Journal of Computational Mathematics, Vol.21, No.5, 2003, 613–624.

COMPOSITE-STEP LIKE FILTER METHODS FOR EQUALITY CONSTRAINT PROBLEMS *1)

Pu-yan Nie
(*LSEC, ICMSEC, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, Beijing 100080, China*)

Abstract

In a composite-step approach, a step s_k is computed as the sum of two components v_k and h_k. The normal component v_k, which is called the vertical step, aims to improve the linearized feasibility, while the tangential component h_k, which is also called horizontal step, concentrates on reducing a model of the merit functions. As a filter method, it reduces both the infeasibility and the objective function. This is the same property of these two methods. In this paper, one concerns the composite-step like filter approach. That is, a step is tangential component h_k if the infeasibility is reduced. Or else, s_k is a composite step composed of normal component v_k and tangential component h_k.

Key words: Composite-step like approaches, Filter methods, Equality constraints, Sequential quadratic programming(SQP) algorithms, Normal component, Tangential component, Convergence.

1. Introduction

In this paper, we consider the problem of minimizing a (linear or nonlinear) function f of n real variables x, which satisfy a set of (linear or nonlinear) constraints $c_i(x) = 0, i = 1, \cdots, m$, namely

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & c(x) = 0, \end{array} \tag{1.1}$$

where $x \in R^n$. The functions $c : R^n \to R^m, m < n, f : R^n \to R$ are assumed to be continuously differentiable. Then, we introduce composite-step like approaches and filter methods respectively.

1.1 Composite-Step Like Methods

An approach, whose every step s_k is consisted of two components v_k and h_k, is termed composite-step method. Where the normal component v_k is to degrade the degree of constraint violation, while the tangential component h_k aims to reduce a model of the merit functions. There are two kinds of composite-step like approaches. One is Vardi-like methods. The other is Byrd-Omojokun-like approaches.

A: Vardi-like methods

From (1.1) one recognizes that the set

$$F_k = \{d | c(x_k) + A(x_k)d = 0 \ \text{ and } \|d\| \le \Delta_k\}, \tag{1.2}$$

may be empty, where $A(x_k) = \bigtriangledown c(x_k)$. Vardi[19] and Byrd, Schnable and Schultz[5] instead relax the linearized constraints so that $\alpha_k c(x_k) + A(x_k)d = 0$ for some $0 < \alpha_k \le 1$ for which

$$F_k(\alpha_k) = \{d | \alpha_k c(x_k) + A(x_k)d = 0 \text{ and } \|d\| \le \Delta_k\}, \tag{1.3}$$

* Received April 4, 2001; final revised March 3, 2002.
1) Supported partially by Chinese NNSF grants 19731010 and the knowledge innovation program of CAS.

is not empty. Clearly, $F_k(0)$ and any $\alpha_k \le \alpha_{max}$ is not empty, where α_{max} is the solution of the following problem.

$$\max_{\alpha\in(0,1]}\min_{\|d\|\le\Delta_k}\|\alpha c(x_k)+A(x_k)d\|=0.$$

Certainly, it may be expensive to find α_{max}. In practice, to obtain the normal component an approximation v_k^c to $v^c(x_k)$ for which $c(x_k)+A(x_k)v=0$ may be computed instead, and α_k subsequently found so that $v_k=\alpha_k v_k^c$ lies in the trust region.

The normal step found, the tangential component is chosen to reduce a model of the merit function. Specially, if we consider a merit function of the form

$$\phi(x,\sigma)=f(x)+\sigma\|c(x)\|. \tag{1.4}$$

Let $m_k(x_k+s)=q(x_k+s)+\sigma m_k^N(x_k+s)$ where

$$q(x_k+s)=f(x_k)+s^Tg(x_k)+\frac{1}{2}s^TH_ks \text{ and } m_k^N(x_k+s)=\|c(x_k)+A(x_k)s\|. \tag{1.5}$$

In a tangential component, the following conditions are satisfied

$$m_k^N(x_k+v_k+h_k)=m_k^N(x_k+v_k) \text{ and } q(x_k+v_k+h_k)<q(x_k+v_k).$$

Thus, the tangential component is obtained by approximately solving the problem.

$$\begin{aligned}\text{minimize}\quad & h^T(g(x_k)+H_kv_k)+\frac{1}{2}h^TH_kh\\ \text{subject to}\quad & A(x_k)h=0\\ & \|h\|\le\Delta_k-\|v_k\|.\end{aligned} \tag{1.6}$$

Of course, the above requirements may readily be satisfied using suitable conjugate-gradient methods.

B: Byrd-Omojokun-like approaches

Byrd-Omojokun-like approach is proposed by Omojokun [14] and Byrd, Gilbert and Nocedal[3]. And it forms ETR, NITRO and BECTR algorithms, which is given by Lalee, Nocedal and Plantenga[13], Byrd, Hribar and Nocedal[4], and Plantenga[15], respectively. In a Byrd-Omojokun method, it is not required that the linearized constraints should be compatible. That is, the main difference lies in the computation of the normal step. Instead of shifting the linearized constraints, to obtain v_k one solves the following subproblem approximately.

$$\begin{aligned}\text{minimize}\quad & \|c(x_k)+A(x_k)v_k\|\\ \text{subject to}\quad & \|v_k\|\le\xi^N\Delta_k,\end{aligned} \tag{1.7}$$

for some $0<\xi^N<1$.

