中国科学

1982 年 11 月　　A 辑　　第 11 期

线性约束凸规划的既约变尺度法

赖炎连　吴　方　桂湘云

（中国科学院应用数学所，北京）

摘　　要

本文将既约梯度法与无约束最优化中的变尺度法相结合，给出了一个新的算法．在与文献 [3] 相同的假设下，证明了方法的收敛性．又在目标函数一致凸与其它假设下证明方法具有超线性速度．

一、引　　言

对于线性约束凸规划的求解问题，现在已有许多解法．但在现有的许多方法中，或者没有讨论收敛速度，或者收敛速度是线性的．在文献 [1] 中，对于包含线性等式与不等式约束的凸规划问题，我们结合变尺度与梯度投影法给出了一个解法，证明了方法的收敛性，并在目标函数一致凸与其它一些假设下，证明了算法产生的点列超线性收敛于最优解．对于线性等式与变量非负约束的凸规划问题，Wolfe 在文献 [2] 中首先提出了所谓的既约梯度法，通过降低维数以减少计算量．然而他并没有对方法给出满意的收敛性证明．在文献 [3] 中，越民义和韩继业提出了一个新的转轴方法，在此基础上，运用摄动的技巧，给出了另一个既约梯度法，并证明了方法的收敛性．其后，McCormick 在文献 [4] 中给出了另一个既约梯度法，方法的计算程序以及收敛性证明都比较简单，并且不用摄动的技巧．可是对约束条件的假设则比文献 [3] 强得多，它要求约束系数矩阵 A 的任何 m 阶子矩阵都是可逆的．另外，在文献 [3, 4] 中，都没有收敛速度的讨论．一般说来，这些方法的收敛速度都是线性的．在这里，我们对线性等式与变量非负约束的凸规划问题，在与文献 [3] 相同的关于线性约束的假设下，以文献 [2] 的转轴方法为基础，结合变尺度法，给出了一个既约变尺度法，证明了方法的收敛性；又在目标函数一致凸与其它一些假设下，证明了方法具有超线性的收敛速度．这里关于约束条件的假设远较文献 [4] 中为弱，而因无需摄动技巧，所以算法比较简单．

二、问题、假设及记号

我们讨论下面的非线性规划求解问题：

$$\min_{x\in R} f(x), \tag{2.1}$$

这里 $x=(x_1, x_2, \cdots, x_n)^T\in E^n$，可行集

$$R=\{x\mid Ax=b,\ x\geqslant 0\}, \tag{2.2}$$

本文 1980 年 7 月 31 日收到．

其中 A 为一秩 m 的 $m\times n$ 矩阵, $b=(b_1, b_2, \cdots b_m)^T$.

在本文中,恒假设

1°) 可行集 R 满足非退化假定,也即它的任何基本可行解都有 m 个非零分量,从而任何可行解都至少有 m 个正的分量.

2°) $f(x)$ 为一具有有界二阶连续偏导数的凸函数,并且水平集 $\{x|f(x)\leqslant f(x^1)\}$ 有界.

又为了证明超线性的收敛速度,在证明定理 4.4 的结论 2) 与证明定理 4.6 时,我们进一步假设

3°) 设 $f(x)$ 在 R 中为一致凸,从而保证了 $f(x)$ 在 R 中有唯一的最优解 x^*.

4°) $f(x)$的二阶偏导数满足 Lipschitz 条件

$$f_{ij}(x)-f_{ij}(x^*)=O(\|x-x^*\|), \tag{2.3}$$

5°) 在最优解 x^* 处,严格互补条件成立,也即我们有

$$\nabla f(x^*)=A^T\mu+\beta,\ \beta\geqslant 0. \tag{2.4}$$

这里 $\nabla f(x)$ 表示 $f(x)$ 在 x 处的梯度向量, $\mu\in E^m$, $\beta\in E^n$, 并且 $\beta_j(j=1, 2, \cdots n)$ 当 $x_j^*>0$ 时为零,而当 $x_j^*=0$ 时取正值.

在下文中,常用 I, J, Q, P 表示下标集,而用 A_I, A_J, $\cdots$表示由下标属于 I, J, $\cdots$的 A 的列向量组成的子矩阵. 若下标集 I 中恰有 m 个元素,并且 A_I 非异,则称 I 为一基,而命

$$J=\{1, 2, \cdots n\}\backslash I.$$

又我们常用 x, y, s, d, $\cdots$表示一些列向量, 而用 x_j, y_j, $\cdots$表示 x, y, $\cdots$的第 j 个分量,用 x_I, x_Q 等表示由下标属于 I 或 Q 的 x 的分量组成的子向量.

对于任何 $x\in R$, 与任意基 I, 则由 $Ax=b$, 可见

$$x_I=A_I^{-1}b-A_I^{-1}A_Jx_J, \tag{2.5}$$

也即基变量 x_I 的数值由非基变量 x_J 唯一确定.

