ERROR ESTIMATES FOR LARGE-SCALE ILL-POSED PROBLEMS

L. REICHEL*, G. RODRIGUEZ†, AND S. SEATZU‡

In memory of Gene H. Golub.

Abstract. The computation of an approximate solution of linear discrete ill-posed problems with contaminated data is delicate due to the possibility of severe error propagation. Tikhonov regularization seeks to reduce the sensitivity of the computed solution to errors in the data by replacing the given ill-posed problem by a nearby problem, whose solution is less sensitive to perturbation. This regularization method requires that a suitable value of the regularization parameter be chosen. Recently, Brezinski et al. [Numer. Algorithms (2008), in press] described new approaches to estimate the error in approximate solutions of linear systems of equations and applied these estimates to determine a suitable value of the regularization parameter in Tikhonov regularization when the approximate solution is computed with the aid of the singular value decomposition. This paper discusses applications of these and related error estimates to the solution of large-scale ill-posed problems when approximate solutions are computed by Tikhonov regularization based on partial Lanczos bidiagonalization of the matrix. The connection between partial Lanczos bidiagonalization and Gauss quadrature is utilized to determine inexpensive bounds for a family of error estimates.

Key words. Ill-posed problem, Tikhonov regularization, error estimate, Lanczos bidiagonalization, Gauss quadrature.

1. Introduction.

This paper considers the approximate solution of least-squares problems

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|A\boldsymbol{x}-\boldsymbol{b}\| \tag{1.1}$$

with a matrix $A \in \mathbb{R}^{m\times n}$ of ill-determined rank and a vector $\boldsymbol{b} \in \mathbb{R}^m$, which is contaminated by an error $\boldsymbol{e} \in \mathbb{R}^m$. In particular, A is severely ill-conditioned and may be rank-deficient. For ease of presentation, we will assume that $m \geq n$; however, our method also can be applied when $m < n$. Throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm or the associated induced matrix norm.

Least-squares problems with a matrix of ill-determined rank are commonly referred to as discrete ill-posed problems. They arise in science and engineering when one is interested in determining the cause of an observed effect. The vector $\boldsymbol{b}$ represents the observations; the error $\boldsymbol{e}$ in $\boldsymbol{b}$ may be caused by measurement inaccuracies, transmission errors of the data, and discretization errors. We will refer to $\boldsymbol{e}$ as "noise."

Let $\hat{\boldsymbol{b}} \in \mathbb{R}^m$ denote the unavailable noise-free vector associated with $\boldsymbol{b}$, i.e.,

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

and let $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ be the minimal-norm least-squares solution of the unavailable minimization problem with noise-free right-hand side,

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|A\boldsymbol{x}-\hat{\boldsymbol{b}}\|. \tag{1.3}$$

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

†Dipartimento di Matematica e Informatica, Università di Cagliari, Viale Merello 92, 09123 Cagliari, Italy. E-mail: rodriguez@unica.it. This work was supported by MIUR under the PRIN grant no. 2006017542-003 and by the University of Cagliari.

‡Dipartimento di Matematica e Informatica, Università di Cagliari, Viale Merello 92, 09123 Cagliari, Italy. E-mail: seatzu@unica.it. This work was supported by MIUR under the PRIN grant no. 2006017542-003 and by the University of Cagliari.

The vector $\hat{\boldsymbol{b}}$ is not required to be in $\mathcal{R}(A)$, the range of A. We would like to determine an approximation of $\hat{\boldsymbol{x}}$ by computing an approximate solution of (1.1). Note that due to the severe ill-conditioning of A and the error $\boldsymbol{e}$ in $\boldsymbol{b}$, the exact solution of (1.1) typically is not a meaningful approximation of $\hat{\boldsymbol{x}}$.

In order to be able to compute a useful approximate solution of (1.1), the minimization problem has to be replaced by a nearby problem that is less sensitive to the error in $\boldsymbol{b}$. Tikhonov regularization replaces (1.1) by the penalized least-squares problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}\{\|A\boldsymbol{x}-\boldsymbol{b}\|^2+\mu\|L\boldsymbol{x}\|^2\}, \tag{1.4}$$

where $\mu > 0$ is a regularization parameter and $L \in \mathbb{R}^{p\times n}$, $p \leq n$, a regularization operator. We will in the present paper use the operator $L = I$; other possible choices of L are discussed in, e.g., [17, 20, 23, 25]. The normal equations associated with (1.4), for $L = I$, are of the form

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

and their solution is given by

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

for any $\mu > 0$. The parameter μ determines how sensitive $\boldsymbol{x}_\mu$ is to the error in $\boldsymbol{b}$ and how close $\boldsymbol{x}_\mu$ is to the desired solution $\hat{\boldsymbol{x}}$ of (1.3). The determination of a suitable positive value of μ is part of the solution process.

Let $\boldsymbol{x} \in \mathbb{R}^n$ be an arbitrary but fixed vector, which we assume not to be explicitly known. Let the associated residual vector

$$\boldsymbol{r} := \hat{\boldsymbol{b}} - A\boldsymbol{x} \tag{1.7}$$

be available. We would like to determining an estimate of the norm of the difference $\boldsymbol{d} = \boldsymbol{x} - \hat{\boldsymbol{x}}$. Brezinski et al. [6] recently derived the following family of estimates,

$$\|\boldsymbol{d}\|^2 \approx \eta_\nu^2 := d_0^{\nu-1}d_1^{5-2\nu}d_2^{\nu-3}, \qquad \nu\in\mathbb{R}, \tag{1.8}$$

where

$$d_0 := \|\boldsymbol{r}\|^2, \qquad d_1 := \|A^T\boldsymbol{r}\|^2, \qquad d_2 := \|AA^T\boldsymbol{r}\|^2, \tag{1.9}$$

and showed that

$$\eta_{\nu_1}^2 < \eta_{\nu_2}^2 \quad \text{for} \quad \nu_1 < \nu_2.$$

The estimates

$$\eta_2 = \frac{\|\boldsymbol{r}\|\,\|A^T\boldsymbol{r}\|}{\|AA^T\boldsymbol{r}\|}, \qquad \eta_3 = \frac{\|\boldsymbol{r}\|^2}{\|A^T\boldsymbol{r}\|} \tag{1.10}$$

are of particular interest to us.

The error estimate η_3 was first discussed by Auchmuty [1]. Brezinski [4] proposed error estimates related to (1.8) for linear systems of equations with a square nonsingular matrix and an error-free right-hand side. Moreover, Brezinski et al. [5] discussed the application of estimates similar to (1.8) to the determination of a suitable value of the regularization parameter μ in Tikhonov regularization when the matrix A is square and nonsingular. Error estimates that are applicable to overdetermined linear systems (1.1) are presented in [6], where both solution by the singular value decomposition and the conjugate gradient method applied to the normal equations (1.5) are discussed. The present paper considers the solution of large-scale problems by Tikhonov regularization using partial Lanczos bidiagonalization and exploits the connection of the latter with Gauss quadrature.

In the application of the error estimates (1.8) to Tikhonov regularization described by Brezinski et al. [5, 6], the vector $\boldsymbol{x}$ in (1.7) is chosen to be the solution (1.6) of the Tikhonov minimization problem (1.4) with $L = I$. Since the noise-free vector $\hat{\boldsymbol{b}}$ is assumed not to be available, $\hat{\boldsymbol{b}}$ is replaced by $\boldsymbol{b}$ in (1.7). This gives the residual vector

$$\boldsymbol{r}_\mu := \boldsymbol{b} - A\boldsymbol{x}_\mu. \tag{1.11}$$

Substituting $\boldsymbol{r} = \boldsymbol{r}_\mu$ into (1.9) yields error estimates $\eta_\nu = \eta_\nu(\mu)$, $\nu \in \mathbb{R}$, which are functions of μ. For a fixed $\nu \in \mathbb{R}$, Brezinski et al. [5, 6] determine a value of the regularization parameter μ that minimizes $\mu \to \eta_\nu(\mu)$.

The residual vector (1.11) and the quantities (1.9) are inexpensive to evaluate for several values of the parameter μ when the singular value decomposition of A is available. This makes the computation of the estimates $\eta_\nu(\mu)$ for several values of μ and ν quite inexpensive. However, when A is large and its singular value decomposition is not available, the computation of the error estimates (1.8) in the context of Tikhonov regularization can be expensive, since $\boldsymbol{r} = \boldsymbol{r}_\mu$ is not explicitly known. We describe how to determine inexpensively computable bounds for the error estimates in this situation.

Our approach to large-scale Tikhonov regularization problems is based on first reducing the matrix A to a fairly small lower bidiagonal matrix by a few steps of Lanczos bidiagonalization. The connection between Lanczos bidiagonalization and Gauss quadrature then is exploited to determine inexpensively computable lower and upper bounds for the error estimates (1.8) for several values of μ. The number of Lanczos bidiagonalization steps is chosen to make the upper and lower bounds sufficiently close in a vicinity of the minimum of the error estimate. We determine μ by minimizing the average of the computed upper and lower bounds for the error estimate.

