广西科学 Guangxi Sciences 2007,14(3):236～238,243

非线性约束优化的一个共轭投影梯度法及其全局收敛性*

A Conjugate Projection Gradient Algorithm with Global Convergence for Nonlinear Constrained Optimization

杨晓辉,朱志斌,唐清干
YANG Xiao-hui,ZHU Zhi-bin,TANG Qing-gan

(桂林电子科技大学计算科学与数学系,广西桂林 541004)
(Department of Computational Science and Mathematics, Guilin University of Electronic Technology,Guilin,Guangxi,541004,China)

摘要:针对非线性等式和不等式约束优化问题,结合罚函数法,提出一个共轭投影梯度法,并证明该方法的全局收敛性,给出有效的数值实验.

关键词:约束优化 共轭投影梯度 全局收敛 罚函数法
中图法分类号:O221.2 **文献标识码**:A **文章编号**:1005-9164(2007)03-0236-03

Abstract: A conjugate project gradient algorithm is presented to solve nonlinear equality and inequality constraints optimization by combining with the method of penalty function. Under some suitable conditions, the global convergence is obtained. Numerical experiment shows that the given method is efficient.

Key words: constrained optimization, conjugate projection gradient, global convergence, method of penalty function

对于约束优化问题(P):

$$\begin{cases}\min f_0(x);\\ s.t.\ x\in X.\end{cases}$$

其中,$X=\{x\in\mathscr{R}^n \mid f_j(x)\leqslant 0, j\in I\cup L\}$,$I=\{1,2,\cdots,m\}$,$L=\{m+1,m+2,\cdots,m+p\}$,自1960年Rosen提出梯度投影法[1]后,该方法一直作为求解非线性规划问题的基本解法之一,受到广泛的讨论,并出现了大量改进的有效算法[2~8].这类方法有的是最速下降法对约束问题的推广,但收敛速度不太理想.为了加快算法的收敛速度,文献[3]将梯度投影法的正交投影推广为其共轭投影,对仅含等式约束优化问题进行研究,结合罚函数法与变尺度法提出了一个投影变尺度法.文献[4]对非线性不等式约束优化的投影梯度法进一步改进,提出了一个具有强收敛的广义投影梯度算法.本文吸取文献[3,4]的思想,进一步讨论带等式和不等式约束的非线性最优化问题,提出一个具有全局收敛的修正共轭投影梯度算法.

收稿日期:2007-02-11
修回日期:2007-05-25
作者简介:杨晓辉(1972-),女,硕士研究生,讲师,主要从事最优化理论及其应用研究。
*国家自然科学基金项目(No.10501009)资助。

1 算法及其理论基础

对问题(P)作如下基本假设:

H_1:可行集 $X\neq\varnothing$,$f_j(x)(j=0,1,\cdots,m+p)$ 至少二阶连续可微;

H_2:$\forall\, x\in X$,向量组$\{\nabla f_j(x)\ j\in J(x)\}$线性无关,其中 $J(x)=\{j|j\in I|\ f_j(x)=0\}\cup L$;

H_3:$\{x^k\}$ 有界,矩阵 $B_k=B(x^k)$ 一致正定.

下面给出与σ-有效集有关的一类乘子函数的定义.

定义1 函数 $u(x)$: $\mathscr{R}^n\rightarrow\mathscr{R}^{m+p}$ 称为问题(P)的乘子函数,若 $u(x)$ 连续,且当 x^* 为问题(P)的 K-T 点,$u(x^*)$ 为相应的 K-T 乘子.

结合罚函数技巧,针对问题(P),在任一点 $x\in\mathscr{R}^n$ 处,构造如下罚函数

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

其中 $c > 0$ 为参数(罚因子). 考虑如下辅助问题(AP)

$$\begin{cases}\min G_c(x);\\ s.t. f_j(x) \leqslant 0, j \in I.\end{cases}$$

对当前近似解 $x^k \in X, \sigma_k > 0$,正定矩阵 $B_k = B(x^k)$,如下定义各量:

$J_k = J_\sigma(x^k) = \{j \in I \mid -\sigma|u_j(x^k)| \leqslant f_j(x) \leqslant 0\}, L_k = J_k \cup L$;

$g_j^k = g_j(x^k) = \nabla f_j(x^k), H_j^k = H_j(x^k) = \nabla^2 f_j(x^k), j \in I \cup L$;

