第 43 卷 第 5 期　　西南师范大学学报(自然科学版)　　2018 年 5 月
Vol. 43 No. 5　　Journal of Southwest China Normal University (Natural Science Edition)　　May. 2018

DOI:10.13718/j.cnki.xsxb.2018.05.005

非线性方程组的一个不使用罚函数和 filter 的算法①

房月华

衡水学院 数学与计算机科学系,河北 衡水 053000

摘要:给出了求解非线性方程组的一个新算法,首先将非线性方程组转化为一个非线性规划,再使用一个不使用罚函数和 filter 的算法求解这个非线性规划,在 Jacobi 矩阵一致列满秩的条件下证明由算法产生序列的极限点是非线性方程组的解.通过在算法中引进二阶校正技术来克服可能的 Maratos 效应,可以证明这个方法是局部超线性收敛的.

关 键 词:罚函数;filter 技术;线搜索;超线性收敛

中图分类号:O221　　**文献标志码:A**　　**文章编号:**1000-5471(2018)05-0023-08

前人提出了求解非线性规划等式约束的一个新方法,这个方法既不使用罚函数、障碍函数,也不使用 filter 算法,而且最终证明该算法全局收敛到原问题的一个阶稳定点,数值试验证明这种方法的数值结果很好. Yamashita 等[1] 通过不使用罚函数和 filter 的信赖域方法求解了一个非线性等式约束和非负约束的优化问题.这 2 个方法都是使用信赖域方法产生新的迭代,求解不同的子问题处理非线性的目标函数和约束. Liu 等[2] 不使用罚函数和 filter 算法求解含有非线性等式约束的非线性规划问题,在温和的假设条件下证明算法至少找到一个 KKT 点,或是找到一个 Fritz-John 点或是不可行的稳定点,并证明算法是局部超线性收敛的.在此基础上,本文考虑了非线性方程组

$$F(x)=\begin{pmatrix}F_1(x)\\F_2(x)\\\cdots\\F_m(x)\end{pmatrix}=0$$

其中 $F_i(x)$ 是 R^n 上的非线性函数.将非线性方程组转化为非线性规划再应用文献[2]的不使用罚函数和 filter 的算法求解该非线性规划.通过该方法可求得非线性方程组的近似解,它也是极小化方程组残差的范数的稳定点,而且在适当的假设下,这个方法是局部超线性收敛的.

1 非线性方程组的一个不使用罚函数和 filter 算法

1.1 算法提出

本文考虑了非线性方程组

$$F(x)=\begin{pmatrix}F_1(x)\\F_2(x)\\\cdots\\F_m(x)\end{pmatrix}=0 \tag{1.1}$$

① 收稿日期:2017-03-18
基金项目:衡水学院院级课题(2010009).
作者简介:房月华(1983-),女,硕士研究生,讲师,主要从事最优算法研究.

其中 $F_i(x)$ 是 R^n 上的非线性方程.

定义集合：$I_0 = \{i: |F_i(x^0)| > \delta\}$，$\varepsilon_0 = \{i: |F_i(x^0)| \leqslant \delta\}$，令 $h(x) = (F_i(x), i \in \varepsilon_0)^T \in R^{|\varepsilon_0|}$，$G(x) = (F_i(x), i \in I_0)^T \in R^{|I_0|}$，其中 $|\varepsilon_0|$ 和 $|I_0|$ 分别表示指标集 ε_0 和 I_0 中元素的个数. 这样就形成了一个带有等式约束的非线性规划

$$\min f(x) = \frac{1}{2}\|G(x)\|^2 \tag{1.2}$$

$$s.\ t.\ h(x) = 0 \tag{1.3}$$

对于非线性规划(1.2)－(1.3)，它的 Larange 函数为 $L(x, \lambda) = f(x) + \lambda^T h(x) = \frac{1}{2}\|G(x)\|^2 + \lambda^T h(x)$ 它的 KKT 条件为

$$\nabla G(x)G(x) + A(x)\lambda = 0 \tag{1.4}$$

$$h(x) = 0 \tag{1.5}$$

其中 $\nabla G(x)G(x)$ 是 $f(x)$ 的梯度，$A(x) = \nabla h(x)$. 对于每个 x 定义约束违背的度量函数 $v(x) = \|h(x)\|$，对于 $d \in R^n$，由于 $h_j: R^n \to R$ 是二次连续可微的，$v(x)$ 沿 d 的方向总有方向导数，令 $v'(x, d)$ 为 $v(x)$ 的方向导数，$\varphi(x, d) = \|h(x) + A^T(x)d\| - \|h(x)\|$，则 $v'(x, d) \leqslant \varphi(x, d)$，由于 $\varphi(x, d) \leqslant 0$，所以 d 为 $v(x)$ 的下降方向.

下面给出具体的算法.

