Global Convergence of a Modified Tri-Dimensional Filter Method

Bei Gao, Ke Su, Zixing Rong

Department of Mathematics and Information Science, Hebei University, Baoding, China
Email: shuiguogaobei@163.com, pigeonsk@163.com, rongzixingcn@163.com

Received 9 January 2015; accepted 27 January 2015; published 3 February 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/

Abstract

In this paper, a tri-dimensional filter method for nonlinear programming was proposed. We add a parameter into the traditional filter for relaxing the criterion of iterates. The global convergent properties of the proposed algorithm are proved under some appropriate conditions.

Keywords

Tri-Dimensional, NCP Function, Global Convergence, QP-Free

1. Introduction

This paper is concerned with finding a solution of a Nonlinear Programming (NLP) problem, as following

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad c(x) \leq 0,
\end{align*}
\]

where \( f(x): R^n \rightarrow R \), \( c(x) = (c_1(x), \cdots, c_m(x)) : R^n \rightarrow R^m \) are second-order continuously differentiable. The Lagrangian function associated with problem (1) is the function

\[
L(x, \lambda) = f(x) + \lambda^T c(x)
\]

where \( \lambda = (\lambda_1, \lambda_2, \cdots, \lambda_m) \in R^m \) is the multiplier vector. For simplicity, we denote the column vector \( (x^T, \lambda^T)^T \) as \( (x, \lambda) \). A point \( (x^*, \lambda^*) \in R^{n+m} \) is called a Karush-Kuhn-Tucker (KKT) point if it satisfies the following conditions:

\[
\nabla L(x^*, \lambda^*) = 0, \quad c(x^*) \leq 0, \quad \lambda^* \geq 0, \quad \lambda^* c(x^*) = 0,
\]

we also say that \( x^* \in D \) is a KKT point of problem (1) if there exists a \( \lambda^* \in R^m \) such that \( (x^*, \lambda^*) \) satisfied

Traditionally, this question has been answered by using penalty function. But it is difficult to find a suitable penalty parameter. In order to avoid the pitfalls of penalty function, Nonlinear programming problems (NLP) filter methods were first proposed by Fletcher in a plenary talk at the SIAM Optimization Conference in Victoria in May 1996; the methods are described in [1]. And soon, Global convergence proof of filter method was given in [2]. Because of good global convergence and numerical results, filter methods have quickly become popular in other areas such as nonsmooth optimization, nonlinear equations and so on [3] [4].

Motivated by the ideas of filter methods above, a tri-dimensional filter method for nonlinear programming was proposed as acceptance criterion to judge whether to accept a trial step in our algorithm. We have following advantages:

1) By enhancing the flexibility of filter, motivated by [5], we increase a dimension by introducing a parameter to relax the criterion of iterates.
2) The Maratos effect that makes good progress toward the solution may be rejected and has been avoided by using tri-dimensional filter method as acceptance criterion.
3) Tri-dimensional filter method can make full use of the information we get along the algorithm process.

This paper is divided into 4 sections. The next section introduces the concept of a Modified tri-dimensional filter and the NCP function. In Section 3, an algorithm of line search filter is given. The global convergence properties are proved in the last section.

2. Preliminaries
2.1. NCP Function

The method based on the Fischer-Burmeister NCP function are efficient, both theoretical results and computational experience. The Fischer-Burmeister function has a very simple structure

\[ \psi(a, b) = \sqrt{a^2 + b^2} - a - b. \]

We know that: \( \psi \) is continuously differentiable everywhere except at the origin, but it is strongly semismooth at the origin. i.e. if \( a \neq 0 \) or \( b \neq 0 \), then \( \psi \) is continuously differentiable at \((a, b) \in \mathbb{R}^2\), and

\[ \nabla \psi(a, b) = \left( \frac{a}{\sqrt{a^2 + b^2}}, -1, \frac{b}{\sqrt{a^2 + b^2}}, -1 \right); \]

if \( a = 0 \) and \( b = 0 \), then the generalized Jacobian of \( \psi \) at \((0, 0)\) is

\[ \partial \psi(0, 0) = \left\{ \xi \in \mathbb{R}^2 | \xi - 1, \eta - 1 \right\}. \]

Let

\[ \phi_i(x, \mu) = \psi(-c_i(x), \mu_i), \quad 1 \leq i \leq m \]

We denote \( \Phi(x, \mu) = \left( (\nabla_s L(x, \mu))^T, (\Phi_1(x, \mu))^T \right)^T \), where \( \Phi_1(x, \mu) = \left( \phi_1(x, \mu), \ldots, \phi_m(x, \mu) \right)^T \).

