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

Received 6 December 2012; Accepted 5 March 2013

Academic Editor: Hadi Nasseri

Copyright © 2013 Ke Su et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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,

\[ \min f(x), \quad \text{s.t. } c_i(x) = 0, \quad i \in I = \{1, 2, \ldots, m\}, \]

(P)

where \( x \in \mathbb{R}^n \), the functions \( f : \mathbb{R}^n \rightarrow R \) and \( c_i(i \in I) : \mathbb{R}^n \rightarrow \mathbb{R} \) 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 unconstrained minimization optimization problem:

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

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

$$\begin{align*}
\min q(d) &= g^T d + \frac{1}{2} d^T Hd, \\
\text{s.t.} \quad ||d|| &\leq \Delta,
\end{align*}$$

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

$$q(0) - q(d) \geq \sigma (q(0) - q(d^P)),$$

where $d^P$ 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

$$q(0) - q(d) \geq \frac{\sigma}{2} \|\nabla f(x)\| \min \bigg\{ \Delta, \frac{\|\nabla f(x)\|}{\|Hd\|} \bigg\}. $$

**Proof (see Powell [19] for the proof).** Now, we turn to explain the complete SQP step. Given an approximate estimate of the solution $x_k$ at $k$th iteration, following Dennis Jr. et al. [2] and M. Ulbrich and S. Ulbrich [12], we obtain the trial step $d_k = d^a_k + d^t_k$ by computing a quasi-normal step $d^a_k$ and a tangential step $d^t_k$. The purpose of the quasi-normal step $d^a_k$ is to improve feasibility. To improve optimality, we seek $d^t_k$ 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 $q_k(d) = q^T_k d + (1/2)d^T H_k d$, where $H_k$ is a symmetric approximation of $\nabla^2 f(x)$.

$d^a_k$ is the solution to the subproblem

$$\begin{align*}
\min & \quad \frac{1}{2} \|s_k + A^T_k d^a\|^2 + \frac{\xi}{2} \|d^a\|^2, \\
\text{s.t.} & \quad \|d^a\| \leq \Delta_k,
\end{align*}$$

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

$$\begin{align*}
\min & \quad q_k(d^t_k + d^a), \\
\text{s.t.} & \quad A^T_k d^t = 0, \\
& \quad \|d^t\| \leq \Delta_k.
\end{align*}$$

Then we get the current trial step $d_k = d^a_k + d^t_k$. Let $d^a_k = W_k d^a_k$, whereas $d^t_k \in \mathbb{R}^{n-m}$ and $W_k \in \mathbb{R}^{n \times (n-m)}$ 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 $d^a_k$ and the tangential step $d^t_k$ are required to satisfy

$$\begin{align*}
\|d^a\| &\leq \kappa \Delta_k, \\
\|d^a_k + d^t\| &\leq \Delta_k,
\end{align*}$$

where $0 < \kappa < 1$. Here, to simplify the proof, we only impose a trust region on $\|d^a\| \leq \Delta_k$ and $\|d^t\| \leq \Delta_k$, 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:

$$\nabla_s (q_k(W_k s)) = W_k^T \nabla q_k(W_k s) = W_k^T \nabla q_k(d^t).$$

Define

$$\bar{g}(x) = W(x)^T g(x).$$

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

$$c(\bar{x}) = 0, \quad \bar{g}(\bar{x}) = 0.$$ 

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
the right half-plane

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

\[
\mathcal{S} = \{ k | x_{k+1} = x^*_k \}.
\]

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

\[
h(x) \leq \beta h_j \text{ or } f(x) \leq f_j - \gamma h_j, \quad \forall (h_j, f_j) \in \mathcal{F},
\]

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

\[
\mathcal{D}(\mathcal{F}) = \left\{ (h, f) \mid h > h_j \text{ and } f > f_j, \ \forall j \in \mathcal{F} \right\}.
\]

