SIAM J. OPTIMIZATION
Vol. 6, No. 2, pp. 418–445, May 1996

AN INTERIOR TRUST REGION APPROACH FOR NONLINEAR MINIMIZATION SUBJECT TO BOUNDS*

THOMAS F. COLEMAN† AND YUYING LI†

Abstract. We propose a new trust region approach for minimizing a nonlinear function subject to simple bounds. Unlike most existing methods, our proposed method does not require that a quadratic programming subproblem, with inequality constraints, be solved in each iteration. Instead, a solution to a trust region subproblem is defined by minimizing a quadratic function subject only to an ellipsoidal constraint. The iterates generated are strictly feasible. Our proposed method reduces to a standard trust region approach for the unconstrained problem when there are no upper or lower bounds on the variables. Global and local quadratic convergence is established. Preliminary numerical experiments are reported indicating the practical viability of this approach.

Key words. trust region method, interior Newton method, interior-point method

AMS subject classifications. 15A15, 15A09, 15A23

1. Introduction.

We consider the problem of computing a local minimizer of a smooth nonlinear function subject to bounds on the variables:

$$\min_{x\in\Re^n} f(x), \quad l \le x \le u, \tag{1.1}$$

where $l \in \{\Re \cup \{-\infty\}\}^n$, $u \in \{\Re \cup \{\infty\}\}^n$, $l < u$, and $f : \Re^n \to \Re^1$. We denote the feasible set $\mathcal{F} \stackrel{\text{def}}{=} \{x : l \le x \le u\}$ and the strict interior $\text{int}(\mathcal{F}) \stackrel{\text{def}}{=} \{x : l < x < u\}$.

We propose a strictly feasible trust region approach for problem (1.1). Global convergence to a second-order point is established under reasonable assumptions, and a local quadratic convergence rate is also obtained.

Minimization problems with upper and/or lower bounds on some of the variables form an important and common class of problems. There are many algorithms for this type of optimization problem (e.g., [1, 4, 5, 7–9, 11, 13–15, 19–21, 25]), some of which are restricted to quadratic (in some cases convex quadratic) objective functions and some of which are more general. Almost all of the existing methods for problem (1.1) are "active set" methods.

Trust region methods form a respected class of algorithms for solving unconstrained minimization problems. Their high regard is partially due to their strong convergence properties, partially due to their naturalness, and partially due to the recent development of reliable, efficient software. The idea behind a trust region method for $\min_{x\in\Re^n} f(x)$ is very simple. The increment $s_k = x_{k+1} - x_k$ is an approximate solution to a quadratic subproblem with a bound on the step:

$$\min_{s\in\Re^n} \left\{ \psi_k(s) \stackrel{\text{def}}{=} g_k^T s + \frac{1}{2} s^T B_k s \;:\; \|\bar{D}_k s\| \le \Delta_k \right\}. \tag{1.2}$$

* Received by the editors May 12, 1993; accepted for publication (in revised form) October 21, 1994. This research was partially supported by the Applied Mathematical Sciences Research Program (KC-04-02) of the Office of Energy Research of the U. S. Department of Energy under grant DE-FG02-86ER25013.A000, in part by the National Science Foundation, Air Force Office of Scientific Research, and Office of Naval Research through grant DMS-8920550, and by the Advanced Computing Research Institute, a unit of the Cornell Theory Center which receives major funding from the National Science Foundation and IBM Corporation, with additional support from New York State and members of its Corporate Research Institute.

† Computer Science Department and Center for Applied Mathematics, Cornell University, Ithaca, NY 14850 (coleman@cs.cornell.edu and yuying@cs.cornell.edu).

Here $g_k \stackrel{\text{def}}{=} \nabla f(x_k)$, B_k is a symmetric approximation to the Hessian matrix $\nabla^2 f(x_k)$, $\bar{D}_k$ is a scaling matrix, and Δ_k is a positive scalar representing the trust region size. Throughout our presentation, $\|\cdot\|$ denotes the 2-norm.

A general scheme for the unconstrained minimization of $f(x)$, based on subproblem (1.2), is described in Figure 1.1. An iteration with $\rho_k^f > \mu$ is said to be *successful.* Otherwise, the iteration is *unsuccessful.* The aim of trust region size updating is to force $\rho_k^f > \mu$ and hence ensure sufficient reduction of the objective function.

Algorithm 0. Unconstrained trust region method

Let $0 < \mu < \eta < 1$.

For $k = 0, 1, \ldots$

1. Compute $f(x_k)$ and the model ψ_k.
2. Define an approximate solution s_k to subproblem (1.2).
3. Compute $\rho_k^f = (f(x_k + s_k) - f(x_k))/\psi_k(s_k)$.
4. If $\rho_k^f > \mu$ then set $x_{k+1} = x_k + s_k$. Otherwise set $x_{k+1} = x_k$.
5. Update the scaling matrix $\bar{D}_k$ and Δ_k.

Updating trust region size

Let $0 < \gamma_1 < 1 < \gamma_2$.

1. If $\rho_k^f \le \mu$ then set $\Delta_{k+1} \in (0, \gamma_1 \Delta_k]$.
2. If $\rho_k^f \in (\mu, \eta)$ then set $\Delta_{k+1} \in [\gamma_1 \Delta_k, \Delta_k]$.
3. If $\rho_k^f \ge \eta$ then set $\Delta_{k+1} \in [\Delta_k, \gamma_2 \Delta_k]$.

FIG. 1.1. *Trust region method for unconstrained minimization.*

Computing a solution to the trust region problem (1.2) in a reliable and efficient way is a nontrivial task. There are several papers on this topic, e.g., [2, 3, 7, 8, 12, 18, 22–24].

Trust region methods have also been developed for the solution of linearly constrained optimization problems (e.g., [9] and [10]). A quadratic trust region subproblem with linear inequalities is usually approximately solved to obtain an improved point. An iterative procedure must be used to solve the subproblem. For example, Fletcher [10] proposes an algorithm for the linearly constrained optimization problem

$$\min\{f(x) \ : \ E^T x \le d\}$$

in which the subproblem is of the form

$$\min_{s \in \Re^n} \left\{ g_k^T s + \frac{1}{2} s^T B_k s : E^T (x_k + s) \le d, \|\bar{D}_k s\|_\infty \le \Delta_k \right\}. \tag{1.3}$$

As pointed out in [17], convergence theory for trust region methods based on quadratic programming subproblems such as (1.3) usually requires that the computed trial step be a global solution to the subproblem. However, the subproblem is typically solved by methods which guarantee local optimality at best. Therefore, there is a mismatch between theory and practice for trust region methods based on quadratic programming subproblems (with linear inequalities).

In this paper, we propose a trust region approach for (1.1) that does not require the solution of a general quadratic programming subproblem at each iteration. Our proposal is related to the line search based reflection methods proposed by Coleman and Li [6]—scaling strategies and the requirement of strict feasibility are common to both approaches. The primary difference is that in [6] a line search based method is proposed, along with a "reflection" strategy to guarantee sufficient descent, whereas here we propose a pure trust region method.

The proposed approach is developed by forming a quadratic model with an appropriate quadratic function and scaling matrix: there is no need to handle the constraints explicitly. It is then possible to obtain an approximate trust region solution which can guarantee second-order convergence by solving a trust region subproblem with a simple 2-norm constraint and then satisfying the strict feasibility requirement by further restricting the step, if necessary. Our proposed approach reduces to a standard trust region algorithm for unconstrained minimization when $l = -\infty$ and $u = +\infty$. Moreover, all the convergence proofs essentially reduce to established proofs for the unconstrained trust region approach when $l = -\infty$ and $u = +\infty$.

We motivate the method in §2 and establish convergence in §3. In §4, preliminary numerical results for small dense problems are presented. A more comprehensive computational investigation of the method, particularly for large problems, will be presented in a subsequent paper.

As a general rule for notation, unless indicated otherwise, the subscript i or j denotes a component of a vector. The subscript k denotes an index for an (infinite) sequence. For any function F, we use the notation F_k to denote $F(x_k)$ and, if F: $\Re^n \to R^n$, F_{ki} to denote $F_i(x_k)$. When clear from the context, we omit the argument, that is, we use the notation F for $F(x)$ or F_i for $F_i(x)$, for instance.

2. Trust region method for bound-constrained problems. In this section, we propose a trust region method for bound-constrained problems. Our method involves choosing a scaling matrix D_k and a quadratic model $\psi_k(s)$ (we reserve $\bar{D}_k$ to denote the "classical" scaling as in (1.2)). We motivate our choice of scaling matrix by examining the optimality conditions for (1.1).

Let $g(x) \stackrel{\text{def}}{=} \nabla f(x)$. We first define a vector function $v(x) : \Re^n \to \Re^n$ as follows.

DEFINITION 2.1. *The vector $v(x) \in \Re^n$ is defined as follows: for each component $1 \le i \le n$,*

(i) *if $g_i < 0$ and $u_i < \infty$ then $v_i \stackrel{\text{def}}{=} x_i - u_i$;*

(ii) *if $g_i \ge 0$ and $l_i > -\infty$ then $v_i \stackrel{\text{def}}{=} x_i - l_i$;*

(iii) *if $g_i < 0$ and $u_i = \infty$ then $v_i \stackrel{\text{def}}{=} -1$;*

(iv) *if $g_i \ge 0$ and $l_i = -\infty$ then $v_i \stackrel{\text{def}}{=} 1$.*

Following Matlab notation, for any $s \in \Re^n$, diag(s) denotes an n-by-n diagonal matrix with the vector s defining the diagonal entries in their natural order. Moreover, for any nonsingular matrix $A \in \Re^{n\times n}$ and any $l > 0$, A^{-l} denotes the inverse of A^l, where A^l is the lth power of A. Using this notation, we define

$$D(x) \stackrel{\text{def}}{=} \text{diag}(|v(x)|^{-\frac{1}{2}}), \tag{2.1}$$

i.e., D^{-2} is a diagonal matrix with the ith diagonal component equal to $|v_i|$.

Optimality conditions for problem (1.1) are well established. Assuming feasibility and $g_* \stackrel{\text{def}}{=} g(x_*)$, *first-order necessary conditions* for x_* to be a local minimizer are

$$\text{first-order:} \begin{cases} g_{*i} = 0 & \text{if } l_i < x_{*i} < u_i, \\ g_{*i} \le 0 & \text{if } x_{*i} = u_i, \\ g_{*i} \ge 0 & \text{if } x_{*i} = l_i. \end{cases} \tag{2.2}$$

Equivalently, $D(x_*)^{-2}g_* = 0$. Second-order conditions involve the Hessian matrix of f. Let $Free_*$ denote the set of indices corresponding to "free" variables at point x_*:

$$Free_* = \{i : l_i < x_{*i} < u_i\}.$$

Second-order necessary conditions can be written:[1] if a feasible point x_* is a local minimizer of (1.1) then $D(x_*)^{-2}g_* = 0$ and $H_*^{Free_*} \geq 0$, where $H_*^{Free_*}$ is the submatrix of $H_* \stackrel{\text{def}}{=} \nabla^2 f(x_*)$ corresponding to the index set $Free_*$.

These conditions are necessary but not sufficient. Sufficiency conditions that are achievable in practice often require a nondegeneracy assumption. This is the case here.

DEFINITION 2.2. *A point $x \in \mathcal{F}$ is* nondegenerate *if, for each index i,*

$$g_i = 0 \implies l_i < x_i < u_i. \tag{2.3}$$

A problem (1.1) *is nondegenerate if* (2.3) *holds for every $x \in \mathcal{F}$.*

Using this definition we can state *second-order sufficiency conditions*: if a nondegenerate feasible point x_* satisfies (2.2) and $H_*^{Free_*} > 0$, then x_* is a local minimizer of (1.1).

Similar to the view expressed in [7], we consider the following diagonal system of nonlinear equations:

$$D(x)^{-2}g(x) = 0. \tag{2.4}$$

It is easy to see that system (2.4) is an equivalent statement of the first-order necessary conditions. System (2.4) is continuous but not everywhere differentiable. Nondifferentiability occurs when $v_i = 0$; we avoid such points by restricting $x_k \in \text{int}(\mathcal{F})$. Discontinuity of v_i may also occur when $g_i = 0$; however, $D(x)^{-2}g(x)$ is continuous at such points. Moreover, Coleman and Li [7] show that it is possible to generate a second-order Newton process for (2.4). The proof in [7] is based on the following observations.