Apparently, (1.7) may have many solutions. Obtaining an exact solution to (1.7) may be costly. A cheaper choice is to calculate an approximate solution giving a reduction in $\|c(x_k)+A(x_k)v\|$ no worse than a fraction of Cauchy point for this problem, which is,

$$v_k^c=-\alpha_k^cA(x_k)^Tc(x_k), \tag{1.8}$$

where

$$\alpha_k^c=\arg_{0\le\alpha\le\xi^N\Delta_k/\|A(x_k)^Tc(x_k)\|}\min\|c(x_k)-\alpha A(x_k)A(x_k)^Tc(x_k)\|.$$

At every step, (1.8) is satisfied if suitable conjugate-gradient method is used. Meanwhile, (1.5) is also met.

For Vardi-like approaches and Byrd-Omojokun-like methods, the superlinear convergence is obtained under suitable assumptions.

1.2. Filter Technique

Filter approach is proposed by Fletcher and Leyffer[8]. It has been applied extensively so far. In [9], filter method is used to SLP(sequential linear programming) and its global convergence to first order critical point is shown. In 1999, filter method is combined with SQP by Fletcher, Leyffer and Toint[10]. In[12], filter method is used to bundle nonsmooth approach. Audet and Dennis[1] present a pattern search filter method for derivative-free nonlinear programming. In 2000, Ulbrich et al.[18] use interior-point filter method to nonconvex programming. Furthermore, the filter idea has proved to be very successful numerically in the SLP/SQP framework[8].

Filter methods have several advantages over penalty function methods. A penalty parameter estimate, which could be problematic to obtain, is not required. Practical experience shows that they exhibit a certain degree of nonmonotonicity which can be beneficial. Then, it is introduced as follows.

Our purpose is to minimize both the objective function f and a nonnegative continuous constraint violation function p where $p(x) \geq 0$, $p(x) = 0$ if and only if x is feasible ($p(x) > 0$ if and only if x is infeasible). The filter will be used as a criterion for accepting or rejecting a step generated by subproblem.

Fletcher et al.'s definition of filter is based on the definition of dominance, which is originated from multiobjective terminology. The definition of dominance is:

Definition 1. *For a pair of ω, ω' with finite components. ω dominates ω', written $\omega \prec \omega'$ if and only if $\omega_i \leq \omega'_i$ for each i and $\omega \neq \omega'$.*

Similarly, we use $\omega \preceq \omega'$ to indicate that either $\omega \prec \omega'$ or that $\omega = \omega'$, which is the notion of dominance in earlier filter papers. Combined with our problem, we define $x \prec_{(p,f)} x'$ if and only if $(p(x), f(x)) \prec (p(x'), f(x'))$ where " $\prec$ " is listed in the Definition 1. In order to simplify the terminology, we use $x \prec x'$ rather than $x \prec_{(p,f)} x'$. As above, $x \preceq x'$ indicates that either $x \prec x'$ or equivalent. A filter is defined as follows:

Definition 2. *A filter $\mathcal{F}$ is a set of points in R^n such that for any $x, x' \in \mathcal{F}$. $x \prec x'$ is not true.*

To acquire the convergence, stronger conditions are required to decide whether to accept a point to the filter or not. There are diverse rules in different papers. We give a rule combining their definitions. A point x' is filtered, if it satisfies:

$$x' \in \bar{\mathcal{F}} = \bigcup_{x \in \mathcal{F}} \{x' : \{x \prec x'\} \text{ or } (p(x') = 0, f(x')) \succeq R(p^I, f^I, f^{\mathcal{F}})\},$$

where R is a monotonic increasing function to $p^I, f^I, f^{\mathcal{F}}$ and

$$f^{\mathcal{F}} = \min_{x \in \mathcal{F}} \{f(x) : p(x) = 0\}, \tag{1.9}$$

$$f^I = \min_{x \in \mathcal{F}} \{f(x) : p(x) > 0\}, \tag{1.10}$$

$$p^I = \min_{x \in \mathcal{F}} \{p(x) : p(x) > 0\}. \tag{1.11}$$

The constraint violation function is defined as follows

$$p_{k+1} = c(x_k + s_k)^T c(x_k + s_k). \tag{1.12}$$

Certainly, in an algorithm, filtered point is rejected by the set $\mathcal{F}$. This means that the trial step does not produce a successful iteration. On the contrary, unfiltered points are accepted. It is pointed out that the filter points are related to when it is generated. In this paper, a point is acceptable to the filter (unfiltered conditions) if it satisfies:

$$\text{either } p_{k+1} \leq \beta p_j \text{ or } f_{k+1} + \gamma p_{k+1} \leq f_j, \tag{1.13}$$

for all $j \in \mathcal{F}_k$ where $0 < \gamma < \beta < 1$ are constants. To obtain the convergent results, we use (1.13). There are some reduction to f or p to guarantee the convergence properties.

In practice, to obtain good properties some additional conditions are necessary. For example, Fletcher and Leyffer[8] define the "envelope" and Audet and Dennis use "poll search" in[1].

The paper is organized as follows: In Section 2 the algorithm is presented. In Section 3 the convergence properties are given. Some numerical results and remarks are given in Section 4.

2. Motivation

The properties of the composite-step approach and filter method motivate us to combine them together. Furthermore, Yuan[20] points out the shortcoming of null-space methods. That is, when the two side reduced Hessian matrix is replaced by a quasi-Newton update, the method in range step is a Newton step while the approach in null-space is quasi-Newton step. Just as the null-space method, the composite-step method has the same disadvantage. In the same case, the normal step is Newton step while the tangential one is quasi-Newton step. In [20], Yuan uses more null-space steps to overcome this unbalance. But, the optimal number of null-space steps is flexible because of the difference of quasi-Newton methods. Meanwhile, in some case, the tangential component increases the constraint violation. In this case, it is not necessary to adopt a normal component with several tangential components.