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 \( \mathcal{D}(\mathcal{F}_k) \) to be the complement of \( \mathcal{D}(\mathcal{F}_k) \). Let

\[
\begin{align*}
h^\mathcal{F}_\min & = \min_{j \in \mathcal{F}_k} h_j, & h^\mathcal{F}_\max & = \max_{j \in \mathcal{F}_k} h_j, \\
f^\mathcal{F}_\min & = \min_{j \in \mathcal{F}_k} f_j, & f^\mathcal{F}_\max & = \max_{j \in \mathcal{F}_k} f_j.
\end{align*}
\]

These four parts are

1. the dominated part of the filter: \( \mathcal{NE}(\mathcal{F}_k) \overset{\text{def}}{=} \mathcal{D}(\mathcal{F}_k) \);
2. the undominated part of lower left corner of the half plane:

\[
\mathcal{SW}(\mathcal{F}_k) \overset{\text{def}}{=} \mathcal{D}(\mathcal{F}_k)^c \cap \left[ 0, h^\mathcal{F}_\max \right] \times \left[ -\infty, f^\mathcal{F}_\max \right],
\]

3. the undominated upper left corner: \( \mathcal{NW}(\mathcal{F}_k) \overset{\text{def}}{=} \left[ 0, h^\mathcal{F}_\min \right] \times (f^\mathcal{F}_\max, +\infty) \);
4. the undominated lower right corner: \( \mathcal{SE}(\mathcal{F}_k) \overset{\text{def}}{=} \left( h^\mathcal{F}_\max, +\infty \right] \times \left[ -\infty, f^\mathcal{F}_\min \right) \).

Consider the trial point \( x^*_k \), if the filter is empty, then define its contribution to the area of the filter by

\[
\alpha(x^*_k, \mathcal{F}_k) \overset{\text{def}}{=} \kappa_F^2,
\]

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 \( (h^*_k, f^*_k) \in \mathcal{SW}(\mathcal{F}_k) \), assume

\[
\alpha(x^*_k, \mathcal{F}_k) \overset{\text{def}}{=} \text{area} \left( \mathcal{D}(\mathcal{F}_k)^c \cap \left[ h^\mathcal{F}_\min, h^\mathcal{F}_\max + \kappa_F \right] \times \left[ f^\mathcal{F}_\min, f^\mathcal{F}_\max + \kappa_F \right] \right).
\]

If \( (h^*_k, f^*_k) \in \mathcal{NW}(\mathcal{F}_k) \), assume

\[
\alpha(x^*_k, \mathcal{F}_k) \overset{\text{def}}{=} \kappa_F \left( h^\mathcal{F}_\min - h^\mathcal{F}_\max \right).
\]

If \( (h^*_k, f^*_k) \in \mathcal{SE}(\mathcal{F}_k) \), assume

\[
\alpha(x^*_k, \mathcal{F}_k) \overset{\text{def}}{=} \kappa_F \left( f^\mathcal{F}_\min - f^\mathcal{F}_\max \right).
\]

If \( (h^*_k, f^*_k) \in \mathcal{NE}(\mathcal{F}_k) \), assume

\[
\alpha(x^*_k, \mathcal{F}_k) \overset{\text{def}}{=} -\text{area} \left( \mathcal{D}(\mathcal{F}_k) \cap \left[ h^\mathcal{F}_\min, h^\mathcal{F}_\max \right] \times \left[ f^\mathcal{F}_\min, f^\mathcal{F}_\max \right] \right),
\]

where \( \mathcal{F}_k = \{(h, f) \in \mathcal{F}_k | h_j < h^\mathcal{F}_k, f_j < f^\mathcal{F}_k \} \), and \( h^\mathcal{F}_\min \overset{\text{def}}{=} \min_{j \in \mathcal{F}_k} h_j, f^\mathcal{F}_\min \overset{\text{def}}{=} \min_{j \in \mathcal{F}_k} f_j \).

