SIAM J. MATRIX ANAL. APPL.
Vol. 16, No. 2, pp. 520–535, April 1995

014

ANALYSIS OF A QR ALGORITHM FOR COMPUTING SINGULAR VALUES*

S. CHANDRASEKARAN† AND I.C.F. IPSEN‡

Abstract. We extend the Golub–Kahan algorithm for computing the singular value decomposition of bidiagonal matrices to triangular matrices R. Our algorithm avoids the explicit formation of R^TR or RR^T.

We derive a relation between left and right singular vectors of triangular matrices and use it to prove monotonic convergence of singular values and singular vectors. The convergence rate for singular values equals the square of the convergence rate for singular vectors. The convergence behaviour explains the occurrence of deflation in the interior of the matrix.

We analyse the relationship between our algorithm and rank-revealing QR and URV decompositions. As a consequence, we obtain an algorithm for computing the URV decomposition, as well as a divide-and-conquer algorithm that computes singular values of dense matrices and may be beneficial on a parallel architecture. Our perturbation result for the smallest singular values of a triangular matrix is stronger than the traditional results because it guarantees high *relative* accuracy in the smallest singular values after an off-diagonal block of the matrix has been set to zero.

Key words. singular value decomposition, eigenvalue decomposition, QR decomposition, rank revealing QR decomposition, URV decomposition, deflation

AMS subject classifications. 15A18, 15A23, 15A42, 65F15, 65F25, 65W05

1. Introduction.

We present an algorithm for computing the singular value decomposition (SVD) of a real upper triangular matrix R that is based on the repeated QR decomposition of R.

1.1. The algorithm.

In 1965 Golub and Kahan [21] introduced an algorithm for the computation of the singular values and vectors of a real upper bidiagonal matrix B. The algorithm is based on the QR algorithm for computing eigenvalues but avoids the explicit formation of the tridiagonal matrix B^TB. An Algol implementation of this algorithm was proposed by Golub and Reinsch in 1970 [22].

The following extension of the unshifted Golub–Kahan algorithm from bidiagonal matrices to triangular matrices was proposed in [28]. It determines a new iterate from the QR decomposition of the transpose of the old iterate,

$$(*) \qquad R^{(0)} = R, \qquad [R^{(i)}]^T = Q^{(i+1)}R^{(i+1)}, \qquad i \geq 0,$$

and so avoids the explicit formation of R^TR or RR^T. Since the iterates $R^{(i)}$ are related to each other by orthogonal equivalence transformations, they all have the same singular values. We show in §3 that this algorithm computes the singular values of R.

The repeated transformation from lower to upper triangular form by means of orthogonal transformations was motivated by an algorithm for computing partial correlation coefficients [11], [12]. F. Chatelin and A. Ruhe pointed out to us that (*) had already been proposed by Fadeev, Kublanovskaya, and Fadeeva in 1966 [17],

* Received by the editors August 31, 1992; accepted for publication (in revised form) by F. T. Luk, February 25, 1994. The work presented in this paper was supported by National Science Foundation grant CCR-9102853.

† Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106-9560 (shir@ece.ucsb.edu).

‡ Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27695-8205 (ipsen@math.ncsu.edu).

where it is formulated as applying an LQ iteration to $R^{(2i)}$ and a QR iteration to $R^{(2i+1)}$.

Fernando and Parlett [19] derive a version of Rutishauser's differential QD algorithm based on (*) that computes singular values of bidiagonal matrices to high relative accuracy. Mathias and Stewart [31] use (*) to update rank-revealing URV and ULV decompositions, while Dowling, Ammann, and DeGroat [16] use it to develop a systolic real-time algorithm for computing the SVD. Moonen, Van Dooren, and Vanpouke [32] insert permutations to turn (*) into a Jacobi-type algorithm. In a second paper, Fernando and Parlett [18] incorporate shifts into (*) to compute singular values and vectors and to derive lower bounds on the smallest singular value.

This paper concentrates on the unshifted algorithm. At this point we do not advocate (*) as a practical method for computing singular values of dense matrices. Our motivation is to obtain insight into the behaviour of the Golub–Kahan algorithm for bidiagonal matrices [21] and into the unshifted QR algorithm [23], [33], [38], [41] for computing eigenvalues of symmetric matrices.

1.2. Overview. In §2 we derive a relation between left and right singular vectors of triangular matrices. It provides the basis for a simple analysis in §3 of the monotonic convergence of (*). In particular, we show that the tangent of the angle between certain canonical spaces and the singular vector subspaces of the iterates $R^{(i)}$ decreases monotonically at the usual rate; and that the convergence rate of the singular values is equal to the square of that of the singular vectors. These results explain the occurrence of deflation in the interior of the matrix.

Our analysis helps to understand the relation between algorithms that produce a complete SVD and those that produce a partial SVD, such as rank-revaling QR (RRQR) decompositions [8] and URV decompositions [24], [30], [36]. In §4 we show that with respect to a particular block partitioning of the matrix R, (*) proceeds in two phases: a rank-revealing phase where the large singular values are separated from the small ones, and a monotonic phase, where the iterates converge monotonically to block-diagonal form. Hence, preceding (*) with a rank-revealing algorithm accomplishes two things: it reverses the grading of inappropriately graded matrices and so enhances subsequent convergence; and, it forces premature deflation of a particular off-diagonal block and thus amounts to the computation of a URV decomposition. Based on this observation, we sketch a divide-and-conquer algorithm for computing singular values of dense matrices, which may be advantageous on a parallel architecture.

Section 5 derives a simple perturbation result for the smallest singular values of a triangular matrix. It is stronger than the traditional results because it guarantees high *relative* accuracy in the smallest singular values after an off-diagonal block of the matrix has been set to zero.

Some of the material in §§2 and 5 has appeared in preliminary form in [6], [7].

1.3. Relation to other algorithms. Two successive iterations of (*) are mathematically equivalent to one iteration of the unshifted QR algorithm for computing eigenvalues [23], [33], [38], [41] applied to both $R^{(i)}[R^{(i)}]^T$ and $[R^{(i)}]^T R^{(i)}$, as

$$R^{(i+2)}[R^{(i+2)}]^T = [Q^{(i+2)}]^T \left(R^{(i)}[R^{(i)}]^T \right) Q^{(i+2)}$$

and

$$[R^{(i+2)}]^T R^{(i+2)} = [Q^{(i+1)}]^T \left([R^{(i)}]^T R^{(i)} \right) Q^{(i+1)}.$$

This is also observed in [31]. If $R^{(0)}$ is upper bidiagonal, so are all iterates $R^{(i)}$, and two successive iterations amount to applying one iteration of the Golub–Kahan algorithm [21].

Fadeev, Kublanoskaya, and Fadeeva [17] and Fernando and Parlett [19] observe that one iteration of (*) is mathematically equivalent to one iteration of the Cholesky LR algorithm [41] applied to $A^{(i)} \equiv R^{(i)}[R^{(i)}]^T$. This is because $A^{(i)}$ has the upper-lower Cholesky factorisation $A^{(i)} = [R^{(i+1)}]^T R^{(i+1)}$. A subsequent multiplication of the Cholesky factors in reverse order gives the next iterate $A^{(i+1)} \equiv R^{(i+1)}[R^{(i+1)}]^T$.

