A Deterministic Algorithm for Eigenvalue Separation in Tridiagonal Matrices

Wenqi Zhu* Ping Nie†

November 2026

Abstract

We develop a deterministic algorithm for separating the eigenvalues of Hermitian tridiagonal matrices by small diagonal perturbations. For $\|T\|_2 \le 1$, $0 < \delta < 1/2$, and $X = \mathrm{diag}(1, \dots, n)$, our algorithm selects $t \in [0, \delta/n]$ such that $D = tX$ satisfies

$$\|D\|_2 \le \delta, \qquad \mathrm{gap}(T + D) > c\delta^2 n^{-12}$$

where $c > 0$ is a constant. The construction uses a classical convex weighted eigenvalue sum and its finite-scale slope variation. A tridiagonal localization argument turns near-degenerate eigenvalues into separated responses to the fixed diagonal direction, yielding a quantitative gap certificate. An adaptive interval selector then identifies a certifying window in at most $\lceil \log_2(16n^3) \rceil$ refinements. Coupled with a suitable spectral-function oracle, this gives a nearly linear exact-arithmetic complexity bound. A structure-exploiting implementation uses rigorous Sturm eigenvalue enclosures and adaptive precision to certify the stopping condition on six families of test matrices. The directly certified gaps are substantially larger than the worst-case analytical bounds, illustrating both the effectiveness of the construction and the conservatism of the universal certificate. Our results address the nearly linear deterministic Minami perturbation question for tridiagonal matrices.

1 Introduction

Spectral separation plays an important role in numerical linear algebra. When eigenvalues are tightly clustered, computing their individual eigenvectors and associated invariant subspaces can become sensitive to perturbations. Spectral gaps also influence the accuracy, convergence, and deflation criteria of many eigenvalue algorithms [23, 18, 12]. A natural approach is to perturb the matrix slightly so that its eigenvalues become better separated. The challenge is to construct such a perturbation efficiently, without first computing an expensive eigendecomposition.

Diagonal perturbations are particularly attractive for tridiagonal matrices because they preserve the matrix structure and the cost of matrix-vector products. In random spectral theory, Minami-type estimates provide probabilistic bounds on the occurrence of closely clustered eigenvalues under suitable independent diagonal perturbations [17, 5]. Whether comparable separation can be achieved deterministically is a more challenging question. It was posed as Problem 3.2, *Deterministic Minami Perturbation*, in the report of the Simons Institute workshop *Linear Systems and Eigenvalue Problems*, held in Berkeley in October 2025 [1]. The KE-02 entry in the Open Problems in NLA catalogue further asks for a nearly linear deterministic construction for arbitrary normalized Hermitian tridiagonal matrices [20].

*Corresponding author. Mathematical Institute, University of Oxford, Andrew Wiles Building, Woodstock Road, Oxford OX2 6GG, United Kingdom. wenqi.zhu@maths.ox.ac.uk

†David R. Cheriton School of Computer Science, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada. ping.nie@uwaterloo.ca

Let $n \geq 2$, $0 < \delta < 1/2$, and let $T \in \mathbb{C}^{n\times n}$ be Hermitian tridiagonal with $\|T\|_2 \leq 1$. Thus

$$T = \begin{pmatrix} d_1 & \beta_1 & & & \\ \overline{\beta_1} & d_2 & \beta_2 & & \\ & \overline{\beta_2} & d_3 & \ddots & \\ & & \ddots & \ddots & \beta_{n-1} \\ & & & \overline{\beta_{n-1}} & d_n \end{pmatrix}, \qquad d_i \in \mathbb{R}. \tag{1}$$

We write its eigenvalues in nondecreasing order and define its minimum eigenvalue gap by

$$\operatorname{gap}(T) := \min_{1\leq j<n} \big(\lambda_{j+1}(T) - \lambda_j(T)\big). \tag{2}$$

In particular, $\operatorname{gap}(T) = 0$ if T has a repeated eigenvalue. The objective is to construct a real diagonal matrix D such that

$$\|D\|_2 \leq \delta, \qquad \operatorname{gap}(T + D) \geq c_0(\delta/n)^p, \tag{3}$$

for absolute constants $c_0, p > 0$, at a total cost of $O\big(n[1+\log(n/\delta)]^q\big)$ exact arithmetic operations for a constant q. This cost concerns the construction of D, including any internal spectral evaluations; it does not require outputting an eigenvector matrix. Arithmetic complexity and finite-precision complexity are separate questions.

Several related approaches provide insight into this problem. Classical work by Ye [27] and Parlett [23] explains how closely spaced eigenvalues of tridiagonal matrices are associated with the localization of their invariant subspaces. Nakatsukasa [18] uses bounds on eigenvector components to study the sensitivity of eigenvalues to structured perturbations. More recently, deterministic constructions have been obtained for special classes, including weakly coupled tridiagonal matrices and matrices with constant diagonal and equal off-diagonal magnitudes [20]. These results provide important structural insights, but do not establish the requested nearly linear construction for the unrestricted tridiagonal case.

In this paper, we give a deterministic construction for the full Hermitian tridiagonal class. Our approach takes a different perspective: rather than estimating individual eigenvalue movements directly, we use the convexity of a weighted spectral function to detect and avoid regions of strong spectral sensitivity. In particular, we restrict the perturbation to a single predetermined diagonal direction:

$$\boxed{X = \operatorname{diag}(1, \ldots, n), \qquad H(t) = T + tX, \qquad 0 \leq t \leq \delta/n.} \tag{4}$$

Since $\|tX\|_2 = nt \leq \delta$, every parameter in this interval respects the perturbation budget. The matrix search therefore reduces to selecting one scalar t. The remaining issue is to certify spectral separation at the selected parameter without first computing the minimum gap accurately.

Main results and contributions. This paper delivers two main results.

First, we develop a convex spectral-gap certificate. We consider a weighted sum of the ordered eigenvalues of $H(t)$, viewed as a function $F(t)$ of the perturbation parameter. This spectral function is convex, strictly increasing, and has bounded slopes with respect to the perturbation size t. Up to an affine trace term, it measures the total pairwise spectral distance after perturbation. By the Hellmann–Feynman formula, its first derivative is a weighted sum of the responses of individual eigenvalues to the diagonal perturbation. Moreover, classical second-order perturbation formulas show that the curvature contains pairwise contributions of the form

$$\underbrace{\text{curvature contribution}}_{\text{for a pair of eigenvalues}} \propto \frac{|\text{perturbation coupling}|^2}{\text{eigenvalue gap}}.$$

Thus closely spaced eigenvalues can produce a strong change in spectral slope, provided that their responses to the perturbation are sufficiently different.

The difficulty is that a small gap alone does not guarantee large curvature: the corresponding coupling may be weak or vanish. To overcome this, we exploit the tridiagonal structure. A discrete Wronskian identity

and a small-leakage cut convert a nearly degenerate eigenvalue pair into two spatially separated approximate eigenvectors. Their responses to the fixed diagonal direction X are quantitatively different. Using a nearby auxiliary matrix and the variational representation of the spectral function, we turn this response separation into a finite-scale gap certificate.

This proof combines ideas from classical eigenvector localization and structured perturbation theory [27, 23, 18] with convex spectral analysis [22, 13, 16]. The key is to use these classical tools for a different purpose: instead of bounding the sensitivity of a given eigenvalue or optimizing a spectral objective, we certify a lower bound on the minimum spectral gap.

Second, we construct an adaptive deterministic interval-selection algorithm. Starting from the admissible parameter interval $[0, \delta/n]$, the algorithm repeatedly selects one of two shorter windows, each retaining slightly less than half of the current interval. Convexity allows us to compare their endpoint slope variations and choose a window with a smaller spectral response. At every iteration, we directly test whether the centered slope variation is sufficiently small to certify spectral separation. Once this condition is satisfied, the algorithm terminates.

Each refinement requires only a constant number of evaluations of the spectral function. In particular, the selection rule does not require prior knowledge of the eigenvectors or a full eigendecomposition at each step. We prove that the algorithm terminates after at most

$$K = \lceil \log_2(16n^3) \rceil = O(\log n)$$

refinements. The returned perturbation satisfies

$$\boxed{\|tX\|_2 \le \delta, \qquad \operatorname{gap}(T + tX) \ge c\,\frac{\delta^2}{n^{12}},} \tag{5}$$

where $c > 0$ is a constant. Including the cost of the spectral-function oracle, the construction requires $O(n[1 + \log(n/\delta)]^q)$ exact arithmetic operations for a constant q. Thus the algorithm meets the nearly linear arithmetic-complexity target of KE-02 for general Hermitian tridiagonal matrices. The two contributions are complementary: the first provides a local certificate relating spectral response to eigenvalue separation, while the second uses convexity to find a certifying perturbation parameter efficiently.

Numerical validation. We implement the adaptive algorithm using double-precision window selection, rigorous Sturm-based eigenvalue enclosures, and adaptive precision. Experiments on six families of tridiagonal matrices show that the certified stopping condition is often satisfied well before the theoretical iteration bound. The implementation has also been tested at dimensions of order 10^3, while the consistently audited strict-Sturm benchmarks are reported separately in Section 5. For several random and nearly reducible matrix families, the observed minimum gaps decrease approximately as n^{-2} over the tested dimension ranges, much more slowly than the universal n^{-12} guarantee. Directly certified eigenvalue gaps are also many orders of magnitude larger than the analytical lower bounds. These results demonstrate the practical effectiveness of the construction on the tested instances, without implying an improved worst-case exponent.

Related work: localization and structured perturbation. Classical work on close eigenvalues of tridiagonal matrices provides an important foundation for our geometric argument. Ye [27] relates closely spaced eigenvalues to complementary principal submatrices, even when the off-diagonal entry connecting the two blocks is not small. Parlett [23] studies the localization of invariant subspaces associated with tightly clustered eigenvalues, while Vömel and Parlett [26] investigate how such localization can be detected.

These results show that small spectral gaps are closely connected to the spatial structure of the corresponding eigenspaces. Our localization argument follows this viewpoint, but requires a specific quantitative conclusion: two approximate eigenvectors supported on opposite sides of a coordinate cut, with small residuals and separated responses to the predetermined direction X. Nakatsukasa [18] develops a complementary perturbation approach for Hermitian block tridiagonal matrices. By bounding eigenvector components through the block recurrence,

he obtains upper bounds on the movement of individual eigenvalues under structured perturbations, with applications to Wilkinson matrices and aggressive early deflation. In contrast, we use the localization of a near-degenerate pair to establish a lower bound on its collective spectral response. This response is then converted into a minimum-gap certificate through convexity. Thus our work connects the geometric ideas of tridiagonal localization to a different spectral-certification problem.

Related work: convex eigenvalue optimization. Weighted eigenvalue sums and diagonal matrix corrections have a long history in spectral optimization, including graph partitioning and nondifferentiable optimization [7, 6, 15]. Ky Fan's variational principle and subsequent work of Overton and Womersley, Lewis, and Lewis–Sendov establish the convexity, subdifferential, and second-order calculus of spectral functions [9, 21, 22, 13, 14, 16]. These classical results are key to our spectral function. In eigenvalue optimization, such functions have typically been used to characterize optimality, derive subgradients, or construct optimization algorithms. Here we use the same variational structure in a different direction. In particular, our proof uses the existence of distinct supporting spectral slopes at a nearby matrix with a repeated eigenvalue. The difference between these slopes is controlled by the responses of two localized eigenvectors to X. We then transfer this information back to the original matrix using stability of the convex spectral function. This is the central connection between structured eigenvalue perturbation and nonsmooth convex analysis in our proof.

Rather than minimizing the spectral function, we search for a window on which its finite-scale slope variation is small. Such a window certifies that the spectrum is sufficiently separated. The interval-selection mechanism is also related to adaptive sampling of convex functions. Simchowitz et al. [24], for example, use midpoint secant errors to guide the approximation of a convex function. Their midpoint error is directly related to our centered slope variation.

Notation. We use $\|\cdot\|_2$ for the Euclidean vector norm and the matrix operator norm, $\|\cdot\|_{\mathrm{F}}$ for the Frobenius norm, and $\|\cdot\|_*$ for the nuclear norm, the sum of the singular values. The vector e_i is the ith standard unit vector. At every permitted parameter, $\|H(t)\|_2 \le 1+\delta < 2$, and its off-diagonal entries have absolute value at most one.

Remark 1.1 (Reduction to real symmetric matrices)**.** *A diagonal unitary matrix S can be chosen so that S^*TS is real symmetric. Choose its phases successively along each chain of nonzero off-diagonal entries; zero entries separate the chains. Since $S^*XS = X$, the entire family $T+tX$ is preserved up to unitary similarity. Thus the eigenvalues, gaps, and norms are unchanged, and the real symmetric arguments below also apply to Hermitian tridiagonal matrices.*

The remainder of the paper is organized as follows. Section 2 develops the convex spectral function and its finite-scale slope variation. Section 3 uses tridiagonal localization to establish the local gap certificate. Section 4 develops a basic and an improved adaptive window selector and analyzes spectral-function evaluations. Section 5 presents rigorously certified numerical experiments. Section 6 summarizes the results and discusses possible extensions.

2 The convex spectral function and its slope variation

Throughout this section, the results apply to any Hermitian matrix, not only to tridiagonal matrices. We use the notation T for simplicity.

The spectral function. For the fixed family $H(t) = T + tX$, define the spectral function

$$\boxed{F(t) := \sum_{j=1}^{n} j\lambda_j(H(t)).} \tag{6}$$

The map $A \mapsto \sum_{j=1}^{n} j\lambda_j(A)$ is a convex spectral function, meaning that it is convex in a Hermitian matrix A and depends only on its eigenvalues. Its basic properties follow from Ky Fan's maximum principle [9], which provides a variational characterization of sums of the largest eigenvalues. Overton and Womersley [21] further studied these eigenvalue sums and their convex-analytic properties, while Lewis [13] developed a general convex-analytic framework for spectral functions on Hermitian matrices.

