(e_a + f_a).$$

For two selected cells, positivity and their exact masses give

$$\begin{aligned} |(G_a G_b - R_a R_b)a| &\leq 2b(e_a + f_a) + 2a(e_b + f_b), \\ G_a G_b - R_a R_b &= (G_a - R_a)G_b + R_a(G_b - R_b). \end{aligned} \quad (12)$$

The entry enclosure is the auxiliary double sum, with the error in (12), divided by $\sqrt{ab}$. With one singleton coordinate, the corresponding allowance is $2(e_a + f_a)/\sqrt{a}$. Singleton-to-singleton entries are exact kernel values. Each error bound remains valid when e_a is replaced by any certified upper bound.

2.2 A Laplace approximation and a resolvent bound

Let

$$r = H_B q - \mu q, \quad Q = I - qq^T, \quad C = QH_B Q|_{q^\perp}.$$

The Schur complement gives

$$s - \mu = \langle r, (s - C)^{-1} r \rangle.$$

We bound this quadratic form using a finite-rank Laplace approximation of B . The resulting resolvent is evaluated through a linear system whose order is twice the approximation rank. The estimates below bound the approximation error and the error from evaluating the resolvent at a fixed shift.

For $x \geq 1$, put

$$\rho(x) = \sqrt{2x - \frac{7}{4}}, \quad f(x) = x - \frac{1}{2} - \rho(x), \quad u_x(t) = e^{-xt}, \quad v_x(t) = e^{-(x-1/2)t} \frac{\sinh(\rho(x)t)}{\rho(x)}.$$

Direct integration gives

$$a(x, y) = 2 \int_0^\infty u_x(t) v_y(t) dt. \quad (13)$$

Indeed, the denominator of this integral is $(x + y - 1/2)^2 - \rho(y)^2 = (x + y - 1)^2 + x - y + 1$. Define the sampled vectors

$$u(t) = (\sqrt{w_i} u_{c_i}(t))_i, \quad v(t) = (\sqrt{w_i} v_{c_i}(t))_i.$$

Thus $B = 2 \int_0^\infty u(t) v(t)^T dt$.

We use the logarithmic substitution $t = e^w$ and strip estimates for operator quadrature as in Bonito, Lei, and Pasciak [11, §3]. The factorization and error constants below are derived for the present kernel; the explicit trapezoidal bound is Theorem 5.1 of Trefethen and Weideman [14].

Choose $h > 0$ and integers $M, L \geq 0$. For $-M \leq \ell \leq L$, set $t_\ell = e^{\ell h}$ and form the columns

$$U_\ell^0 = \sqrt{2ht_\ell} u(t_\ell), \quad V_\ell^0 = \sqrt{2ht_\ell} v(t_\ell).$$

Choose a boundary $D_\ell \geq 2$ between complete cells of the sampled measure, and set the entries with $c_i \geq D_\ell$ to zero in both columns. Denote the resulting columns by U_ℓ, V_ℓ . The choice $D_\ell = N + 1$ retains every entry. With $m = M + L + 1$, write

$$\mathcal{U} = (U_{-M}, \dots, U_L), \quad \mathcal{V} = (V_{-M}, \dots, V_L), \quad \tilde{B} = \mathcal{U} \mathcal{V}^T. \quad (14)$$

The rank is at most m .

Here and below $\|X\|_{\text{tr}}$ denotes the nuclear norm, the sum of the singular values of a matrix. In particular, $\|ab^T\|_{\text{tr}} = \|a\| \|b\|$, also for complex vectors. The error bound for $B - \tilde{B}$ will be an operator-norm bound.

Bounds for the Laplace factors. All sampled nodes belong to $\{1\} \cup [2, \infty)$. On this set,

$$0 \leq v_x(t) \leq \min(t, 1) e^{-f(x)t}, \quad t > 0. \quad (15)$$

The two bounds follow from $\sinh(\rho t) \leq \rho t e^{\rho t}$ and $\sinh(\rho t) \leq e^{\rho t}/2$, since $2\rho(x) \geq 1$. Moreover, $f(1) = 0$ and $f(x) \geq (x-2)/3$ for $x \geq 2$. The latter inequality follows by squaring $\rho(x) \leq (4x+1)/6$; the difference of the squares is $4(x-2)^2/9$.

For $b > 0$, Gaussian quadrature on each cell underestimates e^{-bx} : its derivative of even order $2p_a$ is positive, so the Hermite interpolation remainder is nonnegative. Consequently, for any boundary D between complete cells,

$$\sum_i w_i e^{-bc_i} \leq \frac{1}{e^b - 1}, \quad \sum_{c_i \geq D} w_i e^{-b(c_i - D)} \leq \frac{1}{1 - e^{-b}}. \quad (16)$$

The same argument applies to the exact singleton cells. Equations (15)–(16) give

$$\|u(t)\|^2 \leq \frac{1}{e^{2t} - 1}, \quad \|v(t)\|^2 \leq \min(t, 1)^2 \left(1 + \frac{1}{1 - e^{-2t/3}}\right). \quad (17)$$

The logarithmic quadrature error. Choose $0 < \gamma < 2$, $0 < d < \pi/2$, and a cell boundary $2 \leq J \leq n_0 + 1$, so that the nodes $1, \dots, J-1$ are exact singletons. Define the Hurwitz zeta value $\zeta(2, z) = \sum_{k=0}^{\infty} (k+z)^{-2}$ for $z > 0$, and put

$$U_\gamma = \frac{1}{4} \zeta(2, 1 - \gamma/2), \quad \alpha_J = 1 - \frac{1}{\rho(J)} > 0, \quad (18)$$

$$V_{\gamma, J}^{\text{tail}} = \frac{1}{4\alpha_J^2} \zeta\left(2, \frac{\gamma + 2f(J)}{2\alpha_J}\right),$$

$$b_j = j - \frac{1}{2} + \gamma/2, \quad I_j(\gamma, d) = \frac{1}{4\rho(j)^2} \log \frac{b_j^2 + \rho(j)^2 \tan^2 d}{b_j^2 - \rho(j)^2},$$

$$V_{\gamma, d, J} = \sum_{j=1}^{J-1} I_j(\gamma, d) + \sec^2 d V_{\gamma, J}^{\text{tail}}. \quad (19)$$

Every logarithm has positive numerator and denominator, since $b_j > \rho(j)$ for $\gamma > 0$.

For $z = \tau(1 + i \tan d)$, $\tau > 0$, the exact identity

$$|v_j(z)|^2 = e^{-2(j-1/2)\tau} \frac{\cosh(2\rho(j)\tau) - \cos(2\rho(j)\tau \tan d)}{2\rho(j)^2}$$

yields

$$\int_0^\infty e^{-\gamma\tau} |v_j(z)|^2 \frac{d\tau}{\tau} = I_j(\gamma, d). \quad (20)$$

To evaluate it, expand the hyperbolic cosine into exponentials and use $\int_0^\infty (e^{-at} - e^{-bt}) dt/t = \log(b/a)$ for $a, b > 0$. The remaining identity $\int_0^\infty e^{-at} (1 - \cos(bt)) dt/t = \frac{1}{2} \log(1 + b^2/a^2)$ follows by differentiation in b .

The function f is convex, and hence $f(x) \geq f(J) + \alpha_J(x - J)$ for $x \geq J$. Also $|v_x(\tau(1 + i \tan d))| \leq \sec d v_x(\tau)$, as follows from the integral of $\cosh(\rho(x)s)$ along the straight segment from 0 to z . Using $v_x(\tau) \leq \tau e^{-f(x)\tau}$ and (16), termwise integration of a positive geometric series gives

$$\int_0^\infty \tau e^{\gamma\tau} \|u(\tau)\|^2 d\tau \leq U_\gamma, \quad \int_0^\infty e^{-\gamma\tau} \|v(\tau(1 + i \tan d))\|^2 \frac{d\tau}{\tau} \leq V_{\gamma, d, J}.$$

For $\mathcal{G}(w) = 2e^w u(e^w) v(e^w)^T$, the substitution $\tau = e^x \cos d$ and Cauchy–Schwarz therefore imply

$$\int_{\mathbb{R}} \|\mathcal{G}(x \pm id)\|_{\text{tr}} dx \leq 2 \sec d \sqrt{U_\gamma V_{\gamma, d, J}} =: M_d.$$

The same bound holds at every smaller absolute angle, since its right-hand side increases with d . The function $\mathcal{G}$ is analytic in $|\text{Im } w| < \pi/2$ and tends to zero uniformly at both ends of each

closed narrower strip, by (17) and the complex comparison above. Theorem 5.1 of Trefethen and Weideman [14], applied to linear functionals Λ satisfying $|\Lambda(X)| \leq \|X\|_{\text{tr}}$, now gives

$$\left\| B - 2h \sum_{\ell \in \mathbb{Z}} t_\ell u(t_\ell) v(t_\ell)^T \right\|_{\text{tr}} \leq \delta_{\text{strip}} := \frac{4 \sec d \sqrt{U_\gamma V_{\gamma, d, J}}}{e^{2\pi d/h} - 1}. \quad (21)$$

The two omitted ends of the sum. Put $\theta = e^{-h}$ and $t_- = e^{-(M+1)h}$. Choose $0 < \varepsilon < 1$ and define

$$a_\varepsilon = 1 - \varepsilon, \quad b_\varepsilon = \frac{1}{2} + \frac{1}{2\varepsilon} - \frac{7\varepsilon}{8}.$$

The identity

$$f(x) - a_\varepsilon x + b_\varepsilon = \frac{\varepsilon}{2} (\rho(x) - 1/\varepsilon)^2$$

gives a positive entrywise majorant for the omitted terms at $t_i = t_- \theta^i$, $i \geq 0$:

$$2h \sum_{i \geq 0} t_i u(t_i) v(t_i)^T \leq 2h e^{b_\varepsilon t_-} \sum_{i \geq 0} t_i^2 u(t_i) u(a_\varepsilon t_i)^T.$$

Here the inequality is entrywise. The Gram matrices of the two families $t_i u(t_i)$ and $t_i u(a_\varepsilon t_i)$ are bounded entrywise by $\mathcal{C}$ and $\mathcal{C}/a_\varepsilon$, respectively, where

$$\mathcal{C}_{ij} = \frac{t_i t_j}{t_i + t_j}.$$

This follows from (16) and $(e^s - 1)^{-1} \leq s^{-1}$. Entrywise domination of nonnegative matrices implies domination of their operator norms, by applying the majorant to the absolute value of each vector. For finite families, the squared norm of their outer-product sum is the spectral radius of the product of these two Gram matrices. Positivity of all entries and monotonicity of the spectral radius give the bound $\|\mathcal{C}\|^2/a_\varepsilon$. The infinite limit follows because the sums of squared column norms are bounded by the summable diagonals of $\mathcal{C}$ and $\mathcal{C}/a_\varepsilon$; the outer-product sums converge in nuclear norm.