From

$$A^{(i)} = R^{(i)}[R^{(i)}]^T = [R^{(i+1)}]^T R^{(i+1)}$$

it follows that $R^{(i)}$ is the factor from the upper-lower Cholesky factorisation of $A^{(i)}$, while $R^{(i+1)}$ is the factor from its lower-upper Cholesky factorisation. Hence the two factors are related through the orthogonal transformation $Q^{(i+1)}$. The fact that the two Cholesky factors of a matrix are related by an orthogonal transformation is used in [11], [12] to compute partial correlation coefficients. It is a consequence of the more general result that $M = M_1^T M_1 = M_2^T M_2$ for a positive-definite matrix M implies the existence of an orthogonal matrix W with $M_2 = W M_1$, cf. the exercise beneath [27, Coro. 7.2.8] and [19, §3].

Notation. The norm $\|\cdot\|$ represents the Euclidean two-norm. The identity matrix of order k is denoted by I_k and its ith column by e_i.

2. SVD of triangular matrices. To understand why (*) makes progress in every iteration we establish a relation between left and right singular vectors of triangular matrices. Let $R = U\Sigma V^T$ be the SVD of the upper triangular matrix

$$R = \begin{array}{c} \\ k \\ n-k \end{array} \begin{array}{c} \begin{array}{cc} k & n-k \end{array} \\ \begin{pmatrix} R_{11} & R_{12} \\ & R_{22} \end{pmatrix} \end{array},$$

where

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}, \qquad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

are orthogonal matrices, and

$$\Sigma = \begin{pmatrix} \Sigma_1 & \\ & \Sigma_2 \end{pmatrix}, \qquad \text{with} \qquad \Sigma_1 = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} \sigma_{k+1} & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix},$$

is a diagonal matrix whose diagonal contains the singular values of R in descending order,

$$\sigma_1 \geq \cdots \geq \sigma_k \geq \sigma_{k+1} \geq \cdots \geq \sigma_n.$$

The following theorem implies that if the singular values are well separated, then the left singular vectors are almost always closer to canonical form than the right singular vectors. By "canonical form" we mean a matrix $\binom{Z}{0}$ with Z orthogonal. The columns of such a canonical form span what we casually call the "canonical space $\binom{I_k}{0}$," that is, the column space of $\binom{I_k}{0}$.

THEOREM 2.1. *If R_{11} is nonsingular, and if U_{11} or V_{11} is also nonsingular, then*

$$\tan\theta_{u,k} \le \frac{\sigma_{k+1}}{\sigma_k} \tan\theta_{v,k},$$

where $\theta_{u,k}$ is the largest principal angle between $\binom{U_{11}}{U_{21}}$ and $\binom{I_k}{0}$; and $\theta_{v,k}$ is the largest principal angle between $\binom{V_{11}}{V_{21}}$ and $\binom{I_k}{0}$.

Proof. From the SVD $U^T R = \Sigma V^T$ one gets $U_{11}^T R_{11} = \Sigma_1 V_{11}^T$. Hence the nonsingularity of R_{11} implies that U_{11} is nonsingular whenever V_{11} is nonsingular. According to the CS decomposition [23, Thm. 2.6.1] for orthogonal matrices, V_{22} and U_{22} must also be nonsingular. Furthermore, the $(2,1)$ block in $R = U\Sigma V^T$ yields

$$U_{21}\Sigma_1 V_{11}^T + U_{22}\Sigma_2 V_{12}^T = 0$$

and

$$\|U_{22}^{-1}U_{21}\| \le \frac{\sigma_{k+1}}{\sigma_k} \|V_{11}^{-1}V_{12}\|.$$

Again, from the CS decomposition,

$$\sin\theta_{u,k} = \|U_{12}\| = \|U_{21}\| = \sqrt{1 - \frac{1}{\|U_{11}^{-1}\|^2}} = \sqrt{1 - \sigma_{\min}^2(U_{11}^T I_k)}.$$

Since the square root term represents the distance between the column space of $\binom{U_{11}}{U_{21}}$ and the canonical space $\binom{I_k}{0}$ [23, Coro. 2.6.2], the angle $\theta_{u,k}$ must be the largest principal angle [23, §12.4.3] between these two spaces. Moreover,

$$\|U_{11}^{-1}U_{12}\| = \|U_{22}^{-1}U_{21}\| = \tan\theta_{u,k}.$$

Substituting this in the above inequality gives

$$\tan\theta_{u,k} \le \frac{\sigma_{k+1}}{\sigma_k} \tan\theta_{v,k},$$

where $\theta_{v,k}$ is the analogous angle for V. □

COROLLARY 2.2. *If R is nonsingular, and if U_{11} or V_{11} is also nonsingular, then*

$$\frac{\sigma_k}{\sigma_{k+1}} \tan\theta_{u,k} \le \tan\theta_{v,k} \le \frac{\sigma_1}{\sigma_n} \tan\theta_{u,k}.$$

3. Monotonic convergence results. We determine some of the quantities that undergo monotonic change during one iteration of (*). Partition the iterates as in §2

$$R^{(i)} = \begin{array}{c} \\ k \\ n-k \end{array} \begin{array}{c} \begin{array}{cc} k & n-k \end{array} \\ \begin{pmatrix} R_{11}^{(i)} & R_{12}^{(i)} \\ & R_{22}^{(i)} \end{pmatrix} \end{array},$$

and denote their SVDs by $R^{(i)} = U^{(i)}\Sigma {V^{(i)}}^T$, where

$$U^{(i)} = \begin{pmatrix} U_{11}^{(i)} & U_{12}^{(i)} \\ U_{21}^{(i)} & U_{22}^{(i)} \end{pmatrix}, \qquad V^{(i)} = \begin{pmatrix} V_{11}^{(i)} & V_{12}^{(i)} \\ V_{21}^{(i)} & V_{22}^{(i)} \end{pmatrix}$$

are orthogonal.

For simplicity we assume that the initial matrix R is nonsingular. If this is not the case then the zero singular values can be extracted in two iterations. First perform a QR decomposition with column pivoting [5], [20], [23] that moves the zeros to the bottom of the matrix:

$$R = Q_P \begin{pmatrix} \hat{R}_{11} & \hat{R}_{12} \\ 0 & 0 \end{pmatrix} P^T,$$

where Q_P is orthogonal, P is a permutation matrix, and $\hat{R}_{11}$ is nonsingular upper triangular. In the next iteration eliminate the off-diagonal block,

$$\begin{pmatrix} \hat{R}_{11}^T & 0 \\ \hat{R}_{12}^T & 0 \end{pmatrix} = Q \begin{pmatrix} R_{11} & 0 \\ 0 & 0 \end{pmatrix}.$$

Our algorithm (*) can now be applied to the nonsingular triangular matrix R_{11}.

