线性约束非线性规划问题的两个算法*

赵 福 安
(曲阜师大运筹学研究所)

TWO ALGORITHMS FOR LINEARLY CONSTRAINED OPTIMIZATION

Zhao Fu-an
(Institute of O.R., QUfU Normal University)

Abstract

This paper presents two algorithms for the linearly constrained optimization problem. The convergence of the algorithms are proved and the convergence rates estimated. One is linear and the other is superlinear.

§1. 符号和假设

本文讨论的问题为

$$(\mathrm{LNP}) \qquad \min\{f(x)\,|\,x\in R=\{x\in R^n\,|\,a_i^Tx\leqslant b_i, i=1,\cdots,m\}\}.$$

对 $x\in R$，记 $I(x)=\{i\,|\,a_i^Tx=b_i,\ i=1,\cdots,m\}$，$I_k=I(x^k)$，$|I|$ 表示 I 中元素的个数。用 N_I 表示以 $a_i(i\in I)$ 为行向量构成的矩阵。令 $\Omega\subseteq\{1,2,\cdots,n\}$ 为任一指标集，N_I^Ω 表示矩阵 N_I 中指标属于 Ω 的行构成的子矩阵。

设 $\Omega\subseteq\{1,\cdots,n\}$，记 $\bar\Omega=\{1,\cdots,n\}\backslash\Omega$。如果 $|\Omega|=|I|$，N_I^Ω 可逆，则称 Ω 为 N_I 的一组基。此时记

$$\bar f(x_{\bar\Omega})=f((N_I^\Omega)^{-1}b-[T_I^{\bar\Omega}(\Omega)]x_{\bar\Omega},\ x_{\bar\Omega}),$$

其中 $T_I^{\bar\Omega}(\Omega)=(N_I^\Omega)^{-1}N_I^{\bar\Omega}$，$x_{\bar\Omega}=(x_{j_1},\ \cdots,\ x_{j_{n-|I|}})^T(j_i\in\bar\Omega,\ i=1,2,\cdots,n-|I|)$。

如果 $f(x)\in c^1$，则

$$\nabla\bar f(x_{\bar\Omega})=\nabla_{\bar\Omega}f(x)-T^{\bar\Omega}(\Omega)\nabla_\Omega f(x),$$

其中 $\nabla_\Omega f(x)=\left(\dfrac{\partial f(x)}{\partial x_i},\ i\in\Omega\right)$。

全文假设

(H_1)：对任意的 $x\in R$，$\{a_i\,|\,i\in I(x)\}$ 线性无关。

* 1985 年 9 月 21 日收到。

(H_2): $f(x)\in c^1$.

称 $x\in R$ 是问题（LNP）的 K-T 点，如果

$$\nabla f(x)=\sum_{i\in I(x)}\lambda_i a_i,\ \lambda_i\leqslant 0,\ i\in I(x).$$

用 $G>0$ 表示 G 为正定矩阵。

§2. 算 法 I

下面是解（LNP）的一个改进投影方法

算法 I

1° 任取 $x^0\in R$，计算 $I_0=I(x^0)$，$g_0=\nabla f(x^0)$，$z_0^1=(N_{I_0}N_{I_0}^T)^{-1}N_{I_0}$，求出 N_{I_0} 的一个基 Ω_0. 令

$$z_0^2=\begin{bmatrix}-(N_{I_0}^{\Omega})^{-1}N_{I_0}^{\bar{\Omega}}\\ E_0\end{bmatrix},$$

其中 E_0 是 $(n-|I_0|)$ 阶单位矩阵，$\beta_0=0$，$0<\delta<\dfrac{1}{2}$，令 $k:=0$.

2° 计算 $\lambda_{I_k}=(z_k^1)^Tg_k$，$\bar{\lambda}_k=(z_k^2)^Tg_k$，确定 r、t，使

$$\lambda_k^{(r)}=\max_{i\in I_k}\lambda_i,\quad \bar{\lambda}_k^{(t)}=\max_{1\leqslant i\leqslant n-|I_0|}|\bar{\lambda}_{k_i}|.$$

如果 $|I_k|=0$，令 $\lambda_k^{(r)}=0$；如果 $|I_k|=n$，令 $\bar{\lambda}_k^{(t)}=0$.

