SIAM J. Sci. Comput.
Vol. 23, No. 6, pp. 1842–1860

A TRUST-REGION APPROACH TO THE REGULARIZATION OF LARGE-SCALE DISCRETE FORMS OF ILL-POSED PROBLEMS*

MARIELBA ROJAS† AND DANNY C. SORENSEN‡

Abstract. We consider large-scale least squares problems where the coefficient matrix comes from the discretization of an operator in an ill-posed problem, and the right-hand side contains noise. Special techniques known as regularization methods are needed to treat these problems in order to control the effect of the noise on the solution. We pose the regularization problem as a quadratically constrained least squares problem. This formulation is equivalent to Tikhonov regularization, and we note that it is also a special case of the trust-region subproblem from optimization. We analyze the trust-region subproblem in the regularization case and we consider the nontrivial extensions of a recently developed method for general large-scale subproblems that will allow us to handle this case. The method relies on matrix-vector products only, has low and fixed storage requirements, and can handle the singularities arising in ill-posed problems. We present numerical results on test problems, on an inverse interpolation problem with field data, and on a model seismic inversion problem with field data.

Key words. regularization, constrained quadratic optimization, trust region, Lanczos method, ill-posed problems, inverse problems, seismic inversion

AMS subject classifications. 86A22, 65K10, 90C06

PII. S1064827500378167

1. Introduction.

Discrete forms of ill-posed problems arise when we discretize the continuous operator in an ill-posed problem and introduce experimental data contaminated by noise. One of the main sources of ill-posed problems are inverse problems, where we want to determine the internal structure of a system from the observed behavior of the system. Inverse problems arise in many important applications such as image processing [2], seismic inversion [39], and medical and seismic tomography [30], [32]. Discrete forms of ill-posed problems are usually formulated as linear systems or least squares problems. The focus of this paper is the numerical treatment of large-scale discrete forms of ill-posed least squares problems.

We are interested in recovering x_{LS}, the minimum norm solution of

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|, \tag{1}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $m \geq n$. Throughout the paper we assume that A comes from the discretization of a continuous operator in an ill-posed problem, and instead of the *exact* data vector b, only a *perturbed* data vector $\bar{b}$ is available. Specifically, we regard $\bar{b}$ as $\bar{b} = b + s$, where s is a random vector of uncorrelated

*Received by the editors September 15, 2000; accepted for publication (in revised form) September 29, 2001; published electronically February 20, 2002.
http://www.siam.org/journals/sisc/23-6/37816.html

†CERFACS, 42 avenue Gaspard Coriolis, 31057 Toulouse CEDEX 1, France. Current address: Department of Mathematics, Wake Forest University, 302 Calloway Hall, Winston-Salem, NC 27109 (mrojas@math.wfu.edu). This author was supported in part by NSF cooperative agreement CCR-9120008.

‡Department of Computational and Applied Mathematics, Rice University, 6100 Main St., Houston, TX 77005-1892 (sorensen@caam.rice.edu). This author was supported in part by NSF grant CCR-9988393 and in part by the Los Alamos National Laboratory Computer Science Institute (LACSI) through LANL contract 03891-99-23 as part of the prime contract (W-7405-ENG-36) between the Department of Energy and the Regents of the University of California.

noise. The norm is the Euclidean norm throughout the paper, and it will be denoted by $\|\cdot\|$.

We will assume that the matrix A is large and might not be available explicitly but that we can compute the action of A and A^T on vectors of the appropriate dimensions. We will also assume that errors in A, due to discretization or finite-precision representation, are small in comparison to the noise in $\bar{b}$. Finally, we will not assume any particular structure for A.

Given the fact that only $\bar{b}$ is available, we could formulate the problem

$$\min_{x\in\mathbb{R}^n} \|Ax-\bar{b}\| \tag{2}$$

and use its minimum norm solution, denoted by $\bar{x}_{LS}$, to approximate x_{LS}. Unfortunately, as we shall see, the two solutions might differ considerably.

If we use a reasonably accurate discretization to obtain A, this matrix will be highly ill-conditioned with a singular spectrum that decays to zero gradually, a large cluster of small singular values, and high-frequency components of the singular vectors associated with small singular values. If, in addition, the discrete Picard condition [16] holds, we will have that the expansion coefficients of the exact data vector b in the left singular vectors basis decay to zero faster than the singular values of A, while the expansion coefficients of the noise vector s remain constant. Therefore, those components of $\bar{x}_{LS}$ corresponding to small singular values are magnified by the noise.

As a consequence of the ill conditioning of the matrix A and the presence of noise in the right-hand side, standard numerical algebra methods such as the ones discussed in [3], [13, Chap. 5], and [26] applied to problem (2) produce meaningless solutions with very large norm. Therefore, to solve these problems, we need special techniques known as *regularization* or *smoothing methods*. These methods aim to recover information about the desired solution of the unknown problem with exact data from the solution of a better conditioned problem that is related to the problem with noisy data but incorporates additional information about the desired solution. The formulation of the new problem involves a special parameter known as the *regularization parameter*, used to control the effect of the noise on the solution. The conditioning of the new problem depends on the choice of the regularization parameter. Excellent surveys on regularization methods can be found, for example, in [15], [20] and more recently in [31].

While there are many alternatives for solving small- to medium-scale problems, this is not the case in the large-scale setting. However, in recent years interesting methods for large-scale ill-posed problems have been proposed. Among these are Golub and von Matt [14], Björck, Grimme, and Van Dooren [4], Calvetti, Reichel, and Zhang [8], Rojas, Santos, and Sorensen [35], as well as several variants of the conjugate gradient method on the normal equations (CGLS), including the use of preconditioners chosen according to the structure of the problem [22], [24], [29]. In spite of these developments, the efficient solution of large-scale discrete forms of ill-posed problems remains a challenge.

In practice, the most common approach is to apply the conjugate gradient method to the normal equations associated with problem (2), taking advantage of what seems to be an intrinsic regularization property of this method. It has been observed that, at early stages, CGLS generates iterates with components in the direction of right singular vectors associated with *large* singular values, while components associated with small singular values come into play at later stages. This observation leads to the heuristic that the number of iterations acts as a regularization parameter.

Thus, we could compute a regularized solution by stopping the iteration before the unwanted components contaminate the current approximation. The success of this approach depends, in the first place, on the reliability of the heuristic and, second, on accurately determining when to stop the iteration, which is a difficult problem in itself, and most practical termination strategies rely on visual inspection. We are not aware of any systematic termination criterion for this approach. Alternatively, we could use CGLS on the Tikhonov regularization problem, discussed in the next section, which can be formulated as a damped least squares problem. The success of this approach depends on a good choice of the damping parameter, and also on the availability of a preconditioner, and good general preconditioners have not emerged yet. The approach we propose here does not depend on either a heuristic or a preconditioner.

In this paper, we formulate the regularization problem as a quadratically constrained least squares problem. It is well known that this approach is equivalent to Tikhonov regularization (cf. [10] and the references therein), and we also observe that the problem is a special case of the problem of minimizing a quadratic subject to a quadratic constraint, which is known in optimization as the trust-region subproblem arising in trust-region methods (see also [3, sections 5.3 and 9.2.3]). The connection between the trust-region subproblem and the regularization problem is well known, but the specific nature of the numerical difficulties for solving the regularization problem as a trust-region subproblem was first studied extensively in [34]. We discuss the properties of the trust-region subproblem in the regularization case and apply the recently developed method LSTRS [35] for the large-scale trust-region subproblem to the regularization of discrete forms of ill-posed problems from a variety of applications. The method relies on matrix-vector products only, has low and fixed storage requirements and robust stopping criteria, and computes both a solution and the corresponding Tikhonov regularization parameter. Moreover, LSTRS can efficiently handle the high-degree singularities associated with ill-posed problems. Most of the results presented here are based on [34].

