
Minimization of functionals on the solution of a large-scale discrete ill-posed problem*

David R. Martin · Lothar Reichel

Received: date / Accepted: date

Abstract In this work we study the minimization of a linear functional defined on a set of approximate solutions of a discrete ill-posed problem. The primary application of interest is the computation of confidence intervals for components of the solution of such a problem. We exploit the technique introduced by Eldén in 1990, utilizing a parametric programming reformulation involving the solution of a sequence of quadratically constrained least squares problems. Our iterative method, which uses the connection between Lanczos bidiagonalization and Gauss-type quadrature rules to bound certain matrix functionals, is well-suited for large-scale problems, and offers a significant reduction in matrix-vector product evaluations relative to available methods.

Keywords ill-posed problem · trust region · quadrature · confidence intervals

Mathematics Subject Classification (2010) 65D32 · 65F10 · 65F22 ·
65R30

1 Introduction

We are concerned with the solution of least-squares problems

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|, \quad (1.1)$$

* Research supported in part by NSF grant DMS-1115385.

David R. Martin
Department of Mathematical Sciences
Kent State University, Kent, OH 44242, USA
E-mail: dmarti49@kent.edu

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

where $A \in \mathbb{R}^{m \times n}$ is a highly ill-conditioned matrix with $m \geq n$. The matrix is assumed to have many singular values of different orders of magnitude close to the origin. Here $\|\cdot\|$ denotes the Euclidean vector norm. Problems of this kind are referred to as linear discrete ill-posed problems. We assume the available data vector $\mathbf{b} \in \mathbb{R}^m$ in (1.1) is contaminated by an unknown error \mathbf{e} that can be modeled as white Gaussian noise. Thus,

$$\mathbf{b} = \mathbf{b}_{\text{true}} + \mathbf{e},$$

where \mathbf{b}_{true} represents the unavailable error-free vector associated with \mathbf{b} . The linear system of equations

$$A\mathbf{x} = \mathbf{b}_{\text{true}}$$

with the (unknown) error-free right-hand side is assumed to be consistent. One would like to determine its solution \mathbf{x}_{true} of minimal Euclidean norm by computing an approximate solution of the available least-squares problem (1.1). However, due to the clustering of the singular values of A at the origin, the solution of (1.1) is very sensitive to the error \mathbf{e} in \mathbf{b} . Let A^\dagger denote the Moore-Penrose pseudoinverse of A . Then the solution

$$A^\dagger \mathbf{b} = A^\dagger \mathbf{b}_{\text{true}} + A^\dagger \mathbf{e} = \mathbf{x}_{\text{true}} + A^\dagger \mathbf{e}$$

of (1.1) generally is dominated by the propagated noise $A^\dagger \mathbf{e}$ and is of very large norm. Therefore, straightforward solution of (1.1) does not yield a meaningful approximation of \mathbf{x}_{true} . Regularization methods replace the original problem (1.1) with a “nearby” problem that is less sensitive to perturbations in the data \mathbf{b} in an attempt to compute a meaningful solution in spite of the noisy data. In this paper, we explore an alternative to standard regularization methods, namely the computation of confidence intervals for components of \mathbf{x}_{true} .

This paper is concerned with the minimization of a linear functional defined on a set of approximate solutions of a discrete ill-posed problem. We consider the problem

$$\min_{\mathbf{x}} \mathbf{w}^T \mathbf{x} \quad \text{subject to} \quad \|A\mathbf{x} - \mathbf{b}\| \leq \varepsilon, \quad \|\mathbf{x} - \mathbf{d}\| \leq \delta, \quad (1.2)$$

where \mathbf{w} is an arbitrary unit vector, and ε and δ are two positive constants. The vector \mathbf{d} may be an a priori estimate of the solution; a simple change of variables allows us to assume that $\mathbf{d} = \mathbf{0}$.

In the primary application of interest, the computation of confidence intervals for components of the solution of an ill-posed problem, the functional is chosen to be a coordinate of \mathbf{x} by letting $\mathbf{w} = \mathbf{e}_i$, the i th vector of the standard basis of \mathbb{R}^n .

The first constraint in (1.2) ensures that the least squares problem (1.1) is approximately satisfied, while the second constraint prevents the amplified noise from dominating the solution. The constants ε and δ are a priori estimates of $\|\mathbf{e}\|$ and $\|\mathbf{x}_{\text{true}} - \mathbf{d}\|$, respectively.

We utilize a transformation, also applied by Eldén [6], that replaces the minimization problem (1.2) by a nonlinear equation, for which each evaluation

requires the solution of a quadratically constrained least squares problem. A similar approach was used earlier in [15] for a problem similar in form to (1.2) but with a nonnegativity constraint on \mathbf{x} . Indeed, according to Eldén, the minimum value $\theta = \mathbf{w}^T \mathbf{x}$ attained in (1.2) is the smallest solution of the nonlinear equation

$$L(\theta) = \varepsilon^2 \quad (1.3)$$

in the interval $[\mathbf{w}^T \mathbf{d} - \delta, \mathbf{w}^T \mathbf{d} + \delta]$, where the function $L(\theta)$ is defined by

$$L(\theta) := \min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|^2 \quad \text{subject to} \quad \|\mathbf{x} - \mathbf{d}\| = \delta, \quad \mathbf{w}^T \mathbf{x} = \theta. \quad (1.4)$$

This result assumes that both constraints in (1.2) are “active” at the solution, i.e., that equality holds in both of the inequalities. Otherwise, analysis of the problem is easier; see Section 4 for more details. Equation (1.4) is solved by iterating in θ with a robust zero-finder, after applying the following simplifications. Making the substitution $\mathbf{x}' = \mathbf{x} - \mathbf{d}$ yields the equivalent problem

$$L(\theta) = \min_{\mathbf{x}'} \|A\mathbf{x}' - \mathbf{b}'\|^2 \quad \text{subject to} \quad \|\mathbf{x}'\| = \delta, \quad \mathbf{w}^T \mathbf{x}' = \theta',$$

where $\mathbf{b}' = \mathbf{b} - A\mathbf{d}$ and $\theta' = \theta - \mathbf{w}^T \mathbf{d}$. Following [6], we next eliminate the linear constraint with the substitution $\mathbf{x}' = \theta' \mathbf{w} + H\mathbf{y}$, where $Q = [\mathbf{w}, H]$ is an orthogonal matrix, which can be chosen to be a Householder reflector. This yields the equivalent problem

$$L(\theta) = \min_{\mathbf{y}} \|\bar{A}\mathbf{y} - \bar{\mathbf{b}}\|^2 \quad \text{subject to} \quad \|\mathbf{y}\| = \bar{\delta}, \quad (1.5)$$

where

$$\bar{A} := AH, \quad (1.6)$$

$$\bar{\mathbf{b}} := \mathbf{b}' - \theta' A\mathbf{w}, \quad (1.7)$$

$$\bar{\delta}^2 := \delta^2 - (\theta')^2.$$

Eldén [6] relied on a complete factorization of A , allowing very efficient solution of the constrained least squares problems (1.5) for computing $L(\theta)$. We are interested in the case when A is so large that a complete factorization is not attractive.

Eldén et al. [7] proposed to solve the quadratically constrained least squares problems by the LSTRS method [19, 20]. This method is designed to solve large-scale quadratic minimization problems arising in the context of trust-region methods for optimization by recasting them as parameterized eigenvalue problems, whose solutions are computed by the implicitly restarted Arnoldi method. However, this approach requires a large number of matrix-vector product evaluations with the matrix A for each evaluation of $L(\theta)$.

The current paper seeks to keep the number of matrix-vector product evaluations small by using the connection between partial Lanczos bidiagonalization and Gauss-type quadrature rules. Other applications of this connection to the solution of discrete ill-posed problems are described in [1–4, 11–13]; see also [9, 10].

As mentioned above, when A is ill-conditioned and \mathbf{b} is corrupted by error, the unregularized least squares solution $\mathbf{y}_0 = \bar{A}^\dagger \bar{\mathbf{b}}$ associated with (1.5) is generally very large in norm, much larger than $\bar{\delta}$. Let

$$\varphi(\mu) := \|\mathbf{y}_\mu\|^2, \quad (1.8)$$

where

$$\mathbf{y}_\mu := (\bar{A}^T \bar{A} + \mu I)^{-1} \bar{A}^T \bar{\mathbf{b}} \quad (1.9)$$

and I denotes the identity matrix. Then, \mathbf{y}_μ minimizes $\mathbf{y} \mapsto \|\bar{A}\mathbf{y} - \bar{\mathbf{b}}\|^2 + \mu \|\mathbf{y}\|^2$. We will show in Section 2 that for any choice of τ with $0 < \tau < \|\mathbf{y}_0\|^2$, there exists $0 < \mu < \infty$ such that $\varphi(\mu) = \tau$. It is therefore not hard to see that the solution to (1.5) is the Tikhonov solution (1.9) of norm $\bar{\delta}$. Thus, for each θ we first iterate in μ to determine the solution μ_θ of

$$\varphi(\mu) = \bar{\delta}^2 \quad (1.10)$$

and then compute $L(\theta) = \psi(\mu_\theta)$, where

$$\psi(\mu) := \|\bar{A}\mathbf{y}_\mu - \bar{\mathbf{b}}\|^2. \quad (1.11)$$

To make this approach feasible, upper and lower bounds for both $\varphi(\mu)$ and $\psi(\mu)$ can be evaluated cheaply by using Gauss-type quadrature rules and partial Lanczos bidiagonalization as shown in [4, 9–11]. Therefore, all of the work when iterating to solve (1.10) during the computation of $L(\theta)$ can be done within Krylov subspaces of low dimension. We also show that with the proper decomposition of the functions φ and ψ , we need not compute new bidiagonal factors when iterating in θ , even though the right-hand side in (1.5) depends on θ . Therefore, matrix-vector products with A need not be repeated for different values of θ , and the dimension of the bidiagonal factors is increased only in the event that the current quadrature-related bounds for the matrix functionals cannot guarantee solution of (1.10) within specified tolerance.