These variational characterizations have played an important role in eigenvalue optimization. In particular, Overton and Womersley [22] developed optimality conditions and duality theory for minimizing sums of the largest eigenvalues. Closely related weighted trace objectives also appear in Brockett's work on matrix gradient flows, eigenvalue ordering, and diagonalization [2]. The weights $1, \dots, n$ in our spectral function correspond to such a weighted trace objective through (9).

While these spectral functions and their convexity properties are classical, our interest is different: we study their variation along a fixed diagonal perturbation and use this variation to certify quantitative eigenvalue separation.

Why this spectral function? A useful interpretation follows by collecting coefficients:

$$\boxed{F(t) = \frac{n+1}{2} \operatorname{tr} H(t) + \frac{1}{2} \sum_{i<j} \big(\lambda_j(t) - \lambda_i(t)\big).} \tag{7}$$

Thus F measures the total pairwise spectral distance, up to an affine trace term. Example 2.1 illustrates its decomposition into sums of the largest eigenvalues and shows why we instead use its change under perturbation to control the smallest gap.

Example 2.1 (Tail sums and spectral spread)**.** *When $n = 3$, the tail-sum decomposition in (9) is*

$$F(t) = \lambda_1(t) + 2\lambda_2(t) + 3\lambda_3(t) = \underbrace{\lambda_1(t) + \lambda_2(t) + \lambda_3(t)}_{\operatorname{tr} H(t)} + \underbrace{\lambda_2(t) + \lambda_3(t)}_{\textit{two largest}} + \underbrace{\lambda_3(t)}_{\textit{largest}}.$$

The significance of F lies in its variation under the perturbation tX. As t varies, the eigenvalues move at different rates, and their pairwise gaps change accordingly. When two eigenvalues approach one another, their relative motion may change rapidly, producing curvature in F. This motivates studying the second difference of F, which removes the affine trace contribution and measures the variation of the spectral spread over a finite perturbation window. We will use this variation to derive a certificate for the minimum eigenvalue gap.

The following lemma collects the technical properties of the spectral function needed for our analysis. Note that the variational principle and convexity in Lemma 2.1 are classical [9, 21, 13]. We give the proof for completeness.

Lemma 2.1 (Properties of the spectral function)**.** *For every Hermitian T, the function F is **strictly increasing** and **convex**. For $u < v$, its secant slopes satisfy*

$$\boxed{\frac{n(n+1)}{2} \le \frac{F(v) - F(u)}{v - u} \le \frac{n^2(n+1)}{2}.} \tag{8}$$

Moreover, it admits the variational representation

$$F(t) = \max_{U^*U = I} \operatorname{tr}\big(UXU^*(T + tX)\big) = \operatorname{tr}(T + tX) + \sum_{j=2}^{n} \sum_{\ell=j}^{n} \lambda_\ell(t). \tag{9}$$

Proof. Write $\lambda_\ell(t) = \lambda_\ell(T + tX)$. Counting the coefficient of each eigenvalue gives $F(t) = \operatorname{tr}(T + tX) + \sum_{j=2}^{n} \sum_{\ell=j}^{n} \lambda_\ell(t)$. Ky Fan's maximum principle [9] states that

$$\sum_{\ell=n-k+1}^{n} \lambda_\ell(t) = \max_{\substack{V \in \mathbb{C}^{n \times k} \\ V^*V = I_k}} \operatorname{tr}\big(V^*(T + tX)V\big).$$

For fixed V, the expression on the right is affine in t. Its pointwise maximum is therefore convex, and (9) proves the convexity of F. To obtain the first equality in (9), expand the weighted trace in the same way into a trace term and tail sums of Rayleigh quotients. Each tail sum is bounded by the corresponding eigenvalue sum, and all bounds are attained simultaneously in an ordered orthonormal eigenbasis. When T is real symmetric, the maximizing matrix may be chosen real orthogonal. Finally, since $I \preceq X \preceq nI$, each ordered eigenvalue increases by an amount between $v-u$ and $n(v-u)$ for $u < v$. Thus F is strictly increasing with bounded secant slopes. □

Spectral sensitivity under perturbation. The connection between the spectral function and eigenvalue separation can also be understood through classical perturbation theory. Suppose that $H(t)$ has simple eigenvalues, with normalized eigenvectors $u_j(t)$. The Hellmann–Feynman formula gives

$$\lambda_j'(t) = u_j(t)^* X u_j(t), \qquad F'(t) = \sum_{j=1}^{n} j\, u_j(t)^* X u_j(t). \tag{10}$$

Thus the derivative of F is a weighted sum of the responses of individual eigenvalues to the diagonal perturbation. Differentiating again and pairing the eigenvalue terms yields

$$\boxed{F''(t) = 2\sum_{j<k} \frac{(k-j)|u_j(t)^* X u_k(t)|^2}{\lambda_k(t) - \lambda_j(t)}.} \tag{11}$$

Schematically, each nonzero pairwise contribution obeys

$$\underbrace{\text{curvature contribution}}_{\text{for a pair of eigenvalues}} \propto \frac{|\text{perturbation coupling}|^2}{\text{eigenvalue gap}},$$

with the precise weight $2(k-j)$ in (11). The numerator measures the coupling between the two eigenvectors induced by X, while the denominator is their eigenvalue separation. Thus small gaps can produce large curvature when the corresponding perturbation coupling is non-negligible. Conversely, a small gap need not produce large curvature if this coupling is weak. These observations are related to the differential theory of spectral functions [14, 16, 8].

This motivates a finite-scale measure of spectral sensitivity that remains meaningful when eigenvalues are repeated and classical derivatives are unavailable.

Computing and interpreting slope variation.

Definition 2.1 (Centered slope variation)**.** *For $h > 0$, define the centered slope variation of F at t by*

$$\boxed{\mathcal{S}_h(t) := \frac{F(t+h) - F(t)}{h} - \frac{F(t) - F(t-h)}{h} \geq 0.} \tag{12}$$

The quantity $\mathcal{S}_h(t)$ requires three spectral-function evaluations and measures the difference between the right and left secant slopes over the window $[t-h, t+h]$.

Equivalently, the slope variation can be written directly in terms of the ordered eigenvalues:

$$\boxed{\mathcal{S}_h(t) = \sum_{j=1}^{n} j \left(\frac{\lambda_j(t+h) - \lambda_j(t)}{h} - \frac{\lambda_j(t) - \lambda_j(t-h)}{h} \right).} \tag{13}$$

Hence $\mathcal{S}_h(t)$ measures the change in the weighted average eigenvalue slopes between the left and right halves of the window. When the eigenvalues respond approximately linearly to the perturbation, this variation is small. Rapid changes in their relative motion, including changes in the ordering of continuously tracked eigenvalue branches, can instead produce a large slope variation.

If F is twice continuously differentiable on the window, the slope variation can also be expressed as

$$\boxed{\mathcal{S}_h(t) = \int_{-h}^{h} \left(1 - \frac{|s|}{h}\right) F''(t+s)\, ds.} \tag{14}$$

Thus $\mathcal{S}_h(t)$ is a weighted integral of the curvature over the prescribed window.

The perturbation scale offers another interpretation. If the family is extended to large positive t, the term tX dominates the fixed matrix T, and the eigenvalues satisfy $\lambda_j(t) = jt + O(1)$. In fact, F becomes asymptotically affine, so $\mathcal{S}_h(t) \to 0$ for each fixed $h > 0$ as $t \to +\infty$. This is a useful limiting picture, but it lies outside the small-perturbation regime required by our problem.

Here $0 \le t \le \delta/n$, so domination by tX cannot be assumed. Instead, the convexity and bounded total slope variation of F allow us to search for a window with sufficiently stable spectral response. The local certificate in Proposition 3.1 shows that, for our normalized tridiagonal family,

$$\mathcal{S}_h(t) < \frac{1}{2} \quad \Longrightarrow \quad \operatorname{gap}(H(t)) > \frac{h^2}{2^{20} n^4}.$$

Thus small slope variation certifies quantitative separation at the window center. The procedure in Section 4 then finds such a window.

Example 2.2 (A collision produces a corner)**.** *Let* $T = \operatorname{diag}(0.1, 0)$ *and* $X = \operatorname{diag}(1, 2)$*. Then* $H(t) = \operatorname{diag}(0.1 + t, 2t)$*, whose eigenvalues coincide at* $t = 0.1$*. Ordering them gives*

$$F(t) = \begin{cases} 0.2 + 4t, & t \le 0.1, \\ 0.1 + 5t, & t \ge 0.1, \end{cases} \qquad F(t) = 0.15 + 4.5t + \frac{1}{2}|t - 0.1|. \tag{15}$$

The slope increases from 4 *to* 5*. Dividing the centered second difference by* h *gives the centered slope variation*

$$\boxed{\mathcal{S}_h(t) = \frac{1}{2h}\big(|t + h - 0.1| + |t - h - 0.1| - 2|t - 0.1|\big) = \max\left\{1 - \frac{|t - 0.1|}{h}, 0\right\}.} \tag{16}$$

Thus $\mathcal{S}_h$ *is a triangular peak centered at* $t = 0.1$*; see Figure 1. Since affine terms cancel in the slope difference, the trace contribution in* (7) *does not affect* $\mathcal{S}_h(t)$*.*

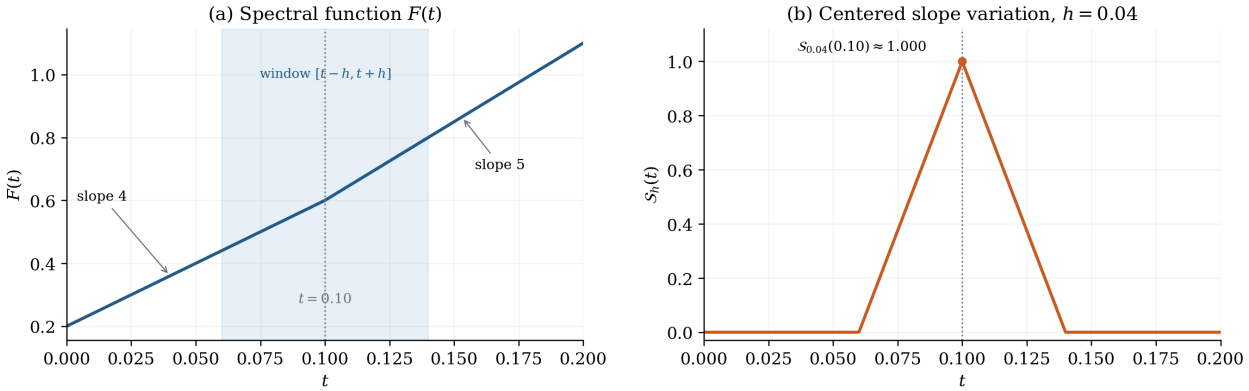


Figure 1: Example 2.2 with $h = 0.04$. Left: the spectral function $F(t)$, with the window $[t-h, t+h]$ highlighted around $t = 0.1$. The slope increases from 4 to 5 at the center. Right: the corresponding centered slope variation $\mathcal{S}_h(t)$, which forms a triangular peak of height 1 at $t = 0.1$.

Example 2.3 (A tridiagonal matrix with avoided crossings)**.** *Let*

$$T = \begin{pmatrix} 0.16 & 0.003 & 0 \\ 0.003 & 0.10 & 0.003 \\ 0 & 0.003 & 0 \end{pmatrix}, \qquad X = \operatorname{diag}(1, 2, 3). \tag{17}$$

Both off-diagonal edges are nonzero, so $T + tX$ has simple eigenvalues for every t. Thus F and $\mathcal{S}_h(t)$ are smooth. Figure 2 shows the local bend and the slope variation over a larger range of centers. There are three visible peaks near $0.06, 0.08, 0.10$, the crossing locations of the diagonal branches when the edges are set to zero. With the nonzero edges in (17), these are smooth peaks, not exact corners. These computations illustrate the slope variation; they are not certified numerical bounds.

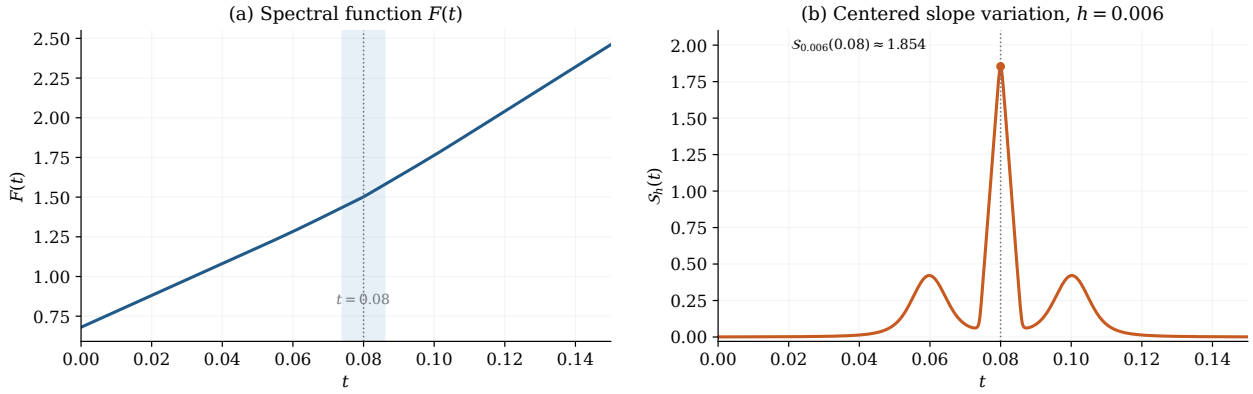


Figure 2: Example 2.3 with $h = 0.006$. Left: the spectral function $F(t)$, with the perturbation window highlighted around $t = 0.08$. Right: the corresponding centered slope variation $\mathcal{S}_h(t)$, showing three smooth peaks. At $t = 0.08$, the slope variation is approximately 1.8541.

