Electronic Transactions on Numerical Analysis.
Volume 31, pp. 12-24, 2008.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

A FAST ALGORITHM FOR SOLVING REGULARIZED TOTAL LEAST SQUARES PROBLEMS*

JÖRG LAMPE† AND HEINRICH VOSS†

Abstract. The total least squares (TLS) method is a successful approach for linear problems if both the system matrix and the right hand side are contaminated by some noise. For ill-posed TLS problems Renaut and Guo [SIAM J. Matrix Anal. Appl., 26 (2005), pp. 457–476] suggested an iterative method based on a sequence of linear eigenvalue problems. Here we analyze this method carefully, and we accelerate it substantially by solving the linear eigenproblems by the Nonlinear Arnoldi method (which reuses information from the previous iteration step considerably) and by a modified root finding method based on rational interpolation.

Key words. Total least squares, regularization, ill-posedness, Nonlinear Arnoldi method.

AMS subject classifications. 15A18, 65F15, 65F20, 65F22.

1. Introduction.

Many problems in data estimation are governed by overdetermined linear systems

$$Ax \approx b, \quad A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m, m \geq n, \tag{1.1}$$

where both the matrix A and the right hand side b are contaminated by some noise. An appropriate approach to this problem is the total least squares (TLS) method which determines perturbations $\Delta A \in \mathbb{R}^{m\times n}$ to the coefficient matrix and $\Delta b \in \mathbb{R}^m$ to the vector b such that

$$\|[\Delta A, \Delta b]\|_F^2 = \min! \quad \text{subject to } (A + \Delta A)x = b + \Delta b, \tag{1.2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix; see, e.g., [7, 18].

In this paper we consider ill-conditioned problems which arise, for example, from the discretization of ill-posed problems such as integral equations of the first kind; see, e.g., [4, 8, 11]. Then least squares or total least squares methods for solving (1.1) often yield physically meaningless solutions, and regularization is necessary to stabilize the solution.

Motivated by Tikhonov regularization a well established approach is to add a quadratic constraint to the problem (1.2) yielding the regularized total least squares (RTLS) problem

$$\|[\Delta A, \Delta b]\|_F^2 = \min! \quad \text{subject to } (A + \Delta A)x = b + \Delta b,\ \|Lx\| \leq \delta, \tag{1.3}$$

where (as in the rest of the paper) $\|\cdot\|$ denotes the Euclidean norm, $\delta > 0$ is a regularization parameter, and $L \in \mathbb{R}^{k\times n}$, $k \leq n$ defines a (semi-) norm on the solution through which the size of the solution is bounded or a certain degree of smoothness can be imposed on the solution. Stabilization by introducing a quadratic constraint was extensively studied in [2, 6, 9, 15, 16, 17]. Tikhonov regularization was considered in [1].

Based on the singular value decomposition of $[A, b]$, methods were developed for solving the TLS problem (1.2) [7, 18], and even a closed formula for its solution is known. However, this approach can not be generalized to the RTLS problem (1.3). Golub, Hansen and O'Leary [6] presented and analyzed the properties of regularization of TLS. Inspired by the fact that quadratically constrained least squares problems can be solved by a quadratic eigenvalue problem [5], Sima, Van Huffel, and Golub [16, 17] developed an iterative method for solving (1.3), where at each step the right–most eigenvalue and corresponding eigenvector of a

* Received December 17, 2007. Accepted April 25, 2008. Published online on September 10, 2008. Recommended by Zdeněk Strakoš.

†Institute of Numerical Simulation, Hamburg University of Technology, D-21071 Hamburg, Germany ({joerg.lampe,voss}@tu-hamburg.de)

ETNA
Kent State University
etna@mcs.kent.edu

quadratic eigenproblem has to be determined. Using a minimax characterization for nonlinear eigenvalue problems [20], we analyzed the occurring quadratic eigenproblems and taking advantage of iterative projection methods and updating techniques we accelerated the method substantially [13]. Its global convergence was proved in [14]. Beck, Ben-Tal, and Teboulle [2] proved global convergence for a related iteration scheme by exploiting optimization techniques.

A different approach was presented by Guo and Renaut [9, 15] who took advantage of the fact that the RTLS problem (1.3) is equivalent to the minimization of the Rayleigh quotient of the augmented matrix $M := [A, b]^T[A, b]$ subject to the regularization constraint. For solving the RTLS problem a real equation $g(\theta) = 0$ has to be solved where at each step of an iterative process the smallest eigenvalue and corresponding eigenvector of the matrix

$$B(\theta) = M + \theta N, \quad \text{with } N := \begin{bmatrix} L^T L & 0 \\ 0 & -\delta^2 \end{bmatrix},$$

is determined by Rayleigh quotient iteration. The iterates θ_k are the current approximations to the root of g. To enforce convergence the iterates θ_k are modified by backtracking such that they are all located on the same side of the root which hampers the convergence of the method.

Renaut and Guo [15] tacitly assume in their analysis of the function g that the smallest eigenvalue of $B(\theta)$ is simple which is in general not the case. In this paper we alter the definition of g, and we suggest two modifications of the approach of Guo and Renaut thus accelerating the method considerably. We introduce a solver for $g(\theta) = 0$ based on a rational interpolation of g^{-1} which exploits the known asymptotic behavior of g. We further take advantage of the fact that the matrices $B(\theta_k)$ converge as θ_k approaches the root of g. This suggests solving the eigenproblems by an iterative projection method thus reusing information from the previous eigenproblems.

The paper is organized as follows. In Section 2 we briefly summarize the mathematical formulation of the RTLS problem, we introduce the modified function g, and we analyze it. Section 3 presents the modifications of Renaut's and Guo's method which were mentioned in the last paragraph. Numerical examples from the "Regularization Tools" [10, 12] in Section 3 demonstrate the efficiency of the method.

In this paper the minimum of a Rayleigh quotient on a subspace appears at many instances. It goes without saying that the zero element is excluded of the mentioned subspace in all of these cases.

2. Regularized Total Least Squares. It is well known (cf. [18], and [2] for a different derivation) that the RTLS problem (1.3) is equivalent to

$$\frac{\|Ax - b\|^2}{1 + \|x\|^2} = \min! \quad \text{subject to } \|Lx\|^2 \le \delta^2. \tag{2.1}$$

We assume that the problem (2.1) is solvable, which is the case if the row rank of L is n or if $\sigma_{\min}([AF, b]) < \sigma_{\min}(AF)$ where the columns of the matrix F form an orthonormal basis of the null space $\mathcal{N}(L)$ of L, and $\sigma_{\min}(\cdot)$ denotes the minimal singular value of its argument; cf. [1].

