

A New Approach to Generalized Singular Value Decomposition

Shmuel Friedland *

Department of Mathematics, Statistics and Computer Science

University of Illinois at Chicago

Chicago, Illinois 60607-7045 †

and

Institute for Mathematics and its Applications

400 Lind Hall, 207 Church St., S.E.

Minneapolis, MN 55455-0436

December 8, 2004

Abstract

We propose a new algorithm to find the generalized singular value decompositions of two matrices with the same number of columns. We discuss in detail the sensitivity of our algorithm to errors in the entries of the matrices and suggest a way to suppress this sensitivity.

2000 Mathematical Subject Classification: 15A18, 92D10

Keywords: Gene expression matrix, singular value decomposition (SVD), extended singular value decomposition (ESVD), generalized singular value decomposition (GSVD).

1 Introduction

The Singular Value Decomposition (SVD) used in mathematics [6] and in numerical computations [5] is a very useful and versatile tool. It is used in statistics in Principal Component Analysis (PCA) [7] and recently became very useful in analysis of DNA microarrays [1].

*Partially supported by NSF

†Tel.: +1-312-996-3041; fax: +1-312-996-1491; email: friedlan@uic.edu

Let \mathbb{R}^m , $M_{mn}(\mathbb{R})$, $O_{md}(\mathbb{R})$, $M_n(\mathbb{R})$, $S_n(\mathbb{R})$, $GL_n(\mathbb{R})$, $O_n(\mathbb{R})$ be the linear space of real column vectors with m coordinates, the linear space of real $m \times n$ matrices, the subset of $m \times d$ matrices whose d ($\leq m$) columns is an orthonormal system, the algebra of $n \times n$ real matrices, the subspace of $n \times n$ real symmetric matrices, the group of $n \times n$ real invertible matrices and the subgroup of $n \times n$ orthogonal matrices. As usual, for $S \in S_n(\mathbb{R})$ we let $S \geq 0$ if S is nonnegative definite and $S > 0$ if S is positive definite. As in [5], denote by $\text{diag}(d_1, \dots, d_{\min(m,n)}) \in M_{mn}(\mathbb{R})$ the diagonal matrix with the diagonal entry d_i on the (i, i) position for $i = 1, \dots, \min(m, n)$ and all other entries are equal to zero.

For $A \in M_{mn}(\mathbb{R})$ the SVD of A given by $A = U\Sigma V^T$, where the following variations are popular: For the standard SVD $U \in O_m(\mathbb{R})$, $V \in O_n(\mathbb{R})$, $\Sigma = (s_{ij})_{i,j=1}^{m,n} \in M_{mn}(\mathbb{R})$, where Σ is a diagonal matrix with $s_{11} = \sigma_1 \geq \dots \geq s_{dd} = \sigma_d > 0 = s_{ii} = \sigma_i = 0$, $i = d + 1, \dots, \min(m, n)$. For the reduced SVD

$$A = U\Sigma V^T, \quad U \in O_{md}(\mathbb{R}), \quad V \in O_{nd}(\mathbb{R}), \quad 0 < \Sigma = \text{diag}(\sigma_1, \dots, \sigma_d) \in S_d(\mathbb{R}). \quad (1.1)$$

Note that d is the rank of A , denoted by $\text{rank } A$. The columns of U and the rows of V^T form an orthonormal basis of the column and the row space of A respectively. In this paper we will use mostly the reduced SVD, which has all the information available in A . In many applications as DNA analysis, image processing or data analysis, one can compress the data by observing that there is a small number of *significant* singular values $\sigma_1 \geq \dots \geq \sigma_t > 0$, while other singular values σ_{t+1}, \dots , are much smaller than σ_t or equal to zero numerically. (We call t the numerical rank of A .) Then the *compressed* SVD of A will be given by

$$A_1 = U_1 \Sigma_1 V_1^T, \quad U_1 \in O_{mt}(\mathbb{R}), \quad V_1 \in O_{nt}(\mathbb{R}), \quad \Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_t) > 0. \quad (1.2)$$

A_1 can be viewed also the *noise reduction* of A . This approach is used in [4] for finding the missing entries of the matrix A .

Let $A \in M_{mn}(\mathbb{R})$ and $B \in M_{ln}(\mathbb{R})$. Then the *Generalized Singular Value Decomposition* (GSVD) of A and B [5] is given by

$$A = F\Gamma R, \quad B = G\Delta R, \quad F \in O_m(\mathbb{R}), \quad G \in O_l(\mathbb{R}), \quad R \in GL(n, \mathbb{R}), \quad (1.3)$$

and $\Gamma \in M_{mn}(\mathbb{R})$, $\Delta \in M_{ln}(\mathbb{R})$ are diagonal matrices with the diagonal elements, called the generalized singular values, $\gamma_1, \dots, \gamma_{\min(m,n)} \geq 0$ and $\delta_1, \dots, \delta_{\min(l,n)} \geq 0$ respectively. (In general it is impossible to arrange the both sets of the singular values in a decreasing order.) GSVD became important recently in DNA microarrays analysis as a tool to compare two sets of DNA microarrays of different organisms [2] and [3]. (See §2 for more details.)

The numerical difficulties obtaining a stable GSVD decomposition can be partially attributed to the fact that the nonzero pair $(\gamma_i, \delta_i) \neq (0, 0)$ is determined up to a multiple by a positive scalar [5], [8] and [9]. The main feature to the existing algorithms for GSVD is the observation that the eigenvalues of the pencil $\delta^2 A^T A - \gamma^2 B^T B$ are the generalized singular values of the pair A, B .

The aim of this paper is to give a new robust algorithm to compute the *reduced* GSVD of A and B . The main feature of our approach is to consider the eigenvalues of the pencil $A^T A - \phi^2(A^T A + B^T B)$. Furthermore, to reduce the "noise" in the data represented by A, B we have to replace $P = A^T A + B^T B$ by a lower rank matrix \tilde{P} , using the *significant* singular values of P . Once \tilde{P} is chosen the low rank approximations of \tilde{A}, \tilde{B} of A, B , such that $\tilde{A}^T \tilde{A} + \tilde{B}^T \tilde{B} = \tilde{P}$, are determined. Thus our algorithm computes GSVD of the pencil $\tilde{A}^T \tilde{A} - \tilde{\phi}^2 \tilde{P}$. In applications, as microarrays analysis, the choice of \tilde{P} is left to the judgement of the user.

