BIT manuscript No.
(will be inserted by the editor)

Projected Nonstationary Iterated Tikhonov Regularization

Guangxin Huang · Lothar Reichel · Feng Yin

Received: date / Accepted: date

Dedicated to Heinrich Voss on the occasion of his 70th birthday.

Abstract This paper presents a nonstationary iterated Tikhonov regularization method for the solution of large-scale Tikhonov minimization problems in general form. The method projects the large-scale problem into a sequence of generalized Krylov subspaces of low dimension. The regularization parameter is determined by the discrepancy principle. Numerical examples illustrate the effectiveness of the method.

Keywords discrete ill-posed problem · Tikhonov regularization · generalized Krylov subspace

Mathematics Subject Classification (2010) 65F15, 65F22, 65F30

1 Introduction

We are concerned with the solution of large least-squares problems of the form

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|, \qquad A\in\mathbb{R}^{m\times n}, \quad b\in\mathbb{R}^m. \tag{1.1}$$

The singular values of the matrix A are assumed to gradually decay to zero without a significant gap. In particular, A is severely ill-conditioned and may

Guangxin Huang
Geomathematics Key Laboratory of Sichuan, Chengdu University of Technology, Chengdu, 610059, P. R. China. E-mail: huangx@cdut.edu.cn

Lothar Reichel
Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA.
E-mail: reichel@math.kent.edu

Feng Yin
Geomathematics Key Laboratory of Sichuan, Chengdu University of Technology, Chengdu, 610059, P. R. China; School of Sciences, Sichuan University of Science and Engineering, Zigong, 643000, P. R. China. E-mail: fyin@suse.edu.cn

be singular. Minimization problems (1.1) with a matrix of this kind are commonly referred to as discrete ill-posed problems. They arise, for instance, from the discretization of ill-posed problems such as Fredholm integral equations of the first kind; see, e.g., [5,9]. Throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm or the spectral matrix norm.

The vector b in (1.1) represents measured data that are contaminated by an error $e \in \mathbb{R}^m$, which may stem from measurement inaccuracies or discretization errors. Let $b_{\text{true}} \in \mathbb{R}^m$ denote the unknown error-free vector associated with b. Then b can be written as

$$b = b_{\text{true}} + e. \tag{1.2}$$

We will assume that a fairly sharp bound for the norm of the error in (1.2) is available,

$$\|e\| \leq \delta, \tag{1.3}$$

and that the linear system of equations

$$Ax = b_{\text{true}} \tag{1.4}$$

associated with the least-squares problem (1.1) is consistent. These assumptions allow us to apply the discrepancy principle; see below.

We would like to determine an accurate approximation of the solution of minimal Euclidean norm, x_{true}, of (1.4) by computing a suitable approximate solution of (1.1). Due to the ill-conditioning of A and the error e in b, straightforward solution of (1.1), in general, does not yield a meaningful approximation of x_{true}. Therefore, one often replaces (1.1) by a nearby problem, whose solution is less sensitive to the error in b than the solution of (1.1). This replacement is known as regularization. The possibly most popular regularization method is due to Tikhonov. This method replaces (1.1) by a penalized least-squares problem of the form

$$\min_{x\in\mathbb{R}^n} \left\{ \|Ax - b\|^2 + \mu^{-1}\|Lx\|^2 \right\}, \tag{1.5}$$

where $L \in \mathbb{R}^{p\times n}$ is a referred to as a regularization matrix and the scalar $\mu > 0$ as a regularization parameter. When L is the identity matrix I, the Tikhonov minimization problem (1.5) is said to be in *standard form*; otherwise it is in *general form*. The choice of regularization matrix may significantly affect how close the solution of (1.5) for a suitable $\mu > 0$ is to x_{true}; see Example 4.2 as well as [21] for illustrations. The use of μ^{-1} instead of μ in (1.5) is commented on in Remark 3.1 below.

We note for future reference that the normal equations associated with (1.5) are given by

$$(A^T A + \mu^{-1} L^T L)x = A^T b, \tag{1.6}$$

where the superscript ^{T} denotes transposition. Assume that L is chosen so that

$$\mathcal{N}(A) \cap \mathcal{N}(L) = \{0\}, \tag{1.7}$$

where $\mathcal{N}(M)$ denotes the null space of the matrix M. Then (1.5) has a unique solution

$$x_\mu = (A^T A + \mu^{-1} L^T L)^{-1} A^T b \tag{1.8}$$

for any $\mu > 0$. The value of μ determines how sensitive x_μ is to the error e and how close x_μ is to x_{true}; see, e.g., Engl et al. [5], Groetsch [7], and Hansen [9] for discussions on Tikhonov regularization. There are many approaches to determining a suitable value of μ, including the L-curve criterion, generalized cross validation, and the discrepancy principle; see, e.g., [14,19] for discussions and illustrations. The availability of the bound (1.3) allows us to apply the discrepancy principle, which we will use in the computed examples of Section 4. However, the solution method for (1.5) of this paper also can be combined with other approaches to determine μ.

We propose a new method for computing an approximate solution of large-scale Tikhonov regularization problems (1.5) with a general regularization matrix. It is based on nonstationary iterated Tikhonov regularization. The kth approximate solution, $x_k \in \mathbb{R}^n$, determined by this method solves the minimization problem

$$\min_{x\in\mathbb{R}^n} \{\|Ax - b\|^2 + \mu_k^{-1}\|L(x - x_{k-1})\|^2\}, \qquad k = 1, 2, \dots \ , \tag{1.9}$$

where $x_0 \in \mathbb{R}^n$ is an initial approximate solution. Thus, x_k satisfies

$$(A^T A + \mu_k^{-1} L^T L)x_k = \mu_k^{-1} L^T L x_{k-1} + A^T b, \qquad k = 1, 2, \dots \ , \tag{1.10}$$

which are the normal equations asssociated with (1.9) for $x = x_k$. To the best of our knowledge, nonstationary iterated Tikhonov regularization has been described in the literature only in the special case when $L = I$. For instance, Hanke and Groetsch [8] establish the rate of convergence of the iterates x_k defined by (1.10) for $L = I$ when b is contaminated by error and the number of iterations is determined by the discrepancy principle. Recently, Donatelli and Hanke [2] introduced a preconditioned iterative scheme related to (1.10), in which $L = I$ and A is assumed to have a Toeplitz-type structure and in the left-hand side is replaced by a circulant-type matrix; see also Engl et al. [5] for discussions on nonstationary iterated Tikhonov regularization.

Our interest in a solution method for (1.10) stems from the fact that approximations determined by nonstationary iterated Tikhonov regularization with $L = I$ give more accurate approximations of x_{true} than approximations determined by standard Tikhonov regularization (1.5) with $L = I$ when μ is determined by the discrepancy principle and $\delta \searrow 0$; see, e.g., [5, Section 5.1] or [8]. Numerical examples presented in Section 4 demonstrate that nonstationary iterated Tikhonov regularization (1.10) with $L \neq I$ also yields more accurate approximations of x_{true} than standard Tikhonov regularization (1.5).

We compute approximations of the iterates x_k determined by (1.10) by projecting the iterates and the matrices A and L into low-dimensional generalized Krylov subspaces. These subspaces are determined by an Arnoldi-type algorithm. A similar reduction approach has previously been applied by Lampe

et al. [15] to Tikhonov regularization problems in general form (1.5). A recent application to the solution of dual regularized least-squares problems is described by Lampe and Voss [18]. This type of reduction method was first proposed by Voss [23] for the solution of nonlinear eigenvalue problems.

Several other methods for the solution of (1.5) are available. For instance, small to medium-sized problems can be solved conveniently by first computing the generalized singular value decomposition (GSVD) or a related decomposition of the matrix pair $\{A, L\}$; see [3,9]. When the regularization matrix L is banded and has a known null space, it may be attractive to transform the Tikhonov minimization problem (1.5) to standard form by using the substitution $y = Lx$. The matrix A in (1.5) then is replaced by $AL_A^\dagger$, where

$$L_A^\dagger := (I - (A(I - L^\dagger L))^\dagger A)L^\dagger$$

is the A-weighted generalized inverse of L. Here $L^\dagger$ denotes the Moore–Penrose pseudoinverse of L; see Eldén [4] for details. For general large matrices L, this transformation is too expensive to be attractive. Kilmer et al. [13] proposed an inner-outer iterative method that determines a reduced GSVD of the matrix pair $\{A, L\}$. This approach yields nice results, but can be expensive when many inner iterations are required. Methods that reduce the pair of large matrices $\{A, L\}$ to a pair of small matrices are described in [12,20,22]. The GSVD can then be applied to the reduced problem obtained. A scheme in which the solution subspace is independent of the matrix L is described in [11]. For instance, A may be reduced by partial Golub–Kahan bidiagonalization and L be projected into the solution Krylov subspace

$$\mathcal{K}_\ell(A^TA, A^Tb) = \text{span}\{A^Tb, (A^TA)A^Tb, \ldots, (A^TA)^{\ell-1}A^Tb\} \tag{1.11}$$

so obtained. A careful comparison of these methods is outside the scope of the present paper, but we note that some comments on how the methods compare can be found in [12,22]. The method of this paper is the only one that is based on nonstationary iterated Tikhonov regularization (1.10). A nice recent survey of many Krylov subspace methods for the solution of linear discrete ill-posed problems is provided by Gazzola et al. [6].