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

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

(21)

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

$$\mathcal{F}_{k+1} \leftarrow (\mathcal{F}_k \setminus \mathcal{P}_k) \cup (h_{\min}^{\rho_k}, f_k) \cup (h_k, f_k^\rho) .$$

(22)

We now return to the question of deciding whether a trial point $x_k^r$ 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 $x_{k+1} = x_k^r$. If we consider an iterate $x_k$, there must exist a predecessor iteration such that $x_{p(k)}^r = x_{p(k)+1} = x_k$. Under the monotonic situation, a trial point $x_k^r$ would be accepted whenever it results in an sufficient increase in the dominated area of the filter, that means $x_k^r$ would be accepted whenever

$$\alpha_k \geq \gamma_\varphi(h_k^r)^2,$$

(23)

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

$$\max \left\{ \alpha_k, \sum_{j \in \mathcal{U}} \lambda_{p(j)} \alpha_{p(j)} \right\} \geq \gamma_\varphi(h_k^r)^2,$$

(24)

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

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

According to condition (24), it is possible to accept $x_k^r$ even though it may be dominated. Then $x_k^r$ 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

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

(25)

and the actual reduction

$$\text{ared} (d_k) = f (x_k) - f (x_k^r).$$

(26)

Moreover, let $r_k = \text{ared} (d_k) / \text{pred} (d_k)$, if there exists a nonzero constant $\eta_3$ such that $r_k \geq \eta_3$ and condition (23) and (24) hold, the trial point $x_k^r$ will be called acceptable. Then the next trial point $x_{k+1}^r$ is obtained, and for its feasibility, we consider the condition

$$h_{k+1} \leq \eta_3 \min \{\mu, \alpha_1 \Delta_k^{2+\alpha_2}\},$$

(27)

is true or not, where $\eta_3$ and $\mu$ are all positive constants, if it is not true, then turn to the feasibility restoration phase and define

$$r^j_k = \frac{\|c(x_k^j)\|^2}{\|c(x_k^j + A(x_k^j)^T d_k^j)\|^2},$$

(28)

A formal description of the algorithm is given as follows.

**Algorithm A**

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

*Step 1.* Compute $c_\varphi, A_\varphi, W_k, f_k, h_k, g_k, \hat{g}_k = W_k^T g_k$.

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

*Step 3.* Solve the subproblems (5) and (6) to get the quasi-normal step $d_k^0$ and the tangential step $d_k^T$. Let $d_k = d_k^0 + d_k^T$, $x_k = x_k + d_k$.

*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 $x_{k+1} = x_k + d_k$ and update the filter according to (21) and (22), the trust region radius and $H_k$, then get the corresponding $h_{k+1} = f_k^r$. 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 $x_0^0 = x_k, \Delta_0^0 = \Delta_k, j = 0, \eta_2, \eta_3 \in (0, 1)$, $\alpha_1, \alpha_2 \in [0, 1]$, $\mu > 0$.

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

*Step 2.* Compute

$$\min \frac{\|c(x_k^r) + A(x_k^r)^T d_k^r\|^2}{\|c(x_k^r) + A(x_k^r)^T d_k^r\|^2},$$

(29)

s.t. $\|d_k^r\| \leq \Delta_k^j$ to get $d_k^r$, then compute $r_k^j$.

*Step 3.* If $r_k^j \leq \eta_2, x_k^{j+1} = x_k^r, \Delta_k^{j+1} = \Delta_k^j / 2, j = j + 1$ 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 \( h_k \neq 0 \) is defined in \( [h_1, h_2, \ldots, h_m] \) as the lower bound of the trust region radius. Given \( \eta_1 > 0 \), \( 0 < \gamma_j < \gamma_1 < 1 \leq \gamma_2 < \gamma_3 \leq 2 \), we have the following.

1. If \( r_k < \eta_1 \) or \( x_k^j \) is acceptable to the filter, \( \Delta_k^{j+1} \in [\gamma_2 \Delta_k, \gamma_3 \Delta_k] \).
2. If \( x_k^j \) is acceptable to the filter but does not satisfy \( \Delta_k^j \leq \gamma_2 \Delta_k \).
3. If \( x_k^j \) is acceptable to the filter and satisfies \( \Delta_k^{j+1} \in [\gamma_2 \Delta_k, \gamma_3 \Delta_k] \).

