ON A VARIATIONAL FORMULATION OF THE GENERALIZED SINGULAR VALUE DECOMPOSITION

MOODY T. CHU*, ROBERT. E. FUNDERLIC† AND GENE H. GOLUB‡

Abstract. A variational formulation for the generalized singular value decomposition (GSVD) of a pair of matrices $A \in R^{m\times n}$ and $B \in R^{p\times n}$ is presented. In particular, a duality theory analogous to that of the SVD provides new understanding of left and right generalized singular vectors. It is shown that the intersection of row spaces of A and B plays a key role in the GSVD duality theory. The main result that characterizes left GSVD vectors involves a generalized singular value deflation process.

Key words. Generalized Eigenvalue and Eigenvector, Generalized Singular Value and Singular Vector, Stationary Value and Stationary Point, Deflation, Duality.

AMS(MOS) subject classifications. 65F15, 65H15.

1. Introduction.

The singular value decomposition (SVD) of a given matrix $A \in R^{m\times n}$ is

$$U^T AV = S = \mathrm{diag}\{\sigma_1, \ldots, \sigma_q\},\ q = \min\{m, n\} \tag{1}$$

where $U \in R^{m\times m}$ and $V \in R^{n\times n}$ are orthogonal matrices, $S \in R^{m\times n}$ is zero except for the real nonnegative elements $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_q = 0$ on the leading diagonal with $r = \mathrm{rank}(A)$. The $\sigma_i, i = 1, \ldots, q$ are the singular values of A. Of the many ways to characterize the singular values of A, the following variational property is of particular interest [4].

THEOREM 1.1. *Consider the optimization problem*

$$\max_{x\neq 0} \frac{\|Ax\|}{\|x\|}, \tag{2}$$

where $\|\cdot\|$ denotes the 2-norm of a vector. Then the singular values of A are precisely the stationary values, i.e., the functional evaluations at the stationary points, of the objective function $\|Ax\|/\|x\|$ with respect to $x \neq 0$.

We note that the stationary points $x \in R^n$ in the problem (2) are the right singular vectors of A. At each of such points, it follows from the usual duality theory that there exists a vector $y \in R^m$ of unit Euclidean length such that $y^T Ax$ is equal to the corresponding stationary value. This y is the corresponding left singular vector of A.

The main purpose of this paper is to delineate a similar variational principle that leads to the generalized singular value decomposition (GSVD) of a pair of matrices $A \in R^{m\times n}$ and $B \in R^{p\times n}$. While the variational formula analogous to (2) for the GSVD

* Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205. This research was supported in part by the National Science Foundation under grants DMS-9123448 and DMS-9422280.

† Department of Computer Science, North Carolina State University, Raleigh, NC 27695-8206.

‡ Department of Computer Science, Stanford University, Stanford, CA 94305. This research was supported in part by National Science Foundation under grant CCR-9505393.

is well-known, the corresponding duality theory has apparently not been developed. (See [1] for a related treatise.) The purpose of this note is to fill the duality theory gap for the GSVD problem.

Let $\mathcal{R}(M)$ and $\mathcal{N}(M)$ denote, respectively, the range space and the null space of any given matrix M. We will see that the intersection of row spaces of A and B,

$$\mathcal{R}(A^T)\bigcap\mathcal{R}(B^T) = \left\{z \in R^n \mid z^T = x^T A = y^T B \text{ for some } x \in R^m \text{ and } y \in R^p\right\},$$

plays a fundamental role in the duality theory of the associated GSVD. The equivalence

$$C\begin{bmatrix} x \\ -y \end{bmatrix} = [A^T, B^T]\begin{bmatrix} x \\ -y \end{bmatrix} = 0 \Longleftrightarrow x^T A = y^T B$$

suggests that the null space of the matrix $C := [A^T, B^T]$,

$$\mathcal{N}(C) = \left\{\begin{bmatrix} x \\ -y \end{bmatrix} \in R^{m+p} \mid C\begin{bmatrix} x \\ -y \end{bmatrix} = 0\right\}, \tag{3}$$

