Journal of Computational Mathematics, Vol.19, No.3, 2001, 309–322.

A ROBUST TRUST REGION ALGORITHM FOR SOLVING GENERAL NONLINEAR PROGRAMMING*[1)]

Xin-wei Liu
(*State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing 100080, China*)
(*Faculty of Science and Arts, Hebei University of Technology, Tianjin 300130, China*)

Ya-xiang Yuan
(*State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing 100080, China*)

Abstract

The trust region approach has been extended to solving nonlinear constrained optimization. Most of these extensions consider only equality constraints and require strong global regularity assumptions. In this paper, a trust region algorithm for solving general nonlinear programming is presented, which solves an unconstrained piecewise quadratic trust region subproblem and a quadratic programming trust region subproblem at each iteration. A new technique for updating the penalty parameter is introduced. Under very mild conditions, the global convergence results are proved. Some local convergence results are also proved. Preliminary numerical results are also reported.

Key words: Trust region algorithm, Nonlinear programming.

1. Introduction

Trust region methods are iterative. As a strategy of globalization, the trust region approach was introduced into solving unconstrained optimization and proved to be efficient and robust. An excellent survey was given by Moré(1983). The associated research with trust region methods for unconstrained optimization can be found in Fletcher(1980), Powell(1975), Sorensen(1981), Shultz, Schnabel and Byrd(1985), Yuan(1985). The solution of the trust region subproblem is still an active studying area, see Stern and Wolkowicz(1994), Peng and Yuan(1997) et al.

Since the 80's the trust region approach has been extended to solving nonlinear constrained optimization. Most of these extensions consider only equality constraints, and the global convergence theories are based on strong global regularity assumptions, for example, see Byrd, Schnabel and Shultz(1987), Vardi(1985), Omojokun(1989), Powell and Yuan(1991), Dennis, El-Alem and Maciel(1997), Dennis and Vicente(1997). At each iteration of an algorithm given by Omojokun(1989), the trial step consists of a normal direction step and a null space step. Similarly, Dennis, El-Alem and Maciel(1997) considered the method which replaced the normal component by a quasi-normal direction and developed its global convergence theory. Dennis and Vicente(1997) proved that under suitable conditions their method will converge to the second-order optimal point. For general constrained optimization, Fletcher(1981) proposed a trust region method which is based on the L_1 nonsmooth exact penalty function. Burke and

* Received March 3, 1999.
[1)] Research partially supported by Chinese NSF grants 19525101, 19731001, and by State key project 96-221-04-02-02. It is also partially supported by Hebei provincial doctoral fund.

Han(1989), Liu and Yuan(1998) have extented Fletcher's approach to other penalty functions. Burke(1992) presented a general framework for trust region algorithms for constrained problems. Without requiring any regularity assumption, Burke proved that his method converges to the points which satisfies certain first-order optimality conditions. Similar to Fletcher(1981) and Burke(1992), Yuan(1995) proposed a new trust region algorithm for solving the optimization with equality and inequality constraints. Under mild conditions, Yuan(1995) proved the global convergence of his algorithm and established local convergence results.

In this paper, we consider the general nonlinear programming problem

$$\min \quad f(x) \tag{1.1}$$

$$s.t. \quad c_i(x) = 0, \quad i \in E, \tag{1.2}$$

$$c_i(x) \geq 0, \quad i \in I, \tag{1.3}$$

where $f(x)$, $c_i(x)(i \in E \cup I)$ are real valued continuously differentiable functions on $\Re^n$, $E = \{1, 2, \cdots, m_e\}$ and $I = \{m_e+1, \cdots, m\}$ are two index sets with the integers m_e and m satisfying $m \geq m_e \geq 0$. If $m_e = m > 0$, (1.1)-(1.3) is the optimization with only equality constraints.

Successive quadratic programming (SQP) methods are very efficient for solving problem (1.1)–(1.3), see Han(1977), Powell(1978), Burke and Han(1989), Burke(1989). At each iteration, the original SQP method, developed by Wilson, Han and Powell, generates a new approximate to the solution by the procedure

$$x^+ = x + sd, \tag{1.4}$$

where x is the current point, d is a search direction which minimizes a quadratic model subject to linearized constraints and s is the steplength along the direction and is decided by some line search procedure. Under certain conditions, SQP methods converge superlinearly. The requisite consistency of the linearized constraints of the QP subproblem, however, is its serious limitation. In order to handle the inconsistency of the linearized constraints, Liu and Yuan(1998) presented a modified SQP algorithm which solves an unconstrained piecewise quadratic subproblem and a quadratic programming subproblem at each iteration. The algorithm is a natural extension of the original SQP method since it solves the same subproblems as the original SQP method at the feasible points of the original problem, and it coincides with the original method when the iterates are sufficiently close to the solution. Moreover, in order to ensure the fast rate convergence, it seems reasonable to use the second-order information to generate the normal direction, instead of using the first-order term only (for example, see Burke(1989) and Burke and Han(1989)). For optimization with only equality constraints, the normal direction and the null space direction are independent, so the search direction can be computed parallelly (see Liu(1998)).

In this paper, we present a new trust region algorithm for problem (1.1)-(1.3). The new algorithm is based on the SQP method of Liu and Yuan(1998). The trial step is computed by solving an unconstrained piecewise quadratic trust region subproblem and a quadratic programming trust region subproblem at each iteration. A motivation for using trust region techniques is that trust region approach is robust and it can applied to ill-conditioned problems. Our algorithm is similar to Burke(1992) and Yuan(1995), but there remain fundmental differences. For equality constrained case, our method is also similar to the null space and range space approach analyzed by Dennis, El-Alem and Maciel(1997) and Dennis and Vicente(1997). A new technique for updating the penalty parameter is introduced. Under very mild conditions, the global convergence results are proved. Local superlinear convergence results are also proved. Preliminary numerical results are also reported.

This paper is organized as follows. In section 2 we present our algorithm. Some global convergence results of our algorithm are proved in section 3. The local analyses are given in section 4. In section 5, we report some preliminary numerical results.

Throughout this paper, we use the following notations: $g_k = \nabla f(x_k)$, $c_k = c(x_k)$, $\nabla c_k = \nabla c(x_k)$, et al.

2. The Algorithm

Define

$$P(x,\mu) = f(x) + \mu\|c(x)_-\|, \tag{2.1}$$

where $\|\cdot\|$ is any norm on $\Re^n$, $\mu > 0$ is a penalty parameter. It is well known that (2.1) is an exact penalty function for problem (1.1)-(1.3). If $\|\cdot\|$ is selected to be the L_1 norm, then (2.1) is the merit function used in Han(1977), Powell(1978).

