Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 6 (5): 1085-1089, 2006
ISSN 1812-5654

A General Projection Gradient Method for Linear Constrained Optimization with Superlinear Convergence

Zhibin Zhu and Binliang Zhang
Department of Computational Science and Mathematics,
Guilin Institute of Electronic Technology, Guilin, 541004, People's Republic of China

Abstract: In this study, by using the technique of the projection and the idea of the conjugate projection, a new algorithm is presented to solve the linear equality and inequality constrained optimization. Under some suitable conditions which are weaker than those in corresponding references, it is proved that the proposed method is global convergence as well as superlinear convergence.

Key words: Linear constrained optimization, conjugate projection gradient algorithm, global convergence, superlinear convergence

INTRODUCTION

We consider the following linear constrained optimization:

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t} \quad & x \in X \end{aligned} \tag{1}$$

where:

$$X = \{x \in R^n \mid f_j(x) = a^T_j x - b_j \le 0, j \in I;\ f_j(x) = a^T_j x - b_j = 0, j \in E\},$$
$$I \cup E = \{1, 2, ..., m\},\ I \cap E = \Phi,\ I(x) = \{j \in I \mid f_j(x) = 0\}.$$

Since the gradient projection method was proposed by Rosen (1960), it become one of basic methods to solve nonlinear programming, and some authors were absorbed in research on this method (Zhang, 1979; Jian and Zhang, 1999). However, being lack of the information of twice derivatives, this type of methods converges slowly.

In order to quicken the rate of convergence, recently, it is arisen some improved algorithms (Han, 1976; Panier and Tits, 1987; Facchinel and Lucidi, 1995). In two references (Shi, 1996; Zhang and Wang, 1999) by generalizing the conjugate projection from the gradient projection, a new projection variable metric algorithm is presented which is combined the penalty function method with the variable metric algorithm. While, even under some strong conditions (for example, the sequence $\{x^k\}$ converges to the optimum solution $\{u^k\}$ and the corresponding multiplier vector sequence $\{u^k\}$ converges to the optimum multipliers u^*), it is only proved that the sequence $\{x^k, u^k\}$ converges superlinearly to (x^*, u^*), instead of the sequence $\{x^k\}$ itself.

In this study, by taking advantage of the projection gradient technique, a new general projection gradient method is present to improve those methods (Shi, 1996; Zhang and Wang, 1999). Under some weaker suitable conditions, it is proved that the sequence $\{x^k\}$ generated by the algorithm is superlinear convergent to the optimum solution x^*.

DESCRIPTION OF ALGORITHM

The following assumptions are true throughout the study.

H 1: The feasible set $X \neq \emptyset$ and the function $f_0(x)$ is twice differentiable;

H 2: $\forall x \in X$ vectors $\{a_j, j \in I(x) \cup E\}$ are linear independent.

Definition 1: The function $\mu(x)$: $R^n \to R^m$ is called a multiplier function, if $\mu(x)$ is continuous and $\mu(x^*)$ is the corresponding K-T multiplier vector for the K-T point x^* of (1).

For the current approximate solution $x^k \in X$, $\sigma_k > 0$, a positive definite matrix $B_k = B(x^k)$, the set $L_k \subseteq IUE$, we define:

$$\begin{aligned} F(x^k) &= (f_j(x^k), j \in L_k),\ A_k = A(x^k) = (a_j, j \in L_k) \\ Q_k &= Q(x^k) = (A^T_k B^{-1} A_k)^{-1} A^T_k B^{-1}_k,\ P_k = P(x^k) \\ &= B^{-1}_k (E_n - A_k Q_k) \\ \pi^k &= \pi(x^k) = -Q_k \nabla f_0(x^k),\ d^k_0 = -P_k \nabla f_0(x^k) + Q^T_k V^k \end{aligned} \tag{2}$$

