Regularized total least squares based on quadratic eigenvalue problem solvers [1]

Diana M. Sima[2] Sabine Van Huffel[2] Gene H. Golub[3]

January 2004
Accepted for publication in BIT Numerical Mathematics.

December 2004
Published in BIT Numerical Mathematics, **44**(4), p. 793 – 812.

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/dsima/reports/SISTA3-07.ps.gz*

[2]ESAT-SISTA, K.U. Leuven, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium, (diana.sima,sabine.vanhuffel@esat.kuleuven.ac.be). This work was supported by **Research Council KUL:** GOA-Mefisto 666, IDO/99/003 (Predictive computer models for medical classification problems using patient data and expert knowledge), several PhD/postdoc & fellow grants; **Flemish Government: FWO:** PhD/postdoc grants, projects, G.0200.00 (damage detection in composites by optical fibers), G.0078.01 (structured matrices), G.0407.02 (support vector machines), G.0269.02 (magnetic resonance spectroscopic imaging), G.0270.02 (nonlinear Lp approximation), research communities (ICCoS, ANMMM); **AWI:** Bil. Int. Collaboration Hungary/Poland; **IWT:** PhD Grants, **Belgian Federal Government:** DWTC (IUAP IV-02 (1996-2001) and IUAP V-22 (2002-2006): Dynamical Systems and Control: Computation, Identification & Modelling); **EU:** NICONET, INTERPRET, PDT-COIL, MRS/MRI signal processing (TMR); **Contract Research/agreements:** Data4s, IPCOS.

[3]Department of Computer Science, Stanford University, Stanford, California 94305-9025 (golub@sccm.stanford.edu) This work was supported in part by National Science Foundation under Grant No. CCR-9971010.

Abstract

This paper presents a new computational approach for solving the Regularized Total Least Squares problem. The problem is formulated by adding a quadratic constraint to the Total Least Square minimization problem. Starting from the fact that a quadratically constrained Least Squares problem can be solved via a quadratic eigenvalue problem, an iterative procedure for solving the regularized Total Least Squares problem based on quadratic eigenvalue problems is presented. Discrete ill-posed problems are used as simulation examples in order to numerically validate the method.

REGULARIZED TOTAL LEAST SQUARES BASED ON QUADRATIC EIGENVALUE PROBLEM SOLVERS

DIANA M. SIMA[1], SABINE VAN HUFFEL[1] and GENE H. GOLUB[2]

[1] *ESAT-SISTA, K.U. Leuven, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium.*
email: {diana.sima,sabine.vanhuffel}@esat.kuleuven.ac.be

[2] *Department of Computer Science, Stanford University, Stanford, CA 94305-9025, USA.*
email: golub@sccm.stanford.edu

Abstract.

This paper presents a new computational approach for solving the Regularized Total Least Squares problem. The problem is formulated by adding a quadratic constraint to the Total Least Square minimization problem. Starting from the fact that a quadratically constrained Least Squares problem can be solved via a quadratic eigenvalue problem, an iterative procedure for solving the regularized Total Least Squares problem based on quadratic eigenvalue problems is presented. Discrete ill-posed problems are used as simulation examples in order to numerically validate the method.

AMS subject classification: 65F20, 65F30.

Key words: Quadratic eigenvalue problem, regularization, Total Least Squares.

1 Introduction

Total Least Squares (TLS) is a technique for solving overdetermined linear systems of equations

$$Ax \approx b, \qquad A \in \Re^{m\times n}, \quad b \in \Re^m, \quad x \in \Re^n, \quad (m > n), \tag{1.1}$$

in which both the coefficient matrix A and the right-hand side b are subject to errors. Mathematically, it solves the optimization problem:

$$\min_{x,\overline{A},\overline{b}} \|(A\ b) - (\overline{A}\ \overline{b})\|_F^2 \text{ subject to } \overline{A}x = \overline{b}, \tag{1.2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix.

The ordinary Least Squares (LS) method assumes the coefficient matrix A to be error-free and all errors are confined to the right-hand side b. In many practical applications, all data are contaminated by noise, which motivates the use of TLS. Efficient and reliable numerical methods to compute the TLS solution were developed in the literature ([10], [35]), and they are based on the singular value decomposition (SVD). Methods are developed for the problem when only some of the columns of the matrix A are contaminated by noise and the rest are noise-free ([9], [35]).

An important special class of problems that appear in practice involves *structured* data matrices (Toeplitz, Hankel, etc). The Structured Total Least Squares problem has the same formulation as the TLS problem (1.2), with the additional constraint that $(\overline{A}\ \overline{b})$ has the same structure as $(A\ b)$ ([23], [24]).

Least squares, total least squares or other classical methods for solving an (overdetermined) linear system (1.1), when the coefficient matrix A is ill-conditioned, might provide a solution that is physically meaningless for the given problem. This happens when A is (nearly) rank-deficient with no significant gap in the singular values. Typical examples are encountered when system (1.1) is the discretization of a continuous ill-posed problem [17].

Regularization is needed in order to decrease the effect due to the intrinsic ill-conditioning of the problem and due to the noise in the data and stabilize the solution. Various formulations have been used in the literature for the purpose of introducing regularization in ill-posed problems ([32], [17], [7], [4], [2]). In Section 2, some of the most important methods are shortly reviewed.

The Regularized Total Least Squares (RTLS) problem formulation considered in this paper consists in imposing a quadratic constraint on the solution vector x in the TLS problem (1.2) ([7], [14]). In this manner, the values that the components of the solution vector may take are bounded and a certain degree of smoothness can be imposed on the solution. The new constrained problem cannot be solved with simple and elegant SVD-based methods, as the classical TLS problem. Therefore, a different computational approach based on iteratively solving quadratic eigenvalue problems (QEP) is proposed. The development of this new algorithm was inspired by the fact that quadratically constrained least squares (also named Regularized Least Squares (RLS)) can be solved by a quadratic eigenvalue problem [6]. Due to the more complicated nature of the RTLS formulation, one QEP cannot solve the problem, but it is shown in this paper that the solution can be approximated in a few iterations, each consisting in solving a QEP.

Two of the main advantages exhibited by this new computational approach are its *robustness* with respect to the initialization of the iterative procedure and its *efficiency* in solving large problems. Experiments showed that the global minimum of the RTLS nonconvex optimization problem can be attained even when using random starting vectors. Quadratic eigenvalue problems equivalent to the RLS formulation [6] can be solved efficiently even for large problem sizes, as it was recently shown in [25]. In the RTLS algorithm presented in this paper, the same kind of QEPs appears at every iteration; therefore, the efficient solver described in [25] can also be used for RTLS. Moreover, it is not necessary to compute the whole spectrum, since only one eigenpair is needed.

Many application areas benefit from the use of RTLS methods. Unstructured ill-posed problems suited for applying RTLS techniques can arise from discretizations of integral or differential operators, in applications from medical imaging [1] (electrical impedance tomography, X-ray tomography, optical tomography), geophysical applications [3] (seismology, radar or sonar imaging), etc. Regularization can be required for the structured TLS problem. Structured RTLS

is useful for solving deconvolution problems in image deblurring [30], in medical applications (renography) [26] and in signal restoration [36]. A fast reliable method for regularized structured TLS is presented in [26], while structure and other norms are also taken into account in a regularized total least norm formulation in [30].

The remainder of the paper is organized as follows. Section 2 deals with mathematical formulations of regularization methods. In Section 3, the proposed algorithm for RTLS is presented and the mathematical transformation from the original problem into a quadratic eigenvalue problem is also detailed. A theorem that gives insight into the convergence of the iterative procedure is also proved at this point. Numerical results using examples built with the "Regularization Tools" [16] are presented in Section 4.

2 Problem formulations

Throughout this section, the estimation of a solution x to problem (1.1) is considered, assuming that A and $(A\ b)$ are ill-conditioned or even rank-deficient with the singular values decreasing to zero without significant gaps.