We now survey briefly the content of our paper. In §2 we discuss our main algorithm for GSVD. It is given in terms of matrices and it is self contained. In §3 we discuss two numerical algorithms to obtain stable GSVD decompositions of A, B . In §4 we discuss a random example of $A_0 \in M_{8,7}(\mathbb{R}), B_0 \in M_{9,7}(\mathbb{R})$ both of rank 2, such that the intersection of the row space of A_0 and B_0 is a subspace of dimension 1. Then the rank of $P_0 = A_0^T A_0 + B_0^T B_0$ is 3, and A_0, B_0 have three nonzero pairs of singular values: $(1.0, 0.0), (0.681, 0.732), (0.0, 1.0)$, up to 3 significant digits. In the terminology of microarrays we deduce that the two different organisms have three distinct functions, with exactly one common function.

Next we consider the random perturbation of A_0, B_0 given by matrices A, B . We replace this perturbation by A_1, B_1 of ranks 2 using the SVD of A_1, B_1 . The matrix $P := A_1^T A_1 + B_1^T B_1$ has three large singular values of magnitude 10^8 , the fourth singular value is of order 10^4 and the rest three singular values are less than 10^{-2} . (We used floating point precision rounded off to 10 digits.) We first assumed that P has $r = 3$ significant singular values. Then GSVD decomposition of the appropriate approximations \tilde{A}, \tilde{B} of rank 2 of A_0, B_0 is reasonably close to the original GSVD decomposition of A_0, B_0 . Next we assume that P has $r = 4$ significant singular values. (The maximal possible number of nonzero singular values.) Then the four pairs of generalized singular values of A_1, B_1 are $(1, 0), (1, 0), (0, 1), (0, 1)$. That is the pair $(0.681, 0.732)$ of generalized singular values of A_0, B_0 split to the two pairs $(1, 0), (0, 1)$ of generalized singular values of A_1, B_1 . In this case the microarrays interpretation yields that the two organisms have each two distinct functions, and no function in common! We explain the reason for this phenomenon in §3.

In the Appendix (§5) we give a short summary of SVD for a linear operator that maps one finite dimensional inner product space to another finite dimensional inner product space, over the complex numbers \mathbb{C} . This approach can be considered as a

base free approach to the SVD. For one matrix $A \in M_{mn}(\mathbb{C})$ it yields an extended singular value decomposition, called ESVD, obtained by introducing any two inner products on the linear space \mathbb{C}^m and \mathbb{C}^n . Then the GSVD of $A \in M_{mn}(\mathbb{C}), B \in M_{ln}(\mathbb{C})$ described in §2 is the ESVD of A and B obtained by choosing a special inner product on \mathbb{C}^n and the standard inner products on \mathbb{C}^m and \mathbb{C}^l . We hope that the ESVD introduced here will have more applications in the near future.

2 An exact algorithm for GSVD

We now describe briefly the main steps in our algorithm for the reduced GSVD for a given pair $A \in M_{mn}(\mathbb{R}), B \in M_{ln}(\mathbb{R})$. (In this section we assume that A, B have no "noise".) First we compute the three symmetric nonnegative definite matrices

$$P_A := A^T A, P_B := B^T B, P := P_A + P_B \in S_n(\mathbb{R}). \quad (2.1)$$

Assume that the reduced singular value decomposition of P , having rank r , is

$$P = O\Omega^2 O^T, \quad O \in O_{nr}(\mathbb{R}), \quad \Omega = \text{diag}(\omega_1, \dots, \omega_r), \quad \omega_1 \geq \dots \geq \omega_r > 0. \quad (2.2)$$

(The k -th column of O is an eigenvector of P corresponding to the eigenvalue ω_k^2 for $k = 1, \dots, r$.) Then

$$Q_A := \Omega^{-1} O^T P_A O \Omega^{-1}, \quad Q_B := \Omega^{-1} O^T P_B O \Omega^{-1} \in S_r(\mathbb{R}) \quad (2.3)$$

are nonnegative definite and $Q_A + Q_B$ is $r \times r$ identity matrix I_r . The spectral decompositions of Q_A and Q_B are given by

$$Q_A = T\Phi^2 T^T, \quad \Phi = \text{diag}(\phi_1, \dots, \phi_r), \quad \phi_i \geq 0, \quad i = 1, \dots, r, \quad T \in O_r(\mathbb{R}), \quad (2.4)$$

$$Q_B = T\Psi^2 T^T, \quad \Psi = \text{diag}(\psi_1, \dots, \psi_r), \quad \psi_i \geq 0, \quad i = 1, \dots, r, \quad (2.5)$$

$$\phi_i^2 + \psi_i^2 = 1, \quad \phi_i \geq 0, \psi_i \geq 0, \quad i = 1, \dots, r. \quad (2.6)$$

Let

$$V = O\Omega T \in M_{nr}(\mathbb{R}). \quad (2.7)$$

Then the GSVD of A and B is given by

$$A = U\Phi V^T, \quad U \in O_{mr}(\mathbb{R}), \quad B = W\Psi V^T, \quad W \in O_{lr}(\mathbb{R}). \quad (2.8)$$

The matrices U and W are easily obtained from the equalities

$$U\Phi = A O \Omega^{-1} T, \quad W\Psi = B O \Omega^{-1} T. \quad (2.9)$$

Theorem 2.1 *Let $A \in M_{mn}(\mathbb{R}), B = B \in M_{ln}(\mathbb{R})$ and assume that $r = \text{rank}(A^T A + B^T B)$. Then there exists a GSVD given by (2.8), where $\Phi, \Psi \in M_r(\mathbb{R})$ are nonnegative diagonal matrices satisfying $\Phi^2 + \Psi^2 = I_r$ and $V \in M_{nr}(\mathbb{R}), \text{rank } V = r$. Every such GSVD can be found as follows: Let P be defined by (2.1); let O and Ω be given by the reduced SVD of P (2.2); let Q_A and Q_B be defined in (2.3); let the spectral decomposition of Q_A and Q_B be given by (2.4)-(2.6); and let V is given by (2.7). The columns of U and W in (2.8) corresponding to the nonzero ϕ_i and ϕ_j are uniquely determined by (2.9). Other columns of U and W are any orthonormal systems in the orthogonal complement of the subspaces spanned by the determined columns of U and W respectively.*

