优化问题的线性逼近与罚函数搜索算法及其收敛性*[1)]

韦增欣
(广西大学)
赖炎连
(中国科学院应用数学所)

AN OPTIMIZATION ALGORITHM USING LINEAR APPROXIMATION AND PENALTY FUNCTION AND ITS CONVERGENCE

Wei Zeng-xin
(*Guangxi University*)
Lai Yan-lian
(*Institute of Applied Mathematics, Academia Sinica*)

Abstract

In this paper, we consider the optimization problem with nonlinear constraints and combine linear programming with the penalty function to give an algorithm with arbitrary initial point. In every iterative step, we solve the linear programming to get the iterative direction d(x) which is the descending direction of the penalty function. The parameter of the penalty function is given by the simplex multiplier and becomes a constant number when k is sufficiently large. Finaly, we give the proof of global convergence of this algorithm.

§1. 引　言

我们讨论一般的非线性约束的优化问题 (NP):

$$(NP)\quad \begin{array}{l}\min f(x),\\ x\in R,\end{array}$$

$$x\in E_n,\ R=\{x\mid g_i(x)\leqslant 0,\ i=\overline{1,m};\ h_j(x)=0,\ j=\overline{1,l}\}.$$

* 1989 年 6 月 17 日收到
1) 国家自然科学基金资助项目

在本文中,恒假定 (H_1): $f(x)$, $g_i(x)$, $1 \leqslant i \leqslant m$, $h_j(x)$, $1 \leqslant j \leqslant l$ 为一阶连续可微函数.

上述 (NP) 问题,若用可行方向法[5,6]等方法求解时,初始点必须是可行点,且在每一步迭代中,为了得到目标函数值下降而又可行的点,除进行一维搜索外往往需要增加辅助的计算. [7] 对只有不等式约束的问题给出了一个线性逼近与效益函数相结合的方法,但只证明了整体收敛于 Fritz-John 点,也没有给出罚参数的选取方法,而是满足一定条件的理论值. [1]—[4] 是在每一迭代步求解一个二次规划,得到迭代方向,罚参数依赖于二次规划的 K-T 乘子. 受此启发,本文给出了一个线性逼近与罚函数相结合的方法,证明了算法全局收敛于 Kuhn-Tucker 点,罚参数的计算公式通过线性子规划的单纯形乘子给出,并在迭代过程中逐步调整而成为常数. 因而克服了 [7] 中的不足. 算法简便,初始点可任取,是一个完整而实用的方法.

§2. 预备知识

在本文的算法中,迭代的每一步将 $f(x)$ 与 $g_i(x)$, $h_j(x)$ 线性化并考虑线性子规划问题 $LP(x)$:

$$
LP(x):\quad \begin{array}{ll} \min & \nabla f(x)^T d, \\ \text{s.t} & g_i(x) + \nabla g_i(x)^T d \leqslant 0, \quad i = 1,2,\cdots m. \\ & \begin{cases} h_j(x) + \nabla h_j(x)^T d = 0, & j = 1,2,\cdots l, \\ -1 \leqslant d_i \leqslant 1, & d = (d_1, d_2 \cdots d_n). \end{cases} \end{array} \tag{1}
$$

对于每一迭代点 x^k, 解 $LP(x^k)$, 记其最优解为 d^k.

由于初始点 x^0 可以是非可行点, x^k 亦然,因此,当 x^k 不为 K-T 点时,我们以罚函数 $F_c(x)$ 作为从 x^k 出发沿 d^k 方向作线搜索的目标函数,以求出下一个迭代点. 由此产生方法的迭代点列 $\{x^k\}$. $F_c(x)$ 及沿方向 d 的方向导数 $DF_c(x, d)$ 由下式定义:

$$
F_c(x) = f(x) + c\left\{\sum_{i=1}^{m} \max(0, g_i(x)) + \sum_{j=1}^{l} |h_j(x)|\right\},
$$

$$
DF_c(x, d) = \lim_{t \to 0^+} \frac{F_c(x + td) - F_c(x)}{t}. \tag{2}
$$

为保证方法可行,假定 (H_2): 对于每一点 x^k, $LP(x^k)$ 有解. 因而约束集 (1) 非空。

为了论证需要,我们考虑 $LP(x)$ 的对偶规划问题 $LD(x)$ 如下:

$$
LD(x):\quad \begin{array}{ll} \max & u^{(1)} g(x) - u^{(2)} e - u^{(3)} e + u^{(4)} h(x), \\ \text{s.t:} & -u^{(1)} \nabla g(x)^T - u^{(2)} I_n + u^{(3)} I_n - u^{(4)} \nabla h(x)^T = \nabla f(x), \\ & u^{(i)} \geqslant 0, \quad i = 1,2,3, \\ & u^{(4)} \text{ 不受限制}, \end{array} \tag{3}
$$