$F(x^k) = (f_j(x^k), j \in L_k), A_k = A(x^k) = (g_j(x^k), j \in L_k)$; (1)

$Q_k = Q(x^k) = (A_k{}^T B_k{}^{-1} A_k)^{-1} A_k{}^T B_k{}^{-1}$,

$P_k = P(x^k) = B_k{}^{-1}(E_n - A_k Q_k)$; (2)

$\pi^k = \pi(x^k) = -Q_k g_0(x^k), d_0^k = d_0(x^k) = -P_k g_0(x^k) + Q_k^T V^k$; (3)

$V^k = V(x^k) = (V_j{}^k, j \in J^*)^T, V_j^k =$

$$\begin{cases}-f_j(x^k), \pi_j^k > 0 \text{ 且 } j \in J_k,\\ \pi_j^k, \quad \pi_j{}^k \leqslant 0 \text{ 且 } j \in J_k, \\ -f_j(x^k), j \in L.\end{cases} \quad (4)$$

$\rho_k = -DG_{c^k}(x^k, d_0^k), d_2^k = \dfrac{-\rho_k}{1 + 2|e^T\pi^k|} Q_k^T e, q^k = \rho_k(d_0^k + d_2^k)$. (5)

由文献[3]知 $G_c(x)$ 在 x 处沿方向 d 的方向导数为

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

引理 1 $\forall x \in X, d \in E^n$,有

$$DG_c(x,d) = g_0(x)^T d + c \sum_{\substack{j \in I \\ f_j(x) > 0}} A_j^T d + c \sum_{\substack{j \in I \\ f_j(x) = 0}} |A_j^T d| - c \sum_{\substack{j \in I \\ f_j(x) < 0}} A_j^T d.$$

引理 1 的证明参考文献[3].

算法

步骤 0:选 $x^1 \in X, u(x^1) \in \mathscr{R}^{m+p}$, B_1 正定, $\sigma_0, \zeta, \sigma, v \in (0,1), \alpha \in (0, \frac{1}{2}), \tau \in (2,3), \delta > 2, c_\epsilon > 0, k = 1$;

步骤 1:令 $i = 0, \sigma_{k,i} = \sigma_0$;

步骤 2:若 $det(A_k^T A_k) \geqslant \sigma_{k,i}$ 令 $J_k = J_{\sigma_{k,i}} = \{j \in I \mid -\sigma_{k,i}|u_j(x_j^k)| \leqslant f_j(x^k) \leqslant 0\}$,转步骤 3,否则令 $i = i + 1, \sigma_{k,i} = \dfrac{\sigma_{k,i-1}}{2}$,转步骤 2;

步骤 3:令

$t_k = \max\{|\pi_j^k|, j \in L\} + c_\epsilon, c_k =$

$$\begin{cases}\max\{t_k, c_{k-1}\}, c_{k-1} < t_k,\\ c_{k-1}, \qquad c_{k-1} \geqslant t_k.\end{cases}$$

步骤 4:按(1)、(2)、(3)、(4) 式计算 $d_0^k, DG_{c_k}(x^k, d_0^k)$,若 $DG_{c_k}(x^k, d_0^k) = 0$,则 x^k 为问题(P)之 $K\text{-}T$ 点,停;

步骤 5:作非精确线搜索,由(5)式计算 q^k,求 $\{1, \frac{1}{2}, \frac{1}{4}, \cdots\cdots\}$ 中满足以下各式:

$$G_{c_k}(x^k + \beta q^k) - G_{c_k}(x^k) \leqslant v\beta DG_{c_k}(x^k, q^k), \quad (6)$$

$$f_j(x^k + \beta q^k) \leqslant 0, j \in I, \quad (7)$$

的最大值 β_k,令 $d^k = q^k, \lambda_k = \beta_k$;

步骤 6:令 $x^{k+1} = x^k + \lambda_k d^k$,用某种方法修正 $B_{k+1}, k = k + 1$,返回步骤 1.

引理 2 算法的步骤 2 不会产生无限次循环.

证明 反证法,假设算法在步骤 2 循环无限多次,有 $det(A_k^T A_k) \leqslant \frac{1}{2^i}\sigma_{k,i}, i = 1,2,\cdots\cdots$ 由 $J_{\sigma_{k,i}}$ 定义可知 $J_{\sigma_{k,i}} \supseteq J_{\sigma_{k,i+1}}$. 注意到集合 I 的子集只有有限个,故对充分大的 i,必有 $J_{\sigma_{k,i}} = J_{\sigma_{k,i+1}}$. 不妨记 $J_k{}^*$. 现令 $i \to \infty$ 由此, $det(A_k{}^T A_k) = 0, J_k{}^* = J^*$. 这与假设 H_2 中 $\nabla f_j(x) j \in J^*$ 线性无关矛盾. 本引理得证.

