ELSEVIER Linear Algebra and its Applications 309 (2000) 121–151

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

Relatively robust representations of symmetric tridiagonals

Beresford N. Parlett [a,*], Inderjit S. Dhillon [b]

[a] *Mathematics Department and Computer Science Division, EECS Department, University of California, Berkeley, CA 94720, USA*

[b] *IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120-6099, USA*

Received 12 February 1999; accepted 5 December 1999

Submitted by J.L. Barlow

Abstract

Let LDL^t be the triangular factorization of an unreduced symmetric tridiagonal matrix $T - \tau I$. Small *relative* changes in the nontrivial entries of L and D may be represented by diagonal scaling matrices Δ_1 and Δ_2; $LDL^t \longrightarrow \Delta_2 L \Delta_1 D \Delta_1 L^t \Delta_2$. The effect of Δ_2 on the eigenvalues $\lambda_i - \tau$ is benign. In this paper we study the inner perturbations induced by Δ_1. Suitable condition numbers govern the *relative* changes in the eigenvalues $\lambda_i - \tau$. We show that when $\tau = \lambda_j$ is an eigenvalue then the *relative* condition number of $\lambda_m - \lambda_j$, $m \neq j$, is the same for all n twisted factorizations, one of which is LDL^t, that could be used to represent $T - \tau I$. See Section 2.

We prove that as $\tau \longrightarrow \lambda_j$ the smallest eigenvalue has relative condition number relcond $= 1 + O(|\tau - \lambda_j|)$. Each relcond is a rational function of τ. We identify the poles and then use orthogonal polynomial theory to develop upper bounds on the sum of the relconds of *all* the eigenvalues. These bounds require $O(n)$ operations for an $n \times n$ matrix. We show that the sum of all the relconds is bounded by κ trace $(L|D|L^t)$ and conjecture that $\kappa < n/\|LDL^t\|$. The quantity trace$(L|D|L^t)/\|LDL^t\|$ is a natural measure of element growth in the context of this paper.

An algorithm for computing numerically orthogonal eigenvectors without recourse to the Gram–Schmidt process is sketched. It requires that there exist values of τ close to each cluster of close eigenvalues such that all the relconds belonging to the cluster are modest (say $\leqslant 10$), the sensitivity of the other eigenvalues is not important. For this reason we develop $O(n)$ bounds on the sum of the relconds associated with a cluster. None of our bounds makes

* Corresponding author.
E-mail address: parlett@math.berkeley.edu (B.N. Parlett).

PII: S 0 0 2 4 - 3 7 9 5 (9 9) 0 0 2 6 2 - 1

reference to the nature of the distribution of the eigenvalues within a cluster which can be very complicated.

Keywords: Eigenvalue; Symmetric tridiagonal matrix

1. Discussion and summary

A real symmetric tridiagonal matrix T permits triangular factorization $T = L_+ D_+ L_+^{\mathrm{t}}$ provided that no proper leading principal submatrix of T is singular. The main goal of this paper is to show that the entries in L_+ and D_+ determine the very small eigenvalues of T to high *relative* accuracy except in a few easily recognized cases. This is in sharp contrast to eigenvalue dependence on the entries of T except for special classes such as scaled diagonally dominant T's [2]. An illustration and precise statement of some of our results are given at the end of this section but first it is proper to step back and explain why this recondite result in Perturbation Theory is of general interest.

Current methods for diagonalizing T use the QR algorithm for the eigenvalues and inverse iteration for the eigenvectors and have been considered very satisfactory. They require only $O(n^2)$ operations for T's of order n except for certain cases. The existence of such cases was first noted (by Dr. George Fann of Pacific Northwest National Laboratories) in the early 1990s. When T has a large cluster of, say, 100 or more eigenvalues all agreeing to 4 or more decimal places then the execution time dramatically increases. The cause is the $O(n^3)$ Gram–Schmidt process invoked to make sure that all computed eigenvectors associated with the cluster are orthogonal to working accuracy.

On the other hand the 'true' eigenvectors of T are orthogonal and so if we can approximate them very accurately (error angle $O(\varepsilon)$) then orthogonality to working precision follows automatically. In [3,5] we have shown how to compute, despite roundoff errors, an accurate approximation to λ's eigenvector under two conditions:

(i) λ has few ($\leqslant 3$) decimal digits in common with its neighbors;
(ii) λ is approximated to high *relative* accuracy (all bits but the last few must be correct).

In order to achieve (i), the origin must be shifted close to each cluster, i.e., one uses $T - \tau I$ instead of T. To achieve (ii), the shifted eigenvalues $\lambda_i - \tau$ in the cluster must be *defined* to high *relative* accuracy by $T - \tau I$. The trouble is that, in general, this is not the case. So one must either give up this approach or find a new representation of $T - \tau I$ that does define its very small eigenvalues to the desired accuracy. Our finding is that triangular factorization of $T - \tau I$ has the desired property except in rare situations that can be detected in $O(n)$ operations. We show that when there is little element growth then all eigenvalues are usually defined to high relative accuracy. Since τ may be chosen anywhere in a small interval on either side of a cluster there is a continuum of τ's that satisfy both (i) and (ii) for the whole cluster. See the illustration below.

In 1967, Kahan discovered a tricky proof that the Cholesky factors LL^{t} of a positive-definite T have the required property: small *relative* changes in the entries of L cause small *relative* changes in each eigenvalue of $LL^{\mathrm{t}} = T$ however small it may be. That is what is meant by saying that L *defines* the eigenvalues to high *relative* accuracy and that is the meaning of our title's phrase 'relatively robust representation' of T. Only in the late 1990s have simple explanations of Kahan's result been found.

Our task is to investigate the indefinite case. In Section 2, we introduce a condition number relcond ($\geqslant 1$) for each eigenvalue $\lambda_i - \tau$. In the definite case all relconds are unity. We give a variety of small indefinite examples and show that when τ is an eigenvalue then all possible twisted factorizations of $T - \tau I$ give the same value for relcond$(\lambda_i - \tau)$. That is why we stay with the familiar $L_+ D_+ L_+^{\mathrm{t}}$ and ignore $U_- D_- U_-^{\mathrm{t}}$. Here ends the motivation for our study.

An illustration: The matrix W_{21}^+ was introduced by Wilkinson [16] in the 1960s: diag $= (10, 9, 8, \ldots, 1, 0, 1, \ldots, 8, 9, 10)$, the next to diagonal entries are all 1. The eigenvalues are ordered $\lambda_1 < \lambda_2 < \cdots < \lambda_{21}$. Eigenvalues λ_{20} and λ_{21} are near 10.75 and differ by 10^{-13}, λ_{18} and λ_{19} are near 9.21 and differ by $2\delta = 5.6 \times 10^{-11}$, λ_{16} and λ_{17} are near 8.1 and differ by 10^{-9}. In Table 1, we exhibit some condition numbers, relcond (defined in Section 2), when the shift τ is close to $\{\lambda_{18}, \lambda_{19}\}$. The top row in Table 1 is the index of the unshifted eigenvalue. When $\tau = \lambda_{18}$ triangular factorization does not exist but nevertheless $\tau = \lambda_{18} - \delta$ gives an excellent representation. The relconds shown for $\lambda_{19} + \delta$ do not change as $\tau \longrightarrow \lambda_{19}$.

One of our results, a realistic bound on the relative condition numbers for an interior cluster, is given in Theorem 5, Section 8, but a crude corollary that establishes our claim may be quoted here.

> *Consider a cluster $\mathscr{C}$ of $\#(\mathscr{C})$ close eigenvalues with reasonable gaps on its left end* (*gap-left*) *and on its right end* (*gap-right*) *separating it from the rest of the spectrum of T. Let τ be chosen very close to, or at, the left end of $\mathscr{C}$, let $T - \tau I = L_+ D_+ L_+^{\mathrm{t}}$, let $\omega_k = \operatorname{sign} D_+(k,k)$ then*
>
> $$\sum_{\lambda_i \in \mathscr{C}} \text{relcond}(\lambda_i - \tau) \leqslant \#(\mathscr{C}) + \frac{2}{\text{gap-left}} \sum_{\omega_k = -1} \left(L_+ |D_+| L_+^{\mathrm{t}}\right)_{kk},$$
>
> *where $|D_+| = (D_+^2)^{1/2}$. The point is that neither $\#(\mathscr{C})$ nor tiny gaps inside the cluster influence the second term. There is a similar result for the right end of $\mathscr{C}$.*

Table 1
Condition numbers for selected λ_i

τ	1	16	17	18	19	20	21
$\lambda_{19} + \delta$	1.00	1.26	1.26	1.00	1.00	1.35	1.35
$\lambda_{18} - \delta$	1.87	1.26	1.26	1.51	2.54	1.35	1.35
$\lambda_{18} - \delta/100$	101.9	1.26	1.26	1.99	198.8	1.35	1.35

Our final result is a bound on $\sum_{\lambda_i \in \mathscr{C}} \text{relcond}(\lambda_i - \tau)$ but it makes no reference to element growth and is computable in $O(n)$ operations. It illustrates a mechanism by which the tiny eigenvalues can have small relconds while the large ones have huge relconds.

We also show that, if triangular factorization exists with $\tau = \lambda_j$, then, as $\tau \longrightarrow \lambda_j$,

$$\text{relcond}(\lambda_j - \tau) = 1 + O\left(|\lambda_j - \tau|\right).$$

The limit case, when $\tau = \lambda_j$, is well known; when $D_+(n,n) = 0$ then no *relative* perturbation can disturb its singularity nor make large relative changes to the eigenvector entries.

Let us sketch the sequential algorithm that is based on the results of this paper. Suppose that T is positive-definite. Compute the Cholesky factorization $LL^t = T$ and find all eigenvalues of LL^t to high relative accuracy. Next compute the eigenvectors for all $\lambda - \tau$ with large relative gaps by the method in [5]. If some eigenvalues remain without eigenvectors, then pick a new shift τ at, or close to, one end of the remaining spectrum. Perform a careful factorization $\bar{L}\bar{D}\bar{L}^t = LL^t - (\text{new } \tau)I$ using dqds algorithms described in [5] and monitor the bounds mentioned above. If necessary perturb τ (away from the cluster) until the bounds are acceptable. Then refine, to high relative accuracy, the shifted small eigenvalues with large relative gaps and compute their eigenvectors. Repeat the process with suitable shifts τ until all eigenvectors have been computed.

Our results do not provide easy reading but the analysis has been shortened significantly by invoking kernel polynomials and the Christoffel–Darboux identity. Thus, Sections 3 and 4 present background material that may not be familiar to some readers. Our analysis begins in Section 5, where we study the vector whose squared norm is a relcond. Section 6 is an important digression to prove a conjecture made by one of us in [3]. Section 7 shows clearly how the indefinite case differs from the definite one, see (38), and leads us to a conjecture that gives an elegant bound on the sum of all relconds in terms of element growth.

Our computable bounds for individual clusters, Theorems 5 and 6, are given in Section 8.

2. Relative condition numbers

Consider the eigenvector equation for any eigenvector $\boldsymbol{s}_m$ of T, $\|\boldsymbol{s}_m\| = 1$,

$$L_+ D_+ L_+^t \boldsymbol{s}_m = (T - \tau I)\boldsymbol{s}_m = \boldsymbol{s}_m(\lambda_m - \tau). \tag{1}$$

The eigenvalues have been shifted by τ and it is the robustness of these shifted values that is our concern here.

