Center for Pure and Applied Mathematics

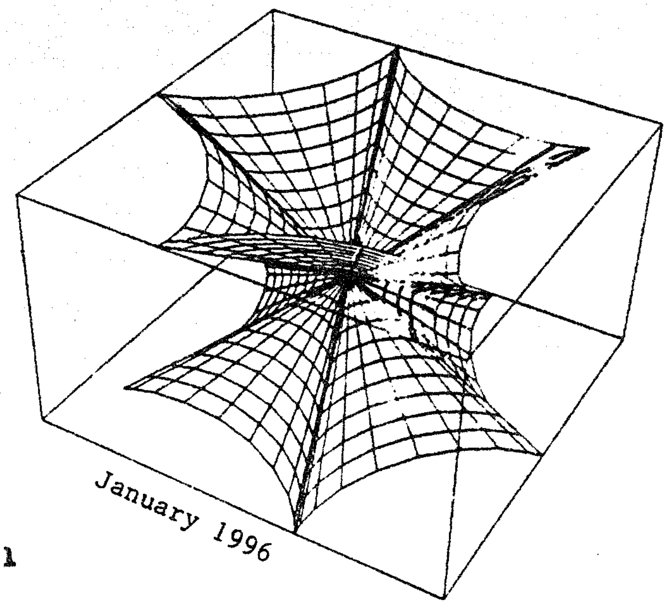
University of California at Berkeley

PAM-664

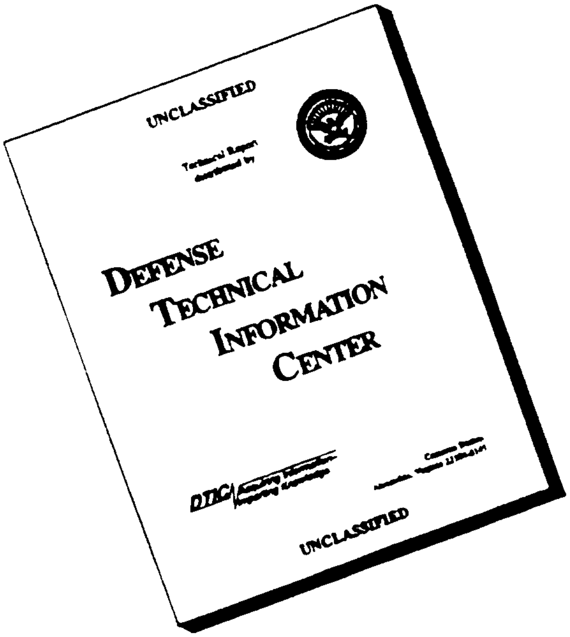
THE CONSTRUCTION OF ORTHOGONAL EIGENVECTORS FOR TIGHT CLUSTERS BY USE OF SUBMATRICES

Beresford Parlett

Department of Mathematics
University of California
Berkeley, California 94720



January 1996



This report was done with support from the Center for Pure and Applied Mathematics. Any conclusions or opinions in this report represent solely those of the author(s) and not necessarily those of the Center for Pure and Applied Mathematics or the Department of Mathematics.

The Construction of Orthogonal Eigenvectors for Tight Clusters by Use of Submatrices

Beresford N. Parlett *

January 18, 1996

Abstract

The goal is to compute eigenvectors of a symmetric tridiagonal matrix T that are orthogonal to working accuracy. Consider a cluster of m very close eigenvalues that are reasonably well separated from the remaining spectrum. We show here that there are m principal submatrices of T such that only the nearest neighbors overlap and each submatrix has a simple, isolated eigenvalue in the convex hull of the cluster with eigenvector having small entries in the first and last positions. This eigenvector is padded with zero entries, above and below, to make it conform to T. The set of vectors, one from each submatrix, forms a good basis for the invariant subspace. Each basis vector may be modified, if necessary, by its nearest neighbors to produce an orthonormal basis.

The only communication that may be needed, in such situations, is between nearest neighbors.

We give a good bound on the dot product of nearest neighbors. A variety of examples illustrate the theory.

The ideas in this paper were presented at the ENUMATH meeting at CNRS, Paris, in September 1995 and at the ILAY workshop at Cerfacs in October, 1995.

*Mathematics Department and Computer Science Division, EECS Department, University of California, Berkeley, CA 94720, USA. Supported by ONR, Contract N000014-90-J-1372.

Contents

1 Introduction

> 'Inverse Iteration gives a very satisfactory solution to the problem as far as reasonably well separated eigenvalues are concerned. The problem of determining reliably full digital information in the subspace spanned by eigenvectors corresponding to coincident or pathologically close eigenvalues has never been satisfactorily solved.'
>
> J. H. Wilkinson
> (from 'The Algebraic Eigenvalue Problem', 1965,
> Chapter 5, p. 344.)

This quotation is over 30 years old and yet most experts would agree that its claim is still true in 1995.

Our goal is to build up an orthonormal basis for the invariant subspace associated with a reasonably isolated cluster of very close eigenvalues of a symmetric tridiagonal matrix T. This is achieved by the QR algorithm and it is only the relatively high cost of accumulating all the plane rotations that drives the search for other techniques. Certain codes (e. g., xstein) currently used in LAPACK, and other libraries such as NAG and IMSL and ESSL, decline in both efficiency and quality of output as eigenvalues get closer to each other. The reason is that these codes are based on inverse iteration and it is extremely difficult to choose automatically suitable right hand sides to ensure both the spanning property (accuracy) and orthogonality.

One way out of the difficulty is to discard an appropriate set of rows from the top and bottom of T and to work with the remaining submatrix to obtain a basis vector. The well known test matrix W_{21}^{+} that is discussed in Section 5 has its largest two eigenvalues equal to single (and double) precision. In this easy case it suffices to compute the eigenvectors of the submatrices in rows 1 to 19 and 3 to 21, append zero entries at the bottom of one and the top of the other, and finally deliver the internal and external bisectors of these two vectors. Even the tiny entries are computed to high relative accuracy.

The guiding principle behind this approach is that an eigenvector associated with a simple, isolated eigenvalue is easy to compute. So, for a cluster of m close eigenvalues, the task is to find m different submatrices of T each of which has an isolated eigenvalue in the convex hull of the cluster. That is not enough. The eigenvector of any internal submatrix must have small entries in its first and last positions. It is not obvious that these specifications can always be met and the goal of this paper is to provide the theory that supports our use of submatrices.

It turns out that the case of close pairs is fundamental. The general case may be reduced to considering a number of pairs. A surprising outcome of these investigations is that the difficulty of computing an accurate orthogonal basis is not properly measured by the gap between eigenvalues. The *support* of the vectors (the positions holding nonnegligible values) plays an important role. Strictly speaking zero entries in eigenvectors are extremely rare and, when they do occur, they are isolated. Since we ignore isolated zero entries we might conclude that the support of all the eigenvectors we seek is the full index set. To take such a pedantic view would be to miss an important fact about many, but not all, $n \times n$ tridiagonal matrices as n grows: the active part of the eigenvector is confined to a small part of the domain, perhaps only 30 or 40 consecutive positions. The remaining entries carry a tiny bit of noise. Consequently we use the term support somewhat informally. If the support of two eigenvectors is disjoint then they are orthogonal however close the associated eigenvalues may be.

Closely related to the idea of support is the concept of the *overlap* of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$;

$$Overlap(\boldsymbol{x},\boldsymbol{y}) := \frac{|\boldsymbol{x}| \cdot |\boldsymbol{y}|}{\|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\|},$$

the cosine between the vectors of absolute values. Our two main results are:
1. Each normalized vector $\boldsymbol{x}$ produced from an appropriate submatrix by appending zeros has $\mu := \boldsymbol{x}^* T \boldsymbol{x}$ in the cluster interval and

$$\|(T - \mu I)\boldsymbol{x}\| = O(cluster\ \ length)$$

2. Any two normalized vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ from overlapping submatrices satisfy

$$Overlap(\boldsymbol{x},\boldsymbol{y}) = O\left(\left(\frac{cluster\ \ length}{gap}\right)^{3/4}\right)$$

where *gap* is the separation of the cluster from the remaining spectra of the two submatrices.

Submatrices only overlap their nearest neighbors. Physicists would say that the overlap matrix is tridiagonal and close to the identity. However the subspace spanned by these subvectors is more accurate than indicated by the residual norms of the subvectors separately. That is the content of Section 9. What Results 1 and 2 show is that the subvectors yield a sparse, nearly orthogonal basis for a good subspace.

The results developed here are pure matrix theory, there is no reference to machine precision. They depend strongly on the tridiagonal from.

The first result is somewhat surprising because experience with the Lanczos algorithm had suggested that the best that could be guaranteed for any vector $\boldsymbol{x}$ obtained from a submatrix was $\|(T-\mu I)\boldsymbol{x}\| = O(\sqrt{cluster\ \ length})$ and that is not strong enough for our purposes.

Tight clusters of close eigenvalues are not the only spectral distributions to pose challenges to numerical analysts. Consider eigenvalues that form a geometric progression. This case is not troublesome if the eigenvalues are computed to high relative relative accuracy. A different problem is posed by small perturbations of the identity matrix I. How large must the perturbation be before the user will not accept the columns of I as suitable eigenvectors? Shifting by I and scaling by the largest remaining entry makes the perturbations seem important and the shifted and scaled eigenvalues seem well separated. One would then compute a set of eigenvectors very different from the columns of I. It would seem that only the user can choose between the two and we shall not pursue this question here.

We begin our analysis in Section 6 by introducing the envelope of a cluster and showing how it guarantees Result 1 with a constant bounded by $\sqrt{n/2}$ but normally much smaller. The envelope also reveals good choices for submatrices. Then we take a different approach to show how the submatrix indices affect the constants behind the $O(cluster\ \ length)$. Our results make heavy use of detailed properties of tridiagonal matrices and that material is gathered in Section 3 and Section 10.

The reader is urged to read Section 2 on notation.

2 Notation

$$T = tridiag \left(\begin{array}{ccccccccc} & \beta_1 & & \beta_2 & & \bullet & & \beta_{n-1} & \\ \alpha_1 & & \alpha_2 & & \bullet & & \bullet & & \alpha_n \\ & \beta_1 & & \beta_2 & & \bullet & & \beta_{n-1} & \end{array} \right)$$

$$\text{Eigenvalues:} \quad \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n. \tag{1}$$

$$\text{Normalized eigenvectors:} \quad \boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n. \tag{2}$$

When $\beta_i \neq 0$, $i = 1, \ldots, n-1$ then T is *unreduced.* In that case the inequalities in (1) are strict.

The principal submatrix of T in rows $j, j+1, \ldots, k$ is denoted by either $T^{j:k}$ or, simply, by $(j : k)$. The eigenvalues of a submatrix $(j : k)$ are called Ritz values or R-values:

$$\theta_1^{j:k} < \theta_2^{j:k} < \ldots < \theta_{k-j+1}^{j:k}. \tag{3}$$

The normalized eigenvector of $\theta_l^{j:k}$ is $\boldsymbol{s}_l^{j:k}$. The mth entry is $s_l^{j:k}(m)$. The natural way to index the entries of $\boldsymbol{s}_l^{j:k}$ is from j to k, not from 1 to $k-j+1$. In this way $\boldsymbol{s}_l^{j:k}$ is embedded in a vector conformable to T and we consider entries in positions $1 : j-1$ and $k+1 : n$ as zero. Sometimes we write $\theta_l^{1:j}$ as $\theta_l^{(j)}$. To simplify further we write $\theta^{(j)}$ for $\theta_i^{(j)}$ when $\theta_i^{(j)}$ lies in $[\lambda_-, \lambda_+]$, our cluster.

The characteristic polynomial of $T^{j:k}$, or $(j : k)$, is defined by

$$\chi^{j:k}(\tau) := det\,(\tau I - T^{j:k}).$$

We write χ_j for $\chi^{1:j}$. In general column vectors are denoted by lower case Roman letters in boldface type: $\boldsymbol{v}, \boldsymbol{z}, \ldots$, and individual entries are $v(j), z(l), \ldots$. The size of the identity matrix I is given by the context and its columns are

$$\boldsymbol{e}_i = (0, \ldots 0, 1, 0 \ldots, 0)^t, \quad 1 \text{ in position } i.$$

3 Eigenvectors and Error Estimates

The unreduced $n \times n$ matrix T is uniquely determined (up to signs) by its spectrum $\{\lambda_1, \ldots, \lambda_n\}$ and $(n-1)$ appropriate extra items. This extra information may take various forms: the spectrum of submatrix $(1 : n-1)$ or $(2 : n)$, or the squares of the top (or bottom) entries of the normalized eigenvectors. See [4] or [2].

Consequently there are numerous expressions for the eigenvectors of T and we give some of them here.

Here we assume that T is *unreduced*: $\beta_i \neq 0$, $i = 1, \ldots, n-1$.

Perhaps the oldest formulae for the unnormalized eigenvector for λ are

$$\left(1, \frac{\chi^{1:1}(\lambda)}{\beta_1}, \frac{\chi^{1:2}(\lambda)}{\beta_1\beta_2}, \ldots, \frac{\chi^{1:n-1}(\lambda)}{\beta_1\beta_2\ldots\beta_{n-1}}\right)^t, \tag{4}$$

$$\left(\frac{\chi^{2:n}(\lambda)}{\beta_1\ldots\beta_{n-1}}, \frac{\chi^{3:n}(\lambda)}{\beta_2\ldots\beta_{n-1}}, \ldots, \frac{\chi^{n:n}(\lambda)}{\beta_{n-1}}, 1\right)^t \tag{5}$$

The catch to using these formulae is that they can overflow very easily and they may sometimes be extremely sensitive to tiny changes in λ.

There are attractive formulae for the magnitudes of the entries of normalized eigenvectors $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$. For each $i = 1, 2, \ldots, n$, and $\chi = \chi^{1:n}$

$$\begin{aligned} s_i(1)s_i(n)\chi'(\lambda_i) &= \beta_1\ldots\beta_{n-1} && (6)\\ s_i(1)^2\chi'(\lambda_i) &= \chi^{2:n}(\lambda_i), && (7)\\ s_i(n)^2\chi'(\lambda_i) &= \chi^{1:n-1}(\lambda_i), && (8)\\ s_i(j)^2\chi'(\lambda_i) &= \chi^{1:j-1}(\lambda_i)\chi^{j+1:n}(\lambda_i). && (9) \end{aligned}$$

All these formulae come from a result in [4] that

$$adj\,(\lambda_i I - T) = \boldsymbol{s}_i\boldsymbol{s}_i^*\,\chi'(\lambda_i)$$

where $adj(M)$ is the classical adjugate of M.

These results raise the question of how to find the correct signs of the entries and that leads to the next few observations.

Given any unreduced $\tilde{T}$ there is a unique $\Delta = diag(\pm1, \pm1, \ldots, \pm1)$ such that $T = \Delta\tilde{T}\Delta$ has positive off-diagonal entries. Observe that

$$T\boldsymbol{s} = \boldsymbol{s}\lambda \Longleftrightarrow \tilde{T}(\Delta\boldsymbol{s}) = (\Delta\boldsymbol{s})\lambda.$$

Consequently it is possible to normalize any symmetric T so that it becomes a direct sum of unreduced tridiagonals each of which has positive off-diagonal entries. From now on we assume that $\beta_i > 0$, $i = 1, \ldots, n-1$.

These Ts are oscillation matrices, a term coined by Krein and Gantmacher [1]. The number of sign changes among consecutive entries of the eigenvector $\boldsymbol{s}_i$ is $n - i$. Recall that we label eigenvalues so that

$$\lambda_1 < \lambda_2 < \ldots < \lambda_n.$$

For example, for the second difference SINE matrix $(1, -2, 1)$, the eigenvalues lie in $-4 < \lambda_i < 0$ and the fundamental mode is given by $\boldsymbol{s}_n$, associated with the rightmost eigenvalue. We use right and left rather than large or small in order to be translation invariant. For some people it is unnatural to say that -1 is larger than -2 because they think of magnitudes when the word large is used.

The correct signs may be attached to magnitudes when certain information is available. A warning is in order here. Although the characteristic polynomial is defined by $\chi(\tau) = det(\tau I - T)$ most people compute with $T - \tau I$. For example, the pivots in Gaussian elimination are usually computed from the recurrence

$$d_{i+1} = (\alpha_{i+1} - \tau) - \beta_i^2/d_i$$

and so

$$\chi^{1:j}(\tau) = (-1)^j d_1 \dots d_j.$$

If $\tau = \lambda_i$ (4) may be rewritten as

$$\left(1, -\left(\frac{d_1}{\beta_1}\right), \left(\frac{d_1}{\beta_1}\right)\left(\frac{d_2}{\beta_2}\right), \dots, (-1)^{n-1}\left(\frac{d_1}{\beta_1}\right)\cdots\left(\frac{d_{n-1}}{\beta_{n-1}}\right)\right)^t.$$

However, in practice $\tau \neq \lambda_i$.

We turn now to approximations and how to assess them. Let

$$T^{j:k}\boldsymbol{s}_l^{j:k} = \boldsymbol{s}_l^{j:k}\theta_l^{j:k}, \quad \|\boldsymbol{s}_l^{j:k}\| = 1,$$

define an eigenvector of a submatrix. Consider $\boldsymbol{s}_l^{j:k}$ as an approximate eigenvector of T and append zeros to $\boldsymbol{s}_l^{j:k}$ to make it conform. Now drop the subscript l and observe that

$$\begin{aligned} (T - \theta^{j:k}I)\boldsymbol{s}^{j:k} &= \boldsymbol{e}_{j-1}\beta_{j-1}s^{j:k}(j) + \boldsymbol{e}_{k+1}\beta_k s^{j:k}(k) && (10) \\ \|(T - \theta^{j:k}I)\boldsymbol{s}^{j:k}\| &= \left(\beta_{j-1}^2 s^{j:k}(j)^2 + \beta_k^2 s^{j:k}(k)^2\right)^{1/2}. && (11) \end{aligned}$$

Paige's Persistence Theorem. *Write $\theta_i^{(j)}$ for $\theta_i^{1:j}$. For any j, $1 \leq j \leq n$, and any i, $1 \leq i \leq j$, the closed interval*

$$[\theta_i^{(j)} - \beta_j\omega_i, \theta_i^{(j)} + \beta_j\omega_i]$$

contains at least one Ritz value $\theta_\bullet^{(j+l)}$ *for each* $l = 1, 2, \ldots, n-j$. *Here* $\omega_i = |s_i^{1:j}(j)|$.

Proof. By partitioning T one sees that

$$\left(T^{1:j+l} - \theta_i^{(j)} I_{j+l}\right) \begin{pmatrix} s_i^{(j)} \\ 0 \end{pmatrix} = e_{j+1} \beta_j \, s_i^{(j)}(j).$$

So, by a standard theorem, see Chap. 4 in [4], there is a $\theta_\bullet^{(j+l)}$ such that

$$|\theta_\bullet^{(j+l)} - \theta_i^{(j)}| \le \| \left(T^{1:j+l} - \theta_i^{(j)} I_{j+l}\right) \begin{pmatrix} s_i^{(j)} \\ 0 \end{pmatrix} \| = \beta_j \omega_i.$$

□

Frequently the bound $\beta_j \omega_i$ is a severe overestimate. Whenever $\theta_i^{(j)}$ is isolated from the other Ritz values $\theta_k^{(j+l)}$ then one defines

$$gap(i; j, l) = gap(i) := \min_{k \ne \bullet} |\theta_k^{(j+l)} - \theta_i^{(j)}|.$$

Gap Theorem. *For* $l = 1, 2, \ldots, n-j$ *there is a* $\theta_\bullet^{(j+l)}$ *such that*

$$|\theta_\bullet^{(j+l)} - \theta_i^{(j)}| \le (\beta_j \omega_i)^2 / gap(i; j, l).$$

See [4, Chap. 11].

This is a huge improvement over $\beta_j \omega_i$ in most cases.

