THE CORE PROBLEM WITHIN A LINEAR APPROXIMATION PROBLEM $AX \approx B$ WITH MULTIPLE RIGHT-HAND SIDES[*]

IVETA HNĚTYNKOVÁ[†], MARTIN PLEŠINGER[‡], AND ZDENĚK STRAKOŠ[§]

Abstract. This paper focuses on total least squares (TLS) problems $AX \approx B$ with multiple right-hand sides. Existence and uniqueness of a TLS solution for such problems was analyzed in the paper [I. Hnětynková, M. Plešinger, D. M. Sima, Z. Strakoš, and S. Van Huffel (2011)]. For TLS problems with single right-hand sides the paper [C. C. Paige and Z. Strakoš (2006)] showed how necessary and sufficient information for solving $Ax \approx b$ can be revealed from the original data through the so-called core problem concept. In this paper we present a theoretical study extending this concept to problems with multiple right-hand sides. The data reduction we present here is based on the singular value decomposition of the system matrix A. We show minimality of the reduced problem; in this sense the situation is analogous to the single right-hand side case. Some other properties of the core problem, however, can not be extended to the case of multiple right-hand sides.

Key words. total least squares problem, multiple right-hand sides, core problem, linear approximation problem, error-in-variables modeling, orthogonal regression, singular value decomposition.

AMS subject classifications. 15A06, 15A18, 15A21, 15A24, 65F20, 65F25.

1. Introduction. Consider a linear approximation problem

$$AX \approx B, \qquad \text{or, equivalently,} \qquad [B|A]\begin{bmatrix} -I_d \\ X \end{bmatrix} \approx 0, \tag{1.1}$$

where $A \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^{n \times d}$, $B \in \mathbb{R}^{m \times d}$, without any further assumption on the positive integers m, n, d, and $A^T B \neq 0$ (this eliminates the trivial case where it does not make sense to approximate B by a linear combination of the columns of A; see also [14]). The equivalent $[B|A]$ form in (1.1) has been chosen so that our transformations will take it to block form (as for example in (2.1) below for the $d = 1$ case) revealing the core problem most simply. We will focus on incompatible problems, i.e. $\mathcal{R}(B) \not\subset \mathcal{R}(A)$. If $\mathcal{R}(B) \subset \mathcal{R}(A)$, then the system $AX = B$ can be solved using standard methods. Consider changes of the coordinate systems in $\mathbb{R}^m$, $\mathbb{R}^n$, and $\mathbb{R}^d$ represented by orthogonal transformations

$$\widehat{A}\widehat{X} \equiv (P^T A Q)(Q^T X R) \approx (P^T B R) \equiv \widehat{B}, \tag{1.2}$$

where $P^{-1} = P^T$, $Q^{-1} = Q^T$, $R^{-1} = R^T$; or, equivalently,

$$[\widehat{B}|\widehat{A}]\begin{bmatrix} -I_d \\ \widehat{X} \end{bmatrix} \equiv \left(P^T [B|A] \begin{bmatrix} R & 0 \\ 0 & Q \end{bmatrix} \right) \left(\begin{bmatrix} R^T & 0 \\ 0 & Q^T \end{bmatrix} \begin{bmatrix} -I_d \\ X \end{bmatrix} R \right) \approx 0. \tag{1.3}$$

[†]Charles University in Prague, Faculty of Mathematics and Physics, Czech Republic, and Institute of Computer Science, AS CR (e-mail: `hnetynkova@cs.cas.cz`).

[‡]Technical University of Liberec, Department of Mathematics, Czech Republic, and Institute of Computer Science, AS CR (e-mail: `martin.plesinger@tul.cz`).

[§]Charles University in Prague, Faculty of Mathematics and Physics, Czech Republic (e-mail: `strakos@karlin.mff.cuni.cz`).

We require that X solves (1.1) if and only if $\widehat{X} = Q^T X R$ solves (1.2) and call such problems *orthogonally invariant*. The *total least squares problem* (TLS)

$$\min_{X,E,G} \|[G|E]\|_F \qquad \text{subject to} \qquad (A+E)X = B+G \qquad (1.4)$$

serves as an important example; see [12], [13], [7], [5, section 6], [18]. Mathematically equivalent problems have been independently investigated under the names *orthogonal regression* and *errors-in-variables modeling*; see [20], [21].

In [8] it is shown that even with $d = 1$ (which gives $Ax \approx b$, where b is an m-vector) the TLS problem may not have a solution and, when the solution exists, it may not be unique; see also [6, pp. 324–326]. In order to resolve this difficulty, the classical book [19] introduces the so-called *nongeneric solution*. This book also extends the TLS theory to problems with multiple right-hand sides, i.e. for $d > 1$. The existence and uniqueness of a TLS solution with $d > 1$ is then discussed in full generality in the recent paper [10], giving a new classification of all possible cases.

The sequence of papers [12], [13], and [14] by C. C. Paige and Z. Strakoš investigates, using a unified framework, different least squares formulations for problems with $d = 1$. The last paper [14] introduced the so-called *core problem* that separates the necessary and sufficient information for solving the problem from the rest. It gives the necessary and sufficient condition for existence of the TLS solution, explains when the TLS solution exists and when it is unique, and clarifies the meaning of the nongeneric solution. For a brief summary see also the recent paper [10, pp. 752–753], which also shows that there is a class of problems for which the classical TLS approach described in [19], [22], [23], [9, Chapter 6.3, pp. 320–327] (in particular the so-called *classical TLS algorithm*; see [19, Chapter 3.6.1, pp. 87–90], [10, Algorithm 1, p. 767]) is unable to find the existing TLS solutions.

The first steps in generalizing the core problem theory for $d > 1$ were done by Å. Björck in the series of talks [1], [2], [3], and also in the unpublished manuscript [4], and by D. M. Sima and S. Van Huffel in [16], [17]. In a theoretical study presented in this paper we further develop the data reduction suggested in [15] which gives the core problem, we investigate its properties and prove its minimality. We do not advocate a computational technique for solving the problem. This is a matter of further work and results will be published elsewhere.

The organization of this paper is as follows. Section 2 recalls the core problem concept for a *single* right-hand side. Section 3 describes the data reduction for *multiple* right-hand sides, shows how to assemble the transformation matrices, and discusses basic properties of the reduced problem. Section 4 proves the minimality of the reduced problem, and thus justifies the definition of the core problem. Section 5 concludes the paper.

Throughout the paper, $\mathcal{R}(M)$ and $\mathcal{N}(M)$ denote the range and null space of a matrix M, respectively; I_ℓ (or just I) denotes an $\ell \times \ell$ identity matrix; and $0_{\ell,\xi}$ (or just 0) denotes an $\ell \times \xi$ zero matrix. The matrices A, B, $[B|A]$, and X from (1.1) are called the *system matrix*, the *right-hand sides (or the observation) matrix*, the *extended (or data) matrix*, and the *matrix of unknowns*, respectively.

