

Signal processing computations using the generalized singular value decomposition

Jeffrey M. Speiser  
 Naval Ocean Systems Center  
 San Diego, California 92152

Charles Van Loan  
 Department of Computer Science, Cornell University  
 Ithaca, New York 14853

Abstract

The ordinary Singular Value Decomposition (SVD) is widely used in statistical and signal processing computation, both for the insight it provides into the structure of a linear operator, and as a technique for reducing the computational word length required for least-squares solutions and certain Hermitian eigensystem decompositions by roughly a factor of two, via computing directly on a data matrix, rather than on the corresponding estimated correlation or covariance matrix. Although the SVD has long been utilized as a method of off-line or non-real-time computation, parallel computing architectures for its implementation in near real time have begun to emerge.

The Generalized Singular Value Decomposition (GSVD) bears the same relationship to the computation of certain Hermitian generalized eigensystem decompositions that the ordinary SVD bears to the corresponding ordinary eigensystem decompositions.

This paper discusses methods for computing the GSVD via a sequence of more familiar computations and indicates the relation of the GSVD to the MUSIC algorithm of R. Schmidt.

Introduction

Numerical linear algebra forms a large part of the computational load for modern signal processing. Previous papers have discussed the applications<sup>[1-2]</sup> and selected parallel implementations<sup>[1-8]</sup> for matrix-vector multiplication, matrix multiplication, orthogonal-triangular factorization and least-squares solution, eigensystem solution, and singular value decomposition.

This paper considers the less familiar tasks of generalized eigensystem solution and computation of the generalized singular value decomposition. In order to have a representative signal processing task to guide our computational considerations, we begin by examining a representative classical and modern beamforming/direction finding techniques.

1. Beamforming

Classical beamforming

Classical delay-and-sum or phase-shift and sum beamforming provides both a baseline against which to compare more recent techniques and a convenient context for introducing the signal propagation models used in beamforming. Figure 1 illustrates a signal source located at position  $\underline{R}$ , radiating a signal  $s(t)$ . Receiving elements (antennas or acoustic transducers, as appropriate) are located at  $\underline{R}_1, \dots, \underline{R}_n$ .

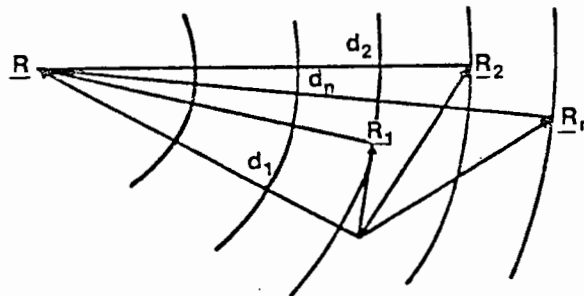


Figure 1. Propagation geometry for an arbitrary array.

This analysis will assume straight line propagation and equal attenuation to all array elements. The signal component at the  $n$ th array element is

$$s_n(t) = s(t - \|\underline{R} - \underline{R}_n\|/c) \tag{1-1}$$

It is desired to form

$$b(t) = \sum_n s_n(t + \|\underline{R} - \underline{R}_n\|/c) = \sum_n s_n(t - T_n) \tag{1-2}$$

$$\|\underline{R} - \underline{R}_n\| = (\|\underline{R} - \underline{R}_n, \underline{R} - \underline{R}_n\|)^{1/2} = (\|\underline{R}\|^2 - 2\|\underline{R}, \underline{R}_n\| + \|\underline{R}_n\|^2)^{1/2} \quad (1-3)$$

From equation (1-3), it may be seen that a frequency-domain or phase-shift and sum beamformer requires primarily matrix-vector multiplication for its implementation. For a broadband signal,  $S_n(f)$  may be interpreted as the Fourier transform of the  $n$ th element output at frequency  $f$ , and such a matrix-vector multiplication will be required at each resolved frequency. For a narrowband application,  $S_n(f)$  may be replaced by the complex envelope of the  $n$ th element's output. This complex envelope may be obtained from a quadrature demodulator for each element output, with a common local oscillator used for all the quadrature demodulators.

