求解单调非线性方程组的一种全局收敛的梯度型算法

杨　芳[1]，周伟军[2]

（1. 湖南广播电视大学 网络工程职业学院，湖南 长沙 410004；2. 长沙理工大学 数学系，湖南 长沙 410004）

摘　要：提出了一种求解单调非线性方程组的梯度型算法，在适当条件下，证明了该方法具有全局收敛性。通过实例与牛顿型算法进行比较，结果表明：该方法结构简单，适合求解大型问题。

关键词：单调方程组；梯度型算法；全局收敛；投影方法

中图分类号：O221.1　　**文献标志码**：A　　**文章编号**：1673-9833(2011)06-0015-03

A Global Convergent Gradient Algorithm for Monotone Nonlinear Equations

Yang Fang[1]，Zhou Weijun[2]

（1. Network Engineering Vocational School，Hunan Radio and TV University，Changsha 410004，China；
2. Department of Mathematics，Changsha University of Science and Technology，Changsha 410004，China）

Abstract：Presents a gradient algorithm for solving monotone nonlinear equations. Proves that the method has global convergence property under suitable conditions. Comparing with Newton type algorithm, numerical results show that the proposed method is simple and suitable for solving large scale problems.

Keywords：monotone equations; gradient algorithm; global convergent; projection method

0　引言

牛顿型算法是求解非线性方程组

$$F(x)=0, \tag{1}$$

的重要方法[1-3]，其中F：$\mathbf{R}^n\to\mathbf{R}^n$是连续且单调的映射，即

$$\langle F(x)-F(y),x-y\rangle\geqslant 0,\ \forall x,y\in\mathbf{R}^n。$$

对给定点$\boldsymbol{x}$和搜索方向$\boldsymbol{d}$，按$\boldsymbol{x}'=\boldsymbol{x}+\boldsymbol{d}$生成下一次迭代，且$\boldsymbol{d}$是线性方程组$\nabla F(\boldsymbol{x})\boldsymbol{d}=-F(\boldsymbol{x})$的解，其中$\nabla F(\boldsymbol{x})$是$F$在$\boldsymbol{x}$处的Jacobian矩阵。

如果初始点足够好，$\nabla F(\bar{x})$非奇异且在$F(\boldsymbol{x})=0$的点$\bar{\boldsymbol{x}}$附近满足Lipschitz条件，则牛顿法具有局部的超线性甚至二阶收敛速度，这种收敛的结果对一般的半光滑方程组也成立[4-6]。为了扩大牛顿法的收敛域，需要某种全局化策略，最一般的全局化方法是阻尼牛顿法$\boldsymbol{x}'=\boldsymbol{x}+\alpha\boldsymbol{d}$，其中$\alpha>0$由某种线性搜索得到，例如Armijo-type线性搜索。

尽管牛顿法收敛速度快，但每次迭代都需要计算和存储矩阵，这不适合求解大型问题。基于Solodov和Svaiter超平面投影的思想，本文提出一种梯度型算法，它只需利用F的值，而不需利用其Jacobian矩阵，而且在适当的条件下，该方法具有全局收敛性。

收稿日期：2011-09-17
基金项目：国家自然科学基金资助项目（10901026）
作者简介：杨　芳（1967-），女，湖南桑植人，湖南广播电视大学副教授，主要从事应用数学的教学与研究，E-mail：465804345@qq.com
通信作者：周伟军（1977-），男，湖南隆回人，长沙理工大学副教授，博士，主要研究方向为最优化理论和算法，E-mail：weijunzhou@126.com

1 梯度型算法及其收敛性分析

1.1 算法

Step 0 给出初始点$\boldsymbol{x}_0 \in \mathbf{R}^n$，常数$\beta \in (0,1)$，$\rho \in \left(0,\frac{1}{2}\right), \alpha > 0$，置$k$=0。

Step 1 按式（2）计算$\boldsymbol{d}_k \in \mathbf{R}^n$：

$$\boldsymbol{d}_k = \begin{cases} -F(\boldsymbol{x}_k),\ k=0; \\ -F(\boldsymbol{x}_k)+\beta_k \boldsymbol{d}_{k-1},\ k \geqslant 0; \end{cases} \tag{2}$$

