Research Article

Global Convergence of a New Nonmonotone Filter Method for Equality Constrained Optimization

Ke Su, Wei Liu, and Xiaoli Lu

College of Mathematics and Computer Science, Hebei University, Baoding 071002, China

Correspondence should be addressed to Ke Su; pigeonsk@163.com

A new nonmonotone filter trust region method is introduced for solving optimization problems with equality constraints. This method directly uses the dominated area of the filter as an acceptability criterion for trial points and allows the dominated area decreasing nonmonotonically. Compared with the filter-type method, our method has more flexible criteria and can avoid Maratos effect in a certain degree. Under reasonable assumptions, we prove that the given algorithm is globally convergent to a first order stationary point for all possible choices of the starting point. Numerical tests are presented to show the effectiveness of the proposed algorithm.

1. Introduction

We analyze an algorithm for solving optimization problems where a smooth objective function is to be minimized subject to smooth nonlinear equality constraints. More formally, we consider the problem,

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & c_i(x) = 0, \quad i \in I = \{1, 2, \ldots, m\}, \end{aligned} \quad (P)$$

where $x \in R^n$, the functions $f : R^n \rightarrow R$ and $c_i (i \in I) : R^n \rightarrow R$ are all twice continuously differentiable. For convenience, let $g(x) = \nabla f(x)$, $c(x) = (c_1(x), c_2(x), \ldots, c_m(x))^T$ and $A(x) = (\nabla c_1(x), \nabla c_2(x), \ldots, \nabla c_m(x))$, and f_k refers to $f(x_k)$, c_k to $c(x_k)$, g_k to $g(x_k)$ and A_k to $A(x_k)$, and so forth.

There are many trust region methods for equality constrained nonlinear programming (P), for example, Byrd et al. [1], Dennis Jr. et al. [2] and Powell and Yuan [3], but in these works, a penalty or augmented Lagrange function is always used to test the acceptability of the iterates. However, there are several difficulties associated with the use of penalty function, and in particular the choice of the penalty parameter. Hence, in 2002, Fletcher and Leyffer [4] proposed a class of filter method, which does not require any penalty parameter and has promising numerical results. Consequently, filter technique has been employed to many approaches, for instance, SLP methods [5], SQP methods [6–8], interior point approaches [9], bundle techniques [10], and so on.

Filter technique, in fact, exhibits a certain degree of nonmonotonicity. The nonmonotone technique was proposed by Grippo et al. in 1986 [11] and combined with many other methods. M. Ulbrich and S. Ulbrich [12] proposed a class of penalty-function-free nonmonotone trust region methods for nonlinear equality constrained optimization without filter technique. Su and Pu [13] introduced a nonmonotone trust region method which used the nonmonotone technique in the traditional filter criteria. Su and Yu [14] presented a nonmonotone method without penalty function or filter. Gould and Toint [15] directly used the dominated area of the filter as an acceptability criteria for trial points and obtained the global convergence properties. We refer the reader [16–18] for some works about this issue.

Motivated by the ideas and methods above, we propose a modified nonmonotone filter trust region method for solving problem (P). Similar to the Byrd-Omojokun class of algorithms, each step is decomposed into the sum of two distinct components, a quasi-normal step and a tangential step. The main contribution of our paper is to employ the nonmonotone idea to the dominated area of the filter so that the new and more flexible criteria is given, which is different from that of Gould and Toint [15] and Su and Pu [13].

Under usual assumptions, we prove that the given algorithm is globally convergent to first order stationary points.

This paper is organized as follows. In Section 2, we introduce the fraction of Cauchy decrease and the composite SQP step. The new nonmonotone filter technique is given in Section 3. In Section 4, we propose the new nonmonotone filter method and present the global convergence properties in Section 5. Some numerical results are reported in the last section.

2. The Fraction of Cauchy Decrease and the Composite SQP Step

Consider the following unconstraint minimization optimization problem:

$$\min_{x \in R^n} f(x), \tag{1}$$

where $f : R^n \to R$ is a continuously differentiable function. A trust region algorithm for solving the above problem is an iterate procedure that computes a trial step as an approximate solution to the following subproblems:

$$\min \quad q(d) = g^T d + \frac{1}{2} d^T H d,$$
$$\text{s.t.} \quad \|d\| \le \Delta, \tag{2}$$

where H is the Hessian matrix $\nabla^2 f(x)$ or an approximate to it and $\Delta > 0$ is a given trust region radius.

To assure the global convergence, the step is required only to satisfy a fraction of Cauchy decrease condition. This means that d must predict via the quadratic model function $q(d)$ at least as much as a fraction of the decreased given by the Cauchy step on $q(d)$; that is, there exists a constant $\sigma > 0$ fixed across all iterations, such that

$$q(0) - q(d) \ge \sigma \left(q(0) - q\left(d^{cp} \right) \right), \tag{3}$$

where d^{cp} is the steepest descent step for $q(d)$ inside the trust region.

Lemma 1. *If the trial step d satisfies a fraction of Cauchy decrease condition, then*

$$q(0) - q(d) \ge \frac{\sigma}{2} \|\nabla f(x)\| \min \left\{ \Delta, \frac{\|\nabla f(x)\|}{\|H\|} \right\}. \tag{4}$$

Proof (see Powell [19] for the proof). Now, we turn to explain the composite SQP step. Given an approximate estimate of the solution x_k at kth iteration, following Dennis Jr. et al. [2] and M. Ulbrich and S. Ulbrich [12], we obtain the trial step $d_k = d_k^n + d_k^t$ by computing a quasi-normal step d_k^n and a tangential step d_k^t. The purpose of the quasi-normal step d_k^n is to improve feasibility. To improve optimality, we seek d_k^t in the tangential space of the linearized constraints in such a way that it provides sufficient decrease for a quadratic model of the objective function $f(x)$. Let $q_k(d) = g_k^T d + (1/2) d^T H_k d$, where H_k is a symmetric approximation of $\nabla^2 f(x)$.

d_k^n is the solution to the subproblem

$$\min \quad \frac{1}{2} \left\| c_k + A_k^T d^n \right\|^2 + \frac{\xi}{2} \|d^n\|^2,$$
$$\text{s.t.} \quad \|d^n\| \le \Delta_k, \tag{5}$$

where Δ_k is a trust region radius and $A_k = \nabla c(x_k) \in R^{n \times m}$, $\xi > 0$. In order to improve the value of the objective function, we solve the following subproblem to get d_k^t:

$$\min \quad q_k \left(d_k^n + d^t \right),$$
$$\text{s.t.} \quad A_k^T d^t = 0 \tag{6}$$
$$\|d^t\| \le \Delta_k.$$

Then we get the current trial step $d_k = d_k^n + d_k^t$. Let $d_k^t = W_k \overline{d}_k^t$, where $\overline{d}_k^t \in R^{n-m}$ and $W_k \in R^{n \times (n-m)}$ denote a matrix whose columns form a basis of the null space of A_k^T. We refer to [2] for a more detailed discussion of this issue.