Proof. Clearly the matrices P_A, P_B and P are nonnegative definite. The reduced SVD of P is the spectral decomposition of P corresponding to the positive eigenvalues of P . Observe

$$\Omega^2 = O^T P O = O^T (P_A + P_B) O = O^T P_A O + O^T P_B O.$$

Hence

$$I_r = Q_A + Q_B, \quad Q_A \geq 0, Q_B \geq 0. \quad (2.10)$$

Let (2.4) be the spectral decomposition of Q_A . Then (2.10) implies (2.5) and (2.6). Let $\hat{U} := A O \Omega^{-1} T$. Then $\hat{U}^T \hat{U} = T^T Q_A T = \Phi^2$. Assume for simplicity of the exposition that

$$\phi_1 \geq \dots \geq \phi_{r_A} > 0 = \phi_{r_A+1} = \dots = \phi_r = 0, \quad (2.11)$$

where $r_A = \text{rank } A$. Then the last $r - r_A$ columns of \hat{U} are zero, while first r_A columns of \hat{U} is an orthogonal system, where the norm of column i is ϕ_i for $i = 1, \dots, r_A$. We define $U \in M_{nr}(\mathbb{R})$ as follows. The i -th column of U is the i -th column of \hat{U} divided by ϕ_i for $i = 1, \dots, r_A$. The last $r - r_A$ columns of U is any orthonormal system in the orthogonal complement of the column space of \hat{U} . Hence $U \in O_{nr}(\mathbb{R}), \hat{U} = U \Phi$ and $A = U \Phi V^T$. The second equality in (2.8) is established similarly.

We now show that any GSVD given (2.8), with the conditions $\Phi, \Psi \in M_r(\mathbb{R})$ are nonnegative diagonal matrices satisfying $\Phi^2 + \Psi^2 = I_r$ and $V \in M_{nr}(\mathbb{R}), \text{rank } V = r$, can be obtained by using the algorithm given in the theorem. Clearly $P = V V^T$. So the column space of V is equal to the columns space of O . Since $\text{rank } V = \text{rank } O = r$ it follows that $V = O \Omega T$ for some invertible $T \in M_r(\mathbb{R})$. Use (2.2) and the equality $P = V V^T$ to deduce that $T T^T = I_r$. Hence $Q_A = T \Phi^2 T^T, Q_B = T \Psi^2 T^T$ and the rest of the theorem follows. \square

We conclude this section with the following remarks. First, we consider the interpretation of GSVD in comparing two sets of DNA microarrays of different organisms as discussed in [2]. As mentioned in §1 a_{ij}, b_{pj} entries of $A = (a_{ij})_{i,j=1}^{m,n}, B = (b_{pj})_{p,j=1}^{l,n}$ represent the measurements for the i and p genome of the two different organisms, represented by the matrices A and B , in the measurement $j = 1, \dots, n$. (We assume for the simplicity of the exposition that A, B do not have errors (noise). We address the problem of noise filtering of A, B in the next section.) The number r appearing in the GSVD (2.8) is the number of total functions of the DNA of the two organisms observed in the n experiments is k . If $\frac{\phi_i}{\psi_i} \approx 1$ then the function i is similarly expressed in both organisms. If $\frac{\phi_i}{\psi_i} \gg 1$ ($\frac{\psi_i}{\phi_i} \gg 1$) then the function i expressed only in the first organism (the second organism).

Second, from the proof of Theorem 2.1 it follows that the information given in (2.8) is equivalent to the following

$$A = U_1 \Phi_1 V_1^T, \quad U_1 \in O_{mr_A}(\mathbb{R}), \quad V_1 \in O_{nr_A}(\mathbb{R}), \quad 0 < \Phi_1 = \text{diag}(\phi_{i_1}, \dots, \phi_{i_{r_A}}) \in S_{r_A}(\mathbb{R}), \quad (2.12)$$

$$B = W_1 \Psi_1 V_2^T, \quad W_1 \in O_{mr_B}(\mathbb{R}), \quad V_2 \in O_{nr_B}(\mathbb{R}), \quad 0 < \Psi_1 = \text{diag}(\psi_{j_1}, \dots, \psi_{j_{r_B}}) \in S_{r_B}(\mathbb{R}).$$

Here $\phi_{i_1}, \dots, \phi_{i_{r_A}}$ and $\psi_{j_1}, \dots, \psi_{j_{r_B}}$ are all positive entries of Φ and Ψ appearing (2.8). (Note that $r_A = \text{rank } A, r_B = \text{rank } B$.) V_1 and V_2 are the submatrices of V corresponding to the columns i_1, \dots, i_{r_A} and j_1, \dots, j_{r_B} respectively. U_1 and W_1 are the submatrices of U and W corresponding to the columns i_1, \dots, i_{r_A} and j_1, \dots, j_{r_B} . Equivalently (2.12) can be viewed as the reduced ESVD of A and B with respect to the standard inner products in \mathbb{R}^m and \mathbb{R}^l and the inner product in \mathbf{V} , the subspace spanned by the rows of A and B , induced by P . See §5.