The Average-β Result

Lemma 1 *For* $i = 1, 2, \ldots, n$,

$$(\beta_1 \cdot \ldots \cdot \beta_{n-1})^{1/(n-1)} < \left(\prod_{k \ne i} |\lambda_i - \lambda_k| \right)^{1/(n-1)}.$$

Proof. For each i, $1 \le i \le n$

$$s_i(1)\, s_i(n)\, \chi'(\lambda_i) = \beta_1 \cdot \ldots \cdot \beta_{n-1}$$

and

$$\chi'(\lambda_i) = \prod_{k \ne i} (\lambda_i - \lambda_k).$$

Since the geometric mean is majorized by the arithmetic mean

$$|s_i(1)\, s_i(n)| \le \frac{1}{2}\, \left(s_i^2(1) + s_i^2(n)\right) < \frac{1}{2}.$$

Hence

$$\beta_1 \cdot \ldots \cdot \beta_{n-1} \le \frac{1}{2}|\chi'(\lambda_i)|$$

and

$$(\beta_1 \cdot \ldots \cdot \beta_{n-1})^{1/(n-1)} \le |\chi'(\lambda_i)|^{1/(n-1)}/2^{1/(n-1)} < |\chi'(\lambda_i)|^{1/(n-1)}.$$

as claimed. □

Quantities of the form β_j/gap will be $O(1)$ on average and they occur in our analysis. Speaking loosely we may say that the average β is bounded by the average distance of any eigenvalue from all the others.

4 On First and Last Entries

Let λ_- and λ_+ be two adjacent eigenvalues of T well separated from the remaining spectrum. Let

$$T\boldsymbol{s}_\pm = \boldsymbol{s}_\pm \lambda_\pm, \quad \|\boldsymbol{s}_\pm\| = 1.$$

From (6) in Section 3,

$$s_\pm(1) s_\pm(n) \chi'(\lambda_\pm) = \beta_1 \cdots \beta_{n-1}$$

Thus

$$\begin{aligned}
\frac{s_+(1)s_+(n)}{s_-(1)s_-(n)} &= \frac{\chi'(\lambda_-)}{\chi'(\lambda_+)} \\
&= \frac{(\lambda_- - \lambda_+)\prod_j''(\lambda_- - \lambda_j)}{(\lambda_+ - \lambda_-)\prod_j''(\lambda_+ - \lambda_j)}, \quad \left(\prod_j{}'' \text{ means } \lambda_j \ne \lambda_\pm\right) \\
&= -\prod_j{}'' \left(1 - \frac{\lambda_+ - \lambda_-}{\lambda_+ - \lambda_j}\right) \\
&= -1 + (\lambda_+ - \lambda_-) \sum_j{}'' \frac{1}{\lambda_+ - \lambda_j} + \cdots \\
&= -1 + O\left(\frac{\lambda_+ - \lambda_-}{gap}\right)
\end{aligned}$$

where $gap = \min_j |\lambda_+ - \lambda_j|$. So, when $\lambda_+ - \lambda_- \ll gap$,

$$\left|\frac{s_+(1)}{s_-(1)}\right| \approx \left|\frac{s_-(n)}{s_+(n)}\right|.$$

If $|s_+(1)| \gg |s_-(1)|$ then the support of $\boldsymbol{s}_+$ is concentrated in the top of an n-vector whereas the support of $\boldsymbol{s}_-$ is concentrated in the lower section. Thus the supports may be nearly disjoint, however close λ_- and λ_+ may be. When $|s_+(1)| \approx |s_-(1)|$ then $\boldsymbol{s}_+$ and $\boldsymbol{s}_-$ have the same support and it is easier to approximate the bisectors $(\boldsymbol{s}_+ \pm \boldsymbol{s}_-)/\sqrt{2}$.

5 An Example: W_{21}^+

$$W = W_{21}^+ = tridiag\left(\begin{array}{cccccccccc} & 1 & & 1 & & \bullet & & & 1 & \\ 10 & & 9 & & \bullet & & \bullet & 9 & & 10 \\ & 1 & & 1 & & \bullet & & & 1 & \end{array}\right).$$

This matrix, and its companion W_{21}^-, were designed by Wilkinson, see Ch. 6 in [6], to illustrate some subtle points concerning the computation of eigenvectors. Although the eigenvalues of W are distinct in exact arithmetic there are several pairs that are equal to single precision ($\epsilon = 1.2 \times 10^{-7}$). However the smaller eigenvalue pairs are not so close and the only negative eigenvalue is well separated from the rest. Table 1 gives selected eigenvalues. Note that the W_{21}^+ is persymmetric: invariant under reversal.

> 'If we separate W_{21}^+ into a direct sum by putting $\beta_{11} = 0$ then we obtain independent orthogonal vectors which span the subspace corresponding to the pathologically close pair of eigenvalues λ_{20} and λ_{21} to a very high accuracy. It is therefore quite possible that such a decomposition is always permissible, and what is needed is some reliable method of deciding when and where to decompose.'
>
> J. H. Wilkinson
> (from 'The Algebraic Eigenvalue Problem', 1965,
> Chapter 5, p. 330.)

This quotation shows that Wilkinson considered the possibility of using *disjoint* submatrices to obtain orthogonal basis vectors. Our investigations

λ_{21}	10.746194182903393..
λ_{20}	10.746194182903322..
.	.
.	.
λ_{13}	6.0002340..
λ_{12}	6.0002175..
.	.
.	.
λ_3	0.9475..
λ_2	0.2538..
λ_1	-1.1254..

Table 1: Selected Eigenvalues of W_{21}^{+}

show that this idea only works in extreme, and easy, cases. What we illustrate here is that the submatrices must be allowed to overlap.

In a nice recent study, see [7], Ye has turned Wilkinson's idea into some theorems relating a pair of eigenvalues of T to an eigenvalue of a certain submatrix and an eigenvalue of the complementary submatrix. However he does not give results on the related eigenvectors and that is our concern.

The largest pair: λ_{20} and λ_{21} (equal to single precision at value λ)
We drop the last 2 rows and approximate the Ritz vector $\boldsymbol{z}_+$ by solving

$$(W^{1:19} - \lambda I)\boldsymbol{z}_+ = \boldsymbol{e}_1\gamma_1, \quad \|\boldsymbol{z}_+\| = 1.$$

We drop the first 2 rows and approximate the Ritz vector $\boldsymbol{z}_-$ by solving

$$(W^{3:21} - \lambda I)\boldsymbol{z}_- = \boldsymbol{e}_{21}\gamma_{21}, \quad \|\boldsymbol{z}_-\| = 1.$$

We insert zero entries to make $\boldsymbol{z}_-$ and $\boldsymbol{z}_+$ conform to W. It turns out that

$$\boldsymbol{z}_+ \cdot \boldsymbol{z}_- = 1.22 \times 10^{-13}\,!, \quad \gamma_1 = -8.33 \times 10^{-7}, \qquad \gamma_{21} = -8.33 \times 10^{-7}.$$

Thus $\boldsymbol{z}_-$, and $\boldsymbol{z}_+$ pass both requirements for good eigenvectors, orthogonality and small residuals. However an equally valid basis for the dominant invariant 2-space is $\{(\boldsymbol{z}_+ - \boldsymbol{z}_-)/\sqrt{2},\ (\boldsymbol{z}_+ + \boldsymbol{z}_-)/\sqrt{2}\}$, the bisectors. It is gratifying that this *computed* basis delivers the exact eigenvectors correct to single precision, even the smallest entries.

The results from dropping three rows instead of two were indistinguishable from the ones above. In fact this case is so easy that one can use submatrices $(1 : j)$ and $(22 - j : 21)$ for $j = 14, \ldots, 20$, for satisfactory results.

The pair near 6: λ_{12} **and** λ_{13}

The shift invariant measure of a symmetric matrix is its *spread* defined as $\lambda_{max} - \lambda_{min}$. In this case

$$\begin{aligned} spread(W) &= \lambda_{21} - \lambda_1 = 11.87, \\ interval\ width &= \lambda_{13} - \lambda_{12} = 11.66 \cdot \epsilon \cdot spread. \end{aligned}$$

The two best choices are (1 : 15), (7 : 21) and (1 : 17), (5 : 21). The outputs from the two choices are are barely distinguishable in single precision. The Ritz value μ for (1 : 17) and (5 : 21) is not exactly at the mean but, in single precision, appears to be so:

$$\mu = 6.000226.$$

Approximate the corresponding Ritz vectors by solving

$$(W^{1:17} - \mu I)\boldsymbol{z}_+ = \boldsymbol{e}_6\gamma_6, \quad (W^{5:21} - \mu I)\boldsymbol{z}_- = \boldsymbol{e}_{15}\gamma_{15}$$

in single precision to find

$$\|(W^{1:17} - \mu I)\boldsymbol{z}_+\| = \|(W^{5:21} - \mu I)\boldsymbol{z}_-\| = 1.4 \times 10^{-5} = 10 \cdot \epsilon \cdot spread$$

and

$$\boldsymbol{z}_+ \cdot \boldsymbol{z}_- = 2.1 \times 10^{-5} = 8.4 \cdot n \cdot \epsilon.$$

Section 9 shows how the subspace span $\{\boldsymbol{z}_-, \boldsymbol{z}_+\}$ is slightly more accurate than indicated by the residual norms for $\boldsymbol{z}_-$ and $\boldsymbol{z}_+$ separately.

In several ways $\{\boldsymbol{z}_-, \boldsymbol{z}_+\}$ is a satisfactory basis but the pair of bisectors $\{(\boldsymbol{z}_+ - \boldsymbol{z}_-)/\sqrt{2},\ (\boldsymbol{z}_+ + \boldsymbol{z}_-)/\sqrt{2}\}$ is even better. Just as for the pair $\lambda_{20}, \lambda_{21}$ the *computed* bisectors deliver the exact eigenvectors to single precision, even the smallest entries were correct to 6 decimals. Figure 1 shows $\boldsymbol{z}_-$ and

$\boldsymbol{z}_+$ while Figure 2 shows the bisectors. An unorthodox representation has been used in order to emphasize the small entries. Instead of $v(i)$ we plot $(-1/\log_{10}|v(i)|)\,\mathrm{sign}(v(i))$ and the vector is normalized so that the maximum value is 1. To show the distortion we also plot $(\boldsymbol{z}_+ + \boldsymbol{z}_-)/\sqrt{2}$ in the standard manner at the top of Figure 2.

A careful look at $\boldsymbol{z}_-$ and $\boldsymbol{z}_+$ shows that Wilkinson's idea of using *disjoint* submatrices could not be satisfactory: entries $z_+(12), \ldots, z_+(15)$ and $z_-(5), \ldots, z_-(8)$ are not negligible. However $\{\boldsymbol{z}_-, \boldsymbol{z}_+\}$ is the most sparse basis of the invariant subspace for $\lambda_{12}, \lambda_{13}$. So there is no basis in which half the entries in each vector are negligible, in contrast to the case for $\lambda_{20}, \lambda_{21}$. The overlap of the submatrices is needed to obtain the middle entries accurately.

These remarks do not contradict the fact the Wilkinson constructed W_{21}^+ so that each pair of eigenvectors could be built out of two small vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. Here

$$(W^{1:10} - \lambda_{12}I)\boldsymbol{u} = \mathbf{0}, \quad (\tilde{W}^{1:11} - \lambda_{13}I)\boldsymbol{v} = \mathbf{0}.$$

Where $\tilde{W}^{1:11}$, which is not symmetric, differs from $W^{1:11}$ by replacing entry $(11, 10)$ by 2. The eigenvectors for λ_{12} and λ_{13} are

$$\begin{array}{lll} (u(1), \ldots, u(10), & 0, & -u(10), \ldots, -u(1))^t, \\ (v(1), \ldots, v(10), & v(11), & v(10), \ldots, v(1))^t. \end{array}$$

Wilkinson's idea of using (1:11) and (12:21) (with $\beta_{11} = 0$) will not work here because $u(7:10) \neq v(7:10)$ to working accuracy in contrast to the case of $\lambda_{20}, \lambda_{21}$.

Inverse iteration, even with well chosen right hand sides, gives poor results unless the Gram-Schmidt process is used heavily. The output from the LAPACK code sstein was not as accurate as our bisectors.

The following sections show that well chosen submatrices yield good bases in all cases when the cluster is reasonably isolated from the remaining spectrum.

6 The Envelope of the Invariant Subspace

Let $\{\lambda_l, \lambda_{l+1}, \ldots, \lambda_{l+m-1}\}$ be a set of eigenvalues of T well separated from the rest of the spectrum. Let

$$T\boldsymbol{z}_i = \boldsymbol{z}_i\lambda_i, \quad \|\boldsymbol{z}_i\|_2 = 1,$$

and define $\mathcal{I} := [\lambda_l, \lambda_{l+m-1}]$ and

$$\mathcal{S}_{\mathcal{I}} = span\{\boldsymbol{z}_l, \ldots, \boldsymbol{z}_{l+m-1}\}.$$

The *envelope vector* of $\mathcal{S}_{\mathcal{I}}$ is $\mathcal{E}_{\mathcal{I}}$ given by

$$\mathcal{E}(j) = \mathcal{E}_{\mathcal{I}}(j) := \max\{v(j) : \ \boldsymbol{v} \in \mathcal{S}_{\mathcal{I}}, \ \|\boldsymbol{v}\|_2 = 1\}.$$

The *extremal vectors* $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(n)}$ in $\mathcal{S}_{\mathcal{I}}$ are characterized by

$$\boldsymbol{y}^{(j)}(j) = \mathcal{E}(j), \quad \|\boldsymbol{y}^{(j)}\|_2 = 1, \quad j = 1 \text{ and } n.$$

We consider the case $m = 2$ (close pairs) here although some of the results may be extended to larger clusters. The subvectors that are useful in computing an orthogonal basis for $\mathcal{S}_{\mathcal{I}}$ may be understood as approximations to these extremal vectors. A little more notation is needed before the results of this section can be described. By Lemma 2 (proved below)

$$\mathcal{E}(1) = (z_l(1)^2 + z_{l+1}(1)^2)^{1/2}$$

and $\boldsymbol{y}^{(1)}$ may be expressed as

$$\boldsymbol{y}^{(1)} = \boldsymbol{z}_l \cos\varphi + \boldsymbol{z}_{l+1} \sin\varphi, \qquad \tan\varphi = \frac{z_{l+1}(1)}{z_l(1)},$$

and φ plays an important role in this section. We may assume that $z_l(1) > 0$, $z_{l+1}(1) > 0$. The Rayleigh quotient of $\boldsymbol{y}^{(1)}$ is

$$\rho_1 = \boldsymbol{y}^{(1)} \cdot T\boldsymbol{y}^{(1)} = \lambda_l \cos^2\varphi + \lambda_{l+1} \sin^2\varphi. \tag{12}$$

The *residual* of $\boldsymbol{y}^{(1)}$ and the residual norm are

$$\begin{aligned} T\boldsymbol{y}^{(1)} - \boldsymbol{y}^{(1)}\rho_1 &= \boldsymbol{z}_l \cos\varphi\,(\lambda_l - \rho_1) + \boldsymbol{z}_{l+1} \sin\varphi\,(\lambda_{l+1} - \rho_1), \\ \nu_1^2 := \|T\boldsymbol{y}^{(1)} - \boldsymbol{y}^{(1)}\rho_1\|^2 &= (\lambda_l - \rho_1)^2 \cos^2\varphi + (\lambda_{l+1} - \rho_1)^2 \sin^2\varphi \\ &= \left(\sin 2\varphi \cdot \frac{\lambda_{l+1} - \lambda_l}{2}\right)^2, \quad \text{by (12).} \end{aligned} \tag{13}$$

defining the residual norm ν_1 which occurs throughout this section.

Our goal is to find an index j, $1 < j < n$, such that the eigenvector $\boldsymbol{s}^{1:j}$, $T^{1:j}\boldsymbol{s}^{1:j} = \boldsymbol{s}^{1:j}\theta^{1:j}$, $\theta^{1:j} \in \mathcal{I}$, satisfies $\beta_j|s^{1:j}(j)| = O(\lambda_{l+1} - \lambda_l)$. The connection of such j to the envelope vector $\boldsymbol{\mathcal{E}}$ is given by the principal results of this section which we summarize first.

For each j such that $T^{1:j}$ has an eigenvalue $\theta^{1:j} \in \mathcal{I}$, Theorem 1 shows that

$$\beta_j \left|s^{1:j}(j)\right| \mathcal{E}(j+1) \approx \left[\left(\sin 2\varphi \cdot \frac{\lambda_{l+1} - \lambda_l}{2}\right)^2 + (\theta^{1:j} - \rho_1)^2\right]^{1/2} \approx \nu_1,$$

provided that $(\lambda_{l+1} - \lambda_l)/gap \ll 1$, where

$$gap = \min\{\lambda_{l+2} - \theta^{1:j}, \theta^{1:j} - \lambda_{l-1}\}. \tag{14}$$

For W_{21}^+ and the dominant pair $\lambda_{20}, \lambda_{21}$ one finds $\mathcal{E}(1) = \mathcal{E}(21) \approx 0.78$ and $\mathcal{E}(2) = \mathcal{E}(20) \approx 0.58$. Thus $j = 19$ or 20 is a good choice.

If $T^{1:j}$ does not have an eigenvalue in $\mathcal{I}$ then, by Lemma 6, $T^{j+2:n}$ has an eigenvalue in $\mathcal{I}$ and

$$\beta_{j+1} \left|s^{j+2:n}(j+2)\right| \mathcal{E}(j+1) \approx \left[\left(\sin 2\psi \cdot \frac{\lambda_{l+1} - \lambda_l}{2}\right)^2 + (\theta^{j+2:n} - \rho_n)^2\right]^{1/2}$$

where $\boldsymbol{y}^{(n)} = \boldsymbol{z}_l \cos\psi - \boldsymbol{z}_{l+1} \sin\psi$, $\rho_n = \boldsymbol{y}^{(n)} \cdot T\boldsymbol{y}^{(n)}$.

Note that $\|\boldsymbol{\mathcal{E}}\|_2 = \sqrt{2}$ and the average value of $\boldsymbol{\mathcal{E}}$'s entries is $\sqrt{\frac{2}{n}}$. Theorem 1 tells us to locate $\boldsymbol{\mathcal{E}}$'s maximal entries to get small values of $\beta_j\left|s^{1:j}(j)\right|$. It turns out that the vector $\begin{pmatrix} \boldsymbol{s}^{1:j} \\ \boldsymbol{0} \end{pmatrix}$, for suitable j, is a cheap approximation to $\boldsymbol{y}^{(1)}$. Throughout this section we abbrevaiate $\chi^{1:j}$ by χ_j. For any j that keeps $|\theta^{1:j} - \rho_1|/(\lambda_{l+1} - \lambda_l)$ small, Lemma 8 says

$$y^{(1)}(i+1) = \left[\frac{\chi_i(\rho_1)}{\beta_1 \cdots \beta_i} + \frac{1}{2}\nu_1^2 \frac{\chi_i''(\theta^{1:j})}{\beta_1 \cdots \beta_i} + O\left(\left(\frac{\lambda_{l+1} - \lambda_l}{gap}\right)^3\right)\right] \mathcal{E}(1),$$

whereas

$$s^{(j)}(i+1) = \begin{cases} \chi_i(\theta^{1:j}) s^{(j)}(1)/(\beta_1 \cdots \beta_i), & i < j, \\ 0, & i \geq j. \end{cases}$$

Note that $\theta^{1:j}$ only occurs in the second (and small) term in $y^{(1)}(i+1)$, not the dominant first term. The difference between $\boldsymbol{s}^{(j)}$ and $\boldsymbol{y}^{(1)}$ depends on the two ratios, $(\lambda_{l+1}-\lambda_l)/gap$ and $|\theta^{1:j}-\rho_1|/(\lambda_{l+1}-\lambda_l)$. In the same vein we show (following Lemma 7) that, for the appropriate j values,

$$s^{(j)}(1)=\mathcal{E}(1)\left[1+O\left(\frac{\lambda_{l+1}-\lambda_l}{gap(\theta^{1:j})}\right)\right], \qquad \text{if } \mathcal{E}(1) \text{ is not too small,}$$

where $gap(\theta^{1:j})=\min|\theta-\theta^{1:j}|$ over all eigenvalues θ of $T^{1:j}$ other than $\theta^{1:j}$, and is slightly different from gap defined in (14).

