DELFT UNIVERSITY OF TECHNOLOGY

REPORT 10-12

Large-Scale Eigenvalue Problems in Trust-Region Calculations

Marielba Rojas, Bjørn H. Fotland, and Trond Steihaug

ISSN 1389-6520

Reports of the Department of Applied Mathematical Analysis

Delft 2010

Large-Scale Eigenvalue Problems in Trust-Region Calculations

Marielba Rojas* Bjørn H. Fotland† Trond Steihaug‡

April 8, 2010

Abstract

The Trust-Region Subproblem of minimizing a quadratic function subject to a norm constraint arises in the context of trust-region methods in optimization and in the regularization of discrete forms of ill-posed problems. In recent years, methods and software for large-scale trust-region subproblems have been developed that require the solution of a large bordered eigenvalue problem at each iteration. In this work, we describe the bordered eigenvalue problems, the computational challenges in solving them, and present some approaches for their efficient solution by means of Krylov subspace methods for linear and nonlinear eigenvalue problems.

1 The Problem

We consider the eigenvalue problem

$$\begin{pmatrix} \alpha_{\mathbf{k}} & g^T \\ g & H \end{pmatrix} \begin{pmatrix} \nu \\ u \end{pmatrix} = \lambda \begin{pmatrix} \nu \\ u \end{pmatrix} \tag{1}$$

where H is an $n \times n$ real, symmetric matrix, g is an n-dimensional non-zero real vector, and α_k is a real parameter that is iteratively updated such that

*Delft Institute of Applied Mathematics, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands (marielba.rojas@tudelft.nl).

†This author's contributions were developed while he was with the Department of Informatics, University of Bergen in Bergen, Norway. The author is currently with WesternGeco in Stavanger, Norway.

‡Department of Informatics, University of Bergen, Postboks 7803, N-5020 Bergen, Norway (trond@ii.uib.no).

$\{\alpha_k\}$ is a convergent sequence. We assume that H is large and not explicitly available but that matrix-vector multiplications with H can be computed efficiently. We are interested in computing the algebraically smallest eigenvalue and a corresponding eigenvector with non-zero first component.

Eigenvalue problems of type (1) are related to the problem

$$\begin{aligned} \min \quad & \tfrac{1}{2} x^T H x + g^T x \\ s.t. \quad & \|x\| \leq \Delta \end{aligned} \tag{2}$$

with H, g as above, Δ a positive scalar, and $\|\cdot\|$ the Euclidean norm.

Problem (2) is known in optimization as the trust-region subproblem arising in connection with the trust-region globalization strategy. Methods that use such strategy are known as trust-region methods and are state-of-the-art iterative techniques for solving nonlinear programming problems. Trust-region methods require the solution of a problem of type (2) at each iteration. A special case of (2) can be used as a regularization approach for the treatment of discrete forms of ill-posed problems. In this case, $H = A^T A$ and $g = -A^T b$, with A a discretized form of an operator in an ill-posed problem and b a data vector perturbed by noise such that b is not in the range of A. It is well-known (cf. [1, 16]) that the resulting regularization approach is equivalent to Tikhonov regularization [22]. In particular, the Lagrange multiplier associated with the constraint in (2) corresponds to the Tikhonov regularization parameter.

Optimality conditions for problem (1) are presented in Lemma 1.1 from [18]. For a proof, we refer the reader to [18].

Lemma 1.1 ([18]) *A feasible vector $x_* \in \mathbb{R}^n$ is a solution to (2) with corresponding Lagrange multiplier λ_* if and only if x_*, λ_* satisfy*

(i) $(H - \lambda_* I)x_* = -g$.

(ii) $H - \lambda_* I$ is positive semidefinite.

(iii) $\lambda_* \leq 0$.

(iv) $\lambda_*(\Delta - \|x_*\|) = 0$.

