On the Structure and Geometry of the Product Singular Value Decomposition*

Bart L. R. De Moor†
ESAT
Department of Electrical Engineering
Katholieke Universiteit Leuven
Kardinaal Mercierlaan 94
B-3001 Leuven, Belgium

Submitted by Richard A. Brualdi

ABSTRACT

The product singular value decomposition is a factorization of two matrices, which can be considered as a generalization of the ordinary singular value decomposition, at the same level of generality as the quotient (generalized) singular value decomposition. A constructive proof of the product singular value decomposition is provided, which exploits the close relation with a symmetric eigenvalue problem. Several interesting properties are established. The structure and the nonuniqueness properties of the so-called contragredient transformation, which appears as one of the factors in the product singular value decomposition, are investigated in detail. Finally, a geometrical interpretation of the structure is provided in terms of principal angles between subspaces.

1. INTRODUCTION

The ordinary singular value decomposition (OSVD) has become an important tool in the analysis and numerical solution of numerous problems. Not only does it allow for an elegant problem formulation, but at the same time it provides geometrical and algebraic insight accompanied by a numerically stable implementation of the solution. Several algorithms and applica-

*Research supported in part by the U.S. Army under contract DAAL03-87-K-0095, an Advanced Research Fellowship in Science and Technology of the NATO Science Fellowships Programme, and a grant from IBM.

†Research Associate of the Belgian National Fund for Scientific Research. Telephone: 32/16/220931; fax: 32/16/221855; E-mail: demoor@esat.kuleuven.ac.be.

655 Avenue of the Americas, New York, NY 10010 0024-3795/92/$5.00

tions are discussed in e.g. [7], [12], and the references therein. Recently, several generalizations of the singular value decomposition have been derived and analysed. The best-known example is the so called "generalized" singular value decomposition of Van Loan [19] and Paige and Saunders [18]. In [4], we propose to call it the *quotient singular value decomposition* (QSVD), as opposed to the *product singular value decomposition* (PSVD), which was introduced in its explicit form by Fernando and Hammarling in [8] (who called it the ΠSVD). In [20], Zha introduced yet another generalization of the OSVD, this time for three matrices, which was called the *restricted singular value decomposition* (RSVD) in [4] and [3]. In [4] we have proposed a standardized nomenclature for generalizations of the OSVD, and we shall use it in this paper.

A common feature of all these generalizations is that they are related to the OSVD on the one hand and to generalized eigenvalue problems on the other hand. While a lot of their properties and structure can be established by exploiting these relationships, the explicit forms of the generalizations themselves are important in their own right: Not only do they possess a richer structure than their corresponding generalized eigenvalue problems, but it is expected that their direct numerical computation will be better behaved than the computation via transformation to a generalized eigenvalue or OSVD problem. The reason is that, typically, generalizations of the OSVD are related to the OSVD or to generalized eigenvalue problems by AA^t squaring type of operations or matrix (pseudo)inversions, which may cause nontrivial losses of numerical accuracy when implemented on a finite precision machine.

The PSVD is a generalization for two matrices of the OSVD. In this respect, it is a kind of "dual" generalization of the OSVD with respect to the QSVD. For instance, we have shown in [3] that both the PSVD and the QSVD play an important role in the construction of the RSVD, which is a generalization of the OSVD for three matrices. Hence, it can be expected that the structural and geometrical properties of both the PSVD and the QSVD will play an important role in the future work on formulations, numerical implementations, and applications of other generalizations of the OSVD.[1]

While the geometrical properties and numerical implementations of the OSVD and QSVD are by now well understood, similar knowledge for the PSVD is less well developed. It is one of the goals of this paper to provide some more insight into the *structure and geometry* of the PSVD.

[1]As a matter of fact, recently, Zha Hongyuang and the author in [5] have established a most interesting result that both the PSVD and the QSVD are "parents" of an infinite chain of generalizations of the OSVD for any number of matrices.

Algorithmic ideas to actually implement the PSVD in a numerically robust way can be found in [8] and [13]. Applications include the orthogonal Procrustes problem [12], computing balancing transformations for state space systems [8, 16], and computing the Kalman decomposition of a linear system [9]. The PSVD could also be applied in the computation of approximate intersections between subspaces in the stochastic realization problem [1], as an alternative to canonical correlation analysis. The main difference between the two approaches lies in the fact that canonical correlation analysis first performs a normalization of the data, hence normalizing the relevant signal energy and the pure noise energy to the same level, while the PSVD can be considered as a way of decomposing the cross-covariance matrix into canonical directions, without an *a priori* normalization. However, these issues will not be discussed in this paper.

The main results of this paper concentrate around two constructive proofs of the PSVD. The first one, in Section 2, exploits the close relationship of the PSVD to the OSVD and several eigenvalue problems. In the second proof, given in Section 3, we provide a profound analysis of the nonuniqueness properties of the so-called contragredient transformation which appears as one of the factors in the PSVD. Surprisingly enough, this turns out to be a considerably complicated problem. In essence, our result is a parametrization of all contragredient transformations for two symmetric nonnegative definite matrices of the form A^tA and B^tB in terms of matrices that can be derived from the OSVDs of the two matrices A and B.

NOTATION AND ABBREVIATIONS. All matrices and vectors in this paper are real. Matrices are denoted by capitals, and vectors by lowercase letters other than $i, j, k, l, m, n, p, q, r$, which are nonnegative integers. Scalars are denoted by Greek letters. The range (column space) of a matrix A will be denoted by $R(A)$, its row space by $R(A^t)$, its null space by $N(A)$. The orthogonal projection of the column space of a matrix B onto the column space of a matrix A is denoted by $\Pi_A R(B)$. The orthogonalization of the column space of a matrix B to the column space of a matrix A is denoted by $\Pi_A^{\perp} R(B)$. The subspace that is the intersection of the column spaces of two matrices A and B is denoted by $R(A) \cap R(B)$. The direct sum of two mutually orthogonal subspaces $R(U_1)$ and $R(U_2)$ $(U_1^tU_2 = 0)$ is denoted by $R(U_1) \oplus R(U_2)$. The dimension of a subspace is abbreviated as dim; hence $\dim R(A) = \operatorname{rank} A = \dim R(A^T)$. By $\#\{\sigma(A) = 1\}$ we denote the number of singular values of A equal to 1.

2. THE PRODUCT SINGULAR VALUE DECOMPOSITION

In this section, we shall first state the main theorem and provide a constructive proof of the PSVD, which is based on some results that relate the OSVD of the matrix AB^tBA^t to the eigenvalue decomposition of the matrices B^tBA^tA and A^tAB^tB. We shall also prove a lemma that permits us to express the PSVD of the matrix pair A,B in terms of their OSVDs when $AB^t=0$. In Section 2.2, we shall provide a variational characterization of the PSVD.

2.1. *A Constructive Proof of the PSVD*

THEOREM 1 (The PSVD). *Every pair of real matrices A $(m\times n)$ and B $(p\times n)$ can be factorized as*

$$A=U_AS_AX^t,$$

$$B=U_BS_BX^{-1}.$$

All the matrices are real. The matrices U_A, U_B are square orthonormal, and X is square nonsingular. S_A and S_B have the following structure:

$$S_A=\begin{array}{c} \\ r_1 \\ r_a-r_1 \\ m-r_a \end{array}\begin{array}{c} \begin{array}{cccc} r_1 & r_a-r_1 & r_b-r_1 & n-r_a-r_b+r_1 \end{array} \\ \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array},$$

$$S_B=\begin{array}{c} \\ r_1 \\ r_b-r_1 \\ p-r_b \end{array}\begin{array}{c} \begin{array}{cccc} r_1 & r_a-r_1 & r_b-r_1 & n-r_a-r_b+r_1 \end{array} \\ \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array},$$

where S_1 is square diagonal with positive elements and $r_1=$ rank(AB^t).

While some related eigenvalue problems were discussed in [13] and [16], the explicit formulation of the PSVD is in Theorem 1 was given for the first time by Fernando and Hammarling in [8], who called it the ΠSVD.[2]

[2] In [8], also a constructive proof was provided. It is however based on a lemma (Lemma 1 in [8]) of which the proof is not correct. To give a counterexample to the proof, consider the pair of matrices

$$A=\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \end{pmatrix},$$

$$B=\begin{pmatrix} 3 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Throughout the paper, we shall also use the matrix Y defined as $Y = X^{-t}$. In [8], the factorization is presented in a slightly different form, where a QR factorization of X is used. While this may be preferable in analysing *numerical* issues related to the PSVD, such an additional factorization is not relevant for our present purpose, which is the detailed exploration of structural and geometrical properties. We propose to call the pairs of nonzero elements of S_A and S_B in Theorem 1 the *product singular value pairs*, and their product the *product singular values*. Obviously, the pairs contain more structural information than the product singular values. There are four possibilities: There are r_1 pairs of the form $(\sqrt{\sigma_i}, \sqrt{\sigma_i})$ with corresponding product singular value σ_i, $i = 1,\dots,r_1$. By convention, they are ordered so that $\sigma_i \geqslant \sigma_{i+1}$. There are $r_a - r_1$ pairs $(1,0)$ with corresponding product singular value 0. There are $r_b - r_1$ pairs $(0,1)$ with corresponding product singular value 0. There are $n - r_a - r_b + r_1$ pairs $(0,0)$, which we shall call the trivial product singular value pairs, in analogy with the trivial quotient singular value pairs [4].

In the constructive proof of Theorem 1, we shall need the following four lemmas:

LEMMA 1. *Let the OSVD of a matrix A be given as*

$$A = (U_{a1} \quad U_{a2}) \begin{pmatrix} S_{a1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_{a1}^t \\ V_{a2}^t \end{pmatrix}.$$

Then the set of solutions of the consistent matrix equation $AX = B$ is characterized by $X = V_{a1} S_{a1}^{-1} U_{a1}^t B + V_{a2} T$, where T is an arbitrary matrix.

The first term is nothing else than $A^\dagger B$, where $A^\dagger$ is the Moore-Penrose pseudoinverse of A. It is also the unique minimum Frobenius norm solution. Recall that $A^\dagger$ is the Moore-Penrose inverse of A if it is the unique solution $T = A^\dagger$ of

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

$$TAT = T, \tag{2}$$

$$(AT)^t = AT, \tag{3}$$

$$(TA)^t = TA. \tag{4}$$

This pair of matrices satisfies the condition required by the lemma in [8] that AB^t is diagonal. With the notation of [8], we have that $i = 1$, $j = 1$, $k = 2$, $r = 5$. While the proof of the lemma states that $r - i - j = k$, this is not true in general, because for our example $k < r - i - j$. Hence, the proof of Lemma 1 in [8] is not correct.

In Section 3, we shall also use the notions of a 1-2-3-inverse of the matrix A, which is any matrix T satisfying (1), (2), (3).

LEMMA 2. *For any pair of $m \times n$ matrices A and B, the nonzero eigenvalues of AB^t and B^tA are the same.*

An immediate consequence of Lemma 2 is the following:

COROLLARY 1. *Denote by $\lambda(\cdot)$ the nonzero eigenvalue spectrum of a matrix. Then $\lambda(AB^tBA^t) = \lambda(BA^tAB^t) = \lambda(A^tAB^tB) = \lambda(B^tBA^tA)$.*

Another result we shall need concerns the PSVD of two matrices in the special case that their row spaces are orthogonal, i.e. $AB^t = 0$.