2.1 Tikhonov regularization

The most commonly used regularization method is due to Tikhonov [31], [32]. It amounts to solving the problem:

$$\min_x \|Ax - b\|_2^2 + \lambda \|Lx\|_2^2, \tag{2.1}$$

where λ is a fixed, properly chosen regularization parameter that controls the "size" of the solution vector x, and L is a matrix that defines a (semi)norm on the solution through which the "size" is measured.

Consider the singular value decomposition (SVD) of the coefficient matrix A: $A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$ $(r = \text{rank}(A))$. Then, the (unstable) least squares solution is given by

$$x_{\text{LS}} = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i. \tag{2.2}$$

Tikhonov regularization in the form (2.1) with $L = I_n$ provides a solution

$$x_T = \sum_{i=1}^{r} \frac{\sigma_i^2}{\sigma_i^2 + \lambda} \frac{u_i^T b}{\sigma_i} v_i. \tag{2.3}$$

Several methods for choosing λ are popular in the literature: the discrepancy principle [27], generalized cross-validation [8], [12], the L-curve [15], [19], [20]. Other methods are described and analyzed, for instance, in [21], [28], [34]. For large scale problems, efficient computation of the regularization parameter λ for several different methods is discussed in [13].

REMARK 2.1. *Formula (2.3) illustrates a more general characteristic of regularized solutions; many regularization methods compute solutions of the form:*

$$x_R = \sum_{i=1}^{r} f_i \frac{u_i^T b}{\sigma_i} v_i,$$

where the filter factors f_i are meant to "filter out" the contribution of the noise. For Tikhonov regularization, these factors are computed as $f_i(\lambda) = \frac{\sigma_i^2}{\sigma_i^2+\lambda}$. See [2] for a regularization method based on a filter function of Gaussian type.

2.2 Truncation methods

Straightforward truncation methods (truncated SVD, truncated generalized SVD, truncated TLS, [4], [16]) involve computing the SVD of a certain matrix and using only the information corresponding to several of the largest singular values. For instance, the truncated SVD solution is obtained taking the sum of the first $k < r$ terms in (2.2). Note that the truncation level k is not obvious when dealing with a discrete ill-posed problem, because the singular values decay smoothly towards zero; k can be considered as the unknown regularization parameter. Similarly, the Truncated Total Least Squares (TTLS) solution is computed as the ordinary TLS solution [35], but a lower value k is used instead of the numerical rank r of $(A\ b)$, i.e., $x_{TTLS} = \sum_{i=1}^{k} \frac{\sigma_i(u_i^T b)}{\sigma_i^2 - \sigma_{min}^2(A\ b)} v_i$, $k < r$, where $\sigma_{min}(M)$ denotes the minimal singular value of a matrix M. Experiments in [4] show that in some cases TTLS outperforms other regularization methods such as Tikhonov regularization, truncated SVD or the LSQR method [29].

2.3 Quadratically constrained formulations

Regularization is often introduced by adding a quadratic constraint to the LS or TLS optimization problems ([11, §12.1], [7]). In this subsection, the mathematical form of the problem under study is presented. As in [7], it is important to first stress the differences and the connections between the Regularized Least Squares (RLS) and the Regularized Total Least Squares (RTLS) problems.

The RLS problem is defined as follows:

$$\min_x \|Ax - b\|_2^2 \text{ subject to } \|Lx\|_2^2 \leq \delta^2, \tag{2.4}$$

where $L \in \Re^{p \times n}$, $p \leq n$, and $\delta > 0$. For $\delta > 0$ small enough (i.e., $\delta < \|Lx_{\mathrm{LS}}\|_2$), an appropriate value of parameter λ can be fixed (depending in a nonlinear way of δ), such that the solution of (2.4) coincides with the one of the Tikhonov regularization problem (2.1) and of the normal system of equations

$$(A^T A + \lambda L^T L)x = A^T b. \tag{2.5}$$

The RTLS problem statement is

$$\min_{x,\overline{A},\overline{b}} \|(A\ b) - (\overline{A}\ \overline{b})\|_F^2 \text{ subject to } \overline{A}x = \overline{b},\ \|Lx\|_2^2 \leq \delta^2. \tag{2.6}$$

It is known that the TLS objective function (i.e., the Frobenius norm of the extended correction matrix) can be replaced by the orthogonal distance $\frac{\|Ax-b\|_2^2}{1+\|x\|_2^2}$ [35, §2.4.2]. Then the previous formulation can be rewritten as

$$\min_x \frac{\|Ax-b\|_2^2}{1+\|x\|_2^2} \text{ subject to } \|Lx\|_2^2 \leq \delta^2. \tag{2.7}$$

For δ small enough (i.e., $\delta < \|Lx_{\mathrm{TLS}}\|_2$), there exists a value of parameter λ such that the solution of (2.7) coincides with the solution of:

$$\min_x \frac{\|Ax-b\|_2^2}{1+\|x\|_2^2} + \lambda\|Lx\|_2^2. \tag{2.8}$$

The paper [7] is one of the first references where Tikhonov regularization has been recast for the TLS framework. In [7], the RTLS problem is also described in the following form:

$$(A^TA + \lambda_I I_n + \lambda_L L^TL)x = A^Tb, \tag{2.9}$$

where

$$\lambda_I = -\frac{\|Ax-b\|_2^2}{1+\|x\|_2^2} \text{ and } \lambda_L = -\frac{1}{\delta^2}\left(b^T(Ax-b) - \lambda_I\right). \tag{2.10}$$

This formulation is considered in two recent methods in the literature. In [7], λ_L is considered as free parameter; a corresponding value is computed for λ_I and system (2.9) is solved in an efficient way. An 'optimal' regularized solution is sought by varying λ_L. In [14], a shifted inverse power method is used to obtain an eigenpair $\left(-\lambda_I, \begin{pmatrix} x^T & -1 \end{pmatrix}^T\right)$ for the problem

$$D(x)\begin{pmatrix} x \\ -1 \end{pmatrix} = -\lambda_I \begin{pmatrix} x \\ -1 \end{pmatrix},$$

with

$$D(x) = \begin{pmatrix} A^TA + \lambda_L L^TL & A^Tb \\ b^TA & -\lambda_L\delta^2 + b^Tb \end{pmatrix},$$

where λ_I and λ_L are also given by (2.10).

REMARK 2.2. *As also noted in [7] and [14], choosing the parameter δ smaller than the value $\|Lx_{\mathrm{TLS}}\|_2$ implies that the quadratic constraint $\|Lx\|_2 \leq \delta$ is* active *at the solution for optimization problems (2.6) or (2.7). For an ill-conditioned problem (1.1), the norms $\|Lx_{\mathrm{LS}}\|_2$ and $\|Lx_{\mathrm{TLS}}\|_2$ are very big (therefore, the need for regularization); the assumption that δ is small enough can be considered as guaranteed in practice.*

In view of Remark 2.2, the regularization problems considered subsequently will be *equality constrained* problems. For further reference, their formulations are stated below:

$$\text{RLS problem:} \qquad \min_x \|Ax-b\|_2^2 \text{ subject to } \|Lx\|_2^2 = \delta^2, \tag{2.11}$$

$$\text{RTLS problem:} \qquad \min_x \frac{\|Ax-b\|_2^2}{1+\|x\|_2^2} \text{ subject to } \|Lx\|_2^2 = \delta^2. \tag{2.12}$$

The case when matrix L is the identity matrix I_n is called the *standard case* and both RLS and RTLS yield the same solution if δ is small enough ($\delta < \|x_{\mathrm{LS}}\|_2$ and $\delta < \|x_{\mathrm{TLS}}\|_2$), and the inequality constraints are replaced by equalities, as in (2.11) and (2.12). Indeed, in formulation (2.12) with $L = I_n$, the denominator $1 + \|x\|_2^2$ may be replaced by the constant $1 + \delta^2$ and a problem equivalent to RLS formulation (2.11) is obtained.

In the *general case*, matrix L may be rectangular and in practice it is usually chosen as an approximation of the first or second order derivative operators in order to impose a certain degree of smoothness on the solution. In this case, (2.11) and (2.12) are distinct problems.