2 算法的全局收敛性

由算法的步骤 3 中 c_k 的定义易知以下结论成立.

引理 3 若算法产生的点列 $\{x^k\}$ 包含于某一紧集 X 中,且假设 H_1, H_2, H_3 成立,则 $\exists k_0$,当 $k \geqslant k_0$ 时, $c_k \equiv c$.

引理 3 的证明参考文献[3].

定理 1 (1) 若 $x^k \in X, DG_{c_k}(x^k, d_0^k) = 0 \Leftrightarrow x^k$ 为问题(P) 的 $K\text{-}T$ 点;

(2) 若 x^k 不是问题(P) 的 $K\text{-}T$ 点,则 $DG_{c_k}(x^k, d_0^k) \neq 0$,有

$DG_{c_k}(x^k, d_0^k) < 0, (g^k)^T d_0^k \leqslant 0, j \in J(x^k)$,

$DG_{c_k}(x^k, q^k) < 0, (g^k)^T q^k < 0, j \in J(x^k)$.

证明 先证明定理 1(1) 成立. 由引理 1 得

$$DG_{c_k}(x^k, d_0^k) = g_0(x^k)^T d_0^k + c_k \sum_{\substack{j \in L \\ f_j(x) > 0}} A_j^T d_0^k + c_k \sum_{\substack{j \in L \\ f_j(x) = 0}} |A_j^T d_0^k| - c_k \sum_{\substack{j \in L \\ f_j(x) < 0}} A_j^T d_0^k. \quad (8)$$

由简单推理知:

$Q_k A_k = E,\ P_k A_k = 0, P_k B_k P_k = P_k$,

$A_j^T d_0^k = A_j^T(-P_k g_0(x^k) + Q_k^T V^k) = V^k =$

$$\begin{cases}-f_j(x^k), \pi_j{}^k > 0, j \in J_k,\\ \pi_j^k \quad \pi_j^k \leqslant 0, j \in J_k, \\ -f_j(x^k), j \in L.\end{cases}$$

$g_0(x^k)^T d_0^k = -g_0(x^k)^T P_k g_0^k + g_0(x^k) Q_k^T V^k = -g_0(x^k)^T P_k g_0^k - \pi_j^T V^k$.

故

$DG_{c_k}(x^k, d_0^k) = -g_0(x^k)^T P_k g_0^k - \sum_{\substack{j \in J_k \\ \pi_j^k \leqslant 0}} (\pi_j^k)^2 + \sum_{\substack{j \in J_k \\ \pi_j^k > 0}} \pi_j^k f_j(x^k) - \sum_{\substack{j \in L \\ f_j(x^k) > 0}} (\pi_j^k + c_k) f_j(x^k) + \sum_{\substack{j \in L \\ f_j(x^k) < 0}} (c_k - \pi_j^k) f_j(x^k).$ (9)

由 c_k 的定义可知 $c_k \geqslant |\pi_j^k|$. 由(9)式知, $DG_{c^k}(x^k, d_0^k) \leqslant 0$. 若 $DG_{c^k}(x^k, d_0^k) = 0$, 则必有 $P_k g_0(x^k) = 0$, $f_j(x^k) = 0, j \in L_k, \pi_j^k f_j(x^k) = 0, \pi_j^k \geqslant 0, j \in L_k$. 由 $P_k g_0(x^k) = 0$, B_k 正定知 $g_0(x^k) + A_k \pi_j^k = 0, \pi_j^k \geqslant 0$, $j \in L_k$. 故有 $g_0(x^k) + \sum_{j \in I} \pi_j^k A_k + \sum_{j \in L} \pi_j^k A_k = 0$, $\pi_j^k f_j(x^k) = 0, j \in I \cup L, \pi_j^k \geqslant 0, j \in I$.

这说明 x^k 为问题(P)的 K-T 点.