We assume that the regularization parameter $\delta > 0$ is less than $\|Lx_{TLS}\|$, where x_{TLS} denotes the solution of the total least squares problem (1.2) (otherwise no regularization would be necessary). Then at the optimal solution of (2.1) the constraint $\|Lx\| \le \delta$ holds with equality, and we may replace (2.1) by

$$\phi(x) := \frac{\|Ax - b\|^2}{1 + \|x\|^2} = \min! \quad \text{subject to } \|Lx\|^2 = \delta^2. \tag{2.2}$$

The following first order condition was proved in [6]:

THEOREM 2.1. *The solution x_{RTLS} of RTLS problem (2.2) solves the problem*

$$(A^TA + \lambda_I I_n + \lambda_L L^T L)x = A^T b, \tag{2.3}$$

where

$$\lambda_I = \lambda_I(x_{RTLS}) = -\phi(x_{RTLS}), \tag{2.4}$$

$$\lambda_L = \lambda_L(x_{RTLS}) = -\frac{1}{\delta^2}(b^T(Ax_{RTLS} - b) + \phi(x_{RTLS})). \tag{2.5}$$

We take advantage of the following first order conditions which were proved in [15]:

THEOREM 2.2. *The solution x_{RTLS} of RTLS problem (2.2) satisfies the augmented eigenvalue problem*

$$B(\lambda_L)\begin{bmatrix} x_{RTLS} \\ -1 \end{bmatrix} = -\lambda_I \begin{bmatrix} x_{RTLS} \\ -1 \end{bmatrix},$$

where λ_L and λ_I are given in (2.4) and (2.5).

Conversely, if $((\hat{x}^T, -1)^T, -\hat{\lambda})$ is an eigenpair of $B(\lambda_L(\hat{x}))$ where $\lambda_L(\hat{x})$ is recovered according to (2.5), then $\hat{x}$ satisfies (2.3), and $\hat{\lambda} = -\phi(\hat{x})$.

Theorem 2.2 suggests the following approach to solving the regularized total least squares problem (2.2) (as proposed by Renaut and Guo [15]): determine θ such that the eigenvector $(x_\theta^T, -1)^T$ of $B(\theta)$ corresponding to the smallest eigenvalue satisfies the constraint $\|Lx_\theta\|^2 = \delta^2$, i.e., find a non-negative root θ of the real function

$$g(\theta) := \frac{\|Lx_\theta\|^2 - \delta^2}{1 + \|x_\theta\|^2} = 0. \tag{2.6}$$

Renaut and Guo claim that (under the conditions $b^TA \neq 0$ and $\mathcal{N}(A) \cap \mathcal{N}(L) = \{0\}$) the smallest eigenvalue of $B(\theta)$ is simple, and that (under the further condition that the matrix $[A, b]$ has full rank) g is continuous and strictly monotonically decreasing. Hence, $g(\theta) = 0$ has a unique root θ_0, and the corresponding eigenvector (scaled appropriately) yields the solution of the RTLS problem (2.2).

Unfortunately these assertions are not true. The last component of an eigenvector corresponding to the smallest eigenvalue of $B(\theta)$ need not be different from zero, and in that case $g(\theta)$ is not necessarily defined. A problem of this type is given in the following example.

EXAMPLE 2.3. Let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \; b = \begin{bmatrix} 1 \\ 0 \\ \sqrt{3} \end{bmatrix}, \; L = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 1 \end{bmatrix}, \; \delta = 1.$$

Then the conditions '$[A, b]$ has full rank', '$b^TA = (1, 0) \neq 0$', and '$\mathcal{N}(A) \cap \mathcal{N}(L) = \{0\}$' are satisfied. Furthermore,

$$B(\theta) = \begin{bmatrix} 1 + 2\theta & 0 & 1 \\ 0 & 1 + \theta & 0 \\ 1 & 0 & 4 - \theta \end{bmatrix},$$

and the smallest eigenvalues $\lambda_{\min}(B(0.5)) = 1.5$ and $\lambda_{\min}(B(1)) = 2$ of

$$B(0.5) = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 1.5 & 0 \\ 1 & 0 & 3.5 \end{bmatrix} \quad \text{and} \quad B(1) = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix}$$

ETNA
Kent State University
etna@mcs.kent.edu

have multiplicity 2, and for $\theta \in (0.5, 1)$ the last component of the eigenvector $y_\theta = (0, 1, 0)^T$ corresponding to the minimal eigenvalue $\lambda_{\min}(B(\theta)) = 1 + \theta$ is equal to 0. This means that g is undefined in the interval $(0.5, 1)$ because $(x_\theta^T, -1)^T$ cannot be an eigenvector.

To fill this gap we generalize the definition of g in the following way:

DEFINITION 2.4. *Let $\mathcal{E}(\theta)$ denote the eigenspace of $B(\theta)$ corresponding to its smallest eigenvalue, and let $N := \begin{bmatrix} L^T L & 0 \\ 0 & -\delta^2 \end{bmatrix}$. Then*

$$g(\theta) := \min_{y \in \mathcal{E}(\theta)} \frac{y^T N y}{y^T y} = \min_{(x^T, x_{n+1})^T \in \mathcal{E}(\theta)} \frac{\|Lx\|^2 - \delta^2 x_{n+1}^2}{\|x\|^2 + x_{n+1}^2} \tag{2.7}$$

is the minimal eigenvalue of the projection of N to $\mathcal{E}(\theta)$.

This extends the definition of g to the case of eigenvectors with zero last components.

THEOREM 2.5. *Assume that $\sigma_{\min}([AF, b]) < \sigma_{\min}(AF)$ holds, where the columns of $F \in \mathbb{R}^{n,n-k}$ form an orthonormal basis of the null space of L. Then $g : [0, \infty) \to \mathbb{R}$ has the following properties:*

(i) if $\sigma_{\min}([A, b]) < \sigma_{\min}(A)$, then $g(0) > 0$;
(ii) $\lim_{\theta\to\infty} g(\theta) = -\delta^2$;
(iii) if the smallest eigenvalue of $B(\theta_0)$ is simple, then g is continuous at θ_0;
(iv) g is monotonically not increasing on $[0, \infty)$;
(v) let $g(\hat\theta) = 0$ and let $y \in \mathcal{E}(\hat\theta)$ be such that $g(\hat\theta) = y^T N y/\|y\|^2$, then the last component of y is different from 0;
(vi) g has at most one root.