The convergence properties of the unshifted QR algorithm are well known [23], [33], [34], [38], [40], [41]. They are usually derived from the fact that one iteration of the QR algorithm is mathematically equivalent to one nested subspace iteration, applied to particular starting spaces, cf. in particular [34], [38], [40]. The subspace iterates converge linearly to eigenspaces with an asymptotic convergence rate equal to a ratio of adjacent singular values (in fact, the distance between the iterates and the eigenspace decreases from the start [40]). In [39] these results are extended to the computation of the SVD of R from R^TR and RR^T. The monotonic convergence of the eigenvalues during nested subspace iteration is proved through the connection to Toda flows [29].

3.1. Convergence of the singular vectors. We show that the angles between the invariant subspaces and the canonical spaces almost always decrease monotonically during (*). First we prove that the convergence rate for the leading k columns of the singular vector matrices depends on the gap between the kth and $(k+1)$st singular values.

THEOREM 3.1. *Let $R^{(0)}$ be nonsingular; $U_{11}^{(0)}$ or $V_{11}^{(0)}$ be nonsingular; and $\sigma_k > \sigma_{k+1}$.*

Then $U_{11}^{(i)}$ and $V_{11}^{(i)}$ are nonsingular for all $i \geq 0$; and convergence is monotonic in the sense that

$$\tan\theta_{v,k}^{(i+1)} \leq \frac{\sigma_{k+1}}{\sigma_k} \tan\theta_{v,k}^{(i)}, \qquad \tan\theta_{u,k}^{(i+1)} \leq \frac{\sigma_{k+1}}{\sigma_k} \tan\theta_{u,k}^{(i)},$$

where $\theta_{v,k}^{(i)}$ is the largest principal angle between the canonical space $\binom{I_k}{0}$ and the space spanned by the leading k columns of $V^{(i)}$; and $\theta_{u,k}^{(i)}$ is the analogous angle for $U^{(i)}$.

Proof. Consider one iteration $R^T = Q\hat{R}$. From the SVD $U^TR = \Sigma V^T$ follows $U_{11}^T R_{11} = \Sigma_1 V_{11}^T$. The nonsingularity of $R^{(0)}$ implies that U_{11} is nonsingular whenever V_{11} is. Applying Theorem 2.1 to R gives

$$\tan\theta_{u,k} \leq \frac{\sigma_{k+1}}{\sigma_k} \tan\theta_{v,k}.$$

Let $\hat{R} = \hat{U}\Sigma\hat{V}^T$ be the SVD of $\hat{R}$, where $\hat{U} = Q^TV$ and $\hat{V} = U$. The analogous relation for $\hat{R}$ is $\hat{U}_{11}^T\hat{R}_{11} = \Sigma_1\hat{V}_{11}^T$. Another application of Theorem 2.1, this time to $\hat{R}$, yields

$$\tan\theta_{\hat{u},k} \leq \frac{\sigma_{k+1}}{\sigma_k} \tan\theta_{\hat{v},k}.$$

Putting the two inequalities together via $\hat{V} = U$ results in

$$\tan\theta_{\hat{u},k} \leq \frac{\sigma_{k+1}}{\sigma_k} \tan\theta_{u,k}. \qquad \square$$

Now we prove that the rate of convergence for interior columns of the singular vector matrices depends on the gaps with the adjacent distinct singular values. This result holds if the initial singular vector matrices are strongly nonsingular.[1] The fact that the strong nonsingularity is preserved throughout the iteration (*) follows already from the convergence results of the eigenvalue QR algorithms [34], [38]–[41].

THEOREM 3.2. *Let $R^{(0)}$ be nonsingular, $U^{(0)}$ or $V^{(0)}$ be strongly nonsingular, and*

$$\sigma_k > \sigma_{k+1} = \cdots = \sigma_{k+m} > \sigma_{k+m+1}.$$

Then $U^{(i)}$ and $V^{(i)}$ are strongly nonsingular, for all $i \geq 0$, and columns $k+1, \ldots, k+m$ of $U^{(i)}$ and $V^{(i)}$ converge to a $n \times m$ matrix of the form

$$\begin{array}{c} \\ k \\ m \end{array} \begin{array}{c} \begin{array}{c} m \end{array} \\ \begin{pmatrix} 0 \\ Z \\ 0 \end{pmatrix} \end{array},$$

where Z is orthogonal, at the rate

$$\rho_k = \max\left\{\frac{\sigma_{k+1}}{\sigma_k}, \frac{\sigma_{k+m+1}}{\sigma_{k+1}}\right\}.$$

Proof. The strong nonsingularity of $U^{(i)}$ and $V^{(i)}$ can be proved as in Theorem 3.1. According to the convergence results for distinct singular values in Theorem 3.1, the singular vector matrices converge at the rate σ_{k+1}/σ_k to the canonical form

$$\begin{array}{c} \begin{array}{cc} k & \end{array} \\ \begin{pmatrix} X & X & & & & \\ X & X & & & & \\ & & X & X & X & X \\ & & X & X & X & X \\ & & X & X & X & X \\ & & X & X & X & X \end{pmatrix} \end{array} \begin{array}{c} k \\ \\ \\ \\ \\ \\ \end{array},$$

while they converge at the rate $\sigma_{k+m+1}/\sigma_{k+m}$ to the canonical form

$$\begin{array}{c} \begin{array}{cccc} k & + & m & \end{array} \\ \begin{pmatrix} X & X & X & X & & \\ X & X & X & X & & \\ X & X & X & X & & \\ X & X & X & X & & \\ & & & & X & X \\ & & & & X & X \end{pmatrix} \end{array} \begin{array}{c} k \\ + \\ m \\ \\ \\ \\ \end{array}.$$

[1] A square matrix is called "strongly nonsingular" if all its leading principal submatrices are nonsingular.

Here X represents a matrix element that may be nonzero. Thus the singular vector matrices converge to the form

$$\begin{array}{c} \\ k \\ \\ m \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{cc} k & \quad m \end{array} \\ \begin{pmatrix} X & X & & & & \\ X & X & & & & \\ & & X & X & & \\ & & X & X & & \\ & & & & X & X \\ & & & & X & X \end{pmatrix} \end{array}$$

at the rate $\max\{\sigma_{k+1}/\sigma_k,\ \sigma_{k+m+1}/\sigma_{k+m}\}$. □

Therefore, if all singular values of the matrix $R^{(0)}$ are distinct and if the singular vector matrices are strongly nonsingular, then the singular vector matrices converge to the identity matrix monotonically at the rate $\max_k \sigma_{k+1}/\sigma_k$. In general, the singular vector matrices converge to a block-diagonal matrix whose diagonal blocks are orthogonal. The convergence rate is equal to the largest ratio of adjacent distinct singular values. The size of the kth diagonal block equals the multiplicity of the kth distinct singular value, and the columns making up the block represent an orthogonal basis for the associated invariant subspace.

3.2. Convergence of the singular values. From the convergence rate of the singular vector matrices we can in turn estimate the convergence rate for the singular values. First we show that the singular values converge monotonically. The inequalities in the lemma below are also derived in [31, Thm. 2.1]. They are special cases of the monotonicity properties of eigenvalues during subspace iteration [29].