In usual way that impose a trust region in step-decomposition methods, the quasi-normal step d_k^n and the tangential step d_k^t are required to satisfy

$$\|d^n\| \le \kappa \Delta_k, \qquad \|d_k^n + d^t\| \le \Delta_k, \tag{7}$$

where $0 < \kappa < 1$. Here, to simplify the proof, we only impose a trust region on $\|d^n\| \le \Delta_k$ and $\|d^t\| \le \Delta_k$, which is natural.

Note that $W_k^T \nabla q_k(d^t)$ is the reduced gradient of q_k in terms of the representation $d^t = W_k s$ of the tangential step:

$$\nabla_s \left(q_k \left(W_k s \right) \right) = W_k^T \nabla q_k \left(W_k s \right) = W_k^T \nabla q_k \left(d^t \right). \tag{8}$$

Define

$$\widehat{g}(x) = W(x)^T g(x). \tag{9}$$

Then the first order necessary optimality conditions (Karush-Kuhn-Tucker or KKT conditions) at a local solution $\overline{x} \in R^n$ of problem (P) can be written as

$$c(\overline{x}) = 0, \qquad \widehat{g}(\overline{x}) = 0. \tag{10}$$
□

3. A New Nonmonotone Filter Technique

In filter method, originally proposed by Fletcher and Leyffer [4], the acceptability of iterates is determined by comparing the value of constraint violation and the objective function with previous iterates collected in a filter. Define the violation function $h(x)$ by $h(x) = \|c(x)\|_2^2$, it is easy to see that $h(x) = 0$ if and only if x is a feasible point, so a trial point should reduce either the value of constraint violation or that of the objective function f.

In the process of the algorithm, we need to decide whether the trial point x_k^+ is any better than x_k as an approximate solution to the problem (P). If we decide that this is the case, we say that the iteration k is successful and choose x_k^+ as
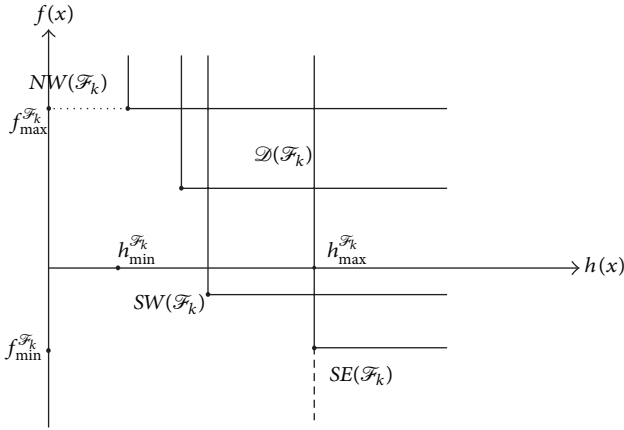
FIGURE 1

the next iterate. Let us denote by $\mathcal{S}$ the set of all successful iterations, that is,

$$\mathcal{S} = \left\{ k \,\middle|\, x_{k+1} = x_k^+ \right\}. \tag{11}$$

In traditional filter method, a point x is called acceptable to the filter if and only if

$$h(x) \leq \beta h_j \quad \text{or} \quad f(x) \leq f_j - \gamma h_j, \quad \forall \left(h_j, f_j \right) \in \mathcal{F}, \tag{12}$$

where $0 < \gamma < \beta < 1$, $\mathcal{F}$ denotes the filter set. Define

$$\mathcal{D}(\mathcal{F}) = \left\{ (h, f) \,\middle|\, h > h_j \text{ and } f > f_j, \ \exists j \in \mathcal{F} \right\}. \tag{13}$$

A trial point x_k^+ is accepted if and only if $(h_k^+, f_k^+) \notin \mathcal{D}(\mathcal{F}_k)$.

Now, similar to the idea of Gould and Toint [15], we give a new modified nonmonotone filter technique. For any (h, f)-pair, define an area that represents its contribution to the area of $\mathcal{D}(\mathcal{F})$, we hope this contribution is positive; that is, the area of $\mathcal{D}(\mathcal{F})$ is increasing. For convenience, we partition the right half-plane $[0, +\infty] \times [-\infty, +\infty]$ into four different regions (see Figure 1). Define $\mathcal{D}(\mathcal{F}_k)^c$ to be the complement of $\mathcal{D}(\mathcal{F}_k)$. Let

$$h_{\min}^{\mathcal{F}_k} \stackrel{\text{def}}{=} \min_{j \in \mathcal{F}_k} h_j, \qquad h_{\max}^{\mathcal{F}_k} \stackrel{\text{def}}{=} \max_{j \in \mathcal{F}_k} h_j,$$
$$f_{\min}^{\mathcal{F}_k} \stackrel{\text{def}}{=} \min_{j \in \mathcal{F}_k} f_j, \qquad f_{\max}^{\mathcal{F}_k} \stackrel{\text{def}}{=} \max_{j \in \mathcal{F}_k} f_j. \tag{14}$$

These four parts are

(1) the dominated part of the filter: $NE(\mathcal{F}_k) \stackrel{\text{def}}{=} \mathcal{D}(\mathcal{F}_k)$;

(2) the undominated part of lower left corner of the half plane:

$$SW(\mathcal{F}_k) \stackrel{\text{def}}{=} \mathcal{D}(\mathcal{F}_k)^c \cap \left[0, h_{\max}^{\mathcal{F}_k} \right] \times \left[-\infty, f_{\max}^{\mathcal{F}_k} \right], \tag{15}$$

(3) the undominated upper left corner: $NW(\mathcal{F}_k) \stackrel{\text{def}}{=} [0, h_{\min}^{\mathcal{F}_k}) \times (f_{\max}^{\mathcal{F}_k}, +\infty]$;

(4) the undominated lower right corner: $SE(\mathcal{F}_k) \stackrel{\text{def}}{=} (h_{\max}^{\mathcal{F}_k}, +\infty] \times [-\infty, f_{\min}^{\mathcal{F}_k})$.
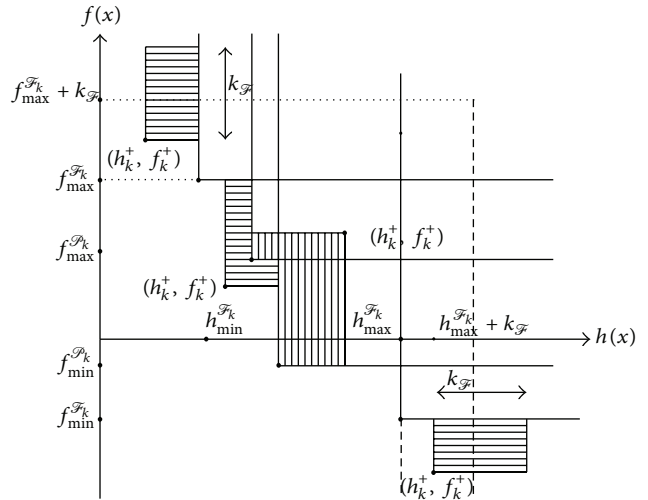

FIGURE 2

Consider the trial point x_k^+, if the filter is empty, then define its contribution to the area of the filter by

$$\alpha\left(x_k^+, \mathcal{F}_k \right) \stackrel{\text{def}}{=} \kappa_F^{\ 2}, \tag{16}$$

where $\kappa_F > 0$ is a constant. If the filter is not empty, then define the contribution of x_k^+ to the area of the filter by four different formulae.

