非线性等式约束优化的信赖域算法

TRUST REGION METHODS FOR NONLINEAR EQUALITY CONSTRAINED OPTIMIZATION

唐 健* 段虞荣
Tang Jian Duan Yurong
（重庆建筑工程学院） （重庆大学）

摘 要 提出了求解一般非线性等式约束优化的信赖域算法。运用了不同方法在信赖域内求解原优化问题的二次近似模型的解，通过收敛性分析，获得了算法的整体及局部超线性收敛等结果，并给出了算法的执行细节。

关键词 数学规划；最优化；二次规划/信赖域算法

中国图书资料分类法分类号 O221.2

ABSTRACT This paper presents two trust region algorithms for solving general nonlinear equality constrained optimization problems. The different techniques are used in minimizing a quadrdtic approximate model of the original optimization problems in trust region. By the convergence analysis, the global convergences and local superlinear convergences of the algorithms are obtained. Finally, the implementations of the algorithms are given.

KEY WORDS mathematical programming; optimization; quadratic programming/trust region method.

0 引 言

在过去几年里，求解无约束优化问题的信赖域算法被证明是有效的方法。人们自然想将其推广到约束优化中去。在国内外不少综述文章中，信赖域算法被认为是将会成为非线性规划中有效而富有竞争力的一种方法。随着人们对序贯二次规划（sequential quadratic programming）的研究，信赖域方法也会渐渐地成熟起来。

在约束优化中，信赖域算法的研究直到近几年才有些进展，不过迄今为止发表的论文很少，理论和算法均不成熟。可以说把信赖域方法由无约束推广到约束有不少困难，研究工作尚待深入。

非线性规划的信赖域算法研究已成为当前研究工作比较活跃的方向之一。对约束优化及不可微优化问题的信赖域算法值得进一步研究，需要对相应子模型问题构造有效的下降条件和发展新方法，有关评述，可参见[5,6]。

本文采用不同方法，提出了两个关于等式约束优化的信赖域算法，并证明了其整体及局部

* 收文日期 1989－09－06
*原系我校硕士研究生

超线性收敛定理，给出了算法实施过程的细节。

1 算法一

考虑等式约束优化问题：

$$\min f(x), \qquad s.t. \qquad h(x)=0, \tag{1}$$

其中 $x\in R^n$， $f:R^n\to R, h:R^n\to R^m$， $f\in c^2$， $h\in C^1$， 且 $\nabla h(x)$ 列线性无关，并假定 $m\leqslant n$。

考虑问题(1)的近似模型

$$\begin{cases} \min\{f(x_k)+g_k^Td+\frac{1}{2}d^TB_kd\} \\ s.t.\quad h(x_k)+\nabla h(x_k)^Td=0 \\ \qquad \|d\|\leqslant\triangle_k \end{cases} \tag{2}$$

其中 x_k 为每次迭代的近似解，$g_k=\nabla f(x_k)$，B_k 为 f 的近似 Hesse 阵，记 $f_k=f(x_k)$，其它符号均采用数学规划中的标准记号，$\|\cdot\|$为某种范数。

令 $\Psi_k(x)=g_k^Tx+\frac{1}{2}x^TB_kx$

我们用信赖域方法求解此子问题，算法如下

1. 给定 $x_0, B_0^{-1}, \underline{\triangle}\leqslant\triangle_0\leqslant\overline{\triangle}, 0<r_3<r_2<r<1<r_1$。

其中$\overline{\triangle}$、$\underline{\triangle}$分别为预先选定的信赖域半径$\triangle_k$的上下界，$\underline{\triangle}\leqslant\triangle_k\leqslant\overline{\triangle}$，根据计算情况可调节。

2. 令 $k=0$，计算 $f_0, g_0, h_0, \nabla h_0$的值。
3. (i)用最小范数法求解二次规划

$$\begin{cases} \min\{f_k+g_k^Td+\frac{1}{2}d^TB_kd\} \\ s.t.\qquad h_k+\nabla h_k^Td=0, \end{cases} \tag{3}$$

