Electronic Transactions on Numerical Analysis.
Volume 14, pp. 20-35, 2002.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

L-CURVE CURVATURE BOUNDS VIA LANCZOS BIDIAGONALIZATION *

D. CALVETTI†, P. C. HANSEN‡, AND L. REICHEL§

Abstract. The L-curve is often applied to determine a suitable value of the regularization parameter when solving ill-conditioned linear systems of equations with a right-hand side contaminated by errors of unknown norm. The location of the vertex of the L-curve typically yields a suitable value of the regularization parameter. However, the computation of the L-curve and of its curvature is quite costly for large problems; the determination of a point on the L-curve requires that both the norm of the regularized approximate solution and the norm of the corresponding residual vector be available. Recently, the L-ribbon, which contains the L-curve in its interior, has been shown to be suitable for the determination of the regularization parameter for large-scale problems. In this paper we describe how techniques similar to those employed for the computation of the L-ribbon can be used to compute a "curvature-ribbon," which contains the graph of the curvature of the L-curve. Both curvature- and L-ribbon can be computed fairly inexpensively by partial Lanczos bidiagonalization of the matrix of the given linear system of equations. A suitable value of the regularization parameter is then determined from these ribbons, and we show that an associated approximate solution of the linear system can be computed with little additional work.

Key words. Ill-posed problem, regularization, L-curve, Gauss quadrature.

AMS subject classifications. Ill-posed problem, regularization, L-curve, Gauss quadrature.

1. Introduction.

Consider the linear system of equations

$$Ax = \mathbf{b}, \quad A \in \mathbb{R}^{m\times n}, \quad \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{b} \in \mathbb{R}^m, \tag{1.1}$$

with a matrix of ill-determined rank, i.e., A is assumed to have many singular values of different orders of magnitude close to the origin; some singular values may vanish. Matrices of this kind arise when discretizing linear ill-posed problems; they are severely ill-conditioned. We refer to (1.1) as a linear discrete ill-posed problem. The linear system (1.1) is not required to be consistent.

Linear discrete ill-posed problems arise when seeking to determine the internal structure of a system from external measurements. The measurements are represented by the right-hand side $\mathbf{b}$. Typically, the right-hand side of linear discrete ill-posed problems that arise in the sciences and engineering is contaminated by an error $\mathbf{e} \in \mathbb{R}^m$, i.e.,

$$\mathbf{b} = \hat{\mathbf{b}} + \mathbf{e},$$

where $\hat{\mathbf{b}} \in \mathbb{R}^m$ denotes the unknown error-free right-hand side. In the present paper, we assume that neither the error $\mathbf{e}$ nor its norm are known. The linear system of equations with the unknown error-free right-hand side

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

is assumed to be consistent. We denote its solution of minimal Euclidean norm by $\hat{\mathbf{x}}$. We would like to determine an approximation of $\hat{\mathbf{x}}$ by computing an approximate solution to the available linear system (1.1). Due to the error $\mathbf{e}$ in $\mathbf{b}$ and the severe ill-conditioning of A,

*Received August 8, 2001. Accepted for publication May 20, 2002. Communicated by Sven Ehrich.

†Department of Mathematics, Case Western Reserve University, Cleveland, OH 44106. E-mail: dxc57@po.cwru.edu. Research supported in part by NSF grant DMS-9806702.

‡Informatics and Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark. E-mail: pch@imm.dtu.dk.

§Department of Mathematical Sciences, Kent State University, Kent, OH 44242. E-mail: reichel@math.kent.edu. Research supported in part by NSF grant DMS-9806413.

ETNA
Kent State University
etna@mcs.kent.edu

the least-squares solution of minimal Euclidean norm of (1.1) typically is not a meaningful approximation of $\hat{\mathbf{x}}$. One therefore often replaces the discrete ill-posed problem (1.1) by a related problem, whose solution is less sensitive to errors in the right-hand side $\mathbf{b}$, and then solves the new system obtained for an approximation of $\hat{\mathbf{x}}$. This replacement is referred to as regularization.

One of the most popular regularization methods is Tikhonov regularization, which in its simplest form replaces the linear system (1.1) by the minimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n}\{\|A\mathbf{x}-\mathbf{b}\|^2+\mu^2\|\mathbf{x}\|^2\}. \tag{1.3}$$

Here μ is the regularization parameter. Throughout this paper $\|\cdot\|$ denotes the Euclidean norm. The solution $\mathbf{x}_\mu$ of (1.3) solves the linear system

$$(A^TA+\mu^2I)\mathbf{x}=A^T\mathbf{b}, \tag{1.4}$$

i.e.,

$$\mathbf{x}_\mu=(A^TA+\mu^2I)^{-1}A^T\mathbf{b}, \tag{1.5}$$

where we define $\mathbf{x}_0 := \lim_{\mu\to 0}\mathbf{x}_\mu = A^\dagger\mathbf{b}$, with $A^\dagger$ denoting the Moore-Penrose pseudo-inverse of A.

A proper choice of the value of the regularization parameter μ is essential for the success of this regularization method. Specifically, we would like to determine $\mu \geq 0$, so that $\mathbf{x}_\mu$ is an accurate approximation of the minimal-norm solution $\hat{\mathbf{x}}$ of (1.2). A popular method for choosing a suitable value of the regularization parameter, when the norm of the error $\mathbf{e}$ is not explicitly known, is based on the curve

$$\mathcal{L} := \{(\log\|\mathbf{x}_\mu\|, \log\|A\mathbf{x}_\mu-\mathbf{b}\|) : \mu\geq 0\}. \tag{1.6}$$

This curve is usually referred to as the L-curve, because for a large class of problems it is shaped like the letter "L." Hansen and O'Leary [10, 13] propose to choose the value of the parameter μ that corresponds to the point at the "vertex" of the "L," where the vertex is defined to be the point on the L-curve with curvature κ_μ of largest magnitude. We denote this value of the regularization parameter by μ_L. A heuristic motivation for choosing the value μ_L is that when $\mu \geq 0$ is "tiny," the associated solution $\mathbf{x}_\mu$ of (1.3) is likely to be severely contaminated by propagated errors due to the error $\mathbf{e}$ in the right-hand side $\mathbf{b}$ of (1.1). On the other hand, when μ is large, the associated solution $\mathbf{x}_\mu$ of (1.3) is a poor approximation of $\hat{\mathbf{x}}$. The choice $\mu = \mu_L$ seeks to balance these sources of errors. Hansen [12] provides an insightful discussion on the properties of this choice of regularization parameter, and addresses both advantages and shortcomings.

A computational difficulty that arises, when seeking to determine the value μ_L of the regularization parameter for large linear discrete ill-posed problems, is that it requires the computation of the curvature at many points on the L-curve. These computations can be quite expensive. It is the purpose of the present paper to describe how upper and lower bounds for the curvature κ_μ can be computed quite inexpensively for several values of μ by incomplete Lanczos bidiagonalization of the matrix A and by the application of Gauss-type quadrature rules. The method proposed allows us to determine a ribbon that contains the curve $\{(\mu, \kappa_\mu) : \mu \geq 0\}$ in its interior. We refer to this ribbon as the curvature-ribbon. Its width depends on the number of Lanczos bidiagonalization steps carried out; typically the width decreases monotonically as the number of Lanczos steps increases.

The techniques used to determine the curvature-ribbon are similar to those applied by Calvetti et al. [2] for inexpensively computing rectangular regions that contain points on the L-curve. The union of these rectangular regions for all $\mu \geq 0$ is in [2] referred to as the L-ribbon; see also Section 4. It is attractive to compute both the curvature-ribbon and the L-ribbon in order to determine an approximation of the value μ_L of the regularization parameter.

This paper is organized as follows. Section 2 expresses the coordinates of points on the L-curve and the curvature of the L-curve at these points in terms of certain matrix functionals that depend on the regularization parameter μ. We can determine upper and lower bounds for the abscissa and ordinate for each point on the L-curve, as well as for the curvature, by computing upper and lower bounds for these matrix functionals. Section 3 describes how Gauss and Gauss-Radau quadrature rules can be employed to compute such bounds inexpensively. Section 4 describes algorithms for constructing the L-ribbon and the curvature-ribbon. A few illustrative numerical examples are presented in Section 5, and Section 6 contains concluding remarks.

