Algorithm 873: LSTRS: MATLAB Software for Large-Scale Trust-Region Subproblems and Regularization

Rojas Larrazabal, Marielba de la Caridad; Santos, Sandra A.; Sorensen, Danny C.

Published in:
ACM Transactions on Mathematical Software

Link to article, DOI:
10.1145/1326548.1326553

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Algorithm 873: LSTRS: MATLAB Software for Large-Scale Trust-Region Subproblems and Regularization

MARIELBA ROJAS
Technical University of Denmark
SANDRA A. SANTOS
State University of Campinas
and
DANNY C. SORENSEN
Rice University

A MATLAB 6.0 implementation of the LSTRS method is presented. LSTRS was described in Rojas et al. [2000]. LSTRS is designed for large-scale quadratic problems with one norm constraint. The method is based on a reformulation of the trust-region subproblem as a parameterized eigenvalue problem, and consists of an iterative procedure that finds the optimal value for the parameter. The adjustment of the parameter requires the solution of a large-scale eigenvalue problem at each step. LSTRS relies on matrix-vector products only and has low and fixed storage requirements, features that make it suitable for large-scale computations. In the MATLAB implementation, the Hessian matrix of the quadratic objective function can be specified either explicitly, or in the form of a matrix-vector multiplication routine. Therefore, the implementation preserves the matrix-free nature of the method. A description of the LSTRS method and of the MATLAB software, version 1.2, is presented. Comparisons with other techniques and applications of the method are also included. A guide for using the software and examples are provided.

Categories and Subject Descriptors: G.4 [**Mathematical Software**]—*Algorithm design and analysis*

General Terms: Algorithms, Design

M. Rojas was supported in part by NSF cooperative agreement CCR-9120008, the Research Council of Norway, and the Science Research Fund of Wake Forest University. S. A. Santos was supported by FAPESP (06/53768-0), CNPq (302412/2004-2), FINEP, and FAEP-UNICAMP. D. C. Sorensen was supported in part by NSF Grants CCR-0306503, ACI-0325081, CCF-0634902.
Authors' addresses: M. Rojas, Department of Informatics and Mathematical Modelling, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark; email: mr@imm.dtu.dk; S. A. Santos, Department of Applied Mathematics, State University of Campinas, SP, Brazil; email: sandra@ime.unicamp.br; D. C. Sorensen, Department of Computational and Applied Mathematics, Rice University, 6100 Main St., Houston, TX 77005-1892; email: sorensen@caam.rice.edu.

Additional Key Words and Phrases: ARPACK, constrained quadratic optimization, Lanczos method, regularization, MATLAB, trust-region

ACM Reference Format:

Rojas, M., Santos, S. A., and Sorensen, D. C. 2008. Algorithm 873: LSTRS: MATLAB software for large-scale trust-region subproblems and regularization. ACM Trans. Math. Softw., 34, 2, Article 11 (March 2008), 28 pages. DOI = 10.1145/1326548.1326553 http://doi.acm.org/10.1145/1326548.1326553

1. INTRODUCTION

We describe version 1.2 of a MATLAB 6.0 implementation of the LSTRS method [Rojas et al. 2000] for large-scale quadratic problems with a quadratic constraint, or trust-region subproblems:

$$\min \quad \frac{1}{2}x^T Hx + g^T x \quad \text{subject to (s.t.)} \quad \|x\| \leq \Delta, \tag{1}$$

where H is an $n \times n$, real, symmetric matrix, g is an n-dimensional real vector, and Δ is a positive scalar. In (1), and throughout the article, $\|\cdot\|$ denotes the Euclidean norm. The following notation is also used throughout the article: δ_1 denotes the algebraically smallest eigenvalue of H, $\mathcal{S}_1 \equiv \mathcal{N}(H - \delta_1 I)$ denotes the corresponding eigenspace, $\mathcal{N}(\cdot)$ denotes the nullspace of a matrix, and $\dagger$ denotes the pseudoinverse.

Problem (1) arises in connection with the trust-region globalization strategy in optimization. A special case of problem (1), namely, a least squares problem with a norm constraint, is equivalent to Tikhonov regularization [Tikhonov 1963] for discrete forms of ill-posed problems. The Lagrange multiplier associated with the constraint is the Levenberg-Marquardt parameter in optimization and the Tikhonov parameter in regularization. A constraint of the form $\|Cx\| \leq \Delta$ for a matrix $C \neq I$ is not considered in this work. The matrix C can be used, for example, as a scaling matrix in optimization or to impose a smoothness condition on the solution in regularization. Note that when C is nonsingular, a change of variables can be used to reduce the problem to the case we are considering.

The trust-region subproblem has very interesting theoretical properties that lead to the design of efficient solution methods. In particular, if it is possible to compute the Cholesky factorization of matrices of the form $H - \lambda I$, the method of choice is probably the one proposed by Moré and Sorensen [1983]. The algorithm uses Newton's method to find a root of a scalar function that is almost linear on the interval of interest. The authors also proposed a computationally-efficient strategy for dealing with a special and usually difficult case, known since then in the optimization literature as the *hard case*. The hard case is discussed in detail in Section 2.

If the matrix H is very large or not explicitly available, factoring or even forming the matrices $H - \lambda I$ may be prohibitive and a different approach is needed to solve the problem. Possibly, the most popular method for the large-scale trust-region subproblem is the one of Steihaug [1983] and Toint [1981]. The method computes the solution to the problem in a Krylov space and is

efficient in conjunction with optimization methods. An improvement upon the Steihaug-Toint approach, based on the truncated Lanczos idea, was proposed by Gould et al. [1999]. Hager [2001] adopts an SQP approach to solve the trust-region subproblem in a special Krylov subspace. New properties of the trust-region subproblem that provide useful tools for the development of new classes of algorithms in the large-scale scenario are presented by Lucidi et al. [1998]. Other authors that have considered large-scale problems are Golub and von Matt [1991], Sorensen [1997], Rendl and Wolkowicz [1997] (revisited by Fortin and Wolkowicz [2004]), Rojas et al. [2000] and Pham Dinh and Le Thi [1998]. The theory of Gauss quadrature, matrix moments and Lanczos diagonalization is used in Golub and von Matt [1991] to compute bounds for the optimal Lagrange multiplier and solution. The hard case is not analyzed in Golub and von Matt [1991]. The algorithm in Pham-Dinh and Le-Thi [1998] is based on differences of convex functions, and is inexpensive due to its projective nature. However, a restarting mechanism is needed in order to guarantee convergence to a global solution. The approaches in Rendl and Wolkowicz [1997], Rojas et al. [2000], and Sorensen [1997] recast the trust-region subproblem as a parameterized eigenvalue problem and design an iteration to find an optimal value for the parameter. A primal-dual semidefinite framework is proposed in Rendl and Wolkowicz [1997], with a dual simplex-type method for the basic iteration and a primal simplex-type method for the hard-case iteration. In Rojas et al. [2000] and Sorensen [1997], two different rational interpolation schemes are used for the update of the scalar parameter. In Sorensen, a superlinearly-convergent scheme is developed for the adjustment of the parameter, as long as the hard case does not occur. In the presence of the hard case, the algorithm in Sorensen [1997] is linearly convergent. In Rojas et al. [2000], a unified iterative scheme is proposed which converges superlinearly in all cases.

It is possible to classify methods for the trust-region subproblem based on the properties of the computed solution. We will call an approximation to an optimal solution of problem (1) (see Section 2.1), a *nearly-exact solution*; and any other approximation, an *approximate solution*. Accordingly, we can make a distinction between nearly-exact methods and approximate methods. The methods in Fortin and Wolkowicz [2004], Golub and von Matt [1991], Moré and Sorensen [1983], Pham-Dinh and Le-Thi [1998], Rendl and Wolkowicz [1997], Rojas et al. [2000], and Sorensen [1997] are nearly exact, while the methods in Gould et al. [1999], Hager [2001], Steihaug [1983], and Toint [1981] are approximate. Approximate solutions (and methods) are of particular interest in the context of trust-region methods for optimization. In regularization, nearly exact solutions are often required.

In this article, we describe a set of MATLAB 6.0 routines implementing the nearly-exact method LSTRS from Rojas et al. [2000]. LSTRS is suitable for large-scale computations since it relies on matrix-vector products only and has low and fixed storage requirements. As mentioned above, LSTRS is based on a reformulation of problem (1) as a parameterized eigenvalue problem. The goal of the method is to compute the optimal value for a scalar parameter, which is then used to compute a solution for problem (1). The method requires the solution of an eigenvalue problem at each step. LSTRS can handle all instances

of the trust-region subproblem, including those arising in the regularization of ill-posed problems. The method has been successfully used for computing regularized solutions of large-scale inverse problems in several areas [Eldén et al. 2005; Rojas 1998; Rojas and Sorensen 2002; Rojas and Steihaug 2002].

The MATLAB implementation of LSTRS described in this article allows the user to specify the matrix H explicitly, a feature that can be useful for small test problems, and implicitly, in the form of a matrix-vector multiplication routine; hence preserving the matrix-free nature of the original method. Several options are available for the solution of the eigenvalue problems, namely: the MATLAB routine `eig` (QR method), a slightly modified version of `eigs` (a MEX-file interface for ARPACK [Lehoucq et al. 1998]) a combination of `eigs` with a Tchebyshev Spectral Transformation as in Rojas and Sorensen [2002], or a user-provided routine.

The article is organized as follows. In Section 2, we present the properties of the trust-region subproblem and its connection with regularization. In Section 3, we describe the method LSTRS from Rojas et al. [2000]. We discuss the use of the software for regularization problems in Section 4. In Section 5, we present comparisons of LSTRS with other methods for the large-scale trust-region subproblem. In Section 6, we discuss the use of LSTRS on large-scale problems and present an application from image restoration. We present concluding remarks in Section 7.

2. TRUST REGIONS AND REGULARIZATION

In this section, we describe the trust-region subproblem as well as its connection with the regularization of discrete forms of ill-posed problems. We present the properties of the trust-region subproblem in Section 2.1 and discuss regularization issues in Section 2.2.

2.1 The Structure of the Trust-Region Subproblem

The trust-region subproblem always has a solution that lies either in the interior or on the boundary of the feasible set $\{x \in \mathbb{R}^n,\ \|x\| \leq \Delta\}$. A characterization of the solutions of problem (1), found independently by Gay [1981] and Sorensen [1982], is given in the following lemma where we have followed Sorensen [1997] in the nonstandard but notationally more convenient use of a nonpositive multiplier.

LEMMA 2.1 [SORENSEN 1982]. *A feasible vector $x_* \in \mathbb{R}^n$ is a solution to (1) with corresponding Lagrange multiplier λ_* if and only if x_*, λ_* satisfy $(H - \lambda_* I)x_* = -g$ with $H - \lambda_* I$ positive semidefinite, $\lambda_* \leq 0$ and $\lambda_*(\Delta - \|x_*\|) = 0$.*

PROOF. For the proof see Sorensen [1982]. □

Lemma 2.1 implies that all solutions to the trust-region subproblem are of the form $x = -(H - \lambda I)^{\dagger} g + z$, for some $z \in \mathcal{N}(H - \lambda I)$. If the Hessian matrix H is positive definite and $\|H^{-1} g\| < \Delta$, problem (1) has a unique interior solution given by $x = -H^{-1} g$, with Lagrange multiplier $\lambda = 0$. If the Hessian is positive semidefinite or indefinite, there exist boundary solutions satisfying $\|x\| = \Delta$ with $\lambda \leq \delta_1$.