The matrix $\mathcal{C}$ is the Gram matrix of $t_i e^{-t_i x}$ in $L^2(0, \infty)$. Splitting the corresponding sum of positive rank-one operators after its first $K \geq 1$ summands gives

$$\|\mathcal{C}\| \leq \|\mathcal{C}_K\| + \frac{t_- \theta^K}{2(1 - \theta)}.$$

The Collatz–Wielandt upper ratio bound [12] applies here. Let $p \in \mathbb{R}^K$ have strictly positive entries, and put

$$(D_K)_{ij} = \frac{\theta^{i+j}}{\theta^i + \theta^j}, \quad 0 \leq i, j < K, \quad \Lambda_K = \max_{0 \leq i < K} \frac{(D_K p)_i}{p_i}. \quad (22)$$

The componentwise inequality $D_K p \leq \Lambda_K p$ implies $\|D_K\| \leq \Lambda_K$: conjugation by $\text{diag}(p)$ bounds the spectral radius by the maximum row sum, and D_K is positive semidefinite. Thus the negative end has operator norm at most

$$\delta_- = \frac{2h e^{b_\varepsilon t_-} t_-}{\sqrt{a_\varepsilon}} \left(\Lambda_K + \frac{\theta^K}{2(1 - \theta)} \right). \quad (23)$$

For the positive end put $t_+ = e^{(L+1)h} > 1$ and

$$\mathcal{K}(t) = \left(\frac{1 + (1 - e^{-2t/3})^{-1}}{1 - e^{-2t}} \right)^{1/2}.$$

Equation (17) gives $2\|u(t)\|\|v(t)\| \leq 2\mathcal{K}(t)e^{-t}$. The function $\mathcal{K}$ decreases, as does te^{-t} for $t > 1$. The first term plus the integral of this decreasing function on the logarithmic mesh bounds the positive end in nuclear norm by

$$\delta_+ = 2h\mathcal{K}(t_+)(t_+ + 1/h)e^{-t_+}. \quad (24)$$

Truncation of the columns. For $t = t_\ell$ and $D = D_\ell$, set

$$\begin{aligned} n_u(t)^2 &= \frac{2ht}{e^{2t} - 1}, & n_v(t)^2 &= 2ht \min(t, 1)^2 \left(1 + \frac{1}{1 - e^{-2t/3}}\right), \\ e_u(t, D)^2 &= \frac{2hte^{-2Dt}}{1 - e^{-2t}}, & e_v(t, D)^2 &= \frac{2ht \min(t, 1)^2 e^{-2(D-2)t/3}}{1 - e^{-2t/3}}. \end{aligned}$$

For $D = N + 1$, set both e_u and e_v to zero. These quantities bound the full and omitted column norms, by (16)–(17). Given a tolerance $\tau_f > 0$, a cutoff rule is to take the smallest power of two $D \leq N + 1$ for which both omitted norm bounds are at most τ_f , using $N + 1$ if needed. The individual cutoffs D_ℓ are recorded with the numerical data. Since $U_\ell^0(V_\ell^0)^T - U_\ell V_\ell^T = (U_\ell^0 - U_\ell)(V_\ell^0)^T + U_\ell(V_\ell^0 - V_\ell)^T$, their total nuclear-norm error is at most

$$\delta_{\text{cut}} = \sum_{\ell=-M}^L [e_u(t_\ell, D_\ell)n_v(t_\ell) + n_u(t_\ell)e_v(t_\ell, D_\ell)]. \quad (25)$$

Combining the four estimates, using the operator norm for the negative end and the nuclear norm for the other terms, proves

$$\|B - \tilde{B}\| \leq \delta_{\text{app}} := \delta_{\text{strip}} + \delta_- + \delta_+ + \delta_{\text{cut}}. \quad (26)$$

A spectral bound for the compression. With β from Section 2.1 and the upper bound u_* from Section 2, define

$$\eta = \frac{R_B}{(\mu - \beta)^2}, \quad \kappa = \beta + (u_* - \beta)\eta. \quad (27)$$

Then $\lambda_{\max}(C) \leq \kappa$. Indeed, for any unit $z \perp q$, (7) implies $|\langle z, q_1 \rangle|^2 \leq \eta$. The spectral bound on $q_1^\perp$ therefore gives $\langle z, H_B z \rangle \leq \beta + (s - \beta)|\langle z, q_1 \rangle|^2 \leq \kappa$. The calculation below requires $\kappa < \mu$.

Form $\tilde{H} = \begin{pmatrix} 0 & \tilde{B} \\ \tilde{B}^T & 0 \end{pmatrix}$ and $C_0 = Q\tilde{H}Q|_{q^\perp}$. Then $\|C - C_0\| \leq \delta_{\text{app}}$. Choose an exact shift $\xi > \kappa + \delta_{\text{app}}$ and set

$$v_0 = (\xi - C_0)^{-1}r, \quad \phi_0 = \langle r, v_0 \rangle, \quad \nu_L = \|(v_0)_L\|^2, \quad \nu_R = \|(v_0)_R\|^2. \quad (28)$$

The subscripts L, R denote the two coordinate halves of the dilation.

Evaluation by a small linear system. Put

$$Z = \text{diag}(\mathcal{U}, \mathcal{V}), \quad \mathcal{J} = \begin{pmatrix} 0 & I_m \\ I_m & 0 \end{pmatrix}, \quad G_0 = Z^T Z = \text{diag}(G_U, G_V), \quad w = Z^T q.$$

Thus $\tilde{H} = Z\mathcal{J}Z^T$. Choose an exact positive scale α , and let $g = Z^T r/\alpha$. Solve the system of order $2m$

$$(\xi\mathcal{J} - G_0 + ww^T)z = g, \quad \sigma = w^T z. \quad (29)$$

The Woodbury identity [13, (3)] on $q^\perp$ gives

$$v_0 = \frac{r + \alpha(Zz - q\sigma)}{\xi}, \quad \phi_0 = \frac{R_B + \alpha^2 g^T z}{\xi}. \quad (30)$$

With $W = QZ$ one has $W^T W = G_0 - ww^T$ and $C_0 = W\mathcal{J}W^T$. Substitution of (29) into (30) gives $(\xi - C_0)v_0 = r$, proving the formula. To check nonsingularity, suppose $(\xi\mathcal{J} - W^T W)z = 0$ with $z \neq 0$. Then $Wz \neq 0$, and $C_0 Wz = \xi Wz$, contradicting $\lambda_{\max}(C_0) \leq \kappa + \delta_{\text{app}} < \xi$.

Define the squared residuals of the two singular-vector equations

$$\mathcal{R}_L = \|B\hat{y} - \mu\hat{x}\|^2, \quad \mathcal{R}_R = \|B^T \hat{x} - \mu\hat{y}\|^2.$$

Since $q = (\hat{x}, \hat{y})/\sqrt{2}$, the exact relations are

$$\|q_L\|^2 = \|q_R\|^2 = \frac{1}{2}, \quad \|r_L\|^2 = \mathcal{R}_L/2, \quad \|r_R\|^2 = \mathcal{R}_R/2, \quad \langle r_L, q_L \rangle = \langle r_R, q_R \rangle = 0.$$

The two vanishing inner products both equal $(\hat{x}^T B\hat{y} - \mu)/2$. Expanding the two norms in (30) now yields

$$\begin{aligned} \xi^2 \nu_L &= \frac{\mathcal{R}_L}{2} + \alpha^2 \left(2g_L^T z_L + z_L^T G_U z_L - 2\sigma w_L^T z_L + \frac{\sigma^2}{2} \right), \\ \xi^2 \nu_R &= \frac{\mathcal{R}_R}{2} + \alpha^2 \left(2g_R^T z_R + z_R^T G_V z_R - 2\sigma w_R^T z_R + \frac{\sigma^2}{2} \right). \end{aligned} \quad (31)$$

Thus $G_U, G_V, w, g, \mathcal{R}_L, \mathcal{R}_R$ determine ϕ_0, ν_L, ν_R . Interval arithmetic encloses the solution of (29) and evaluates (30) and (31).

The singular-value enclosure. Define

$$\begin{aligned} e_1 &= 2\delta_{\text{app}}\sqrt{\nu_L \nu_R}, & e_2 &= \frac{\delta_{\text{app}}^2(\nu_L + \nu_R)}{\xi - \kappa}, \\ e_{\text{shift}} &= \frac{R_B|\xi - \mu|}{(\min(\xi, \mu) - \kappa)^2}, & e_{\text{nl}} &= \frac{R_B^2}{(\mu - \kappa)^3}. \end{aligned} \quad (32)$$

Then

$$\mu + \phi_0 - e_1 - e_{\text{shift}} - e_{\text{nl}} \leq s \leq \mu + \phi_0 + e_1 + e_2 + e_{\text{shift}}. \quad (33)$$

To prove this, write $E = C - C_0$ and $F_C(t) = \langle r, (t - C)^{-1} r \rangle$ for $t > \kappa$. Two successive resolvent identities give

$$F_C(\xi) = \phi_0 + \langle v_0, E v_0 \rangle + \langle E v_0, (\xi - C)^{-1} E v_0 \rangle.$$

The last term is nonnegative and at most e_2 . Since $v_0 \perp q$, the middle term is the quadratic form of $H_B - \tilde{H}$. Its diagonal blocks vanish, so

$$|\langle v_0, E v_0 \rangle| = |2\langle (v_0)_L, (B - \tilde{B})(v_0)_R \rangle| \leq e_1.$$

Also $|F'_C(t)| \leq R_B/(t - \kappa)^2$, which bounds $|F_C(\xi) - F_C(\mu)|$ by e_{shift} . The Schur complement of $s - C$ in $s - H_B$ gives $s - \mu = F_C(s)$. Positivity and monotonicity of F_C imply

$$0 \leq F_C(\mu) - (s - \mu) \leq \frac{(s - \mu)R_B}{(\mu - \kappa)^2} \leq \frac{R_B^2}{(\mu - \kappa)^3} = e_{\text{nl}},$$

which proves (33).

Intersect (33) with $[\mu, u_*]$ and call the result $[l_B, u_B]$. In interval arithmetic, positive error terms are bounded upward and positive denominators downward; the full enclosure of μ is retained in both endpoints.

Define the weighted sums

$$S_x = \sum_i \sqrt{w_i} x_i, \quad S_y = \sum_i \sqrt{w_i} y_i. \quad (34)$$

If l_0 is a lower endpoint for the trial Rayleigh quotient μ , then each trial half differs from its exact singular vector by at most $2\sqrt{R_B}/(l_0 - \beta_0)$, provided $l_0 > \beta_0$. Hence each weighted trial sum has error at most

$$e_S = \frac{2\sqrt{NR_B}}{l_0 - \beta_0}. \quad (35)$$

Here $l_0 \leq \mu$, so the denominator in (7) is at least $l_0 - \beta_0$.

3 Comparison with the integer section

Write $s_N = \|A_N\|$. Define

$$U_{ji} = a(j, c_i)\sqrt{w_i}, \quad V_{ji} = a(c_i, j)\sqrt{w_i}, \quad 1 \leq j \leq N.$$

Applying the sampled singular vectors to the kernel at integer arguments gives Uy and Vx , which, after division by s , are their Nyström extensions evaluated at the integers. Sloan's iterated Galerkin method [8] similarly improves a discrete eigenelement by reapplying the operator. Allouch and Tahrichi [9, §§2, 5] analyze modified Nyström operators and their iterated eigenspaces. We evaluate the combined Rayleigh quotient of Uy and Vx and bound its single and double quadrature errors. Form the vector $z = (Uy, Vx)/\sqrt{2}$. With $\mathcal{H}_N = \begin{pmatrix} 0 & A_N \\ A_N^T & 0 \end{pmatrix}$, put

$$d = \|z\|^2, \quad g = d - s^2, \quad r_{\text{lift}} = \mathcal{H}_N z - sz.$$

The following identity gives the Rayleigh quotient of $z/\sqrt{d}$:

$$\frac{\langle z, \mathcal{H}_N z \rangle}{d} = s + \frac{sg}{s^2 + g} + \frac{\omega}{s^2 + g}, \quad (36)$$

where

$$M = U^T A_N V - U^T U B^T - B^T V^T V + B^T B B^T, \\ \omega = y^T M x.$$

To verify (36), use $B^T x = sy$ and $By = sx$ to obtain $y^T U^T A_N V x = 2sd - s^3 + \omega$. The differences $U^T U - B^T B$ and $V^T V - BB^T$ replace one quadrature sum by an integer sum; their quadratic forms give g . The four terms defining M make this replacement in both summation variables. Indeed, if I is counting measure on $\{1, \dots, N\}$ and G is the composite Gaussian rule, then

$$M_{ij} = \sqrt{w_i w_j} (I_z - G_z)(I_w - G_w) [a(z, c_i) a(z, w) a(c_j, w)]. \quad (37)$$

We next enclose g , ω , and $\|r_{\text{lift}}\|^2$. Let I_a denote counting measure on $[a, 2a - 1]$ and G_a its Gaussian rule. For an active cell define the positive scalar errors

$$E_{r,a} = (I_a - G_a)(t - \frac{1}{2})^{-r}, \quad r > 0, \quad \tau_a = \frac{p_a + 2}{2a - 1}. \quad (38)$$