The method described in the present paper is well suited for linear discrete ill-posed problems (1.1) with a matrix A of order $m \times n$ when m is larger than or of the same order of magnitude as n. The method also can be applied when m is considerably smaller than n; however, there are alternative approaches that require less computer storage when $m \ll n$. The computation of L-ribbons for discrete ill-posed problems (1.1) with $m \ll n$ is discussed in [3, 4]. The latter methods can be extended to allow the efficient computation of curvature-ribbons for discrete ill-posed problems with $m \ll n$. We also note that problems and techniques related to those of the present paper are discussed by Golub and von Matt [8, 9].

2. The L-curve, its Curvature, and Matrix Functionals. Let $\mu > 0$ be a given value of the regularization parameter for Tikhonov regularization and let $P_\mu = (\log \|\mathbf{x}_\mu\|, \log \|A\mathbf{x}_\mu - \mathbf{b}\|)$ be the corresponding point on the L-curve. It follows from (1.5) that

$$\|\mathbf{x}_\mu\|^2 = \mathbf{x}_\mu^T \mathbf{x}_\mu = \mathbf{b}^T A (A^T A + \mu^2 I)^{-2} A^T \mathbf{b}, \tag{2.1}$$

and

$$\|A\mathbf{x}_\mu - \mathbf{b}\|^2 = \mu^4 \mathbf{b}^T (AA^T + \mu^2 I)^{-2} \mathbf{b}.$$

Introduce the function

$$\phi_\mu(t) := (t + \mu^2)^{-2}, \tag{2.2}$$

and define the matrix functionals

$$\eta_\mu := (A^T\mathbf{b})^T \phi_\mu(A^T A) A^T \mathbf{b}, \qquad \rho_\mu := \mu^4 \mathbf{b}^T \phi_\mu(AA^T)\mathbf{b}. \tag{2.3}$$

Then

$$\|\mathbf{x}_\mu\| = \eta_\mu^{1/2}, \qquad \|A\mathbf{x}_\mu - \mathbf{b}\| = \rho_\mu^{1/2}. \tag{2.4}$$

and therefore $P_\mu = \frac{1}{2}(\log \eta_\mu, \log \rho_\mu)$. Let

$$\hat{\eta}_\mu := \log \eta_\mu, \qquad \hat{\rho}_\mu := \log \rho_\mu. \tag{2.5}$$

Then the curvature of the L-curve at the point P_μ is given by

$$\kappa_\mu = 2 \frac{\hat{\rho}_\mu'' \hat{\eta}_\mu' - \hat{\rho}_\mu' \hat{\eta}_\mu''}{((\hat{\rho}_\mu')^2 + (\hat{\eta}_\mu')^2)^{3/2}}, \tag{2.6}$$

ETNA
Kent State University
etna@mcs.kent.edu

where $'$ denotes differentiation with respect to μ. It follows from (2.5) that

$$\hat{\eta}''_\mu = \frac{\eta''_\mu \eta_\mu - (\eta'_\mu)^2}{\eta^2_\mu}, \qquad \hat{\rho}''_\mu = \frac{\rho''_\mu \rho_\mu - (\rho'_\mu)^2}{\rho^2_\mu}. \tag{2.7}$$

Furthermore, since $\rho'_\mu = -\mu^2 \eta'_\mu$, we have

$$\rho''_\mu = -2\mu\eta'_\mu - \mu^2\eta''_\mu, \tag{2.8}$$

Substituting (2.7) and (2.8) into (2.6) yields

$$\kappa_\mu = 2\frac{\eta_\mu \rho_\mu}{\eta'_\mu} \frac{\mu^2\eta'_\mu\rho_\mu + 2\mu\eta_\mu\rho_\mu + \mu^4\eta_\mu\eta'_\mu}{(\mu^4\eta^2_\mu + \rho^2_\mu)^{3/2}}. \tag{2.9}$$

Notice that only the first derivative of η_μ is required.

This paper describes how equation (2.9) can be used to inexpensively compute upper and lower bounds of κ_μ for many values of μ. Introduce the function

$$\psi_\mu(t) := \frac{d\phi_\mu}{d\mu} = -4\mu(t+\mu^2)^{-3}. \tag{2.10}$$

Then we can express η'_μ as

$$\eta'_\mu = (A^T\mathbf{b})^T\psi_\mu(A^TA)(A^T\mathbf{b}). \tag{2.11}$$

The key to obtaining computable bounds for ρ_μ, η_μ and η'_μ is to express these quantities as Stieltjes integrals. Assume for ease of notation that $m \geq n$ and introduce the Singular Value Decomposition (SVD) of the matrix A,

$$A = \widetilde{U}\begin{pmatrix}\widetilde{\Sigma}\\ 0\end{pmatrix}\widetilde{V}^T, \qquad \widetilde{U} \in \mathbb{R}^{m\times m}, \qquad \widetilde{\Sigma}, \widetilde{V} \in \mathbb{R}^{n\times n}, \tag{2.12}$$

with $\widetilde{U}$ and $\widetilde{V}$ orthogonal, and $\widetilde{\Sigma} = \mathrm{diag}(\widetilde{\sigma}_1, \widetilde{\sigma}_2, \ldots, \widetilde{\sigma}_n)$. Then

$$AA^T = \widetilde{U}\Lambda\widetilde{U}^T, \qquad A^TA = \widetilde{V}\hat{\Lambda}\widetilde{V}^T,$$

where

$$\Lambda := \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n, 0, \ldots, 0] \in \mathbb{R}^{m\times m}, \qquad \hat{\Lambda} := \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n] \in \mathbb{R}^{n\times n},$$

and $\lambda_i = \widetilde{\sigma}^2_i$ for $1 \leq i \leq n$. Now let

$$\mathbf{h} = [h_1, h_2, \ldots, h_m]^T := \mu^2\widetilde{U}^T\mathbf{b}, \qquad \hat{\mathbf{h}} = [\hat{h}_1, \hat{h}_2, \ldots, \hat{h}_n]^T := \widetilde{V}^TA^T\mathbf{b}$$

and substitute the SVD of A (2.12) into (2.3) and (2.11). We obtain

$$\rho_\mu = \mathbf{h}^T\phi_\mu(\Lambda)\mathbf{h} = \sum_{k=1}^{n}\phi_\mu(\lambda_k)h^2_k + \phi_\mu(0)\sum_{k=n+1}^{m}h^2_k = \int_{-\infty}^{\infty}\phi_\mu(t)d\omega(t), \tag{2.13}$$

$$\eta_\mu = \hat{\mathbf{h}}^T\phi_\mu(\hat{\Lambda})\hat{\mathbf{h}} = \sum_{k=1}^{n}\phi_\mu(\lambda_k)\hat{h}^2_k = \int_{-\infty}^{\infty}\phi_\mu(t)d\hat{\omega}(t), \tag{2.14}$$

$$\eta'_\mu = \hat{\mathbf{h}}^T\psi_\mu(\hat{\Lambda})\hat{\mathbf{h}} = \sum_{k=1}^{n}\psi_\mu(\lambda_k)\hat{h}^2_k = \int_{-\infty}^{\infty}\psi_\mu(t)d\hat{\omega}(t). \tag{2.15}$$

The distribution functions ω and $\hat{\omega}$, defined by (2.13) and (2.14), respectively, are nondecreasing step functions with jump discontinuities at the eigenvalues λ_k. Moreover, the function ω generally has a jump discontinuity at the origin when $m > n$. We show in the next section how Gauss and Gauss-Radau quadrature rules can be applied to cheaply compute upper and lower bounds of the quantities ρ_μ, η_μ, and η'_μ by using their representations (2.13), (2.14) and (2.15), respectively, in terms of Stieltjes integrals.

3. Gauss Quadrature and Lanczos Bidiagonalization.

The computation of the quantities ρ_μ, η_μ and η'_μ introduced in the previous section for several values of μ is very expensive when the matrix A is very large. This section discusses how upper and lower bounds of these quantities can be evaluated by using Gauss and Gauss-Radau quadrature rules. These bounds can be substituted into (2.9) to determine upper and lower bounds of the curvature κ_μ. We remark that the application of Gauss-type quadrature rules to compute bounds for certain matrix functionals is well-known; see the paper by Golub and Meurant [7] and references therein. A discussion on how to use these techniques to compute an L-ribbon that contains the L-curve (1.6) is presented in [2].

