Accurate computation of singular values and eigenvalues of symmetric matrices*

Ivan Slapničar†

Abstract. *We give the review of recent results in relative perturbation theory for eigenvalue and singular value problems and highly accurate algorithms which compute eigenvalues and singular values to the highest possible relative accuracy.*

Key words: *symmetric eigenvalue problem, singular value problem, perturbation theory, relative perturbation theory, relative accuracy*

Sažetak. **Točno računanje singularnih i svojstvenih vrijednosti simetričnih matrica**. *Dan je pregled recentnih rezultata o relativnoj teoriji smetnje za probleme svojstvenih i singularnih vrijednosti, te algoritama visoke točnosti koji računaju svojstvene i singularne vrijednosti s najvećom mogućom relativnom točnošću.*

Ključne riječi: *simetrični problem svojstvenih vrijednosti, problem singularnih vrijednosti, teorija smetnje, teorija relativnih smetnji, relativna točnost*

1. Introduction

The eigenvalue problem [**?**, **?**, **?**] reads

$$Hx = \lambda x.$$

The scalar λ is the eigenvalue, and the vector x is the corresponding eigenvector of the matrix H. If H is a symmetric or a Hermitian matrix of order n, then

*The lecture presented at the Mathematical Colloquium in Osijek organized by Croatian Mathematical Society – Division Osijek, May 17, 1996.

†University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, R. Boškovića b.b., HR-21000 Split, Croatia, e-mail: ivan.slapnicar@fesb.hr, URL: http://adria.fesb.hr/∼slap

H has exactly n real eigenvalues, and the corresponding eigenvectors span the bases of the n-dimensional space. More precisely, for a symmetric matrix H we have

$$Q^T H Q = \Lambda,$$

where $\Lambda = \operatorname{diag}(\lambda_i)$ is a diagonal matrix with eigenvalues of H on the diagonal, and Q is an orthogonal matrix whose columns are the corresponding eigenvectors. For Hermitian matrix H we have $Q^* H Q = \Lambda$, where the eigenvector matrix Q is unitary. Similarly, *the singular value problem* for the general matrix $G \in \mathbb{C}^{m\times n}$ reads [**?**]

$$U^* G V = \Sigma,$$

where $\Sigma = \operatorname{diag}(\sigma_i)$, $\Sigma \in R^{m\times n}$, $\sigma_i \geq 0$, and the matrices U and V are unitary. The columns of the matrix U are the left singular vectors, and the columns of the matrix V are the right singular vectors. If, for example, $m \geq n$, then obviously

$$\begin{aligned} V^* G^* G V &= \operatorname{diag}(\sigma_1^2, \cdots, \sigma_n^2) \in R^{n\times n}, \\ U^* G G^* U &= \operatorname{diag}(\sigma_1^2, \cdots, \sigma_n^2, 0, \cdots, 0) \in R^{m\times m}. \end{aligned}$$

We see that the eigenvalue and singular value problems are closely related, that is, the singular values of the matrix G can be obtained as the roots of the eigenvalues of Hermitian matrices G^*G and GG^*.

Solution of many problems in technical applications is reduced to solving eigenvalue and singular value problems. Thus, these problems attract considerable attention and represent one of the most important areas of numerical linear algebra. The first method for solving the eigenvalue problem for symmetric matrices is the Jacobi method [**?**, **?**, **?**, **?**] which dates back in 1846. The Jacobi method constructs a sequence of matrices

$$H_1 = H, \qquad H_{k+1} = R_k^T H_k R_k,$$

which converges to the eigenvalue matrix Λ, while the sequence of products $R_1 R_2 \cdots R_k$ converges to the eigenvector matrix Q. Matrices R_k are the orthogonal plane rotation matrices chosen to annihilate one off-diagonal element of the matrix H_k. Due to the finite arithmetic of the computer this infinite iterative procedure stops after a finite number of steps.

In 1950-ties and 60-ties the QR methods [**?**, **?**, **?**] were developed by many authors. These methods first reduce the symmetric matrix H to the tridiagonal matrix T by using orthogonal similarity transformations, and then use QR iterations to solve the eigenvalue problem for the matrix T. Although both methods require $O(n^3)$ floating-point operations, the QR methods are on sequential (single processor) computers about five times faster than Jacobi type methods. Other methods for solving the eigenvalue problem [**?**, **?**, **?**] are LR methods, iterative methods like the power method, the inverse iterations, the method

of Krylov subspaces, the Lanczos method, the subspace iteration method, and the bisection method and the divide-and-conquer method which are particularly efficient for tridiagonal matrices.

Iterative methods are especially suitable for large matrices, sparse matrices, and when only some of the values or vectors are needed. The bisection locates an eigenvalue by using the Sturm sequence, and the corresponding eigenvectors can be computed by inverse iteration. The divide-and-conquer method is very suitable for multi-processor computers. The method first partitions the staring matrix into blocks, then solves smaller eigenvalue problems, and finally connects all solutions. We conclude that the choice of the method depends upon the structure and the size of the matrix, on the requirements for speed and accuracy, whether all or just some values/vectors are required, and the available hardware.