3 Two numerical algorithms

In this section point out two numerical algorithms to compute GSVD for $A \in M_{mn}(\mathbb{R}), B \in M_{ln}(\mathbb{R})$. In the first algorithm we replace A, B by A_1, B_1 using the significant singular values of A, B of their corresponding SVD as explained in §1. (A_1, B_1 are the compressed SVD of A, B .) Then we form the matrices

$$P_A := A_1^T A_1, \quad P := A_1^T A_1 + B_1^T B_1. \quad (3.1)$$

Next we consider the SVD of P , i.e. its spectral decomposition, and assume that P has \tilde{r} *significant* singular values $\tilde{r} \leq \text{rank } P$. Let \tilde{P} be the *compressed* SVD of P :

$$\tilde{P} = \tilde{O} \tilde{\Omega}^2 \tilde{O}^T, \quad \tilde{O} \in O_{n\tilde{r}}(\mathbb{R}), \quad \tilde{\Omega} = \text{diag}(\tilde{\omega}_1, \dots, \tilde{\omega}_{\tilde{r}}), \quad \tilde{\omega}_1 \geq \dots \geq \tilde{\omega}_{\tilde{r}} > 0. \quad (3.2)$$

(The k -th column of \tilde{O} is an eigenvector of P corresponding to the eigenvalue $\tilde{\omega}_k^2$ for $k = 1, \dots, \tilde{r}$.) Form \tilde{Q}_A and compute its spectral decomposition, which determines $\tilde{\Phi}$ and $\tilde{\Psi}$:

$$\tilde{Q}_A := \tilde{\Omega}^{-1} \tilde{O}^T P_A \tilde{O} \tilde{\Omega}^{-1} \in S_{\tilde{r}}(\mathbb{R}) \quad (3.3)$$

$$\tilde{Q}_A = \tilde{T} \tilde{\Phi}^2 \tilde{T}^T, \quad \tilde{T} \in O_{\tilde{r}}(\mathbb{R}), \quad \tilde{\Phi} = \text{diag}(\tilde{\phi}_1, \dots, \tilde{\phi}_{\tilde{r}}), \quad \tilde{\phi}_1 \geq \dots \geq \tilde{\phi}_{\tilde{r}} \geq 0, \quad (3.4)$$

$$\tilde{\Psi} = \text{diag}(\tilde{\psi}_1, \dots, \tilde{\psi}_{\tilde{r}}), \quad 0 \leq \tilde{\psi}_1 \leq \dots \leq \tilde{\psi}_{\tilde{r}}, \quad \tilde{\Phi}^2 + \tilde{\Psi}^2 = I_{\tilde{r}}. \quad (3.5)$$

Then the first rank $\tilde{\Phi}$ columns of \tilde{U} and the last rank $\tilde{\Psi}$ columns of \tilde{W} are effectively determined by the equalities

$$\tilde{U} \tilde{\Phi} := A_1 \tilde{O} \tilde{\Omega}^{-1} \tilde{T}, \quad \tilde{U} \in O_{m\tilde{r}}(\mathbb{R}), \quad \tilde{W} \tilde{\Psi} = B_1 \tilde{O} \tilde{\Omega}^{-1} \tilde{T}, \quad \tilde{W} \in O_{l\tilde{r}}(\mathbb{R}). \quad (3.6)$$

We now explain why we need the additional noise reduction in P , after we already performed the noise reduction in A, B . Assume that the ideal data was given by $A_0 \in M_{mn}(\mathbb{R}), B_0 \in M_{ln}(\mathbb{R})$. Suppose that $\dim A_0^T \mathbb{R}^m \cap B_0^T \mathbb{R}^l = c > 0$. Let $P_0 := A_0^T A_0 + B_0^T B_0$. Then $\text{rank } P_0 = \text{rank } A_0 + \text{rank } B_0 - c$. Assume that interesting case $\text{rank } P_0 < n$.

The given matrices A, B are perturbations of A_0, B_0 . Suppose that $\text{rank } A_1 = \text{rank } A_0$, $\text{rank } B_1 = \text{rank } B_0$. However, as we shall see in the next section, with high probability $\text{rank } P = \min(\text{rank } A_0 + \text{rank } B_0, n) > \text{rank } P_0$. Still P will have rank P_0 significant singular values. Thus \tilde{P} is the correct approximation of P_0 , and \tilde{A}, \tilde{B} are the approximation of A_0, B_0 with the property $\dim \tilde{A}^T \mathbb{R}^m \cap \tilde{B}^T \mathbb{R}^l = c$.

In the second algorithm we simply let $A_1 = A$ and $B_1 = B$, i.e. we do not perform the noise reduction in A, B . We just perform the noise reduction in P . Then \tilde{A}, \tilde{B} are the noise reductions of A, B with the properties as above.

4 Numerical examples

One way to obtain a random matrix $E \in M_{pq}(\mathbb{R})$ of rank $t \leq \min(p, q)$, in general, is to use the following procedure:

$$E := \sum_{i=1}^t \mathbf{x}_i \mathbf{y}_i^T, \quad \mathbf{x}_1, \dots, \mathbf{x}_t \in \mathbb{R}^p, \quad \mathbf{y}_1, \dots, \mathbf{y}_t \in \mathbb{R}^q, \quad (4.1)$$

where $\mathbf{x}_1, \dots, \mathbf{x}_t, \mathbf{y}_1, \dots, \mathbf{y}_t$ are chosen at random. Suppose we use the procedure (4.1) to generate random matrix $A_0 \in M_{mn}(\mathbb{R})$ with $t := a_0 \leq \min(m, n)$ and random matrix $B_0 \in M_{ln}(\mathbb{R})$ with $t := b_0 \leq \min(l, n)$. Then generically $r_{A_0} = a_0$ and $r_{B_0} = b_0$ respectively. Assume that $a_0 + b_0 \leq n$. Then almost surely P_{A_0} and P_{B_0}

do not have a common nullspace, and the rank of the matrix $P = A_0^T A_0 + B_0^T B_0$ is $r = r_{A_0} + r_{B_0}$. In order to generate random matrices $A_0 \in M_{mn}(\mathbb{R}), B_0 \in M_{ln}(\mathbb{R})$ of ranks r_{A_0}, r_{B_0} , such that the matrix P will have in general rank $r \in [\max(r_{A_0}, r_{B_0}), r_{A_0} + r_{B_0}]$ we modify the procedure (4.1) as follows. Given $rc \geq 1$ we generate two random matrices $F \in M_n(\mathbb{R}), G \in M_{ln}(\mathbb{R})$ both of rank rc such that $F^T \mathbb{R}^m \cap G^T \mathbb{R}^l$ have in general dimension rc :

