Computation of the Generalized Singular Value Decomposition Using Mesh-Connected Processors

R.P. Brent, F.T. Luk, and C. Van Loan

TR 83-563
July 1983

Department of Computer Science
Cornell University
Ithaca, New York 14853

Invited paper to appear in Proceedings of SPIE Volume 431, Real Time Signal Processing VI (1983).

Computation of the Generalized Singular Value Decomposition Using Mesh-Connected Processors

Richard P. Brent

Centre for Mathematical Analysis
Australian National University
Canberra, A.C.T. 2601
Australia

Franklin T. Luk

Department of Computer Science
Cornell University
Ithaca, New York 14853
U.S.A.

Charles Van Loan

Department of Computer Science
Cornell University
Ithaca, New York 14853
U.S.A.

ABSTRACT

This paper concerns the systolic array computation of the generalized singular value decomposition. Numerical algorithms for both one- and two-dimensional systolic architectures are discussed.

Keywords and Phrases: Systolic arrays, QR-decomposition, singular value decomposition, generalized singular value decomposition, real-time computation, VLSI.

Invited paper to appear in *Proceedings of SPIE Volume 431, Real Time Signal Processing VI* (1983).

Introduction

Two of the most important ways to decompose a given matrix $A \in \mathbb{R}^{m \times n}$ $(m \geq n)$ are the Q-R factorization:

$$A = QR \ , \tag{1}$$

where $Q \in \mathbb{R}^{m \times n}$ has orthonormal columns and $R \in \mathbb{R}^{n \times n}$ is upper triangular, and the singular value decomposition (SVD):

$$A = U \Sigma V^T \ , \tag{2}$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal, and $\Sigma \in \mathbb{R}^{m \times n} = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$, with $\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0$ and $r = \mathrm{rank}(A)$. See Golub and Van Loan [1] and Dongarra et al.[2] for details.

The systolic array computation of these decompositions has recently attracted a great deal of attention. QR-arrays are discussed in Bojanczyk, Brent and Kung [3], Gentleman and Kung [4] and Heller and Ipsen [5]; SVD arrays in Brent and Luk [6], Brent, Luk and Van Loan [7], Finn, Luk and Pottle [8], Heller and Ipsen [9] and Schreiber [10]. In this paper we discuss the systolic array computation of the generalized singular value decomposition (GSVD). It has been suggested (see Speiser and Whitehouse [11]) that real-time computation of this decomposition is important in modern signal processing.

The GSVD amounts to a simultaneous diagonalization of a pair of matrices $A \in \mathbb{R}^{m \times n}$ $(m \geq n)$ and $B \in \mathbb{R}^{p \times n}$:

$$\begin{pmatrix} U^T & 0 \\ 0 & V^T \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} X = \begin{pmatrix} D_A \\ D_B \end{pmatrix} , \tag{3}$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{p \times p}$ are orthogonal, $X \in \mathbb{R}^{n \times n}$ is nonsingular, $D_A = \mathrm{diag}(\alpha_1, \ldots, \alpha_n) \geq 0$, $D_B = \mathrm{diag}(\beta_1, \ldots, \beta_q) \geq 0$ and $q = \min\{p, n\}$. We call (α_i, β_i) a singular value pair of A and B. Note that when B is square and nonsingular, the singular values of AB^{-1} are α_i / β_i, for $i = 1, \cdots, n$, and when $B = I_n$ these ratios are just the singular values of A. For a general B, we may refer to α_i / β_i as the generalized singular values of A with respect to B, although some of these values may be infinite or undefined. The use of singular value pairs, however, avoids the distinction between A and B. The GSVD was first introduced by Van Loan [12] and further discussed in Paige and Saunders [13]. The decomposition is useful for certain constrained and generalized least squares problems (see Golub and Van Loan [1]).

We briefly discuss the computation of the GSVD. Suppose that the null spaces of A and B intersect trivially, i.e., $N(A) \bigcap N(B) = \{0\}$. Let

$$E \equiv \begin{pmatrix} A \\ B \end{pmatrix} , \tag{4}$$

and compute its QR-factorization:

$$E = QR \ .$$

By assumption, the matrix R is nonsingular. Partition Q in the form

$$Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} ,$$

such that $Q_1 \in \mathrm{R}^{m \times n}$ and $Q_2 \in \mathrm{R}^{p \times n}$. Then we can find orthogonal matrices $U \in \mathrm{R}^{m \times m}$, $V \in \mathrm{R}^{p \times p}$ and $W \in \mathrm{R}^{n \times n}$ such that