We review some facts about Gauss quadrature rules and their connection with the Lanczos process. Define the inner product induced by the distribution function $\hat{\omega}$ introduced in (2.14),

$$\langle f, g\rangle := \int_{-\infty}^{\infty} f(t)g(t)d\hat{\omega}(t) = \sum_{k=1}^{n} f(\lambda_k)g(\lambda_k)\hat{h}_k^2 = \hat{\mathbf{h}}^T f(\hat{\Lambda})g(\hat{\Lambda})\hat{\mathbf{h}}, \tag{3.1}$$

and let $\hat{q}_0, \hat{q}_1, \hat{q}_2, \ldots$ be the family of orthonormal polynomials with positive leading coefficients with respect to this inner product; thus,

$$\langle \hat{q}_k, \hat{q}_j\rangle = \begin{cases} 0, & k \neq j, \\ 1, & k = j. \end{cases} \tag{3.2}$$

The polynomials $\hat{q}_k$ satisfy a three-term recurrence relation of the form

$$t\hat{q}_{k-1}(t) = \hat{\beta}_k\hat{q}_k(t) + \hat{\alpha}_k\hat{q}_{k-1}(t) + \hat{\beta}_{k-1}\hat{q}_{k-2}(t), \qquad k = 1, 2, 3, \ldots, \tag{3.3}$$

where $\hat{q}_{-1}(t) := 0$, $\hat{q}_0(t) := \langle 1, 1\rangle^{-1/2}$ and $\hat{\beta}_0 := 0$. We remark that since the distribution function $\hat{\omega}$ has at most n distinct points of increase, there are at most n orthonormal polynomials $\hat{q}_j$.

It is well-known that ℓ steps of the Lanczos algorithm applied to the matrix $A^T A$ with initial vector $A^T\mathbf{b}$ yields the symmetric positive definite or semidefinite tridiagonal matrix

$$\hat{T}_\ell := \begin{bmatrix} \hat{\alpha}_1 & \hat{\beta}_1 & & & \\ \hat{\beta}_1 & \hat{\alpha}_2 & \hat{\beta}_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \hat{\beta}_{\ell-2} & \hat{\alpha}_{\ell-1} & \hat{\beta}_{\ell-1} \\ & & & \hat{\beta}_{\ell-1} & \hat{\alpha}_\ell \end{bmatrix}, \tag{3.4}$$

whose entries are the first $2\ell - 1$ coefficients in the recurrence relation (3.3); see, e.g., [7] as well as the end of this section. We assume here that ℓ is sufficiently small to secure that $\hat{\beta}_j > 0$ for $1 \leq j < \ell$.

The ℓ-point Gauss quadrature rule associated with the distribution function $\hat{\omega}$ can be expressed in terms of the tridiagonal matrix $\hat{T}_\ell$ as follows:

$$\hat{\mathcal{G}}_\ell(f) = \|A^T\mathbf{b}\|^2\mathbf{e}_1^T f(\hat{T}_\ell)\mathbf{e}_1. \tag{3.5}$$

Here and throughout this paper $\mathbf{e}_j$ denotes the jth axis vector.

Introduce the Cholesky factor of $\hat{T}_\ell$,

$$\hat{C}_\ell := \begin{bmatrix} \hat{\gamma}_1 & & & & \\ \hat{\delta}_2 & \hat{\gamma}_2 & & & \\ & \ddots & \ddots & & \\ & & \hat{\delta}_{\ell-1} & \hat{\gamma}_{\ell-1} & \\ & & & \hat{\delta}_\ell & \hat{\gamma}_\ell \end{bmatrix} \tag{3.6}$$

as well as the matrix

$$\bar{\hat{C}}_{\ell-1} = \begin{bmatrix} \hat{C}_{\ell-1} \\ \hat{\delta}_\ell \mathbf{e}_{\ell-1}^T \end{bmatrix} \in \mathbb{R}^{\ell \times (\ell-1)}, \tag{3.7}$$

which is made up of the $\ell - 1$ first columns of $\hat{C}_\ell$.

In the cases of particular interest to us, when $f := \phi_\mu$ or $f := \psi_\mu$, where ϕ_μ and ψ_μ are defined by (2.2) and (2.10), respectively, formula (3.5) can be evaluated easily by using the Cholesky factor $\hat{C}_\ell$ of the matrix $\hat{T}_\ell$; see Section 4. A discussion on how to compute the matrix $\hat{C}_\ell$ directly, without determining $\hat{T}_\ell$, by partial Lanczos bidiagonalization of A is provided at the end of this section.

We use the matrix $\bar{\hat{C}}_{\ell-1}$ in a representation analogous to (3.5) of the ℓ-point Gauss-Radau quadrature rule

$$\hat{\mathcal{R}}_\ell(f) = \sum_{k=1}^{\ell} f(\tilde{\theta}_k)\tilde{\eta}_k, \tag{3.8}$$

associated with the distribution function $\hat{\omega}(t)$ with one assigned node, say $\tilde{\theta}_1$, at the origin. One can show that this quadrature rule can be expressed as

$$\hat{\mathcal{R}}_\ell(f) = \|A^T\mathbf{b}\|^2 \mathbf{e}_1^T f(\bar{\hat{C}}_{\ell-1}\bar{\hat{C}}_{\ell-1}^T)\mathbf{e}_1; \tag{3.9}$$

see, e.g., [2] for a proof.

Let f be a 2ℓ times differentiable function. Introduce the quadrature error

$$\mathcal{E}_{\mathcal{Q}_\ell}(f) := \int_{-\infty}^{\infty} f(t)d\hat{\omega}(t) - \mathcal{Q}_\ell(f),$$

where $\mathcal{Q}_\ell = \hat{\mathcal{G}}_\ell$ or $\mathcal{Q}_\ell = \hat{\mathcal{R}}_\ell$. In the former case, there exists a constant $\theta_{\hat{\mathcal{G}}_\ell}$, with $\lambda_1 \leq \theta_{\hat{\mathcal{G}}_\ell} \leq \lambda_n$, such that

$$\mathcal{E}_{\hat{\mathcal{G}}_\ell}(f) = \frac{f^{(2\ell)}(\theta_{\hat{\mathcal{G}}_\ell})}{(2\ell)!} \int_{-\infty}^{\infty} \prod_{i=1}^{\ell} (t - \theta_i)^2 d\hat{\omega}(t), \tag{3.10}$$

where $f^{(j)}$ denotes the jth derivative of the function f. On the other hand, if $\mathcal{Q}_\ell = \hat{\mathcal{R}}_\ell$, then there is a constant $\theta_{\hat{\mathcal{R}}_\ell}$, with $\lambda_1 \leq \theta_{\hat{\mathcal{R}}_\ell} \leq \lambda_n$, such that

$$\mathcal{E}_{\hat{\mathcal{R}}_\ell}(f) = \frac{f^{(2\ell-1)}(\theta_{\hat{\mathcal{R}}_\ell})}{(2\ell-1)!} \int_{-\infty}^{\infty} t \prod_{i=2}^{\ell} (t - \tilde{\theta}_i)^2 d\hat{\omega}(t), \tag{3.11}$$

where we recall that $\tilde{\theta}_1 = 0$. We refer to Golub and Meurant [7] for details.

PROPOSITION 3.1. *Assume that $\mu > 0$ and let ϕ_μ and ψ_μ be defined by (2.2) and (2.10), respectively. Then, for $\ell \geq 1$,*

$$\mathcal{E}_{\hat{\mathcal{G}}_\ell}(\phi_\mu) > 0, \qquad \mathcal{E}_{\hat{\mathcal{R}}_\ell}(\phi_\mu) < 0 \tag{3.12}$$

and

$$\mathcal{E}_{\hat{\mathcal{G}}_\ell}(\psi_\mu) < 0, \qquad \mathcal{E}_{\hat{\mathcal{R}}_\ell}(\psi_\mu) > 0. \tag{3.13}$$

Proof. Analogous results have been shown, e.g., by Golub and Meurant [7]. The jth derivative of $\phi_\mu(t)$ with respect to t is given by $\phi_\mu^{(j)}(t) = (-1)^j(j+1)!(t+\mu)^{-j-2}$ and it follows that $\phi_\mu^{(2\ell-1)}(t) < 0$ and $\phi_\mu^{(2\ell)}(t) > 0$ for $t \geq 0$. Substituting these inequalities into (3.10) and (3.11), and using the fact that all the jump discontinuities of $\omega(t)$ are on the nonnegative real axis, gives (3.12). The inequalities (3.13) can be shown analogously. □

It follows from Proposition 3.1 that upper and lower bounds for the quantities η_μ and η'_μ can be determined by replacing the Stieltjes integrals in (2.14) and (2.15) by Gauss and Gauss-Radau quadrature rules.