算法 1.1.1

步 1：给定 x^0，$\delta > 0$，$\sigma \in \left(0, \frac{1}{2}\right)$，$\xi_1 > 0$，$\xi_2 > 0$，$0 < c_1 < c_2 < 1$，$0 < \varepsilon < 0.5$，令 $v_{\max}^0 = 0$，$r_0 = 0.9$，$k = 0$；

步 2：将方程违背的量 $F_i(x^0)$ 按从小到大排列，取 δ 为由方程违背的量 $F_i(x^0)$ 组成的数组中间的数，形成了下面的非线性规划

$$\min f(x) = \frac{1}{2}\|G(x)\|^2$$

$$s.\ t.\ h(x) = 0$$

计算 $\nabla G(x)G(x)$，h^0，A_0，B_0；

步 3：当 x^k 不满足 KKT 条件时(进入循环)

求解下面的子问题

$$\min(\nabla G(x^k)G(x^k))^T d + \frac{1}{2}d^T B_k d \tag{1.6}$$

$$s.\ t.\ h_j^k + (a_j^k)^T d = 0 \tag{1.7}$$

令 d^k 为该问题的解.

步 4：步长 $\alpha_k \in (0, 1]$，使之满足不等式

$$f(x^k + \alpha_k d^k) - f(x^k) \leqslant \min\{\sigma\alpha_k (g^k)^T d, -\xi_1 v(x^k + \alpha_k d^k)\} \tag{1.8}$$

$$v(x^k + \alpha_k d^k) \leqslant \max\{(r_k + 1)/2, 0.95\} v_{\max}^k v_{\max}^k \neq 0 \tag{1.9}$$

或者

$$v(x^k + \alpha_k d^k) - v(x^k) \leqslant \min\{\sigma\alpha_k \varphi(x^k, d_k), -\xi_2 \alpha_k^2 \|d^k\|^2\} \tag{1.10}$$

令 $x^{k+1} = x^k + \alpha_k d^k$

步 5：如果(1.10)在 x^{k+1} 处成立而在 x^k 外不成立，令 $v_{\max}^{k+1} = v(x^k)$，否则 $v_{\max}^{k+1} = v_{\max}^k$；计算 $\nabla G(x)G(x)$，h^{k+1}，A_{k+1}，迭代 B_k 到 B_{k+1}. 如果(1.10)成立，$r_k = v(x^{k+1})/v(x^k)$，否则的话，有 $r_{k+1} = r_k$，令 $k = k + 1$，回步 3；

(循环结束)

对于非线性方程组(1.1)，这个算法最终在迭代点是非线性规划(1.2)－(1.3)一个 KKT 点，即 $\|\nabla G(x^k)G(x^k) + A(x^k)\lambda^k\| = 0$ 及 $\|h(x^k)\| = 0$ 的情况下终止. 由于 $\nabla G(x^k)G(x^k) + A(x^k)\lambda^k = 0$，

可以写为$(\nabla G(x^k)A(x^k))\begin{pmatrix}G(x^k)\\ \lambda^k\end{pmatrix}=0$. 在$(\nabla G(x^k)A(x^k))$满足列满秩的情况下，有$\begin{pmatrix}G(x^k)\\ \lambda^k\end{pmatrix}=0$，也就是$G(x^k)=0$和$\lambda^k=0$，若方程组(1.1)的 Jacobi 矩阵$\nabla F(x)$一致列满秩，KKT 条件的满足与$F(x^k)=0$等价. 为了使算法有限终止，在算法运行的过程中需要引入 KKT 条件残差忍受度的界.

我们知道，如果B_k在A^T的零空间上正定，子问题(1.6)－(1.7)有唯一解且d^k满足下面个条件

$$\nabla G(x^k)G(x^k)+B_kd^k+A(x^k)\lambda^k=0 \tag{1.11}$$

$$h_j^k+(a_j^k)^Td=0 \tag{1.12}$$

其中λ^k是相应的乘子. 如果$\|d^k\|=0$，则有

$$\nabla G(x^k)G(x^k)+A(x^k)\lambda^k=0 \tag{1.13}$$

$$h(x^k)=0 \tag{1.14}$$

显然，此时x^k是非线性规划(1.2)－(1.3)的一个 KKT 点.

1.2　全局收敛性分析

假设 1.2.1

(1) $F(x)$的每个非线性函数$F_i(x)$在R^n上是二次连续可微的；

(2) $F(x)=0$的 Jacobi 矩阵$\nabla F(x)$满足一致列满秩的条件；

(3) $\lim\limits_{k\to\infty}f^k>-\infty$，$C=\{x\mid x\in R^n,\ v(x)\leqslant v^0\}$有界，$v^0$是在开始$h$-型迭代时$v(x)$的值. 如果所有的迭代都是$f$-型迭代，$v^0$就是一个给定的常数；

(4) 对于某个正常数$\beta>0$，$\|B_k\|\leqslant\beta$，对$\forall d\neq0$，$A_k^Td=0$，有$d^TB_kd\geqslant\gamma\|d\|^2$，即对每个$k$，$B_k$在$A^T$的零空间上有界正定.

