Applied Mathematics, 2015, 6, 235-241
Published Online February 2015 in SciRes. http://www.scirp.org/journal/am
http://dx.doi.org/10.4236/am.2015.62023

Scientific Research Publishing

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

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

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

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

$$L(x,\lambda) = f(x) + \lambda c(x)$$

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

$$\nabla_x L(x^*, \lambda^*) = 0,\ \ c(x^*) \le 0,\ \lambda^* \ge 0,\ \lambda^* c(x^*) = 0, \tag{2}$$

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^*, λ^*) satisfied

How to cite this paper: Gao, B., Su, K. and Rong, Z.X. (2015) Global Convergence of a Modified Tri-Dimensional Filter Method. *Applied Mathematics*, **6**, 235-241. http://dx.doi.org/10.4236/am.2015.62023

(2).

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 nonliner 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 that 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: ψ 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 ψ is continuously differentiable at $(a,b)\in R^2$, 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 ψ at $(0,0)$ is

$$\partial\psi(0,0)=\left\{\xi-1,\eta-1\,\middle|\,\xi^2+\eta^2=1\right\}.$$

Let

$$\phi_i(x,\mu)=\psi(-c_i(x),\mu_i),\ 1\le i\le m$$

We denote $\Phi(x,\mu)=\left(\left(\nabla_x L(x,\mu)\right)^{\mathrm{T}},\left(\Phi_1(x,\mu)\right)^{\mathrm{T}}\right)^{\mathrm{T}}$, where $\Phi_1(x,\mu)=\left(\phi_1(x,\mu),\cdots,\phi_m(x,\mu)\right)^{\mathrm{T}}$.

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 ϕ_i is continuously differentiable at $(x,\mu)\in R^{n+m}$. In this case, we have

$$\nabla_x\phi_i=\left(\frac{-c_i(x)}{\sqrt{(c_i(x))^2+\mu_i^2}}+1\right)\nabla c_i(x);\ \nabla_\mu\phi_i=\left(\frac{\mu_i}{\sqrt{(c_i(x))^2+\mu_i^2}}-1\right)e_i$$

where $e_i=(0,\cdots,0,1,0\cdots,0)^{\mathrm{T}}\in R^m$ 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 $\mu_i=0, 1\le i\le m$, then $\phi_i(x,\mu)$ is strongly semismooth and directionally differentiable at (x,μ). We have

$$\partial_x \phi_i(x,\mu) = \{(\xi+1)\nabla c_i(x) | -1 \le \xi \le 1\}$$

and

$$\partial_{\mu_i} \phi_i(x,\mu) = \{(\xi-1) | -1 \le \xi \le 1\}.$$

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

$$\nabla_x L(x^*,\lambda^*,\mu^*) = 0,\ \Phi_1(x^*,\mu^*) = 0,$$

where $\lambda = (\lambda_1,\lambda_2,\cdots,\lambda_p)^{\mathrm{T}} \in R^p$ and $\mu = (\mu^1,\mu^2,\cdots,\mu^m)^{\mathrm{T}} \in R^m$ are the multiplier vectors, $\phi_i(x,\mu_j) = \psi(-g_i(x),\mu_j)$, $\phi(x^*,\mu^*) = (\phi_1(x,\mu_1^*),\phi_2(x,\mu_2^*),\cdots,\phi_m(x,\mu_m^*))$.

Replace the violation constrained function $p(G(x))$ in filter F of Fletcher and Leyffer method, we use the violation constrained function $p(G(x),\mu) = \|\Phi_1(x,\mu)\|^2$.

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

$$\xi_j^k = \xi_j(x^k,\mu^k) = \frac{-c_j^k}{\sqrt{(c_j^k)^2 + (\mu_j^k)^2}} + 1;\ \eta_j^k = \eta_j(x^k,\mu^k) = \frac{\mu_j^k}{\sqrt{(c_j^k)^2 + (\mu_j^k)^2}} - 1;$$

otherwise we denote

$$\xi_j^k = \xi_j(x^k,\mu^k) = 1 + \frac{\sqrt{2}}{2};\ \eta_j^k = \eta_j(x^k,\mu^k) = -1 + \frac{\sqrt{2}}{2}.$$