The connection between the symmetric Lanczos process and Gauss quadrature was first discussed by Golub [13], and a nice survey of this technique with a few applications are presented by Golub and Meurant [15]. A book on this topic by Golub and Meurant is in preparation. Recently López Lagomasino et al. [18] extended this approach to rational Gauss rules. Many formulas relevant for the development in the present paper can be found in [7], and we will refer to the latter work for some details.

If an accurate estimate of $\|\boldsymbol{e}\|$ is explicitly known and $\hat{\boldsymbol{b}} \in \mathcal{R}(A)$, then a suitable value of μ often can be determined by the discrepancy principle; see, e.g., [8, 10]. We are concerned with the approximate solution of large-scale minimization problems (1.1) when no accurate estimate of $\|\boldsymbol{e}\|$ is available or for which $\hat{\boldsymbol{b}} \notin \mathcal{R}(A)$. We note that since our criterion for determining μ does not use $\|\boldsymbol{e}\|$, it may fail for some problems; see [3] for a discussion. Nevertheless, numerous numerical experiments, some of which are reported in Section 5, show the proposed method to perform well for many problems and to be competitive with other schemes that do not use $\|\boldsymbol{e}\|$ for determining μ.

This paper is organized as follows. Section 2 expresses the quantities (1.9) as Stieltjes integrals and discusses how Gauss quadrature rules can be used to evaluate upper and lower bounds. The connection between partial Lanczos bidiagonalization of A and Gauss quadrature rules for these Stieltjes integrals is described in Section 3. Our algorithm is discussed in Section 4 and computed examples are presented in Section 5. Concluding remarks can be found in Section 6.

2. Stieltjes integrals and Gauss quadrature. Let the vectors $\boldsymbol{x}_\mu$ and $\boldsymbol{r}_\mu$ be given by (1.6) and (1.11), respectively. In order to represent $d_0 = d_0(\mu)$ as a Stieltjes integral, we first observe that the relation $A(A^TA + \mu I)^{-1} = (AA^T + \mu I)^{-1}A$ leads to the following representation of the norm of the residual vector

$$\|\boldsymbol{r}_\mu\|^2 = \|\boldsymbol{b} - A\boldsymbol{x}_\mu\|^2 = \mu^2 \boldsymbol{b}^T(AA^T + \mu I)^{-2}\boldsymbol{b}. \tag{2.1}$$

Introduce the spectral decomposition

$$AA^T = W\Lambda W^T \tag{2.2}$$

with $W \in \mathbb{R}^{m\times m}$, $W^TW = I$, and

$$\Lambda = \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_m], \qquad 0 \le \lambda_1 \le \lambda_2 \le \ldots \le \lambda_m.$$

If $m > n$, then $\lambda_1 = \lambda_2 = \ldots = \lambda_{m-n} = 0$. Substitute the spectral decomposition (2.2), the vector $\boldsymbol{w} = [w_1, w_2, \ldots, w_m]^T := W^T\boldsymbol{b}$, and the function

$$\phi_\mu(t) := \left(\frac{\mu}{t+\mu}\right)^2 \tag{2.3}$$

into the right-hand side of (2.1) to obtain

$$\begin{aligned} d_0(\mu) = \|\boldsymbol{r}_\mu\|^2 &= \mu^2 \boldsymbol{w}^T(\Lambda + \mu I)^{-2}\boldsymbol{w} \\ &= \sum_{j=1}^m \phi_\mu(\lambda_j)w_j^2 = \int_{-\infty}^{\infty} \phi_\mu(t)dw(t), \end{aligned} \tag{2.4}$$

where $dw(t)$ is a nonnegative measure. The associated distribution function, $w(t)$, is nondecreasing and piecewise constant with jumps at the eigenvalues λ_j; when λ_j is distinct, $w(t)$ has a jump of size w_j^2 at λ_j. The right-hand side of (2.4) is a Stieltjes integral with an integrand that allows the integral to be bracketed by Gauss quadrature rules; see below.

The norm of $\boldsymbol{x}_\mu$ also can be expressed as a Stieltjes integral. We have

$$\|\boldsymbol{x}_\mu\|^2 = \boldsymbol{x}_\mu^T\boldsymbol{x}_\mu = \boldsymbol{b}^TA(A^TA + \mu I)^{-2}A^T\boldsymbol{b}. \tag{2.5}$$

Consider the spectral decomposition

$$A^TA = \hat{W}\hat{\Lambda}\hat{W}^T \tag{2.6}$$

with $\hat{W} \in \mathbb{R}^{n\times n}$, $\hat{W}^T\hat{W} = I$, and

$$\hat{\Lambda} = \mathrm{diag}[\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_n], \qquad 0 \le \hat{\lambda}_1 \le \hat{\lambda}_2 \le \ldots \le \hat{\lambda}_n,$$

where $\hat{\lambda}_j = \lambda_{j+m-n}$ for $1 \le j \le n$. Substituting the decomposition (2.6), $\hat{\boldsymbol{w}} = [\hat{w}_1, \hat{w}_2, \ldots, \hat{w}_n]^T := \hat{W}^TA^T\boldsymbol{b}$, and the function (2.3) into the right-hand side of (2.5) yields

$$\|\boldsymbol{x}_\mu\|^2 = \hat{\boldsymbol{w}}^T(\hat{\Lambda} + \mu I)^{-2}\hat{\boldsymbol{w}} = \sum_{j=1}^n \frac{1}{\mu^2}\phi_\mu(\hat{\lambda}_j)\hat{w}_j^2 = \int_{-\infty}^{\infty} \frac{1}{\mu^2}\phi_\mu(t)d\hat{w}(t), \tag{2.7}$$

where $d\hat{w}(t)$ is a nonnegative measure analogous to $dw(t)$; the associated distribution function, $\hat{w}(t)$, is nondecreasing, piecewise constant, and has jumps at the eigenvalues $\hat{\lambda}_j$. The Stieltjes integral in the right-hand side of (2.7) can be bounded from above and below by Gauss quadrature rules; see below.

It follows from (1.5) that

$$A^T \boldsymbol{r}_\mu = \mu \boldsymbol{x}_\mu. \tag{2.8}$$

Therefore, $d_1 = d_1(\mu)$ can be expressed in terms of the Stieltjes integral (2.7),

$$d_1(\mu) := \|A^T \boldsymbol{r}_\mu\|^2 = \int_{-\infty}^{\infty} \phi_\mu(t) d\hat{w}(t),$$

and be bracketed in terms of Gauss quadrature rules.

Finally, application of the spectral decomposition (2.2) give

$$\begin{aligned} d_2(\mu) &= \|AA^T \boldsymbol{r}_\mu\|^2 = \mu^2 \|A\boldsymbol{x}_\mu\|^2 = \mu^2 \boldsymbol{b}^T AA^T (AA^T + \mu I)^{-2} AA^T \boldsymbol{b} \qquad (2.9) \\ &= \sum_{j=1}^{m} \phi_\mu(\lambda_j) \breve{w}_j^2 = \int_{-\infty}^{\infty} \phi_\mu(t) d\breve{w}(t), \qquad (2.10) \end{aligned}$$

where $\breve{\boldsymbol{w}} = [\breve{w}_1, \breve{w}_2, \ldots, \breve{w}_m]^T := W^T AA^T \boldsymbol{b}$ and the measure $d\breve{w}(t)$ is defined analogously as $dw(t)$.

Let $\mathcal{G}_{\ell_1}$, $\hat{\mathcal{G}}_{\ell_1}$, and $\breve{\mathcal{G}}_{\ell_1}$ denote ℓ_1-point Gauss quadrature rules associated with the measures $dw(t)$, $d\hat{w}(t)$, and $d\breve{w}(t)$, respectively, and let $\mathcal{R}_{\ell_2}$, $\hat{\mathcal{R}}_{\ell_2}$, and $\breve{\mathcal{R}}_{\ell_2}$ denote the corresponding ℓ_2-point Gauss-Radau rules with a prescribed node at the origin. The derivatives (with respect to t) of even order of the integrand (2.3) are positive on the positive real axis, which contains the support of the measures $dw(t)$, $d\hat{w}(t)$, and $d\breve{w}(t)$, and the derivatives of odd order are negative on the positive real axis. The remainder formulas for Gauss and Gauss-Radau rules therefore show these quadrature rules to provide upper or lower bounds for the integrals. Specifically,

$$\begin{array}{ccccccccc} \mathcal{G}_{\ell_1-1}(\phi_\mu) & < & \mathcal{G}_{\ell_1}(\phi_\mu) & < & d_0(\mu) & < & \mathcal{R}_{\ell_2}(\phi_\mu) & < & \mathcal{R}_{\ell_2-1}(\phi_\mu), \\ \hat{\mathcal{G}}_{\ell_1-1}(\phi_\mu) & < & \hat{\mathcal{G}}_{\ell_1}(\phi_\mu) & < & d_1(\mu) & < & \hat{\mathcal{R}}_{\ell_2}(\phi_\mu) & < & \hat{\mathcal{R}}_{\ell_2-1}(\phi_\mu), \\ \breve{\mathcal{G}}_{\ell_1-1}(\phi_\mu) & < & \breve{\mathcal{G}}_{\ell_1}(\phi_\mu) & < & d_2(\mu) & < & \breve{\mathcal{R}}_{\ell_2}(\phi_\mu) & < & \breve{\mathcal{R}}_{\ell_2-1}(\phi_\mu), \end{array} \tag{2.11}$$

where ℓ_1 and ℓ_2 are arbitrary positive integers smaller than the number of mass points of the measures; see, e.g., [7, 15, 18] for details.

