Journal of Mathematical Research & Exposition
Vol. 17, No. 3, 375-381, Aug. 1997

Projected Gradient Type Method of Centers for Constrained Optimization*

Shi Baochang　*Zhou Xiaoyang*
(Dept. of Math., Huazhong Univ. of Sci. & Tech., Wuhan 430074)

Hu Xinsheng
(Shenzhen Radio and Television University, Shenzhen 518008)

Abstract In this paper, a new algorithm-projected gradient type method of centers for constrained optimization is presented. Under the assumptions of continuous differentiability and nondegeneracy, the global convergence of the algorithm is proved. The method here is simple in computation and flexible in form.

Key words constrained optimization, nondegeneracy, projected gradient type method of centers; global convergence.

Classification AMS(1991) 40C30, 49M37/CCL O221.2, O224

1. Introduction

Method of centers is a class of important algorithms for nonlinear programming, which has the advantages of feasible directions method and penalty function method, and can overcome some of their shortcomings such as the requirement of feasibility of initial point for the former and the uncertainty of penalty factor for the later[1, 2]. However, the existing methods of centers only consider inequality constraints and use the subproblems of linear/quadratic programming to generate the search directions.

In this paper, we consider the following problem:

$$\text{(NP)}\quad \min f(x),\ \text{s.t. } x \in R = \{x \in R^n \,|\, g_j(x) \le 0, j \in L;\ a_i^T x - b_i = 0,\ i \in M\},$$

where $f, g_j \in C^1\ (j \in L)$. L and M are finite index sets. Since the linear constraints can be treated directly, and some of the constraints may be required to be satisfied in practice, we divide L into two subsets: $L = L_1 \cup L_2$ such that $L_1 \cap L_2 = \emptyset$, and use only $g_j\ (j \in L_2)$ to construct the merit (distance) function of (NP) with the parameter $y \in R^n$:

$$f(x, y) = \max\{f(x) - f(y) - r\varphi(y), g_j(x) - \varphi(y), j \in L_2\}, \tag{1.1}$$

* Received Nov. 25, 1994.

where $r>0$, $\varphi(y)=\max\{0, g_j(x), j\in L_2\}$. Then for the current iteration point

$$x^k\in R_1=\{x\in R^n \mid g_j(x)\le 0, j\in L_1;\ a_i^Tx-b_i=0, i\in M\}, \tag{1.2}$$

by using the projected gradient direction to generate the descent feasible direction of the parametric programming (P_{x^k}): $\min\{f(x,x^k) \mid x\in R_1\}$ at $x=x^k$, a projected gradient type method of centers for (NP) is obtained. Under the assumption of nondegeneracy, the global convergence of the method is proved. The method is simple in computation and flexible in form.

2 Definitions and Notations

Definition 2.1 *Given $y\in R^n$, $J\subset L_2$, define the generalized pseudo directional derivative of $f(\cdot, y)$ at x along $d\in R^n$ with respect to J as follows*

$$f_J^*(x,y;d)=\max\{f(x)+\nabla f(x)^Td-f(y)-r\varphi(y);\ g_j(x)+\nabla g_j(x)^Td - \varphi(y), j\in J\}-f(x,y). \tag{2.1}$$

Since $f(x,x)=0$, we obtain

$$f_J^*(x,x;d)=\max\{\nabla f(x)^Td-r\varphi(x);\ g_j(x)+\nabla g_j(x)^Td-\varphi(x), j\in J\}. \tag{2.2}$$

Denote $f'(x,y;d)$, $f^0(x,y;d)$ respectively for the directional derivative and the generalized directional derivative[3] of $f(\cdot,y)$ at x along $d\in R^n$ and $f'(x,y;d,p)$, $f^0(x,y;d,p)$ for those of $f(\cdot,\cdot)$ at (x,y) along $(d,p)\in R^{2n}$; $I_1(x)=\{j\in L_1 \mid g_j(x)=0\}$, $I_2(x)=\{j\in L_2 \mid g_j(x)=\varphi(x)\}$, $I(x)=\{j\in L \mid g_j(x)=0\}$ and $J(x)=I_1(x)\cup I_2(x)$, the following lemma is obvious. See [3] in detail.

