GAUSS QUADRATURE APPLIED TO TRUST REGION COMPUTATIONS

DANIELA CALVETTI * AND LOTHAR REICHEL †

Dedicated to Gene Golub on the occasion of his 70th birthday.

Abstract. This paper discusses the application of the Lanczos process to the solution of large-scale trust-region subproblems. Techniques pioneered by Golub based on Gauss quadrature are applied to derive inexpensively computable upper and lower bounds for quantities of interest. These bounds help determine how many steps of the Lanczos process should be carried out.

AMS subject classifications. 65K05, 65F10, 90C53, 90C06.

Key words. Optimization, quasi-Newton method, Lanczos process, bounds.

1. Introduction.

Let the function $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable and bounded below. We consider the problem of determining a local minimizer x_* of f by the model trust-region approach and discuss the application of the Lanczos process to reduce large-scale problems to more manageable problems of small to medium size. For notational simplicity, we will assume that x_* is the unique local minimizer of f; however, the method described is applicable also when f has several local minima and then is guaranteed to determine one of them.

Let x_c denote the currently available approximation of x_*, and let $\nabla f(x_c) \in \mathbb{R}^n$ and $\nabla^2 f(x_c) \in \mathbb{R}^{n\times n}$ denote the gradient and Hessian of f at $x = x_c$, respectively. The Hessian $\nabla^2 f(x_c)$ is symmetric positive definite in a neighborhood of the unique minimizer x_*. Away from x_*, the Hessian might be indefinite. Throughout this paper $H_c \in \mathbb{R}^{n\times n}$ denotes a symmetric matrix that approximates $\nabla^2 f(x_c)$; in particular, we may let $H_c := \nabla^2 f(x_c)$.

A common approach to the design of globally convergent minimization methods is to approximate the function f in a neighborhood of x_c by a quadratic model of the form

$$m_c(s) := f(x_c) - g_c^T s + \frac{1}{2} s^T H_c s, \qquad g_c := -\nabla f(x_c), \tag{1.1}$$

see, e.g., [2, 3, 11, 14] and references therein. When H_c is positive definite, the minimum of $m_c(s)$ is achieved at

$$s_+ := H_c^{-1} g_c, \tag{1.2}$$

and we obtain the new approximation

$$x_+ := x_c + s_+ \tag{1.3}$$

of x_*. The correction s_+ of x_c also can be obtained by applying Newton's method to the solution of the equation $\nabla m_c(s) = 0$. We therefore refer to s_+ as the Newton step.

In many situations x_+ is a better approximation of x_* than x_c, in particular when x_c is close to x_*. However, it is easy to construct examples for which x_+ is further

*Department of Mathematics, Case Western Reserve University, Cleveland, OH 44106. E-mail: dxc57@po.cwru.edu. Research supported in part by NSF grant DMS-0107841.

†Department of Mathematical Sciences, Kent State University, Kent, OH 44242. E-mail: reichel@math.kent.edu. Research supported in part by NSF grant DMS-0107858.

away from x_* than x_c. The model trust-region approach to minimization seeks to avoid the latter situation by imposing a constraint on the size of the correction s_+ of x_c. Thus, we replace the unconstrained minimization of the quadratic model (1.1) by the constrained minimization problem

$$\min_{\|s\|\leq\delta_c} m_c(s), \tag{1.4}$$

where δ_c is a positive constant. Here and throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm.

The minimization problem (1.4) is commonly referred to as the *trust-region subproblem*, the set $\{x : \|x - x_c\| \leq \delta_c\}$ as the *trust-region* and the constant δ_c as the *trust-region radius*. Recent discussions of the trust-region subproblem and of numerical methods for its solution are presented by Conn et al. [2], Kelley [11, Section 3.3], and Nocedal and Wright [14, Chapter 4].

The present paper is concerned with the computation of an approximate solution of trust-region subproblems (1.4) of size so large that factorization of the matrix H_c is unfeasible or undesirable. Large-scale trust region subproblems arise in many applications, see, e.g., Conn et al. [2, Section 1.3], and a variety of numerical methods are available for their approximate solution; see, e.g., Conn et al. [2], Gould et al. [8], Hager [9], Kelley [11, Section 3.3], Nocedal and Wright [14, Chapter 4], Rojas et al. [15], Steihaug [17] and Toint [18]. Several of these methods are based on the Lanczos process and require the computed approximate solution of (1.4) to live in a Krylov subspace of the form

$$\mathcal{K}_\ell(H_c, g_c) = \mathrm{span}\{g_c, H_c g_c, H_c^2 g_c, \ldots, H_c^{\ell-1} g_c\}. \tag{1.5}$$

Thus, the n-dimensional minimization problem (1.4) is replaced by the ℓ-dimensional minimization problem

$$\min_{\substack{s\in\mathcal{K}_\ell(H_c,g_c)\\ \|s\|\leq\delta_c}} m_c(s), \tag{1.6}$$

where in general ℓ is chosen to be much smaller than n.

It is the purpose of the present paper to describe how application of the truncated Lanczos method allows the computation of upper and lower bounds for the quadratic model $m_c(s)$ and of the norm of the solution $s(\mu)$. These bounds are helpful for determining how large the dimension ℓ of the Krylov subspace $\mathcal{K}_\ell(H_c, g_c)$ should be chosen in (1.6).

The bounds for $m_c(s)$ and $s(\mu)$ are computed by using the connection between the Lanczos process and Gauss quadrature. This technique to determine bounds was proposed by Golub and von Matt [7] in their discussion of equality constrained minimization problems for quadratic forms (1.1) with a positive definite matrix H_c. Extensions of this technique are presented by Golub and Meurant [5].

This paper is organized as follows. Section 2 reviews properties of the trust-region subproblem (1.4) and discusses Steihaug's method [17] for the approximate solution of large-scale trust-region subproblems. Section 3 outlines the Lanczos process and its application to the solution of large linear systems of equations. Upper and lower bounds for the norm of the solution $s(\mu)$ are expressed in terms of certain quadrature rules of Gauss-type in Section 4. This allows us to define a "trust-ribbon," i.e., a ribbon-like set that contains the curve

$$\{(\mu, \|s(\mu)\|) : \mu \geq 0\} \tag{1.7}$$

in its closure. Section 5 applies Gauss-type quadrature rules to determine upper and lower bounds for $m_c(s(\mu))$. These bounds allow us to define a "model-ribbon" that contains the curve

$$\{(\mu, m_c(s(\mu))) : \mu \geq 0\} \tag{1.8}$$

in its closure. The widths of the trust- and model-ribbons decrease as the number of Lanczos steps ℓ increases. Section 6 discusses how the ribbons can be used to help determine how many steps ℓ of the Lanczos process to carry out. A few computed examples are reported in Section 7, and Section 8 contains concluding remarks.

Most of this paper is concerned with the situation when the matrix H_c is positive definite, and unless explicitly stated otherwise we assume that H_c has this property. We will comment on how to handle indefinite matrices H_c in Sections 6 and 7; see also Conn et al. [2, Section 7.5.4] for a discussion of this topic.

