Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 854890, 6 pages
http://dx.doi.org/10.1155/2013/854890

Research Article

A Generalized Gradient Projection Filter Algorithm for Inequality Constrained Optimization

Wei Wang,[1] Shaoli Hua,[2] and Junjie Tang[1]

[1] *Department of Mathematics, East China University of Science and Technology, Shanghai 200237, China*
[2] *China UnionPay Merchant Services Co., Ltd., Shanghai 200135, China*

Correspondence should be addressed to Wei Wang; wangwei@ecust.edu.cn

Received 27 May 2013; Revised 22 August 2013; Accepted 5 September 2013

Academic Editor: Igor Andrianov

A generalized gradient projection filter algorithm for inequality constrained optimization is presented. It has three merits. The first is that the amount of computation is lower, since the gradient matrix only needs to be computed one time at each iterate. The second is that the paper uses the filter technique instead of any penalty function for constrained programming. The third is that the algorithm is of global convergence and locally superlinear convergence under some mild conditions.

1. Introduction

The optimal problems are often discovered in the field of management, engineering design, traffic transportation, national defence, and so on. The efficient algorithms for these problems are important. We will consider the following nonlinear inequality constrained optimization problem:

$$\begin{aligned} &\min \quad f(x) \\ &\text{s.t.} \quad c_j(x) \le 0, \end{aligned} \tag{1}$$

where $I = \{1, 2, \ldots, m\}$ and $x \in R^n$; assume that $f : R^n \to R$ and $c_j (j \in I) : R^n \to R$ are continuously differentiable.

In 2002, Fletcher and Leyffer [1] had proposed a filter method for nonlinear inequality constrained optimization, which did not require any penalty function. The main idea is that a trial point is accepted if it improves either the objective function or the constraint violation. Fletcher et al. [2, 3] and Gonzaga et al. [4] had proved that the method was of global convergence. More recently, this method has been extended by Wächter and Biegler [5, 6] and Chin [7] to line search method and by Su [8] to the SQP method.

In this paper, we modify the method given by Wang et al. [9] and propose a generalized gradient projection filter algorithm for inequality constrained optimization with arbitrary initial point. It is organized as follows. In Section 2, we first review the filter method and some definitions of generalized gradient projection and then introduce an algorithm for problem (1). The convergence and the rate of convergence on the algorithm are discussed in Sections 3 and 4, respectively. In the last section, we shall list the numerical tests.

2. Preliminaries and a Filter Algorithm

Let $h(x)$ be a violation function; that is,

$$h(x) = \max\left\{0, c_j(x),\ j \in I\right\}. \tag{2}$$

Definition 1. A pair $(h(x_k), f(x_k))$ obtained on iteration k dominates another pair $(h(x_l), f(x_l))$ if and only if $h(x_k) \le h(x_l)$ and $f(x_k) \le f(x_l)$ hold.

Definition 2. A filter is a list of pairs $(h(x_k), f(x_k))$ such that no pair dominates any other. A pair $(h(x_k), f(x_k))$ is said to be acceptable for the filter if it is not dominated by any point in the filter.

We use $F^{(k)}$ to denote the set of iterations indices $j\,(j < k)$ such that $(h(x_j), f(x_j))$ is an entry in the current filter. A point x is said to be "acceptable for the filter" if and only if

$$h(x) \le \left(1 - \alpha^2\eta\right) h\left(x_j\right) \quad \text{or} \quad f(x) \le f\left(x_j\right) - \gamma h\left(x_j\right) \tag{3}$$

holds for all $j \in F^{(k)}$, where $\gamma, \eta \in (0,1)$ is close to zero and α is the step size. We may also "update the filter" which means that the pair $(h(x), f(x))$ is added to the list of pairs in the filter, and any pairs in the filter that are dominated by $(h(x), f(x))$ are removed.

However, relying solely on this criterion would result in convergence to a feasible but nonoptimal point. In order to prevent this, we employ the following sufficient reduction criterion.

We denote $\Delta f_k = f(x_k) - f(x_k + \alpha d_k)$ and $\Delta l_k = -\alpha \nabla f(x_k)^T d_k$ as actual reduction and linear reduction, respectively, at $f(x_k)$. The sufficient reduction condition for $f(x_k)$ takes the form

$$\Delta l_k \geq 0, \qquad \Delta f_k \geq \sigma \Delta l_k, \tag{4}$$

where $\sigma \in (0, 1/2)$ is a preassigned parameter.

