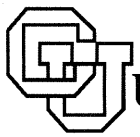
A Tangent Algorithm for Computing the Generalized Singular Value Decomposition*

Zlatko Drmac

CU-CS-815-96



University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE

* This research was supported by National Science Foundation grants ACS-9357812 and ASC-9625912, Department of Energy grant DE-FG03-94ER25215 and the Intel Corporation. Parts of this work were presented at the symposium *Accuracy Issues in Eigenvalue Problems* at the International Congress on Industrial and Applied Mathematics (ICIAM 95) July 3-7, 1995 Hamburg, Germany, and at the XIII Householder Symposium on Numerical Algebra, June 17-21, 1996 Pontresina, Switzerland. The author acknowledges financial support by SIAM (travel grant award to the ICIAM 95) and by the Householder Organizing Committee.

A TANGENT ALGORITHM FOR COMPUTING THE GENERALIZED SINGULAR VALUE DECOMPOSITION*

ZLATKO DRMAČ†

November 20, 1996

Abstract. We present two new algorithms for floating–point computation of the generalized singular values of a real pair (A, B) of full column rank matrices and for floating–point solution of the generalized eigenvalue problem $Hx = \lambda Mx$ with symmetric, positive definite matrices H and M. The pair (A, B) is replaced with an equivalent pair (A', B'), and the generalized singular values are computed as the singular values of the explicitly computed matrix $F = A'B'^{-1}$. The singular values of F are computed using the Jacobi method. The relative accuracy of the computed singular value approximations does not depend on column scalings of A and B, that is, the accuracy is nearly the same for all pairs (AD_1, BD_2), with D_1, D_2 arbitrary diagonal, nonsingular matrices. Similarly, the pencil $H - \lambda M$ is replaced with an equivalent pencil $H' - \lambda M'$, and the eigenvalues of $H - \lambda M$ are computed as the squares of the singular values of $G = L_H L_M^{-1}$, where L_H, L_M are the Cholesky factors of H', M', respectively, and the matrix G is explicitly computed as the solution of a linear system of equations. For the computed approximation $\lambda + \delta\lambda$ of any exact eigenvalue λ, the relative error $|\delta\lambda|/\lambda$ is of order $p(n)\varepsilon \max\{\min_{\Delta\in\mathcal{D}} \kappa_2(\Delta H\Delta), \min_{\Delta\in\mathcal{D}} \kappa_2(\Delta M\Delta)\}$, where $p(n)$ is modestly growing polynomial of the dimension of the problem, ε is the roundoff unit of floating–point arithmetic, $\mathcal{D}$ denotes the set of diagonal nonsingular matrices and $\kappa_2(\cdot)$ is the spectral condition number. Furthermore, floating–point computation corresponds to an exact computation with $H + \delta H$, $M + \delta M$, where, for all i, j, $|\delta H_{ij}|/\sqrt{H_{ii}H_{jj}}$ and $|\delta M_{ij}|/\sqrt{M_{ii}M_{jj}}$ are of order of ε times a modest function of n.

Key words. generalized singular value decomposition, generalized eigenvalue problem, Jacobi method, relative accuracy, singular value decomposition

AMS subject classifications. 65F15, 65F25, 65G05

1. Introduction.

In this paper, we propose two new algorithms for floating–point computation of the generalized singular value decomposition of a matrix pair

$$(A, B) \in \mathbf{R}^{m\times n} \times \mathbf{R}^{p\times n}, \quad \operatorname{rank}(A) = \operatorname{rank}(B) = n, \tag{1.1}$$

and for floating–point solution of the generalized eigenvalue problem

$$Hx = \lambda Mx, \quad H, M \in \mathbf{R}^{n\times n} \text{ symmetric and positive definite.} \tag{1.2}$$

The generalized singular value decomposition (GSVD) is introduced by Van Loan [55], [56], and it naturally arises in applications like equality constrained least squares [56], [57], [7], the general Gauss–Markov linear model [44], linear matrix equations [10], or, for example, the Kronecker canonical form [37].

THEOREM 1.1. *Let $(A, B) \in \mathbf{C}^{m\times n} \times \mathbf{C}^{p\times n}$, and let $m \geq n$, $q = \min\{p, n\}$, $r = \operatorname{rank}(B)$. There exist unitary matrices U, V and nonsingular X of dimensions $m \times m$, $p \times p$ and $n \times n$, respectively, such that $U^*AX = \Sigma_A = \operatorname{diag}(\alpha_1, \ldots, \alpha_n)$, $\alpha_i \geq 0$, $V^*BX = \Sigma_B = \operatorname{diag}(\beta_1, \ldots, \beta_q)$, and $\beta_1 \geq \cdots \geq \beta_r > \beta_{r+1} = \cdots = \beta_q = 0$. If $\alpha_j = 0$ for any j, $r + 1 \leq j \leq n$, then the set of generalized singular values of (A, B) is $\sigma(A, B) = \{\sigma \in \mathbf{R} : \sigma \geq 0\}$. Otherwise, $\sigma(A, B) = \{\alpha_i/\beta_i : i = 1, \ldots, r\}$.*

If in Theorem 1.1 the matrix B is square and nonsingular, the GSVD of (A, B) is equivalent to the singular value decomposition (SVD) of AB^{-1}, and if $B = I$ we have the SVD of A. Furthermore, the equivalence transformation with the matrix X diagonalizes the pencil $A^*A - \lambda B^*B$, that

---

* Technical report CU-CS-815-96, Department of Computer Science, University of Colorado at Boulder.

† Department of Computer Science, University of Colorado, Boulder CO 80309–0430. (zlatko@cs.colorado.edu)
This research was supported by National Science Foundation grants ACS–9357812 and ASC–9625912, Department of Energy grant DE–FG03–94ER25215 and the Intel Corporation. Parts of this work were presented at the symposium *Accuracy Issues in Eigenvalue Problems* at the International Congress on Industrial and Applied Mathematics (ICIAM 95) July 3–7, 1995 Hamburg, Germany, and at the XIII Householder Symposium on Numerical Algebra, June 17–21, 1996 Pontresina, Switzerland. The author acknowledges financial support by SIAM (travel grant award to the ICIAM 95) and by the Householder Organizing Committee.

is, $X^*(A^*A - \lambda B^*B)X$ is diagonal matrix. Hence, if the matrices H and M in relation (1.2) are factored as $H = A^*A$, $M = B^*B$, we can solve the eigenvalue problem (1.2) implicitly by computing the GSVD of (A, B).

Van Loan notes that, if A^*A and B^*B commute, it is possible to choose X unitary. Hence, if $Q = [A^*, B^*]^*$ has orthonormal columns, there exist unitary matrices U, V and W such that $U^*AW = \Sigma_A$ and $V^*BW = \Sigma_B$ are diagonal matrices and $\Sigma_A^*\Sigma_A + \Sigma_B^*\Sigma_B = I$. This is the cosine–sine decomposition (CSD) of the partitioned orthonormal matrix Q. The CSD is defined by Stewart [51] and it is implicitly contained in Davis and Kahan's paper [11]. Paige and Saunders [46] remove the minor constraint $m \geq n$ in Theorem 1.1, and reformulate the decomposition to avoid the use of nonorthogonal transformations.

The relation between the GSVD and the CSD is the basis for two backward stable algorithms, proposed by Stewart [51] and Van Loan [58]. If

$$G \equiv \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} R_G, \quad Q \equiv \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} U & \mathbf{O} \\ \mathbf{O} & V \end{bmatrix} \begin{bmatrix} S \\ C \end{bmatrix} W^* \tag{1.3}$$

are the QR factorization of G and the CSD of Q, respectively, with $C = \mathrm{diag}\,(c_i)$, $S = \mathrm{diag}\,(s_i)$, then the GSVD of (A, B) can be written as $U^*A(R_G^{-1}W) = S$, $V^*B(R_G^{-1}W) = C$, provided that $\mathrm{rank}\,(G) = n$. A nice property of this approach is that the generalized singular values $\{s_i/c_i\}$ can be computed using only orthogonal transformations. It immediately implies that the computed generalized singular values are the exact generalized singular values of $(A + \delta A, B + \delta B)$, where the backward errors δA, δB are small in the Frobenius norm sense. More precisely, $\|\delta A\|_F/\|A\|_F$ and $\|\delta B\|_F/\|B\|_F$ are bounded by the product of the machine roundoff unit ε and a moderate polynomial of the matrix dimensions. Shougen and Shuqin [49] use this approach for the solution of the eigenvalue problem (1.2), where A and B are obtained as the Cholesky factors of H and M, respectively.

Paige [45] generalizes the Kogbetliantz algorithm for SVD computation to matrix pairs and gives an elegant implementation that initially transforms A and B to a pair of triangular matrices and preserves the triangular form in the iterative phase by using suitable plane rotations. A variation of Paige's algorithm is given by Bai and Demmel [4]. This algorithm is also norm–wise backward stable, and it is implemented as the LAPACK [2] procedure `STGSJA()`.

Deichmöller and Veselić [14], [13] and Drmač [21] use an implicit variant of the Falk–Langemeyer method [26], [30] and show that, in certain well–conditioned cases, the generalized singular values of a pair (A, B) of full column rank matrices are computed with high relative accuracy. A nice property of this method is that the relative accuracy of the computed generalized singular value approximations is nearly the same for all pairs of the form (AD_1, BD_2), where D_1 and D_2 are arbitrary diagonal nonsingular matrices. An important difference between this method and the methods from [51], [58] and [45] is that, in this method, the elementary transformations are nonorthogonal.

The possibility of using the SVD of $T = AB^{-1}$ in case of square nonsingular B is not considered to be a numerically attractive approach. The main reason is that ill–conditioning of B with respect to inversion may produce an inaccurate T in floating–point arithmetic. Our goal in this paper is to show that the pair from relation (1.1) can be replaced, in a numerically stable and efficient way, with an equivalent pair (A', R) such that R is triangular and nonsingular and such that the generalized singular values of (A', R) can be accurately computed by computing the SVD of the explicitly computed matrix $F = A'R^{-1}$. To obtain A' and R, we use certain matrix column scalings and the QR factorization with column pivoting. The matrix F is computed as the solution of triangular systems of linear equations, and the SVD of F is computed using the Jacobi SVD algorithm. Using backward error analysis, we show that floating–point implementation of this procedure is equivalent to exact computation with $A + \delta A$ and $B + \delta B$ such that, for all i,

$$\|\delta Ae_i\|_2 \leq f(n, R)\varepsilon\|Ae_i\|_2, \quad \|\delta Be_i\|_2 \leq g(p, n)\varepsilon\|Be_i\|_2, \tag{1.4}$$

where $f(n,R)$ is usually a modestly growing function, and $g(p,n)$ is a linear polynomial in p and n. (In relation (1.4), e_i is the ith column of the identity matrix I and $\|\cdot\|_2$ denotes the Euclidean vector norm. We use $\|\cdot\|_2$ to denote the spectral matrix norm as well.) Furthermore, we show that the generalized singular values $\sigma_1 \geq \cdots \geq \sigma_n$ of (A,B) are computed with an error bound of the form

$$\max_{1\leq i\leq n} \frac{|\delta\sigma_i|}{\sigma_i} \leq f(n,R)\sqrt{n}\epsilon\|A_c^{\dagger}\|_2 + g(p,n)\sqrt{n}\epsilon\|B_c^{\dagger}\|_2 + O(\epsilon^2),$$

where A_c and B_c are obtained from A and B, respectively, by scaling their columns to have unit Euclidean lengths. The symbol † denotes the Moore–Penrose generalized inverse of a matrix. We have two theoretical bounds for $f(n,R)$. The first one guarantees that $f(n,R)$ is moderate if $\|B_c^{\dagger}\|_2$ is such. The second one guarantees that $f(n,R)$ is bounded by a modestly growing function of n, independent of B. If we use strong rank–revealing QR factorization of Gu and Eisenstat [31], the theoretical bound for $f(n,R)$ is comparable to the Wilkinson's bound for pivot growth in Gaussian elimination with complete pivoting (see [62]).

We apply a similar procedure for the solution of the eigenvalue problem (1.2): we replace H and M with an equivalent pair H', M', then we compute the Cholesky factors A' and B' of H', M', respectively, and, finally, we compute the SVD of the explicitly computed matrix $F = A'B'^{-1}$. We show that floating–point implementation of these operations is equivalent to exact computation with the symmetric pencil $(H+\delta H) - \lambda(M+\delta M)$, where, for all i, j, the values of $|\delta H_{ij}|/\sqrt{H_{ii}H_{jj}}$ and $|\delta M_{ij}|/\sqrt{M_{ii}M_{jj}}$ are bounded by ϵ times a moderate function of n. Furthermore, the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ are approximated with an error bound of the form

$$\max_{1\leq i\leq n} \frac{|\delta\lambda_i|}{\lambda_i} \leq h(n)\epsilon(\|H_s^{-1}\|_2 + \|M_s^{-1}\|_2),$$

where $(H_s)_{ij} = H_{ij}/\sqrt{H_{ii}H_{jj}}$, $(M_s)_{ij} = M_{ij}/\sqrt{M_{ii}M_{jj}}$, $1 \leq i,j \leq n$, and $h(n)$ is a modestly growing function of n. The only condition for the element–wise backward stability and the high relative accuracy is that floating–point Cholesky factorizations of H and M are guaranteed to complete without breakdown. Using the results of Demmel [15], Demmel and Veselić [17] and Veselić and Slapničar [60], we show that floating-point solution of the eigenvalue problem (1.2) is numerically feasible only if $\|H_s^{-1}\|_2$ and $\|M_s^{-1}\|_2$ are moderate.

We call the new algorithms *tangent algorithms*. For the sake of simplicity, let the matrix B in (1.3) be square and nonsingular. Since, in relation (1.3), $Q_1 = AR_G^{-1}$ and $Q_2 = BR_G^{-1}$, it follows that $Q_1Q_2^{-1} = AB^{-1} = F$ and the SVD of F is, in a sense, a *tangent decomposition* of Q. Indeed, if $[\Sigma^T, \mathbf{O}]^T = \Omega^* F \Xi$ is the SVD of F, then the CSD of the matrix Q in relation (1.3) is

$$Q_1 = \Omega\begin{bmatrix} \Sigma(I+\Sigma^2)^{-1/2} \\ \mathbf{O} \end{bmatrix}\Upsilon, \quad Q_2 = \Xi(I+\Sigma^2)^{-1/2}\Upsilon, \tag{1.5}$$

where $\Upsilon = (I+\Sigma^2)^{1/2}\Xi^* Q_2$ is unitary. With a suitably chosen ordering of the singular values we have $C = (I+\Sigma^2)^{-1/2}$, $S = \Sigma C$, and we easily recognize sines and cosines expressed as functions of tangents of certain angles. Recall that the mapping $\tan\theta \mapsto (\sin\theta, \cos\theta)$ is well–conditioned for all θ. On the other hand, the mappings $\sin\theta \mapsto \cos\theta$, $\sin\theta \mapsto \tan\theta$ ($\cos\theta \mapsto \sin\theta$, $\cos\theta \mapsto \tan\theta$) become ill–conditioned as $\sin\theta$ ($\cos\theta$) approaches one.

The paper is organized as follows. In § 2, we give detailed analysis of the new algorithm for the GSVD computation. We show that the new algorithm is backward stable and that it computes with nearly the same accuracy the generalized singular values of all regular pairs (AD_1, BD_2), where D_1, D_2 are arbitrary diagonal nonsingular matrices. In § 3, we analyze the new algorithm for computing the eigenvalues of the positive definite symmetric pencil $H - \lambda M$. We show that the new algorithm computes the eigenvalues with high relative accuracy and that the computed eigenvalues are the exact eigenvalues of a symmetric pencil $(H+\delta H) - \lambda(M+\delta M)$, where, for

all i, j, the values of $|\delta H_{ij}|/\sqrt{H_{ii}H_{jj}}$ and $|\delta M_{ij}|/\sqrt{M_{ii}M_{jj}}$ are bounded by ϵ times a moderate function of the matrix dimension. As a corollary of our analysis, we obtain sharp element–wise backward stability of the Jacobi method for eigenvalue computation of symmetric positive definite matrices. In § 4, we briefly analyze two similar algorithms for the solution of the generalized singular value and the generalized eigenvalue problems (1.1) and (1.2). Finally, in § 5, we present results of extensive numerical testing of floating point implementation of the tangent algorithm. Our experiments show that the tangent algorithm software runs as predicted by the theory.