This paper is organized as follows. In Section 2 we review basic results regarding the relationship between partial Lanczos bidiagonalization, Gauss-type quadrature rules, and matrix functionals. Section 3 describes an algorithm for solving the quadratically constrained least squares problems (1.5). In Section 4 we provide details regarding the iteration in θ to solve (1.3), including an estimate for the derivative $L'(\theta)$ using already computed quantities, affording a significant reduction by means of Newton's method of the number of iterates θ_i required to solve (1.3). In Section 5 some computed examples are presented, and Section 6 contains concluding remarks.

2 Matrix functionals, Gauss-quadrature, and partial Lanczos bidiagonalization

This section reviews results on the connection between partial Lanczos bidiagonalization, Gauss-type quadrature rules, and the computation of upper and

lower bounds for the functions (1.8) and (1.11). Related results can be found in [1–3, 9–12].

Using the representation (1.9), the functions (1.8) and (1.11) can be expressed as

$$\begin{aligned}\varphi(\mu) &= (\bar{A}^T \bar{\mathbf{b}})^T (\bar{A}^T \bar{A} + \mu I)^{-2} (\bar{A}^T \bar{\mathbf{b}}), \\ \psi(\mu) &= \mu^2 \bar{\mathbf{b}}^T (\bar{A} \bar{A}^T + \mu I)^{-2} \bar{\mathbf{b}},\end{aligned}$$

for $\mu > 0$; the expression for $\psi(\mu)$ can be derived by inserting the singular value decomposition (SVD) of \bar{A} into (1.11). Letting

$$f_\mu(t) := (t + \mu)^{-2}, \quad (2.1)$$

we have

$$\varphi(\mu) = (\bar{A}^T \bar{\mathbf{b}})^T f_\mu(\bar{A}^T \bar{A}) (\bar{A}^T \bar{\mathbf{b}}), \quad \psi(\mu) = \mu^2 \bar{\mathbf{b}}^T f_\mu(\bar{A} \bar{A}^T) \bar{\mathbf{b}}. \quad (2.2)$$

Inserting the SVD of \bar{A} , it is not hard to show that $\varphi(\mu)$ is strictly decreasing and convex for $\mu > 0$. It follows that the equation

$$\varphi(\mu) = \tau$$

has a unique solution μ with $0 < \mu < \infty$, for any τ with $0 < \tau < \|\bar{A}^\dagger \bar{\mathbf{b}}\|^2$. In particular, equation (1.10) has a unique solution under the reasonable assumption that $\bar{\delta} < \|\bar{A}^\dagger \bar{\mathbf{b}}\|$. Therefore, for each value of θ we can compute $L(\theta)$ by first solving (1.10) for μ_θ and then evaluating $\psi(\mu_\theta)$.

The vectors $\bar{\mathbf{b}}$, but not the matrices \bar{A} , in the representations (2.2) depend on θ ; cf. (1.6) and (1.7). The evaluation of upper and lower bounds for the functions (2.2) is simplified by writing these functions as linear combinations of functions that are independent of θ . We propose the following decompositions:

$$\begin{aligned}\varphi(\mu) &= (\mathbf{b}' - \theta' A \mathbf{w})^T \bar{A} f_\mu(\bar{A}^T \bar{A}) \bar{A}^T (\mathbf{b}' - \theta' A \mathbf{w}) \\ &= [1 + \theta'] \varphi_1(\mu) + [(\theta')^2 + \theta'] \varphi_2(\mu) - \theta' \varphi_3(\mu),\end{aligned} \quad (2.3)$$

where

$$\begin{aligned}\varphi_1(\mu) &:= (\bar{A}^T \mathbf{b}')^T f_\mu(\bar{A}^T \bar{A}) (\bar{A}^T \mathbf{b}'), \\ \varphi_2(\mu) &:= (\bar{A}^T A \mathbf{w})^T f_\mu(\bar{A}^T \bar{A}) (\bar{A}^T A \mathbf{w}), \\ \varphi_3(\mu) &:= [\bar{A}^T (\mathbf{b}' + A \mathbf{w})]^T f_\mu(\bar{A}^T \bar{A}) [\bar{A}^T (\mathbf{b}' + A \mathbf{w})],\end{aligned} \quad (2.4)$$

and

$$\begin{aligned}\psi(\mu) &= \mu^2 (\mathbf{b}' - \theta' A \mathbf{w})^T f_\mu(\bar{A} \bar{A}^T) (\mathbf{b}' - \theta' A \mathbf{w}) \\ &= [1 + \theta'] \psi_1(\mu) + [(\theta')^2 + \theta'] \psi_2(\mu) - \theta' \psi_3(\mu),\end{aligned} \quad (2.5)$$

where, analogously,

$$\begin{aligned}\psi_1(\mu) &:= \mu^2 (\mathbf{b}')^T f_\mu(\bar{A} \bar{A}^T) (\mathbf{b}'), \\ \psi_2(\mu) &:= \mu^2 (A \mathbf{w})^T f_\mu(\bar{A} \bar{A}^T) (A \mathbf{w}), \\ \psi_3(\mu) &:= \mu^2 (\mathbf{b}' + A \mathbf{w})^T f_\mu(\bar{A} \bar{A}^T) (\mathbf{b}' + A \mathbf{w}).\end{aligned} \quad (2.6)$$

The functions in (2.4) and (2.6) are independent of θ . We now describe how to inexpensively compute upper and lower bounds for these functions. Thus, we are interested in the computation of upper and lower bounds for expressions of the form

$$\xi = (M^T \mathbf{h})^T f(M^T M)(M^T \mathbf{h}), \quad \zeta = \mathbf{h}^T f(M M^T) \mathbf{h}. \quad (2.7)$$

We will achieve this by using the connection between orthogonal polynomials, Gauss quadrature, and partial Lanczos bidiagonalization of M . Here $M \in \mathbb{R}^{m \times n}$ is a large matrix, $m \geq n$, $\mathbf{h} \in \mathbb{R}^m$, and f is a function defined on the spectrum of MM^T (and thus also on the spectrum of $M^T M$). Details about this approach can be found in [1, 4, 9, 10].

First note that the quantities ξ and ζ in (2.7) can be represented as Stieltjes integrals. To see this, introduce the spectral factorization

$$M^T M = W \Lambda W^T,$$

where $\Lambda = \text{diag}[\lambda_1, \dots, \lambda_n]$ and the matrix $W \in \mathbb{R}^{n \times n}$ is orthogonal. Defining $\mathbf{k} = [k_1, \dots, k_n] := W^T M^T \mathbf{h}$ yields the representation

$$\xi = \mathbf{k}^T f(\Lambda) \mathbf{k} = \sum_{i=1}^n f(\lambda_i) k_i^2 = \int_0^\infty f(t) d\omega(t), \quad (2.8)$$

where the distribution ω can be chosen to be a nondecreasing step function with a jump of size k_i^2 at the eigenvalue λ_i of $M^T M$. Note that ω is determined by M and \mathbf{h} , but is independent of f .