今设 S 为一 n 维列向量,为了使 S 成为点 x 处的可行方向,由 $Ax=b$ 及 $A(x+\lambda s)=b$, $\lambda>0$ 可见必须

$$S_I=-A_I^{-1}A_JS_J, \tag{2.6}$$

也即 S_I 由 S_J 完全确定;又必须有 $\lambda_0>0$, 使

$$x_J+\lambda S_J\geqslant 0\quad (0<\lambda\leqslant\lambda_0). \tag{2.7}$$

这时若 $x_I=A_I^{-1}b-A_I^{-1}A_Jx_J>0$, 自然对于充分小的 $\lambda>0$, 都有

$$x_I+\lambda S_I=x_I-\lambda A_I^{-1}A_JS_J\geqslant 0, \tag{2.8}$$

于是 S 就是 x 处的一个可行方向. 而当 $x_I>0$ 不成立也即有 $i\in I$, 使 $x_i=0$ 时, 就不能保证(2.8)式的成立. 所以我们希望对于任何可行点 $x\in R$, 能够选取基 I, 使 $x_I>0$, 这时只要有 $\lambda_0>0$ 使 (2.7) 式成立, 那末 $S=\begin{pmatrix}-A_I^{-1}A_J\\ I\end{pmatrix}S_J$ 就成为 x 处的可行方向.

在 R 满足非退化假定的假设 1°) 下,文献 [3] 中给出了一种新的转轴运算,根据这种转轴运算,对于任何可行点 $x\in R$ 与任何基 I, 如果 $x_I>0$ 不成立,那末一定可以经过有限多次转轴运算得到一个新的基 I' 使 $x_{I'}>0$ 成立. 转轴运算的具体步骤如下: 设 x 为一可行解, I 为一基, $\varepsilon<1$ 为任一正数, $D\subset J=\{1, 2, \cdots n\}\backslash I$ 为任一下标集,命

$$E(x, \varepsilon)=\{j|x_j>\varepsilon, 1\leqslant j\leqslant n\}, \tag{2.9}$$

于是

1) 若秩 $(A_{E(x,\varepsilon)}) = m$，则到 3)，否则到 2).

2) 命 $\varepsilon := \varepsilon \times$ (x 的最小正分量)，回到 1).

3) 命 $x_r = \min\limits_{i\in I} x_i$. 若 $x_r > \varepsilon/2$，就命 $I' = I$，$\varepsilon' = \varepsilon$，$D' = D$，停止转轴，否则若 $x_r \leqslant \varepsilon/2$，到 4).

4) 因为 $r \in I$，所以由 (2.5) 式可有 $x_r = u - v^T x_J$，这里 u 为一数，v 为一个 $n-m$ 维列向量. 若有 $j \in J\backslash D$ 使 $x_j > \varepsilon/2$ 及 $v_j \neq 0$，就取

$$x_s = \max\{x_j \mid j \in J\backslash D,\ x_j > \varepsilon/2,\ v_j \neq 0\},$$

$I = I \cup \{S\}\backslash\{r\}$，$D = D \cup \{r\}$，回到 3)，否则到 5).

5) 取

$$x_s = \max\{x_j \mid j \in J,\ x_j > \varepsilon/2,\ v_j \neq 0\},$$

$I = I \cup \{s\}\backslash\{r\}$，$D = \{r\}$，回到 3).

对于上述的转轴运算，文献 [3] 中证明了下面的性质：

任取初始基 I_0，正数 $\varepsilon_0 < 1$，下标集 $D_0 \subset J_0$，又设 $\{x^k\}(k = 1, 2, \cdots)$ 为一列可行点. 在假设 1°) 下，对于 $k \geqslant 1$，由 x^k，I_{k-1}，ε_{k-1}，D_{k-1} 依上述转轴步骤，最后得到 I'_{k-1}，ε'_{k-1}，D'_{k-1}，命 $I_k = I'_{k-1}$，$\varepsilon_k = \varepsilon'_{k-1}$，$D_k = D'_{k-1}$，则有[1)]

$$x_i^k > \varepsilon_k/2 \quad (i \in I_k).$$

又若 $\{x^{k'}\}$ 为 $\{x^k\}$ 的一个收敛子序列，$x^{k'} \to x^*$ $(k' \to \infty)$，并且所有的 $I_{k'}$ 都等于同一个 I，则必有[2)]

$$x_i^* > 0 \quad (i \in I).$$

又若 $\{x^k\}$ 为一收敛的可行点列，则在假设 1°) 下，转轴次数有限，也即对于充分大的 k，都有[3)] $I_k = I_{k-1}$.

