SIAM J. SCI. COMPUT.
Vol. 26, No. 1, pp. 224–239
© 2004 Society for Industrial and Applied Mathematics

TIKHONOV REGULARIZATION WITH A SOLUTION CONSTRAINT*

DANIELA CALVETTI† AND LOTHAR REICHEL‡

Abstract. Many numerical methods for the solution of linear ill-posed problems apply Tikhonov regularization. This paper presents a modification of a numerical method proposed by Golub and von Matt for quadratically constrained least-squares problems and applies it to Tikhonov regularization of large-scale linear discrete ill-posed problems. The method is based on partial Lanczos bidiagonalization and Gauss quadrature. Computed examples illustrating its performance are presented.

Key words. ill-posed problem, regularization, Gauss quadrature

AMS subject classifications. 65F10, 65F22, 65R30

DOI. 10.1137/S1064827502412280

1. Introduction. The solution of large-scale linear discrete ill-posed problems has been the topic of numerous investigations over the past decade. Linear discrete ill-posed problems are 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 matrix of ill-determined rank. The singular values of the matrix accumulate at the origin, and this makes the matrix severely ill-conditioned and possibly singular. Linear discrete ill-posed problems arise from the discretization of ill-posed problems, such as Fredholm integral equations of the first kind with a smooth kernel.

Many linear discrete ill-posed problems that arise in applications have a right-hand side $\boldsymbol{b}$ that is contaminated by an error $\boldsymbol{e} \in \mathbb{R}^m$, i.e., $\boldsymbol{b} = \hat{\boldsymbol{b}} + \boldsymbol{e}$, where $\hat{\boldsymbol{b}}$ denotes the unavailable error-free right-hand side. The error $\boldsymbol{e}$ may stem from measurement or discretization errors. The error-free linear system of equations

$$A\boldsymbol{x} = \hat{\boldsymbol{b}} \tag{1.2}$$

is assumed to be consistent.

Let $A^\dagger$ denote the Moore–Penrose pseudoinverse of the matrix A. We would like to determine the least-squares solution of minimal Euclidean norm of the unavailable linear system of equations (1.2), given by

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

by computing an approximate solution of the available linear system of equations (1.1). Note that, in general, the least-squares solution of minimal Euclidean norm of (1.1), given by

$$\boldsymbol{x}_0 := A^\dagger \boldsymbol{b}, \tag{1.4}$$

*Received by the editors July 29, 2002; accepted for publication (in revised form) February 2, 2004; published electronically November 9, 2004.
http://www.siam.org/journals/sisc/26-1/41228.html
†Department of Mathematics, Case Western Reserve University, Cleveland, OH 44106 (dxc57@po.cwru.edu). This author's research was supported in part by NSF grant DMS-0107841.
‡Department of Mathematical Sciences, Kent State University, Kent, OH 44242 (reichel@math.kent.edu). This author's research was supported in part by NSF grant DMS-0107858.

is not a meaningful approximation of $\hat{\boldsymbol{x}}$ due to the error $\boldsymbol{e}$ and the ill-conditioning of A; typically

$$\|\hat{\boldsymbol{x}}\| \ll \|\boldsymbol{x}_0\| \approx \|A^\dagger \boldsymbol{e}\|; \tag{1.5}$$

i.e., the norm of $\boldsymbol{x}_0$ is essentially determined by the norm of propagated and amplified noise. Throughout this paper, $\|\cdot\|$ denotes the Euclidean vector norm or the associated induced matrix norm.

Several techniques have been proposed in the literature to modify the linear system (1.1) so that the sensitivity of the solution to the error $\boldsymbol{e}$ is reduced. These techniques are commonly referred to as regularization methods, among which Tikhonov regularization is one of the most popular. In its simplest form, Tikhonov regularization replaces the solution of the linear system (1.1) by the minimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \{\|\boldsymbol{b} - A\boldsymbol{x}\|^2 + \mu\|\boldsymbol{x}\|^2\} \tag{1.6}$$

for a suitable positive value of the regularization parameter μ. We remark that for any $\mu > 0$, the minimization problem (1.6) has the unique solution

$$\boldsymbol{x}_\mu := (A^T A + \mu I)^{-1} A^T \boldsymbol{b}. \tag{1.7}$$

Moreover, $\boldsymbol{x}_0 = \lim_{\mu \searrow 0} \boldsymbol{x}_\mu$.

The choice of a suitable value of μ is an essential part of Tikhonov regularization. The value of μ determines how sensitive the solution $\boldsymbol{x}_\mu$ of (1.6) is to the error $\boldsymbol{e}$ and how close $\boldsymbol{x}_\mu$ is to the solution $\hat{\boldsymbol{x}}$ of (1.2). The condition number of the matrix $A^T A + \mu I$, defined by $\kappa(A^T A + \mu I) := \|A^T A + \mu I\|\|(A^T A + \mu I)^{-1}\|$, is a decreasing function of μ, and therefore the solution $\boldsymbol{x}_\mu$ is generally more sensitive to the error $\boldsymbol{e}$, the smaller $\mu > 0$ is.

When the norm of the error is known a priori, μ may be chosen so that

$$\|\boldsymbol{b} - A\boldsymbol{x}_\mu\| = \|\boldsymbol{e}\|. \tag{1.8}$$

This criterion for choosing the value of the regularization parameter is usually referred to as the discrepancy principle. Numerical methods for large-scale problems using the discrepancy principle to determine μ are presented in [6, 9]. Other criteria for choosing a suitable value of μ must be employed when the norm of $\boldsymbol{e}$ is not known, such as criteria based on the L-curve and generalized cross validation. Applications of the latter to large-scale linear discrete ill-posed problems are discussed in [3, 4, 9, 13].

In this paper, we assume that the quantity

$$\Delta := \|\hat{\boldsymbol{x}}\| \tag{1.9}$$

is available, where $\hat{\boldsymbol{x}}$ is defined by (1.3), and we require that the norm of the computed approximate solution of (1.1) is bounded by Δ. Thus, we replace the linear system (1.1) by the constrained minimization problem

$$\min_{\|\boldsymbol{x}\|\le\Delta} \|\boldsymbol{b} - A\boldsymbol{x}\|. \tag{1.10}$$

This approach to regularization has recently been discussed by Rojas [17] and Rojas and Sorensen [19]. In the optimization literature, minimization problems of the form (1.10) arise in the context of trust-region methods. The following proposition characterizes the solutions of (1.10). In view of (1.5), we may assume that $\Delta < \|\boldsymbol{x}_0\|$. The

proposition shows that the constrained minimization problem (1.10) is equivalent to Tikhonov regularization.

PROPOSITION 1.1. *Assume that* $\Delta < \|\boldsymbol{x}_0\|$. *Then the constrained minimization problem* (1.10) *has a unique solution* $\boldsymbol{x}_{\mu_\Delta}$ *of the form* (1.7). *The value* μ_Δ *of the parameter* μ *is positive and such that* $\|\boldsymbol{x}_{\mu_\Delta}\| = \Delta$.

Proof. A proof for the case when the minimization in (1.10) is over the set $\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| = \Delta\}$ is presented by Golub and von Matt [8]. It is easy to establish that there is a solution of (1.10) of norm Δ. □

Rojas [17] and Rojas and Sorensen [19] proposed to solve the constrained minimization problem (1.10) by the LSTRS (large-scale trust-region subproblem) method. This is a method for the solution of large-scale quadratic minimization problems that arise in the context of trust-region methods for optimization. The LSTRS method, developed by Rojas, Santos, and Sorensen [18], solves these quadratic minimization problems by recasting them as parameterized eigenvalue problems whose solutions are computed by the implicitly restarted Arnoldi method. The LSTRS method can be applied to more difficult quadratic minimization problems than (1.10); in particular the method can handle the so-called hard case, which can arise when the Hessian of the quadratic minimization problem is indefinite; see [18, 20] for insightful discussions on this situation.