The case $\lambda = \delta_1$ is usually called the *hard case* in the literature [Moré and Sorensen 1983]. The hard case can only occur when $\delta_1 \leq 0$, $g \perp \mathcal{S}_1$ and $\|(H - \delta_1 I)^{\dagger} g\| \leq \Delta$. For most problems of interest, solving the trust-region problem in the hard case can be an expensive and difficult task since it requires the computation of 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 difficulties in this case. As in Rojas [1998] and Rojas and Sorensen [2002], 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, depending on the value of Δ. Therefore we call this situation a *potential hard case*.

The occurrence of the exact, near, or potential hard case is structural: it depends on the relationships among the matrix H, the vector g, and the scalar Δ. Although not too common in optimization, the near hard case is rather frequent in regularization. Indeed, it was shown in Rojas [1998] and Rojas and Sorensen [2002] that the potential hard case is precisely the common case for discrete forms of ill-posed problems, where it occurs in a multiple instance in which the vector g is orthogonal or nearly orthogonal to several eigenspaces of H. We discuss these issues in Section 2.2.

2.2 The Trust-Region Approach to Regularization

In this section, we first describe the properties of discrete forms of ill-posed problems and show how they lead to the use of regularization. We then discuss the connection of trust regions and regularization. Finally, we describe the properties of the trust-region subproblem in the regularization context.

Discrete forms of linear ill-posed problems consist of linear systems or linear least squares problems in which the coefficient matrices come from the discretization of the continuous operator in an ill-posed problem and the right-hand side contains experimental data contaminated by noise. The discretization of continuous problems in inversion [Bertero and Bocacci 1998; Natterer 1986; Nolet 1987; Symes 1993] usually leads to highly ill-conditioned problems, called discrete forms of ill-posed problems, or discrete ill-posed problems, in the literature. Reasonably accurate discretizations will produce coefficient matrices whose properties are the discrete analogs of those of the continuous operators. In particular, the matrices will be highly ill-conditioned with singular spectra that decay to zero gradually with no particular gap, and will have a large cluster of very small singular values Hansen [1998]. Moreover, as observed in Hansen [1998], the high-frequency components (those with more sign changes) of the singular vectors will usually correspond to the smallest singular values.

We consider the problem of recovering x_{LS}, the minimum-norm solution to:

$$\begin{aligned} &\min \quad \|Ax - b\| \\ &x \in \mathbb{R}^n, \end{aligned}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $m \geq n$, when the *exact* data vector b is not known, and instead, 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 noise.

Considering that only $\bar{b}$ is available, we could try to approximate x_{LS} by $\bar{x}_{LS}$, the minimum-norm solution to:

$$\begin{aligned} &\min \quad \|Ax - \bar{b}\|. \\ &x \in \mathbb{R}^n. \end{aligned} \tag{2}$$

Unfortunately, as we now show, the two solutions might differ considerably.

Let $A = U\Sigma V^T$ be a Singular Value Decomposition (SVD) of A, where $U \in \mathbb{R}^{m\times n}$ has orthormal columns u_i, $V \in \mathbb{R}^{n\times n}$ is orthogonal with columns v_i, and Σ is a diagonal matrix with elements $\sigma_1, \sigma_2, \ldots, \sigma_n$. The σ_i's are the singular values of A in nonincreasing order. The solution of problem (2) in terms of the SVD of A is given by:

$$\bar{x}_{LS} = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i + \sum_{i=1}^{n} \frac{u_i^T s}{\sigma_i} v_i. \tag{3}$$

As usual in the analysis of discrete forms of ill-posed problems, we assume that the Discrete Picard Condition (DPC) [Hansen 1990] holds: that the values $|u_i^T b|$ overall decay to zero faster than σ_i as the index i increases. Assuming that the DPC holds, the first term in the right-hand side of (3) is bounded. However, the second term might become very large since the expansion coefficients of the uncorrelated noise vector $(u_i^T s)$ remain constant while the singular values decay to zero. Therefore, the components of $\bar{x}_{LS}$ corresponding to small singular values are magnified by the noise and $\bar{x}_{LS}$ might be dominated by the high-frequency components. Consequently, standard methods such as those in Björck [1996], Golub and Van Loan [1996, Ch. 5], and Lawson and Hanson [1995] applied to problem (2) usually produce meaningless solutions with very large norms. Note that even in the noise-free case, the ill conditioning of the matrix A will pose difficulties to most numerical methods. Therefore, to solve these problems, a special technique known as *regularization* is needed.

In regularization, we aim to recover an approximation to 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 conditioning of the new problem depends on the choice of a special parameter known as the *regularization parameter*. Excellent surveys on regularization methods can be found for example in Hanke and Hansen [1993], Hansen [1998], and Neumaier [1998].

One of the most popular regularization approaches is Tikhonov regularization [Tikhonov 1963], which consists of adding a penalty term to problem (2) to obtain:

$$\min_{x \in \mathbb{R}^n} \quad \|Ax - \bar{b}\|^2 + \mu\|x\|^2, \tag{4}$$

where $\mu > 0$ is the Tikhonov regularization parameter. It is well known [Eldén 1977; Rojas and Sorensen 2002] that this approach is equivalent to a special instance of the trust-region subproblem, namely, to a least squares problem

with a quadratic constraint:

$$\min \ \|Ax - \bar{b}\|^2 \ \text{ s.t. } \ \|x\| \leq \Delta, \tag{5}$$

where $H = A^T A$ and $g = -A^T \bar{b}$. Therefore, in principle, methods for the trust-region subproblem could be used to solve regularization problems of type (5), where instead of specifying a value for the Tikhonov parameter as required for (4), we need to prescribe a bound on the norm of the desired solution. However, as we shall see, the trust-region subproblem (5) has special properties in the regularization context and these properties should be taken into consideration when developing solution methods. The following analysis is based on Rojas [1998] and Rojas and Sorensen [2002].

We now show that the potential (near) hard case is the common case for ill-posed problems, where it occurs in a multiple instance, with g nearly orthogonal to the eigenspaces associated with several of the smallest eigenvalues of H. This was first shown in Rojas [1998]. Assume that the singular values of A are not zero and that σ_n, the smallest singular value, has multiplicity k. Let $n - k + 1 \leq i \leq n$ and let v_i be a right-singular vector of A associated with σ_n. Then:

$$g^T v_i \ = \ -\bar{b}^T U \Sigma V^T v_i = -\sigma_n u_i^T \bar{b} = -\sigma_n \big(u_i^T b + u_i^T s\big).$$

If there is no noise in the data ($s = 0$) and if the DPC holds, the coefficients $u_i^T b$, for $n - k + 1 \leq i \leq n$, are small and since σ_n is also small, it follows that g is nearly orthogonal to v_i in this case. For noisy data, $g^T v_i$ might not be small due to the possible contribution of the term $u_i^T s$. However, for severely ill-conditioned problems, the smallest singular value σ_n is so close to zero that even if $u_i^T s$ is large, g will still be nearly orthogonal to v_i. Since v_i is an eigenvector corresponding to $\delta_1 = {\sigma_n}^2$, the smallest eigenvalue of $A^T A$, we have that g will be nearly orthogonal to the eigenspace corresponding to the smallest eigenvalue and therefore, the potential (near) hard case will occur.

Observe that in ill-posed problems, the matrix A usually has a large cluster of singular values very close to zero. Therefore, following the previous argument, we see that the vector g will be orthogonal, or nearly orthogonal, to the eigenspaces corresponding to several of the smallest eigenvalues of the matrix $A^T A$, and the potential hard case will occur in multiple instances. The numerical experimentation presented in Rojas [1998] and Rojas and Sorensen [2002] indicates that the LSTRS algorithm can efficiently handle the multiple instances of orthogonality (or near orthogonality) based on the complete characterization of the hard case given in Rojas [1998].

3. THE LSTRS METHOD

In this section, we present a description of the LSTRS method with special emphasis on the computational aspects. For more details, as well as for the theoretical foundations and the convergence properties of the method, we refer the reader to Rojas [1998] and Rojas et al. [2000].

LSTRS is based on a reformulation of the trust-region subproblem (1) as a parameterized eigenvalue problem. The new formulation is based on the fact that

there exists a value of a scalar parameter α such that problem (1) is equivalent to:

$$\begin{array}{ll} \min & \frac{1}{2} y^T B_\alpha y \\ \text{s.t.} & y^T y \leq 1 + \Delta^2, \quad e_1^T y = 1, \end{array} \tag{6}$$

where B_α is the *bordered matrix* $B_\alpha = \left(\begin{smallmatrix} \alpha & g^T \\ g & H \end{smallmatrix}\right)$, and e_1 is the first canonical vector in $\mathbb{R}^{n+1}$. The optimal value for α is given by $\alpha_* = \lambda_* - g^T x_*$, with λ_*, x_* the optimal pair in Lemma 2.1. Observe that if we knew α_*, we could compute a solution to the trust-region subproblem from the algebraically smallest eigenvalue of B_{α_*} and a corresponding eigenvector with a special structure. The solution would consist of the last n components of the eigenvector and the Lagrange multiplier would be the eigenvalue. LSTRS starts with an initial guess for α and iteratively adjusts this parameter toward the optimal value. This is accomplished by solving a sequence of eigenvalue problems for B_α, for different α's, as we now show.

Let α be a scalar, let λ be the algebraically smallest eigenvalue of B_α, and assume that there exists a corresponding eigenvector that can be safely normalized to have the first component equal to one. For such an eigenvector, $(1, x^T)^T$, we have:

$$\begin{pmatrix} \alpha & g^T \\ g & H \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ x \end{pmatrix} \quad \Leftrightarrow \quad \begin{array}{rcl} \alpha - \lambda & = & -g^T x \\ (H - \lambda I)x & = & -g \end{array} \tag{7}$$

and consequently, two of the optimality conditions in Lemma 2.1 are automatically satisfied by the pair λ, x. Namely, $(H - \lambda I)x = -g$ with $H - \lambda I$ positive semidefinite. The latter holds by the Cauchy Interlace Theorem [Parlett 1980], which states that the eigenvalues of H interlace the eigenvalues of B_α, for any value of α. In particular, λ, the algebraically smallest eigenvalue of B_α is a lower bound for δ_1, the algebraically smallest eigenvalue of H, and therefore $H - \lambda I$ is positive semidefinite.