2. The GSVD of a regular pair (A, B). In this section, we describe and analyze in detail the new algorithm for the GSVD computation. We consider a real matrix pair $(A, B) \in \mathbf{R}^{m\times n} \times \mathbf{R}^{p\times n}$, where $\mathrm{rank}\,(A) = \mathrm{rank}\,(B) = n$. Our goal is to approximate the singular values $\sigma_1 \geq \cdots \geq \sigma_n$ of (A, B) with high relative accuracy in floating–point arithmetic. In § 2.1, we analyze the sensitivity of the singular values of (A, B). We show that, for certain small relative perturbations that occur in floating–point computation, the relative error in the perturbed generalized singular values is determined by $\|A_c^\dagger\|_2$ and $\|B_c^\dagger\|_2$, where A_c and B_c are defined by

$$A = A_c\mathrm{diag}\,(\|Ae_i\|_2), \quad B = B_c\mathrm{diag}\,(\|Be_i\|_2). \tag{2.6}$$

In § 2.2, we define the new algorithm, and in § 2.3, we analyze conditions for backward and forward stability of the new algorithm. We show that the algorithm has nearly the same accuracy properties for all pairs $\{(AD_1, BD_2),\ D_1, D_2$ diagonal matrices $\}$. In § 2.4, we describe a modification of the new algorithm and give sharp backward and forward error bounds. In § 2.5, we describe how the Jacobi rotations can be used as a preconditioner for the new algorithm.

2.1. Sensitivity of the generalized singular values. Let (A, B) be a pair of full column rank matrices. Our goal is to estimate how the generalized singular values $\sigma_1 \geq \cdots \geq \sigma_n$ of (A, B) change, if A and B are perturbed to $A+\delta A$, $B+\delta B$, respectively. We are interested in the relative size of the perturbation, that is, we seek an uniform bound for $|\delta\sigma_i|/\sigma_i$, $1 \leq i \leq n$. The obtained bound is used in the analysis of the new algorithm, where the perturbations δA and δB are the roundoff errors.

In the following theorem, we use multiplicative representation of $A + \delta A$ and $B + \delta B$ and apply the variational characterization of the generalized singular values to estimate the relative distance between the true and the perturbed singular values.

THEOREM 2.1. *Let $A \in \mathbf{R}^{m\times n}$, $B \in \mathbf{R}^{p\times n}$ have full column rank, let $\tilde{A} = A+\delta A$, $\tilde{B} = B+\delta B$ with $\|\delta AA^\dagger\|_2 < 1$, $\|\delta BB^\dagger\|_2 < 1$. If $\sigma_1 \geq \cdots \geq \sigma_n$ and $\tilde{\sigma}_1 \geq \cdots \geq \tilde{\sigma}_n$ are the generalized singular values of (A, B) and $(\tilde{A}, \tilde{B})$, respectively, then*

$$\max_{1\leq i\leq n} \frac{|\tilde{\sigma}_i - \sigma_i|}{\sigma_i} \leq \frac{\|\delta AA^\dagger\|_2 + \|\delta BB^\dagger\|_2}{1 - \|\delta BB^\dagger\|_2}. \tag{2.7}$$

Proof. Note that the assumptions of the theorem imply that $\tilde{A} = (I + \delta AA^\dagger)A$ and $\tilde{B} = (I + \delta BB^\dagger)B$ are full column rank matrices. Furthermore, it holds, for all nonzero vectors x, that

$$\frac{\|Ax\|_2}{\|Bx\|_2}\frac{1 - \|\delta AA^\dagger\|_2}{1 + \|\delta BB^\dagger\|_2} \leq \frac{\|\tilde{A}x\|_2}{\|\tilde{B}x\|_2} \leq \frac{\|Ax\|_2}{\|Bx\|_2}\frac{1 + \|\delta AA^\dagger\|_2}{1 - \|\delta BB^\dagger\|_2}, \tag{2.8}$$

and relation (2.7) follows from the variational characterization of the generalized singular values. (Cf. [33], [34], [28].) □

Hence, if $\|\delta A\|_2/\|A\|_2$ and $\|\delta B\|_2/\|B\|_2$ are sufficiently small, relation (2.7) implies

$$\max_{1\leq i\leq n} \frac{|\tilde{\sigma}_i - \sigma_i|}{\sigma_i} \leq \frac{\frac{\|\delta A\|_2}{\|A\|_2}\kappa_2(A) + \frac{\|\delta B\|_2}{\|B\|_2}\kappa_2(B)}{1 - \frac{\|\delta B\|_2}{\|B\|_2}\kappa_2(B)}. \tag{2.9}$$

The error bound in relation (2.9) can be improved if information about δA, δB is refined.

COROLLARY 2.2. *Let in Theorem 2.1* $\|\delta Ae_i\|_2 \le \varepsilon_A\|Ae_i\|_2$, $\|\delta Be_i\|_2 \le \varepsilon_B\|Be_i\|_2$, $1 \le i \le n$, *and let* q *denote the maximal number of nonzero entries in any row of* δA *and* δB. *If* $\sqrt{q}\varepsilon_A\|A_c^\dagger\|_2 < 1$, $\sqrt{q}\varepsilon_B\|B_c^\dagger\|_2 < 1$, *then*

$$\max_{1\le i\le n}\frac{|\tilde\sigma_i - \sigma_i|}{\sigma_i} \le \sqrt{q}\frac{\varepsilon_A\|A_c^\dagger\|_2 + \varepsilon_B\|B_c^\dagger\|_2}{1 - \|B_c^\dagger\|_2\sqrt{n}\varepsilon_B}. \tag{2.10}$$

Proof. Note that $\|\delta AA^\dagger\|_2 \le \sqrt{q}\varepsilon_A\|A_c^\dagger\|_2$ and $\|\delta BB^\dagger\|_2 \le \sqrt{q}\varepsilon_B\|B_c^\dagger\|_2$, and apply Theorem 2.1. □

REMARK 2.3. A similar relative error bound is derived by Demmel and Veselić [17] under assumption that $\|\delta Ax\|_2 \ll \|Ax\|_2$, $\|\delta Bx\|_2 \ll \|Bx\|_2$ for all $x \ne \mathbf{0}$. For a more general theory, see [52], [53], [43], [3], [39].

REMARK 2.4. An advantage of the estimate (2.10) over (2.9) is seen from the relation $\|A_c^\dagger\|_2 \le \kappa_2(A_c) \le \sqrt{n}\min_D \kappa_2(AD)$, where the minimum is taken over the set of diagonal nonsingular matrices (cf. [54]). Hence, $\kappa_2(A_c)$ is never much larger than $\kappa_2(A)$, and it can be much smaller. If δA and δB are small rounding errors in the entries of A and B, respectively, an application of Theorem 2.1 yields the following corollary.

COROLLARY 2.5. *If in Theorem 2.1* $|\delta A| \le \epsilon_A|A|$, $|\delta B| \le \epsilon_B|B|$, *where* $\epsilon_A\| \,|A|\cdot|A^\dagger|\,\|_2 < 1$, $\epsilon_B\| \,|B|\cdot|B^\dagger|\,\|_2 < 1$, *then*

$$\max_{1\le i\le n}\frac{|\tilde\sigma_i - \sigma_i|}{\sigma_i} \le \frac{\epsilon_A\| \,|A|\cdot|A^\dagger|\,\|_2 + \epsilon_B\| \,|B|\cdot|B^\dagger|\,\|_2}{1 - \epsilon_B\| \,|B|\cdot|B^\dagger|\,\|_2}. \tag{2.11}$$

2.2. The tangent algorithm. The new algorithm has two major stages. In the first stage, the pair (A, B) is reduced to a single matrix F. In the second stage, the algorithm computes the SVD of F. We use the fact that the generalized singular values of (A, B) are invariants of the equivalence transformation

$$(A, B) \longmapsto (A', B') = (U^\tau AS, V^\tau BS),$$

where U, V are arbitrary orthogonal matrices and S is an arbitrary nonsingular matrix. The pairs (A, B) and (A', B') are by definition equivalent.

ALGORITHM 2.6.

Input $A \in \mathbf{R}^{m\times n}$, $B \in \mathbf{R}^{p\times n}$, $m \ge n$, $\operatorname{rank}(B) = n$.

Step 1 Compute $\Delta_A = \operatorname{diag}(\|Ae_i\|_2)$ and $A_c = A\Delta_A^{-1}$, $B_1 = B\Delta_A^{-1}$.

Step 2 Compute the QR factorization with column pivoting, $\begin{bmatrix} R \\ \mathbf{O} \end{bmatrix} = Q^\tau B_1\Pi$.

Step 3 Compute $F = A_c\Pi R^{-1}$ by solving the equation $FR = A_c\Pi$.

Step 4 Compute the SVD of F using the Jacobi SVD algorithm, $\begin{bmatrix} \Sigma \\ \mathbf{O} \end{bmatrix} = V^\tau FU$.

Step 5 Compute the matrices $X = \Delta_A^{-1}\Pi R^{-1}U$ and $W = Q\begin{bmatrix} U & \mathbf{O} \\ \mathbf{O} & I_{p-n} \end{bmatrix}$.

Output The GSVD of (A, B) reads $\begin{bmatrix} V^\tau A \\ W^\tau B \end{bmatrix}X = \begin{bmatrix} (\Sigma, \mathbf{O})^\tau \\ (I, \mathbf{O})^\tau \end{bmatrix}$.

The first three steps in Algorithm 2.6 can be implemented efficiently using the LAPACK [2] and the Level 3 BLAS [19] libraries. In Step 1, we assume that A has nonzero columns. We show in § 5 that zero columns of A (and the corresponding zero generalized singular values of (A, B)) can be deflated

without error. The Jacobi SVD algorithm can be implemented as reliable mathematical software using [17], [20]. Since Algorithm 2.6 computes the matrix X in factored form as $X = \Delta_A^{-1}\Pi R^{-1}U$, it can optionally return X^{-1} as $X^{-1} = U^T R\Pi^T\Delta_A$.

2.3. Error analysis. Now we show that, in floating–point arithmetic, Algorithm 2.6 is numerically stable. We use standard model of floating point arithmetic,

$$\boldsymbol{fl}(a \odot b) = (a \odot b)(1+\xi), \quad \boldsymbol{fl}(\sqrt{c}) = \sqrt{c}(1+\zeta), \quad |\xi|, |\zeta| \leq \boldsymbol{\varepsilon}, \tag{2.12}$$

where a, b and c are floating–point numbers, $\odot$ denotes any of the four elementary operations $+$, $-$, $\cdot$ and $\div$, and $\boldsymbol{\varepsilon}$ is the round–off unit. From relation (2.12) it follows that the Euclidean length $\|x\|_2$ of a floating–point vector $x \in \mathbf{R}^m$ is computed as

$$\boldsymbol{fl}(\|x\|_2) = \|x\|_2(1+\epsilon), \quad |\epsilon| \leq \varepsilon_{\ell_2}(m) \leq (1+\varepsilon)^{(m+2)/2} - 1. \tag{2.13}$$

Using double precision accumulation or compensated summation (cf. [32, Chapter 4]), the bound for $\varepsilon_{\ell_2}(m)$ can be reduced to $O(1)\varepsilon$ for all $m \leq 1/\varepsilon$. Furthermore, using (2.12) and an elegant technique of Gentleman, we can prove the following backward error estimate for the floating–point QR factorization in Step 2 of Algorithm 2.6. (Cf. [64], [27], [5], [21], [32].)

PROPOSITION 2.7. *Let the QR factorization of $B_1 \in \mathbf{R}^{p\times n}$ be computed by a sequence of Givens rotations in some prescribed order. Let all rotations be divided into $\wp$ sets, where each set contains rotations that can be applied simultaneously to different pairs of matrix rows. If $\tilde{R}$ is the computed triangular factor, then there exist an orthogonal matrix Q' and a backward error δB_1 such that*

$$(B_1 + \delta B_1) = Q'\begin{bmatrix}\tilde{R}\\ \mathbf{O}\end{bmatrix}, \quad \|\delta B_1 e_i\|_2 \leq \varepsilon_{QR}(p,n)\|B_1 e_i\|_2, \quad 1 \leq i \leq n,$$

where $\varepsilon_{QR}(p,n) \leq ((1+6\varepsilon)^{\wp} - 1)$. For the usual column–wise ordering of Givens rotations we have $\wp = p + n - 3$. For some more sophisticated strategies as, for example, in [42], for large $p \gg n$ it holds that $\wp \approx \log_2 p + (n-1)\log_2\log_2 p$.

Next we show that, in the case of moderate $\|B_c^{\dagger}\|_2$, the reduction of the pair (A, B) to the single matrix F is backward stable in a certain very strong sense.

THEOREM 2.8. *Let in Algorithm 2.6 $\tilde{R}$ and $\tilde{F}$ be the computed floating–point approximations of the matrices R and F, respectively. Let $\eta_B = \varepsilon_{QR}(p,n)(1+\varepsilon) + \varepsilon$, where $\varepsilon_{QR}(p,n)$ is defined in Proposition 2.7, and let $\sqrt{n}\eta_B\|B_c^{\dagger}\|_2 < 1$ and $n\varepsilon\| \,|\tilde{R}^{-1}|\cdot|\tilde{R}|\,\|_1 < 1$. Then there exist backward perturbations δA_b and δB_b such that the pairs $(\tilde{F}, \begin{bmatrix}I\\ \mathbf{O}\end{bmatrix})$ and $(A+\delta A_b, B+\delta B_b)$ are equivalent. Furthermore, it holds, for all i, that*

$$\|\delta A_b e_i\|_2 \leq \eta_A\|Ae_i\|_2, \quad \eta_A = \frac{\varepsilon_T(n)\| \,|\tilde{R}^{-1}|\cdot|\tilde{R}|\,\|_1}{1-\varepsilon_T(n)\| \,|\tilde{R}^{-1}|\cdot|\tilde{R}|\,\|_1}\frac{1+\varepsilon_{\ell_2}(m)}{1-\varepsilon_{\ell_2}(m)}(1+\varepsilon)+\varepsilon \tag{2.14}$$

$$\|\delta B_b e_i\|_2 \leq \eta_B\|Be_i\|_2, \quad \eta_B = \varepsilon_{QR}(p,n)(1+\varepsilon)+\varepsilon, \tag{2.15}$$

where $\varepsilon_T(n) \leq n\varepsilon$. Hence, if $\sqrt{n}\eta_A\|A_c^{\dagger}\|_2 < 1$, and if $\sigma_1 \geq \cdots \geq \sigma_n$ and $\tilde{\sigma}_1 \geq \cdots \geq \tilde{\sigma}_n$ are the generalized singular values of (A,B) and the singular values of $\tilde{F}$, respectively, then

$$\max_{1\leq i\leq n}\frac{|\tilde{\sigma}_i - \sigma_i|}{\sigma_i} \leq \sqrt{n}\frac{\eta_A\|A_c^{\dagger}\|_2 + \eta_B\|B_c^{\dagger}\|_2}{1-\sqrt{n}\eta_B\|B_c^{\dagger}\|_2}. \tag{2.16}$$

Proof. Let $\tilde{\Delta}_A$, $\tilde{A}_c$ and $\tilde{B}_1$ be the computed approximations of Δ_A, A_c and B_1, respectively, and let $\tilde{R}$ be the computed triangular factor of $\tilde{B}_1\tilde{\Pi}$, where $\tilde{\Pi}$ is the computed permutation. Then there exist an orthogonal matrix Q_B and a backward error $\delta\tilde{B}_1$ such that

$$(\tilde{B}_1 + \delta\tilde{B}_1)\tilde{\Pi} = Q_B\begin{bmatrix}\tilde{R}\\ \mathbf{O}\end{bmatrix}, \quad \|\delta\tilde{B}_1 e_i\|_2 \leq \varepsilon_{QR}(p,n)\|\tilde{B}_1 e_i\|_2, \quad 1 \leq i \leq n. \tag{2.17}$$

Note that the matrix $\tilde{R}$ is nonsingular. (Cf. relation (2.20) below.) Let $\tilde{F}$ be the matrix computed by solving the matrix equation $F\tilde{R} = \tilde{A}_c\tilde{\Pi}$ in floating–point arithmetic. Using Wilkinson's analysis [63], we know that $\tilde{F}$ satisfies the following equations

$$e_k^T\tilde{F}(\tilde{R} + \delta\tilde{R}^{(k)}) = e_k^T\tilde{A}_c\tilde{\Pi}, \quad |\delta\tilde{R}_{ij}^{(k)}| \leq (j-i+1)\varepsilon|\tilde{R}_{ij}|, \ 1 \leq i \leq j \leq n, \ 1 \leq k \leq m.$$

Hence, $\tilde{F}\tilde{R} - \tilde{A}_c\tilde{\Pi} = \mathcal{E}$, $|\mathcal{E}| \leq \varepsilon_T(n)|\tilde{F}| \cdot |\tilde{R}|$, $\varepsilon_T(n) \leq n\varepsilon$, and we can write $\tilde{F}$ as

$$\tilde{F} = (\tilde{A}_c + \delta\tilde{A}_c)\tilde{\Pi}\tilde{R}^{-1}, \ \ \delta\tilde{A}_c = \mathcal{E}\tilde{\Pi}^T.$$