We remark that the right-hand side of (2.9) can be expressed as a Stieltjes integral in several ways. For instance, we have the alternative representation

$$d_2(\mu) = \int_{-\infty}^{\infty} \psi_\mu(t) dw(t), \qquad \psi_\mu(t) := \left(\frac{t}{t+\mu}\right)^2. \tag{2.12}$$

However, higher order derivatives of $t \to \psi_\mu(t)$ are not of constant sign on the support of the measure for all values of $\mu > 0$ of interest. Pairs of Gauss and Gauss-Radau quadrature rules applied to the integral (2.12) therefore are not guaranteed to bracket $d_2(\mu)$. For this reason we have chosen to express $d_2(\mu)$ in terms of the Stieltjes integral (2.10).

3. Partial Lanczos bidiagonalization. We describe how Gauss and Gauss-Radau quadrature rules for the Stieltjes integrals introduced in the previous section can be determined via partial Lanczos bidiagonalization of A. A detailed discussion of Lanczos bidiagonalization is provided by Paige and Saunders [21]; properties relevant to the present paper also are discussed in [7].

Application of ℓ steps of Lanczos bidiagonalization to A with initial vector $\boldsymbol{b}$, with $\ell < n$ sufficiently small, gives the decompositions

$$AV_\ell = U_\ell C_\ell + \sigma_{\ell+1}\boldsymbol{u}_{\ell+1}\boldsymbol{e}_\ell^T, \qquad A^T U_\ell = V_\ell C_\ell^T, \tag{3.1}$$

where $U_\ell = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_\ell] \in \mathbb{R}^{m\times\ell}$ and $V_\ell = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_\ell] \in \mathbb{R}^{n\times\ell}$ satisfy

$$U_\ell^T U_\ell = I, \qquad V_\ell^T V_\ell = I, \qquad \boldsymbol{u}_1 = \boldsymbol{b}/\|\boldsymbol{b}\|, \qquad U_\ell^T \boldsymbol{u}_{\ell+1} = \boldsymbol{0}.$$

Moreover, $\|\boldsymbol{u}_{\ell+1}\| = 1$ and $\sigma_{\ell+1} \geq 0$. We will assume that $\sigma_{\ell+1} > 0$; otherwise the computations simplify. The rare situation when $\sigma_{\ell+1} = 0$ will be commented on at the end of this section. The matrix

$$C_\ell = \begin{bmatrix} \rho_1 & & & & 0 \\ \sigma_2 & \rho_2 & & & \\ & \ddots & \ddots & & \\ & & \sigma_{\ell-1} & \rho_{\ell-1} & \\ 0 & & & \sigma_\ell & \rho_\ell \end{bmatrix} \in \mathbb{R}^{\ell\times\ell} \tag{3.2}$$

is lower bidiagonal. It is convenient also to define the rectangular lower bidiagonal matrix

$$\bar{C}_\ell = \begin{bmatrix} C_\ell \\ \sigma_{\ell+1}\boldsymbol{e}_\ell^T \end{bmatrix} \in \mathbb{R}^{(\ell+1)\times\ell},$$

where $\boldsymbol{e}_j = [0, \ldots, 0, 1, 0, \ldots, 0]^T$ denotes the jth axis vector. For future reference, we note that

$$\mathcal{R}(V_\ell) = \mathbb{K}_\ell(A^T A, A^T\boldsymbol{b}), \tag{3.3}$$

where

$$\mathbb{K}_\ell(A^T A, A^T\boldsymbol{b}) := \text{span}\{A^T\boldsymbol{b}, (A^T A)A^T\boldsymbol{b}, \ldots, (A^T A)^{\ell-1}A^T\boldsymbol{b}\} \tag{3.4}$$

is a Krylov subspace.

Combining the equations (3.1) shows that

$$AA^T U_\ell = U_\ell C_\ell C_\ell^T + \sigma_{\ell+1}\rho_\ell \boldsymbol{u}_{\ell+1}\boldsymbol{e}_\ell^T, \tag{3.5}$$

where $\rho_\ell = \boldsymbol{e}_\ell^T C_\ell \boldsymbol{e}_\ell$.

The matrix $T_\ell := C_\ell C_\ell^T$ is symmetric and tridiagonal. It follows from (3.5) that the columns $\boldsymbol{u}_j$ of U_ℓ, as well as $\boldsymbol{u}_{\ell+1}$, satisfy a three-term recurrence relation, and that for certain polynomials $p_{j-1} \in \Pi_{j-1}$,

$$\boldsymbol{u}_j = p_{j-1}(AA^T)\boldsymbol{b}, \qquad 1 \leq j \leq \ell + 1.$$

Introduce the inner product

$$(f, g) := \int_{-\infty}^{\infty} f(t)g(t)dw(t).$$

Then

$$
\begin{aligned}
(p_{j-1}, p_{k-1}) &= \int_{-\infty}^{\infty} p_{j-1}(t) p_{k-1}(t) dw(t) = \boldsymbol{b}^T p_{j-1}(AA^T) p_{k-1}(AA^T) \boldsymbol{b} \\
&= \boldsymbol{u}_j^T \boldsymbol{u}_k = \left\{ \begin{array}{ll} 1, & j = k, \\ 0, & j \neq k, \end{array} \right.
\end{aligned}
$$

i.e., the p_j are orthonormal polynomials with respect to this inner product. The nontrivial entries of the matrix T_ℓ are recursion coefficients for the p_j. Hence, the ℓ-point Gauss rule associated with the measure $dw(t)$ can be expressed as

$$
\mathcal{G}_\ell(f) = \|\boldsymbol{b}\|^2 \boldsymbol{e}_1^T f(T_\ell) \boldsymbol{e}_1;
$$

see [7, 15] for details. In particular,

$$
\mathcal{G}_\ell(\phi_\mu) = \mu^2 \|\boldsymbol{b}\|^2 \boldsymbol{e}_1^T \phi_\mu(C_\ell C_\ell^T) \boldsymbol{e}_1 = \mu^2 \|\boldsymbol{b}\|^2 \boldsymbol{e}_1^T (C_\ell C_\ell^T + \mu I)^{-2} \boldsymbol{e}_1.
$$

Let $\boldsymbol{y}_\ell(\mu) := \mu(C_\ell C_\ell^T + \mu I)^{-1} \boldsymbol{e}_1$. Then

$$
\mathcal{G}_\ell(\phi_\mu) = \|\boldsymbol{b}\|^2 \boldsymbol{y}_\ell^T(\mu) \boldsymbol{y}_\ell(\mu).
$$

We compute $\boldsymbol{y}_\ell(\mu)$ as the solution of the least-squares problem

$$
\min_{\boldsymbol{y} \in \mathbb{R}^\ell} \left\| \left[\begin{array}{c} C_\ell^T \\ \mu^{1/2} I_\ell \end{array} \right] \boldsymbol{y} - \mu^{1/2} \boldsymbol{e}_{\ell+1} \right\|. \tag{3.6}
$$

Eldén [9] has described an algorithm for solving this kind of least-squares problem in only $\mathcal{O}(\ell)$ arithmetic floating point operations for each value of μ.

The $(\ell+1)$-point Gauss-Radau rule for the measure $dw(t)$ with a preassigned node at the origin, applied to the integrand ϕ_μ, can be expressed as

$$
\mathcal{R}_{\ell+1}(\phi_\mu) = \|\boldsymbol{b}\|^2 \boldsymbol{e}_1^T \phi_\mu(\bar{C}_\ell \bar{C}_\ell^T) \boldsymbol{e}_1;
$$

see, e.g., [7] for a proof. We compute $\mathcal{R}_{\ell+1}(\phi_\mu)$ for different values of $\mu > 0$ by solving a least-squares problem analogous to (3.6).