3 Quadratic eigenvalue problems for RTLS

In this section it is shown how the RTLS problem can be solved numerically with an iterative method that requires at each iteration the solution of a quadratic eigenvalue problem (QEP).

It is known that the quadratically constrained least squares problem

$$\min_x \|Ax-b\|_2^2 \text{ subject to } \|x\|_2^2 = \delta^2$$

can be solved via one QEP [6]:

$$(\lambda^2 I + 2\lambda H + H^2 - \delta^{-2} g g^T) y = 0, \tag{3.1}$$

where $H = A^T A$, and $g = A^T b$.

For RTLS, the difficulty is that a single QEP cannot solve the problem. Therefore, an iterative procedure to approximate the solution by solving at each step a QEP is proposed here. Numerically, it was observed that very few iterations are needed in order to achieve a desired accuracy of the solution.

In the newly proposed algorithm, solving a QEP of the form (3.1) will be the most important computation at each iteration. A thorough survey on quadratic eigenvalue problems, their properties and solvers can be found in [33]. Recently, a fast method for solving QEPs was proposed in [25]. It can be successfully applied to monic QEPs of the form $(\lambda^2 I + \lambda A + B)y = 0$, where matrices A and B satisfy the condition that some linear combination $\zeta A + \xi B$ is of low rank. The special form of the QEP (3.1) allows the application of a Lanczos process that simultaneously projects the matrices H and gg^T into a common subspace. After k steps of the Lanczos process, the projections of H and gg^T can be approximated by two $k \times k$ symmetric banded matrices with bandwidth 2. Using the reduced matrices, a lower dimensional QEP that approximates the original one can be solved by standard methods (i.e., linearization and eigenvalue decomposition). Further details are given in [25].

3.1 RTLS algorithm

Consider the RTLS problem (2.12) and write the Lagrangean

$$\mathcal{L}(x,\lambda) = \frac{\|Ax-b\|_2^2}{1+\|x\|_2^2} + \lambda(\|Lx\|_2^2 - \delta^2).$$

The first order optimality conditions are:

$$B(x)x + \lambda L^T L x = d(x), \qquad \|Lx\|_2^2 = \delta^2, \tag{3.2}$$

where

$$B(x) = \frac{A^T A}{1+\|x\|_2^2} - \frac{\|Ax-b\|_2^2}{(1+\|x\|_2^2)^2} I_n, \qquad d(x) = \frac{A^T b}{1+\|x\|_2^2}. \tag{3.3}$$

System (3.2) will be solved iteratively by applying the following method:
ALGORITHM 3.1 (RTLSQEP).

initialization *Let x^0 be a starting vector. Compute $B_0 := B(x^0)$ and $d_0 := d(x^0)$ from (3.3). Set $k = 0$.*

step k *Find x^{k+1} and λ_{k+1}, which solve the system in x and λ:*

$$B_k x + \lambda L^T L x = d_k, \qquad \|Lx\|_2^2 = \delta^2, \tag{3.4}$$

corresponding to the largest λ (using an equivalent quadratic eigenvalue problem).
Compute $B_{k+1} := B(x^{k+1})$ and $d_{k+1} := d(x^{k+1})$ from (3.3).

stopping criterion *If $\|B_{k+1}x^{k+1} + \lambda_{k+1}L^T L x^{k+1} - d_{k+1}\|_2 < \epsilon$, where ϵ is a specified tolerance, then STOP; else $k \leftarrow k+1$ and go to step k.*

The focus will be on solving system (3.4). In subsection 3.2 it is shown how this can be done using a monic QEP.

REMARK 3.1. *Notice that at* **step** *k, the solution with largest λ is selected from the set of solutions of system (3.4). This choice is needed for the convergence of the algorithm (see Lemma 3.1 and Theorem 3.3).*

3.2 Quadratic eigenvalue problem derivation

Consider the system

$$Bx + \lambda L^T L x = d, \qquad \|Lx\|_2^2 = \delta^2, \tag{3.5}$$

with B a symmetric matrix.

Case 1: L square and invertible

If L is invertible, a change of variable, $z = Lx$, gives

$$L^{-T}BL^{-1}z + \lambda z = L^{-T}d, \qquad z^T z = \delta^2. \tag{3.6}$$

Therefore, one is led to a system of the form:

$$Wz + \lambda z = h, \qquad z^T z = \delta^2, \tag{3.7}$$

with symmetric matrix $W = L^{-T}BL^{-1}$, which can be solved using a QEP [6]. Indeed, assuming $\lambda > 0$ large enough (such that $W + \lambda I$ is positive definite) and denoting by $u = (W + \lambda I)^{-2}h$, one has $h^T u = z^T z = \delta^2$; noticing that $h = \delta^{-2}hh^T u$, the condition

$$(W + \lambda I)^2 u = h$$

can be equivalently written as the QEP

$$(\lambda^2 I + 2\lambda W + W^2 - \delta^{-2}hh^T)u = 0. \tag{3.8}$$

This QEP is solved in order to find the largest (right-most) eigenvalue λ and a corresponding eigenvector u (scaled such that $h^T u = \delta^2$).

REMARK 3.2. *It is known [33] that a QEP having a nonsingular coefficient matrix for the second order term λ^2 (in particular, monic QEP) has a full set of* finite *eigenvalues. When all coefficient matrices are real and symmetric, the quadratic eigenvalues are real or come in complex conjugate pairs; moreover, the special structure of the QEP (3.8) will enforce the right-most eigenvalue to be* real *and* positive. *Therefore, expressing system (3.4) as a monic QEP is also a guarantee that a solution corresponding to the largest $\lambda > 0$ can be found.*

The solution of the original problem is recovered by setting first $z = (W+\lambda I)u$, then $x = L^{-1}z$.

Case 2: Generalization for nonsquare L

If L is not square (for example, when L is an approximation matrix of the first or second order derivative operator), then $L^T L$ is singular. Let $L^T L = USU^T$ be the eigenvalue decomposition of $L^T L$. An equivalent form for (3.5) is

$$U^T BUy + \lambda Sy = U^T d, \qquad y^T Sy = \delta^2, \tag{3.9}$$

where $y = U^T x$. Let $r = \text{rank}(S)$ and $S_1 = S_{1:r,1:r}$. Partitioning the elements of system (3.9) according to the rank r:

$$U^T BU = \begin{pmatrix} T_1 & T_2 \\ T_2^T & T_4 \end{pmatrix},\ S = \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix},\ U^T d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix},\ y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

the system to be solved becomes:

$$\begin{cases} T_1 y_1 + T_2 y_2 + \lambda S_1 y_1 &= d_1, \\ T_2^T y_1 + T_4 y_2 &= d_2, \end{cases} \qquad y_1^T S_1 y_1 = \delta^2. \tag{3.10}$$

Under the assumption that T_4 is invertible (otherwise its pseudoinverse may still be used instead of its inverse),

$$y_2 = T_4^{-1}(d_2 - T_2^T y_1) \tag{3.11}$$

is substituted into the first equation of (3.10):

$$(T_1 - T_2 T_4^{-1} T_2^T + \lambda S_1) y_1 = (d_1 - T_2 T_4^{-1} d_2).$$

For $W = S_1^{-\frac{1}{2}}(T_1 - T_2 T_4^{-1} T_2^T) S_1^{-\frac{1}{2}}$ and $h = S_1^{-\frac{1}{2}}(d_1 - T_2 T_4^{-1} d_2)$, a system in the same form as (3.7) for the variable $z = S_1^{\frac{1}{2}} y_1$ is obtained; it can be solved as described before, in order to find a solution (λ, z), corresponding to the largest λ.

The solution (y_1, y_2) of (3.10) is given by

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} S_1^{-\frac{1}{2}} z \\ T_4^{-1}(d_2 - T_2^T S_1^{-\frac{1}{2}} z) \end{pmatrix}. \tag{3.12}$$

Therefore, the solution for the original problem is $x = Uy$.

3.3 Convergence of the method

In this subsection, a convergence theorem is proved. The notation $f(x) := \dfrac{\|Ax - b\|_2^2}{1 + \|x\|_2^2}$ will be used.

