An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds

Thomas F. Coleman
Yuying Li*

TR 93-1342
May 1993

Department of Computer Science
Cornell University
Ithaca, NY 14853-7501

*Research 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, and in part by NSF, AFOSR, and ONR 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.

An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds

Thomas F. Coleman and Yuying Li[1]
Computer Science Department
Cornell University
Ithaca, New York 14853

April 28, 1993

Abstract. We propose a new trust region approach for minimizing a nonlinear function subject to simple bounds. By choosing an appropriate quadratic model and scaling matrix at each iteration, we show that it is not necessary to solve a quadratic programming subproblem, with linear inequalities, to obtain an improved step using the trust region idea. Instead, a solution to a trust region subproblem is defined by minimizing a quadratic function subject only to an ellipsoidal constraint. The iterates generated by these methods are always strictly feasible. Our proposed methods reduce to a standard trust region approach for the unconstrained problem when there are no upper or lower bounds on the variables. Global and quadratic convergence of the methods is established; preliminary numerical experiments are reported.

[1] Research 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, and in part by NSF, AFOSR, and ONR 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.

1. Introduction. In this paper, we consider the problem of locating 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 $int(\mathcal{F}) \stackrel{\text{def}}{=} \{x : l < x < u\}$.

We propose a new strictly feasible trust region approach for problem (1.1). Global convergence to a second-order point is achieved, 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, some of which are restricted to quadratic (in some cases convex quadratic) objective functions and some are more general (e.g., [1, 4, 5, 7, 8, 9, 10, 12, 14, 15, 16, 20, 21, 22, 26]). 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 a solution to a quadratic subproblem with a bound on the step:

$$\min_{s\in\Re^n} \{\psi_k(s) \stackrel{\text{def}}{=} g_k^T s + \frac{1}{2} s^T B_k s \; : \; \|\bar{D}_k s\|_2 \le \Delta_k\}. \tag{1.2}$$

Here 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.

A general scheme for unconstrained minimization of $f(x)$ is described in Figure 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.

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, 13, 19, 23, 24, 25].

Trust region methods have also been developed for the solution of linearly constrained optimization problems (e.g., [9] and [11]). A quadratic trust region subproblem with linear inequalities is usually approximately solved to obtain an improved point [9]. An iterative procedure must to be used to solve the subproblem. For example, Fletcher [11] proposed an algorithm for the linearly constrained optimization subproblem

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

in which the subproblem is of the form

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

Algorithm 0: Unconstrained Trust Region Method
For $k = 0, 1, \cdots$
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 model ψ_k, the scaling matrix $\bar{D}_k$ and Δ_k.

Updating Trust Region Size
Let $0 < \mu < \eta < 1$ and $\gamma_1 < 1 < \gamma_2$ be given
1. If $\rho_k^f \leq \mu$ 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 $\rho_k^f \geq \eta$ then set $\Delta_{k+1} \in [\Delta_k, \gamma_2 \Delta_k]$.

FIG. 1. *Trust Region Method for Unconstrained Minimization*

As pointed out in [18], convergence theory for trust region methods based on quadratic programming subproblems – such as (1.3) – usually require that the computed trial step be a global solution to the subproblem. However, the subproblem is usually solved by methods which guarantee local optimality at best. Therefore, there is a mismatch between theory and practise for trust region methods based on quadratic programming subproblems (with linear inequalities).

In this paper, we propose trust region methods for (1.1) without the need to solve a general quadratic programming subproblem at each iteration. These methods are related to the line search based reflection methods proposed by Coleman and Li [7] – scaling strategies and requirement of strict feasibility are common to both approaches. The primary difference is that in [7] 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 methods are developed by forming a quadratic model with an appropriate quadratic function and scaling matrix such that there is no need to handle the constraints explicitly. In particular, it is possible to obtain an approximate trust region solution which can guarantee second-order convergence by simply solving an unconstrained trust region subproblem and then satisfying the feasibility requirement through further restricting the step if necessary. Our proposed methods become standard trust region algorithms 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 the convergence in §3. In §4, preliminary numerical results for small dense problems are presented. The computational investigation of the method for large problems will be presented in a subsequent paper.

2. Trust Region Method for Bound-Constrained Problems. In this section, we propose a trust region method for bound-constrained problems by choosing an appropriate scaling matrix $\bar{D}_k$ and quadratic model $\psi_k(x)$. We motivate our choice by examining the optimality conditions for (1.1).

First, we define a vector function $v(x): \Re^n \to \Re^n$ as follows.

DEFINITION 1. *The vector $v(x) \in \Re^n$ is defined:*

(i). *If $\nabla f(x)_i < 0$ and $u_i < \infty$ then $v_i \stackrel{\text{def}}{=} x_i - u_i$.*
(ii). *If $\nabla f(x)_i \geq 0$ and $l_i > -\infty$ then $v_i \stackrel{\text{def}}{=} x_i - l_i$.*
(iii). *If $\nabla f(x)_i < 0$ and $u_i = \infty$ then $v_i \stackrel{\text{def}}{=} -1$.*
(iv). *If $\nabla f(x)_i \geq 0$ 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. In addition, 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 i^{th} diagonal component equal to $|v_i(x)|$.