Introduce the inner product for polynomials of sufficiently low degree,

$$\langle f, g \rangle := \int_0^\infty f(t) g(t) d\omega(t) = \sum_{i=1}^n f(\lambda_i) g(\lambda_i) k_i^2 = \mathbf{k}^T f(\Lambda) g(\Lambda) \mathbf{k},$$

defined by the measure $d\omega(t)$. Let $\{q_i\}_{i=0}^{n-1}$ be a family of orthonormal polynomials with positive leading coefficient with respect to this inner product. Thus, $\langle q_i, q_j \rangle = \delta_{ij}$, where δ_{ij} is the Kronecker delta function. The q_i satisfy a three-term recurrence relation of the form

$$t q_{i-1}(t) = \beta_i q_i(t) + \alpha_i q_{i-1}(t) + \beta_{i-1} q_{i-2}(t), \quad i = 1, 2, \dots, \quad (2.9)$$

where $q_{-1}(t) := 0$ and $q_0(t) := \langle 1, 1 \rangle^{-1/2}$. Application of ℓ steps of the standard Lanczos algorithm to the matrix $M^T M$ with initial vector $M^T \mathbf{h}$ yields the symmetric positive definite or semidefinite tridiagonal matrix

$$T_\ell := \begin{bmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \beta_{\ell-2} & \alpha_{\ell-1} & \beta_{\ell-1} \\ & & & \beta_{\ell-1} & \alpha_\ell \end{bmatrix}, \quad (2.10)$$

whose entries are the first $2\ell - 1$ coefficients in the recurrence relation (2.9). The orthonormal Lanczos vectors determined by the Lanczos algorithm can be expressed as $\mathbf{v}_i = q_i(M^T M)M^T \mathbf{h}$ and the expression

$$\mathcal{G}_\ell(f) = \|M^T \mathbf{h}\|^2 \mathbf{e}_1^T f(T_\ell) \mathbf{e}_1 \quad (2.11)$$

is the ℓ -point Gauss quadrature rule associated with the measure $d\omega(t)$; see [1, 10] for details. Thus,

$$\mathcal{G}_\ell(p) = \int_0^\infty p(t) d\omega(t) \quad \forall p \in \mathbb{P}_{2\ell-1},$$

where \mathbb{P}_j denotes the set of all polynomials of degree at most j .

Introduce the Cholesky factor

$$C_\ell := \begin{bmatrix} \rho_1 & & & \\ \sigma_2 & \rho_2 & & \\ & \ddots & \ddots & \\ & & \sigma_{\ell-1} & \rho_{\ell-1} \\ & & & \sigma_\ell & \rho_\ell \end{bmatrix}$$

of T_ℓ , as well as the matrix

$$C_{\ell,\ell-1} := \begin{bmatrix} C_{\ell-1} \\ \sigma_\ell \mathbf{e}_{\ell-1}^T \end{bmatrix} \in \mathbb{R}^{\ell \times (\ell-1)},$$

which consists of the the first $\ell - 1$ columns of C_ℓ . When $f = f_\mu$ is defined by (2.1), the quadrature rule (2.11) can be evaluated efficiently when the Cholesky factor C_ℓ of T_ℓ is available. This Cholesky factor can be computed directly, without forming T_ℓ ; see below.

Let $\mathcal{R}_\ell(f)$ denote the ℓ -point Gauss-Radau quadrature rule associated with the measure $d\omega(t)$ with a prescribed node at the origin. Then

$$\mathcal{R}_\ell(p) = \int_0^\infty p(t) d\omega(t) \quad \forall p \in \mathbb{P}_{2\ell-2}.$$

It is convenient to evaluate this rule by using the representation

$$\mathcal{R}_\ell(f) = \|M^T \mathbf{h}\|^2 \mathbf{e}_1^T f(C_{\ell,\ell-1} C_{\ell,\ell-1}^T) \mathbf{e}_1; \quad (2.12)$$

see, e.g., [1, 9, 10, 13].

Assume that f has 2ℓ continuous derivatives on the convex hull of the support of $d\omega$, and consider the quadrature error

$$\mathcal{E}_{\mathcal{Q}_\ell}(f) := \int_0^\infty f(t) d\omega(t) - \mathcal{Q}_\ell(f),$$

where $\mathcal{Q}_\ell = \mathcal{G}_\ell$ or $\mathcal{Q}_\ell = \mathcal{R}_\ell$. In the former case, there exists $\theta_{\mathcal{G}_\ell} \in [\lambda_1, \lambda_n]$ such that

$$\mathcal{E}_{\mathcal{G}_\ell}(f) = \frac{f^{(2\ell)}(\theta_{\mathcal{G}_\ell})}{(2\ell)!} \int_0^\infty \prod_{i=1}^{\ell} (t - \theta_i)^2 d\omega(t), \quad (2.13)$$

where $f^{(j)}$ denotes the j th derivative of the function f and the θ_i are the Gaussian nodes, i.e., the eigenvalues of the matrix T_ℓ . For $\mathcal{Q}_\ell = \mathcal{R}_\ell$, there exists $\theta_{\mathcal{R}_\ell} \in [0, \lambda_n]$ such that

$$\mathcal{E}_{\mathcal{R}_\ell}(f) = \frac{f^{(2\ell-1)}(\theta_{\mathcal{R}_\ell})}{(2\ell-1)!} \int_0^\infty t \prod_{i=2}^{\ell} (t - \tilde{\theta}_i)^2 d\omega(t), \quad (2.14)$$

where the $\tilde{\theta}_i$ are the nodes of the Gauss-Radau rule with $\tilde{\theta}_1 = 0$. We refer to Golub and Meurant [9, 10] or Gautschi [8] for details. The error formulas (2.13) and (2.14) yield the following result.

Proposition 2.1 *Let f_μ be defined by (2.1) and assume that $\mu > 0$. Then, for $\ell \geq 1$,*

$$\mathcal{E}_{\mathcal{G}_\ell}(f_\mu) > 0, \quad \mathcal{E}_{\mathcal{R}_\ell}(f_\mu) < 0. \quad (2.15)$$

Proof We have $f_\mu^{(2\ell-1)}(t) < 0$ and $f_\mu^{(2\ell)}(t) > 0$ for $t \geq 0$. Substituting these inequalities into (2.13) and (2.14) yields the desired bounds. \square

It follows from the proposition that

$$\mathcal{G}_\ell(f_\mu) < \int_0^\infty f_\mu(t) d\omega(t) < \mathcal{R}_\ell(f_\mu).$$

Setting $M = \bar{A}$ and $\mathbf{h} = \mathbf{b}'$, we obtain the following lower and upper bounds for the function φ_1 in (2.4):

$$\begin{aligned} \varphi_{1,\ell}^-(\mu) &:= \|\bar{A}^T \mathbf{b}'\|^2 \mathbf{e}_1^T f_\mu(T_\ell) \mathbf{e}_1 = \|\bar{A}^T \mathbf{b}'\|^2 \mathbf{e}_1^T f_\mu(C_\ell C_\ell^T) \mathbf{e}_1, \\ \varphi_{1,\ell}^+(\mu) &:= \|\bar{A}^T \mathbf{b}'\|^2 \mathbf{e}_1^T f_\mu(C_{\ell,\ell-1} C_{\ell,\ell-1}^T) \mathbf{e}_1. \end{aligned}$$

Similar bounds, denoted by $\varphi_{2,\ell}^\pm(\mu)$ and $\varphi_{3,\ell}^\pm(\mu)$, for $\varphi_2(\mu)$ and $\varphi_3(\mu)$, respectively, can be derived, and these bounds together yield upper and lower bounds for $\varphi(\mu)$; see below for details.

Applying virtually the same argument to the quantity ζ as defined in (2.7), analogous bounds can be obtained. In the same manner as before, a Stieltjes integral representation similar to (2.8) can be obtained, and the associated ℓ -point Gauss quadrature rule is given by

$$\hat{\mathcal{G}}_\ell(f) = \|\mathbf{h}\|^2 \mathbf{e}_1^T f(\hat{T}_\ell) \mathbf{e}_1, \quad (2.16)$$

where \hat{T}_ℓ is the symmetric tridiagonal matrix analogous to (2.10) obtained by applying ℓ steps of the standard Lanczos algorithm to the matrix MM^T with initial vector \mathbf{h} . The associated ℓ -point Gauss-Radau quadrature rule with one node assigned at the origin, similar to (2.12), can be evaluated as

$$\hat{\mathcal{R}}_\ell(f) = \|\mathbf{h}\|^2 \mathbf{e}_1^T f(\hat{C}_{\ell,\ell-1} \hat{C}_{\ell,\ell-1}^T) \mathbf{e}_1, \quad (2.17)$$

where the matrix $\hat{C}_{\ell,\ell-1}$ consists of the first $\ell - 1$ columns of the Cholesky factor \hat{C}_ℓ of \hat{T}_ℓ . We obtain analogously to (2.15) that

$$\mathcal{E}_{\hat{\mathcal{G}}_\ell}(f_\mu) > 0, \quad \mathcal{E}_{\hat{\mathcal{R}}_\ell}(f_\mu) < 0.$$

Thus, when $f = f_\mu$, lower and upper bounds for the quantity ζ defined by (2.7) are determined by the Gauss (2.16) and Gauss-Radau (2.17) quadrature rules, respectively. Again setting $M = \bar{A}$ and $\mathbf{h} = \mathbf{b}'$, we obtain the following lower and upper bounds for the function ψ_1 in (2.4):

$$\begin{aligned}\varphi_{1,\ell}^-(\mu) &:= \mu^2 \|\mathbf{b}'\|^2 \mathbf{e}_1^T f_\mu(\hat{T}_\ell) \mathbf{e}_1 = \mu^2 \|\mathbf{b}'\|^2 \mathbf{e}_1^T f_\mu(\hat{C}_\ell \hat{C}_\ell^T) \mathbf{e}_1, \\ \psi_{1,\ell}^+(\mu) &:= \mu^2 \|\mathbf{b}'\|^2 \mathbf{e}_1^T f_\mu(\hat{C}_{\ell,\ell-1} \hat{C}_{\ell,\ell-1}^T) \mathbf{e}_1.\end{aligned}$$

Similar bounds can be computed for $\psi_2(\mu)$ and $\psi_3(\mu)$, and these bounds together are used to produce bounds for $\psi(\mu)$.