$$\begin{pmatrix} U^T & 0 \\ 0 & V^T \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} W = \begin{pmatrix} C \\ S \end{pmatrix} , \tag{5}$$

where $C = \mathrm{diag}(c_1, \ldots, c_n) \geq 0$, $S = \mathrm{diag}(s_1, \ldots, s_q) \geq 0$ and $C^T C + S^T S = I_n$. The decomposition (5) is referred to as the CS-decomposition. It says that the SVD's of the blocks in a partitioned orthonormal matrix are related. The CS-decomposition first appears in Stewart [14], where it is pointed out that the result is implicit in Davis and Kahan [15]. Van Loan [16] shows how this decomposition can be used to analyze certain important problems involving orthogonal matrices. If we set

$$D_A = C \ , \ D_B = S \text{ and } X = R^{-1} W \ ,$$

we obtain a GSVD of A and B.

If the null spaces of A and B intersect nontrivially, or nearly so, then it is advisable to compute an SVD of the matrix E:

$$\begin{pmatrix} A \\ B \end{pmatrix} = Q \Sigma Z^T \equiv \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Z_1^T \\ Z_2^T \end{pmatrix} .$$

Here, $\Sigma_r = \mathrm{diag}(\sigma_1, \ldots, \sigma_r) \in \mathrm{R}^{r\times r}$, $Q_{11} \in \mathrm{R}^{m\times r}$, $Q_{21} \in \mathrm{R}^{p\times r}$, $Z_1 \in \mathrm{R}^{n\times r}$ and $r = \mathrm{rank}(E)$.

Let

$$\begin{pmatrix} \tilde{U}^T & 0 \\ 0 & \tilde{V}^T \end{pmatrix} \begin{pmatrix} Q_{11} \\ Q_{21} \end{pmatrix} \tilde{W} = \begin{pmatrix} \tilde{C} \\ \tilde{S} \end{pmatrix}$$

be a CS-decomposition of Q_{11} and Q_{21}. Then

$$A = Q_{11}\Sigma_r Z_1^T = \tilde{U}(\tilde{C},0) \begin{pmatrix} \tilde{W}^T\Sigma_r & 0 \\ 0 & I_{n-r} \end{pmatrix} Z^T$$

and

$$B = Q_{21}\Sigma_r Z_1^T = \tilde{V}(\tilde{S},0) \begin{pmatrix} \tilde{W}^T\Sigma_r & 0 \\ 0 & I_{n-r} \end{pmatrix} Z^T .$$

A GSVD results by setting $D_A = (\tilde{C},0)$, $D_B = (\tilde{S},0)$ and $X = Z\begin{pmatrix} \Sigma_r^{-1}\tilde{W} & 0 \\ 0 & I_{n-r} \end{pmatrix}$.

From the above discussion, we see that the key problem confronting us is the systolic array calculation of the CS-decomposition.

Stewart's algorithm

We desire a CS-decomposition of a partitioned orthonormal matrix

$$Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} ,$$

where $Q_1 \in \mathrm{R}^{m \times n}$ $(m \geq n)$ and $Q_2 \in \mathrm{R}^{p \times n}$. First, an SVD of Q_1 may be determined via standard techniques:

$$U^T Q_1 W = C \ .$$

Since

$$Q_1^T Q_1 + Q_2^T Q_2 = I_n \ ,$$

the nonnull columns of the matrix

$$\overline{Q}_2 = Q_2 W$$

are orthogonal. Suppose that $\overline{Q}_2$ has rank $= r$ and that its first r columns are nonzero. These columns can be normalized to yield

$$\overline{Q}_2 = (V_1, 0) \begin{pmatrix} S_1 \\ 0 \end{pmatrix} ,$$

where $V_1 \in \mathrm{R}^{p \times r}$ is orthonormal and $S_1 = \mathrm{diag}(s_1, \ldots, s_r) \geq 0$. Let $V = (V_1, V_2) \in \mathrm{R}^{p \times p}$ be an orthogonal matrix. Then we have

$$V^T Q_2 W = \begin{pmatrix} S_1 \\ 0 \end{pmatrix} \equiv S \ ,$$

an SVD of Q_2.

Unfortunately, the preceding procedure is numerically unsound. Troubles may arise when some columns of $\overline{Q}_2$ have euclidean lengths less than $\epsilon^{1/2}$, where ϵ denotes the machine precision. Numerical examples are given in Stewart [17,18]. To simplify our presentation, let us assume from here on that Q_2 has full column rank, i.e.,

$$\mathrm{rank}(Q_2) = n \leq p \ . \tag{6}$$

Stewart [17,18] presents the following cleanup procedure:

1. Determine an orthogonal matrix J such that the columns of $\overline{Q}_2 J$ can be normalized to give a matrix V whose columns are then orthogonal to working accuracy.