2. Data reduction in the single right-hand side case. Consider the linear approximation problem (1.1) with $d = 1$. In [14] it was shown that there exist

orthogonal matrices P, Q, that transform the original problem into the block form

$$P^T[b|A] \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q^T \end{bmatrix} \begin{bmatrix} -1 \\ x \end{bmatrix} = \left[\begin{array}{c||c|c} b_1 & A_{11} & 0 \\ \hline 0 & 0 & A_{22} \end{array} \right] \begin{bmatrix} -1 \\ x_1 \\ x_2 \end{bmatrix} \approx 0, \quad (2.1)$$

where b_1 and A_{11} are of *minimal dimensions*. Such transformation can be obtained using the singular value decomposition (SVD) of the system matrix A,

$$A = U\Sigma V^T, \qquad U \in \mathbb{R}^{m \times m}, \quad \Sigma \in \mathbb{R}^{m \times n}, \quad V \in \mathbb{R}^{n \times n}, \tag{2.2}$$

where $U^{-1} = U^T$, $V^{-1} = V^T$. Let A have k *distinct* nonzero singular values

$$\sigma_1 > \sigma_2 > \ldots > \sigma_k > 0, \tag{2.3}$$

and let their multiplicities be m_j, $j = 1, \ldots, k$; $\sum_{j=1}^{k} m_j = r \equiv \text{rank}(A)$. Then

$$\Sigma = \text{diag}(\sigma_1 I_{m_1}, \ldots, \sigma_k I_{m_k}, 0_{m-r, n-r}). \tag{2.4}$$

Consider the partitioning $U = [U_1, \ldots, U_k, U_{k+1}]$, $U_j \in \mathbb{R}^{m \times m_j}$, $j = 1, \ldots, k$, and $U_{k+1} \in \mathbb{R}^{m \times m_{k+1}}$, where $m_{k+1} \equiv m - r$ is the dimension of the null space $\mathcal{N}(A^T)$. Columns of U_j represent an orthonormal basis of the jth left singular vector subspaces of A. Then

$$U^T[b|A] \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} = U^T[b|AV] \equiv [f|\Sigma], \tag{2.5}$$

where

$$f \equiv U^T b = [f_1^T, \ldots, f_k^T, f_{k+1}^T]^T, \quad f_j \equiv U_j^T b, \quad j = 1, \ldots, k+1,$$

and

$$\varphi_j \equiv \|f_j\| \geq 0. \tag{2.6}$$

Note that $\varphi_j = 0$ if and only if b is orthogonal to the jth left singular vector subspace. In order to be conformal with the multiple right-hand sides case, we keep (unlike in [14]) the zero and nonzero components f_j together until the last permutation. Let $S_j \in \mathbb{R}^{m_j \times m_j}$, $S_j^{-1} = S_j^T$, be a Householder reflection matrix such that

$$S_j^T f_j = e_1 \varphi_j, \quad e_1 = [1, 0, \ldots, 0]^T \in \mathbb{R}^{m_j}, \quad j = 1, \ldots, k+1, \tag{2.7}$$

and let

$$S_\oplus = \text{diag}(S_1, \ldots, S_k), \quad S_L = \text{diag}(S_\oplus, S_{k+1}), \quad S_R = \text{diag}(S_\oplus, I_{n-r}).$$

Note that $S_L^T \Sigma S_R = \Sigma$. The orthogonal transformation

$$(US_L)^T[b|A(VS_R)] = S_L^T[f|\Sigma S_R] = [S_L^T f|\Sigma]$$

maximizes the number of zero entries in the right-hand side vector

$$S_L^T f = [\varphi_1 e_1^T, \ldots, \varphi_k e_1^T, \varphi_{k+1} e_1^T]^T$$

(as mentioned above, we may have $\varphi_j = 0$ for some j). Let b have nonzero components f_{j_ℓ} in $\overline{n}$ left singular vector subspaces corresponding to nonzero singular values σ_j with indices $j_1, \ldots, j_{\overline{n}}$, $1 \leq \overline{n} \leq k$. The component f_{k+1} (the component of b in the null space of A^T) is nonzero due to the fact that the problem is incompatible. Consider the row permutation Π_L of the matrix $[S_L^T f | \Sigma]$ such that

$$\Pi_L^T S_L^T f = [b_1^T, 0]^T \equiv [\varphi_{j_1}, \ldots, \varphi_{j_{\overline{n}}}, \varphi_{k+1}, 0, \ldots, 0]^T,$$

i.e., all entries of

$$b_1 = [\varphi_{j_1}, \ldots, \varphi_{j_{\overline{n}}}, \varphi_{k+1}]^T = [\|f_{j_1}\|, \ldots, \|f_{j_{\overline{n}}}\|, \|f_{k+1}\|]^T \in \mathbb{R}^{\overline{n}+1} \qquad (2.8)$$

are positive. Then there exists a column permutation Π_R of the matrix $\Pi_L^T \Sigma$ such that

$$\Pi_L^T \Sigma \Pi_R = \left[\begin{array}{c|c} A_{11} & 0 \\ \hline 0 & A_{22} \end{array} \right],$$

where the block $A_{11} \in \mathbb{R}^{(\overline{n}+1) \times \overline{n}}$ is (with $b \notin \mathcal{R}(A)$) rectangular, containing *at most one* copy of each nonzero singular value σ_j on its diagonal and having the zero last row. All the other singular values are moved to the diagonal of the second block A_{22}, which can be of any shape, or nonexistent. Summarizing, we obtain

$$P^T [b | AQ] = \left[\begin{array}{c||c|c} b_1 & A_{11} & 0 \\ \hline 0 & 0 & A_{22} \end{array} \right], \quad P \equiv U S_L \Pi_L, \quad Q \equiv V S_R \Pi_R,$$

where $[b_1 | A_{11}]$ and A_{11} are of minimal dimensions; see [14]. The corresponding transformation and conformal splitting of vector of unknowns is

$$Q^T x = (V S_R \Pi_R)^T x = \left[\begin{array}{c} x_1 \\ x_2 \end{array} \right].$$

The subproblem $[b_1 | A_{11}]$, or $A_{11} x_1 \approx b_1$, contains the necessary and sufficient information for solving the original problem $Ax \approx b$, and it is called the *core problem*. The solution of the second subproblem $A_{22} x_2 \approx 0$ with the maximally dimensioned block $A_{22} \in \mathbb{R}^{(m-\overline{n}-1) \times (n-\overline{n})}$ can be considered to be $x_2 = 0$ (see the discussion in [14]) giving

$$x = Q \left[\begin{array}{c} x_1 \\ x_2 \end{array} \right] = V S_R \Pi_R \left[\begin{array}{c} x_1 \\ 0 \end{array} \right]. \qquad (2.9)$$

The core problem (in a different form) can also be revealed by the Golub-Kahan bidiagonalization; see [14, Section 3] and [11].

For $d = 1$ the core problem has a unique TLS solution, see [14]. Using (2.9), the core problem defines the minimum norm TLS solution if it exists, or the minimum norm nongeneric solution of (1.1); see [14] and [10, section 3.1, pp. 752–753]. Computationally (assuming exact arithmetic) the solution (2.9) is for $d = 1$ therefore identical to the output of the classical TLS algorithm given in [19, Chapter 3.6.1, pp. 87–90]; see also [10, Algorithm 1, p. 767]. Unlike in the classical TLS algorithm, the solution (2.9) is constructed in a straightforward way by separating the TLS-meaningful part $A_{11} x_1 \approx b_1$ of the problem $Ax \approx b$ from the redundant and irrelevant part represented by $A_{22} x_2 \approx 0$, $x_2 = 0$. In the rest of this paper it will be shown how to generalize the SVD-based data reduction and obtain the core problem for $d > 1$, i.e., for the problem $AX \approx B$.

3. Data reduction in the multiple right-hand side case. Consider the problem (1.1) with $d > 1$. In this section, we construct orthogonal matrices P, Q, R, that transform the original data matrix $[B|A]$ into the block form

$$P^T[B|A]\begin{bmatrix} R & 0 \\ 0 & Q \end{bmatrix} = [P^T BR | P^T AQ] = \left[\begin{array}{c|c||c|c} B_1 & 0 & A_{11} & 0 \\ \hline 0 & 0 & 0 & A_{22} \end{array}\right], \qquad (3.1)$$

where B_1 and A_{11} are of minimal dimensions (the proof of minimality will be given in section 4). The orthogonal transformation (3.1) is done in four successive steps: preprocessing of the right-hand side B (section 3.1), transformation of the system matrix A (SVD of A) (section 3.2), transformation of the right-hand side B (section 3.3), and final permutation (section 3.4).