3 Small slope variation certifies a spectral gap

The main result of this section shows that a sufficiently small centered slope variation certifies a quantitative lower bound on the minimum eigenvalue gap. We first state the certificate and then develop its two ingredients: a localization argument that uses the tridiagonal structure, and a stability argument that transfers the slope variation from a nearby auxiliary matrix to the original perturbation family.

Throughout, fix $t \in [0, \delta/n]$ and write $H = H(t) = T + tX$ when the parameter is fixed. We work in real symmetric coordinates by Remark 1.1; in particular, $\|H\|_2 \le 2$ and $|H_{i,i+1}| \le 1$.

Proposition 3.1 (Local gap certificate)**.** *For $t \in [0, \delta/n]$ and $0 < h \le 1$,*

$$\boxed{\mathcal{S}_h(t) < \frac{1}{2} \quad \Longrightarrow \quad \operatorname{gap}(H(t)) > \frac{h^2}{2^{20} n^4}.} \tag{18}$$

We prove the contrapositive. A small eigenvalue gap forces two approximate eigenvectors to be separated by a cut in the coordinate basis of X. We then make them exact eigenvectors at a common eigenvalue of a nearby matrix. Their distinct responses to X force a slope variation of at least one, and the perturbation error is too small to remove this signal.

3.1 Localization: a near-degenerate pair splits across a cut

The key structural step is to construct two approximate eigenvectors whose projections lie in disjoint coordinate subspaces. This is where the tridiagonal assumption enters; note that the tridiagonal assumption is only used in this subsection. Example 3.1 illustrates how two nearby eigenvalues of a tridiagonal matrix give rise to approximate eigenvectors supported in disjoint coordinate subspaces, with distinct responses to the diagonal perturbation.

Example 3.1 (An illustrative example of a small-leakage cut)**.** *Consider an 8×8 real symmetric tridiagonal matrix*

$$H = \operatorname{diag}(d_1, \dots, d_8) + \sum_{i=1}^{7} \beta_i (e_i e_{i+1}^{\mathsf{T}} + e_{i+1} e_i^{\mathsf{T}}),$$

where, with $\gamma = 10^{-5}$,

$$(d_1, \dots, d_8) = \left(\frac{1}{4} - \frac{\gamma}{2}, \frac{1}{2} - \frac{\gamma}{2}, \frac{1}{4} - \frac{\gamma}{2}, \frac{1}{2}, \frac{1}{2}, \frac{\gamma}{2}, \frac{\gamma}{2}, \frac{\gamma}{2}\right),$$
$$(\beta_1, \dots, \beta_7) = \left(-\frac{1}{4}, -\frac{1}{4}, 0.0005, 0.1, 0.0005, \frac{1}{4}, \frac{1}{4}\right).$$

All off-diagonal entries are nonzero, and $\|H\|_2 < 1$. *Two eigenvalues are* $\lambda \approx -5.173792 \times 10^{-6}$ *and* $\mu \approx 4.739763 \times 10^{-6}$, *satisfying* $|\lambda - \mu| < \gamma$. *Their corresponding orthonormal eigenvectors are approximately*

$$u_1 \approx (0.577344, 0.577345, 0.577346, -0.000602, 0.000123, -0.003033, 0, 0.003033)^\mathsf{T},$$
$$u_2 \approx (0.002477, 0.002477, 0.002477, 0.000145, -0.000736, 0.707100, 0, -0.707100)^\mathsf{T}.$$

We choose the cut $k = 4$, *separating the coordinate subspaces* $\operatorname{span}\{e_1, \dots, e_4\}$ *and* $\operatorname{span}\{e_5, \dots, e_8\}$. *We call* $k = 4$ *a small-leakage cut. We obtain two orthonormal approximate eigenvectors*

$$v_1 \approx (0.577350, 0.577350, 0.577351, -0.000601, 0, 0, 0, 0)^\mathsf{T},$$
$$v_2 \approx (0, 0, 0, 0, -0.000737, 0.707107, 7.37 \times 10^{-7}, -0.707107)^\mathsf{T}.$$

Their coordinate supports are disjoint, and direct computation gives

$$\left\| \left(H - \frac{\lambda + \mu}{2} I\right) v_1 \right\|_2 \approx 6.15 \times 10^{-5}, \qquad \left\| \left(H - \frac{\lambda + \mu}{2} I\right) v_2 \right\|_2 \approx 7.53 \times 10^{-5}.$$

Moreover, for $X = \operatorname{diag}(1, \dots, 8)$,

$$v_2^\mathsf{T} X v_2 - v_1^\mathsf{T} X v_1 \approx 5.0000 \geq 1.$$

How do we obtain v_1, v_2***?*** *We first construct two linear combinations of the original eigenvectors* u_1, u_2. *Choose* $p = 2$ *and* $q = 6$, *corresponding to two rows with substantial norms and distinct directions, on opposite sides of the cut. Write* $r_i = ((u_1)_i, (u_2)_i) \in \mathbb{R}^{1\times 2}$ *and define*

$$w_1 = \frac{u_1[p]u_1 + u_2[p]u_2}{\sqrt{u_1[p]^2 + u_2[p]^2}} = \frac{\gamma_p^T}{\|\gamma_p\|}(u_1, u_2), \qquad w_2 = \frac{u_1[q]u_1 + u_2[q]u_2}{\sqrt{u_1[q]^2 + u_2[q]^2}} = \frac{\gamma_q^T}{\|\gamma_q\|}(u_1, u_2),$$

where $u_j[i]$ *denotes the* i*th entry of* u_j. *Let* P_L *and* P_R *be the orthogonal projections onto* $\operatorname{span}\{e_1, \dots, e_4\}$ *and* $\operatorname{span}\{e_5, \dots, e_8\}$, *respectively. We set*

$$v_1 = \frac{P_L w_1}{\|P_L w_1\|_2}, \qquad v_2 = \frac{P_R w_2}{\|P_R w_2\|_2}.$$

Lemma 3.1 (Spatially separated approximate eigenvectors)**.** *Suppose* λ, μ *are two eigenvalues of* $H(t)$, *counted with multiplicity, satisfying* $0 \leq |\lambda - \mu| \leq \gamma \leq 1/(256n^2)$ *with* $\gamma > 0$. *Then there exist a cut* $1 \leq k < n$ *and orthonormal unit vectors* $v_1 \in \operatorname{span}\{e_1, \dots, e_k\}$ *and* $v_2 \in \operatorname{span}\{e_{k+1}, \dots, e_n\}$ *such that*

$$\boxed{\left\| \left(H(t) - \frac{\lambda + \mu}{2} I\right) v_i \right\|_2 \leq 16n\sqrt{\gamma} \quad (i = 1, 2), \qquad v_2^\mathsf{T} X v_2 - v_1^\mathsf{T} X v_1 \geq 1.} \tag{19}$$

Proof. Fix t and write $H = H(t)$. Choose orthonormal eigenvectors u_1, u_2 associated with λ, μ, respectively, and write $r_i = ((u_1)_i, (u_2)_i) \in \mathbb{R}^{1\times 2}$. By orthogonality, $\sum_{i=1}^n r_i^\mathsf{T} r_i = I_2$.

Step 1. Two substantial rows with separated directions. Let $J = \{i : \|r_i\|_2 \geq 1/(2\sqrt{n})\}$. The discarded rows have total squared norm at most $1/4$. Consequently, $\sum_{i\in J} r_i^\mathsf{T} r_i \succeq (3/4) I_2$ and $\sum_{i\in J} \|r_i\|_2^2 \leq 2$.

Choose any $p \in J$ and a unit vector $e \in \mathbb{R}^2$ perpendicular to r_p. Then $\sum_{i\in J} |r_i e|^2 \geq 3/4$, so there exists $q \in J$ satisfying $|r_q e|^2 / \|r_q\|_2^2 \geq 3/8$. After exchanging p, q if necessary, assume $p < q$.

For every nonzero row, let $\Pi_i = r_i^\mathsf{T} r_i / \|r_i\|_2^2$, the rank-one orthogonal projector onto its direction. Since $\|\Pi_i - \Pi_j\|_2 = |\det(r_i, r_j)| / (\|r_i\|_2 \|r_j\|_2)$, we obtain

$$\boxed{\|r_p\|_2, \|r_q\|_2 \geq \frac{1}{2\sqrt{n}}, \qquad \|\Pi_p - \Pi_q\|_2 \geq \sqrt{\frac{3}{8}} > \frac{1}{2}.} \tag{20}$$

Thus two rows have substantial norms but distinctly different directions.

Step 2. Wronskian identity. Write $d_i = H_{ii}$ and $\beta_i = H_{i,i+1}$. The eigenvector equations at coordinate i are

$$\beta_{i-1}(u_1)_{i-1} + d_i(u_1)_i + \beta_i(u_1)_{i+1} = \lambda(u_1)_i,$$
$$\beta_{i-1}(u_2)_{i-1} + d_i(u_2)_i + \beta_i(u_2)_{i+1} = \mu(u_2)_i.$$

Multiplying the first equation by $(u_2)_i$, the second by $(u_1)_i$, subtracting, and summing from 1 to i gives the discrete Wronskian identity

$$\boxed{\beta_i \det \begin{pmatrix} (u_1)_i & (u_1)_{i+1} \\ (u_2)_i & (u_2)_{i+1} \end{pmatrix} = \beta_i \det \begin{pmatrix} r_i \\ r_{i+1} \end{pmatrix} = (\mu - \lambda) \sum_{j=1}^{i} (u_1)_j (u_2)_j.} \tag{21}$$

Let $P_{\le i} = \sum_{j=1}^{i} e_j e_j^{\mathsf{T}}$. Since $u_1^{\mathsf{T}} u_2 = 0$, Cauchy–Schwarz gives

$$\left| \sum_{j=1}^{i} (u_1)_j (u_2)_j \right| = \left| u_1^{\mathsf{T}} \left(P_{\le i} - \frac{1}{2} I \right) u_2 \right| \le \frac{1}{2}.$$

Thus $|\beta_i|\,|\det(r_i, r_{i+1})| \le \gamma/2$.

Step 3. Finding a small-leakage cut. Set $\tau = 4\sqrt{n\gamma}$. We claim that there exists $p \le k < q$ such that

$$\boxed{|\beta_k| \max\{\|r_k\|_2, \|r_{k+1}\|_2\} \le \tau.} \tag{22}$$

Suppose otherwise. Call a row *large* if $\|r_i\|_2 > \tau$, and *small* otherwise. Since $\tau \le 1/(4\sqrt{n})$, the endpoint rows r_p, r_q are large.

We distinguish three cases.

(a) Two adjacent large rows. If r_i, r_{i+1} are both large, the failure of (22) implies $|\beta_i| \|r_i\|_2 \|r_{i+1}\|_2 > \tau^2$. The Wronskian bound therefore yields

$$\|\Pi_i - \Pi_{i+1}\|_2 = \frac{|\det(r_i, r_{i+1})|}{\|r_i\|_2 \|r_{i+1}\|_2} \le \frac{\gamma}{2\tau^2}.$$

Thus two adjacent large rows must have nearly parallel directions.

(b) Two adjacent small rows. If both r_i, r_{i+1} are small, then $|\beta_i| \le 1$ gives $|\beta_i| \max\{\|r_i\|_2, \|r_{i+1}\|_2\} \le \tau$, contradicting our assumption. Hence small rows cannot occur consecutively.

(c) A large–small–large configuration. Suppose r_{i-1}, r_{i+1} are large but r_i is small. The direction of r_i itself need not be well controlled, so we compare the directions of r_{i-1} and r_{i+1} directly.

The three-term recurrence at row i gives

$$\beta_{i-1} r_{i-1} + \left(d_i - \frac{\lambda + \mu}{2} \right) r_i + \beta_i r_{i+1} = r_i \operatorname{diag}\left(\frac{\lambda - \mu}{2}, \frac{\mu - \lambda}{2} \right).$$

Taking the determinant with r_{i-1}, using the Wronskian bound and $|d_i - (\lambda + \mu)/2| \le 4$, yields

$$\|\Pi_{i-1} - \Pi_{i+1}\|_2 \le \frac{2\gamma}{|\beta_{i-1}\beta_i| \|r_{i-1}\|_2 \|r_{i+1}\|_2} + \frac{\gamma \|r_i\|_2}{2|\beta_i| \|r_{i+1}\|_2} \le \frac{5\gamma}{\tau^2}.$$

Indeed, the failure of (22) implies $|\beta_{i-1}| \|r_{i-1}\|_2 > \tau$ and $|\beta_i| \|r_{i+1}\|_2 > \tau$, while $\|r_i\|_2 \le \tau \le 1$.

We now follow the large rows from p to q, skipping any isolated small rows. By cases (a)–(c), the total change in direction satisfies

$$\|\Pi_p - \Pi_q\|_2 \le \frac{5n\gamma}{\tau^2} = \frac{5}{16} < \frac{1}{2}.$$

This contradicts (20), proving the existence of the cut (22).

Step 4. Projection, truncation, and normalization. Using the row r_p, define the unit vector

$$w_1 = \frac{(u_1)_p u_1 + (u_2)_p u_2}{\|r_p\|_2}.$$

By construction, $(w_1)_p = \|r_p\|_2$, $|(w_1)_i| \le \|r_i\|_2$, and $\left\|(H - \frac{\lambda+\mu}{2} I)w_1\right\|_2 \le \gamma/2$.