Therefore, the composite-step like filter is brought out. Our method is: the tangential step is computed if the constraint violation is reduced. Or else, a restoration step is presented to make the value of constraint violation reduced, which is the idea of filter method. Certainly, we can use the normal step as a restoration step.

A straightforward way to solve equality constraint problems is to add the trust-region constraint to the (QP) subproblem to restrict the size of the step. That is, at each step, we solve the following trust-region subproblem

$$\begin{array}{ll} \text{minimize} & q_c(s) = s^T g_c + \frac{1}{2} s^T H_c s + f_c \\ \text{subject to} & A_c s + c_c = 0 \\ & \|s\| \leq \Delta_c. \end{array} \tag{2.1}$$

However, as pointed out in section 1, this problem may be inconsistent. In this paper, we take the technique of Vardi's or Byrd-Omojokun's to deal with it. But there is difference in dealing with the trust region radius. Namely, the following subproblem is dealt with:

$$\begin{array}{ll} \text{minimize} & q_c(s) = s^T g_c + \frac{1}{2} s^T H_c s + f_c \\ \text{subject to} & A_c s = 0 \\ & \|s\| \leq \Delta_c. \end{array} \tag{2.2}$$

Then, to compare the reduction of model with the actual reduction, we denote $ared(s_k) = f(x_k + s_k) - f(x_k)$ and $pred(s_k) = q_k(s_k) - f(x_k)$. If

$$r_k = \frac{ared(s_k)}{pred(s_k)} > \eta_1 > 0, \tag{2.3}$$

where η_1 is constant and $x_k + s_k$ satisfied the unfiltered conditions (1.13), then the trial point is accepted. That is, $x_{k+1} = x_k + s_k$. Namely, if (2.3) and (1.13) are satisfied, $x_k + s_k$ is acceptable to the filter $\mathcal{F}_k$.

Then, the next trial step is considered. In our algorithm, our major object is to guarantee the feasibility. Thus, the following inequality is the rule to decide the way of next trial step.

$$p_{k+1} \leq \eta_3 \min\{\mu, \alpha_1 \Delta_k^{2+\alpha_2}\}. \tag{2.4}$$

where η_3 and μ are positive constants. If (2.4) is satisfied, the tangential component is still used in the next step. Or else, the next step is obtained by restoration algorithm. we get s by solving the following subproblem:

$$\begin{array}{ll} \text{minimize} & \|\alpha c(x_k) + A(x_k)s_k\| \\ \text{subject to} & \|s_k\| \leq \Delta_k, \end{array} \tag{2.5}$$

where α satisfies the conditions in section 1. Namely, the normal step of Vardi's is used. Certainly, we can also use Byrd-Omojukun like approach to get the next iteration under the conditions analyzed above

$$\begin{array}{ll} \text{minimize} & \|c(x_k) + A(x_k)s_k\| \\ \text{subject to} & \|s_k\| \leq \xi^N \Delta_k, \end{array} \tag{2.6}$$

for some $0 < \xi^N < 1$. Then, the accepting criterion is listed as follows for the restoration algorithm. That is $p(x_k + s_k) \leq \eta_3 \min\{p_k^I, \alpha_1 \Delta_k^{1+\alpha_2}\}$, where $0 < \eta_3 < 1$. After a restoration algorithm, the trust region radius is kept unchanged. Of course, the sufficient reduction is required

$$\|c(x_k)\|^2 - \|c(x_k) + A(x_k)s_k\|^2 \geq u_1 \|c(x_k)\| \min\{u_2 \|c(x_k)\|, \xi^N \Delta_k\}, \tag{2.7}$$

where u_1, u_2 are positive constants independent of k. Then, a restoration algorithm is listed as follows. In brief, a restoration algorithm generates a point which has the following properties

(1) $p(x_k + s_k) \leq \eta_3 \min\{p_k^I, \alpha_1 \Delta_k^{2+\alpha_2}\}$ where $0 < \eta_3 < 1$.

(2) the trust region radius is not changed.

(3) $x_k + s_k$ is acceptable to the filter.

In the restoration algorithm, we define the criterion of estimate the trial point.

$$r_k^j = \frac{\|c(x_k^j)\| - \|c(x_k^j + s_k^j)\|}{\|c(x_k^j)\| - \|c(x_k^j) + A(x_k^j)s_k^j\|}.$$

Algorithm 1. *Restoration Algorithm*

0. Let $x_k^0 := x_k, \Delta_k^0 := \Delta_k, \alpha_1, \alpha_2 \in [0,1], j := 0, \eta_2, \eta_3 \in (0,1)$.

1. If $p(x_k^j) \leq \eta_3 \min\{ p_k^I, \alpha_1 \Delta_k^{2+\alpha_2}\}$ *and* x_k^j *is acceptable to the filter, then let* $x_k^r := x_k^j$ *and stop.*

2. Compute

$$\begin{array}{ll} \textit{minimize} & \|c(x_k^j) + A(x_k^j)s_k^j\| \\ \textit{subject to} & \|s_k^j\| \leq \Delta_k^j \end{array}$$

to get s_k^j. *Calculate* r_k^j.