If $(h_k^+, f_k^+) \in SW(\mathcal{F}_k)$, assume

$$\alpha\left(x_k^+, \mathcal{F}_k \right) \stackrel{\text{def}}{=} \text{area} \left(\mathcal{D}(\mathcal{F}_k)^c \cap \left[h_k^+, h_{\max}^{\mathcal{F}_k} + \kappa_F \right] \right.$$
$$\left. \times \left[f_k^+, f_{\max}^{\mathcal{F}_k} + \kappa_F \right] \right). \tag{17}$$

If $(h_k^+, f_k^+) \in NW(\mathcal{F}_k)$, assume

$$\alpha\left(x_k^+, \mathcal{F}_k \right) \stackrel{\text{def}}{=} \kappa_F \left(h_{\min}^{\mathcal{F}_k} - h_k^+ \right). \tag{18}$$

If $(h_k^+, f_k^+) \in SE(\mathcal{F}_k)$, assume

$$\alpha\left(x_k^+, \mathcal{F}_k \right) \stackrel{\text{def}}{=} \kappa_F \left(f_{\min}^{\mathcal{F}_k} - f_k^+ \right). \tag{19}$$

If $(h_k^+, f_k^+) \in NE(\mathcal{F}_k) = \mathcal{D}(\mathcal{F}_k)$, assume

$$\alpha\left(x_k^+, \mathcal{F}_k \right) \stackrel{\text{def}}{=} -\text{area} \left(\mathcal{D}(\mathcal{F}_k) \cap \left[h_k^+ - h_{\min}^{\mathcal{P}_k} \right] \right.$$
$$\left. \times \left[f_k^+ - f_{\min}^{\mathcal{P}_k} \right] \right), \tag{20}$$

where $\mathcal{P}_k = \{(h, f) \in \mathcal{F}_k | h_j < h_k^+, f_j < f_k^+\}$, and $h_{\min}^{\mathcal{P}_k} \stackrel{\text{def}}{=} \min_{j \in \mathcal{P}_k} h_j$, $f_{\min}^{\mathcal{P}_k} \stackrel{\text{def}}{=} \min_{j \in \mathcal{P}_k} f_j$.

Figure 2 illustrate the corresponding areas in the filter. Horizontally dashed surfaces indicate a positive contribution and vertically dashed ones a negative contribution. Note that $\alpha(x, \mathcal{F})$ is a continuous function of $(h(x), f(x))$.

Next, we should consider the updating of the filter. If $(h_k, f_k) \notin \mathcal{D}(\mathcal{F}_k)$, then

$$\mathcal{F}_{k+1} \longleftarrow \mathcal{F}_k \cup (h_k, f_k). \qquad (21)$$

If $(h_k, f_k) \in \mathcal{D}(\mathcal{F}_k)$, then

$$\mathcal{F}_{k+1} \longleftarrow (\mathcal{F}_k \setminus \mathcal{P}_k) \cup (h_{\min}^{\mathcal{P}_k}, f_k) \cup (h_k, f_{\min}^{\mathcal{P}_k}). \qquad (22)$$

We now return to the question of deciding whether a trial point x_k^+ is acceptable for the filter or not. We will insist that this is a necessary condition for the iteration k to be successful in the sense that $x_{k+1} = x_k^+$. If we consider an iterate x_k, there must exist a predecessor iteration such that $x_{p(k)}^+ = x_{p(k)+1} = x_k$. Under the monotonic situation, a trial point x_k^+ would be accepted whenever it results in an sufficient increase in the dominated area of the filter, that means x_k^+ would be accepted whenever

$$\alpha_k \geq \gamma_{\mathcal{F}} (h_k^+)^2, \qquad (23)$$

where $\alpha_k \stackrel{\text{def}}{=} \alpha(x_k^+, \mathcal{F}_k)$, $\gamma_{\mathcal{F}} \in (0, 1)$ is a constant. Under the nonmonotonic situation, we relax condition (23) to be

$$\max \left\{ \alpha_k, \sum_{\substack{j=r(k)+1 \\ j \in \mathcal{U}}}^{k} \lambda_{p(j)} \alpha_{p(j)} \right\} \geq \gamma_{\mathcal{F}} (h_k^+)^2, \qquad (24)$$

where $\alpha_{p(j)} \stackrel{\text{def}}{=} \alpha(x_j, \mathcal{F}_{p(j)})$, $\mathcal{U} = \{k|$ filter is updated for $(h_k, f_k)\}$, $r(k) \leq k$ is some reference iteration for $\mathcal{U}$, $r(k) \in \mathcal{U}$, $\mathcal{U} \subseteq \mathcal{S}$, $\lambda_{p(j)} \in [0, 1]$, $\sum_{j=r(k)+1, j \in \mathcal{U}}^{k} \lambda_{p(j)} = 1$.

Compared to condition (2.21) [15], our condition (24) is more flexible if α_k is negative.

According to condition (24), it is possible to accept x_k^+ even though it may be dominated. Then x_k^+ will be accepted if either (23) or (24) holds.

4. The New Nonmonotone Filter Trust Region Algorithm

Our algorithm is based on the usual trust region technique; define the predict reduction for the function $q_k(x)$ to be

$$\text{pred}(d_k) = q_k(0) - q_k(d_k) \qquad (25)$$

and the actual reduction

$$\text{ared}(d_k) = f(x_k) - f(x_k^+). \qquad (26)$$

Moreover, let $r_k = \text{ared}(d_k)/\text{pred}(d_k)$, if there exists a nonzero constant η_1 such that $r_k \geq \eta_1$ and condition (23) and (24) hold, the trial point x_k^+ will be called acceptable. Then the next trial point x_{k+1} is obtained, and for its feasibility, we consider the condition

$$h_{k+1} \leq \eta_3 \min \left\{ \mu, \alpha_1 \Delta_k^{2+\alpha_2} \right\} \qquad (27)$$

is true or not, where η_3 and μ are all positive constants, if it is not true, then turn to the feasibility restoration phase and define

$$r_k^j = \frac{\left\| c(x_k^j) \right\|^2 - \left\| c(x_k^j + d_k^j) \right\|^2}{\left\| c(x_k^j) \right\|^2 - \left\| c(x_k^j) + A(x_k^j)^T d_k^j \right\|^2}. \qquad (28)$$

A formal description of the algorithm is given as follows.