Indeed very good approximations to the eigenvectors $\boldsymbol{z}_l$ and $\boldsymbol{z}_{l+1}$ are given by the internal and the external bisectors of $\boldsymbol{y}^{(1)}$ and $\boldsymbol{y}^{(n)}$. The extremal vectors $\boldsymbol{y}^{(1)}$ and $\boldsymbol{y}^{(n)}$ are not quite orthogonal:

$$\boldsymbol{y}^{(1)}\cdot\boldsymbol{y}^{(n)}\approx(\lambda_{l+1}-\lambda_l)\sum_{i\neq l,l+1}(\lambda_l-\lambda_i)^{-1}.$$

Proofs

To establish all these results in a simple way a sequence of lemmata will prove useful. Recall that $\mathcal{S}_{\mathcal{I}}=span\{\boldsymbol{z}_l,\ldots,\boldsymbol{z}_{l+m-1}\}$.

Lemma 2

$$\mathcal{E}(j)^2=\sum_{i=l}^{l+m-1} z_i(j)^2.$$

Proof:

$$\begin{aligned}
\mathcal{E}(j)^2 &= \max v(j)^2\\
&= \max_{\boldsymbol{\gamma}}\left\{\left(\sum_{i=l}^{l+m-1} z_i(j)\gamma_i\right)^2 : \|\gamma\|=1\right\}\\
&\leq \max_{\boldsymbol{\gamma}}\left(\sum_{i=l}^{l+m-1}\gamma_i^2\cdot\sum_{i=l}^{l+m-1} z_i(j)^2\right)=\sum_{i=l}^{l+m-1} z_i(j)^2,
\end{aligned}$$

Equality is attained when $\boldsymbol{\gamma}$ is a multiple of $(z_l(j),\ldots,z_{l+m-1}(j))$. □

Corollary 1

$$\|\mathcal{E}\ \|^2 = \sum_{j=1}^{n}\sum_{i=l}^{l+m-1} z_i(j)^2 = \sum_{i=l}^{l+m-1}\sum_{j=1}^{n} z_i(j)^2 = \sum_{i=l}^{l+m-1} 1 = m.$$

Lemma 3 *With $\rho_1 = \boldsymbol{y}^{(1)} \cdot T\boldsymbol{y}^{(1)} = \lambda_l \cos^2\varphi + \lambda_{l+1}\sin^2\varphi$, then for any ξ,*

$$\begin{aligned} \nu(\xi)^2 &:= \|T\boldsymbol{y}^{(1)} - \boldsymbol{y}^{(1)}\xi\|^2 \\ &= (\lambda_l - \xi)^2\cos^2\varphi + (\lambda_{l+1} - \xi)^2\sin^2\varphi \\ &= \nu_1^2 + (\xi - \rho_1)^2, \\ \nu_1 &= \sin 2\varphi \cdot \left(\frac{\lambda_{l+1} - \lambda_l}{2}\right). \end{aligned}$$

Proof.

$$\begin{aligned} \nu(\xi) &= (\lambda_l - \rho_1 + \rho_1 - \xi)^2\cos^2\varphi + (\lambda_{l+1} - \rho_1 + \rho_1 - \xi)^2\sin^2\varphi \\ &= \nu_1^2 + 2(\rho_1 - \xi)[\lambda_l\cos^2\varphi + \lambda_{l+1}\sin^2\varphi - \rho_1] + (\rho_1 - \xi)^2 \\ &= \nu_1^2 + (\rho_1 - \xi)^2. \end{aligned}$$

□

Lemma 4 *For any polynomial χ of degree d with an isolated zero θ*

$$\chi(\xi) = (\xi - \theta)\chi'(\theta)\left\{1 + \frac{1}{2}\frac{(\xi - \theta)}{gap(\theta)} \cdot \left(\frac{gap(\theta)\chi''(\theta)}{\chi'(\theta)}\right) + O\left(\frac{\xi - \theta}{gap(\theta)}\right)^2\right\}$$

where $gap(\theta)$ equals to the separation of θ from the rest of χ's zeros. Also

$$\left|\frac{gap(\theta)\chi''(\theta)}{\chi'(\theta)}\right| < 2(d-1).$$

Proof. A Taylor series expansion of χ around θ yields

$$\chi(\xi) = 0 + (\xi - \theta)\chi'(\theta) + \frac{1}{2}(\xi - \theta)^2\chi''(\theta) + \frac{1}{6}(\xi - \theta)^3\chi'''(\eta)$$

for some η in the convex hull of ξ and θ. There is a simple expression for $\chi''(\theta)/\chi'(\theta)$ if we denote the zeros of χ by $\theta_1, \ldots, \theta_d$ and $\theta = \theta_j$. By logarithmic differentiation,

$$\frac{\chi''(\theta_j)}{\chi'(\theta_j)} = 2\sum_{k \neq j}(\theta_j - \theta_k)^{-1}$$

whence

$$\left|\frac{gap(\theta_j)\chi''(\theta_j)}{\chi'(\theta_j)}\right| = 2\sum_{k \neq j}\frac{\theta_j - \theta_{j\pm 1}}{\theta_j - \theta_k} < 2(d-1).$$

□

Remark. For evenly spaced zeros the upper bound is $2\ln(d-1)$ not $2(d-1)$.

Lemma 5 *For each j, $1 \leq j \leq n$, either $\chi^{1:j-1}$ has a zero in $[\lambda_l, \lambda_{l+1}]$ or $\chi^{j+1:n}$ has a zero in $[\lambda_l, \lambda_{l+1}]$.*

Proof. The characteristic polynomial of the submatrix of T obtained by deleting row and column j is $\chi^{1:j-1}(\lambda)\chi^{j+1:n}(\lambda)$. By Cauchy's interlace theorem one of these polynomials (at least) has a zero in $[\lambda_i, \lambda_{i+1}]$ for $i = 1, \ldots, n-1$. It can be shown that if (and only if) both polynomials have a zero in $[\lambda_i, \lambda_{i+1}]$ then it is either λ_i or λ_{i+1}. □

Now we can establish the principal result of this section.

Theorem 1 *With the notation developed in this section, if χ_j has a zero $\theta^{1:j}$ in $(\lambda_l, \lambda_{l+1})$ then*

$$\begin{aligned}\beta_j|s^{(j)}(j)|\mathcal{E}(j+1) &= \frac{\mathcal{E}(1)}{s^{(j)}(1)}\left[\left(\sin 2\varphi \cdot \frac{\lambda_{l+1} - \lambda_l}{2}\right)^2 + \left(\theta^{1:j} - \rho_1\right)^2\right]^{1/2} \\ &\quad \cdot\left\{1 + O\left(\frac{\lambda_{l+1} - \lambda_l}{gap(\theta^{1:j})}\right)\right\}.\end{aligned}$$

Proof. By Lemma 2 and (4) in Section 3

$$\begin{aligned}\mathcal{E}(j+1)^2 &= z_l(j+1)^2 + z_{l+1}(j+1)^2, \\ &= \left[z_l(1)^2\chi_j(\lambda_l)^2 + z_{l+1}(1)^2\chi_j(\lambda_{l+1})^2\right]/(\beta_1 \cdots \beta_j)^2.\end{aligned}$$

Square Lemma 4, for $\chi^{1:j} = \chi_j$ and its zero $\theta^{1:j}$

$$\begin{aligned}[\beta_1\cdots\beta_j\,\mathcal{E}(j+1)]^2 &= \chi_j'(\theta^{1:j})^2 z_l(1)^2(\lambda_l-\theta^{1:j})^2\left[1+O\left(\frac{\lambda_l-\theta^{1:j}}{gap(\theta^{1:j})}\right)\right]\\ &+ \chi_j'(\theta^{1:j})^2 z_{l+1}(1)^2(\lambda_{l+1}-\theta^{1:j})^2\left[1+O\left(\frac{\lambda_{l+1}-\theta^{1:j}}{gap(\theta^{1:j})}\right)\right]\end{aligned}$$

and

$$\begin{aligned}\beta_1 \quad \cdots \quad &\beta_j\,\mathcal{E}(j+1) = \left|\chi_j'(\theta^{1:j})\right|\\ \cdot\quad &\left\{\left[z_l(1)^2(\lambda_l-\theta^{1:j})^2 + z_{l+1}(1)^2(\lambda_{l+1}-\theta^{1:j})^2\right]^{1/2}\left[1+O\left(\frac{\lambda_{l+1}-\lambda_l}{gap(\theta^{1:j})}\right)\right]\right\}.\end{aligned}$$

By Lemma 3,

$$\begin{aligned}\beta_1\cdots\beta_j\,\mathcal{E}(j+1) &= \left|\chi_j'(\theta^{1:j})\right|\mathcal{E}(1)\nu\left(\theta^{1:j}\right)\left[1+O\left(\frac{\lambda_{l+1}-\lambda_l}{gap(\theta^{1:j})}\right)\right]\\ &= \left|\chi_j'(\theta^{1:j})\right|\mathcal{E}(1)\left[\nu_1^2+(\theta^{1:j}-\rho_1)^2\right]^{1/2}\\ &\cdot\quad\left[1+O\left(\frac{\lambda_{l+1}-\lambda_l}{gap(\theta^{1:j})}\right)\right].\end{aligned}$$

To relate $\mathcal{E}(j+1)$ to the submatrix $T^{1:j}$ and $\theta^{1:j}$'s eigenvector $s^{1:j}$, recall (6) in Section 3

$$s^{(j)}(1)s^{(j)}(j)\chi_j'(\theta^{1:j}) = \beta_1\cdots\beta_{j-1}.$$

Substitute this expression into the equation above and cancel the nonzero derivative to obtain

$$\beta_j\left|s^{(j)}(j)\right|\mathcal{E}(j+1) = \frac{\mathcal{E}(1)}{s^{(j)}(1)}\nu\left(\theta^{1:j}\right)\left[1+O\left(\frac{\lambda_{l+1}-\lambda_l}{gap(\theta^{1:j})}\right)\right]$$

as claimed. □

Lemma 6 *For all j such that $\mathcal{E}(j+1)$ is nearly maximal and χ_j has a zero in $[\lambda_l, \lambda_{l+1}]$*

$$\begin{aligned}s^{(j)}(1)^2 &= \left[\mathcal{E}(1)^2+\frac{\mathcal{E}(1)}{\mathcal{E}(j+1)}\cdot\tau\nu\left(\theta^{1:j}\right)\,sign\left(s^{(j)}(j)\right)\right]\left[1+O\left(\frac{\lambda_{l+1}-\lambda_l}{gap(\theta^{1:j})}\right)\right],\\ gap(j) &= \min\{\lambda_{l+2}-\theta^{1:j},\theta^{1:j}-\lambda_{l-1}\},\quad \tau=\tau(\theta^{1:j})\quad \textit{is given in } (15).\end{aligned}$$

Proof. To analyze $\boldsymbol{s}^j$ in terms of T rather than $T^{1:j}$ note that

$$\left(T - \theta^{1:j} I\right) \begin{pmatrix} \boldsymbol{s}^{(j)} \\ \boldsymbol{0} \end{pmatrix} = \boldsymbol{e}_{j+1} \beta_j s^{(j)}(j).$$

Use $T = Z \Lambda Z^t$ to obtain

$$\begin{pmatrix} \boldsymbol{s}^{(j)} \\ \boldsymbol{0} \end{pmatrix} = Z \left(\Lambda - \theta^{1:j} I\right)^{-1} Z^t \boldsymbol{e}_{j+1} \beta_j s^{(j)}(j)$$

and

$$s^{(j)}(1) = \sum_{k=1}^{n} \frac{z_k(1) z_k(j+1)}{\lambda_k - \theta^{1:j}} \beta_j s^{(j)}(j)$$

is taken positive. The terms corresponding to $k = l, l+1$ usually dominate the sum and it may be written

$$s^{(j)}(1) = \left(\frac{z_l(1) z_l(j+1)}{\lambda_l - \theta^{1:j}} + \frac{z_{l+1}(1) z_{l+1}(j+1)}{\lambda_{l+1} - \theta^{1:j}} + \tau \right) \beta_j s^{(j)}(j) \qquad (15)$$

and

$$\tau = \sum_{k \neq l, l+1} \frac{z_k(1) z_k(j+1)}{\lambda_k - \theta^{1:j}}.$$

The key fact is that the first two terms in (15) combine to give $\mathcal{E}(1)^2$. Use (15) to obtain

$$\begin{aligned} & \frac{z_l(1) z_l(j+1)}{\lambda_l - \theta^{1:j}} + \frac{z_{l+1}(1) z_{l+1}(j+1)}{\lambda_{l+1} - \theta^{1:j}} = \\ & \left(z_l(1)^2 \chi_j'(\theta^{1:j}) + z_{l+1}(1)^2 \chi_j'(\theta^{1:j})\right) \cdot \left[1 + O\left(\frac{\lambda_{l+1} - \lambda_l}{gap(\theta^{1:j})}\right)\right] \frac{1}{\beta_1 \cdots \beta_j} \\ = \quad & \mathcal{E}(1)^2 \cdot \frac{\chi_j'(\theta^{1:j})}{\beta_1 \cdots \beta_j} \cdot \left[1 + O\left(\frac{\lambda_{l+1} - \lambda_l}{gap(\theta^{1:j})}\right)\right]. \end{aligned} \qquad (16)$$

Now (6) and Theorem 1 give

$$\left| \frac{\beta_1 \cdots \beta_j}{\chi'(\theta^{1:j})} \right| = \left| s^{(j)}(1) s^{(j)}(j) \beta_j \right| = \frac{\mathcal{E}(1)}{\mathcal{E}(j+1)} \nu\left(\theta^{1:j}\right) \cdot \left[1 + O\left(\frac{\lambda_{l+1} - \lambda_l}{gap(\theta^{1:j})}\right)\right]. \qquad (17)$$

Multiply (15) by $s^{(j)}(1)$ and then use (16) to find

$$s^{(j)}(1)^2 = \mathcal{E}(1)^2 \frac{\chi'(\theta^{1:j})s^{(j)}(1)s^{(j)}(j)}{\beta_1 \cdots \beta_{j-1}} \cdot \left[1 + O\left(\frac{\lambda_{l+1} - \lambda_l}{gap(\theta^{1:j})}\right)\right] + \tau s^{(j)}(1)s^{(j)}(j)\beta_j.$$

Now use (17) to simplify both terms

$$\begin{aligned} s^{(j)}(1)^2 &= \mathcal{E}(1)^2 \left[1 + O\left(\frac{\lambda_{l+1} - \lambda_l}{gap(\theta^{1:j})}\right)\right] \\ &+ \frac{\mathcal{E}(1)}{\mathcal{E}(j+1)}\nu\left(\theta^{1:j}\right) \cdot \tau \,\mathrm{sign}\left(s^{(j)}(j)\right) \left[1 + O\left(\frac{\lambda_{l+1} - \lambda_l}{gap(\theta^{1:j})}\right)\right] \\ &= \left[\mathcal{E}(1)^2 + \mathcal{E}(1)\nu\left(\theta^{1:j}\right)\frac{\tau \,\mathrm{sign}\left(s^{(j)}(j)\right)}{\mathcal{E}(j+1)}\right]\left[1 + O\left(\frac{\lambda_{l+1} - \lambda_l}{gap(\theta^{1:j})}\right)\right]. \end{aligned}$$

□

For well chosen j the first term $\mathcal{E}(1)^2$ usually dominates the second. When $\mathcal{E}(1)$ is small then j must be large to ensure that $\theta^{1:j}$ lies in $[\lambda_l, \lambda_{l+1}]$ and it is possible that $\mathcal{E}(1)^2$ always dominates but we have not proved that yet.

Note that

(i) $\nu\left(\theta^{1:j}\right) \cdot |\tau| \le \sum_{k \neq l, l+1} |z_k(1)z_k(j+1)|\, \nu\left(\theta^{1:j}\right)/gap(j) < \nu\left(\theta^{1:j}\right)/gap(j)$ is assumed to be very small.

(ii) $\mathcal{E}(j+1) \ge \sqrt{\frac{2}{n}}$ (above average, by choice of j).

(iii) By Lemma 7, (i) and (ii), whenever

$$\mathcal{E}(1)\mathcal{E}(j+1) > 2\nu\left(\theta^{1:j}\right)/gap(j)$$

then

$$s^{(j)}(1) = \mathcal{E}(1)\left(1 + \frac{\nu\, \tau \,\mathrm{sign}\left(s^{(j)}(j)\right)}{\mathcal{E}(1)\mathcal{E}(j+1)}\right)\left[1 + O\left(\frac{\lambda_{l+1} - \lambda_l}{gap(\theta^{1:j})}\right)\right]$$

i. e. $s^{(j)}(1) \approx \mathcal{E}(1)$.

In Figures 3 and 4 we show typical envelopes.

Figure 3 shows snapshots of the envelope of a cluster of 108 close eigenvalues, each differing from its neighbors by $O(\epsilon \cdot \mathit{spread})$, from a matrix of order $n = 2053$ supplied by George Fann, Pacific Northwest Laboratory, Richland, WA. The matrix arose in the application of the self-consistent field Hartree-Fock method for solving the nonlinear Schroedinger equation for Zeolite ZSM-5. The top shows the first 50 entries in $\mathcal{E}$ and the bottom shows entries 470–550. The cluster is completely determined by submatrix (1:515). The supports of all the eigenvectors belonging to this cluster lie in (1:515).

Figure 4 shows the envelope of $span(\boldsymbol{z}_{12}, \boldsymbol{z}_{13})$ for W_{21}^{+}, see Section 5. The eigenvalues are close to 6.

The 'humps' in most envelopes are unimodal but this case is an exception. As shown in Section 5 either peak (4 or 6, 16 or 18) will serve for choosing a submatrix. However it is vital to realize that indices 4 and 6 belong to the same hump.