We turn to the evaluation of Gauss rules for the measure $d\hat{w}(t)$. Let $U_{\ell+1} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{\ell+1}] \in \mathbb{R}^{m \times (\ell+1)}$. Then the left-hand side equation in (3.1) can be written in the form

$$
AV_\ell = U_{\ell+1} \bar{C}_\ell. \tag{3.7}
$$

Multiplying this equation by A^T yields

$$
A^T A V_\ell = A^T U_{\ell+1} \bar{C}_\ell = V_{\ell+1} C_{\ell+1}^T \bar{C}_\ell = V_\ell \bar{C}_\ell^T \bar{C}_\ell + \rho_{\ell+1} \sigma_{\ell+1} \boldsymbol{v}_{\ell+1} \boldsymbol{e}_\ell^T, \tag{3.8}
$$

where $\rho_{\ell+1}$ is the last diagonal entry of $C_{\ell+1}$ and $\boldsymbol{v}_{\ell+1}$ is the last column of $V_{\ell+1}$. Determine the QR-factorization

$$
\bar{C}_\ell = \bar{Q}_\ell \hat{C}_\ell^T, \tag{3.9}
$$

e.g., by application of ℓ Givens rotations. This factorization, with $\bar{Q}_\ell \in \mathbb{R}^{(\ell+1) \times \ell}$ represented as a product of Givens rotations, can be computed in only $\mathcal{O}(\ell)$ arithmetic floating point operations. Thus, $\bar{Q}_\ell$ has orthonormal columns and $\hat{C}_\ell^T \in \mathbb{R}^{\ell\times\ell}$ is upper bidiagonal. Equation (3.8) now can be written as

$$A^T A V_\ell = V_\ell \hat{C}_\ell \hat{C}_\ell^T + \rho_{\ell+1}\sigma_{\ell+1}\boldsymbol{v}_{\ell+1}\boldsymbol{e}_\ell^T.$$

This relation is analogous to (3.5). The matrix $\hat{T}_\ell := \hat{C}_\ell \hat{C}_\ell^T$ is tridiagonal. Therefore the columns $\boldsymbol{v}_j$ of V_ℓ and $\boldsymbol{v}_{\ell+1}$ satisfy a three-term recurrence relation. The first column of V_ℓ is proportional to $A^T\boldsymbol{b}$. It follows that the polynomials $\hat{p}_{j-1} \in \Pi_{j-1}$ determined by the relations

$$\boldsymbol{v}_j = \hat{p}_{j-1}(A^T A)A^T\boldsymbol{b}, \qquad 1 \le j \le \ell+1,$$

are orthonormal with respect to the inner product

$$\langle f, g\rangle := \int_{-\infty}^{\infty} f(t)g(t)d\hat{w}(t) = \boldsymbol{b}^T A f(A^T A)g(A^T A)A^T\boldsymbol{b}.$$

The nontrivial entries of the matrix $\hat{T}_\ell$ are recurrence coefficients for the polynomials $\hat{p}_j$. It follows that the ℓ-point Gauss rule with respect to the measure $d\hat{w}(t)$ is given by

$$\hat{\mathcal{G}}_\ell(f) = \|A^T\boldsymbol{b}\|^2\boldsymbol{e}_1^T f(\hat{T}_\ell)\boldsymbol{e}_1.$$

In particular,

$$\hat{\mathcal{G}}_\ell(\phi_\mu) = \mu^2\|A^T\boldsymbol{b}\|^2\boldsymbol{e}_1^T(\hat{C}_\ell\hat{C}_\ell^T + \mu I)^{-2}\boldsymbol{e}_1.$$

Note that $\|A^T\boldsymbol{b}\| = \|\boldsymbol{b}\|\rho_1$, where ρ_1 is the leading diagonal entry of C_ℓ; see (3.2). Thus, $\|A^T\boldsymbol{b}\|$ can be determined cheaply without additional matrix-vector product evaluations. We compute $\hat{\mathcal{G}}_\ell(\phi_\mu)$ by solving a least-squares problem analogous to (3.6). The corresponding ℓ-point Gauss-Radau rule with one fixed node at the origin is given by

$$\hat{\mathcal{R}}_\ell(\phi_\mu) = \mu^2\|A^T\boldsymbol{b}\|^2\boldsymbol{e}_1^T(\bar{\hat{C}}_{\ell-1}(\bar{\hat{C}}_{\ell-1})^T + \mu I)^{-2}\boldsymbol{e}_1,$$

where $\bar{\hat{C}}_{\ell-1} \in \mathbb{R}^{\ell\times(\ell-1)}$ is obtained by removing the last column of $\hat{C}_\ell$. Also this quadrature rule is evaluated by solving a least-squares problem similar to (3.6).

We are in a position to discuss the computation of Gauss rules for the measure $d\breve{w}(t)$. Consider the QR-factorization of the lower bidiagonal matrix (3.2),

$$C_\ell = Q'_\ell R'_\ell. \tag{3.10}$$

The orthogonal matrix $Q'_\ell \in \mathbb{R}^{\ell\times\ell}$ can be represented as a product of $\ell-1$ Givens rotations, which shows that Q'_ℓ is of upper Hessenberg form. The matrix $R'_\ell \in \mathbb{R}^{\ell\times\ell}$ is upper bidiagonal and differs from the matrix $\hat{C}_\ell^T$ in (3.9) only in the last diagonal entry. Indeed, if the matrix $\hat{C}_\ell^T$ is computed with the aid of ℓ Givens rotations, then R'_ℓ is available when $\ell-1$ of these rotations have been applied.

Substitution of (3.10) into (3.5), with $U'_\ell := U_\ell Q'_\ell$, yields

$$AA^T U_\ell = U'_\ell R'_\ell C_\ell^T + \sigma_{\ell+1}\rho_\ell \boldsymbol{u}_{\ell+1}\boldsymbol{e}_\ell^T. \tag{3.11}$$

Assume that $\ell > 1$ and multiply this equation by $\boldsymbol{e}_1$. We obtain

$$AA^T\boldsymbol{u}_1 = U'_\ell R'_\ell C_\ell^T\boldsymbol{e}_1 = \gamma_\ell U'_\ell\boldsymbol{e}_1,$$

where the last equality follows from the fact that both matrices R'_ℓ and C_ℓ^T are upper triangular; the constant γ_ℓ is the product of the leading diagonal entries of R'_ℓ and C_ℓ^T. We conclude that the first column of U'_ℓ is proportional to $AA^T\boldsymbol{b}$.

Multiplying (3.11) from the right-hand side by Q'_ℓ gives

$$AA^T U'_\ell = U'_\ell R'_\ell (R'_\ell)^T + \sigma_{\ell+1}\rho_\ell \boldsymbol{u}_{\ell+1}\boldsymbol{e}_\ell^T Q'_\ell. \tag{3.12}$$

This is not a decomposition of the form (3.5), because the last two entries of the vector $\boldsymbol{e}_\ell^T Q'_\ell$ are nonvanishing. We bring the decomposition (3.12) into the form of (3.5) in two steps, the first of which entails the computation of the QR-factorization

$$(R'_\ell)^T = Q''_\ell R''_\ell, \tag{3.13}$$

where $Q''_\ell \in \mathbb{R}^{\ell\times\ell}$ is orthogonal and $R''_\ell \in \mathbb{R}^{\ell\times\ell}$ is upper bidiagonal. Substituting (3.13) into (3.12) yields

$$AA^T U'_\ell = U'_\ell (R''_\ell)^T R''_\ell + \sigma_{\ell+1}\rho_\ell \boldsymbol{u}_{\ell+1}\boldsymbol{e}_\ell^T Q'_\ell. \tag{3.14}$$

The second steps consists of removing the last column in every term of (3.14). Let the matrix $\breve{U}_{\ell-1} \in \mathbb{R}^{m\times(\ell-1)}$ be the leading $m \times (\ell-1)$ submatrix of U'_ℓ, and let $\breve{R}_{\ell-1}$ be the leading $(\ell-1)\times(\ell-1)$ principal submatrix of R''_ℓ. Then (3.14) yields

$$AA^T \breve{U}_{\ell-1} = \breve{U}_{\ell-1}\breve{R}_{\ell-1}^T \breve{R}_{\ell-1} + \boldsymbol{f}_{\ell-1}\boldsymbol{e}_{\ell-1}^T,$$

where the vector $\boldsymbol{f}_{\ell-1} \in \mathbb{R}^m$ is orthogonal to the columns of $\breve{U}_{\ell-1}$. This is the desired decomposition. It is analogous to (3.5) and shows that the $(\ell-1)$-point Gauss quadrature rule associated with the measure $d\breve{w}(t)$ is given by

$$\breve{\mathcal{G}}_{\ell-1}(f) = \|AA^T\boldsymbol{b}\|^2 \boldsymbol{e}_1^T f(\breve{R}_{\ell-1}^T \breve{R}_{\ell-1})\boldsymbol{e}_1. \tag{3.15}$$

The corresponding Gauss-Radau rule with one prescribed node at the origin can be expressed as

$$\breve{\mathcal{R}}_{\ell-1}(f) = \|AA^T\boldsymbol{b}\|^2 \boldsymbol{e}_1^T f(\breve{R}_{\ell-1,0}^T \breve{R}_{\ell-1,0})\boldsymbol{e}_1, \tag{3.16}$$

where the matrix $\breve{R}_{\ell-1,0} \in \mathbb{R}^{(\ell-2)\times(\ell-1)}$ is obtained by removing the last row from $\breve{R}_{\ell-1}$. Equivalently, we can define $\breve{R}_{\ell-1,0}$ by setting the last diagonal entry of $\breve{R}_{\ell-1}$ to zero. It follows from (3.1)-(3.2) that

$$\|AA^T\boldsymbol{b}\|^2 = \|\boldsymbol{b}\|^2 \rho_1^2(\rho_1^2 + \sigma_2^2).$$

Thus, $\|AA^T\boldsymbol{b}\|^2$ can be evaluated inexpensively by using entries of C_ℓ. Note that the computation of bounds for $d_2 = d_2(\mu)$ only requires the quadrature rules (3.15) and (3.16), not the matrices U'_ℓ and $\breve{U}_{\ell-1}$. We therefore do not have to compute the latter.