Lemma 2.1 (1) $f'(x,x;d)\le f^*_{I_2(x)}(x,x;d)\le f_J^*(x,x;d)$, $\forall J\supseteq I_2(x)$;

(2) $f'(x,y;d)=f'(x,y;d,0)=f^0(x,y;d,0)=f^0(x,y;d)$; $J(x)=I(x)$, $\forall x\in R$;

(3) *For* $x,y,d,p\in R^n$, $t\ge 0$, $\exists\theta\in(0,1)$ *such that*

$$f(x+td, y+tp)-f(x,y)\le tf^0(x+\theta td, y+\theta tp; d,p)=tf'(x+\theta td, y+\theta tp; d,p).$$

If $p=0$, then by (2), we obtain

$$f(x+td,y)-f(x,y)\le tf^0(x+\theta td, y; d)=tf'(x+\theta td, y; d);$$

(4) $\overline{\lim}_{(x,y,d)\to(x^*,y^*,d^*)} f'(x,y;d)=\overline{\lim}_{(x,y,d)\to(x^*,y^*,d^*)} f^0(x,y;d)\le f^0(x^*,y^*;d^*)=f'(x^*,y^*;d^*)$.

Now, in order to obtain the search direction of projected gradient type, we assume

(H) $\forall x\in R_1$, $\mathrm{Rank}\{\nabla g_j(x), j\in J(x), a_i, i\in M\}=|J(x)\cup M|$.

Lemma 2.2 *(H) holds if and only if for any bounded subset $S\subseteq R_1$, there exists $\epsilon_S>0$ such that*

$$\forall x\in S, \epsilon\in(0,\epsilon_S],\ \det N_{J(x,\epsilon)}(x)^T N_{J(x,\epsilon)}(x)\ge \epsilon_S, \tag{2.3}$$

where

$$N_J(x)=\{\nabla g_j(x), j\in J; a_i, i\in M\},\ \text{for } J\subseteq L;\ J(x,\epsilon)=I_1(x,\epsilon)\cup I_2(x,\epsilon),$$
$$I_1(x,\epsilon)=\{j\in L_1 \mid g_j(x)\ge -\epsilon\},\ I_2(x,\epsilon)=\{j\in L_2 \mid g_j(x)-\varphi(x)\ge-\epsilon\}.$$

Proof The proof is similar to that of [4, Th. 1] and is omitted.

Let $x \in R_1, J_1 \supseteq I_1(x), J_2 \supseteq I_2(x)$, $J = J_1 \cup J_2$ such that $\det N_J(x)^T N_J(x) > 0$. Denote $B_J(x) = N_J(x)[N_J(x)^T N_J(x)]^{-1}$, $P_J(x) = E - N_J(x) B_J(x)^T$, $u_J(x) = -B_J(x)^T \nabla f(x)$, define the following directions

$$d_J(x) = -P_J(x)\nabla f(x) + B_J(x)[v_J(x) - \rho(x)\delta_J], \tag{2.4}$$

where for $j \in J = J_1 \cup J_2$,

$$v_J^j(x) = \begin{cases} u_J^j(x), & \text{if } u_J^j(x) < 0, \\ -g_j(x), & \text{if } u_J^j(x) \geq 0 \text{ and } g_j(x) \leq 0, \\ \varphi(x) - g_j(x), & \text{if } u_J^j(x) \geq 0 \text{ and } g_j(x) > 0. \end{cases} \tag{2.5}$$

and for $i \in M$, $v_J^i(x) = 0$;

$$\delta_J^j = \begin{cases} 1, & \text{if } j \in J, \\ 0, & \text{if } j \in M; \end{cases} \quad \rho(x) \geq 0. \tag{2.6}$$

Lemma 2.3 (1) *Let* $\alpha(J, x) = \|P_J(x)\nabla f(x)\|^2 + u_J^T(x) v_J(x) + r\varphi(x)$, *then* $\alpha(J, x) \geq 0$, *and* $\alpha(J, x) = 0$ *implies that* x *is a Kuhn-Tucker (K-T) point of* (NP);

(2) *If* $\alpha(J, x) > 0$ *and* $\rho(x) > 0$ *such that* $-\alpha(J, x) + \rho(x) u_J^T(x)\delta_J < 0$, *then*

$$f'_{J_2}(x, x; d_J(x)) < 0 \text{ and } a_i^T d_J(x) = 0,\ \forall i \in M,\ \nabla g_j(x)^T d_J(x) < 0,\ j \in I_1(x), \tag{2.7}$$

which means that $d_J(x)$ *is a descent feasible direction of* (P_x) *at* x.