Optimality conditions for problem (1.1) are well-established. Assuming feasibility, *first-order necessary conditions* for x_* to be a local minimizer are:

$$\text{first order:} \begin{cases} \nabla f(x_*)_i = 0 & \text{if } l_i < (x_*)_i < u_i, \\ \nabla f(x_*)_i \leq 0 & \text{if } (x_*)_i = u_i, \\ \nabla f(x_*)_i \geq 0 & \text{if } (x_*)_i = l_i. \end{cases} \tag{2.2}$$

Equivalently, $D_*^{-2} g_* = 0$. Second-order conditions involve the Hessian matrix of f, $H = H(x) \stackrel{\text{def}}{=} \nabla^2 f(x)$. 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[2]: if a feasible point x_* is a local minimizer of (1.1) then $D_*^{-2} g_* = 0$ and $H_*^{Free_*} \geq 0$ where $H_*^{Free_*}$ is the submatrix of $H_* = H(x_*)$ corresponding to the index set $Free_*$

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

DEFINITION 2. *A point $x \in \mathcal{F}$ is* **nondegenerate** *if, for each index i:*

$$\nabla f(x)_i = 0 \implies l_i < x_i < u_i. \tag{2.3}$$

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

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

With 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 [6], we consider the following diagonal system of nonlinear equations:

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

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 $x_k \in int(\mathcal{F})$. Strictly speaking v_i may not be differentiable at a point where $\nabla f_i = 0$; however, $D^{-2}(x)\nabla f(x)$ is continuous at such points. Moreover, Coleman and Li [6] show that it is easy to define a Jacobian matrix at such points to allow for a second-order Newton process.

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

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

where $J^v(x) \in \Re^{n\times n}$ plays the role of the "Jacobian" matrix of $|v(x)|$. If all the components of l and u are finite, we define $J^v = \text{diag}(\text{sgn}(\nabla f))$. If variable x_i has a finite lower bound and an infinite upper bound (or vice-versa) and $\nabla f_i = 0$, we define $J_{ii}^v = 0$ at such a point.

In our presentation, $B(x)$ is an approximation to $H(x) \stackrel{\text{def}}{=} \nabla^2 f(x)$ and $g(x) \stackrel{\text{def}}{=} \nabla f(x)$. Based on the Newton step for system (2.4), we define our quadratic model in the same way as in [6]:

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

where

$$\begin{aligned} C(x) &\stackrel{\text{def}}{=} D(x)\text{diag}(g(x))J^v(x)D(x), \\ M(x) &\stackrel{\text{def}}{=} B(x) + C(x). \end{aligned}$$

Note that $C(x)$ is a positive semi-definite diagonal matrix.

Define

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

The following lemma can be easily proved.

LEMMA 1. *Assume that $x_* \in \mathcal{F}$ and $B(x_*) = H(x_*)$.*
(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 semi-definite and $\hat{g}_* = 0$ if and only if the second order necessary conditions are satisfied.*

The results of Lemma 1 imply that x_* is a local minimizer of (1.1) if and only if $w = 0$ is a solution to

$$\min_w \{\hat{\psi}_k(w) : \|w\|_2 \le \Delta_k\} \tag{2.7}$$

where $x_k = x_*$. Hence solving the subproblem (2.7) is a reasonable step to attempt when x_k is not a local minimizer. Let $s = D_k^{-1} w$. Subproblem (2.7) is equivalent to the following problem in the original variable space:

$$\min_s \{\psi_k(s) : \|D_k s\|_2 \le \Delta_k\}. \tag{2.8}$$

Moreover, in the neighborhood of a local minimizer, the Newton step to (2.4) is a solution to the trust region subproblem (2.8) if the trust region size Δ_k is sufficiently large.

The purpose of the scaling matrix D_k in (2.8) is distinctively different from that in an unconstrained trust region method. The scaling matrix D_k measures the distance to the boundary of the feasible region. Its purpose is to prevent a step directly towards a boundary point. In contrast, a scaling matrix used in the unconstrained trust region method is usually employed for numerical reasons: the scaling matrix helps to improve the conditioning of the problem.

If a bound–constrained problem (1.1) is badly scaled, the subproblem (2.8) can be replaced by

$$\min_s \{\psi_k(s) : \|D_k \bar{D}_k s\|_2 \le \Delta_k\},$$

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. To emphasize the role of the scaling matrix D_k, we assume that $\bar{D}_k = I$ in our presentation.

Next we illustrate that it is possible to develop trust region methods for the bound-constrained problem (1.1) based on (2.8). First we introduce a few notations and assumptions.

In our presentation, p_k denotes a solution to (2.8). Assume that $d_k \in \Re^n$. The scalar α_k denotes the stepsize along d_k to the boundary:

$$\alpha_k = \min\{\max\{\frac{l_i - x_{ki}}{d_{ki}}, \frac{u_i - x_{ki}}{d_{ki}}\} : 1 \le i \le n\}. \tag{2.9}$$

If problem (1.1) is unconstrained, i.e., $l = -\infty$ and $u = \infty$, we define $\alpha_k = +\infty$. 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.,

$$\psi_k^*[d_k] \stackrel{\text{def}}{=} \psi_k(s_k^*) = \min\{\psi_k(s) : s = \tau d_k, \|D_k s\| \le \Delta_k, x_k + s \in \mathcal{F}\}. \tag{2.10}$$

Since we always require $x_k \in int(\mathcal{F})$, a possible step-back may be necessary to stay strictly feasible. 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 given below. Let τ_k^* denote the minimizer along d_k within the feasible trust region, i.e., $\tau_k^* = \mathrm{argmin}_\tau\{\psi_k(\tau d_k) : \|\tau D_k d_k\| \le \Delta_k, x_k + \tau d_k \in \mathcal{F}\}$, $\theta_k \in [\theta_l, 1]$ for some $0 < \theta_l < 1$ and $\theta_k - 1 = O(\|d_k\|)$. Then

$$\alpha_k^*[d_k] \stackrel{\text{def}}{=} \theta_k \tau_k^* d_k \stackrel{\text{def}}{=} \begin{cases} \tau_k^* d_k & \text{if } x_k + \tau_k^* d_k \in int(\mathcal{F}), \\ \theta_k \tau_k^* d_k & \text{otherwise.} \end{cases} \tag{2.11}$$

The above definition implies that $\theta_k = 1$ if $x_k + \tau_k^* d_k \in int(\mathcal{F})$.