其中, $g(x) = (g_1(x), g_2(x) \cdots g_m(x))^T$, $h(x) = (h_1(x), h_2(x), \cdots h_m(x))^T$, $-1 \leqslant d_j \leqslant 1, 1 \leqslant j \leqslant n$ 写成 $I_n d \leqslant e, -I_n d \leqslant e$, $e = (1, 1, \cdots 1)^T$, I_n 为单位 n 阶阵.

引理 1. 设 $d(x)$ 为 $LP(x)$ 的最优解，$U(x)=(u^{(1)}(x), u^{(2)}(x), u^{(3)}(x), u^{(4)}(x))$ 为 $LD(x)$ 的最优解，则

$$\nabla f(x)^T d(x) \leqslant u^{(1)}(x)g(x)+u^{(4)}(x)h(x). \tag{4}$$

证明. 由 $LP(x)$ 有最优解 $d(x)$，则由线性规划的对偶定理，知对偶规划必有最优解，设为 $U(x)=(u^{(1)}(x), u^{(2)}(x), u^{(3)}(x), u^{(4)}(x))$，且它们的最优目标函数值相等，即

$$\nabla f(x)^T d(x)=u^{(1)}(x)g(x)-u^{(2)}(x)e-u^{(3)}(x)e+u^{(4)}(x)h(x),$$

由于 $u^{(2)}(x)\geqslant 0$，$u^{(3)}(x)\geqslant 0$，$e=(1,1,\cdots 1)^T$，即得 (4) 式.

由线性规划的性质可知，当我们用单纯形方法求解线性规划 $LP(x)$ 时，其最优解对应的单纯形乘子即为对偶规划的最优解 $U(x)$。因此 $U(x)$ 在解 $LP(x)$ 时同时得到而不必去解 $LD(x)$.

引理 2. 设 $x^*\in R$为可行解，当线性规划$LP(x^*)$ 的最优目标函数值 $\nabla f(x^*)^T d(x^*)=0$ 时，则 x^* 为问题 (NP) 的 K-T 点.

证明. 由于 $d(x^*)$ 是 $LP(x^*)$ 的最优解，因而相应的 K-T 条件成立，即存在乘子 u_i，v_j，使得

$$\begin{aligned}
&\nabla f(x^*)+\sum_{i=1}^{m} u_i\nabla g_i(x^*)+\sum_{j=1}^{l} v_j\nabla h_j(x^*)=0,\\
&u_i\geqslant 0,\ u_i(g_i(x^*)+\nabla g_i(x^*)^T d(x^*))=0,\\
&g_i(x^*)+\nabla g_i(x^*)^T d(x^*)\leqslant 0,\\
&h_j(x^*)+\nabla h_j(x^*)^T d(x^*)=0.
\end{aligned} \tag{5}$$

$1\leqslant i\leqslant m$，$1\leqslant j\leqslant l$ 成立.

若 $d(x^*)=0$，则由 (5) 可知 x^* 为可行点及它为原问题 (NP) 的 K-T 点.

若 $d(x^*)\neq 0$ 但 $\nabla f(x^*)^T d(x^*)=0$，则由 $d(x^*)$ 是 $LP(x^*)$ 的最优解，可得对所有满足 $LP(x^*)$ 的约束

$$\begin{aligned}
&g_i(x^*)+\nabla g_i(x^*)^T d\leqslant 0,\ 1\leqslant i\leqslant m,\\
&h_j(x^*)+\nabla h_j(x^*)^T d=0,\ 1\leqslant j\leqslant l
\end{aligned} \tag{6}$$

的 d，均有

$$\nabla f(x^*)^T d\geqslant 0. \tag{7}$$

因此，若 $x^*\in R$，记 $I(x^*)=\{i\,|\,g_i(x^*)=0,\ 1\leqslant i\leqslant m\}$，则对所有满足

$$\begin{aligned}
&-g_i(x^*)-\nabla g_i(x^*)^T d=-\nabla g_i(x^*)^T d\geqslant 0,\ i\in I(x^*),\\
&h_j(x^*)+\nabla h_j(x^*)^T d=\nabla h_j(x^*)^T d=0,\qquad j=1,2,\cdots l
\end{aligned} \tag{8}$$