The verifier evaluates $E_{2,a}$, $E_{4,a}$, and $E_{6,a}$ through a finite Jacobi continued fraction. In the coordinate $u = (t - 1/2)/a$, the $p_a \times p_a$ Jacobi matrix J_a for normalized counting measure has constant diagonal $d_a = (3a - 2)/(2a)$ and squared off-diagonal entries

$$b_k^2 = \frac{k^2(a^2 - k^2)}{4a^2(4k^2 - 1)}, \quad 1 \leq k < p_a.$$

These follow by rescaling (4). The Gaussian nodes are its eigenvalues, and the weights divided by a are the squared first components of its normalized eigenvectors [6, §2, (2.6)]. Thus

$$F_a(z) = e_1^T (J_a + zI)^{-1} e_1 = \sum_{i \text{ in cell } a} \frac{w_i/a}{(c_i - 1/2)/a + z}. \quad (39)$$

Successive Schur complements evaluate this function by $F_{p_a}(z) = (d_a + z)^{-1}$ and $F_k(z) = (d_a + z - b_k^2 F_{k+1}(z))^{-1}$, with $F_a = F_1$. For even r , its coefficient of z^{r-1} is $-a^{r-1} G_a(t - 1/2)^{-r}$. The integer sum is

$$I_a(t - \frac{1}{2})^{-r} = \zeta(r, a - \frac{1}{2}) - \zeta(r, 2a - \frac{1}{2}).$$

Series arithmetic through degree 5, followed by this subtraction, encloses all three scaled errors $a^{r-1} E_{r,a}$ without computing the Gaussian nodes. Positivity follows from the Gaussian Hermite remainder. If subtraction gives a negative lower endpoint, it may therefore be replaced by zero.

Lemma 3. *For $a \geq 3$ and $t, u \geq 1$, either orientation of the kernel has a function $\Phi_a(t)$ satisfying*

$$0 < \Phi_a(t) \leq \min\{1, 4a^2/(t + a)^2\},$$

such that

$$|(I_a - G_a)[a(t, \cdot)a(u, \cdot)]| \leq 4E_{4,a}\Phi_a(t)\Phi_a(u). \quad (40)$$

The bounds also hold with either kernel transposed, with the corresponding orientation of Φ_a in each factor. Moreover, $\sum_{j=1}^N \Phi_a(j)^2 \leq 16a/3$ and $\sum_i w_i \Phi_a(c_i)^2 \leq 16a/3$.

Proof. Both orientations factor as $2/[(z + \alpha_1)(z + \alpha_2)]$. For $a(t, z)$ the shifts are $t - 3/2 \pm i\sqrt{8t - 1}/2$; for $a(z, t)$ they are $t - 1/2 \pm \sqrt{8t - 7}/2$. Their real parts are at least $-1/2$. The divided difference of a reciprocal is $(-1)^n / \prod_{j=0}^n (z_j + \alpha)$. The product rule for divided differences expresses a product of r reciprocals as a sum of terms with $n + r$ denominator factors. At least one factor contains each shift. For $z \in [a, b]$, $b = 2a - 1$, bound those selected factors using

$$\frac{|z + \alpha|}{z - 1/2} \geq \frac{b + \Re \alpha}{b - 1/2} \geq 1.$$

For the other reciprocal factors use $|z + \alpha|^{-1} \leq (z - 1/2)^{-1}$. At even order $n = 2p_a$ the comparison divided difference of $(z - 1/2)^{-r}$ is positive. Hermite interpolation at the repeated Gaussian nodes and subsequent summation prove the product error bound, with

$$\Phi_a^f(t) = \frac{(b - 1/2)^2}{(b + t - 3/2)^2}, \quad \Phi_a^t(t) = \frac{(b - 1/2)^2}{(b + t - 1/2)^2 - (8t - 7)/4}.$$

Direct expansion proves the asserted pointwise bounds. Finally $\sum_{j \geq 1} (j + a)^{-4} \leq 1/(3a^3)$. The same estimate holds for Gaussian measure: Hermite interpolation implies $G_a f \leq I_a f$ whenever $f^{(2p_a)} \geq 0$, and in particular for $f(t) = (t + a)^{-4}$. $\square$

Set

$$P = \frac{S_x^2 + S_y^2}{2}, \quad T_4 = \sum_a E_{4,a}, \quad T_{1/2,4} = \sum_a \sqrt{a} E_{4,a}.$$

In any upper bound, S_x, S_y may be replaced by upper endpoints. The singular equations give

$$g = \frac{1}{2} \left[y^T (U^T U - B^T B) y + x^T (V^T V - B B^T) x \right],$$

and

$$r_{\text{lift}} = \frac{1}{\sqrt{2}} \begin{pmatrix} (A_N V - U B^T) x \\ (A_N^T U - V B) y \end{pmatrix}.$$

Each matrix entry on the right is a quadrature error for a product of two kernels. Lemma 3 therefore gives

$$|g| \leq 4PT_4, \quad \|r_{\text{lift}}\|^2 \leq \frac{256}{3} PT_{1/2,4}^2 =: \mathcal{R}. \quad (41)$$

For example, a residual coordinate from cell a is bounded by $4E_{4,a}S_y\Phi_a(j)$ in one half, and by $4E_{4,a}S_x\Phi_a(j)$ in the other. Apply the last assertion of Lemma 3, sum the norms over cells, and average the two squared half norms. This proves the second bound.

3.1 A signed bound for the Gram error

On an active cell, the contribution to the squared-norm difference is

$$g_a = \frac{1}{2}(I_a - G_a) \left[\left(\sum_i \sqrt{w_i} y_i a(t, c_i) \right)^2 + \left(\sum_i \sqrt{w_i} x_i a(c_i, t) \right)^2 \right], \quad g = \sum_a g_a.$$

The bounds below give an interval for each g_a ; summing these intervals encloses g . We call $a(t, c_i)$ the row orientation and $a(c_i, t)$ the column orientation. Their factorizations in t have real and complex conjugate shifts, respectively. The estimates use scalar quadrature errors and the first coordinates of the sampled singular vectors.

Bounds from the trial vectors. Let $H_0 \leq n_0$ and retain the first H_0 coordinates of the normalized full trial vectors $\hat{x}, \hat{y}$. Positivity and (7) give, for either half,

$$\|(v_j - \hat{v}_j)_{j=1}^{H_0}\| \leq d_h := \frac{2\sqrt{R_B}}{l_0 - \beta}, \quad v = x \text{ or } y, \quad (42)$$

where l_0 is a lower endpoint for μ and $l_0 > \beta$. Indeed, if $\alpha > 0$ is the overlap of the two unit dilation vectors, then $1 - \alpha^2 \leq R_B/(l_0 - \beta)^2$ and $2(1 - \alpha) \leq 2(1 - \alpha^2)$. Multiplication by $\sqrt{2}$ bounds each half, and restriction to the first H_0 coordinates decreases the norm.

For nonnegative coefficients $b_1, \dots, b_H$, $H \leq H_0$, define

$$\mathcal{L}_v(b; H) = \max \left\{ 0, \sum_{j=1}^H b_j \hat{v}_j - d_h \left(\sum_{j=1}^H b_j^2 \right)^{1/2} \right\}. \quad (43)$$

Then $\sum_{j=1}^H b_j v_j \geq \mathcal{L}_v(b; H)$. The saved coordinate intervals and the distance bound are both retained when evaluating this expression.

The singular equations and the bounds $a(t, 1) \leq 2/t^2$ and $a(1, t) \leq 16/(7t^2)$ imply

$$x_i \leq K_x \frac{\sqrt{w_i}}{c_i^2}, \quad y_i \leq K_y \frac{\sqrt{w_i}}{c_i^2}, \quad K_x = \frac{2S_y}{s}, \quad K_y = \frac{16S_x}{7s}. \quad (44)$$

For the second kernel bound, use $1 - 1/t + 2/t^2 \geq 7/8$; monotonicity in the other coordinate gives the bounds for every source node.

For $v = x$ or y , put

$$C_v(\gamma) = \sum_i \sqrt{w_i} v_i \min\{\gamma c_i, 1\}, \quad \gamma \geq 0.$$

Let $H \leq H_0$, let $b \geq H$ be a boundary after a retained singleton or a complete quadrature cell, and set $\chi_j = \min\{\gamma j, 1\}$ and $D_{\gamma, H} = (\sum_{j=1}^H \chi_j^2)^{1/2}$. Gaussian domination of t^{-1} and t^{-2} gives

$$C_v(\gamma) \leq \min \left\{ S_v, \min \left\{ D_{\gamma, H}, \sum_{j=1}^H \chi_j \hat{v}_j + d_h D_{\gamma, H} \right\} \right. \\ \left. + K_v [\gamma(H_b - H_H) + \zeta(2, b+1)] \right\}. \quad (45)$$

Here $H_b = \sum_{j=1}^b j^{-1}$ and $H_0 = 0$. The first minimum bounds the retained terms by Cauchy–Schwarz. For the remaining terms up to b , use $\min\{\gamma c_i, 1\} \leq \gamma c_i$ and (44); beyond b , use $\min\{\gamma c_i, 1\} \leq 1$. We minimize (45) over the chosen integers H and admissible boundaries b . For fixed H , the variable part changes by $\gamma/(b+1) - (b+1)^{-2}$ when b increases by one. For $\gamma > 0$, it therefore suffices to check the admissible boundaries immediately below and above $1/\gamma$, subject to $b \geq H$. For $\gamma = 0$, $C_v(0) = 0$.

Basic signed estimates. Write $E_r = E_{r,a}$, $p = p_a$, and $\tau = \tau_a$ as in (38). The factorization used in Lemma 3 gives

$$g_a \geq \max \left\{ [2(S_x^2 + S_y^2) - 4S_y C_y(2\tau) - 8S_x C_x(4\tau/3)] E_4, -2S_x^2 E_4 \right\}, \quad (46)$$

$$g_a \leq 2(S_x^2 + \theta_a S_y^2) E_4, \quad \theta_a = \left(\frac{2a - 3/2}{2a - 1} \right)^{2p+4}. \quad (47)$$

To see the lower bound, write each product of two kernels as four reciprocal factors relative to $h = t - 1/2$ and subtract $4h^{-4}$. Telescoping the factors and applying the divided-difference comparison to the resulting five-factor products gives

$$|(I_a - G_a)[a(t, j)a(t, k)] - 4E_4| \leq 8(j+k)E_5, \\ |(I_a - G_a)[a(j, t)a(k, t)] - 4E_4| \leq (32/3)(j+k)E_5.$$

The first error is also at most $4E_4$, and the second at most $8E_4$. The positive Peano kernel and the derivative ratio give $E_5/E_4 \leq (2p+4)/(4a-2) = \tau$. Use $\min\{u+v, 1\} \leq \min\{u, 1\} + \min\{v, 1\}$ and contract the two bounds. The real row-oriented error is nonnegative, whereas the column-oriented error is at least $-4E_4$; this gives the second lower bound in (46). For (47), all four real row-oriented shifts are at least $1/2$ on the sampled coordinates $\{1\} \cup [2, \infty)$. Comparison with the derivative of t^{-4} gives the factor θ_a . Lemma 3 bounds the other orientation by $4E_4$.

