GREEDY TIKHONOV REGULARIZATION FOR LARGE LINEAR ILL-POSED PROBLEMS

L. REICHEL *, H. SADOK †, AND A. SHYSHKOV ‡

Abstract. Several numerical methods for the solution of large linear ill-posed problems combine Tikhonov regularization with an iterative method based on partial Lanczos bidiagonalization of the operator. This paper discusses the determination of the regularization parameter and the dimension of the Krylov subspace for this kind of methods. A method that requires a Krylov subspace of minimal dimension is referred to as greedy.

Key words. ill-posed problem, inverse problem, regularization, Lanczos bidiagonalization, discrepancy principle.

1. Introduction.

This paper is concerned with the solution of linear systems of equations

$$A\boldsymbol{x} = \boldsymbol{b}, \qquad A \in \mathbb{R}^{m\times n}, \qquad \boldsymbol{x} \in \mathbb{R}^n, \qquad \boldsymbol{b} \in \mathbb{R}^m, \tag{1.1}$$

with a large matrix of ill-determined rank. In particular, A is severely ill-conditioned and has many singular values of different orders of magnitude close to zero; some singular values may be vanishing. We allow $m \neq n$; if the linear system (1.1) is inconsistent, then we consider the associated least-squares problem.

Linear systems of equations with a matrix of ill-determined rank are often referred to as linear discrete ill-posed problems. They are obtained, for instance, when discretizing linear ill-posed problems, such as Fredholm integral equations of the first kind with a smooth kernel. This type of integral equations arises in science and engineering, when one seeks to determine the cause (the solution) of an observed effect represented by $\boldsymbol{b}$. Since the right-hand side is obtained through observation, it is typically contaminated by a measurement error $\boldsymbol{e} \in \mathbb{R}^m$, which we refer to as *noise*. Let $\hat{\boldsymbol{b}} \in \mathbb{R}^m$ denote the unavailable error-free right-hand side associated with $\boldsymbol{b}$, i.e.,

$$\boldsymbol{b} = \hat{\boldsymbol{b}} + \boldsymbol{e}. \tag{1.2}$$

Many solution methods for linear discrete ill-posed problems of small to moderate size compute the singular value decomposition of A; see, e.g., Hansen [18] for discussions of several such methods. We are concerned with the solution of problems (1.1) that are too large to allow the computation of the singular value decomposition of the matrix, and focus on iterative solution methods based on partial Lanczos bidiagonalization of A.

The linear system of equations with the unavailable noise-free right-hand side,

$$A\boldsymbol{x} = \hat{\boldsymbol{b}}, \tag{1.3}$$

is assumed to be consistent. Let $\hat{\boldsymbol{x}}$ denote the solution of minimal Euclidean norm of (1.3); it can be expressed as $\hat{\boldsymbol{x}} = A^\dagger \hat{\boldsymbol{b}}$, where $A^\dagger$ denotes the Moore-Penrose pseudo-

*Department of Mathematical Sciences, Kent State University, Kent, OH 44242. E-mail: reichel@math.kent.edu. Research supported in part by an OBR Research Challenge Grant.

†Laboratoire de Mathématiques Pures et Appliquées, Université du Littoral, Centre Universtaire de la Mi-Voix, Batiment H. Poincarré, 50 Rue F. Buisson, BP 699, 62228 Calais cedex, France. E-mail: Hassane.Sadok@lmpa.univ-littoral.fr.

‡Department of Mathematical Sciences, Kent State University, Kent, OH 44242. E-mail: ashyshko@math.kent.edu. Research supported in part by an OBR Research Challenge Grant.

inverse of A. We seek to determine an approximation of $\hat{\boldsymbol{x}}$ by computing an approximate solution of the available linear system of equations (1.1). The norm of the noise

$$\varepsilon = \|\boldsymbol{e}\|, \tag{1.4}$$

or a fairly accurate estimate thereof, is assumed to be known. Here and throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm. Note that due to the error $\boldsymbol{e}$ in $\boldsymbol{b}$ and the ill-conditioning of A, the vector $A^\dagger \boldsymbol{b} = \hat{\boldsymbol{x}} + A^\dagger \boldsymbol{e}$ generally does not furnish a useful approximation of $\hat{\boldsymbol{x}}$.

In order to be able to determine a meaningful approximation of $\hat{\boldsymbol{x}}$, one typically replaces the linear system (1.1) by a nearby system that is less sensitive to perturbations of the right-hand side, and considers the solution of the latter system an approximation of $\hat{\boldsymbol{x}}$. This replacement is commonly referred to as regularization. Tikhonov regularization and truncated iteration are the most popular regularization methods; see Engl et al. [9], Groetsch [13], Hanke [14], and Hansen [18] for detailed discussions of these methods.

Tikhonov regularization in its simplest form replaces (1.1) by the linear system of equations

$$(A^T A + \mu^{-1} I)\boldsymbol{x} = A^T \boldsymbol{b} \tag{1.5}$$

with a positive regularization parameter μ. The value of μ determines how sensitive the solution

$$\boldsymbol{x}_\mu = (A^T A + \mu^{-1} I)^{-1} A^T \boldsymbol{b} \tag{1.6}$$

of the regularized system (1.5) is to the error $\boldsymbol{e}$, and how close $\boldsymbol{x}_\mu$ is to the solution $\hat{\boldsymbol{x}}$ of (1.3). The *discrepancy principle* suggests that the regularization parameter μ be chosen so that the discrepancy

$$\boldsymbol{d}_\mu = \boldsymbol{b} - A\boldsymbol{x}_\mu$$

satisfies

$$\|\boldsymbol{d}_\mu\| = \eta\varepsilon \tag{1.7}$$

for some constant $\eta > 1$, whose size reflects the uncertainty in the estimate (1.4). It can be shown that for any fixed $\eta > 1$, $\mu = \mu(\varepsilon)$ determined by (1.7), and $\boldsymbol{x}_\mu$ given by (1.6), we have $\boldsymbol{x}_\mu \to \hat{\boldsymbol{x}}$ as $\varepsilon \searrow 0$; see, e.g., Engl et al. [9] or Groetsch [13] for proofs in Hilbert space settings.

Several iterative methods for the solution of the Tikhonov equation (1.5) are based on partial Lanczos bidiagonalization of A; see, e.g., Björck [2, 3], Calvetti et al. [4, 5, 6], Golub and von Matt [12], Hanke [15], and O'Leary and Simmons [24]. These methods determine approximations of the vector (1.6) in a Krylov subspace of the form

$$\mathbb{K}_\ell(A^T A, A^T \boldsymbol{b}) = \mathrm{span}\{A^T \boldsymbol{b}, A^T A A^T \boldsymbol{b}, \ldots, (A^T A)^{\ell-1} A^T \boldsymbol{b}\}. \tag{1.8}$$

Note that this subspace is independent of μ. A recent survey of methods that first apply a few steps of an iterative method, such as partial Lanczos bidiagonalization, to obtain a reduced problem, and then regularize the latter by some technique, is presented by Kilmer and O'Leary [19].

When A is large, the major computational effort required by all of these methods is the evaluation of matrix-vector products with the matrices A and A^T; the determination of a vector in $\mathbb{K}_\ell(A^TA, A^T\boldsymbol{b})$ may require up to $2\ell - 1$ matrix-vector product evaluations, $\ell - 1$ with A and ℓ with A^T. It is desirable to determine an approximate solution of (1.5) in a Krylov subspace (1.8) of low dimension in order to keep the required number of matrix-vector product evaluations small. We remark that in some applications, A has a structure, such as Toeplitz or Hankel, that makes it possible to reduce the arithmetic work required for evaluating matrix-vector products; see, e.g., Ng [23, Section 3.4].

