文章编号: 1000-341X(2007)04-0728-05 **文献标识码**: A

利用 Fisher 函数解非线性不等式约束优化问题的梯度投影算法

赵 岩 [1,3], 陈翠玲 [2,3], 韦增欣 [3]
(1. 上海理工大学理学院，上海 200093; 2. 广西师范大学数学科学学院，广西 桂林 541004;
3. 广西大学数学与信息科学学院，广西 南宁 530004)
(E-mail: zhaoyantt@126.com)

摘 要: 本文将利用梯度投影与 Fisher 函数提出一个新的二阶段搜索方向，给出相应的解非线性不等式约束优化问题的梯度投影算法，并证明了该算法具有全局收敛性.

关键词: 非线性不等式约束；梯度投影； Fisher 函数；全局收敛性.
MSC(2000): 90C30
中图分类: O221.2

1 引 言

Rosen 在文献 [1,2] 中利用梯度投影建立了带约束非线性规划问题的可行方向算法，称为梯度投影方法. 由于此方法简单易行，计算的每一步都是显示迭代，而不必去求解复杂的线性规划或二次规划问题，因此颇受人们注意 [3−10]. 文献 [4] 对不等式约束问题提出了一个一般模型，当在这个模型中选取不同参数时，就会得到不同算法. 文献 [5] 利用广义投影构造了一个二阶段搜索方向，给出了一个有效的梯度类算法，因每步迭代时无需转轴运算，故大大减少了计算量. 受文献 [5] 的启发，再结合 Fisher 函数 $\varphi(a,\ b)=\sqrt{a^2+b^2}-(a+b)$, $\forall a,\ b\in\mathbf{E}$ 的特殊性质：

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

本文将构造一个新的二阶段搜索方向，给出相应的解非线性不等式约束优化问题的梯度投影算法，并将证明该算法具有全局收敛性.

2 问题和算法

本文考虑的问题如下：

$$(\mathrm{NP})\qquad \begin{array}{ll}\min & f(x),\\ \text{s.t.} & h_j(x)\leq 0,\quad j\in I=\{1,2,\cdots,m\},\ x\in E^n.\end{array}$$

并假设

(H$_1$) $f(x)\in C^1,\ h_j(x)\in C^1, j\in I.$

(H$_2$) $R=\{x\mid h_j(x)\leq 0,\ j\in I,\ x\in E^n\}\neq\emptyset$ 且对 $\forall x\in R$, $\{\nabla h_j(x)\mid j\in J_0(x)\}$ 线性无关，其中 $J_0(x)=\{j\mid h_j(x)=0,\ j\in I\}$.

收稿日期: 2005-04-21; **接受日期**: 2006-12-10
基金项目: 上海市高等学校科学技术发展基金 (07ZZ83); 广西省自然科学基金 (0542043) 和广西师范大学青年基金.

记 $J_\delta(x)=\{j\mid 0\le -h_j(x)\le\delta,\ j\in I\}$, $\delta>0$, $g(x)=-\nabla f(x)$, $N_j(x)=\nabla h_j(x)$.

假设对 $\forall x\in R$, $\exists\delta_1>0$, 使 $J=J_{\delta_1}(x)\subseteq I$ 满足下述算法步骤 1 的要求，因而有矩阵 $N_J(x)=(N_j(x),\,j\in J)$ 列满秩. 因此再记:

$$B_J(x)=(N_J(x)^{\mathrm T}N_J(x))^{-1}N_J(x)^{\mathrm T},\quad U_J(x)=(u_j(x),\ j\in J)=B_J(x)g(x),$$
$$P_J(x)=I-N_J(x)(N_J(x)^{\mathrm T}N_J(x))^{-1}N_J(x)^{\mathrm T},$$

其中 I 为单位矩阵.

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

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

因此构造如下二阶段搜索方向

$$\begin{cases} s(x)=P_J(x)g(x)+\rho(x)B_J(x)^{\mathrm T}V_J(x),\\ d(x)=s(x)+\tau(x)B_J(x)^{\mathrm T}W,\end{cases}\tag{2.1}$$

其中 $\rho(x)=\sum_{j\in J}(\sqrt{u_j^2(x)+h_j^2(x)}-(u_j(x)-h_j(x)))^2$, $\tau(x)=\frac{g(x)^{\mathrm T}s(x)}{2|U_J(x)^{\mathrm T}W|+1}$, $W=(w_j,\ j\in J)$, $w_j=-1$, $j\in J$, $V_J(x)=(v_j(x),\ j\in J)$,

$$v_j(x)=\begin{cases}-1+h_j(x), & u_j(x)<0;\\ -h_j(x), & u_j(x)\ge 0.\end{cases}\tag{2.2}$$

现在，给出相应的算法:

算法:

步骤 0: 取初始点 $x_0\in R,\ \delta_0>0,\ g_0=-\nabla f(x_0),\ \beta>1,\ \sigma\in(0,\ 1)$, 令 $J_{\delta_0}=\{j\mid 0\le -h_j(x_0)\le\delta_0\},\ k:=0$.