This paper is organized as follows. Section 2 describes our projected nonstationary iterated Tikhonov regularization method. Implementation details are discussed in Section 3, and Section 4 presents numerical examples. Concluding remarks can be found in Section 5.

2 The solution method

This section describes an algorithm for carrying out projected nonstationary iterated Tikhonov regularization in generalized Krylov subspaces. Implementation details are provided in the following section.

Introduce the subspace $\mathcal{V}_k \subset \mathbb{R}^n$ of small dimension $k + \ell \ll n$, and let the columns of the matrix $V_k \in \mathbb{R}^{n\times(k+\ell)}$ form an orthonormal basis for this

subspace. We will let ℓ be the (small) dimension of an initial Krylov subspace (1.11), which will be contained in all the solution subspaces $\mathcal{V}_k$, $k = 1, 2, \dots$. The initial approximate solution, x_0, in (1.9) and (1.10) will be determined in this Krylov subspace.

Let $y_k \in \mathbb{R}^{k+\ell}$ and substitute $x_k = V_k y_k$ into (1.10). Multiplying (1.10) by V_k^T from the left then yields

$$\left((AV_k)^T(AV_k) + \mu_k^{-1}(LV_k)^T(LV_k)\right) y_k = \mu_k^{-1}(LV_k)^T LV_{k-1}y_{k-1} + (AV_k)^T b.$$

Due to (1.7), the matrix on the left-hand side is nonsingular. Therefore,

$$y_k = \bar{y}_k + \left((AV_k)^T(AV_k) + \mu_k^{-1}(LV_k)^T(LV_k)\right)^{-1} (AV_k)^T \bar{r}_k, \qquad (2.1)$$

where

$$\bar{y}_k = \left[y_{k-1}^T, 0\right]^T, \qquad \bar{r}_k = b - (AV_k)\bar{y}_k$$

for $k = 1, 2, \dots$, with $y_0 = 0 \in \mathbb{R}^\ell$. The dimension of the solution subspace $\mathcal{V}_k$ is increased by one in each iteration. The expansion of $\mathcal{V}_k$ is carried out similarly as in [15]; see below. We choose the initial subspace $\mathcal{V}_0$ to be a Krylov subspace (1.11) of low dimension ℓ, and let $y_0 = 0$. The first column of the matrices V_k, $k = 0, 1, 2, \dots$, is

$$V_k e_1 = A^T b / \|A^T b\|; \qquad (2.2)$$

cf. (1.11). Here and throughout this paper e_j denotes the jth axis vector.

The error bound (1.3) allows us to determine a suitable value of the regularization parameter μ_k in each iteration step (1.10) by the discrepancy principle, which prescribes that $\mu_k = \mu_k(\delta)$ be chosen as the zero of the function

$$\varphi(\mu, V_k) := \|Ax_k - b\|^2 - \eta^2\delta^2, \qquad (2.3)$$

where $\eta \geq 1$ is a user-specified constant that is independent of δ and $x_k = x_k(\mu)$ is an approximate solution of the kth iterate of (1.10). Hence, the solution of (1.5) entails both the determination of the value μ_k of the regularization parameter and the computation of an associated approximation $x_k = x_k(\mu_k)$ of the solution of (1.5) that satisfies (2.3).

We compute the zero of $\varphi(\mu, V_k)$ with a zero-finder described in [15]. The zero-finder is based on four-point rational inverse interpolation. We remark that the function $\varphi(\mu, V_k)$ can be evaluated by using a low-dimensional projection of A; it is therefore not necessary to evaluate matrix-vector products with A when computing $\varphi(\mu, V_k)$.

Having determined μ_k, we can compute y_k by (2.1). Then the search subspace $\mathcal{V}_k$ is expanded by the gradient of the functional in (1.5) evaluated at $x_k = V_k y_k$. After expansion, a new value, μ_{k+1}, of the regularization parameter is calculated. This process is repeated until a stopping criterion is satisfied; see Algorithm 2.1 for details. The solution subspaces generated generally are not Krylov subspaces; see Section 3.

Algorithm 2.1 Projected nonstationary iterated Tikhonov regularization (PNITR).

1. **Input:** $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $L \in \mathbb{R}^{p\times n}$, $\eta > 1$, *and* δ;
2. **Initial:** *let the columns of* V_0 *form an orthonormal basis for the Krylov subspace* $\mathcal{K}_\ell(A^TA, A^Tb)$ *of low dimension; let* $y_0 = 0 \in \mathbb{R}^\ell$;
3. **for** $k = 1, 2, \ldots$ **until convergence**
4. *Compute* $\bar{y}_k = \left[y_{k-1}^T, 0\right]^T$ *and* $\bar{r}_k = b - (AV_k)\bar{y}_k$
5. *Find the zero* μ_k *of* $\varphi(\mu, V_k)$
6. *Compute* y_k *by* (2.1)
7. *Compute* $\tilde{r}_k = A^TAV_ky_k + \mu_k^{-1}L^TLV_k(y_k - \bar{y}_k) - A^Tb$
8. *Reorthogonalize (optional)* $\hat{r}_k = (I - V_kV_k^T)\tilde{r}_k$
9. *Normalize* $v_{k+1} = \hat{r}_k/\|\hat{r}_k\|$
10. *Enlarge search space* $V_{k+1} = [V_k, v_{k+1}]$
11. **end for**
12. **Output:** *approximate solution* $x_k = V_ky_k$ *of* (1.5) *with* $\mu = \mu_k$.

The computations of our solution method are outlined by Algorithm 2.1. The algorithm adjusts the regularization parameter in each step and builds up the solution subspaces $\mathcal{V}_k = \text{range}(V_k)$. Generally, "convergence" is achieved already for solution subspaces of fairly small dimension. This is illustrated in Section 4. Most of the computational work is carried out in line 7, since determining the zero of $\varphi(\mu, V_k)$ in line 5 and solving the projected problem in line 6 are quite inexpensive; see below.

3 Implementation details

This section describes an implementation of Algorithm 2.1 based on QR factorization of the matrices AV_k and LV_k. Subsection 3.1 considers the computation of the iterates y_k and their derivatives y_k' with respect to the regularization parameter. Some properties of the function $\varphi(\mu, V_k)$ are discussed in Subsection 3.2, and the zero-finder applied to enforce the discrepancy principle is reviewed in Subsection 3.3. Finally, the construction of the generalized Krylov subspaces $\mathcal{V}_k$ is described in Subsection 3.4. Several of the properties shown in this section are analogous to but different from results in [15]. For notational simplicity, we assume throughout this section that $V_k \in \mathbb{R}^{n\times k}$. Property (2.2) is assumed to hold.

3.1 Computation of y_k and y_k'

Introduce the QR factorizations

$$AV_k = Q_{A,k}R_{A,k} \quad \text{with} \quad Q_{A,k} \in \mathbb{R}^{m\times k}, \quad R_{A,k} \in \mathbb{R}^{k\times k}, \tag{3.1}$$

$$LV_k = Q_{L,k}R_{L,k} \quad \text{with} \quad Q_{L,k} \in \mathbb{R}^{p\times k}, \quad R_{L,k} \in \mathbb{R}^{k\times k}, \tag{3.2}$$

where the matrices $Q_{A,k}$ and $Q_{L,k}$ have orthonormal columns and the matrices $R_{A,k}$ and $R_{L,k}$ are upper triangular. We assume that $k \ll \min\{m, n, p\}$. Substituting (3.1) and (3.2) into (2.1) produces

$$(R_{A,k}^T R_{A,k} + \mu_k^{-1} R_{L,k}^T R_{L,k})(y_k - \bar{y}_k) = R_{A,k}^T d_k^{\bar{r}}, \tag{3.3}$$

where $d_k^{\bar{r}} := Q_{A,k}^T \bar{r}_k$. Using (3.3), we compute the vector y_k in (2.1) by solving the reduced least-squares problem

$$\left\| \begin{bmatrix} R_{A,k} \\ \mu_k^{-1/2} R_{L,k} \end{bmatrix} (y_k - \bar{y}_k) - \begin{bmatrix} Q_{A,k}^T \bar{r}_k \\ 0 \end{bmatrix} \right\|^2 = \min! \tag{3.4}$$