Classical beamformers are relatively simple and robust with respect to assumptions about the signal and noise model, but they have several well-known limitations: a) The array geometry and aperture shading force a tradeoff between resolution and sidelobe suppression. b) It is difficult to do any meaningful form of shading for an array with an irregular or time-variant geometry.

Adaptive beamformers provide for the cancellation of strong interfering signals. Typical recent approaches to adaptive beamforming use the direct solution of a linearly constrained least-squares problem to provide a minimum variance distortionless-response estimate of the arriving signal, with rapid adaptation<sup>[9]</sup>. However, adaptive interference cancellation per se does not provide improved spatial localization, and it is therefore of interest to consider methods for improving angular resolution. Several such techniques have been derived by analogy with corresponding spectral analysis methods such as the maximum likelihood, maximum entropy, and Pisarenko methods. The spectrum analysis methods have been surveyed by Kay and Marple<sup>[10]</sup> and maximum-likelihood and maximum entropy methods have been applied to beamforming/direction finding by McDonough<sup>[11]</sup>. The eigenvector spectrum analysis method of Pisarenko<sup>[12]</sup> has been extended and applied to direction finding by Schmidt<sup>[13]</sup>, and others<sup>[14-18]</sup>, and appears to offer better resolution than previous methods.

While classical beamforming provides no mechanism for utilizing prior knowledge about the signal and noise fields, the high resolution spectrum analysis methods and their beamforming analogs incorporate prior information concerning the signals, and Schmidt's MUSIC algorithm also incorporates prior information about the noise field.

### The MUSIC direction finding method

The MUSIC method assumes that we have available estimates or measurements of two spatial covariance matrices: The total signal-plus-noise-plus-interference spatial covariance matrix, and the noise-plus-interference covariance matrix. The underlying signal model and assumptions are:

$$\underline{x} = \underline{A} \underline{f} + \underline{w}$$

where

$\underline{x}$  = data vector = set of spatial observations (complex envelopes or one frequency bin)

$\underline{f}$  = vector of source complex amplitudes

$\underline{A}$  = phase shift matrix corresponding to the propagation geometry.

$\underline{R}_x = E \underline{x} \underline{x}^H$  = spatial covariance matrix for total array output

$\underline{R}_n = (1/\lambda) E \underline{w} \underline{w}^H$  = scaled noise covariance matrix

$\underline{P} = E \underline{f} \underline{f}^H$  = source covariance matrix

$\underline{x}^H$  denotes the Hermitian transpose of  $\underline{x}$ , and  $E$  denotes statistical expectation.

It is assumed that:

- 1)  $\underline{P}$  is positive definite
- 2)  $\underline{A}$  is of full rank
- 3) The number of arriving signal wavefronts  $D$ , including multipaths, is less than the number of array elements  $M$ .

In this case, Schmidt has shown that the multiplicity  $N$  of the smallest generalized eigenvalue of  $\underline{R}_x \underline{e} = \lambda \underline{R}_n \underline{e}$  is equal to  $M - D$ , and that the corresponding generalized eigenvectors are orthogonal to the propagation phase shift vectors from the source arrival directions.

The procedure for MUSIC is then:

- a) Collect data, estimate  $\underline{R}_x = E \underline{x} \underline{x}^H$
- b) Solve the generalized eigensystem  $\underline{R}_x \underline{e} = \lambda \underline{R}_n \underline{e}$
- c) Find the multiplicity  $\hat{N}$  of the smallest generalized eigenvalue  $\lambda_{\min}$  and estimate the number of arriving signal wavefronts as  $\hat{D} = M - \hat{N}$

d) Evaluate the so-called direction of arrival spectrum  $P_{MU}(\theta)$

$$P_{MU}(\theta) = 1 / \left( \sum_k | \underline{e}_k \cdot \underline{a}(\theta) |^2 \right)$$

where the  $\underline{e}_k$  are an orthonormal basis for the generalized eigenvectors corresponding to the minimum generalized eigenvalues (the "noise subspace") and  $\underline{a}(\theta)$  is the classical steering vector associated with a plane or spherical wave arriving from direction  $\theta$ .

e) Pick  $\hat{D}$  peaks of  $P_{MU}(\theta)$ , and identify  $\theta_1, \dots, \theta_{\hat{D}}$  as the arrival directions. The arrival phase shift matrix can then be estimated as  $A = (\underline{a}(\theta_1), \dots, \underline{a}(\theta_{\hat{D}}))$ .

f) The source covariance matrix can be estimated as

$$P = (A^H A)^{-1} A^H (R_x - \lambda_{\min} R_N) A (A^H A)^{-1}$$

The diagonal entries of  $P$  provide estimates of arrival power from the selected directions.