(ii)如 $d_k=0$，停止执行。

(iii)确定 $\tau_k=\sigma^j (j=0,1,2,\cdots\cdots)$为使下式成立的最大正数

$$\begin{cases} \|\tau_kd_k\|\leqslant\triangle_k, \\ f(x_k+\tau_kd_k)\leqslant f_k+\gamma\tau_kg_k^Td_k. \end{cases}$$

4. 令 $x_{k+1}=x_k+\tau_kd_k$，计算 $f_{k+1}, g_{k+1}, h_{k+1}, \nabla h_{k+1}, p_k=\tau_kd_k$ 的值。
5. 修正 B_k^{-1} 得到 B_{k+1}^{-1}，计算

$$\rho_k=\frac{f_{k+1}-f_k}{\psi_k(p_k)},$$

$$\triangle_{k+1}=\begin{cases} \min(\overline{\triangle}, r_1\triangle_k), \text{如 } \rho_k\geqslant r_3 \\ \max(\underline{\triangle}, r_2\triangle_k) \text{否则} \end{cases}$$

6. 令 $k=k+1$，转3。

算法5中 B_k 的修正见第3部分。

下面我们给出本算法的收敛性分析。

引理1： 设 B,D 均为 Banach 空间，$T:B\to D$ 是有界线性算子，$u_0\in B$ 为给定向量，

$$J(u) = \|u - u_0\|。$$

则$\forall$ $b\in T(B)$，$\exists|$ $u_b\in B$，使得 $T(u_b)=b$，且满足 $J(u_b)=\min_{u\in B}\|u-u_0\|$。

此引理作为算子投影定理的特例，见于众多教科书中。

引理2. 设 $\Omega\subset R^n$ 为闭凸集，$x_0\notin\Omega$。则存在 $a>0$，及向量 g_0，$\|g_0\|=1$，使得 $\langle x_0-x,-g_0\rangle\leqslant -a$，$\forall$ $x\in\Omega$。

这里 $\langle\cdot,\cdot\rangle$ 为 R^n 内的内积，$a=\|x_0-z_0\|$，$g_0=\frac{1}{a}(x_0-z_0)$，其中 $z_0\in\Omega$，并使得 $\|z_0-x_0\|=\min_{z\in\Omega}\|z-x_0\|$。

推论1： 在引理2假设下，$\forall$ $y\in\Omega$，有 $\langle z_0-y,z_0-x_0\rangle\leqslant 0$。

我们可将上述结论中的欧氏内积、范数分别改为矩阵内积和范数 $\langle\cdot,\cdot\rangle_B$，$\|\cdot\|_B$，这里 B 为正定对称阵，结论同样成立。$\langle x,y\rangle_B=x^TBy$，$\|x\|_B^2=x^TBx$，$\forall$ $x,y\in R^n$。

把规划(3)改写为

$$\begin{cases}\min \frac{1}{2}\|d + B_k^{-1}g_k\|_{B_k}^2 + (f_k - \frac{1}{2}g_k^TB_k^{-1}g_k) \\ \text{s.t.} \quad h_k + \nabla h_k^Td = 0\end{cases} \tag{4}$$

设 $R_k=\{d\,|\,h_k+\nabla h_k^Td=0，\|d\|\leqslant\overline{\triangle}\}$，适当选取$\overline{\triangle}$，可保证 R_k 非空，显然 R_k 为有界闭凸集。由引理1可知，问题(4)存在唯一解，其等价问题为

$$d_k = \bar{x}_{k+1} - x_k = \operatorname{argmin}\{\frac{1}{2}\|d + B_k^{-1}g_k\|_{B_k}^2, d \in R_k\}$$

由引理2推出

推论2： 对任何对称正定阵 B_k 和 $y\in R_k$，有

$$\langle\bar{x}_{k+1} - y, \bar{x}_{k+1} - x_k + B_k^{-1}g_k\rangle \leqslant 0$$

定理1： 设算法一中 B_k 正定对称，则 $\langle g_k,d_k\rangle\leqslant -\|d_k\|_{B_k}^2$，从而 d_k 为 f 的下降方向。