LEMMA 3. *Let A $(m \times n)$ and B $(p \times n)$ be such that $AB^t = 0$. Assume that A and B have OSVDs:*

$$A = \begin{pmatrix} U_{a1} & U_{a2} \end{pmatrix} \begin{pmatrix} S_{a1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_{a1}^t \\ V_{a2}^t \end{pmatrix}, \tag{5}$$

$$B = \begin{pmatrix} U_{b1} & U_{b2} \end{pmatrix} \begin{pmatrix} S_{b1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_{b1}^t \\ V_{b2}^t \end{pmatrix}, \tag{6}$$

where S_{a1} is $r_a \times r_a$ ($r_a = \operatorname{rank} A$) and S_{b1} is $r_b \times r_b$ ($r_b = \operatorname{rank} B$). Assume that the common null space is generated by the columns of the orthonormal matrix V_{ab2}:

$$\begin{pmatrix} A \\ B \end{pmatrix} V_{ab2} = 0.$$

Then a PSVD of A, B is given by

$$A = \begin{pmatrix} U_{a1} & U_{a2} \end{pmatrix} \overset{\begin{matrix} r_a & \quad r_b & \quad n - r_a - r_b \end{matrix}}{\begin{pmatrix} I_{r_a} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}} \begin{pmatrix} S_{a1} V_{a1}^t \\ S_{b1}^{-1} V_{b1}^t \\ V_{ab2}^t \end{pmatrix}$$

$$B = \begin{pmatrix} U_{b1} & U_{b2} \end{pmatrix} \overset{\begin{matrix} r_a & \quad r_b & \quad n - r_a - r_b \end{matrix}}{\begin{pmatrix} 0 & I_{r_b} & 0 \\ 0 & 0 & 0 \end{pmatrix}} \begin{pmatrix} S_{a1}^{-1} V_{a1}^t \\ S_{b1} V_{b1}^t \\ V_{ab2}^t \end{pmatrix}$$

We have written "a" PSVD instead of "the" PSVD because of the nonuniqueness of V_{ab2} (which for instance can be postmultiplied by any orthonormal matrix) and possibly of $U_{a1}, U_{a2}, V_{a1}, V_{a2}, U_{b1}, U_{b2}, V_{b1}, V_{b2}$ from the (non)uniqueness properties of the OSVD. A detailed analysis of the nonuniqueness properties of the PSVD in general is the subject of Section 3.

Proof. Observe that, because of the orthogonality of the row spaces of A and B, it follows that

$$\operatorname{rank}\begin{pmatrix} A \\ B \end{pmatrix} = r_a + r_b.$$

Hence, the dimension of the common null space is $n - r_a - r_b$. It is straightforward to find that V_{a2} and V_{b2} can be chosen as

$$V_{a2}^t = \begin{pmatrix} V_{b1}^t \\ V_{ab2}^t \end{pmatrix},$$

$$V_{b2}^t = \begin{pmatrix} V_{a1}^t \\ V_{ab2}^t \end{pmatrix}.$$

The theorem then follows. The matrices S_{a1}^{-1} and S_{b1}^{-1} are inserted because the right hand factors of A and B must be related to each other as X^{-1} and X^t (see Theorem 1). ■

The central idea of the proof of Theorem 1 is to exploit the close connection between the OSVD of AB^t and the eigenvalue decompositions of B^tBA^tA and A^tAB^tB, which is the subject of the following lemma:

LEMMA 4. *Let the OSVD of AB^t be given as*

$$AB^t = UD_1V^t \tag{7}$$

$$= \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^t \\ V_2^t \end{pmatrix}, \tag{8}$$

where S_1 ($r_1 \times r_1$ with $r_1 = \operatorname{rank}(AB^t)$) contains the nonzero singular values *of AB^t. Consider the eigenvalue problem*

$$(B^tBA^tA)Y = YD_2. \tag{9}$$

Consider also the OSVD of A as in (5). *Then all possible matrices of eigenvectors Y can be written as*

$$Y = \begin{pmatrix} Y_1 & Y_2 & Y_3 \end{pmatrix} = \begin{pmatrix} A^\dagger & V_{a2} \end{pmatrix} \begin{pmatrix} U_1 & U_3 & U_4 \\ T_1 & T_3 & T_4 \end{pmatrix},$$

where $T_1 = V_{a2}^t B^t B A^t U_1 S_1^{-2}$, U_3 is any matrix such that $R(A) = R(U_1) \oplus R(U_3)$, U_4 is any matrix such that $N(A^t) = R(U_4)$, T_3 and T_4 are arbitrary matrices that can be chosen to ensure that $\operatorname{rank}(Y) = n$, *and*

$$D_2 = \begin{pmatrix} S_1^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Proof. From Corollary 1 it follows that the nonzero eigenvalues of AB^tBA^t and B^tBA^tA are the same. We shall show that there exist $r_1 = \operatorname{rank}(AB^t)$ eigenvectors corresponding to S_1^2. These will form the $n \times r_1$ matrix Y_1. Then we shall show that it is possible to choose an $n \times (r_a - r_1)$ matrix Y_2 and an $n \times (n - r_a)$ matrix Y_3, both containing eigenvectors corresponding to zero eigenvalues, such that the $n \times n$ matrix $Y = (Y_1\ Y_2\ Y_3)$ is nonsingular.

Proof for Y_1: From the fact that $r_1 = \operatorname{rank}(AB^t) \leqslant r_a = \operatorname{rank} A$, it follows that $R(U_1) \subset R(A)$, so that $AA^\dagger U_1 = U_1$. The matrix Y_1 will contain eigenvectors corresponding to S_1^2 if

$$(B^tBA^tA)Y_1 = Y_1 S_1^2. \tag{10}$$

Premultiply this expression with A to find $(AB^tBA^t)AY_1 = AY_1S_1^2$. But from the OSVD (8) of AB^t, it follows then that we can put $AY_1 = U_1$, and using Lemma 1 it follows that $Y_1 = A^\dagger U_1 + V_{a2}T_1$. The matrix T_1 is however not arbitrary, because Y_1 has to satisfy (10). Substituting the last equation into (10) results in $B^tBA^tA(A^\dagger U_1 + V_{a2}T_1) = (A^\dagger U_1 + V_{a2}T_1)S_1^2$. Premultiplying this with V_{a2}^t results in $T_1 = V_{a2}^t B^t B A^t U_1 S_1^{-2}$. Hence we find that $Y_1 = V_{a1}S_{a1}^{-1}U_{a1}^t U_1 + V_{a2}V_{a2}^t B^t B A^t U_1 S_1^{-2}$. Let us now verify that Y_1 satisfies (10). First observe that from the OSVD of AB^t (8) and the OSVD of A (5) it follows that $V_{a1}^t B^t B A^t U_1 = S_{a1}^{-1} U_{a1}^t U_1 S_1^2$. Together with the expression for T_1, this implies the following identity:

$$\begin{pmatrix} V_{a1}^t \\ V_{a2}^t \end{pmatrix} B^t B A^t U_1 = \begin{pmatrix} V_{a1}^t \\ V_{a2}^t \end{pmatrix} (V_{a1} S_{a1}^{-1} U_{a1}^t U_1 + V_{a2} T_1) S_1^2.$$

But because $(V_{a1} \;\; V_{a2})$ is nonsingular, it follows from that $B^t B A^t U_1 = (A^\dagger U_1 + V_{a2} T_1) S_1^2 = Y_1 S_1^2$. It can be verified that $U_1 = A Y_1$. Substitute this to find $B^t B A^t A Y_1 = Y_1 S_1^2$, which proves that Y_1 contains the eigenvectors corresponding to the eigenvalues that are diagonal elements of S_1^2.

Proof for Y_2: Observe that $R(A) = R(U_1) \oplus R(U_3)$ implies that $U_1^t U_3 = 0$. Furthermore, because $R(U_3) \subset R(A)$, it follows that $A A^\dagger U_3 = U_3$. Let Y_2 be given as $Y_2 = A^\dagger U_3 + V_{a2} T_3$, where T_3 is an arbitrary matrix. Then

$$B^t B A^t A Y_2 = B^t B A^t A \left(A^\dagger U_3 + V_{a2} T_3 \right) = B^t B A^t U_3 = B^t V_1 S U_1^t U_3 = 0.$$

Hence, the column vectors of Y_2 belong to the null space of $B^t B A^t A$, and $\operatorname{rank} Y_2 = \operatorname{rank} U_3 = r_a - r_1$.

Proof for Y_3: Assume that $Y_3 = A^\dagger U_4 + V_{a2} T_4 = V_{a2} T_4$. It follows that $B^t B A^t A Y_3 = B^t B A^t A V_{a2} T_4 = 0$. This implies that the column vectors of Y_3 belong to the null space of $B^t B A^t A$, and obviously $\operatorname{rank} Y_3 = \operatorname{rank} V_{a2} = n - r_a$ if T_4 is nonsingular.

Finally, we have to verify that with fixed U_1, U_3, U_4, and T_1, we can always chose T_3 and T_4 to make the matrix

$$Y = \begin{pmatrix} Y_1 & Y_2 & Y_3 \end{pmatrix} = \begin{pmatrix} A^\dagger & V_{a2} \end{pmatrix} \begin{pmatrix} U_1 & U_3 & U_4 \\ T_1 & T_3 & T_4 \end{pmatrix} \qquad (11)$$

of full rank. Rewrite (11), using the OSVD of A (5), as

$$Y = \begin{pmatrix} V_{a1} S_{a1}^{-1} & V_{a2} \end{pmatrix} \begin{pmatrix} U_{a1}^t U_1 & U_{a1}^t U_3 & U_{a1}^t U_4 \\ T_1 & T_3 & T_4 \end{pmatrix}. \qquad (12)$$

The matrix Y is now written as a product of two factors: The first factor $(V_{a1} S_{a1}^{-1} \;\; V_{a2})$ is square nonsingular. Obviously, the second factor can always be made nonsingular by an appropriate choice of T_3 and T_4. ∎

An immediate consequence of Lemma 4 is:

COROLLARY 2. *Consider the eigenvalue problem for B^tBA^tA as in* (9): $(B^tBA^tA)Y = YD_2$, *where Y is chosen as described in Lemma* 4. *Then $X = Y^{-t}$ contains the eigenvectors of A^tAB^tB*: $(A^tAB^tB)X = XD_2$.

Proof. The proof follows from the nonsingularity of Y and from transposing (9). ■

Obviously, the column vectors of X are the left eigenvectors of B^tBA^tA. We are now ready to prove Theorem 1:

Proof of Theorem 1. The proof consists of three steps:

Step 1. First we'll show that A and B can be decomposed as

$$A = U\begin{pmatrix} A'_{11} & 0 \\ 0 & A'_{22} \end{pmatrix} X^t,$$

$$B = V\begin{pmatrix} B'_{11} & 0 \\ 0 & B'_{22} \end{pmatrix} Y^t$$

with $X^tY = I$.

Step 2. Then it will be shown that A'_{11} and B'_{11} are diagonal.

Step 3. It will be shown that $A'_{22}B'^t_{22} = 0$. This orthogonality of the row spaces of A'_{22} and B'_{22} allows us to apply Lemma 3 to the pair (A'_{22}, B'_{22}).

Combining steps 1, 2, 3 will then prove the theorem.