2. The trust-region subproblem.

The following characterization of the solution of the constrained minimization problem (1.4), due to Gay [4] and Sorensen [16], is well known:

PROPOSITION 2.1. *Let $f(x_c) \in \mathbb{R}$ and $g_c := -\nabla f(x_c) \in \mathbb{R}^n$, and assume that the matrix $H_c \in \mathbb{R}^{n\times n}$ is symmetric. The vector $s \in \mathbb{R}^n$ solves the constrained minimization problem (1.4) if and only if there is a scalar $\mu \geq 0$, such that*

$$(H_c + \mu I)s = g_c, \qquad (\delta_c - \|s\|)\mu = 0, \tag{2.1}$$

and the matrix $H_c + \mu I$ is positive semidefinite. If H_c is positive definite, then the constrained minimization problem (1.4) has a unique solution.

Proof. See, e.g., Sorensen [16, Section 2]. □

Assume that H_c is positive definite, and that the global minimizer s_+ of the quadratic model (1.1) is outside the trust-region. Define the family of vectors

$$s(\mu) := (H_c + \mu I)^{-1} g_c, \qquad \mu \geq 0. \tag{2.2}$$

Then the solution of the trust-region subproblem (1.4) is given by $s(\mu_+)$, where $\mu_+ > 0$ is the largest solution of the equation

$$\|s(\mu)\| = \delta_c. \tag{2.3}$$

One of the simplest and most elegant methods for the approximate solution of large-scale trust-region subproblems is due to Steihaug [17]. A closely related method also has been proposed by Toint [18]. Steihaug and Toint apply the conjugate gradient method to the minimization of the quadratic model (1.1). Starting with the initial approximate solution $s_0 := 0$, the conjugate gradient method generates a sequence of approximate solutions of the trust-region subproblem (1.4), denoted by $s_1, s_2, s_3, \ldots$, of increasing norm until an approximate solution, say s_{j+1}, of norm larger than δ_c has been determined. Steihaug [17] then chooses a convex combination of s_j and s_{j+1} of norm δ_c, denoted by $\hat{s}_{j+1}$, as an approximate solution of the trust-region subproblem (1.4). Having determined $\hat{s}_{j+1}$, either the trust-region radius is modified or a new trust-region subproblem with a trust-region centered at $x_+ := x_c + \hat{s}_{j+1}$ is solved. Recent discussions of this method are provided by Conn et al. [2, Sections 7.5.1-7.5.2], Nocedal and Wright [14, Chapter 4] and Yuan [19].

For $1 \leq \ell \leq j$, the approximate solution s_ℓ determined by the methods of Steihaug and Toint solves the minimization problem (1.6). Convergence of the methods towards a local minimum of f follows from the fact that the negative gradient g_c is in every Krylov subspace. However, under some circumstances the convergence may be slow. For instance, note that

(i) when the Newton step s_+ is inside the trust-region, all approximate solutions s_ℓ of the trust-region subproblem determined by the conjugate gradient method lie inside the trust-region. In this situation, one has to decide on a suitable number of steps to be carried out by the conjugate gradient method before turning to the solution of the next trust-region subproblem. The computation of many conjugate gradient steps for one trust-region subproblem may increase the computational work unnecessarily.

(ii) when s_+ is outside the trust-region, but the number of steps, $j+1$, of the conjugate gradient method required to determine an approximate solution s_{j+1} not in the trust-region is large, it may be advantageous to carry out fewer than $j+1$ steps before turning to the next trust-region subproblem.

(iii) when H_c is indefinite, Steihaug's method typically determines a correction of x_c in a Krylov subspace $\mathcal{K}_{j+1}(H_c, g_c)$ of low dimension. For instance, when the Cauchy point is outside the trust-region, we may find that $\|s_1\| > \delta_c$. Then $j = 0$ and Steihaug's method simplifies to a steepest descent method. If this happens for many trust-region subproblems (1.4), then the convergence of Steihaug's method towards a local minimum of $f(x)$ can be slow.

The above aspects of Steihaug's method have been discussed recently by Conn et al. [2, p. 206 and p. 221]. The difficulty (iii) motivates Conn et al. [2, Section 7.5.4] to describe an approach based on the truncated Lanczos method. This approach allows the dimension ℓ of the Krylov subspace in the minimization problem (1.4) to be chosen larger than in Steihaug's method. The larger dimension of the Krylov subspace can increase the rate of convergence significantly. We illustrate how the computed bounds for $m_c(s)$ and $s(\mu)$ help determine when to terminate the iterations in the cases (i) and (ii), or equivalently how to the dimension ℓ in the minimization problem (1.6).

We remark that in software for the minimization of nonlinear functions f based on the trust-region method, the norm in the trust-region subproblem (1.4) is sometimes chosen to be a weighted Euclidean norm $\|s\|_D := (s^T D s)^{1/2}$, $s \in \mathbb{R}^n$, where D is a positive definite diagonal matrix; see, e.g., Dennis and Schnabel [3, Chapter 7] and Moré [12] for discussions on the choice of D. For ease of notation, we present our method for the case when D is the identity matrix; however, it is straightforward to incorporate a positive definite diagonal matrix D by rescaling the variables.

3. The Lanczos process. We briefly review the Lanczos process and its application to the computation of approximate solutions of linear systems of equations of the form

$$(H_c + \mu I)s = g_c, \qquad g_c := -\nabla f(x_c), \tag{3.1}$$

where $\mu \geq 0$. Application of ℓ steps of the Lanczos process to the matrix H_c with initial vector $g_c \neq 0$ yields the Lanczos decomposition

$$H_c V_\ell = V_\ell T_\ell + f_\ell e_\ell^T, \tag{3.2}$$

where $V_\ell = [v_1, v_2, \ldots, v_\ell] \in \mathbb{R}^{n\times\ell}$ and $f_\ell \in \mathbb{R}^n$ satisfy $V_\ell^T V_\ell = I$, $V_\ell^T f_\ell = 0$, $v_1 = g_c/\|g_c\|$ and e_ℓ denotes the ℓth axis vector. The matrix

$$T_\ell = \begin{bmatrix} \alpha_1 & \beta_1 & & & & & 0 \\ \beta_1 & \alpha_2 & \beta_2 & & & & \\ & \beta_2 & \alpha_3 & & & & \\ & & & \ddots & & & \\ & & & \ddots & \ddots & \beta_{\ell-2} & \\ & & & & \beta_{\ell-2} & \alpha_{\ell-1} & \beta_{\ell-1} \\ 0 & & & & & \beta_{\ell-1} & \alpha_\ell \end{bmatrix} \in \mathbb{R}^{\ell\times\ell} \tag{3.3}$$

is symmetric and tridiagonal. It follows from (3.2) that

$$T_\ell = V_\ell^T H_c V_\ell, \tag{3.4}$$

and since H_c is positive definite, so is T_ℓ. We may assume that T_ℓ has nonvanishing subdiagonal entries, because otherwise the Lanczos process breaks down and the solution of (3.1) can be computed as a linear combination of the columns v_j of V_ℓ generated before break down; see, e.g., Golub and Van Loan [6, Chapter 9] for details on the Lanczos process. It follows from the relation (3.2) that the columns of V_ℓ form an orthonormal basis of the Krylov subspace $\mathcal{K}_\ell(H_c, g_c)$.