We require one more fact regarding the quadrature approximations $\varphi_{i,\ell}^\pm(\mu)$ and $\psi_{i,\ell}^\pm(\mu)$, namely that these bounds become monotonically tighter as ℓ increases. This can be seen by viewing the Gauss rule $\mathcal{G}_{\ell+1}$ as an integral with respect to a discrete measure and observing that the ℓ -point rule \mathcal{G}_ℓ also is a Gauss quadrature rule with respect this measure. Therefore, in the same manner as above, we have the inequalities

$$\varphi_{i,\ell}^-(\mu) < \varphi_{i,\ell+1}^-(\mu), \quad \psi_{i,\ell}^-(\mu) < \psi_{i,\ell+1}^-(\mu). \quad (2.18)$$

A similar argument shows that

$$\varphi_{i,\ell+1}^+(\mu) < \varphi_{i,\ell}^+(\mu), \quad \psi_{i,\ell+1}^+(\mu) < \psi_{i,\ell}^+(\mu). \quad (2.19)$$

Returning to the original functions decomposed in (2.3) and (2.5), let $\varphi_\ell^\pm(\mu)$ and $\psi_\ell^\pm(\mu)$ be the derived bounds for $\varphi(\mu)$ and $\psi(\mu)$ obtained in the obvious manner from the bounds $\varphi_{i,\ell}^\pm(\mu)$ and $\psi_{i,\ell}^\pm(\mu)$. For example, when $\theta' > 0$ we have

$$\varphi_\ell^+(\mu) := [1 + \theta'] \varphi_{1,\ell}^+(\mu) + [(\theta')^2 + \theta'] \varphi_{2,\ell}^+(\mu) - \theta' \varphi_{3,\ell}^-(\mu); \quad (2.20)$$

the formulas for $\varphi_\ell^-(\mu)$ and $\psi_\ell^\pm(\mu)$ are analogous. The details change slightly depending on the signs of θ' and $1 + \theta'$. With these definitions, we have

$$\varphi_\ell^-(\mu) < \varphi(\mu) < \varphi_\ell^+(\mu), \quad \psi_\ell^-(\mu) < \psi(\mu) < \psi_\ell^+(\mu) \quad (2.21)$$

for each ℓ . Moreover, the inequalities in (2.18) and (2.19) together imply that

$$\varphi_\ell^-(\mu) < \varphi_{\ell+1}^-(\mu), \quad \psi_\ell^-(\mu) < \psi_{\ell+1}^-(\mu), \quad (2.22)$$

and

$$\varphi_{\ell+1}^+(\mu) < \varphi_\ell^+(\mu), \quad \psi_{\ell+1}^+(\mu) < \psi_\ell^+(\mu). \quad (2.23)$$

The method in Section 3 for computing $L(\theta)$ requires the evaluation of $\varphi_{i,\ell}^\pm(\mu)$ and $\psi_{i,\ell}^\pm(\mu)$ for several values of the argument μ , as well as their derivatives $\frac{d}{d\mu} [\varphi_{i,\ell}^\pm](\mu)$ and $\frac{d}{d\mu} [\psi_{i,\ell}^\pm](\mu)$, for $i = 1, 2, 3$. We evaluate $\varphi_{1,\ell}^+(\mu)$ by first determining the solution \mathbf{z}_ℓ of the linear system of equations

$$(C_{\ell,\ell-1} C_{\ell,\ell-1}^T + \mu I_\ell) \mathbf{z} = \mathbf{e}_1 \quad (2.24)$$

and then computing

$$\varphi_{1,\ell}^+(\mu) = \|\bar{A}^T \mathbf{b}'\|^2 \mathbf{z}_\ell^T \mathbf{z}_\ell.$$

Note that (2.24) are the normal equations associated with the least squares problem

$$\min_{\mathbf{z} \in \mathbb{R}^\ell} \left\| \begin{bmatrix} C_{\ell,\ell-1}^T \\ \mu^{1/2} I_\ell \end{bmatrix} \mathbf{z} - \mu^{-1/2} \mathbf{e}_\ell \right\|,$$

and we compute \mathbf{z}_ℓ by solving the latter by a method described by Eldén [5] and Paige and Saunders [17]. This method uses orthogonal Givens transformations and requires only $\mathcal{O}(\ell)$ arithmetic floating point operations for each value of μ . All functions $\varphi_{i,\ell}^\pm(\mu)$ and $\psi_{i,\ell}^\pm(\mu)$ are evaluated analogously.

The derivative

$$\frac{d}{d\mu} \left[\varphi_{1,\ell}^+ \right] (\mu) = -2 \|\bar{A}^T \mathbf{b}'\|^2 \mathbf{e}_1^T (C_{\ell,\ell-1} C_{\ell,\ell-1}^T + \mu I_\ell)^{-3} \mathbf{e}_1$$

is computed by solving the least squares problem

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

Denote the solution by \mathbf{s}_ℓ . The value of the derivative is then given by

$$\frac{d}{d\mu} \left[\varphi_{1,\ell}^+ \right] (\mu) = -2 \|\bar{A}^T \mathbf{b}'\|^2 \mathbf{z}_\ell^T \mathbf{s}_\ell.$$

The other derivatives $\frac{d}{d\mu} \left[\varphi_{i,\ell}^\pm \right] (\mu)$ are evaluated analogously, as are

$$\frac{d}{d\mu} \left[\psi_{1,\ell}^+ \right] (\mu) = \frac{2}{\mu} \psi_{1,\ell}^+(\mu) - 2\mu^2 \|\mathbf{b}'\|^2 \mathbf{e}_1^T (\hat{C}_{\ell,\ell-1} \hat{C}_{\ell,\ell-1}^T + \mu I_\ell)^{-3} \mathbf{e}_1,$$

and the other derivatives $\frac{d}{d\mu} \left[\psi_{i,\ell}^\pm \right] (\mu)$.

We turn to the computation of the lower bidiagonal Cholesky factors C_ℓ and \hat{C}_ℓ of the tridiagonal matrices T_ℓ and \hat{T}_ℓ , respectively. They can be computed without explicitly forming the latter matrices. Application of $\ell \leq \min\{m, n\}$ steps of Lanczos bidiagonalization to the matrix M with initial vector $\mathbf{u}_1 = \mathbf{h}/\|\mathbf{h}\|$ yields the decompositions

$$MV_\ell = U_{\ell+1} \hat{C}_{\ell+1,\ell}, \quad M^T U_\ell = V_\ell \hat{C}_\ell^T, \quad \mathbf{h} = \sigma_1 U_\ell \mathbf{e}_1, \quad (2.25)$$

where the matrices $V_\ell = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_\ell] \in \mathbb{R}^{n \times \ell}$ and $U_{\ell+1} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{\ell+1}] \in \mathbb{R}^{m \times (\ell+1)}$ have orthonormal columns. The evaluation of the partial Lanczos bidiagonalization (2.25) requires ℓ matrix-vector product evaluations with each of the matrices M and M^T . We refer to Paige and Saunders for details [16].

The equations (2.25) imply that

$$MM^T U_\ell = U_\ell \hat{C}_\ell \hat{C}_\ell^T + \text{scalar} \cdot \mathbf{u}_{\ell+1} \mathbf{e}_\ell^T$$

and, therefore, \hat{C}_ℓ is the Cholesky factor of the symmetric tridiagonal matrix \hat{T}_ℓ in (2.16), and the columns of U_ℓ are the associated Lanczos vectors. The

lower bidiagonal Cholesky factor C_ℓ of T_ℓ in (2.10) can be determined, without forming T_ℓ , from the QR factorization $\hat{C}_{\ell+1,\ell} = Q_{\ell+1,\ell}R_\ell$, where $Q_{\ell+1,\ell} \in \mathbb{R}^{(\ell+1) \times \ell}$ satisfies $Q_{\ell+1,\ell}^T Q_{\ell+1,\ell} = I$ and $R_\ell \in \mathbb{R}^{\ell \times \ell}$ is upper bidiagonal. Indeed, the equations (2.25) show that

$$M^T M V_\ell = V_\ell R_\ell^T R_\ell + \text{scalar} \cdot v_{\ell+1} e_\ell^T,$$

which implies that R_ℓ^T is the lower bidiagonal Cholesky factor C_ℓ of the tridiagonal matrix T_ℓ , and the columns of V_ℓ are the associated Lanczos vectors. We note that the matrix $Q_{\ell+1,\ell}$ can be represented by ℓ Givens rotations. Therefore, R_ℓ , and thus C_ℓ , can be computed from $\hat{C}_{\ell+1,\ell}$ in only $\mathcal{O}(\ell)$ arithmetic floating point operations.