对于任何基 I，由 (2.5) 式可见：n 个变量的函数 $f(x) = f(x_I, x_J)$ 在 R 中实际上是 $n-m$ 个变量 $x_j (j \in J)$ 的函数 $\bar{f}(x_J)$，也即

$$\bar{f}(x_J) = f(A_I^{-1}b - A_I^{-1}A_J x_J,\ x_J). \tag{2.10}$$

因此 n 维的规划问题 (2.1) 变成了 $n-m$ 维的规划问题，

$$\min_{\substack{A_I^{-1}b - A_I^{-1}A_J x_J \geqslant 0 \\ x_J \geqslant 0}} \bar{f}(x_J). \tag{2.11}$$

用 $\nabla\bar{f}(x_J)$ 表 $\bar{f}(x_J)$ 在 x_J 处的梯度向量，则有

$$\nabla\bar{f}(x_J) = \nabla_J f(x) - (A_I^{-1}A_J)^T \nabla_I f(x), \tag{2.12}$$

$\nabla\bar{f}(x_J)$ 称为 $f(x)$ 的既约梯度，以下记

$$d(x_J) = -\nabla\bar{f}(x_J). \tag{2.13}$$

今设 $S = \begin{pmatrix} S_I \\ S_J \end{pmatrix}$ 为 x 处的一个可行方向，则因

$$\nabla f(x)^T S = -\nabla_I f(x)^T (A_I^{-1}A_J S_J) + \nabla_J f(x)^T S_J = -d(x_J)^T S_J, \tag{2.14}$$

---

1) 见文献 [3] Theorem 2.　2) 见文献 [3] Proposition 4.　3) 见文献 [3] Theorem 7.

所以当

$$d(x_J)^T S_J > 0 \tag{2.15}$$

时，可行方向 S 是一下降方向，又若 x' 为 $f(x)$ 在直线 $x' + \lambda S$ 上的一个局部最优点，则必

$$d(x'_J)^T S_J = -\nabla f(x')^T S = 0. \tag{2.16}$$

三、既约变尺度法

在这一节中，我们将以转轴运算作基础结合变尺度法与既约梯度法提出以下的既约变尺度算法，并作了说明. 既约变尺度法的具体步骤如下.

0) 任取初始可行点 x', 基 I_0, 正数 $\varepsilon_0 < 1$, 下标集 $D_0 \subset J_0$. 命 $k = 1$.

1) 对 x^k, I_{k-1}, ε_{k-1}, D_{k-1} 作上节中所述的转轴运算，最后得到 I_k, $\varepsilon_k > 0$, $D_k \subset J_k$, 并且有 $x_i^k > \varepsilon_k/2$ (对任何 $i \in I_k$). 命

$$Q_k = \{j \mid x_j^k = 0,\ 1 \leqslant j \leqslant n\}, \tag{3.1}$$

则必 $Q_k \subset J_k$. 又由 (2.12) 及 (2.13) 式确定负既约梯度

$$d^k = d(x^k_{J_k}) = -\nabla \bar{f}(x^k_{J_k}).$$

2) 定义 $n - m$ 维向量 y^k 如下：对于 $j \in J_k$, 命

$$y_j^k = \begin{cases} d_j^k, & 若\ x_j^k \leqslant d_j^k, \\ x_j^k d_j^k, & 若\ x_j^k > d_j^k. \end{cases} \tag{3.2}$$

若 $y^k = 0$, 由定理 4.1, x^k 即为凸规划问题 (2.1) 的最优解，计算停止；否则到 3).

3) 若 $k = 1$ 或 $k > 1$, 但 $I_k \neq I_{k-1}$, 或 $k > 1$, $I_k = I_{k-1}$, 且有 $j \in Q_k$, 使 $d_j^k > 0$, 则命

$$S^k_{J_k} = y^k, \tag{3.3}$$

而到 6). 若 $k > 1$, $I_k = I_{k-1}$, 而 $d^k_{Q_k} \leqslant 0$, 则到 4).

4) 若 $I_{k-1} \neq I_{k-2}$ 或 $I_{k-1} = I_{k-2}$, 但 $Q_k \neq Q_{k-1}$, 则命

$$H_k = P_{Q_k}, \tag{3.4}$$

而到 5). 这里 P_{Q_k} 是一个 $n - m$ 阶投影矩阵. 设 v 为一个 $n - m$ 维列向量，其分量的下标只属于 J_k, 则 $w = P_{Q_k} v$ 的分量 w_j $(j \in J_k)$ 满足

$$w_j = 0\ (j \in Q_k),\quad w_j = v_j\ (j \notin Q_k).$$

若 $I_{k-1} = I_{k-2}$, 而 $Q_k = Q_{k-1}$ (由引理 3.1 的后一半，$S^{k-1}_{J_{k-1}}$ 一定由 (3.7) 式所定义，所以 H_{k-1} 已有定义)，则命

$$H_k = H_{k-1} - \frac{H_{k-1}\Delta d^{k-1}(\Delta d^{k-1})^T H_{k-1}}{(\Delta d^{k-1})^T H_{k-1}\Delta d^{k-1}} - \frac{\Delta x^{k-1}_{J_k}(\Delta x^{k-1}_{J_k})^T}{(\Delta x^{k-1}_{J_k})^T \Delta d^{k-1}}, \tag{3.5}$$

这里

$$\Delta x^{k-1}_{J_k} = x^k_{J_k} - x^{k-1}_{J_{k-1}},\quad \Delta d^{k-1} = d^k - d^{k-1}, \tag{3.6}$$

而到 5).