At the current iterate x_k, define that $J(x_k) = \{j \in I : \epsilon \leq c_j(x_k) - h(x_k) \leq 0\}$, $A_k = (\nabla c_j(x_k), j \in J(x_k))$, and $c_{J_k} = (c_j(x_k), j \in J(x_k))^T$, and then

$$\begin{gathered} d_k^0 = -P_k \nabla f(x_k) - B_k^T c_{J_k}, \\ \lambda_k = -B_k \nabla f(x_k) + \left(A_k^T H_k A_k\right)^{-1} c_{J_k} = \lambda_k^1 + \lambda_k^2, \end{gathered} \tag{5}$$

where H_k is a given symmetric positive definite matrix, $\lambda_k^1 = -B_k \nabla f(x_k)$, $\lambda_k^2 = (A_k^T H_k A_k)^{-1} c_{J_k}$, $B_k = (A_k^T H_k A_k)^{-1} A_k^T H_k$, and $P_k = H_k - H_k A_k B_k$.

Let $U_k = (u_{k_j}, j \in J(x_k)^T)$, where $u_{k_j} = \begin{cases} \lambda_{k_j}^1 & \lambda_{k_j}^1 < 0 \\ 0 & \lambda_{k_j}^1 \geq 0 \end{cases}$. Set $d_k^1 = -P_k \nabla f(x_k) + B_k^T U_k$ and $d_k^2 = -P_k \nabla f(x_k) + B_k^T \|d_k^1\| e$, where $e = (1, \ldots, 1)^T$. Then

$$d_k = (1 - \rho_k) d_k^1 + \rho_k d_k^2, \tag{6}$$

where $\rho_k = \max\{\rho \in (0,1] : \nabla f(x_k)^T((1-\rho) d_k^1 + \rho d_k^2) \leq \theta \nabla f(x_k)^T d_k^1\}$, $\theta \in (1/2, 1)$. We use correction direction d_k if a trial point has been rejected.

The following is the algorithm.

Algorithm

(S0) Given start point $x_0 \in R^n$, $\epsilon_0, \epsilon_1 > 0$, $\mu = h(x_0)$, $\eta, \gamma \in (0,1)$, and $\beta, t, \sigma \in (0, 1/2)$. Initialize the filter $\Phi_0 = \{(\mu, +\infty) \in R^2\}$ and $F^{(0)} = \emptyset$. Set $k = 0$.

(S1) Inner loop A:

- (S1.1) set $i = 0$ and $\epsilon_{k_i} = \epsilon_0$;
- (S1.2) if $\det(A_{k_i}^T A_{k_i}) \geq \epsilon_{k_i}$, where $A_{k_i} = (\nabla c_j(x_k) : j \in J_{k_i})$ and $J_{k_i} = \{j \in I : \epsilon_{k_i} \leq c_j(x_k) - h(x_k) \leq 0\}$, then set $J(x_k) = J_{k_i}$, $A_k = A_{k_i}$, and $\epsilon_k = \epsilon_{k_i}$, and go to S2;
- (S1.3) let $i = i + 1$, $\epsilon_{k_i} = \epsilon_{k_{i-1}}/2$, and go to S1.2.

(S2) Compute d_k^0, λ_k by (5). If $d_k^0 = 0$ and $\lambda_k \geq 0$, then stop.

(S3) Test direction d_k^0:

- (S3.1) if $\lambda_{kj} \geq \epsilon_1$, and $x_k + d_k^0$ is acceptable for the filter, then go to S3.2; otherwise, go to S4;
- (S3.2) if $h(x_k) > 0$, let $x_{k+1} = x_k + d_k^0$, and go to S7; otherwise, go to S3.3;
- (S3.3) if $x_k + d_k^0$ satisfies the sufficient reduction condition (4), then let $x_{k+1} = x_k + d_k^0$, and go to S7; otherwise, go to S4.

(S4) Compute d_k by (6) and set $\alpha = 1$.

(S5) Inner loop B:

- (S5.1) if $x_k + \alpha d_k$ is acceptable for the filter, go to S5.2; otherwise, go to S5.3;
- (S5.2) if $\Delta f_k < \sigma \Delta l_k$, go to S5.3; otherwise, go to S6;
- (S5.3) set $\alpha = t\alpha$, and go to S5.1.

(S6) Set $\alpha_k = \alpha$ and $x_{k+1} = x_k + \alpha d_k$.