An easy calculation shows that $|\delta\tilde{A}_c|(I - \varepsilon_T(n)\tilde{\Pi}|\tilde{R}^{-1}| \cdot |\tilde{R}|\tilde{\Pi}^T) \leq \varepsilon_T(n)|\tilde{A}_c\tilde{\Pi}\tilde{R}^{-1}| \cdot |\tilde{R}|\tilde{\Pi}^T$. Since $I - \varepsilon_T(n)\tilde{\Pi}|\tilde{R}^{-1}| \cdot |\tilde{R}|\tilde{\Pi}^T$ is an M–matrix, we have

$$|\delta\tilde{A}_c| \leq \varepsilon_T(n)|\tilde{A}_c| \ (\tilde{\Pi}|\tilde{R}^{-1}| \cdot |\tilde{R}|\tilde{\Pi}^T)(I - \varepsilon_T(n)\tilde{\Pi}|\tilde{R}^{-1}| \cdot |\tilde{R}|\tilde{\Pi}^T)^{-1}. \tag{2.18}$$

From relation (2.18) it follows, for all i, that

$$\|\delta\tilde{A}_c e_i\|_2 \leq \max_{1\leq k\leq n} \|\tilde{A}_c e_k\|_2 \frac{\varepsilon_T(n)\| \ |\tilde{R}^{-1}| \cdot |\tilde{R}| \ \|_1}{1 - \varepsilon_T(n)\| \ |\tilde{R}^{-1}| \cdot |\tilde{R}| \ \|_1}, \tag{2.19}$$

where $\max_{1\leq k\leq n} \|\tilde{A}_c e_k\|_2 \leq \frac{1+\varepsilon}{1-\varepsilon\iota_2(m)}$. On the other hand, since there exist small element–wise perturbations δA_e, δB_e such that $\tilde{A}_c = (A + \delta A_e)\tilde{\Delta}_A^{-1}$, $|\delta A_e| \leq \varepsilon|A|$, and $\tilde{B}_1 = (B + \delta B_e)\tilde{\Delta}_A^{-1}$, $|\delta B_e| \leq \varepsilon|B|$, we can write $\tilde{F}$ and $\tilde{R}$, respectively, as

$$\begin{bmatrix} \tilde{R} \\ \mathbf{O} \end{bmatrix} = Q_B^T(B + \delta B_b)\tilde{\Delta}_A^{-1}\tilde{\Pi}, \quad \delta B_b = \delta B_e + \delta\tilde{B}_1\tilde{\Delta}_A, \tag{2.20}$$

$$\tilde{F} = (A + \delta A_b)\tilde{\Delta}_A^{-1}\tilde{\Pi}\tilde{R}^{-1}, \quad \delta A_b = \delta A_e + \delta\tilde{A}_c\tilde{\Delta}_A. \tag{2.21}$$

Hence, $\tilde{F}$ is the exact result of the computation (2.20), (2.21), and the estimates (2.14), (2.15) follow from (2.17), (2.19). Relation (2.16) follows from Corollary 2.2. □

From relation (2.15), it follows that relative norm–wise backward error in each column of B is of the order of $(p+n)\varepsilon$. On the other hand, the relative backward error in the columns of A is of order of $n\varepsilon\| \ |\tilde{R}^{-1}| \cdot |\tilde{R}| \ \|_1$. Moreover, if we use double precision accumulation, the bound for η_A in relation (2.14) reduces to the order of $\varepsilon\| \ |\tilde{R}^{-1}| \cdot |\tilde{R}| \ \|_1$. For comparison, computing the QR factorization of A, using Householder reflections, introduces backward error δA such that, for all i, $\|\delta Ae_i\|_2 \leq O(mn)\varepsilon\|Ae_i\|_2$. Hence, for $\| \ |\tilde{R}^{-1}| \cdot |\tilde{R}| \ \|_1$ not too much larger than mn, the bound in relation (2.14) is comparable to the backward error bound for the Householder QR factorization. We now show that the condition number

$$\beta_1(\tilde{R}) = \| \ |\tilde{R}^{-1}| \cdot |\tilde{R}| \ \|_1 \tag{2.22}$$

is usually of moderate size. A bound for the matrix $|\tilde{R}^{-1}| \cdot |\tilde{R}|$ generally depends on the column pivoting in the QR factorization. From Theorem 2.8, it follows that the column pivoting should be chosen to minimize $\| \ |\tilde{R}^{-1}| \cdot |\tilde{R}| \ \|_1$. In our implementation of Algorithm 2.6, we use column pivoting so that

$$\tilde{R}_{ii}^2 \geq \sum_{k=i}^{j} \tilde{R}_{kj}^2, \ \ 1 \leq i \leq j \leq n, \tag{2.23}$$

cf. [9]. (Due to round–off, relation (2.23) holds up to small relative error which we ignore.) In the next proposition, we use relation (2.23) to show that $\| \ |\tilde{R}^{-1}| \cdot |\tilde{R}| \ \|_1$ is modest if $\|B_r^\dagger\|_2$ is such. Furthermore, we show that $\| \ |\tilde{R}^{-1}| \cdot |\tilde{R}| \ \|_1$ is bounded by a function of n, independent of B.

PROPOSITION 2.9. *Let $\tilde{R}$ be as in (2.23) and let $\tilde{R} = \tilde{R}_c \mathrm{diag}\,(\|\tilde{R}e_i\|_2) = \mathrm{diag}\,(\|\tilde{R}^T e_i\|_2)\tilde{R}_r$. Then $|\tilde{R}_r^{-1}| \le \sqrt{n}|\tilde{R}_c^{-1}|$, and, hence, $\| \, |\tilde{R}_r^{-1}| \, \|_2 \le \sqrt{n} \| \, |\tilde{R}_c^{-1}| \, \|_2$ and*

$$\| \, |\tilde{R}^{-1}| \cdot |\tilde{R}| \, \|_1 \le n^{3/2} \|\tilde{R}_c^{-1}\|_1 \le n^2 \|\tilde{R}_c^{-1}\|_2 \le n^2 \|B_c^\dagger\|_2 \frac{1+\sqrt{n}\eta_B}{1-\sqrt{n}\eta_B\|B_c^\dagger\|_2}. \tag{2.24}$$

Furthermore, it holds $\| \, |\tilde{R}^{-1}| \cdot |\tilde{R}| \, \|_1 \le 2^n - 1$, independent of B.

Proof. From relation (2.23) it follows that $|\tilde{R}_r^{-1}|_{ii} \le \sqrt{n-i+1}$, $1 \le i \le n$, and that

$$|(\tilde{R}_r^{-1})_{ij}| = \frac{\|\tilde{R}^T e_j\|_2}{\|\tilde{R}e_i\|_2} |\tilde{R}_c^{-1}|_{ij} \le \sqrt{n-j+1} \frac{|\tilde{R}_{jj}|}{|\tilde{R}_{ii}|} |\tilde{R}_c^{-1}|_{ij}, \quad 1 \le i \le j \le n.$$

Hence, $|\tilde{R}_r^{-1}| \le \sqrt{n}|\tilde{R}_c^{-1}|$ and relation (2.24) follows from the properties of matrix norms and from Corollary 2.2. Furthermore, from [24], [23] (see also [38]), we have

$$|\tilde{R}^{-1}|e_i \le \frac{1}{|\tilde{R}_{ii}|} t(i), \quad t(i) = (2^{i-2}, 2^{i-3}, \ldots, 2, 1, 1, 0, \ldots, 0)^T, \quad 2 \le i \le n. \tag{2.25}$$

Hence, it follows that, for all i, $|\tilde{R}^{-1}| \cdot |\tilde{R}|e_i \le \sum_{k=1}^{i} t(k) \le (2^{n-1}, 2^{n-2}, \ldots, 2, 1)^T$. □

REMARK 2.10. In the practice, $\| \, |\tilde{R}^{-1}| \cdot |\tilde{R}| \, \|_1$ usually behaves like $O(n)$. Gu and Eisenstat [31] describe a column pivoting that can be used to replace the exponential factor 2^n in Proposition 2.9 with Wilkinson's $O(n^{1+(1/4)\log n})$ factor that bounds the pivot growth in Gaussian elimination with complete pivoting. Furthermore, note that $B_1 = B_c(\Delta_B \Delta_A^{-1})$, $\Delta_B = \mathrm{diag}\,(\|Be_i\|_2)$, and that the bound for $\beta_1(\tilde{R})$ depends on B_c and not on $\Delta = \Delta_B \Delta_A^{-1}$. In fact, it follows from the analyses from [59], [17] and the proof of Proposition 2.9 that an ill-conditioned Δ may only help to reduce $\beta_1(\tilde{R})$.

The last step in Algorithm 2.6 is computation of the SVD of $\tilde{F}$. We choose the implicit Jacobi SVD algorithm because of its ability to compute very accurate approximations of all singular values. Demmel and Veselić [17] have shown that the Jacobi SVD algorithm is more accurate than any other method that begins by reducing the matrix to bidiagonal form.

PROPOSITION 2.11. *Let $\tilde{F}^{(0)} \equiv \tilde{F}$, let $\tilde{\sigma}_1 \ge \cdots \ge \tilde{\sigma}_n > 0$ be the singular values of $\tilde{F}$, and let*

$$\tilde{F}^{(k+1)} = (\tilde{F}^{(k)} + \delta\tilde{F}^{(k)})\tilde{U}^{(k)}, \quad k = 0, 1, \ldots \tag{2.26}$$

be the matrices computed by floating-point Jacobi SVD algorithm. In relation (2.26), $\tilde{U}^{(k)}$ is the plane rotation that transforms the columns $\tilde{F}^{(k)}e_p$ and $\tilde{F}^{(k)}e_q$, where $(p,q) = (p(k), q(k))$ is pivot position in kth step. The matrix $\delta\tilde{F}^{(k)}$ is the backward error in kth step. Let

$$\tilde{F}^{(k)}_{c_{1,1,n-2}} = [\, \frac{\tilde{F}^{(k)}e_{p(k)}}{\|\tilde{F}^{(k)}e_{p(k)}\|_2}, \; \frac{\tilde{F}^{(k)}e_{q(k)}}{\|\tilde{F}^{(k)}e_{q(k)}\|_2}, \; \Omega^{(k)} \,], \tag{2.27}$$

where $\Omega^{(k)}$ is the left singular vector matrix of $[\, \tilde{F}^{(k)}e_j : j \notin \{p(k), q(k)\} \,]$. Let $\tilde{F}^{(\ell)}$ be the first matrix in the sequence $(\tilde{F}^{(k)}, k \ge 0)$ that satisfies

$$\max_{i,j} |\boldsymbol{fl}(\cos \angle(\tilde{F}^{(\ell)}e_i, \tilde{F}^{(\ell)}e_j))| \le \mathtt{tol}, \tag{2.28}$$

where $\mathtt{tol} \approx m\varepsilon$ is given threshold. If $\tilde{\sigma}'_1 \ge \cdots \ge \tilde{\sigma}'_n$ are the sorted floating-point values of the Euclidean column norms of $\tilde{F}^{(\ell)}$, then

$$\max_{1 \le i \le n} \frac{|\tilde{\sigma}'_i - \tilde{\sigma}_i|}{\tilde{\sigma}_i} \le \left(\frac{1+\varepsilon_{\ell_2}(m)}{1-n\tau(m)} (1+2.5\varepsilon)^\ell \prod_{k=0}^{\ell} (1 + 5.1\sqrt{2}\varepsilon \|(\tilde{F}^{(k)}_{c_{1,1,n-2}})^\dagger\|_2) \right) - 1,$$

where $\max_{i,j} |\cos \angle(\tilde{F}^{(\ell)}e_i, \tilde{F}^{(\ell)}e_j)| \le \tau(m) = \mathtt{tol} + O(m\varepsilon)$.

Proof. The proof is based on perturbation analysis of a single floating–point Jacobi rotation. For the sake of simplicity, we consider the Jacobi rotation that transforms the first two columns of $\tilde{F}$. (Any other pivot pair can be by error–free permutation transformed to the pair $(1,2)$.) Let $t = \tan\tilde{\phi}$ be the computed approximation of the tangent of Jacobi angle. Let $\tilde{f}_1 = \tilde{F}e_1$, $\tilde{f}_2 = \tilde{F}e_2$ be the pivot column pair. It can be shown that the computed columns $\tilde{f}_1^{(1)}$ and $\tilde{f}_2^{(1)}$ satisfy

$$[\tilde{f}_1^{(1)}, \tilde{f}_2^{(1)}] = [\tilde{f}_1 + \delta\tilde{f}_1, \tilde{f}_2 + \delta\tilde{f}_2]\begin{bmatrix} \cos\tilde{\phi} & \sin\tilde{\phi} \\ -\sin\tilde{\phi} & \cos\tilde{\phi} \end{bmatrix}\begin{bmatrix} 1+\varepsilon_c & 0 \\ 0 & 1+\varepsilon_c \end{bmatrix},$$

where

$$\|\delta\tilde{f}_i\|_2 \leq 5.1\varepsilon\|\tilde{f}_i\|_2, \quad i = 1,2; \quad \boldsymbol{fl}((1+t^2)^{-1/2}) = (1+t^2)^{-1/2}(1+\varepsilon_c), \quad |\varepsilon_c| \leq 2.5\varepsilon.$$

Now, write the transformed matrix as $\tilde{F}^{(1)} = \tilde{F}'((1+\varepsilon_c)I_2 \oplus I_{n-2})$, where

$$\tilde{F}' = (\tilde{F} + \delta\tilde{F})(\begin{bmatrix} \cos\tilde{\phi} & \sin\tilde{\phi} \\ -\sin\tilde{\phi} & \cos\tilde{\phi} \end{bmatrix} \oplus I_{n-2}), \quad \delta\tilde{F}e_i = \delta\tilde{f}_i, \quad i = 1,2; \quad \delta F e_i = \mathbf{0}, \; i \neq 1,2.$$

The relative distance between the singular values of $\tilde{F}^{(1)}$ and $\tilde{F}'$ is at most $|\varepsilon_c|$. It remains to compare the singular values of $\tilde{F}'$ and $\tilde{F}$. Let $\hat{F} = [\tilde{F}e_3, \ldots, \tilde{F}e_n] = \Omega D T^T$ be the SVD of $\hat{F}$ and consider the matrix

$$\tilde{F}'(I_2 \oplus T) = (\tilde{F}(I_2 \oplus T) + \delta\tilde{F})(\begin{bmatrix} \cos\tilde{\phi} & \sin\tilde{\phi} \\ -\sin\tilde{\phi} & \cos\tilde{\phi} \end{bmatrix} \oplus I_{n-2}). \tag{2.29}$$

We can equivalently consider the singular value perturbation of the matrix $\tilde{F}(I_2 \oplus T)$ to obtain the singular value variation bound (cf. Corollary 2.2)

$$\max_{1\leq i\leq n} \frac{|\delta\tilde{\sigma}_i|}{\tilde{\sigma}_i} \leq 5.1\sqrt{2}\varepsilon\|(\tilde{F}_{c_{1,1,n-2}})^\dagger\|_2, \quad \tilde{F}_{c_{1,1,n-2}} = [\, \frac{\tilde{f}_1}{\|\tilde{f}_1\|_2}, \frac{\tilde{f}_2}{\|\tilde{f}_2\|_2}, \, \Omega \,]. \tag{2.30}$$

Because of floating–point computation in (2.28), a somewhat weaker estimate

$$\max_{i,j} |\cos\angle(\tilde{F}^{(\ell)}e_i, \tilde{F}^{(\ell)}e_j)| \leq \tau(m) = \mathtt{tol} + O(m\varepsilon) \tag{2.31}$$

holds instead of (2.28), and the column norms of $\tilde{F}^{(\ell)}$ approximate its singular values with a relative error bounded by $n\tau(m)/(1-n\tau(m))$. Finally, the Euclidean norm of the ith column is computed with a relative error not larger than $\varepsilon_{\ell_2}(m)$. □

The error bound in Proposition 2.11 is similar to the one of Demmel and Veselić [17], where the matrix $\tilde{F}_c^{(k)} = \tilde{F}^{(k)}\mathrm{diag}\,(\|\tilde{F}^{(k)}e_i\|_2)^{-1}$ is used instead of $\tilde{F}_{c_{1,1,n-2}}^{(k)}$. Note, however, that

$$\|(\tilde{F}_c^{(k)})^\dagger\|_2 \leq \kappa_2(\tilde{F}_c^{(k)}) \leq \sqrt{n}\min_D \kappa_2(\tilde{F}^{(k)}D),$$

while (see [18])

$$\|(\tilde{F}_{c_{1,1,n-2}}^{(k)})^\dagger\|_2 \leq \kappa_2(\tilde{F}_{c_{1,1,n-2}}^{(k)}) \leq \sqrt{3}\min_{\Delta,T} \kappa_2(\tilde{F}^{(k)}(\Delta \oplus T)),$$

where D denotes an arbitrary $n \times n$ diagonal nonsingular matrix, Δ denotes an arbitrary 2×2 diagonal nonsingular matrix, and T is an arbitrary $(n-2) \times (n-2)$ nonsingular matrix. Hence, it is possible that $\|(\tilde{F}_{c_{1,1,n-2}}^{(k)})^\dagger\|_2 \ll \|(\tilde{F}_c^{(k)})^\dagger\|_2$.