The Peano measure. Set

$$n_a = 2p, \quad m = n_a + 4, \quad h = t - \frac{1}{2}, \quad h_- = a - \frac{1}{2}, \quad h_+ = 2a - \frac{3}{2}.$$

Let $K(t) \geq 0$ be the order- n_a Gaussian Peano kernel, so that $(I_a - G_a)f = \int K(t)f^{(n_a)}(t) dt$. The exact E_4 is positive: $p_a < a$ leaves a counting point outside the Gaussian nodes, where the Hermite remainder for h^{-4} is strictly positive. Normalize $K(t)h^{-m} dt$ to a probability measure ν . The derivatives of the reciprocal powers give

$$M_2 := \int h^2 d\nu = \frac{(m-1)(m-2)}{6} \frac{E_2}{E_4}, \quad M_{-2} := \int h^{-2} d\nu = \frac{20}{m(m+1)} \frac{E_6}{E_4}. \quad (48)$$

These intervals are intersected with the support bounds. In particular,

$$M_{-2} \leq \min \left\{ h_-^{-2}, \frac{h_-^2 + h_+^2 - M_2}{h_-^2 h_+^2} \right\}, \quad (49)$$

by the chord bound for $z \mapsto 1/z$ on $[h_-^2, h_+^2]$. If an enclosed denominator E_4 has zero lower endpoint, the corresponding ratio is replaced by its support bound.

The following reciprocal identity is the Genocchi–Hermite formula [7, §9, (52)]. Its Laplace representation gives the normalization used here. For four nonnegative real shifts $\delta_1, \dots, \delta_4$, let ϑ be uniform on the simplex $\{\vartheta_i \geq 0 : \sum_i \vartheta_i = 1\}$, and put $S_\vartheta = \sum_i \vartheta_i \delta_i$. The Laplace representations of the four reciprocals, followed by the substitution $t_i = r\vartheta_i$, give

$$\prod_{i=1}^4 (h + \delta_i)^{-1} = \mathbb{E}_\vartheta (h + S_\vartheta)^{-4}.$$

The simplex has volume $1/6$, which cancels the radial integral $\Gamma(4) = 6$. Differentiating under the integral yields the normalized derivative

$$R_\delta(h) := \frac{[\prod_i (h + \delta_i)^{-1}]^{(n_a)}}{(h^{-4})^{(n_a)}} = \mathbb{E}_\vartheta (1 + S_\vartheta/h)^{-m}. \quad (50)$$

All integrands are bounded on the positive compact h interval.

The complex orientation. For a source coordinate $j \geq 1$, write

$$f_j(h) = [(h + j - 1)^2 + d_j]^{-1}, \quad q_j(h) = (h + j - 1)^{-2}, \quad d_j = 2j - \frac{1}{4}.$$

Then $a(j, t) = 2f_j(h)$ and

$$f_j f_k - q_j q_k = -d_j q_j f_j f_k - d_k q_k q_k f_k.$$

Each term is a product of six reciprocals whose shifts have nonnegative real part. The derivative comparison therefore gives

$$|(I_a - G_a)[4f_j f_k - 4q_j q_k]| \leq 4(d_j + d_k)E_6 \leq 8(j + k)E_6.$$

Its absolute value is also at most $8E_4$. Discarding the imaginary parts of the reciprocal shifts bounds the absolute derivative of $f_j f_k$ by the positive derivative of $q_j q_k$. Hence the real comparison bounds the original quadrature error from above. If $\gamma_6 \geq E_6/E_4$, contraction gives the following total loss in the half-averaged Gram expression:

$$0 \leq g_a^{\text{real}} - g_a \leq C_a := 8S_x C_x(\gamma_6)E_4. \quad (51)$$

Here g_a^{real} uses $2q_j(h)$ for the column-oriented kernel. An upper bound for M_{-2} from (48)–(49), multiplied by $m(m+1)/20$, supplies γ_6 .

Lower bounds for the real contractions. Let $\bar{\delta} = \frac{1}{4} \sum_i \delta_i$. Convexity in S_ϑ gives $R_\delta(h) \geq (1 + \bar{\delta}/h)^{-m}$. For $\delta \geq 0$,

$$\frac{d^2}{dz^2}(1 + \delta\sqrt{z})^{-m} = \frac{m\delta[1 + (m+2)\delta\sqrt{z}]}{4z^{3/2}(1 + \delta\sqrt{z})^{m+2}} \geq 0.$$

Thus, for any upper bound $c \geq \sqrt{M_-}$,

$$\int R_\delta(h) d\nu \geq (1 + c\bar{\delta})^{-m}.$$

For the row orientation, the shifts from source j are $j \pm \sqrt{8j-7}/2$. For the real column comparison, they are $j-1, j-1$. A pair j, k therefore has $\bar{\delta} = (j+k-2o)/2$, with $o=0$ for the row orientation and $o=1$ for the column comparison. Using $1+u+v \leq (1+u)(1+v)$ separates the two source variables. Define

$$b_o(j) = (1 + c(j-o)/2)^{-m}, \quad \tilde{b}_o(j) = \exp[-mc(j-o)/2].$$

Both choices give a separable lower bound, since $b_o(j) \geq \tilde{b}_o(j)$. Also $1 - b_o(j) \leq \min\{mcj/2, 1\}$ and $1 - \tilde{b}_o(j) \leq \min\{mcj/2, 1\}$. For $v = y, o = 0$ or $v = x, o = 1$, a lower bound for the resulting weighted sum is therefore

$$L_{v,o} = \max \left\{ \mathcal{L}_v(b_o; H), \mathcal{L}_v(\tilde{b}_o; H), S_v - C_v(mc/2), 0 \right\}, \quad H \leq H_0.$$

After contracting the ordered source pairs,

$$g_a \geq 2E_4(L_{y,0}^2 + L_{x,1}^2) - \mathcal{C}_a. \quad (52)$$

A second use of the Peano measure supplies the endpoint-chord bound. The preceding Jensen inequality and $1+u \leq e^u$ give $R_\delta(h) \geq \exp[-m\bar{\delta}/h]$. If $mH \leq 3h_-$, then $z \mapsto \exp[-m(j+k-2o)/(2\sqrt{z})]$ is concave on $[h_-^2, h_+^2]$ for $j, k \leq H$. Put

$$\alpha = \frac{h_+^2 - M_2}{h_+^2 - h_-^2}, \quad Z_{v,o}(h) = \mathcal{L}_v((\exp[-m(j-o)/(2h)])_{j=1}^H; H).$$

The endpoint chord gives a lower bound

$$\alpha Z_{v,o}(h_-)^2 + (1 - \alpha) Z_{v,o}(h_+)^2$$

for the corresponding real contraction divided by $2E_4$. For the complete source sum, $h \geq h_-$ also gives the lower bound

$$[e^{\sigma\gamma} \max\{0, S_v - C_v(\gamma)\}]^2, \quad \gamma = \frac{m}{2h_-}.$$

Take the larger of these two bounds for each orientation, add them, multiply by $2E_4$, and subtract $\mathcal{C}_a$. Both endpoints of an interval for α are tested.

Upper bounds and the Dirichlet variance. Convexity in (50) gives

$$R_\delta(h) \leq \frac{1}{4} \sum_{i=1}^4 \phi_h(\delta_i), \quad \phi_h(t) = (1 + t/h)^{-m}.$$

The second derivative of $z \mapsto (1 + \delta/\sqrt{z})^{-m}$ has the sign of $(m-2)\delta - 3\sqrt{z}$. The function is therefore concave on the Peano support when $(m-2)\delta \leq 3h_-$. Consequently an upper bound for its mean is

$$T(\delta) = \begin{cases} (\sqrt{M_2}/(\sqrt{M_2} + \delta))^m, & (m-2)\delta \leq 3h_-, \\ (h_+/(h_+ + \delta))^m, & \text{otherwise.} \end{cases}$$

In the first case an upper endpoint for M_2 is used. Let $\delta_j^\pm = j \pm \sqrt{8j-7}/2$ and set

$$a_{y,j} = 2 - T(\delta_j^+) - T(\delta_j^-), \quad a_{x,j} = 1 - T(j-1).$$

These coefficients are nonnegative. Contracting the upper bound and retaining the subtracted terms for $j \leq H_0$ gives

$$g_a \leq A_a := E_4 \left[2(S_x^2 + S_y^2) - S_y \mathcal{L}_y(a_y; H_0) - 2S_x \mathcal{L}_x(a_x; H_0) \right]. \quad (53)$$

There is an explicit lower bound for the difference in the convexity inequality. If $0 \leq \delta_i \leq \Delta$, then

$$\phi_h''(t) \geq K(h, \Delta) := \frac{m(m+1)}{h^2} \left(\frac{h}{h+\Delta} \right)^{m+2}, \quad 0 \leq t \leq \Delta.$$

Subtract the quadratic of curvature $K(h, \Delta)$ before applying convexity. Since uniform simplex weights satisfy $\mathbb{E}\vartheta_i = 1/4$, $\mathbb{E}\vartheta_i^2 = 1/10$, and $\mathbb{E}\vartheta_i\vartheta_j = 1/20$ for $i \neq j$, one obtains

$$\frac{1}{4} \sum_i \phi_h(\delta_i) - R_\delta(h) \geq \frac{K(h, \Delta)}{10} \sum_i (\delta_i - \bar{\delta})^2. \quad (54)$$

For singleton source indices $j, k \leq H$, $1 \leq H \leq H_0$, the squared-deviation sum is

$$(j-k)^2 + 4(j+k) - 7 \quad \text{in the row orientation,} \quad (j-k)^2 \quad \text{in the real column comparison.}$$

Both are nonnegative. Take $\Delta_y(H) = H + \sqrt{8H-7}/2$ and $\Delta_x(H) = H - 1$. The function $h^{-2}(h/(h+\Delta))^{m+2}$ has derivative with the sign of $m\Delta - 2h$, so its minimum is attained at a support endpoint. Define

$$k_v(H) = \frac{m(m+1)}{10} \min_{h \in \{h_-, h_+\}} h^{-2} \left(\frac{h}{h + \Delta_v(H)} \right)^{m+2}.$$

Choose fixed coordinate lower bounds $0 \leq v_j^- \leq \max\{0, \hat{v}_j - d_h\}$ and put $s_r = \sum_{j=1}^H j^r v_j^-$, $r = 0, 1, 2$. The two source contractions of the squared deviations are

$$P_y(H) = 2s_0s_2 - 2s_1^2 + 8s_0s_1 - 7s_0^2, \quad P_x(H) = 2s_0s_2 - 2s_1^2.$$

Each polynomial uses the moments of the same fixed lower vector and is evaluated as a whole. It equals a sum with nonnegative pair coefficients, so it bounds the contraction over the first H coordinates from below. For any chosen set of integers H , put $\Gamma_v = \max_H k_v(H)P_v(H)$. Integration of (54) gives

$$g_a \leq A_a - 2E_4(\Gamma_x + \Gamma_y). \quad (55)$$

The final cell upper bound is the minimum of (47) and (55). The cell lower bound is the maximum of the proved lower bounds above. Group the cells into batches. Within each batch, sum the cell intervals; each lower or upper formula also gives a bound when summed over a before interval evaluation. Intersect these bounds and sum the resulting batch intervals. The supplied data contain 21 batches covering all 670 active cells. The default verifier checks this coverage and repeats the intersections and sums; scalar recomputation also evaluates every cell bound from the preceding formulas.

3.2 The two-variable quadrature error and the finite norm

We bound the two-variable error in (37). On Bernstein ellipses of parameter 5 for active cells beginning at $a, b \geq 64$, the three kernel factors have respective bounds

$$\frac{50}{(c_i + a)^2}, \quad \frac{100}{(a + b)^2}, \quad \frac{50}{(c_j + b)^2}.$$

For the middle factor, $s_0 = a + b \geq 128$ gives

$$\Re(z + w - 1) \geq s_0/5 + 3/5, \quad |z - w + 1| \leq 14s_0/5 - 13/5,$$

so its denominator in (1) has modulus at least $s_0^2/50$. Each quadrature-error functional has norm at most twice its cell mass and annihilates degrees below twice its order. Summing the two tails of the tensor Chebyshev series bounds an entry of M , after removing $\sqrt{w_i w_j}$, by

$$\frac{6250000 ab 5^{-2p_a - 2p_b}}{(c_i + a)^2 (a + b)^2 (c_j + b)^2}.$$

Contracting against x, y and using $(a + b)^2 \geq 4ab$ proves

$$|\omega| \leq 1562500 S_x S_y \left(\sum_a a^{-2} 5^{-2p_a} \right)^2 =: \mathcal{M}. \quad (56)$$

The one-variable Chebyshev coefficient bound used here is $|a_n| \leq 2M\rho^{-n}$ for a function bounded by M on the ellipse [15, Theorem 8.1]; applying its contour proof twice gives the tensor bound $4M\rho^{-n-m}$.