LSQR is an implementation of the conjugate gradient method applied to the normal equations

$$A^TA\boldsymbol{x} = A^T\boldsymbol{b} \tag{1.9}$$

associated with (1.1); see Björck [3] or Paige and Saunders [25] for properties of this method. Let the initial iterate be $\boldsymbol{x}^{(0)} = \boldsymbol{0}$. Then the ℓth iterate, $\boldsymbol{x}^{(\ell)}$, determined by LSQR lives in the Krylov subspace (1.8). Section 2 uses properties of LSQR to derive a scheme for computing an approximate solution of (1.5) with discrepancy $\eta\varepsilon$ in a Krylov subspace of the form (1.8) of minimal dimension. Because of the latter property, we refer to this scheme as a *greedy Tikhonov method*. Section 3 describes a modification of this method, and Section 4 presents a few numerical examples. Concluding remarks can be found in Section 5.

2. A greedy Tikhonov method.

Introduce the function

$$\phi(\mu) = \|\boldsymbol{d}_\mu\|^2. \tag{2.1}$$

The following properties of ϕ have been shown in [6].

PROPOSITION 2.1. *The function (2.1) allows the representation*

$$\phi(\mu) = \boldsymbol{b}^T(\mu AA^T + I)^{-2}\boldsymbol{b}. \tag{2.2}$$

Assume that $A^T\boldsymbol{b} \neq \boldsymbol{0}$. Then ϕ is strictly decreasing and convex for $\mu \geq 0$. The equation

$$\phi(\mu) = \tau$$

has a unique solution μ, such that $0 < \mu < \infty$, for any τ that satisfies $\|\boldsymbol{b}_0\|^2 < \tau < \|\boldsymbol{b}\|^2$, where $\boldsymbol{b}_0$ denotes the orthogonal projection of $\boldsymbol{b}$ onto the null space of AA^T.

Proof. The representation (2.2) follows from the identity

$$I - A(A^TA + \mu^{-1}I)^{-1}A^T = (\mu AA^T + I)^{-1}. \tag{2.3}$$

The remaining properties of ϕ follow from this representation, for details see [6, Theorem 2.1]. □

Equation (1.7) is equivalent to

$$\phi(\mu) = \eta^2\varepsilon^2. \tag{2.4}$$

We will assume that

$$\|\boldsymbol{b}_0\| < \eta\|\boldsymbol{e}\| < \|\boldsymbol{b}\| \tag{2.5}$$

holds, where $\boldsymbol{b}_0$ is defined in Proposition 2.1. Then it follows from the proposition that equation (2.4) has a unique solution, μ_*, such that $0 < \mu_* < \infty$.

The methods of this paper are based on the partial Lanczos bidiagonalization algorithm by Paige and Saunders [25, 26]. Application of ℓ steps of the algorithm yields the decompositions

$$AV_\ell = U_{\ell+1}\bar{C}_\ell, \qquad A^T U_\ell = V_\ell C_\ell^T, \tag{2.6}$$

where the matrices $U_{\ell+1} \in \mathbb{R}^{m\times(\ell+1)}$ and $V_\ell \in \mathbb{R}^{n\times\ell}$ have orthonormal columns, and $U_{\ell+1}\boldsymbol{e}_1 = \boldsymbol{b}/\|\boldsymbol{b}\|$, where $\boldsymbol{e}_1$ denotes the first axis vector. Moreover, $U_\ell \in \mathbb{R}^{m\times\ell}$ consists of the ℓ first columns of $U_{\ell+1}$, and

$$\bar{C}_\ell = \begin{bmatrix} \rho_1 & & & 0 \\ \sigma_2 & \rho_2 & & \\ & \ddots & \ddots & \\ & & \sigma_\ell & \rho_\ell \\ 0 & & & \sigma_{\ell+1} \end{bmatrix} \in \mathbb{R}^{(\ell+1)\times\ell}$$

is lower bidiagonal with $\rho_j > 0$ and $\sigma_j > 0$ for all j. The matrix C_ℓ is the leading $\ell \times \ell$ submatrix of $\bar{C}_\ell$. Moreover,

$$\mathrm{range}(V_\ell) = \mathbb{K}_\ell(A^TA, A^T\boldsymbol{b}); \tag{2.7}$$

thus, the columns of V_ℓ furnish an orthonormal basis of (1.8). We assume that ℓ is chosen small enough so that the decompositions (2.6) with the stated properties exists. The computations with the regularization methods of the present paper simplify when $\sigma_{j+1} = 0$. We will comment on this situation at the end of this section. Assume for the moment that the decomposition (2.6) and a suitable value of μ are available. Then it is convenient to determine an approximate solution $\boldsymbol{x}_\mu^{(\ell)} \in \mathbb{K}_\ell(A^TA, A^T\boldsymbol{b})$ of the Tikhonov equation (1.5) by solving the associated Galerkin equation

$$V_\ell^T(A^TA + \mu^{-1}I)V_\ell\boldsymbol{y} = V_\ell^TA^T\boldsymbol{b} \tag{2.8}$$

for $\boldsymbol{y}_\mu^{(\ell)} \in \mathbb{R}^\ell$ and letting

$$\boldsymbol{x}_\mu^{(\ell)} = V_\ell\boldsymbol{y}_\mu^{(\ell)}. \tag{2.9}$$

Using (2.6), we can simplify (2.8) to obtain

$$(\bar{C}_\ell^T\bar{C}_\ell + \mu^{-1}I_\ell)\boldsymbol{y} = \|\boldsymbol{b}\|\bar{C}_\ell^T\boldsymbol{e}_1, \tag{2.10}$$

which are the normal equations associated with the least-squares problem

$$\min_{\boldsymbol{y}\in\mathbb{R}^\ell}\left\| \begin{bmatrix} \mu^{1/2}\bar{C}_\ell \\ I_\ell \end{bmatrix}\boldsymbol{y} - \|\boldsymbol{b}\|\mu^{1/2}\boldsymbol{e}_1 \right\|. \tag{2.11}$$

We compute the solution $\boldsymbol{y}_\mu^{(\ell)}$ of (2.8) by solving (2.11). This can be done in only $\mathcal{O}(\ell)$ arithmetic floating point operations by application of a judiciously chosen sequence of Givens rotations; see Eldén [8] or Paige and Saunders [26] for descriptions of such solution methods.

We turn to the computation of a suitable value of μ. Introduce, analogously to (2.1), the function

$$\bar{\phi}_\ell(\mu) = \|\boldsymbol{b} - A\boldsymbol{x}_\mu^{(\ell)}\|^2,$$

where $\boldsymbol{x}_\mu^{(\ell)}$ is given by (2.9). It follows from the matrix identity (2.3), with A replaced by $\bar{C}_\ell$, that analogously to (2.2), we have

$$\bar{\phi}_\ell(\mu) = \|\boldsymbol{b}\|^2 \boldsymbol{e}_1^T(\mu \bar{C}_\ell \bar{C}_\ell^T + I_{\ell+1})^{-2}\boldsymbol{e}_1.$$

Using this representation, we evaluate the function $\bar{\phi}_\ell(\mu)$ by solving a least-squares problem related to (2.11). Thus, we determine the vector $\boldsymbol{z}_\mu^{(\ell+1)} = \|\boldsymbol{b}\|(\bar{C}_\ell\bar{C}_\ell^T + I_{\ell+1})^{-1}\boldsymbol{e}_1$ by first solving

$$\min_{\boldsymbol{z}\in\mathbb{R}^\ell} \left\| \begin{bmatrix} \mu^{1/2}\bar{C}_\ell^T \\ I_{\ell+1} \end{bmatrix} \boldsymbol{z} - \|\boldsymbol{b}\|\boldsymbol{e}_{\ell+1} \right\| \tag{2.12}$$

and then evaluating $\bar{\phi}_\ell(\mu) = (\boldsymbol{z}_\mu^{(\ell+1)})^T \boldsymbol{z}_\mu^{(\ell+1)}$. The solution of (2.12) requires only $\mathcal{O}(\ell)$ arithmetic floating point operations for each value of $\mu > 0$, similarly as the solution of (2.11). The numerical methods to be described are based on the following properties of $\bar{\phi}_\ell$.

