

文章编号: 1674-8085(2017)03-0001-05

# 求解非线性方程组的一种新共轭梯度法

王博朋, \*袁功林, 胡午杰

(广西大学数学与信息科学学院, 广西, 南宁 530004)

**摘要:** 在求解非线性方程组问题的过程中, 由已知的三项共轭梯度法的基础上设计出了一种新的共轭梯度法 WW, 并在适当条件下证明了其充分下降性及全局收敛性。数值实验结果表明, 在与现有的一些共轭梯度法的对比中, WW 方法有较强的竞争性。

**关键词:** 共轭梯度法; 下降性; 全局收敛性

中图分类号: O224

文献标识码: A

DOI:10.3969/j.issn.1674-8085.2017.03.001

## A NEW CONJUGATE GRADIENT METHOD FOR SOLVING NONLINEAR EQUATIONS

WANG Bo-peng, \*YUAN Gong-lin, HU Wu-jie

(College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi 530004, China)

**Abstract:** In the process of solving nonlinear equations, a new conjugate gradient method is derived from the known DY conjugate gradient method. Furthermore, the sufficient descent property and global convergence are proved under suitable conditions. The numerical results also show that there is a strong competitive advantage in comparison with some existing conjugate gradient methods.

**Key words:** gradient algorithm; descent property; global convergence

### 0 引言

非线性方程组的公式可表示为:

$$g(x) = 0, x \in \mathbb{R}^n \quad (1)$$

其中  $g(x) = 0$  连续可微且单调。下令

$F(x) = \frac{1}{2} \|g(x)\|^2$ , 其中  $\|\bullet\|$  为欧几里得范数。那么,

式 (1) 中的问题等价于以下问题的全局最优:

$$\min F(x), x \in \mathbb{R}^n \quad (2)$$

求解非线性方程组的一般迭代公式为:

$$x_{k+1} = x_k + \alpha_k d_k$$

其中  $d_k$ : 搜索方向,  $\alpha_k$ : 沿搜索方向上的步长。目前求解  $d_k$  及  $\alpha_k$  的方法有许多种<sup>[1-5]</sup>, 如标准的 Wolf 线搜索:

$$F(x_k + \alpha_k d_k) - F(x_k) \leq \delta \alpha_k g_k^T d_k$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k$$

其中  $0 < \delta < \sigma < 1$ 。本文采用 Qi 和 Li<sup>[6]</sup>提出的一种新的线搜索

$$-g(x_k + \alpha_k d_k)^T d_k \geq \sigma \alpha_k \|g(x_k + \alpha_k d_k)\| \|d_k\| \quad (3)$$

其中  $\alpha_k = \max\{s, \rho s, \rho^2 s, \dots\}$ ,  $\sigma, s > 0, \rho \in (0, 1)$ 。此方法在一定条件下有全局收敛性及超线性收敛等性质, 比一般线搜索更具竞争性。

Zhang<sup>[7]</sup>提出了一种三项共轭梯度算法:

收稿日期: 2017-03-11; 修改日期: 2017-04-02

基金项目: 国家自然科学基金项目(11261006); 广西杰出青年科学基金项目(2015GXNSFGA139001)

作者简介: 王博朋(1991-), 男, 河南周口人, 硕士生, 主要从事最优化理论与方法研究(Email:15277112359@163.com);

\*袁功林(1976-), 男, 河南商丘人, 教授, 博士, 博士生导师, 主要从事最优化理论与方法研究(E-mail:yuangl0417@126.com);

胡午杰(1993-), 男, 安徽安庆人, 硕士生, 主要从事最优化理论与方法研究(Email:1094782552@qq.com).

$$d_k = \begin{cases} -g_k + \beta_k^{PRP} d_{k-1} - v_k y_k & \text{if } k \geq 1 \\ -g_k & \text{if } k = 0 \end{cases}$$

其中:  $g_k = g(x_k)$ ,  $g_{k-1} = g(x_{k-1})$ ,  $\beta_k^{PRP} = \frac{g_{k-1}^T y_k}{\|g_{k-1}\|^2}$ ,

$y_k = g_k - g_{k-1}$ ,  $v_k = \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2}$ 。如果  $g$  是映射

