Numerical Analysis Project
Manuscript NA-89-06

May 1989

On the Structure and Geometry of the Product Singular Value Decomposition

by

Bart L.R. De Moor

Numerical Analysis Project
Computer Science Department
Stanford University
Stanford, California 94305

On the structure and geometry of the product singular value decomposition.*

Bart L.R. De Moor †
Department of Computer Sciences
Stanford University
CA 94305 Stanford
tel: 415-723-1923
email: demoor@patience.stanford.edu
fax:(415)-725-7411

May 24, 1989

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 **eigen**-value problem. Several interesting properties are established.

*Research supported in part by the US-Army under contract **DAAL03-87-K-0095**

†**Dr.** De Moor is on leave from the Katholieke Universiteit Leuven, Belgium. He is now with the Department of Computer Sciences and with the Information Systems Lab, Department of Electrical Engineering, Stanford University. He is supported partially by an Advanced Research Fellowship in Science and Technology of the NATO Science Fellowships Programme and a grant from IBM.

The structure and the non-uniqueness 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.
Keywords: (Generalized) singular value decompositions, contragredient transformation.

1 Introduction

The *ordinary singular value decomposition* (OSVD) has become an important tool in the analysis and numerical solutions 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 applications are discussed in e.g. [5] [10] and the references therein.
Recently, several generalizations of the singular value decomposition have been derived and analysed. The most well known example is the so called '*generalized*' *singular value decomposition* of Van Loan [17] and Paige and Saunders [16]. 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 [6] (who called it the IISVD). In [18], Zha introduced yet another generalization of the OSVD, this time for 3 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 these 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 is 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 operations or matrix-(pseuclo)-inversions, which may cause non-trivial losses of numerical accuracy when implemented on a finite precision machine.

The PSVD is a generalization for 2 matrices of the OSVD. In this respect, it is a kind of 'dual' generalization of the OSVD compared 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, a similar knowledge for the PSVD is less well developed. It is one of the goals of this paper to provide some more insight in the *structure and geometry* of the PSVD. Algorithmic ideas to actually implement the PSVD in a numerically robust way can be found in [6] and [11]. Applications include the orthogonal Procrustes problem [10], computing balancing transformations for state space systems [6][14] and computing the Kalman decomposition of a linear system [7]. The PSVD could also be applied in the computation of approximate intersections between subspaces in the stochastic realization problem [1], as an alternative for canonical correlation analysis. The main difference between the 2 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 normalisation. However, these issues will not be discussed in this paper.

The main results of this paper concentrate around *2 constructive proofs* of the PSVD. The first one exploits the close relationship of the PSVD to the OSVD and several eigenvalue problems. In the second proof, we provide a profound analysis of the non-uniqueness 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 con-

[1] As a matter of fact, recently, Zha Hongyuang and the author have established a most interesting result that both the PSVD and the QSVD are 'parents' of an infinite chain of generalizations of the OSVD. This result will be published in due course.

tragredient transformations for 2 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 2 matrices A and B.

The main results and organisation of this paper are as follows:

- The constructive proof of the PSVD of 2 matrices A and B in section 2 exploits the connection between the OSVD of the matrix AB^tBA^t and the eigenvalue decomposition of the matrix A^tAB^tB.
- In section 2, we also investigate the connection of the PSVD with the QSVD and give a variational interpretation.
- The structure of the so called contragredient transformation is investigated-m section 3. We summarize some known results for existence and uniqueness of a contragredient transformation for pairs of symmetric matrices, where one of the matrices is positive definite and the other is nonnegative definite. The results in sections 3.2-3.3 give a precise account of the structure of this transformation for 2 symmetric nonnegative definite matrices. It will be demonstrated that the question of characterizing the non-uniqueness issues of the PSVD is not an easy one. First, it will be shown in section 3.3 how a certain 'canonical' PSVD can be explicitly constructed from some OSVDs of the matrices A and B. The complete description of the non-uniqueness is given in section *3.4.*
- The geometrical interpretation given in section 4 concentrates on the relation with principal angles between certain subspaces of the 2 matrices.

Notations and Abbreviations

All matrices and vectors in this paper are real. Matrices are denoted by capitals, vector by lower case 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 2 matrices A and B is denoted by $R(A) \bigcap R(B)$. The direct sum of 2 mutually orthogonal subspaces $R(U_1)$ and $R(U_2)$ ($U_1^t U_2 = 0$) is denoted by $R(U_1) \oplus R(U_2)$. The dimension of a subspace is abbreviated as *dim*, hence $dim(R(A)) = rank(A) = dim(R(A^t))$. By $\#\{\sigma(A) = 1\}$ we denote the number of singular values of A equal to 1.

It is assumed that, whenever a dimension indicating number becomes zero, the corresponding matrix, block row or block column can be omitted in all expressions where it appears. This convention allows for an elegant treatment of several possible cases at once. Dimensions of identity matrices are omitted if they are obvious from the context.

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 proof a lemma that permits 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 and investigate a relation between the PSVD and the QSVD.

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:

$$\begin{aligned} A &= U_A S_A X^t \\ B &= U_B S_B X^{-1} \end{aligned}$$

All 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:*

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

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

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

Before proving the theorem, let us first give the following remarks:

- While **some related** eigenvalue **problems were discussed in** [11] **and** [14], **the explicit formulation of the PSVD as n theorem 1, was given for the first time by Fernando and Hammarling in** [6], **who called it the** ΠSVD.[2]

- **Throughout the paper, we shall also use the matrix** Y **defined as** $Y = X^{-t}$.

- **In** [6], **the factorization is presented in a slightly different form, where a QR-factorization of X is used. While this may be preferable in**

[2] In [6] also a constructive proof was provided. It is however based on a lemma (lemma 1 in [6]), the proof of which is not correct. To give a counterexample to the proof, consider the pair of matrices:

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \end{pmatrix}$$
$$B = \begin{pmatrix} 3 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

This pair of matrices satisfies the condition required by the lemma in [6] that AB^t is diagonal. With the notations of [6], 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 [6] is not correct.

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.

- **Here are some examples of possible structures of S_A and S_B in the PSVD of theorem 1:**

$m = 4, p = 4, n = 7, r_a = 3, r_b = 4, r_1 = 2$:

$$S_A = \begin{pmatrix} 0 & \sqrt{\sigma_2} & 0 & 0 & 0 & 0 & 0 \\ \sqrt{\sigma_1}0 & 00(0 & 0 & 0 & 0 & 001000 & 0000 \end{pmatrix}$$

$$S_B = \begin{pmatrix} \sqrt{\sigma_1} & 0\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 & 0 \\ 0 & \sqrt{\sigma_2} & 0 & 0 & 010 & 0 \end{pmatrix}$$

$m = 4, p = 5, n = 4, r_a = 4, r_b = 3, r_1 = 3$:

$$S_A = \begin{pmatrix} \sqrt{\sigma_1} & 0 & 0 & 0 \\ 0 & \sqrt{\sigma_2} & 0 & 0 \\ 0 & 0 & \sqrt{\sigma_3} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$S_B = \begin{pmatrix} \sqrt{\sigma_1} & 0 & 0 & 0 \\ 0 & \sqrt{\sigma_2} & 0 & 0 \\ 0 & 0 & \sqrt{\sigma_3} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

- **Of course, the PSVD ressembles closely the QSVD of 2 matrices, at least in appearance:**

Theorem 2 *The quotient (generalized) SVD* **(QSVD)**
Every pair of real matrices A, $m \times n$, and B, $p \times n$ can be factorized as:

$$\begin{aligned} A &= U_A S_A X^{-1} \\ B &= U_B S_B X^{-1} \end{aligned}$$

All matrices are real. The matrix U_a is $m \times m$ orthonormal, U_B is $p \times p$ orthonormal, X is $n \times n$ nonsingular. With $r_{ab} = rank \begin{pmatrix} A \\ B \end{pmatrix}$, the matrices S_A $(m \times n)$ and S_B $(p \times n)$ have the following structure:

$$S_A = \begin{array}{c} \\ r_{ab}-r_b \\ r_a+r_b-r_{ab} \\ m-r_a \end{array} \begin{array}{c} \begin{array}{cccc} r_{ab}-r_b & r_a+r_b-r_{ab} & r_{ab}-r_a & n-r_{ab} \end{array} \\ \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$$S_B = \begin{array}{c} \\ p-r_b \\ r_a+r_b-r_{ab} \\ r_{ab}-r_a \end{array} \begin{array}{c} \begin{array}{cccc} r_{ab}-r_b & r_a+r_b-r_{ab} & r_{ab}-r_a & n-r_{ab} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & S & 0 & 0 \\ 0 & 0 & I & 0 \end{pmatrix} \end{array}$$

where C and S are $(r_a + r_b - r_{ab}) \times (r_a + r_b - r_{ab})$ diagonal matrices with positive diagonal elements, satisfying:

$$C^2 + S^2 = I_{r_a+r_b-r_{ab}}$$

and $r_a = rank(A), r_b = rank(B)$.