Let $[g_-, g_+]$ enclose $g = \sum_a g_a$. The calculation below has $g_- \geq 0$. For $s > 0$ and $g \geq 0$, define

$$f_{\pm}(s, g) = s + \frac{sg \pm \mathcal{M}}{s^2 + g}.$$

If $\mathcal{M} \leq l_B^3/2$, these functions increase in both variables on $[l_B, u_B] \times [g_-, g_+]$. Indeed,

$$\frac{\partial f_{\pm}}{\partial g} = \frac{s^3 \mp \mathcal{M}}{(s^2 + g)^2}, \quad \frac{\partial f_{\pm}}{\partial s} = \frac{s^4 + s^2 g + 2g^2 \mp 2s\mathcal{M}}{(s^2 + g)^2}.$$

Consequently the lifted Rayleigh quotient lies in

$$l_R = f_-(l_B, g_-), \quad u_R = f_+(u_B, g_+).$$

The value $d_- = l_B^2 + g_- > 0$ bounds its squared normalization below. Since $s_N \geq l_R$, equation (2) bounds its second singular value by $\beta_N = \sqrt{\zeta(2) - l_R^2}$. Require $l_R > \beta_N$. Centering the normalized residual at its Rayleigh quotient can only decrease its norm. Equations (6) and (41) give

$$l_R \leq s_N \leq u_R + \frac{\mathcal{R}}{d_-(l_R - \beta_N)}. \quad (57)$$

4 An explicit correction for the infinite tail

Let u, v be the positive unit singular vectors of A_N , with $A_N v = s_N u$, and define

$$s_u = \sum_{k=1}^N u_k, \quad s_v = \sum_{k=1}^N v_k, \quad m_u = \sum_{k=1}^N k u_k, \quad m_v = \sum_{k=1}^N k v_k.$$

Appendix C encloses these four quantities from the sampled trial data. Write $H = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix}$, extend $q = (u, v)/\sqrt{2}$ by zero, and set

$$r = Hq - s_N q, \quad R = \|r\|^2, \quad J = \langle r, Hr \rangle.$$

The residual is supported beyond N in both halves. We regard A_N and its dilation as operators on the infinite spaces by extending them by zero.

Proposition 4. *Suppose $N \geq 8$ and*

$$\epsilon^2 = \frac{2}{N(N+1)}, \quad b = \sqrt{\zeta(2) - s_N^2} + \epsilon.$$

Then, for $\lambda = T$,

$$\lambda - s_N = \frac{R}{\lambda} + \frac{J}{\lambda^2} + E, \quad 0 \leq E \leq \frac{\epsilon^2 R}{s_N^2(s_N - b)}, \quad (58)$$

$$0 \leq \frac{R}{s_N} + \frac{J}{s_N^2} - \frac{R}{\lambda} - \frac{J}{\lambda^2} \leq \frac{R}{s_N - b} \left(\frac{R}{s_N^2} + \frac{2J}{s_N^3} \right). \quad (59)$$

Proof. Every entry outside the N -by- N section has denominator greater than $N(N+1)/2$. Summing reciprocals squared shows $\|H - \mathcal{H}_N\| \leq \|A - A_N\|_{\text{HS}} \leq \epsilon$. Let $Q = I - qq^T$ and $K = QHQ$ on $q^\perp$. The largest spectral value of $Q\mathcal{H}_NQ$ is at most $\sqrt{\zeta(2) - s_N^2}$, whence $K \leq bI$. Since $s_N \geq 1$ and $\epsilon \leq 1/6$,

$$s_N - b \geq 1 - \sqrt{\zeta(2) - 1} - 1/6 > 0.$$

The leading eigenvalue of H exceeds s_N , since $r \neq 0$. Thus $\lambda > s_N > b$, and $\lambda - K$ is positive and invertible. The Schur complement gives

$$\lambda - s_N = \langle r, (\lambda - K)^{-1} r \rangle \leq R/(s_N - b).$$

The identity

$$(\lambda - K)^{-1} = \lambda^{-1}I + \lambda^{-2}K + \lambda^{-2}K(\lambda - K)^{-1}K$$

gives the first two terms in (58) and the remainder $\lambda^{-2}\langle Kr, (\lambda - K)^{-1}Kr \rangle$. Here $\langle r, Kr \rangle = J$, since $r \perp q$. The remainder is nonnegative. Since $\mathcal{H}_N r = 0$, $\|Kr\| \leq \epsilon\|r\|$, which proves its upper bound. Positivity of the kernel gives $J \geq 0$. The mean value theorem applied to $R/t + J/t^2$, together with the Schur-complement upper bound, proves (59). $\square$

It remains to bound R and J accurately enough. For $j > N$, let

$$f_v(j) = \sum_{k \leq N} a(j, k)v_k, \quad f_u(j) = \sum_{k \leq N} a(k, j)u_k.$$

The two nonzero tails of r are $f_v/\sqrt{2}$ and $f_u/\sqrt{2}$, so

$$R = \frac{1}{2} \sum_{j>N} (f_v(j)^2 + f_u(j)^2), \quad J = \sum_{j,k>N} f_v(j)a(j,k)f_u(k).$$

We expand each tail through its j^{-3} terms. The ratio N/j ranges over $(0, 1)$, so the function $h(N/j)$ is retained explicitly; the constants E_u, E_v bound the remaining j^{-4} errors. Put

$$\begin{aligned} c_u &= 2s_v/s_N, \quad c_v = 2s_u/s_N, \quad H_* = 1 + \log N, \\ a_v &= 2s_v - 4m_v, \quad a_u = 6s_u - 4m_u, \quad h(t) = 4\log(1+t) + \frac{2t}{1+t}, \\ E_v &= (70H_* + 6)c_v + \frac{20s_v}{s_N^2}(6H_*^2 + 4H_* + 12), \\ E_u &= (70H_* + 6)c_u + \frac{40s_u}{s_N^2}(6H_*^2 + 4H_* + 12). \end{aligned}$$

The following bounds hold for every $j > N$:

$$f_v(j) = \frac{2s_v}{j^2} + \frac{a_v + c_v h(N/j)}{j^3} + \varepsilon_v(j), \quad |\varepsilon_v(j)| \leq E_v j^{-4}, \quad (60)$$

$$f_u(j) = \frac{2s_u}{j^2} + \frac{a_u + c_u h(N/j)}{j^3} + \varepsilon_u(j), \quad |\varepsilon_u(j)| \leq E_u j^{-4}. \quad (61)$$

We derive the constants from the following entry estimates. For integers $1 \leq k \leq j$,

$$|a(j, k) - 2j^{-2}|, |a(k, j) - 2j^{-2}| \leq 6kj^{-3}.$$

In the first orientation the denominator is $j^2 + (2k-1)j + (k-1)(k-2)$; its excess over j^2 lies between 0 and $3kj$. In the second orientation the same bounds hold for $k \geq 2$. At $k = 1$, its denominator is $j^2 - j + 2 \geq j^2/2$, and direct subtraction gives the stated estimate. For $k > j$, both entries lie between 0 and $2/j^2$, so each difference from $2/j^2$ has magnitude at most $2/j^2$.

For $1 \leq k \leq N < j$, a further division gives

$$\begin{aligned} |a(j, k) - 2j^{-2} + (4k-2)j^{-3}| &\leq 20k^2j^{-4}, \\ |a(k, j) - 2j^{-2} + (4k-6)j^{-3}| &\leq 20k^2j^{-4}, \\ |a(j, k) - 2(j+k)^{-2} - 2j^{-3}| &\leq 70kj^{-4}, \\ |a(k, j) - 2(j+k)^{-2} - 6j^{-3}| &\leq 70kj^{-4}. \end{aligned}$$

For the last pair, write the denominator as $(j+k)^2 - d$, where $d = j + 3k - 2$ or $3j + k - 2$. Since $d \leq 3(j+k)$ and $j+k \geq 9$, the error after $2/(j+k)^2 + 2d/(j+k)^4$ is at most $36/(j+k)^4$. Comparison of the second term with $2/j^3$ or $6/j^3$ adds at most $18k/j^4$ or $30k/j^4$.

Write $H_j = \sum_{k=1}^j k^{-1}$. The singular equations and the first entry bound also show, for $1 \leq j \leq N$,

$$\begin{aligned} |u_j - c_u j^{-2}| &\leq (24H_j + 8)s_u/(s_N^2 j^3), \\ |v_j - c_v j^{-2}| &\leq (12H_j + 4)s_v/(s_N^2 j^3). \end{aligned}$$

To see this, use $u_k \leq 2s_v/(s_N k^2)$ and $v_k \leq 4s_u/(s_N k^2)$. In the equation for u_j , the terms $k \leq j$ contribute at most $24s_u H_j/(s_N^2 j^3)$ to the error, and the terms $k > j$ contribute at most $8s_u/(s_N^2 j^3)$,

since $\sum_{k>j} k^{-2} \leq 1/j$. The equation for v_j gives the other bound. Define $H_n^{(2)} = \sum_{k=1}^n k^{-2}$, $d_v(k) = v_k - c_v k^{-2}$, and $q_v(j, k) = a(j, k) - 2j^{-2} + (4k - 2)j^{-3}$. Then

$$f_v(j) = \frac{2s_v}{j^2} + \frac{a_v}{j^3} + c_v \sum_{k \leq N} \frac{q_v(j, k)}{k^2} + \sum_{k \leq N} d_v(k) q_v(j, k).$$

The last sum has absolute value at most

$$\frac{20s_v}{s_N^2 j^4} \sum_{k \leq N} \frac{12H_k + 4}{k} \leq \frac{20s_v(6H_*^2 + 4H_* + 12)}{s_N^2 j^4},$$

using $\sum_{k \leq N} H_k/k = (H_N^2 + H_N^{(2)})/2$ and $H_N^{(2)} < 2$. For the other sum, the exact identity

$$\sum_{k=1}^N \frac{2}{k^2(j+k)^2} = \frac{2(H_N^{(2)} + H_{j+N}^{(2)} - H_j^{(2)})}{j^2} - \frac{4(H_N + H_j - H_{j+N})}{j^3}$$

and the entry bound with denominator $(j+k)^2$ give

$$\sum_{k \leq N} \frac{q_v(j, k)}{k^2} = \frac{2(H_{j+N}^{(2)} - H_j^{(2)})}{j^2} + \frac{4(H_{j+N} - H_j)}{j^3} + e_v(j), \quad |e_v(j)| \leq 70H_* j^{-4}.$$

Integral comparison replaces $H_{j+N} - H_j$ by $\log(1 + N/j)$ with error at most $1/j$, and $H_{j+N}^{(2)} - H_j^{(2)}$ by $N/[j(j+N)]$ with error at most $1/j^2$. The two substitutions contribute at most $6j^{-4}$ and leave $h(N/j)/j^3$. This proves (60) with the stated E_v . For u , use $d_u(k) = u_k - c_u k^{-2}$ and $q_u(j, k) = a(k, j) - 2j^{-2} + (4k - 6)j^{-3}$. The same calculation replaces $12H_k + 4$ by $24H_k + 8$, giving the factor 40 in E_u and proving (61).

Let $Z_r = \zeta(r, N+1)$ and put

$$I_* = 2 \log 2 - 7/12, \quad h_1 = 4 \log 2 + 1, \quad W = \sum_{j>N} j^{-5} h(N/j).$$

The summand is positive and decreasing. Integration gives $I_* N^{-4} - h_1 N^{-5} \leq W \leq I_* N^{-4}$. Squaring (60)–(61) and averaging gives the enclosure

$$R = 2(s_u^2 + s_v^2)Z_4 + 2(s_v a_v + s_u a_u)Z_5 + 2(s_v c_v + s_u c_u)W + \Delta_R, \quad (62)$$

$$|\Delta_R| \leq \frac{1}{5N^5} \left[2(s_v E_v + s_u E_u) + \frac{(D_v + E_v/(N+1))^2 + (D_u + E_u/(N+1))^2}{2} \right],$$

where $D_v = |a_v| + c_v h_1$, $D_u = |a_u| + c_u h_1$. We used $Z_6 \leq 1/(5N^5)$ for all omitted terms.