Clearly, the KKT optimality conditions (2) can be equivalently reformulated as the nonsmooth equations \( \Phi(x, \mu) = 0 \).

If \( (c_i(x), \mu_i) \neq (0, 0) \), then \( \phi_i \) is continuously differentiable at \((x, \mu) \in \mathbb{R}^{n+m}\). In this case, we have

\[ \nabla_s \phi_i = \left( \frac{-c_i(x)}{\sqrt{(c_i(x))^2 + \mu_i^2}} + 1 \right) \nabla c_i(x); \quad \nabla \mu \phi_i = \left( \frac{\mu_i}{\sqrt{(c_i(x))^2 + \mu_i^2}} - 1 \right) e_i \]

where \( e_i = (0, \ldots, 0, 1, 0 \ldots, 0)^T \in \mathbb{R}^n \) is the \( i \)th column of the unit matrix, its \( i \)th element is 1, and other elements are 0.

If \( c_i(x) = 0 \) and \( \mu_i = 0, 1 \leq i \leq m \), then \( \phi_i(x, \mu) \) is strongly semismooth and directionally differentiable at \((x, \mu)\). We have
\[ \partial_x \phi(x, \mu) = \{(\xi + 1) \nabla c_i(x)| -1 \leq \xi \leq 1\} \]

and

\[ \partial_{\lambda} \phi_i(x, \mu) = \{(\xi - 1)| -1 \leq \xi \leq 1\}. \]

We may reformulated the KKT (at point \( x^*, \lambda^*, \mu^* \)) conditions as a system of equations.

\[ \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_p)^T \in \mathbb{R}^p \quad \text{and} \quad \mu = (\mu_1, \mu_2, \ldots, \mu_m)^T \in \mathbb{R}^m \]

are the multiplier vectors,

\[ \phi_i(x, \mu_j) = \psi(-g_i(x, \mu_j)) \quad \phi(x^*, \mu^*) = \left( \phi_1(x^*, \mu_1), \phi_2(x^*, \mu_2), \ldots, \phi_m(x^*, \mu_m) \right). \]

Replace the violation constrained function \( p(G(x)) \) in filter F of Fletcher and Leyffer method, we use the violation constrained function \( p(G(x), \mu) = \| \Phi_i(x, \mu) \| \).

If \( c(x^k, \mu^k) \neq (0, 0) \), let

\[ \xi^* = \xi(x^k, \mu^k) = \frac{-c_j}{\sqrt{(c_j^2 + (\mu_j^k)^2)}} + 1; \quad \eta^* = \eta(x^k, \mu^k) = \frac{\mu_j^k}{\sqrt{(c_j^2 + (\mu_j^k)^2)}} - 1; \]

otherwise we denote

\[ \xi^* = \xi(x^k, \mu^k) = 1 + \frac{\sqrt{2}}{2}; \quad \eta^* = \eta(x^k, \mu^k) = -1 + \frac{\sqrt{2}}{2}. \]

Let

\[ V_k = \begin{pmatrix} V_{11}^k & V_{12}^k \\ V_{21}^k & V_{22}^k \end{pmatrix} = \begin{pmatrix} H_k^k & \nabla c_k^2 \\ \text{diag}(\xi^*) & \text{diag}(\eta^*) \end{pmatrix}. \]

where \( H_k \) is a positive matrix which may be modified by BFGS update. \( \text{diag}(\xi^*) \) or \( \text{diag}(\eta^*) \) denotes the diagonal matrix whose \( j \)-diagonal element is \( \xi^* \) or \( \eta^* \) respectively.

**Definition 1.1** [1] A pair \( (f_j, h_j) \) is said to dominate another pair \( (f_l, h_l) \) if and only if both \( f_j \leq f_l \) and \( h_j \leq h_l \).

**Definition 1.2** [1] A filter is a list of pairs \( (f_j, h_j) \) such that no pair dominates any other. A point \( (f_j, h_j) \) is said to be acceptable for inclusion in the filter if it is not dominated by any point in the filter.

**Definition 1.3 NCP pair and NCP functions** [6] We call a pair \( (a, b) \in \mathbb{R}^2 \) to be an NCP pair if \( a \geq 0 \), \( b \geq 0 \) and \( ab = 0 \) a function \( \psi: \mathbb{R}^2 \to \mathbb{R} \) is called an NCP function if \( (a, b) = 0 \) if and only if \( (a, b) \) is an NCP pair.