Problem (2) can be efficiently solved by means of the Newton iteration proposed in [10], when the Cholesky factorization of matrices $H - \sigma I$ can be efficiently computed. In the large-scale case, different strategies are required. Until the mid 1990's, the only efficient method for solving large-scale trust-region subproblems was the Conjugate-Gradient-based method

proposed in [21], and this is still one of the best choices in the context of trust-region methods. However, the method is a so-called *approximate* technique since it does not aim to satisfy the optimality conditions, a property that may be desirable in practice. Moreover, regularization problems can prove challenging for the technique.

In recent years, new *nearly-exact* iterative methods have appeared [2, 12, 14, 15, 20] that aim to compute solutions for large-scale trust-region subproblems that satisfy the optimality conditions. At the same time, new approximate methods have also been proposed in [4, 5]. Most of the new techniques are also suitable for solving regularization problems.

One approach to develop large-scale methods is based on the following result. It can be shown (see [13]) that there always exists an optimal value of a scalar parameter α such that a solution x to (2) can be computed from a solution $y = (1, x^T)^T$ of

$$\begin{array}{ll} \min & \frac{1}{2} y^T B_\alpha y \\ s.t. & y^T y \leq 1+\Delta^2 \\ & e_1^T y = 1 \end{array} \tag{3}$$

where $B_\alpha = \begin{pmatrix} \alpha & g^T \\ g & H \end{pmatrix}$.

Notice that a solution to (3) is an eigenvector with non-zero first component corresponding to the smallest eigenvalue of B_α. Since the optimal value of α is not known in general, this suggests that we can design an iteration for updating α based on the solution of eigenvalue problems for B_α. LSTRS is one such iteration. Notice, moreover, that the eigenvalues of H and B_α are related. Due to Cauchy Interlace Theorem (cf. [11]), the eigenvalues of B_α interlace the eigenvalues of H. In particular, the smallest eigenvalue of B_α is a lower bound for the smallest eigenvalue of H. This implies that solving an eigenvalue problem for the smallest eigenvalue of B_α and a corresponding eigenvector with non-zero first component, produces x and λ that automatically satisfy optimality conditions (i) and (ii) in Lemma 1.1, for any value of α.

In [2, 12, 14, 15, 20], the main computation per iteration is the solution of an eigenvalue problem of type (1) for an eigenpair associated with the smallest eigenvalue of the *bordered* matrix B_{α_k}. In [15], an additional eigenpair (corresponding to another eigenvalue) is needed. In both families of methods, the eigenpairs are used in a rational interpolation scheme for updating α_k.

In this work, we discuss the features of eigenvalue problems of type (1) in the context of methods for large-scale trust-region subproblems, including those arising in regularization. The presentation is organized as follows. In Section 2, we discuss methods for solving (1). In Section 3, we discuss the features of problem (1) and show their implication for the numerical solution of the problem. Section 4 presents concluding remarks.

2 Methods

Large-scale symmetric eigenvalue problems of type (1) can be efficiently solved with several techniques. Here, we mention three approaches: the Implicitly Restarted Lanczos Method (IRLM) [9, 19], the Jacobi-Davidson Method [17], and the Nonlinear Lanczos (Arnoldi) Method (NLLM) [23]. All methods are matrix-free in the sense that they use the matrix in matrix-vector multiplications only.

At this moment, the performance of the Jacobi-Davidson Method for solving the eigenproblems of interest in trust-region calculations is yet to be studied. Therefore, we focus on the IRLM and the NLLM. The main features of the IRLM include: limited-memory, the ability to compute several eigenpairs at a time, and the possibility of choosing *one* initial vector. Features of the NLLM include: limited-memory through the use of restarts, the ability to compute only one eigenpair at a time, and the possibility of choosing *several* initial vectors.

To the best of our knowledge, the use of eigensolvers other than the IRLM in combination with the methods in [2, 12] has not been tested. For this reason, our discussion focuses on the efficient solution of problem (1) within LSTRS. The IRLM and NLLM have been used for this purpose in [15] and [8], respectively. To efficiently solve problems of type (1), any matrix-free eigensolver must be able to handle certain issues arising in the context of trust-region calculations. These issues are the subject of the next section.

