第12卷第1期　　数 学 研 究 与 评 论　　Vol. 12 No. 2
1992年5月　　JOURNAL OF MATHEMATICAL RESEARCH AND EXPOSITION　　May 1992

非线性约束优化问题的一个新的全局收敛算法*

韦 增 欣
(广西大学数学系，南宁 530004)

摘　要

对于非线性约束的优化问题，最近给出的各种SQP算法均采用罚函数技巧以保证算法的全局收敛性，因而都必须小心地调整惩罚参数。本文给出一个不依赖于惩罚参数、每步迭代的校正矩阵也不需正定且仍具有全局收敛性的SQP方法，而且罚函数形式简单、具有和约束函数同阶的光滑性.

§1 引　言

考虑非线性规划问题(NLP)，$\min\limits_{x\in R} f(x)$. 其中 $x\in E^n$. $R=\{x\,|\,g_i(x)\leqslant 0,i=\overline{1,m};h_j(x)=0,j=\overline{1,l}\}$. 且始终假定：$(H_1)$：$f(x)$、$g_i(x)(i=\overline{1,m})$，$h_j(x)(j=\overline{1,l})$为一阶连续可微函数.

目前用SQP方法求解上述(NLP)时，为了保证算法的全局收敛性，均需采用带罚参数的罚函数作为下降函数[1-9]，而且除了依赖域方法[6,7]外，每步迭代中的校正矩阵至少要求在约束梯度的零空间中正定. 本文吸取[2,6,8]等的思想，先解一个线性子规划以产生依赖域的最小半径，然后再解一个二次规划产生迭代点处的搜索方向. 由于步长的选取只须使下一个迭代点更为靠近可行域，故不需要对带有罚参数的罚函数进行搜索，而且算法保持了依赖域方法不要求校正矩阵正定而只要求其有界，而又具有全局收敛性的优点. 特别还克服了[2]等算法不能用标准的Armijo型求步长等缺陷.

§2 算　法

设 $\varphi(\cdot):E^{1^+}=[0,+\infty]\to E^{1+}$ 的连续数且满足：$\forall\, t\in E^{1^+}$，$0\leqslant\varphi(t)\leqslant 1$，且 $\varphi(t)=0\Leftrightarrow t=0$. 设 $l_k>0$，$x^k\in E^n$，线性子规划$(LP)(x^k,l_k)$、二次规划$(QP)(x^k,l_k)$分别为如下两个问题：$d=(d_1,\cdots,d_n)$.

$$(LP)(x^k,l_k)\begin{cases}\min & t\\ \text{s.t:} & \varphi(|g_i(x^k)|)g_i(x^k)+\nabla g_i(x^k)^T d\leqslant 0 \quad (i=\overline{1,m})\\ & \varphi(|h_i(x^k)|h_j(x^k)\nabla h_j(x^k)^T d\leqslant 0 \quad (l=\overline{1,l})\\ & -t\leqslant d_p\leqslant t \quad (p=\overline{1,n})\\ & t\geqslant l_k\end{cases}$$

并用 t_k 记$(LP)x^k,l_k)$的最优解$(t_k,\bar{d}^k)$的第一个分量. 采用线性规划知识易证，即使$(LP)(x^k,l_k)$的

* 1990年5月4日收到，广西大学科学基金资助课题.

最优解不唯一，但第一个分量必相同：即若$(\hat{l}_K,\hat{d}^k)$也为(LP)(x^k,l_k)的最优解，则$l_k=\hat{l}_k$.

$$
(\mathrm{QP})(x^k,l_k)\begin{cases}\min \quad \nabla f(x^k)^T d+\dfrac{1}{2}d^T B_k d \\ \quad \varphi(|g_i(x^k)|)g_i(x^k)+\nabla g_i(x^k)^T d\leqslant 0 \quad (i=\overline{1,m}) \\ \quad \varphi(|h_j(x^k)|h_j(x^k)+\nabla h_j(x^k)^T=0 \quad (l=\overline{1,l}) \\ \quad \dfrac{1}{2}d^T d\leqslant nl_k^2\end{cases}
$$

容易验证：若$(l_k,\bar{d}^k)$为(LP)(x^k,l_k)的最优解，则$\bar{d}^k$为(QP)(x^k,l_k)的一个可行解，注意到目标函数为d的连续函数，QP$({}^k,l_k)$的可行集非空且有界(因为$\frac{1}{2}d^T\leqslant nl_k^2$)可得QP$(x^k,l_k)$必有最优解(不一定唯一)．设$d^k$为QP$(x^k,l_k)$的一个最优解(或K－T点)．

记

$$
\psi(x)=\sum_{i=1}^{m}(\max\{0,g_i(x)\})^q+\sum_{j=1}^{l}|h_j(x)|^q \quad (q\geqslant 1)
$$

$$
D\psi(x,d)=\lim_{t\to 0}\frac{\psi(x+td)-\psi(x)}{t}.
$$

有了上面记号，我们可以给出本文算法：

步骤 0：任取一序列$\{l_k>0\}$，$l_k\to 0$(一般取l_k，使$\frac{l_{k+1}}{l_k}\to 0$)，$x^1\in E^n$，B_1为$n\times n$矩阵(一般取B_1为n阶单位阵)．$k:=1$.