Corresponding Author: Zhibin Zhu, Department of Computational Science and Mathematics,
Guilin Institute of Electronic Technology, Guilin, 541004, People's Republic of China

$$V^k = V(x^k) = (V_j^k, j \in L_k), V_j^k = \begin{cases} -f_j(x^k), \pi_j^k > 0, j \notin E, \\ \pi_j^k, \pi_j^k \le 0, j \notin E, \\ -f_j(x^k), j \notin E, \end{cases} \quad (3)$$

$$d_1^k = -Q_k^T\left(\| d_0^k \|^\tau e + F(x^k + d_0^k)\right), e = (1,\cdots,1) \in R^{|L_k|}, d^k = d_0^k + d_1^k$$

Consider the following auxiliary problem:

$$\min G_c(x) \qquad \text{S.t. } x \in R^n \quad (4)$$

Where

$$G_c(x) = f_0(x) + c\sum_{j \in E} |f_j(x)|$$

define the directional derivative along dat xas follows:

$$DG_c(x,d) = \lim_{t \to 0^+} \frac{G_c(x+td) - G_c(x)}{t}$$

It is easy to see that

$$DG_c(x,d) = \nabla f_0(x)^T d + c \sum_{\substack{f_j(x)>0 \\ j \in E}} a_j^T d - c \sum_{\substack{f_j(x)=0 \\ j \in E}} |a_j^T d| - c \sum_{\substack{f_j(x)<0 \\ j \in E}} a_j^T d$$

The following algorithm is proposed for solving problem (1):

Algorithm A

Step 0:

$$x^1 \in X,\ \mu(x^1) \in R^m,\ B_1 \in R^{n \times n}$$

Parameters $\alpha \in (0, 1/2)$, θ_0, ε, ξ, $\nu \in (0, 1)$

$$\delta > 2, \tau \in (2,3), c_\varepsilon > 0, 0 < \mu_j^1 < \bar{\mu}, j = 1,\cdots, m; \text{ Set } k = 1$$

Step 1: Let $i = 0$, $\sigma_{k,i} = \sigma_0$

Step 2: If $\det(A_k^T A_k) \ge \sigma_{k,i}$, set $L_k = L_{k,i}$, $i_k = i$ and go to Step 4, otherwise, go to Step 3,where

$$L_{k,i} = \{j \in I \mid -\sigma_{k,i} |\mu_j^k| \le f_j(x^k) \le 0\} \cup E,\ A_{k,i} = (a_j, j \in L_{k,i})$$

Step 3: Let $i = i + 1$, $\sigma_{k,i} = 1/2\sigma_{k,i-1}$, go to step 2.

Step 4: Compute

$$t_k = \max\{|\pi_j^k|, j \in E\} + c_\varepsilon,\ c_k = \begin{cases} \max\{t_k, c_{k-1}\}, & c_{k-1} < t_k \\ c_{k-1}, & c_{k-1} \ge t_k \end{cases}$$

Step 5: Compute d_0^k. If $d_0^k = 0$, STOP; Otherwise, compute d_0^k. If

$$DG_{c_k}(x^k, d_0^k) \le \min\left\{-\xi\|d_0^k\|^\delta, -\xi\|d^k\|^\delta\right\} \quad (5)$$

go to Step 6, otherwise goto Step 7;

Step 6: Let $\lambda = 1$.
1) If

$$G_{c_k}(x^k + \lambda d^k) \le G_{c_k}(x^k) + \alpha\lambda DG_{c_k}(x^k, d_0^k) \quad (6)$$
$$f_j(x^k + \lambda d^k) \le 0, j \in I$$

set $\lambda_k = \lambda$, go to Step 8, otherwise go to 2).
2) Let $\lambda = 1/2\lambda$. If $\lambda < \epsilon$, go to Step 7, otherwise go to 1) of Step 6.