3° 令

$$s^k=\begin{cases}z_k^2\bar{\lambda}_k, & 若\ |\bar{\lambda}_k^{(t)}|\geqslant\delta\lambda_k^{(r)}\ 或\ \beta_k=1,\ \bar{\lambda}_k^{(t)}=0,\ 此时令\ r_k=0,\\ \lambda_k^{(r)}z_k^{(r)}, & 其它情况，此时令\ r_k=1,\end{cases}$$

其中 $z_k^{(r)T}$ 是矩阵 $(N_{I_k}N_{I_k}^T)^{-1}N_{I_k}$ 的第 r 行向量. 如果 $s^k=0$，则停，否则转 4°.

4° 令

$$\bar{a}^k=\min_{i\notin I_k}\left\{\frac{b_i-a_i^Tx^k}{-a_i^Ts^k}\ \Big|\ -a_i^Ts^k>0\right\},\tag{2.1}$$

若 $\{i\mid -a_i^Ts^k>0\}=\phi$，则令 $\bar{a}^k=+\infty$. 利用下述方法之一确定步长 a^k:

方式 1 令 $\hat{a}^k=\min\{1,\ a^k\}$，$v_k\geqslant 0$ 是满足 $\phi_k(0.5^{v_k}a^k)\geqslant\delta$ 的最小非负整数，其中

$$\phi_k(a)=\frac{f(x^k)-f(x^k-as^k)}{ag_k^Ts^k},\ a>0.\tag{2.2}$$

置 $a^k=0.5^{v_k}\hat{a}^k$，$x^{k+1}=x^k-a^ks^k$，转 5°.

方式 2 求 a^k，使 a^k 满足

$$f(x^k)-f(x^k-a^ks^k)\geqslant\rho a^kg_k^Ts^k,\tag{2.3}$$

$$g(x^k-a^ks^k)^Ts^k\leqslant\sigma g_k^Ts^k,\tag{2.4}$$

$$0<a^k\leqslant\bar{a}^k,$$

其中 $\rho\in\left(0,\ \dfrac{1}{2}\right)$，$\sigma\in(\rho,\ 1)$ 是实数. 若无满足(2.3)和(2.4)的 a^k，则令 $a^k=\bar{a}^k$. 置

$x^{k+1}=x^k-a^ks^k$，转 5°。

5° 置

$$\beta_{k+1}=\begin{cases}1, & a^k=\bar{a}^k,\\ 0, & a^k<\bar{a}^k.\end{cases}$$

6° 若 $\beta_{k+1}=\gamma_k=0$，令 $I_{k+1}=I_k$，$z_{k+1}^1=z_k^1$，$z_{k+1}^2=z_k^2$，$k:=k+1$ 转 2°；否则计算 I_{k+1}，g_{k+1}，求出 $N_{I_{k+1}}$ 和 $N_{I_{k+1}}$ 的一组基 Q_{k+1}，令

$$z_{k+1}^1=(N_{I_{k+1}}N_{I_{k+1}}^T)^{-1}N_{I_{k+1}},$$
$$z_{k+1}^2=\begin{pmatrix}-(N_{I_{k+1}}^{Q_{k+1}})^{-1}N_{I_{k+1}}^{\bar{Q}}\\ E_{k+1}\end{pmatrix}, \tag{2.5}$$

其中 E_{k+1} 是 $n-|I_{k+1}|$ 阶单位矩阵，令 $k:=k+1$，转 2°。

引理 2.1 若 $s^k\neq 0$，则 $-s^k$ 是可行下降方向。

引理 2.2 若 $s^k=0$，则 x^k 是问题（LNP）的 K-T 点。

引理 2.3 设 $\{s^k\}$ 是由算法 I 产生的方向序列，并且 $s^k\neq 0$，$k=0,1,\cdots$，则存在常数 δ_i，$i=1,2,3,4$，使

