SIAM J. MATRIX ANAL. APPL.
Vol. 31, No. 1, pp. 75–91

PERTURBATION SPLITTING FOR MORE ACCURATE EIGENVALUES*

RUI RALHA†

Abstract. Let T be a symmetric tridiagonal matrix with entries and eigenvalues of different magnitudes. For some T, small entrywise relative perturbations induce small errors in the eigenvalues, independently of the size of the entries of the matrix; this is certainly true when the perturbed matrix can be written as $\widetilde{T} = X^T T X$ with small $||X^T X - I||$. Even if it is not possible to express in this way the perturbations in every entry of T, much can be gained by doing so for as many as possible entries of larger magnitude. We propose a technique which consists of splitting multiplicative and additive perturbations to produce new error bounds which, for some matrices, are much sharper than the usual ones. Such bounds may be useful in the development of improved software for the tridiagonal eigenvalue problem, and we describe their role in the context of a mixed precision bisection-like procedure. Using the very same idea of splitting perturbations (multiplicative and additive), we show that when T defines well its eigenvalues, the numerical values of the pivots in the usual decomposition $T - \lambda I = LDL^T$ may be used to compute approximations with high relative precision.

Key words. symmetric tridiagonal matrices, eigenvalues, perturbation theory

AMS subject classifications. 15A15, 15A09, 15A23

DOI. 10.1137/070687049

1. Introduction.

Let A and E be n-by-n symmetric matrices. Let $\lambda_1 \leq \cdots \leq \lambda_n$ and $\widetilde{\lambda}_1 \leq \cdots \leq \widetilde{\lambda}_n$ be the eigenvalues of A and $\widetilde{A} = A + E$, respectively. Then $|\lambda_k - \widetilde{\lambda}_k| \leq \|E\|_2$. This is a classical result in the perturbation theory (see [44, pp. 101–102]), which is usually referred to as Weyl's theorem (see, for instance, [9, p. 198]).

Weyl's theorem can be used to get error bounds for the eigenvalues computed by any backward stable algorithm since such an algorithm computes eigenvalues $\widetilde{\lambda}_k$ that are the exact eigenvalues of $\widetilde{A} = A + E$, where $\|E\|_2 = O(\epsilon)\|A\|_2$. (Here and throughout the paper we will use ϵ to denote the rounding error unit.) This is a very satisfactory error bound for large eigenvalues, especially those of magnitude close to $\|A\|_2$, but eigenvalues much smaller than $\|A\|_2$ will have fewer correct digits (eventually none in extreme cases).

The decade starting in 1990 was fertile in new results on bounds for relative errors of eigenvalues and several authors have contributed to this [1], [4], [5], [16], [17], [22], [32], [33], [34], [35], [42]. In [22], Ipsen presents a good survey of the work done until 1998. Not surprisingly, many of the published results are for the Hermitian positive definite case. For an Hermitian indefinite matrix A and, more generally, for normal matrices, the Hermitian positive-semidefinite factor H in the polar decomposition $A = HU$, with U unitary, may be used to derive bounds for the eigenvalues of A (see [22, Theorems 2.4 and 2.10] and the references therein).

The first relative perturbation bound for eigenvalues is due to Ostrowski. Let $\widehat{A} = XAX^*$, with X nonsingular, be a multiplicative perturbation of an Hermitian matrix

*Received by the editors April 2, 2007; accepted for publication (in revised form) by Z. Drmač March 3, 2008; published electronically February 27, 2009. This research was supported by Portuguese Foundation for Science and Technology research program POCI 2010.
http://www.siam.org/journals/simax/31-1/68704.html
†Departamento de Matemática, Universidade do Minho, 4710-057 Braga, Portugal (r_ralha@math.uminho.pt).

A; for the eigenvalues λ_k and $\widehat{\lambda}_k$, of A and $\widehat{A}$, respectively, we have [21, Theorem 4.5.9]

$$\lambda_k \cdot \lambda_{\min}(XX^*) \leq \widehat{\lambda}_k \leq \lambda_k \cdot \lambda_{\max}(XX^*).$$

This result is at the heart of high relative accuracy theory for the eigenvalues of Hermitian matrices (and singular values). An immediate consequence, for real symmetric matrices, is the following (Theorem 2.1 in [16]): let A have eigenvalues λ_k and $\widehat{A} = X^T AX$ have eigenvalues $\widehat{\lambda}_k$. Then $|\widehat{\lambda}_k - \lambda_k| \leq |\lambda_k| \; ||X^T X - I||_2$. Following Demmel [9, p. 208], we will refer to this result as the relative Weyl's theorem.

Some types of matrices are known to define well their eigenvalues and/or singular values. In 1990, Demmel and Kahan [4] showed that small relative perturbations in the entries of any bidiagonal matrix cause small relative errors in the singular values, independent of their magnitudes. They also proposed the zero-shifted QR algorithm to compute such singular values with high relative accuracy. Another remarkable development in this area of fast and highly accurate computation of the singular values of bidiagonal matrices was the dqds algorithm [18], [38]. Furthermore, any matrix with an acyclic graph (bidiagonals and many others) defines well its singular values, and these may be computed to high accuracy using bisection [6].

In [11], Demmel et al. showed that it is possible to compute efficiently a highly accurate SVD of a dense rectangular matrix A from a rank-revealing decomposition (RRD) $A = XDY^T$, i.e., a decomposition where D is diagonal and X and Y are well conditioned (but otherwise arbitrary); furthermore, also in [11], a variety of matrix classes were described for which a special form of Gaussian elimination with complete pivoting does provide the necessary accuracy of the computed factors $\widetilde{X}$, $\widetilde{D}$, and $\widetilde{Y}$. For some structured matrices (these include, among others, Cauchy matrices, Vandermonde matrices, M-matrices, and totally nonnegative matrices), forward stable algorithms have been proposed for the computation of highly accurate RRD. See [10], [11], [12], [15], and [26], [27], [28], [29].

Congruence transformations play an important role in the perturbation theory of the eigenvalues of an Hermitian positive-definite matrix A (see [22, Corollary 2.2] and [34, Theorem 2.4]). For scaled diagonally dominant (*sdd*) matrices, diagonal congruence transformations may be used to pull the grading out of the matrix [1], [5], [35], [9]. If A is indefinite, the error bounds are the same as the error bounds for the eigenvalues of the best scaled version of the positive-definite polar factor of A (see [22, Corollary 2.6] and [42, Theorem 2.13]).

Symmetric tridiagonal matrices do not always define well their eigenvalues, not even in the positive-definite case. In this paper, we focus our attention on symmetric tridiagonal matrices with entries of different magnitudes. Our matrices, however, are not necessarily *sdd*.

