UMIACS-TR-91-38 March 1991
CS-TR 2626 Revised May 1992

A Block QR Algorithm and the Singular Value Decomposition

R. Mathias*
G. W. Stewart†

ABSTRACT

In this note we consider an iterative algorithm for moving a triangular matrix toward diagonality. The algorithm is related to algorithms for refining rank-revealing triangular decompositions and in a variant form to the QR algorithm. It is shown to converge if there is a sufficient gap in the singular values of the matrix, and the analysis provides a new approximation theorem for singular values and singular subspaces.

†Institute for Mathematics and its Applications, University of University of Minnesota, Minneapolis, MN 55455

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This work was supported in part by the Air Force Office of Scientific Research under Contract AFOSR-87-0188.

A Block QR Algorithm and the Singular Value Decomposition

R. Mathias*
G. W. Stewart†

1. Introduction

Let R_0 be an $n \times n$ block triangular matrix of the form

$$R_0 = \begin{pmatrix} S_0 & H_0 \\ 0 & E_0 \end{pmatrix}, \tag{1.1}$$

where H_0 and E_0 are small compared to the smallest singular value of S_0. In this paper we will be concerned with the following two-stage iteration. For the first step, let Q_0 be a unitary matrix such that

$$R_1 \equiv R_0 Q_0 = \begin{pmatrix} S_1 & 0 \\ H_1 & E_1 \end{pmatrix} \tag{1.2}$$

is block lower triangular. Then let Q_1 be a unitary matrix such that

$$R_2 \equiv Q_1^{\mathrm{H}} R_1 = \begin{pmatrix} S_2 & H_2 \\ 0 & E_2 \end{pmatrix},$$

is block upper triangular, like R_1. The iteration is continued in the obvious way. Note that the matrices Q_0 and Q_1 are not unique; for example, Q_0 can be any unitary matrix of the form $Q = (Q_1\ Q_2)$, where the columns of Q_2 are orthogonal to the rows of $(S_0\ H_0)$.

This iteration arises in two connections. The one that motivated this paper is a refinement step in updating rank-revealing URV and ULV decompositions [3, 2]. Here H_0 is a vector, and E_0 is a scalar and the purpose of the iteration is to make H small, so that R_2 is nearer a diagonal matrix.

The second connection is with a variant of the (unshifted) QR algorithm for Hermitian matrices. Specifically, suppose that in addition to the above requirements, we demand that $R_0, R_2, \ldots$ be upper triangular and that $R_1, R_3, \ldots$ be lower triangular. Then

$$A_0 \equiv R_0^{\mathrm{H}} R_0 = (R_0^{\mathrm{H}} R_1) Q^{\mathrm{H}}$$

is a factorization of the Hermitian matrix A_0 into the product of a lower triangular matrix and a unitary matrix — the first step of the LQ variant of the QR algorithm.

If we perform the second step of the LQ algorithm by multiplying the factors in the reverse order, we get

$$\begin{aligned} Q^{\mathrm{H}}(R_0^{\mathrm{H}} R_1) &= R_1^{\mathrm{H}} R_1 \\ &= (R_1^{\mathrm{H}} P)(P^{\mathrm{H}} R_1) \\ &= R_2^{\mathrm{H}} R_2 \\ &\equiv A_2. \end{aligned}$$

Thus R_0 is the Cholesky factor of the Hermitian matrix A_0, and R_2 is the Cholesky factor of the matrix A_2 obtained by applying a step of the LQ algorithm to A_0. Since, under mild restrictions on A_0, the LQ algorithm converges to a diagonal matrix whose diagonal elements are the eigenvalues of A in descending order, the matrices $R_0, R_2, \ldots$ will converge to diagonal matrices whose diagonal elements are the singular values of R in descending order.

In this paper we will chiefly be concerned with the block variant of the algorithm, although our results will say something about the triangular LQ variant. In the next section we will analyze the convergence of the matrices H_i, an analysis which answers our concerns with the algorithm for refining rank-revealing decompositions. However, in the following section we will go on to show how our analysis can be applied to give a new approximation theorem for singular values and their associated subspaces.

Throughout the paper $\sigma_i(R)$ will denote the ith singular value of R in descending order. The quantity $\|R\|_2 = \sigma_1(R)$ is the spectral norm of R, $\|R\|$ denotes any unitarily invariant norm of R, and $\inf(R)$ is the smallest singular value of R.

