CORE PROBLEM WITHIN 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 a 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 an extension of this concept to problems with multiple right-hand sides. The presented data reduction 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. Some other properties of the the core problem, however, can not be extended for the 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 a sense to approximate B by a linear combination of the columns of A; see also [13]). 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,

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

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)

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

*This work has been supported by the GAAV grant IAA100300802. The research of M. Plešinger has been supported by the research project ESF No. CZ.1.07/2.3.00/09.0155, “Constitution and improvement of a team for demanding technical computations on parallel computers at TU Liberec”. The research of Z. Strakoš has been supported by the GACR grant 201/09/0917.

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

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

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

serves as an important example; see [11], [12], [7], [5, section 6], [17]. Mathematically equivalent problems have been independently investigated under the names *orthogonal regression* and *errors-in-variables modeling*; see [19], [20].

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 [18] 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 [9], giving a new classification of all possible cases.

The sequence of papers [11], [12], and [13] by C. C. Paige and Z. Strakoš investigates, using an unified framework, different least squares formulations for problems with $d = 1$. The last paper [13] introduces the so-called *core problem* separating the information sufficient and necessary 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 [9, pp. 752–753].

The first steps in generalization of the core problem theory for $d > 1$ were done by Å. Björk 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 [15], [16]. In this paper we further develop the data reduction suggested in [14] which gives the core problem, we investigate its properties and prove its minimality.

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 discuses 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 ℓ-by-ℓ identity matrix; and $0_{\ell,\xi}$ (or just 0) denotes an ℓ-by-ξ 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 [13] it was shown that there exist orthogonal matrices P, Q, that transform the original problem into the block form

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

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, $j = 1, \ldots, k$; $\sum_{j=1}^{k} m_j = r \equiv \mathrm{rank}(A)$. Then

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

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 [13]) 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

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

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

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

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 [13]. 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 $x_2 = 0$ (see the discussion in [13]) giving

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

The core problem (in a different form) can also be revealed by the Golub-Kahan bidiagonalization; see [13, Section 3] and [10].

The core problem has the unique TLS solution, see [13]. Using (2.9), the core problem defines the minimum norm TLS solution, if it exists, or the minimum norm nongeneric solution; see [13] and [9, section 3.1, pp. 752–753]. In the rest of this paper it will be shown how to generalize the SVD-based data reduction and obtain the core problem for $d > 1$.

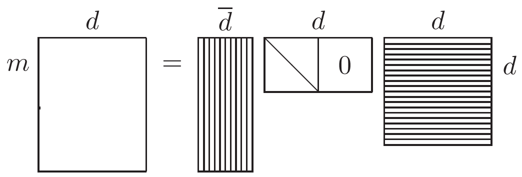
3. Data reduction in the multiple right-hand side case. Consider the problem (1.1) with $d > 1$. In this section, we construct orthogonal matrices P, Q, R, that transform the original data matrix $[B|A]$ into the block form

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

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 subsequent 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 $\overline{d} \equiv \operatorname{rank}(B) \leq \min\{m, d\}$. Consider the SVD of B in the form

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



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

$$A(XR) \approx BR, \tag{3.3}$$

where

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

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

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

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

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

where the second problem is homogenous. Following the argumentation in [13], we consider the meaningful solution $Y' \equiv 0$. In this way, the approximation problem (1.1) reduces to

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

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),

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

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

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

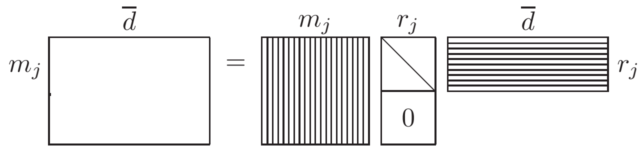
The approximation problem $\Sigma Z \approx F$ has the full column rank right-hand side matrix F with mutually orthogonal columns and the system matrix Σ in a diagonal form.