证：因 $\|d_k\|_{B_k}^2=\langle\bar{x}_{k+1}-x_k+B_k^{-1}g_k,d_k\rangle_{B_k}-\langle g_k,d_k\rangle$

由推论1

$$\|d_k\|_{B_k}^2 \leqslant -\langle g_k,d_k\rangle, \qquad \forall\ k$$

$$\langle g_k,d_k\rangle \leqslant -\|d_k\|_{B_k}^2$$

证毕

定理2： 对算法一有 $\psi_k(p_k)\leqslant -\frac{\tau_k}{2}\|d\|_{B_k}^2$

证： 由推论2及定理1，

因 $x_{k+1}=x_k+\tau_kd_k(\varepsilon\leqslant\tau_k\leqslant 1)$

$p_k=\tau_kd_k=x_{k+1}-x_k$

∴ $\psi_k(p_k)=\psi_k(\tau_kd_k)$

$$=\tau_k g_k^T d_k+\frac{1}{2}\tau_k^2 d_k^T B_k d_k$$

$$\leqslant \tau_k(g_k^T d_k+\frac{1}{2}d_k^T B_k d_k)$$

$$\leqslant \tau_k(-\|d_k\|_{B_k}^2+\frac{1}{2}d_k^T B_k d_k)$$

$$=-\frac{\tau_k}{2}\|d_k\|_{B_k}^2$$

证毕

在算法中我们采用 Boggs 和 Tolle 提出的 DFP 方法修正 B_k^{-1}，所以$\{B_k\}$总是对称正定的。下面我们直接给出整体收敛定理。

定理3：设 $Z_0=\{x\mid f(x)\leqslant f(x_0)\}$有界，$\|B_k\|\geqslant\frac{1}{v}$，$v>0$为常数，且问题(1)在 Z_0内 K—T 点有限。则由算法导出的$\{x_k\}$收敛于问题(1)的 K—T 点。

证：由算法和定理1

$$f_{k+1}\leqslant f_k+r\tau_k g_k^T d_k\leqslant f_k\leqslant\cdots\cdots\leqslant f(x_0)$$

所以　$x_{k+1}\in z_0$，$\forall k$　有$\{x_k\}\subset z_0$。

由于$\{f_k\}$单调下降，又 f 在 z_0有下界，故而

$$\lim_{k\to\infty} f_k=\bar{f}$$

$$\therefore \lim_{k\to\infty}\tau_k g_k^T d_k=0$$

由[5]引理1得

$$\lim_{k\to\infty}\|d_k\|_{B_k}^2=0$$

又　$d_k^T B_k d_k\geqslant\frac{1}{v}\|d_k\|^2$

$\therefore$　$\lim_{k\to\infty}\|d_k\|=0$

因$\{x_k\}\subset z_0$(紧集)，故至少有一聚点 x^*，设子列$\{x_{k_j}\}\subset\{x_k\}$收敛于 x^*。

$$\therefore\quad h_k+\nabla h_k^T d_k=0,\qquad h(x^*)=0$$

又

$$f(x^*)=\lim_{k_j\to\infty} f(x_{k_j})=\lim_{k\to\infty} f_k=\bar{f}$$

因 f_k 单调下降，x^*不可能是 f 的局部极大值点。设 $\bar{u}_k$ 为规划(3)的 Lagrange 乘子向量，则

$$g_k+B_k d_k=\nabla h_k\bar{u}_k$$

故　$$g(x^*)=\nabla h(x^*)\bar{u}^*$$

即$\{x_k\}$的聚点为问题(1)的 K—T 点，由假设，$\{x_k\}$的聚点有限，类似于文[5]定理证明，可推出 x^*唯一，即 $\lim_{k\to\infty}x_k=x^*$。　证毕

在一定条件下我们可以证明算法具有超线性收敛性。设 $H(x_k)=\nabla^2 f(x_k)$。

定理4： 设 $f\in C^2$，$H(x)$为 f 的 Hesse 阵，$\{x_k\}$收敛于 x^*。B_k 为$\nabla^2 f(x_k)$的好近似，$\|B_k\|\geqslant\frac{1}{v}$ 如果 $\lim\limits_{k\to\infty}\|(B_k-H(x^*))d_k\|=0$。则 $\lim\limits_{k\to\infty}\tau_k=1$。