(S7) Update filter $F^{(k)}$ to $F^{(k+1)}$. Update H_k to H_{k+1} by a quasi-Newton method. Set $k = k + 1$, and back to S1.

3. Global Convergence of the Algorithm

In this section, we assume that the following conditions hold.

(A1) $\{\nabla c_j(x), j \in J(x)\}$ is linearly independent of any $x \in R^n$.

(A2) For any k and d, $a\|d\|^2 \leq d^T H_k^{-1} d \leq b\|d\|^2$ holds, where $0 < a \leq b$ are constants.

(A3) Sequence $\{x_k\}$ generated by the algorithm remains in a closed, bounded subset $\Omega \subset R^n$.

(A4) $f(x)$ and $c_i(x)$ $(i = 1, 2, \ldots, m)$ are twice differentiable in Ω; that is, $M_{\min}^f \leq \lambda(\nabla^2 f(x)) \leq M_{\max}^f$, $M_{\min}^c \leq \lambda(\nabla^2 c_i(x)) \leq M_{\max}^c$.

Similar to [9], the following theorem and lemma hold.

Theorem 3. *If $d_k^0 = 0$ and $\lambda_k \geq 0$ hold, then x_k is a KKT point of problem* (1).

Lemma 4. *Consider*

$$d_k^0 = 0, \qquad \lambda_k \geq 0 \Longleftrightarrow d_k^1 = 0. \tag{7}$$

According to [8], the following lemma holds.

Lemma 5. *The inner loop A will terminate in finite times.*

Lemma 6. *If x_k is not a KKT point of problem* (1)*, there must exist $\nabla f(x_k)^T d_k < 0$ and $\nabla c_j(x_k)^T d_k < 0$, $j \in J(x_k)$.*

Proof. Since x_k is not a KKT point, we have either $d_k^0 \neq 0$ or $j \in J(x_k)$ such that $\lambda_{kj}^1 < 0$. Thus

$$\begin{aligned} \nabla f(x_k)^T d_k &\leq \theta \nabla f(x_k)^T d_k^1 \\ &\leq \theta \left[-\left(d_k^0\right)^T H_k^{-1} d_k^0 - \sum_{\lambda_{k_j}^1 < 0} \left(\lambda_{k_j}^1\right)^2 \right] < 0 \end{aligned} \tag{8}$$

holds. From Lemma 4, we know that $d_k^1 \neq 0$. Therefore

$$A_{J_k}^T d_k^1 = U_k \leq 0, \qquad A_{J_k}^T d_k^2 = -\left\|d_k^1\right\| e < 0. \tag{9}$$

That is, $\nabla c_j(x_k)^T d_k < 0$ $(j \in J(x_k))$ hold. □

Lemma 7. *Let x^∞ be the cluster point of $\{x_k\}$ generated by algorithm. If x^∞ is not the KKT point of problem* (1)*, there exists $\overline{\alpha} > 0$, such that $\Delta f_k \geq \sigma \Delta l_k$ holds when $\alpha \leq \overline{\alpha}$.*

Proof. From the definition of ρ and the assumption (A2), we have

$$\begin{aligned}\Delta l_k &\geq -\alpha\theta \nabla f(x_k)^T d_k^1 \\ &= -\alpha\theta\left[-\left(d_k^0\right)^T H_k^{-1} d_k^0 - \sum_{\lambda_{k_j}^1 < 0}\left(\lambda_{k_j}^1\right)^2\right] \\ &\geq \alpha\theta\left(d_k^0\right)^T H_k^{-1} d_k^0 \geq \frac{\alpha\theta a}{2}\left\|d_k^0\right\|^2.\end{aligned} \tag{10}$$

Since

$$\begin{aligned}\left|\Delta f_k - \Delta l_k\right| &= \left|f(x_k) - f(x_k + \alpha d_k) + \alpha \nabla f(x_k)^T d_k\right| \\ &\leq \frac{1}{2}\alpha^2 M_{\max}^f \left\|d_k\right\|^2,\end{aligned} \tag{11}$$

we have

$$\left|\frac{\Delta f_k - \Delta l_k}{\Delta l_k}\right| \leq \frac{(1/2)\,\alpha^2 M_{\max}^f \left\|d_k\right\|^2}{(\alpha\theta a/2)\left\|d_k^0\right\|^2} = \frac{\alpha M_{\max}^f \left\|d_k\right\|^2}{\theta a \left\|d_k^0\right\|^2}. \tag{12}$$