LEMMA 3.1. *If (λ_{k+1}, x^{k+1}) is a solution of system (3.4) corresponding to the largest λ_{k+1}, then x^{k+1} is the global minimizer of the quadratically constrained quadratic optimization problem:*

$$\min_x \; x^T B_k x - 2 d_k^T x \quad \textit{subject to } \|Lx\|_2^2 = \delta^2, \tag{3.13}$$

if such a global minimum exists.

PROOF. The proof follows the ideas in [5, Th. 1]. □

LEMMA 3.2. *The optimization problem (3.13) admits a global minimum if and only if the vector x^k satisfies*

$$\min_{x \in \mathrm{Null}\,(L^T L),\, x \neq 0} \frac{x^T A^T A x}{x^T x} \geq f(x^k) \tag{3.14}$$

PROOF. In the case L is square and nonsingular, the feasibility region $\mathcal{F} = \{x \,|\, \|Lx\|_2^2 = \delta^2\}$ is a nondegenerate ellipsoid, therefore any quadratic function attains its global minimum on $\mathcal{F}$. On the other hand, $\mathrm{Null}\,(L^T L) = \{0\}$, so any vector x^k satisfy (3.14).

If $L^T L$ is singular, any x can be uniquely written as $x_1 + x_2$, with $x_1^T x_2 = 0$, $x_1 \in \mathrm{Range} L^T L$, $x_2 \in \mathrm{Null}\,(L^T L)$. It is easy to see that the unboundedness of a quadratic function of x can only occur from the contribution of the x_2 part. In

order to attain a global minimum of the function in (3.13), it must be imposed that $x_2^T B_k x_2 \geq 0$, for any $x_2 \in \text{Null}\,(L^T L)$. From the definition formula (3.3) of $B_k = B(x^k)$, the relation (3.14) is readily obtained. □

THEOREM 3.3. *For a starting vector x^0 that satisfies (3.14) (with $k = 0$), Algorithm 3.1 provides a sequence of vectors $\{x^k\}_{k=1,2,\ldots}$, for which the function f is monotonically decreasing:*

$$0 \leq f(x^{k+1}) \leq f(x^k), \qquad \forall\ k = 0, 1, 2, \ldots. \tag{3.15}$$

Any limit point $(\overline{\lambda}, \overline{x})$ of the sequence $\{(\lambda_k, x^k)\}_k$ is a solution of the system (3.2).

PROOF. For a fixed value of k, denote by $g_k(x)$ the objective function in the minimization (3.13):

$$g_k(x) = x^T B_k x - 2 d_k^T x.$$

Suppose, for now, that the iterate x^k satisfies the assumption (3.14), thus, by Lemma 3.2, g_k admits a global minimum.

Taking into account the definition formulas (3.3) for $B_k = B(x^k)$ and $d_k = d(x^k)$, simple algebraic manipulations give

$$g_k(x) = \frac{1}{1 + \|x^k\|_2^2} \left(x^T A^T A x - f(x^k) x^T x - 2 b^T A x \right).$$

Disregarding the constant factor, the minimization of $g_k(x)$ with respect to x (subject to the quadratic constraint $\|Lx\|_2^2 = \delta^2$) is equivalent to minimizing

$$\begin{aligned} x^T A^T A x - f(x^k) x^T x - 2 b^T A x &= \|Ax - b\|_2^2 - f(x^k) x^T x - b^T b \\ &= (1 + \|x\|_2^2)(f(x) - f(x^k)) + f(x^k) - b^T b, \quad (\text{s.t. } \|Lx\|_2^2 = \delta^2). \end{aligned}$$

Therefore, the following equivalent problem is derived:

$$\min_x\ (1 + \|x\|_2^2)(f(x) - f(x^k)) \quad \text{subject to } \|Lx\|_2^2 = \delta^2. \tag{3.16}$$

Let $\overline{g}_k(x) := (1 + \|x\|_2^2)(f(x) - f(x^k))$. The iterate x^{k+1} is a solution for system (3.4), and, by Lemma 3.1, it is the global minimizer of $g_k(x)$ under the quadratic constraint $\|Lx\|_2^2 = \delta^2$. Therefore, x^{k+1} is also the optimal solution for (3.16). It implies that for any $x \in \Re^n$, $\overline{g}_k(x^{k+1}) \leq \overline{g}_k(x)$. In particular, for $x := x^k$,

$$\overline{g}_k(x^{k+1}) \leq \overline{g}_k(x^k) \Leftrightarrow (1 + \|x^{k+1}\|_2^2)(f(x^{k+1}) - f(x^k)) \leq 0 \Leftrightarrow f(x^{k+1}) \leq f(x^k).$$

Since x^0 satisfies the assumption (3.14), all the arguments above hold for the case $k = 0$. Therefore, $f(x^1) \leq f(x^0)$, and x^1 satisfies (3.14), too. By induction, all iterates x^k satisfy (3.14), and the proof holds for any k.

The second part of the theorem is trivial: making $k \to \infty$ for any convergent subsequence of $\{(\lambda_k, x^k)\}_k$, the relation

$$B_{k-1} x^k + \lambda_k L^T L x^k = d_{k-1},$$

implies $B(\overline{x})\overline{x} + \overline{\lambda} L^T L \overline{x} = d(\overline{x})$. In addition, the quadratic relation $\|Lx^k\|_2^2 = \delta^2$ is guaranteed at every iteration, and it is preserved for $\overline{x}$. Therefore, $\overline{x}$ is a solution for (3.2). □

3.4 Initial vector

Theorem 3.3 says that, when L is square and nonsingular, any random vector can be used as a starting vector in the RTLSQEP algorithm, whereas if L is rectangular, the starting vector x^0 should satisfy the condition (3.14). Equivalently, in the latter case, if N is a matrix whose columns generate Null $(L^T L)$, x^0 should satisfy

$$f(x^0) \leq \min_{y \neq 0} \frac{y^T N^T A^T A N y}{y^T N^T N y} = \sigma^2_{\min}(AN, N),$$

where $\sigma_{\min}(M, N)$ denotes the minimum generalized singular value of the matrix pair (M, N).

In the common case when L is taken as the approximation matrix of the first order derivative operator, $L = \begin{pmatrix} -1 & 1 & 0 & 0 \\ & \ddots & \ddots & \\ 0 & 0 & -1 & 1 \end{pmatrix} \in \Re^{(n-1)\times n}$, this condition is rewritten as $f(x^0) \leq \frac{1}{n} \mathbf{1}_n^T A^T A \mathbf{1}_n$, where $\mathbf{1}_n$ is the vector of all ones.

3.5 Computational remarks

For solving quadratic eigenvalue problems at every iteration, it is possible to choose between several computational methods (see [33], [25]). In particular, one could either *linearize* the QEP and then solve the (generalized) eigenvalue problem, or use a *direct* method for QEPs.

One way to linearize the QEP (3.8) is:

$$\begin{bmatrix} -2W & -W^2 + \delta^{-2} h h^T \\ I & 0 \end{bmatrix} \begin{bmatrix} v \\ u \end{bmatrix} = \lambda \begin{bmatrix} v \\ u \end{bmatrix}. \tag{3.17}$$

This eigenvalue problem should be solved in order to find the largest real eigenvalue λ and an associated eigenvector. An expensive approach consists in computing the complete eigenvalue decomposition. Such a technique is actually used by Matlab's function `polyeig`, which solves quadratic (or polynomial) eigenvalue problems by linearization. For efficiency, it is preferable to restrict the computations to finding only the rightmost eigenvalue and an associated eigenvector. The rightmost eigenvalue is not necessarily the eigenvalue of largest magnitude, therefore it is not possible to apply the power iteration method directly. Polynomial or rational preconditioning should be used in order to transform the search for the rightmost eigenvalue into a search for the largest magnitude eigenvalue.