Let $P_L = \sum_{i=1}^{k} e_i e_i^\mathsf{T}$ be the orthogonal projection onto the first k coordinate directions. Since $p \le k$, $\|P_L w_1\|_2 \ge \|r_p\|_2 \ge 1/(2\sqrt{n})$.

By tridiagonality, the commutator $[H, P_L] = HP_L - P_L H$ has only two boundary contributions:

$$[H, P_L]w_1 = \beta_k (w_1)_k e_{k+1} - \beta_k (w_1)_{k+1} e_k.$$

The small-leakage bound (22) therefore gives $\|[H, P_L]w_1\|_2 \le \sqrt{2}\tau$.

Define $v_1 = P_L w_1 / \|P_L w_1\|_2$. Combining the residual and boundary estimates,

$$\left\| \left(H - \frac{\lambda+\mu}{2} I \right) v_1 \right\|_2 \le 2\sqrt{n} \left(\frac{\gamma}{2} + \sqrt{2}\tau \right) \le 4\sqrt{n}\tau = 16n\sqrt{\gamma}.$$

Similarly, using the row r_q, define $w_2 = ((u_1)_q u_1 + (u_2)_q u_2)/\|r_q\|_2$. Let $P_R = \sum_{i=k+1}^{n} e_i e_i^\mathsf{T}$ and set $v_2 = P_R w_2 / \|P_R w_2\|_2$. The same argument gives $\left\|(H - \frac{\lambda+\mu}{2} I)v_2\right\|_2 \le 16n\sqrt{\gamma}$.

The vectors v_1, v_2 have disjoint coordinate supports and are therefore orthonormal. Finally, since $X = \mathrm{diag}(1, \ldots, n)$, $v_1^\mathsf{T} X v_1 \le k$ and $v_2^\mathsf{T} X v_2 \ge k+1$, proving $v_2^\mathsf{T} X v_2 - v_1^\mathsf{T} X v_1 \ge 1$. □

Example 3.2 (Identifying the small-leakage cut in Example 3.1)**.** *Returning to Example 3.1, write* $[u_1 \;\; u_2]^\mathsf{T} = [\gamma_1 \;\cdots\; \gamma_8]$*, where* $\gamma_i = ((u_1)_i, (u_2)_i)^\mathsf{T} = r_i^\mathsf{T}$*. The determinant in (21) is the* 2×2 *minor* $\det[\gamma_i \;\gamma_{i+1}]$*. The dashed vertical line below marks the selected cut* $k = 4$*, between columns* 4 *and* 5*.*

$$[u_1 \;\; u_2]^\mathsf{T} = [\gamma_1 \;\cdots\; \gamma_8] \approx \left[\begin{array}{cccc|cccc} 0.577344 & 0.577345 & 0.577346 & -0.000602 & 0.000123 & -0.003033 & -1.23 \cdot 10^{-7} & 0.003033 \\ 0.002477 & 0.002477 & 0.002477 & 0.000145 & -0.000736 & 0.707100 & 7.36 \cdot 10^{-7} & -0.707100 \end{array}\right]. \quad (23)$$

For $\gamma = 10^{-5}$ *and* $n = 8$*, the threshold is* $\tau = 4\sqrt{n\gamma} \approx 0.0358$*. At any cut* k*, we must check the* leakage $L_k := |\beta_k| \max\{\|\gamma_k\|_2, \|\gamma_{k+1}\|_2\}$*, not just the determinant of the two columns. The values of* L_k *for all seven cuts are:*

k	1	2	3	4	5	6	7
row types	L − L	L − L	L − S	S − S	S − L	L − S	S − L
L_k	.1443	.1443	.000289	.0000746	.000354	.1768	.1768
$L_k \le \tau$?	*No*	*No*	*Yes*	*Yes*	*Yes*	*No*	*No*

Here L *and* S *indicate whether the row norm is larger or smaller than* τ*, respectively.*

(a) Two adjacent large rows (blue box). *At* $k = 1$*, both columns have norms approximately* 0.577*, and* $|\det[\gamma_1 \;\gamma_2]| \approx 5.67 \times 10^{-8}$*. Nevertheless,* $L_1 \approx 0.1443 > \tau$*, so this is* not *a small-leakage cut. The columns are nearly parallel; this is why a small Wronskian determinant alone does not produce the cut.*

(b) Two adjacent small rows (orange box). *At* $k = 4$*, the column norms are approximately* 6.19×10^{-4} *and* 7.46×10^{-4}*. Consequently,* $L_4 \approx 7.46 \times 10^{-5} < \tau$*. Thus* the orange block gives a valid cut*. In fact, the adjacent cuts* $k = 3$ *and* $k = 5$ *also satisfy the condition, owing to their small off-diagonal couplings.*

(c) A large–small–large configuration (green boxes). *The columns* γ_6 *and* γ_8 *have norms about* 0.7071*, while* $\|\gamma_7\|_2 \approx 7.46 \times 10^{-7}$*. Yet* $L_6 \approx L_7 \approx 0.1768 > \tau$*: neither adjacent cut is valid. This is precisely why the contradiction argument uses the three-term recurrence to compare the directions of* γ_6 *and* γ_8 *directly, skipping the intermediate small column.*

The three cases are therefore not three ways of finding a cut. They explain how the proof rules out the possibility that every *cut has leakage greater than* τ*. The corresponding vectors are separated at the same cut:*

$$\begin{pmatrix} v_1^\mathsf{T} \\ v_2^\mathsf{T} \end{pmatrix} \approx \left(\begin{array}{cccc|cccc} 0.577350 & 0.577350 & 0.577351 & -0.000601 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.000737 & 0.707107 & 7.37 \cdot 10^{-7} & -0.707107 \end{array}\right). \quad (24)$$

Where the structure enters. Only the localization lemma uses the tridiagonal recurrence: through the Wronskian identity, the small-leakage cut, and the boundary terms created by coordinate truncation. The rank-two auxiliary construction and the nuclear-norm transfer apply to general Hermitian matrices whenever appropriate approximate eigenvectors with separated X-responses can be found.

Remark 3.1 (Role of the tridiagonal structure)**.** *The tridiagonal assumption enters at three points: the discrete Wronskian identity, the three-term recurrence used to control directions across an intermediate small row, and the two boundary terms created by truncation at a cut. The initial orthogonality argument and the projection onto coordinate subspaces are not specific to tridiagonal matrices.*

Thus the structural role of tridiagonality is to guarantee a cut with small boundary leakage. Once the approximate eigenvectors in Lemma 3.1 have been constructed, the auxiliary-matrix argument and the subsequent slope-variation estimates do not require the auxiliary matrix to remain tridiagonal.

3.2 From the cut to the gap certificate

The preceding lemma is the only step that uses the tridiagonal recurrence. We now replace the two localized approximate eigenvectors by exact eigenvectors with a common eigenvalue in a nearby auxiliary matrix. This gives a uniform slope response, which we then transfer back to the original family.

Lemma 3.2 (A nearby matrix with large slope variation)**.** *Under the assumptions of Lemma 3.1, there is a real symmetric matrix $\widetilde{H}$ for which*

$$\boxed{\widetilde{H}v_i = \frac{\lambda+\mu}{2}v_i \quad (i=1,2), \qquad \left\|H(t)-\widetilde{H}\right\|_* \le 96n\sqrt{\gamma},} \tag{25}$$

where $\|\cdot\|_*$ denotes the nuclear norm. The nuclear-norm bound measures its distance from $H(t)$ and will allow us to transfer the slope-variation estimate back to the original perturbation family. The auxiliary matrix is not required to be tridiagonal. Define $\widetilde{H}(s) = \widetilde{H} + (s-t)X$ and $F^{\mathrm{aux}}(s) = \sum_{j=1}^{n} j\lambda_j(\widetilde{H}(s))$. Then the centered slope variation of the auxiliary family satisfies

$$\boxed{\mathcal{S}_h^{\mathrm{aux}}(t) := \frac{F^{\mathrm{aux}}(t+h) + F^{\mathrm{aux}}(t-h) - 2F^{\mathrm{aux}}(t)}{h} \ge 1 \qquad (h>0).} \tag{26}$$

Proof. Take v_1, v_2 from Lemma 3.1 and set $Q = v_1v_1^{\mathsf{T}} + v_2v_2^{\mathsf{T}}$ and $\widetilde{H} = (I-Q)H(t)(I-Q) + (\lambda+\mu)Q/2$. Because $Qv_i = v_i$, both are exact eigenvectors of $\widetilde{H}$ at the common eigenvalue $(\lambda+\mu)/2$. Moreover, expanding $H(t)-\widetilde{H}$ gives

$$H(t)-\widetilde{H} = \big(H(t)-\tfrac{\lambda+\mu}{2}I\big)Q + Q\big(H(t)-\tfrac{\lambda+\mu}{2}I\big) - Q\big(H(t)-\tfrac{\lambda+\mu}{2}I\big)Q.$$

Since Q has rank two, the nuclear–Frobenius norm inequality and (19) yield the compact estimate

$$\left\|H(t)-\widetilde{H}\right\|_* \le 3\sqrt{2}\left(\sum_{i=1}^{2}\left\|\big(H(t)-\tfrac{\lambda+\mu}{2}I\big)v_i\right\|_2^2\right)^{1/2} \le 96n\sqrt{\gamma}.$$

For the slope variation, choose an ordered orthonormal eigenbasis of $\widetilde{H}$ in which v_1, v_2 occupy adjacent places $j, j+1$ in their common eigenspace. Assign the corresponding weights $j, j+1$ first to v_1, v_2, then in the opposite order. Both choices maximize the variational representation (9) at $s=t$. The corresponding matrices G_+, G_- satisfy

$$\operatorname{tr}(G_+\widetilde{H}) = \operatorname{tr}(G_-\widetilde{H}) = F^{\mathrm{aux}}(t), \qquad G_+ - G_- = v_2v_2^{\mathsf{T}} - v_1v_1^{\mathsf{T}}.$$

Using G_+ at $t+h$ and G_- at $t-h$ gives

$$\mathcal{S}_h^{\mathrm{aux}}(t) \ge \operatorname{tr}\big((G_+-G_-)X\big) = v_2^{\mathsf{T}}Xv_2 - v_1^{\mathsf{T}}Xv_1 \ge 1.$$

□

Proof of Proposition 3.1. We prove the contrapositive. Suppose $\mathrm{gap}(H(t)) \le \gamma := h^2/(2^{20}n^4)$. Because $0 < h \le 1$, we have $0 < \gamma \le 1/(256n^2)$. Choose two eigenvalues λ, μ realizing the minimum gap, including $\lambda = \mu$ if the gap is zero. By Lemmas 3.1 and 3.2, we have $\left\|H(t) - \widetilde{H}\right\|_* \le 96n\sqrt{\gamma}$ and $\mathcal{S}_h^{\mathrm{aux}}(t) \ge 1$.

By the variational representation (9) and operator–nuclear norm duality, for every s,

$$\boxed{|F(s) - F^{\mathrm{aux}}(s)| \le n \left\|H(s) - \widetilde{H}(s)\right\|_* = n \left\|H(t) - \widetilde{H}\right\|_*.} \tag{27}$$

The original and auxiliary families differ by the constant matrix $H(t) - \widetilde{H}$, since $H(s) - \widetilde{H}(s) = H(t) - \widetilde{H}$ for all s. Applying (27) at $t-h, t, t+h$ yields

$$\mathcal{S}_h(t) \ge \mathcal{S}_h^{\mathrm{aux}}(t) - \frac{4n}{h}\left\|H(t) - \widetilde{H}\right\|_* \ge 1 - \frac{384n^2\sqrt{\gamma}}{h}.$$

At $\gamma = h^2/(2^{20}n^4)$, the right-hand side equals $1 - 384/1024 = 5/8 > 1/2$, proving (18). □

4 The algorithm: adaptive selection of the perturbation parameter

The purpose of this section is to construct a diagonal perturbation $D = tX$ that guarantees quantitative spectral separation. By Proposition 3.1, it suffices to find a parameter t and a window radius h such that

$$\mathcal{S}_h(t) < \frac{1}{2} \quad \Longrightarrow \quad \mathrm{gap}(H(t)) > \frac{h^2}{2^{20}n^4}.$$

The key observation is that the convex spectral function F has a bounded total increase in slope. Although its slope may vary rapidly over some windows, repeated interval refinement allows us to find a window with sufficiently small slope variation.

We first establish a simple interval-selection algorithm and then improve its contraction and the resulting gap bound. Both algorithms check the local certificate at each selected window and terminate as soon as it is verified. We prove explicit worst-case iteration bounds and finally explain how the required spectral-function evaluations can be carried out efficiently.

4.1 A basic selector

We represent each interval by $I = [t-h, t+h]$, where t is its center and h is its radius. At each iteration, we compare the centered slope variations on the two middle subintervals

$$I_{\mathrm{L}} = [t - h/2, t], \qquad I_{\mathrm{R}} = [t, t + h/2].$$

Their centers are $t - h/4$ and $t + h/4$, respectively, and both have radius $h/4$.

The basic procedure selects the subinterval with smaller computed slope variation. At each iteration, we also check whether the selected window already satisfies the local gap certificate. We will show that the algorithm must terminate after at most $O(\log n)$ refinements, while maintaining an inverse-polynomial lower bound on the window radius.