THEOREM 2.2. *Let the integer $\ell \geq 2$ be such that the decompositions (2.6) with the stated properties exists. Then*

$$\phi(\mu) < \bar{\phi}_\ell(\mu) < \bar{\phi}_{\ell-1}(\mu), \qquad \mu > 0. \tag{2.13}$$

The function $\bar{\phi}_\ell(\mu)$ is decreasing and convex for $\mu \geq 0$. Moreover, $\bar{\phi}_\ell(0) = \|\boldsymbol{b}\|^2$ and

$$\lim_{\mu\to\infty} \bar{\phi}_\ell(\mu) = \omega_{0,\ell+1}, \tag{2.14}$$

where $\omega_{0,\ell+1}$ is the weight for the node at the origin of an $(\ell+1)$-point Gauss-Radau quadrature rule.

Proof. Substituting the spectral decomposition of AA^T into (2.2) yields a representation of ϕ in terms of a Stieltjes integral,

$$\phi(\mu) = \int_0^\infty \psi_\mu(t)d\omega(t), \qquad \psi_\mu(t) = (\mu t^2 + 1)^{-2},$$

where the distribution function $\omega(t)$ is a step function with jumps at the eigenvalues. The function $\bar{\phi}_\ell(\mu)$ can be interpreted as an $(\ell+1)$-point Gauss-Radau quadrature rule,

$$\bar{\phi}_\ell(\mu) = \sum_{j=0}^{\ell} \psi_\mu(\theta_{j,\ell+1})\omega_{j,\ell+1}, \tag{2.15}$$

associated with the distribution function ω and with the prescribed node $\theta_{0,\ell+1} = 0$. The nodes $\theta_{j,\ell+1}$, $0 \leq j \leq \ell$, are the eigenvalues of the symmetric singular positive semidefinite tridiagonal matrix $\bar{C}_\ell\bar{C}_\ell^T$. Since ρ_j and σ_j are assumed to be positive for all j, the matrix $\bar{C}_\ell\bar{C}_\ell^T$ has positive subdiagonal entries and, therefore, distinct eigenvalues. It follows that $\theta_{j,\ell+1} > 0$ for $j \geq 1$. The limit (2.14) follows from (2.15) and this property of the eigenvalues.

The error formula for Gauss-Radau quadrature shows that $\phi(\mu) - \bar{\phi}_\ell(\mu)$ is of the same sign as the $(2\ell+1)$st derivate of $\psi_\mu(t)$ with respect to t. This derivative is negative for $\mu > 0$ and the left-hand side inequality in (2.13) follows. For further details, e.g., on the interpretation of $\bar{\phi}_\ell$ as a Gauss-Radau rule; see Calvetti et al. [4] or Golub and Meurant [11].

The right-hand side inequality in (2.13) can be shown by considering $\bar{\phi}_{\ell-1}(\mu)$ an ℓ-point Gauss-Radau quadrature formula associated with a Stieltjes integral that is defined by $\bar{\phi}_\ell(\mu)$; see Hanke [16] for details. ☐

Corollary 2.3. *Let $\omega_{0,\ell+1}$ be the same as in Theorem 2.2 and assume that $\ell \geq 1$ is sufficiently large, so that*

$$\omega_{0,\ell+1} < \eta^2\varepsilon^2. \tag{2.16}$$

Then there is a unique value μ_ℓ of the regularization parameter with $0 < \mu_\ell < \infty$, such that the associated approximate solution $\boldsymbol{x}_{\mu_\ell}^{(\ell)}$ of (1.5), determined by solving the Galerkin equations (2.8) and using (2.9) with $\mu = \mu_\ell$, satisfies

$$\|\boldsymbol{b} - A\boldsymbol{x}_{\mu_\ell}^{(\ell)}\| = \eta\varepsilon. \tag{2.17}$$

Proof. It follows from Theorem 2.2, (2.5), and (2.16) that

$$\bar{\phi}_\ell(0) = \|\boldsymbol{b}\|^2 > \eta^2\varepsilon^2 > \omega_{0,\ell+1} = \lim_{\mu\to\infty} \bar{\phi}_\ell(\mu).$$

Since $\bar{\phi}_\ell$ is continuous and decreasing, the equation $\bar{\phi}_\ell(\mu) = \eta^2\varepsilon^2$ has a unique solution μ_ℓ with $0 < \mu_\ell < \infty$. ☐

The value μ_ℓ of Corollary 2.3 can be computed conveniently by Newton's method applied to $\hat{\phi}_\ell(\mu) = \bar{\phi}_\ell(\mu) - \eta^2\varepsilon^2$. Because $\hat{\phi}_\ell(\mu)$ is decreasing and convex, cf. Theorem 2.2, Newton's method yields monotonic and quadratic convergence to μ_ℓ for any initial iterate smaller than μ_ℓ, e.g., zero.

Denote the minimal value of ℓ, such that the inequality (2.16) holds by ℓ_ε. This value is of interest, since it allows the computation of an approximate solution of the Tikhonov equation (1.5) that satisfies (2.17) with the minimal number of matrix-vector product evaluations. We now relate ℓ_ε to the LSQR method applied to the solution of (1.9).

Let the initial iterate be $\boldsymbol{x}^{(0)} = \boldsymbol{0}$. The ℓth iterate, $\boldsymbol{x}^{(\ell)}$, computed by LSQR then satisfies

$$\|A\boldsymbol{x}^{(\ell)} - \boldsymbol{b}\| = \min_{\boldsymbol{x}\in\mathbb{K}_\ell(A^TA, A^T\boldsymbol{b})} \|A\boldsymbol{x} - \boldsymbol{b}\|, \tag{2.18}$$

i.e., LSQR yields the residual vector

$$\boldsymbol{r}^{(\ell)} = \boldsymbol{b} - A\boldsymbol{x}^{(\ell)} \tag{2.19}$$

of smallest norm over the Krylov subspace (1.8).

Corollary 2.4. *Let $\ell \geq 1$ be an integer, such that the decompositions (2.6) with the stated properties exist. Then the residual vector (2.19) associated with the LSQR iterate $\boldsymbol{x}^{(\ell)}$ satisfies*

$$\|\boldsymbol{r}^{(\ell)}\|^2 = \omega_{0,\ell+1}, \tag{2.20}$$

where $\omega_{0,\ell+1}$ is the Gauss-Radau weight introduced in Theorem 2.2.