For some constructive proofs based upon several OSVDs, see e.g. [10], [16]. The name QSVD is proposed in [4].

- **While the structure of the PSVD and QSVD seems similar, their geometrical properties are completely different.**

- **We propose to call the pairs of nonzero elements of S_A and S_B in theorem 1, the *product singular values pairs* and their product *the product singular values.* Obviously, the pairs contain more structural information than the product singular values. There are 4 possibilities:**

  1. **There are r_1 pairs of the form $(\sqrt{\sigma_i}, \sqrt{\sigma_i})$ with corresponding product singular value $\sigma_i, i = 1, \ldots, r_1$. By convention, they are ordered such that $\sigma_i \geq \sigma_{i+1}$.**

  2. **There are $r_a - r_1$ pairs $(1, 0)$ with corresponding product singular value 0.**

3. There are $r_b - r_1$ pairs $(0,1)$ with corresponding product singular value 0.
4. 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]. The corresponding product singular values are undefined.

In the constructive proof of theorem 1, we shall need the following 4 lemmas:

Lemma 1

On the general solution of a consistent linear matrix equation

The set of solutions of the consistent matris equation

$$A\ X = B$$

is generated by

$$X = X_{part} + A^{\perp}T$$

where X_{part} *is a particular solution satisfying* $AX_p = B$, $A^{\perp}$ *is a matris of* maximal *rank such that* $AA^{\perp} = 0$ *and* T is *an arbitrary matris.*
In particular, let the OSVD *of* A *be given as:*

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

then

$$X = V_{a1} S_{a1}^{-1} U_{a1}^t B + V_{a2} T$$

is a solution for every matrix T.

- Observe that the lemma states that all solutions X can be written as the sum of a particular solution and the general solution to the homogeneous equation $AX = 0$.
- The first term is nothing else than $A^+ B$ where A^+ is the Moore-Penrose pseudo-inverse of A. It is also the unique minimum Frobenius norm solution. Recall that A^+ is the Moore-Penrose inverse of A if it is the unique solution $T = A^+$ *of:*

$$1.\quad ATA = A \tag{1}$$
$$2.\quad TAT = T \tag{2}$$
$$3.\quad (AT)' = AT \tag{3}$$
$$4.\quad (TA)' = TA \tag{4}$$

In section 3, we shall also use the notion of an **1-2-3**-inverse of the matrix A, which is any mat xix T satisfying (1)-(2)-(3).

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

Proof: Consider the following matrix identities:

$$\begin{pmatrix} AB^t & 0 \\ B^t & 0 \end{pmatrix} \begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix} = \begin{pmatrix} AB^t & AB^tA \\ B^t & B^tA \end{pmatrix}$$

and

$$\begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix} \begin{pmatrix} 0 & 0 \\ B^t & B^tA \end{pmatrix} = \begin{pmatrix} AB^t & AB^tA \\ B^t & B^tA \end{pmatrix}$$

Since the matrix

$$\begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix}$$

is nonsingular, we find that:

$$\begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix}^{-1} \begin{pmatrix} AB^t & 0 \\ B^t & 0 \end{pmatrix} \begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ B^t & B^tA \end{pmatrix}$$

Hence, the matrices

$$\begin{pmatrix} AB^t & 0 \\ B^t & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 \\ B^t & B^tA \end{pmatrix}$$

are similar. The first matrix has as its eigenvalues the eigenvalues of AB^t and n eigenvalues 0. The second matrix has as its eigenvalues the eigenvalues of B^tA and m eigenvalues 0. □

An immediate consequence of lemma 2 is the following:

Corollary 1 *Denote by $\lambda(.)$ the nonzero eigenvalue spectrum of a matrix. Then:*

$$\begin{aligned} \lambda(AB^tBA^t) &= \lambda(BA^tAB^t) \\ &= \lambda(A^tAB^tB) \\ &= \lambda(B^tBA^tA) \end{aligned}$$

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
PSVD of A, B **if** $AB^t = 0$
Let A, m x n and B, p x n be such that:

$$AB^t = 0$$

Assume that A and B have **OSVD***s:*

$$A = \begin{pmatrix} U_{a1} & U_{a2} \end{pmatrix} \begin{pmatrix} S_{a1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_{a1}^t \\ V_{a2}^t \end{pmatrix} \quad (5)$$

$$B = \begin{pmatrix} U_{b1} & U_{b2} \end{pmatrix} \begin{pmatrix} S_{b1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_{b1}^t \\ V_{b2}^t \end{pmatrix} \quad (6)$$

where S_{a1} *is* r_a x r_a (r_a *= rank(A)) and* S_{b1} *is* r_b x r_b (r_b *= 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:*

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

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

We have used 'a' PSVD instead of '*the*' PSVD because of the non-uniqueness 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 non-uniqueness 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:

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

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:

$$\begin{aligned} V_{a2}^t &= \begin{pmatrix} V_{b1}^t \\ V_{ab2}^t \end{pmatrix} \\ V_{b2}^t &= \begin{pmatrix} V_{a1}^t \\ V_{ab2}^t \end{pmatrix} \end{aligned}$$

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
The relation between the OSVD of AB^t and the eigenvalue decomposition of B^tBA^tA
Let the OSVD of AB^t be given as:

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

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

where S_1 is $r_1 \times r_1$ with $r_1 = rank(AB^t)$ and 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:*

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

where

- $D_2 = \begin{pmatrix} S_1^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
- $T_1 = V_{a2}^t B^t B A^t U_1 S_1^{-2}$
- *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 $\tilde{T}_4$ are arbitrary matrices that can be chosen to ensure that $rank(Y) = n$.*

Proof: First observe that 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 $r_1 = rank(AB^t)$ 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 a $n \times (r_a - r_1)$ matrix Y_2 and a $n \times (n - r_a)$ 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 $r_1 = rank(AB^t) \leq r_a = rank(A)$, it follows that:

$$R(U_1) \subset R(A)$$

so that

$$AA^+U_1 = U_1 \tag{10}$$

The matrix Y_1 will contain eigenvectors corresponding to S_1^2 if:

$$(B^tBA^tA)Y_1 = Y_1S_1^2 \tag{11}$$

Premultiply this expression with A:

$$(AB^tBA^t)AY_1 = AY_1S_1^2 \tag{12}$$

But from the OSVD (8) of AB' it follows then that we can put:

$$AY_1 = U_1$$

and using lemma 1 it follows that

$$Y_1 = A^+U_1 + V_{a2}T_1 \tag{13}$$

The matrix T_1 is however not arbitrary because it has to satisfy (11). Substituting (13) into (11) results in:

$$B^tBA^tA(A^+U_1 + V_{a2}T_1) = (A^+U_1 + V_{a2}T_1)S_1^2 \tag{14}$$

Premultiplying (14) with V_{a2}^t results in:

$$T_1 = V_{a2}^tB^tBA^tU_1S_1^{-2} \tag{15}$$

Hence we find that:

$$Y_1 = V_{a1}S_{a1}^{-1}U_{a1}^tU_1 + V_{a2}V_{a2}^tB^tBA^tU_1S_1^{-2} \tag{16}$$

Let us now verify that Y_1 as given by (16) satisfies (11). Hereto, first observe that from the OSVD of *AB'(8)* and the OSVD of *A (5)* it follows that :

$$V_{a1}^tB^tBA^tU_1 = S_{a1}^{-1}U_{a1}^tU_1S_1^2 \tag{17}$$

Together with the expression for T_1 (15), this implies the following identity:

$$\begin{pmatrix} V_{a1}^t \\ V_{a2}^t \end{pmatrix} B^tBA^tU_1 = \begin{pmatrix} V_{a1}^t \\ V_{a2}^t \end{pmatrix} (V_{a1}S_{a1}^{-1}U_{a1}^tU_1 + V_{a2}T_1)S_1^2 \tag{18}$$

But because $(V_{a1} \quad V_{a2})$ is nonsingular, it follows from (18) that:

$$\begin{aligned} B^tBA^tU_1 &= (A^+U_1 + V_{a2}T_1)S_1^2 & (19) \\ &= Y_1S_1^2 & (20) \end{aligned}$$

It can be verified from (16) and (10) that:

$$U_1 = AY_1 \tag{21}$$

Substitute this in (20) to find that:

$$B^tBA^tAY_1 = Y_1S_1^2 \tag{22}$$

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 $AA^+U_3 = U_3$. Let Y_2 be given as: $Y_2 = A^+U_3 + V_{a2}T_3$ where T_3 is an arbitrary matrix. Then:

$$\begin{aligned} B^tBA^tAY_2 &= B^tBA^tA(A^+U_3 + V_{a2}T_3) \\ &= B^tBA^tU_3 \\ &= B^tV_1SU_1^tU_3 \\ &= 0 \end{aligned}$$

Hence, the column vectors of Y_2 belong to the null space of B^tBA^tA and $rank(Y_2) = rank(U_3) = r_a - r_1$.

