On the Singular Value Decomposition of Triangular Matrices

Shivkumar Chandrasekaran
Ilse C.F. Ipsen
Department of Computer Science
Yale University
New Haven, CT 06520, USA

Our objective is to obtain more insight into the properties of the singular value decomposition of triangular matrices. These properties can then be used to analyse the convergence of a QR algorithm that computes singular values of triangular matrices.

First we derive relationships between certain angles formed by the left and by the right singular vectors. Suppose the singular values of the triangular matrix are ordered in descending order. We show that a subspace spanned by the first k left singular vectors is closer to the canonical space than the subspace spanned by the corresponding right singular vectors. This difference between the angles of left and right singular vectors increases with the distance between the kth and $(k+1)$st singular values.

Second we present a simple perturbation result that determines when an off-diagonal block of a triangular matrix may be eliminated without causing a large relative change in the smallest singular values. It turns out that a small off-diagonal block may be safely eliminated if one of the adjacent diagonal blocks is well-conditioned and its singular values are well-separated from the remaining ones.

We use $\|\cdot\|$ to represent the Euclidean two-norm. The identity matrix of order k is denoted by I_k.

The Singular Value Decomposition

Let $R = U\Sigma V^T$ be the singular value decomposition (SVD) of the non-singular block upper triangular matrix

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

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 Σ is a diagonal matrix whose diagonal contains the singular values of R,

$$\sigma_1 \geq \ldots \geq \sigma_k \geq \sigma_{k+1} \geq \ldots \geq \sigma_n,$$

The work presented in this paper was supported by NSF grant CCR-9102853.

in descending order,

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

The first theorem shows that if the singular values are well-separated then the left singular vector matrix is almost always closer to canonical form than the right singular vector matrix. By 'canonical form' we mean an orthogonal matrix whose columns span the canonical space $\begin{pmatrix} I_k \\ 0 \end{pmatrix}$, that is, a matrix of the form $\begin{pmatrix} Z \\ 0 \end{pmatrix}$ where Z is orthogonal.

Theorem 1 *If the respective principal submatrices U_{22} and V_{11} of the singular vector matrices U and V are non-singular then*

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

where $\theta_{v,k}$ is the angle between the leading k columns of V and the canonical subspace $\begin{pmatrix} I_k \\ 0 \end{pmatrix}$; and $\theta_{u,k}$ is the corresponding angle for U.

Proof: From $RV = U\Sigma$ one gets

$$R_{22}V_{21} = U_{21}\Sigma_1, \qquad R_{22}V_{22} = U_{22}\Sigma_2,$$

and from $U^T R = \Sigma V^T$,

$$U_{11}^T R_{11} = \Sigma_1 V_{11}^T, \qquad U_{12}^T R_{11} = \Sigma_2 V_{12}^T.$$

Furthermore, $R_{21} = 0$ in $R = U\Sigma V^T$ yields

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

If V_{11} and U_{22} are non-singular then the last equality can be written as

$$U_{22}^{-1}U_{21} = -\Sigma_2 V_{12}^T V_{11}^{-T}\Sigma_1^{-1}.$$

In order to derive a geometric interpretation for this equality, we first establish some properties of orthogonal matrices.

The CS decomposition, Theorem 2.6.1 in [1], of an orthogonal matrix

$$Z \quad = \quad \begin{array}{c} \\ k \\ \\ \end{array}\!\!\begin{array}{c} \quad k \\ \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix}, \end{array}$$

implies for non-singular Z_{11}

$$\|Z_{12}\| = \|Z_{21}\| = \sin\theta, \qquad \|Z_{11}^{-1}Z_{12}\| = \|Z_{22}^{-1}Z_{21}\| = \tan\theta,$$

where θ is some angle, and

$$\sqrt{1 - \frac{1}{\|Z_{11}^{-1}\|^2}} = \|Z_{21}\|.$$

These equalities give, together with $\sigma_{min}(Z_{11}) = 1/\|Z_{11}^{-1}\|$,

$$\sin\theta = \sqrt{1 - \sigma_{min}^2(Z_{11}^T I_k)}.$$

But the square-root term represents the distance between the column space of $\begin{pmatrix} Z_{11} \\ Z_{21} \end{pmatrix}$ and the canonical space $\begin{pmatrix} I_k \\ 0 \end{pmatrix}$, Corollary 2.6.2 in [1]. Hence, the angle θ from the CS decomposition of Z must be the largest principal angle, Section 12.4.3 in [1], between these two subspaces.