The use of $\|(\tilde{F}_{c_{1,1,n-2}}^{(k)})^\dagger\|_2$ in Proposition 2.11 is possible because of the special structure of the error matrices $\delta\tilde{F}^{(k)}$. In the practice, the input matrix usually has initial uncertainty and the

appropriate condition number is $\|\tilde{F}_c^\dagger\|_2$. Our next goal is to estimate the condition number of the column scaled matrix F. For the sake of simplicity, we consider matrices from an exact application of Algorithm 2.6. The difference between condition number is of minor importance because it contributes only to the higher order terms in the error bound.

PROPOSITION 2.12. *Let A_c, R and F be the matrices from Algorithm 2.6 and let $\Delta_F = \mathrm{diag}\,(\|Fe_i\|_2)$, $F_c = F\Delta_F^{-1}$. Then $\|F_c^\dagger\|_2 \le \|\Delta_F R\|_2 \|A_c^\dagger\|_2$, where $\|\Delta_F R\|_2 \le \|\mathrm{diag}\,(\|R^{-1}e_i\|_1)|R|\|_2$.*

Proof. Set $Z = R^{-1}\Delta_F^{-1}$. Then $F_c = A_c\Pi Z$, and we can use the variational characterization of the minimal singular value $\sigma_{\min}(\cdot)$ to obtain

$$\sigma_{\min}(F_c) = \min_{x \neq 0} \frac{\|A_c\Pi Z x\|_2}{\|x\|_2} = \min_{y \neq 0} \frac{\|A_c\Pi y\|_2}{\|Z^{-1}y\|_2} \ge \frac{\sigma_{\min}(A_c)}{\|Z^{-1}\|_2}. \tag{2.32}$$

This implies that $\|F_c^\dagger\|_2 \le \|\Delta_F R\|_2 \|A_c^\dagger\|_2$. Furthermore, from

$$(\Delta_F)_{ii} = \|\sum_{k=1}^{i} (R^{-1})_{ki} A_c\Pi e_k\|_2 \le \sum_{k=1}^{i} |R^{-1}|_{ki}, \quad 1 \le i \le n,$$

it follows that

$$|Z^{-1}|_{ij} \le |R|_{ij} \sum_{k=1}^{i} |R^{-1}|_{ki}, \quad 1 \le i \le j \le n, \tag{2.33}$$

and, hence, $\|Z^{-1}\|_2 \le \| \, |Z^{-1}| \, \|_2 \le \|\mathrm{diag}\,(\|R^{-1}e_i\|_1)|R| \, \|_2$. □

COROLLARY 2.13. *If the matrix R in Algorithm 2.6 satisfies $R_{ii}^2 \ge \sum_{k=i}^{j} R_{kj}^2$, $1 \le i \le j \le n$, then*

$$\|F_c^\dagger\|_2 \le \sqrt{n(n+1)/2}\,\|A_c^\dagger\|_2 \|R_C^{-1}\|_1, \quad R_C = RD^{-1}, \tag{2.34}$$

where $D = \mathrm{diag}\,(|R_{ii}|)$ or $D = \mathrm{diag}\,(\|Re_i\|)$, $\|\cdot\| \in \{\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty\}$. Furthermore, it holds that

$$\|F_c^\dagger\|_2 \le \frac{\sqrt{4^{n+1} - 3n - 4}}{3} \|A_c^\dagger\|_2. \tag{2.35}$$

Proof. The inequality (2.33) implies

$$|Z^{-1}|_{ij} \le \sum_{k=1}^{i} \left(\frac{|R|_{ij}}{D_{kk}} \right) |R_C^{-1}|_{ki}, \quad 1 \le i \le j \le n, \tag{2.36}$$

which together with (2.32) gives (2.34). Furthermore, the inequalities $(\Delta_F)_{ii} \le \|R^{-1}e_i\|_1$, $1 \le i \le n$, and relations (2.25) and (2.33) imply

$$|Z^{-1}|_{ij} \le \frac{|R_{ij}|}{|R_{ii}|}(1 + \sum_{k=0}^{i-2} 2^k) \le 2^{i-1}, \quad 1 \le i \le j \le n,$$

and, hence, $\| \, |Z^{-1}| \, \|_2 \le \sqrt{4^{n+1} - 3n - 4}/3$. □

EXAMPLE 2.14. In the case $n = 2$, we have $\|F_c^\dagger\|_2 \le \sqrt{6}\|A_c^\dagger\|_2$, and in the case $n = 3$, we have $\|F_c^\dagger\|_2 \le \sqrt{27}\|A_c^\dagger\|_2$.

2.4. Modification of the tangent algorithm. The efficiency and the accuracy of Algorithm 2.6 can be improved in two ways. *First*, if $m \gg n$, we can compute the QR factorization of A, $A = Q_A R_A$, and then apply Algorithm 2.6 to the pair (R_A, B). In this way, the matrix F and its SVD are computed more efficiently. *Second*, we can compute the QR factorization with column pivoting of F, $F\Pi_F = Q_F R_F$, and then apply the Jacobi SVD algorithm to the matrix R_F^τ. In this way, we apply the accelerated Jacobi SVD algorithm (cf. [59], [17]) which converges faster than the Jacobi SVD algorithm applied to F. Moreover, we obtain sharper bounds for forward and backward errors. The new bound is based on Proposition 2.7 and its application to the sequence of Jacobi rotations (cf. [21]).

PROPOSITION 2.15. *Let the Jacobi SVD algorithm be applied to $G \in \mathbf{R}^{n\times n}$, and let $\tilde\sigma_1 \geq \cdots \geq \tilde\sigma_n$ be the singular values of G. Let the pivot columns be chosen so that one sweep ($n(n-1)/2$ rotations with different pivot pairs) can be completed in $\wp$ parallel steps, as described in Proposition 2.7. If the stopping criterion (2.28) is satisfied by $\tilde G^{(\ell)}$ in the sth sweep, then there exist a backward error δG and an orthogonal matrix U such that*

$$\tilde G^{(\ell)} = (G+\delta G)U, \quad \|(\delta G)^\tau e_i\|_2 \leq \varepsilon_J(n)\|G^\tau e_i\|_2, \quad 1 \leq i \leq n, \quad \varepsilon_J(n) \leq ((1+6\varepsilon)^{(\mathbf{s}-1)\wp} - 1),$$

If the matrix $G_r = \operatorname{diag}(\|G^\tau e_i\|_2)^{-1} G$ satisfies $\sqrt{n}\varepsilon_J(n)\|G_r^\dagger\|_2 < 1$, and if $\tilde\sigma_1' \geq \cdots \geq \tilde\sigma_n'$ are the computed values of the Euclidean norms of the columns of $\tilde G^{(\ell)}$, then

$$\max_{1\leq i\leq n} \frac{|\tilde\sigma_i' - \tilde\sigma_i|}{\tilde\sigma_i} \leq (1+\sqrt{n}\varepsilon_J(n)\|G_r^\dagger\|_2)\frac{1+\varepsilon_{\ell_2}(n)}{1-n\tau(n)} - 1,$$

where $\tau(n)$ is as in Proposition 2.11.

THEOREM 2.16. *Let the assumptions of Theorem 2.8 hold, and let $\tilde R_F$ be the computed upper triangular factor in the floating-point QR factorization of $\tilde F$. Let the Jacobi SVD algorithm be applied on $G = \tilde R_F^\tau$, and let $\tilde F^{(\ell)} = (\tilde G^{(\ell)})^\tau$, where $\tilde G^{(\ell)}$ is as in Proposition 2.15. Furthermore, let $\tilde R_{r,1} = \operatorname{diag}(\|\tilde R^{-1} e_i\|_1)\tilde R$ and let*

$$\eta(m,n) = \varepsilon_{QR}(m,n) + \varepsilon_J(n) + \varepsilon_{QR}(m,n)\varepsilon_J(n), \quad \eta(\tilde R) = \frac{\varepsilon_T(n)\beta_1(\tilde R)}{1-\varepsilon_T(n)\beta_1(\tilde R)}. \tag{2.37}$$

There exist backward perturbations δA, δB such that the diagram in Figure 1 commutes. Furthermore, it holds, for all i, that

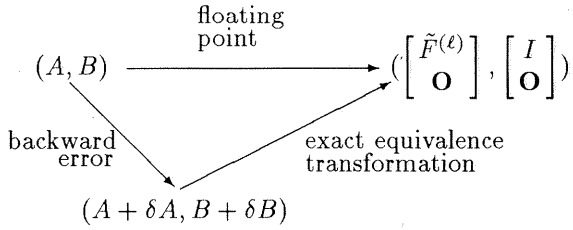


FIG. 1. *Commutative diagram for the modified algorithm.*

$$\|\delta A e_i\|_2 \leq \bar\eta_A \|A e_i\|_2, \quad \|\delta B e_i\|_2 \leq \eta_B \|B e_i\|_2, \tag{2.38}$$

where η_B is as in Theorem 2.8 and

$$\bar\eta_A = \frac{1+\varepsilon_{\ell_2}(m)}{1-\varepsilon_{\ell_2}(m)}(1+\varepsilon)\left\{\eta(\tilde R) + \eta(m,n)\|\tilde R_{r,1}\|_1\left(1+\eta(\tilde R)\right)\right\} + \varepsilon. \tag{2.39}$$

(Note that the assumption $\operatorname{rank}(A) = n$ *is not necessary for the backward error estimate in relation (2.39).) Hence, if $\sqrt{n}\bar\eta_A\|A_c^\dagger\|_2 < 1$, and if $\sigma_1 \geq \cdots \geq \sigma_n$ are the generalized singular values of*

(A,B) and if $\tilde\sigma_1' \geq \cdots \geq \tilde\sigma_n'$ are the sorted floating-point values of the Euclidean norms of the rows of $\tilde F^{(\ell)}$, then

$$\max_{1\leq i\leq n}\frac{|\tilde\sigma_i'-\sigma_i|}{\sigma_i} \leq \frac{1+\sqrt{n}\bar\eta_A\|A_c^\dagger\|_2}{1-\sqrt{n}\eta_B\|B_c^\dagger\|_2}\frac{1+\varepsilon_{\ell_2}(n)}{1-n\tau(n)}-1, \tag{2.40}$$

where $\tau(\cdot)$ is defined in Proposition 2.11.

Proof. From Proposition 2.7 and Proposition 2.15 it follows that

$$\begin{bmatrix}\tilde R_F\\ \mathbf{0}\end{bmatrix} = Q_F^T(\tilde F+\delta\tilde F),\quad \tilde F^{(\ell)} = U(\tilde R_F+\delta\tilde R_F),$$

where Q_F and U are certain orthogonal matrices and the backward errors $\delta\tilde F$ and $\delta\tilde R_F$ satisfy $\|\delta\tilde F e_i\|_2 \leq \varepsilon_{QR}(m,n)\|\tilde F e_i\|_2$, $\|\delta\tilde R_F e_i\|_2 \leq \varepsilon_J(n)\|\tilde R_F e_i\|_2$, $1\leq i\leq n$. Hence,

$$\begin{bmatrix}\tilde F^{(\ell)}\\ \mathbf{O}\end{bmatrix} = \begin{bmatrix}U & \mathbf{O}\\ \mathbf{O} & I_{m-n}\end{bmatrix} Q_F^T(\tilde F+\delta\tilde F'),\quad \delta\tilde F' = \delta\tilde F + Q_F\begin{bmatrix}\delta\tilde R_F\\ \mathbf{0}\end{bmatrix},$$

where, for all i, $\|\delta\tilde F' e_i\|_2 \leq (\varepsilon_{QR}(m,n)+\varepsilon_J(n)+\varepsilon_{QR}(m,n)\varepsilon_J(n))\|\tilde F e_i\|_2$. Furthermore, with the notation from the proof of Theorem 2.8, we have

$$\begin{bmatrix}\tilde F^{(\ell)}\\ \mathbf{O}\end{bmatrix} = \begin{bmatrix}U & \mathbf{O}\\ \mathbf{O} & I_{m-p}\end{bmatrix} Q_F^T(\tilde A_c+\delta\tilde A_c')\tilde\Pi\tilde R^{-1},\quad \delta\tilde A_c' = \mathcal{E}\tilde\Pi^T+\delta\tilde F'\tilde R\tilde\Pi^T$$

where, for all i and i' such that $\tilde\Pi^T e_i = e_{i'}$,

$$\|\delta\tilde F'\tilde R\tilde\Pi^T e_i\|_2 \leq (\varepsilon_{QR}(m,n)+\varepsilon_J(n)+\varepsilon_{QR}(m,n)\varepsilon_J(n))\sum_{k=1}^{i'}\|\tilde F e_k\|_2|\tilde R_{ki'}|.$$

Note that, for all k,

$$\|\tilde F e_k\|_2 \leq \max_{1\leq j\leq n}\|(\tilde A_c\tilde\Pi+\mathcal{E})e_j\|_2\|\tilde R^{-1}e_k\|_1 \tag{2.41}$$

$$\leq \frac{1+\varepsilon}{1-\varepsilon_{\ell_2}(m)}\left(1+\frac{\varepsilon_T(n)\beta_1(\tilde R)}{1-\varepsilon_T(n)\beta_1(n)}\right)\|\tilde R^{-1}e_k\|_1. \tag{2.42}$$

On the other hand, as in Theorem 2.8, we can write

$$\begin{bmatrix}\tilde R\\ \mathbf{O}\end{bmatrix} = Q_B^T(\tilde B_1+\delta\tilde B_1)\tilde\Pi,\quad \|\delta\tilde B_1 e_i\|_2 \leq \varepsilon_{QR}(p,n)\|\tilde B_1 e_i\|_2,\ 1\leq i\leq n.$$

where Q_B is an orthogonal matrix. Hence, the matrix $\tilde F^{(\ell)}$ is the result of the following computation:

$$\begin{aligned} A^{(1)} &= (A+\delta A_e+\delta\tilde A_c'\tilde\Delta_A)\tilde\Delta_A^{-1},\quad B^{(1)} = (B+\delta B_e+\delta\tilde B_1\tilde\Delta_A)\tilde\Delta_A^{-1},\\ \begin{bmatrix}\tilde R\\ \mathbf{O}\end{bmatrix} &= Q_B^T B^{(1)}\tilde\Pi,\\ \begin{bmatrix}\tilde F^{(\ell)}\\ \mathbf{O}_{m-n,n}\end{bmatrix} &= \begin{bmatrix}U & \mathbf{O}\\ \mathbf{O} & I_{m-n}\end{bmatrix} Q_F^T A^{(1)}\tilde\Pi\tilde R^{-1}, \end{aligned}$$

where $|\delta A_e|\leq\varepsilon|A|$, $|\delta B_e|\leq\varepsilon|B|$ and, for all i,

$$\|\delta\tilde A_c'\tilde\Delta_A e_i\|_2 \leq \frac{1+\varepsilon_{\ell_2}(m)}{1-\varepsilon_{\ell_2}(m)}(1+\varepsilon)\left\{\eta(\tilde R)+\eta(m,n)\|\tilde R_{r,1}\|_1\left(1+\eta(\tilde R)\right)\right\}\|Ae_i\|_2.$$

In other words, the matrix $\tilde{F}^{(\ell)}$ is the result of a tangent algorithm with input matrices $A+\delta A_e+\delta\tilde{A}'_c\tilde{\Delta}_A$ and $B+\delta B_e+\delta\tilde{B}_1\tilde{\Delta}_A$. □

REMARK 2.17. From the analysis in § 2.3 and from Theorem 2.16, it follows that the backward and forward error bounds for Algorithm 2.6 are nearly the same for all matrix pairs (AD_1, BD_2), where D_1 and D_2 are arbitrary diagonal nonsingular matrices. Furthermore, since $\tilde{R}_{r,1}$ is invariant under row scaling of $\tilde{R}$, the behavior of $\|\tilde{R}_{r,1}\|_1$ is similar to the behavior of $\beta_1(\tilde{R})$ (cf. Remark 2.10).

Let us state, without proof, the following interesting corollary of Theorem 2.16.