Assume that $x_k \in \text{int}(\mathcal{F})$. A Newton step for (2.4) satisfies

$$(D_k^{-2}\nabla^2 f(x_k) + \text{diag}(g_k)J_k^v)d_k = -D_k^{-2}g_k, \tag{2.5}$$

where $J^v(x) \in \Re^{n\times n}$ is the Jacobian matrix of $|v(x)|$ whenever $|v(x)|$ is differentiable. Since $|v_k| > 0$ by strict feasibility, following Definition 2.1, the only nondifferentiable points of possible concern occur when $g_{k_i} = 0$ for some index i. If $g_i = 0$, we define the ith row J_i^v of J^v to be zero, i.e., $J_i^v \stackrel{\text{def}}{=} 0$. Nondifferentiability of this type is not a cause for concern because, for such a component, the value of v_i is not significant—local quadratic convergence can be achieved with nonlinear systems of this type [7]. It is clear that $J^v(x)$ is a diagonal matrix. Moreover, if all the components of l and u are finite, $J^v = \text{diag}(\text{sgn}(g))$. If a variable x_i has a finite lower bound and an infinite upper bound (or vice versa) and $g_i = 0$, then $J_{ii}^v = 0$.

Let $B(x)$ be an approximation to $\nabla^2 f(x)$. Based on the Newton step (2.5) for system (2.4), we define our quadratic model in the same way as in [7]:

$$\psi_k(s) \stackrel{\text{def}}{=} g_k^T s + \frac{1}{2}s^T M_k s, \tag{2.6}$$

[1] Notation: if a matrix A is a symmetric matrix then we write $A > 0$ to mean that A is positive definite; $A \geq 0$ means that A is positive semidefinite.

where

$$M_k \stackrel{\text{def}}{=} B_k + C_k, \tag{2.7}$$

$$C_k \stackrel{\text{def}}{=} D_k \text{diag}(g_k) J_k^v D_k. \tag{2.8}$$

It is clear that C_k is a positive semidefinite diagonal matrix.

Define

$$\hat{g}_k \stackrel{\text{def}}{=} D_k^{-1} g_k = \text{diag}(|v_k|^{\frac{1}{2}}) g_k,$$
$$\hat{M}_k \stackrel{\text{def}}{=} D_k^{-1} M_k D_k^{-1} = \text{diag}(|v_k|^{\frac{1}{2}}) B_k \text{diag}(|v_k|^{\frac{1}{2}}) + \text{diag}(g_k) J_k^v,$$
$$\hat{\psi}_k(\hat{s}) \stackrel{\text{def}}{=} \hat{g}_k^T \hat{s} + \frac{1}{2} \hat{s}^T \hat{M}_k \hat{s}.$$

The following lemma can be easily proved.

LEMMA 2.3. *Assume that $x_* \in \mathcal{F}$ and $B(x_*) = H_*$. Then*

(a) *$\hat{g}_* = 0$ if and only if (2.2) is satisfied;*

(b) *$\hat{M}_*$ is positive definite and $\hat{g}_* = 0$ if and only if the second-order sufficiency conditions are satisfied at x_*;*

(c) *$\hat{M}_*$ is positive semidefinite and $\hat{g}_* = 0$ if and only if the second-order necessary conditions are satisfied.*

Lemma 2.3 implies that x_k is a local minimizer of (1.1) if and only if $\hat{s} = 0$ is a solution to

$$\min_{\hat{s} \in \Re^n} \{\hat{\psi}_k(\hat{s}) : \|\hat{s}\| \leq \Delta_k\}. \tag{2.9}$$

Therefore, a solution of the subproblem (2.9) should yield a reasonable (trial) step when x_k is not a local minimizer. Let $s = D_k^{-1}\hat{s}$. Subproblem (2.9) is equivalent to the following problem in the original variable space:

$$\min_{s \in \Re^n} \{\psi_k(s) : \|D_k s\| \leq \Delta_k\}. \tag{2.10}$$

Moreover, in the neighborhood of a local minimizer, the Newton step defined by (2.5) for (2.4) is a solution to the trust region subproblem (2.10) if the trust region size Δ_k is sufficiently large.

The purpose of the scaling matrix D_k in (2.10) is distinctively different from the scaling matrix $\bar{D}_k$ used in unconstrained trust region methods, e.g., (1.2). The scaling matrix D_k is related to the distance to the boundary of the feasible region. Its purpose is to prevent a step directly toward a boundary point. In contrast, the scaling matrix $\bar{D}_k$, used in unconstrained trust region methods, is usually employed for numerical reasons—the scaling matrix $\bar{D}_k$ improves the conditioning of the problem.

If a bound-constrained problem (1.1) is badly scaled, the subproblem (2.10) can be replaced with

$$\min_{s \in \Re^n} \{\psi_k(s) : \|D_k \bar{D}_k s\| \leq \Delta_k\},$$

where $\bar{D}_k$ is chosen to improve the scaling and is a diagonal matrix with the property that $\{\bar{D}_k^{-1}\}$ is bounded and $\{\bar{D}_k\}$ is uniformly bounded. However, to emphasize the role of the scaling matrix D_k, we assume that $\bar{D}_k = I$.

Next we illustrate that it is possible to develop trust region methods for the bound-constrained problem (1.1) based on (2.10). First we introduce some notation and assumptions.

In the subsequent presentation, p_k denotes a global solution to (2.10). Assume that $d_k \in \Re^n$. The scalar α_k denotes the stepsize along d_k to the boundary:

$$\alpha_k \stackrel{\text{def}}{=} \min\left\{\max\left\{\frac{l_i - x_{k_i}}{d_{k_i}}, \frac{u_i - x_{k_i}}{d_{k_i}}\right\} : 1 \le i \le n\right\}, \tag{2.11}$$

and $\frac{l_i - x_{k_i}}{d_{k_i}} \equiv \frac{u_i - x_{k_i}}{d_{k_i}} \stackrel{\text{def}}{=} +\infty$ if $d_{k_i} = 0$. (This is reasonable since $l_i < x_{k_i} < u_i$ always holds.) If problem (1.1) is unconstrained, i.e., $l = -\infty$ and $u = \infty$, we also define $\alpha_k \stackrel{\text{def}}{=} +\infty$. We use $\psi_k^*[d_k]$ to denote the minimum value of $\psi_k(s)$ along the direction d_k within the feasible trust region, i.e.,

$$\psi_k^*[d_k] \stackrel{\text{def}}{=} \psi_k(\tau_k^* d_k) \stackrel{\text{def}}{=} \min\{\psi_k(\tau d_k) : \|\tau D_k d_k\| \le \Delta_k, x_k + \tau d_k \in \mathcal{F}\}. \tag{2.12}$$

Since we always require $x_k \in \text{int}(\mathcal{F})$, a possible step-back may be necessary to stay strictly feasible. Assume $\theta_l \in (0, 1)$. We require that[2]

$$\theta_k \in [\theta_l, 1), \quad \theta_k - 1 = O(\|d_k\|), \quad \text{and} \quad \theta_k = 1 \text{ if } x_k + \tau_k^* d_k \in \text{int}(\mathcal{F}). \tag{2.13}$$

We use $\alpha_k^*[d_k]$ to denote the step obtained from d_k with a possible step-back. The exact definition of $\alpha_k^*[d_k]$ is

$$\alpha_k^*[d_k] \stackrel{\text{def}}{=} \theta_k \tau_k^* d_k. \tag{2.14}$$

We now state a few assumptions.

(AS.1) Given an initial point $x_0 \in \mathcal{F}$, it is assumed that $\mathcal{L}$ is compact, where $\mathcal{L}$ is the level set, i.e., $\mathcal{L} = \{x : x \in \mathcal{F} \text{ and } f(x) \le f(x_0)\}$.

(AS.2) There exists a positive scalar χ_B such that $\|B_k\| \le \chi_B$ for all k.

(AS.3) There exists a positive scalar χ_g such that for $x \in \mathcal{L}$, $\|g(x)\|_\infty < \chi_g$.

Assumption (AS.2) is also required in the convergence analysis of trust region methods for unconstrained problems. Assumption (AS.1) is needed for the boundedness of the scaling matrices $\{D_k^{-1}\}$. Condition (AS.3) is weak. It is satisfied, for example, if the gradient $g(x)$ is continuous on $\mathcal{L}$. Assumptions (AS.1) and (AS.2) imply that there exist positive scalars χ_D, χ_M such that

$$\|D_k^{-1}\| \le \chi_D, \quad \|\hat{M}_k\| \le \chi_M.$$

Note that $\{M_k\}$ is unbounded in general.

Next we present two trust region algorithms for the bound-constrained problem (1.1). The first, called the double-trust region method, is theoretically interesting. It illustrates that bound constraints can be handled by adjusting the trust region size for a 2-norm trust region subproblem with appropriate scaling matrices and the quadratic model. The second method, which we describe in §2.2, represents a more efficient approach.

2.1. The double-trust region method. Our objective is to develop a trust region method for (1.1) based on the trust region subproblem (2.10): a solution p_k to the trust region subproblem (2.10) is obtained and then truncated, i.e., $s_k = \alpha_k^*[p_k]$, to ensure strict feasibility.

The essential idea behind trust region methods is to adjust the trust region size to ensure a sufficient decrease of the objective function. Consider the unconstrained

[2] The notation $\eta_k = O(\rho_k)$ means that there exists a constant $\chi > 0$ such that $|\eta_k| \le \chi|\rho_k|$.

setting: the trust region size is updated to ensure that the reduction of the nonlinear function $f(x)$ is at least a fraction of the reduction of the quadratic model within the trust region. Specifically, the updating of the trust region size forces the condition

$$\rho_k^f = \frac{f(x_k + s_k) - f(x_k)}{\psi_k(s_k)} > \mu$$

for some constant $\mu > 0$. (We use the superscript f to emphasize the dependence on $f(x)$.) Since our quadratic model $\psi_k(s)$ is defined to include the constraint information, a natural extension of the definition of ρ_k^f to a bound-constrained problem (1.1) is given by

$$\rho_k^f \stackrel{\text{def}}{=} \frac{f(x_k + s_k) - f(x_k) + \frac{1}{2}s_k^T C_k s_k}{\psi_k(s_k)}.$$

Similar to unconstrained trust region methods, ρ_k^f measures the agreement between the nonlinear function $f(x)$ and its quadratic approximation.

To obtain first-order convergence of unconstrained trust region methods, a sufficient reduction of the quadratic model $\psi_k(s)$ within the trust region is guaranteed if

$$\text{(2.15)} \quad \psi_k(s_k) < \beta \min\{\psi_k(s) : s = -\tau \bar{D}_k^{-1}\bar{D}_k^{-T} g_k, \|\bar{D}_k s\| \leq \Delta_k\}, \ \|\bar{D}_k s_k\| \leq \beta_0 \Delta_k$$

for some constants $\beta, \beta_0 > 0$. In our notation, (2.15) is equivalent to

$$\text{(2.16)} \qquad \psi_k(s_k) < \beta \psi_k^*[-\bar{D}_k^{-1}\bar{D}_k^{-T} g_k], \quad \|\bar{D}_k s_k\| \leq \beta_0 \Delta_k.$$

For problem (1.1), as we will prove in Lemma 3.1 in §3, if s_k satisfies

$$\text{(2.17)} \qquad \psi_k(s_k) < \beta \psi_k^*[-D_k^{-2} g_k], \ \|D_k s_k\| \leq \beta_0 \Delta_k, \ x_k + s_k \in \text{int}(\mathcal{F}),$$

the trust region model is sufficiently reduced to yield first-order convergence for our approach.

Unfortunately, a truncated step along an exact solution p_k of (2.10) may not sufficiently reduce the quadratic model $\psi_k(s)$ because of the effect of truncation, i.e., (2.17) is not guaranteed when $s_k = \alpha_k^*[p_k]$ and p_k solves (2.10). However, for any trust region subproblem with a nonzero gradient, the trust region solution approaches the gradient direction if the trust region size is reduced to zero. Since a step along the scaled steepest descent direction $-D_k^{-2} g_k$ does produce a sufficient reduction of $\psi_k(s)$, adjustment of the trust region size Δ_k can be used to ultimately guarantee a sufficient decrease. In particular, trust region size reduction can be used to force (2.17), i.e.,

$$\rho_k^c \stackrel{\text{def}}{=} \frac{\psi_k(s_k)}{\psi_k^*[-D_k^{-2} g_k]} > \beta.$$

Hence, we can adjust the trust region size so that both the quadratic model function $\psi_k(s)$ and the nonlinear function $f(x)$ are sufficiently reduced. This gives us the trust region algorithm, Algorithm 1, described in Figure 2.1. We call it the *double-trust region method* because the trust region size is adjusted for both nonlinearity and feasibility. For this method, an iteration is *successful* if both $\rho_k^f > \mu$ and $\rho_k^c > \beta$. Otherwise, an iteration is *unsuccessful*.