Proof. In view of (2.7), the iterate $\boldsymbol{x}^{(\ell)}$ can be expressed as

$$\boldsymbol{x}^{(\ell)} = V_\ell \boldsymbol{y}^{(\ell)}, \tag{2.21}$$

where $\boldsymbol{y}^{(\ell)}$ solves the least-squares problem

$$\min_{\boldsymbol{y}\in\mathbb{R}^\ell} \|AV_\ell \boldsymbol{y} - \boldsymbol{b}\|. \tag{2.22}$$

This minimization problem is equivalent to (2.18). The normal equations associated with (2.22) can be written as

$$\bar{C}_\ell^T \bar{C}_\ell \boldsymbol{y} = \|\boldsymbol{b}\| \bar{C}_\ell^T \boldsymbol{e}_1, \tag{2.23}$$

where we have use the partial Lanczos bidiagonalizations (2.6). Thus, $\boldsymbol{y}^{(\ell)}$ solves (2.23). Comparing (2.23) and (2.10) shows that $\lim_{\mu\to\infty} \boldsymbol{y}_\mu^{(\ell)} = \boldsymbol{y}^{(\ell)}$, and, hence, $\lim_{\mu\to\infty} \boldsymbol{x}_\mu^{(\ell)} = \boldsymbol{x}^{(\ell)}$. It follows that

$$\lim_{\mu\to\infty} \bar{\phi}_\ell(\mu) = \lim_{\mu\to\infty} \|\boldsymbol{b} - A\boldsymbol{x}_\mu^{(\ell)}\|^2 = \|\boldsymbol{r}^{(\ell)}\|^2.$$

Equation (2.20) now is a consequence of (2.14). □

Application of LSQR to the solution of linear discrete ill-posed problems has been investigated by Nemirovskii [22]: see also Hanke [14] for a discussion. Let ℓ_ε be the smallest index, such that the residual vector (2.19) with $\ell = \ell_\varepsilon$ satisfies

$$\|\boldsymbol{r}^{(\ell_\varepsilon)}\| \le \eta\varepsilon. \tag{2.24}$$

Note that ℓ_ε increases as ε decreases to zero. Nemirovskii [22] and Hanke [14] showed in a Hilbert space setting that $\boldsymbol{x}^{(\ell_\varepsilon)} \to \hat{\boldsymbol{x}}$ as $\varepsilon \searrow 0$.

Termination of the iterations at step ℓ_ε constitutes a regularization method, since the reduced system solved (2.23) is less ill-conditioned than (1.9). The termination criterion (2.24) therefore commonly is referred to as regularization by truncated iteration.

The greedy Tikhonov method amounts to carrying out ℓ_ε Lanczos bidiagonalization steps and then computing the value μ_{ℓ_ε} of the regularization parameter, such that (2.17) holds for $\ell = \ell_\varepsilon$ and $\mu_\ell = \mu_{\ell_\varepsilon}$. The computations are summarized in the following algorithm.

Algorithm 1 (Greedy Tikhonov Algorithm).
Input: $\boldsymbol{b} \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$, ε, η;
Output: *Approximate solution* $\tilde{\boldsymbol{x}}$, *regularization parameter* $\tilde{\mu}$, *number of Lanczos bidiagonalization steps* $\tilde{\ell} = \ell_\varepsilon$.

1. *Compute the decompositions (2.6) with* $\ell = \ell_\varepsilon$, *where* ℓ_ε, *the smallest integer such that (2.24) holds, also is to be determined.*
2. *Let* $\tilde{\ell} := \ell_\varepsilon$; $s := \|\boldsymbol{r}^{(\tilde{\ell})}\|$;
3. **if** $s = \eta^2\varepsilon^2$ **then** *compute the LSQR solution* $\boldsymbol{x}^{(\tilde{\ell})}$ *determined by (2.21) and (2.22) with* $\ell = \tilde{\ell}$. *Let* $\tilde{\boldsymbol{x}} := \boldsymbol{x}^{(\tilde{\ell})}$; $\tilde{\mu} := \infty$. *Exit.*
4. **if** $s < \eta^2\varepsilon^2$ **then** *determine* $\tilde{\mu} > 0$, *such that*

$$\bar{\phi}_{\tilde{\ell}}(\tilde{\mu}) = \eta^2\varepsilon^2. \tag{2.25}$$

Compute the approximate Tikhonov solution $\boldsymbol{x}_{\tilde{\mu}}^{(\tilde{\ell})}$ *determined by (2.9) and (2.11) with* $\ell = \tilde{\ell}$ *and* $\mu = \tilde{\mu}$. *Let* $\tilde{\boldsymbol{x}} := \boldsymbol{x}_{\tilde{\mu}}^{(\tilde{\ell})}$; *Exit.* □

Step 1 of the algorithm requires the evaluation of the norm of the residual errors (2.22) for increasing values of ℓ until the norm is bounded by $\eta\varepsilon$. The evaluation of these norms can be carried out easily by using the QR-factorization of the bidiagonal matrix $\bar{C}_\ell = Q_{\ell+1}\bar{R}_\ell$. Here $Q_{\ell+1} \in \mathbb{R}^{(\ell+1)\times(\ell+1)}$ is orthogonal and $\bar{R}_\ell \in \mathbb{R}^{(\ell+1)\times\ell}$ has an upper bidiagonal leading $\ell \times \ell$ submatrix and a vanishing last row. We have, using (2.6) and (2.22),

$$\|\boldsymbol{r}^{(\ell)}\| = \|\bar{C}_\ell \boldsymbol{y}^{(\ell)} - \|\boldsymbol{b}\|\boldsymbol{e}_1\| = \|\bar{R}_\ell \boldsymbol{y}^{(\ell)} - \|\boldsymbol{b}\|Q_{\ell+1}^T\boldsymbol{e}_1\| = \|\boldsymbol{b}\||\boldsymbol{e}_{\ell+1}^T Q_{\ell+1}^T \boldsymbol{e}_1|.$$

Since $\bar{C}_\ell$ is bidiagonal, the matrix $Q_{\ell+1}^T$ can be represented by a product of ℓ Givens rotations, and the evaluation of $\boldsymbol{e}_{\ell+1}^T Q_{\ell+1}^T \boldsymbol{e}_1$, given $\boldsymbol{e}_\ell^T Q_\ell^T \boldsymbol{e}_1$, can be carried out by the application of only one Givens rotation.

Algorithm 1 tacitly assumes that the existence of the decomposition (2.6) with $\ell = \ell_\varepsilon$. If this is not the case, then an invariant subspace of A has been found and the solution of the Tikhonov equation (1.5) lives in this subspace. It is quite straightforward to determine a value of the regularization parameter, such that (1.7) holds in this case. We will not dwell on the details of the computations, since the occurrence of an invariant subspace is rare. Computed examples in Section 4 illustrate that Algorithm 1 may yield more accurate approximations of $\hat{\boldsymbol{x}}$ than LSQR.

3. Enlarging the solution subspace. Algorithm 1 determines a regularized approximate solution $\tilde{\boldsymbol{x}}$ of (1.1) both by i) requiring $\tilde{\boldsymbol{x}}$ to live in the Krylov subspace $\mathbb{K}_{\ell_\varepsilon}(A^TA, A^T\boldsymbol{b})$ of typically fairly small dimension $\ell_\varepsilon \ll \min\{m,n\}$, and ii) determining a Tikhonov regularization parameter $\tilde{\mu} > 0$. The algorithm is designed to carry out the minimal number of bidiagonalization steps. However, for certain systems (1.1) application of $\tilde{\ell} > \ell_\varepsilon$ bidiagonalization steps may increase the quality of the computed approximation of $\hat{\boldsymbol{x}}$. This is illustrated by Example 4.5 in Section 4. An increase in $\tilde{\ell}$ can be achieved by replacing $\tilde{\ell} := \ell_\varepsilon$ by

$$\tilde{\ell} := \ell_\varepsilon + \ell_\delta \tag{3.1}$$

in Step 2 of Algorithm 1. An appropriate value of the regularization parameter $\tilde{\mu}$ is obtained by solving equation (2.25) in Step 4. We next discuss how the regularization parameter changes as the number of Lanczos bidiagonalization steps is increased.

