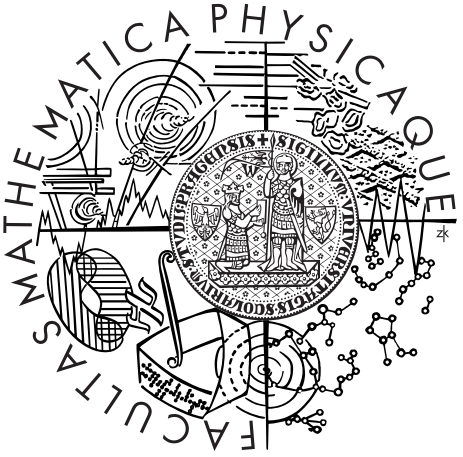
Charles University
Faculty of Mathematics and Physics

HABILITATION THESIS



Iveta Hnětynková

Error contaminated linear approximation problems: Analysis and methods

Department of Numerical Mathematics

Prague 2019

WWW: http://www.ms.mff.cuni.cz/~hnetynka
e-mail: iveta.hnetynkova@mff.cuni.cz

Contents

Preface

This thesis studies linear approximation problems

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

where the aim is to approximate the unknown matrix $X \in \mathbb{R}^{n\times d}$ from the given model A and the observation (measurement) B. Such problems arise in various areas and are known under various names, e.g., linear data fitting or errors-in-variables modeling. Depending on the particular application, the system above typically suffers from the presence of various types of errors in the observation matrix B and/or the matrix A representing the model. Furthermore, many troublesome properties complicating the solution can be present, e.g., sensitivity of the solution X to small changes of data, severe ill-conditioning of the matrix A, etc. Thus large effort has been devoted to the analysis and development of solution approaches for such classes of problems. This thesis presents a collection of 11 papers contributing to this area coauthored by Iveta Hnětynková in various collaborations, in particular with M. Plešinger (TU Liberec) and Z. Strakoš (Charles University). Other coauthors are M. Kubínová (Czech Academy of Sciences), J. Mead (Boise State University), R. Renaut (Arizona State University), D. M. Sima (KU Leuven), S. Van Huffel (KU Leuven), J. Žáková (TU Liberec).

The principal difficulty in data fitting problems is the presence of factors (e.g. errors in the data) that make the problem incompatible. Thus a general question arises in which sense the problem should be solved. Least squares methods [3], [29] are the most widely used, while the total least squares (TLS) [14] is the method of choice when the errors are present both in A and B. Since TLS looks simultaneously for the minimal correction of A and B making the corrected problem compatible, its analysis and solution is significantly more complicated than for the well known ordinary least squares (LS) and this holds even for the case of single observation. Already in the early work [14] it was shown that the TLS solution may not exist. The TLS problem for $d = 1$ was there analyzed through the singular value decompositions (SVD) of the matrices A and $[b, A]$ leading to a sufficient (but not necessary) GVL condition for existence of a TLS solution. Having multiple observations for the same model A available, i.e. $d > 1$, the TLS problem complicates even more significantly both in terms of existence of the solution and computational approaches. The additional difficulty is related to the fact that the individual observations (columns of B) can be closely correlated with different subsets of columns of A. The TLS problem with $d > 1$ was studied in [52] assuming special distribution of singular values of the matrix $[B, A]$, followed by [57], [58] and many others. The core reduction introduced in [36], [37] for $d = 1$ brought in a completely different concept. It suggested to extract the necessary and sufficient

information for solving the problem contained in the data A and b, while removing the redundant and/or irrelevant information. Thus it represents a crucial step towards understanding, analysis and solution of TLS problems with a single right-hand side. While [36], [37] presents a complete theory including necessary and sufficient conditions for the existence of the TLS solution for a single observation vector, extensions for the multiple observation case ($d > 1$) remained unresolved.

The first part of this thesis summarizes our main contributions to the analysis of existence and uniqueness of solutions of TLS problems with multiple observations, and to generalizations of the core problem concept to such problems. It includes seven papers listed below:

[C1] 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.
DOI: 10.1137/100813348

[C2] I. Hnětynková, M. Plešinger, and Z. Strakoš, *The core problem within a linear approximation problem $AX \approx B$ with multiple right-hand sides*, SIAM J. Matrix Anal. Appl. 34 (2013), pp. 917–931.
DOI: 10.1137/120884237

[C3] I. Hnětynková, M. Plešinger, and Z. Strakoš, *Band generalization of the Golub–Kahan bidiagonalization, generalized Jacobi matrices, and the core problem*, SIAM J. Matrix Anal. Appl. 36 (2015), pp. 417–434.
DOI: 10.1137/140968914

[C4] I. Hnětynková, and M. Plešinger, *Complex wedge-shaped matrices: A generalization of Jacobi matrices*, Linear Algebra Appl. 487 (2015), pp. 203–219.
DOI: 10.1016/j.laa.2015.09.017

[C5] I. Hnětynková, M. Plešinger, and D. M. Sima, *Solvability of the core problem with multiple right-hand sides in the TLS sense*, SIAM J. Matrix Anal. Appl. 37(3) (2016), pp. 861–876.
DOI: 10.1137/15M1028339

[C6] I. Hnětynková, M. Plešinger, and J. Žáková, *TLS formulation and core reduction for problems with structured right-hand sides*. Linear Algebra Appl. 555 (2018), pp. 241–265.
DOI: 10.1016/j.laa.2018.06.016

[C7] I. Hnětynková, M. Plešinger, and J. Žáková, *On TLS formulation and core reduction for data fitting with generalized models*. Linear Algebra Appl. 577 (2019), pp. 1–20.
DOI: 10.1016/j.laa.2019.04.018

The paper [C1] provides a full classification of TLS problems with respect to the (non)existence of the solution covering all possible cases. The paper [C2] presents generalization of the core reduction to problems with $d > 1$. Its determination by a

band (or block) generalization of the Golub-Kahan iterative bidiagonalization with deflations is described in [C3]. This paper also introduces generalized Jacobi matrices that can be used to study band algorithms. The paper [C4] provides analysis of eigenvalues and eigenvectors of these generalized Jacobi matrices. Existence of TLS solutions of core problems with multiple right-hand sides is studied in [C5]. The papers [C6] and [C7] present further generalizations of the TLS formulation and core reduction to approximation problems where the observation matrix B or the matrix A of the model has tensor (or some other special) structure.

The second part of this thesis is devoted to linear approximation problems $AX \approx B$ that are ill-posed. Such problems arise in various applications, e.g. in image processing, seismology, etc., see [15], [18] for an overview. It was observed in the early work [46] that the solution X is then extremely sensitive to errors present in the data, in particularly in the observation matrix B. Thus classical methods such as LS or TLS produce for this class of problems outputs dominated by the amplified errors (noise), and regularization is necessary. In truncated LS (called also TSVD) or truncated TLS [10] (called TTLS or RTLS), regularization is achieved by replacing the matrix A (in TSVD) or $[B, A]$ (in TTLS) by their low rank approximations. Alternatively, a Tikhonov regularization can be used in the LS formulation, see [46] and [47]. As shown in [11], [20, Chap. 6] for $d = 1$, these methods can be interpreted as dumping the unwanted SVD components of the LS solution by the so called filter factors. While generalization of filter factors to $d > 1$ for TSVD or Tikhonov regularization is straightforward, such representation for TTLS was not known. For large scale and also matrix-free problems, where the matrix is not given explicitly but only in terms of a function allowing for matrix-vector multiplication, iterative regularization is needed. Widely used class of such methods including CRAIG [7], LSQR [35], LSMR [12] (and also their hybrid variants [24], [8], etc.), constructs a sequence of approximate solutions on Krylov subspaces of increasing dimensions, with the bases computed by the Golub-Kahan iterative bidiagonalization [13]. As explained in [18], early termination of the iterations is necessary, because the larger dimensional Krylov subspaces tend to be contaminated by amplified noise. This well known phenomenon called semiconvergence requires to study propagation of noise into bidiagonalization vectors and into the approximate solutions.

The second part of this thesis summarizes our main contributions to the analysis of regularization properties of the TTLS and bidiagonalization-based methods for the solution of linear ill-posed problems. It includes four papers listed below:

[C8] I. Hnětynková, M. Plešinger, and Z. Strakoš, *The regularizing effect of the Golub-Kahan iterative bidiagonalization and revealing the noise level in the data*, BIT 49 (4) (2009), pp. 669–696.
DOI: 10.1007/s10543-009-0239-7

[C9] R. Renaut, I. Hnětynková, and J. Mead, *Regularization parameter estimation for large-scale Tikhonov regularization using a priori information*, Computational Statistics and Data Analysis 54 (2010), pp. 3430–3445.
DOI: 10.1016/j.csda.2009.05.026

[C10] I. Hnětynková, M. Kubínová, M. Plešinger, *Noise representation in residuals of CRAIG, LSQR and LSMR regularization*, Linear Algebra Appl. 533 (2017), pp. 357–379.
DOI: 10.1016/j.laa.2017.07.031

[C11] I. Hnětynková, M. Plešinger, and J. Žáková, *Filter factors of truncated TLS regularization with multiple observations*. Applications of Mathematics 62 (2017), pp. 105–120.
DOI: 10.21136/AM.2017.0228-16

The papers [C8]-[C10] are devoted to regularization based on the Golub-Kahan iterative bidiagonalization for problems with $d = 1$. Propagation of noise from the observation vector b to the basis vectors generated by the bidiagonalization is analyzed in [C8]. Furthermore, a method for cheap approximation of the unknown noise level (amount of noise in b) is presented. The paper [C9] introduces a hybrid regularization method combining the outer Golub-Kahan iterative bidiagonalization with the inner regularization of the projected system for the solution of Tikhonov minimization problem with a general weighted norm. Regularization properties of CRAIG, LSQR and LSMR are studied in [C10] by deriving explicit relations between their residuals and noise-contaminated bidiagonalization vectors. The paper [C11] shows that TTLS for problems with $d > 1$ can be represented as a filtering method and derives the corresponding filter factors.