Extremal Vectors in $\mathcal{S}_{\mathcal{I}}$

Lemma 7 *For any $\theta \in (\lambda_l, \lambda_{l+1})$*

$$\begin{aligned} (\beta_1 \cdots \beta_i)\frac{y^{(1)}(i+1)}{y^{(1)}(1)} &= \chi_i(\rho_1) + \frac{1}{2}\nu_1^2 \chi_i''(\theta) + O\left((\lambda_{l+1} - \lambda_l)^3\right), \\ \rho_1 &= \boldsymbol{y}^{(1)} \cdot T\boldsymbol{y}^{(1)}, \\ \nu_1 &= \sin 2\varphi \cdot \frac{\lambda_{l+1} - \lambda_l}{2}. \end{aligned}$$

Proof. From the beginning of this section

$$\boldsymbol{y}^{(1)} = \boldsymbol{z}_l \cos\varphi + \boldsymbol{z}_{l+1} \sin\varphi, \quad \varphi \in [0, \pi/2],$$

By (4) in Section 3

$$y^{(1)}(i+1) = \frac{z_l(1)\chi_i(\lambda_l)}{\beta_1 \cdots \beta_i} \cdot \cos\varphi + \frac{z_{l+1}(1)\chi_i(\lambda_{l+1})}{\beta_1 \cdots \beta_i} \cdot \sin\varphi.$$

Since $\cos\varphi = z_l(1)/\mathcal{E}(1)$ and $\mathcal{E}(1) = y^{(1)}(1)$,

$$\frac{y^{(1)}(i+1)}{y^{(1)}(1)} = \frac{\chi_i(\lambda_l)\cos^2\varphi + \chi_i(\lambda_{l+1})\sin^2\varphi}{\beta_1 \cdots \beta_i}.$$

Expand χ_i about any $\theta \in (\lambda_l, \lambda_{l+1})$ and use Lemma 3 to find

$$\begin{aligned}\chi_i(\lambda_l)\cos^2\varphi &+ \chi_i(\lambda_{l+1})\sin^2\varphi = \chi_i(\theta)(\cos^2\varphi + \sin^2\varphi) \\ &+ \chi_i'(\theta)[\cos^2\varphi\,(\lambda_l - \theta) + \sin^2\varphi\,(\lambda_{l+1} - \theta)] \\ &+ \frac{1}{2}\chi_i''(\theta)\nu\,(\theta)^2 + O\left((\lambda_{l+1} - \lambda_l)^3\right).\end{aligned}$$

The coefficient of $\chi_i'(\theta)$ is just $\rho_1 - \theta$ and part of $\nu\,(\theta)^2$ is $(\rho_1 - \theta)^2$. Thus the right side simplifies to

$$\chi_i(\rho_1) + \frac{1}{2}\nu_1^2\chi_i''(\theta) + O\left((\lambda_{l+1} - \lambda_l)^3\right).$$

as claimed. □

For comparison

$$(\beta_1 \cdots \beta_i)\frac{s^{(j)}(i+1)}{s^{(j)}(1)} = \begin{cases} \chi_i(\theta^{1:j}), & i < j, \\ 0, & i \geq j. \end{cases}$$

The closer is $\theta^{1:j}$ to ρ_1 the closer is $\begin{pmatrix} s^{(j)} \\ 0 \end{pmatrix}$ to $\boldsymbol{y}^{(1)}$. Let us consider the magnitude of $y^{(1)}(j+1)$ which we approximate by 0. By (6) in Section 3

$$\frac{\chi_j'(\theta^{1:j})}{\beta_1 \cdots \beta_j} = \frac{1}{s^{(j)}(1)s^{(j)}(j)\beta_j}. \tag{18}$$

Hence, with $\theta = \theta^{1:j}$ in Lemma 7,

$$\begin{aligned}\frac{y^{(1)}(j+1)}{y^{(1)}(1)} &= \frac{\rho_1 - \theta^{1:j}}{s^{(j)}(1)s^{(j)}(j)\beta_j} \\ &+ \frac{1}{2}\frac{\nu_1}{gap(\theta^{1:j})} \cdot \frac{\nu_1\chi_j'(\theta^{1:j})}{\beta_1 \cdots \beta_j} \cdot \frac{gap(\theta^{1:j})\chi_j''(\theta^{1:j})}{\chi_j'(\theta^{1:j})} \\ &+ \text{ higher order terms.}\end{aligned}$$

By Theorem 1, for the best values of j,

$$\beta_j s^{(j)}(j)\mathcal{E}(j+1) \approx \nu\left(\theta^{1:j}\right) = \left[\nu_1^2 + (\rho_1 - \theta^{1:j})^2\right]^{1/2}.$$

Recall that $y^{(j)}(1) = \mathcal{E}(1)$ and $\mathcal{E}(1) \leq s^{(j)}(1)$ (but $\approx$ if $\mathcal{E}(1)$ is not small). Thus

$$\begin{aligned} y^{(1)}(j+1) &= \left[\frac{\rho_1 - \theta^{1:j}}{\nu\left(\theta^{1:j}\right)} \cdot \mathcal{E}(j+1)\right. \\ &+ \frac{1}{2} \cdot \frac{\nu_1}{gap(\theta^{1:j})} \cdot \frac{\nu_1}{\nu\left(\theta^{1:j}\right)} \cdot \mathcal{E}(j+1) M_2 \\ &+ \left. \text{higher order terms}\right] \frac{\mathcal{E}(1)}{s^{(j)}(1)} \\ &\leq \mathcal{E}(j+1) \cdot \left[\frac{\rho_1 - \theta^{1:j}}{\nu\left(\theta^{1:j}\right)} + \frac{1}{2} \cdot \frac{\nu_1}{gap(\theta^{1:j})} \cdot M_2\right] + \ldots \qquad (19) \end{aligned}$$

where

$$M_2 = \left| \frac{gap(\theta^{1:j}) \chi_j''(\theta^{1:j})}{\chi_j'(\theta^{1:j})} \right| < 2(j-1).$$

The relation (19) tells us that it is essential that both ratios

$$\frac{\rho_1 - \theta^{1:j}}{\nu\left(\theta^{1:j}\right)} \text{ and } \frac{\nu_1}{gap(\theta^{1:j})} = \frac{(\lambda_{l+1} - \lambda_l) \sin 2\varphi}{2\, gap(\theta^{1:j})}$$

be small.

Next consider $\boldsymbol{y}^{(1)} \cdot \boldsymbol{y}^{(n)}$. Recall that

$$\boldsymbol{y}^{(1)} = \boldsymbol{z}_l \cos\varphi + \boldsymbol{z}_{l+1} \sin\varphi, \quad \boldsymbol{y}^{(n)} = \boldsymbol{z}_l \cos\psi - \boldsymbol{z}_{l+1} \sin\psi,$$

where $\varphi \in (0, \pi/2)$. This choice of sign for $\boldsymbol{y}^{(n)}$ is deliberate and yields $\psi \in (0, \pi/2)$. Since $\boldsymbol{z}_l \cdot \boldsymbol{z}_{l+1} = 0$

$$\boldsymbol{y}^{(1)} \cdot \boldsymbol{y}^{(n)} = \cos\psi \cos\varphi - \sin\psi \sin\varphi = \cos(\psi + \varphi).$$

By (6) in Section 3,

$$z_l(1) z_l(n) \chi'(\lambda_l) = \beta_1 \cdots \beta_{n-1} = z_{l+1}(1) z_{l+1}(n) \chi'(\lambda_{l+1})$$

where $\chi = \chi^{1:n}$. Hence

$$\begin{aligned} \tan\psi &= \frac{-z_{l+1}(n)}{z_l(n)} = \frac{-z_l(1)}{z_{l+1}(1)} \cdot \frac{\chi'(\lambda_l)}{\chi'(\lambda_{l+1})} \\ &= -\frac{1}{\tan\varphi} \cdot \frac{(\lambda_l - \lambda_{l+1}) \prod_i''(\lambda_l - \lambda_i)}{(\lambda_{l+1} - \lambda_l) \prod_i''(\lambda_{l+1} - \lambda_i)} \end{aligned}$$

where $\prod'' = \prod_{i=1, i\neq l,l+1}^{n}$. So

$$\begin{aligned}
\tan\varphi \cdot tan\psi &= \prod^{''} \left(1 + \frac{\lambda_{l+1} - \lambda_l}{\lambda_l - \lambda_i}\right) \\
&= 1 + (\lambda_{l+1} - \lambda_l) \sum^{''} (\lambda_l - \lambda_i)^{-1} + O\left((\lambda_{l+1} - \lambda_l)^2\right) \\
1 - \tan\varphi \cdot tan\psi &= O\left(\frac{\lambda_{l+1} - \lambda_l}{gap}\right), \\
gap &= \min |\lambda_i - \lambda_l|, \ i \neq l, l+1.
\end{aligned}$$

This establishes

Lemma 8

$$\boldsymbol{y}^{(1)} \cdot \boldsymbol{y}^{(n)} = \cos\psi \cos\varphi \,(1 - \tan\varphi \cdot tan\psi) = O\left(\frac{\lambda_{l+1} - \lambda_l}{gap}\right).$$

7 The Submatrix Theorems

We consider the case when a pair of adjacent eigenvalues, call them λ_- and λ_+ are much closer together than they are to any other eigenvalues.

We claim that there is a submatrix $T^{1:j}$ with two properties

(a) It has a well isolated eigenvalue θ in (λ_-, λ_+).

(b) The normalized eigenvector s for θ has the property that $\begin{pmatrix} s \\ 0 \end{pmatrix}$ is very close to the invariant subspace (under T) associated with λ_- and λ_+.

If we apply this property to the trailing submatrices $T_{k,n}$, for some k, we obtain another vector $\begin{pmatrix} 0 \\ t \end{pmatrix}$ that is also very close to the desired invariant subspace for $\lambda_\pm$. Thus $\begin{pmatrix} s \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ t \end{pmatrix}$ make an excellent basis for this subspace.

The smaller is $\lambda_+ - \lambda_-$, the smaller is the dot product $\begin{pmatrix} s \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ t \end{pmatrix}$.

We now prepare to make precise the preceeding claims.

Paige's Persistence theorem (in Section 3) says that there is at least one Ritz value $\theta_\bullet^{(j+l)}$ in the Ritz interval

$$\mathcal{I}_i^{(j)} := [\theta_i^{(j)} - \beta_j\omega_i,\, \theta_i^{(j)} + \beta_j\omega_i], \quad \omega_i = \omega_i^{(j)} = \left|s_i^{(j)}(j)\right|. \tag{20}$$

for all $l = 1, 2, \ldots, n-j$. When $\beta_j\omega_i$ is relatively large, as happens when j is small, then $\mathcal{I}_i^{(j)}$ may contain several eigenvalues. However as ω_i decreases then $\mathcal{I}_i^{(j)}$ becomes disjoint from all other $\mathcal{I}_m^{(j)}$ and, as j increases, shrinks onto a single eigenvalue.

Our interest now centers on those rare cases when $\mathcal{I}_i^{(j)}$ is isolated from its neighbors and small and yet 2 eigenvalues get into $\mathcal{I}_i^{(j)}$. How close can they be to each other? We will show that $\beta_j\omega_i$ acts as a barrier in the sense that the two values, λ_- and λ_+, satisfy

$$\lambda_+ - \lambda_- > \mathcal{O}(\beta_j\omega_i).$$

So, if $\lambda_+ - \lambda_-$ is tiny then the vector $\begin{pmatrix} s_i^{(j)} \\ 0 \end{pmatrix}$ has a small residual norm since

$$\|T\begin{pmatrix} s_i^{(j)} \\ 0 \end{pmatrix} - \begin{pmatrix} s_i^{(j)} \\ 0 \end{pmatrix}\theta_i^{(j)}\| = \beta_j\omega_i.$$

More notation is needed to state Theorem 2.

Secular Equation

For some $j < n$, write

$$\begin{pmatrix} T^{1:j} & \bullet & \\ \bullet & \alpha_{j+1} & \bullet \\ & \bullet & T^{j+2:n} \end{pmatrix}, \qquad \bullet = \beta_j \text{ or } \beta_{j+1}.$$

If ζ is not an eigenvalue of $T^{1:j}$ nor of $T^{j+2:n}$ then

$$\begin{aligned} det(T - \zeta I) \;&=\; det(T^{1:j} - \zeta I)[\alpha_{j+1} - \zeta - \beta_j^2 e_j^t (T^{1:j} - \zeta I)^{-1} e_j \\ &\quad -\beta_{j+1}^2 e_1^t (T^{j+2:n} - \zeta I)^{-1} e_1] \cdot det[T^{j+2:n} - \zeta I]. \end{aligned}$$

Consequently the eigenvalues of T that are not eigenvalues of $T^{1:j}$ nor of $T^{j+2:n}$ must satisfy the nonlinear secular equation

$$\sigma(\lambda) := \alpha_{j+1} - \lambda - \beta_j^2 e_j^t (T^{1:j} - \lambda I)^{-1} e_j - \beta_{j+1}^2 e_1^t (T^{j+2:n} - \lambda I)^{-1} e_1 = 0.$$

The middle term will be written as

$$\beta_j^2 e_j^t (T^{1:j} - \zeta I)^{-1} e_j = \frac{\beta_j^2 \omega_l^2}{\theta_l^{(j)} - \zeta} + \psi_l^{(j)}(\zeta)$$

where $\omega_l = \omega_l^{(j)} = |s_l^{(j)}(j)|$ and $\theta_l^{(j)} \in (\lambda_-, \lambda_+)$, and

$$\psi_l^{(j)}(\zeta) := \beta_j^2 \sum_{i \neq l} \frac{s_i^{(j)}(j)^2}{\theta_i^{(j)} - \zeta} \tag{21}$$

and the final term as

$$\beta_{j+1}^2 e_1^t (T_{j+2,n} - \zeta I)^{-1} e_1 = \tau_{j+2,n}(\zeta). \tag{22}$$

The following results bound the residual norm $\beta_j \omega_l^{(j)}$ in terms of the separation of $\theta_l^{(j)}$ from the other zeros of χ_j. Sometimes, these gaps are greater than separation of $\theta_l^{(j)}$ from eigenvalues λ other than $\lambda_\pm$.

Theorem 2 (Double Occupancy) *Let T_n denote a symmetric unreduced tridiagonal matrix. Let λ_- and λ_+ be two adjacent eigenvalues. The notation developed in this section is in force. Consider those indices j, $1 < j < n$, that satisfy the hypothesis*

(H) *There is a single Ritz value $\theta_l^{(j)}$ in the open interval (λ_-, λ_+) and neither end point is a Ritz value.*

For such j test the condition for double occupancy of $\mathcal{I}_l^{(j)}$, in terms of (21) and (22):

($\mathcal{DO}$) $\psi_l^{(j)}(\lambda_-) + \tau_{j+2,n}(\lambda_-) < \alpha_{j+1} - \theta_l^{(j)} < \psi_l^{(j)}(\lambda_+) + \tau_{j+2,n}(\lambda_+)$.

If ($\mathcal{DO}$) does not hold (only one eigenvalue in $\mathcal{I}_l^{(j)}$) then
$\lambda_+ - \lambda_- > \beta_j \omega_l^{(j)}$.

If ($\mathcal{DO}$) holds (both eigenvalues in $\mathcal{I}_l^{(j)}$) then
$2\beta_j \omega_l^{(j)} \geq \lambda_+ - \lambda_- \geq 2\beta_j \omega_l^{(j)} / \sqrt{1 + 2G}$,

where

$$G = G_j(\lambda_-, \lambda_+) = \frac{1}{2} \left\{ \frac{\psi_l^{(j)}(\lambda_+) - \psi_l^{(j)}(\lambda_-)}{\lambda_+ - \lambda_-} + \frac{\tau_{j+2,n}(\lambda_+) - \tau_{j+2,n}(\lambda_-)}{\lambda_+ - \lambda_-} \right\}.$$

Remark 1 Theorem 2 does not give lower bounds on $\lambda_+ - \lambda_-$ because G is a function of λ_- and λ_+.

Remark 2 $G_j = \frac{1}{2}\left\{\psi_l^{(j)'}(\eta_1) + \tau'_{j+2,n}(\eta_2)\right\}$ for some values η_1 and η_2 in (λ_-, λ_+). Moreover $\psi_l^{(j)}$ and $\tau_{j+2,n}$ are rational functions that are monotonic increasing between poles. By (H) (λ_-, λ_+) is between poles of $\psi_l^{(j)}$ and also (λ_-, λ_+) is between poles of $\tau_{j+2,n}$. Hence $G_j > 0$ on (λ_-, λ_+). Our interest is in those j for which G_j is smallest.

Proof (of Theorem 2). Partition $T - \zeta I$ into a 3×3 structure with blocks $1 : j, j+1 : j+1, j+2 : n$ and then form the Schur complement $\sigma(\zeta)$ of entry $(j+1, j+1)$. Thus $T - \zeta I$ is congruent to

$$T^{1:j} - \zeta I \oplus \sigma(\zeta) \oplus T^{j+2:n} - \zeta I$$

where

$$\begin{aligned} \sigma(\zeta) = \alpha_{j+1} \quad &- \quad \zeta - \beta_j^2 e_j^t (T^{1:j} - \zeta I)^{-1} e_j \\ &- \quad \beta_{j+1}^2 e_1^t (T^{j+2:n} - \zeta I)^{-1} e_1. \end{aligned}$$

$\sigma = 0$ is sometimes called the secular equation.

Observe that $T^{1:j} \bigoplus T^{j+2:n}$ is the submatrix of T obtained by deleting row and column $j+1$. By Cauchy's Interlace theorem the Ritz values from $T^{1:j}$ and $T^{j+2:n}$ interlace the eigenvalues of T. By (H) (λ_-, λ_+) contains $\theta_l^{(j)}$ and so cannot contain any Ritz value of $T^{j+2:n}$. By Theorem 5 in [5] $\lambda_\pm$ can only be Ritz values if both $\theta_l^{(j)}$ and a Ritz value of $T^{j+2:n}$ coincide at $\lambda_\pm$. Thus (H) rules out this possibility and so

$$[\lambda_-, \lambda_+], \text{ the closed interval, contains no Ritz values of } T^{j+2:n}. \tag{23}$$

The set of indices j that satisfy (H) includes $j = n-1$, by Cauchy's Interlace theorem and so is nonempty. When j is too small then it is a Ritz value of $T^{j+2:n}$ that lies in (λ_-, λ_+) and not $\theta_l^{(j)}$.

Let $\theta_l^{(j)}$, $1 \leq l \leq j$ denote the Ritz value in (λ_-, λ_+) and separate it from the other j-level Ritz values. Thus

$$\beta_j^2 e_j^t (T^{1:j} - \zeta I)^{-1} e_j = \frac{\beta_j^2 \omega_l^2}{\theta_l^{(j)} - \zeta} + \psi_l^{(j)}(\zeta)$$

where $\psi_l^{(j)}(\zeta)$ is defined by (21).

Recall that $(s_i^{(j)}(1), \ldots, s_i^{(j)}(j))$ is a normalized Ritz vector for $\theta_i^{(j)}$ and $\omega_l = \omega_l^{(j)} = |s_l^{(j)}(j)|$. The final term in $\sigma(\zeta)$ is $\tau_{j+2,n}(\zeta)$ and is defined in (22).

By (H) λ_- and λ_+ are not eigenvalues of T_j. By (23) they are not eigenvalues of $T_{j+2,n}$. Consequently λ_- and λ_+ must satisfy the secular equation

$$\sigma(\lambda_\pm) := \alpha_{j+1} - \lambda_\pm - \frac{\beta_j^2 \omega_l^{(j)2}}{\theta_l^{(j)} - \lambda_\pm} - \psi_l^{(j)}(\lambda_\pm) - \tau_{j+2,n}(\lambda_\pm) = 0.$$

To simplify the analysis that follows write

$$\theta = \theta_l^{(j)}, \quad \omega = \omega_l^{(j)}, \quad \beta = \beta_j, \quad \alpha = \alpha_{j+1}, \quad \psi_l = \psi_l^{(j)}, \quad \tau = \tau_{j+2,n}.$$

Next rewrite the above equations in terms of positive terms $\lambda_+ - \theta$ and $\theta - \lambda_-$ to find

$$(\lambda_+ - \theta)^2 + 2E(\lambda_+ - \theta) = \beta^2\omega^2, \tag{24}$$

$$(\theta - \lambda_-)^2 + 2F(\theta - \lambda_-) = \beta^2\omega^2, \tag{25}$$

where

$$E = E(\lambda_+) = [\psi_l(\lambda_+) + \tau(\lambda_+) + \theta - \alpha]/2,$$
$$F = F(\lambda_-) = [\alpha - \theta - \psi_l(\lambda_-) - \tau(\lambda_-)]/2,$$

and

$$E + F = [\psi_l(\lambda_+) - \psi_l(\lambda_-) + \tau(\lambda_+) - \tau(\lambda_-)]/2.$$

By Remark 2 both ψ_l and τ are monotone increasing in (λ_-, λ_+) and so

$$E + F > 0.$$

The quadratic $x^2 + 2Ex - \beta^2\omega^2 = 0$ has two real roots whose product is $-\beta^2\omega^2$. First we establish the case when $(\mathcal{DO})$ does not hold.

If $E < 0$ then the positive root is the larger (in magnitude) and must exceed $\beta\omega$. Hence

$$\lambda_+ - \theta > \beta\omega, \quad \theta - \lambda_- > 0$$

together imply

$$\lambda_+ - \lambda_- > \beta\omega.$$

Similarly, if $F < 0$, then $\theta - \lambda_- > \beta\omega$, $\lambda_+ - \theta > 0$, and, again

$$\lambda_+ - \lambda_- > \beta\omega,$$

as claimed. Differences smaller than $\beta\omega$ can occur only when both $E = E(\lambda_+) > 0$ and $F = F(\lambda_-) > 0$; that is Condition $(\mathcal{DO})$ holds. This is the case considered in the remaining analysis.