Corollary 3.1. *Let $\ell > \ell_\varepsilon$ be an integer, such that the decompositions (2.6) with the stated properties exist. Let the regularization parameter values $\mu_{\ell-1}$, μ_ℓ, and μ_* satisfy*

$$\bar{\phi}_{\ell-1}(\mu_{\ell-1}) = \bar{\phi}_\ell(\mu_\ell) = \phi(\mu_*) = \eta^2\varepsilon^2.$$

Then

$$\mu_{\ell-1} > \mu_\ell > \mu_* > 0. \tag{3.2}$$

Proof. The first two inequalities follow from (2.13); the last inequality from Proposition 2.1. □

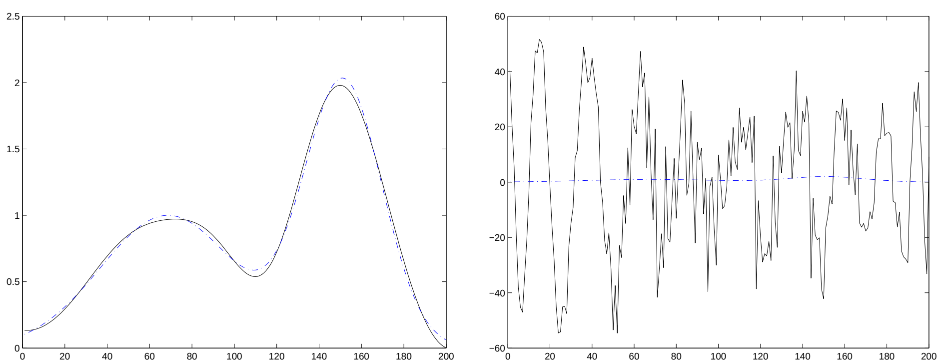
A smaller value of the Tikhonov regularization parameter yields more regularization than a larger value. When the solution is sought in a Krylov subspace (1.8) of small dimension ℓ, the original Tikhonov equations (1.5) are projected onto the Galerkin equation (2.8) of small size. This projection entails regularization; the fewer Lanczos bidiagonalization steps ℓ, the more regularization. Therefore the projected equations have to be regularized less than the Tikhonov equation (1.5); see, e.g., Natterer [21] for a discussion on regularization by projection.

4. Computed examples. All computations are carried out in Matlab with about 16 significant decimal digits. We compare the methods discussed in this paper with the Tikhonov regularization method described in [6]. The latter method determines the number of Lanczos bidiagonalization steps and a suitable value of the regularization parameter using the discrepancy principle, similarly as Algorithm 1, however, by different criteria. A comparison of the method in [6] with methods proposed by Frommer and Maas [10] and Golub and von Matt [12] reported in [6], shows the former method to require the fewest matrix-vector product evaluations. The examples of this section show Algorithm 1 to demand fewer matrix-vector product computations than the approach of [6]. All examples are based on code from the Matlab package Regularization Tools by Hansen [17].

In all examples, we let $\eta = 1 + 1 \cdot 10^{-14}$, i.e., we assume that the norm of the noise (1.4) is accurately known. The relative error in the right-hand side,

$$\Delta = \frac{\varepsilon}{\|\hat{\boldsymbol{b}}\|},$$

is referred to as the noise level. The matrices in all examples are of ill-determined rank.



(a) (b)

Fig. 4.1. *Example 4.1: The continuous black graph in (a) depicts the approximate solution* $\tilde{\boldsymbol{x}} = \boldsymbol{x}_{\tilde{\mu}}^{(16)}$ *determined by Algorithm 1, and the continuous black graph in (b) displays the approximate solution* $\boldsymbol{x}^{(16)}$ *computed by LSQR. The desired solution* $\hat{\boldsymbol{x}}$ *of the noise-free problem is displayed by blue dash-dotted graphs in both (a) and (b).*

Example 4.1. We consider the Fredholm integral equation of the first kind,

$$\int_{-\pi/2}^{\pi/2} \kappa(\sigma,\tau)x(\sigma)d\sigma = b(\tau), \qquad -\frac{\pi}{2} \le \tau \le \frac{\pi}{2}, \tag{4.1}$$

where

$$\kappa(\sigma,\tau) = (\cos(\sigma)+\cos(\tau))^2 \left(\frac{\sin(\xi)}{\xi}\right)^2, \qquad \xi = \pi(\sin(\sigma)+\sin(\tau)),$$

and the right-hand side $b(\tau)$ is chosen so that the solution $x(\sigma)$ is a sum of two Gaussian functions. This integral equation is discussed by Shaw [28]. We use the code `shaw` from [17] to discretize (4.1) by a quadrature rule with 200 nodes. This yields the matrix $A \in \mathbb{R}^{200\times 200}$ and right-hand side $\hat{\boldsymbol{b}} \in \mathbb{R}^{200}$. We generate a "noise vector" $\boldsymbol{e} \in \mathbb{R}^{200}$ of noise level $\Delta = 1 \cdot 10^{-3}$ with normally distributed zero-mean components, such that $\boldsymbol{e}$ is orthogonal to the eigenspace of $A^T A$ associated with the 20 largest eigenvalues; $\boldsymbol{e}$ is numerically in the null space of A.

Algorithm 1 carries out 16 Lanczos bidiagonalization steps to determine the regularization parameter $\tilde{\mu}$ and the approximate solution $\tilde{\boldsymbol{x}} = \boldsymbol{x}_{\tilde{\mu}}^{(16)}$ of (1.5), shown in Figure 4.1(a), with error $\|\boldsymbol{x}_{\tilde{\mu}}^{(16)} - \hat{\boldsymbol{x}}\| = 4.8 \cdot 10^{-1}$. LSQR yields the iterate $\boldsymbol{x}^{(16)}$, displayed in Figure 4.1(b), with error $\|\boldsymbol{x}^{(16)} - \hat{\boldsymbol{x}}\| = 3.7 \cdot 10^{2}$. We used the Matlab implementation lsqr in [17] with modified Gram-Schmidt reorthogonalization (parameter reorth$= 1$).

While the displayed behavior of LSQR is not typical, it nevertheless illustrates that LSQR may fail in a black-box setting when one has little control of the properties of the noise. Algorithm 1 is seen to provide a meaningful approximation of $\hat{\boldsymbol{x}}$.

Generally, LSQR is applied without reorthogonalization. The code lsqr without reorthogonalization and the stopping criterion (2.24) yields the iterate $\boldsymbol{x}^{(158)}$ with error $\|\boldsymbol{x}^{(158)} - \hat{\boldsymbol{x}}\| = 1.8 \cdot 10^{1}$. Hence, loss of orthogonality may increase the computational work significantly and give a poor approximate solution. We apply Algorithm 1 and lsqr with Gram-Schmidt reorthogonalization in all examples.

We remark that LSQR with Gram-Schmidt reorthogonalization is able to determine a quite accurate approximation of $\hat{\boldsymbol{x}}$; the difficulty is to select the appropriate iterate. For instance, the heuristic stopping criterion based on common behavior of semi-convergent series described in [20], yields the optimal iterate $\boldsymbol{x}^{(13)}$ with error $\|\boldsymbol{x}^{(13)} - \hat{\boldsymbol{x}}\| = 5.2 \cdot 10^{-2}$. $\square$