It is easy to learn that $\Delta f_k \geq \sigma \Delta l_k$ holds when $\alpha \leq \overline{\alpha} = (1 - \sigma)\theta a \|d_k^0\|^2 / M_{\max}^f \|d_k\|^2$. □

Lemma 8. *The inner loop B will end in finite times.*

Proof. From Lemma 7, we have that $\Delta f_k \geq \sigma \Delta l_k$ holds when $\alpha \leq \overline{\alpha}$. By contradiction, if the conclusion is false, then the algorithm will run infinitely between S5.1 and S5.3, so we have $\alpha \to 0$ and $x_k + \alpha d_k$ not acceptable for the filter. We consider it in the following two cases.

Case 1 ($h(x_k) = 0$). From Lemma 6, we have $\nabla f(x_k)^T d_k < 0$ and $\nabla c_j(x_k)^T d_k < 0$, $j \in J(x_k)$. So when

$$\alpha \leq \min\left\{\frac{-\nabla f(x_k)^T d_k}{(1/2)\,M_{\max}^f \left\|d_k\right\|^2}, \min_{j \in J(x_k)}\left\{\frac{-\nabla c_j(x_k)^T d_k}{(1/2)\,M_{\max}^c \left\|d_k\right\|^2}\right\}\right\}, \tag{13}$$

it is easy to get that

$$\begin{aligned}f(x_k + \alpha d_k) &\leq f(x_k) + \alpha \nabla f(x_k)^T d_k \\ &\quad + \frac{1}{2}\alpha^2 M_{\max}^f \left\|d_k\right\|^2 \leq f(x_k) \\ &= f(x_k) - \gamma h(x_k), \\ h(x_k + \alpha d_k) &\leq \max\left\{0, c_j(x_k) + \alpha \nabla c_j(x_k)^T d_k \right. \\ &\quad \left. + \frac{1}{2}\alpha^2 M_{\max}^c \left\|d_k\right\|^2\right\} \\ &\leq \max\left\{0, c_j(x_k)\right\} \\ &= \left(1 - \alpha^2\eta\right)\max\left\{0, c_j(x_k)\right\} \\ &= \left(1 - \alpha^2\eta\right) h(x_k).\end{aligned} \tag{14}$$

It proves that $x_k + \alpha d_k$ is acceptable for the filter.

Case 2 ($h(x_k) > 0$). Similarly, when

$$\alpha \leq \min_{j \in J(x_k)}\left\{\frac{-\nabla c_j(x_k)^T d_k}{(1/2)\,M_{\max}^c \left\|d_k\right\|^2 + \eta c_j(x_k)}\right\}, \tag{15}$$

it is easy to learn that

$$\begin{aligned}&h(x_k + \alpha d_k) \\ &\quad \leq \max\left\{0, c_j(x_k) + \alpha \nabla c_j(x_k)^T d_k + \frac{1}{2}\alpha^2 M_{\max}^c \left\|d_k\right\|^2\right\} \\ &\quad \leq \left(1 - \alpha^2\eta\right) h(x_k).\end{aligned} \tag{16}$$

Since x_k is acceptable for the filter, so for all $j \in F^{(k-1)}$, $h(x_k) \leq h(x_j)$ or $f(x_k) \leq f(x_j) - \gamma h(x_j)$ holds. From $x_k + \alpha d_k$ that is not acceptable for the filter, we have

$$h(x_k + \alpha d_k) > \left(1 - \alpha^2\eta\right) h(x_j), \tag{17}$$

$$f(x_k + \alpha d_k) > f(x_j) - \gamma h(x_j) \tag{18}$$

hold. If $h(x_k) \leq h(x_j)$, then

$$h(x_k + \alpha d_k) \leq \left(1 - \alpha^2\eta\right) h(x_k) \leq \left(1 - \alpha^2\eta\right) h(x_j), \tag{19}$$

which contradicts (17). If $f(x_k) \leq f(x_j) - \gamma h(x_j)$, then when $\alpha \leq -\nabla f(x_k)^T d_k / (1/2) M_{\max}^f \|d_k\|^2$, it is easy to learn that

$$\begin{aligned}f(x_k + \alpha d_k) &\leq f(x_k) + \alpha \nabla f(x_k)^T d_k \\ &\quad + \frac{1}{2}\alpha^2 M_{\max}^f \left\|d_k\right\|^2 \leq f(x_k) \\ &\leq f(x_j) - \gamma h(x_j),\end{aligned} \tag{20}$$

which contradicts (18).