步骤 1：求解$(LP)(x^k,l_k)$，设其最优解的第一个分量为l_k．解(QP)(x^k,l_k)．任取它的一个K－T点．

步骤 2：若$d^k=0$，停止，x^k为$(N$LP)的K－T点．

步骤 3：若$x^k\in R$(此时当然$d^k\neq 0$)，则令$x^{k+1}=x^k$，$k:=k+1$，返回步1.

步骤 4．求步长λ_k满足如下方式之一：

$$
(\mathrm{I}):\begin{cases}\psi(x^k+\lambda_k d^k)\leqslant\psi(x^k) \\ \psi(x^k+\lambda_k d^k)\leqslant\min\{\psi(x^k+\lambda d^k)\,|\,0\leqslant\lambda\leqslant\delta\}+\varepsilon_k\end{cases}
$$

其中$\delta>0$为固定正数，数列$\varepsilon_k\geqslant 0$且满足$\sum_{k=1}^{\infty}\varepsilon_k<+\infty$.

(II)：λ_k为$\{1,\beta,\beta^2,\cdots\}$中满足下式之最大者：

$$
\psi(x^k+\lambda_k d^k)-\psi(x^k)\leqslant C\lambda_k D\psi(x^k,d^k),
$$

其中$C\in(0,\frac{1}{2})$，$\beta\in(0,1)$.

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

注　对(II)，一般仅在$q>1$时采用，此时$\psi(x)$连接可微，$D\psi_{(x^k,d^k)}=\nabla\psi_{(x^k)}{}^T d^k$；而(I)对$q\geqslant 1$均适用．

引理 2.1　$\forall\, x\in E^n$，$\forall\, d\in E^n$

i)$q=1$时，$D\psi(x;d)=\sum_{g_i(x)>0}\nabla g_i(x)^T d+\sum_{g_i(x)=0}\nabla g_i(x)^T d+\sum_{h_j(x)>0}\nabla h_j(x)^T d+$

$$\sum_{h_j(x)=0}|\nabla h_j(x)^T d|+\sum_{h_j(x)<0}(-\nabla h_j(x)^T d);$$

ii) $q>1$ 时，$D\psi(x;d)=\nabla\psi(x)^T d=q\{\sum_{i=1}^{m}(\max\{0,g_i(x)\})^{q-1}\nabla g_i(x)^T d$

$$+\sum_{h_j(x)\geqslant 0}|h_j(x)|^{q-1}\nabla h_j(x)^T d+\sum_{h_j(x)<0}|\nabla h_j(x)|^{q-1}(-\nabla h_j(x)^T d)\}.$$

证明 i)见[10]，ii)仿[2]可证.

定理 2.2 若在算法或迭代过程中，$d^k=0$，则 x^k 为原问题的 K－T 点. 若出现 $d^k\neq 0$ 且 $x^k\overline{\in}R$，则 $D\psi(x^k;d_k)<0$.

证明 因 d^k 为(QP)(x^k,t_k)的 K－T 点，故有如下 K－T 方程组：(2.1)－(2.8).

$$\nabla f(x^k)+B_k d^k+\sum_{i=1}^{m}u_i^k\nabla g_i(x^k)+\sum_{j=1}^{l}v_j^k\nabla h_j(x^k)+u_0^k d^k=0 \tag{2.1}$$

$$\varphi(|g_i(x^k)|)g_i(x^k)+\nabla g_i(x^k)^T d^k\leqslant 0 \tag{2.2}$$

$$u_i^k[\varphi(|g_i(x^k)|)g_i(x^k)+\nabla g_i(x^k)^T d^k]=0 \tag{2.3}$$

$$u_i^k\geqslant 0 \qquad (i=\overline{1,m}) \tag{2.4}$$

$$\varphi(|h_j(x^k)|)h_j(x^k)+\nabla h_j(x^k)^T d^k=0 \tag{2.5}$$

$$\frac{1}{2}d^{k^T}d_k-nt_k^2\leqslant 0 \tag{2.6}$$

$$u_0^k[\frac{1}{2}]d^{k^T}d^k-nt_k^2]=0 \tag{2.7}$$

$$u_0^k\geqslant 0 \tag{2.8}$$

由 $d^k=0$ 及 $\varphi(\cdot)$的定义，从(2.1)－(2.5)可得：

$$\nabla f(x^k)+\sum_{i=1}^{m}u_i^k\nabla g_i(x^k)+\sum_{j=1}^{l}v_j^k\nabla h_j(x^k)=0 \tag{2.1$'$}$$