Suppose that x_k is the current iterate. Similar to Liu and Yuan(1998), let

$$I_k = \{i \in I : \ c_i(x_k) \le 0\}, \tag{2.2}$$
$$J_k = I_k \cup E, \tag{2.3}$$
$$\bar{J}_k = \{i \in I : \ i \notin J_k\}. \tag{2.4}$$

In practice, the condition $c_i(x_k) \le 0$ is relaxed by $c_i(x_k) \le \epsilon$, where ϵ is a very small positive tolerance number. Let Δ_k be the current trust region radius and B_k be the actual or approximate Hessian of Lagrangian

$$L(x,\lambda) = f(x) + \lambda^T c(x) \tag{2.5}$$

at x_k, where $\lambda \in \Re^m$ is a multiplier vector. Firstly we solve an unconstrained trust region subproblem

$$\min \ \ \psi_k(d) = \frac{1}{2} d^T B_k d + \mu_k \|(c_{J_k} + \nabla c_{J_k}^T d)_-\| \tag{2.6}$$
$$s.t. \ \ \|d\|_2 \le \delta \Delta_k, \tag{2.7}$$

where $0 < \delta < 1$ is a constant, and the norm used in (2.6) is the same as that in (2.1). For the rest of this paper, if it is not specified, the norm $\|\cdot\|$ is also the same as that in (2.1). The L_2 norm in (2.7) can also be replaced by any other norm. If the norm $\|\cdot\|$ in (2.6) is the L_1 or L_2 norm, subproblem (2.6)–(2.7) is to minimize a piecewise quadratic function within a ball.

Suppose that d_{k1} is a solution of (2.6)-(2.7), we choose $\tau_k \le 1$ close to 1 as much as possible such that

$$c_i(x_k) + \nabla c_i(x_k)^T (\tau_k d_{k1}) \ge 0, \ \ i \in \bar{J}_k, \tag{2.8}$$

where $\bar{J}_k$ is defined as (2.4). Then we solve the following quadratic programming trust region subproblem

$$\min \ \ \varphi_k(d) = g_k^T d + \frac{1}{2} d^T B_k d \tag{2.9}$$
$$s.t. \ \ \nabla c_i(x_k)^T d = 0, \ \ i \in E \tag{2.10}$$
$$\nabla c_i(x_k)^T d \ge 0, \ \ i \in I_k \tag{2.11}$$
$$\hat{c}_i(x_k) + \nabla c_i(x_k)^T d \ge 0, \ \ i \in \bar{J}_k \tag{2.12}$$
$$\|d + \tau_k d_{k1}\|_2 \le \Delta_k, \tag{2.13}$$

where $g_k = \nabla f(x_k)$, $\hat{c}_i(x_k) = c_i(x_k) + \nabla c_i(x_k)^T (\tau_k d_{k1})$. Let d_{k2} be a solution of (2.9)-(2.13), the trial step of our algorithm is defined by

$$s_k = \tau_k d_{k1} + d_{k2}. \tag{2.14}$$

To define the predicted reduction, we use the following function

$$\phi_{k+1}(s_k) = g_k^T s_k + \frac{1}{2} s_k^T B_k s_k + \mu_{k+1} \|(c_k + \nabla c_k^T s_k)_-\|, \tag{2.15}$$

where μ_{k+1} is the updated value of the penalty parameter at the kth iteration. The predicted reduction is defined by $Pred_k = \phi_{k+1}(0) - \phi_{k+1}(s_k)$. The actual reduction of the penalty function (2.1) from x_k to $x_k + s_k$ is

$$Ared_k = P(x_k, \mu_{k+1}) - P(x_k + s_k, \mu_{k+1}). \tag{2.16}$$

Let

$$r_k = \frac{Ared_k}{Pred_k}, \tag{2.17}$$

which plays an important role in accepting or rejecting the trial step and in updating the next trust region radius.

Now our algorithm is stated as follows.

Algorithm 2.1.

Step 0. Given $x_0 \in \Re^n$, $B_0 \in \Re^{n\times n}$,

$\Delta_0 \in \Re_+$, $\mu_0 \in \Re_+$ *and positive constants* δ, ϵ. *Evaluate* $f(x_0)$, $c(x_0)$, g_0, $\nabla c(x_0)$. *Let* $k = 0$;

Step 1. Identify the subscript subsets I_k *and* J_k *by (2.2) and (2.3). Calculate the trial step* s_k *by solving subproblems (2.6)-(2.7) and (2.9)-(2.13). If* $\|s_k\|_2 \le \epsilon$, *stop;*

Step 2. Update the value of the penalty parameter. If

$$\phi_k(0) - \phi_k(s_k) \ge \frac{\tau_k}{2}[\psi_k(0) - \psi_k(d_{k1})], \tag{2.18}$$

let $\mu_{k+1} = \mu_k$. *Otherwise, update* μ_k *by* μ_{k+1} *such that* $\mu_{k+1} \ge 2\mu_k$ *and*

$$\phi_{k+1}(0) - \phi_{k+1}(s_k) \ge \frac{\tau_k}{2}[\psi_k(0) - \psi_k(d_{k1})]; \tag{2.19}$$

Step 3. Computing r_k *by (2.17). If* $r_k > 0$, *let* $x_{k+1} = x_k + s_k$. *Otherwise,* $x_{k+1} = x_k$. *The trust region radius* Δ_k *is updated as follows:*

$$\Delta_{k+1} = \begin{cases} \max\{\Delta_k,\ 2\|s_k\|_2\}, & r_k \ge 0.9 \\ \Delta_k, & 0.1 \le r_k < 0.9, \\ \min\{\frac{1}{4}\Delta_k,\ \frac{1}{2}\|s_k\|_2\}, & r_k < 0.1; \end{cases} \tag{2.20}$$

Step 4. Evaluate $f(x_{k+1})$, $c(x_{k+1})$, $\nabla f(x_{k+1})$, $\nabla c(x_{k+1})$. *Update* B_k. *Let* $k = k+1$ *and go to Step 1.*

3. Global Convergence of the Algorithm

In this section, we study the global convergence of the algorithm.

Lemma 3.1. *Let* $\gamma_k = \|(c_{J_k})_-\| - \min_{\|d\|_2 \le \delta\Delta_k} \|(c_{J_k} + \nabla c_{J_k}^T d)_-\|$. *If* d_{k1} *is a solution of (2.6)-(2.7), then*

$$\psi_k(0) - \psi_k(d_{k1}) \ge \frac{1}{2}\mu_k\gamma_k \min\{1,\ \frac{\mu_k\gamma_k}{\|B_k\|_2\delta^2\Delta_k^2}\}. \tag{3.1}$$

Proof. Suppose that $\|\hat{d}_{k1}\|_2 \le \delta\Delta_k$ and

$$\|(c_{J_k} + \nabla c_{J_k}^T \hat{d}_{k1})_-\| = \min_{\|d\|_2 \le \delta\Delta_k} \|(c_{J_k} + \nabla c_{J_k}^T d)_-\|. \tag{3.2}$$