The organization of the paper is as follows. In section 2 we describe our regularization approach and show its connection with Tikhonov regularization and with the trust-region subproblem. In section 3 we describe the trust-region subproblem and show its special properties in the discrete ill-posed case. In section 4 we describe the method LSTRS from [35] and discuss the issues related to ill-posed problems. In section 5 we present numerical results of LSTRS on regularization problems, including test problems from the Regularization Tools package [19], an inverse interpolation problem with field data, and a model seismic inversion problem with field data. We present some conclusions in section 6.

2. Regularization through trust regions.

As we mentioned before, regularization involves the formulation of a problem related to both the original problem with exact data and the problem with noisy data, where we incorporate a priori information such as the size or smoothness of the desired solution, the noise level in the data, or the statistical properties of the noise process.

One of the most popular regularization approaches is the classical Tikhonov regularization approach [40]

$$\min_{x \in \mathbb{R}^n} \ \|Ax - \bar{b}\|^2 + \varepsilon^2 \|x\|^2, \tag{3}$$

where $\varepsilon^2 > 0$ is the regularization parameter, and where the term $\|x\|^2$ could also be of the form $\|Lx\|^2$ for a general square or rectangular matrix L. This matrix could be, for example, the identity matrix as in (3), or a discrete form of first derivative. In the first case, the regularization parameter ε acts as a penalty parameter on the size of the solution, while in the latter case ε acts as a penalty parameter on the smoothness of the solution. Notice that if L is any square and nonsingular matrix, then a change of variable will reduce the problem to the form in (3). We can also accomplish such a transformation when L is a full-rank rectangular matrix by means of the methods in [10], [15]. Throughout the paper we assume that this transformation is possible, and therefore we consider only the case in which L is the identity matrix.

Observe that given ε, problem (3) becomes a damped least squares problem that we can solve with standard numerical linear algebra techniques for medium- and large-scale problems (cf. [3], [13]). However, determining an optimal value for the Tikhonov regularization parameter ε^2 can be as difficult as the original problem, and most of the methods currently available require the solution of several problems of type (3) for different values of ε. This approach might be very expensive in the large-scale setting. Recent and promising methods for computing the Tikhonov regularization parameter for large-scale problems have been proposed in [7], [8], and [25]. Calvetti, Reichel, and Zhang [8] propose a very elegant way of computing the parameter from the noise level in the data; Calvetti, Golub, and Reichel [7] propose a strategy based on the L-curve (see [17], [18]); while Kilmer and O'Leary [25] propose several strategies for computing the parameter from a problem in an appropriate subspace of smaller dimension.

In this work we will not assume a priori knowledge of the noise level or noise properties. Instead, we will assume that some information about the size or smoothness of the desired solution is available, and we formulate the regularization problem as

$$\min_{s.t. \|x\| \leq \Delta} \|Ax - \bar{b}\| \tag{4}$$

with $\Delta > 0$. As we show next, this formulation is equivalent to Tikhonov regularization.

Observe that if $\bar{b} \notin \mathcal{R}(A)$, where $\mathcal{R}(A)$ is the range of A, any solution of problem (4) is a regular point, and therefore the Karush–Kuhn–Tucker conditions for a feasible point x_* to be a solution of problem (4) with corresponding Lagrange multiplier λ_* are $(A^T A - \lambda_* I)x_* = -A^T \bar{b}$, $\lambda_* \leq 0$, and $\lambda_*(\|x_*\| - \Delta) = 0$. Further, since (4) is a convex quadratic problem, these conditions are both necessary and sufficient. Equivalence with problem (3) follows directly, since a solution x_* to problem (4) is also a solution to problem (3) corresponding to $\varepsilon^2 = -\lambda_*$. Conversely, if x_ε is a solution of (3) for a given ε, then x_ε solves problem (4) for $\Delta = \|x_\varepsilon\|$.

While Tikhonov regularization involves the computation of a parameter that does not necessarily have a physical meaning in most problems, the quadratically constrained least squares formulation has the advantage that, in some applications, the physical properties of the problem either determine or make it easy to estimate an optimal value for the norm constraint Δ. This is the case, for example, in image restoration where Δ represents the energy of the target image (cf. [2]).

Another example is the following problem closely related to (4):

$$\min_{s.t. \|Ax - \bar{b}\| \leq \rho} \|x\|,$$

where ρ is an estimate of the noise level in the data. For A nonsingular, the problem can be transformed into the form (4) by means of a change of variable. It is possible to do this in some special applications where an effective approximation to the inverse of A is available. This is the case in the example presented in section 5.3.

An additional advantage of the quadratically constrained least squares formulation is that it is a special case of a well-known problem in optimization, namely, that of minimizing a quadratic on a sphere or the trust-region subproblem

$$\min_{s.t. \|x\| \le \Delta} \frac{1}{2} x^T H x + g^T x, \tag{5}$$

where $H \in \mathbb{R}^{n \times n}$, $H = H^T$, $g \in \mathbb{R}^n$, and $\Delta > 0$. Problem (4) is a special case of (5) when $H = A^T A$ and $g = -A^T \bar{b}$.

The high degree of structure of the trust-region subproblem leads to strong theoretical properties and makes it possible to design efficient solution methods. For this reason we shall formulate the regularization problem as a trust-region subproblem.

3. The trust-region subproblem.

In this section we present the properties of the trust-region subproblem. In section 3.1, we consider the problem when H is any symmetric matrix in $\mathbb{R}^{n \times n}$, and g is any vector in $\mathbb{R}^n$. In section 3.2, we focus on the special case when $H = A^T A$, $g = -A^T \bar{b}$, and in addition A is a discretized version of a continuous operator in an ill-posed problem and $\bar{b}$ contains noise.

3.1. Structure of the problem.

A first observation about the trust-region subproblem is that it always has a solution. A not-so-obvious and quite remarkable fact about the problem is the existence of a characterization of its solutions, discovered independently by Gay [11] and Sorensen [36]. The result is contained in the following lemma.

LEMMA 3.1 (see [36]). *A feasible vector* $x_* \in \mathbb{R}^n$ *is a solution to* (5) *with corresponding Lagrange multiplier* λ_* *if and only if* x_*, λ_* *satisfy* $(H - \lambda_* I)x_* = -g$ *with* $H - \lambda_* I$ *positive semidefinite,* $\lambda_* \le 0$*, and* $\lambda_*(\Delta - \|x_*\|) = 0$.

Proof. See [36] for the proof. □

The optimality conditions imply that all the solutions of the trust-region subproblem are of the form $x = -(H - \lambda I)^\dagger g + z$ for $z \in \mathcal{N}(H - \lambda I)$, where $\mathcal{N}(\cdot)$ denotes the null space of a matrix and † denotes pseudoinverse. These solutions may lie in the interior or on the boundary of the set $\{x \in \mathbb{R}^n \mid \|x\| \le \Delta\}$. There are no solutions on the boundary if and only if H is positive definite and $\|H^{-1} g\| < \Delta$ (see [28]). In this case, the unique interior solution is $x = -H^{-1} g$ with Lagrange multiplier $\lambda = 0$. Boundary solutions satisfy $\|x\| = \Delta$ with $\lambda \le \delta_1$, where δ_1 is the smallest eigenvalue of H. The case $\lambda = \delta_1$ can only occur if $\delta_1 \le 0$, $g \perp \mathcal{S}_1$, where $\mathcal{S}_1 \equiv \mathcal{N}(H - \delta_1 I)$, and $\|(H - \delta_1 I)^\dagger g\| \le \Delta$. This corresponds to the so-called *hard case*, which poses great difficulties for the numerical solution of the trust-region subproblem since in this case it is necessary to compute an approximate eigenvector associated with the smallest eigenvalue of H. Moreover, in practice g will be nearly orthogonal to $\mathcal{S}_1$, and we can expect greater numerical problems in this case. We call this situation a *near hard case*. Note that whenever g is nearly orthogonal to $\mathcal{S}_1$ there is the possibility for the hard case or near hard case to occur. Therefore we call this a *potential hard case*. We show in section 3.2 that the potential hard case is precisely the common case for discrete ill-posed problems.