A performance comparison of MUSIC with other direction finding techniques using simulated data<sup>[13]</sup> reproduced in Figure 2 indicated substantial improvement in resolution compared to classical beamforming and the maximum likelihood method, without the bias errors which can be observed in the maximum entropy method.

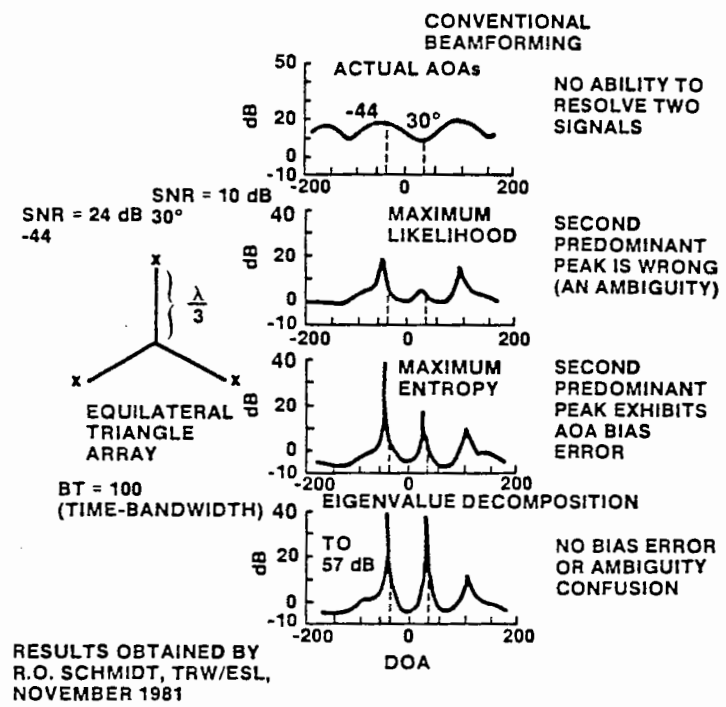


Figure 2. Performance comparison of algorithms for direction finding.

Spatial covariance matrix estimation

In order to avoid measurement of the noise + interference only covariance matrix, some researchers have used a guess for  $R_N$ , typically assuming incoherent noise from element to element<sup>[17]</sup>. However, this can provide false prior to the MUSIC algorithms, and can induce a bias error<sup>[17,18]</sup>.

For this reason, it is preferable to estimate both  $R_x$  and  $R_N$  from observed data. In an active echolocation system,  $R_N$  can be estimated while not transmitting. In a passive system it may be desirable to estimate  $R_N$  using data from adjacent frequency bins.

In order to consider estimation of  $R_x$ , we let  $x(t,m)$  denote the output (complex envelope or Fourier component) of the  $m$ th array element at observation sample  $t$ , where  $m = 1 \dots M$  and  $t = 1 \dots T$

As an estimate of  $R_x$ , we use  $\hat{R}(m_1, m_2) = (1/T) \sum_{t=1}^T x(t, m_1) x^*(t, m_2)$

where the asterisk \*, denotes complex conjugation.

If we define  $A(t, m) = (T^{-1/2}) x^*(t, m)$

then  $\hat{R}_x = A^H A$ . Do not confuse this matrix A with the arrival phase shift matrix used earlier in the paper. For the remainder of the paper, A will denote a scaled, conjugated matrix of observed signal-plus-noise-plus-interference data. The corresponding noise-plus-interference matrix data for use in estimating the spatial covariance of the noise-plus-interference will be denoted by B.

Computation Task

The preceding discussion leads to the following computational requirement:

- a) find the smallest generalized eigenvalue of  $A^H A \underline{e} = \lambda B^H B \underline{e}$
- b) find an orthonormal basis for generalized eigenvectors corresponding to this minimum generalized eigenvalue. In Schmidt's notation this orthonormal basis is used to form the columns of a matrix called  $E_N$ , and  $P_{MU}(\theta) = 1/(a^H(\theta) E_N E_N^H a(\theta))$ .