may be interpreted as a "representation" of $\mathcal{R}(A^T)\bigcap\mathcal{R}(B^T)$. But this representation is not unique in that different values of $\begin{bmatrix} x \\ -y \end{bmatrix} \in \mathcal{N}(C)$ may give rise to the same $z \in \mathcal{R}(A^T)\bigcap\mathcal{R}(B^T)$. In particular, all points in the subspace

$$\mathcal{S} := \left\{\begin{bmatrix} g \\ -h \end{bmatrix} \in R^{m+p} \mid g^T A = h^T B = 0\right\} \tag{4}$$

collapse into the zero vector in $\mathcal{R}(A^T)\bigcap\mathcal{R}(B^T)$. For a reason to be discussed in the sequel (see (16) and the argument thereafter), the subspace $\mathcal{S}$ should be taken out of consideration. More precisely, define the *homomorphism* $H : \mathcal{N}(C) \longrightarrow \mathcal{R}(A^T)\bigcap\mathcal{R}(B^T)$ by

$$z = H(\begin{bmatrix} x \\ -y \end{bmatrix}) \Longleftrightarrow z^T = x^T A = y^T B,$$

and define, for every $\begin{bmatrix} x \\ -y \end{bmatrix} \in \mathcal{N}(C)$, the *quotient map* $\pi(\begin{bmatrix} x \\ -y \end{bmatrix})$ to be the coset of $\mathcal{S}$ containing $\begin{bmatrix} x \\ -y \end{bmatrix}$, i.e.,

$$\pi(\begin{bmatrix} x \\ -y \end{bmatrix}) := \begin{bmatrix} x \\ -y \end{bmatrix} + S. \tag{5}$$

Then the first homomorphism theorem for vector spaces (See, for example, [5, Theorem 4.a]) states that $\mathcal{R}(A^T)\bigcap\mathcal{R}(B^T)$ is *isomorphic* to the quotient space $\mathcal{N}(C)/\mathcal{S}$ where

$$\mathcal{N}(C)/\mathcal{S} := \left\{\pi(\begin{bmatrix} x \\ -y \end{bmatrix}) \mid \begin{bmatrix} x \\ -y \end{bmatrix} \in \mathcal{N}(C)\right\}. \tag{6}$$

It is in this quotient space that we establish the duality theory.

Recall that linearly independent vectors in $\mathcal{N}(C)$ that are not in $\mathcal{S}$ will generate naturally linearly independent vectors in the quotient space $\mathcal{N}(C)/\mathcal{S}$ through the quotient map. Thus the simplest way to represent $\mathcal{N}(C)/\mathcal{S}$ is through the orthogonal complement $\mathcal{S}^\perp$ of $\mathcal{S}$ in $\mathcal{N}(C)$. Define $N(A^T)$ and $N(B^T)$ to be matrices so that their columns span, respectively, the null spaces $\mathcal{N}(A^T)$ and $\mathcal{N}(B^T)$. Define

$$Z := \begin{bmatrix} A^T & B^T \\ N(A^T)^T & 0 \\ 0 & N(B^T)^T \end{bmatrix}. \tag{7}$$

Then $\mathcal{N}(C)/\mathcal{S}$ can be uniquely represented by the subspace

$$\mathcal{N}(Z) = \left\{ \begin{bmatrix} x \\ -y \end{bmatrix} \in R^{m+p} \mid Z \begin{bmatrix} x \\ -y \end{bmatrix} = 0 \right\}. \tag{8}$$

We shall have the dimension counted carefully in §2.