An attractive property of tridiagonals is that arbitrary *relative* perturbations to the $n-1$ parameters in L_+ and the n parameters in D_+ may be represented as

$$L_+ \longrightarrow EL_+E^{-1} \quad \text{and} \quad D_+ \longrightarrow D_+F^2$$

for appropriately chosen diagonal scaling matrices close to I. See [5] for details. The tridiagonal matrix changes from $L_+D_+L_+^{\mathrm{t}}$ to

$$EL_+E^{-1}FD_+FE^{-1}L_+^{\mathrm{t}}E.$$

Outer perturbations corresponding to E have been studied by several authors [6,7, 9–11,15] and are known to cause small *relative* changes in each eigenvalue. A preliminary study of inner perturbations, corresponding to $E^{-1}F$, was made by Dhillon [13], and in his thesis he has introduced a single condition number for inner perturbations. Let us write

$$E^{-1}F = I + \Delta, \quad \|\Delta\| \leqslant \eta. \tag{2}$$

He applies standard first-order additive perturbation theory to

$$L_+(I+\Delta)D_+(I+\Delta)L_+^{\mathrm{t}} = L_+D_+L_+^{\mathrm{t}} + 2L_+\Delta D_+L_+^{\mathrm{t}} + L_+\Delta^2 D_+L_+^{\mathrm{t}}.$$

The change $\delta\lambda_j$ to $\lambda_j - \tau$ is given by a Rayleigh quotient

$$\begin{aligned}\delta\lambda_j &= 2\boldsymbol{s}_j^{\mathrm{t}}L_+\Delta D_+L_+^{\mathrm{t}}\boldsymbol{s}_j + \mathrm{O}(\eta^2),\\ |\delta\lambda_j| &\leqslant 2\eta\boldsymbol{s}_j^{\mathrm{t}}L_+|D_+|L_+^{\mathrm{t}}\boldsymbol{s}_j + \mathrm{O}(\eta^2),\end{aligned}$$

since, by (2),

$$\left|\boldsymbol{v}^{\mathrm{t}}\Delta D_+\boldsymbol{v}\right| \leqslant \eta\boldsymbol{v}^{\mathrm{t}}|D_+|\boldsymbol{v} \quad \text{for all } \boldsymbol{v}.$$

So

$$\frac{|\delta\lambda_j|}{|\lambda_j-\tau|} \leqslant 2\eta\frac{\boldsymbol{s}_j^{\mathrm{t}}L_+|D_+|L_+^{\mathrm{t}}\boldsymbol{s}_j}{|\lambda_j-\tau|} + \mathrm{O}(\eta^2) = 2\eta\frac{\boldsymbol{s}_j^{\mathrm{t}}L_+|D_+|L_+^{\mathrm{t}}\boldsymbol{s}_j}{|\boldsymbol{s}_j^{\mathrm{t}}L_+D_+L_+^{\mathrm{t}}\boldsymbol{s}_j|} + \mathrm{O}(\eta^2).$$

Dhillon defines the condition number for $\lambda_j - \tau$ under small relative changes in the entries of L_+ and D_+ as

$$\mathrm{relcond}(\lambda_j-\tau) := \mathrm{relcond}(\lambda_j-\tau;\, L_+,\, D_+) := \frac{\left\| |D_+|^{1/2}L_+^{\mathrm{t}}\boldsymbol{s}_j\right\|^2}{|\lambda_j-\tau|}. \tag{3}$$

In (3) the explicit reference to τ reminds the reader that the shift is τ.

Our main interest is in values of τ close to or even equal to certain eigenvalues of T. Consequently, D_+ may be either ill-conditioned or singular and so we now derive an alternative expression for relcond which reveals that relcond is independent of D_+.

From (1) with $m \leftarrow j$,

$$D_+L_+^{\mathrm{t}}\boldsymbol{s}_j = L_+^{-1}\boldsymbol{s}_j(\lambda_j-\tau)$$

and from the expression for $\delta\lambda_j$ above

$$\begin{aligned}\delta\lambda_j &= 2\boldsymbol{s}_j^{\mathrm{t}}L_+\Delta D_+L_+^{\mathrm{t}}\boldsymbol{s}_j + \mathrm{O}(\eta^2)\\ &= 2\boldsymbol{s}_j^{\mathrm{t}}L_+\Delta L_+^{-1}\boldsymbol{s}_j(\lambda_j-\tau) + \mathrm{O}(\eta^2).\end{aligned}$$

For any positive-definite diagonal scaling matrix Γ

$$\delta\lambda_j = 2\boldsymbol{s}_j^{\mathrm{t}} L_+ \Gamma \Delta \Gamma^{-1} L_+^{-1} \boldsymbol{s}_j (\lambda_j - \tau)$$

so that, using (2)

$$\begin{aligned}\frac{|\delta\lambda_j|}{|\lambda_j - \tau|} &\leqslant 2 \left\| \boldsymbol{s}_j^{\mathrm{t}} L_+ \Gamma \right\| \cdot \|\Delta\| \left\| \Gamma^{-1} L_+^{-1} \boldsymbol{s}_j \right\| \\ &\leqslant 2\eta \left\| \boldsymbol{s}_j^{\mathrm{t}} L_+ \Gamma \right\| \cdot \left\| \Gamma^{-1} L_+^{-1} \boldsymbol{s}_j \right\| \\ &\leqslant 2\eta \|L_+\Gamma\| \cdot \|(L_+\Gamma)^{-1}\| \\ &= 2\eta \ \mathrm{cond}(L_+\Gamma).\end{aligned}$$

Thus,

$$\mathrm{relcond}(\lambda_j - \tau) = \left\| \boldsymbol{s}_j^{\mathrm{t}} L_+ \Gamma \right\| \cdot \left\| \Gamma^{-1} L_+^{-1} \boldsymbol{s}_j \right\| \tag{4}$$

$$\leqslant \min_{\Gamma} \ \mathrm{cond}(L_+\Gamma). \tag{5}$$

One of us has shown that $\mathrm{cond}(L_+)$ can be computed in a stable way (no overflows or underflows) in $\mathrm{O}(n)$ operations. See [4]. We say more about the best scaling matrix Γ in Section 2.2.

For the analysis in the remaining sections it is convenient to introduce an alternate notation. Define

$$L := L_+ |D_+|^{1/2}, \quad \Omega := \mathrm{sign}(D_+).$$

In the event that D_+ is singular, i.e., $(D_+)_{n,n} = 0$, we need a convention and choose $\Omega_{n,n} = 1$, $L_{n,n} = 0$. Thus, $\Omega^2 = I$. Now

$$T - \tau I = L_+ D_+ L_+^{\mathrm{t}} = L\Omega L^{\mathrm{t}} \tag{6}$$

and Ω will not be perturbed.

It is worth mentioning that there is an unsymmetric eigenvalue problem closely related to $L\Omega L^{\mathrm{t}} \boldsymbol{s}_m = \boldsymbol{s}_m(\lambda_m - \tau)$, namely,

$$\begin{aligned}\Omega L^{\mathrm{t}} L \left(\Omega L^{\mathrm{t}} \boldsymbol{s}_m\right) &= \left(\Omega L^{\mathrm{t}} \boldsymbol{s}_m\right)(\lambda_m - \tau), \\ \left(\boldsymbol{s}_m^{\mathrm{t}} L\right) \Omega L^{\mathrm{t}} L &= (\lambda_m - \tau)\boldsymbol{s}_m^{\mathrm{t}} L.\end{aligned}$$

Inner perturbations of $L\Omega L^{\mathrm{t}}$ become outer perturbations of $\Omega L^{\mathrm{t}} L$. Now the ordinary (absolute) condition number of $\lambda_m - \tau$ for $\Omega L^{\mathrm{t}} L$ equals the relative condition number given in (3):

$$\begin{aligned}\mathrm{relcond}(\lambda_m - \tau) &= \frac{\boldsymbol{s}_m^{\mathrm{t}} L L^{\mathrm{t}} \boldsymbol{s}_m}{|\boldsymbol{s}_m^{\mathrm{t}} L \Omega L^{\mathrm{t}} \boldsymbol{s}_m|} \\ &= \mathrm{secant} \angle \left(L^{\mathrm{t}} \boldsymbol{s}_m, \Omega L^{\mathrm{t}} \boldsymbol{s}_m\right) \\ &:= \mathrm{cond}\left(\lambda_m - \tau; \Omega L^{\mathrm{t}} L\right).\end{aligned}$$

In [16, Chapter 2], Wilkinson showed that it is impossible to have just one large condition number among the eigenvalues of an unsymmetric matrix and so the same is true for our relconds.

2.1. Examples

Here we give the reader a guide to our relative condition numbers by studying various examples.

Example 1. Consider the Toeplitz matrix

$$\begin{bmatrix} 2 & 1 & & \\ 1 & 2 & 1 & \\ & 1 & 2 & 1 \\ & & 1 & 2 \end{bmatrix}$$

with well-separated eigenvalues

$$\lambda_1 = 0.3820, \quad \lambda_2 = 1.3820, \quad \lambda_3 = 2.6180, \quad \lambda_4 = 3.6180.$$

Take $\tau = \lambda_3(1+\varepsilon)$, $\varepsilon \approx 2.2 \times 10^{-16}$, to form $T - \tau I = L\Omega L^{\mathrm{t}}$. Thus,

$$L = \begin{bmatrix} 0.786151 & & & \\ -1.272010 & 0.999999 & & \\ & 1.000000 & 1.272010 & \\ & & -0.786151 & 3.65002 \times 10^{-8} \end{bmatrix}$$

and

$$\Omega = \mathrm{diag}(-1, 1, -1, -1).$$

At first glance, we might fear that

$$\mathrm{relcond}(\lambda_3 - \tau) = \frac{\boldsymbol{s}_3^{\mathrm{t}} L L^{\mathrm{t}} \boldsymbol{s}_3}{|\lambda_3 - \tau|}$$

may be large since $|\lambda_3 - \tau| = 4.44 \times 10^{-16}$, L is nearly singular and the Cholesky-like bound (5) with $\Gamma = |D_+|^{1/2}$ gives

$$\mathrm{relcond}(\lambda_j - \tau) \leqslant \mathrm{cond}(L) = 9.01 \times 10^7, \quad 1 \leqslant j \leqslant 4.$$

A closer look at L shows that its rank is revealed by its $(4,4)$ element and thus by D_+ in the $L_+ D_+ L_+^{\mathrm{t}}$ decomposition. The bound (5) with $\Gamma = I$ gives

$$\mathrm{relcond}(\lambda_j - \tau) \leqslant \mathrm{cond}(L_+) = 6.975, \quad 1 \leqslant j \leqslant 4.$$

Despite the near singularity of L, $L\Omega L^{\mathrm{t}}$ determines all its eigenvalues to high relative accuracy. In fact, the relative condition numbers for all the eigenvalues $\lambda_j - \tau$ are 1.00, 1.89, 1.00 and 1.89, respectively.

Example 2. When there are close eigenvalues, the choice of τ can be critical in getting a relatively robust representation. For example, consider the 21×21 Wilkinson matrix W_{21}^+, which has several pairs of close eigenvalues. The pair $(\lambda_{14}, \lambda_{15})$ is near 7 and has separation $\lambda_{15} - \lambda_{14} = 4.1 \times 10^{-7}$. Consider two factorizations, one with $\tau_1 = \lambda_{14} - \varepsilon$ and the other with $\tau_2 = \lambda_{15} + \varepsilon$ ($\varepsilon \approx 2 \times 10^{-16}$):

$$W_{21}^+ - \tau_1 I = L_{14}\Omega_{14}L_{14}^{\mathrm{t}}, \qquad W_{21}^+ - \tau_2 I = L_{15}\Omega_{15}L_{15}^{\mathrm{t}}.$$

There is a large element growth in L_{14} ($\|L_{14}\|^2 = 1.9 \times 10^9$), whereas there is no element growth in forming L_{15}. The large element growth leads to some large relative condition numbers:

$$\begin{aligned}&\mathrm{relcond}(\lambda_{15} - \tau_1; L_{14}) = 2.8 \times 10^8,\\ &\mathrm{relcond}(\lambda_1 - \tau_1; L_{14}) = 1.4 \times 10^8.\end{aligned}$$

Note that here $\lambda_{15} - \tau_1 = 4.1 \times 10^{-7}$ whereas $\lambda_1 - \tau_1 = -8.129$. Due to the element growth eigenvalues as large as -8.12 are not determined to relative or absolute accuracy (with respect to $\|W_{21}^+\|$) by L_{14}. Similarly, $\lambda_3 - \tau_1$, $\lambda_5 - \tau_1$, $\lambda_7 - \tau_1$, $\lambda_9 - \tau_1$, $\lambda_{11} - \tau_1$, and $\lambda_{13} - \tau_1$ have large relconds. Somewhat surprisingly, the smallest eigenvalue $\lambda_{14} - \tau_1$ is determined to high relative accuracy with

$$\mathrm{relcond}(\lambda_{14} - \tau_1; L_{14}) = 2.15.$$

See Section 6 for more on the relcond of the smallest eigenvalue as $\tau \longrightarrow \lambda_j$. On the other hand, near λ_{15} there is no element growth and *all* eigenvalues of $L_{15}\Omega L_{15}^{\mathrm{t}}$ are relatively robust. In particular,

$$\begin{aligned}&\mathrm{relcond}(\lambda_{14} - \tau_2; L_{15}) = 1.0,\\ &\mathrm{relcond}(\lambda_{15} - \tau_2; L_{15}) = 1.0\end{aligned}$$

and the largest relcond is less than 2.1.