Algorithm A Step 0. Choose an initial point $x_0 \in R^n$, a symmetric matrix $H_0 \in R^{n \times n}$, let $\Delta_0 > 0$, $\epsilon_t > 0$, $\eta_2, \eta_3 \in (0, 1)$, $\alpha_1, \alpha_2 \in [0, 1]$, $\eta_1 > 0$, $0 < \gamma_0 < \gamma_1 < 1 \leq \gamma_2 < \gamma_3 \leq 2$, $0 < \xi < 1$, $\gamma_{\mathcal{F}} \in (0, 1)$, $\mu > 0$, compute $f(x_0)$, $h(x_0)$, let $k = 0$, $\mathcal{F}_0 = \emptyset$.

Step 1. Compute $c_k, A_k, W_k, f_k, h_k, g_k, \hat{g}_k = W_k^T g_k$.

Step 2. If $\| \hat{g}_k \| + h_k \leq \epsilon_t$, stop.

Step 3. Solve the subproblems (5) and (6) to get the quasi-normal step d_k^n and the tangential step d_k^t. Let $d_k = d_k^n + d_k^t$, $x_k^+ = x_k + d_k$.

Step 4. If $r_k < \eta_1$, let $x_{k+1} = x_k$, then go to Step 8.

Step 5. If $x_k + d_k$ is not acceptable to the filter, go to Step 8. Otherwise $x_{k+1} = x_k^+$ and update the filter according to (21) and (22), the trust region radius and H_k, then get the corresponding h_{k+1}, f_{k+1}.

Step 6. If $h_{k+1} \leq \eta_3 \min\{\mu, \alpha_1 \Delta_k^{2+\alpha_2}\}$, $k = k + 1$ and go to Step 1, otherwise go to Step 7.

Step 7. By restoration Algorithm B to get d_k^r, then the trial point $x_k^r = x_k + d_k^r$.

Step 8. Update the trust region radius by Algorithm C, let $k = k + 1$ and go to Step 3.

We aim to reduce the value of $h(x)$ in the restoration Algorithm B, that is to get $c(x_k^r) = 0$ by Newton-type method.

Algorithm B Step 0. Let $x_k^0 = x_k$, $\Delta_k^0 = \Delta_k$, $j = 0$, $\eta_2, \eta_3 \in (0, 1)$, $\alpha_1, \alpha_2 \in [0, 1]$, $\mu > 0$.

Step 1. If $h(x_k^j) \leq \eta_3 \min\{h^{\mathcal{F}_k}, \alpha_1 \Delta_k^{2+\alpha_2}\}$ and x_k^j is acceptable by the filter, then $x_k^r = x_k^j$, stop.

Step 2. Compute

$$\begin{aligned} \min \quad & \left\| c(x_k^j) + A(x_k^j)^T d_k^j \right\|^2, \\ \text{s.t.} \quad & \left\| d_k^j \right\| \leq \Delta_k^j \end{aligned} \qquad (29)$$

to get d_k^j, then compute r_k^j.

Step 3. If $r_k^j \leq \eta_2$, $x_k^{j+1} = x_k^j$, $\Delta_k^{j+1} = \Delta_k^j/2$, $j = j + 1$ and go to Step 2.

Step 4. If $r_k^j > \eta_2$, $x_k^{j+1} = x_k^j + d_k^j$, $\Delta_k^{j+1} = 2\Delta_k^j$, $j = j + 1$, compute A_k^{j+1} and go to Step 1, where $h_k^{\mathscr{F}_k} = \min\{\min_{j \in \mathscr{F}_k} \{h_j | h_j > 0\}, \mu\}$.

Algorithm C (updating the trust region radius). Given $\eta_1 > 0$, $0 < \gamma_0 < \gamma_1 < 1 \le \gamma_2 < \gamma_3 \le 2$, we have the following.

(1) If $r_k < \eta_1$ or x_k^+ is not acceptable to the filter, $\Delta_{k+1} \in [\gamma_0 \Delta_k, \gamma_1 \Delta_k]$.

(2) If x_k^+ is acceptable to the filter but does not satisfiy condition (27), $\Delta_{k+1} \in (\Delta_k, \gamma_2 \Delta_k)$.

(3) If x_k^+ is acceptable to the filter and satisfies (27), $\Delta_{k+1} \in [\gamma_2 \Delta_k, \gamma_3 \Delta_k]$.

From the description above and the idea of the algorithm, we can see that our algorithm is more flexible. Every successful iterate must be any better than the predecessor one in some degree according to the traditional filter method. But our algorithm relaxes this demand by using the nonmonotone technique and also avoids Maratos effect in a certain degree. Moreover, Algorithm C allows a relatively wide choice of the trust region.

5. The Convergence Properties

In this section, to present a proof of global convergence of algorithm, we always assume that the following conditions hold.

Assumption

(A1) The objective function f and the constraint functions c_i ($i \in I = \{1, 2, \ldots, m\}$) are twice continuously differentiable.

(A2) For all k, x_k, and $x_k + d_k$ all remain in a closed, bounded convex subset $\Omega \subset R^n$.

(A3) The matrix $A(x) = \nabla c(x)$ is nonsingular matrix for all $x \in \Omega$.

(A4) The matrices $(A(x)^T A(x))^{-1}$, $W(x)$, $(W^T W)^{-1}$ are uniformly bounded in Ω, where $W(x)$ denotes a matrix whose columns form a basis of the null space of $A(x)^T$.

(A5) The matrix H_k is uniformly bounded.

By the assumptions, we can suppose there exist constants v_0, v_1, v_2, v_3 such that $\| f(x) \| \le v_0$, $\| g(x) \| \le v_0$, $\| c(x) \| \le v_0$, $\| A(x) \| \le v_0$, $\| A(x)^T A(x)^{-1} \| \le v_1$, $\| W(x) \| \le v_2$, $\| H_k \| \le v_3$, $\| W_k^T H_k \| \le v_3$, $\| W_k^T H_k W_k \| \le v_3$.

By (A1) and (A2), it holds

$$f_{\min} \le f_k, \quad 0 \le h_k \le h_{\max} \quad \forall k, \tag{30}$$

where $f_{\min}, h_{\max} > 0$, hence in the (h, f)-plane, the (h, f)-pair lies in the area $[0, h_{\max}] \times [f_{\min}, +\infty]$.