Denote \( h(x) = \| \Phi_i(x, \mu) \| \) in the following context. It is straightforward to see that the constraint (1) is equivalent to the following equation: \( h(x) = 0 \).

### 2.2. Tri-Dimensional Filter

A two dimensional filter is often used in traditional filter method, some information about convergent like the positions of iterates are neglected. Therefore, we aim to enhance its flexibility of filter. Motivated by [5], we adopt \( (h, f, \delta) \) in which a parameter \( \delta \) is used to relax the criterion of iterates. We denote the filter by \( F_k \) for each iteration \( k \). Flexible exact penalty function is introduced to promote convergence refer to [7]. Given a prescribed interval, penalty parameter can be chosen as any number from it and it is extends classical penalty function methods. We generalized the idea to filter which we called Tri-dimensional filter. Different from the original two dimensional filter, we increase a dimension by introducing a parameter.

We use pairs \( (h_j, f_j, \delta_j) \) to constitute the elements of filter, where \( \delta_j \) is a non-negative parameter. Our strategy for setting \( \delta_j \) depends on the region in \( h - f - \delta \) space to which \( s_k \) moves into. **Figure 1** is Distinct regions defined by the current iterate.

If \( s_k \) moves into region \( I \), which is defined as

\[ I = \{(h, f, \delta) : h > 1.1h_k \text{ and } f + \delta h < f_k + \delta h_k, \delta_k \geq 0\}, \]
We say that the algorithm does not make good improvement since we do not want to accept points with larger constraint violation. Thus, we try to impose stricter acceptance criterion. Meanwhile, we do not permit \( \delta_k \) larger than \( \sigma_k \). In our algorithm, we increase \( \delta^k \) in the following way

\[
\delta_{k+1} = \min \left\{ \sigma_k, \delta_k + \max \left\{ 0.001, 0.1 \left( \frac{f_k - f(x_k + s_k)}{h_k - h(x_k + s_k)} - \delta_k \right) \right\} \right\}. \tag{3}
\]

If \( s_k \) moved into region \( \Pi \) which is defined as

\[
\Pi = \{ (h, f, \delta) : h < 0.9h_k \text{ and } f + \delta_k h < f_k + \delta_k h_k, \delta_k \geq 0 \},
\]

We say that the algorithm makes good improvement since it reduces not only the constraint violation, but also the penalty function value. So, we may loosen the acceptance criterion to wish more improvement. Here, we achieve this goal by reducing \( \delta_k \) by setting

\[
\delta_{k+1} = \max \left\{ 0, \sigma_k - \max \left\{ 0.001, 0.1 \left( \frac{f_k - f(x_k + s_k)}{h_k - h(x_k + s_k)} - \delta_k \right) \right\} \right\}. \tag{4}
\]

In our algorithm, the trial step \( s_k \) is accepted by filter if

\[
h(x_k + s_k) \leq \gamma h_j \quad \text{or} \quad f(x_k + s_k) + \delta_k h(x_k + s_k) \leq \gamma (f_j + \delta_j) \quad \text{and} \quad \delta_k \geq 0 \tag{5}
\]

For all \( (h_j, f_j, \delta_j) \in \mathcal{F}_j \). The parameter \( \gamma \in (0, 1) \) is a constant close to 1 which sets an “envelope” around the border of the dominated part of the \( (h, f, \delta) \)-space in which the trial step is rejected. And also in the filter if

\[
h_j > h_k \text{ and } h_j + \delta_j f_j > h_k + \delta_k f_k \text{ and } \delta_k \geq 0 \tag{6}
\]

then we say \( x_j \) is dominated by \( x_k \).

### 3. Description of the Algorithm

In this section we hope that the Lagrange multiplier \( \lambda_k \) will converge to the Lagrange multiplier \( \lambda^* \) at the solution \( x^* \). From the KKT system of (1), a good estimate of the Lagrange multiplier is the least square solution of \( c(x) - A(x) \lambda = 0 \), namely \( \lambda = (A(x)^T c(x)) \). In our algorithm, \( \lambda_k \) is updated only after a trial step is accepted, and is set componentwise as

\[
\lambda'_k = \begin{cases} 
A'_k c'_k, & l' = u', \\
\max \left\{ (A'_k c'_k)^T \right\}, & l' = 0, u' = +\infty, \\
\min \left\{ (A'_k c'_k)^T \right\}, & l' = -\infty, u' = 0,
\end{cases} \tag{7}
\]

Now, we consider how to update the penalty parameter. Let \( x' \) be a solution of (1) at which the LICQ is
satisfied, and the second order sufficient conditions are satisfied. Then when \( \sigma > \| x^* \| \), \( x^* \) is the strict local minimizer of penalty function. So we force the condition at each iteration: \( \sigma_{k+1} \geq \| x_{k+1} \| \).

