BIT
2002, Vol. 43, No. 2, pp. 001–004

0006-3835/00/4004-0001 $15.00

AN ITERATIVE LAVRENTIEV REGULARIZATION METHOD *

S. MORIGI[1], L. REICHEL[2,†] and F. SGALLARI[3‡]

In Memory of Germund Dahlquist.

[1] *Dipartimento di Matematica, Universitá degli Studi di Bologna Piazza Porta S. Donato 5, 40127 Bologna, Italy. email: morigi@dm.unibo.it*

[2] *Department of Mathematical Sciences, Kent State University Kent, OH 44242, USA. email: reichel@math.kent.edu*

[3] *Dipartimento di Matematica CIRAM, Universitá degli Studi di Bologna Via Saragozza 8, 40123 Bologna, Italy. email: sgallari@dm.unibo.it*

Abstract.

This paper presents an iterative method for the computation of approximate solutions of large linear discrete ill-posed problems by Lavrentiev regularization. The method exploits the connection between Lanczos tridiagonalization and Gauss quadrature to determine inexpensively computable lower and upper bounds for certain functionals. This approach to bound functionals was first described in a paper by Dahlquist, Eisenstat, and Golub. A suitable value of the regularization parameter is determined by a modification of the discrepancy principle.

AMS subject classification: 65R30, 65R32, 65F10.

Key words: Ill-posed problem, regularization, Lanczos tridiagonalization, Gauss quadrature, discrepancy principle.

1 Introduction

We are concerned with the iterative solution of linear systems of equations of the form

$$Hx = g, \qquad H \in \mathbb{R}^{n\times n}, \qquad x \in \mathbb{R}^n, \qquad g \in \mathbb{R}^n, \tag{1.1}$$

where H is a large, symmetric positive semidefinite matrix of ill-determined rank. The eigenvalues of H are assumed to "cluster" at the origin. In particular, H is severely ill-conditioned and may be singular. In the latter case, we replace the linear system (1.1) by a least-squares minimization problem.

Linear systems of equations with a matrix of ill-determined rank are commonly referred to as linear discrete ill-posed problems. Such systems arise, e.g., from the

*Received XXXXXXXXX. Revised XXXXXXXXX. Communicated by XXXXXXXX.
†Research supported in part by NSF grant DMS-0107858.
‡Research supported in part by PRIN-2004 grant 2004014411_005.

discretization of linear ill-posed problems, such as Fredholm integral equations of the first kind with a smooth kernel.

Many linear ill-posed problems are so-called inverse problems, whose solution yields the cause of an observed effect. The data, i.e., the observed effect, is represented by the right-hand side vector g. Typically the available data is contaminated by a measurement error $e \in \mathbb{R}^n$, i.e.,

$$g = \hat{g} + e, \tag{1.2}$$

where $\hat{g}$ is the unknown error-free right-hand side vector associated with g. The vector $\hat{g}$ is assumed to be in the range of H.

We would like to determine the least-squares solution of minimal Euclidean norm, denoted by $\hat{x} \in \mathbb{R}^n$, of the linear system of equation

$$Hx = \hat{g} \tag{1.3}$$

with error-free right-hand side. It is given by

$$\hat{x} = H^{\dagger}\hat{g}, \tag{1.4}$$

where $H^{\dagger}$ denotes the Moore-Penrose pseudo-inverse of H.

Since $\hat{g}$ is not available, we seek to determine an approximation of $\hat{x}$ by computing an approximate solution of (1.1). Note that due to the ill-conditioning of H, the vector $H^{\dagger}e$ typically is of very large norm, and therefore the exact solution of (1.1), given by $H^{\dagger}g = \hat{x} + H^{\dagger}e$, generally is a very poor approximation of $\hat{x}$. A better approximation of $\hat{x}$ can be determined by replacing the linear system (1.1) by a nearby system, whose solution is less sensitive to the error e in g, and use the solution of the latter system as an approximation of $\hat{x}$. This replacement is commonly referred to as regularization.

Lavrentiev regularization replaces (1.1) by the linear system of equations

$$(H + \frac{1}{\beta}I)x = g, \tag{1.5}$$

where $\beta > 0$ is a suitably chosen scalar; see, e.g., Bakushinsky and Goncharsky [1, Section 4.2.1] for a discussion of this method. We refer to β as the regularization parameter.

Many applications of Lavrentiev regularization are reported in the literature, including magnetic resonance imaging, see, e.g., Liang and Lauterbur [21], and function approximation by neural networks; see, e.g., Girosi et al. [12]. An application to image deblurring is discussed in Section 5. In the literature $\lambda = 1/\beta$ often is referred to as the regularization parameter, however, for our method it is more convenient to consider β the regularization parameter; see below.

The solution

$$x_\beta = (H + \frac{1}{\beta}I)^{-1}g \tag{1.6}$$

of the Lavrentiev equation (1.5) depends on the value β; for instance

$$\lim_{\beta \searrow 0} x_\beta = 0, \qquad \lim_{\beta \to \infty} x_\beta = \begin{cases} H^{\dagger}g & \text{if (1.1) is consistent,} \\ \infty & \text{if (1.1) is inconsistent.} \end{cases} \tag{1.7}$$

Generally, none of the limits[1] (1.7) of x_β yields a meaningful approximation of $\hat{x}$. It is an important part of the computation of an approximate solution of (1.1) to determine a suitable value of β.

In the present paper, we will assume that the norm of the error

$$\varepsilon = \|e\|, \tag{1.8}$$

or an estimate thereof, are available, and we describe how a suitable value of β can be computed by using the connection between partial Lanczos tridiagonalization of H and Gauss quadrature rules. Throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm or the induced matrix norm.

Our approach for determining β is based on a technique first described by Dahlquist et al. [9] for the computation of error bounds for approximate solutions determined by an iterative method. Dahlquist et al. apply this technique to the Jacobi iterative method in [9] and to the conjugate gradient method in [10]. Extensions and further applications have been presented by Golub and von Matt [16], and Golub and Meurant [13, 14]. Applications to the determination of a suitable value of the regularization parameter in Tikhonov regularization are described by Golub and von Matt [17] and Calvetti et al. [2, 3, 4, 6, 7]. Recent applications to the computation of error bounds for iterative methods, using the Euclidean norm, can be found in [5, 14]. Further applications are described in the references of the cited works. Thus, the technique pioneered by Dahlquist et al. [9, 10] has over the years lead to the development of many numerical methods.

This paper is organized as follows. Section 2 introduces a parameter choice method that can be used with Lavrentiev regularization, Section 3 exploits the connection between the Lanczos process and Gauss-type quadrature to help determine a suitable value of the regularization parameter, and Section 4 discusses the computation of regularized approximate solutions. Numerical examples are presented in Section 5 and concluding remarks can be found in Section 6.