LEMMA 3.3. *If $R^{(0)}$ is nonsingular then*

$$\|R_{11}^{(i+1)^{-1}}\| \le \|R_{11}^{(i)^{-1}}\|, \qquad \|R_{22}^{(i+1)}\| \le \|R_{22}^{(i)}\|.$$

Proof. From one iteration $R^T = Q\hat{R}$ follows that $R_{11}^T = Q_{11}\hat{R}_{11}$ and $\hat{R}_{22} = Q_{22}^T R_{22}^T$. Hence

$$\|\hat{R}_{11}^{-1}\| \le \|R_{11}^{-1}\|, \qquad \|\hat{R}_{22}\| \le \|R_{22}\|. \qquad \square$$

Now we derive the rate of convergence of the extreme singular values of the leading and trailing principal submatrices.

THEOREM 3.4. *Let $R^{(0)}$ be nonsingular, $U_{11}^{(0)}$ or $V_{11}^{(0)}$ be nonsingular, and $\sigma_k > \sigma_{k+1}$.*

Then convergence of the singular values is monotonic in the sense that

$$\frac{\|R_{22}^{(i)}\| - \sigma_{k+1}}{\sigma_{k+1}} \le \frac{\sigma_1}{\sigma_k} \tan^2 \theta_{v,k}^{(i)}, \qquad \frac{\|R_{11}^{(i)^{-1}}\| - \frac{1}{\sigma_k}}{\frac{1}{\sigma_k}} \le \frac{\sigma_{k+1}}{\sigma_n} \tan^2 \theta_{v,k}^{(i)}.$$

Proof. Consider one iteration $R^T = Q\hat{R}$. The SVD $R = U\Sigma V^T$ gives

$$R_{22} = U_{21}\Sigma_1 V_{21}^T + U_{22}\Sigma_2 V_{22}^T = U_{22}(\Sigma_2 + U_{22}^{-1}U_{21}\Sigma_1 V_{21}V_{22}^{-T})V_{22}^T.$$

Following the proof of Theorem 2.1,

$$\|R_{22}\| \le \sigma_{k+1} + \sigma_1 \tan\theta_{u,k} \tan\theta_{v,k}$$

and substituting

$$\tan\theta_{u,k} \leq \frac{\sigma_{k+1}}{\sigma_k} \tan\theta_{v,k}$$

for $\tan\theta_{u,k}$ gives

$$\frac{\|R_{22}\| - \sigma_{k+1}}{\sigma_{k+1}} \leq \frac{\sigma_1}{\sigma_k} \tan^2\theta_{v,k}.$$

The second inequality is derived analogously from $R^{-1} = V\Sigma^{-1}U^T$. □

Theorem 3.4 implies that the relative distance of $\|R_{11}^{-1}\|$ from $1/\sigma_k$ is bounded above by the condition number of Σ_2, as well as the square of the angle between the leading k columns of the right singular vector matrix and the corresponding canonical space. Similarly, the relative distance of $\|R_{22}\|$ from σ_{k+1} is bounded above by the condition number of Σ_1 and the square of the same angle. Hence if V_{11} is well conditioned and the spread of singular values in Σ_1 is small then $\|R_{22}\|$ is close to σ_{k+1}.

Furthermore, the rate of convergence of the singular values is approximately the square of that of the associated singular vectors. B. Parlett pointed out that this is a result of Rayleigh's principle.

COROLLARY 3.5. *The following convergence estimates hold:*

$$\frac{\|\hat{R}_{22}\| - \sigma_{k+1}}{\|R_{22}\| - \sigma_{k+1}} \approx \left(\frac{\tan\theta_{\hat{v},k}}{\tan\theta_{v,k}}\right)^2 \leq \left(\frac{\sigma_{k+1}}{\sigma_k}\right)^2$$

and

$$\frac{\|\hat{R}_{11}^{-1}\| - \frac{1}{\sigma_k}}{\|R_{11}^{-1}\| - \frac{1}{\sigma_k}} \approx \left(\frac{\tan\theta_{\hat{v},k}}{\tan\theta_{v,k}}\right)^2 \leq \left(\frac{\sigma_{k+1}}{\sigma_k}\right)^2.$$

Now we estimate the convergence of an interior principal submatrix. The theorem below implies that the iterates $R^{(i)}$ converge to a diagonal matrix with the singular values in sorted order along the diagonal.

THEOREM 3.6. *Let $R^{(0)}$ be nonsingular, $V^{(0)}$ or $U^{(0)}$ be strongly nonsingular, and*

$$\sigma_k > \sigma_{k+1} = \cdots = \sigma_{k+m} > \sigma_{k+m+1}.$$

Then the principal submatrix of order m of $R^{(i)}$,

$$\begin{pmatrix} R^{(i)}_{k+1,k+1} & \cdots & R^{(i)}_{k,k+m} \\ & \ddots & \vdots \\ & & R^{(i)}_{k+m,k+m} \end{pmatrix},$$

converges to $\sigma_{k+1}I_m$ at approximately the rate ρ_k^2, where

$$\rho_k = \max\left\{\frac{\sigma_{k+1}}{\sigma_k}, \frac{\sigma_{k+m+1}}{\sigma_{k+1}}\right\}.$$

Proof. Partition the iterates as in the proof of Theorem 3.2,

$$R^{(i)} = \begin{pmatrix} R^{(i)}_{11} & X & X \\ & R^{(i)}_{22} & X \\ & & R^{(i)}_{33} \end{pmatrix},$$

where $R_{11}^{(i)}$ is of order k and $R_{22}^{(i)}$ is of order m.

Corollary 3.5 implies that the convergence of $\|{R_{11}^{(i)}}^{-1}\|$ to $1/\sigma_k$ and the convergence of

$$\left\| \begin{pmatrix} R_{22}^{(i)} & X \\ & R_{33}^{(i)} \end{pmatrix} \right\|$$

to σ_{k+1} occur at approximately the rate $\sigma_{k+1}^2/\sigma_k^2$, while the convergence of

$$\left\| \begin{pmatrix} R_{11}^{(i)} & X \\ & R_{22}^{(i)} \end{pmatrix}^{-1} \right\|$$

to $1/\sigma_{k+1}$ and the convergence of $\|R_{33}^{(i)}\|$ to σ_{k+m+1} occurs at approximately the rate $\sigma_{k+1}^2/\sigma_{k+m+1}^2$.

Consider the essential limit of the iterates, which we define as $R^{(\infty)} = U^{(\infty)}\Sigma V^{(\infty)}$, and partition their singular vector matrices like $R^{(i)}$:

$$U^{(\infty)} = \begin{pmatrix} U_{11}^{(\infty)} & & \\ & U_{22}^{(\infty)} & \\ & & U_{33}^{(\infty)} \end{pmatrix}, \qquad V^{(\infty)} = \begin{pmatrix} V_{11}^{(\infty)} & & \\ & V_{22}^{(\infty)} & \\ & & V_{33}^{(\infty)} \end{pmatrix},$$

where the diagonal blocks $U_{ii}^{(\infty)}$ and $V_{ii}^{(\infty)}$ are orthogonal and

$$\Sigma = \begin{pmatrix} \Sigma_1 & & \\ & \sigma_{k+1} I_m & \\ & & \Sigma_3 \end{pmatrix}, \quad \Sigma_1 = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix}, \quad \Sigma_3 = \begin{pmatrix} \sigma_{k+m+1} & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix}.$$