The relationship $\alpha = \lambda - g^T x$ could provide a way of updating α. Indeed, LSTRS uses this relationship to adjust the parameter. Note that from (7), $-g^T x = g^T (H - \lambda I)^\dagger g = \phi(\lambda)$, which is a rational function in λ with poles at the distinct eigenvalues of H. Therefore, the first equation in (7) can be written as $\alpha = \lambda + \phi(\lambda)$. Since ϕ is expensive to compute, instead of using this function directly to update α, LSTRS uses a rational interpolant for ϕ. The interpolation points are obtained by solving an eigenvalue problem for the algebraically smallest eigenvalue of B_α and a corresponding eigenvector, since the eigenpair provides suitable values for λ, $\phi(\lambda)$ and also for $\phi'(\lambda) = g^T((H - \lambda I)^\dagger)^2 g = x^T x$. The value of α is then computed as $\alpha = \widehat{\lambda} + \widehat{\phi}(\widehat{\lambda})$, where $\widehat{\phi}$ is the rational interpolant, and $\widehat{\lambda}$ satisfies $\widehat{\phi}'(\widehat{\lambda}) = \Delta^2$. One could regard the LSTRS iteration as translating the line $\alpha - \lambda$ until it intersects the graph of ϕ at the point where ϕ has slope Δ^2, as Figure 1 illustrates. Each new value of α replaces the 1,1 entry of B_α and an eigenvalue problem is solved for each new bordered matrix. A safeguarding strategy is used to ensure the convergence of α to its optimal value.

The procedure we just described relies on the assumption that there exists an eigenvector corresponding to the algebraically smallest eigenvalue of

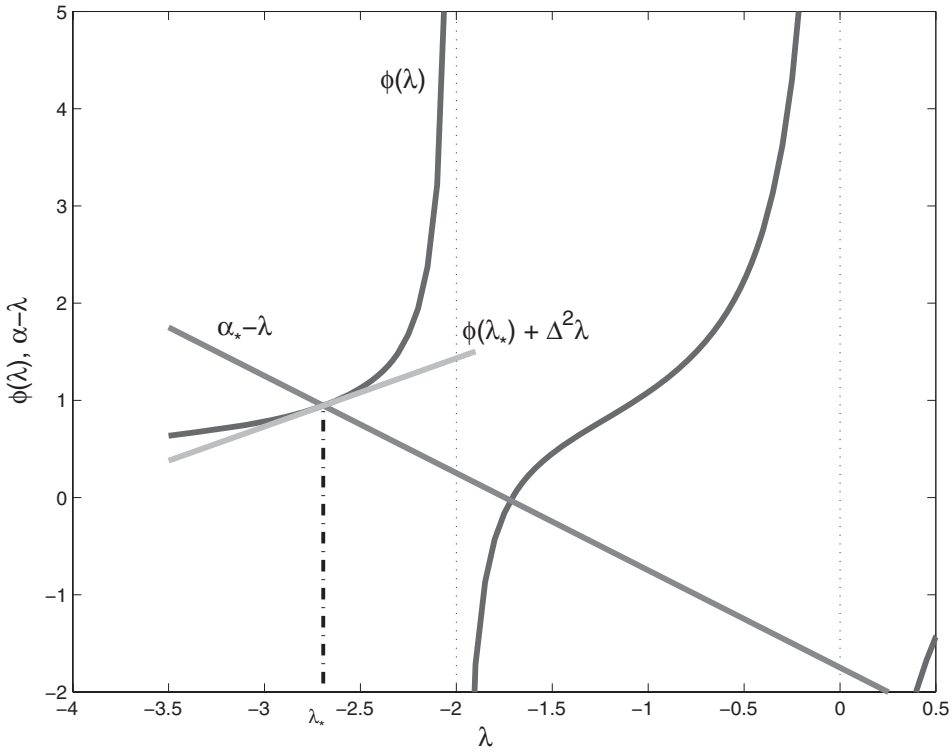

Fig. 1. LSTRS method: the standard case.

the bordered matrix that can be safely normalized to have its first component equal to one. The strategy breaks down in the presence of a zero or very small first component. This situation is equivalent to one of the conditions for the hard case and is illustrated in Figure 2. The eigenvector of interest will have a first component zero or nearly zero if and only if the vector g is orthogonal or nearly orthogonal to $\mathcal{S}_1$, the eigenspace corresponding to the algebraically smallest eigenvalue of H. Therefore, a small first component indicates the potential occurrence of the hard case. In terms of the function ϕ, this means that δ_1 is not a pole or a very weak one, and ϕ will be very steep around such a pole, causing difficulties in the interpolation procedure. LSTRS handles this case by computing *two* eigenpairs of the bordered matrix at each step: one corresponding to the algebraically smallest eigenvalue of B_α, and the other, corresponding to another eigenvalue of B_α. Under certain conditions, both eigenpairs can be used to construct an approximate solution for the trust-region subproblem.

We will now describe the main components of LSTRS: the computation of initial values, the interpolation schemes, the safeguarding strategies, and the stopping criteria. We will also describe the different tolerances needed by the method. We will focus on the results leading to the computational formulas and omit their derivations. We refer the reader to Rojas [1998] and Rojas et al. [2000] for more details.

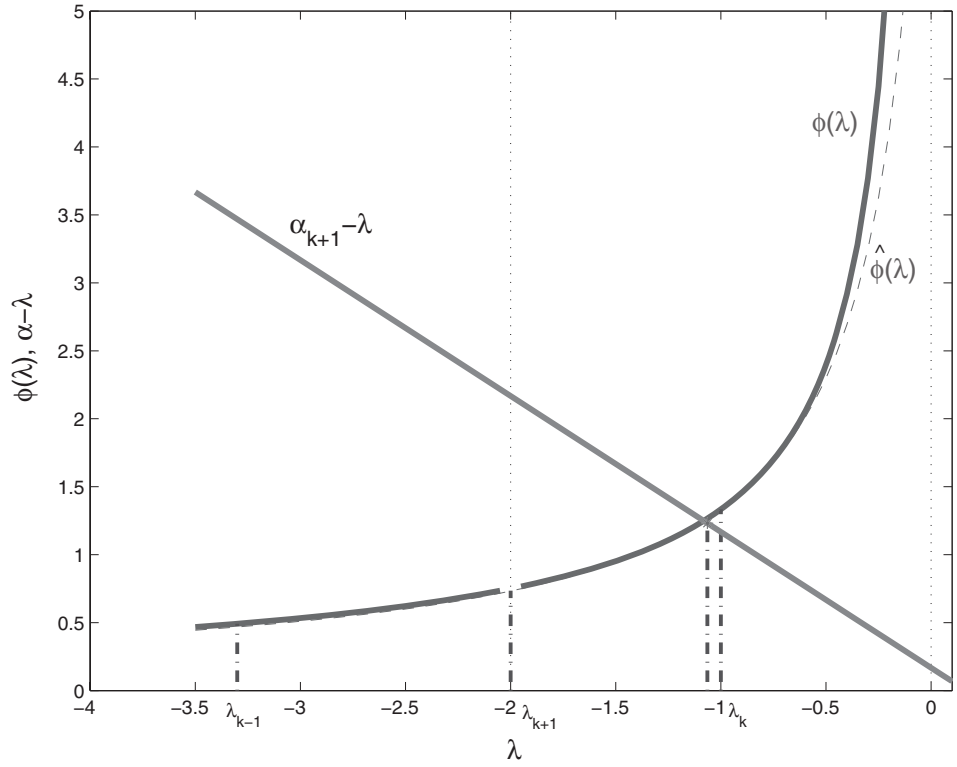

Fig. 2. LSTRS method: the near hard case. ϕ solid, $\widehat{\phi}$ dashed.

In the remainder of this section, λ_1 refers to the algebraically smallest eigenvalue of B_α and λ_i to any of the remaining ones. An eigenvector of B_α is denoted by $(v, u^T)^T$, where v is a scalar and u is an n-dimensional vector.

3.1 Initial Values

Initial values are needed for δ_L, δ_U, α_L, α_U, and α. The values δ_L and δ_U are lower and upper bounds for δ_1, the algebraically smallest eigenvalue of H. The values α_L, α_U, are lower and upper bounds for α_*, the optimal value for the parameter α.

Initial values are computed as in Rojas et al. [2000]: δ_U is chosen as either the Rayleigh quotient $\frac{u^T H u}{u^T u}$, for a random vector u, or as the minimum diagonal element of H; α_U is set to $\delta_U + \|g\|\Delta$. An initial value for α can be chosen as either $\alpha^{(0)} = \min\{0, \alpha_U\}$ or $\alpha^{(0)} = \delta_U$. The value $\alpha^{(0)}$ is used to construct a first bordered matrix $B_{\alpha^{(0)}}$ for which two eigenpairs, corresponding to λ_1 and to λ_i, are computed. As discussed before, the algebraically smallest eigenvalue is a lower bound for δ_1, and consequently we set $\delta_L = \lambda_1$. A lower bound for α is given by $\alpha_L = \delta_L - \frac{\|g\|}{\Delta}$. It was shown in Rojas et al. [2000] that the interval $[\alpha_L, \alpha_U]$ contains α_* and therefore, it is used as the initial safeguarding interval for the parameter α. We remark that an adjusting procedure is applied to $\alpha^{(0)}$ in order to ensure that one of the two eigenvectors of $B_{\alpha^{(0)}}$ can be safely normalized to have

first component one. The existence of an eigenvector with this special structure is guaranteed by the theory [Rojas et al. 2000]. This eigenvector, $(\nu, u^T)^T$ and the corresponding eigenvalue λ provide an initial iterate $\{\lambda^{(0)}, x^{(0)}\}$, with $\lambda^{(0)} = \lambda$ and $x^{(0)} = \frac{u}{\nu}$. This iterate will be used in the computation of $\alpha^{(1)}$ by the 1-point rational interpolation scheme [Rojas et al. 2000], used to interpolate the pair $(\lambda^{(0)}, \phi(\lambda^{(0)}))$. The scheme yields:

$$\alpha^{(1)} = \widehat{\lambda} + \widehat{\phi}(\widehat{\lambda}) = \alpha^{(0)} + \frac{\alpha^{(0)} - \lambda^{(0)}}{\|x^{(0)}\|} \left(\frac{\Delta - \|x^{(0)}\|}{\Delta} \right) \left(\Delta + \frac{1}{\|x^{(0)}\|} \right), \tag{8}$$

where $\widehat{\lambda} = \frac{(x^{(0)})^T H x^{(0)}}{(x^{(0)})^T x^{(0)}} + \frac{g^T x^{(0)}}{\|x^{(0)}\|\Delta}$.

The value $\alpha^{(1)}$ is used to construct a second bordered matrix $B_{\alpha^{(1)}}$ for which two eigenpairs are computed. As before, an adjusting procedure is applied to $\alpha^{(1)}$ to ensure the availability of an eigenvector with the required structure. This eigenvector, $(\nu, u^T)^T$ and the corresponding eigenvalue λ provide the new iterate $\{\lambda^{(1)}, x^{(1)}\}$, with $\lambda^{(1)} = \lambda$ and $x^{(1)} = \frac{u}{\nu}$. Observe that from the k-th LSTRS iterate we have $\lambda = \lambda^{(k)}$, $\phi(\lambda) = -g^T x^{(k)}$, and $\phi'(\lambda) = (x^{(k)})^T x^{(k)}$. Therefore, the first two iterates, $\{\lambda^{(0)}, x^{(0)}\}$ and $\{\lambda^{(1)}, x^{(1)}\}$, provide the first six values required in the 2-point rational interpolation scheme used to construct an interpolant for ϕ, which in turn is used to update the parameter α in the main iteration of LSTRS.