证： 由中值定理

$$f_{k+1}-f_k=\langle g_k,x_{k+1}-x_k\rangle+\frac{1}{2}\langle x_{k+1}-x_k,B_k(x_{k+1}-x_k)\rangle$$
$$+\int_0^1(1-t)\langle x_{k+1}-x_k,(H(x_k+t(x_{k+1}-x_k))-B_k))(x_{k+1}-x_k)\rangle dt$$

记 $u_k=\int_0^1(1-t)\|H(x^*)-H(x_k+t(x_{k+1}-x_k))\|dt$

由定理1

$$f_{k+1}-f_k\leqslant\tau_k g_k^Td_k-\frac{1}{2}\tau_k^2g_k^Td_k+u_k\tau_k^2\|d_k\|^2+\frac{1}{2}\|(B_k-H(x^*))(x_{k+1}-x_k)\|$$
$$/\|x_{k+1}-x_k\|$$

又 $x^TB_kx\geqslant\frac{1}{v}\|x\|^2 \qquad \forall\, x\in R^n$

故 $f_{k+1}-f_k\leqslant\tau_kg_k^Td_k[1-\frac{\tau_k}{2}-\frac{\tau_k}{v}(u_k+\|(B_k-H(x^*))d_k\|/2\|d_k\|)]$

算法一3中 r 决定了 τ_k，使$\|\tau_kd_k\|\leqslant\triangle_k$，

且
$$r\leqslant[1-\frac{\tau_k}{2}-\frac{\tau_k}{v}(u_k+\|(B_k-H(x^*))d_k\|/2\|d_k\|)]\leqslant\frac{1}{2} \tag{5}$$

如 τ_k 使不等式(5)成立，那么 τ_k 也满足步骤3中的不等式

因 $\lim\limits_{k\to\infty}x_k=x^*$，故 $u_k\searrow0(k\to\infty$时)，由假设 $\lim\limits_{k\to\infty}\|(B_k-H(x^*))d_k\|/\|d_k\|=0$，所以存在 k_0，$\forall\, k\geqslant k_0$ 有

$$\frac{1}{v}(u_k+\|(B_k-H(x^*))d_k\|/2\|d_k\|)\leqslant\frac{1}{2}-r$$

由定理3知 $\lim\limits_{k\to\infty}d_k=0$

$$\therefore\quad \|d_k\|\leqslant\overline{\triangle}\leqslant\triangle_k$$

即$\forall\, k\geqslant k_0$时， $\tau_k\equiv1$ 证毕

容易推得

定理5： 设 $\lim\limits_{k\to\infty}x_k=x^*$，$\nabla h(x^*)$列满秩，约束优化的二阶充分性条件成立。令 P_k 是 R^n 到子空间 $w_k=\{s|s^T\nabla h_k=0\}$的投影阵，且 $\lim\limits_{k\to\infty}\|P_k(B_k-H(x^*)d_k)\|//\|d_k\|=0$。则算法有 Q 超线性收敛性，从而避免了 Maratos 效应发生。

算法一的具体执行细节在第3部分讨论。

2 算法二

利用 Fletcher 可微精确罚函数作步长搜索，建立一种新的信赖域算法。

考虑等式约束问题(1)，第二部分中相应假定成立，不再赘述。

设 $L(x,c)=f(x)-\lambda^T h(x)+\frac{c}{2}\|h(x)\|^2$ (6)

其中 $\lambda(x)=(\nabla h^T(x)\nabla h(x))^{-1}\nabla h^T(x)f(x)$，其导数 $\nabla\lambda(x)$ 我们采用差商逼近。即有

$$\nabla_x L(x,c)^T d = g_k^T d - \frac{1}{\tau_k}(\lambda_{k+1}-\lambda_k)^T h_k - \lambda_k^T \nabla h_k^T d + c h_k^T \nabla h_k^T d \tag{7}$$