Based on the above analysis, we can see that the claim holds. □

By the above statement, we can see that the algorithm is implementable. Now we turn on to prove the global convergence of the algorithm.

Theorem 9. *Let the assumptions hold and $M^f_{min} > 0$. Suppose x^∞ be the cluster point of $\{x_k\}$ generated by algorithm. There exist two possible cases. (i) The iteration terminates at a KKT point. (ii) Any accumulation point of $\{x_k\}$ is a KKT point.*

Proof. we only need to proof case (ii). Since x^∞ is the cluster point generated by algorithm, let $\{x_k\}_{k\in K}$ be any thinner subsequences converging to x^∞.

We will first show that x^∞ is a feasible point. Assume that $h(x_k) \to h(x^\infty) > 0$ for $k \in K$. Let i and j be any two adjacent indices in K where $i < j$. If $h(x^\infty) > 0$, then there exists $k' \in K$ such that for all $i \geq k'$ and because x_j is acceptable to the filter, we have

$$f\left(x_j\right) \leq f\left(x_i\right) - \gamma h\left(x_i\right). \tag{21}$$

Since $\{f(x_k)\}_{k\in K}$ is a monotonically decreasing subsequence for $k \geq k'$ and is bounded below, therefore for $i, j \in K$, $i, j \geq k'$, and $i < j$,

$$\sum_{i,j\in K} \Delta f_{ij} = \sum_{i,j\in K} \left(f\left(x_i\right) - f\left(x_j\right)\right) \tag{22}$$

is bounded above. However, since $f(x_j) \leq f(x_i) - \gamma h(x_i)$, therefore by summing over all indices $i, j \in K$, $i, j \geq k'$, and $i < j$,

$$\sum_{i,j\in K} \Delta f_{ij} \geq \gamma \sum_{i\in K} h\left(x_i\right) \longrightarrow +\infty, \tag{23}$$

which contradicts the fact that $\sum_{i,j\in K} \Delta f_{ij}$ is bounded above. Thus $h(x^\infty) = 0$, hence x^∞ is feasible.

Next we need to show that x^∞ is a KKT point. By the construction of algorithm, there are two cases: one generates the sequence $\{x_k\}$ from $x_{k+1} = x_k + d^0_k$, and the other generates it from $x_{k+1} = x_k + \alpha d_k$. We prove that claim according to the two cases.

Case 1. Suppose that there are infinite points gotten by $x_{k+1} = x_k + d^0_k$. Since $\Delta f_k \geq \sigma \Delta l_k$, we have

$$\begin{aligned} f\left(x_k\right) - f\left(x_k + d^0_k\right) &= -\nabla f\left(x_k\right)^T d^0_k \\ &\quad - \frac{1}{2}\left(d^0_k\right)^T \nabla^2 f\left(y\right) d^0_k \\ &\geq -\sigma \nabla f\left(x_k\right)^T d^0_k. \end{aligned} \tag{24}$$

Thus $\nabla f(x_k)^T d^0_k \leq -(1/2)(d^0_k)^T \nabla^2 f(y) d^0_k/(1-\sigma)$ holds. Since f is bounded below, then

$$\begin{aligned} +\infty &> \sum_{k=0}^{\infty} f\left(x_k\right) - f\left(x_{k+1}\right) \geq -\sum_{k=0}^{\infty} \nabla f\left(x_k\right)^T d^0_k \\ &\geq \frac{1}{2}\sum_{k=0}^{\infty} \frac{\left(d^0_k\right)^T \nabla^2 f\left(y\right) d^0_k}{1-\sigma} \geq \frac{M^f_{\min}}{2\left(1-\sigma\right)} \sum_{k=0}^{\infty} \left\|d^0_k\right\|^2. \end{aligned} \tag{25}$$

Thus $\sum_{k=0}^{\infty} \|d^0_k\|^2 < +\infty$, which means $\|d^0_k\| \to 0$. Since x^∞ is a feasible point, x^∞ is a KKT point.