2 A parameter choice method for Lavrentiev regularization

Introduce the function

$$\phi(\beta) = \|Hx_\beta - g\|^2, \tag{2.1}$$

where x_β is given by (1.6). We can express ϕ in the form

$$\phi(\beta) = g^T(\beta H + I)^{-2} g. \tag{2.2}$$

Substituting the spectral factorization of H into (2.2) shows the following properties.

PROPOSITION 2.1. *Let the function ϕ be defined by (2.1) and assume that $Hg \neq 0$. Then ϕ is decreasing and strictly convex for $\beta \geq 0$ with*

$$\phi(0) = \|g\|^2, \qquad \lim_{\beta\to\infty} \phi(\beta) = \|g_0\|^2,$$

where g_0 denotes the orthogonal projection of g onto the null space of H. In particular, if H is nonsingular, then $\lim_{\beta\to\infty} \phi(\beta) = 0$.

[1] The limit ∞ denotes that norm of the limiting vector is unbounded.

We remark that our reason for choosing the regularization parameter β, rather than $\lambda = 1/\beta$, is to secure convexity of ϕ. The analog of the function ϕ for the parameter λ is not guaranteed to be convex.

It follows from the proposition that the equation

$$\phi(\beta) = \tau$$

has a unique solution β, such that $0 < \beta < \infty$, for any τ with $\|g_0\|^2 < \tau < \|g\|^2$.

Let $0 < s < 1$ be fixed and let $\beta = \beta_\varepsilon$ solve the equation

$$\phi(\beta) = \varepsilon^{2s}, \tag{2.3}$$

where we assume that $\|g_0\|^2 < \varepsilon^{2s} < \|g\|^2$. We will show below that this choice of regularization parameter gives solutions x_{β_ϵ} of the Lavrentiev equation (1.5) that converge to $\hat{x}$ as ε decreases to zero. Groetsch and Guacamene [18] present an analysis in Hilbert space which shows that convergence cannot be guaranteed when $s = 1$, and they propose that the regularization parameter β should be chosen so that

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

We focus on the parameter choice method (2.3) because it is better suited for the numerical method of Section 3 than the choice (2.4).

The following theorem can be shown in a similar fashion as the convergence proof of the parameter choice method (2.4) by Groetsch and Guacamene [18]. Our proof is more elementary and sheds light on the role of the parameter s. We therefore believe it to be of interest. Theorem 2.2 is formulated in finite dimensions, however, the proof carries over to an (infinite-dimensional) Hilbert space setting.

Theorem 2.2. *Let $0 < s < 1$ be fixed and let $\beta = \beta_\varepsilon$ satisfy (2.3). Then x_{β_ε}, defined by (1.6) with $\beta = \beta_\varepsilon$, satisfies*

$$\lim_{\varepsilon \searrow 0} \sup_{\|g-\hat{g}\| \leq \varepsilon} \|x_{\beta_\varepsilon} - \hat{x}\| = 0, \tag{2.5}$$

where $\hat{x}$ is given by (1.4).

Proof. We note for future reference that

$$\|(H + \frac{1}{\beta}I)^{-1}\| \leq \beta, \qquad \|(H + \frac{1}{\beta}I)^{-1}H\| \leq 1, \qquad \beta > 0, \tag{2.6}$$

where $\|\cdot\|$ denotes the spectral norm. It follows from (1.6) that for any $\beta > 0$,

$$Hx_\beta - g = Hx_\beta - (H + \frac{1}{\beta}I)x_\beta = -\frac{1}{\beta}x_\beta,$$

and letting $\beta = \beta_\varepsilon$ yields

$$\|x_{\beta_\varepsilon}\| = \beta_\varepsilon\|Hx_{\beta_\varepsilon} - g\| = \beta_\varepsilon\varepsilon^s. \tag{2.7}$$

Let $\hat{x}_\beta$ for $\beta > 0$ denote the regularized approximate solutions of (1.3) associated with the exact data $\hat{g}$,

$$\hat{x}_\beta = (H + \frac{1}{\beta}I)^{-1}\hat{g}. \tag{2.8}$$

It follows from (1.2), (1.8), and (2.6) that

$$\|x_{\beta_\varepsilon} - \hat{x}_{\beta_\varepsilon}\| = \|(H + \frac{1}{\beta_\varepsilon} I)^{-1}(g - \hat{g})\| \leq \beta_\varepsilon \varepsilon. \tag{2.9}$$

We now derive a bound for the right-hand side of (2.9). Note that β_ε grows as ε decreases. Using (2.9) and (2.7), in order, we obtain

$$\|x_{\beta_\varepsilon}\| \leq \|\hat{x}_{\beta_\varepsilon}\| + \|x_{\beta_\varepsilon} - \hat{x}_{\beta_\varepsilon}\| \leq \|\hat{x}_{\beta_\varepsilon}\| + \beta_\varepsilon \varepsilon = \|\hat{x}_{\beta_\varepsilon}\| + \|x_{\beta_\varepsilon}\| \varepsilon^{1-s}.$$

Thus,

$$\|x_{\beta_\varepsilon}\|(1 - \varepsilon^{1-s}) \leq \|\hat{x}_{\beta_\varepsilon}\|, \tag{2.10}$$

where we may assume that $0 < \varepsilon < 1$.

Equation (2.8) yields

$$\hat{x}_\beta = (H + \frac{1}{\beta} I)^{-1} H \hat{x}$$

and it follows from (2.6) that

$$\|\hat{x}_\beta\| \leq \|\hat{x}\|, \qquad \beta > 0. \tag{2.11}$$

This inequality together with (2.10) shows that

$$\|x_{\beta_\varepsilon}\| \leq \frac{1}{1 - \varepsilon^{1-s}} \|\hat{x}\|$$

and, therefore, in view of (2.7),

$$\beta_\varepsilon \varepsilon \leq \frac{\varepsilon^{1-s}}{1 - \varepsilon^{1-s}} \|\hat{x}\|. \tag{2.12}$$

We are in a position to show (2.5). The inequalities (2.9) and (2.12) yield

$$\|x_{\beta_\varepsilon} - \hat{x}\| \leq \|x_{\beta_\varepsilon} - \hat{x}_{\beta_\varepsilon}\| + \|\hat{x}_{\beta_\varepsilon} - \hat{x}\| \leq \frac{\varepsilon^{1-s}}{1 - \varepsilon^{1-s}} \|\hat{x}\| + \|\hat{x}_{\beta_\varepsilon} - \hat{x}\| \tag{2.13}$$

and, using (2.11), we obtain

$$\|\hat{x}_{\beta_\varepsilon} - \hat{x}\|^2 = \|\hat{x}_{\beta_\varepsilon}\|^2 - 2\hat{x}_{\beta_\varepsilon}^T \hat{x} + \|\hat{x}\|^2 \leq 2\|\hat{x}\|^2 - 2\hat{x}_{\beta_\varepsilon}^T \hat{x}. \tag{2.14}$$