反之, 若 x^k 是问题(P)的 K-T 点, 则由 L_k 的定义知, 相应与约束部分 $I \cup L \backslash L_k$ 的 K-T 乘子为0. 故存在向量 $\alpha = (\alpha_j, j \in L_k)$ 使

$$g_0(x^k) + A_k \alpha = 0, \alpha_j \geqslant 0, \alpha_j f_j(x^k) = 0, j \in L_k. \tag{10}$$

由算法的步骤2及 B_k 正定知 $(A_k^T B_k^T A_k)^{-1}$ 存在, 故

$\alpha = -(A_k^T B_k^T A_k)^{-1} A_k^T B_k^{-1} g_0(x^k) = -Q_k g_0(x^k) = \pi^k$. 由(4)式和(10)式知

$V^k = 0, B_k^{-1} g_0(x^k) - B_k^{-1} A_k Q_k g_0(x^k) = 0$, $P_k g_0(x^k) = 0$, 所以 $d_0^k = 0$, 从而 $DG_{c_k}(x^k, d_0^k) = 0$.

然后证明定理1(2)成立. 由(9)式知 $DG_{c_k}(x^k, d_0^k) \leqslant 0$. 又由定理1(1)知 $DG_{c_k}(x^k, d_0^k) \neq 0$, 故 $DG_{c_k}(x^k, d_0^k) < 0$. 由 $A_k^T d_0^k = V^k$ 有 $(g_j^k)^T d_0^k = V_j^k \leqslant 0$, $j \in L_k, DG_{c_k}(x^k, q^k) = \rho_k DG_{c_k}(x^k, d_0^k + d_2^k) \leqslant -\frac{1}{2}\rho_k^2 < 0$.

由 $A_k^T q^k = \rho_k A_k^T (d_0^k + d_2^k) = \rho_k (V^k - \frac{-\rho_k}{1 + 2|e^T \pi^k|}) \leqslant \frac{-\rho_k^2}{1 + 2|e^T \pi^k|} e$, 知 $(g_j^k)^T q^k \leqslant -\frac{-\rho_k^2}{1 + 2|e^T \pi^k|} < 0, j \in L_k$.

定理2 若 H_1, H_2, H_3 成立, 算法产生的点列 $\{x^k, k \in \mathcal{N}\}$ 包含于某一紧集 X 中, 则算法或有限步终止于问题(P)的 K-T 点, 或产生无穷点列 $\{x^k, k \in \mathcal{N}\}$. 其任意聚点都是问题(P)的 K-T 点.

证明 定理的前半部分已由定理1得到. 设算法产生无穷点列 $\{x^k\}$, x^* 为其任意给定聚点, 不妨再设存在无穷子集 K, 使 $x^k \to x^*, B_k \to B_*, L_k \to L_* \equiv L_0, k \in K$. 其中 L_0 为某一固定指标集. 记 $A_*, Q_*, P_*, \pi^*, V^*, d_0^*, q^*$ 分别为(2)~(5)等各式在 x^* 处以 B_k 代替 $B(x^k)$ 的相应各量, 则 $A_k \to A_*, Q_k \to Q_*, P_k \to P_*, d_0^k \to d_0^*, q^k \to q^*, c_k \equiv c, k \in K, k \to \infty$.

因 $\{G_c(x^k)\}$ 为单调下降序列, 由 $G_c(x)$ 的连续性得 $\{G_c(x^k)\}$. 反设 x^* 不是问题(P)的 K-T 点, 此时, 参考定理1易知

$DG_c(x^*, d_0^*) < 0, \beta_k \geqslant \beta_* = inf\{\beta_k, k \in K\} > 0, k \in K$. 故由(6)式知

$$0 = \lim_{k \in K}(G_c(x^{k+1}) - G_c(x^k)) \leqslant \lim_{k \in K} v \beta_k DG_c(x^k, q^k) \leqslant \frac{1}{2} v \beta_* DG_c(x^*, d_0^*) < 0,$$

得出矛盾, 故 x^* 为问题(P)的 K-T 点.

3 数值实验

为进一步从数值计算的角度分析算法的有效性, 通过 Matlab 6.5 编程将算法在计算机上进行数值实验, 所得结果表明算法是有效的. 这里, $\zeta = 0.001$, $\upsilon = 0.01, \delta = 2.5, \alpha = 0.25, \tau = 2.25, B_1 = I$, B_k 由 BFGS公式修正[9], $B_{k+1} = \mathrm{BFGS}(B_k, s^k, y^k)$, 其中 $s^k = x^{k+1} - x^k$, $\hat{y}^k = \nabla f(x^{k+1}) - \nabla f(x^k) + \sum_{j=1}^m \pi_j^k (\nabla g_j(x^{k+1}) - \nabla g_j(x^k))$, $y^k = \theta \hat{y}^k + (1 - \theta) B_k s^k$, $\theta = \begin{cases} 1, \text{如果} (\hat{y}^k)^T s^k \geqslant 0.2 (s^k)^T B_k s^k, \\ \frac{0.8 (s^k)^T B_k s^k}{(s^k)^T B_k s^k - (\hat{y}^k)^T s^k}, \text{否则}. \end{cases}$