3.3. Transformation of the right-hand side. Similarly to the single right-hand side case, we now transform the right-hand side matrix 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 $r_j \equiv \mathrm{rank}(F_j) \leq \min\{m_j, \overline{d}\}$. Consider the SVD of F_j in the form

$$F_j = S_j \Theta_j W_j^T, \qquad S_j \in \mathbb{R}^{m_j \times m_j}, \quad \Theta_j \in \mathbb{R}^{m_j \times r_j}, \quad W_j \in \mathbb{R}^{\overline{d} \times r_j}, \tag{3.10}$$



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 rows, 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 identically denoted matrix in section 2; see (2.7). Consider the block diagonal orthogonal matrices

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

and recall that

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

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

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

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

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

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 [13, Section 2]), by moving the rows of (3.13) with zero blocks in $\Theta_j W_j^T$ (see (3.14)) to the bottom of the whole matrix with subsequent moving of the corresponding columns with the diagonal blocks in the bottom to the right,

$$
\begin{aligned}
&\Pi_L^T \left[\begin{array}{c|cccc}
\Theta_1 W_1^T & \sigma_1 I_{m_1} & & 0 & 0 \\
\vdots & & \ddots & & \vdots \\
\Theta_k W_k^T & 0 & & \sigma_k I_{m_k} & 0 \\
\Theta_{k+1} W_{k+1}^T & 0 & \cdots & 0 & 0
\end{array}\right]
\left[\begin{array}{cc} I_{\overline{d}} & 0 \\ 0 & \Pi_R \end{array}\right] \\
&= \left[\begin{array}{c||ccc|cccc}
\Phi_1 & \sigma_1 I_{r_1} & & 0 & 0 & \cdots & 0 & 0 \\
\vdots & & \ddots & & \vdots & & \vdots & \vdots \\
\Phi_k & 0 & & \sigma_k I_{r_k} & 0 & \cdots & 0 & 0 \\
\Phi_{k+1} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
\hline
0 & 0 & \cdots & 0 & \sigma_1 I_{m_1 - r_1} & & 0 & 0 \\
\vdots & \vdots & & \vdots & & \ddots & & \vdots \\
0 & 0 & \cdots & 0 & 0 & & \sigma_k I_{m_k - r_k} & 0 \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0
\end{array}\right] \\
&\equiv \left[\begin{array}{c||c|c}
B_1 & A_{11} & 0 \\
\hline
0 & 0 & A_{22}
\end{array}\right]. \qquad (3.15)
\end{aligned}
$$

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

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

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

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

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 SVD 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).

Consider

$$P \equiv U \mathrm{diag}(S_\oplus, S_{k+1})\, \Pi_L, \qquad Q \equiv V \mathrm{diag}(S_\oplus, I_{n-r})\, \Pi_R, \tag{3.18}$$

and recall the orthogonal matrix R defined by the SVD (3.2) of B. Clearly $P^{-1} = P^T$, $Q^{-1} = Q^T$, $R^{-1} = R^T$, and

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

is of the form (3.1), where

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

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

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

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

$$Q^T X R = \begin{array}{c} \left[\begin{array}{cc} X_1 & X_1' \\ X_2 & X_2' \end{array}\right] \begin{array}{l} \}\ \overline{n} \\ \}\ n-\overline{n} \end{array} \\ \begin{array}{cc} \underbrace{\quad}_{\overline{d}} & \underbrace{\quad}_{d-\overline{d}} \end{array} \end{array}, \tag{3.21}$$

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 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 argumentation in [13], $X_2 \equiv 0$, $X_1' \equiv 0$, $X_2' \equiv 0$. Only the subproblem

$$A_{11} X_1 \approx B_1, \qquad \text{or, equivalently,} \qquad [B_1 | A_{11}] \begin{bmatrix} -I_{\overline{d}} \\ X_1 \end{bmatrix} \approx 0, \tag{3.22}$$

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. Giving its solution X_1, the solution X of the original problem $AX \approx B$ is expressed as

$$X \equiv Q \begin{bmatrix} X_1 & 0 \\ 0 & 0 \end{bmatrix} R^T. \tag{3.23}$$

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, equal, 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-by-$\overline{d}$ 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 SVD(s) (3.10) of F_j (section 3.3) one can use QR decomposition(s) (producing a matrix with r_j-by-$\overline{d}$ 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*

$$\widetilde{P}^T [B_1 | A_{11}] \begin{bmatrix} \widetilde{R} & 0 \\ 0 & \widetilde{Q} \end{bmatrix} = [\widetilde{B}_1 | \widetilde{A}_{11}]. \tag{3.24}$$

Properties (CP1)–(CP3) are invariant with respect to any orthogonal transformation of the form (3.24).