It follows from (1.7) that the $2k \times k$ matrix of this minimization problem is nonsingular for $0 < \mu_k < \infty$. The solution $y_k - \bar{y}_k$ yields the desired vector

$$y_k = \bar{y}_k + (R_{A,k}^T R_{A,k} + \mu_k^{-1} R_{L,k}^T R_{L,k})^{-1} R_{A,k}^T d_k^{\bar{r}}. \tag{3.5}$$

To emphasize that this vector is a function of the regularization parameter μ_k, we write $y_k = y_k(\mu_k)$ and obtain for the derivative with respect to μ the expression

$$y_k'(\mu) = \mu^{-2}(R_{A,k}^T R_{A,k} + \mu^{-1} R_{L,k}^T R_{L,k})^{-1} R_{L,k}^T R_{L,k}(y_k(\mu) - \bar{y}_k). \tag{3.6}$$

The derivative $y_k'(\mu_k)$ can be computed by solving the least-squares problem,

$$\left\| \begin{bmatrix} R_{A,k} \\ \mu_k^{-1/2} R_{L,k} \end{bmatrix} y_k'(\mu_k) - \begin{bmatrix} 0 \\ \mu_k^{-3/2} R_{L,k}(y_k(\mu) - \bar{y}_k) \end{bmatrix} \right\| = \min! \tag{3.7}$$

Note that the system matrices in (3.4) and (3.7) are the same. Therefore, the QR factorization of this matrix can be used to solve both (3.4) and (3.7).

We conclude this subsection by providing some asymptotic properties of $y_k(\mu)$ and $y_k'(\mu)$. The limits of $y_k(\mu)$ are given by

$$\begin{aligned} y_k(\infty) &= \lim_{\mu\to\infty} y_k(\mu) = \lim_{\mu\to\infty} (\bar{y}_k + (R_{A,k}^T R_{A,k} + \mu^{-1} R_{L,k}^T R_{L,k})^{-1} R_{A,k}^T d_k^{\bar{r}}) \\ &= \bar{y}_k + (R_{A,k}^T R_{A,k})^\dagger R_{A,k}^T d_k^{\bar{r}} = \bar{y}_k + R_{A,k}^\dagger d_k^{\bar{r}}, \\ y_k(0) &= \lim_{\mu\to 0} y_k(\mu) = \lim_{\mu\to 0} (\bar{y}_k + \mu(\mu R_{A,k}^T R_{A,k} + R_{L,k}^T R_{L,k})^{-1} R_{A,k}^T d_k^{\bar{r}}) = \bar{y}_k \end{aligned}$$

and the limits of $y_k'(\mu)$ are

$$\begin{aligned} y_k'(\infty) &= \lim_{\mu\to\infty} y_k'(\mu) \\ &= \lim_{\mu\to\infty} \mu^{-2}(R_{A,k}^T R_{A,k} + \mu^{-1} R_{L,k}^T R_{L,k})^{-1} R_{L,k}^T R_{L,k}(y_k(\mu) - \bar{y}_k) = 0, \\ y_k'(0) &= \lim_{\mu\to 0} y'(\mu) \\ &= \lim_{\mu\to 0} (\mu R_{A,k}^T R_{A,k} + R_{L,k}^T R_{L,k})^{-1} R_{L,k}^T R_{L,k} \\ &\quad \times (\mu R_{A,k}^T R_{A,k} + R_{L,k}^T R_{L,k})^{-1} R_{A,k}^T d_k^{\bar{r}} \\ &= (R_{L,k}^T R_{L,k})^\dagger R_{L,k}^T R_{L,k} (R_{L,k}^T R_{L,k})^\dagger R_{A,k}^T d_k^{\bar{r}} = (R_{L,k}^T R_{L,k})^\dagger R_{A,k}^T d_k^{\bar{r}}. \end{aligned}$$

3.2 Properties of $\varphi(\mu, V_k)$

Let $\varepsilon := \eta\delta$. Substituting the factorizations (3.1) and (3.2) into $\varphi(\mu_k, V_k)$, we obtain

$$\begin{aligned}\varphi(\mu_k, V_k) &= \|AV_k y_k - b\|^2 - \varepsilon^2 \\ &= b^T b + (R_{A,k} y_k)^T (R_{A,k} y_k) - 2(R_{A,k} y_k)^T d_k^b - \varepsilon^2,\end{aligned}$$

where

$$d_k^b := Q_{A,k}^T b. \tag{3.8}$$

Moreover, the relation $R_{A,k}^T d_k^b = V_k^T A^T b = \|A^T b\| e_1$ gives

$$\varphi(\mu_k, V_k) = b^T b + (R_{A,k} y_k)^T (R_{A,k} y_k) - 2\|A^T b\| e_1^T y_k - \varepsilon^2.$$

Thus, we can evaluate $\varphi(\mu_k, V_k)$ by solving a small least-squares problem with a $2k \times k$ matrix consisting of two stacked triangular matrices; cf. (3.4). The following theorem yields some properties of the function $\varphi(\mu, V_k)$ under a mild condition on the search space $\mathcal{V}_k$.

Theorem 3.1 *Define the GSVD of the matrix pair $\{R_{A,k}, R_{L,k}\}$, i.e.,*

$$R_{A,k} = \hat{U}_k \Sigma_k Y_k^T, \tag{3.9}$$

$$R_{L,k} = \hat{V}_k \Omega_k Y_k^T, \tag{3.10}$$

where the matrices $\hat{U}_k, \hat{V}_k \in \mathbb{R}^{k\times k}$ are orthogonal, and the nontrivial entries of the matrices $\Sigma_k = \mathrm{diag}[\sigma_1, \sigma_2, \ldots, \sigma_k] \in \mathbb{R}^{k\times k}$ and $\Omega_k = \mathrm{diag}[\omega_1, \omega_2, \ldots, \omega_k] \in \mathbb{R}^{k\times k}$ are ordered according to $0 \le \sigma_1 \le \ldots \le \sigma_k \le 1$ and $1 \ge \omega_1 \ge \ldots \ge \omega_k \ge 0$ with $\sigma_j^2 + \omega_j^2 = 1$ for $1 \le j \le k$. If

$$(Q_{A,k}\hat{U}_k)^T \bar{r}_k \neq 0, \tag{3.11}$$

then the function $\varphi(\mu, V_k)$ is monotonically decreasing and convex for $\mu > 0$. Moreover, the equation

$$\varphi(\mu, V_k) = 0 \tag{3.12}$$

has a unique solution $0 < \mu < \infty$ for any ε with

$$\|\mathrm{P}_{\mathcal{N}(R_{A,k}^T)}(Q_{A,k}^T \bar{r}_k)\|^2 < \varepsilon^2 < \|\bar{r}_k\|^2, \tag{3.13}$$

where $\mathrm{P}_{\mathcal{N}(R_{A,k}^T)}$ denotes the orthogonal projector onto the null space of $R_{A,k}^T$.

Proof. We first show monotonicity and convexity. Substituting the factorizations (3.1) and (3.2), as well as (2.1), (3.9), and (3.10) into the right-hand side of (2.3), we obtain

$$\begin{aligned}&\varphi(\mu, V_k) + \varepsilon^2 = \|AV_k(y_k - \bar{y}_k) - (b - AV_k\bar{y}_k)\|^2 \\ &= \|Q_{A,k}R_{A,k}(R_{A,k}^T R_{A,k} + \mu^{-1}R_{L,k}^T R_{L,k})^{-1}(Q_{A,k}R_{A,k})^T \bar{r}_k - \bar{r}_k\|^2 \\ &= \|(R_{A,k}(R_{A,k}^T R_{A,k} + \mu^{-1}R_{L,k}^T R_{L,k})^{-1}R_{A,k}^T - I)Q_{A,k}^T \bar{r}_k\|^2 \\ &= ((Q_{A,k}\hat{U}_k)^T \bar{r}_k)^T \Omega_k^4 (\mu\Sigma_k^2 + \Omega_k^2)^{-2} (Q_{A,k}\hat{U}_k)^T \bar{r}_k \\ &= \sum_{i=1}^{k} \frac{\omega_i^4}{(\mu\sigma_i^2 + \omega_i^2)^2} \left|(Q_{A,k}\hat{U}_k)_i^T \bar{r}_k\right|^2.\end{aligned}$$

It follows that the derivative of φ with respect to μ is given by

$$\varphi'(\mu, V_k) = -\sum_{i=1}^{k} \frac{2\sigma_i^2 \omega_i^4}{(\mu\sigma_i^2 + \omega_i^2)^3} \left| (Q_{A,k}\hat{U}_k)_i^T \bar{r}_k \right|^2, \tag{3.14}$$

and the second derivative is

$$\varphi''(\mu, V_k) = \sum_{i=1}^{k} \frac{6\sigma_i^4 \omega_i^4}{(\mu\sigma_i^2 + \omega_i^2)^4} \left| (Q_{A,k}\hat{U}_k)_i^T \bar{r}_k \right|^2. \tag{3.15}$$

Due to (3.11), there is an index i such that $(Q_{A,k}\hat{U}_k)_i^T \bar{r}_k \neq 0$. This implies that $\varphi'(\mu, V_k) < 0$ and $\varphi''(\mu, V_k) > 0$. Thus, $\varphi(\mu, V_k)$ is monotonically decreasing and convex for $\mu > 0$ on the subspace $\mathcal{V}_k$.