We will later use the following lemma to obtain good relative bounds on all the singular values of R. It can be proved from the min-max characterization of singular values [1, Theorem 7.3.10], and a proof is outlined in [1, Problem 7.3.18].

Lemma 1.1. *Let A and B be n by n matrices. Then*

$$\inf(A)\sigma_i(B) \leq \sigma_i(AB) \leq \|A\|_2\sigma_i(B).$$

This result can be used to prove that for any unitarily invariant norm

$$\|AB\| \leq \|A\|_2\|B\|. \tag{1.3}$$

See, for example, [1, Example 7.4.54] for a proof.

2. Convergence of the Iteration

It turns out that the analysis of the passage from R_{2i} to R_{2i+1} of the refinement algorithm is *mutatis mutandis* the same as the analysis of the passage from R_{2i+1} to R_{2i+2}. We will therefore confine ourselves to the former, and in particular to the passage from R_0 to R_1. For notational convenience we will drop the subscripts and attach a prime to quantities associated with R_1.

Let

$$\begin{aligned} \epsilon &= \|E\|_2, \\ \eta &= \|H\|_2, \\ \gamma &= \inf(S), \end{aligned} \tag{2.1}$$

and assume that

$$\rho \equiv \frac{\epsilon}{\gamma} < 1. \tag{2.2}$$

Partition Q conformally with R and write

$$\begin{pmatrix} S & H \\ 0 & E \end{pmatrix} \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} = \begin{pmatrix} S' & 0 \\ H' & E' \end{pmatrix}. \tag{2.3}$$

Now it is easily verified from the orthogonality of Q that

$$\|Q_{22}\|_2 \leq 1$$

and that

$$Q_{21}^{\mathrm{H}} Q_{21} = I - Q_{11}^{\mathrm{H}} Q_{11} \quad \text{and} \quad Q_{12} Q_{12}^{\mathrm{H}} = I - Q_{11} Q_{11}^{\mathrm{H}}.$$

Since $I - Q_{11}^{\mathrm{H}} Q_{11}$ and $I - Q_{11} Q_{11}^{\mathrm{H}}$ have the same eigenvalues so do $Q_{21}^{\mathrm{H}} Q_{21}$ and $Q_{12} Q_{12}^{\mathrm{H}}$. Hence Q_{12} and Q_{21} have the same singular values, and so $\|Q_{12}\| = \|Q_{21}\|$ for any unitarily invariant norm. Consequently, from the equation

$$SQ_{12} + HQ_{22} = 0,$$

we obtain by two applications of (1.3)

$$\|Q_{21}\| = \|Q_{12}\| \leq \|S^{-1}\|_2 \|H\| \|Q_{22}\|_2 \leq \frac{\|H\|}{\gamma}.$$

It follows from the equation

$$H' = EQ_{21}$$

that
$$\|H'\| \le \frac{\epsilon\|H\|}{\gamma} = \rho\|H\|.$$
Since by assumption $\rho < 1$, the norm of H' is less than the norm of H by a factor of at least ρ.

But more is true. Let the quantities ϵ', η', and ρ' be defined in analogy with (2.1) and (2.2). We have already shown that $\eta' < \eta$. From (2.3) it follows that
$$E' = EQ_{22}, \tag{2.4}$$
and hence $\sigma_i(E') \le \sigma_i(E)$ by Lemma 1.1. In particular, $\epsilon' \le \epsilon$. Similarly, from $R' = RQ$ we have $R = R'Q^{\mathrm{H}}$, which implies
$$S'Q_{11}^{\mathrm{H}} = S. \tag{2.5}$$
Thus, $\sigma_i(S') \ge \sigma_i(S)$. From this it follows that $\rho' \le \rho < 1$. Since $\rho' < 1$, we may repeat the above argument to show that the passage from R_1 to R_2 will produce a matrix $H'' = H_2$ satisfying $\|H''\| \le \rho'\|H'\| \le \rho\|H'\|$; i.e., the left iteration reduces the norm of the off-diagonal block by at least ρ. The same is obviously true of subsequent iterations. Hence we have proved the following theorem. Here we drop the primes in favor of subscripts, with the convention that unadorned quantities refer to R_0.