Proof. (i): Let $y \in \mathcal{E}(0)$. From $\sigma_{\min}([A, b]) < \sigma_{\min}(A)$ it follows that $y_{n+1} \neq 0$ and $x_{TLS} := -y(1 : n)/y_{n+1}$ solves the total least squares problem (1.2); see [18]. Hence, $\delta < \|Lx_{TLS}\|$ implies $g(0) > 0$.

(ii): $B(\theta)$ has exactly one negative eigenvalue for sufficiently large θ (see [15]) and the corresponding eigenvector converges to the unit vector e_{n+1} having one in its last component. Hence, $\lim_{\theta\to\infty} g(\theta) = -\delta^2$.

(iii): If the smallest eigenvalue of $B(\theta_0)$ is simple for some θ_0, then in a neighborhood of θ_0 the smallest eigenvalue of $B(\theta)$ is simple as well, and the corresponding eigenvector y_θ depends continuously on θ if it is scaled appropriately. Hence, g is continuous at θ_0.

(iv): Let $y_\theta \in \mathcal{E}(\theta)$ be such that $g(\theta) = y_\theta^T N y_\theta / y_\theta^T y_\theta$. For $\theta_1 \neq \theta_2$ it holds that

$$\frac{y_{\theta_1}^T B(\theta_1) y_{\theta_1}}{\|y_{\theta_1}\|^2} \leq \frac{y_{\theta_2}^T B(\theta_1) y_{\theta_2}}{\|y_{\theta_2}\|^2} \quad \text{and} \quad \frac{y_{\theta_2}^T B(\theta_2) y_{\theta_2}}{\|y_{\theta_2}\|^2} \leq \frac{y_{\theta_1}^T B(\theta_2) y_{\theta_1}}{\|y_{\theta_1}\|^2}.$$

Adding these inequalities and subtracting equal terms on both sides yields

$$\theta_1 \frac{y_{\theta_1}^T N y_{\theta_1}}{\|y_{\theta_1}\|^2} + \theta_2 \frac{y_{\theta_2}^T N y_{\theta_2}}{\|y_{\theta_2}\|^2} \leq \theta_1 \frac{y_{\theta_2}^T N y_{\theta_2}}{\|y_{\theta_2}\|^2} + \theta_2 \frac{y_{\theta_1}^T N y_{\theta_1}}{\|y_{\theta_1}\|^2},$$

i.e.,

$$(\theta_1 - \theta_2)(g(\theta_1) - g(\theta_2)) \leq 0,$$

which demonstrates that g is monotonically not increasing.

(v): Assume that there exists $y = \begin{bmatrix} \hat x \\ 0 \end{bmatrix} \in \mathcal{E}(\hat\theta)$ with

$$0 = \frac{y^T N y}{y^T y} = \frac{\hat x^T L^T L \hat x}{\hat x^T \hat x}.$$

Then $\hat{x} \in \mathcal{N}(L)$, and the minimum eigenvalue $\lambda_{\min}(\hat{\theta})$ of $B(\hat{\theta})$ satisfies

$$\begin{aligned}\lambda_{\min}(B(\hat{\theta})) &= \min_{z\in\mathbb{R}^{n+1}} \frac{z^T B(\hat{\theta}) z}{z^T z} = \frac{[\hat{x}^T, 0]\, B(\hat{\theta}) \begin{bmatrix}\hat{x}\\0\end{bmatrix}}{\hat{x}^T\hat{x}} \\ &= \min_{x\in\mathcal{N}(L)} \frac{[x^T, 0]\, B(\hat{\theta}) \begin{bmatrix}x\\0\end{bmatrix}}{x^T x} = \min_{x\in\mathcal{N}(L)} \frac{x^T A^T A x}{x^T x}.\end{aligned}$$

That is,

$$\lambda_{\min}(B(\hat{\theta})) = (\sigma_{\min}(AF))^2. \tag{2.8}$$

On the other hand, for every $x \in \mathcal{N}(L)$ and $\alpha \in \mathbb{R}$, it holds that

$$\begin{aligned}\lambda_{\min}(B(\hat{\theta})) &= \min_{z\in\mathbb{R}^{n+1}} \frac{z^T B(\hat{\theta}) z}{z^T z} \le \frac{[x^T, \alpha]\, B(\hat{\theta}) \begin{bmatrix}x\\\alpha\end{bmatrix}}{x^T x + \alpha^2} \\ &= \frac{[x^T, \alpha]\, M \begin{bmatrix}x\\\alpha\end{bmatrix}}{x^T x + \alpha^2} - \hat{\theta}\frac{\alpha^2\delta^2}{x^T x + \alpha^2} \le \frac{[x^T, \alpha]\, M \begin{bmatrix}x\\\alpha\end{bmatrix}}{x^T x + \alpha^2},\end{aligned}$$

which implies

$$\lambda_{\min}(B(\hat{\theta})) \le \min_{x\in\mathcal{N}(L),\ \alpha\in\mathbb{R}} \frac{[x^T, \alpha]\, M \begin{bmatrix}x\\\alpha\end{bmatrix}}{x^T x + \alpha^2} = (\sigma_{\min}([AF, b]))^2,$$

contradicting (2.8) and the solvability condition $\sigma_{\min}([AF, b]) < \sigma_{\min}(AF)$ of the RTLS problem (2.2).