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 \( \Omega \subset \mathbb{R}^n \).

\( (A3) \) The matrix \( A(x) = Vc(x) \) 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 \( \Omega \), 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 \( v_0 \), \( v_1 \), \( v_2 \), \( v_3 \), \( v_4 \) such that \( \| f(x) \| \leq v_0 \), \( \| g(x) \| \leq v_0 \), \( \| c(x) \| \leq v_0 \), \( \| A(x) \| \leq v_0 \), \( \| A(x)^T A(x)^{-1} \| \leq v_1 \), \( \| W(x) \| \leq v_2 \), \( \| H_k \| \leq v_3 \), \( \| W_k^T H_k \| \leq v_3 \), \( \| W_k^T W_k \| \leq v_3 \).

By \( (A1) \) and \( (A2) \), it holds

\[
0 \leq h_k \leq h_{\text{max}} \quad \forall k.
\]

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

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

\[
|f(x_k + d_k) - f_k| \leq \bar{\varphi} \Delta_k^j.
\]

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

\[
\| d_k^j \| \leq \alpha_k \| c_k \|.
\]

**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

\[
\| q_k \|^2 - \left\| q_k^j + A_k^j d_k^j \right\|^2 \geq k_2 \| c_k \| \min \left\{ k_3 \| c_k \|, \Delta_k \right\}.
\]

**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

\[
K = \left\{ j \mid r_k^j = \left\| c_k^j - c_k^j + (A_k^j)^T d_k^j \right\|^2 > \eta_2 > 0 \right\},
\]

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

\[
\sum_{j=1}^{\infty} (h_k^{j-1} - h_k^j) \geq \sum_{j=1}^{\infty} \left( \| c_k^j \|^2 - \| c_k^j + (A_k^j)^T d_k^j \|^2 \right)
\]

\[
\geq \eta_2 \sum_{j=1}^{\infty} \min \left\{ k_4 \| c_k \|, \Delta_k^j \right\}.
\]

By \( h_k^j \to 0 \), \( h_k^j \to 0 \) for \( j \in K \). From Algorithm B, we can obtain that \( \Delta_k^j \to 0 \) for all \( j \).

On the other side,

\[
\| c_k^j \|^2 - \left\| c_k^j + (A_k^j)^T d_k^j \right\|^2 = \| c_k^j \|^2 - \left\| c_k^j + (A_k^j)^T d_k^j \right\|^2 + \Delta_k^j
\]

for \( \Delta_k^j \to 0 \). By the algorithm, the radius \( \Delta_k^j \) 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 \( x_{n(k)} \). 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,
\[
i(k) = p(k) \quad \text{if } p(k) \notin \mathcal{A}
\]  
(37)
the other is by the criteria (24), that is,
\[
i(k) = p(k) \quad \text{if } p(k) \notin \mathcal{A}.
\]  
(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 \( \mathcal{C}_k = \{l_1, l_2, \ldots, l_q \} \) for each \( k \), such that
\[
\begin{align*}
x_{i_1} &= x_0, \quad x_{i_q} = x_k, \quad x_{i_j} = x_{i(l_{j+1})}, \quad j = 1, 2, \ldots, q - 1, \\
x_{i_q} &= x_k.
\end{align*}
\]  
(39)
where \( l_i \) 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 \( \mathcal{C}_k = \{l_1, l_2, \ldots, l_q \} \), it holds
\[
\operatorname{area}(\mathcal{D}(\mathcal{F}_k)) \geq \gamma_{\mathcal{F}} \sum_{j=1}^{q} h_{l_j}^2.
\]  
(40)