其中$\beta_k = -\rho \frac{\langle F(\boldsymbol{x}_k), \boldsymbol{d}_{k-1} \rangle}{\|\boldsymbol{d}_{k-1}\|}$。如果$\boldsymbol{d}_k=\mathbf{0}$，则算法停止。

Step 2 求$\boldsymbol{y}_k=\boldsymbol{x}_k+\alpha_k \boldsymbol{d}_k$，其中$\alpha_k=\beta^{m_k}$，$m_k$是满足式（3）的最小的非负整数$m$，

$$-\langle F(\boldsymbol{x}_k+\beta^m \boldsymbol{d}_k), \boldsymbol{d}_k \rangle \geqslant \alpha \beta^m \|\boldsymbol{d}_k\|^2。 \tag{3}$$

Step 3 计算

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \frac{\langle F(\boldsymbol{y}_k), \boldsymbol{x}_k - \boldsymbol{y}_k \rangle}{\|F(\boldsymbol{y}_k)\|^2} F(\boldsymbol{y}_k), \tag{4}$$

置k=k+1；转Step1。

1.2 算法的收敛性

引理 1 假设F是单调的，且$\boldsymbol{x}, \boldsymbol{y} \in \mathbf{R}^n$使得$\langle F(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y} \rangle > 0$，令$\boldsymbol{x}^+ = \boldsymbol{x} - \frac{\langle F(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y} \rangle}{\|F(\boldsymbol{y})\|^2} F(\boldsymbol{y})$，则对任意满足$F(\bar{\boldsymbol{x}})$=0 的$\bar{\boldsymbol{x}} \in \mathbf{R}^n$，有

$$\|\boldsymbol{x}^+ - \bar{\boldsymbol{x}}\|^2 \leqslant \|\boldsymbol{x} - \bar{\boldsymbol{x}}\|^2 - \|\boldsymbol{x}^+ - \boldsymbol{x}\|^2。$$

本文假设方程组（1）的解集非空。

定理 1 设F单调且连续，对任意满足$F(\bar{\boldsymbol{x}})$=0 的$\bar{\boldsymbol{x}}$，总有$\|\boldsymbol{x}_{k+1} - \bar{\boldsymbol{x}}\|^2 \leqslant \|\boldsymbol{x}_k - \bar{\boldsymbol{x}}\|^2 - \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2$；$\{\boldsymbol{x}_k\}$有界，$\lim\limits_{k\to\infty}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\| = 0$，且$\{\boldsymbol{x}_k\}$收敛到$F(\boldsymbol{x})$=0 的某个解$\bar{\boldsymbol{x}}$。其中$\{\boldsymbol{x}_k\}$是由1.1节中算法产生的序列。

证 如果算法在第k次迭代终止，则由算法有$\boldsymbol{d}_k=\mathbf{0}$。

由式（2）知，当$k \geqslant 1$时，$\boldsymbol{d}_k=-F(\boldsymbol{x}_k)+\beta_k \boldsymbol{d}_{k-1}$，有$F(\boldsymbol{x}_k)=\beta_k \boldsymbol{d}_{k-1}$，又因为$\beta_k = -\rho \frac{\langle F(\boldsymbol{x}_k), \boldsymbol{d}_{k-1} \rangle}{\|\boldsymbol{d}_{k-1}\|^2}$，因此

$$\langle F(\boldsymbol{x}_k), \boldsymbol{d}_{k-1} \rangle = \beta_k \|\boldsymbol{d}_{k-1}\|^2 = -\rho \langle F(\boldsymbol{x}_k), \boldsymbol{d}_{k-1} \rangle,$$

从而$\langle F(\boldsymbol{x}_k), \boldsymbol{d}_{k-1} \rangle = \beta_k = 0$。

再由式（2）有$F(\boldsymbol{x}_k)$=0，这表明$\boldsymbol{x}_k$就是$F(\boldsymbol{x})$=0 的一个解。

