A Deterministic Algorithm for Eigenvalue Gap Separation of Tridiagonal Matrices

Sketch of proof

Wenqi Zhu[1] and Ping Nie[2]

[1]Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK. wenqi.zhu@maths.ox.ac.uk

[2]David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

ping.nie@uwaterloo.ca

The problem and our answer. Let $T \in \mathbb{C}^{n\times n}$ be Hermitian tridiagonal with $\|T\|_2 \le 1$, let $0 < \delta < 1/2$, and write $\mathrm{gap}(T) = \min_j(\lambda_{j+1}(T) - \lambda_j(T))$ for its minimum eigenvalue gap. Minami-type estimates say that a *random* diagonal perturbation of size δ separates the eigenvalues with high probability. The *deterministic Minami perturbation* problem (Amsel et al., Problem 3.2 [1]; catalogue problem KE-02 [2]) asks for a deterministic diagonal D with

$$\|D\|_2 \le \delta, \qquad \mathrm{gap}(T + D) \ge c_0(\delta/n)^p,$$

computable in nearly linear time $O(n[1+\log(n/\delta)]^q)$, without first computing an eigendecomposition. We give such an algorithm:

$$\boxed{D = tX, \quad X = \mathrm{diag}(1,\dots,n), \quad t \in [0, \delta/n] \ \text{ chosen so that } \ \mathrm{gap}(T+D) > 2^{-34}\,\frac{\delta^2}{n^{12}},}$$

after at most $K = \lceil \log_2(16n^3) \rceil$ refinements, each costing a constant number of evaluations of one convex spectral function. The proof has two halves: a *local certificate* (Steps 2–4) saying that small slope variation of the spectral function implies a gap, and an *adaptive search* (Steps 5–6) that finds a window where the certificate holds. By a diagonal unitary similarity we may assume T is real symmetric.

1. One direction suffices

Role. Reduce the search over diagonal matrices to a search over one scalar.

Fix the single tilt direction $X = \mathrm{diag}(1,\dots,n)$ and the family

$$\boxed{H(t) = T + tX, \qquad 0 \le t \le \delta/n.} \tag{1}$$

Then $\|tX\|_2 = nt \le \delta$ for every admissible t, so the budget is automatic, $\|H(t)\|_2 < 2$, and every off-diagonal entry has modulus at most one. What remains is to *certify* separation at a chosen t without computing $\mathrm{gap}(H(t))$ accurately.

2. A convex spectral function and its slope variation

Role. Replace the nonsmooth quantity gap by a convex function of t whose finite differences we can control.

Define the weighted eigenvalue sum

$$\boxed{F(t) := \sum_{j=1}^{n} j\,\lambda_j(H(t)) = \max_{U^*U=I} \operatorname{tr}\big(UXU^*H(t)\big).} \tag{2}$$

By Ky Fan's principle F is convex and strictly increasing, with secant slopes between $n(n+1)/2$ and $n^2(n+1)/2$; up to the affine trace term it is the total pairwise spectral distance $\sum_{j<\ell}(\lambda_\ell - \lambda_j)$. For a window $[t-h, t+h]$ define the *centered slope variation*

$$\boxed{\mathcal{S}_h(t) := \frac{F(t+h)-F(t)}{h} - \frac{F(t)-F(t-h)}{h} = \sum_j j\Big(\tfrac{\lambda_j(t+h)-\lambda_j(t)}{h} - \tfrac{\lambda_j(t)-\lambda_j(t-h)}{h}\Big) \;\geq 0.} \tag{3}$$

It needs three evaluations of F. Second-order perturbation theory shows that the curvature of F contains terms $|\text{coupling}|^2/(\text{gap})$: a near-degenerate pair *can* produce large $\mathcal{S}_h$, but only if the two eigenvalues respond differently to X. Steps 3–4 show that for tridiagonal T they always do.

3. Localization: a near-degenerate pair splits across a cut

Role. This is where the tridiagonal structure enters. Two close eigenvalues force two approximate eigenvectors living on opposite sides of a coordinate cut, hence with different X-responses.

Let λ, μ be eigenvalues of $H = H(t)$ with $|\lambda - \mu| \leq \gamma \leq 1/(256n^2)$, with orthonormal eigenvectors u_1, u_2, and put $r_i = ((u_1)_i, (u_2)_i)$. The eigen-equations at coordinate i give the *discrete Wronskian identity*

$$\boxed{\beta_i \det \begin{pmatrix} r_i \\ r_{i+1} \end{pmatrix} = (\mu-\lambda)\sum_{j\leq i}(u_1)_j(u_2)_j, \qquad \text{so} \qquad |\beta_i|\,|\det(r_i, r_{i+1})| \leq \gamma/2.} \tag{4}$$

With $\tau = 4\sqrt{n\gamma}$ there is a cut k with $|\beta_k| \max\{\|r_k\|, \|r_{k+1}\|\} \leq \tau$ (*small leakage*): otherwise adjacent large rows are nearly parallel, small rows cannot be consecutive, and a large–small–large pattern is still nearly parallel, so all rows would be nearly parallel and $u_1 \perp u_2$ would be violated. Cutting at k and renormalising gives orthonormal $v_1 \in \operatorname{span}\{e_1, \ldots, e_k\}$, $v_2 \in \operatorname{span}\{e_{k+1}, \ldots, e_n\}$ with

$$\boxed{\Big\|(H - \tfrac{\lambda+\mu}{2}I)v_i\Big\|_2 \leq 16n\sqrt{\gamma}, \qquad v_2^{\mathsf{T}}Xv_2 - v_1^{\mathsf{T}}Xv_1 \geq 1.} \tag{5}$$

The last inequality is just that X is increasing along the diagonal.