In spite of the fact that the presented analysis assumes exact arithmetic (whose simulation would require, e.g., reorthogonalization in the Golub-Kahan bidiagonalization), most of the results are applicable to practical computations and this is supported by the numerical experiments included in the papers. Rigorous numerical analysis of rounding errors represents a nontrivial challenge for the future.

Chapter 1

Analysis of TLS and core reduction for problems with multiple observations

1.1 Introduction to least squares methods

In the most common situation we are given for the linear model A one observation (measurement) vector b leading to the single (or vector) right-hand side formulation

$$Ax \approx b, \quad \text{where} \quad A \in \mathbb{F}^{m\times n}, \quad x \in \mathbb{F}^n, \quad b \in \mathbb{F}^m \tag{1.1}$$

and $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. For some models, however, repeated measurements are possible that can be organized into an observation matrix of the form $B = [b_1, \ldots, b_d]$. In that case we arrive at a more general multiple (or matrix) right-hand side problem

$$AX \approx B, \quad \text{where} \quad A \in \mathbb{F}^{m\times n}, \quad X \in \mathbb{F}^{n\times d}, \quad B \in \mathbb{F}^{m\times d}. \tag{1.2}$$

Each column of X then represents the unknown vector corresponding to the individual observation in the column of B. Several nonrestrictive assumptions simplifying the exposition can be made. If not stated otherwise, we assume $\mathbb{F} = \mathbb{R}$, range of B is not in the range of A (otherwise the system is compatible and we can look for the exact solution), $A^T B \neq 0$ (otherwise the model and the observations are uncorrelated and the only meaningful solution is a zero matrix $X = 0$), and $m \geq n + d$ (otherwise we can formally add zero rows).

Assuming errors both in A and B, the TLS formulation searches for correction matrices E of A and G of B minimizing the Frobenius norm in

$$\min_{X,E,G} \left\| \left[\begin{array}{c|c} G & E \end{array}\right] \right\|_F \qquad \text{subject to} \qquad (A+E)X = B + G, \tag{1.3}$$

and any X solving the corrected system above is called the TLS solution. (Minimization in other norms is also possible, see [30] or [56].) The TLS formulation differs from the commonly known (ordinary) LS

$$\min_{X,G} \| \, G \, \|_F \qquad \text{subject to} \qquad AX = B + G, \tag{1.4}$$

where the correction is restricted only to the observation matrix B.

The ordinary LS can be directly reformulated as a search for X minimizing the residual norm $\|B - AX\|_F = \|R\|_F$, which immediately implies that a solution always exists and can be found by various methods; for an overview see [3], [29]. If A has full column rank, this solution is moreover unique, otherwise a minimum norm solution can be constructed. Furthermore, the problem (1.4) is equivalent to the set of the individual LS problems within

$$Ax_i \approx b_i, \quad i = 1, \ldots, d,$$

since the corrections of the individual columns of B are independent. This important property allows to analyze and solve the multiple right-hand side problem (1.2) in the LS sense using the tools for single right-hand side problems (1.1). This is however not true for TLS, where the correction E is common to all observations.

1.2 Analysis of existence and uniqueness of TLS solution

The TLS problem has been investigated for decades under various names such as orthogonal regression or errors-in-variables modeling, see [54] for some examples. Consider first the single right-hand side problem (1.1). Rewriting this equivalently as

$$[b|A] \begin{bmatrix} -1 \\ x \end{bmatrix} \approx 0$$

reveals that TLS looks for the minimal correction $[g|E]$ of the extended matrix $[b|A]$ such that there exists a vector with the nonzero first entry in the null space of the corrected matrix $[b + g|A + E]$. G. Golub and C. Van Loan shown already in [14] that a TLS solution for $d = 1$ may not exist and it may not be unique. They also proved that if the smallest singular value of A is strictly larger than the smallest singular value σ_{n+1} of $[b|A]$ (the so called GVL condition), then σ_{n+1} is simple and the corresponding right singular vector v_{n+1} has a nonzero first component. Thus the vector

$$x = -\frac{1}{v^{(1)}} v^{(2)}, \quad \text{where} \quad v_{n+1} = \begin{bmatrix} v^{(1)} \\ v^{(2)} \end{bmatrix} \begin{matrix} \}1 \\ \}n \end{matrix},$$

represents the unique TLS solution. It is worth to emphasize that the GVL condition is sufficient for existence of a TLS solution, but not necessary. General analysis requires to study the SVD of $[b|A]$, where the key role is played by the singular values and the corresponding right singular subspaces. In the case of nonexistence of a TLS solution, S. Van Huffel and J. Vandewalle defined in the book [52, Sect. 3.4] a vector called "nongeneric solution". This vector can be interpreted as a minimizer of the TLS problem extended by an additional constraint, see [52, Def. 3.2]. The TLS problem with $d = 1$ was then analyzed by C. Paige and Z. Strakoš in [36] revealing principal difficulties connected with the generic-nongeneric terminology and clarifying the meaning of the nongeneric solution. This was done by introducing the core reduction concept, which we explain in the next section. Analysis of problems with $d > 1$ remained, however,

unfinished.

The multiple right-hand side problem (1.2) can be similarly rewritten into the form

$$[B|A]\begin{bmatrix} -I_d \\ X \end{bmatrix} \approx 0,$$

where $I_d \in \mathbb{R}^{d\times d}$ is the identity matrix, revealing that in TLS we are looking for the minimal correction of the extended matrix $[B|A]$ such that its null space has special properties. The book [52] analyzed existence and uniqueness of a solution of (1.3), but only for two particular cases with a prescribed distribution of singular values of the extended data matrix $[B|A]$. To be more precise, consider the SVD

$$[B|A] = U\Sigma V^T, \qquad U \in \mathbb{R}^{m\times m},\ V \in \mathbb{R}^{(n+d)\times(n+d)},\ \Sigma \in \mathbb{R}^{m\times(n+d)}, \tag{1.5}$$

where U and V are unitary and Σ is a diagonal matrix with the singular values

$$\sigma_1 \geq \ldots \geq \sigma_n \geq \sigma_{n+1} \geq \ldots \geq \sigma_{n+d} \geq 0,$$

on its diagonal. The solvability of the TLS problem depends particularly on the singular value σ_{n+1} and the corresponding right singular subspace. In [52] it was shown that if $\sigma_n > \sigma_{n+1}$ or if there exists some index $p \leq n$ such that $\sigma_p = \cdots = \sigma_{n+1} = \cdots = \sigma_{n+d}$, then the problem (1.3) has a TLS solution. The "nongeneric solution" for $d > 1$ was also defined in [52, Def. 3.3]. See also [57] and [58] studying TLS problems with nonunique solutions and providing various relations between LS and TLS solutions.

Our paper [C1] revealed that full analysis of existence and uniqueness of a TLS solution requires careful handling of the multiplicity of σ_{n+1}. Thus we define integers q, $n \geq q \geq 0$, and e, $d \geq e \geq 1$, as

$$\sigma_{n-q} > \underbrace{\sigma_{n-q+1} = \ldots = \sigma_n}_{q} = \underbrace{\sigma_{n+1} = \ldots = \sigma_{n+e}}_{e} > \sigma_{n+e+1}, \tag{1.6}$$

(σ_{n-q} or σ_{n+e+1} may be undefined when $q = n$ or $e = d$, respectively). Consider the corresponding partitioning of the matrix of right singular vectors,

$$V = \begin{array}{c} \left[\begin{array}{ccc} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \end{array}\right] \begin{array}{l} \}d \\ \}n \end{array} \\ \underbrace{}_{n-q}\ \underbrace{}_{q+e}\ \underbrace{}_{d-e} \end{array}. \tag{1.7}$$

In words, the block-columns correspond to singular values strictly larger than, equal to, and strictly smaller than σ_{n+1}, respectively (again V_{11}, V_{21} or V_{13}, V_{23} may be undefined). The novel classification introduced in [C1] distinguishes four disjoint sets $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$, and $\mathcal{S}$ of TLS problems based on ranks of individual blocks in (1.7) covering all possible cases. For each of the classes, it is analyzed whether a TLS solution exists and whether it is unique. In particular:

$\mathcal{F}$: If $rank([V_{12}, V_{13}]) = d$, then the problem belongs to the 1st class, where:

- $\mathcal{F}_1$: If $rank(V_{12}) = e$, $rank(V_{13}) = d - e$, then a TLS solution exists (unique or nonunique). If it is nonunique, the TLS solution minimal in the Frobenius and at the same time in the 2-norm can be found.

