利用 Fisher 函数解约束优化问题的广义梯度投影算法*

A Generalized Gradient Projection Algorithm with the Fisher Function for Solving the Constrained Optimization Problem

韦增欣,赵　岩,陈翠玲

WEI Zeng-xin,ZHAO Yan,CHEN Cui-ling

(广西大学数学与信息科学学院,广西南宁　530004)

(College of Mathematics and Information Science,Guangxi University,Nanning,Guangxi,530004,China)

摘要:通过引入一个辅助优化问题,将广义投影与罚函数技巧和 Fisher 函数 $\phi(a,b)=\sqrt{a^2+b^2}-(a+b),\forall a,b\in E^n$ 的特殊性质:$\sqrt{a^2+b^2}-(a+b)=0\Leftrightarrow a\geqslant 0,b\geqslant 0,ab=0$ 结合起来,给出处理非线性等式、不等式约束问题的广义梯度投影算法,并证明该算法是全局收敛的.该算法不仅保持文献[6]的优点,而且还扩大了初始点的选择范围.

关键词:约束优化　梯度投影　罚函数　Fisher 函数　全局收敛性

中图法分类号:O221.2　　**文献标识码**:A　　**文章编号**:1005-9164(2006)02-0102-03

Abstract:In this paper,we give a generalized gradient projection algorithm for solving equality and inequality constrained optimization problem by introducing an assistant optimization problem and combining the following three facets:1)generalized projection,2)the sbill of penalty function,3) Fisher function $\phi(a,b)=\sqrt{a^2+b^2}-(a+b)$, which has its special property: $\sqrt{a^2+b^2}-(a+b)=0\Leftrightarrow a\geqslant 0,b\geqslant 0,ab=0$. Furthermore,we prove that this algorithm is globally convergent. This algorithm preserves not only the advantage of reference [6],but also extends the choice range of initial point.

Key words:constrained optimization,gradient projection,penalty function,Fisher function,global convergence

梯度投影法是解决约束优化问题的一类有效算法,在最优化领域中占有重要地位.自 1960 年 Rosen[1,2]提出梯度投影方法起,国内外学者做了大量工作.由于 Rosen 方法对于线性约束是非常好的算法,但对于非线性约束却不总是很有效,因此国内外学者对非线性约束的梯度算法进行大量研究,他们或利用广义投影或将广义投影与罚函数技巧相结合[3~5].

文献[6]通过引入辅助优化问题,给出一个解带等式和不等式约束优化问题的广义梯度投影算法.基于该文思想,本文引入一个辅助优化问题,将广义投影与罚函数技巧和 Fisher 函数 $\phi(a,b)=\sqrt{a^2+b^2}-(a+b),\forall a,b\in E^n$ 的特殊性质:

$$\sqrt{a^2+b^2}-(a+b)=0\Leftrightarrow a\geqslant 0,b\geqslant 0,ab=0$$

结合起来,给出处理非线性不等式、等式约束问题的广义梯度投影算法,并证明该算法是全局收敛的.

1　问题和算法

我们引入的问题如下:

$(NP)\quad \min f(x)$

$\text{s.t.}\ h_j(x)\leqslant 0,j\in I=\{1,2,\cdots,m\},$

$h_j(x)=0,j\in L=\{m+1,m+2,\cdots,m+p\},x\in E^n.$

记 $Z=\{x|h_j(x)\leqslant 0,j\in I\},R=\{x|h_j(x)\leqslant 0,j\in I,h_j(x)=0,j\in L,x\in E^n\},$

$g(x)=-\nabla f(x),N(x)=(\nabla h_j(x),j\in I\cup L),H(x)=diag(H_j(x),j\in I\cup L),$

收稿日期:2005-05-27

作者简介:韦增欣(1963-),男,广西武鸣人,教授,主要从事最优化理论与方法研究。

*国家自然科学基金(10161002)和广西自然科学基金(30542043)资助。

其中

$$H_j(x)=\begin{cases}h_j(x), j\in I;\\ 0, \quad j\in L.\end{cases} \tag{2.1}$$

当 $x\in Z$ 时，显然 $H(x)\leqslant 0$.

我们始终假设：

(H1) $f(x)\in C^1, h_j(x)\in C^1, j\in I\cup L$;

(H2) 对 $\forall\, x\in Z, \{\nabla h_j(x)\mid j\in J_0(x)\cup L\}$ 线性无关，其中 $J_0(x)=\{j\mid h_j(x)=0, j\in I\}$.

引理 1 对 $\forall\, x\in Z$，矩阵 $(N(x)^TN(x)-H(x))$ 是对称正定阵. 证明类似文献[3].