3.2 Update of α

The 2-point interpolation scheme [Rojas et al. 2000] used to compute $\alpha^{(k+1)}$, $k \geq 1$, yields:

$$\begin{aligned} \alpha^{(k+1)} \;=\; & \omega\alpha^{(k-1)} + (1-\omega)\alpha^{(k)} \\ & + \frac{\|x^{(k-1)}\|\,\|x^{(k)}\|\left(\|x^{(k)}\| - \|x^{(k-1)}\|\right)}{\omega\|x^{(k)}\| + (1-\omega)\|x^{(k-1)}\|} \frac{\left(\lambda^{(k-1)} - \widehat{\lambda}\right)\left(\lambda^{(k)} - \widehat{\lambda}\right)}{\left(\lambda^{(k)} - \lambda^{(k-1)}\right)}, \end{aligned} \tag{9}$$

where $\omega = \frac{\lambda^{(k)} - \widehat{\lambda}}{\lambda^{(k)} - \lambda^{(k-1)}}$, $\alpha^{(k-1)} = \lambda^{(k-1)} + \phi(\lambda^{(k-1)})$ and $\alpha^{(k)} = \lambda^{(k)} + \phi(\lambda^{(k)})$, and where

$$\widehat{\lambda} = \frac{\lambda^{(k-1)}\|x^{(k-1)}\|\left(\|x^{(k)}\| - \Delta\right) + \lambda^{(k)}\|x^{(k)}\|\left(\Delta - \|x^{(k-1)}\|\right)}{\Delta\left(\|x^{(k)}\| - \|x^{(k-1)}\|\right)}.$$

3.3 Adjustment of α

Each computed value of $\alpha^{(k)}$, $k \geq 0$, is adjusted to ensure that one of the two eigenpairs of $B_{\alpha^{(k)}}$ has an eigenvector that can be safely normalized to have first component equal to one. As previously mentioned, the existence of such an eigenvector is guaranteed by the theory [Rojas et al. 2000, Theorem 3.1]. This eigenvector is needed to construct the rational interpolants used to derive the updates (8) and (9), and continue the iterations of LSTRS. Figure 3 presents the adjusting procedure.

3.4 Safeguarding of α

The use of an interpolant for the update of α might yield values that are far from the desired optimal value α_*. Therefore, a safeguarding interval $[\alpha_L, \alpha_U]$,

Adjust α.

Input: $\varepsilon_\nu, \varepsilon_\alpha \in (0,1)$, α_L, α_U, α with $\alpha \in [\alpha_L, \alpha_U]$,
eigenpairs $\{\lambda_1, (\nu_1, u_1^T)^T\}$ and $\{\lambda_i, (\nu_i, u_i^T)^T\}$ of B_α
Output: $\alpha, \{\lambda_1, (\nu_1, u_1^T)^T\}$ and $\{\lambda_i, (\nu_i, u_i^T)^T\}$.

while
$\|g\||\nu_1| \le \varepsilon_\nu \sqrt{1-\nu_1{}^2}$ **and** $\|g\||\nu_i| \le \varepsilon_\nu \sqrt{1-\nu_i{}^2}$
and $|\alpha_U - \alpha_L| > \varepsilon_\alpha * \max\{|\alpha_L|, |\alpha_U|\}$ **do**
$\alpha_U = \alpha$
$\alpha = (\alpha_L + \alpha_U)/2$
Compute eigenpairs $\{\lambda_1, (\nu_1, u_1^T)^T\}$ and $\{\lambda_i, (\nu_i, u_i^T)^T\}$ of B_α
end while

Fig. 3. Adjustment of α.

Safeguard α.

Input: α, $\delta_U \ge \delta_1$, α_L, α_U,
$\phi_i = -g^T x^{(i)}$ and $\phi_i' = \|x^{(i)}\|^2$, for $i = k-1, k$.
Output: $\alpha \in [\alpha_L, \alpha_U]$.

if $\alpha \notin [\alpha_L, \alpha_U]$
if $k = 0$ **then** $\alpha = \delta_U + \phi_k + \phi_k'(\delta_U - \lambda^{(k)})$
else if $\|x^{(k)}\| < \|x^{(k-1)}\|$ **then** $\alpha = \delta_U + \phi_k + \phi_k'(\delta_U - \lambda^{(k)})$
else $\alpha = \delta_U + \phi_{k-1} + \phi_{k-1}'(\delta_U - \lambda^{(k-1)})$
end if
if $\alpha \notin [\alpha_L, \alpha_U]$ **then** set $\alpha = (\alpha_L + \alpha_U)/2$ **end if**
end if

Fig. 4. Safeguarding of α.

containing α_*, is maintained and updated throughout the iterations, and each value of α is safeguarded so it belongs to this interval. The safeguarding strategy is presented in Figure 4.

3.5 Stopping Criteria

3.5.1 *Boundary Solution.* We detect a boundary solution, according to Lemma 2.1, whenever the following two optimality conditions are satisfied:

$$\left(\left| \left\| \frac{u_1}{\nu_1} \right\| - \Delta \right| \le \varepsilon_\Delta * \Delta \right) \quad \text{and} \quad (\lambda_1 \le 0)$$

for a given $\varepsilon_\Delta \in (0,1)$. It suffices to check these two conditions since, as shown in the analysis of (7), the other two optimality conditions are satisfied by the eigenpair corresponding to the algebraically smallest eigenvalue of each of the bordered matrices generated in LSTRS. The solution to (1) in this case is $\lambda_* = \lambda_1$ and $x_* = \frac{u_1}{\nu_1}$.

3.5.2 *Interior Solution.* We detect an interior solution when

$$(\|u_1\| < \Delta|v_1|) \quad \text{and} \quad (\lambda_1 > -\varepsilon_{Int}),$$

for a given $\varepsilon_{Int} \in [0, 1)$. In this case, the solution is $\lambda_* = 0$ and x_* such that $Hx_* = -g$, with H positive definite. An unpreconditioned version of the Conjugate Gradient Method is used to solve the system in this case.

3.5.3 *Quasi-Optimal Solution.* Let $\psi(x) = \frac{1}{2}x^T Hx + g^T x$ be the quadratic objective function of problem (1). Then, we say that a vector $\widetilde{x}$ is a *quasi-optimal* or *nearly-optimal* solution for problem (1), if $\|\widetilde{x}\| = \Delta$ and if $\psi(\widetilde{x})$ is sufficiently close to $\psi(x_*)$, the value of the objective function at the true solution of (1): if for a given tolerance $\eta \in (0, 1)$,

$$|\psi(\widetilde{x}) - \psi(x_*)| \leq \frac{\eta}{1+\eta}|\psi(x_*)|. \tag{10}$$

A quasi-optimal solution can only occur in the hard case or near hard case. A sufficient condition for (10) to hold is the basis for the stopping criterion in the hard case. The condition has the same flavor as Lemmas 3.4 and 3.13 in Moré and Sorensen [1983], and was established in Theorem 3.2 and Lemmas 3.5 and 3.6 of Rojas et al. [2000]. Theorem 3.2 establishes that, under certain conditions, the last n components of a special linear combination of eigenvectors of B_α form a nearly-optimal solution for problem (1). Lemma 3.5 establishes the conditions under which the special linear combination can be computed, and Lemma 3.6 shows how to compute it. The three results combined yield the stopping criterion presented in Figure 5.

3.5.4 *The Safeguarding Interval is Too Small.* If the safeguarding interval for α, $[\alpha_L, \alpha_U]$, satisfies $|\alpha_U - \alpha_L| \leq \varepsilon_\alpha \max\{|\alpha_L|, |\alpha_U|\}$ for a given tolerance $\varepsilon_\alpha \in (0, 1)$, then the interval cannot be further decreased and we stop the iteration. If $(v, u^T)^T$ is one of the two available eigenvectors of the bordered matrix such that v is not small, then $x = \frac{u}{v}$ and $\lambda = \lambda_1$ are in general, good approximations to x_*, λ_*, and we return them as the approximate solution pair. If, in addition, $\|x\| < \Delta$ (hard case) and if an approximate eigenvector corresponding to the algebraically smallest eigenvalue of H is available, we add to x, a component in the direction of this eigenvector to obtain x_* such that $\|x_*\| = \Delta$. This strategy was thoroughly described in Moré and Sorensen [1983] and Sorensen [1997, Section 5], and was also adopted in Rojas [1998] and Rojas et al. [2000]. Note that the necessary eigenvector will usually be available from the LSTRS iteration. The *updated* x_* is returned along with λ_1 as a solution pair.

3.5.5 *Maximum Number of Iterations Reached.* The user may specify the maximum number of LSTRS iterations allowed and the method will stop when this number is reached.

3.6 Tolerances

LSTRS requires a few tolerances for the stopping criteria and also for some computations. The different tolerances and their meanings are summarized

Conditions for a Quasi-Optimal Solution

Input: $\lambda_1,\ (\nu_1, u_1^T)^T,\ \lambda_i,\ (\nu_i, u_i^T)^T,\ \varepsilon_{HC} \in (0,1)$
Output: True or False: quasi-optimal condition found or not.
In case a solution has been found, $\widetilde{\lambda},\ \widetilde{x}$ are also returned.

found = **false**, $\eta = \dfrac{\varepsilon_{HC}}{1-\varepsilon_{HC}}$
if $(1+\Delta^2)({\nu_1}^2 + {\nu_i}^2) > 1$ **then**

$$\tau_1 = \frac{\nu_1 - \nu_i\sqrt{(1+\Delta^2)({\nu_1}^2+{\nu_i}^2)-1}}{({\nu_1}^2+{\nu_i}^2)\sqrt{(1+\Delta^2)}},\quad \tau_2 = \frac{\nu_i + \nu_1\sqrt{(1+\Delta^2)({\nu_1}^2+{\nu_i}^2)-1}}{({\nu_1}^2+{\nu_i}^2)\sqrt{1+\Delta^2}}$$

else

$$\tau_1 = \frac{\nu_1}{\sqrt{{\nu_1}^2+{\nu_i}^2}},\quad \tau_2 = \frac{\nu_i}{\sqrt{{\nu_1}^2+{\nu_i}^2}}$$

end if

$$\widetilde{x} = \frac{\tau_1 u_1 + \tau_2 u_i}{\tau_1\nu_1 + \tau_2\nu_i},\quad \widetilde{\lambda} = {\tau_1}^2\lambda_1 + {\tau_2}^2\lambda_i,\quad \psi(\widetilde{x}) = \tfrac{1}{2}\widetilde{x}^T H\widetilde{x} + g^T\widetilde{x}$$

if $(\lambda_i - \lambda_1){\tau_2}^2(1+\Delta^2) \le -2\eta\psi(\widetilde{x})$ **then**
$\widetilde{\lambda},\ \widetilde{x}$ is a nearly-optimal pair, found = **true**
else
if $(1+\Delta^2)({\nu_1}^2+{\nu_i}^2) > 1$ **then**

$$\tau_1 = \frac{\nu_1 + \nu_i\sqrt{(1+\Delta^2)({\nu_1}^2+{\nu_i}^2)-1}}{({\nu_1}^2+{\nu_i}^2)\sqrt{(1+\Delta^2)}},\quad \tau_2 = \frac{\nu_i - \nu_1\sqrt{(1+\Delta^2)({\nu_1}^2+{\nu_i}^2)-1}}{({\nu_1}^2+{\nu_i}^2)\sqrt{1+\Delta^2}}$$

$$\widetilde{x} = \frac{\tau_1 u_1 + \tau_2 u_i}{\tau_1\nu_1 + \tau_2\nu_i},\quad \widetilde{\lambda} = {\tau_1}^2\lambda_1 + {\tau_2}^2\lambda_i,\quad \psi(\widetilde{x}) = \tfrac{1}{2}\widetilde{x}^T H\widetilde{x} + g^T\widetilde{x}$$

if $(\lambda_i - \lambda_1){\tau_2}^2(1+\Delta^2) \le -2\eta\psi(\widetilde{x})$ **then**
$\widetilde{\lambda},\ \widetilde{x}$ is a nearly-optimal pair, found = **true**
end if
end if
end if

Fig. 5. Conditions for a quasi-optimal solution.

in Table I. The MATLAB implementation of LSTRS provides a set of default values for the tolerances.