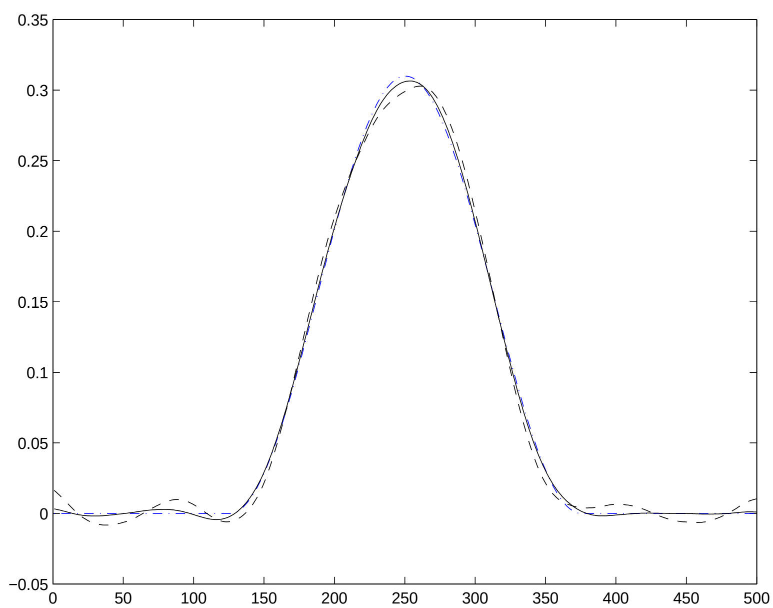


FIG. 4.2. *Example 4.2: Solution $\hat{\boldsymbol{x}}$ of the error-free linear system (1.3) (blue dash-dotted graph), approximate solution $\tilde{\boldsymbol{x}} = \boldsymbol{x}_{\tilde{\mu}}^{(8)}$ determined by Algorithm 1 (black continuous graph), and LSQR solution $\boldsymbol{x}^{(8)}$ (black dashed graph).*

Example 4.2. Consider the Fredholm integral equation of the first kind

$$\int_{-6}^{6} \kappa(\tau,\sigma)x(\sigma)d\sigma = b(\tau), \qquad -6 \leq \tau \leq 6, \tag{4.2}$$

discussed by Phillips [27]. Its solution, kernel, and right-hand side are given by

$$\begin{aligned} x(\sigma) &= \begin{cases} 1+\cos(\frac{\pi}{3}\sigma), & \text{if } |\sigma|<3, \\ 0, & \text{otherwise,} \end{cases} \\ \kappa(\tau,\sigma) &= x(\tau-\sigma), \\ b(\tau) &= (6-|\tau|)(1+\frac{1}{2}\cos(\frac{\pi}{3}\tau)) + \frac{9}{2\pi}\sin(\frac{\pi}{3}|\tau|). \end{aligned}$$

We use the Matlab code phillips from [17] to discretize (4.1) by a Galerkin method with orthonormal box functions as test and trial functions to obtain the matrix $A \in \mathbb{R}^{500\times 500}$ and right-hand side $\hat{\boldsymbol{b}} \in \mathbb{R}^{500}$ of (1.3). A noise vector $\boldsymbol{e} \in \mathbb{R}^{500}$ of noise level $\Delta = 1\cdot 10^{-2}$ with normally distributed zero-mean components is added to $\hat{\boldsymbol{b}}$ to obtain $\boldsymbol{b}$, cf. (1.2). Algorithm 1 yields the approximate solution $\tilde{\boldsymbol{x}} = \boldsymbol{x}_{\tilde{\mu}}^{(8)}$ of (1.1) with $\|\boldsymbol{x}_{\tilde{\mu}}^{(8)} - \hat{\boldsymbol{x}}\| = 5.1\cdot 10^{-2}$ and $\tilde{\mu} = 4.7\cdot 10^{1}$. The corresponding LSQR iterate, $\boldsymbol{x}^{(8)}$, has a larger error; $\|\boldsymbol{x}^{(8)} - \hat{\boldsymbol{x}}\| = 1.6\cdot 10^{-1}$. Figure 4.2 displays the vectors $\boldsymbol{x}_{\tilde{\mu}}^{(8)}$, $\boldsymbol{x}^{(8)}$, and $\hat{\boldsymbol{x}}$.

We also compare Algorithm 1 with the method described in [6], which requires 14 Lanczos bidiagonalization steps to determine an approximate solution $\breve{\boldsymbol{x}}$ of (1.1) that satisfies the discrepancy principle (1.7). The error in $\breve{\boldsymbol{x}}$ is $\|\breve{\boldsymbol{x}} - \hat{\boldsymbol{x}}\| = 5.2\cdot 10^{-2}$. Thus, $\boldsymbol{x}_{\breve{\mu}}^{(8)}$ and $\breve{\boldsymbol{x}}$ furnish approximations of $\hat{\boldsymbol{x}}$ of about the same accuracy, but the computation of $\boldsymbol{x}_{\tilde{\mu}}^{(8)}$ requires only about half the number of matrix-vector product evaluations. □

Method	Bidiag. Steps	Reg. Param.	Error in Comp. Solution
Algorithm 1	5	$2.6\cdot 10^{2}$	$2.0\cdot 10^{0}$
Algorithm 1, $\ell_\delta = 1$	6	$2.5\cdot 10^{2}$	$2.0\cdot 10^{0}$
Algorithm 1, $\ell_\delta = 2$	7	$2.5\cdot 10^{2}$	$2.0\cdot 10^{0}$
[6]	9	$2.5\cdot 10^{3}$	$2.0\cdot 10^{0}$

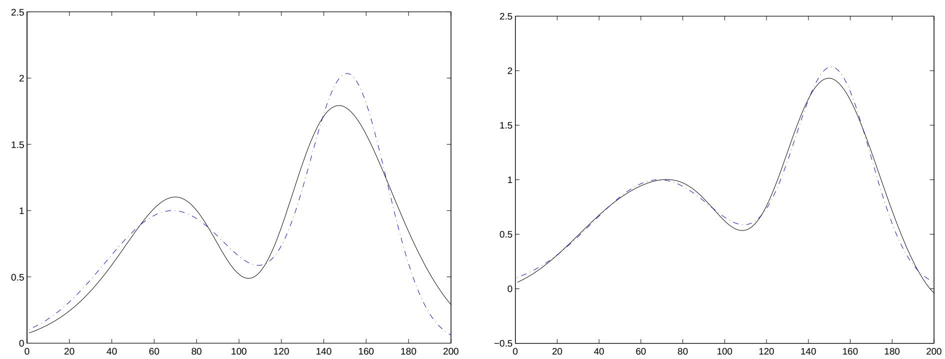
TABLE 4.1

Example 4.3: Comparison of Algorithm 1, Algorithm 1 with modification (3.1) of Step 2, and the method described in [6], applied to the solution of a discretization of (4.1) with noise level $\Delta = 1\cdot 10^{-2}$.

Method	Bidiag. Steps	Reg. Param.	Error in Comp. Solution
Algorithm 1	7	$9.0\cdot 10^{3}$	$7.3\cdot 10^{-1}$
Algorithm 1, $\ell_\delta = 1$	8	$9.0\cdot 10^{3}$	$7.4\cdot 10^{-1}$
Algorithm 1, $\ell_\delta = 2$	9	$9.0\cdot 10^{3}$	$7.4\cdot 10^{-1}$
[6]	10	$9.0\cdot 10^{3}$	$7.4\cdot 10^{-1}$

TABLE 4.2

Example 4.3: Comparison of Algorithm 1, Algorithm 1 with modification (3.1) of Step 2, and the method described in [6], applied to the solution of a discretization of (4.1) with noise level $\Delta = 1\cdot 10^{-3}$.



(a) (b)

FIG. 4.3. *Example 4.3: (a) The continuous black graph shows the approximate solution* $\tilde{\boldsymbol{x}} = \boldsymbol{x}_{\tilde{\mu}}^{(5)}$ *determined by Algorithm 1 for noise level* $\Delta = 1 \cdot 10^{-2}$. *(b) The continuous black graph depicts the approximate solution* $\tilde{\boldsymbol{x}} = \boldsymbol{x}_{\tilde{\mu}}^{(7)}$ *determined by Algorithm 1 for noise level* $\Delta = 1 \cdot 10^{-3}$. *The blue dash-dotted graphs in both figures display the desired solution* $\hat{\boldsymbol{x}}$.

