Advances in Applied Mathematics 应用数学进展, **2019, 8(5), 869-875**
Published Online May 2019 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2019.85097

A New Three-Term Conjugate Gradient Method for Solving Nonlinear Equations

Ruosha Liao

College of Mathematics and Information Science, Guangxi University, Nanning Guangxi
Email: sai_2015@163.com

Received: Apr. 16th, 2019; accepted: May 7th, 2019; published: May 14th, 2019

Abstract

Based on the existing three-term conjugate gradient method, this paper designs a new conjugate gradient method JG to solve the problem of nonlinear equations, and proves the sufficient descent and global convergence of JG algorithm under certain assumptions. From the results of numerical experiments, we can see that JG algorithm has better properties than PRP algorithm.

Keywords

Conjugate Gradient Method, Sufficient Descent Property, Global Convergence

一种新的三项共轭梯度法求解非线性方程组

廖若沙

广西大学，数学与信息科学学院，广西 南宁
Email: sai_2015@163.com

收稿日期：2019年4月16日；录用日期：2019年5月7日；发布日期：2019年5月14日

摘　要

本文在现有的三项共轭梯度法的基础上，设计了一种新的共轭梯度法JG求解非线性方程组问题，并在一定的假设条件下证明了JG算法的充分下降性和全局收敛性。通过数值实验的结果我们可以看到，JG算法与PRP算法相比具有更好的性质。

关键词

共轭梯度法，充分下降性，全局收敛性

文章引用: 廖若沙. 一种新的三项共轭梯度法求解非线性方程组[J]. 应用数学进展, 2019, 8(5): 869-875.
DOI: 10.12677/aam.2019.85097

1. 引言

我们考虑如下非线性方程组：

$$h(x)=0, x\in R^n \tag{1}$$

其中 $h(x)=0$ 单调并连续可微。令 $F(x)=\frac{1}{2}\|h(x)\|^2$，则(1)式等价于求解(2)式的全局最优解：

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

通常求解非线性方程组的迭代公式为：

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

其中 α_k 为沿搜索方向上的步长，d_k 为搜索方向。

随着共轭梯度法的发展，产生了一系列的求解 d_k 和 α_k 的方法[1] [2] [3] [4] [5]，例如标准的 Wolfe 线搜索：

$$F(x_{k+1})-F(x_k)\le c_1\alpha_k g_k^{\mathrm{T}} d_k$$
$$g_{k+1}^{\mathrm{T}} d_k \ge c_2 g_k^{\mathrm{T}} d_k$$

其中 $0<c_1<c_2<1$ 为任意常数，$g_k := \nabla F(x_k)$。

我们利用 Yuan 和 Lu [6]给出的一种新的线搜索：

$$-g(x_k+\alpha_k d_k)^{\mathrm{T}} d_k \ge \sigma\alpha_k \|g(x_k+\alpha_k d_k)\|\|d_k\|^2 \tag{3}$$

其中 $\alpha_k=\max\{s,\rho s,\rho^2 s,\cdots\}$，$\sigma, s>0$，$\rho\in(0,1)$。在一定的假设条件下，得到全局收敛性和超线性收敛。

Zhang [7]提出了一种三项共轭梯度算法：

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

其中 $g_k=g(x_k)$，$\beta_k^{PRP}=\frac{g_{k-1}^{\mathrm{T}} y_k}{\|g_{k-1}\|^2}$，$v_k=\frac{g_k^{\mathrm{T}} d_{k-1}}{\|g_{k-1}\|^2}$，$y_k=g_k-g_{k-1}$。由 d_k 的定义易得：

$$d_k^{\mathrm{T}} g_k=-\|g_k\|^2$$

根据上述算法及线搜索，本文给出一种新的三项共轭梯度算法 JG：

$$d_k=\begin{cases}-g_k+\dfrac{g_k^{\mathrm{T}} y_k^* d_{k-1}-g_k^{\mathrm{T}} d_{k-1} y_k^*}{\mu\|d_{k-1}\|\|y_k^*\|+v\|y_k^*\|^2+\|g_{k-1}\|^2+\eta\|g_{k-1}\|\|d_{k-1}\|+r\|d_{k-1}\|^2} & \text{if } k\ge 1\\ -g_k & \text{if } k=0\end{cases} \tag{4}$$

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

2. 算法

新的三项共轭梯度算法 JG 的步骤如下：

Step 0：令初始点 $x_0 \in R$， $\mu > 0$， $v > 0$， $\eta > 0$， $r > 0$； $\rho, \varepsilon \in (0,1)$。 $k := 1$；