From (24) and (25)

$$\begin{aligned} \lambda_+ - \theta &= \beta^2\omega^2/\{\sqrt{\beta^2\omega^2 + E^2} + E\}, \\ \theta - \lambda_- &= \beta^2\omega^2/\{\sqrt{\beta^2\omega^2 + F^2} + F\}. \end{aligned} \tag{26}$$

Thus

$$\frac{\lambda_+ - \lambda_-}{\beta^2\omega^2} = \frac{1}{\sqrt{\beta^2\omega^2 + E^2} + E} + \frac{1}{\sqrt{\beta^2\omega^2 + F^2} + F}. \tag{27}$$

To simplify this equation note that the function

$$f(x) := (x + \sqrt{x^2 + \beta^2\omega^2})^{-1} = \frac{\sqrt{x^2 + \beta^2\omega^2} - x}{\beta^2\omega^2}$$

is monotone decreasing and concave upward ($f'' > 0$) for $x \geq 0$. By concavity and the positivity of E and F,

$$\frac{f(E) + f(F)}{2} \geq f\left(\frac{E+F}{2}\right),$$

in other words, from (27),

$$\frac{\lambda_+ - \lambda_-}{2\beta^2\omega^2} = \frac{f(E) + f(F)}{2} \geq f\left(\frac{E+F}{2}\right). \tag{28}$$

To simplify this relation write

$$\begin{aligned} \frac{E+F}{2} &= \frac{1}{4}\left[\psi_l(\lambda_+) - \psi_l(\lambda_-) + \tau(\lambda_+) - \tau(\lambda_-)\right] \\ &= \left(\frac{\lambda_+ - \lambda_-}{2}\right) G \end{aligned}$$

where $G = G(\lambda_-, \lambda_+)$ is given in the statement of the theorem.

Let $r = \frac{\lambda_+ - \lambda_-}{2\beta\omega}$ and multiply (28) by $\beta\omega$ to find

$$\begin{aligned} r &\geq \frac{\beta\omega}{\sqrt{\beta^2\omega^2 + \left(\frac{\lambda_+ - \lambda_-}{2}\right)^2 G^2} + \left(\frac{\lambda_+ - \lambda_-}{2}\right) G}, \\ &= \frac{1}{\sqrt{1 + (rG)^2} + rG}, \\ &= \sqrt{1 + (rG)^2} - rG. \end{aligned}$$

Add rG to each side, square and subtract r^2G^2 to find

$$r^2(1 + 2G) \geq 1$$

as claimed. □

Remark 3 In the cases of interest to us $\beta\omega$ is small and

$$G \approx \frac{1}{2}\left(\psi_l'(\theta) + \tau'(\theta)\right).$$

The next task is to turn the inequalities of Theorem 2 into lower bounds on $\lambda_+ - \lambda_-$ by obtaining upper bounds on G. These bounds depend on the gap between $\theta_l^{(j)}$ and the other Ritz values of $T^{1:j}$ and $T^{j+2:n}$.

Theorem 3 *Let λ_- and λ_+ be two consecutive eigenvalues of a symmetric unreduced tridiagonal matrix T_n. The notation developed in this section is in force. For each j, $1 < j < n$, that satisfies the two hypotheses*

(H) *There is a single Ritz value $\theta_l^{(j)}$ in the open interval (λ_-, λ_+)*

(GAP) $2\beta_j\omega_l < gap(l)$, *where*

$$gap(l) = \min\left\{\theta_{l+1}^{(j)} - \theta_l^{(j)}, \theta_l^{(j)} - \theta_{l-1}^{(j)}, |\theta_l^{(j)} - \theta_i^{(j+2,n)}|,\ \ i = 1, n-j-1\right\},$$

$$\omega_l = |s_l^{(j)}(j)|$$

then,

either $(\lambda_-, \lambda_+) \not\subset \mathcal{I}_l^{(j)} := (\theta_l^{(j)} - \beta_j\omega_l, \theta_l^{(j)} + \beta_j\omega_l)$,
in which case $\lambda_+ - \lambda_- > \beta_j\omega_l$

or $(\lambda_-, \lambda_+) \subset \mathcal{I}_l^{(j)}$,
in which case , $\lambda_+ - \lambda_- > 2\beta_j\omega_l / \sqrt{1 + 2\Gamma_j}$,

where

$$\Gamma_j = (\beta_j^2(1 - \omega_l^2) + \beta_{j+1}^2)/2[gap(l) - \beta_j\omega_l]^2.$$

Remark 4 *The observations at the end of Section 3 say that, on average, Γ_j is $O(1)$; the geometric mean of the β's is less than the geometric mean of the gaps $|\theta_l - \theta_i|$, $i \neq l$.*

Proof. It is only necessary to bound the quantity G from Theorem 2 and the same notation will be adopted.

For some $\eta \in (\lambda_-, \lambda_+)$,

$$\begin{aligned} \frac{\psi_l^{(j)}(\lambda_+) - \psi_l^{(j)}(\lambda_-)}{\lambda_+ - \lambda_-} &= \psi_l^{(j)\prime}(\eta) = \beta_j^2 \sum_{i \neq l} \frac{{s_i^{(j)}(j)}^2}{(\theta_i^{(j)} - \eta)^2}, \\ &= \beta_j^2(1 - \omega_l^2)/\mathcal{H}_l^{(j)}, \qquad (29) \end{aligned}$$

where $\mathcal{H}_l^{(j)} = \mathcal{H}_l^{(j)}(\eta)$ is a weighted harmonic mean of $\{(\theta_i^{(j)} - \eta)^2\}$. Since the minimum value never exceeds any mean

$$\begin{aligned} \mathcal{H}_l^{(j)} &\geq \min_{i \neq l}(\theta_i^{(j)} - \eta)^2 \\ &= \min\{(\theta_{l+1}^{(j)} - \eta)^2, (\theta_{l-1}^{(j)} - \eta)^2\}. \end{aligned}$$

There are two cases. If $(\lambda_-, \lambda_+) \not\subset \mathcal{I}_l^{(j)}$ then $\lambda_+ - \lambda_- > \beta_j \omega_l$ for the same reasons given in Theorem 2. In the contrary case, one has

$$\eta \in [\lambda_-, \lambda_+] \subset \mathcal{I}_l^{(j)}.$$

Now (GAP) guarantees that

$$\min\{\theta_{l+1}^{(j)} - \eta, \eta - \theta_{l-1}^{(j)}\} \geq \min\{\theta_{l+1} - \theta_l, \theta_l - \theta_{l-1}\} - \beta\omega \geq \beta\omega.$$

Similarly, for some $\zeta \in (\lambda_-, \lambda_+)$

$$\begin{aligned} \frac{\tau_{j+2,n}(\lambda_+) - \tau_{j+2,n}(\lambda_-)}{\lambda_+ - \lambda_-} &= \tau_{j+2,n}'(\zeta) \\ &= \beta_{j+1}^2 \sum_i \left[\frac{s_i^{(j+2,n)}(i)}{\theta_i^{(j+2,n)} - \zeta}\right]^2 \\ &= \beta_{j+1}^2/\mathcal{H}^{(j+2,n)}, \qquad (30) \end{aligned}$$

where $\mathcal{H}^{(j+2,n)} = \mathcal{H}^{(j+2,n)}(\zeta)$ is a weighted harmonic mean of $\{(\theta_i^{(j+2,n)} - \zeta)^2\}$ and so (GAP) yields

$$\begin{aligned} \mathcal{H}^{(j+2,n)} &\geq \min_i(\theta_i^{(j+2,n)} - \zeta)^2 \\ &\geq (\min_i |\theta_i^{(j+2,n)} - \theta_l| - \beta\omega)^2 \\ &\geq \beta^2\omega^2. \end{aligned}$$

The definition of $gap(l)$ is made so that

$$\mathcal{H}_l^{(j)} \geq (gap(l) - \beta\omega)^2, \quad \mathcal{H}^{(j+2,n)} \geq (gap(l) - \beta\omega)^2,$$

and thus, using (29) and (30),

$$\begin{aligned} 2G &= \psi_l^{(j)'}(\eta) + \tau^{(j+2,n)'}(\zeta) \\ &\leq \frac{\beta_j^2(1 - \omega_l^2) + \beta_{j+1}^2}{(gap(l) - \beta_j\omega_l)^2}, \\ &:= 2\Gamma_j \end{aligned}$$

and application of this inequality into the second inequality of Theorem 2 establishes Theorem 3. □

Remark 5 When we first established Theorems 2 and 3 we were concerned that the gap quantities involved Ritz values $\theta_l^{1:j}$ etc. and not eigenvalues λ_i. However the study of some challenging examples has shown that, in fact, the gaps involving Ritz values do lead to an important extension. In examples with 100 or 200 eigenvalues in a cluster. each separated from its neighbor by about 10 or 20 ulps (units in the last place) combine to form a cluster of nonnegligible width (e. g. 1500 ulps). We find that for each submatrix, taken with its nearest neighbor, the gaps of Theorem 3 are vastly greater than quantities such as $\theta_\bullet^{(j)} - \lambda_{l-1}$ and $\lambda_{l+2} - \theta_\bullet^{(j)}$. In other words, Theorem 3 can be applied, submatrix by submatrix, to clusters whose total width is greater than $O(\epsilon \cdot spread)$.

Recall that the envelope bound in Section 6 shows that there are j-values for which our gap quantities G_j satisfy

$$2G_j \approx \frac{1}{\mathcal{E}(j+1)^2} < \frac{n}{2}.$$

In fact each hump in an envelope contributes 1 to $\|\mathcal{E}\|^2$, on average, and if the hump is spiky $\mathcal{E}(j+1) \approx 1$ while if four consecutive entries contribute to the hump then the greatest of the four exceeds 1/2.

8 The Overlap Theorems

We need expressions for the entries of the vectors constructed to span the invariant subspace corresponding to a close pair of eigenvalues λ_- and λ_+ of $T = T^{1:n}$. Recall that $\chi^{l:m}(\zeta)$ is the characteristic polynomial of the submatrix $T^{l:m}$.

The index m (not unique) is determined so that $T^{1:m}$ has a well isolated Ritz value $\theta^{1:m}$ in (λ_-, λ_+) whose normalized Ritz vector $\boldsymbol{s}^{1:m}$ has a small last entry. More precisely

$$\beta_m |s^{1:m}(m)| = O(\lambda_+ - \lambda_-).$$

Trailing submatrices are used to find a suitable index l (not unique) so that $T^{l:n}$ has a well isolated Ritz value $\theta^{l:n}$ in (λ_-, λ_+) whose normalized Ritz vector $\boldsymbol{s}^{l:n}$ has a small first entry. More precisely

$$\beta_{l-1} |s^{l:n}(l)| = O(\lambda_+ - \lambda_-).$$

The spanning vectors are

$$p = \begin{pmatrix} \boldsymbol{s}^{1:m} \\ 0 \end{pmatrix}, \quad q = \begin{pmatrix} 0 \\ \boldsymbol{s}^{l:n} \end{pmatrix}$$

and the indices $1 < l < n$, $1 < m < n$, play a crucial role in the analysis. To avoid index troubles the entries in $\boldsymbol{s}^{l:n}$ are labelled from l to n, not from 1 to $n - l + 1$.

There are cases in which $m < l$ and then p and q are orthogonal because their supports are disjoint. So we turn to the other cases when

$$1 < l < m < n.$$

and study $|p(j)q(j)|$ for $l < j < m$. We will show that $\boldsymbol{p}$ and $\boldsymbol{q}$ are nearly orthogonal. There are formulae for the magnitudes of $p(i)$ and $q(i)$

$$p(i)^2 = \chi_{1,i-1}(\theta)\chi_{i+1,m}(\theta)/\chi'_{1,m}(\theta), \quad i \leq m,$$

where $\theta = \theta^{1:m}$ is p's Ritz value. Similarly

$$q(i)^2 = \chi_{l,i-1}(\varphi)\chi_{i+1,n}(\varphi)/\chi'_{l,n}(\varphi), \quad i \geq l,$$

where $\varphi = \theta^{l:n}$ is q's Ritz value.

It is readily verified that for tridiagonal matrices

$$\chi_{1\langle i\rangle n}(\zeta) := \chi_{1,i-1}(\zeta)\chi_{i+1,n}(\zeta) \tag{31}$$

is the characteristic polynomial of the submatrix of $T^{1:n}$ obtained by deleting row and column i. Hence, by Cauchy's Interlace theorem the Ritz values of $T^{1:i-1}$ and $T^{i+1:n}$ together interlace (weakly) the Ritz values of $T^{1:n}$. In particular, for each i, the open interval (λ_-, λ_+) contains either a Ritz value of $T^{1:i-1}$ or a Ritz value of $T^{i+1:n}$ but not both. In an exceptional case the open interval is free of Ritz values just when Ritz values of $T^{1:i-1}$ and $T^{i+1:n}$ coincide at either λ_- or λ_+ (but not both) and in that case $p(i)q(i) = 0$.

We can simplify some expressions considerably by the following convention:

If a Ritz value $\theta_i^{j:k} \in [\lambda_-, \lambda_+]$ then the index i is omitted. Further $\theta_+^{j:k}$ denotes the smallest Ritz value of $T^{j:k}$ exceeding λ_+ and $\theta_-^{j:k}$ denotes the largest Ritz value of $T^{j:k}$ less than λ_-.

We need a result from [3].

Theorem 4 *With the notation developed above let $Ts_i = s_i\lambda_i$, $s_i^* s_i = 1$ be an eigenvector equation. Let $T^{\langle j\rangle}$ denote the submatrix obtained from T by deleting row and column j; its spectrum is $\theta_1^{\langle j\rangle} < \ldots < \theta_{n-1}^{\langle j\rangle}$. Then for $1 < i < n$,*

$$s_i(j)^2 < \frac{\lambda_i - \theta_{i-1}^{\langle j\rangle}}{\lambda_i - \lambda_{i-1}} \cdot \frac{\theta_i^{\langle j\rangle} - \lambda_i}{\lambda_{i+1} - \lambda_i} \cdot \frac{\theta_{i+1}^{\langle j\rangle} - \lambda_i}{\lambda_{i+2} - \lambda_i}.$$

The bound for $i = 1$ and $i = n$ is obtained by omitting quotients with out-of-bounds indices.

Proof. For tridiagonal matrices, see [4],

$$s_i(j)^2 = \frac{\chi^{\langle j\rangle}(\lambda_i)}{\chi'(\lambda_i)} = \frac{\prod_{k=1}^{n-1}(\lambda_i - \theta_k^{\langle j\rangle})}{\prod_{m=1,m\neq i}^{n}(\lambda_i - \lambda_m)}.$$

By Cauchy's Interlace theorem

$$\lambda_{k-1} \leq \theta_{k-1}^{\langle j\rangle} \leq \lambda_k \leq \theta_k^{\langle j\rangle}.$$

So, for each $k < i$, the quotient $\frac{\lambda_i-\theta_k^{\langle j\rangle}}{\lambda_i-\lambda_k} < 1$ and, for each $k > i$, the quotient $\frac{\theta_k^{\langle j\rangle}-\lambda_i}{\lambda_{k+1}-\lambda_i} < 1$. As $|k-i|$ increases the quotients become close to 1. By discarding all but the smallest two or three quotients the upper bound is obtained. □

In what follows we shall apply Theorem 4 to submatrices such as $T^{1:m}$ and $T^{l:n}$. One of the principle concerns in the theorem proved below is the location of Ritz values such as $\theta_{\pm}^{1\langle j\rangle m}$ or $\theta_{\pm}^{l\langle j\rangle n}$. By (31) $\theta_{+}^{1\langle j\rangle m}$ is either $\theta_{+}^{1:j-1}$ or $\theta_{+}^{j+1:m}$ and we shall be concerned with both cases. Recall that only one of $\theta_{+}^{1:j-1}$ and $\theta_{+}^{j+1:n}$ lies in $(\lambda_+, \lambda_{++})$ and the other exceeds λ_{++}, the next eigenvalue of T greater than λ_+.

Figure 5 is worth contemplating before reading the Overlap theorem. It shows the intervals in which $\theta_{+}^{1:j-1}$ and $\theta_{+}^{j+1:n}$ will lie as j varies.

Theorem 5 (Overlap) *Let T be $n \times n$, symmetric, unreduced, and tridiagonal. Suppose that adjacent eigenvalues λ_- and λ_+ of T are well enough separated from the remaining spectrum to yield the indices l and m and the vectors $\mathbf{p}$ and $\mathbf{q}$ described at the beginning of the section.*
For each j, $l < j < m$,

$$|p(j)q(j)| < 2^{1/4}\sqrt{\frac{\beta}{gap}}\left(\frac{\lambda_+ - \lambda_-}{gap}\right)^{3/4} + O\left(\frac{\lambda_+ - \lambda_-}{gap}\right)$$

where

$$\begin{aligned} \beta &= \min\{\beta_{j-1}, \beta_j\}, \\ gap &= \min\{gap(l), gap(m)\}, \\ gap(l) &= \min\{\theta^{l:n} - \theta_-^{l:n}, \theta_+^{l:n} - \theta^{l:n}\}, \\ gap(m) &= \min\{\theta^{1:m} - \theta_-^{1:m}, \theta_+^{1:m} - \theta^{1:m}\}. \end{aligned}$$

Proof. First confine attention to those j values such that $\theta^{1:j-1} \in (\lambda_-, \lambda_+)$. Consider the expression for $p(j)^2$ in Theorem 4 using $T^{1:m}$ and extract the three smallest terms in the product to find

$$p(j)^2 \le \begin{cases} \frac{\theta^{1:m}-\theta^{1\langle j\rangle m}}{\theta^{1:m}-\theta^{1:m}_{--}} \cdot \frac{\theta^{1:m}-\theta^{1:j-1}}{\theta^{1:m}-\theta^{1:m}_{-}} \cdot \frac{\theta^{1\langle j\rangle m}_{+}-\theta^{1:m}}{\theta^{1:m}_{+}-\theta^{1:m}}, & \text{if} \quad \theta^{1:j-1} < \theta^{1:m}, \\ \\ \frac{\theta^{1:m}-\theta^{1\langle j\rangle m}}{\theta^{1:m}-\theta^{1:m}_{-}} \cdot \frac{\theta^{1:m}-\theta^{1:j-1}}{\theta^{1:m}_{+}-\theta^{1:m}} \cdot \frac{\theta^{1\langle j\rangle m}-\theta^{1:m}}{\theta^{1:m}_{++}-\theta^{1:m}}, & \text{otherwise.} \end{cases} \tag{32}$$

Thus the middle term itself is bounded by $(\lambda_+ - \lambda_-)/gap(m)$. A smaller bound emerges by considering either a neighboring term from $p(j)^2$ or the smallest term in $q(j)^2$. Without loss of generality we suppose that the closest Ritz values outside (λ_-, λ_+) are on the right. By Cauchy's Interlace theorem the open interval $(\lambda_+, \lambda_{++})$ contains either $\theta_+^{1:j-1}$ or $\theta_+^{j+1:n}$ but not both.

Case 1: $\theta_+^{1:j-1} \in (\lambda_+, \lambda_{++})$.