From (A1), (A2), and (A3), it exists a constant $\bar{v}$ such that

$$|f(x_k + d_k) - q_k(d_k)| \le \bar{v} \Delta_k^2. \tag{31}$$

Lemma 2. *At the current iterate x_k, let the trial point component d_k^n actually be normal to the tangential space. Under the problem assumptions, there exists a constant $k_1 > 0$ independent of the iterates such that*

$$\|d_k^n\| \le \alpha_1 \|c_k\|. \tag{32}$$

Proof. It is similar to the proof of Lemma 2 in [13]. □

Lemma 3. *Under Assumptions, there exist positive constants k_2, k_3, k_4 independent of the iterates such that*

$$\|c_k\|^2 - \|c_k + A_k^T d_k^n\|^2 \ge k_2 \|c_k\| \min\{k_3 \|c_k\|, \Delta_k\},$$

$$q_k(d_k^n) - q_k(d_k)$$
$$\ge \frac{\sigma}{2} \|W_k^T \nabla q_k(d_k^n)\| \min\{k_4 \|W_k^T \nabla q_k(d_k^n)\|, \Delta_k\}. \tag{33}$$

Proof. The proof is an application of Lemma 1 to the two subproblems (5) and (6). □

Lemma 4. *Suppose that Assumptions hold, then restoration Algorithm B is well defined.*

Proof. The conclusion is obvious, if $h_k^j \to 0$. Otherwise it exists $\epsilon > 0$ such that for all j, it holds $h_k^j > \epsilon$. Consider the set

$$K = \left\{ j \,\middle|\, r_k^j = \frac{\|c_k^j\|^2 - \|c(x_k^j + d_k^j)\|^2}{\|c_k^j\|^2 - \|c_k^j + (A_k^j)^T d_k^j\|^2} > \eta_2 > 0 \right\}, \tag{34}$$

where $c_k^j = c(x_k^j)$, $A_k^j = A(x_k^j)$. By Lemma 3 and the definition of h_k, we have

$$+\infty > \sum_{j=1}^{\infty} \left(h_k^{j-1} - h_k^j\right) \ge \sum_{j \in K} \left(\|c_k^j\|^2 - \|c_k^j + (A_k^j)^T d_k^j\|^2 \right)$$
$$\ge \eta_2 k_2 \sum_{j \in K} \|c_k\| \min\{k_3 \|c_k\|, \Delta_k^j\}. \tag{35}$$

By $h_k^j > \epsilon$, it holds $\Delta_k^j \to 0$ for $j \in K$. From Algorithm B, we can obtain that $\Delta_k^j \to 0$ for all j.

On the other side,

$$\|c_k^j\|^2 - \|c(x_k^j + d_k^j)\|^2 = \|c_k^j\|^2 - \|c_k^j + (A_k^j)^T d_k^j\|^2 + o(\Delta_k^j) \tag{36}$$

for $\Delta_k^j \to 0$. By the algorithm, the radius Δ_k^j should be satisfied $\Delta_k^{j+1} > \Delta_k^j$, that is contradicted to $\Delta_k^j \to 0$. The proof is complete. □

Now, we analyze the impact of the criteria (23) and (24). Once a trial point is accepted as a new iterate, it must be provided some improvement, and we formalize this by saying that iterate $x_k = x_{p(k)+1}$ improves on iterate $x_{i(k)}$. That is

the trial point x_k is accepted at iterate $p(k)$; it happens under two situations, one is by the criteria (23), that is,

$$i(k) = p(k) \quad \text{if } p(k) \notin \mathscr{A} \tag{37}$$

the other is by the criteria (24), that is,

$$i(k) = p(k) \quad \text{if } p(k) \notin \mathscr{A}. \tag{38}$$

Now consider any iterate x_k, it improved on $x_{i(k)}$, which was itself accepted because it improved on $x_{i(i(k))}$, and so on, until back to x_0. Hence we may construct a chain of successful iterations indexed by $\mathscr{C}_k = \{l_1, l_2, \ldots, l_q\}$ for each k, such that

$$x_{l_1} = x_0, \quad x_{l_q} = x_k, \quad x_{l_j} = x_{i(l_{j+1})}, \quad j = 1, 2, \ldots, q-1, \tag{39}$$

where l_1 is the smallest index in the chain of successful iterations.

Lemma 5. *Suppose that Assumptions hold and Algorithm A does not terminate finitely, apply Algorithm A to the problem (P), then for all k and $\mathscr{C}_k = \{l_1, l_2, \ldots, l_q\}$, it holds*

$$\text{area}\left(\mathscr{D}\left(\mathscr{F}_k\right)\right) \geq \gamma_{\mathscr{F}} \sum_{j=1}^{q} h_{l_j}^2. \tag{40}$$

Proof. For $\forall l_j \in \mathscr{C}_k$, if $p(l_j) \in \mathscr{A}$, by (24), $i(l_j) = r(p(l_j)) = l_{j-1}$, then

$$\max\left\{\alpha_{p(l_j)}, \sum_{\substack{i=l_{j-1}+1 \\ i \in \mathscr{U}}}^{l_j} \lambda_{p(i)} \alpha_{p(i)}\right\} \geq \gamma_{\mathscr{F}} h_{l_j}^2. \tag{41}$$

If $\max\{\alpha_{p(l_j)}, \sum_{i=l_{j-1}+1, i \in \mathscr{U}}^{l_j} \lambda_{p(i)} \alpha_{p(i)}\} = \alpha_{p(l_j)}$,

$$\alpha_{p(l_j)} \geq \gamma_{\mathscr{F}} h_{l_j}^2. \tag{42}$$

If $\max\{\alpha_{p(l_j)}, \sum_{i=l_{j-1}+1, i \in \mathscr{U}}^{l_j} \lambda_{p(i)} \alpha_{p(i)}\} = \sum_{i=l_{j-1}+1, i \in \mathscr{U}}^{l_j} \lambda_{p(i)} \alpha_{p(i)}$,

$$\sum_{\substack{i=l_{j-1}+1 \\ i \in \mathscr{U}}}^{l_j} \alpha_{p(i)} \geq \gamma_{\mathscr{F}} h_{l_j}^2. \tag{43}$$

If $p(l_j) \notin \mathscr{A}$, $l_{j-1} = p(l_j)$. By $\mathscr{U} \subseteq \mathscr{S}$, it holds

$$\{l_{j-1}+1, \ldots, l_j\} \cap \mathscr{U} \subseteq \{l_{j-1}+1, \ldots, l_j\} \cap \mathscr{S} = \{l_j\}. \tag{44}$$

Then from (23), $\alpha_{p(i)} \geq \gamma_{\mathscr{F}} h_{l_j}^2$. It implies (43). Moreover

$$\text{area}\left(\mathscr{D}\left(\mathscr{F}_k\right)\right) \geq \sum_{\substack{i=0 \\ i \in \mathscr{U}}}^{k} \alpha_{p(i)} = \sum_{j=0}^{q}\left(\sum_{\substack{i=l_{j-1}+1 \\ i \in \mathscr{U}}}^{l_j} \alpha_{p(i)}\right). \tag{45}$$

Together with (42) and (43), the result follows. □

Lemma 6. *Suppose that Assumptions hold. If Algorithm A does not terminate finitely and the filter contains infinite iterates, then $\lim_{k \to \infty} h_k = 0$.*

Proof. Suppose by contradiction that there exists a constant $\epsilon > 0$ and infinite sequence $\{k_i\} \subseteq \mathscr{S}$ such that $h_{k_i} \geq \epsilon$ for all i. Because there are infinite iterations in the filter, we have $|\mathscr{S}| = \infty$, then $h_{l_q} \geq \epsilon$ for $\forall q$.

$$\text{area}\left(\mathscr{D}\left(\mathscr{F}_k\right)\right) \geq \gamma_{\mathscr{F}} \cdot q \cdot \epsilon^2. \tag{46}$$