〈1〉 $s^k\in U_k$ 时，$g_k^Ts^k=\bar{g}_k^Ts^k\geqslant\delta_1\|\bar{g}_k\|^2$，$\|s^k\|\leqslant\delta_2\|\bar{g}_k\|$；

〈2〉 $s^k\notin U_k$ 时，$g_k^Ts^k\geqslant\delta_3(\bar{\lambda}_k^{(t)})^2$，$\|s^k\|\leqslant\delta_4\lambda_k^{(r)}$，

其中

$$U_k=\{x\,|\,a_i^Tx=0,\ i\in I_k\}, \tag{2.6}$$

$\bar{g}_k$ 是 g_k 在 U_k 上的投影。

引理 2.4 由方式 1 确定步长的算法 I，如果存在 $\varepsilon>0$，对某 $J\subset\{0,1,\cdots\}$，$g_k^Ts^k\geqslant\varepsilon(k\in J)$，有 $f(x^k)-f(x^k-a^ks^k)\to 0(k\to\infty,k\in J)$，则对充分大的 k，$a^k=\bar{a}^k$，且 $\bar{a}^k\to 0(k\to\infty,k\in J)$.

§3. 算法 I 的收敛性质

定理 3.1 设 $f(x)$ 在包含 R 的某开集上一阶连续可微，$L_0=\{x\,|\,f(x)\leqslant f(x^0),x\in R\}$ 有界，则算法 I 或在有限步终止于（LNP）的 K-T 点，或产生一无穷点列 $\{x^k\}\subset L_0$，满足：

(i) $\lim\limits_{k\to\infty}\|x^{k+1}-x^k\|=0$；

(ii) $\{x^k\}$ 的任一极限点都是（LNP）的 K-T 点；

(iii) 当 $\{x^k\}$ 只有一个极限点 x^* 时，$\lim\limits_{k\to\infty}x^k=x^*$.

证明 如果算法在有限步终止，由引理 2.2 终点是（LNP）的 K-T 点。下设产生无穷点列 $\{x^k\}$.

(i) 若存在 $\varepsilon>0$、无穷集 $J\subset\{0,1,\cdots\}$，使 $\|x^{k+1}-x^k\|\geqslant\varepsilon(k\in J)$，则 $\|a^ks^k\|\geqslant\varepsilon$，即对某 $\varepsilon_1>0$，$\varepsilon_2>0$，有 $a^k\geqslant\varepsilon_1$，$\|s^k\|\geqslant\varepsilon_2$. 由引理 2.3，若 $s^k\in U_k$，则

$$g_k^Ts^k\geqslant\delta_1\|\bar{g}_k\|^2\geqslant\frac{\delta_3}{\delta_4}\ \|s^k\|^2\geqslant\frac{\delta_1}{\delta_2}\ \varepsilon_2^2;$$

若 $s^k\notin U_k$，则

$$g_k^T s^k \geqslant \frac{\delta_3}{\delta_4}\varepsilon_2^2.$$

故由步长的取法.

$$f(x^k)-f(x^{k+1})\geqslant \delta a_k g_k^T s^k \geqslant \delta\varepsilon_2\varepsilon_3 \quad (\varepsilon_3>0)$$

或者

$$f(x^k)-f(x^{k+1})\geqslant \rho a_k g_k^T s^k \geqslant \varepsilon_4 \ (\varepsilon_4>0),$$

再由 $f(x^k)$ 的单调下降性及 L_0 有界

$$\lim_{k\to\infty} f(x^k)-f(x^{k+1})=0$$

得矛盾.

(ii) 下面我们只考虑由方式 1 确定步长的算法 I,至于另一种方式确定步长的情况,本性质的证明可参见定理 4.1.