To obtain a bound better than $\lambda_+ - \lambda_-$ on $|\theta^{1:m} - \theta^{1:j-1}|$ consider $\theta^{1:j-1}$ as an approximation to $\theta^{1:m}$ and apply the Gap theorem for Rayleigh quotients, see [4].

$$|\theta^{1:m} - \theta^{1:j-1}| \le \min\left\{\lambda_+ - \lambda_-,\ \frac{\|\boldsymbol{r}\|^2}{gap(m, j-1)}\right\} \tag{33}$$

where

$$\|\boldsymbol{r}\| = \|(T^{1:m} - \theta^{1:j-1} I_m)\tilde{\boldsymbol{s}}\| = \beta_{j-1}\omega_{j-1},$$

$$\tilde{\boldsymbol{s}} = \begin{pmatrix} \boldsymbol{s}^{1:j-1} \\ 0 \end{pmatrix}, \qquad \omega_{j-1} = |s^{1:j-1}(j-1)|,$$

and

$$gap(m, j-1) = \min\left\{\theta_+^{1:m} - \theta^{1:j-1},\ \theta^{1:j-1} - \theta_-^{1:m}\right\}$$

For future application note that

$$gap(m, j-1) = gap(m)\left[1 + O\left(\frac{\lambda_+ - \lambda_-}{gap(m)}\right)\right] \tag{34}$$

since $\theta^{1:j-1}$ and $\theta^{1:m}$ lie in (λ_-, λ_+). To bound $\|\boldsymbol{r}\|$ we apply the Double Occupancy Theorem, proved above, not to $T^{1:m}$ but to T. Single occupancy guarantees that $\|\boldsymbol{r}\| < \lambda_+ - \lambda_-$ and hence

$$|\theta^{1:m} - \theta^{1:j-1}| \le \frac{(\lambda_+ - \lambda_-)^2}{gap(m)}\left[1 + O\left(\frac{\lambda_+ - \lambda_-}{gap(m)}\right)\right],$$

so that $|p(j)| < (\lambda_+ - \lambda_-)/gap$. This is already tighter than the bound to be established. Double occupancy yields a weaker bound:

$$\|\boldsymbol{r}\|^2 < (1 + 2G_{j-1})(\lambda_+ - \lambda_-)^2 \tag{35}$$

where

$$\begin{aligned} G_{j-1} &= \frac{1}{2}\left\{\psi'_{1:j-1}(\mu) + \tau'_{j+1:n}(\mu)\right\}\left[1 + O\left(\frac{\lambda_+ - \lambda_-}{gap(m)}\right)^2\right], \\ \mu &= \frac{1}{2}(\lambda_+ + \lambda_-). \end{aligned}$$

Moreover, from Remark 3 in Section 7 and (29), (30),

$$\psi'(\mu) < \left(\frac{\beta_{j-1}}{\theta_+^{1:j-1} - \mu}\right)^2, \tag{36}$$

$$\tau'(\mu) < \left(\frac{\beta_j}{\theta_+^{j+1:n} - \mu}\right)^2. \tag{37}$$

In Case 1, $\psi' > \tau'$ and putting (35) into (33) yields

$$\frac{|\theta^{1:m} - \theta^{1:j-1}|}{gap(m)} < \frac{\lambda_+ - \lambda_-}{gap(m)} \min\left\{1, (1 + 2G_{j-1})\frac{\lambda_+ - \lambda_-}{gap(m)}\right\}\left[1 + O\left(\frac{\lambda_+ - \lambda_-}{gap(m)}\right)\right]. \tag{38}$$

Now we must use the right hand term in (32), noting that

$$\frac{\theta_+^{1(j)m} - \theta^{1:m}}{\theta_+^{1:m} - \theta^{1:m}} = \frac{\theta_+^{1:j-1} - \theta^{1:m}}{\theta_+^{1:m} - \theta^{1:m}} = \left(\frac{\theta_+^{1:j-1} - \mu}{\theta_+^{1:m} - \theta^{1:m}}\right)\left[1 + O\left(\frac{\lambda_+ - \lambda_-}{gap(m)}\right)\right]. \tag{39}$$

In the analysis to follow we shall drop the 1 from $1 + 2G_{j-1}$ because it contributes only a higher order term to the bounds. Insert (36) into (38) and multiply by (39) to find

$$\begin{aligned} p(j)^2 &< \frac{|\theta^{1:m} - \theta^{1:j-1}|}{\theta^{1:m} - \theta_-^{1:m}} \cdot \frac{\theta_+^{1(j)m} - \theta^{1:m}}{\theta_+^{1:m} - \theta^{1:m}} \\ &< \frac{(\lambda_+ - \lambda_-)}{gap(m)} \min\left\{1, \frac{2\beta_{j-1}^2}{(\theta_+^{1:j-1} - \mu)^2}\frac{(\lambda_+ - \lambda_-)}{gap(m)}\right\} \end{aligned}$$

$$
\begin{aligned}
& \cdot \left(\frac{\theta_+^{1:j-1}-\mu}{gap(m)}\right)\left[1+O\left(\frac{\lambda_+-\lambda_-}{gap(m)}\right)\right]+O\left[\left(\frac{\lambda_+-\lambda_-}{gap(m)}\right)^2\right] \\
= & \frac{(\lambda_+-\lambda_-)}{gap(m)}\cdot\min\left\{\frac{\theta_+^{1:j-1}-\mu}{gap(m)},\frac{2\beta_{j-1}^2}{\theta_+^{1:j-1}-\mu}\cdot\frac{\lambda_+-\lambda_-}{gap(m)^2}\right\} \\
\cdot & \left[1+O\left(\frac{\lambda_+-\lambda_-}{gap(m)}\right)\right]+O\left[\left(\frac{\lambda_+-\lambda_-}{gap(m)}\right)^2\right]. \qquad (40)
\end{aligned}
$$

Bound the min by the geometric mean to find

$$
\begin{aligned}
p(j)^2 &< \sqrt{2}\,\frac{\lambda_+-\lambda_-}{gap(m)}\,\frac{\beta_{j-1}}{gap(m)}\left(\frac{\lambda_+-\lambda_-}{gap(m)}\right)^{1/2} \\
&\cdot \left[1+O\left(\frac{\lambda_+-\lambda_-}{gap(m)}\right)\right]+O\left[\left(\frac{\lambda_+-\lambda_-}{gap(m)}\right)^2\right]. \qquad (41)
\end{aligned}
$$

Since $q(j)^2 < 1$ the claimed bound holds in Case 1.

Case 2: $\theta_+^{j+1:n} \in (\lambda_+, \lambda_{++})$.

The argument is similar to Case 1 but now it is $q(j)^2$ that offsets a large value for G_{j-1}. In Case 2, $\tau' > \psi'$ and so (38) yields

$$
\begin{aligned}
p(j)^2 &< \frac{(\lambda_+-\lambda_-)}{gap(m)}\min\left\{1,\frac{2\beta_j^2}{(\theta_+^{j+1:n}-\mu)^2}\frac{\lambda_+-\lambda_-}{gap(m)}\right\} \\
&\cdot \left[1+O\left(\frac{\lambda_+-\lambda_-}{gap(m)}\right)\right]+O\left[\left(\frac{\lambda_+-\lambda_-}{gap(m)}\right)^2\right].
\end{aligned}
$$

The smallest term in $q(j)^2$ gives

$$
\begin{aligned}
q(j)^2 &< \frac{\theta_+^{j+1:n}-\theta_+^{l:n}}{\theta_+^{l:n}-\theta^{l:n}} \\
&= \frac{\theta_+^{j+1:n}-\mu}{gap(l)}\left[1+O\left(\frac{\lambda_+-\lambda_-}{gap(l)}\right)\right] \qquad (42)
\end{aligned}
$$

Now take the product of (40) and (42) and bound min by the geometric mean to find

$$
p(j)^2q(j)^2 < \frac{\lambda_+-\lambda_-}{gap(m)}\min\left\{\frac{\theta_+^{j+1:n}-\mu}{gap(l)},\frac{2\beta_j^2}{\theta_+^{j+1:n}-\mu}\cdot\frac{\lambda_+-\lambda_-}{gap(m)gap(l)}\right\}
$$

$$\cdot \left[1 + O\left(\frac{\lambda_+ - \lambda_-}{gap}\right)\right] + O\left(\frac{\lambda_+ - \lambda_-}{gap}\right)^2$$

$$< \sqrt{2}\,\frac{\lambda_+ - \lambda_-}{gap(m)} \cdot \frac{\beta_j}{gap(l)} \cdot \left(\frac{\lambda_+ - \lambda_-}{gap(m)}\right)^{1/2} + O\left(\frac{\lambda_+ - \lambda_-}{gap}\right)^2 \quad (43)$$

We see that (43) is an instance of the bound claimed in the theorem.

The analysis for j-values in which $\theta^{j+1:n} \in (\lambda_-, \lambda_+)$ is the dual of what is given above. Attention concentrates on $q(j)^2$ instead of $p(j)^2$, $gap(l)$ replaces $gap(m)$ and the roles of β_j and β_{j-1} are exchanged. By Cauchy's Interlace theorem either $\theta^{1:j-1} \in (\lambda_-, \lambda_+)$ or $\theta^{j+1:n} \in (\lambda_-, \lambda_+)$ for each j, $1 \le j \le n$, and so the proof is complete. □

Remark. The analysis shows that the configuration of Ritz values has to be quite special to yield a value for $|p(j)q(j)|$ as large as $O((\lambda_+ - \lambda_-)/gap)^{3/4}$. If $\theta_+^{1:j-1}$, in Case 1, or $\theta_+^{j+1:n}$, in Case 2, is close to λ_+ or to $\theta_+^{1:m}$ then the minima in (40) and (41) are $O((\lambda_+ - \lambda_-)/gap)$ and that same bound holds for $|p(j)q(j)|$. Since there is usually a little freedom in the choice of l and m we can expect that for j close to l and to m

$$|p(j)q(j)| = O\left(\frac{\lambda_+ - \lambda_-}{gap}\right).$$

It follows that

$$Overlap(\boldsymbol{p}, \boldsymbol{q}) := |\boldsymbol{p}| \cdot |\boldsymbol{q}| = O\left[\left(\frac{\lambda_+ - \lambda_-}{gap}\right)^{3/4}\right].$$

9 Accuracy of Subspaces

Consider the basis $\{\boldsymbol{p}, \boldsymbol{q}\}$ produced by the use of submatrices. Here $\boldsymbol{p}$'s support is $1 : m$, and $\boldsymbol{q}$'s support is $l : n$ and l, m are chosen so that $|\beta_m p(m)|$ and $|\beta_{l-1} q(l)|$ are small. In addition $\|\boldsymbol{p}\| = \|\boldsymbol{q}\| = 1$. We consider how well $span(\boldsymbol{p}, \boldsymbol{q})$ approximates the invariant subspace for λ_- and λ_+.

Let $\theta^{1:m} = \boldsymbol{p}^* T \boldsymbol{p}$ be $\boldsymbol{p}$'s Rayleigh quotient and let $\theta^{l:n} = \boldsymbol{q}^* T \boldsymbol{q}$ be $\boldsymbol{q}$'s Rayleigh quotient. We have

$$T\boldsymbol{p} = \boldsymbol{p}\theta^{1:m} + \boldsymbol{e}_{m+1}\beta_m p(m), \quad T\boldsymbol{q} = \boldsymbol{q}\theta^{l:n} + e_{l-1}\beta_{l-1}q(l). \qquad (44)$$

There is no loss in shifting T to the mean μ of the two Ritz values $\theta^{1:m}$ and $\theta^{l:n}$. Also let

$$\delta = \frac{\theta^{1:m} - \theta^{l:n}}{2}.$$

There are two expressions for $\kappa = \boldsymbol{p}^*(T - \mu I)\boldsymbol{q}$:

$$\begin{aligned} \kappa &= \boldsymbol{p}^*\boldsymbol{q}\delta + q(m+1)\beta_m p(m) \\ \kappa &= -\boldsymbol{p}^*\boldsymbol{q}\delta + p(l-1)\beta_{l-1}q(l). \end{aligned} \tag{45}$$

Also

$$[\boldsymbol{p},\boldsymbol{q}]^*[\boldsymbol{p},\boldsymbol{q}] = \begin{pmatrix} 1 & \boldsymbol{p}^*\boldsymbol{q} \\ \boldsymbol{q}^*\boldsymbol{p} & 1 \end{pmatrix},$$

$$[\boldsymbol{p},\boldsymbol{q}]^*T[\boldsymbol{p},\boldsymbol{q}] = \begin{pmatrix} \theta^{1:m} & \kappa \\ \kappa & \theta^{l:n} \end{pmatrix}.$$

Then

$$R := (T - \mu I)[\boldsymbol{p},\boldsymbol{q}] - [\boldsymbol{p},\boldsymbol{q}]\begin{pmatrix} \delta & \kappa \\ \kappa & -\delta \end{pmatrix}.$$

By (44) and (45),

$$R = [\boldsymbol{e}_{m+1}\beta_m p(m) - \boldsymbol{q}\kappa,\ \boldsymbol{e}_{l-1}\beta_{l-1}q(l) - \boldsymbol{p}\kappa].$$

Recall the supports of $\boldsymbol{p}$ and $\boldsymbol{q}$ and use (45) to find

$$R^*R = \begin{pmatrix} \beta_m^2 p(m)^2 - \kappa^2 + 2\boldsymbol{p}^*\boldsymbol{q}\delta\kappa, & \boldsymbol{p}^*q\kappa^2 \\ \boldsymbol{p}^*\boldsymbol{q}\kappa^2, & \beta_{l-1}^2 q(l)^2 - \kappa^2 - 2\boldsymbol{p}^*\boldsymbol{q}\delta\kappa \end{pmatrix}.$$

Since $\boldsymbol{p}$ and $\boldsymbol{q}$ are not orthogonal the proper measure of $[\boldsymbol{p},\boldsymbol{q}]$'s residual is

$$\sigma_{\max}\left[R\begin{pmatrix} 1 & \boldsymbol{p}^*\boldsymbol{q} \\ \boldsymbol{p}^*\boldsymbol{q} & 1 \end{pmatrix}^{-1/2}\right].$$

Thus $\sigma_{\max}^2$ is the largest zero of

$$det\ \left[R^*R - \sigma^2\begin{pmatrix} 1 & \boldsymbol{p}^*\boldsymbol{q} \\ \boldsymbol{p}^*\boldsymbol{q} & 1 \end{pmatrix}\right].$$

So

$$[1-(\boldsymbol{p}^*\boldsymbol{q})^2]\sigma^4 - [\beta_m^2 p(m)^2 + \beta_{l-1}^2 q(l)^2 - 2\kappa^2 - 2\kappa^2(\boldsymbol{p}^*\boldsymbol{q})^2]\sigma^2 + const = 0.$$

$\sigma_{\max}$ is majorized by the sum of the roots

$$\sigma_{\max}^2 < \frac{\beta_m^2 p(m)^2 + \beta_{l-1}^2 q(l)^2 - 2\kappa^2}{1-(\boldsymbol{p}^*\boldsymbol{q})^2}.$$

Now (45) may be rewritten as $2\kappa = \mathcal{L} + \mathcal{M}$, defining $\mathcal{L}$ and $\mathcal{M}$ in a natural way. Since

$$2\kappa^2 = \mathcal{L}^2 + \mathcal{M}^2 - \frac{1}{2}(\mathcal{L}-\mathcal{M})^2$$

we obtain

$$\begin{aligned} \sigma_{\max}^2 &< \{\beta_m^2 p(m)^2(1-q(m+1)^2) + \beta_{l-1}^2 q(l)^2(1-p(l-1)^2) \\ &+ \frac{1}{2}[\beta_m p(m)q(m+1) - \beta_{l-1}q(l)p(l-1)]^2\} \cdot [1+(\boldsymbol{p}^*\boldsymbol{q})^2]. \quad (46) \end{aligned}$$

This is an easily computed bound. The closer is $|q(m+1)|$ to $\|\boldsymbol{q}\|_\infty$ and $|p(l-1)|$ to $\|\boldsymbol{p}\|_\infty$ the lower is the bound on $\sigma_{\max}$. From standard gap theorems in [4] the sine of the error angle is less than $\sigma_{\max}/gap$, where gap is the separation of $[\lambda_-, \lambda_+]$ from the rest of the spectrum.

The General Case

Given are $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_\#$ with the support of $\boldsymbol{p}_j$ on $(l_j : m_j)$. Also $\boldsymbol{p}_i \cdot \boldsymbol{p}_j = 0$ if $|i-j| > 1$ and $\|\boldsymbol{p}_j\| = 1$, $j = 1, \ldots, \#$. By construction,

$$T\boldsymbol{p}_j = \boldsymbol{p}_j\theta_j + \boldsymbol{e}_{l_j-1}\beta_{l_j-1}p_j(l_j) + \boldsymbol{e}_{m_j+1}\beta_{m_j}p_j(m_j), \quad \theta_j = \theta^{l_j:m_j}. \qquad (47)$$

In order to simplify expressions it is convenient to shift T to the mean value μ of the Ritz values $\theta_1, \ldots, \theta_\#$. Let $\theta_i = \mu + \delta_i$, $i = 1, \ldots, \#$.

There are two expressions for $\kappa_i := \boldsymbol{p}_i^*(T-\mu I)\boldsymbol{p}_{i+1}$:

$$\kappa_i = \begin{cases} \boldsymbol{p}_{i+1} \cdot \boldsymbol{p}_i\, \delta_i + p_{i+1}(m_i+1)\beta_{m_i} p_i(m_i), \\ \boldsymbol{p}_i \cdot \boldsymbol{p}_{i+1}\, \delta_{i+1} + p_i(l_{i+1}-1)\beta_{l_{i+1}-1}\, p_{i+1}(l_{i+1}), \end{cases} \qquad (48)$$

because of the disjoint supports.

Define

$$\boldsymbol{r}_i := (T-\mu I)\boldsymbol{p}_i - \delta_i\boldsymbol{p}_i - \kappa_{i-1}\boldsymbol{p}_{i-1} - \kappa_i\boldsymbol{p}_{i+1} \qquad (49)$$

where $\kappa_0 = \kappa_\# = 0$. Let $R = [\boldsymbol{r}_1, \ldots, \boldsymbol{r}_\#]$. By (47),

$$\boldsymbol{r}_i = \boldsymbol{e}_{l_i-1}\beta_{l_i-1}p_i(l_i) + \boldsymbol{e}_{m_i+1}\beta_{m_i}p_i(m_i) - \boldsymbol{p}_{i-1}\kappa_{i-1} - \boldsymbol{p}_{i+1}\kappa_i.$$

Again the disjoint supports show that

$$\begin{aligned} \boldsymbol{r}_i \cdot \boldsymbol{r}_{i+1} &= \boldsymbol{p}_{i-1} \cdot \boldsymbol{p}_i \, \kappa_{i-1}\kappa_i + \boldsymbol{p}_i \cdot \boldsymbol{p}_{i+1} \, \kappa_i^2 + \boldsymbol{p}_{i+1} \cdot \boldsymbol{p}_{i+2} \, \kappa_i\kappa_{i+1}, & (50) \\ \boldsymbol{r}_i \cdot \boldsymbol{r}_{i+2} &= \kappa_i\kappa_{i+1} & (51) \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{r}_i \cdot \boldsymbol{r}_i &= \beta_{l_i-1}^2 p_i(l_i)^2 + \beta_{m_i}^2 p_i(m_i)^2 + \kappa_{i-1}^2 + \kappa_i^2 \\ &\quad -2\kappa_{i-1}p_{i-1}(l_i-1)\beta_{l_i-1}p_i(l_i) - 2\kappa_i p_{i+1}(m_i+1)\beta_{m_i}p_i(m_i). \end{aligned}$$

By (48),

$$\begin{aligned} & \kappa_i^2 - 2\kappa_i p_{i+1}(m_i+1)\beta_{m_i}p_i(m_i) \\ = \; & [\kappa_i - p_{i+1}(m_i+1)\beta_{m_i}p_i(m_i)]^2 - p_{i+1}(m_i+1)^2\beta_{m_i}^2 p_i(m_i)^2 \\ = \; & (\boldsymbol{p}_i \cdot \boldsymbol{p}_{i+1} \, \delta_i)^2 - p_{i+1}(m_i+1)^2\beta_{m_i}^2 p_i(m_i)^2. \end{aligned}$$

Thus, for $i = 1, 2, \ldots \#$,

$$\begin{aligned} \boldsymbol{r}_i \cdot \boldsymbol{r}_i &= (1 - p_{i-1}(l_i-1)^2)\,\beta_{l_i-1}^2 \, p_i(l_i)^2 \\ &+ (1 - p_{i+1}(m_i+1)^2)\beta_{m_i}^2 p_i(m_i)^2 & (52) \\ &+ (\boldsymbol{p}_{i-1} \cdot \boldsymbol{p}_i \, \delta_{i-1})^2 + (\boldsymbol{p}_i \cdot \boldsymbol{p}_{i+1} \, \delta_i)^2, \end{aligned}$$

with out of range terms set to zero when $i = 1$ and $\#$. Let

$$\overset{o}{I} := tridiag \left(\begin{array}{ccccccccc} & \boldsymbol{p}_1 \cdot \boldsymbol{p}_2 & & \boldsymbol{p}_2 \cdot \boldsymbol{p}_3 & & \bullet & & \boldsymbol{p}_{\#-1} \cdot \boldsymbol{p}_\# & \\ 1 & & 1 & & \bullet & & \bullet & & 1 \\ & \boldsymbol{p}_1 \cdot \boldsymbol{p}_2 & & \boldsymbol{p}_2 \cdot \boldsymbol{p}_3 & & \bullet & & \boldsymbol{p}_{\#-1} \cdot \boldsymbol{p}_\# & \end{array} \right).$$

Then the measure of the quality of Range $[\boldsymbol{p}_1, \ldots, \boldsymbol{p}_\#]$ as an invariant subspace is given by $\sigma_{\max}$ where $\sigma_{\max}^2$ is the largest zero of

$$det\,[R^*R - \sigma_{\max}^2 \, \overset{o}{I}]. \qquad (53)$$

some terms in (48), (50), and (52) are much smaller than others.

