具有全局收敛性的非单调不精确牛顿法[①]

陈　静[②]　李正锋
（基础科学部）

摘　要　对大规模非线性方程组 $F=0$（其中 $F:R^n \to R^n$ 连续可微）提出 2 种非单调不精确牛顿法。在算法迭代过程中，每步求出 F 的局部线性化模型的一个近似解，而不要求 F 的某种范数单调递减，因此具有不精确牛顿法的优点，并且对非常病态的非线性方程组是有效的。在合理假设下证明此算法仍具有全局收敛性。

关键词　非线性方程组；病态问题；非单调；不精确牛顿算法；全局收敛性
中图分类号　O221.2

Nonmonotone Inexact Newton Methods With Global Convergence

Chen Jing　Li Zhengfeng
(Department of Basic Sciences)

Abstract　Two nonmonotone inexact Newton methods are proposed for finding a zero of $F:R^n \to R^n$. In the methods the solution is forced to satisfies the linear Newton equation approximately at each step and it is not necessary to reduce the norm of F. Therefore, they share the advantageous properties of inexact Newton methods, and are very effective for the ill-posed problems. Under reasonable assumptions, it is proved that these two methods are globally convergent.

Key words　nonlinear equations; ill-posed problems; nonmonotone; inexact Newton method; global convergence

牛顿方法是求解非线性方程组 $F(x)=0$ 的有效算法，这里 $x\in R^n$，$F:R^n \to R^n$ 是连续可微的；但如果 n 很大，且近似解 x_k 远离 F 的精确解，计算线性牛顿方程的精确解并不是必要的。文[1]提出的不精确牛顿法需求出线性牛顿方程的一个近似解；文[2]在此基础上提出了具有全局收敛性的不精确牛顿法，它要求迭代产生的点列 $\{\|F(x_k)\|\}$ 单调递减，因此使用该算法求解某些非常病态问题的迭代点列 $\{x_k\}$ 时，只能迂回地接近 F 的精确解 x_*。本文中提出 2 种非

收稿日期：1995-11-21
①国家自然科学基金和北京市自然科学基金资助项目
②陈　静，北京清华东路 17 号中国农业大学（东校区）71 信箱，100083

单调不精确牛顿法，它们不要求$\{\|F(x_k)\|\}$单调递减，并在合理假设下证明算法是全局收敛的。

1　2 种非单调不精确牛顿算法

算法 1

1)取初始点为 x_0，$t\in(0,1)$，$k=0$。

2)选取 $\eta_k(\in(0,1))$及 s_k，满足

$$\|F(x_k)+F'(x_k)s_k\|\leqslant\eta_k\|F(x_k)\|$$

及

$$\|F(x_k+s_k)\|\leqslant[1-t(1-\eta_k)]\|F(x_l(k))\| \tag{1}$$

其中

$$\|F(x_l(k))\|=\max_{0\leqslant j\leqslant m(k)}\{\|F(x_{k-j})\|\}$$

这里$\|\cdot\|$表示向量的范数。$m(k)$的定义为：$m(0)=0$ 且 $0\leqslant m(k)\leqslant\min\{m(k-1)+1,M\}$，$k\geqslant1$，$M$ 为一事先给定的非负整数。

3)令 $x_{k+1}=x_k+s_k$，若$\|F(x)\|=0$ 则停，否则继续。

4)令 $k=k+1$，转至 2)。

在算法 1 中，若令 $M=0$，则得到文[2]中的算法 GIN。

算法 2

1)取初始点为 x_0，$k=0$。

2)选取 $\eta_k(\in(0,1))$、参数 t_k 和 s_k，使得

$$\|F(x_k)+F'(x_k)s_k\|\leqslant\eta_k\|F(x_k)\|$$

和

$$\|F(x_k+s_k)\|\leqslant[1-t_k(1-\eta_k)]\|F(x_k)\| \tag{2}$$

成立。

3)令 $x_{k+1}=x_k+s_k$，若$\|F(x)\|=0$ 则停，否则继续。

4)令 $k=k+1$，转至 2)。

若令 $t_k\equiv t\in(0,1)$，则算法 2 也退化为文[2]中的算法 GIN。

2　算法的全局收敛性

假设 $F:\mathrm{R}^n\to\mathrm{R}^n$ 是一次连续可微的，令 $N_\delta(x)\equiv\{y\mid\|y-x\|<\delta\}$且 $\delta>0$。