Step 1: Combining the OSVD (8) of AB^t and the eigenvalue decomposition (9) results in $B^tBA^tAY = B^t(BA^t)AY = B^t(VD_1^tU^t)AY = YD_2$. Premultiplying with A results in $AB^t(VD_1^tU^t)AY = AYD_2$ whence $(UD_1V^t)(VD_1^tU^t)AY = AYD_2$, whence $(D_1D_1^t)(U^tAY) = (U^tAY)D_2$, or with the block structure of D_1 and D_2,

$$\begin{pmatrix} S_1^2 & 0 \\ 0 & 0 \end{pmatrix}(U^tAY) = (U^tAY)\begin{pmatrix} S_1^2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Now call $A' = U^tAY$, and partition A' according to the block structure of D_1

and D_2 as

$$A' = \begin{array}{c} \\ r_1 \\ m-r_1 \end{array} \begin{array}{c} \begin{array}{cc} r_1 & n-r_1 \end{array} \\ \begin{pmatrix} A'_{11} & A'_{12} \\ A'_{21} & A'_{22} \end{pmatrix} \end{array}.$$

Then obviously

$$\begin{pmatrix} S_1^2 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} A'_{11} & A'_{12} \\ A'_{21} & A'_{22} \end{pmatrix} = \begin{pmatrix} A'_{11} & A'_{12} \\ A'_{21} & A'_{22} \end{pmatrix}\begin{pmatrix} S_1^2 & 0 \\ 0 & 0 \end{pmatrix},$$

which implies that $S_1^2 A'_{11} = A'_{11} S_1^2$, $A'_{12} = 0$, and $A'_{21} = 0$. Recall from Lemma 4 that Y is nonsingular. Hence the matrix $A = UA'Y^{-1}$ can be written as

$$A = U\begin{pmatrix} A'_{11} & 0 \\ 0 & A'_{22} \end{pmatrix} Y^{-1}. \tag{13}$$

Because U and Y are nonsingular matrices, we have that

$$\operatorname{rank} A'_{11} + \operatorname{rank} A'_{22} = \operatorname{rank} A. \tag{14}$$

Using Corollary 2 and applying a similar derivation to the matrix A^tAB^tB results in a decomposition of the matrix B as

$$B = V\begin{pmatrix} B'_{11} & 0 \\ 0 & B'_{22} \end{pmatrix} Y^t, \tag{15}$$

where $B' = V^tBY^{-t}$ and B'_{11} is the upper $r_1 \times r_1$ block of B'. Moreover,

$$\operatorname{rank} B'_{11} + \operatorname{rank} B'_{22} = \operatorname{rank} B. \tag{16}$$

Step 2. Carrying out the multiplication AB^t with the two factorizations (13) and (15) results in

$$AB^t = U\begin{pmatrix} A'_{11}B'^t_{11} & 0 \\ 0 & A'_{22}B'^t_{22} \end{pmatrix} V^t, \tag{17}$$

but from the uniqueness properties of the OSVD (8), it follows immediately that we can put $A'_{11}B'^{t}_{11}=S_1$. Hence, we have $\operatorname{rank} A'_{11}=\operatorname{rank} B'_{11}=r_1$, so that $B'^{t}_{11}=(A'_{11})^{-1}S_1$. When we require that $A'_{11}=B'_{11}$, one can always take $A'_{11}=B'_{11}=S_1^{1/2}$. In the case that the elements of S_1 are distinct, this solution is unique. If some of the elements coincide, the solution is unique up to block diagonal orthonormal matrices, which can however by incorporated into the orthonormal matrices U and V in the factorization of AB^t (17).

Step 3: It follows from the (non)uniqueness properties of the OSVD in (17) and (8) that $A'_{22}B'^{t}_{22}=0$. Moreover, from (14) and (16), it follows that

$$\operatorname{rank} A'_{22}=\operatorname{rank} A-r_1=r_a-r_1,$$

$$\operatorname{rank} B'_{22}=\operatorname{rank} B-r_1=r_b-r_1.$$

The proof is now straightforward by applying Lemma 3 to the pair A'_{22}, B'_{22} and inserting the corresponding factorizations for A'_{22} and B'_{22} into (13) and (15). ■

2.2. *A Variational Characterization*

Note that, from Theorem 1, Lemma 4, and Corollary 2, it follows that there are four eigenvalue decompositions that can be related to the PSVD:

$$(A^tAB^tB)X=X(S_A^tS_AS_B^tS_B),$$

$$(B^tBA^tA)Y=Y(S_B^tS_BS_A^tS_A),$$

$$(AB^tBA^t)U_A=U_A(S_AS_B^tS_BS_A^t),$$

$$(BA^tAB^t)U_B=U_B(S_BS_A^tS_AS_B^t).$$

The last two of them are OSVDs. Let us now derive a variational interpretation of the PSVD. Consider the following optimization problem:

Maximize over all vectors x and y

$$(y^tA^tAy)(x^tB^tBx) \tag{18}$$

subject to

$$x^ty=1. \tag{19}$$

Assume that the maximum is achieved for some vectors x_1 and y_1. Then consider the following set of problems:

Find the vectors x^k, y^k, $k = 2,3,\ldots,$ that maximize

$$\left[(y^k)^t A^t A y^k\right]\left[(x^k)^t B^t B x^k\right] \tag{20}$$

subject to

$$(x^k)^t y^k = 1, \tag{21}$$

$$(x^k)^t y^j = 0, \qquad j = 1,\ldots,k-1, \tag{22}$$

$$(x^i)^t y^k = 0, \qquad i = 1,\ldots,k-1. \tag{23}$$

It can be shown that the PSVD delivers the solution: The maximum of (18) is achieved for the first column vectors of X and Y and is equal to the largest product singular value. The other column vectors of X and Y provide the solutions to (20)–(23).

3. ON THE STRUCTURE OF THE CONTRAGREDIENT TRANSFORMATION

In this section, we shall investigate in detail the structure of the matrix X, including its (non)uniqueness properties. As a matter of fact, already in Lemma 4 we have provided a parametrization of possible matrices $X = Y^{-t}$ in terms of matrices U_3, T_3, U_4, and T_4. In this section, however, we shall make a more detailed analysis of the nonuniqueness.

First, in Section 3.1, we summarize some known results on contragredient and balancing transformations of pairs of symmetric matrices, one of which is positive definite and the other nonnegative or positive definite. Then, in Section 3.2, it is shown how certain submatrices of the contragredient transformation matrix X are solutions of a set of nonlinear matrix equations. A solution of these is provided in Section 3.3 (a constructive derivation can be found in the appendix). These "basic" solutions, which themselves contain a certain degree of nonuniqueness, are then used to parametrize all possible PSVDs of a pair of matrices, which is the subject of Section 3.4.

In summary, the main result of this section is a complete characterization and description of the nonuniqueness properties of the PSVD, and in particular, of a contragredient transformation for two nonnegative definite matrices.

3.1. *Contragredient and Balancing Transformations*

In order to introduce the notion of a contragredient transformation, observe that it follows from Theorem 1 that

$$A^tA = X(S_A^tS_A)X^t,$$

$$B^tB = X^{-t}(S_B^tS_B)X^{-1},$$

so that

$$X^{-1}A^tAX^{-t} = S_A^tS_A,$$

$$X^tB^tBX = S_B^tS_B.$$

Hence X^{-1} diagonalizes the matrix A^tA, while X^t diagonalizes the matrix B^tB. A double congruence transformation of this kind for a pair of matrices is called *contragredient* [16].

Definition 1 (Contragredient transformation). The nonsingular $n\times n$ matrix T is a *contragredient transformation* for a pair of matrices F,G if both $T^{-1}FT^{-t}$ and T^tGT are real diagonal.

If both diagonal matrices are equal, we have:

Definition 2 (Balancing contragredient transformation). A contragredient transformation T is called *balancing* if $T^{-1}FT^{-t}=T^tGT$ is real diagonal.

Applications of (balancing) contragredient transformations can be found in system and control theory (open loop balancing of stable plants [8, 16, 17] and unstable systems [15]; closed loop balancing [14]; model reduction [11]; and H_∞ controller design [10]).

An immediate consequence of Definition 2 is of course that balancedness can only occur if F and G have the same inertia, because T is a congruence transformation on F and G, which preserves inertia. Obviously, a necessary condition for the existence of a contragredient transformation for the pair

F,G is that the product FG must be similar to a real diagonal matrix. An example of a pair F,G for which no contragredient transformation exists is

$$F=\begin{pmatrix}3 & 1\\ 1 & -1\end{pmatrix},\qquad G=\begin{pmatrix}2 & -2\\ -2 & 0\end{pmatrix}.$$

The eigenvalues of FG are $1\pm j\sqrt{15}$; hence FG is not similar to a real diagonal matrix. In case F and G are nonnegative definite (NND) and/or positive definite (PD), a contragredient transformation always exists. This is shown in Lemma 7, where F and G are both PD, and in Lemma 8, where F is PD and G is NND. The case where both F and G are NND is analysed in detail in Sections 3.2–3.4. These conditions of positive and nonnegative definiteness are sufficient but not necessary. As an example, consider

$$F=\begin{pmatrix}2 & 2\\ 2 & 1\end{pmatrix},\qquad G=\begin{pmatrix}-1 & -1\\ -1 & 1\end{pmatrix}.$$

Both F and G are indefinite. It is easy to check that

$$T=\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}$$

is a contragredient transformation.

LEMMA 5 (Existence of a contragredient transformation for positive definite matrices). *Suppose $F=F^t$ and $G=G^t$ are both positive definite. Let F and G have Cholesky factorization $F=L_FL_F^t$ and $G=L_GL_G^t$. Let $L_G^tL_F$ have singular value decomposition $L_G^tL_F=U\Sigma V^t$. Then $T=L_FV\Sigma^{-1/2}$ is a contragredient balancing transformation. Also $T^{-1}=\Sigma^{-1/2}U^tL_G^t$.*

Proof. [16, Theorem 1]. ∎

The next theorem addresses the case where one of F and G is nonnegative definite, say G. In this case, the contragredient transformation cannot be balancing because F and G do not have the same inertia.

LEMMA 6 (Existence of a contragredient transformation for positive definite F, nonnegative definite G). *Let $F=F^t$ be positive definite and $G=G^t$ be nonnegative definite. Let F have Cholesky factorization $F=L_FL_F^t$, and $G=L_GL_G^t$ be a Cholesky-like factorization where L_G is $n\times r_G=$*

rank(G). *Let the OSVD of* $L_F^t L_G$ *be* $L_F^t L_G = U\Sigma V^t$. *Then* $T = L_F U$ *is a contragredient transformation.*

Proof. [16]. ∎

Observe that a contragredient transformation can only be unique up to a diagonal matrix, because if T is contragredient, then TD, where D is nonsingular diagonal, will also be contragredient. In case F and G are positive definite, a *balancing* contragredient transformation is essentially unique if the eigenvalues of FG are distinct. In case two or more eigenvalues of FG are repeated, their corresponding eigenvectors can be rotated arbitrarily in the corresponding eigenspace. In case F is positive definite and G nonnegative definite, similar statements apply. If however, both F and G are nonnegative definite, nonuniqueness for balancing contragredient transformations arises even in the distinct eigenvalue case, as is evident from the following example, borrowed from [16]:

$$F = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad G = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then

$$FG = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

has distinct eigenvalues at 1 and 0. But the transformation

$$T = \begin{pmatrix} \beta & 0 \\ \beta & \gamma \end{pmatrix}$$

is contragredient for any nonzero β and γ, and balancing if $\beta = 1$ and γ nonzero. From Theorem 1, it can be seen that the PSVD provides a contragredient transformation for the matrix pair A^tA and B^tB, and the conditions for this transformation to be balancing are obvious from the structure of the matrices S_A and S_B in Theorem 1.