因 $\{x^k\}\subset L_0$, 不妨设 x^+ 为 $\{x^k\}$ 的一个极限点,并且

$$a_i^T x^+ = b_i, \quad i=1,\cdots,q, \tag{3.1a}$$

$$a_i^T x^+ < b_i, \quad i=q+1,\cdots,m. \tag{3.1b}$$

由定义,存在无穷集 $J\subset\{0,1,\cdots\}$, $x^k\to x^+(k\to\infty, k\in J)$, 设 $I\subset\{1,\cdots,m\}$ 使 $|I|$ 是满足下述条件之最大数:

存在 $J_1\subset\{0,1,\cdots\}$, $\alpha^k\to x^+(k\to\infty, k\in J_1)$, 且 $x^k\in T$

$$T=\{x\,|\,a_i^T x=b_i,\ i\in I\}, \tag{3.2}$$

即令 $\Gamma=\{\tilde I\,|$ 存在无限多个 k, $I_k=\tilde I\}$. $|I|=\max\{|\tilde I|\,|\,\tilde I\in\Gamma\}$ 由 (3.1) 可得 $I\subset\{1,\cdots,q\}$.

若 $\nabla f(x^+)$ 到

$$U^+=\{x\,|\,a_i^T x=0,\ i\in I\} \tag{3.3}$$

的投影不为 0,由 $\nabla f(x)$ 的连续性,存在 $\varepsilon>0$, 使对充分大的 $k\in J$, $\|Pg_k\|\geqslant\varepsilon$, 其中 Pg_k 为 g_k 在 U^+ 上的投影. 因此对充分大的 $k\in J$

$$s^{k-1}\in U^+ \text{ 或 } s^k\in U^+,$$

存在 $J_1\subset\{0,1,\cdots\}$, 使

$$x^k\in T,\ s^k\in U^+,\ \text{但 } a^k<\bar a^k,\ k\in J.$$

由 $g_k^T s^k\geqslant\varepsilon$, $x^k\to x^+(k\in J, k\to\infty)$, $a^k<\bar a^k$ 和引理 2.4, $f(x^k)-f(x^{k+1})\not\to 0$ $(k\to\infty,\ k\in J_1)$ 矛盾. 故存在 α_i, $i\in I$, 使

$$\nabla f(x^+)=\sum_{i\in I}\alpha_i a_i. \tag{3.4}$$

下面我们证明 $\alpha_i\leqslant 0(i\in I)$. 若不然,则 $\alpha=\max\{\alpha_i|i\in I\}>0$, 记

$$g_k=g_k^*+\hat g_k+\bar g_k,$$

其中

$$g_k^*=\sum_{\lambda_k^{(i)}\leqslant 0}\lambda_k^{(i)}a_i\ (i\in I),\quad \hat g_k=\sum_{\lambda_k^{(i)}>0}\lambda_k^{(i)}a_i\ (i\in I).$$

由于 $\bar g_k\to 0(k\to\infty)$, 故对充分大的 $k\in J$ 及某 $\varepsilon>0$, $\|\hat g_k\|\geqslant\varepsilon$, 事实上由 $\alpha>0$, 对充分大的 $k\in J$, $|\bar\lambda_k^{(+)}|>\delta_0\lambda_k^{(r)}$ 不成立.

取 I 是前述之最大指标集,如果对无限多个 $k\in J$, $s^k\in U^+$, 则 $a^k<\bar{a}^k$, 即 $\beta_{k+1}=0$, $s^{k+1}\notin U^+$, 但 $x^{k+1}\in T$, $x^{k+1}\to x^+(k\to\infty)$, 故设对充分大的 $k\in J$, $x^k\in T$, $s^k\notin U^+$, 此时 $g_k^Ts^k\geqslant\delta_3(\lambda_k^{(r)})^2>0$. 由 $f(x^k)-f(x^{k+1})\to 0(k\to\infty)$ 及引理 2.4 得 $a^k=\bar{a}^k$. 因此存在 $t\notin I$, $t\leqslant q$, $r\in I$, 使 $\alpha_r>0$, 且

$$\{x\,|\,a_i^Tx=b_i,\ i\in(I\backslash\{r\})\cup\{t\}\}$$

中包含了 $\{x^k\}$ 的无限多个元素 $\{x^k\}_{J_2}$, $x^k\to x^+$, $(k\to\infty, k\in J_2)$. 重复(3.4)式的证明可得

$$\nabla f(x^+)=\sum\beta_i a_i,\ i\in(I\cup\{t\})\backslash\{r\}. \tag{3.5}$$

比较(3.4)与(3.5)可得

$$\sum_{i\in I\backslash\{r\}}(\alpha_i-\beta_i)a_i+\alpha_r a_r-\beta_t a_t=0.$$

由条件 (H_1), $\alpha_r=0$, 矛盾,故 $\alpha_i\leqslant 0$, $i\in I$.