(vi): Assume that the function $g(\theta)$ has two roots $g(\theta_1) = g(\theta_2) = 0$ with $\theta_1 \neq \theta_2$. With $\mathcal{E}(\theta)$ being the invariant subspace corresponding to the smallest eigenvalue of $B(\theta)$ it holds that

$$g(\theta_j) = \min_{y\in\mathcal{E}(\theta_j)} \frac{y^T N y}{y^T y} =: \frac{y_j^T N y_j}{y_j^T y_j} = 0,\ \ j = 1, 2.$$

Hence, the smallest eigenvalue $\lambda_{\min,1}(B(\theta_1))$ satisfies

$$\lambda_{\min,1}(B(\theta_1)) = \min_y \frac{y^T B(\theta_1) y}{y^T y} = \frac{y_1^T M y_1}{y_1^T y_1} + \theta_1 \frac{y_1^T N y_1}{y_1^T y_1} = \frac{y_1^T M y_1}{y_1^T y_1},$$

which implies

$$\frac{y_1^T M y_1}{y_1^T y_1} = \min_y \frac{y^T B(\theta_1) y}{y^T y} \le \frac{y_2^T B(\theta_1) y_2}{y_2^T y_2} = \frac{y_2^T M y_2}{y_2^T y_2}. \tag{2.9}$$

Interchanging θ_1 and θ_2 we see that both sides in (2.9) are equal, and therefore it holds that $\lambda_{\min} := \lambda_{\min,1}(B(\theta_1)) = \lambda_{\min,2}(B(\theta_2))$.

By part (v) the last components of y_1 and y_2 are different from zero. So, let us scale $y_1 = [x_1^T, -1]^T$ and $y_2 = [x_2^T, -1]^T$.

We distinguish two cases:

Case 1: Two solution pairs $(\theta_1, [x^T, -1]^T)$ and $(\theta_2, [x^T, -1]^T)$ exist, with $\theta_1 \neq \theta_2$ but with the same vector x. Then the last row of

$$B(\theta_j)\begin{bmatrix} x \\ -1 \end{bmatrix} = \begin{bmatrix} A^TA + \theta L^TL & A^Tb \\ b^TA & b^Tb - \theta_j\delta^2 \end{bmatrix}\begin{bmatrix} x \\ -1 \end{bmatrix} = \lambda_{\min}\begin{bmatrix} x \\ -1 \end{bmatrix}, \quad j = 1, 2, \tag{2.10}$$

yields the contradiction $\theta_1 = \frac{1}{\delta^2}(b^TAx - b^Tb + \lambda_{\min}) = \theta_2$.

Case 2: Two solution pairs $(\theta_1, [x_1^T, -1]^T)$ and $(\theta_2, [x_2^T, -1]^T)$ exist, with $\theta_1 \neq \theta_2$ and $x_1 \neq x_2$. Then we have

$$\begin{aligned} \lambda_{\min} &= \min_y \frac{y^TB(\theta_1)y}{y^Ty} = \frac{y_1^TB(\theta_1)y_1}{y_1^Ty_1} = \frac{y_1^TMy_1}{y_1^Ty_1} + \theta_1\frac{y_1^TNy_1}{y_1^Ty_1} \\ &= \frac{y_1^TMy_1}{y_1^Ty_1} = \frac{y_1^TMy_1}{y_1^Ty_1} + \theta_2\frac{y_1^TNy_1}{y_1^Ty_1} = \frac{y_1^TB(\theta_2)y_1}{y_1^Ty_1} = \min_y \frac{y^TB(\theta_2)y}{y^Ty}. \end{aligned}$$

Therefore, y_1 is an eigenvector corresponding to the smallest eigenvalue $\lambda_{\min}$ of both, $B(\theta_1)$ and $B(\theta_2)$, which yields (according to Case 1) the contradiction $\theta_1 = \theta_2$. ☐

Theorem 2.5 demonstrates that if $\hat{\theta}$ is a positive root of g, then $x := -y(1:n)/y(n+1)$ solves the RTLS problem (2.2) where y denotes an eigenvector of $B(\hat{\theta})$ corresponding to its smallest eigenvalue.

However, g is not necessarily continuous. If the multiplicity of the smallest eigenvalue of $B(\theta)$ is greater than 1 for some θ_0, then g may have a jump discontinuity at θ_0, and this may actually occur; cf. Example 2.3 where g is discontinuous for $\theta_0 = 0.5$ and $\theta_0 = 1$. Hence, the question arises whether g may jump from a positive value to a negative one, such that it has no positive root.

The following Theorem demonstrates that this is not possible for the standard case $L = I$.

THEOREM 2.6. *Consider the standard case $L = I$, where $\sigma_{\min}([A, b]) < \sigma_{\min}(A)$ and $\delta^2 < \|x_{TLS}\|^2$.*

Assume that the smallest eigenvalue of $B(\theta_0)$ is a multiple one for some θ_0 . Then it holds that

$$0 \notin [\lim_{\theta\to\theta_0-} g(\theta), g(\theta_0)].$$

Proof. Let $\mathcal{V}_{\min}$ be the space of right singular vectors of A corresponding to $\sigma_{\min}(A)$, and $v_1, \ldots, v_r$ be an orthonormal basis of $\mathcal{V}_{\min}$.

Since $\lambda_{\min}(B(\theta_0))$ is a multiple eigenvalue of

$$B(\theta_0) = \begin{bmatrix} A^TA + \theta_0 I & A^Tb \\ b^TA & b^Tb - \theta_0\delta^2 \end{bmatrix},$$

it follows from the interlacing property that it is also the smallest eigenvalue of $A^TA + \theta_0 I$. Hence, $\lambda_{\min}(B(\theta_0)) = \sigma_{\min}(A)^2 + \theta_0$, and for every $v \in \mathcal{V}_{\min}$ we obtain $\begin{bmatrix} v \\ 0 \end{bmatrix} \in \mathcal{E}(\theta_0)$ with $A^Tb \perp \mathcal{V}_{\min}$.

If the last component of an element of $\mathcal{E}(\theta_0)$ does not vanish, then it can be scaled to $\begin{bmatrix} x \\ -1 \end{bmatrix}$. Hence,

$$\begin{bmatrix} A^TA + \theta_0 I & A^Tb \\ b^TA & b^Tb - \theta_0\delta^2 \end{bmatrix}\begin{bmatrix} x \\ -1 \end{bmatrix} = (\sigma_{\min}(A)^2 + \theta_0)\begin{bmatrix} x \\ -1 \end{bmatrix}$$