Let $\hat{z} = H^\dagger \hat{x}$. Since $\hat{x}$ is in the range of H, we have $\hat{x} = H\hat{z}$. The vector $\hat{z}$ always exists in a finite-dimensional setting. In an infinite-dimensional Hilbert space, the existence of $\hat{z}$ of finite norm may be considered a smoothness requirement on $\hat{x}$. We obtain

$$\begin{aligned} \|\hat{x}\|^2 - \hat{x}_{\beta_\varepsilon}^T \hat{x} &= (\hat{x} - x_{\beta_\varepsilon})^T H \hat{z} = (H\hat{x} - g + g - Hx_{\beta_\varepsilon})^T \hat{z} \\ &\leq \|H\hat{x} - g\|\|\hat{z}\| + \|Hx_{\beta_\varepsilon} - g\|\|\hat{z}\| \leq (\varepsilon + \varepsilon^s)\|\hat{z}\|. \end{aligned}$$

Combining this inequality with (2.13) and (2.14) shows the theorem. □

REMARK 2.1. *The above proof suggests that the choice of the parameter s may be important for the success of the regularization method for finite values of ε. Computed examples reported in Section 5 illustrate that this, indeed, is the case.*

REMARK 2.2. *The proof by Groetsch and Guacamene [18, Theorem] of convergence of the parameter choice method (2.4) circumvents the introduction of the vector $\hat{z}$, used in the proof above, by applying properties of weakly convergent subsequences. This approach also can be used to show Theorem 2.2.*

3 Lanczos tridiagonalization and Gauss quadrature

Application of ℓ steps of Lanczos tridiagonalization to H with initial vector g yields the partial Lanczos decomposition

$$HV_\ell = V_\ell T_\ell + f_\ell e_\ell^T, \tag{3.1}$$

where $T_\ell \in \mathbb{R}^{\ell\times\ell}$ is a symmetric tridiagonal matrix. Further, $f_\ell \in \mathbb{R}^n$ and $V_\ell \in \mathbb{R}^{n\times\ell}$ satisfy $V_\ell^T V_\ell = I_\ell$, $V_\ell e_1 = g/\|g\|$, and $V_\ell^T f_\ell = 0$. Throughout this paper I_ℓ denotes the $\ell \times \ell$ identity matrix and $e_j = [0,\dots,0,1,0,\dots,0]^T$ the jth axis vector of appropriate dimension. We assume that ℓ is sufficiently small so that the decomposition (3.1) with the stated properties exists; see, e.g., [15, Chapter 9] for details on the Lanczos process.

Introduce the Krylov subspace

$$\mathcal{K}_\ell(H,g) = \mathrm{span}\{g, Hg, \dots, H^{\ell-1}g\}. \tag{3.2}$$

We note for future reference that

$$\mathrm{range}(V_\ell) = \mathcal{K}_\ell(H,g). \tag{3.3}$$

Since T_ℓ is an orthogonal projection of H, and H is positive semidefinite, so is T_ℓ. The following theorem shows that unless $f_\ell = 0$, the matrix T_ℓ is positive definite.

Theorem 3.1. *Let the matrix $H \in \mathbb{R}^{n\times n}$ be symmetric and positive semidefinite. Then the symmetric tridiagonal matrix $T_\ell \in \mathbb{R}^{\ell\times\ell}$ in (3.1) is positive definite if $f_\ell \neq 0$.*

Proof. The matrices T_j, $j = 1,2,3,\dots$, are orthogonal projections of H. They therefore have no negative eigenvalues. Assume that T_ℓ is singular and $f_\ell \neq 0$. Then one more step of the Lanczos process can be carried out to yield the symmetric tridiagonal matrix $T_{\ell+1}$ with leading principal $\ell\times\ell$ submatrix T_ℓ and positive subdiagonal entries. The eigenvalues of $T_{\ell+1} \in \mathbb{R}^{(\ell+1)\times(\ell+1)}$ strictly interlace the eigenvalues of T_ℓ. This follows, e.g., from the connection between the eigenvalues of symmetric tridiagonal matrices and zeros of orthogonal polynomials. For a proof of the interlacing of the latter; see, e.g., Gautschi [11] or Szegő [24, Section 3.3]. It follows from the strict interlacing of the eigenvalues that $T_{\ell+1}$ has a negative eigenvalue. This contradiction shows that T_ℓ is positive definite. □

If $f_\ell = 0$, then the computations simplify. We comment on this situation at the end of this section. For now, we will assume that $f_\ell \neq 0$.

Introduce the symmetric tridiagonal matrix

$$T_{\ell+1}^{(0)} = \begin{bmatrix} T_\ell & \|f_\ell\| e_\ell \\ \|f_\ell\| e_\ell^T & \alpha \end{bmatrix} \in \mathbb{R}^{(\ell+1)\times(\ell+1)},$$

where $\alpha = \|f_\ell\|^2 e_\ell^T T_\ell^{-1} e_\ell$. The matrices T_ℓ and $T_{\ell+1}^{(0)}$ can be applied to compute inexpensive lower and upper bounds for the function ϕ, by using the connection between the Lanczos process and Gauss-type quadrature rules. Substituting the spectral decomposition of H into (2.2) and expressing the resulting sum as a Stieltjes integral yields

$$\phi(\beta) = \int_0^\infty \psi(t)\,dw(t), \qquad \psi(t) = (\beta t + 1)^{-2}. \tag{3.4}$$

It can be shown, using the techniques first described by Dahlquist et al. [9, 10], and further developed by Golub and Meurant [13], that

$$\phi_\ell(\beta) = \|g\|^2 e_1^T(\beta T_\ell + I_\ell)^{-2} e_1 \tag{3.5}$$

is an ℓ-point Gauss quadrature rule for the approximation of the integral (3.4). Since the derivative of order 2ℓ of the integrand ψ is positive for $t \geq 0$, the remainder formula for Gauss quadrature yields that

$$\phi_\ell(\beta) < \phi(\beta), \qquad \beta > 0;$$

see, e.g., [6, 9, 10, 13] for details. Similarly, one can show that

$$\bar{\phi}_\ell(\beta) = \|g\|^2 e_1^T(\beta T_{\ell+1}^{(0)} + I_{\ell+1})^{-2} e_1 \tag{3.6}$$

is an $(\ell+1)$-point Gauss-Radau quadrature rule with a prescribed node at the origin for the approximation of (3.4). Since the derivative of order $2\ell+1$ of the integrand ψ is negative for $t \geq 0$, it follows from the remainder formula for Gauss-Radau quadrature that

$$\bar{\phi}_\ell(\beta) > \phi(\beta), \qquad \beta > 0.$$

PROPOSITION 3.2. *The function ϕ_ℓ defined by (3.5) is strictly decreasing and convex for $\beta \geq 1$. Moreover,*

$$\phi_\ell(0) = \|g\|^2, \qquad \lim_{\beta\to\infty} \phi_\ell(\beta) = 0.$$

PROOF. Let ϕ_ℓ' and ϕ_ℓ'' denote the first and second derivatives of ϕ_ℓ. It follows from (3.5) that $\phi_\ell(\beta) > 0$, $\phi_\ell'(\beta) < 0$, and $\phi_\ell''(\beta) > 0$ for $\beta \geq 0$. Moreover, by Theorem 3.1 the matrix T_ℓ is positive definite, which establishes the limit of ϕ_ℓ as $\beta \to \infty$. □

Proposition 3.2 shows that the equation

$$\phi_\ell(\beta) = \tau$$

has a unique solution β, such that $0 < \beta < \infty$, for any τ with $0 < \tau < \|g\|^2$.