4. The local gap certificate

Role. Convert the separated responses into a lower bound on the gap, through a nearby matrix and the variational formula for F.

Make v_1, v_2 exact eigenvectors of a nearby symmetric $\widetilde{H}$ with common eigenvalue $(\lambda+\mu)/2$ and $\|H-\widetilde{H}\|_* \le 96n\sqrt{\gamma}$ (nuclear norm). For the auxiliary family $\widetilde{H}+(s-t)X$ the two supporting slopes at t differ by at least $v_2^{\mathsf{T}}Xv_2 - v_1^{\mathsf{T}}Xv_1 \ge 1$, so $\mathcal{S}_h^{\mathrm{aux}}(t) \ge 1$ for every $h>0$. Operator–nuclear duality in the variational formula gives $|F(s)-F^{\mathrm{aux}}(s)| \le n\|H-\widetilde{H}\|_*$ for all s, hence

$$\mathcal{S}_h(t) \;\ge\; 1 - \frac{4n}{h}\cdot 96n\sqrt{\gamma} = 1 - \frac{384n^2\sqrt{\gamma}}{h}.$$

Taking $\gamma = h^2/(2^{20}n^4)$ makes the right side $5/8$. Contrapositively:

$$\boxed{\mathcal{S}_h(t) < \tfrac{1}{2} \quad\Longrightarrow\quad \mathrm{gap}(H(t)) > \frac{h^2}{2^{20}n^4} \qquad (t\in[0,\delta/n],\ 0<h\le 1).} \tag{6}$$

A small, cheaply computable number certifies a gap.

5. Adaptive selection of the window

Role. Find t, h with $\mathcal{S}_h(t) < 1/2$ and h not too small, using convexity to decide where to look.

Start from $[0,\delta/n]$. At each refinement compare the slope variations of the two candidate half-windows and keep the smaller one. For the basic rule (child radius $h/4$) convexity gives the contraction

$$\boxed{\min\Big\{\mathcal{S}_{h/4}(t-\tfrac{h}{4}),\ \mathcal{S}_{h/4}(t+\tfrac{h}{4})\Big\} \le \tfrac{2}{3}\,\mathcal{S}_h(t),} \tag{7}$$

and with F evaluated to accuracy $h/256$ the inexact iteration satisfies $\mathcal{S}_{h_{j+1}}(t_{j+1}) \le \frac{2}{3}\mathcal{S}_{h_j}(t_j) + \frac{1}{32}$. Since $\mathcal{S}_h(t) \le n^2(n+1)/2 - n(n+1)/2 < n^3$ initially, after $O(\log n)$ steps the certified test $\widehat{\mathcal{S}}_h(t) + \frac{1}{64} < \frac{1}{2}$ must succeed. The improved rule compares *endpoint* slope variations $\mathcal{S}_h^{(\theta)}$ with $\theta = 1/(4K)$ and keeps a child of radius $(\frac{1}{2}-\theta)h$, contracting by $\frac{1}{2}$ instead of $\frac{2}{3}$ while losing only a constant factor in h. Altogether:

$$\boxed{K = \lceil \log_2(16n^3)\rceil, \qquad \mathcal{S}_h(t) < \tfrac{1}{2}, \qquad h > \frac{\delta}{128n^4}, \qquad \mathrm{gap}(T+tX) > 2^{-34}\frac{\delta^2}{n^{12}},} \tag{8}$$

using at most $3+8K$ evaluations of F. The algorithm stops as soon as the certificate holds, which in practice is long before K steps.

6. Evaluating F in nearly linear time, and what the experiments show

Role. Make each evaluation cheap without treating an eigendecomposition as a unit-cost operation, and check the whole scheme numerically.

The characteristic polynomial of $H(t)$ comes from the three-term determinant recurrence; balanced polynomial multiplication and Graeffe root-squaring then approximate the ordered eigenvalues, giving an oracle

$$\boxed{|\widehat{F}(t) - F(t)| \le \varepsilon \quad \text{in} \quad O\big(n[1+\log(n/\varepsilon)]^3\big) \text{ exact operations.}} \tag{9}$$

The smallest accuracy ever requested is $\varepsilon_{\min} > \delta/(2^{16}Kn^4)$, so $\log(n/\varepsilon_{\min}) = O(1+\log(n/\delta))$, and the total cost is $O(n[1+\log(n/\delta)]^4)$: the nearly linear target of KE-02 in the exact-arithmetic model.

Experiments. A structure-exploiting implementation with rigorous Sturm eigenvalue enclosures and adaptive precision certifies the stopping test on six families (random tridiagonal, nearly reducible, two coupled blocks, Wilkinson W_n^+, Toeplitz, and repeated diagonal). The certified stopping condition is usually met well before K refinements; the directly certified gaps are many orders of magnitude larger than the universal $2^{-34}\delta^2 n^{-12}$ bound and decrease roughly like n^{-2} on the tested ranges; and the deterministic selector is competitive with random perturbations. The constants and the exponent 12 are therefore artefacts of the worst-case analysis, not of the construction.

References

[1] N. Amsel et al. *Linear Systems and Eigenvalue Problems: Open Questions from a Simons Workshop.* arXiv:2602.05394 (2026), Problem 3.2.

[2] Open Problems in Numerical Linear Algebra, problem KE-02. https://github.com/ajt60gaibb/OpenProblemsInNLA/tree/main/eigenvalues-and-inverse-problems/KE-02.