We turn to bounds for the quantity ρ_μ defined by (2.3). Let $T_\ell \in \mathbb{R}^{\ell\times\ell}$ denote the tridiagonal matrix obtained by ℓ steps of the Lanczos process applied to AA^T with initial vector $\mathbf{b}$, and let C_ℓ denote the lower bidiagonal Cholesky factor of T_ℓ. Then analogous to the representation (3.5), the ℓ-point Gauss quadrature rule, with respect to the distribution function $\omega(t)$ defined by (2.13), can be written as

$$\mathcal{G}_\ell(\phi_\mu) = \mu^4\|\mathbf{b}\|^2\mathbf{e}_1^T\phi_\mu(T_\ell)\mathbf{e}_1. \tag{3.14}$$

Similarly to (3.9), the ℓ-points Gauss-Radau rule, associated with the distribution function $\omega(t)$ with one node assigned at the origin, can be evaluated as

$$\mathcal{R}_\ell(\phi_\mu) = \mu^4\|\mathbf{b}\|^2\mathbf{e}_1^T\phi_\mu(\bar{C}_{\ell-1}(\bar{C}_{\ell-1})^T)\mathbf{e}_1, \tag{3.15}$$

where the matrix $\bar{C}_{\ell-1}$ consists of the $\ell-1$ first columns of the matrix C_ℓ. We obtain analogously to (3.12) that

$$\mathcal{E}_{\mathcal{G}_\ell}(\phi_\mu) > 0, \qquad \mathcal{E}_{\mathcal{R}_\ell}(\phi_\mu) < 0. \tag{3.16}$$

Thus, lower and upper bounds for the quantity ρ_μ defined by (2.3) can be determined by replacing the Stieltjes integral in (2.13) by the Gauss and Gauss-Radau quadrature rules (3.14) and (3.15), respectively.

We conclude this section with some remarks on the computation of the Cholesky factors used for the evaluation of the quadrature rules. Application of ℓ steps of the Lanczos bidiagonalization algorithm bidiag1, described by Paige and Saunders [14] to the matrix A with initial vector $\mathbf{b}$, yields the decompositions

$$AV_\ell = U_\ell C_\ell + \delta_{\ell+1}\mathbf{u}_{\ell+1}\mathbf{e}_\ell^T, \qquad A^TU_\ell = V_\ell C_\ell^T, \qquad \mathbf{b} = \delta_1 U_\ell \mathbf{e}_1, \tag{3.17}$$

where the matrices $U_\ell \in \mathbb{R}^{m\times\ell}$ and $V_\ell \in \mathbb{R}^{n\times\ell}$ satisfy $U_\ell^TU_\ell = I_\ell$ and $V_\ell^TV_\ell = I_\ell$, and the unit vector $\mathbf{u}_{\ell+1} \in \mathbb{R}^m$ is such that $U_\ell^T\mathbf{u}_{\ell+1} = \mathbf{0}$. The matrix $C_\ell \in \mathbb{R}^{\ell\times\ell}$ is lower bidiagonal. Combining the equations (3.17) suitably shows that

$$AA^TU_\ell = U_\ell C_\ell C_\ell^T + \delta_{\ell+1}\gamma_\ell\mathbf{u}_{\ell+1}\mathbf{e}_\ell^T, \tag{3.18}$$

where $\gamma_\ell = \mathbf{e}_\ell^T C_\ell \mathbf{e}_\ell$.

Let $\mathbf{u}_j := U_\ell \mathbf{e}_j$, $1 \le j \le \ell$. It is straightforward to verify that $\{\mathbf{u}_j\}_{j=1}^{\ell+1}$ are Lanczos vectors and that $T_\ell := C_\ell C_\ell^T$ is the tridiagonal matrix obtained when applying ℓ steps of the Lanczos algorithm to the matrix AA^T with initial vector $\mathbf{b}$. In particular, the bidiagonal matrix C_ℓ in (3.17) is a Cholesky factor of T_ℓ.

Let $\mathbf{v}_j := V_\ell \mathbf{e}_j$, $1 \le j \le \ell$. Multiplying the equation on the left-hand side in (3.17) from the left by A^T shows that $\{\mathbf{v}_j\}_{j=1}^{\ell}$ are Lanczos vectors that can be generated by applying ℓ steps of the Lanczos algorithm to the matrix $A^T A$ with starting vector $A^T\mathbf{b}$. The corresponding tridiagonal matrix is given by $\hat{T}_\ell := \bar{C}_\ell^T \bar{C}_\ell$, where the lower bidiagonal $(\ell+1)\times\ell$ matrix $\bar{C}_\ell$ is defined analogously to (3.7).

We remark that the lower bidiagonal Cholesky factor $\hat{C}_\ell$ of $\hat{T}_\ell$ can be computed, without forming $\hat{T}_\ell$, from the QR-factorization $\bar{C}_\ell = \bar{Q}_\ell \hat{C}_\ell^T$, where $\bar{Q}_\ell \in \mathbb{R}^{(\ell+1)\times\ell}$ satisfies $\bar{Q}_\ell^T \bar{Q}_\ell = I_\ell$ and $\hat{C}_\ell^T$ is upper bidiagonal. Note that the matrix $\bar{Q}_\ell$ can be represented by ℓ Givens rotation and, therefore, $\hat{C}_\ell$ can be computed from $\bar{C}_\ell$ in only $\mathcal{O}(\ell)$ arithmetic operations.

4. The L- and Curvature-Ribbons. We show how to compute upper and lower bounds of the abscissa and ordinate of points on the L-curve and of the curvature of the L-curve at these points. These bounds yield the L-ribbon and the curvature-ribbon defined below. The evaluation of the bounds requires that the bidiagonal matrix C_ℓ and the coefficient $\delta_{\ell+1}$ in (3.17) be available; the width of the ribbons decreases as ℓ increases. Results from [2] show how to cheaply compute an approximate solution $\mathbf{x}_{\mu,\ell}$ of (1.3) and (1.4) associated with a desired value of μ

We first discuss the computation of bounds for ρ_μ defined by (2.3). It follows from (3.16) that

$$\mathcal{G}_\ell(\phi_\mu) \le \rho_\mu \le \mathcal{R}_{\ell+1}(\phi_\mu) \tag{4.1}$$

and (3.14) and (3.15) yield

$$\mathcal{G}_\ell(\phi_\mu) = \mu^4 \|\mathbf{b}\|^2 \mathbf{e}_1^T (C_\ell C_\ell^T + \mu^2 I_\ell)^{-2} \mathbf{e}_1, \tag{4.2}$$

$$\mathcal{R}_{\ell+1}(\phi_\mu) = \mu^4 \|\mathbf{b}\|^2 \mathbf{e}_1^T (\bar{C}_\ell \bar{C}_\ell^T + \mu^2 I_{\ell+1})^{-2} \mathbf{e}_1. \tag{4.3}$$

We evaluate the Gauss quadrature rule (4.2) by first determining the vector $\mathbf{y}_\mu := (C_\ell C_\ell^T + \mu^2 I_\ell)^{-1}\mathbf{e}_1$, which is computed as the solution of the least-squares problem

$$\min_{\mathbf{y}\in\mathbb{R}^\ell} \left\| \begin{bmatrix} C_\ell^T \\ \mu I_\ell \end{bmatrix} \mathbf{y} - \mu^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{e}_1 \end{bmatrix} \right\|, \qquad \mathbf{0} \in \mathbb{R}^\ell.$$

Eldén [5] describes how this problem can be solved efficiently with the aid of Givens rotations. Note that the matrix C_ℓ is independent of the regularization parameter μ. Therefore, given this matrix, the Gauss rule (4.2) can be evaluated in only $\mathcal{O}(\ell)$ arithmetic floating point operations for each value of μ. The evaluation of the Gauss-Radau rule (4.3) can be carried out similarly.