Example 4.1 (One step of interval selection)**.** *Consider the spectral function in Example 2.2, whose centered slope variation is*

$$\mathcal{S}_h(t) = \max\left\{1 - \frac{|t - 0.1|}{h}, 0\right\}.$$

Starting with the illustrative window $[0.06, 0.14]$*, centered at* $t = 0.1$ *with* $h = 0.04$*, we have* $\mathcal{S}_{0.04}(0.1) = 1$*.*

After one refinement, the two candidate windows are $[0.08, 0.10]$ *and* $[0.10, 0.12]$*, with centers* 0.09 *and* 0.11 *and radius* 0.01*. Since* F *is affine on each window,*

$$\mathcal{S}_{0.01}(0.09) = \mathcal{S}_{0.01}(0.11) = 0.$$

Thus either candidate already satisfies $\mathcal{S}_h(t) < 1/2$*. More generally, the algorithm selects the candidate with smaller slope variation, moving towards a window where the spectral response is more stable.*

Algorithm 1: Basic selection of a perturbation parameter

Input: T, $n \geq 2$, and $0 < \delta < 1/2$.

Output: A perturbation parameter t and radius h.

Set $t = \delta/(2n)$ and $h = \delta/(2n)$

Let K_{b} be the smallest nonnegative integer such that $(3/2)^{K_{\mathrm{b}}} \geq 16n^3$

Evaluate $F(t-h)$, $F(t)$, and $F(t+h)$ to absolute error at most $h/256$

if $\widehat{\mathcal{S}}_h(t) + 1/64 < 1/2$ **then**

  return t, h

for $j = 0, \ldots, K_{\mathrm{b}} - 1$ **do**

  Set $h = h/4$

  Evaluate F at $t-2h, t-h, t, t+h, t+2h$ to absolute error at most $h/256$

  Compute $\widehat{\mathcal{S}}_h(t-h)$ and $\widehat{\mathcal{S}}_h(t+h)$

  if $\widehat{\mathcal{S}}_h(t-h) \leq \widehat{\mathcal{S}}_h(t+h)$ **then**

    Set $t = t - h$

  else

    Set $t = t + h$

  if $\widehat{\mathcal{S}}_h(t) + 1/64 < 1/2$ **then**

    return t, h

We next combine the exact contraction property of centered slope variation with the effect of approximate spectral-function evaluations.

Lemma 4.1 (Contraction of centered slope variation)**.** *For a parent interval $I = [t-h, t+h]$, let $I_{\mathrm{L}} = [t-h/2, t]$ and $I_{\mathrm{R}} = [t, t+h/2]$ be its two middle subintervals. Then*

$$\boxed{\min\left\{\mathcal{S}_{h/4}\left(t - \frac{h}{4}\right), \mathcal{S}_{h/4}\left(t + \frac{h}{4}\right)\right\} \leq \frac{2}{3}\mathcal{S}_h(t).} \tag{28}$$

Moreover, if each value of F is evaluated with absolute error at most one 256th of the child radius, the selected center and radius after j iterations of Algorithm 1 satisfy

$$\boxed{\mathcal{S}_{h_{j+1}}(t_{j+1}) \leq \frac{2}{3}\mathcal{S}_{h_j}(t_j) + \frac{1}{32}.} \tag{29}$$

Proof. Recall from (14) that the centered slope variation is a triangularly weighted integral of F'' whenever F is twice continuously differentiable. For general convex F, the same representation holds with its distributional second derivative interpreted as a nonnegative measure ν.

Thus

$$\mathcal{S}_h(t) = \int_{\mathbb{R}} w_{t,h}(x)\, d\nu(x), \qquad w_{t,h}(x) = \left(1 - \frac{|x-t|}{h}\right)_+.$$

A direct comparison of the triangular weights gives

$$w_{t-h/4,h/4}(x) + w_{t+h/4,h/4}(x) \leq \frac{4}{3} w_{t,h}(x).$$

Integrating against the nonnegative measure ν and taking half the sum proves (28).

For approximate evaluations, let $r = h/4$ be the child radius. Since the centered slope variation uses the coefficients $1, -2, 1$, evaluating each of its three function values within $r/256$ gives

$$\left|\widehat{\mathcal{S}}_r(s) - \mathcal{S}_r(s)\right| \leq \frac{(1+2+1)(r/256)}{r} = \frac{1}{64}.$$

Selecting the smaller computed child variation therefore incurs an additional error of at most $2/64 = 1/32$. Combining this estimate with (28) proves (29). □

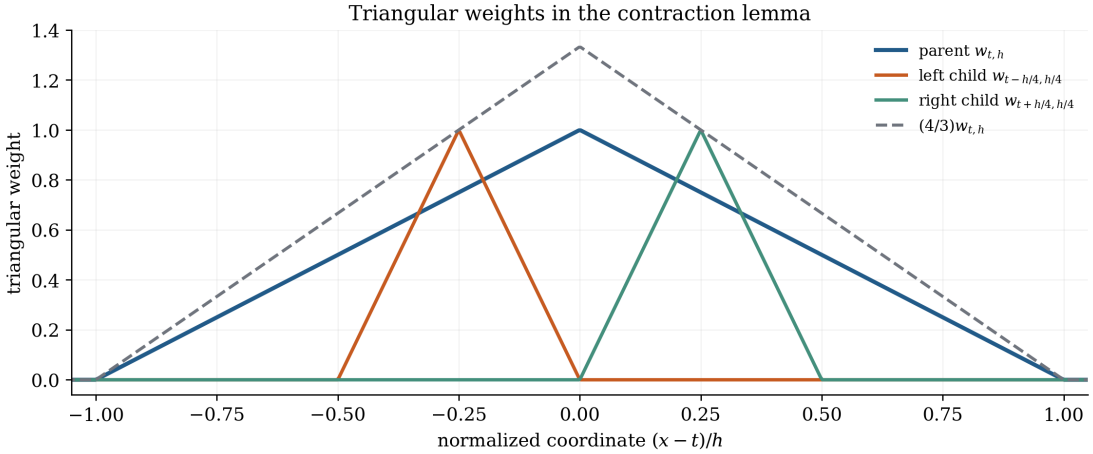


Figure 3: Triangular weights for Lemma 4.1, plotted in the normalized coordinate $(x-t)/h$. The two child weights correspond to the middle subintervals $[t-h/2, t]$ and $[t, t+h/2]$. Their sum is bounded pointwise by $(4/3)$ times the parent weight.

The contraction estimate guarantees that the slope variation eventually becomes small enough for the certified stopping test. We now bound the number of required refinements.

Theorem 4.1 (Basic iteration bound and certified scale)**.** *Algorithm 1 terminates after at most*

$$K_{\mathrm{b}} = \left\lceil \log_{3/2}(16n^3) \right\rceil = O(\log n)$$

refinements and returns t, h satisfying

$$\boxed{[t-h, t+h] \subseteq [0, \delta/n], \qquad \mathcal{S}_h(t) < \frac{1}{2}, \qquad h \geq \frac{\delta}{2n} 4^{-K_{\mathrm{b}}}.} \tag{30}$$

It requires at most $3 + 5K_{\mathrm{b}}$ spectral-function evaluations.

Consequently, the returned perturbation $D = tX$ satisfies $\|D\|_2 \leq \delta$ and

$$\boxed{\mathrm{gap}(T+D) > \frac{h^2}{2^{20} n^4}.}$$

Proof. Initially $t_0 = h_0 = \delta/(2n)$, so the initial interval is $[0, \delta/n]$. Each refinement selects a subinterval contained in its parent, with radius one quarter of the parent radius. Thus, if the algorithm stops after j refinements,

$$h_j = \frac{\delta}{2n} 4^{-j}, \qquad 0 \leq j \leq K_{\mathrm{b}}.$$

By the secant-slope bounds in Lemma 2.1, the initial slope variation satisfies $\mathcal{S}_{h_0}(t_0) < n^3$. Iterating (29) gives

$$\mathcal{S}_{h_j}(t_j) < \left(\frac{2}{3}\right)^j n^3 + \frac{1}{32} \sum_{\ell=0}^{j-1} \left(\frac{2}{3}\right)^\ell < \left(\frac{2}{3}\right)^j n^3 + \frac{3}{32}.$$

At $j = K_{\mathrm{b}}$, the first term is at most $1/16$, so

$$\mathcal{S}_{h_{K_{\mathrm{b}}}}(t_{K_{\mathrm{b}}}) < \frac{5}{32}.$$

Every computed slope variation has absolute error at most $1/64$. Consequently, at the K_{b}th refinement,

$$\widehat{\mathcal{S}}_{h_{K_{\mathrm{b}}}}(t_{K_{\mathrm{b}}}) + \frac{1}{64} < \frac{5}{32} + \frac{2}{64} < \frac{1}{2}.$$

Hence the certified stopping test must succeed by this iteration. If it succeeds earlier, the returned window is larger.

The initial test uses three evaluations of F, and each refinement uses at most five. Finally, $\|tX\|_2 = nt \leq \delta$ and Proposition 3.1 gives the claimed gap bound. □

Remark 4.1 (A lower bound on the window radius)**.** *Although the algorithm repeatedly reduces the window radius, it performs only $O(\log n)$ refinements. Let $\alpha = \log_{3/2} 4$. By the minimality of K_{b},*

$$\boxed{h > \frac{\delta}{8n}(16n^3)^{-\alpha}.} \tag{31}$$

Thus the window radius remains bounded below by an inverse polynomial in n. The next subsection improves this bound by retaining longer intervals at each refinement.

4.2 An improved selector that keeps almost half of each window

The basic algorithm is simple, but each iteration retains only one quarter of the parent interval. Since the local gap certificate depends on h^2, this repeated reduction leads to a conservative bound. We now improve the selection rule while keeping the spectral function F and the local certificate of Proposition 3.1 unchanged. By comparing slopes near the endpoints, we can retain almost half of the parent interval at each iteration.

Endpoint slope variation. For a window $I = [t-h, t+h]$ and $0 < \theta \le 1/4$, define

$$\boxed{\mathcal{S}_h^{(\theta)}(t) := \frac{F(t+h) - F(t+(1-2\theta)h)}{2\theta h} - \frac{F(t-(1-2\theta)h) - F(t-h)}{2\theta h}.} \tag{32}$$

This measures the difference between the secant slopes near the two endpoints and is nonnegative by convexity.

The endpoint slopes also control the centered slope variation. Indeed, for $0 < \theta \le 1/2$, convexity gives

$$\frac{F(t-h+2\theta h) - F(t-h)}{2\theta h} \le \frac{F(t) - F(t-h)}{h}, \qquad \frac{F(t+h) - F(t)}{h} \le \frac{F(t+h) - F(t+h-2\theta h)}{2\theta h}.$$

Subtracting yields

$$\boxed{\mathcal{S}_h(t) \le \mathcal{S}_h^{(\theta)}(t).} \tag{33}$$

Thus endpoint slope variation can be used for selection, while the stopping test remains the original condition $\mathcal{S}_h(t) < 1/2$.

Let $K = \lceil \log_2(16n^3) \rceil$ and $\theta = 1/(4K)$. For a current interval $I = [t-h, t+h]$, consider

$$I_{\mathrm{L}} = [t-(1-2\theta)h, t], \qquad I_{\mathrm{R}} = [t, t+(1-2\theta)h].$$

Both candidates have radius $h' = (1/2-\theta)h$, rather than $h/4$ as in the basic algorithm.

The following lemma is the endpoint analogue of Lemma 4.1. It combines the exact contraction and the effect of approximate evaluations.

Lemma 4.2 (Contraction of endpoint slope variation)**.** *Let $I = [t-h, t+h]$, $h' = (1/2-\theta)h$, and $0 < \theta \le 1/4$. Then*

$$\boxed{\min\left\{\mathcal{S}_{h'}^{(\theta)}(t-h'), \mathcal{S}_{h'}^{(\theta)}(t+h')\right\} \le \frac{1}{2}\mathcal{S}_h^{(\theta)}(t).} \tag{34}$$

Moreover, the selected centers and radii of Algorithm 2 satisfy

$$\boxed{\mathcal{S}_{h_{j+1}}^{(\theta)}(t_{j+1}) \le \frac{1}{2}\mathcal{S}_{h_j}^{(\theta)}(t_j) + \frac{1}{32}.} \tag{35}$$

Proof. As in Lemma 4.1, convexity gives a contraction between the parent and its two candidate windows. Here the endpoint secants are ordered, and the two child slope differences correspond to disjoint increments between the parent endpoint slopes. Hence

$$\mathcal{S}_{h'}^{(\theta)}(t-h') + \mathcal{S}_{h'}^{(\theta)}(t+h') \le \mathcal{S}_h^{(\theta)}(t),$$

which proves (34).

The inexact estimate follows as in the proof of Lemma 4.1. Each endpoint slope variation uses four evaluations of F with denominator $2\theta h_{j+1}$. The prescribed absolute error $2\theta h_{j+1}/256$ therefore gives an error of at most $1/64$ in each computed variation. Selecting the smaller computed value adds at most $2/64 = 1/32$, proving (35). □

Algorithm 2: Improved selection of a separating perturbation

Input: T, $n \geq 2$, and $0 < \delta < 1/2$.

Output: A perturbation parameter t and certification radius h.