We turn to the second part of the theorem and observe that

$$\begin{aligned} &\lim_{\mu\to 0} \varphi(\mu, V_k) + \varepsilon^2 = \lim_{\mu\to 0} \|AV_k y_k - b\|^2 \\ &= \lim_{\mu\to 0} \|(R_{A,k}(R_{A,k}^T R_{A,k} + \mu^{-1} R_{L,k}^T R_{L,k})^{-1} R_{A,k}^T - I) Q_{A,k}^T \bar{r}_k\|^2 \\ &= \lim_{\mu\to 0} \|(\mu R_{A,k}(\mu R_{A,k}^T R_{A,k} + R_{L,k}^T R_{L,k})^{-1} R_{A,k}^T - I) Q_{A,k}^T \bar{r}_k\|^2 \\ &= \|Q_{A,k}^T \bar{r}_k\|^2 = \|\bar{r}_k\|^2 \end{aligned} \tag{3.16}$$

and

$$\begin{aligned} &\lim_{\mu\to\infty} \varphi(\mu, V_k) + \varepsilon^2 \\ &= \lim_{\mu\to\infty} \|(R_{A,k}(R_{A,k}^T R_{A,k} + \mu^{-1} R_{L,k}^T R_{L,k})^{-1} R_{A,k}^T - I) Q_{A,k}^T \bar{r}_k\|^2 \\ &= \lim_{\mu\to\infty} (\|Q_{A,k}^T \bar{r}_k\|^2 + \|R_{A,k}(R_{A,k}^T R_{A,k} + \mu^{-1} R_{L,k}^T R_{L,k})^{-1} R_{A,k}^T Q_{A,k}^T \bar{r}_k\|^2 \\ &- 2(Q_{A,k}^T \bar{r}_k)^T R_{A,k}(R_{A,k}^T R_{A,k} + \mu^{-1} R_{L,k}^T R_{L,k})^{-1} R_{A,k}^T Q_{A,k}^T \bar{r}_k) \\ &= \|Q_{A,k}^T \bar{r}_k\|^2 - (Q_{A,k}^T \bar{r}_k)^T R_{A,k} R_{A,k}^\dagger Q_{A,k}^T \bar{r}_k \\ &= (Q_{A,k}^T \bar{r}_k)^T (I - R_{A,k} R_{A,k}^\dagger) Q_{A,k}^T \bar{r}_k. \end{aligned} \tag{3.17}$$

Since $\varphi(\mu, V_k)$ is monotonically decreasing and convex for $\mu > 0$ on the subspace $\mathcal{V}_k$, the above limits imply that equation (3.12) has a unique solution when (3.13) holds. This completes the proof. □

Remark 3.1 The convexity is a result of using $1/\mu$ in (1.5) instead of μ.

Remark 3.2 The conditions (3.13) determine whether the function (2.3) has a zero. These conditions will be used when designing a zero-finder; see Subsection 3.3.

In the remainder of this subsection, we discuss the asymptotic behavior of the functions φ, φ', and φ''. It follows from (3.16) and (3.17) that

$$\lim_{\mu\to 0}\varphi(\mu,V_k) = \|\bar{r}_k\|^2-\varepsilon^2, \tag{3.18}$$

$$\begin{aligned}\lim_{\mu\to\infty}\varphi(\mu,V_k) &= (Q_{A,k}^T\bar{r}_k)^T(I-R_{A,k}R_{A,k}^\dagger)Q_{A,k}^T\bar{r}_k-\varepsilon^2\\ &= \|\bar{r}_k\|^2-\|P_{\mathcal{R}(R_{A,k})}d_k^{\bar{r}}\|^2-\varepsilon^2.\end{aligned} \tag{3.19}$$

Assume that (3.11) holds. Since $\varphi(\mu,V_k)$ is monotonically decreasing and convex for $\mu>0$, the limit (3.19) has to be negative for φ to have a finite zero. Therefore, if

$$\|\bar{r}_k\|^2-(d_k^{\bar{r}})^TR_{A,k}R_{A,k}^\dagger d_k^{\bar{r}}\geq\varepsilon^2,$$

then we increase k in order for a finite zero to exist. In general, $\|P_{\mathcal{N}(R_{A,k}^T)}d_k^{\bar{r}}\|^2\ll\varepsilon^2$ and $\varphi(\mu,V_k)$ has a finite zero already for k small. We remark that if $\lim_{\mu\to\infty}\varphi(\mu,V_{k_0})<0$ for some $k_0\geq 1$, then this inequality also holds for all $k>k_0$.

The expressions (3.6) and (2.3) yield

$$\begin{aligned}\varphi'(\mu,V_k) &= 2((R_{A,k}(y_k(\mu)-\bar{y}_k))^TR_{A,k}(y_k(\mu)-\bar{y}_k)'\\ &\quad -\mu^{-2}(R_{L,k}(y_k(\mu)-\bar{y}_k))^TR_{L,k}(y_k(\mu)-\bar{y}_k)),\\ \varphi''(\mu,V_k) &= 6\|R_{A,k}y_k'(\mu)\|^2.\end{aligned} \tag{3.20}$$

We are in a position to show the asymptotic behavior of φ' and φ''. It follows from (3.20) that

$$\begin{aligned}\lim_{\mu\to\infty}\varphi'(\mu,V_k) &= \lim_{\mu\to\infty}2((R_{A,k}(y_k(\infty)-\bar{y}_k))^TR_{A,k}y_k'(\infty)\\ &\quad -\mu^{-2}(R_{L,k}(y_k(\infty)-\bar{y}_k))^TR_{L,k}(y_k(\infty)-\bar{y}_k)),\\ &= \lim_{\mu\to\infty}2\Big(0-\mu^{-2}(R_{L,k}R_{A,k}^\dagger d_k^{\bar{r}})^T(R_{L,k}R_{A,k}^\dagger d_k^{\bar{r}})\Big)=0.\end{aligned}$$

This limit is independent of V_k. Since

$$y_k(0)=\bar{y}_k,\quad y_k'(0)=(R_{L,k}^TR_{L,k})^\dagger R_{A,k}^Td_k^{\bar{r}}$$

and

$$\begin{aligned}&\lim_{\mu\to 0}\mu^{-2}(R_{L,k}(y_k(\mu)-\bar{y}_k))^TR_{L,k}(y_k(\mu)-\bar{y}_k)\\ &=\lim_{\mu\to\infty}\left(R_{L,k}(\mu R_{A,k}^TR_{A,k}+R_{L,k}^TR_{L,k})^{-1}R_{A,k}^Td_k^{\bar{r}}\right)^T\\ &\quad\times\left(R_{L,k}(\mu R_{A,k}^TR_{A,k}+R_{L,k}^TR_{L,k})^{-1}R_{A,k}^Td_k^{\bar{r}}\right)\\ &=\left(R_{L,k}(R_{L,k}^TR_{L,k})^\dagger R_{A,k}^Td_k^{\bar{r}}\right)^T\left(R_{L,k}(R_{L,k}^TR_{L,k})^\dagger R_{A,k}^Td_k^{\bar{r}}\right)\\ &=\|(R_{L,k})^\dagger R_{A,k}^Td_k^{\bar{r}}\|^2,\end{aligned}$$

it follows that

$$\begin{aligned}\lim_{\mu\to 0}\varphi'(\mu,V_k) &= \lim_{\mu\to 0} 2((R_{A,k}(y_k(0)-\bar{y}_k))^T R_{A,k}y_k'(0)\\ &\quad -\mu^{-2}(R_{L,k}(y_k(\mu)-\bar{y}_k))^T R_{L,k}(y_k(\mu)-\bar{y}_k)),\\ &= 2(0-\|(R_{L,k})^\dagger R_{A,k}^T d_k^{\bar{r}}\|^2)\\ &= -2\|(R_{L,k})^\dagger R_{A,k}^T(d_k^b - R_{A,k}\bar{y}_k)\|^2. \end{aligned} \tag{3.21}$$

Finally, for the limits of the second derivative, we have

$$\begin{aligned}\lim_{\mu\to 0}\varphi''(\mu,V_k) &= 6\|R_{A,k}y_k'(0)\|^2\\ &= 6\|R_{A,k}(R_{L,k}^T R_{L,k})^\dagger R_{A,k}^T d_k^{\bar{r}}\|^2\\ &= 6\|R_{A,k}(R_{L,k}^T R_{L,k})^\dagger R_{A,k}^T(d_k^b - R_{A,k}\bar{y}_k)\|^2\end{aligned}$$

and

$$\lim_{\mu\to\infty}\varphi''(\mu,V_k) = 6\|R_{A,k}y_k'(\infty)\|^2 = 0.$$

The asymptotic behavior of $\varphi(\mu,V_k)$ will be used below.