We determine bounds for $\|\mathbf{x}_\mu\|$ by computing bounds for the quantity η_μ, defined in (2.3). It follows from (3.12) that, analogously to (4.1),

$$\hat{\mathcal{G}}_\ell(\phi_\mu) \le \eta_\mu \le \hat{\mathcal{R}}_\ell(\phi_\mu), \tag{4.4}$$

and similarly to the representations (4.2) and (4.3), we have

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

We evaluate these quadrature rules similarly as (4.2).

Combining (2.4) and the inequalities (4.1) and (4.4) yields the following bounds

$$\begin{aligned}(\mathcal{G}_\ell(\phi_\mu))^{1/2} \leq & \|\mathbf{b} - A\mathbf{x}_\mu\| \leq (\mathcal{R}_{\ell+1}(\phi_\mu))^{1/2},\\ (\hat{\mathcal{G}}_\ell(\phi_\mu))^{1/2} \leq & \quad \|\mathbf{x}_\mu\| \quad \leq (\hat{\mathcal{R}}_\ell(\phi_\mu))^{1/2}.\end{aligned}$$

Introduce the quantities

$$\begin{aligned}\rho_\mu^- &:= \mathcal{G}_\ell(\phi_\mu), & \rho_\mu^+ &:= \mathcal{R}_{\ell+1}(\phi_\mu),\\ \eta_\mu^- &:= \hat{\mathcal{G}}_\ell(\phi_\mu), & \eta_\mu^+ &:= \hat{\mathcal{R}}_\ell(\phi_\mu),\end{aligned} \tag{4.5}$$

as well as

$$\hat{\rho}_\mu^- := \log \rho_\mu^-, \quad \hat{\rho}_\mu^+ := \log \rho_\mu^+, \quad \hat{\eta}_\mu^- := \log \eta_\mu^-, \quad \hat{\eta}_\mu^+ := \log \eta_\mu^+, \tag{4.6}$$

for $\mu > 0$. We define the L-ribbon as the union of rectangular regions with vertices determined by $\hat{\rho}_\mu^\pm$ and $\hat{\eta}_\mu^\pm$,

$$\bigcup_{\mu>0} \{(\hat{\eta}, \hat{\rho}) : \hat{\eta}_\mu^- \leq \hat{\eta} \leq \hat{\eta}_\mu^+, \hat{\rho}_\mu^- \leq \hat{\rho} \leq \hat{\rho}_\mu^+\}. \tag{4.7}$$

We turn to the derivation of bounds for the curvature κ_μ of the L-curve. Our bounds are based on formula (2.9) and require that bounds for η_μ' be evaluated. Equation (2.15) and the inequalities (3.13) yield

$$\hat{\mathcal{R}}_\ell(\psi_\mu) \leq \eta_\mu' \leq \hat{\mathcal{G}}_\ell(\psi_\mu),$$

where

$$\begin{aligned}\hat{\mathcal{G}}_\ell(\psi_\mu) &= -4\mu\|A^T\mathbf{b}\|^2 \mathbf{e}_1^T(\hat{C}_\ell \hat{C}_\ell^T + \mu^2 I_\ell)^{-3}\mathbf{e}_1,\\ \hat{\mathcal{R}}_\ell(\psi_\mu) &= -4\mu\|A^T\mathbf{b}\|^2 \mathbf{e}_1^T(\bar{\hat{C}}_{\ell-1}(\bar{\hat{C}}_{\ell-1})^T + \mu^2 I_\ell)^{-3}\mathbf{e}_1.\end{aligned} \tag{4.8}$$

The quadrature rules (4.8) can be evaluated analogously as (4.2).

For future reference, we introduce

$$(\eta_\mu')^- := \hat{\mathcal{R}}_\ell(\psi_\mu), \qquad (\eta_\mu')^+ := \hat{\mathcal{G}}_\ell(\psi_\mu). \tag{4.9}$$

The bounds (4.5) and (4.9) of ρ_μ, η_μ and η_μ' allow us to derive easily computable lower and upper bounds for the curvature κ_μ at the point $P_\mu = \frac{1}{2}(\hat{\eta}_\mu, \hat{\rho}_\mu)$. Define the auxiliary quantities

$$\begin{aligned}\tau_\mu &:= 2\frac{\eta_\mu \rho_\mu}{(\mu^4 \eta_\mu^2 + \rho_\mu^2)^{3/2}},\\ \xi_\mu &:= \mu^2 \rho_\mu + \mu^4 \eta_\mu + 2\mu\rho_\mu \frac{\eta_\mu}{\eta_\mu'}.\end{aligned}$$

Then, by (2.9),

$$\kappa_\mu = \tau_\mu \xi_\mu. \tag{4.10}$$

PROPOSITION 4.1. *Assume that* $\rho_\mu^- \geq 0$, $\eta_\mu^- \geq 0$, $(\eta_\mu')^+ < 0$ *and* $\mu^4(\eta_\mu^-)^2 + (\rho_\mu^-)^2 > 0$. *Let*

$$\tau_\mu^- := 2\frac{\eta_\mu^- \rho_\mu^-}{(\mu^4(\eta_\mu^+)^2 + (\rho_\mu^+)^2)^{3/2}}, \qquad \tau_\mu^+ := 2\frac{\eta_\mu^+ \rho_\mu^+}{(\mu^4(\eta_\mu^-)^2 + (\rho_\mu^-)^2)^{3/2}}, \tag{4.11}$$

ETNA
Kent State University
etna@mcs.kent.edu

and

$$\xi_\mu^- := \mu^2\rho_\mu^- + \mu^4\eta_\mu^- + 2\mu\rho_\mu^+ \frac{\eta_\mu^+}{(\eta'_\mu)^+}, \quad \xi_\mu^+ := \mu^2\rho_\mu^+ + \mu^4\eta_\mu^+ + 2\mu\rho_\mu^- \frac{\eta_\mu^-}{(\eta'_\mu)^-}. \tag{4.12}$$

Then

$$0 \le \tau_\mu^- \le \tau_\mu \le \tau_\mu^+, \qquad \xi_\mu^- \le \xi_\mu \le \xi_\mu^+. \tag{4.13}$$

Proof. The inequalities (4.13) follow from the bounds (4.5) of ρ_μ and η_μ, and the bounds (4.9) of η'_μ. Note that the first two terms in the expressions for $\xi_\mu^\pm$ are nonnegative, while the last term is nonpositive. □

PROPOSITION 4.2. *Assume that $\rho_\mu^- \ge 0$, $\eta_\mu^- \ge 0$, $(\eta'_\mu)^+ < 0$ and $\mu^4(\eta_\mu^-)^2 + (\rho_\mu^-)^2 > 0$. Let $\tau_\mu^\pm$ and $\xi_\mu^\pm$ be defined by (4.11) and (4.12), respectively. Then*

$$\kappa_\mu^- \le \kappa_\mu \le \kappa_\mu^+, \tag{4.14}$$

where

$$\kappa_\mu^- := \begin{cases} \tau_\mu^- \xi_\mu^-, & \text{if} \quad \xi_\mu^- \ge 0, \\ \tau_\mu^+ \xi_\mu^-, & \text{if} \quad \xi_\mu^- < 0, \end{cases} \tag{4.15}$$

$$\kappa_\mu^+ := \begin{cases} \tau_\mu^+ \xi_\mu^+, & \text{if} \quad \xi_\mu^+ \ge 0, \\ \tau_\mu^- \xi_\mu^+, & \text{if} \quad \xi_\mu^+ < 0. \end{cases} \tag{4.16}$$

Proof. The inequalities (4.14) follow from (4.10) and Proposition 4.1). □

A curvature-ribbon, which contains the graph $\{(\mu, \kappa_\mu) : \mu > 0\}$ can be defined as the union of intervals with endpoints (4.15) and (4.16),

$$\bigcup_{\mu>0} \{(\mu, \kappa) : \kappa_\mu^- \le \kappa \le \kappa_\mu^+\}.$$

The curvature κ_μ as defined by (2.6) is negative at the "vertex" of the L-curve. Generally, we find it more pleasing to plot the curvature-ribbon associated with the negative curvature $-\kappa_\mu$. This ribbon is given by

$$\bigcup_{\mu>0} \{(\mu, -\kappa) : -\kappa_\mu^+ \le -\kappa \le -\kappa_\mu^-\} \tag{4.17}$$

and will also be referred to as a curvature-ribbon. We display the latter curvature-ribbon in the computed examples of Section 5.