Example 4.3. This example presents some more computations with the integral equation (4.1). Discretization is carried out in the same way as in Example 4.1, but the noise vector $\boldsymbol{e}$ is not in an invariant subspace of A; its components are normally distributed with zero mean. For Tables 4.1 and 4.2, $\boldsymbol{e}$ is normalized to yield the noise levels $1 \cdot 10^{-2}$ and $1 \cdot 10^{-3}$, respectively. In the tables "Algorithm 1, $\ell_\delta = j$" denotes Algorithm 1 with the modification $\tilde{\ell} := \ell_\varepsilon + j$ in Step 2; cf. (3.1). Thus, the algorithm carries out $\tilde{\ell} = \ell_\varepsilon + j$ bidiagonalization steps. In the present example, this modification decreases the value of the regularization parameter very little, and does not improve the accuracy in the computed approximate solutions. Finally, Tables 4.1 and 4.2 also show results obtained with the method discussed in [6]. In summary, all the Tikhonov regularization methods yield about the same accuracy, with the greedy Tikhonov method requiring the least matrix-vector product evaluations. Also, LSQR furnishes approximations of $\hat{\boldsymbol{x}}$ of about the same accuracy.

Figures 4.3(a) and (b) show the solutions $\tilde{\boldsymbol{x}}$ computed by Algorithm 1 for the noise levels $1 \cdot 10^{-2}$ and $1 \cdot 10^{-3}$, as well as the desired solution $\hat{\boldsymbol{x}}$ of the noise-free problem. □

Method	Bidiag. Steps	Reg. Param.	Error in Comp. Solution
Algorithm 1	3	$2.8 \cdot 10^3$	$2.1 \cdot 10^{-1}$
Algorithm 1, $\ell_\delta = 1$	4	$2.7 \cdot 10^3$	$2.1 \cdot 10^{-1}$
Algorithm 1, $\ell_\delta = 2$	5	$2.7 \cdot 10^3$	$2.1 \cdot 10^{-1}$
[6]	6	$2.7 \cdot 10^3$	$2.1 \cdot 10^{-1}$

TABLE 4.3

Example 4.4: Comparison of Algorithm 1, Algorithm 1 with modification (3.1) of Step 2, and the method described in [6], applied to the solution of a discretization of (4.3) with noise level $\Delta = 1 \cdot 10^{-2}$.

Example 4.4. The Fredholm integral equation of the first kind,

$$\int_0^{\pi/2} \kappa(\sigma,\tau)x(\sigma)d\sigma = b(\tau), \qquad 0 \leq \tau \leq \pi, \tag{4.3}$$

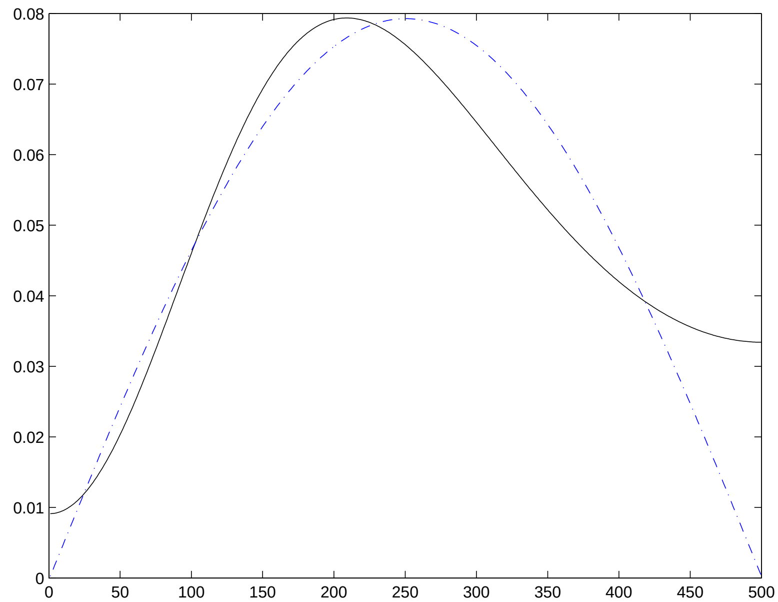


FIG. 4.4. *Example 4.4: Solution $\hat{\boldsymbol{x}}$ of the error-free linear system (1.3) (blue dash-dotted graph) and approximate solution $\tilde{\boldsymbol{x}} = \boldsymbol{x}_{\tilde{\mu}}^{(3)}$ determined by Algorithm 1 (black continuous graph).*

with $\kappa(\sigma,\tau) = \exp(\sigma\cos(\tau))$, $b(\tau) = 2\sinh(\tau)/\tau$, and solution $x(\tau) = \sin(\tau)$, is discussed by Baart [1]. We use the Matlab code baart from [17] to discretize (4.3) by a Galerkin method with 500 orthonormal box functions as test and trial functions. This yields the linear system (1.3) with $A \in \mathbb{R}^{500\times500}$ and the right-hand side vector $\hat{\boldsymbol{b}} \in \mathbb{R}^{500}$. An error vector $\boldsymbol{e} \in \mathbb{R}^{500}$ with normally distributed zero-mean components and noise level $1 \cdot 10^{-2}$ is added to $\hat{\boldsymbol{b}}$ to yield the right-hand side $\boldsymbol{b}$ in the system (1.1) to be solved. The computations are reported in Table 4.3, which shows all the Tikhonov regularization methods to give about the same accuracy, with the greedy Tikhonov method requiring the fewest matrix-vector product evaluations. The accuracy is improved insignificantly by the modification (3.1) with $\ell_\delta > 0$ of Algorithm 1. LSQR yields about the same accuracy.

Figure 4.4 displays the approximate solution $\tilde{\boldsymbol{x}}$ determined by Algorithm 1 and the solution $\hat{\boldsymbol{x}}$. Reduction of the noise level or application of a suitable regularization operator, different from the identity, would give more accurate approximations of $\hat{\boldsymbol{x}}$. Computations with finite difference-based regularization operators can be found in [7]. $\square$

Example 4.5. We modify the right-hand side of the Fredholm integral equation (4.2) to determine the solution

$$x(\sigma) = -\sin(\frac{\pi}{2}\sigma).$$

Discretization is carried out like in Example 4.2. This yields the matrix $A \in \mathbb{R}^{500\times500}$ and the error-free right-hand side $\hat{\boldsymbol{b}} \in \mathbb{R}^{500}$. The "noise vector" $\boldsymbol{e} \in \mathbb{R}^{500}$ with normally distributed zero-mean components of noise level $\Delta = 1 \cdot 10^{-3}$ is added to $\hat{\boldsymbol{b}}$ to give the right-hand side $\boldsymbol{b}$ of (1.1).