Example 3. In all examples we have tried, absence of element growth in the triangular factorization has given relative robustness, see Section 7. However, the converse is not always true. Consider the tridiagonal [3, p. 114]

$$T = \begin{bmatrix} & \sqrt{2}/2 & & \sqrt{2}/2 & & \eta & \\ 1+\eta & & 1-3\eta & & 1+3\eta & & 1+2\eta \\ & \sqrt{2}/2 & & \sqrt{2}/2 & & \eta & \end{bmatrix}.$$

With $\varepsilon \approx 2.2 \times 10^{-16}$ and $\eta = \sqrt{\varepsilon}$ the eigenvalues are $\lambda_1 \approx \varepsilon$, $\lambda_2 \approx 1 + \sqrt{\varepsilon}$, $\lambda_3 \approx 1 + 2\sqrt{\varepsilon}$, $\lambda_4 \approx 2.0$. Forming $T - I = L\Omega L^{\mathrm{t}}$ gives

$$L = \begin{bmatrix} 1.057 \times 10^{-4} & & & \\ 6.688 \times 10^3 & 6.688 \times 10^3 & & \\ & -1.057 \times 10^{-4} & 2.114 \times 10^{-4} & \\ & & 4.983 \times 10^{-5} & 1.409 \times 10^{-4} \end{bmatrix}$$

with

$$\Omega = \mathrm{diag}(1, -1, 1, 1).$$

Now, $\|L\|^2 = 8.95 \times 10^7$ and $\mathrm{cond}(L) = 1.37 \times 10^8$. The rather large element growth suggests that the eigenvalues $\{\lambda_j - \tau\}$ may not be relatively robust. Indeed

$$\mathrm{relcond}(\lambda_1 - 1) \approx \mathrm{relcond}(\lambda_4 - 1) \approx 4.5 \times 10^7.$$

However, the two *smallest* eigenvalues $\lambda_2 - 1$ and $\lambda_3 - 1$ are relatively well conditioned:

$$\mathrm{relcond}(\lambda_2 - 1) = 1.666, \quad \mathrm{relcond}(\lambda_3 - 1) = 2.333. \tag{7}$$

Here we have the remarkable situation in which the large eigenvalues are not relatively robust, while the small eigenvalues are determined to high relative accuracy. The nice relconds in (7) *are* explained by small second components in the eigenvectors $\boldsymbol{s}_2$ and $\boldsymbol{s}_3$ – both $s_2(2)$ and $s_3(2)$ are $\mathrm{O}(10^{-8})$ and neutralize the large elements in the second column of L when forming $L^{\mathrm{t}}\boldsymbol{s}$, see (3).

Example 4. There are cases where no eigenvalue is relatively robust. For example,

$$L = \begin{bmatrix} 0.7451 & & & \\ -0.6967 & 2.01 \times 10^{-7} & & \\ & 1.81 \times 10^6 & 1.81 \times 10^6 & \\ & & 1.51 \times 10^{-14} & 0.2744 \end{bmatrix}$$

with

$$\Omega = \mathrm{diag}(-1, 1, -1, -1).$$

Eigenvalues of $L\Omega L^{\mathrm{t}}$ are

$$-1.075, \quad -0.075, \quad -0.075, \quad 0.924$$

with relative condition numbers

$$1.1 \times 10^{11}, \quad 6.4 \times 10^{12}, \quad 6.8 \times 10^7, \quad 6.5 \times 10^{12}.$$

In our primary application (computing orthogonal eigenvectors), we have no interest in the above situation where $\|L\|$ is large and no eigenvalue of $L\Omega L^{\mathrm{t}}$ is small (like ε). On the contrary, we must choose τ so that $L\Omega L^{\mathrm{t}}$ is nearly singular.

2.2. Twisted factorizations

If $T - \tau I$ permits triangular factorization in both directions, from top to bottom and from bottom to top, then

$$T - \tau I = L_+ D_+ L_+^{\mathrm{t}} = U_- D_- U_-^{\mathrm{t}}.$$

It is an interesting property of tridiagonal matrices that from these two representations one can create a one parameter family of (twisted) factorizations $\bar{N}_k \bar{D}_k \bar{N}_k^{\mathrm{t}}$ with essentially no extra work. Using Matlab notation,

$$\bar{N}_k(1{:}k, 1{:}k) = L_+(1{:}k, 1{:}k),$$
$$\bar{N}_k(k{:}n, k{:}n) = U_-(k{:}n, k{:}n)$$

and these equations are consistent because

$$L_+(k,k) = U_-(k,k) = 1.$$

Finally,

$$\bar{D}_k(i,i) = \begin{cases} D_+(i,i), & i < k, \\ D_-(i,i), & i > k, \\ \gamma_k, & i = k. \end{cases}$$

There are various formulas for γ_k. The most symmetrical is

$$\gamma_k = D_+(k) + D_-(k) - (T(k,k) - \tau).$$

We say that k is the twist index.

For theoretical purposes it is convenient to define $N_k = \bar{N}_k|\bar{D}_k|^{1/2}$ and $\Omega_k = \operatorname{sign}(\bar{D}_k)$.

At first sight the existence of these extra factorizations seems to complicate the search for relatively robust representations. For each shift τ we must consider the best among the twisted factorizations. The following surprising result eases the situation significantly.

Theorem 1. *Let T be an unreduced symmetric tridiagonal matrix with eigenpairs $(\lambda_j, \boldsymbol{s}_j)$, $j = 1, \ldots, n$. If, and only if, $\boldsymbol{s}_j$ has no zero entries, then $T - \lambda_j I$ permits a twisted factorization $T - \lambda_j I = N_k \Omega_k N_k^{\mathrm{t}}$ for each $k = 1, \ldots, n$ and* relcond$(\lambda_m - \lambda_j; N_k)$, $m \neq j$, *is the same for all k.*

Proof. By the convention introduced for formula (6), $\Omega_k(k,k) = 1$ and $N_k(k,k) = 0$. If $\boldsymbol{e}_k$ denotes the kth column of I then, because of the twist,

$$N_k \boldsymbol{e}_k = \boldsymbol{e}_k \cdot 0.$$

The existence of the twisted factorizations is an immediate consequence of well-known formulae for L_+, U_-, etc. From (12) and (16) with $\beta_i = T(i+1, i)$:

$$D_+(i,i) = -\beta_i \frac{s_j(i+1)}{s_j(i)}, \quad i < n,$$
$$D_+(n,n) = 0,$$
$$D_-(i,i) = -\beta_{i-1} \frac{s_j(i-1)}{s_j(i)}, \quad i > 1,$$
$$D_-(1,1) = 0.$$

If no entry of $\boldsymbol{s}_j$ vanishes then both sets of pivots are nonzero until the end and L_+, D_+, U_-, D_- are well defined and $\gamma_k = 0$.

The claim for the relconds holds because all the twisted factors N_k in $T - \lambda_j I = N_k \Omega_k N_k^{\mathrm{t}}$ are closely related. Write $L = N_n$ by columns as

$$L = (\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \ldots, \boldsymbol{\ell}_{n-1}, \boldsymbol{o})$$

and

$$T - \lambda_j I = L\Omega L^{\mathrm{t}} = \sum_{i=1}^{n-1} \boldsymbol{\ell}_i \omega_i \boldsymbol{\ell}_i^{\mathrm{t}} + 0, \quad \omega_i = \Omega(i,i).$$

Recall that $\boldsymbol{\ell}_i$ is null except in positions i and $i+1$. The crucial step in the proof is to push columns k through $n-1$ of L to the right for each $k = 1, \ldots, n-1$ to get

$$N_k = (\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_{k-1}, \boldsymbol{o}, \boldsymbol{\ell}_k, \ldots, \boldsymbol{\ell}_{n-1}),$$

$$\Omega_k = (\omega_1, \ldots, \omega_{k-1}, 1, \omega_k, \ldots, \omega_{n-1}).$$

Note that N_k has its twist at index k. Thus,

$$N_k \Omega_k N_k^{\mathrm{t}} = \sum_{i=1}^{n-1} \boldsymbol{\ell}_i \omega_i \boldsymbol{\ell}_i^{\mathrm{t}} + 0 = T - \lambda_j I.$$

So, by the analogue of (3),

$$\begin{aligned} \mathrm{relcond}(\lambda_m - \lambda_j; N_k) &= \frac{\left\| N_k^{\mathrm{t}} \boldsymbol{s}_m \right\|^2}{|\lambda_m - \lambda_j|} \\ &= \frac{\sum_{i=1}^{n-1} \left(\boldsymbol{\ell}_i^{\mathrm{t}} \boldsymbol{s}_m \right)^2}{|\lambda_m - \lambda_j|} \end{aligned}$$

and the right-hand side is independent of k. □

Since our interest is in values of τ very close to or at eigenvalues we conclude that we are not going to miss a good representation by staying with $L_+ D_+ L_+^{\mathrm{t}}$.

However, the twists are relevant to obtaining good bounds. Replace $L_+ D_+ L_+^{\mathrm{t}}$ by $\bar{N}_k \bar{D}_k \bar{N}_k^{\mathrm{t}}$ in (4) and set $\Gamma = I$ to see that

$$\begin{aligned} \mathrm{relcond}(\lambda_m - \lambda_j) &= \left\| \boldsymbol{s}_m^{\mathrm{t}} \bar{N}_k \right\| \left\| \bar{N}_k^{-1} \boldsymbol{s}_m \right\| \\ &\leqslant \mathrm{cond}\left(\bar{N}_k \right). \end{aligned}$$

We conjecture that if k is chosen so that $|s_j(k)| = \|\boldsymbol{s}_j\|_\infty$, then $\mathrm{cond}(\bar{N}_k)$ is close to $\min \mathrm{cond}(L_+\Gamma)$ over all scaling matrices Γ. Given L_+, D_+, U_-, D_- a suitable k may be found in $O(n)$ operations [12]. Theorem 1 justifies the notation $\mathrm{relcond}(\lambda_m - \lambda_j)$ to replace $\mathrm{relcond}(\lambda_m - \lambda_j; \bar{N}_k, \bar{D}_k)$.

3. Associated orthogonal polynomials

The material in this section is essential to our analysis.

Consider triangular factorization as a function of a real parameter τ. If τ is not an eigenvalue of a proper leading principal submatrix of T, then

$$T - \tau I = L_+ D_+ L_+^{\mathrm{t}},$$

where L_+ is lower bidiagonal and D_+ is diagonal. The symbol $+$ indicates that the elimination is made with increasing indices. It is convenient to write the factorization in an unconventional way that ought to be called the Cholesky factorization of $T - \tau I$, namely,

$$T - \tau I = L \Omega L^{\mathrm{t}}, \tag{8}$$

where

$$L = L_+ |D_+|^{1/2}, \quad \Omega = \operatorname{sign}(D_+). \tag{9}$$

The dependence of L and Ω on τ is suppressed. By this change of representation we confine our concern with relative changes to the entries of one matrix L instead of two matrices L_+ and D_+.

Let $p_0(\tau) = 1$ and define the vector $\boldsymbol{p} = \boldsymbol{p}^{1:n}(\tau) = (p_0(\tau), p_1(\tau), \ldots, p_{n-1}(\tau))^{\mathrm{t}}$ by

$$(T - \tau I)\boldsymbol{p} = -\boldsymbol{e}_n p_n(\tau), \tag{10}$$

where $\boldsymbol{e}_j$ denotes column j of I. Let $\alpha_i = T(i,i)$, $\beta_i = T(i, i+1) > 0$. Apply (10) for $j = 1, 2, \ldots, n-1$, to find

$$p_1(\tau) = \frac{\tau - \alpha_1}{\beta_1},$$
$$p_2(\tau) = \frac{(\alpha_2 - \tau) p_1 + \beta_1 p_0}{-\beta_2} = \frac{\det[T_2 - \tau I_2]}{\beta_1 \beta_2},$$

where $T_i = T(1{:}i, 1{:}i)$. Hence, by induction, for $k < n$,

$$p_k(\tau) = (-1)^k \frac{\det[T_k - \tau I_k]}{\beta_1 \beta_2 \cdots \beta_k} \tag{11}$$

and (11) holds for $k = n$ as well if $\beta_n := 1$. We shall see that these polynomials p_i are intimately related to the matrix L in (9). The leading coefficient of p_j is $1/(\beta_1 \beta_2 \cdots \beta_j) > 0$, $j < n$, while that of p_n is $1/(\beta_1 \beta_2 \cdots \beta_{n-1}) > 0$. Note that when $\tau = \lambda_j$ then $p_n(\tau) = 0$ and the normalized eigenvector $\boldsymbol{s}_j$ satisfies

$$p_{k-1}(\lambda_j) = \frac{s_j(k)}{s_j(1)}. \tag{12}$$

Following Matlab notation let $\boldsymbol{v}(1{:}k)$ denote the subvector of $\boldsymbol{v}$ in positions $1, 2, \ldots, k$. If $p_k(\tau) = 0$, then $\boldsymbol{p}(1:k)$ is an eigenvector of the leading principal submatrix T_k. In general,

$$(T_k - \tau I_k)\boldsymbol{p}(1{:}k) = -\boldsymbol{e}_k p_k(\tau)\beta_k. \tag{13}$$

For future reference we note that, for $k < n$,

$$(T - \tau I) \begin{pmatrix} \boldsymbol{p}(1{:}k) \\ \boldsymbol{o} \end{pmatrix} = (-\boldsymbol{e}_k p_k(\tau) + \boldsymbol{e}_{k+1} p_{k-1}(\tau))\, \beta_k. \tag{14}$$