(iii) 本结论是结论 (i) 的推论.

定理 3.2 若 $\{x^k\}$ 是由算法 I 产生的点列, $x^k\to x^+$, $k\to\infty$, (3.1) 式成立,并且在 x^+ 处满足严格互补松弛条件,则对充分大的 k,

$$\begin{cases}a_i^Tx^k=b_i, & i=1,\cdots,q,\\ a_i^Tx^k<b_i, & i=q+1,\cdots,m.\end{cases} \tag{3.6}$$

证明 令 I 是由定理 3.1 证明中 (ii) 所定义之指标集, T, U^+ 的定义见 (3.2), (3.3).

由定理 3.1, $\nabla f(x^+)=\sum\limits_{i\in I}\alpha_i a_i$, $\alpha_i\leqslant 0$, $i\in I$, 利用严格互补松弛条件, $I=\{1,\cdots,q\}$.

记 $\bar{g}_k$ 是 g_k 到 U^+ 的投影,

$$g_k=\sum_{i=1}^{q}\lambda_k^{(i)}a_i+\bar{g}_k.$$

由 $g_k\to\nabla f(x^+)(k\to\infty)$, 故对充分大的 k, $\lambda_k^{(i)}<0(i=1,\cdots,q)$. 因此存在 k_0, $k\geqslant k_0$, $x^k\in T$, $s^k\in U^+$, 即 $x^{k+1}\in T$. 由 $a_i^Tx^k<b_i$, $i=q+1,\cdots,m$, 对充分大的 k, (3.6) 式成立.

引理 3.1 在定理 3.2 的条件下, $\varliminf\limits_{k\to\infty}\bar{a}^k\geqslant a>0$, 其中 a 是某常数.

引理 3.2 (1) 设矩阵 $G>0$, 则对任意的矩阵 $(M^T, I)_{n\times p}^T$, $p<n$,

$$\bar{G}=(M^T, I)G\begin{pmatrix}M\\ I\end{pmatrix}$$

也是正定的.

(2) $\bar{f}(y)=f((N_I^{\Omega})^{-1}b-[T^{\bar{\Omega}}(\Omega)]^Ty, y)$

在 $R^{n-|I|}$ 上也是凸的. 且当 $f(x)$ 在 R 上是严格凸时, $\bar{f}(y)$ 在 $R^{n-|I|}$ 上也是严格凸的.

定理 3.3 若算法 I 产生的点列 $x^k\to x^+(k\to\infty)$, 在 x^+ 处满足严格互补松弛条件, $f(x)\in c^2$, $G(x^+)>0$, 则

(1) $\varlimsup\limits_{k\to\infty}\dfrac{f(x^{k+1})-f(x^+)}{f(x^k)-f(x^+)}\leqslant l<1\ (l\geqslant 0)$.

进一步，如果算法是由方式 2 确定步长，则

$$\overline{\lim_{k\to\infty}} \frac{f(x^{k+1}) - f(x^+)}{f(x^k) - f(x^+)} \leqslant 1 - 8\rho(1-\sigma)\frac{\mu_1\mu_2}{(\mu_1+\mu_2)^2},$$

其中 $\mu_1\mu_2$ 分别为 $\bar{G}(x^+)$ 的最小和最大特征值。

(2)

$$\overline{\lim_{k\to\infty}} \frac{\|x_{\bar{\Omega}}^{k+1} - x_{\bar{\Omega}}^+\|}{\|x_{\bar{\Omega}}^k - x_{\bar{\Omega}}^+\|} < 1,$$