In conclusion, although we require the computation of six different lower bidiagonal matrices, the matrices used for bounding the functions φ_i defined in (2.4) can be obtained with very little work from their counterparts used for bounding ψ_i in (2.6). In particular, the Lanczos bidiagonalization algorithm has to be executed only three times (with different initial vectors).

3 Computing $L(\theta)$

We exploit the quadrature results of the previous section to derive a numerical method for the evaluation of $L(\theta)$. The primary task is the computation of an approximate solution μ_θ of equation (1.10). We would like to determine a value of μ such that

$$|\varphi(\mu) - \bar{\delta}^2| \leq \eta \bar{\delta}^2, \quad (3.1)$$

where the user-specified parameter $0 < \eta < 1$ is the maximum allowable relative error when solving (1.10). We determine such a value of μ by computing a pair $\{\ell, \mu\}$, such that

$$(1 - \eta) \bar{\delta}^2 \leq \varphi_\ell^-(\mu), \quad \varphi_\ell^+(\mu) \leq (1 + \eta) \bar{\delta}^2, \quad (3.2)$$

where as usual ℓ denotes the number of Lanczos bidiagonalization steps. It follows from the first part of (2.21) that the inequalities (3.2) imply (3.1). For many linear discrete ill-posed problems, the value of ℓ in a pair $\{\ell, \mu\}$ that satisfies (3.2) can be chosen fairly small, and therefore the effort required to compute such a pair is typically much smaller than the effort necessary to solve (1.10) to high accuracy.

We assume that A is so large that the dominating computational work when determining an approximate solution of (1.10) is the evaluation of the ℓ matrix-vector products with each of the matrices \bar{A} and \bar{A}^T required for the determination of each of three partial Lanczos bidiagonalizations of \bar{A} . Our numerical method is designed to keep the number of Lanczos bidiagonalization steps ℓ small. The method starts with $\ell = 2$ and then increases ℓ only if necessary.

For a given value of $\ell \geq 2$, we seek a sequence of approximations $\mu_\ell^{(j)}$, $j = 0, 1, 2, \dots$, of the largest zero μ_ℓ of the function

$$h_\ell^+(\mu) := \varphi_\ell^+(\mu) - (1 + (1 - \nu)\eta)\bar{\delta}^2, \quad (3.3)$$

terminating the iterations when an approximation $\mu_\ell^{(p)}$ of μ_ℓ that satisfies

$$|h_\ell^+(\mu_\ell^{(p)})| \leq \nu\eta\bar{\delta}^2 \quad (3.4)$$

has been found. Here $\varphi_\ell^+(\mu)$ is defined as in (2.20) and $0 < \nu < 1$ is a user-specified parameter. In the numerical experiments reported in Section 5, we let $\nu := 0.01$; use of a larger factor tends to reduce the number of iterations with the zero-finder (small-scale work), but may increase the number of Lanczos bidiagonalization steps ℓ (large-scale work). By assumption, the square of the norm of the unregularized least squares solution of (1.5) is much larger than $\bar{\delta}^2$. It follows that $\varphi_\ell^+(0) > \varphi(0) > (1 + (1 - \nu)\eta)\bar{\delta}^2$. Moreover, it is also clear from the representations (2.20) and (2.4) that $\varphi_\ell^+(\mu) \rightarrow 0$ as $\mu \rightarrow \infty$. Since φ_ℓ^+ is continuous, it follows that (3.3) has at least one solution.

In [4], the monotonically and quadratically convergent zero-finder described by Golub and von Matt in [11] is used to determine a sequence of approximations $\mu_\ell^{(j)}$ converging monotonically to μ_ℓ from the right. However, due to the details of the decomposition (2.20), it is not clear that the same performance can be obtained here. Newton's method works remarkably well for solving (3.3), contingent on a good initial choice of μ ; see Section 5.

We should take into consideration that the tolerance η from (3.1) for solving (1.10) may not guarantee a small relative error in the computed value $\psi(\mu)$. That is, a priori, a small relative change $\frac{\Delta\varphi}{\varphi}$ could correspond to a large relative change $\frac{\Delta\psi}{\psi}$. We therefore outline a way to (almost) ensure a small relative error in $\psi(\mu)$.

Immediately after finding an approximation $\mu_\ell^{(p)}$ of μ_ℓ satisfying (3.4), we compute an approximate upper bound for the relative change $\frac{\Delta\psi}{\psi}$ that is caused by the relative error, at most η , when solving $\varphi(\mu) = \bar{\delta}^2$. An estimate of $\frac{\Delta\psi}{\psi}$, obtained by linearization of both φ and ψ , is given by

$$\left| \frac{\Delta\psi}{\psi} \right| \approx \left| \frac{\Delta\varphi}{\frac{d\varphi}{d\mu}} \frac{\frac{d\psi}{d\mu}}{\psi} \right| = \left| \frac{\bar{\delta}^2\eta}{\frac{d\varphi}{d\mu}(\mu_\ell^{(p)})} \frac{\frac{d\psi}{d\mu}(\mu_\ell^{(p)})}{\psi(\mu_\ell^{(p)})} \right|. \quad (3.5)$$

Utilizing the available quadratures, an approximate upper bound for $\frac{\Delta\psi}{\psi}$ can be computed as

$$\left| \frac{\bar{\delta}^2\eta}{\left[\frac{d\varphi}{d\mu} \right]_\ell^+ (\mu_\ell^{(p)})} \frac{\frac{d}{d\mu} [\psi_\ell^+] (\mu_\ell^{(p)})}{\psi_\ell^- (\mu_\ell^{(p)})} \right|, \quad (3.6)$$

where $\left[\frac{d\varphi}{d\mu} \right]_\ell^+$ is a quadrature-related upper bound for $\frac{d\varphi}{d\mu}$, as described below. Strictly speaking, the expression (3.6) need not be an upper bound for

$\frac{\Delta\psi}{\psi}$, due to the approximate nature of (3.5) and because $\frac{d}{d\mu} [\psi_\ell^+] (\mu_\ell^{(p)})$ could be less than $\frac{d\psi}{d\mu} (\mu_\ell^{(p)})$. Nevertheless, the quantity (3.6) provides a useful estimate of $\frac{\Delta\psi}{\psi}$. The use of lower and upper bounds for ψ and $\frac{d\varphi}{d\mu}$, respectively, usually provides a larger (hence safer) estimate of $\frac{\Delta\psi}{\psi}$, since $\psi(\mu) \geq 0$ and $\frac{d\varphi}{d\mu} \leq 0$ for $\mu > 0$ (and their approximations usually satisfy these inequalities). Now, differentiation with respect to μ changes only the function at which the quadratures are being evaluated (not the associated measure), and therefore commutes with computing quadratures. Hence, $\frac{d}{d\mu} [\varphi_{i,\ell}]$ and $\frac{d}{d\mu} [\varphi_{i,\ell}^+]$ are the Gauss and Gauss-Radau quadratures, respectively, of $\frac{d\varphi_i}{d\mu}$. The discussion in the previous section regarding the quadrature error shows that lower and upper bounds for $\frac{d\varphi_i}{d\mu}$ are thus given by

$$\left[\frac{d\varphi_i}{d\mu} \right]_\ell^- := \frac{d}{d\mu} [\varphi_{i,\ell}^+], \quad \left[\frac{d\varphi_i}{d\mu} \right]_\ell^+ := \frac{d}{d\mu} [\varphi_{i,\ell}^-]$$

for $i = 1, 2, 3$, and these bounds are used together to compute $\left[\frac{d\varphi}{d\mu} \right]_\ell^+$ as in (2.20).

Unfortunately, the details are somewhat more complicated for $\frac{d\psi}{d\mu}$, so rather than seeking an upper bound for $\frac{d\psi}{d\mu}$, we simply insert the derivative of the upper bound with respect to μ , $\frac{d}{d\mu} [\psi_\ell^+]$.

Letting 2γ represent the maximum allowable relative error in $\psi(\mu)$, we require that (3.6) be less than γ , or equivalently that

$$\eta \leq \left| \frac{\gamma \left[\frac{d\varphi}{d\mu} \right]_\ell^+ (\mu_\ell^{(p)}) \psi_\ell^- (\mu_\ell^{(p)})}{\bar{\delta}^2 \frac{d}{d\mu} [\psi_\ell^+] (\mu_\ell^{(p)})} \right| \quad (3.7)$$

holds. If this inequality already holds, no change need be made. Otherwise, we redefine η to satisfy (3.7) and continue iterating until we have a new solution of the now tighter inequality (3.4).

A separate issue is the error introduced by the quadrature approximations used to estimate $\psi(\mu)$. We ensure that the relative error introduced here is small by requiring the inequality

$$\left| \frac{\psi_\ell^+ (\mu_\ell^{(p)}) - \psi_\ell^- (\mu_\ell^{(p)})}{\psi_\ell^- (\mu_\ell^{(p)})} \right| < \gamma \quad (3.8)$$

to hold.