的 d，(由假定 (H_2)，(6) 中 d 的集合非空)，均有

$$\nabla f(x^*)^T d\geqslant 0 \tag{9}$$

成立. 由 Farkas 定理，知存在 $u_i\geqslant 0$，$i\in I(x^*)$ 及 v_j，$j=1,2,\cdots l$，并令 $u_i=0$，$i\notin I(x^*)$，使得

$$\nabla f(x^*)+\sum_{i=1}^{m} u_i\nabla g_i(x^*)+\sum_{j=1}^{l} v_j\nabla h_j(x^*)=0 \tag{10}$$

成立. 由 $I(x^*)$ 的定义，可得 $u_i g_i(x^*)=0$，$1\leqslant i\leqslant m$，这说明 x^* 为原问题 (NP) 的 K-T 点. 对于条件$-1\leqslant d_j\leqslant 1$，对任意的 d，总可取常数 $t>0$，使得$-1\leqslant td_j\leqslant 1$

成立。而在(8)、(9)中同乘以 t，不影响(10)的成立，即 x^* 为 (NP) 的 K-T 点。

记 $S(x)=\{d \mid g(x)+\nabla g(x)^T d \leqslant 0, h(x)+\nabla h(x)^T d=0, I_n d \leqslant e, -I_n d \leqslant e\}$，有如下引理：

引理 3. i). 对 $\forall d \in E_n$，$DF_c(x, d)=\nabla f(x)^T d+c\Big\{\sum\limits_{g_i(x)>0} \nabla g_i(x)^T d+$

$$\sum_{g_i(x)=0}(\nabla g_i(x)^T d)_{+}+\sum_{h_j(x)>0} \nabla h_j(x)^T d+\sum_{h_j(x)=0}|\nabla h_j(x)^T d|+\sum_{h_j(x)<0}(-\nabla h_j(x)^T d)\Big\}.$$

ii). 对 $\forall d \in S(x)$，$DF_c(x, d)=\nabla f(x)^T d+c\Big\{\sum\limits_{g_i(x)>0} \nabla g_i(x)^T d-\sum\limits_{j=1}^{l}|h_j(x)|\Big\}$.

证明. i). 证明见 [8]. 其中 $a_{+}=\max(0, a)$.

ii). 由 $d \in S(x)$，以 $\nabla g_i(x)^T d \leqslant -g_i(x)$，$\nabla h_j(x)^T d=-h_j(x)$ 代入 i) 中即知 ii) 成立.

引理 4. 设 $x \notin R$，$\bar{c}(x)=\max\{u_i^{(1)}(x),\ 1 \leqslant i \leqslant m,\ |u_j^{(4)}(x)|,\ 1 \leqslant j \leqslant l\}+\varepsilon_0$，则当 $C \geqslant \bar{c}(x)$ 时，$DF_c(x, d(x))<0$，其中 $U(x)=(u^{(1)}(x), u^{(2)}(x), u^{(3)}(x), u^{(4)}(x))$，$d(x)$ 分别为 $LD(x)$ 与 $LP(x)$ 的最优解，$\varepsilon_0>0$ 为常数.

证明. 因 $d(x) \in S(x)$. 由引理 3 及引理 1，得

$$\begin{aligned} DF_c(x, d(x)) \leqslant\ & u^{(1)}(x) g(x)+u^{(4)}(x) h(x) \\ & -c\Big\{\sum_{g_i(x)>0} g_i(x)+\sum_{j=1}^{l}|h_j(x)|\Big\} \leqslant \sum_{g_i(x)>0} u_i^{(1)}(x) g_i(x) \\ & -c \sum_{g_i(x)>0} g_i(x)-\sum_{j=1}^{l}(c-|u_j^{(4)}(x)|)|h_j(x)| \\ =\ & -\Big\{\sum_{g_i(x)>0}(c-u_i^{(1)}(x)) g_i(x)+\sum_{j=1}^{l}(c-|u_j^{(4)}(x)|)|h_j(x)|\Big\}. \end{aligned}$$

因 $x \notin R$，$c \geqslant \bar{c}(x)$，即 $c-u_i^{(1)}(x)>0$，$c-|u_j^{(4)}(x)|>0$，上式右端至少有一项不为 0，从而 $DF_c(x, d(x))<0$.