Proof By (2.5) and (2.6), we have

$$u_J^T(x) v_J(x) = \sum_{j \in J_1 \cup J_2, u_J^j(x) < 0} (u_J^j(x))^2 + \sum_{\substack{j \in J_1, u_J^j(x) \geq 0 \\ g_j(x) \leq 0}} -g_j(x) u_J^j(x)$$
$$+ \sum_{\substack{j \in J_2, u_J^j(x) \geq 0 \\ g_j(x) > 0}} (\varphi(x) - g_j(x)) u_J^j(x) \geq 0 \tag{2.8}$$

hence $\alpha(J, x) \geq 0$. If $\alpha(J, x) = 0$, then $P_J(x)\nabla f(x) = 0$, $u_J^T(x) v_J(x) = 0$ and $\varphi(x) = 0$, we obtain

$$\nabla f(x) + N_J(x) u_J(x) = 0,$$
$$u_J^T(x) v_J(x) = \sum_{j \in J_1 \cup J_2, u_J^j(x) < 0} (u_J^j(x))^2 + \sum_{\substack{j \in J_1 \cup J_2, u_J^j(x) \geq 0 \\ g_j(x) \leq 0}} -g_j(x) u_J^j(x) = 0$$

and $x \in R$, which means $u_J^j(x) \geq 0$, $u_J^j(x) g_j(x) = 0$, $\forall j \in J_1 \cup J_2 \supseteq I(x) = \{j \in L \mid g_j(x) = 0\}$, that is, x is a K-T point of (NP). (1) is proved.

(2.7) can be obtained easily from (2.2), (2.4)-(2.6). The proof is complete.

For convenience, we take $\rho(x) = \alpha(J, x) / (|u_J^T(x)\delta_J| + 1)$ in the rest of the paper. It is clear that if $\alpha(J, x) > 0$, then the conditions in Lemma 3 (2) hold.

3. Algorithm and Its Convergence

Algorithm

Step 0 Given $x^0 \in R_1$; $\alpha_1, \alpha_2, \beta \in (0, 1)$, $\delta_0 > 0$; $\{\epsilon_k\}$: $\epsilon_k \geq 0$, $\lim_{k\to\infty}\epsilon_k = 0$. Set $k := 0$.

Step 1 Set $J_{i,k} = I_i(x^k, \delta_k)$, $i = 1, 2$; $J_k = J_{1,k} \cup J_{2,k}$.

Step 2 If $\det N_{J_k}(x^k)^T N_{J_k}(x^k) \geq \delta_k$, then go to Step 3; else, set $\delta_k = \alpha_1 \delta_k$ go back to Step 1.

Step 3 Compute $\alpha(J_k, x^k)$. If $\alpha(J_k, x^k) = 0$, stops; else, go to Step 4.

Step 4 Compute $d^k = d_{J_k}(x^k)$ by the formula (2.4).

Step 5 Compute stepsize $t_k > 0$ by one of the following rules:

Rule 1 $x^k + t_k d^k \in R_1$, $f(x^k + t_k d^k, x^k) \leq f(x^k, x^k) = 0$, and

$$f(x^k + t_k d^k, x^k) \leq \min_{t \geq 0}\{f(x^k + td^k, x^k) \,|\, x^k + td^k \in R_1\} + \epsilon_k;$$

Rule 2 $x^k + t_k d^k \in R_1$, $f(x^k + t_k d^k, x^k) \leq f(x^k, x^k) = 0$, $|t_k - t_k^*| \leq \epsilon_k$, where t_k^* is an optimal solution of the problem $\min_{0 \leq t \leq \tau}\{f(x^k + td^k, x^k) \,|\, x^k + td^k \in R_1\}$, $\tau > 0$;

Rule 3 $t_k = \max\{t \in \Gamma \,|\, f(x^k + td^k, x^k) \leq f(x^k, x^k) + \alpha_2 t f^*_{J_{2,k}}(x^k, x^k; d^k),\ x^k + td^k \in R_1\}$, where $\Gamma = \{\beta^0, \beta^1, \beta^2, \cdots\}$.