Algorithm 1. The double-trust region method
Let $0 < \mu, \beta < \eta < 1$, $x_0 \in \mathrm{int}(\mathcal{F})$.
For $k = 0, 1, \ldots$

1. Compute $f(x_k)$, g_k, B_k and C_k; define the quadratic model $\psi_k(s) = g_k^T s + \frac{1}{2} s^T (B_k + C_k) s$.
2. Compute p_k, a solution to (2.10).
3. Compute

$$s_k = \alpha_k^*[p_k],$$

$$\rho_k^c = \frac{\psi_k(s_k)}{\psi_k^*[-D_k^{-2} g_k]},$$

$$\rho_k^f = \frac{f(x_k + s_k) - f(x_k) + \frac{1}{2} s_k^T C_k s_k}{\psi_k(s_k)}.$$

4. If $\rho_k^f > \mu$ and $\rho_k^c > \beta$ then set $x_{k+1} = x_k + s_k$. Otherwise set $x_{k+1} = x_k$.
5. Update the scaling matrix D_k and Δ_k as specified.

Updating Δ_k for the double-trust region method
Let $0 < \gamma_1 < 1 < \gamma_2$.

1. If $\rho_k^f \le \mu$ or $\rho_k^c \le \beta$ then set $\Delta_{k+1} \in (0, \gamma_1 \Delta_k]$.
2. If $\rho_k^f \in (\mu, \eta)$ then set $\Delta_{k+1} \in [\gamma_1 \Delta_k, \Delta_k]$.
3. If $\rho_k^f \ge \eta$ then
   if $\rho_k^c \ge \eta$ set $\Delta_{k+1} \in [\Delta_k, \gamma_2 \Delta_k]$,
   if $\beta < \rho_k^c < \eta$ set $\Delta_{k+1} \in [\gamma_1 \Delta_k, \Delta_k]$,
   if $\rho_k^c < \beta$ set $\Delta_{k+1} \in (0, \gamma_1 \Delta_k]$.

FIG. 2.1. *Double-trust region method for minimization subject to bounds.*

In §3, we will prove that the double-trust region method has reasonable convergence properties under the nondegeneracy assumption. Although the nondegeneracy assumption is impractical, we believe that the double-trust region method in Figure 2.1 is of theoretical interest. It indicates that if we allow the trust region size to be adjusted according to both accuracy of the quadratic approximation to the nonlinear objective function *and* how well the bound restriction is handled, we can have a trust method for the bound-constrained problem (1.1) based on a 2-norm trust region subproblem. In other words, bound constraints are dealt with *implicitly* using an approach modeled on trust region methods for unconstrained problems. Moreover, the double-trust region method motivates our allowance for possible trust region size reduction in the more practical model algorithm which is presented next, even when $f(x)$ is well represented by its quadratic approximation, i.e., $\rho_k^f > \eta$.

2.2. A practical trust region method. In the last section, we proposed a double-trust region method for bound-constrained problems (1.1) by solving an ellipsoidal trust region subproblem (2.10). In this algorithm, a sufficient decrease of the quadratic model is achieved by monitoring the ratios ρ_k^c and ρ_k^f and adjusting the trust region size accordingly. Since ρ_k^c is determined by $\psi_k(s_k)$, an exact solution to the subproblem (2.10) is assumed in order for ρ_k^c to be reliable. However, for large problems, the assumption that s_k be in the direction of the exact solution of the trust region problem (2.10) is impractical. Moreover, the convergence of the method is established under the assumption that problem (1.1) is nondegenerate. In this section, we suggest a more efficient and practical model algorithm.

As we will see, a sufficient reduction of the quadratic model within the feasible trust region is not difficult to achieve; for example, moving along the scaled gradient $-D_k^{-2}g_k$ guarantees this. If we assume the availability of a step which sufficiently decreases the quadratic function within the feasible trust region, the trust region size is needed only to force the condition $\rho_k^f > \mu$ in the next iteration.

Algorithm 2. A more practical model

Let $0 < \mu < \eta < 1$ and $x_0 \in \mathrm{int}(\mathcal{F})$.

For $k = 0, 1, \ldots$

1. Compute $f(x_k)$, g_k, B_k, and C_k; define the quadratic model $\psi_k(s) = g_k^T s + \frac{1}{2}s^T(B_k + C_k)s$.
2. Compute s_k, based on (2.10), such that $x_k + s_k \in \mathrm{int}(\mathcal{F})$.
3. Compute

$$\rho_k^f = \frac{f(x_k + s_k) - f(x_k) + \frac{1}{2}s_k^T C_k s_k}{\psi_k(s_k)}.$$

4. If $\rho_k^f > \mu$ then set $x_{k+1} = x_k + s_k$. Otherwise set $x_{k+1} = x_k$.
5. Update the scaling matrix D_k and Δ_k as specified.

Updating trust region size Δ_k

Let $0 < \gamma_1 < 1 < \gamma_2$ and $\Lambda_l > 0$ be given.

1. If $\rho_k^f \le \mu$ then set $\Delta_{k+1} \in (0, \gamma_1\Delta_k]$.
2. If $\rho_k^f \in (\mu, \eta)$ then set $\Delta_{k+1} \in [\gamma_1\Delta_k, \Delta_k]$.
3. If $\rho_k^f \ge \eta$ then
   - if $\Delta_k > \Lambda_l$ then
     - set $\Delta_{k+1} \in$ either $[\gamma_1\Delta_k, \Delta_k]$ or $[\Delta_k, \gamma_2\Delta_k]$;
   - otherwise,
     - set $\Delta_{k+1} \in [\Delta_k, \gamma_2\Delta_k]$.

FIG. 2.2. *Trust region method for minimization subject to bounds.*

In Figure 2.2, we describe a trust region method for bound-constrained problems in which the trust region size is primarily updated according to ρ_k^f. However, motivated by the double-trust region method in Figure 2.1, we allow more freedom than usual in the adjustment of Δ_k to permit further reduction in Δ_k even when $\rho_k^f \ge \eta$, thus encouraging the use of the trust region step (2.10).

To satisfy the first-order necessary conditions, given two positive constants β and β_0, it is required that the approximate trust region solution s_k satisfy

$$\text{(AS.4)} \qquad \begin{array}{l} \psi_k(s_k) < \beta\psi_k^*[-D_k^{-2}g_k], \\ \|D_k s_k\| \le \beta_0\Delta_k, \ \ x_k + s_k \in \mathrm{int}(\mathcal{F}). \end{array}$$

In other words, we require that $\psi_k(s_k)$ be less than a fraction of the minimum of $\psi_k(s)$ along the scaled gradient $-D_k^{-2}g_k$ within the feasible trust region. We point out that condition (AS.4) is satisfied for every successful iteration of Algorithm 1. For Algorithm 2, an iteration is *successful* if the condition $\rho_k^f > \mu$ holds. Otherwise, an iteration is *unsuccessful.*

Condition (AS.4) can be easily satisfied for $0 < \beta < 1$. Let d_k be the solution to $\min\{\psi_k(s) : D_k^2 s = -\nu g_k,\ \|D_k s\| \le \Delta_k,\ x_k + s \in \mathcal{F}\}$. Then $s_k = d_k$ satisfies (AS.4), except for the possible violation of $x_k + s_k \in \operatorname{int}(\mathcal{F})$. Assume that $x_k + d_k \notin \operatorname{int}(\mathcal{F})$. Since $\psi_k(s)$ is continuous, a small step-back $s_k = \theta_k d_k$, where $\theta_l < \theta_k < 1$, can ensure both the condition $x_k + s_k \in \operatorname{int}(\mathcal{F})$ and $\psi_k(s_k) < \beta\psi_k^*[-D_k^{-2}g_k]$.

Assumption (AS.4) will not necessarily guarantee a solution at which the second-order necessary conditions are satisfied. To achieve this we make the following stronger assumptions on the quadratic model and the approximate solution:

(AS.5) $$\psi_k(s) = g_k^T s + \frac{1}{2}s^T(\nabla^2 f(x_k) + C_k)s, \text{i.e., } B_k = \nabla^2 f(x_k).$$

(AS.6) Assume that p_k is a solution to $\min_{s\in\Re^n}\{\psi_k(s) : \|D_k s\| \le \Delta_k\}$ and β^q and β_0^q are two positive constants. Then s_k satisfies $\psi_k(s_k) < \beta^q\psi_k^*[p_k], \|D_k s_k\| \le \beta_0^q\Delta_k,\ x_k + s_k \in \operatorname{int}(\mathcal{F})$.

Since both conditions (AS.4) and (AS.6) can be satisfied by simply solving a quadratic trust region subproblem $\min_{s\in\Re^n}\{\psi_k(s) : \|D_k s\| \le \Delta_k\}$, it is *not* necessary to solve a quadratic programming subproblem to achieve convergence. For example, one can first compute a solution p_k to the unconstrained trust region problem

$$\min_{s\in\Re^n}\{\psi_k(s) : \|D_k s\| \le \Delta_k\}$$

and then choose s_k so that $x_k + s_k \in \operatorname{int}(\mathcal{F})$ and $\psi_k(s_k)$ is the minimum of the values $\psi_k^*[p_k]$ and $\psi_k^*[-D_k^{-2}g_k]$. However, requirements (AS.4) and (AS.6) are not very restrictive. There are many ways of computing such approximations. As another example, one can consider the reflection techniques used in [7]. It is also possible to have a subspace adaptation of this trust region approach in which low-dimension subspace trust region problems are solved. We leave the investigation of these computational issues to a subsequent paper.

Before we study the convergence properties of the two trust region methods proposed, we make the following important observation. If we assume that $l = -\infty$ and $u = +\infty$, then $C_k = 0$ and $D_k = I$ and the quadratic model is the same as that for unconstrained problems. Moreover, the conditions (AS.4) and (AS.6) are the same as the conditions required for unconstrained trust region methods (e.g., [17]) since the feasibility constraints are always satisfied.

3. Convergence properties. The convergence proofs for the double-trust region algorithm in Figure 2.1 and the practical algorithm in Figure 2.2 follow the same main steps. Lemmas 3.2 and 3.3 make it possible to present the proofs for both algorithms simultaneously in a clean fashion. The major results are Theorems 3.4, 3.5, 3.10, and 3.11.

The main difference between the practical method and the double-trust region method is that the condition (AS.4) is assumed by the first but satisfied for the latter by adjustment of the trust region size by monitoring the ratio ρ_k^c. However, a common property of the two methods is that, for any successful iteration, (AS.4) is satisfied. Moreover, condition (AS.6) is always satisfied for the double-trust region method.

The convergence results of the two methods are similar. However, the assumptions required by the double-trust region method are stronger—for Algorithm 1 we assume that problem (1.1) is nondegenerate. This nondegeneracy assumption is not needed for Algorithm 2.

The following result is required to express (AS.4) in a manageable form. It is similar to Lemma (4.8) in [17].

LEMMA 3.1. *Assume that* (AS.1)–(AS.3) *are satisfied. If s_k satisfies* (AS.4) *then*

$$-\psi_k(s_k) \geq \frac{1}{2}\beta\|\hat{g}_k\| \min\left\{\Delta_k, \frac{\|\hat{g}_k\|}{\|\hat{M}_k\|}, \frac{\|\hat{g}_k\|}{\|g_k\|_\infty}\right\}.$$

Proof. Define $\phi(\tau) : \Re \to \Re$ by setting $d_k = -D_k^{-1}\frac{\hat{g}_k}{\|\hat{g}_k\|}$ and

$$\phi(\tau) \stackrel{\text{def}}{=} \psi_k(\tau d_k).$$

Let τ_k^* be the minimizer of ϕ on $[0, \min\{\Delta_k, \alpha_k\}]$, where α_k is the stepsize along d_k to the boundary as defined in (2.11):

$$\alpha_k = \min\left\{\max\left\{\frac{l_i - x_{ki}}{d_{ki}}, \frac{u_i - x_{ki}}{d_{ki}}\right\} : 1 \leq i \leq n\right\}.$$

Since $\alpha_k > 0$ (recall that $x_k \in \text{int}(\mathcal{F})$) and the components of d_k have the same sign as that of $-g_k$, it is easy to verify that

$$\alpha_k = \frac{|v_{kj}|}{|d_{kj}|} = \frac{|v_{kj}|\|\hat{g}_k\|}{|v_{kj}||g_{kj}|} \quad \text{for some } j.$$

Hence

$$\alpha_k \geq \frac{\|\hat{g}_k\|}{\|g_k\|_\infty}.$$

By the definition of $\phi(\tau)$,

$$\phi(\tau) = -\tau\|\hat{g}_k\| + \frac{1}{2}\tau^2\mu_k, \quad \mu_k \stackrel{\text{def}}{=} \frac{\hat{g}_k^T\hat{M}_k\hat{g}_k}{\|\hat{g}_k\|^2}.$$