If we explicitly multiply to form  $A^H A = \hat{R}_x$  and  $B^H B = \hat{R}_N$ , then the computation is called a generalized eigensystem – more specifically a Hermitian symmetric generalized eigensystem. If we would like to perform the computation using A and B directly, we have a generalized singular value problem.

Since high speed parallel implementations of the computation may force the use of a short computation word length, it is desirable to avoid the potential loss of accuracy in the explicit computation of  $A^H A$  and  $B^H B$ .

For an ordinary eigensystem of the form  $A^H A \underline{e} = \lambda \underline{e}$ , it is well known that one can utilize the singular value decomposition  $A = U S V^H$ , where U and V are unitary matrices and S is diagonal. The columns of V are an orthonormal set of eigenvectors of  $A^H A$ , and the squares of the diagonal entries of S are the eigenvalues of  $A^H A$ , since the  $A^H A = V S^H U^H U S V^H = V S^2 V^H$ .

Numerically stable methods for computing the singular value decomposition are well known, and a substantial number of approaches to its parallel computation are now known<sup>[5-8]</sup>. The remainder of this paper deals with the extension of these techniques to the generalized singular value decomposition in the form needed for the MUSIC algorithm.

In section 2 we review the generalized singular value decomposition. A stable algorithm for computing  $\lambda_{min}$  and  $E_N$  is then given in section 3. Some concluding remarks are offered in section 4.

2. The generalized singular value decomposition

The approach that we have taken to the multiple signal classification problem leads to a generalized eigenproblem of the form

$$A^H A \underline{e} = \lambda B^H B \underline{e} \tag{2-1}$$

This is referred to as the generalized singular value problem because it involves aspects of the (ordinary) singular value problem

$$A^H A \underline{x} = \lambda \underline{x} \quad A \in C^{m \times n} \tag{2-2}$$

and the generalized eigenvalue problem

$$G \underline{x} = \lambda H \underline{x} \quad G, H \in C^{n \times n} \tag{2-3}$$

Stable methods for (2-2) avoid the formation of  $A^H A$  because cross products can destroy valuable information. For example, if

$$A = \begin{bmatrix} 1 & 1 \\ \sqrt{\epsilon} & 0 \\ 0 & \sqrt{\epsilon} \end{bmatrix} \quad \text{rank}(A) = 2$$

where  $\epsilon$  is the machine precision, then the completed version of  $A^H A$ , which we designate by  $\Omega[A^H A]$ , satisfies

$$\Omega[A^H A] = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

a singular matrix. For a given machine precision, cross-product techniques (because they square condition) narrow the class of solvable problems. This can be an acute problem in many real-time signal processing situations where speed, size, weight, and power constraints may force the use of a short word length. See [19, pp. 142ff] for a discussion of the numerical issues.

Stable methods for (2-3) avoid the formation of  $H^{-1}$ . For example, attempts to convert the symmetric generalized eigensystem  $Gx = \lambda Hx$  to the standard eigenproblem  $(H^{-1}G)x = \lambda x$  are prone to serious inaccuracies since  $H^{-1}G$  is generally nonsymmetric and the resulting eigensystem can be sensitive to perturbations. See [19, pp. 251-266, 313-321].

If  $H$  is ill-conditioned, then numerical difficulties may be encountered even with symmetric reductions such as

$$(H^{-1/2} G H^{-1/2}) H^{1/2} x = \lambda H^{1/2} x$$

or

$$(L^{-1} G (L^H)^{-1}) L^H x = \lambda L^H x$$

where

$$H = LL^H \dots$$

Our experience with singular value and generalized eigenvalue problems suggests two things if we are to stably deal with (2-1):

- avoid the cross products of  $A^H A$  and  $B^H B$
- avoid inverses that involve  $B$  such as  $(B^H B)^{-1}$ .

This turns out to be possible by carefully exploiting the generalized singular value decomposition.