PROPOSITION 3.3. *The function $\bar{\phi}_\ell$ defined by (3.6) is strictly decreasing and convex for $\beta \geq 1$. Moreover,*

$$\bar{\phi}_\ell(0) = \|g\|^2, \qquad \lim_{\beta\to\infty} \bar{\phi}_\ell(\beta) = \omega_0,$$

where $\omega_0 > 0$ is the weight of the $(\ell+1)$-point Gauss-Radau rule for the node at the origin.

PROOF. Substituting the spectral decomposition of $T_\ell^{(0)}$ into (3.6) and using the relation between the first eigenvector components and the weights gives the limit of $\bar{\phi}_\ell(\beta)$ as $\beta \to \infty$; see, e.g., Gautschi [11, Section 3.1] for details on this relation. The remaining properties are established similarly as in the proof for Proposition 3.2. □

THEOREM 3.4. *Let β_j and $\bar{\beta}_j$ satisfy $\phi_j(\beta_j) = \varepsilon^{2s}$ and $\bar{\phi}_j(\bar{\beta}_j) = \varepsilon^{2s}$ for $j = \ell-1, \ell$. Then*

$$\beta_{\ell-1} < \beta_\ell < \beta_\varepsilon < \bar{\beta}_\ell < \bar{\beta}_{\ell-1},$$

where β_ε satisfies (2.3).

PROOF. The inequalities $\beta_\ell < \beta_\varepsilon < \bar{\beta}_\ell$ follow from Propositions 3.2 and 3.3. The other inequalities follow from Hanke [19, Theorem]. □

Assume that $0 < \varepsilon^{2s} < \|g\|^2$ and let β_ℓ denote the unique solution of

$$\phi_\ell(\beta) = \varepsilon^{2s} \tag{3.7}$$

for $\ell \geq 1$. We will for increasing values of ℓ, starting with $\ell = 1$, compute β_ℓ until a sufficiently accurate approximation of β_ε has been determined. A condition for sufficiency is discussed below.

It is convenient to solve (3.7) for $\ell \geq 1$ by Newton's method with initial approximate solution $\beta_{\ell-1}$, where we define $\beta_0 = 0$. The monotonic decrease and convexity of ϕ_ℓ, cf. Proposition 3.2, secure monotonic and quadratic convergence of Newton's method.

For each fixed $\beta > 0$, the function $\phi_\ell(\beta)$ can be evaluated by first computing the solution $\tilde{y}_\ell$ of the linear system of equations

$$(\beta T_\ell + I_\ell)y = \|g\|e_1. \tag{3.8}$$

and then evaluating $\phi_\ell(\beta) = \tilde{y}_\ell^T \tilde{y}_\ell$. The Cholesky factorization of the matrix, and therefore the computation of $\tilde{y}_\ell$, require only $\mathcal{O}(\ell)$ arithmetic floating point operations.

Having computed β_ℓ, we check whether an associated solution x_ℓ of (1.5) with $\beta = \beta_\ell$ satisfies the inequality

$$\|Hx_\ell - g\| \leq \eta\varepsilon^s \tag{3.9}$$

for a user-specified constant $\eta > 1$. Section 4 discusses how this can be done without computing x_ℓ. If the inequality (3.9) holds, then we evaluate x_ℓ and terminate the computations; otherwise, we increase the value of ℓ. Details are presented in Section 4.

We finally comment on the case when f_ℓ vanishes in (3.1). One can show that in this situation $\phi(\beta) = \phi_\ell(\beta)$ for all $\beta \geq 0$ and, therefore, the computations simplify. We will not dwell on this case further since it is very rare.

4 Computation of a regularized approximate solution

Let β_ℓ be a computed approximation of β_ε. We discuss two Galerkin methods for computing an approximate solution of the Lavrentiev equation (1.5) with $\beta = \beta_\ell$.

Method 1 determines an approximate solution x_ℓ of (1.5) with $\beta = \beta_\ell$ in the Krylov subspace (3.2). We express x_ℓ as

$$x_\ell = V_\ell y_\ell, \tag{4.1}$$

cf. (3.3), where y_ℓ denotes the solution of the Galerkin equation

$$V_\ell^T \left(H + \frac{1}{\beta_\ell} I \right) V_\ell y = V_\ell^T g.$$

Using (3.1), this equation simplifies to

$$\left(T_\ell + \frac{1}{\beta_\ell} I_\ell \right) y = e_1 \|g\|. \tag{4.2}$$

A comparison of the systems (3.8) and (4.2) shows that $y_\ell = \beta_\ell \tilde{y}_\ell$, where $\tilde{y}_\ell$ solves (3.8). The vector $\tilde{y}_\ell$ can be assumed already to be available from the computation of β_ℓ and, therefore, y_ℓ can be evaluated simply by scaling $\tilde{y}_\ell$.

THEOREM 4.1. *Let y_ℓ solve (4.2) and let x_ℓ be given by (4.1). Then*

$$\|Hx_\ell - g\|^2 = \phi_\ell(\beta_\ell) + \|f_\ell\|^2 (e_\ell^T y_\ell)^2,$$

where ϕ_ℓ is defined by (3.5).