3 Numerical Issues and Challenges

In this section, we discuss the main issues in solving (1) efficiently and accurately in the context of trust-region calculations. We observe that most of the issues discussed here also apply to [2, 12]. Next, we discuss the issues and describe our strategies to handle them in LSTRS.

3.1 Eigenvalues close to zero

The solution of (1) is particularly challenging for methods based on matrix-vector multiplications when the eigenvalues of interest are close to zero since every matrix-vector product will annihilate components of the resulting vector precisely in the desired directions. In regularization, it is often the case that the eigenvalues of interest are close to zero.

In [15], this situation is handled by means of a Tchebyshev Spectral Transformation. We construct a Tchebyshev polynomial T that is as large as possible on λ_1 and as small as possible on an interval containing the remaining eigenvalues of B_α. We then compute the eigenvalues of $T(B_\alpha)$ instead of the eigenvalues of B_α. In LSTRS, a polynomial of degree ten is used. Hence, the number of matrix-vector products increases accordingly. However, the convergence of IRLM is usually enhanced in this way and in the context of LSTRS for regularization this is often the only way to handle certain challenging ill-posed problems (cf. [16]). After convergence, the eigenvalues of B_α are recovered via the Rayleigh quotients with the converged eigenvectors. No special strategy is used to handle this case when the NLLM is used.

3.2 Clustered eigenvalues

In regularization problems, the singular values of A are usually clustered and very close to zero with no apparent gap (cf. [6]). The eigenvalues of $H = A^T A$ inherit this feature. The interlacing properties discussed in Section 1 imply that if the smallest q eigenvalues of H are small and clustered then, eigenvalues 2 through q of B_α will also be small and clustered.

Recall now that λ_1, the smallest eigenvalue of B_α, is a lower bound for δ_1, the smallest eigenvalue of H. The distance between λ_1 and δ_1 depends on the value of α, which in turn depends on the value of Δ in LSTRS. For values of Δ smaller than a certain critical value, the smallest eigenvalue of B_α is well-separated from the rest and Lanczos-type methods can compute it very efficiently. For larger values of Δ, λ_1 is well-separated from δ_1 only at early iterations. As the process converges (and α_k approaches the optimal value), λ_1 gets closer to δ_1 and, in regularization, to the cluster. Figure 1 illustrate this situation for a test problem from [7].

Large values of Δ are common in regularization. In practice, this often means that the number of vectors used by the IRLM or by the NLLM must be increased. This is the only strategy we follow at this moment.