If $\tau_k^* \in [0, \min\{\Delta_k, \alpha_k\})$, then $\mu_k \geq 0$ (thus $\hat{g}_k^T\hat{M}_k\hat{g}_k \geq 0$) and $\tau_k^* = \frac{\|\hat{g}_k\|}{\mu_k}$. Thus

$$\phi(\tau_k^*) = -\frac{1}{2}\frac{\|\hat{g}_k\|^2}{\mu_k} \leq -\frac{1}{2}\frac{\|\hat{g}_k\|^2}{\|\hat{M}_k\|}.$$

Assume $\tau_k^* = \Delta_k$. Since $\mu_k\Delta_k \leq \|\hat{g}_k\|$ when $\mu_k > 0$, and $\phi(\tau_k^*) \leq -\Delta_k\|\hat{g}_k\|$ otherwise, we have

$$\phi(\tau_k^*) = \phi(\Delta_k) \leq -\frac{1}{2}\Delta_k\|\hat{g}_k\|.$$

Assume $\tau_k^* = \alpha_k$. Since $\mu_k\alpha_k \leq \|\hat{g}_k\|$ when $\mu_k > 0$, and $\phi(\tau_k^*) \leq -\alpha_k\|\hat{g}_k\|$ otherwise, we have

$$\phi(\tau_k^*) = \phi(\alpha_k) \leq -\frac{1}{2}\alpha_k\|\hat{g}_k\| \leq -\frac{1}{2}\frac{\|\hat{g}_k\|^2}{\|g_k\|_\infty}.$$

Since $\psi_k(s_k) \leq \beta\phi(\tau_k^*)$ by (AS.4), the result follows from the two previous estimates. □

Assume that the kth iteration is successful from either Algorithm 1 or Algorithm 2. From Lemma 3.1,

$$f(x_k) - f(x_{k+1}) - \frac{1}{2} s_k^T C_k s_k > -\mu \psi_k(s_k)$$
$$\geq \frac{1}{2} \beta \mu \|\hat{g}_k\| \min \left\{ \Delta_k, \frac{\|\hat{g}_k\|}{\|\hat{M}_k\|}, \frac{\|\hat{g}_k\|}{\|g_k\|_\infty} \right\}.$$

Hence, under assumptions (AS.1), (AS.2), (AS.3), and (AS.4),

$$f(x_k) - f(x_{k+1}) > \frac{1}{2} s_k^T C_k s_k + \frac{1}{2} \beta \mu \|\hat{g}_k\| \min \left\{ \Delta_k, \frac{\|\hat{g}_k\|}{\chi_M}, \frac{\|\hat{g}_k\|}{\chi_g} \right\}.$$

Note that $s_k^T C_k s_k \geq 0$. The reduction in f is guaranteed to be better than a multiple of the reduction achieved in the (negative) scaled gradient direction, i.e.,

$$f(x_k) - f(x_{k+1}) > \frac{1}{2} \beta \mu \|\hat{g}_k\| \min \left\{ \Delta_k, \frac{\|\hat{g}_k\|}{\chi_M}, \frac{\|\hat{g}_k\|}{\chi_g} \right\}. \tag{3.1}$$

This inequality is important for the convergence proof.

Next, in Theorems 3.4 and 3.5, we prove that the first-order necessary conditions are satisfied at every limit point of $\{x_k\}$. Several technical results are required first.

Recall that p_k is a global solution to the trust region subproblem (2.10). Using Theorem (3.11) in [17], there exists a parameter λ_k and an upper triangular matrix $R_k \in \Re^{n \times n}$ such that

$$\hat{M}_k + \lambda_k I = R_k^T R_k, \quad (\hat{M}_k + \lambda_k I) D_k p_k = -\hat{g}_k, \quad \lambda_k \geq 0, \tag{3.2}$$

with $\lambda_k(\Delta_k - \|D_k p_k\|) = 0$. Equivalently, p_k is the solution to

$$(\lambda_k I + D_k^{-1} C_k D_k^{-1}) p_k = -D_k^{-2}(g_k + B_k p_k). \tag{3.3}$$

LEMMA 3.2. *Suppose that $\{x_k\}$ is a sequence generated by Algorithm 1. Assume that problem (1.1) is nondegenerate; (AS.1), (AS.2), and (AS.3) hold; and $\{x_k\}$ converges. If $\{\Delta_k\}$ converges to zero and $\liminf_{k\to\infty} \|\hat{g}_k\| > 0$, then $\rho_k^c \geq 1$ for sufficiently large k.*

Proof. Since $\{\Delta_k\}$ converges to zero and $\|D_k p_k\| \leq \Delta_k$, it follows that $\{D_k p_k\}$ converges to zero. But $\|p_k\| \leq \chi_D \|D_k p_k\|$. Hence $\{p_k\}$ converges to zero. Using (AS.2), $\{B_k p_k\}$ converges to zero. From (2.8), $D_k^{-1} C_k D_k^{-1} = \mathrm{diag}(g_k) J_k^v$ is positive semidefinite and bounded. Together with the assumption $\liminf_{k\to\infty} \|\hat{g}_k\| > 0$ and (3.3), it is clear that $\{\lambda_k\}$ converges to $+\infty$.

Assume that α_k is the stepsize to the boundary of the constraints along p_k. From (2.11),

$$\alpha_k = \min \left\{ \max \left\{ \frac{l_i - x_{ki}}{p_{ki}}, \frac{u_i - x_{ki}}{p_{ki}} \right\} : 1 \leq i \leq n \right\}.$$

Consider a limit point of $\{x_k\}$. Since $\{p_k\}$ converges to zero and the problem (1.1) is nondegenerate, it is clear that

$$\lim_{k\to\infty} \max \left\{ \frac{l_j - x_{kj}}{p_{kj}}, \frac{u_j - x_{kj}}{p_{kj}} \right\} = +\infty \quad \forall j \text{ with } g_{*j} = 0.$$

Assume now that $g_{*j} \neq 0$. Since $\{B_k p_k\}$ converges to zero and using (3.3), for k sufficiently large, $-g_{k_j}$ and p_{k_j} have the same sign. Hence, if

$$\alpha_k = \max\left\{\frac{l_j - x_{k_j}}{p_{k_j}}, \frac{u_j - x_{k_j}}{p_{k_j}}\right\},$$

for some j with $g_{*j} \neq 0$, $\alpha_k = \frac{|v_{k_j}|}{|p_{k_j}|}$. Using $\alpha_k = \frac{|v_{k_j}|}{|p_{k_j}|}$, (3.3), the boundedness of $\{(g_k + B_k p_k)\}$, and $\{\lambda_k\}$ converging to $+\infty$, we conclude that $\lim_{k\to\infty} \alpha_k = +\infty$.

Subsequently, $s_k = \alpha_k^*[p_k] = p_k$. Thus, for k sufficiently large,

$$\min\{\psi_k(s) : s = \tau p_k,\ \|D_k s\| \le \Delta_k,\ x_k + s \in \mathcal{F}\} = \min\{\psi_k(s) : s = \tau p_k,\ \|D_k s\| \le \Delta_k\}.$$

Hence

$$\psi_k(p_k) = \min\{\psi_k(s) : s = \tau p_k,\ \|D_k s\| \le \Delta_k,\ x_k + s \in \mathcal{F}\} \le \psi_k^*[-D_k^{-2} g_k],$$

and therefore

$$\rho_k^c = \frac{\psi_k(s_k)}{\psi_k^*[-D_k^{-2} g_k]} = \frac{\psi_k(p_k)}{\psi_k^*[-D_k^{-2} g_k]} \ge 1. \qquad \square$$

LEMMA 3.3. *Assume that* $\{\Delta_k\}$ *is updated by Algorithm* 2. *If* $\rho_k^f \ge \eta$ *for sufficiently large* k, *then* $\{\Delta_k\}$ *is bounded away from zero.*

Proof. By assumption, there exists $\bar{k}$ such that when $k \ge \bar{k}$, $\rho_k^f \ge \eta$. We prove, by induction, that for $k \ge \bar{k}$,

$$\Delta_k \ge \min\{\gamma_1 \Lambda_l, \Delta_{\bar{k}}\}. \tag{3.4}$$

First, it is clear that (3.4) is true when $k = \bar{k}$.

Assume that (3.4) is true for $k \ge \bar{k}$, i.e., $\Delta_k \ge \min\{\gamma_1 \Lambda_l, \Delta_{\bar{k}}\}$. If $\Delta_k \le \Lambda_l$, $\Delta_{k+1} \ge \Delta_k \ge \min\{\gamma_1 \Lambda_l, \Delta_{\bar{k}}\}$. If $\Delta_k > \Lambda_l$, $\Delta_{k+1} \ge \min\{\gamma_1 \Lambda_l, \Delta_{\bar{k}}\}$.

Hence (3.4) is true for all $k \ge \bar{k}$ and $\{\Delta_k\}$ is bounded away from zero. $\square$

The proof of the following theorem is a slight modification of Theorem (4.10) in [17].

THEOREM 3.4. *Assume that* $f : \Re^n \to \Re$ *is continuously differentiable on* $\mathcal{F}$ *and* (AS.1), (AS.2), *and* (AS.3) *hold. For Algorithm* 2, *if* $\{s_k\}$ *satisfies* (AS.4), *then*

$$\liminf_{k\to\infty} \|\hat{g}_k\| = 0. \tag{3.5}$$

For Algorithm 1, (3.5) *is true under the further assumption that problem* (1.1) *is nondegenerate.*

Proof. We must show that $\{\|\hat{g}_k\|\}$ is not bounded away from zero. Assume that there is an $\epsilon > 0$ such that $\|\hat{g}_k\| \ge \epsilon$ for all sufficiently large k. We now show that

$$\sum_{k=1}^{\infty} \Delta_k < +\infty. \tag{3.6}$$

If there are a finite number of successful iterations, then $\Delta_{k+1} \le \gamma_1 \Delta_k$ for all k sufficiently large and then (3.6) clearly holds. Assume now that there is an infinite sequence $\{k_i\}$ of successful iterations. Since $\{f(x_k)\}$ is nonincreasing and bounded below,

$$0 \le \sum_{k=0}^{\infty} (f(x_k) - f(x_{k+1})) < +\infty.$$

This last inequality, (3.1), and $\|\hat{g}_k\| \ge \epsilon$ imply that

$$\sum_{i=1}^{\infty} \Delta_{k_i} < +\infty.$$

Similar to the updating rule in Figure 1.1 for unconstrained trust region methods, the updating rules of Algorithms 1 and 2 specify that $\Delta_{k+1} \le \gamma_1 \Delta_k$ for an unsuccessful iteration and $\Delta_{k+1} \le \gamma_2 \Delta_k$ for a successful iteration. It is easy to verify that the following holds:

$$\sum_{k=1}^{\infty} \Delta_k \le \left(1 + \frac{\gamma_2}{1-\gamma_1}\right) \sum_{i=1}^{\infty} \Delta_{k_i},$$

and thus (3.6) holds in this case as well.

Next we prove that (3.6) implies that $\{|\rho_k^f - 1|\}$ converges to zero. First,

$$\|x_{k+1} - x_k\| \le \|s_k\| \le \chi_D \beta_0 \Delta_k,$$

and hence (3.6) shows that $\{x_k\}$ converges. Now (AS.1) and (AS.2) imply that

$$\begin{aligned} \left|\psi_k(s_k) - g_k^T s_k - \frac{1}{2} s_k^T C_k s_k\right| &= \left|\frac{1}{2} s_k^T B_k s_k\right| \\ &\le \frac{1}{2} \chi_B \chi_D^2 \|D_k s_k\|^2. \end{aligned}$$

But $\|D_k s_k\| \le \beta_0 \Delta_k$. Therefore, using $f(x_k + s_k) - f(x_k) = \bar{g}_k^T s_k$, where $\bar{g}_k \stackrel{\text{def}}{=} \nabla f(x_k + \xi_k s_k)$ with $0 \le \xi_k \le 1$, we have

$$\begin{aligned} &\left|f(x_k + s_k) - f(x_k) + \frac{1}{2} s_k^T C_k s_k - \psi_k(s_k)\right| \\ &\le \left|\psi_k(s_k) - g_k^T s_k - \frac{1}{2} s_k^T C_k s_k\right| + |(\bar{g}_k - g_k)^T s_k| \\ &\le \frac{1}{2} \chi_B \beta_0^2 \chi_D^2 \Delta_k^2 + \chi_D \beta_0 \Delta_k \|g_k - \bar{g}_k\|. \end{aligned}$$

This inequality, the continuity of $g(x)$, and the fact that $\{x_k\}$ converges indicate that there exists a sequence $\{\epsilon_k\}$ converging to zero such that

$$\left|f(x_k + s_k) - f(x_k) + \frac{1}{2} s_k^T C_k s_k - \psi_k(s_k)\right| \le \epsilon_k \Delta_k.$$

Since Lemma 3.1 implies that

$$-\psi_k(s_k) \ge \frac{1}{2} \beta \epsilon \Delta_k,$$

we readily obtain that $\{|\rho_k^f - 1|\}$ converges to zero.