The eigenvector matrix S of T is defined by $S(k,i) = s_i(k)$ and the orthogonality of rows k and m of S yields, by (12),

$$0 = \sum_{i=1}^{n} S(k,i)S(m,i) = \sum_{i=1}^{n} s_i(1)^2 p_{k-1}(\lambda_i) p_{m-1}(\lambda_i), \quad k \neq m.$$

The $\{p_i\}$ are not just orthogonal but form an orthonormal system for the inner product on the space of polynomials of degree less than n given by

$$\langle \varphi, \psi \rangle := \sum_{i=1}^{n} s_i(1)^2 \varphi(\lambda_i)\psi(\lambda_i). \tag{15}$$

In what follows the expression $p_i(\tau)$ will often be abbreviated by p_i. From (11)

$$\begin{aligned} d_k(\tau) := D_+(k) &= \frac{\det[T_k - \tau I_k]}{\det[T_{k-1} - \tau I_{k-1}]} \\ &= -\beta_k \frac{p_k}{p_{k-1}} \end{aligned} \tag{16}$$

and from (9) the entries of L are given by:

$$l_{kk} = |d_k|^{1/2} = \left| \beta_k \frac{p_k}{p_{k-1}} \right|^{1/2}, \tag{17}$$

$$l_{k+1,k} = \frac{|d_k|^{1/2}\beta_k}{d_k} = \omega_k \left| \beta_k \frac{p_{k-1}}{p_k} \right|^{1/2}, \tag{18}$$

$$\omega_k = \operatorname{sign}(d_k),$$

so that

$$L\boldsymbol{e}_k = \sqrt{\beta_k} \left(\boldsymbol{e}_k \left| \frac{p_k}{p_{k-1}} \right|^{1/2} + \boldsymbol{e}_{k+1}\omega_k \left| \frac{p_{k-1}}{p_k} \right|^{1/2} \right),$$

$$\|L\boldsymbol{e}_k\|^2 = \beta_k \left(\frac{p_{k-1}^2 + p_k^2}{|p_{k-1}p_k|} \right) \geqslant 2\beta_k. \tag{19}$$

Expressions (15)–(19) are used in subsequent sections.

4. Kernel polynomials

Our results have been simplified by the Christoffel–Darboux formula (20) that we now derive.

For a vector $\boldsymbol{v}$ let $\boldsymbol{v}(i:j)$ denote the subvector of $\boldsymbol{v}$ having entries i through j. We continue to abbreviate $p_i(\tau)$ by p_i.

Premultiply (14) by $\boldsymbol{s}_m^{\mathrm{t}}$ to find

$$\boldsymbol{s}_m^{\mathrm{t}}(T-\tau I)\begin{pmatrix}\boldsymbol{p}(1:k)\\ \boldsymbol{o}\end{pmatrix}=\beta_k[-s_m(k)p_k+s_m(k+1)p_{k-1}].$$

On the other hand $\boldsymbol{s}_m^{\mathrm{t}}(T-\tau I)=(\lambda_m-\tau)\boldsymbol{s}_m^{\mathrm{t}}$, so

$$\boldsymbol{s}_m^{\mathrm{t}}(T-\lambda_j I)\begin{pmatrix}\boldsymbol{p}(1:k)\\ \boldsymbol{o}\end{pmatrix}=(\lambda_m-\tau)[\boldsymbol{s}_m(1:k)^{\mathrm{t}}\boldsymbol{p}(1:k)].$$

Equate the two expressions on the right and divide by $s_m(1)$ to obtain a remarkable formula,

$$(\lambda_m-\tau)\sum_{i=0}^{k-1}p_i(\tau)p_i(\lambda_m)$$
$$=\beta_k\det\begin{bmatrix}p_{k-1} & p_{k-1}(\lambda_m)\\ p_k & p_k(\lambda_m)\end{bmatrix},\quad k=1,2,\ldots,n-1. \tag{20}$$

Formula (20) is the Christoffel–Darboux relation for orthogonal polynomials. See [1, Chapter 1, Theorem 4.5].

Following standard notation, for fixed τ and variable ξ define the polynomials,

$$K_j(\xi,\tau):=\sum_{i=0}^{j}p_i(\xi)p_i(\tau). \tag{21}$$

This function is called the reproducing kernel. For the space of polynomials φ of degree not exceeding j endowed with the inner product given in (15), K_j plays the role of the Dirac delta function,

$$\langle K_j(\cdot,\tau),\varphi(\cdot)\rangle=\varphi(\tau). \tag{22}$$

In particular

$$\langle K_j(\cdot,\tau),K_j(\cdot,\tau)\rangle=K_j(\tau,\tau)=\|\boldsymbol{p}(1:j+1)\|^2. \tag{23}$$

It is known [1, Chapter 1] that $\varphi=K_j(\cdot,\tau)$ minimizes $\langle\varphi,\varphi\rangle$ over all polynomials of degree $\leqslant j$ that satisfy

$$\varphi(\tau)=K_j(\tau,\tau).$$

The zeros of $K_j(\cdot,\tau)$ interlace those of p_j and p_{j-1} in a special way.

In terms of K_j the Christoffel–Darboux relation becomes

$$K_{k-1}(\tau,\lambda_m)=\beta_k\frac{\det\begin{bmatrix}p_{k-1} & p_{k-1}(\lambda_m)\\ p_k & p_k(\lambda_m)\end{bmatrix}}{\lambda_m-\tau} \tag{24}$$

and this is an identity in λ_m. Let $\lambda_m\longrightarrow\tau$, to find

$$K_{k-1}(\tau,\tau)=\beta_k\det\begin{bmatrix}p_{k-1} & p'_{k-1}\\ p_k & p'_k\end{bmatrix}.$$

5. Expressions for $L^t s$

Since relcond$(\lambda - \tau)$ depends on $\|L^t s\|^2$ we develop expressions for $L^t s$ to be used in later sections.

First we use the pivots $d_k(\tau)$ defined in Section 3. Recall that $\omega_k = \text{sign}(d_k)$, $d_k = D_+(k,k)$.

Theorem 2. *Let $T - \tau I = L\Omega L^t$ exist and let (λ, s) be any eigenpair of T. Then*

$$\begin{aligned}
\left(L^t s\right)(1) &= \frac{\omega_1(\lambda-\tau)s(1)}{|\alpha_1-\tau|^{1/2}},\\
\left(L^t s\right)(k) &= \omega_k s(k)\frac{d_k(\tau)-d_k(\lambda)}{|d_k(\tau)|^{1/2}}, \quad 1 < k < n,\\
\left(L^t s\right)(n) &= s(n)|d_n(\tau)|^{1/2}.
\end{aligned}$$

Proof. By (12), $s(k+1) = p_k(\lambda)s(1)$. Thus, for $k = 1, \ldots, n-1$, use (17) and (18) to find

$$\begin{aligned}
(L^t s)(k) &= l_{kk}s(k) + l_{k+1,k}s(k+1)\\
&= |d_k(\tau)|^{1/2}s(k) + \left(\frac{\omega_k\beta_k}{|d_k(\tau)|^{1/2}}\right)s(k+1)\\
&= \frac{\omega_k s(k)}{|d_k(\tau)|^{1/2}}\left[d_k(\tau) + \beta_k\frac{p_k(\lambda)}{p_{k-1}(\lambda)}\right] \quad \text{(by (12))}\\
&= \frac{\omega_k s(k)}{|d_k(\tau)|^{1/2}}(d_k(\tau) - d_k(\lambda)) \quad \text{(using (16))}.
\end{aligned}$$

For $k = n$, $(L^t s)(n) = |d_n(\tau)|^{1/2}s(n)$ since L^t is upper bidiagonal. □

Corollary 1. *Let p_k denote $p_k(\tau)$. In terms of the polynomials $p_k(\xi)$, for $k < n$,*

$$\left(L^t s\right)(k) = \text{sign}(p_{k-1})s(1)|\beta_k p_k p_{k-1}|^{1/2}\left(\frac{p_{k-1}(\lambda)}{p_{k-1}} - \frac{p_k(\lambda)}{p_k}\right), \tag{25}$$

$$\left(L^t s\right)(n) = s(1)\left|\frac{p_n}{p_{n-1}}\right|^{1/2}\cdot p_{n-1}(\lambda).$$

Also

$$\left(L^t s\right)(n) = \text{sign}(p_{n-1}(\lambda))\left|\frac{p_{n-1}(\lambda)}{p_n'(\lambda)}\cdot\frac{p_n}{p_{n-1}}\right|^{1/2}.$$

Proof. Use (16) to rewrite Theorem 2. For the case $k = n$ use

$$s(n)^2 = p_{n-1}(\lambda)/p_n'(\lambda) \tag{26}$$

from [14, Corollary 7.9.1]. □

The kernel polynomials let us rewrite Corollary 1 in a convenient form.

Theorem 3. *Let $T - \tau I = L\Omega L^{\mathrm{t}}$ and $T\boldsymbol{s} = \boldsymbol{s}\lambda$. In terms of the vector $\boldsymbol{p} = \boldsymbol{p}(\tau)$ defined in* (10)*, for $k < n$,*

$$\left(L^{\mathrm{t}}\boldsymbol{s}\right)(k) = \frac{-\operatorname{sign}(p_k)}{|\beta_k p_k p_{k-1}|^{1/2}} s(1) K_{k-1}(\tau, \lambda)(\lambda - \tau).$$

The result holds for $k = n$ if β_n is defined as 1.

Proof. Recall the Christoffel–Darboux identity

$$\beta_k \det\begin{bmatrix} p_{k-1} & p_{k-1}(\lambda) \\ p_k & p_k(\lambda) \end{bmatrix} = (\lambda - \tau) K_{k-1}(\tau, \lambda).$$

Expand (25) from Corollary 1 to find, for $k < n$,

$$\begin{aligned}(L^{\mathrm{t}}\boldsymbol{s})(k) &= \operatorname{sign}(p_{k-1}) s(1) \frac{|\beta_k p_k p_{k-1}|^{1/2}}{p_k p_{k-1}} \det\begin{bmatrix} p_{k-1}(\lambda) & p_{k-1} \\ p_k(\lambda) & p_k \end{bmatrix} \\ &= \frac{\operatorname{sign}(p_k) s(1) \beta_k}{|\beta_k p_k p_{k-1}|^{1/2}} \det\begin{bmatrix} p_{k-1}(\lambda) & p_{k-1} \\ p_k(\lambda) & p_k \end{bmatrix} \\ &= \frac{-\operatorname{sign}(p_k) s(1)}{|\beta_k p_k p_{k-1}|^{1/2}} K_{k-1}(\tau, \lambda)(\lambda - \tau).\end{aligned}$$

For $k = n$ use Corollary 1 and $\beta_n = 1$ and note that $p_n(\lambda) = 0$. □

Finally, take the product of Theorem 3 and Corollary 1 to find a third representation. From (16), $-\operatorname{sign}(p_k \cdot p_{k-1}) = \omega_k$:

$$\omega_k \frac{\left(L^{\mathrm{t}}\boldsymbol{s}\right)(k)^2}{\lambda - \tau} = s(1)^2 \left(\frac{p_{k-1}(\lambda)}{p_{k-1}} - \frac{p_k(\lambda)}{p_k}\right) K_{k-1}(\tau, \lambda). \tag{27}$$

6. The case $\tau \longrightarrow \lambda_j$

This section shows that, as $\tau \longrightarrow \lambda_j$, $L^{\mathrm{t}}\boldsymbol{s}_j/|\lambda_j - \tau|^{1/2} \longrightarrow \boldsymbol{e}_n + \mathrm{O}(|\tau - \lambda_j|^{1/2})$. It follows that $\operatorname{relcond}(\lambda_j - \tau) = 1 + \mathrm{O}(|\tau - \lambda_j|)$ and thus proves Conjecture 2 in Section 5.2.3 of [3]. We exhibit the constant hidden by O.