总结以上，可得如下定理：

定理 1. 若 $x \in R$，则当 $DF_c(x, d(x))=0$ 时，x 为 (NP) 问题的 K-T 点；当 $x \notin R$ 时，$LP(x)$ 的解 $d(x)$ 为 $F_c(x, d(x))$ 的下降方向.

§3. 算 法

本文算法的迭代步骤如下：

步骤 0. 任取 $x^1 \in E_n$，数值 $\delta>0$，$c_0>0$，$k:=1$.

步骤 1. 解线性规划 $LP(x^k)$，得最优解 d^k 以及单纯形乘子 $U(x^k)$，计算 $\bar{c}(x^k)$.

若 $x^k \notin R$，当 $\bar{c}(x^k)>c_{k-1}$ 时，令 $c_k=\max\{\bar{c}(x^k), c_{k-1}+\delta\}$；当 $\bar{c}(x^k) \leqslant c_{k-1}$ 时，令 $c_k=c_{k-1}$.

若 $x^k \in R$，令 $c_k=c_{k-1}$. 转到步骤 2.

步骤 2. 若 $x^k \in R$ 且 $DF_c(x^k, d^k)=0$，则 x^k 为 (NP) 的 K-T 点. 否则，转

到步骤 3.

步骤 3. 求线搜索步长 λ_k，满足

$$F_{c_k}(x^k+\lambda_k d^k)\leqslant F_{c_k}(x^k),$$
$$|\lambda_k-\lambda_k^*|\to 0,$$

其中 $F_{c_k}(x^k+\lambda_k^* d^k)=\min\limits_{\lambda\geqslant 0}\{F_{c_k}(x^k+\lambda d^k)\}$.

令 $x^{k+1}=x^k+\lambda_k d^k$，$k:=k+1$，转回步骤 1.

§4. 算法的收敛性

由算法可知，若迭代于 x^k 终止，则 x^k 为 K-T 点，否则将产生无穷点列 $\{x^k\}$. 我们将证明，$\{x^k\}$ 的任一有限极限点均为 (NP) 的 K-T 点.

引理 5 设 $\{x^k\}$ 为算法产生的点列，且存在某紧集 S_0，使 $x^k\in S_0$，$k=1,2,\cdots$成立，则必存在 k_0，使得对 $\forall k\geqslant k_0$，均有 $c_k\equiv c_{k_0}$.

证明. 设 x^* 为 $\{x^k\}$ 的任一有限极限点，$x^*\in S_0$，且子列 $x^k\to x^*$. 由于对任一 k，$LP(x^k)$ 有最优解，由线性规划的稳定性理论，可知 $LP(x^*)$ 亦有最优解. 因而对偶规划的最优解 $U(x^k)$ 必有界. 否则，若 $\|U(x^k)\|\to+\infty$，则由 $LD(x^*)$ 无有界解得 $LP(x^*)$ 无最优解. 此为矛盾. 由 $\|U(x^k)\|<+\infty$，可知在算法的步骤 1 中 $\bar{c}(x^k)>c_{k-1}$ 的情形不可能无穷次出现，否则由 $c_k>c_{k-1}$，得 $c_k\to+\infty$，$(k\to\infty)$，但 $c_k\leqslant\bar{c}(x^{k+1})\leqslant\operatorname{Sup}\|U(x^k)\|<+\infty$，此亦矛盾. 因此存在 k_0，当 $k\geqslant k_0$ 时，$c_k\equiv c_{k_0}$.

定理 2. 设算法产生的点列为 $\{x^k\}$，$\{d^k\}$ 为相应的搜索方向，若 S_0 为某紧集，$x^k\in S_0$，$k=1,2,\cdots$，$x^k\to x^*$，则 x^* 为 (NP) 的 K-T 点.

证明. 由 $x^k\to x^*$，$-1\leqslant d_i^k\leqslant 1$，不妨设 $d^k\to d^*$. 由线性规划的稳定性理论知 d^* 为 $LP(x^*)$ 的最优解. 另一方面，由引理 5，对充分大的 k，$c_k\equiv c$ 且 $c\geqslant\bar{c}(x^*)$. 若 x^* 不是 (NP) 的 K-T 点，则由定理 1 及当 $x^*\in R$ 时，由 $d=0\in S(x^*)$，可知 $\nabla f(x^*)^T d\leqslant 0$，从而得知 $DF_c(x^*,\ d^*)<0$. 因此，若记 $\bar{\lambda}$ 为满足：$F_c(x^*+\bar{\lambda}d^*)=\min\limits_{\lambda\geqslant 0}\{F_c(x^*+\lambda d^*)\}$，则必有 $\bar{\lambda}>0$，且 $b\triangleq F_c(x^*)-F_c(x^*+\bar{\lambda}d^*)>0$. 又当 $k\to\infty$ 时，$x^k+\bar{\lambda}d^k\to x^*+\bar{\lambda}d^*$，而 k 充分大时，

$$F_c(x^k+\bar{\lambda}d^k)+\frac{1}{2}b<F_c(x^*)\tag{11}$$

成立.