The measure $d\breve{w}(t)$ is obtained by multiplying $dw(t)$ by the factor t. Several derivations of algorithms for modifying the symmetric tridiagonal matrix T_ℓ associated with $dw(t)$ to obtain the symmetric tridiagonal matrix $\breve{T}_{\ell-1} := \breve{R}_{\ell-1}^T \breve{R}_{\ell-1}$ associated with the measure $d\breve{w}(t)$ are available; see Gautschi [11, 12] and Golub and Kautsky [14]. Our derivation is suitable in the context of bounding matrix functionals and differs from the approaches in the references mentioned.

THEOREM 3.1. *The following upper and lower bounds for the quantities (1.9) can be computed after ℓ steps of Lanczos bidiagonalization applied to A with initial vector $\boldsymbol{b}$:*

$$\begin{array}{rcccl} \mathcal{G}_\ell(\phi_\mu) & < & d_0(\mu) & < & \mathcal{R}_{\ell+1}(\phi_\mu), \\ \hat{\mathcal{G}}_\ell(\phi_\mu) & < & d_1(\mu) & < & \hat{\mathcal{R}}_\ell(\phi_\mu), \\ \breve{\mathcal{G}}_{\ell-1}(\phi_\mu) & < & d_2(\mu) & < & \breve{\mathcal{R}}_{\ell-1}(\phi_\mu). \end{array} \tag{3.17}$$

These bounds allow the computation of upper and lower bounds for the error estimates (1.8) for any $\nu \in \mathbb{R}$. In particular, writing $\eta_\nu = \eta_\nu(\mu)$ in order to explicitly indicate the dependence of the estimates (1.10) on μ, we obtain

$$\begin{array}{ccccc} \left(\dfrac{\mathcal{G}_\ell(\phi_\mu)\,\hat{\mathcal{G}}_\ell(\phi_\mu)}{\breve{\mathcal{R}}_{\ell-1}(\phi_\mu)}\right)^{1/2} & < & \eta_2(\mu) & < & \left(\dfrac{\mathcal{R}_{\ell+1}(\phi_\mu)\,\hat{\mathcal{R}}_\ell(\phi_\mu)}{\breve{\mathcal{G}}_{\ell-1}(\phi_\mu)}\right)^{1/2}, \\ \dfrac{\mathcal{G}_\ell(\phi_\mu)}{(\hat{\mathcal{R}}_\ell(\phi_\mu))^{1/2}} & < & \eta_3(\mu) & < & \dfrac{\mathcal{R}_{\ell+1}(\phi_\mu)}{(\hat{\mathcal{G}}_\ell(\phi_\mu))^{1/2}}. \end{array} \tag{3.18}$$

Proof. The discussion of this section shows that the quadrature rules in (3.17) and (3.18) can be evaluated when ℓ Lanczos bidiagonalization steps have been carried out. The bounds follow from the inequalities (2.11). ∎

We finally consider the situation when $\sigma_{\ell+1} = 0$ in (3.1). Then $\boldsymbol{b}^T p(AA^T)\boldsymbol{b} = \|\boldsymbol{b}\|^2 \boldsymbol{e}_1^T p(T_\ell)\boldsymbol{e}_1$ for all polynomials p, and therefore also for all functions f that can be approximated arbitrarily well by polynomials.

4. The numerical method. The QR-factorizations (3.9), (3.10), and (3.13) are computed by the application of a judiciously chosen sequence of simple orthogonal transformations. For instance, the upper bidiagonal matrices

$$R'_\ell = \begin{bmatrix} \delta_1 & \gamma_2 & & 0 \\ & \ddots & \ddots & \\ & & \delta_{\ell-1} & \gamma_\ell \\ 0 & & & \bar{\delta}_\ell \end{bmatrix} \quad \text{and} \quad \hat{C}_\ell^T = \begin{bmatrix} \delta_1 & \gamma_2 & & 0 \\ & \ddots & \ddots & \\ & & \delta_{\ell-1} & \gamma_\ell \\ 0 & & & \delta_\ell \end{bmatrix}$$

only differ in the last diagonal entry. Assuming that R'_ℓ is nonsingular, we can determine $\hat{C}_\ell^T$ and $R'_{\ell+1}$ by first adding a new row to R'_ℓ and then updating the extended matrix by applying the reflection

$$\begin{bmatrix} c_\ell & s_\ell \\ s_\ell & -c_\ell \end{bmatrix} \begin{bmatrix} \bar{\delta}_\ell & 0 \\ \sigma_{\ell+1} & \rho_{\ell+1} \end{bmatrix} = \begin{bmatrix} \delta_\ell & \gamma_{\ell+1} \\ 0 & \bar{\delta}_{\ell+1} \end{bmatrix},$$

where

$$\delta_\ell := c_\ell \bar{\delta}_\ell + s_\ell \sigma_{\ell+1}, \qquad \gamma_{\ell+1} := s_\ell \rho_{\ell+1}, \qquad \bar{\delta}_{\ell+1} := -c_\ell \rho_{\ell+1}.$$

The constants c_ℓ and s_ℓ are chosen to annihilate the $(\ell+1,\ell)$-entry of $C_{\ell+1}$. The process is initialized by setting $\bar{\delta}_1 = \rho_1$.

The upper bidiagonal factor R''_ℓ in (3.13) is computed by application of a sequence of $\ell - 1$ Givens rotations. Their product, in reverse order of application, forms the orthogonal matrix Q''_ℓ. Note that the latter matrix does not have to be explicitly formed. All the QR-factorizations (3.9), (3.10), and (3.13) can be determined in only $\mathcal{O}(\ell)$ arithmetic floating point operations. It follows that the bounds (3.1) can be evaluated in only $\mathcal{O}(\ell)$ arithmetic floating point operations for each value of $\mu > 0$.

For a chosen value of ν for the error estimate $\eta_\nu = \eta_\nu(\mu)$ defined by (1.8), we determine suitable values of the number of Lanczos bidiagonalization steps, ℓ, and of the regularization parameter, μ, in the following manner. We start with a grid of q log-equispaced values of μ,

$$0 < \mu_1 < \mu_2 < \ldots < \mu_q, \tag{4.1}$$

in an interval $[\mu_{\min}, \mu_{\max}]$. The grid points are partitioned into two sets, $\mathcal{C}$ and $\mathcal{N}$, depending on whether the upper and lower bounds $\eta_\nu^l(\mu_j)$ and $\eta_\nu^u(\mu_j)$, respectively, of $\eta_\nu(\mu_j)$ are close enough to be considered converged; see below. Note that the integrand (2.3) is analytic in the extended complex plane, except at the point $t = -\mu$. This singularity is further away from the interval of integration (a subset of the positive real axis) when μ is large than when $\mu > 0$ is small. Therefore, the Gauss and Gauss-Radau rules deliver more accurate bounds for large values of μ than for small ones. It follows that for a fixed value of ℓ, the set $\mathcal{C}$ of grid points with converged bounds typically is made up of some of the larger values of μ_j in the sequence (4.1).

Our algorithm consists of two phases. The first phase computes a rough estimate of the minimum μ^* of $\eta_\nu(\mu)$. This estimate is improved in the second phase. We initialize $\mathcal{N} := \{\mu_1, \ldots, \mu_q\}$ and $\mathcal{C} := \emptyset$ and carry out a few Lanczos bidiagonalization steps. At step ℓ, we perform the following additional computations:

(i) For each μ_j in $\mathcal{N}$, compute the lower and upper bounds $\eta_\nu^l(\mu_j)$ and $\eta_\nu^u(\mu_j)$, respectively, as well as their averages $\bar{\eta}_\nu(\mu_j)$.

(ii) Determine the smallest index k, such that

$$\left|\eta_\nu^u(\mu_j) - \eta_\nu^l(\mu_j)\right| < \beta \cdot \bar{\eta}_\nu(\mu_j), \qquad j = k, k+1, \ldots, \#\mathcal{N}, \tag{4.2}$$

where $\#\mathcal{N}$ denotes the cardinality of the set $\mathcal{N}$ and $\beta > 0$ is a user-specified tolerance.

(iii) If condition (4.2) holds for $k \leq \#\mathcal{N}$, then move the parameters $\{\mu_k, \ldots, \mu_{\#\mathcal{N}}\}$ from the set $\mathcal{N}$ to the set $\mathcal{C}$; this indicates that we consider $\bar{\eta}_\nu(\mu_j)$ a converged approximation of $\eta_\nu(\mu_j)$ for $j = k, \ldots, \#\mathcal{N}$.

(iv) Let $\ell := \ell + 1$ in (3.1), i.e., carry out another Lanczos bidiagonalization step and repeat the computations of steps (i)-(iii) until

$$\bar{\eta}_\nu(\mu_k) > \bar{\eta}_\nu(\mu_{k+1})$$

for some pair $\{\mu_k, \mu_{k+1}\} \subset \mathcal{C}$. Thus, we carry out an increasing number of Lanczos bidiagonalization steps until we have determined a local minimum μ^* of the function $\bar{\eta}_\nu(\mu)$ at an internal point of $\mathcal{C}$. This ends the first phase of our scheme. The dominating computational work for this phase is the evaluation of matrix-vector products with the matrices A and A^T. Each Lanczos bidiagonalization step requires the evaluation of one matrix-vector product with A and one with A^T.