Using Lemma 3.3, $\{\Delta_k\}$ cannot converge to zero in Algorithm 2. This contradicts (3.6) and establishes the result.

For Algorithm 1, using Lemma 3.2, Δ_k is not decreased for sufficiently large k. Thus $\{\Delta_k\}$ cannot converge to zero, contradicting (3.6). The result follows. □

The next theorem establishes that $\{D_k^{-1} g_k\}$ converges to zero. This result is obtained despite the fact that $\{D_k\}$ is not uniformly bounded. This may be somewhat surprising since the analogous convergence result in the unconstrained setting requires the sequence of diagonal scaling matrices to be uniformly bounded. The proof of this theorem is similar to that of Theorem (4.14) in Moré [17].

THEOREM 3.5. *Assume* (AS.1) *and* (AS.2) *hold and* $g(x)$ *is continuous on* $\mathcal{F}$. *If* $\{x_k\}$ *is generated by Algorithm* 2 *and* (AS.4) *holds for* s_k, *then*

$$\lim_{k\to\infty} \|D_k^{-1} g_k\| = 0. \tag{3.7}$$

Result (3.7) *also holds for Algorithm* 1 *when problem* (1.1) *is nondegenerate.*

Proof. The proof is by contradiction and is the same for Algorithms 1 and 2. (The nondegeneracy assumption is needed for Algorithm 1 because the proof uses Theorem 3.4.)

Let ϵ_1 in $(0,1)$ be given and assume that there is a sequence $\{m_i\}$ such that $\|\hat{g}_{m_i}\| \geq \epsilon_1$. Theorem 3.4 guarantees that for any ϵ_2 in $(0,\epsilon_1)$ there is a subsequence of $\{m_i\}$ (without loss of generality we assume that it is the full sequence) and a sequence $\{l_i\}$ such that

$$\|\hat{g}_k\| \geq \epsilon_2, \quad m_i \leq k < l_i, \quad \|\hat{g}_{l_i}\| < \epsilon_2. \tag{3.8}$$

If the kth iteration is successful, then according to (3.1),

$$f(x_k) - f(x_{k+1}) > \frac{1}{2}\beta\mu\epsilon_2 \min\left\{\Delta_k, \frac{\epsilon_2}{\chi_M}, \frac{\epsilon_2}{\chi_g}\right\}, \quad m_i \leq k < l_i.$$

Since $f(x)$ is bounded below on $\mathcal{L}$ and the sequence $\{f(x_k)\}$ is nonincreasing, $\{f(x_k)\}$ converges and $\{f(x_k) - f(x_{k+1})\}$ converges to zero. From $\|x_{k+1} - x_k\| \leq \beta_0 \chi_D \Delta_k$, it follows that, for sufficiently large i,

$$f(x_k) - f(x_{k+1}) \geq \epsilon_3 \|x_{k+1} - x_k\|, \quad m_i \leq k < l_i, \tag{3.9}$$

where $\epsilon_3 = (\frac{1}{2}\beta\mu\epsilon_2)/(\beta_0\chi_D)$. Using (3.9) and the triangle inequality,

$$f(x_{m_i}) - f(x_{k_i}) \geq \epsilon_3 \|x_{k_i} - x_{m_i}\|, \quad m_i \leq k_i \leq l_i.$$

The uniform continuity of $g(x)$ (since $\mathcal{L}$ is compact) and the convergence of $\{f(x_k)\}$ can now be used to deduce that

$$\|g_{m_i} - g_{l_i}\| \leq \epsilon_2 \tag{3.10}$$

for i sufficiently large.

Consider a subsequence of l_i (without loss of generality assume that it is the full sequence) such that $\{x_{l_i}\}$ converges to x_*. Then $\{x_{m_i}\}$ converges to x_*. Based on the definition of $v(x)$, if the jth component of g_* is nonzero, then, for i sufficiently large, the corresponding component of $||v_{m_i}| - |v_{l_i}||$ is no greater than that of $|x_{m_i} - x_{l_i}|$. Thus $\{\mathrm{diag}(|v_{m_i}|^{\frac{1}{2}} - |v_{l_i}|^{\frac{1}{2}}) g_{l_i}\}$ converges to zero. Therefore, for i sufficiently large,

$$\|(D_{m_i}^{-1} - D_{l_i}^{-1}) g_{l_i}\| = \|\mathrm{diag}(|v_{m_i}|^{\frac{1}{2}} - |v_{l_i}|^{\frac{1}{2}}) g_{l_i}\| \leq \epsilon_2. \tag{3.11}$$

Using the triangle inequality for any m and l,

$$\|\hat{g}_m\| \leq \|D_m^{-1}\|\|g_m - g_l\| + \|(D_m^{-1} - D_l^{-1})g_l\| + \|\hat{g}_l\|. \tag{3.12}$$

Combining (3.12) with (3.8), (3.10), and (3.11), we obtain that

$$\epsilon_1 \leq (\chi_D + 2)\epsilon_2.$$

Since ϵ_2 can be any number in $(0, \epsilon_1)$, this is a contradiction. □

Next we consider the second-order necessary conditions. As mentioned in (AS.5) of §2.2, we assume the following quadratic model: $\psi_k(s) = g_k^T s + \frac{1}{2}s^T(\nabla^2 f(x_k) + C_k)s$. Moreover, in addition to (AS.4), the condition (AS.6) holds, i.e., the reduction of the quadratic model satisfies

$$\psi_k(s_k) < \beta^q \min\{\psi_k(s) : s = \tau p_k,\ \|D_k s\| \leq \Delta_k,\ x_k + s \in \mathcal{F}\},$$
$$\|D_k s_k\| \leq \beta_0^q \Delta_k,\ x_k + s_k \in \operatorname{int}(\mathcal{F}),$$

where p_k is a global solution to the unconstrained subproblem

$$\min_{s \in \Re^n} \{\psi_k(s) : \|D_k s\| \leq \Delta_k\}.$$

Before we state the second-order convergence result, several technical lemmas are required. First, we quote Lemma (4.10) in [18].

LEMMA 3.6. *Let x_* be an isolated limit point of a sequence $\{x_k\}$ in $\Re^n$. If $\{x_k\}$ does not converge, then there is a subsequence x_{l_j} which converges to x_* and an $\epsilon > 0$ such that*

$$\|x_{l_j+1} - x_{l_j}\| \geq \epsilon.$$

Now we examine the consequences of (AS.6) in greater detail. Recall (3.2): there exists a parameter λ_k such that

$$(\hat{M}_k + \lambda_k I) = R_k^T R_k, \quad (\hat{M}_k + \lambda_k I)D_k p_k = -\hat{g}_k,\ \lambda_k \geq 0$$

with $\lambda_k(\Delta_k - \|D_k p_k\|) = 0$. Furthermore, as mentioned before, p_k satisfies

$$(\lambda_k I + D_k^{-2} C_k)p_k = -D_k^{-2}(g_k + B_k p_k),$$

which is equation (3.3). These equations will be used repeatedly in the subsequent proofs.

LEMMA 3.7. *Assume that* (AS.6) *is satisfied. Then*

$$-\psi_k(s_k) \geq \beta^q \psi_k^*[p_k] \geq \frac{\beta^q}{2}[\min\{1, \alpha_k^2\}\lambda_k \Delta_k^2 + \min\{1, \alpha_k\}\|R_k D_k p_k\|^2],$$

where α_k is the stepsize along p_k to the boundary and p_k is a global solution to the trust region subproblem (2.10).

Proof. Let $\phi(\tau) \stackrel{\text{def}}{=} \psi_k(\tau p_k)$ and $\tau \in [0, \min\{1, \alpha_k\}]$, where α_k is the stepsize along p_k to the boundary.

It is easy to see that

$$\begin{aligned}
\phi(\tau) &= \tau g_k^T p_k + \frac{1}{2}\tau^2 p_k^T M_k p_k \\
&= \tau \hat{g}_k^T D_k p_k + \frac{1}{2}\tau^2 (D_k p_k)^T \hat{M}_k D_k p_k \\
&= \tau \hat{g}_k^T D_k p_k - \frac{1}{2}\tau^2 \hat{g}_k{}^T D_k p_k - \frac{1}{2}\tau^2 \lambda_k \|D_k p_k\|^2 \quad \text{(from (3.2))} \\
&= -\tau\|R_k D_k p_k\|^2 + \frac{1}{2}\tau^2\|R_k D_k p_k\|^2 - \frac{1}{2}\tau^2 \lambda_k \Delta_k^2 \qquad \text{(from (3.2))}.
\end{aligned}$$

But $\tau^2 \le \tau$ since $\tau \le 1$. Let τ_k^* be the minimizer of $\phi(\tau)$ in $[0, \min\{1, \alpha_k\}]$. Then $\tau_k^* = \min\{1, \alpha_k\}$ since p_k solves the trust region subproblem (2.10). From (AS.6),

$$-\psi_k(s_k) \ge -\beta^q \phi(\tau_k^*) \ge \frac{\beta^q}{2}[\min\{1, \alpha_k^2\}\lambda_k \Delta_k^2 + \min\{1, \alpha_k\}\|R_k D_k p_k\|^2]. \qquad \square$$

The following lemma provides estimates of the reductions in the objective function and the quadratic model. We emphasize that the results hold for *any subsequence* generated by the algorithms. (Consequently, they hold for the entire sequence as well.)

LEMMA 3.8. *Assume that the conditions of Theorem* 3.5 *and* (AS.6) *hold. Furthermore,* $\{x_k\}$ *is any subsequence generated by either Algorithm* 1 *or Algorithm* 2. *If every limit point of* $\{x_k\}$ *is nondegenerate, then there exists* $0 < \epsilon_0 < 1$ *such that, for* k *sufficiently large,*

$$-\psi_k(s_k) \ge \frac{\beta^q}{2} \min\left\{1, \frac{\lambda_k^2 \epsilon_0^2}{[(\chi_g + \Delta_k \chi_B \chi_D)\chi_D^2]^2}, \frac{\lambda_k^2}{(\chi_g + \Delta_k \chi_B \chi_D)^2}\right\} \lambda_k \Delta_k^2,$$

and if the k*th iteration is successful, then*

$$f(x_k) - f(x_{k+1}) > \frac{\beta^q}{2}\mu \min\left\{1, \frac{\lambda_k^2 \epsilon_0^2}{[(\chi_g + \Delta_k \chi_B \chi_D)\chi_D^2]^2}, \frac{\lambda_k^2}{(\chi_g + \Delta_k \chi_B \chi_D)^2}\right\} \lambda_k \Delta_k^2.$$

Proof. Using Lemma 3.7,

$$-\psi_k(s_k) \ge \frac{\beta^q}{2}\min\{1, \alpha_k^2\}\lambda_k \Delta_k^2, \tag{3.13}$$

where α_k is the stepsize along p_k to the boundary as defined in (2.11):

$$\alpha_k = \min\left\{\max\left\{\frac{l_i - x_{k_i}}{p_{k_i}}, \frac{u_i - x_{k_i}}{p_{k_i}}\right\} : 1 \le i \le n\right\}. \tag{3.14}$$

Since the problem is nondegenerate at every limit point and $\{x_k\}$ is bounded, there exists $0 < \epsilon_0 < 1$ and $2\epsilon_0 < \min(u - l)$, such that, for sufficiently large k,

$$\min(x_k - l, u - x_k) + |g_k| > 2\epsilon_0 e, \quad e = [1, \ldots, 1]^T \in \Re^n. \tag{3.15}$$

(Otherwise, there would be a degenerate limit point of $\{x_k\}$.)

Following Theorem 3.5, $\{D_k^{-1} g_k\}$ converges to zero. Hence, for sufficiently large k,

$$\|D_k^{-2} g_k\|_\infty < \epsilon_0^2. \tag{3.16}$$

Assume that k is sufficiently large and

$$\alpha_k = \max\left\{\frac{l_j - x_{k_j}}{p_{k_j}}, \frac{u_j - x_{k_j}}{p_{k_j}}\right\} \quad \text{for some } j. \tag{3.17}$$

Recall that

$$\operatorname{diag}(g_k) J_k^v \ge 0 \tag{3.18}$$

and J_k^v is a diagonal matrix.