5) 命

$$S^k_{J_k} = H_k d^k, \tag{3.7}$$

而到 6).

6) 命

$$S^k_{I_k} = -A^{-1}_{I_k} A_{J_k} S^k_{J_k},\quad S^k = \begin{pmatrix} S^k_{I_k} \\ S^k_{J_k} \end{pmatrix}. \tag{3.8}$$

由下面定理 4.2，S^k 为 x^k 的一个可行下降方向．用 x^{k+1} 表示 $f(x)$ 在直线段

$$x = x^k + \lambda S^k,\ \lambda > 0,\ x \in R \tag{3.9}$$

上的最小点，然后命 k 为 $k+1$，回到 1)．

以上是既约变尺度法的计算步骤，现在对它作些说明．

从算法可见：$S^k_{J_k}$ 由 (3.3) 或 (3.7) 式所定义，后者适用于 $k>1$，$I_k = I_{k-1}$，而 $d^k_{Q_k} \leqslant 0$ 的情形，而前者适用于其它情形；又 $S^k_{I_k}$ 则都由 (3.8) 式所定义．

引理 3.1. 若 $I_k = I_{k-1}$，且 $S^k_{J_k}$ 由(3.3)式定义，则必 $Q_{k+1} \not\supset Q_k$；故若 $I_k = I_{k-1}$，且 $Q_{k+1} \supset Q_k$，则 $S^k_{J_k}$ 一定由 (3.7) 式定义，所以一定有 $d^k_{Q_k} \leqslant 0$．

证．按照算法，对于 $I_k = I_{k-1}$ 的情形，当且仅当有 $j \in Q_k$，使 $d^k_j > 0$ 时，$S^k_{J_k}$ 才由 (3.3) 式所定义，对于此 j，仍由算法可见

$$S^k_j = y^k_j = d^k_j > 0,$$

因此

$$x^{k+1}_j = x^k_j + \lambda S^k_j > 0,$$

从而 $j \notin Q_{k+1}$，所以 $Q_{k+1} \not\supset Q_k$．引理 3.1 的后一半是前一半的自然推论．

引理 3.2. 若 $S^k_{J_k}$ 由 (3.7) 式所定义，则必

$$S^k_{Q_k} = 0, \tag{3.10}$$

从而 $Q_{k+1} \supset Q_k$．

证．用 v 表任何分量下标只属于 J_k 的 $n-m$ 向量，而命 $w^k = H_k v$，如能证明 $w^k_{Q_k} = 0$，(3.10) 式便成立，从而得到引理的结果．

今用 $l(\leqslant k)$ 表示这样一个正整数，即：$S^h_{J_h}$ $(l \leqslant h \leqslant k)$ 都由 (3.7) 式定义，且 $H_l = P_{Q_l}$，而 $H_h (l < h \leqslant k)$ 则都由 (3.5) 式所定义者．根据算法可有

$$I_l = I_{l+1} = \cdots = I_k,\ Q_l = Q_{l+1} = \cdots = Q_k. \tag{3.11}$$

因为 $J_l = J_{l+1} = \cdots = J_k$，所以由 P_{Q_l} 的定义可见 $w^k_{Q_k} = w^l_{Q_l} = 0$，又因 Q_l 的定义(3.1) 式及 $Q_{l+1} = Q_l$，所以 $x^{l+1}_{Q_{l+1}} = x^l_{Q_l} = 0$，从而

$$\Delta x^l_{Q_{l+1}} = x^{l+1}_{Q_{l+1}} - x^l_{Q_l} = 0,$$

因此由 (3.5) 式可见，由 $w^l_{Q_l} = 0$ 可以推出 $w^{l+1}_{Q_{l+1}} = 0$，同样，又由 $w^{l+1}_{Q_{l+1}} = 0$ 可以推出 $w^{l+2}_{Q_{l+2}} = 0$，$\cdots$，直至 $w^k_{Q_k} = 0$．引理证毕．