The above algorithm assumes that $\bar{\eta}_\nu(\mu)$ is convex in an interval around the minimal value of $\eta_\nu(\mu)$. In the computed examples of Section 5, we used the parameter values $q = 10$, $\mu_{\min} := 1 \cdot 10^{-24}$, $\mu_{\max} := 1 \cdot 10^{4}$, $\mu_1 := \mu_{\min}$, $\mu_q := \mu_{\max}$, and

$\beta = 1 \cdot 10^{-2}$. We found these parameter values to be suitable for a large number of discrete ill-posed problems.

The second phase of our scheme improves the approximate local minimum μ^* of $\bar{\eta}_\nu(\mu)$ determined in phase one. This is achieved by refining the grid $\mathcal{C}$ by adding points around μ^*. We do this by bisection in log-scale and update the value μ^* until the minimum distance between adjacent grid points around μ^* in $\mathcal{C}$ reaches a prescribed (small) tolerance. The computations of this phase typically do not require that additional Lanczos bidiagonalization steps be carried out and therefore are inexpensive.

Assume now that a suitable value $\mu = \mu^*$ of the regularization parameter has been determined. We turn to the computation of an approximate solution of (1.5). The partial Lanczos decomposition (3.1) of A thus is available and we describe a Galerkin method, which uses this decomposition to compute an approximation $\boldsymbol{x}_{\mu,\ell}$ of $\boldsymbol{x}_\mu$ in the Krylov subspace (3.4). In particular, $\boldsymbol{x}_{\mu,\ell} \in \mathcal{R}(A^T)$ and, therefore, $\boldsymbol{x}_{\mu,\ell}$ is orthogonal to the null space of A.

In view of (3.3), we have $\boldsymbol{x}_{\mu,\ell} = V_\ell \boldsymbol{y}$ for some $\boldsymbol{y} \in \mathbb{R}^\ell$, and we determine $\boldsymbol{x}_{\mu,\ell}$ by imposing the Galerkin condition

$$V_\ell^T (A^T A + \mu I) V_\ell \boldsymbol{y} = V_\ell^T A^T \boldsymbol{b}.$$

Using (3.1) and (3.7), this equation can be reduced to

$$(\bar{C}_\ell^T \bar{C}_\ell + \mu I)\boldsymbol{y} = \|\boldsymbol{b}\| \bar{C}_\ell^T \boldsymbol{e}_1. \tag{4.3}$$

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

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

We compute the solution $\boldsymbol{y}_{\mu,\ell}$ of the reduced Galerkin equation (4.3) by solving this least-squares problem, whose solution can be computed in the same manner as (3.6). When doing this, we may use the QR-factorization (3.9) of $\bar{C}_\ell$.

Theorem 4.1. *The error estimate $\eta_3 = \eta_3(\mu)$ can be written as*

$$\eta_3(\mu) = \frac{\|\boldsymbol{r}_\mu\|^2}{\mu \|\boldsymbol{x}_\mu\|}. \tag{4.4}$$

Let $\boldsymbol{x}_{\mu,\ell} := V_\ell \boldsymbol{y}_{\mu,\ell}$ be the computed approximate solution of (1.1) with $\boldsymbol{y}_{\mu,\ell}$ the solution of (4.3). Introduce

$$\eta_{3,\ell}(\mu) := \frac{\|\boldsymbol{r}_{\mu,\ell}\|^2}{\mu \|\boldsymbol{x}_{\mu,\ell}\|}, \tag{4.5}$$

where $\boldsymbol{r}_{\mu,\ell} := \boldsymbol{b} - A\boldsymbol{x}_{\mu,\ell}$. Then $\eta_{3,\ell}(\mu)$ achieves the upper bound (3.18) for $\eta_3(\mu)$.

Proof. The representation (4.4) is a consequence of (1.10) and (2.8). It follows from [7, Theorem 5.1] that

$$\|\boldsymbol{x}_{\mu,\ell}\|^2 = \frac{1}{\mu^2} \hat{\mathcal{G}}_\ell(\phi_\mu), \qquad \|\boldsymbol{r}_{\mu,\ell}\|^2 = \mathcal{R}_{\ell+1}(\phi_\mu).$$

Substituting these expressions into the right-hand side of (4.5) yields

$$\eta_{3,\ell}(\mu) = \frac{\mathcal{R}_{\ell+1}(\phi_\mu)}{(\hat{\mathcal{G}}_\ell(\phi_\mu))^{1/2}}.$$

The right-hand side is equal to the upper bound (3.18) for $\eta_3(\mu)$. $\square$

We remark that, generally, $A^T \boldsymbol{r}_{\mu,\ell} - \mu \boldsymbol{x}_{\mu,\ell} \neq 0$, i.e., relation (2.8) does not hold for the computed solution $\boldsymbol{x}_{\mu,\ell}$ and the associated residual $\boldsymbol{r}_{\mu,\ell}$.

5. Computed examples. The algorithm described in the previous section has been implemented in MATLAB in order to assess its performance. The MATLAB code is available, upon request, from the authors.

The algorithm is applied to a selection of discrete ill-posed problems, most of which are standard test problems from [16]. The noise-free right-hand side $\hat{\boldsymbol{b}}$, see (1.2), is available in our test problems, and the "noise-vector" with noise-level κ is of the form $\boldsymbol{e} = \kappa \|\hat{\boldsymbol{b}}\| \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is a vector of normally distributed random variables with zero mean and of unit variance.

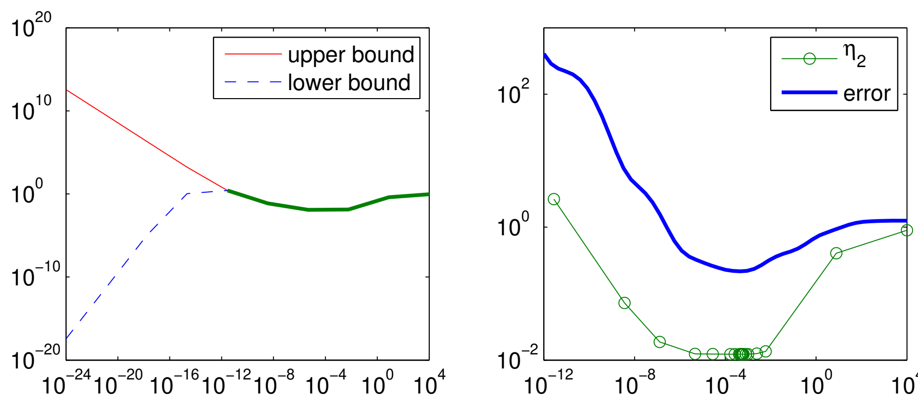

FIG. 5.1. *The left-hand side graphs illustrate the first phase of our algorithm for determining μ^* and the right-hand side graphs the second phase. The μ-values are given by the horizontal axes.*

In our first test problem, we are concerned with the solution of a Fredholm integral equation of the first kind with a nonsymmetric analytic kernel described by Baart [2]. We use MATLAB code from [16] to discretize the integral equation to obtain a discrete ill-posed problem (1.1) with a matrix $A \in \mathbb{R}^{200 \times 200}$. The noise-level is $\kappa = 1 \cdot 10^{-2}$. We refer to this problem as *Baart*. The regularization parameter μ is determined by minimizing $\eta_2(\mu)$.

The left-hand side graphs of Figure 5.1 illustrate the first phase of the algorithm. The graphs show computed upper and lower bounds for $\eta_2(\mu)$ obtained with 7 steps of Lanczos bidiagonalization, which were required to determine the local minimum μ^* of $\bar{\eta}_2(\mu)$ on the coarse grid. The thick graph depicts the averages $\bar{\eta}_2(\mu)$ in an interval containing $\mathcal{C}$. These averages are considered converged. The upper and lower bounds for $\eta_2(\mu)$, i.e., $\eta_2^u(\mu)$ and $\eta_2^l(\mu)$, are displayed in an interval containing $\mathcal{N}$ by the thin continuous and dashed graphs, respectively. We apply Lanczos bidiagonalization with reorthogonalization. The latter can reduce the number of bidiagonalization steps necessary significantly.

The right-hand side graphs of Figure 5.1 display the refinement of the minimum in the second phase of the algorithm. The circles show where $\bar{\eta}_2(\mu)$ is evaluated. Adjacent circles are connected by straight lines for visual clarity; however, the lines are of no significance. The thick curve depicts the error

$$e(\mu) := \|\hat{\boldsymbol{x}} - \boldsymbol{x}_{\mu,\ell}\| \tag{5.1}$$

for $\ell = 7$ as a function of μ. The figure shows that $\bar{\eta}_2(\mu)$ and $e(\mu)$ achieve their minima for close μ-values, even though $\bar{\eta}_2(\mu)$ is much smaller than $e(\mu)$.