It is convenient to introduce the scalar

$$\beta_\ell := \|f_\ell\|. \tag{3.5}$$

We will assume that

$$\beta_\ell > 0, \tag{3.6}$$

because otherwise the solution $s = s(\mu)$ of (3.1) can be expressed as a linear combination of the columns of V_ℓ.

We determine an approximate solution of (3.1) of the form

$$s_\ell(\mu) := V_\ell z_\ell(\mu), \qquad z_\ell(\mu) \in \mathbb{R}^\ell, \tag{3.7}$$

by solving the Galerkin equations

$$V_\ell^T (H_c + \mu I) V_\ell z_\ell(\mu) = V_\ell^T g_c$$

associated with the linear system (3.1) and the Krylov subspace $\mathcal{K}_\ell(H_c, g_c)$. Using properties of the Lanczos decomposition (3.2), the Galerkin equations simplify to

$$(T_\ell + \mu I) z_\ell(\mu) = \|g_c\| e_1. \tag{3.8}$$

In typical applications, $\ell \ll n$ and the computation of the Lanczos decomposition (3.2), the solution of the linear system (3.8) and the evaluation of $s_\ell(\mu)$ by (3.7) can be carried out with much less computational effort than the computation of the vector $s(\mu)$ defined by (2.2).

4. Gauss quadrature and the trust-ribbon. This section discusses the application of quadrature rules of Gauss-type to the computation of lower and upper bounds for $\|s(\mu)\|$ and $m_c(\mu)$, using techniques advocated by Golub and his collaborators; see, e.g., [1, 5, 7].

Define the functions

$$\eta(\mu) := \|s(\mu)\|, \tag{4.1}$$

$$\psi_\mu(t) := (t+\mu)^{-2}, \tag{4.2}$$

where $\mu \geq 0$ and $t \geq 0$. It follows from (2.2) that

$$\eta^2(\mu) = s(\mu)^T s(\mu) = g_c^T (H_c + \mu I)^{-2} g_c = g_c^T \psi_\mu(H_c) g_c. \tag{4.3}$$

Introduce the spectral factorization

$$H_c = W \Lambda W^T, \tag{4.4}$$

where

$$\Lambda = \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n] \in \mathbb{R}^{n\times n}, \qquad W \in \mathbb{R}^{n\times n}, \qquad W^T W = I,$$

and order the eigenvalues according to

$$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n.$$

Substituting (4.4) and $w = [w_1, w_2, \ldots, w_n]^T := W^T g_c$ into the right-hand side of (4.3) allows us to express $\eta^2(\mu)$ as a Stieltjes integral,

$$\eta^2(\mu) = w^T \psi_\mu(\Lambda) w = \sum_{k=1}^{n} \psi_\mu(\lambda_k) w_k^2 = \int_0^\infty \psi_\mu(t) d\omega(t). \tag{4.5}$$

The distribution function $\omega(t)$ is chosen to be a nondecreasing step function with jumps of height w_k^2 at the eigenvalues λ_k. Thus, the support of the measure $d\omega(t)$ so defined is in the interval $[\lambda_1, \lambda_n]$.

The positive definiteness of H_c implies that the support of the measure $d\omega(t)$ introduced in (4.5) is on the positive real axis. Let $\mathcal{G}_\ell$ denote the ℓ-point Gauss quadrature rule associated with the distribution function $\omega(t)$. The following theorem is concerned with the quadrature error

$$\mathcal{E}_\ell(\psi_\mu) := \int_0^\infty \psi_\mu(t) d\omega(t) - \mathcal{G}_\ell(\psi_\mu). \tag{4.6}$$

THEOREM 4.1. *Assume that (3.6) holds. Then the quadrature error (4.6) satisfies*

$$\mathcal{E}_{\ell-1}(\psi_\mu) > \mathcal{E}_\ell(\psi_\mu) > 0, \qquad \mu \geq 0. \tag{4.7}$$

Moreover, for fixed ℓ the quadrature error $\mathcal{E}_\ell(\psi_\mu)$ decreases monotonically to zero as μ increases.

Proof. The positivity of $\mathcal{E}_\ell(\psi_\mu)$ is a consequence of the following well-known expression for the quadrature error

$$\mathcal{E}_\ell(\psi_\mu) = \frac{\psi_\mu^{(2\ell)}(\tau)}{(2\ell)!} \int_0^\infty \prod_{j=1}^{\ell} (t - t_j)^2 d\omega(t), \tag{4.8}$$

where τ satisfies $\lambda_1 < \tau < \lambda_n$, $\psi_\mu^{(2\ell)}(t)$ denotes the 2ℓth derivative of the function $\psi_\mu(t)$ with respect to t, and the t_j are the nodes of the Gauss quadrature rule. Since $\psi_\mu^{(2\ell)}(t) > 0$ for all $t > 0$ and $\mu \geq 0$, it follows from (4.8) that $\mathcal{E}_\ell(\psi_\mu) > 0$ for $\mu \geq 0$; see, e.g., [1, 5, 7] for details.

The derivative of $\mathcal{E}_\ell(\psi_\mu)$ with respect to μ can be expressed as

$$\frac{d}{d\mu}\mathcal{E}_\ell(\psi_\mu) = \int_0^\infty \frac{d}{d\mu}\psi_\mu(t)d\omega(t) - \mathcal{G}_\ell(\frac{d}{d\mu}\psi_\mu),$$

i.e., the derivative of $\mathcal{E}_\ell(\psi_\mu)$ is the quadrature error obtained when approximating the integral of

$$\frac{d}{d\mu}\psi_\mu(t) = -2(t+\mu)^{-3} \tag{4.9}$$

by the Gauss quadrature rule $\mathcal{G}_\ell(d\psi_\mu/d\mu)$. Replacing the function ψ_μ by its derivative (4.9) in (4.8), and using the fact that the 2ℓth derivative with respect to t of the function (4.9) is negative for $t > 0$, yields

$$\frac{d}{d\mu}\mathcal{E}_\ell(\psi_\mu) < 0, \qquad \mu \geq 0.$$

Since $\lim_{\mu\to\infty} \psi_\mu(t) = 0$ uniformly for $0 \leq t < \infty$, it follows that

$$\lim_{\mu\to\infty} \int_0^\infty \psi_\mu(t)d\omega(t) = 0, \qquad \lim_{\mu\to\infty} \mathcal{G}_\ell(\psi_\mu) = 0.$$

This shows that $\lim_{\mu\to\infty} \mathcal{E}_\ell(\psi_\mu) = 0$. Hanke [10] recently pointed out that the k-point Gauss rule $\mathcal{G}_k$ associated with the measure $d\omega(t)$ for $k < \ell$ also is a Gauss rule associated with the discrete measure defined by the nodes and weights of $\mathcal{G}_\ell$. It follows that $\mathcal{E}_k(\psi_\mu) > \mathcal{E}_\ell(\psi_\mu)$. $\square$