Step 7: Compute

$$\rho_k = -DG_{c_k}(x^k, d_0^k), q^k = \rho_k d_0^k \quad (8)$$

Find out β_k, the first number β in the sequence {1, 1/2, 1/4, ...} satisfying

$$G_{c_k}(x^k + \beta q^k) \le G_{c_k}(x^k) + \nu\beta DG_{c_k}(x^k, q^k) \quad (9)$$

$$f_j(x^k + \beta q^k) \le 0, j = I \quad (10)$$

Set $d^k = q^k$, $\lambda_k = \beta_k$.

Step 8: Obtain B_{k+1} by updating the positive definite matrix B_k using some quasi-Newton formulas. Set

$$x^{k+1} = x^k + \lambda_k d^k, \mu_j^{k+1} = \min\left\{\max\left\{\pi_j^k, \|d_0^k\|\right\}, \bar{\mu}\right\}, j = 1,\cdots,m$$

Set $k = k + 1$. Go back to Step 1.

CONVERGENCE OF ALGORITHM

Here, firstly, it is shown that Algorithm A is well defined.

H 3: The sequence $\{x^k\}$ is bounded, and the sequence $\{B_k\}$ is positive definite.

Lemma 1: For any iteration, there is no infinite cycle between Step 1 and Step 3. Moreover, if $\{x^k\}_{k \in K} \to x^*$, then there exits a constant $\bar{\varepsilon} > 0$, such that $\varepsilon_{k,i_k} \ge \bar{\varepsilon}$, for $k \in K$, k large enough.

Proof: The proof is refereed to Lemma 1 in reference (Zhu *et al*., 2003).

Theorem 1: $\forall k$, if $d^k_0 = 0$, then x^k is a K-T point of (1), else, it holds that

$$DG_{c_k}\left(x^k, d_0^k\right) < 0, DG_{c_k}\left(x^k, q^k\right) < 0, a_j^T d_0^k \le 0, a_j^T q^k \le 0, j \in I\left(x^k\right) \tag{11}$$

Proof: Firstly, it is easy to see that

$$P_k A_k = 0, P_k B_k P_k = P_k, Q_k A_k = E_{|L_k|}$$

If $d^k_0 = 0$, it holds that

$$0 = A^T_k d^k_0 = V^k, P_k \nabla f_0(x^k) = 0$$

So, form (2), (3) and H 3, we have

$$\nabla f_0\left(x^k\right) + A_k \pi^k = 0, f_j\left(x^k\right) = 0, j \in E, \pi_j^k f_j\left(x^k\right) = 0, j \in I\left(x^k\right)$$

which shows that x^k is a K-T point of (1).

If $d^k_0 = 0$, form the definition of V^k, it is obvious that

$$a^T_j d^k_0 = -f_j(x^k), j \in E$$

So, it holds that

$$\begin{aligned} DG_{c_k}\left(x^k, q^k\right) &= \nabla f_0\left(x^k\right)^T d_0^k + c_k \sum_{\substack{f_j(x^k)>0 \\ j\in E}} a_j^T d_0^k - c_k \sum_{\substack{f_j(x^k)=0 \\ j\in E}} \left|a_j^T d_0^k\right| - c_k \sum_{\substack{f_j(x^k)<0 \\ j\in E}} a_j^T d_0^k \\ &= -\nabla f_0\left(x^k\right)^T P_k \nabla f_0\left(x^k\right) - \pi^T V^K - c_k \sum_{\substack{f_j(x^k)>0 \\ j\in E}} f_j\left(x^k\right) + c_k \sum_{\substack{f_j(x^k)<0 \\ j\in E}} f_j\left(x^k\right) \\ &= -\nabla f_0\left(x^k\right)^T P_k \nabla f_0\left(x^k\right) + \sum_{\substack{\pi_j^k>0 \\ j\in I(x^k)}} \pi_j^k f_j\left(x^k\right) - \sum_{\substack{\pi_j^k\le 0 \\ j\in I(x^k)}} \left(\pi_j^k\right)^2 \\ &\quad + \sum_{\substack{f_j(x^k)>0 \\ j\in E}} \left(\pi_j^k - c_k\right) f_j\left(x^k\right) + c_k \sum_{\substack{f_j(x^k)<0 \\ j\in E}} \left(\pi_j^k + c_k\right) f_j\left(x^k\right) \end{aligned}$$