$$F := \sum_{i=1}^{rc} \mathbf{x}_i \mathbf{y}_i^T, G := \sum_{i=1}^{rc} \mathbf{z}_i \mathbf{y}_i^T, \mathbf{x}_1, \dots, \mathbf{x}_{rc} \in \mathbb{R}^m, \mathbf{y}_1, \dots, \mathbf{y}_{rc} \in \mathbb{R}^n, \mathbf{z}_1, \dots, \mathbf{z}_{rc} \in \mathbb{R}^l, \quad (4.2)$$

where $\mathbf{x}_1, \dots, \mathbf{x}_{rc}, \mathbf{y}_1, \dots, \mathbf{y}_{rc}, \mathbf{z}_1, \dots, \mathbf{z}_{rc}$ are chosen at random. After that we generate the random matrices $E_1 \in M_{mn}(\mathbb{R}), E_2 \in M_{ln}(\mathbb{R})$ of ranks $r_{A_0} - rc, r_{B_0} - rc$ using the procedure (4.1). Then $A_0 = F + E_1, B_0 = G + E_2$ are of the ranks r_{A_0}, r_{B_0} and rank $P = r_{A_0} + r_{B_0} - rc$. (If $rc = 0$ then $F = 0, G = 0$.)

We first generated $A_0 \in M_{8,7}(\mathbb{R}), B_0 \in M_{9,7}(\mathbb{R})$ with $rc = 1, r_{A_0} = r_{B_0} = 2$ as explained above:

$$A_0 = \begin{pmatrix} 1826 & 846 & 1516 & 1831 & 3060 & -577 & 1368 \\ -3452 & -1752 & -2182 & -2827 & -5970 & 1199 & -2236 \\ 5765 & 3573 & 745 & 2032 & 10755 & -2461 & 2250 \\ -202 & -1818 & 7558 & 6964 & -2430 & 1286 & 3804 \\ 3873 & 1353 & 5193 & 5718 & 5955 & -911 & 3914 \\ -5206 & -2862 & -2306 & -3350 & -9270 & 1964 & -2868 \\ -2060 & 1224 & -11470 & -11119 & -810 & -893 & -6540 \\ -2630 & -726 & -4390 & -4684 & -3810 & 482 & -3100 \end{pmatrix},$$

$$B_0 = \begin{pmatrix} -3652 & -3486 & 640 & 2833 & -321 & 1424 & -1731 \\ -8657 & -7471 & -2665 & 3283 & 1354 & 2669 & -6371 \\ 2420 & 2122 & 568 & -1063 & -289 & -776 & 1685 \\ -3927 & -4161 & 2865 & 4833 & -1446 & 1899 & -681 \\ 253 & -873 & 5837 & 4631 & -2952 & 895 & 3309 \\ -4620 & -2044 & -11676 & -6664 & 5908 & -308 & -8960 \\ 2596 & 2388 & 20 & -1624 & -12 & -932 & 1488 \\ -8624 & -7722 & -1180 & 4481 & 603 & 2908 & -5547 \\ -7964 & -5438 & -10024 & -3195 & 5075 & 1176 & -9967 \end{pmatrix}.$$

(Since we used Maple routine to generate random vectors and matrices with integer entries in the range $[-99, 99]$, the matrices A_0 and B_0 have integer entries.) The first three singular values of A_0 are and B_0 are

$$27455.5092631633888, 17374.6830503566089, 3.14050409246786192 \times 10^{-12},$$

29977.5429571960522, 19134.3838220483449, $3.52429226420727071 \times 10^{-12}$,

i.e. the ranks of A_0 and B_0 are 2. The four first singular values of P are

$$1.32179857269680762 \times 10^9, 6.04366385186753988 \times 10^8, \\ 3.94297368116438210 \times 10^8, 1.34609524647135614 \times 10^{-7}.$$

The 3 generalized singular values of A_0 and B_0 given by Theorem 2.1 are:

$$\phi_1 = 1, \phi_2 = 0.6814262563, \phi_3 = 3.777588180 \times 10^{-9}, \\ \psi_1 = 0, \psi_2 = 0.7318867789, \psi_3 = 1.$$

The matrix $V \in M_{7,3}(\mathbb{R})$ given by Theorem 2.1 is:

$$V = \begin{pmatrix} 9975.15570726070292 & 2218.77772608917530 & -16518.7090029761894 \\ 4258.08009519515508 & 5446.09078420315382 & -13181.0851653484933 \\ 9910.16789114064704 & -17951.9288836113556 & -10755.8389196556800 \\ 11513.1007966750912 & -16136.5652888949844 & 1610.05490646888757 \\ 16275.4438487944208 & 9076.81797277147780 & 5451.31112120093894 \\ -2894.44180376545456 & -3832.43425556036664 & 4134.75009336168569 \\ 8306.91358787289937 & -8471.69677689182754 & -15231.5628696020595 \end{pmatrix}.$$

U_1 : the first two columns of $U \in M_{8,3}(\mathbb{R})$, and W_1 : the last two columns of $W \in M_{9,3}(\mathbb{R})$ in the decomposition (2.8) are given by:

$$U_1 = \begin{pmatrix} .179765121701646796 & 0.0217039456257290890 \\ -.332293103754918662 & -0.0908271155631820566 \\ .522953081333910608 & .362755122878894332 \\ 0.0653691351259460541 & -.564884208286237532 \\ 403109666836316383 & -0.0979419423157340264 \\ -.490268513741756395 & -.208670062565787018 \\ -.310503391981747257 & .686088211972394224 \\ -.283266252367416982 & .129387295795073296 \end{pmatrix},$$

$$W_1 = \begin{pmatrix} -.212587926393571491 & .200183908075783706 \\ -.209341218852211240 & .503492859705758855 \\ 0.0709866110946431456 & -.139522117749788622 \\ -.381933946402143232 & .200183908075783734 \\ -.399563630497640044 & -0.0545956112933954560 \\ .610572448157044589 & .339706025825572078 \\ .117661178466181738 & -.145588296782388132 \\ -.312433467319077507 & .491360501640559999 \\ .340824522437284283 & .515625217770957822 \end{pmatrix}.$$