$F: \mathfrak{R}^n \rightarrow \mathfrak{R}$  的梯度, 不难得出:

$$d_k^T g_k = -\|g_k\|^2$$

笔者根据此种方法提出另一种三项共轭梯度算法:

$$d_k = \begin{cases} -g_k + \frac{g_k^T y_k d_{k-1} - g_k^T d_{k-1} y_k^*}{\mu \|d_{k-1}\| \|y_k^*\| + v \|y_k^*\|^2 + \|g_{k-1}\|^2 + \eta \|g_{k-1}\| \|d_{k-1}\|} & \text{if } k \geq 1 \\ -g_k & \text{if } k = 0 \end{cases} \quad (4)$$

其中  $\mu > 0, v > 0, \eta > 0, \lambda > 0$ ,  $y_k^* = g_k - \frac{\|g_{k-1}\|}{\|g_k\|} g_{k-1}$ 。

## 1 算法

算法 WW 步骤如下:

Step 0: 给定初始点  $x_0 \in \mathfrak{R}$ , 常数  $\mu > 0, v > 0,$

$\eta > 0, \lambda > 0; \rho, \varepsilon \in (0, 1)$ 。令  $k := 1$ ;

Step 1: 若  $\|g(x)\| \leq \varepsilon$ , 停止; 否则转 step 2;

Step 2: 通过式 (4) 计算搜索方向  $d_k$ ;

Step 3: 选出满足条件 (3) 的步长  $\alpha_k$ ;

Step 4: 令迭代公式为  $z_k = x_k + \alpha_k d_k$ ;

Step 5: 若  $\|g(x)\| \leq \varepsilon$ , 停止, 令  $x_{k+1} = z_k$ ; 否则令

$$x_{k+1} = x_k - \frac{g(z_k)^T (x_k - z_k)}{\|g(z_k)\|^2} g(z_k) \quad (5)$$

Step 6: 令  $k := k + 1$ , 转 Step 1。

备注: 式 (5) 出自文献[8], 它是  $x_k$  在超平面  $H_k = \{x \in \mathfrak{R}^n | \langle g(z_k), x - z_k \rangle = 0\}$  上的投影。

## 2 充分下降性及全局收敛性

先给出两个假设条件:

假设 1 问题 (1) 的解集不是空集。

假设 2  $g(x)$  在  $\mathfrak{R}^n$  上 Lipschitz 连续, 那么存在  $L > 0$ , 使得

$$\|g(x) - g(y)\| \leq L \|x - y\|, \quad \forall x, y \in \mathfrak{R}^n \quad (6)$$

由此得出

$$\|g_k\| \leq \varsigma$$

其中  $\varsigma > 0$ 。

引理 1  $d_k$  由式 (4) 给出, 则有:

$$g_{k+1} d_{k+1} \leq -\|g_{k+1}\|^2 \quad (7)$$

和

$$\|g_k\| \leq \|d_k\| \leq (1 + \frac{2}{\mu}) \|g_k\| \quad (8)$$

成立。

证明: 式 (7) 的证明显而易见, 式 (7) 可直接推出式 (8) 的左半部分, 下面给出式 (8) 的右半部分的证明。

$$\begin{aligned} \|d_k\| &= \left\| -g_k + \frac{g_k^T y_k d_{k-1} - g_k^T d_{k-1} y_k^*}{\mu \|d_{k-1}\| \|y_k^*\| + v \|y_k^*\|^2 + \|g_{k-1}\|^2 + \eta \|g_{k-1}\| \|d_{k-1}\|} \right\| \\ &\leq \|g_k\| + \left\| \frac{g_k^T y_k d_{k-1} - g_k^T d_{k-1} y_k^*}{\mu \|d_{k-1}\| \|y_k^*\| + v \|y_k^*\|^2 + \|g_{k-1}\|^2 + \eta \|g_{k-1}\| \|d_{k-1}\|} \right\| \\ &\leq \|g_k\| + \frac{2 \|g_k^T y_k d_{k-1}\|}{\mu \|d_{k-1}\| \|y_k^*\|} = (1 + \frac{2}{\mu}) \|g_k\| \end{aligned}$$