Under the conditions of Proposition 1.1, which are reasonable when solving linear discrete ill-posed problems with a perturbed right-hand side, the hard case cannot occur. We therefore propose that the minimization problem (1.10) be solved by a modification of the method developed by Golub and von Matt [8]. The purpose of the modification is to make the method better suited to the situation when A is of ill-determined rank.

This paper is organized as follows. Section 2 reviews results on the connection between partial Lanczos bidiagonalization and Gauss quadrature. Related results are discussed in [3, 4, 5, 8, 9, 10]. This connection allows us to inexpensively compute upper and lower bounds for $\|\boldsymbol{x}_\mu\|$. Section 3 discusses how to compute an approximation of the largest solution of the nonlinear equation

$$h(\mu) = 0, \qquad h(\mu) := \|\boldsymbol{x}_\mu\|^2 - \Delta^2, \tag{1.11}$$

using inexpensively computable bounds for $\|\boldsymbol{x}_\mu\|$. A few computed examples are presented in section 4. They illustrate the behavior of the method of the present paper, compare it with the method described by Rojas and Sorensen [19], and seek to shed light on when it is appropriate to compute an approximate solution of (1.1) by solving the constrained minimization problem with Δ defined by (1.9).

2. Lanczos bidiagonalization, Gauss quadrature, and matrix functionals. This section reviews results on the connection between partial Lanczos bidiagonalization, Gauss quadrature rules, and the computation of upper and lower bounds for the function

$$\phi(\mu) := \|\boldsymbol{x}_\mu\|^2, \qquad \mu \geq 0. \tag{2.1}$$

Related results can be found in [3, 4, 5, 8, 9, 10].

PROPOSITION 2.1. *Assume that* $A^\dagger \boldsymbol{b} \neq \boldsymbol{0}$. *The function* (2.1) *can be expressed as*

$$\phi(\mu) = \boldsymbol{b}^T A (A^T A + \mu I)^{-2} A^T \boldsymbol{b}, \qquad \mu > 0, \tag{2.2}$$

which shows that $\phi(\mu)$ is strictly decreasing and convex for $\mu > 0$. Moreover, the equation

$$\phi(\mu) = \tau \tag{2.3}$$

has a unique solution μ, such that $0 < \mu < \infty$, for any τ that satisfies $0 < \tau < \|A^\dagger \boldsymbol{b}\|^2$.

Proof. Substituting (1.7) into (2.1) yields (2.2). The stated properties of $\phi(\mu)$ and (2.3) can be shown by substituting the singular value decomposition of A into (2.2). □

If a fairly accurate approximation of $\|\hat{\boldsymbol{x}}\|$ is available, then this approximation is, in view of (1.5), generally much smaller than $\|A^\dagger \boldsymbol{b}\|$. Therefore, (2.3) can be solved with τ being the square of the norm of the available estimate of $\|\hat{\boldsymbol{x}}\|$ for almost all linear discrete ill-posed problems of practical significance.

Our numerical method is based on partial Lanczos bidiagonalization of A with initial vector $\boldsymbol{b}$. Application of $\ell \leq \min\{m, n\}$ steps of Lanczos bidiagonalization yields the decompositions

$$AV_\ell = U_{\ell+1} C_{\ell+1,\ell}, \qquad A^T U_\ell = V_\ell C_\ell^T, \qquad \boldsymbol{b} = \sigma_1 U_\ell \boldsymbol{e}_1, \tag{2.4}$$

where $V_\ell \in \mathbb{R}^{n\times\ell}$ and $U_{\ell+1} \in \mathbb{R}^{m\times(\ell+1)}$ satisfy $V_\ell^T V_\ell = I_\ell$ and $U_{\ell+1}^T U_{\ell+1} = I_{\ell+1}$. Further, $\sigma_1 = \|\boldsymbol{b}\|$. Throughout this paper, I_j denotes the $j \times j$ identity matrix and $\boldsymbol{e}_j$ is the jth axis vector of appropriate dimension. The matrix $C_{\ell+1,\ell}$ is bidiagonal,

$$C_{\ell+1,\ell} = \begin{bmatrix} \rho_1 & & & 0 \\ \sigma_2 & \rho_2 & & \\ & \ddots & \ddots & \\ & & \sigma_\ell & \rho_\ell \\ 0 & & & \sigma_{\ell+1} \end{bmatrix} \in \mathbb{R}^{(\ell+1)\times\ell}, \tag{2.5}$$

with positive subdiagonal entries σ_j, $2 \leq j \leq \ell+1$, and C_ℓ denotes the leading $\ell \times \ell$ principal submatrix of $C_{\ell+1,\ell}$. The evaluation of the partial Lanczos bidiagonalization (2.4) requires ℓ matrix-vector product evaluations with both the matrices A and A^T; see, e.g., Paige and Saunders for details [14].

We tacitly assume that the number of Lanczos bidiagonalization steps ℓ is chosen small enough so that the decompositions (2.4) with the stated properties exist. Generically, $\sigma_{\ell+1}$ is positive; however, if $\sigma_{\ell+1}$ vanishes, then the development simplifies. The latter case is discussed at the end of this section.

Let $C_{\ell+1,\ell} = Q_{\ell+1,\ell} R_\ell$ be the QR factorization of $C_{\ell+1,\ell}$; i.e., $Q_{\ell+1,\ell} \in \mathbb{R}^{(\ell+1)\times\ell}$ has orthonormal columns and $R_\ell \in \mathbb{R}^{\ell\times\ell}$ is upper bidiagonal. We obtain from (2.4) that

$$V_\ell^T A^T A V_\ell = C_{\ell+1,\ell}^T C_{\ell+1,\ell} = R_\ell^T R_\ell. \tag{2.6}$$

The columns of the matrix V_ℓ are the Lanczos vectors and $R_\ell^T R_\ell$ is the tridiagonal Lanczos matrix that would be obtained by applying ℓ steps of the Lanczos algorithm to the symmetric matrix $A^T A$ with initial vector $A^T \boldsymbol{b}$; see, e.g., [2].

We remark that the application of the Lanczos bidiagonalization (2.4) of A with initial vector $\boldsymbol{b}$, rather than using the initial vector $A^T \boldsymbol{b}$, has the advantage that it allows the computation of auxiliary bounds, e.g., the bounds required for the L-ribbon, that may be helpful when determining a suitable value of μ; see [3, 4] for discussions

on the computation and application of the L-ribbon. Lanczos bidiagonalization with initial vector $\boldsymbol{b}$ also has better numerical stability properties; see [14].

Let $R_{\ell-1,\ell}$ denote the leading $(\ell-1)\times\ell$ principal submatrix of R_ℓ, and introduce the functions

$$\phi_\ell^-(\mu) := \|A^T\boldsymbol{b}\|^2 \boldsymbol{e}_1^T (R_\ell^T R_\ell + \mu I_\ell)^{-2}\boldsymbol{e}_1, \tag{2.7}$$

$$\phi_\ell^+(\mu) := \|A^T\boldsymbol{b}\|^2 \boldsymbol{e}_1^T (R_{\ell-1,\ell}^T R_{\ell-1,\ell} + \mu I_\ell)^{-2}\boldsymbol{e}_1. \tag{2.8}$$

These functions approximate $\phi(\mu)$ in the following manner. Substituting the spectral factorization

$$A^TA = W\Lambda W^T, \quad \Lambda = \mathrm{diag}[\lambda_1,\lambda_2,\ldots,\lambda_n], \quad W\in\mathbb{R}^{n\times n}, \quad W^TW = I,$$

into (2.2) yields

$$\phi(\mu) = \hat{\boldsymbol{b}}^T(\Lambda+\mu I)^{-2}\hat{\boldsymbol{b}} = \sum_{j=1}^m \frac{\hat\beta_j^2}{(\lambda_j+\mu)^2} \equiv \int_0^\infty \frac{1}{(\lambda+\mu)^2}\mathrm{d}\omega(\lambda), \tag{2.9}$$

where $\hat{\boldsymbol{b}} = [\hat\beta_j]_{j=1}^m := W^TA^T\boldsymbol{b}$ and the distribution function $\omega(\lambda)$ associated with the nonnegative measure $\mathrm{d}\omega(\lambda)$ is piecewise constant with jump discontinuities of height $\hat\beta_j^2$ at the eigenvalues λ_j. Thus, the integral in the right-hand side of (2.9) is a Stieltjes integral defined by the spectral factorization of A^TA and the vector $A^T\boldsymbol{b}$.