其中 Ω 为 N_I 的某个基，而

$$\bar{G}(x^+) = (-((N_I^{\Omega})^{-1}N_I^{\bar{\Omega}})^T, E)G(x^+)\begin{pmatrix} -(N_I^{\Omega})^{-1}N_I^{\bar{\Omega}} \\ E \end{pmatrix},$$

E 是 $n-|I|$ 阶单位矩阵。

证明　由定理 3.2，对充分大的 k，$I_k = I$，故设 $I_k = I = \{1, \cdots, q\}$。通过简单的代数变换，在充分大的 k 之后，算法 I 与求解

$$(P)\quad \begin{array}{l} \min \bar{f}(x_{\bar{\Omega}}) \\ x_{\bar{\Omega}} \in R^{n-|I|} \end{array}$$

的算法 III 等价(见附录)，故结论(1)成立。利用[2]中命题(2)可成立。

§4. 算法 II 及其收敛性质

把算法 I 中的技巧与变尺度方法的思想结合起来，就得到如下的算法 II。

算法 II

1° 同算法 I 中 1°，只要补充定义 $H_0 = E_{n\times n}$。

2° 同算法 I 中 2°。

3° 令

$$s^k = \begin{cases} z_k^2((z_k^2)^T H_k z_k^2)\bar{\lambda}_k, & \text{若} |\bar{\lambda}_k^{(+)}| \geqslant \delta_0\lambda_k^{(r)} \text{ 或 } \beta_k = 1 \text{ 且 } \bar{\lambda}_k^{(+)} = 0, \text{定义 } \nu_k = 0, \\ \lambda_k^{(r)} z_k^{(r)}, & \text{其它情况，定义 } \nu_k = 1. \end{cases}$$

若 $s^k = 0$，停，否则进入 4°。

4° 同算法 I 中4°(利用方法 2 确定步长)。

5° 同算法 I 中 5°。

6° 利用某种方式修正 H_k，得 H_{k+1}，使 $\{H_k\}$ 一致正定。

7° 同算法 I 中 6°。

引理 4.1　对算法 II，引理 2.1、引理 2.2 的结论成立，特别当 $\{H_k\}$ 一致正定时，引理 2.3 也成立。

定理 4.1　设 $f(x)\in c^1$，L_0 有界，$\{H_k\}$ 一致正定，则算法 II 或在有限步终止于 (LNP) 的 K-T 点，或产生一无穷点列 $\{x^k\}$，满足

(i) $\lim\limits_{k\to\infty}\|x^{k+1} - x^k\| = 0$;

(ii) $\{x^k\}$ 的任一极限点均为 (LNP) 的 K-T 点；

(iii) 当 $\{x^k\}$ 只有一个极限点时，$\{x^k\}$ 收敛。

证明 我们只证明 (ii).

类似于定理 3.2 的证明，若 $\nabla f(x^+)$ 到 U^+ 的投影不为 0，由 $\nabla f(x)$ 的连续性存在 $\varepsilon > 0$，使对充分大的 $k \in J$，$\|g_k\| \geqslant \varepsilon$. 因而，当 $k \in J$ 充分大时，$s^{k-1} \in U^+$ 或 $s^k \in U^+$. 由此可得无穷指标集 $J_1 \subset \{0, 1, 2, \cdots\}$ 使

$$x^k \in T, \ x^k \in U^+, a^k < \bar{a}^k, \ k \in J_1.$$

由引理 4.1 得

$$g_k^T s^k \geqslant \delta_1 \|g_k\|^2 \geqslant \varepsilon_1 \ (\text{某 } \varepsilon_1 > 0).$$

a^k 必须满足(2.3)和(2.4)，故

$$g_{k+1}^T s^k < \sigma g_k^T s^k.$$

若 $a^k \to 0(k \to \infty, k \in J_1)$，则 $g_{k+1} \to g^+ = \nabla f(x^+)$，$g_k \to g^+(k \to \infty, k \in J_1)$，而 $s^k \to s^+$. 设 $\Omega_k = \Omega_{k_0}$，对 $k \geqslant k_0, k, k_0 \in J_1$，$s^+ = (I - N_I^T (N_I N_I^T)^{-1} N_I^T) g^+$ 但 $g_k^T s^k \not\to 0$ $(k \to \infty, k \in J_1)$，故有

$$g^{+T} s^+ < \sigma g^{+T} s^+.$$

所以由 $\sigma < 1$ 矛盾.