Form the definition of c_k, it holds that $DG_{c_k}\left(x^k, d_0^k\right) < 0$ since $A^T_k d^k_0 = V^k$, it holds that

$$a_j^T d_0^k = V_j^k, j \in I\left(x^k\right) \subseteq L_k$$

So, we have

$$DG_{c_k}\left(x^k, q^k\right) = DG_{c_k}\left(x^k, \rho_k d_0^k\right) = \rho_k DG_{c_k}\left(x^k, d_0^k\right) = -\rho_k^2 < 0$$
$$\rho_k > 0, a_j^T q^k = \rho_k V_j^k \le 0, j \in I(x^k)$$

The conclusion holds.

Lemma 2: There exists a constant k_0, such that $c_k \equiv c_{k0} \triangleq c, \forall k \ge k_0$.

In the sequel , we always assume that $c_k \equiv c$.

Theorem 2: The algorithm either stops at the K-T point x^k of the problem (1) in finite iteration, or generates an infinite sequence $\{x^k\}$, any accumulation point x^* of which is a K-T point of the problem (1).

Proof: The first statement is obvious, the only stopping point being step 5. Thus, suppose that $\{x^k\}_{k\in K} \to x^*$, $d^k_0 \to 0$, $k \in K$. From (5), (6), (9) and Theorem 1, it is easy to see that $\{G_c(x^k)\}$ is decreasing. So, it holds that

$$G_c(x^k) \to G_c(x^*), k \to \infty \tag{12}$$

If there exists $K_1 \subseteq K$ ($|K_1| = \infty$), such that for all $k \in K_1$, $x^{k+1} = x^k + \lambda_k d^k$ is generated by step 6 and step 8, then from (5), (6), we get

$$0 = \lim_{k\in K_1}\left(G_c\left(x^{k+1}\right) - G_c\left(x^k\right)\right) \le \lim_{k\in K_1} \alpha\lambda_k DG_c\left(x^k, d_0^k\right) \le \lim_{k\in K_1}\left(-\alpha\varepsilon\xi\left\|d_0^k\right\|^\delta\right) \le 0$$

So, $d^k_0 \to 0$, $k \in K_1$ since $d^k_0 \to d^*_0$, $k \in K$ it is clear that $d^*_0 = 0$ i.e., $d^k_0 \to 0$, $k \in K$. So, according to Theorem 1, it is obvious that x^k is a K-T point of (1).

Now, we might as well assume that, for all $k \in K$, $x^{k+1} = x^k + \lambda_k d^k$ is generated for by step 7 and step 8, Suppose that the desired conclusion is false, i.e., $d^k_0 \ne 0$. Imitating the proof of Theorem 1,we have $DG_c(x^*, q^*) < 0$ and we can conclude that the step-size β_k obtained by the linear search in step 7 is bounded away from zero on K, i.e.,

$$\beta_k \geq \beta_* = \inf\{\beta_k, k \in K\} > 0, k \in K$$

So, from (9) and Theorem 1, it holds that

$$0 = \lim_{k\in K}\left(G_c\left(x^{k+1}\right) - G_c\left(x^k\right)\right) \leq \lim_{k\in K} v\beta_k DG_c\left(x^k, q^k\right)$$
$$\leq \frac{1}{2} v\beta_* DG_C\left(x^*, q^*\right) < 0$$

It is a contradiction, which shows that $d^k_0 \to 0, k \in K$, $k \to \infty$. So, according to Theorem 1. It is easy to see that x^k is a K-T point of (1).

In order to obtain superlinear convergence, we also make the following additional assumptions.