It can be shown that $\phi_\ell^-(\mu)$, defined by (2.7), is the ℓ-point Gauss quadrature rule associated with the distribution function ω applied to the function

$$\psi_\mu(t) := (t+\mu)^{-2}; \tag{2.10}$$

i.e., we can express $\phi_\ell^-(\mu)$ as

$$\phi_\ell^-(\mu) = \sum_{j=1}^\ell \psi_\mu(t_j^{(\ell)})w_j^{(\ell)}, \tag{2.11}$$

where $t_j^{(\ell)}$ and $w_j^{(\ell)}$ denote Gaussian nodes and weights, respectively; see, e.g., [3, 8, 10] for details.

Similarly, $\phi_\ell^+(\mu)$, defined by (2.8), is the ℓ-point Gauss–Radau quadrature rule with a node at the origin associated with the distribution function ω applied to the function (2.10). Since, for any positive fixed value of μ, the derivatives of $\psi_\mu(t)$ with respect to t of odd order are strictly negative and the derivatives of even order are strictly positive for $t\geq 0$, the remainder terms for Gauss and Gauss–Radau quadrature rules yield

$$\phi_\ell^-(\mu) < \phi(\mu) < \phi_\ell^+(\mu), \qquad \mu>0; \tag{2.12}$$

see, e.g., [3].

Let $\omega_\ell(t)$ denote the piecewise constant distribution function determined by the ℓ-point Gauss quadrature rule (2.11); i.e., $\omega_\ell(t)$ has a jump discontinuity of height $w_j^{(\ell)}$ at the node $t_j^{(\ell)}$ for $1\leq j\leq\ell$. The function $\phi_{\ell-1}^-(\mu)$, defined by (2.7) with ℓ replaced by $\ell-1$, can be considered an $(\ell-1)$-point Gauss quadrature rule associated with the distribution function $\omega_\ell(t)$, and, similarly to the left-hand side inequality in (2.12), we obtain

$$\phi_{\ell-1}^-(\mu) < \phi_\ell^-(\mu), \qquad \mu>0; \tag{2.13}$$

see Hanke [11] for a discussion. Analogously, one also can show that

$$\phi_\ell^+(\mu) < \phi_{\ell-1}^+(\mu), \qquad \mu > 0. \tag{2.14}$$

The zero-finder discussed in section 3 requires the evaluation of the functions $\phi_\ell^\pm(\mu)$ and the derivative $\frac{d}{d\mu}\phi_\ell^+(\mu)$ for several values of the argument μ. We evaluate $\phi_\ell^+(\mu)$ by first determining the solution $\boldsymbol{z}_\ell$ of the linear system of equations

$$(R_{\ell-1,\ell}^T R_{\ell-1,\ell} + \mu I_\ell)\boldsymbol{z} = \boldsymbol{e}_1 \tag{2.15}$$

and then evaluating

$$\phi_\ell^+(\mu) = \|A^T\boldsymbol{b}\|^2 \boldsymbol{z}_\ell^T \boldsymbol{z}_\ell.$$

Note that (2.15) are the normal equations associated with the least-squares problem

$$\min_{\boldsymbol{z}\in\mathbb{R}^\ell} \left\| \begin{bmatrix} R_{\ell-1,\ell} \\ \mu^{1/2} I_\ell \end{bmatrix} \boldsymbol{z} - \mu^{-1/2}\boldsymbol{e}_\ell \right\|, \tag{2.16}$$

and we compute $\boldsymbol{z}_\ell$ by solving the latter by a method described by Eldén [7] and Paige and Saunders [15]. This method uses orthogonal transformations and requires only $\mathcal{O}(\ell)$ arithmetic floating-point operations for each value of μ. The function $\phi_\ell^-(\mu)$ is evaluated analogously.

The derivative

$$\frac{d}{d\mu}\phi_\ell^+(\mu) = -2\|A^T\boldsymbol{b}\|^2 \boldsymbol{e}_1^T (R_{\ell-1,\ell}^T R_{\ell-1,\ell} + \mu I_\ell)^{-3}\boldsymbol{e}_1 \tag{2.17}$$

is evaluated by solving the least-squares problem

$$\min_{\boldsymbol{s}\in\mathbb{R}^\ell} \left\| \begin{bmatrix} R_{\ell-1,\ell} \\ \mu^{1/2} I_\ell \end{bmatrix} \boldsymbol{s} - \mu^{-1/2} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{z}_\ell \end{bmatrix} \right\|.$$

Denote the solution by $\boldsymbol{s}_\ell$. The value of the derivative is then given by

$$\frac{d}{d\mu}\phi_\ell^+(\mu) = -2\|A^T\boldsymbol{b}\|^2 \boldsymbol{z}_\ell^T \boldsymbol{s}_\ell.$$

Finally, we note that when the entry $\sigma_{\ell+1}$ of the matrix $C_{\ell+1,\ell}$ vanishes, we have $\phi_\ell^-(\mu) = \phi(\mu)$ for $\mu > 0$. Hence, we can evaluate the function $\phi(\mu)$ inexpensively by computing the value of $\phi_\ell^-(\mu)$. The computation of a suitable value of the regularization parameter and an associated solution of (1.6) is fairly straightforward in this situation and will not be discussed further.

3. Computation of an approximate solution of specified norm. We exploit the relation between the function (2.1) and Gauss quadrature rules discussed in the previous section to derive a numerical method for the computation of an approximate solution of the minimization problem (1.10). Our method differs from the one described by Golub and von Matt [8] in that the computed approximations $\mu_\ell^{(j)}$ of μ_Δ converge monotonically to μ_Δ from the right. Note that the condition number, i.e., the ratio of the largest and the smallest singular values, of the matrix in (2.16) is a decreasing function of $\mu > 0$. The fact that $\mu_\ell^{(j)} \geq \mu_\Delta$ secures that none of the matrices

$$\begin{bmatrix} R_{\ell-1,\ell} \\ \left(\mu_\ell^{(j)}\right)^{1/2} I_\ell \end{bmatrix} \in \mathbb{R}^{(2\ell-1)\times\ell}, \tag{3.1}$$

which arise when evaluating $\phi_\ell^+(\mu_\ell^{(j)})$ for nonacceptable intermediate approximations $\mu_\ell^{(j)}$ of μ_Δ by solving (2.16) with $\mu = \mu_\ell^{(j)}$, are more ill-conditioned than the matrix associated with the computed acceptable approximation. This property is important, because since A is of ill-determined rank, the matrix (3.1) may be severely ill-conditioned when $\mu_\ell^{(j)} > 0$ is tiny.

We would like to determine a value of μ such that

$$\Delta^2\eta^2 \le \phi(\mu) \le \Delta^2, \tag{3.2}$$

where the constant $0 < \eta < 1$ determines how close to Δ we require the norm of the computed approximate solution of (1.10) to be. It follows from Proposition 2.1 that any μ which satisfies (3.2) is bounded below by μ_Δ. We determine such a value of μ by computing a pair $\{\ell, \mu\}$, such that

$$\Delta^2\eta^2 \le \phi_\ell^-(\mu), \qquad \phi_\ell^+(\mu) \le \Delta^2, \tag{3.3}$$

where as usual ℓ denotes the number of Lanczos bidiagonalization steps. It follows from (2.12) that the inequalities (3.3) imply (3.2). For many linear discrete ill-posed problems, the value of ℓ in a pair $\{\ell, \mu\}$ that satisfies (3.3) can be chosen fairly small, and therefore the effort required to compute such a pair is typically much smaller than the effort necessary to solve (1.11) to high accuracy.