Then by (31), area$(\mathscr{D}(\mathscr{F}_k))$ is upper bounded for each k. That means it exists $\kappa_{\mathscr{F}}^{\max} \geq 0$ such that area$(\mathscr{D}(\mathscr{F}_k)) \leq \kappa_{\mathscr{F}}^{\max}$, so $i \leq (\kappa_{\mathscr{F}}^{\max}/\gamma_{\mathscr{F}} \epsilon^2)$. Hence i must be finite, it contradicts to the infinity of $\{k_i\}$. The proof is complete. □

Lemma 7. *Suppose that Assumptions hold and Algorithm A terminate finitely, then $h_k = 0$.*

Proof. From the Algorithm A and the definition of filter, the conclusion follows. □

Lemma 8. *For any trial point $x_{k+1} \neq x_k$, there must be one accepted by the filter.*

Lemma 9. *Suppose that Assumptions hold, there exists $k_5 > 0$ independent of the iterates such that*

$$q_k(0) - q_k\left(d_k^n\right) \geq -k_5 \|c_k\|. \tag{47}$$

Proof. By (32), the assumptions and $\|d_k^n\| \leq \Delta_{\max}$, it is obvious that

$$
\begin{aligned}
q_k(0) - q_k\left(d_k^n\right) &= -g_k^T d_k^n - \frac{1}{2}\left(d_k^n\right)^T H_k d_k^n \\
&\geq -\|d_k^n\|\left(\|g_k\| + \frac{1}{2}\|H_k\|\|d_k^n\|\right) \\
&\geq -k_1 \|c_k\|\left(\nu_0 + \nu_3 \Delta_{\max}\right) \\
&\stackrel{\text{def}}{=} -k_5 \|c_k\|.
\end{aligned}
\tag{48}
$$

The proof is complete. □

Lemma 10. *Suppose that Assumptions hold and $\|\widehat{g}_k\| \geq \epsilon_t$, if*

$$\Delta_k \leq \min\left\{\left(\frac{\epsilon_t}{2k_1\nu_3\sqrt{\eta_3\alpha_1}}\right)^{2/(2+\alpha_2)}, \frac{k_4\epsilon_t}{2}, \left(\frac{\nu_0}{k_1\nu_3\sqrt{\eta_3\alpha_1}}\right)^{2/(2+\alpha_2)}, \left(\frac{\sigma\epsilon_t}{16k_1\nu_0\sqrt{\eta_3\alpha_1}}\right)^{2/\alpha_2}\right\} \stackrel{\text{def}}{=} \delta_1 \tag{49}$$

one can deduce

$$q_k(0) - q_k(d_k) \geq \frac{\sigma\epsilon_t}{8}\Delta_k \stackrel{\text{def}}{=} \overline{\sigma}\Delta_k. \tag{50}$$

Proof. By the assumptions and the definition of d_k^n, it holds

$$
\left\| W_k^T \nabla q_k(d_k^n) \right\| = \left\| W_k^T (g_k + H_k d_k^n) \right\| \geq \|\hat{g}_k\| - \left\| W_k^T H_k d_k^n \right\|
$$

$$
\geq \epsilon_t - \nu_3 k_1 \|c_k\|
$$

$$
\geq \epsilon_t - \nu_3 k_1 \sqrt{\eta_3 \alpha_1} \Delta_k^{1+(\alpha_2/2)} \geq \frac{\epsilon_t}{2}.
\tag{51}
$$

From Lemma 3, $q_k(d_k^n) - q_k(d_k) \geq (\sigma/4)\epsilon_t \min\{\Delta_k, k_4 \epsilon_t/2\} \geq (\sigma \epsilon_t/4)\Delta_k$. Together with

$$
q_k(0) - q_k(d_k^n) = -g_k^T d_k^n - \frac{1}{2}(d_k^n)^T H_k d_k^n
$$

$$
\geq -\|g_k\| \|d_k^n\| - \frac{1}{2}\|H_k\| \|d_k^n\|^2
$$

$$
\geq -k_1 \nu_0 \sqrt{\eta_3 \alpha_1} \Delta_k^{1+(\alpha_2/2)} - k_1^2 \nu_3 \eta_3 \alpha_1 \Delta_k^{2+\alpha_2}
$$

$$
\geq -2k_1 \nu_0 \sqrt{\eta_3 \alpha_1} \Delta_k^{1+(\alpha_2/2)}
$$

$$
\geq -\frac{\sigma \epsilon_t}{8}\Delta_k
\tag{52}
$$

then $q_k(0) - q_k(d_k) \geq -(\sigma\epsilon_t/8)\Delta_k + (\sigma\epsilon_t/4)\Delta_k = (\sigma\epsilon_t/8)\Delta_k = \bar{\sigma}\Delta_k$. It is the conclusion. $\square$

Lemma 11. *Suppose the conditions of Lemma 10 hold, if*

$$
\Delta_k \leq \min\left\{\delta_1, \frac{(1-\eta_1)\bar{\sigma}}{\bar{\nu}}\right\} \overset{\text{def}}{=} \delta_2
\tag{53}
$$

then $r_k \geq \eta_1$.

Proof. From the definition of r_k and Lemma 10, together with (31), we have

$$
|r_k - 1| = \frac{|f(x_k + d_k) - q_k(d_k)|}{|q_k(0) - q_k(d_k)|} \leq \frac{\bar{\nu}\Delta_k^2}{\bar{\sigma}\Delta_k} \leq 1 - \eta_1.
\tag{54}
$$

It is obvious that $r_k \geq \eta_1$. $\square$

Lemma 12. *Suppose the conditions of Lemmas 10 and 11 hold, if*

$$
h_k \leq (\eta_3 \alpha_1)^{-1/(1+\alpha_2)}\left(\frac{\eta_1 \bar{\sigma}}{\sqrt{\gamma_{\mathscr{F}}}}\right)^{(2+\alpha_2)/(1+\alpha_2)}
\tag{55}
$$

then

$$
f(x_k^+) \leq f(x_k) - \sqrt{\gamma_{\mathscr{F}}} h_k.
\tag{56}
$$

Proof. By Lemmas 3, 10, and 11, together with $\|\hat{g}_k\| \geq \epsilon_t$, it holds

$$
f(x_k) - f(x_k^+) \geq \eta_1(q_k(0) - q_k(d_k)) \geq \eta_1 \bar{\sigma}\Delta_k.
\tag{57}
$$

From the Algorithm, $h_k \leq \eta_3 \alpha_1 \Delta_k^{2+\alpha_2}$, then

$$
f(x_k) - f(x_k^+) \geq \eta_1 \bar{\sigma}\left(\frac{h_k}{\eta_3 \alpha_1}\right)^{1/(2+\alpha_2)}.
\tag{58}
$$