Theorem 2.1. *Let the matrices R_i $(i = 0, 1, \ldots)$ be partitioned in analogy with* (1.1) *or* (1.2) *according as i is even or odd. Assume that*
$$\rho \equiv \frac{\|E\|_2}{\inf(S)} < 1. \tag{2.6}$$
Then
$$\begin{aligned}
\|H_i\| &\le \rho^i\|H\|, && \quad (2.7)\\
\sigma_j(S_{i+1}) &\ge \sigma_j(S_i), \qquad j = 1, \ldots, k, && \quad (2.8)\\
\sigma_j(E_{i+1}) &\le \sigma_j(E_i), \qquad j = 1, \ldots, n-k. && \quad (2.9)
\end{aligned}$$

The condition (2.6) is necessary; for if we start with the matrix
$$\begin{pmatrix} 1 & \eta \\ 0 & 1 \end{pmatrix},$$

then the first iteration produces the matrix

$$\begin{pmatrix} 1 & 0 \\ \eta & 1 \end{pmatrix},$$

and the next iteration restores the original matrix.

In practice one may not know $\inf(S)$ but may know $\sigma_k(R)$. In this case, one can still apply Theorem 2.1 since the theorem is true with ρ replaced by

$$\hat{\rho} = \frac{\|E\|_2}{\sigma_k(R) - \|H\|_2},$$

as we will now show. Suppose that $\hat{\rho} < 1$, then

$$\sigma_k(R) - \sigma_i(E) \geq \sigma_k(R) - \|E\|_2 > \|H\|_2. \tag{2.10}$$

We know that the singular values of R can be paired with those of S and E in such a way that the difference between the pairs is at most $\|H\|_2$. In view of (2.10) the k largest singular values of R must be paired with the singular values of S, and in particular, $|\sigma_i(R) - \sigma_k(S)| \leq \|H\|_2$ for some $i \leq k$. Thus, $\sigma_k(S) \geq \sigma_i(R) - \|H\|_2 \geq \sigma_k(R) - \|H\|_2$. Thus if $\hat{\rho} < 1$, then $\rho \leq \hat{\rho} < 1$, and the theorem holds with ρ replaced by $\hat{\rho}$.

3. Approximation Results

We now turn to the problem of assessing the accuracy of the singular values of

$$\hat{R} = \begin{pmatrix} S & 0 \\ 0 & E \end{pmatrix}$$

as approximations to singular values of R. We know from standard perturbation theory that they differ from singular values of R by quantities no greater than $\|H\|_2$. We will now show that under the condition (2.6) $\sigma_i(R)/\sigma_i(\hat{R}) = 1 + O(\|H\|_2^2)$.

The basic idea is to follow the iterates R_i of the iteration as the H_i approach zero. However, the approach is complicated by the fact that the R_i need not converge. Nonetheless, from the fact that $\|H_i\|_2 \to 0$ and from (2.8) and (2.9), we know that the singular values of S_i and E_i converge to those of R. Because

$$\sigma_k(S_i) \geq \sigma_k(S) > \sigma_1(E) \geq \sigma_1(E_i)$$

it follows that $\lim_{i\to\infty} \sigma_k(S_i) > \lim_{i\to\infty} \sigma_1(E_i)$, and hence

$$\begin{aligned} \sigma_j(R) &= \lim_{i\to\infty} \sigma_j(S_i), \qquad j = 1, \ldots, k, \\ \sigma_{k+j}(R) &= \lim_{i\to\infty} \sigma_j(E_i), \qquad j = 1, \ldots, n-k. \end{aligned}$$