Our discussion is based upon the following formulation of the GSVD for A and B by Paige and Saunders [6] (or QSVD in [3]) that generalizes the original concept in [9]:

DEFINITION 1.1. *Assume rank*$(C) = k$, *then there exist orthogonal* $U \in R^{m\times m}$ *and* $V \in R^{p\times p}$ *and an invertible* $X \in R^{n\times n}$ *such that*

$$\begin{bmatrix} U^T & 0 \\ 0 & V^T \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} X = \begin{bmatrix} \Omega_A & 0 \\ \Omega_B & 0 \end{bmatrix} \tag{9}$$

with $\Omega_A \in R^{m\times k}$ *and* $\Omega_B \in R^{p\times k}$ *given by*

$$\Omega_A = \begin{array}{c} \begin{array}{ccc} r & s & k-r-s \end{array} \\ \begin{bmatrix} I_A & & \\ & S_A & \\ & & O_A \end{bmatrix} \end{array} \begin{array}{l} \\ r \\ s \\ m-r-s \end{array}$$

$$\Omega_B = \begin{array}{c} \begin{array}{ccc} r & s & k-r-s \end{array} \\ \begin{bmatrix} O_B & & \\ & S_B & \\ & & I_B \end{bmatrix} \end{array} \begin{array}{l} \\ p-k+r \\ s \\ k-r-s \end{array}$$

where I_A *and* I_B *are identity matrices,* O_A *and* O_B *are zero matrices with possibly no rows or no columns, and* $S_A = diag\{\omega_A^{(1)}, \ldots, \omega_A^{(s)}\}$ *and* $S_B = diag\{\omega_B^{(1)}, \ldots, \omega_B^{(s)}\}$ *satisfy*

$$1 > \omega_A^{(1)} \geq \ldots \geq \omega_A^{(s)} > 0, \qquad 0 < \omega_B^{(1)} \leq \ldots \leq \omega_B^{(s)} < 1,$$
$$\omega_A^{(i)^2} + \omega_B^{(i)^2} = 1.$$

The quotients

$$\lambda_i := \frac{\omega_A^{(i)}}{\omega_B^{(i)}},\ i = 1, \ldots, s$$

are called the generalized singular values of (A, B) *for which we make use of the notation* $\Lambda := diag\{\lambda_1, \ldots, \lambda_s\}$. *The values of* r *and* s *are defined internally by the matrices* A *and* B.

Suppose we partition X into four blocks of columns $X = [X_1, X_2, X_3, X_4]$ with column sizes $r, s, k-r-s$ and $n-k$, respectively. Correspondingly, suppose we partition U into $U = [U_1, U_2, U_3]$ with column sizes $r, s, m-r-s$ and V into $V = [V_1, V_2, V_3]$ with column sizes $p-k+r, s, k-r-s$, respectively. Observe that

$$\begin{aligned} X^T A^T A X &= \begin{bmatrix} I_A & & & 0 \\ & S_A^2 & & \vdots \\ & & O_A^T O_A & \\ 0 & \ldots & & 0 \end{bmatrix}, \\ X^T B^T B X &= \begin{bmatrix} O_B^T O_B & & & 0 \\ & S_B^2 & & \vdots \\ & & I_B & \\ 0 & \ldots & & 0 \end{bmatrix} \end{aligned}$$

where, for simplicity, we have used "0" to denote various zero matrices with appropriate sizes. Upon examining the second column block, we notice that

$$A^T A X_2 = B^T B X_2 \Lambda^2.$$

That is, $\{\lambda_i^2 | i = 1, \ldots, s\}$ is a subset of the eigenvalues of the symmetric pencil

$$A^T A - \mu B^T B. \tag{10}$$

Similarly, we point out the following remarks to include all other cases [3, 6]:

1. If $k < n$, then $A^T A X_4 = B^T B X_4 = 0$ implies that *every* complex number is an eigenvalue of (10). This is the case that is considered of little interest. We will refer to eigenvalues of this type as *defective*.
2. Since $A^T A X_3 = 0$ and $B^T B X_3 \neq 0$, the pencil (10) has 0 as an eigenvalue with multiplicity $k-r-s$.
3. Since $B^T B X_1 = 0$ and $A^T A X_1 \neq 0$, we may regard that the pencil (10) has ∞ as an eigenvalue with multiplicity r.