这样对 $\forall\, x\in Z$，再记

$B(x)=(N(x)^TN(x)-H(x))^{-1}N(x)^T$,

$U(x)=(u_j(x), j\in I\cup L)=B(x)g(x)$,

$P(x)=I-N(x)(N(x)^TN(x)-H(x))^{-1}N(x)^T$,

其中 I 为单位矩阵.

定义辅助规划：

(AP) $\min \quad G_c(x)=f(x)+c\sum_{j\in L}|h_j(x)|$

s. t. $h_j(x)\leqslant 0, j\in I, c>0$, c 为参数.

定义 $DG_c(x,d)$ 为 x 处关于方向 d 的方向导数.

由于 Fisher 函数 $\psi(a,b)=\sqrt{a^2+b^2}-(a+b)$, $\forall\, a,b\in E^n$ 具有如下特殊性质：

$$\sqrt{a^2+b^2}-(a+b)=0\Leftrightarrow a\geqslant 0, b\leqslant 0, ab=0.$$

因此我们构造如下搜索方向

$$d(x)=s(x)+\tau(x)B(x)^TW,$$

其中

$s(x)=P(x)g(x)-\rho(x)B(x)^TV(x)$,

$\rho(x)=\sum_{j\in I}(\sqrt{u_j^2(x)+h_j^2(x)}-(u_j(x)-h_j(x)))^2+\sum_{j\in L}|h_j(x)|$,

$\tau(x)=\dfrac{|DGc(x,s)|}{2|U(x)^TW|+1}$, $W=(w_j, j\in I\cup L)$, $w_j=-1, j\in J$,

$V(x)=(v_j(x), j\in I\cup L)$,

$$v_j(x)=\begin{cases}1-h_j(x), u_j(x)<0, j\in I;\\ h_j(x), u_j(x)\geqslant 0, j\in I;\\ h_j(x), j\in L.\end{cases} \tag{2.2}$$

现在，我们给出相应的算法.

算法 算法的步骤如下.

步骤 0：取初始点 $x_0\in Z, g_0=-\nabla f(x_0), \varepsilon>0, \beta>1, \sigma\in(0,1)$. 记 $\bar{c}(x)=\max\{|u_j(x)|, j\in L\}+\bar{c}_0$, $\bar{c}_0$ 是一个充分小的正常数.

步骤 1：计算 $g(x_k), \rho(x_k), B(x_k), U(x_k), P(x_k)$. 若 $P(x_k)g(x_k)=0, \rho(x_k)=0$ 同时成立，停止；否则到步骤 2.

步骤 2：计算 $s_k=s(x_k), d_k=d(x_k)$.

步骤 3：若(1) $\bar{c}(x_k)>c(x_{k-1})$，则令 $c(x_k)=\max\{\bar{c}(x_k), c(x_{k-1})+\varepsilon\}$；(2) $\bar{c}(x_k)\leqslant c(x_{k-1})$，则令 $c(x_k)=c(x_{k-1})$.

步骤 4：令 $x_{k+1}=x_k+\lambda_k d_k$，其中 λ_k 是 $\{1, \frac{1}{\beta}, \frac{1}{\beta^2}, \cdots\}$ 中满足 $x_k+\lambda d_k\in Z$ 及 $G_{c_k}(x_k+\lambda d_k)-G_{c_k}(x_k)\leqslant\sigma\lambda DG_{c_k}(x_k,s_k)$ 的最大者. 令 $k:=k+1$，转向步骤 1.

容易验证下面 2 个引理成立.

引理 2 对 $\forall\, x\in E^n, d\in E^n$，有

$$DG_c(x,d)=\nabla f(x)^Td+c\sum_{j\in L,h_j(x)>0}\nabla h_j(x)^Td+c\sum_{j\in L,h_j(x)>0}|\nabla h_j(x)^Td|-c\sum_{j\in L,h_j(x)<0}\nabla h_j(x)^Td.$$

引理 3 对 $\forall\, x\in Z$，有 $g(x)^TP(x)g(x)\geqslant\|P(x)g(x)\|^2$.

定理 1 $x_k\in Z$ 为 (NP) 的一个 K-T 点的充要条件是 $P(x_k)g(x_k)=0, \rho(x_k)=0$ 同时成立.

证明 (1) 必要性. 若 $x_k\in Z$ 为 (NP) 的一个 K-T 点，则 $\exists\,\lambda=(\lambda_j, j\in I\cup L)$，其中 $\lambda_j\geqslant 0, j\in I$，使得

$$g(x_k)=N(x_k)\lambda, \lambda_jh_j(x_k)=0, j\in I, h_j(x_k)=0, j\in L$$

成立. 由 $H(x)$ 的定义易证 $\lambda=U(x_k)$，即 $P(x_k)g(x_k)=0$. 再由 Fisher 函数的性质可得 $\rho(x_k)=0$.