引理 3.3. 由 (3.4)，(3.5) 式所定义的 H_k 是一个 $n-m$ 阶半正定对称矩阵．命 $P_k = J_k \backslash Q_k$，而用 v 表任何分量下标只属于 J_k，并且 $v_{P_k} \neq 0$ 的 $n-m$ 维向量，则必

$$v^T H_k v > 0.$$

证．若 H_k 由 (3.4) 式所定义，则由投影矩阵的性质，可见 H_k 具有要求的性质；而若 H_k 由 (3.5) 定义，则 H_k 的有关性质可以通过对传统的证明稍加变化，而从 H_{k-1} 的对应性质推出，这里从略．

四、既约变尺度法的收敛性

本节将证明：在第二节的假设 1°)，2°) 下，由算法产生的点列 $\{x^k\}$ 的任何极限点都是问题(2.1)的最优点；而在假设 1°) —5°) 下，x^k 超线性收敛于问题 (2.1) 的唯一最优点．

定理 4.1. 设 $x \in R$，I 是一基，$x_I > 0$，则 x 是问题(2.1)的 K-T 点的充要条件，是依(3.2)

式所定义的 $n-m$ 维向量

$$y=0, \tag{4.1}$$

y 的分量 y_j 只当 $j\in J$ 时才有定义.

证. x 是 K-T 点的充要条件是有常向量 $\mu\in E^m$ 及 $\beta\in E^n$, 使

$$\nabla f(x)=A^T\mu+\beta,\ \beta^Tx=0,\ \beta\geqslant 0. \tag{4.2}$$

因为 $x_I>0$, 所以必须 $\beta_I=0$. 因此 (4.2) 式又等价于

$$\nabla_I f(x)=A_I^T\mu,\ \nabla_J f(x)=A_J^T\mu+\beta_J,$$
$$\beta_J^Tx_J=0,\ \beta_J\geqslant 0,$$

或即 ($\mu=(A_I^T)^{-1}\nabla_I f(x)$)

$$\beta_J=-d(x_J)=\nabla_J f(x)-(A_I^{-1}A_J)^T\nabla_I f(x)\geqslant 0,\ d(x_J)^Tx_J=0, \tag{4.3}$$

而这又等价于 (4.1) 式.

当 $f(x)$ 为凸函数时, x 是问题 (2.1) 的最小点的充要条件是 x 也是一个 K-T 点, 所以 (4.1) 式也是 x 是最小点的充要条件. 定理 4.1 证毕.

定理 4.2. 设 $y^k\neq 0$, 则由 (3.8) 式定义的 S^k 是 x^k 处的可行下降方向.

证. 因为 I_k 是经过转轴而得的基, 所以 $x_{I_k}^k>0$. 又由 (2.6) 式, $S_{I_k}^k=-A_{I_k}^{-1}A_{J_k}S_{J_k}^k$, 所以为了证明 S^k 的可行性, 只要证明: 存在 $\lambda_0>0$, 使对 $0<\lambda\leqslant\lambda_0$ 中的 λ 都有 (2.7) 式, 或即

$$x_{J_k}+\lambda S_{J_k}^k\geqslant 0\ \ (0<\lambda\leqslant\lambda_0) \tag{4.4}$$

成立. 为此又只须证明: 对于 $S_j^k<0$ 的 $j\in J_k$ 有 $x_j>0$; 又为了证明 S^k 的下降性, 只要证明 (2.15) 式或

$$(d^k)^TS_{J_k}^k>0 \tag{4.5}$$

成立.

按照算法, 若 $S_{J_k}^k$ 由 (3.3) 式定义, 由 (3.2) 式可见: 对于 J_k 中的 j, $S_j^k=y_j^k$ 只在 $x_j^k>0$ 时才能为负, 所以 S^k 为可行方向; 又

$$(d^k)^TS_{J_k}^k=\sum_{d_j^k\geqslant x_j^k}(d_j^k)^2+\sum_{d_j^k<x_j^k}x_j^k(d_j^k)^2\geqslant 0,$$

并且上式只在 $y^k=0$ 时为零, 所以对于这种情形, S^k 的确是 x^k 处的可行下降方向.

又若 $S_{J_k}^k$ 由 (3.7) 式定义, 由引理 3.2 可知, $S_{Q_k}^k=0$; 所以对于这种情形, 使 (4.4) 式成立的 λ_0 一定存在. 又因 $y^k\neq 0$, 且 $d_{Q_k}^k\leqslant 0$, 故必有 $j\notin Q_k$, 使 $d_j^k\neq 0$, 而由引理 3.3, 一定有

$$(d^k)^TS_{J_k}^k=(d^k)^TH_kd^k>0,$$

从而 S^k 也是 x^k 处的可行下降方向. 定理 4.2 证毕.

定理 4.3. 设 $\{x^{k'}\}$ 为 $\{x^k\}$ 的一个收敛子列, $x^{k'}\to x^*$; 又设对每一 k', $S_{J_{k'}}^{k'}$ 都由 (3.3) 式所定义, 则 x^* 必为凸规划问题 (2.1) 的最优解.