We view the relationships

$$\begin{aligned} U_2^T A X_2 &= S_A, \\ V_2^T B X_2 &= S_B. \end{aligned}$$

as the fundamental and most important components of (9). We refer to the corresponding columns of U_2 and V_2 as the left generalized singular vectors of A and B, respectively. Note that there are two such left vectors for each generalized singular value, one for A, and one for B.

Similar to Theorem 1.1, we have the following variational formulation.

THEOREM 1.2. *Consider the optimization problem*

$$\max_{Bx \neq 0} \frac{\|Ax\|}{\|Bx\|}. \tag{11}$$

Then the generalized singular values, $\lambda_1, \ldots, \lambda_s$, of (A, B) are precisely the nonzero finite stationary values of the objective function in (11).

Proof. The stationary values of $\|Ax\|/\|Bx\|$ are square roots of those of the function

$$f(x) := \frac{\langle Ax, Ax \rangle}{\langle Bx, Bx \rangle}$$

where $\langle \cdot, \cdot \rangle$ is the standard Euclidean inner product. It is not difficult to see that the gradient of f at x where $Bx \neq 0$ is given by

$$\nabla f(x) = \frac{2}{\langle Bx, Bx \rangle} \left(A^T Ax - f(x) B^T Bx \right).$$

The theorem follows from comparing the first order condition $\nabla f(x) = 0$ with (10). $\square$

Obviously, the corresponding stationary points $x \in R^n$ for the problem (11) are related to columns of the matrix X_2 (up to scalar multiplications), which are also eigenvectors of the pencil (10). What is not clear are the roles that U_2 and V_2 play in terms of the variational formula (11). In this note we present some new insights in this regard.

In the usual SVD duality theory the left singular vectors can be obtained from the optimization problem

$$\max_{y \neq 0} \frac{\|y^T A\|}{\|y\|}, \tag{12}$$

a formula similar to (2). Thus one might first guess that the duality theory analogous to (11) would be the problem

$$\max_{y^T B \neq 0} \frac{\|y^T A\|}{\|y^T B\|}.$$

However, this is certainly not a correct form as a single row vector y^T is not compatible for left multiplication on both A and B. We will see correct dual forms for the GSVD in (18) and (23).

2. Duality Theory.

For convenience, we denote

$$\begin{aligned} U_2 &= [u_1^{(2)}, \ldots, u_s^{(2)}] \\ V_2 &= [v_1^{(2)}, \ldots, v_s^{(2)}]. \end{aligned}$$

It follows from

$$U_2^T AX = S_A S_B^{-1} V_2^T BX = \Lambda V_2^T BX \tag{13}$$

that $U_2^T A = \Lambda V_2^T B$ or equivalently

$$C \begin{bmatrix} U_2 \\ -V_2 \Lambda \end{bmatrix} = 0. \tag{14}$$

Note that the columns of both U_2 and V_2 are unit vectors. Given any $\begin{bmatrix} x \\ -y \end{bmatrix}$ in the null space of C with $\|x\| \neq 0$ and $\|y\| \neq 0$, we observe that

$$C \begin{bmatrix} \frac{x}{\|x\|} \\ -\frac{\|y\|}{\|x\|} \frac{y}{\|y\|} \end{bmatrix} = \frac{1}{\|x\|} C \begin{bmatrix} x \\ -y \end{bmatrix} = 0 \tag{15}$$

where $x/\|x\|$ and $y/\|y\|$ are also unit vectors. Comparing (15) with the relationship (14), we are motivated to consider the role that each generalized singular value λ_i plays in the optimization problem:

$$\max_{C \begin{bmatrix} x \\ -y \end{bmatrix} = 0, x \neq 0} \frac{\|y\|}{\|x\|}. \tag{16}$$