The rest of this paper is devoted to a detailed analysis of the case of nonnegative definite F and G, in case $F = A^tA$ and $G = B^tB$. When for instance both the matrices A and B have more columns than rows, both A^tA and B^tB are nonnegative definite. In particular, we shall analyse in detail all possible causes of the nonuniqueness of the contragredient transformation X that occurs in the PSVD of Theorem 1. Obviously, the results will also apply to the case where F and G are nonnegative definite, but not given explicitly as $F = A^tA$ and $G = B^tB$ for some A and B. A suitable A and B can always be obtained from (for instance) a Cholesky-like factorization as in Lemma 8. The results of this section can then be applied to the Cholesky factors.

3.2. *Expressing the PSVD via OSVDs*

First, we shall show how to deflate a common null space of the matrices A and B. This will allow us to assume without loss of generality that A and B do not have a common null space. Then we shall relate the PSVD of the matrix pair A, B to several OSVDs in Sections 3.2.2 and 3.2.3. This leads to a set of nonlinear equations, which will be solved in Section 3.3.

3.2.1. *Deflating the Common Null Space.* Assume that the OSVD of the concatenation of A and B is given by

$$(A \quad B) = (U_{ab1} \quad U_{ab2})\begin{pmatrix} S_{ab1} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} V_{ab1}^t \\ V_{ab2}^t \end{pmatrix},$$

where S_{ab1} is $r_{ab} \times r_{ab}$ diagonal and

$$r_{ab} = \operatorname{rank}\begin{pmatrix} A \\ B \end{pmatrix}.$$

The common null space of A and B is then generated by the column vectors of the $n \times (n - r_{ab})$ matrix V_{ab2}. Define the matrices A_0 $(m \times r)$ and B_0 $(p \times r)$ as

$$\begin{pmatrix} A \\ B \end{pmatrix} V_{ab} = \begin{pmatrix} A_0 & 0 \\ B_0 & 0 \end{pmatrix}$$

with $V_{ab} = (V_{ab1} \; V_{ab2})$. Obviously, A_0 and B_0 don't have a common null space. Now assume that a PSVD of the pair A_0, B_0 is given as

$$A_0 = U_{A_0} S_{A_0} X_0^t,$$

$$B_0 = U_{B_0} S_{B_0} X_0^{-1},$$

where S_{A_0} is $m\times r_{ab}$, S_{B_0} is $p\times r_{ab}$, and X_0 is $r_{ab}\times r_{ab}$. It follows immediately that a PSVD of the pair A, B is given by

$$A = U_{A_0}\begin{pmatrix} S_{A_0} & 0_{m\times(n-r_{ab})}\end{pmatrix}\begin{pmatrix} X_0^t & 0 \\ 0 & W_1^t \end{pmatrix}V_{ab}^t, \tag{24}$$

$$B = U_{B_0}\begin{pmatrix} S_{B_0} & 0_{p\times(n-r_{ab})}\end{pmatrix}\begin{pmatrix} X_0^{-1} & 0 \\ 0 & W_1^{-1} \end{pmatrix}V_{ab}^t, \tag{25}$$

where W_1 is an arbitrary but nonsingular $(n-r_{ab})\times(n-r_{ab})$ matrix. This matrix represents the first source of possible nonuniqueness of the contragredient transformation.

We assume from now on throughout the rest of Sections 3.2 and 3.3, without loss of generality, that the matrices A and B do not have a common null space and that

$$r_{ab} = \operatorname{rank}\begin{pmatrix} A \\ B \end{pmatrix} = n.$$

Only in Sections 3.4 and 4 shall we again consider the possibility of A and B having a common null space.

3.2.2. *The OSVD of the Product.* Let the OSVDs of A $(m\times r_{ab})$ and B $(p\times r_{ab})$ be

$$A = \begin{pmatrix} U_{a1} & U_{a2}\end{pmatrix}\begin{pmatrix} S_{a1} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} V_{a1}^t \\ V_{a2}^t \end{pmatrix}, \tag{26}$$

$$B = \begin{pmatrix} U_{b1} & U_{b2}\end{pmatrix}\begin{pmatrix} S_{b1} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} V_{b1}^t \\ V_{b2}^t \end{pmatrix}, \tag{27}$$

where $r_a = \operatorname{rank} A$, $r_b = \operatorname{rank} B$, and S_{a1} is $r_a\times r_a$ and S_{b1} is $r_b\times r_b$ diagonal, the matrices of left and right singular vectors being partitioned accordingly. Then the product can be written as

$$AB^t = \begin{pmatrix} U_{a1} & U_{a2}\end{pmatrix}\begin{pmatrix} S_{a1}V_{a1}^tV_{b1}S_{b1} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} U_{b1}^t \\ U_{b2}^t \end{pmatrix}.$$

Consider the OSVD of the $r_a \times r_b$ matrix

$$S_{a1}V_{a1}^t V_{b1}S_{b1} = (P_1 \quad P_2)\begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} Q_1^t \\ Q_2^t \end{pmatrix} \tag{28}$$

with $r_1 = \mathrm{rank}(AB^t)$ and S_1 $(r_1 \times r_1)$ diagonal with the nonzero singular values of AB^t. Again, the matrices of left and right singular vectors are partitioned in an obvious way; e.g., P_2 is an $r_a \times (r_a - r_1)$ matrix. The OSVD of AB^t can then be written as

$$AB^t = (U_{a1}P_1 \quad U_{a1}P_2 \quad U_{a2})\begin{pmatrix} S_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} Q_1^t U_{b1}^t \\ Q_2^t U_{b2}^t \\ U_{b2}^t \end{pmatrix}. \tag{29}$$

Obviously, $r_1 \leqslant \min(r_a, r_b)$. Observe that if $S_{a1} = I_{r_a}$ and $S_{b1} = I_{r_b}$, the OSVD of $V_{a1}^t V_{b1}$ is nothing else than performing a *canonical correlation analysis* between the row spaces of the matrices A and B [2]. In other words, the OSVD of $S_{a1}V_{a1}^t V_{b1}S_{b1}$ could be considered as a *weighted canonical correlation analysis*.

Let A $(m \times r_{ab})$ and B $(p \times r_{ab})$ be matrices with no common null space. Referring to (29) and the PSVD theorem of Section 2, introduce two nonsingular $r_{ab} \times r_{ab}$ matrices X and Y and rewrite A and B as

$$A = (U_{a1}P_1 \quad U_{a1}P_2 \quad U_{a2})\begin{pmatrix} S_1^{1/2} & 0 & 0 \\ 0 & I_{r_a - r_1} & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} X_1^t \\ X_2^t \\ X_3^t \end{pmatrix}, \tag{30}$$

$$B = (U_{b1}Q_1 \quad U_{b1}Q_2 \quad U_{b2})\begin{pmatrix} S_1^{1/2} & 0 & 0 \\ 0 & 0 & I_{r_b - r_1} \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} Y_1^t \\ Y_2^t \\ Y_3^t \end{pmatrix}, \tag{31}$$

where X_1 is $r_{ab} \times r_1$, X_2 is $r_{ab} \times (r_a - r_1)$, X_3 is $r_{ab} \times (r_{ab} - r_a)$, Y_1 is $r_{ab} \times r_1$, Y_2 $r_{ab} \times (r_b - r_1)$, and Y_3 is $r_{ab} \times (r_{ab} - r_b)$.

Then obviously X will be a contragredient transformation if

$$X^tY = \begin{pmatrix} X_1^t \\ X_2^t \\ X_3^t \end{pmatrix} \begin{pmatrix} Y_1 & Y_2 & Y_3 \end{pmatrix} = I_r. \tag{32}$$

From the expressions (26) and (30) for A and (27) and (31) for B it is obvious that

$$X_1^t = = S_1^{-1/2} P_1^t S_{a1} V_{a1}^t, \tag{33}$$

$$X_2^t = P_2^t S_{a1} V_{a1}^t \tag{34}$$

and

$$Y_1^t = S_1^{-1/2} Q_1^t S_{b1} V_{b1}^t, \tag{35}$$

$$Y_3^t = Q_2^t S_{b1} V_{b1}^t. \tag{36}$$

Obviously, rank $X_1 = r_1 =$ rank Y_1, rank $X_2 = r_a - r_1$, and rank $Y_3 = r_b - r_1$. Moreover, it follows immediately that

$$X_1^t Y_1 = I_{r_1}, \tag{37}$$

$$X_2^t Y_1 = 0, \tag{38}$$

$$X_1^t Y_3 = 0, \tag{39}$$

$$X_2^t Y_3 = 0. \tag{40}$$

Because P_2 and Q_2, containing singular vectors corresponding to nondistinct zero singular values, are not unique, X_2 and Y_3 are not unique. They are only determined up to orthonormal matrices W_2 and W_3 as

$$X_2 = V_{a1} S_{a1} P_2 W_2, \tag{41}$$

$$Y_3 = V_{b1} S_{b1} Q_2 W_3, \tag{42}$$

with $W_2^t W_2 = I_{r_a - r_1} = W_2 W_2^t$ and $W_3^t W_3 = I_{r_b - r_1} = W_3 W_3^t$. The fact that W_2 and W_3 must be orthonormal also follows from (30) and (31): If X_2^t (Y_3^t) is premultiplied there by W_2^t (W_3^t), then $U_{a1} P_2$ ($U_{b2} Q_2$) must be postmultiplied by W_2^{-t} (W_3^{-t}) but must remain orthonormal. In what follows, we shall choose $W_2 = I_{r_a - r_1}$ and $W_3 = I_{r_b - r_1}$, until Section 3.4, where we discuss nonuniqueness issues in detail.

3.2.3. Refinement of the Block Structure. Let's now have a closer look at the dimensions of the blocks of the matrix product $X^t Y$:

$$X^t Y = \begin{array}{c} \\ r_1 \\ r_a - r_1 \\ r_{ab} - r_a \end{array} \begin{array}{c} \begin{array}{ccc} r_1 & r_{ab} - r_b & r_b - r_1 \end{array} \\ \begin{pmatrix} X_1^t Y_1 & X_1^t Y_2 & X_1^t Y_3 \\ X_2^t Y_1 & X_2^t Y_2 & X_2^t Y_3 \\ X_3^t Y_1 & X_3^t Y_2 & X_3^t Y_3 \end{pmatrix} \end{array}.$$

The requirement that this product must be equal to the identity matrix imposes the following structure. Since we know already that $X_2^t Y_3 = 0$, it follows that $r_{ab} - r_a \geq r_b - r_1$, or

$$r_{ab} \geq r_a + r_b - r_1. \tag{43}$$

This follows also from $X_2^t Y_1 = 0$. The lower $(r_b - r_1) \times (r_b - r_1)$ matrix of $X_3^t Y_3$ is the identity matrix $I_{r_b - r_1}$. The left $(r_a - r_1)$ part of $X_2^t Y_2$ equals $I_{r_a - r_1}$. The upper right corner of $X_3^t Y_2$ equals $I_{r_{ab} - r_a - r_b + r_1}$.