The ℓ-point Gauss rule can be expressed as

$$\mathcal{G}_\ell(\psi_\mu) = \|g_c\|^2 e_1^T \psi_\mu(T_\ell)e_1, \tag{4.10}$$

and the right-hand side can be evaluated in only $\mathcal{O}(\ell)$ arithmetic floating point operations for each value of $\mu \geq 0$; see, e.g., [1, 5, 7] for details. The evaluation method discussed in these references uses the Cholesky factorization of T_ℓ to compute the vector

$$y_\ell(\mu) := (T_\ell + \mu I)^{-1}e_1. \tag{4.11}$$

Let $\hat{\mathcal{G}}_\ell$ denote the $(\ell+1)$-point Gauss-Radau quadrature rule associated with the distribution function $\omega(t)$ and with a node at the origin, and define the quadrature error

$$\hat{\mathcal{E}}_\ell(\psi_\mu) := \int_0^\infty \psi_\mu(t)d\omega(t) - \hat{\mathcal{G}}_\ell(\psi_\mu). \tag{4.12}$$

THEOREM 4.2. *Assume that (3.6) holds. Then the quadrature error (4.12) satisfies*

$$\hat{\mathcal{E}}_{\ell-1}(\psi_\mu) < \hat{\mathcal{E}}_\ell(\psi_\mu) < 0, \qquad \mu \geq 0. \tag{4.13}$$

Moreover for fixed ℓ the quadrature error (4.12) increases monotonically to zero as μ increases.

Proof. The negativity of $\hat{\mathcal{E}}_\ell(\psi_\mu)$ is a consequence of the following well-known expression for the quadrature error

$$\hat{\mathcal{E}}_\ell(\psi_\mu) = \frac{\psi_\mu^{(2\ell+1)}(\hat{\tau})}{(2\ell+1)!} \int_0^\infty t \prod_{j=1}^{\ell} (t - \hat{t}_j)^2 d\omega(t), \tag{4.14}$$

where $\hat{\tau}$ satisfies $0 < \hat{\tau} < \lambda_n$, and the $\hat{t}_j$ are the positive nodes of the Gauss-Radau rule. Using the fact that $\psi_\mu^{(2\ell+1)}(t) < 0$ for all $t > 0$ and $\mu \geq 0$, the theorem can be shown similarly as Theorem 4.1. □

It follows from the inequalities (4.7) and (4.13) that

$$\mathcal{G}_{\ell-1}(\psi_\mu) < \mathcal{G}_\ell(\psi_\mu) < \eta^2(\mu) < \hat{\mathcal{G}}_\ell(\psi_\mu) < \hat{\mathcal{G}}_{\ell-1}(\psi_\mu). \tag{4.15}$$

Thus, Gauss and Gauss-Radau rules provide lower and upper bounds, respectively, of $\eta^2(\mu)$.

We turn to the evaluation of the Gauss-Radau rule $\hat{\mathcal{G}}_\ell(\psi_\mu)$. Introduce the matrix

$$\hat{T}_{\ell+1} = \begin{bmatrix} \alpha_1 & \beta_1 & & & & 0 \\ \beta_1 & \alpha_2 & \beta_2 & & & \\ & \beta_2 & \alpha_3 & & & \\ & & & \ddots & & \\ & & \ddots & \ddots & \beta_{\ell-1} & \\ & & & \beta_{\ell-1} & \alpha_\ell & \beta_\ell \\ 0 & & & & \beta_\ell & \hat{\alpha}_{\ell+1} \end{bmatrix} \in \mathbb{R}^{(\ell+1)\times(\ell+1)}, \tag{4.16}$$

where the entry $\hat{\alpha}_{\ell+1}$ is chosen to make the matrix positive semidefinite with one vanishing eigenvalue. The entry $\hat{\alpha}_{\ell+1}$ can be determined in $\mathcal{O}(\ell)$ arithmetic floating point operations. Analogously to (4.10), we have

$$\hat{\mathcal{G}}_\ell(\psi_\mu) = \|g_c\|^2 e_1^T \psi_\mu(\hat{T}_{\ell+1}) e_1, \tag{4.17}$$

and the right-hand side can be evaluated in $\mathcal{O}(\ell)$ arithmetic floating point operations for each value of $\mu \geq 0$; see, e.g., [1] for details.

Let

$$\eta_\ell^-(\mu) := (\mathcal{G}_\ell(\psi_\mu))^{1/2}, \qquad \eta_\ell^+(\mu) := (\hat{\mathcal{G}}_\ell(\psi_\mu))^{1/2}, \qquad \mu \geq 0, \tag{4.18}$$

and introduce the set

$$\bigcup \{(\mu, \eta) : \eta_\ell^-(\mu) \leq \eta \leq \eta_\ell^+(\mu),\ \mu \geq 0\}, \tag{4.19}$$

which we refer to as the *trust-ribbon*. It follows from the inequalities (4.15), that the trust-ribbon contains the curve (1.7). By Theorems 4.1 and 4.2 the width of the ribbon, $\eta_\ell^+(\mu) - \eta_\ell^-(\mu)$, decreases as μ or ℓ increase.

Let $s_\ell(\mu)$ be the approximate solution of (1.4) defined by (3.7) and (3.8). Then it follows from (4.10) that

$$\|s_\ell(\mu)\|^2 = \|z_\ell(\mu)\|^2 = \|g_c\|^2 e_1^T (T_\ell + \mu I)^{-2} e_1 = \mathcal{G}_\ell(\psi_\mu).$$

Hence, the approximate solution $s_\ell(\mu)$ achieves the lower bound $\mathcal{G}_\ell(\psi_\mu)$ in (4.15).

5. The model-ribbon. We describe how the techniques discussed in Section 4 can be applied to the determination of lower and upper bounds for the function $m_c(s)$ defined by (1.1) with s given by (2.2). These bounds will help us decide how to choose the number of steps ℓ of the Lanczos process when computing the Lanczos decomposition (3.2).