The conditions in Lemma 3.1 are computationally attractive since they provide a means for reducing the problem of computing boundary solutions for the trust-region subproblem from an n-dimensional problem to a zero-finding problem in one variable. We can accomplish this, for example, by solving the following equation in λ, known as the *secular equation*:

$$\Delta - \|x_\lambda\| = 0, \tag{6}$$

where $x_\lambda = -(H - \lambda I)^{-1} g$, and λ is monitored to ensure that $H - \lambda I$ is positive definite. Newton's method is particularly efficient for solving an equation equivalent to (6), and this approach, due to Moré and Sorensen [28], is the method of choice whenever it is affordable to compute the Cholesky factorization of matrices of the form $H - \lambda I$. However, in some applications this computation may be prohibitive either because of storage considerations or because the matrix H is not explicitly available. Therefore, we need other strategies to treat the problem in those cases. We describe one such strategy in section 4.

3.2. Discrete ill-posed trust-region subproblem. We now study the trust-region subproblem in the special case when $H = A^T A$ and $g = -A^T \bar{b}$, where A comes from the discretization of a continuous operator in an ill-posed problem, and $\bar{b}$ contains noise. We will show that the potential hard case is the common case for these problems and also that it will occur in a *multiple* instance, where g is orthogonal to the eigenpaces associated with several of the smallest eigenvalues of H. This was first shown in [34] and is a consequence of the following result.

LEMMA 3.2. *Let $H = A^T A$ and $g = -A^T \bar{b}$, with $\bar{b} = b + s$. Suppose σ_k is the kth largest singular value of A with multiplicity m_k. Suppose u_j, v_j, $1 \le j \le m_k$, are left and right singular vectors associated with σ_k. Then*

$$g^T v_j = -\sigma_k (u_j^T b + u_j^T s) \ , \quad 1 \le j \le m_k.$$

Proof. The result follows directly assuming $A = U \Sigma V^T$ is a singular value decomposition of A, since this yields $g = -V \Sigma U^T \bar{b}$ with V orthogonal. ∎

Since $H = A^T A = V \Sigma^2 V^T$, we see that for discrete ill-posed problems, the hard case is always present in an extreme form. Lemma 3.2 implies that whenever σ_k is very small then, for any reasonable noise level in $\bar{b}$, g will be nearly orthogonal to the subspace spanned by the right singular vectors associated with σ_k. This is precisely the case in discrete ill-posed problems, where the matrix A has a large cluster of very small singular values, and therefore we can expect g to be nearly orthogonal to the right singular vectors associated with such singular values. Since these vectors are eigenvectors corresponding to the smallest eigenvalues of $A^T A$, then g will be orthogonal to the eigenspaces corresponding to several of the smallest eigenvalues of $A^T A$ and the potential hard case will occur in a *multiple* instance. Figure 1 illustrates this situation for problem **foxgood** from the Regularization Tools package by Hansen [19]. The problem is of dimension 300, and in the logarithmic plot we observe that $g^T v_k$ is of order 10^{-15} for approximately 292 of the right singular vectors of A.

Observe that for large noise level and σ_k not so small, g will not be nearly orthogonal to the eigenspace corresponding to the smallest eigenvalue of $A^T A$ and the hard case will not occur. Therefore, in this case a high noise level implies a less difficult trust-region subproblem. However, we do not expect to compute a good approximation in the presence of large noise.

4. LSTRS for discrete ill-posed problems. In this section we give a brief description of the method LSTRS from [35]. We present the method for a general symmetric matrix H and nonzero vector g and discuss the advantages of using this method for the special case of large-scale discrete ill-posed trust-region subproblems of type (4). LSTRS is based on formulating the trust-region subproblem as a parameterized eigenvalue problem. Such formulation comes from the observation that

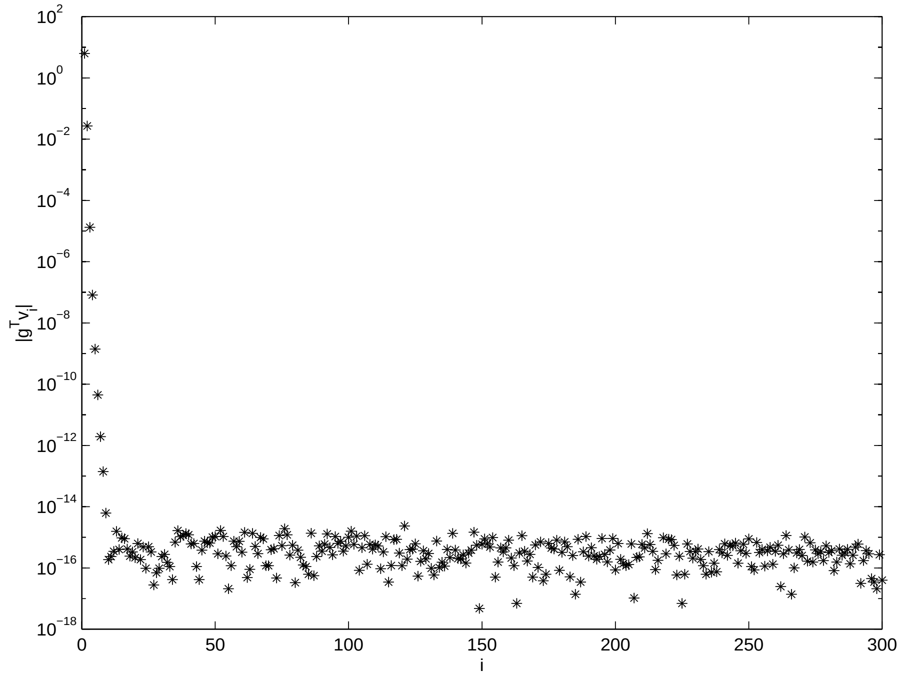

FIG. 1. *Orthogonality of g with respect to right singular vectors of a discretized operator in an ill-posed problem.*

if x_*, λ_* solve problem (5), then for $\alpha = \lambda_* - g^T x_*$, problem (5) is equivalent to

$$\begin{array}{ll} \min & \frac{1}{2} y^T B_\alpha y \\ \text{s.t.} & y^T y \le 1 + \Delta^2, \; e_1^T y = 1, \end{array} \tag{7}$$

where e_1 is the first canonical unit vector in $\mathbb{R}^{n+1}$ and $B_\alpha = \left(\begin{smallmatrix} \alpha & g^T \\ g & H \end{smallmatrix}\right)$. The solution of the trust-region subproblem consists of the last n components of the solution of problem (7).

Problem (7) suggests that if we know the optimal value for α, we can solve the trust-region subproblem by solving an eigenvalue problem for the smallest eigenvalue of B_α and an eigenvector with special structure. To see this, observe that if $\{\lambda, (1, x^T)^T\}$ is an eigenpair of B_α, then

$$\begin{pmatrix} \alpha & g^T \\ g & H \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} = \begin{pmatrix} 1 \\ x \end{pmatrix} \lambda,$$

which is equivalent to

$$\alpha - \lambda = -g^T x \quad \text{and} \quad (H - \lambda I)x = -g \; . \tag{8}$$

If λ is the smallest eigenvalue of B_α and since the eigenvalues of H interlace the eigenvalues of B_α by the Cauchy interlace theorem (cf. [33]), then $H - \lambda I$ is positive semidefinite. Therefore, two of the optimality conditions in Lemma 3.1 are automatically satisfied in this case. If, in addition, $\lambda \le 0$ and $\|x\| = \Delta$, we will have a solution for the trust-region subproblem.