If $\hat{d}_{k1} = 0$, then $\gamma_k = 0$. Thus, (3.1) holds. Assume that $\hat{d}_{k1} \neq 0$, then for any $0 < \tau \leq 1$, we have $\|\tau \hat{d}_{k1}\|_2 \leq \delta\Delta_k$ and

$$\begin{aligned}\psi_k(d_{k1}) &\leq \frac{1}{2}\tau^2 \hat{d}_{k1}^T B_k \hat{d}_{k1} + \mu_k \|(c_{J_k} + \tau\nabla c_{J_k}^T \hat{d}_{k1})_-\| \\ &\leq \frac{1}{2}\tau^2 \|B_k\|_2 \|\hat{d}_{k1}\|_2^2 + \tau\mu_k(\|(c_{J_k} + \nabla c_{J_k}^T \hat{d}_{k1})_-\| - \|(c_{J_k})_-\|) \\ &\quad + \mu_k \|(c_{J_k})_-\|. \end{aligned} \tag{3.3}$$

Thus, $\psi_k(0) - \psi_k(d_{k1}) \geq -\frac{1}{2}\tau^2 \|B_k\|_2 \delta^2 \Delta_k^2 + \tau\mu_k\gamma_k, \ \ \forall 0 < \tau \leq 1.$

If $\tilde{\tau}_k = \frac{\mu_k\gamma_k}{\|B_k\|_2 \delta^2 \Delta_k^2} < 1$, then

$$\begin{aligned}\psi_k(0) - \psi_k(d_{k1}) &\geq \tilde{\tau}_k \mu_k \gamma_k - \frac{1}{2}(\tilde{\tau}_k)^2 \|B_k\|_2 \delta^2 \Delta_k^2 \\ &\geq \frac{1}{2}\frac{\mu_k^2\gamma_k^2}{\|B_k\|_2 \delta^2 \Delta_k^2}. \end{aligned} \tag{3.4}$$

Otherwise, $\mu_k\gamma_k \geq \|B_k\|_2 \delta^2 \Delta_k^2$. Thus we have

$$\psi_k(0) - \psi_k(d_{k1}) \geq \frac{1}{2}\mu_k\gamma_k. \tag{3.5}$$

The result of the lemma follows from (3.4) and (3.5).

In order to prove the global convergence results of the algorithm, we need to make the following assumptions:

Assumption 3.2. *(1) $f(x)$ and $c(x)$ are twice continuously differentiable on $\Re^n$; (2) $\{B_k\}$ is uniformly bounded, i.e. there exists a positive constant M such that $\|B_k\|_2 \leq M$ for all positive integer k; (3) $\{\Delta_k\}$ and $\{x_k\}$ are uniformly bounded.*

It should be noted that we do not make positive definiteness assumption on B_k. The trust region algorithm can circumvent the difficulties of SQP associated with indefinite Lagrangian Hessian or its approximation. If B_k is the exact Hessian of Lagrangian function and if the multipliers are bounded, Assumption 3.2(2) is implied by Assumption 3.2(1) and (3). On the other hand, the assumption on uniformly boundedness of $\{\Delta_k\}$ is not strict since in practice it is quite common to use $\Delta_{k+1} = \min\{\Delta^*, \ \max\{\Delta_k, \ 2\|s_k\|_2\}\}$ whenever $r_k \geq 0.9$, where $\Delta^* > 0$ is a constant.

In the analyses of this section, we do not preassume any regularity of constraints. Thus, Algorithm 2.1 may converge to some points other than Kuhn-Tucker points. Our analyses are similar to Burke(1992) and Yuan(1995), and they developed the convergence theories for their methods without the assumption of regularity.

By our definition, it is easy to see that either μ_k will remain unchanged for all large k or $\lim_{k\to\infty} \mu_k = \infty$. Firstly we have the following result:

Lemma 3.3. *Under Assumption 3.2, if* $\lim_{k\to\infty} \mu_k = \infty$*, then there exists a finite number c^* such that $c^* = \lim_{k\to\infty} \|(c_k)_-\|$.*

The proof of the above result is nearly the same as that of Lemma 4.2 of Yuan (1995), and therefore it is omitted. Based on the above result, we have the following lemma:

Lemma 3.4. *If $\lim_{k\to\infty} \mu_k = \infty$ and $c^* \neq 0$, then there is a convergent subsequence of $\{x_k\}$, its limit x^* is infeasible for (1.2)-(1.3) and is a stationary point of $\|c(x)_-\|$, that is, $0 \in \partial q(x^*)$ for $q(x) = \|c(x)_-\|$.*

Proof. Because $c^* \neq 0$, any accumulation point of $\{x_k\}$ is an infeasible point of the original problem (1.1)-(1.3). Let S be the set of all accumulation points of $\{x_k\}$. Then there must have a $x^* \in S$, which satisfies

$$\min_{d \in \Re^n} \|(c(x^*) + \nabla c(x^*)^T d)_-\| = \|c(x^*)_-\|, \tag{3.6}$$

from this equation the result of the lemma is followed. Otherwise, for any $x \in S$ we have

$$\min_{d \in \Re^n} \|(c(x) + \nabla c(x)^T d)_-\| < \|c(x)_-\|. \tag{3.7}$$

By the convexity of the norm, $\min_{\|d\|_2 \le 1} \|(c(x) + \nabla c(x)^T d)_-\| < \|c(x)_-\|$. Thus, it follows from the continuity of the norm that there is a $\omega > 0$ such that for sufficiently large k,

$$\min_{\|d\|_2 \le 1} \|(c_k + \nabla c_k^T d)_-\| \le \|(c_k)_-\| - \omega. \tag{3.8}$$

Suppose that $\|\hat{d}_k\|_2 \le 1$ such that $\|(c_k + \nabla c_k^T \hat{d}_k)_-\| = \min_{\|d\|_2 \le 1} \|(c_k + \nabla c_k^T d)_-\|$, since $\|(c_{J_k} + \nabla c_{J_k}^T \hat{d}_k)_-\| \le \|(c_k + \nabla c_k^T \hat{d}_k)_-\|$, we have

$$\|(c_{J_k} + \nabla c_{J_k}^T \hat{d}_k)_-\| \le \|(c_{J_k})_-\| - \omega. \tag{3.9}$$

Let $t_k = \min\{1, \delta\Delta_k\}$, then $\|t_k \hat{d}_k\|_2 \le \delta\Delta_k$. Thus, for sufficiently large k,

$$\gamma_k \ge \|c_{J_k -}\| - \|(c_{J_k} + \nabla c_{J_k}^T (t_k \hat{d}_k))_-\| \ge t_k \omega, \tag{3.10}$$

where γ_k is defined in Lemma 3.1. Since

$$\begin{aligned} &\phi_k(0) - \phi_k(s_k) - \frac{\tau_k}{2}(\psi_k(0) - \psi_k(d_{k1})) \\ &\ge \frac{\tau_k}{2}(\psi_k(0) - \psi_k(d_{k1})) - g_k^T s_k - \frac{1}{2} s_k^T B_k s_k + \frac{1}{2}\tau_k d_{k1}^T B_k d_{k1}, \end{aligned} \tag{3.11}$$

and by selection of τ_k, there is a $\tau_0 > 0$ such that $\tau_k \ge \tau_0$ for sufficiently large k, it follows from Assumption 3.2, Lemma 3.1 and $\mu_k \to \infty$ that for sufficiently large k,

$$\phi_k(0) - \phi_k(s_k) \ge \frac{\tau_k}{2}(\psi_k(0) - \psi_k(d_{k1})). \tag{3.12}$$

(3.12) contradicts Step 2 of Algorithm 2.1. This contradiction implies that the lemma is true.