The following algorithm determines rectangles associated with the L-ribbon and intervals associated with the curvature-ribbon for the parameter values μ_j, $1 \le j \le p$.

ALGORITHM 1 (L- and Curvature-Ribbon Algorithm).
Input: $\mathbf{b} \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$, ℓ, $\{\mu_j\}_{j=1}^p$;
Output: $\{\hat\rho_{\mu_j}^+\}_{j=1}^p$, $\{\hat\rho_{\mu_j}^-\}_{j=1}^p$, $\{\hat\eta_{\mu_j}^+\}_{j=1}^p$, $\{\hat\eta_{\mu_j}^-\}_{j=1}^p$, $\{\kappa_{\mu_j}^+\}_{j=1}^p$, $\{\kappa_{\mu_j}^-\}_{j=1}^p$;

i) *Compute the entries of the bidiagonal matrix $\bar C_\ell$. Determine $\hat C_\ell$ by QR-factorization of $\bar C_\ell$.*

ETNA
Kent State University
etna@mcs.kent.edu

ii) **for** $j = 1, 2, \ldots, p$ **do**
Evaluate the quadrature rules $\mathcal{G}_\ell(\phi_{\mu_j})$, $\mathcal{R}_{\ell+1}(\phi_{\mu_j})$, $\hat{\mathcal{G}}_\ell(\phi_{\mu_j})$, $\hat{\mathcal{R}}_\ell(\phi_{\mu_j})$, $\hat{\mathcal{G}}_\ell(\psi_{\mu_j})$ *and* $\hat{\mathcal{R}}_\ell(\psi_{\mu_j})$ *to determine the bounds (4.5)-(4.6) and (4.15)-(4.16);*
end j

$\square$

We conclude this section by noting that an approximate solution of (1.4), and thereby of (1.1), that corresponds to a point in the L-ribbon (4.7), can easily be evaluated using the partial Lanczos bidiagonalization (3.17). Let $\mu_* > 0$ be the chosen value of the regularization parameter and consider the Galerkin equation

$$V_\ell^T(A^TA + \mu_*^2 I_n)V_\ell \mathbf{y} = V_\ell^T A^T \mathbf{b}. \tag{4.18}$$

The solution $\mathbf{y}_{\mu_*,\ell}$ of (4.18) yields the approximate solution $\mathbf{x}_{\mu_*,\ell} = V_\ell \mathbf{y}_{\mu_*,\ell}$ of (1.4). It is shown in [2] that

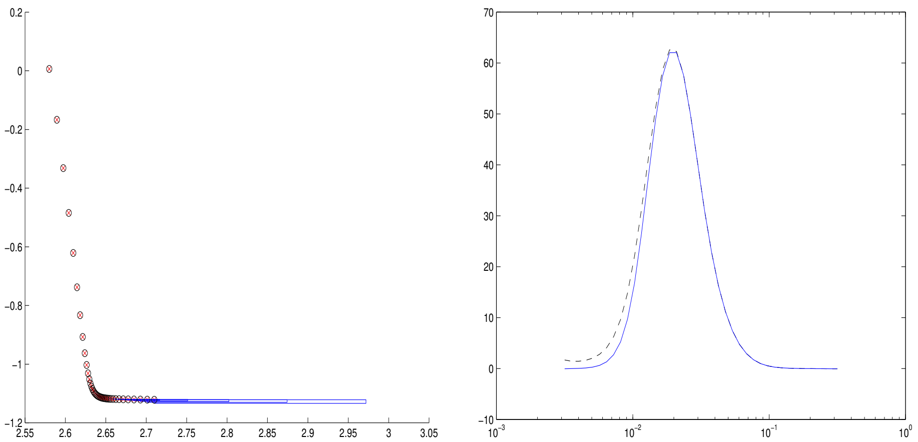
$$\log \|\mathbf{x}_{\mu_*,\ell}\| = \hat{\eta}_{\mu_*}^-, \qquad \log \|A\mathbf{x}_{\mu_*,\ell} - \mathbf{b}\| = \hat{\rho}_{\mu_*}^+.$$

We refer to [2] for details on the computation of $\mathbf{y}_{\mu_*,\ell}$.

5. Computed Examples. We illustrate the techniques from the previous sections with two numerical examples, both of which are Fredholm integral equations of the first kind,

$$\int_\alpha^\beta K(s,t)\,x(t)\,dt = b(s), \qquad \alpha \le s \le \beta. \tag{5.1}$$

All computations were performed with Matlab, version 5.3, on a personal computer with about 16 significant decimal digits.



(a) (b)

FIG. 5.1. *Example 5.1: 8 Lanczos bidiagonalization steps: (a) L-ribbon (4.7), (b) curvature-ribbon (4.17).*

Example 5.1. We consider the test problem shaw from the REGULARIZATION TOOLS package [11]. This problem is an integral equation of the form (5.1) with

$$\begin{aligned} K(s,t) &:= (\cos(s) + \cos(t))^2 \left(\frac{\sin(u(s,t))}{u(s,t)}\right)^2, \\ u(s,t) &:= \pi(\sin(s) + \sin(t)), \\ x(t) &:= 2\,\exp\left(-6(t-0.8)^2\right) + \exp\left(-2(t+0.5)^2\right) \end{aligned} \tag{5.2}$$

ETNA
Kent State University
etna@mcs.kent.edu

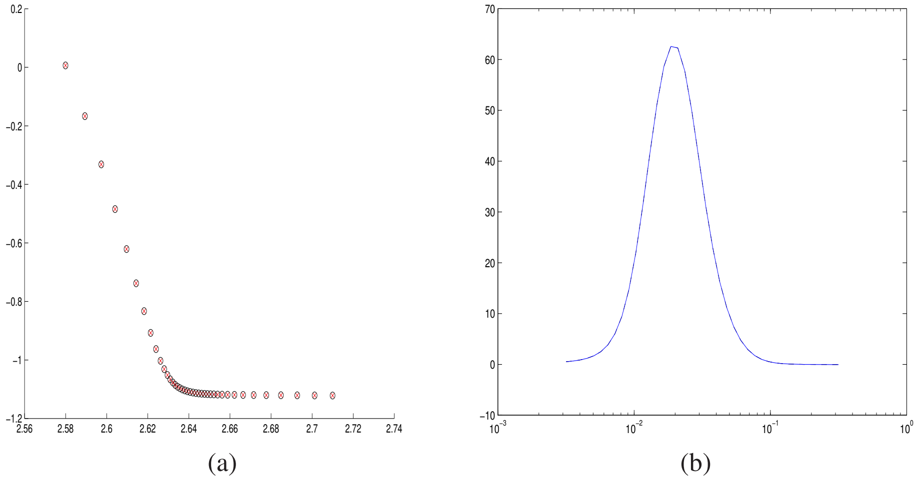


FIG. 5.2. *Example 5.1: 9 Lanczos bidiagonalization steps: (a) L-ribbon (4.7), (b) curvature-ribbon (4.17).*

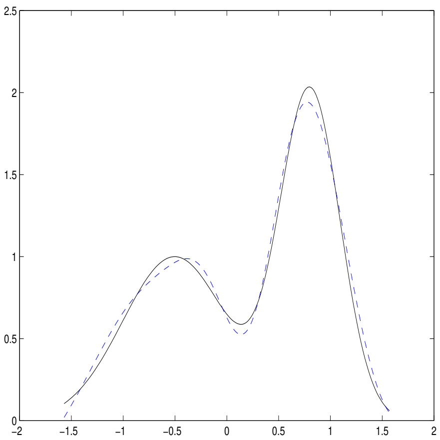


FIG. 5.3. *Example 5.1: 9 Lanczos bidiagonalization steps: Computed solution (blue dashed curve) and exact solution to noise-free problem (black solid curve).*

and $\alpha := -\pi/2$, $\beta := \pi/2$. The right-hand side function b in (5.1) is determined by the kernel K and the solution x.

The matrix A of the linear system (1.1) is obtained by discretizing the integral in (5.1) by a Nyström method. Specifically, the integral is replaced by the composite midpoint rule with nodes

$$t_i := \frac{2i - 201}{400}\pi, \qquad 1 \le i \le 200,$$

and the discretized integral so obtained is required to equal the right-hand side function b of (5.1) at the points $s_i := t_i$, $1 \le i \le 200$. This determines the matrix $A \in \mathbb{R}^{200\times 200}$, which is symmetric, indefinite and numerically singular.