Theorem 2.1 (GSVD)

If  $A \in C^{m \times n}$  ( $m \geq n$ ) and  $B \in C^{p \times n}$  ( $p \geq n$ ) are given, then there exist unitary  $U_A \in C^{m \times m}$  and  $U_B \in C^{p \times p}$  and nonsingular

$$X = [x_1, \dots, x_n] \quad x_i \in C^n$$

such that

$$U_A^H A X = D_A \equiv \text{diag}(c_1, \dots, c_n)$$

$$U_B^H B X = D_B \equiv \text{diag}(s_1, \dots, s_n)$$

It can be assumed that  $0 < c_1 < \dots < c_n$  and  $s_1 > \dots > s_n > 0$  with  $c_1^2 + s_1^2 = 1$ .

Proof

See [20]. The theorem also holds if  $p < n$  but since we are not interested in this case, it has been suppressed.  $\square$

If we set  $B = I$  in the GSVD then  $U_B^H X = D_B$

and so

$$U_A^H A U_B = U_A^H (X D_B^{-1}) = D_A D_B^{-1} \equiv \Sigma$$

This is just the "ordinary" singular value decomposition (SVD). See [19].

The GSVD displays quantities that are very important to us in "generalized MUSIC" algorithm. Since

$$X^H (A^H A) X = \text{diag}(c_1^2, \dots, c_n^2) \in R^{n \times n}$$

and

$$X^H (B^H B) X = \text{diag}(s_1^2, \dots, s_n^2) \in R^{n \times n}$$

it follows that

$$\begin{aligned} A^H A x_i &= c_i^2 y_i \\ B^H B x_i &= s_i^2 y_i \end{aligned} \quad i = 1, \dots, n$$

where

$$(X^H)^{-1} \equiv Y = [y_1, \dots, y_n]$$

Thus, if

$$\mu_i = c_i/s_i \quad s_i \neq 0$$

then

$$A^H A \underline{x}_i = \mu_i^2 B^H B \underline{x}_i .$$

This is why the  $\underline{x}_i$  (the columns of  $X$ ) are referred to as "generalized singular vectors" and the  $\mu_i$  as "generalized singular values".

Given our remarks earlier, let us evaluate the following procedure for computing generalized singular values and vectors:

(a) Compute the QR factorization

$$B = QR \quad Q^H Q = I_n, R \in \mathbb{C}^{m \times n} \text{ upper triangular.}$$

(b) Solve  $A_1 R = A$  for  $A_1 \in \mathbb{C}^{m \times n}$ .

(c) Compute the SVD

$$U^H A_1 V = \text{diag}(\mu_1, \dots, \mu_n) \quad U, V \text{ unitary, } \mu_i \geq 0$$

(d) Solve  $RX = V$  for  $X = [\underline{x}_1, \dots, \underline{x}_n]$

Note that in principle this procedure should work so long as  $B$  has full rank:

$$X^H (B^H B) X = (BR^{-1}V)^H (BR^{-1}V) = V^H Q^H Q V = I_n$$

$$X^H (A^H A) X = (AR^{-1}V)^H (AR^{-1}V) = V^H (A^H A) V = \text{diag}(\mu_i^2).$$

Unfortunately, if  $B$  is ill-conditioned, then  $R$  will be ill-conditioned and a totally inaccurate  $A_1$  can result. A way around this is given in the next section.

### 3. An algorithm [21]

A stable rendering of the GSVD can be had by the following two-step procedure:

Step 1. Compute the SVD

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

where  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ ,  $Z \in \mathbb{C}^{n \times n}$  is unitary, and  $Q_1 \in \mathbb{C}^{m \times n}$  and  $Q_2 \in \mathbb{C}^{p \times n}$  satisfy  $Q_1^H Q_1 + Q_2^H Q_2 = I_n$ .

Step 2. Compute unitary  $U_A$ ,  $U_B$ , and  $V$  such that

$$U_A^H Q_1 V = \text{diag}(c_1, \dots, c_n) \quad c_i \geq 0$$

$$U_B^H Q_2 V = \text{diag}(s_1, \dots, s_n) \quad s_i \geq 0$$

Step 3. Set  $X = [\underline{x}_1, \dots, \underline{x}_n] = Z \Sigma^{-1} V$

The "double" SVD computed in Step 2 is referred to as the "CS decomposition" of the pair  $(Q_1, Q_2)$ . The stable computation of these SVD's is nontrivial; see [21].