The value of the quadratic model (1.1) at the solution (2.2) of the trust-region subproblem (1.4) is given by

$$m_c(s(\mu)) = f(x_c) - \frac{1}{2}g_c^T(H_c + \mu I)^{-1}g_c - \frac{\mu}{2}g_c^T(H_c + \mu I)^{-2}g_c. \tag{5.1}$$

Using the functions $\phi_\mu(t) := (t+\mu)^{-1}$ and $\psi_\mu(t)$, we can express (5.1) in the form

$$m_c(s(\mu)) = f(x_c) - \frac{1}{2}g_c^T\phi_\mu(H_c)g_c - \frac{\mu}{2}g_c^T\psi_\mu(H_c)g_c. \tag{5.2}$$

The 2ℓth and $(2\ell+1)$st derivatives of $\phi_\mu(t)$ satisfy $\phi_\mu^{(2\ell)}(t) > 0$ and $\phi_\mu^{(2\ell+1)}(t) < 0$ for $t > 0$ and $\mu \geq 0$. Therefore, one can show similarly as (4.15) that

$$\mathcal{G}_{\ell-1}(\phi_\mu) < \mathcal{G}_\ell(\phi_\mu) < g_c^T\phi_\mu(H_c)g_c < \hat{\mathcal{G}}_\ell(\phi_\mu) < \hat{\mathcal{G}}_{\ell-1}(\phi_\mu). \tag{5.3}$$

Let

$$m_{c,\ell}^-(\mu) := f(x_c) - \frac{1}{2}\hat{\mathcal{G}}_\ell(\phi_\mu) - \frac{\mu}{2}\hat{\mathcal{G}}_\ell(\psi_\mu), \tag{5.4}$$

$$m_{c,\ell}^+(\mu) := f(x_c) - \frac{1}{2}\mathcal{G}_\ell(\phi_\mu) - \frac{\mu}{2}\mathcal{G}_\ell(\psi_\mu). \tag{5.5}$$

The inequalities (4.15) and (5.3) then yield

$$m_{c,\ell-1}^-(\mu) < m_{c,\ell}^-(\mu) < m_c(s(\mu)) < m_{c,\ell}^+(\mu) < m_{c,\ell-1}^+(\mu). \tag{5.6}$$

These bounds are valid for all $\mu \geq 0$. We remark that the quadrature rule $\mathcal{G}_\ell(\phi_\mu)$ in (5.5) is straightforward to evaluate when the vector $y_\ell(\mu)$ defined by (4.11) is available, because

$$\mathcal{G}_\ell(\phi_\mu) = \|g_c\|^2 e_1^T y_\ell(\mu). \tag{5.7}$$

The value $\hat{\mathcal{G}}_\ell(\phi_\mu)$ can be computed analogously.

We refer to the set

$$\bigcup\{(\mu, m) : m_{c,\ell}^-(\mu) \leq m \leq m_{c,\ell}^+(\mu),\ \mu \geq 0\} \tag{5.8}$$

as the *model-ribbon*. The inequalities (5.6) show that the model-ribbon contains the curve (1.8), and that the width of the model-ribbon, $m_{c,\ell}^+(\mu) - m_{c,\ell}^-(\mu)$, decreases as ℓ increases. Moreover, one can show that the width decreases as μ increases. The proof is similar to the proof of the analogous result for the trust-ribbon. After having determined a suitable value of μ we can use the bounds (5.6) to determine whether the number of Lanczos steps ℓ should be increased. This is discussed in the following section.

Finally, let the approximate solution $s_\ell(\mu)$ of the trust-region subproblem (1.4) be defined by equations (3.7) and (3.8). Substituting $s_\ell(\mu)$ into (1.1) yields

$$m_c(s_\ell(\mu)) = f(x_c) - \frac{1}{2}\mathcal{G}_\ell(\phi_\mu) - \frac{\mu}{2}\mathcal{G}_\ell(\psi_\mu).$$

Thus, $m_c(s)$ achieves the upper bound $m_{c,\ell}^+(\mu)$ in (5.6) for $s = s_\ell(\mu)$.

6. Application of the model- and trust-ribbons. The following algorithm indicates how the model- and trust-ribbons can be used to help determine how many steps, ℓ, to carry out with the Lanczos process when solving the trust-region subproblem (1.4). We assume that the matrix H_c is positive definite and, thus, address the points (i) and (ii) of Section 1. The parameters Δ_η and Δ_{m_c} are prescribed bounds for the relative differences $(\eta_\ell^+(\mu) - \eta_\ell^-(\mu))/\eta_\ell^+(\mu)$ and $(m_{c,\ell}^+(\mu) - m_{c,\ell}^-(\mu))/m_{c,\ell}^+(\mu)$, respectively.

ALGORITHM 6.1. *Trust-Region Subproblem Algorithm with Bounds*

1. *Input:* $f(x_c)$, $g_c := -\nabla f(x_c) \in \mathbb{R}^n$, $H_c \in \mathbb{R}^{n\times n}$, δ_c, ℓ, Δ_η, Δ_{m_c}.
   Output: Parameter value μ *and associated approximate solution* $s_\ell(\mu)$ *of (1.4), bounds* $\eta_\ell^\pm(\mu)$, $m_{c,\ell}^\pm(\mu)$.
2. *Apply the Lanczos process to the matrix* H_c *with initial vector* g_c *to compute the entries of the tridiagonal matrix* T_ℓ *in the Lanczos decomposition (3.2) and the coefficient* β_ℓ *given by (3.5), which is assumed to be positive.*
3. *Let* $\mu := 0$ *and compute the Cholesky factorization* $T_\ell = L_\ell L_\ell^T$. *If this factorization does not exist, then* `exit`.
4. *Solve equation (3.8) for* $z_\ell(\mu)$ *using the Cholesky factorization of* T_ℓ.
5. *If* $\|z_\ell(\mu)\| > \delta_c$, *then compute the largest zero of the equation* $\|z_\ell(\mu)\| = \delta_c$, *e.g., by using one of the zero-finders discussed in [2, Section 7.3.3-7.3.4], [3, p. 136] or [7, 13]. Denote the computed zero by* μ.
6. *Evaluate* $\mathcal{G}_\ell(\psi_\mu)$, $\hat{\mathcal{G}}_\ell(\psi_\mu)$ *and the bounds* $\eta_\ell^-(\mu)$ *and* $\eta_\ell^+(\mu)$ *of* $\|s(\mu)\|$. *If* $\eta_\ell^+(\mu) - \eta_\ell^-(\mu) > \Delta_\eta \eta_\ell^+(\mu)$, *then increase* ℓ *by* 1 *and goto Step 2.*
7. *Evaluate* $\mathcal{G}_\ell(\phi_\mu)$, $\hat{\mathcal{G}}_\ell(\phi_\mu)$ *and the bounds* $m_{c,\ell}^-(\mu)$ *and* $m_{c,\ell}^+(\mu)$ *of* $m_c(s(\mu))$. *If* $m_{c,\ell}^+(\mu) - m_{c,\ell}^-(\mu) > \Delta_{m_c} m_{c,\ell}^+(\mu)$, *then increase* ℓ *by* 1 *and goto Step 2.*
8. *Compute* $s_\ell(\mu) := V_\ell z_\ell(\mu)$.

Properties of Krylov subspace based trust-region methods are discussed by Conn et al. [2, Section 7.5.4]; in particular, an outline of a complete algorithm for a Lanczos-based trust-region method for function minimization is provided.