引理 1.2.2　若假设 1.2.1 成立，如果对于所有的$k\geqslant0$，$\|A_kh^k\|\geqslant\eta\|h^k\|$，其中$\eta\in(0,1)$上常数，$\xi_2\leqslant\dfrac{\sigma}{1-\sigma}M$，$\xi_1\leqslant\min\left\{\dfrac{\sigma}{2(1-\sigma)},\ \min\{0.1,(1-\sigma)/8\}\sigma k_2\eta^2\gamma^0M\right\}$，其中$M$，$\sigma$，$k_2$，$\gamma^0$均为正常数，那么存在$\tilde t\in(0,1)$使得$\tilde t\leqslant\alpha\leqslant1$，当$\alpha_k$有界且$\alpha_k\neq0$，有(1.8)和(1.9)成立或是(1.10)成立.

定理 1.2.3　假设引理 1.2.2 的条件成立，那么每有限次迭代有(1.8)和(1.9)成立或者是$\lim\limits_{k\to\infty}v(x^k)=0$且$\lim\limits_{k\to\infty}\|d^k\|=0$

证　如果在第k次迭代时，(1.10)成立，由引理 1.2.2 则有

$$v(x^{k+1})-v(x^k)\leqslant\sigma\alpha_k\varphi(x^k,d_k)\leqslant\sigma\alpha_kk_2\eta^2\|h^k\|\leqslant-\eta'\|h^k\|$$

其中$\eta'\in(0,1)$，是一个常数. 所以有

$$v(x^{k+1})\leqslant(1-\eta')v(x^k)$$

再假设$\eta''=1-\eta'$，则$\eta''\in(0,1)$，$v(x^{k+1})\leqslant\eta''v(x^k)$.

如果(1.8)和(1.9)在任何有限次迭代都成立，不失一般性，假设对于整个迭代序列(1.10)都成立. 由$v(x^{k+1})\leqslant(1-\eta')v(x^k)$可知，$\lim\limits_{k\to\infty}v(x^k)=0$. 再由(1.10)式，

$$v(x^{k+1})-v(x^k)\leqslant-\xi_2\alpha_k^2\|d^k\|\leqslant-\xi_2\tilde t^2\|d^k\|^2\leqslant0$$

对于上式两边取极限，得到$\lim\limits_{k\to\infty}\|d^k\|=0$，证毕.

定理 1.2.4　假设引理 1.2.2 的条件成立，且对于充分大的正整数k，(1.8)式成立，那么有$\lim\limits_{k\to\infty}v(x^k)=0$且$\lim\limits_{k\to\infty}\|d^k\|=0$.

证　如果对于充分大的k有(1.8)式成立，那么有限次迭代之后$\{f(x^k)\}$是非单调递减的.

由于

$$f(x^{k+1})-f(x^k)\leqslant-\xi_1v(x^{k+1})$$

对上面的不等式两边取极限，则有$\lim\limits_{k\to\infty}v(x^k)=0$.

如果对于某一个$k\in K$，$d^k\not\to0$其中K是指标集的无限子集，由于$\lim\limits_{k\to\infty}v(x^k)=0$，且对某个常数$\gamma>0$，$(\nabla G(x^k)G(x^k))^Td^k=-\gamma_0\|d^k\|^2$，那么对于充分大的$k\in K$，

$$f(x^{k+1})-f(x^k)\leqslant\sigma\alpha_k(\nabla G(x^k)G(x^k))^T d^k\leqslant\sigma\tilde{t}\gamma_0\|d^k\|^2$$

又由于$\{f(x^k)\}$非单调递减且有下界，对上式两边取极限，则有$\lim\limits_{k\to\infty,\,k\in K}\|d^k\|=0$，与假设矛盾，从而结论成立，证毕.

定理 1.2.5 假设引理 1.2.2 的条件成立，那么迭代序列的每一个极限点都是问题(1.2)-(1.3)的一个可行点，在这些极限点中有一个 KKT 点.

证 令x^*是序列$\{x^k\}$的任一极限点. 如果$\lim\limits_{k\to\infty}v(x^k)=0$和$\lim\limits_{k\to\infty}\|d^k\|=0$，对(1.11)的两边取极限，则有

$$\nabla G(x^*)G(x^*)+A(x^*)\lambda^*=0 \tag{1.15}$$

其中$\nabla G(x^*)G(x^*)$是目标函数的梯度，$A(x^*)=(a_i(x^*))$，$A(x^*)\lambda^*$是$\{A(x^k)\lambda^k\}$的一个极限. 由下面极限$\lim\limits_{k\to\infty}v(x^k)=0$可知，$h^*=0$. 由定理 1.2.3 和 1.2.4 可知，如果除了有限次迭代都是f-型迭代或是h-型迭代，那么$\{x^k\}$的每个极限点x^*都是(1.2)-(1.3)的一个 KKT 点. 因此我们只需考虑$\{x^k\}$是由有限次f-型迭代和是h-型迭代轮流执行组成的情况. 换句话说，首先是有限次h-型迭代，再是有限次f-型迭代，然后又是有限次h-型迭代，以此类推. 不是一般性，我们假设从第$(k_{2l-1}+1)$次迭代到第(k_{2l})次迭代和第$(k_{2l+1}+1)$次迭代到第(k_{2l+2})次迭代都是h-型迭代，第$(k_{2l}+1)$次迭代到第(k_{2l+1})次迭代都是f-型迭代.