(2) 充分性. 若 $x_k\in Z$，有 $P(x_k)g(x_k)=0, \rho(x_k)=0$ 同时成立. 则由 $P(x_k)g(x_k)=0$，可得

$$g(x_k)=N(x_k)U(x_k).$$

再利用 Fisher 函数的性质知

$$u_j(x_k)\geqslant 0, u_j(x_k)h_j(x_k)=0, j\in I, h_j(x_k)=0, j\in L.$$

K-T 条件成立，所以 x_k 为 (NP) 的一个 K-T 点. 证毕.

引理 4 (1) 对 $x_k\in Z$，有 $DG_{c_k}(x_k,s_k)\leqslant 0$, $\nabla h_j^{kT}s_k\leqslant 0, \forall\, j\in J_0(x_k)$.

(2) 当 $x_k\in Z$ 为非 K-T 点时，则 $DG_{c_k}(x_k,s_k)<0$.

证明 (1) 因为

$\nabla f_k^Ts_k=-\nabla f_k^TP_k\nabla f_k+\rho_kU_k^TV_k\leqslant-\|P_k\nabla f_k\|^2+\rho_k(\sum_{u_j^k\geqslant 0,j\in I}(u_j^k-u_j^kh_j^k)+\sum_{u_j^k\geqslant 0,j\in I}u_j^kh_j^k+\sum_{j\in L}u_j^kh_j^k)$，且 $\forall\, j\in L$，有 $\nabla h_j^{kT}s_k=-\rho_kv_j^k=-\rho_kh_j^k$，所以

$DG_{c_k}(x_k,s_k)\leqslant-\|P_k\nabla f_k\|^2+\rho_k(\sum_{u_j^k<0,j\in I}(u_j^k-u_j^kh_j^k)+\sum_{u_j^k\geqslant 0,j\in I}u_j^kh_j^k)+\rho_k(\sum_{h_j^k>0,j\in L}(u_j^k-c_k)h_j^k+\sum_{h_j^k<0,j\in L}(u_j^k+c_k)h_j^k)$.

由算法知 $c_k > |u_j^k|, j \in L$，所以 $DG_{c_k}(x_k, s_k) \leqslant 0$.

对 $\forall_j \in J_0(x_k)$，有 $\nabla h_j^{kT} s_k = -\rho_k v_j^k \leqslant 0$.

(2) 当 $x_k \in Z$ 为非 K-T 点时，则或 $P(x_k)g(x_k) \neq 0$ 或 $\rho(x_k) \neq 0$. 但不论何种情况，显然 $DG_{c_k}(x_k, s_k) < 0$ 成立. 证毕.

引理 5 当 $x_k \in Z$ 为非 K-T 点时，则 $DG_{c_k}(x_k, d_k) < 0, \nabla h_j(x_k)^T d(x_k) < 0, \forall j \in J_0(x_k)$.

证明 因为

$$DG_{c_k}(x_k, d_k) = DG_{c_k}(x_k, s_k) - \tau_k U_k^T W - c_k p \tau_k,$$

又由引理 4(2) 可得 $DG_{c_k}(x_k, d_k) < \frac{1}{2} DG_{c_k}(x_k, s_k) < 0$.

对 $\forall j \in J_0(x_k)$，由引理 4 得

$$\nabla h_j^{kT} d_k = \nabla h_j^{kT} s_k + \tau_k \nabla h_j^{kT} B_k^T W \leqslant \tau_k w_j = -\tau_k < 0.$$ 证毕.

上述引理说明由算法步骤 2 产生的方向 d_k 为 (AP) 的一个可行下降方向，即算法中的 λ_k 是存在的，因此算法的每一步都是有意义的.

2 算法的全局收敛性

证明算法的全局收敛性就是证明算法或在有限步终止于 (NP) 的一个 K-T 点，或产生无穷点列，其任意极限点都为 (NP) 的 K-T 点.

由算法步骤及定理 1 知，当算法在有限步终止于 x_k 时，x_k 为 (NP) 的 K-T 点. 下面讨论算法产生无穷点列 $\{x_k\}$ 的情况.

设 x^* 为 $\{x_k\}$ 的任一极限点，则 $\exists K$ 使得 $\{x_k\}K \to x^*$. 因为 $h_j(x^*) \leqslant 0, j \in I$，且 $g(x), \nabla h_j(x), j \in I \cup L$ 连续，所以 $(N(x^*)^T N(x^*) - H(x^*))^{-1}$ 存在，且 $B(x_k), P(x_k), U(x_k), \rho(x_k)$ 的极限均存在，分别记它们的极限为 B^*, P^*, U^*, ρ^*，则 $B(x_k) \to B^*, P(x_k) \to P^*, U(x_k) \to U^*, \rho(x_k) \to \rho^*, k \to +\infty, k \in K$.