至此，不妨假设对任意的$k \neq 0$，$\boldsymbol{d}_k \neq \mathbf{0}$。

欲证1.1节中算法有定义，只需证明式（3）有定义。事实上，如果式（3）不成立，则对某个固定的k和任意的非负整数m，有

$$-\langle F(\boldsymbol{x}_k+\beta^m \boldsymbol{d}_k), \boldsymbol{d}_k \rangle = -\langle F(\boldsymbol{x}_k+\beta^m \boldsymbol{d}_k), -F(\boldsymbol{x}_k) \rangle + \rho \frac{\langle F(\boldsymbol{x}_k), \boldsymbol{d}_{k-1} \rangle}{\|\boldsymbol{d}_{k-1}\|^2} \langle F(\boldsymbol{x}_k+\beta^m \boldsymbol{d}_k), \boldsymbol{d}_{k-1} \rangle < \alpha \beta^m \|\boldsymbol{d}_k\|^2。$$

又由F的连续性，令$m \to \infty$，有

$$\|F(\boldsymbol{y}_k)\|^2 + \rho \frac{\langle F(\boldsymbol{x}_k), \boldsymbol{d}_{k-1} \rangle^2}{\|\boldsymbol{d}_{k-1}\|^2} \leqslant 0,$$

这就导致了矛盾，因而算法是有定义的。

由式（3）有

$$\langle F(\boldsymbol{y}_k), \boldsymbol{x}_k - \boldsymbol{y}_k \rangle = -\alpha_k \langle F(\boldsymbol{y}_k), \boldsymbol{d}_k \rangle \geqslant \alpha \alpha_k^2 \|\boldsymbol{d}_k\|^2 > 0。 \tag{5}$$

设$\bar{\boldsymbol{x}}$是任意满足$F(\bar{\boldsymbol{x}})$=0 的点，由式（4）和式（5）及引理1，有

$$\|\boldsymbol{x}_{k+1} - \bar{\boldsymbol{x}}\|^2 \leqslant \|\boldsymbol{x}_k - \bar{\boldsymbol{x}}\|^2 - \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2。 \tag{6}$$

因此，序列$\{\|\boldsymbol{x}_k - \bar{\boldsymbol{x}}\|\}$单调递减且收敛，从而$\{\boldsymbol{x}_k\}$是有界的，并且

$$\lim_{k\to\infty}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\| = 0。 \tag{7}$$

因为$\{\boldsymbol{x}_k\}$是有界的，由F的连续性可知$\{F(\boldsymbol{x}_k)\}$也是有界的。

由式（2）得

$$\|\boldsymbol{d}_k\| \leqslant \|F(\boldsymbol{x}_k)\| + |\beta_k| \|\boldsymbol{d}_{k-1}\| \leqslant \|F(\boldsymbol{x}_k)\| + \rho \|F(\boldsymbol{x}_k)\| = (1+\rho)\|F(\boldsymbol{x}_k)\|,$$

这表明$\{\boldsymbol{d}_k\}$有界，因此$\{\boldsymbol{y}_k\}$也有界。

由F的连续性及式（4）和式（5）可知，存在常数M>0 使$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\| = \frac{\langle F(\boldsymbol{y}_k), \boldsymbol{x}_k - \boldsymbol{y}_k \rangle}{\|F(\boldsymbol{y}_k)\|} \geqslant M \alpha_\alpha^2 \|\boldsymbol{d}_k\|^2$。

再结合式（7）有

$$\lim \alpha_k \|\boldsymbol{d}_k\| = 0。 \tag{8}$$

假设$\liminf\limits_{k\to\infty}\|F(\boldsymbol{x}_k)\| > 0$，则由式（2）得

$$\|\boldsymbol{d}_k\| \geqslant \|F(\boldsymbol{x}_k)\| - \rho \|F(\boldsymbol{x}_k)\| = (1-\rho)\|F(\boldsymbol{x}_k)\|,$$

因此，$\liminf\limits_{k\to\infty}\|\boldsymbol{d}_k\| > 0$。

再由式（8）有$\lim\limits_{k\to\infty}\alpha_k = 0$，这表明$m_k \to \infty$。

由步长准则知，对β^{m_k-1}式（3）不成立，即

$$-\langle F(\boldsymbol{x}_k+\beta^{m_k-1} \boldsymbol{d}_k), \boldsymbol{d}_k \rangle < \alpha \beta^{m_k-1} \|\boldsymbol{d}_k\|^2。$$