Using the computers in solving eigenvalue and singular value problems has lead to two aspects of research: *speed* and *accuracy*. Due to need to solve larger and larger problems, the first aspect of research is finding faster algorithms and the analysis of their speed of convergence. This subject is beyond the scope of this review. Computers use a discrete subset of rational (real) numbers and every real number is represented by the closest approximation in that subset [**?**, **?**, **?**, **?**]. Numbers are usually represented with 8 (single precision) or 16 (double precision) significant digits. The question of accuracy can be stated very simply: *how many accurate digits does the computed eigenvalue have?* In applications four kinds of errors appear: errors of the model, since the chosen mathematical model may not completely describe the actual real world system; errors in data, since the data are most often acquired by measurements which are not absolutely accurate; errors in storing the matrix into the computer due to previously mentioned approximations; and the errors generated by the computational method. Here we shall deal with the two latter sources of errors, although by using the perturbation theory, which we describe later, one may try to estimate the effect of the first two sources of error on the final solution. When storing the matrix H into the computer, instead of the element H_{ij} we store the element

$$H_{ij} + \delta H_{ij}, \qquad |\delta H_{ij}| \leq \epsilon |H_{ij}|,$$

where ϵ is the machine precision, $\epsilon \approx 10^{-8}$ or $\epsilon \approx 10^{-16}$. Therefore, the last stored digit does not need to be correct and instead of H we store some $H + \delta H$. *The condition* is defined as the number κ which tells us how many times does the error in original data increase. If λ_i is the i-th largest eigenvalue of the matrix H, and $\lambda_i + \delta\lambda_i$ is the i-th largest computed eigenvalue, then the answer to the question about accuracy generally has the form

$$|\delta\lambda_i| \leq \kappa \|\delta H\| |\lambda_i|.$$

Here $\|\cdot\|$ represents some matrix norm or some other way of measuring the size of the perturbation which does not necessarily has to have all properties of the norm. The condition κ depends on the matrix, but also on the computational

method which we use. From this exposition it follows that we shall obtain the answer to the question of how many accurate digits does the computed value have when we answer the following two questions:

(A) *Is the matrix "well behaved", that is, do small relative changes in matrix elements cause small relative changes in eigen/singular values?*

(B) *If the matrix is well-behaved which algorithm computes eigen/singular values with this accuracy?*

In general, to answer the question (A) an appropriate perturbation theory for the given type of problem needs to be developed, while the answer to (B) is given by the numerical analysis of the algorithm. Many authors have noticed that for some problems different methods give answers with widely varying accuracy. For example, in 1968 Rosanoff et al. [**?**] performed experimental analysis of many structural models and noticed that the Jacobi method often computed tiny eigenvalues much more accurate than the QR method. The authors had many excellent observations and gave interesting explanations for facts which were much later established with complete mathematical rigor. We also need to mention the important paper by Kahan [**?**]. In 1980-ties many articles appear and the intensive research is still going on. First Demmel and Kahan [**?**] analyzed the singular value problem for bidiagonal matrices. Then the symmetric (Hermitian) eigenvalue problems were analyzed by Barlow and Demmel [**?**] for scaled diagonally dominant matrices, by Demmel and Veselić [**?**] for positive definite matrices, as well as the SVD, and by Veselić and Slapničar [**?**, **?**, **?**] for indefinite matrices. These works are followed by many others which we will describe in the final section.

2. Symmetric eigenvalue problem

The classical answer to the question of how big the relative changes in eigenvalues of the non-singular symmetric matrix H are when its elements are relatively perturbed, $|\delta H_{ij}| \leq \epsilon |H_{ij}|$, follows from Weyl's theorem [**?**, **?**, **?**],

$$|\delta\lambda_i| \leq \|\delta H\|_2.$$

From this it follows

$$|\delta\lambda_i| \leq \epsilon \||H|\|_2 \frac{|\lambda_i|}{|\lambda_i|} \leq \epsilon\sqrt{n}\|H\|_2\||H^{-1}\|_2|\lambda_i| \equiv \epsilon\sqrt{n}\kappa_2(H)|\lambda_i|.$$

Here $\kappa_2(H) \equiv \|H\|_2\|H^{-1}\|_2$ denotes the spectral condition of the matrix H. We see that tiny eigenvalues, which are the most important in many applications, are the most sensitive. This bound is almost attainable, but it is also in many

cases inappropriate. The two extreme cases are illustrated by the following examples:

$$H = \begin{pmatrix} 1 & 1 \\ 1 & 1 + 10^{-10} \end{pmatrix}, \qquad H + \delta H = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

where

$$\lambda_{min} \approx 0.5 \cdot 10^{-10}, \quad \lambda_{min} + \delta\lambda_{min} = 0, \quad \frac{|\delta\lambda_{min}|}{|\lambda_{min}|} = 1, \quad \kappa_2(H) \approx 4 \cdot 10^{10},$$

and

$$H = \begin{pmatrix} 10^{10} & 0 \\ 0 & -1 \end{pmatrix}, \qquad H + \delta H = \begin{pmatrix} 10^{10}(1 + 10^{-10}) & 0 \\ 0 & -1 + 10^{-10} \end{pmatrix},$$

where

$$\frac{|\delta\lambda_{min}|}{|\lambda_{min}|} = 10^{-10}, \qquad \kappa_2(H) = 10^{10}.$$

Both, QR and Jacobi method always compute the eigenvalues at least as accurately as predicted by the above perturbation bound [**?**, **?**, **?**]. However, as we have just seen there exist matrices where the above bounds are inappropriate, and there also exist matrices where not all methods attain the same accuracy.