Step 6 Set $x^{k+1} := x^k + t_k d^k$, $\delta_{k+1} := \delta_k$ or δ_0, $k := k + 1$, go back to Step 1.

Lemma 3.1 *After entering Step 1 from Step 2 finite times, the algorithm must go to Step 3.*

Lemma 3.2 *If the algorithm generates infinite sequence $\{x^k\}$ which has a cluster point x^*, then* $\lim_{k\to+\infty} f(x^{k+1}, x^k) = \lim_{k\to+\infty}[f(x^{k+1}, x^k) - f(x^k, x^k)] = 0$.

Theorem 3.1 *The algorithm either stops at a K-T point x^k of (NP) after finite iterations or generates an infinite sequence $\{x^k\}$ of which each cluster point is a K-T point of (NP).*

Proof By Lemma 2.3, we need only to prove the second conclusion.

Let x^* be a cluster of $\{x^k\}$ such that $\{x^k\} \xrightarrow{K} x^*$, then by Lemma 2.2, there exists $\epsilon > 0$ such that $\delta_k \geq \epsilon > 0$, $\forall k \in K$. Since $J_{1,k}, J_{2,k}$ are the subsets of the finite index sets L_1, L_2 respectively, we can assume that $J_{i,k} = J_i$, $\forall k \in K$ $(i = 1, 2)$, thus

$$J_k = J = J_1 \cup J_2,\ \forall k \in K, \text{ and } J_i \supseteq I_i(x^*),\ i = 1, 2;\ J \supseteq J(x^*).$$

Since $f, g_j \in C^1 (j \in L)$, we have

$$P_J(x^k)\nabla f(x^k) \xrightarrow{K} P_J(x^*)\nabla f(x^*), B_J(x^k) \xrightarrow{K} B_J(x^*), u_J(x^k) \xrightarrow{K} u_J(x^*). \tag{3.1}$$

Furthermore, $\{v_J(x^k)\}$ is bounded, hence there exists an infinite subset K_1 of K such that $v_J(x^k) \xrightarrow{K_1} v_J$. Therefore, $\alpha(J, x^k) \xrightarrow{K_1} \alpha(J, x^*) = \|P_J(x^*)\nabla f(x^*)\|^2 + u_J(x^*)^T v_J(x^*) + r\varphi(x^*) \geq 0$, $\rho(x^k) \xrightarrow{K_1} \rho(x^*) = \alpha(J, x^*)/(|u_J(x^*)^T v_J| + 1)$ and $d^k = d_J(x^k) \xrightarrow{K_1} d^* = d_J(x^*) = -P_J(x^*)\nabla f(x^*) + B_J(x^*)[v_J - \rho(x^*)\delta_J]$.

Now, we prove that $\alpha(J, x^*) = 0$.

If $\alpha(J, x^*) > 0$, then $\rho(x^*) > 0$ and $-\alpha(J, x^*) + \rho(x^*) u_J(x^*)^T \delta_J < 0$. By (2.4)-(2.6), we have

$$g_j(x^k) + \nabla g_j(x^k)^T d^k \leq -\rho(x^k)\delta_j, \quad \forall k \in K, j \in J_1,$$
$$g_j(x^k) + \nabla g_j(x^k)^T d^k - \varphi(x^k) \leq -\rho(x^k)\delta_j, \forall k \in K, j \in J_2.$$

Let $k \xrightarrow{K_1} +\infty$, we obtain

$$g_j(x^*) + \nabla g_j(x^*)^T d^* \leq -\rho(x^*)\delta_j, \forall j \in J_1, g_j(x^*) + \nabla g_j(x^*)^T d^* - \varphi(x^*) \leq -\rho(x^*)\delta_j, \forall j \in J_2.$$

Therefore, similar way to the proof of Lemma 2.3 (2), we have

$$f_{J_2}^*(x^*, x^*; d^*) < 0, \nabla g_j(x^*)^T d^* < 0, \forall j \in I_1(x^*) \subseteq J_1, a_i^T d^* = 0, \forall i \in M. \tag{3.2}$$

From (3.2), we can easily prove that

$$\exists \delta > 0, \forall t \in (0, \delta), \exists k_t \text{ such that } x^k + td^k \in R_1, \forall k \geq k_t, k \in K_1. \tag{3.3}$$

According to the definition of t_k, we deduce a contradiction for Rule 1-3 respectively.