步骤 1: 对 $x_k,\ \delta_k,\ J_k=J_{\delta_k},\ N_{J_k}=(N_j^k,\ j\in J_k)$, $|\det(N_{J_k}(x_k)^{\mathrm T}N_{J_k}(x_k))|\ge\delta_k$? 若不成立，到步骤 2; 否则到步骤 3.

步骤 2: 令 $\delta_k:=\frac{1}{2}\delta_k$, 重回步骤 1.

步骤 3: 计算 $g(x_k),\ \rho(x_k),\ B_{J_k}(x_k),\ U_{J_k}(x_k),\ P_{J_k}(x_k)$. 若 $P_{J_k}(x_k)g(x_k)=0,\ \rho(x_k)=0$ 同时成立，停止；否则到步骤 4.

步骤 4: 计算 $s_k=s(x_k),\ d_k=d(x_k)$.

步骤 5: 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 $x_k+\lambda d_k\in R$ 及 $f(x_k+\lambda d_k)-f(x_k)\le\sigma\lambda\nabla f(x_k)^{\mathrm T}d_k$ 的最大者. 令 $\delta_{k+1}:=\delta_0,\ k:=k+1$, 转向步骤 1.

定理 1 $x_k\in R$ 为 K-T 点的充要条件是算法步骤 1 中对 x_k 确定的 J_k 有 $P_{J_k}(x_k)g(x_k)=0,\ \rho(x_k)=0$ 同时成立.

证明 必要性. 当 $x_k\in R$ 为 K-T 点时，则 $\exists\bar\lambda=(\bar\lambda_j,\ j\in I)\ge 0$, 使得 $g(x_k)=N(x_k)\bar\lambda$, $\bar\lambda_j h_j(x_k)=0$, $j\in I$. 由于对 $\forall j\notin J_k$, 有 $h_j(x_k)<-\delta_k$, 所以 $\bar\lambda_j=0$, $j\notin J_k$, 即 $g(x_k)=N_{J_k}(x_k)\bar\lambda_{J_k}$. 随后可得 $\bar\lambda_{J_k}=U_{J_k}(x_k)$. 因此 $P_{J_k}(x_k)g(x_k)=0,\ u_j(x_k)\ge 0,\ u_j(x_k)h_j(x_k)=0,\ j\in J_k$. 最后由 Fisher 函数的性质有 $\rho(x_k)=0$ 成立.

充分性. 当 $P_{J_k}(x_k)g(x_k)=0,\ \rho(x_k)=0$ 时， $\rho(x_k)=0$, 即

$$\sqrt{u_j(x_k)^2+h_j(x_k)^2}-(u_j(x_k)-h_j(x_k))=0,\ j\in J_k.$$

利用 Fisher 函数的性质得 $u_j(x_k)\ge 0,\ u_j(x_k)h_j(x_k)=0,\ j\in J_k$. 因为

$$P_{J_k}(x_k)g(x_k)=g(x_k)-N_{J_k}(x_k)U_{J_k}(x_k),$$

所以由 $P_{J_k}(x_k)g(x_k)=0$ 可得 $g(x_k)=N_{J_k}(x_k)U_{J_k}(x_k)$. 令 $u_j(x_k)=0,\ j\notin J_k$, 即得 x_k 点处 K-T 条件. □

引理 1 (1) 若 $x_k\in R$, 则 $g(x_k)^{\mathrm T}s(x_k)\ge 0,\ \nabla h_j(x_k)^{\mathrm T}s(x_k)\le 0,\ \forall j\in J_0(x_k)$.

(2) 当 $x_k\in R$ 为非 K-T 点时，有 $g(x_k)^{\mathrm T}s(x_k)>0$.