证明算法 1 和算法 2 的全局收敛性需要下述引理。

引理 1[3]　假设 $F'(x)$是可逆的，则对任何 $\varepsilon>0$ 都存在 $\delta>0$，使得当 $y\in N_\delta(x)$时 $F'(y)$是可逆的，且

$$\|F'(y)^{-1}-F'(x)^{-1}\|<\varepsilon$$

成立。

引理 2[3]　对任何 x 和 $\varepsilon(>0)$都存在 $\delta>0$，使得当 $y,z\in N_\delta(x)$时$\|F(z)-F(y)-F'(y)\cdot(z-y)\|\leqslant\varepsilon\|z-y\|$成立。

用定理 1 可证明算法 1 的全局收敛性。

定理 1　假设$\{x_k\}$是一序列且使 $F(x_k)\to 0$，当 $k\to\infty$时，对每一 k，有

$$\|F(x_k)+F'(x_k)s_k\|\leqslant\eta\|F(x_k)\|$$

及

$$\|F(x_{k-1})\|\leqslant\|F(x_l(k))\|$$

式中

$$\|F(x_l(k))\|=\max_{0\leqslant j\leqslant m(k)}\{\|F(x_{k-j})\|\}\tag{3}$$

$$\|s_k\|\leqslant c\|F(x_k)\|\tag{4}$$

其中 $c>0$，为一常数，$s_k=x_{k+1}-x_k$；η 是不依赖于 k 的常数。如果 x_* 是子序列$\{x_{l(k)}\}$的极限点且 $F'(x_*)$可逆，则 $F(x_*)=0$ 且 $x_k\to x_*$。

证明　首先证明$\{\|F(x_{l(k)}\|\}$是非增的。事实上，考虑 $m(k+1)\leqslant m(k)+1$，有

$$\|F(x_{l(k+1)})\|=\max_{0\leqslant j\leqslant m(k+1)}\{\|F(x_{k+1-j})\|\}\leqslant\max_{0\leqslant j\leqslant m(k)+1}\{\|F(x_{k+1-j})\|\}=$$
$$\max\{\|F(x_{l(k)})\|,\|F(x_{k+1})\|\}=\|F(x_{l(k)})\|$$

即

$$\|F(x_{l(k-1)}\|\leqslant\|F(x_{l(k)})\|\tag{5}$$

另外，从式(3)可得：对于 $K>M$，有

$$\|F(x_{l(k)})\|\leqslant\max_{0\leqslant j\leqslant m(l(k)-1)}\{\|F(x_{l(k)-1-j})\|\}=\|F(x_{l(l(k)-1)})\|$$

即

$$\|F(x_{l(k)})\|\leqslant\|F(x_{l(l(k)-1)})\|$$

由假设，显然有 $F(x_*)=0$，设 $K=\|F'(x_*)^{-1}\|$，δ 足够小，使 $F'(y)^{-1}$存在且

$$\|F'(y)^{-1}\|\leqslant 2K$$
$$\|F(y)-F(x_*)-F'(x_*)(y-x_*)\|\leqslant\|y-x_*\|/2K$$

当 $y\in N_\delta(x_*)$时，这里 δ 的存在性由引理 1 和引理 2 保证。

若 $y\in N_\delta(x_*)$则

$$\|F(y)\|\geqslant\|F'(x_*)(y-x_*)\|-\|F(y)-F(x_*)-F'(x_*)(y-x_*)\|\geqslant$$
$$\|y-x_*\|/\|F'(x_*)^{-1}\|-\|y-x_*\|/2K=\|y-x_*\|/2K$$

因此，当 $y\in N_\delta(x_*)$时

$$\|y-x_*\|\leqslant 2K\|F(y)\|\tag{6}$$

设给定 $\varepsilon\in(0,\delta/4(M+2))$，因为 x_*是$\{x_{l(k)}\}$的一个极限点且 $F(x_*)=0$，所以存在一个充分大的 k，使得

$$x_{l(k)}\in S_\varepsilon\equiv\{y\mid\|y-x_*\|<\delta/2\text{ 且}\|F(y)\|<\varepsilon/[(1+K)(1+c)(1+\eta)]\}$$

则由式(4)有

$$\|s_{l(k)}\|\leqslant c\|F(x_{l(k)})\|\leqslant c\varepsilon/[(1+K)(1+c)(1+\eta)]<\varepsilon<\delta/2$$

因此

$$\|x_{l(k)}-x_*\|\leqslant\|x_{l(k)}-x_*\|+\|s_{l(k)}\|<\delta/2+\delta/2=\delta$$

又由式(5)有

$$\|F(x_{l(k-1)})\|\leqslant\|F(x_{l(k)})\|<\varepsilon/[(1+K)(1+c)(1+\eta)]$$

从而由式(6)有

$$\|x_{l(k+1)}-x_*\|\leqslant 2K\|F(x_{l(k+1)})\|<2K\epsilon/[(1+\eta)(1+c)(1+K)]<\delta/2(M+2)<\delta/2$$