2. Determine an orthogonal matrix K such that $K^T CJ$ is diagonal.

If we replace W by WJ and U by UK, and normalize the columns of $\overline{Q}_2 J$ to get V, we obtain

$$\begin{pmatrix} U^T Q_1 \\ V^T Q_2 \end{pmatrix} W = \begin{pmatrix} K^T CJ \\ V^T \overline{Q}_2 \end{pmatrix} .$$

Since $K^T CJ$ and $V^T \overline{Q}_2$ are diagonal, we have computed a CS-decomposition of Q.

Stewart chooses J and K by working with the matrix

$$F \equiv \overline{Q}_2^T \overline{Q}_2 ,$$

and using the Jacobi method, as implemented by Rutishauser [20], to determine J such that $J^T FJ$ is diagonal. Stewart then shows why we may take $K = J$, so long as certain unnecessary rotations are not performed in the Jacobi method. Specifically, a Jacobi rotation R_{ij} in the (i,j)-plane will be suppressed if

$$c_i + c_j \leq \tau ,$$

where c_i and c_j are the i-th and j-th diagonal elements of C and τ is some preset tolerance. A value of $\tau = 0.7$ is proposed, for if

$$c_i + c_j = 0.7 ,$$

then the error made in accepting $R_{ij}^T C R_{ij}$ as a diagonal matrix is roughly equal to the error made in accepting the i-th and j-th columns of $\overline{Q}_2$ as orthogonal. Finally, Stewart proves that, because of the suppression, the diagonal entries of C are effectively unchanged in the passage to $J^T CJ$.

Linear arrays

Brent and Luk [6] present a systolic array of $O(n)$ linearly-connected processors for computing an SVD of an $l \times n$ matrix, say M. Their array implements a one-sided orthogonalization method due to Hestenes [21]. The idea is to determine an orthogonal matrix V such that the non-null columns of MV are mutually orthogonal. These columns are normalized to give a matrix $\tilde{U}$ with orthonormal columns and a nonnegative diagonal matrix Σ. We have thus determined an SVD of M:

$$M = \tilde{U} \Sigma V^T .$$

The orthogonal transformation V is constructed as a sequence of plane rotations; the rotations are generated to orthogonalize column pairs of M. Hence the Hestenes method is mathematically equivalent to the serial Jacobi procedure for finding an eigenvalue decomposition of $M^T M$. For the sake of parallel computing, Brent and Luk discard the classical scheme of rotating column pairs in the order:

$$(1,2),(1,3), \ldots ,(1,n),(2,3), \ldots ,(2,n),(3,4), \ldots ,(3,n), \ldots ,(n-1,n) ,$$

in preference for a new ordering that allows $\lfloor n/2 \rfloor$ simultaneous rotations. Their new ordering is amply illustrated by the $n = 8$ case:

$$\begin{aligned}
(p,q) \;=\; & (1,2) , (3,4) , (5,6) , (7,8) , \\
& (1,4) , (2,6) , (3,8) , (5,7) , \\
& (1,6) , (4,8) , (2,7) , (3,5) , \\
& (1,8) , (6,7) , (4,5) , (2,3) , \\
& (1,7) , (8,5) , (6,3) , (4,2) , \\
& (1,5) , (7,3) , (8,2) , (6,4) , \\
& (1,3) , (5,2) , (7,4) , (8,6) .
\end{aligned}$$

Note that the rotation pairs associated with each "row" of the above can be calculated concurrently. Brent and Luk [22] conjecture that this Jacobi approach would require $O(\log n)$ sweeps for convergence. Their algorithm for computing an SVD of an $l \times n$ matrix thus requires $O(nl \log n)$ time.

We may compute the GSVD using the linear systolic array of Brent and Luk [6] as follows:

1. Compute an SVD of

$$\begin{pmatrix} A \\ B \end{pmatrix} = Q \Sigma Z^T \equiv \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} \Sigma Z^T$$

2. Compute an SVD of

$$Q_1 = UCW^T ,$$

and apply the appropriate transformations to get

$$\overline{Q}_2 \equiv Q_2 W .$$

3. Initiate Stewart's algorithm. (We note that the Jacobi procedure applied to F is equivalent to the Hestenes method applied to $\overline{Q}_2$.)

Our procedure requires time $O((m+p)n\log n)$.