To show the robustness of our GSVD decomposition we perturb the matrices A_0 and B_0 by random matrices of the maximal rank by generating matrices $X \in M_{8,7}(\mathbb{R})$, $Y \in \mathbb{R}_{9,7}(\mathbb{R})$ with random entries and relatively small ℓ_2 norm (the first singular value) with respect to the ℓ_2 norms of A and B respectively. The random matrices X, Y , with integer entries in $[-99, 99]$, are:

$$X = \begin{pmatrix} -14 & -73 & 65 & 3 & -14 & 16 & 9 \\ -10 & 8 & 90 & -94 & -22 & -24 & 0 \\ 78 & 32 & -48 & -6 & 80 & -18 & -63 \\ 66 & -13 & 88 & -45 & -92 & -69 & -43 \\ 32 & 9 & 41 & -95 & -28 & -90 & -63 \\ -23 & -72 & -84 & -84 & -58 & -37 & -40 \\ 35 & 14 & -29 & 76 & -62 & -82 & -5 \\ 18 & -40 & -51 & 11 & 87 & 66 & -46 \end{pmatrix},$$

$$Y = \begin{pmatrix} -14 & -83 & 65 & -22 & -80 & 43 & -56 \\ -55 & -50 & 68 & 66 & 9 & -26 & -58 \\ -63 & -32 & -25 & 96 & 90 & -5 & 28 \\ -49 & 31 & -17 & -27 & 46 & -5 & 37 \\ 81 & -99 & -98 & 22 & 58 & -68 & -37 \\ 59 & 57 & 65 & -64 & 65 & 84 & 41 \\ -68 & 36 & -63 & 7 & -58 & 53 & 90 \\ 95 & -27 & 54 & -16 & -46 & -18 & 46 \\ 23 & 10 & -64 & 58 & 58 & -73 & 97 \end{pmatrix}$$

The singular values of X, Y rounded off to three significant digits are:

$$(266, 183, 165, 151, 99.1, 36.0, 14.1), (259, 229, 198, 153, 116, 86.8, 46.2).$$

Note that $\|X\| \sim 0.01\|A_0\|$, $\|Y\| \sim 0.01\|B_0\|$. Form the matrices $A := A_0 + X$, $B := B_0 + Y$. These matrices have the full ranks with corresponding singular values rounded off to three significant digits at least:

$$(27490, 17450, 233, 130, 119, 70.0, 18.2), (29884, 19183, 250, 187, 137, 102, 19.7).$$

We now replace A, B by A_1, B_1 of rank two using the first two singular values and the corresponding singular vectors in the SVD decompositions of A, B . (We do the noise reduction described in the Introduction.) Then two nonzero singular values of A_1, B_1 are $(27490, 17450), (29883, 19183)$, rounded to five significant digits. The singular values of the corresponding $P = A_1^T A_1 + B_1^T B_1$ are up to 3 significant digits:

$$(1.32 \times 10^9, 6.07 \times 10^8, 3.96 \times 10^8, 1.31 \times 10^4, 0.068, 9.88 \times 10^{-3}, 6.76 \times 10^{-3}). \quad (4.3)$$

Assume that $\tilde{r} = 3$, i.e. P has three significant singular values. We now apply our first numerical algorithm for GSVD, as described in §3. The three generalized singular values of A_1, B_1 are

$$(1.000000000, .6814704276, 0.7582758358 \times 10^{-8}), \quad (0., .7318456506, 1.0).$$

These result match the generalized singular values of A_0, B_0 at least up to four significant digits. Let $\hat{V}, \hat{U}_1, \hat{W}_1$ be the matrix V , the first two columns of U , the last two columns of W , which are computed for A_1, B_1 . Then

$$\frac{\|V - \hat{V}\|}{\|V\|} \sim 0.0061, \quad \|U_1 - \hat{U}_1\| \sim 0.0093, \quad \|W_1 - \hat{W}_1\| \sim 0.0098.$$

We now implement the second numerical algorithm to the matrices $A \in M_{8,7}, B \in M_{9,7}(\mathbb{R})$, as described in §3. The singular values of $P = A^T A + B^T B$, up to three significant digits, are

$$(1.31 \times 10^9, 6.07 \times 10^8, 3.96 \times 10^8, 6.91 \times 10^4, 6.61 \times 10^4, 2.84 \times 10^4, 1.19 \times 10^4).$$

As expected, P has $\tilde{r} = 3$ significant singular values. Then the 3 generalized singular eigenvalues of \tilde{A}, \tilde{B} are:

$$(.9999667639, .6814699415, 0.005726580138), (0.008152974917, .7318461033, .9999836030).$$

Let $\tilde{V}, \tilde{U}_1, \tilde{W}_1$ be the corresponding matrices for GSVD of \tilde{A}, \tilde{B} . Then

$$\frac{\|V - \tilde{V}\|}{\|V\|} \sim 0.0061, \quad \|U_1 - \tilde{U}_1\| \sim 0.0091, \quad \|W_1 - \tilde{W}_1\| \sim 0.011.$$

This shows that the two "noise reduction" algorithms are comparable in these examples.

Finally we discuss the critical issue of choosing correctly the number of significant singular values of noised matrices A, B and the corresponding matrix P . First we revisit our example with A_1, B_1 . Recall the values of the singular values of $P = A_1^T A_1 + B_1^T B_1$ given by (4.3). Assume now that $\tilde{r} = 4$. Then the four generalized singular values of \tilde{A}, \tilde{B} up to six significant digits are $(1, 1, 0, 0), (0, 0, 1, 1)$!

We now replace A, B by A_2, B_2 of rank three using the first three singular values and the corresponding singular vectors in the SVD decompositions of A, B . The singular values of the matrix P up to three significant digits are:

$$(1.32 \times 10^9, 6.07 \times 10^8, 3.96 \times 10^8, 4.74 \times 10^4, 3.83 \times 10^4, 5.39 \times 10^3, 9.70 \times 10^{-3}).$$

Assume first that $\tilde{r} = 6$. It then turns out that the generalized singular values of A_2, B_2 are $(1, 1, 1, 0, 0, 0)$ and $(0, 0, 0, 1, 1, 1)$ up to six significant digits!. Assume finally in this example that $\tilde{r} = 3$. It then turns out that the generalized singular values of \tilde{A}, \tilde{B} are

$(.9999796224, .6814701987, 0.005232470265), (0.006383948621, .7318458638, .9999863106)$.