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 $x \in \mathcal{L}$, $\|g\|_\infty < \chi_g$.

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 matrix $\{D_k^{-1}\}$. Condition (AS.3) is a weak assumption and is satisfied if the gradient $\nabla f(x)$ 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 will 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 the essential ideas. The second method 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 (2.8): a solution p_k to the trust region subproblem (2.8) is obtained and then truncated, i.e., $s_k = \alpha_k^*[p_k]$, to ensure strict feasibility.

The essential idea of trust region methods is to use the trust region size to ensure sufficient decrease of the objective function. Consider the trust region approach to unconstrained problems: 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 $f(x)$.) To obtain first order convergence of unconstrained trust region methods, a sufficient reduction of the quadratic model ψ_k within the trust region is guaranteed if

$$\psi_k(s_k) \le \beta \min\{\psi_k(s) : s = \tau \bar{D}_k^{-T} g_k, \|\bar{D}_k s\| \le \Delta_k\}, \quad \|\bar{D}_k s_k\| \le \beta_0 \Delta_k \tag{2.12}$$

for some constants $\beta, \beta_0 > 0$. In our notation, (2.12) is the same as

$$\psi_k(s_k) \leq \beta\psi_k^*[-\bar{D}_k^{-T} g_k], \quad \|\bar{D}_k s_k\| \leq \beta_0 \Delta_k. \tag{2.13}$$

Since our quadratic model $\psi_k(s)$ is defined to include the constraint information, a natural extension of the definition of ρ_k^f is given by

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

Similar to unconstrained trust region methods, ρ_k^f measures the agreement between the nonlinear function $f(x)$ and its quadratic approximation. As we will prove from Lemma 2 in §3, for our trust region approaches, if s_k satisfies

$$\psi_k(s_k) < \beta\psi_k^*[-D_k^{-2} g_k], \tag{2.14}$$

the trust region model is sufficiently reduced for first order convergence.

Unfortunately a truncated step along an exact solution p_k of (2.8) may not sufficiently reduce the quadratic model ψ_k, due to the effect of truncation, i.e., (2.14) is not guaranteed when $s_k = \alpha_k^*[p_k]$ and p_k solves (2.8). However, a step along the scaled steepest descent direction $-D_k^{-2} g_k$ does produce enough reduction of $\psi_k(s)$. Since, for any trust region subproblem with nonzero gradient, the trust region solution approaches the gradient direction if the trust region size is reduced to zero, the trust region size can be used to ultimately guarantee sufficient decrease. In particular, the trust region size updating can be used to force the condition

$$\rho_k^c = \frac{\psi_k(s_k)}{\psi_k^*[-D_k^{-2} g_k]} > \beta, \quad i.e., \quad \psi_k(s_k) < \beta\psi_k^*[-D_k^{-2} g_k].$$

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 described in Figure 2. 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 $\rho_k^f > \eta$ and $\rho_k^c > \eta$. Otherwise, an iteration is *unsuccessful.*

In §3, we will prove that the double-trust region method has reasonable convergence properties under the nondegeneracy assumption.

2.2. A Practical Trust Region Method. In the last section, we proposed a double-trust region method for bound-constrained problems (1.1) through solving an ellipsoidal trust region subproblem (2.8). In this algorithm, sufficient decrease of the quadratic model is achieved through monitoring the ratio ρ_k^c and adjusting the trust region size accordingly. Since ρ_k^c is determined by $\psi_k(s_k)$, an exact solution to the subproblem (2.8) 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 trust region problem (2.8) 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 practical model algorithm.

Algorithm 1: The Double-Trust Region Method
$x_0 \in int(\mathcal{F})$
For $k = 0, 1, \cdots$

1. Compute $f(x_k)$, g_k, H_k, and C_k; Define the quadratic model $\psi_k(s) = g_k^T s + \frac{1}{2} s^T (H_k + C_k) s$.
2. Compute p_k, a solution to (2.8).
3. Compute

$$s_k = \alpha_k^*[p_k],$$
$$\rho_k^c = \frac{\psi_k(s_k)}{\psi_k^*[-D_k^{-2} g_k]},$$
$$\rho_k^f = \frac{f(x_k + s_k) - f(x_k) + \frac{1}{2} s_k^T C(x_k) s_k}{\psi_k(s_k)}.$$

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 model ψ_k, the scaling matrix D_k and Δ_k as specified.

Updating Δ_k for the Double-Trust Region Method
Let $0 < \mu, \beta < \eta < 1$ and $\gamma_1 < 1 < \gamma_2$ be given

1. if $\rho_k^f \leq \mu$ 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 $\rho_k^f \geq \eta$ then
   - if $\rho_k^c \geq \eta$, set $\Delta_k \in [\Delta_k, \gamma_2 \Delta_k]$,
   - if $\beta < \rho_k^c < \eta$, set $\Delta_k \in [\gamma_1 \Delta_k, \Delta_k]$,
   - if $\rho_k^c < \beta$, set $\Delta_k \in (0, \gamma_1 \Delta_k]$.

FIG. 2. *Double-Trust Region Method for Minimization Subject to Bounds*

It is clear that 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 only needed to force the condition $\rho_k^f > \mu$.

In Figure 3, 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, we have allowed more freedom than usual in the adjustment of Δ_k to permit further reduction in Δ_k thus encouraging the use of the trust region step (2.8).