3.3 **Determining the regularization parameter**

We use the simplified notation $f(\mu)=\varphi(\mu,V_k)$, and similarly for the derivatives. The easiest way to evaluate $f(\mu)$, $f'(\mu)$, and $f''(\mu)$ is to first compute the GSVD of the matrix pair $\{AV_k, LV_k\}$. However, it is expensive to update the GSVD when k is increased. We therefore seek to avoid the use of the GSVD. Lampe et al. [15] compared several zero-finders, including Newton's method, and found a four-point zero-finder based on rational inverse interpolation, also described in [16,17], to be superior. We will use this zero-finder for the computed examples in Section 4. The formulas differ somewhat from those in [15]. We therefore provide a brief description.

The four-point zero-finder uses a rational function $h(f)$ to approximate f^{-1}. The function h is determined by two function values $(\mu^i, f(\mu^i))$ and two derivative values $(\mu^i, f'(\mu^i))$, $i=1,2$; see below. We evaluate h at the origin to compute $\mu_{\text{new}} = h(0) \approx f^{-1}(0)$. Specifically,

$$h(f) := \frac{p(f)}{f - f_\infty} \approx f^{-1}, \tag{3.22}$$

where $p(f)=\sum_{i=0}^3 a_i f^i$ is a polynomial. The pole is given by

$$f_\infty := \|\bar{r}_k\|^2 - (d_k^{\bar{r}})^T(R_A R_A^\dagger)d_k^{\bar{r}} - \varepsilon^2.$$

The two values μ^i are chosen so that the function values $f(\mu^i)$ do not have the same sign, i.e., we require that

$$\mu^1 < \mu^2 \quad \text{and} \quad f(\mu^1) > 0 > f(\mu^2).$$

The derivatives $f'(\mu^i)$, $i = 1, 2$, are evaluated according to (3.20), using (3.4) and (3.7). The polynomial p in (3.22) is determined by the interpolation conditions

$$h(f(\mu^i)) = \mu^i, \qquad h'(f(\mu^i)) = 1/f'(\mu^i), \qquad i = 1, 2,$$

which give rise to a linear system of equations with a 4×4 matrix for the coefficients of the polynomial. The new parameter value $\mu_{\text{new}} = h(0)$ replaces the μ^i on the same side of the root. In case $\mu_{\text{new}} \notin (\mu^1, \mu^2)$ (e.g., due to round-off errors), a bisection step is carried out. Depending on the sign of $f((\mu^1 + \mu^2)/2)$, the appropriate μ^i is updated.

3.4 **Enlarging the solution subspace**

When the vector y_k, which defines the approximate solution $x_k = V_k y_k$, and the regularization parameter μ_k have been computed, the solution subspace is enlarged as follows. The residual in Algorithm 2.1 can be evaluated by using

$$\tilde{r}_k := (A^T A V_k) y_k + \mu_k^{-1} (L^T L V_k)(y_k - \bar{y}_k) - A^T b. \tag{3.23}$$

To reduce the computational effort, we store the matrices AV_k, $A^T AV_k$, and $L^T LV_k$, and update them when k is increased; see below. The matrices $A^T AV_k$ and $L^T LV_k$ are of the same size as V_k. Since $k \ll n$, the computational effort required to evaluate matrix-vector products with the matrices $A^T AV_k$ and $L^T LV_k$ is negligible compared with the effort needed to compute a matrix-vector product with the matrix A.

In exact arithmetic, the residual vector (3.23) is orthogonal to the solution subspace $\mathcal{V}_k$. To enforce orthogonality in the presence of round-off errors, we reorthogonalize $\tilde{r}_k$ against $\mathcal{V}_k$ and then include the reorthogonalized vector in $\mathcal{V}_k$. Thus, we let

$$\hat{r}_k := (I - V_k V_k^T)\tilde{r}_k, \qquad v_{k+1} := \hat{r}_k / \|\hat{r}_k\|, \qquad V_{k+1} := [V_k, v_{k+1}].$$

Note that the spaces $\mathcal{V}_k$, $k = 1, 2, \ldots$, in general, are not Krylov subspaces when $L \neq I$, because the μ_k are updated at each step k and, therefore , so is the matrix $A^T A + \mu_k^{-1} L^T L$. We refer to the search spaces $\mathcal{V}_k$ as *generalized Krylov subspaces*.

When the new vector v_{k+1} is added to the solution subspace, the matrices AV_k, $A^T AV_k$, LV_k, and $L^T LV_k$ have to be updated to obtain AV_{k+1}, $A^T AV_{k+1}$, LV_{k+1}, and $L^T LV_{k+1}$, respectively. This requires the computation of a matrix-vector product with each one of matrices A and A^T. Specifically, we evaluate Av_{k+1}, $A^T(Av_{k+1})$, Lv_{k+1}, and $L^T(Lv_{k+1})$. The QR factorizations $AV_k = Q_{A,k} R_{A,k}$ and $LV_k = Q_{L,k} R_{L,k}$ are updated as follows:

$$A[V_k, v_{k+1}] = [Q_{A,k}, \tilde{q}_{A,k+1}] \begin{bmatrix} R_{A,k} & r_{A,k+1} \\ 0 & \rho_{A,k+1} \end{bmatrix}, \tag{3.24}$$

$$L[V_k, v_{k+1}] = [Q_{L,k}, \tilde{q}_{L,k+1}] \begin{bmatrix} R_{L,k} & r_{L,k+1} \\ 0 & \rho_{L,k+1} \end{bmatrix}; \tag{3.25}$$

see Daniel et al. [1] for a detailed discussion on updating and downdating of QR factorizations. The new vectors and scalars in the above factorizations are obtained by

$$\begin{aligned} r_{A,k+1} &= Q_{A,k}^T(Av_{k+1}), \quad q_{A,k+1} = Av_{k+1} - r_{A,k+1}, \\ \rho_{A,k+1} &= \|q_{A,k+1}\|, \quad \tilde{q}_{A,k+1} = q_{A,k+1}/\rho_{A,k+1}; \\ r_{L,k+1} &= Q_{L,k}^T(Lv_{k+1}), \quad q_{L,k+1} = Lv_{k+1} - r_{L,k+1}, \\ \rho_{L,k+1} &= \|q_{L,k+1}\|, \quad \tilde{q}_{L,k+1} = q_{L,k+1}/\rho_{L,k+1}. \end{aligned}$$

When $\rho_{A,k+1}$ vanishes, the expression for $\tilde{q}_{A,k+1}$ is replaced by an arbitrary unit vector such that $Q_{A,k}^T\tilde{q}_{A,k+1} = 0$. We proceed analogously when $\rho_{L,k+1} = 0$. Updating the factorizations (3.24) and (3.25) requires evaluation of matrix-vector products with the matrices $Q_{A,k}^T \in \mathbb{R}^{k\times m}$ and $Q_{L,k}^T \in \mathbb{R}^{k\times p}$. Since k is small, the computational expense for these evaluations is negligible. The vector (3.8) is updated according to

$$d_{k+1}^b = \begin{bmatrix} d_k^b \\ \tilde{q}_{A,k+1}^T b \end{bmatrix}.$$

In summary, the computational cost for increasing the dimension of the solution subspace by one is dominated by one matrix-vector product evaluation with each one of the matrices A and A^T. This also is the dominating cost for evaluating one matrix-vector product with the matrix $A^TA+\mu_k^{-1}L^TL$. We remark that commonly used regularization matrices L are discrete approximations of a differential operator. They therefore are sparse and matrix-vector product evaluations with L and L^T are inexpensive. The cost for enlarging the solution subspace therefore is acceptable.

