A Convergence Analysis on URV Refinement

Limin Wu
Department of Applied Mathematics
Florida Institute of Technology
Melbourne, FL. 32901-6988
August 1997

Abstract

Recently, Stewart gave an algorithm for computing a rank revealing URV decomposition of a rectangular matrix. His method makes use of a refinement iteration to achieve an improved estimate of the smallest singular value and its corresponding singular vectors of the matrix. Here, a new proof is given for the convergence of the refinement iteration. This analysis is carried out under slightly weaker assumptions than those of Mathias and Stewart.

1. Introduction.

In [4], Stewart gave an updating algorithm for subspace tracking. His algorithm makes use of a refinement iteration, called URV refinement in the literature, to achieve an improved estimate of the smallest singular value and its corresponding singular vectors of a nonsingular upper triangular matrix. The URV refinement can be briefly described as follows.

Consider a real $n \times n$ nonsingular upper triangular matrix $R$. Let $R^{(0)} = R$ be partitioned as

$$R^{(0)} = \begin{bmatrix} S^{(0)} & h^{(0)} \\ 0 & e^{(0)} \end{bmatrix},$$

where $S^{(0)}$ is an $(n - 1) \times (n - 1)$ upper triangular matrix, $h^{(0)}$ is an $(n - 1)$-vector, and $e^{(0)}$ is a scalar. Then a sequence of orthogonal matrices, $Q^{(1)}, Q^{(2)}, \ldots, Q^{(2k-1)}, Q^{(2k)}$, each

1
determined as products of Givens rotations, is constructed such that, for $k \geq 1$,

$$
R^{(2k-1)} \equiv R^{(2k-2)}[Q^{(2k-1)}]^T = \begin{bmatrix}
S^{(2k-1)} & 0 \\
0 & e^{(2k-1)}
\end{bmatrix},
$$

(2)

$$
R^{(2k)} \equiv Q^{(2k)}R^{(2k-1)} = \begin{bmatrix}
S^{(2k)} & 0 \\
0 & e^{(2k)}
\end{bmatrix},
$$

(3)

where $S^{(2k-1)}, S^{(2k)}$ are $(n-1) \times (n-1)$ upper triangular matrices.

The URV refinement is identified by Chandrasekaran and Ipsen [1] as an incomplete version of the QR algorithm for computing the singular value decomposition of an upper triangular matrix. Stewart and Mathias [5, 3] discussed the URV refinement in a broader framework of block QR iterations, where $S^{(l)}$ are allowed to be $k \times k$ ($1 \leq k < n$) matrices, not necessarily upper triangular, and the $e^{(l)}$ are then $(n-k) \times (n-k)$ matrices. They established error bounds and derived convergence properties for the singular values of $S^{(l)}$ and $e^{(l)}$. In particular, for the special case considered in this paper, they proved that, if $|e^{(0)}|/\sigma_{\min}(S^{(0)}) < 1$, then the URV refinement computes the smallest singular value. We have used $\sigma_{\min}(\cdot)$ to denote the smallest singular value of a matrix. We will also use $\sigma_i(\cdot)$ to denote the $i$th largest singular value of a matrix and $\| \cdot \|$ to denote the 2-norm of a matrix throughout the paper.

To facilitate comparison with the new convergence proof given here, we restate a theorem from [3] for the case $k = n - 1$.

**Theorem 1 (Mathias and Stewart, 1993)** Let $S^{(l)}, e^{(l)},$ and $h^{(l)}$ be defined as in (1)-(3). For $l \geq 1$, we have

1. $|e^{(l)}| \leq |e^{(l-1)}|$;

2. $\sigma_j(S^{(l)}) \geq \sigma_j(S^{(l-1)}), \quad j = 1, \ldots, n - 1$;

3. $\|h^{(l)}\| \leq \rho^{(l)} \cdots \rho^{(0)} \|h^{(0)}\| \leq (\rho^{(0)})^l \|h^{(0)}\|$, where $\rho^{(l)} \equiv |e^{(l)}|/\sigma_{\min}(S^{(l)})$;

and if $\rho^{(0)} < 1$, then
4. \( \lim_{t \to \infty} |e^{(l)}| = \sigma_n(R) \);

5. \( \lim_{t \to \infty} \sigma_j(S^{(l)}) = \sigma_j(R), \ j = 1, \ldots, n - 1. \)

The assumption \( \rho^{(0)} < 1 \) is needed for the method of proof used to establish parts 4–5 of the above theorem, but is not a necessary condition for the convergence of the algorithm. An example which illustrates this fact is

\[
R = \begin{bmatrix}
1 & 0 & 10^{-6} \\
0 & 2 & 10^{-6} \\
0 & 0 & 10
\end{bmatrix}.
\]

For this example, a MATLAB implementation of the URV refinement yields an approximation to the smallest singular value of \( R \) after 14 iterations as \( e^{(28)} = 9.9 \cdots 948e-01 \) in double precision. This is very close to the smallest singular value of \( R, s_3 = 9.9 \cdots 950e-01 \) computed using the MATLAB SVD routine.