Then $R_{22}^{(\infty)} V_{22}^{(\infty)} = \sigma_{k+1} U_{22}^{(\infty)}$, so that $R_{22}^{(\infty)} = \sigma_{k+1} U_{22}^{(\infty)} {V_{22}^{(\infty)}}^T$ is a multiple of an orthogonal matrix. But $R_{22}^{(\infty)}$ is also upper triangular. Therefore $R_{22}^{(\infty)} = \sigma_{k+1} I_m$ is a scalar matrix, and $U_{22}^{(\infty)} = V_{22}^{(\infty)}$ (where we have assumed that $R^{(\infty)}$ has positive diagonal elements). Hence a principal submatrix $R_{22}^{(i)}$ associated with a singular value σ_{k+1} of multiplicity m converges to $\sigma_{k+1} I_m$ at approximately the rate $\max\{\sigma_{k+1}^2/\sigma_k^2, \sigma_{k+m+1}^2/\sigma_{k+1}^2\}$. □

3.3. Consequences. Our upper bounds on the relative distance between $\|{R_{11}^{(i)}}^{-1}\|$ and $\|R_{22}^{(i)}\|$ to the respective singular values depend on the spreads $\sigma_1 \ldots \sigma_k$ and $\sigma_{k+1} \ldots \sigma_n$, and the conditioning of the leading principal submatrices of order k of $V^{(i)}$. The number of iterations required to reduce the relative distance between $\|R_{22}\|$ and σ_{k+1} to ϵ can thus be estimated as

$$\frac{\log \sigma_1/\sigma_k - \log \epsilon + \log \tan \theta_{v,k}^{(0)}}{\log \sigma_k/\sigma_{k+1}}.$$

An analogous estimate can be made for $\|R_{11}^{-1}\|$.

According to [34, §2.2], the QR algorithm tends to converge to the small eigenvalues first. According to our analysis, though, there is no preference of (*) for small singular values over larger ones. However, such a preference may be enforced by a suitable choice of shifts [33], [41].

The bounds on the relative distance also explain why (*) and the QR algorithm have such a hard time with graded matrices whose elements increase in size towards the bottom, cf. [14, §5] and [19, Thm. 5]. These matrices have a large spread in the spectrum and very ill-conditioned leading principal submatrices. One of the simplest examples of a graded matrix is

$$R^{(0)} = \begin{pmatrix} 1 & \epsilon \\ & \alpha \end{pmatrix},$$

where $\epsilon \ll 1 \ll \alpha$. One iteration of (*) gives

$$R^{(1)} = \frac{1}{\sqrt{1+\epsilon^2}} \begin{pmatrix} 1+\epsilon^2 & \alpha\epsilon \\ & \alpha \end{pmatrix},$$

whose off-diagonal element has increased from ϵ to about $\alpha\epsilon$. But the diagonal elements have only changed marginally, and it is obvious that many iterations are needed to arrive at a diagonal matrix with diagonal elements in descending order (a similar example was used in [41, §8.7] to illustrate slow convergence of the LR algorithm). Section 4.2 illustrates how to force fast convergence on such graded matrices without the need to decide between QR- and QL-type algorithms as in [14], [19].

The leading principal submatrices of the singular vector matrices are almost always nonsingular [38, p. 430] but may be very ill conditioned, in which case the convergence is slow.

4. RRQR and URV decompositions. We discuss the connections between (*) on the one hand and RRQR decompositions [8] and URV decompositions [36] on the other hand.

4.1. Two phases in the algorithm. For each partitioning index k of the nonsingular matrix $R^{(0)}$ define

$$\gamma_k^{(i)} \equiv \|{R_{11}^{(i)}}^{-1}\| \, \|R_{22}^{(i)}\|.$$

When $\gamma_k^{(i)} < 1$ then $\|R_{22}^{(i)}\| < 1/\|{R_{11}^{(i)}}^{-1}\|$, which means that all singular values of $R_{11}^{(i)}$ are larger than the singular values of $R_{22}^{(i)}$, and a partial ordering of the singular values of $R^{(i)}$ has occurred. Lemma 3.3 implies that

$$\gamma_k^{(i+1)} \leq \gamma_k^{(i)},$$

so the separation between singular values of $R_{11}^{(i)}$ and $R_{22}^{(i)}$ never decreases throughout the iterations (*). Furthermore, if $\sigma_{k+1}/\sigma_k < 1$ then $\gamma_k^{(i)} \to \sigma_{k+1}/\sigma_k$ as $i \to \infty$, provided $U^{(0)}$ and $V^{(0)}$ are strongly nonsingular. Because the convergence of $\gamma_k^{(i)}$ to $\sigma_{k+1}/\sigma_k < 1$ is monotone, there exists a number i_k such that $\gamma_k^{(i)} < 1$ for all $i \geq i_k$. It makes sense therefore to distinguish, for each k, two phases of (*) depending on the value of $\gamma_k^{(i)}$.

1. A rank-revealing phase, where $\gamma_k^{(i)} > 1$, during which the singular values of $R_{11}^{(i)}$ and $R_{22}^{(i)}$ are in the process of separating.

2. A monotonic phase, where $\gamma_k^{(i)} \leq 1$, during which *all* quantities of interest converge monotonically.

4.2. The rank-revealing phase. The name for the first phase comes from its resemblance to RRQR decompositions. Given a matrix R and a specific k (usually determined by the number of singular values of R that are smaller than a certain threshold), RRQR algorithms try to find a permutation matrix P so that the triangular matrix $\bar{R}$ in the QR decomposition $RP = Q\bar{R}$ has a $(1,1)$ block with maximal smallest singular value, and/or a $(2,2)$ block with minimal largest singular value [8]. The existence of RRQR decompositions was proved in [26], and one of the most accurate RRQR algorithms is Hybrid III(k) [8], which finds a permutation matrix P so that $RP = Q\bar{R}$ with

$$\bar{\gamma}_k \equiv \|\bar{R}_{11}^{-1}\| \, \|\bar{R}_{22}\| \leq (k+1)(n-k+1) \, \frac{\sigma_{k+1}}{\sigma_k}.$$

In practice, though, the cheaper and possibly less accurate forms of column pivoting, such as QR with column pivoting [5], [20], [23], tend to work quite well (an attempt at explaining the practical effectiveness of the simple column pivoting strategies, regardless of their potential failures, is made in [8]).

Therefore, if the singular values σ_k and σ_{k+1} are well separated then one can try to enforce the onset of the monotonic phase for a particular k by preceding $(*)$ with an RRQR decomposition. This also reverses the grading in a matrix all of whose large elements are at the bottom, thus obviating the need for a decision between an algorithm of QR or of QL type [14], [19].

The idea of permuting rows or columns of the iterates during eigenvalue computations is not new. Pivoting, in the form of row exchanges, has been suggested for the LR algorithm, [41, §8.13] and [34, §2.7], to enhance numerical stability in those cases where the orthodox LR algorithm fails to converge. A preliminary pivoting step has also been suggested for Jacobi methods: Hari and Veselič use QR with column pivoting [25] and Cholesky decomposition with symmetric pivoting [37], while Demmel and Veselič [15, Algorithm 4.4] propose to compute the eigendecomposition of a symmetric positive-definite matrix A by first determining the Cholesky factor R of A with complete pivoting, followed by the application of a one-sided Jacobi method to R.