$$g_i(x^k)\leqslant 0 \tag{2.2$'$}$$

$$u_i^k g_i(x^k)=0 \qquad (i=\overline{1,m}) \tag{2.3$'$}$$

$$u_i^k\geqslant 0 \tag{2.4$'$}$$

$$h_j(x^k)=0 \qquad (j=1,l) \tag{2.5$'$}$$

此式表明：x^k(NLP)的 K－T 点.

若 $d^k\neq 0$ 且 $x^k\overline{\in}R$，记 $J_k=\{i,j|g_i(x^k)>0,(i=\overline{1,m}),h_j(x_k)\neq 0(j=\overline{1,l})\}$，则 $J_k\neq\varnothing$.

由引理 2.1 及 d^k 为(QP)(x^k,t_k)的一个可行解及(2.2)，(2.5)可得：当 $q=1$ 时，

$$D\psi(x^k,d^k)=\sum_{g_i(x^k)>0}\nabla g_i(x^k)^T d^k+\sum_{j=1}^{1}(-|h_j(x^k)|\varphi_|h_j(x^k))$$

$$\leqslant-\{\sum_{g_i(x^k)>0}\varphi(|g_i(x^k)|)g_i(x^k)+\sum_{j=1}^{l}|h_j(x^k)|\varphi(|h_j(x^k)|)\}.$$

因 $J_k\neq\varnothing$，故上式括号中至少有一项不为 0，从而 $D\varphi(x^k;d^k),<0$. 当 $q>1$ 时，

$$D\varphi(x^k;d^k)=q\{\sum_{=1}^{m}(\max\{0,g_i(x^k)\})^{q-1}\nabla g_i(x^k)^T d^k$$

$$+\sum_{h_j(x^k)>0}|h_j(x^k)|^{q-1}\nabla h_j(x^k)^T d^k+\sum_{h_j(x^k)<0}|h_j(x^k)|^{q-1}(-\nabla h_j(x^k)^T d^k)\}$$

$$\leqslant -q\{\sum_{g_i(x^k)>0}(g_i(x^k)^q\psi(|g_i(x^k)|)+\sum_{j=1}^{l}|\varphi h_j(x^k)|\}.$$

由 $J_k\neq\varnothing$ 得 $D\psi(x^k;d^k)<0$. 证毕.

由上述定理,所给算法是可以执行的,或者有限步后终止,或者产生一无穷点列.

§3 算法的全局收敛性

定理 3.1 设$\{x^k\}$为算法产生的无穷点列,相应的搜索方向序列为$\{d_k\}$. 如果下列条件满足:

(H_2):$\{B_k\}$有界;

(H_3):存在一个紧集 X,使$\{x^k\}\subseteq X$;

(H_4):存在紧集 $Y\subseteq E^n\times E^l\times E^1$,使得对每个 k,QP(x^k,t_k)的 K−T 型向量$(d^k,u^k,v^k,u_0^k)\in Y$. 则$\{x^k\}$的任一聚点均为(NLP)的 K−T 点.

证明 设$\{x^k\}_{\mathscr{K}}\to x^*$. 由$(H_4)$,不妨设$\{d^k\}_{\mathscr{K}}\to d^*$,$\{u^k\}_{\mathscr{K}}\to u^*$,$\{v^k\}_{\mathscr{K}}\to v^*$,$\{u_0^k\}\to u_0^*$.

<一>:先证 $x^*\in R$.

情形 1:若$\{x^k\}_{\mathscr{K}}$为有限集,即算法在 $\mathscr{K}$ 中仅执行步骤 4 有限次,由步 3,必有 k 充分大后,$\forall\ k\in\mathscr{K}$,$x^{k+1}=x^k\in R$,故 $x^*\in R$.