Thus if $w :=$ cluster width

$$\begin{aligned}
(R^*R)_{ii} &= (1 - p_{i-1}(l_i - 1)^2)\,\beta^2_{l_i-1}\,p_i(l_i)^2 \\
&\quad +(1 - p_{i+1}(m_i + 1)^2)\beta^2_{m_i}p_i(m_i)^2 + O(w^{7/2}) \\
(R^*R)_{i,i+1} &= 0 + O(w^{11/4}) \\
(R^*R)_{i,i+2} &= \kappa_i\kappa_{i+1}.
\end{aligned}$$

Using this approximation it is straightforward to approximate the largest zero of (53) by bisection. The largest diagonal entry of R^*R is a reasonable approximation to $\sigma^2_{\max}$.

10 Counting Ritz Values

In order to justify the selection of submatrices some background material is needed. Recall that $\theta_\bullet^{l\langle j\rangle m}$ denotes a Ritz value from $T^{l\langle j\rangle m}$, the submatrix obtained by deleting row and column j from $T^{l:m}$.

- Cauchy's Interlace theorem (true for symmetric matrices):
  Let $\theta^{\langle j\rangle} := \theta^{1\langle j\rangle n}$. For each $j = 1, \dots, n$

$$\lambda_i \le \theta_i^{\langle j\rangle} \le \lambda_{i+1} \le \theta_{i+1}^{\langle j\rangle}.$$

- Unreduced Tridiagonal Interlace theorem: For each $j < n, i < j$, and $1 < l \le n-j$ there is a Ritz value $\theta_\bullet^{1:j+l}$ in the closed interval $[\theta_i^{1:j}, \theta_{i+1}^{1:j}]$. For a proof see [3].

- The Window Count.
  Let $\mathcal{I}$ be any fixed closed interval on the real line. Let $\#_{\mathcal{I}}(j:k)$ be the number of Ritz values of $T^{j:k}$ in $\mathcal{I}$. Then $\#_{\mathcal{I}}(j:k)$ is 'nearly' monotone increasing with k. More precisely,

$$-1 \le \#_{\mathcal{I}}(j:k+1) - \max_{1\le i\le k} \#_{\mathcal{I}}(j:i) \le 1, \quad \text{for all } k.$$

  This result is a direct corollary of the Tridiagonal Interlace Theorem.

Lemma 9 (Window Count) *Let $\mathcal{I}$ be the convex hull of a cluster of adjacent eigenvalues of unreduced, symmetric, tridiagonal $T = T^{1:n}$. If the window count $\#_{\mathcal{I}}(1:j)$ is well defined for $j = 2, \dots, n-1$ then*

$$\#_{\mathcal{I}}(1:j-1) + \#_{\mathcal{I}}(j+1:n) = \#_{\mathcal{I}}(1:n) - 1, \quad j = 2, \dots, n+1.$$

Proof. Since the end points of $\mathcal{I}$ are the extreme eigenvalues of the cluster there are exactly $\#_{\mathcal{I}}(1:n)-1$ abutting subintervals in $\mathcal{I}$ of the form $[\lambda_i, \lambda_{i+1}]$. If the window count is well defined there are no zero pivots in the triangular factorization (up or down) with shifts at the end points of $\mathcal{I}$. Hence no Ritz values of $T^{1:j-1}$ or $T^{j+1:n}$ fall at the end points of $\mathcal{I}$. By Cauchy's Interlace theorem (one can assign a Ritz value of $T^{1(j)n}$ to each subinterval) there are either $\#_{\mathcal{I}}(1:n)-1$ or $\#_{\mathcal{I}}(1:n)$ Ritz values of $T^{1(j)n}$ in $\mathcal{I}$, for each j. By the 'coincidence' property of tridiagonals there can only be $\#_{\mathcal{I}}(1:n)$ Ritz values in $\mathcal{I}$ if one of $\mathcal{I}$'s end points is a Ritz value and this is ruled out by the assumption on the window count. □

11 Submatrix Selection

At present we have no preferred method for choosing submatrices automatically. Below we present two methods that have been satisfactory so far.

Mid-Point Selection

T and $\mathcal{I}$ are given. $\mathcal{I}$ is the convex hull of a cluster. Let $\# = \#_{\mathcal{I}}(1:n)$. Define, for $j = 0, 1, \ldots, \#$,

$$\bar{l}_j = \max\{i : \#_{\mathcal{I}}(1:i) = j\}. \tag{54}$$

Define, for $j = 1, \ldots, \#$,

$$m_j = \lceil (\bar{l}_{j-1} + \bar{l}_j)/2 \rceil.$$

Note that $\bar{l}_{\#} = n$. Let $\bar{l}_{-1} = -1$. Take as initial submatrices

$$(\bar{l}_{j-2} + 2 : m_j), \quad j = 1, \ldots, \#.$$

Justification. Since $\#_{\mathcal{I}}(1:i)$ is nearly monotone increasing in i the indices $\{\bar{l}_j\}$ are strictly monotone increasing in j thanks to the max in their definition. Hence

$$m_j \le \bar{l}_j < \bar{l}_j + 2, \; j = 0, 1, \ldots, \#. \tag{55}$$

Hence

$$m_{j-2} \le \bar{l}_{j-2} < \bar{l}_{j-2} + 2.$$

Thus the supports of the submatrices are disjoint except, possibly, for nearest neighbors.

Next we show that there are 'enough' Ritz values in each submatrix. Since $\mathcal{I}$ is not the convex hull of the Ritz values of the submatrix $(\bar{l}_{j-2}+2:n)$ in $\mathcal{I}$ the window count lemma is not applicable. By Cauchy's Interlace theorem

$$\#_{\mathcal{I}}(\bar{l}_{j-2}+2:n)-1 \leq \#_{\mathcal{I}}(\bar{l}_{j-2}+2:m_j)+\#_{\mathcal{I}}(m_j+2:n) \leq \#_{\mathcal{I}}(\bar{l}_{j-2}+2:n)+1. \tag{56}$$

However Lemma 9 may be invoked twice to obtain

$$\#_{\mathcal{I}}(m_j+2:n) = (\#-1) - \#_{\mathcal{I}}(1:m_j) = \#-j-1, \tag{57}$$

and

$$\#_{\mathcal{I}}(\bar{l}_{j-2}+2:n) = (\#-1) - \#_{\mathcal{I}}(1:\bar{l}_{j-2}) = \#-j+1. \tag{58}$$

Use (57) and (58) in (56) to obtain

$$3 \geq \#_{\mathcal{I}}(\bar{l}_{j-2}+2:m_j) \geq 1. \tag{59}$$

In the exceptional case that $\#_{\mathcal{I}}(\bar{l}_{j-2}+2:m_j) > 1$, for some j, the support of the jth submatrix may be reduced from either or both ends until the count is exactly 1. □

There is a dual algorithm using trailing submatrices that delivers $\underline{m}_j$, $j=1,\ldots,\#+1$ such that

$$\underline{m}_j = \min\{i : \#_{\mathcal{I}}(i:n) = \#-j+1\}$$

and mid-points l_j, $j=1,\ldots,\#$, with $l_j = \lfloor(\underline{m}_j+\underline{m}_{j+1})/2\rfloor$. This process yields more balanced submatrices

$$(l_j : m_j), \quad j=1,\ldots,\#$$

with the same bounds and disjoint support properties.

We used a selection very close to $\{(l_j : m_j)\}$ in 1989 before our analysis of close pairs was developed. The performance was very satisfactory but there is no reason why the mid-points in the ranges $\{\bar{l}_{j-1}+1,\ldots,\bar{l}_j\}$ should give the smallest coefficient of $|\mathcal{I}|$ in the residual norm bounds. We want the index i in $\{\bar{l}_{j-1}+1,\ldots,\bar{l}_j\}$ that gives a minimal, or small, value to the G_i of the Double Occupancy Theorem.

Selection by Pairs

The previous results for close pairs of eigenvalues may be used in a systematic way to produce appropriate submatrices for isolated clusters containing any number of eigenvalues. The goal here is to show the existence of the submatrices, not to produce an efficient algorithm.

From Section 7 if $\#(1:m) = 2$ then there exist suitable indices μ and ν (by no means unique) such that $\#(1,\nu) = 1$ and $\#(\mu,m) = 1$. Usually $\mu < \nu$ but that is not necessary in what follows. There are basis vectors with supports on $(1,\nu)$ and on (μ,m) whose residual norms are proportional to the separation of the two Ritz values that cause $\#(1:m) = 2$.

Now suppose that

$$\# := \#(1,n) = \#_{\mathcal{I}}(1,n) > 2.$$

Let h be maximal such that $\#(1,h) = 1$. For the submatrix $(1:h+1)$ let the optimal indices (μ and ν) be j_2 and k_1. The Ritz vector (with Ritz value in $\mathcal{I}$) for submatrix $(1:k_1)$ is the first basis vector. Set $j_1 = 1$. The Ritz vector for $(j_2:h+1)$ is not used but the index j_2 will play a role. Check that

$$\#(j_2:n) = \#(1:n) - 1. \tag{60}$$

If not, adjust j_2 until (60) holds. Having peeled off a submatrix $j_1:k_1$ from the top and then discarded rows $1:j2-1$ from T we proceed in the same way at the bottom of T.

Let p be minimal such that $\#(p,n) = 1$. For the submatrix $(p-1:n)$ let the optimal indices be $j_\#$ and $k_{\#-1}$. The Ritz vector (with Ritz value in $\mathcal{I}$) for submatrix $(j_\#,n)$ is the final ($\#$) basis vector. Set $k_\# = n$. The Ritz vector for submatrix $(p-1:k_{\#-1})$ is not used but $k_{\#-1}$ plays a role. Check that

$$\#(j_2,k_{\#-1}) = \# - 2. \tag{61}$$

If not, adjust $k_{\#-1}$ until (61) holds. If $\# - 2 > 2$ then repeat the procedure just described on $(j_2:k_{\#-1})$ to obtain two new submatrices (j_2,k_2), $(j_{\#-1},k_{\#-1})$ and, possibly, a remaining submatrix with fewer Ritz values in $\mathcal{I}$. Eventually one obtains $\#$ submatrices $(j_i:k_i)$, $i = 1,\ldots,\#$ each of which, by construction, has a simple Ritz value in $\mathcal{I}$. The associated Ritz vectors, with zeros appended to make n-vectors, constitute a good basis for $\mathcal{I}$'s invariant subspace.

Should it ever occur that $j_{i+1} < k_{i-1}$ we are at liberty to increase j_{i+1} or decrease k_{i-1} a little subject only to the constraint that $\#(j_i : k_i) = 1$ for each i. Our goal is to have

$$k_{i-1} \leq \text{ max entry in } \boldsymbol{s}^{j_i:k_i} \leq j_{i+1}, \quad i = 2, \ldots, \# - 1.$$

Selection by Envelope

A way to choose submatrices is suggested by Section 6. First find the envelope vector $\mathcal{E}$ and then find # entries of $\mathcal{E}$ that are local maxima. Suppose first that there is a unique set of such positions $k_1, k_2, \ldots, k_\#$. For $1 < j < \#$ the jth submatrix is

$$(k_{j-1} + 1 : k_{j+1} - 1). \tag{62}$$

Usually the first and the last are $(1 : k_2 - 1)$ and $(k_{\#-1} + 1 : n)$. However in general we must be more careful and define k_0 and $k_{\#+1}$. Let the first *nonnegligible* entry of $\mathcal{E}$ be in position $k_0 + 1$ and let the last *nonnegligible* entry be in position $k_{\#+1} - 1$. Now (62) also gives the submatrices for $j = 1$ and $j = \#$.

In case the location of the ith summit is given by several indices then take k_i to be that set and interpret $k_i + 1$ as $1 + \max k_i$ and $k_i - 1$ as $-1 + \min k_i$. To quantify the adjective negligible we propose a threshold of $macheps \cdot \|\mathcal{E}\| = macheps \cdot \sqrt{\#}$.

The approximation of $\mathcal{E}$ is an implementation issue that will not be addressed here.

12 More Examples

A Glued Wilkinson Matrix

In Section 5 we studied W_{21}^+. Here we use W_{25}^+ but take 4 copies and connect them by an off-diagonal entry e which is called 'the glue' in the matrix W_{100}. In an obvious extension of our notation in Section 2

$$W_{100} = tridiag \begin{pmatrix} & e & & e & & e & \\ W_{25}^+ & & W_{25}^+ & & W_{25}^+ & & W_{25}^+ \\ & e & & e & & e & \end{pmatrix}.$$

If e is too small, like 10^{-3}, then W_{100} is too close to a direct sum of 4 matrices and calculation of orthogonal eigenvectors is not hard. If e is too large (> 2) then the eigenvalues are sufficiently well separated to be treated as isolated. We use

$$e = 0.3$$

and show the largest 8 computed eigenvalues in Table 2. Without the use of Gram-Schmidt orthogonalization inverse iteration gives neither small residuals nor adequate orthogonality.

λ_{93}	12.577864
λ_{94}	12.577870
λ_{95}	12.577881
λ_{96}	12.746191
λ_{97}	12.746193
λ_{98}	12.939114
λ_{99}	12.939115
λ_{100}	12.939117

Table 2: Selected Eigenvalues of W_{100}

Figures 6, 7 and 8 show the submatrix indices used and the basis vectors they yield. The vectors are plotted on a logarithmic scale with the correct sign attached. All entries less than 10^{-9} are treated as 0. The reason for using log scale is the to focus attention on the smaller entries.

In this example the submatrices overlap by only one or two indices. The residual norms are the magnitudes of the first and last entries and are all less than $\mathit{macheps} \cdot \mathit{spread}$. The dot products between vectors in each group are almost zero because the supports are almost disjoint. On the other hand the supports of the vectors $\boldsymbol{x}_{93}$ and $\boldsymbol{x}_{98}$ are identical and orthogonality comes from cancellation. These dot products are less than $30 \cdot \mathit{macheps}$.

An Example from the Lanczos Algorithm

The Lanczos Algorithm with no reorthogonalization was run in double precision on a diagonal matrix of order 205

$$D = diag(1, 2, \ldots, 200, 400, 400, 400, 400, 600)$$

with starting vector

$$e = (1, 1, \ldots, 1)^t.$$

The run stopped at step 87 and the resulting tridiagonal matrix $T_{87} = T^{1:87}$ had 5 copies of 600, four copies of 400, and a single eigenvalue at 200, to single precision.

Figure 9 shows the four vectors corresponding to the cluster at 400. All these calculations were in single precision. Note that the overlap of the supports is greater than in the glued Wilkinson matrix. However all nonzero products were about 10^{-14}, much les than *macheps*.

Acknowledgement I am grateful to Yin Sun Feng for making some of the figures, in 1989, and Inderjit Dhillon for creating the recent ones, for a careful reading of the manuscript and for many useful suggestions.

References

[1] F. P. Gantmacher and M. G. Krein. *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems.* US AEC Translation Series, Published in Moscow, 1950.

[2] W. B. Gragg and W. J Harrod. The numerically stable reconstruction of Jacobi spectral data. *Numer. Math.*, 44:317–336, 1984.

[3] R. O. Hill, Jr. and B. N. Parlett. Refined interlacing properties. *SIAM J. on Matrix Anal. and Appl.*, 13:239–247, 1992.

[4] B. N. Parlett. *The Symmetric Eigenvalue Problem.* Prentice-Hall, Englewood Cliffs, NJ, 1980.

[5] B. N. Parlett and I. S. Dhillon. Fernando's method to find the most redundant equation in a tridiagonal system. *J. Lin. Alg. & Appls.*, (Accepted for publication).

[6] J. H. Wilkinson. *The Algebraic Eigenvalue Problem.* Clarendon Press, Oxford, 1965.

[7] Q. Ye. On close eigenvalues of tridiagonal matrices. *Numer. Math.*, 70:507–514, 1995.

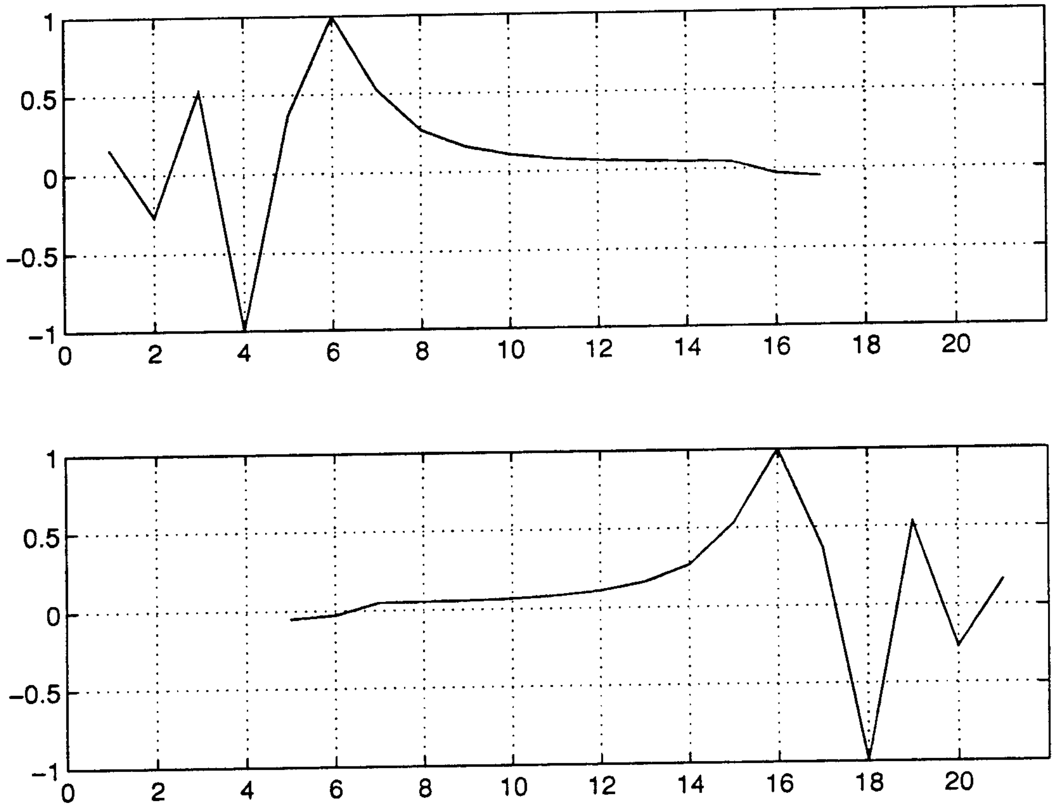


Figure 1: Vectors $\boldsymbol{z}_+$ and $\boldsymbol{z}_-$ for the pair near 6 on a log scale