We turn to the situation when H_c is indefinite. Then the bounds for $m_c(s(\mu))$ and $\|s(\mu)\|$ derived in Sections 4 and 5 are not valid. First assume that in Step 3 of Algorithm 6.1 we find that T_ℓ is indefinite and let θ_1 denote the smallest eigenvalue of T_ℓ and u_1 an associated eigenvector. Since T_ℓ has nonvanishing subdiagonal entries, the eigenvalue θ_1 is simple. Let $M^\dagger$ denote the Moore-Penrose pseudoinverse of the matrix M. It follows from Proposition 2.1 that when $\|g_c\|\|(T_\ell + \theta_1 I)^\dagger e_1\| \geq \delta_c$, the solution of the reduced trust-region subproblem (1.6) is given by

$$s_\ell(\mu) = \|g_c\| V_\ell (T_\ell + \mu I)^\dagger e_1,$$

where $\mu > -\theta_1$ is chosen so that $\|s_\ell(\mu)\| = \delta_c$. If, instead, $\|g_c\|\|(T_\ell + \theta_1 I)^\dagger e_1\| < \delta_c$, then the solution of (1.6) is of the form

$$s_\ell(\tau) = \|g_c\| V_\ell((T_\ell - \theta_1 I)^\dagger e_1 + \tau u_1),$$

where the parameter τ is chosen so that τu_1 is a descent direction and $\|s_\ell(\tau)\| = \delta_c$.

One of the inherent difficulties with Krylov subspace based trust-region methods is that the matrix T_ℓ may be positive definite, when H_c is not. We are particularly concerned about this situation when we in Step 5 of Algorithm 6.1 find that

$$\|z_\ell(0)\| \ll \delta_c. \tag{6.1}$$

The correction $s_\ell = V_\ell z_\ell(0)$ to x_c then is much smaller than the correction of the trust-region subproblem (1.4), which, since H_c is indefinite, is of norm δ_c, and this may hamper progress towards the minimizer x_* of $f(x)$. We remark that the inequality (6.1) also may hold when H_c is positive definite and the Newton correction of x_c is inside the trust-region.

In order to avoid to frequently take unnecessarily small steps, we proceed as follows. When $\|z_\ell(0)\|/\delta_c$ is smaller than a prescribed tolerance, we periodically apply ℓ steps of the Lanczos process to H_c with a random initial vector, $g_c^\perp$, that is orthogonal to g_c. Let $T_\ell^\perp$ denote the symmetric tridiagonal matrix determined by the Lanczos method. If $T_\ell^\perp$ is indefinite, then we compute an eigenvector $u_1^\perp$ associated with the smallest eigenvalue of $T_\ell^\perp$ and determine an approximate solution of the reduced trust-region subproblem (1.6) of the form

$$s_\ell(\tau) = \|g_c\| V_\ell((T_\ell - \theta_1 I)^\dagger e_1 + \tau u_1^\perp), \tag{6.2}$$

where the parameter τ is chosen so that $\tau u_1^\perp$ is a descent direction and $s_\ell(\tau)$ is of length δ_c. This approach is illustrated in Example 6.3 below.

7. Computed examples. This section compares Algorithm 6.1 to the method of Steihaug outlined in Section 1. We also illustrate the application of the approximate solution (6.2) of the trust-region subproblem (1.4). All computations were carried out in Matlab with about 16 significant decimal digits.

Let $H \in \mathbb{R}^{1024\times 1024}$ be defined by discretizing the negative Laplace operator on the unit square on a uniform mesh with 32×32 grid points using the standard five-point stencil multiplied by the square of the mesh size. Thus, the diagonal entries of the matrix H so obtained are all equal to 4. Let the vector $g \in \mathbb{R}^{1024}$ have uniformly distributed randomly generated entries in the interval $[0, 1]$. The value $f(x_c)$ is not required by Algorithm 6.1 and is assumed to be zero.

Example 6.1. Let $H_c := H$ and $g_c := g$, where H and g are defined above. We display the trust- and model-ribbons that are determined with $\ell = 5$ steps of the Lanczos process applied to the matrix H_c with initial vector g_c. Figure 7.1 shows the upper bound $\eta_5^+(\mu)$ of $\|s(\mu)\|$ as a function of μ (dash-dotted curve), the lower bound $\eta_5^-(\mu)$ of $\|s(\mu)\|$ (dashed curve), and $\|s(\mu)\|$ (solid curve). The trust-ribbon is the area between the dash-dotted and dashed curves. Figure 7.1 also displays $\delta_c = 140$ by a horizontal line. Let μ_5 denote the μ-coordinate of the intersection between this horizontal line and the dashed curve. It is the largest solution of the equation $\|z_5(\mu)\| = \delta_c$, and its value, $5.91 \cdot 10^{-2}$, is computed in Step 5 of Algorithm 6.1. The intersection between the horizontal line and the solid curve is $\mu_+ = 8.020 \cdot 10^{-2}$.

Figure 7.2 shows the upper bound $m_{c,5}^+(\mu)$ of $m_c(s(\mu))$ as a function of μ (dash-dotted curve), the lower bound $m_{c,5}^-(\mu)$ of $m_c(s(\mu))$ (dashed curve), and $m_c(s(\mu))$ (solid curve). The model-ribbon is the area between the dash-dotted and dashed curves. We have $m_{c,5}^+(\mu_5) = -1.69 \cdot 10^3$, $m_{c,5}^-(\mu_5) = -3.15 \cdot 10^3$ and $m_c(s(\mu_5)) = -2.25 \cdot 10^3$. We remark that $m_c(s(\mu_+)) = -1.85 \cdot 10^3$. We can associate the following relative errors with the computed solution $s_5(\mu_5)$ defined by (3.7) and (3.8):

$$\frac{m_{c,5}^+(\mu_5) - m_c(s(\mu_5))}{m_c(s(\mu_5))} = -2.5 \cdot 10^{-1}, \qquad \frac{m_{c,5}^+(\mu_5) - m_c(s(\mu_+))}{m_c(s(\mu_+))} = -8.5 \cdot 10^{-2}$$

and

$$\frac{\|s_5(\mu_5) - s(\mu_+)\|}{\|s(\mu_+)\|} = 2.1 \cdot 10^{-1}.$$

For comparison, we note that the method of Steihaug [17] applied to this problem gave the approximate solution $\hat{s}_3$ of the trust-region subproblem after three iterations

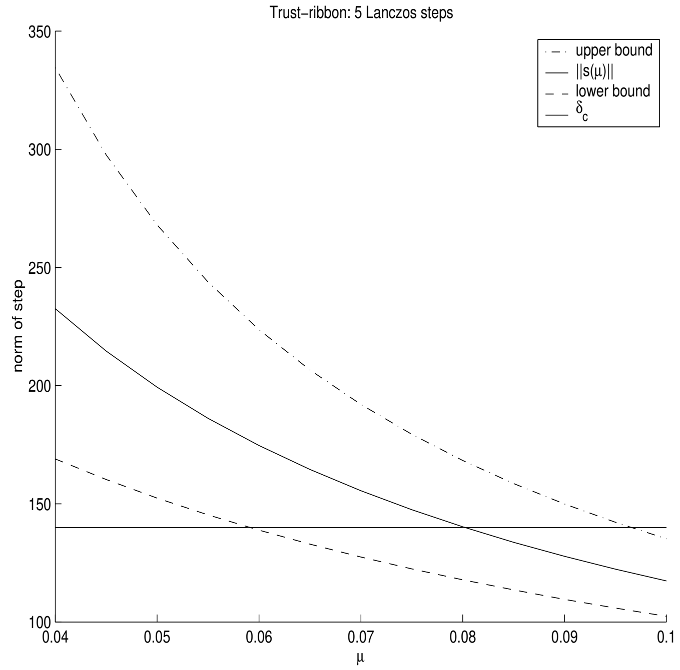

FIG. 7.1. *Example 6.1: Upper bound $\eta_5^+(\mu)$, $\|s(\mu)\|$ and lower bound $\eta_5^+(\mu)$ for $4 \cdot 10^{-2} \leq \mu \leq 1 \cdot 10^{-1}$ and δ_c.*