3.7 Algorithm

The strategies and procedures described in Sections 3.1 through 3.5 are the building blocks for the LSTRS method, shown in Figure 6.

4. LSTRS FOR REGULARIZATION

A few considerations are in order when using the LSTRS MATLAB software for regularization problems, for which $H = A^TA$ and $g = -A^Tb$.

In the first place, since the potential near hard case is the common case for these problems (cf. Section 2.2), the solution will usually be quasi-optimal. However, it might happen that the exact hard case is detected, which will cause a correction term to be added to the iterate so it can have the desired norm (see Section 3.5.4). The correction term is in the direction of an eigenvector

Table I. Tolerances for LSTRS

ε_Δ	The desired relative accuracy in the norm of the trust-region solution. A boundary solution x satisfies $\frac{\lvert\lVert x\rVert-\Delta\rvert}{\Delta} \le \varepsilon_\Delta$.
ε_{HC}	The desired accuracy of a quasi-optimal solution. If x_* is the true solution and $\widetilde{x}$ is the quasi-optimal solution, then $\psi(x_*) \le \psi(\widetilde{x}) \le (1-\varepsilon_{HC})\psi(x_*)$, where $\psi(x) = \frac{1}{2}x^T Hx + g^T x$.
ε_{Int}	Used to declare the algebraically smallest eigenvalue of B_α positive in the test for an interior solution: λ_1 is considered positive if $\lambda_1 > -\varepsilon_{Int}$.
ε_α	The minimum relative length of the safeguarding interval for α. The interval is too small when $\lvert\alpha_U - \alpha_L\rvert \le \varepsilon_\alpha * \max\{\lvert\alpha_L\rvert, \lvert\alpha_U\rvert\}$.
ε_ν	The minimum relative size of an eigenvector component. The component ν is small when $\lvert\nu\rvert \le \varepsilon_\nu \dfrac{\lVert u\rVert}{\lVert g\rVert}$.
maxiter	The maximum number of iterations allowed.

LSTRS

Input: $H \in \mathbb{R}^{n\times n}$, $g \in \mathbb{R}^n$, $\Delta > 0$,
Tolerances: $\varepsilon_\Delta, \varepsilon_\nu, \varepsilon_{HC}, \varepsilon_\alpha \in (0,1)$, $\varepsilon_{Int} \in [0,1)$.
Output: λ_*, x_* satisfying conditions of Lemma 2.1,
or $\widetilde{\lambda}$, $\widetilde{x}$, a quasi-optimal pair as in Figure 5.

1. Initialization
1.1 Compute: $\delta_U \ge \delta_1$, $\alpha_U \ge \alpha_*$, $\alpha^{(0)}$ as in Section 3.1
1.2 Compute eigenpairs $\{\lambda_1, (\nu_1, u_1^T)^T\}$ and $\{\lambda_i, (\nu_i, u_i^T)^T\}$ of $B_{\alpha^{(0)}}$
1.3 Compute $\delta_L \le \delta_1$, $\alpha_L \le \alpha_*$ as in Section 3.1
1.4 Set $k = 0$
2. repeat
2.1 Update $\delta_U = \min\left\{\delta_U, \dfrac{u_1^T H u_1}{u_1^T u_1}\right\}$
2.2 Adjust $\alpha^{(k)}$ using procedure in Figure 3
2.3 if $\lVert g\rVert\lvert\nu_1\rvert > \varepsilon_\nu\sqrt{1-\nu_1{}^2}$ **then**
set $\lambda^{(k)} = \lambda_1$ and $x^{(k)} = \dfrac{u_1}{\nu_1}$
if $\lVert x^{(k)}\rVert < \Delta$ **then** $\alpha_L = \alpha^{(k)}$ **end if**
if $\lVert x^{(k)}\rVert > \Delta$ **then** $\alpha_U = \alpha^{(k)}$
else set $\lambda^{(k)} = \lambda_i$, $x^{(k)} = \dfrac{u_i}{\nu_i}$ and $\alpha_U = \alpha^{(k)}$ **end if**
end if
2.4 Compute $\alpha^{(k+1)}$ by rational interpolation schemes using (8) from Section 3.1 if $k = 0$, and (9) from Section 3.2, otherwise
2.5 Safeguard $\alpha^{(k+1)}$ using procedure in Figure 4
2.6 Compute eigenpairs $\{\lambda_1, (\nu_1, u_1^T)^T\}$ and $\{\lambda_i, (\nu_i, u_i^T)^T\}$ of $B_{\alpha^{(k+1)}}$
2.7 Set $k = k+1$
until convergence

Fig. 6. LSTRS: an algorithm for large-scale trust-region subproblems.

corresponding to the smallest eigenvalue of H. In general, such correction term is not desirable in regularization, since it might bring high-frequency components into the solution (cf. Section 2.2). Therefore we recommend not adding the correction term for regularization problems. To indicate this choice, the

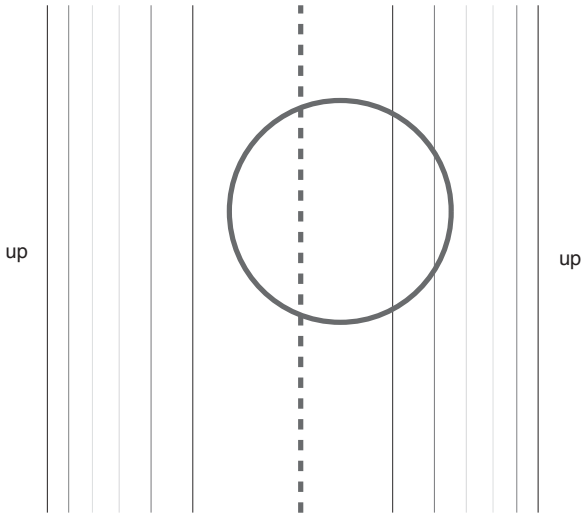

Fig. 7. Interior and boundary solutions for a positive semidefinite Hessian.

user should set the input parameter **lopts.correction** to a string that does not begin with **'y'** or **'Y'**. Note that, if the correction term is not added, the solution provided by LSTRS is not a solution to the original trust-region subproblem, but a regularized solution corresponding to a smaller value of Δ.

The second consideration concerns interior solutions. For problem (2), an interior solution corresponds to the unconstrained least squares solution. Such solution is unique if the Hessian matrix is positive definite. If the Hessian is positive semidefinite and singular, it might be that both interior and boundary solutions exist, as illustrated in Figure 7, where we can see that there are infinite solutions along the dashed line in the "valley." The least squares solution is of no interest in regularization since it is contaminated by the noise in the data. In general however, we have no means of distinguishing between the positive definite and the positive semidefinite cases. Therefore, we recommend not computing the least squares (interior) solution, when such a solution is detected for a regularization problem. To indicate this choice, the user should set the input parameter **lopts.interior** to a string that does not begin with **'y'** or **'Y'**. In this case, a message is displayed advising the user to decrease Δ. The current iterate is returned since it might be useful in some situations. Note that if λ is small and an interior solution is detected, then the current iterate $x^{(k)}$ is an approximation to $x = -H^{-1}g = (A^TA)^{-1}(A^T\bar{b})$, which is the interior solution. If in addition, the noise level in $\bar{b}$ is low, $x^{(k)}$ will be a reasonable approximation to the desired solution.

A final note on regularization problems concerns the eigensolver. As we saw in Section 2.2, for these problems, the smallest eigenvalues of the Hessian matrix are usually 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 can 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 Implicitly Restarted Lanczos Method (IRLM) [Sorensen 1992] will converge much faster toward the eigenvector q_1 corresponding to λ_1. As in Rojas and Sorensen [2002], we use a Tchebyshev 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. The convergence of the 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. We provide the routine **tcheigs_lstrs_gateway**, implementing a Tchebyshev Spectral Transformation with a polynomial of degree $\ell = 10$. We recommend the use of this routine for most regularization problems.

A less expensive alternative to the Tchebyshev Spectral Transformation consists of using zero as eigenvalue and the available Lanczos vectors (not converged eigenvectors) returned by the eigensolver (e.g. `eigs_lstrs`) to construct an LSTRS iterate. The rationale behind this heuristics is that zero is a lower bound for the smallest eigenvalue of $H = A^T A$, and that the Lanczos vectors are rich in the direction of the eigenvectors of interest. Note that the safeguarding mechanisms of LSTRS guarantee convergence also in this case. This option is available through the input parameter **lopts.heuristics** and can only be used in combination with the eigensolver `eigs_lstrs`. If the heuristics does not yield satisfactory results, then **lopts.heuristics** must be set to zero and the Tchebyshev Spectral Transformation must be used.

5. COMPARISONS

In this section, we compare LSTRS with other methods for the large-scale trust-region subproblem. The methods used for comparisons were the Sequential Subspace Method (SSM) of Hager [2001], the Semidefinite Programming approach (SDP) of Fortin and Wolkowicz [2004], and the Generalized Lanczos Trust Region method (GLTR) of Gould et al. [1999]. Note that only LSTRS, SSM and SDP are limited-memory methods. For SSM, results are reported only for the two matrix-free variants SSM and SSM$_d$. The methods are described in Section 5.1.

We used MATLAB implementations of LSTRS, SSM, and SDP, and a Fortran 90 implementation of GLTR. It is important to note that the four codes are at different stages of maturity. The GLTR code is the routine HSL_VF05 of the HSL library [HSL 2004]. The SSM and SDP codes are initial implementations, not yet released publicly, that were kindly provided by their authors for the purpose of these comparisons. In particular, the SDP code is still under development. The double precision version of HSL_VF05 was used whereas the MATLAB codes were run under MATLAB 6.0. The experiments were carried out on a SUN Ultra-250 with a 400 MHZ processor and 2048 Megabytes of RAM, running

Solaris 5.8. The floating point arithmetic was IEEE standard double precision with machine precision $2^{-52} \approx 2.2204 \cdot 10^{-16}$.

We ran the codes on three different families of problems whose Hessian matrices were as follows: the 2-D Discrete Laplacian, UDU^T with U orthogonal and D diagonal, and a discretized operator from the inverse heat equation. The Laplacian is a frequently used model problem in CFD applications, the UDU^T matrix allows for the exploration of ill conditioning and the effect on the hard case, while the inverse heat equation is a well-known ill-posed problem.