Suppose that we are given a symmetric matrix A which has entries of different orders of magnitude and assume small relative perturbations of size $O(\varepsilon)$ in its entries (or, at least, small relative perturbations in the entries of larger size). With $\widetilde{A} = A+E$, it is clear that $\|E\|_2$ is proportional to the size of the largest entries in A, and the classical error bound, provided by Weyl's theorem, may not be very satisfactory for small eigenvalues, if they arise. For this reason, we attack $\widetilde{A}$ with a congruence $X^T \widetilde{A} X$ to get $\widehat{A} = A+F$ with $\|F\|_2 < \|E\|_2$ and $||X^T X - I||_2$ of size $O(\varepsilon)$; the relative Weyl's theorem gives

$$|\widetilde{\lambda}_k - \widehat{\lambda}_k| \leq ||X^T X - I||_2 \cdot |\widetilde{\lambda}_k|, \tag{1.1}$$

and we get

$$|\widetilde{\lambda}_k - \lambda_k| \le |\widetilde{\lambda}_k - \widehat{\lambda}_k| + |\widehat{\lambda}_k - \lambda_k| \le ||X^T X - I||_2 \cdot |\widetilde{\lambda}_k| + \|F\|_2 \,, \tag{1.2}$$

which, in some cases, is a much sharper bound than

$$|\widetilde{\lambda}_k - \lambda_k| \le \|E\|_2 \,. \tag{1.3}$$

In the following sections, we exploit this idea in the context of symmetric tridiagonal matrices, although it can also be applied to dense symmetric matrices. In section 2, we analyze the perturbation of the eigenvalues of affine transformations of Golub–Kahan matrices. Section 3 contains the main perturbation result, Theorem 3.1, which states that a symmetric tridiagonal matrix T, with diagonals a_j, defines well the eigenvalues whose magnitude is not much smaller than $\max |a_j|$. In section 4 we present a detailed numerical example to show that for matrices with entries of different magnitudes, depending upon the location of the entries of larger size, the eigenvalues may or may not be all well defined. In section 5 we describe a fast procedure that will produce an estimate for the value of $\|F\|_2$ in the bound (1.2). In sections 6 and 7 we present applications of our perturbation results; in section 6 we show that the numerical values of the pivots in the decomposition $T - \lambda I = LDL^T$, computed in the usual way, may be used to determine the eigenvalues with high relative accuracy, if the matrix T defines them well, and in section 7 we show that our results are useful in the context of a mixed precision bisection algorithm.

2. Constant main diagonal. It is well known that, for n even, the eigenvalues of the Golub–Kahan matrix

$$T(0) = \begin{bmatrix} 0 & b_1 & & & \\ b_1 & 0 & b_2 & & \\ & b_2 & 0 & \ddots & \\ & & \ddots & \ddots & b_{n-1} \\ & & & b_{n-1} & 0 \end{bmatrix} \tag{2.1}$$

are

$$-\sigma_1 \le \cdots \le -\sigma_{\frac{n}{2}} \le \sigma_{\frac{n}{2}} \le \cdots \le \sigma_1, \tag{2.2}$$

where σ_k $(k = 1, \ldots, \frac{n}{2})$ are the singular values of

$$B = \begin{bmatrix} b_1 & b_2 & & \\ & b_3 & \ddots & \\ & & \ddots & b_{n-2} \\ & & & b_{n-1} \end{bmatrix} \tag{2.3}$$

(see, for instance, Lemma 5.5 in [9]). This relation may be used in both directions; that is, one may compute singular values of B as the corresponding positive eigenvalues of $T(0)$ or one may compute eigenvalues of $T(0)$ from the corresponding singular values of B. This last option may also be used for the computation of the eigenvalues of a skew-symmetric tridiagonal matrix with high relative accuracy (see [41]). We will therefore be interested in matrices with the structure given in (2.1), with n even or odd. We have the following result.

PROPOSITION 2.1. *Let $T(0)$ be as given in* (2.1)*, and let D_{2k-1} ($k = 1, \ldots, \frac{n}{2}$ if n is even and $k = 1, \ldots, \frac{n+1}{2}$ if n is odd) and D_{2k} ($k = 1, \ldots, \frac{n}{2}$ if n is even and $k = 1, \ldots, \frac{n-1}{2}$ if n is odd) be the principal minors of $T(0)$ of order odd and even, respectively. We have*

$$D_{2k-1} = 0, \;\; D_{2k} = (-1)^k \cdot \prod_{j=1}^{k} b_{2j-1}^2. \tag{2.4}$$

Thus, for n even, $T(0)$ is singular if and only if $b_{2j-1} = 0$ for some j, $1 \leq j \leq k$; if n is odd, then $D_n = 0$; i.e., $T(0)$ is always singular.

Proof. The proof follows easily from $D_1 = 0$, $D_2 = -b_1^2$, and the relation $D_j = -b_{j-1}^2 \times D_{j-2}$ for $j \geq 3$. ∎

When n is odd, we may keep relating $T(0)$ to a bidiagonal matrix. For this, we construct a matrix of even order by adding a row and a column of zeros to $T(0)$. The resulting matrix has a double eigenvalue equal to zero. The corresponding bidiagonal in (2.3) is now replaced by the singular matrix with diagonal entries $b_1, \ldots, b_{n-2}$, $b_n = 0$ and superdiagonal entries $b_2, \ldots, b_{n-1}$.

Small relative perturbations of the off-diagonal pairs of $T(0)$ may be expressed in terms of a congruence transformation $X^T T(0) X$ with X diagonal very close to identity (see [1], [16], and [22, Example 5.1]). Therefore, $T(0)$ defines well its eigenvalues (even when n is odd, because the zero eigenvalue is unchanged by perturbations in the off-diagonal entries). From [9, Theorem 5.13] we may conclude the following.

COROLLARY 2.2. *Let $T(0)$ be as given in* (2.1)*, and let $\widetilde{T}(0)$ be the tridiagonal matrix which results from $T(0)$ by replacing each b_k with $\widetilde{b}_k = b_k(1 + \delta_k)$ with $|\delta_k| \leq \varepsilon \ll 1$. Let $\lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of $T(0)$ and $\widetilde{\lambda}_1 \leq \cdots \leq \widetilde{\lambda}_n$ the eigenvalues of $\widetilde{T}(0)$. For every eigenvalue (even if zero) we can write*

$$|\widetilde{\lambda}_k - \lambda_k| \leq \xi\,(n, \varepsilon)\,|\lambda_k|\,, \tag{2.5}$$

where

$$\xi\,(n, \varepsilon) = (2n - 1)\varepsilon + O\left(\varepsilon^2\right). \tag{2.6}$$

Now, we consider affine transformations of $T(0)$. If $T(c)$ is a symmetric tridiagonal matrix whose main diagonal entries are equal to a constant c, then $T(0) = T(c) - cI$ has zeros in the main diagonal and Corollary 2.2 does apply. We have the following.

PROPOSITION 2.3. *Let $\lambda_k(0)$ and $\lambda_k(c)$ be the eigenvalues of $T(0)$ and $T(c) = T(0) + cI$, respectively; let $\widetilde{\lambda}_k(0)$ be the eigenvalues of $\widetilde{T}(0)$, as defined in Corollary* 2.2*, and $\widetilde{\lambda}_k(c)$ the eigenvalues of $\widetilde{T}(c) = \widetilde{T}(0) + cI$. For $\lambda_k(c) \neq 0$ we have*

$$|\widetilde{\lambda}_k(c) - \lambda_k(c)| \leq \xi(n, \varepsilon) \left|1 - \frac{c}{\lambda_k(c)}\right| |\lambda_k(c)|\,, \tag{2.7}$$

where $\xi(n, \varepsilon)$ is as given in (2.6).

Proof. Since $\lambda_k(c) = \lambda_k(0) + c$ and $\widetilde{\lambda}_k(c) = \widetilde{\lambda}_k(0) + c$, we have $\widetilde{\lambda}_k(c) - \lambda_k(c) = \widetilde{\lambda}_k(0) - \lambda_k(0)$; using (2.5), we get

$$|\widetilde{\lambda}_k(c) - \lambda_k(c)| \leq \xi(n, \varepsilon)\,|\lambda_k(0)|\,,$$

which, for $\lambda_k(c) \neq 0$, can be written as

$$|\widetilde{\lambda}_k(c) - \lambda_k(c)| \leq \xi(n, \varepsilon) \left|\frac{\lambda_k(0)}{\lambda_k(c)}\right| |\lambda_k(c)|\,. \tag{2.8}$$

Replacing $\lambda_k(0)$ with $\lambda_k(c) - c$ gives (2.7). ∎

Small relative perturbations in the off-diagonal entries of $T(c)$ cause relative errors in the eigenvalues which depend upon the ratio

$$\frac{\lambda_k(0)}{\lambda_k(c)} = 1 - \frac{c}{\lambda_k(c)}. \tag{2.9}$$

Therefore, we see that the relative errors will be small except for those eigenvalues $\lambda_k(c)$ such that $|\lambda_k(0)| \gg |\lambda_k(c)|$, i.e.,

$$|\lambda_k(c)| \ll |c|\,. \tag{2.10}$$

Furthermore, (2.7) shows that the relative error of $\widetilde{\lambda}_k(c)$ approaches zero when $\lambda_k(c)$ gets close to c.