证. 因为只有有限多个可能的基, 故在 $\{x^{k'}\}$ 中必可选出一个子列, 使相应的 $I_{k'}$ 都相等, 因此不妨假设所有的 $I_{k'}$ 都等于同一个 I, 由第二节中所述的转轴性质可知

$$x_I^*>0.$$

今若 x^* 非最优解, 先证必有正数 $\delta>0$ 及自然数 K_1, 使对一切 $k'\geqslant K_1$ 都有

$$(d^{k'})^TS_J^{k'}=(d^{k'})^Ty^{k'}\geqslant\delta>0. \tag{4.6}$$

事实上, 用 ε 表示最小的非零 $|d_j^*|$ 与最小的非零 x_j^* 中的较小者, 不妨假设 $\varepsilon<1$. 因 x^* 非最

优解，由定理 4.1 可知，$y^*\neq 0$，所以或者存在 $j\in J$ 使 $d_j^*>0$ 及 $x_j^*=0$，或者存在 $j\in J$，使 $d_j^*\neq 0$ 及 $x_j^*>0$. 因为偏导数的连续性，所以 $d_j^{k'}\to d_j^*$，$x_j^{k'}\to x_j^*$. 故对前一种情形，当 k' 充分大时，$d_j^{k'}>\varepsilon/2>x_j^{k'}$，所以 $y_j^{k'}=d_j^{k'}$，而有

$$(d^{k'})^Ty^{k'}\geqslant(d_j^{k'})^2\geqslant\varepsilon^2/4;$$

又对后一种情形，当 k' 充分大时，$|d_j^{k'}|>\varepsilon/2$，$x_j^{k'}>\varepsilon/2$，所以无论 $y_j^{k'}=d_j^{k'}$ 或 $y_j^{k'}=x_j^{k'}d_j^{k'}$，都有

$$(d^{k'})^Ty^{k'}\geqslant\varepsilon^3/8.$$

故若取 $\delta=\varepsilon^3/8$，就有 (4.6) 式成立.

再证：存在与 k' 无关的 $\bar{\lambda}>0$ 及自然数 K_2，使对 $k'>K_2$ 与 $\lambda\in[0,\bar{\lambda}]$，恒有 $x^{k'}+\lambda S^{k'}\in R$.

因为 $S_J^{k'}=y^{k'}$，$S_I^{k'}=-A_I^{-1}A_JS_J^{k'}$，所以

$$x_j^{k'}+\lambda S_j^{k'}=\begin{cases}x_j^{k'}-\lambda v_j^TS_J^{k'}, & 若\ j\in I,\\ x_j^{k'}+\lambda d_j^{k'}, & 若\ j\in J\ 且\ x_j^{k'}\leqslant d_j^{k'},\\ x_j^{k'}+\lambda x_j^{k'}d_j^{k'}, & 若\ j\in J\ 且\ x_j^{k'}>d_j^{k'}.\end{cases}$$

式中 v_j^T 为 $A_I^{-1}A_J$ 的某一行向量. 今命

$$\bar{\lambda}_{k'}=\min\{\min_{j\in I}(x_j^{k'}/|v_j^TS_J^{k'}|),\ \min_{j\in J}1/|d_j^{k'}|\},$$

则当 $\lambda\in[0,\bar{\lambda}_{k'}]$ 时，$x^{k'}+\lambda S^{k'}\geqslant 0$，所以属于 R. 因为 v_j^T 与 k' 无关，$x_j^{k'}\to x_j^*$，$d_j^{k'}\to d_j^*$，所以 $|d_j^{k'}|$ 与 $|v_j^TS_{J_{k'}}|$ 都有界，因此 $\bar{\lambda}_{k'}$ 有正的下界，命之为 $\bar{\lambda}$.

最后，由 $\nabla f(x)$ 在 x^* 附近的一致连续性可得

$$\begin{aligned}f(x^{k'})-f(x^{k'+1})&\geqslant f(x^{k'})-f(x^{k'}+\bar{\lambda}S^{k'})\\&=-\bar{\lambda}\nabla f(x^{k'}+\theta\bar{\lambda}S^{k'})^TS^{k'}\geqslant-\bar{\lambda}\ \left\{\nabla f(x^{k'})^TS^{k'}+\frac{\delta}{2}\right\}\\&=\bar{\lambda}\left\{(d^{k'})^TS_J^{k'}-\frac{\delta}{2}\right\}\geqslant\frac{1}{2}\delta\bar{\lambda}>0.\end{aligned}\tag{4.7}$$

又因 $f(x^k)$ 为非增序列，所以由 (4.7) 式将会得出 $f(x^{k'})\to-\infty$，而与 $f(x^{k'})\to f(x^*)$ 矛盾，故 x^* 必为问题 (2.1) 的最优解.