证明 (1) 由 $s(x)$ 的定义，有

$$\begin{aligned}g(x_k)^{\mathrm T}s(x_k)=&\|P_{J_k}(x_k)g(x_k)\|^2+\rho(x_k)(\sum_{u_j(x_k)<0}-u_j(x_k)+\\&\sum_{u_j(x_k)<0}u_j(x_k)h_j(x_k)+\sum_{u_j(x_k)\ge 0}-u_j(x_k)h_j(x_k)).\end{aligned}$$

因为 $x_k\in R$, 所以 $g(x_k)^{\mathrm T}s(x_k)\ge 0$.

因为 $\nabla h_j(x_k)^{\mathrm T}s(x_k)=\rho(x_k)v_j(x_k)$, 又 $\forall j\in J_0(x_k)$ 有 $h_j(x_k)=0$, 所以 $v_j(x_k)\le 0$, 即得 $\nabla h_j(x_k)^{\mathrm T}s(x_k)\le 0$.

(2) 当 $x_k\in R$ 为非 K-T 点时，则或 $P_{J_k}(x_k)g(x_k)\ne 0$ 或 $\rho(x_k)\ne 0$. 但不论何种情况 $g(x_k)^{\mathrm T}s(x_k)>0$ 都显然成立. □

引理 2 当 $x_k\in R$ 为非 K-T 点时，则 $g(x_k)^{\mathrm T}d(x_k)>0$, $\nabla h_j(x_k)^{\mathrm T}d(x_k)<0$, $\forall j\in J_0(x_k)$.

证明 因为

$$g(x_k)^{\mathrm T}d(x_k)=g(x_k)^{\mathrm T}s(x_k)+\tau(x_k)U_{J_k}(x_k)^{\mathrm T}W=g(x_k)^{\mathrm T}s(x_k)+\frac{g(x_k)^{\mathrm T}s(x_k)}{2|U_{J_k}(x_k)^{\mathrm T}W|+1}U_{J_k}(x_k)^{\mathrm T}W,$$

又由引理 1(2), 所以可得 $g(x_k)^{\mathrm T}d(x_k)>\frac{1}{2}g(x_k)^{\mathrm T}s(x_k)>0$.

对 $\forall j\in J_0(x_k)$, 由引理 1 得

$$\nabla h_j(x_k)^{\mathrm T}d(x_k)=\nabla h_j(x_k)^{\mathrm T}s(x_k)+\tau(x_k)\nabla h_j(x_k)^{\mathrm T}B_{J_k}(x_k)^{\mathrm T}W\le\tau(x_k)w_j=-\tau(x_k)<0.$$

上述引理说明由算法步骤 3 产生的方向 d_k 为 (NP) 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.

3 算法的全局收敛性

在上述算法中，若有限步内 $P_{J_k}(x_k)g(x_k)=0$, $\rho(x_k)=0$ 同时成立，则由定理 1 知 x_k 为 (NP) 的一个 K-T 点；否则算法将产生无穷点列 $\{x_k\}$. 下面考虑算法产生无穷点列 $\{x_k\}$ 的情况：

由假设条件 (H_2) 知，算法步骤 1 与 2 之间不会产生无限循环，因此有

$$K=\{k\in N\mid |\det(N_{J_k}(x_k)^{\mathrm T}N_{J_k}(x_k))|\ge\delta_k\}$$

为 N 的一个无穷子列，并且 $\{x_k\}$ 与 $\{x_k\}_K$ 有相同的聚点 (若把相同的点看成一个点，则 $\{x_k\}$ 与 $\{x_k\}_K$ 相同). 设 x^* 为 $\{x_k\}$ 的任一有限聚点，又因为 $J_k\subseteq I$ 只有有限多种选择，所以存在无穷子列 $K_1\subseteq K$, 使得

(1) $\{x_k\}_{K_1} \to x^*, \quad k \to +\infty$.

(2) $\forall k \in K_1$, J_k 与 k 无关，记作 $J = J_k$.

由文献 [3], 有如下引理

引理 3 $\exists \delta^* > 0$, 使 $\delta_k > \delta^*$, 对 $\forall k \in K_1$ 成立.