H 4: The sequence generated by the algorithm possesses an accumulation point x^*.

H 5: $B_k \to B_*, k \to \infty$.

H 6: The second-order sufficiency conditions with strict complementary slackness are satisfied at the K-T point x^* and the corresponding multiplier vector u^*.

According to Lemma 5 in reference (Zhu, 2005), we have the following conclusion.

Lemma 3: The entire sequence $\{x^k\}$ converges to x^*, i.e., $x^k \to x^*, k \to \infty$ and for k large enough , it holds that

$$L_k \equiv I\left(x^*\right) \bigcup E, d^k_0 \to 0, \pi^k \to \left(u^*_j, j \in I\left(x^*\right) \bigcup E\right), k \to \infty$$

Lemma 4: Denote $\tilde{u}^k = \pi^k + (A^T_k B^{-1}_k A_k)^{-1} F(x^k)$. Under above mentioned conditions, for k large enough , it holds that

$$\nabla f_0(x^k) + B_k d^k_0 + A_k \tilde{u}^k = 0, F(x^k) + A^T_k d^k_0$$

Proof: According to Lemma 3 and H 6, it holds, for k large enough, that $\pi^k > 0$. So, from the definition of V^k, we have

$$A^T_k d^k_0 = V^k = -F(x^k), F(x^k) + A^T_k d^k_0 = 0$$

While,

$$\begin{aligned}\nabla f_0(x^k) + B_k d^k_0 + A_k \tilde{u}^k = {} & \nabla f_0(x^k) - (E_n - A_k Q_k)\nabla f_0(x^k) + \\ & A_k(A^T_k B^{-1}_k A_k)^{-1} V^k + \\ & A_k(\pi^k + (A^T_k B^{-1}_k A_k)^{-1} F(x^k)) \\ = {} & -A_k\pi^k - A_k(A^T_k B^{-1}_k A_k)^{-1} F(x^k) + \\ & A_k(\pi^k + (A^T_k B^{-1}_k A_k)^{-1} F(x^k)) \\ = {} & 0\end{aligned}$$

The conclusion holds.

Lemma 5: For k large enough, there exists a constant b>0, such that

$$DG_c\left(x^k, d^k_0\right) \leq -b\left\|d^k_0\right\|^2, \left\|d^k\right\| \sim \left\|d^k_0\right\|, \left\|d^k_1\right\| = o\left(\left\|d^k_0\right\|^2\right)$$

Proof : Since $x^k \to x^*$and for k large enough, $L_k \equiv I(x^* U) E$, it holds that

$$F\left(x^k\right) \to \left(f_j\left(x^*\right), j \in L_k\right) = 0, \tilde{u} \to \pi^*, k \to \infty$$

Thereby, there exists some $\eta>0$, such that

$$\sum_{j\in L_k} \tilde{u}^k_j f_j\left(x^k\right) \leq \sum_{j\in L_k} \tilde{u}^k_j \left|f_j\left(x^k\right)\right| \leq \eta\left\|F\left(x^k\right)\right\|$$

while, from Lemma 4 it holds that

$$\begin{aligned}DG_c\left(x^k, d^k_0\right) &= \nabla f_0\left(x^k\right)^T d^k_0 + \sum_{\substack{f_j(x^k)>0\\ j\in E}} a^T_j d^k_0 - \\ &\quad c\sum_{\substack{f_j(x^k)=0\\ j\in E}} \left|a^T_j d^k_0\right| - c\sum_{\substack{f_j(x^k)<0\\ j\in E}} a^T_j d^k_0 \\ &= \nabla f_0\left(x^k\right)^T d^k_0 - c\sum_{\substack{f_j(x^k)>0\\ j\in E}} f_j\left(x^k\right) + c\sum_{\substack{f_j(x^k)<0\\ j\in E}} f_j\left(x^k\right) \\ &\leq \nabla f_0\left(x^k\right)^T d^k_0 = -\left(d^k_0\right)^T B_k d^k_0 - \sum_{j\in L_k} \tilde{u}^k_j f_j\left(x^k\right) \\ &\leq -b\left\|d^k_0\right\|^2 - \eta\left\|F\left(x^k\right)\right\| \leq -b\left\|d^k_0\right\|^2\end{aligned}$$