$\min f(x) = (x_1 - 2)^2 + (x_2 - 1)^2$, 问题 $s.t.$ $g_1(x) = \frac{x_1^2}{4} + x_2^2 \leqslant 1, g_2 = (x) x_1 - 2x_2 = -1, x_1, x_2, x_3 \geqslant 0$.

分别给定初始点(2,2), (1,1). 迭代次数分别为10,3. 近似解为0.8229, 0.9114.

参考文献:

[1] ROSEN J B. The gradient projection method for nonlinear programming, Part 1, Linear constraints[J]. SIAM Journal on Applied Mathematics, 1960, 8: 181-217.

[2] 简金宝, 张可村. 不等式约束优化一个具有强收敛的强次可行方向法[J]. 西安交通大学学报, 1999, 33(8): 88-91.

[3] 时贞军. 一类全局收敛的共轭投影梯度法及其超线性收敛性[J]. 计算数学, 1996, 11(4): 411-421.

[4] 朱志斌. 一个新的共轭投影梯度算法及其超线性收敛性[J]. 应用数学学报, 2004, 27(1): 149-161.

[5] PAINIER E R, TITS A L. A superlinearly convergent feasible method for the solution of inequality constrained optimization problems[J]. SIAM J Control and Opti, 1987, 25(4): 934-950.

(下转第243页 Continue on page 243)

参考文献：

[1] DAI Y H. Conjugate gradient methods with armijo-type line search [J]. Acta Mathematical Applicate Sinica (English Series), 2002, 18(1): 123-130.

[2] POLAK B T. The conjugate gradient method in extreme problems [J]. USSR Comput Math and Math Phys, 1969, 9: 94-112.

[3] POLAK E, RIBIRE G. Note surla convergence de directions conjugees [J]. Rev Francaise Informat Recherche Operatinelle 3e Annee, 1969, 16: 35-43.

[4] HESTENES M R, STIEFEL E. Method of conjugate gradient for solving linear equations[J]. J Res Nat Bur stand, 1952, 49: 409-436.

[5] DAI Y H, YUAN Y. A nonlinear conjugate gradient method with a strong global convergence property[J]. SIAM J Optim, 1999, 10: 177-182.

[6] WEI Z X, LI G H, QI L Q. On the global convergence of the polak-ribiere-polyak conjugate gradient method with an armijo-type inexact line search for nonconvex unconstrained optimization problems [M]. Hong Kong: The Hong Kong Polytechnic University, 2002.

[7] WEI Z X, YAO S W, LIU L Y. The convergence properties of some new conjugate gradient methods[J]. Applied Mathematics and Computation, 2006 (183): 1341-1350.

[8] YU G H, GUAN L T, WEI Z X. A globally convergent polak-ribiere-polyak conjugate gradient method with armijo-type line search [J]. Numerical Mathematics, 2006, 15(4): 357-367.

[9] DAI Y H, YUAN Y X. A nonlinear conjugate gradient method with a strong global convergence property[J]. SIAM Journal on Optimization, 1999, 10(1): 177-182.

（责任编辑：邓大玉）

（上接第 238 页 Continue from page 238）

[6] JIAN J B, TANG C M. An SQP feasible descent algorithm for nonlinear inequality constrained optimization without strict complementarity [J]. Computers and Mathematics with Applications, 2005, 49 (2～3): 223-238.

[7] JIAN J B, ZHENG H Y, TANG C M, et al. A new superlinearly convergent norm-relaxed method of strong sub-feasible direction for inequality constrained optimization[J]. Applied Mathematics and Computation, 2006, 182(2): 955-976.

[8] JIAN J B, ZHENG H Y, HU Q J, et al. A new norm-relaxed method of strongly sub-feasible direction for inequality constrained optimization [J]. Applied Mathematics and Computation, 2005, 168(1): 1-28.

[9] 袁亚湘，孙文瑜. 最优化理论与方法[M]，北京：科学出版社，1997.

（责任编辑：邓大玉）