4. Core problem. Now we show that the reduced problem $[B_1 | A_{11}]$ is minimal over all orthogonal transformations yielding the block structure (3.1). In analogy with the single right-hand side case we call $[B_1 | A_{11}]$ a *core problem* within $[B|A]$.

Let $[B_1 | A_{11}]$ be the subproblem obtained by the transformation (3.18), (3.19) as described above. The subproblem $[B_1 | A_{11}]$ has the properties (CP1)–(CP3), $B_1 \in \mathbb{R}^{\overline{m} \times \overline{d}}$, $A_{11} \in \mathbb{R}^{\overline{m} \times \overline{n}}$. Assume an arbitrary orthogonal transformation of the original problem such that

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

where $\widehat{P}^{-1} = \widehat{P}^T$, $\widehat{Q}^{-1} = \widehat{Q}^T$, $\widehat{R}^{-1} = \widehat{R}^T$, $\widehat{B}_1 \in \mathbb{R}^{\widehat{m} \times \widehat{d}}$, $\widehat{A}_{11} \in \mathbb{R}^{\widehat{m} \times \widehat{n}}$. We prove that $\widehat{d} \geq \overline{d}$, $\widehat{m} \geq \overline{m}$, and $\widehat{n} \geq \overline{n}$.

4.1. Number of columns of the reduced observation matrix. First we prove $\widehat{d} \geq \overline{d}$. From (3.19),

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

Substitution into (4.1) gives

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

yielding

$$(P^T \widehat{P})^T \left[\begin{array}{c|c} B_1 & 0 \\ \hline 0 & 0 \end{array} \right] (R^T \widehat{R}) = \left[\begin{array}{c|c} \widehat{B}_1 & 0 \\ \hline 0 & 0 \end{array} \right]. \quad (4.3)$$

From (CP2), B_1 is of full column rank $\overline{d}$. Therefore the matrix $\widehat{B}_1 \in \mathbb{R}^{\widehat{m} \times \widehat{d}}$ is also of rank $\overline{d}$ and thus $\widehat{d} \geq \overline{d}$, i.e., the observation matrix B_1 from the reduction (3.19) has the minimal number of columns of all possible transformations (4.1).

4.2. Number of rows of the reduced data matrix. Before proving $\widehat{m} \geq \overline{m}$, it will be useful to show that the orthogonal matrix $R^T \widehat{R}$ has a special structure. Consider the partitioning

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

Then (4.3) gives

$$(P^T \widehat{P})^T \left[\begin{array}{c|c} \underbrace{\begin{array}{c} B_1 R'_{11} \\ \hline 0 \end{array}}_{\widehat{d}} & \underbrace{\begin{array}{c} B_1 R'_{12} \\ \hline 0 \end{array}}_{d-\widehat{d}} \end{array} \right] = \left[\begin{array}{c|c} \underbrace{\begin{array}{c} \widehat{B}_1 \\ \hline 0 \end{array}}_{\widehat{d}} & \underbrace{\begin{array}{c} 0 \\ \hline 0 \end{array}}_{d-\widehat{d}} \end{array} \right]. \quad (4.4)$$

Since B_1 is of full column rank, this gives $R'_{12} = 0$. Consequently

$$R^T \widehat{R} = \left[\begin{array}{cc} R'_{11} & 0 \\ R'_{21} & R'_{22} \end{array} \right], \quad (4.5)$$

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

Now we prove $\widehat{m} \geq \overline{m}$. Consider the SVD 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 orthogonal matrices $\widehat{U}_1 \in \mathbb{R}^{\widehat{m} \times \widehat{m}}$, $\widehat{U}_2 \in \mathbb{R}^{(m-\widehat{m}) \times (m-\widehat{m})}$, $\widehat{V}_1 \in \mathbb{R}^{\widehat{n} \times \widehat{n}}$, and $\widehat{V}_2 \in \mathbb{R}^{(n-\widehat{n}) \times (n-\widehat{n})}$, and diagonal matrices $\widehat{\Sigma}_1 \in \mathbb{R}^{\widehat{m} \times \widehat{n}}$, $\widehat{\Sigma}_2 \in \mathbb{R}^{(m-\widehat{m}) \times (n-\widehat{n})}$, both with standard (descending) ordering of singular values; i.e.