Finally, with $J_* = 3/4 - \log 2$,

$$\left| J - \frac{8s_u s_v J_*}{N^4} \right| \leq \frac{J_*}{N^5} \left[56s_u s_v + 48(s_v m_u + s_u m_v) + \frac{144m_u m_v}{N} \right]. \quad (63)$$

Indeed, $\int_1^\infty \int_1^\infty [s^2 t^2 (s+t)^2]^{-1} ds dt = J_*$. Monotone integral comparison changes the corresponding sum by at most $4J_*/N^5$. The entry error $0 \leq a(j, k) - 2/(j+k)^2 \leq 12/(j+k)^3$ and $|f_v(j) - 2s_v/j^2| \leq 6m_v/j^3$ (and its transpose) give (63). Specifically, sum–integral comparison contributes $32s_u s_v J_*/N^5$, and the kernel correction contributes at most $24s_u s_v J_*/N^5$. Using $a(j, k) \leq 4/(j+k)^2$ for $j, k > N$, the two linear residual terms and their product give the terms with coefficients 48

and 144. Intersect the enclosures of R, J with $[0, \infty)$ and substitute them in Proposition 4. For an explicit final enclosure, write $s_N \in [l_N, u_N]$, $R \in [R_-, R_+]$, and $J \in [J_-, J_+]$. Put

$$b_+ = \sqrt{\zeta(2) - l_N^2} + \epsilon, \quad d_0 = l_N - b_+ > 0,$$

$$D_+ = \frac{R_+}{d_0} \left(\frac{R_+}{l_N^2} + \frac{2J_+}{l_N^3} \right), \quad E_+ = \frac{\epsilon^2 R_+}{l_N^2 d_0}.$$

Proposition 4 gives

$$\max \left\{ l_N, l_N + \frac{R_-}{u_N} + \frac{J_-}{u_N^2} - D_+ \right\} \leq T \leq u_N + \frac{R_+}{l_N} + \frac{J_+}{l_N^2} + E_+. \quad (64)$$

The endpoints in Theorem 1 are outward decimal roundings of this enclosure.

5 Choosing the parameters

The discretization is specified by e , r_0 , and the cell orders p_a . The trial data, expansion order, and working precisions determine the finite enclosures used in the estimates. In this section d_{dec} denotes a target error exponent: we seek an interval of width at most $10^{-d_{\text{dec}}}$.

5.1 Balancing the errors

The truncation exponent e controls the infinite tail, while r_0 and the orders p_a control the comparison with the integer section. For example, Appendix B gives the sufficient bound

$$|g| \leq C A_3, \quad C = \frac{C_\rho L_2}{l_0^2}, \quad A_3 = \sum_a a^{-3} \rho^{-2p_a},$$

where $5 \leq \rho < 3 + \sqrt{8}$ and $l_0 > 0$. To make this bound at most ε_q , one may allocate ε_q/ℓ to each active cell. The resulting condition is

$$2p_a \log \rho + 3 \log a \geq \log(C\ell/\varepsilon_q). \quad (65)$$

One integer choice is

$$p_a = \max \left\{ 1, \left\lceil \frac{\log(C\ell/\varepsilon_q) - 3 \log a}{2 \log \rho} \right\rceil \right\}.$$

Use this choice where $p_a < a$; enlarge the exact prefix to absorb earlier cells requiring full summation, and recompute the bound with the resulting value of ℓ . Maintain $r_0 < e$, increasing e if the prefix would absorb every active cell. For the new discretization, evaluate C with a certified lower bound l_0 for its trial Rayleigh value. The orders decrease by approximately

$$\frac{3 \log 2}{2 \log \rho}$$

per doubling of a . This allocation concerns the absolute Gram bound. The signed estimate in Section 3 can give a narrower interval for g . The bounds involving A_0, A_1, A_2 in Appendix C must also be evaluated when choosing the orders.

For fixed t , scaling $[a, 2a - 1]$ to $[-1, 1]$ sends the kernel poles to -3 as a tends to infinity. The Bernstein ellipse through -3 has parameter $3 + \sqrt{8}$. As ρ increases toward $3 + \sqrt{8}$, the factor ρ^{-2p_a}

decreases, while $C_\rho = 16\rho/[(\rho-1)k_\rho^4]$ diverges because k_ρ tends to zero. Choose ρ using the product of these factors.

For the tail estimate, the finite singular vectors have bounded sums and first moments of order $\log N$. The Cauchy–Schwarz argument in Appendix B bounds the sums, and the inverse-square entry bounds in Section 4 then bound the moments. Consequently $R = O(N^{-3})$ and $J = O(N^{-4})$. With exact finite-vector data, the uncomputed terms in (62)–(63) and Proposition 4 have total size

$$O((\log N)^2 N^{-5}). \quad (66)$$

The constants can be chosen uniformly for $N \geq 8$, by the lower bound for $s_N - b$ in the proof of Proposition 4. For error $10^{-d_{\text{dec}}}$, the leading sufficient truncation exponent is therefore

$$e = \frac{d_{\text{dec}}}{5 \log_{10} 2} + O(\log d_{\text{dec}}). \quad (67)$$

The explicit remainders determine the additive allowance. Errors in the transported vector sums and moments contribute separately. Their radii contain factors $N^{1/2}$ and $N^{3/2}$, respectively, but the corresponding coefficients in (62) are of orders N^{-3} and N^{-4} . At the fixed shift ξ , the error caused by replacing C with C_0 is bounded in absolute value by

$$2\delta_{\text{app}}\sqrt{\nu_L\nu_R} + \frac{\delta_{\text{app}}^2(\nu_L + \nu_R)}{\xi - \kappa}.$$

The shift and nonlinear terms e_{shift} and e_{nl} in (32), and the arithmetic enclosures of μ and the small linear system, also contribute to the final interval. If $\xi - \kappa - \delta_{\text{app}}$ stays bounded below by a positive constant, both ν terms are $O(R_B)$. Thus reducing the trial residual or the Laplace approximation error can sharpen the sampled bound. The mesh and truncation indices control δ_{app} through the four explicit estimates in Section 2.2; the finite solve must be enclosed accurately enough that its radius is below the target width.

The far expansions in Appendix D have remainders proportional to $(p_{\text{far}} + 1)6^{-p_{\text{far}}}$, with the stated cell and vector factors. If the logarithms of these factors are $O(d_{\text{dec}})$, an order $p_{\text{far}} = O(d_{\text{dec}})$ suffices for error $10^{-d_{\text{dec}}}$. Working precision must also cover cancellation in polynomial evaluation and in the scalar quadrature errors. The implemented root brackets start with radius $2^{-b_{\text{work}}+2p_a+32}$; the final sign checks and product enclosures determine whether the precision suffices.

5.2 Operation counts and storage

Equation (3) gives the exact sampled dimension. In a family with $e = O(d_{\text{dec}})$, $p_a = O(d_{\text{dec}})$, and $n_0 = O(d_{\text{dec}})$, it is $n = O(d_{\text{dec}}^2)$. When the required orders are $O(d_{\text{dec}})$, a compatible exact prefix has $n_0 = O(d_{\text{dec}})$: choose its endpoint on the corresponding dyadic scale, with $r_0 = \log_2 d_{\text{dec}} + O(1)$. Only the logarithm of the integer-section size enters this count. A compression with m row and column coordinates and a k -column trial matrix requires $O(m^2k + mk^2 + k^3)$ arithmetic operations once its entries are enclosed. The present implementation proposes the trial matrix by a dense singular-value decomposition, with $O(m^3)$ binary64 cost. The compression dimension, the trial rank, and the number of retained trial-vector coordinates are independent parameters.

In the present rule constructor, a recurrence of length p_a is evaluated at each of p_a roots. Each Newton sweep and each full sign check costs $O(p_a^2)$ operations. With $O(\log b_{\text{work}})$ precision stages and a bounded number of sweeps per stage, refinement and sign checks cost $O(p_a^2 \log b_{\text{work}})$ operations at precisions up to b_{work} . The initial guesses are computed with a dense symmetric eigensolver on the tridiagonal Jacobi matrix, costing $O(p_a^3)$ binary64 operations per cell. Weight evaluation costs another $O(p_a^2)$ operations at working precision.

For the products, group the exact prefix into its dyadic intervals and put

$$m_* = \max\{2^{r_0-1}, \max_a p_a, p_{\text{far}}\}.$$

There are e groups. Each interacts directly with at most seven nearby groups; the remaining interactions use accumulated moments. Let $M(m, b)$ denote the cost of multiplying two polynomials of length at most m with b -bit coefficients. Product and remainder trees give the model

$$O(e M(m_*, b_{\text{work}}) \log m_*) \quad (68)$$

for one matrix product, including the two far expansions. The linear moment accumulation is covered by this bound. The polynomial operations used by the verifier are described in the FLINT documentation [17]. For the Lanczos procedure used to generate the saved trial, k_{it} steps use $2k_{\text{it}} + 1$ matrix products at proposal precision. Repeated orthogonalization adds $O(k_{\text{it}}^2 n)$ scalar operations, followed by a projected eigensystem of order k_{it} and $O(k_{\text{it}} n)$ reconstruction work. The public verifier starts from the saved exact trial and uses two products at the final precision. The scalar quadrature stage evaluates the Jacobi continued fraction (39) through degree 5 and uses six Hurwitz-zeta evaluations per cell. For the fixed powers 2, 4, and 6, its series arithmetic costs $O(p_a)$ operations per cell. Through degree $r_{\text{max}} - 1$, ordinary series arithmetic costs $O(p_a r_{\text{max}}^2)$ operations.

A Laplace approximation with $k_L = M + L + 1$ terms requires inner products of the Laplace columns and a solve of order $2k_L$. If F_L denotes the total number of retained entries in these columns, forming their Gram matrices costs at most $O(k_L F_L)$ operations, and the solve costs $O(k_L^3)$. These operations use the precision of the small interval calculation.

A full-precision vector needs about nb_{work} bits for its midpoint mantissas. Polynomial trees use $O(em_* \log m_*)$ coefficients, and retaining k_{it} basis vectors adds $O(k_{\text{it}} n)$ numbers. Exponents, interval radii, object headers, and temporary arrays add to this storage. Forked workers inherit the polynomial data, but allocate their own temporaries and return serialized interval values. Increasing the worker count therefore changes both the elapsed time and the memory requirement.

Under $e, m_*, b_{\text{work}} = O(d_{\text{dec}})$, vector mantissa storage is $O(d_{\text{dec}}^3)$ bits. Under the refinement assumptions above, summing the rule-construction counts gives $O(\sum_a p_a^2 \log b_{\text{work}})$ operations at precisions up to b_{work} and $O(\sum_a p_a^3)$ binary64 operations for the dense node proposals. With $O(d_{\text{dec}})$ cells of order $O(d_{\text{dec}})$, these bounds are $O(d_{\text{dec}}^3 \log b_{\text{work}})$ and $O(d_{\text{dec}}^4)$, respectively. The cost of each multiple-precision operation must also be included.

6 The 870-place calculation

The discretization uses

$$e = 680, \quad r_0 = 10, \quad p_{2^r} = 660 - \left\lfloor \frac{78(r-7)}{100} \right\rfloor, \quad 10 \leq r \leq 679.$$

Thus $N = 2^{680} - 1$, the first 1,023 integers are exact nodes, and there are 670 active cells, with orders from 658 down to 136. Equation (3) gives $n = 267173$. The analytic transport estimates use $\rho = 1161/200$ and $\beta_0 = 1/5$. The compression has 519 coordinates: the first 511 singleton coordinates and eight normalized averages over the dyadic cells starting at 512, 1024, $\dots$, 65536. Its trial matrix has 16 columns. The auxiliary Gaussian rules have order 16. With l equal to the lower endpoint of μ , it gives $\beta < 0.142487952162$. The exact trial matrix and auxiliary rules are in `certificate/compression.json`.