3. If $r_k^j \leq \eta_2$, *then let* $x_k^{j+1} := x_k^j, \Delta_k^{j+1} := \frac{1}{2}\Delta_k^j, j := j+1$ *and goto step 2.*

4. $x_k^{j+1} := x_k^j + s_k^j, \Delta_k^{j+1} := 2\Delta_k^j, j := j+1$. *Compute* A_k^{j+1} *goto step 1.*

As for the update of the trust region radius, it observes the following rules: when the trial step is "good", the trust region radius is increased. Or else, it is not changed or reduced. Then we list in a algorithm.

Algorithm 2. *The Update of Trust-region Radius*

0. Give $v_1 \in (0.25, 0.75], v_2 \in [1.0, 1.25), v_3 \in [1.25, 2.0]$ *are constant.*

1. If $x_k + s_k$ *is rejected by the filter, then*

$$\Delta_{k+1} = v_1 \Delta_k;$$

2. If s_k *is obtained by (2.5) or (2.6), then*

$$\Delta_{k+1} = \Delta_k;$$

3. If s_k *is obtained by (2.2) and (2.4) is satisfied, then*

$$\Delta_{k+1} = v_3 \Delta_k;$$

4. If s_k *is obtained by (2.2) and (2.4) is not satisfied,then*

$$\Delta_{k+1} = v_2 \Delta_k.$$

Then, our algorithm is presented as follows.

Algorithm 3. *Composite-Step-Like Filter Algorithm*

Step 0: Choose $v_1 \in (0.25, 0.75], v_2 \in [1.0, 1.25), v_3 \in [1.25, 2.0]$, $x_0, \Delta_0 > 0, \alpha_1, \alpha_2 \in [0,1], \eta_2, \eta_3 \in (0,1), \eta_2 > 0$. *Set* $k := 0, \mathcal{F}_0 = \{x_0\}$. *Compute* p_0 *by (1.12).*

Step 1: Compute p_k^I, f_k^I, f_k^F.

Step 2: Compute s_k *by (2.2). If* $s_k \neq 0$ *goto Step 3. If* $p_k = 0$ *then stop. Or else, goto Step 7.*

Step 3: Compute $r_k = \frac{ared(s_k)}{pred(s_k)}$.

Step 4: If $r_k < \eta_1$, *then* $x_{k+1} := x_k$ *goto Step 8.*

Step 5: Compute $f(x_k + s_k)$ *and* $\hat{p}_k$. *And decide whether* $x_k + s_k$ *is acceptable to the filter. If it is not, go to Step 8. Or else,* $x_{k+1} = x_k + s_k$. *Then*

(1)let x_{k+1} *enter the filter and* $p_{k+1} = \hat{p}_k, f_{k+1} = f(x_k + s_k)$.

(2)remove the point dominated by x_{k+1}, *update the trust region radius and* H_k.

Step 6: If $p_{k+1} \leq \eta_3 \min\{\mu, \alpha_1 \Delta_k^{2+\alpha_2}\}$ *then* $k := k+1$ *goto Step 1. Or else,* $k := k+1$ *goto Step 7.*

Step 7: Use Restoration Algorithm to get the point $x_k^r = x_k + s_k^r$, *where* s_k^r *is obtained by Restoration Algorithm, then let* $x_k = x_k^r$ *and goto Step 2.*

Step 8: Using Algorithm 2 to update the trust region radius. Then, $k := k+1$ *and goto Step 2.*

The composite-step-like filter methods have several advantages over composite-step-like approaches.

(1) If the approximate Hessian is used, then normal step is obtained by Newton method, while the tangential step is quasi-Newton step. Thus, the convergence rates are inconsistent. Combined with filter technique, the shortcoming is overcome to a certain degree.

(2) As for the composite-step-like methods, if it fails in some step, we have to recompute v_k and h_k. That is, some useful information is lost. While composite-step-like filter methods avoids this.

(3) It is an infeasible method. The difficulty to find a feasible point, which is thought to be as difficult as solving the problem, is avoided.

(4) Finally, this is a nonmonotonic method which is beneficial to a certain degree. When the restoration step is used, the objective function value may be increased.

Furthermore, in filter technique, to obtain the optimization point we pay more attention to the value of constraint violation.

3. The Convergence of the Algorithm

Just as [1,8,9,10,18], our analysis of the convergence properties to Algorithm 3 is based on the assumption as follows. Meanwhile, the sufficient reduction condition plays a key role in getting the convergence.

Assumption 1.

(1) The set $\{x_k\} \in X$ *is nonempty and bounded.*

(2) The function $f(x)$ *and* $c(x)$ *are twice continuously differentiable on an open set containing* X.

(3) The matrix sequence $\{B_k\}$ *is bounded.*

(4) When (2.2) is solved, we have $q_k(0) - q_k(s_k) \geq \beta_0 \|\hat{g}_k\| \min\{\|\hat{g}_k\|, \Delta_k\}$ *where* $\beta_0 > 0$ *is fixed and* $\hat{g}_k = Z_k^T g_k$, *where* $A_k Z_k = 0$ *and* $Z_k^T Z_k = I$. *For (2.6), we have* $\|c_k\| - \|A_k s_k^n + c_k\| \geq u_2 \min\{\|c_k\|, \xi^N \Delta_k\}$.

(1) and (2) are the standard assumption. (3) plays important role to obtain the convergence result. But it has minor effect to obtain local convergence rate. (4) is the sufficient reduction

conditions, which is reasonable because many algorithms satisfy these conditions because of (3). For example, Cauchy step satisfies it. In trust region method, it guarantees the global convergence. The following results are based on these assumptions.

As for the restoration algorithm, it should be terminated finitely. Otherwise, it is impossible to make our algorithm terminate finitely. Then, we give this conclusion as follows.