2. Convergence Analysis.

The singular value decomposition (SVD) of \( R^{(l)} \)'s provides a basis for our convergence analysis. Let \( R^{(0)} \) have the SVD \( R^{(0)} = U^{(0)}\Sigma[V^{(0)}]^T \), where \( \Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n) \), with

\[
\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0.
\]

Defining \( V^{(1)} = Q^{(1)}V^{(0)} \) and \( U^{(2)} = Q^{(2)}U^{(0)} \), then \( R^{(1)} \) and \( R^{(2)} \) have SVD's, \( R^{(1)} = U^{(0)}\Sigma[V^{(1)}]^T \) and \( R^{(2)} = U^{(2)}\Sigma[V^{(1)}]^T \), respectively. For \( k \geq 2 \), define

\[
V^{(2k-1)} = Q^{(2k-1)}V^{(2k-3)} = G^{(2k-1)}V, \text{ where } G^{(2k-1)} = Q^{(2k-1)} \cdots Q^{(1)},
\]

and

\[
U^{(2k)} = Q^{(2k)}U^{(2k-2)} = G^{(2k)}U, \text{ where } G^{(2k)} = Q^{(2k)} \cdots Q^{(2)}.
\]

Then \( R^{(2k-1)} \) and \( R^{(2k)} \) have SVD's:

\[
R^{(2k-1)} = U^{(2k-2)}\Sigma[V^{(2k-1)}]^T,
\]

\[
R^{(2k)} = U^{(2k)}\Sigma[V^{(2k)}]^T.
\]
and

\[ R^{(2k)} = U^{(2k)} \Sigma [V^{(2k-1)}]^T. \]

Denote \( R^{(l)} = [r_{ij}^{(l)}] \), \( V^{(2k-1)} = [v_{ij}^{(2k-1)}] \), and \( U^{(2k)} = [u_{ij}^{(2k)}] \); then \( r_{nn}^{(l)} = e^{(l)} \). Let \( v^{(0)} \) and \( u^{(0)} \) be the last columns of matrices \( V^{(0)} \) and \( U^{(0)} \), respectively. Let \( g^{(2k-1)} \) and \( g^{(2k)} \) contain the last rows of \( G^{(2k-1)} \) and \( G^{(2k)} \), respectively. Then the following theorem, first given in [6], holds.

**Theorem 2** Assuming in the URV refinement that \( r_{nn}^{(l)} \) are kept positive, then we have:

1. if \( v_{nn}^{(0)} \neq 0 \), then \( r_{nn}^{(l)} \) converges to \( \sigma_n \) monotonically;

2. if \( v_{nn}^{(0)} \neq 0 \) and \( \sigma_{n-1} > \sigma_n \), then \( \langle g^{(2k-1)}, v^{(0)} \rangle \equiv [g^{(2k-1)}]^T v^{(0)} \) and \( \langle g^{(2k)}, u^{(0)} \rangle \equiv [g^{(2k)}]^T u^{(0)} \) converge to \( \pm 1 \) monotonically;

3. if \( \langle g^{(2k-1)}, v^{(0)} \rangle \to 1 \), then \( v_{nn}^{(0)} \neq 0 \).

The condition \( \sigma_{n-1} > \sigma_n \) in part 2 of the theorem says that the smallest singular value \( \sigma_n \) is not repeated, that is, it is simple. Therefore it has unique left and right singular vectors associated with it. To prove this theorem, we need the following lemmas.

**Lemma 1** The smallest singular value of a square nonsingular triangular matrix is not greater than the absolute value of any diagonal element of the matrix.

**Proof.** see Lawson and Hanson [2, p.29, (6.3)].

**Lemma 2** The sequence \( \{ |r_{nn}^{(l)}| \} \) obtained from the URV refinement is nonincreasing and converges. In particular, \( \{ r_{nn}^{(l)} \} \) is nonincreasing and converges if \( r_{nn}^{(l)} \) are kept positive in the URV refinement.