ETNA
Kent State University
etna@mcs.kent.edu

from which we obtain $(A^TA - \sigma_{\min}(A)^2 I)x = A^Tb$. Therefore, it holds that $x = x_{\min} + z$ where $x_{\min} = (A^TA - \sigma_{\min}(A)^2 I)^\dagger A^Tb$ denotes the pseudonormal solution and $z \in \mathcal{V}_{\min}$. Hence,

$$V := \left[\begin{bmatrix} v_1 \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} v_r \\ 0 \end{bmatrix}, \begin{bmatrix} x_{\min} \\ -1 \end{bmatrix}\right]$$

is a basis of $\mathcal{E}(\theta_0)$ with orthogonal columns and

$$g(\theta_0) = \min_{y \in \mathcal{E}(\theta)} \frac{y^TNy}{y^Ty} = \min_{w \in \mathbb{R}^{r+1}} \frac{w^TV^TNVw}{w^TV^TVw} = \min_{w \in \mathbb{R}^{r+1}} \frac{w^T \begin{bmatrix} I_r & 0 \\ 0 & \|x_{\min}\|^2 - \delta^2 \end{bmatrix} w}{w^T \begin{bmatrix} I_r & 0 \\ 0 & \|x_{\min}\|^2 + 1 \end{bmatrix} w}$$

demonstrating that

$$g(\theta_0) = \min\left\{1, \frac{\|x_{\min}\|^2 - \delta^2}{\|x_{\min}\|^2 + 1}\right\} = \frac{\|x_{\min}\|^2 - \delta^2}{\|x_{\min}\|^2 + 1} \geq \frac{\|x_{TLS}\|^2 - \delta^2}{\|x_{\min}\|^2 + 1}.$$

The active constraint assumption $\|x_{RTLS}\|^2 = \delta^2 < \|x_{TLS}\|^2$ finally yields that g stays positive at θ_0. □

For general regularization matrices L it may happen that g does not have a root, but it jumps below zero at some θ_0.

REMARK 2.7. A jump discontinuity of $g(\theta)$ can only appear at a value θ_0 if $\lambda_{\min}(B(\theta_0))$ is a multiple eigenvalue of $B(\theta_0)$. By the interlacing theorem $\lambda_{\min}$ is also the smallest eigenvalue of $A^TA + \theta_0 L^TL$. Hence there exists an eigenvector v of $A^TA + \theta_0 L^TL$ corresponding to the smallest eigenvalue $\lambda_{\min}$, such that $\bar{v} = (v^T, 0)^T \in \mathcal{E}(\theta_0)$ is an eigenvector of $B(\theta_0)$. The Rayleigh quotient $R_N(\bar{v}) = (\bar{v}^TN\bar{v})/(\bar{v}^T\bar{v})$ of N at $\bar{v}$ is positive, i.e., $R_N(\bar{v}) = \|Lv\|^2$.

If $g(\theta_0) < 0$, there exists some $w \in \mathcal{E}(\theta_0)$ with $R_N(w) = g(\theta_0) < 0$ and w has a non vanishing last component. In this case of a jump discontinuity below zero it is still possible to construct a RTLS solution: A linear combination of $\bar{v}$ and w has to be chosen such that $R_N(\alpha\bar{v} + \beta w) = 0$. Scaling the last component of $\alpha\bar{v} + \beta w$ to -1 yields a solution of the RTLS problem (2.2), which is not unique in this case.

EXAMPLE 2.8. Let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \; b = \begin{bmatrix} 1 \\ 0 \\ \sqrt{5} \end{bmatrix}, \; L = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 1 \end{bmatrix}, \; \delta = \sqrt{3}. \tag{2.11}$$

Then, $1 = \sigma_{\min}(A) > \tilde{\sigma}_{\min}([A, b]) \approx 0.8986$ holds, and the corresponding TLS problem has the solution

$$x_{TLS} = (A^TA - \tilde{\sigma}^2_{\min} I)^{-1} A^Tb \approx \begin{bmatrix} 5.1926 \\ 0 \end{bmatrix}.$$

Since $\mathcal{N}(L) = \{0\}$, the RTLS problem (2.11) is solvable and the constraint at the solution is active, because $\delta^2 = 3 < 53.9258 \approx \|Lx_{TLS}\|_2^2$ holds.

Figure 2.1 shows the corresponding function $g(\theta)$ with two jumps, one at $\theta = 0.25$, and another one at $\theta = 1$, which falls below zero.

The matrix $B(\theta)$ for $\theta_0 = 1$

$$B(1) = [A, b]^T[A, b] + 1 \begin{bmatrix} L^TL & 0 \\ 0 & -\delta^2 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix}$$

ETNA
Kent State University
etna@mcs.kent.edu

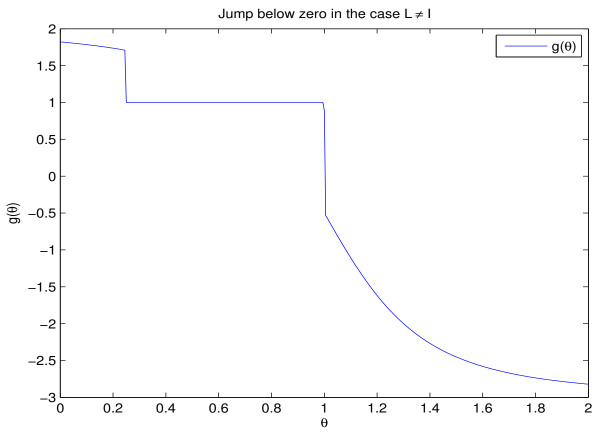

FIGURE 2.1. *Jump below zero of $g(\theta)$ in the case $L \neq I$*

has the double smallest eigenvalue $\lambda_{\min} = 2$ and corresponding eigenvectors $v = (0,1,0)^T$ and $w = (1,0,-1)^T$. The eigenvectors with $R_N(x) = 0$ and last component -1 are $x_1 = v+w = (1,1,-1)^T$ and $x_2 = -v+w = (1,-1,-1)^T$ yielding the two solutions $x_{RTLS,1} = (1,1)^T$ and $x_{RTLS,2} = (1,-1)^T$ of the RTLS problem (2.11).

REMARK 2.9. Consider a jump discontinuity at $\hat{\theta}$ below zero and let the smallest eigenvalue of $(A^TA + \hat{\theta}L^TL)$ have a multiplicity greater than one (in Example 2.8 it is equal to one). Then there exist infinitely many solutions of the RTLS problem (2.2), all satisfying $R_N([x_{RTLS}^T, -1]^T) = 0$.