Case 1 t_k is defined by Rule 1.

For fixed $t \in (0, \delta]$, by Lemma 2.1, $\exists \theta_t^k \in (0, 1)$ such that

$$\begin{aligned} f(x^{k+1}, x^k) - f(x^k, x^k) &\leq f(x^k + td^k, x^k) - f(x^k, x^k) + \epsilon_k \\ &\leq tf'(x^k + \theta_t^k td^k, x^k; d^k) + \epsilon_k, \quad \forall k \geq k_t, k \in K_1. \end{aligned}$$

Without loss of generality, assume that $\theta_t^k \xrightarrow{K_1} \theta_t \in [0, 1]$. By Lemma 3.2 and 2.1, we get

$$\begin{aligned} 0 = \lim_{K_1} t^{-1}(f(x^{k+1}, x^k) - f(x^k, x^k)) &\leq \overline{\lim_{K_1}}[f'(x^k + \theta_t^k td^k, x^k; d^k) + \epsilon_k] \\ &\leq f'(x^* + \theta_t td^*, x^*; d^*). \end{aligned}$$

Hence, $0 \leq \overline{\lim}_{t\to 0^+} f'(x^* + \theta_t td^*, x^*; d^*) \leq f'(x^*, x^*; d^*) \leq f_{J_2}^*(x^*, x^*; d^*)$, this contradicts (3.2).

Case 2 t_k is defined by Rule 2.

From the definition of t_k and (3.3), for fixed $t \in (0, \delta) \cap (0, \tau]$, $k \geq k_t$, $k \in K_1$, we have

$$\begin{aligned} & f(x^{k+1}, x^k) - f(x^k, x^k) \\ &= f(x^k + t_k d^k, x^k) - f(x^k + t_k^* d^k, x^k) + f(x^k + t_k^* d^k, x^k) - f(x^k, x^k) \\ &\leq f(x^k + t_k d^k, x^k) - f(x^k + t_k^* d^k, x^k) + f(x^k + t d^k, x^k) - f(x^k, x^k). \end{aligned}$$

Let $f(x^k + t_k d^k, x^k) - f(x^k + t_k^* d^k, x^k) = \epsilon_k^*$, we can show that $\overline{\lim}_{K_1} \epsilon_k^* \leq 0$. Thus, replacing ϵ_k in Case 1 for ϵ_k^*, we still obtain that $f_{J_2}^*(x^*, x^*; d^*) \geq 0$, which contradicts (3.2).

Case 3 t_k is defined by Rule 3.

In this case, one and only one of the following cases will occur:

(i) $\exists\, t^* \in (0, \delta] \cap \Gamma$ and $k^* \geq k_{t^*}$ such that

$$f(x^k + t^* d^k, x^k) - f(x^k, x^k) \leq \alpha_2 t^* f_{J_2}^*(x^k, x^k; d^k), \ \forall k \geq k^*, k \in K_1;$$

(ii) $\forall t \in (0, \delta] \cap \Gamma$, $\exists\, K_t \subseteq K_1$ such that

$$f(x^k + t d^k, x^k) - f(x^k, x^k) > \alpha_2 t f_{J_2}^*(x^k, x^k; d^k), \ \forall k \in K_t.$$

If (i) happens, then by the definition of t_k, $t_k \geq t^* > 0$, $\forall k \geq k^*$, $k \in K_1$ and

$$f(x^{k+1}, x^k) - f(x^k, x^k) \leq \alpha_2 t_k f_{J_2}^*(x^k, x^k; d^k) \leq \alpha_2 t^* f_{J_2}^*(x^k, x^k; d^k), \ \forall k \geq k^*, k \in K_1,$$

hence by Lemma 3.2, we have

$$0 = \lim_{K_1} [f(x^{k+1}, x^k) - f(x^k, x^k)] / \alpha_2 t^* \leq \lim_{K_1} f_{J_2}^*(x^k, x^k; d^k) = f_{J_2}^*(x^*, x^*; d^*).$$