We have shown in (2.4) that after one step of refinement $E_1 = EQ_{22}$. Since Q is unitary, $Q_{22}Q_{22}^{\mathrm{H}} = I - Q_{21}Q_{21}^{\mathrm{H}}$, and

$$\inf(Q_{22}) = (1 - \|Q_{21}\|_2^2)^{1/2} \geq (1 - (\eta/\gamma)^2)^{1/2}.$$

Hence, by Lemma 1.1,

$$\sigma_j(E) \geq \sigma_j(E_1) \geq \sigma_j(E)(1 - (\eta/\gamma)^2)^{1/2}.$$

Iterating this argument and recalling that $\|H_i\|_2 \leq \rho^i \eta$, we obtain the following lower bound on the smallest singular values of R:

$$\begin{aligned} \sigma_{k+j}(R) &= \lim_{i\to\infty} \sigma_j(E_i) \\ &\geq \lim_{i\to\infty} \left[\prod_{l=0}^{i} (1 - (\rho^l \eta/\gamma)^2)^{1/2}\right] \sigma_j(E) \\ &\geq \left(1 - (\eta/\gamma)^2 \sum_{l=0}^{\infty} \rho^{2l}\right)^{1/2} \sigma_j(E) \\ &= \left(1 - \frac{\eta^2}{\gamma^2(1-\rho^2)}\right)^{1/2} \sigma_j(E). \end{aligned}$$

One can show that

$$\left(1 - \frac{\eta}{\gamma^2(1-\rho^2)}\right)^{-1/2} \sigma_j(S) \geq \sigma_j(R) \geq \sigma_j(S)$$

in the same way.

We can also bound the perturbation of the singular subspace associated with the smallest $n-k$ singular values. Here it is convenient to choose a specific unitary $Q = Q_0$ of the form

$$Q = \begin{pmatrix} I & -P \\ P^{\mathrm{H}} & I \end{pmatrix} \begin{pmatrix} (I + PP^{\mathrm{H}})^{-1/2} & 0 \\ 0 & (I + P^{\mathrm{H}}P)^{-1/2} \end{pmatrix} \tag{3.1}$$

where $P = S^{-1}H$. It is easy to check that

$$\begin{aligned}
\|Q - I\|_2 &\le \left\| \begin{pmatrix} 0 & -P \\ P^{\mathrm{H}} & 0 \end{pmatrix} \right\|_2 \\
&\quad + \left\| \begin{pmatrix} (I + PP^{\mathrm{H}})^{-1/2} - I & 0 \\ 0 & (I + P^{\mathrm{H}}P)^{-1/2} - I \end{pmatrix} \right\|_2 \\
&\le 2\frac{\eta}{\gamma}. \qquad (3.2)
\end{aligned}$$

Let $Q_1, Q_2, Q_3, \ldots$ be defined analogously, and let $V_k = Q_0 Q_2 \cdots Q_{2k}$. Then by (2.7) and the analogue of (3.2),

$$\|V_{k+1} - V_k\|_2 = \|I - Q_{2(k+1)}\|_2 \le 2\rho^{2(k+1)}\eta/\gamma.$$

Thus, $V_0, V_1, \ldots$ has a limit V, which is also unitary. Similarly, the limit $U = Q_1 Q_3 \cdots$ exists and is unitary.

Let

$$U^{\mathrm{H}} R V = R_\infty = \begin{pmatrix} S_\infty & 0 \\ 0 & E_\infty \end{pmatrix},$$

and let

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}.$$

What we require is a bound on the canonical angles between the spaces spanned by

$$\begin{pmatrix} 0 \\ I \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} V_{21}^{\mathrm{H}} \\ V_{22}^{\mathrm{H}} \end{pmatrix}.$$

If Θ_{R} denotes the matrix of canonical angles, it is known that $\|\sin\Theta_{\mathrm{R}}\| = \|V_{21}\|$. To get a bound on $\|V_{21}\|$, recall that $\|(Q_i)_{21}\| \le \rho^i \|H\|/\gamma$. Hence by an easy induction

$$\|\sin\Theta_{\mathrm{R}}\| = \|V_{21}\| \le \frac{1}{(1-\rho^2)} \frac{\|H\|}{\gamma}.$$

In much the same way, one can obtain a bound for the matrix Θ_{L} of canonical angles for the left singular subspace corresponding to E_∞:

$$\|\sin\Theta_{\mathrm{L}}\| \le \frac{\rho}{(1-\rho^2)} \frac{\|H\|}{\gamma}.$$

The extra factor ρ arises because the first iteration, which does not affect U, reduces $\|H\|$ by a factor of ρ.

We now summarize what we have proved.