Example 1. Consider the matrix

$$T(1) = \begin{bmatrix} 1 & 10^6 & & & & \\ 10^6 & 1 & 1 & & & \\ & 1 & 1 & 1 & & \\ & & 1 & 1 & 1 & \\ & & & 1 & 1 & 10^6 \\ & & & & 10^6 & 1 \end{bmatrix}. \tag{2.11}$$

The function eig of MATLAB (version 7.4) produces the following approximations for the eigenvalues (note that with a previous version of MATLAB we got much worse values for $\widetilde{\lambda}_3(1)$ and $\widetilde{\lambda}_4(1)$):

$$\begin{array}{ll} \widetilde{\lambda}_1(1) = -9.999990000005000e{+}005, & \widetilde{\lambda}_2(1) = -9.999990000005000e{+}005, \\ \widetilde{\lambda}_3(1) = 1.139421890172798e{-}012, & \widetilde{\lambda}_4(1) = 1.999999999999141e{+}000, \\ \widetilde{\lambda}_5(1) = 1.000001000000500e{+}006, & \widetilde{\lambda}_6(1) = 1.000001000000500e{+}006. \end{array}$$

The classical error analysis gives us, with $\epsilon = 2^{-52}$, for all $k = 1, \ldots, 6$,

$$|\widetilde{\lambda}_k(1) - \lambda_k(1)| \le O(\epsilon)\,\|T(1)\|_2 = O(10^{-10}).$$

Thus, for $k \neq 3$ and $k \neq 4$, $\widetilde{\lambda}_k(1)$ is an accurate approximation of the corresponding true eigenvalue $\lambda_k(1)$, and $\widetilde{\lambda}_4(1)$ has at least 9 or 10 correct decimal digits. Interestingly, we may improve upon the computed values $\widetilde{\lambda}_3(1)$ and $\widetilde{\lambda}_4(1)$. Since we know the exact value $\det(T(1)) = 2 \times 10^{12} - 1$, we use the relation

$$\lambda_3(1) = \det(T(1)) / \prod_{k=1, k\neq 3}^{6} \lambda_k(1)$$

to compute an approximation

$$\overline{\lambda}_3(1) = fl\left(\det(T(1)) / \prod_{k=1, k\neq 3}^{6} \widetilde{\lambda}_k(1)\right) \tag{2.12}$$

which has at least nine correct decimal significant digits. We have

$$\overline{\lambda}_3(1) = \det(T(1))/\left(\prod_{k=1,k\neq 3}^{6} \lambda_k(1)\,(1+\phi_k)\right)(1+\kappa\epsilon)$$
$$= \lambda_3(1)\cdot\left((1+\kappa\epsilon)\prod_{k=1,k\neq 3}^{6}(1+\phi_k)^{-1}\right),$$

where ϕ_k for $k \neq 3$ is the relative error in $\widetilde{\lambda}_k(1)$ and the term $\kappa\epsilon$, with $\kappa \leq 5.05$, accounts for the rounding errors in the four multiplications and one division. Since the relative errors ϕ_k in the four eigenvalues of larger size are all bounded by $O\left(2^{-52}\right)$, the size of the relative error in $\overline{\lambda}_3(1)$ is determined essentially by the size of ϕ_4, which we know to be bounded by $O\left(10^{-10}\right)$. The computation of (2.12) in MATLAB produces $\overline{\lambda}_3(1) = 9.999999999999297e{-}013$. Since the interval $[\lambda_3(1),\ \lambda_4(1)]$ of the true eigenvalues is known to be centered in $c = 1$, we compute $\overline{\lambda}_4(1) = 2 - \overline{\lambda}_3(1) = 1.999999999999000e{+}000$ with 16 correct digits. Again, we may use (2.12), replacing $\widetilde{\lambda}_4(1)$ with $\overline{\lambda}_4(1)$ to compute $\overline{\overline{\lambda}}_3(1) = 1.000000000000000e{-}012$ with a relative error bounded by $O(\epsilon)$. Now, according to Proposition 2.3, if MATLAB could deliver the exact eigenvalues of a matrix differing from $T(1)$ by relative perturbations of size $O(\epsilon)$ in the off-diagonal entries,[1] $\widetilde{\lambda}_4(1)$ would be closer to $\overline{\lambda}_4(1)$, and for $\widetilde{\lambda}_3(1)$ we would have

$$|\widetilde{\lambda}_3(1) - \lambda_3(1)| \leq \left|1 - \frac{1}{\lambda_3(1)}\right| O(\epsilon)\,|\lambda_3(1)| \approx 10^{-4}\,|\lambda_3(1)|\,,$$

and such approximation, although not as good as $\overline{\overline{\lambda}}_3(1)$ or even $\overline{\lambda}_3(1)$, is significantly better than the computed $\widetilde{\lambda}_3(1)$. It is also worth mentioning that in MATLAB, svd(T) and [L,U]=lu(T); eig(U*L), where T is the matrix in our example, both produce approximations $\widetilde{\lambda}_4(1)$ and $\widetilde{\lambda}_3(1)$ which do satisfy the error bound (2.7).[2]

We conclude this section by emphasizing that the matrix in our example is not *sdd* and the theory of Barlow and Demmel does not apply here.

3. A perturbation theory result. In the previous section, we showed that small relative changes in the off-diagonal entries of a symmetric tridiagonal matrix with constant main diagonal c do not cause too much perturbation in those eigenvalues of magnitude not much smaller than the constant $|c|$. To this end, we have used a simple affine transformation of the given matrix to produce a Golub–Kahan matrix whose relative perturbations in the off-diagonal pairs may be entirely expressed in terms of a congruence transformation $X^T T(0) X$, with X very close to identity. A similar result may be obtained without the affine transform by directly expressing the perturbations in the off-diagonal entries in terms of a congruence transformation. This is a more general procedure since it applies to any symmetric tridiagonal matrix. We have the following theorem.

[1] This is what the bisection method can actually deliver; see section 6.
[2] Zlatko Drmač has brought to our attention the accuracy of these approximations.

THEOREM 3.1. *Let*

$$T = \begin{bmatrix} a_1 & b_1 & & \\ b_1 & \ddots & \ddots & \\ & \ddots & \ddots & b_{n-1} \\ & & b_{n-1} & a_n \end{bmatrix} \tag{3.1}$$

and

$$\widetilde{T} = \begin{bmatrix} a_1\,(1+\eta_1) & b_1\,(1+\delta_1) & & \\ b_1\,(1+\delta_1) & \ddots & \ddots & \\ & \ddots & \ddots & b_{n-1}\,(1+\delta_{n-1}) \\ & & b_{n-1}\,(1+\delta_{n-1}) & a_n\,(1+\eta_n) \end{bmatrix}, \tag{3.2}$$

where δ_k and η_k are tiny quantities such that $|\delta_k| \le \varepsilon$ and $|\eta_k| \le \varepsilon$. Denoting by λ_k and $\widetilde{\lambda}_k$ the ordered eigenvalues of T and $\widetilde{T}$, respectively, the following relation holds, for each $k = 1, \ldots, n$:

$$|\lambda_k - \widetilde{\lambda}_k| < 2.02 n\varepsilon \left(\max_j |a_j| + |\widetilde{\lambda}_k| \right). \tag{3.3}$$