Hence $f(x_k) - f(x_k^+) \geq \sqrt{\gamma_{\mathscr{F}}} h_k$. $\square$

Theorem 13. *Suppose the assumptions hold, there must exist $\Delta_{\min} > 0$ such that for each k, it holds*

$$
\Delta_k \geq \Delta_{\min}.
\tag{59}
$$

Proof. Let $\bar{k}_1$ be large enough such that $h_{\bar{k}_1} \leq \epsilon_t$, it is true by Lemmas 6 and 7. Suppose by contradiction that the index j is the first one after $\bar{k}_1$, which satisfies

$$
\Delta_j \leq \gamma_0 \min\left\{\delta_2, \left(\frac{(1-\sqrt{\gamma_{\mathscr{F}}})h^F}{\eta_3 \alpha_1}\right)^{1/(2+\alpha_2)}, \Delta_{\bar{k}_1}\right\} \overset{\text{def}}{=} \gamma_0 \delta_3,
\tag{60}
$$

where $h^F \overset{\text{def}}{=} \min_{i \in \mathscr{U}} h_i$ is the smallest value of violation function in filter. Then $\Delta_j \leq \gamma_0 \Delta_{\bar{k}_1}$. By the above analysis, we know $j \geq \bar{k}_1 + 1$, that is $j - 1 \geq \bar{k}_1$. From the Algorithm and (60), it concludes

$$
\Delta_{j-1} \leq \frac{1}{\gamma_0}\Delta_j \leq \delta_3.
\tag{61}
$$

By (60) and (61), (53) can be obtained. In Lemma 11, let $j-1$ instead of k, it deduces

$$
r_{j-1} \geq \eta_1.
\tag{62}
$$

Based on Lemma 12, together with (60), (61), and the algorithm, we can see

$$
h_{j-1}^+ \leq \eta_3 \alpha_1 \Delta_{j-1}^{2+\alpha_2} \leq (1 - \sqrt{\gamma_{\mathscr{F}}}) h^F.
\tag{63}
$$

It can be seen that (53) is true for $j - 1 \geq k$, with (55), we can deduce

$$
f_{j-1}^+ \leq f_{j-1} - \sqrt{\gamma_{\mathscr{F}}} h_{j-1}.
\tag{64}
$$

That means x_{j-1}^+ can be accepted by the filter. From above and (55), we know $\Delta_j \geq \Delta_{j-1}$. Hence the index j is not the first one after $\bar{k}_1$ which satisfied (60), that is a contradiction. So, for any $k > \bar{k}_1$, it holds $\Delta_k \geq \gamma_0 \delta_3$. Define

$$
\Delta_{\min} = \min\left\{\Delta_0, \ldots, \Delta_{\bar{k}_1}, \gamma_0 \delta_3\right\}
\tag{65}
$$

we can see that

$$
\Delta_k \geq \Delta_{\min}
\tag{66}
$$

holds for each k. The proof is complete. $\square$

Lemma 14. *Suppose that Assumptions hold and Algorithm A does not terminate finitely, then $\liminf_{k \to \infty} \|\hat{g}_k\| = 0$.*

Proof. Suppose by contradiction that for ϵ_t, there exists a constant $\bar{k} > 0$ such that $\|\hat{g}_k\| \geq \epsilon_t$.

By Assumption (A3) and (A4), $\left\| W_k^T \nabla q_k(d_k^n) \right\| \geq \|\hat{g}_k\| - \nu_3 k_1 \|c_k\|$. From Lemma 6, we know $h_k \to 0$. Hence there exists $\bar{k} > 0$ such that

$$
\|c_k\| \leq \frac{2\epsilon_t}{3\nu_3 k_1}
\tag{67}
$$

for $k > \tilde{k}$. Then $\| W_k^T \nabla q_k(d_k^n) \| \geq (1/3) \| \widehat{g}_k \| \geq (1/3)\epsilon_t$ for $k > \widehat{k} \overset{\text{def}}{=} \max\{\widehat{k}, \tilde{k}\}$.

It is obvious that

$$q_k(0) - q_k(d_k) = q_k(0) - q_k(d_k^n) + q_k(d_k^n) - q_k(d_k). \tag{68}$$

By the proof of Lemma 9, it holds $|q_k(0) - q_k(d_k^n)| \leq \nu_0 \| d_k^n \| + (1/2)\nu_3 \| d_k^n \|^2$. Together with Lemma 6 and the definition of d_k, we have

$$\lim_{k \to \infty} \left(q_k(0) - q_k(d_k^n) \right) = 0. \tag{69}$$

By the Algorithm, we can get

$$+\infty > \sum_{k=0}^{\infty} (f_k - f_{k+1}) \geq \eta_1 \sum_{k=0}^{\infty} \left(q_k(0) - q_k(d_k) \right). \tag{70}$$

Then

$$\lim_{k \to \infty} \left(q_k(0) - q_k(d_k) \right) = 0. \tag{71}$$

By (68), (69), and (71), it deduces

$$\lim_{k \to \infty} \left(q_k(d_k^n) - q_k(d_k) \right) = 0. \tag{72}$$

Based on the assumptions, Lemma 3 and Theorem 13, for $k > \widehat{k}$, it holds

$$
\begin{aligned}
& q_k(d_k^n) - q_k(d_k) \\
& \geq \frac{\sigma}{2} \left\| W_k^T \nabla q_k(d_k^n) \right\| \min \left\{ \Delta_k, k_4 \left\| W_k^T \nabla q_k(d_k^n) \right\| \right\} \\
& \geq \frac{\sigma \epsilon_t}{6} \min \left\{ \Delta_{\min}, \frac{k_4 \epsilon_t}{6} \right\} > 0,
\end{aligned}
\tag{73}
$$

which contradicts (72). The conclusion follows. $\qquad\square$

Theorem 15. *Suppose the assumptions hold, and apply the algorithm to problem* (P), *then*

$$\liminf_{k \to \infty} \left(h_k + \|\widehat{g}_k\| \right) = 0, \tag{74}$$

where $\widehat{g}_k = W_k^T g_k$, $g_k = \nabla f(x_k)$, $W(x)$ *denotes a matrix whose columns form a basis of the null space of* $A(x)^T$.

Proof. If the algorithm terminates finitely, it is obvious that it holds. Otherwise, by Lemmas 6 and 14, the conclusion also can be obtained. $\qquad\square$

Theorem 16. *Suppose the assumptions hold, and* $\{x_k\}$ *is the infinite sequence obtained by the algorithm, then there must exist a subsequence such that*

$$\lim_{j \to \infty} x_{k_j} = x^* \tag{75}$$

and x^* *satisfies the one order KKT condition of* (P).