We assume that the linear discrete ill-posed problem (1.1) is so large that the dominating computational work when determining an approximate solution of (1.10) is the evaluation of 2ℓ matrix-vector products with the matrices A and A^T required to compute the partial Lanczos bidiagonalization (2.4). Our numerical method is designed to keep the number of Lanczos bidiagonalization steps ℓ small. The method starts with $\ell = 2$ and then increases ℓ if necessary.

For a given value of $\ell \ge 2$, we determine approximations $\mu_\ell^{(j)}$, $j = 0, 1, 2, \ldots$, of the largest zero μ_ℓ of the function

$$h_\ell^+(\mu) := \phi_\ell^+(\mu) - \Delta^2.$$

The following proposition collects some properties of $\phi_\ell^+(\mu)$.

PROPOSITION 3.1. *The function $\phi_\ell^+(\mu)$, defined by* (2.8), *is strictly decreasing and convex for $\mu > 0$. The equation*

$$\phi_\ell^+(\mu) = \tau \tag{3.4}$$

has a unique solution μ, such that $0 < \mu < \infty$, for any positive finite value of τ.

Proof. The properties stated follow from the representation (2.8) and the fact that the orthonormal eigenvectors of the singular unreduced symmetric tridiagonal matrix $R_{\ell-1,\ell}^T R_{\ell-1,\ell}$ have nonvanishing first components. □

The proposition shows that (3.4) has a unique solution whenever (2.3) has one. Since $\phi_\ell^+(\mu)$ and $\phi(\mu)$ are decreasing convex functions, it follows from (2.12) and (2.14) that

$$\mu_\Delta \le \mu_\ell < \mu_{\ell-1}, \qquad \ell = 2, 3, 4, \ldots . \tag{3.5}$$

We use the monotonically and quadratically convergent zero-finder described by Golub and von Matt [8, equations (75)–(78)] to determine a sequence of approximations $\mu_\ell^{(j)}$, $j = 1, 2, 3, \ldots$, of μ_ℓ. The zero-finder requires repeated evaluation of the function (2.8) and its derivative (2.17).

We assume that the initial approximation $\mu_\ell^{(0)}$ of μ_ℓ satisfies $\mu_\ell \leq \mu_\ell^{(0)}$. The computed approximations $\mu_\ell^{(j)}$ then satisfy $\mu_\ell \leq \mu_\ell^{(j)} \leq \mu_\ell^{(j-1)}$ for $j = 1, 2, 3, \ldots$. We terminate the iterations with the zero-finder as soon as an approximation $\mu_\ell^{(p)}$ of μ_ℓ has been found, such that

$$\frac{1}{10}(\eta^2 - 1)\Delta^2 + \Delta^2 \leq \phi_\ell^+(\mu_\ell^{(p)}) \leq \Delta^2. \tag{3.6}$$

This stopping criterion is used in the numerical experiments reported in section 4; however, we remark that factors other than 1/10 for the negative term in (3.6) could have been used instead. The factor has to be positive and strictly smaller than unity. The use of a larger factor may reduce the number of iterations with the zero-finder but could increase the number of Lanczos bidiagonalization steps ℓ.

If, in addition to (3.6), $\mu_\ell^{(p)}$ also satisfies

$$\Delta^2\eta^2 \leq \phi_\ell^-(\mu_\ell^{(p)}), \tag{3.7}$$

then both inequalities (3.3) hold for $\mu = \mu_\ell^{(p)}$, and we accept $\mu_\ell^{(p)}$ as an approximation of μ_Δ. It follows from (2.12), (2.13), and (2.14) that the difference $\phi_\ell^+(\mu) - \phi_\ell^-(\mu)$ decreases monotonically to zero as ℓ increases for any fixed $\mu > 0$. Therefore, since

$$\Delta^2\eta^2 < \frac{1}{10}(\eta^2 - 1)\Delta^2 + \Delta^2,$$

the left-hand side inequality of (3.6) implies (3.7) when ℓ is sufficiently large.

The above discussion suggests that if inequality (3.7) does not hold, then the number of Lanczos bidiagonalization steps ℓ should be increased. Specifically, when $\mu_\ell^{(p)}$ satisfies (3.6) but not (3.7), we carry out one more Lanczos bidiagonalization step, replacing ℓ by $\ell+1$, and seek to determine an approximation of the largest zero, denoted by $\mu_{\ell+1}$, of the function

$$h_{\ell+1}^+(\mu) := \phi_{\ell+1}^+(\mu) - \Delta^2,$$

using the same zero-finder as above with initial approximate solution $\mu_{\ell+1}^{(0)} := \mu_\ell^{(p)}$. It follows from (3.6) and Proposition 3.1 that $\mu_{\ell+1}^{(0)} \geq \mu_\ell$, and in view of (3.5), with ℓ replaced by $\ell + 1$, we have that $\mu_{\ell+1}^{(0)} > \mu_{\ell+1}$. Thus, if $\mu_2^{(0)} \geq \mu_2$, then $\mu_\ell^{(j)} \geq \mu_\ell$ for all $j \geq 0$ and $\ell \geq 2$. It is easy to determine whether $\mu_2^{(0)} \geq \mu_2$ by evaluating $\phi_2(\mu_2^{(0)})$; Proposition 3.1 shows that $\phi_2(\mu_2^{(0)}) \leq 0$ implies $\mu_2^{(0)} \geq \mu_2$.

Assume that $\mu_\ell^{(p)}$ satisfies (3.6) and (3.7). We then compute the solution $\boldsymbol{y}_\ell$ of

$$(R_\ell^T R_\ell + \mu_\ell^{(p)} I_\ell)\boldsymbol{y} = \rho_1\sigma_1\boldsymbol{e}_1 \tag{3.8}$$

by solving a least-squares problem analogous to (2.16) and evaluate the approximation $\boldsymbol{x}_\ell := V_\ell\boldsymbol{y}_\ell$ of $\boldsymbol{x}_{\mu_\Delta}$. The matrix R_ℓ and the constants ρ_1 and σ_1 are defined by (2.4), (2.5), and (2.6). The vector $\boldsymbol{x}_\ell$ is a Galerkin approximation of $\boldsymbol{x}_{\mu_\ell^{(p)}}$. It satisfies $\|\boldsymbol{x}_\ell\| = (\phi_\ell^-(\mu_\ell^{(p)}))^{1/2}$; see [3, Theorem 5.1] for details.

Instead of determining the largest zero of $h_\ell^+(\mu)$, Golub and von Matt [8] compute the largest zero of the function

$$h_\ell^-(\mu) := \phi_\ell^-(\mu) - \Delta^2. \tag{3.9}$$

The latter zero is smaller than μ_Δ. Our discussion in the beginning of this section showed that this may lead to numerical difficulties when A is of ill-determined rank, because the evaluation of $\phi_\ell^-(\mu)$ at approximations μ of μ_Δ that are smaller than μ_Δ may require the solution of least-squares problems with severely ill-conditioned matrices. The computation of the zero of $h_\ell^-(\mu)$ is discussed in Example 4.4 of the following section.

4. Computed examples. We illustrate how the numerical method described in the previous sections performs on a few linear discrete ill-posed problems. All computations are carried out using Matlab with approximately 16 decimal digits of accuracy. We choose the initial values $\ell = 2$ and $\mu_2^{(0)} = 10$, which in all examples give $\phi_2(\mu_2^{(0)}) < 0$. Hence, $\mu_2^{(0)} > \mu_2$. The error vectors $\boldsymbol{e}$ used in the examples have normally distributed random entries with zero mean; the variance is chosen so that $\boldsymbol{e}$ is of desired norm. The matrices A and right-hand sides $\hat{\boldsymbol{b}}$ of the linear system (1.2) are generated by Matlab programs in the Regularization Tools package by Hansen [12]. These programs also determine discretizations $\hat{\boldsymbol{x}}$ of the solution of the ill-posed problems whose discretizations yield (1.2). We refer to $\hat{\boldsymbol{x}}$ as the "exact solution." The value of Δ is determined by (1.9). The examples and norms of the error $\boldsymbol{e}$ are chosen to enable comparison with results reported by Rojas [17] and Rojas and Sorensen [19] for their LSTRS-based method.