**Proof.** Since orthogonal matrices preserve the 2-norm of vectors, we have

\[
\begin{bmatrix}
  r_{1n}^{(0)} \\
  \vdots \\
  r_{nn}^{(0)}
\end{bmatrix}
\geq
\begin{bmatrix}
  r_{nn}^{(0)} = \| (r_{n1}^{(1)}, \ldots, r_{nn}^{(1)})^T \| \\
  \vdots \\
  r_{nn}^{(0)}
\end{bmatrix}
\geq
\begin{bmatrix}
  r_{1n}^{(2)} \\
  \vdots \\
  r_{nn}^{(2)}
\end{bmatrix}
\geq
\begin{bmatrix}
  |r_{nn}^{(2)}| = \cdots > 0.
\end{bmatrix}
\]
Thus, $\{r_{nn}^{(l)}\}_{1}^{\infty}$ is nonincreasing and bounded below by 0. It follows that this sequence has a limit. If we choose Givens rotations in the refinement process in such a way that $r_{nn}^{(l)}$ are kept positive, then the sequence $\{r_{nn}^{(l)}\}_{1}^{\infty}$ has a limit.

**Lemma 3** If $r_{nn}^{(l)}$ are kept positive in the URV refinement, then

1. $u_{nn}^{(0)} = v_{nn}^{(2k-1)} = u_{nn}^{(2k)} = 0$, for $k \geq 1$, provided $v_{nn}^{(0)} = 0$.
2. $u_{nn}^{(0)} > 0$, $v_{nn}^{(2k-1)} > 0$, and $u_{nn}^{(2k)} > 0$, for $k \geq 1$, provided $v_{nn}^{(0)} > 0$.
3. $u_{nn}^{(0)} < 0$, $v_{nn}^{(2k-1)} < 0$, and $u_{nn}^{(2k)} < 0$, for $k \geq 1$, provided $v_{nn}^{(0)} < 0$.

**Proof.** Write the SVD of $R^{(0)}$ as

$$R^{(0)}V^{(0)} = U^{(0)}\Sigma. \tag{8}$$

Equating the corner elements at the $(n, n)$-position on both sides (8) gives

$$r_{nn}^{(0)}v_{nn}^{(0)} = \sigma_n u_{nn}^{(0)}. \tag{9}$$

Also, SVD (6) can be written as

$$[U^{(2k-2)}]^{T} R^{(2k-1)} = \Sigma[V^{(2k-1)}]^{T}. \tag{10}$$

Since $R^{(2k-1)}$ is of form (2), it is easy to see that

$$u_{nn}^{(2k-2)}r_{nn}^{(2k-1)} = \sigma_n v_{nn}^{(2k-1)}, \text{ for } k \geq 1. \tag{11}$$

Similarly, writing (7) as

$$R^{(2k)}V^{(2k-1)} = U^{(2k)}\Sigma, \tag{12}$$

we have

$$r_{nn}^{(2k)}v_{nn}^{(2k-1)} = u_{nn}^{(2k)}\sigma_n, \text{ for } k \geq 1. \tag{13}$$

Since we have assumed that $r_{nn}^{(l)} > 0$ and $\sigma_n > 0$, the conclusions are easily drawn using equations (9), (11), and (13). \[\square\]
Lemma 4  If \(v^{(0)}_{nn} \neq 0\) and \(r^{(l)}_{nn}\) are kept positive in the URV refinement, then \(\{v^{(2k-1)}_{nn}\}^\infty_1\) and \(\{u^{(2k)}_{nn}\}^\infty_1\) converge monotonically to the same nonzero limit.

Proof. We first assume \(v^{(0)}_{nn} > 0\). According to Lemma 3, \(u^{(2k)}_{nn}\) are also positive for \(k \geq 1\). By manipulating (11) and (13) we obtain

\[
\frac{u^{(2k-2)}_{nn}}{u^{(2k)}_{nn}} = \frac{\sigma^2_n}{r^{(2k-1)}_{nn} r^{(2k)}_{nn}}.
\]

By Lemma 1 and Lemma 2 we have \(r^{(2k)}_{nn} \geq \sigma_n\) and \(r^{(2k-1)}_{nn} \geq r^{(2k)}_{nn}\). It follows that the right hand side of (14) is less than or equal to one. Therefore, \(\{u^{(2k)}_{nn}\}^\infty_1\) is a nondecreasing sequence. The orthogonality of \(U^{(2k)}\) means that \(u^{(2k)}_{nn}\) is bounded from above by one. Hence, \(\{u^{(2k)}_{nn}\}^\infty_1\) has a positive limit. Also, the relation (11) tells us that \(\{v^{(2k-1)}_{nn}\}^\infty_1\) has the same limit as \(\{u^{(2k)}_{nn}\}^\infty_1\) does. For the case \(v^{(0)}_{nn} < 0\) the proof is similar. ■

Proof of Theorem 2.