Proof for Y_3: Assume that $Y_3 = A^+U_4 + V_{a2}T_4 = V_{a2}T_4$. It follows that

$$B^tBA^tAY_3 = B^tBA^tAV_{a2}T_4 = 0$$

This implies that the column vectors of Y_3 belong to the null space of B^tBA^tA and obviously $rank(\&) = rank(\&) = n - r_a$, 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

$$Y = (Y_1 \; Y_2 \; Y_3) = (A^+ \; V_{a2}) \begin{pmatrix} U_1 & U_3 & U_4 \\ T_1 & T_3 & T_4 \end{pmatrix} \tag{23}$$

of full rank. Hereto, rewrite (23), using the OSVD of A (5) as:

$$Y = (V_{a1}S_{a1}^{-1} \; V_{a2}) \begin{pmatrix} U_{a1}^tU_1 & U_{a1}^tU_3 & U_{a1}^t U_4 \\ T_1 & T_3 & T_4 \end{pmatrix} \tag{24}$$

The matrix Y is now written as a product of 2 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:

$$(A^tAB^tB)X = XD_2 \tag{25}$$

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 3 steps:

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

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

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 $A'_{22}{B'_{22}}^t = 0$. 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 step 1, 2, 3 will then prove the theorem.

Step 1:

Combining the OSVD (8) of AB^t and the eigenvalue decomposition (9) results in:

$$\begin{aligned} B^tBA^tAY &= B^t(BA^t)AY \\ &= B^t(VD_1^tU^t)AY \\ &= YD_2 \end{aligned}$$

Premultiplying with A results in:

$$\begin{aligned} AB^t(VD_1^tU^t)AY &= AYD_2 \\ (UD_1V^t)(VD_1^tU^t)AY &= AYD_2 \\ (D_1D_1^t)(U^tAY) &= (U^tAY)D_2 \end{aligned}$$

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 $A' = U'AY$ and partition A' according to the block structure of D_1 and D_2 as:

$$A' = \begin{matrix} & r_1 & n-r_1 \\ r_1 & A'_{11} & A'_{12} \\ m-r_1 & A'_{21} & A'_{22} \end{matrix}$$

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:

$$\begin{aligned} S_1^2A'_{11} &= A'_{11}S_1^2 \\ A'_{12} &= 0 \\ A;, &= 0 \end{aligned}$$

Recall from lemma 4 that Y is nonsingular. Hence the matrix $A = UA'Y^{-1}$ can be written as:

$$A = U\begin{pmatrix} A'_{11} & 0 \\ 0 & A'_{22} \end{pmatrix}Y^{-1} \qquad (26)$$

Because U and Y are nonsingular matrices, we have that:

$$rank(A'_{11}) + rank(A'_{22}) = rank(A) \tag{27}$$

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

$$B = V \begin{pmatrix} B'_{11} & 0 \\ 0 & B'_{22} \end{pmatrix} Y^t \tag{28}$$

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

$$rank(B'_{11}) + rank(B'_{22}) = rank(B) \tag{29}$$

Step 2:

Carrying out the multiplication AB^t with the two factorizations (26) and (28) results in:

$$AB^t = U \begin{pmatrix} A'_{11}{B'_{11}}^t & 0 \\ 0 & A'_{22}{B'_{22}}^t \end{pmatrix} V^t \tag{30}$$

but from the uniqueness properties of the OSVD (8) it follows immediately that we can put:

$$A'_{11}{B'_{11}}^t = S_1 \tag{31}$$

Hence, we have that:

$$rank(A'_{11}) = rank(B'_{11}) = r_1$$

so that:

$${B'_{11}}^t = (A'_{11})^{-1}S_1 \tag{32}$$

When we require that $A'_{11} = B'_{11}$, one can always write a solution to (32) as:

$$A'_{11} = B'_{11} = S_1^{1/2} \tag{33}$$

In case that the elements of S_1 are distinct, this solution is unique. If some of the elements are coinciding, the solution is unique up to block diagonal orthonormal matrices that can however be incorporated into the orthonormal matrices U and V in the factorization of AB^t (30).

Step *3:*

It follows from the (non-)uniqueness properties of the OSVD in (30) and *(8)* that:

$$A'_{22}B'^{\,t}_{22} = 0 \qquad (34)$$

Moreover, from *(27)* and (29), it follows that:

$$\begin{aligned} rank(A'_{22}) &= rank(A) - r_1 = r_a - r_1 \\ rank(B'_{22}) &= rank(B) - r_1 = r_b - r_1 \end{aligned}$$

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 *(26)* and (28). □

2.2 A variational characterization and the relation with the QSVD

Note that, from theorem 1, lemma 4 and corollary 2, it follows that there are 4 eigenvalue decompositions that can be related to the PSVD:

$$\begin{aligned} (A^tAB^tB)X &= X(S_A^tS_AS_B^tS_B) \\ (B^tBA^tA)Y &= Y(S_B^tS_BS_A^tS_A) \\ (AB^tBA^t)U_A &= U_A(S_AS_B^tS_BS_A^t) \\ (BA^tAB^t)U_B &= U_B(S_BS_A^tS_AS_B^t) \end{aligned}$$

The last two of them are OSVDs.

Let us now derive a variational interpretation of the PSVD. Hereto, consider the optimization problem:

Maximize over all vectors x and y:

$$(y^tA^tAy)(x^tB^tBx) \qquad (35)$$

subject to

$$x^ty = 1 \qquad (36)$$

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:*

$$((y^k)^t A^t A y^k)((x^k)^t B^t B x^k) \tag{37}$$

subject to:

$$(x^k)^t y^k = 1 \tag{38}$$

$$(x^k)^t y^j = 0 \quad j = 1, \ldots, k-1 \tag{39}$$

$$(x^i)^t y^k = 0 \quad i = 1, \ldots, k-1 \tag{40}$$

It can be shown that the PSVD delivers the solution: The maximum of (35) 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 (37)-(40).

In order to derive a relation of the PSVD with the QSVD, we need the following lemma, relating a factorization of a matrix to its pseudo-inverse.

Lemma 5
Pseudo-inverse of a factorization
Let A of rank r_a be factorized as:

$$A = PSQ^t = (P_1 \; P_2) \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Q_1^t \\ Q_2^t \end{pmatrix}$$

where S_1 is r_a x r_a non-singular diagonal and P, Q, which are square non-singular, are partitioned conformally. Then:

$$A^+ = Q^{-*}\left(S^+ P^{-1} = (\right)^{-*} \begin{matrix} S^{-1} & 0 \\ 0 & 0 \end{matrix} P^{-1}$$

if and only if:

$$P_1^t P_2 = 0 \; \textit{and} \; Q_1^t Q_2 = 0$$

Proof: The proof follows immediately from substitution of the proposed factorization of A^+ into the relations (1)-(4). □

The lemma includes the special cases where A is of full column- and/or row rank, and the cases where P and/or Q is unitary.
We are now ready to establish the connection between the PSVD and the QSVD.

Lemma 6 *Let A, m x n and B, p x n have a* **PSVD** as in *theorem 1:*

$$\begin{aligned} A &= U_A S_A \begin{pmatrix} X_1^t \\ X_2^t \end{pmatrix} \\ B &= U_B S_B X^{-1} \end{aligned}$$

where the partitioning of X is *according to the zero-nonzero* diagonal struc*ture of S_A. Then (up* to a *reordening of* rows *of $(XD)^{-1}$ and columns of U_A and U_B and a corresponding reorganization of S_AD and S_BD), the* **QSVD** *of (A+)*; B is given by:*

$$\begin{aligned} (A^+)^* &= U_A((S_A^+)^t D)(XD)^{-1} \\ B &= U_B(S_B D)(XD)^{-1} \end{aligned}$$

if A is of full column rank or $X_1^t X_2 = 0$ where D is a non-singular diagonal *matrix given by:*

$$D = \begin{pmatrix} \frac{S_1^{1/2}}{\sqrt{I_{r_1}+S_1^2}} & 0 \\ 0 & I_{n-r_1} \end{pmatrix}$$

Proof: The proof is an immediate consequence of lemma 5. The matrix D is a diagonal scaling matrix, which ensures that the sum of squares of the diagonal elements equals 1 as required by theorem 2. □

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 non-uniqueness.