PROOF. Equations (3.1) and (4.2) yield

$$Hx_\ell - g = V_\ell T_\ell y_\ell + f_\ell e_\ell^T y_\ell - V_\ell e_1 \|g\| = -\frac{1}{\beta_\ell} V_\ell y_\ell + f_\ell e_\ell^T y_\ell,$$

and it follows from $V_\ell^T f_\ell = 0$ that

$$\begin{aligned}\|Hx_\ell - g\|^2 &= \frac{1}{\beta_\ell^2} y_\ell^T y_\ell + \|f_\ell\|^2 (e_\ell^T y_\ell)^2 = \tilde{y}_\ell^T \tilde{y}_\ell + \|f_\ell\|^2 (e_\ell^T y_\ell)^2 \\ &= \phi_\ell(\beta_\ell) + \|f_\ell\|^2 (e_\ell^T y_\ell)^2.\end{aligned}$$

□

The computations with Method 1 proceed as follows. For each $\ell = 1, 2, 3, \ldots$, we determine the root β_ℓ of equation (3.7) as described in Section 3. The solution $\tilde{y}_\ell$ of (3.8) is then available from which we determine the solution y_ℓ of (4.2). The inequality (3.9) with x_ℓ defined by (4.1) can be expressed as

$$\phi_\ell(\beta_\ell) + \|f_\ell\|^2 (e_\ell^T y_\ell)^2 \leq \eta^2 \varepsilon^{2s},$$

cf. Theorem 4.1, and using $\phi_\ell(\beta_\ell) = \varepsilon^{2s}$ yields

$$\|f_\ell\| \, |e_\ell^T y_\ell| \leq (\eta^2 - 1)^{1/2} \varepsilon^s. \tag{4.3}$$

If this inequality holds, then we evaluate x_ℓ from (4.1) and accept this vector as an approximate solution of (1.1); otherwise we increase ℓ by one and repeat the computations. We remark that the convergence of the left-hand side of (4.3) towards zero as ℓ increases is primarily driven by the typically rapid convergence of $|e_\ell^T y_\ell|$ towards zero. This completes the discussion of Method 1.

We turn to Method 2. When the desired solution $\hat{x}$ represents a smooth function, it can be beneficial to require the computed solution x_ℓ to live in a Krylov subspace of the form

$$\mathcal{K}_\ell(H, Hg) = \text{span}\{Hg, H^2 g, \ldots, H^\ell g\}, \tag{4.4}$$

instead of in a Krylov subspace of the type (3.2). The reason for this is that in many applications Hg represents a smoother function than g. Method 2 is a Galerkin method for determining an approximate solution of (1.5) in a Krylov subspace of the form (4.4). Since the Krylov subspace (4.4) is in the range of H, the computed approximate solutions are orthogonal to the null space of H, like the desired solution $\hat{x}$ given by (1.4).

The following manipulations are closely related to those required for the implicitly restarted Lanczos method; see, e.g., [8, 23] for further details on the latter. Let T_ℓ and V_ℓ be the matrices in the Lanczos decomposition (3.1), and introduce the QR-factorization

$$T_\ell = Q_\ell R_\ell, \tag{4.5}$$

i.e., $Q_\ell \in \mathbb{R}^{\ell\times\ell}$ is orthogonal and $R_\ell \in \mathbb{R}^{\ell\times\ell}$ is upper triangular. Multiplication of (3.1) by Q_ℓ from the right-hand side yields

$$HV'_\ell = V'_\ell T'_\ell + f_\ell e_\ell^T Q_\ell, \tag{4.6}$$

where $T'_\ell = R_\ell Q_\ell$ is symmetric and tridiagonal, and $V'_\ell = V_\ell Q_\ell$ has orthonormal columns. The property

$$V'_\ell e_1 = Hg/\|Hg\|, \qquad \ell \geq 2, \tag{4.7}$$

can be established as follows. Substitute the QR-factorization (4.5) into (3.1) and multiply the resulting expression from the right-hand side by e_1. We obtain

$$HV_\ell e_1 = V'_\ell R_\ell e_1 + f_\ell e_\ell^T e_1,$$

which for $\ell \geq 2$ yields (4.7).

Since T_ℓ is tridiagonal, the orthogonal matrix Q_ℓ in the factorization (4.5) is of upper Hessenberg form. It follows that all but the last two components of the vector $e_\ell^T Q_\ell$ in (4.6) are guaranteed to vanish. Hence, the decomposition (4.6) differs from a partial Lanczos decomposition only in that the last two columns of the matrix $f_\ell e_\ell^T Q_\ell$ may be nonvanishing. Removing the last column from each term in the decomposition (4.6) gives

$$HV''_{\ell-1} = V''_{\ell-1} T''_{\ell-1} + f''_{\ell-1} e_{\ell-1}^T, \tag{4.8}$$

where $V''_{\ell-1} \in \mathbb{R}^{n\times(\ell-1)}$ is made up of the first $\ell-1$ columns of the matrix V'_ℓ; in particular, the columns of $V''_{\ell-1}$ are orthonormal, and by (4.7), $V''_{\ell-1}e_1 = Hg/\|Hg\|$. Moreover, $T''_{\ell-1}$ is the leading $(\ell-1)\times(\ell-1)$ principal submatrix of T'_ℓ. The vector f''_ℓ is a linear combination of f_ℓ and the last column of V'_ℓ and therefore satisfies $(V''_{\ell-1})^T f''_{\ell-1} = 0$. Thus, the decomposition (4.8) is a partial Lanczos decomposition of H with

$$\text{range}(V''_{\ell-1}) = \mathcal{K}_{\ell-1}(H, Hg).$$

The decomposition (4.8) can be computed incrementally for increasing values of ℓ. Only the last few columns of the matrix V_ℓ have to be available for computing the last column of $V''_{\ell-1}$. Method 2 is based on the partial Lanczos decomposition (4.8). For every value of ℓ the method requires storage of the matrix $V''_{\ell-1}$ and of the last few columns of V_ℓ.

Let β_ℓ denote the current approximation of β_ε. The Galerkin equations

$$(V''_{\ell-1})^T \left(H + \frac{1}{\beta_\ell} I\right) V''_{\ell-1} y = (V''_{\ell-1})^T g,$$

simplify to

$$\left(T''_{\ell-1} + \frac{1}{\beta_\ell} I_{\ell-1}\right) y = (V''_{\ell-1})^T g. \tag{4.9}$$

The solution $y''_{\ell-1}$ of (4.9) determines the approximate solution

$$x''_{\ell-1} = V''_{\ell-1} y''_{\ell-1} \in \mathcal{K}_{\ell-1}(H, Hg) \tag{4.10}$$

of (1.5) with $\beta = \beta_\ell$.

The norm of the residual error associated with $x''_{\ell-1}$ can be computed fairly inexpensively by evaluating the right-hand side of

$$\|Hx''_{\ell-1} - g\| = \|V''_\ell \bar{T}''_{\ell-1} y''_{\ell-1} - g\|,$$

where $\bar{T}''_{\ell-1}$ is the leading $\ell \times (\ell - 1)$ submatrix of T''_ℓ. Method 2 increases the value of ℓ until a solution $y''_{\ell-1}$ of (4.9) that satisfies

$$\|V''_\ell \bar{T}''_{\ell-1} y''_{\ell-1} - g\| \leq \eta \varepsilon^s \tag{4.11}$$

has been determined. Then $x''_{\ell-1}$ is evaluated by (4.10) and accepted as an approximate solution of (1.5) with $\beta = \beta_\ell$.

5 Numerical results

All computed examples were carried out in Matlab with machine epsilon $2 \cdot 10^{-16}$.