Quadratic arrays

An array for computing an SVD of an $l \times l$ matrix is proposed in Brent, Luk and Van Loan [7]. It requires $O(l^2)$ processors and $O(l \log l)$ time to execute. The array implements a two-sided Jacobi procedure that is detailed in Forsythe and Henrici [23]. In essence, the off-diagonal elements of the given matrix are reduced to zero by a sequence of plane rotations that are determined by solving carefully chosen two-by-two SVD's. The algorithm is very similar to the classical Jacobi algorithm for the symmetric eigenvalue problem, for which a systolic array has been proposed by Brent and Luk [22]. Briefly, the new ordering of Brent and Luk [6], illustrated in the previous section, is extended in an obvious manner to allow the simultaneous computations of $\lfloor l/2 \rfloor$ two-by-two SVD's. In addition, a staggering of computations allows the execution of the equivalence transformations without requiring that the rotation parameters be broadcasted. For details see Brent et al. [7,22].

If we want an SVD of a matrix $M \in \mathrm{R}^{m \times n}$, where $m,n \leq l$, we feed the matrix

$$\hat{M} = \begin{pmatrix} M & 0 \\ 0 & 0 \end{pmatrix} \in \mathrm{R}^{l \times l}$$

into the array of Brent, Luk and Van Loan [7]. An SVD:

$$\hat{M} = \begin{pmatrix} U & 0 \\ 0 & I_{l-m} \end{pmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V & 0 \\ 0 & I_{l-n} \end{pmatrix}^T$$

will emerge, and we see that $M = U \Sigma V^T$, as desired.

Let us point out how we can compute a nonsquare CS-decomposition using a "square" l-by-l hardware. Suppose that $Q_1 \in \mathrm{R}^{m \times n}$, $Q_2 \in \mathrm{R}^{p \times n}$, $Q_1^T Q_1 + Q_2^T Q_2 = I_n$ and $l \geq m,n,p$. If

$$\hat{Q}_1 = \begin{pmatrix} Q_1 & 0 \\ 0 & 0 \end{pmatrix} \in \mathrm{R}^{l \times l} \quad \text{and} \quad \hat{Q}_2 = \begin{pmatrix} Q_2 & 0 \\ 0 & 0 \end{pmatrix} \in \mathrm{R}^{l \times l} ,$$

then

$$\hat{Q}_1^T \hat{Q}_1 + \hat{Q}_2^T \hat{Q}_2 = \begin{pmatrix} I_n & 0 \\ 0 & 0 \end{pmatrix} .$$

It is not hard to show that there exist orthogonal matrices of the form

$$\hat{U}_1 = \begin{pmatrix} U_1 & 0 \\ 0 & I_{l-m} \end{pmatrix}, \ \hat{U}_2 = \begin{pmatrix} U_2 & 0 \\ 0 & I_{l-p} \end{pmatrix} \text{ and } \hat{W} = \begin{pmatrix} W & 0 \\ 0 & I_{l-n} \end{pmatrix},$$

such that

$$\hat{U}_1^T \hat{Q}_1 \hat{W} = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix}, \ \hat{U}_2^T \hat{Q}_2 \hat{W} = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} \text{ and } C^T C + S^T S = I_n .$$

Thus, applying Stewart's algorithm to $\hat{Q}_1$ and $\hat{Q}_2$ will produce a CS-decomposition of Q_1 and Q_2 .

We now outline how we may compute a GSVD of A and B using a QR-array, a matrix-matrix multiply array (see, e.g., Kung and Leiserson [24]) and an SVD array:

1. Compute a QR-decomposition of

$$\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} R .$$

2. Compute $T = Q_2^T Q_2$.

3. Set $\hat{Q}_1 = \begin{pmatrix} Q_1 & 0 \\ 0 & 0 \end{pmatrix}$, $\hat{Q}_2 = \begin{pmatrix} Q_2 & 0 \\ 0 & 0 \end{pmatrix}$ and $\hat{T} = \begin{pmatrix} T & 0 \\ 0 & 0 \end{pmatrix}$,

   so that they are all $l \times l$ matrices.

4. Compute an SVD of $\hat{Q}_1$ and apply the appropriate transformations to $\hat{Q}_2$ and $\hat{T}$.

5. Initiate Stewart's procedure.

The complete procedure requires time $O(l \log n)$.

Acknowledgements

The work of the second author was supported in part by the National Science Foundation under grant MCS-8213718, and in part by the Centre for Mathematical Analysis, Australian National University.