Lemma 3.5. *Suppose that* $\lim_{k \to \infty} \mu_k = \infty$ *and* $c^* = 0$. *If* $\|d_{k1}\| = O(\|c_{k-}\|)$, *then any accumulation point of* $\{x_k\}$ *is a Fritz-John point of the original problem (1.1)-(1.3)(which is not necessarily a Kuhn-Tucker point of (1.1)-(1.3)).*

Proof. If $\|d_{k1}\| = O(\|c_{k-}\|)$, it follows from $\lim_{k \to \infty} \mu_k = \infty$ that there exists an infinitely set K such that x_k is infeasible for all $k \in K$. Because $c^* = 0$, any accumulation point of $\{x_k : k \in K\}$ is feasible for problem (1.1)-(1.3).

If

$$\lim_{k \to \infty} \frac{\min_{d \in Re^n} \|(c_k + \nabla c_k^T d)_-\|}{\|(c_k)_-\|} = 1, \tag{3.13}$$

then for any accumulation point x^* of $\{x_k\}$, we have $0 \in \partial q(x^*)$ for $q(x^*) = \|c(x)_-\|$. Since x^* is a feasible point, x^* is also a Fritz-John point of (1.1)-(1.3).

Suppose the result of the lemma does not hold. Then for sufficiently large $k \in K$, we have

$$\frac{\min_{d \in \Re^n} \|(c_k + \nabla c_k^T d)_-\|}{\|(c_k)_-\|} < 1. \tag{3.14}$$

Let $\eta_k = \min\{\delta\Delta_k, \|c_{k-}\|\}$, it follows from the above inequality and the convexity of $\|(c_k + \nabla c_k^T d)_-\|$ that

$$\frac{\min_{\|d\|_2 \le \eta_k} \|(c_k + \nabla c_k^T d)_-\|}{\|(c_k)_-\|} < 1. \tag{3.15}$$

If $\bar{d}_k \in \Re^n$ such that

$$\min_{\|d\|_2 \le \eta_k} \|(c_k + \nabla c_k^T d)_-\| = \|(c_k + \nabla c_k^T \bar{d}_k)_-\|, \tag{3.16}$$

then for sufficiently large $k \in K$, by (3.11),

$$\phi_k(s_k) - \phi_k(0) - \frac{\tau_k}{2}(\psi_k(d_{k1}) - \psi_k(0)) \le \frac{\tau_k}{2}(\psi_k(\bar{d}_k) - \psi_k(0)) + O(\|d_{k1}\|). \tag{3.17}$$

Suppose that $\{x_k : k \in \bar{K}\}(\bar{K} \subseteq K)$ is any convergent subsequence of $\{x_k : k \in K\}$. Then there must exist a $\tau_0 > 0$ such that for sufficiently large $k \in \bar{K}$, $\tau_k \ge \tau_0$. Thus, by (3.15), for sufficiently large $k \in \bar{K}$,

$$\begin{aligned} &\frac{\phi_k(s_k) - \phi_k(0) - \frac{\tau_k}{2}(\psi_k(d_{k1}) - \psi_k(0))}{\mu_k\|c_{k-}\|} \\ &\le \frac{\tau_0}{2}\frac{\|(c_k + \nabla c_k^T \bar{d}_k)_-\| - \|(c_k)_-\|}{\|(c_k)_-\|} + o(1) < 0, \end{aligned} \tag{3.18}$$

which contradicts $\mu_k \to \infty$.

Lemma 3.4 and Lemma 3.5 show that if $\lim_{k\to\infty} \mu_k = \infty$, Algorithm 2.1 may converge to some points other than Kuhn-Tucker points of (1.1)-(1.3). The next theorem illustrates that Algorithm 2.1 can converge to a Kuhn-Tucker point of the original problem if $\lim_{k\to\infty} \mu_k = \mu$ ($\mu > 0$ is a constant).

We suppose that $\|d_{k1}\|_2 \to 0$ when $\lim_{k\to\infty} \|c_{k-}\| = 0$, that is, the norm of the quasi-normal component of the trial step closes to zero as iterate closes to the feasible region. It can be seen that our assumption is weaker than a similar assumption of Dennis, El-Alem and Maciel(1997) and Dennis and Vicente(1997), where they require that the norm of the quasi-normal component of the trial step is not more than a fraction of the norm of the constraint violations.

Theorem 3.6. *Suppose that* $\lim_{k\to\infty} \mu_k = \mu$ *(*$\mu > 0$ *is a constant),* $\{x_k : k \in K\}$ *is a convergent subsequence of* $\{x_k\}$ *and* x^* *is its limite. If* $\|c(x^*)_-\| = 0$*, and the Mangasarian-Fromovitz constraint qualification conditions hold at* x^**, then* x^* *is a Kuhn-Tucker point of the original problem (1.1)-(1.3).*

Proof. Suppose that the theorem is not true. Then we claim that

$$\lim_{k\to\infty, k\in K} \Delta_{k+1} = 0. \tag{3.19}$$

Otherwise, there exists a constant $u > 0$ such that for sufficient large $k \in K$, we have

$$\Delta_k \ge u, \quad r_k \ge 0.1. \tag{3.20}$$

For sufficient large $k \in K$, there is a $\tau_0 > 0$ such that $\tau_k \ge \tau_0$, and $\mu_k = \mu$. Thus, for large $k \in K$,

$$Ared_k \ge 0.1 Pred_k \ge 0.05\tau_0[\psi_k(0) - \psi_k(d_{k1})] \tag{3.21}$$

Let d^* is the solution of the problem

$$\min \quad \bar{\varphi}^*(d) = g(x^*)^T d + \frac{1}{2} M \|d\|_2^2 \tag{3.22}$$

$$s.t. \quad \nabla c_i(x^*)^T d = 0, \;\; i \in E \tag{3.23}$$

$$\nabla c_i(x^*)^T d \ge 0, \;\; i \in I^* \tag{3.24}$$

$$c_i(x^*) + \nabla c_i(x^*)^T d \ge 0, \;\; i \in \bar{I}^* \tag{3.25}$$

$$\|d\|_2 \le \frac{u}{2}, \tag{3.26}$$

where $I^* = \{i \in I : \ c_i(x^*) = 0\}$. The supposition that x^* is not a Kuhn-Tucker point of the original problem implies that $d^* \neq 0$ and $\bar{\varphi}^*(d^*) < 0$. Thus, by the fact that $x_k \to x^*$, $d_{k1} \to 0$ and the perturbed lemma of quadratic programming (see Daniel(1973)), we have