And also, since the penalty term aims to reduce the constraint violation we double the penalty parameter if the constraint violation could not reduce by half, that is

\[
\sigma_{k+1} = 2\sigma_k, \text{ if } h(x_k + s_k) \geq 0.5h_k.
\]

To summarize, we update the penalty parameter in the following formula:

\[
\sigma_{k+1} = \begin{cases} 
\max \left\{ 2\sigma_k, \| x_{k+1} \| \right\}, & h(x_k + s_k) \geq 0.5h_k, \\
\max \left\{ \sigma_k, \| x_{k+1} \| \right\}, & \text{otherwise}.
\end{cases}
\]

(8)

The improved algorithm is presented as following.

**Algorithm**

**Step 0.** Initialization: Give a starting point \( x_0 \in \mathbb{R}^n \), \( \mu_0 \), \( \lambda_0 \) and a initial positive definite matrix \( H_0 \), \( \tau \in (0, 1) \), \( k = 0 \). compute \( h_0, f_0, g_0, A_0 \).

**Step 1.** Termination test. If \( h_k + \| g_k - A_k \lambda_k \| < \varepsilon \) then returning \( x_k \) as a solution and stop.

**Step 2.** Computation of the search direction. compute \( d^{k0} \) and \( \lambda^{k0} \) by solving the following linear system in \((d, \lambda)\):

\[
V_k \left( \begin{array}{c} d \\ \lambda \end{array} \right) = \left( \begin{array}{c} -\nabla f_k \\ 0 \end{array} \right).
\]

(9)

where \( \nabla f_k = \nabla f(x^k) \).

If \( d^{k0} = 0 \), then stop otherwise, compute \( (d^{k1}, \lambda^{k1}) \) by solving the following linear system in \((d, \lambda)\):

\[
V_k \left( \begin{array}{c} d \\ \lambda \end{array} \right) = \left( \begin{array}{c} -\nabla L_k \\ -\Phi_k^1 \end{array} \right).
\]

(10)

where \( \nabla L_k = \nabla L(x^k, \lambda^k) \) and \( \Phi_k^1 = \Phi_1(x^k, \lambda^k) \).

**Step 3.** Liner search with filter

If \( \Phi_k^1 = 0 \) then let \( b^k = 1 \) and \( \rho^k = 0 \), otherwise if \( d^{k0} = 0 \) then let \( b^k = 0 \) and \( \rho^k = 1 \), otherwise denote \( b^k = (1 - \rho^k) \) and

\[
\rho^k = \begin{cases} 
1 & \text{if } (d^{k1})^T \nabla f_k \leq \theta (d^{k0})^T \nabla f_k \\
(1-\theta) \frac{(d^{k0})^T \nabla f_k}{(d^{k0} - d^{k1})^T \nabla f_k} & \text{otherwise}
\end{cases}
\]

(11)

and let

\[
\left( \begin{array}{c} d^k \\ \lambda^k \end{array} \right) = b^k \left( \begin{array}{c} d^{k0} \\ \lambda^{k0} \end{array} \right) + \rho^k \left( \begin{array}{c} d^{k1} \\ \lambda^{k1} \end{array} \right).
\]

**Step 4.** Acceptance criterion of the trial step

Let \( x^* = x_k + s_k \), evaluate \( h^* \) and \( f^* \) and \( \delta^* \); If \( x^* \) is accepted by filter, \( x_{k+1} = x^* \) and go to step 5; \( x_{k+1} = x_k \), and \( k = k+1 \); go to step 2.

**Step 5.** Parameter update

Update \( \lambda_{k+1} \) by (7); Update \( \sigma_{k+1} \) by (8); Update \( \delta_{k+1} \) by (3) or (4); \( k = k+1 \) go to step 1.