情形 2:$\{x^k\}_{\mathscr{K}}$为无限集,用反证法,设 $x^*\in R$. 由(2.2),(2.5)令 $k\to\infty$,$k\in\mathscr{K}$ 可得:

$$\psi(|g_i(x^*)|)g_i(x^*)+\nabla g_i(x^*)^Td^*\leqslant 0\quad(i=\overline{1,m}).$$

$$\psi_{|h_j(x^*)|}h_j(x^*)+\nabla h_j(x^*)^Td^*\leqslant 0\quad(j=\overline{1,l}).$$

因 $x^*\notin R$,故 $d^*\neq 0$,完全相同于定理 2.2 的证明可证得:

$$D\psi(x^*,d^*)<0 \tag{3.1}$$

1°) 若步长选取采用方式(I)($q\geqslant 1$ 均适用). 由 $\lim\limits_{t\to 0}\dfrac{\psi(x^*+td^*)-\psi(x^*)}{t}=D\psi(x^*;d^*)<0$ 可得 $\bar\lambda>0$ 及 $0\leqslant\bar\lambda\leqslant\delta$,$\psi(x^*+\bar\lambda d^*)<\psi(x^*)$.

记 $a=\psi(x^*)-\psi(x^*+\lambda d^*)$,则 $a<0$. 由 $x^k+\bar\lambda d^k\to x^*+\bar\lambda d^*$ 知存在 k_1,当 $k\in\mathscr{K}$,$k\geqslant k_1$ 时.

$$\psi(x^k+\bar\lambda d^k)+\frac{1}{2}a<\psi(x^*). \tag{3.2}$$

但对充分大的 k,$\psi(x^{k+1})<\psi(x^k)+\varepsilon_k$ 及 $\sum\limits_{i=k}^{\infty}\varepsilon_i<\dfrac{1}{2}a$,因而:$\psi(x^*)<\psi(x^{k+1})+\sum\limits_{i=k+1}^{\infty}\varepsilon_i\leqslant\min\{\psi(x^k+\lambda d^k)\mid 0\leqslant\lambda\leqslant\delta\}+\varepsilon_k+\sum\limits_{i=k+1}^{\infty}\varepsilon_i<\psi(x^k+\bar\lambda d^k)+\dfrac{1}{2}a$.

此式与(3.2)矛盾,故 $x^*\in R$.

2°). 若步长选取用方式(II)($q>1$)

由(3.1),必存在 $k_2>1$,$c<\tau_1<1$,使得 $\psi(x^*+\beta^{k_2}d^*)-\psi(x^*)\leqslant\tau_1\beta^{k_2}D\psi(x^*;d^*)$. 由 $q>1$,$\psi(x)$连续可微,故 $D\psi(x^k;d^k)\to D\psi(x^*;d^*)$,从而必存在 $\tau_2>1$,$c>\tau_1\tau_2^{-1}<1$,$k_3\geqslant k_2$,使得$\forall\ k\in\mathscr{K}$,$k\geqslant k_3$ 均有 $D\psi(x^k;d^k)\geqslant\tau_2D\psi(x^*;d^*)$. 再由 $\psi(x)$的连续性,存在 $k_4\geqslant k_3$,$\tau_3<1$,$c\leqslant\tau_1\tau_2^{-1}\tau_3<1$,使得 $\forall\ k\geqslant k_4$,$k\in\mathscr{K}$,均有 $\psi(x^k+\beta^{k_2}d^k)-\psi(x^k)\leqslant\tau_3[\psi(x^*+\beta^{k_2}d^*)-\psi(x^*)]$,从而当 $k\geqslant k_4$,$k\in\mathscr{K}$ 时,

$$\begin{aligned}\psi(x^k+\beta^{k_2}d^k)-\psi(x^k)&\leqslant\tau_3[\psi(x^*+\beta^{k_2}d^*)-\psi(x^*)]\\&\leqslant\tau_1\tau_3\beta^{k_2}D\psi(x^*;d^*)\leqslant\tau_1\tau_3\tau_2^{-1}\beta^{k_2}D\psi(x^k;d^k)\leqslant c\beta^{k_2}D\psi(x^k;d^k)\end{aligned}$$

上式表明,当 $k\geqslant k_4$ 时,

$$\psi(x^k+\beta^{k_2}d^k)-\psi(x^k)\leqslant c\beta^{k_2}D\psi(x^k;d^k) \qquad k\in\mathscr{K}.$$

按照方式(II)，当 $k\geqslant k_4, k\in\mathscr{K}$ 时，$\lambda_k\geqslant\beta^{k_2}$，故

$$\inf\{\lambda_k \mid k\in \mathrm{k}\}\geqslant \beta_2^k>0 \tag{3.3}$$

由 $\psi(x_k+\lambda_k d^k)-\psi(x^k)\leqslant c\lambda_k D\psi(x^k;d^k)$ 及 $\{\psi(x^k)\}$ 单调下降因而有极限 ψ^* 得

$$\psi^*-\psi\leqslant c\lim_{\substack{k\in\mathscr{K}\\ k\to\infty}}\lambda_k D\psi(x^*;d^*).$$

由(3.3)式可得：$D\psi(x^*;d^*)\geqslant 0$. 这与(3.1)矛盾. 从而 $x^*\in R$. 至此<一>证毕.