3. Numerical method. Assuming that g is continuous and strictly monotonically decreasing, Renaut and Guo [15] derived the following update

$$\theta_{k+1} = \theta_k + \frac{\theta_k}{\delta^2} g(\theta_k)$$

for solving $g(\theta) = 0$, where at step k $(x_{\theta_k}^T, -1)^T$ is the eigenvector of $B(\theta_k)$ corresponding to $\lambda_{\min}(B(\theta_k))$, and g is defined in (2.6). Additionally, backtracking was introduced to make the method converge, i.e., the update was modified to

$$\theta_{k+1} = \theta_k + \iota \frac{\theta_k}{\delta^2} g(\theta_k) \tag{3.1}$$

where $\iota \in (0,1]$ was reduced until the sign condition $g(\theta_k)g(\theta_{k+1}) \geq 0$ was satisfied. Thus, Renaut and Guo considered the method described in Algorithm 1.

Although in general the assumptions in [15] (continuity and strict monotonicity of g as defined in (2.6)) are not satisfied, the algorithm may be applied to the modified function g in (2.7) since generically the smallest eigenvalue of $B(\theta)$ is simple and solutions of the RTLS problem correspond to the root of g.

However, the method as suggested by Renaut and Guo suffers two drawbacks: The suggested eigensolver in line 7 of Algorithm 1 for finding the smallest eigenpair of $B(\theta_{k+1})$ is the Rayleigh quotient iteration (or inverse iteration in an inexact version, where the eigensolver is terminated as soon as an approximation to $(x_{k+1}^T, -1)^T$ is found satisfying the sign condition). Due to the required LU factorizations at each step this method is very costly. An approach of this kind does not take account of the fact that the matrices $B(\theta_k)$ converge as

ETNA
Kent State University
etna@mcs.kent.edu

Algorithm 1 RTLSEVP [Renaut and Guo [15]]

Require: Initial guess $\theta_0 > 0$

1: compute the smallest eigenvalue $\lambda_{\min}(B(\theta_0))$, and the corresponding eigenvector $(x_0^T, -1)^T$
2: compute $g(\theta_0)$, and set $k = 1$
3: **while** not converged **do**
4:     $\iota = 1$
5:     **while** $g(\theta_k)g(\theta_{k+1}) < 0$ **do**
6:         update θ_{k+1} by (3.1)
7:         compute the smallest eigenvalue $\lambda_{\min}(B(\theta_{k+1}))$, and the corresponding eigenvector $(x_{k+1}^T, -1)^T$
8:         **if** $g(\theta_k)g(\theta_{k+1}) < 0$ **then**
9:             $\iota = \iota/2$
10:         **end if**
11:     **end while**
12:     $k = k + 1$
13: **end while**
14: $x_{RTLS} = x_{k+1}$

θ_k approaches the root $\hat{\theta}$ of g. We suggest a method which takes advantage of information acquired in previous iteration steps by thick starts. Secondly, the safeguarding by backtracking hampers the convergence of the method considerably. We propose to replace it by an enclosing algorithm that generates a sequence of shrinking intervals, all of which contain the root. The algorithm further utilizes the asymptotic behavior of g.

3.1. Nonlinear Arnoldi. A method which is able to make use of information from previous iteration steps when solving a convergent sequence of eigenvalue problems is the Nonlinear Arnoldi method, which was introduced in [19] for solving nonlinear eigenvalue problems.

As an iterative projection method it computes an approximation to an eigenpair from a projection to a subspace of small dimension, and it expands the subspace if the approximation does not meet a given accuracy requirement. These projections can be easily reused when changing the parameter θ_k A similar technique has been successfully applied in [13, 14] for accelerating the RTLS solver in [17] which is based on a sequence of quadratic eigenvalue problems.

Let $T_k(\mu) = M + \theta_k N - \mu I$, then Algorithm 2 is used in step 7 of Algorithm 1. The Nonlinear Arnoldi method allows thick starts in line 1, i.e., solving $T_k(\lambda)u = 0$ in step k of RTLSEVP we start Algorithm 2 with the orthonormal basis V that was used in the preceding iteration step when determining the solution $u_{k-1} = Vz$ of $V^T T_{k-1}(\lambda)Vz = 0$.

The projected problem

$$V^T T_k(\mu)Vz = (([A, b]V)^T([A, b]V)z + \theta_k V^T NV - \mu I)z = 0 \tag{3.2}$$

can be determined efficiently, if the matrices V, $[A, b]V$ and $LV(1:n, :)$ are known. These are obtained on-the-fly appending one column to the current matrix, at every iteration step of the Nonlinear Arnoldi method. Notice that the explicit form of the matrices M and N are not needed to execute these multiplications. Moreover, we can take advantage for the following updates on θ by computing the product $([A, b]V)^T([A, b]V)$ in advance.

For the preconditioner, it is appropriate to choose $PC \approx N^{-1}$, that stays constant throughout the whole algorithm. This can be computed cheaply, since L and N are typi-

ETNA
Kent State University
etna@mcs.kent.edu

Algorithm 2 Nonlinear Arnoldi

Start with initial basis V, $V^TV = I$
For fixed θ_k find smallest eigenvalue μ of $V^TT_k(\mu)Vz = 0$ and corresponding eigenvector z
Determine a preconditioner $PC \approx T_k^{-1}(\mu)$
set $u = Vz$, $r = T_k(\mu)u$
while $\|r\|/\|u\| > \epsilon$ **do**
  $v = PCr$
  $v = v - VV^Tv$
  $\tilde{v} = v/\|v\|$, $V = [V, \tilde{v}]$
  Find smallest eigenvalue μ of $V^TT_k(\mu)Vz = 0$ and corresponding eigenvector z
  Set $u = Vz$, $r = T_k(\mu)u$
end while

cally banded matrices. Otherwise a coarse approximation is also good enough; in most cases even the identity matrix (that means no preconditioner) is also sufficient.

3.2. Root-Finding algorithm. Figure 3.1 shows the typical behavior of the graph of a function g close to its root $\hat{\theta}$. On the left of $\hat{\theta}$ its slope is often very steep, while on the right of $\hat{\theta}$ it approaches its limit $-\delta^2$ quite quickly. This makes it difficult to determine $\hat{\theta}$ by Newton's method, and this made the backtracking in [15] necessary to enforce convergence.