3.1. Preprocessing of the right-hand side. Let $\overline{d} \equiv \text{rank}(B) \le \min\{m, d\}$. Consider the SVD of B in the form

$$B = S\Theta R^T, \qquad S \in \mathbb{R}^{m \times \overline{d}}, \quad \Theta \in \mathbb{R}^{\overline{d} \times d}, \quad R \in \mathbb{R}^{d \times d}, \qquad (3.2)$$

where S has mutually orthonormal columns, i.e. $S^T S = I_{\overline{d}}$, Θ is of *full row rank*, and R is square, i.e. $R^{-1} = R^T$. (Note that the schema illustrates the case $m > d > \overline{d}$; other cases can be illustrated analogously.) We will see later that this R plays the role of the transformation matrix R in (3.1). If $\overline{d} < d$, then B contains linearly dependent columns representing redundant information that can be removed from the original problem (1.1). Multiplication of (1.1) from the right by R gives

$$A(XR) \approx BR, \qquad (3.3)$$

where

$$BR = S\Theta \equiv [C, 0] \in \mathbb{R}^{m \times d}, \qquad C \in \mathbb{R}^{m \times \overline{d}}, \qquad \text{and}$$
$$XR \equiv [Y, Y'] \in \mathbb{R}^{n \times d}, \qquad Y \in \mathbb{R}^{n \times \overline{d}}; \qquad (3.4)$$

if $d = \overline{d}$, then $BR = C$, $XR = Y$. With this notation

$$[BR|A]\begin{bmatrix} -I_d \\ XR \end{bmatrix} = [C, 0|A]\begin{bmatrix} -I_{\overline{d}} & 0 \\ 0 & -I_{d-\overline{d}} \\ Y & Y' \end{bmatrix} = [AY - C|AY'] \approx 0. \qquad (3.5)$$

The original problem (1.1) is in this way split into two subproblems

$$AY \approx C \qquad \text{and} \qquad AY' \approx 0, \qquad (3.6)$$

where the second problem is homogeneous. Following the arguments in [14], we consider the meaningful solution $Y' \equiv 0$. In this way, the approximation problem (1.1) reduces to

$$AY \approx C, \qquad \text{or, equivalently,} \qquad [C|A]\begin{bmatrix} -I_{\overline{d}} \\ Y \end{bmatrix} \approx 0, \qquad (3.7)$$

where $A \in \mathbb{R}^{m \times n}$, $Y \in \mathbb{R}^{n \times \overline{d}}$, and $C \in \mathbb{R}^{m \times \overline{d}}$ is of *full column rank*. From (3.2)–(3.4) it follows that the right-hand side matrix C has *mutually orthogonal columns*.

3.2. Transformation of the system matrix. Consider the SVD of A given by (2.2) with (2.3) and (2.4). The problem (3.7) can then be transformed analogously to (2.5),

$$(U^T A V)(V^T Y) = \Sigma Z \approx F, \tag{3.8}$$

where $Z = V^T Y$, $F = U^T C$. Equivalently,

$$[F|\Sigma] \begin{bmatrix} -I_{\overline{d}} \\ Z \end{bmatrix} \approx 0. \tag{3.9}$$

The approximation problem $\Sigma Z \approx F$ has the full column rank right-hand side matrix F with mutually orthogonal columns and the system matrix Σ in a diagonal form.

3.3. Transformation of the right-hand side. Similarly to the single right-hand side case, we now transform the right-hand side matrix of (3.8) in order to get as many zero rows as possible. Consider a partitioning of F into the block-rows with respect to the multiplicities of the singular values of the system matrix A, i.e.

$$F = [F_1^T, \ldots, F_k^T, F_{k+1}^T]^T, \qquad \text{where} \qquad F_j \in \mathbb{R}^{m_j \times \overline{d}}, \qquad j = 1, \ldots, k, k+1.$$

Let $r_j \equiv \operatorname{rank}(F_j) \le \min\{m_j, \overline{d}\}$. Consider the SVD of F_j in the form

$$F_j = S_j \Theta_j W_j^T, \qquad S_j \in \mathbb{R}^{m_j \times m_j}, \quad \Theta_j \in \mathbb{R}^{m_j \times r_j}, \quad W_j \in \mathbb{R}^{\overline{d} \times r_j}, \qquad (3.10)$$
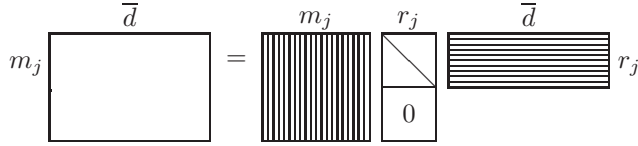
where S_j is square, i.e. $S_j^{-1} = S_j^T$, Θ_j is of *full column rank*, and W_j has mutually orthonormal columns, i.e. $W_j^T W_j = I_{r_j}$, $j = 1, \ldots, k, k+1$. (Note that, as above, the schema illustrates the case $r_j < m_j < \overline{d}$.) The matrix S_j generalizes the role of the identically denoted matrix in section 2; see (2.7). Consider the block diagonal orthogonal matrices

$$S_\oplus = \operatorname{diag}(S_1, \ldots, S_k), \quad S_L = \operatorname{diag}(S_\oplus, S_{k+1}), \quad S_R = \operatorname{diag}(S_\oplus, I_{n-r}), \quad (3.11)$$

where as in (2.4) $r = \operatorname{rank}(A)$, and recall that

$$\Sigma = \operatorname{diag}(\sigma_1 I_{m_1}, \ldots, \sigma_k I_{m_k}, 0_{m-r, n-r}) \in \mathbb{R}^{m \times n}, \qquad \sigma_1 > \sigma_2 > \ldots > \sigma_k > 0,$$

see (2.3), (2.4). Then $\Sigma Z \approx F$ from (3.8) can be transformed to

$$(S_L^T \Sigma S_R)(S_R^T Z) \approx S_L^T F \qquad \text{and} \qquad S_L^T \Sigma S_R = \Sigma. \tag{3.12}$$

Equivalently, (3.9) becomes

$$[S_L^T F|\Sigma] \begin{bmatrix} -I_{\overline{d}} \\ S_R^T Z \end{bmatrix} \approx 0,$$

and the extended (data) matrix has the form

$$[S_L^T F|\Sigma] = \begin{bmatrix} \Theta_1 W_1^T & \sigma_1 I_{m_1} & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ \Theta_k W_k^T & 0 & & \sigma_k I_{m_k} & 0 \\ \Theta_{k+1} W_{k+1}^T & 0 & \cdots & 0 & 0 \end{bmatrix} \in \mathbb{R}^{m \times (n+\overline{d})}. \tag{3.13}$$

If $m_j > r_j$, then the block $S_j^T F_j = \Theta_j W_j^T$ contains zero rows at the bottom, see (3.10). Denote

$$\Theta_j W_j^T \equiv \begin{bmatrix} \Phi_j \\ 0 \end{bmatrix}, \qquad \Phi_j \in \mathbb{R}^{r_j \times \overline{d}}, \tag{3.14}$$

where Φ_j is the block of nonzero rows (if $r_j = 0$, then the block Φ_j has *no rows*). For $r_j = m_j$ we simply have $\Theta_j W_j^T \equiv \Phi_j$. It follows from (3.10) that Φ_j has *mutually orthogonal rows*. The matrix Φ_j generalizes the role of the number φ_j; see (2.6), (2.8).