Applying this to the above inequality relating the two singular vector matrices of R yields

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

where $\theta_{u,k}$ is the largest principal angle between the leading k columns of U and the canonical subspace $\begin{pmatrix} I_k \\ 0 \end{pmatrix}$; while $\theta_{v,k}$ is the analogous angle for V. $\blacksquare$

Deflation Criterion

Stewart [4] shows that if $\|R_{12}\| + \|R_{22}\| \leq \sigma_k$ then

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

The theorem below presents a deflation criterion that holds without any assumptions on the size of $\|R_{12}\|$.

Theorem 2 *If V_{11} is non-singular and if* $\text{gap}_k = \sigma_{min}(R_{11}) - \|R_{22}\| > 0$ *then the singular values $\sigma_j(R_{22})$ of R_{22} satisfy*

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

Proof: Let $R^T = Q\hat{R}$ be the QR decomposition of the block lower triangular matrix R^T, where

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}, \qquad \hat{R} = \begin{pmatrix} \hat{R}_{11} & \hat{R}_{12} \\ & \hat{R}_{22} \end{pmatrix}$$

are respective orthogonal and block upper triangular matrices. 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 $|\sigma_j(A+E) - \sigma_j(A)| \leq \|E\|$, Corollary 8.3.2 in [1], 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 for the singular values of R_{22} and $\hat{R}_{22}$

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

Moreover,

$$\hat{R}_{12} = Q_{21}^T R_{22}^T, \qquad R_{12}^T = Q_{21}\hat{R}_{11}, \qquad R_{11}^T = Q_{11}\hat{R}_{11}$$

imply $\hat{R}_{12} = Q_{11}^T R_{11}^{-1} R_{12} R_{22}^T$ and

$$\|\hat{R}_{12}\| \le \gamma_k \|R_{12}\|, \qquad \text{where} \qquad \gamma_k = \|R_{11}^{-1}\|\,\|R_{22}\|.$$

Therefore

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

We also make use of the fact that the γ_k are monotonically non-increasing,

$$\hat{\gamma}_k \le \gamma_k, \qquad \text{where} \qquad \hat{\gamma}_k = \|\hat{R}_{11}^{-1}\|\,\|\hat{R}_{22}\|,$$

due to

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

which represent special cases of the monotonicity properties of eigenvalues during subspace iteration [2].

Let $R^{(0)} = R$ and for $i \ge 0$ define the iteration

$$R^{(i)^T} = Q^{(i)} R^{(i+1)},$$

where $Q^{(i)}$ is orthogonal and $R^{(i+1)}$ is block upper triangular. Then $\gamma_k^{(i+1)} \le \gamma_k^{(i)}$, and the difference between two successive iterations is

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

As for the difference between iteration $i+2$ and i, we employ the idea in Stewart's proof in [3],

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

When V_{11} is non-singular then the singular values of $R_{22}^{(i)}$ converge to the singular values $\sigma_{k+1}, \ldots, \sigma_n$ of Σ_2 as $i \to \infty$. By assumption $\gamma_k < 1$, hence extrapolation to the limit gives

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

as $\sum_{l=1}^{\infty} \gamma_k^l = \frac{1}{1-\gamma_k} - 1$. Using $\gamma_k = \|R_{11}^{-1}\|\,\|R_{22}\|$ and $\sigma_{min}(R_{11}) = 1/\|R_{11}^{-1}\|$ yields

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

■

Theorem 2 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_{k+1} to high relative accuracy whenever the off-diagonal block is small and the singular values of R_{11} are much larger than those of R_{22}. The condition $\text{gap}_k > 0$ is not satisfied, for instance, for a graded matrix whose elements increase in size towards the bottom, regardless of how small R_{12} is.

In general, the bound in Theorem 2 suggests using the simple deflation criterion

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

in order 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, Corollary 8.3.2 in [1],

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

Relative accuracy η for all singular values of R_{22} is guaranteed if

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

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

References

[1] G. Golub and C. van Loan, *Matrix Computations*, The Johns Hopkins Press, 1989.

[2] J. Lagarias, *Monotonicity properties of the Toda flow, the QR flow, and subspace iteration*, SIAM J. Matrix Anal. Appl., 12 (1991), pp. 449–62.

[3] G. Stewart, *An updating algorithm for subspace tracking*, IEEE Transactions on Signal Processing, SP-40 (1992), pp. 1535–41.

[4] ——, *Updating a rank-revealing ULV decomposition*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 494–99.