$\mathcal{F}_2$: If $rank(V_{12}) > e$, $rank(V_{13}) = d - e$, then there exist infinitely many TLS solutions. The TLS solutions minimal in the Frobenius and 2-norms may not be the same.

$\mathcal{F}_3$: If $rank(V_{12}) > e$, $rank(V_{13}) < d - e$, then a TLS solution does not exist.

$\mathcal{S}$: If $rank([V_{12}, V_{13}]) < d$, then the problem belongs to the 2nd class. A TLS solution does not exist.

Note that single right-hand side problems (1.1) belong either to $\mathcal{F}_1$ or $\mathcal{S}$. This classification again reflects several difficulties (revealed for $d = 1$ in [36]) with the generic-nongeneric terminology used originally for TLS problems, that still appear in some publications. For example, the generic problems are often considered as TLS solvable. However, since they correspond to the problems of the 1st class, their solvability is not ensured unless they satisfy additional properties (see the class $\mathcal{F}_3$).

The widely used computational tool for solving TLS problems is the TLS algorithm [51]. It was shown in [C1] that for problems in the set $\mathcal{F}_2$ the TLS algorithm cannot reach the existing TLS solutions. The output matrix X of the algorithm is not a solution of (1.3) but a solution of a problem modified by a specific constraint, see [C1, Sect. 6].

1.3 Core reduction and block generalization of the Golub-Kahan bidiagonalization

As illustrated in the previous section, classification of TLS problems is complicated requiring distinguishing various cases based on the SVD information. Different point of view was given by the core reduction concept introduced by C. Paige and Z. Strakoš in [36], [37] for problems with single right-hand side, i.e. (1.1). In was shown how the original problem can be orthogonally transformed to a special block form allowing for reduction of the data by separating the necessary and sufficient information for solving the problem (1.1) from the rest. More precisely, for any data $[b|A]$ there exist orthogonal matrices $P \in \mathbb{R}^{m\times m}, Q \in \mathbb{R}^{n\times n}$, such that

$$P^T[b|A]\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & Q^T \end{bmatrix}\begin{bmatrix} -1 \\ x \end{bmatrix} = \left[\begin{array}{c|c|c} b_1 & A_{11} & 0 \\ \hline 0 & 0 & A_{22} \end{array}\right]\begin{bmatrix} -1 \\ x_1 \\ x_2 \end{bmatrix} \approx 0, \tag{1.8}$$

where the matrix A_{11} has smallest possible dimensions among all matrices that can be obtained by the orthogonal transformation of the form above. Now, since the TLS problem (1.3) is orthogonally invariant, instead of solving (1.1) we can solve

$$\left[\begin{array}{c|c} A_{11} & 0 \\ \hline 0 & A_{22} \end{array}\right]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \approx \left[\begin{array}{c} b_1 \\ \hline 0 \end{array}\right]. \tag{1.9}$$

Because of its block structure, the problem (1.9) decomposes into two independent subproblems $A_{11}x_1 \approx b_1$ and $A_{22}x_2 \approx 0$, where for the second one it is reasonable to put $x_2 = 0$ and only the first one called the core problem has to be solved. The vector x can then be obtained simply by the back transformation of variables

$$x = Q\begin{bmatrix} x_1 \\ 0 \end{bmatrix}. \tag{1.10}$$

Moreover, it was proved in [37] that the core problem has various properties implying that it satisfies the GVL condition. Consequently, it always has the unique TLS solution. In this way, core reduction simplifies the TLS approach by separating the only TLS-meaningful information from A and b.

As shown in [37] and [23], the core problem for $d = 1$ can be constructed either directly by using the SVD or iteratively by the Golub-Kahan iterative bidiagonalization [13]. Putting the starting vectors $w_0 = 0$ and $s_1 = b/\beta_1$, where $\beta_1 = \|b\|$, the algorithm computes for $k = 1, 2, \ldots$

$$\alpha_k w_k = A^T s_k - \beta_k w_{k-1}\,, \qquad \|w_k\| = 1, \tag{1.11}$$

$$\beta_{k+1} s_{k+1} = A w_k - \alpha_k s_k\,, \qquad \|s_{k+1}\| = 1, \tag{1.12}$$

where $\alpha_k \geq 0$, $\beta_{k+1} \geq 0$ are the normalization coefficients. The vectors $s_k \in \mathbb{R}^m$ are mutually orthogonal and the same holds for the vectors $w_k \in \mathbb{R}^m$. The process continues until $\alpha_k = 0$ for incompatible problems (1.1), or until $\beta_{k+1} = 0$ for compatible problems. In the incompatible case of our interest, the core data are then given by the coefficients

$$[b_1|A_{11}] = \left[\begin{array}{c|ccc} \beta_1 & \alpha_1 & & \\ & \beta_2 & \ddots & \\ & & \ddots & \alpha_{k-1} \\ & & & \beta_k \end{array}\right].$$

Noting that the matrices $A_{11}A_{11}^T$ and $[b_1|A_{11}]^T[b_1|A_{11}]$ are Jacobi matrices, i.e. symmetric tridiagonal matrices with positive subdiagonal entries, various properties of the core problem can be proved through spectral properties of Jacobi matrices, see our paper [23] (not included in this thesis).

In the view of these results the natural question arises whether a similar data reduction can be done in the multiple right-hand side case (1.2). The first attempts can be found in [4], [44], and [38]. The task was then resolved in our paper [C2], where the SVD-based core reduction was presented, followed by the paper [C3] providing iterative algorithm for construction of the multi-observation core problem. It was proved that for any data $[B|A]$ there exist (in this case three) orthogonal matrices $P \in \mathbb{R}^{m\times m}, Q \in \mathbb{R}^{n\times n}, R \in \mathbb{R}^{d\times d}$ such that

$$P^T[B|A]\begin{bmatrix} R & 0 \\ 0 & Q \end{bmatrix} = P^T[BR|AQ] \equiv \left[\begin{array}{c|c||c|c} B_1 & 0 & A_{11} & 0 \\ \hline 0 & 0 & 0 & A_{22} \end{array}\right], \tag{1.13}$$

where A_{11} and B_1 are minimally dimensioned subject to the orthogonal transformation above. Taking the advantage of orthogonal invariance of TLS, we can again solve the problem

$$\left[\begin{array}{c|c} A_{11} & 0 \\ \hline 0 & A_{22} \end{array}\right]\begin{bmatrix} X_1 & Z_1 \\ X_2 & Z_2 \end{bmatrix} \approx \left[\begin{array}{c|c} B_1 & 0 \\ \hline 0 & 0 \end{array}\right],$$

decomposing into four independent subproblems

$$A_{11}X_1 \approx B_1, \qquad \text{and} \qquad A_{11}Z_1 \approx 0, \quad A_{22}X_2 \approx 0, \quad A_{22}Z_2 \approx 0.$$

Putting naturally $X_2 \equiv 0$, $Z_1 \equiv 0$, and $Z_2 \equiv 0$, only the core problem $A_{11}X_1 \approx B_1$ has to be solved, complemented by the back transformation

$$X \equiv Q \begin{bmatrix} X_1 & 0 \\ 0 & 0 \end{bmatrix} R^T. \tag{1.14}$$

Consequently, the solution of (1.2) is again fully determined by the solution of the extracted minimal subproblem (1.14). Note that any problem $A_{11}X_1 \approx B_1$ is a core problem if and only if it satisfies the following properties (see [C2, Sect. 4]):

(CP1) The matrix A_{11} is of *full column rank*.

(CP2) The matrix B_1 is of *full column rank*.

(CP3) Let A_{11} have ξ distinct nonzero singular values with multiplicities μ_i, let $\mu_{\xi+1}$ be the dimension of the null space of A_{11}^T, and let U_i be matrices having orthonormal bases of left singular vector subspaces of A_{11} as their columns.

Then the matrices $U_i^T B_1$ are of *full row rank* μ_i, for $i = 1, \ldots, \xi, \xi+1$.

Thus (CP1)-(CP3) characterize fully the core problem.