我们首先证明$\lim\limits_{k\to\infty}v(x^k)=0$. 由引理 1.2.2 可知，存在常数$\tilde{t}>0$，对于所有的$k\geqslant 0$使得$\alpha_k\geqslant\tilde{t}$，那么有对于在$x^j$处任一$h$-型迭代，(1.10)可知，

$$v(x^{j+1})-v(x^j)\leqslant\sigma\alpha_j\varphi(x^j,d^j)\leqslant-\sigma\tilde{t}k_2\eta^2 v(x^j)$$

因而对所有的$k\geqslant 0$，

$$r_k\leqslant\hat{r}=1-\sigma\tilde{t}k_2\eta^2<1,\ \max\{(r_k+1)/2,0.95\}=\max\{(\hat{r}_k+1)/2,0.95\}<1$$

若令$\tilde{r}=\max\{(\hat{r}_k+1)/2,0.95\}$，根据算法有

$$v(x^{k_{2l-1}})>v(x^{k_{2l-1}+1})>\cdots>v(x^{k_{2l}}) \tag{1.16}$$

$$v(x^{k_{2l+1}})>v(x^{k_{2l+1}+1})>\cdots>v(x^{k_{2l+2}}) \tag{1.17}$$

且$f(x^{k_{2l}+1})>f(x^{k_{2l}+2})>\cdots>f(x^{k_{2l+1}})$

$$v(x^{k_{2l}+1})\leqslant\tilde{r}v(x^{k_{2l-1}}),\ v(x^{k_{2l}+2})\leqslant\tilde{r}v(x^{k_{2l-1}})\cdots,\ v(x^{k_{2l+1}})\leqslant\tilde{r}v(x^{k_{2l-1}}) \tag{1.18}$$

即对于h-型迭代序列，约束违背的度量$v(x^k)$是单调下降的，对于f-型迭代序列，尽管约束违背的l_2-范数不是单调下降的，但他们都小于前一次h-型迭代的最大值.

下面我们考虑每有限步非单调下降条件(1.10)和(1.8)-(1.9)交替执行的情形. 如果有限步(1.10)和(1.8)-(1.9)交替执行，那么对于整数l，序列$\{\cdots,v(x^{k_{2l-1}}),v(x^{k_{2l+1}})\cdots\}$是单调下降的，且$v(x^{k_{2l+1}})\leqslant\tilde{r}v(x^{k_{2l-1}})$，最终有$\lim\limits_{k\to\infty}v(x^{k_{2l+1}})=0$. 由(1.16)与(1.17)和(1.18)及$v(x^k)$的非负性，则有$\lim\limits_{k\to\infty}v(x^k)=0$

下面证明存在序列$\{d^k:k\in K\}$，$\lim\limits_{k\in K,\,k\to\infty}d^k=0$. 假设$d^k\not\to 0$，则存在常数$\varepsilon>0$，对于充分大的$k$，使得$\|d^k\|^2\geqslant\varepsilon$. 由引理 1.2.2 可知，

$$-\xi_2\alpha_k^2\|d^k\|^2\leqslant-\xi_2\tilde{t}^2\varepsilon^2 \tag{1.19}$$

另一方面，$\sigma\alpha_k\varphi(x^k,d^k)\geqslant-\sigma\alpha_k\|h^k\|\geqslant-\sigma v(x^k)$，所以如果$v(x^k)\leqslant\xi_2\tilde{t}^2\varepsilon^2/\sigma$，那么

$$\sigma\alpha_k\varphi(x^k,d^k)\geqslant-\xi_2\alpha_k^2\|d^k\|^2 \tag{1.20}$$

如果(1.10)所立，由(1.19)和(1.20)，有

$$v(x^{k+1})-v(x^k)\leqslant-\xi_2\alpha_k^2\|d^k\|^2\leqslant-\xi_2\tilde{t}^2\varepsilon^2$$

与$\lim\limits_{k\to\infty}v(x^k)=0$矛盾，也就是说对于充分大的$k$，(1.8)和(1.9)成立，即

$$f(x^{k+1})-f(x^k)\leqslant\sigma\alpha_k(\nabla G(x^k)G(x^k))^T d^k\leqslant\sigma\tilde{t}\gamma_0\varepsilon^2$$

与$\{f(x^k)\}$有下界矛盾，从而在某一指标集K上，$\lim\limits_{k\in K,\,k\to\infty}\|d^k\|=0$，即存在指标集$K'\subseteq K$使得当$k\to\infty$

时，序列$\{x^k: k \in K'\}$收敛到问题(1.2)－(1.3)的一个 KKT 点，证毕.