Example 4.1. Consider the Fredholm integral equation of the first kind,

$$\int_{-6}^{6} k(\tau,\sigma)x(\sigma)d\sigma = b(\tau), \qquad -6 \le \tau \le 6, \tag{4.1}$$

discussed by Phillips [16]. Its solution, kernel, and right-hand side are given by

$$x(\sigma) := \begin{cases} 1+\cos(\frac{\pi}{3}\sigma) & \text{if } |\sigma| < 3, \\ 0 & \text{otherwise,} \end{cases}$$

$$k(\tau,\sigma) := x(\tau-\sigma),$$

$$b(\tau) := (6-|\tau|)\left(1+\frac{1}{2}\cos\left(\frac{\pi}{3}\tau\right)\right) + \frac{9}{2\pi}\sin\left(\frac{\pi}{3}|\tau|\right).$$

We use the code phillips from Regularization Tools [12] to discretize (4.1) by a Galerkin method with orthonormal box functions as test and trial functions to obtain the symmetric indefinite matrix $A \in \mathbb{R}^{300\times 300}$, right-hand side $\hat{\boldsymbol{b}} \in \mathbb{R}^{300}$, and exact solution $\hat{\boldsymbol{x}} \in \mathbb{R}^{300}$. Then, by (1.9), $\Delta = 2.9999$. The exact solution $\hat{\boldsymbol{x}}$ is a scaled discretization of the solution x of (4.1). The condition number of A, defined by $\kappa(A) := \|A\|\|A^{-1}\|$, is $2.1 \cdot 10^8$.

An error vector $\boldsymbol{e}$ of norm $9.9409 \cdot 10^{-2}$ is added to $\hat{\boldsymbol{b}}$ to yield the right-hand side $\boldsymbol{b}$ of (1.1). This gives the relative error $\|\boldsymbol{e}\|/\|\hat{\boldsymbol{b}}\| = 6.5013 \cdot 10^{-3}$. The method of the present paper with $\eta = 0.999$ terminates after 8 Lanczos bidiagonalization steps, and the computed approximate solution $\boldsymbol{x}_8$, defined by (3.8), has relative error $\|\boldsymbol{x}_8 - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 1.7143 \cdot 10^{-2}$ and norm 2.9973. The computation of $\boldsymbol{x}_8$ requires only 16 matrix-vector product evaluations with A or A^T. The method determines the approximation $1.2748 \cdot 10^{-2}$ of μ_Δ. Figure 4.1 displays $\boldsymbol{x}_8$ and $\hat{\boldsymbol{x}}$. We remark that the unregularized solution (1.4) of (1.1) is of norm $3.4 \cdot 10^5$, which is much larger than Δ.

Rojas [17] and Rojas and Sorensen [19] describe numerical experiments for this example for their LSTRS-based method. Rojas [17, p. 104] reports that the

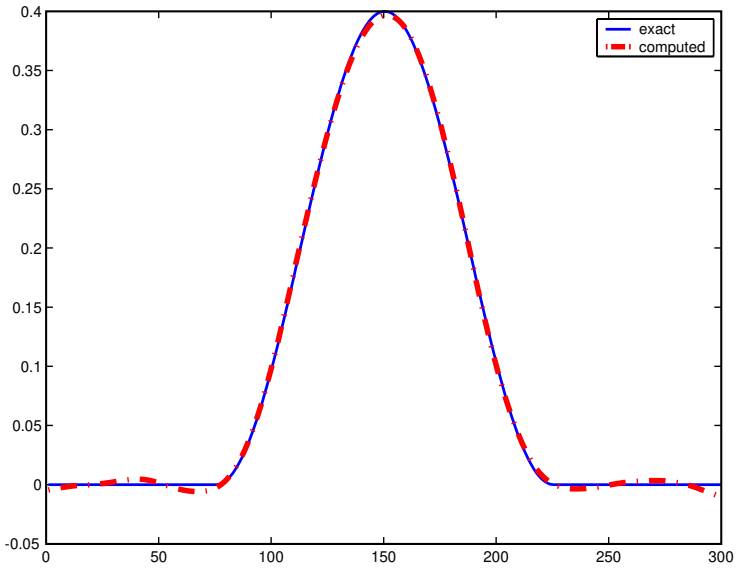

FIG. 4.1. *Example* 4.1: *Exact solution* $\hat{\boldsymbol{x}}$ *(continuous curve) and computed approximate solution* $\boldsymbol{x}_8$ *(dash-dotted curve) for relative error* $\|\boldsymbol{e}\|/\|\hat{\boldsymbol{b}}\| = 6.5013 \cdot 10^{-3}$ *in* $\boldsymbol{b}$.

computation of an approximate solution $\boldsymbol{x}'$, such that $\|\boldsymbol{x}' - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 1.9405 \cdot 10^{-2}$ and $\|\boldsymbol{x}'\| = 2.9999$, requires 691 matrix-vector product evaluations with the matrices A or A^T; Rojas and Sorensen [19, p. 1852] claim that the computation of a less accurate approximate solution $\boldsymbol{x}''$, such that $\|\boldsymbol{x}'' - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 2.6883 \cdot 10^{-2}$ and $\|\boldsymbol{x}''\| = 2.9869$, requires 521 matrix-vector product evaluations with A or A^T.

We now increase the norm of the error $\boldsymbol{e}$ in the right-hand side $\boldsymbol{b}$ to obtain the relative error $\|\boldsymbol{e}\|/\|\hat{\boldsymbol{b}}\| = 0.1$. Our method with $\eta = 0.999$ then requires 9 Lanczos bidiagonalization steps, i.e., 18 matrix-vector product evaluations with the matrices A or A^T, to determine the solution $\boldsymbol{x}_9$, which satisfies $\|\boldsymbol{x}_9 - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 8.2190 \cdot 10^{-2}$ and $\|\boldsymbol{x}_9\| = 2.9986$. The computed approximate solution $\boldsymbol{x}_9$ is shown in Figure 4.2, which illustrates that, in the presence of a fairly large relative error $\|\boldsymbol{e}\|/\|\hat{\boldsymbol{b}}\|$, the constraint $\|\boldsymbol{x}\| \leq \Delta$, with Δ given by (1.9), does not prevent the computed approximate solution from oscillating. The value of μ determined is $5.9299 \cdot 10^{-2}$. Since the choice (1.9) of Δ is independent of $\|\boldsymbol{e}\|/\|\hat{\boldsymbol{b}}\|$, it is not surprising that the value of μ used in the computation of $\boldsymbol{x}_9$ is not always the most appropriate. For this problem, the norm of the unregularized solution (1.4) is $5.2 \cdot 10^6$.