The core problem for $d > 1$ can be determined iteratively by the so-called band (or block) generalization of the Golub-Kahan iterative bidiagonalization, see [C3]. Since the algorithm is in its complete form complicated, we explain only the main idea. Instead of the vectors and numbers in (1.11)-(1.12), the generalized algorithm works with blocks of vectors and with matrices. Starting with the whole matrix B, it produces in each step a set of maximally d vectors w_k and maximally d vectors s_k such that w_k are mutually orthogonal and s_k are mutually orthogonal. Normalization steps in (1.11)-(1.12) are replaced by QR-factorizations of square blocks of the size maximally $d \times d$, where their rank must be controlled. Here the so called (upper and lower) deflation plays a crucial role due to possible zero entries present in these blocks reducing the band shape of the resulting matrix. As derived in [C3, Sect. 3], after d deflations we obtain the core data in the form of a band (or equivalently block bidiagonal) matrix, e.g.,

$$[B_1||A_{11}] = \left[\begin{array}{ccc||ccc|cc|c} \beta_1 & \delta_{1,2} & \delta_{1,3} & \alpha_1 & & & & & \\ & \beta_2 & \delta_{2,3} & \delta_{2,1} & \alpha_2 & & & & \\ & & \beta_3 & \delta_{3,1} & \delta_{3,2} & \alpha_3 & & & \\ \hline & & & \beta_4 & \delta_{4,2} & \delta_{4,3} & \alpha_4 & & \\ & & & & & \beta_5 & \delta_{5,4} & \alpha_5 & \\ \hline & & & & & & \beta_6 & \delta_{6,5} & \\ & & & & & & & \beta_7 & \alpha_6 \end{array}\right].$$

In order to study properties of the core problem above, motivated by the derivations for $d = 1$, we look at the matrices $A_{11}A_{11}^T$ and $[B_1|A_{11}]^T[B_1|A_{11}]$. For the example above

$$A_{11}A_{11}^T = \left[\begin{array}{ccc|c|cc|c} \heartsuit & \heartsuit & \heartsuit & \alpha_1\beta_4 & & & \\ \heartsuit & \heartsuit & \heartsuit & \heartsuit & & & \\ \heartsuit & \heartsuit & \heartsuit & \heartsuit & \alpha_3\beta_5 & & \\ \hline \alpha_1\beta_4 & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \alpha_4\beta_6 & \\ & & \alpha_3\beta_5 & \heartsuit & \heartsuit & \heartsuit & \\ \hline & & & \alpha_4\beta_6 & \heartsuit & \heartsuit & \alpha_5\beta_7 \\ \hline & & & & & \alpha_5\beta_7 & \heartsuit \end{array}\right] \equiv \begin{bmatrix} \Delta_1 & \Xi_1^T & & \\ \Xi_1 & \Delta_2 & \Xi_2^T & \\ & \Xi_2 & \Delta_3 & \Xi_3^T \\ & & \Xi_3 & \Delta_4 \end{bmatrix}, \tag{1.15}$$

where $\heartsuit$ are in general nonzero entries. The blocks Δ_ℓ are symmetric, and Ξ_ℓ are full row rank blocks in upper triangular row echelon forms. Consequently, the matrices (1.15) can be considered a structural generalization of Jacobi matrices. We showed in [C3] that they are related to the band matrices produced by the generalized band algorithm in a similar way as the bidiagonal matrices produced by (1.11)-(1.12) are related to the Jacobi matrices. We called the matrices of the form (1.15) *wedge-shaped* and analyzed their spectral properties in a separate paper [C4] revealing bounds for the multiplicities of their eigenvalues, special nonzero structure of entries of their eigenvectors, etc. Note that various generalized band algorithms are used also in other contexts, e.g., for the solution of systems of linear algebraic equations with a square nonsingular matrix A and a matrix right-hand side. Deflation is in analysis of their convergence typically neglected. We believe that wedge-shaped matrices can be used as a geral tool to study behavior of band algorithms including deflation.

As showed in [36], for $d = 1$ the core problem is always uniquely TLS solvable. In the classification of TLS problems (see [C1]), it belongs to the set $\mathcal{F}_1$. Solvability of the core problem for $d > 1$ was investigated later in [C5]. We proved that the core problem has the unique TLS solution if and only if it belongs to the set $\mathcal{F}_1$. We further showed by construction that it is possible to find a core problem (i.e. a problem satisfying (CP1)-(CP3)) in any of the classification sets $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \mathcal{S}$. Consequently, a core problem with multiple right-hand sides may not have a TLS solution. To understand the cases where this happens, we studied possible internal structure of core problems. It was shown that a core problem with $d > 1$ can be composed of several uncorrelated subproblems, e.g.

$$A_{11}X_1 \approx B_1, \quad \text{where} \quad A_{11} = \begin{bmatrix} A_{11}^{\alpha} & 0 \\ 0 & A_{11}^{\beta} \end{bmatrix}, \quad B_1 = \begin{bmatrix} B_1^{\alpha} & 0 \\ 0 & B_1^{\beta} \end{bmatrix}. \tag{1.16}$$

Even when the subproblems (so called components)

$$A_{11}^{\alpha}X_1^{\alpha} \approx B_1^{\alpha}, \quad A_{11}^{\beta}X_1^{\beta} \approx B_1^{\beta}$$

are uniquely TLS solvable, depending on the relations among their singular values, the composed core problem (1.16) still may not have a TLS solution. In that case, it would be reasonable to consider the individual components separately, as the data are obviously independent. Computational technique for separation of these components from the data A_{11}, B_1 is, however, still an open question.

Even though the core reduction for $d > 1$ does not necessarily lead to $\mathcal{F}_1$ problems, it still represents a tool for maximal possible reduction of the dimensions of the data without loss of information. In [C5] we also proved that the output matrix of the TLS algorithm [51] for the core problem with $d \geq 1$ after the change of variables (1.14) equals the output of this algorithm for the original problem (1.2), which is an important consistency result. (For $d = 1$ this follows already from [37]). Here it is particularly interesting to look at the outputs of the TLS algorithm for composed core problems and their components. Depending on the properties of components, the output matrix for the core problem can be, e.g., a direct sum of the outputs for individual components, or it can be fully determined by the output for one of the components if this component is in some sense significantly more important (for example in terms of the size of its

singular values). This can be seen as an automatic inner regularization in the TLS algorithm, where some part of information is removed from $[B|A]$ during the solution process, see [C5] for detailed explanation.

1.4 Generalizations of the problem setting

Some applications lead to linear approximation problems having more complicated structure, see for example [40] (where the model depends on multiple parameters), [34], or [45] (where observations depend on time). Assume first that the observation vector depends on $k-1$ parameters, each of them is sampled to d_j samples and for each combination we save the observation vector. In fact we have a set of multiple observations forming naturally a k-way tensor (multidimensional array) $\mathcal{B}$ of dimensions $m \times d_2 \times \cdots \times d_k$. This yields a linear approximation problem of the form

$$A \times_1 \mathcal{X} \approx \mathcal{B}, \tag{1.17}$$

where A is a matrix and $\mathcal{X}$ is a tensor of unknowns. The 1-mode matrix-tensor product "$\times_1$" above is defined with reversed order of operands in comparison to the standard notation given in [2], to keep the ordering of objects in equations $Ax \approx b$, $AX \approx B$, and $A \times_1 \mathcal{X} \approx \mathcal{B}$ consistent.

In order to solve the problem (1.17) in terms of TLS, it seems enough to unfold simultaneously $\mathcal{B}$ and $\mathcal{X}$ into a matrix by putting the individual column vectors behind each other and use approaches derived for the matrix problem (1.2). However, our analysis of (1.17) in [C6] showed why removing the parameter-given tensor structure may not be appropriate. We introduced the TLS minimization directly for the tensor observation problem as follows

$$\min_{\substack{\mathcal{G} \in \mathbb{F}^{m \times d_2 \times \cdots \times d_k} \\ E \in \mathbb{F}^{m \times n}}} (\|\mathcal{G}\|^2 + \|E\|_F^2)^{\frac{1}{2}} \quad \text{subject to} \quad (A+E) \times_1 \mathcal{X} = \mathcal{B} + \mathcal{G}, \tag{1.18}$$

where $\|\mathcal{G}\|$ is a tensor norm defined as the square root of the sum of squares of all entries of $\mathcal{G}$, i.e., a straightforward generalization of the matrix F-norm. For a tensor $\mathcal{C}$ denote by $\mathcal{C}^{\{1\}}$ a matrix obtained by its 1-mode matricization. Then the tensor right-hand side TLS problem (1.18) is equivalent to the matrix right-hand side TLS problem (1.3) for

$$AX \approx B, \quad \text{where} \quad B \equiv \mathcal{B}^{\{1\}}, \quad d \equiv \prod_{j=2}^{k} d_j. \tag{1.19}$$

This means that $\mathcal{X}$ is a tensor TLS solution of (1.18) if and only if $X \equiv \mathcal{X}^{\{1\}}$ is a matrix TLS solution of (1.3), see [C6, Sect. 4.1]. While the basic solvability analysis can be easily translated from (1.3) to (1.18) using this equivalence, this is not true for the core reduction.

In [C6] an orthogonal reduction is derived extracting the necessary and sufficient information from (1.17) into a small subproblem

$$A_{11} \times_1 \mathcal{X}_1 \approx \mathcal{B}_1, \tag{1.20}$$

while maintaining the tensor structure. This tensor core reduction uses heavily the Tucker decomposition [48]-[50] generalizing the SVD to tensors. The tensor core problem can be characterized by three properties, similarly as the matrix core problem:

(CP1) The matrix A_{11} is of full column rank.