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 non-uniqueness, 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 non-uniqueness properties of the PSVD, and in particular, of a contragredient transformation for 2 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:

$$\begin{aligned} A^tA &= X(S_A^tS_A)X^t \\ B^tB &= X^{-t}(S_B^tS_B)X^{-1} \end{aligned}$$

or that:

$$\begin{aligned} X^{-1}A^tAX^{-t} &= (S_A^tS_A) \\ X^tB^tBX &= (S_B^tS_B) \end{aligned}$$

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* [14].

Definition 1 Contragredient transformation
The nonsingular $n \times n$ matrix T is a contragredient transformation for a pair of matrices F, G if:

$$\begin{aligned} T^{-1}FT^{-t} &= \textit{real diagonal} \\ T^{t}GT &= \textit{real diagonal} \end{aligned}$$

If both diagonal matrices are equal, we have:

Definition 2 Balancing contragredient transformation
A contragredient transformation T is called balancing if:

$$T^{-1}FT^{-t} = T^{t}GT = \textit{real diagonal}$$

Applications of (balancing) contragredient transformations can be found in system and control theory (open loop balancing of stable plants [6] [14] [15] and unstable systems [13] and closed loop balancing [12], model reduction [9] and H_∞ controller design [8]).

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 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 (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:

$$F = \begin{pmatrix} 2 & 2 \\ 2 & 1 \end{pmatrix} \quad G = \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix}$$

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 7

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 $F = L_F L_F^t$ and $G = L_G L_G^t$. Let $L_G^t L_F$ have singular value decomposition $L_G^t L_F = U\Sigma V^t$. Then $T = L_F V \Sigma^{-1/2}$ is a contragredient balancing transformation. Also $T^{-1} = \Sigma^{-1/2} U^t L_G^t$.

Proof: [14], 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 can not be balancing because F and G do not have the same inertia.

Lemma 8

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 $F = L_F L_F^t$ and $G = L_G L_G^t$ be a Cholesky-like factorization where L_G is $n \times r_G$ with $r_G = rank(G)$. 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: [14].

Observe that a contragredient transformation can only be unique up to a diagonal matrix, because if T is contragredient, 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 2 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, non-uniqueness for balancing contragredient transformations arises even in the distinct eigenvalue case, as is evident from the following example, borrowed from [14].

$$F = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad G = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Then

$$F\,G \; \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

has distinct eigenvalues at 1 and 0. But the transformation

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

is contragredient for any non-zero 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 casu $F = A^tA$ 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 non-uniqueness 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 $F = A'A$ and $G = B'B$ 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:

$$\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} U_{ab1} & U_{ab2} \end{pmatrix} \begin{pmatrix} S_{ab1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V^t_{ab1} \\ V^t_{ab2} \end{pmatrix}$$

where S_{ab1} is r_{ab} x r_{ab} diagonal and $r_{ab} = rank \begin{pmatrix} A \\ B \end{pmatrix}$. 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:

$$\begin{aligned} A_0 &= U_{A_0} S_{A_0} X_0^t \\ B_0 &= U_{B_0} S_{B_0} X_0^{-1} \end{aligned}$$

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:

$$A = U_{A_0} \begin{pmatrix} S_{A_0} & 0_{m\times(n-r_{ab})} \end{pmatrix} \begin{pmatrix} X_0^t & 0 \\ 0 & W_1^t \end{pmatrix} V_{ab}^t \quad (41)$$

$$B = U_{B_0} \begin{pmatrix} S_{B_0} & 0_{p\times(n-r_{ab})} \end{pmatrix} \begin{pmatrix} X_0^{-1} & 0 \\ 0 & W_1^{-1} \end{pmatrix} V_{ab}^t \quad (42)$$

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

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

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

Only in section 3.4 and 4 we shall 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

$$A = \begin{pmatrix} U_{a1} & U_{a2} \end{pmatrix} \begin{pmatrix} S_{a1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_{a1}^t \\ V_{a2}^t \end{pmatrix} \quad (43)$$

$$B = \begin{pmatrix} U_{b1} & U_{b2} \end{pmatrix} \begin{pmatrix} S_{b1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_{b1}^t \\ V_{b2}^t \end{pmatrix} \quad (44)$$

with $r_a = rank(A)$, $r_b = rank(B)$ 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:

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

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

$$S_{a1} V_{a1}^t V_{b1} S_{b1} = \begin{pmatrix} P_1 & P_2 \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Q_1^t \\ Q_2^t \end{pmatrix} \quad (45)$$

with $r_1 = rank(AB^t)$ and S_1 is $r_1 \times r_1$ diagonal with the non-zero 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:

$$AB^t = \begin{pmatrix} U_{a1}P_1 & U_{a1}P_2 & U_{a2} \end{pmatrix} \begin{pmatrix} S_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} Q_1^t U_{b1}^t \\ Q_2^t U_{b2}^t \\ U_{b2}^t \end{pmatrix} \quad (46)$$

Obviously, $r_1 \leq min(r_a, r_b)$. 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 (46) and the PSVD-theorem of section 2, introduce two non-singular $r_{ab} \times r_{ab}$ matrices X and Y and rewrite A and B as:

$$A = \begin{pmatrix} U_{a1}P_1 & U_{a1}P_2 & U_{a2} \end{pmatrix} \begin{pmatrix} S_1^{1/2} & 0 & 0 \\ 0 & I_{r_a - r_1} & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_1^t \\ X_2^t \\ X_3^t \end{pmatrix} \quad (47)$$

$$B = \begin{pmatrix} U_{b1}Q_1 & U_{b1}Q_2 & U_{b2} \end{pmatrix} \begin{pmatrix} S_1^{1/2} & 0 & 0 \\ 0 & 0 & I_{r_b - r_1} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} Y_1^t \\ Y_2^t \\ Y_3^t \end{pmatrix} \quad (48)$$

where Xi is $r_{ab} \times r_1$, X_2 is $r_{ab} \times (r_a - r_1)$, X_3 is $r_{ab} \times (r_{ab} - r_a)$ and 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:

$$X^t Y = \begin{pmatrix} X_1^t \\ X_2^t \\ X_3^t \end{pmatrix} \begin{pmatrix} Y_1 & Y_2 & Y_3 \end{pmatrix} = I_r \quad (49)$$

From the expressions (43) and (47) for A and (44) and (48) for B it is obvious that:

$$X_1^t = S_1^{-1/2} P_1^t S_{a1} V_{a1}^t \quad (50)$$

$$X_2^t = P_2^t S_{a1} V_{a1}^t \quad (51)$$

and

$$
\begin{aligned}
Y_1^t &= S_1^{-1/2} Q_1^t S_{b1} V_{b1}^t &\quad (52)\\
Y_3^t &= Q_2^t S_{b1}\, V_{b1}^t &\quad (53)
\end{aligned}
$$

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 :

$$
\begin{aligned}
X_1^t Y_1 &= I_{r_1} &\quad (54)\\
X_2^t Y_1 &= 0 &\quad (55)\\
X_1^t Y_3 &= 0 &\quad (56)\\
X_2^t Y_3 &= 0 &\quad (57)
\end{aligned}
$$

Because P_2 and Q_2, containing singular vectors corresponding to non-distinct zero singular values, are not unique, X_2 and Y_3 are non-unique. Hence, they are only determined up to orthonormal matrices W_2 and W_3 as:

$$
\begin{aligned}
X_2 &= V_{a1} S_{a1} P_2 W_2 &\quad (58)\\
Y_3 &= V_{b1} S_{b1} Q_2 W_3 &\quad (59)
\end{aligned}
$$

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 (47) and (48): If X_2^t (Y_3^t) is premultiplied there with 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 in detail non-uniqueness issues.

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^tY:

r_1	$r_{ab} - r_b$	$r_b - r_1$	
$X_1^tY_1$	$X_1^tY_2$	$X_1^tY_3$	r_1
$X_2^tY_1$	$X_2^tY_2$	$X_2^tY_3$	$r_a - r_1$
$X_3^tY_1$	$X_3^tY_2$	$X_3^tY_3$	$r_{ab} - r_a$

The requirement that this product must be equal to the identity matrix, imposes the following structure:

- **Since we know already that $X_2^tY_3 =$ [illegible] it follows that:**

$$r_{ab} - r_a \geq r_b - r_1$$

or

$$r_{ab} \geq (r_a + r_b - r_1) \tag{60}$$

This follows also from $X_2^tY_1 = 0$.