In order 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

(AS.4) $\psi_k(s_k) < \beta \psi_k^*[-D_k^{-2} g_k]$
$\|D_k s_k\| \leq \beta_0 \Delta_k$, $x_k + s_k \in int(\mathcal{F})$.

In other words, the condition $\rho_k^c > \beta$ is assumed. More explicitly, we require that $\psi_k(s_k)$ to be less

Algorithm 2: A More Practical Model
$x_0 \in int(\mathcal{F})$
For $k = 0, 1, \cdots$

1. Compute $f(x_k)$, g_k, H_k, and C_k; Define the quadratic model $\psi_k(s) = g_k^T s + \frac{1}{2} s^T (H_k + C_k) s$.
2. Compute s_k, based on (2.8), such that $x_k + s_k \in int(\mathcal{F})$.
3. Compute

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

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 model ψ_k, the scaling matrix D_k and Δ_k as specified.

Updating Trust Region Size Δ_k
Let $0 < \mu < \eta < 1$, $\gamma_1 < 1 < \gamma_2$ and $0 < \Lambda_l$ 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. 3. *Trust Region Method for Minimization Subject to Bounds*

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 step of Algorithm 1. For Algorithm 2, an iteration is *successful* if the condition $\rho_k^f > \eta$ holds. Otherwise, an iteration is *unsuccessful*.

Condition (AS.4) can be easily satisfied for $0 < \beta < 1$. Let w_k be the solution to $\min\{\psi(s) : D_k^2 s = \nu g_k,\ \|D_k s\| \le \Delta_k,\ x_k + s \in \mathcal{F}\}$. Then $s_k = w_k$ satisfies (AS.4) except for the possible violation of $x_k + s_k \in int(\mathcal{F})$. Assume that $x_k + w_k \notin int(\mathcal{F})$. Since $\psi_k(s)$ is continuous, a small step-back $s_k = \theta w_k$ where $0 < \theta < 1$ can ensure both the condition $x_k + s_k \in int(\mathcal{F})$ 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 condition is satisfied. To achieve this we make the following stronger assumptions on the quadratic model and the approximate solution:

(AS.5) $\psi_k(s) = \nabla f(x_k)^T s + \frac{1}{2} s^T (\nabla^2 f(x_k) + C(x_k)) s$.

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

Since both conditions (AS.4) and (AS.6) can be satisfied by simply solving a quadratic trust region subproblem $\min_s\{\psi_k(s) : \|D_k s\| \le \Delta_k\}$, 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 following 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 $\psi_k(s_k)$ 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 restrictive. There are many ways of computing such approximations. As another example, one can consider the reflection techniques used in [6]. 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 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 $C(x) = 0$ and $D(x) = I$ 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., [18]) since the feasibility constraints are always satisfied.

3. Convergence Properties. The convergence proofs for the double-trust region algorithm in Figure 2, and the practical algorithm in Figure 3, follow the same main steps. Lemmas 3 and 4 make it possible to present the proofs for both algorithms simultaneously in a clean fashion. The major results are Theorems 5, 6, 11, and 12.

The main difference between the double-trust region method and the practical method is that condition (AS.4) is satisfied through the ratio ρ_k^c for the first while assumed by the latter. However, a common property of the two methods is that, for any successful step, (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.

Notational note: In all expressions to follow, the norm symbol without subscript, $\|\cdot\|$, refers to the 2-norm.

The following result is required to express (AS.4) in a manageable form. It is similar to Lemma (4.8) in [18].

LEMMA 2. *Assume that (AS.1)-(AS.3) are satisfied. If s_k satisfies (AS.4) then*

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

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

$$\phi(\tau) \stackrel{\text{def}}{=} \psi[\tau d_k].$$

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

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

Since $\alpha_k > 0$ and the components of d_k have the same sign as that of $-g_k$, the absolute value of the numerator equals $|v_i|$, i.e.,

$$\alpha_k = \frac{|v_{ki}|}{|d_{ki}|} = \frac{|v_{ki}| \|\hat{g}_k\|}{|v_{ki}| |g_{ki}|}.$$

Hence

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

By 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 $\tau_k^* = \|\hat{g}_k\|/\mu_k$ and thus

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

If $\tau_k^* = \Delta_k$ then $\mu_k \Delta_k \le \|\hat{g}_k\|$ and hence

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

If $\tau_k^* = \alpha_k$, $\mu_k\alpha_k \le \|\hat{g}_k\|$ and hence

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

Since $\psi_k(s_k) \le \beta\phi(\tau^*)$, the result follows from the above estimates. ∎

Assume that s_k is a successful step from either Algorithm 1 or Algorithm 2. From Lemma 2,

$$\begin{aligned} f(x_k) - f(x_{k+1}) &- \frac{1}{2}s_k^T C_k s_k \ge -\mu\psi_k(s_k) \\ &\ge \frac{1}{2}\beta\mu\|\hat{g}_k\|\min\{\Delta_k, \frac{\|\hat{g}_k\|}{\|\hat{M}_k\|}, \frac{\|\hat{g}_k\|}{\|g_k\|_\infty}\}. \end{aligned}$$

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

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

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

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

This inequality is important for the convergence proof.