Thus, the most important step in finding the GSVD decomposition of noised data given by \tilde{A}, \tilde{B} is the choice of \tilde{r} , the significant number of singular values of P .

5 Appendix: SVD on inner product spaces

In this appendix we discuss briefly the standard notion of the SVD decomposition of a linear operator that maps one finite dimensional inner product space to another finite dimensional inner product space. For the utmost generality we consider here the inner products over the complex numbers \mathbb{C} . The proofs of the facts stated here are either standard or straightforward, and are left to the reader.

Let \mathbf{U}_i be an m_i -dimensional inner product space over \mathbb{C} , given by $\langle \cdot, \cdot \rangle_i$ for $i = 1, 2$. Let $T : \mathbf{U}_1 \rightarrow \mathbf{U}_2$ be a linear operator. Let $T^* : \mathbf{U}_2 \rightarrow \mathbf{U}_1$ be the adjoint operator of T , i.e. $\langle T\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, T^*\mathbf{y} \rangle$ for all $\mathbf{x} \in \mathbf{U}_1, \mathbf{y} \in \mathbf{U}_2$. Equivalently, let $[\mathbf{a}_1, \dots, \mathbf{a}_{m_1}]$ and $[\mathbf{b}_1, \dots, \mathbf{b}_{m_2}]$ are orthonormal bases of \mathbf{U}_1 and \mathbf{U}_2 respectively. Let $A \in M_{m_2 m_1}(\mathbb{C})$ be the representation matrix of T in these bases: $[T\mathbf{a}_1, \dots, T\mathbf{a}_{m_1}] = [\mathbf{b}_1, \dots, \mathbf{b}_{m_2}]A$. Then the matrix $A^* := \overline{A}^T$ represents T^* in these bases. The SVD decomposition of $A = U\Sigma V^*$, where U, V are unitary matrices of dimensions m_2, m_1 respectively and $\Sigma \in M_{m_2 m_1}(\mathbb{R})$ is a diagonal matrix with the diagonal entries $\sigma_1 \geq \dots \geq \sigma_{\min(m_2, m_1)} \geq 0$, corresponds to the following *base free* concepts of T .

Consider the operators $S_1 := T^*T : \mathbf{U}_1 \rightarrow \mathbf{U}_1$ and $S_2 := TT^* : \mathbf{U}_2 \rightarrow \mathbf{U}_2$. Then S_1, S_2 are self-adjoint, i.e. $S_1^* = S_1, S_2^* = S_2$ and nonnegative definite: $\langle S_i \mathbf{x}_i, \mathbf{x}_i \rangle \geq 0$ for all $\mathbf{x}_i \in \mathbf{U}_i$ for $i = 1, 2$. The positive eigenvalues of S_1 and S_2 , counted with their multiplicities and arranged in a decreasing order are $\sigma_1^2 \geq \dots \geq \sigma_r^2 > 0$, where $r = \text{rank } T = \text{rank } T^*$. Let

$$S_1 \mathbf{v}_i = \sigma_i^2 \mathbf{v}_i, \quad i = 1, \dots, r, \quad \langle \mathbf{v}_i, \mathbf{v}_j \rangle_1 = \delta_{ij}, \quad i, j = 1, \dots, r.$$

Then $\mathbf{u}_i := \sigma_i^{-1} T \mathbf{v}_i$ for $i = 1, \dots, r$ is an orthonormal set of the eigenvectors of S_2 corresponding corresponding to the eigenvalue σ_i^2 for $i = 1, \dots, r$. Complete the orthonormal systems $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ and $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ to orthonormal bases $[\mathbf{v}_1, \dots, \mathbf{v}_{m_1}]$ and $[\mathbf{u}_1, \dots, \mathbf{u}_{m_2}]$ in \mathbf{U}_1 and \mathbf{U}_2 respectively. Then the unitary matrices U, V are the transition matrices from basis $[\mathbf{b}_1, \dots, \mathbf{b}_{m_2}]$ to $[\mathbf{u}_1, \dots, \mathbf{u}_{m_2}]$ and basis $[\mathbf{a}_1, \dots, \mathbf{a}_{m_1}]$ to

$[\mathbf{v}_1, \dots, \mathbf{v}_{m_1}]$:

$$[\mathbf{u}_1, \dots, \mathbf{u}_{m_2}] = [\mathbf{b}_1, \dots, \mathbf{b}_{m_2}]U, \quad [\mathbf{v}_1, \dots, \mathbf{v}_{m_1}] = [\mathbf{a}_1, \dots, \mathbf{a}_{m_1}]V.$$

Let $A \in M_{m_2 m_1}(\mathbb{C})$. Then A can be viewed as a linear operator $A : \mathbb{C}^{m_1} \rightarrow \mathbb{C}^{m_2}$, where $\mathbf{x} \rightarrow A\mathbf{x}$ for any $\mathbf{x} \in \mathbb{C}^{m_1}$. Let P_i be $m_i \times m_i$ hermitian positive definite matrix for $i = 1, 2$. We define the following inner product on \mathbb{C}^{m_1} and \mathbb{C}^{m_2} :

$$\langle \mathbf{x}, \mathbf{y} \rangle_i := \mathbf{y}^* P_i \mathbf{x}, \quad \mathbf{x}, \mathbf{y} \in \mathbb{C}^{m_i}, \quad i = 1, 2. \quad (5.1)$$

It is straightforward to show that the SVD decomposition of A , viewed as the above operator, with respect to the inner products given by (5.1) is

$$A = U\Sigma V^*, \quad U^* P_2 U = I_{m_2}, \quad V^* P_1^{-1} V = I_{m_1}, \quad (5.2)$$

where Σ is an $m_2 \times m_1$ diagonal matrix with nonnegative diagonal entries in a decreasing order. A simple way to deduce this decomposition is to observe that $\langle \mathbf{x}, \mathbf{y} \rangle_i = (P_i^{\frac{1}{2}} \mathbf{y})^* (P_i^{\frac{1}{2}} \mathbf{x})$, where $P_i^{\frac{1}{2}}$ is the unique hermitian positive definite square root of P_i . The decomposition (5.2) is called the *extended singular value decomposition* of A , ESVD for short. The diagonal entries of Σ are called the *extended singular values* of A . ESVD of A corresponds to the standard SVD decomposition of $P_2^{\frac{1}{2}} A P_1^{-\frac{1}{2}}$.