1.3　局部收敛性

为了获得快速的局部收敛速率，克服 Marotos 效应，在算法 1.1.1 中引入二阶矫正技术(soc)，得到如下算法.

算法 1.3.1

步 1：给定 x^0，$\delta > 0$，$\sigma \in \left(0, \frac{1}{2}\right)$，$\xi_1 > 0$，$\xi_2 > 0$，$0 < c_1 < c_2 < 1$，$0 < \varepsilon < 0.5$ 和 v_{soc} 令 $v_{\max}^0 = 0$，$r_0 = 0.9$，$k = 0$；

步 2：将方程的违背的量 $F_i(x^0)$ 按从小到大排列，取 δ 为由方程的违背的量 $F_i(x^0)$ 组成的数组的中间的数，形成了下面的非线性规划

$$\min f(x) = \frac{1}{2}\| G(x) \|^2 \tag{1.21}$$

$$s.t.\ h(x) = 0 \tag{1.22}$$

计算 $\nabla G(x^0)G(x^0)$，h^0，A_0，B_0；

步 3：当 $\| d^k \| \neq 0$ 且不满足 KKT 条件时(进入循环)

求解下面的子问题

$$\min(\nabla G(x^k)G(x^k))^T d + \frac{1}{2}d^T B_k d \tag{1.23}$$

$$s.t.\ h_j^k + (a_k^j)^T d = 0 \tag{1.24}$$

令 d^k 为该问题的解.

步 4：如果

$$f(x^k + d^k) - f(x^k) \leqslant \min\{\sigma\alpha_k (g^k)^T d, -\xi_1 v(x^k + d^k)\} \tag{1.25}$$

$$v(x^k + d^k) \leqslant \max\{(r_k + 1)/2, 0.95\} v_{\max}^k \qquad v_{\max}^k \neq 0 \tag{1.26}$$

或者

$$v(x^k + d^k) - v(x^k) \leqslant \min\{\sigma\varphi(x^k, d_k), -\xi_2 \| d^k \|^2\} \tag{1.27}$$

成立. 令 $x^{k+1} = x^k + d^k$，否则(1.12)成立，且 $v(x^k) \leqslant v_{soc}$，求矫正子问题

$$\min(\nabla G(x^k)G(x^k))^T (d^k + d) + \frac{1}{2}(d^k + d)^T B_k (d^k + d) \tag{1.28}$$

$$s.t.\ h_j^k + (a_k^j)^T (d^k + d) = 0 \tag{1.29}$$

令 $\widetilde{d}^k$ 为该问题的解.

步 5：检验是否有下面不等式成立

$$f(x^k + d^k + \widetilde{d}^k) \leqslant \min\{\sigma (g^k)^T d, -\xi_1 v(x^k + d^k + \widetilde{d}^k)\} \tag{1.30}$$

$$v(x^k + d^k + \widetilde{d}^k) \leqslant \max\{(r_k + 1)/2, 0.95\} v_{\max}^k \qquad v_{\max}^k \neq 0 \tag{1.31}$$

或者

$$v(x^k + d^k + \widetilde{d}^k) - v(x^k) \leqslant \min\{\sigma\varphi v(x^k + d^k + \widetilde{d}^k), -\xi_2 \| d^k \|^2\} \tag{1.32}$$

如果(1.30)和(1.31)或者(1.32)成立，令 $x^{k+1} = x^k + d^k + \widetilde{d}^k$. 否则选取尽可能大的步长 $\alpha_k \in (0, 1]$，使之满足不等式

$$f(x^k + \alpha_k d^k) - f(x^k) \leqslant \min\{\sigma\alpha_k (g^k)^T d, -\xi_1 v(x^k + \alpha_k d^k)\} \tag{1.33}$$

$$v(x^k + \alpha_k d^k) \leqslant \max\{(r_k + 1)/2, 0.95\} v_{\max}^k \qquad v_{\max}^k \neq 0 \tag{1.34}$$

或者

$$v(x^k + \alpha_k d^k) - v(x^k) \leqslant \min\{\sigma\alpha_k \varphi(x^k, d_k), -\xi_2 \alpha_k^2 \| d^k \|^2\} \tag{1.35}$$

令 $x^{k+1} = x^k + \alpha_k d^k$.

步 6：如果(1.27)或是(1.35)在 x^{k+1} 处成立而在 x^k 处不成立，令 $v_{\max}^{k+1} = v(x^k)$，否则就有 $v_{\max}^{k+1} = v_{\max}^k$；计算 $\nabla G(x^{k+1})G(x^{k+1})$，$h^{k+1}$，$A_{k+1}$，迭代 B_k 到 B_{k+1}. 如果(1.34)成立，$r_k = v(x^{k+1})/v(x^k)$，否则 $r_{k+1} = r_k$，令 $k = k + 1$，回步 3；

(循环结束)