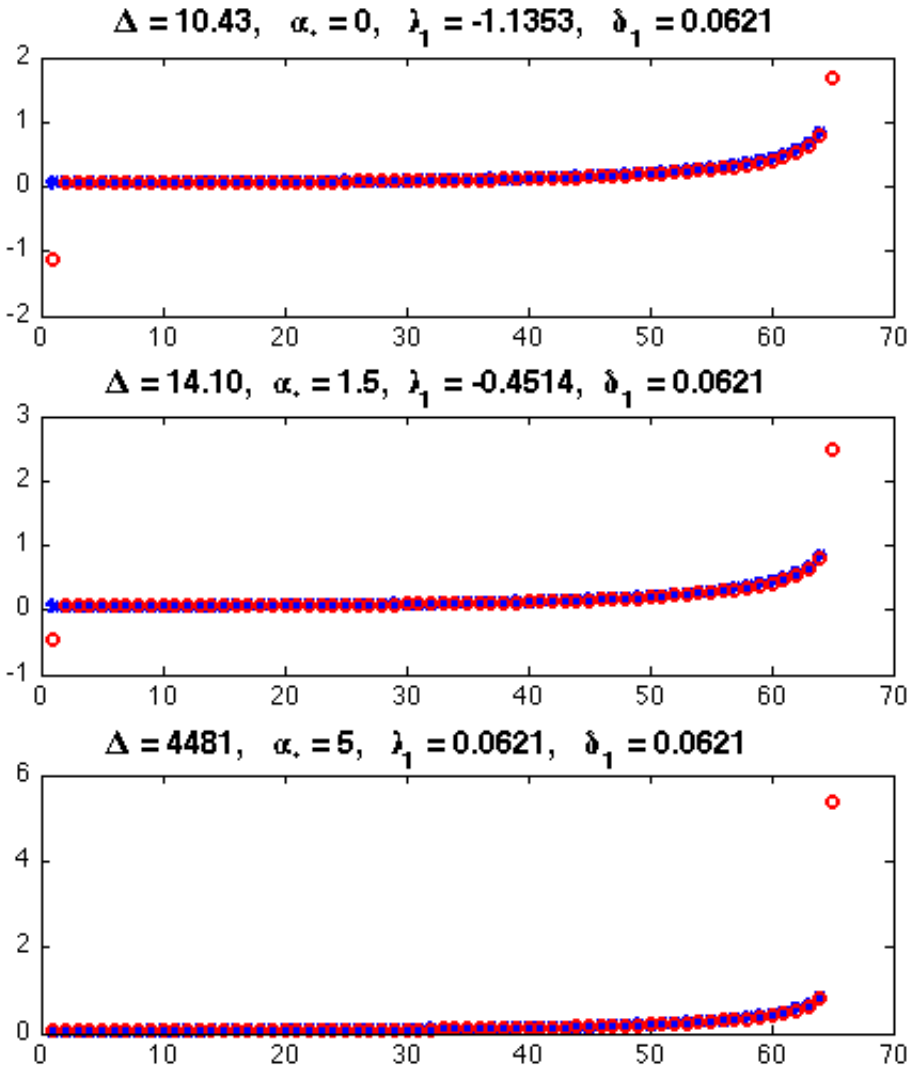

Figure 1: Eigenvalues of H (dot) and B_α (circle) for different values of α_k. Problem **heat** from [7].

3.3 Accuracy, Stability, and Performance

In this section, we investigate accuracy, stability, and performance of LSTRS in the presence of perturbed (inexact) eigenpairs. Since the eigenvalue problems are embedded in an outer iteration, it is important to know how accurately the problems must be solved and how sensitive the LSTRS solution is to perturbations in the eigenpairs. It is also interesting to investigate the effect of perturbations in the eigenpairs on the performance of the method. A preliminary numerical study was conducted in [3] to investigate these issues. Here, we have extended that study.

We performed a set of experiments in MATLAB, in which perturbations were added to the eigenvalue and eigenvector computed at each iteration of LSTRS. The values of Δ and of LSTRS parameters were chosen to force boundary solutions for problem (2). The test problems were taken from [7].

Instances of problem (2) were solved with LSTRS for $H = A^T A$ and $g = -A^T b$, with A and $m \times n$ matrix and b an m-dimensional vector such that $b = b_o + s$, with b_o a noise-free data vector available from the test set, and s a vector of Gaussian noise. The noise level $\|s\|/\|b\|$ was 10^{-2}. Note that although several problems from the test set were tried, the results were similar for all of them. Therefore, we report results only for problem **shaw** with $m = n = 100$.

The results correspond to perturbations that were *relative* to the unperturbed quantities. They were generated as follows. Given the *unperturbed* eigenvalue λ_1, the *perturbed* eigenvalue $\widetilde{\lambda_1}$ was constructed as $\widetilde{\lambda_1} = \lambda_1(1+\varepsilon\rho)$, with $\varepsilon \in (0, 1)$ and $\rho \in (-\varepsilon/2, 1 - \varepsilon/2)$ a random number from a uniform distribution. The eigenvector perturbations were generated in a similar way. Given the *unperturbed* eigenvector y_1, the *perturbed* eigenvector $\widetilde{y_1}$ satisfied $\widetilde{y_1} = y_1 + \varepsilon y_1 \times r$, with $\varepsilon \in (0, 1)$ and r a random vector with components in $(-\varepsilon/2, 1 - \varepsilon/2)$ from a uniform distribution. Here, $\times$ denotes componentwise multiplication. For eigenvalue perturbations, we generated a set of random numbers ρ_i, $i = 1, \cdots, 100$ and we sorted them in increasing order. For eigenvector perturbations, we generated a set of random vectors r_i, $i = 1, \cdots, 100$ and we sorted them according to increasing order of their 2-norm. Notice that *absolute* perturbations, ie. perturbations not related to the size of the perturbed numbers, were also used and they yielded similar results for all experiments.