We conclude this section by considering the GSVD of $A \in M_{m_2 m_1}(\mathbb{C})$ and $B \in M_{m_3 m_1}(\mathbb{C})$ given by

$$A = U\Phi V^*, \quad U \in U_{m_2 r}(\mathbb{C}), \quad B = W\Psi V^*, \quad W \in U_{m_3 r}(\mathbb{C}), \quad V \in U_{m_1 r}(\mathbb{C}). \quad (5.3)$$

Here $U_{mr}(\mathbb{C}) := \{Z \in M_{mr}(\mathbb{C}) : Z^* Z = I_r\}$, and the matrices U, V, W, Φ, Ψ are given by the formulas as in §2, except that the transposed matrices appearing there are replaced by the conjugate transposed matrices. We claim that (5.3) are ESVD of A and B with respect to the following corresponding hermitian positive definite matrices $P_i \in M_{m_i}(\mathbb{C}), i = 1, 2, 3$. First let $P_2 = I_{m_2}$ and $P_3 = I_{m_3}$. Let $P = A^* A + B^* B$ be an $m_1 \times m_1$ hermitian nonnegative definite matrix. Assume first that $P > 0$. Then choose $P_1 = P$. Let $E := AP^{-\frac{1}{2}}, F := BP^{-\frac{1}{2}}$. Observe that $E^* E + F^* F = I_{m_1}$. Then the spectral decomposition of $E^* E$ and $F^* F$ is given by

$$E^* E = V_0 \Phi^2 V_0^*, \quad F^* F = V_0 \Psi^2 V_0^*, \quad V_0^* V_0 = I_{m_1}, \quad \Phi^2 + \Psi^2 = I_{m_1},$$

where Φ, Ψ are diagonal nonnegative definite matrices. This establishes the decomposition (5.3) with $V^* = V_0^* P^{\frac{1}{2}}$.

Assume now that P is not positive definite, i.e. $\text{rank } P < m_1$. Let $\mathbf{V} \subset \mathbb{C}^{m_1}$ be the range of P of dimension r . Note that $A\mathbb{C}^{m_1} = A\mathbf{V}$, $B\mathbb{C}^{m_1} = B\mathbf{V}$. Hence we can view the matrices A, B as the following operators $A : \mathbf{V} \rightarrow \mathbb{C}^{m_2}$. Let $\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{y}^* P \mathbf{x}$ be an inner product on \mathbf{V} . Then the ESVD of the operators A, B gives the GSVD (5.3). Alternatively, let P_1 be any $m_1 \times m_1$ hermitian matrix such that \mathbf{V} is an invariant subspace of P_1 and the restriction of P_1 to \mathbf{V} is equal to the restriction of P to \mathbf{V} . Let E, F be defines as above. Then $E^*E + F^*F$ is the identity operator on \mathbf{V} and is equal to zero operator on the orthogonal complement of \mathbf{V} in \mathbb{C}^{m_1} . Hence E^*E and F^*F commute. Thus E^*E and F^*F have a common orthonormal basis of eigenvectors. Then the reduced ESVD of A and B yields the decomposition (5.3).

6 Conclusions

In this paper we discuss a new algorithm to find the generalized singular value decomposition (GSVD) of two real matrices A, B with the same number of rows. The main novelty of our approach is to consider the spectral decomposition of the nonnegative symmetric matrix $P := A^T A + B^T B$ and use it to derive the GSVD of the pair A, B . Roughly speaking the GSVD is closely related to the spectral decomposition of the pencils $A^T A - \phi^2 P, B^T B - \psi^2 P$.

To obtain stable numerical computations of GSVD of A, B one needs to reduce the "noise" in P by approximating P with a low rank matrix \tilde{P} , which is called the *numerical* rank of P . The numerical examples clearly showed that the stability of GSVD is very influenced by the choice of the numerical rank of P . This may have a significant impact on the interpretation of the comparative analysis of genome-scale expression data sets of two different organisms.

References

- [1] O. Alter, P.O. Brown and D. Botstein, Singular value decomposition for genome-wide expression data processing and modelling, *Proc. Nat. Acad. Sci. USA* 97 (2000), 10101-10106.
- [2] O. Alter, P.O. Brown and D. Botstein, Generalized singular decomposition for comparative analysis of genome-scale expression data sets of two different organisms, *Proc. Nat. Acad. Sci. USA* 100 (2003), 3351-3356.
- [3] O. Alter, G.H. Golub, P.O. Brown and D. Botstein, Novel genome-scale correlation between DNA replication and RNA transcription during the cell cycle in yeast is predicted by data-driven models, 2004 *Miami Nature Winter Symposium*, Jan. 31 - Feb. 4, 2004, to appear.
- [4] S. Friedland, A. Niknejad and L. Chihara, A simultaneous reconstruction of missing data in DNA microarrays, submitted. <http://www.ima.umn.edu/preprints/dec2003/dec2003.html>
- [5] G.H. Golub and C.F. Van Loan, *Matrix Computation*, John Hopkins Univ. Press, 3rd Ed., 1996.
- [6] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge Univ. Press, 1987.
- [7] R.A. Johnson, D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice Hall, New Jersey, 4th edition (1998).
- [8] C.C. Paige and M. A. Saunders, Towards a generalized singular value decomposition, *SIAM J. Numer. Anal.* 18 (1981), 398-405.
- [9] G.W. Stewart, A method for computing the generalized singular value decomposition, *Matrix Pencils*, B. Kagström and A. Ruhe, *Lecture Notes in Mathematics*, 973 (1982), 207-220.