Next, in Theorems 5 and 6, 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 solution to the trust region subproblem (2.8). Using Theorem (3.11) in [18], there exists a parameter λ_k and upper triangular matrix $R_k \in \Re^{n\times n}$ such that

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

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

$$(\lambda_k I + D_k^{-1} C_k D_k^{-1}) p_k = -D_k^{-2}(g_k + B_k p_k). \tag{3.2}$$

LEMMA 3. *Suppose that $\{x_k\}$ is a sequence generated by Algorithm 1. Assume that problem (1.1) is nondegenerate, (AS.1) and (AS.2) 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 $\{D_k p_k\}$ converges to zero and $\|p_k\| \leq \chi_D \|D_k p_k\|$, $\{p_k\}$ converges to zero. By the assumption that $\liminf_{k\to\infty}\{\|D_k^{-1} g_k\|\} > 0$ and (3.2), it is clear that $\{\lambda_k\}$ converges to $+\infty$. But $\{D_k^{-2}(g_k + B_k p_k)\}$ is bounded and the problem is nondegenerate. We claim that $\lim_{k\to\infty} \alpha_k = +\infty$ where α_k is the stepsize to the boundary of the constraints. This can be easily seen from the additional fact that, for k sufficiently large, the components of $-g_k$ and p_k have the same sign if the corresponding component of the limit point g_* is nonzero. Subsequently, $s_k = \alpha_k^*[p_k] = p_k$. Thus, for k sufficiently large

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

Hence

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

and therefore

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

■

LEMMA 4. *Assume that $\{\Delta_k\}$ is updated by Algorithm 2. If $\rho_k^f > \eta$ for sufficiently large k then $\{\Delta_k\}$ is bounded away from zero.*

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

$$\Delta_k \geq \min\{\gamma_1 \Delta_l, \Delta_{\bar{k}}\}. \tag{3.3}$$

First, it is clear that (3.3) is true when $k = \bar{k}$.

Assume that (3.3) is true for k, i.e., $\Delta_k \geq \min\{\gamma_1\Delta_l, \Delta_{\bar{k}}\}$. If $\Delta_k < \Delta_l$, $\Delta_{k+1} \geq \Delta_k \geq \min\{\gamma_1\Delta_l, \Delta_{\bar{k}}\}$. If $\Delta_k \geq \Delta_l$, $\Delta_{k+1} \geq \min\{\gamma_1\Delta_l, \Delta_{\bar{k}}\}$.

Hence (3.3) is true for all $k \geq \bar{k}$. Therefore $\{\Delta_k\}$ is bounded away from zero. ■

The proof of the following theorem is a slight modification of Theorem (4.10) in [18].

THEOREM 5. *Assume that $f : \Re^n \to \Re$ is continuously differentiable and bounded below on $\mathcal{L}$, (AS.1) and (AS.2) hold. For Algorithm 2, if $\{s_k\}$ satisfies (AS.4), then*

$$\liminf_{k\to\infty} \|\mathrm{diag}(|v_k|^{\frac{1}{2}})g_k\| = 0. \tag{3.4}$$

For Algorithm 1, (3.4) is true under the further assumption that problem (1.1) is nondegenerate.

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

$$\sum_{k=1}^{\infty} \Delta_k < +\infty. \tag{3.5}$$

If there are a finite number of successful iterations then $\Delta_{k+1} \leq \gamma_1\Delta_k$ for all k sufficiently large and then (3.5) clearly holds. If there is an infinite sequence $\{k_i\}$ of successful iterations then inequality (3.1) shows that

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

Now the updating rules of Algorithm 1 & 2 imply that

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

and thus (3.5) holds in this case as well.

Next we prove that (3.5) implies that $\{|\rho_k^f - 1|\}$ converges to zero. First,

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

and hence (3.5) shows that $\{x_k\}$ converges. Now (AS.1) and (AS.2) imply that

$$\begin{aligned} |\psi_k(s_k) - \nabla f(x_k)^T s_k - \frac{1}{2}s_k^T C_k s_k| &= |g_k^T s_k + \frac{1}{2}s_k^T B_k s_k - \nabla f(x_k)^T s_k| \\ &\leq \frac{1}{2}\chi_B\chi_D^2\|D_k s_k\|^2. \end{aligned}$$

But $\|D_k s_k\| \leq \Delta_k$. Therefore

$$|f(x_k + s_k) - f(x_k) + \frac{1}{2}s_k^T C_k s_k - \psi_k(s_k)| \leq \frac{1}{2}\chi_B\chi_D^2\Delta_k^2$$

Since Lemma 2 implies that

$$-\psi(s_k) \geq \frac{1}{2}\beta\epsilon\Delta_k$$

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

For Algorithm 2, using Lemma 4, $\{\Delta_k\}$ cannot converge to zero. This contradicts (3.5) and establishes the result.

For Algorithm 1, using Lemma 3, Δ_k is not decreased for sufficiently large k. Thus $\{\Delta_k\}$ cannot converge to zero. This contradicts (3.5) and establishes the result. ■

The next theorem establishes that $\{\text{diag}(|v_k|^{\frac{1}{2}})\nabla f(x_k)\}$ converges to zero. This result is established despite the fact that $\{D_k\}$ is not uniformly bounded. This may be somewhat surprising, perhaps, since the analogous convergence result in the unconstrained setting requires the sequence of diagonal scaling matrices to be uniformly bounded.

THEOREM 6. *Assume (AS.1-2) hold and $\nabla f(x)$ is uniformly continuous on $\mathcal{L}$. If $\{x_k\}$ is generated by Algorithm 2 and (AS.4) holds for s_k. Then*

$$\lim_{k\to\infty} \|\text{diag}(|v_k|^{\frac{1}{2}})\nabla f(x_k)\| = 0. \tag{3.6}$$

Result (3.6) holds for Algorithm 1 *when problem (1.1) is nondegenerate.*

Proof. The proof is by contradiction and is the same for Algorithm 1 and Algorithm 2. (The nondegeneracy assumption is needed for Algorithm 2 because the proof uses Theorem 5.)