For larger dimensions one must avoid even forming matrix W (and, of course, W^2). From the definition of W (in subsection 3.2), it is clear that matrix-vector products with W can be executed in a fast way for particular forms of nonsingular L (banded Toeplitz, for instance). For large scale problems, it is advantageous to have A sparse, but all tests presented later in this paper have dense A. Using only matrix-vector products of the matrix in (3.17), one can apply Arnoldi method for computing the largest real eigenvalue and corresponding eigenvector.

For the numerical tests in Section 4, a distinction will be made between two implementations of the RTLSQEP method. The first, which will be referred to as `rtlsqep`, uses linearization and computes the rightmost eigenpair with the ARPACK implementation [22] included in Matlab (namely the function `eigs`). The second, referred to as `frtlsqep`, uses the fast solver described in [25], which solves lower dimensional approximations of the original QEPs.

4 Numerical results

4.1 Test problems description

In order to test the performance of the proposed method, several problems from the "Regularization Tools" [16] were employed. All of them are discretizations of continuous ill-posed problems of the Fredholm integral type [18, §2], constructed by quadrature. The functions return the elements of a square system $A_{\text{true}}x_{\text{true}} = b_{\text{true}}$, with matrix A_{true} singular or very ill-conditioned.

In the classical context (i.e., without additional regularization), it is known that TLS gives more accurate results than LS when increasing the degree of overdetermination, provided entries of $(A\, b)$ are affected by independent identically distributed errors of zero mean and equal variance [35, §8]. For this reason, the RTLS approach is also tested for rectangular systems. Some example functions from "Regularization Tools" were easily modified to construct rectangular problems (by using a rectangular discretization grid instead of a square one or by "partitioning" a square system in order to form an overdetermined system).

In the following, denote by $A_{\text{true}}x_{\text{true}} = b_{\text{true}}$ a (square or rectangular) ill-conditioned example system and let σ be a noise level. By adding white noise to the data:

$$A = A_{\text{true}} + \sigma E, \qquad b = b_{\text{true}} + \sigma e,$$

with $E =$ `randn(m,n)`, $e =$ `randn(n,1)`, the problem to be solved becomes $Ax \approx b$.

The matrix $L \in \Re^{(n-1)\times n}$ is set to approximate the first order derivative operator. For the simulations below, the exact solutions are known; it is straightforward to consider as regularization condition the equality $\|Lx\|_2 = \delta := \|Lx_{\text{true}}\|_2$. In this case, it is expected to obtain a regularized solution x_R of the original ill-conditioned problem which is close to the exact solution x_{true}.

4.2 Comparison with other regularization solvers

The purpose of these tests is to numerically validate the new RTLS method, but also to compare its performance with other existing methods. The solvers employed in the tests are described in Table 4.1. Results were obtained in Matlab 6 on an i686 PC.

For several noise levels σ, relative errors $\|x_R - x_{\text{true}}\|/\|x_{\text{true}}\|$ are averaged in 200 random simulations. Table 4.2 shows results for the square problem ($m = n = 20$) and Table 4.3, results concerning the rectangular problem ($m = 200$, $n = 20$).

Table 4.1: Solvers for regularized least squares and regularized TLS

Solver	Description
`tikhonov`	Tikhonov regularization (from Hansen's "Regularization Tools" [16])
`rlsqep`	RLS solved by a QEP (with Matlab's `polyeig`)
`frlsqep`	RLS solved by a QEP (with fast Lanczos method [25])
`ttls`	Truncated Total Least Squares (from Hansen's "Regularization Tools" [16])
`rtlseig1`	Guo and Renaut's eigenvalue method for RTLS [14] (random starting vector)
`rtlseig2`	Guo and Renaut's eigenvalue method for RTLS [14] (starting vector - the `frlsqep` solution)
`rtlsqep`	RTLS by iterative QEPs (each solved by linearization, with Matlab's `eigs`)
`frtlsqep`	RTLS by iterative QEPs (each solved with fast Lanczos method [25])

The relative errors (averaged over 200 random simulations) are illustrated also in Figures 4.1 and 4.2 for two of the problems.

For the square problem or for small noise levels, there is no significant improvement of the RTLS solutions in comparison with the 'traditional' methods; in the overdetermined case and for increasing noise level, as expected, `rtlsqep` or `frtlsqep` gave the most accurate results (or as accurate as the other solvers) for many of the experiments. For the iterative RTLS methods `rtlseig1`, `rtlsqep` and `frtlsqep`, random starting vectors were used for each one of the problems. For the `rtlseig` method in [14], starting from the same random vector gave bad results (as shown in the column corresponding to `rtlseig1`); nevertheless, using the regularized least squares solution given by `frlsqep` as starting vector improved the results of the same solver (shown in the `rtlseig2` column).

In Table 4.4, average timings and number of iterations for several problem dimensions are reported. In order to have a fair comparison, the same starting vector and the same termination criterion were used for all implementations. Namely, the initial vector was set to the regularized least squares solution, and the convergence test was: $\|x_{k+1} - x_k\|/\|x_k\| < \texttt{1e-4}$.

For `rtlseig2`, each iteration involves solving an $n \times n$ linear system, requiring, in general, $\mathcal{O}(n^3)$ operations (or $\mathcal{O}(n^2)$, if A and L are first transformed via Generalized Singular Value decomposition; however, this preprocessing is of cubic complexity).

The `rtlsqep` implementation solves at each iteration a quadratic eigenproblem via linearization. The total number of matrix-vector products with matrix W in (3.17) is also shown. This implementation is in general the fastest and allows solving large problems. Results in Table 4.4 are obtained with L set to the $n \times n$ approximation matrix for the first order derivative operator. In this situation, one takes advantage of the fact that the number of *flops* for solving a system with

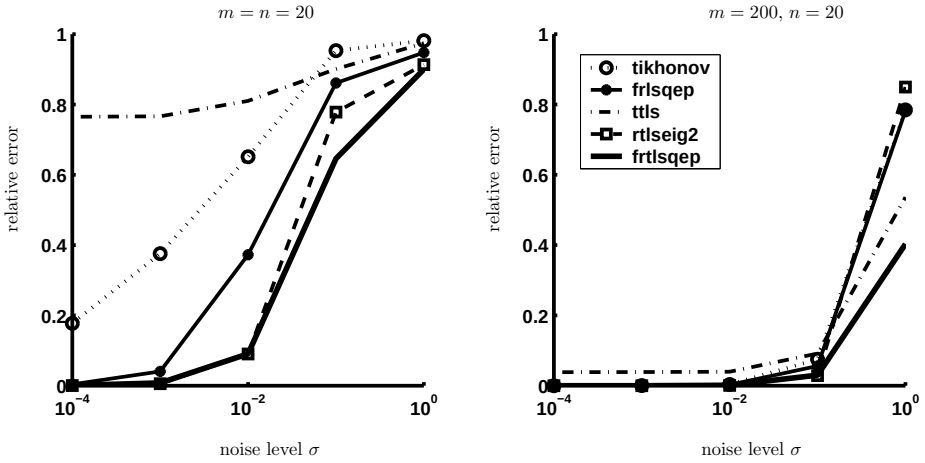


Figure 4.1: Average relative errors for example `ilaplace`

Table 4.2: The relative errors $\|x_R - x_{\text{true}}\|/\|x_{\text{true}}\|$ in several example problems, for all methods and several noise levels σ; square case, with $m = n = 20$. The smallest errors for each problem set are indicated in underlined bold numbers.