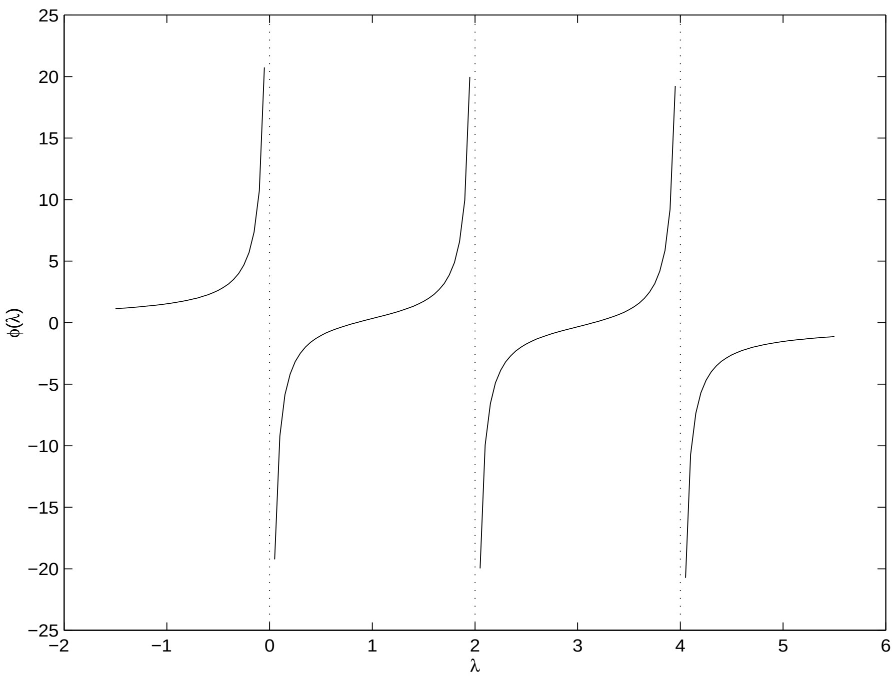

FIG. 2. *Secular function* $\phi(\lambda)$.

LSTRS consists of iteratively adjusting the parameter α to drive it towards the optimal value $\alpha_* = \lambda_* - g^T x_*$. This is accomplished in the following way. Let $\phi(\lambda) = -g^T x$ for x satisfying $(H - \lambda I)x = -g$, and note that $\phi'(\lambda) = x^T x$. Both ϕ and ϕ' are rational functions with poles at a subset of the eigenvalues of H. Figure 2 illustrates the typical behavior of $\phi(\lambda)$ when H is a 3×3 matrix with eigenvalues $0, 2, 4$. The LSTRS iteration is based on approximately solving the secular equation (6), using rational interpolation on ϕ and ϕ'. Observe that, in view of (8), we can obtain convenient interpolation points by solving eigenvalue problems for the smallest eigenvalue of B_α for different values of the parameter α. LSTRS computes the interpolation points in this way, using the implicitly restarted Lanczos method (IRLM) [37] as implemented in ARPACK [27] to solve the eigenvalue problems. The IRLM has fixed storage requirements and relies upon matrix-vector products only, features that make it suitable for large-scale problems.

The strategy described above works as long as the smallest eigenvalue of B_α has a corresponding eigenvector that can be safely normalized to have first component one. The adjustment of the parameter becomes very difficult in the hard case and near hard case since in these situations the smallest eigenvalue of B_α might not have a corresponding eigenvector with the desired structure (see [34], [35], [38]). Moreover, in the near hard case δ_1 is a weak pole of $\phi(\lambda)$ and the function becomes very steep around this value, as Figure 3 illustrates. This makes the interpolation problem very ill-conditioned. The situation is considerably more difficult for ill-posed problems where *several* of the smallest eigenvalues of H are weak poles of $\phi(\lambda)$. We illustrate this case in Figure 4 which shows $\phi(\lambda)$ for the test problem to be discussed in section 5.3. LSTRS relies on the complete characterization of the hard case given in [34] to proceed with the iteration even when the desired eigenvector cannot be normalized

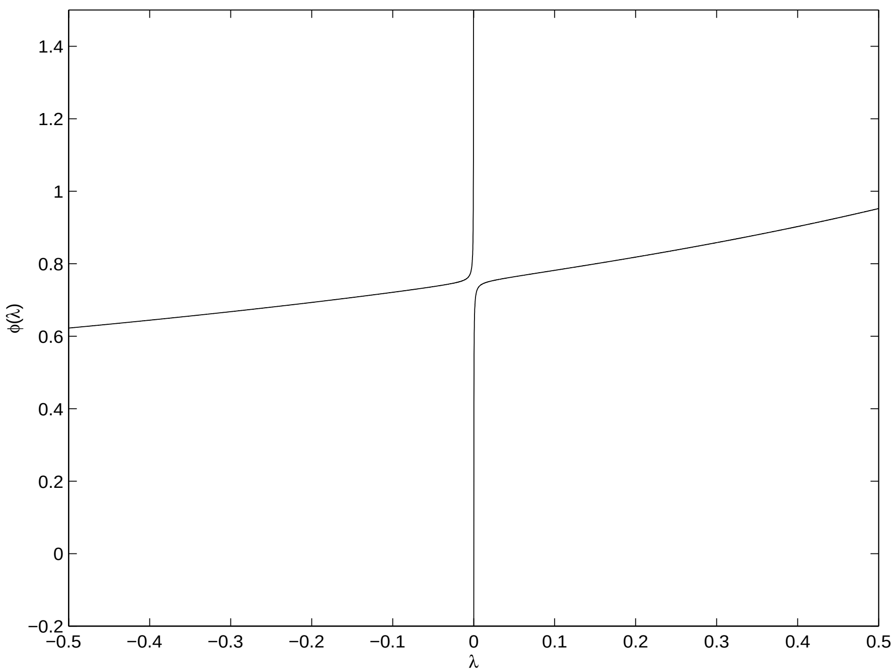

FIG. 3. $\phi(\lambda)$ *in the near hard case.*

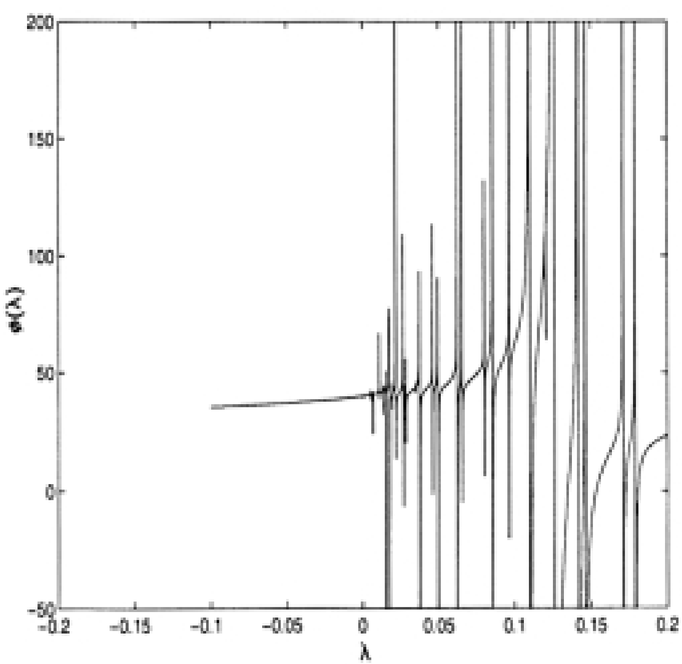

FIG. 4. $\phi(\lambda)$ *for the viscoacoustic wave equation,* $n = 500$.

to have first component one, and to compute nearly optimal solutions in any instance of the hard case including *multiple* occurrences as in ill-posed problems.

A detailed description of the results concerning the hard case and the elaborate algorithmic techniques derived from those results are beyond the scope of this paper. We refer the reader to [34] and [35] for more details.