- **The lower $(r_b - r_1) \times (r_b - r_1)$ matrix of $X_3^tY_3$ is the identity matrix $I_{r_b - r_1}$.**
- **The left $(r_a - r_1)$ part of $X_2^tY_2$ equals $I_{r_a - r_1}$.**
- **The upper right corner of $X_3^tY_2$ equals $I_{r_{ab} - r_a - r_b + r_1}$.**

According to these requirements, the block structure is refined as:

$$X^tY = \begin{pmatrix} X_1^t \\ X_2^t \\ X_{31}^t \\ X_{32}^t \end{pmatrix} \begin{pmatrix} Y_1 & Y_{21} & Y_{22} & Y_3 \end{pmatrix} = I_{r_{ab}} \tag{61}$$

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 (47) and (48) (Recall that, for the time being, there is no common null space):

$$D_A = \begin{array}{c} \\ r_1 \\ r_a - r_1 \\ m - r_a \end{array} \begin{array}{c} \begin{array}{cccc} r_1 & r_a - r_1 & r_{ab} - r_a - r_b + r_1 & r_b - r_1 \end{array} \\ \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \quad (62)$$

$$D_B = \begin{array}{c} \\ r_1 \\ r_b - r_1 \\ p - r_b \end{array} \begin{array}{c} \begin{array}{cccc} r_1 & r_a - r_1 & r_{ab} - r_a - r_b + r_1 & r_b - r_1 \end{array} \\ \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 \\ 0 & 0 & 0 & I \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \quad (63)$$

It. follows from the refined block structure (61) that the matrices $X_{31}, X_{32}, Y_{21}, Y_{22}$ will be solutions to the following set of nonlinear matrix equations:

$$\begin{pmatrix} Y_1^t \\ Y_3^t \end{pmatrix} \begin{pmatrix} X_{31} & X_{32} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & I_{r_b - r_1} \end{pmatrix} \quad (64)$$

$$\begin{pmatrix} X_1^t \\ X_2^t \end{pmatrix} \begin{pmatrix} Y_{21} & Y_{22} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ I_{r_a - r_1} & 0 \end{pmatrix} \quad (65)$$

subject to the orthogonality constraints:

$$\begin{pmatrix} X_{31}^t \\ X_{32}^t \end{pmatrix} \begin{pmatrix} Y_{21} & Y_{22} \end{pmatrix} = \begin{pmatrix} 0 & I_{r - r_a - r_b + r_1} \\ 0 & 0 \end{pmatrix} \quad (66)$$

where the matrices Xi, X_2, Y_1, Y_3 are given by (50), (51), (52), (53).

A solution for the set of equations (64)-(66) will be obtained in the next section.

3.3 A solution to the set of nonlinear matrix equations.

In this section, we present a solution of the set of nonlinear matrix equations (64)-(66). 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 notations.
Recall the expressions (50)-(53) for X_1,X_2,Y_1,Y_3. Define the new matrices:

$$\begin{aligned}
\overline{X}_1 &= V_{a1}S_{a1}^{-1}P_1S_1^{1/2} && (67)\\
\overline{X}_2 &= V_{a1}S_{a1}^{-1}P_2 && (68)\\
\overline{Y}_1 &= V_{b1}S_{b1}^{-1}Q_1S_1^{1/2} && (69)\\
\overline{Y}_3 &= V_{b1}S_{b1}^{-1}Q_2 && (70)
\end{aligned}$$

Then, we have the following properties:

Lemma 9
Properties of $\overline{X}_1, \overline{X}_2, \overline{Y}_1, \overline{Y}_3$

- *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.*
- *They satisfy the following properties:*

$$\begin{aligned}
X_1^t\overline{X}_1 &= I_{r_1} && (71)\\
X_2^t\overline{X}_2 &= I_{r_a-r_1} && (72)\\
Y_1^t\overline{Y}_1 &= I_{r_1} && (73)\\
Y_3^t\overline{Y}_3 &= I_{r_b-r_1} && (74)
\end{aligned}$$

- *There are also the orthogonality relations:*

$$\begin{aligned}
X_1^t\overline{X}_2 &= 0 && (75)\\
X_2^t\overline{X}_1 &= 0 && (76)\\
Y_1^t\overline{Y}_3 &= 0 && (77)\\
Y_3^t\overline{Y}_1 &= 0 && (78)
\end{aligned}$$

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 (43), (44) and (45) 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 (43)-(45), properties (71)-(74) follow immediately. The ort hogonality relations (75)-(78) follow from the OSVD (45). □

We shall now show how a PSVD can be constructed from the OSVDs (43)-(45) and the **1-2-3**-inverses of X_1, X_2, Y_1, Y_3 as in (67)-(70).

Theorem 3
An explicit construction of the PSVD
Assume that A and B do not have a common null space and let their **OSVDs** *be:*

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

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

Define a 'weighted canonical correlation' **OSVD** as:

$$s_{a1} V_{a1}^t V_{b1} s_{b1} = \begin{pmatrix} P_1 & P_2 \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Q_1^t \\ Q_2^t \end{pmatrix}$$

and a canonical correlation **OSVD** *as:*

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

Furthermore, consider the **1-2-3**-*inverses* as in *(67)-(70).*
Then, a **PSVD** *of A and B* is given *by:*

$$A = \begin{pmatrix} U_{a1}P_1 & U_{a1}P_2 & U_{a2} \end{pmatrix} \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 \\ 0 & I_{r_a - r_1} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_1^t \\ X_2^t \\ X_{31}^t \\ X_{32}^t \end{pmatrix} \quad (79)$$

$$B = \begin{pmatrix} U_{b1}Q_1 & U_{b2}Q_2 & U_{b2} \end{pmatrix} \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{r_b - r_1} \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} Y_1^t \\ Y_{21}^t \\ Y_{22}^t \\ Y_3^t \end{pmatrix} \quad (80)$$

where the submatrices of X and Y are given by:

$$
\begin{aligned}
X_1 &= V_{a1}S_{a1}P_1S_1^{-1/2} \\
X_2 &= V_{a1}S_{a1}P_2 \\
X_{31} &= V_{b2}Q_3S_3^{-1/2} \\
X_{32} &= \overline{Y}_3 + V_{b2}V_{b2}^t(X_1(\overline{Y}_1^t\overline{Y}_1)^{-1}\overline{Y}_1^t\overline{Y}_3 + X_2W_5) \\
&= X_1(\overline{Y}_1^t\overline{Y}_1)^{-1}\overline{Y}_1^t\overline{Y}_3 + X_2W_5 + V_{a2}P_4P_4^tV_{a2}^t\overline{Y}_3
\end{aligned}
$$

$$
\begin{aligned}
Y_1 &= V_{b1}S_{b1}Q_1S_1^{-1/2} \\
Y_{21} &= \overline{X}_2 + V_{a2}V_{a2}^t(Y_1(\overline{X}_1^t\overline{X}_1)^{-1}\overline{X}_1^t\overline{X}_2 + Y_3W_6) \\
&= Y_1(\overline{X}_1^t\overline{X}_1)^{-1}\overline{X}_1^t\overline{X}_2 + Y_3W_6 + V_{b2}Q_4Q_4^tV_{b2}^t\overline{X}_2 \\
Y_{22} &= V_{a2}P_3S_3^{-1/2} \\
Y_3 &= V_{b1}S_{b1}Q_2
\end{aligned}
$$

*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:***

$$W_5^t + W_6 = \overline{Y}_3^t\overline{Y}_1(\overline{Y}_1^t\overline{Y}_1)^{-1}(\overline{X}_1^t\overline{X}_1)^{-1}\overline{X}_1^t\overline{X}_2 \tag{81}$$

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 9 and (81). □

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

3.4 Non-uniqueness properties of the PSVD

In case A and B do have a common null space, it is straightforward to combine the result of theorem 3 with the result of section 3.2.1.
A PSVD of any matrix pair A, B is given by:

$$A = (U_{A1}\ U_{A2}\ U_{A3}) \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 & 0 \\ 0 & I_{r_a - r_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_1^t \\ X_2^t \\ X_{31}^t \\ X_{32}^t \\ X_4^t \end{pmatrix} \tag{82}$$

$$B = (U_{B1}\, U_{B2}\, U_{B3}) \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{r_b - r_1} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} Y_1^t \\ Y_{21}^t \\ Y_{22}^t \\ Y_3^t \\ Y_4^t \end{pmatrix} \quad (83)$$

The matrices $U_{A1}, U_{A2}, U_{A3}, U_{B1}, U_{B2}, U_{B3}$ can be identified from (80) 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:

$$\begin{pmatrix} A \\ B \end{pmatrix} X_4 = \begin{pmatrix} A \\ B \end{pmatrix} Y_4 = 0 \qquad X_4^t Y_4 = I_{n-r_{ab}} \quad (84)$$

The question of non-uniqueness can now be analysed as follows:
Insert nonsingular square matrices R, T, W, Z into the above PSVD (83) as:

$$A = U_A W D_A T^t X^t \quad (85)$$
$$B = U_B Z D_B R^t Y^t \quad (86)$$

with appropriate partitionings of the matrices W, T, Z, R corresponding to the block structure of S_A and S_B.
This will correspond to another valid P SVD 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** $Z D_B R^t = D_B$.
- $$T R = I \quad (87)$$

Let us analyse these requirements in more detail:

- **From equations (50) and (52) it follows that X_1 and Y_1 are essentially unique (i.e. apart from (non-generic) non-uniqueness arising from non-distinct non-zero singular values in one of the OSVDs (43), (44) and (45)).**

- The non-uniqueness for X_2 and Y_3 is described in (58) and (59). 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 non-uniqueness of the choice of basis is characterized by the nonsingular matrix W_1 in (41) and (42).