Let the vector $\hat{\mathbf{x}} = [x(t_1), x(t_2), \ldots, x(t_{200})]^T$ be the tabulation of the solution (5.2) at the nodes t_i and compute the "noisy" right-hand side vector $\mathbf{b}$ in (1.1) according to $\mathbf{b} := A\hat{\mathbf{x}} + \mathbf{e}$, where $\mathbf{e}$ is an unbiased Gaussian noise vector with $\|\mathbf{e}\|/\|A\hat{\mathbf{x}}\| = 1 \cdot 10^{-2}$. We seek to determine an accurate approximation of the vector $\hat{\mathbf{x}}$ by computing a suitable approximate solution of the linear system (1.1).

While the linear system of equations (1.1), determined in this manner, is fairly small, it illustrates the behavior of the L- and curvature-ribbons as the number of Lanczos bidiagonalization steps ℓ is increased.

Figure 5.1 shows the L-ribbon (4.7) and curvature-ribbon (4.17) obtained with $\ell = 8$ Lanczos bidiagonalization steps. Figure 5.1(a) displays rectangles of the L-ribbon obtained for 40 logarithmically equispaced values μ_j of μ in the interval $\frac{1}{\sqrt{10}}[1 \cdot 10^{-2}, 1]$. The upper left corner $\{\hat{\eta}^-_{\mu_j}, \hat{\rho}^+_{\mu_j}\}$ of each rectangle, associated with the Galerkin solution of (4.18) for $\mu_* = \mu_j$, is marked by a red cross. When the rectangles are tiny, only this cross is visible of the rectangle. Points associated with the exact solution (1.5) for $\mu = \mu_j$ of the minimization problem (1.3) are marked by black circles.

The bounds are always tight for large values of μ, which correspond to imposing a large amount of regularization on the problem, while they deteriorate as μ gets smaller. The latter is due to the singularity of the integrands ϕ_μ and ψ_μ in (2.13)-(2.15) at $t = -\mu^2$; since the support of the measure is on the nonnegative real axis, as μ decreases the singularity moves closer to the support, thus making the quadrature rules less accurate. In Figure 5.1(a) the rectangles of the L-ribbon (shown in blue) are invisible for values of μ close to $1/\sqrt{10}$, but fairly large for values of μ near $10^{-2.5}$.

Similarly, the curvature-ribbon in Figure 5.1(b) is "thinner," indicating that the curvature bounds are tighter, for large values of μ than for small values. Here the dashed (blue) graph shows $-\kappa^-_\mu$ and the continuous (black) graph shows $-\kappa^+_\mu$.

When ℓ Lanczos bidiagonalization steps are carried out, the Gauss and Gauss-Radau rules determined have ℓ or $\ell+1$ nodes. We therefore expect the computed bounds to improve as ℓ increases. This, indeed, can be observed in Figures 5.2, which is analogous to Figure 5.1, and correspond to $\ell = 9$ Lanczos bidiagonalization steps. The rectangles of the L-ribbon can be seen to be significantly smaller in Figure 5.2(a) than in Figure 5.1(a). Moreover, the upper and lower bounds of the curvature-ribbon are so close that they cannot be distinguished in Figure 5.2(b).

It is natural to increase the number of Lanczos bidiagonalization steps ℓ until the upper and lower curvature bounds are within a prescribed tolerance in an interval around the location of the maximum absolute value of the curvature. In the present example, this condition may be satisfied when either $\ell = 8$ or $\ell = 9$ Lanczos bidiagonalization steps have been carried out, depending on the prescribed tolerance. Figure 5.2(b) indicates that $-\kappa_\mu$ is maximal for $\mu \approx 2 \cdot 10^{-2}$. The Galerkin solution obtained from (4.18) with $\mu_* = 2 \cdot 10^{-2}$ and $\ell = 9$ is shown in Figure 5.3 (blue dashed curve). The figure also displays the exact solution (5.2) of the noise-free continuous problem (black continuous curve).

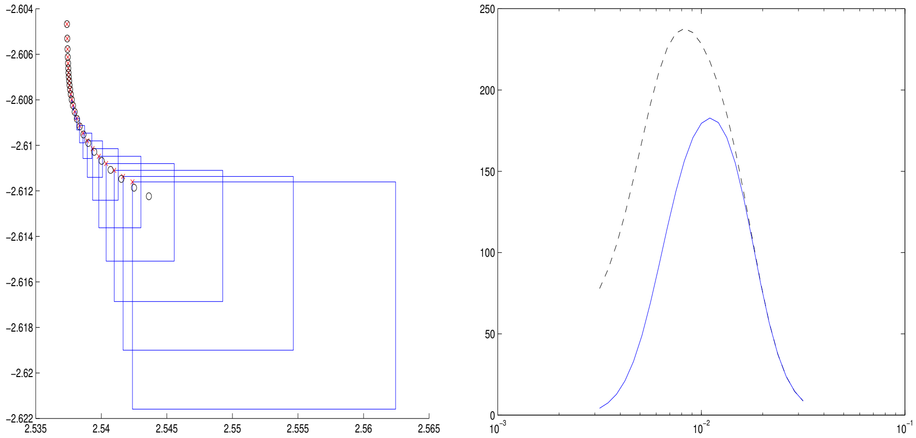
This example illustrates that by using both the L- and curvature-ribbons, it is easy to determine a suitable number of Lanczos bidiagonalization steps and a value of the regularization parameter that corresponds to a point in the vicinity of the vertex of the L-curve. □

Example 5.2. The integral equation (5.1) in this example is a test problem from geomagnetic prospecting (see [15]), namely, a deconvolution problem with

$$K(s,t) := \frac{d}{(d^2 + (s-t)^2)^{3/2}}, \qquad d := \frac{1}{4},$$
$$x(t) := \sin(\pi t) + \frac{1}{2}\sin(2\pi t), \tag{5.3}$$

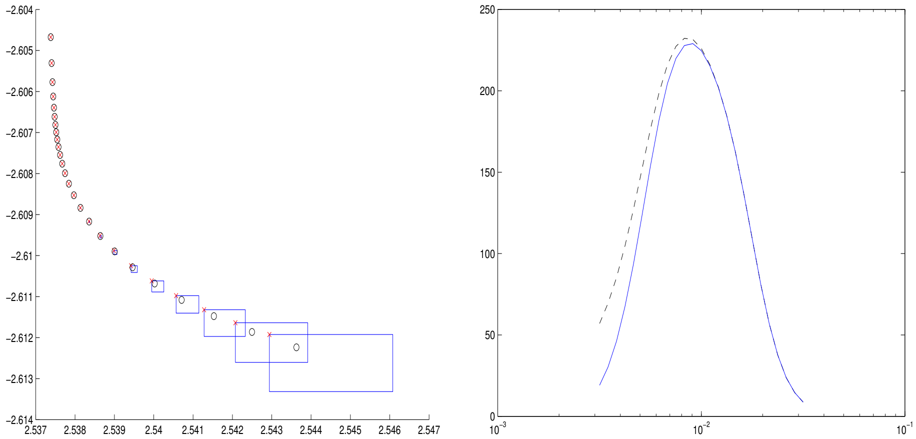
and $\alpha := 0$, $\beta := 1$. The right-hand side function b in (5.1) models the vertical component of the magnetic field from a source distribution x of magnetic dipoles at depth d. Dis-

ETNA
Kent State University
etna@mcs.kent.edu



(a) (b)

FIG. 5.4. *Example 5.2: 12 Lanczos bidiagonalization steps: (a) L-ribbon (4.7), (b) curvature-ribbon (4.17).*



(a) (b)