COROLLARY 2.18. *Let $A \in \mathbf{R}^{m\times n}$, $m \geq n$, and let the SVD of A be computed using the following Jacobi SVD algorithm of Veselić and Hari [59]:*

Step 1 *Compute the QR factorization with column pivoting of A, $A\Pi = Q_A \begin{bmatrix} R_A \\ \mathbf{O} \end{bmatrix}$.*

Step 2 *Apply the Jacobi SVD algorithm to the matrix $G = R_A^\tau$.*

If $G^{(\ell)}$ is defined as in Proposition 2.15, then there exist orthogonal matrices $Q \in \mathbf{R}^{m\times m}$ and $U \in \mathbf{R}^{n\times n}$ and a backward error δA such that

$$(G^{(\ell)})^\tau = \begin{bmatrix} U^\tau & \mathbf{O} \\ \mathbf{O} & I_{m-n} \end{bmatrix} Q(A+\delta A)\Pi,$$

where, for all i, $\|\delta A e_i\|_2 \leq (\varepsilon_{QR}(m,n)+\varepsilon_J(n)+\varepsilon_{QR}(m,n)\varepsilon_J(n))\|Ae_i\|_2$.

2.5. Preconditioner for the tangent algorithm. The accuracy and error estimates of Algorithm 2.6 can be further improved by preconditioning. The aim of preconditioning is to reduce $\| \, |\tilde{R}^{-1}|\cdot|\tilde{R}| \, \|_1$ by reducing $\|B_c^\dagger\|_2$ (cf. Proposition 2.9). We can reduce $\|B_c^\dagger\|_2$ using Jacobi rotations that simultaneously transform A_c and B_1 in Step 1 of Algorithm 2.6. The motivation for using Jacobi rotations is their ability to reduce the condition number of the column–scaled matrix (cf. [35], [36], [25], [17]). This property of Jacobi rotations is the crucial factor in the high accuracy of the Jacobi SVD algorithm, see [17]. We apply one sweep ($n(n-1)/2$ rotations), where the pivot pairs are chosen using de Rijk's pivot strategy, see [12]. The threshold for the application of the Jacobi rotation is set higher than the usual $O(p\varepsilon)$ value. For example, we can choose to apply Jacobi rotation only if the computed cosine of the angle between pivot columns is larger than $1/\sqrt{n}$ (say). Since our algorithm achieves high relative accuracy for moderate $\|B_c^\dagger\|_2$ (cf. Corollary 2.2), we expect that, in that case, the reduction of $\|B_c^\dagger\|_2$ is such that $\| \, |\tilde{R}^{-1}|\cdot|\tilde{R}| \, \|_1$ can be considered to be bounded by a moderate polynomial of n. This computation is especially effective if $\|B_1e_1\|_2 \gg \cdots \gg \|B_1e_n\|_2$ because in that case one sweep of Jacobi rotations with de Rijk's pivoting behaves like the modified Gram–Schmidt orthogonalization, see [21]. The new computed pair of matrices is given as input to Algorithm 2.6. If J is the product of the Jacobi rotations used to reduce the condition number of the column scaled matrix B_1, then the matrix A_c is changed to A_cJ. Let $(A_cJ)_c$ be the matrix A_cJ with columns scaled to have unit Euclidean norm. Using [54], we conclude that $\kappa_2((A_cJ)_c) \leq \sqrt{n}\kappa_2(A_c)$. Hence, the possible growth of the condition number of the column scaled matrix A_cJ is moderate.

REMARK 2.19. We can first compute the matrix $\tilde{R}$ and estimate $\beta_1(\tilde{R})$ using an $O(n^2)$ condition estimator. When the computed estimate of $\beta_1(\tilde{R})$ is larger than some given tolerance, we apply Jacobi rotations to $\tilde{R}$ instead of B_c.

REMARK 2.20. If $\|B_c^\dagger\|_2$ is so large that our analysis cannot guarantee relative accuracy, then the preconditioning step is replaced with an algorithm from [22]. The algorithm from [22] replaces (A, B) with an equivalent pair (A', B') such that the condition number of the column scaled matrix B' is moderate, and the condition number of the column scaled matrix A' is not much larger than $\kappa_2(A_c)$.

3. Eigenvalue computation of positive definite pencils. In this section, we consider the generalized eigenvalue problem

(3.43) $\qquad Hx = \lambda Mx, \quad H$ and M $n\times n$ symmetric positive definite matrices.

We present an algorithm that replaces the eigenvalue problem with the GSVD computation of the pair (L_H, L_M) of the Cholesky factors L_H, L_M of H and M, respectively. The GSVD of (L_H, L_M) is computed using an algorithm similar to Algorithm 2.6.

REMARK 3.1. In some applications, certain factors A and B of H and M, respectively, can be derived directly from the application that leads to the eigenvalue problem (3.43). In that case, we can solve the eigenvalue problem (3.43) without having to compute the matrices H and M. Sometimes, this implicit formulation is the most important step in the overall solution process, see [48], [8], [56].

In § 3.1, we analyze the sensitivity of the eigenvalues in relation (3.43). We describe the set of positive definite pencils for which the eigenvalues can be computed with high relative accuracy in floating–point arithmetic. The details of the new algorithm are given in § 3.2. In § 3.3, we analyze the numerical properties of the new algorithm and we show that it attains the optimal accuracy described in § 3.1.

3.1. Sensitivity of the eigenvalues. If H and M in (3.43) change, how do the eigenvalues of $H-\lambda M$ change? We are particularly interested in δH and δM that naturally arise in floating–point computation. Such δH, δM include element–wise perturbations that satisfy

$$|\delta H_{ij}| \leq \epsilon |H_{ij}|, \quad |\delta M_{ij}| \leq \epsilon |M_{ij}|. \tag{3.44}$$

A more general example comes from floating–point Cholesky factorization: the computed Cholesky factor $\tilde{L}_H$ of H satisfies (cf. [15], [17])

$$\tilde{L}_H \tilde{L}_H^T = H + \delta H, \ \max_{i,j} \frac{|\delta H_{ij}|}{\sqrt{H_{ii}H_{jj}}} \leq \varepsilon_C(n), \ \varepsilon_C(n) \leq (n+5)\varepsilon. \tag{3.45}$$

Using double precision accumulation, the bound for $\varepsilon_C(n)$ can be reduced to $O(1)\varepsilon$ for all $n \leq 1/\varepsilon$.

A necessary condition for relative accuracy of the eigenvalues of $H - \lambda M$ is that the positive definiteness of H and M is not changed by δH and δM. Demmel [15] shows that smallest ϵ in (3.44) such that $H + \delta H$ is singular is between $\|H_s^{-1}\|_2^{-1}/n$ and $\|H_s^{-1}\|_2^{-1}$, where $H_s = \Delta_H H \Delta_H$, $\Delta_H = \text{diag}\,(H_{ii})^{-1/2}$. Furthermore, it is shown in [15] that, in the case $\|H_s^{-1}\|_2 > 1/\varepsilon$, there exist rounding errors ($|\delta H_{ij}| \leq \varepsilon |H_{ij}|$) such that $H + \delta H$ is not positive definite, and that if $\|H_s^{-1}\|_2 < 1/(n\varepsilon_C(n))$ the Cholesky factorization is guaranteed to succeed in floating–point arithmetic. Hence, in this section we make a reasonable assumption that the matrices

$$H_s = \Delta_H H \Delta_H, \ \text{ and } \ M_s = \Delta_M M \Delta_M, \tag{3.46}$$

where $\Delta_H = \text{diag}\,(H_{ii})^{-1/2}$, $\Delta_M = \text{diag}\,(M_{ii})^{-1/2}$, have inverses bounded by a modest constant (in the spectral norm).

THEOREM 3.2. *Let H and M be positive definite and let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of $H - \lambda M$. Let δH and δM be symmetric perturbations such that $\|(\delta H)_s\|_2 \|H_s^{-1}\|_2 < 1$, $\|(\delta M)_s\|_2 \|M_s^{-1}\|_2 < 1$, where $(\delta H)_s = \Delta_H \delta H \Delta_H$, $(\delta M)_s = \Delta_M \delta M \Delta_M$. If $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$ are the eigenvalues of $(H + \delta H) - \lambda(M + \delta M)$, then*

$$\max_{1 \leq i \leq n} \frac{|\tilde{\lambda}_i - \lambda_i|}{\lambda_i} \leq \frac{\|(\delta H)_s\|_2 \|H_s^{-1}\|_2 + \|(\delta M)_s\|_2 \|M_s^{-1}\|_2}{1 - \|(\delta M)_s\|_2 \|M_s^{-1}\|_2}. \tag{3.47}$$

Proof. Let L_H, L_M be the Cholesky factors of H and M, respectively. Note that we can write

$$H + \delta H = L_H \left(\sqrt{I + L_H^{-1} \delta H L_H^{-T}} \right)^2 L_H^T, \tag{3.48}$$

because $\|L_H^{-1} \delta H L_H^{-T}\|_2 \leq \|(\delta H)_s\|_2 \|H_s^{-1}\|_2 < 1$. If we factor $M + \delta M$ in the same way, we can consider the generalized singular values of the pair

$$(\sqrt{I + L_H^{-1} \delta H L_H^{-T}}\, L_H^T, \sqrt{I + L_M^{-1} \delta M L_M^{-T}}\, L_M^T).$$

Using the proof of Theorem 2.1, we conclude that, for all i,

$$\frac{1-\|L_H^{-1}\delta H L_H^{-\tau}\|_2}{1+\|L_M^{-1}\delta M L_M^{-\tau}\|_2} \leq \frac{\tilde{\lambda}_i}{\lambda_i} \leq \frac{1+\|L_H^{-1}\delta H L_H^{-\tau}\|_2}{1-\|L_M^{-1}\delta M L_M^{-\tau}\|_2}. \tag{3.49}$$

□

If δH is as in relation (3.45) and $(\delta H)_s = \Delta_H \delta H \Delta_H$, then $\max_{i,j} |(\delta H)_s|_{ij} \leq \varepsilon_C$, and we conclude that the right–hand side in relation (3.47) is less than one if $\max\{\|H_s^{-1}\|_2, \|M_s^{-1}\|_2\} \leq 1/(3n\varepsilon_C)$. For similar results in the case of symmetric definite scaled diagonally dominant pencil $H - \lambda M$ see [6]. In the next theorem, we use the results of Veselić and Slapničar [60] to show that the bound (3.47) is sharp.

THEOREM 3.1. *Let H and M be as in Theorem 3.2, and let $\kappa > 1$. If for all $\epsilon < 1/\kappa$ and all symmetric perturbations as in (3.44) the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ and $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$ of $H - \lambda M$ and $(H + \delta H) - \lambda(M + \delta M)$, respectively, satisfy*

$$\max_{1 \leq i \leq n} \frac{|\tilde{\lambda}_i - \lambda_i|}{\lambda_i} \leq \kappa\epsilon, \tag{3.50}$$

then $\max\{\|H_s^{-1}\|_2, \|M_s^{-1}\|_2\} \leq (1+\kappa)/2$.

Proof. Let $\delta H = \mathbf{O}$ and $|\delta M_{ij}| \leq \epsilon |M_{ij}|$, $1 \leq i, j \leq n$. Then $M + \delta M$ must remain positive definite and, for all $\epsilon < 1/\kappa$, $\|M_s^{-1}\|_2 \leq (1+\epsilon)/(2\epsilon)$ (cf. [60, Lemma 2.20]). This implies $\|M_s^{-1}\|_2 \leq (1+\kappa)/2$ (cf. [60, Corollary 2.23]). Now choose $\delta M = \mathbf{O}$ and $|\delta H_{ij}| \leq \epsilon |H_{ij}|$, $1 \leq i, j \leq n$. □

3.2. The algorithm. Perturbation analysis in § 3.1 shows that floating–point Cholesky factorization is a numerically stable way to replace the eigenvalue problem (3.43) with the GSVD problem. In the following algorithm, we exploit that stable relationship by combining the Cholesky factorization with an accurate generalized singular value computation by Algorithm 2.6.

ALGORITHM 3.3.

Input Symmetric, positive definite matrices H and M.
Step 1 Compute $\Delta_H = \operatorname{diag}(H_{ii})^{-1/2}$, $H_s = \Delta_H H \Delta_H$, and $M_1 = \Delta_H M \Delta_H$.
Step 2 Compute the Cholesky factorizations $A^\tau A = H_s$, $R^\tau R = \Pi^\tau M_1 \Pi$ (with pivoting).
Step 3 Compute $F = A\Pi R^{-1}$ by solving the equation $FR = A\Pi$.
Step 4 Compute the SVD of F using thc Jacobi SVD algorithm, $\Sigma = V^\tau F U$.
Step 5 Compute $X = \Delta_H \Pi R^{-1} U$.
Output X and Σ satisfy $HX = MX\Sigma^2$.

In Step 4, we compute the SVD of F by an application of the Jacobi SVD algorithm to the matrix F^τ. In that case, the Jacobi SVD algorithm computes the SVD as $U\Sigma = F^\tau V$, where V denotes the accumulated product of Jacobi rotations and $U\Sigma$ is the limit matrix. Since for the eigenvector matrix X we need only the matrix U, there is no need to accumulate the Jacobi rotations. Since the accumulation of Jacobi rotations is about 40 percent of overall computation in the Jacobi SVD algorithm, this procedure is more efficient than the Jacobi SVD computation of the matrix F.
We also note that we can obtain an upper triangular F if we modify Step 2 and Step 3 as follows:
Step 2' Compute the Cholesky factorization with pivoting $R^\tau R = \Pi^\tau M_1 \Pi$, and then the Cholesky factorization of $\Pi^\tau H_s \Pi$, $A^\tau A = \Pi^\tau H_s \Pi$.
Step 3' Compute $F = AR^{-1}$.

3.3. Error analysis. Let now $\tilde{H}_s$ and $\tilde{M}_1$ be the computed approximations of H_s and M_1, respectively, where the diagonals of $\tilde{H}_s$ are explicitly set to one. Then there exist backward errors δH_e and δM_e such that

$$\tilde{H}_s = \Delta_H(H + \delta H_e)\Delta_H, \quad \tilde{M}_1 = \Delta_H(M + \delta M_e)\Delta_H,$$

where, for all i, j, $(\delta H_e)_{ii} = 0$ and

$$|\delta H_e|_{ij} \leq \varepsilon_1 |H_{ij}|, \quad |\delta M_e|_{ij} \leq \varepsilon_1 |M_{ij}|, \quad \varepsilon_1 \leq \frac{1+\varepsilon}{(1-\varepsilon)^{3/2}} - 1. \tag{3.51}$$

On the other hand, the computed Cholesky factors $\tilde{A}$ and $\tilde{R}$ satisfy

$$\tilde{A}^T \tilde{A} = \tilde{H}_s + \delta \tilde{H}_s, \quad |\delta \tilde{H}_s|_{ij} \leq \varepsilon_C(n), \tag{3.52}$$

$$\tilde{R}^T \tilde{R} = \Pi^T (\tilde{M}_1 + \delta \tilde{M}_1) \Pi, \quad |\delta \tilde{M}_1|_{ij} \leq \varepsilon_C(n) \sqrt{(\tilde{M}_1)_{ii} (\tilde{M}_1)_{jj}}, \tag{3.53}$$

where $\varepsilon_C(n)$ is as in relation (3.45). Hence,

$$\begin{aligned} \tilde{A}^T \tilde{A} &= \Delta_H (H + \delta H_e + \Delta_H^{-1} \delta \tilde{H}_s \Delta_H^{-1}) \Delta_H, \quad |\Delta_H^{-1} \delta \tilde{H}_s \Delta_H^{-1}|_{ij} \leq \varepsilon_C(n) \sqrt{H_{ii} H_{jj}}, \\ \Pi \tilde{R}^T \tilde{R} \Pi^T &= \Delta_H (M + \delta M_e + \Delta_H^{-1} \delta \tilde{M}_1 \Delta_H^{-1}) \Delta_H, \quad |\Delta_H^{-1} \delta \tilde{M}_1 \Delta_H^{-1}| \leq \varepsilon_C(n)(1+\varepsilon_1) \sqrt{M_{ii} M_{jj}}. \end{aligned}$$

Now, using Theorem 3.2 we obtain the following error bound.

PROPOSITION 3.4. *Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the true eigenvalues of $H - \lambda M$ and let $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$ be the squared singular values of the exact product $\tilde{A} \Pi \tilde{R}^{-1}$, where $\tilde{A}$ and $\tilde{R}$ are the computed (upper triangular) Cholesky factors in Step 2 of Algorithm 3.3. Then*

$$\max_{1 \leq i \leq n} \frac{|\tilde{\lambda}_i - \lambda_i|}{\lambda_i} \leq n \tilde{\varepsilon}_C(n) \frac{\|H_s^{-1}\|_2 + \|M_s^{-1}\|_2}{1 - n \tilde{\varepsilon}_C(n) \|M_s^{-1}\|_2}, \quad \tilde{\varepsilon}_C(n) = \varepsilon_C(n) + \varepsilon_1 + \varepsilon_C(n) \varepsilon_1, \tag{3.54}$$

where $\varepsilon_C(n)$ and ε_1 are as in relation (3.45) and (3.51), respectively. Using double precision accumulation in the Cholesky factorization, the bound (3.54) can be reduced to $O(n\varepsilon)(\|H_s^{-1}\|_2 + \|M_s^{-1}\|_2)$ for all $n \leq 1/\varepsilon$.