Case 2. Suppose that there are infinite points gotten by $x_{k+1} = x_k + \alpha d_k$. Since $\Delta f_k \geq \sigma \Delta l_k$, we have

$$\begin{aligned} 0 &= \lim_{k\to\infty} f\left(x_k\right) - f\left(x_k + \alpha d_k\right) \\ &\geq -\lim_{k\to\infty} \sigma\alpha \nabla f\left(x_k\right)^T d_k \geq -\lim_{k\to\infty} \overline{\alpha}\sigma \nabla f\left(x_k\right)^T d_k \geq 0, \end{aligned} \tag{26}$$

which means that $\nabla f(x_k)^T d_k \to 0$. Since

$$\begin{aligned} \nabla f\left(x_k\right)^T d_k &\leq \theta \nabla f\left(x_k\right)^T d^1_k \\ &\leq \theta\left[-\left(d^0_k\right)^T H^{-1}_k d^0_k - \sum_{\lambda^1_{kj}<0} \left(\lambda^1_{kj}\right)^2\right] < 0 \end{aligned} \tag{27}$$

we have $\|d^0_k\| \to 0$ and $\lambda_{kj} \geq 0$, and since x^∞ is a feasible point, x^∞ is a KKT point.

Combined Case 1 and Case 2, we can see that the claim holds. □

4. The Rate of Convergence

In this section, we discuss the convergent rate of the algorithm. We need the following strong assumptions.

(A5) The second-order sufficiently conditions hold, that is, $d^T \nabla^2_{xx} L(x^\infty, \lambda^\infty) d$, for all $d \in \ker \nabla c_{\widehat{J}(x^\infty)} \setminus \{0\}$, where $L(x,\lambda) = f(x) + \lambda^T c(x)$, $c(x) = (c_1(x), \ldots, c_m(x))^T$, $\widehat{J}(x^\infty) = \{j \in J(x^\infty) : (\lambda^\infty)_j > 0\}$, and $(x^\infty, \lambda^\infty)$ is the KKT pair of problem (1).

(A6) Consider $\|(H^{-1}_k - \nabla^2_{xx} L(x^\infty, \lambda^\infty))d^0_k\| = o(\|d^0_k\|)$.

Theorem 10. *Suppose that assumptions (A1)–(A6) hold; then $x_{k+1} = x_k + d^0_k$ for large enough k. Therefore the algorithm is superlinearly convergent.*

Proof. Suppose that x_k is acceptable for the filter; we will show that for large enough k, $x_{k+1} = x_k + d^0_k$ is acceptable for the filter and satisfies the sufficient reduction condition.

First we need to prove that $x_{k+1} = x_k + d^0_k$ is acceptable for the filter. If $h(x_k + d^0_k) \leq (1-\eta)h(x_k)$, then $x_{k+1} = x_k + d^0_k$ is already acceptable for the filter. Else we need to show that $f(x_k + d^0_k) \leq f(x_k) - \gamma h(x_k)$. Let $s_k = f(x_k + d^0_k) - f(x_k) + \gamma h(x_k)$; it holds that

$$\begin{aligned} s_k &\leq \nabla f\left(x_k\right)^T d^0_k + \frac{1}{2}\left(d^0_k\right)^T \nabla^2 f\left(x_k\right) d^0_k \\ &\quad + \frac{\gamma h\left(x_k + d^0_k\right)}{1-\eta} + o\left(\left\|d^0_k\right\|^2\right) \leq \nabla f\left(x_k\right)^T d^0_k \\ &\quad + \frac{1}{2}\left(d^0_k\right)^T \nabla^2 f\left(x_k\right) d^0_k \\ &\quad + \frac{\gamma}{2\left(1-\eta\right)} \sum_{j=1}^{m} \left(d^0_k\right)^T \nabla^2 c_j\left(x_k\right) d^0_k + o\left(\left\|d^0_k\right\|^2\right). \end{aligned} \tag{28}$$

From $\nabla f(x_k)^T d_k^0 = \lambda_k^T c_{J_k} - (d_k^0)^T H_k^{-1} d_k^0$, we have

$$\begin{aligned} s_k \le{}& \lambda_k^T c_{J_k} - \left(d_k^0\right)^T H_k^{-1} d_k^0 + \frac{1}{2}\left(d_k^0\right)^T \nabla^2 f\left(x_k\right) d_k^0 \\ &+ \frac{\gamma}{2(1-\eta)} \sum_{j=1}^{m} \left(d_k^0\right)^T \nabla^2 c_j\left(x_k\right) d_k^0 + o\left(\left\|d_k^0\right\|^2\right). \end{aligned} \tag{29}$$