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

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

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

where $T_{22} = W_2$ (see equation (58)) and $R_{44} = W_3$ (see equation (59)) are arbitrary but orthonormal.
Similarly, the matrices W and Z have the following structure:

$$W = \begin{array}{c} \\ r_1 \\ r_a - r_1 \\ m - r_a \end{array} \begin{array}{c} \begin{array}{ccc} r_1 & r_a - r_1 & m - r_a \end{array} \\ \begin{pmatrix} I_{r} & 0 & 0 \\ 0 & T_{22} & 0 \\ 0 & 0 & W_{33} \end{pmatrix} \end{array} \tag{88}$$

$$Z = \begin{array}{c} \\ r_1 \\ r_b - r_1 \\ p - r_b \end{array} \begin{array}{c} \begin{array}{ccc} r_1 & r_b - r_1 & p - r_b \end{array} \\ \begin{pmatrix} I_{r_1} & 0 & 0 \\ 0 & R_{44} & 0 \\ 0 & 0 & Z_{33} \end{pmatrix} \end{array} \tag{89}$$

where W_{33} and Z_{33} are arbitrary but orthonormal.
From condition (87), 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:

$$T = \begin{pmatrix} I & O & 0 & 0 & 0 \\ 0 & T_{22} & T_{23} & T_{24} & 0 \\ 0 & 0 & T_{33} & T_{34} & 0 \\ 0 & 0 & 0 & T_{44} & 0 \\ 0 & 0 & 0 & 0 & T_{55} \end{pmatrix} \quad (90)$$

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

$$R = \begin{pmatrix} I & 0 & 0 & 0 & 0 \\ 0 & T_{22} & 0 & 0 & 0 \\ 0 & -T_{33}^{-t}T_{23}^{t}T_{22}^{-t} & T_{33}^{-t} & 0 & 0 \\ 0 & -T_{44}^{-t}(T_{24}^{t} - T_{34}^{t}T_{33}^{-t}T_{23}^{t})T_{22}^{-t} & -T_{44}^{-t}T_{34}^{t}T_{33}^{-t} & T_{44} & 0 \\ 0 & 0 & 0 & 0 & T_{55}^{-t} \end{pmatrix} \quad (91)$$

The conclusion is summarized in the following:

Theorem *4*

On the non-uniqueness of the PSVD

If a **PSVD** *of A, B* **is** *given by:*

$$A = \begin{pmatrix} U_{A1} & U_{A2} & U_{A3} \end{pmatrix} \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_1^t \\ X_2^t \\ X_{31}^t \\ X_{32}^t \\ X_4^t \end{pmatrix} \quad (92)$$

$$B = \begin{pmatrix} U_{B1} & U_{B2} & U_{B3} \end{pmatrix} \begin{pmatrix} S_1^{1/2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} Y_1^t \\ Y_{21}^t \\ Y_{22}^t \\ Y_3^t \\ Y_4^t \end{pmatrix} \quad (93)$$

then the following **is** *also a* **PSVD:**

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

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

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* (87) *and are given in (91). The matrices W_{33} and Z_{33} are arbitraryorthonormal.*

In order to conclude this section, observe that we have characterized the non-uniqueness of the PSVD on a double level:

- **In theorem 3, we have derived an explicit construction of the PSVD from 4 OSVDs that could be obtained from the matrices *A* and *B*. Together with the observation of section 3.2.1 about a common null space, it became clear that the matrices X and Y are partitioned in 5 submatrices. Even here there is already some non-uniqueness parametrized by the matrices W_5 and W_6, which are arbitrary apart from the constraint (81).**
- **In theorem 4, it is shown that, once a PSVD is known with the corresponding partitioning in 5 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 (88) and (89). The mat rices *T* and *R* have the block triangular structure of (90) and (91). This block triangular structure will be important in the geometrical interpretation of the submatrices of X and Y in theorem 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 4, solve the set of nonlinear equations (64)-(66). Hence, theorem 4 also gives all solutions to this set of equations whereas theorem 3 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 $r_a = rank(A)$, $r_b = rank(B)$ and the OSVD of A and B be as in (43) and (44). let r_{ab} be defined as:

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

Then, it is well known that:

$$r_{ab} = r_a + r_b - dim(R(A^t) \bigcap R(B')) \tag{94}$$

Let r_1 be defined as in (45):

$$r_1 = rank(S_{a1} V_{a1}^t V_{b1} S_{b1}) = rank(V_{a1}^t V_{b1})$$

where the second equality follows from the nonsingularity of S_a and S_b. From the definition of angles between subspace as e.g. in [2], it follows immediately that r_1 is the number of canonical angles different from 90^0, between the row spaces of A and B:

$$r_1 = dim(\Pi_{A^t} R(B^t)) = dim(\Pi_{B^t} R(A^t)) \tag{95}$$

Hence $r_1 = 0$ only if 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 $r_{c2} = r_1 - r_{c1}$ others are not. Obviously:

$$r_{c1} = dim(R(A^t) \bigcap R(B^t))$$

Hence:

$$r_{ab} = r_a + r_b - r_{c1}$$

and

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

This is nothing else than inequality (60), which was derived from a structural requirement, whereas the derivation here is based on a geometrical argument.
Because r_{c2} is the number of non-zero canonical angles, different from 90^0, between the row spaces of A and B, it is also the number of non-zero canonical angles different from 90^0 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 4, it is straightforward to derive the following

generalized dyadic decomposition

$$A = U_{A1} S_1^{1/2} X_1^t + U_{A2} X_2^t \quad (96)$$

$$B = U_{B1} S_1^{1/2} Y_1^t + U_{B3} Y_3^t \quad (97)$$

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

From the fact that $X^tY = Y^tX = I_n$, it follows that:

$$A(Y_1 \; Y_{21} \; Y_{22} \; Y_3 \; Y_4) = (U_{A1} S_1^{1/2} \; U_{A2} \; 0 \; 0 \; 0) \quad (98)$$

$$B(X_1 \; X_2 \; X_{31} \; X_{32} \; X_4) = (U_{B1} S_1^{1/2} \; 0 \; 0 \; U_{B3} \; 0) \quad (99)$$

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 2 subspaces:

– $R(X_2)$ forms a subspace of $R(A^t)$, which is orthogonal to $R(B^t)$. It can be verified that

$$rank(X_2) = r_a - r_1 = \#\{\sigma(V_{b2}^t V_{a1}) = 1\} \qquad (100)$$

– $R(X_1)$ 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 *(95):*

$$r_1 = \#\{\sigma(V_{a1}^t V_{b1}) > 0\} \qquad (101)$$

- 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:

$$r_{ab} - r_a - r_b + r_1 = \#\{0 < \sigma(V_{a1}^t V_{b2}) < 1\} \qquad (102)$$

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

$$n - r_{ab} = \#\{\ \sigma(V_{a2}^t V_{b2}) = 1\} \qquad (103)$$

  Also, it follows immediately that:

$$X_4^t X_1 = 0 \qquad (104)$$
$$X_4^t X_2 = 0 \qquad (105)$$

- $R(B^t)$ is generated by the columns of Y_1 and Y_3. Hence, the row space of the matrix *B* can be split into 2 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').* It can be verified that:

$$rank(\&) = r_b - r_1 = \#\{\sigma(V_{a2}^t V_{b1}) = 1\} \qquad (106)$$

- *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:

$$r_{ab} - r_a - r_b + r_1 = \#\{0 < \sigma(V_{a2}^t V_{b1}) < 1\} \tag{107}$$

– $R(Y3)$ is orthogonal to $R(A^t)$ and also a subspace of $R(B^t)$.