Lemma 1. *The restoration algorithm terminates finitely.*

Proof. If $p_k^j \to 0$, the result is true clearly. Then we consider the cases when $p_k^j \to 0$ is not true. That is, there exists $\epsilon > 0$ and $p_k^j > \epsilon$ for all j. Denote

$$K = \{j | r_k^j = \frac{\|c(x_k^j)\| - \|c(x_k^j + s_k^j)\|}{\|c(x_k^j)\| - \|c(x_k^j) + A(x_k^j)s_k^j\|} > \eta_2 > 0\}. \tag{3.1}$$

From the above set K, (2.6) and the Assumption 1, we have

$$+\infty > \sum_{j=1}^{\infty}((p_k^{j-1})^{1/2} - (p_k^j)^{1/2}) \geq \sum_K \eta_2(\|c(x_k^j)\| - \|c(x_k^j) + A(x_k^j)s_k^j\|)$$
$$\geq \eta_2 \sum_K u_2 \min\{\|c(x_k)\|, \xi^N \Delta_k^j\}.$$

Therefore, $\Delta_k^j \to 0$ for $j \in K$. From the restoration algorithm, $\Delta_k^j \to 0$ for all j. On the other hand, it is apparent that

$$\|c(x_k^j)\| - \|c(x_k^j + s_k^j)\| = \|c(x_k^j)\| - \|c(x_k^j) + A(x_k^j)s_k^j\| + o(\Delta_k^j),$$

when $\Delta_k^j \to 0$. Thus, according to the algorithm, the trust region radius Δ_k^j will be increased. That is, $\Delta_k^{j+1} > \Delta_k^j$, which contradicts with our assumption. When (2.5) is used, in the same way do we obtain the result. Therefore, the result is true.

For (1.13), the left hand inequality is an apparent way to define a sufficient reduction to p, while the second inequality is to obtain a sufficient reduction to f. In such a way does it guarantee the iterates toward feasibility.

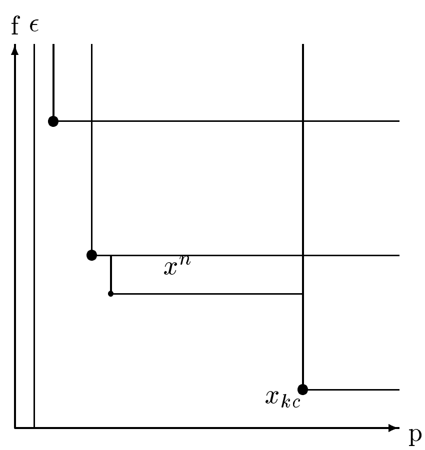

Figure 1

Lemma 2. *Suppose there are infinitely many points added to the filter. Then*

$$\lim_{i\to\infty} p(x_i) = 0. \tag{3.2}$$

Proof. If the theorem were not true, there would have infinite components in K_1, which is defined as follows.

$$K_1 = \{k|p_k > \epsilon\},$$

Because of the Assumption 1 we assume that $|f_k| < M$ for all k without loss of generality, where M is a positive constant. Then we analyze with two cases.

(1) If $min_{i\in K_1}\{f_i\}$ exists. Let $f_{kc} = \min_{i\in K_1}\{f_i\}$. And p_{kc} is the corresponding value related to (1.12). Then, according to the definition of the filter, the other components, which lie behind x_{kc} in the filter, satisfy:

$$p_k \leq p_{kc}, \text{ and } f_k \geq f_{kc}.$$

Then, all the filter points, which enter the filter behind x_{kc}, can be covered with a square, whose area is no more than $2Mp_{kc}$. We consider the area lies to the south-west of the filter in this square. When a new point x_{kc} enters the filter, the next point x_{kc+1} should lies south-west of the points in the filter $\mathcal{F}_{kc}$. and the area which lies south-west of the $\mathcal{F}_{kc+1}$ in the square is smaller than that of $\mathcal{F}_{kc}$. Therefore, we think that the area is reduced if a new point enter the filter. If a new point enters K_1 of the filter, the area of this square, more than $(1-\beta)\gamma\epsilon^2$, will be reduced (When a point is added to the filter, its p is less than every point, which lies to the left of this point, to more than $(1-\beta)\epsilon$. its f is less than every point, which lies to the right of this point, to more than $\gamma\epsilon$. Therefore, the area of this square, more than $(1-\beta)\gamma\epsilon^2$ will be reduced). (See Figure 1, there is a symbol 'x^n' inside this area which is reduced.) Thus, the area will be reduced to 0 after finite times. When the area is zero, it means that a point can not enter K_1, which is contradicted with the infiniteness of K_1.

(2) If $min_{i\in K_1}\{f_i\}$ doesn't exist. From the conditions in this Lemma, let $f_c = \inf_{i\in K_1}\{f_i\}$. From the definition of inf there exists $f_{kc} \geq f_c$ and $f_{kc} \leq (f_c + \gamma\epsilon)$. Then, according to the definition of the filter, the other components, which lie behind x_{kc} in the filter, satisfy:

$$p_k \leq p_{kc}, \text{ and } f_k \geq (f_{kc} - \gamma\epsilon).$$

Using the same techniques as that in (1), the result is gotten.

Thus, the conclusion is obtained.

As for the case of finitely many points added to the filter, it is apparent that the following result be true.

Lemma 3. *Suppose there are finitely many points added to the filter. Then $p(x_k) = 0$ where k is the number of points added to the filter.*

Proof. The result is apparent from the terminating condition of Step 2 in Algorithm 3.