For comparison, we determine an approximate solution by the numerical method described in [6]. This method computes an approximate solution $\boldsymbol{x}_{\text{discr}}$ and a value μ_{discr} of the regularization parameter, such that $\boldsymbol{x}_{\text{discr}}$ satisfies the discrepancy principle (1.8). Similarly as the method of the present paper, the scheme described in [6] determines a partial Lanczos bidiagonalization of A and computes bounds for certain pertinent quantities by the use of Gauss-type quadrature rules. An auxiliary parameter η is set to 1.001. This parameter is analogous to the parameter η of the present paper; roughly η of the present paper corresponds to $1/\eta$ in [6]. With the same relative error 0.1 in $\boldsymbol{b}$ as above, the method of [6] required 8 Lanczos bidiagonalization steps to compute the approximate solution $\boldsymbol{x}_{\text{discr}}$, which satisfies $\|\boldsymbol{x}_{\text{discr}} - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 4.4174 \cdot 10^{-2}$. This error is about half the error achieved by the

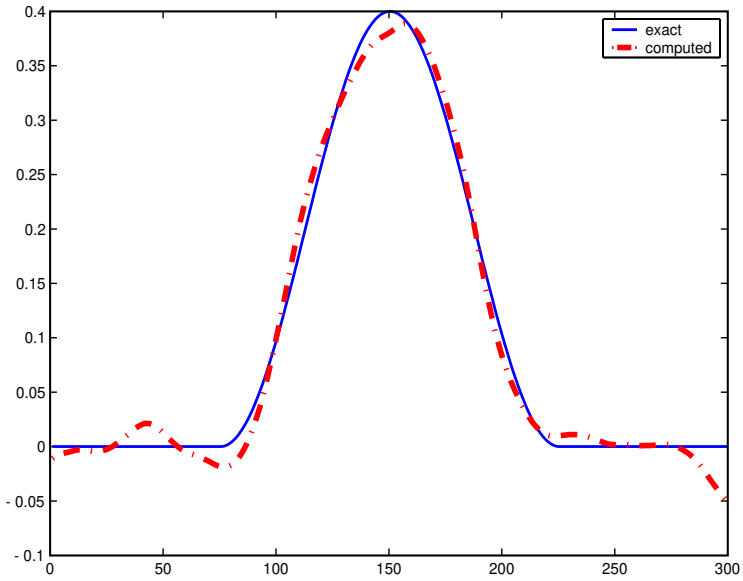

FIG. 4.2. *Example* 4.1: *Exact solution* $\hat{\boldsymbol{x}}$ *(continuous curve) and computed approximate solution* $\boldsymbol{x}_9$ *(dash-dotted curve) for relative error* $\|\boldsymbol{e}\|/\|\hat{\boldsymbol{b}}\| = 0.1$ *in* $\boldsymbol{b}$.

method of the present paper, and the associated value of the regularization parameter $\mu_{\rm discr} = 3.1504 \cdot 10^{-1}$ is much larger than the value $\mu = 5.9299 \cdot 10^{-2}$ determined by the method of this paper.

This comparison as well as numerical examples with other discrete ill-posed problems indicate that the discrepancy principle gives more accurate approximations of $\hat{\boldsymbol{x}}$ than the method of the present paper when the relative error in the right-hand side $\boldsymbol{b}$ is large. This depends on the value of μ determined by the method of the present paper being independent of the error in $\boldsymbol{b}$.

Finally, we use the code phillips to determine the symmetric indefinite matrix $A \in \mathbb{R}^{1000\times 1000}$, the right-hand side $\hat{\boldsymbol{b}} \in \mathbb{R}^{1000}$, and the exact solution $\hat{\boldsymbol{x}} \in \mathbb{R}^{1000}$. Then Δ, defined by (1.9), has the value 3.0000. An error vector $\boldsymbol{e}$ of norm $9.9409 \cdot 10^{-2}$ is added to $\hat{\boldsymbol{b}}$ to give $\boldsymbol{b}$. This gives the relative error $\|\boldsymbol{e}\|/\|\hat{\boldsymbol{b}}\| = 6.5012 \cdot 10^{-3}$. The method of the present paper with $\eta = 0.999$ terminates after 9 Lanczos bidiagonalization steps with the computed approximate solution $\boldsymbol{x}_9$ and $\mu = 1.2095 \cdot 10^{-2}$. We have $\|\boldsymbol{x}_9 - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 1.0230 \cdot 10^{-2}$ and $\|\boldsymbol{x}_9\| = 2.9971$. The unregularized solution (1.4) is of norm $2.3 \cdot 10^7$.

Rojas [17, p. 104] reports that for this problem the LSTRS-based method requires 751 matrix-vector product evaluations with the matrices A or A^T to compute an approximate solution $\boldsymbol{x}'$, such that $\|\boldsymbol{x}' - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 2.6030 \cdot 10^{-2}$ and $\|\boldsymbol{x}'\| = 3.0000$. Rojas and Sorensen [19, p. 1852] state that the computation of a less accurate approximate solution $\boldsymbol{x}''$, such that $\|\boldsymbol{x}'' - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 3.3607 \cdot 10^{-2}$ and $\|\boldsymbol{x}''\| = 2.9839$, requires the evaluation of 575 matrix-vector products.

We found that, for all computations reported in this example, the number of iterations required and the approximation errors achieved are the same up to at least 5 significant digits whether or not reorthogonalization of the columns of the matrices $U_{\ell+1}$ and V_ℓ in the partial Lanczos bidiagonalization (2.4) is carried out. We will comment on this further below. □

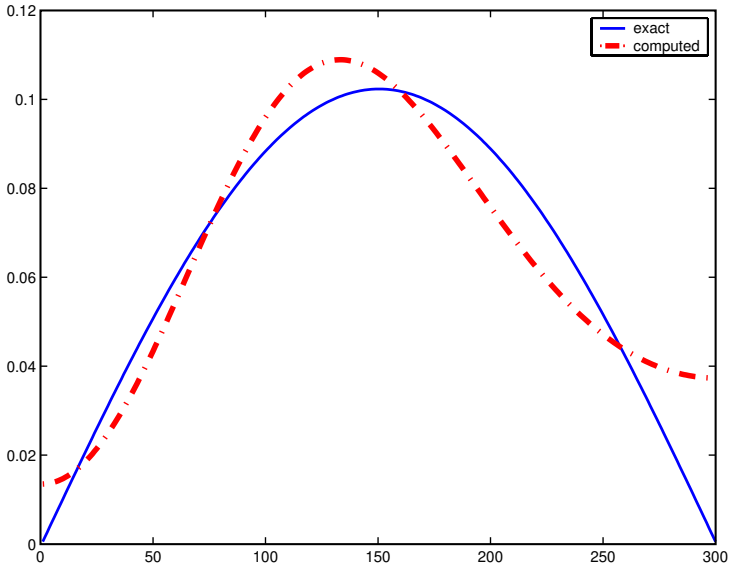

FIG. 4.3. *Example* 4.2: *Exact solution* $\hat{\boldsymbol{x}}$ *(continuous curve) and computed approximate solution* $\boldsymbol{x}_4$ *(dash-dotted curve) for relative error* $\|\boldsymbol{e}\|/\|\hat{\boldsymbol{b}}\| = 3.4315 \cdot 10^{-2}$ *in* $\boldsymbol{b}$.

Example 4.2. Consider the equation

$$\int_0^{\pi/2} k(\sigma,\tau)x(\sigma)d\sigma = b(\tau), \qquad 0 \le \tau \le \pi, \tag{4.2}$$

where $k(\sigma,\tau) := \exp(\sigma\cos(\tau))$ and $b(\tau) := 2\sinh(\tau)/\tau$. This integral equation is discussed by Baart [1]. It has the solution $x(\tau) = \sin(\tau)$. We use the Matlab code baart from [12] to discretize (4.2) by a Galerkin method with 300 orthonormal box functions as test and trial functions. This yields the matrix $A \in \mathbb{R}^{300\times 300}$, the right-hand side $\hat{\boldsymbol{b}} \in \mathbb{R}^{300}$, and the exact solution $\hat{\boldsymbol{x}}$. The latter is a scaled discretization of the solution of (4.2); we have $\|\hat{\boldsymbol{x}}\| = 1.2533$. Adding an error vector $\boldsymbol{e}$ of norm $9.9409\cdot 10^{-2}$ to $\hat{\boldsymbol{b}}$ yields the right-hand side $\boldsymbol{b}$ of (1.1) with relative error $\|\boldsymbol{e}\|/\|\hat{\boldsymbol{b}}\| = 3.4315 \cdot 10^{-2}$. Moreover, Matlab yields $\kappa(A) = 3.3 \cdot 10^{18}$ and $\|\boldsymbol{x}_0\| = 1.1 \cdot 10^{10}$, where $\boldsymbol{x}_0$ is given by (1.4).