Recall from (8) that

$$T - \tau I = L\Omega L^{\mathrm{t}}, \quad \Omega = \operatorname{diag}(\pm 1).$$

Let $(\lambda_j, \boldsymbol{s}_j)$ be an eigenpair of T such that $\boldsymbol{s}_j$ has no zero entries and $\|\boldsymbol{s}_j\| = 1$. Recall that

$$s_j(k) = s_j(1)p_{k-1}(\lambda_j).$$

Consider $L^t \boldsymbol{s}_j$ for τ close to λ_j. Take the last entry first. By Corollary 1,

$$\left(L^t \boldsymbol{s}_j\right)(n)^2 = \left|\frac{p_n(\tau)}{p_{n-1}(\tau)} \cdot \frac{p_{n-1}(\lambda_j)}{p_n'(\lambda_j)}\right|.$$

Define $p_n^{\langle j\rangle}(\tau)$ by $p_n(\tau) = (\tau - \lambda_j)p_n^{\langle j\rangle}(\tau)$ so that

$$\frac{\left(L^t \boldsymbol{s}_j\right)(n)^2}{|\tau - \lambda_j|} = \left|\frac{p_n^{\langle j\rangle}(\tau)}{p_n'(\lambda_j)}\right| \left|\frac{p_{n-1}(\lambda_j)}{p_{n-1}(\tau)}\right|. \tag{28}$$

By (11)

$$\begin{aligned} p_n^{\langle j\rangle}(\tau) &= \frac{\prod_{i\neq j}(\tau - \lambda_i)}{\prod_{i=1}^{n-1}\beta_i} \\ &= p_n^{\langle j\rangle}(\lambda_j) + (\tau - \lambda_j)p_n^{\langle j\rangle\,\prime}(\lambda_j) + \mathrm{O}((\tau - \lambda_j)^2). \end{aligned}$$

Also

$$\begin{aligned} p_n^{\langle j\rangle}(\lambda_j) &= p_n'(\lambda_j), \\ p_n^{\langle j\rangle\prime}(\tau) &= p_n^{\langle j\rangle}(\tau)\sum_{i\neq j}(\tau - \lambda_i)^{-1}, \\ p_{n-1}'(\tau) &= p_{n-1}(\tau)\sum_{i=1}^{n-1}(\tau - \theta_i)^{-1}, \end{aligned}$$

where $p_{n-1}(\theta_i) = 0$, $i = 1, \ldots, n-1$. Thus,

$$\frac{p_n^{\langle j\rangle}(\tau)}{p_n'(\lambda_j)} = 1 + (\tau - \lambda_j)\sum_{i\neq j}(\lambda_j - \lambda_i)^{-1} + \mathrm{O}((\tau - \lambda_j)^2),$$

$$\frac{p_{n-1}(\tau)}{p_{n-1}(\lambda_j)} = 1 + (\tau - \lambda_j)\sum_{i=1}^{n-1}(\lambda_j - \theta_i)^{-1} + \mathrm{O}\left((\tau - \lambda_j)^2\right)$$

and, as $\tau \longrightarrow \lambda_j$,

$$\begin{aligned} \frac{\left(L^t \boldsymbol{s}_j\right)(n)^2}{|\tau - \lambda_j|} &= 1 + (\tau - \lambda_j)\left[\sum_{i\neq j}(\lambda_j - \lambda_i)^{-1} - \sum_{i=1}^{n-1}(\lambda_j - \theta_i)^{-1}\right] \\ &\quad + \mathrm{O}\left((\tau - \lambda_j)^2\right), \end{aligned} \tag{29}$$

It remains to show that, for $k < n$, $(L^t \boldsymbol{s}_j)(k) = \mathrm{O}(\tau - \lambda_j)$.

By Theorem 3, for $k < n$,

$$\frac{|(L^t \boldsymbol{s}_j)(k)|}{|\tau - \lambda_j|} = \frac{s_j(1)|K_{k-1}(\tau, \lambda_j)|}{|\beta_k p_k p_{k-1}|^{1/2}}$$

$$\longrightarrow g_j(k) := \frac{s_j(1)|K_{k-1}(\lambda_j, \lambda_j)|}{|\beta_k p_k(\lambda_j) p_{k-1}(\lambda_j)|^{1/2}} \quad \text{as } \tau \to \lambda_j. \tag{30}$$

Let $g_j(n) = 0$ to complete the definition of g_j. Combine (29) and (30) to see that, as $\tau \to \lambda_j$,

$$\frac{L^t \boldsymbol{s}_j}{|\tau - \lambda_j|^{1/2}} = \boldsymbol{e}_n + |\tau - \lambda_j|^{1/2} \boldsymbol{g}_j + O(|\tau - \lambda_j|)$$

as claimed, and

$$\begin{aligned} &\text{relcond}(\lambda_j - \tau) \\ &\quad = 1 + |\tau - \lambda_j| \left[\|\boldsymbol{g}_j\|^2 + \sum_{i \neq j} (\lambda_j - \lambda_i)^{-1} - \sum_{i=1}^{n-1} (\lambda_j - \theta_i)^{-1} \right] \\ &\qquad + O\left(|\tau - \lambda_j|^{3/2} \right). \end{aligned} \tag{31}$$

It is useful to see how $(L^t \boldsymbol{s}_m)(n) \longrightarrow 0$ for $m \neq j$. By (28),

$$\frac{\left(L^t \boldsymbol{s}_m\right)(n)^2}{|\tau - \lambda_j|} = \left| \frac{p_n^{\langle j \rangle}(\tau)}{p_{n-1}(\tau)} \Big/ \frac{p_n'(\lambda_m)}{p_{n-1}(\lambda_m)} \right|$$

$$\longrightarrow \left| \frac{p_{n-1}(\lambda_m)}{p_n'(\lambda_m)} \Big/ \frac{p_{n-1}(\lambda_j)}{p_n'(\lambda_j)} \right| = \left(\frac{s_m(n)}{s_j(n)} \right)^2 \tag{32}$$

as $\tau \to \lambda_j$, using (26).

It is clear from (32) that the larger is $|s_j(n)|$ then the larger is the asymptotic region in which $(L^t \boldsymbol{s}_j)(n) \to 1$ and $(L^t \boldsymbol{s}_m)(n) \to 0$ as $\tau \to \lambda_j$. In Section 2.2, it was shown that for $\tau = \lambda_j$ all the twisted factorizations yield the same relconds. Nevertheless, for $\tau \approx \lambda_j$ some twisted factorizations will be more rank revealing than others. In particular twists at the location of maximal entries in $\boldsymbol{s}_j$ ensure that the critical diagonal entry $\bar{D}_k(k, k) = \gamma_k$ satisfies $|\gamma_k| \leqslant n|\tau - \lambda_j|$. See [12].

7. Summing the relconds

Now we employ the expressions in Sections 4 and 5 to obtain bounds on relcond $(\lambda_m - \tau)$ for all the λ_m's, not necessarily the one closest to τ which was discussed in Section 6. The natural fear is that the eigenvalues in a tight cluster will be highly sensitive to small changes in L. The matrix L is determined by the vector $\boldsymbol{p}(\tau)$ defined

in (10) and its approximations $\binom{\boldsymbol{p}^k}{\boldsymbol{o}}$ defined in (14), where $\boldsymbol{p}^k$ abbreviates $\boldsymbol{p}(1:k)$. See (17) and (18). Here are the pertinent relations. When the argument of a function is τ it will be omitted, $p_i = p_i(\tau)$.

The tridiagonal form of T shows, in (14), that for $k < n$,

$$(T - \tau I)\begin{pmatrix}\boldsymbol{p}^k\\ \boldsymbol{o}\end{pmatrix} = \beta_k(0,\ldots,0,-p_k,p_{k-1},0,\ldots,0)^{\mathrm{t}}, \tag{33}$$

$$\left\|(T - \tau I)\begin{pmatrix}\boldsymbol{p}^k\\ \boldsymbol{o}\end{pmatrix}\right\|^2 = \beta_k^2\left(p_{k-1}^2 + p_k^2\right). \tag{34}$$

It will be useful to express (34) in terms of the kernel functions $K_j(\sigma,\tau) := \sum_{i=0}^{j} p_i(\sigma)p_i(\tau)$ from Section 4. Rewrite the left-hand side of (33) using the spectral decomposition

$$\begin{aligned}(T - \tau I)\begin{pmatrix}\boldsymbol{p}^k\\ \boldsymbol{o}\end{pmatrix} &= S(\Lambda - \tau I)S^{\mathrm{t}}\begin{pmatrix}\boldsymbol{p}^k\\ \boldsymbol{o}\end{pmatrix}\\ &= \sum_{i=1}^{n}\boldsymbol{s}_i(\lambda_i - \tau)s_i(1)K_{k-1}(\lambda_i,\tau),\end{aligned} \tag{35}$$

since $s_i(j) = s_i(1)p_{j-1}(\lambda_i)$. Hence, by (33) and (35),

$$\begin{aligned}\begin{pmatrix}\boldsymbol{p}^k\\ \boldsymbol{o}\end{pmatrix}^{\mathrm{t}}(T - \tau I)\begin{pmatrix}\boldsymbol{p}^k\\ \boldsymbol{o}\end{pmatrix} &= -\beta_k p_{k-1}p_k\\ &= \sum_{i=1}^{n}(\lambda_i - \tau)(s_i(1)K_{k-1}(\lambda_i,\tau))^2.\end{aligned} \tag{36}$$

Now recall Theorem 3 in Section 5 and replace $\beta_k p_{k-1}p_k$ by (36). For $k < n$,

$$\begin{aligned}\frac{\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)(k)^2}{|\lambda_m - \tau|} &= \frac{|\lambda_m - \tau|}{|\beta_k p_{k-1}p_k|}(s_m(1)K_{k-1}(\lambda_m,\tau))^2\\ &= \frac{|\lambda_m - \tau|(s_m(1)K_{k-1}(\lambda_m,\tau))^2}{\left|\sum_{i=1}^{n}(\lambda_i - \tau)(s_i(1)K_{k-1}(\lambda_i,\tau))^2\right|}.\end{aligned} \tag{37}$$

Let $\tau \longrightarrow \lambda_m$ in (37) to recover (30) in Section 6.

To give meaning to (37) we sum over m, not k, to find

$$\begin{aligned}\sum_{m=1}^{n}\frac{\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)(k)^2}{\|\lambda_m - \tau|} &= \frac{\sum_{m=1}^{n}|\lambda_m - \tau|s_m(1)^2K_{k-1}(\lambda_m,\tau)^2}{\left|\sum_{i=1}^{n}(\lambda_i - \tau)s_i(1)^2K_{k-1}(\lambda_i,\tau)^2\right|}\\ &= \frac{P_k + N_k}{|P_k - N_k|},\end{aligned} \tag{38}$$

where

$$N_k := \sum_{\lambda_i < \tau} (\tau - \lambda_i) s_i(1)^2 K_{k-1}(\lambda_i, \tau)^2 > 0.$$

When $\tau \leqslant \lambda_1$, then, for each k, $N_k = 0$ and $(P_k + N_k)/(P_k - N_k) = 1$. When $\tau \geqslant \lambda_n$ then for each k, $P_k = 0$ and $(P_k + N_k)/(P_k - N_k) = -1$. That is why, interchanging the order of summation,

$$\sum_{m=1}^{n} \operatorname{relcond}(\lambda_m - \tau) = \sum_{k=1}^{n} \sum_{m=1}^{n} \frac{\left(L^{\mathrm{t}} \boldsymbol{s}_m\right)(k)^2}{|\lambda_m - \tau|} = \sum_{k=1}^{n} 1 = n,$$

in the definite case. In addition (38) shows that, in the indefinite case, if there is catastrophic cancellation between P_k and N_k, even for one k, then some relconds will be large. Now we analyze the indefinite case.

The denominator in (38) is $|\beta_k p_{k-1} p_k|$ and vanishes when, and only when, $p_{k-1}(\tau) p_k(\tau) = 0$ since T is assumed to be unreduced ($\beta_k > 0$). We doubt that there is a closed expression for the numerator in terms of $\boldsymbol{p}^{(k)}$ and we are forced to find a bound. To this end we define two quantities that measure how close $\binom{\boldsymbol{p}^k}{\boldsymbol{o}}$ is to an eigenvector of T. The first is a Rayleigh quotient:

$$\begin{aligned} \rho_k = \rho_k(\tau) &:= \frac{\begin{pmatrix} \boldsymbol{p}^k \\ \boldsymbol{o} \end{pmatrix}^{\mathrm{t}} (T - \tau I) \begin{pmatrix} \boldsymbol{p}^k \\ \boldsymbol{o} \end{pmatrix}}{\|\boldsymbol{p}^k\|^2} \\ &= \frac{-\beta_k p_{k-1} p_k}{\|\boldsymbol{p}^k\|^2} \quad \text{(by (36))}. \end{aligned} \tag{39}$$

The second is a normalized residual

$$\begin{aligned} \boldsymbol{r}_k = \boldsymbol{r}_k(\tau) &:= \frac{\left\| (T - \tau I) \begin{pmatrix} \boldsymbol{p}^k \\ \boldsymbol{o} \end{pmatrix} \right\|}{\left\|\boldsymbol{p}^k\right\|} \\ &= \frac{\beta_k \left(p_{k-1}^2 + p_k^2\right)^{1/2}}{\|\boldsymbol{p}^k\|} \quad \text{(by (34))}. \end{aligned} \tag{40}$$

These expressions remain valid for $k = n$ if we take $\beta_n = 1$ but we do not exploit this fact.

Both (39) and (40) are easily computed for all k, in $O(n)$ operations using the three-term recurrence for the $\{p_i(\tau)\}$. Note that

$$\rho_1 = T_{11} - \tau, \quad \|\boldsymbol{r}_1\| = \left((T_{11} - \tau)^2 + T_{21}\right)^{1/2},$$

$$\min_i \ |\lambda_i - \tau| \leqslant \|\boldsymbol{r}_k\| \leqslant \max_i \ |\lambda_i - \tau| = \|L \Omega L^{\mathrm{t}}\|.$$

A lengthy calculation shows that

$$\rho_k'(\tau) = \begin{cases} +1, & p_{k-1}(\tau) = 0, \\ -1, & p_k(\tau) = 0. \end{cases}$$

See Remark 4 at the end of Section 8.