Proof. We use a diagonal congruence to account for all the off-diag perturbations and then just see what it does to the diagonal entries: lo and behold, it makes just a few more changes from what was there initially. Concretely, if we write

$$\widehat{T} = X^T \widetilde{T} X \tag{3.4}$$

with X diagonal, $X(1,1) = 1$, $X(2,2) = (1+\delta_1)^{-1}$, and

$$X(j,j) = (1+\delta_{j-1})^{-1}\, X(j-1, j-1)^{-1}, \qquad j = 3, \ldots, n, \tag{3.5}$$

we get $\widehat{T}(i,j) = T(i,j)$ for $i \ne j$, $\widehat{T}(1,1) = a_1\,(1+\eta_1)$, and

$$\widehat{T}(j,j) = a_j\,(1+\eta_j) \cdot X(j,j)^2, \qquad j = 2, \ldots, n. \tag{3.6}$$

We write

$$X(j,j)^2 = 1 + \phi_j, \tag{3.7}$$

and since $\phi_1 = 0$ and $|\delta_j| \le \varepsilon$, from (3.5), assuming that $2\,(n-1)\,\varepsilon \le 0.01$, we get

$$|\phi_j| \le 2.02\,(j-1)\,\varepsilon, \qquad j = 2, \ldots, n, \tag{3.8}$$

and $||X^T X - I||_2 \le \max_j |\phi_j| < 2.02 n\varepsilon$. From (3.6)–(3.8) and taking into account that $|\eta_j| \le \varepsilon$, we may write, for each $j = 1, \ldots, n$, assuming that $(2n-1)\,\varepsilon \le 0.01$, $\widehat{T}(j,j) = a_j\,(1+\theta_j)$ with $|\theta_j| \le 1.01\,(2j-1)\,\varepsilon$. Therefore, we have $\widehat{T} = T + F$ with F a diagonal matrix such that

$$\|F\|_2 = \max_j |a_j|\,|\theta_j| < 2.02 n\varepsilon \cdot \max_j |a_j|. \tag{3.9}$$

Applying the relative Weyl's theorem to matrices $\widehat{T}$ and $\widetilde{T}$ in (3.4), we get $|\widehat{\lambda}_k - \widetilde{\lambda}_k| \leq |\widetilde{\lambda}_k| \cdot ||X^T X - I||_2$, and we may finally write $|\lambda_k - \widetilde{\lambda}_k| \leq |\lambda_k - \widehat{\lambda}_k| + |\widehat{\lambda}_k - \widetilde{\lambda}_k| \leq \|F\|_2 + |\widetilde{\lambda}_k| \cdot ||X^T X - I||_2$, which, after some simplifications, gives (3.3). ∎

If $\widetilde{\lambda}_k \neq 0$, the bound (3.3) may be written as

$$\frac{\left|\lambda_k - \widetilde{\lambda}_k\right|}{\left|\widetilde{\lambda}_k\right|} < 2.02n\varepsilon \left(1 + \frac{\max_j |a_j|}{\left|\widetilde{\lambda}_k\right|}\right). \tag{3.10}$$

Part of the novelty of Theorem 3.1 for relative perturbation theory is that, as expressed in (3.10), a general symmetric tridiagonal matrix T defines well those eigenvalues whose magnitude is not much smaller than $\max |a_j|$.

For the case of a matrix with zeros in the main diagonal, we get from (3.3)

$$|\lambda_k - \widetilde{\lambda}_k| \leq 2.02n\varepsilon |\widetilde{\lambda}_k| \tag{3.11}$$

and we note that this is essentially the bound given in (2.5), with $|\lambda_k|$ replaced with $|\widetilde{\lambda}_k|$.

It must be observed that there are many distinct congruences X which are able to produce $\widehat{T}$ with unperturbed off-diagonal entries. We have used X with $X(1,1) = 1$, but it is possible to use a different X, setting $X(k,k) = 1$, for any $k = 1, \ldots, n$; then, we choose the values of $X(k-1,k-1), \ldots, X(1,1)$ to remove perturbations from entries $\widetilde{b}_{k-1}, \ldots, \widetilde{b}_1$, by this order, and $X(k+1,k+1), \ldots, X(n,n)$ to remove perturbations from entries $\widetilde{b}_k, \ldots, \widetilde{b}_{n-1}$. In particular, by choosing $k = n/2$ we may reduce the bounds (3.3) and (3.10) by a factor of 2.

Finally, we remark that there is a diagonal X which, besides the off-diagonal perturbations, also expresses, in multiplicative terms, the perturbation in any diagonal entry $\widetilde{a}_k$: $X(k,k)$ is chosen to remove the perturbation in $\widetilde{a}_k$, and the remaining entries of X are determined as we have just described. So, in the bounds (3.3) and (3.10) we may replace $\max |a_j|$ with the second largest absolute value of the diagonal entries of T.

4. More general perturbations: An example. There are matrices for which the bound (3.10) is sharp. This is the case with matrices of constant main diagonal c since, as we have seen in section 2, the relative error in $\widetilde{\lambda}_k(c)$ depends upon the ratio $c/\lambda_k(c)$. In discussing the relative errors of small eigenvalues computed with the bisection method, Wilkinson also observed (see [44, p. 307]) that the method (which we know to be able to compute accurately the eigenvalues if the matrix defines them well) could not compute accurately the small eigenvalues of such a matrix.

However, we know that there are other matrices for which the bound (3.10) is too pessimistic. This is the case of the *sdd* matrices. We now show that there are other matrices, not *sdd*, which define well their eigenvalues, even in cases where their size is much smaller than that of some of the diagonal entries.

In the previous section, we expressed the perturbations in the off-diagonal entries in terms of a diagonal congruence,

$$\widehat{T} = X^T \widetilde{T} X. \tag{4.1}$$

Although X does not account for the perturbations in the diagonal entries, the key point of our analysis is based upon the fact that

$$\widehat{T} = T + F \tag{4.2}$$

with $\|F\|_2$ independent of the size of the off-diagonal entries.

In a more general situation, T may have entries of different order of magnitude, and we are interested in expressing the perturbations in the entries of larger size, independently of their location, in terms of the transformation expressed in (4.1). We point out that in the general case, F in (4.2) does not need to be a diagonal matrix. Again, we start with a numerical example to motivate the general procedure that will be proposed in the next section.

Example 2. Consider the matrices

$$T_1 = \begin{bmatrix} 1 & 10^5 & 0 \\ 10^5 & 10^5 & 10^5 \\ 0 & 10^5 & 1 \end{bmatrix}, \quad T_2 = \begin{bmatrix} 10^5 & 10^5 & 0 \\ 10^5 & 10^5 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

The approximations for the eigenvalues of T_1 and T_2, computed with MATLAB, are

$$\begin{aligned} \lambda_1(T_1) &= -9.999933333407408e{+}004, \\ \lambda_2(T_1) &= 1.000000000014616e{+}000, \\ \lambda_3(T_1) &= 2.000003333340741e{+}005 \end{aligned}$$

and

$$\begin{aligned} \lambda_1(T_2) &= -3.660259320914954e{-}001, \\ \lambda_2(T_2) &= 1.366023432085007e{+}000, \\ \lambda_3(T_2) &= 2.000000000025000e{+}005. \end{aligned}$$