另一方面，由 $F_c(x)$ 的连续性及 $|\lambda_k-\lambda_k^*|\to 0$，当 k 充分大时，有

$$F_c(x^k+\lambda_k d^k)\leqslant F_c(x^*+\lambda_k^* d^k)+\frac{1}{4}b\tag{12}$$

成立. 注意到 λ_k^* 的极小性及 (11)，得

$$F_c(x^k+\lambda_k d^k)+\frac{1}{4}b\leqslant F_c(x^*+\lambda_k^* d^*)+\frac{1}{2}b\leqslant F_c(x^*+\bar{\lambda}d^*)$$
$$+\frac{1}{2}b<F_c(x^*).$$

由于 $F_c(x^k)$ 单调下降,得 $F_c(x^*)+\frac{1}{4}b<F_c(x^*)$. 但 $F_c(x^*)$ 是有限数,得 $b\leqslant 0$, 此与 $b>0$ 矛盾,从而 x^* 为 (NP) 的 K-T 点.

在算法与证明过程中,线搜索可改用其他方式. 例如,当 λ_k 的计算满足

$$F_{c_k}(x^k+\lambda_k d^k)\leqslant F_{c_k}(x^k),$$
$$F_{c_k}(x^k+\lambda_k d^k)\leqslant \min_{\lambda\geqslant 0}F_{c_k}(x^k+\lambda d^k)+\varepsilon_k.$$

$\varepsilon_k\geqslant 0$, $\sum_{k=1}^{\infty}\varepsilon_k<+\infty$ 时,上述论证仍然成立.

计算实例: 令 $x=(x_1,x_2,x_3)$,

$$(NP)_0\qquad \begin{aligned} &\min f(x)=1000-x_1^2-2x_2^2-x_3^2-x_1x_2-x_1x_3,\\ &\text{s.t.}\quad h_1(x)=x_1^2+x_2^2+x_3^2-25=0,\\ &\qquad\quad h_2(x)=8x_1+14x_2+7x_3-56=0.\end{aligned}$$

已知问题 $(NP)_0$ 的最优解 $x^*=(4.846,1.231,0.000)$, $f(x^*)=967.524$.

今取初始点 $x^0=(2,2,2)$, $f(x^0)=976$ 开始计算,经 19 次迭代得近似最优解 $\tilde{x}=(4.835,1.234,0.015)$, $f(\tilde{x})=967.538$.

参 考 文 献

[1] S. P. Han, A globally convergent method for nonlinear programming, Journal of optimization theory and Applications, 22: 23(1977) 297—309.

[2] Hirsohi Yamashita., A globally convergent constrained Quasi-Newton method with augmented Lagrangian type penalty function., Mathematical programming 23(1982), 75—86.

[3] 张连生,对于非线性等式约束的增广 L-罚函数的拟牛顿法,计算数学,1(1986),90—94.

[4] M. SAHBA., Globally convergent algorithm for nonlinearly constrained optimization problems, Journal of optimization Theory and applications, Vol. 52, No. 2. February, 1987. 291—309.

[5] M. S. Bazaraa, C. M. Shetty., Nonlinear programming, Theory and Algorithms., John wiley and Sons. Inc. 1979.

[6] M. Avriel., Nonlinear programming (Analysis and methods), pretice Hall. Inc. Englewed Cliffs. New Jersy, 1976.

[7] 杜 藏,线性逼近——效益函数法及其收敛性分析,南开大学学报, No. 2, Jun, (1987), 56—59.

[8] G. Di Pillo, L. Grippo, On the Exactness of a class of nondifferentiable penlty functions. Journal of optimization theory and Application., Vol. 57, No. 3, Jun. (1988). 399—410.

[9] 马仲蕃,魏权龄,赖炎连,数学规划讲义,中国人民大学出版社,1981.