Since $\lambda_{kj} \ge \epsilon_1$, set $\epsilon_1 = \gamma/(1-\eta)$, and then

$$\begin{aligned} s_k \le{}& \lambda_k^T c_{J_k} - \left(d_k^0\right)^T H_k^{-1} d_k^0 + \frac{1}{2}\left(d_k^0\right)^T \nabla^2 f\left(x_k\right) d_k^0 \\ &+ \frac{1}{2}\sum_{j=1}^{m} \lambda_{kj} \left(d_k^0\right)^T \nabla^2 c_j\left(x_k\right) d_k^0 + o\left(\left\|d_k^0\right\|^2\right) \\ ={}& -\left(d_k^0\right)^T H_k^{-1} d_k^0 + \lambda_k^T c_{J_k} \\ &+ \frac{1}{2}\left(d_k^0\right)^T \nabla_{xx}^2 L\left(x_k, \lambda_k\right) d_k^0 + o\left(\left\|d_k^0\right\|^2\right). \end{aligned} \tag{30}$$

According to $x_k \to x^\infty$, $\lambda_k \to \lambda^\infty \ge 0$, and $c_j(x_k) \to c_j(x^\infty) \le 0$ and assumptions (A2), (A3), and (A5), then

$$\begin{aligned} s_k \le{}& -\frac{a}{2}\left\|d_k^0\right\|^2 \\ &+ \frac{1}{2}\left(d_k^0\right)^T \left(\nabla_{xx}^2 L\left(x_k, \lambda_k\right) - \nabla_{xx}^2 L\left(x^\infty, \lambda^\infty\right)\right) d_k^0 \\ &+ o\left(\left\|d_k^0\right\|^2\right) + \frac{1}{2}\left(d_k^0\right)^T \left(\nabla_{xx}^2 L\left(x^\infty, \lambda^\infty\right) - H_k^{-1}\right) d_k^0 \\ \le{}& -\frac{a}{2}\left\|d_k^0\right\|^2 + o\left(\left\|d_k^0\right\|^2\right) \le 0. \end{aligned} \tag{31}$$

Hence, for large enough k, $x_{k+1} = x_k + d_k^0$ is acceptable for the filter.

Now we are going to show that when k is large enough, $x_{k+1} = x_k + d_k^0$ satisfies the sufficient reduction condition $\Delta f_k \ge \sigma \Delta l_k$. Let $t_k = f(x_k) - f(x_k + d_k^0) - \sigma \Delta l_k$; then we have

$$\begin{aligned} t_k ={}& (\sigma - 1)\left(\lambda_k^T c_{J_k} - \left(d_k^0\right)^T H_k^{-1} d_k^0\right) \\ &- \frac{1}{2}\left(d_k^0\right)^T \nabla^2 f\left(x_k\right) d_k^0 - o\left(\left\|d_k^0\right\|^2\right) \\ \ge{}& (\sigma - 1)\left(\lambda_k^T c_{J_k} - \left(d_k^0\right)^T H_k^{-1} d_k^0\right) \\ &- \frac{1}{2}\left(d_k^0\right)^T \nabla_{xx}^2 L\left(x_k, \lambda_k\right) d_k^0 - o\left(\left\|d_k^0\right\|^2\right). \end{aligned} \tag{32}$$

Since $c_j(x_k) \to c_j(x^\infty) \le 0$ and assumptions (A3), and (A5), then

$$\begin{aligned} t_k \ge{}& (\sigma - 1)\lambda_k^T c_{J_k} - \left(\sigma - \frac{1}{2}\right)\left(d_k^0\right)^T H_k^{-1} d_k^0 \\ &- \frac{1}{2}\left(d_k^0\right)^T \left(\nabla_{xx}^2 L\left(x_k, \lambda_k\right) - \nabla_{xx}^2 L\left(x^\infty, \lambda^\infty\right)\right) d_k^0 \\ &- \frac{1}{2}\left(d_k^0\right)^T \left(\nabla_{xx}^2 L\left(x^\infty, \lambda^\infty\right) - H_k^{-1}\right) d_k^0 - o\left(\left\|d_k^0\right\|^2\right) \\ \ge{}& \frac{a}{2}\left(\frac{1}{2} - \sigma\right)\left\|d_k^0\right\|^2 - o\left(\left\|d_k^0\right\|^2\right) \ge 0. \end{aligned} \tag{33}$$

Hence, for large enough k, $x_{k+1} = x_k + d_k^0$ satisfies the sufficient reduction condition. □

Based on Theorem 10, we can see, when k is large enough that the algorithm will implement the Newton steps and will not change; thus the algorithm is superlinearly convergent.