If (ii) happens, then by Lemma 2.1, $\exists\, \theta^k \in (0, 1)$ such that

$$\alpha_2 t f_{J_2}^*(x^k, x^k; d^k) < f(x^k + t d^k, x^k) - f(x^k, x^k) \leq t f'(x^k + \theta^k t d^k, x^k; d^k), \ \forall k \in K_t.$$

Similar to the proof of Case 1, we have $\alpha_2 t f_{J_2}^*(x^*, x^*; d^*) \leq t f'(x^* + \theta t d^*, x^*; d)$, hence

$$\alpha_2 f_{J_2}^*(x^*, x^*; d^*) \leq \overline{\lim_{t \to 0^+}} f'(x^* + \theta t d^*, x^*; d^*) \leq f'(x^*, x^*; d^*) \leq f_{J_2}^*(x^*, x^*; d^*),$$

i.e., $(1 - \alpha_2) f_{J_2}^*(x^*, x^*; d^*) \geq 0$, which implies that $f_{J_2}^*(x^*, x^*; d^*) \geq 0$ since $\alpha_2 \in (0, 1)$.

We also deduce a contradiction if Case 3 happens.

From the discussions above, we obtain that $\alpha(J, x^*) = 0$, hence

$$\varphi(x^*) = 0, \quad u_J(x^*)^T v_J = 0, \quad P_J(x^*) \nabla f(x^*) = 0. \tag{3.4}$$

Consider the function $h(u, g, \varphi): R^3 \to R^1$ defined as

$$h(u, g, \varphi) = \begin{cases} u^2, & u < 0 \\ -ug, & u \geq 0, g \leq 0 \\ u(\varphi - g), & u \geq 0, g > 0 \end{cases}$$

It is clear that h is continuous at $(u, g, 0)$, hence

$$u_J^j(x^k)v_J^j(x^k) = h(u_J^j(x^k), g_j(x^k), \varphi(x^k)) \xrightarrow{K_1} h(u_J^j(x^*), g_j(x^*), 0)$$
$$= \begin{cases} u_J^j(x^*)^2, & u_J^j(x^*) < 0 \\ -u_J^j(x^*)g_j(x^*), & u_J^j(x^*) \geq 0 \end{cases} = u_J^j(x^*)v_J^j(x^*) = u_J^j(x^*)v_J^j,$$

since $\varphi(x^*) = 0$.

Therefore, $u_J(x^*)v_J = 0$ i.e. $u_J^j(x^*)v_J^j(x^*) = u_J^j(x^*)v_J^j = 0, \forall j \in J$ implies that

$$u_J^j(x^*) \geq 0, \ u_J^j(x^*)g_j(x^*) = 0, \forall j \in J. \tag{3.5}$$

From (3.4) and (3.5), we know that x^* is a K-T point of (NP). The theorem is true.

Remark When we take $\delta_{J_k}^j = 0$ for linear constraint $g_j(j \in L_1)$ or use the curvilinear search instead of the line search in the algorithm, the results are still true.

References

[1] E. Polak, *Computational methods in optimization: A Unified Approach*, New York: Academic Press, 1971.

[2] E. Polak and L. He, *Unified steerable phase I-phase II method of feasible directions for semi-infinite optimization*, JOTA, **69**: **1** (1991), 83-107.

[3] F. H. Clarke, *Optimization and nonsmooth analysis*, New York: Wiley-Inter-Science, 1983.

[4] Chen Guangjun, *A gradient projection algorithm for optimization problems with general constraints*, J. Comput. Math. (in Chinese), **9**: **4** (1987), 356-364.

[5] Shi Baochang, *A family of perturbed gradient projection algorithms for nonlinear constraints*, Acta Math. Appl., Sinica (in Chinese), **12**: **2** (1989), 190-195.

解约束优化问题的投影梯度型中心方法

施保昌　周晓阳

(华中理工大学数学系, 武汉430074)

胡 新 生

(深圳广播电视大学, 深圳518008)

摘　要

本文提出了一种求解约束优化问题的新算法—投影梯度型中心方法. 在连续可微和非退化的假设条件下, 证明了其全局收敛性. 本文算法计算简单且形式灵活.