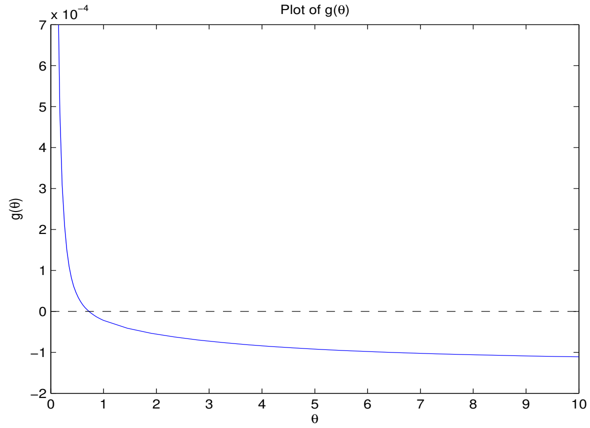

FIGURE 3.1. *Plot of a typical function* $g(\theta)$

Instead, we apply an enclosing algorithm that incorporates the asymptote of g, since it turned out that this has still dominant influence on the behavior of g close to its root. Given three pairs $(\theta_j, g(\theta_j))$, $j = 1, 2, 3$, with

$$\theta_1 < \theta_2 < \theta_3 \quad \text{and} \quad g(\theta_1) > 0 > g(\theta_3) \tag{3.3}$$

we determine the rational interpolation

$$h(\gamma) = \frac{p(\gamma)}{\gamma + \delta^2}$$

ETNA
Kent State University
etna@mcs.kent.edu

where p is a polynomial of degree 2 and p is chosen such that $h(g(\theta_j)) = \theta_j$, $j = 1, 2, 3$. If g is strictly monotonically decreasing in $[\theta_1, \theta_3]$ then this is a rational interpolation of $g^{-1} : [g(\theta_3), g(\theta_1)] \to \mathbb{R}$. As our next iterate we choose $\theta_4 = h(0)$. In exact arithmetic $\theta_4 \in (\theta_1, \theta_3)$, and we replace θ_1 or θ_3 by θ_4 such that the new triple satisfies (3.3).

It may happen, due to nonexistence of the inverse g^{-1} on $[g(\theta_3), g(\theta_1)]$ or due to rounding errors very close to the root $\hat{\theta}$, that θ_4 is not contained in the interval (θ_1, θ_3). In this case we perform a bisection step such that the interval is guaranteed to still contain the root. If two positive values $g(\theta_i)$ are present, then set $\theta_1 = (\theta_2 + \theta_3)/2$ otherwise, in the case of two negative values $g(\theta_i)$ set $\theta_3 = (\theta_1 + \theta_2)/2$.

If a discontinuity is encountered at the root, or close to it, then a very small $\epsilon = \theta_3 - \theta_1$ appears with relatively large $g(\theta_1) - g(\theta_3)$. In this case we terminate the iteration and determine the solution as described in Example 2.8.

4. Numerical Example. To evaluate the performance of the RTLSEVP method for large dimensions we use test examples from Hansen's *Regularization Tools* [10]. The eigenproblems are solved by the Nonlinear Arnoldi method according to Section 3.1 and the root-finding algorithm from the Section 3.2 is applied.

Two functions *phillips* and *deriv2*, which are both discretizations of Fredholm integral equations of the first kind, are used to generate matrices $A_{\text{true}} \in \mathbb{R}^{n\times n}$, right hand sides $b_{\text{true}} \in \mathbb{R}^n$ and solutions $x_{\text{true}} \in \mathbb{R}^n$ such that

$$A_{\text{true}}\, x_{\text{true}} = b_{\text{true}}.$$

In all cases the matrices A_{true} and $[A_{\text{true}}, b_{\text{true}}]$ are ill-conditioned.

To construct a suitable TLS problem, the norm of b_{true} is scaled such that $\|b_{\text{true}}\|_2 = \max_i \|A_{\text{true}}(:,i)\|_2$ holds. The vector x_{true} is scaled by the same factor. The noise added to the problem is put in relation to the maximal element of the augmented matrix

$$\text{maxval} = \max\left(\max\left(\text{abs}[A_{\text{true}}, b_{\text{true}}]\right)\right).$$

We add white noise of level 1-10% to the data, setting $\sigma = \text{maxval} \cdot (0.01, \ldots, 0.1)$, and obtain the systems $Ax \approx b$ to be solved where $A = A_{\text{true}} + \sigma E$ and $b = b_{\text{true}} + \sigma e$, and the elements of E and e are independent random variables with zero mean and unit variance. The matrix $L \in \mathbb{R}^{(n-1)\times n}$ approximates the first order derivative, and δ is chosen to be $\delta = 0.9\|Lx_{\text{true}}\|_2$.

TABLE 4.1
Example phillips, aver. CPU time in sec.

noise	1%			10%		
n	1000	2000	4000	1000	2000	4000
CPU time	0.06	0.15	0.57	0.05	0.14	0.54
MatVecs	19.8	19.0	20.0	18.8	18.2	18.9

TABLE 4.2
Example deriv2, aver. CPU time in sec.

noise	1%			10%		
n	1000	2000	4000	1000	2000	4000
CPU time	0.07	0.20	0.69	0.07	0.19	0.68
MatVecs	24.9	24.6	24.1	23.6	23.4	23.6

The numerical test were run on a 3.4 GHz PentiumR4 computer with 8GB RAM under MATLAB R2007a. Tables 4.1 and 4.2 contain the CPU times in seconds and the number of matrix-vector products averaged over 100 random simulations for dimensions $n = 1000$, $n = 2000$, and $n = 4000$ with noise levels 1% and 10% for *phillips* and *deriv2*, respectively. The outer iteration was terminated if the residual norm of the first order condition was less than 10^{-8}. The preconditioner was calculated with UMFPACK [3], i.e., MATLAB's $[L, U, P, Q] = lu(N)$, with a slightly perturbed N to make it regular.

It turned out that a suitable start basis V for the Nonlinear Arnoldi is an orthonormal basis of the Krylov space $\mathcal{K}(M, e_{n+1})$ of M with initial vector e_{n+1} of small dimension complemented by the vector $e := \text{ones}(n+1, 1)$ of all ones. In the starting phase we determine three values θ_i such that not all $g(\theta_i)$ have the same sign. Multiplying the θ_i either by 0.01 or 100 depending on the sign of $g(\theta_i)$ leads after very few steps to an interval that contains the root of $g(\theta)$.

Figure 4.1 shows the convergence history of the RTLSEVP algorithm. The problem is *phillips* from [10], with a dimension of $n = 2000$ and a noise level of 1%.