From the above presentation we see that LSTRS has desirable features for solving large-scale trust-region subproblems in general, and for handling discrete ill-posed problems in particular. This is not surprising since the regularization of discrete forms of ill-posed problems was part of the motivation for developing the method. There are, however, some issues that must be taken into account when implementing LSTRS to treat ill-posed problems. As we saw in section 2, for these problems the smallest eigenvalues of H are clustered and close to zero, and because of the interlacing property the smallest eigenvalues of B_α will also be clustered and small for certain values of Δ. Computing a clustered set of small eigenvalues with a method that relies only on matrix-vector products with the original matrix is likely to fail since the multiplication will annihilate components precisely in the direction of the eigenvectors of interest. This difficulty may be overcome through the use of a spectral transformation. Instead of trying to find the smallest eigenvalue of B_α directly, we work with a matrix function $T(B_\alpha)$ and use the fact that $B_\alpha q = q\lambda \iff T(B_\alpha)q = qT(\lambda)$. If we are able to construct T so that $|T(\lambda_1)| \gg |T(\lambda_j)|,\ j > 1$, then a Lanczos-type method such as the IRLM will converge much faster towards the eigenvector q_1 corresponding to λ_1. We use a Chebyshev polynomial T_ℓ of degree ℓ constructed to be as large as possible on λ_1 and as small as possible on an interval containing the remaining eigenvalues of B_α. Convergence of IRLM is often greatly enhanced through this spectral transformation strategy. After convergence, the eigenvalues of B_α are recovered via the Rayleigh quotients with the converged eigenvectors.

Finally, the occurrence of an interior solution when $H = A^T A$ is positive definite in regularization problems deserves a special comment. In this case the solution of the trust-region subproblem corresponds to the least squares solution of the original problem. This solution is contaminated by noise and is of no interest. When we detect an interior solution we have taken the simple approach of reducing the trust-region radius and restarting the method. It is worth noting that if we knew that the noise level in the data was low, then if λ is close to zero when we detect an interior solution, we could approximate the least squares solution by x satisfying (8) since this would be a reasonable approximation to $x = -H^{-1}g$. Note that in this case it would not be necessary to solve a linear system to obtain the solution.

5. Numerical results. In this section we present numerical experiments to illustrate the performance of LSTRS on regularization problems from different sources, including both test problems and real applications. We used a MATLAB version of LSTRS running under MATLAB 5.3 using Mexfile interfaces to access ARPACK [27] and also the routines to compute matrix-vector products in some of the examples. Notice that the capabilities of ARPACK have been incorporated into MATLAB 6 and are now available through the routine `eigs`. We ran our experiments on a SUN Ultrasparc 2 with a 200 MHz processor and 256 Megabytes of RAM running Solaris 5.6. The floating point arithmetic was IEEE standard double precision with machine precision $2^{-52} \approx 2.2204 \cdot 10^{-16}$.

We present three sets of experiments. In section 5.1 we describe the results obtained on test problems from the Regularization Tools package [19]. In section 5.2 we present an inverse interpolation problem with field data. In section 5.3 we present

Table 1
Results of LSTRS on test problems from the Regularization Tools package.

Problem	Dim.	Δ	$\|x\|$	$\frac{\|x-x_{IP}\|}{\|x_{IP}\|}$	MV Prods.	Iter.
Ill **heat**	300	4.2631	4.2527	3.5684e-01	1721	8
Ill **heat**	1000	7.7829	7.7497	2.6900e-01	967	8
Well **heat**	300	4.2631	4.2958	9.1853e-02	1049	4
ilaplace	195	2.7629	2.7362	1.8537e-01	349	4
parallax	300	5.0000	5.0421	–	869	10
phillips	300	2.9999	2.9869	2.6883e-02	521	6
phillips	1000	3.0000	2.9839	3.3607e-02	575	6
shaw	300	17.2893	17.2467	6.0625e-02	510	6
shaw	1000	31.5659	31.6002	5.2847e-02	423	5

a model seismic inversion problem using a standard data set. The various stopping tolerances on $\frac{|\|x\|-\Delta|}{\Delta}$ were chosen (as they often are in practice) in an ad hoc fashion after some trial runs.

5.1. Problems from the Regularization Tools package. In this section we will present the results of LSTRS on problems from the Regularization Tools package [19]. This package consists of a set of MATLAB routines for the analysis of discrete ill-posed problems along with test problems that are easy to generate. All the test problems come from the discretization of a Fredholm integral equation of the first kind

$$\int_a^b K(s,t)f(t)dt = g(s),$$

and the problem is to compute the unknown function $f(t)$ given $g(s)$ and $K(s,t)$.

In all cases we solved a quadratically constrained least squares problem (4) where A came from the discretization of the kernel $K(s,t)$ and $b = g(s_i)$ at discrete points $s_i \in [a,b]$, where $i = 1,\ldots,n$ and n is the dimension of the problem. For some of the problems the exact solution $f(t)$ was available and in those cases we used $x_{IP} = f(t_i)$ for comparison purposes, where $t_i \in [a,b]$, $i = 1,\ldots,n$. Note that in general Ax_{IP} is different from b. Unless otherwise specified, we used $\Delta = \|x_{IP}\|$ as trust-region radius. In ARPACK, we used nine Lanczos basis vectors with seven shifts on each implicit restart. The required accuracy for the eigenpairs was 10^{-2}. The initial vector for the Lanczos factorization was a randomly generated vector that remained fixed in all the experiments. We solved the trust-region subproblems to a relative accuracy of $|\frac{\|x\|-\Delta}{\Delta}| < 10^{-2}$. We also solved the problems to a higher accuracy but this was computationally more expensive and did not seem to improve the accuracy of the regularized solution x with respect to the exact solution x_{IP} for this particular set of problems. In Table 1, we present the results for a subset of problems from [19].

Several observations are in order concerning Table 1. The third and fourth columns indicate that in all cases LSTRS solved the trust-region subproblem to the prescribed accuracy. The quality of the regularized solution or a measure of how well this solution approximates the exact solution x_{IP} is reported in the fifth column, where a dash indicates that x_{IP} was not available. We see that, generally, there is a reasonable agreement between computed and exact solutions, with relative errors of order 10^{-2}. The number of matrix-vector products is reported in column six, and the last column shows the number of LSTRS iterations.

The primary purpose of these tests was simply to verify that our approach would compute reasonably accurate regularized solutions to well-known examples of discrete forms of ill-posed problems. We cannot draw any conclusions about computational cost or expected number of matrix-vector products from these small examples. However, our limited experience would indicate that the number of matrix-vector products does not increase significantly with the dimension of the problem, and we give examples of this in sections 5.2 and 5.3.

Finally, we remark that some modifications probably would have been possible to improve the accuracy of the computed solutions to this set of examples. For problems ill-conditioned **heat** (inverse heat equation) and **ilaplace** (inverse Laplace transformation), the relative error is probably higher than one would like. It turns out that in these cases the solutions are highly oscillatory. This suggests that we should have solved the trust-region subproblem with a constraint of the form $\|Lx\|$, where L is a discrete form of first derivative. Since our goal here was a basic verification of LSTRS on such problems, we did not analyze each case separately, nor did we pursue more elaborate formulations.

5.2. An inverse interpolation problem. The two-dimensional (2-D) linear interpolation problem consists of using a linear interpolant to find the values of a function at arbitrary points given the values of the function at equally spaced points. A more interesting problem is the inverse interpolation problem: finding the values of the function on a regular grid of points from which we can extract given values of the function at irregularly spaced points by linear interpolation. We can pose the 2-D inverse interpolation problem as a least squares problem,

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|,$$

where $A \in \mathbb{R}^{m\times n}$ is the 2-D linear interpolant and $b \in \mathbb{R}^m$ contains the function values at irregularly spaced points.

To illustrate the performance of LSTRS on this kind of problem we will use the example of constructing a depth map of the Sea of Galilee on a regular grid of points, given depth measurements at irregularly spaced points. The data consists of triplets $v_i,\ w_i,\ b_i, i = 1,\dots,132044$, representing coordinates on the plane and depth, respectively. The data was collected from a ship using an echo sounder. The data contains noise coming from different sources, including malfunctioning equipment that reported zero depths at points in the middle of the lake and the fact that the measurements were taken at different times of the year and therefore varied greatly from rainy season to dry season. See [1] for a complete description of the data acquisition process. In Figure 5 we show a view from above of a three-dimensional (3-D) plot of the original data. The straight lines we observe in the figure are the tracks of the ship. Therefore, the data acquisition process was an additional source of noise. As Clærbout points out [9], an image of the sea should not include those lines.