这样可避免计算 f 和 h 的二阶导数。算法二如下：

1^0. 给定 $x_1, B_1^{-1}, c_0, u, \overline{\overline{\triangle}} \leqslant \triangle_1 \leqslant \overline{\triangle}, 0<r_3<r_2<r<1<r_1, k=1$

2^0. 解
$$\begin{cases}\min\{f_k+g_k^T d+\frac{1}{2}d^T B_k d\} \\ \text{s.t.}\quad h_k+\nabla h_k^T d=0\end{cases} \tag{8}$$

如果 $d_k=0$，停止；否则，如果 $\|d_k\|>\triangle_k$，取 $d_k=\frac{\triangle_k}{\|d_k\|}d_k$；

3^0. 取 $c_k=c_{k-1}$，考察下列不等式是否成立

$$\nabla_x L_k^T d_k \leqslant \frac{1}{2}(d_k^T B_k d_k + c_k\|h_k\|^2) \tag{9}$$

如否，取 $c_k=2c_{k-1}$

4^0. 取 $\tau_k=\sigma^j(j=0,1,2,\cdots\cdots)$ 为使下式成立的最大正数

$$L(x_k+\tau_k d_k, c_k) \leqslant L(x_k, c_k) + u\tau_k \nabla_x L^T(x_k,c_k) d_k \tag{10}$$

5^0. $x_{k+1}=x_k+\tau_k d_k$，计算 $f_{k+1}, g_{k+1}, h_{k+1}, \nabla h_{k+1}, \lambda_{k+1}$；

6^0. 修正 B_k^{-1}，得 B_{k+1}^{-1}。

计算 $\rho_k=\frac{f_{k+1}-f_k}{\Psi_k(\tau_k d_k)}$， $\triangle_{k+1}=\begin{cases}\min(\overline{\triangle}, r_1\triangle_k)，如\ \rho_k\geqslant r_3 \\ \max(\overline{\overline{\triangle}}, r_2\triangle_k)，否则\end{cases}$

7^0. $k=k+1$，转 2^0。

算法二中二次规划(8)的求解采用 Wolfe 法，而 B_k^{-1} 的修正仍采用正定的 DFP 校正形式。

在本节里，假定 $\{x_k\}$，$\{B_k\}$ 有界，$\nabla h(x_k)$ 列满秩，对满足 $\nabla h_k^T x=0, x\in R^n$ 的 x 有

$$x^T B_k x \geqslant \frac{1}{v}\|x\|^2, v>0$$

为一常数。下面我们进行收敛性分析。

定理1. 设(9)式成立。则算法二导出的序列满足 $\nabla_x L_k^T d_k \leqslant -\eta\|d_k\|^2, \eta>0$ 为常数。

证： 由于 $\{B_k\}$ 有界，存在 $\varepsilon>0$，使得

$$(2\varepsilon+\varepsilon^2)\|B_k\| \leqslant \frac{1}{2v} \qquad \forall\ k$$

将 d_k 分解成 $d_k=e_k+E_k, e_k\in R(\nabla h_k), E_k\in N(\nabla h_k)$。因此 e_k 为满足二次规划(8)之等式约束而模最小的 d，即 $e_k=(\nabla h_k^T)^+(-h_k)$，$(\nabla h_k^T)^+$ 为 Moore—Penrose 广义逆。因 $h\in C^2$，存在 $M_0>0$，使得

$$\|e_k\| \leqslant M_0\|h_k\|, \forall\ k。$$

因为$\{x_k\}$有界，因此$\{(\nabla h_k^T)^+\}$有界。

又$\nabla h_k^T E_k=0$，有 $E_k^T B_k E_k \geq \frac{1}{v}\|E_k\|^2$

如果 $\|e_k\| \geq \varepsilon \|E_k\|$，则

$$\|h_k\| \geq \frac{1}{M_0}\|e_k\| \geq \frac{\|e_k\|+\|E_k\|}{M_0(1+\varepsilon^{-1})} \geq \frac{\|d_k\|}{M_0(1+\varepsilon^{-1})} \tag{11}$$