Application of Algorithm 1 to (1.1), with the iterations terminated as soon as the

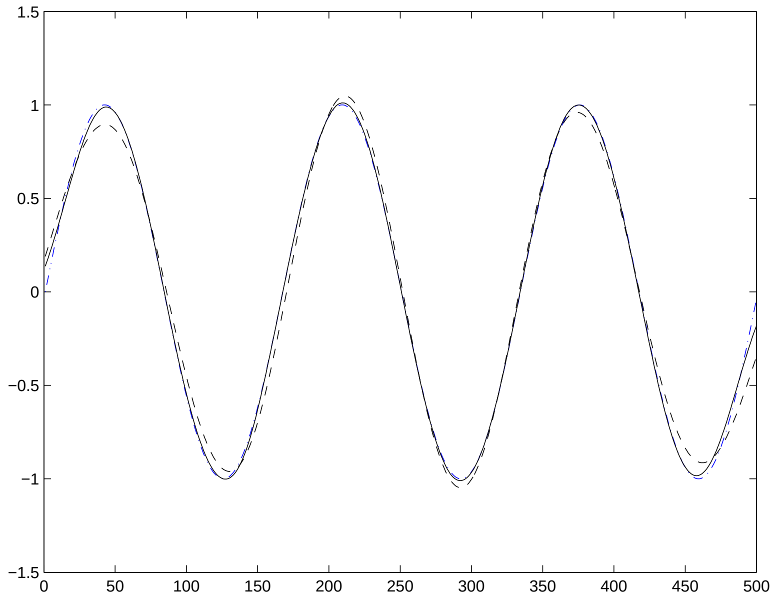


FIG. 4.5. *Example 4.5: Solution $\hat{\boldsymbol{x}}$ of the error-free linear system (blue dash-dotted graph), the approximate solution $\boldsymbol{x}_{\tilde{\mu}}^{(13)}$ determined by Algorithm 1 with the modification (3.1) with $\ell_\delta = 6$ (black continuous graph), and the approximate solutions $\boldsymbol{x}_{\tilde{\mu}}^{(7)}$ computed by Algorithm 1 (black dashed graph), and $\boldsymbol{x}^{(7)}$ determined by LSQR (black dashed graph). The graphs for $\boldsymbol{x}_{\tilde{\mu}}^{(7)}$ and $\boldsymbol{x}^{(7)}$ cannot be distinguished.*

residual vector satisfies (2.24) with $\eta = 1.1$, yields the approximate solution $\tilde{\boldsymbol{x}} = \boldsymbol{x}_{\tilde{\mu}}^{(7)}$ with error $\|\boldsymbol{x}_{\tilde{\mu}}^{(7)} - \hat{\boldsymbol{x}}\| = 1.6$. Similarly, LSQR determines the iterate $\boldsymbol{x}^{(7)}$ with error $\|\boldsymbol{x}^{(7)} - \hat{\boldsymbol{x}}\| = 1.6$. Algorithm 1 with Step 2 modified according to (3.1) with $\ell_\delta = 6$ gives the approximate solution $\tilde{\boldsymbol{x}} = \boldsymbol{x}_{\tilde{\mu}}^{(13)}$ with error $\|\boldsymbol{x}_{\tilde{\mu}}^{(13)} - \hat{\boldsymbol{x}}\| = 4.9 \cdot 10^{-1}$. Thus, for this example it is beneficial to choose $\ell_\delta > 0$.

Figure 4.5 displays the exact solution $\hat{\boldsymbol{x}}$ of the error-free system (1.3)R, and the computed solutions $\tilde{\boldsymbol{x}} = \boldsymbol{x}_{\tilde{\mu}}^{(13)}$, $\tilde{\boldsymbol{x}} = \boldsymbol{x}_{\tilde{\mu}}^{(7)}$, and $\boldsymbol{x}^{(7)}$. The graphs for the latter computed approximate solutions cannot be distinguished. □

5. Conclusions. A unified approach to Tikhonov regularization and LSQR has been presented. For many problems all the methods in our comparison yield about the same accuracy, with LSQR and Algorithm 1 requiring the least number of matrix-vector product evaluations. Algorithm 1 can in some instances provide meaningful approximate solutions when LSQR does not. Moreover, Algorithm 1 allows the number of Lanczos bidiagonalization steps to be increased. For some problems, such an increase improves the accuracy in the computed approximate solution.

REFERENCES

[1] M. L. Baart, *The use of auto-correlation for pseudo-rank determination in noisy ill-conditioned least-squares problems*, IMA J. Numer. Anal., 2 (1982), pp. 241–247.
[2] Å. Björck, *A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations*, BIT, 28 (1988), pp. 659–670.
[3] Å. Björck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, 1996.

[4] D. Calvetti, G. H. Golub, and L. Reichel, *Estimation of the L-curve via Lanczos bidiagonalization*, BIT, 39 (1999), pp. 603–619.
[5] D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari, *Tikhonov regularization and the L-curve for large, discrete ill-posed problems*, J. Comput. Appl. Math., 123 (2000), pp. 423–446.
[6] D. Calvetti and L. Reichel, *Tikhonov regularization of large linear problems*, BIT, 43 (2003), pp. 263–283.
[7] D. Calvetti, L. Reichel, and A. Shuibi, *Invertible smoothing preconditioners for linear discrete ill-posed problems*, Appl. Numer. Math., 54 (2005), pp. 135–149.
[8] L. Eldén, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT, 17 (1977), pp. 134–145.
[9] H. W. Engl, M. Hanke, and A. Neubauer, Regularization of Inverse Problems, Kluwer, Dordrecht, 1996.
[10] A. Frommer and P. Maass, *Fast CG-based methods for Tikhonov-Phillips regularization*, SIAM J. Sci. Comput., 20 (1999), pp. 1831–1850.
[11] G. H. Golub and G. Meurant, *Matrices, moments and quadrature*, in Numerical Analysis 1993, eds. D. F. Griffiths and G. A. Watson, Longman, Essex, 1994, pp. 105–156.
[12] G. H. Golub and U. von Matt, *Tikhonov regularization for large scale problems*, in Workshop on Scientific Computing, eds. G. H. Golub, S. H. Lui, F. Luk, and R. Plemmons, Springer, New York, 1997, pp. 3–26.
[13] C. W. Groetsch, The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind, Pitman, Boston, 1984.
[14] M. Hanke, Conjugate Gradient Type Methods for Ill-Posed Problems, Longman Scientific and Technical, Essex, England, 1995.
[15] M. Hanke, *On Lanczos based methods for the regularization of discrete ill-posed problems*, BIT, 41 (2001), pp. 1008–1018.
[16] M. Hanke, *A note on Tikhonov regularization of large linear problems*, BIT, 43 (2003), pp. 449–451.
[17] P. C. Hansen, *Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algor., 6 (1994), pp. 1–35. Software is available in Netlib at `http://www.netlib.org`.
[18] P. C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia, 1998.
[19] M. E. Kilmer and D. P. O'Leary, *Choosing regularization parameters in iterative methods for ill-posed problems*, SIAM J. Matrix Anal. Appl., 22 (2001), pp. 1204–1221.
[20] S. Morigi, L. Reichel, F. Sgallari, and F. Zama, *Iterative methods for ill-posed problems and semiconvergent sequences*, J. Comput. Appl. Math., 193 (2006), pp. 157–167.
[21] F. Natterer, *Regularisierung schlecht gestellter Probleme durch Projektionsverfahren*, Numer. Math., 28 (1977), pp. 329–341.
[22] A. S. Nemirovskii, *The regularization properties of the adjoint gradient method in ill-posed problems*, USSR Comput. Math. Math. Phys., 26 (1986), pp. 7–16.
[23] M. K. Ng, Iterative Methods for Toeplitz Systems, Oxford Univ. Press, Oxford, 2004.
[24] D. P. O'Leary and J. A. Simmons, A bidiagonalization-regularization procedure for large-scale discretizations of ill-posed problems, SIAM J. Sci. Statist. Comput., 2 (1981), pp. 474–489.
[25] C. C. Paige and M. A. Saunders, *LSQR: An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software, 8 (1982), pp. 43–71.
[26] C. C. Paige and M. A. Saunders, *Algorithm 583 LSQR: Sparse linear equations and least squares problems*, ACM Trans. Math. Software, 8 (1982), pp. 195–209.
[27] D. L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, J. ACM, 9 (1962), pp. 84–97.
[28] C. B. Shaw, Jr., *Improvements of the resolution of an instrument by numerical solution of an integral equation*, J. Math. Anal. Appl., 37 (1972), pp. 83–112.