In both cases, we know that the absolute errors in these approximations have a bound of size $O\left(10^{-11}\right)$ because the norm of the matrices is $O\left(10^5\right)$ and ϵ is $O\left(10^{-16}\right)$. To gain insight into the influence of perturbations, we used again the function eig of MATLAB to compute the eigenvalues of the matrices

$$\widetilde{T}_1 = \begin{bmatrix} 1(1+\eta_1) & 10^5(1+\delta_1) & 0 \\ 10^5(1+\delta_1) & 10^5(1+\eta_2) & 10^5(1+\delta_2) \\ 0 & 10^5(1+\delta_2) & 1(1+\eta_3) \end{bmatrix}$$

and

$$\widetilde{T}_2 = \begin{bmatrix} 10^5(1+\eta_1') & 10^5(1+\delta_1') & 0 \\ 10^5(1+\delta_1') & 10^5(1+\eta_2') & 1(1+\delta_2') \\ 0 & 1(1+\delta_2') & 1(1+\eta_3') \end{bmatrix}$$

with $\eta_k, \eta_k', \delta_k$, and δ_k' randomly generated, all bounded by $\varepsilon = 10^{-7}$ in absolute value. We got the errors

$$\begin{aligned} \lambda_1(T_1) - \lambda_1(\widetilde{T}_1) &\approx \quad 3.1e{-}008, \\ \lambda_2(T_1) - \lambda_2(\widetilde{T}_1) &\approx -9.9e{-}009, \\ \lambda_3(T_1) - \lambda_3(\widetilde{T}_1) &\approx \quad 8.9e{-}010 \end{aligned}$$

and

$$\begin{aligned} \lambda_1(T_2) - \lambda_1(\widetilde{T}_2) &\approx -7.1e{-}003, \\ \lambda_2(T_2) - \lambda_2(\widetilde{T}_2) &\approx \quad 5.1e{-}004, \\ \lambda_3(T_2) - \lambda_3(\widetilde{T}_2) &\approx -1.4e{-}008, \end{aligned}$$

which are clearly due to the perturbations, not to the numerical errors in the function eig. We see that the eigenvalues of $\widetilde{T}_1$ exhibit absolute errors much smaller than $\|T_1\|_2\,\varepsilon \approx 2\times 10^{-2}$, which do correspond to relative errors smaller than $\varepsilon = 10^{-7}$, but the error in $\lambda_1(\widetilde{T}_2)$ is close to $\|T_2\|_2\,\varepsilon \approx 2\times 10^{-2}$. Why does T_1 define well its eigenvalues? First, we note that T_1 is not *sdd*; therefore [1, Theorem 4] does not apply. Furthermore, we computed the polar factor H of $T1 = T_1$, in MATLAB, from $[V, D] = \mathrm{eig}(T1); H = V * \mathrm{abs}(D) * V'$, and observed that the results of [22, section 2.8] are also unable to explain the good results obtained for T_1. Now, take

$$X = \begin{bmatrix} (1+\delta_1)^{-1}(1+\eta_2)^{1/2} & & \\ & (1+\eta_2)^{-1/2} & \\ & & (1+\delta_2)^{-1}(1+\eta_2)^{1/2} \end{bmatrix}$$

and verify that for $\widehat{T}_1 := X^T\widetilde{T}_1 X$ we get

$$\widehat{T}_1 = \begin{bmatrix} (1+\eta_1)(1+\delta_1)^{-2}(1+\eta_2) & 10^5 & 0 \\ 10^5 & 10^5 & 10^5 \\ 0 & 10^5 & (1+\eta_3)(1+\delta_2)^{-2}(1+\eta_2) \end{bmatrix}.$$

As in the example given in section 2, we have managed to produce a matrix $\widehat{T}_1$ with no perturbations in the entries of larger size and, as a consequence, we have $\widehat{T}_1 = T_1 + F$ with $\|F\|_2$ much smaller than $||T_1 - \widetilde{T}_1||_2$; furthermore, since X is close to the identity matrix, the relative Weyl's theorem guarantees that the eigenvalues of $\widehat{T}_1$ and $\widetilde{T}_1$ are close. The situation is quite different with T_2 because it is not possible to express the perturbations in all the larger entries $T_2(1,1)$, $T_2(2,2)$, $T_2(1,2)$, and $T_2(2,1)$ in terms of a multiplicative perturbation $X^T\widetilde{T}_2 X$, with some X close to the identity matrix.

5. A fast procedure to compute the error bound. In general, given a symmetric tridiagonal T with entries of different magnitudes and small relative perturbations, as expressed in $\widetilde{T}$ given in (3.2), we want to find a diagonal matrix X, with entries very close to the unity, such that the relations (4.1) and (4.2) hold, with $\|F\|_2$ as small as possible.

The example in the previous section shows that the rate of success of the procedure depends upon the locations of the entries of larger magnitude relatively to each other. Since our goal is to minimize, as much as possible, the size of the perturbed entries in $\widehat{T}$, we start by producing a sequence of $2n-1$ numbers, sorting the entries of T by decreasing order of their absolute values and "clean" as many entries as possible in this sequence. To simplify the presentation, we say that we clean the entry (i,j) when, in the course of the transformation (4.1), we get $\widehat{T}(i,j) = T(i,j)$, getting rid of the perturbation in $\widetilde{T}(i,j)$. In practice, we do not carry out such an operation, we just need to assume that it has been done. (This is in fact a combinatorial task and does not require any arithmetic at all.) By "operation of index k," $k = 1,\ldots,n$, we will mean the transformation that multiplies the kth row and the kth column of $\widetilde{T}$, i.e., the diagonal congruence associated with $X(k,k)$ in (4.1). We illustrate the cleaning procedure with the following example.

Example 3. Suppose that our matrix T, of order $n = 5$, is such that

$$|a_1| \geq |b_3| \geq |b_2| \geq |a_4| \geq |b_4| \geq |a_2| \geq |b_1| \geq |a_5| \geq |a_3|\,. \tag{5.1}$$

First, we remove the perturbation from $a_1(1+\eta_1)$, the entry of largest size, by setting $X(1,1) = (1+\eta_1)^{-1/2}$; because we want a_1 to remain unperturbed, we close the index 1; i.e., it is removed from the set of indices allowed for subsequent operations. Next, to clean $b_3(1+\delta_3)$, we have two options: an operation of index 3 or an operation of index 4. Note that after cleaning $b_3(1+\delta_3)$ the indices 3 and 4 will be closed; therefore, before cleaning $b_3(1+\delta_3)$, we clean $a_4(1+\eta_4)$, since $|a_4| \geq |a_3|$, and close index 4. Then, we clean $b_3(1+\delta_3)(1+\eta_4)^{-1/2}$ and close index 3. The next entry in (5.1) is b_2, and the set of indices still open is $\{2,5\}$. So, we clean $b_2(1+\delta_2)(1+\delta_3)^{-1}(1+\eta_4)^{1/2}$. At this point, it is still possible to clean $b_4(1+\delta_4)(1+\eta_4)^{-1/2}$, which is next in (5.1), and this is the last entry to be cleaned. To summarize, with X diagonal such that

$$\begin{aligned}
X(1,1) &= (1+\eta_1)^{-1/2},\\
X(2,2) &= (1+\delta_2)^{-1}(1+\delta_3)(1+\eta_4)^{-1/2},\\
X(3,3) &= (1+\delta_3)^{-1}(1+\eta_4)^{1/2},\\
X(4,4) &= (1+\eta_4)^{-1/2},\\
X(5,5) &= (1+\delta_4)^{-1}(1+\eta_4)^{1/2},
\end{aligned}$$

we get the following entries for $\widehat{T} = X^T\widetilde{T}X$:

$$\begin{aligned}
&\widehat{a}_1 = a_1, \quad \widehat{a}_4 = a_4, \quad \widehat{b}_2 = b_2, \quad \widehat{b}_3 = b_3, \quad \widehat{b}_4 = b_4,\\
&\widehat{a}_2 = a_2(1+\eta_2)(1+\delta_2)^{-2}(1+\delta_3)^2(1+\eta_4)^{-1},\\
&\widehat{a}_3 = a_3(1+\eta_3)(1+\delta_3)^{-2}(1+\eta_4),\\
&\widehat{a}_5 = a_5(1+\eta_5)(1+\delta_4)^{-1}(1+\eta_4)^{1/2},\\
&\widehat{b}_1 = b_1(1+\delta_1)(1+\eta_1)^{-1/2}(1+\delta_2)^{-1}(1+\delta_3)(1+\eta_4)^{-1/2}.
\end{aligned}$$

Therefore, we may write $\widehat{T} = T + F$ with

$$F = \begin{bmatrix}
0 & b_1\delta_1' & & & \\
b_1\delta_1' & a_2\eta_2' & 0 & & \\
 & 0 & a_3\eta_3' & 0 & \\
 & & 0 & 0 & 0\\
 & & & 0 & a_5\eta_5'
\end{bmatrix},$$

where δ_1', η_2', η_3', and η_5' are all of magnitude $O(\varepsilon)$ and the null entries do correspond to those positions that have been cleaned. In our example, if $|a_1| \gg |a_2|$ (remember that a_2 is the entry of largest size that has not been possible to clean), then $||T-\widetilde{T}||_2$ is much larger than $\|F\|_2$ and the bound (1.2) will be much sharper than the bound (1.3) for the eigenvalues of size significantly smaller than $\|T\|_2$. The gain, in terms of the sharpness of the bound that we get for the absolute errors in the eigenvalues, depends roughly on how large the ratio $|a_1|/|a_2|$ is.