假设 1.3.2 ① 当 $k \to \infty$，$x^k \to x^*$，其中 x^* 是问题(1.21)－(1.22)的一个 KKT 点，λ^* 是相应的 Lagrange 乘子向量；② $f(x)$ 和 $h(x)$ 是二次可微的，它们的二阶导数在 x^* 处是 Lipschitz 连续的；③ $J^k = J^*$，其中 J^k 和 J^* 分别是 x^k 和 x^* 处的 Jacobi 矩阵的线性无关的最大行数；④ 对任意的 j 不属于 J^*，$\lambda^{j*}=0$ 且对任意的 j 不属于 J^k，$\lambda^{jk}=0$；(5) $d^T \nabla_{xx}^2 L(x^*, \lambda^*) d \geqslant \gamma \| d \|^2$，对于任意的 $d \in \{d \mid d \neq 0, \nabla h_j(x^*)^T d = 0, j \in J^*\}$，其中 $L(x,\lambda) = f(x) + \lambda h(x)$，$\gamma > 0$ 是一个常数.

定理 1.3.3 如果假设 1.3.2 成立，$\| h^k + A_k^T d_p^k \| \leqslant (1-\eta_k) \| h^k \|$，$(0 < \eta_k \leqslant 1)$；

1) 若 $\| (B_k - \nabla_{xx}^2 L(x^*, \lambda^*)) d^k \| = o(\| d^k \|)$，$1-\eta_k = O(1)$，则 $\lim\limits_{k\to\infty} \| x^k + d^k - x^* \| / \| x^k - x^* \| = 0$.

2) 若 $B_k = \nabla_{xx}^2 L(x^k, \lambda^k)$，$1-\eta_k = O(1)$，那么有 $\| x^k + d^k - x^* \| = O(\| x^k - x^* \|^2)$.

假设 1.3.4 假设 $\widehat{\lambda_k^k} = 0$，对任意的 j 不属于 J^k，$\widehat{\lambda^{jk}} \in R^m$ 是(1.31)－(1.32)相应的 Lagrange 乘子.

定理 1.3.5 若假设 1.3.2 和假设 1.3.4 成立，如果对于每一个 $d \in R^n$，都有下面的等式成立：

$\| (B_k - \nabla_{xx}^2 L(x^*, \lambda^*)) d^k \| = o(\| d^k \|)$，$\| h^k + A_k^T d_p^k \| \leqslant (1-\eta_k) \| h^k \|$，$(0 < \eta_k \leqslant 1)$，$1-\eta_k = O(1)$，那么有 $\| \widehat{d^k} \| = o(\| d^k \|)$ 成立.

定理 1.3.6 若假设 1.3.2 和假设 1.3.4 成立，如果对于每一个 $d \in R^n$，都有下面的等式成立：

$\| (B_k - \nabla_{xx}^2 L(x^*, \lambda^*)) d^k \| = o(\| d^k \|)$，$\| h^k + A_k^T d_p^k \| \leqslant (1-\eta_k) \| h^k \|$，$(0 < \eta_k \leqslant 1)$，且有 $1-\eta_k = O(1)$，那么对于充分大的 k，有 $x^{k+1} = x^k + d^k$ 或者 $x^{k+1} = x^k + d^k + \widehat{d^k}$ 成立.

由以上的分析可知，算法 1.3.1 中引入的二阶校正步无论是使得(1.30)和(1.31)成立，还是使得(1.32)成立都可使算法是局部二次收敛的.

2 数值结果及分析

本文算法是在 Windows XP 系统下使用 MATLAB7.0 版本运行的，子问题使用的是 Matlab 工具箱中的 M-函数 quadprog 来求解，并且在求解子问题之前对 A 作 QR 分解去掉多余的约束，与多余的约束相应的乘子设为 0. 使用最小二乘法求解时使用的是 Matlab 工具箱中 M-函数 Lsqnonlin 来求解的.

本文的测试例子取自文献[9]标准初始点下，笔者将本文算法分别与一般求解非线性方程组的非线性最小二乘法做了比较(表 1).

表 1 与最小二乘法结果比较

PROB	N	M	FLSQ	FNLP	IFLSQ	IFNLP	TLSQ	TNLP
BBF	2	2	1.972 2e-016	2.775 6e-015	23	5	0.078 1	0.046 9
PSF	4	4	2.498 3e-006	1.412 5e-006	27	16	0.656 3	0.146 2
PBSF	2	2	2.994 8e-006	0.001 9e-008	18	18	0.052 8	0.937 5
WOF	4	6	7.877 0	1.028 3	15	11	0.218 8	0.115 6
HVF	3	3	0.316 9	1.262 7e-003	35	14	0.871 9	0.253 8
WAF	6	31	1.876e-004	3.246 3e-005	40	41	1.684 5	1.796 9
BALF	12	12	9.583 7e-011	3.718 5e-014	10	4	0.932 5	0.328 1
VDF	12	14	1.324e-009	1.511 9e-008	18	10	0.511 6	0.043 9
DBVF	12	12	1.193 5e-005	4.513 0e-005	42	6	1.150 7	0.856 3
TRF	12	12	8.204 8e-006	1.907 3e-006	34	30	1.578 1	1.546 9
BTF	12	12	3.101 6e-011	4.048 3e-009	12	10	0.078 1	0.061 5
DIEF	12	12	2.032 2e-004	2.506 8e-004	11	8	0.085 4	0.051 3