Example 5.1. The Fredholm integral equation of the first kind

$$\int_{-6}^{6} \kappa(\tau, \sigma) x(\sigma) d\sigma = g(\tau), \qquad -6 \leq \tau \leq 6, \tag{5.1}$$

with kernel, right-hand side, and solution given by

$$\begin{aligned} \kappa(\tau, \sigma) &= x(\tau - \sigma), \\ g(\tau) &= (6 - |\tau|)(1 + \frac{1}{2}\cos(\frac{\pi}{3}\tau)) + \frac{9}{2\pi}\sin(\frac{\pi}{3}|\tau|), \\ x(\sigma) &= \begin{cases} 1 + \cos(\frac{\pi}{3}\sigma), & \text{if } |\sigma| < 3, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \tag{5.2}$$

respectively, is discussed by Phillips [22]. We discretize this integral equation by a Galerkin method using the Matlab code phillips from the program package Regularization Tools by Hansen [20] to obtain the symmetric matrix $H \in \mathbb{R}^{200\times 200}$, i.e., $n = 200$ in (1.1). The matrix is of ill-determined rank and has condition number $\kappa(H) = \|H\|\|H^{-1}\| = 4.23 \cdot 10^7$. Most of the eigenvalues of H are positive, but some are not. Lavrentiev regularization can be applied because the tridiagonal matrices in the reduced problems (4.2) and (4.9) have positive semidefinite matrices T_ℓ and $T''_{\ell-1}$, respectively. We comment on this further below. The code phillips also determines a scaled discretization of the solution (5.2), which we consider the desired solution $\hat{x} \in \mathbb{R}^n$ of the linear system (1.3) with the error-free right-hand side. The latter is determined by $\hat{g} = H\hat{x}$. An error vector e with normally distributed random entries with zero mean is added to $\hat{g}$ to yield the right-hand side g of (1.1); cf. (1.2). The vector e is scaled to correspond to a specified noise level

$$\gamma = \frac{\|e\|}{\|g\|}. \tag{5.3}$$

Table 5.1: Example 5.1: Discretization of the integral equation (5.1).

γ	method	ℓ	β_ℓ	$\|x_\ell - \hat{x}\|/\|\hat{x}\|$	$\|x_{\beta_\ell} - \hat{x}\|/\|\hat{x}\|$
$1 \cdot 10^{-1}$	1	3	$3.46 \cdot 10^{0}$	$5.09 \cdot 10^{-1}$	$6.44 \cdot 10^{-1}$
	2	4	$3.81 \cdot 10^{0}$	$1.55 \cdot 10^{-1}$	$7.03 \cdot 10^{-1}$
$1 \cdot 10^{-2}$	1	4	$2.10 \cdot 10^{1}$	$1.40 \cdot 10^{-1}$	$4.12 \cdot 10^{-1}$
	2	5	$2.25 \cdot 10^{1}$	$3.77 \cdot 10^{-2}$	$1.54 \cdot 10^{0}$
$1 \cdot 10^{-3}$	1	4	$1.33 \cdot 10^{2}$	$3.03 \cdot 10^{-2}$	$2.82 \cdot 10^{-1}$
	2	5	$1.35 \cdot 10^{2}$	$2.45 \cdot 10^{-2}$	$2.89 \cdot 10^{-1}$
$1 \cdot 10^{-4}$	1	7	$8.46 \cdot 10^{2}$	$6.18 \cdot 10^{-2}$	$1.72 \cdot 10^{-1}$
	2	10	$8.48 \cdot 10^{2}$	$5.93 \cdot 10^{-2}$	$1.73 \cdot 10^{-1}$

Table 5.2: Example 5.1: Discretization of the integral equation (5.1). Results for Methods 1 and 2 for $\gamma = 1 \cdot 10^{-3}$, $\eta = 1.1$, and several values of s.

mthd	s	$\|e\|^{2s}$	ℓ	β_ℓ	$\|x_\ell - \hat{x}\|/\|\hat{x}\|$	$\|x_{\beta_\ell} - \hat{x}\|/\|\hat{x}\|$
1	0.999	$1.01 \cdot 10^{-6}$	12	$1.48 \cdot 10^{3}$	$2.61 \cdot 10^{0}$	$4.34 \cdot 10^{1}$
1	0.900	$3.98 \cdot 10^{-6}$	10	$3.04 \cdot 10^{2}$	$5.48 \cdot 10^{-1}$	$5.62 \cdot 10^{-1}$
1	0.800	$7.94 \cdot 10^{-5}$	4	$1.33 \cdot 10^{2}$	$3.03 \cdot 10^{-2}$	$2.82 \cdot 10^{-1}$
1	0.700	$6.31 \cdot 10^{-5}$	4	$6.65 \cdot 10^{1}$	$3.13 \cdot 10^{-2}$	$1.90 \cdot 10^{-1}$
1	0.500	$1.00 \cdot 10^{-3}$	3	$1.62 \cdot 10^{1}$	$6.62 \cdot 10^{-2}$	$8.65 \cdot 10^{-2}$
2	0.999	$1.01 \cdot 10^{-6}$	10	$1.26 \cdot 10^{3}$	$6.48 \cdot 10^{-2}$	$3.79 \cdot 10^{0}$
2	0.900	$3.98 \cdot 10^{-6}$	8	$2.81 \cdot 10^{2}$	$3.97 \cdot 10^{-2}$	$5.15 \cdot 10^{-1}$
2	0.800	$1.58 \cdot 10^{-5}$	5	$1.35 \cdot 10^{2}$	$2.45 \cdot 10^{-2}$	$2.89 \cdot 10^{-1}$
2	0.700	$6.31 \cdot 10^{-5}$	5	$6.68 \cdot 10^{1}$	$2.60 \cdot 10^{-2}$	$1.96 \cdot 10^{-1}$
2	0.500	$1.00 \cdot 10^{-3}$	4	$1.63 \cdot 10^{1}$	$9.17 \cdot 10^{-2}$	$8.68 \cdot 10^{-2}$

Table 5.1 presents results obtained by Methods 1 and 2 for several noise levels γ. The parameters η and s in (3.9), (4.3), and (4.11) are chosen to be 1.1 and 0.8, respectively. The table shows ℓ to increase as the noise level γ decreases. Moreover, β_ℓ is seen to increase with ℓ, in agreement with Theorem 3.4. The computed approximate solutions of (1.1) determined by by Methods 1 and 2 are denoted by x_ℓ.

We have observed Method 2 to generally yield better approximations of the desired solution $\hat{x}$ than Method 1 when $\hat{x}$ is the discretization of a smooth function. This is the case in the present example. Indeed, for all noise levels reported, Method 2 gives more accurate approximations of $\hat{x}$ than Method 1.