Theorem 4. *Assume that* $\boldsymbol{p}^n = (p_0(\tau), \ldots, p_{n-1}(\tau))^{\mathrm{t}}$ *has no zero entries. Let* λ_j *be the eigenvalue of* T *closest to* τ. *The factorization* $T - \tau I = L\Omega L^{\mathrm{t}}$ *exists and, in terms of* $\boldsymbol{r}_k$ *and* ρ_k *defined above*,

$$\sum_{m=1}^{n} \operatorname{relcond}(\lambda_m - \tau) \leqslant 1 + \sum_{k=1}^{n-1} \frac{\|L\boldsymbol{e}_k\|^2}{\|\boldsymbol{r}_k\|} + \mathrm{O}(|\tau - \lambda_j|), \tag{41}$$

$$\sum_{m=1}^{n} \operatorname{relcond}(\lambda_m - \tau) \leqslant 1 + \|L\Omega L^{\mathrm{t}}\| \sum_{k=1}^{n-1} |\rho_k|^{-1} + \mathrm{O}(|\tau - \lambda_j|). \tag{42}$$

Proof. In Section 6 it was shown that, for the case $k = n$, $(L^{\mathrm{t}}\boldsymbol{s}_j)(n)^2/|\lambda_j - \tau| = 1 + \mathrm{O}(|\tau - \lambda_j|)$ and $(L^{\mathrm{t}}\boldsymbol{s}_m)(n)^2/|\lambda_m - \tau| = \mathrm{O}(|\tau - \lambda_j|)$. Hence,

$$\sum_{m=1}^{n} \frac{\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)(n)^2}{|\lambda_m - \tau|} = 1 + \mathrm{O}(|\tau - \lambda_j|).$$

For $k < n$ we begin from (38). The numerator may be majorized by the Cauchy–Schwartz inequality $(\sum w_i|\lambda_i - \tau|)^2 \leqslant \sum w_i(\lambda_i - \tau)^2 \sum w_i$. Use (34) and (35) to find

$$\begin{aligned}
\sum_{m=1}^{n} \frac{\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)(k)}{|\lambda_m - \tau|} &\leqslant \frac{\beta_k \left(p_{k-1}^2 + p_k^2\right)^{1/2} \left\|\boldsymbol{p}^k\right\|}{|\beta_k p_{k-1} p_k|} \\
&= \beta_k \left(\frac{p_{k-1}^2 + p_k^2}{|p_{k-1} p_k|}\right) \frac{\left\|\boldsymbol{p}^k\right\|}{\beta_k \left(p_{k-1}^2 + p_k^2\right)^{1/2}} \\
&= \frac{\|L\boldsymbol{e}_k\|^2}{\|\boldsymbol{r}_k\|} \quad \text{(by (19) and (40))}.
\end{aligned} \tag{43}$$

Recall that

$$\sum_{m=1}^{n} \operatorname{relcond}(\lambda_m - \tau) = \sum_{m=1}^{n} \sum_{k=1}^{n} \frac{\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)(k)^2}{|\lambda_m - \tau|}.$$

Reverse the order of summation and apply (43) to obtain (41). Instead of the Cauchy–Schwartz inequality we can take out of the numerator in (36) $\max_i |\lambda_i - \tau| = \|T - \tau I\| = \|L\Omega L^{\mathrm{t}}\|$ and obtain

$$\sum_{m=1}^{n} \frac{\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)(k)^2}{|\lambda_m - \tau|} \leqslant \frac{\left\|L\Omega L^{\mathrm{t}}\right\| \cdot \left\|\boldsymbol{p}^k\right\|^2}{|\beta_k p_{k-1} p_k|} = \frac{\left\|L\Omega L^{\mathrm{t}}\right\|}{|\rho_k|} \quad \text{(by (39))}. \tag{44}$$

Reverse the order of summation in $\sum_{m=1}^{n}\sum_{k=1}^{n}(L^{t}\boldsymbol{s}_m)(k)^2/|\lambda_m-\tau|$ and apply (44) to obtain (42). □

Remark 1. If $|p_{n-1}(\tau)|=\|\boldsymbol{p}^n\|_\infty$, then the denominators $\|\boldsymbol{r}_k\|$ in (41) cannot become arbitrarily small. Let

$$\theta_k := \angle\left(\boldsymbol{p}^n, \begin{pmatrix}\boldsymbol{p}^k\\ \boldsymbol{o}\end{pmatrix}\right).$$

Then

$$\cos\,\theta_k = \frac{\left\|\boldsymbol{p}^k\right\|^2}{\left\|\boldsymbol{p}^k\right\|\cdot\|\boldsymbol{p}^n\|} = \frac{\left\|\boldsymbol{p}^k\right\|}{\|\boldsymbol{p}^n\|} \leqslant \left(1-\frac{1}{n}\right)^{1/2},$$

$$|\sin\,\theta_k| \geqslant \frac{1}{\sqrt{n}}.$$

By the fundamental gap theorem [14, Chapter 11],

$$\frac{\mathrm{gap}(\tau)}{\sqrt{n}} \leqslant |\sin\,\theta_k|\,\mathrm{gap}(\tau) \leqslant \|\boldsymbol{r}_k\|, \tag{45}$$

where $\mathrm{gap}(\tau):=\min_{i\neq j}\,|\tau-\lambda_i|$, where λ_j is the eigenvalue closest to τ.

By Theorem 1 in Section 2.2, we can choose any twisted factorization of $(T-\tau I)$, when τ is an eigenvalue, without changing $\mathrm{relcond}(\lambda_i-\tau)$. If a largest entry in τ's eigenvector occurs in position k, then we may analyze relcond for the factorization with twist at k. The same lower bound (45) on the residual norms will hold in this case too.

Neither bound in Theorem 4 can be attained. So we now derive an exact expression for $\sum_{m=1}^{n}\mathrm{relcond}(\lambda_m-\tau)$ that displays the role of element growth. Recall (19), for $k<n$,

$$\begin{aligned}\|L\boldsymbol{e}_k\|^2 &= \beta_k\frac{p_{k-1}^2+p_k^2}{|p_{k-1}p_k|} = \frac{\beta_k^2\left(p_{k-1}^2+p_k^2\right)}{|\beta_k p_{k-1}p_k|}\\ &= \frac{\sum_{i=1}^{n}(\lambda_i-\tau)^2(s_i(1)K_{k-1}(\lambda_i,\tau))^2}{\sum_{i=1}^{n}(\lambda_i-\tau)(s_i(1)K_{k-1}(\lambda_i,\tau))^2}.\end{aligned}$$

Next multiply numerator and denominator of (38) by $\sum_{i=1}^{n}(\lambda_i-\tau)^2(s_i(1)K_{k-1}(\lambda_i,\tau))^2$ and rearrange to find

$$\sum_{m=1}^{n}\frac{\left(L^{t}\boldsymbol{s}_m\right)(k)^2}{|\lambda_m-\tau|} = \frac{\|L\boldsymbol{e}_k\|^2}{\pi_k(\tau)}, \tag{46}$$

where

$$\pi_k(\tau) := \frac{\sum_{i=1}^{n}|\lambda_i-\tau|\left(|\lambda_i-\tau|s_i(1)^2K_{k-1}(\lambda_i,\tau)^2\right)}{\sum_{i=1}^{n}|\lambda_i-\tau|s_i(1)^2K_{k-1}(\lambda_i,\tau)^2}.$$

For $k = n$ we have already seen that

$$\sum_{m=1}^{n} \frac{\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)(n)^2}{|\lambda_m - \tau|} = 1 + \mathrm{O}(|\tau - \lambda_j|),$$

where λ_j is the eigenvalue closest to τ. Again reversing the order of summation

$$\sum_{m=1}^{n} \mathrm{relcond}(\lambda_m - \tau) = \sum_{k=1}^{n-1} \frac{\|L\boldsymbol{e}_k\|^2}{\pi_k(\tau)} + 1 + \mathrm{O}(|\tau - \lambda_j|).$$

The ratios $\pi_k(\tau)$ seem difficult to analyze but on our test bed of examples $\pi_k(\tau) \geqslant \pi_1(\tau)$ for $2 \leqslant k \leqslant n$. We conjecture that

$$\sum_{m=1}^{n} \mathrm{relcond}(\lambda_m - \tau) \leqslant \frac{1}{\pi_1(\tau)} \operatorname{trace}\left(LL^{\mathrm{t}}\right) + 1 + \mathrm{O}(|\tau - \lambda_j|).$$

Consider the case when all the $|s_i(1)|$ are equal. Suppose that $\max|\lambda_i - \tau| = |\lambda_1 - \tau|$ and define $r_i = |\lambda_i - \tau|/|\lambda_1 - \tau|$, $i = 2, \ldots, n$. Then

$$\begin{aligned}
\pi_1(\tau) &= \frac{\sum_i (\lambda_i - \tau)^2}{\sum_i |\lambda_i - \tau|} \\
&= (\tau - \lambda_1)\frac{1 + r_2^2 + \cdots + r_n^2}{1 + r_2 + \cdots + r_n} \\
&> \frac{\tau - \lambda_1}{n} = \frac{\left\|L\Omega L^{\mathrm{t}}\right\|}{n}.
\end{aligned}$$

By Theorem 1 there is no need to consider an extreme case in which $|\tau - \lambda_j| \leqslant \varepsilon$ and $|s_j(1)| \approx 1$. We may assume that if λ_j is the closest eigenvalue to τ then $|s_j(1)| \leqslant 1/\sqrt{2}$. Finally, we conjecture that in all cases

$$\sum_{m=1}^{n} \mathrm{relcond}(\lambda_m - \tau) \leqslant n\, \frac{\operatorname{trace}\left(LL^{\mathrm{t}}\right)}{\|L\Omega L^{\mathrm{t}}\|}.$$

8. Bounds for an interior cluster

In Section 1, an algorithm was described for computing orthogonal eigenvectors of T. When a new shift is chosen and a new factorization performed the only new eigenvectors to be computed are those with large relative gaps. In other words, eigenvectors for eigenvalues close to the shift. Consequently, it is desirable to have bounds on $\mathrm{relcond}(\lambda_i - \tau)$ just for cluster of eigenvalues λ_i close to τ. We now derive $\mathrm{O}(n)$ computable bounds for such cases.

From the eigenvector equation

$$(T - \tau I)\boldsymbol{s}_m = L\Omega L^{\mathrm{t}}\boldsymbol{s}_m = \boldsymbol{s}_m(\lambda_m - \tau)$$

it follows that, for $\lambda_m \neq \tau$,

$$1 = \frac{\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)^{\mathrm{t}}\Omega\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)}{\lambda_m - \tau} = \sum_{k=1}^{n} \omega_k \frac{\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)(k)^2}{\lambda_m - \tau},$$

where $\omega_k = \mathrm{sign}(-\beta_k p_{k-1} p_k) = \mathrm{sign}(d_k)$ for the pivots d_k, see (16). Consequently for $\lambda_m > \tau$

$$\begin{aligned}\mathrm{relcond}(\lambda_m - \tau) &:= \sum_{k=1}^{n} \frac{\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)(k)^2}{|\lambda_m - \tau|} \\ &= \sum_{\omega_k=+1} \frac{\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)(k)^2}{|\lambda_m - \tau|} + \sum_{\omega_k=-1} \frac{\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)(k)^2}{|\lambda_m - \tau|} \\ &= 1 + 2\sum_{\omega_k=-1} \frac{\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)(k)^2}{|\lambda_m - \tau|}. \end{aligned} \tag{47}$$

The representation (47) lets us focus on a subset of indices k. From (36) in the previous section

$$\begin{aligned}-\beta_k p_{k-1} p_k &= \sum_{\tau \leqslant \lambda_i} (\lambda_i - \tau) s_i(1)^2 K_{k-1}(\lambda_i, \tau)^2 \\ &\quad - \sum_{\lambda_i < \tau} (\tau - \lambda_i) s_i(1)^2 K_{k-1}(\lambda_i, \tau)^2 \\ &= P_k - N_k \quad \text{(as in (38))}.\end{aligned}$$

So the cases $\omega_k = -1$ are characterized by $N_k > P_k$ or $2N_k > P_k + N_k$; so

$$\sum_{\lambda_i < \tau} (\tau - \lambda_i) s_i(1)^2 K_{k-1}(\lambda_i, \tau)^2 > \frac{1}{2}\sum_{i=1}^{n} |\lambda_i - \tau| s_i(1)^2 K_{k-1}(\lambda_i, \tau)^2. \tag{48}$$

Now suppose that λ_j is the left end of a cluster of close eigenvalues so that $\lambda_j - \lambda_{j-1}$ is not small. In many cases $\lambda_j - \lambda_{j-1}$ is of the order of the average gap $(\lambda_n - \lambda_1)/(n-1)$ but that is not necessary to the analysis that follows. Consider the shift $\tau \leqslant \lambda_j$ and very close, if not equal to λ_j.