推论 4.1. 若有无穷多个 k 使 $S_{J_k}^k$ 都由 (3.3) 式所定义，则 $\{x^k\}$ 的任何极限点一定是问题 (2.1) 的最优解.

事实上，由假设 2°)，$\{x^k\}$ 为一有界点列，因此必有 $S_{J_{k'}}^{k'}$ 都由 (3.3) 式所定义的收敛子列 $\{x^{k'}\}$. 由定理 4.3，$\{x^{k'}\}$ 的极限 x^* 是问题 (2.1) 的最优解. 但因 $f(x^k)$ 非增，所以 $\{x^k\}$ 的任何极限点都是问题 (2.1) 的最优解.

定理 4.4. 若从某一 k 开始，所有的 I_k 与 Q_k 都相等，则

1) 在假设 1°)，2°) 下，$f(x^k)$ 收敛于 $f(x)$ 在 R 上的最小值，x^k 的任何极限点都是 $f(x)$ 在 R 上的最优点，

2) 在假设 1°)—5°) 下，x^k 超线性收敛于 $f(x)$ 在 R 上的唯一最优点.

证. 假设对充分大的 k，$I_k=I$，$J_k=J$，$Q_k=Q$. 用 $|Q|$ 表 Q 中所含元素个数. 通过调换各个变量的次序，不妨假设

$$Q=\{1,2,\cdots,|Q|\},\ J=\{1,2,\cdots,n-m\},\\ I=\{n-m+1,n-m+2,\cdots,n\}. \tag{4.8}$$

设 $f(x)$ 的 Hessian 矩阵为 G, 则由 (2.10), (2.12) 式,易证 $\bar{f}(x_J)=\bar{f}(x_Q, x_{J\setminus Q})$ 的 Hessian 矩阵为:

$$\bar{G}=(I_{n-m},\ -(A_I^{-1}A_J)^T)\,G\begin{pmatrix} I_{n-m} \\ -A_I^{-1}A_J \end{pmatrix}, \tag{4.9}$$

这里 I_{n-m} 表 $n-m$ 阶单位矩阵.

现在证明: 若 $f(x)$ 是凸函数或一致凸函数, 则 $\bar{f}(x_J)$ 从而 $\bar{f}(0, x_{J\setminus Q})$ 也分别是 x_J 与 $x_{J\setminus Q}$ 的凸或一致凸函数. 事实上, 若 $f(x)$ 是凸或一致凸, 则有 $\varepsilon=0$ (若 $f(x)$为凸)或 $\varepsilon>0$ (若 $f(x)$为一致凸),使对任何 $x^T=(x_J^T, x_I^T)$ 都有

$$x^TGx\geqslant\varepsilon x^Tx\geqslant\varepsilon x_J^Tx_J.$$

于是由 (4.9) 式可见,

$$x_J^T\bar{G}x_J=(x_J^T,\ -(A_I^{-1}A_Jx_J)^T)G\begin{pmatrix} x_J \\ -A_I^{-1}A_Jx_J \end{pmatrix}\geqslant\varepsilon x_J^Tx_J.$$

因此 $\bar{f}(x_J)$ 从而 $\bar{f}(0, x_{J\setminus Q})$ 也是凸或一致凸函数.

再来考察第三节中的算法, (3.4) 式中的投影矩阵 P_Q 为:

$$P_Q=\begin{pmatrix} 0 & 0 \\ 0 & I_{n-m-|Q|} \end{pmatrix},$$

$I_{n-m-|Q|}$ 为 $n-m-|Q|$ 阶单位矩阵,又由 (3.5) 式定义的 H_k 一定形如

$$H_k=\begin{pmatrix} 0 & 0 \\ 0 & D_k \end{pmatrix},$$

其中 D_k 是一个 $n-m-|Q|$ 阶的对称正定矩阵,它由

$$D_k=D_{k-1}-\frac{D_{k-1}\Delta d_{J\setminus Q}^{k-1}(\Delta d_{J\setminus Q}^{k-1})^TD_{k-1}}{(\Delta d_{J\setminus Q}^{k-1})^TD_{k-1}(\Delta d_{J\setminus Q}^{k-1})}-\frac{\Delta x_{J\setminus Q}^{k-1}(\Delta x_{J\setminus Q}^{k-1})^T}{(\Delta x_{J\setminus Q}^{k-1})^T\Delta d_{J\setminus Q}^{k-1}}$$

定义,这里

$$\Delta x_{J\setminus Q}^{k-1}=x_{J\setminus Q}^k-x_{J\setminus Q}^{k-1},\ \Delta d_{J\setminus Q}^{k-1}=d_{J\setminus Q}^k-d_{J\setminus Q}^{k-1}.$$