这表明 $x_{l(k+1)}\in S_\epsilon$。

令 $\hat{l}(k)=l(k+M+2)$。证明对所有充分大的 k 有$\|x_{k+1}-x_*\|<\delta/2$，即 $x_{k+1}\in S_\epsilon$。

首先用归纳法证明。对任何给定的 $j\geqslant 0$，有

$$\|s_{\hat{l}(k)-j}\|<\delta/2(M+2) \tag{7}$$

(这里不失一般性，假定迭代下标 k 充分大，能保证下标是正的，即 $k\geqslant j$)

若 $j=0$，因为$\{\hat{l}(k)\}\subset\{l(k)\}$，所以式(7)能够成立。

假设式(7)对所给定的 j 都成立，则由式(4)知

$$\|s_{\hat{l}(k)-j-1}\|\leqslant c\|F(x_{\hat{l}(k)-j-1})\|\leqslant c\|F(x_{l(\hat{l}(k)-j-2)})\|<c\epsilon/[(1+K)(1+c)(1+\eta)]<\epsilon<\delta/2(M+2)$$

故对所有给定的 $j\geqslant 0$，式(7)都成立。

现在对任何充分大的 k 有

$$x_{k+1}=x_{\hat{l}(k)}-\sum_{j=1}^{\hat{l}(k)-k-1}(s_{\hat{l}(k)-j})$$

$$\|x_{k+1}-x_*\|\leqslant\|x_{\hat{l}(k)}-x_*\|+\sum_{j=1}^{\hat{l}(k)-k-1}\|s_{\hat{l}(k)-j}\|<\delta/2(M+2)+\delta(\hat{l}(k)-k-1)/2(M+2)$$

由 $l(k)\leqslant k$ 得

$$\hat{l}(k)-k-1=l(k+M+2)-k-1\leqslant k+M+2-k-1\leqslant M+1$$

所以$\|x_{k+1}-x_*\|<\delta(M+1+1)/2(M+2)=\delta/2$。这表明 $x_{k+1}\in S_\epsilon$。

因此对所有充分大的 k 有

$$x_k\in S_\epsilon\subseteq N_\delta(x_*)$$

且由于$\|F(x_k)\|\to 0$ 和式(6)知 $x_k\to x_*$。

定理 2 给出了算法 2 收敛的条件。

定理 2 假定算法 1 适定，若 $\sum_{k\in T}(1-\eta_k)$发散，则 $F(x_k)\to 0$，其中 $T=\{l(k),k=0,1,2,\cdots\}$。另外，若 x_*是$\{x_k\}$的一个极限点并使得 $F'(x_*)$可逆，则 $F(x_*)=0$ 且 $x_k\to x_*$。

证明 由式(1)和(5)有

$$\|F(x_{k+1})\|\leqslant[1-t(1-\eta_k)]\|F(x_{l(k)})\|\leqslant\|F(x_0)\|\prod_{0\leqslant j\leqslant k\in T}[1-t(1-\eta_j)]\leqslant\|F(x_0)\|\exp[-t\sum_{\substack{0\leqslant j\leqslant k\\ k\in T}}(1-\eta_j)$$

因为 $t>0$ 及 $1-\eta_k>0$，$\sum_{k\in T}(1-\eta_k)$发散，所以有 $F(x_{k+1})\to 0$，$k\to\infty$。令定理 1 中 $\eta=1$，定理 2 的剩余部分可得证。

下面证明算法 2 的全局收敛性。

定理 3 假定算法 2 适定，若 $\sum_{j>0}t_j(1-\eta_j)$发散，则 $k\to\infty$时 $F(x_k)\to 0$。

证明　由式（2）有

$$
\begin{aligned}
\|F(x_k)\| &\leqslant [1-t_k(1-\eta_k)]\|F(x_{k-1})\|\leqslant \\
&\|F(x_0)\|\prod_{0\leqslant j\leqslant k}[1-t_j(1-\eta_j)]\leqslant \\
&\|F(x_0)\|\exp[-\sum_{0\leqslant j\leqslant k}t_j(1-\eta_j)]
\end{aligned}
$$

由于 $\sum_{j>0}t_j(1-\eta_j)$ 发散，因此当 $k\to\infty$ 时，有 $F(x_k)\to 0$。

参 考 文 献

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

2 Eisenstat S C, Homer W F. Globally convergent inexact Newton methods. SIAM J Optimization, 1994, 4 (2): 303～422

3 Ypma T J. Local convergency of inexact Newton methods. SIAM J Numer Anal, 1984, 21(3): 583～590