4.3. The monotonic phase. Since $\|R_{12}^{(i+1)}\| \leq \gamma_k^{(i)} \|R_{12}^{(i)}\|$, this implies for the monotonic phase $\|R_{12}^{(i+1)}\| < \|R_{12}^{(i)}\|$. Hence the off-diagonal blocks $R_{12}^{(i)}$ decrease monotonically; and convergence to block-diagonal form is fast once the monotonic phase has been reached. Since $\gamma_k^{(i)} \to \sigma_{k+1}/\sigma_k$, the blocks corresponding to well-separated singular values may decrease faster and deflation[2] is likely to set in earlier.

4.4. A divide and conquer algorithm. The previous sections showed that once the rank-revealing phase has been completed for some k, the iterates converge rapidly to block diagonal form. Hence preceding $(*)$ with an RRQR algorithm tends to force completion of the rank-revealing phase and the start of deflation for that k. This observation leads to a divide and conquer algorithm for computing singular values of dense or banded matrices A, which may be advantageous on a parallel architecture. Below is a rough sketch.

1. Select a k and apply an RRQR algorithm to $AP = Q\bar{R}$ so that $\bar{\gamma}_k < 1$.
2. Set $R^{(0)} = \bar{R}$ and iterate $(*)$ until $\|R_{12}^{(i)}\|$ is small enough.
3. Apply Steps 1 and 2 recursively to $R_{11}^{(i)}$ and to $R_{22}^{(i)}$.

[2] The splitting of a matrix into two or more independent diagonal blocks due to almost zero off-diagonal blocks is called "deflation."

There are several ways to determine the index k in Step 1 where the matrix is to be split. The simplest option is to set $k = n/2$ and choose Hybrid III($n/2$) as the RRQR algorithm to break the matrix into equally sized blocks and ensure load balance with regard to parallel execution. But the separation of the singular values $\sigma_{n/2}$ and $\sigma_{n/2+1}$ may not be large enough. Alternatively one can apply QR with column pivoting and select as k that index for which $|\bar{r}_{k+1,k+1}|/|\bar{r}_{kk}| \approx \bar{\gamma}_k$ is smallest. A third possibility is to estimate the norm of $\|\bar{R}_{11}^{-1}\|$ by an incremental condition estimator [1]–[3]. We have not yet gathered enough computational experience to judge whether the algorithm presents a viable alternative to other methods that operate on dense matrices, such as Jacobi methods [4], [9], for instance.

4.5. Computation of the URV decomposition. We show how to compute a URV decomposition by means of (*). The URV decomposition was introduced by Hanson and Lawson, [24] and [30, Thm. (3.19)], to solve (rank deficient) least squares problems. Stewart [36] emphasizes its use for computing the null space of a matrix that is repeatedly updated. If R has rank $k < n$ then there exist orthogonal matrices U and V and a nonsingular upper triangular matrix $\bar{R}$ of order k such that

$$R = U \begin{pmatrix} \bar{R} & 0 \\ 0 & 0 \end{pmatrix} V^T.$$

In practice, R is often only of numerical rank k, where the singular values $\sigma_{k+1}, \ldots, \sigma_n$ are small. In this case one would like to find a decomposition

$$R = U \begin{pmatrix} \bar{R}_{11} & \bar{R}_{12} \\ & \bar{R}_{22} \end{pmatrix} V^T,$$

where $\|\bar{R}_{11}^{-1}\| \approx 1/\sigma_k$ is large and where $\|\bar{R}_{22}\| \approx \sigma_{k+1}$ and $\|\bar{R}_{12}\| \approx \sigma_{k+1}$ are small.

Hanson and Lawson, [24] and [30, §14], as well as Stewart and Mathias [31], [35], [36], compute a URV decomposition by determining orthogonal matrices P and Q such that $RP = Q\bar{R}$ where $\|\bar{R}_{11}^{-1}\| \approx 1/\sigma_k$ and $\|\,(\,\bar{R}_{12}^T \quad \bar{R}_{22}^T\,)\,\| \approx \sigma_{k+1}$. Stewart and Mathias [31], [35] then perform several of the following "refinement steps" on $R^{(0)} = \bar{R}$ to further decrease the size of the $(1,2)$ block: first determine an orthogonal matrix $Q^{(1)}$ so that ${R^{(1)}}^T = R^{(0)}Q^{(1)}$ is lower triangular and, second, determine an orthogonal matrix $Q^{(2)}$ so that $R^{(2)} = {Q^{(2)}}^T {R^{(1)}}^T$ is upper triangular. In [36] Stewart proposes an incomplete version of these refinement steps: reduce only the last column of $R^{(0)}$ to e_n, and in this resulting matrix in turn reduce only the last row to e_n^T.

Note that in the beginning these algorithms accomplish more than an RRQR decomposition. Due to the rotations performed on both sides of the matrix the off-diagonal block also ends up being small. Hence the following result from [35] applies. If

$$\|R_{12}^{(0)}\| + \|R_{22}^{(0)}\| < \sigma_k$$

then the first part of the refinement steps in [35], [36] causes a monotonic decrease $\|R_{12}^{(1)}\| < \|R_{12}^{(0)}\|$ in the (1,2) block, and so does, of course, the second part of the refinement step. The refinement step in [35] represents two iterations of (*)

$$[R^{(0)}]^T = Q^{(1)}R^{(1)}, \qquad [R^{(1)}]^T = Q^{(2)}R^{(2)},$$

while the refinement step in [36] amounts to one incomplete iteration of (*) where $R_{11}^{(1)}$ of order $n-1$ remains lower triangular.

Section 4.3 showed that generally no assumption on the (1,2) block is necessary to ensure monotonic decrease provided the singular values of $R_{11}^{(i)}$ and $R_{22}^{(i)}$ are well separated: if $\gamma_k^{(i)} < 1$ then $\|R_{12}^{(i+1)}\| < \|R_{12}^{(i)}\|$. This is true regardless of whether σ_{k+1} is small or not. However, if $\|R_{22}^{(i)}\|$ is small then $\|R_{12}^{(i+1)}\|$ is as small—regardless of the relation between $R_{11}^{(i)}$ and $R_{22}^{(i)}$—because $R_{12}^{(i+1)} = Q_{21}^{(i+1)} {R_{22}^{(i)}}^T$, so $\|R_{12}^{(i+1)}\| \leq \|R_{22}^{(i)}\|$.

Therefore, one can compute a URV decomposition of R by determining an RRQR decomposition $RP = Q\bar{R}$ and then applying several iterations of $(*)$ to $\bar{R}$, which then converges monotonically to the desired URV decomposition.

5. Deflation criteria. We extend some of the existing convergence and deflation critera for computing singular values of bidiagonal matrices to triangular matrices.

Demmel and Kahan [14] and Deift et al. [10] have shown that, in floating point arithmetic, a particular implementation of the Golub–Kahan algorithm for bidiagonal matrices computes small singular values to high relative accuracy. This implementation is based on deflation and convergence criteria that preserve high relative accuracy of the computed singular values.