4 Numerical examples

We illustrate the performance of Algorithm 2.1 with three examples. The first two examples are from Hansen's *Regularization Tools* [10] and the last one is concerned with the restoration of a blur- and noise-contaminated image. All examples are discrete ill-posed problems. The initial value for the regularization parameter is chosen to be $\mu_0 = 10$ in all examples. The parameter η in (2.3) in the first two examples is set to 1.1. Our method, projected nonstationary iterative Tikhonov regularization (PNITR), is compared to the generalized Krylov subspace-based Tikhonov regularization (GKSTR) method described in [15]. Both methods use the discrepancy principle to determine the regularization parameter.

In our computed examples, the vector b_{true} in (1.4) is available. The error e in the vector b in (1.1) models white Gaussian noise. Given b_{true}, we add e to determine b; cf. (1.2). We refer to the quotient

$$\sigma := \frac{\|e\|}{\|b_{\text{true}}\|}$$

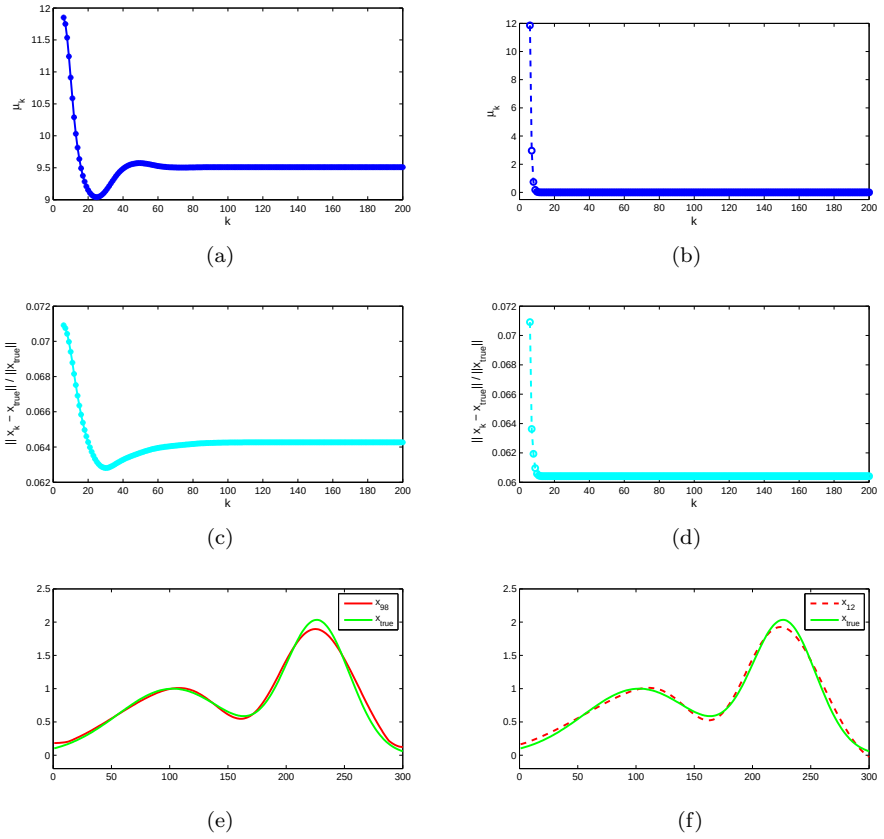


Fig. 4.1 Example 4.1: Convergence of the Tikhonov parameters for GKSTR (a) and PNITR (b), relative errors of GKSTR (c) and PNITR (d), exact and computed approximate solutions by the methods GKSTR (e) and PNITR (f) with $L = L_1$.

as the *noise level*. All experiments are carried out on an Intel Core i5-3230M 2.60GHz computer with 8GB RAM. The computations were carried out in MATLAB with about 15 significant decimal digits.

Example 4.1. We solve an overdetermined linear system of equations with coefficient matrix A and right-side of vector b. Let $\hat{A} \in \mathbb{R}^{300\times 300}$, x_{true} and $\hat{b} := \hat{A}x_{\text{true}}$ be produced by the MATLAB function `shaw` in [10], and define the stacked matrix A and vector b_{true} by

$$A = \begin{bmatrix} \hat{A} \\ \hat{A} \end{bmatrix}, \qquad b_{\text{true}} = \begin{bmatrix} \hat{b} \\ \hat{b} \end{bmatrix},$$

and the error-contaminated vector b described above with different noise levels $\sigma = 1\cdot 10^{-3}, 5\cdot 10^{-3}, 1\cdot 10^{-2}, 5\cdot 10^{-2}$. Stacked problems of this kind arise when two measurements b with different errors are available.

The condition number $\kappa(A) := \|A\|\|A^{\dagger}\|$ of A exceeds $4{\cdot}10^{12}$. Thus, the matrix is numerically singular. The regularization matrix L is chosen to be scaled

discrete first and second order derivative operators in one space-dimension, i.e.,

$$L_1 = \begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{bmatrix} \in \mathbb{R}^{(n-1)\times n}, \tag{4.1}$$

or

$$L_2 = \begin{bmatrix} -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \end{bmatrix} \in \mathbb{R}^{(n-2)\times n} \tag{4.2}$$

with $n = 300$.

The initial search space is the Krylov subspace (1.11) with $\ell = 6$. When $\ell < 6$, the condition $\lim_{\mu\to\infty} \varphi(\mu, V_0) < 0$ is violated; cf. (3.13). An index ℓ such that $\lim_{\mu\to\infty} \varphi(\mu, V_0) < 0$ generally is not known a priori, but can be determined during the computations. We would like to illustrate the convergence of several interesting quantities and therefore do not terminate the computations with Algorithm 2.1 until the solution subspace is of dimension 200.

In Table 4.1, we compare the relative error and iteration number using GKSTR and PNITR with different noise levels $\sigma = 1\cdot10^{-3}, 5\cdot10^{-3}, 1\cdot10^{-2}, 5\cdot10^{-2}$ and regularization matrices I, L_1 in (4.1) and L_2 in (4.2). In each bracket, the left item is the number of iteration required until the computed solution does not change any more or reaches the maximum iteration number, while the right one is the corresponding relative error.

Figure 4.1 displays the convergence of iterated Tikhonov parameters $\{\mu_k\}$, the relative errors $\{\|x_k - x_{\text{true}}\|/\|x_{\text{true}}\|\}$, as well as exact and computed approximate solutions determined by the methods GKSTR and PNITR with L given by (4.1) and $\sigma = 1 \cdot 10^{-3}$. The left-hand side column of Figure 4.1 displays results for GKSTR while the right-hand side column is for PNITR. Figure 4.1(a) shows the regularization parameter values $\{\mu_k\}$ for GKSTR to decrease quickly, then increase and approach the value $\bar{\mu} = 9.508$ as the dimension k of the solution subspace $\mathcal{V}_k$ increases. Figure 4.1(b) is analogous for PNITR. The convergence of the relative errors $\{\|x_k - x_{\text{true}}\|/\|x_{\text{true}}\|\}$ achieved with GKSTR and PNITR is displayed in Figure 4.1(c) and (d), respectively. The exact solutions x_{true} together with the computed approximations x_{98} determined by GKSTR and x_{12} by PNITR are shown in Figures 4.1 (e) and (f), respectively.

Figure 4.2 shows the convergence of the Tikhonov parameter, relative error of PNITR, and exact and computed approximate solutions by the methods GKSTR and PNITR with $L = L_2$ defined by (4.2). The results in Figure 4.2 is analogous to Figure 4.1. Both in Figures 4.1 and 4.2, the regularization parameter for the PNITR method is seen to converge faster than for the GKSTR method. Moreover, the relative errors converge faster for PNITR and the computed solutions are of higher quality than for the GKSTR method. Figures 4.1-4.2, as well as Table 4.1, illustrate the benefit of using the PNITR method. □

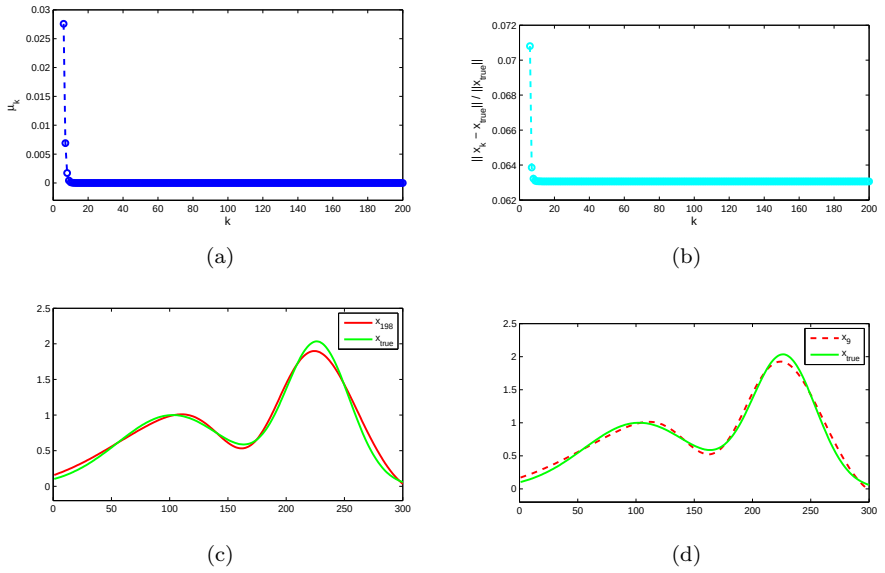


Fig. 4.2 Example 4.1: Convergence of the Tikhonov parameter (a), relative error (b) of PNITR, and exact and computed approximate solutions by the methods GKSTR (c) and PNITR (d) with $L = L_2$.