$$\bar{\varphi}_k(0) - \bar{\varphi}_k(\bar{d}_{k2}) \geq -\frac{1}{2}\eta \tag{3.27}$$

for large $k \in K$, where $\eta = \bar{\varphi}^*(d^*)$, $\bar{d}_{k2}$ minimizes $g_k^T d + \frac{1}{2}M\|d\|_2^2$ on the feasible region of the subproblem (2.9)-(2.13). Thus, by

$$\begin{aligned}\phi_k(0) - \phi_k(s_k) &= \mu_k(\|(c_k)_-\| - \|(c_k + \nabla c_k^T s_k)_-\|) - g_k^T s_k - \frac{1}{2}s_k^T B_k s_k \\ &\geq \varphi_k(0) - \varphi_k(d_{k2}) + \tau_k[\psi_k(0) - \psi_k(d_{k1})] \\ &\quad -\tau_k d_{k1}^T(g_k + B_k d_{k2}) + \frac{1}{2}\tau_k(1-\tau_k)d_{k1}^T B_k d_{k1} \\ &\geq \bar{\varphi}_k(0) - \bar{\varphi}_k(\bar{d}_{k2}) + \tau_k[\psi_k(0) - \psi_k(d_{k1})] + O(\|d_{k1}\|_2),\end{aligned} \tag{3.28}$$

we have for sufficiently large $k \in K$, $Pred_k \geq -\frac{1}{4}\eta$. $r_k \geq 0.1$ implies that

$$P(x_k, \mu_{k+1}) - P(x_k + s_k, \mu_{k+1}) \geq -0.025\eta \tag{3.29}$$

for sufficiently large $k \in K$. Since $\lim_{k\to\infty, k\in K} P(x_k, \mu_{k+1}) = f(x^*)$, (3.29) can not hold for infinitely many k. This contradiction implies that (3.19) holds when the theorem is not true.

Now we suppose that (3.19) holds. Thus, by (2.20), $r_k < 0.1$ for sufficiently large $k \in K$ and

$$\lim_{k\to\infty, k\in K} s_k = 0. \tag{3.30}$$

If x^* is not a Kuhn-Tucker point of (1.1)-(1.3), $\tilde{d}^*$ is the solution of the subproblem

$$\min \quad \bar{\varphi}^*(d) = g(x^*)^T d + \frac{1}{2}M\|d\|_2^2 \tag{3.31}$$

$$s.t. \quad \nabla c_i(x^*)^T d = 0, \ \ i \in E \tag{3.32}$$

$$\nabla c_i(x^*)^T d \geq 0, \ \ i \in I^* \tag{3.33}$$

$$c_i(x^*) + \nabla c_i(x^*)^T d \geq 0, \ \ i \in \bar{I}^* \tag{3.34}$$

$$\|d\|_2 \leq 1, \tag{3.35}$$

then $\tilde{d}^* \neq 0$, $\tilde{\eta} = \bar{\varphi}^*(\tilde{d}^*) < 0$, and $\min\{1, \Delta_k\}\tilde{d}^*$ is a feasible solution of the problem

$$\min \quad \bar{\varphi}^*(d) = g(x^*)^T d + \frac{1}{2}M\|d\|_2^2 \tag{3.36}$$

$$s.t. \quad \nabla c_i(x^*)^T d = 0, \ \ i \in E \tag{3.37}$$

$$\nabla c_i(x^*)^T d \geq 0, \ \ i \in I^* \tag{3.38}$$

$$c_i(x^*) + \nabla c_i(x^*)^T d \geq 0, \ \ i \in \bar{I}^* \tag{3.39}$$

$$\|d\|_2 \leq \Delta_k. \tag{3.40}$$

Hence, $\bar{\varphi}^*(\tilde{d}_k) \leq \tilde{\eta}\min\{1, \Delta_k\}$, where $\tilde{d}_k$ solves the problem (3.36)-(3.40). Therefore, the quadratic programming perturbation lemma and (3.28) imply that

$$Pred_k \geq -0.25\tilde{\eta}\min\{1, \Delta_k\}. \tag{3.41}$$

If there exists an infinite subset $\tilde{K} \subseteq K$ such that $\Delta_k < 1$ for all $k \in \tilde{K}$, then

$$\lim_{k\to\infty, k\in\tilde{K}} r_k = 1, \tag{3.42}$$

since

$$|Ared_k - Pred_k| \leq o(\Delta_k), \tag{3.43}$$

which contradicts $r_k < 0.1$; Otherwise, $\Delta_k \geq 1$ for all sufficiently large $k \in K$ and

$$Pred_k \geq -0.25\tilde{\eta}. \tag{3.44}$$

By (3.30) and $|Ared_k - Pred_k| \leq o(\|s_k\|_2)$, we have $r_k \to 1 (k \in K, k \to \infty)$, which is a contradiction with $r_k < 0.1$. Thus, (3.19) is not true. The contradiction proves the theorem.

4. Local Discussion

For analyses of local convergence of the algorithm, we need the following assumption:

Assumption 4.1. *(1) $x_k \to x^*$, where x^* is a vector such that $\|c(x^*)_-\| = 0$; (2) $\nabla c_i(x^*)(i \in E \cup I^*)$ are linearly independent; (3) $\mu_k = \mu$ for sufficiently large k; (4) $\{B_k\}$ is bounded uniformly.*

Under Assumption 4.1, $\bar{I}_k \supset \bar{I}_{k+1}$ for sufficiently large k since I is a finite set. Thus, for sufficiently large k, we have $J_k = E \cup I^*$. d_{k1} is a solution of the subproblem (2.6)-(2.7), which implies that

$$B_k d_{k1} + \mu_k \nabla c_{J_k}(x_k)\lambda_{k1} + \beta_{k1}\eta_{k1} = 0, \tag{4.1}$$

$$\lambda_{k1} \in \partial\|y\||_{y=(c_{J_k}+\nabla c_{J_k}^T d_{k1})_-}, \tag{4.2}$$

$$\beta_{k1} \geq 0, \eta_{k1} \in \partial\|d\||_{d=d_{k1}}, \tag{4.3}$$

$$\beta_{k1}(\|d_{k1}\|_2 - \delta\Delta_k) = 0. \tag{4.4}$$

Lemma 4.2. *Under Assumption 4.1, let $\bar{K} = \{k : r_k > 0\}$, then*

$$\liminf_{k\to\infty, k\in\bar{K}} \beta_{k1}\|d_{k1}\|_2 = 0. \tag{4.5}$$

Proof. The fact that d_{k2} is a solution of the subproblem (2.9)-(2.13) implies that d_{k2} satisfies the constraints (2.10)-(2.13). Thus we have

$$g_k + B_k d_{k2} - \nabla c_k \lambda_{k2} + \beta_{k2}\eta_{k2} = 0, \tag{4.6}$$

$$(\lambda_{k2})_i \geq 0, i \in I; (\lambda_{k2})_i \nabla c_i(x_k)^T d_{k2} = 0, i \in I_k, \tag{4.7}$$

$$(\lambda_{k2})_i(c_i(x_k) + \nabla c_i(x_k)^T s_k) = 0, i \in \bar{I}_k, \tag{4.8}$$

$$\beta_{k2} \geq 0, \eta_{k2} \in \partial\|d\||_{d=s_k}, \tag{4.9}$$

$$\beta_{k2}(\|s_k\|_2 - \Delta_k) = 0. \tag{4.10}$$

Therefore,

$$d_{k2}^T B_k d_{k1} + \beta_{k1} d_{k2}^T \eta_{k1} = -\mu_k(\nabla c_{J_k}^T d_{k2})^T \lambda_{k1} \geq 0. \tag{4.11}$$