Barlow and Demmel [**?**] analyzed symmetric scaled diagonally dominant matrices of the form

$$H = D(J + N)D,$$

where D is a diagonal positive definite matrix, J is a diagonal matrix, $J_{ii} \in \{-1, 1\}$, and $\|N\|_2 < 1$. They showed that symmetric relative perturbations $|\delta H_{ij}| \leq \epsilon |H_{ij}|$ imply

$$|\delta\lambda_i| \leq \frac{n^2\epsilon}{1 - \|N\|_2} |\lambda_i|.$$

It is important to notice that this bound depends upon $\|N\|_2$ *independently* of the condition $\kappa_2(H)$. The authors also showed that a version of bisection on a full matrix, a time consuming algorithm which needs $O(n^4)$ floating-point operations, computes the eigenvalues with this accuracy, and that inverse iteration produces highly accurate eigenvectors in a norm-wise sense.

Demmel and Veselić [**?**] analyzed symmetric positive definite matrices. Let the scaled matrix A be defined as $H = DAD$, where D is a diagonal positive definite matrix such that $A_{ii} = 1$. Then relative perturbations of matrix elements imply

$$|\delta\lambda_i| \leq \frac{n\epsilon}{\lambda_{min}(A)} \leq n\epsilon\kappa_2(A)\lambda_i.$$

This bound is, of course, meaningful only if the quotient on the right-hand side is less than one. According to the result by van der Sluis [**?**]

$$\kappa_2(A) \leq n \min_D \kappa_2(DHD),$$

where the minimum is taken over all diagonal positive definite matrices. Thus, $\kappa_2(A) \leq n\kappa_2(H)$, so the above bound is never much worse than the classical bound. On the other hand, it is possible that $\kappa_2(A) \ll \kappa_2(H)$ (the trivial example is when H is diagonal) in which case the above bound is much better than the classical one. The authors showed that the Jacobi method computes the eigenvalues with this accuracy. The QR method can fail both in the tridiagonalization phase as well as during QR iterations. Another algorithm which is also as accurate as predicted by perturbation bound consists of two steps:

1. *H is factorized by the Cholesky factorization [?] $H = LL^T$, where L is a lower triangular matrix, or by the Cholesky factorization with diagonal pivoting $H = PLL^TP^T$, where L is a lower triangular and P is a permutation matrix.*

2. *One-sided implicit Jacobi method is applied from the right to the factor L or PL.*

Let us explain the second step of this algorithm. First note that H and L^TL have identical eigenvalues and closely related eigenvectors: if $U^TL^TLU = \Lambda$ is the eigenvalue decomposition of the matrix L^TL, then the orthogonal matrix $Q = LU\Lambda^{-1/2}$ is the eigenvector matrix of H, that is, $Q^THQ = \Lambda$. Applying the Jacobi method to the matrix L^TL would yield the sequence

$$L_{k+1}^TL_{k+1} = R_k^TL_k^TL_kR_k.$$

In the implicit method we multiply by rotation matrices just the factor from the right, which creates the sequence

$$L_1 = L, \qquad L_{k+1} = L_kR_k.$$

In order to do this, in each step three elements of the implicitly defined matrix $L_k^TL_k$ which are needed to compute the rotation matrix R_k need to be computed. Only one scalar product suffices since the diagonal of the sequence $L_k^TL_k$ can be updated in a separate vector. (for details see [?]). From this we see that the convergence of the implicitly defined sequence $L_k^TL_k$ to the eigenvalue matrix Λ is equivalent to the convergence of the sequence L_k to the matrix $Q\Lambda^{1/2}$. Therefore, when the infinite iterative process stops on some matrix L_M due to the finite precision of the computer, then the squares of the norms of the columns of L_M are the computed eigenvalues of the matrix H and the normalized columns of L_M are the computed corresponding eigenvectors.

Veselić and Slapničar [?, ?] generalized the above results to indefinite matrices. Their main result is the relative perturbation bound for the generalized eigenvalue problem

$$Hx = \lambda Kx,$$

where H and K are Hermitian matrices and K is positive definite. Application of this result to a single indefinite Hermitian matrix H gives the following bound:

let $H = Q\Lambda Q^*$ be the eigenvalue decomposition of H. The spectral absolute value of H is defined as

$$H^\dagger = Q|\Lambda|Q^T = \sqrt{H^2}.$$

Let the matrix A be defined as $H^\dagger = DAD$, where D is a diagonal positive definite matrix such that $A_{ii} = 1$. Then ϵ-relative changes of the matrix elements imply

$$|\delta\lambda_i| \leq n\epsilon\kappa_2(A)|\lambda_i|.$$

Note that for a positive definite matrix $H^\dagger = H$ in which case this result reproduces the bound by Demmel and Veselić. It was also shown that this bound reproduces the bound by Barlow and Demmel for scaled diagonally dominant matrices.