If we compare the bound (3.54) with the estimates in § 3.1, we conclude that the accuracy is about the best possible we can expect in floating-point computation. Next we show that the eigenvalue approximations computed in Step 3 and Step 4 of Algorithm 3.3 satisfy an analogous error bound as in Proposition 3.4.

THEOREM 3.5. *Let $\tilde{F}$ be the computed floating-point value of the matrix $\tilde{A} \Pi \tilde{R}^{-1}$ and let $\tilde{R}_{r,1} = \operatorname{diag}(\|\tilde{R}^{-1} e_i\|_1) \tilde{R}$. Furthermore, let the Jacobi SVD algorithm be applied to the matrix $G = \tilde{F}^T$ and let $\tilde{F}^{(\ell)} = (\tilde{G}^{(\ell)})^T$, where $\tilde{G}^{(\ell)}$ is as in Proposition 2.15. Let $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$ be as in Proposition 3.4, and let $\tilde{\lambda}'_1 \geq \cdots \geq \tilde{\lambda}'_n$ be the floating-point values of the squared Euclidean norms of the rows of the matrix $\tilde{F}^{(\ell)}$. Then, for all i,*

$$\frac{|\tilde{\lambda}'_i - \tilde{\lambda}_i|}{\tilde{\lambda}_i} \leq \left((1+\eta) \frac{1 + \varepsilon_{\ell_2}(n)}{1 - n\tau(n)} \right)^2 - 1, \tag{3.55}$$

where $\tau(\cdot)$ is as in Proposition 2.11 and

$$\eta = \varepsilon_T(n) \frac{\|\tilde{A}^{-1}\|_2 \| \, |\tilde{A}| \, \|_2 \beta_2(\tilde{R})}{1 - \varepsilon_T(n) \beta_2(\tilde{R})} + \sqrt{n} \|\tilde{A}^{-1}\|_2 \|\tilde{R}_{r,1}\|_1 \varepsilon_J(n) \frac{\sqrt{1 + \varepsilon_C(n)}}{1 - \varepsilon_T(n) \beta_1(\tilde{R})}$$

with $\beta_2(\tilde{R}) = \| \, |\tilde{R}^{-1}| \cdot |\tilde{R}| \, \|_2$. Furthermore, there exist symmetric backward errors δH_b and δM_b such that $\tilde{F}^{(\ell)}$ is the result of exact computation with the matrices $H + \delta H_b$ and $M + \delta M_b$, and such that

$$\max_{i,j} \frac{|\delta H_b|_{ij}}{\sqrt{H_{ii} H_{jj}}} \leq f(n)\varepsilon, \quad \max_{i,j} \frac{|\delta M_b|_{ij}}{\sqrt{M_{ii} M_{jj}}} \leq O(n\varepsilon),$$

where $f(\cdot)$ is a modestly growing function of matrix dimension.

Proof. From Proposition 2.15 and Theorem 2.8, it follows that $\tilde{F}^{(\ell)}$ is the result of exact application of the Jacobi SVD algorithm to the matrix $\tilde{F}' = \tilde{A}\Pi\tilde{R}^{-1} + \mathcal{E}\tilde{R}^{-1} + \delta\tilde{F}$, where

$$|\mathcal{E}| \leq \varepsilon_T(n)|\tilde{F}| \cdot |\tilde{R}|, \quad \|\delta\tilde{F}e_i\|_2 \leq \varepsilon_J(n)\|(\tilde{A}\Pi\tilde{R}^{-1} + \mathcal{E}\tilde{R}^{-1})e_i\|_2, \quad 1 \leq i \leq n.$$

The matrix $\tilde{F}'$ can be written as

$$\tilde{F}' = (I + \Omega)\tilde{A}\Pi\tilde{R}^{-1}, \quad \Omega = \mathcal{E}\Pi^T\tilde{A}^{-1} + \delta\tilde{F}\tilde{R}\Pi^T\tilde{A}^{-1},$$

where

$$\|\mathcal{E}\Pi^T\tilde{A}^{-1}\|_2 \leq \varepsilon_T(n)\|\tilde{A}^{-1}\|_2\| \, |\tilde{A}| \, \|_2 \frac{\| \, |\tilde{R}^{-1}| \cdot |\tilde{R}| \, \|_2}{1 - \varepsilon_T(n)\| \, |\tilde{R}^{-1}| \cdot |\tilde{R}| \, \|_2}. \tag{3.56}$$

Furthermore, using relation (3.52) and an estimate similar to (2.41), (2.42), we have

$$\begin{aligned}
\|\delta\tilde{F}\tilde{R}\Pi^T\tilde{A}^{-1}\|_F &\leq \|\tilde{A}^{-1}\|_2\sqrt{n}\max_{1\leq j\leq n}\|\delta\tilde{F}\tilde{R}e_j\|_2 \\
&\leq \|\tilde{A}^{-1}\|_2\sqrt{n}\varepsilon_J(n)\max_{1\leq j\leq n}\sum_{k=1}^{j}\|\tilde{F}e_k\|_2|\tilde{R}_{kj}| \\
&\leq \|\tilde{A}^{-1}\|_2\sqrt{n}\varepsilon_J(n)\left(1 + \frac{\varepsilon_T(n)\beta_1(\tilde{R})}{1 - \varepsilon_T(n)\beta_1(\tilde{R})}\right)\max_{1\leq j\leq n}\|\tilde{A}e_j\|_2\|\tilde{R}_{r,1}\|_1 \\
&\leq \sqrt{n}\|\tilde{A}^{-1}\|_2\|\tilde{R}_{r,1}\|_1\varepsilon_J(n)\frac{\sqrt{1+\varepsilon_C(n)}}{1 - \varepsilon_T(n)\beta_1(\tilde{R})}
\end{aligned} \tag{3.57}$$

Now, an application of Theorem 2.1 and Proposition 2.15 implies relation (3.55).

Now we show that the matrix $\tilde{F}^{(\ell)}$ is a backward stable function of H and M. In other words, there are small backward perturbations δH_b, δM_b such that $\tilde{F}^{(\ell)}$ is the result of exact computation with the matrices $H + \delta H_b$, $M + \delta M_b$. To prove this, first note that the matrix $\tilde{F}'$ can be written as

$$\tilde{F}' = (\tilde{A} + \delta\tilde{A})\Pi\tilde{R}^{-1}, \quad \delta\tilde{A} = \mathcal{E}\Pi^T + \delta\tilde{F}\tilde{R}\Pi^T,$$

where, for all i,

$$\|\delta\tilde{A}e_i\|_2 \leq \zeta_A \equiv \sqrt{1+\varepsilon_C(n)}\left(\frac{\varepsilon_T(n)\beta_1(\tilde{R})}{1 - \varepsilon_T(n)\beta_1(\tilde{R})} + \frac{\varepsilon_J(n)\|\tilde{R}_{r,1}\|_1}{1 - \varepsilon_T(n)\beta_1(\tilde{R})}\right)$$

Hence,

$$(\tilde{A} + \delta\tilde{A})^T(\tilde{A} + \delta\tilde{A}) = \tilde{H}_s + \delta\tilde{H}_s + \delta\tilde{H}'_s, \quad |\delta\tilde{H}'_s|_{ij} \leq 2\sqrt{1+\varepsilon_C(n)}\zeta_A + \zeta_A^2.$$

(Note that $\tilde{A} + \delta\tilde{A}$ is not triangular.) If we define

$$\begin{aligned}
\delta H_b &= \delta H_e + \Delta_H^{-1}\delta\tilde{H}_s\Delta_H^{-1} + \Delta_H^{-1}\delta\tilde{H}'_s\Delta_H^{-1}, \\
\delta M_b &= \delta M_e + \Delta_H^{-1}\delta\tilde{M}_1\Delta_H^{-1},
\end{aligned}$$

then, for all i, j,

$$|\delta H_b|_{ij} \leq (\varepsilon_1 + \varepsilon_C(n) + 2\sqrt{1+\varepsilon_C(n)}\zeta_A + \zeta_A^2)\sqrt{H_{ii}H_{jj}}, \quad |\delta M_b|_{ij} \leq (\varepsilon_1 + \varepsilon_C(n)(1+\varepsilon_1))\sqrt{M_{ii}M_{jj}},$$

and $\tilde{F}^{(\ell)}$ is the result of an exact computation with the matrices $H + \delta H_b$ and $M + \delta M_b$. Note that the backward errors are given element–wise and that the error in M is $O(n\varepsilon)$, while the error

in H contains an additional factor that depends on $\|\tilde{R}_{r,1}\|_2$. The discussion in § 2.3 indicates that the backward errors in H can be taken as $f(n)\epsilon$, where $f(n)$ is a moderate function of n. □

Next we show that the relative error bound in Theorem 3.5 is newer much larger than the bound in Proposition 3.4, and that it can be much smaller. For the sake of simplicity, we use the following notation: For an arbitrary matrix Y and a positive definite matrix Z we define the scaled matrices Y_c, Y_r and Z_s by

$$Y = Y_c \mathrm{diag}\,(\|Ye_i\|_2) = \mathrm{diag}\,(\|Y^T e_i\|_2) Y_r, \quad Z = \mathrm{diag}\,(Z_{ii})^{1/2} Z_s \mathrm{diag}\,(Z_{ii})^{1/2}.$$

Define $\delta H = \delta H_e + \Delta_H^{-1}\delta\tilde{H}_s\Delta_H^{-1}$, $\delta M = \delta M_e + \Delta_H^{-1}\delta\tilde{M}_1\Delta_H^{-1}$ and note that

$$\tilde{R}^T\tilde{R} = (\Pi^T\Delta_H\Pi)\Pi^T(M+\delta M)\Pi(\Pi^T\Delta_H\Pi) \tag{3.58}$$

and

$$\|(\tilde{R}^T\tilde{R})_s^{-1}\|_2 = \|\tilde{R}_c^{-1}\|_2^2 = \|(M+\delta M)_s^{-1}\|_2. \tag{3.59}$$

Furthermore, since $\tilde{R}$ is computed with complete pivoting, we can use relation (2.23) and Proposition 2.9 to conclude that $\| \, |\tilde{R}_r^{-1}| \, \|_2 \leq \sqrt{n}\| \, |\tilde{R}_c^{-1}| \, \|_2$. On the other hand, it holds that

$$\| \, |\tilde{R}^{-1}| \cdot |\tilde{R}| \, \|_2 = \| \, |\tilde{R}_r^{-1}| \cdot |\tilde{R}_r| \, \|_2 \leq n\| \, |\tilde{R}_c^{-1}| \, \|_2 \leq n^{3/2}\sqrt{\|(M+\delta M)_s^{-1}\|_2}. \tag{3.60}$$

Hence, we can bound $\|\mathcal{E}\Pi^T\tilde{A}^{-1}\|_2$ in relation (3.56) by

$$\|\mathcal{E}\tilde{\Pi}^T\tilde{A}^{-1}\|_2 \leq n^2 \frac{\epsilon_T(n)\sqrt{\|(H+\delta H)_s^{-1}\|_2\|(M+\delta M)_s^{-1}\|_2}}{1-\epsilon_T(n)\| \, |\tilde{R}^{-1}| \cdot |\tilde{R}| \, \|_2}. \tag{3.61}$$

Similarly, since $\|\tilde{R}_{r,1}\|_1 \leq n^2\|\tilde{R}_c^{-1}\|_2$, the quantity $\|\delta\tilde{F}\tilde{R}\Pi^T\tilde{A}^{-1}\|_F$ in relation (3.57) can be bounded by

$$\|\delta\tilde{F}\tilde{R}\Pi^T\tilde{A}^{-1}\|_F \leq n^{5/2}\epsilon_J(n)\sqrt{1+\epsilon_C(n)}\frac{\sqrt{\|(H+\delta H)_s^{-1}\|_2\|(M+\delta M)_s^{-1}\|_2}}{1-\epsilon_T(n)\| \, |\tilde{R}^{-1}| \cdot |\tilde{R}| \, \|_1}. \tag{3.62}$$

Note that in Algorithm 3.3 we can use preconditioning as described in § 2.5. Also, note that we can first compute the Cholesky factorization of M_1 without pivoting, and then compute the strong rank–revealing QR factorization [31] of the computed Cholesky factor. In this way, we can improve the bounds for $\beta_1(\tilde{R})$ and $\|\tilde{R}_{r,1}\|_1$.

As a corollary of Theorem 2.16, we obtain the following strong element–wise backward stability of a variant of the Jacobi algorithm for eigenvalue computation of symmetric positive definite matrices.

COROLLARY 3.6. *Let H be an $n \times n$ symmetric and positive definite matrix, and let the eigenvalues of H be computed by the accelerated Jacobi algorithm (cf. [59], [17]):*
Step 1 *Compute the Cholesky factorization with complete pivoting, $\Pi^T H\Pi = GG^T$.*
Step 2 *Compute the SVD of G using the Jacobi SVD algorithm.*
Let the Cholesky factorization complete without breakdown, and let $\tilde{G}^{(\ell)}$ be as in Proposition 2.15. Then there exists a backward perturbation δH such that $\tilde{G}^{(\ell)}(\tilde{G}^{(\ell)})^T = \Pi^T(H+\delta H)\Pi$, where

$$\max_{i,j}\frac{|\delta H_{ij}|}{\sqrt{H_{ii}H_{jj}}} \leq \epsilon_C(n) + 2\epsilon_J(n)(1+\epsilon_C(n)) + (\epsilon_J(n)(1+\epsilon_C(n)))^2 \approx O(n\epsilon).$$

Hence, the accelerated Jacobi algorithm computes the eigenvalues of a symmetric positive definite matrix with a symmetric backward error δH such that $\max_{i,j}|\delta H_{ij}|/\sqrt{H_{ii}H_{jj}} \leq O(n\epsilon)$.

4. Other tangent algorithms. There are two algorithms in the literature that influenced the development of Algorithm 2.6. The first one is given by Lawson and Hanson [38] and it is also mentioned by Van Loan in [56].

ALGORITHM 4.1.

Step 1 Compute the QR factorization with column pivoting, $\begin{bmatrix} R \\ \mathbf{O} \end{bmatrix} = Q^T B\Pi$.

Step 2 Compute $F_{LH} = A\Pi R^{-1}$ and the SVD of F_{LH}, $\begin{bmatrix} \Sigma \\ \mathbf{O} \end{bmatrix} = V^T F_{LH} U$.

Step 3 Compute $X = \Pi R^{-1} U$, $W = Q \begin{bmatrix} U & \mathbf{O} \\ \mathbf{O} & I_{p-n} \end{bmatrix}$.

The difference between Algorithm 2.6 and Algorithm 4.1 is is illustrated in the following example.

EXAMPLE 4.2. Let α be small parameter, $|\alpha| < \varepsilon/2$, and let

$$A = \begin{bmatrix} 1 & \alpha \\ 2 & -\alpha \end{bmatrix}, \quad B = R = \begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix}, \quad (\|B\|_F \|B^{-1}\|_F = 3).$$

Then the matrices F_{LH} and F, computed by Algorithm 4.1 and Algorithm 2.6, respectively, are

$$F_{LH} = AR^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} + \alpha \\ 1 & 1 - \alpha \end{bmatrix}, \quad F = \begin{bmatrix} \frac{-\alpha}{\sqrt{2}} & \frac{1+\alpha}{\sqrt{2}} \\ \frac{\alpha}{\sqrt{2}} & \frac{2-\alpha}{\sqrt{2}} \end{bmatrix}.$$

Note that rounding errors of order $|\alpha|$ can make the matrix F_{LH} exactly singular, and that the floating–point approximation of F_{LH} contains no information about α. On the other hand, the matrix F defines the generalized singular values of (A, B) as well as the pair (A, B) does.

The second algorithm is a variant of an algorithm of Martin and Wilkinson [40], described by Barlow and Demmel [6]. Barlow and Demmel consider a scaled diagonally dominant pencil $H - \lambda M$, where M is positive definite. More precisely, $H = \Delta_H(J_H + N_H)\Delta_H$, $M = \Delta_M(I + N_M)\Delta_M$, where Δ_H and Δ_M are diagonal scalings, $|J_H| = I$, N_H and N_M have zero diagonals and $\|N_H\|_2, \|N_M\|_2 \ll 1$.

ALGORITHM 4.3.

Step 1 Compute $\Delta_M = \text{diag}\,(M_{ii})^{-1/2}$, $H_1 = \Delta_H H \Delta_H$, $M_S = \Delta_M M \Delta_M$.