From (48) comes a useful estimate. Define an average of τ's distance from the eigenvalues to its left,

$$\begin{aligned}\mathscr{A}_k^- &:= \frac{\sum_{i=1}^{j-1} (\tau - \lambda_i)^2 s_i(1)^2 K_{k-1}(\lambda_i, \tau)^2}{\sum_{i=1}^{j-1} (\tau - \lambda_i) s_i(1)^2 K_{k-1}(\lambda_i, \tau)^2} \\ &\leqslant \frac{\sum_{i=1}^{n} (\tau - \lambda_i)^2 s_i(1)^2 K_{k-1}(\lambda_i, \tau)^2}{\sum_{i=1}^{j-1} (\tau - \lambda_i) s_i(1)^2 K_{k-1}(\lambda_i, \tau)^2}.\end{aligned}$$

From (48) for $\omega_k = -1$,

$$\frac{\mathscr{A}_k^-}{2} \leqslant \frac{\sum_{i=1}^n (\tau-\lambda_i)^2 s_i(1)^2 K_{k-1}(\lambda_i,\tau)^2}{\sum_{m=1}^n |\tau-\lambda_m| s_m(1)^2 K_{k-1}(\lambda_m,\tau)^2}. \tag{49}$$

We use (49) later. Clearly, $\mathscr{A}_k^- \geqslant \tau - \lambda_{j-1}$.

Theorem 5. *Consider a cluster $\mathscr{C}$ of $\#(\mathscr{C})$ close eigenvalues in the interior of the spectrum of T. Choose a shift τ close to but not exceeding the left end of $\mathscr{C}$. Define $\mathscr{A}_k^-$ as above. If the factorization $T-\tau I = L\Omega L^{\mathrm{t}}$ exists, then*

$$\sum_{\lambda_m\in\mathscr{C}} \mathrm{relcond}(\lambda_m-\tau) \leqslant \#(\mathscr{C}) + 2\sum_{\omega_k=-1} \frac{\|L\boldsymbol{e}_k\|^2}{\mathscr{A}_k^-}.$$

Proof. Invoke (47) for $\lambda_m \in \mathscr{C}$ to find that

$$\sum_{\lambda_m\in\mathscr{C}} \mathrm{relcond}(\lambda_m-\tau) = \#(\mathscr{C}) + 2\sum_{\lambda_m\in\mathscr{C}}\sum_{\omega_k=-1} \frac{\left(L^{\mathrm{t}}\boldsymbol{s}_m\right)(k)^2}{|\lambda_m-\tau|}.$$

Reverse the summations and invoke (47) for each $\lambda_m \in \mathscr{C}$, where $\lambda_m - \tau > 0$,

$$\sum_{\lambda_m\in\mathscr{C}} \mathrm{relcond}(\lambda_m-\tau)$$
$$= \#(\mathscr{C}) + 2\sum_{\omega_k=-1} \|L\boldsymbol{e}_k\|^2 \frac{\sum_{\lambda_m\in\mathscr{C}} |\lambda_m-\tau| s_m(1)^2 K_{k-1}(\lambda_m,\tau)^2}{\sum_{i=1}^n (\lambda_i-\tau)^2 s_i(1)^2 K_{k-1}(\lambda_i,\tau)^2}.$$

The numerator above comes from a subset of the terms defining P_k and $P_k < N_k$ when $\omega_k = -1$. Thus,

$$\sum_{\lambda_m\in\mathscr{C}} \mathrm{relcond}|\lambda_m-\tau|$$
$$\leqslant \#(\mathscr{C}) + \sum_{\omega_k=-1} \|L\boldsymbol{e}_k\|^2 \frac{\sum_{m=1}^n |\lambda_m-\tau| s_m(1)^2 K_{k-1}(\lambda_m,\tau)^2}{\sum_{i=1}^n (\lambda_i-\tau)^2 s_i(1)^2 K_{k-1}(\lambda_i,\tau)^2}$$
$$\leqslant \#(\mathscr{C}) + 2\sum_{\omega_k=-1} \frac{\|L\boldsymbol{e}_k\|^2}{\mathscr{A}_k^-} \quad \text{(by (49))}. \qquad \square$$

Corollary 2. *With the hypotheses of Theorem* 5 *and*

$$\text{gap-left} := \min\{\tau-\lambda_i : \lambda_i < \tau\}$$

then

$$\sum_{\lambda_m\in\mathscr{C}} \mathrm{relcond}(\lambda_m-\tau) \leqslant \#(\mathscr{C}) + \frac{2}{\text{gap-left}} \sum_{\omega_k=-1} \left(L^{\mathrm{t}}L\right)_{kk}.$$

Proof. $\mathscr{A}_k^-$ is a weighted average of $\tau - \lambda_i$, $\lambda_i < \tau$, and so gap-left $\leqslant \mathscr{A}_k^-$. Now substitute gap-left for $\mathscr{A}_k^-$ in Theorem 5. □

At the right end of the cluster the ks with $\omega_k = +1$ are used.

Corollary 3. *If a shift τ' is chosen close to but not less than the right end of a cluster $\mathscr{C}$, if $T - \tau' I = L'\Omega'(L')^{\mathrm{t}}$ and*

$$\text{gap-right} := \min\{\lambda_j - \tau' : \lambda_j > \tau'\},$$

then

$$\sum_{\lambda_m \in \mathscr{C}} \text{relcond}\left(\lambda_m - \tau'\right) \leqslant \#(\mathscr{C}) + \frac{2}{\text{gap-right}} \sum_{\omega_k' = +1} \left[\left(L'\right)^{\mathrm{t}} L'\right]_{kk}.$$

Remark 2. The bound in Theorem 5 is a sum of two expressions. The term 2 $\sum_{\omega_k=-1} \|L\boldsymbol{e}_k\|^2/\mathscr{A}_k^-$ applies to all $\lambda_m > \tau$, not just those in cluster. Hence,

$$\sum_{\lambda_m > \tau} \text{relcond}(\lambda_m - \tau) \leqslant \#\{\lambda_m > \tau\} + 2 \sum_{\omega_k = -1} \frac{\|L\boldsymbol{e}_k\|^2}{\mathscr{A}_k^-}. \tag{50}$$

Similarly,

$$\sum_{\lambda_m < \tau} \text{relcond}(\lambda_m - \tau) \leqslant \#\{\lambda_m < \tau\} + 2 \sum_{\omega_k = +1} \frac{\|L\boldsymbol{e}_k\|^2}{\mathscr{A}_k^+}, \tag{51}$$

where

$$\mathscr{A}_k^+ := \frac{\sum_{\lambda_i > \tau} (\lambda_i - \tau)^2 \boldsymbol{s}_i(1)^2 K_{k-1}(\lambda_i, \tau)^2}{\sum_{\lambda_i > \tau} (\lambda_i - \tau) \boldsymbol{s}_i(1)^2 K_{k-1}(\lambda_i, \tau)^2}.$$

Example 3 exhibits a factorization $T - \tau I = L\Omega L^{\mathrm{t}}$ with large element growth. The relconds of the large eigenvalues are large but the relconds of the two tiny eigenvalues are bounded by 2.5 and so the representation is relatively robust for the cluster. Next we give a computable bound for the cluster nearest 0 that is independent of element growth in L. Recall from Section 7,

$$\rho_k(\tau) := -\frac{\beta_k p_k p_{k-1}}{\left\|\boldsymbol{p}^k\right\|^2}, \quad k = 1, \ldots, n. \tag{52}$$

which is the Rayleigh quotient of $\binom{\boldsymbol{p}^k}{\boldsymbol{o}}$ with respect to $T - \tau I$. The result is valid for any cluster but our interest is only in the one closest to zero.

Theorem 6. *Let $\mathscr{C} = \{\lambda_j, \lambda_{j+1}, \ldots\}$ be a cluster of eigenvalues. Then, using* (52), *as $\tau \to \lambda_j$,*

$$\sum_{\lambda_m \in \mathscr{C}} \text{relcond}(\lambda_m - \tau) \leqslant \max_{\lambda_m \in \mathscr{C}} |\lambda_m - \tau| \sum_{k=1}^{n-1} |\rho_k(\tau)|^{-1} + 1 + \mathrm{O}(|\tau - \lambda_j|).$$

Proof. From Theorem 3, for $k < n$,

$$\frac{\left(L^{\mathrm{t}} \boldsymbol{s}_m\right)(k)^2}{|\lambda_m - \tau|} = \frac{|\lambda_m - \tau|}{|\beta_k p_k p_{k-1}|} \left(s_m(1) K_{k-1}(\lambda_m, \tau)\right)^2.$$

From (32) in Section 6, as $\tau \to \lambda_j$,

$$\frac{\left(L^{\mathrm{t}} \boldsymbol{s}_m\right)(n)^2}{|\lambda_m - \tau|} \longrightarrow \left(\frac{s_m(n)}{s_j(n)}\right)^2 \left|\frac{\tau - \lambda_j}{\tau - \lambda_m}\right|.$$

Using (52), for $k < n$,

$$\frac{\left(L^{\mathrm{t}} \boldsymbol{s}_m\right)(k)^2}{|\lambda_m - \tau|} = \frac{|\lambda_m - \tau|}{|\rho_k(\tau)|} \left(\frac{s_m(1) K_{k-1}(\lambda_m, \tau)}{\|\boldsymbol{p}^k\|}\right)^2. \tag{53}$$

Now sum (53) over the cluster

$$\begin{aligned}
\sum_{\lambda_m \in \mathscr{C}} \frac{\left(L^{\mathrm{t}} \boldsymbol{s}_m\right)(k)^2}{|\lambda_m - \tau|} &= \sum_{\lambda_m \in \mathscr{C}} \frac{|\lambda_m - \tau|}{|\rho_k(\tau)|} \left(\frac{s_m(1) K_{k-1}(\lambda_m, \tau)}{\|\boldsymbol{p}^k\|}\right)^2 \\
&\leqslant \frac{\max_{\mathscr{C}} |\lambda_m - \tau|}{|\rho_k(\tau)|} \cdot \frac{\sum_{m \in \mathscr{C}} \left(s_m(1) K_{k-1}(\lambda_m, \tau)\right)^2}{\|\boldsymbol{p}^k\|^2} \\
&\leqslant \frac{\max_{\mathscr{C}} |\lambda_m - \tau|}{|\rho_k(\tau)|},
\end{aligned} \tag{54}$$

since the eigenvector matrix S yields

$$\begin{pmatrix} \boldsymbol{p}^k \\ \boldsymbol{o} \end{pmatrix} = S S^{\mathrm{t}} \begin{pmatrix} \boldsymbol{p}^k \\ \boldsymbol{o} \end{pmatrix} = \sum_{i=1}^{n} \boldsymbol{s}_i s_i(1) K_{k-1}(\lambda_i, \tau).$$

For the last terms, as $\tau \to \lambda_j$,

$$\begin{aligned}
\sum_{m \in \mathscr{C}} \frac{\left(L^{\mathrm{t}} \boldsymbol{s}_m\right)(n)^2}{|\lambda_m - \tau|} &\longrightarrow \sum_{m \in \mathscr{C}} \left(\frac{s_m(n)}{s_j(n)}\right)^2 \left|\frac{\tau - \lambda_j}{\tau - \lambda_m}\right| \\
&\longrightarrow \begin{cases} 1 + \mathrm{O}(|\tau - \lambda_j|), & j \in \mathscr{C}, \\ \mathrm{O}(|\tau - \lambda_j|), & j \notin \mathscr{C}. \end{cases}
\end{aligned} \tag{55}$$

Sum (54) for $k = 1, \ldots, n-1$ and then add (55) to obtain Theorem 6. □

The bound on $\max_{\mathscr{C}} |\lambda_m - \tau| \sum_{k=1}^{n-1} |\rho_k(\tau)|^{-1}$ may be accumulated in $\mathrm{O}(n)$ operations when $T - \tau I$ is factored using the three term recurrence for $\{p_i\}$.

Theorem 6 should be compared with (42), the term $\max_{\mathscr{C}} |\lambda_m - \tau|$ compensates for $\sum_{k=1}^{n-1} |\rho_k|^{-1}$ which may be large when there is element growth in $L\Omega L^{\mathrm{t}}$.