The Laplace approximation uses

$$h = 2/5, \quad M = 40, \quad L = 8, \quad \gamma = 3/20, \quad d = 149/100.$$

There are 49 terms. The strip estimate sums 1,023 individual coordinates before applying the tangent bound from $J = 1024$ onward. The negative-index estimate uses $\varepsilon = 288/1048576$ and $K = 64$. The four operator-error contributions satisfy

$$\begin{aligned} \delta_{\text{strip}} &< 6.908133 \cdot 10^{-9}, & \delta_- &< 8.012231 \cdot 10^{-8}, \\ \delta_+ &< 5.640933 \cdot 10^{-15}, & \delta_{\text{cut}} &< 6.087384 \cdot 10^{-20}. \end{aligned}$$

Their full interval sum gives $\delta_{\text{app}} < 8.70304480975 \cdot 10^{-8}$. The positive vector used in (22) and the column cutoffs are supplied with the resolvent inputs. The residual scale in the small system is $\alpha = 2^{-1426}$, and its exact shift is

$$\xi = \frac{73772538405909641761574433319504103052238386202356115753712293790811912807911}{2^{255}}.$$

The compression satisfies $\kappa < 0.142487952162$; both $\mu - \kappa$ and $\xi - \kappa$ exceed 1.131736200659 . The enclosure of ϕ_0 has radius below $3.158 \cdot 10^{-940}$ and lies in $(1.3564, 1.3565) \cdot 10^{-859}$. The two squared norms and the error allowances satisfy

$$\begin{aligned} \nu_L &< 2.997 \cdot 10^{-870}, & \nu_R &< 1.065 \cdot 10^{-859}, \\ e_1 &< 9.831309867752023 \cdot 10^{-872}, & e_2 &< 7.124568594195790 \cdot 10^{-874}, \\ e_{\text{shift}} &< 8.317 \cdot 10^{-938}, & e_{\text{nl}} &< 2.061 \cdot 10^{-1718}. \end{aligned}$$

In particular, $2e_1 + e_2 < 1.973386542144601 \cdot 10^{-871}$. The full intervals used in (33) are included in `certificate/resolvent.json`. The column cutoff tolerance is $\tau_f = 2^{-64}$. The cutoffs retain 279,474 entries in each orientation, covering 14,037 sampled nodes. The Gram correction uses the first 1,023 coordinates of the normalized full trial pair. All positivity, monotonicity, and spectral-gap conditions in the preceding estimates are checked with outward arithmetic.

The environment pins Python 3.14, python-flint 0.9.0, and NumPy 2.5.1. Floating-point proposals are stored as exact dyadic numbers and then checked with Arb. The final sampled products used 6,656 bits and $p_{\text{far}} = 1330$. The trial products used 5,120 bits and $p_{\text{far}} = 720$. The inner products of the Laplace columns used 256 bits; the retained trial-vector coordinates used 1,280 bits. The cell Gram calculations used 1,024 bits and the final endpoint arithmetic used 8,192 bits.

The sampled calculation took 6,744.80 seconds on a cloud allocation of 16 CPUs and 96 GiB of memory. Peak aggregate proportional set size (PSS) was 71.23 GiB. Projections onto the Laplace columns took 5.06 seconds. The sampled calculation includes trial construction and the final enclosed products.

The numerical data in `certificate/` include the saved trial pair, enclosures derived from the sampled products, inner products of the Laplace columns, compression data, scalar quadrature errors, and cell Gram intervals. The command `uv run -frozen python verify.py` checks file integrity, reconstructs the compression and Laplace approximation bounds, solves the small resolvent system, and evaluates the final scalar bounds. It checks that the cell data cover every active cell exactly once. The `-recompute-scalars` option also reconstructs the scalar Gaussian errors and the cell Gram bounds. With `-full`, the verifier reconstructs the Gaussian rules and recomputes the enclosed large products and projections onto the Laplace columns before applying the scalar estimates. The README gives the commands and recorded resource use. The complete public recomputation path has been tested on a 158-node example, including dense-product checks; a full 267,173-node repeat with the public program has not been run. The verification-status record in `certificate/` specifies the inputs and completed checks for each mode.

Quantity	Bound or enclosure
Sampled squared residual R_B	$< 1.729 \cdot 10^{-859}$
Squared-norm difference g	$(1.3094, 1.3095) \cdot 10^{-814}$
Width of the interval for g	$< 9.084 \cdot 10^{-1005}$
Bound $\mathcal{M}$ for $ \omega $	$< 1.523 \cdot 10^{-1191}$
Bound $\mathcal{R}$ for $\ r_{\text{lift}}\ ^2$	$< 3.573 \cdot 10^{-1424}$
Finite first-moment error, each vector	$< 5.022 \cdot 10^{-123}$
Tail shift $T - s_N$	$(3.8066, 3.8067) \cdot 10^{-614}$
Width of the sampled norm interval	$< 1.974 \cdot 10^{-871}$
Width of the tail-shift interval	$< 2.297 \cdot 10^{-940}$

Table 1: Outward bounds used in the interval calculation. The accompanying records contain the full intervals.

7 Computational limits

The sampled resolvent enclosure accounts for nearly all of the final width. The uncertainty in the Gram correction is below 10^{-1004} , and that of the infinite-tail correction is below 10^{-939} . At the current parameters, a smaller residual or a more accurate Laplace resolvent calculation can therefore improve the result before a larger integer section is needed.

A dense matrix at $n = 267173$ and $b_{\text{work}} = 6656$ would require about 54 TiB for its midpoint mantissas. The product implementation stores polynomial and moment data instead. Table 2 extrapolates n proportionally to d_{dec}^2 and b_{work} proportionally to d_{dec} from the completed calculation.

d_{dec}	Sampled dimension n	Bits b_{work}	Two vectors (GiB)
870	267173	6656	0.414
1000	$3.53 \cdot 10^5$	7651	0.629
3000	$3.18 \cdot 10^6$	22952	17.0
10000	$3.53 \cdot 10^7$	76506	629

Table 2: Storage model $n = 267173\alpha^2$, $b_{\text{work}} = 6656\alpha$, with $\alpha = d_{\text{dec}}/870$. The first row is the completed calculation; the other rows are projections. The last column is $2nb_{\text{work}}/(8 \cdot 2^{30})$ and covers two packed midpoint arrays.

For $d_{\text{dec}} = 10000$, this model gives about 35 million nodes and 629 GiB for two vectors. Polynomial caches, interval metadata, retained basis vectors, and worker temporaries add to this storage. Proportional scaling of the cell count and orders multiplies the quadratic recurrence and weight-evaluation count by about 1,519 and the dense node-proposal count by about 17,455. Higher-precision arithmetic and any increase in trial iterations add further cost.

The recorded peak of 71.23 GiB is about 172 times the 0.414 GiB mantissa storage for two vectors. If this ratio remained constant under the model in Table 2, peak memory at $d_{\text{dec}} = 10000$ would exceed 100 TiB. For comparison, documented high-memory EC2 instances provide up to 32 TiB [18]. A tridiagonal eigensolver can reduce the cost of the initial Gaussian-node estimates. Releasing polynomial trees in groups can reduce peak memory at the expense of recomputation. The present measurements do not determine total runtime or peak memory at that target.

A The exact endpoints

Let D be the integer formed by concatenating the following lines, including the initial digit 1. Spaces and line breaks are separators only.

```

12742 24152 82122 81882 12340 63972 50780 99472 46945 66385 03917 70381 03174 97939
67672 67660 60681 89521 68758 55877 74980 11044 33713 32140 78133 71070 34873 20368
96846 77183 56086 00622 97082 95741 56243 25822 18960 80661 28846 25082 04853 66710
99248 49570 06452 16126 15480 43926 90280 75880 55894 38735 33138 75269 02929 29551
29722 58745 33840 04646 50896 17013 66450 13367 53502 70723 03590 77147 19905 78080
29395 78036 63710 78360 93339 16935 14861 23752 82735 23709 55862 96407 80893 47518
88053 04130 45399 49746 56213 71127 75208 71865 10386 61554 48754 88046 89356 96951
65041 79011 58021 61905 03608 02142 11375 55044 35692 31888 27976 34009 40167 10958
18572 18143 44452 58754 56012 08518 28787 78355 29576 52651 05445 95100 14676 32712
55125 92183 16629 41059 89696 72955 69635 86271 97863 72840 68628 03050 95549 80387
26138 26369 41880 30243 79468 52654 21444 23311 71104 28331 86831 67027 10239 42026
30682 90396 63271 04535 16611 33264 22961 90097 47717 15556 16199 29408 30239 46005
38614 55091 15646 07125 85227 52899 2

```

Then D has 871 digits, and the exact endpoints are

$$L = \frac{10^{12}D + 235132526703}{10^{882}}, \quad U = \frac{10^{12}D + 432471180919}{10^{882}}.$$

Both endpoints lie strictly between $D/10^{870}$ and $(D+1)/10^{870}$. The files `endpoints.json` and `digits.txt` in `certificate/` contain the same endpoints and shared prefix.

B Analytic bounds on the quadrature error

The coarse analytic bounds in this appendix are used for the sampled Frobenius norm and for transporting vector sums. Set

$$5 \leq \rho < 3 + \sqrt{8}, \quad R_\rho = (\rho + \rho^{-1})/2, \quad k_\rho = (3 - R_\rho)/2 > 0, \quad A_r = \sum_{\text{active } a} a^{-r} \rho^{-2p_a},$$

and define

$$C_\rho = \frac{16\rho}{(\rho-1)k_\rho^4}, \quad J_\rho = \frac{8\rho}{(\rho-1)k_\rho^2}, \quad L_2 = 103/3. \quad (69)$$

On the filled Bernstein ellipse of $[a, 2a-1]$, $\Re z \geq k_\rho a + (R_\rho - 1)/2$. The factors used in Lemma 3 satisfy $\Re(z + \alpha_\pm) \geq k_\rho(t+a)$, $t \geq 1$. For the complex shifts this follows by setting $t = 1$, where the remaining difference is $(2R_\rho - 5)/2 > 0$. For the real shifts it follows from

$$(4/5)t + 3/10 \geq \sqrt{8t-7}/2;$$

the difference of the squared sides is $(16t^2 - 38t + 46)/25 > 0$. Thus either kernel is bounded on the ellipse by $2/[k_\rho^2(t+a)^2]$.

Chebyshev approximation through degree $2p_a - 1$, Gaussian exactness, and positivity of the weights give

$$|(I_a - G_a)a(t, \cdot)| \leq J_\rho a \rho^{-2p_a} / (t+a)^2, \\ |(I_a - G_a)[a(t, \cdot)a(u, \cdot)]| \leq C_\rho a \rho^{-2p_a} / [(t+a)^2(u+a)^2].$$

The same formulas hold for transposed factors. Gaussian domination of $(t+a)^{-4}$ and Minkowski's inequality imply that each of

$$A_N V - U B^T, \quad A_N^T U - V B, \quad U^T U - B^T B, \quad V^T V - B B^T$$

has norm at most $\delta = C_\rho A_2/3$. Telescoping the two quadratures in $a(s, t)^2$ also gives

$$\|B\|_{\text{HS}}^2 \leq \zeta(2) + 2\delta =: F. \quad (70)$$

To bound sums of singular-vector components, note that $a(s, t), a(t, s) \leq 2/(s+t-2)^2$ for $s+t > 2$. For $t \geq 2$, summing either orientation in s against either measure gives at most $2/(t-1)^2 + 2/(t-1)$. At $t=1$, the orientation $a(s, 1) = 2/[s(s+1)]$ is completely monotone and its integer sum telescopes to 2, so its Gaussian sum is at most 2. For $a(1, s)$, the first two singleton terms sum to $3/2$; for $s \geq 3$, use $a(1, s) \leq 2/(s-1)^2$. Gaussian domination and integral comparison bound the remaining sum by $2 \sum_{k \geq 2} k^{-2} \leq 3/2$. Thus either sum at $t=1$ is at most 3. Squaring and summing gives the bound

$$9 + 4\zeta(4) + 8\zeta(3) + 4\zeta(2) < 103/3 = L_2.$$

The same bound holds for Gaussian measure, by domination of the positive reciprocal powers. Cauchy–Schwarz now gives

$$S_x, S_y \leq \sqrt{L_2}/s, \quad \|Uy\|_1, \|Vx\|_1 \leq \sqrt{L_2}. \quad (71)$$

In particular,

$$|d - s^2| \leq C_\rho L_2 A_3 / s^2. \quad (72)$$

This follows by contracting the product error with the positive vectors, bounding each weighted source sum divided by $(t+a)^2$ by $\sqrt{L_2}/(sa^2)$.