如果$\|e_k\|<\varepsilon\|E_k\|$则

$$\begin{aligned} d_k^T B_k d_k &= (e_k+E_k)^T B_k (e_k+E_k) \\ &\geq \|E_k\|^2\left(\frac{1}{v}-2\varepsilon\|B_k\|-\varepsilon^2\|B_k\|^2\right) \\ &\geq \frac{1}{2v}\|E_k\|^2 \\ &> \frac{(\|e_k\|+\|E_k\|)^2}{2v(1+\varepsilon)^2} \geq \frac{\|d_k\|^2}{2v(1+\varepsilon)^2} \end{aligned} \tag{12}$$

利用不等式(9)及$\{B_k\}$正定，

$$\nabla_x L_k^T d_k \leq -\frac{1}{2}(d_k^T B_k d_k + c_k\|h_k\|^2) \leq -\frac{c_k}{4}\|h_k\|^2$$

利用此不等式首尾两项及不等式(11)，或不等式(9)及(12)均可推出：

$$\nabla_x L_k^T d_k \leq -\eta\|d_k\|^2,$$

这里 $\eta=\frac{1}{4}\min\left[\frac{c_0}{M_0(1+\varepsilon^{-1})^2},\frac{1}{v(1+\varepsilon)^2}\right]$ 证毕。

定理1说明，只要(9)式成立，d_k就是L_k的下降方向，以下定理说明(9)式是成立的。

定理2： 存在K_0，$\forall\ k\geq K_0$，当取$C_k=C_{k_0}=\bar{C}$时，(9)式恒成立。

证： 因 $h_k+\nabla h_k^T d_k=0$，

$$\nabla_x L_k^T d_k = g_k^T d_k - \frac{1}{\tau}(\lambda_{k+1}-\lambda_k)^T h_k - \lambda_k^T \nabla h_k^T d_k - C_k\|h_k\|^2$$

设$\bar{u}_k$是(8)式的 Lagrange 乘子，则

$$g_k + B_k d_k = \nabla h_k \bar{u}_k \tag{13}$$

又 $$\begin{aligned}\|g_k^T d_k-\lambda_k^T\nabla h_k^T d_k+d_k^T B_k d_k\| &= \|(\bar{u}_k-\lambda_k)^T\nabla h_k^T d_k\| \\ &= \|(\lambda_k-\bar{u}_k)^T h_k\|\end{aligned}$$

因 $\nabla h_k\lambda_k=\nabla h_k(\nabla h_k^T\nabla h_k)^{-1}\nabla h_k^T g_k$，而$\nabla h_k(\nabla h_k^T\nabla h_k)^{-1}\nabla h_k^T$为$R^n$到$R(\nabla h_k)$的正投影矩阵，因而

$$\|\nabla h_k\lambda_k - g_k\| \leq \|\nabla h_k\bar{u}_k - g_k\| = \|B_k d_k\|$$

又∇h_k列满秩。

$$\begin{aligned}\|\lambda_k-\bar{u}_k\| &= O(\|\nabla h_k(\lambda_k-\bar{u}_k)\|)\\ &\leqslant O(\|\nabla h_k\lambda_k-g_k\|+\|\nabla h_k\bar{u}_k-g_k\|)\\ &= O(\|B_k\|\|d_k\|)\end{aligned}$$

故　$\|g_k^T d_k-\lambda_k^T\nabla h_k^T d_k+d_k^T B_k d_k\|\leqslant\|\bar{u}_k-\lambda_k\|\|h_k\|=O(\|h_k\|\|d_k\|)$

因　$f,h\in C^2$,

$$\left\|\frac{1}{\tau_k}(\lambda_{k+1}-\lambda_k)^T h_k\right\|=O(\|h_k\|\|d_k\|)$$

要选择 C_{k_0}，使(9)式成立，即要下式成立

$$-\frac{1}{2}d_k^T B_k d_k+M_1\|d_k\|\|h_k\|\leqslant\frac{1}{2}C_{k_0}\|h_k\|^2 \tag{14}$$