Let

$$V_k = \begin{pmatrix} V_{11}^k & V_{12}^k \\ V_{21}^k & V_{22}^k \end{pmatrix} = \begin{pmatrix} H^k & \nabla c^k \\ \operatorname{diag}(\xi^k)(\nabla c^k)^{\mathrm{T}} & \operatorname{diag}(\eta^k) \end{pmatrix}.$$

where H^k is a positive matrix which may be modified by BFGS update. $\operatorname{diag}(\xi^k)$ or $\operatorname{diag}(\eta^k)$ denotes the diagonal matrix whose *j* diagonal element is ξ^k or η^k 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 $f_j \le f_l$ and $h_j \le h_l$.

Definition 1.2 [1] A filter is a list of pairs (f_l,h_l) 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 $(a,b) \in R^2$ to be an NCP pair if $a \ge 0$, $b \ge 0$ and $ab = 0$ a function $\psi: R^2 \to R$ is called an NCP function if $(a,b) = 0$ if and only if (a,b) is an NCP pair.

Denote $h(x) = \|\Phi_1(x,\mu)\|^2$ 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,δ) in which a parameter δ is used to relax the criterion of iterates. We denote the filter by $\mathcal{F}_k$ 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,δ_j) to constitute the elements of filter, where δ_j is a non-negative parameter. Our strategy for setting δ_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

$$I = \{(h,f,\delta): h > 1.1h_k \text{ and } f + \delta_k h < f_k + \delta_k h_k,\ \delta_k \ge 0\},$$

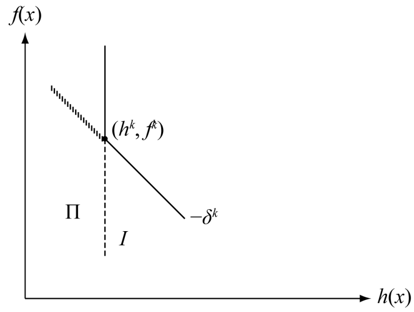

Figure 1. Distinct regions defined by the current iterate.

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 δ_k larger than σ_k. In our algorithm, we increase δ^k in the following way

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

If s_k moved into region Π which is defined as

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

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 δ_k by setting

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

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

$$h\left(x_k+s_k\right)\leq\gamma h_j \text{ or } f\left(x_k+s_k\right)+\delta_j h\left(x_k+s_k\right)\leq\gamma\left(f_j+\delta_j\right) \text{ and } \delta_k\geq 0 \tag{5}$$

For all $\left(h_j,f_j,\delta_j\right)\in\mathcal{F}_k$. 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 $\left(h,f,\delta\right)$-space in which the trial step is rejected. And also in the filter if

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

then we say x_j is dominated by x_k.

3. Description of the Algorithm

In this section we hope that the Lagrange multiplier λ_k will converge to the Lagrange multiplier λ^* 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 $\lambda=\left(A(x)\right)^+c(x)$. In our algorithm, λ_k is updated only after a trial step is accepted, and is set componentwise as

$$\lambda_k^i=\begin{cases}A_k^+c_k^i, & l^i=u^i,\\ \max\left\{\left(A_k^+c_k\right)^i,0\right\}, & l^i=0,u^i=+\infty,\\ \min\left\{\left(A_k^+c_k\right)^i,0\right\}, & l^i=-\infty,u^i=0,\end{cases} \tag{7}$$

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 $\sigma > \left\|\lambda^*\right\|_1$, x^* is the strict local minimizer of penalty function. So we force the condition at each iteration: $\sigma_{k+1} \ge \left\|\lambda_{k+1}\right\|_1$.

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

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

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

$$\sigma_{k+1} = \begin{cases} \max\left\{2\sigma_k, \left\|\lambda_{k+1}\right\|_1\right\}, & h(x_k + s_k) \ge 0.5h_k, \\ \max\left\{\sigma_k, \left\|\lambda_{k+1}\right\|_1\right\}, & \text{otherwise.} \end{cases} \tag{8}$$

The improved algorithm is presented as following.

Algorithm

Step 0. Initialization: Give a starting point $x_0 \in R^n$, μ_0, λ_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. Terimination test. If $h_k + \left\|g_k - A_k\lambda_k\right\|_\infty < \varepsilon$ then returing x_k as a solution and stop.