Set $K = \lceil \log_2(16n^3) \rceil$ and $\theta = 1/(4K)$

Set $t = \delta/(2n)$ and $h = \delta/(2n)$

Evaluate $F(t-h), F(t), F(t+h)$ to absolute error at most $h/256$

if $\widehat{\mathcal{S}}_h(t) + 1/64 < 1/2$ **then**

  return t, h

for $j = 0, \ldots, K-1$ **do**

  Set $h = (1/2 - \theta)h$

  Evaluate the values of F required for $\mathcal{S}_h^{(\theta)}(t-h)$ and $\mathcal{S}_h^{(\theta)}(t+h)$, each to absolute error at most $2\theta h/256$

  Compute $\widehat{\mathcal{S}}_h^{(\theta)}(t-h)$ and $\widehat{\mathcal{S}}_h^{(\theta)}(t+h)$

  if $\widehat{\mathcal{S}}_h^{(\theta)}(t-h) \leq \widehat{\mathcal{S}}_h^{(\theta)}(t+h)$ **then**

    Set $t = t - h$

  else

    Set $t = t + h$

  Evaluate the additional midpoint value $F(t)$ to absolute error at most $2\theta h/256$

  Compute $\widehat{\mathcal{S}}_h(t)$

  if $\widehat{\mathcal{S}}_h(t) + 1/64 < 1/2$ **then**

    return t, h

Theorem 4.2 (Improved constructive spectral separation)**.** *Algorithm 2 terminates after at most* $K = \lceil \log_2(16n^3) \rceil = O(\log n)$ *refinements and returns* t, h *satisfying*

$$\boxed{\mathcal{S}_h(t) < \frac{1}{2}, \qquad h > \frac{\delta}{128n^4}.} \tag{36}$$

Consequently, $D = tX$ *satisfies*

$$\boxed{\|D\|_2 \leq \delta, \qquad \operatorname{gap}(T + D) > 2^{-34}\frac{\delta^2}{n^{12}}.} \tag{37}$$

The algorithm requires at most $3 + 8K$ *spectral-function evaluations.*

Proof. The argument follows that of Theorem 4.1, with two improvements: the contraction factor is $1/2$ instead of $2/3$, and each refinement retains a fraction $1/2 - \theta$ rather than $1/4$ of the parent radius.

By Lemma 2.1 and (35),

$$\mathcal{S}_{h_K}^{(\theta)}(t_K) \leq 2^{-K}n^3 + \frac{1}{32}\sum_{j=0}^{K-1} 2^{-j} < \frac{1}{8}.$$

By (33), we also have $\mathcal{S}_{h_K}(t_K) < 1/8$. The endpoint values and the additional midpoint evaluation give

$$\left|\widehat{\mathcal{S}}_{h_K}(t_K) - \mathcal{S}_{h_K}(t_K)\right| \leq \frac{\theta}{32} \leq \frac{1}{128}.$$

Thus $\widehat{\mathcal{S}}_{h_K}(t_K) + 1/64 < 1/2$, so the certified stopping test succeeds by iteration K.

It remains to bound the radius. As in Theorem 4.1, termination at an earlier iteration only increases the returned radius. Since

$$L_K = \frac{\delta}{n}\left(\frac{1}{2} - \theta\right)^K = \frac{\delta}{n}2^{-K}\left(1 - \frac{1}{2K}\right)^K > \frac{\delta}{64n^4},$$

where we used Bernoulli's inequality and $2^K < 32n^3$, the returned radius satisfies $h \geq L_K/2 > \delta/(128n^4)$. Applying Proposition 3.1 yields

$$\operatorname{gap}(T + tX) > \frac{h^2}{2^{20}n^4} > 2^{-34}\frac{\delta^2}{n^{12}}.$$

Finally, the initial test uses three evaluations of F, and each refinement uses at most eight. □

Remark 4.2 (Improvement over the basic selector)**.** *Compared with Algorithm 1, the improved selector achieves a contraction factor of* $1/2$ *while retaining almost half of the parent interval. Both algorithms use the same centered-slope stopping test and the same local gap certificate. The larger retained radius gives the stronger* n^{-12} *bound in* (37)*, without changing the localization argument of Section 3.*

4.3 Evaluating the spectral function in nearly linear time

We finally explain how the required values of F can be computed without treating an eigendecomposition as a unit-cost operation. Both selection algorithms use the same spectral-function oracle; the improved selector changes only the evaluation points and their required absolute accuracy.

For the tridiagonal family $H(t) = T+tX$, the characteristic polynomial can be computed using the three-term determinant recurrence. Balanced polynomial multiplication and Graeffe root-squaring then provide approximations to the ordered eigenvalues. The algebraic details are given in Section A.

Lemma 4.3 (Spectral-function oracle)**.** *For* $H(t) = T + tX$*,* $0 \leq t \leq \delta/n$*, and any* $0 < \varepsilon < 1$*, one can compute* $\widehat{F}(t)$ *satisfying*

$$\boxed{|\widehat{F}(t) - F(t)| \leq \varepsilon} \tag{38}$$

using $O(n[1 + \log(n/\varepsilon)]^3)$ *exact arithmetic operations and comparisons, where nonnegative square roots are permitted.*

Proof. The construction uses the tridiagonal determinant recurrence, balanced polynomial multiplication, and Graeffe root-squaring. We refer to Section A for the details. □

Alternatively, the spectral-function evaluation can be related to existing algorithms for computing all eigenvalues of a symmetric tridiagonal matrix. For example, Sobczyk [25] establishes nearly linear arithmetic complexity for eigenvalues-only computation under the assumptions of the corresponding fast multipole framework. Indeed, if approximations $\widehat{\lambda}_j(t)$ satisfy $|\widehat{\lambda}_j(t) - \lambda_j(t)| \leq \eta$, then

$$\left|\sum_{j=1}^{n} j\widehat{\lambda}_j(t) - F(t)\right| \leq \frac{n(n+1)}{2}\eta.$$

Thus choosing $\eta = 2\varepsilon/[n(n+1)]$ yields the required accuracy for F. This provides an alternative route to the oracle under the assumptions of that eigenvalue algorithm; our stated oracle bound uses the construction in Section A.

For the improved selector, the required absolute accuracy in each endpoint comparison is $2\theta h/256$, where h is the updated candidate radius. The additional centered-slope test uses the same accuracy and does not require a stronger spectral-function oracle.

Since the final search radius after K refinements is greater than $\delta/(128n^4)$, the smallest requested accuracy satisfies

$$\boxed{\varepsilon_{\min} > \frac{\delta}{2^{16}Kn^4}.} \tag{39}$$

Consequently,

$$\log(n/\varepsilon_{\min}) = O(1 + \log(n/\delta)).$$

Combining the at most $K = O(\log n)$ refinements, including the certified stopping tests, with Lemma 4.3, the total arithmetic cost is $O(n[1 + \log(n/\delta)]^q)$ for an absolute constant q, for example $q = 4$.

The basic selector also requires at most $O(\log n)$ refinements and inverse-polynomial evaluation accuracy, so the same oracle gives a nearly linear arithmetic bound with a polylogarithmic factor for that algorithm.

The stopping and selection tests above assume the stated absolute evaluation-error guarantees. An implementation using ordinary floating-point approximations without such guarantees may still produce a valid perturbation whenever its final centered-slope test is rigorously certified, but the worst-case iteration bound requires the endpoint-selection errors to be controlled. These complexity statements concern the exact arithmetic model. They do not imply identical finite-precision or bit-complexity bounds.

Related work: eigenvalue algorithms and complexity. The efficiency of spectral-function evaluation is essential to the overall algorithm. Although our selection procedure requires only $O(\log n)$ refinements, this does not by itself imply a nearly linear algorithm unless the required spectral evaluations can also be performed efficiently.

Classical structured eigensolvers, including the divide-and-conquer algorithm of Gu and Eisenstat [11], provide important computational foundations. Sobczyk [25] gives deterministic complexity bounds for Hermitian eigenproblems, including nearly linear eigenvalues-only computation for symmetric tridiagonal matrices under the stated fast-multipole assumptions. Such results motivate the use of a spectral oracle that computes the weighted eigenvalue sum without requiring a complete eigenvector matrix.

Related algorithmic developments include the randomized subspace methods of Nakatsukasa and Tropp [19], the contour and subspace approaches studied by Horning and Townsend [12], and rigorous spectral computations developed by Colbrook and Townsend [3, 4]. These works address different eigenvalue-computation problems, but highlight the importance of scalable algorithms and reliable spectral information.

Our contribution is complementary: we construct a small deterministic perturbation that gives a guaranteed spectral gap, while explicitly accounting for the cost of selecting and certifying that perturbation.

5 Numerical experiments

The theoretical guarantee of Theorem 4.2 is conservative. We investigate how the improved selector behaves in practice, whether its centered-slope stopping test can be certified in finite precision, and how the resulting spectral gaps compare with their analytical lower bounds.

Throughout, we use $\delta = 0.48$ and the fixed perturbation direction $X = \operatorname{diag}(1, \dots, n)$. The experiments distinguish three quantities: the worst-case iteration bound K, the actual termination iteration j_*, and the certified minimum gap of the returned matrix.

Test matrices and certified implementation. We use six real symmetric tridiagonal matrix families, summarized in Table 1, and normalize each input so that $\|T\|_2 \le 1$. The real symmetric reduction in Remark 1.1 explains why separate complex Hermitian tests are not required for the spectral claims.

Our implementation of the selection procedure uses double-precision tridiagonal eigensolvers to propose successive windows. Each proposed window is then tested using rigorous enclosures of the centered slope variation. If the available enclosure is insufficient to decide the stopping condition, the required spectral evaluations are refined using arbitrary-precision arithmetic.

For certification, we use a structured spectral oracle based on Sturm counts. Double-precision eigenvalue approximations provide initial search locations, but the final eigenvalue brackets are validated against the input matrix using interval arithmetic. Individual eigenvalues are enclosed by two-sided Sturm bracketing, while unresolved clusters are enclosed and counted collectively. The resulting ordered eigenvalue enclosures give rigorous bounds on $F(t)$, $\mathcal{S}_h(t)$, and the minimum spectral gap.

This practical oracle is distinct from the nearly linear exact-arithmetic oracle of Lemma 4.3. The structured Sturm implementation has quadratic arithmetic cost in the matrix order for a complete set of eigenvalue brackets. It is used here to demonstrate rigorous finite-precision certification rather than the asymptotic complexity of Section 4.

Table 1: Test families and the spectral features they probe.

Family	Construction	Main purpose
Random tridiagonal	Independent random diagonal and first off-diagonal entries.	Typical inputs.
Nearly reducible	Random tridiagonal matrices with selected first off-diagonal entries reduced to 10^{-9}.	Weakly coupled blocks.
Two coupled blocks	Two identical tridiagonal blocks connected by a 10^{-7} edge.	Nearly repeated pairs.
Wilkinson W_n^+	Diagonal entries $\lvert i-(n+1)/2\rvert$ and unit off-diagonals, followed by scaling.	Extremely tight clusters.
Toeplitz	Zero diagonal and constant off-diagonal entries $1/2$.	Known structured spectrum.
Repeated diagonal	Diagonal matrices with deliberately repeated entries.	Exact multiplicity.

The reported experiments precompute the double-precision selection path and then find the first window whose centered slope variation is rigorously certified below $1/2$. The measured termination index is therefore the stopping index of the adaptive algorithm along that path; the reported runtime also includes the full preliminary double-precision path construction. A separate rigorous fallback implementation controls the endpoint comparisons required by the worst-case termination proof.

Iteration counts and guaranteed termination. Algorithm 2 has the worst-case iteration bound

$$K = \lceil \log_2(16n^3) \rceil.$$

In practice, the direct centered-slope test often succeeds substantially earlier. We denote by j_* the first refinement for which the certified upper bound satisfies

$$\overline{\mathcal{S}}_{h_{j_*}}(t_{j_*}) < \frac{1}{2}.$$

Figure 4 compares the observed termination indices with the theoretical bound. For a random tridiagonal matrix of order 100, the algorithm terminates after 8 refinements, compared with $K = 24$. For the Toeplitz matrix of the same order, it requires 14 refinements. Both cases satisfy the rigorous stopping condition.

The observed number of refinements remains substantially below the worst-case bound on the tested matrices. The Toeplitz family generally requires more refinements than the random families, providing a useful structured test of the selection mechanism. These observations concern the tested instances; they do not improve the worst-case iteration bound.

Analytical versus directly certified spectral gaps. For the returned parameter and radius (t_*, h_*), the local certificate of Proposition 3.1 gives

$$g_{\mathrm{S}} := \frac{h_*^2}{2^{20} n^4} < \operatorname{gap}(T + t_* X). \tag{40}$$

Because the spectral oracle provides rigorous eigenvalue enclosures $\lambda_j(T + t_* X) \in [\underline{\lambda}_j, \overline{\lambda}_j]$, we can also compute the independent lower bound

$$g_{\mathrm{dir}} := \min_{1 \le j < n} (\underline{\lambda}_{j+1} - \overline{\lambda}_j) \le \operatorname{gap}(T + t_* X). \tag{41}$$

The second bound uses certified eigenvalue enclosures and is not required by the interval-selection theorem.