According to these requirements, the block structure is refined as

$$X^t Y = \begin{pmatrix} X_1^t \\ X_2^t \\ X_{31}^t \\ X_{32}^t \end{pmatrix} \begin{pmatrix} Y_1 & Y_{21} & Y_{22} & Y_3 \end{pmatrix} = I_{r_{ab}}. \tag{44}$$

Here, X_{31} is $r_{ab} \times (r_{ab} - r_a - r_b + r_1)$, X_{32} $r_{ab} \times (r_b - r_1)$, Y_{21} $r_{ab} \times (r_a - r_1)$, and Y_{22} $r_{ab} \times (r_{ab} - r_a - r_b + r_1)$.

This leads to the following refinement of the structure of the matrices S_A and S_B in (30) and (31) (recall that, for the time being, there is no common null space):

$$D_a = \begin{array}{c} r_1 \\ r_a - r_1 \\ m - r_a \end{array} \overset{\begin{array}{cccc} r_1 & r_a - r_1 & r_{ab} - r_a - r_b + r_1 & r_b - r_1 \end{array}}{\begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}}, \tag{45}$$

$$D_B = \begin{array}{c} r_1 \\ r_b - r_1 \\ p - r_b \end{array} \overset{\begin{array}{cccc} r_1 & r_a - r_1 & r_{ab} - r_a - r_b + r_1 & r_b - r_1 \end{array}}{\begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 \\ 0 & 0 & 0 & I \\ 0 & 0 & 0 & 0 \end{pmatrix}}. \tag{46}$$

It follows from the refined block structure (44) that the matrices $X_{31}, X_{32}, Y_{21}, Y_{22}$ will be solutions to the following set of nonlinear matrix equations:

$$\begin{pmatrix} Y_1^t \\ Y_3^t \end{pmatrix} \begin{pmatrix} X_{31} & X_{32} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & I_{r_b - r_1} \end{pmatrix}, \tag{47}$$

$$\begin{pmatrix} X_1^t \\ X_2^t \end{pmatrix} \begin{pmatrix} Y_{21} & Y_{22} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ I_{r_a - r_1} & 0 \end{pmatrix}, \tag{48}$$

subject to the orthogonality constraints

$$\begin{pmatrix} X_{31}^t \\ X_{32}^t \end{pmatrix} \begin{pmatrix} Y_{21} & Y_{22} \end{pmatrix} = \begin{pmatrix} 0 & I_{r - r_a - r_b + r_1} \\ 0 & 0 \end{pmatrix}, \tag{49}$$

where the matrices X_1, X_2, Y_1, Y_3 are given by (33), (34), (35), (36).

A solution for the set of equations (47)–(49) will be obtained in the next subsection.

3.3. *A Solution to the Set of Nonlinear Matrix Equations*

In this subsection, we present a solution of the set of nonlinear matrix equations (47)–(49). For a constructive derivation, the interested reader is referred to the appendix. In order to simplify our expressions below, we shall first introduce some new notation.

Recall the expressions (33)–(36) for X_1, X_2, Y_1, Y_3. Define the new matrices

$$\overline{X}_1 = V_{a1} S_{a1}^{-1} P_1 S_1^{1/2}, \tag{50}$$

$$\overline{X}_2 = V_{a1} S_{a1}^{-1} P_2, \tag{51}$$

$$\overline{Y}_1 = V_{b1} S_{b1}^{-1} Q_1 S_1^{1/2}, \tag{52}$$

$$\overline{Y}_3 = V_{b1} S_{b1}^{-1} Q_2. \tag{53}$$

Then we have the following properties:

LEMMA 7 (Properties of $\overline{X}_1, \overline{X}_2, \overline{Y}_1, \overline{Y}_3$).

(a) *The matrices* $\overline{X}_1, \overline{X}_2, \overline{Y}_1, \overline{Y}_3$ *are* 1-2-3-*inverses of the matrices* $X_1^t, X_2^t, Y_1^t, Y_3^t$. *They are all of full column rank.*

(b) *They satisfy the following properties*:

$$X_1^t \overline{X}_1 = I_{r_1}, \tag{54}$$

$$X_2^t \overline{X}_2 = I_{r_a - r_1}, \tag{55}$$

$$Y_1^t \overline{Y}_1 = I_{r_1}, \tag{56}$$

$$Y_3^t \overline{Y}_3 = I_{r_b - r_1}. \tag{57}$$

(c) *There are also the orthogonality relations*

$$X_1^t \overline{X}_2 = 0, \tag{58}$$

$$X_2^t \overline{X}_1 = 0, \tag{59}$$

$$Y_1^t \overline{Y}_3 = 0, \tag{60}$$

$$Y_3^t \overline{Y}_1 = 0. \tag{61}$$

Because each of the matrices involved is of full column rank, these relations express the fact that the corresponding column spaces are complementary, e.g., the columns of $\overline{X}_2$ generate the kernel of X_1^t.

Proof. Use the OSVDs (26), (27), and (28) to show that $\overline{X}_1$ is a solution $T = \overline{X}_1$ to $X_1^t T X_1^t = X_1^t$, $T X_1^t T = T$, $(X_1^t T)^t = X_1^t T$, which are the defining relations for a 1-2-3-inverse. The same argument applies for X_2^t, Y_1^t, Y_3^t. From the OSVDs (26)–(28), the properties (54)–(57) follow immediately. The orthogonality relations (58)–(61) follow from the OSVD (28). ∎

We shall now show how a PSVD can be constructed from the OSVDs (26)–(28) and the 1-2-3-inverses of X_1, X_2, Y_1, Y_3 as in (50)–(53).

THEOREM 2 (An explicit construction of the PSVD). *Assume that A and B do not have a common null space, and let their OSVDs be*

$$A = (U_{a1} \quad U_{a2}) \begin{pmatrix} S_{a1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_{a1}^t \\ V_{a2}^t \end{pmatrix},$$

$$B = (U_{b1} \quad U_{b2}) \begin{pmatrix} S_{b1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_{b1}^t \\ V_{b2}^t \end{pmatrix}.$$

Define a weighted canonical correlation OSVD as

$$S_{a1} V_{a1}^t V_{b1} S_{b1} = (P_1 \quad P_2) \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Q_1^t \\ Q_2^t \end{pmatrix},$$

and a canonical correlation OSVD as

$$V_{a2}^t V_{b2} = (P_3 \quad P_4) \begin{pmatrix} S_3 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Q_3^t \\ Q_4^t \end{pmatrix}.$$

Furthermore, consider the 1-2-3-*inverses as in* (50)–(53). *Then a PSVD of A* *and B is given by*

$$A = \begin{pmatrix} U_{a1}P_1 & U_{a1}P_2 & U_{a2} \end{pmatrix} \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 \\ 0 & I_{r_a - r_1} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_1^t \\ X_2^t \\ X_{31}^t \\ X_{32}^t \end{pmatrix}, \tag{62}$$

$$B = \begin{pmatrix} U_{b1}Q_1 & U_{b2}Q_2 & U_{b2} \end{pmatrix} \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{r_b - r_1} \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} Y_1^t \\ Y_{21}^t \\ Y_{22}^t \\ Y_3^t \end{pmatrix}, \tag{63}$$

where the submatrices of X and Y are given by

$$X_1 = V_{a1}S_{a1}P_1S_1^{-1/2},$$

$$X_2 = V_{a1}S_{a1}P_2,$$

$$X_{31} = V_{b2}Q_3S_3^{-1/2},$$

$$X_{32} = \bar{Y}_3 + V_{b2}V_{b2}^t\left[X_1\left(\bar{Y}_1^t\bar{Y}_1\right)^{-1}\bar{Y}_1^t\bar{Y}_3 + X_2W_5\right]$$

$$= X_1\left(\bar{Y}_1^t\bar{Y}_1\right)^{-1}\bar{Y}_1^t\bar{Y}_3 + X_2W_5 + V_{a2}P_4P_4^tV_{a2}^t\bar{Y}_3;$$

$$Y_1 = V_{b1}S_{b1}Q_1S_1^{-1/2},$$

$$Y_{21} = \bar{X}_2 + V_{a2}V_{a2}^t\left[Y_1\left(\bar{X}_1^t\bar{X}_1\right)^{-1}\bar{X}_1^t\bar{X}_2 + Y_3W_6\right]$$

$$= Y_1\left(\bar{X}_1^t\bar{X}_1\right)^{-1}\bar{X}_1^t\bar{X}_2 + Y_3W_6 + V_{b2}Q_4Q_4^tV_{b2}^t\bar{X}_2,$$

$$Y_{22} = V_{a2}P_3S_3^{-1/2},$$

$$Y_3 = V_{b1}S_{b1}Q_2.$$

The matrix W_5 *is* $(r_a - r_1)\times(r_b - r_1)$, *while* W_6 *is* $(r_b - r_1)\times(r_a - r_1)$. *Both* *are arbitrary except for the constraint*

$$W_5^t + W_6 = \overline{Y}_3^t\overline{Y}_1\left(\overline{Y}_1^t\overline{Y}_1\right)^{-1}\left(\overline{X}_1^t\overline{X}_1\right)^{-1}\overline{X}_1^t\overline{X}_2. \tag{64}$$

Proof. The only fact to be proved is that the matrices X and Y satisfy $X^tY = I_{r_{ab}}$, which is straightforward by exploiting the properties of Lemma 7 and (64). ■

A detailed derivation of the expressions for the submatrices of X and Y can be found in the appendix.

3.4. *Nonuniqueness Properties of the PSVD*

In case A and B do have a common null space, it is straightforward to combine the result of Theorem 2 with the result of Section 3.2.1.

A PSVD of any matrix pair A, B is given by

$$A = \begin{pmatrix} U_{A1} & U_{A2} & U_{A3} \end{pmatrix} \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 & 0 \\ 0 & I_{r_a - r_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_1^t \\ X_2^t \\ X_{31}^t \\ X_{32}^t \\ X_4^t \end{pmatrix}, \tag{65}$$

$$B = \begin{pmatrix} U_{B1} & U_{B2} & U_{B3} \end{pmatrix} \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{r_b - r_1} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} Y_1^t \\ Y_{21}^t \\ Y_{22}^t \\ Y_3^t \\ Y_4^t \end{pmatrix}. \tag{66}$$

The matrices $U_{A1}, U_{A2}, U_{A3}, U_{B1}, U_{B2}, U_{B3}$ can be identified from (63), and the expressions for the submatrices of X and Y are given in Theorem 3. The matrices X_4 and Y_4 are such that

$$\begin{pmatrix} A \\ B \end{pmatrix} X_4 = \begin{pmatrix} A \\ B \end{pmatrix} Y_4 = 0, \qquad X_4^t Y_4 = I_{n - r_{ab}}. \tag{67}$$

The question of nonuniqueness can now be analysed as follows: Insert nonsingular square matrices R, T, W, Z into the above PSVD (66) as

$$A = U_A W D_A T^t X^t, \tag{68}$$

$$B = U_B Z D_B R^t Y^t, \tag{69}$$

with appropriate partitioning of the matrices W, T, Z, R corresponding to the block structure of S_A and S_B. This will correspond to another valid PSVD if the following conditions are satisfied: The matrix $U_A W$ is orthonormal; hence W should be orthonormal. The matrix $U_B Z$ is orthonormal; hence Z should be orthonormal. $W D_A T^t = D_A$ and $Z D_B R^t = D_B$, and finally

$$T^t R = I. \tag{70}$$

Let us analyse these requirements in more detail: From Equations (33) and (35) it follows that X_1 and Y_1 are essentially unique [i.e. apart from (nongeneric) nonuniqueness arising from nondistinct nonzero singular values in one of the OSVDs (26), (27), and (28)]. The nonuniqueness for X_2 and Y_3 is described in (41) and (42). They are unique up to orthonormal matrices W_2 and W_3. The common null space of A and B is also uniquely determined. The nonuniqueness of the choice of basis is characterized by the nonsingular matrix W_1 in (24) and (25).