We should remark that the described procedure is not optimal for symmetric tridiagonal matrices whose entries satisfy the condition $\max|a_j| < \min|b_j|$. In fact, by closing indices $1,\ldots,n$, in this ordering, we may clean all off-diagonal elements, as we did in Theorem 3.1; however, the procedure, as presented before, will clean first the off-diagonal entries of larger size and will not allow, in general, all off-diagonal entries to be cleaned. There are other cases for which our cleaning algorithm is not optimal and where it may be possible to use combinatorial analysis to improve the technique.

It should be noted that it is not possible to clean every entry of a submatrix

$$\begin{bmatrix} a_j & b_j \\ b_j & a_{j+1} \end{bmatrix}$$

for any $j = 1, \ldots, n-1$. Therefore, the error bound in (1.2) will never be smaller than $M \cdot O(\varepsilon)$, where

$$M = \max_{1 \le j \le n-1} \min \left\{ |a_j|, |b_j|, |a_{j+1}| \right\}.$$

In particular, for the matrix T_2 in the example given in section 4, we have $M = 10^5$.

We finish this section by noting that our procedure can be readily adapted for general symmetric matrices A to clean up to n entries. As for the tridiagonal case, after ordering the nonzero entries of A, in decreasing absolute values, we clean as many entries as possible in this sequence. To clean a pair of off-diagonal entries, say, $A(i,j)$ and $A(j,i)$, there may be a choice for the index to use (if both i and j are open). Because after cleaning $A(i,j)$ and $A(j,i)$, both indices will be closed, we may, similarly to the procedure that we have used in the tridiagonal case, clean first $A(i,i)$ or $A(j,j)$, the one of larger absolute value. A better solution may consist in looking at the size of the remaining entries in the ith and jth columns (or rows) and trying to clean the one of bigger size. Let this be the pair $A(i,p)$ and $A(p,i)$ for some $p \neq i$ and $p \neq j$. If the index p is already close, then it is certainly a good decision to clean entries $A(i,p)$ and $A(p,i)$ before cleaning entries $A(i,j)$ and $A(j,i)$. However, if index p is still open, the cleaning of the entries $A(i,p)$ and $A(p,i)$ closes p, and this may prevent the eventual cleaning of a bigger entry in the pth column (row). For this reason, it appears to be sensible to clean the pair $A(i,q)$ and $A(q,i)$ such that

$$|A(i,q)| = |A(q,i)| = \max_{r \in C} \left\{ |A(r,i)|, |A(r,j)| \right\},$$

where C denotes the set of indices which are already closed at this point. As it happens with the tridiagonal case, we cannot claim that this always produces the best possible X. Nevertheless, this procedure is very fast and may improve significantly the error bounds for the eigenvalues.

6. Accurate computation of the pivots. Using the very same idea of combining additive perturbations with multiplicative perturbations, we now show that the numerical values of the pivots of a symmetric tridiagonal matrix, computed through the formulae (6.1), may be used to determine eigenvalues with high relative accuracy. This may be of interest in the practical development of a parallel implementation of an algorithm which combines bisection with a faster zerofinder. Even in the context of sequential processing, there may still be room for new codes to take advantage of special features of matrices like those exploited in this paper. For instance, the state-of-the-art dqds algorithm, described in [40] and now implemented in the DSTEMR routine of the latest release of LAPACK, cannot guarantee high relative accuracy for the eigenvalues of symmetric tridiagonal indefinite matrices that define well their eigenvalues. In such cases, the only LAPACK routine that warrants full precision is DSTEBZ which implements the bisection method.

For a matrix T as given in Theorem 3.1, bisection (and related methods) is based upon the decomposition $T - \lambda I = LDL^T$, where L is unit lower bidiagonal and $D = \mathrm{diag}(q_1, \ldots, q_n)$ is diagonal. The numbers q_k are computed through

$$\begin{aligned} & q_1(\lambda) = a_1 - \lambda, \\ & q_k(\lambda) = a_k - \lambda - b_{k-1}^2 / q_{k-1}(\lambda), \qquad k = 2, \ldots, n. \end{aligned} \tag{6.1}$$

For each λ, the inertia of $T - \lambda I$, which is given by the signs of the $q_k(\lambda)$, can be used to locate eigenvalues. It is well known (see [25, p. 35] and [9, p. 230]) that the bisection method is able to compute the eigenvalues of a symmetric matrix which is very close to the exact one. In fact, the values $q_k(\lambda)$ computed with (6.1) in floating point arithmetic have the same signs as the values $\widetilde{q}_k(\lambda)$ that would be obtained if exact arithmetic was carried out with the matrix $\widetilde{T}$ such that[3]

$$\begin{aligned} &\widetilde{a}_k = a_k, \\ &\widetilde{b}_k = b_k\left(1+\delta_k\right), \text{ where } |\delta_k| \leq 2.5\epsilon + O\left(\epsilon^2\right). \end{aligned} \tag{6.2}$$

However, if one is to use not only the signs but also the numerical values of $q_k(\lambda)$, in the context of a method with a faster convergence rate, the previous result does not apply because it does not guarantee that the computed values of $q_k(\lambda)$ do correspond to a matrix $\widetilde{T}$ with entries satisfying the relations (6.2). In the context of the computation of singular values of bidiagonal matrices with relative accuracy, Demmel and Kahan [4, p. 24] briefly mentioned the possible use of zerofinders, different from simple bisection, to refine intervals; however, no details were given on the accuracy of the computed values $q_k(\lambda)$.

It is not true, in general, that the computed pivots are the exact ones for a matrix with small relative changes in its entries. However, an analysis similar to that used by Wilkinson for the leading principal minors (see [44, p. 303]) allows us to show that the computed values $q_k(\lambda)$ are the exact ones corresponding to off-diagonal entries with small relative perturbations and diagonal entries with additive perturbations of size $(a_k - \lambda)\, O(\epsilon)$. Writing the perturbed diagonal entries in the form

$$\widetilde{a}_k = a_k\left(1 + O(\epsilon)\right) - \lambda O(\epsilon),$$

we see that the computed $q_k(\lambda)$ do correspond to a matrix with small relative perturbations in its entries plus a diagonal additive perturbation of size $|\lambda|\, O(\epsilon)$. More precisely, we have the next theorem.