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 5 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

$$\|\hat{g}_k\| \geq \epsilon_2, \quad m_i \leq k < l_i, \quad \|\hat{g}_{l_i}\| < \epsilon_2. \tag{3.7}$$

If the k-th iteration is successful then

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

Since $\{f(x_k)\}$ converges, $\{f(x_k) - f(x_{k+1})\}$ converges to zero. From $\|x_{k+1} - x_k\| \leq \beta_0\chi_D\Delta_k$, it follows that, for sufficiently large i,

$$f(x_k) - f(x_{k+1}) \geq \epsilon_3\|x_{k+1} - x_k\|, \quad m_i \leq k < l_i, \tag{3.8}$$

where $\epsilon_3 = (\frac{1}{2}\beta\mu\epsilon_2)/(\beta_0\chi_D)$. Using (3.8) 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 ∇f and the convergence of $\{f(x_k)\}$ can now be used to deduce that

$$\|g_{m_i} - g_{l_i}\| \leq \epsilon_2, \tag{3.9}$$

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 j-th component of $g(x_*)$ 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 $\{\text{diag}(|v_{m_i}|^{\frac{1}{2}} - |v_{l_i}|^{\frac{1}{2}})g_{l_i}\}$ converges to zero. Therefore, for i sufficiently large,

$$\|D_{m_i}^{-1}(D_{m_i} - D_{l_i})D_{l_i}^{-1}g_{l_i}\| = \|\text{diag}(|v_{l_i}|^{\frac{1}{2}} - |v_{m_i}|^{\frac{1}{2}})g_{l_i}\| \leq \epsilon_2. \tag{3.10}$$

Using the triangle inequality for any m and l

$$\|\hat{g}_m\| \leq \|D_m^{-1}\|\|g_m - g_l\| + \|D_m^{-1}(D_m - D_l)D_l^{-1}g_l\| + \|\hat{g}_l\|. \tag{3.11}$$

Combining (3.11) with (3.7), (3.9) and (3.10), 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 §2, we assume the following quadratic model: $\psi_k(s) = \nabla f(x_k)^T s + \frac{1}{2}s^T(\nabla^2 f(x_k) + C(x_k))s$. Moreover, in addition to (AS.4), the condition (AS.6) holds, i.e., the reduction of the quadratic model satisfies

$$\begin{aligned} &\psi_k(s_k) \leq \beta \min\{\psi_k(s) : s = \tau p_k,\ \|D_k s\| \leq \Delta_k,\ x_k + s \in \mathcal{F}\}, \\ &\|D_k s_k\| \leq \beta_0 \Delta_k,\ x_k + s_k \in int(\mathcal{F}), \end{aligned}$$

where p_k is a solution to the unconstrained subproblem

$$\min_s\{\psi_k(s) : \|D_k s\| \leq \Delta_k\}.$$

Before we state the second order convergence result, several technical lemmas are required. First, we quote Lemma (4.10) in [19] below.

LEMMA 7. *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. Using Theorem (3.11) in [18], if p_k is a solution to (2.8), there exists a parameter λ_k such that

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

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

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

LEMMA 8. *Assume that (AS.6) is satisfied. Then*

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

where α_k is the stepsize along p_k to the boundary and p_k is a solution to the trust region subproblem (2.8).

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. Let τ^* to be the minimizer of $\phi(\tau)$ in $[0, \min\{1, \alpha_k\}]$.

It is easy to see that

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

But $\tau^2 \leq \tau$ since $\tau \leq 1$. Hence, from (AS.6),

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

∎

The following lemma provides estimates of the reductions in the objective function and quadratic model.

LEMMA 9. *Assume that the conditions of Theorem 6 and (AS.6) hold. Furthermore, $\{x_k\}$ is any sequence 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,*

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

and if s_k is successful,

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

Proof. Using Lemma 8,

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

where α_k is the stepsize along p_k to the boundary:

$$\alpha_k = \min\{\max\{\frac{l_i - x_{ki}}{p_{ki}}, \frac{u_i - x_{ki}}{p_{ki}}\} : 1 \leq i \leq n\}. \tag{3.12}$$

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,

$$\min(x_k - l, u - x_k) + |g(x_k)| > 2\epsilon_0 e, \quad e = [1, \cdots, 1]^T \in \Re^n.$$

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

Following Theorem 6, $\{\text{diag}(|v_k|^{\frac{1}{2}})g(x_k)\}$ converges to zero. Hence, for sufficiently large k,

$$\|\text{diag}(|v_k|)g(x_k)\|_\infty < \epsilon_0^2.$$

Assume that k is sufficiently large. If $|g_{k_i}| \leq \epsilon_0$, then

$$\min\{x_{k_i} - l_i, u_i - x_{k_i}\} > \epsilon_0.$$

Hence, from (3.2), (3.12) and

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

we obtain:

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

If $|g_{k_i}| > \epsilon_0$, then $|v_{k_i}| \leq \epsilon_0$. If $\alpha_k = |v_{k_i}|/|p_{k_i}|$, then from (3.2) and (3.12),

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

If $\alpha_k \neq |v_{k_i}|/|p_{k_i}|$, the numerator determining α_k is greater than ϵ_0. Hence, using (3.12)

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

■

When $\hat{M}_k$ is positive definite, we denote the Newton step for (2.8) by

$$s_k^N \stackrel{\text{def}}{=} -D_k^{-1}\hat{M}_k^{-1}\hat{g}_k, \quad \text{i.e.} \quad \hat{M}_k D_k s_k^N = -\hat{g}_k. \tag{3.13}$$

LEMMA 10. *If the sequence of trust region subproblem (2.8) solution $\{p_k\}$ converges to zero, $\{x_k\}$ converges to x_* and $\hat{M}_*$ is positive definite, then*

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

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:

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

Since p_k is a solution to (2.8), $\tau_* = \min\{1, \alpha_k\}$ in (2.11). By definition (2.11), $\theta_l \leq \theta_k \leq 1, \theta_k - 1 = O(\|p_k\|)$. Since $\hat{M}_k$ is positive definite for sufficiently large k, we have that $p_k^T M_k p_k > 0$. Therefore,

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

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

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

If there exist variables on the boundary at x_*, since $\{p_k\}$ converges to zero, $\alpha_k = \frac{|v_{kj}|}{p_{kj}} = \frac{|g_{kj}|+\lambda_k}{|g_{kj}+(B_k p_k)_j|}$ for some $v_j(x_*) = 0$ and $g_j^* \neq 0$. This means that

$$\liminf_{k\to\infty}\{\alpha_k\} \geq 1. \tag{3.14}$$

Assume that $\epsilon_0 > 0$ is a lower bound on the eigenvalue of $\hat{M}_*$. Since $p_k = s_k^N$ if $\lambda_k = 0$, using Lemma 8,

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

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

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

In addition,

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

Hence

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

■

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

THEOREM 11. *Let the level set $\mathcal{L} = \{x \in \Re^n : f(x) \le f(x_0),\ x \in \mathcal{F}\}$ be compact and $f : \mathcal{F} \to \Re$ be twice continuously differentiable on $\mathcal{L}$. 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*

(a) The sequence $\{\hat{g}_k\}$ converges to zero.

(b) If every limit point is nondegenerate, then there is a limit point x_* with $\hat{M}_*$ positive semidefinite.

(c) If x_* is an isolated nondegenerate limit point, then $\hat{M}_*$ is positive semidefinite.

(d) 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.

(a) The sequence $\{\hat{g}_k = \mathrm{diag}(|v_k|^{\frac{1}{2}})\nabla f(x_k)\}$ converges to zero: this has been proved in Theorem 6.

(b) If $\{\lambda_k\}$ is not bounded away from zero, the result immediately follows.
We prove that $\{\lambda_k\}$ is not bounded away from zero by contradiction. Assume that $\lambda_k \ge \epsilon > 0$. First we show that $\{\Delta_k\}$ converges to zero.
From Lemma 9, we have that, for sufficiently large k,

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

where

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

Moreover, for sufficiently large k and successful steps

$$f(x_k) - f(x_{k+1}) \ge \frac{1}{2}\beta^q\mu\hat{\epsilon}_k\epsilon\Delta_k^2. \tag{3.15}$$

The rest of arguments are similar to the proof of Lemma 5. If there are finite number of successful steps, $\{\Delta_k\}$ converges to zero. Otherwise, let $\{k_i\}$ be the infinite sequence of successful iterations. Inequality (3.15) implies that there exists a constant $\epsilon_1 > 0$ such that $\hat{\epsilon}_{k_i} > \epsilon_1$. This fact and inequality (3.15) imply that

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

Now the updating rules of Algorithm 1 & 2 imply that

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

Hence $\{\Delta_k\}$ converges to zero. Since $\|s_k\| \leq \|D_k^{-1} D_k s_k\| \leq \chi_D \beta_0^q \Delta_k$ 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,

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

Hence

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

Now a standard estimate is that

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

and thus the last two inequalities 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 4, $\{\Delta_k\}$ 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:

$$\alpha_k = \min\{\max\{\frac{l_i - x_{ki}}{p_{ki}}, \frac{u_i - x_{ki}}{p_{ki}}\}.$$

If all variables are free at a limit point x_*, then $\alpha_* = +\infty$. Otherwise, consider a limit point x_* with some variables at their bounds. Since this limit point is nondegenerate and $\lim_{k\to\infty} \operatorname{diag}(|v_k|) g_k = 0$, using (3.2), we have

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

for sufficiently large k with $g_{*j} \neq 0$. But $\lambda_k > \epsilon$. Thus the corresponding limit $\alpha_* > 1$. 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. This is again a contradiction. In conclusion, there is a limit point with $\hat{M}_*$ positive semidefinite.

(c) If $\{x_k\}$ converges to x_*, the result follows from (b). If $\{x_k\}$ does not converge then Lemma 7 applies. Thus, if $\{x_{l_j}\}$ is the subsequence guaranteed by Lemma 7 then $\Delta_{l_j} \geq \frac{1}{\chi_D \beta_0^q}\epsilon$. From Lemma 9, $\{\lambda_{l_j}\Delta_{l_j}\}$ converges to zero. Hence $\{\lambda_{l_j}\}$ converges to zero. Thus $\hat{M}_*$ is positive semidefinite.

(d) 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 (c). Since $\psi_k(s_k) = g_k^T s_k + \frac{1}{2}s_k^T M_k s_k < 0$ we have that

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

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 $\{D_k^{-1}g_k = \mathrm{diag}(|v_k|^{\frac{1}{2}})\nabla f(x_k)\}$ converges to zero. Following Lemma 7, $\{x_k\}$ 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 $\{\Delta_k\}$ is bounded away from zero. Assume that $\epsilon > 0$ is a lower bound on the eigenvalues of $\hat{M}_k$. Using Lemma 10, for sufficiently large k,

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

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

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

Hence, 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

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

yield that $\rho_k^f > \eta$ for k sufficiently large.

For Algorithm 2, using Lemma 4, we immediately conclude that $\{\Delta_k\}$ is bounded away from zero.

Now we consider Algorithm 1. Using Lemma 10, 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 and $\{\Delta_k\}$ is bounded away from zero.