(CP2) The j-mode matricization $\mathcal{B}_1^{\{j\}}$ is of full row rank, or equivalently, all j-mode co-fibers of $\mathcal{B}$ are linearly independent, $j = 2, \ldots, k$.

(CP3) Let A_{11} have ξ distinct nonzero singular values with multiplicities μ_i, let $\mu_{\xi+1}$ denote the dimension of the null space of A_{11}^T, and let U_i be matrices having orthonormal bases of left singular vector subspaces of A_{11} as their columns.

Then the matrices $U_i^T \mathcal{B}_1^{\{1\}}$ are of full row rank μ_i, for $i = 1, \ldots, \xi, \xi + 1$.

The tensor core problem (1.20) is generally different from the matrix core problem for the matricized data (1.19), the dimensions of (1.20) are typically larger, see [C6] for a detailed discussion.

Another extension of the matrix approximation problem (1.2) is discussed in [C7]. Using the vectorization of B and X (i.e. the columns of a matrix are stacked into a long vector), (1.2) rearranges to

$$(I \otimes A)\,\mathrm{vec}(X) \approx \mathrm{vec}(B), \tag{1.21}$$

where $\otimes$ is the Kronecker product. The corrected problem in (1.3) similarly takes the form

$$\Big(\underbrace{(I \otimes A)}_{\mathcal{A}} + \underbrace{(I \otimes E)}_{\mathcal{E}} \Big)\,\mathrm{vec}(X) = \mathrm{vec}(B) + \mathrm{vec}(G)$$

revealing that the perturbation $\mathcal{E}$ in TLS follows the Kronecker-product structure of the model $\mathcal{A}$. This restriction of the search set is the key factor limiting the TLS solvability of (1.3), see [C1]. In [C7], we showed how the search set is enriched when moving from matrix to more general models, and how this influences the TLS formulation. We staid with the matrix observation B and studied various extensions of the model A:

- the bilinear model represented by a pair of matrices (for applications see [26], [27]),
- the model of higher Kronecker rank,
- and the fully tensor model $\mathcal{A}$.

In particular, for fully general tensor models the search set is so rich that by vectorization the problem can be transformed to the basic single right-hand side TLS problem. Consequently, core reduction leading to a core problem with the unique TLS solution is available here. Fully tensor problems, where both the model and observations are represented by tensors $\mathcal{A}$ and $\mathcal{B}$ are currently under investigation.

Chapter 2

Analysis and development of regularization methods for linear ill-posed problems

2.1 Introduction to ill-posed problems

In many applications one finds inverse problems modeled as Fredholm integral equations of the first kind with a kernel having smoothing properties; see, e.g., [15] and [18] for various examples in image processing, geophysics, etc. After discretization, we obtain a problem (1.1), where typically:

- Singular values of A decay gradually to zero without a noticable gap. Thus A is ill-conditioned, often numerically close to singular (for $n = m$) and its numerical rank is not well defined.
- The matrix A (and also A^T) has smoothing properties, meaning that high frequency components of Av are significantly reduced compared to high frequency components of the vector $v, v \neq 0$.

Furthermore, the observation vector b is contaminated by noise. Assuming this noise is additive and the problem (1.1) for the noise-free observation is compatible, we can write

$$b = b^{\text{exact}} + \eta, \quad b^{\text{exact}} \equiv Ax^{\text{exact}}, \tag{2.1}$$

where x^{exact} denotes the unknown exact solution and η is the noise vector. The noise as well as the noise level in the data

$$\delta_{\text{noise}} \equiv \frac{\|\eta\|}{\|b^{\text{exact}}\|}, \quad \delta_{\text{noise}} \ll 1 \tag{2.2}$$

is in most applications unknown. Typically, the properties of the underlying inverse problem imply that:

- The vector b^{exact} is smooth and satisfies the discrete Picard condition (DPC), i.e. sizes of its projections to left singular subspaces of A decay (on average) faster than the corresponding singular values of A.

- The vector η does not satisfy DPC. Its properties vary depending on the particular application. For example, the most common random noise (called white) is modeled by a vector η, where each of its entries comes from the same Gaussian distribution with the zero mean.

See [18] for the detailed discussion on the DPC.

Properties above cause that data fitting approaches such as LS or TLS in their basic form can fail to approximate x^{exact}. For a closer look, consider the SVD

$$A = U\Sigma V^T, \qquad U = [u_1, \ldots, u_m],\ V = [v_1, \ldots, v_n],\ \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0, \tag{2.3}$$

where $r = rank(A)$. Then the LS solution of (1.1) with the right-hand side as in (2.1) satisfies

$$x^{\text{LS}} = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i \tag{2.4}$$

$$= \sum_{i=1}^{r} \frac{u_i^T b^{\text{exact}}}{\sigma_i} v_i + \sum_{i=1}^{r} \frac{u_i^T \eta}{\sigma_i} v_i, \tag{2.5}$$

where the left sum in (2.5) represents the unknown exact solution and the right sum is caused by the presence of noise η. Now the question arises how the size of the second sum compares to the size of the first sum. For example for white noise, sizes of projections $u_i^T \eta$ are for all i roughly comparable. Because of DPC, sizes of projections $u_i^T b^{\text{exact}}$ decay with the increasing i. Thus $u_i^T b$ is dominated by exact data for small i, but by noise for large i. Now, since σ_i decay to zero with the increasing i, the noisy components in x^{LS} are amplified destroying fully the solution x^{LS}. Such problems must be solved by regularization techniques allowing to suppress the influence of noise on the approximate solution.

2.2 Regularization via SVD filtering

Regularized least squares methods include TSVD (also called truncated least squares), TTLS (also called regularized total least squares), or Tikhonov regularization; see, e.g., [17], [21] for a summary and references. These methods are often interpreted as filtering of the unwanted components of the SVD expansion (2.4) by the so called filter factors.

Since noise is amplified in components corresponding to large i, the TSVD regularized solution is defined as

$$x_k^{\text{TSVD}} = \sum_{i=1}^{k} \frac{u_i^T b}{\sigma_i} v_i,$$

$$= \sum_{i=1}^{r} f_i \frac{u_i^T b}{\sigma_i} v_i, \qquad f_1 = \cdots = f_k = 1, \quad f_{k+1} = \cdots = f_r = 0,$$

where f_i are the filter factors and $1 \leq k \leq r$ is the regularization parameter controlling the balance between the regularization and approximation error. The vector x_k^{TSVD} is thus the LS solution of the nearby problem

$$A_k x_k \approx b, \tag{2.6}$$

where A_k is the best rank-k approximation of A in the sense of the Eckart-Young-Mirsky theorem [9]. For the basic Tikhonov regularization [46] minimizing

$$\min_{x\in\mathbb{R}^n} \Big(\|Ax-b\|^2 + \lambda \|x\|^2 \Big), \tag{2.7}$$

where $\lambda > 0$ is the regularization parameter, filter factors are given by [20, Chap. 6]

$$f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}, \qquad i = 1, \ldots, r.$$

For multiple observation problems with the right-hand side

$$B = B^{\text{exact}} + H, \quad B^{\text{exact}} \equiv AX^{\text{exact}},$$

TSVD or Tikhonov filter factors can be obtained directly from the filter factors for $d = 1$, see [C11, Sect. 2].

When we expect significant errors to be present also in A, TTLS regularization is more appropriate. We describe it for the multiple observation problem (1.2). The basic idea is to chose the regularization parameter $\epsilon > 0$ such that the influence of singular values of $[B|A]$ smaller than ϵ is suppressed similarly as in (2.6); see [11, Sect. 2]. To use the notation consistent with the paper [C11], consider the SVD (2.3) of A and denote

$$[B|A] = \widehat{U}\widehat{\Sigma}\widehat{V}^T, \quad \widehat{\sigma}_1 \geq \cdots \geq \widehat{\sigma}_{n+d} \geq 0, \tag{2.8}$$

the SVD and singular values of $[B|A]$. For $0 \leq t \leq n$ define the partitioning

$$\widehat{V} = \underbrace{\begin{bmatrix} \widehat{V}_{11} & \widehat{V}_{12} \\ \widehat{V}_{21} & \widehat{V}_{22} \end{bmatrix}}_{\begin{matrix} n-t & t+d \end{matrix}} \begin{matrix} \} d \\ \} n \end{matrix}, \quad \text{with} \quad \widehat{V}_{12} = \begin{bmatrix} \widehat{v}_{1,(n-t)+1} & \cdots & \widehat{v}_{1,n+d} \\ \vdots & \ddots & \vdots \\ \widehat{v}_{d,(n-t)+1} & \cdots & \widehat{v}_{d,n+d} \end{bmatrix}. \tag{2.9}$$

In order to ensure existence of the solution (see also Section 1.2), the regularization parameter $\epsilon > 0$ must be chosen such that:

(i) $\widehat{\sigma}_{n-t} > \epsilon > \widehat{\sigma}_{(n-t)+1}$ holds for some index t, $0 \leq t \leq n$, and

(ii) $\widehat{V}_{12}$ in the corresponding partitioning of V is of full column rank.