Theorem 3.1. *Let Θ_{L} and Θ_{R} be the matrices of canonical angles between the left and right singular subspace corresponding to the smallest $n-k$ singular values of R and $\hat{R}$. Under the hypotheses of Theorem 2.1,*

$$1 \geq \frac{\sigma_i(E)}{\sigma_{k+i}(R)} \geq \left(1 - \frac{\|H\|_2^2}{(1-\rho^2)\gamma^2}\right)^{1/2}, \quad i = 1, \ldots, n-k \tag{3.3}$$

$$1 \geq \frac{\sigma_i(R)}{\sigma_i(S)} \geq \left(1 - \frac{\|H\|_2^2}{(1-\rho^2)\gamma^2}\right)^{1/2}, \quad i = 1, \ldots, k \tag{3.4}$$

and for any unitarily invariant norm

$$\|\sin\Theta_{\mathrm{R}}\| \leq \frac{\|H\|}{(1-\rho^2)\gamma} \tag{3.5}$$

$$\|\sin\Theta_{\mathrm{L}}\| \leq \frac{\rho\|H\|}{(1-\rho^2)\gamma}. \tag{3.6}$$

There are a number of comments to be made about this theorem. First, the bounds (3.3) and (3.4) are remarkable in that they show that the *relative* error in the singular values is $O(\|H\|_2^2)$. Ordinarily, an off-diagonal perturbation of size H would obliterate singular values smaller than $\|H\|_2^2$. The fact that even the smallest singular values retain their accuracy is a consequence of the block triangularity of the starting matrix, as we shall see in a moment.

Second, Wedin ([5], [4, Theorem V.4.1]) has given a bound on the perturbation of singular subspaces. His bound does not assume that the matrix is in block triangular form, merely that the off diagonal blocks are small; but this bound, when specialized to our situation, gives the weaker inequality

$$\max\{\|\sin\Theta_R\|, \|\sin\Theta_L\|\} \leq \frac{\|H\|}{(1-\rho)\gamma}.$$

Third, the result can be cast as more conventional residual bounds for approximate null spaces. Specifically, suppose that for a given matrix B we have a unitary matrix $(V_1\ V_2)$ such that the residual

$$\|BV_2\|$$

is small. If $(U_1\ U_2)$ is a unitary matrix such that the column space of U_1 is the same as the column space of AV_1, then

$$U^{\mathrm{T}}AV = \begin{pmatrix} S & H \\ 0 & E \end{pmatrix},$$

where

$$\left\| \begin{pmatrix} H \\ E \end{pmatrix} \right\| = \|BV_2\|.$$

Thus if the singular values of S are not too small, the above theorem applies to bound the singular values and subspaces of B in terms of the residual norm $\|BV_2\|$.

Finally, we note that the approach yields approximation bounds for matrices of the form

$$\begin{pmatrix} \hat{S} & \hat{H} \\ \hat{G} & \hat{E} \end{pmatrix}, \qquad \|\hat{G}\| = O(\|\hat{H}\|)$$

by the expedient of first premultiplying to reduce the matrix to the form (1.1): i.e.,

$$\begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} \hat{S} & \hat{H} \\ \hat{G} & \hat{E} \end{pmatrix} = \begin{pmatrix} S & H \\ 0 & E \end{pmatrix}.$$

However, in this case the matrix

$$E = P_{21}\hat{H} + P_{22}\hat{E} = \hat{E} + O(\|\hat{H}\|_2^2)$$

is an $O(\|H\|_2^2)$ *additive* perturbation of of $\hat{E}$, so that the small singular values of $\hat{E}$ are no longer give $O(\|H\|_2^2)$ relative approximations to the small singular values of R. The block triangularity of R is really necessary.

Acknowledgements

The authors are grateful to W. Kahan for pointing out the relation of the iteration to the QR algorithm and to Marc Moonen for illuminating it.

References

[1] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1985.

[2] G. W. Stewart. *Updating a Rank-Revealing ULV Decomposition*. Technical Report CS-TR 2627, Department of Computer Science, University of Maryland, 1991.

[3] G. W. Stewart. *An Updating Algorithm for Subspace Tracking.* Technical Report CS-TR 2494, Department of Computer Science, University of Maryland, 1990. To appear in IEEE Transactions on Signal Processing.

[4] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory.* Academic Press, Boston, 1990.

[5] P.-Å. Wedin. Perturbation bounds in connection with singular value decomposition. *BIT*, 12:99–111, 1972.