In the remainder of the section, x_p and x_u denote the solutions computed by LSTRS using perturbed and unperturbed eigenvalues (or eigenvectors), respectively.

3.3.1 Experiment 1

In the first part of the experiment, we investigated the effect of eigenvalue perturbations on the solution computed by LSTRS. We tested two eigensolvers: `eig` (QR method) and `eigs` (MATLAB's interface to ARPACK). We used $\varepsilon = 10^{-2}$, which was the highest value for which we could compute boundary solutions for all the (100) trust-region subproblems. For $i = 1, \cdots, 100$, we used $(\widetilde{\lambda_1})_i$ corresponding to perturbation ρ_i and the unperturbed eigenvector y_1 for each iteration of LSTRS. In Figure 2, we show the distribution of $\|x_p\|$, the relative error in x_p with respect to x_u, and the standard deviation of the perturbations over the LSTRS iterations for each experiment, when `eig` is used as eigensolver.

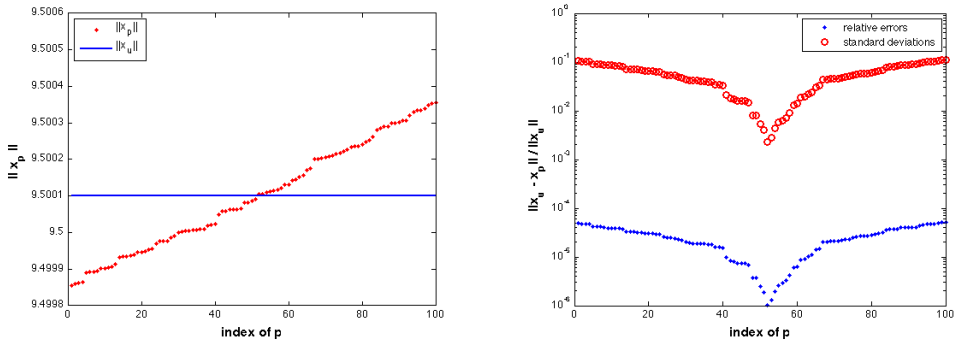

Figure 2: Distribution of $\|x_p\|$ (left) and of $\|x_p - x_u\|/\|x_u\|$ (right). On the right, relative errors (dots) and standard deviations of perturbations (circle). Eigensolver: `eig`.

On the left plot, we can observe a linear dependence of the norm of $\|x_p\|$ on the value of the perturbation. On the right plot, we observe the distribution of the relative error in x_p with respect to x_u. Interestingly enough, this distribution (dots) mimics the distribution of the standard deviations of the perturbations (circles). The relative errors were all under the order of 10^{-4}. Therefore, according to these results, LSTRS is stable with respect to perturbations in the smallest eigenvalue of B_α when `eig` is used as eigensolver.

In Figure 3, we show the results obtained for `eigs`. In this case, except for one outlier, the behavior of the method was similar to the case when `eig` was used. The relative errors were all under the order of 10^{-3}. For both `eig` and `eigs`, other values of ε yielded similar results.