For the first two families of problems, we report the average number of matrix-vector products (MVP), of the number of vectors (STORAGE), and of the value of the optimality measure $\frac{\|(H-\lambda\ I)x+g\|}{\|g\|}$, from a sample of ten related problems in each family. Time is not reported, due to the different nature of the implementations: three MATLAB interpreted codes and one Fortran 90 standalone code. Since the methods use different stopping criteria, the tolerances were adjusted so that the methods computed a solution with $\frac{\|(H-\lambda\ I)x+g\|}{\|g\|}$ of the order of 10^{-6}. Finally, the results used to compute the averages correspond to the choice of options for which each method required the lowest number of matrix-vector products. For the third family of problems, we report the number of matrix-vector products, the storage, the optimality measure, and the relative error of the solution with respect to the true solution to the inverse problem. We report the best results in terms of this relative error out of several trials with each method.

This section is organized as follows. A brief description of the methods used for comparisons is presented in Section 5.1. A detailed description of the three families of problems, the settings for each of the codes, and the results, are presented in Sections 5.2, 5.3, and 5.4, respectively. The discussion of the results is presented in Section 5.5.

5.1 Methods for Large-Scale Trust-Region Subproblems

SSM

SSM considers subproblems restricted to a Krylov subspace and by imposing that this subspace contains the iterate generated by one step of the sequential quadratic programming (SQP) algorithm applied to the trust region subproblem, a locally quadratic convergent scheme is obtained. The SQP method is equivalent to Newton's method applied to the nonlinear system

$$\begin{aligned}(H - \lambda\ I)x + g &= 0\\ \frac{1}{2}x^T x - \frac{\Delta^2}{2} &= 0.\end{aligned}$$

The use of the minimum residual solution ensures locally quadratic convergence even for degenerate problems with multiple solutions and a singular Jacobian for the first order optimality conditions. Hager observed in his experiments that appropriate small-dimensional subspaces could be generated by combining preconditioned Krylov spaces with minimum residual techniques [Hager 2001]. Two preconditioned schemes corresponding to a diagonal preconditioner

(SSM_d) and an SSOR-type of approach (SSM_{ssor}) were suggested in Hager [2001]. Note that the SSM_{ssor} variant is not matrix-free.

SDP

The SDP method is an extension of the semi-definite programming approach of Rendl and Wolkowicz [1997]. The current algorithm maintains the primal-dual philosophy of the previous and introduces a novel strategy for the hard case, which combines shifting of the eigenvalues and deflation. The equality-constrained trust-region subproblem is, due to duality, equivalent to an unconstrained concave maximization problem in one variable. The evaluation of the objective function of this problem depends on the determination of the algebraically smallest eigenvalue of a parameterized eigenvalue problem, similar to the idea employed by LSTRS. In the MATLAB working code provided by the authors, the eigenvalue is computed with MATLAB's subroutine **eigs**.

GLTR

The GLTR approach is based on the Lanczos tridiagonalization of the matrix H and on the solution of a sequence of problems restricted to Krylov spaces, inspired by the Steihaug-Toint algorithm [Steihaug 1983; Toint 1981]. GLTR uses a weighted ℓ_2 norm that defines the trust region and plays the role of preconditioning. The method is an alternative to the Steihaug-Toint algorithm, which further investigates the trust-region boundary whenever it is reached, and keeps the efficiency of the preconditioned conjugate gradient method inside the trust region.

Although GLTR does not require any factorization of H, the Lanczos vectors are needed to recover the minimizer of the original problem. Therefore, the Lanczos vectors should be either stored or regenerated [Gould et al. 1999, p. 509] and limited-storage requirements may be lost.

The software package is available as the Fortran 90 module HSL_VF05 in the Harwell Subroutine Library [HSL 2004], and it is also part of the GALAHAD Optimization Library, version 1.0 [Gould et al. 2002].

5.2 The 2-D Discrete Laplacian Family

In this family of problems, the Hessian matrix was $H = L - 5I$, where L is the standard 2-D discrete Laplacian on the unit square based upon a 5-point stencil with equally-spaced mesh points. The diagonal shift of -5 was introduced to make H indefinite. The order of H was $n = 1024$. We used the four trust-region solvers to solve a sequence of ten related problems, differing only by the vector g, randomly generated with entries uniformly distributed on $(0, 1)$. The trust-region radius was fixed at $\Delta = 100$.

We studied problems with and without the hard case. To generate the hard case, we orthogonalized the random vectors g against the eigenvector q corresponding to the algebraically smallest eigenvalue of H. We accomplished this by setting $g := g - q(q^T g)$. For the easy and hard cases we added a noise vector to g, of norm 10^{-8}.

Table II. Average Results for the 2-D Laplacian, $n = 1024$

METHOD	MVP	STORAGE	$\frac{\|(H-\lambda\ I)x+g\|}{\|g\|}$
LSTRS	127.1	10	2.32×10^{-6}
SSM	67.3	10	9.53×10^{-7}
SSM_d	67.3	10	9.53×10^{-7}
SDP	595	10	3.17×10^{-5}
GLTR	81.6	41.3	8.56×10^{-6}

(a) Easy Case

METHOD	MVP	STORAGE	$\frac{\|(H-\lambda\ I)x+g\|}{\|g\|}$
LSTRS	252.6	10	6.91×10^{-6}
SSM	377.9	10	1.42×10^{-6}
SSM_d	377.9	10	1.42×10^{-6}
SDP	2023.8	10	5.76×10^{-2}
GLTR	151.8	76.4	8.37×10^{-6}

(b) Hard Case

For the limited-memory methods (LSTRS, SSM, SSM_d, and SDP) the number of vectors was fixed at 10, and 8 shifts were applied in each implicit restart in ARPACK. Other parameters were as follows. LSTRS: for the easy case, $\varepsilon_\Delta = 10^{-5}$, $\varepsilon_{HC} = 10^{-11}$; for the hard case, $\varepsilon_{HC} = \varepsilon_\Delta = 10^{-11}$. In both cases, $\delta_U =$ 'mindiag' and $\alpha^{(0)} = \delta_U$ as in Rojas et al. [2000]. The initial vector for ARPACK was $e/\sqrt{n+1}$, where e is the vector of all ones. Default values were used for the remainder of the parameters. SSM, SSM_d: $\|(H - \lambda I)x + g\|$ was required to be less than or equal to $tol = 10^{-5}$; one of the initial vectors in a relevant Krylov subspace was chosen as the vectors of all ones. SDP: the tolerance for the duality gap was set to 10^{-10} in the easy case and 10^{-9} in the hard case. GLTR: the tolerance for the optimality measure was set to 10^{-5} and the required fraction of the optimal value of the objective function was set to 1. The desired optimality level could not be achieved for lower fractions of the optimal objective value. This was also the case in the other experiments.

The results are presented in Table II. For the easy case, SSM required about 50 and GLTR required a relatively low number of matrix-vector products but used more than three times the storage of the limited-memory methods. For the hard case, LSTRS required about 30 SSM. GLTR required the lowest number of matrix-vector products but more than six times the storage of the limited-memory methods.

5.3 The UDU^T Family

In these problems, the matrix H was of the form $H = UDU^T$, with D a diagonal matrix with elements $d_1, \ldots, d_n$ and $U = I - 2uu^T$, with $u^T u = 1$. The elements of D were randomly generated with a uniform distribution on $(-5, 5)$, then sorted in nondecreasing order and d_1 set to -5. Both vectors u and g were randomly generated with entries selected from a uniform distribution on $(-0.5, 0.5)$ and then u was normalized to have unit length. The order of H was $n = 1000$. There was a total of ten problems.

In this case, the eigenvectors of the matrix H are of the form $q_i = e_i - 2uu_i$, $i = 1, \ldots, n$ with e_i the i-th canonical vector in $\mathbb{R}^n$, and u_i the i-th component of the vector u. The vector g was orthogonalized against $q_1 = e_1 - 2uu_1$, and a noise vector was added to g. Finally, g was normalized to have unit norm. The noise vectors had norms 10^{-2} and 10^{-8}, for the easy and hard cases, respectively. To ensure that the hard case really occurred, we computed $\Delta_{min} = \|(H - d_1 I)^{\dagger} g\|$, and set $\Delta = 0.1\Delta_{min}$ for the easy case and $\Delta = 5\Delta_{min}$ for the hard case.

For the limited-memory methods (LSTRS, SSM, SSM$_d$ and SDP), the number of vectors was fixed at 10 in the easy case and 24 in the hard case. Other parameters were as follows. LSTRS: for the easy case, δ_U was set to the minimum of the diagonal of UDU^T, $\alpha^{(0)} = \delta_U$, the adaptive tolerance for the eigenpairs was 0.2, and 8 shifts were applied in each implicit restart; for the hard case, $\delta_U = -4.5$, $\alpha^{(0)} =$ 'min', the adaptive tolerance for the eigenpairs was 0.03, and 12 shifts were applied in each implicit restart; $\varepsilon_{\Delta} = 10^{-4}$, $\varepsilon_{HC} = 10^{-10}$. More basis vectors were needed in the hard case since the eigenvalues were computed to a higher accuracy. The initial vector for ARPACK was $e/\sqrt{n+1}$, where e is the vector of all ones. Default values were used for the remainder of the parameters. SSM, SSM$_d$: $tol = 10^{-6}$ and one of the initial vectors in a relevant Krylov subspace was chosen as the vectors of all ones. SDP: the tolerance for the duality gap was set to 10^{-11} in the easy case and 10^{-12} in the hard case. GLTR: the tolerance for the KKT (Karush-Kuhn-Tucker) condition was set to 10^{-5} for the easy case and 10^{-7} for the hard case. The fraction of the optimal value of the objective function was set to 1.

The results are reported in Table III. The SSM methods outperformed all methods for the hard case. GLTR was comparable in both computations and storage for the easy case, but more expensive than the SSM methods for the hard case. These results are consistent with those reported in Hager [2001] and show that SSM performs extremely well on this class of problems. LSTRS was the second best of the limited-memory methods. For these problems, we found that LSTRS was sensitive to the choice of some initial parameters such as δ_U and $\alpha^{(0)}$. This behavior, as well as the features of this particular class of problems, should be further investigated.