Our method with $\eta = 0.99$ and Δ given by (1.9) determines the approximate solution $\boldsymbol{x}_4$ shown in Figure 4.3; its computation requires 4 steps of Lanczos bidiagonalization, i.e., 8 matrix-vector product evaluations with A or A^T. We have $\|\boldsymbol{x}_4 - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 1.4803 \cdot 10^{-1}$ and $\|\boldsymbol{x}_4\| = 1.2466$. The computed value of μ is $1.6300 \cdot 10^{-5}$.

The number of iterations required for this example and the approximation errors are the same to at least 5 significant digits whether or not reorthogonalization of the columns of the matrices U_5 and V_4 in the partial Lanczos bidiagonalization (2.4) with $\ell = 4$ is carried out.

Rojas [17, p. 104] reports that the LSTRS-based method requires 491 matrix-vector product evaluations to determine a solution $\boldsymbol{x}'$, such that $\|\boldsymbol{x}' - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 1.7723 \cdot 10^{-1}$ and $\|\boldsymbol{x}'\| = 1.2533$. □

Example 4.3. The discrete ill-posed problem of this example is generated with the Matlab code foxgood from Regularization Tools [12]. The symmetric indefinite

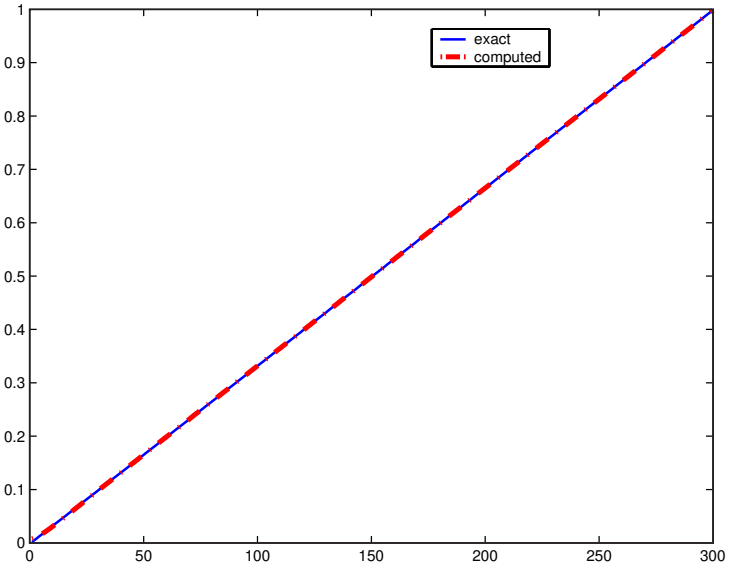

FIG. 4.4. *Example* 4.3: *Exact solution* $\hat{\boldsymbol{x}}$ *(continuous curve) and computed approximate solution* $\boldsymbol{x}_6$ *(dash-dotted curve). No error vector* $\boldsymbol{e}$ *is added to the right-hand side* $\hat{\boldsymbol{b}}$ *of* (1.2)*; i.e., the right-hand side* $\boldsymbol{b}$ *in* (1.1) *is* $\hat{\boldsymbol{b}}$.

matrix $A \in \mathbb{R}^{300\times 300}$ generated by the code is of low numerical rank; its norm is 0.81, and it has only 28 eigenvalues of magnitude larger than $1 \cdot 10^{-14}$. The right-hand side $\boldsymbol{b} \in \mathbb{R}^{300}$ is also determined by the code foxgood. No error vector $\boldsymbol{e}$ is added to $\boldsymbol{b}$. The vector $\hat{\boldsymbol{x}}$ denotes the exact solution generated by foxgood. Matlab yields $\kappa(A) = 7.2 \cdot 10^{20}$ and $\|\hat{\boldsymbol{x}}\| = 1.0000 \cdot 10^{1}$, and we let $\Delta = 10$.

Rojas and Sorensen [19] point out that the minimization problem (1.10) is difficult to solve because it is close to the "hard case." Rojas [17, p. 104] reports that 389 matrix-vector products are required to compute an approximate solution $\boldsymbol{x}'$, such that $\|\boldsymbol{x}' - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 4.3303 \cdot 10^{-2}$ and $\|\boldsymbol{x}'\| = 9.9999$.

The method of the present paper with $\eta = 0.999999$ requires 6 Lanczos bidiagonalization steps, i.e., 12 matrix-vector product evaluations with the matrices A or A^T, and gives the approximate solution $\boldsymbol{x}_6$, which satisfies $\|\boldsymbol{x}_6 - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 8.8996 \cdot 10^{-4}$ and $\|\boldsymbol{x}_6\| = 1.0000 \cdot 10^{1}$. The vectors $\boldsymbol{x}_6$ and $\hat{\boldsymbol{x}}$ are displayed in Figure 4.4. We remark that the computations were carried out with reorthogonalization of the columns of the matrices U_7 and V_6 in (2.4). The computed value of μ is $2.1721 \cdot 10^{-8}$.

When no reorthogonalization is carried out, our method terminates after 9 Lanczos bidiagonalization steps. We obtain the approximate solution $\boldsymbol{x}_9$, which satisfies $\|\boldsymbol{x}_9 - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 8.8965 \cdot 10^{-4}$ and $\|\boldsymbol{x}_9\| = 1.0000 \cdot 10^{1}$. The computed value of μ is $2.1701 \cdot 10^{-8}$.

We now add an error vector $\boldsymbol{e}$ of norm $9.90409 \cdot 10^{-2}$ to $\hat{\boldsymbol{x}}$ to obtain the right-hand side $\boldsymbol{b}$ of (1.1) with relative error $\|\boldsymbol{e}\|/\|\hat{\boldsymbol{x}}\| = 1.2828 \cdot 10^{-2}$. Our method applied to this system with $\eta = 0.999$ requires 3 Lanczos bidiagonalization steps to determine the approximate solution $\boldsymbol{x}_3$ of norm 9.9997 and to have relative error $\|\boldsymbol{x}_3 - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 2.7289 \cdot 10^{-4}$. The computed value of μ is $2.2987 \cdot 10^{-4}$. These values are the same up to at least 5 significant digits whether or not reorthogonalization of the columns of U_4 and V_3 is carried out. □

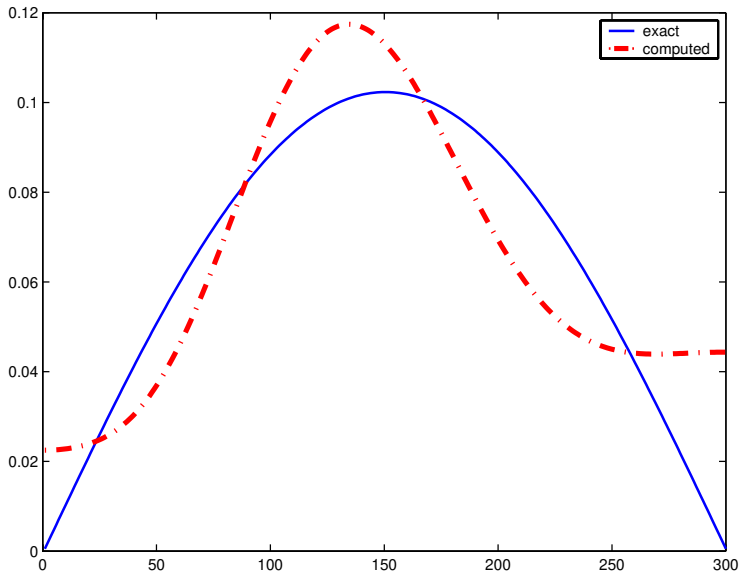

FIG. 4.5. *Example* 4.4: *Exact solution* $\hat{\boldsymbol{x}}$ *(continuous curve) and computed approximate solution* $\boldsymbol{x}_5$ *(dash-dotted curve) determined by the method described in* [8].