3.4. Final permutation. Now the aim is to find a permutation of (3.13) that reveals the block diagonal structure (3.1). This can be done analogously to the single right-hand side case (see also [14, Section 2]), by moving the rows of (3.13) with zero blocks in $\Theta_j W_j^T$ (see (3.14)) to the bottom submatrix of the whole matrix, see the matrix in the middle row of (3.15), with subsequent moving of the corresponding columns with the diagonal blocks $\sigma_j I_{m_j - r_j}$ in the bottom to the right,

$$\Pi_L^T \begin{bmatrix} \Theta_1 W_1^T & \sigma_1 I_{m_1} & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ \Theta_k W_k^T & 0 & & \sigma_k I_{m_k} & 0 \\ \Theta_{k+1} W_{k+1}^T & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} I_{\overline{d}} & 0 \\ 0 & \Pi_R \end{bmatrix}$$

$$= \left[\begin{array}{c||cccc|cccc} \Phi_1 & \sigma_1 I_{r_1} & & 0 & 0 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots & & \vdots & \vdots \\ \Phi_k & 0 & & \sigma_k I_{r_k} & 0 & \cdots & 0 & 0 \\ \Phi_{k+1} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \hline 0 & 0 & \cdots & 0 & \sigma_1 I_{m_1 - r_1} & & 0 & 0 \\ \vdots & \vdots & & \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & & \sigma_k I_{m_k - r_k} & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \end{array} \right]$$

$$\equiv \left[\begin{array}{c||c|c} B_1 & A_{11} & 0 \\ \hline 0 & 0 & A_{22} \end{array} \right]. \tag{3.15}$$

Here $\Pi_L \in \mathbb{R}^{m \times m}$ is given by

$$\Pi_L \equiv \left[\begin{array}{cccc|cccc} \begin{bmatrix} I_{r_1} \\ 0 \end{bmatrix} & & 0 & 0 & \begin{bmatrix} 0 \\ I_{m_1 - r_1} \end{bmatrix} & & 0 & 0 \\ & \ddots & & \vdots & & \ddots & & \vdots \\ 0 & & \begin{bmatrix} I_{r_k} \\ 0 \end{bmatrix} & 0 & 0 & & \begin{bmatrix} 0 \\ I_{m_k - r_k} \end{bmatrix} & 0 \\ 0 & \cdots & 0 & \begin{bmatrix} I_{r_{k+1}} \\ 0 \end{bmatrix} & 0 & \cdots & 0 & \begin{bmatrix} 0 \\ I_{m_{k+1} - r_{k+1}} \end{bmatrix} \end{array} \right] \tag{3.16}$$

and it permutes the block-rows starting with Φ_j up, while moving the block-rows starting with zero blocks down. Analogously, $\Pi_R \in \mathbb{R}^{n \times n}$ given by

$$\Pi_R \equiv \left[\begin{array}{ccc|ccc} \begin{bmatrix} I_{r_1} \\ 0 \end{bmatrix} & & 0 & \begin{bmatrix} 0 \\ I_{m_1 - r_1} \end{bmatrix} & & 0 & 0 \\ & \ddots & & & \ddots & & \vdots \\ 0 & & \begin{bmatrix} I_{r_k} \\ 0 \end{bmatrix} & 0 & & \begin{bmatrix} 0 \\ I_{m_k - r_k} \end{bmatrix} & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & I_{n-r} \end{array} \right] \tag{3.17}$$

rearranges the block-columns of the system matrix. Note that if for some j we have $r_j = m_j$, then the block $I_{m_j - r_j}$ and the corresponding block-rows and block-columns vanish; if $r_j = 0$, then the block I_{r_j} and the corresponding block-rows and block-columns vanish. Let us briefly summarize the whole transformation.

3.5. Summary of the transformation. Using the SVD $B = S\Theta R^T$, defined in (3.2), the original approximation problem (1.1)

$$AX \approx B, \qquad A \in \mathbb{R}^{m \times n}, \quad X \in \mathbb{R}^{n \times d}, \quad B \in \mathbb{R}^{m \times d},$$

is transformed to

$$AY \approx C, \qquad A \in \mathbb{R}^{m \times n}, \quad Y \in \mathbb{R}^{n \times \overline{d}}, \quad C \in \mathbb{R}^{m \times \overline{d}},$$

with *the full column rank right-hand side*. Then using the SVD $A = U\Sigma V^T$ defined in (2.2)–(2.4), the problem is further transformed to

$$\Sigma Z \approx F, \qquad \Sigma \in \mathbb{R}^{m \times n}, \quad Z \in \mathbb{R}^{n \times \overline{d}}, \quad F \in \mathbb{R}^{m \times \overline{d}},$$

with *the diagonal system matrix* Σ. Using singular value decompositions $F_j = S_j \Theta_j W_j^T$ of block-rows of F, see (3.10), and the orthogonal matrix $S_\oplus$ given by (3.11), the right-hand side matrix gets the structure with the full row rank block-rows Φ_j and the zero block-rows, while the diagonal system matrix Σ stays unchanged. Finally, the permutation matrices Π_L, Π_R given by (3.16) and (3.17) are used to collect the full row rank blocks, and to transform the system matrix to the block diagonal form with two diagonal (in general rectangular) blocks, see (3.15).

We give a quick summary of the mathematical transformations, each preceded by the relevant equation numbers:

$$
\begin{aligned}
\text{(1.1), (3.2):} \qquad & AX \approx B = S\Theta R^T = [C,\ 0]R^T, \\[4pt]
\text{(2.2), (3.4)–(3.8):} \qquad & \left\{ \begin{aligned} (U^T AV)(V^T XR) &\approx U^T BR, \\ \Sigma(V^T XR) &\approx [U^T C,\ 0] = [F,\ 0], \end{aligned} \right. \\[4pt]
\text{(3.10)–(3.14):} \qquad & \left\{ \begin{aligned} (S_L^T U^T AV S_R)(S_R^T V^T XR) &\approx S_L^T U^T BR, \\ \Sigma(S_R^T V^T XR) &\approx [S_L^T F,\ 0], \end{aligned} \right. \\[4pt]
\text{(3.15)–(3.17):} \qquad & \left\{ \begin{aligned} (\Pi_L^T S_L^T U^T AV S_R \Pi_R)(\Pi_R^T S_R^T V^T XR) &\approx \Pi_L^T S_L^T U^T BR, \\ (\Pi_L^T \Sigma \Pi_R)(\Pi_R^T S_R^T V^T XR) &\approx [\Pi_L^T S_L^T F,\ 0], \\ \operatorname{diag}(A_{11}, A_{22})(\Pi_R^T S_R^T V^T XR) &\approx \operatorname{diag}(B_1, 0), \end{aligned} \right.
\end{aligned}
$$

so that the full transformations of A, X, and B can be summarized as

$$(P^T AQ)(Q^T XR) \approx P^T BR, \qquad P \equiv US_L \Pi_L, \qquad Q \equiv VS_R \Pi_R. \tag{3.18}$$