由于 J 固定及 $g(x)$, $\nabla h_j(x)$, $j \in I$ 连续, 又因引理 3 保证了 $|\det(N_J(x^*)^{\mathrm{T}} N_J(x^*))| \geq \delta^* > 0$, 从而 $N_J(x^*)^{\mathrm{T}} N_J(x^*)$ 的逆存在. 因此 $B_J(x_k)$, $P_J(x_k)$, $U_J(x_k)$, $\rho(x_k)$ 的极限均存在，分别记它们的极限为 B_J^*, P_J^*, U_J^*, ρ^*, 则

$$B_J(x_k) \to B_J^*,\ P_J(x_k) \to P_J^*,\ U_J(x_k) \to U_J^*,\ \rho(x_k) \to \rho^*, \quad k \to +\infty,\ k \in K_1.$$

由 $h_j(x) \in C^1$, $j \in I$, 可知 $\{V_J(x_k)\}_{K_1}$ 是有界序列, 故可取出收敛子列 $\{V_J(x_k)\}_{K_2} \to V_J^*$, $K_2 \subseteq K_1$. 定义：

$$s^* = P_J^* g^* + \rho^* B_J^{*\mathrm{T}} V_J^*, \quad \tau^* = \frac{g^{*\mathrm{T}} s^*}{2|U_J^{*\mathrm{T}} W| + 1}, \quad d^* = s^* + \tau^* B_J^{*\mathrm{T}} W.$$

因此有 $s_k \to s^*$, $\tau_k \to \tau^*$, $d_k \to d^*$, $k \to +\infty$, $k \in K_2$.

引理 4 若 $x^* \in R$ 为非 K-T 点，则 $\tau^* > 0$, $g^{*\mathrm{T}} d^* > 0$.

证明见文献 [5].

引理 5 若 $x^* \in R$ 为非 K-T 点，则 $\exists \mu > 0$, 对 $\forall \lambda \in [0,\ \mu]$ 和充分大的 $k \in K_2$, 有 $x_k + \lambda d_k \in R$ 成立.

证明 (1) 若 $j \notin J_0(x^*)$, 则 $h_j(x^*) < 0$, 所以 $\exists \mu_1 > 0$, 对 $\forall \lambda \in [0,\ \mu_1]$ 和充分大的 $k \in K_2$ 有 $h_j(x_k + \lambda d_k) < 0$ 成立.

(2) 当 $j \in J_0(x^*)$, 有 $\nabla h_j(x_k)^{\mathrm{T}} d_k = \nabla h_j(x_k)^{\mathrm{T}} s_k - \tau(x_k)$. 令 $k \to +\infty$, $k \in K_2$, 由引理 1(1) 得 $\nabla h_j(x^*)^{\mathrm{T}} d^* = \nabla h_j(x^*)^{\mathrm{T}} s^* - \tau^* \leq -\tau^*$, 所以对 $\forall j \in J_0(x^*)$ 和充分大的 $k \in K_2$, 有 $\nabla h_j(x_k)^{\mathrm{T}} d_k \leq -\frac{1}{2}\tau^*$. 进而 $\exists \mu_2 > 0$ 使对 $\forall \lambda \in [0,\ \mu_2]$ 和充分大的 $k \in K_2$, 有

$$\nabla h_j(x_k + \lambda d_k)^{\mathrm{T}} d_k \leq -\frac{1}{4}\tau^*. \tag{3.1}$$

另一方面，由中值定理得 $h_j(x_k + \lambda d_k) \leq \lambda \nabla h_j(x_k + \eta_{jk}\lambda d_k)^{\mathrm{T}} d_k$, 其中 $\eta_{jk} \in (0,\ 1)$. 由 (3.1) 式知 $\exists \mu_2 > 0$, 对 $\forall \lambda \in [0,\ \mu_2]$ 和充分大的 $k \in K_2$, 有 $h_j(x_k + \lambda d_k) \leq \frac{1}{4}\lambda\tau^* \leq 0$ 成立. 令 $\mu = \min\{\mu_1,\ \mu_2\}$, 则该引理证毕. □

定理 2 假设 (H_1), (H_2) 成立，则算法或有限步终止于 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 $\{f(x_k)\}$ 是单调下降序列, 所以 $\lim_{k\to+\infty}(f(x_k) - f(x_{k+1})) = 0$.