Fernando and Parlett [19] introduce a modification of Rutishauser's differential QD algorithm for bidiagonal matrices that is faster than the current implementations of the Golub–Kahan algorithm. Their deflation criterion for shifted matrices continues to preserve high relative accuracy for the singular values. Demmel and Gragg [13] extend the criterion from [10] to biacyclic matrices.

In [35] Stewart proves a deflation criterion that bounds the relative accuracy of the smallest singular value and can be considered an extension of Criterion 2a in [14] to triangular matrices: If the off-diagonal block is small enough with regard to the singular value separation,

$$\|R_{12}\| < \sigma_k - \|R_{22}\|,$$

then

$$\frac{|\sigma_{k+i} - \sigma_i(R_{22})|}{\sigma_1(R_{22})} \leq \frac{\|R_{12}\|^2}{\delta^2 - \|R_{22}\|^2}, \qquad \delta = \sigma_k - \|R_{12}\|.$$

In [31] Mathias and Stewart prove a deflation criterion for the eigenvalues of RR^T that bounds the relative accuracy of the $n-k$ smallest eigenvalues,

$$\frac{\sigma_i^2(R_{22}) - \sigma_{k+i}^2}{\sigma_i^2(R_{22})} \leq \frac{\|R_{12}\|^2}{\mathrm{gap}_k(\sigma_{\min}(R_{11}) + \|R_{22}\|)}, \qquad \mathrm{gap}_k = 6\mathrm{min}(R_{11}) - \|R_{22}\|$$

(a similar theorem in [31] also bounds the relative accuracy of the largest k eigenvalues of R^TR). It implies a first-order bound on the relative accuracy of the $n-k$ smallest singular values of R,

$$\frac{\sigma_i(R_{22}) - \sigma_{k+i}}{\sigma_i(R_{22})} \leq \frac{\|R_{12}\|^2}{2\mathrm{gap}_k(\sigma_{\min}(R_{11}) + \|R_{22}\|)} + O\left(\frac{\|R_{12}\|^4}{\mathrm{gap}_k^2}\right).$$

If $\|R_{12}\|$ is small then this criterion permits earlier deflation than the one from [35].

Our deflation criterion below guarantees high relative accuracy in $\sigma_1(R_{22})$ and holds without any assumptions on the size of $\|R_{12}\|$.

THEOREM 5.1. *If R is nonsingular, V is strongly nonsingular and* $\mathrm{gap}_k = \sigma_{\min}(R_{11}) - \|R_{22}\| > 0$ *then*

$$\frac{|\sigma_{k+j} - \sigma_j(R_{22})|}{\sigma_1(R_{22})} \leq \frac{\|R_{12}\|}{\mathrm{gap}_k}.$$

Proof. From $R^T = Q\hat{R}$ it follows that

$$\begin{pmatrix} 0 \\ R_{22}^T \end{pmatrix} = Q \begin{pmatrix} 0 \\ \hat{R}_{22} \end{pmatrix} + Q \begin{pmatrix} \hat{R}_{12} \\ 0 \end{pmatrix}.$$

Using $|\sigma_j(A+E) - \sigma_j(A)| \leq \|E\|$ [23, Cor. 8.3.2] with

$$A = Q \begin{pmatrix} 0 \\ \hat{R}_{22} \end{pmatrix}, \qquad A + E = Q \begin{pmatrix} 0 \\ \hat{R}_{22} \end{pmatrix} + Q \begin{pmatrix} \hat{R}_{12} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ R_{22}^T \end{pmatrix}$$

yields

$$|\sigma_j(R_{22}) - \sigma_j(\hat{R}_{22})| \leq \|\hat{R}_{12}\|.$$

This implies, together with

$$\|\hat{R}_{12}\| \leq \gamma_k \|R_{12}\|, \qquad \gamma_k \equiv \|R_{11}^{-1}\| \, \|R_{22}\|$$

from §4.3, that

$$|\sigma_j(R_{22}) - \sigma_j(\hat{R}_{22})| \leq \gamma_k \|R_{12}\|.$$

Now set $R^{(0)} \equiv R$ and apply $(*)$ to $R^{(0)}$. According to §4.1 $\gamma_k^{(i+1)} \leq \gamma_k^{(i)}$, so the difference between two successive iterations is

$$|\sigma_j(R_{22}^{(i+1)}) - \sigma_j(R_{22}^{(i)})| \leq \|R_{12}^{(i+1)}\| \leq \gamma_k^{(i)} \|R_{12}^{(i)}\| \leq \gamma_k^i \|R_{12}\|.$$

As for the difference between iteration $i+2$ and i, we make use of Stewart's idea [36],

$$\begin{aligned} |\sigma_j(R_{22}^{(i+2)}) - \sigma_j(R_{22}^{(i)})| &\leq |\sigma_j(R_{22}^{(i+2)}) - \sigma_j(R_{22}^{(i+1)})| + |\sigma_j(R_{22}^{(i+1)}) - \sigma_j(R_{22}^{(i)})| \\ &\leq \left(\gamma_k^{i+1} + \gamma_k^i\right) \|R_{12}\|. \end{aligned}$$

Because V is strongly nonsingular, Theorem 3.6 implies that the singular values of $R_{22}^{(i)}$ converge to the singular values $\sigma_{k+1}, \ldots, \sigma_n$ as $i \to \infty$. The assumption $\gamma_k < 1$ allows extrapolation to the limit

$$|\sigma_{k+j} - \sigma_j(R_{22})| \leq \|R_{12}\| \sum_{l=1}^{\infty} \gamma_k^l = \|R_{12}\| \, \frac{\gamma_k}{1-\gamma_k}$$

as $\sum_{l=1}^{\infty} \gamma_k^l = \frac{1}{1-\gamma_k} - 1$. Hence

$$\frac{|\sigma_{k+j} - \sigma_j(R_{22})|}{\sigma_1(R_{22})} \leq \frac{\|R_{11}^{-1}\|}{1-\gamma_k} \|R_{12}\| = \frac{\|R_{12}\|}{\sigma_{\min}(R_{11}) - \|R_{22}\|}. \qquad \square$$

Theorem 5.1 is most valuable for the case $k = n-m$, where m is the multiplicity of the smallest singular value σ_n, because it assures that $\|R_{22}\|$ approximates σ_n to high relative accuracy whenever the norm of the off-diagonal block is small and the singular values of R_{11} are much larger than those of R_{22}. Since the requirement $\text{gap}_k > 0$ is equivalent to $\gamma_k < 1$, Theorem 5.1 can be applied as a deflation criterion for $(*)$ only once the monotonic phase for k has set in. Note that $\text{gap}_k > 0$ is not satisfied for a graded matrix whose elements increase in size towards the bottom, regardless of how small R_{12} is.

In general, Theorem 5.1 suggests using the simple deflation criterion

$$\|R_{12}\| \leq \eta \frac{\text{gap}_k}{\|R_{22}\|}$$

to guarantee absolute accuracy η for all singular values of R_{22}. If $\|R_{22}\|$ is small and if the singular values of R_{11} and R_{22} are well separated then this criterion permits earlier deflation than the traditional criterion [23, Coro. 8.3.2]

$$\|R_{12}\| \leq \eta.$$

Relative accuracy η for *all* singular values is achieved if

$$\|R_{12}\| \leq \eta \frac{\text{gap}_k}{\kappa(R_{22})},$$

where $\kappa(R_{22}) = \|R_{22}\| \, \|R_{22}^{-1}\|$ is the condition number of R_{22}.