THEOREM 6.1. *Let T be a tridiagonal matrix as in* (3.1). *For a given λ, the values of $q_1(\lambda), \ldots, q_n(\lambda)$ computed with the formulae* (6.1) *are the exact values corresponding to a matrix having diagonal entries $\widetilde{a}_k = a_k(1+\eta_k) - \lambda\eta_k$ and off-diagonal entries $\widetilde{b}_{k-1} = b_{k-1}(1+\delta_{k-1})$, where*

$$\begin{cases} |\eta_k| \leq 2.02\epsilon, \\ |\delta_{k-1}| \leq 3.03\epsilon. \end{cases} \tag{6.3}$$

Proof. The proof is by induction. The result is obviously true for $k = 1$. Let us assume that the computed

$$\widetilde{q}_1(\lambda), \ldots, \widetilde{q}_{r-1}(\lambda)$$

are exact for a matrix having modified elements up to $\widetilde{a}_{r-1}$ and $\widetilde{b}_{r-2}$ and then show that the computed $\widetilde{q}_r(\lambda)$ is the exact value for a matrix having those modified elements and also the elements $\widetilde{a}_r$ and $\widetilde{b}_{r-1}$. If we assume that $\widetilde{q}_{r-1}(\lambda) \neq 0$ and represent by ε_1, ε_2, ε_3, and ε_4 the individual errors in the four operations involved in (6.1), we get

[3] In [6], it is shown that a similar result holds for symmetric matrices with acyclic graphs.

for the computed value of $q_r(\lambda)$

$$\begin{aligned}\widetilde{q}_r(\lambda) &= \left[(a_r-\lambda)(1+\varepsilon_1) - \frac{b_{r-1}^2(1+\varepsilon_2)}{\widetilde{q}_{r-1}(\lambda)}(1+\varepsilon_3)\right](1+\varepsilon_4)\\ &= a_r(1+\eta_r) - \lambda(1+\eta_r) - \frac{b_{r-1}^2(1+\delta_{r-1})}{\widetilde{q}_{r-1}(\lambda)}, \end{aligned} \tag{6.4}$$

where $\eta_r = (1+\varepsilon_1)(1+\varepsilon_4) - 1$ and $\delta_{r-1} = (1+\varepsilon_2)(1+\varepsilon_3)(1+\varepsilon_4) - 1$, so that we get

$$\begin{cases} |\eta_r| \leq 2.02\epsilon, \\ |\delta_{r-1}| \leq 3.03\epsilon. \end{cases} \tag{6.5}$$

Now, we may get $\widetilde{q}_{r-1}(\lambda) = 0$. If the arithmetic can handle the division by zero, as IEEE arithmetic does, then it gives $\widetilde{q}_r(\lambda) = -\infty$, independently of the value of $b_{r-1} \neq 0$, and we can write, in this case,

$$\eta_r = \delta_{r-1} = 0, \tag{6.6}$$

which, of course, satisfy the bounds (6.3). Furthermore, with $\widetilde{q}_r(\lambda) = -\infty$ in (6.1), we get that

$$\widetilde{q}_{r+1}(\lambda) = (a_{r+1}-\lambda)(1+\eta_{r+1})$$

does not depend upon the value of b_r and we can write

$$|\eta_{r+1}| \leq \epsilon, \ \delta_r = 0. \tag{6.7}$$

In case the arithmetic in use does not handle the division by zero, we may replace $\widetilde{q}_{r-1}(\lambda) = 0$ with $\widetilde{q}_{r-1}(\lambda) = a_{r-1}\epsilon$ since this corresponds to perturbing a_{r-1} to $a_{r-1}(1+\epsilon)$, in (6.1), for $k = r-1$. ∎

So, for a given λ, the computed $\widetilde{q}_k(\lambda)$ do correspond to a matrix

$$\widetilde{T} = \widehat{T} + D, \tag{6.8}$$

where $\widehat{T}$ differs from T by small relative perturbations in its (diagonal and off-diagonal) entries and D is a diagonal matrix with entries of size bounded by $2.02\epsilon|\lambda|$. Therefore, if T defines well its eigenvalues so that for $\lambda_k \neq 0$ and some small constant γ, we may write, denoting by $\widehat{\lambda}_k$ the eigenvalues of $\widehat{T}$,

$$\left|\lambda_k - \widehat{\lambda}_k\right| \leq \gamma\epsilon|\lambda_k|,$$

we get, denoting by $\widetilde{\lambda}_k$ the eigenvalues of $\widetilde{T}$ and taking into account that $\|D\| \leq 2.02\epsilon|\lambda|$,

$$\left|\lambda_k - \widetilde{\lambda}_k\right| \leq \gamma\epsilon|\lambda_k| + 2.02\epsilon|\lambda|$$

or

$$\left|\lambda_k - \widetilde{\lambda}_k\right| \leq \left(\gamma + 2.02\frac{|\lambda|}{|\lambda_k|}\right)\epsilon|\lambda_k|, \tag{6.9}$$

which shows that the relative error in $\widetilde{\lambda}_k$ is small whenever the ratio $\frac{|\lambda|}{|\lambda_k|}$ is not large. In practice, the bisection method, based upon the inertia of $T - \lambda I$, is used until we have a good approximation for the target eigenvalue; therefore, if one starts using the numerical values $\widetilde{q}_k(\lambda)$ only when a few significant digits are correct, we have that $\frac{|\lambda|}{|\lambda_k|} \approx 1$ and, in this case, the bound in (6.9) guarantees small relative errors. Note that from the point of view of convergence speed, it is premature to switch from bisection to a method with a better asymptotic rate of convergence before we have an approximation with a few correct digits anyway. So we claim that the numerical values of the pivots may be used to compute the eigenvalues with high relative accuracy whenever T defines them well.

7. Toward a mixed precision bisection algorithm. Another practical application that we envisage for our results is a mixed precision bisection algorithm. Processors are arriving on the market that are much faster for single precision floating point operations than for double precision arithmetic. Examples include the Intel Pentium IV and M processors, AMD's Opteron architectures, and the IBM Cell Broad Engine processor. When working in single precision, floating point operations can be performed up to two times faster on the Pentium and up to 10 times faster on the Cell than for double precision [31]. This technological change is likely to have a significant impact in the design of many numerical algorithms. Some work has already been carried out in the context of iterative refinement for linear systems (see [2], [30], [31]).

In an implementation of the bisection method, tailored for such processors, single precision arithmetic may be used to deliver intervals that are refined using double precision arithmetic. Because each interval produced in single precision is not guaranteed to contain the desired eigenvalue (unless some form of interval arithmetic is implemented), it cannot be accepted blindly and may need to be corrected in double precision.

Now, a critical issue is to decide when to switch from single to double precision. If we switch too soon, we will be using expensive double precision arithmetic that could have been carried out in the single format; on the other hand, if we go too far in single precision, an incorrect interval will be produced and we pay a penalty for correcting the interval. It is for this reason that a good stopping criterion for the single precision phase is much more important than a stopping criterion in the usual situation where double precision is used from the very beginning.

For a matrix T with diagonal elements a_j of size much smaller than $\|T\|$, we may, taking the relation (3.3) into account, switch from single to double precision immediately after locating an eigenvalue in the interval $[y, z]$ such that

$$z - y \leq O(\epsilon_s) \max |a_j|, \tag{7.1}$$

where ϵ_s denotes the single precision roundoff error unit. More generally, for a matrix with entries of different magnitudes, we may use the procedure described in section 5 to compute the largest size M of the entries that cannot be cleaned and replace $\max |a_j|$ with M in (7.1). For *sdd* matrices, this does not provide a good stopping criteria; therefore, a different test would be required in conjunction with the one proposed here.

8. Conclusions and further work. We have combined well-known results of the perturbation theory to derive new error bounds for the eigenvalues of symmetric tridiagonal matrices. Our bounds are sharper than the usual bounds in the case of certain matrices with entries and eigenvalues of varying size. As an application of this

idea, we have shown that a symmetric tridiagonal matrix T, with diagonal entries a_j, defines well the eigenvalues whose magnitude is not much smaller than $\max |a_j|$. This can be understood as a generalization of the well-known fact that a Golub–Kahan matrix defines well all its eigenvalues. As a practical application of our perturbation technique, we have shown that the numerical values and not only the signs of the pivots, computed in the usual way, may be used to find, with high relative accuracy, those eigenvalues which are well defined. Also, we have briefly considered a mixed precision bisection algorithm and have shown that our perturbation technique may help in the critical issue of determining when to switch from single to double precision. We are currently working in this line of research.