The TTLS solution of (1.2) is then defined as

$$X_\epsilon^{\text{TTLS}} = -\widehat{V}_{22}\widehat{V}_{12}^\dagger, \tag{2.10}$$

where $\widehat{V}_{12}^\dagger$ is the Moore-Penrose pseudoinverse. For $d = 1$, filter factors for TTLS were derived in [11] yielding

$$f_i = \sum_{\ell=(n-t)+1}^{n+1} \frac{\widehat{v}_{1,\ell}^2}{\|\widehat{V}_{12}\|_F^2} \frac{\sigma_i^2}{\sigma_i^2 - \widehat{\sigma}_\ell^2}, \qquad i = 1, \ldots, r. \tag{2.11}$$

The derivation employs relations between eigendecompositions of AA^T and $[b, A][b, A]^T$ using the fact that $[b, A][b, A]^T = AA^T + bb^T$ can be interpreted as a rank-one update of AA^T.

Since TLS and thus also TTLS can not be rewritten as a set of problems for the individual observations (as explained in Section 1.2), filter factors for problems with $d > 1$ do not follow directly from the filter factors above. Motivated by [10], [11], we studied in the paper [C11] eigenvalues and eigenvectors of rank-d updates of symmetric matrices. Based on these results, we proved that TTLS for $d > 1$ can also be expressed as a filtering method since the lth column of the matrix $X_\epsilon^{\mathrm{TTLS}}$ satisfies

$$X_\epsilon^{\mathrm{TTLS}} e_l = \sum_{i=1}^{r} \sum_{j=1}^{d} f_{i,j,l} \frac{u_i^T b_j}{\sigma_i} v_i \,, \qquad l = 1, \ldots, d,$$

where b_j is the jth column of B. Explicit formulas for the filter factors $f_{i,j,l}$ generalizing the formulas for $d = 1$ are given in [C11, Theorem 5.3]. In this way we provided a tool for future studying of regularization properties of the TTLS method.

2.3 Noise in the Golub-Kahan iterative bidiagonalization

Since direct approaches mentioned above require computation of (at least a part of) the SVD, iterative Krylov subspace regularization can be more beneficial for some problems; see [31] for an overview of Krylov subspace methods and [18] for basic principles of their application in regularization. Among the most widely used we can find LSQR [35], LSMR [12], or their hybrid variants (see, e.g., [16], [8], and many others) combing the iterative outer regularization with a direct inner regularization of the projected problem. All of these methods are based on computation of bases of underlying Krylov subspaces by the Golub-Kahan iterative bidiagonalization (1.11)-(1.12). The bidiagonalization thus represents one of the fundamental algorithms in iterative regularization. A closer look at its properties when applied to ill-posed problems was provided in our paper [C8].

Vectors $s_1, \ldots, s_k$, and $w_1, \ldots, w_k$, computed by the three-term recurrences (1.11)-(1.12) with the starting vectors $w_0 = 0$ and $s_1 = b/\beta_1, \beta_1 = \|b\|$, form orthonormal bases of the Krylov subspaces $\mathcal{K}_k(AA^T, b)$ and $\mathcal{K}_k(A^TA, A^Tb)$, respectively. Assume the process does not terminate before the step k. Denote $S_k \equiv [s_1, \ldots, s_k] \in \mathbb{R}^{m\times k}$, $W_k \equiv [w_1, \ldots, w_k] \in \mathbb{R}^{n\times k}$ and

$$L_k \equiv \begin{bmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_k & \alpha_k \end{bmatrix} \in \mathbb{R}^{k\times k}, \qquad L_{k+} \equiv \begin{bmatrix} L_k \\ e_k^T \beta_{k+1} \end{bmatrix} \in \mathbb{R}^{(k+1)\times k}. \tag{2.12}$$

Then

$$A^T S_k = W_k L_k^T, \qquad A W_k = S_{k+1} L_{k+},$$

implying

$$(AA^T) S_k = S_{k+1}(L_{k+} L_k^T), \qquad (A^TA) W_k = W_{k+1}(L_{k+1}^T L_{k+}). \tag{2.13}$$

The equations (2.13) give the well known connection between the bidiagonalization algorithm and the Lanczos tridiagonalization [28] of AA^T with the starting vector s_1 and of A^TA with the starting vector w_1, described in various ways in [1], [5], [6], [13], etc.

The bidiagonalization vectors s_{k+1} and w_{k+1} can thus be written in terms of the Lanczos polynomials as

$$s_{k+1} = \varphi_k(AA^T)b, \qquad w_{k+1} = \psi_k(A^TA)A^Tb, \qquad \varphi_k, \psi_k \in \mathcal{P}_k, \tag{2.14}$$

where $\mathcal{P}_k$ denotes the set of polynomials of degree less or equal to k. Using this representation, we analyzed in [C8] how white noise η from the measured right-hand side b propagates in the bidiagonalization vectors. In particular, while all the vectors w_k are smooth, this is not true for s_k. Since $s_1 = b/\beta_1 = (b^{\text{exact}} + \eta)/\beta_1$, then

$$s_{k+1} = \varphi_k(AA^T)b = \varphi_k(AA^T)b^{\text{exact}} + \varphi_k(AA^T)\eta.$$

Denoting

$$\varphi_k(AA^T) = \zeta_k(AA^T)^k + \cdots + \zeta_1(AA^T) + \zeta_0,$$

where $\zeta_k, \ldots, \zeta_0 \in \mathbb{R}$ are the coefficients of the polynomial φ_k, gives

$$\varphi_k(AA^T)\eta = \left[\zeta_k(AA^T)^k + \cdots + \zeta_1(AA^T)\right]\eta + \zeta_0\eta.$$

Multiplication by AA^T smooths the noise vector η, except of the last summand corresponding to the constant term $\zeta_0 \equiv \varphi_k(0)$ of the Lanczos polynomial φ_k. Consequently, we can write

$$s_{k+1} = \tilde{s}_{k+1} + \varphi_k(0)\eta,$$

where

$$\tilde{s}_{k+1} = \varphi_k(AA^T)b^{\text{exact}} + \left[\varphi_k(AA^T) - \varphi_k(0)\right]\eta$$

is for smaller k smooth and thus $\varphi_k(0)\eta$ is approximately the noisy component of s_{k+1}. Moreover, as shown in [C8, Sect. 3], the relative size of the noisy part in s_{k+1} in comparison to the smooth part $\tilde{s}_{k+1}$ increases until the so called noise revealing iteration k^{noise}, where noise fully dominates. Then, a part of noise is projected out resulting in a smoother left bidiagonalization vector. Figure 2.1 adopted from [C8, Fig. 5.1] shows individual entries of several bidiagonalization vectors s_k for the standard benchmark problem `shaw(400)` available in the Regularization Toolbox [19] with the noise level $\delta_{\text{noise}} = 10^{-4}$. The relative size of noise increases until $k^{\text{noise}} = 7$, the vector s_8 is dominated by a high-frequency part of noise η.

The noise amplification factor $\varphi_k(0)$ can be easily controlled during the iterations, since it can be computed from the bidiagonalization coefficients

$$\varphi_k(0) = (-1)^k \frac{1}{\beta_{k+1}} \prod_{j=1}^{k} \frac{\alpha_j}{\beta_j}. \tag{2.15}$$

Other way of detecting k^{noise} based on left singular vectors of L_k was also presented in [C8, Sect. 4].

This result can be used in various ways. First of all, it indicates that approximately after the iteration step k^{noise}, noise significantly propagates to the bidiagonalization

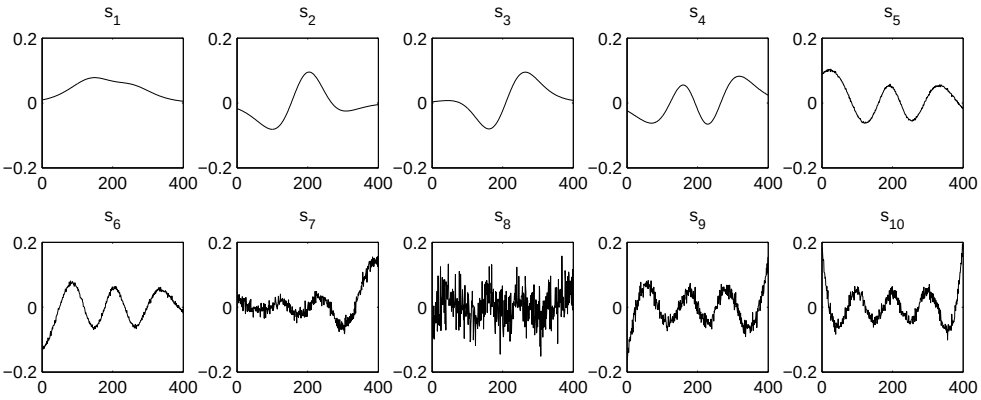


Figure 2.1: Individual entries of several left bidiagonalization vectors s_k computed by the double reorthogonalized Golub-Kahan iterative bidiagonalization for the problem `shaw(400)` with the noise level $\delta_{\text{noise}} = 10^{-4}$. The vector s_8 is dominated by a high-frequency part of noise η.

and thus we have to stop (unless inner regularization is used in a hybrid framework), see, e.g., [39] for application of this stopping criterion. Furthermore, in the iteration k^{noise},

$$\eta \approx \varphi_k(0)^{-1} s_{k+1}, \tag{2.16}$$

in the sense that the right term represents an estimate of the high-frequency part of unknown noise η. Thus an estimate of the unknown noise level is given by

$$\delta_{\text{noise}} \approx \frac{\|\eta\|}{\|b\|} \approx \frac{1}{\beta_1} \, |\varphi_k(0)^{-1}|,$$

see, e.g., our paper [22] (not included in this thesis) for an application in image deblurring on problems with various noise distributions not restricted to white noise.