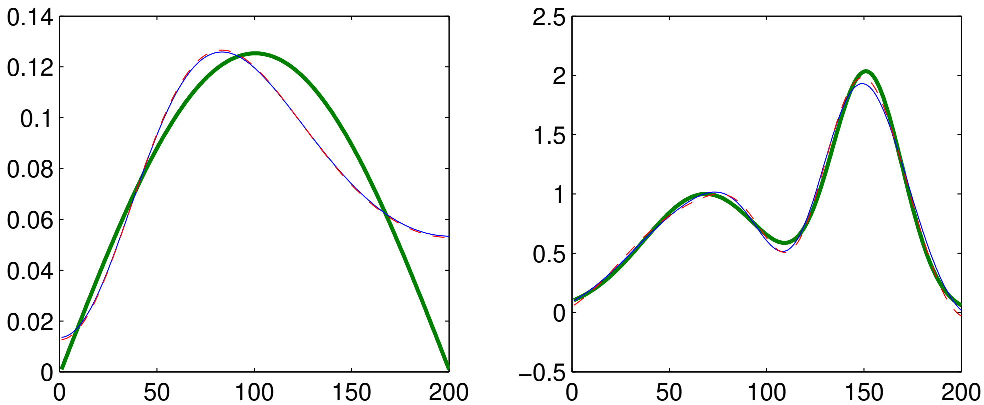

FIG. 5.2. *Computed and desired solutions for the* Baart *and* Shaw *test problems.*

The left-hand side graphs of Figure 5.2 display the solution $\hat{\boldsymbol{x}}$ of the noise-free problem (1.3) (thick curve) and the solution determined by our algorithm (thin curve, $\mu = 5.6 \cdot 10^{-4}$). We also show the solution $\boldsymbol{x}_{\mu,7}$, which is obtained by minimizing the error (5.1) over $\mu > 0$ for $\ell = 7$; the dashed curve of the figure shows this solution. The associated regularization parameter is $\mu = 4.4 \cdot 10^{-4}$. Note that, despite that $\bar{\eta}_2(\mu)$ is not a good approximation of $e(\mu)$, the thin continuous and dashed graphs are very close.

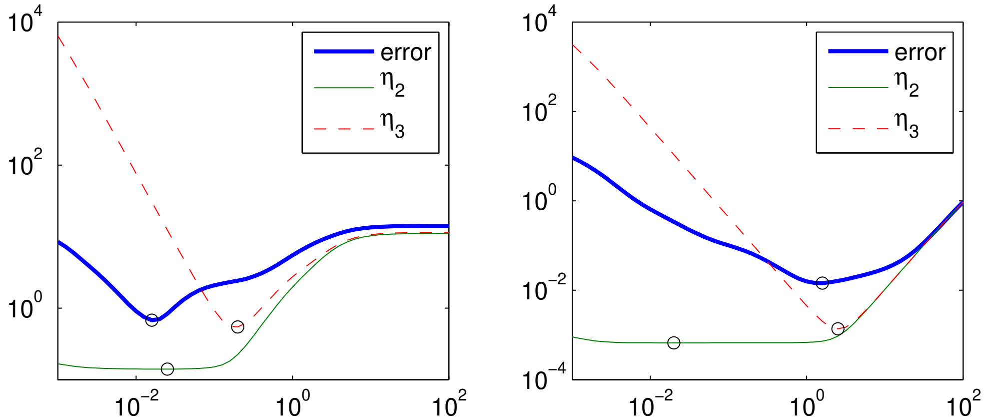

FIG. 5.3. *Error $e(\mu)$ (thick curves) and error estimates $\bar{\eta}_2(\mu)$ and $\bar{\eta}_3(\mu)$ for the* Shaw *test problem (left-hand side graphs) and the* Gauss *test problem (right-hand side graphs).*

Our second test problem is a Fredholm integral equation of the first kind with an analytic kernel described by Shaw [26]. MATLAB code for its discretization is available in [16]. We use this code to determine a discrete ill-posed problem with a matrix $A \in \mathbb{R}^{200\times 200}$. The error $\boldsymbol{e}$ with noise-level $\kappa = 1 \cdot 10^{-2}$ is defined similarly as above. We refer to this problem as *Shaw*. The left-hand side graph of Figure 5.3 displays the error $e(\mu)$ (thick curve) as well as the error estimates $\bar{\eta}_2(\mu)$ and $\bar{\eta}_3(\mu)$. The minima of these curves are marked by circles. The curves show the minimum of

$\bar{\eta}_2(\mu)$ to provide a better approximation of the minimum of $e(\mu)$ than the minimum of $\bar{\eta}_3(\mu)$.

The right-hand side graphs of Figure 5.2 show the solution $\hat{\boldsymbol{x}}$ of the noise-free problem (1.3) (thick curve) and the solution determined by our algorithm (thin curve, $\mu = 5.4 \cdot 10^{-4}$). The computation of the latter required 10 Lanczos bidiagonalization steps. The thin dashed curve is obtained by minimizing (5.1) over $\mu > 0$ for $\ell = 10$. The μ-value determined in this manner is $\mu = 3.2 \cdot 10^{-4}$; it is in Table 5.1 referred to as μ_{opt}. The thin continuous and dashed curves are close.

κ	10^{-2}	10^{-4}	10^{-6}	10^{-8}	10^{-10}	10^{-12}
ℓ	10	14	15	16	18	19
μ	$5.4 \cdot 10^{-4}$	$5.4 \cdot 10^{-8}$	$5.3 \cdot 10^{-12}$	$2.2 \cdot 10^{-15}$	$5.3 \cdot 10^{-20}$	$5.2 \cdot 10^{-24}$
$e(\mu)$	$7.8 \cdot 10^{-1}$	$3.5 \cdot 10^{-1}$	$8.3 \cdot 10^{-1}$	$1.2 \cdot 10^{-1}$	$3.8 \cdot 10^{-1}$	$5.2 \cdot 10^{-1}$
μ_{opt}	$3.2 \cdot 10^{-4}$	$3.0 \cdot 10^{-8}$	$1.1 \cdot 10^{-9}$	$4.2 \cdot 10^{-14}$	$4.4 \cdot 10^{-18}$	$1.1 \cdot 10^{-19}$
$e(\mu_{\mathrm{opt}})$	$6.7 \cdot 10^{-1}$	$3.3 \cdot 10^{-1}$	$2.8 \cdot 10^{-1}$	$7.9 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$	$1.2 \cdot 10^{-2}$

Table 5.1

Results for the Shaw *test problem for several noise-levels κ. The parameter ℓ shows the number of Lanczos bidiagonalization steps required, the computed value of the regularization parameter is denoted by μ, and the value of the regularization parameter that minimizes the error $e(\mu)$ is referred to as μ_{opt}.*

Table 5.1 illustrates the performance of our method for different noise-levels. The table shows that the method also is applicable for small noise-levels. The number of Lanczos bidiagonalization steps, ℓ, is seen to increase when the noise-level κ decreases, because the discrete ill-posed problems (1.1) are solved more accurately for small noise-levels.

In the above test problems, the singular values of the matrices decrease rapidly with increasing index. We now illustrate the performance of our method for a test problem for which the singular values decay quite slowly. Specifically, we consider discrete ill-posed problems with a Gaussian matrix with entries

$$A = [a_{ij}] \in \mathbb{R}^{m \times n}, \qquad a_{ij} = \sqrt{\frac{\pi}{2\alpha}}\, \exp(\frac{\alpha}{2}(i-j)^2), \qquad \alpha = 1 \cdot 10^{-2}.$$

The asymptotic condition number is of order 10^{214}; see [19] for details. We refer to this test problem as *Gauss*.

Letting $m = 400$ and $n = 200$ yields the matrix in (1.1). The desired solution $\hat{\boldsymbol{x}}$ is a sampling of the function $\sin(t)$ at 200 equidistant points in the interval $[0, \pi]$. The right-hand side $\hat{\boldsymbol{b}}$ in (1.3) is chosen to make the noise-free discrete ill-posed problem (1.3) consistent or inconsistent. Let

$$\tau := \|\hat{\boldsymbol{b}} - A\hat{\boldsymbol{x}}\|.$$

Thus, the minimization problem (1.3) is consistent when $\tau = 0$. Vectors $\hat{\boldsymbol{b}}$ that give an inconsistent problem with a prescribed positive value of τ are constructed as described in [24, p. 744]. A noise-vector $\boldsymbol{e}$ is determined similarly as in the above test problems and scaled to correspond to the noise-level $\kappa = 1 \cdot 10^{-2}$.