Figure 5 compares these two bounds with the universal guarantee $2^{-34}\delta^2 n^{-12}$. The differences are substantial. For the random tridiagonal matrix with $n = 100$, we obtain approximately

$$g_{\mathrm{S}} = 5.98 \times 10^{-25}, \qquad g_{\mathrm{dir}} = 5.31 \times 10^{-4}.$$

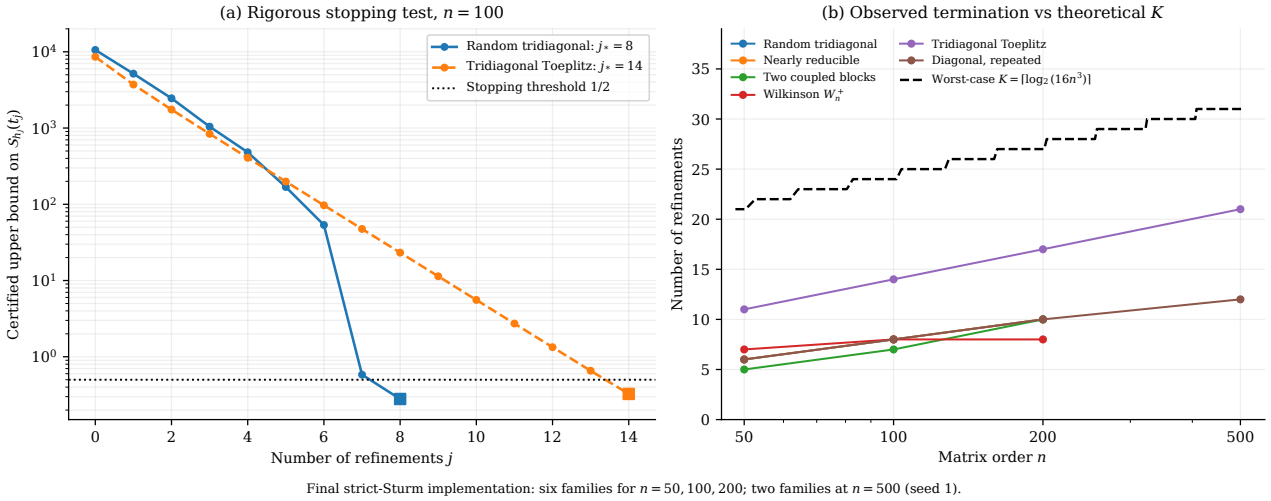


Figure 4: Certified stopping behavior. Left: rigorously enclosed centered slope variation along representative selection paths at $n = 100$, with the stopping threshold $1/2$. Right: the observed termination iteration j_* compared with the theoretical upper bound $K = \lceil \log_2(16n^3) \rceil$. The final strict-Sturm dataset includes all six families for $n = 50, 100, 200$ and two families for $n = 500$.

Thus the directly certified gap is more than twenty orders of magnitude larger than the slope-based lower bound.

The experiments reveal two distinct sources of conservatism. First, the worst-case iteration count is often unnecessary: termination at $j_* < K$ retains a larger window and hence a stronger slope-based certificate. Second, even at this larger radius, the universal local certificate remains much smaller than the actual certified gap. The second effect is particularly pronounced for structured matrices, including the Toeplitz family.

These observations identify the localization-to-gap estimate of Section 3 as a natural target for further theoretical improvement. They do not, by themselves, imply a better worst-case exponent than n^{-12}.

Finite-precision performance and scalability. The principal numerical difficulty is the evaluation of slope differences over increasingly short windows. Subtraction of nearby spectral-function values can substantially amplify absolute evaluation errors. This explains why the original double-precision stopping estimates became unreliable at moderate matrix orders.

The structured implementation avoids relying on a double-precision error model for the final certificate. Sturm bracketing supplies rigorous spectral enclosures, and additional precision is requested only when the current enclosure cannot decide the stopping condition. A bisection fallback ensures that the refinement does not depend solely on Newton convergence.

Figure 6 reports the resulting runtime. For the random matrix with $n = 100$, the complete recorded workflow takes approximately 0.94 seconds. At $n = 200$, it takes about 4.0 seconds. The Toeplitz family is more demanding: its $n = 200$ instance takes about 17.7 seconds, and the completed $n = 500$ instance takes approximately 238 seconds.

These times include the preliminary double-precision selection path and the subsequent rigorous certification. They should not be interpreted as timings for a nearly linear implementation of the theoretical oracle.

The results show that strict numerical certification is feasible well beyond the sizes accessible to the original dense arbitrary-precision implementation. The remaining bottleneck is the cost of refining many eigenvalue enclosures, especially for structured instances requiring more selection steps.

Comparison with alternative perturbations. We finally compare the deterministic selection rule with several perturbations under the common budget $\|D\|_2 \leq \delta$. These include no perturbation, the fixed ramp $(\delta/n)X$, a random parameter along the same direction, independent random diagonal perturbations, and the best parameter on a finite grid along the admissible segment.

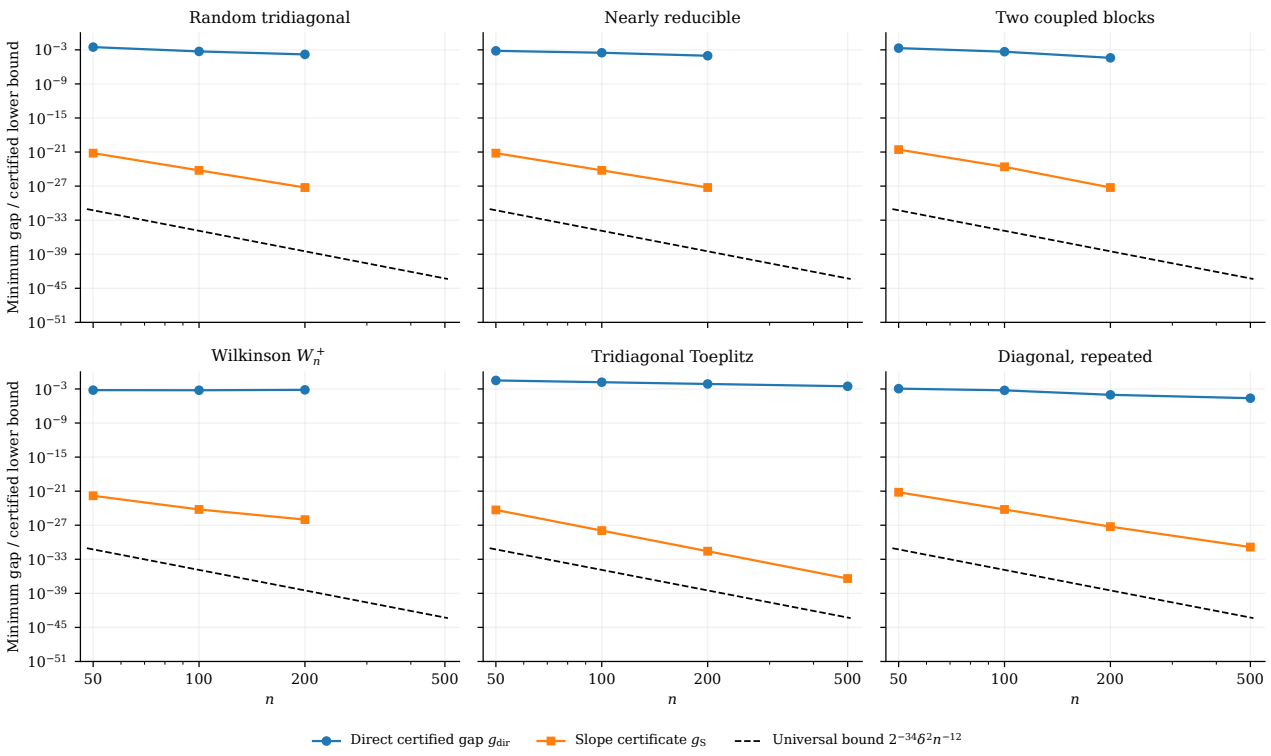


Figure 5: Comparison of spectral-gap lower bounds for the six test families. The solid curves give the directly certified minimum gap; the square markers give the slope-based certificate g_{S} at the actual stopping window. Dashed curves show the universal lower bound $2^{-34}\delta^2 n^{-12}$. All displayed points use the completed strict-Sturm dataset.

Figure 7 summarizes the comparison at $n = 100$. The paired baseline distributions come from the original double-precision sweep, while the overlaid strict points give the directly certified gaps for representative instances of the improved selector.

The deterministic selector is competitive with random perturbations on the tested matrix families. In particular, the structured Toeplitz examples illustrate that a suitable diagonal direction can outperform unstructured diagonal noise. The grid search can yield larger gaps, but it uses many more trial parameters and does not supply the same worst-case selection guarantee.

The purpose of our algorithm is not to maximize the minimum gap along the admissible segment. Rather, it constructs a parameter at which spectral separation can be certified within a controlled number of refinements.

For random tridiagonal matrices and weakly coupled blocks, the selected perturbation typically achieves a minimum gap two to three times larger than the random baselines.

The advantage is particularly pronounced for structured inputs. On the Wilkinson and Toeplitz examples, the selected perturbation achieves substantially larger gaps than independent random diagonal perturbations. For the Toeplitz matrix, its performance is also close to the best parameter found by grid search.

These comparisons illustrate the benefit of structured perturbations and deterministic parameter selection. The objective of our method, however, is not to maximize the minimum gap, but to construct a perturbation with a rigorous spectral-separation certificate and a worst-case iteration guarantee.

Summary. The experiments support three conclusions. First, the centered-slope stopping test often terminates well before the theoretical limit K. Second, rigorous structure-aware spectral computations make certified separation practical at substantially larger dimensions than the original dense high-precision implementation. Third, the directly certified gaps are typically many orders of magnitude larger than the analytical lower bounds.

The main limitation of the theoretical gap guarantee thus appears to lie in the conservatism of the universal

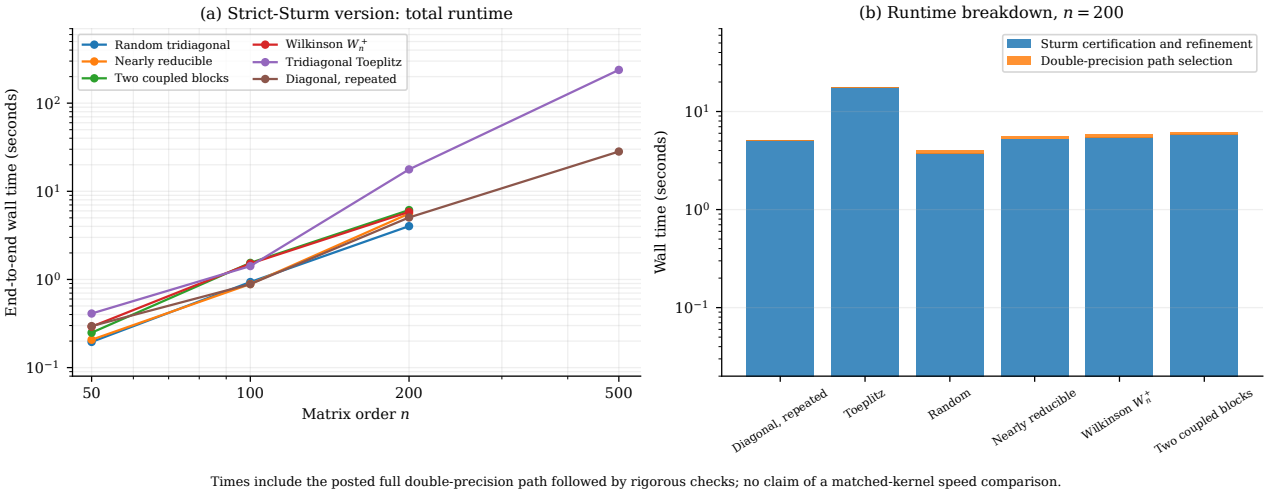


Figure 6: Runtime of the strict-Sturm implementation. Left: end-to-end runtime across completed matrix families and sizes. Right: time spent on preliminary double-precision selection and rigorous certification at $n = 200$. The experiments demonstrate the practical benefit of structure-exploiting certification, without claiming the arithmetic complexity of the oracle in Lemma 4.3.

Table 2: Representative runs of the strict-Sturm implementation with $\delta = 0.48$. The index j_* is the first rigorously certified stopping iteration. Runtime includes both selection path construction and certification.

Family and n	K	j_*	g_S	g_{dir}	Time (s)
Random, 100	24	8	5.98×10^{-25}	5.31×10^{-4}	0.94
Toeplitz, 100	24	14	1.13×10^{-28}	1.54×10^{-2}	1.42
Random, 200	27	10	5.63×10^{-28}	1.58×10^{-4}	4.02
Wilkinson, 200	27	8	9.71×10^{-27}	6.97×10^{-4}	5.84
Toeplitz, 200	27	17	2.65×10^{-32}	7.57×10^{-3}	17.72
Diagonal, 500	31	12	1.42×10^{-31}	2.37×10^{-5}	28.25
Toeplitz, 500	31	21	4.04×10^{-37}	2.99×10^{-3}	238.35

local certificate, rather than in the ability of the selector to find a separating perturbation on these test matrices.

6 Conclusions and outlook

We have given a deterministic construction of a small diagonal perturbation that produces inverse-polynomial spectral separation for arbitrary Hermitian tridiagonal matrices. The proof combines tridiagonal invariant-subspace localization with the variational geometry of a convex weighted spectral function. Its finite-scale slope variation provides a local gap certificate, while an adaptive, error-tolerant window selector yields the global perturbation and a quantitative bound on the number of refinements. The resulting construction addresses the nearly linear arithmetic version of the deterministic Minami perturbation problem, subject to the stated spectral-oracle model.

Rigorous numerical experiments show that the stopping criterion often succeeds well before the theoretical iteration bound. The structured Sturm implementation also provides a practical means of certifying individual eigenvalue enclosures and, consequently, lower bounds on the actual minimum gap. These direct bounds are far larger than the universal slope-based certificate on the tested matrices. Thus the finite-scale selection mechanism performs well on these instances, whereas the localization-to-gap constants and dimension dependence remain natural targets for further analysis.