Step 2. Computation of the search direction. compute d^{k_0} and λ^{k_0} by solving the following linear system in (d, λ):

$$V_k \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f^k \\ 0 \end{pmatrix}. \tag{9}$$

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

If $d^{k_0} = 0$, then stop otherwise, compute (d^{k_1}, λ^{k_1}) by solving the following linear system in (d, λ):

$$V_k \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla L^k \\ -\Phi_1^k \end{pmatrix}. \tag{10}$$

where $\nabla L^k = \nabla L(x^k, \lambda^k)$ and $\Phi_1^k = \Phi_1(x^k, \lambda^k)$.

Step3. Liner search with filter

If $\Phi_1^k = 0$ then let $b^k = 1$ and $\rho^k = 0$, otherwise if $d^{k_0} = 0$ then let $b^k = 0$ and $\rho^k = 1$, otherwise denote $b^k = (1 - \rho^k)$ and

$$\rho^k = \begin{cases} 1 & \text{if } \left(d^{k_1}\right)^{\mathrm{T}} \nabla f^k \le \theta\left(d^{k_0}\right)^{\mathrm{T}} \nabla f^k \\ (1-\theta)\dfrac{\left(d^{k_0}\right)^{\mathrm{T}} \nabla f^k}{\left(d^{k_0} - d^{k_1}\right)^{\mathrm{T}} \nabla f^k} & \text{otherwise} \end{cases} \tag{11}$$

and let

$$\begin{pmatrix} d^k \\ \lambda^k \end{pmatrix} = b^k \begin{pmatrix} d^{k_0} \\ \lambda^{k_0} \end{pmatrix} + \rho^k \begin{pmatrix} d^{k_1} \\ \lambda^{k_1} \end{pmatrix},$$

Step 4. Acceptance criterion of the trial step

Let $x^+ = x_k + s_k$, evalute h^+ and f^+ and δ^+; 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. Paramenters update

Update λ_{k+1} by (7); Update σ_{k+1} by (8); Update δ_{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 $\left\{x \middle| F(x) \le f\left(x^0\right)\right\}$ is bounded, and for sufficiently large *k*, $\left\|\mu^k + \lambda^{k_0} + \lambda^{k_1}\right\| < \mu$

A2 f and g_i are twice Lipschitz continuously differentiable, and for all $y, z \in R^{n+m}$,

$$\left\|\nabla L(y)-\nabla L(z)\right\| \le m_3 \|y-z\|, \quad \left\|\Phi(y)-\Phi(z)\right\| \le m_3 \|y-z\|,$$

where $m_3 > 0$ is the Lipschitz constant.

A3 H^k is positive definite and there exist positive numbers m_1 and m_2 such that

$$m_1 \|d\|^2 \le d^{\mathrm{T}} H^k d \le m_2 \|d\|^2$$

for all $d \in R^n$ and all k.

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

Proof. If $V_k \begin{pmatrix} u \\ \vartheta \end{pmatrix} = 0,$ for some $(u, \vartheta) \in R^n$, where $\vartheta = (\vartheta_1, \cdots, \vartheta)^{\mathrm{T}}$, $u = (u_1, \cdots, u_n)^{\mathrm{T}}$, then we have

$$H^k u + \nabla c^k v = 0 \tag{12}$$

and

$$\operatorname{diag}(\xi^k)(\nabla c^k)^{\mathrm{T}} u + \operatorname{diag}(\eta^k) v = 0 \tag{13}$$

From the definition of ξ_j^k and η_j^k, we know that $\xi_j^k \ge 0$ and $\eta_j^k \ne 0$ for all *j*. So, diag η^k is nonsingular. We have

$$v = -\left(\operatorname{diag}(\eta^k)\right)^{-1} \operatorname{diag}(\xi^k)(\nabla c^k)^{\mathrm{T}} u \tag{14}$$

Putting (14) into (12), we have

$$u^{\mathrm{T}}\left(H^k u + \nabla c^k v\right) = u^{\mathrm{T}} H^k u - u^{\mathrm{T}} \nabla c^k \operatorname{diag}(\xi^k)\left(\operatorname{diag}(\eta^k)\right)^{-1}(\nabla c^k)^{\mathrm{T}} u = 0$$