Part 1. Since both \(r^{(l)}_{nn}\) and \(u^{(2k)}_{nn}\) converge, taking the limit on both sides of (14) yields \(\lim_{k \to \infty} r^{(l)}_{nn} = \sigma_n\) and the convergence is monotone by Lemma 2.

Part 2. By (11) and (15), \(\langle g^{(2k-1)}, v^{(0)} \rangle = v^{(2k-1)}_{nn}\) and \(\langle g^{(2k)}, u^{(0)} \rangle = u^{(2k)}_{nn}\). We prove that \(\lim_{k \to \infty} v^{(2k-1)}_{nn} = \lim_{k \to \infty} u^{(2k)}_{nn} = \pm 1\) under the assumption. Suppose \(\lim_{k \to \infty} u^{(2k)}_{nn} = a\).

Apparenty \(|a| \leq 1\). By Lemma 4, \(\lim_{k \to \infty} v^{(2k-1)}_{nn} = a\). Equating the last rows in both sides of (12) gives

\[
r^{(2k)}_{nn} (v^{(2k-1)}_{n1}, \ldots, v^{(2k-1)}_{nn}) = (\sigma_1 u^{(2k)}_{n1}, \ldots, \sigma_n u^{(2k)}_{nn}).
\]

Taking the 2-norm of the above equation and then squaring both sides gives

\[
\langle r^{(2k)}_{nn} \rangle^2 = \sigma_1^2 (u^{(2k)}_{n1})^2 + \cdots + \sigma_n^2 (u^{(2k)}_{nn})^2.
\]

Rewriting the above equation and considering the ordering of \(\sigma_i\)'s we get

\[
\langle r^{(2k)}_{nn} \rangle^2 - \sigma_n^2 (u^{(2k)}_{nn})^2 = \sigma_1^2 (u^{(2k)}_{n1})^2 + \cdots + \sigma_{n-1}^2 (u^{(2k)}_{n,n-1})^2.
\]
Taking the limit on both sides of the above equation yields

$$\sigma_n^2(1 - a^2) \geq \sigma_{n-1}^2(1 - a^2)$$

Since we have assumed $\sigma_{n-1} > \sigma_n$, the only way that this inequality can hold is if $a = \pm 1$.

Part 3. Since $v_{nn}^{(2k-1)} = \langle g^{(2k)}, v^{(0)} \rangle \to 1$, in view of Lemma 3, it is obvious that $v_{nn}^{(0)} \neq 0$. ■

Note It is a consequence of the standard theory of inner product space that

$$\lim_{k \to \infty} \langle g^{(2k-1)}, v^{(0)} \rangle = 1 \text{ if and only if } \lim_{k \to \infty} \| g^{(2k-1)} - v^{(0)} \| = 0.$$  

Since $v_{nn}^{(0)} \neq 0$ is vital for the convergence of the URV refinement when $\sigma_n$ is simple, it is desirable to know under what conditions the nonsingular upper triangular matrix $R$ has a simple smallest singular value and nonzero $v_{nn}^{(0)}$ in its SVD. A sufficient condition is given by the following theorem. In the proof of the theorem we drop the superscript $(0)$ for $R, V$, and $U$ and related quantities. Let $R_1$ be the matrix consisting of the first $n - 1$ columns of $R$.

**Lemma 5** $\sigma_{n-1} \geq \sigma_{\min}(R_1) \geq \sigma_n$.

**Proof.** see Lawson and Hanson [2, p.26, (5.12)]. ■

**Theorem 3** If $\sigma_{\min}(S) > \sigma_n$, then $\sigma_n$ is simple and $v_{nn} \neq 0$.

**Proof.** Since $\sigma_{\min}(S) = \sigma_{\min}(R_1)$, it follows by Lemma 5 that $\sigma_n$ is simple. To prove the second part, we will show that $v_{nn} = 0$ implies $\sigma_{\min}(S) = \sigma_n$. First, since $\sigma_{\min}(S) = \sigma_{\min}(R_1)$, the inequality $\sigma_{\min}(S) \geq \sigma_n$ follows from Lemma 5. We now establish the reverse inequality. Let $R$ have the SVD $R = U\Sigma V^T$. Let $u$ be the last column of $U$, $r$ the last column of $R$. Write the SVD of $R$ as

$$(15) \quad U^T R = \Sigma V^T.$$
Equating the corner elements at the (n,n) position on both sides of (15) gives $u^T r = \sigma_n v_{nn}$.