The vectors x_{β_ℓ} in Table 5.1 denote solutions (1.6) of the Lavrentiev equation (1.5) with $\beta = \beta_\ell$. Note that x_{β_ℓ} does not approximate $\hat{x}$ as well as the corresponding approximate solutions x_ℓ determined by Methods 1 and 2. This depends on that Methods 1 and 2 regularize the linear system (1.1) both by the Lavrentiev method and by requiring the approximate solutions x_ℓ to live in an ℓ-dimensional Krylov subspace. We have observed this behavior in many examples, e.g., also in Example 5.2 below.

Table 5.3: Example 5.1: Discretization of the integral equation (5.1). Results for Method 2 for $\gamma = 1 \cdot 10^{-3}$, $\eta = 1.1$, $s = 0.8$ and several values of n.

n	ℓ	β_ℓ	$\|x_\ell - \hat{x}\|/\|\hat{x}\|$	$\|x_{\beta_\ell} - \hat{x}\|/\|\hat{x}\|$
100	4	$1.33 \cdot 10^2$	$3.14 \cdot 10^{-2}$	$2.54 \cdot 10^{-1}$
200	4	$1.33 \cdot 10^2$	$3.02 \cdot 10^{-2}$	$2.82 \cdot 10^{-1}$
300	4	$1.33 \cdot 10^2$	$3.07 \cdot 10^{-2}$	$2.75 \cdot 10^{-1}$
400	4	$1.33 \cdot 10^2$	$2.99 \cdot 10^{-2}$	$2.48 \cdot 10^{-1}$
500	4	$1.33 \cdot 10^2$	$3.08 \cdot 10^{-2}$	$2.54 \cdot 10^{-1}$
600	4	$1.33 \cdot 10^2$	$3.13 \cdot 10^{-2}$	$2.69 \cdot 10^{-1}$
800	4	$1.33 \cdot 10^2$	$3.06 \cdot 10^{-2}$	$2.71 \cdot 10^{-1}$
1000	4	$1.33 \cdot 10^2$	$3.12 \cdot 10^{-2}$	$2.62 \cdot 10^{-1}$

Theorem 2.2 guarantees convergence of x_{β_ε} to $\hat{x}$ as $\varepsilon \to 0$ for any fixed $0 < s < 1$. However, the difference $x_{\beta_\varepsilon} - \hat{x}$ for a fixed positive value of ε depends on the choice of s. This is illustrated by Table 5.2. The table also shows the differences $x_\ell - \hat{x}$, with x_ℓ determined by Methods 1 or 2. Table 5.2 suggests $s = 0.8$ to be a suitable choice for both Methods 1 and 2. We have found that for many linear discrete ill-posed problems with a smooth solution s should be between 0.8 and 0.9.

We return to the indefiniteness of the matrices H determined by the discretization of the integral equation (5.1). Table 5.3 shows computed results achieved by Method 2 for linear discrete ill-posed problems (1.1) for several values of n and $\gamma = 1 \cdot 10^{-3}$, $\eta = 1.1$, $s = 0.8$. The matrices $H \in \mathbb{R}^{n \times n}$ used for Table 5.3 all have negative eigenvalues, however, the computed symmtric tridiagonal matrices $T''_{\ell-1}$ in (4.9) generated by Method 2 are positive definite. Therefore, Lavrentiev regularization can be applied. The present example illustrates that the reduction of the original linear system (1.1) by the Lanczos process enlarges the class of problems to which Lavrentiev regularization can be applied; it suffices that the symmetric tridiagonal matrices T_ℓ and $T''_{\ell-1}$ in the reduced problems (4.2) and (4.9), respectively, are positive definite. □

Example 5.2. Discretization of the Fredholm integral equation of the first kind

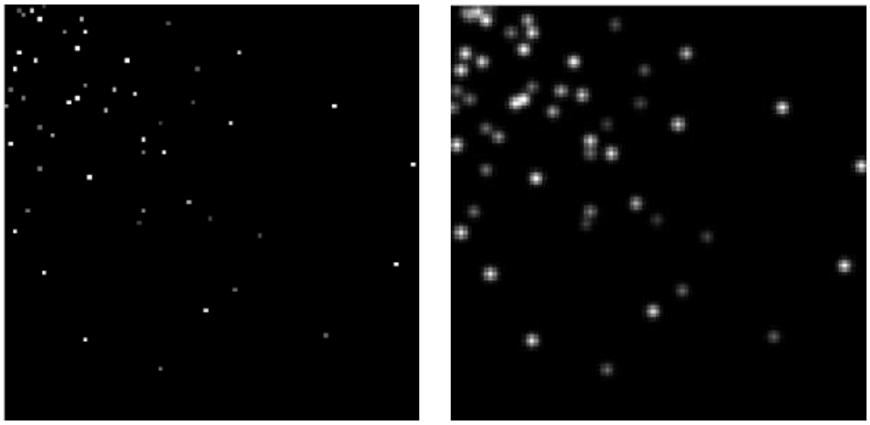
$$\int_0^\pi \exp(s\cos(t))x(s)ds = 2\frac{\sinh(t)}{t}, \qquad 0 \le t \le \frac{\pi}{2}, \tag{5.4}$$

by the Matlab function baart from [20] yields the matrix $A \in \mathbb{R}^{200\times 200}$ and a scaled discretization $\hat{x} \in \mathbb{R}^{200}$ of the solution $x(s) = \sin(s)$ of (5.4). The symmetric positive semidefinite matrix of the linear system (1.1) is given by $H = (AA^T)^{1/2}$. Define $\hat{g} = H\hat{x}$. The right-hand side g of (1.1) is computed analogously as in Example 5.1.

Table 5.4 is analogous to Table 5.1 and reports results obtained by Methods 1 and 2 for two noise levels with $s = 0.9$ and $\eta = 1.1$. Table 5.4 shows Method 2 to yield approximate solutions x_ℓ of higher accuracy than Method 1, similarly as in Example 5.1. The approximate solutions determined by Methods 1 and 2 can be seen to be better approximations of $\hat{x}$ than the vectors (1.6) with $\beta = \beta_\ell$. □

Table 5.4: Example 5.2: Discretization of the integral equation (5.4).

γ	method	ℓ	β	$\|x_\ell - \hat{x}\|/\|\hat{x}\|$	$\|x_{\beta_\ell} - \hat{x}\|/\|\hat{x}\|$
$1 \cdot 10^{-2}$	1	3	$3.82 \cdot 10^{1}$	$3.57 \cdot 10^{-1}$	$6.56 \cdot 10^{-1}$
	2	4	$4.46 \cdot 10^{1}$	$1.45 \cdot 10^{-1}$	$7.55 \cdot 10^{-1}$
$1 \cdot 10^{-3}$	1	3	$2.93 \cdot 10^{2}$	$6.15 \cdot 10^{-2}$	$4.89 \cdot 10^{-1}$
	2	4	$3.37 \cdot 10^{2}$	$4.75 \cdot 10^{-2}$	$5.62 \cdot 10^{-1}$



(a) (b)

Figure 5.1: Example 5.3: (a) Image represented by $\hat{x}$, (b) Available blurred and noisy image represented by g.