with the conjugate gradient method, with $m_c(\hat{s}_3) = -1.24 \cdot 10^3$. We obtain the relative errors

$$\frac{m_c(\hat{s}_3) - m_c(s(\mu_5))}{m_c(s(\mu_5))} = -4.5 \cdot 10^{-1}, \qquad \frac{m_c(\hat{s}_3) - m_c(s(\mu_+))}{m_c(s(\mu_+))} = -3.3 \cdot 10^{-1}$$

and

$$\frac{\|\hat{s}_3 - s(\mu_+)\|}{\|s(\mu_+)\|} = 3.2 \cdot 10^{-1}.$$

Thus, the errors achieved by Algorithm 6.1 are smaller than the errors obtained by Steihaug's method. □

Example 6.2. This example differs from Example 6.1 only in that the number of Lanczos steps is increased to $\ell = 10$ in Algorithm 6.1. Figure 7.3 is analogous to Figure 7.1. A comparison between these figures shows that increasing the number of Lanczos steps makes the computed bounds tighter; the trust-ribbon is narrower in Figure 7.3 than in Figure 7.1. In particular, the graph for $\eta_{10}^-(\mu)$ in Figure 7.3 is above the graph for $\eta_5^-(\mu)$ in Figure 7.1 and, therefore, the largest solution of $\|z_{10}(\mu)\| = \delta_c$, denoted by μ_{10} is larger than the zero μ_5 of Example 6.1. We have $\mu_{10} = 7.79 \cdot 10^{-2}$.

Figure 7.4 is analogous to Figure 7.2. The values $m_{c,10}^+(\mu_{10}) = -1.84 \cdot 10^3$ and $m_{c,10}^-(\mu_{10}) = -1.99 \cdot 10^3$ bracket $m_{c,10}(s(\mu_{10})) = -1.89 \cdot 10^3$. The computed solution $s_{10}(\mu_{10})$ corresponds to the value $m_{c,10}^+(\mu_{10})$ of the model. We obtain the relative errors

$$\frac{m_{c,10}^+(\mu_{10}) - m_c(s(\mu_{10}))}{m_c(s(\mu_{10}))} = -2.6 \cdot 10^{-2}, \qquad \frac{m_{c,10}^+(\mu_{10}) - m_c(s(\mu_+))}{m_c(s(\mu_+))} = -7.1 \cdot 10^{-3}$$

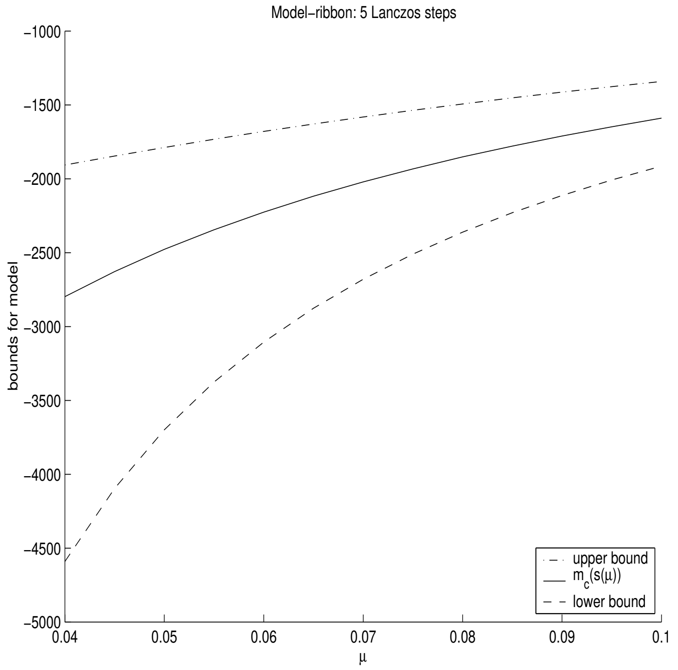

FIG. 7.2. *Example 6.1: Upper bound $m_{c,5}^{+}(\mu)$, $m_c(s(\mu))$ and lower bound $m_{c,5}^{-}(\mu)$ for* $4\cdot 10^{-2} \leq \mu \leq 1\cdot 10^{-1}$.

and

$$\frac{\|s_{10}(\mu_{10}) - s(\mu_+)\|}{\|s(\mu_+)\|} = 7.4\cdot 10^{-2}.$$

A comparison with the corresponding errors reported in Example 6.1 shows that by increasing the number of Lanczos steps, we obtain an improved approximation of the minimum of the trust-region subproblem. Increasing the number of Lanczos steps further reduces the errors even more. For instance, application of Algorithm 6.1 with $\ell = 20$ Lanczos steps yields the approximate solution $s_{20}(\mu_{20})$ of the trust-region subproblem (1.4) with relative error

$$\frac{\|s_{20}(\mu_{20}) - s(\mu_+)\|}{\|s(\mu_+)\|} = 3.5\cdot 10^{-3},$$

where $\mu_{20} = 8.019\cdot 10^{-2}$ is the largest solution of $\|z_{20}(\mu)\| = \delta_c$. $\Box$

Example 6.3. We illustrate the handling of indefinite matrices H_c outlined in Section 6. Let H and g be defined as in the beginning of this section, and let w_1 be an eigenvector of unit length associated with the smallest eigenvalue of H. The spectrum of H lives in the interval $[1.8\cdot 10^{-2}, 8.0]$. Define

$$H_c := H - \frac{1}{10}I, \qquad g_c := \frac{g - w_1^T g\, w_1}{\|g - w_1^T g\, w_1\|}. \tag{7.1}$$

Then H_c is indefinite with 6 negative and 1018 positive eigenvalues, and g_c is orthogonal to w_1. Let $\delta_c := 10$ and apply $\ell = 5$ Lanczos steps to the matrix H_c with initial