2.4 Iterative regularization via projection

With understanding on how noise contaminates the bidiagonalization vectors, we can investigate methods CRAIG, LSQR and LSMR based on the Golub-Kahan bidiagonalization. Considering for simplicity the zero initial guess, they search for the approximation in the form

$$x_k = W_k y_k \in \mathcal{K}_k(A^T A, A^T b),$$

where

$$\begin{aligned} L_k y_k^{\text{CRAIG}} &= \beta_1 e_1, \\ y_k^{\text{LSQR}} &= \operatorname*{argmin}_{y \in \mathbb{R}^k} \|\beta_1 e_1 - L_{k+} y\|, \\ y_k^{\text{LSMR}} &= \operatorname*{argmin}_{y \in \mathbb{R}^k} \|\beta_1 \alpha_1 e_1 - L_{k+1}^T L_{k+} y\|. \end{aligned}$$

In words, the approximation x_k on $\mathcal{K}_k(A^TA, A^Tb)$ is selected such that CRAIG minimizes the distance of x_k from the LS solution, LSQR minimizes the norm of the residual r_k and LSMR minimizes the norm of A^Tr_k.

We explained in Section 2.1, that for ill-posed problems searching a solution minimizing the residual norm (LS solution) is not adequate. Instead, one should look for a vector x_k such that (in an appropriate sense)

$$r_k \equiv b - Ax_k \approx \eta.$$

This idea is behind various stopping criteria in iterative regularization, e.g., the Morozov's discrepancy principle [33] (comparing residual norm to an a priori known noise level estimate), the cumulative residual periodogram [41], [42] (assuming white noise). However, a priori information is rarely available in practice. In the paper [C10] (see also the PhD thesis [25] for further discussion), we considered a general noise setting and studied the match between the actual residual r_k and the unknown noise vector η for LSQR, LSMR, and CRAIG without any such a priori information. We derived explicit relations between residual vectors and the vectors s_k with the coefficients given by the constant terms of the Lanzos polynomials (2.14). In particular, it was revealed that for CRAIG,

$$r_k^{\mathrm{CRAIG}} = \varphi_k(0)^{-1}s_{k+1}, \quad k = 1, 2, \ldots$$

Consequently, x_k^{CRAIG} is the exact solution of a compatible problem

$$Ax = \tilde{b}, \quad \text{where} \quad \tilde{b} = b - \varphi_k(0)^{-1}s_{k+1},$$

i.e. currently available estimate (2.16) of the noise vector is subtracted from the original right-hand side. In LSQR and LSMR, the residual is a combination of all previously computed vectors s_k, where the coefficients reflect the amount of propagated noise in each of s_k vectors. For LSQR,

$$r_k^{\mathrm{LSQR}} = \frac{1}{\sum_{l=0}^{k}\varphi_l(0)^2}\sum_{l=0}^{k}\varphi_l(0)s_{l+1},$$

for LSMR see [C10, Sect. 3.3]. In this way LSQR and LSMR provide generally solutions with lower true error (i.e. the norm of the difference between the approximation x_k and the exact solution x^{exact}) than CRAIG. (Note that because of their mathematical equivalence to selected methods above, the results apply also to CGNE and CGME [43].)

The Golub-Kahan bidiagonalization can be used also to solve iteratively the Tikhonov problem (2.7), see, e.g., [17] and the references therein, [24], [8], etc. Using the well known equivalence between regularize-then-project and project-then-regularize approaches described in [18], one of the possibilities is to employ hybrid LSQR, where the outer Golub-Kahan iterations for the data A and b are combined with the inner Tikhonov regularization of the projected problem

$$L_{k+}y_k \approx \beta_1 e_1.$$

Here in fact two regularization parameters are tuned - the number of bidiagonalization iterations k and the Tikhonov parameter λ. Some ill-posed problems must be

approached by a more general form of the Tikhonov regularization (also known as weighted regularized LS)

$$\min_{x\in\mathbb{R}^n} \Big(\|Ax-b\|_W^2 + \lambda \|D(x-x_0)\|^2 \Big), \tag{2.17}$$

incorporating general weighted norm defined by a weighting matrix $W \in \mathbb{R}^{m\times m}$ and the initial approximation $x_0 \in \mathbb{R}^n$. The matrix $D \in \mathbb{R}^{p\times n},\ p \le n$, comes often from a discretization of the first or second order derivative operator, see [18, Chap. 8]. In the paper [C9] we derived a hybrid approach based on the Golub-Kahan bidiagonalization for the solution of (2.17) particularly effective for large scale problems. The developments were motivated by the algorithm presented in [32] that required costly evaluation of the generalized SVD [55]. In [C9], we assume that an estimate on the covariance structure of noise in the measured data b is given. The two regularization parameters in the hybrid method are then selected automatically based on the χ^2 principle.

Bibliography

[1] A. Björck, *A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations*, BIT 28 (1988), pp. 659–670.

[2] B. W. Bader and T. G. Kolda, *Algorithm 862: MATLAB tensor classes for fast algorithm prototyping*, ACM Transactions on Mathematical Software 32(4) (2006), pp. 635–653.

[3] A. Björck, *Numerical Methods for Least Squares problems*, SIAM, Philadelphia, PA, 1996.

[4] A. Björck, *Block bidiagonal decomposition and least squares problems with multiple right-hand sides*, unpublished manuscript (2008).

[5] A. Björck, E. Grimme, and P. Van Dooren, *An implicit shift bidiagonalization algorithm for ill-posed systems*, BIT 34 (1994), pp. 510–534.

[6] D. Calvetti, G. H. Golub, and L. Reichel, *Estimation of the L-curve via Lanczos bidiagonalization*, BIT 39 (1999), pp. 603–619.

[7] E. J. Craig, *The N-step iteration procedures*, J. Math. Phys. 34 (1955), pp. 64–73.

[8] J. Chung, and K. Palmer, *A hybrid LSMR algorithm for large-scale Tikhonov regularization*, SIAM J. Sci. Comput. 37 (5) (2015), pp. 562–580.

[9] G. Eckart, and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrica 1 (1936), pp. 211–218.

[10] R. D. Fierro, and J. R. Bunch, *Collinearity and total least squares*, SIAM J. Matrix Anal. Appl. 15 (1994), pp. 1167–1181.

[11] R. D. Fierro, G. H. Golub, P. C. Hansen, and D. P. O'Leary, *Regularization by truncated total least squares*, SIAM J. Sci. Comp. 18(4) (1997), pp. 1223–1241.

[12] D. C.-L. Fong, and M. Saunders, *LSMR: an iterative algorithm for sparse least-squares problems*, SIAM J. Sci. Comput. 33 (5) (2011), pp. 2950–2971.

[13] G. Golub, and W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, SIAM: Series B, Numerical Analysis 2 (1965), pp. 205–224.

[14] G. H. Golub, and C. F. Van Loan, *An analysis of the total least squares problem*, Numer. Anal. 17 (1980), pp. 883–893.

[15] M. Hanke, *A Taste of Inverse Problems*, SIAM Publications, Philadelphia, PA, 2017.

[16] M. Hanke, *On Lanczos based methods for the regularization of discrete ill-posed problems*, BIT 41 (5, suppl.) (2001), pp. 1008–1018.

[17] P. C. Hansen, *Rank Deficient and Discrete Ill-posed Problems: Numerical Aspects of Linear Inversion*, SIAM Publications, Philadelphia, PA, 1996.

[18] P. C. Hansen, *Discrete Inverse Problems: Insight and Algorithms*, SIAM Fundamentals of Algorithms, Philadelphia, PA, 2010.

[19] P. C. Hansen, *Regularization Tools – version 3.2 for MATLAB 6.0, a package for analysis and solution of discrete ill-posed problems*, (`http://www2.imm.dtu.dk/~pch/Regutools/index.html`).

[20] P. C. Hansen, J. Nagy, and D. P. O'Leary: *Deblurring images: Matrices, spectra, and filtering*, SIAM Publications, Philadelphia, PA, 2006.

[21] P. C. Hansen, V. Pereyra, and G. Scherer, *Least Squares Data Fitting with Applications*, SIAM Publications, Philadelphia, PA, 2013.