The fact that $-\nabla c^k \operatorname{diag}(\xi^k)\left(\operatorname{diag}(\eta^k)\right)^{-1}(\nabla c^k)^{\mathrm{T}}$ is positive semidefinite implies $u = 0$, and then $v = 0$ by (14). V^k is nonsingular. And if (x^*, μ^*) is an accumulation point of $\{(x^k, \mu^k)\}$, $\{(x^k, \mu^k)\} \to (x^*, \mu^*)$, $\Phi^k \to \Phi^*$ and $V^K \to V^*$. If $\Phi^* \ne 0$ then Φ^* 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* $\left\{\left\|\Phi_1^k\right\|^2, f^k\right\}$ *entered into filter, where* $\left\|\Phi_1^k\right\|^2 > 0$ *and* $\{f^k\}$ *is bounded below. It follows that* $\Phi_1^k \to 0$.

Proof. Suppose the theorem is not true, then exists an $\varepsilon > 0$ and an infinitely members of index set K such that either $\left\|\Phi_1(x^k, \mu^k)\right\| \ge \varepsilon > 0$ and $\left\|\Phi_1(x^{k+1}, \mu^{k+1})\right\| \le \eta \Phi_1(x^k, \mu^k)$ for any $k \in K$. then we obtain that $\left\{\left\|\Phi_1(x^k, \mu^k)\right\|\right\}_{k\in K} \to 0$, or $\{f_k\}_{k\in K}$ is monotonically decreasing, then lemma 5.1 implies $\left\|\Phi_1(x^k, \mu^k)\right\| \to 0$. So, the lemma holds. □

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

Lemma 4. $d^k \to 0$.

Lemma 5. *If* (x^*, μ^*) *is an accumulation point of* $\{(x^k, \mu^k)\}$ *then* $d^* = 0$, *and* d^*, λ^* *is the solution of*:

$$V_* \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f^* \\ 0 \end{pmatrix}.$$

and $\nabla L(x^*, \mu^*) = 0$.

Theorem 1. *If* (x^*, μ^*) *is an accumulation point of* $\{(x^k, \mu^k)\}$ *then* x^* *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

[1] Fletcher, R. and Leyyfer, S. (2002) Nonlinear Programming without a Penalty Function. *Mathematical Programming*, **91**, 239-269. http://dx.doi.org/10.1007/s101070100244

[2] Fletcher, R., Leyffer, S. and Toint, P.L. (1998) On the Global Convergence of an SLP-Filter Algorithm. Numerical Analysis Report NA/183, University of Dundee, Dundee.

[3] Fletcher, R., Leyffer, S., *et al.* (2006) A Brief History of Filter Methods. Mathematics and Computer Science Division, Preprint ANL/MCSP1372-0906, Argonne National Laboratory.

[4] Chin, C.M., Rashid, A.H.A. and Nor, K.M. (2007) Global and Local Convergence of a Filter Line Search Method for Nonlinear Programming. *Optimization Method Software*, **22**, 365-390. http://dx.doi.org/10.1080/10556780600565489

[5] Wang, X. (2010) A New Filter Trust Region Method for Nonlinear Programming. *Journal of the Operations Research of China*, **10**, 133-140.

[6] Zhou, Y. and Pu, D. (2007) A New QP-Free Feasible Method for Inequality Constrained Optimization. *OR Transactions*, **11**, 31-43.

[7] Curtis, F.E. and Nocedal, J.(2008) Flexible Penalty Function for Nonlinear Constrained Optimization. *IMA Journal of Numerical Analysis*, **28**, 749-769. http://dx.doi.org/10.1093/imanum/drn003

[8] Su, K. (2008) A New Globally and Superlinearly Convergent QP-Free Method for Inequality Constrained Optimization. *Journal of Tongji University*, **36**, 265-272.

[9] Pu, D.G., Li, K. and Xue, W. (2005) Convergence of QP-Free Infeasible Methods for Nonlinear Inequality Constrained Optimization Problems. *Journal of Tongji University*, **33**, 525-529.