However, we need to hastily point out a subtle flaw in the formulation of (16). Consider a given point $\begin{bmatrix} x \\ -y \end{bmatrix} \in \mathcal{S}$ with $\|x\| \neq 0$ and $\|y\| \neq 0$. Then $\begin{bmatrix} \alpha x \\ -\beta y \end{bmatrix} \in \mathcal{S}$ for arbitrary $\alpha, \beta \in R$. In this case, the optimization subproblem

$$\max_{x^T A = y^T B = 0, x \neq 0} \frac{\|y\|}{\|x\|} \tag{17}$$

becomes the problem

$$\max_{\alpha, \beta \in R, \alpha \neq 0} \frac{|\beta|}{|\alpha|}$$

that obviously has no stationary point at all and has maximum infinity. The trouble persists so long as $\begin{bmatrix} x \\ -y \end{bmatrix}$ contains components from $\mathcal{S}$. It is for this reason that the subspace $\mathcal{S}$ should be taken out of consideration. We should, instead of (16), consider the modified optimization problem (See (7) and (8))

$$\max_{\begin{bmatrix} x \\ -y \end{bmatrix} \in \mathcal{N}(Z), x \neq 0} \frac{\|y\|}{\|x\|}. \tag{18}$$

We will prove that each λ_i corresponds to a stationary value for the problem (18). But first it is worthy to point out some interesting remarks:

1. The optimization problem (18) is consistent with the ordinary singular value problem where $B = I$. In this case, $Z = \begin{bmatrix} A^T & I \\ N(A^T)^T & 0 \end{bmatrix}$. Thus $\begin{bmatrix} x \\ -y \end{bmatrix} \in \mathcal{N}(Z)$ implies that $y = A^T x \neq 0$. The forbidden situation $x^T A = y^T = 0$ in (18) is not a concern in this case because of the homogeneity in x, and only implies that 0 is a stationary value (or equivalently A has a zero singular value.) Thus in the case of the ordinary SVD the problem (18) reduces to (12).
2. It is clear that $\dim(\mathcal{N}(C)) = m + p - k$ since we assume $\operatorname{rank}(C) = k$. The structure involved in (9) implies that for $\mathcal{S}$ defined in (4) it must be

   $$\mathcal{S} = \mathcal{R}(U_3) \oplus \mathcal{R}(V_1).$$

   That is, the size of $N(A^T)$ and $N(B^T)$ should be $m \times (m - r - s)$ and $p \times (p - k + r)$, respectively. It follows that $\dim(\mathcal{S}) = m + p - k - s$. The space we are interested in is the quotient space $\mathcal{N}(C)/\mathcal{S}$. It is known from the homomorphism theorem that $\dim(\mathcal{N}(C)/\mathcal{S}) = \dim(\mathcal{N}(C)) - \dim(\mathcal{S})$ [5, Lemma 4.8]. Thus $\dim(\mathcal{N}(C)/\mathcal{S}) = s$. We will see below that this dimension count agrees with our assumption that there are s generalized singular values.

The following theorem is critical to the study of stationary values and stationary points of the optimization problem (18).

THEOREM 2.1. *Let the columns of the matrix* $\begin{bmatrix} \Phi \\ \Psi \end{bmatrix}$ *with* $\Phi \in R^{m\times s}$ *and* $\Psi \in R^{p\times s}$ *be a basis for the subspace* $\mathcal{N}(Z)$. *Then the non-defective finite nonzero eigenvalues of the symmetric pencil of (10),*

$$A^T A - \mu B^T B,$$

are the same as those of the pencil

$$\Psi^T\Psi - \lambda\Phi^T\Phi. \tag{19}$$

Proof. Suppose $A^T Ax = \mu B^T Bx$. Since μ is non-defective and nonzero, $A^T Ax = \mu B^T Bx \neq 0$. That is, $\begin{bmatrix} Ax \\ -\mu Bx \end{bmatrix}$ represents a nonzero element in $\mathcal{N}(C)/\mathcal{S}$. Thus there exist vectors $v \in R^s$, $v \neq 0$, $\xi_A \in R^{m-r-s}$, $\xi_B \in R^{p-k+r}$ such that

$$\begin{aligned} Ax &= \Phi v + N(A^T)\xi_A \\ -\mu Bx &= \Psi v + N(B^T)\xi_B. \end{aligned}$$