$$\left[\begin{array}{c|c} \widehat{A}_{11} & 0 \\ \hline 0 & \widehat{A}_{22} \end{array} \right] = \left[\begin{array}{cc} \widehat{U}_1 & 0 \\ 0 & \widehat{U}_2 \end{array} \right] \left[\begin{array}{cc} \widehat{\Sigma}_1 & 0 \\ 0 & \widehat{\Sigma}_2 \end{array} \right] \left[\begin{array}{cc} \widehat{V}_1 & 0 \\ 0 & \widehat{V}_2 \end{array} \right]^T . \quad (4.6)$$

Comparing of (2.2), (3.19), and (4.1) with (4.6), gives

$$\begin{aligned} A = U\Sigma V^T &= P \left[\begin{array}{c|c} A_{11} & 0 \\ \hline 0 & A_{22} \end{array} \right] Q^T \\ &= \left(\widehat{P} \begin{bmatrix} \widehat{U}_1 & 0 \\ 0 & \widehat{U}_2 \end{bmatrix} \right) \begin{bmatrix} \widehat{\Sigma}_1 & 0 \\ 0 & \widehat{\Sigma}_2 \end{bmatrix} \left(\begin{bmatrix} \widehat{V}_1 & 0 \\ 0 & \widehat{V}_2 \end{bmatrix}^T \widehat{Q}^T \right). \end{aligned} \tag{4.7}$$

Recall that $A_{11} \in \mathbb{R}^{\overline{m} \times \overline{n}}$ and $A_{22} \in \mathbb{R}^{(m-\overline{m})\times(n-\overline{n})}$ are diagonal matrices; see (3.15). Thus (4.7) represents three different SVD decompositions of A (the last two with non-standard forms of matrices containing singular values). Because the singular values and their multiplicities are *unique* for any given matrix A, there exist permutation matrices Π_L, Π_R, $\widehat{\Pi}_L$, and $\widehat{\Pi}_R$ such that

$$\Sigma = \Pi_L \left[\begin{array}{c|c} A_{11} & 0 \\ \hline 0 & A_{22} \end{array} \right] \Pi_R^T = \widehat{\Pi}_L \begin{bmatrix} \widehat{\Sigma}_1 & 0 \\ 0 & \widehat{\Sigma}_2 \end{bmatrix} \widehat{\Pi}_R^T. \tag{4.8}$$

Permutation matrices Π_L and Π_R are given by (3.16) and (3.17), respectively. The structure of the other two matrices $\widehat{\Pi}_L$ and $\widehat{\Pi}_R$ is fully analogous. In particular

$$\widehat{\Pi}_L \equiv \left[\begin{array}{cccc|cccc} \left[\begin{smallmatrix} I_{\widehat{r}_1} \\ 0 \end{smallmatrix}\right] & & 0 & 0 & \left[\begin{smallmatrix} 0 \\ I_{m_1-\widehat{r}_1} \end{smallmatrix}\right] & & 0 & 0 \\ & \ddots & & \vdots & & \ddots & & \vdots \\ 0 & & \left[\begin{smallmatrix} I_{\widehat{r}_k} \\ 0 \end{smallmatrix}\right] & 0 & 0 & & \left[\begin{smallmatrix} 0 \\ I_{m_k-\widehat{r}_k} \end{smallmatrix}\right] & 0 \\ 0 & \cdots & 0 & \left[\begin{smallmatrix} I_{\widehat{r}_{k+1}} \\ 0 \end{smallmatrix}\right] & 0 & \cdots & 0 & \left[\begin{smallmatrix} 0 \\ I_{m_{k+1}-\widehat{r}_{k+1}} \end{smallmatrix}\right] \end{array} \right], \tag{4.9}$$

where m_j are the multiplicities of distinct nonzero singular values of Σ, $j = 1, \ldots, k$, $m_{k+1} = \dim(\mathcal{N}(\Sigma^T))$, see (2.3), (2.4), and $\widehat{r}_j$ are the multiplicities of distinct nonzero singular values of $\widehat{\Sigma}_1$, $j = 1, \ldots, k$, $\widehat{r}_{k+1} = \dim(\mathcal{N}(\widehat{\Sigma}_1^T))$, $\sum_{j=1}^{k+1} \widehat{r}_j = \widehat{m}$. Combination of (4.7) and (4.8) yields the last two SVD decompositions from (4.7) with the same (and standard) ordering of singular values placed in the matrix Σ, i.e.