(1) 当 $\inf_{K_2} \lambda_k > 0$ 时，由于 $f(x_{k+1}) - f(x_k) \leq \sigma\lambda_k \nabla f(x_k)^{\mathrm{T}} d_k$, 令 $k \to +\infty$, $k \in K_2$, 可得

$$\nabla f(x^*)^{\mathrm{T}} d^* \geq 0.$$

与引理 4 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

$$f(x_k + \beta\lambda_k d_k) - f(x_k) > \sigma\beta\lambda_k \nabla f(x_k)^{\mathrm{T}} d_k.$$

由中值定理得 $\beta\lambda_k\nabla f(x_k+\theta\beta\lambda_k d_k)^{\mathrm{T}}d_k>\sigma\beta\lambda_k\nabla f(x_k)^{\mathrm{T}}d_k$, 其中 $\theta\in(0,\ 1)$. 令 $k\to+\infty$, $k\in K_2$, 得 $\nabla f(x^*)^{\mathrm{T}}d^*\geq\sigma\nabla f(x^*)^{\mathrm{T}}d^*$. 因为 $\sigma\in(0,\ 1)$, 所以 $\nabla f(x^*)^{\mathrm{T}}d^*\geq 0$. 与引理 4 矛盾. □

参考文献:

[1] ROSEN J B. *The gradient projection method for nonlinear programming. I. Linear constraints* [J]. J. Soc. Indust. Appl. Math., 1960, **8**: 181–217.

[2] ROSEN J B. *The gradient projection method for nonlinear programming. II. Nonlinear constraints* [J]. J. Soc. Indust. Appl. Math., 1961, **9**: 514–532.

[3] 堵丁柱. 非线性约束条件下的梯度投影方法 [J]. 应用数学学报, 1985, **8**(1): 7–16.
DU Ding-zhu. *A gradient projection algorithm for nonlinear constraints* [J]. Acta Math. Appl. Sinica, 1985, **8**(1): 7–16. (in Chinese)

[4] 韦增欣. 非线性约束条件下梯度投影法的一个统一途径 [J]. 应用数学学报, 1990, **13(3)**: 304–313.
WEI Zeng-xin. *A unified approach to the gradient projection algorithm under nonlinear constraints* [J]. Acta Math. Appl. Sinica, 1990, **13**(3): 304–313. (in Chinese)

[5] 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, **19**: 1444–1447.
GAO Zi-you, HE Guo-ping. *A generalized gradient projection method for constrained optimization problem* [J]. Chinese Science Bull., 1991, **19**: 1444–1447.(in Chinese)

[6] LAI Yan-lian, WEI Zeng-xin. *A unified approach to the method of gradient projection with arbitrary initial point* [J]. Systems Sci. Math. Sci., 1991, **4**(3): 215–224.

[7] 陈广军. 一个解带线性或非线性约束最优化问题的梯度投影方法 [J]. 计算数学, 1987, **4**: 356–364.
CHEN Guang-jun. *A gradient projection algorithm for optimization problems with general constraints* [J]. Math. Numer. Sin., 1987, **4**: 356–364. (in Chinese)

[8] 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.
LAI Yan-lian, WEI Zeng-xin. *A gradient projection method with arbitrary initial point and its global convergence* [J]. Chinese Sci. Bull., 1990, **20**: 1536–1539. (in Chinese)

[9] 赖炎连. 非线性规划的法向与梯度组合方向算法及其收敛性 [J]. 系统科学与数学, 1990, **10**(2): 181–188.
LAI Yan-lian. *An algorithm for nonlinear programming using a combination of the normal direction and the gradient direction, and its convergence* [J]. J. Systems Sci. Math. Sci., 1990, **10**(2): 181–188. (in Chinese)

[10] 高自友, 于伟. 任意初始点下的广义梯度投影方法 [J]. 科学通报, 1992, **20**: 1832–1836.
GAO Zi-you, YU Wei. *A generalized gradient projection method with arbitrary initial point* [J]. Chinese Sci. Bull., 1992, **20**: 1832–1836. (in Chinese)

A Gradient Projection Algorithm Using the Fisher Function for Solving the Nonlinear Inequality Constrained Optimization Problem

ZHAO Yan[1,3], *CHEN Cui-ling*[2,3], *WEI Zeng-xin*[3]

(1. College of Science, University of Shanghai for Science and Technology, Shanghai 200093, China;
2. College of Mathematics Science, Guangxi Normal University, Guangxi 541004, China;
3. College of Mathematics and Information Science, Guangxi University, Guangxi 530004, China)

Abstract: A new two-stage search direction is proposed that combines the gradient projection and the Fisher function. Based on the new direction, this paper proposes a gradient projection algorithm for solving nonlinear inequality constrained optimization problem. The algorithm is shown to be globally convergent.

Key words: nonlinear inequality constriants; gradient projection; the Fisher function; global convergence.