Clearly $P^{-1} = P^T$, $Q^{-1} = Q^T$, $R^{-1} = R^T$, and

$$P^T[B|A]\begin{bmatrix} R & 0 \\ 0 & Q \end{bmatrix} = \left[\begin{array}{c|c||c|c} B_1 & 0 & A_{11} & 0 \\ \hline 0 & 0 & 0 & A_{22} \end{array} \right] \begin{array}{l} \} \overline{m} \\ \} m - \overline{m} \end{array} \tag{3.19}$$

$$\underbrace{}_{\overline{d}} \underbrace{}_{d - \overline{d}} \underbrace{\phantom{A_{11}}}_{\overline{n}} \underbrace{\phantom{A_{22}}}_{n - \overline{n}}$$

is of the form (3.1), where from (3.15)

$$[B_1 | A_{11}] \equiv \left[\begin{array}{c|ccc} \Phi_1 & \sigma_1 I_{r_1} & & 0 \\ \vdots & & \ddots & \\ \Phi_k & 0 & & \sigma_k I_{r_k} \\ \Phi_{k+1} & 0 & \cdots & 0 \end{array} \right] \in \mathbb{R}^{\overline{m} \times (\overline{n} + \overline{d})}, \qquad (3.20)$$

and $\overline{m} \equiv \sum_{j=1}^{k+1} r_j$, $\overline{n} \equiv \sum_{j=1}^{k} r_j$, $\overline{d} \equiv \mathrm{rank}(B)$. The block A_{22} has the form

$$A_{22} \equiv \mathrm{diag}(\sigma_1 I_{m_1 - r_1}, \ldots, \sigma_k I_{m_k - r_k}, 0_{m-r-r_{k+1}, n-r}). \qquad (3.21)$$

Thus the original problem $AX \approx B$ is transformed into the block form

$$\left[\begin{array}{c|c} A_{11} & 0 \\ \hline 0 & A_{22} \end{array} \right] (Q^T X R) \approx \left[\begin{array}{c|c} B_1 & 0 \\ \hline 0 & 0 \end{array} \right],$$

compare with (1.2). Using a conformal partitioning of the matrix of unknowns

$$Q^T X R = \left[\begin{array}{cc} X_1 & X_1' \\ X_2 & X_2' \end{array} \right] \begin{array}{c} \} \, \overline{n} \\ \} \, n - \overline{n} \end{array}, \qquad (3.22)$$
$$\underbrace{}_{\overline{d}} \underbrace{}_{d - \overline{d}}$$

where $Q \left[\begin{smallmatrix} X_1 \\ X_2 \end{smallmatrix} \right] = Y$, $Q \left[\begin{smallmatrix} X_1' \\ X_2' \end{smallmatrix} \right] = Y'$, and $[Y, Y'] = XR$, see (3.4), the block diagonal structure of the system matrix and the right-hand sides matrix allows us to split the original problem $AX \approx B$ into (generally four) subproblems

$$A_{11} X_1 \approx B_1, \qquad \text{and} \qquad A_{22} X_2 \approx 0, \quad A_{11} X_1' \approx 0, \quad A_{22} X_2' \approx 0.$$

The last three subproblems are homogenous and we consider, following the arguments in [14], $X_2 \equiv 0$, $X_1' \equiv 0$, $X_2' \equiv 0$. Only the subproblem

$$A_{11} X_1 \approx B_1, \qquad \text{or, equivalently,} \qquad [B_1 | A_{11}] \left[\begin{array}{c} -I_{\overline{d}} \\ X_1 \end{array} \right] \approx 0, \qquad (3.23)$$

where $A_{11} \in \mathbb{R}^{\overline{m} \times \overline{n}}$, $X_1 \in \mathbb{R}^{\overline{n} \times \overline{d}}$, $B_1 \in \mathbb{R}^{\overline{m} \times \overline{d}}$, has to be solved. If its solution is X_1, the solution X of the original problem $AX \approx B$ is then

$$X \equiv Q \left[\begin{array}{cc} X_1 & 0 \\ 0 & 0 \end{array} \right] R^T. \qquad (3.24)$$

The dimensions of the reduced problem satisfy

$$\max\{\overline{n}, \overline{d}\} \leq \overline{m} \equiv \overline{n} + r_{k+1} \leq \overline{n} + \overline{d}.$$

Note that $\overline{d}$ can be smaller than, equal to, or even larger than $\overline{n}$. From the construction we immediately have the following properties:

(CP1) The matrix $A_{11} \in \mathbb{R}^{\overline{m} \times \overline{n}}$ is of *full column rank* equal to $\overline{n} \leq \overline{m}$.
(CP2) The matrix $B_1 \in \mathbb{R}^{\overline{m} \times \overline{d}}$ is of *full column rank* equal to $\overline{d} \leq \overline{m}$.
(CP3) The matrices $\Phi_j \in \mathbb{R}^{r_j \times \overline{d}}$ are of *full row rank* equal to $r_j \leq \overline{d}$, $j = 1, \ldots, k+1$.

REMARK 3.1. *Note that instead of the SVD preprocessing (3.2) of the right-hand side B (section 3.1), one can use an LQ decomposition (producing a matrix with $m \times \overline{d}$ lower triangular block in the column echelon form and an orthogonal matrix), or another decomposition giving a full column rank matrix multiplied by an orthogonal matrix from the right. Similarly, instead of the singular value decompositions (3.10) of F_j (section 3.3) one can use QR decomposition(s) (producing a matrix with an $r_j \times \overline{d}$ upper triangular block in the row echelon form and an orthogonal matrix), or another decomposition(s) giving a full row rank matrix multiplied by an orthogonal matrix from the left. Such modifications lead, in general, to a different subproblem $[\widetilde{B}_1|\widetilde{A}_{11}]$ with the same dimension as (3.20) and satisfying (CP1)–(CP3). Moreover, there exist orthogonal matrices $\widetilde{P}^{-1} = \widetilde{P}^T$, $\widetilde{Q}^{-1} = \widetilde{Q}^T$, $\widetilde{R}^{-1} = \widetilde{R}^T$ such that*

$$\widetilde{P}^T[B_1|A_{11}]\left[\begin{array}{cc} \widetilde{R} & 0 \\ 0 & \widetilde{Q} \end{array}\right] = [\widetilde{B}_1|\widetilde{A}_{11}]. \tag{3.25}$$

Properties (CP1)–(CP3) are invariant with respect to any orthogonal transformation of the form (3.25).

4. The Core Problem. Let $[B_1|A_{11}]$ be the subproblem of the given problem $[B|A]$ obtained by the transformation (3.18)–(3.19). The subproblem $[B_1|A_{11}]$, $B_1 \in \mathbb{R}^{\overline{m} \times \overline{d}}$, $A_{11} \in \mathbb{R}^{\overline{m} \times \overline{n}}$, has the properties (CP1)–(CP3). Consider now *arbitrary* orthogonal transformations of the form in (1.2) of the original problem analogous to (3.18)–(3.19) such that

$$\widehat{P}^T[B|A]\left[\begin{array}{cc} \widehat{R} & 0 \\ 0 & \widehat{Q} \end{array}\right] = \left[\begin{array}{cc|cc} \widehat{B}_1 & 0 & \widehat{A}_{11} & 0 \\ \hline 0 & 0 & 0 & \widehat{A}_{22} \end{array}\right], \tag{4.1}$$

where $\widehat{P}^{-1} = \widehat{P}^T$, $\widehat{Q}^{-1} = \widehat{Q}^T$, $\widehat{R}^{-1} = \widehat{R}^T$, and $\widehat{B}_1 \in \mathbb{R}^{\widehat{m} \times \widehat{d}}$, $\widehat{A}_{11} \in \mathbb{R}^{\widehat{m} \times \widehat{n}}$. Substituting for $[B|A]$ from (3.19) gives

$$(P^T\widehat{P})^T\left[\begin{array}{cc|cc} B_1 & 0 & A_{11} & 0 \\ \hline 0 & 0 & 0 & A_{22} \end{array}\right]\left[\begin{array}{cc} R^T\widehat{R} & 0 \\ 0 & Q^T\widehat{Q} \end{array}\right] = \left[\begin{array}{cc|cc} \widehat{B}_1 & 0 & \widehat{A}_{11} & 0 \\ \hline 0 & 0 & 0 & \widehat{A}_{22} \end{array}\right]. \tag{4.2}$$

We will show that the subproblem $[B_1|A_{11}]$ on the left side of (4.2) has minimal dimensions over all possible subproblems $[\widehat{B}_1|\widehat{A}_{11}]$ on the right-hand side; i.e. $\overline{d} \leq \widehat{d}$, $\overline{m} \leq \widehat{m}$, and $\overline{n} \leq \widehat{n}$. In analogy with the single right-hand side case, this justifies calling $[B_1|A_{11}]$ a *core problem* within $[B|A]$.