在表 1 中，PROB 表示数值试验所要测试的问题，每个测试问题使用它的首字母缩写来简写，如第一个测试问题 RBF 表示的是 Rosenbrock funcuion. FLSQ 与 FNLP 在最优解 x^* 处最小二乘法的方程组残差和本文算法方程组的残差值，IFLSQ 和 IFNLP 分别表示最小二乘法和本文算法在求解时迭代的次数. TLSQ 和 TNLP 分别最小二乘法求解和本文算法求解时所使用的 CPU 时间. 第二列中 N 代表非线性方程组中变量的维数，第三列中 M 代表非线性方程组中方程的个数. 这里将使用最小二乘法和本文算法求解非线性方程组的残差值、迭代次数和 CPU 时间(时间单位为秒)分别作了比较.

表 2 是通过改变问题 6—问题 12 变量的维数和方程的个数而得到的数值结果. 表 2 与表 1 中的测试问题属于同一类，但是变量的维数不同，方程的个数也不同，并将使用最小二乘法和本文算法求解非线性方程组时的残差值、迭代次数和 CPU 时间(时间单位为秒)分别作比较.

表 2　改变条件与最小二乘法的比较

PROB	N	M	FLSQ	FNLP	IFLSQ	IFNLP	TLSQ	TNLP
WAF	9	31	3.078 2e-004	9.808 4e-006	90	29	2.859 4	1.528 1
WAF	12	31	6.274 3e-005	2.021 3e-005	13	9	0.171 9	0.140 6
BALF	10	10	1.451 3e-010	1.005 9e-015	9	5	0.031 3	0.031 3
BALF	30	30	6.071 0e-015	1.677 7e-015	7	5	0.171 9	0.156 3
BALF	40	40	2.870 5e-009	4.001 2e-010	2	3	0.140 6	0.046 9
VDF	20	22	1.578 4e-010	1.670 5e-011	15	9	0.078 1	0.051 1
DBVF	10	10	8.380 1e-006	4.556 8e-008	39	9	1.125 0	0.571 9
DBVF	20	20	6.798 2e-009	3.793 5e-008	3	2	0.062 5	0.043 9
TRF	20	20	3.812 1e-004	2.167 5e-006	45	28	1.390 6	0.943 1
BTF	10	10	7.898 0e-007	1.557 6e-005	11	5	0.062 9	0.015 6
BTF	20	20	2.624 4e-011	6.126 2-009	20	9	0.046 3	0.171 5
DIEF	10	10	1.949 7e-011	2.793 0e-009	3	3	0.057 1	0.057 1
DIEF	20	20	2.081 2e-001	3.723 6e-009	19	14	0.945 1	0.721 1

从表 1 中可以看出，对于问题 3(Powell badly scaled function)，本文算法在最优解处得到的残差值比最小二乘法更精确，迭代次数也相同，但是本文算法使用的时候却比最小二乘法的时间要长一些，这是可以接受的. 一般来说，如果算法得到更高的精度或是更快的计算速度，必然以牺牲某一方面为代价. 为了避免由于初始点的选取而造成的偏差，更好地说明最小二乘法与本文的算法在问题 3 中的表现，我们改变初始点或是变量的维数. 问题 3 中变量的维数是给定的，从而我们改变它的初始点来比较 2 个方法. 表 1 中，标准初始点(0，1)且 f 在(1.09 8e-5，9.106)处有 $f=0$，即最优解为(1.098e-5，9.106). 给定初始点(0.000 4，8.1)，本文算法的残差值、迭代次数和 CPU 时间分别为 2.259 6e-8，4 次，0.046 9(s)，求得最小值点为(0.000 0，9.042 9)，最小二乘的结果为 2.755 7e-7，6 次，0.064 09(s)，求得最小值点为(0.000 0，8.721 1). 我们可以发现，越是接近最优解，本文算法比最小二乘法算得越好. 对于问题 4，我们会发现无论是最小二乘求得的残差值还是本文算法求得的残差值都不为零，我们改变它的初始值. 在表 1 中，问题 4 标准初始点为(—3，—1，—3，—1)，且 f 在(1，1，1，1)处有 $f=0$，即最优化解为(1，1，1，1). 我们给定一个新的初始点(0，0.5，0，0.5)本文算法的残差值、迭代次数和 CPU 时间分别为 1.513 8e-4，10 次，0.240 1(s)，求得最小值点为(0.958 4，0.986 9，1.001 6，1.003 3). 最小二乘法的结果为 0.009 3，7 次，0.171 4(s)，求得最小值点为(1.000 0，0.998 3，1.000 5，1.000 9)，越是接近最优解. 对于其他的问题，使用本文的算法求解非线性方程组无论是最优解处的残差值、迭代次数还是 CPU 时间都要比最小二乘法要好得多，因此本文的算法有效.