ilaplace	tikhonov	rlsqep	frlsqep	ttls	rtlseig1	rtlseig2	rtlsqep	frtlsqep
$\sigma =$ 1e-4	1.8e-1	4.1e-3	4.1e-3	7.6e-1	1.0e+0	**5.8e-4**	8.5e-4	8.5e-4
$\sigma =$ 1e-3	3.8e-1	4.1e-2	4.1e-2	7.7e-1	1.0e+0	**5.6e-3**	8.0e-3	8.0e-3
$\sigma =$ 1e-2	6.5e-1	3.7e-1	3.7e-1	8.1e-1	1.0e+0	**9.0e-2**	9.1e-2	9.1e-2
$\sigma =$ 1e-1	9.5e-1	8.6e-1	8.6e-1	9.0e-1	9.0e-1	7.8e-1	**6.5e-1**	**6.5e-1**
$\sigma =$ 1e+0	9.8e-1	9.5e-1	9.5e-1	9.7e-1	9.3e-1	9.1e-1	**9.0e-1**	**9.0e-1**
baart	tikhonov	rlsqep	frlsqep	ttls	rtlseig1	rtlseig2	rtlsqep	frtlsqep
$\sigma =$ 1e-4	1.6e-2	2.4e-2	2.1e-2	**1.3e-2**	1.0e+0	1.7e-2	1.7e-2	1.6e-2
$\sigma =$ 1e-3	2.9e-2	2.5e-2	2.5e-2	2.7e-2	1.0e+0	**2.4e-2**	**2.4e-2**	**2.4e-2**
$\sigma =$ 1e-2	3.0e-1	**8.1e-2**	**8.1e-2**	8.3e-2	1.0e+0	8.2e-2	8.2e-2	8.2e-2
$\sigma =$ 1e-1	9.5e-1	**2.6e-1**	**2.6e-1**	3.1e-1	8.1e-1	3.2e-1	3.2e-1	3.2e-1
$\sigma =$ 1e+0	9.9e-1	7.4e-1	7.4e-1	8.7e-1	8.0e-1	7.4e-1	**7.3e-1**	**7.3e-1**
shaw	tikhonov	rlsqep	frlsqep	ttls	rtlseig1	rtlseig2	rtlsqep	frtlsqep
$\sigma =$1e-4	1.1e-2	5.0e-3	**2.3e-3**	2.2e-2	1.0e+0	3.2e-3	2.6e-3	**2.3e-3**
$\sigma =$1e-3	1.2e-1	9.1e-3	9.3e-3	2.2e-2	1.0e+0	**7.3e-3**	9.2e-3	9.3e-3
$\sigma =$1e-2	3.5e-1	1.1e-1	1.1e-1	**3.0e-2**	1.0e-0	2.0e-1	2.0e-1	2.0e-1
$\sigma =$1e-1	6.2e-1	1.9e-1	1.9e-1	**1.4e-1**	9.9e-1	4.1e-1	4.1e-1	4.1e-1
$\sigma =$1e+0	9.6e-1	7.5e-1	7.5e-1	8.0e-1	8.9e-1	7.5e-1	**7.2e-1**	**7.2e-1**
deriv2	tikhonov	rlsqep	frlsqep	ttls	rtlseig1	rtlseig2	rtlsqep	frtlsqep
$\sigma =$1e-4	6.1e-4	**1.5e-4**	**1.5e-4**	5.5e-2	3.2e-1	**1.5e-4**	**1.5e-4**	**1.5e-4**
$\sigma =$1e-3	1.0e-2	**8.8e-4**	**8.8e-4**	5.6e-2	3.2e-1	8.9e-4	8.9e-4	8.9e-4
$\sigma =$ 1e-2	8.1e-2	7.4e-3	7.4e-3	8.7e-2	3.2e-1	**7.0e-3**	**7.0e-3**	**7.0e-3**
$\sigma =$ 1e-1	9.1e-1	**6.8e-2**	**6.8e-2**	2.6e-1	2.2e-1	6.9e-2	6.9e-2	6.9e-2
$\sigma =$ 1e+0	9.7e-1	7.1e-1	7.1e-1	8.8e-1	5.4e-1	6.4e-1	**4.9e-1**	**4.9e-1**

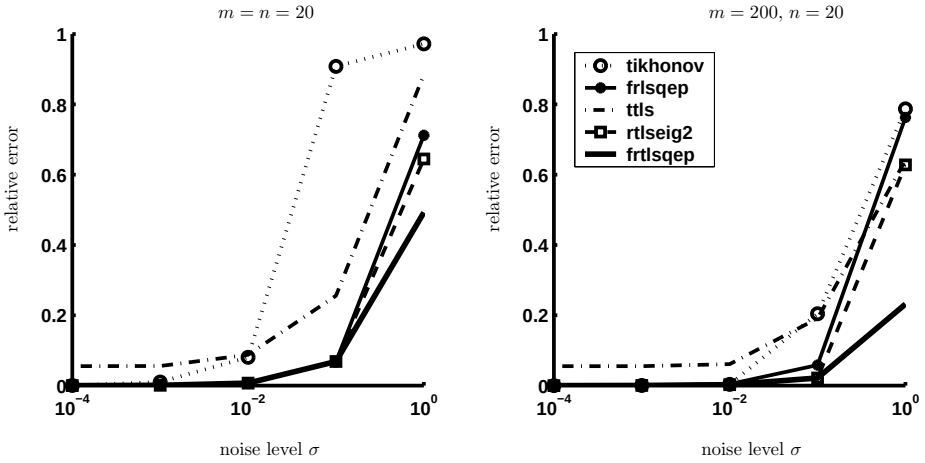


Figure 4.2: Average relative errors for example `deriv2`

Table 4.3: The relative errors $\|x_R - x_{\text{true}}\|/\|x_{\text{true}}\|$ in several example problems, for all methods and several noise levels σ; overdetermined problems, with $m = 200$, $n = 20$. The smallest errors for each problem set are indicated in underlined bold numbers.

ilaplace	tikhonov	rlsqep	frlsqep	ttls	rtlseig1	rtlseig2	rtlsqep	frtlsqep
$\sigma =$ 1e-4	2.2e-4	6.7e-6	**3.0e-6**	3.9e-2	3.2e-3	3.5e-5	6.8e-6	3.1e-6
$\sigma =$ 1e-3	4.8e-4	8.3e-5	8.4e-5	3.9e-2	3.2e-3	**4.4e-5**	8.1e-5	8.1e-5
$\sigma =$ 1e-2	3.4e-3	2.8e-3	2.8e-3	4.0e-2	3.3e-3	**6.9e-4**	1.3e-3	1.3e-3
$\sigma =$ 1e-1	7.5e-2	5.6e-2	5.6e-2	9.2e-2	3.6e-2	**2.9e-2**	3.0e-2	3.0e-2
$\sigma =$ 1e+0	7.8e-1	7.8e-1	7.8e-1	5.4e-1	8.5e-1	8.5e-1	**4.0e-1**	**4.0e-1**
baart	tikhonov	rlsqep	frlsqep	ttls	rtlseig1	rtlseig2	rtlsqep	frtlsqep
$\sigma =$ 1e-4	2.5e-2	3.7e-2	2.2e-2	**2.1e-2**	1.0e+0	**2.1e-2**	**2.1e-2**	**2.1e-2**
$\sigma =$ 1e-3	1.8e-1	4.8e-2	4.8e-2	**3.5e-2**	1.0e+0	4.0e-2	4.0e-2	4.0e-2
$\sigma =$ 1e-2	2.3e-1	1.8e-1	1.8e-1	**1.4e-1**	9.9e-1	2.7e-1	2.7e-1	2.7e-1
$\sigma =$ 1e-1	7.3e-1	7.1e-1	7.1e-1	6.2e-1	6.5e-1	7.1e-1	**5.3e-1**	**5.3e-1**
$\sigma =$ 1e+0	9.6e-1	9.5e-1	9.5e-1	9.5e-1	8.9e-1	9.0e-1	**8.1e-1**	**8.1e-1**
shaw	tikhonov	rlsqep	frlsqep	ttls	rtlseig1	rtlseig2	rtlsqep	frtlsqep
$\sigma =$ 1e-4	1.8e-3	2.8e-3	2.8e-3	**2.0e-4**	4.3e-1	3.6e-3	3.6e-3	3.6e-3
$\sigma =$ 1e-3	7.9e-2	1.6e-2	1.6e-2	**4.2e-3**	3.5e-1	3.9e-2	3.9e-2	3.9e-2
$\sigma =$ 1e-2	1.6e-1	7.3e-2	7.3e-2	**3.9e-2**	1.7e-1	1.6e-1	1.5e-1	1.5e-1
$\sigma =$ 1e-1	9.3e-1	8.9e-1	8.9e-1	9.4e-1	7.0e-1	7.4e-1	**6.7e-1**	**6.7e-1**
$\sigma =$ 1e+0	9.6e-1	9.4e-1	9.4e-1	9.6e-1	7.3e-1	7.8e-1	**7.0e-1**	**7.0e-1**
deriv2	tikhonov	rlsqep	frlsqep	ttls	rtlseig1	rtlseig2	rtlsqep	frtlsqep
$\sigma =$ 1e-4	1.2e-4	**7.5e-5**	8.5e-5	5.5e-2	3.2e-1	8.3e-5	8.9e-5	8.9e-5
$\sigma =$ 1e-3	**3.2e-4**	**3.2e-4**	**3.2e-4**	5.5e-2	3.2e-1	3.7e-4	3.8e-4	3.8e-4
$\sigma =$ 1e-2	3.5e-3	4.4e-3	4.4e-3	6.1e-2	3.2e-1	**2.7e-3**	**2.7e-3**	**2.7e-3**
$\sigma =$ 1e-1	2.0e-1	5.8e-2	5.8e-2	1.9e-1	3.2e-1	**2.2e-2**	**2.2e-2**	**2.2e-2**
$\sigma =$ 1e+0	7.9e-1	7.6e-1	7.6e-1	6.4e-1	2.8e-1	6.3e-1	**2.3e-1**	**2.3e-1**