$$A = P(\Pi_L^T \Sigma \Pi_R) Q^T = \left(\widehat{P} \begin{bmatrix} \widehat{U}_1 & 0 \\ 0 & \widehat{U}_2 \end{bmatrix} \right) (\widehat{\Pi}_L^T \Sigma \widehat{\Pi}_R) \left(\begin{bmatrix} \widehat{V}_1 & 0 \\ 0 & \widehat{V}_2 \end{bmatrix}^T \widehat{Q}^T \right). \tag{4.10}$$

Because the left singular vector spaces and their dimensions are *unique* for any given matrix A, there exist orthogonal matrices

$$H = \mathrm{diag}(H_1, \ldots, H_k, H_{k+1}), \qquad H_j \in \mathbb{R}^{m_j \times m_j}, \quad j = 1, \ldots, k, k+1, \tag{4.11}$$

transforming bases of these singular vector spaces such that

$$\left(P\Pi_L^T \right) H = \left(\widehat{P} \begin{bmatrix} \widehat{U}_1 & 0 \\ 0 & \widehat{U}_2 \end{bmatrix} \widehat{\Pi}_L^T \right). \tag{4.12}$$

(Analogously for the right singular vector subspaces.) Since

$$\widehat{P}^T A \widehat{Q} = \left[\begin{array}{c|c} \widehat{A}_{11} & 0 \\ \hline 0 & \widehat{A}_{22} \end{array} \right],$$

see (4.1), multiplication of (4.10) by $\widehat{P}^T$ and $\widehat{Q}$ from the left and right, respectively, gives the following two SVD decompositions of $\mathrm{diag}(\widehat{A}_{11}, \widehat{A}_{22})$,

$$\left[\begin{array}{c|c} \widehat{A}_{11} & 0 \\ \hline 0 & \widehat{A}_{22} \end{array}\right] = (P^T\widehat{P})^T(\Pi_L^T\Sigma\Pi_R)(Q^T\widehat{Q}) = \begin{bmatrix} \widehat{U}_1 & 0 \\ 0 & \widehat{U}_2 \end{bmatrix} (\widehat{\Pi}_L^T\Sigma\widehat{\Pi}_R) \begin{bmatrix} \widehat{V}_1 & 0 \\ 0 & \widehat{V}_2 \end{bmatrix}^T.$$

For the matrices with the left singular vectors we have

$$\left((P^T\widehat{P})^T\Pi_L^T\right)H = \left(\begin{bmatrix} \widehat{U}_1 & 0 \\ 0 & \widehat{U}_2 \end{bmatrix}\widehat{\Pi}_L^T\right), \tag{4.13}$$

i.e. (4.12) multiplied by $\widehat{P}^T$ from the left. Using (4.13) and (4.3) gives

$$\begin{aligned} \widehat{\Pi}_L \begin{bmatrix} \widehat{U}_1 & 0 \\ 0 & \widehat{U}_2 \end{bmatrix}^T \left[\begin{array}{c|c} \widehat{B}_1 & 0 \\ \hline 0 & 0 \end{array}\right] &= \widehat{\Pi}_L \left[\begin{array}{c|c} \widehat{U}_1^T\widehat{B}_1 & 0 \\ \hline 0 & 0 \end{array}\right] \\ &= H^T\Pi_L(P^T\widehat{P})\left((P^T\widehat{P})^T\left[\begin{array}{c|c} B_1 & 0 \\ \hline 0 & 0 \end{array}\right](R^T\widehat{R})\right) \\ &= H^T\Pi_L\left[\begin{array}{c|c} B_1 & 0 \\ \hline 0 & 0 \end{array}\right](R^T\widehat{R}) = H^T\Pi_L\left[\begin{array}{c|c} B_1R'_{11} & 0 \\ \hline 0 & 0 \end{array}\right], \end{aligned}$$

where the last equality exploits the structure of $R^T\widehat{R}$ matrix; see (4.5). This gives

$$\left[\begin{array}{c|c} \widehat{U}_1^T\widehat{B}_1 & 0 \\ \hline 0 & 0 \end{array}\right] = \widehat{\Pi}_L^T H^T\Pi_L\left[\begin{array}{c|c} B_1R'_{11} & 0 \\ \hline 0 & 0 \end{array}\right]. \tag{4.14}$$