[22] I. Hnětynková, M. Kubínová, M. Plešinger, *Notes on performance of bidiagonalization-based estimator in image deblurring*, Proceedings of Algoritmy 2016 - 20th Conference on Scientific Computing, Publishing House of Slovak University of Technology (2016), pp. 333–342.

[23] I. Hnětynková, and Z. Strakoš, *Lanczos tridiagonalization and core problems*, Linear Algebra Appl. 421 (2007), pp. 243–251.

[24] M. Kilmer, P. C. Hansen, and M. Español: *A projection-based approach to general-form Tikhonov regularization*, SIAM J. Sci. Comp. 29(1) (2007), pp. 315–330.

[25] M. Kubínová, *Numerical Methods in Discrete Inverse Problems*, PhD Thesis, Charles University, Prague, Czech Republic, 2018.

[26] A. Kukush, I. Markovsky, and S. Van Huffel, *On consistent estimators in linear and bilinear multivariate errors-in-variables models*, Total Least Squares and Errors-in-Variables Modeling (S. Van Huffel and P. Lemmerling, eds.), Kluwer Academic Publishers, Dordrecht, 2002, pp. 155–164.

[27] A. Kukush, I. Markovsky, and S. Van Huffel, *Consistent estimation in the bilinear multivariate errors-in-variables model*, Metrika 57(3) (2003), pp. 254–285.

[28] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bur. Standards 45 (1950), pp. 255–282.

[29] C. L. Lawson, and R. J. Hanson, *Solving Least Squares Problem*, SIAM Classics in Applied Mathematics, Philadelphia, PA, 1995.

[30] C.-K. Li, X.-G. Liu, and X.-F. Wang, *Extension of the total least square problem using general unitarily invariant norms*, Linear and Multilinear Algebra 55(1) (2007), pp. 71–79.

[31] J. Liesen, and Z. Strakoš, *Krylov subspace methods: Principles and Analysis*, Oxford University Press, Oxford, 2013.

[32] J. Mead, and R. Rosemary, *A Newton root-finding algorithm for estimating the regularization parameter for solving ill-conditioned least squares problems*, Inverse Problems 25 (2009), art. no. 025002.

[33] V. A. Morozov, *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl. 7 (1966), pp. 414–417.

[34] B. Ottersten, D. Asztély, M. Kristensson, and S. Parkvall, *A statistical approach to subspace based estimation with applications in telecommunications*, Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling (S. Van Huffel, ed.), SIAM Publications, Philadelphia, 1997, pp. 285–294.

[35] C. C. Paige, and M. A. Saunders, *LSQR: an algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software 8 (1) (1982), pp. 43–71.

[36] C. C. Paige, and Z. Strakoš, *Scaled total least squares fundamentals*, Numerische Mathematic 91 (2006), pp. 117–146.

[37] C. C. Paige, and Z. Strakoš, *Core problems in linear algebraic systems*, SIAM J. Matrix Anal. Appl. 27 (2006), pp. 861–875.

[38] M. Plešinger, *The Total Least Squares Problem and Reduction of Data in $AX \approx B$*, PhD Thesis, Technical University of Liberec, Liberec, Czech Republic, 2008.

[39] R. Renaut, S. Vatankhah, and V. Ardestani, *Hybrid and iteratively reweighted regularization by unbiased predictive risk and weighted GCV for projected systems*, SIAM J. Sci. Comp. 39 (2017), pp. B221–B243.

[40] J. B. Rosen, H. Park, and J. Glick, *Total least norm for linear and nonlinear structured problems*, Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling (S. Van Huffel, ed.), SIAM Publications, Philadelphia, 1997, pp. 203–214.

[41] B. W. Rust, *Parameter selection for constrained solutions to ill-posed problems*, Computing Science and Statistics 32 (2000), pp. 333–347.

[42] B. W. Rust, and D. P. O'Leary, *Residual periodograms for choosing regularization parameters for ill-posed problems*, Inverse Problems 24 (2008), art. no. 034005.

[43] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd Edition, SIAM, Philadelphia, PA, 2003.

[44] D. M. Sima, *Regularization Techniques in Model Fitting and Parameter Estimation*, PhD Thesis, Katholieke Universiteit Leuven, Leuven, Belgium, 2006.

[45] T. Söderström, U. Soverini, and K. Mahata, *Perspectives on errors-in-variables estimation for dynamic systems*, Total Least Squares and Errors-in-Variables Modeling (S. Van Huffel and P. Lemmerling, eds.), Kluwer Academic Publishers, Dordrecht, 2002, pp. 271–280.

[46] A. Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, Soviet Math. Dokl. 4 (1963), pp. 1035–1038.

[47] A. Tikhonov, and V. Y. Arsenin, *Solution of Ill-posed Problems*, Winston and Sons, Washington, 1977.

[48] L. R. Tucker, *Implications of factor analysis of three-way matrices for measurement of change*, Problems in Measuring Change (C. W. Harris ed.), University of Wisconsin Press, 1963, pp. 122–137.

[49] L. R. Tucker, *The extension of factor analysis to three-dimensional matrices*, Contributions to Mathematical Psychology (H. Gulliksen, N. Frederiksen eds), Holt, Rinehardt & Winston, New York, 1964, pp. 110–127.

[50] L. R. Tucker, *Some mathematical notes on three-mode factor analysis*, Psychometrika 31(3) (1966), pp. 279–311.

[51] S. Van Huffel, *Documented Fortran 77 programs of the extended classical total least squares algorithm, the partial singular value decomposition algorithm and the partial total least squares algorithm*, Internal. Report ESAT-KUL 88/1, ESAT Lab., Dept. of Electrical Engrg., Katholieke Universiteit Leuven (1988).

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

[53] 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, SIAM Publications, Philadelphia, 1997.

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

[55] C. F. Van Loan, *Generalizing the singular value decomposition*, SIAM J. Numer. Anal. 13 (1976), pp. 76–83.

[56] X.-F. Wang, *Total least squares problem with the arbitrary unitarily invariant norms*, Linear and Multilinear Algebra 65(3) (2017), pp. 438–456.

[57] 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.

[58] M. Wei, *Algebraic relations between the total least squares and least squares problems with more than one solution*, Numerische Mathematik 62 (1992), pp. 123–148.

Reprints of papers

Here we include 11 papers in the following order (the ordering is the same as in the preface of this thesis):

[C1] 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.
DOI: 10.1137/100813348

[C2] I. Hnětynková, M. Plešinger, and Z. Strakoš, *The core problem within a linear approximation problem $AX \approx B$ with multiple right-hand sides*, SIAM J. Matrix Anal. Appl. 34 (2013), pp. 917–931.
DOI: 10.1137/120884237

[C3] I. Hnětynková, M. Plešinger, and Z. Strakoš, *Band generalization of the Golub–Kahan bidiagonalization, generalized Jacobi matrices, and the core problem*, SIAM J. Matrix Anal. Appl. 36 (2015), pp. 417–434.
DOI: 10.1137/140968914

[C4] I. Hnětynková, and M. Plešinger, *Complex wedge-shaped matrices: A generalization of Jacobi matrices*, Linear Algebra Appl. 487 (2015), pp. 203–219.
DOI: 10.1016/j.laa.2015.09.017

[C5] I. Hnětynková, M. Plešinger, and D. M. Sima, *Solvability of the core problem with multiple right-hand sides in the TLS sense*, SIAM J. Matrix Anal. Appl. 37(3) (2016), pp. 861–876.
DOI: 10.1137/15M1028339

[C6] I. Hnětynková, M. Plešinger, and J. Žáková, *TLS formulation and core reduction for problems with structured right-hand sides*. Linear Algebra Appl. 555 (2018), pp. 241–265.
DOI: 10.1016/j.laa.2018.06.016

[C7] I. Hnětynková, M. Plešinger, and J. Žáková, *On TLS formulation and core reduction for data fitting with generalized models*. Linear Algebra Appl. 577 (2019), pp. 1–20.
DOI: 10.1016/j.laa.2019.04.018

[C8] I. Hnětynková, M. Plešinger, and Z. Strakoš, *The regularizing effect of the Golub-Kahan iterative bidiagonalization and revealing the noise level in the data*, BIT 49 (4) (2009), pp. 669–696.
DOI: 10.1007/s10543-009-0239-7

[C9] R. Renaut, I. Hnětynková, and J. Mead, *Regularization parameter estimation for large-scale Tikhonov regularization using a priori information*, Computational Statistics and Data Analysis 54 (2010), pp. 3430–3445.
DOI: 10.1016/j.csda.2009.05.026

[C10] I. Hnětynková, M. Kubínová, M. Plešinger, *Noise representation in residuals of CRAIG, LSQR and LSMR regularization*, Linear Algebra Appl. 533 (2017), pp. 357–379.
DOI: 10.1016/j.laa.2017.07.031

[C11] I. Hnětynková, M. Plešinger, and J. Žáková, *Filter factors of truncated TLS regularization with multiple observations*. Applications of Mathematics 62 (2017), pp. 105–120.
DOI: 10.21136/AM.2017.0228-16