Combining these observations, it turns out that we can impose the following block structure on the matrices T, R, W, and Z:

$$T = \begin{array}{r} r_1 \\ r_a - r_1 \\ r_{ab} - r_a - r_b + r_1 \\ r_b - r_1 \\ n - r_{ab} \end{array} \overset{\begin{array}{ccccc} r_1 & r_a - r_1 & r_{ab} - r_a - r_b + r_1 & r_b - r_1 & n - r_{ab} \end{array}}{\begin{pmatrix} I & 0 & T_{13} & T_{14} & 0 \\ 0 & T_{22} & T_{23} & T_{24} & 0 \\ 0 & 0 & T_{33} & T_{34} & 0 \\ 0 & 0 & T_{43} & T_{44} & 0 \\ 0 & 0 & 0 & 0 & T_{55} \end{pmatrix}},$$

$$R = \begin{array}{r} r_1 \\ r_a - r_1 \\ r_{ab} - r_a - r_b + r_1 \\ r_b - r_1 \\ n - r_{ab} \end{array} \overset{\begin{array}{ccccc} r_1 & r_a - r_1 & r_{ab} - r_a - r_b + r_1 & r_b - r_1 & n - r_{ab} \end{array}}{\begin{pmatrix} I & R_{12} & R_{13} & 0 & 0 \\ 0 & R_{22} & R_{23} & 0 & 0 \\ 0 & R_{32} & R_{33} & 0 & 0 \\ 0 & R_{42} & R_{43} & R_{44} & 0 \\ 0 & 0 & 0 & 0 & R_{55} \end{pmatrix}},$$

where $T_{22} = W_2$ [see Equation (41)] and $R_{44} = W_3$ [see Equation (42)] are arbitrary but orthonormal. Similarly, the matrices W and Z have the following structure:

$$W = \begin{array}{c} \\ r_1 \\ r_a - r_1 \\ m - r_a \end{array} \!\!\begin{array}{c} \begin{array}{ccc} r_1 & r_a - r_1 & m - r_a \end{array} \\ \begin{pmatrix} I_{r_1} & 0 & 0 \\ 0 & T_{22} & 0 \\ 0 & 0 & W_{33} \end{pmatrix} \end{array} \tag{71}$$

$$Z = \begin{array}{c} \\ r_1 \\ r_b - r_1 \\ p - r_b \end{array} \!\!\begin{array}{c} \begin{array}{ccc} r_1 & r_b - r_1 & p - r_b \end{array} \\ \begin{pmatrix} I_{r_1} & 0 & 0 \\ 0 & R_{44} & 0 \\ 0 & 0 & Z_{33} \end{pmatrix} \end{array} \tag{72}$$

where W_{33} and Z_{33} are arbitrary but orthonormal. From the condition (70), it is straightforward to show that $T_{13}, T_{14}, T_{43}, R_{12}, R_{13}, R_{23}$ must all be zero and that T_{33} and T_{55} are nonsingular. Hence

$$T = \begin{pmatrix} I & 0 & 0 & 0 & 0 \\ 0 & T_{22} & T_{23} & T_{24} & 0 \\ 0 & 0 & T_{33} & T_{34} & 0 \\ 0 & 0 & 0 & T_{44} & 0 \\ 0 & 0 & 0 & 0 & T_{55} \end{pmatrix}, \tag{73}$$

and from $R = T^{-t}$ it follows that

$$R = \begin{pmatrix} I & 0 & 0 & 0 & 0 \\ 0 & T_{22} & 0 & 0 & 0 \\ 0 & -T_{33}^{-t}T_{23}^{t}T_{22}^{-t} & T_{33}^{-t} & 0 & 0 \\ 0 & -T_{44}^{-t}(T_{24}^{t} - T_{34}^{t}T_{33}^{-t}T_{23}^{t})T_{22}^{-t} & -T_{44}^{-t}T_{34}^{t}T_{33}^{-t} & T_{44} & 0 \\ 0 & 0 & 0 & 0 & T_{55}^{-t} \end{pmatrix}. \tag{74}$$

The conclusion is summarized in the following:

THEOREM 3 (On the nonuniqueness of the PSVD). *If a PSVD of* A, B *is given by*

$$A = \begin{pmatrix} U_{A1} & U_{A2} & U_{A3} \end{pmatrix} \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_1^t \\ X_2^t \\ X_{31}^t \\ X_{32}^t \\ X_4^t \end{pmatrix}, \tag{75}$$

$$B = \begin{pmatrix} U_{B1} & U_{B2} & U_{B3} \end{pmatrix} \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} Y_1^t \\ Y_{21}^t \\ Y_{22}^t \\ Y_3^t \\ Y_4^t \end{pmatrix}, \tag{76}$$

then the following is also a PSVD:

$$A = \begin{pmatrix} U_{a1} & U_{a2}T_{22} & U_{A3}W_{33} \end{pmatrix} \times \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_1^t \\ T_{22}^t X_2^t \\ T_{23}^t X_2^t + T_{33}^t X_{31}^t \\ T_{24}^t X_2^t + T_{34}^t X_{31}^t + T_{44}^t X_{32}^t \\ T_{55}^t X_4^t \end{pmatrix},$$

$$B = \begin{pmatrix} U_{B1} & U_{B2}T_{44} & U_{B3}Z_{33} \end{pmatrix} \times \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} Y_1^t \\ R_{22}^t Y_{21}^t + R_{23}^t Y_{22}^t + R_{24}^t Y_3^t \\ R_{33}^t Y_{22}^t + R_{34}^t Y_3^t \\ R_{44}^t Y_3^t \\ T_{55}^{-1} Y_4^t \end{pmatrix}.$$

The blocks T_{ij} are arbitrary except for T_{22} and T_{44}, which should be orthonormal, and T_{33} and T_{55}, which should be nonsingular. The blocks R_{ij} are determined by (70) *and are given in* (74). *The matrices W_{33} and Z_{33} are arbitrary orthonormal.*

To conclude this section, observe that we have characterized the nonuniqueness of the PSVD on a double level: In Theorem 2, we have derived an explicit construction of the PSVD from four OSVDs that could be obtained from the matrices A and B. With the observation of Section 3.2.1 about a common null space, it then became clear that the matrices X and Y are partitioned into five submatrices. Even here there is already some nonuniqueness parametrized by the matrices W_5 and W_6, which are arbitrary apart from the constraint (64). In Theorem 3, it is shown that, once a PSVD is known with the corresponding partitioning into five submatrices for X and Y, all other PSVDs for the matrix pair can be obtained by inserting some matrices W, Z, T, and R. The matrices W and Z have a block diagonal structure as in (71) and (72). The matrices T and R have the block triangular structure of (73) and (74). This block triangular structure will be important in the geometrical interpretation of the submatrices of X and Y in Section 4. It is an interesting exercise to show that the matrices XT and YR, where T and R have the required block structure from Theorem 3, solve the set of nonlinear equations (47)–(49). Hence, Theorem 3 also gives all solutions to this set of equations, whereas Theorem 2 only described one particular solution.

4. GEOMETRICAL INTERPRETATION OF THE STRUCTURE

In this section, we shall relate the structure of the contragredient transformation, as derived in the previous section, to the geometry of subspaces related to A and B.

Let $r_a = \operatorname{rank} A$, $r_b = \operatorname{rank} B$, and the OSVD of A and B be as in (26) and (27). Let r_{ab} be defined as

$$r_{ab} = \operatorname{rank}\begin{pmatrix} A \\ B \end{pmatrix}$$

Then it is well known that $r_{ab} = r_a + r_b - \dim[R(A^t) \cap R(B^t)]$. Let r_1 be defined as in (28), where $r_1 = \operatorname{rank}(S_{a1}V_{a1}^tV_{b1}S_{b1}) = \operatorname{rank}(V_{a1}^tV_{b1})$, where the second equality follows from the nonsingularity of S_{a1} and S_{b1}. From the definition of angles between subspaces, as e.g. in [2], it follows immediately that r_1 is the number of canonical angles different from 90° between the row spaces of A and B:

$$r_1 = \dim[\Pi_{A^t} R(B^t)] = \dim[\Pi_{B^t} R(A^t)]. \tag{77}$$

Hence $r_1 = 0$ only of the row spaces of A and B are orthogonal, as was the case in Lemma 3. Assume that r_{c1} of these canonical angles are zero while the $r_{c2} = r_1 - c_{c1}$ others are not. Obviously,

$$r_{c1} = \dim[R(A^t) \cap R(B^t)].$$

Hence $r_{ab} = r_a + r_b - r_{c1}$ and $r_{ab} \geqslant r_a + r_b - r_1$. This is nothing else than the inequality (43), which was derived from a structural requirement, whereas the derivation here is based on a geometrical argument.

Because r_{c2} is the number of nonzero canonical angles different from 90° between the row spaces of A and B, it is also the number of nonzero canonical angles different from 90° between the ranges of V_{a2}, V_{b2}. Hence

$$r_{c2} = r_1 - r_{c1} = r_1 + r_{ab} - r_a - r_b = \#\{0 < \sigma(V_{a2}^t V_{b2}) < 1\}.$$

Now consider the partitioning of X and Y as derived in Section 3, which is repeated here for convenience:

$$X = \begin{pmatrix} \overset{r_1}{X_1} & \overset{r_a - r_1}{X_2} & \overset{r_{ab} - r_a - r_b + r_1}{X_{31}} & \overset{r_b - r_1}{X_{32}} & \overset{n - r_{ab}}{X_4} \end{pmatrix},$$

$$Y = \begin{pmatrix} \overset{r_1}{Y_1} & \overset{r_a - r_1}{Y_{21}} & \overset{r_{ab} - r_a - r_b + r_1}{Y_{22}} & \overset{r_b - r_1}{Y_3} & \overset{n - r_{ab}}{Y_4} \end{pmatrix}.$$

With an obvious partitioning of the orthonormal matrices U_A and U_B as in Theorem 3, it is straightforward to derive the following *generalized dyadic decomposition*:

$$A = U_{A1} S_1^{1/2} X_1^t + U_{A2} X_2^t, \tag{78}$$

$$B = U_{B1} S_1^{1/2} Y_1^t + U_{B3} Y_3^t, \tag{79}$$

which can be written out as a sum of rank one terms.