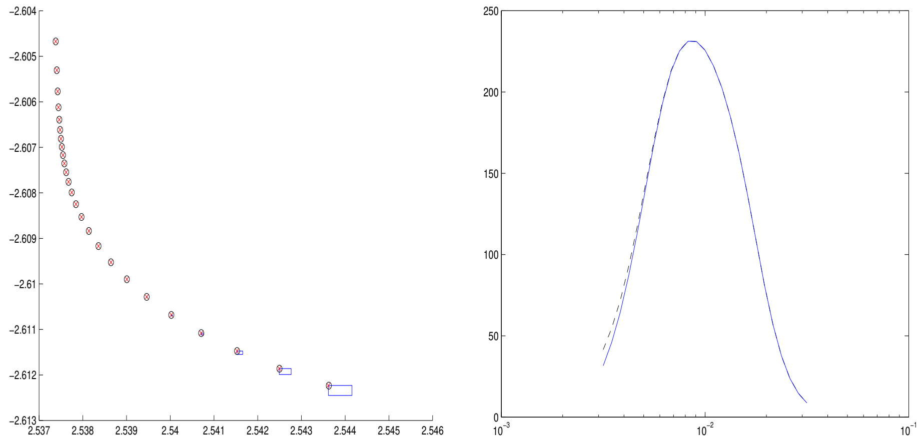
FIG. 5.5. *Example 5.2: 13 Lanczos bidiagonalization steps: (a) L-ribbon (4.7), (b) curvature-ribbon (4.17).*

cretization by a simple Nyström method gives the symmetric, numerically singular, matrix $A \in \mathbb{R}^{256\times 256}$ of the linear system (1.1), similarly as in Example 5.1. Let $\hat{\mathbf{x}} \in \mathbb{R}^{256}$ be the tabulation of the solution (5.3) at the nodes of the quadrature rule, and define the "noisy" right-hand side vector of (1.1) by $\mathbf{b} := A\hat{\mathbf{x}} + \mathbf{e}$, where $\mathbf{e} \in \mathbb{R}^{256}$ is an unbiased Gaussian noise vector with $\|\mathbf{e}\|/\|A\hat{\mathbf{x}}\| = 1 \cdot 10^{-3}$. We seek to compute an accurate approximation of $\hat{\mathbf{x}}$ by determining a suitable approximate solution of the linear system (1.1).

Figure 5.4 shows the L-ribbon (4.7) and curvature-ribbon (4.17) obtained with $\ell = 12$ Lanczos bidiagonalization steps for $10^{-2.5} \leq \mu \leq 10^{-1.5}$. The figure is analogous to Figure 5.1. The large rectangles of the L-ribbon shown in Figure 5.4(a) and the large distance between the curves of $-\kappa_\mu^+$ and $-\kappa_\mu^-$ in Figure 5.4(b) near the maximum of $-\kappa_\mu$ suggests that more Lanczos bidiagonalization steps should be carried out.

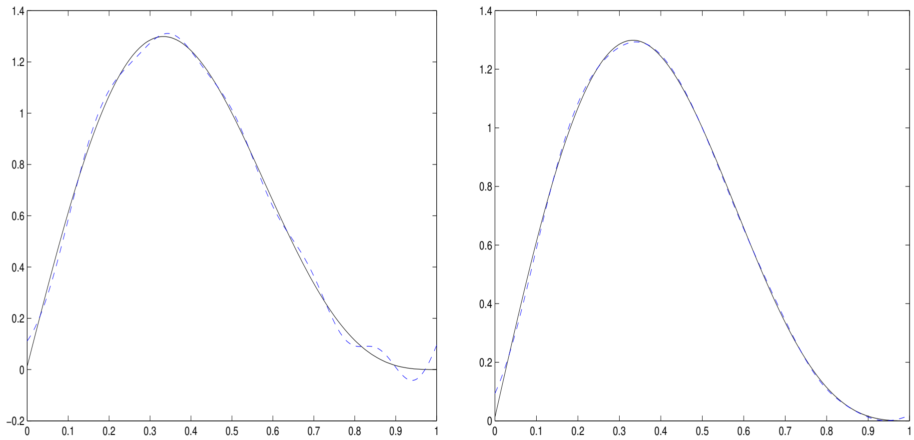
Figure 5.5 is analogous to Figure 5.4 and shows the L- and curvature-ribbons obtained with $\ell = 13$ Lanczos bidiagonalization steps. The rectangles in Figure 5.5(a) are smaller than

ETNA
Kent State University
etna@mcs.kent.edu



(a) (b)

FIG. 5.6. *Example 5.2: 14 Lanczos bidiagonalization steps: (a) L-ribbon (4.7), (b) curvature-ribbon (4.17).*



(a) (b)

FIG. 5.7. *Example 5.2: 14 Lanczos bidiagonalization steps: (a) Computed solution associated with $\mu = 9 \cdot 10^{-3}$ (blue dashed curve) and exact solution to the noise-free continuous problem (black solid curve), (b) Computed solution associated with $\mu = 3 \cdot 10^{-2}$ (blue dashed curve) and exact solution to the noise-free continuous problem (black solid curve).*

in Figure 5.4(a), and the distance between the curves of $-\kappa_\mu^+$ and $-\kappa_\mu^-$ is smaller in Figure 5.5(b) than in Figure 5.4(b). The former figure suggests that $-\kappa_\mu$ achieves its maximum for about $\mu = 9 \cdot 10^{-3}$. This is confirmed by Figure 5.6, which shows the L- and curvature-ribbons obtained with $\ell = 14$ Lanczos bidiagonalization steps.

The Galerkin solution obtained from (4.18) with $\mu_* = 9 \cdot 10^{-3}$ and $\ell = 14$ (blue dashed curve) and the exact solution (5.3) of the noise-free continuous problem (black continuous curve) are shown in Figure 5.7(a). We remark that the computed Galerkin solutions for $\mu_* = 9 \cdot 10^{-3}$ and $\ell = 12$ or $\ell = 13$ look the same as the Galerkin solution displayed.

The value of the regularization parameter μ associated with the vertex of the L-curve often is somewhat too small in the sense that for many linear discrete ill-posed problems (1.1) a better approximation of the solution of the associated, and generally unknown, noise-

ETNA
Kent State University
etna@mcs.kent.edu

free problem (1.2) can be obtained by choosing a value of μ that is somewhat larger than the value associated with the vertex. For instance, the Galerkin solution obtained from (4.18) with $\mu_* = 3\cdot 10^{-2}$, shown in Figure 5.7(b), yields a better approximation of the exact solution (5.3) of the noise-free continuous problem than the Galerkin solution shown in Figure 5.7(a). Nevertheless, the L- and curvature-ribbons provide valuable estimates of the size of suitable values of the regularization parameter for many linear discrete ill-posed problems (1.1). □

6. Conclusion. We have derived computable bounds for the log residual norm, the log solution norm, and the curvature of the L-curve. The bounds are inexpensive to compute during a Lanczos bidiagonalization process, thus providing a convenient way for using the L-curve as an aid for choosing the regularization parameter. In addition, the bounds provide information about when to stop the Lanczos iterations, namely, when the bounds are sufficiently close for regularization parameters in an interval around the optimal one. This is an important step towards making Lanczos bidiagonalization a practical general-purpose regularization algorithm.

Acknowledgement. We would like to thank Abdallah Shuibi for helpful comments.

REFERENCES

[1] Å. Björck, *A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations*, BIT 18 (1988), pp. 659–670.
[2] D. Calvetti, G. H. Golub and L. Reichel, *Estimation of the L-curve via Lanczos bidiagonalization*, BIT 39 (1999), pp. 603–619.
[3] 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.
[4] D. Calvetti, S. Morigi, L. Reichel and F. Sgallari, *An L-ribbon for large underdetermined linear discrete ill-posed problems*, Numer. Algorithms 25 (2000), pp. 89–107.
[5] L. Eldén, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT 17 (1977), pp. 134–145.
[6] G. H. Golub, *Some modified matrix eigenvalue problems*, SIAM Review 15 (1973), pp. 318–334.
[7] 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.
[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, eds. G. H. Golub, S. H. Lui, F. Luk and R. Plemmons, Springer, New York, 1997.
[10] P. C. Hansen, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Review 34 (1992), pp. 561–580.
[11] P. C. Hansen, *Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms 6 (1994), pp. 1–35. Software is available in Netlib at the web site `http://www.netlib.org`.
[12] P. C. Hansen, *The L-curve and its use in the numerical treatment of inverse problems*, in Computational Inverse Problems in Electrocardiology, ed. P. Johnston, Advances in Computational Bioengineering, vol. 4, WIT Press, Southampton, 2000, pp. 119–142.
[13] P. C. Hansen and D. P. O'Leary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput. 14 (1993), pp. 1487–1503.
[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] G. M. Wing and J. D. Zahrt, A Primer on Integral Equations of the First Kind, SIAM, Philadelphia, 1991.