**Proof.** For \( \forall l_j \in \mathcal{C}_k \), if \( p(l_j) \notin \mathcal{A} \), by (24), \( i(l_j) = r(p(l_j)) = l_{j-1} \), then
\[
\max \left\{ \alpha_{p(l_j)}, \sum_{i=l_{j-1}+1}^{l_j} \lambda_{p(i)} \alpha_{p(i)} \right\} \geq \gamma_{\mathcal{F}} h_{l_j}^2.
\]  
(41)
If \( \max \{ \alpha_{p(l_j)}, \sum_{i=l_{j-1}+1}^{l_j} \lambda_{p(i)} \alpha_{p(i)} \} = \alpha_{p(l_j)} \),
\[
\alpha_{p(l_j)} \geq \gamma_{\mathcal{F}} h_{l_j}^2.
\]  
(42)
If \( \alpha_{p(l_j)}, \sum_{i=l_{j-1}+1}^{l_j} \lambda_{p(i)} \alpha_{p(i)} \) = \sum_{i=l_{j-1}+1}^{l_j} \lambda_{p(i)}
\[
\sum_{i=l_{j-1}+1}^{l_j} \alpha_{p(i)} \geq \gamma_{\mathcal{F}} h_{l_j}^2.
\]  
(43)
If \( p(l_j) \notin \mathcal{A} \), \( l_{j-1} = p(l_j) \). By \( \mathcal{U} \subseteq \mathcal{A} \), it holds
\[
\{l_{j-1} + 1, \ldots, l_j\} \cap \mathcal{U} \subseteq \{l_{j-1} + 1, \ldots, l_j\} \cap \mathcal{A} = \{l_j\}.
\]  
(44)
Then from (22), \( \alpha_{p(0)} \geq \gamma_{\mathcal{F}} h_{l_j}^2 \). It implies (43). Moreover
\[
\operatorname{area}(\mathcal{D}(\mathcal{F}_k)) \geq \sum_{i=0}^{k} \alpha_{p(i)} \geq \sum_{i=0}^{k} \left( \sum_{i=l_{j-1}+1}^{l_j} \alpha_{p(i)} \right).
\]  
(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 \( \lim_{k \to \infty} \delta_k = 0 \).

**Proof.** Suppose by contradiction that there exists a constant
\( \varepsilon > 0 \) and infinite sequence \( \{k_i\} \subseteq \mathcal{A} \) such that \( \delta_{k_i} \geq \varepsilon \)
for all \( i \). Because there are infinite iterations in the filter, we have
\( |\delta| = \infty \), then \( \delta_{k_i} \geq \varepsilon \) for \( \forall q \).
\[
\operatorname{area}(\mathcal{D}(\mathcal{F}_k)) \geq \gamma_{\mathcal{F}} \cdot q \cdot \varepsilon^2.
\]  
(46)
Then by (31), \( \operatorname{area}(\mathcal{D}(\mathcal{F}_k)) \) is upper bounded for each \( k \). That
means it exists \( \kappa_{\mathcal{F}} \geq 0 \) such that \( \operatorname{area}(\mathcal{D}(\mathcal{F}_k)) \leq \kappa_{\mathcal{F}} \), so
\( i \leq \kappa_{\mathcal{F}} / \gamma_{\mathcal{F}} \varepsilon^2 \). 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 \( \delta_k = 0 \).

**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
\[
q_k(0) - q_k(d_k^r) \geq -k_5 \|c_k\|.
\]  
(47)

**Proof.** By (32), the assumptions and \( \|d_k^r\| \leq \Delta_{r_{\max}} \), it is obvious that
\[
q_k(0) - q_k(d_k^r) = -g_k^T d_k^r - \frac{1}{2} (d_k^r)^T H_k d_k^r
\]  
\[
\geq - \|d_k^r\| \left( \|g_k\| + \frac{1}{2} \|H_k\| \|d_k^r\| \right)
\]  
\[
\geq -k_5 \|c_k\| (\eta_0 + \nu_3 \Delta_{r_{\max}})
\]  
\[
def = -k_5 \|c_k\|.
\]  
(48)
The proof is complete.