It follows from (4.6) that

$$g_k^T d_{k2} + d_{k2}^T B_k d_{k2} - \sum_{i\in\bar{I}_k} (\lambda_{k2})_i \nabla c_i(x_k)^T d_{k2} + \beta_{k2}\eta_{k2}^T d_{k2} = 0. \tag{4.12}$$

Hence,

$$\frac{1}{2} d_{k2}^T B_k d_{k2} + \beta_{k2} d_{k2}^T \eta_{k2} \geq \sum_{i\in\bar{I}_k} (\lambda_{k2})_i \nabla c_i(x_k)^T d_{k2}. \tag{4.13}$$

Since $\lim_{k\to\infty,k\in\bar{K}} s_k = \lim_{k\to\infty,k\in\bar{K}}(x_{k+1} - x_k) = 0$, by (4.8), we have $(\lambda_{k2})_i = 0$ for $i \in \bar{I}_k$ and sufficiently large $k \in \bar{K}$.

Now suppose lemma is not true. Then there must exist a set $\tilde{K} \in \bar{K}$ such that $\lim_{k\to\infty,k\in\tilde{K}} d_{k1} = d^*$ for some $d^* \neq 0$. Without loss of generality, we assume that $\lim_{k\to\infty,k\in\tilde{K}} \tau_k = \tau^*$, $\lim_{k\to\infty,k\in\tilde{K}} B_k = B^*$ and

$$\lim_{k\to\infty,k\in\tilde{K}} \beta_{k1} = \liminf_{k\to\infty,k\in\tilde{K}} \beta_{k1} = \beta_1^*,$$

$$\lim_{k\to\infty,k\in\tilde{K}}\beta_{k2}=\liminf_{k\to\infty,k\in\tilde{K}}\beta_{k2}=\beta_2^*.$$

Then $\tau^*>0$ and

$$\lim_{k\to\infty,k\in\tilde{K}} d_{k2}=\lim_{k\to\infty,k\in\tilde{K}}(s_k-\tau_k d_{k1})=-\tau^* d^*. \tag{4.14}$$

Thus, by (4.11),

$$-\tau^* d^{*T}B^*d^*-\beta_1^*\|d^*\|_2\geq 0, \tag{4.15}$$

that is, $d^{*T}B^*d^*\leq 0$. By (4.13), $d^{*T}B^*d^*\geq 0$. Thus, $d^{*T}B^*d^*=0$. Using (4.15) we have $\beta_1^*\|d^*\|_2=0$, which completes the proof.

The analyses below show that similar results to Yuan(1995) also hold for our algorithm. Apart from Assumption 4.1, we also need to assume that $d_{k1}\to 0(k\to\infty)$. Thus, for all sufficiently large k, $\tau_k=1$.

Lemma 4.3. *Under above assumptions, there exists a large integer k_0 such that for $k\geq k_0$, if $(c_{J_k}+\nabla c_{J_k}^T d_{k1})_-\neq 0$, then*

$$\|d_{k1}\|_2=\delta\Delta_k, \tag{4.16}$$

$$\psi_k(0)-\psi_k(d_{k1})\geq\hat{\delta}\Delta_k, \tag{4.17}$$

$$\phi_k(0)-\phi_k(s_k)\geq\frac{1}{4}\hat{\delta}\Delta_k, \tag{4.18}$$

$$r_k\to 1, \tag{4.19}$$

$$x_{k+1}=x_k+s_k, \tag{4.20}$$

where $\hat{\delta}>0$ is a constant.

Proof. Suppose that $(d_{k1},\lambda_{k1},\beta_{k1})$ satisfies the first-order Kuhn-Tucker conditions of (2.6)-(2.7). Then we have (4.1), and $\|\eta_k\|_2=1$, $(\lambda_{k1})_i\leq 0(i\in I_k)$. If $(c_{J_k}+\nabla c_{J_k}^T d_{k1})_-\neq 0$, we have $\|\lambda_{k1}\|_0=1$, where $\|\cdot\|_0$ is the dual norm of the norm $\|\cdot\|$.

Let $\hat{K}=\{k:\ (c_{J_k}+\nabla c_{J_k}^T d_{k1})_-\neq 0\}$, similar to Lemma 4.2 of Liu(1998), we can prove that

$$\beta_{k1}=\|B_k d_{k1}+\mu_k\nabla c_{J_k}\lambda_{k1}\|_2\geq\beta_0>0 \tag{4.21}$$

for $k\geq k_0$ and $k\in\hat{K}$, where $\beta_0>0$ is a constant and k_0 is a large positive integer. Hence, $\|d_{k1}\|_2=\delta\Delta_k$ for $k\in\hat{K}$ and $k\geq k_0$. Therefore,

$$\lim_{k\to\infty,k\in\hat{K}}\Delta_k=\frac{1}{\delta}\lim_{k\to\infty,k\in\hat{K}}\|d_{k1}\|_2=0. \tag{4.22}$$

By using (4.1),

$$d_{k1}^T B_k d_{k1}+\mu_k(\nabla c_{J_k}^T d_{k1})^T\lambda_{k1}+\beta_{k1}\|d_{k1}\|_2=0, \tag{4.23}$$

and since

$$(\nabla c_{J_k}^T d_{k1})^T\lambda_{k1}=(c_{J_k}+\nabla c_{J_k}^T d_{k1})^T\lambda_{k1}-c_{J_k}^T\lambda_{k1} \tag{4.24}$$

$$\geq\|(c_{J_k}+\nabla c_{J_k}^T d_{k1})_-\|-\|(c_{J_k})_-\|, \tag{4.25}$$

we have

$$\psi_k(0)-\psi_k(d_{k1})\geq\beta_{k1}\|d_{k1}\|_2+\frac{1}{2}d_{k1}^T B_k d_{k1}\geq\hat{\delta}\Delta_k. \tag{4.26}$$

Thus,

$$\phi_k(0)-\phi_k(s_k)\geq\frac{1}{4}[\psi_k(0)-\psi_k(d_{k1})]\geq\frac{1}{4}\hat{\delta}\Delta_k. \tag{4.27}$$

It follows from $r_k=1+\frac{o(\|s_k\|_2)}{Pred_k}$ and (4.27) that $r_k\to 1$ for sufficiently large $k\in\hat{K}$. By Algorithm 2.1, $x_{k+1}=x_k+s_k$.