TABLE 1

Problem	n	m	NF	NG	L's NF	L's NG
HS6	2	1	11	11	20	12
HS7	2	1	9	3	15	11
HS8	2	2	7	4	—	—
HS9	2	1	6	6	6	6
HS26	3	1	24	24	42	24
HS39	4	2	15	9	26	19
HS40	4	3	7	5	7	5
HS42	4	2	8	8	11	11
HS78	5	3	6	6	8	7

Proof. By Assumption (A1), there exist a subsequence $\{k_j\}$ and x^*, such that $\lim_{j \to \infty} x_{k_j} = x^*$. Together with Assumption (A3) and (A4), it hods $\lim_{j \to \infty} W_{k_j}^T g_{k_j} = 0$, which means for large enough j, g_{k_j} lies in the space spaned by the columns of $A_{k_j}^T$. That is there exists λ_{k_j} such that

$$\lim_{j \to \infty} g_{k_j} + A_{k_j}^T \lambda_{k_j} = 0. \tag{76}$$

The conclusion follows. $\qquad\square$

6. Some Numerical Experiments

(1) Updating of H_k is done by $H_{k+1} = H_k + (y_k^T y_k / y_k^T s_k) - (H_k s_k s_k^T H_k / s_k^T H_k s_k)$, where $y_k = \theta_k \widehat{y}_k + (1 - \theta_k) H_k s_k$

$$
\theta_k = \begin{cases} 1 & s_k^T \widehat{y}_k \geq 0.2 s_k^T H_k s_k \\ \dfrac{0.8 s_k^T H_k s_k}{s_k^T H_k s_k - s_k^T \widehat{y}_k} \end{cases}
\tag{77}
$$

and $\widehat{y}_k = g_{k+1} - g_k + (A_{k+1} - A_k)\rho_k$, $s_k = x_{k+1} - x_k \rho_k$ is the multipluser of corresponding quadratic subproblems.

(2) We assume the error toleration is 10^{-5}.

(3) The algorithm parameters were set as follows: $H_0 = I \in R^{n \times n}, \eta_1 = 0.25, \eta_2 = 0.25, \eta_3 = 0.1, \alpha_1 = \alpha_2 = 0.5, \gamma = 0.02, \rho = 0.5, \xi = 10^{-6}, \gamma_0 = 0.1, \gamma_1 = 0.5, \gamma_2 = 2, \gamma_F = 10^{-4}, \Delta_0 = 1$. The program is written in Matlab.

The numerical results for the test problems are listed in Table 1.

In Table 1, the problems are numbered in the same way as in Schittkowski [20] and Hock and Schittkowski [21]. For example, "S216" is the problem (216) in Schittkowski [20] and "HS6" is the problem (6) in Hock and Schittkowski [21]. NF, NG represent the numbers of function and gradient calculations and "L's" is the solution in [22]. The numerical results show that the our algorithm is more effective than the L's for most test examples. Moreover, the higher the level of nonmonotonic, the better the numerical results. The results show that the new algorithm is robust and effective, and more flexible for the acceptance of the the trial iterate.

Acknowledgments

References

[1] R. H. Byrd, R. B. Schnabel, and G. A. Shultz, "A trust region algorithm for nonlinearly constrained optimization," *SIAM Journal on Numerical Analysis*, vol. 24, no. 5, pp. 1152–1170, 1987.

[2] J. E. Dennis Jr., M. El-Alem, and M. C. Maciel, "A global convergence theory for general trust-region-based algorithms for equality constrained optimization," *SIAM Journal on Optimization*, vol. 7, no. 1, pp. 177–207, 1997.

[3] M. J. D. Powell and Y. Yuan, "A trust region algorithm for equality constrained optimization," *Mathematical Programming A*, vol. 49, no. 2, pp. 189–211, 1991.

[4] R. Fletcher and S. Leyffer, "Nonlinear programming without a penalty function," *Mathematical Programming A*, vol. 91, no. 2, pp. 239–269, 2002.

[5] C. M. Chin and R. Fletcher, "On the global convergence of an SLP-filter algorithm that takes EQP steps," *Mathematical Programming A*, vol. 96, no. 1, pp. 161–177, 2003.

[6] R. Fletcher, N. I. M. Gould, S. Leyffer, P. L. Toint, and A. Wächter, "Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming," *SIAM Journal on Optimization*, vol. 13, no. 3, pp. 635–659, 2002.

[7] R. Fletcher, S. Leyffer, and P. L. Toint, "On the global convergence of a filter-SQP algorithm," *SIAM Journal on Optimization*, vol. 13, no. 1, pp. 44–59, 2002.

[8] S. Ulbrich, "On the superlinear local convergence of a filter-SQP method," *Mathematical Programming B*, vol. 100, no. 1, pp. 217–245, 2004.

[9] M. Ulbrich, S. Ulbrich, and L. N. Vicente, "A globally convergent primal-dual interior-point filter method for nonlinear programming," *Mathematical Programming A*, vol. 100, no. 2, pp. 379–410, 2004.

[10] R. Fletcher and S. Leyffer, "A bundle filter method for nonsmooth nonlinear optimization," Tech. Rep. NA/195, Department of Mathematics, University of Dundee, Dundee, Scotland, December 1999.

[11] L. Grippo, F. Lampariello, and S. Lucidi, "A nonmonotone line search technique for Newton's method," *SIAM Journal on Numerical Analysis*, vol. 23, no. 4, pp. 707–716, 1986.

[12] M. Ulbrich and S. Ulbrich, "Non-monotone trust region methods for nonlinear equality constrained optimization without a penalty function," *Mathematical Programming B*, vol. 95, no. 1, pp. 103–135, 2003.

[13] K. Su and D. Pu, "A nonmonotone filter trust region method for nonlinear constrained optimization," *Journal of Computational and Applied Mathematics*, vol. 223, no. 1, pp. 230–239, 2009.

[14] K. Su and Z. Yu, "A modified SQP method with nonmonotone technique and its global convergence," *Computers & Mathematics with Applications*, vol. 57, no. 2, pp. 240–247, 2009.

[15] N. I. M. Gould and P. L. Toint, "Global convergence of a nonmonotone trust-region filter algorithm for nonlinear programming," in *Multiscale optimization methods and applications*, vol. 82, pp. 125–150, Springer, New York, NY, USA, 2006.

[16] Z. Chen and X. Zhang, "A nonmonotone trust-region algorithm with nonmonotone penalty parameters for constrained optimization," *Journal of Computational and Applied Mathematics*, vol. 172, no. 1, pp. 7–39, 2004.

[17] Z. W. Chen, "A penalty-free-type nonmonotone trust-region method for nonlinear constrained optimization," *Applied Mathematics and Computation*, vol. 173, no. 2, pp. 1014–1046, 2006.

[18] F. A. M. Gomes, M. C. Maciel, and J. M. Martínez, "Nonlinear programming algorithms using trust regions and augmented Lagrangians with nonmonotone penalty parameters," *Mathematical Programming A*, vol. 84, no. 1, pp. 161–200, 1999.

[19] M. J. D. Powell, "Convergence properties of a class of minimization algorithms," in *Nonlinear Programming*, pp. 1–27, Academic Press, New York, NY, USA, 1974.

[20] K. Schittkowski, *More Test Examples for Nonlinear Programming Codes*, vol. 282, Springer, Berlin, Germany, 1987.

[21] W. Hock and K. Schittkowski, *Test Examples for Nonlinear Programming Codes*, vol. 187, Springer, Berlin, Germany, 1981.

[22] M. Lalee, J. Nocedal, and T. Plantenga, "On the implementation of an algorithm for large-scale equality constrained optimization," *SIAM Journal on Optimization*, vol. 8, no. 3, pp. 682–706, 1998.