**Lemma 10.** Suppose that Assumptions hold and \( \|\delta_k\| \geq \varepsilon \), if
\[
\Delta_k \leq \min \left\{ \left( \frac{\epsilon_k}{2k_1 \gamma_3 \sqrt{\eta_3 \alpha_1}} \right)^{2(2+\alpha_3)}, \frac{k_1 \epsilon_k}{2}, \right\}
\]  
(49)
\[
\Delta_k \leq \min \left\{ \left( \frac{\gamma_0}{k_1 \gamma_3 \sqrt{\eta_3 \alpha_1}} \right)^{2(2+\alpha_3)}, \frac{\sigma \epsilon_k}{16k_1 \gamma_3 \sqrt{\eta_3 \alpha_1}} \right\} \def = \delta_k
\]  
(49)
one can deduce
\[
q_k(0) - q_k(d_k^r) \geq \frac{\sigma \epsilon_k}{8} \Delta_k \def = \sigma \Delta_k.
\]  
(50)
**Proof.** By the assumptions and the definition of $d^n_k$, it holds
\[
\|W_k^TV_k q_k (d^n_k)\| = \|W_k^T (g_k + H_k d^n_k)\| \geq \|\hat{g}_k\| - \|W_k^T H_k d^n_k\|
\geq \epsilon_1 - \gamma_3 k_1 \sqrt{\eta_3 \alpha_1 \Delta_k} 1^{1+\alpha/2} \geq \frac{\epsilon_1}{2} .
\]
(51)

From Lemma 3, $q_k(d^n_k) - q_k(d_k) \geq (\sigma/4) \epsilon_2 \min \{\Delta_k, k_4 \epsilon_2 / 2\} \geq \sigma \epsilon_2 / 4 \Delta_k$. Together with
\[
q_k(0) - q_k(d^n_k) = -g_k^T d^n_k - \frac{1}{2} (d^n_k)^T H_k d^n_k
\geq - \|g_k\| \|d^n_k\| - \frac{1}{2} \|H_k\| \|d^n_k\|^2
\geq - k_1 \gamma_0 \sqrt{\eta_3 \alpha_1} \Delta_k^{1+\alpha/2} - k_1 \gamma_2 \gamma_3 \alpha_3 \Delta_k^{2+\alpha/2}
\geq - 2 k_1 \gamma_0 \sqrt{\eta_3 \alpha_1} \Delta_k^{1+\alpha/2}
\geq - \frac{\sigma \epsilon_2}{8} \Delta_k
\]
which together with $q_k(0) - q_k(d_k) \geq -(\sigma \epsilon_2 / 4) \Delta_k + (\sigma \epsilon_2 / 4) \Delta_k = (\sigma \epsilon_2 / 8) \Delta_k = \overline{\sigma} \Delta_k$. It is the conclusion.

**Lemma 11.** Suppose the conditions of Lemma 10 hold, if
\[
\Delta_k \leq \min \left\{ \delta_1, \left( \frac{1 - \eta_1}{V} \right) \right\} \overset{\text{def}}{=} \delta_2
\]
then $r_k \geq \eta_1$.

**Proof.** From the definition of $r_k$ and Lemma 10, together with (31), we have
\[
|r_k - 1| = \left| \frac{f(x_k + d_k) - q_k(d_k)}{q_k(0) - q_k(d_k)} \right| \leq \frac{\overline{\sigma} \Delta_k^2}{\overline{\sigma} \Delta_k} \leq 1 - \eta_1.
\]
(52)

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

**Lemma 12.** Suppose the conditions of Lemmas 10 and 11 hold, if
\[
h_k \leq (\eta_3 \alpha_3)^{-1/(1+\alpha_3)} \left( \frac{\eta_3 \overline{\sigma}}{\sqrt{\gamma F}} \right)^{(2+\alpha_3)/(1+\alpha_3)}
\]
then
\[
f(x_k) \leq f(x_k) - \sqrt{\gamma F} h_k .
\]
(53)