Recall that $\widehat{B}_1 \in \mathbb{R}^{\widehat{m}\times\widehat{d}}$ and $B_1 \in \mathbb{R}^{\overline{m}\times\overline{d}}$, where using (3.20) and (CP3),

$$B_1 = [\Phi_1^T, \ldots, \Phi_k^T, \Phi_{k+1}^T]^T, \qquad \Phi_j \in \mathbb{R}^{r_j\times\overline{d}}, \qquad \mathrm{rank}(\Phi_j) = r_j.$$

In particular, B_1 has $\overline{m}$ nonzero rows. The matrix $R'_{11} \in \mathbb{R}^{\overline{d}\times\widehat{d}}$, $\widehat{d} \geq \overline{d}$ is of full row rank, see (4.5), thus also $(\Phi_jR'_{11}) \in \mathbb{R}^{r_j\times\widehat{d}}$ are of full row rank r_j, and the matrix $B_1R'_{11}$ has $\overline{m}$ nonzero rows. The matrix Π_L permutes rows such that

$$\Pi_L\left[\begin{array}{c|c} B_1R'_{11} & 0 \\ \hline 0 & 0 \end{array}\right] = \begin{array}{c} \left[\begin{array}{c|c} \Phi_1R'_{11} & 0 \\ 0 & 0 \\ \hline \vdots & \vdots \\ \hline \Phi_kR'_{11} & 0 \\ 0 & 0 \\ \hline \Phi_{k+1}R'_{11} & 0 \\ 0 & 0 \end{array}\right] \end{array} \begin{array}{l} \left.\vphantom{\begin{array}{c}a\\a\end{array}}\right\} m_1 \\ \quad\vdots \\ \left.\vphantom{\begin{array}{c}a\\a\end{array}}\right\} m_k \\ \left.\vphantom{\begin{array}{c}a\\a\end{array}}\right\} m_{k+1} \end{array},$$

where each block of m_j rows correspond to one singular value σ_j (with the multiplicity m_j) of the matrix Σ. The matrix H^T represents an orthogonal linear combination of rows in each of the blocks; see (4.11). Because each $\Phi_jR'_{11}$ is of full row rank r_j, multiplication by the matrix H^T

$$H^T \Pi_L \left[\begin{array}{c|c} B_1 R'_{11} & 0 \\ \hline 0 & 0 \end{array}\right] = \left[\begin{array}{c} H_1^T \begin{bmatrix} \Phi_1 R'_{11} & 0 \\ 0 & 0 \end{bmatrix} \\ \vdots \\ H_k^T \begin{bmatrix} \Phi_k R'_{11} & 0 \\ 0 & 0 \end{bmatrix} \\ H_{k+1}^T \begin{bmatrix} \Phi_{k+1} R'_{11} & 0 \\ 0 & 0 \end{bmatrix} \end{array}\right] \begin{array}{l} \} \, m_1 \\ \vdots \\ \} \, m_k \\ \} \, m_{k+1} \end{array}, \tag{4.15}$$

can not reduce the number of nonzero rows. In particular, each block of m_j rows of (4.15) contains at least r_j nonzero rows. The permutation matrix $\widehat{\Pi}_L^T$ then moves all the nonzero rows (and possibly also some zero rows) of (4.15) to the top block $[\widehat{U}_1^T \widehat{B}_1, 0] \in \mathbb{R}^{\widehat{m} \times d}$ of (4.14). Because $\widehat{U}_1$ is square, the number of rows of $\widehat{B}_1 \in \mathbb{R}^{\widehat{m} \times \widehat{d}}$ is larger than or equal to the number of rows of $B_1 \in \mathbb{R}^{\overline{m} \times \overline{d}}$, i.e. $\widehat{m} \geq \overline{m}$.