In our experiments the size of the grid was $n = 201 \times 201 = 40401$, which is also the number of unknowns when the 2-D grid is represented as a one-dimensional vector. The number of rows in A was $m = 132044$. This matrix was ill-conditioned and was not available explicitly, but we could compute the action of A and A^T on vectors by means of FORTRAN routines. In all the experiments, we solved the trust-region subproblems to a relative accuracy of $|\frac{\|x\|-\Delta}{\Delta}| \le 10^{-3}$. The size of the Lanczos basis was five, and we applied three shifts on each implicit restart. Therefore, the storage requirement was essentially five vectors of length 40401.

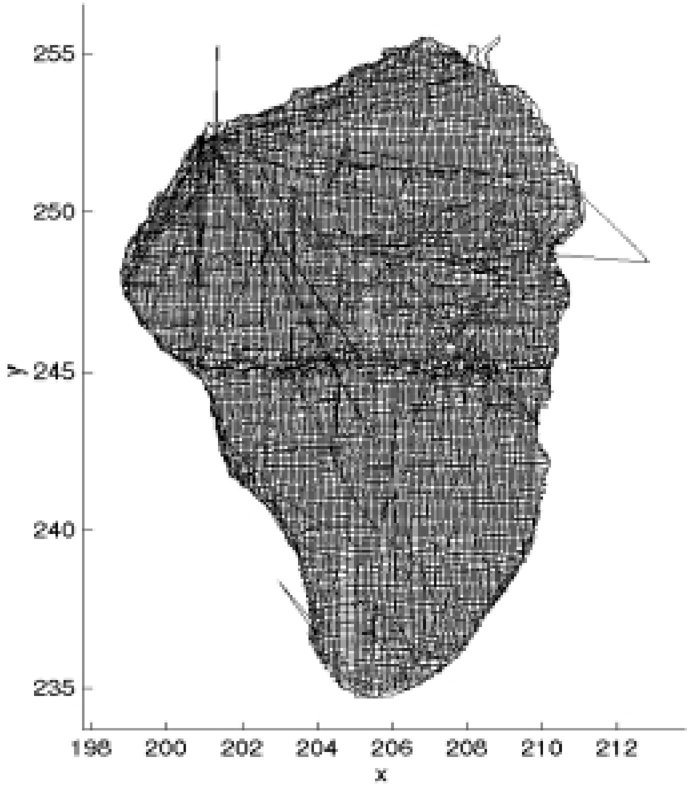

FIG. 5. *Sea of Galilee from original data.*

We posed the trust-region subproblem as

$$\min_{s.t.\|Lx\|\leq\Delta} \frac{1}{2}x^TA^TAx - (A^Tb)^Tx,$$

where L was either the $n \times n$ identity matrix I, or the matrix M_p, a discretization of the following pth power of the (scaled) 2-D Helmholtz operator

$$(\mathcal{I} - \nabla^T * \text{diag}(s) * \nabla)^p, \tag{9}$$

where $\mathcal{I}$, ∇, and ∇^T are the identity, gradient, and divergence operators, respectively; the vector $s > 0$ is a 2-D vector of scales; the expression $\text{diag}(s)$ denotes a diagonal matrix with the components of the vector s on the diagonal; and p is a real scalar.

We ran several experiments for different trust-region radii since in this application we did not have a priori information about the size or smoothness of the desired solution. Figure 6 shows the result for $\Delta = 6000$ and $L = I$. This image still shows the tracks of the ship and does not reveal any known features of the depth distribution of the lake. The contour plot (not shown) is very rough with highly oscillatory contours. This suggested the need to introduce a constraint on the smoothness of the solution.

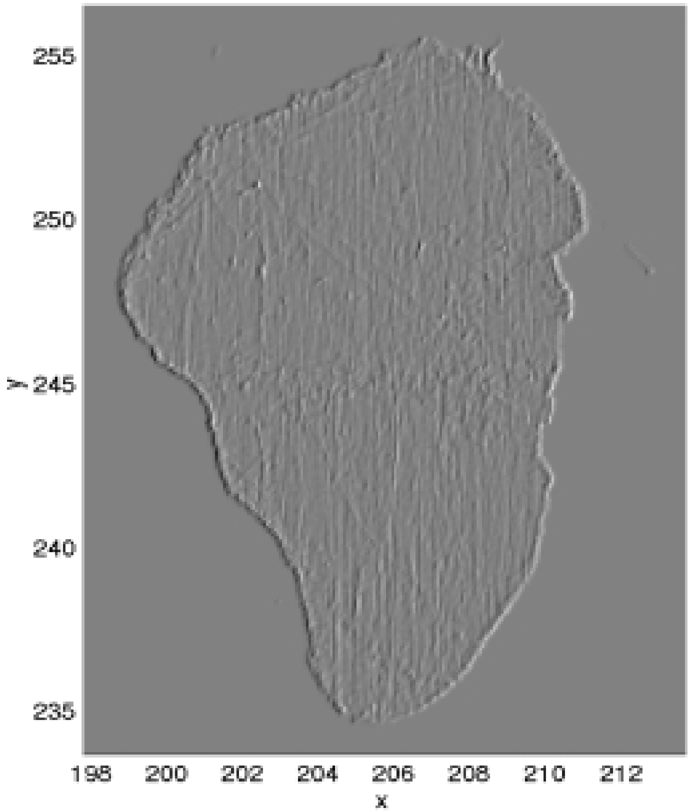

FIG. 6. *Sea of Galilee, regularizing with constraint on size of solution.*

We then solved the trust-region subproblem with $L = M_p$, a discretization of (9), and $\Delta = 26000$. Figure 7 shows the solution for $p = 0.3$. In this image we were able to identify some of the features reported in [1], such as the shallow areas in the northeast, a scarp in the southeast, and a more prominent scarp in the southwest.

We also tried the approach of solving the trust-region subproblem with $L = I$ and applying the Helmholtz operator a posteriori. We call this approach *postsmoothing*. We tried the postsmoothing approach with $p = -1$ for $\Delta = 23423$, which is the norm of x when $\|M_p x\| = 26000$, and we obtained an image very similar to the one in Figure 7. We also used this approach for $\Delta = 6000$, obtaining the image in Figure 8, which clearly shows the features mentioned before. Table 2 shows that the postsmoothing approach is less expensive than using a constraint on the smoothness. There are two reasons for this difference in efficiency. The first one is that the matrix-vector products when $L = M_p$ are more expensive than the matrix-vector products when $L = I$. The second reason is that the smallest eigenvalues of B_α are close to the smallest eigenvalues of $L^{-T} A^T A L^{-1}$, and these are more clustered for $L = M_p$ than for $L = I$. This causes slow convergence of the IRLM. We note, however, that the cost is not too high in either approach relative to the dimension of the problem.