引理 6 $\exists k_0 > 0$，使当 $k \geqslant k_0$ 时，$c_k = c_{k_0} \doteq c$. 证明见文献[5].

由 $h_j(x) \in C^1, j \in I \cup L$，可知 $\{V(x_k)\}_K$ 是有界序列，故可取出收敛子列 $\{V(x_k)\}_{K_1} \to V^*, K_1 \subseteq K$.

定义 $s^* = P^* g^* - \rho^* B^{*T} V^*, \tau^* = \dfrac{|DG_c(x^*, s^*)|}{2|U^{*T} W| + 1}, d^* = s^* + \tau^* B^{*T} W$.

因此有 $s^k \to s^*$. 又由于 L 有限，所以 $\exists K_2 \subseteq K_1$，使得 $J_{11} = \{h_j(x_k) > 0, j \in L\}, J_{10} = \{h_j(x_k) = 0, j \in L\}, J_{12} = \{h_j(x_k) < 0, j \in L\}$ 与 k 无关. 因此 $DG_c(x_k, s_k) \to DG_c(x^*, s^*), k \to +\infty, k \in K_2$，所以 $\tau_k \to \tau^*, d_k \to d^*, DG_c(x_k, d_k) \to DG_c(x^*, d^*), k \to +\infty, k \in K_2$.

引理 7 若 $x^* \in Z$ 为非 K-T 点，则 $DG_c(x^*, d^*) < 0$.

类似引理 4，容易证明引理 7 成立.

定理 2 假设$(H_1)(H_2)$成立，则算法或有限步终止于 (NP) 的 K-T 点，或产生一个无穷点列 $\{x_k\}$，其任一极限点都是 (NP) 的 K-T 点.

证明 设 $\{x_k\}_{K_2} \to x^*$，且假设 x^* 不是 (NP) 的 K-T 点. 由于 $\{G_c(x_k)\}$ 是单调下降序列，所以 $\lim_{k \to +\infty}(G_c(x_k) - G_c(x_{k+1})) = 0$.

(1) 当 $\inf_{K_2} \lambda_k > 0$ 时，由于 $G_c(x_{k+1}) - G_c(x_k) \leqslant \sigma \lambda_k DG_c(x_k, d_k)$，令 $k \to +\infty, k \in K_2$，可得

$$DG_c(x^*, d^*) \geqslant 0$$

与引理 7 矛盾.

(2) 当 $\inf_{K_2} \lambda_k = 0$ 时，不妨设 $\lim_{k \to +\infty} \lambda_k = 0, k \in K_2$，由 λ_k 的取法，可知

$$G_c(x_k + \beta \lambda_k d_k) - G_c(x_k) > \sigma \beta \lambda_k DG_c(x_k, d_k).$$

由中值定理得

$$\beta \lambda_k DG_c(x_k + \theta \beta \lambda_k d_k, d_k) > \sigma \beta \lambda_k DG_c(x_k, d_k),$$

其中 $\theta \in (0,1)$. 令 $k \to +\infty, k \in K_2$，得 $DG_c(x^*, d^*) \geqslant \sigma DG_c(x^*, d^*)$. 因为 $\sigma \in (0,1)$，所以

$$DG_c(x^*, d^*) \geqslant 0$$

与引理 7 矛盾. 证毕.

3 结束语

因为广义投影与罚函数技巧都是求解约束优化问题的基本方法，又由于 Fisher 函数具有特殊的性质，所以本文通过引入一个辅助优化问题将三者结合起来，给出一个解带等式和不等式约束优化问题的广义梯度投影算法，并证明该算法是全局收敛的. 该算法的优点在于不仅保持文献[6]的优点，而且还扩大了初始点的选择范围.

参考文献：

[1] ROSEN J B. The gradient projection method for nonlinear programming: Part Ⅰ[J]. Linear Constraints, J SIAM, 1960, 1(8): 181-217.

[2] ROSEN J B. The gradient projection method for nonlinear programming: Part Ⅱ[J]. Nonlinear Constraints, J SIAM, 1960, 9(4): 514-532.

[3] 高自友，贺国平. 约束优化问题的一个广义梯度投影法[J]. 科学通报，1991，36(19)：1444-1447.

[4] 赖炎连，韦增欣. 初始点任意且全局收敛的梯度投影法[J]. 科学通报，1990，20：1536-1539.

[5] 赖炎连，简金宝. 初始点任意的一个非线性优化的广义梯度投影法[J]. 系统科学与数学，1995，15(4)：374-380.

[6] 高自友，刘仍奎. 非线性约束条件下的广义投影梯度法[J]. 北方交通大学学报，1995，19(2)：203-208.

（责任编辑：邓大玉）