Then, the global convergence of Algorithm 3 is obtained under Assumption 1 as follows. In the proof process, the matrix Z_k is appeared, where $A_kZ_k = 0$ and $Z_k^TZ_k = I$. By the way, Z_k is not unique. But it is not appeared in our algorithm. Similar to Dennis et al's, we also make some assumptions to the normal component.

Assumption 2. *A_k has full rank and $(A_kA_k^T)^{-1}$ is bounded.*

Under Assumption 2, the following results are reasonable for some algorithms to solve (2.6).

$$\|s_k^n\| \leq k_1\|c_k\|, \tag{3.3}$$

and

$$\|c_k\|^2 - \|A_ks_k^n + c_k\|^2 \geq k_2\|c_k\|\min\{k_3\|c_k\|, \Delta_k\}. \tag{3.4}$$

where k_1, k_2 and k_3 are positive constants independent of k. Furthermore, several ways of computing s_k^n satisfy (3.3) and (3.4) in [6]. For example, conjugate-gradient method of Steihaug and Lanczos bidiagonalizatioon approaches do. Meanwhile, the Newton method is used to compute $c(x)^Tc(x) = 0$. Without loss of generality we assume $\|c(x_k + s_k^n)\| \leq k_5\|c(x_k)\|$ where $k_5 \in (0, 1)$.

Theorem 1. *Under Assumption 1, (3.3) and (3.4), then*

$$\lim_{k\to\infty}\inf(\|\hat{g}_k\|+\|p_k\|)=0, \tag{3.5}$$

where $\hat{g}_k = Z_k^T g_k$, $A_k Z_k = 0$ *and* $Z_k^T Z_k = I$.

Proof. From Assumption 1, we have that $\hat{g}_k, g_k$ are bounded. We prove the theorem by contradiction. If (3.5) were not true, there would exist

$$\|\hat{g}_k\|+\|p_k\|>\epsilon, \tag{3.6}$$

for all k.

If the algorithm terminates finitely, (3.5) is true apparently. Then, we assume that the algorithm terminates infinitely.

If there are infinitely or finitely many steps to enter the filter, from Lemma 3 and Lemma 2 we have $p_j \to 0$. Then $\|\hat{g}_k\| > \epsilon$ for all sufficient large k. Firstly, we analyze the Restoration Algorithm. We consider every normal component s_k^n.

$$q_k(0)-q_k(s_k^n) = -[\frac{1}{2}(s_k^n)^T B_k s_k^n + g_k^T s_k^n] \geq -k_1\|c_k\|(\|g_k\|+\|B_k\|\|s_k^n\|) \geq -k_4\|c_k\|.$$

where k_4 is positive constant independent of k. Let $\|c(x_k+s_k^n)\| = \|c_k^1\|$. If the terminate condition of the restoration algorithm is satisfied after t normal steps, then

$$\begin{aligned}
q_k(0)-q_k(s_k^r) &= q_k(0)-q_k((s_k^n)^1+\cdots+(s_k^n)^t) = \\
&-\frac{1}{2}((s_k^n)^1+\cdots+(s_k^n)^t)^T B_k((s_k^n)^1+\cdots+(s_k^n)^t) - g_k^T((s_k^n)^1+\cdots+(s_k^n)^t) \\
&\geq -k_1(\|c_k\|+\|c_k^1\|+\cdots+\|c_k^{t-1}\|)[\|g_k\|+\|B_k\|(\|c_k\|+\|c_k^1\|+\cdots+\|c_k^{t-1}\|)] \\
&\geq -k_1(\|c_k\|+k_5\|c_k\|+\cdots+k_5^{t-1}\|c_k\|)[\|g_k\|+\|B_k\|(\|c_k\|+k_5\|c_k\|+\cdots+k_5^{t-1}\|c_k\|)] \\
&\geq -\frac{k_1}{1-k_5}\|c_k\|(\|g_k\|+\|B_k\|\|c_k\|) \\
&\geq -4k_4\|c_k\|,
\end{aligned}$$

where k_4 is a positive constant. Certainly, the above relation is also true if there is no normal component in some step. We denote the tangential component in the k-th step as s_k^t. Then we consider every successful step. If $\|\hat{g}_k\| > \epsilon$ were true, we would have

$$q_k(s_k^r) - q_k(s_k^r + s_k^t) \geq \beta_0 \epsilon \min\{\epsilon, \Delta_k\}. \tag{3.7}$$

Thus,

$$q_k(0) - q_k(s_k^r + s_k^t) \geq \beta_0 \epsilon \min\{\epsilon, \Delta_k\} - 4k_4\|c_k\|. \tag{3.8}$$

On the other hand,

$$\begin{aligned}
\infty > \sum_{k=1}^{\infty}[f(x_{k-1})-f(x_k)] &\geq \eta_1 \sum_{k=0}^{\infty}[q_k(0)-q_k(s_k^r+s_k^t)] \\
&\geq \eta_1 \sum_{k=0}^{\infty}(\beta_0\epsilon\min\{\epsilon,\Delta_k\} - 4k_4\|c_k\|).
\end{aligned}$$

Thus, we have $\Delta_k \to 0$ because $\|c_k\| \to 0$.