The postsmoothing approach has the drawback that we do not know its physical meaning, and a closer look at the result shows that the postsmoothing is causing a high degree of perturbation on the depths since the lowest point is known to be around

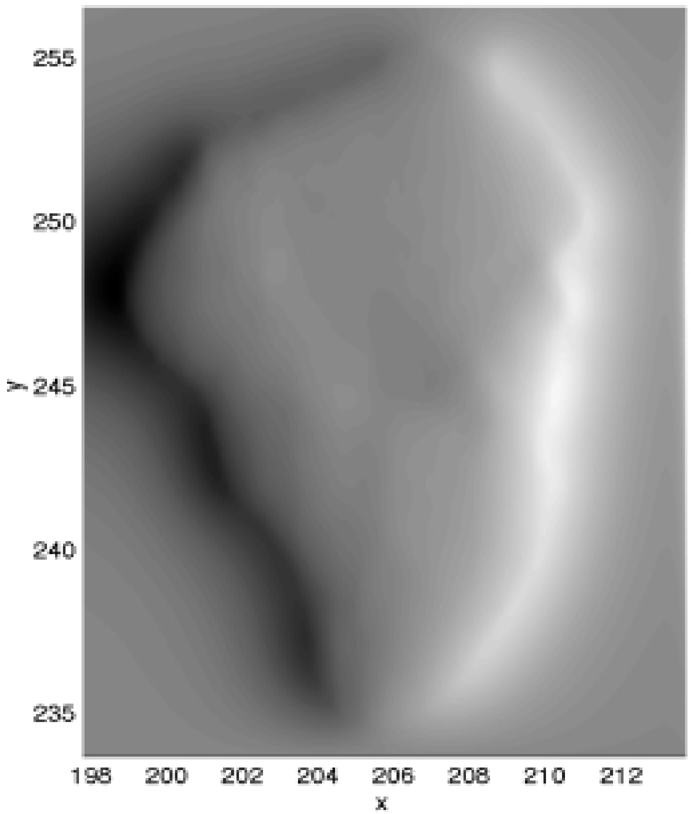

FIG. 7. *Sea of Galilee, regularizing with constraint on smoothness.*

256m below sea level, and the lowest point in Figure 8 is -45m. Another interesting aspect of this particular regularization approach is that it produces very smooth solutions, which was also noted in [21], and this causes, for example, the resulting regular grid to miss the true deepest point located approximately at coordinates $(207, 247)$ and yields a deepest point located at approximately $(205, 245)$ for Figure 8. It is quite remarkable, though, how the known features of the lake are clearly present in this image.

5.3. A model seismic inversion problem. We also solved the quadratically constrained least squares problem (4) where the problem comes from the discretization of the linear viscoacoustic model. As explained in [6], this model describes the behavior of an anelastic fluid, in which the strain response to a change of stress is linear but not completely instantaneous. A relaxation function $G(t,\vec{x})$ is used to express the stress-strain relation. The equations of motion relate $G(t,\vec{x})$, the material density $\rho(\vec{x})$, the pressure (stress) $p(t,\vec{x})$, the particle velocity $\vec{v}(t,\vec{x})$, and the body force source $f(t,\vec{x})$ in the following way:

$$
\begin{aligned}
p_{,t}(t,\vec{x}) &= -\dot{G}(t,\vec{x}) * \nabla \cdot \vec{v}(t,\vec{x}) + f(t,\vec{x}), \\
\vec{v}_{,t}(t,\vec{x}) &= -\frac{1}{\rho(\vec{x})}\nabla p(t,\vec{x}),
\end{aligned}
\tag{10}
$$

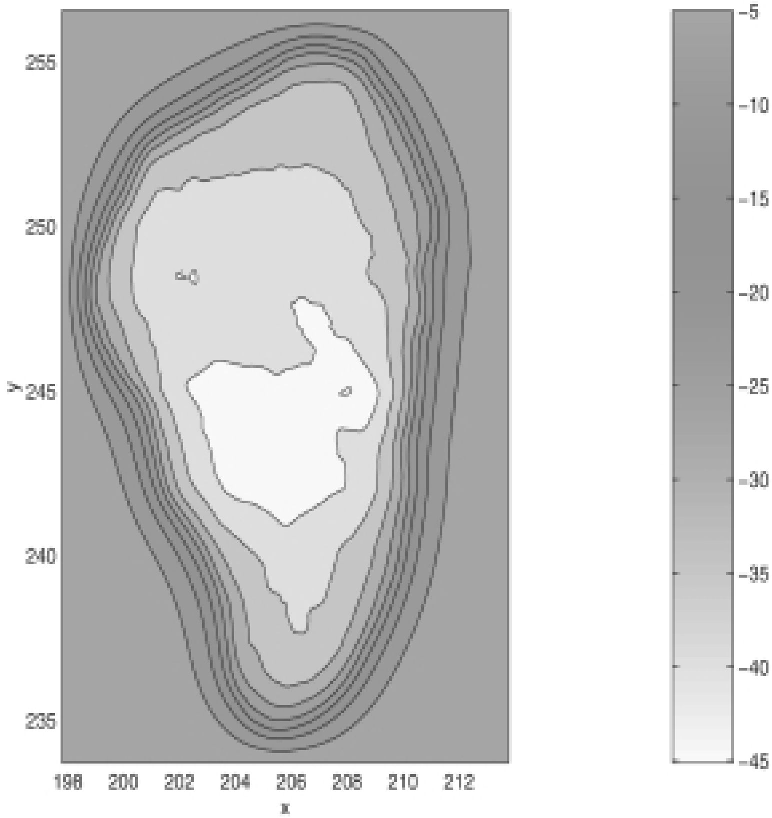

FIG. 8. *Sea of Galilee, regularizing with constraint on size of solution and postsmoothing.*

TABLE 2
Performance of LSTRS on an inverse interpolation problem.

Dimension: 40401 Storage: 5 vectors	Δ	$\|x_*\|$	LSTRS Iter.	MV Prods.	CPU time (min.)
TRS with postsmoothing	23423	23423	4	206	1.40
Constraint on smoothness	26000	25980.61	15	723	22.05

where $p = 0$, $\vec{x} = 0$ for $t \ll 0$. In (10), t denotes time and $\vec{x}$ denotes position.

These equations are used in [6] to model the propagation of waves in marine media using the relaxation function for a standard linear fluid. The matrix A in our quadratically constrained least squares problem corresponds to DF, a linearized version of the forward map or prediction operator. The matrix A^T corresponds to the adjoint of DF denoted by DF^*. The operators DF and DF^* were not explicitly available, but their action on vectors was obtained by solving a simplified (layered) and linearized version of (10). See [5, Chap. 5] and [6] for more details. The data vector $\bar{b}$ is a seismogram containing velocities of waves measured in oil wells in the North Sea. The data is part of the Mobil AVO data set [23], a standard data set for testing inversion methods. The parameters to be estimated in the experiment are the

TABLE 3
Performance of LSTRS on the viscoacoustic wave equation.

Dimension	Δ	$\|x\|$	LSTRS Iter.	MV Prods.	Storage
121121	0.5	0.5	2	15	4 vectors

stress-strain ratio under simple hydrostatic pressure and the material density. The quadratically constrained least squares problems arise in the context of sophisticated nonlinear inversion strategies [12], where the norm constraint was of the form $\|x\| \leq \Delta$. The quadratically constrained least squares problems were obtained from problems of type

$$\min_{s.t. \|Ax - \bar{b}\| \leq \rho} \|x\|$$

by means of a change of variable made possible by the availability of an effective approximation to the inverse of A. The parameter ρ is an estimate of the noise level in $\bar{b}$ and is known a priori in this case. The reformulation of the problem yields $\Delta = \rho$.

The dimensions of the problem are $m = n = 121121$. Table 3 shows the result obtained when we used LSTRS to solve the trust-region subproblem to an accuracy of $|\frac{\|x\|-\Delta}{\Delta}| < 10^{-5}$ using four Lanczos basis vectors. The method is very efficient for small Δ since in this case the smallest eigenvalue of B_α is well separated from the rest and the IRLM converges rapidly to such eigenvalue. For larger Δ, the eigenvalue of interest belongs to a cluster and the IRLM needs more iterations to compute it.

In Table 3 we can observe the low storage and low number of matrix-vectors products required to solve the problem to a very high accuracy.

6. Conclusions.

We considered the problem of regularizing large-scale discrete forms of ill-posed problems arising in several applications. We posed the regularization problem as a quadratically constrained least squares problem, showed the relationship of this approach to Tikhonov regularization and to the trust-region subproblem, and analyzed the latter in the ill-posed case.

We have presented numerical results obtained when we used the recently developed method LSTRS for the large-scale trust-region subproblem to solve regularization problems from a wide variety of applications including problems with field data. The method requires solving a sequence of large-scale eigenvalue problems, which is accomplished with a variant of the Lanczos method. An important feature of LSTRS is that it computes both the solution and the Tikhonov regularization parameter from the prescribed norm.