4.3. Number of columns of the reduced system matrix. Before proving $\widehat{n} \geq \overline{n}$, we show that the orthogonal matrix $P^T \widehat{P}$ has also a special structure, analogous to (4.5). Consider the partitionings

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

Then (4.14) can be rewritten as

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

yielding the condition

$$\left(\widehat{\Pi}_2^T H^T \Pi_1\right)(B_1 R'_{11}) = 0. \tag{4.16}$$

Structures of the matrices Π_L, H, and $\widehat{\Pi}_L$ (see (3.16), (4.11), respectively (4.9)) reveal that

$$\widehat{\Pi}_2^T H^T \Pi_1$$
$$= \begin{bmatrix} [0, I_{m_1 - \widehat{r}_1}] H_1^T \left[\begin{smallmatrix} I_{r_1} \\ 0 \end{smallmatrix}\right] & 0 & 0 \\ & \ddots & \vdots \\ 0 & [0, I_{m_k - \widehat{r}_k}] H_k^T \left[\begin{smallmatrix} I_{r_k} \\ 0 \end{smallmatrix}\right] & 0 \\ 0 \quad \cdots & 0 & [0, I_{m_{k+1} - \widehat{r}_{k+1}}] H_{k+1}^T \left[\begin{smallmatrix} I_{r_{k+1}} \\ 0 \end{smallmatrix}\right] \end{bmatrix}$$

is block diagonal. Thus the condition (4.16) can be split into $k+1$ sub-conditions

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

Since $\Phi_j R'_{11}$ are of full row rank, the sub-conditions give

$$[0, I_{m_j - \widehat{r}_j}] H_j^T \begin{bmatrix} I_{r_j} \\ 0 \end{bmatrix} = 0$$

and thus also $\widehat{\Pi}_2^T H^T \Pi_1 = 0$. Consequently, from (4.13),

$$\begin{aligned} P^T \widehat{P} &= \left(\begin{bmatrix} \widehat{U}_1 & 0 \\ 0 & \widehat{U}_2 \end{bmatrix} \widehat{\Pi}_L^T H^T \Pi_L \right)^T \\ &= \begin{bmatrix} \widehat{U}_1(\widehat{\Pi}_1^T H^T \Pi_1) & \widehat{U}_1(\widehat{\Pi}_1^T H^T \Pi_2) \\ 0 & \widehat{U}_2(\widehat{\Pi}_2^T H^T \Pi_2) \end{bmatrix}^T \equiv \begin{array}{c} \begin{bmatrix} P'_{11} & 0 \\ P'_{21} & P'_{22} \end{bmatrix} \\ \begin{array}{cc} \underbrace{\quad\;}_{\widehat{m}} & \underbrace{\quad\;}_{m-\widehat{m}} \end{array} \end{array} \begin{array}{l} \} \,\overline{m} \\ \} \,m - \overline{m} \\ \\ \end{array} \end{aligned} \tag{4.17}$$

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

Finally we prove $\widehat{n} \geq \overline{n}$. Using (4.2),

$$(P^T \widehat{P})^T \left[\begin{array}{c|c} A_{11} & 0 \\ \hline 0 & A_{22} \end{array} \right] (Q^T \widehat{Q}) = \left[\begin{array}{c|c} \widehat{A}_{11} & 0 \\ \hline 0 & \widehat{A}_{22} \end{array} \right]. \tag{4.18}$$

Exploiting the structure of $P^T \widehat{P}$, see (4.17), the first $\widehat{m}$ rows of (4.18) are

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

Because $(P'_{11})^T$ has orthonormal columns and $Q^T \widehat{Q}$ is a square orthogonal matrix,

$$\begin{aligned} \operatorname{rank}(A_{11}) = \operatorname{rank}((P'_{11})^T A_{11}) &\leq \operatorname{rank}([(P'_{11})^T A_{11}, (P'_{21})^T A_{22}]) \\ &= \operatorname{rank}([\widehat{A}_{11}, 0]) = \operatorname{rank}(\widehat{A}_{11}). \end{aligned}$$

Recall that $A_{11} \in \mathbb{R}^{\overline{m} \times \overline{n}}$ is, by (CP1), of full column rank $\overline{n}$, and $\widehat{A}_{11} \in \mathbb{R}^{\widehat{m} \times \widehat{n}}$ has $\widehat{n}$ columns. Thus $\widehat{n} \geq \overline{n}$.