On the other hand, when Δ_k and $\|s_k^r\|$ are small enough, from Taylor expansion we have

$$f(x_k) - f(x_k+s_k^r+s_k^t) = q_k(0) - q_k(s_k^r+s_k^t) + o(\|\Delta_k\|).$$

We consider $r_k \geq \eta_1$ when k is large enough. If $s_k^r = 0$ then $f(x_k) - f(x_{k-1}) > 0$ from (3.7). If $s_k^r \neq 0$, then $\eta_3\alpha_1\Delta_k^{2+\alpha_2} \leq h_k \leq 64\eta_3\alpha_1\Delta_k^{2+\alpha_2}$, or $(\eta_3\alpha_1)^{\frac{1}{2}}\Delta_k^{1+\frac{\alpha_2}{2}} \leq \|c_k\| \leq 8(\eta_3\alpha_1)^{\frac{1}{2}}\Delta_k^{1+\frac{\alpha_2}{2}}$. (We will show this result.) From (3.8) we have

$$f(x_k) - f(x_k + s_k^r + s_k^t) \geq \eta_1(q_k(0) - q_k(s_k^r + s_k^t))$$

$$\geq \eta_1(\beta_0\epsilon\min\{\epsilon, \Delta_k\} - 4k_4\|c_k\|) \geq \eta_1(\beta_0\epsilon\min\{\epsilon, \Delta_k\} - 32k_4(\eta_3\alpha_1)^{\frac{1}{2}}\Delta_k^{1+\frac{\alpha_2}{2}}) > 0,$$

when $\Delta_k \leq \min\{\epsilon, (\frac{\beta_0\epsilon}{32k_4(\eta_3\alpha_1)^{\frac{1}{2}}})^{\frac{2}{\alpha_2}}\}$. Thus, every trial step will be accepted by the filter. Therefore, the trust region radius will not be reduced if k is large enough, which is contradicted with $\Delta_k \to 0$. Thus, the conclusion is met.

Factually, $\|c(x_k + s_k^n)\| \leq k_5\|c(x_k)\|$ can be obtained from (3.4) where $k_5 \in (0, 1)$ is constant. Thus, the result is reasonable.

From the above analysis, we get the global convergent property under Assumption 1. The following results are deduced immediately:

Theorem 2. *Under Assumption 1, Assumption 2, if $\{x_k\}$ is a bounded sequence, then x_k has an accumulation point satisfying the first-order KKT conditions.*

Proof. From the Assumption 1, Assumption 2 and Z_k, there exists a subsequence k_j such that $\lim_{j\to\infty} Z_{k_j}^T g_{k_j} = 0$. Thus, when j is large enough, g_{k_j} almost lies in the subspace which is generated by the columns of $A_{k_j}^T$. Namely, there exists a vector λ_{k_j} which satisfies

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

Therefore, our result is obtained because $\{x_k\}$ is a bounded.

Now, we show $\eta_3\alpha_1\Delta_k^{2+\alpha_2} < h_k \leq 64\eta_3\alpha_1\Delta_k^{2+\alpha_2}$.

Theorem 3. *If $\Delta_k \to 0$, k is large enough, $\|\nabla^2 c_i\| \leq M_2$ for $i = 1, 2, \cdots, m$ and $s_k^r \neq 0$ where M_2 is a positive constant, then,*

$$\eta_3\alpha_1\Delta_k^{2+\alpha_2} \leq h_k \leq 64\eta_3\alpha_1\Delta_k^{2+\alpha_2}. \tag{3.9}$$

Proof. The first part is obvious. We show the second part. If $x_k := x_{k-1}$, then $h_k = h_{k-1}$ and $\Delta_k \geq \frac{1}{4}\Delta_{k-1}$. Thus, $h_k = h_{k-1} \leq \eta_3\alpha_1\Delta_{k-1}^{2+\alpha_2} \leq 64\eta_3\alpha_1\Delta_k^{2+\alpha_2}$.

If $x_k \neq x_{k-1}$, then, from Taylor expansion and our algorithm we have $\Delta_{k-1} \leq \Delta_k$, $A(x_{k-1} + s_{k-1}^r)(s_{k-1} - s_{k-1}^r) = 0$ and

$$\|c(x_k)\| \leq \|c(x_{k-1} + s_{k-1}^r) + A(x_{k-1} + s_{k-1}^r)(s_{k-1} - s_{k-1}^r)\| + 2mM_2\|s_{k-1} - s_{k-1}^r\|$$

$$\leq (\eta_3\alpha_1)^{\frac{1}{2}}\Delta_{k-1}^{1+\frac{\alpha_2}{2}} + 2mM_2\Delta_{k-1}^2 \leq (\eta_3\alpha_1)^{\frac{1}{2}}\Delta_k^{1+\frac{\alpha_2}{2}} + 2mM_2\Delta_k^2.$$

Thus,

$$h_k \leq ((\eta_3\alpha_1)^{\frac{1}{2}}\Delta_k^{1+\frac{\alpha_2}{2}} + 2mM_2\Delta_k^2)^2 \leq 64\eta_3\alpha_1\Delta_k^{2+\alpha_2}$$

if $\Delta \leq (\frac{7(\eta_3\alpha_1)^{\frac{1}{2}}}{2mM_2})^{\frac{2}{2-\alpha_2}}$. Therefore, the result holds.

4. Numerical Results and Remarks

Algorithm 3 is also a kind of SQP filter method. For our algorithm, some promising numerical results are obtained. Especially, when the initial point is infeasible, our algorithm works efficiently. Some numerical results are listed in Table 1.