LSTRS is particularly suitable for large-scale discrete forms of ill-posed problems, for which it computed regularized solutions close to the desired exact solutions using limited storage and moderate computational effort in general. For real applications the method required a low number of matrix-vector products with respect to the dimension of the problem, and storage comparable to or less than the conjugate gradient method. Our approach also had the desirable feature of providing systematic stopping criteria.

We are currently investigating various approaches to preconditioning, aiming to generate eigenvalue problems that can be solved more efficiently. Although further improvement is needed, LSTRS proved to be a promising method for the numerical treatment of large-scale discrete forms of ill-posed problems in which the norm of the desired solution is prescribed.

Acknowledgments. We would like to thank Chao Yang for making his MATLAB code for the Chebyshev spectral transformation available to us. We thank Ricardo von Borries for his help in recovering what seemed to be a hopelessly lost illustration. We would also like to express our deep gratitude to Bill Symes who provided us with the data codes for the 2-D linear interpolant and interpretation of the results in section 5.2. Bill also provided the model seismic problem in section 5.3 and the AVO data set. We also thank Professor Zvi Ben-Avraham from Tel-Aviv University for making the data of the Sea of Galilee available and for providing important references. Finally, we would like to thank the two anonymous referees for their careful reading of the original manuscript, and for their valuable and relevant suggestions, which helped improve the original version of this paper.

REFERENCES

[1] Z. Ben-Avraham, G. Amit, A. Golan, and Z.B. Begin, *The bathymetry of Lake Kinneret and its structural significance*, Israel J. Earth Sci., 39 (1992), pp. 77–84.
[2] M. Bertero and P. Bocacci, *Introduction to Inverse Problems in Imaging*, Institute of Physics, Bristol, UK, 1998.
[3] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
[4] Å. Björck, E. Grimme, and P. Van Dooren, *An implicit shift bidiagonalization algorithm for ill-posed systems*, BIT, 34 (1994), pp. 510–534.
[5] J.O. Blanch, *A Study of Viscous Effects in Seismic Modeling, Imaging, and Inversion: Methodology, Computational Aspects, and Sensitivity*, Ph.D. thesis, Department of Geology and Geophysics, Rice University, Houston, 1995.
[6] J.O. Blanch, W.W. Symes, and R.J. Versteeg, *A numerical study of linear viscoacoustic inversion*, in Comparison of Seismic Inversion Methods on a Single Real Data Set, R.G. Keys and D.J. Foster, eds., Society of Exploration Geophysicists, Tulsa, OK, 1998, pp. 13–44.
[7] D. Calvetti, G.H. Golub, and L. Reichel, *Estimation of the L-curve via Lanczos bidiagonalization*, BIT, 39 (1999), pp. 603–619.
[8] D. Calvetti, L. Reichel, and Q. Zhang, *Iterative exponential filtering for large discrete ill-posed problems*, Numer. Math., 83 (1999), pp. 535–556.
[9] J. Clærbout, *Geophysical Estimation by Example: Environmental Soundings Image Enhancement: Multidimensional Autoregression*, available online from http://sepwww.stanford.edu/sep/jon/index.html.
[10] L. Eldén, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT, 17 (1977), pp. 134–145.
[11] D. M. Gay, *Computing optimal locally constrained steps*, SIAM J. Sci. Statist. Comput., 2 (1981), pp. 186–197.
[12] M.S. Gockenbach and W.W. Symes, *Duality for inverse problems in wave propagation*, in Large Scale Optimization, IMA Vol. Math. Appl. 92, L. Biegler, T. Coleman, F. Santosa, and A. Conn, eds., Springer-Verlag, New York, 1997, pp. 37–61.
[13] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, 1996.
[14] G.H. Golub and U. von Matt, *Quadratically constrained least squares and quadratic problems*, Numer. Math., 59 (1991), pp. 561–580.
[15] M. Hanke and P.C. Hansen, *Regularization methods for large-scale problems*, Surveys Math. Indust., 3 (1993), pp. 253–315.
[16] P.C. Hansen, *The discrete Picard condition for discrete ill-posed problems*, BIT, 30 (1990), pp. 658–672.
[17] P.C. Hansen, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev., 34 (1992), pp. 561–580.
[18] P.C. Hansen and D.P. O'Leary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput., 14 (1993), pp. 1487–1503.
[19] P.C. Hansen, *Regularization Tools: A MATLAB package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35.
[20] P.C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, Doctoral Dissertation, UNI•C, Technical University of Denmark, Lyngby, Denmark, 1996.
[21] P.C. Hansen, M. Jacobsen, J.M. Rasmussen, and H. Sørensen, *The PP-TSVD algorithm*

for image restoration problems, in Methods and Applications of Inversion, Lecture Notes in Earth Sci. 92, P.C. Hansen, B.H. Jacobsen, and K. Mosegaard, eds., Springer-Verlag, Berlin, 2000, pp. 171–186.

[22] L. Kaufman and A. Neumaier, *Image Reconstruction through Regularization by Envelope Guided Conjugate Gradients*, AT&T Bell Laboratories, 4-14, 1994; available online from http:cm.bells-labs.com/cm/cs/doc/94/4-14.ps.gz.

[23] R.G. Keys and D.J. Foster, *A data set for evaluating and comparing seismic inversion methods*, in Comparison of Seismic Inversion Methods on a Single Real Data Set, R.G. Keys and D.J. Foster, eds., Society of Exploration Geophysicists, Tulsa, OK, 1998, pp. 1–12.

[24] M.E. Kilmer and D.P. O'Leary, *Pivoted Cauchy-like preconditioners for regularized solution of ill-posed problems*, SIAM J. Sci. Comput., 21 (1999), pp. 88–110.

[25] M.E. Kilmer and D.P. O'Leary, *Choosing regularization parameters in iterative methods for ill-posed problems*, SIAM J. Matrix Anal. Appl., 22 (2001), pp. 1204–1221 .

[26] C.L. Lawson and R.J. Hanson, *Solving Least Squares Problems*, Classics Appl. Math. 15, SIAM, Philadelphia, 1995.

[27] R.B. Lehoucq, D.C. Sorensen, and C. Yang, *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems by Implicitly Restarted Arnoldi Methods*, SIAM, Philadelphia, 1998.

[28] J.J. Moré and D.C. Sorensen, *Computing a trust region step*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 553–572.

[29] J. Nagy, V. Pauca, R. Plemmons, and T. Torgersen, *Space-varying restoration of optical images*, Opt. Soc. Amer. A, 14 (1997), pp. 3162–3174.

[30] F. Natterer, *The Mathematics of Computerized Tomography*, John Wiley, New York, 1986.

[31] A. Neumaier, *Solving ill-conditioned and singular linear systems: A tutorial on regularization*, SIAM Rev., 40 (1998), pp. 636–666.

[32] G. Nolet, ed., *Seismic Tomography, with Applications in Global Seismology and Exploration Geophysics*, D. Reidel, Dordrecht, 1987.

[33] B.N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[34] M. Rojas, *A Large-Scale Trust-Region Approach to the Regularization of Discrete Ill-Posed Problems*, Ph.D. thesis, Technical Report TR98-19, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1998.

[35] 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.

[36] D.C. Sorensen, *Newton's method with a model trust region modification*, SIAM J. Numer. Anal., 19 (1982), pp. 409–426.

[37] D.C. Sorensen, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 357–385.

[38] D.C. Sorensen, *Minimization of a large-scale quadratic function subject to a spherical constraint*, SIAM J. Optim., 7 (1997), pp. 141–161.

[39] W.W. Symes, *A differential semblance criterion for inversion of multioffset seismic reflection data*, J. Geophysical Research, 98 (1993), pp. 2061–2073.

[40] A.N. Tikhonov, *Regularization of incorrectly posed problems*, Soviet Math., 4 (1963), pp. 1624–1627.