Table 4.1 Example 4.1: Comparison of relative error and iteration number using GKSTR and PNITR with different noise levels and regularization matrices. In each bracket, the left and right items are the number of iterations required until the computed solution does not change much any more or reaches the maximum iteration number, and the corresponding relative error, respectively.

Method	$L = I$	$L = L_1$	$L = L_2$
$\sigma = 1 \cdot 10^{-3}$			
GKSTR	(9,0.0648)	(98,0.0643)	(198,0.0724)
PNITR	(11,0.0550)	(12,0.0604)	(9,0.0631)
$\sigma = 5 \cdot 10^{-3}$			
GKSTR	(8,0.1396)	(88,0.2318)	(193,0.1935)
PNITR	(11,0.1314)	(11,0.1555)	(15,0.1558)
$\sigma = 1 \cdot 10^{-2}$			
GKSTR	(6,0.1579)	(71,0.4062)	(199,0.3170)
PNITR	(12,0.1430)	(11,0.3034)	(7,0.3021)
$\sigma = 5 \cdot 10^{-2}$			
GKSTR	(6,0.1945)	(88,0.5262)	(200,0.3285)
PNITR	(13,0.1787)	(13,0.3734)	(12,0.2924)

Example 4.2. In this example we solve another overdetermined linear system of equations obtained by stacking the matrix $\hat{A} \in \mathbb{R}^{500\times 500}$ and $\hat{b} := \hat{A}x_{\text{true}}$ of the inverse heat equation problem `heat` from [10] with parameter $\kappa = 5$, i.e.,

$$A = \begin{bmatrix} \hat{A} \\ \hat{A} \end{bmatrix}, \qquad b_{\text{true}} = \begin{bmatrix} \hat{b} \\ \hat{b} \end{bmatrix},$$

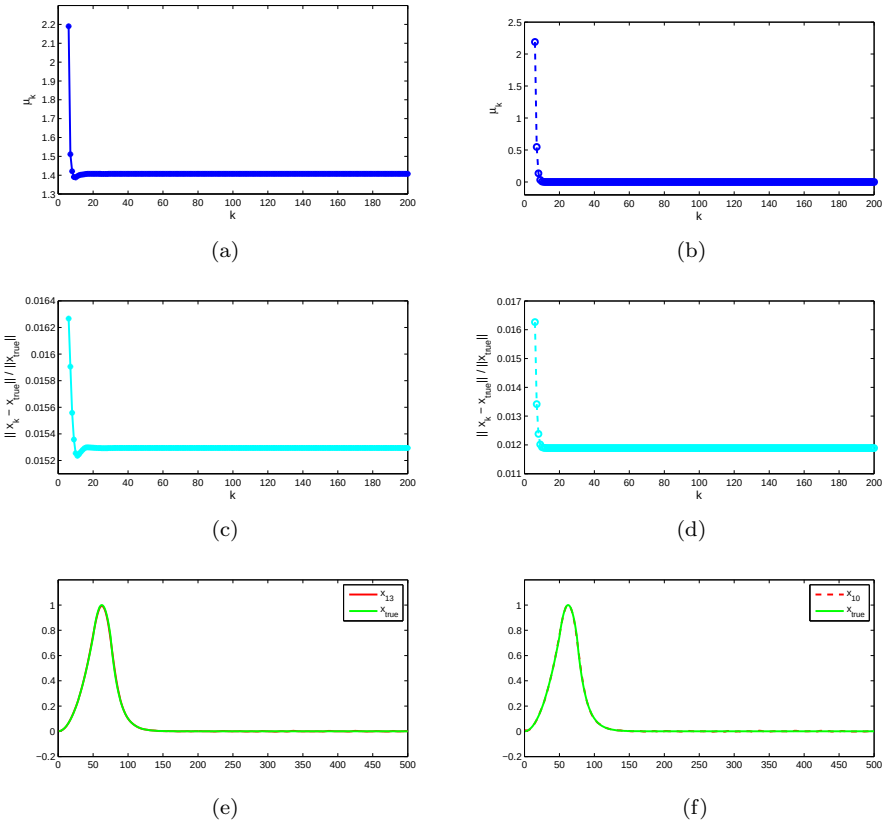


Fig. 4.3 Example 4.2: Convergence of the Tikhonov parameters for GKSTR (a) and PNITR (b), relative errors of GKSTR (c) and PNITR (d), exact and computed approximate solutions by the methods GKSTR (e) and PNITR (f) with $L = L_1$.

with the matrix $A \in \mathbb{R}^{1000\times 500}$ and the error-contaminated vector b defined as described above with noise level $\sigma = 1 \cdot 10^{-2}$. The initial solution subspace is $\mathcal{V}_0 = \mathcal{K}_6(A^TA, A^Tb)$. Similarly to Example 4.1, we do not terminate the computations with Algorithm 2.1 until the dimension of the solution subspace reaches 200 in order to see the convergence of several interesting quantities.

Figure 4.3 is analogous to Figure 4.1 and shows results for the regularization matrix L_1. Figures 4.3 (e) and (f) display the computed approximate solutions x_{13} determined with GKSTR and x_{10} with PNITR, respectively, as well as the exact solution x_{true}. The relative errors in these computed solutions and the iteration number required until the computed solution does not change any more or reaches the maximum iteration number with different regularization matrices are reported in Table 4.2. Figure 4.4 is similar to Figure 4.2. Figures 4.4(c) and (d) show the exact solution x_{true} and computed approximate solutions x_{91} determined by GKSTR and x_9 by PNITR, respectively. Figures 4.3-4.4 and Table 4.2 illustrate that both the regularization parame-

ter and relative error converge faster for PNITR than for GKSTR, and the computed solutions determined by PNITR are of higher quality than those obtained with the GKSTR method. □

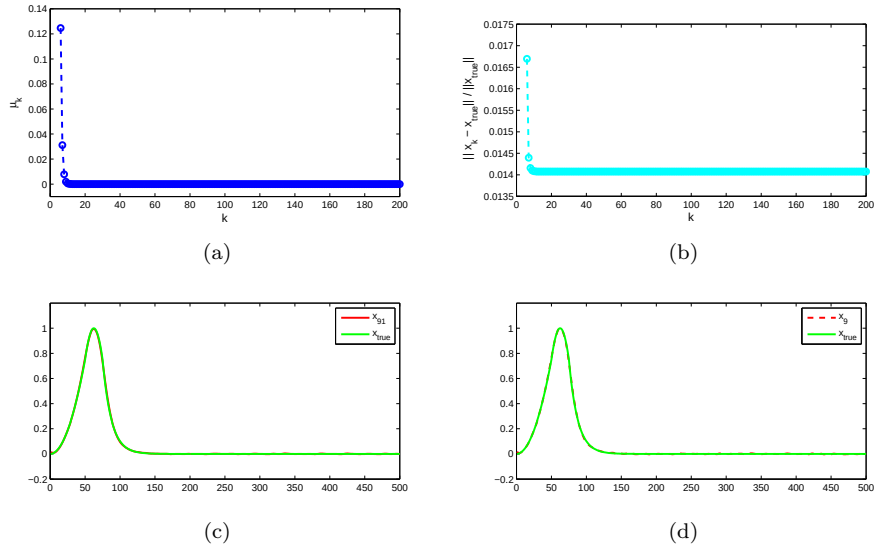
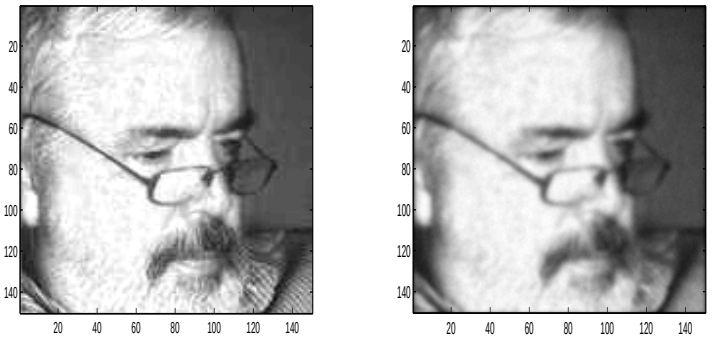


Fig. 4.4 Example 4.2: Convergence of the Tikhonov parameter (a), relative error of approximate solutions determined by PNITR (b), and exact and computed approximate solutions obtained with GKSTR (c) and PNITR (d) with $L = L_2$.