Veselić [**?**] proposed the following two-step algorithm in analogy to the positive definite case:

1. *H is factorized by the symmetric indefinite factorization with complete pivoting $H = PGJG^TP^T$, where G is a lower block-triangular matrix with 1×1 and 2×2 diagonal blocks and has a full column rank (this algorithm works if H is singular, as well, but in this case the above perturbation bound does not hold), P is a permutation matrix, and J is a diagonal matrix, $J_{ii} \in \{-1, 1\}$.*

2. *One-sided implicit J-orthogonal Jacobi method is applied to the pair (G, J) from the right.*

The above factorization is a modification of the well-known method by Bunch and Parlett [**?**]. If H is a positive definite matrix, then this factorization becomes the Cholesky factorization with diagonal pivoting. Error analysis of the factorization is given by Slapničar [**?**, **?**]. It was shown that the factors G and J computed in the floating-point arithmetic with machine precision ϵ are the exact factors of the perturbed matrix $H + \delta H$, where

$$|\delta H| \leq O(n)\epsilon(|H| + P|G||G^T|P^T).$$

Here $|H|$ is defined by $|H|_{ij} = |H_{ij}|$. Note that the same type of bound holds for the Cholesky factorization with or without pivoting as well as for LU factorization [**?**, **?**]. The difference between the implicit Jacobi method and the implicit J-orthogonal Jacobi method is that the J-orthogonal method uses J-orthogonal plane rotations for which $R_k^T J R_k = J$. If J_{ii} and J_{jj} have the same sign, where (i, j) is the pivot pair in the k-th step, then the matrix R_k performs the trigonometric (orthogonal) plane rotation as in the ordinary Jacobi method. If J_{ii} and J_{jj} have different signs, then a hyperbolic rotation is performed. Implicit J-orthogonal Jacobi method forms a sequence of matrices

$$G_1 = G, \qquad G_{k+1} = G_k R_k,$$

which converges to the matrix $Q|\Lambda|^{1/2}$. Thus, G_M is the last matrix on which the algorithm has stopped, then $\|G_{M,:i}\|^2 J_{ii}$ is the i-th computed eigenvalue and the normalized i-th column is the corresponding eigenvector. The convergence of the method was proved by Veselić [**?**], and the quadratical convergence was proved by Drmač and Hari [**?**]. Slapničar [**?**] proved that this algorithm computes the eigenvalues as accurately as predicted by the perturbation theory. On the other hand, the QR method and the standard two-sided Jacobi method sometimes do not achieve this accuracy. We shall illustrate this on the following example: let

$$H = \begin{pmatrix} 1600 & -300 & 14 & 300000 \\ -300 & 43.5 & -4.75 & -423212 \\ 14 & -4.75 & 0.1875 & 19800 \\ 300000 & -423212 & 19800 & 3207938 \cdot 10^3 \end{pmatrix}.$$

All elements are the sums of the powers of 2 and are exactly stored in the IEEE single precision [**?**], $\epsilon \approx 10^{-8}$. Since

$$\kappa_2(A) \approx 18, \qquad \kappa_2(H) \approx 10^{10},$$

we expect 6-7 accurate digits from the above algorithm in single precision. The eigenvalues of H are

$$\begin{array}{lcllcl} \lambda_1 & = & -54.043364 & \lambda_2 & = & -0.0283096849 \\ \lambda_3 & = & 1613.74866 & \lambda_4 & = & 3207938084.0105 \end{array}$$

Here the digits common to the above algorithm and the LAPACK QR routine `dsyev.f` [**?**] computed IEEE double precision , $\epsilon \approx 10^{-16}$, are shown. Our algorithm, LAPACK QR algorithm `ssyev.f` and the two-sided Jacobi method computed the following eigenvalues in single precision:

	Ouralg.	`ssyev.f`	Jacobi
λ_1	-54.043369	-55.990593	-54.043369
λ_2	-0.02830968	-0.0326757	-0.02830995
λ_3	1613.7487	1651.6652	1613.7486
λ_4	3207938000	3207938000	3207938000

We see that our algorithm behaves as predicted, the QR method has completely missed the tiniest eigenvalue and two more are insufficiently accurate, while the Jacobi method is somewhat less accurate than the our algorithm.

Let us now discuss perturbations of the eigenvectors. Let λ_i and $\lambda_i + \delta\lambda_i$ be simple eigenvalues in the same order and let x_i and $x_i + \delta x_i$ be the corresponding eigenvectors, respectively. Then the classical perturbation theory applied to relative changes of matrix elements gives [**?**]

$$\|\delta x_i\|_2 \leq \frac{n\epsilon\|H\|_2}{\min_{j\neq i}|\lambda_i - \lambda_j|} + O(\epsilon^2).$$

Therefore, the perturbation of the eigenvectors is proportional to the norm of the perturbation and inversely proportional to the distance between the eigenvalues and the rest of the spectrum. For scaled diagonally dominant matrices Barlow and Demmel [**?**] proved the following result:

$$\|\delta x_i\|_2 \leq \frac{n\epsilon}{(1-\|N\|_2)\mathrm{relgap}(\lambda_\mathrm{i})} + O(\epsilon^2), \quad \mathrm{relgap}(\lambda_\mathrm{i}) = \min_{\mathrm{j}\neq\mathrm{i}} \frac{|\lambda_\mathrm{i}-\lambda_\mathrm{j}|}{|\lambda_\mathrm{i}\lambda_\mathrm{j}|^{1/2}}.$$