由于$\{\boldsymbol{x}_k\}, \{\boldsymbol{d}_k\}$都有界，选取适当的子列，并令$k \to \infty$，可得$-\langle F(\bar{\boldsymbol{x}}), \bar{\boldsymbol{d}} \rangle \leqslant 0$，其中$\bar{\boldsymbol{x}}, \bar{\boldsymbol{d}}$是相应子列的极限。

由式（2）有

$$-\langle F(\boldsymbol{x}_k),\boldsymbol{d}_k\rangle=\langle \boldsymbol{d}_k,\boldsymbol{d}_k\rangle-\beta_k\langle \boldsymbol{d}_k,\boldsymbol{d}_{k-1}\rangle\geqslant$$
$$\|\boldsymbol{d}_k\|^2-\rho\|F(\boldsymbol{x}_k)\|\|\boldsymbol{d}_k\|,$$

再从$\{\boldsymbol{x}_k\},\{\boldsymbol{d}_k\}$中选取适当的子列，并令$k\to\infty$，有$\|\bar{\boldsymbol{d}}\|^2-\rho\|F(\bar{\boldsymbol{x}})\|\|\bar{\boldsymbol{d}}\|\leqslant 0$，即

$$\|\bar{\boldsymbol{d}}\|\leqslant\rho\|F(\bar{\boldsymbol{x}})\|。\tag{9}$$

又由式（2）有

$$\|\boldsymbol{d}_k\|\geqslant\|F(\boldsymbol{x}_k)\|-|\beta_k|\|\boldsymbol{d}_{k-1}\|\geqslant(1-\rho)\|F(\boldsymbol{x}_k)\|。$$

利用与证明$\|\bar{\boldsymbol{d}}\|\leqslant\rho\|F(\bar{\boldsymbol{x}})\|$相似的方法，可得

$$\|\bar{\boldsymbol{d}}\|\geqslant(1-\rho)\|F(\bar{\boldsymbol{x}})\|。\tag{10}$$

由于$\rho\in\left(0,\frac{1}{2}\right)$，则式（9）与式(10)矛盾，即$\liminf\limits_{k\to\infty}\|F(\boldsymbol{x}_k)\|>0$是不可能的。因此，

$$\liminf_{k\to\infty}\|F(\boldsymbol{x}_k)\|=0。$$

由F的连续性及$\{\boldsymbol{x}_k\}$的有界性，蕴涵$\{\boldsymbol{x}_k\}$有满足$F(\bar{\boldsymbol{x}})=0$的聚点$\bar{\boldsymbol{x}}$。由序列$\{\|\boldsymbol{x}_k-\bar{\boldsymbol{x}}\|\}$收敛，且$\bar{\boldsymbol{x}}$是$\{\boldsymbol{x}_k\}$的一个聚点，因此$\{\boldsymbol{x}_k\}$一定收敛到$\bar{\boldsymbol{x}}$。证毕。

2　梯度型和牛顿型算法的比较

先构造几个单调的非线性方程组问题1~3。

问题1　F: $\mathbf{R}^n\to\mathbf{R}^n$，$F_i(\boldsymbol{x})=2x_i-\sin x_i$，$i=1,2\cdots,n$。

问题2　F: $\mathbf{R}^n\to\mathbf{R}^n$，当$i$为奇数时，$F_i(\boldsymbol{x})=x_{i+1}+x_i-i$；当$i$为偶数时，$F_i(\boldsymbol{x})=x_{i-1}+x_i-i$；$i=1,2\cdots,n$。

问题3　F: $\mathbf{R}^n\to\mathbf{R}^n$，$F_i(\boldsymbol{x})=\sqrt{i}(x_i-i)$，$i=1,2\cdots,n$。