由 (3.7) 式可见, $S_Q^k=0$, 所以 $x_Q^{k+1}=0$. x^{k+1} 是 $f(x)$ 在直线段 (3.9) 式上的最小点. 因为 $I_{k+1}=I_k$, $Q_{k+1}=Q_k$, 所以 $x\in R$ 这一限制在 (3.9) 式中不起作用,也就是说 x^{k+1} 也是 $f(x)$ 在半直线

$$x=x^k+\lambda S^k,\ \lambda>0$$

上的最小点,所以 x^{k+1} 由

$$\nabla f(x^{k+1})^TS^k=0$$

所确定, 但 (2.14) 式也等价于

$$(d^{k+1})^TS_J^k=0\ \text{或}\ (d_{J\setminus Q}^{k+1})^TS_{J\setminus Q}^k=0.$$

而因 $\bar{f}(0, x_{J\setminus Q})$ 是凸函数,所以 $x_{J\setminus Q}^{k+1}$ 也是 $\bar{f}(0, x_{J\setminus Q})$ 在半直线

$$x_{J\setminus Q}=x_{J\setminus Q}^k+\lambda S_{J\setminus Q}^k,\ \lambda>0$$

上的最小点. 正因为如此,在 I_k, Q_k, 不变的假设下将第三节中的算法应用于问题 (2.1) 产生的点列 x^k, 其相应的 $x_{J\setminus Q}^k$ 就是将 Davidon-Fletcher-Powell 方法应用于无约束问题

$$\min \bar{f}(0,\ x_{J\setminus Q})$$

所产生的点列. 反之，有了 $x^k_{J\setminus Q}$，配上 $x^k_Q = 0$，然后由 (2.5) 式定出 x^k_I，从而 x^k 也就完全确定.

在文献 [5] 中的定理 1 曾证明：若 $F(x)$ 是 x 的凸函数，并且具有连续的二阶偏导数，且水平集 $\{x \mid F(x) \leqslant F(x^1)\}$ 为有界，则将 DFP 算法应用于 $F(x)$ 所得的点列 $\{x^k\}$，相应的 $F(x^k)$ 一定收敛于 $F(x)$ 的最小值.

由此立刻推知 $\{x^k\}$ 的任何极限点都是 $F(x)$ 的最小点. 于是在 1°)，2°) 的假设下，由文献 [5] 的定理 1 可知，$x^k_{J\setminus Q}$ 的任何极限点 $x^*_{J\setminus Q}$ 都是 $\bar{f}(0,\ x_{J\setminus Q})$ 的最小点. 因此 $d^*_{J\setminus Q} = 0$；又因 $x^k_Q = 0$，并由算法知，$d^k_Q \leqslant 0$，所以由连续性得 $x^*_Q = 0$ 及 $d^*_Q \leqslant 0$. 于是由 (3.2) 式可知 $y^* = 0$. 再由定理 4.1 可知，由 x^*_J 决定的 x^*，也即 x^k 的任何极限点 x^* 都是 $f(x)$ 在 R 上的最小点.

在文献 [6] 的定理 4 中曾证明：若 $F(x)$ 是 x 的一致凸函数，且具有连续的二阶偏导数，并且有 $L > 0$，使

$$\left|\frac{\partial^2 F(x)}{\partial x_i \partial x_j} - \frac{\partial^2 F(x^*)}{\partial x_i \partial x_j}\right| \leqslant L\|x - x^*\|$$

成立，则由 DFP 方法得到的点列 $\{x^k\}$ 超线性收敛于 $F(x)$ 的唯一最优点 x^*.

于是，在 1°)—5°) 的假设下，应用上述定理可知 $x^k_{J\setminus Q}$ 超线性收敛于 $\bar{f}(0,\ x_{J\setminus Q})$ 的唯一最小点 $x^*_{J\setminus Q}$，从而 x^k 超线性收敛于 $f(x)$ 在 R 上的唯一最小点. 定理 4.4 证毕.

参　考　文　献

[1] Kwei Hsiang-yuin, Wu Fang & Lai Yan-nian, *O. R.* 78 (Ed. Haley, K. B.), Noth-Holland Publishing Company, 1978, 955—974.

[2] Wolfe, P., In *Recent Advances in Mathematical Programming* (Ed. Grares, R. L. & Wolfe, P.), 1963.

[3] Yue Ming-yi & Han Ji-ye, *Scientia Sinica*, **XXII** (1977), 10: 1099—1113.

[4] McCormick G. P., *Management Science Theory*, **10**(1970 a), 146—160.

[5] Powell, M. J. D., In *Numerical Methods for non-linear Optimization* (Ed. Lootsma, F. A.), 1971.

[6] ——————, *J Inst. Math. Applic.*, **7**(1971), 21—36.