5.4 Regularization Problems

The third family of problems comes from the Regularization Tools package by Hansen [1994]. We chose problem **heat** of dimension $n = 1000$. This problem is

Table III. Average Results for UDU^T, $n = 1000$

METHOD	MVP	STORAGE	$\frac{\|(H-\lambda\ I)x+g\|}{\|g\|}$
LSTRS	90.2	10	2.95×10^{-6}
SSM	35.2	10	1.35×10^{-6}
SSM_d	24.1	10	9.90×10^{-7}
SDP	950.4	10	9.65×10^{-7}
GLTR	36.8	18.9	7.37×10^{-6}

(a) Easy Case

METHOD	MVP	STORAGE	$\frac{\|(H-\lambda\ I)x+g\|}{\|g\|}$
LSTRS	954.1	24	9.65×10^{-6}
SSM	445.1	24	1.91×10^{-6}
SSM_d	130.4	24	9.59×10^{-7}
SDP	1720.8	24	7.86×10^{-6}
GLTR	634.6	317.8	7.64×10^{-6}

(b) Hard Case

a discretized version of the Inverse Heat Equation, which arises, for example, in the inverse heat conduction problem of determining the temperature on the surface of a body from transient measurements of the temperature at a fixed location in the interior [Carasso 1982]. The equation is a Volterra integral equation:

$$\gamma(y) = \int_0^1 \mathcal{K}(y,t)\phi(t)\,dt, \quad 0 \le y \le 1, \tag{11}$$

where $\mathcal{K}(y,t) = k(y-t)$, with $k(t) = \frac{t^{-3/2}}{2\kappa\sqrt{\pi}} \exp(-\frac{1}{4\kappa^2 t^2})$. The parameter κ controls the degree of ill posedness. We performed experiments with a mildly ill-posed problem ($\kappa = 5$) and a severely ill-posed one ($\kappa = 1$).

To compute regularized solutions for problem (11), we solve the following quadratically constrained least squares problem:

$$\min\ \frac{1}{2}\|Ax-b\|^2\ \text{ s.t. } \|x\|\ \le\ \Delta.$$

The MATLAB routine **heat** provided the matrix A, the vector b, and X_{IP}, a discretized version of the analytical solution of the continuous problem. For the trust-region problem, $H = A^TA$, $g = -A^Tb$, and $\Delta = \|X_{IP}\|$. Twenty percent of the singular values of the matrix A were zero to working precision. No noise

Table IV. Results for the Inverse Heat Equation, $n = 1000$

METHOD	MVP	STORAGE	$\frac{\|(H-\lambda I)x+g\|}{\|g\|}$	$\frac{\|x-X_{IP}\|}{\|X_{IP}\|}$
LSTRS	265	8	9.12×10^{-7}	6.13×10^{-4}
SSM	700	8	2.99×10^{-9}	2.41×10^{-4}
SSM_d	649	8	2.74×10^{-9}	4.57×10^{-4}
SDP	5700	8	2.73×10^{-7}	3.63×10^{-4}

(a) Mildly Ill-Posed Case

METHOD	MVP	STORAGE	$\frac{\|(H-\lambda I)x+g\|}{\|g\|}$	$\frac{\|x-X_{IP}\|}{\|X_{IP}\|}$
LSTRS	552	8	7.05×10^{-6}	5.49×10^{-2}
SSM	512	8	1.81×10^{-7}	3.75×10^{-2}
SSM_d	215	8	2.04×10^{-7}	2.25×10^{-2}
SDP	4600	8	2.27×10^{-4}	2.08×10^{-1}

(b) Severely Ill-Posed Case

was added to the vector b since, as discussed in Section 2.2; the absence of noise yields a more difficult trust-region problem.

Since this problem is implemented as a MATLAB routine, we only tested the methods for which a MATLAB implementation was available: LSTRS, SSM, SSM_d, and SDP. Several options were tried for all methods. We report the best results in terms of the relative error in the solution to the trust-region problem with respect to X_{IP}, the exact solution to the inverse problem. Note that a bound on the optimality measure was not prescribed for these problems.

The number of vectors was fixed at 8 since this choice produced the best results for all methods. Other settings were as follows. For LSTRS: for the mildly ill-posed problem, $\varepsilon_\Delta = 10^{-3}$ and `lopts.maxeigentol = 0.7`; for the severely ill-posed problem, $\varepsilon_\Delta = 10^{-2}$ and `lopts.maxeigentol = 0.4`. The initial vector for ARPACK was $e/\sqrt{n+1}$, where e is the vector of all ones. The parameter `lopts.heuristics` was set to 1. Default values were used for the remainder of the parameters. SSM, SSM_d: $tol = 10^{-8}$ for the mildly ill-posed problem and 10^{-7} for the severely ill-posed problem, one of the initial vectors in a relevant Krylov subspace was chosen as the vector of all ones. SDP: the tolerance for the duality gap was set to 10^{-7} for the mildly ill-posed problem and 10^{-10} for the severely ill-posed problem.

The results are reported in Table IV. In the mildly ill-posed case, all the computed solutions are indistinguishable from the exact solution to the inverse problem. Plots of the solutions in the severely ill-posed case are shown in Figure 8. The results in Table IV show that LSTRS required the lowest number of matrix-vector products for the mildly ill-posed problem. For the severely

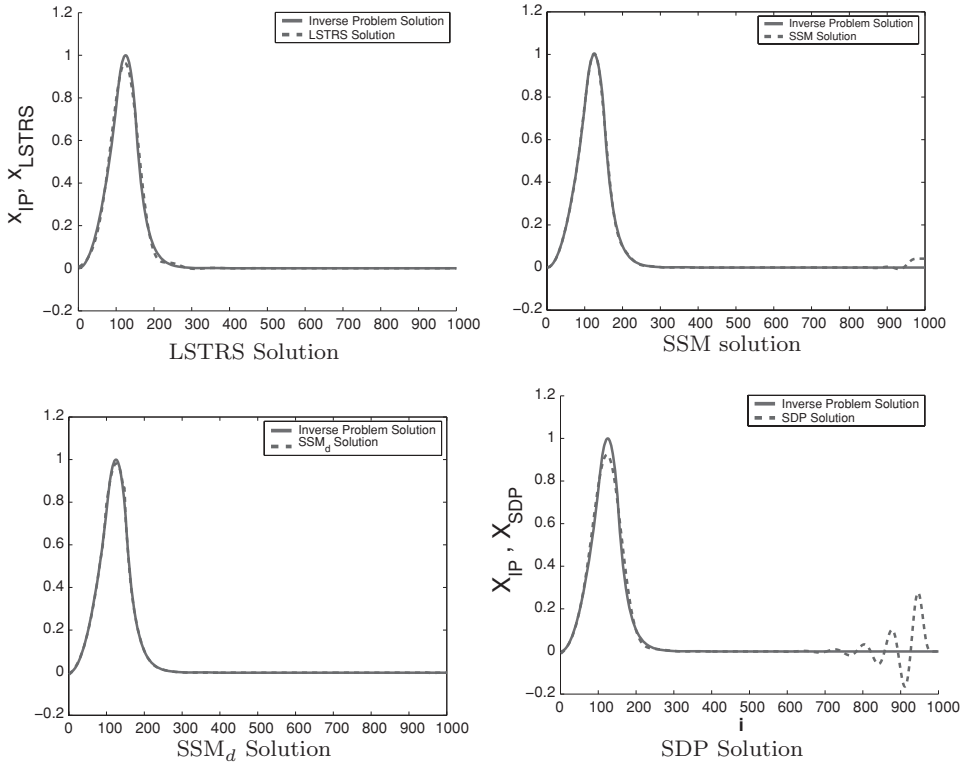

Fig. 8. Regularized solutions (dashed) and the exact solution (solid) to the inverse problem for a severely ill-posed Inverse Heat Equation, $n = 1000$.

ill-posed problem, the number of matrix-vector products required by LSTRS was comparable to, or less than, for the unpreconditioned methods, while the diagonally-preconditioned version of SSM had the best performance on this problem. In Figure 8, we can observe the oscillatory pattern in the SDP computed approximation for the severely ill-posed problem. The oscillations are probably due to high-frequency components and indicate that the desired regularizing effect could not be achieved.

5.5 Discussion

We have compared four methods for the large-scale trust-region subproblem on a set of different problems. Although more experiments should be performed, our results seem to indicate that LSTRS is competitive with state-of-the-art techniques for some classes of trust-region problems, including regularization problems. The latter was expected since part of the motivation for the development of the method came from the regularization of large-scale discrete forms of ill-posed problems.

The Sequential Subspace Methods performed extremely well on most problems and we believe that a publicly available, matrix-free and portable, MEX-file-free, implementation will be of interest. As we pointed out before, the

Semidefinite Programming approach is still under development and we expect that some refinement of the codes and of the interesting deflation strategy will greatly improve the performance of the method. Moreover, the use of a Tchebyshev Spectral Transformation might help compute small eigenvalues to higher accuracy. Our tests with SDP were run in the summer of 2004. The authors have since reported improvements. Recently, they have also used their approach for regularization [Grodzevich and Wolkowicz 2007].

Except for regularization problems, the GLTR approach had the best performance in terms of the number of matrix-vector products. However, the memory requirements were larger than for the other methods. The method is a good choice when storage is not an issue and stand-alone software is desired. To the best of our knowledge, the method is yet to be tested on regularization problems.

Finally, we remark that both SDP and LSTRS required the solution of a low number of eigenproblems in all tests. As pointed out in Hager [2001], the parameterized eigenvalue approach would probably benefit from improvements in the eigenvalue computation such as the introduction of some sort of preconditioning together with ARPACK, or the use of other techniques such as the Jacobi-Davidson method [Sleijpen and van der Vorst 1996]. For LSTRS, these possibilities are yet to be investigated and can be easily incorporated into the software.

6. APPLICATIONS

A MATLAB 5.3 implementation of LSTRS has been successfully used in several large-scale applications. In Rojas [1998] and Rojas and Sorensen [2002], the code was used in the study of the bathymetry of the Sea of Galilee. This study required the regularization of a linear inverse interpolation problem of dimension **40401**. The same version of the code was used to compute regularized solutions for a problem arising in the solution of the visco-acoustic wave equation in marine oil exploration with field data. The problem was of dimension **121121**. The code was also used as an inner solver in the TRUST$_\mu$ method for non-negative image restoration [Rojas and Steihaug 2002], where the typical images are digital arrays of 256×256 pixels which give rise to trust-region problems of dimension **65536**.

The MATLAB 5.3 and the current versions used the same computational routines. The versions differ only on the interface.

The current code was used as the inner solver in the iterative method MLFIP [Eldén et al. 2005], which has been applied to the computation of confidence intervals for regularized solutions of the severely ill-posed sideways inverse heat equation. These experiments were carried out in MATLAB 6.5.

The performance of LSTRS in the context of trust-region methods for nonlinear programming is yet to be investigated. An important aspect to take into account is that nearly-exact solutions are acceptable in this context, as long as they provide a fraction of the Cauchy-point reduction [Conn et al. 2000; Nocedal and Wright 1999]. We expect that LSTRS can compute such solutions with lower computational effort than in regularization, where good approximations to the exact solutions are usually required.