Step 2 Compute $H_2 = \Pi^T H_1 \Pi$, $M_2 = \Pi^T M_1 \Pi$, where Π is permutation such that $|H_2|_{11} \leq \cdots \leq |H_2|_{nn}$.

Step 3 Compute the Cholesky factorization $R^T R = M_2$.

Step 4 Compute $K = R^{-T} H_2 R^{-1}$ and the eigendecomposition of K.

It is shown in [6] that, for sufficiently small $\|N_M\|_2$, the matrix K is also scaled diagonally dominant. However, the condition on $\|N_M\|_2$ is so restrictive that $I+N_M$ has to be almost diagonal. Algorithm 4.3 is also analyzed by Wang [61], where both H and M are positive definite.

5. Numerical examples. Before we start with presentation of results of extensive numerical testing of tangent algorithm software, we briefly discuss some issues of reliable implementation. In the first step of Algorithm 2.6, we first simulate scaling by checking the range of $\|Be_i\|_2/\|Ae_i\|_2$, $i = 1, \ldots, n$. If necessary, B is implicitly replaced with βB, where a constant β is chosen to prevent overflow and to avoid underflow without causing overflow. Each column Be_i of B is scaled by $\beta/\|Ae_i\|_2$, using the LAPACK's [2] `SLASCL()` procedure. If the ith column of A is zero, the corresponding column of B is not scaled. The QR factorization is performed by the LAPACK's `SGEQPF()` procedure. The columns of B that correspond to zero columns of A (if any) are permuted to the front of the array B, the remaining columns of B are *free* columns (cf. [2]). Instead of `SGEQPF()` we can use, for example, the strong rank–revealing QR factorization of Gu and Eisenstat [31]. For more reliable QR factorization one can use Householder reflectors with row pivoting or Givens QR factorization with careful implementation of plane rotations. This is important if the matrix has differently scaled rows. For more details see [47], [38]. For triangular system solution we use the Level 3 BLAS [19] procedure `STRSM()`. Instead of `STRSM()`

we can use the SLATRS() procedure [1]. Jacobi SVD computation is implemented as in [17], [20]. Here we note that the floating–point Jacobi SVD computation without modifications from [20] generally depends on $\kappa_2(\Delta_F)$, if $\kappa_2(\Delta_F)$ is sufficiently large (of the order of overflow threshold). The following example illustrates this.

EXAMPLE 5.1. Let $A = [a_1, a_2]$, $\|a_1\|_2 = 10^{10}$, $\|a_2\|_2 = 10^{-10}$, $(a_1, a_2) = 1/2$, and let $B = \mathrm{diag}\,(10^{-10}, 10^{10})$. In this simple case, we have $F = [f_1, f_2]$ with $\|f_1\|_2 = 10^{20}$ and $\|f_2\|_2 = 10^{-20}$. The Jacobi rotation that orthogonalizes f_1 and f_2 is computed by

$$\zeta \equiv \cot 2\phi = \frac{\|f_2\|_2^2 - \|f_1\|_2^2}{2(f_1, f_2)}, \quad \tan\phi = \frac{\mathrm{sign}\,\zeta}{|\zeta| + \sqrt{1+\zeta^2}}.$$

Thus, $\cot 2\phi \approx -10^{40}$, $\tan\phi \approx -10^{-40}/2$. Hence, $\boldsymbol{fl}(\cot 2\phi)$ overflows and $\boldsymbol{fl}(\tan\phi)$ underflows. Using the denormalized value $\boldsymbol{fl}(\tan\phi)$ does not fit into the error analyses [17], [50], [41]. Moreover, if $\boldsymbol{fl}(\tan\phi) = 0$ due to underflow, then the Jacobi procedure may not converge. Note that a small Jacobi angle ϕ does not necessarily mean that pivot columns are nearly orthogonal.
We test the accuracy properties only of Algorithm 2.6. Testing of Algorithm 3.3 for diagonalization of positive definite pencils is not presented because Algorithm 3.3 is based on Algorithm 2.6. We do not use the preconditioning from § 2.5.

5.1. Test matrix generation. We generate random matrices A_c and B_c with given $\kappa_2(A_c)$ and $\kappa_2(B_c)$, and apply scalings $A = A_c\Delta_A$, $B = B_c\Delta_B$, where Δ_A, Δ_B are random diagonal, nonsingular with given spectral condition numbers. The 4–tuple $(\kappa_2(A_c), \kappa_2(\Delta_A), \kappa_2(B_c), \kappa_2(\Delta_B))$ is chosen from the set

$$\mathcal{C} = \left\{\kappa_{ijkl} = (10^i, 10^j, 10^k, 10^l) \;:\; (i,j,k,l) \in \mathcal{I}\times\mathcal{J}\times\mathcal{K}\times\mathcal{L} \subset \mathbf{N}^4\right\},$$

where $\mathcal{I}, \mathcal{J}, \mathcal{K}, \mathcal{L}$ are determined at the very beginning of the test and kept fixed. For each fixed κ_{ijkl}, we generate a set of test pairs, using the LAPACK's DLATM1 procedure [16] as follows. We let the 4–tuple $(\mu_{i'}, \mu_{j'}, \mu_{k'}, \mu_{l'})$ of distributions of the singular values of $(A_c, \Delta_A, B_c, \Delta_B)$ take all values from the set

$$\mathcal{M} = \{\mu_{i'j'k'l'} = (\mu_{i'}, \mu_{j'}, \mu_{k'}, \mu_{l'})\} \subseteq \mathcal{P}_1\times\mathcal{P}_2\times\mathcal{P}_3\times\mathcal{P}_4 \subseteq \{\pm 1, \ldots, \pm 6\}^4,$$

where the sets of indices $\mathcal{P}_1, \ldots, \mathcal{P}_4$ contain admissible values of parameter MODE in the DLATM1 procedure. For each fixed $(\kappa_{ijkl}, \mu_{i'j'k'l'})$ we generate random pairs using random number generators with distributions chosen from the set $\mathcal{R} \subseteq \{\mathcal{U}(-1,1), \mathcal{U}(0,1), \mathcal{N}(0,1)\}$. For each fixed distribution $\chi \in \mathcal{R}$ we generate a set $\mathcal{E}^{\chi}_{\kappa_{ijkl},\mu_{i'j'k'l'}}$ of different pairs, with the cardinality of $\mathcal{E}^{\chi}_{\kappa_{ijkl},\mu_{i'j'k'l'}}$ being fixed at the beginning of the test. This makes a total of

$$\tau \equiv |\mathcal{I}|\,|\mathcal{J}|\,|\mathcal{K}|\,|\mathcal{L}|\,|\mathcal{M}|$$

different classes and $\tau\prod_{\chi\in\mathcal{R}}|\mathcal{E}^{\chi}_{\kappa_{ijkl},\mu_{i'j'k'l'}}|$ different pairs. Each test pair is generated in double precision and its generalized singular values are computed using a double precision procedure. The generalized singular values computed by double precision procedure are then taken as reference for single precision procedure that runs on original pair rounded to single precision.
A random matrix $A = A_c\Delta_A$ with given $\kappa_2(A_c)$ and $\kappa_2(\Delta_A)$ is generated using the following algorithm. (Cf. [29, P.8.5.3 and P.8.5.4], [17], [50].)

ALGORITHM 5.2.
1. $A :=$ diag (a_{ii}), where $a_{11}, \ldots, a_{nn}$ are generated using DLATM1() with parameters chosen in accordance with the current node in $\mathcal{C}\times\mathcal{M}\times\mathcal{R}$.
2. $A := \cdots(U_i(\cdots(U_1AV_1)\cdots)V_j)\cdots$, where U_i, V_j are random plane rotations.
3. $A := \cdots((\cdots(AW_1)\cdots)W_k)\cdots$, where W_k, $k = 1, \ldots$ are plane rotations designed to equilibrate columns of A. On output, all columns of A have about the same Euclidean length.
4. The diagonal matrix Δ_A is generated by DLATM1() with parameters chosen accordingly.
5. $A := A\Delta_A$.

Before the GSVD computation, both $\kappa_2(A_c)$ and $\kappa_2(B_c)$ are computed using the singular values computed by LAPACK's `SGESVD()` procedure applied to A_c, B_c, respectively. The computed values are used in theoretical estimates for the relative error in the computed singular values. The computed condition numbers are comparable with the desired values in κ_{ijkl}.

5.2. Test results. The tests in Example 5.3 and Example 5.4 were done on an Intel 486DX processor. We used Microsoft Fortran Power–Station with the *improve floating–point consistency* compiler option. The computation in Example 5.5 was done using Sun Fortran on a SUN SPARC 20 workstation.

EXAMPLE 5.3. In this example, we use LAPACK's `DGGSVD()` procedure [2] as reference for testing our procedure `SGGSVT()`. The input parameters for the test are

$$\begin{aligned} \mathcal{I} &= \{2,3,5,7\},\ \mathcal{J} = \{3,4,7,8\},\ \mathcal{K} = \{2,4,6,7\},\ \mathcal{L} = \{2,3,6,7\}, \\ \mathcal{M} &= \{(5,4,-5,3),\ (3,-4,5,-3)\},\ \mathcal{R} = \{\mathcal{U}(-1,1), \mathcal{U}(0,1), \mathcal{N}(0,1)\}. \end{aligned}$$

In Figure 2, we display the quotients between the computed spectral condition numbers of the generated matrices A_c, B_c and the desired values of $\kappa_2(A_c)$, $\kappa_2(B_c)$. All values are in the interval $(0.383, 11.277)$. This figure confirms that test pairs with desired condition numbers are used.

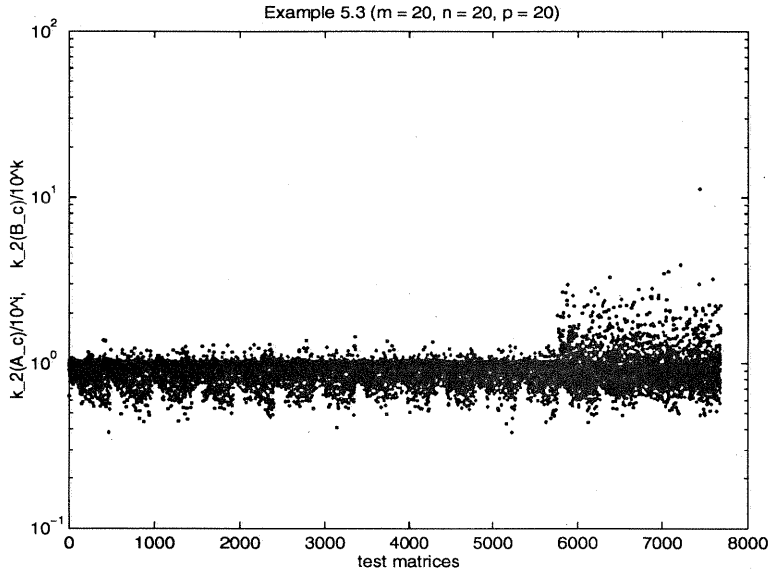


FIG. 2. *The values of $\kappa_2(A_c)/10^i$, $\kappa_2(B_c)/10^k$ for all generated test matrices.*

For each fixed point of condition×distribution mesh $\mathcal{C} \times \mathcal{M}$ we compute

$$\epsilon(\kappa_{i,j,k,l}, \mu_{i',j',k',l'}) = \max_{(A,B) \in \bigcup_{\chi \in \mathcal{R}} \mathcal{E}^{\chi}_{\kappa_{ijkl}, \mu_{i'j'k'l'}}} \frac{\max_{\sigma \in \sigma(A,B)} \frac{|\delta\sigma|}{\sigma}}{\max\{\kappa_2(A_c), \kappa_2(B_c)\}}.$$

Our analysis predicts the values of $\epsilon(\cdot)$ of the order of single precision roundoff unit times a moderate function of the dimensions. The computed values of $\epsilon(\kappa_{i,j,k,l}, \mu_{i',j',k',l'})$ are given in Figure 3.

An increase of $\kappa_2(\Delta_A)$, $\kappa_2(\Delta_B)$ with fixed $\kappa_2(A_c)$, $\kappa_2(B_c)$ does not impact the accuracy of `SGGSVT()`. This property is not shared by `DGGSVD()`. While `SGGSVT()` runs in accordance with error estimates from § 2.3 for the values of $\kappa_2(\Delta_A)$, $\kappa_2(\Delta_B)$ as large as 10^{16}, the double precision procedure `DGGSVD()` often returns inaccurate generalized singular values, as $\kappa_2(\Delta_A)$, $\kappa_2(\Delta_B)$

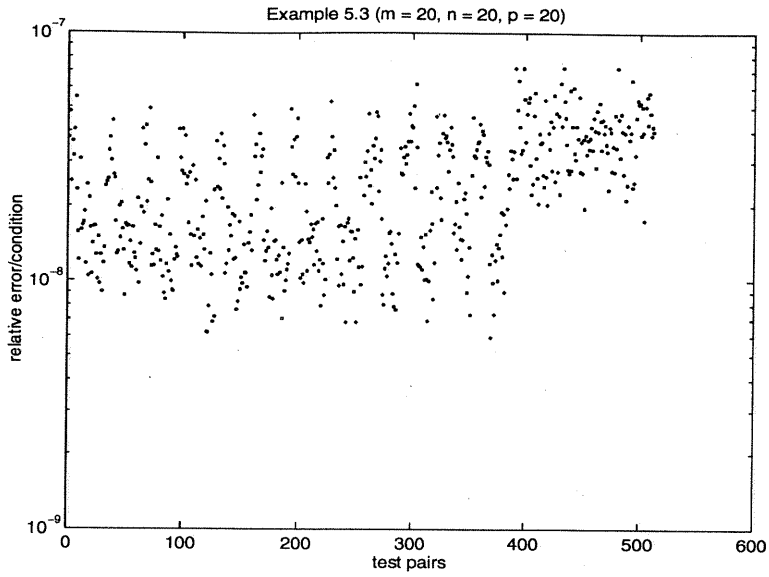


FIG. 3. *The values of* $\epsilon(\kappa_{i,j,k,l}, \mu_{i',j',k',l'})$. *Maximal value of* $\epsilon(\cdot)$ *is less than* $7.25 \cdot 10^{-8}$. *The values of* $\epsilon(\cdot)$ *below* 10^{-8} *indicate an overestimate of the condition number.*

approach 10^{16}. Therefore, in our next example, we use a double precision implementation of Algorithm 2.6 (`DGGSVT()`) as reference for `SGGSVT()`.

EXAMPLE 5.4. In this example, the input parameters are

$$\begin{aligned} \mathcal{I} &= \{1,\ldots,7\},\ \mathcal{K}=\mathcal{I}, \\ \mathcal{J} &= \{2,4,6,8,10,12,14,16\},\ \mathcal{L}=\{3,5,7,9,11,13,14,16\}, \\ \mathcal{M} &= \{(5,4,-5,3),(3,-4,5,-3),(4,5,3,-4)\},\ \mathcal{R}=\{\mathcal{U}(-1,1),\mathcal{U}(0,1),\mathcal{N}(0,1)\}. \end{aligned}$$

For each node of $\mathcal{C} \times \mathcal{M} \times \mathcal{R}$ we perform three tests on randomly generated pairs. This makes a total of 84672 test pairs. In Figure 4, we display, in $\log_{10}$ scale, the values of

$$\varepsilon(i,k) = \max_{\mathcal{J},\mathcal{L}} \max_{\mathcal{M}} \max_{(A,B)\in\bigcup_{\chi\in\mathcal{R}} \mathcal{E}^{\chi}_{\kappa_{ijkl},\mu_{i'j'k'l'}}} \max_{\sigma\in\sigma(A,B)} \frac{|\delta\sigma|}{\sigma}, \quad (i,k)\in\mathcal{I}\times\mathcal{K}.$$

Note that the relative accuracy of the computed generalized singular values depends on

$$\max\{\kappa_2(A_c), \kappa_2(B_c)\}$$

and not on $\kappa_2(\Delta_A)$, $\kappa_2(\Delta_B)$. Hence, the approximate number of correct digits in the computed generalized singular values is roughly $-\log_{10}\varepsilon(i,k) \approx 7 - \max\{i,k\}$. The behavior of the actual error in the computed values indicates that the error bound in Corollary 2.2 is sharp.