If, in addition to (3.4) and (3.8), $\mu_\ell^{(p)}$ also satisfies

$$(1 - \eta) \bar{\delta}^2 \leq \varphi_\ell^- (\mu_\ell^{(p)}), \quad (3.9)$$

then both inequalities (3.2) hold for $\mu = \mu_\ell^{(p)}$, and we accept $\mu_\ell^{(p)}$ as an approximation of μ_θ .

It follows from (2.21), (2.22), and (2.23) that the differences $\varphi_\ell^+(\mu) - \varphi_\ell^-(\mu)$ and $\psi_\ell^+(\mu) - \psi_\ell^-(\mu)$ decrease monotonically to zero as ℓ increases for any fixed $\mu > 0$. Therefore, (3.8) and (3.9) hold for sufficiently large ℓ .

The above discussion suggests that if either one of the inequalities (3.8) and (3.9) does not hold, then the number of Lanczos bidiagonalization steps ℓ should be increased. Specifically, when $\mu_\ell^{(p)}$ satisfies (3.4) but fails to satisfy either (3.8) or (3.9), we carry out one more Lanczos bidiagonalization step, seeking to approximate the largest zero $\mu_{\ell+1}$ of the function

$$h_{\ell+1}^+(\mu) := \varphi_{\ell+1}^+(\mu) - (1 + (1 - \nu)\eta)\bar{\delta}^2,$$

using the same zero-finder as above with initial approximate solution $\mu_{\ell+1}^{(0)} := \mu_\ell^{(p)}$.

Finally, when $\mu_\ell^{(p)}$ has been found satisfying (3.4), (3.8), and (3.9), we approximate $L(\theta)$ by the Gauss quadrature $\psi_\ell^-(\mu_\ell^{(p)})$ of $\psi(\mu_\ell^{(p)})$.

4 Solving $L(\theta) = \varepsilon^2$

The problem (1.2) is much easier to solve when only one of the constraints is active at the solution, in which case the transformation (1.3) is not needed (and is not valid). Consider the vectors

$$\mathbf{x}_1 := \operatorname{argmin}_{\mathbf{w}} \mathbf{w}^T \mathbf{x} \quad \text{subject to} \quad \|\mathbf{x} - \mathbf{d}\| = \delta \quad (4.1)$$

and

$$\mathbf{x}_2 := \operatorname{argmin}_{\mathbf{w}} \mathbf{w}^T \mathbf{x} \quad \text{subject to} \quad \|A\mathbf{x} - \mathbf{b}\| = \varepsilon \quad (4.2)$$

obtained by applying the constraints from (1.2) separately. If the solution \mathbf{x}_1 of (4.1) satisfies $\|A\mathbf{x}_1 - \mathbf{b}\| \leq \varepsilon$, then it is not hard to see that \mathbf{x}_1 solves (1.2). Moreover, if the solution \mathbf{x}_2 of (4.2) satisfies $\|\mathbf{x}_2 - \mathbf{d}\| \leq \delta$, then \mathbf{x}_2 solves (1.2); see [6] for more discussion of these cases.

The numerical method therefore begins by computing $\mathbf{x}_1 = \mathbf{d} - \delta \mathbf{w}$. If $\|A\mathbf{x}_1 - \mathbf{b}\| \leq \varepsilon$, then the solution of (1.2) is $\mathbf{w}^T \mathbf{d} - \delta$, and we are done. In the case of interest, when A is of ill-determined rank, the norm of \mathbf{x}_2 from (4.2) is almost sure to be much larger than δ for the following reason. Because many of the singular values of A are clustered about the origin, many of the semi-axes of the ellipsoid $\|A\mathbf{x} - \mathbf{b}\| = \varepsilon$ are very long. For \mathbf{x}_2 to satisfy $\|\mathbf{x}_2 - \mathbf{d}\| \leq \delta$, \mathbf{w} would have to be numerically orthogonal to nearly all of these semi-axes. This is very unlikely. For the numerical experiments of Section 5, we therefore make the reasonable assumption that $\|\mathbf{x}_2 - \mathbf{d}\| > \delta$. It should be noted that when A is well conditioned, this assumption may not be valid.

If \mathbf{x}_1 from (4.1) does not solve (1.2), then the transformation (1.3) is valid and is applied. The function $L(\theta)$ is defined at any θ for which the constraint set is nonempty, or at any θ in the interval

$$[\theta_*, \theta^*] := \{\mathbf{w}^T \mathbf{x} : \|\mathbf{x} - \mathbf{d}\| = \delta\} = [\mathbf{w}^T \mathbf{d} - \delta, \mathbf{w}^T \mathbf{d} + \delta].$$

In [7], the authors prove that the function $L(\theta)$ from (1.3) is continuous and convex. This implies that the number of zeros of the function $L(\theta) - \varepsilon^2$ is either 0, 1, or 2. Since we are assuming that the solution \mathbf{x}_1 of (4.1) satisfies $\|A\mathbf{x}_1 - \mathbf{b}\| > \varepsilon$, it follows that $L(\theta_*) - \varepsilon^2 > 0$. Due to the convexity of L , assuming the existence of at least one zero, application of Newton's method with $\theta_0 = \theta_*$ or the secant method with $\theta_* = \theta_0 < \theta_1 \ll \theta^*$ produces a sequence of iterates θ_i that converges monotonically to the smallest zero of $L(\theta) - \varepsilon^2$, provided that sufficiently accurate approximations of L and L' are computed. The convergence is quadratic for Newton's method and super linear for the secant method. Although the computed function values $L(\theta_i)$ are only approximations obtained by working with a projected problem of greatly reduced dimension, monotonic convergence was nevertheless observed in the numerical experiments of Section 5, with both zero-finders, thanks to the small relative errors in the approximations of $L(\theta)$ ensured by (3.7) and (3.8).

In the numerical experiments of the next section, both Newton's method and the secant method are used, terminating when (1.3) has been solved within some prescribed relative error τ , i.e., when

$$\left| \frac{L(\theta_i) - \varepsilon^2}{\varepsilon^2} \right| < \tau$$

holds.

The test problems are chosen in such a manner that a solution exists, but of course this cannot be guaranteed in practical applications when the exact norms of the solution and error are not known. In such situations, we utilize part of the *safeguarded secant* algorithm for iterating in θ , described in [7]. This method seeks to find the smallest zero of $L(\theta) - \varepsilon^2$, if one exists, or otherwise confirms nonexistence of a zero. If $L(\theta) - \varepsilon^2$ has no zero in the interval $[\theta_*, \theta^*]$, which happens precisely when the two constraint sets from (1.2) are disjoint, then one of two alternatives must eventually happen: either an iterate $\theta_i > \theta^*$ is obtained, or a backward step occurs, $\theta_i < \theta_{i-1}$. In either case, the algorithm then iterates to minimize L , thereby bracketing the solution if a negative value is found, and otherwise confirming nonexistence of a zero.

We now derive an estimate for the derivative $L'(\theta)$ which allows solution of (1.3) by Newton's method. At the end of each evaluation of $L(\theta)$, an estimate of the Tikhonov parameter $\mu(\theta)$ is available. This allows estimation of $L'(\theta)$ at almost no cost using already computed quantities. Indeed,

$$L(\theta) = \psi(\mu) = \mu^2 (\mathbf{b}' - \theta' A \mathbf{w})^T (\bar{A} \bar{A}^T + \mu I)^{-2} (\mathbf{b}' - \theta' A \mathbf{w}),$$

where $\mu = \mu(\theta)$ satisfies the constraint

$$\varphi(\mu) = (\mathbf{b}' - \theta' A \mathbf{w})^T \bar{A} (\bar{A}^T \bar{A} + \mu I)^{-2} \bar{A}^T (\mathbf{b}' - \theta' A \mathbf{w}) = \delta^2 - (\theta')^2. \quad (4.3)$$

Implicit differentiation of (4.3) with respect to θ yields

$$-2(A \mathbf{w})^T \bar{A} (\bar{A}^T \bar{A} + \mu I)^{-2} \bar{A}^T (\mathbf{b}' - \theta' A \mathbf{w}) + \varphi'(\mu) \frac{d\mu}{d\theta} = -2\theta',$$

implying that

$$\begin{aligned}\frac{d\mu}{d\theta} &= \frac{2(A\mathbf{w})^T \bar{A}(\bar{A}^T \bar{A} + \mu I)^{-2} \bar{A}^T (\mathbf{b}' - \theta' A\mathbf{w}) - 2\theta'}{\varphi'(\mu)} \\ &= \frac{\varphi_3(\mu) - \varphi_1(\mu) - [2\theta' + 1]\varphi_2(\mu) - 2\theta'}{\varphi'(\mu)},\end{aligned}$$

where quadrature estimates of all of these quantities already have been computed for other reasons. Finally,

$$\begin{aligned}L'(\theta) &= -2\mu^2(A\mathbf{w})^T(\bar{A}\bar{A}^T + \mu I)^{-2}(\mathbf{b}' - \theta' A\mathbf{w}) + \psi'(\mu)\frac{d\mu}{d\theta} \\ &= \psi_1(\mu) + [1 + 2\theta']\psi_2(\mu) - \psi_3(\mu) + \psi'(\mu)\frac{d\mu}{d\theta}.\end{aligned}$$

In the numerical implementation, appropriate quantities in this expression are approximated by Gauss quadratures.