像证明定理1一样将 d_k 分解，(14)式要成立，

只须　$-\frac{1}{2v}\|E_k\|^2+M_1M_0\|h_k\|^2+M_2\|E_k\|\|h_k\|\leqslant\frac{1}{2}C_{k_0}\|h_k\|^2$,

即　$-\frac{1}{v}\left(\frac{\|E_k\|}{\|h_k\|}\right)^2+2M_2\left(\frac{\|E_k\|}{\|h_k\|}\right)+2M_1M_0\leqslant C_{k_0}$

因上式左端为$\frac{\|E_k\|}{\|h_k\|}$的二次多项式，有最大值 $vM_2^2+M_1M_0$。故只要取 $\sigma=vM_2^2+M_1M_0$，由算法，存在 K_0，使 $C_{k_0-1}<\sigma\leqslant C_{k_0}$，此 C_{k_0} 及 K_0 即为所求，$\forall\, k\geqslant K_0$，取 $C_k\equiv C_{k_0-1}=\bar{C}$ 时，(9)式恒成立。证毕。

定理3：$\lim\limits_{k\to\infty}\|d_k\|=0$

证：由定理2，存在 K_0，$\forall\, k\geqslant K_0$，$C_k\equiv C_{k_0}=\bar{C}$,

$$L(x_k,C_k)=L(x_k,\bar{C}),\forall\, k\geqslant K_0$$

由定理1和(10)式知$\{L(x_k,\bar{C})\}$单调下降；由于$\{x_k\}$有界，故$\{L_k\}$收敛，而 $L_{k+1}\leqslant L_k+u\tau_k\nabla_x L_k^T d_k$。由定理1

$$L_{k+1}\leqslant L_k-u\tau_k\eta\|d_k\|^2$$

即　$\lim\limits_{k\to\infty}\tau_k\|d_k\|^2=0$

由文[5]引理1

$$\lim_{k\to\infty}\|d_k\|=0$$

证毕

定理4：$\{x_k\}$的聚点为问题(1)的 K—T 点。

证：由定理3和(8)式及(13)式

$$\lim_{k\to\infty}\|h_k\|=0,\qquad \lim_{k\to\infty}(g_k-\nabla h_k\bar{u}_k)=0$$

如 x^* 为$\{x_k\}$之聚点，显然

$$h(x^*)=0,\qquad g(x^*)-\nabla h(x^*)\lambda^*=0$$

即 x^* 为问题(1)的 K—T 点。 证毕

因 $\{x_k\}$ 有界，因此有聚点存在。下面推出算法的超线性收敛性。

定理5： 设 $x_k \to x^*$，且 $\|x_k+d_k-x^*\|=O(\|x_k-x^*\|)$，则当 k 充分大时 $\tau_k=1$。

证： 不妨设 $k \geqslant K_0$， 则 $C_k \equiv C_{k_0}=\bar{C}$，

由(10)式，只须证明

$$L(x_k+d_k,\bar{c}) \leqslant L(x_k,\bar{c})+u\nabla_x L(x_k,\bar{c})^T d_k$$

由 Taylor 展式

$$f(x_k+d_k)=f_k+\frac{1}{2}(g_k+g_{k+1})^T d_k+O(\|d_k\|^2)$$

$$=f_k+\frac{1}{2}(g_k+g(x^*))^T d_k+O(\|d_k\|^2)$$

最后一个等式由假设推得。同样

$$h(x_k+d_k)=h_k+\frac{1}{2}(\nabla h_k+\nabla h(x^*))^T d_k+O(\|d_k\|^2)$$

注意 $\|h_{k+1}\|^2=O(\|d_k\|^2)$

$$L(x_k+d_k,\bar{c})-L(x_k,\bar{c})=\frac{1}{2}\nabla_x L_k^T d_k+\frac{1}{2}d_k^T(g(x^*)-\nabla h(x^*)\lambda_{k+1})+O(\|d_k\|^2)$$

又 $g(x^*)-\nabla h(x^*)\lambda(x^*)=0$

因此 $L(x_k+d_k,\bar{c})-L(x_k,\bar{c})=\frac{1}{2}\nabla_x L_k^T d_k+O(\|d_k\|^2)$