显然，上述3个问题都有解。

将3个问题中n取不同维数，取初始点$\boldsymbol{x}_0=(1,1,\cdots,1)$，以$\|F(\boldsymbol{x}_k)\|<10^{-6}$作为终止准则，参数设置为：$\rho=0.4$，$\beta=0.4$，$\alpha=0.4$。通过用MATLAB编程来实现牛顿型和梯度型2种算法，数值结果见表1~3。

表1　问题1中2种算法结果比较

Table 1　Two algorithms comparison in question 1

n	算法	iter1	iter2	norm(F)	t/s
10	Newton	18	18	$9.976\,8\times10^{-7}$	0.040
	Gradient	14	14	$9.984\,4\times10^{-7}$	0.030
100	Newton	20	32	$9.790\,3\times10^{-7}$	0.651
	Gradient	16	29	$9.821\,6\times10^{-7}$	0.461
1 000	Newton	28	37	$9.894\,2\times10^{-7}$	3.254
	Gradient	23	34	$9.963\,5\times10^{-7}$	2.564
10 000	Newton	47	49	$9.967\,8\times10^{-7}$	4.304
	Gradient	41	39	$9.975\,9\times10^{-7}$	3.044

注：iter1为总迭代次数；iter2为总的函数值次数；t为CPU时间；norm(F)为算法终止时$F(x_k)$的范数。下同。

表2　问题2中2种算法结果比较

Table 2　Two algorithms comparison in question 2

n	算法	iter1	iter2	norm(F)	t/s
10	Newton	28	28	$8.506\,8\times10^{-7}$	0.090
	Gradient	24	24	$8.508\,7\times10^{-7}$	0.050
100	Newton	36	36	$7.905\,0\times10^{-7}$	0.300
	Gradient	29	29	$7.917\,0\times10^{-7}$	0.100
1 000	Newton	41	41	$7.132\,7\times10^{-7}$	0.580
	Gradient	34	34	$7.143\,6\times10^{-7}$	0.240
10 000	Newton	48	48	$6.406\,7\times10^{-7}$	5.571
	Gradient	39	39	$6.408\,5\times10^{-7}$	4.227

表3　问题3中2种算法结果比较

Table 3　Two algorithms comparison in question 3

n	算法	iter1	iter2	norm(F)	t/s
10	Newton	24	47	$4.873\,2\times10^{-7}$	0.120
	Gradient	17	34	$4.884\,1\times10^{-7}$	0.030
100	Newton	49	108	$7.244\,5\times10^{-7}$	0.350
	Gradient	33	99	$7.255\,4\times10^{-7}$	0.170
1 000	Newton	932	4 571	$9.832\,1\times10^{-7}$	7.412
	Gradient	673	3 364	$9.842\,6\times10^{-7}$	4.647
10 000	Newton	2 723	19 376	$9.960\,5\times10^{-7}$	138.396
	Gradient	1 710	10 259	$9.974\,1\times10^{-7}$	93.665

表1~3表明：梯度型算法不仅能求解问题1~3，而且同一个问题在相同维数下，总的迭代次数和总的函数值次数都优于牛顿型算法。

从理论和实例的数值结果表明：本文提出的全局收敛的求解单调非线性方程组的梯度型方法，其结构简单，存储量小，适合求解大型问题。

参考文献：

[1] Ortega J M，Rheinboldt W C. Iterative Solution of Nonlinear Equations in Several Variables[M]. Philadelphia PA USA：Academic Press，1997：356-412.

[2] Dennis J E，Schnable R B. Numerical Methods for Unconstrained Optimization and Nonlinear Equations[M]. Englewood Cliffs N. J.：Prenrice-Hall，1983：276-293.

[3] Bertsekas D P. Nonlinear Programming[D]. Belmont Massachusetts：Athena Scientific，1995.

[4] Qi Liqun. Convergence Analysis of Some Algorithms for Solving Nonsmooth Equations[J]. Mathematics of Operations Research，1993，18(1)：227-244.

[5] Qi Liqun，Sun Jie. A Nonsmooth Version of Newton′s Method[J]. Mathematical Programming，1993，58(1/2/3)：353-367.

[6] Pang J S，Qi L. Nonsmooth Equations: Motivation and Algorithms[J]. SIAM Journal on Optimization，1993(3)：443-465.

（责任编辑：邓光辉）