Figure 5.4 displays three computed solutions corresponding to $\tau = 0$, $\tau = 1 \cdot 10^2$, and $\tau = 1 \cdot 10^4$, as well as the desired solution $\hat{\boldsymbol{x}}$. The latter is depicted by a thick

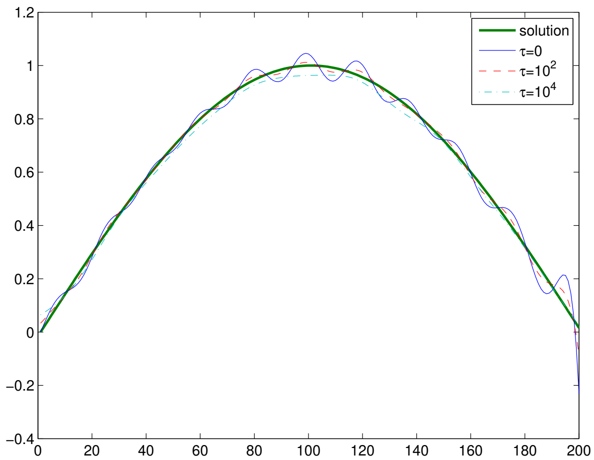

FIG. 5.4. *Computed and desired solutions for the* Gauss *test problem.*

curve and referred to as “solution” in the legend. The figure shows our method to produce meaningful approximate solutions also for inconsistent problems. The method requires 30 Lanczos bidiagonalization steps for $\tau = 0$ and $\tau = 1 \cdot 10^2$, and 26 steps for $\tau = 10^4$. The computed values of the regularization parameters are, for increasing values of τ, $\mu = 7.7$, $\mu = 4.2 \cdot 10^1$, and $\mu = 3.0 \cdot 10^3$. We remark that, when $\tau = 0$, the algorithm without reorthogonalization requires 101 Lanczos bidiagonalization steps.

The right-hand side graph of Figure 5.3 displays the error $e(\mu)$ (thick curve) as well as the error estimates $\bar{\eta}_2(\mu)$ and $\bar{\eta}_3(\mu)$. The minima of these curves are marked by circles. The curves show the minimum of $\bar{\eta}_3(\mu)$ to provide a better approximation of the minimum of $e(\mu)$ than the minimum of $\bar{\eta}_2(\mu)$. We found in numerous numerical examples the minimum of at least one of the error estimates $\bar{\eta}_2(\mu)$ and $\bar{\eta}_3(\mu)$ to furnish a useful value of the regularization parameter.

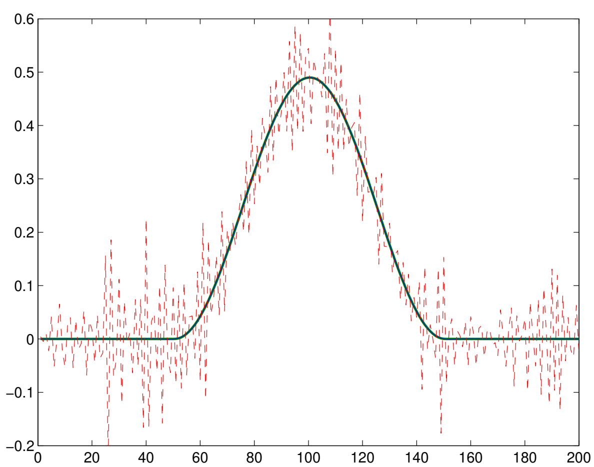

FIG. 5.5. *Desired and approximate solutions determined by the L-curve criterion and our method.*

The final example compares our method to the L-curve criterion. The latter is a popular approach to determining μ when no estimate of the norm of the noise is available. A thorough discussion of the L-curve criterion and many illustrations are provided by Hansen [17]. We consider a discretization of the integral equation *Phillips* described in [22] and use MATLAB code from [16] to determine $A \in \mathbb{R}^{200\times 200}$ and $\hat{\boldsymbol{b}} \in \mathbb{R}^{200}$. A noise-vector $\boldsymbol{e}$ of the same type as above, with noise-level $\kappa = 1\cdot 10^{-6}$, is added to $\hat{\boldsymbol{b}}$ to give the vector $\boldsymbol{b}$ in (1.1). Figure 5.5 shows the desired solution $\hat{\boldsymbol{x}}$ (thick curve) and the approximate solution $\boldsymbol{x}_\mu$ determined by our method based on determining upper and lower bounds for $\eta_3(\mu)$ (thin continuous curve). This approach yields $\mu = 1\cdot 10^{-5}$. The thick and thin curves are indistinguishable; we have $\|\boldsymbol{x}_\mu - \hat{\boldsymbol{x}}\| = 2.1\cdot 10^{-3}$.

Let $\boldsymbol{x}_\mu$ and $\boldsymbol{r}_\mu$ be the solution (1.6) of the Tikhonov equation and the associated residual vector (1.11), respectively. The L-curve criterion determines the point on the curve

$$\mu \to \{\log\|\boldsymbol{r}_\mu\|, \log\|\boldsymbol{x}_\mu\|\}, \qquad \mu > 0,$$

where the curvature is of largest magnitude. The L-curve often looks like the letter "L", and the desired μ-value, which we refer to as μ_L, corresponds to the point at the "vertex." The choice $\mu = \mu_L$ seeks to balance the sizes of $\|\boldsymbol{x}_\mu\|$ and $\|\boldsymbol{r}_\mu\|$; see [17] for details.

In order to be able to compute μ_L by publicly available and tested software, we first determine the singular value decomposition of A. The MATLAB code *l_corner* from [16] applies this decomposition to determine the location of the vertex. The value of the regularization parameter determined in this manner is $\mu_L = 1\cdot 10^{-11}$. The dashed graph of Figure 5.5 shows the associated solution $\boldsymbol{x}_{\mu_L}$. It is clear from this graph that μ_L is too small; we have $\|\boldsymbol{x}_{\mu_L} - \hat{\boldsymbol{x}}\| = 1.0$. This example illustrates that for some problems our method yields a more appropriate value of the regularization parameter than the L-curve criterion.

6. Conclusion. A new method for determining a suitable value of the regularization parameter in Tikhonov regularization when no information about the norm of the noise is available is described. Computed examples show the method to give meaningful approximate solutions for a few discrete ill-posed problems. The method also has performed well in numerous other computed examples.

Acknowledgment. Work by LR was supported by the University of Cagliari. LR would like to thank G. Rodriguez and S. Seatzu for an enjoyable visit to Cagliari, during which work on the present paper was begun.

REFERENCES

[1] G. Auchmuty, *A posteriori error estimates for linear equations*, Numer. Math., 61 (1992), pp. 1–6.

[2] 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.

[3] A. B. Bakushinskii, *Remarks on choosing a regularization parameter using quasi-optimality and ratio criterion*, USSR Comp. Math. Math. Phys., 24(4) (1984), pp. 181–182.

[4] C. Brezinski, *Error estimates for the solution of linear systems*, SIAM J. Sci. Comput., 21 (1999), pp. 764–781.

[5] C. Brezinski, G. Rodriguez, and S. Seatzu, *Error estimates for linear systems with applications to regularization*, Numer. Algorithms, in press.

[6] C. Brezinski, G. Rodriguez, and S. Seatzu, *Error estimates for the regularization of least squares problems*, submitted for publication.

[7] D. Calvetti, G. H. Golub, and L. Reichel, *Estimation of the L-curve via Lanczos bidiagonalization*, BIT, 39 (1999), pp. 603–619.

[8] D. Calvetti and L. Reichel, *Tikhonov regularization of large linear problems*, BIT, 43 (2003), pp. 263–283.
[9] L. Eldén, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT, 17 (1977), pp. 134–145.
[10] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
[11] W. Gautschi, *The interplay between classical analysis and (numerical) linear algebra – a tribute to Gene H. Golub*, Electron. Trans. Numer. Anal., 13 (2002), pp. 119–147.
[12] W. Gautschi, *Orthogonal Polynomials: Computation and Approximation*, Oxford University Press, Oxford, 2004.
[13] G. H. Golub, *Bounds for matrix moments*, Rocky Mountain J. Math., 4 (1974), pp. 207–211.
[14] G. H. Golub and J. Kautsky, *Calculation of Gauss quadratures with multiple free and fixed nodes*, Numer. Math., 41 (1983), pp. 147–163.
[15] G. H. Golub and G. Meurant, *Matrices, moments and quadrature*, in Numerical Analysis 1993, eds. D. F. Griffiths and G. A. Watson, Longman, Essex, England, 1994, pp. 105–156.
[16] P. C. Hansen, *Regularization Tools version 4.0 for MATLAB 7.3*, Numer. Algorithms, 46 (2007), pp. 189–194. Software is available in Netlib at `http://www.netlib.org`.
[17] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.
[18] G. López Lagomasino, L. Reichel, and L. Wunderlich, *Matrices, moments, and rational quadrature*, Linear Algebra Appl., 429 (2008), pp. 2540–2554.
[19] C. V. M. van der Mee and S. Seatzu, *A method for generating infinite positive self-adjoint test matrices and Riesz bases*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 1132–1149.
[20] S. Morigi, L. Reichel, and F. Sgallari, *Orthogonal projection regularization operators*, Numer. Algorithms, 44 (2007), pp. 99–114.
[21] 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.
[22] D. L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, J. ACM, 9 (1962), pp. 84–97.
[23] L. Reichel and Q. Ye, *Simple square smoothing regularization operators*, submitted for publication.
[24] G. Rodriguez, *Fast solution of Toeplitz- and Cauchy-like least-squares problems*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 724–748.
[25] S. Seatzu, *A remark on the solution of linear inverse problems with discrete data*, Inverse Problems, 2 (1986), pp. L27–L30.
[26] 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.