From the fact that $X^t Y = Y^t X = I_n$, it follows that

$$A\begin{pmatrix} Y_1 & Y_{21} & Y_{22} & Y_3 & Y_4 \end{pmatrix} = \begin{pmatrix} U_{A1}S_1^{1/2} & U_{A2} & 0 & 0 & 0 \end{pmatrix}, \quad (80)$$

$$B\begin{pmatrix} X_1 & X_2 & X_{31} & X_{32} & X_4 \end{pmatrix} = \begin{pmatrix} U_{B1}S_1^{1/2} & 0 & 0 & U_{B3} & 0 \end{pmatrix}. \quad (81)$$

From these, the following geometrical characterizations can be derived. $R(A^t)$ is generated by the columns of X_1 and X_2. Hence, the row space of the matrix A can be split into two subspaces; $R(X_2)$ forms a subspace of $R(A^t)$, which is orthogonal to $R(B^t)$. It can be verified that

$$\operatorname{rank} X_2 = r_a - r_1 = \#\{\sigma(V_{b2}^t V_{a1}) = 1\}. \quad (82)$$

$R(X_1)$ forms a subspace of the row space of A, which is not orthogonal to the row space of B. Its dimension is r_1, as follows also from (77):

$$r_1 = \#\{\sigma(V_{a1}^t V_{b1}) > 0\}. \quad (83)$$

$N(B)$ is generated by the columns of X_2, X_{31}, X_4. Hence, the null space of B can be decomposed into three subspaces: $R(X_2)$ is a subspace of $R(A^t)$. $R(X_{31})$ is orthogonal to $R(B^t)$, hence a subspace of $N(B)$, but is not contained in $R(A^t)$. Hence

$$r_{ab} - r_a - r_b + r_1 = \#\{0 < \sigma(V_{a1}^t V_{b2}) < 1\}. \quad (84)$$

$R(X_4)$ is the common null space of A and B. Obviously,

$$n - r_{ab} = \#\{\sigma(V_{a2}^t V_{b2}) = 1\}. \quad (85)$$

Also, it follows immediately that

$$X_4^t X_1 = 0, \quad (86)$$

$$X_4^t X_2 = 0. \quad (87)$$

$R(B^t)$ is generated by the columns of Y_1 and Y_3. Hence, the row space of the matrix B can be split into two subspaces: $R(Y_1)$ forms a subspace of $R(B^t)$, which is not orthogonal to $R(A^t)$. Its dimension is r_1. $R(Y_3)$ forms a subspace of $R(B^t)$, which is orthogonal to $R(A^t)$. It can be verified that

$$\operatorname{rank} Y_3 = r_b - r_1 = \#\{\sigma(V_{a2}^t V_{b1}) = 1\}. \tag{88}$$

$N(A)$, the null space of A, is generated by the columns of Y_{22}, Y_3, Y_4. $R(Y_{22})$ is orthogonal to $R(A^t)$ but not contained in $R(B^t)$. Hence

$$r_{ab} - r_a - r_b + r_1 = \#\{0 < \sigma(V_{a2}^t V_{b1}) < 1\}. \tag{89}$$

$R(Y_3)$ is orthogonal to $R(A^t)$ and also a subspace of $R(B^t)$. $R(Y_4)$ is the common null spaces of A and B. Hence

$$Y_4^t Y_1 = 0, \tag{90}$$

$$Y_4^t Y_3 = 0. \tag{91}$$

Moreover,

$$R(X_4) = R(Y_4). \tag{92}$$

It can be verified that these geometrical results are independent of the nonuniqueness of the matrices X and Y as described in Theorem 4. The reason for this independence is precisely the block triangular structure of the matrices T (73) and R (74).

In order to appreciate this observation, compare the structure of the matrix X with that of the matrix XT in Theorem 4. Take for instance the matrix X_{31}. The matrix X_{31} undergoes an affine transformation of the form $X_{31} \to X_{31}T_{33} + X_2T_{23}$. It is easy to check from $Y^tX = I_n$ that $R(X_{31}T_{33} + X_2T_{23})$ is orthogonal to $R(B^t)$. Moreover, because T_{33} is nonsingular, $X_{31}T_{33} + X_2T_{23}$ will never be contained in the row space of A because X_{31} isn't either. In summary, all statements for X_{31} remain true for $X_{31}T_{33} + X_2T_{23}$. The same applies for the other submatrices of X and Y.

5. CONCLUSIONS

In this paper, we have investigated the structural properties of the product singular value decomposition (PSVD) of two matrices A and B. First, we have derived a constructive proof, which exploits the close relation of the PSVD with the OSVD of AB^tBA^t and the eigenvalue decompositions of AA^tBB^t and BB^tAA^t. Next, we have provided a detailed analysis of the structural and geometrical properties of the so-called contragredient transformation of the two symmetric matrices A^tA and B^tB, both of which are nonnegative and/or positive definite. A complete characterization and description of the nonuniqueness was obtained. The geometry of the structure was interpreted in terms of principal angles between subspaces.

Recently, some more elegant constructive proofs for the PSVD and other generalizations (such as generalized QR decompositions) have been obtained. They are reported in [6].

APPENDIX. A SOLUTION OF THE NONLINEAR MATRIX EQUATIONS THAT DEFINE THE CONTRAGREDIENT TRANSFORMATION

Observe that the linear equations (47)–(48) form an underdetermined set. With the factorizations of X_1, X_2, Y_1, and Y_3 in (33)–(36) one can apply Lemma 1 to obtain the general solution to the underdetermined equations as

$$\begin{pmatrix} X_{31} & X_{32} \end{pmatrix} = V_{b1}S_{b1}^{-1}\begin{pmatrix} Q_1S_1^{-1/2} & Q_2 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & I_{r_b - r_1} \end{pmatrix} + V_{b2}\begin{pmatrix} Z_1^x & Z_2^x \end{pmatrix}, \quad (93)$$

$$\begin{pmatrix} Y_{21} & Y_{22} \end{pmatrix} = V_{a1}S_{a1}^{-1}\begin{pmatrix} P_1S_1^{-1/2} & P_2 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ I_{r_a - r_1} & 0 \end{pmatrix} + V_{a2}\begin{pmatrix} Z_1^y & Z_2^y \end{pmatrix}, \quad (94)$$

where $Z_1^x, Z_2^x, Z_1^y, Z_2^y$ are arbitrary matrices of appropriate dimensions. The first term in (93) and (94) is a particular solution, while the second term is the general solution to the homogeneous equations obtained from (47) and (48). The determination of X_{31}, X_{32}, Y_{21}, and Y_{21} reduces to the determination of $Z_1^x, Z_2^x, Z_1^y, Z_2^y$ in

$$X_{31} = V_{b2}Z_1^x, \quad (95)$$

$$X_{32} = V_{b1}S_{b1}^{-1}Q_2 + V_{b2}Z_2^x, \quad (96)$$

$$Y_{21} = V_{a1}S_{a1}^{-1}P_2 + V_{a2}Z_1^y, \quad (97)$$

$$Y_{22} = V_{a2}Z_2^y \quad (98)$$

subject to the conditions

$$X_{31}^t Y_{21} = 0, \tag{99}$$

$$X_{32}^t Y_{21} = 0, \tag{100}$$

$$X_{32}^t Y_{22} = 0, \tag{101}$$

$$X_{31}^t Y_{22} = I_{r_{ab}-r_a-r_b+r_1}. \tag{102}$$

Observe that this is a set of *nonlinear* equations in the unknown matrices $Z_1^x, Z_2^x, Z_1^y, Z_2^y$.

Determination of X_{31} and Y_{22}: Canonical Correlation

Substituting the expressions for X_{31} (95) and Y_{22} (98) into the last constraint (102) results in

$$(Z_1^x)^t V_{b2}^t V_{a2} Z_2^y = I_{r-r_a-r_b+r_1}. \tag{103}$$

Since both V_{a2} and V_{b2} are orthonormal matrices, the OSVD of the product $V_{a_2}^t V_{b2}$ corresponds to a canonical correlation analysis between the kernels of the matrices A and B. It can be shown that the number of nonzero singular values of $V_{a_2}^t V_{b2}$ must be equal to $r_{ab} - r_a - r_b + r_1$, because the number of nonzero singular values of $V_{a1}^t V_{b1}$ is equal to r_1. Hence, Z_1^x and Z_2^y can be determined from the OSVD of $V_{a2}^t V_{b2}$:

$$V_{a2}^t V_{b2} = \begin{pmatrix} P_3 & P_4 \end{pmatrix} \begin{pmatrix} S_3 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Q_3^t \\ Q_3^t \end{pmatrix}, \tag{104}$$

where S_3 is an $(r_{ab} - r_a - r_b + r_1) \times (r_{ab} - r_a - r_b + r_1)$ nonsingular diagonal matrix and the matrices of left and right singular vectors are partitioned accordingly. One possible solution for X_{31} and Y_{22} follows immediately from this OSVD as

$$X_{31} = V_{b2} Q_3 S_3^{-1/2}, \tag{105}$$

$$Y_{22} = V_{a2} P_3 S_3^{-1/2}. \tag{106}$$

Observe that this is *not* the most general solution to (95)–(102), but only a specific one.

The Determination of X_{32} and Y_{21}

Having determined expressions for X_{31} (105) and Y_{22} (106) from a canonical correlation analysis between the kernels of A and B, the orthogonality conditions (99)–(102) permit us to derive two other equations for X_{32} and Y_{21}.

First observe that from (26) and (27), and from (104), it follows that

$$Q_3^t V_{b2}^t \begin{pmatrix} V_{b1} & V_{b2} Q_4 \end{pmatrix} = 0, \tag{107}$$

$$P_3^t V_{a2}^t \begin{pmatrix} V_{a1} & V_{a2} P_4 \end{pmatrix} = 0. \tag{108}$$

From Equations (105) and (100) it follows that

$$X_{31}^t Y_{21} = S_3^{-1/2} Q_3^t V_{b2}^t Y_{21} = 0, \tag{109}$$

while from (106) and (101) it follows that

$$Y_{22}^t X_{32} = S_3^{-1/2} P_3^t V_{a2}^t X_{32} = 0 \tag{110}$$

The combination of equations (107) together with (109) permits us to conclude via Lemma 1 that there must exist matrices Z_3^y, Z_4^y, of appropriate size, such that

$$Y_{21} = V_{b1} Z_3^y + V_{b2} Q_4 Z_4^y. \tag{111}$$

Similarly, it follows from (108) and (110) that

$$X_{32} = V_{a1} Z_3^x + V_{a2} P_4 Z_4^x. \tag{112}$$

Hence, there are two equations for X_{32}, namely (96) and (112), and two equations for Y_{21}, (97) and (111). These are now repeated for convenience:

$$Y_{21} = V_{a1} S_{a1}^{-1} P_2 + V_{a2} Z_1^y \tag{113}$$

$$= V_{b1} Z_3^y + V_{b2} Q_4 Z_4^y \tag{114}$$

and

$$X_{32} = V_{b1}S_{b1}^{-1}Q_2 + V_{b2}Z_2^x \tag{115}$$

$$= V_{a1}Z_3^x + V_{a2}P_4Z_4^x. \tag{116}$$

From these four equations, we shall eliminate all unknown matrices in four steps:

Step 1: *Elimination of Z_1^y and Z_4^y.* Recall the OSVD of $V_{a2}^tV_{b2}$ (104). Premultiplication of the expressions for Y_{21} (113)–(114) with V_{a2}^t results in

$$Z_1^y = V_{a2}^tV_{b1}Z_3^y, \tag{117}$$

and with $Q_4^tV_{b2}^t$ results in

$$Z_4^y = Q_4^tV_{b2}^tV_{a1}S_{a1}^{-1}P_2. \tag{118}$$

Upon substitution in (113) and (114), this gives

$$Y_{21} = V_{a1}S_{a1}^{-1}P_2 + V_{a2}V_{a2}^tV_{b1}Z_3^y \tag{119}$$

$$= V_{b1}Z_3^y + V_{b2}Q_4Q_4^tV_{b2}V_{a1}S_{a1}^{-1}P_2. \tag{120}$$

If these expressions are premultiplied with V_{b1}^t, we get a set of linear equations for Z_3^y:

$$\left(I_{r_b} - V_{b1}^tV_{a2}V_{a2}^tV_{b1}\right)Z_3^y = V_{b1}^tV_{a1}S_{a1}^{-1}P_2.$$