Table 4.4: Average timings in two example problems, for several RTLS methods, and for various problem dimensions. 'Iter' = number of iterations; 'CPU' = CPU time (in seconds); '$W \times v$' = number of matrix-vector products (with matrix W); '–' denotes problems that were not solved in comparable time (> 30 min).

		`rtlseig2`		`rtlsqep`			`frtlsqep`	
baart		Iter	CPU	Iter	CPU	$W \times v$	Iter	CPU
$m = n$	$n = 50$	406.8	0.62	4.0	0.17	135	4.0	0.15
$m = 2n$	$n = 50$	241.7	0.41	4.1	0.18	140	4.1	0.17
$m = n$	$n = 500$	777.0	364.67	4.0	1.56	135	4.0	19.00
$m = 2n$	$n = 500$	–	–	4.0	3.14	135	4.0	19.53
$m = 2n$	$n = 1000$	–	–	4.0	18.17	135	–	–
$m = 2n$	$n = 2500$	–	–	4.0	108.21	135	–	–
$m = n$	$n = 5000$	–	–	4.0	132.77	135	–	–
$m = 2n$	$n = 5000$	–	–	4.0	255.27	135	–	–
deriv2		Iter	CPU	Iter	CPU	$W \times v$	Iter	CPU
$m = n$	$n = 50$	35.9	0.06	4.0	0.14	135	4.1	0.16
$m = 2n$	$n = 50$	23.0	0.04	4.0	0.15	135	4.0	0.15
$m = n$	$n = 500$	34.3	19.18	4.0	1.38	135	4.0	18.30
$m = 2n$	$n = 500$	34.6	20.03	4.3	3.26	148	4.0	18.86
$m = 2n$	$n = 1000$	–	–	4.0	15.03	135	–	–
$m = 2n$	$n = 2500$	–	–	4.0	105.10	135	–	–

L is linear in n. Therefore, each matrix-vector product with $W = L^{-T} B_k L^{-1}$ is of order $\mathcal{O}(2mn + 3n)$ operations.

Note that for both `rtlsqep` and `frtlsqep`, a very small number of iterations (under 5) is required for each example, *i.e.*, at most 5 quadratic eigenvalue problems were solved. Moreover, for various problem sizes, almost the same number of matrix-vector products were performed (for the `rtlsqep` approach). This number is actually linked to the convergence of the Arnoldi method (ARPACK's `eigs`) and shows a stability of the number of Arnoldi steps with respect to problem dimensions.

4.3 Comparison between RTLS optimization solvers

The newly proposed RTLS method using quadratic eigenvalue problems is, in fact, a method for solving a quadratically constrained nonconvex optimization problem; so is the `rtlseig` method of [14]. Numerical experiments confirm that classical optimization methods are not as suited for the RTLS problem as the tailored RTLS method described herein. For the quasi-Newton method (function `fmincon` from Matlab), a very good initial approximation must be used in order to have convergence. Decreasing the default tolerance values (even with a 'good' initial vector) has also the effect of non-convergence (after 10^5 iterations).

In Table 4.5, the average over 200 random simulations of the objective function value $f(\cdot)$ is reported for several examples and methods. For reference, the function f is also evaluated in the exact solution x_{true} (which is the exact solution of the unperturbed example system $A_{\text{true}} x_{\text{true}} = b_{\text{true}}$, and not the optimal solution of the RTLS problem!). The vector x_{true} is actually used as the

Table 4.5: Average of the function value in 200 random simulations ($m = 200$, $n = 20$, $\sigma = 0.001$). $f(x_{\text{true}})$ is given just for reference.

	`ilaplace`	`baart`	`shaw`	`deriv2`
$f(x_{\text{true}})$	1.983e-4	2.005e-4	2.003e-4	2.018e-4
$f(x_{\texttt{rtlseig2}})$	1.932e-4	*1.902*e-4	*1.944*e-4	1.936e-4
$f(x_{\texttt{frtlsqep}})$	*1.920*e-4	*1.902*e-4	*1.944*e-4	*1.932*e-4
$f(x_{\texttt{fmincon}})$	1.961e-4	1.991e-4	1.982e-4	1.983e-4

Table 4.6: Average of the quadratic constraint violation in 200 random simulations ($m = 200$, $n = 20$, $\sigma = 0.001$).

	`ilaplace`	`baart`	`shaw`	`deriv2`
$x_{\texttt{rtlseig2}}$	8.51e-04	1.99e-07	6.47e-09	1.76e-04
$x_{\texttt{frtlsqep}}$	3.08e-13	1.58e-13	2.24e-17	4.30e-14
$x_{\texttt{fmincon}}$	2.53e-07	2.48e-08	1.88e-08	7.46e-03

initial approximation for `fmincon`. Note that the solution provided by `fmincon` improves just a little bit the value at the initial approximation. In contrast, the objective function value at the `frtlsqep` solution is the smallest, for all examples.

Another remark on the RTLSQEP method is that the quadratic constraint $\|Lx\|_2^2 = \delta^2$ is preserved at each iteration (at least in exact arithmetic). In Table 4.6, the constraint violation $|\|Lx\|_2^2 - \delta^2|$ is averaged in 200 simulations for the RTLS solutions computed by three methods (`rtlseig2`, `frtlsqep`, and `fmincon`, the latter with initial approximation x_{true}). Again, `frtlsqep` is much more accurate than the other two solvers. The results in this section show the good numerical performances of the RTLSQEP algorithm as a specialized nonlinear optimization solver. In practice, however, the merit that the constraint is satisfied with high accuracy is not so important, because the parameter δ might not be known exactly.

4.4 Importance of the starting vector

The implementations of the RTLS method using either the fast QEP solver [25] or the Arnoldi method for the linearized eigenvalue problem are quite robust with respect to the chosen starting vector. For many of the problems, there is no need to seek for a certain initial vector, because random vectors satisfy the condition required for convergence of the method. To a certain extent, the convergence depends also on the problem dimensions (i.e., square or rectangular) and on the noise level. In Table 4.7, percentages of 'good' solutions for the test problem `ilaplace` using the solver `frtlsqep` are shown. The tolerances to which the relative errors were compared are also shown.

5 Conclusions

From both theoretical and practical points of view, the regularized total least squares problem is a necessary extension of the regularized least squares for-

Table 4.7: (a) Percentage of `frtlsqep` solutions close to (within a given tolerance) the exact solution in example `ilaplace` with $n = 30$, for several noise levels σ and dimensions m, in 1000 runs with different random starting vectors `rand(n,1)`.
(b) Tolerances for the relative errors between `frtlsqep` solutions and the exact solution.