If $|g_{kj}| \le \epsilon_0$, using (3.15), we have

$$\min\{x_{kj} - l_j, u_j - x_{kj}\} > \epsilon_0.$$

Hence, from (3.3), (3.17), (3.18), and

$$\|g_k + B_k p_k\|_\infty \le \chi_g + \chi_B \|D_k^{-1}\| \|D_k p_k\| \le \chi_g + \chi_B \chi_D \Delta_k,$$

we obtain

$$\alpha_k \ge \frac{\lambda_k \epsilon_0}{(\chi_g + \Delta_k \chi_B \chi_D)\chi_D^2}.$$

If $|g_{kj}| > \epsilon_0$, then $|v_{kj}| < \epsilon_0$ because of (3.16). If $\alpha_k = \frac{|v_{kj}|}{|p_{kj}|}$, then from (3.17), (3.3), and (3.18), and noting that $|v_{kj}|$ in the numerator is cancelled with D_{kjj}^{-2} in the denominator,

$$\alpha_k \ge \frac{\lambda_k}{\chi_g + \Delta_k \chi_B \chi_D}.$$

Assume that $\alpha_k \ne \frac{|v_{kj}|}{|p_{kj}|}$. Since $|v_{kj}| < \epsilon_0$, $u_j - l_j > 2\epsilon_0$, and $\alpha_k \ne \frac{|v_{kj}|}{|p_{kj}|}$, the magnitude of the numerator determining α_k is greater than ϵ_0. Hence, using (3.3), (3.17), and (3.18),

$$\alpha_k \ge \frac{\lambda_k \epsilon_0}{(\chi_g + \Delta_k \chi_B \chi_D)\chi_D^2}.$$

Using (3.13), we have

$$-\psi_k(s_k) \ge \frac{\beta^q}{2} \min\left\{1, \frac{\lambda_k^2 \epsilon_0^2}{[(\chi_g + \Delta_k \chi_B \chi_D)\chi_D^2]^2}, \frac{\lambda_k^2}{(\chi_g + \Delta_k \chi_B \chi_D)^2}\right\} \lambda_k \Delta_k^2.$$

If the kth iteration is successful, $\rho_k^f > \mu$. Hence

$$\begin{aligned} f(x_k) - f(x_{k+1}) &\ge -\mu \psi_k(s_k) + \frac{1}{2} s_k^T C_k s_k \\ &\ge -\mu \psi_k(s_k) \qquad (\text{since } s_k^T C_k s_k \ge 0) \\ &> \frac{\beta^q}{2} \mu \min\left\{1, \frac{\lambda_k^2 \epsilon_0^2}{[(\chi_g + \Delta_k \chi_B \chi_D)\chi_D^2]^2}, \frac{\lambda_k^2}{(\chi_g + \Delta_k \chi_B \chi_D)^2}\right\} \lambda_k \Delta_k^2. \end{aligned}$$

The proof is completed. □

When $\hat{M}_k$ is positive definite, we denote the Newton step for (2.10) by

$$s_k^N \stackrel{\text{def}}{=} -D_k^{-1} \hat{M}_k^{-1} \hat{g}_k, \quad \text{i.e.,} \quad \hat{M}_k D_k s_k^N = -\hat{g}_k. \tag{3.19}$$

LEMMA 3.9. *Assume* (AS.1), (AS.4), (AS.5), *and* (AS.6) *hold and $f(x)$ is twice continuously differentiable on $\mathcal{L}$. If the sequence of trust region subproblem* (2.10) *solutions $\{p_k\}$ converges to zero, $\{x_k\}$ converges to x_*, and $\hat{M}_*$ is positive definite, then*

$$\liminf_{k \to \infty} \frac{\psi_k(\alpha_k^*[p_k])}{\psi_k^*[p_k]} \ge 1, \quad \liminf_{k \to \infty} \frac{\psi_k^*[p_k]}{\psi_k(p_k)} \ge 1.$$

Moreover, for sufficiently large k,

$$|\psi_k^*[p_k]| \geq \epsilon \min\{\Delta_k^2, \|D_k s_k^N\|^2\}$$

for some constant $\epsilon > 0$.

Proof. Let α_k be the stepsize along p_k to the boundary as defined in (2.11):

$$\alpha_k = \min\left\{\max\left\{\frac{l_i - x_{k_i}}{p_{k_i}}, \frac{u_i - x_{k_i}}{p_{k_i}}\right\} : \ 1 \leq i \leq n\right\}.$$

Since p_k is a solution to (2.10), $\tau_k^* = \min\{1, \alpha_k\}$ in (2.12). By condition (2.13) on θ_k, $\theta_l \leq \theta_k < 1, \theta_k - 1 = O(\|p_k\|)$. Since $\hat{M}_k$ is positive definite for sufficiently large k, we have that $p_k^T M_k p_k > 0$. In addition, $\psi_k^*[p_k] < 0$. Therefore

$$\begin{aligned}
\liminf_{k\to\infty} \frac{\psi_k(\alpha_k^*[p_k])}{\psi_k^*[p_k]} &= \liminf_{k\to\infty} \frac{\tau_k^*\theta_k g_k^T p_k + \frac{1}{2}\tau_k^{*2}\theta_k^2 p_k^T M_k p_k}{\tau_k^* g_k^T p_k + \frac{1}{2}\tau_k^{*2} p_k^T M_k p_k} \\
&\geq \liminf_{k\to\infty} \frac{\tau_k^*\theta_k g_k^T p_k + \frac{1}{2}\tau_k^{*2}\theta_k p_k^T M_k p_k}{\tau_k^* g_k^T p_k + \frac{1}{2}\tau_k^{*2} p_k^T M_k p_k} \quad (\text{using } \theta_k^2 \leq \theta_k \text{ and } \psi_k^*[p_k] < 0) \\
&\geq \lim_{k\to\infty} \theta_k \\
&= 1.
\end{aligned}$$

The last equality $\lim_{k\to\infty} \theta_k = 1$ comes from $\lim_{k\to\infty} p_k = 0$ (by assumption).

Since $\hat{M}_*$ is positive definite, x_* is nondegenerate. If all variables are free at the limit point x_*, then from the assumption that $\{p_k\}$ converges to zero it is clear that

$$\liminf_{k\to\infty} \alpha_k = +\infty.$$

Assume now that there exist variables on the boundary at x_*. Recall (3.3):

$$(\lambda_k I + D_k^{-2} C_k) p_k = -D_k^{-2}(g_k + B_k p_k).$$

Since $\{B_k p_k\}$ converges to zero (note that $B_k = \nabla^2 f_k$ under the assumption (AS.5)), $\lambda_k I + D_k^{-2} C_k$ is a positive semidefinite diagonal matrix, and x_* is nondegenerate with $D_*^{-1} g_* = 0$ (because of Theorem 3.5), for any i with $v_{*i} = 0$, p_{k_i} and g_{k_i} have the same sign for k sufficiently large. Hence, if α_k is defined by some $v_{*j} = 0$ and $g_{*j} \neq 0$, then $\alpha_k = \frac{|v_{k_j}|}{|p_{k_j}|}$ for sufficiently large k. Using (3.3) again,

$$\alpha_k = \frac{|g_{k_j}| + \lambda_k}{|g_{k_j} + (B_k p_k)_j|}.$$

This means that

$$\liminf_{k\to\infty} \alpha_k \geq 1. \tag{3.20}$$

Using Lemma 3.7, we have

$$-\psi_k^*[p_k] \geq \frac{1}{2}\min\{1, \alpha_k\}\|R_k D_k p_k\|^2.$$

Assume that $\epsilon_0 > 0$ is a lower bound on the eigenvalues of $\hat{M}_*$. From (3.2), we have

$$\|R_k D_k p_k\|^2 \geq \epsilon_0 \|D_k p_k\|^2 + \lambda_k \|D_k p_k\|^2.$$

But $\|D_k p_k\| \le \Delta_k$ and, for sufficiently large k, $p_k = s_k^N$ if $\|D_k p_k\| < \Delta_k$. Hence

$$|\psi_k^*[p_k]| \ge \frac{1}{2}\epsilon_0 \min\{1, \alpha_k\} \min\{\Delta_k^2, \|D_k s_k^N\|^2\},$$

where $D_k s_k^N = -\hat{M}_k^{-1}\hat{g}_k$. Let $0 < \epsilon < \frac{1}{2}\epsilon_0$. Then, using (3.20), for k sufficiently large,

$$|\psi_k^*[p_k]| \ge \epsilon \min\{\Delta_k^2, \|D_k s_k^N\|^2\}.$$

In addition, from $\tau_k^{*2} \le \tau_k^*$ (since $0 \le \tau_k^* = \min\{\alpha_k, 1\} \le 1$), $p_k^T M_k p_k > 0$ for sufficiently large k, and $\psi_k^*(p_k) < 0$, we have

$$\begin{aligned}
\liminf_{k\to\infty} \frac{\psi_k^*[p_k]}{\psi_k(p_k)} &= \liminf_{k\to\infty} \frac{\tau_k^* g_k^T p_k + \frac{1}{2}\tau_k^{*2} p_k^T M_k p_k}{g_k^T p_k + \frac{1}{2} p_k^T M_k p_k} \\
&\ge \liminf_{k\to\infty} \min\{\alpha_k, 1\} \\
&= 1.
\end{aligned}$$

Hence

$$\liminf_{k\to\infty} \frac{\psi_k^*[p_k]}{\psi_k(p_k)} \ge 1. \qquad \square$$

The next theorem indicates that the first-order and second-order necessary conditions can be satisfied.

THEOREM 3.10. *Assume* (AS.1) *holds and* $f : \mathcal{F} \to \Re$ *is twice continuously differentiable on* $\mathcal{F}$. *Let* $\{x_k\}$ *be the sequence generated by Algorithm* 2 *under assumption* (AS.5) *on the model* ψ_k *and under assumptions* (AS.4) *and* (AS.6) *on the step* s_k. *Then*

(i) *the sequence* $\{\hat{g}_k\}$ *converges to zero*;

(ii) *if every limit point is nondegenerate, then there is a limit point* x_* *with* $\hat{M}_*$ *positive semidefinite*;

(iii) *if* x_* *is an isolated nondegenerate limit point, then* $\hat{M}_*$ *is positive semidefinite*;

(iv) *if* $\hat{M}_*$ *is nonsingular for some limit point* x_* *of* $\{x_k\}$, *then* $\hat{M}_*$ *is positive definite*, $\{x_k\}$ *converges to* x_*, *all iterations are eventually successful, and* $\{\Delta_k\}$ *is bounded away from zero.*

Under the additional assumption that problem (1.1) *is nondegenerate, equivalent results hold for the sequence generated by Algorithm* 1.

Proof. We prove each result in order.

(i) The sequence $\{\hat{g}_k = D_k^{-1} g_k\}$ converges to zero—this was proved in Theorem 3.5.

(ii) First we consider the case when $\liminf_{k\to\infty} \lambda_k = 0$. Let $\lambda_{\min}(M_k)$ denote the minimum eigenvalue of M_k. Since $\lambda_k \ge \max(-\lambda_{\min}(M_k), 0)$, it is clear that, when $\liminf_{k\to\infty} \lambda_k = 0$, there must exist a limit point x_* at which M_* is positive semidefinite.

Next we prove by contradiction that $\liminf_{k\to\infty} \lambda_k = 0$. Assume that $\lambda_k \ge \epsilon > 0$ for all k sufficiently large. First we show that $\{\Delta_k\}$ converges to zero.

From Lemma 3.8, we have that, for sufficiently large k,

$$-\psi_k(s_k) \ge \frac{\beta^q}{2}\hat{\epsilon}_k \epsilon \Delta_k^2,$$

where

$$\hat{\epsilon}_k \stackrel{\text{def}}{=} \min\left\{1, \frac{\epsilon^2\epsilon_0^2}{[(\chi_g + \Delta_k\chi_D\chi_B)\chi_D^2]^2}, \frac{\epsilon^2}{(\chi_g + \Delta_k\chi_B\chi_D)^2}\right\}.$$

Moreover, for sufficiently large k and successful iterations,

$$f(x_k) - f(x_{k+1}) \geq \frac{1}{2}\beta^q\mu\hat{\epsilon}_k\epsilon\Delta_k^2. \tag{3.21}$$

The remaining arguments are similar to the proof of Theorem 3.4. If there is a finite number of successful iterations, $\{\Delta_k\}$ converges to zero. Otherwise, let $\{k_i\}$ be the infinite sequence of successful iterations. The definition of $\hat{\epsilon}_k$, inequality (3.21), and the convergence of $\{f(x_k)-f(x_{k+1})\}$ to zero imply that there exists a constant $\epsilon_1 > 0$ such that $\hat{\epsilon}_{k_i} > \epsilon_1$. This fact and inequality (3.21) imply that

$$\sum_{i=1}^{\infty}\Delta_{k_i}^2 < \infty.$$

Similar to the updating rule in Figure 1.1 for unconstrained trust region methods, the updating rules of Algorithms 1 and 2 specify that $\Delta_{k+1} \leq \gamma_1\Delta_k$ for an unsuccessful iteration and $\Delta_{k+1} \leq \gamma_2\Delta_k$ for a successful iteration. It is easy to verify that the following holds:

$$\sum_{k=1}^{\infty}\Delta_k^2 \leq \left(1 + \frac{\gamma_2^2}{1-\gamma_1}\right)\sum_{i=1}^{\infty}\Delta_{k_i}^2.$$

Hence $\{\Delta_k\}$ converges to zero. Since $\|s_k\| = \|D_k^{-1}D_ks_k\| \leq \chi_D\beta_0^q\Delta_k$ and $\|p_k\| \leq \chi_D\Delta_k$, we conclude that both $\{s_k\}$ and $\{p_k\}$ converge to zero.