– $R(Y_4)$ is the common null space of A and B. Hence:

$$Y_4^t Y_1 = 0 \tag{108}$$
$$Y_4^t Y_3 = 0 \tag{109}$$

Moreover:

$$R(X_4) = R(Y_4) \tag{110}$$

It can be verified that these geometrical results are independent of the non-uniqueness of the matrices X and Y as described in theorem *4.* The reason for this independency is precisely the block triangular structure of the matrices T *(90)* and R (91).
In order to appreciate this observation, compare the structure of the matrix X to 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 $X_{31} \rightarrow X_{31}T_{33}+X_2T_{23}$. It is easy to check from $Y^tX = I$,, 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 neither. 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 2 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. We have also investigated the connection with the QSVD and discussed several interesting properties and special cases.
Next, we have provided a detailed analysis of the structural and geometrical properties of the so called contragredient transformation of the *2* symmetric

matrices A^tA and B^tB, both of which are nonnegative and/or positive definite. A complete characterization and description of the non-uniqueness was obtained.
The geometry of the structure was interpreted in terms of principal angles between subspaces.

In a future publication, we shall show how the PSVD and the QSVD lie at the basis of an infinite number of generalizations of the OSVD. One of these, the RSVD, has already been analysed in detail in [3] and [18].

Acknowledgement

I would like to thank professor Gene Golub of the Department of Computer Science and professor *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] Arun K.S., Kung S.Y., Generalized principal component analysis and its application in approximate stochastic realization, in *Modelling and application of stochastic processes*, (U.B. Desai ed.), Kluwer Academic Publishers, 1986, pp.75-104.

[2] Bjorck A., Golub G., Numerical methods for computing angles between linear subspaces, *Mathematics of computation*, Vol.27, no.123, pp.579-594 (1973).

[3] De Moor B., Golub G.H., The restricted singular value decomposition: Properties and Applications, Numerical Analysis Project Report 89-03, Department of Computer Science, Stanford University, April 1989.

[4] De Moor B., Golub G. H., Generalized singular value decompositions: A proposal for a standardized nomenclature, Numerical Analysis Project Report 89-04, Department of Computer Science, Stanford University, April 1989.

[5] Deprettere E. (editor), *SVD and Signal Processing: Algorithms, Applications and Architectures,* North Holland, 1988.

[6] Fernando K.V., Hammarling S.J., A Product Induced Singular Value Decomposition for two matrices and balanced realisation, NAG Technical Report, TR8/87, 1987.

[7] Fernando K.V.,The Kalman Reachability/Observability Canonical Form and the IISVD, NAG Technical Report TR9/87, July 1987.

[8] Francis B. *A course in H_∞ control theory,* Lecture Notes in Control and Information Sciences (Eds. M. Thoma and A. Wynes), Springer Verlag, Berlin, 1987.

[9] Glover K., All optimal Hankel norm approximations of linear multivariable sys terns and their L_∞ error bounds, *Int. J. Control,* Vo1.39, no.6, pp.1115-1193 (1984).

[10] Golub G.H., Van Loan C.F., Matrix *Computations*, North Oxford Academic, Oxford, 1983.

[11] Heath MT., Laub A. J., Paige C.C., Ward R.C., Computing the singular value decomposition of a product of two matrices, *SIAM J. Sci. Stat. Comput.,* Vol.7, no.4, (1986).

[12] Jonckheere E., Silverman L., A new set of invariants for linear systems, applications to reduced order compensator design, *IEEE Trans. Aut. Control,* Vol.AC-28, no.1, (1983).

[13] Kenney C., Hewer G., Necessary and sufficient conditions for balancing unstable systems, *IEEE Trans. Aut. Control,* Vol.AC-32, no.2, (1987).

[14] Laub A. J., Heath M.T., Paige C.C., Ward R.C., Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms, *IEEE Trans. Aut. Control,* Vol.AC-32, no.2, (1987).

[15] Moore B.C., Principal component analysis in linear systems: controllability, observability and model reduction. *IEEE Trans. Aut. Control.,* Vol.AC-26, no.1, pp.17-32, (1981).

[16] Paige C.C., Saunders MA., Towards a generalized singular value decomposition, *SIAM J. Numer. Anal.*, 18, pp.398-405, (1981).

[17] Van Loan C.F., Generalizing the singular value decomposition, *SIAM J. Numer. Anal.*, 13, pp.76083, (1976).

[18] Zha H., Restricted SVD for matrix triplets and rank determination of matrices, Scientific Report 89-2, ZIB, Berlin.

Appendix: A solution of the nonlinear matrix equations that define the contragredient transformation

Observe that the linear equations (64)-(65) form an underdetermined set. With the factorizations of Xi, X_2, Y_1 and Y_3 (50)-(53) one can apply lemma 1 to obtain the general solution to the underdetermined equations as:

$$(X_{31} \; X_{32}) = V_{b1} \, S_{b1}^{-1} (\, Q_1 S_1^{-1/2} \;\; Q_2 \,) \begin{pmatrix} 0 & 0 \\ 0 & I_{r_b - r_1} \end{pmatrix} + V_{b2} (\; Z_1^x \; Z_2^x \;) \quad (111)$$

$$(Y_{21} \;\; Y_{22}) = V_{a1} S_{a1}^{-1} (\, P_1 S_1^{-1/2} \;\; P_2 \,) \begin{pmatrix} 0 & 0 \\ I_{r_a - r_1} & 0 \end{pmatrix} + V_{a2} (\; Z_1^y \; Z_2^y \;) \quad (112)$$

where $Z_1^x, Z_2^x, Z_1^y, Z_2^y$ are arbitrary matrices of appropriate dimensions. The first term in (111) and (112) is a particular solution while the second term is the general solution to the homogeneous equations obtained from (64) and (65).

The determination of X_{31}, X32, Y_{21} and Y_{22} reduces to the determination of $Z_1^x, Z_2^x, Z_1^y, Z_2^y$ in:

$$X_{31} = V_{b2} Z_1^x \quad (113)$$

$$X_{32} = V_{b1} S_{b1}^{-1} Q_2 + V_{b2} Z_2^x \quad (114)$$

$$Y_{21} = V_{a1} S_{a1}^{-1} P_2 + V_{a2} Z_1^y \quad (115)$$

$$Y_{22} = V_{a2} Z_2^y \quad (116)$$

subject to the conditions:

$$X_{31}^t Y_{21} = 0 \quad (117)$$

$$X_{32}^t Y_{21} = 0 \quad (118)$$

$$X_{32}^t Y_{22} = 0 \quad (119)$$

$$X_{31}^t Y_{22} = I_{r_{ab} - r_a - r_b + r_1} \quad (120)$$

Observe that this is a set of *non-linear* 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} (113) and Y_{22} (116) into the last constraint (120), results in:

$$(Z_1^x)^t V_{b2}^t V_{a2} Z_2^y = I_{r-r_a-r_b+r_1} \tag{121}$$

Since both V_{a2} and V_{b2} are orthonormal matrices, the OSVD of the product $V_{a2}^t V_{b2}$ corresponds to a canonical correlation analysis between the kernels of the matrices A and B. It can be shown that the number of non-zero singular values of $V_{a2}^t V_{b2}$ must be equal to $r_{ab} - r_a - r_b + r_1$ because the number of non-zero 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 $V_{a2}^t V_{b2}$:

$$V_{a2}^t V_{b2} = \begin{pmatrix} P_3 & P_4 \end{pmatrix} \begin{pmatrix} S_3 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Q_3^t \\ Q_4^t \end{pmatrix} \tag{122}$$

where S_3 is a $(r_{ab} - r_a - r_b + r_1)$ X $(r_{ab} - r_a - r_b + r_1)$ non-singular 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:

$$X_{31} = V_{b2} Q_3 S_3^{-1/2} \tag{123}$$
$$Y_{22} = V_{a2} P_3 S_3^{-1/2} \tag{124}$$

Observe that this is *not* the most general solution to (113)-(116)-(120) but only a specific one.

The determination of X32 and Y_{21}

Having determined expressions for X_{31} (123) and Y_{22} (124) from a canonical correlation analysis between the kernels of A and B, the orthogonality conditions (117)-(120) permit to derive two other equations for X32 and Y_{21}.