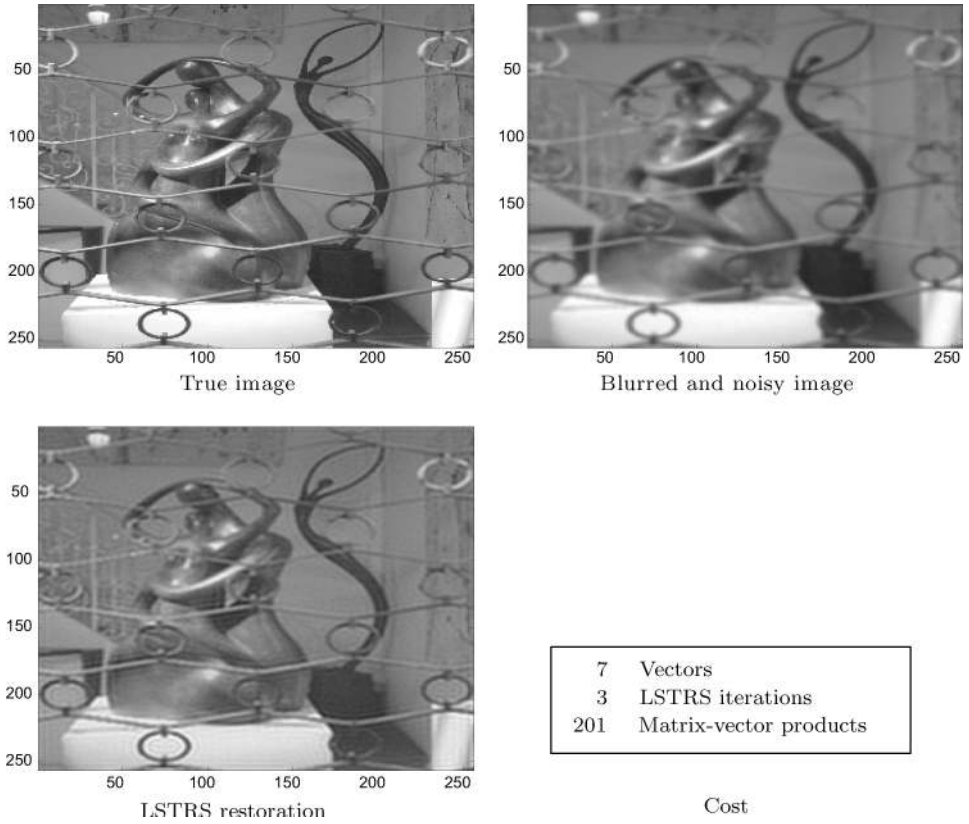

Fig. 9. Restoration of the photograph of an art gallery in Paris. Dimension: **65536**.

Next, we present a large-scale example in image restoration. The problem is that of recovering an image from blurred and noisy data. The problem was constructed in the following way. A digital photograph of an art gallery in Paris was blurred with the routine **blur** from the Regularization Tools package [Hansen 1994]. Then, a random Gaussian noise vector was added to the blurred image. The regularization problem was a constrained least squares problem of type (5), where A was the blurring operator returned by the routine **blur**, and $\bar{b}$ was the blurred and noisy image generated as above and stored columnwise as a one-dimensional array. The noise level in $\bar{b}$ was 10^{-2}. The dimension of the problem is **65536**.

The following options were used in LSTRS: **epsilon.Delta** $= 10^{-2}$ and **epsilon.HC** $= 10^{-4}$. The eigensolver was **tcheigs_lstrs_gateway** with initial vector equal to the vectors of all ones. The results are shown in Figure 9. Default values were used for the remainder of the parameters. LSTRS required 201 matrix-vector products and 7 vectors of storage to compute a quasi-optimal solution with a relative error of 1.06×10^{-1} with respect to the true solution. The optimality measure was 1.01×10^{-3}. The waves or ripples observed in the LSTRS restoration are due to the famous Gibbs phenomenon [Parks and Burrus 1987] and are characteristic of least squares restorations.

7. CONCLUDING REMARKS

We have presented a MATLAB implementation of the LSTRS method and compared it with state-of-the-art methods for large-scale trust-region subproblems. Our results seem to indicate that LSTRS is competitive with existing techniques, especially in the presence of singularities such as in regularization problems.

ACKNOWLEDGMENTS

The authors would like to thank Chao Yang for providing the code for the Tchebyshev matrix-vector multiplication. We also thank Bill Hager and Henry Wolkowicz for providing MATLAB implementations of their methods so that comparisons with LSTRS could be performed. M. Rojas would like to thank Iain Duff and the rest of the Algo Team at CERFACS for welcoming her for several extended visits, during which part of this work was done. The final revision of this work was done while M. Rojas was visiting Professor Cees Roos and the Optimization Group of the Software Technology Department at the Delft University of Technology. She would like to thank Professor Roos and his group for their hospitality. We are indebted to Martin van Gijzen for helpful discussions and suggestions. Finally, we thank the three anonymous referees and the editor for numerous helpful suggestions that greatly improved an earlier version of this manuscript.

REFERENCES

BERTERO, M. AND BOCACCI, P. 1998. *Introduction to Inverse Problems in Imaging*. Institute of Physics, Bristol.

BJÖRCK, Å. 1996. *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia.

CARASSO, A. 1982. Determining surface temperatures from interior observations. *SIAM J. Appl. Math. 42*, 558–574.

CONN, A. R., GOULD, N. N., AND TOINT, P. L. 2000. *Trust Region Methods*. SIAM, Philadelphia.

ELDÉN, L. 1977. Algorithms for the regularization of ill-conditioned least squares problems. *BIT 17*, 134–145.

ELDÉN, L., HANSEN, P. C., AND ROJAS, M. 2005. Minimization of linear functionals defined on solutions of large-scale discrete ill-posed problems. *BIT 45,* 2, 329–340.

FORTIN, C. AND WOLKOWICZ, H. 2004. The trust-region subproblem and semidefinite programming. *Optim. Methods Softw. 19,* 1, 41–67.

GAY, D. 1981. Computing optimal locally constrained steps. *SIAM J. Sci. Stat. Comput. 2,* 2, 186–197.

GOLUB, G. H. AND VAN LOAN, C. F. 1996. *Matrix Computations*, Third ed. John Hopkins University Press, Baltimore.

GOLUB, G. H. AND VON MATT, U. 1991. Quadratically constrained least squares and quadratic problems. *Numer. Math. 59*, 561–580.

GOULD, N. N., LUCIDI, S., ROMA, M., AND TOINT, P. L. 1999. Solving the trust-region subproblem using the Lanczos method. *SIAM J. Optim. 9,* 2, 504–525.

GOULD, N. N., ORBAN, D., AND TOINT, P. L. 2002. GALAHAD—A library of thread-safe Fortran 90 packages for large-scale nonlinear optimization. Tech. Rep. RAL-TR-2002-014, Rutherford Appleton Laboratory, Chilton, England. Available from `http://galahad.rl.ac.uk/galahad-www`.

GRODZEVICH, O. AND WOLKOWICZ, H. 2007. Regularization using a parameterized trust region subproblem. *Math. Programming*.

HAGER, W. W. 2001. Minimizing a quadratic over a sphere. *SIAM J. Optim. 12*, 188–208.

HANKE, M. AND HANSEN, P. C. 1993. Regularization methods for large-scale problems. *Surv. Math. Ind. 3*, 253–315.

HANSEN, P. C. 1990. The Discrete Picard Condition for discrete ill-posed problems. *BIT 30*, 658–672.

HANSEN, P. C. 1994. Regularization Tools: A MATLAB package for analysis and solution of discrete ill-posed problems. *Numer. Algo. 6*, 1–35.

HANSEN, P. C. 1998. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM, Philadelphia.

HSL. 2004. HSL 2004: A collection of Fortran codes for large scale scientific computation. Available from `www.cse.clrc.ac.uk/nag/hsl`.

LAWSON, C. L. AND HANSON, R. J. 1995. *Solving Least Squares Problems*. SIAM, Philadelphia.

LEHOUCQ, R. B., SORENSEN, D. C., AND YANG, C. 1998. *ARPACK User's Guide: Solution of Large Scale Eigenvalue Problems by Implicitly Restarted Arnoldi Methods*. SIAM, Philadelphia.

LUCIDI, S., PALAGI, L., AND ROMA, M. 1998. On some properties of quadratic programs with a convex quadratic constraint. *SIAM J. Optim. 8*, 105–122.

MORÉ, J. J. AND SORENSEN, D. C. 1983. Computing a trust region step. *SIAM J. Sci. Stat. Comput. 4,* 3, 553–572.

NATTERER, F. 1986. *The Mathematics of Computerized Tomography*. John Wiley & Sons, New York.

NEUMAIER, A. 1998. Solving ill-conditioned and singular linear systems: a tutorial on regularization. *SIAM Review 40,* 3, 636–666.

NOCEDAL, J. AND WRIGHT, S. J. 1999. *Numerical Optimization*. Springer, New York.

NOLET, G. 1987. *Seismic Tomography, with Applications in Global Seismology and Exploration Geophysics*. D. Reidel Publishing Company, Dordrecht.

PARKS, T. W. AND BURRUS, C. S. 1987. *Digital Filter Design*. John Wiley & Sons, New York.

PARLETT, B. N. 1980. *The Symmetric Eigenvalue Problem*. Prentice-Hall, Englewood Cliffs NY.

PHAM-DINH, T. AND LE-THI, H. A. 1998. A D.C. optimization algorithm for solving the trust-region subproblem. *SIAM J. Optim. 8,* 2, 476–505.

RENDL, F. AND WOLKOWICZ, H. 1997. A semidefinite framework for trust region subproblems with applications to large scale minimization. *Math. Prog. 77,* 2, 273–299.

ROJAS, M. 1998. A large-scale trust-region approach to the regularization of discrete ill-posed problems. Ph.D. dissertation, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, Tech. Rep. TR98-19 (May).

ROJAS, M., SANTOS, S. A., AND SORENSEN, D. C. 2000. A new matrix-free algorithm for the large-scale trust-region subproblem. *SIAM J. Optim. 11,* 3, 611–646.

ROJAS, M. AND SORENSEN, D. C. 2002. A trust-region approach to the regularization of large-scale discrete forms of ill-posed problems. *SIAM J. Sci. Comput. 23,* 6, 1843–1861.

ROJAS, M. AND STEIHAUG, T. 2002. An interior-point trust-region-based method for large-scale nonnegative regularization. *Inverse Problems 18,* 5, 1291–1307.

SLEIJPEN, G. L. AND VAN DER VORST, H. A. 1996. A Jacobi-Davidson iteration method for linear eigenvalue problems. *SIAM J. Matrix Anal. Appl. 17*, 401–425.

SORENSEN, D. C. 1982. Newton's method with a model trust region modification. *SIAM J. Numer. Anal. 19,* 2, 409–426.

SORENSEN, D. C. 1992. Implicit application of polynomial filters in a k-step Arnoldi method. *SIAM J. Matrix Anal. Appl. 13,* 1, 357–385.

SORENSEN, D. C. 1997. Minimization of a large-scale quadratic function subject to a spherical constraint. *SIAM J. Optim. 7,* 1, 141–161.

STEIHAUG, T. 1983. The conjugate gradient method and trust regions in large scale optimization. *SIAM J. Numer. Anal. 20,* 3, 626–637.

SYMES, W. W. 1993. A differential semblance criterion for inversion of multioffset seismic reflection data. *J. Geophys. Res. 98*, 2061–2073.

THE MATHWORKS, INC. 2000. *MATLAB: The Language of Technical Computing. Using MATLAB Version 6*. Natick, Massachussetts.

TIKHONOV, A. N. 1963. Regularization of incorrectly posed problems. *Soviet Math. 4*, 1624–1627.

TOINT, P. L. 1981. Towards an efficient sparsity exploiting Newton method for minimization. In *Sparse Matrices and Their Uses*, I. Duff, Ed., Academic Press, London, 57–88.

Received August 2003; revised October 2004, October 2006, March 2007; accepted April 2007