**4. The Convergence Properties**

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

**A1** The level set \( \{ x | F(x) \leq f(x^0) \} \) is bounded, and for sufficiently large \( k \), \( \| x^k + \lambda^{k0} + \lambda^{k1} \| < \mu \)
A2 $f$ and $g_i$ are twice Lipschitz continuously differentiable, and for all $y, z \in R^{n \times m}$,
$$
|\nabla L(y) - \nabla L(z)| \leq m_1 \|y - z\|, \quad \|\Phi(y) - \Phi(z)\| \leq m_2 \|y - z\|
$$
where $m_1 > 0$ is the Lipschitz constant.
A3 $H^k$ is positive definite and there exist positive numbers $m_1$ and $m_2$ such that
$$
m_1 |d|^2 \leq d^T H^k d \leq m_2 \|d\|^2
$$
for all $d \in R^n$ and all $k$.

**Lemma 1.** If $\Phi^i \neq 0$ then $V^k$ and $V^*$ are nonsingular.

**Proof.** If $V_k(u, \vartheta) = 0$, for some $(u, \vartheta) \in R^n$, where $\vartheta = (\vartheta_1, \ldots, \vartheta_n)^T$, and $u = (u_1, \ldots, u_n)^T$, then we have
$$
H^k u + \nabla c^k v = 0
$$
and
$$
diag(\xi^k)(\nabla c^k)^T u + diag(\eta^k)v = 0
$$
From the definition of $\xi^j$ and $\eta^j$, we know that $\xi^j \geq 0$ and $\eta^j \neq 0$ for all $j$. So, $diag(\eta^k)$ is nonsingular. We have
$$
v = -\left(diag(\eta^k)\right)^{-1} diag(\xi^k)(\nabla c^k)^T u
$$
Putting (14) into (12), we have
$$
u^T (H^k u + \nabla c^k v) = u^T H^k u - u^T \nabla c^k \left(diag(\xi^k)\right)^{-1} (\nabla c^k)^T u = 0
$$
The fact that $-\nabla c^k \left(diag(\xi^k)\right)^{-1} (\nabla c^k)^T$ is positive semidefinite implies $u = 0$, and then $v = 0$ by (14). $V^k$ is nonsingular. And if $(x^*, \mu^*)$ is an accumulation point of $(x^k, \mu^k)$, $(x^*, \mu^*) \rightarrow (x^*, \mu^*)$, $\Phi^i \rightarrow \Phi^*$ and $V^k \rightarrow V^*$. If $\Phi^* \neq 0$ then $\Phi^*$ is nonsingular. This lemma holds. □

The lemma 2 hold (see [8] Lemma 2)

**Lemma 2.** If $d^{k_0} = 0$, then $\nabla f(x^k) = 0$, and $x^k$ is KKT point of problem (NLP).

**Lemma 3.** Consider an infinite sequence iterations on which $\left\{\|\Phi^i\|^2, f^i\right\}$ entered into filter, where $\|\Phi^i\|^2 > 0$ and $\{f^i\}$ is bounded below. It follows that $\Phi^i \rightarrow 0$.

**Proof.** Suppose the theorem is not true, then exists an $\varepsilon > 0$ and an infinitely members of index set $K$ such that either $\|\Phi^i(x^k, \mu^k)\| \geq \varepsilon > 0$ and $\|\Phi^i(x^{k+1}, \mu^{k+1})\| \leq \eta \Phi^i(x^k, \mu^k)$ for any $k \in K$. Then we obtain that $\|\Phi^i(x^k, \mu^k)\| \rightarrow 0$, or $\{f_k\}_{k \in K}$ is monotonically decreasing, then lemma 5.1 implies $\|\Phi^i(x^k, \mu^k)\| \rightarrow 0$.

So, the lemma holds. □

The following lemma 4 - 5 hold (see [9])

**Lemma 4.** $d^k \rightarrow 0$.

**Lemma 5.** If $(x^k, \mu^k)$ is an accumulation point of $(x^k, \mu^k)$, then $d^* = \mu^*$ and $d^*, \lambda^*$ is the solution of:
$$
\lambda^* \left(\begin{array}{c}
v^* \\
0
\end{array}\right)
= \frac{-\nabla f^*}{0}
and
V^k(x^*, \mu^*) = 0$.

**Theorem 1.** If $(x^k, \mu^k)$ is an accumulation point of $(x^k, \mu^k)$, then $x^k$ is a KKT point of Problem (NLP).

It is obviously to prove the conclusion holds according to the above lemmas.

**Acknowledgements**

We thank the Editor and the referee for their comments. This work is supported by the National Natural Science...
Foundation of China (No. 11101115), the Natural Science Foundation of Hebei Province (No. 2014201033) and the Science and Technology project of Hebei province (No. 13214715).

References


Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.