4.1. Proof of minimality. The proof consists of five successive steps presented, for an easy orientation, in separate subsections. The first gives $\overline{d} \leq \widehat{d}$. The second step describes the structure of the orthogonal matrix $R^T\widehat{R}$. The proof of the inequality $\overline{m} \leq \widehat{m}$ is then based on the projections of B to the left singular subspaces of A. The fourth step describes the structure of the orthogonal matrix $P^T\widehat{P}$ which is needed for finalizing the proof by showing $\overline{n} \leq \widehat{n}$.

4.1.1. Number of columns of the reduced observation matrix. In (4.1) $\widehat{B}_1 \in \mathbb{R}^{\widehat{m} \times \widehat{d}}$, and

$$\widehat{P}^T B\widehat{R} = \left[\begin{array}{c|c} \widehat{B}_1 & 0 \\ \hline 0 & 0 \end{array}\right] \quad \text{has} \quad \text{rank}(\widehat{B}_1) = \text{rank}(B) = \overline{d}, \quad \text{so that} \quad \overline{d} \leq \widehat{d}.$$

4.1.2. Structure of the $R^T \widehat{R}$ matrix. Consider the partitioning

$$R^T \widehat{R} = \begin{bmatrix} R'_{11} & R'_{12} \\ R'_{21} & R'_{22} \end{bmatrix}, \qquad R'_{11} \in \mathbb{R}^{\overline{d} \times \widehat{d}}, \quad R'_{22} \in \mathbb{R}^{(d-\overline{d}) \times (d-\widehat{d})}.$$

Then (4.2) gives

$$(P^T \widehat{P})^T \left[\begin{array}{c|c} B_1 & 0 \\ \hline 0 & 0 \end{array} \right] (R^T \widehat{R}) = (P^T \widehat{P})^T \left[\begin{array}{c|c} B_1 R'_{11} & B_1 R'_{12} \\ \hline 0 & 0 \end{array} \right] = \left[\begin{array}{c|c} \widehat{B}_1 & 0 \\ \hline 0 & 0 \end{array} \right].$$
$$\underbrace{\phantom{B_1 R'_{11}}}_{\widehat{d}} \quad \underbrace{\phantom{B_1 R'_{12}}}_{d-\widehat{d}} \qquad \underbrace{\phantom{\widehat{B}_1}}_{\widehat{d}} \quad \underbrace{}_{d-\widehat{d}}$$

Because B_1 is of full column rank, then $R'_{12} = 0$. Consequently

$$R^T \widehat{R} = \begin{bmatrix} R'_{11} & 0 \\ R'_{21} & R'_{22} \end{bmatrix}, \tag{4.3}$$

and therefore R'_{11} has $\overline{d}$ orthonormal rows.

4.1.3. Number of rows of the reduced data matrix. The number $\overline{m} = \sum_{j=1}^{k+1} r_j$ of rows of $[B_1 | A_{11}]$ is the sum of dimensions of intersections of the range of B with the individual left singular subspaces of A, and these dimensions are invariant with respect to the orthogonal transformation of the form (3.18)–(3.19); see also (3.20). The desired inequality $\overline{m} \leq \widehat{m}$ will be shown by comparing three different forms of the SVD of the system matrix A. Recall that $A = U\Sigma V^T$ (see (2.2)–(2.4)) is the standard SVD of A. Now consider the singular value decompositions

$$\widehat{A}_{11} = \widehat{U}_1 \widehat{\Sigma}_1 \widehat{V}_1^T, \qquad \widehat{A}_{22} = \widehat{U}_2 \widehat{\Sigma}_2 \widehat{V}_2^T,$$

with square orthogonal matrices $\widehat{U}_1$, $\widehat{U}_2$, $\widehat{V}_1$, and $\widehat{V}_2$, and diagonal matrices $\widehat{\Sigma}_1$, $\widehat{\Sigma}_2$. Then, using (4.1),

$$A = \left(\widehat{P} \begin{bmatrix} \widehat{U}_1 & 0 \\ 0 & \widehat{U}_2 \end{bmatrix} \right) \begin{bmatrix} \widehat{\Sigma}_1 & 0 \\ 0 & \widehat{\Sigma}_2 \end{bmatrix} \left(\widehat{Q} \begin{bmatrix} \widehat{V}_1 & 0 \\ 0 & \widehat{V}_2 \end{bmatrix} \right)^T \tag{4.4}$$

represents another SVD of the system matrix A. Furthermore, A_{11}, A_{22} in (3.20), (3.21) are diagonal matrices. The transformation (3.18)–(3.19) gives

$$A = P \left[\begin{array}{c|c} A_{11} & 0 \\ \hline 0 & A_{22} \end{array} \right] Q^T, \qquad \text{where} \qquad P = U S_L \Pi_L, \tag{4.5}$$

which also represents the SVD of A. Because the singular values (and their multiplicities) of A are unique, for some permutation matrices $\widehat{\Pi}_L$, $\widehat{\Pi}_R$ (analogous to Π_L, Π_R in (3.16), (3.17))

$$\Sigma = \widehat{\Pi}_L \begin{bmatrix} \widehat{\Sigma}_1 & 0 \\ 0 & \widehat{\Sigma}_2 \end{bmatrix} \widehat{\Pi}_R^T = \Pi_L \left[\begin{array}{c|c} A_{11} & 0 \\ \hline 0 & A_{22} \end{array} \right] \Pi_R^T.$$

The left singular vector subspaces of A can be expressed using (4.4), (4.5) and some orthogonal matrices

$$\widehat{S}_L = \text{diag}(\widehat{S}_1, \ldots, \widehat{S}_k, \widehat{S}_{k+1}), \qquad \widehat{S}_j \in \mathbb{R}^{m_j \times m_j}, \quad j = 1, \ldots, k, k+1, \tag{4.6}$$

as

$$U = \widehat{P} \begin{bmatrix} \widehat{U}_1 & 0 \\ 0 & \widehat{U}_2 \end{bmatrix} \widehat{\Pi}_L^T \widehat{S}_L^T = P \Pi_L^T S_L^T. \tag{4.7}$$

Using (4.1) and (3.19), the projections of B in the left singular subspaces of A are

$$U^T B = \widehat{S}_L \widehat{\Pi}_L \left[\begin{array}{c|c} \widehat{U}_1^T \widehat{B}_1 & 0 \\ \hline 0 & 0 \end{array} \right] \widehat{R}^T = S_L \Pi_L \left[\begin{array}{c|c} B_1 & 0 \\ \hline 0 & 0 \end{array} \right] R^T,$$

and, with (4.3),

$$\left[\begin{array}{c|c} \widehat{U}_1^T \widehat{B}_1 & 0 \\ \hline 0 & 0 \end{array} \right] = \widehat{\Pi}_L^T \widehat{S}_L^T S_L \Pi_L \left[\begin{array}{c|c} B_1 R'_{11} & 0 \\ \hline 0 & 0 \end{array} \right]. \tag{4.8}$$