如果 $a^k \not\to 0(k \to \infty, k \in J_1)$，则与 $f(x^k)$ 在 L_0 处有下界，$\{f(x^k)\}$ 单调下降矛盾，因此(3.4)式成立.

若 $\alpha = \max\{\alpha_i | i \in I\} > 0$，记 $g_k = g_k^* + \hat{g}_k + \bar{g}_k$，其中 $\hat{g}_k$ 和 g_k^* 的意义同§3.由 $\nabla f(x)$ 的连续性可得，对充分大的 $k \in J$，$\|\hat{g}_k\| \geqslant \varepsilon$ (某正数). 与§3 定理 3.2 的证明类似设

$$x^k \in T, \ s^k \notin U^+, \ k \in J.$$

此时存在某 $\varepsilon_1 > 0$，使 $g_k^T s^k \geqslant \delta_3 (\lambda_k^{(r)})^2 \geqslant \varepsilon_1$，如果对充分大的 $k \in J$，恒有 $a^k < \bar{a}^k$，则可得 $g^{+T} s^+ \leqslant \sigma g^{+T} s^+$，矛盾(因为 $\sigma < 1$). 因此设

$$x^k \in T, \ s^k \notin U^+, \ a^k = \bar{a}^k, \ k \in J,$$

于是存在 $t \in I$，$r < q$，$r \notin I$，使 $\alpha_t > 0$. 在集合 $\{x \in R | a_i^T x = b_i, \ i \in (I \cup \{r\}) \backslash \{t\}\}$ 中包含了 $\{x^k\}$ 的无限多个元素 $\{x^k\}_{J_3}$，并且 $x^k \to x^+(k \to \infty, k \in J_3)$. 利用前面的证明可以得到

$$\nabla f(x^+) = \sum \beta_i a_i, \quad i \in (I \cup \{r\}) \backslash \{t\},$$

因此 $\alpha_t = 0$.

引理 4.2 设 $\{x^k\}$ 是由算法 II 产生的点列，$x^k \to x^+(k \to \infty)$，定理 3.2 的条件成立，且 H_k 一致正定，则对充分大的 k，有

$$a_i^T x^k = b_i, \quad i = 1, \cdots, q,$$
$$a_i^T x^k < b_i, \quad i = q + 1, \cdots, m.$$

引理 4.3 在引理 4.2 的假设下，$\varliminf_{k \to \infty} \bar{a}^k \geqslant a > 0$.

定理 4.2 设 $f(x)$ 在包含 R 的某开集上二阶连续可微，其 Hessen 矩阵满足：存在 Δ_1，Δ_2，$0 < \Delta_1 \leqslant \Delta_2$，对任意的 x，$y \in L_0$，

$$\Delta_1 \|y\|^2 \leqslant y^T G(x) y \leqslant \Delta_2 \|y\|^2.$$

对于 $x \in L_0$，存在 $l_0 > 0$，使

$$\|G(x)-G(x^*)\| \leqslant l_0\|x-x^*\|,$$

其中 x^* 为 (LNP) 的唯一整体最优解. $\{H_k\}$ 由 BFGS 方法修正，满足一致正定，在 x^* 处满足严格互补松弛条件，则由算法 II 产生的点列 $\{x^k\}$ 超线性收敛于 x^*.