Acknowledgments. The author is indebted to Beresford Parlett, Zlatko Drmač, and the anonymous referees for several criticisms and suggestions that improved the paper significantly. The author also expresses his gratitude to Froilán Dopico for fruitful discussions and acknowledges that he proved Theorem 3.1 independently.

REFERENCES

[1] J. Barlow and J. Demmel, *Computing accurate eigensystems of scaled diagonally dominant matrices*, SIAM J. Numer. Anal., 27 (1990), pp. 762–791.
[2] A. Buttari, J. Dongarra, J. Kurzak, P. Luszcez, and S. Tomov, *Using Mixed Precision for Sparse Matrix Computations to Enhance the Performance while Achieving* 64*-Bit Accuracy*, LAPACK Working Note 180, 2006.
[3] P. Deift, J. Demmel, L.-C. Li, and C. Tomei, *The bidiagonal singular value decomposition and Hamiltonian mechanics*, SIAM J. Numer. Anal., 28 (1991), pp. 1463–1516.
[4] J. Demmel and W. Kahan, *Accurate singular values of bidiagonal matrices*, SIAM J. Sci. Statist. Comput., 11 (1990), pp. 873–912.
[5] J. Demmel and K. Veselić, *Jacobi's method is more accurate than QR*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 1204–1245.
[6] J. Demmel and W. Gragg, *On computing accurate singular values and eigenvalues of acyclic matrices*, Linear Algebra Appl., 185 (1993), pp. 203–218.
[7] J. Demmel, *The Inherent Inaccuracy of Implicit Tridiagonal QR*, LAPACK Working Note 45, 1992.
[8] J. Demmel, I. Dhillon, and H. Ren, *On the correctness of some bisection-like parallel eigenvalue algorithms in floating point arithmetic*, Electronic Trans. Numer. Anal., 3 (1995), pp. 116–149.
[9] J. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
[10] J. Demmel, *Accurate SVDs of structured matrices*, SIAM J. Matrix Anal. Appl., 21 (1999), pp. 562–508.
[11] J. Demmel, M. Gu, S. Eisenstat, I. Slapničar, K. Veselić, and Z. Drmač, *Computing the singular value decomposition with high relative accuracy*, Linear Algebra Appl., 299 (1999), pp. 21–80.
[12] J. Demmel and P. Koev, *Accurate SVDs of weakly diagonally dominant M-matrices*, Numer. Math., 98 (2004), pp. 99–104.
[13] I. Dhillon and B. Parlett, *Multiple representations to compute orthogonal eigenvectors of symmetric matrices*, Linear Algebra Appl., 387 (2004), pp. 1–28.
[14] I. Dhillon and B. Parlett, *Orthogonal eigenvectors and relative gaps*, SIAM J. Matrix Anal. Appl., 25 (2004), pp. 858–899.
[15] F. Dopico and P. Koev, *Accurate symmetric rank revealing and eigendecompositions of symmetric structured matrices*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 1126–1156.
[16] S. Eisenstat and I. Ipsen, *Relative perturbation techniques for singular value problems*, SIAM J. Numer. Anal., 32 (1995), pp. 1972–88.
[17] S. Eisenstat and I. Ipsen, *Absolute Perturbation Bounds for Matrix Eigenvalues Imply Relative Bounds*, Technical report CRSC-TR97-16, Center for Research in Scientific Computation, Department of Mathematics, North Carolina State University, Raleigh, NC, 1997.
[18] V. Fernando and B. Parlett, *Accurate singular values and differential qd algorithms*, Numer. Math., 67 (1994), pp. 191–229.
[19] B. Grosser and B. Lang, *An $O(n^2)$ algorithm for the bidiagonal SVD*, Linear Algebra Appl., 358 (2003), pp. 45–70.

[20] B. Grosser and B. Lang, *On symmetric eigenproblems induced by the bidiagonal SVD*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 599–620.
[21] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
[22] I. Ipsen, *Relative perturbation bounds for matrix eigenvalues and singular values*, in Acta Numerica, 1998, Acta Numer. 7, Cambridge University Press, Cambridge, UK, 1998, pp. 151–201.
[23] I. Ipsen, *A note on unifying absolute and relative perturbation bounds*, Linear Algebra Appl., 358 (2003), pp. 239–53.
[24] I. C. F. Ipsen and B. Nadler, *Refined perturbation bounds for eigenvalues of Hermitian and non-Hermitian matrices*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 40–53.
[25] W. Kahan, *Accurate Eigenvalues of a Symmetric Tridiagonal Matrix*, Technical Report CS41, Computer Science Department, Stanford University, Palo Alto, CA, 1966.
[26] P. Koev, *Accurate and Efficient Computations with Structured Matrices*, Ph.D. thesis, University of California, Berkeley, CA, 2002.
[27] P. Koev, *Accurate eigenvalues and SVDs of totally nonnegative matrices*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 1–23.
[28] P. Koev and F. Dopico, *Accurate eigenvalues of certain sign regular matrices*, Linear Algebra Appl., 424 (2007), pp. 435–447.
[29] P. Koev, *Accurate computations with totally nonnegative matrices*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 731–751.
[30] J. Kurkzak and J. Dongarra, *Implementation of a Mixed-Precision High Performance LINPACK Benchmark on the Cell Processor*, LAPACK Working Note 177, 2006.
[31] J. Langou, J. Langou, P. Luszczek, J. Kurzak, A. Buttari, and J. Dongarra, *Exploiting the Performance of* 32 *Bit Floating Point Arithmetic in Obtaining* 64 *Bit Accuracy*, LAPACK Working Note 175, 2006.
[32] R. Li, *Relative Perturbation Theory:* (I) *Eigenvalue Variations*, LAPACK Working Note 84, 1994.
[33] C. Li and R. Mathias, *On the Lidskii-Mirsky-Wielandt Theorem*, Technical report, Department of Mathematics, College of William and Mary, Williamsburg, VA, 1997.
[34] R. Mathias, *Spectral Perturbation Bounds for Positive Definite Matrices*, Technical report, Department of Mathematics, College of William and Mary, Williamsburg, VA, 1994.
[35] R. Mathias, *Accurate eigensystem computations by Jacobi methods*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 977–1003.
[36] B. Parlett, *The Symmetric Eigenvalue Problem*, Prentice–Hall, New York, 1980.
[37] B. Parlett and B. Nour-Omid, *The use of a refined error bound when updating eigenvalues of tridiagonals*, Linear Algebra Appl., 68 (1985), pp. 179–219.
[38] B. Parlett, *The new qd algorithms*, in Acta Numerica, 1995, Acta Numer. 4, Cambridge University Press, Cambridge, UK, 1995, pp. 459–491.
[39] B. Parlett and I. Dhillon, *Relatively robust representations of symmetric tridiagonals*, Linear Algebra Appl., 309 (2000), pp. 121–151.
[40] B. Parlett and O. Marques, *An implementation of the dqds algorithm (positive case)*, Linear Algebra Appl., 309 (2000), pp. 217–259.
[41] S. Singer and S. Singer, *Skew-symmetric differential qd algorithm*, Appl. Numer. Anal. Comput. Math., 2 (2005), pp. 134–151.
[42] K. Veselić and I. Slapničar, *Floating-point perturbations of Hermitian matrices*, Linear Algebra Appl., 195, pp. 81–116.
[43] P. Willems, B. Lang, and C. Vomel, *Computing the Bidiagonal SVD Using Multiple Relatively Robust Representations*, LAPACK Working Note 166, 2005.
[44] J. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, London, 1965.