In addition, it holds, for k large enough, that $\pi^k_j > 0$, $A^T_k d^k_0 = V^k = -f^j(x^k), j \in L_k$. So

$$f_j\left(x^k + d^k_0\right) = f_j\left(x^k\right) + a^T_k d^k_0 =$$
$$0, F\left(x^k + d^k_0\right) = 0, d^k_1 = -Q^T_k\left\|d^k_0\right\|^\tau e$$

From τ (2, 3), we have $\|d^k\| \sim \|d^k_0\|$, $\|d^k_1\| = o(\|d^k_0\|^2)$.

In order to obtain superlinear convergence, a crucial requirement is that a unit step size be used in a neighborhood of the solution. This can be achieved if the following assumption is satisfied.

H 7: Let

$$\left\|\tilde{P}_k\left(B_k - \nabla^2_{xx} f_0\left(x^k\right)\right) d^k_0\right\| = o\left(\left\|d^k_0\right\|\right)$$

where

$$\tilde{P}_k = E_n - A_k\left(A_k^T A_k\right)^{-1} A_k^T$$

In view of Lemma 4, imitating the proof of Lemma 4.4 in reference (Zhu, 2005), it is easy to obtain the following conclusion.

Lemma 6: For k large enough, step 7 is no longer performed in the algorithm and the attempted search in step 6 is successful in every iteration, i.e., $\lambda_k \equiv 1$, $x^{k+1} = x^k + d^k$.

Moreover, in view of Lemma 3.8 and the way of Theorem 2 in reference (Panier and Tits, 1987), we may obtain the following theorem:

Theorem 3: Under all above-mentioned assumptions, the algorithm is superlinearly convergent, i.e., the sequence $\{x^k\}$ generated by the algorithm satisfies

$$\left\|x^{k+1} - x^*\right\| = o\left(\left\|x^k - x^*\right\|\right)$$

ACKNOWLEDGMENT

This study was supported in part by the NNSF (10501009, 60471039) of China.

REFERENCES

Facchinel, F. and S. Lucidi, 1995. Quadratically and superlinearly convergent algorithms for the solution of inequality constrained minimization problems. JOTA, 85: 265-289.

Han, S.P., 1976. Superlinearly convergent variable metric algorithm for general nonlinear programming problem. Math Programming, 11: 263-282.

Jian, J.B. and K.C. Zhang, 1999. Subfeasible direction method with strong convergence for Inequality constrained optimization. J. Xi'an Jiaotong Univ., 33: 88-92.

Panier, E.R. and A.L. Tits, 1987. A Superlinearly convergent feasible method for the solution of inequality constrained optimization problems. SIAM J. Control and Optimization, 25: 934-950.

Rosen, J.B., 1960. The gradient projection method for nonlinear programming. Part I, Linear constraint. SIAM J. Applied Math., 8: 181-217.

Shi, Z.J., 1996. A class of global convergent conjugate projection gradient method and its superlinear convergence rate. Math. Numerica Sin., 4: 411-421.

Zhang, X.S., 1979. An improved Rosen-Polak method. Acta Math. Sin., 2: 257-267.

Zhang, J.L. and C. Wang, 1999. A new conjugate projection gradient method. OR Transactions, 2: 61-70.

Zhu, Z.B., K.C. Zhang and J.B. Jian, 2003. An improved SQP algorithm for inequality constrained optimization. Mathematical Methods of Operations Research, 58: 271-282.

Zhu, Z.B. 2005. A globally and superlinearly convergent feasible QP-free method of nonlinear programming. Applied Math. Comp., 168: 519-539.