Therefore, the perturbation of the eigenvector depends upon $\|N\|_2$, the size of the *relative* perturbations of matrix elements and the *relative distance* between the eigenvalue and the rest of the spectrum. As in the case with eigenvalues, this bound is in some cases much better than the classical bound. Demmel and Veselić [**?**] proved the similar result for positive definite matrices:

$$\|\delta x_i\|_2 \leq \frac{\sqrt{n}\epsilon}{\lambda_{min}(A)\mathrm{relgap}(\lambda_\mathrm{i})} + O(\epsilon^2) \leq \frac{\sqrt{n}\epsilon\kappa_2(A)}{\mathrm{relgap}(\lambda_\mathrm{i})} + O(\epsilon^2).$$

Veselić and Slapničar [**?**, **?**, **?**] generalized these results to indefinite matrices. Instead of analyzing perturbations of eigenvectors they stated their results in terms of perturbations of the eigenprojection P_i onto the invariant subspace which corresponds to the eigenvalue λ_i, thus making it possible to deal with multiple eigenvalues. Let $P_i + \delta P_i$ be the spectral projection to the invariant subspace corresponding to those eigenvalues of the matrix $H + \delta H$ which correspond to λ_i. Let $\eta = n\epsilon\kappa_2(A)$. Then

$$\|\delta P_i\|_2 \leq \frac{\eta}{\mathrm{relgap}(\lambda_\mathrm{i})} \cdot \frac{1}{1-\dfrac{\eta}{\mathrm{relgap}(\lambda_\mathrm{i})}}, \qquad \mathrm{relgap}(\lambda_\mathrm{i}) = \min_{\lambda_\mathrm{j}\neq\lambda_\mathrm{i}} \frac{|\sqrt{|\lambda_\mathrm{i}|}-\sqrt{|\lambda_\mathrm{j}|}|}{\max\{\sqrt{|\lambda_\mathrm{i}|},\sqrt{|\lambda_\mathrm{j}|}\}}$$

if the right-hand side is positive. The described highly accurate algorithms compute the eigenvectors and eigenprojections according to these bounds.

3. Singular value problem

Demmel and Kahan [**?**] showed that bidiagonal matrices determine well their singular values in the sense that relative changes of matrix elements cause relative changes of the same order in singular values independent of the magnitude of the matrix elements. They showed that the QR algorithm for bidiagonal matrices with zero-shift computes the singular values with almost complete relative accuracy. Since use of the zero-shift can result in slow convergence, a hybrid algorithm was suggested. As long as there is no danger of generating large relative errors this algorithm uses the standard QR method with shifts, and when the danger of making unrecoverable errors appears, then the algorithm switches to zero-shift. This algorithm is a part of the numerical linear algebra library LAPACK [**?**] as a subroutine `dbdsqr.f`. Another algorithm which was shown to attain almost complete accuracy is the bisection.

By using this result and the accuracy of the Cholesky factorization, Barlow and Demmel [**?**] proposed the following algorithm for highly accurate solution of the tridiagonal positive definite eigenvalue problem: *(1) the matrix is factorized as $H = LL^T$ by Cholesky factorization; (2) the eigenvalue problem is solved by solving the bidiagonal singular value problem for the factor L by the highly accurate QR method for a bidiagonal matrix.* This class of matrices is very important since it arises in many engineering applications.

Demmel and Veselić [**?**] analyzed the singular value problem for the general matrix G of full (column) rank. The results are similar to those for the positive definite eigenvalue problem. Let $G = BD$, where D is a diagonal positive definite matrix such that the columns of B have a unit norm. Then relative perturbations $|\delta G_{ij}| \leq \epsilon |G_{ij}|$ imply relative changes in singular values

$$|\delta\sigma_i| \leq n\epsilon\kappa_2(B)\sigma_i.$$

It was also shown that the already described implicit Jacobi method applied to the matrix G from the right computes the singular values with this accuracy. On the other hand, the QR method which first reduces G to a bidiagonal matrix by using orthogonal transformations and then solves the bidiagonal singular value problem often does not attain the required accuracy. This is due to the fact that the bidiagonal reduction can cause large errors. The relative bounds for the perturbations of the singular vectors u_i and v_i which correspond to a simple singular value are similar to the bounds in the positive definite case: norm of the perturbation is proportional to the size of relative perturbations of matrix elements and the condition of the matrix B, and inversely proportional to the relative distance between the singular value and the rest of the spectrum:

$$\|\delta u_i\|_2, \|\delta v_i\|_2 \leq \frac{\sqrt{n}\epsilon}{\sigma_{min}(B)\mathrm{relgap}(\sigma_\mathrm{i})} + O(\epsilon^2) \leq \frac{\sqrt{n}\epsilon\kappa_2(B)}{\mathrm{relgap}(\sigma_\mathrm{i})} + O(\epsilon^2),$$

where $\mathrm{relgap}(\sigma_\mathrm{i}) = \min_{\mathrm{j}\neq\mathrm{i}} |\sigma_\mathrm{i} - \sigma_\mathrm{j}|/(\sigma_\mathrm{i} + \sigma_\mathrm{j})$.