The matrix $R'_{11} \in \mathbb{R}^{\overline{d} \times \widehat{d}}$, $\widehat{d} \geq \overline{d}$, has linearly independent rows; see (4.3) and section 4.1.1. Now every row of B_1 is nonzero (see (3.20) and (CP3)), so any row of B_1 multiplied by R'_{11} on the right must also be nonzero, and the matrix $B_1 R'_{11}$ has $\overline{m}$ nonzero rows. The matrix Π_L permutes rows (see (3.16)), and matrices S_L, $\widehat{S}_L$ (see (3.11), (4.6)) are orthogonal block diagonal matrices such that (see (3.15), (3.14)),

$$\widehat{S}_L^T S_L \Pi_L \left[\begin{array}{c|c} B_1 R'_{11} & 0 \\ \hline 0 & 0 \end{array} \right] = \begin{bmatrix} \widehat{S}_1^T S_1 \begin{bmatrix} \Phi_1 R'_{11} & 0 \\ 0 & 0 \end{bmatrix} \\ \vdots \\ \widehat{S}_k^T S_k \begin{bmatrix} \Phi_k R'_{11} & 0 \\ 0 & 0 \end{bmatrix} \\ \widehat{S}_{k+1}^T S_{k+1} \begin{bmatrix} \Phi_{k+1} R'_{11} & 0 \\ 0 & 0 \end{bmatrix} \end{bmatrix} \begin{array}{l} \left.\right\} m_1 \\ \vdots \\ \left.\right\} m_k \\ \left.\right\} m_{k+1} \end{array}, \tag{4.9}$$

where each block of m_j rows corresponds to one singular value σ_j (of the matrix Σ) with the multiplicity m_j. Because $\Phi_j R'_{11}$ has all r_j rows nonzero (since the rows of Φ_j are linearly independent, the rows of $\Phi_j R'_{11}$ are, in fact, also linearly independent and therefore each $\Phi_j R'_{11}$ is of full row rank) and $\widehat{S}_j^T S_j$ are orthogonal matrices, the block matrix on the right of (4.9) has at least $\overline{m}$ nonzero rows, see (3.20). The permutation matrix $\widehat{\Pi}_L^T$ then moves all the nonzero rows (and possibly also some zero rows) of (4.9) to the top block $[\widehat{U}_1^T \widehat{B}_1, 0] \in \mathbb{R}^{\widehat{m} \times d}$ of the leftmost matrix in (4.8). Thus $\widehat{U}_1^T \widehat{B}_1$ has at least $\overline{m}$ nonzero rows (and possibly some zero rows). Because $\widehat{U}_1$ is square, $\widehat{B}_1 \in \mathbb{R}^{\widehat{m} \times \widehat{d}}$ has at least $\overline{m}$ rows, so that $\overline{m} \leq \widehat{m}$.

4.1.4. Structure of the $P^T \widehat{P}$ matrix. Consider the partitioning of the permutation matrices Π_L in (3.16) with $\overline{m} = \sum_{j=1}^{k+1} r_j$ as in (3.20), and $\widehat{\Pi}_L$ in (4.8) where $\widehat{U}_1$ is $\widehat{m} \times \widehat{m}$,

$$\Pi_L = [\Pi_1, \Pi_2], \quad \Pi_1 \in \mathbb{R}^{m \times \overline{m}}, \qquad \widehat{\Pi}_L = [\widehat{\Pi}_1, \widehat{\Pi}_2], \quad \widehat{\Pi}_1 \in \mathbb{R}^{m \times \widehat{m}}.$$

Then (4.8) can be rewritten as

$$\left[\begin{array}{c|c} \widehat{U}_1^T \widehat{B}_1 & 0 \\ \hline 0 & 0 \end{array} \right] = \left[\begin{array}{cc} \widehat{\Pi}_1^T \widehat{S}_L^T S_L \Pi_1 & \widehat{\Pi}_1^T \widehat{S}_L^T S_L \Pi_2 \\ \widehat{\Pi}_2^T \widehat{S}_L^T S_L \Pi_1 & \widehat{\Pi}_2^T \widehat{S}_L^T S_L \Pi_2 \end{array} \right] \left[\begin{array}{c|c} B_1 R'_{11} & 0 \\ \hline 0 & 0 \end{array} \right],$$

yielding the condition

$$(\widehat{\Pi}_2^T \widehat{S}_L^T S_L \Pi_1)(B_1 R'_{11}) = 0. \qquad (4.10)$$

Using the structure of the matrices Π_L, S (see (3.11), (3.16)) and analogously for $\widehat{\Pi}_L$ and $\widehat{S}$ (see section 4.1.3)

$$\widehat{\Pi}_2^T \widehat{S}_L^T S_L \Pi_1$$

$$= \begin{bmatrix} [0, I_{m_1 - \widehat{r}_1}]\widehat{S}_1^T S_1 \begin{bmatrix} I_{r_1} \\ 0 \end{bmatrix} & & 0 & & 0 \\ & \ddots & & & \vdots \\ 0 & & [0, I_{m_k - \widehat{r}_k}]\widehat{S}_k^T S_k \begin{bmatrix} I_{r_k} \\ 0 \end{bmatrix} & & 0 \\ 0 & \cdots & 0 & & [0, I_{m_{k+1} - \widehat{r}_{k+1}}]\widehat{S}_{k+1}^T S_{k+1} \begin{bmatrix} I_{r_{k+1}} \\ 0 \end{bmatrix} \end{bmatrix},$$

and thus with (3.20) the condition (4.10) is split into $k + 1$ block parts

$$\left([0, I_{m_j - \widehat{r}_j}]\widehat{S}_j^T S_j \begin{bmatrix} I_{r_j} \\ 0 \end{bmatrix} \right) (\Phi_j R'_{11}) = 0, \qquad j = 1, \ldots, k, k+1.$$

Because $\Phi_j R'_{11}$ are of full row rank, this gives for all indices j

$$[0, I_{m_j - \widehat{r}_j}]\widehat{S}_j^T S_j \begin{bmatrix} I_{r_j} \\ 0 \end{bmatrix} = 0, \qquad \text{and thus also} \qquad \widehat{\Pi}_2^T \widehat{S}_L^T S_L \Pi_1 = 0.$$

Using (4.7) we finally obtain

$$P^T \widehat{P} = \Pi_L^T S_L^T \widehat{S}_L \widehat{\Pi}_L \begin{bmatrix} \widehat{U}_1 & 0 \\ 0 & \widehat{U}_2 \end{bmatrix}^T$$

$$= \begin{bmatrix} \Pi_1^T S_L^T \widehat{S}_L \widehat{\Pi}_1 \widehat{U}_1^T & 0 \\ \Pi_2^T S_L^T \widehat{S}_L \widehat{\Pi}_1 \widehat{U}_1^T & \Pi_2^T S_L^T \widehat{S}_L \widehat{\Pi}_2 \widehat{U}_2^T \end{bmatrix} \equiv \begin{bmatrix} P'_{11} & 0 \\ P'_{21} & P'_{22} \end{bmatrix} \begin{matrix} \} \, \overline{m} \\ \} \, m - \overline{m} \end{matrix} \qquad (4.11)$$

$$\underbrace{\qquad}_{\widehat{m}} \underbrace{\qquad}_{m - \widehat{m}}$$

and therefore P'_{11} has $\overline{m}$ orthonormal rows.