证明　在定理的条件下，可以证明在充分大的 k 以后，算法 II 与某个无约束 BFGS 方法等价[2]，因而结论成立。

附　　录

一、退化情况的讨论

首先我们有

定理 5.1　当 R 有界时，假设 (H_1) 成立的充要条件是 R 的每个极点处的积极约束线性无关。

定理 5.2　令 $b_i(\varepsilon)=b_i-\varepsilon^{i-1}$, $i=1,\cdots,m$, 则存在 $\varepsilon_0>0$, 使对任意的 $\varepsilon\in(0,\varepsilon_0)$, 恒有

$$R(\varepsilon)=\{\alpha \mid Ax \leqslant b(\varepsilon)\}$$

非退化。

令 x^* 是 (LNP) 的最优解，$x(\varepsilon)$ 是问题

$$\min f(x), \quad x\in R(\varepsilon)$$

的最优解，则

$$|f(x(\varepsilon))-f(x^*)|<Ld(\varepsilon),$$

其中 L 为某一正常的。根据需要，取 ε 充分小，即可得到所要的精度。当 ε 很小时，会不会引起算法的不稳定，这个问题有待进一步研究。

二、无约束问题的一个算法

我们考虑如下非线性规划问题

(P) $\min f(x)$, $x\in R^n$.

取 $\{a^k\}$ 满足 $\lim\limits_{k\to\infty}\bar{a}^k=a>0$, 算法如下：

算法 III

1° 任取 $x^0\in R^n$, $k:=0$.

2° 计算 $g_k=\nabla f(x^k)$.

3° 若 $g_k=0$, 停，否则取 $P_k=Ag_k$, 其中 A 为一个正定矩阵。

4° 利用下述方式之一确定步长 a^k:

〈1〉 求 $a^k\in\left\{\bar{a}^k, \dfrac{\bar{a}^k}{2},\cdots\right\}$, 使 a^k 是满足下述不等式之最大正数

$$f(x^k-a^kp_k)-f(x^k)\leqslant -\frac{a^k}{2}g_k^Tp_k; \tag{5.1}$$

〈2〉 求 a^k 使 $a^k\in(0,\bar{a}^k)$,

$$f(x^k-a^kp_k)-f(x^k)\leqslant -\rho\lambda_k g_k^Tp_k, \tag{5.2}$$

$$g_{k+1}^T p_k \leqslant \sigma g_k^T p_k,$$

其中 $\rho \in \left(0, \frac{1}{2}\right)$, $\sigma \in (\rho, 1)$.

5° 令 $x^{k+1} = x^k - a^k p_k$, 置 $k := k+1$, 转 2°.

对于算法 III 可以证明

定理 5.3 设 $f(x) \in c^1$, $L_0 = \{x \mid f(x) \leqslant f(x^0) x \in R^n\}$ 有界,则算法 III 或在有限步终止于 (P) 的一个 K-T 点(即满足 $g(x) = 0$ 的点),或产生一无穷点列 $\{x^k\}$,其每个极限点都是 (P) 的 K-T 点.

定理 5.4 设 $f(x) \in c^2$, L_0 有界, $f(x)$ 在 L_0 上严格凸,且存在 $\Delta_2 \geqslant \Delta_1 > 0$,使对任意的 x, $y \in L_0$,

$$\Delta_1 \|y\|^2 \leqslant y^T G(x) y \leqslant \Delta_2 \|y\|^2,$$

其中 $G(x)$ 为 $f(x)$ 在 x 点的 Hessen 阵,则算法 III 产生的点列线性收敛于 (P)的唯一极小值点.

注 算法 III 仅仅是在理论上有意义,在计算实际问题时,我们只要恒取 $\bar{a}^k = a > 0$ 即可.

参 考 文 献

[1] K. Ritter, Convergence and Superlcnear Convergence of Algorithms for Linear Constrained Minimization Problems, In Nonlinear Optimization Theory and Algorithms (eds: L. C. W. Dixon et. al) 1980.
[2] 吴士泉,赵福安,不作精确线性搜索的算法及其收敛性质(摘要),曲阜师大学报,运筹学专刊(2),1985.
[3] 邓乃杨,无约束最优化计算方法,科学出版社,1982.
[4] P. F. Gill, W. Murry (eds), Numerical Methods for Constrained Optimization, 1974.
[5] 吴方,拟牛顿法讲义,中科院应用数学所运筹室印,1983.