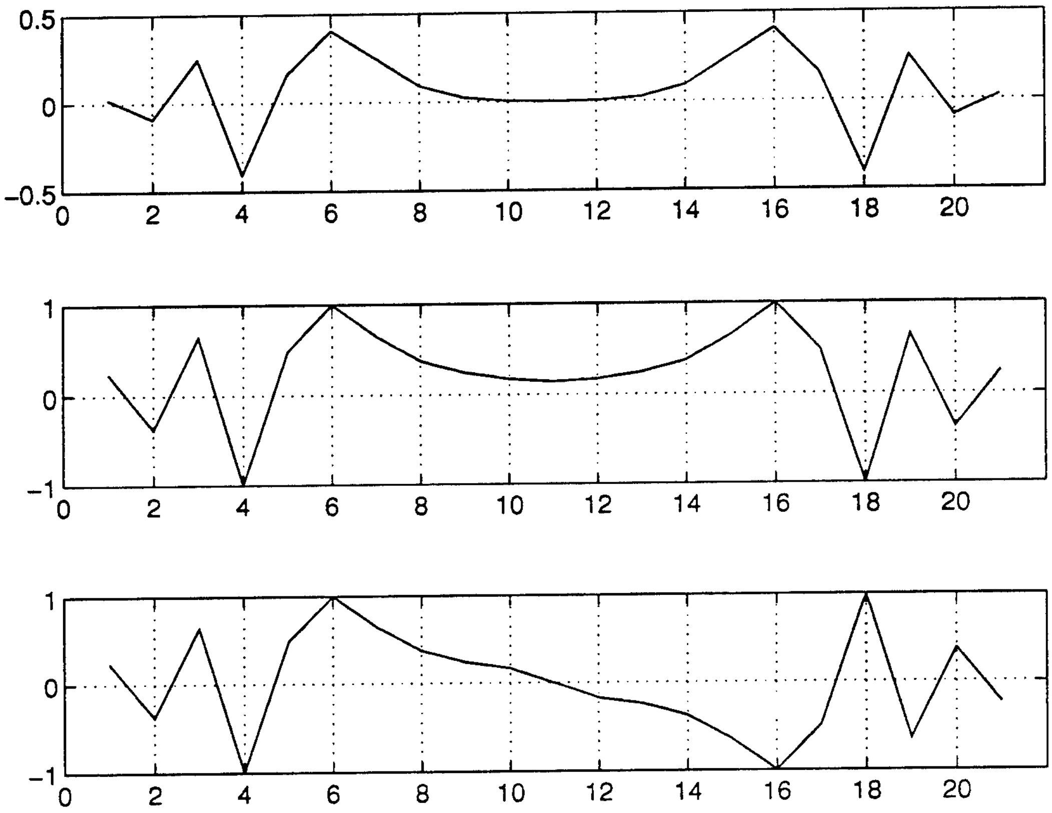


Figure 2: Bisectors of z_+ and z_- on a log scale

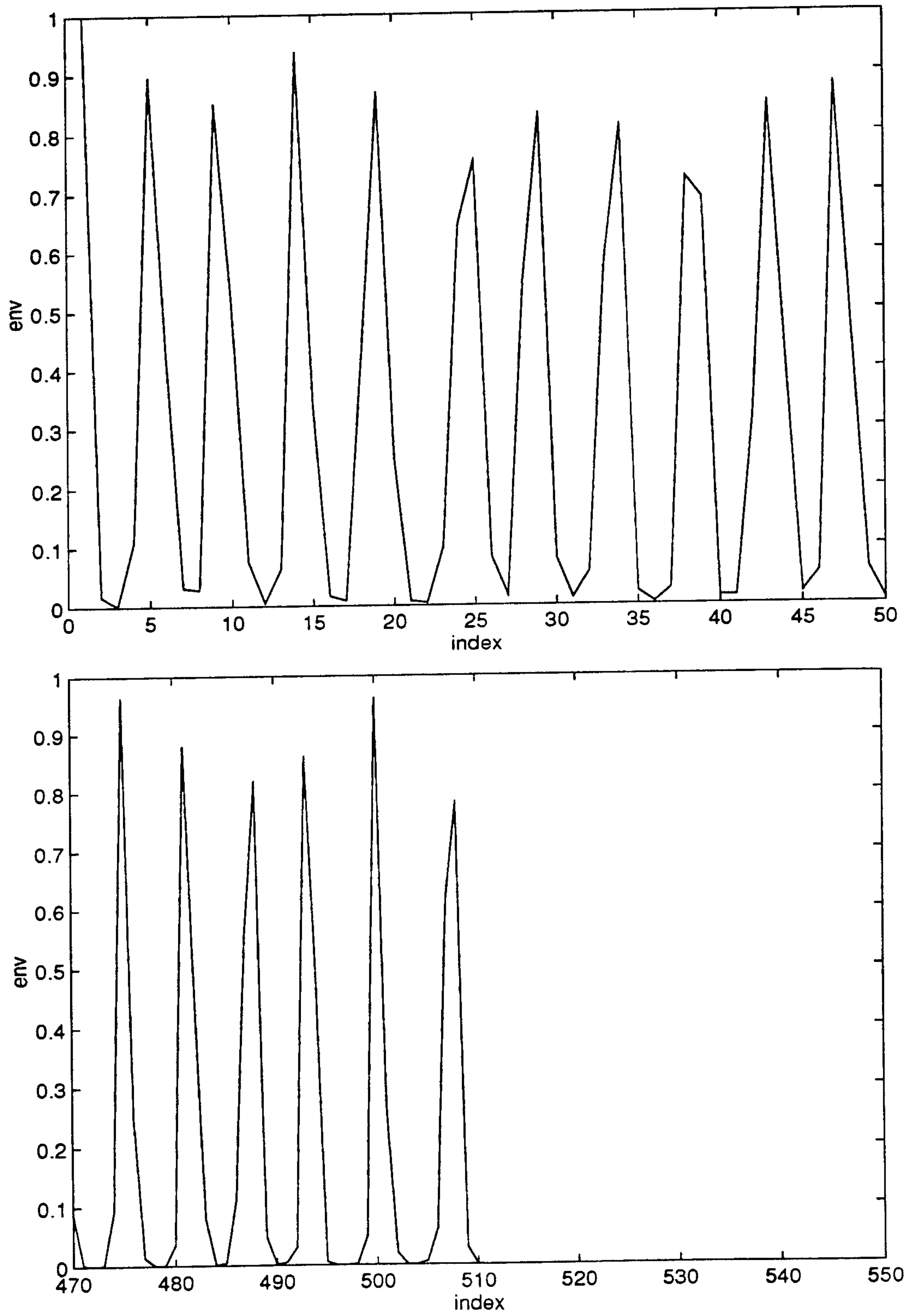


Figure 3: Snapshot of Envelope (108 eigenvalues)

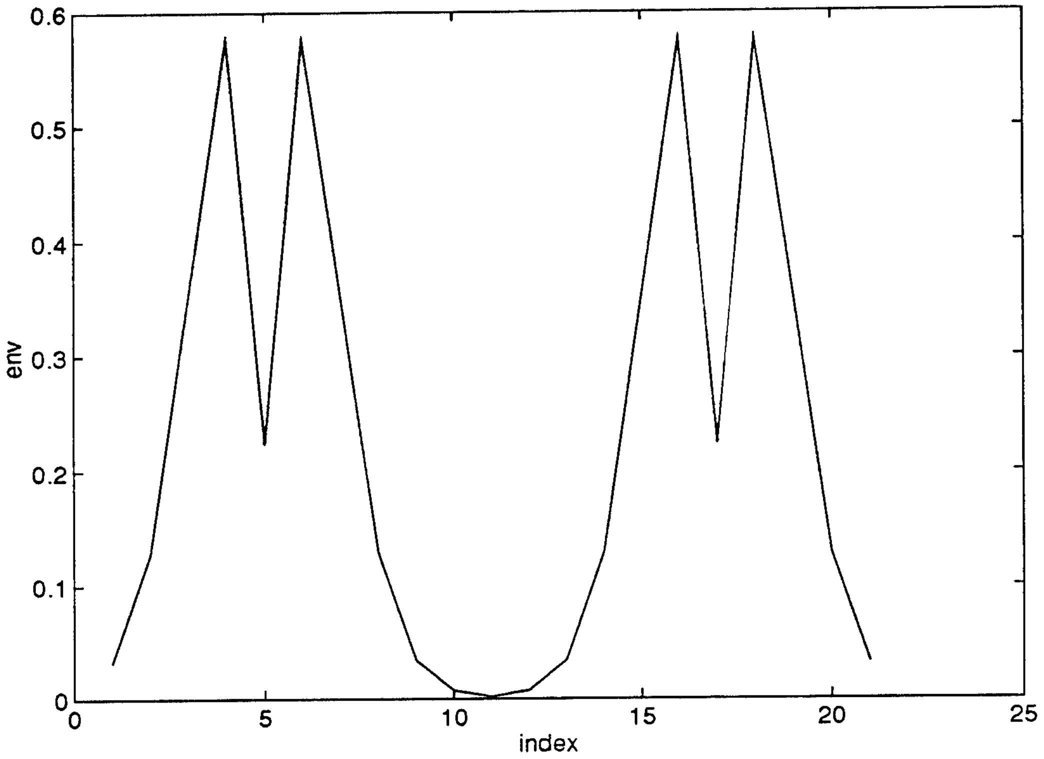


Figure 4: Envelope for $\lambda_{12}, \lambda_{13}$ from W_{21}^{+}

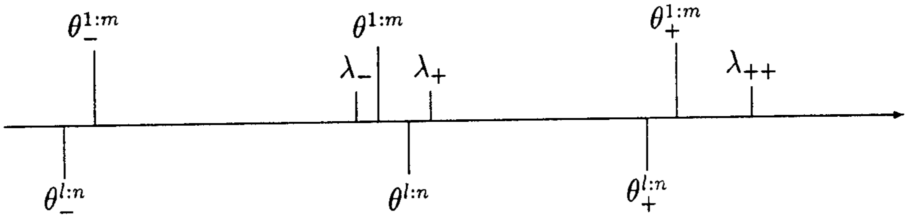


Figure 5: Location of Ritz values

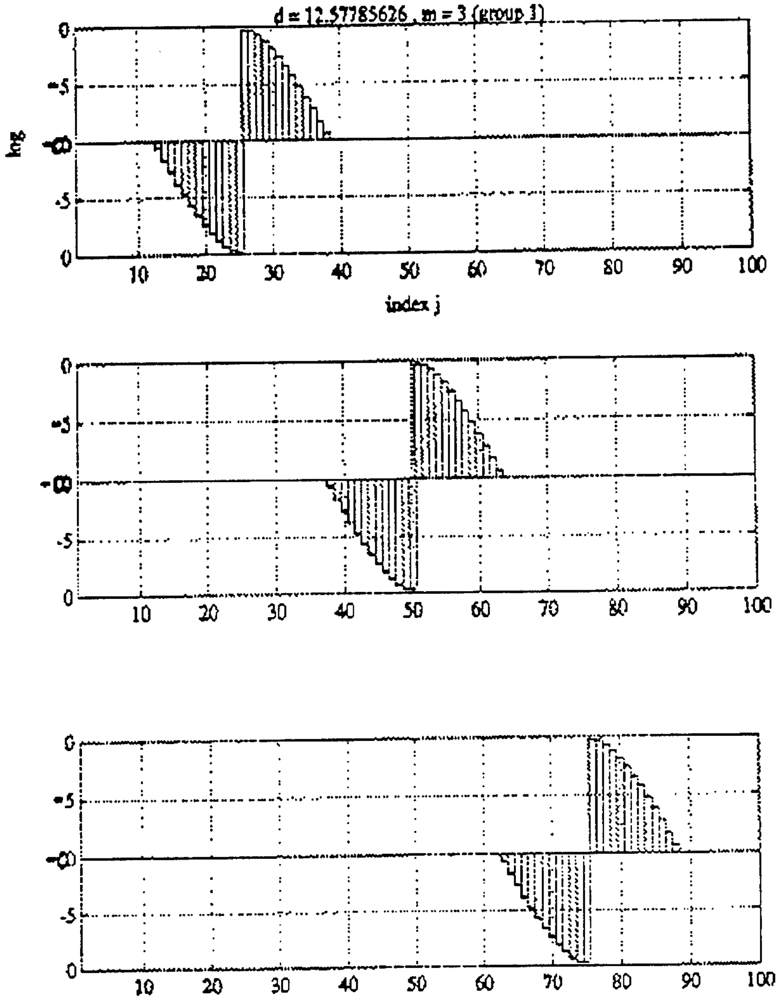


Figure 6: Basis eigenvectors of W_{100} for $\{\lambda_{93}, \lambda_{94}, \lambda_{95}\}$

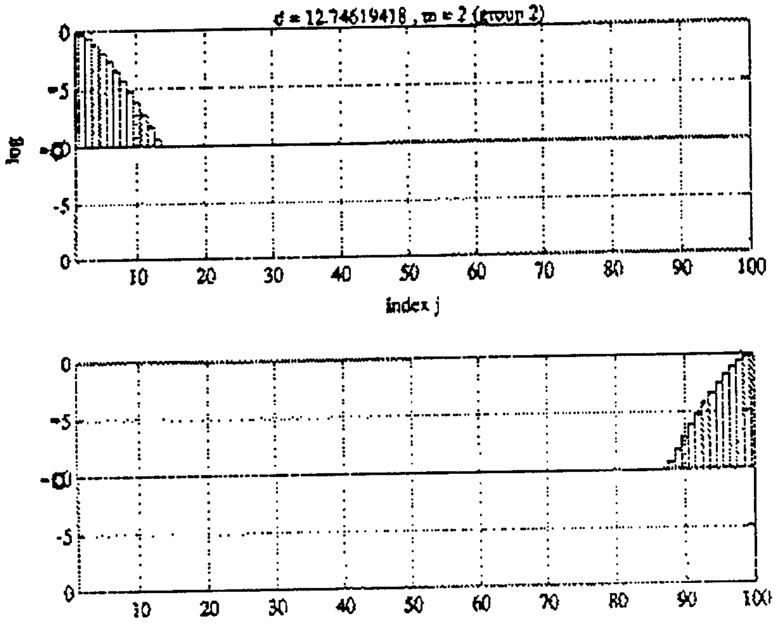


Figure 7: Basis eigenvectors of W_{100} for $\{\lambda_{96},\ \lambda_{97}\}$

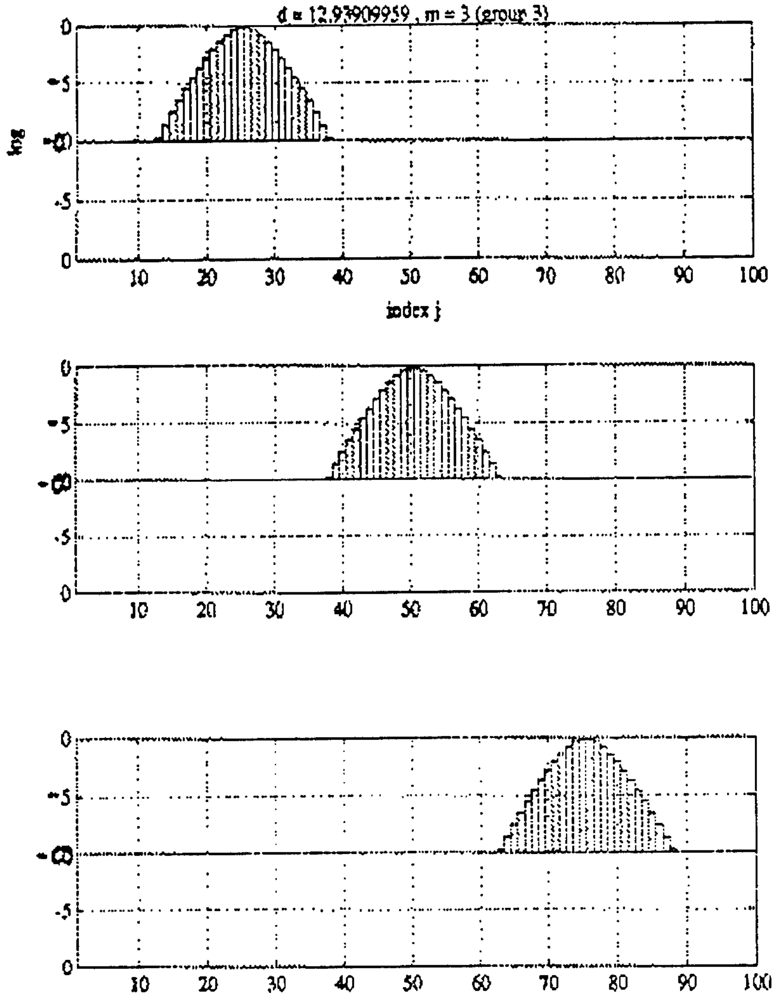


Figure 8: Basis eigenvectors of W_{100} for $\{\lambda_{98},\ \lambda_{99},\ \lambda_{100}\}$

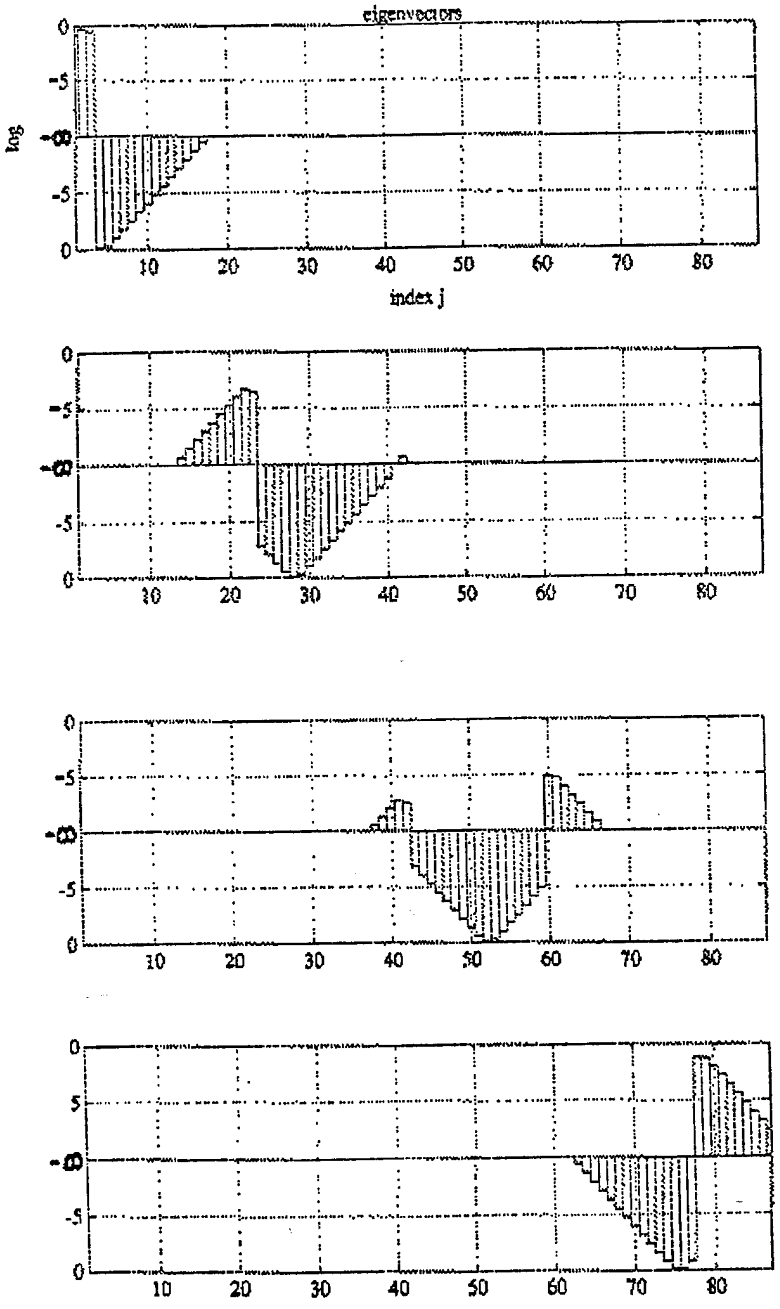


Figure 9: Eigenvectors $\boldsymbol{x}_{79}$, $\boldsymbol{x}_{80}$, $\boldsymbol{x}_{81}$, $\boldsymbol{x}_{82}$ for T_{87} on a log scale