In the second part of the experiment, we investigated the effect of eigen-

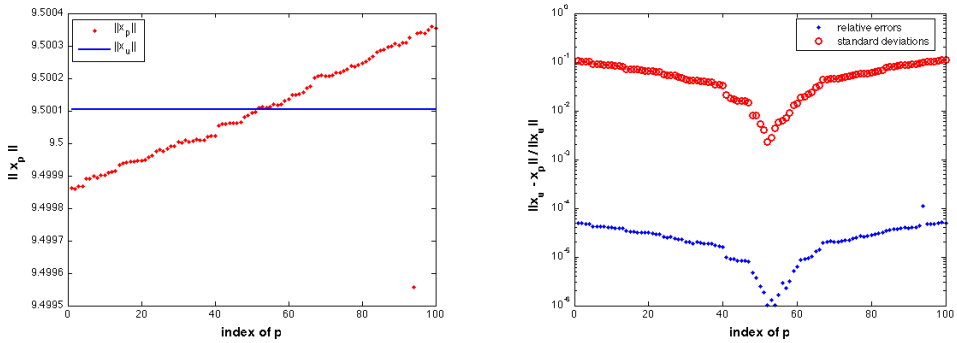

Figure 3: Distribution of $\|x_p\|$ (left) and of $\|x_p - x_u\|/\|x_u\|$ (right). On the right, relative errors (dots) and standard deviations of perturbations (circle). Eigensolver: `eigs`.

vector perturbations on the solution computed by LSTRS. As before, we tested the eigensolvers `eig` and `eigs`. We used $\varepsilon = 10^{-5}$, which was the highest perturbation for which we could compute boundary solutions for the 100 tests problems. For $i = 1, \cdots, 100$, we used $(\widetilde{y_1})_i$ corresponding to perturbation r_i and the corresponding unperturbed eigenvalue λ_1 for each iteration of LSTRS. In this case, we did not observe a correlation between $\|x_p\|$ and the norm of the perturbation, for any of the eigensolvers.

3.3.2 Experiment 2

In this experiment, we repeated Experiment 1 for ε of order 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5}. In Figure 4, we present the results for `eig` and `eigs`, for eigenvector perturbations. We report the value of ε, the average number of LSTRS iterations (Iter), the average number of calls to the eigensolver (Solves), and the maximum relative error in x_p with respect to x_u. For `eigs`, we also report the average number of matrix-vector products (MVP). As a point of reference, we include the data for the unperturbed solution.

We observe a deterioration in performance as the size of the perturbation increases. However, our previous conclusion about stability of LSTRS still holds when the eigenvectors are perturbed since the relative error in x_p remains bounded and is related to the perturbation. In Figure 4, we can see that the maximum relative error in x_p with respect to x_u was of the same order as ε for most perturbations and for both eigensolvers. As the size of the perturbation increases, however, this behavior no longer holds. We can see that the deterioration in the relative error occurs earlier when `eigs`

is used. Moreover, the deterioration in performance is also more dramatic than before. This was expected since, unlike `eig`, `eigs` computes eigenpairs only to a given and, in this case, low accuracy.

ε	Iter	Solves	$\max \frac{\|x_p - x_u\|}{\|x_u\|}$
0	7.0	8.0	0
10^{-5}	7.0	8.0	9.57e-05
10^{-4}	7.0	8.0	9.56e-04
10^{-3}	11.0	11.4	8.53e-03
10^{-2}	16.1	16.2	1.29e-01
10^{-1}	28.0	28.0	5.87e-01

ε	Iter	Solves	MVP	$\max \frac{\|x_p - x_u\|}{\|x_u\|}$
0	7.0	8.0	1026.0	0
10^{-5}	7.1	8.7	1000.5	1.81e-04
10^{-4}	13.8	15.4	2218.1	7.22e-04
10^{-3}	22.8	23.4	3667.5	6.83e-03
10^{-2}	26.4	27.4	4213.8	1.58e-01
10^{-1}	27.6	28.1	4116.5	5.93e-01

Figure 4: Performance results and maximum relative error in x_p for perturbed eigenvectors. Eigensolvers: `eig` (left) and `eigs` (right).

We conducted the same experiment for eigenvalue perturbations. Only negligible variations in performance and in relative error were observed. This was to be expected since the eigenvector y_1 influences the solution in a more direct way than the eigenvalue λ_1. Namely, the last n (normalized) components of y_1 form the approximate solution to problem (2) at each iteration of LSTRS. In particular, in the last LSTRS iteration the last n (normalized) components of y_1 form the solution.

3.4 Efficiency

We now discuss the performance of LSTRS in terms of number of matrix-vector products (MVP). Comparisons of LSTRS with other state-of-the-art methods for large-scale trust-region subproblems seem to indicate an advantage for LSTRS [15], especially for regularization problems. This was to be expected since LSTRS was designed with this kind of problems in mind. Recently [8], great reductions in the number of MVP have been obtained, at a moderate cost in storage, by means of the NLLM. We expect that further improvements are possible, for example, by incorporating the Jacobi-Davidson method in combination with preconditioning. This is the object of current research.