It is easy to show using  $Q_1^H Q_1 + Q_2^H Q_2 = I_n$  that

$$c_i^2 + s_i^2 = 1 \quad c = 1, \dots, n .$$

Moreover,

$$U_A^H A X = U_A^H (Q_1 \Sigma Z^H) X = U_A^H Q_1 V = \text{diag}(c_1, \dots, c_n)$$

$$U_B^H B X = U_B^H (Q_2 \Sigma Z^H) X = U_B^H Q_2 V = \text{diag}(s_1, \dots, s_n)$$

Thus, the method indeed produces the GSVD. The generalized singular values are given by the quotients

$$\mu_i = c_i/s_i = c_i/\sqrt{1 - c_i^2} \quad c_i \neq 1$$

Now in our beamforming application, we do not need the "full" GSVD. Instead, for some designated subset of generalized singular values  $(\mu_{i_1}, \dots, \mu_{i_n})$  we need orthonormal vectors  $\underline{w}_1, \dots, \underline{w}_k$  such that

$$\text{Span}\{\underline{w}_1, \dots, \underline{w}_k\} = \text{Span}\{\underline{x}_{i_1}, \dots, \underline{x}_{i_n}\} .$$

This "lesser aim" enables us to simplify Steps 2 and 3 in the above procedure as follows:

Step 2' Compute the SVD

$$U_A^H Q_1 V = \text{diag}(c_1, \dots, c_n) \quad c_i > 0$$

Step 3' Form  $X^{-1} = V^H \Sigma Z^H \in C^{n \times n}$  and compute a unitary  $W$  such that

$$X^{-1}W = R$$

is upper triangular.

Note that this renders a QR factorization of the matrix of generalized singular vectors:

$$X = WR^{-1}$$

Since  $R^{-1}$  is upper triangular, it follows that

$$\text{Span}\{\underline{x}_1, \dots, \underline{x}_j\} = \text{Span}\{\underline{w}_1, \dots, \underline{w}_j\} \quad j = 1, \dots, n$$

where

$$X = [\underline{x}_1, \dots, \underline{x}_n]$$

and

$$W = [\underline{w}_1, \dots, \underline{w}_n]$$

Thus, if the  $c_i$  are ordered in Step 2' so that the space associated with

$$\mu_i = c_i / \sqrt{1 - c_i^2} \quad i = 1, \dots, k$$

is of interest, then  $\underline{w}_1, \dots, \underline{w}_k$  form the desired orthonormal basis. We mention that when computing SVD's such as the one in Step 2', it is always possible to order the singular values in any chosen fashion.

Regarding Step 3', the computation of  $W$  so  $(V^H \Sigma Z^H)W$  is upper triangular can be accomplished using Givens rotations:

$$T = V^H \Sigma Z^H$$

For  $i = n, \dots, 2$

For  $j = 1, \dots, i-1$

Determine  $c \in R$  and  $s \in C$  such that  $c^2 + |s|^2 = 1$  and

$$(t_{ij} \ t_{i,j+1}) \begin{pmatrix} c & \bar{s} \\ -s & c \end{pmatrix} = (0)^*$$

For  $k = 1, \dots, i$

$$(t_{kj} \ t_{k,j+1}) := (t_{kj} \ t_{k,j+1}) \begin{pmatrix} c & \bar{s} \\ -s & c \end{pmatrix}$$

This algorithm can be adapted to systolic array computation in the spirit of [3,22].

4. Comments

We close with some remarks concerned with the implementation of Steps 1, 2' and 3'. To begin with, rounding errors must be dealt with when attempting to deduce the multiplicity of the minimum generalized singular value. Let  $\hat{c}_1, \dots, \hat{c}_n$  be the computed singular values obtained in Step 2'. Assume that

$$\hat{c}_1 < \dots < \hat{c}_n$$

If  $\epsilon$  is the machine precision (or a small multiple thereof), and  $k$  is the largest integer such that

$$\hat{c}_1 < \hat{c}_2 < \dots < \hat{c}_k < c_1(1 + \epsilon) < \hat{c}_{k+1} < \dots < \hat{c}_n$$

then it is reasonable to regard

$$\hat{\mu}_i = \hat{c}_i / \sqrt{1 - \hat{c}_i^2} \quad i = 1, \dots, k$$