It follows that

$$(\Psi^T\Psi - \mu\Phi^T\Phi)v = -\mu(\Phi^T A + \Psi^T B)x + (\mu\Phi^T N(A^T)\xi_A - \Psi^T N(B^T)\xi_B) = 0.$$

In the above, we have used the fact that $Z \begin{bmatrix} \Phi \\ \Psi \end{bmatrix} = 0$. This shows that μ is an eigenvalue of (19) with v as the corresponding eigenvector.

To complete the eigenvalue (generalized singular value) set equality, suppose now that $(\Psi^T\Psi - \lambda\Phi^T\Phi)v = 0$ with $\lambda \neq 0, \infty$ and $v \neq 0$. We want to show that the equation

$$\begin{bmatrix} A \\ -\lambda B \end{bmatrix} x = \begin{bmatrix} \Phi v \\ \Psi v \end{bmatrix} \tag{20}$$

has a solution x. If this can be done, then since $[A^T, B^T] \begin{bmatrix} \Phi \\ \Psi \end{bmatrix} = 0$ it follows that x is an eigenvector of the pencil $A^TA - \mu B^TB$ with eigenvalue λ.

To show (20) means to show that the vector $\begin{bmatrix} \Phi v \\ \Psi v \end{bmatrix}$ is in the column space of the matrix $\begin{bmatrix} A \\ -\lambda B \end{bmatrix}$. It suffices to show that

$$\begin{bmatrix} \Phi v \\ \Psi v \end{bmatrix} \perp \begin{bmatrix} y \\ z \end{bmatrix} \tag{21}$$

wherever

$$\begin{bmatrix} A^T, & -\lambda B^T \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = 0. \tag{22}$$

Rewrite (22) as $[A^T, B^T] \begin{bmatrix} y \\ -\lambda z \end{bmatrix} = 0$, showing that $\begin{bmatrix} y \\ -\lambda z \end{bmatrix} \in \mathcal{N}(C)$. We therefore must have

$$\begin{aligned} y &= \Phi w + N(A^T)\eta_A \\ -\lambda z &= \Psi w + N(B^T)\eta_B \end{aligned}$$

for some vectors w, η_A and η_B of appropriate size. Substituting y and z into (21) implies

$$\begin{bmatrix} y^T, & z^T \end{bmatrix} \begin{bmatrix} \Phi v \\ \Psi v \end{bmatrix} = w^T(\Phi^T\Phi v - \frac{1}{\lambda}\Psi^T\Psi v) + (\eta_A^T N(A^T)^T\Phi v - \frac{1}{\lambda}\eta_B N(B^T)^T\Psi v = 0.$$

The assertion is therefore proved. □

COROLLARY 2.2. *The generalized singular values λ_i, $i = 1, \ldots, s$, are the stationary values associated with the optimization problem (18).*

Proof. We have already seen in Theorem 1.2 on how the generalized singular values of (A, B) are related to the pencil $A^TA - \mu B^TB$, which are now related to the pencil $\Psi^T\Psi - \lambda\Phi^T\Phi$. By Theorem 1.2 again, we conclude that the generalized singular values of (A, B) can be found from the stationary values associated with the optimization problem

$$\max_{\Phi v \neq 0} \frac{\|\Psi v\|}{\|\Phi v\|}. \tag{23}$$

which is equivalent to (18). □

We now characterize the stationary points of (18). In particular, we prove the following result which completes our duality theory. Aside from the fundamental connection between the GSVD and its duality theory, the eigenvalue *deflation* of the proof should be of special interest in its own right.