4 Concluding Remarks

We have described an important class of eigenvalue problems arising in methods for large-scale trust-region subproblems in optimization and regu-

larization. Our presentation focused on the special features that make these problems particularly difficult to solve in practice, and the approaches we have taken to deal with the difficulties. Future work in this area includes the use of the Jacobi-Davidson method in combination with preconditioning, further experiments concerning accuracy, stability, and performance issues, and the design of strategies to guarantee the computation of all the desired eigenpairs and of at least one eigenvector with the desired structure.

References

[1] L. Eldén. Algorithms for the regularization of ill-conditioned least squares problems. *BIT*, 17:134–145, 1977.

[2] C. Fortin and H. Wolkowicz. The trust-region subproblem and semidefinite programming. *Optim. Methods Softw.*, 19(1):41–67, 2004.

[3] B.H. Fotland. A matrix-free method for the regularisation with unrestricted variables. Master's thesis, Department of Informatics, University of Bergen, Bergen, Norway, June 2008.

[4] N.I.M. Gould, S. Lucidi, M. Roma, and Ph.L. Toint. Solving the trust-region subproblem using the Lanczos method. *SIAM J. Optim.*, 9(2):504–525, 1999.

[5] W.W. Hager. Minimizing a quadratic over a sphere. *SIAM J. Optim.*, 12:188–208, 2001.

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

[7] P.C. Hansen. Regularization Tools version 4.0 for Matlab 7.3. *Numer. Algo.*, 46:189–194, 2007.

[8] J. Lampe, M. Rojas, D.C. Sorensen, and H. Voß. Accelerating the LSTRS algorithm. Technical Report 138, Institute of Numerical Simulation, Hamburg University of Technology, Hamburg-Harburg, Germany, July 2009.

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

[10] J.J. Moré and D.C. Sorensen. Computing a trust region step. *SIAM J. Sci. Stat. Comput.*, 4(3):553–572, 1983.

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

[12] F. Rendl and H. Wolkowicz. A semidefinite framework for trust region subproblems with applications to large scale minimization. *Math. Prog.*, 77(2):273–299, 1997.

[13] M. Rojas. A large-scale trust-region approach to the regularization of discrete ill-posed problems. Ph.D. thesis, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, Technical Report TR98-19, May 1998.

[14] 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(3):611–646, 2000.

[15] M. Rojas, S.A. Santos, and D.C. Sorensen. Algorithm 873: LSTRS: MATLAB software for large-scale trust-region subproblems and regularization. *ACM Trans. Math. Software*, 34(2):11, 2008.

[16] M. Rojas and D.C. Sorensen. A trust-region approach to the regularization of large-scale discrete forms of ill-posed problems. *SIAM J. Sci. Comput.*, 23(3):1843–1861, 2002.

[17] G.L.G. Sleijpen and H.A. Van der Vorst. A Jacobi-Davidson iteration method for linear eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 17:401–425, 1996.

[18] D.C. Sorensen. Newton's method with a model trust region modification. *SIAM J. Numer. Anal.*, 19(2):409–426, 1982.

[19] D.C. Sorensen. Implicit application of polynomial filters in a k-step Arnoldi method. *SIAM J. Matrix Anal. Appl.*, 13(1):357–385, 1992.

[20] D.C. Sorensen. Minimization of a large-scale quadratic function subject to a spherical constraint. *SIAM J. Optim.*, 7(1):141–161, 1997.

[21] T. Steihaug. The conjugate gradient method and trust regions in large scale optimization. *SIAM J. Numer. Anal.*, 20(3):626–637, 1983.

[22] A.N. Tikhonov. Regularization of incorrectly posed problems. *Soviet Math.*, 4:1624–1627, 1963.

[23] H. Voss. An Arnoldi method for nonlinear eigenvalue problems. *BIT*, 44:387–401, 2004.