Veselić and Slapničar [**?**, **?**] showed that the above bound for the perturbation of singular values also holds for the hyperbolic singular value problem for the pair (G, J),

$$U^*GV = \Sigma,$$

where G is a complex matrix with full column rank, U is a unitary matrix, V is a J-unitary matrix, $V^*JV = J$, Σ is a diagonal matrix with positive diagonal elements, and J is a diagonal matrix, $J_{ii} \in \{-1, 1\}$. Slapničar [**?**] proved that for the real matrix G the implicit J-orthogonal Jacobi method described in previous section computes hyperbolic singular values with this accuracy. Slapničar and Veselić [**?**, **?**] also derived relative bounds for norm of perturbations of the orthogonal projections to left invariant subspaces corresponding to possibly multiple singular values. It is interesting to note that J-orthogonal transformations are as stable as the orthogonal transformations. This appears to be

contrary to the established opinion that hyperbolic transformations need to be avoided since the condition of the matrix R_k can be large. In [**?**] it is shown that $\kappa_2(V) \leq \kappa_2(B)$, which implies that in the k-t step

$$\kappa_2(R_k) \leq \kappa_2((\,(B_k)_{:,i} \quad (B_k)_{:,j}\,)) \leq \kappa_2(B_k),$$

where $G_k = B_k D_K$ and D_k is a diagonal positive definite matrix such that B_k has columns of unit norm. Here $(B_k)_{:,i}$ denotes the i-th column of the matrix B_k. Partial theoretical bounds as well as overwhelming numerical evidence [**?**, **?**] show that the growth of the condition $\kappa_2(B_k)$ during Jacobi process is very small, thus we can claim that for each k

$$\kappa_2(R_k), \kappa_2(R_1 R_2 \cdots R_k) \leq c\kappa_2(B),$$

where c is a moderate constant.

The stated perturbation results and the error analysis of implicit Jacobi type methods are the essential components of the proof of the accuracy of the two-step methods. However, in practice it is also important to note that

> *in two-step algorithms the factorization with pivoting almost always results in factors with very low $\kappa_2(B)$. Thus, the main error comes from factorization while the implicit Jacobi type method contributes practically nothing to the final error.*

The multiplicative relative perturbation theory by Eisenstat and Ipsen [**?**, **?**] and later Li [**?**, **?**], where perturbations of matrix elements are given by congruences, $G + \delta G = D_1 G D_2$, are bases for the recent research on highly accurate computation of singular values by Demmel et al. [**?**]. The *rank-revealing factorization*, RRF, of the matrix G is defined as any factorization $G = XDY$, where X and Y are well-conditioned and D is diagonal. Some examples of RRF are the singular values decomposition itself and the LDU factorization (Gaussian elimination) with complete pivoting. If there exists a RRF which is accurately determined by the data then so are the eigen/singular values, and if small relative changes in matrix elements cause large relative changes in D then, eigen/singular values also undergo large relative changes. Further, if there exists a RRF which is accurate in this sense, then the eigen/singular values can be computed to high relative accuracy.

Some classes of matrices which have an accurate RRF are already described in this and the previous section. Further, such classes which are described in [**?**] are: matrices which satisfy some analytic conditions, matrices which satisfy some sparsity and sign pattern conditions, some rationally structured matrices, and some finite element matrices. The first class includes well-scaled positive definite and indefinite matrices and matrices of the form $G = BD$ which are already described, matrices of the form $G = D_1 B D_2$, where D_1 and D_2 are diagonal and non-singular and all minors of B are well conditioned, and matrices of the form $G = D_1 B D_2$, where D_1 and D_2 are diagonal and non-singular

with (nearly) decreasing diagonal elements and all leading minors of B are well conditioned. The second class includes bidiagonal and acyclic matrices (sparsity conditions), and total sign compound matrices (sparsity and sign conditions). The third class includes Cauchy matrices, and the fourth class includes matrices which come from linear mass spring systems (see also [**?**]), two-dimensional trusses, and the Sturm-Liouville problem.

Means of obtaining an accurate RRF for these classes of matrices are different. Generally speaking, one always computes some variant of Gaussian elimination with complete pivoting but the details vary. In the finite element case the factors are obtained by using the natural factor formulation.

Once an accurate RRF is obtained, its singular values can be found to high relative accuracy by several algorithms. We mention two algorithms from [**?**]. The first algorithm uses the J-orthogonal Jacobi method to compute the eigenvalues of the pair

$$\left(\begin{pmatrix} XD^{1/2} & XD^{1/2} \\ Y^T D^{1/2} & -Y^T D^{1/2} \end{pmatrix}, \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \right).$$

The positive eigenvalues of this pair are the singular values of the original matrix G. The second algorithm is based on the algorithm for product singular value decomposition by Drmač [**?**]. The algorithm computes the singular value decomposition $G \equiv XDY^T = U\Sigma V^T$ as follows:

1. *perform QR factorization with pivoting to compute $XD = QRP$, where Q is orthogonal, R is upper triangular and P is a permutation matrix,*
2. *compute a diagonal matrix D' such that $R = D'R'$ and R' is well-conditioned,*
3. *compute the singular value decomposition $D'Z = \bar{U}\Sigma V^T$ by the implicit Jacobi method,*
4. *multiply $U = Q\bar{U}$.*

The first algorithm computes the singular values with the relative error bounded by $O(\epsilon \max\{\kappa(X), \kappa(Y)\})$, and the second algorithm computes the singular values with the relative error bounded by $O(\epsilon \max\{\kappa(X), \kappa(R')\kappa(Y)\})$, where ϵ is machine precision. The norm-wise error bounds for the computed singular vectors are obtained by dividing these bounds by relative gaps, similarly as above.