Table 4.2 Example 4.2: Comparison of relative error and iteration number using GKSTR and PNITR for different regularization matrices. In each bracket, the left and right items are the number of iterations required until the computed solution does not change much any more or reaches the maximum iteration number and the corresponding relative error.

Method	$L = I$	$L = L_1$	$L = L_2$
GKSTR	(27,0.03)	(13,0.0153)	(91,0.0173)
PNITR	(11,0.0111)	(10,0.0119)	(9,0.0141)

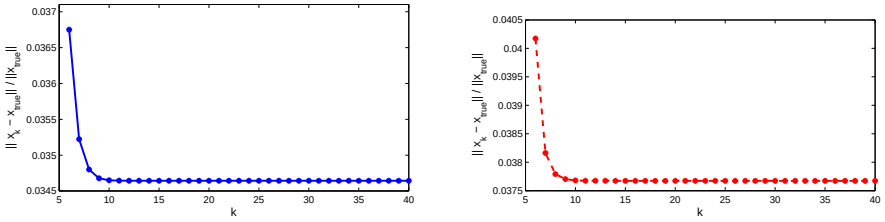
Example 4.3. This example considers the restoration of a gray scale image that is represented by an array of 150×150 pixels. Figure 4.5(a) depicts the original image "Voss", which is assumed not to be available, and Figure 4.5(b) shows the available blur- and noise-contaminated image. The noise is white Gaussian and corresponds to the noise level $\sigma = 1 \cdot 10^{-2}$. The contaminated gray scale image is stored column-wise in the vector $b \in \mathbb{R}^{22500}$. The matrix $A \in \mathbb{R}^{22500 \times 22500}$ is block Toeplitz with Toeplitz blocks and represents the blurring operator. It is generated with the function `blur` from [10] using the



(a) (b)

Fig. 4.5 Example 4.3: Original (a), and blurred and noisy (b) Voss pictures.

parameter values $band = 5$ and $sigma = 1$. The factor η in (2.3) is set to 1.05. We would like to determine an approximation of the uncontaminated image represented by $x_{\text{true}} \in \mathbb{R}^{22500}$ given A and b.



(a) (b)

Fig. 4.6 Example 4.3: Relative errors of the restored images using the PNITR method with $L = I$ (a) and $L_{2,2D}$ (b).

Define the regularization matrix

$$L_{2,2D} = \begin{bmatrix} L_2 \otimes I_n \\ I_n \otimes L_2 \end{bmatrix},$$

which is a discrete approximation of the second order derivative operator in two space-dimensions. Here L_2 is defined by (4.2) with $n = 150$, and I_n is the identity matrix of order $n = 150$.

We would like to recover the blur- and noise-free image using Algorithm 2.1 with the regularization matrices $L = I$ and $L_{2,2D}$. The initial solution subspace is chosen to be $\mathcal{V}_0 = \mathcal{K}_6(A^T A, A^T b)$ and we increase the number of

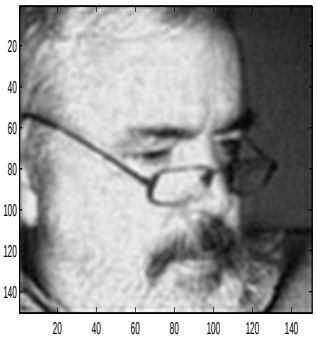


Fig. 4.7 Example 4.3: Restored Voss images using PNITR with $L_{2,2D}$.

steps k until the solution subspace is of dimension 40. Thus, the number of iterations is 34.

Figure 4.6 shows the relative errors of the restored images using PNITR with the regularization matrices $L = I$ and $L_{2,2D}$. The restored image represented by x_{40} determined by PNITR with $L_{2,2D}$ is displayed in Figure 4.7.

5 Conclusion

A projected nonstationary iterated Tikhonov regularization method for large-scale Tikhonov regularization problems with general regularization matrices is presented. Numerical examples show this method to compute approximate solutions of higher quality than a related "stationary" Tikhonov regularization method described in [15].

Acknowledgments. The authors would like to thank the referees for comments that lead to improvements of the presentation. The research of L.R. is supported in part by NSF grant DMS-1115385. The work of G.H. is supported by Fund of China Scholarship Council, the young scientic research backbone teachers of CDUT (KYGG201309). The work of F.Y. is supported by Research Fund Project (NS2014PY08) of SUSE, and Key Natural Science Foundation of Sichuan Education Department (15ZA0234).

References

1. Daniel, J.W., Gragg, W.B., Kaufman, L., Stewart, G.W.: Reorthogonalization and stable algorithms for updating the Gram–Schmidt QR factorization. Math. Comp. **30**, 772–795 (1976)
2. Donatelli, M., Hanke, M.: Fast nonstationary preconditioned iterative methods for ill-posed problems, with application to image deblurring. Inverse Problems **29**, 095008 (2013)

3. Dykes, L., Noschese, S., Reichel, L.: Rescaling the GSVD with application to ill-posed problems. Numer. Algorithms **68**, 531–545 (2015)
4. Eldén, L.: A weighted pseudoinverse, generalized singular values, and constrained least squares problems. BIT **22**, 487–502 (1982)
5. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Kluwer, Dordrecht, The Netherlands (1996)
6. Gazzola, S., Novati, P., Russo, M.R.: On Krylov projection methods and Tikhonov regularization. Electron. Trans. Numer. Anal. **44**, 83–123 (2015)
7. Groetsch, C.W.: The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind. Pitman, Boston (1984)
8. Hanke M., Groetsch, C.W.: Nonstationary iterated Tikhonov regularization. J. Optim. Theory Appl. **436**, 37–53 (1998)
9. Hansen, P.C.: Rank-Deficient and Discrete Ill-Posed Problems. SIAM, Philadelphia (1998)
10. Hansen, P.C.: Regularization tools version 4.0 for Matlab 7.3. Numer. Algorithms **46**, 189–194 (2007)
11. Hochstenbach, M.E., Reichel, L.: An iterative method for Tikhonov regularization with a general linear regularization operator. J. Integral Equations Appl. **22**, 463–480 (2010)
12. Hochstenbach, M.E., Reichel, L., Yu, X.: A Golub–Kahan-type reduction method for matrix pairs. J. Sci. Comput., in press
13. Kilmer, M.E., Hansen, P.C., Español, M.I.: A projection-based approach to general-form Tikhonov regularization. SIAM J. Sci. Comput. **29**, 315–330 (2007)
14. Kindermann, S.: Convergence analysis of minimization-based noise-level-free parameter choice rules for ill-posed problems. Electron. Trans. Numer. Anal. **38**, 233–257 (2011)
15. Lampe, J., Reichel, L., Voss, H.: Large-scale Tikhonov regularization via reduction by orthogonal projection. Linear Algebra Appl. **436**, 2845–2865 (2012)
16. Lampe J., Voss, H.: A fast algorithm for solving regularized total least squares problems. Electron. Trans. Numer. Anal. **31**, 12–24 (2008)
17. Lampe, J., Voss, H.: Solving regularized total least squares problems based on eigenproblems. Taiwanese J. Math. **14**, 885–909 (2010)
18. Lampe, J., Voss, H.: Large-scale dual regularized total least squares. Electron. Trans. Numer. Anal. **42**, 13–40 (2014)
19. Reichel, L., Rodriguez, G.: Old and new parameter choice rules for discrete ill-posed problems. Numer. Algorithms **63**, 65–87 (2013)
20. Reichel, L., Sgallari, F., Ye, Q.: Tikhonov regularization based on generalized Krylov subspace methods. Appl. Numer. Math. **2**, 474–489 (2012)
21. Reichel L., Ye, Q.: Simple square smoothing regularization operators. Electron. Trans. Numer. Anal. **33**, 63–83 (2009)
22. Reichel L., Yu, X.: Tikhonov regularization via flexible Arnoldi reduction. BIT, in press
23. Voss, H.: An Arnoldi method for nonlinear eigenvalue problems. BIT **44**, 387–401 (2004)