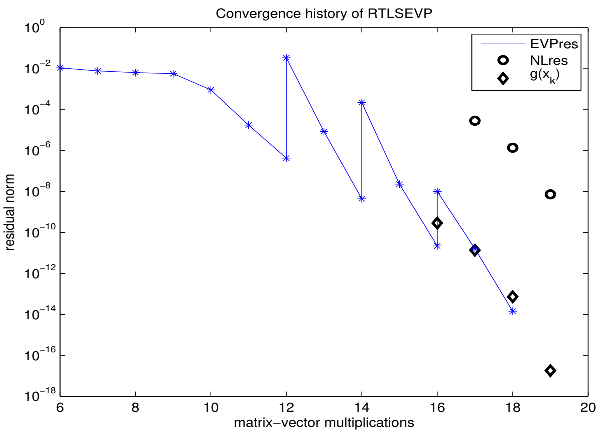

FIGURE 4.1. *Convergence history of RTLSEVP*

This convergence behavior is typical for the RTLSEVP method where the eigenvalue problems in the inner iteration are solved by the Nonlinear Arnoldi method. An asterisk marks the residual norm of an eigenvalue problem in an inner iteration, a circle denotes the residual norm of the first order condition in an outer iteration and a diamond is the value of the function $g(\theta_k)$. The cost of one inner iteration is approximately one matrix-vector product (MatVec), whereas an outer iteration is much cheaper. It only consists of evaluating the function $g(\theta_k)$, solving a 3x3 system of equations for the new θ_{k+1} and evaluating the first order condition. The outer iteration together with the evaluation of the first residual of the next EVP can be efficiently performed by much less than one MatVec.

The size of the starting subspace for the Nonlinear Arnoldi is equal to six, which corresponds to the six MatVecs at the first EVP residual. After 16 MatVecs the three starting pairs $(\theta_i, g(\theta_i))$ are found. This subspace already contains such good information about the solution that only two more MatVecs are needed to obtain the RTLS solution. The main costs of Algorithm 1 with the Nonlinear Arnoldi method used in step 7 and the proposed root-finding algorithm in step 6 are only matrix vector multiplications. The number of MatVecs is much less than the dimension of the problem, hence the computational complexity is of order $\mathcal{O}(n^2)$ with n being the smaller matrix dimension of A.

5. Conclusions. Regularized total least squares problems can be solved efficiently by the RTLSEVP method introduced by Renaut and Guo [15] via a sequence of linear eigenproblems. Since in general no fixed point behavior to a global minimizer can be shown, the function $g(\theta)$ is introduced. A detailed analysis of this function and a suitable root-finding algorithm are presented. For problems of large dimension the eigenproblems can be solved efficiently by the Nonlinear Arnoldi method.

Acknowledgement. The work of Jörg Lampe was supported by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung–BMBF) under grant number 13N9079.

REFERENCES

[1] A. BECK AND A. BEN-TAL, *On the solution of the Tikhonov regularization of the total least squares problem*, SIAM J. Optim., 17 (2006), pp. 98–118.
[2] A. BECK, A. BEN-TAL, AND M. TEBOULLE, *Finding a global optimal solution for a quadratically constrained fractional quadratic problem with applications to the regularized total least squares problem*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 425–445.
[3] T.A. DAVIS AND I.S. DUFF, *Algorithm 832: UMFPACK, an unsymmetric-pattern multifrontal method*, ACM Trans. Math. Softw., 30 (2004), pp. 196–199.
[4] H. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer, Dodrecht, The Netherlands, 1996.
[5] W. GANDER, G.H. GOLUB, AND U. VON MATT, *A constrained eigenvalue problem*, Lin. Alg. Appl., 114–115 (1989), pp. 815–839.
[6] G.H. GOLUB, P.C. HANSEN, AND D.P. O'LEARY, *Tikhonov regularization and total least squares*, SIAM J. Matrix Anal. Appl., 21 (1999), pp. 185–194.
[7] G.H. GOLUB AND C.F. VAN LOAN, *Matrix Computations*, The John Hopkins University Press, Baltimore and London, 3rd ed., 1996.
[8] C. GROETSCH, *Inverse Problems in the Mathematical Sciences*, Vieweg, Wiesbaden, Germany, 1993.
[9] H. GUO AND R. RENAUT, *A regularized total least squares algorithm*, in Total Least Squares and Errors-in-Variable Modelling, S. Van Huffel and P. Lemmerling, eds., Kluwer, Dodrecht, The Netherlands, pp. 57–66, 2002.
[10] P.C. HANSEN, *Regularization tools, a Matlab package for analysis of discrete regularization problems*, Numer. Alg., 6 (1994), pp. 1–35.
[11] ———, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, SIAM, Philadelphia, 1998.
[12] ———, *Regularization tools version 4.0 for Matlab 7.3*, Numer. Alg., 46 (2007), pp. 189–194.
[13] J. LAMPE AND H. VOSS, *On a quadratic eigenproblem occuring in regularized total least squares*, Comput. Stat. Data Anal., 52(2) (2007), pp. 1090–1102.
[14] ———, *Global convergence of RTLSQEP: a solver of regularized total least squares problems via quadratic eigenproblems*, Math. Modelling Anal., 13 (2008), pp. 55–66.
[15] R. RENAUT AND H. GUO, *Efficient algorithms for solution of regularized total least squares*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 457–476.
[16] D. SIMA, *Regularization Techniques in Model Fitting and Parameter Estimation*, PhD thesis, Katholieke Universiteit Leuven, Leuven, Belgium, 2006.
[17] D. SIMA, S. VAN HUFFEL, AND G.H. GOLUB, *Regularized total least squares based on quadratic eigenvalue problem solvers*, BIT, 44 (2004), pp. 793–812.
[18] S. VAN HUFFEL AND J. VANDEVALLE, *The Total Least Squares Problems: Computational Aspects and Analysis*, vol. 9 of Frontiers in Applied Mathematics, SIAM, Philadelphia, 1991.
[19] H. VOSS, *An Arnoldi method for nonlinear eigenvalue problems*, BIT, 44 (2004), pp. 387–401.
[20] H. VOSS AND B. WERNER, *A minimax principle for nonlinear eigenvalue problems with applications to nonoverdamped systems*, Math. Meth. Appl. Sci., 4 (1982), pp. 415–424.