THEOREM 2.3. *Suppose* $\begin{bmatrix} x_1 \\ -y_1 \end{bmatrix} \cdots \begin{bmatrix} x_s \\ -y_s \end{bmatrix}$ *are stationary points for the problem (18) with corresponding stationary values* $\lambda_1, \ldots, \lambda_s$. *Define*

$$u_i := \frac{x_i}{\|x_i\|}, \tag{24}$$

$$v_i := \frac{y_i}{\|y_i\|}. \tag{25}$$

Then the columns of the matrices $\tilde{U} := [u_1, \ldots, u_s]$ *and* $\tilde{V} := [v_1, \ldots, v_s]$ *are the left generalized singular vectors of* A *and* B*, respectively.*

Proof. Suppose $\begin{bmatrix} x_1 \\ -y_1 \end{bmatrix}$ is an associated stationary point of (18) with the stationary value λ_1. (The ordering of which stationary value is found is immaterial in the following discussion. We assume λ_1 is found first.) Taking this vector to be the first basis vector in $\mathcal{N}(Z)$, we may write

$$\begin{bmatrix} \Phi \\ \Psi \end{bmatrix} = \begin{bmatrix} x_1 & \Phi_2 \\ -y_1 & \Psi_2 \end{bmatrix}$$

where $\Phi_2 \in R^{m\times(s-1)}$ and $\Psi_2 \in R^{p\times(s-1)}$ are to be defined below. Consider the stacked matrix

$$Z_2 := \begin{bmatrix} A^T & B^T \\ N(A^T)^T & 0 \\ 0 & N(B^T)^T \\ x_1^T & 0 \end{bmatrix}.$$

Note that, due to the last row in Z_2, the null space of Z_2 is a proper subspace of that of Z with one less dimension. We may therefore use a basis of the null space of Z_2 to form the columns of the matrix $\begin{bmatrix} \Phi_2 \\ \Psi_2 \end{bmatrix}$. In this way, we attain the additional property that

$$x_1^T \Phi_2 = 0.$$

Note that the eigenvector of (19) corresponding to eigenvalue λ_1 is the same as the stationary point for the problem (23) with stationary value λ_1. Since (23) is simply a coordinate representation of (18) and we already assume that $\begin{bmatrix} x_1 \\ -y_1 \end{bmatrix}$ is a stationary point associated with (18), the eigenvector of (19) corresponding to λ_1 must be the unit vector $e_1 \in R^q$. It follows that

$$y_1^T \Psi_2 = 0,$$

and hence

$$\Psi^T\Psi - \lambda\Phi^T\Phi = \begin{bmatrix} y_1^T y_1 - \lambda x_1^T x_1 & 0 \\ 0 & \Psi_2^T\Psi_2 - \lambda\Phi_2^T\Phi_2 \end{bmatrix}.$$

We have thus shown that the eigenvalues of the pencil $\Psi_2^T\Psi_2 - \lambda\Phi_2^T\Phi_2$ are exactly those of the pencil $\Psi^T\Psi - \lambda\Phi^T\Phi$ with λ_1 excluded. Note that the submatrix $\begin{bmatrix} \Phi_2 \\ \Psi_2 \end{bmatrix}$ spans a null subspace of Z that is complementary to the vector $\begin{bmatrix} x_1 \\ -y_1 \end{bmatrix}$. After the first stationary point is found, we may therefore deflate (18) to the problem

$$\max_{Z_2\begin{bmatrix} x \\ -y \end{bmatrix}=0, x\neq 0} \frac{\|y\|}{\|x\|}. \tag{26}$$

A stationary point of (26) will also be a stationary point of (18) since it gives the same stationary value in both problems. This deflation procedure may be continued until all nonzero stationary values are found.

Then, by construction, $\tilde{U}^T\tilde{U} = I$ and $\tilde{V}^T\tilde{V} = I$. Furthermore, we have

$$\tilde{U}^T A = \Lambda \tilde{V}^T B,$$

which completes the proof. □

That is, we have derived two matrices $\tilde{U}$ and $\tilde{V}$ that play the same role as that of U_2 and V_2, in (9) respectively.