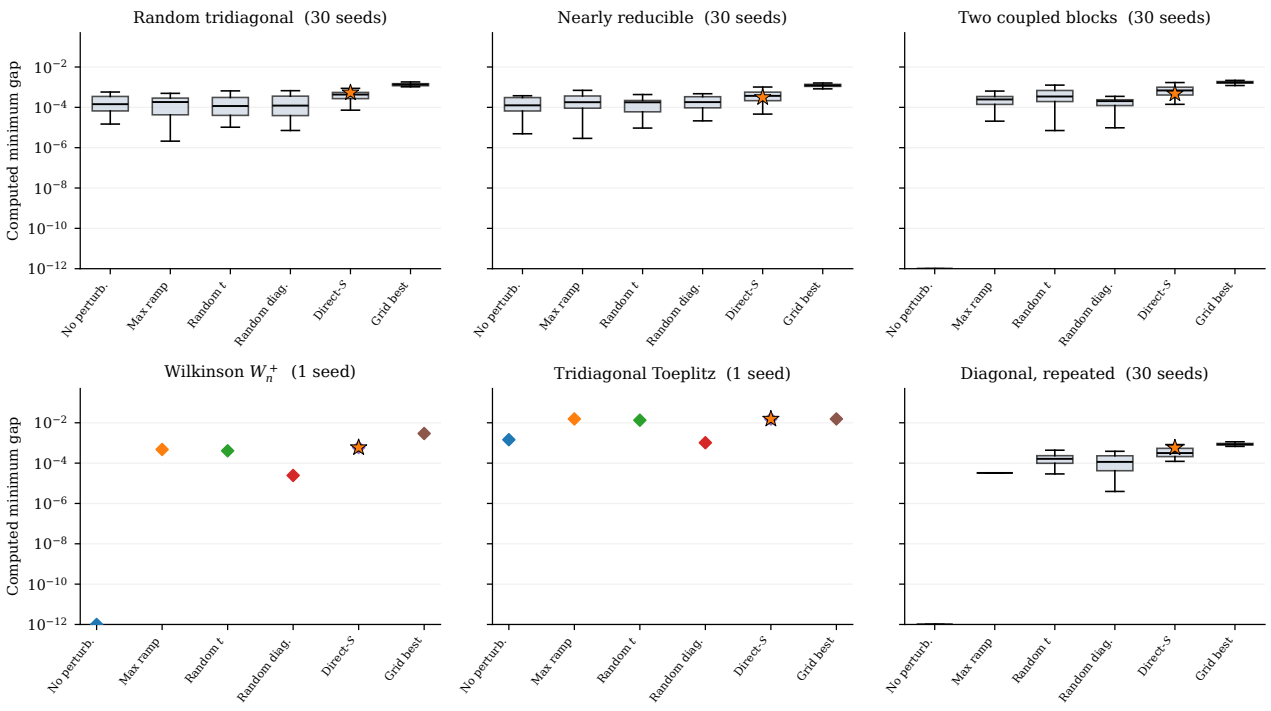


Figure 7: Comparison with alternative perturbations at $n = 100$. The paired distributions are computed in double precision using the original 30-seed baseline sweep. Orange stars indicate the strict-Sturm direct gap for the representative seed-1 output of the adaptive selector. The baseline distributions are numerical comparisons, not formal gap certificates.

Sharper certificates and precision costs. The current local estimate combines an $O(n\sqrt{\gamma})$ quasimode residual from the small-leakage cut with a stability argument for a weighted spectral function. Understanding whether classical localization techniques, including the complementary-block analyses of Ye and Parlett and Nakatsukasa's blockwise sensitivity bounds [27, 23, 26, 18], can yield better dimensional dependence is an immediate question. A second issue is finite-precision complexity. The nearly linear arithmetic guarantee does not itself establish the cost in bits of certified slope computations. An efficient bit-complexity analysis, or a numerically better-conditioned finite-scale observable, would further connect the theory to robust large-scale eigenvalue preprocessing [25, 10].

Banded and block tridiagonal matrices. For bandwidth greater than one, the scalar three-term Wronskian identity and the single-edge commutator are replaced by multi-component boundary interactions. Moreover, the response of a multiple eigenspace to the fixed diagonal ramp X can be scalar even when the eigenspace has dimension two. Thus a direct extension of the one-direction local certificate is not automatic. The block tridiagonal perturbation estimates of Nakatsukasa [18] suggest studying localization and response separation using several diagonal directions, possibly with the number of directions depending on the bandwidth rather than on n.

Multiple directions for general Hermitian matrices. For any two orthonormal vectors $u, v \in \mathbb{C}^n$, consider the coordinate probes $E_i = e_i e_i^*$. Their compressions to $\operatorname{span}\{u, v\}$ have eigenvalue spreads $\Delta_i = |u_i|^2 + |v_i|^2$, and hence

$$\max_i \Delta_i \geq \frac{2}{n}.$$

This shows that a predetermined family of n coordinate directions can distinguish the perturbation responses in every two-dimensional subspace. Similarly, independent random sign diagonal probes satisfy a mean-square

response-separation estimate. These observations concern *detecting* a nearly degenerate eigenspace: they do not yet give a single diagonal perturbation that simultaneously separates the entire spectrum. Whether far fewer directions suffice for banded matrices, or whether adaptive and randomized probes can be converted into an efficient deterministic simultaneous-separation procedure, remains an open direction of investigation. Related questions for larger eigenspace clusters and complex Hermitian matrix families are also natural, but require additional structure beyond the present rank-two tridiagonal argument.

Beyond eigenvalue gaps. Diagonal separation may serve as a preprocessing component in subspace iteration, deflation, and verified spectral computations, including problems where eigenvectors and spectral projectors are more sensitive than ordered Hermitian eigenvalues themselves. The sketched subspace methods of Nakatsukasa and Tropp [19] and the certified spectral approaches of Colbrook and Townsend [3, 4] provide useful, complementary algorithmic contexts. Establishing an end-to-end improvement for a particular parent eigensolver, rather than only a guaranteed separation bound, is a further question.

Statements and Declarations

Funding. The research of Wenqi Zhu was supported by the Engineering and Physical Sciences Research Council under grant EP/Y028872/1, *Mathematical Foundations of Intelligence: An "Erlangen Programme" for AI*. Ping Nie gratefully acknowledges the support of AI4ALPHA and NetMind.AI.

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

Data availability. The certified implementation, the six families of test matrices and the scripts producing the figures in Section 5 are available from the corresponding author on reasonable request.

Use of generative AI and AI-assisted technologies. During the preparation of this work, the authors used AI tools (Claude, Anthropic; Codex, OpenAI; Qwen App, Alibaba) to generate candidate constructions and code, perform computations, draft mathematical arguments, and improve the readability of the text. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication. The theorems are established by the written proofs; the numerical experiments use rigorous Sturm-based eigenvalue enclosures to certify the reported quantities.

A The spectral-function oracle: algebraic details

We provide the arithmetic-model construction used in Lemma 4.3. Let A have diagonal entries a_i and squared off-diagonal magnitudes $c_i = |A_{i,i+1}|^2$. Since $\|A\|_2 < 2$, the numbers $r_j = 2 + \lambda_j(A)$ lie in $(0, 4)$. Form the monic polynomial $P(z) = \det(zI + 2I + A) = \prod_{j=1}^{n}(z + r_j)$. Its tridiagonal determinant recurrence is

$$P_{-1} = 0, \quad P_0 = 1, \qquad P_i = (z + 2 + a_i)P_{i-1} - c_{i-1}P_{i-2}, \quad c_0 = 0.$$

Writing this recurrence as a product of 2×2 polynomial transfer matrices and multiplying with a balanced product tree computes all coefficients of P in $O(n[1 + \log n]^2)$ exact arithmetic operations when degree-m polynomial products cost $O((m+1)[1+\log(m+1)])$. Such multiplication is available through radix-two Fourier transforms. The required roots of unity are constructed in the permitted arithmetic model from nonnegative square roots by successive half-angle formulas.

If $P(z) = E(z^2) + zO(z^2)$, apply the Graeffe transform

$$G(P)(z) = E(-z)^2 + zO(-z)^2 = \prod_{j=1}^{n}(z + r_j^2).$$

After k iterations, $R = 2^k$ and the transformed polynomial is $\prod_j (z + r_j^R)$. Order $r_1 \geq \cdots \geq r_n > 0$ for this argument, and let e_j be its positive elementary symmetric coefficients, with $e_0 = 1$. Setting $p_j = \prod_{i=1}^{j} r_i^R$ gives $p_j \leq e_j \leq 2^n p_j$. Hence

$$\boxed{r_j 2^{-n/R} \leq \left(\frac{e_j}{e_{j-1}}\right)^{1/R} \leq r_j 2^{n/R}.} \tag{42}$$

For $R \geq 8n/\eta$, $0 < \eta < 1$, the approximation $\widehat{r}_j = (e_j/e_{j-1})^{1/R}$ has $|\widehat{r}_j - r_j| \leq 4n/R \leq \eta/2$, using $0 < r_j < 4$ and $2^x - 1 \leq x$ for $0 \leq x \leq 1$. Taking the Rth root requires k successive nonnegative square roots, and $k = O(1+\log(n/\eta))$. The Graeffe iterations each require polynomial squaring, giving $O(n[1+\log n][1+\log(n/\eta)])$ arithmetic operations.

Set $\eta = \varepsilon/(\sum_{j=1}^{n} j)$, reverse the approximate r_j into nondecreasing order and subtract 2. Then $|\widehat{\lambda}_j - \lambda_j(A)| \leq \eta/2$ and $|\sum_j j\widehat{\lambda}_j - \sum_j j\lambda_j(A)| \leq \varepsilon$. The combined arithmetic cost is bounded by $O(n[1+\log(n/\varepsilon)]^3)$. Intermediate coefficient magnitudes and their bit lengths are not charged in the stated exact-arithmetic model.

References

[1] Noah Amsel et al. Linear systems and eigenvalue problems: Open questions from a Simons workshop, 2026.

[2] Roger W. Brockett. Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems. *Linear Algebra and its Applications*, 146:79–91, 1991.

[3] Matthew J. Colbrook and Alex Townsend. Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems. *Communications on Pure and Applied Mathematics*, 77(1):221–283, 2024.

[4] Matthew J. Colbrook and Alex Townsend. Avoiding discretization issues for nonlinear eigenvalue problems. *SIAM Journal on Matrix Analysis and Applications*, 46(1):648–675, 2025.

[5] Jean-Michel Combes, Francois Germinet, and Abel Klein. Generalized eigenvalue-counting estimates for the anderson model. *Journal of Statistical Physics*, 135(2):201–216, 2009.

[6] Jane Cullum, W. E. Donath, and Philip Wolfe. The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices. In M. L. Balinski and Philip Wolfe, editors, *Nondifferentiable Optimization*, volume 3 of *Mathematical Programming Studies*, pages 35–55. Springer, 1975.

[7] W. E. Donath and Alan J. Hoffman. Lower bounds for the partitioning of graphs. *IBM Journal of Research and Development*, 17(5):420–425, 1973.

[8] Dmitriy Drusvyatskiy and Courtney Paquette. Variational analysis of spectral functions simplified. *Journal of Convex Analysis*, 25(1):119–134, 2018.

[9] Ky Fan. On a theorem of Weyl concerning eigenvalues of linear transformations I. *Proceedings of the National Academy of Sciences of the United States of America*, 35(11):652–655, 1949.

[10] Emil Graf and Alex Townsend. Numerical instability of algebraic rootfinders. *Linear Algebra and its Applications*, 729:308–336, 2026.

[11] Ming Gu and Stanley C. Eisenstat. A divide-and-conquer algorithm for the symmetric tridiagonal eigenproblem. *SIAM Journal on Matrix Analysis and Applications*, 16(1):172–191, 1995.

[12] Andrew Horning and Alex Townsend. Feast for differential eigenvalue problems. *SIAM Journal on Numerical Analysis*, 58(2):1239–1262, 2020.

[13] Adrian S. Lewis. Convex analysis on the Hermitian matrices. *SIAM Journal on Optimization*, 6(1):164–177, 1996.

[14] Adrian S. Lewis. Derivatives of spectral functions. *Mathematics of Operations Research*, 21(3):576–588, 1996.

[15] Adrian S. Lewis and Michael L. Overton. Eigenvalue optimization. *Acta Numerica*, 5:149–190, 1996.

[16] Adrian S. Lewis and Hristo S. Sendov. Twice differentiable spectral functions. *SIAM Journal on Matrix Analysis and Applications*, 23(2):368–386, 2001.

[17] Nariyuki Minami. Local fluctuation of the spectrum of a multidimensional Anderson tight binding model. *Communications in Mathematical Physics*, 177(3):709–725, 1996.

[18] Yuji Nakatsukasa. Eigenvalue perturbation bounds for hermitian block tridiagonal matrices. *Applied Numerical Mathematics*, 62(1):67–78, 2012.

[19] Yuji Nakatsukasa and Joel A. Tropp. Fast and accurate randomized algorithms for linear systems and eigenvalue problems. *SIAM Journal on Matrix Analysis and Applications*, 2024.

[20] Open Problems in NLA. KE-02: Construct a separating diagonal perturbation in nearly linear time. https://github.com/ajt60gaibb/OpenProblemsInNLA/tree/main/eigenvalues-and-inverse-problems/KE-02.

[21] Michael L. Overton and Robert S. Womersley. On the sum of the largest eigenvalues of a symmetric matrix. *SIAM Journal on Matrix Analysis and Applications*, 13(1):41–45, 1992.

[22] Michael L. Overton and Robert S. Womersley. Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices. *Mathematical Programming*, 62(2):321–357, 1993.

[23] Beresford N. Parlett. Invariant subspaces for tightly clustered eigenvalues of tridiagonals. *BIT Numerical Mathematics*, 36(3):542–562, 1996.

[24] Max Simchowitz, Kevin Jamieson, Jordan W. Suchow, and Thomas L. Griffiths. Adaptive sampling for convex regression, 2018. Preprint.

[25] Aleksandros Sobczyk. Deterministic complexity analysis of Hermitian eigenproblems, 2025. arXiv:2410.21550v2.

[26] Christof Vömel and Beresford N. Parlett. Detecting localization in an invariant subspace. *SIAM Journal on Scientific Computing*, 33(6):3447–3467, 2011.

[27] Qiang Ye. On close eigenvalues of tridiagonal matrices. *Numerische Mathematik*, 70(4):507–514, 1995.