<二>：证 $\{t_k\}_{\mathscr{K}}\to 0$.

设 $(t_k,\hat{d}^k)$ 为(LP)(x^*,t_k)的最优解. 因 $x^*\to x^*(k\in\mathscr{K})$ 由线性规划的稳定性理论可知，$(t_k,\hat{d}^k)\to(\bar{t},\bar{d})$，其中$(\bar{t},\bar{d})$为(LP)$(x^*,0)$的最优解.

另一方面由<一>$x^*\in R$，故$(t,d)=(0,0)$为(LP)$(x^*,0)$的一个可行解，由(LP)$(x^*,0)$的定义，其最优解的第一个分量 $\bar{t}=0$. 从而 $t_k\to 0$. 故<二>获证.

下证 x^* 为(NLP)的 K－T 点.

利用<一>、<二>的结果，在 K－T 方程组(2.1)－(2.8)令 $k\to+\infty, k\in\mathscr{K}$ 可得：

$$\begin{cases}\nabla f(x^*)+\sum_{i=1}^{m}u_i^*\nabla g_i(x^*)+\sum_{j=1}^{l}v_j^*\nabla h_j(x^*)=0\\ g_i(x^*)\leqslant 0\\ u_i^*g_i(x^*)=0 \qquad (i=\overline{1,m})\\ u_i^*\geqslant 0\\ h_j(x^*)=0 \qquad (j=\overline{1,l})\end{cases}$$

此式表明：x^* 为(NLP)的 K－T 点. 证毕.

注 1 本文定理 3.1 的条件比[9]的收敛性定理的条件弱.

注 2 由于 $q>1$ 时，$\psi(x)$连续可微，采用标准的 Armijo 型步长规则，再加上 t_k 选取的任意性，可望通过取 t_k 快速收敛于 0 而使算法有强收敛性.

中科院应用数学所的赖炎连研究员对本文提出了有益的修改意见，谨此致谢！

参 考 文 献

[1] U. M. Garcia Palmares. and O. L. Mangasarian, Math. Prog. ,Vol. 11. (1976)1－13.

[2] S. P. Han, Journal of Optimuzation Theory and Application, Vol. 22. No. 3, July 1977. 297－309.

[3] R. Flet cher , *Second Order Correction for Nondifferentiable Optimization*. in: G. A. Warson. ed. Numerical Analysis (Spring－Verlag. Berlin, 1982). pp. 85－115.

[4] R. M. Chamberlain, C. Lemarechat, H. C. Pederson and M. J. D. Powell, Math. Pro. 16(1982)1－17.

[5] G. Di Pillo, F. Facchinei and L. Grippo. *A RQP Algorithm Using a Differentiable Exact Penalty Function for Nonlinere Programming With Inequalities*, *to be presented at the 13th International Symposium on Mathematical Programming*, Toyko, 1988.

[6] R. H. Byrd, R. B. Schnabel and G. A. Shultz, *A trust region algorithm for nonlinearly constrained optimization*, Reprot. 1985.

[7] M. R. Celis, J. E. Dennis Jr. and R. A. Tapia, *A Trust Region Strategy for Nonlineal Equality Constrained Optimization*. in: P. T. Boogs, R. H. Byrd and R. B. Schnabl, eds. Numerical Optimization (SIAM,

Philadelphia, 1985)pp, 71—82.
[8] M. Sahba, *Journal of Optimization Theory and Application*. Vol. 52. No. 2, February 1987, 291—309.
[9] Hiroshi Yamashita, Math. Pro. 23. (1982)75—86.
[10] G. Di Pillo and L. Grippo, *Journal of Optimization Theory and Application*, Vol. 57, No. 3, June 1988. 399—410.

A New Gonvergent Algotithm for Nonlinear Constrainted Optimiztion Problems

Wei Zhenxin
(Dept. of Math., Guangxi University, Nanning, China)

Abstract

In order to guarantee the global convergence, the existing SQP methods for the nonlinear constrainted optimization problems all use the technique of penalty functions,so they must correct the penalty parameter carefully. In this paper, we present a SQP method which not only does not depend on the penalty parameter, but also possesses global convergence property. Moreover, the penalty functions are of simple from and of the same smooth order as the constraint functions.