EXAMPLE 5.5. In this example, we take $m = p = 200$ and $n = 100$ and

$$\begin{aligned} \mathcal{I} &= \{2,\ldots,6\},\ \mathcal{K}=\mathcal{I}, \\ \mathcal{J} &= \{2,4,6,8,10,12,14,16\},\ \mathcal{L}=\mathcal{J}, \\ \mathcal{M} &= \{(5,4,-5,3),(3,-4,5,-3),(-5,5,3,-5)\},\ \mathcal{R}=\{\mathcal{U}(-1,1)\}. \end{aligned}$$

For each combination of these parameters, we generate one test pair. This makes the total of 4800 test pairs. We monitor the size of the backward error in the columns of A by evaluating the value of $\beta_1(\tilde{R}) = \| \, |\tilde{R}^{-1}| \cdot |\tilde{R}| \, \|_1$. The computed values of $\beta_1(\tilde{R})$ and the maximal relative errors in the computed generalized singular values are given in Figure 5. (Note that $\beta_1(\tilde{R})$ can be efficiently

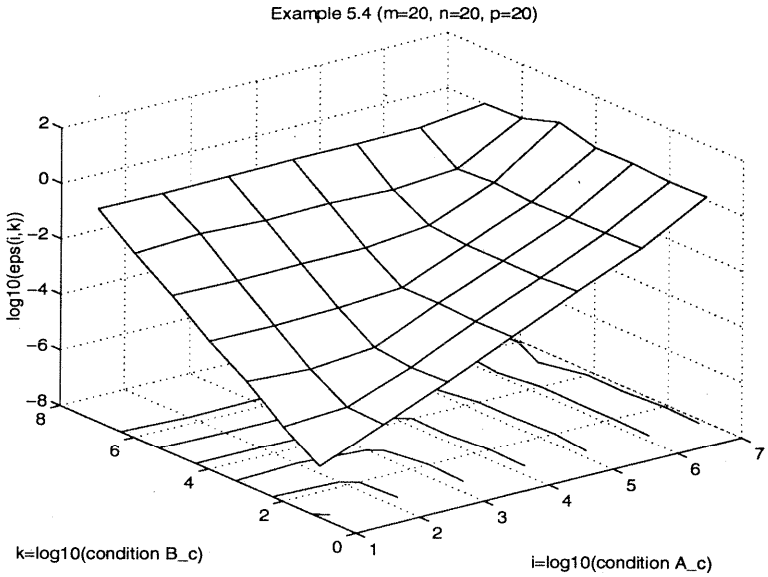


FIG. 4. *The logarithms of the maximal relative errors in the computed generalized singular values* ($-\log_{10}\varepsilon(i,k) \approx$ *the minimal number of correct digits for all test pairs* (A,B) *with* $\kappa_2(A_c) = 10^i$, $\kappa_2(B_c) = 10^k$).

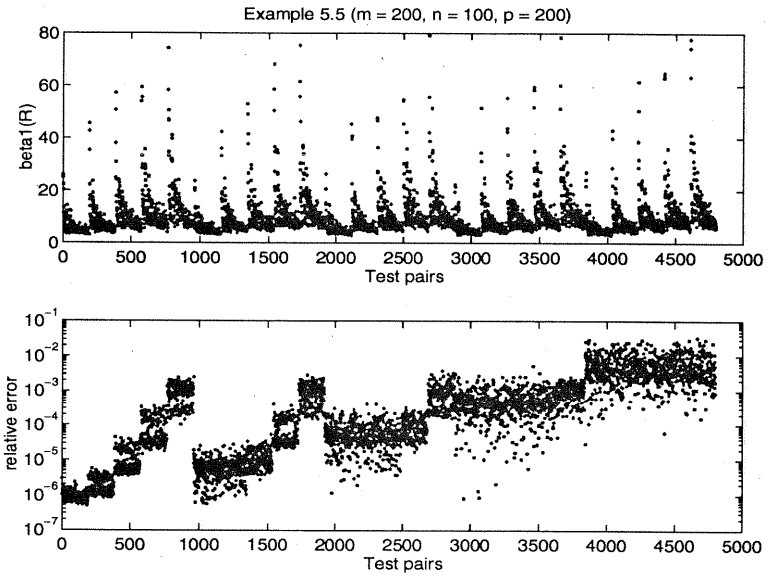


FIG. 5. *The values of* $\beta_1(\tilde{R})$ *and the maximal relative errors in the computed generalized singular values for all test pairs.*

estimated using the procedure `SPOCON()` from LAPACK.) In Figure 6, we display the values of $\log_{10}\varepsilon(i,k)$.

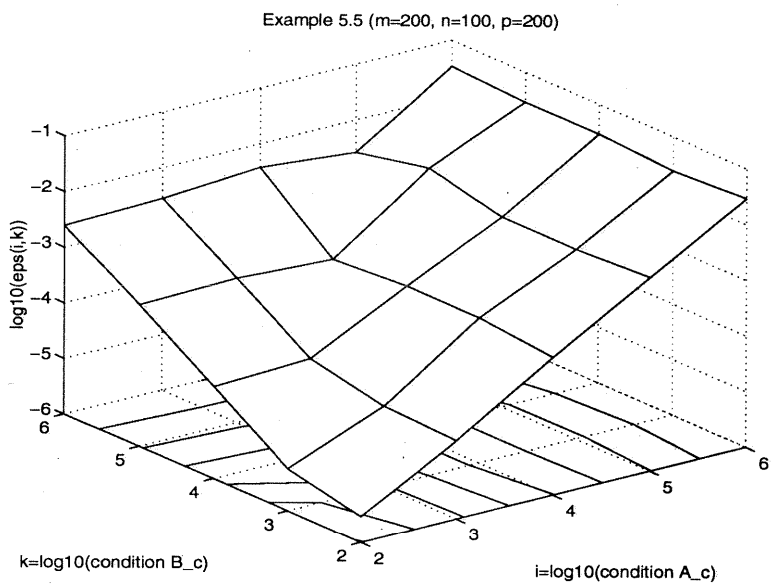


FIG. 6. *The logarithms of the maximal relative errors in the computed generalized singular values* ($-\log_{10}\varepsilon(i,k) \approx$ *the minimal number of correct digits for all test pairs* (A,B) *with* $\kappa_2(A_c) = 10^i$, $\kappa_2(B_c) = 10^k$).

Acknowledgment. The material present in this paper is a part of my Ph.D. thesis [21], written under the supervision of Professor K. Veselić at the Department of Mathematics, University of Hagen. I thank Professor K. Veselić for many valuable discussions. I also thank Professor J. Barlow, Professor J. Demmel, Professor V. Hari, Professor E. R. Jessup, Dr. E. Pietzsch, and Dr. I. Slapničar for many valuable comments. Finally, I am thankful to anonymous referees for their constructive criticism.

REFERENCES

[1] E. Anderson, *Robust triangular solvers for use in condition estimation*, LAPACK Working Note 36, University of Tennessee, Computer Science Department, August 1991.

[2] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. D. Croz, A. Greenbaum, S. Hammarling, A. McKenny, S. Ostrouchov, and D. Sorensen, *LAPACK users' guide, second edition*, SIAM, 1992.

[3] Z. Bai, *The CSD, GSVD, their applications and computations*, Technical Report 958, IMA Preprint Series, April 1992.

[4] Z. Bai and J. Demmel, *Computing the generalized singular value decomposition*, LAPACK Working Note 46, University of Tennessee, Computer Science Department, May 1992.

[5] J. Barlow, *Stability analysis of the G-algorithm and a note on its application to sparse least squares problems*, BIT, 25 (1985), pp. 507–520.

[6] J. Barlow and J. Demmel, *Computing accurate eigensystems of scaled diagonally dominant matrices*, SIAM J. Num. Anal., 27 (1990), pp. 762–791.

[7] Å. Björk, *Numerical Methods for Least Squares Problems*, SIAM, 1996.

[8] O. E. Brönlund, *Computation of the Cholesky factor of a stiffness matrix direct from the factor of its initial quadratic form*, ISD Report 142, Institut für Statik und Dinamik der Luft- und Raumfahrtkonstruktionen, Universität Stuttgart, May 1973.

[9] P. A. Businger and G. H. Golub, *Linear least squares solutions by Householder transformations*, Numer. Math., 7 (1965), pp. 269–276.

[10] K.-W. E. Chu, *Singular value and generalized singular value decompositions and the solution of linear matrix equations*, Linear Algebra Appl., 88 (1987), pp. 83–98.

[11] C. Davis and W. M. Kahan, *The rotation of eigenvectors by perturbation III*, SIAM J. Num. Anal., 7 (1970), pp. 1–46.

[12] P. P. M. de Rijk, *A one-sided Jacobi algorithm for computing the singular value decomposition on a vector computer*, SIAM J. Sci. Stat. Comp., 10 (1989), pp. 359–371.

[13] A. Deichmöller, *Über die Berechnung verallgemeinerter singulärer Werte mittels Jacobi-ähnlicher Verfahren*, PhD thesis, Lehrgebiet Mathematische Physik, Fernuniversität Hagen, 1991.

[14] A. Deichmöller and K. Veselić, *Two algorithms for computing the symmetric positive definite generalized eigenvalue problem and the generalized singular values of full column rank matrices.* Preprint, LG Mathematische Physik, Fernuniversität Hagen, D-58084 Hagen, 1991.

[15] J. Demmel, *On floating point errors in Cholesky*, LAPACK Working Note 14, Computer Science Department, University of Tennessee, October 1989.

[16] J. Demmel and A. McKenney, *A test matrix generation suite*, LAPACK Working Note 9, Courant Institute, New York, March 1989.

[17] J. Demmel and K. Veselić, *Jacobi's method is more accurate than QR*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 1204–1245.

[18] J. W. Demmel, *A Numerical Analyst's Jordan Canonical Form*, PhD thesis, Center for Pure and Applied Mathematics, University of California, Berkeley, 1983.

[19] J. J. Dongarra, J. J. D. Croz, I. Duff, and S. Hammarling, *A set of Level 3 Basic Linear Algebra Subprograms*, ACM Trans. Math. Soft., (1990), pp. 1–17.

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

[21] ———, *Computing the Singular and the Generalized Singular Values*, PhD thesis, Lehrgebiet Mathematische Physik, Fernuniversität Hagen, 1994.

[22] Z. Drmač and E. R. Jessup, *On accurate generalized singular value computation in floating point arithmetic.* University of Colorado at Boulder, Technical Report CU-CS-811-96, submitted to SIAM J. Matrix Anal. Appl., October 1996.

[23] D. K. Faddeev, V. N. Kublanovskaya, and V. N. Faddeeva, *Solution of linear algebraic systems with rectangular matrices*, Proc. Steklov Inst. Math., 96 (1968), pp. 93–111.

[24] ———, *Sur les systemes linearies algebriques de matrices rectangularies et mal-conditionnees*, in Programmation en Mathematiques Numeriques, Editions Centre Nat. Recherche Sci., Paris, VII, 1968, pp. 161–170.

[25] S. Falk, *The stabilization of poorly conditioned systems of linear algebraic equations using Jacobi's method*, Zh. vych. mat., 3 (1963), pp. 358–361.

[26] S. Falk and P. Langemeyer, *Das Jacobische Rotationsverfahren für reellsymmetrische Matrizenpaare I,*

II, Elektronische Datenverarbeitung, (1960), pp. 30–43.
[27] W. M. GENTLEMAN, *Error analysis of QR decompositions by Givens transformations*, Linear Algebra Appl., 10 (1975), pp. 189–197.
[28] S. K. GODUNOV, A. G. ANTONOV, O. P. KIRILYUK, AND V. I. KOSTIN, *Garantirovannaya tochnost resheniya sistem lineĭnykh uravneniĭ v evklidovykh prostranstvakh*, Novosibirsk Nauka, Sibirskoe Otdelenie, 1988.
[29] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations, second edition*, The Johns Hopkins University Press, 1989.
[30] G. GOSE, *Das Jacobi–Verfahren für $Ax = \lambda Bx$*, ZAMM, 59 (1979), pp. 93–101.
[31] M. GU AND S. EISENSTAT, *An efficient algorithm for computing a strong rank-revealing QR factorization*, SIAM J. Sci. Comput., 17 (1996), pp. 848 – 869.
[32] N. J. HIGHAM, *Accuracy and Stability of Numerical Algorithms*, SIAM, 1996.
[33] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, 1990.
[34] ———, *Topics in Matrix Analysis*, Cambridge University Press, 1991.
[35] C. G. J. JACOBI, *Über eine neue Auflösungsart der bei der Methode der kleinsten Quadrate vorkommenden lineären Gleichungen*, Astronomische Nachrichten, 22 (1845), pp. 297–306.
[36] ———, *Über ein leichtes Verfahren die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen*, Crelle's Journal für reine und angew. Math., 30 (1846), pp. 51–95.
[37] B. KÅGSTRÖM, *The generalized singular value decomposition and the general $A - \lambda B$- problem*, BIT, 24 (1984), pp. 568–583.
[38] C. L. LAWSON AND R. J. HANSON, *Solving Least Squares Problems*, Prentice–Hall Inc., Englewood Cliffs, N. J., 1974.
[39] R.-C. LI, *Bounds on perturbations of generalized singular values and of associated subspaces*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 195–234.
[40] R. S. MARTIN AND J. H. WILKINSON, *Reduction of the symmetric problem $Ax = \lambda Bx$ and related problems to standard form*, Numer. Math., 11 (1968), pp. 99–110.
[41] R. MATHIAS, *Accurate eigensystem computations by Jacobi methods*, SIAM J. Matrix Anal. Appl., 16 (1996), pp. 977–1003.
[42] J. J. MODI AND M. R. B. CLARKE, *An alternative Givens ordering*, Numer. Math., 43 (1984), pp. 83–90.
[43] C. C. PAIGE, *A note on a result of Sun Ji-guang: Sensitivity of the CS and GSV decompositions*, SIAM J. Num. Anal., 21 (1984), pp. 186–191.
[44] ———, *The general linear model and the generalized singular value decomposition*, Linear Algebra Appl., 70 (1985), pp. 269–284.
[45] ———, *Computing the generalized singular value decomposition*, SIAM J. Sci. Stat. Comp., 7 (1986), pp. 1126–1146.
[46] C. C. PAIGE AND M. A. SAUNDERS, *Towards a generalized singular value decomposition*, SIAM J. Num. Anal., 18 (1981), pp. 398–405.
[47] M. J. D. POWELL AND J. K. REID, *On applying Householder transformations to linear least squares problems*, in International Federation of Information Processing Congress, 1968, pp. 122–126.
[48] R. A. ROSANOFF, J. F. GLOUDEMAN, AND S. LEVY, *Numerical conditions of stiffness matrix formulations for frame structures*, in Proc.of the Second Conference on Matrix Methods in Structural Mechanics, WPAFB Dayton, Ohio, 1968.
[49] W. SHOUGEN AND Z. SHUQIN, *An algorithm for $Ax = \lambda Bx$ with symmetric positive definite A and B*, SIAM J. Matrix Anal. Appl., 12 (1991), pp. 654–660.
[50] I. SLAPNIČAR, *Accurate Symmetric Eigenreduction by a Jacobi Method*, PhD thesis, Lehrgebiet Mathematische Physik, Fernuniversität Hagen, 1992.
[51] G. W. STEWART, *Computing the CS decomposition of a partitioned orthonormal matrix*, Numer. Math., 40 (1982), pp. 297–306.
[52] J.-G. SUN, *Perturbation analysis for the generalized eigenvalue and the generalized singular value problem*, in Matrix Pencils Proceedings of a Conference, Springer Verlag, 1982.
[53] ———, *Perturbation analysis for the generalized singular value problem*, SIAM J. Num. Anal., 20 (1983), pp. 611–625.
[54] A. VAN DER SLUIS, *Condition numbers and equilibration of matrices*, Numer. Math., 14 (1969), pp. 14–23.
[55] C. F. VAN LOAN, *Generalized Singular Values with Algorithms and Applications*, PhD thesis, University of Michigan, 1973.
[56] ———, *Generalizing the singular value decomposition*, SIAM J. Num. Anal., 13 (1976), pp. 76–83.
[57] ———, *A generalized SVD analysis of some weighting methods for equality constrained least squares*, in Matrix Pencils Proceedings of a Conference, Springer Verlag, 1982.
[58] ———, *Computing the CS and the generalized singular value decomposition*, Numer. Math., 46 (1985), pp. 479–491.
[59] K. VESELIĆ AND V. HARI, *A note on a one-sided Jacobi algorithm*, Numer. Math., 56 (1989), pp. 627–633.
[60] K. VESELIĆ AND I. SLAPNIČAR, *Floating-point perturbations of Hermitian matrices*, Linear Algebra Appl., 195 (1993), pp. 81–116.
[61] X. WANG, *Algorithm for reducing a positive definite pencil to a single matrix*. Unpublished manuscript, Fernuniversität Hagen, 1991.
[62] J. H. WILKINSON, *Error analysis of direct methods of matrix inversion*, J. Assoc. Comput. Mach., 8 (1962),

pp. 281–330.
[63] ———, *Rounding Errors in Algebraic Processes*, Prentice–Hall, Inc., 1963.
[64] ———, *The Algebraic Eigenvalue Problem*, Springer, Berlin Heidelberg New York, 1965.