式 (8) 的右半部分得证。综上, 式 (8) 成立。

引理 2 假设 1, 假设 2 均成立, 且  $\{x_k\}$  和  $\{z_k\}$  是由算法 WW 产生的, 可得:

$$\alpha_k \geq \min \left\{ s, \frac{\rho}{L + \sigma \|g(z_k)\|} \frac{\|g(x_k)\|^2}{\|d_k\|^2} \right\}$$

其中  $z_k' = x_k + \alpha_k' d_k$ ,  $\alpha_k' = \alpha_k \rho^{-1}$ 。

证明:  $\alpha_k$  是由式 (3) 给出的, 若  $\alpha_k \neq s$ , 则  $\alpha_k' = \alpha_k \rho^{-1}$  不满足式 (3), 即有:

$$-g(x_k + \alpha_k d_k)^T d_k < \sigma \alpha_k \|g(x_k + \alpha_k d_k)\| \|d_k\|^2$$

成立。

由 Lipschitz 条件及式 (7) 可得:

$$\begin{aligned} \|g_k\|^2 &\leq -g_k^T d_k \leq (g(z_k') - g(x_k))^T d_k + \\ &\quad \sigma \alpha_k' \|g(z_k')\| \|d_k\|^2 \leq \alpha_k' (L + \sigma \|g(z_k')\|) \|d_k\|^2 \end{aligned}$$

即

$$\alpha_k \geq \frac{c_0 \rho}{L + \sigma \|g(z'_k)\|} \frac{\|g(x_k)\|^2}{\|d_k\|^2}$$

得证。

**引理 3** 假设 1, 假设 2 都成立, 且  $\{x_k\}$  是由算法 WW 产生的, 假设  $x_*$  是问题 (1) 的解, 即满足  $g(x_*)=0$ , 则有:

$$\|x_{k+1} - x_*\|^2 \leq \|x_k - x_*\|^2 - \|x_{k+1} - x_k\|^2 \quad (9)$$

及

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_k\|^2 < \infty \quad (10)$$

成立。

**证明:** 由假设的超平面  $H_k$  和  $g$  的单调性, 有:

$$\langle g(z_k) - g(x_*), x_* - z_k \rangle = \langle g(z_k), x_* - z_k \rangle \leq 0$$

$x_{k+1}$  是  $x_k$  在半平面  $M_k = \{x \in \mathfrak{R}^n \mid \langle g(z_k), x - z_k \rangle \leq 0\}$  上的投影。若  $x_*$  在半平面  $M_k$  上, 由投影算子的性质, 得  $\langle x_k - x_{k+1}, x_{k+1} - x_* \rangle \geq 0$ 。那么有:

$$\begin{aligned} \|x_k - x_*\|^2 &= \|x_k - x_{k+1}\|^2 + \|x_{k+1} - x_*\|^2 + \\ &2 \langle x_k - x_{k+1}, x_{k+1} - x_* \rangle \geq \|x_k - x_{k+1}\|^2 + \|x_{k+1} - x_*\|^2 \end{aligned}$$

显然式 (9) 成立。整理得:

$$\|x_{k+1} - x_k\|^2 \leq \|x_k - x_*\|^2 - \|x_{k+1} - x_*\|^2$$

累加并令  $k \rightarrow +\infty$ , 得式 (10) 成立, 得证。

由式 (3) 的线搜索方法及投影 (5), 得:

$$\begin{aligned} \|x_{k+1} - x_k\| &= \frac{|g(z_k)^T (x_k - z_k)|}{\|g(z_k)\|} = \\ &\frac{-\alpha_k g(z_k)^T d_k}{\|g(z_k)\|} \geq \sigma \alpha_k^2 \|d_k\|^2 \end{aligned}$$

又由式 (10) 知:

$$\|x_{k+1} - x_k\| \rightarrow 0, \quad k \rightarrow +\infty$$

综上, 可得:

$$\lim_{k \rightarrow +\infty} \alpha_k d_k = 0 \quad (11)$$

**定理 1** 假设 1, 假设 2 成立且  $\{\alpha_k, d_k, x_{k+1}, g_{k+1}\}$

由算法 WW 产生, 那么,

$$\liminf_{k \rightarrow +\infty} \|g_k\| = 0 \quad (12)$$

成立。

**证明:** 假设式 (11) 不成立, 则存在  $\tau > 0$  使得对任意  $K \geq 0$  都有  $\|g_k\| \geq \tau$ 。结合式 (8), 有:

$$\tau c \leq c \|g_k\| \leq \|d_k\| \leq c_1 \|g_k\| \leq c_1 \zeta$$

其中  $c_1 = (1 + \frac{2}{\mu})$ 。从引理 1 中, 有:

$$\begin{aligned} \alpha_k \|d_k\| &\geq \min\{\beta, \frac{\rho}{L + \sigma \|g(z'_k)\|} \frac{\|g(x_k)\|^2}{\|d_k\|^2}\} \|d_k\| \geq \\ &\min\{\tau c \beta, \frac{\rho \tau}{c_1 \zeta (L + \sigma \xi)}\} > 0 \end{aligned}$$

这与式 (11) 矛盾。所以式 (12) 成立。得证。

### 3 数值实验

本文将此算法与 PRP 算法进行了比较, 测试函数见文献[9], 本文列举了其中的九个函数, 如表 1:

表 1 测试函数

Table 1 Test functions

函数	初始点
1 Exponential function 1	$x_0 = (\frac{n}{n-1}, \frac{n}{n-1}, \dots, \frac{n}{n-1})^T$
2 Exponential function 2	$x_0 = (\frac{1}{n^2}, \frac{1}{n^2}, \dots, \frac{1}{n^2}, \frac{1}{n^2})^T$
3 Trigonometric function	$x_0 = (\frac{101}{100n}, \dots, \frac{101}{100n})^T$
4 Singular function	$x_0 = (1, 1, \dots, 1)^T$
5 Logarithmic function	$x_0 = (1, 1, \dots, 1)^T$
6 Broyden Tridiagonal function	$x_0 = (-1, -1, \dots, -1)^T$
7 Variable dimensioned function	$x_0 = (1 - \frac{1}{n}, 1 - \frac{2}{n}, \dots, \frac{1}{n}, 0)^T$
8 Discrete boundary value problem	$x_0 = [t(t-1), t(2t-1), \dots, t(nt-1)]^T, t = \frac{1}{n+1}$
9 Troesch problem	$x_0 = (0, 0, \dots, 0)^T$

本次数值实验运行在 MATLAB 2010 上, 电脑配置为: CPU 是 Intel Pentium(R)Dual-Core E5800 3.2 GHz; 内存为 2G; 操作系统是 Windows 10。进行数值实验时的参数设置为:  $\mu = 0.0001$ ,  $\nu = 0.0001, \eta = 0.0001, \varepsilon = 10^{-5}$ , 终止条件为  $\|g(x)\| \leq 10^{-5}$ , 其中各元素定义为 NI: 循环次数;

NF:函数值的计算次数; GN: 试验终止时, 梯度值的范数; Cputime:计算所花费的时间。结果参见表 2。

表 2 数值结果  
Table 2 Numerical results

	Dim	算法 WW			算法 PRP		
		NI/NF	GN	Cputime/s	NI/NF	GN	Cputime/s
1	3000	121/122	9.984965e-06	1.218750	129/130	9.969911e-06	1.062500
	5000	102/103	9.951253e-06	1.406250	109/110	9.912454e-06	1.421875
	10000	80/81	9.993445e-06	2.453125	86/87	9.942374e-06	2.281250
2	3000	53/203	9.445093e-06	1.203125	58/220	9.947840e-06	1.125000
	5000	25/101	9.410726e-06	0.828125	24/97	9.754454e-06	0.906250
	10000	100/463	9.550084e-06	7.859375	105/479	9.816376e-06	6.515625
3	3000	43/86	8.561034e-06	0.687500	48/95	8.647042e-06	0.671875
	5000	42/84	8.283174e-06	1.000000	46/91	9.736889e-06	1.046875
	10000	40/80	9.127624e-06	1.843750	44/87	9.984195e-06	1.875000
4	3000	299/658	3.719317e-02	5.703125	299/584	3.218379e-02	4.796875
	5000	299/721	6.924486e-01	9.171875	299/584	2.866803e-02	7.671875
	10000	299/760	5.063178e-01	17.593750	299/637	2.730928e-02	15.250000
5	3000	5/6	1.009985e-08	0.046875	11/12	1.012266e-08	0.078125
	5000	5/6	6.263918e-09	0.093750	11/12	8.539532e-09	0.156250
	10000	5/6	3.618055e-09	0.156250	11/12	7.621607e-09	0.359375
6	3000	95/190	9.498386e-06	1.140625	104/208	9.243312e-06	1.250000
	5000	97/194	9.049256e-06	2.078125	106/212	9.130520e-06	1.937500
	10000	99/198	9.441814e-06	3.703125	108/216	9.878046e-06	3.843750
7	3000	1/2	0.000000e+00	0.031250	1/2	0.000000e+00	0.000000
	5000	1/2	0.000000e+00	0.000000	1/2	0.000000e+00	0.031250
	10000	1/2	0.000000e+00	0.031250	1/2	0.000000e+00	0.031250
8	3000	35/71	9.254818e-06	0.562500	40/80	9.908999e-06	0.593750
	5000	34/69	8.639879e-06	0.859375	39/78	9.198351e-06	0.937500
	10000	32/65	9.299308e-06	1.671875	37/74	9.416687e-06	1.828125
9	3000	0/1	0.000000e+00	0.000000	0/1	0.000000e+00	0.031250
	5000	0/1	0.000000e+00	0.015625	0/1	0.000000e+00	0.000000
	10000	0/1	0.000000e+00	0.000000	0/1	0.000000e+00	0.031250