4. Concluding remarks

Let us briefly state some of the other results and research concerning the relative perturbation theory and highly accurate algorithms. Demmel and Gragg [**?**] generalized the results by Demmel and Kahan to acyclic matrices, that is,

matrices whose bipartite graph possesses no cycles, and showed that the bisection computes the singular values with almost complete accuracy. Pietzsch [**?**] developed an algorithm for the skew-symmetric eigenvalues problem, and by applying the perturbation theory for symmetric matrices proved the accuracy of the algorithm. Singer [**?**] proved the relative accuracy of the Jacobi method for Hermitian matrices. Deichmöller [**?**] analyzed the implicit variant of the Falk-Langemeyer method for computing the generalized singular values. Drmač [**?**, **?**, **?**] analyzed accurate computation of singular values and various generalized singular values, proved the relative perturbation bounds by using residuals, analyzed the accuracy of the QR factorization which is used as a preprocessing step for the implicit Jacobi method in the case when the matrix G has much more rows than columns, and analyzed numerical aspects of accurate computing such as overflow/underflow and the accuracy of various implementations of plane rotations. Fernando and Parlett [**?**] showed that various variants of the differential QD algorithm compute the singular values of bidiagonal matrices to high relative accuracy, and that this algorithm has some better properties than the zero-shift QR algorithm. Gu and Eisenstat [**?**] further extended the relative perturbation theory for singular values. Eisenstat and Ipsen [**?**, **?**] and Li [**?**, **?**] have given perturbation bounds for the perturbations which are given by congruences. Truhar and Slapničar [**?**, **?**] generalized the perturbation bounds by Veselić and Slapničar to the projection to invariant subspace which corresponds to a set of neighboring eigen/singular values. Barlow and Slapničar [**?**] develop the local bounds for the relative perturbations of eigen/singular values. Namely, all bounds described so far are global in the sense that one bound holds for all values/vectors. These bounds are attainable but only for some values and vectors, so the locally optimal bounds for each value and vector are of great interest. Arbenz and Slapničar [**?**] are among many authors who analyzed the implementation of Jacobi methods on multiprocessor computers. Due to their simplicity, the Jacobi type methods, and in particular implicit methods, are very suitable for such computers and attain almost optimal speedups. New generation of processors prefers block version of matrix algorithms which can also be easily implemented for Jacobi methods. Experiments have shown that the Jacobi methods retain their high relative accuracy when implemented on multiprocessor systems.

Majority of the described theoretical results hold in the complex case as well, but, due to need in applications, mostly real versions of algorithms have been analyzed so far. The existing analysis of complex algorithms [**?**, **?**] is very similar to analysis of their real counterparts, which also indicates that the algorithms for complex matrices are as accurate as algorithms for real matrices.

Let us conclude by saying that the research area of the relative perturbation theory and highly accurate algorithms is, due to its importance, very active. Some basic results are simple but considerable improvements of the classical linear algebra results and throw a new light on the behavior of eigen/singular values and vectors under special types of perturbations of matrix elements which

typically occur in practice. Due to good theoretical and experimental results some algorithms are planned to be implemented in LAPACK [**?**]. Most important open problems are: the problem of speed of the accurate algorithms since Jacobi methods are several times slower than QR type methods, and further application of the results to problems and matrices which appear in engineering applications.

References

[1] E. ANDERSON ET AL., *LAPACK Users' Guide*, Second Edition, SIAM, Philadelphia, 1995.

[2] P. ARBENZ, I. SLAPNIČAR, *On an implementation of a one-sided block Jacobi method on a distributed memory computer*, to appear in "Proc. of The Third International Congress on Industrial and Applied Mathematics ICIAM 95", Hamburg, 1995.

[3] J. BARLOW, J. DEMMEL, *Computing accurate eigensystems of scaled diagonally dominant matrices*, SIAM J. Numer. Anal. **27**(1990), 762–791.

[4] J. BARLOW, I. SLAPNIČAR, *Optimal perturbation bounds for the Hermitian eigenvalue problem*, preprint, The Pennsylvania State University, State College, and University of Split, 1996.

[5] J. R. BUNCH, B. N. PARLETT, *Direct methods for solving symmetric indefinite systems of linear equations*, SIAM J. Numer. Anal. **8**(1971), 639–655.

[6] A. DEICHMÖLLER, *Über die Berechnung verallgemeinerter singulärer Werte mittels Jacobi–ähnlicher Verfahren*, Ph. D. thesis, Fernuniversität Hagen, Germany, 1991.

[7] J. DEMMEL, Z. DRMAČ, S. EISENSTAT, M. GU, I. SLAPNIČAR, K. VESELIĆ, *Notes on computing the SVD with high relative accuracy*, preprint, University of California at Berkeley, 1996.

[8] J. DEMMEL, W. GRAGG, *On computing accurate singular values and eigenvalues of matrices with acyclic graphs*, Linear Algebra Appl. **185**(1993), 203–217.

[9] J. DEMMEL, W. KAHAN, *Accurate singular values of bidiagonal matrices*, SIAM J. Sci. Statist. Comput. **11**(1990), 873–912.

[10] J. DEMMEL, K. VESELIĆ, *Jacobi's method is more accurate than QR*, SIAM J. Matrix Anal. Appl. **13**(1992), 1204–1243.