Hereto, first observe that from (43) and (44), and from (122), it follows that:

$$Q_3^t V_{b2}^t \begin{pmatrix} V_{b1} & V_{b2} Q_4 \end{pmatrix} = 0 \quad (125)$$
$$P_3^t V_{a2}^t \begin{pmatrix} V_{a1} & V_{a2} P_4 \end{pmatrix} = 0 \quad (126)$$

From equations (123) and (118) it follows that:

$$X_{31}^t Y_{21} = S_3^{-1/2} Q_3^t V_{b2}^t Y_{21} = 0 \quad (127)$$

while from (124) and (119) it follows that:

$$Y_{22}^t X_{32} = S_3^{-1/2} P_3^t V_{a2}^t X_{32} = 0 \quad (128)$$

The combination of equations (125) together with (127) permits to conclude via lemma 1 that there must exist matrices Z_3^y, Z_4^y of appropriate size, such that:

$$Y_{21} = V_{b1} Z_3^y + V_{b2} Q_4 Z_4^y \quad (129)$$

Similarly, it follows from (126) and (128) that:

$$X_{32} = V_{a1} Z_3^x + V_{a2} P_4 Z_4^x \quad (130)$$

Hence, there are 2 equations for X32, namely (114) and (130) and 2 equations for Y_{21}, (115) and (129). These are now repeated for convenience:

$$Y_{21} = V_{a1} S_{a1}^{-1} P_2 + V_{a2} Z_1^y \quad (131)$$
$$= V_{b1} Z_3^y + V_{b2} Q_4 Z_4^y \quad (132)$$

and

$$X_{32} = V_{b1} S_{b1}^{-1} Q_2 + V_{b2} Z_2^x \quad (133)$$
$$= V_{a1} Z_3^x + V_{a2} P_4 Z_4^x \quad (134)$$

From these 4 equations, we shall eliminate all unknown matrices in 4 steps:

Step 1: Elimination of Z_1^y and Z_4^y:
Recall the OSVD of $V_{a2}^t V_{b2}$ (122). Premultiplication of the expressions for Y_{21} (131)-(132)

- with V_{a2}^t results in:

$$Z_1^y = V_{a2}^t V_{b1} Z_3^y \tag{135}$$

- with $Q_4^t V_{b2}^t$ results in:

$$Z_4^y = Q_4^t V_{b2}^t V_{a1} S_{a1}^{-1} P_2 \tag{136}$$

Upon substitution in (131) and (132), this gives:

$$\begin{aligned} Y_{21} &= V_{a1} S_{a1}^{-1} P_2 + V_{a2} V_{a2}^t V_{b1} Z_3^y & (137) \\ &= V_{b1} Z_3^y + V_{b2} Q_4 Q_4^t V_{b2} V_{a1} S_{a1}^{-1} P_2 & (138) \end{aligned}$$

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

$$(I_{r_b} - V_{b1}^t V_{a2} V_{a2}^t V_{b1}) Z_3^y = V_{b1}^t V_{a1} S_{a1}^{-1} P_2$$

Observe. that the left hand side expression can be rewritten as:

$$\begin{aligned} (I_{r_b} - V_{b1}^t V_{a2} V_{a2}^t V_{b1}) &= V_{b1}^t (I_r - V_{a2} V_{a2}^t) V_{b1} \\ &= V_{b1}^t V_{a1} V_{a1}^t V_{b1} \end{aligned}$$

Hence, the equation for Z_3^y reads:

$$V_{b1}^t V_{a1} V_{a1}^t V_{b1} Z_3^y = V_{b1}^t V_{a1} S_{a1}^{-1} P_2 \tag{139}$$

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:

$$V_{a1}^t V_{b1} V_{b1}^t V_{a1} Z_3^x = V_{a1}^t V_{b1} S_{b1}^{-1} Q_2 \tag{140}$$

Step 3: A general solution for Z_3^x and Z_3^y

Rewrite equation (139) for Z_3^y, using the 0 SVD of $S_{a1} V_{a1}^t V_{b1} S_{b1} = P_1 S_1 Q_1^t$ (45), 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 9, this can be rewritten more compactly as:

$$(\overline{X}_1^t \overline{X}_1) \overline{Y}_1^t V_{b1} Z_3^y = \overline{X}_1^t \overline{X}_2 \tag{141}$$

The following observations are crucial:

1. The matrix $(\overline{X}_1^t \overline{X}_1)$ is square non-singular.
2. The columns of the matrix Y_3 are complementary to and orthogonal to the columns of the matrix $\overline{Y}_1$ (equation (77)).
3. Recall the relation $\overline{Y}_1^t Y_1 = I_{r_1}$ (equation (73)).

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

$$V_{b1} Z_3^y = Y_1(\overline{X}_1^t \overline{X}_1)^{-1} \overline{X}_1^t \overline{X}_2 + Y_3 W_6 \tag{142}$$

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 (140) as:

$$V_{a1} Z_3^x = X_1(\overline{Y}_1^t \overline{Y}_1)^{-1} \overline{Y}_1^t \overline{Y}_3 + X_2 W_5 \tag{143}$$

where W_5 is an arbitray $(r_a - r_1) \times (r_b - r_1)$ matrix.

However, as will now be shown, that matrices W_5 and W_6 are not independent of each other, because of the orthogonality condition $X_{32}^t Y_{21} = 0$ (118).
Hereto, we shall need the following properties:
Using the properties (71)-(78), itis straightforward to show from (142) and (143) that:

$$\begin{aligned} \overline{X}_2^t V_{a1} Z_3^x &= W_5 && (144) \\ \overline{Y}_3^t V_{b1} Z_3^y &= W_6 && (145) \end{aligned}$$

Also, from multiplying (142) with (143) and using the orthogonality conditions (75)-(78), it follows that:

$$(Z_3^x)^t V_{a1}^t V_{b1} Z_3^y = \overline{Y}_3^t \overline{Y}_1 (\overline{Y}_1^t \overline{Y}_1)^{-1} (\overline{X}_1^t \overline{X}_1)^{-1} \overline{X}_1^t \overline{X}_2 \tag{146}$$

Step 4: The remaining orthogonality condition
So far, we have obtained a general expression for $V_{a1} Z_3^x$ (143) and

$V_{b1}Z_3^y$ (142). The expressions for X_{32} (133)-(134) and Y_{21} (131)-(132) can be rewritten as:

$$\begin{aligned} X_{32} &= V_{a1}Z_3^x + (V_{a2}P_4)(P_4^tV_{a2}^t)\overline{Y}_3 & (147) \\ &= \overline{Y}_3 + V_{b2}V_{b2}^t(V_{a1}Z_3^x) & (148) \\ Y_{21} &= V_{b1}Z_3^y + (V_{b2}Q_4)(Q_4^tV_{b2}^t)\overline{X}_2 & (149) \\ &= \overline{X}_2 + V_{a2}V_{a2}^t(V_{b1}Z_3^y) & (150) \end{aligned}$$

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 ort hogonalit y requirement:

$$X_{32}^tY_{21} = 0$$

induces a constraint between W_5 and W_6. Hereto, we shall substitute the expressions for X32 and Y_{21} into the orthogonality condition:

Equation (147) x equation (149) results in:

$$(Z_3^x)^tV_{a1}^tV_{b1}Z_3^y+(Z_3^x)^tV_{a1}^t(V_{b2}Q_4)(Q_4^tV_{b2}^t)\overline{X}_2+\overline{Y}_3^t(V_{a2}P_4)(P_4^tV_{a2}^t)V_{b1}Z_3^y = 0 \quad (151)$$

Equation (148) x equation (149) results in:

$$\overline{Y}_3^tV_{b1}Z_3^y + (Z_3^x)^tV_{a1}^t(V_{b2}Q_4)(Q_4^tV_{b2}^t)\overline{X}_2 = 0 \quad (152)$$

Equation (147) × equation (150) results in:

$$(Z_3^x)^tV_{a1}^t\overline{X}_2 + \overline{Y}_3^t(V_{a2}P_4)(P_4^tV_{a2}^t)V_{b1}Z_3^y = 0 \quad (153)$$

Equations (152) and (153) permit to simplify equation (151) as:

$$(Z_3^x)^tV_{a1}^tV_{b1}Z_3^y - \overline{Y}_3^tV_{b1}Z_3^y - (Z_3^x)^tV_{a1}^t\overline{X}_2 = 0 \quad (154)$$

Now use equation (144) and (145) to get:

$$(Z_3^x)^tV_{a1}^tV_{b1}Z_3^y = W_5^t + W_6 \quad (155)$$

It follows then from equation (146) that:

$$W_5^t + W_6 = \overline{Y}_3^t \overline{Y}_1 (\overline{Y}_1^t \overline{Y}_1)^{-1} (\overline{X}_1^t \overline{X}_1)^{-1} \overline{X}_1^t \overline{X}_2 \tag{156}$$

This is the constraint between W_5 and W_6 that ensures the orthogonality between X32 and Y_{21}.

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

$$\begin{aligned} \overline{X}_1 x &= \overline{X}_2 \\ \overline{Y}_1 z &= \overline{Y}_3 \end{aligned}$$