但 $\nabla_x L_k^T d_k \leqslant -\eta\|d_k\|^2<0$，又 $\forall\ u<\frac{1}{2}$，$L(x_k+d_k,\bar{c})<L(x_k,\bar{c})+u\nabla L_k^T d_k$

即 $\tau_k=1$ 证毕

定理6： 如 $x_k \to x^*$，$\nabla h(x^*)$ 列满秩，约束优化的二阶最优性条件在 x^* 成立，P_k 是 R_n 到子空间 $\{s|s^T\nabla h_k=0\}$ 的投影阵。则 $\lim\limits_{k\to\infty}\|P_k(B_k-H(x^*))d_k\|/\|d_k\|=0$ 的充要条件是 $\|x_k+d_k-x^*\|=O(\|x_k-x^*\|)$。

由定理5和6知

推论：设定理6假设及必要条件成立。则 $\{x_k\}$ 超线性收敛，从而避免了 Maratos 效应。

3 算法一、二的具体实施

我们先讨论如何求得算法一中二次规划(3)的最小范数解。首先考虑下列问题的最小范数解：

$$\begin{cases} \min \frac{1}{2}\|d\|_B^2 \\ s.t. \quad Ad=b \end{cases}$$

其中　$A\in R^{m\times n}(m\leqslant n)$，且 $r(A)=m$。

用 Lagrange 乘子法求解上述问题，得到

$$d = B^{-1}A^T(AB^{-1}A^T)b$$

于是子问题(4)或(3)的解可直接写为：

$$d_k = B_k^{-1}\nabla h_k(\nabla h_k^T B_k^{-1}\nabla h_k)^{-1}(-h_k+\nabla h_k^T B_k^{-1}g_k) - B_k^{-1}g_k \tag{15}$$

令　　$\bar{b}_k=-h_k+\nabla h_k^T B_k^{-1}g_k,\ A_k=\nabla h_k^T B_k^{-1}\nabla h_k$

求解　$A_k\bar{d}_k=\bar{b}_k$，将其解 $\bar{d}_k$ 代入(15)式得

$$d_k = B_k^{-1}(\nabla h_k\bar{d}_k - g_k) \tag{16}$$

这样我们直接按(16)式求解子问题(3)。

因 A_k 正定，求解线性方程组 $A_k\bar{d}_k=\bar{b}_k$，也可采用 Cholesky 分解方法。

算法二中，$\lambda_k=(\nabla h_k^T\nabla h_k)^{-1}\nabla h_k g_k$　也可采用同样方法求得。

B_k^{-1} 的修正可采用 DFP 的 Powell 校正形式，详见文[5]。

算法一、二中 $\Psi_k(p_k)$的计算采用

$$\Psi_k(p_k) = g_k^T p_k + \frac{1}{2}\tau_k p_k^T(\nabla h_k\bar{d}_k - g_k)$$

避免涉及 B_k，此时 $d_k^T B_k d_k=d_k^T(\nabla h_k\bar{d}_k-g_k)$

两个算法的数值计算实例及进一步的讨论可参见文[5]，这里不再赘述。

参 考 文 献

1 Coleman T F, Conn A R. On the local convergence of a Quasi—Newton method for the nonlinear programming problem. SIAM J. Numer. Anal. ,1984,22:775～769

2 Vardi Avi. A trust region algorithm for equality constrained minimization convergent properties and implementation. SIAM J. Numer. Anal. ,1985,22:575～591

3 Masao F. A successive quadratic programming algorithm with global and superlinear convergence properties. Math. Prog. ,1986,35:253～261

4 Dembo R S. ,Eisenfut S C ,Steihang T. Inexact Newton methods. SIAM J. Numer. ,1982,19:400～408

5 唐健. 非线性规划的信赖域方法. 重庆大学硕士学位论文,1987,

6 张建中. 约束极值问题的 SQP 方法. 高校应用数学学报,1987,(2):87～94

7 Boggs D T, Tolle J W. On the local convergence of Quasi—Newton method for constrained optimization. SIAM Control Optim. ,1982,20:161～171