**Proof.** By Lemmas 3, 10, and 11, together with $\|\hat{g}_k\| \geq \epsilon_1$, it holds
\[
f(x_k) - f(x^+_k) \geq \eta_1 (q_k(0) - q_k(d_k)) \geq \eta_1 \overline{\sigma} \Delta_k .
\]
(54)

From the Algorithm, $h_k \leq \eta_3 \alpha_3 \Delta_k^{2+\alpha_3}$, then
\[
f(x_k) - f(x^+_k) \geq \eta_1 \overline{\sigma} \left( \frac{h_k}{\eta_3 \alpha_3} \right)^{1/(2+\alpha_3)}
\]
(55)

Hence $f(x_k) - f(x^+_k) \geq \sqrt{\gamma F} h_k$.

**Theorem 13.** Suppose the assumptions hold, there must exist $\Delta_{\min} > 0$ such that for each $k$, it holds
\[
\Delta_k \geq \Delta_{\min}.
\]
(56)

**Proof.** Let $\overline{\Delta}_1$ be large enough such that $h_{\overline{\Delta}_1} \leq \epsilon_\gamma$, it is true by Lemmas 6 and 7. Suppose by contradiction that the index $j$ is the first one after $\overline{\Delta}_1$, which satisfies
\[
\Delta_j \leq \gamma_0 \min \left\{ \delta_2, \left( \frac{1 - \sqrt{\gamma F}}{\eta_3 \alpha_1} \right)^{1/(2+\alpha_3)}, \overline{\Delta}_1 \right\}
\]
(60)

where $H^F \overset{\text{def}}{=} \min_{\gamma \in [0,1]} h_j$ is the smallest value of violation function in filter. Then $\Delta_j \leq \gamma_0 \overline{\Delta}_1$. By the above analysis, we know $j \geq \overline{\Delta}_1 + 1$, that is $j - 1 \geq \overline{\Delta}_1$. From the Algorithm and (60), it concludes
\[
\Delta_{j-1} \leq \frac{1}{\gamma_0} \Delta_j \leq \delta_3.
\]
(61)

By (60) and (61), (53) can be obtained. In Lemma 11, let $j - 1$ instead of $k$, it deduces
\[
r_{j-1} \geq \eta_1.
\]
(62)

Based on Lemma 12, together with (60), (61), and the algorithm, we can see
\[
f_{j-1}^+ \leq f_{j-1}^+ - \sqrt{\gamma F} h_{j-1}.
\]
(63)

It can be seen that (53) is true for $j - 1 \geq k$, with (55), we can deduce
\[
f_{j-1}^+ \leq f_{j-1} - \sqrt{\gamma F} h_{j-1} .
\]
(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 $\overline{\Delta}_1$ which satisfied (60), that is a contradiction. So, for any $k > \overline{\Delta}_1$, it holds $\Delta_k \geq \gamma_0 \overline{\Delta}_1$. Define
\[
\Delta_{\min} = \min \{ \Delta_0, \ldots, \Delta_{\overline{\Delta}_1}, \gamma_0 \delta_3 \}
\]
(65)

we can see that
\[
\Delta_k \geq \Delta_{\min}
\]
(66)
holds for each $k$. The proof is complete.

**Lemma 14.** Suppose that Assumptions hold and Algorithm A does not terminate finitely, then $\lim \inf_{k \to \infty} \|g_k\| = 0$.

**Proof.** Suppose by contradiction that for $\epsilon_i$, there exists a constant $\overline{k} > 0$ such that $\|g_k\| \geq \epsilon_i$.