3. Summary. We have discussed a variational formulation for the GSVD of a pair of matrices. In particular, we characterize the role of the left generalized singular vectors in this formulation.

We summarize the analogies between the SVD and the GSVD in the following table. The stationary values in any of the variational formulations below give rise to the corresponding singular values.

There is a close correspondence between the (generalized) eigenvalue problem and the (generalized) singular value problem, as is indicated in Theorem 1.1 and Theorem 1.2. The result in Theorem 2.1 and Theorem 2.3 apparently are new, and shed lights on the understanding of the left singular vectors.

Some of the available numerical methods and approaches for computing the GSVD are available in [2, 7, 8, 10]. The deflation process used in the characterization of the left singular vectors can be carried out effectively by updating techniques [4]. We anticipate that the discussion here might lead to a new numerical algorithm, especially when a few singular values are required and the matrix C is sparse.

Acknowledgment. We want to thank an anonymous referee for the many valuable suggestions that significantly improve this paper.

	Regular Problem	Generalized Problem
Decomposition	$U^T A = S V^T$ See formula (1)	$U^T A X = \Lambda V^T B X$ See Formula (13)
Right Singular Vector	V	X
Variational Formula (including zero σ_i, λ_i)	$\max_{x \neq 0} \frac{\|\|Ax\|\|}{\|\|x\|\|}$ See formula (2)	$\max_{Bx \neq 0} \frac{\|\|Ax\|\|}{\|\|Bx\|\|}$ See formula (11)
Left Singular Vector	U	$[U^T, V^T]^T$
Variational Formula (only positive σ_i, λ_i)	$\max_{\begin{bmatrix} A^T & I \\ N(A^T)^T & 0 \end{bmatrix}\begin{bmatrix} x \\ -y \end{bmatrix}=0, x\neq 0} \frac{\|\|y\|\|}{\|\|x\|\|}$ $(= \max_{A^T x \neq 0, x \neq 0} \frac{\|\|A^T x\|\|}{\|\|x\|\|})$ See formula (12)	$\max_{\begin{bmatrix} A^T & B^T \\ N(A^T)^T & 0 \\ 0 & N(B^T)^T \end{bmatrix}\begin{bmatrix} x \\ -y \end{bmatrix}=0, x\neq 0} \frac{\|\|y\|\|}{\|\|x\|\|}$ See formula (18)

TABLE 1
Comparison of variational formulations between SVD and GSVD.

REFERENCES

[1] B. Ånström, The generalized singular value decomposition and $(A - \lambda B)$-problem, BIT, 24(1984), 568-583.

[2] Z. Bai and J. W. Demmel, Computing the generalized singular value decomposition, SIAM J. Sci. Comput., 14(1993), 1464-1486.

[3] B. De Moor and G. H. Golub, Generalized singular value decomposition: A proposal for a standardized nomenclature, Manuscript NA-89-05, Stanford University, 1989.

[4] G. H. Golub and C. F. Van Loan, Matrix Computations, 2nd ed., The John Hopkins University Press, Baltimore, MD, 1989.

[5] I. N. Herstein, Topics in Algebra, 2nd ed., Xerox College Publishing, Lexington, MS, 1975.

[6] C. C. Paige and M. A. Saunders, Towards a generalized singular value decomposition, SIAM J. Numer. Anal., 18(1981), 398-405.

[7] C. C. Paige, Computing the generalized singular value decomposition, SIAM J. Sci. Stat. Comput., 7(1986), 1126-1146.

[8] G. W. Stewart, Computing the CS-decomposition of a partitioned orthonormal matrix, Numer. Math., 40(1982), 297-306.

[9] C. F. Van Loan, Generalizing the singular value decomposition, SIAM J. Numer. Anal., 13(1976), 76-83.

[10] C. F. Van Loan, Computing the CS and the generalized singular value decomposition, Numer. Math., 46(1985), 479-491.