Table 1

Problem	n	m	NF	NG	Iter	L's NF	L's NG
HS6	2	1	20	11	10	30	13
HS7	2	1	10	7	6	8	7
HS8	2	2	4	4	4	6	6
HS42	4	2	11	5	3	5	5
BT2	3	1	7	6	5	13	13

In Table 1, L's is Lalee et al's result in[13]. We compared our results with that in[13] because the methods in [13] is composite step approaches. Meanwhile, they all use exact Hessian. But our exact Hessian is different from that in [13]. Iter, NF and NG means numbers of iterations, function evaluations and gradient evaluations respectively. The error tolerance is 10^{-7}. In our algorithm, $\beta = 0.98, \gamma = 0.05$.

When the Hessian is updated by BFGS approaches, some results are reported in Table 2, which is promising. Lalee et al's results are obtained by l-BFGS techniques. HS8 has no Hessian for Lalee et al's method.

Table 2

Problem	n	m	NF	NG	L's l-BFGS NF	L's NG
HS8	2	2	4	4	—-	—
HS39	4	2	9	9	26	19
HS60*	3	1	8	4	13	11
HS100*	10	3	30	15	30	24
BT8	5	2	7	6	13	13

*Test problem was created specifically for equality constrained problem.

From Table 1-2, our algorithm needs fewer steps than that of Lalee et. al's. Hence, our algorithm does very well because the constraints and objetive function are balanced. Further, our algorithm is more flexible.

On the other hand, filter method is a nonmonotonic method which may be beneficial. In a filter algorithm, we consider a way to avoid the close relation between f and p. In this way is the algorithm more flexible. About the parameter β, γ where $1 > \beta > \gamma > 0$. β is required to close enough to 1. While γ sufficiently close to 0. There are several advantages: (1) The acceptable criterion is not very strong. (2) The relation of f and p becomes more implicit. Certainly, we can modify some conditions in the algorithm to get the global convergence.

By the way, there are no locally superlinear result to filter methods till now. For our algorithm, superlinear result is not obtained yet. It is still a problem. As for the norm of the constraint violation function, the other forms can be chosen instead, such as l_1-norm or l_2-norm. The method is similar.

Acknowledgment. The author thanks Prof. Yuan Ya-xiang for his carefully reading and some constructive suggestions.

References

[1] C. Audet and J.E. Dennis J., Combining pattern search and filter algorithms for derivative free optimization, TR00-9 Department of Computational and Applied Mathematics, Rice University, Houston, TX, 2000.

[2] M.S. Bazara and C.M. Shetty, Nonlinear Programming theory and algorithms, John Wiley and Sons (1979).

[3] R.H. Byrd, J.Ch. Gilbert and J. Nocedal, A trust region method based on interior point techniques for nonlinear programming, *Math. Prog.,* **89** (2000), 149-185.

[4] R.H. Byrd, M.E.Hribar and J. Nocedal, A interior point algorithm for large scale nonlinear programming, *SIAM J. Optimization,* **9** (1998), 877-900.

[5] R.H. Byrd, R.B. Schnabel and G.A. Schultz, A trust region algorithm for nonlinearly constrained optimization, *SIAM Journal on Numerical Analysis,* **24** (1987), 1152-1170.

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

[7] J.E. Dennis Jr,R.B. Schnable, Numerical methods for unconstrained optimization and nonlinear equations, Prentice-Hall, Englewood Cliffs, NJ.

[8] R. Fletcher and S. Leyffer, Nonlinear programming without a penalty function, *Math. Prog.,* **91**:2 (2002), 239-269.

[9] R. Fletcher, S. Leyffer and P.L. Toint, On the global convergence of an SLP-filter algorithm. Tech. Report. NA/183, Department of Mathematics, University of Dundee, 1998.

[10] R. Fletcher and S. Leyffer, User manual for filter SQP. Tech. Report, NA/181, Department of Mathematics, University of Dundee, 1998.

[11] R. Fletcher, S. Leyffer and Ph.L. Toint, On the golbal convergence of a SQP-filter algorithym, *SIAM J. Optimization,* **13** (2003), 44-59.

[12] R. Fletcher and S. Leyffer, A bundle filter method for nonsmooth nonlinear optimization, Tech. Report, NA/195, Department of Mathematics, University of Dundee, 1999.

[13] M. Lalee, J. Nocedal and T.D. Plantenga, On the implementation of an algorithm for large-scale equality constrained optimization, *SIAM Journal on Optimization,* **8**:3 (1998), 682-706.

[14] E.O. Omojokun, Trust region algorithm for optimization with nonlinear equality and inequality constraints, PhD thesis, University of Colorado, Boulder, Colorado, USA, 1989.

[15] T.D. Plantenga, A trust-region method for nonlinear programming based on primal interior point techniques,*SIAM journal on Scientific Computing,* **20**:1, 285-305.

[16] M.J.D. Powell and Y. Yuan, A recursive quadratic programming algorithm for equality constrained optimization, *Math. Prog.,* **35** (1986), 265-278.

[17] V. Torczon, On the convergence of Pattern search algorithms, *SIAM. J. Optim.,* **7** (1997), 1-25.

[18] M. Ulbrich, S. Ulbrich and L.N. Vicente, A globally convergent primal-dual interior-point filter method for nonconvex nonlinear programming, TR00-12 Department of Computational and Applied Mathematics, Rice University, Houston, TX, 2000.

[19] A. Vardi, A trust region algorithm for equality constrained minimization: convergence and implementation, *SIAM Journal on Numerical Analysis,* **22**:3 (1985), 575-591.

[20] Y. Yuan, A null space algorithm for constrained optimization, Report, In Z.C. Shi et. al eds. Advance in Scientific Computing (Science Press, Beijing, 2001), 210-218.