∎

THEOREM 12. *Assume the conditions of Theorem 11 hold and $\hat{M}_*$ is nonsingular for some limit point x_* of $\{x_k\}$. Let s_k^N be the Newton step (3.13) when it exists. Moreover, $s_k = \alpha_k^*[s_k^N]$ whenever $\|D_k s_k^N\| < \Delta_k$. Then $\{x_k\}$ converges to x_* quadratically.*

Proof. From Theorem 11, $\{\Delta_k\}$ 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:

$$\hat{M}_k D_k^{-1} s_k^N = -\hat{g}_k.$$

Using definition (2.11), for sufficiently large k,

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

where $\tau_k^* = \min\{1, \alpha_k\}$. From Lemma 11 in [6], $|\tau_k^* - 1| = O(\|x_k - x_*\|)$. But $|\theta_k - 1| = O(\|s_k^N\|)$. Hence $\|s_k - s_k^N\| = O(\|x_k - x_*^2\|)$. Using Theorem 10 in [6], $\{x_k\}$ 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 3 in a straightforward manner. Either $s_k = \alpha_k^*[p_k]$ where p_k is a solution to the trust region subproblem (2.8), or $s_k = \alpha_k^*[-D_k^{-2}g_k]$. The exact implementation is described below.

The computed step s_k satisfies the condition

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

and

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

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

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

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 $x_0 \in int(\mathcal{F})$

For $k = 0, 1, \cdots$

1. Compute $f(x_k)$, g_k, H_k, and C_k; Define the quadratic model $\psi_k = g_k^T s_k + \frac{1}{2} s_k^T (H_k + C_k) s_k$.
2. Compute a solution p_k of (2.8). 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

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

4. If $\rho_k^f > \mu$ then $x_{k+1} = x_k + s_k$. Otherwise $x_{k+1} = x_k$.
5. Update the model ψ_k, 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 $\Delta_{k+1} = \max\{\gamma_0 \Delta_k, \gamma_1 \|D_k^{-1} s_k\|\}$.
3. If $\rho_k^f \ge \eta$ then
   if $\rho_k^c > \eta$ then
      $\Delta_{k+1} = \max\{\Delta_k, \gamma_2 \|D_k^{-1} s_k\|\}$
   else
      if $\Delta_k > \Delta_l$ and $\rho_k^c \le \mu$ then
         $\Delta_{k+1} = \max\{\gamma_1 \Delta_k, \|D_k^{-1} s_k\|\}$.

and condition (AS.6) can be replaced with

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

Thus, the implemented method has the convergence properties listed in Theorems 11 and 12.

The experiments were done on a Sun (Sparc) workstation using MATLAB 4.0 [17]. The stopping criteria used are:

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

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:

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

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

In Table 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 cited in the last column as a relative comparison. 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. The stopping accuracy of the method in [9] is not usually as stringent as that used in our experiments – the method in [9] terminates when the norm of the projected gradient is less than 10^{-6}.

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).

The computational experiments reported, 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 – 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 is a subject of current 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 a subject of current research.

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

TABLE 1

Experiments with a Practical Trust Region Method for Bounded Constrained Problems

REFERENCES

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

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

[3] T. F. COLEMAN AND C. HEMPEL, *Computing a trust region step for a penalty function*, SIAM Journal on Scientific and Statistical Computing, 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*, Mathematical Programming, 45 (1989), pp. 373–406.

[5] ———, *A globally and superlinearly convergent algorithm for convex quadratic programs with simple bounds*, Tech. Rep. TR 90-1092, Computer Science Department, Cornell University, February, 1990 (to appear in SIAM Journal on Optimization).

[6] T. F. COLEMAN AND Y. LI, *On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds*, Tech. Rep. TR 92-1314, Computer Science Department, Cornell University, 1992.

[7] ———, *A reflective Newton method for minimizing a quadratic function subject to bounds on the variables*, Tech. Rep. TR 92-1315, Computer Science Department, Cornell University, 1992.

[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 Journal on Numerical Analysis, 25 (1988), pp. 433–460.

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

[10] R. S. DEMBO AND U. TULOWITZKI, *On the minimization of quadratic functions subject to box constraints*, Tech. Rep. B 71, Yale University, 1983.

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

[12] R. FLETCHER AND M. P. JACKSON, *Minimization of a quadratic function of many variables subject only to lower and upper bounds*, Journal of the Institute for Mathematics and its Applications, 14 (1974), pp. 159–174.

[13] D. M. GAY, *Computing optimal locally constrained steps*, SIAM Journal on Scientific and Statistical Computing, 2 (1981), pp. 186–197.

[14] P. GILL AND W. MURRAY, *Minimization subject to bounds on the variables*, Tech. Rep. Report NAC 71, National Physical Laboratory, England, 1976.

[15] J. J. JÚDICE AND F. M. PIRES, *Direct methods for convex quadratic programs subject to box constraints*, departamento de matemática, Universidade de Coimbra, 3000 Coimbra, Portugal, 1989.

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

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

[18] 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.

[19] J. J. MORÉ AND D. SORENSEN, *Computing a trust region step*, SIAM Journal on Scientific and Statistical Computing, 4 (1983), pp. 553–572.

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

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

[22] U. ÖREBORN, *A direct method for sparse nonnegative least squares problems*, PhD thesis, Department of Mathematics, Linköping University, Linköping, Sweden, 1986.

[23] 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 Journal on Numerical Analysis, 22(1) (1985), pp. 47–67.

[24] D. SORENSEN, *Trust region methods for unconstrained optimization*, SIAM Journal on Numerical Analysis, 19 (1982), pp. 409–426.

[25] T. STEIHAUG, *The conjugate gradient methods and trust regions in large scale optimization*, SIAM Journal on Numerical Analysis, 20 (1983), pp. 626–637.

[26] E. K. YANG AND J. W. TOLLE, *A class of methods for solving large convex quadratic programs subject to box constraints*, tech. rep., Department of Operations Research, University of North Carolina, Chapel Hill, North Carolina, 1988.