<u>References</u>

1. G.H. Golub and C.F. Van Loan, *Matrix Computations*, The Johns Hopkins Press, Baltimore, 1983, to appear.
2. J.J. Dongarra, C.B. Moler, J.R. Bunch and G.W. Stewart, *LINPACK Users' Guide*, SIAM, Philadelphia, 1979.
3. A. Bojanczyk, R.P. Brent and H.T. Kung, "Numerically stable solution of dense systems of linear equations using mesh-connected processors," *SIAM J. Sci. Statist. Comput.*, to appear.
4. W.M. Gentleman and H.T. Kung, "Matrix triangularization by systolic arrays", *Proc. SPIE Symp. 1981, Vol. 298, Real-Time Signal Processing IV*, 1981.
5. D.E. Heller and I.C.F. Ipsen, "Systolic networks for orthogonal decompositions", *SIAM J. Sci. Statist. Comput. 4* (1983), 261-269.
6. R.P. Brent and F.T. Luk, "A systolic architecture for the singular value decomposition", Technical Report TR-CS-82-522, Department of Computer Science, Cornell University, 1982.
7. R.P. Brent, F.T. Luk and C.F. Van Loan, "Computation of the singular value decomposition using mesh-connected processors," Technical Report TR-CS-82-528, Department of Computer Science, Cornell University, 1983.
8. A.M. Finn, F.T. Luk, and C. Pottle, "Systolic array computation of the singular value decomposition", *Proc. SPIE Symp. East 1982, Vol. 341, Real-Time Signal Processing V* (1982), 35-43.
9. D.E. Heller and I.C.F. Ipsen, "Systolic networks for orthogonal equivalence transformations and their applications", *Proc. 1982 Conf. on Advanced Research in VSLI*, MIT, Cambridge, Massachusetts (1982), 113-122.
10. R. Schreiber, "A systolic architecture for singular value decomposition", *Proc. 1st Internat. Coll. on Vector and Parallel Computing in Scientific Applications*, Paris, France (Mar. 1983), to appear.
11. J.M. Speiser and H.J. Whitehouse, "A survey of systolic arrays for signal processing," *IEEE Electro '83*, 1983, to appear.
12. C.F. Van Loan, "Generalizing the singular value decomposition," *SIAM J. Numer. Anal. 13* (1976), 76-83.
13. C.C. Paige and M.A. Saunders, "Towards a generalized singular value decomposition," *SIAM J. Numer. Anal. 18* (1981), 398-405.
14. G.W. Stewart, "On the perturbation of pseudo-inverses, projections and linear least squares problems," *SIAM Rev. 19* (1977), 634-662.
15. C. Davis and W.M. Kahan, "The rotation of eigenvectors by a perturbation. III," *SIAM J. Numer. Anal. 7* (1970), 1-46.
16. C.F. Van Loan, "On Stewart's singular value decomposition for partitioned orthogonal matrices," Technical Report STAN-CS-79-767, Department of Computer Science, Stanford University, 1979.
17. G.W. Stewart, "A method for computing the generalized singular value decomposition," in *Lecture Notes in Mathematics 973 : Matrix Pencils*, B. Kagstrom and A. Ruhe, eds., Springer-Verlag, New York (1983), 207-220.
18. G.W. Stewart, "Computing the CS decomposition of a partitioned orthonormal matrix," Technical Report 1159, Department of Computer Science, University of Maryland, College Park, 1982.

19. C.F. Van Loan, "A general matrix eigenvalue problem," *SIAM J. Numer. Anal. 12* (1975), 819-834.

20. H. Rutishauser, "The Jacobi method for real symmetric matrices", in *Handbook for Automatic Compuation, Vol. 2 (Linear Algebra)*, J.H. Wilkinson and C. Reinsch, eds., Springer-Verlag, New York, 1971, 202-211.

21. M.R. Hestenes, "Inversion of matrices by biorthogonalization and related results", *J. Soc. Indust. Appl. Math 6* (1958), 51-90.

22. R.P. Brent and F.T. Luk, "A systolic architecture for almost linear-time solution of the symmetric eigenvalue problem", Technical Report TR-CS-82-525, Department of Computer Science, Cornell University, 1982.

23. G.E. Forsythe and P. Henrici, "The cyclic Jacobi method for computing the principal values of a complex matrix", *Trans. Amer. Math. Soc. 94* (1960), 1-23.

24. H.T. Kung and C.E. Leiserson, "Algorithms for VLSI processor arrays," in *Introduction to VLSI Systems* (by C. Mead and L. Conway), Addison-Wesley, Reading, Massachusetts, 1980, 271-292.