为了更直观地说明以上观点, 我们引用了文献 [9]的作图方法, 画出了函数值的计算次数比较图 (NF):

从图 1 可以看出, 算法 WW 在计算精度相同的条件下, 计算出结果所需要的迭代次数较少, 且能有效求解。所以此算法在求解非线性单调方程组问题中比 PRP 方法更有效。因此, 此算法是有效的。

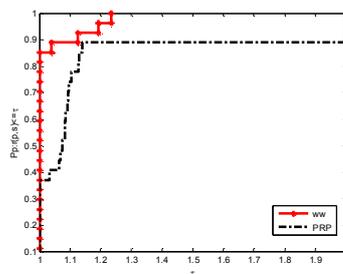


图 1 算法 WW 与算法 PRP 的性能图 (NF)  
Fig.1 Algorithm WW and algorithm PRP performance chart(NF)

从图1中可以看出,算法 WW 能解决大概 85% 的问题,而算法 PRP 只能解决 38%左右的问题;从  $\tau$  值来看,算法 WW 的  $\tau \approx 1.24$ ,算法 PRP 的  $\tau \approx 2$ ,所以算法 WW 的鲁棒性更强。

## 4 结语

笔者设计了一个新的搜索方向来求解非线性单调方程组问题,并在一定的约束条件下证明了其充分下降性和全局收敛性。此算法在与 PRP 算法的数值结果比较中,也显得更有优势。因此,此算法是有效的。

### 参考文献:

- [1] Polyak B T. The conjugate gradient method in extremal problems[J]. *Ussr Computational Mathematics & Mathematical Physics*, 1969, 9(4):94-112.
- [2] Dai Y, Yuan Y. A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property[J]. *Siam Journal on Optimization*, 1999, 10(1):177-182.
- [3] Fletcher R, Reeves. Function minimization by conjugate gradients[J]. *Computer Journal*, 1964, 7(2):149-154.
- [4] Polak E, Ribière G. Note sur la convergence de méthodes de directions conjuguées[J]. *Rev. française Informat. recherche Opérationnelle*, 1968, 16(16):35-43.
- [5] Wei Z, Yao S, Liu L. The convergence properties of some new conjugate gradient methods[J]. *Applied Mathematics & Computation*, 2006, 183(2):1341-1350.
- [6] Yuan G, Lu X. A new backtracking inexact BFGS method for symmetric nonlinear equations[J]. *Computers & Mathematics with Applications*, 2008, 55(1):116-129.
- [7] Zhang L, Zhou W, Li D H. A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence[J]. *IMA Journal of Numerical Analysis*, 2006, 26(4): 629-640.
- [8] Solodov M V, Svaiter B F. A globally convergent inexact Newton method for systems of monotone equations[M]. *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*. Springer US, 1998: 355-369.
- [9] Yuan G, Wei Z, Lu X. A BFGS trust-region method for nonlinear equations[J]. *Computing*, 2011, 92(4):317-333.