By Assumption (A3) and (A4), $\|W_k^TV_k q_k(d^n_k)\| \geq \|\hat{g}_k\| - \|v_k k_1\| \geq \epsilon_i k_1 \|c_k\|$. From Lemma 6, we know $h_k \rightarrow 0$. Hence there exists $\overline{k} > 0$ such that
\[
\|c_k\| \leq \frac{2 \epsilon_i}{3 \gamma_3 k_1}
\]
(67)
for $k > \tilde{k}$. Then $\|W_k^T\nabla q_k(d_k^p)\| \geq (1/3) \| g_k \| \geq (1/3) \epsilon_i$ for $k > \tilde{k}$.

It is obvious that

$$q_k(0) - q_k(d_k) = q_k(0) - q_k(d_k^p) + q_k(d_k^p) - q_k(d_k)$$  \hspace{1cm} (68)

By the proof of Lemma 9, it holds $\|q_k(0) - q_k(d_k^p)\| \leq \nu_0 \| d_k^m \| + (1/2) \nu_5 \| d_k^p \|^2$. Together with Lemma 6 and the definition of $d_k$, we have

$$\lim_{k \to \infty} (q_k(0) - q_k(d_k)) = 0.$$  \hspace{1cm} (69)

By the Algorithm, we can get

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

Then

$$\lim_{k \to \infty} (q_k(0) - q_k(d_k)) = 0.$$  \hspace{1cm} (71)

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

$$\lim_{k \to \infty} (q_k(d_k^p) - q_k(d_k)) = 0.$$  \hspace{1cm} (72)

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

$$q_k(d_k^p) - q_k(d_k)$$

$$\geq \frac{\sigma}{2} \|W_k^T\nabla q_k(d_k^p)\| \min \{\Delta_k, k \epsilon_i \|W_k^T\nabla q_k(d_k^p)\| \}$$

$$\geq \frac{\sigma \epsilon_i}{6} \min \{\Delta_{\min}, \frac{k \epsilon_i}{6} \} > 0,$$

which contradicts (72). The conclusion follows.  \hspace{1cm} \Box

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

$$\lim_{k \to \infty} (h_k + \|g_k\|) = 0,$$

where $g_k = W_k^T \tilde{g}_k, \tilde{g}_k = \langle f(x_k), W(x) \rangle$ 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.  \hspace{1cm} \Box

**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

$$\lim_{j \to \infty} x_k^j = x^*$$

(75)

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

**Table 1.**

<table>
<thead>
<tr>
<th>Problem</th>
<th>n</th>
<th>m</th>
<th>NF</th>
<th>NG</th>
<th>Ls NF</th>
<th>Ls NG</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS6</td>
<td>2</td>
<td>1</td>
<td>11</td>
<td>11</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>HS7</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>9</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>HS8</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS9</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>HS26</td>
<td>3</td>
<td>1</td>
<td>24</td>
<td>24</td>
<td>42</td>
<td>24</td>
</tr>
<tr>
<td>HS39</td>
<td>4</td>
<td>2</td>
<td>15</td>
<td>9</td>
<td>26</td>
<td>19</td>
</tr>
<tr>
<td>HS40</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>HS42</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>HS78</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

**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), the new algorithm is robust and effective, and more flexible for the acceptance of the trial iterate.

### 6. Some Numerical Experiments

1. Updating of $H_k$ is done by $H_{k+1} = H_k + \eta_p \nabla^2 f_k - \eta_k \nabla \tilde{g}_k$.
2. The algorithm parameters were set as follows: $H_0 = 1 \in R^{n \times n}, \eta_1 = 0.25, \eta_2 = 0.25, \eta_3 = 0.1, \alpha_1 = \alpha_2 = 0.5, \gamma = 0.02, \rho = 0.5, \xi = 10^{-6}, \nu_0 = 0.1, \nu_1 = 0.1, \nu_2 = 0.001, \gamma = 10^{-3}, \Delta_0 = 1$. 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 “Ls” is the solution in [22]. The numerical results show that our algorithm is more effective than the Ls 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 trial iterate.
Acknowledgments

This research is supported by the National Natural Science Foundation of China (no. 11011115) and the Natural Science Foundation of Hebei Province (nos. A2010000191, Y2012021).

References