C Bounds for sums and first moments

This appendix supplies the finite vector data needed in Section 4. Use the lower endpoint l_0 for the trial Rayleigh quotient from Section 2, and let l_N be the lower bound in (57). Put

$$G = C_\rho L_2 A_3 / l_0^2, \quad D = l_0^2 - G, \quad \delta = C_\rho A_2 / 3, \quad \beta_N = \sqrt{\zeta(2) - l_N^2}.$$

The following bounds require $D > 0$ and $l_0 > \beta_N$. Equation (8) applied to the lifted vector, whose residual at s is bounded by $\delta/\sqrt{D}$, shows that each of its halves differs from the corresponding finite singular vector by at most $2\delta/[\sqrt{D}(l_0 - \beta_N)]$. The compared vectors are $Uy/\sqrt{d}$ and u , and $Vx/\sqrt{d}$ and v .

The weighted trial sums enclose s_u, s_v after adding a symmetric error of radius

$$\frac{2\sqrt{NR_B}}{l_0 - \beta_0} + \frac{2\sqrt{N}\delta}{\sqrt{D}(l_0 - \beta_N)} + \frac{J_\rho \sqrt{L_2} A_1}{l_0 \sqrt{D}} + \frac{\sqrt{L_2} G}{l_0 \sqrt{D}(l_0 + \sqrt{D})}. \quad (73)$$

The first two terms are the trial and lifted spectral-projection errors. For the third, replace the integer sum of Uy or Vx by Gaussian summation and apply the single-kernel error in Appendix B. The last term bounds normalization from $\sqrt{d}$ to s , using (72) and (71).

For the first moments let $F_N = [N(N+1)(2N+1)/6]^{1/2}$, the norm of the integer first-moment functional. Its sampled counterpart has no larger norm: Gaussian quadrature integrates t^2 exactly

at order at least 2, and the order-1 mean rule underestimates it. The weighted trial first moments enclose m_u, m_v after adding radius

$$\frac{2F_N\sqrt{R_B}}{l_0 - \beta_0} + \frac{2F_N\delta}{\sqrt{D}(l_0 - \beta_N)} + \frac{3J_\rho\sqrt{L_2}A_0}{l_0\sqrt{D}} + \frac{4\sqrt{L_2}(1 + \log N)G}{l_0^2\sqrt{D}(l_0 + \sqrt{D})}. \quad (74)$$

For the third term, $|z| \leq 3a$ on each ellipse, so multiplying the kernel by z multiplies its bound by at most $3a$. For the last, the pointwise inverse-square bounds and Gaussian domination of $1/t$ give a sampled first-moment bound $4\sqrt{L_2}(1 + \log N)/l_0^2$ for either vector.

D Enclosed matrix products

Let $v \in \mathbb{R}^n$ be an input vector for a product with B or B^T . Group all nodes with $b \leq c_i < 2b$ at scale b , a power of two. This includes nodes kept as exact singletons. Put $t_i = c_i/b$ and $r_i = \sqrt{w_i}v_i$. The Cauchy sum has the exact representation

$$F(z) = \sum_i \frac{r_i}{z + t_i} = \frac{\sum_i r_i \prod_{j \neq i} (z + t_j)}{\prod_i (z + t_i)}.$$

A product tree constructs these polynomials. Real or complex interval polynomial evaluation, followed by the kernel factorization in Lemma 3, encloses all interactions from source scales $a/8, \dots, 8a$ for targets $a \leq x < 2a$.

The remaining sources satisfy $c_i < a/8$ or $c_i \geq 16a$. Write the two shifts in the kernel factorization as $\alpha_\pm(x)$. Use the generating series

$$\frac{1}{(1 + \xi z)(1 + \eta z)} = \sum_{k=0}^{\infty} (-z)^k \sum_{m=0}^k \xi^m \eta^{k-m}.$$

For sources below $a/8$, take $z = c_i$, $\xi = \alpha_+(x)^{-1}$, $\eta = \alpha_-(x)^{-1}$, and prefactor $2/[\alpha_+(x)\alpha_-(x)]$. These sources satisfy $c_i \leq a/8 - 1$, and the region is nonempty only for $a \geq 16$. Both shift moduli are at least $a/2$, and also at least $3a/4 - 6$. For the smaller real shift, the latter inequality reduces, after squaring positive sides, to

$$(a + 22)^2 - 4(8a - 7) = a^2 + 12a + 512 \geq 0.$$

For the complex shifts, the modulus is $\sqrt{x^2 - x + 2} \geq x - 1/2$. It follows that $|c_i/\alpha_\pm(x)| \leq 1/6$.

For sources at least $16a$, take $z = c_i^{-1}$, $\xi = \alpha_+(x)$, $\eta = \alpha_-(x)$, and prefactor $2c_i^{-2}$. All targets satisfy $x \leq 2a - 1$. The complex shift moduli are at most $2a$, and the real shift moduli are at most $8a/3$; for the larger real shift the last assertion follows from

$$(4a/3 + 3)^2 - (16a - 15) = \frac{16}{9}(a - 9/4)^2 + 15 > 0.$$

Thus $|\alpha_\pm(x)/c_i| \leq 1/6$ in this region as well.

Retain total degrees $0, \dots, p_{\text{far}} - 1$. In either expansion the term of degree k is bounded by $(k + 1)6^{-k}$, before the prefactor. The identity

$$\sum_{k \geq P} (k + 1)q^k = \frac{q^P[(P + 1) - Pq]}{(1 - q)^2}$$

therefore gives the following remainders before multiplying by the target's $\sqrt{w}$:

$$\frac{128(p_{\text{far}} + 1)6^{-p_{\text{far}}}}{9a^2} \sum_{c_i < a/8} |r_i|, \quad \frac{32(p_{\text{far}} + 1)6^{-p_{\text{far}}}}{9} \sum_{c_i \geq 16a} \frac{|r_i|}{c_i^2}.$$

Indeed, summing the tail at $q = 1/6$ gives the smaller prefactors $288/(25a^2)$ and $72/25$. Absolute weighted sums are accumulated separately from signed moments. For polynomial evaluation, the verifier encloses the arithmetic at exact coefficient and argument midpoints, then adds the variation over the input intervals. Let m_k and z_0 be the coefficient and argument midpoints, with $|c_k - m_k| \leq e_k$ and $|z - z_0| \leq r_z$. Choose R to bound the modulus of every point in the argument interval or rectangle. Then

$$\left| \sum_k c_k z^k - \sum_k m_k z_0^k \right| \leq \sum_k e_k R^k + r_z \sum_{k \geq 1} k |m_k| R^{k-1}.$$

The first sum bounds coefficient variation. Integrating the derivative of $\sum_k m_k z^k$ along the segment from z_0 to z proves the second bound. Nonnegative Horner evaluation encloses both sums; near-field evaluations use a separate R for each argument. Root evaluations, moment sums, midpoint arithmetic, and all added radii use outward interval arithmetic. The product expansion order is at least every source polynomial degree, as required by the cached remainder-tree evaluator. The three source regions partition every product exactly once.

References

- [1] L. N. Trefethen, A hundred-dollar, hundred-digit challenge, *SIAM News* 35(1), 2002. [Original problem statement](#).
- [2] F. Bornemann, D. Laurie, S. Wagon, and J. Waldvogel, *The SIAM 100-Digit Challenge: A Study in High-Accuracy Numerical Computing*, SIAM, 2004. Chapter 3, “How Far Away Is Infinity?”, by J. Waldvogel, pp. 47–76. [doi:10.1137/1.9780898717969.ch3](#).
- [3] G. W. Stewart, A second order perturbation expansion for small singular values, *Linear Algebra and its Applications* 56 (1984), 231–235. [doi:10.1016/0024-3795\(84\)90128-9](#).
- [4] L. N. Trefethen, Ten digit problems, in D. Schleicher and M. Lackmann, eds., *An Invitation to Mathematics: From Competitions to Research*, Springer, 2011. [Author’s manuscript](#).
- [5] NIST Digital Library of Mathematical Functions, §3.3 (interpolation), §3.5 (quadrature), §18.19 (Hahn polynomials), and §18.22 (recurrences). <https://dlmf.nist.gov/>. Accessed September 8, 2026.
- [6] G. H. Golub and J. H. Welsch, Calculation of Gauss quadrature rules, *Mathematics of Computation* 23 (1969), 221–230. [doi:10.1090/S0025-5718-69-99647-1](#).
- [7] C. de Boor, Divided differences, *Surveys in Approximation Theory* 1 (2005), 46–69. <https://surveys-in-approximation-theory.com/papers/2/2.pdf>.
- [8] I. H. Sloan, Iterated Galerkin method for eigenvalue problems, *SIAM Journal on Numerical Analysis* 13 (1976), 753–760. [doi:10.1137/0713061](#).

- [9] C. Allouch and M. Tahrichi, Discrete superconvergent Nyström method for integral equations and eigenvalue problems, *Mathematics and Computers in Simulation* 118 (2015), 17–29. [doi:10.1016/j.matcom.2014.11.010](https://doi.org/10.1016/j.matcom.2014.11.010).
- [10] T. Kato, On the upper and lower bounds of eigenvalues, *Journal of the Physical Society of Japan* 4 (1949), 334–339. [doi:10.1143/JPSJ.4.334](https://doi.org/10.1143/JPSJ.4.334).
- [11] A. Bonito, W. Lei, and J. E. Pasciak, On sinc quadrature approximations of fractional powers of regularly accretive operators, *Journal of Numerical Mathematics* 27 (2019), 57–68. [doi:10.1515/jnma-2017-0116](https://doi.org/10.1515/jnma-2017-0116).
- [12] H. Wielandt, Unzerlegbare, nicht negative Matrizen, *Mathematische Zeitschrift* 52 (1950), 642–648. [doi:10.1007/BF02230720](https://doi.org/10.1007/BF02230720).
- [13] W. W. Hager, Updating the inverse of a matrix, *SIAM Review* 31 (1989), 221–239. [doi:10.1137/1031049](https://doi.org/10.1137/1031049).
- [14] L. N. Trefethen and J. A. C. Weideman, The exponentially convergent trapezoidal rule, *SIAM Review* 56 (2014), 385–458. [doi:10.1137/130932132](https://doi.org/10.1137/130932132).
- [15] L. N. Trefethen, *Approximation Theory and Approximation Practice*, SIAM, 2013, Chapter 8. <https://people.maths.ox.ac.uk/trefethen/ATAP/>.
- [16] F. Johansson, Arb: efficient arbitrary-precision midpoint-radius interval arithmetic, *IEEE Transactions on Computers* 66 (2017), 1281–1292. [doi:10.1109/TC.2017.2690633](https://doi.org/10.1109/TC.2017.2690633).
- [17] FLINT developers, *FLINT documentation*, sections on polynomial product trees and multi-point evaluation. https://flintlib.org/doc/arb_poly.html. Accessed September 8, 2026.
- [18] Amazon Web Services, *Specifications for Amazon EC2 memory optimized instances*, U7i and U7in families. <https://docs.aws.amazon.com/ec2/latest/instancetypes/mo.html>. Accessed September 10, 2026.