4.1.5. Number of columns of the reduced system matrix. The relation (4.2) gives, using the structure of $P^T \widehat{P}$ in (4.11),

$$\begin{bmatrix} P'_{11} & 0 \\ P'_{21} & P'_{22} \end{bmatrix}^T \left[\begin{array}{c|c} A_{11} & 0 \\ \hline 0 & A_{22} \end{array} \right] (Q^T \widehat{Q}) = \left[\begin{array}{c|c} \widehat{A}_{11} & 0 \\ \hline 0 & \widehat{A}_{22} \end{array} \right].$$

Thus

$$[(P'_{11})^T A_{11}, (P'_{21})^T A_{22}](Q^T \widehat{Q}) = [\widehat{A}_{11}, 0].$$

Because $(P'_{11})^T$ has orthonormal columns and $Q^T \widehat{Q}$ is a square orthogonal matrix, $A_{11} \in \mathbb{R}^{\overline{m} \times \overline{n}}$ in (3.20) has rank $\overline{n}$, and $\widehat{A}_{11} \in \mathbb{R}^{\widehat{m} \times \widehat{n}}$ (see (4.1)),

$$\overline{n} = \text{rank}(A_{11}) = \text{rank}((P'_{11})^T A_{11}) \leq \text{rank}([\widehat{A}_{11}, 0]) = \text{rank}(\widehat{A}_{11}) \leq \widehat{n}$$

completing our proof.

5. Summary and concluding remarks. We formulate the minimality property as a theorem.

THEOREM 5.1 (Minimality). *Consider the problem $AX \approx B$ (1.1) and its subproblem $A_{11}X_1 \approx B_1$ (3.23), where $A_{11} \in \mathbb{R}^{\overline{m} \times \overline{n}}$ and $B_1 \in \mathbb{R}^{\overline{m} \times \overline{d}}$ are obtained by the transformation (3.19) described in section 3. The subproblem $A_{11}X_1 \approx B_1$ has minimal dimensions over all subproblems $\widehat{A}_{11}\widehat{X}_1 \approx \widehat{B}_1$, $\widehat{A}_{11} \in \mathbb{R}^{\widehat{m} \times \widehat{n}}$, $\widehat{B}_1 \in \mathbb{R}^{\widehat{m} \times \widehat{d}}$ obtained by an orthogonal transformation of the form (4.1), i.e. $\widehat{d} \geq \overline{d}$, $\widehat{m} \geq \overline{m}$, and $\widehat{n} \geq \overline{n}$.*

This motivates the following definition of the core problem within (1.1).

DEFINITION 5.2 (Core problem). *The subproblem $A_{11}X_1 \approx B_1$ is a core problem within the approximation problem $AX \approx B$ if $[B_1|A_{11}]$ is minimally dimensioned and A_{22} maximally dimensioned subject to (3.1), i.e. subject to the orthogonal transformations of the form*

$$P^T[B|A] \left[\begin{array}{c|c} R & 0 \\ \hline 0 & Q \end{array} \right] = [P^T BR | P^T AQ] \equiv \left[\begin{array}{c|c||c|c} B_1 & 0 & A_{11} & 0 \\ \hline 0 & 0 & 0 & A_{22} \end{array} \right].$$

The core problem obtained by the data reduction based on the SVD described in section 3 is called the core problem in the SVD form.

This extends the core problem definition within the single right-hand side problems formulated by C. C. Paige and Z. Strakoš in [14].

The data reduction presented here is based on the singular value decomposition of the system matrix A. The original paper [14] presents two ways of determining the core problem. The second is based on Golub-Kahan iterative bidiagonalization. This can also be generalized to problems with multiple right-hand sides. It leads to a band algorithm which was for this purpose proposed by Å. Björck; see [1], [2], [3], and also the unpublished manuscript [4].

The core problem concept is a useful tool in understanding the TLS problems, which was the original motivation in [14]. The data reduction and core problem based on the generalization of the Golub-Kahan iterative bidiagonalization is under investigation. The results will be presented in the near future.

REFERENCES

[1] Å. BJÖRCK, *Bidiagonal decomposition and least squares*, Presentation, Canberra (2005).
[2] ——, *A Band-Lanczos generalization of bidiagonal decomposition*, Presentation, Conference in Honor of G. Dahlquist, Stockholm (2006).
[3] ——, *A band-Lanczos algorithm for least squares and total least squares problems*, In: book of abstracts of 4th Total Least Squares and Errors-in-Variables Modeling Workshop, Leuven, Katholieke Universiteit Leuven (2006), pp. 22–23.
[4] ——, *Block bidiagonal decomposition and least squares problems with multiple right-hand sides*, unpublished manuscript.
[5] G. H. GOLUB, *Some modified matrix eigenvalue problems*, SIAM Review, 15 (1973), pp. 318–334.
[6] G. H. GOLUB, A. HOFFMAN, AND G. W. STEWART, *A generalization of the Eckart-Young-Mirsky matrix approximation theorem*, Linear Algebra Appl., 88/89 (1987), pp. 317–327.
[7] G. H. GOLUB AND C. REINSCH, *Singular value decomposition and least squares solutions*, Numerische Mathematik, 14 (1970), pp. 403–420.

[8] G. H. GOLUB AND C. F. VAN LOAN, *An analysis of the total least squares problem*, SIAM J. Num. Anal., 17 (1980), pp. 883–893.

[9] ——, *Matrix Computations, fourth ed.*, The Johns Hopkins University Press, Baltimore and London, 2012.

[10] I. HNĚTYNKOVÁ, M. PLEŠINGER, D. M. SIMA, Z. STRAKOŠ, AND S. VAN HUFFEL, *The total least squares problem in $AX \approx B$: A new classification with the relationship to the classical works*, SIAM J. on Matrix Anal. and Appl., 32 (2011), pp. 748–777.

[11] I. HNĚTYNKOVÁ AND Z. STRAKOŠ, *Lanczos tridiagonalization and core problems*, Linear Algebra Appl., Vol. 421 (2007), pp. 243–251.

[12] C. C. PAIGE AND Z. STRAKOŠ, *Scaled total least squares fundamentals*, Numerische Mathematik, 91 (2002), pp. 117–146.

[13] ——, *Unifying least squares, total least squares and data least squares*, In: Total Least Squares and Errors-in-Variables Modeling, S. Van Huffel and P. Lemmerling, eds., Dordrecht, Kluwer Academic Publishers (2002), pp. 25–34.

[14] ——, *Core problem in linear algebraic systems*, SIAM J. Matrix Anal. Appl., 27 (2006), pp. 861–875.

[15] M. PLEŠINGER, *The Total Least Squares Problem and Reduction of Data in $AX \approx B$*, Ph.D. thesis, Technical University of Liberec, Liberec, Czech Rebuplic, 2008.

[16] D. M. SIMA, *Regularization techniques in model fitting and parameter estimation*, Ph.D. thesis, Dept. of Electrical Engineering, Katholieke Universiteit Leuven, 2006.

[17] D. M. SIMA AND S. VAN HUFFEL, *Core problems in $AX \approx B$*, Technical Report, Dept. of Electrical Engineering, Katholieke Universiteit Leuven (2006).

[18] A. VAN DER SLUIS, *Stability of the solutions of linear least squares problems*, Numerische Mathematik, 23 (1975), pp. 241–254.

[19] S. VAN HUFFEL AND J. VANDEWALLE, *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM Publications, Philadelphia, PA, 1991.

[20] S. VAN HUFFEL, ED., *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling*, Proceedings of the Second Int. Workshop on TLS and EIV, Philadelphia, SIAM Publications, 1997.

[21] S. VAN HUFFEL AND P. LEMMERLING, EDS., *Total Least Squares and Errors-in-Variables Modeling: Analysis, Algorithms and Applications*, Dordrecht, Kluwer Academic Publishers, 2002.

[22] M. WEI, *The analysis for the total least squares problem with more than one solution*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 746–763.

[23] ——, *Algebraic relations between the total least squares and least squares problems with more than one solution*, Numer. Math., 62 (1992), pp. 123–148.