From the fact that $\{\Delta_k\}$ converges to zero,

$$\hat{\epsilon}_k \geq \hat{\epsilon} \qquad \text{for some } \hat{\epsilon} > 0.$$

Hence,

$$-\psi_k(s_k) \geq \frac{\beta^q}{2}\hat{\epsilon}\epsilon\Delta_k^2.$$

Now a standard estimate is that

$$\begin{aligned}&\left|f(x_k + s_k) - f(x_k) + \frac{1}{2}s_k^TC_ks_k - \psi_k(s_k)\right| \\ \leq &\|s_k\|^2 \max_{0\leq\xi\leq1}\|\nabla^2 f(x_k + \xi s_k) - \nabla^2 f(x_k)\|,\end{aligned}$$

and thus the last two inequalities and the fact that $\{s_k\}$ converges to zero (thus $\{\|\nabla^2 f(x_k + \xi s_k) - \nabla^2 f(x_k)\|\}$ converges to zero) show that $\{|\rho_k^f - 1|\}$ converges to zero.

We conclude that the entire sequence $\{\rho_k^f\}$ converges to unity.

For Algorithm 2, using Lemma 3.3, $\{\Delta_k\}$ cannot converge to zero, which is a contradiction.

Now we consider Algorithm 1. Let α_k be the stepsize to the boundary along p_k. According to the definition (2.11),

$$\alpha_k = \min\left\{\max\left\{\frac{l_i - x_{k_i}}{p_{k_i}}, \frac{u_i - x_{k_i}}{p_{k_i}}\right\} : 1 \leq i \leq n\right\}.$$

If all variables are free at a limit point x_*, then it is clear that $\liminf_{k\to\infty} \alpha_k = +\infty$. Otherwise, consider a limit point x_* with some variables at their bounds. Since this limit point is nondegenerate, $\lim_{k\to\infty} D_k^{-2} g_k = 0$, and p_k converges to zero, $-g_{k_j}$ and p_{k_j} have the same sign for any $g_{*j} \neq 0$. Using (3.3), we have

$$\alpha_k = \frac{|v_{k_j}|}{|p_{k_j}|} = \frac{|g_{k_j}| + \lambda_k}{|g_{k_j} + (B_k p_k)_j|}$$

for sufficiently large k with $g_{*j} \neq 0$. But $\lambda_k > \epsilon$. Thus the corresponding limit of α_k is greater than 1. Hence

$$\liminf_{k\to\infty} \alpha_k > 1.$$

In other words, $s_k = \alpha_k^*[p_k] = p_k$ for sufficiently large k. Therefore $\rho_k^c \geq 1$ for sufficiently large k. Hence, $\{\Delta_k\}$ cannot converge to zero, which is a contradiction.

In conclusion, there is a limit point with $\hat{M}_*$ positive semidefinite.

(iii) If $\{x_k\}$ converges to x_*, the result follows from (ii). If $\{x_k\}$ does not converge then Lemma 3.6 applies. Thus, if $\{x_{l_j}\}$ is the subsequence guaranteed by Lemma 3.6, then $\Delta_{l_j} \geq \frac{1}{\chi_D \beta_0^q}\epsilon$ and the l_jth iteration is successful. From Lemma 3.8 (note that Lemma 3.8 holds for *any subsequence*),

$$f(x_{l_j}) - f(x_{l_j+1}) > \frac{\beta^q}{2}\mu \min\left\{1, \frac{\lambda_{l_j}^2 \epsilon_0^2}{[(\chi_g + \Delta_{l_j}\chi_B\chi_D)\chi_D^2]^2}, \frac{\lambda_{l_j}^2}{(\chi_g + \Delta_{l_j}\chi_B\chi_D)^2}\right\} \lambda_{l_j}\Delta_{l_j}^2.$$

From $\Delta_{l_j} \geq \frac{1}{\chi_D\beta_0^q}\epsilon$, it is straightforward to verify that there exist positive constants $\epsilon_1, \epsilon_2, \epsilon_3$ such that

$$f(x_{l_j}) - f(x_{l_j+1}) > \frac{\beta^q}{2}\mu \min\{\epsilon_1\lambda_{l_j}, \epsilon_2\lambda_{l_j}^3, \epsilon_3\lambda_{l_j}^3\}.$$

Since $\{f(x_k)\}$ is monotonically nonincreasing and bounded below, $\{f(x_k) - f(x_{k+1})\}$ converges to zero. Hence $\{\lambda_{l_j}\}$ converges to zero. Thus $\hat{M}_*$ is positive semidefinite.

(iv) If $\hat{M}_*$ is nonsingular at a limit point x_* of $\{x_k\}$, then x_* is an isolated limit point and hence $\hat{M}_*$ is positive definite following (iii). Since $\psi_k(s_k) = g_k^T s_k + \frac{1}{2}s_k^T M_k s_k < 0$, we have that

$$\hat{g}_k^T D_k s_k < -\frac{1}{2}(D_k s_k)^T \hat{M}_k D_k s_k < 0$$

whenever $\hat{M}_k$ is positive definite. Hence

$$\frac{1}{2}(D_k s_k)^T \hat{M}_k D_k s_k < \|D_k s_k\| \|\hat{g}_k\|.$$

But $(D_k s_k)^T \hat{M}_k D_k s_k \geq \frac{1}{\|\hat{M}_k^{-1}\|}\|D_k s_k\|^2$. Therefore

$$\frac{1}{2}\|D_k s_k\| \leq \|\hat{M}_k^{-1}\| \|\hat{g}_k\|$$

whenever $\hat{M}_k$ is positive definite. This means that

$$\frac{1}{2}\|s_k\| \leq \frac{1}{2}\chi_D\|D_k s_k\| \leq \chi_D\|\hat{M}_k^{-1}\|\|\hat{g}_k\|$$

whenever $\hat{M}_k$ is positive definite. But $\{D_k^{-1}g_k\}$ converges to zero. Following Lemma 3.6, $\{x_k\}$ converges. Since $\lim_{k\to\infty}\hat{g}_k = 0$, $\{x_k\}$ converges to x_*, and $\hat{M}_*$ is positive definite, $\{p_k\}$ and $\{s_k\}$ converge to zero.

Next we prove that the iterations are eventually successful and $\{\Delta_k\}$ is bounded away from zero. Using Lemma 3.9, there exists $\epsilon > 0$ such that, for sufficiently large k,

$$|\psi_k^*[p_k]| \geq \epsilon \min\{\Delta_k^2, \|D_k s_k^N\|^2\}. \tag{3.22}$$

But recall that whenever $\hat{M}_k$ is positive definite

$$\frac{1}{2}\|D_k s_k\| \leq \|\hat{M}_k^{-1}\|\|\hat{g}_k\|.$$

Let κ be an upper bound on the condition number of $\hat{M}_k$. From $\hat{g}_k = -\hat{M}_k D_k s_k^N$ based on (3.19),

$$\frac{1}{2}\|D_k s_k\| \leq \kappa\|D_k s_k^N\|.$$

Hence, using (3.22) and (AS.6), there exists $\bar{\epsilon} > 0$ such that

$$-\psi_k(s_k) \geq \bar{\epsilon}\|D_k s_k\|^2 \geq \frac{\bar{\epsilon}}{\chi_D^2}\|s_k\|^2.$$

This estimate and

$$\begin{aligned}&\left|f(x_k+s_k) - f(x_k) + \frac{1}{2}s_k^T C_k s_k - \psi_k(s_k)\right| \\ \leq &\|s_k\|^2 \max_{0\leq\xi\leq1}\|\nabla^2 f(x_k+\xi s_k) - \nabla^2 f(x_k)\|\end{aligned}$$

yield that $\rho_k^f > \eta$ for k sufficiently large since $\{s_k\}$ converges to zero. (Thus $\{\|\nabla^2 f(x_k+\xi s_k) - \nabla^2 f(x_k)\|\}$ converges to zero.)

For Algorithm 2, using Lemma 3.3, we immediately conclude that $\{\Delta_k\}$ is bounded away from zero.

Now we consider Algorithm 1. Since p_k is a solution of the trust region subproblem, it is clear that

$$\psi_k(p_k) \leq \min\{\psi_k(\tau D_k^{-2}g_k) : \|\tau D_k^{-1}g_k\| \leq \Delta_k\} \leq \psi_k^*[-D_k^{-2}g_k] \leq 0.$$

Using Lemma 3.9, for k sufficiently large,

$$\rho_k^c = \frac{\psi_k(p_k)}{\psi_k^*[-D_k^{-2}g_k]} \times \frac{\psi_k^*[p_k]}{\psi_k(p_k)} \times \frac{\psi_k(\alpha_k^*[p_k])}{\psi_k^*[p_k]} > \eta.$$

Therefore all the iterations are eventually successful. According to the trust region size updating rules, $\{\Delta_k\}$ is bounded away from zero. The results are established. $\square$

The next result establishes quadratic convergence of the sequence generated by Algorithm 2, provided the truncated Newton step is chosen when it is admissible. (Note that Theorem 13 in [7] establishes that ultimately the truncated Newton step will always be admissible.)

THEOREM 3.11. *Assume the conditions of Theorem* 3.10 *hold,* $0 < \beta < 1$, *and* $\hat{M}_*$ *is nonsingular for some limit point* x_* *of* $\{x_k\}$. *Let* s_k^N *be the Newton step* (3.19) *when it exists. Further, assume* $s_k = \alpha_k^*[s_k^N]$ *whenever* $\|D_k s_k^N\| < \Delta_k$ *and* $\alpha_k^*[s_k^N]$ *satisfies* (AS.4). *Then* $\{x_k\}$ *converges to* x_* *quadratically.*

Proof. From Theorem 3.10, $\{\Delta_k\}$ is bounded away from zero. But under our assumptions, $\{D_k^{-1} g_k\}$ converges to zero and $\{x_k\}$ converge to x_*. Hence $\{D_k s_k^N\}$ converges to zero where s_k^N is the Newton step:

$$\hat{M}_k D_k s_k^N = -\hat{g}_k.$$

Next we prove that $s_k = \alpha_k^*[s_k^N]$ for sufficiently large k. It suffices to prove that, for sufficiently large k, $\|D_k \alpha_k^*[s_k^N]\| < \Delta_k$ and $\alpha_k^*[s_k^N]$ satisfies (AS.4). Since $\{D_k s_k^N\}$ converges to zero, $\|D_k \alpha_k^*[s_k^N]\| < \Delta_k$ for sufficiently large k. Using definition (2.14), for sufficiently large k,

$$\alpha_k^*[s_k^N] - s_k^N = \tau_k^* \theta_k s_k^N - \tau_k^* s_k^N + \tau_k^* s_k^N - s_k^N,$$

where $\tau_k^* = \min\{1, \alpha_k\}$. From Lemma 11 in [7], $|\tau_k^* - 1| = O(\|x_k - x_*\|)$. But $|\theta_k - 1| = O(\|s_k^N\|)$. Since $0 < \beta < 1$ by assumption, for sufficiently large k, $\beta < \theta_k \tau_k^* \leq 1$ and $\|\alpha_k^*[s_k^N] - s_k^N\| = O(\|x_k - x_*\|^2)$. This implies that $0 \leq \theta_k^2 \tau_k^{*2} \leq \theta_k \tau_k^*$. Using (2.14), for k sufficiently large,

$$\psi_k(\alpha_k^*[s_k^N]) \leq \theta_k \tau_k^* \left(g_k^T s_k^N + \frac{1}{2} s_k^{N^T} M_k s_k^N \right) < \beta \psi_k^*[-D_k^{-2} g_k],$$

i.e., $\alpha_k^*[s_k^N]$ satisfies (AS.4) for sufficiently large k. Hence, under the assumptions, $s_k = \alpha_k^*[s_k^N]$ for k sufficiently large. In addition, following (iv) in Theorem 3.10, $\alpha_k^*[s_k^N]$ yields a successful iteration for sufficiently large k, i.e.,

$$f(x_k + \alpha_k^*[s_k^N]) - f(x_k) < \mu \left(g_k^T s_k^N + \frac{1}{2} s_k^{N^T} M_k s_k^N \right).$$

Using Theorem 13 in [7], $\{x_k\}$ converges quadratically to x_*. □

4. Preliminary numerical experiments. In this section we report on preliminary experiments with the practical trust region method, Algorithm 2, on a set of standard test problems of low dimension. The method solves these problems quite satisfactorily, indicating that this approach has practical potential.