4.4. Summary of the development. We formulate the minimality property as a theorem.

THEOREM 4.1 (Minimality). *Consider the problem $AX \approx B$ (1.1) and its subproblem $A_{11}X_1 \approx B_1$ (3.22), where $A_{11} \in \mathbb{R}^{\overline{m} \times \overline{n}}$, $B_1 \in \mathbb{R}^{\overline{m} \times \overline{d}}$ is obtained by the transformation (3.19) described in section 3. The subproblem $A_{11}X_1 \approx B_1$ has minimal dimensions over all subproblem $\widehat{A}_{11}\widehat{X}_1 \approx \widehat{B}_1$, $\widehat{A}_{11} \in \mathbb{R}^{\widehat{m} \times \widehat{n}}$, $\widehat{B}_1 \in \mathbb{R}^{\widehat{m} \times \widehat{d}}$ obtained by an orthogonal transformation of the form (4.1), i.e. $\overline{d} \leq \widehat{d}$, $\overline{m} \leq \widehat{m}$, and $\overline{n} \leq \widehat{n}$.*

This allows to define the core problem within (1.1).

DEFINITION 4.2 (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 definition extends the definition of the core problem within the single right-hand side problems formulated by C. C. Paige and Z. Strakoš in [13]. It anticipates existence of an original problem and defines the core problem through the relationship between them. But clearly, any problem having the properties (CP1)–(CP3) can be considered as a core problem by itself. We formulate this observation as a corollary.

COROLLARY 4.3. *Any approximation problem $AX \approx B$ having the properties (CP1)–(CP3) does not allow further reduction and it therefore represents by itself a core problem.*

5. Concluding remarks. It has been shown that for any orthogonally invariant linear approximation problem of the form (1.1) there exists the fundamental data decomposition, which reveals the core problem. The presented definition of the core problem, satisfying the properties (CP1)–(CP3), extends the core problem defined for problems with $d = 1$ by C. C. Paige and Z. Strakoš in [13].

The data reduction presented here is based on the singular value decomposition of the system matrix A. The original paper [13] introduced two processes giving the core problem. The second, based on Golub-Kahan iterative bidiagonalization can also be generalized to problems with multiple right-hand sides. It leads to a banded algorithm which was for this purpose proposed by Å. Björck. The core problem concept is a useful tool in understanding the TLS problems, which was the original motivation in [13]. The data reduction and core problem based on the so-called band generalization of the Golub-Kahan iterative bidiagonalization as well as the solvability of core problems with multiple right-hand sides in the TLS sense and the relationship of the presented core problem reduction to the so-called classical TLS algoritm by S. Van Huffel and J. Vandewalle (see [18]) are under investigation and the results will be presented in the near future.

Acknowledgments. The authors thank Diana M. Sima for valuable discussions about the TLS problems theory and the core problems.

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

[11] C. C. Paige and Z. Strakoš, *Scaled total least squares funamentals*, Numerische Mathematik, 91 (2002), pp. 117–146.

[12] ———, *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.

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

[14] M. Plešinger, *The Total Least Squares Problem and Reduction of Data in $AX \approx B$*, Ph.D. thesis, Faculty of Mechatronics and Interdisciplinary Engineering Studies, Technical University of Liberec, Liberec, Czech Rebuplic, 2008.

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

[16] D. M. Sima and S. Van Huffel, *Core problems in $AX \approx B$*, Technical Report, Dept. of Electrical Engineering, Katholieke Universiteit Leuven (2006).

[17] A. van der Sluis, *Stability of the solutions of linear least squares problems*, Numerische Mathematik, 23 (1975), pp. 241–254.

[18] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM Publications, Philadelphia, PA, 1991.

[19] 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.

[20] S. Van Huffel and P. Lemmerling, eds., *Total Least Squares and Errors-in-Variables Modeling*: *Analysis*, *Algorithms and Applications*, Dordrecht, Kluwer Academic Publishers, 2002.