The assumption of $v_{nn} = 0$ implies $u^T r = 0$. Now consider the equation

(16) \[ u^T R R^T u = u^T U \Sigma^2 U^T u, \]

or the equivalent form

(17) \[(u^T R)(u^T R)^T = (u^T U) \Sigma^2 (u^T U)^T = e_n^T \Sigma^2 e_n,\]

where $e_n$ is the unit vector with one in the last component. Since $u^T r = 0$, (17) becomes

(18) \[(u^T R_1)(u^T R_1)^T = \sigma_n^2.\]

Letting $w = (u_{1n}, \ldots, u_{n-1,n})^T$, we have $u^T R_1 = w^T S$. Thus (18) can be further reduced to

(19) \[(w^T S)(w^T S)^T = \sigma_n^2, \text{ or } \|S^T w\| = \sigma_n.\]

By Lemma 3, $u_{nn} = v_{nn} = 0$, thus $\|w\| = \|u\| = 1$. Therefore we have

(20) \[\sigma_{\text{min}}(S) = \sigma_{\text{min}}(S^T) = \min_{\|x\| = 1} \|S^T x\| \leq \|S^T w\| = \sigma_n.\]

This completes the proof that $\sigma_{\text{min}}(S) = \sigma_n$. \[\blacksquare\]

**Corollary 1** If $\sigma_{\text{min}}(S) > \sigma_{\text{min}}(R)$, then the URV refinement converges.

**Remark** The assumption of $\sigma_{\text{min}}(S^{(0)}) > \sigma_{\text{min}}(R^{(0)})$ in the above corollary is weaker than the assumption $\sigma_{\text{min}}(S^{(0)}) > |e^{(0)}|$ used in Theorem 1 because $|e^{(0)}| \geq \sigma_{\text{min}}(R^{(0)})$. We may note, however, that if $\sigma_{\text{min}}(S^{(0)}) > \sigma_{\text{min}}(R^{(0)})$, then the URV refinement will produce, for sufficiently large $l$, an $|e^{(l)}|$ and $\sigma_{\text{min}}(S^{(l)})$ such that $\sigma_{\text{min}}(S^{(l)}) > |e^{(l)}|$. This follows from $\sigma_{\text{min}}(S^{(l)}) \geq \sigma_{\text{min}}(S^{(0)})$ (Theorem 1, Part 2) and the fact that $\sigma_{\text{min}}(S^{(0)}) > \sigma_{\text{min}}(R^{(0)})$ implies $|e^{(l)}| \to \sigma_{\text{min}}(R^{(0)})$ (Corollary 1).
3. Conclusions

In this paper we have shown that $v_n^{(0)} \neq 0$ is sufficient for the convergence of the sequence $\{r_n^{(l)}\}$ to the smallest singular value of $R^{(0)}$ in the URV refinement (Theorem 2, Part 1). The following matrix

$$
R^{(0)} = \begin{bmatrix}
1 & 0 & 0 \\
0 & 9 & 1 \\
0 & 1 & 10
\end{bmatrix}
$$

serves as a convenient example for which $v_{33}^{(0)} = 0$ and the sequence $\{r_{33}^{(l)}\}$ fails to converge to $R^{(0)}$. It is unknown whether $v_n^{(0)} \neq 0$ is a necessary condition for $r_n^{(l)} \to \sigma_{\min}(R^{(0)})$.

Acknowledgments. The author thanks Professor Charles Fulton for helpful discussions and Professor Gary Howell for a suggestion leading to a simplified proof of Theorem 2.

References

[1] S. Chandrasekaran and I. C. F. Ipsen, Analysis of a QR Algorithm for Computing Singular Values, *SIAM J. Matrix Anal. Appl.*, Vol.16, 2:520-535 (1995).

[2] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Classics in Applied Mathematics, SIAM, Philadelphia, 1995.

[3] R. Mathias and G.W. Stewart, A Block QR Algorithm and the Singular Value Decomposition, *Linear Algebra Appl.*, 182:91-100 (1993).

[4] G. W. Stewart, An Updating Algorithm for Subspace Tracking, *IEEE Trans. Signal Processing*, 40:1535-1541 (1992).

[5] G. W. Stewart, On an Algorithm for Refining a Rank-Revealing URV Factorization and a Perturbation Theorem for Singular Values, Tech. Report UMIACS-TR-91-38, University of Maryland, College Park, MD, 1991.
[6] L. Wu, *Regularization Methods and Algorithms for Least Squares and Kronecker Product Least Squares Problems*, Ph.D. thesis, Department of Applied Mathematics, Florida Institute of Technology, Melbourne, FL, 1997.