Remark 3. Suppose $\tau = \lambda_j$ and $\mathscr{C} = \{\lambda_j, \lambda_{j+1}, \ldots, \lambda_{j+\ell}\}$, then $\mathrm{relcond}(\lambda_j - \tau) = 1$. Suppose that

$$\sum_{k=1}^{n-1} |\rho_k(\tau)|^{-1} = \frac{\mu}{\lambda_{j+1} - \lambda_j} \quad \text{(defining } \mu\text{)},$$

then

$$\sum_{\lambda_m \in \mathscr{C}} \mathrm{relcond}(\lambda_m - \lambda_j) \leqslant \mu \frac{\lambda_{j+\ell} - \lambda_j}{\lambda_{j+1} - \lambda_j} + 1.$$

For a fairly uniform distribution in the cluster $\mathscr{C}$ this gives a bound of $\mu \cdot \ell + 1$ and bears out our experience that a cluster near τ has approximately the same relconds for each eigenvalue if $\mu = \mathrm{O}(1)$.

Remark 4. The algebraic function $\rho_k(\tau)$ vanishes at the zeros of p_{k-1} and p_k. It can be shown that

$$\rho_k'(\tau) = \begin{cases} +1 & \text{if } p_{k-1}(\tau) = 0, \\ -1 & \text{if } p_k(\tau) = 0. \end{cases}$$

Suppose that $\mathscr{C} = \{\lambda_j, \lambda_{j+1}, \ldots, \lambda_{j+\ell}\}$ is a cluster of close eigenvalues but $\boldsymbol{s}_j$ has some zero entries. Thus, λ_j is not a valid shift. If, instead, we choose $\tau = \lambda_j - \frac{1}{2}(\lambda_{j+1} - \lambda_j)$, then we can expect $|\rho_k(\tau)| = \mathrm{O}(\lambda_{j+1} - \lambda_j)$, $k = 1, \ldots, n-1$ and the associated factorization $T - \lambda I = L\Omega L^{\mathrm{t}}$ should provide a relatively robust representation for the smallest cluster even if some of the eigenvalues further from τ have large relconds.

9. Sensitivity of eigenvectors

It turns out that the natural definition of a condition number for an eigenvector of $L\Omega L^{\mathrm{t}}$ under *inner* multiplicative perturbations is a complicated combination of the relconds of all the eigenvalues. We derive the formula for $\mathrm{relcond}_i(\boldsymbol{s})$ in this section.

Recall from Section 2 that inner perturbations change $T - \tau I = L\Omega L^{\mathrm{t}} \longrightarrow LD\Omega L^{\mathrm{t}}$ with D diagonal and positive-definite. For small *relative* perturbations to L's entries the perturbation $D = I + \Delta$ with $\|\Delta\| \leqslant 2\eta$, the perturbation level and so gives an additive perturbation

$$L\Omega L^{\mathrm{t}} + L\Delta\Omega L^{\mathrm{t}}.$$

Our Δ here is twice the Δ in Section 2. The change in a specific eigenvector $\boldsymbol{s}_j$ may be expanded in the other eigenvectors as $\sum_{i\neq j} \boldsymbol{s}_i \eta_{ij}$. Standard first order perturbation theory [16, Chapter 2], starts from

$$\left(L\Omega L^{\mathrm{t}} + L\Delta\Omega L^{\mathrm{t}}\right)\left(\boldsymbol{s}_j + \sum_{i\neq j} \boldsymbol{s}_i \eta_{ij}\right) = \left(\boldsymbol{s}_j + \sum_{i\neq j} \boldsymbol{s}_i \eta_{ij}\right)(\lambda_j + \delta_j)$$

and, to first order in η, yields

$$L\Delta\Omega L^{\mathrm{t}}\boldsymbol{s}_j + \sum_{i\neq j} \boldsymbol{s}_i(\lambda_i - \tau)\eta_{ij} = \boldsymbol{s}_j\delta_j + (\lambda_j - \tau)\sum_{i\neq j} \boldsymbol{s}_i\eta_{ij}.$$

Premultiplication by $\boldsymbol{s}_j^{\mathrm{t}}$ gives the material presented in Section 2. Premultiplication by $\boldsymbol{s}_k^{\mathrm{t}}$, $k \neq j$, yields

$$\boldsymbol{s}_k^{\mathrm{t}} L\Delta\Omega L^{\mathrm{t}}\boldsymbol{s}_j + (\lambda_k - \tau)\eta_{kj} = (\lambda_j - \tau)\eta_{kj} + \mathrm{O}(\eta^2). \tag{56}$$

At this point we invoke the definition in Section 2,

$$\left\| L^{\mathrm{t}}\boldsymbol{s}_i \right\| = \sqrt{|\lambda_i - \tau|\ \mathrm{relcond}(\lambda_i - \tau)}. \tag{57}$$

Since $\Omega = \mathrm{diag}(\pm 1)$, $\|\Delta\Omega\| = \|\Delta\| \leqslant 2\eta$, and so, to first order in η

$$\begin{aligned} |(\lambda_j - \lambda_k)\eta_{kj}| &= \left|\boldsymbol{s}_k^{\mathrm{t}} L\Delta\Omega L^{\mathrm{t}}\boldsymbol{s}_j\right| \\ &\leqslant 2\eta\left[|\lambda_k - \tau| \cdot |\lambda_j - \tau|\ \mathrm{relcond}(\lambda_j - \tau)\mathrm{relcond}(\lambda_k - \tau)\right]^{1/2} \\ &\quad + \mathrm{O}(\eta^2). \end{aligned} \tag{58}$$

We mention that Δ may be chosen so that the bound in (58) is attained. In the discussion of eigenvectors it is the angle ψ_j (in radians) between $\boldsymbol{s}_j$ and $\boldsymbol{s}_j + \sum_{i\neq j} \boldsymbol{s}_i\eta_{ij}$ that is of interest. The eigenvectors of $L\Omega L^{\mathrm{t}}$ are orthonormal and so

$$\begin{aligned} \tan(\psi_j) &= \left(\sum_{i\neq j} \eta_{ij}^2\right)^{1/2} \\ &\leqslant 2\eta\sqrt{|\lambda_j - \tau| \cdot \mathrm{relcond}(\lambda_j - \tau)} \\ &\quad \cdot \left(\sum_{i\neq j} \frac{|\lambda_i - \tau| \cdot \mathrm{relcond}(\lambda_i - \tau)}{(\lambda_j - \lambda_i)^2}\right)^{1/2} + \mathrm{O}(\eta^2). \end{aligned} \tag{59}$$

The coefficient of 2η in (59) gives the appropriate expression for $\mathrm{relcond}(\boldsymbol{s}_j) = \mathrm{relcond}(\boldsymbol{s}_j; L\Omega L^{\mathrm{t}})$. It is a somewhat complicated function of the relconds for all the eigenvalues as well as the (relative) separation of the eigenvalues. In order to improve appearances we introduce a little used measure (denoted χ in [9]) of relative separation,

$$\mathrm{relsep}(a, b) := \frac{|a - b|}{\sqrt{|a||b|}} \tag{60}$$

and observe that this function can reach ∞. By this device

$$\operatorname{relcond}(\boldsymbol{s}_j) := \sqrt{\operatorname{relcond}(\lambda_j - \tau)} \cdot \left(\sum_{i \neq j} \frac{\operatorname{relcond}(\lambda_i - \tau)}{\operatorname{relsep}^2(\lambda_j - \tau, \lambda_i - \tau)} \right)^{1/2}. \tag{61}$$

We conclude with some implications of our definition of $\operatorname{relcond}(\boldsymbol{s}_j)$.

Remark 5. If $\boldsymbol{s}_j$ has no zero entries, then $L\Omega L^{\mathrm{t}}$ exists when $\tau = \lambda_j$ and $\operatorname{relcond}(\boldsymbol{s}_j) = 0$ as it should because $L^{\mathrm{t}}\boldsymbol{s}_j = \boldsymbol{o}$ in this case and so $L\Delta\Omega L^{\mathrm{t}}\boldsymbol{s}_j = \boldsymbol{o}$ however large Δ may be.

Remark 6. The unusual definition of $\operatorname{relcond}(\boldsymbol{s}_j)$ in (61) with its range $(0, \infty]$ arises because it concerns the *absolute* change in the angle ψ_j due to *relative* changes in L's entries.

Remark 7. The definition (60) makes our relseps larger than traditional measures such as $(|a - b|/\max\{|a|, |b|\})$. For example, if $\tau = \lambda_j$, then the term $i = j$ makes no contribution to $\operatorname{relcond}(\boldsymbol{s}_k)$, $k \neq j$. When $k = j + 1$ the sensitivity of $\boldsymbol{s}_{j+1}$ is most influenced by the contribution of λ_{j+2} even when $|\lambda_j - \lambda_{j+1}| \ll |\lambda_{j+1} - \lambda_{j+2}|$.

Remark 8. When $T - \tau I$ is definite then all $\operatorname{relcond}(\lambda_i - \tau) = 1$ and

$$\operatorname{relcond}(\boldsymbol{s}_j) = \left(\sum_{i \neq j} \frac{1}{\operatorname{relsep}^2(\lambda_j - \tau, \lambda_i - \tau)} \right)^{1/2}.$$

This shows that *a cluster near the middle of the spectrum* has eigenvectors sensitive to small relative errors in the Cholesky factors because $|\lambda_j - \tau|$ and $|\lambda_{j+1} - \tau|$ will be large. This observation confirms the necessity for taking τ close to each cluster in turn in order to compute orthogonal eigenvectors associated with those clusters.

Remark 9. Consider an interior cluster $\mathscr{C}$ with τ close to one end so that $\operatorname{relsep}(\lambda_i - \tau, \lambda_j - \tau) > 1$ for $\lambda_j \in \mathscr{C}$, $\lambda_i \notin \mathscr{C}$. Then the eigenvalues outside $\mathscr{C}$ contribute little to $\operatorname{relcond}(\boldsymbol{s}_j)$.

For $\lambda_j \in \mathscr{C}$

$$\operatorname{relcond}(\boldsymbol{s}_j) \approx \left(\sum_{\lambda_i \in \mathscr{C}, i \neq j} \frac{1}{\operatorname{relsep}^2(\lambda_i - \tau, \lambda_j - \tau)} \right)^{1/2}.$$

This expression is easily computed and if it is larger than desired the cluster may be split into subclusters. For example, consider a cluster in which all eigenvalues agree to 4 decimals but 7 at one end agree to 6 decimals and 5 at the other end also agree to 6 decimals. It might be profitable to subdivide into a subcluster at each end.

References

[1] T.S. Chihara, An Introduction to Orthogonal Polynomials, Gordon and Breach, London, 1978.
[2] J. Barlow, J. Demmel, Computing accurate eigensystems of scaled diagonally dominant matrices, SIAM J. Numer. Anal. 27 (1990) 762–791.
[3] I.S. Dhillon, A new O(n^2) algorithm for the symmetric tridiagonal eigenvalue/eigenvector problem, Ph.D. thesis, Computer Science Division, Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA, USA, 1997; also available as Computer Science Division Technical Report No. UCB//CSD-97-971.
[4] I.S. Dhillon, Reliable computation of the condition number of a tridiagonal matrix in O(n) time, SIAM J. Matrix Anal. Appl. 19 (3) (1998) 776–796.
[5] I.S. Dhillon, B.N. Parlett, Orthogonal eigenvectors and relative gaps, submitted to SIAM J. Matrix Anal. Appl.
[6] S. Eisenstat, I.C.F. Ipsen, Relative perturbation bounds for eigenspaces and singular vector subspaces, in: J.G. Lewis (Ed.), Proceedings of the Fifth SIAM Conference on Applied Linear Algebra, SIAM, 1994, pp. 62–65.
[7] S. Eisenstat, I.C.F. Ipsen, Relative perturbation techniques for singular value problems, SIAM J. Numer. Anal. 32 (1995).
[8] K.V. Fernando, B.N. Parlett, Accurate singular values and differential qd algorithms, Numer. Math. 67 (2) (March 1994) 191–229.
[9] R.-C. Li, Relative perturbation theory: I. Eigenvalue and singular value variations, SIAM J. Matrix Anal. Appl. 19 (1998) 956–982.
[10] R.-C. Li, Relative perturbation theory: II. Eigenspace and singular subspace variations, SIAM J. Matrix Anal. Appl. 20 (1998) 471–492.
[11] R.-C. Li, Relative perturbation theory: III. More bounds on eigenvalue variation, Linear Algebra Appl. 266 (1997) 337–345.
[12] B.N. Parlett, I.S. Dhillon, Fernando's solution to Wilkinson's problem: an application of double factorization, Linear Algebra Appl. 267 (1997) 247–279.
[13] B.N. Parlett, Spectral sensitivity of products of bidiagonals, Linear Algebra Appl. 275/276 (1998) 417–431.
[14] B.N. Parlett, The Symmetric Eigenvalue Problem, Classics ed., SIAM, Philadelphia, PA, USA, 1998, p. 398.
[15] K. Veselić, I. Slapničar, Floating point perturbations of Hermitian matrices, Linear Algebra Appl. 195 (1993) 81–116.
[16] J.H. Wilkinson, The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, 1965.