5 Numerical experiments

All numerical experiments are conducted in MATLAB R2011a with $\epsilon_{\text{machine}} = 2.22 \times 10^{-16}$. The columns of the Lanczos bidiagonalization matrices U_ℓ and V_ℓ are reorthogonalized after each step to ensure numerical orthogonality. We consider a highly ill-conditioned problem arising from the discretization of a singular Fredholm integral equation of the first kind discussed by Phillips [18], namely

$$\int_{-6}^{+6} K(s, t)f(t)dt = g(s),$$

where the kernel K , the solution f , and the right-hand side g are given by

$$\begin{aligned}K(s, t) &= \kappa(s - t), & f(t) &= \kappa(t), \\ g(s) &= (6 - |s|)(1 + \frac{1}{2}\cos(s\pi/3)) + \frac{9}{2\pi}\sin(|s|\pi/3),\end{aligned}$$

with

$$\kappa(t) = \begin{cases} 1 + \cos(t\pi/3), & |t| < 3, \\ 0, & |t| \geq 3. \end{cases}$$

The function `phillips` from the MATLAB package Regularization Tools by Hansen [14] is used to generate the discretized kernel $A \in \mathbb{R}^{n \times n}$ and the desired solution \mathbf{x}_{true} , as well as $\mathbf{b}_{\text{true}} := A\mathbf{x}_{\text{true}}$. The vector \mathbf{b} of (1.2) is given by $\mathbf{b} = \mathbf{b}_{\text{true}} + \mathbf{e}$, where the error vector \mathbf{e} has independent normally distributed entries with zero mean, and is scaled so that $\|\mathbf{e}\|/\|\mathbf{b}_{\text{true}}\| = 1 \times 10^{-3}$, i.e., \mathbf{b} is contaminated by 0.1% white Gaussian noise. We set $\mathbf{d} = \mathbf{0}$ in (1.2). When an approximation \mathbf{x} of \mathbf{x}_{true} has been obtained, a discrete approximation of f is then given by (t_i, f_i) for $1 \leq i \leq n$, where

$$h = \frac{12}{n}, \quad t_i = -6 + \left(i - \frac{1}{2}\right)h, \quad f_i = x_i/\sqrt{h}, \quad (5.1)$$

and x_i is the i th component of \mathbf{x} . See [14] for more details regarding the discretization.

In applications, the noise \mathbf{e} corrupting the observed data \mathbf{b} may be random in nature, and knowledge of the distribution of the error, as opposed to the size of the error, may be all that is available. In such a case, the distribution can be used to determine an upper bound ε for $\|\mathbf{e}\|$ holding with probability p , where p is a confidence level specified by the user. For instance, if the components of \mathbf{e} are independent, normally distributed random variables with mean 0 and variance σ^2 , then the random variable $\|\mathbf{e}\|^2/\sigma^2$ belongs to the χ^2 distribution with n degrees of freedom. We therefore choose ε so that

$$\int_0^{\varepsilon^2/\sigma^2} \chi_n^2(\rho) d\rho = p$$

holds, where χ_n^2 is the probability density function for the χ^2 distribution with n degrees of freedom. Solving (1.2) with $\mathbf{w} = \pm \mathbf{e}_i$, where \mathbf{e}_i is the i th vector of the standard basis, then provides bounds constituting a confidence interval for x_i holding with probability p . See [15, 6, 7] for further discussions on this topic.

We first discretize the problem with $n = 2^{10} = 1024$ and compare the performance of the current method with that of the method proposed in [7], in which each evaluation of $L(\theta)$ is accomplished by solving (1.5) with the iterative method LSTRS [19, 20], and (1.3) is solved by the secant method. The condition number of A was 2.90×10^{10} , and the norm of the least squares solution $A^\dagger \mathbf{b}$ was 2.01×10^6 . Bounds were computed for 16 equally spaced coordinates of the solution, using $\delta = \|\mathbf{x}_{\text{true}}\| = 3.00$, $\varepsilon = \|\mathbf{e}\| = 1.53 \times 10^{-2}$, and $\tau = 1 \times 10^{-1}$.

Our Method: We used the auxiliary parameters $\nu = 1 \times 10^{-2}$, $\eta = 1 \times 10^{-2}$ (as the default starting value, updated when necessary to satisfy (3.7)) and $\gamma = 1 \times 10^{-4}$. When solving (3.1), the minimum value of η required by (3.7) was about 1×10^{-7} when using the secant method, and 1×10^{-8} when using Newton's method. When an arbitrary initial value of μ was required, we used the largest power of 10 such that $h_\ell^+(\mu) > 0$ in (3.3). This choice ensured successful convergence of Newton's method for solving (3.3) in all cases. Use of the secant method with $\theta_0 = \theta_*$ and $\theta_1 = \theta_* + 2\delta/10^4$ required an average of 24 iterates θ_i to solve (1.3), versus only 16 for Newton's method with $\theta_0 = \theta_* + 2\delta/10^4$. In either case, the average number of Lanczos bidiagonalization steps, ℓ , required to solve (1.2) was about 12, demanding 72 matrix-vector product evaluations with either \bar{A} or \bar{A}^T due to the three separate partial Lanczos bidiagonalizations (plus about two matrix-vector product evaluations for other reasons). Each matrix-vector product evaluation with \bar{A} or \bar{A}^T requires a matrix-vector product evaluation with A or A^T , respectively. It should be noted that the number of Lanczos bidiagonalization steps ℓ , and hence the number of matrix-vector products with \bar{A} and \bar{A}^T , is independent of the number of iterates θ_i . The maximum errors of the upper and lower bounds across all 16 nodes were 8.63×10^{-1} and 8.73×10^{-1} (after transforming via

(5.1)), respectively, when using the secant method, versus 8.77×10^{-1} and 8.88×10^{-1} , respectively, when using Newton's method.

LSTRS: We used default parameters except for `lopts.maxiter = 102`, `epsilon.Delta = 10-4`, `epsilon.HC = 10-6`, `epsilon.nu = 10-4`, `epsilon.Int = 0`, `eigensolverpar.tol = 10-4`, and `eigensolverpar.maxit = 102`. These values were chosen to ensure convergence of the secant method without a backward step (a problem observed for less stringent parameters due to imprecise function evaluations). The average number of matrix-vector products required to solve (1.5) was 4.87×10^3 , and on average about 24 iterates θ_i were required to solve (1.3), for a total of about 1.17×10^5 matrix-vector products per minimization (1.2). Thus, LSTRS is more expensive than our method by a factor of more than 1000 for the current problem and auxiliary parameters. The maximum errors of the upper and lower bounds across all 16 nodes were 8.71×10^{-1} and 8.57×10^{-1} , respectively, nearly identical to the results produced by our method.

We next exhibit the performance of our method for the same problem, now discretized with $n = 2^{12} = 4096$; see Figure 5.1. The condition number of A was 3.21×10^{12} , and the norm of the least squares solution $A^\dagger b$ was 1.21×10^8 . Bounds were computed for 16 equally spaced coordinates of the solution, using $\delta = \|\mathbf{x}_{\text{true}}\| = 3.00$, $\varepsilon = \|\mathbf{e}\| = 1.53 \times 10^{-2}$, and $\tau = 1 \times 10^{-3}$, and with the same auxiliary parameters as above. When solving (3.1), the minimum value of η required by (3.7) was about 1×10^{-8} when using the secant method, and 1×10^{-9} when using Newton's method. Use of the secant method required an average of 29 iterates θ_i to solve (1.3), versus only 20 for Newton's method. In either case, the average number of Lanczos bidiagonalization steps required to solve (1.2) was about 13, demanding 78 matrix-vector product evaluations with either \bar{A} or \bar{A}^T . The maximum errors of the upper and lower bounds across all 16 nodes were both 1.02 (after transforming via (5.1)) when using the secant method, versus 1.01 for both when using Newton's method.

In Figure 5.1, “a priori u.b.” and “a priori l.b.” are the crude upper and lower bounds for each component, $+\delta/\sqrt{h}$ and $-\delta/\sqrt{h}$, respectively, that would be obtained by solving the trivial problem obtained by removing the first constraint in (1.2). It is clear that the use of both constraints provides much tighter bounds. The second graph in each figure depicts the same data but with a zoomed-in vertical scale.

To test the sensitivity of solving (1.2) to the choices of δ and ε , we repeated the same experiment with $\delta = 1.1\|\mathbf{x}_{\text{true}}\| = 3.30$ and $\varepsilon = 1.1\|\mathbf{e}\| = 1.68 \times 10^{-2}$; see Figure 5.2. As expected, the maximum errors of the upper and lower bounds across all 16 nodes were somewhat larger this time, at 25.5 for both of the maximum errors when using either the secant method or Newton's method. Nevertheless, these errors are still superior to the bounds $\pm\delta$ that would be obtained by abandoning the linear model. This time the secant method required an average of 24 iterates θ_i , versus only 16 for Newton's method, but in either case the average number of Lanczos bidiagonalization steps ℓ required to solve (1.2) was about 12, requiring about 72 matrix-vector products with \bar{A} or \bar{A}^T .