对于问题 6，由表 1 可知，使用本文算法所得到的最优解处方程组的残差值显然比用最小二乘法得到的残差值要好一些，但是迭代次数和 CPU 时间要比最小二乘法稍差一些，但这也可以接受. 为了更好地体现本文算法和最小二乘法的表现，我们改变问题中变量的维数. 这里我们之所以改变问题的维数是因为由文献[9]无法知道它的最优解，因而初始点也就无法根据最优解给出. 对于高维的方程组，文献[9]给出了维数 n 为多少时方程组的残差值，而且文献[9]数值试验也是通过改变维数来观察算法表现的. 问题 6 到问题 12 都是一些高维的问题，而且大部分初始值的给定也都与维数 n 有关，例如 $x_0=(1/n, 1/n, \cdots 1/n)$. 由表 2 可知，问题 6—问题 12 方程组的残差值并未因为方程组的个数和变量的维数为大而变得不好，而且有些甚至比维数低的算得还要好. 将表 1 和表 2 比较可以发现，使用本文的算法求解非线性方程组无论是最优解处方程组的残差值、迭代次数还是 CPU 时间都比最小二乘法要好，本文的算法有效.

3　结　　论

本文将求解非线性规划的方法应用于求解非线性方程组，新颖之处在于求解非线性规划时使用了一个

新的不使用罚函数和 filter 的方法求解，这样做的好处是可以不需要假定算法产生的迭代序列有界，同时也不需要象罚函数方法那样选择一个初始的罚参数或像滤子方法那样需要一个恢复阶段. 通过使用这个方法可以求得方程组的近似解，而且在适当的假设下，结合二阶校正技术，这个方法具有局部超线性收敛性.

参考文献：

[1] YAMASHITA Y H. A Globally and Superlinearly Convergent Trust-Region SQP Method Without a Penalty Function for Nonlinearly Constrained Optimization [J]. Math Program, 2005, 102(4): 111-151.

[2] LIU X W, YUAN Y X. A Sequential Quadratic Programming Method Without a Penalty Function or a Filter for Nonlinear Equality Constrained Optimization [J]. Siam Journal on Optimization, 2011, 21(2): 545-571.

[3] GOULD I M, TOINT P H. Nonlinear Programming Without a Penalty Function or a Filter [J]. Math Program, 2010, 122(1): 155-196.

[4] YUAN Y X. Trust Region Algorithm for Nonlinear Equations [J]. Information, 1998(1): 7-20.

[5] FAN J, PAN J. An Improved Trust Region Algorithm for Nonlinear Equations [J]. Computational Optimization & Applications, 2011, 48(1): 59-70.

[6] 黄敬频，陆云双，许克佶. 四元数体上统一代数 Lyapunov 方程的循环解及最佳逼近简 [J]. 西南师范大学学报(自然科学版), 2016, 41(4): 1-3.

[7] ESMAEILI H, KIMIAE M. A Trust Region Method with Improved Adaptive Radius for Systems of Nonlinear Equations [J]. Mathematical Methods of Operation Research, 2016, 83(1): 109-125.

[8] NIE P Y. Composite-Step Like Filter Method for Equality Constraint Problem [J]. Journal of Computational Mathematics, 2003, 21(5): 613-624.

[9] MORE J J, GARBOW B S, HILLSTROM K H. Testing Unconsrrained Optimization Software [J]. ACM Trans Math Software, 1981, 7(1): 17-41.

[10] NOCEDAL J, WRIGHT S. Numerical Optimization [M]. New York: Sprint Verlag, 1999.

[11] LIU X W, YUAN Y X. A Robust Algorithm for Optimization with General Equality and Inequality Constraints [J]. SIAM J SCI COMPUT, 2002, 2(2): 517-534.

[12] FAN J. Convergence Rate of The Trust Region Method for Nonlinear Equations Under Local Error Bound Condition [J]. Computational Optimization & Applications, 2006, 34(2): 215-227.

[13] 袁亚湘，孙文瑜. 最优化理论与方法 [M]. 北京：科学出版社，1997.

[14] 孙娇娇，储昌木. 一类凹凸拟线性椭圆方程解的存在性 [J]. 西南师范大学学报(自然科学版), 2016, 41(8): 21-24.

An Algorithm without A Penalty Function or A Filter for Nonlinear Equations

FANG Yue-hua

Mathematics and computer Department , Hengshui University, Hebei 053000, China

Abstract: We present a new algorithm for solving a system of nonlinear equations. The system of nonlinear equations is reformulated into a nonlinear programming problem firstly, and then we solve the problem by a method without a penalty function or a filter. Under the standard assumption that Jacobi matrices are uniformly full rank, it is proved that every limit point of the sequence generated by the algorithm is a solution of the system of nonlinear equations. Introducing the second-order correction technique to the algorithm for overcoming the so-called Maratos effect, this algorithm will locally have superlinear convergence.

Key words: a penalty function; a filter technique; line search; superlinear convergence

责任编辑　夏　娟