[11] Z. DRMAČ, *Computing the Singular and Generalized Singular Values*, Ph. D. Thesis, Fernuniversität Hagen, Germany, 1994.

[12] Z. DRMAČ, *Implementation of Jacobi rotations for accurate singular value computation in floating point arithmetic*, to appear in SIAM J. Sci. Comp.

[13] Z. DRMAČ, *On the condition behavior in Jacobi method*, SIAM J. Matrix Anal. Appl. **17**(3)(1996).

[14] Z. DRMAČ, *Accurate computation of the product eigenvalue and singular value decompositions with application to system balancing transformations*, preprint, University of Colorado at Boulder, 1995, 1996.

[15] Z. DRMAČ, V. HARI, *On quadratic convergence bounds for the J-symmetric Jacobi method*, Numer. Math. **64**(1993), 147–180.

[16] Z. DRMAČ, K. VESELIĆ, *On the singular value decomposition of matrices generated by finite elements*, preprint, University of Colorado at Boulder and Fernuniversität Hagen, 1995.

[17] S. C. EISENSTAT, I. C. IPSEN, *Relative perturbation techniques for singular value problems*, SIAM J. Numer. Anal. **32**(1995), No. 6.

[18] S. C. EISENSTAT, I. C. F. IPSEN, *Relative Perturbation Bounds for Eigenspaces and Singular Vector Subspaces*, in Applied Linear Algebra, J. G. Lewis, Ed., SIAM, Philadelphia, 1994, 62–66.

[19] K. V. FERNANDO, B. N. PARLETT, *Accurate singular values and differential qd algorithms*, Numer. Math. **67**(1994), 191–229.

[20] D. GOLDBERG, *What every computer scientist should know about floating-point arithmetic*, ACM Computing Surveys **23**(1991), 5–48.

[21] G. H. GOLUB, C. F. VAN LOAN, *Matrix Computations*, Second Ed., The John Hopkins University Press, Baltimore, MD, 1989.

[22] M. GU, S. EISENSTAT, *Relative perturbation theory for eigenvalues*, Computer Science Department Report YALEU/DCS/RR-934, Yale University, 1993.

[23] C. J. G. JACOBI, *Über ein leichtes Verfahren die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen*, Crelle's Journal für reine und angew. Math. **30**(1846), 51–94.

[24] W. KAHAN, *Accurate eigenvalues of a symmetric tridiagonal matrix*, Tech. Report No. CS41, Computer Science Department, Stanford University, Stanford, 1966 (revised 1968).

[25] R.-C. LI, *Relative Perturbation Theory: (i) Eigenvalue Variations*, Computer Science Dept., Technical Report CS-94-252, University of Tennessee, Knoxville, 1994. (LAPACK Working Note # 84.)

[26] R.-C. LI, *Relative Perturbation Theory: (ii) Eigenspace Variations*, Computer Science Dept., Technical Report CS-94-253, University of Tennessee, Knoxville, 1994. (LAPACK Working Note # 85.)

[27] B. PARLETT, *The Symmetric Eigenvalue Problem*, Prentice Hall, Engelwood Cliffs, NJ, 1980.

[28] E. PIETZSCH, *Genaue Eigenwertberechnung nichtsingulärer schiefsymmetrischer Matrizen*, Ph. D. Thesis, Fernuniversität Hagen, Germany, 1993.

[29] R. A. ROSANOFF, J. F. GLOUDEMAN, S. LEVY, *Numerical conditions of stiffness matrix formulations for frame structures*, Proc. of the 2nd Conference on Matrix methods in Structural Mechanics, WPAFB, Dayton, 1968.

[30] S. SINGER, *Computing the Spectral Decomposition of a Positive Definite Matrix*, M. S. Thesis, University of Zagreb, 1993, (in Croatian).

[31] I. SLAPNIČAR, *Accurate Symmetric Eigenreduction by a Jacobi Method*, Ph. D. Thesis, Fernuniversität Hagen, Germany, 1992.

[32] I. SLAPNIČAR, *Componentwise analysis of direct factorization of real symmetric and Hermitian matrix*, submitted to Linear Algebra Appl.

[33] I. SLAPNIČAR, K. VESELIĆ, *Perturbations of the eigenprojections of a factorized Hermitian matrix*, Linear Algebra Appl. **218**(1995), 273–280.

[34] I. SLAPNIČAR, K. VESELIĆ, *Bound for the condition of hyperbolic and symplectic eigenvector matrices*, preprint, University of Split and Fernuniversität Hagen, Germany, 1995.

[35] N. TRUHAR, *Perturbations of Invariant Subspaces*, M. S. Thesis, University of Zagreb, 1995, (in Croatian).

[36] N. TRUHAR, I. SLAPNIČAR, *Relative perturbation bounds for spectral projections*, preprint, University of Osijek and University of Split, 1996.

[37] A. VAN DER SLUIS, *Condition numbers and equilibration of matrices*, Numer. Math. **14**(1969), 14–23.

[38] K. VESELIĆ, *A Jacobi eigenreduction algorithm for definite matrix pairs*, Numer. Math. **64** (1993), 241–269.

[39] K. VESELIĆ, I. SLAPNIČAR, *Floating-point perturbations of Hermitian matrices*, Linear Algebra Appl. **195**(1993), 81–116.

[40] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.