We remark that the sensitivity of the solution of (1.2) to the choice of ε and δ depends on how these choices relate to $\|e\|$ and $\|\mathbf{x}_{\text{true}} - \mathbf{d}\|$. It is instructive to note that the constraint sets $S_1 = \{\|A\mathbf{x} - \mathbf{b}\| \leq \varepsilon\}$ and $S_2 = \{\|\mathbf{x} - \mathbf{d}\| \leq \delta\}$ are an ellipsoid and sphere within \mathbb{R}^n , respectively. The diameters of S_1 and S_2 are increasing functions of ε and δ , respectively, and as exhibited in the last example, the bounds obtained via (1.2) become coarser as these parameters increase. The bounds produced solving (1.2) are valid when $\|e\| \leq \varepsilon$ and $\|\mathbf{x}_{\text{true}} - \mathbf{d}\| \leq \delta$, since in this case $\mathbf{x}_{\text{true}} \in S_1 \cap S_2$. Thus, if ε and δ are chosen too small, the bounds may be invalid. Moreover, if ε and δ are chosen too small, the constraint sets S_1 and S_2 may be disjoint; this event is manifested by the nonexistence of a solution to the equation (1.3).

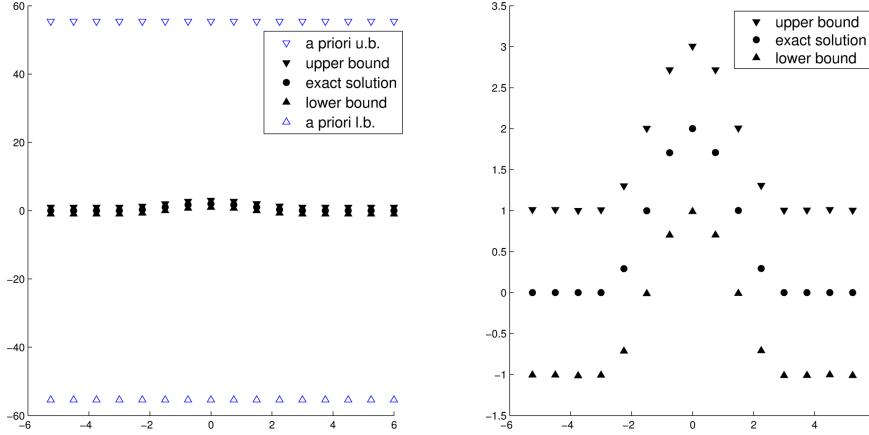


Fig. 5.1 Coordinates $64k, k = 1, \dots, 16$ of $\mathbf{x}_{\text{true}}/\sqrt{h}$, along with the computed upper and lower bounds (produced by Newton's method), the a priori upper bound $+\delta/\sqrt{h}$, and the a priori lower bound $-\delta/\sqrt{h}$ for each component, with $A \in \mathbb{R}^{4096 \times 4096}$ generated by phillips, $\delta = \|\mathbf{x}_{\text{true}}\| = 3.00$, $\varepsilon = \|e\| = 1.53 \times 10^{-2}$, and $\tau = 1 \times 10^{-3}$.

6 Conclusion

The examples in the previous section demonstrate that in some cases, the solution of (1.2) yields much sharper bounds than what could be obtained by using only an estimate of the norm of the solution. Thus, when sufficiently accurate values of the parameters $\|\mathbf{x}_{\text{true}}\|$ and $\|e\|$ are available, the problem (1.2) can be used to compute insightful confidence intervals for components of \mathbf{x}_{true} . However, the quality of the bounds is sensitive to the tightness of the estimates δ and ε of $\|\mathbf{x}_{\text{true}}\|$ and $\|e\|$, as well as to the conditioning of the problem, and the relative orientation of the constraint sets. In the above example, the current method reduces the number of matrix-vector products needed to solve (1.2) by three orders of magnitude relative to the approach presented in [7], while producing nearly identical bounds.

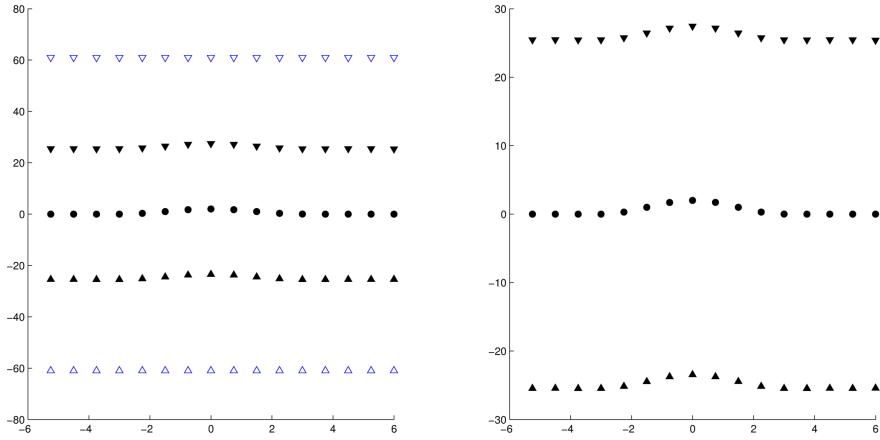


Fig. 5.2 Coordinates $64k, k = 1, \dots, 16$ of $\mathbf{x}_{\text{true}}/\sqrt{h}$, along with the computed upper and lower bounds (produced by Newton's method), the a priori upper bound $+\delta/\sqrt{h}$, and the a priori lower bound $-\delta/\sqrt{h}$ for each component, with $A \in \mathbb{R}^{4096 \times 4096}$ generated by `phillips`, $\delta = 1.1\|\mathbf{x}_{\text{true}}\| = 3.30$, $\varepsilon = 1.1\|\mathbf{e}\| = 1.68 \times 10^{-2}$, and $\tau = 1 \times 10^{-3}$.

References

1. D. Calvetti, G. H. Golub, and L. Reichel, Estimation of the L-curve via Lanczos bidiagonalization, BIT, 39 (1999), pp. 603–619.
2. D. Calvetti, P. C. Hansen, and L. Reichel, L-curve curvature bounds via Lanczos bidiagonalization, Electron. Trans. Numer. Anal., 14 (2002), pp. 134–149.
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 and L. Reichel, Tikhonov regularization with a solution constraint, SIAM J. Sci. Comput., 26 (2004), pp. 224–239.
5. L. Eldén, Algorithms for the regularization of ill-conditioned least squares problems, BIT, 17 (1977), pp. 134–145.
6. L. Eldén, Algorithms for the computation of functionals defined on the solution of a discrete ill-posed problem, BIT, 30 (1990), pp. 466–483.
7. L. Eldén, P. C. Hansen, and M. Rojas, Minimization of linear functionals defined on solutions of large-scale discrete ill-posed problems, BIT, 45 (2005), pp. 329–340.
8. W. Gautschi, Orthogonal Polynomials: Computation and Approximation, Oxford University Press, Oxford, 2004.
9. 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.
10. G. H. Golub and G. Meurant, Matrices, Moments and Quadrature with Applications, Princeton University Press, Princeton, 2010.
11. G. H. Golub and U. von Matt, Quadratically constrained least squares and quadratic problems, Numer. Math., 59 (1991), pp. 561–580.
12. G. H. Golub and U. von Matt, Tikhonov regularization for large scale problems, in Workshop on Scientific Computing, eds. G. H. Golub, S. H. Lui, F. Luk, and R. Plemmons, Springer, New York, 1997, pp. 3–26.
13. G. H. Golub and U. von Matt, Generalized cross-validation for large scale problems, J. Comput. Graph. Stat., 6 (1997), pp. 1–34.
14. P. C. Hansen, Regularization tools version 4.0 for Matlab 7.3, Numer. Algorithms, 46 (2007), pp. 189–194.

15. D. P. O'Leary and B. W. Rust, Confidence intervals for inequality-constrained least squares problems, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 473–489.
16. 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.
17. C. C. Paige and M. A. Saunders, Algorithm 583 LSQR: Sparse linear equations and least squares problems, ACM Trans. Math. Software, 8 (1982), pp. 195–209.
18. D. L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, J. ACM 9 (1962), pp. 84–97.
19. M. Rojas, S. A. Santos, and D. C. Sorensen, A new matrix-free algorithm for the large-scale trust-region subproblem, SIAM J. Optim., 11 (2000), pp. 611–646.
20. M. Rojas, S. A. Santos, and D. C. Sorensen, LSTRS: Matlab software for large-scale trust-region subproblems and regularization. Technical Report 2003-4, Department of Mathematics, Wake Forest University, 2004. Available at <http://web.math.wfu.edu/~mrojas/software.html>.