5. Numerical Test

In this section, we give some numerical results according to our algorithm. We update the matrix H_k by BFGS formulation and the algorithm parameters are set as $H_0 = I \in R^{n\times n}$, $\gamma = 0.1$, $\eta = 0.1$, and $\sigma = 0.01$.

Example 11. One has

$$\begin{aligned} &\min \quad f(x) = 0.1\left\{0.44\frac{x_1^3}{x_2^2} + \frac{10}{x_1} + 0.592\frac{x_1}{x_2^3}\right\} \\ &\text{s.t.} \quad -1 + 8.62\frac{x_2^3}{x_1} \le 0, \end{aligned} \tag{34}$$

where $x_0 = (2.5, 2.5)$, $x^\infty = (1.2867, 0.5305)$, and iterate = 16.

Example 12 (see [8]). Consider

$$\begin{aligned} &\min \quad f(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 \\ &\text{s.t.} \quad 6 - x_1^2 - x_2^2 - x_3^2 - x_4^2 \le 0, \end{aligned} \tag{35}$$

where $x_0 = (2, 2, 2, 2)$, $x^\infty = (1.2247, 1.2247, 1.2247, 1.2247)$, and iterate = 14.

Example 13 (see [10]). One has

$$\begin{aligned} \min \quad & f(x) = -50\left(x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2\right) \\ & \qquad -10.5x_1 - 7.5x_2 - 3.5x_3 - 2.5x_4 \\ & \qquad -1.5x_5 - 10x_6 \\ \text{s.t.} \quad & 6x_1 + 3x_2 + 3x_3 + 2x_4 + x_5 \le 6.5 \\ & 10x_1 + 10x_3 + x_6 \le 20 \\ & 0 \le x_i \le 1,\ i = 1, 2, 3, 4, 5;\ x_6 \ge 0. \end{aligned} \tag{36}$$

$x^\infty = (0, 1, 0, 1, 1, 20)$ is a minimizer with an objective value $f^* = -361.5$. We choose the initial point $x_0 = (1, 1, 1, 1, 1, 10)$, iterate = 6.

Example 14 (see [11]). Consider

$$\begin{aligned} \min \quad & f(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 - 7x_4 \\ \text{s.t.} \quad & x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8 \le 0 \\ & x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 + x_1 - x_4 - 9 \le 0 \\ & 2x_1^2 + x_2^2 + x_3^2 + 2x_4^2 - x_2 - x_4 - 5 \le 0. \end{aligned} \tag{37}$$

We choose the initial point $x_0 = (1, 1, 1, 1)$. $x^{\infty} = (0.2896, 0.9150, 2.1798, 0.6265)$ is a minimizer with an objective value $f^* = -50.1192$, iterate = 40.

Acknowledgment

This research was supported by the National Natural Science Foundation of China (no. 11271128).

References

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

[2] 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.

[3] 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.

[4] C. C. Gonzaga, E. Karas, and M. Vanti, "A globally convergent filter method for nonlinear programming," *SIAM Journal on Optimization*, vol. 14, no. 3, pp. 646–669, 2003.

[5] A. Wächter and L. T. Biegler, "Line search filter methods for nonlinear programming: motivation and global convergence," *SIAM Journal on Optimization*, vol. 16, no. 1, pp. 1–31, 2005.

[6] A. Wächter and L. T. Biegler, "Line search filter methods for nonlinear programming: local convergence," *SIAM Journal on Optimization*, vol. 16, no. 1, pp. 32–48, 2005.

[7] C. M. Chin, "A global convergence theory of a filter line search method for nonlinear programming," *Numerical Optimization Report*, 2002.

[8] K. Su, "A globally and superlinearly convergent modified SQP-filter method," *Journal of Global Optimization*, vol. 41, no. 2, pp. 203–217, 2008.

[9] W. Wang, L.-S. Zhang, and Y.-F. Xu, "A revised conjugate gradient projection algorithm for inequality constrained optimizations," *Journal of Computational Mathematics*, vol. 23, no. 2, pp. 217–224, 2005.

[10] C. A. Floudas, P. M. Pardalos, C. S. Adjiman et al., *Handbook of Test Problems in Local and Global Optimization*, vol. 33, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.

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