Acknowledgments. We thank Stan Eisenstat, Beresford Parlett, and David Watkins for helpful discussions, as well as Françoise Chatelin, Axel Ruhe, and a referee for references to the literature.

REFERENCES

[1] J. Barlow and U. Vemulapati, *Rank detection methods for sparse matrices*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 1279–1297.

[2] C. Bischof, *Incremental condition estimation*, SIAM J. Matrix Anal. Appl., 11 (1990), pp. 312–322.

[3] C. Bischof, J. Lewis, and D. Pierce, *Incremental condition estimation for sparse matrices*, SIAM J. Matrix Anal. Appl., 11 (1990), pp. 644–659.

[4] R. Brent and F. Luk, *The solution of singular-value and symmetric eigenvalue problems on multiprocessor arrays*, SIAM J. Sci. Statist. Comput., 6 (1985), pp. 69–84.

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

[6] S. Chandrasekaran and I. Ipsen, *On the singular value decomposition of triangular matrices*, in Proceedings of the 1992 Shanghai International Numerical Algebra and its Applications Conference.

[7] ———, *A divide and conquer algorithm for computing singular values*, Z. Angew. Math. Mech., 74 (1994), pp. T 532–534.

[8] ———, *On rank-revealing QR factorisations*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 592–622.

[9] J. Charlier, M. Vanbegin, and P. Van Dooren, *On efficient implementations of Kogbetliantz's algorithm for computing the singular value decomposition*, Numer. Math., 52 (1988), pp. 279–300.

[10] P. Deift, J. Demmel, L. Li, and C. Tomei, *The bidiagonal singular value decomposition and Hamiltonian mechanics*, SIAM J. Numer. Anal., 28 (1991), pp. 1463–1516.

[11] J. Delosme and I. Ipsen, *Computing partial correlations from the data matrix*, Research Report 541, Department of Computer Science, Yale University, New Haven, CT, 1987.

[12] ———, *From Bareiss' algorithm to the stable computation of partial correlations*, J. Comput. Appl. Math., 27 (1989), pp. 53–91; also *Parallel Algorithms for Numerical Linear Algebra* (*Advances in Parallel Computing*, 1), H. van der Vorst and P. Van Dooren, eds., North Holland, Amsterdam, 1990.

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

[14] J. Demmel and W. Kahan, *Accurate singular values of bidiagonal matrices*, SIAM J. Sci. Statist. Comput., 11 (1990), pp. 873–912.

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

[16] E. Dowling, L. Ammann, and R. DeGroat, *A TQR-iteration based adaptive SVD for real time angle and frequency tracking*, IEEE Trans. Signal Processing, 42 (1994), pp. 914–926.

[17] D. FADEEV, V. KUBLANOVSKAYA, AND V. FADEEVA, *Sur les systèmes linéaires algébraiques de matrices rectangulaires et mal-conditionnées*, Colloq. Int. du C.N.R.S. Besançon 1966, No. 165 (1968), pp. 161–170.

[18] K. FERNANDO AND B. PARLETT, *Implicit Cholesky algorithm for singular values and vectors*, Tech. Report PAM-587, Center for Pure and Applied Mathematics, University of California, Berkeley, 1993.

[19] ———, *Accurate singular values and differential QD algorithms*, Numer. Math., 67 (1994), pp. 191–229.

[20] G. GOLUB, *Numerical methods for solving linear least squares problems*, Numer. Math., 7 (1965), pp. 206–216.

[21] G. GOLUB AND W. KAHAN, *Calculating the singular values and pseudo-inverse of a matrix*, SIAM J. Numer. Anal. Ser., B2 (1965), pp. 205–224.

[22] G. GOLUB AND C. REINSCH, *Singular value decomposition and least squares solutions*, Numer. Math., 14 (1970), pp. 403–420.

[23] G. GOLUB AND C. VAN LOAN, *Matrix Computations*, 2nd ed., The Johns Hopkins University Press, Baltimore, 1989.

[24] R. HANSON AND C. LAWSON, *Extensions and applications of the Householder algorithms for solving least squares problems*, Math. Comp., 23 (1969), pp. 787–812.

[25] V. HARI AND K. VESELIĆ, *On Jacobi methods for singular value decompositions*, SIAM J. Sci. Statist. Comput., 8 (1987), pp. 741–754.

[26] H. HONG AND C. PAN, *The rank-revealing QR decomposition and* SVD, Math. Comp., 58 (1992), pp. 213–232.

[27] R. HORN AND C. JOHNSON, *Matrix Analysis*, Cambridge University Press, London, 1985.

[28] I. IPSEN, *A close look at an extended Golub–Reinsch algorithm for computing singular values (abstract)*, in The Householder Symposium XI Meeting on Numerical Linear Algebra, Technical Report, Department of Mathematics, Linköping University, Sweden, 1990.

[29] J. LAGARIAS, *Monotonicity properties of the Toda flow, the QR flow, and subspace iteration*, SIAM J. Matrix Anal. Appl., 12 (1991), pp. 449–462.

[30] C. LAWSON AND R. HANSON, *Solving Least Squares Problems*, Prentice-Hall, New York, 1974.

[31] R. MATHIAS AND G. STEWART, *A block QR algorithm and the singular value decomposition*, Linear Algebra Appl., 182 (1993), pp. 91–100.

[32] M. MOONEN, P. VAN DOOREN, AND F. VANPOUKE, *On the QR algorithm and updating the* SVD *and* URV *decomposition in parallel*, Linear Algebra Appl., 188/189 (1993), pp. 549–568.

[33] B. PARLETT, *The Symmetric Eigenvalue Problem*, Prentice-Hall, New York, 1980.

[34] B. PARLETT AND W. POOLE, *A geometric theory for the QR, LU and power iterations*, SIAM J. Numer. Anal., 10 (1973), pp. 389–412.

[35] G. STEWART, *On an algorithm for refining a rank-revealing* URV *factorization and a perturbation theorem for singular values*, Tech. Report UMIACS-TR-91-38, Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, 1991.

[36] ———, *An updating algorithm for subspace tracking*, IEEE Transactions on Signal Processing, SP-40 (1992), pp. 1535–1541.

[37] K. VESELIĆ AND V. HARI, *A note on a one-sided Jacobi algorithm*, Numer. Math, 56 (1989), pp. 627–633.

[38] D. WATKINS, *Understanding the QR algorithm*, SIAM Rev., 24 (1982), pp. 427–440.

[39] D. WATKINS AND L. ELSNER, *Self-equivalent flows associated with the singular value decomposition*, SIAM J. Matrix Anal. Appl., 10 (1989), pp. 244–258.

[40] ———, *Convergence of algorithms of decomposition type for the eigenvalue problem*, Linear Algebra Appl., 143 (1991), pp. 19–47.

[41] J. WILKINSON, *The Algebraic Eigenvalue Problem*, Oxford University Press, Oxford, 1965.