Example 5.3. We illustrate the performance of Methods 1 and 2 when applied to the solution of a linear discrete ill-posed problem with a non-smooth solution. Specifically, we are concerned with the restoration of a discrete image that has been contaminated by blur and noise. The "original" noise- and blur-free image is displayed by Figure 5.1(a); the figure shows a "star cluster" and is assumed to be unavailable. The image is represented by an array of 100×100 floating point numbers (pixels) in the interval $[2 \cdot 10^{-11}, 500]$. The pixel values, ordered row-wise, make up the vector $\hat{x} \in \mathbb{R}^n$ with $n = 100^2$. The Matlab function blur from [20] with parameter $band = 6$ and $sigma = 1$ is applied to generate a blurring operator $H \in \mathbb{R}^{n \times n}$ that models spatially invariant Gaussian blur with variance 1. The vector $\hat{g} = H\hat{x}$ represents a blurred version of the original image $\hat{x}$. Normally distributed "noise" with noise level $\gamma = 1 \cdot 10^{-4}$ is added to $\hat{g}$ to obtain the right-hand side g of (1.1), cf. (1.2). Figure 5.1(b) displays the blurred and noisy image represented by g.

Table 5.5 displays the performance of Methods 1 and 2 for $\eta = 1.001$ and $s = 0.9$. Method 1 can be seen both to yield a better approximation of $\hat{x}$ and to require fewer iterations than Method 2. Figure 5.2 shows the "restored image" determined by Method 1. $\square$

Table 5.5: Example 5.3: Star cluster.

Method	ℓ	β	$\|x_\ell - \hat{x}\|/\|\hat{x}\|$
1	7	$1.95 \cdot 10^1$	$6.26 \cdot 10^{-1}$
2	13	$1.96 \cdot 10^1$	$6.56 \cdot 10^{-1}$

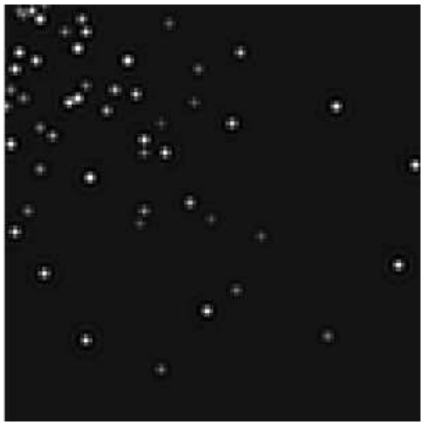


Figure 5.2: Example 5.3: Image represented by the approximation x_7 of $\hat{x}$ determined by Method 1.

6 Conclusion and discussion

The numerical examples show Methods 1 and 2 to determine approximate solutions of (1.1) with fairly few iterations. The dominating work for large-scale problems is the evaluation of matrix-vector products with the matrix H, and each iteration only requires the evaluation of one matrix-vector product. The computational effort is smaller than for standard Tikhonov regularization as described in [6, 17], and the quality of the computed approximation of $\hat{x}$ is better for some problems and worse for others.

REFERENCES

1. A. Bakushinsky and A. Goncharsky, *Ill-Posed Problems: Theory and Applications*, Kluwer, Dordrecht, 1994.
2. D. Calvetti, P. C. Hansen, and L. Reichel, *L-curve curvature bounds via Lanczos bidiagonalization*, Electron. Trans. Numer. Anal., 14 (2002), pp. 20–35.
3. D. Calvetti, B. Lewis, L. Reichel, and F. Sgallari, *Tikhonov regularization with nonnegativity constraint*, Electron. Trans. Numer. Anal., 18 (2004), pp. 153–173.
4. 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.

5. D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari, *An iterative method with error estimators*, J. Comput. Appl. Math., 127 (2001), pp. 93–119.
6. D. Calvetti and L. Reichel, *Tikhonov regularization of large linear problems*, BIT, 43 (2003), pp. 263–283.
7. D. Calvetti and L. Reichel, *Tikhonov regularization with a solution constraint*, SIAM J. Sci. Comput., 26 (2004), pp. 224–239.
8. D. Calvetti, L. Reichel, and D. C. Sorensen, *An implicitly restarted Lanczos method for large symmetric eigenvalue problems*, Electron. Trans. Numer. Anal., 2 (1994), pp. 1–21.
9. G. Dahlquist, S. C. Eisenstat, and G. H. Golub, *Bounds for the error of linear systems of equations using the theory of moments*, J. Math. Anal. Appl., 37 (1972), pp. 151–166.
10. G. Dahlquist, G. H. Golub, and S. G. Nash, *Bounds for the error in linear systems*, in Semi-Infinite Programming, ed. R. Hettich, Lecture Notes in Control and Computer Science # 15, Springer, Berlin, 1979, pp. 154–172.
11. W. Gautschi, *Orthogonal Polynomials: Computation and Approximation*, Oxford University Press, Oxford, 2004.
12. F. Girosi, M. Jones, and T. Poggio, Regularization theory and neural network architecture, Neural Computing, 7 (1995), pp. 219–269.
13. G. H. Golub and G. Meurant, *Matrices, moments and quadrature*, in Numerical Analysis 1993, eds. D. F. Griffiths and G. A. Watson, Longman, Essex, 1994, pp. 105–156.
14. G. H. Golub and G. Meurant, *Matrices, moments and quadrature II: How to compute the norm of the error in iterative methods*, BIT, 37 (1997), pp. 687–705.
15. G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, 1996.
16. G. H. Golub and U. von Matt, *Quadratically constrained least squares and quadratic problems*, Numer. Math., 59 (1991), pp. 561–580.
17. G. H. Golub and U. von Matt, *Tikhonov regularization for large scale problems*, in Workshop on Scientific Computing, eds. G. H. Golub, S. H. Lui, F. Luk, and R. Plemmons, Springer, New York, 1997, pp. 3–26.
18. C. W. Groetsch and J. Guacamene, *Arcangeli's method for Fredholm equations of the first kind*, Proc. Amer. Math. Soc., 99 (1987), pp. 256–260.
19. M. Hanke, *A note on Tikhonov regularization of large linear problems*, BIT, 43 (2003), pp. 449-451.
20. 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 `http://www.netlib.org`.
21. Z. P. Liang and P. C. Lauterbur, An efficient method for dynamic magnetic resonance imaging, IEEE Trans. Med. Imag., 13 (1994), pp. 677–686.
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. D. C. Sorensen, *Numerical methods for large eigenvalue problems*, Acta Numerica, 11 (2002), pp. 519–584.

24. G. Szegő, *Orthogonal Polynomials*, 4th ed., Amer. Math. Soc., Providence, 1975.