Lemma 4.4. *If the conditions of Lemma 4.3 hold, then for sufficiently large $k \in K$,*

$$c_i(x_k) + \nabla c_i(x_k)^T d_{k1} = 0, \quad i \in E, \tag{4.28}$$

$$c_i(x_k) + \nabla c_i(x_k)^T d_{k1} \geq 0, \quad i \in I_k. \tag{4.29}$$

Proof. Let $K^* = \{k : \ (c_{J_k} + \nabla c_{J_k}^T d_{k1})_- = 0\}$, then K^* must be an infinite set; Otherwise, for all large k, $k \in \hat{K}$ and Lemma 4.3 holds, so $\lim_{k\to\infty} \Delta_k = 0$ by (4.18), which contradicts (4.19).

We assume the lemma is not true. Then $\hat{K}$ is also an infinite set. Suppose that $\{k_i : \ i = 1, 2, \cdots\} \subseteq K^*$, $k_i + 1 \in \hat{K}$. If $x_{k_i+1} = x_{k_i}$, then by (4.17), for sufficiently large i,

$$\begin{aligned}\psi_{k_i}(0) - \psi_{k_i}(d_{k_i1}) &\geq \psi_{k_i+1}(0) - \psi_{k_i+1}(d_{(k_i+1)1}) \\ &\geq \hat{\delta}\Delta_{k_i+1} \geq 0.25\hat{\delta}\|s_{k_i}\|_2,\end{aligned} \tag{4.30}$$

so

$$\phi_{k_i}(0) - \phi_{k_i}(s_{k_i}) \geq 0.0625\hat{\delta}\|s_{k_i}\|_2. \tag{4.31}$$

Moreover, our assumptions and (4.30) imply that $\lim_{i\to\infty} \|s_{k_i}\|_2 = 0$. Thus, $r_{k_i} \to 1$, which contradicts Algorithm 2.1. Therefore, $x_{k_i+1} = x_{k_i} + s_{k_i}$ for all large i, which implies

$$\begin{aligned}(c_{J_{k_i}+1})_- &= (c_{J_{k_i}} + \nabla c_{J_{k_i}}^T d_{k_i1} + O(\|d_{k_i1}\|_2^2))_- \\ &= O(\|d_{k_i1}\|_2^2).\end{aligned} \tag{4.32}$$

Hence, for sufficiently large i,

$$\begin{aligned}\psi_{k_i+1}(0) - \psi_{k_i+1}(d_{(k_i+1)1}) &= \mu\|(c_{J_{k_i}+1})_-\| - \mu\|(c_{J_{k_i}+1} + \nabla c_{J_{k_i}+1}^T d_{(k_i+1)1})_-\| \\ &\quad -\frac{1}{2} d_{(k_i+1)1}^T B_{k_i+1} d_{(k_i+1)1} \\ &\leq \mu\|(c_{J_{k_i}+1})_-\| + O(\|d_{(k_i+1)1}\|_2^2) = O(\Delta_{k_i+1}^2).\end{aligned} \tag{4.33}$$

It follows from Lemma 4.4 that $O(\Delta_{k_i+1}^2) \geq \hat{\delta}\Delta_{k_i+1}$, that is, $O(\Delta_{k_i+1}) \geq \hat{\delta}$, which contradicts (4.22). The contradiction implies that $\hat{K}$ is just a finite subset.

By Lemma 4.4, under Assumption 4.1, if $d_{k1} \to 0 (k \to \infty)$, then for sufficiently large k, any solution of the nonsmooth trust region subproblem (2.6)-(2.7) is also a solution of the trust region subproblem

$$\min \quad \frac{1}{2} d^T B_k d \tag{4.34}$$

$$s.t. \quad c_i(x_k) + \nabla c_i(x_k)^T d = 0, i \in E \tag{4.35}$$

$$c_i(x_k) + \nabla c_i(x_k)^T d \geq 0, i \in I_k \tag{4.36}$$

$$\|d\|_2 \leq \delta\Delta_k. \tag{4.37}$$

The merit function $P(x, \mu)$ is nondifferentiable. In order to obtain local superlinear convergence of the algorithm, we generate a second-order correction step by solving the following problem

$$\min \quad \frac{1}{2} d^T B_k d + \mu_k \|(c_{J_k}(x_k + s_k) + \nabla c_{J_k}(x_k)^T d)_-\| \tag{4.38}$$

$$s.t. \quad \|d\|_2 \leq \delta\Delta_k. \tag{4.39}$$

Suppose that x^* is a Kuhn-Tucker point of (1.1)-(1.3). If $\|d_{k1}\|_2 < \delta\Delta_k$ and $\|s_k\|_2 < \Delta_k$ for sufficiently large k, by the discussion of Liu and Yuan(1998), Yuan(1993) and Yuan and Sun(1997), under suitable local conditions, the Algorithm 2.1 with the second-order correction technique will converge to its solution superlinearly.

5. Numerical Results

A FORTRAN subroutine is programmed to test Algorithm 2.1. All test problems are taken from Hock and Schittkowski(1981) and the standard initial points are used.

Table 1.

Problem	n	m	Δ_0	NI-NF-NG	RT	RC
HS6	2	1	1.0	9-11-10	1.20E-10	7.59E-09
HS14	2	2	5.0	4-6-5	8.89E-08	2.09E-07
HS22	2	2	5.0	8-13-9	4.38E-08	3.65E-05
HS28	3	1	5.0	8-10-9	9.70E-08	2.22E-16
HS34	3	8	5.0	7-9-8	4.07E-08	1.06E-05
HS38	4	8	1.0	72-88-73	1.66E-05	0.0
HS43	4	3	5.0	14-19-15	1.26E-06	0.0
HS49	5	2	1.0	25-26-26	3.54E-06	4.44E-15
HS50	5	3	1.0	13-14-14	2.77E-07	4.87E-15
HS52	5	3	5.0	12-14-13	1.98E-06	3.17E-15
HS63	3	5	10.0	7-10-8	7.53E-07	7.87E-11
HS76	4	7	1.0	6-7-7	7.29E-08	0.0
HS77	5	2	1.0	11-13-12	8.84E-08	2.30E-12
HS80	5	13	5.0	9-13-10	3.74E-09	2.65E-12
HS83	5	16	1.0	9-12-10	1.12E-06	0.0
HS86	5	15	10.0	5-8-6	1.14E-05	2.65E-06
HS93	6	8	5.0	22-29-23	6.61E-06	4.78E-08
HS100	7	4	5.0	16-26-17	8.99E-06	0.0
HS108	9	14	10.0	12-17-13	1.46E-07	3.65E-06
HS113	10	8	5.0	13-18-14	7.13E-06	1.39E-05

The numerical results derived by running our trust region algorithm are summarized in Table 1, where Δ_0 is the initial trust region radius, n is the number of variables, m is the number of constraints, NI, NF and NG represent the numbers of iterations, function and gradient calculations respectively, RT and RC are the ℓ_2 norm of the gradients of the Lagrangian and the constraints respectively.