Example 4.4. Golub and von Matt [8] choose μ to be the zero of the function $h_\ell^-(\mu)$ given by (3.9). This example discusses the computation of this zero. Assume that the upper triangular matrix R_ℓ, defined by (2.6), is nonsingular. Then it follows from (2.7) and the fact that $\phi_\ell^-(\mu)$ is decreasing that for $\mu \geq 0$,

$$0 \leq \phi_\ell^-(\mu) \leq \phi_\ell^-(0), \qquad \phi_\ell^-(0) = \|A^T b\|^2 \|R_\ell^{-1} e_1\|^2.$$

For ℓ small, e.g., for $\ell = 2$, the norm of the vector $R_\ell^{-1} e_1$ might be sufficiently small to make the function $h_\ell^-(\mu)$ strictly negative for all $\mu \geq 0$.

Assume that ℓ is large enough so that $h_\ell^-(\mu)$ has a nonnegative zero. Then, analogously to (3.6), we determine a value $\hat{\mu}$, such that

$$\Delta^2 \leq \phi_\ell^-(\hat{\mu}) \leq \frac{1}{10}(\eta^{-2} - 1)\Delta^2 + \Delta^2$$

for some specified value of η in the open interval $(0, 1)$ using the zero-finder of Golub and von Matt [8, p. 576]. We then compute an approximation x_ℓ of $x_{\hat{\mu}}$ by first solving (3.8) with $\mu_\ell^{(p)}$ replaced by $\hat{\mu}$ for y_ℓ and then letting $x_\ell := V_\ell y_\ell$.

We use this approach to determine a suitable value of the regularization parameter for the problem discussed in Example 4.2. Thus, we use the same integral equation, discretization, error vector $\boldsymbol{e}$, and value of η as in Example 4.2. The functions $h_k^-(\mu)$, $2 \leq k \leq 4$, do not have a nonnegative zero, but $h_5^-(\mu)$ does. We determine the approximation $\hat{\mu} = 5.9734 \cdot 10^{-6}$ of this zero. The associated computed approximate solution $\boldsymbol{x}_5$ is shown in Figure 4.5. It satisfies $\|\boldsymbol{x}_5 - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 2.1274 \cdot 10^{-1}$ and $\|\boldsymbol{x}_5\| = 1.2533$. A comparison with Example 4.2 shows that the computed value of the regularization parameter for the present example is smaller, the error $\|\boldsymbol{x}_5 - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\|$ is larger, and one more Lanczos bidiagonalization step is required. Many similar examples can be found. We remark that the computed approximate solution depends on the error vector $\boldsymbol{e}$, both in this example and in Example 4.2.

In summary, the approach of the present paper is advantageous for ill-posed problems because it determines a monotonically decreasing sequence of values of the regularization parameter; see the discussion in the beginning of section 3. Moreover, Proposition 3.1 secures that a solution $\mu > 0$ of the equation $\phi_\ell^+(\mu) = \Delta^2$ exists, while the equation $\phi_\ell^-(\mu) = \Delta^2$ solved in this example might not have a real positive solution when ℓ is small. This makes it somewhat simpler to code the method proposed in the present paper than the scheme used in this example. It therefore may be attractive to use the approach of the present paper to determine a suitable value of the parameter μ also for the problems discussed in [8]. □

Examples 4.1, 4.2, and 4.3 illustrate that the numerical method of the present paper may require significantly fewer matrix-vector product evaluations than the method recently proposed in [17, 19]. The solution of constrained minimization problems (1.10) with Δ given by (1.9) gives pleasing approximations of $\hat{\boldsymbol{x}}$ when the relative error $\|\boldsymbol{e}\|/\|\hat{\boldsymbol{b}}\|$ is small. For large relative errors, the computed solution can be oscillatory, and it may be possible to determine more accurate approximations of the exact solution $\hat{\boldsymbol{x}}$ by other methods for choosing a value of the regularization parameter, such as the discrepancy principle, the L-curve criterion, or generalized cross validation.

The method of this paper has been implemented with and without reorthogonalization of the columns of the matrices $U_{\ell+1}$ and V_ℓ determined by ℓ steps of Lanczos bidiagonalization; cf. (2.4). The examples presented, except for foxgood with $\boldsymbol{e} = \boldsymbol{0}$ of Example 4.3, do not benefit from reorthogonalization. If computer storage is scarce, then it may be attractive to implement the method of this paper without reorthogonalization and so that only the most recently generated columns of the matrices $U_{\ell+1}$ and V_ℓ are stored. Such an implementation keeps the storage requirement low and independent of the number of Lanczos bidiagonalization steps ℓ as this number is increased but requires the columns of the matrix V_ℓ to be recomputed when an appropriate value of the regularization parameter μ has been determined and the approximate solution $\boldsymbol{x}_\ell$ is to be computed as described at the end of section 3. Thus, a very storage-efficient implementation can be achieved by doubling the number of matrix-vector product evaluations. However, we remark that the question of whether reorthogonalization may be required for some problems needs further investigation. Computed examples reported in [5] show that the accuracy of bounds determined by Gauss-type quadrature rules may be affected by loss of orthogonality of the columns of $U_{\ell+1}$ and V_ℓ.

We note that the LSTRS-based method by Rojas [17] and Rojas and Sorensen [19] allows a user to choose the maximum number of vectors stored simultaneously. The choice of this number may affect the number of matrix-vector product evaluations required by the method.

Acknowledgment. We would like to thank Mike Saunders for comments.

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, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
[3] D. Calvetti, G. H. Golub, and L. Reichel, *Estimation of the L-curve via Lanczos bidiagonalization*, BIT, 39 (1999), pp. 603–619.
[4] D. Calvetti, P. C. Hansen, and L. Reichel, *L-curve curvature bounds via Lanczos bidiagonalization*, Electron. Trans. Numer. Anal., 14 (2002), pp. 134–149.
[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] L. Eldén, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT, 17 (1977), pp. 134–145.
[8] G. H. Golub and U. von Matt, *Quadratically constrained least squares and quadratic problems*, Numer. Math., 59 (1991), pp. 561–580.
[9] G. H. Golub and U. von Matt, *Tikhonov regularization for large scale problems*, in Workshop on Scientific Computing, G. H. Golub, S. H. Lui, F. Luk, and R. Plemmons, eds., Springer, New York, 1997, pp. 3–26.
[10] G. H. Golub and G. Meurant, *Matrices, moments and quadrature*, in Numerical Analysis 1993, D. F. Griffiths and G. A. Watson, eds., Longman, Essex, England, 1994, pp. 105–156.
[11] M. Hanke, *A note on Tikhonov regularization of large linear problems*, BIT, 43 (2003), pp. 449–451.
[12] P. C. Hansen, *Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35.
[13] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion,* SIAM Monogr. Math. Model. Comput. 4, SIAM, Philadelphia, 1997.
[14] 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.
[15] 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.
[16] D. L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, J. ACM, 9 (1962), pp. 84–97.
[17] M. Rojas, *A Large-Scale Trust-Region Approach to the Regularization of Discrete Ill-Posed Problems*, Ph.D. thesis, Technical report TR98-19, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1998.
[18] M. Rojas, S. A. Santos, and D. C. Sorensen, *A new matrix-free algorithm for the large-scale trust-region subproblem*, SIAM J. Optim., 11 (2000), pp. 611–646.
[19] M. Rojas and D. C. Sorensen, *A trust-region approach to regularization of large-scale discrete forms of ill-posed problems*, SIAM. J. Sci. Comput., 23 (2002), pp. 1842–1860.
[20] D. C. Sorensen, *Newton's method with a model trust region modification*, SIAM J. Numer. Anal., 19 (1982), pp. 409–426.