We implemented the "practical trust region algorithm" described in Figure 2.2 in a straightforward manner. Either $s_k = \alpha_k^*[p_k]$, where p_k is a solution to the trust region subproblem (2.10) or $s_k = \alpha_k^*[-D_k^{-2} g_k]$. The exact implementation is described in Figure 4.1.

The computed step s_k satisfies the conditions

$$\psi_k(s_k) < \beta \psi_k(\alpha_k^*[-D_k^{-2} g_k]), \ \|D_k s_k\| \leq \Delta_k, \ x_k + s_k \in \text{int}(\mathcal{F})$$

and

$$\psi_k(s_k) < \psi_k(\alpha_k^*[p_k]), \ \|D_k s_k\| \leq \Delta_k, \ x_k + s_k \in \text{int}(\mathcal{F}).$$

Note that it is easy to verify that condition (AS.4) can be replaced with

$$\psi_k(s_k) < \beta \psi_k(\alpha_k^*[-D_k^{-2} g_k]), \ \|D_k s_k\| \leq \beta_0 \Delta_k, \ x_k + s_k \in \text{int}(\mathcal{F})$$

The method implemented
Let $\mu = 0.25, \beta = 0.1, \eta = 0.75, \gamma_0 = 0.0625, \gamma_1 = 0.5, \gamma_2 = 2, \Lambda_l = 1$, and $x_0 \in \mathrm{int}(\mathcal{F})$
For $k = 0, 1, \ldots$

1. Compute $f(x_k)$, g_k, $\nabla^2 f(x_k)$, and C_k; define the quadratic model $\psi_k(s_k) = g_k^T s_k + \frac{1}{2} s_k^T (\nabla^2 f(x_k) + C_k) s_k$.
2. Compute a solution p_k of (2.10). Compute

$$\rho_k^c = \frac{\psi_k(\alpha_k^*[p_k])}{\psi_k(\alpha_k^*[-D_k^{-2} g_k])}.$$

   If $\rho_k^c > \beta$, $s_k = \alpha_k^*[p_k]$. Otherwise, $s_k = \alpha_k^*[-D_k^{-2} g_k]$.
3. Compute

$$\rho_k^f = \frac{f(x_k + s_k) - f(x_k) + \frac{1}{2} s_k^T C_k s_k}{\psi_k(s_k)}.$$

4. If $\rho_k^f > \mu$ then $x_{k+1} = x_k + s_k$. Otherwise $x_{k+1} = x_k$.
5. Update the scaling matrix D_k and Δ_k as specified.

Updating trust region size Δ_k

1. If $\rho_k^f < 0$ then $\Delta_{k+1} = \gamma_0 \Delta_k$.
2. If $0 \le \rho_k^f \le \mu$ then $\Delta_{k+1} = \max\{\gamma_0 \Delta_k, \gamma_1 \|D_k s_k\|\}$.
3. If $\rho_k^f \ge \eta$ then
   if $\rho_k^c > \eta$ then
   $\Delta_{k+1} = \max\{\Delta_k, \gamma_2 \|D_k s_k\|\}$
   else
   if $\Delta_k > \Lambda_l$ and $\rho_k^c \le \mu$ then
   $\Delta_{k+1} = \max\{\gamma_1 \Delta_k, \|D_k s_k\|\}$.
4. Otherwise, $\Delta_{k+1} = \Delta_k$.

FIG. 4.1. *The interior and trust region method implemented.*

and condition (AS.6) can be replaced with

$$\psi_k(s_k) < \beta^q \psi_k(\alpha_k^*[p_k]), \; \|D_k s_k\| \le \beta_0^q \Delta_k, \; x_k + s_k \in \mathrm{int}(\mathcal{F}).$$

Thus, the implemented method has the convergence properties listed in Theorems 3.10 and 3.11.

The experiments were done on a Sun (Sparc) workstation using Matlab 4.0 [16]. The stopping criteria used were

$$\hat{M}_k > 0 \text{ and } \psi_k(s_k) < 0.5 * 10^{-12}.$$

The test problems are taken from [9]. However, the starting points as described in [9] may not be strictly feasible. Assume that x_{start} is the starting point specified in [9]. We modify the starting points as follows:

$$x_{0i} = l_i + 0.1 * (u_i - l_i) \text{ if } x_{start\,i} < l_i + 100\epsilon,$$
$$x_{0i} = u_i - 0.1 * (u_i - l_i) \text{ if } x_{start\,i} > u_i - 100\epsilon,$$

TABLE 4.1
Experiments with a practical trust region method for bounded constrained problems.

Number of Function/Gradient Evaluations				
PROB	n	NEW		CGT
GENROSE U	8	45	38	42
GENROSE C	8	15	14	15
CHAINROSE U	25	23	19	20
CHAINROSE C	25	28	23	18
DEGENROSE U	25	30	30	95
DEGENROSE C	25	28	23	17
GENSING U	20	27	27	10
GENSING C	20	23	22	4
CHAINSING U	20	26	26	18
CHAINSING C	20	22	21	3
DEGENSING U	20	26	26	155
DEGENSING C	20	35	34	3
GENWOOD U	8	68	57	107
GENWOOD C	8	10	9	5
CHAINWOOD U	8	57	48	77
CHAINWOOD C	8	10	9	5
HOSC45 U	10	28	27	19
HOSC45 C	10	10	9	12
BROYDEN1A U	30	14	14	11
BROYDEN1A C	30	14	13	8
BROYDEN1B U	30	7	7	7
BROYDEN1B C	30	9	8	6
BROYDEN2A U	30	19	16	14
BROYDEN2A C	30	24	22	10
BROYDEN2B U	30	9	9	9
BROYDEN2B C	30	15	14	9
TOINTBROY U	30	8	8	8
TOINTBROY C	30	12	11	8
TRIG U	10	13	11	7
TRIG C	10	16	14	8
TOINTTRIG U	10	8	6	13
TOINTTRIG C	10	8	6	10
CRAGGLEVY U	8	33	31	24
CRAGGLEVY C	8	30	29	20
PENALTY U	15	24	24	27
PENALTY C	15	29	29	80
AUGMLAGN U	15	29	26	31
AUGMLAGN C	15	46	44	31
BROWN1 U	10	20	20	27
BROWN1 C	10	31	30	27
BROWN3 U	10	9	9	7
BROWN3 C	10	10	9	6
BVP U	10	21	21	4
BVP C	10	20	20	4
VAR U	20	12	12	6
VAR C	20	12	11	6

where $\epsilon \approx 10^{-16}$ is the machine precision.

In Table 4.1, we report the number of function and gradient evaluations taken by the method to obtain the required accuracy. The number of function evaluations required by the method in [9] is given in the last column as a relative comparison (the exact Hessian is used to obtain these numbers). Our implemented method computes the trust region solutions while the method in [9] uses conjugate gradient. We point out that the starting points and the stopping criteria of the two methods are different. The approximate solutions obtained by the two methods may also differ.

A subspace version of this interior trust region method is currently under investigation for large-scale problems. Preliminary results are very encouraging and will be reported subsequently.

5. Conclusions. We have proposed a trust region approach to the bound-constrained nonlinear minimization problem. This approach generates strictly feasible iterates and possesses strong convergence characteristics. In particular, we have established second-order convergence properties. Moreover, the convergence results match the implementation in the sense that a global solution to a quadratic programming problem, with linear inequality constraints, is not required by the theory. Instead, an approximate minimization of a quadratic function subject to an ellipsoidal constraint is required (and achievable).

Our computational experiments on a well-known test collection of small-dimensional problems indicate that Algorithm 2 has practical potential. However, from a practical computational point of view we believe the real promise of the underlying ideas presented here is in the large-scale setting. The method as described is not directly suitable for large-scale problems—the computation of a (suitably) accurate solution to the trust region problem in high dimensions is probably too costly. Nevertheless, there is considerable scope for modifying and adapting the basic idea, with efficiency in mind, to the large-scale setting. This line of research is currently under investigation.

Finally, we remark that the trust region ideas developed in this paper for box constraints can be extended to the case where there are also linear equality constraints present, i.e., $\min\{f(x) : Ax = b, l \leq x \leq u\}$. This generalization is also the subject of current research.

REFERENCES

[1] A. BJORCK, *A direct method for sparse least squares problems with lower and upper bounds*, Numer. Math., 54 (1988), pp. 19–32.

[2] R. H. BYRD, R. B. SCHNABEL, AND G. SHULTZ, *Approximate solution of the trust region problem by minimization over two-dimensional subspaces*, Math. Programming, 40 (1988), pp. 247–263.

[3] T. F. COLEMAN AND C. HEMPEL, *Computing a trust region step for a penalty function*, SIAM J. Sci. Statist. Comput., 11 (1990), pp. 180–201.

[4] T. F. COLEMAN AND L. A. HULBERT, *A direct active set algorithm for large sparse quadratic programs with simple bounds*, Math. Programming, 45 (1989), pp. 373–406.

[5] ———, *A globally and superlinearly convergent algorithm for convex quadratic programs with simple bounds*, SIAM J. Optim., 3 (1993), pp. 298–321.

[6] T. F. COLEMAN AND Y. LI, *A Reflective Newton Method for Minimizing a Quadratic Function Subject to Bounds on the Variables*, Tech. report TR 92-1315, Computer Science Department, Cornell University, Ithaca, NY, 1992.

[7] ———, *On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds*, Math. Programming, 67 (1994), pp. 189–224.

[8] A. R. CONN, N. I. M. GOULD, AND P. L. TOINT, *Global convergence of a class of trust region algorithms for optimization with simple bounds*, SIAM J. Numer. Anal., 25 (1988), pp. 433–460.

[9] ———, *Testing a class of methods for solving minimization problems with simple bounds on the variables*, Math. Comp., 50 (1988), pp. 399–430.

[10] R. FLETCHER, *An algorithm for solving linearly constrained optimization problems*, Math. Programming, 2 (1972), pp. 133–165.

[11] R. FLETCHER AND M. P. JACKSON, *Minimization of a quadratic function of many variables subject only to lower and upper bounds*, J. Inst. Math. Appl., 14 (1974), pp. 159–174.

[12] D. M. GAY, *Computing optimal locally constrained steps*, SIAM J. Sci. Statist. Comput., 2 (1981), pp. 186–197.

[13] P. GILL AND W. MURRAY, *Minimization Subject to Bounds on the Variables*, Tech. report NAC 71, National Physical Laboratory, England, 1976.

[14] J. J. JÚDICE AND F. M. PIRES, *Direct methods for convex quadratic programs subject to box constraints*, Investigacao Peracional, 9 (1989), pp. 23–56.

[15] P. LOTSTEDT, *Solving the minimal least squares problem subject to bounds on the variables*, BIT, 24 (1984), pp. 206–224.

[16] C. B. MOLER, J. LITTLE, S. BANGERT, AND S. KLEIMAN, *ProMatlab User's guide*, MathWorks, Sherborn, MA, 1987.

[17] J. J. MORÉ, *Recent developments in algorithms and software for trust region methods*, in Mathematical Programming: The State of the Art, M. G. A. Bachem and E. B. Dorte, eds., Springer-Verlag, Berlin, 1983.

[18] J. J. MORÉ AND D. SORENSEN, *Computing a trust region step*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 553–572.

[19] J. J. MORÉ AND G. TORALDO, *Algorithms for bound constrained quadratic programming problems*, Numer. Math., 55 (1989), pp. 377–400.

[20] D. P. O'LEARY, *A generalized conjugate gradient algorithm for solving a class of quadratic programming problems*, Linear Algebra Appl., 34 (1980), pp. 371–399.

[21] U. ÖREBORN, *A Direct Method for Sparse Nonnegative Least Squares Problems*, Ph.D. thesis, Department of Mathematics, Linköping University, Linköping, Sweden, 1986.

[22] G. A. SCHULTZ, R. B. SCHNABEL, AND R. H. BYRD, *A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties*, SIAM J. Numer. Anal., 22 (1985), pp. 47–67.

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

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

[25] E. K. YANG AND J. W. TOLLE, *A class of methods for solving large convex quadratic programs subject to box constraints*, Math. Programming, 51 (1991), pp. 223–228.