The choice of the initial radius of the trust region can affect the efficiency of the algorithm (see Sartenaer(1997)). We tested our algorithm with 1, 5 and 10 three choices of Δ_0 respectively and the best results for the choice are presented in the table. The other initial parameters are $\mu_0 = 1.0$, $\delta = 0.8$ and $\epsilon = 10^{-6}$. The initial Hessian approximation B_0 is taken as the identity matrix, and it is updated in each iteration similar to Powell's procedure (see Powell(1982)) and the details can be found in Liu and Yuan(1998).

For simplicity of calculation, the norm in penalty function and the constraints on trust region are selected to be L_∞ norm, and the piecewise trust region subproblems (2.6)-(2.7) are rewritten into quadratic programming subproblems. It would be better to have an efficient algorithm to solve the piecewise subproblems directly.

Hock and Schittkowski(1981) provided numerical results for six numerical methods for solving nonlinear programming and showed that VF02AD, which is based on Han-Powell SQP method, is superior to other methods. The preliminary numerical results in Table 1 show that our algorithm is comparable to VF02AD. Further computations are needed to study the new algorithm for large scale problems, such as comparing with the famous LANCELOT program on the CUTE problems.

References

[1] R. Byrd, R.B. Schnabel, G.A. Shultz, A trust region algorithm for nonlinearly constrained optimization, *SIAM J. Numer. Anal.,* **24** (1987), 1152-1170.

[2] J.V. Burke, A sequential quadratic programming method for potentially infeasible mathematical programs, *J. Math. Anal. Appl.* **139** (1989), 319-351.

[3] J.V. Burke, A robust trust region method for constrained nonlinear programming problems, *SIAM J. Optim.,* **2** (1992), 325-347.

[4] J.V. Burke, S.P. Han, A robust sequential quadratic programming method, *Math. Prog.,* **43** (1989), 277-303.

[5] J.W. Daniel, Stability of the solutions of definite quadratic programs, *Math. Prog.,* **5** (1973), 41-53.

[6] J.E. Dennis, M. El-Alem, M.C. Maciel, A global convergence theory for general trust-region-based algorithms for equality constrained optimization, *SIAM J. Optim.,* **7** (1997), 177-207.

[7] J.E. Dennis, L.N. Vicente, On the convergence theory of trust-region-based algorithms for equality-constrained optimization, *SIAM J. Optim.,* **7** (1997), 927-950.

[8] R. Fletcher, Practical Methods of Optimization, Vol. 1, Unconstrained Optimization, John Wiley and Sons, Chichester, 1980.

[9] R. Fletcher, Practical Methods of Optimization, Vol. 2, Constrained Optimization, John Wiley and Sons, Chichester, 1981.

[10] S.P. Han, A globally convergent method for nonlinear programming, *JOTA,* **22** (1977), 297-309.

[11] W. Hock, K. Schittkowski, Test Examples for Nonlinear Programming Codes, Lecture Notes in Eco. and Math. Systems 187, 1981.

[12] X. Liu, A Globally Convergent, Locally Superlinearly Convergent Algorithm for Equality Constrained Optimization, in: Numerical Linear Algebra and Optimization, Y. Yuan eds., Science Press, Beijing, New York, 1999, 131-144.

[13] X. Liu, Y. Yuan, A robust algorithm for optimization with general equality and inequality constraints, *SIAM J. Sci. Comput.,* **22** (2000), 517-534.

[14] J.J. Moré, Recent Developments in Algorithms and Software for Trust Region Methods, in: Mathematical Programming, The State of the Art, A. Bachem, M. Grötschel and B. Korte, eds., Springer-Verlag, NY, 1983, 258-287.

[15] E.O. Omojokun, Trust region algorithms for optimization with nonlinear equality and inequality constraints, Ph.D. Thesis, Univ. Colorado, 1989.

[16] J. Peng, Y. Yuan, Optimality conditions for the minimization of a quadratic with two quadratic constraints, *SIAM J. Optim.,* **7** (1997), 574-594.

[17] M.J.D. Powell, Convergence Properties of a Class of Minimization Algorithms, in O.L. Mangasarian, R.R. Meyer, S.M. Robinson, eds., Nonlinear Programming 2, Academic Press, New York, 1975, 1-27.

[18] M.J.D. Powell, A Fast Algorithm for Nonlinearly Constrained Optimization Calculations, in proc. 1977 Dundee Biennial Conference on Numerical Analysis, G.A.Watson, ed., Springer-Verlag, Berlin, 1978, 144-157.

[19] M.J.D. Powell, The Convergence of Variable Metric Methods for Nonlinear Constrained Optimization Calculations, in Nonlinear Programming 3, O.L. Mangasarian, R.R.Meyer and S.M.Robinson, eds., Academic Press, NewYork, 1978, 27-63.

[20] M.J.D. Powell, Y. Yuan, A trust region algorithm for equality constrained optimization, *Math. Prog.,* **49** (1991), 189-211.

[21] A. Sartenaer, Automatic determination of an initial trust region in nonlinear programming, *SIAM J. Sci. Comput.,* **18** (1997), 1788-1803.

[22] G.A. Shultz, R.B. Schnabel, R.H. Byrd, A family of trust-region-based algorithms for unconstrained minimization with strong global convergence, *SIAM J. Numer. Anal.,* **22** (1985), 47-67.

[23] D.C. Sorensen, Trust Region Methods for Unconstrained Optimization, in: M.J.D. Powell, ed., Nonlinear Optimization 1981, Academic Press, London, 1982, 29-38.

[24] R. Stern, H. Wolkowicz, Trust regions and nonsymmetric eigenvalue perturbations, *SIAM J. Matrix Anal. and Appl.,* **15** (1994), 755-778.

[25] A. Vardi, A trust region algorithm for equality constrained minimization: convergence properties and implementation, *SIAM J. Numer. Anal.,* **22** (1985), 575-591.

[26] Y. Yuan, Conditions for convergence of trust region algorithms for nonsmooth optimization, *Math. Prog.,* **31** (1985), 220-228.

[27] Y. Yuan, Numerical Methods for Nonlinear Programming (in Chinese), Modern Mathematics Series, Shanghai Scientific & Technical Publishers, 1993.

[28] Y. Yuan, On the convergence of a new trust region algorithm, *Numer. Math.,* **70** (1995), 515-539.

[29] Y. Yuan, W. Sun, Optimization Theory and Methods (in Chinese), Scientific Press, Beijing, 1997.