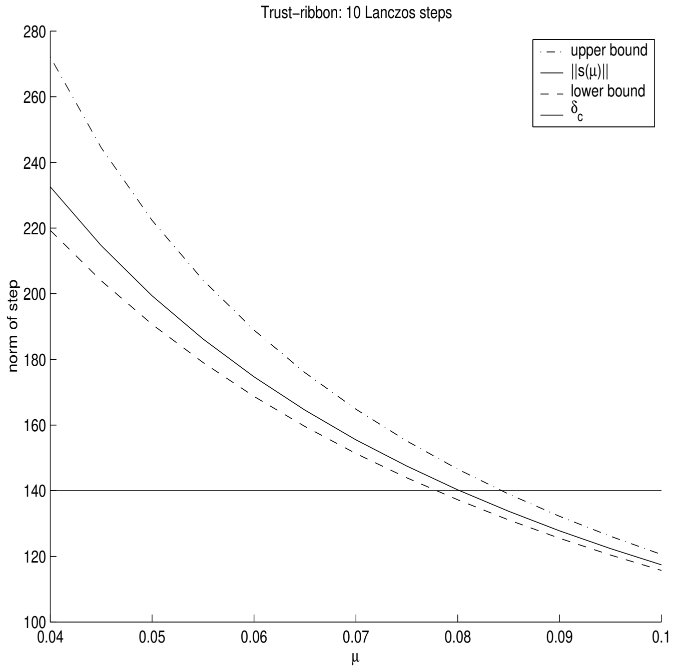

FIG. 7.3. *Example 7.2: Upper bound $\eta_{10}^{+}(\mu)$, $||s(\mu)||$ and lower bound $\eta_{10}^{-}(\mu)$ for $4 \cdot 10^{-2} \leq \mu \leq 1 \cdot 10^{-1}$ and δ_c.*

vector g_c. The tridiagonal matrix T_5 determined by the Lanczos process is positive definite with smallest eigenvalue $7.56 \cdot 10^{-2}$, and the computed Newton step $s_5(0)$ for the reduced problem (1.6) with $\ell = 5$ is of norm 8.73, which is smaller than δ_c. We have $m_c(s_5(0)) = -2.99$.

Instead of accepting the Newton step, we seek to determine whether H_c is indefinite; see the discussion in Section 6. Thus, let $g_c^{\perp}$ be a random vector that is orthogonal to the vector g_c and apply $\ell = 5$ steps of the Lanczos process to the matrix H_c with initial vector $g_c^{\perp}$. This gives the indefinite symmetric tridiagonal matrix $T_5^{\perp}$. We determine the correction $s_5(\tau)$, defined by (6.2), with $\tau = 9.94$ and obtain $m_c(s_5(\tau)) = -4.85$. Hence, the step $s_5(\tau)$ gives a larger reduction in the value of $m_c(s)$ than the Newton step $s_5(0)$.

8. Conclusion. This paper discusses the application of the Lanczos process to reduce large-scale trust-region subproblems to much smaller ones. We show that bounds for the correction $s(\mu)$ and the model value $m_c(s(\mu))$ can be computed inexpensively. These bounds help us determine a suitable number of steps of the Lanczos process. Examples 6.1 and 6.2 show a doubling of the number of Lanczos steps to reduce the length of the intervals in which $s(\mu)$ and $m_c(s(\mu))$ live by about a factor 7. We remark that the size of the reduction is problem dependent. Example 6.3 illustrates an approach to handle indefinite matrices H_c.

Acknowledgement. We would like to thank John Dennis for comments.

REFERENCES

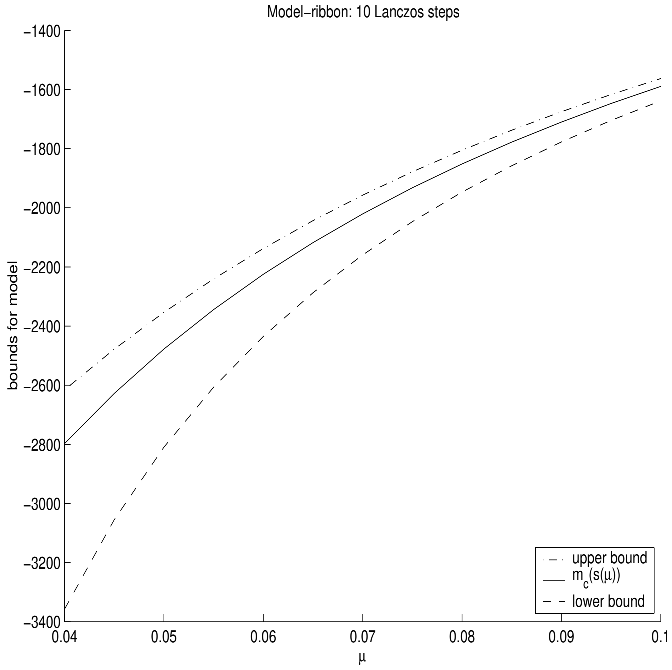

FIG. 7.4. *Example 7.2: Upper bound* $m^+_{c,10}(\mu)$, $m_c(s(\mu))$ *and lower bound* $m^-_{c,10}(\mu)$ *for* $4 \cdot 10^{-2} \leq \mu \leq 1 \cdot 10^{-1}$.

[1] D. Calvetti, G. H. Golub, and L. Reichel, Estimation of the L-curve via Lanczos bidiagonalization, BIT 39 (1999) 603–619.

[2] A. R. Conn, N. I. M. Gould, and Ph. L. Toint, *Trust Region Methods* (SIAM, Philadelphia, 2000).

[3] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstraint Optimization* (SIAM, Philadelphia, 1996).

[4] D. M. Gay, Computing optimal locally constrained steps, SIAM J. Sci. Statist. Comput. 2 (1981) 186–197.

[5] G. H. Golub and G. Meurant, Matrices, moments and quadrature, in: *Numerical Analysis 1993*, eds. D. F. Griffiths and G. A. Watson (Longman, Essex, 1994) pp. 105–156.

[6] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. (John Hopkins University Press, Baltimore, 1996).

[7] G. H. Golub and U. von Matt, Quadratically constrained least squares and quadratic problems, Numer. Math. 59 (1991) 561–580.

[8] 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 (1999) 504–525.

[9] W. W. Hager, Minimizing a quadratic over a sphere, SIAM J. Optim. 12 (2001) 188–208.

[10] M. Hanke, A note on Tikhonov regularization of large linear problems, BIT, to appear.

[11] C. T. Kelley, *Iterative Methods for Optimization* (SIAM, Philadelphia, 1999).

[12] J. J. Moré, Recent developments in algorithms and software for trust region methods: in *Mathematical Programming: The State of the Art*, eds. A. Bachem, M. Grötschel, and B. Korte (Springer, Berlin, 1983) pp. 258–287.

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

[14] J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer, New York, 1999).

[15] M. Rojas, S. A. Santos, and D. C. Sorensen, A new matrix-free algorithm for the large-scale trust-region subproblem, SIAM J. Optim. 11 (2000) 611–646.

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

[17] T. Steihaug, The conjugate gradient method and trust regions in large scale optimization,

SIAM J. Numer. Anal. 20 (1983) 626–637.
[18] Ph. L. Toint, Towards an efficient sparsity exploiting Newton method for minimization, in: *Sparse Matrices and Their Uses*, ed. I. S. Duff (Academic Press, London, 1981) pp. 57–88.
[19] Y. Yuan, On the truncated conjugate gradient method, Math. Program. Ser. A 87 (2000) 561–573.