Observe that the first factor on the left hand side can be rewritten as

$$I_{r_b} - V_{b1}^tV_{a2}V_{a2}^tV_{b1} = V_{b1}^t\left(I_r - V_{a2}V_{a2}^t\right)V_{b1}$$

$$= V_{b1}^tV_{a1}V_{a1}^tV_{b1}.$$

Hence, the equation for Z_3^y reads

$$V_{b1}^tV_{a1}V_{a1}^tV_{b1}Z_3^y = V_{b1}^tV_{a1}S_{a1}^{-1}P_2. \tag{121}$$

Step 2: Elimination of Z_2^x and Z_4^x. In a similar manner, one can derive the following set of linear equations for Z_3^x:

$$V_{a1}^t V_{b1} V_{b1}^t V_{a1} Z_3^x = V_{a1}^t V_{b1} S_{b1}^{-1} Q_2. \tag{122}$$

Step 3: A general solution for Z_3^x and Z_3^y. Rewrite Equation (121) for Z_3^y, using the OSVD of $S_{a1} V_{a1}^t V_{b1} S_{b1} = P_1 S_1 Q_1^t$ (28), as

$$S_{b1}^{-1} Q_1 S_1 P_1^t S_{a1}^{-2} P_1 S_1 Q_1^t S_{b1}^{-1} Z_3^y = S_{b1}^{-1} Q_1 S_1 P_1^t S_{a1}^{-1} P_2.$$

Using the 1-2-3-inverses, defined in Lemma 7, this can be rewritten more compactly as

$$(\bar{X}_1^t \bar{X}_1) \bar{Y}_1^t V_{b1} Z_3^y = \bar{X}_1^t \bar{X}_2. \tag{123}$$

The following observations are crucial:

1. The matrix $(\bar{X}_1^t \bar{X}_1)$ is square nonsingular.
2. The columns of the matrix Y_3 are complementary to and orthogonal to the columns of the matrix $\bar{Y}_1$ [Equation (60)].
3. Recall the relation $\bar{Y}_1^t Y_1 = I_{r_1}$ [Equation (56)].

It follows from Lemma 1 that the general solution for $V_{b1} Z_3^y$ is given by

$$V_{b1} Z_3^y = Y_1 (\bar{X}_1^t \bar{X}_1)^{-1} \bar{X}_1^t \bar{X}_2 + Y_3 W_6, \tag{124}$$

where W_6 is an arbitrary $(r_b - r_1) \times (r_a - r_1)$ matrix. The first term is a particular solution, while the second term is the general solution to the homogeneous equation. In a completely similar way, one obtains the general solution for $V_{a1} Z_3^x$ from (122) as

$$V_{a1} Z_3^x = X_1 (\bar{Y}_1^t \bar{Y}_1)^{-1} \bar{Y}_1^t \bar{Y}_3 + X_2 W_5, \tag{125}$$

where W_5 is an arbitrary $(r_a - r_1) \times (r_b - r_1)$ matrix. However, as will now be shown, the matrices W_5 and W_6 are not independent of each other, because of the orthogonality condition $X_{32}^t Y_{21} = 0$ (100). For this we shall need the following properties: Using (54)–(61), it is straightforward to show from (124) and (125) that

$$\bar{X}_2^t V_{a1} Z_3^x = W_5, \tag{126}$$

$$\bar{Y}_3^t V_{b1} Z_3^y = W_6. \tag{127}$$

Also, from multiplying (124) with (125) and using the orthogonality conditions (58)–(61), it follows that

$$(Z_3^x)^t V_{a1}^t V_{b1} Z_3^y = \bar{Y}_3^t \bar{Y}_1 (\bar{Y}_1^t \bar{Y}_1)^{-1} (\bar{X}_1^t \bar{X}_1)^{-1} \bar{X}_1^t \bar{X}_2. \tag{128}$$

Step 4: The remaining orthogonality condition. So far, we have obtained a general expression for $V_{a1} Z_3^x$ (125) and $V_{b1} Z_3^y$ (124). The expressions for X_{32} (115)–(116) and Y_{21} (113)–(114) can be rewritten as

$$X_{32} = V_{a1} Z_3^x + (V_{a2} P_4)(P_4^t V_{a2}^t) \bar{Y}_3 \tag{129}$$

$$= \bar{Y}_3 + V_{b2} V_{b2}^t (V_{a1} Z_3^x), \tag{130}$$

$$Y_{21} = V_{b1} Z_3^y + (V_{b2} Q_4)(Q_4^t V_{b2}^t) \bar{X}_2 \tag{131}$$

$$= \bar{X}_2 + V_{a2} V_{a2}^t (V_{b1} Z_3^y). \tag{132}$$

The expressions for $V_{a1} Z_3^x$ and $V_{b1} Z_3^y$ contain two arbitrary matrices W_5 and W_6. However, it will now be derived how the only remaining orthogonality requirement,

$$X_{32}^t Y_{21} = 0,$$

induces a constraint between W_5 and W_6. To do so, we shall substitute the expressions for X_{32} and Y_{21} into the orthogonality condition. Equation (129) × Equation (131) results in

$$(Z_3^x)^t V_{a1}^t V_{b1} Z_3^y + (Z_3^x)^t V_{a1}^t (V_{b2} Q_4)(Q_4^t V_{b2}^t) \bar{X}_2$$

$$+ \bar{Y}_3^t (V_{a2} P_4)(P_4^t V_{a2}^t) V_{b1} Z_3^y = 0. \tag{133}$$

Equation (130) × Equation (131) results in

$$\bar{Y}_3^t V_{b1} Z_3^y + (Z_3^x)^t V_{a1}^t (V_{b2} Q_4)(Q_4^t V_{b2}^t) \bar{X}_2 = 0. \tag{134}$$

Equation (129) × Equation (132) results in

$$(Z_3^x)^t V_{a1}^t \bar{X}_2 + \bar{Y}_3^t (V_{a2} P_4)(P_4^t V_{a2}^t) V_{b1} Z_3^y = 0. \tag{135}$$

Equations (134) and (135) permit us to simplify Equation (133) as

$$(Z_3^x)^t V_{a1}^t V_{b1} Z_3^y - \bar{Y}_3^t V_{b1} Z_3^y - (Z_3^x)^t V_{a1}^t \bar{X}_2 = 0. \tag{136}$$

Now use Equations (126) and (127) to get

$$(Z_3^x)^t V_{a1}^t V_{b1} Z_3^y = W_5^t + W_6. \tag{137}$$

It follows then from Equation (128) that

$$W_5^t + W_6 = \bar{Y}_3^t \bar{Y}_1 (\bar{Y}_1^t \bar{Y}_1)^{-1} (\bar{X}_1^t \bar{X}_1)^{-1} \bar{X}_1^t \bar{X}_2. \tag{138}$$

This is the constraint between W_5 and W_6 that ensures the orthogonality between X_{32} and Y_{21}.

Observe that the sum $W_5^t + W_6$ is the product of the least squares solutions to

$$\bar{X}_1 x = \bar{X}_2,$$

$$\bar{Y}_1 z = \bar{Y}_3.$$

I would like to thank Gene Golub of the Department of Computer Science and Thomas Kailath of the Department of Electrical Engineering (Information Systems Lab) for the opportunity they gave me to spend a wonderful year at Stanford University.

REFERENCES

1 K. S. Arun and S. Y. Kung, Generalized principal component analysis and its application in approximate stochastic realization, in *Modelling and Application of Stochastic Processes* (U. B. Desai, Ed.), Kluwer Academic, 1986, pp. 75–104.

2 A. Björck and G. Golub, Numerical methods for computing angles between linear subspaces, *Math. Comp.* 27(123):579–594 (1973).

3 B. De Moor and G. H. Golub, The restricted singular value decomposition: Properties and Applications, *SIAM Matrix Anal. Appl.*, Vol. 12, no. 3, July 1991.

4 B. De Moor and G. H. Golub, Generalized Singular Value Decompositions: A Proposal for a Standardized Nomenclature, Numerical Analysis Project Report 89-04, Dept. of Computer Science, Stanford Univ., Apr. 1989.

5 B. De Moor and H. Zha, A tree of generalizations of the ordinary singular value decomposition, *Linear Algebra Appl.*, special issue on Canonical Forms of Matrices, 147:469–500 (Jan. 1991).

6 B. De Moor and P. Van Dooren, Generalizations of the *QR* and the Singular value Decomposition, ESAT-SISTA Report 1991-03, Dept. of Electrical Engineering, Katholieke Univ. Leuven, Belgium; submitted for publication.

7 E. Deprettere (Ed.), *SVD and Signal Processing: Algorithms, Applications and Architectures*, North Holland, 1988.

8 K. V. Fernando and S. J. Hammarling, A Product Induced Singular Value Decomposition for Two Matrices and Balanced Realisation, NAG Technical Report TR8/87, 1987.

9 K. V. Fernando, The Kalman Reachability/Observability Canonical Form and the ΠSVD, NAG Technical Report TR9/87, July 1987.

10 B. Francis, *A course in H_∞ Control Theory*, Lecture Notes in Control and Inform. Sci. (M. Thoma and A. Wynes, Eds.), Springer-Verlag, Berlin, 1987.

11 K. Glover, All optimal Hankel norm approximations of linear multivariable systems and their L_∞ error bounds, *Internat. J. Control* (6):1115–1193 (1984).

12 G. H. Golub and C. F. Van Loan, *Matrix Computations*, North Oxford Academic, Oxford, 1983.

13 M. T. Heat, A. J. Laub, C. C. Paige, and R. C. Ward, Computing the singular value decomposition of a product of two matrices, *SIAM J. Sci. Statist. Comput.* 8, No. 4 (1986).

14 E. Jonckheere and L. Silverman, A new set of invariants for linear systems, applications to reduced order compensator design, *IEEE Trans. Automat. Control* AC-28, No. 1 (1983).

15 C. Kenney and G. Hewer, Necessary and sufficient conditions for balancing unstable systems, *IEEE Trans. Automat. Control* AC-32, No. 2 (1987).

16 A. J. Laub, M . T. Heath, C. C. Paige, and R. C. Ward, Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms, *IEEE Trans. Automat. Control* AC-32, No. 2 (1987).

17 B. C. Moore, Principal component analysis in linear systems: Controllability, observability and model reduction, *IEEE Trans. Automat. Control*. AC-26 (1):17–32 (1981).

18 C. C. Paige and M. A. Saunders, Towards a generalized singular value decomposition, *SIAM J. Numer. Anal.* 18:398–405 (1981).
19 C. F. Van Loan, Generalizing the singular value decomposition, *SIAM J. Numer. Anal.* 13:76–83 (1976).
20 H. Zha, Restricted SVD for Matrix Triplets and Rank Determination of Matrices, Scientific Report 89-2, ZIB, Berlin; *SIMAX*, to appear.

Received 6 June 1989; final manuscript accepted 13 March 1991