(a)

$\sigma \setminus \frac{m}{n}$	1	5	10	20
0.01	100%	99.9%	98.5%	96.6%
0.1	100%	99.5%	98.1%	96.8%
1	0%	99.7%	98.8%	96.1%
2	0%	99.5%	99.1%	96.6%

(b)

$\sigma \setminus \frac{m}{n}$	1	5	10	20
0.01	5e-4	1e-4	1e-4	5e-5
0.1	5e-4	1e-4	1e-4	1e-4
1	1e+0	5e-2	5e-3	5e-4
2	1e+0	5e-2	1e-2	1e-3

mulation. At present, there are several proposed methods for solving the RTLS problem. In this paper, an alternative approach called RTLSQEP was presented. It involves solving iteratively quadratic eigenvalue problems. An advantage is that either standard or fast methods can be used for solving the specific QEPs. All options were numerically tested and validated in Matlab implementations of the method. Dense problems with dimensions up to several thousands of rows and columns could be solved using the Arnoldi method applied to the linearized QEPs. An important remark is that the RTLSQEP method is *robust*: the initial vector of the iterative algorithm can be arbitrarily chosen in most of the cases.

The drawback of this and similar methods is that they require an exact specification of the constraint parameter δ. In real-life problems, such a parameter is rarely available *a priori*, therefore it should be estimated from given data, using, for instance, cross-validation.

Acknowledgments

The authors would like to thank Ren-Cang Li and Qiang Ye for providing the Matlab code of their fast quadratic eigenvalue problem method.

Dr. Sabine Van Huffel is a full professor and Diana M. Sima is a research assistant at the Katholieke Universiteit Leuven, Belgium. Their research was supported by

Research Council KUL: GOA-Mefisto 666, IDO/99/003 and IDO/02/009 (Predictive computer models for medical classification problems using patient data and expert knowledge), several PhD/postdoc & fellow grants;
Flemish Government:

- **FWO:** PhD/postdoc grants, projects, G.0200.00 (damage detection in composites by optical fibers), G.0078.01 (structured matrices), G.0407.02 (support vector machines), G.0269.02 (magnetic resonance spectroscopic imaging), G.0270.02 (nonlinear Lp approximation), research communities (ICCoS, ANMMM);
- **AWI:** Bil. Int. Collaboration Hungary/Poland;
- **IWT:** PhD Grants;

Belgian Federal Government: DWTC (IUAP IV-02 (1996-2001) and IUAP V-22 (2002-2006): Dynamical Systems and Control: Computation, Identification & Modelling));

EU: NICONET, eTUMOUR, PDT-COIL,BIOPATTERN;
Contract Research/agreements: Data4s, IPCOS.

The work of Gene H Golub was in part supported by the National Science Foundation under Grant No. CCR-9971010.

REFERENCES

1. T. F. Budinger. Medical imaging techniques. `http://imasun.lbl.gov/~budinger/medhome.html`.
2. D. Calvetti and L. Reichel. Lanczos-based exponential filtering for discrete ill-posed problems. *Numerical Algorithms*, 29:45–65, 2002.
3. Y.-H. De Roeck. Sparse linear algebra and geophysical migration: a review of direct and iterative methods. *Numerical Algorithms*, 29:283–322, 2002.
4. 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(1):1223–1241, 1997.
5. W. Gander. Least squares with a quadratic constraint. *Numerische Mathematik*, 36:291–307, 1981.
6. W. Gander, G. H. Golub, and U. von Matt. A constrained eigenvalue problem. *Linear Algebra Appl.*, 114/115:815–839, 1989.
7. G. H. Golub, P. C. Hansen, and D. P. O'Leary. Tikhonov regularization and Total Least Squares. *SIMAX*, 21(1):185–194, 1999.
8. G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 1979.
9. G. H. Golub, A. Hoffman, and G. W. Stewart. A generalization of the Eckhard-Young-Mirsky matrix approximation theorem. *Linear Algebra Appl.*, 88/89:317–327, 1987.
10. G. H. Golub and C. F. Van Loan. An analysis of the total least squares problem. *SIAM J. Numerical Analysis*, 17:883–893, 1980.
11. G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, Maryland, third edition, 1996.
12. G. H. Golub and U. von Matt. Generalized cross-validation for large scale problems. In S. Van Huffel, editor, *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling*, pages 139–148. SIAM, 1997.
13. G. H. Golub and U. von Matt. Tikhonov regularization for large scale problems. Technical Report SCCM-97-03, Stanford University, 1997.
14. H. Guo and R. Renaut. A regularized Total Least Squares algorithm. In S. Van Huffel and P. Lemmerling, editors, *Total Least Squares and Errors-in-Variables Modeling*, pages 57–66. Kluwer, 2002.
15. P. C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 32:561–580, 1992.
16. P. C. Hansen. Regularization Tools, a Matlab package for analysis of discrete regularization problems. *Numerical Algorithms*, 6:1–35, 1994.
17. P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems. Numerical Aspects of Linear Inversion*. SIAM, Philadelphia, 1997.
18. P. C. Hansen. Deconvolution and regularization with Toeplitz matrices. *Numerical Algorithms*, 29:323–378, 2002.

19. P. C. Hansen and D. P. O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comp.*, 14:1487–1503, 1993.
20. P. R. Johnston and R. M. Gulrajani. Selecting the corner in the L-curve approach to Tikhonov regularization. *IEEE Tran. Biomedical Eng.*, 47(9):1293–1296, September 2000.
21. M. E. Kilmer and D. P. O'Leary. Choosing regularization parameters in iterative methods for ill-posed problems. *SIMAX*, 22(3):1204–1221, 2001.
22. R. B. Lehoucq, D. C. Sorensen, and C. Yang. *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*. SIAM, Philadelphia, PA, 1998.
23. P. Lemmerling. *Structured Total Least Squares: Analysis, algorithms and applications*. PhD thesis, Electrical Engineering Department, K.U. Leuven, Belgium, 1999.
24. P. Lemmerling, N. Mastronardi, and S. Van Huffel. Fast algorithm for solving Hankel/Toeplitz structured total least squares problem. *Numerical Algorithms*, 21:371–392, 2000.
25. R.-C. Li and Q. Ye. A Krylov subspace method for quadratic matrix polynomials with application to constrained least squares problems. *SIMAX*, 25(2):405–428, 2003.
26. N. Mastronardi, P. Lemmerling, and S. Van Huffel. Fast regularized structured Total Least Squares problems for solving the basic deconvolution problem. To appear in *Numerical Linear Alg. Appl.*, 2004.
27. V. A. Morozov. On the solution of functional equations by the method of regularization. *Soviet. Math. Dokl.*, 7:414–417, 1966.
28. D. P. O'Leary. Near-optimal parameters for Tikhonov regularization and other regularization methods. Computer Science Department report CS-TR-4004, Institute for Advanced Computer Studies report UMIACS-TR-99-17, University of Maryland, 1999.
29. C. C. Paige and M. A. Saunders. LSQR: an algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Software*, 8:43–71, 1998.
30. A. Pruessner and D. P. O'Leary. Blind deconvolution using a regularized structured total least norm algorithm. Technical Report CS-TR-4287, University of Maryland, Computer Science Department, 2001.
31. A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet. Math. Dokl.*, 4:1035–1038, 1963.
32. A. N. Tikhonov and V. Arsenin. *Solutions of Ill-Posed Problems*. Winston & Sons, Washington, DC, USA, 1977.
33. F. Tisseur and K. Meerbergen. The quadratic eigenvalue problem. *SIAM Review*, 43(2):235–286, 2001.
34. A. M. Urmanov, A. V. Gribok, H. Bozdogan, J. W. Hines, and R. E. Uhrig. Information complexity-based regularization parameter selection for solution of ill conditioned inverse problems. *Inverse Problems*, 18, 2002. Institute of Physics Publishing.
35. S. Van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis*, volume 9 of *Frontiers in Applied Mathematics*. SIAM, Philadelphia, 1991.
36. N. H. Younan and X. Fan. Signal restoration via the regularized constrained Total Least Squares. *Signal Processing*, 71:85–93, 1998.