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

Step 2：通过(4)式计算搜索方向 d_k；

Step 3：选择满足条件(3)的步长 α_k；

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

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

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

Step 6：令 $k := k+1$，转 Step 1。

注：(5)式[8]为 x_k 在超平面 $H_k = \{x \in R^n \mid \langle g(w_k), x - w_k \rangle = 0\}$ 上的投影。

3. 充分下降性和全局收敛性

我们给出如下两个假设条件：

假设 1：(1)式的解集非空。

假设 2： $g(x)$ 在 R^n 上 Lipschitz 连续，即存在 $L > 0$ 使得

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

于是可以得到

$$\|g_k\| \le \varsigma$$

其中 $\varsigma > 0$。

引理 1：(4)式中的 d_k 满足：

$$g_{k+1} d_{k+1} \le -\|g_{k+1}\|^2 \tag{7}$$

和

$$\|g_k\| \le \|d_k\| \le \left(1 + \frac{2}{\mu}\right)\|g_k\| \tag{8}$$

证明：由 d_k 的定义可以直接得(7)式的结果，而利用(7)式可以得到(8)式的左半部分，接下来证明(8)式的右半部分。

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

得证，故(8)式成立。 □

引理 2：若假设 1，假设 2 均成立， $\{x_k\}$ 和 $\{w_k\}$ 是由算法 JG 产生的点列，可得：

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

其中 $\hat{w}_k = x_k + \hat{\alpha}_k d_k$，$\hat{\alpha}_k = \alpha_k \rho^{-1}$。

证明： 这里 α_k 由(3)式给出，如果 $\alpha_k \neq s$，那么 $\hat{\alpha}_k = \alpha_k \rho^{-1}$ 不满足(3)式，即：

$$-g\left(x_k+\alpha_k d_k\right)^{\mathrm{T}} d_k < \sigma\alpha_k\left\|g\left(x_k+\alpha_k d_k\right)\right\|\left\|d_k\right\|^2$$

成立。

由 Lipschitz 条件和(7)式：

$$\left\|g_k\right\|^2 \leq -g_k^{\mathrm{T}} d_k \leq \left(g\left(\hat{w}_k\right)-g\left(x_k\right)\right)^{\mathrm{T}} d_k + \sigma\hat{\alpha}_k\left\|g\left(\hat{w}_k\right)\right\|\left\|d_k\right\|^2 \leq \hat{\alpha}_k\left(L+\sigma\left\|g\left(\hat{w}_k\right)\right\|\right)\left\|d_k\right\|^2$$

则

$$\alpha_k \geq \frac{\rho}{L+\sigma\left\|g\left(\hat{w}_k\right)\right\|}\frac{\left\|g\left(x_k\right)\right\|^2}{\left\|d_k\right\|^2}$$

得证。 □

引理 3： 若假设 1，假设 2 均成立，$\{x_k\}$ 是由算法 JG 产生的点列，若 x_* 是(1)的解，即 $h(x_*)=0$ 成立，那么：

$$\left\|x_{k+1}-x_*\right\|^2 \leq \left\|x_k-x_*\right\|^2 - \left\|x_{k+1}-x_k\right\|^2 \tag{9}$$

和

$$\sum_{k=0}^{\infty}\left\|x_{k+1}-x_k\right\|^2 < \infty \tag{10}$$

均成立。

证明： 由 g 和超平面 H_k 的单调性，可得：

$$\left\langle g\left(w_k\right)-g\left(x_*\right), x_*-w_k\right\rangle = \left\langle g\left(w_k\right), x_*-w_k\right\rangle \leq 0$$

x_{k+1} 为 x_k 在 $M_k = \left\{x\in R^n \mid \left\langle g\left(w_k\right), x-w_k\right\rangle \leq 0\right\}$ 上的投影。如果 x_* 在 M_k 上，则 $\left\langle x_k-x_{k+1}, x_{k+1}-x_*\right\rangle \geq 0$。可得：

$$\left\|x_k-x_*\right\|^2 = \left\|x_k-x_{k+1}\right\|^2 + \left\|x_{k+1}-x_*\right\|^2 + 2\left\langle x_k-x_{k+1}, x_{k+1}-x_*\right\rangle \geq \left\|x_k-x_{k+1}\right\|^2 + \left\|x_{k+1}-x_*\right\|^2$$

整理得：

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

故(9)式成立。令 $k\to+\infty$，对该式进行累加，得(10)式成立，得证。 □

由(3)、(5)二式，得：

$$\left\|x_{k+1}-x_k\right\| = \frac{\left|g\left(w_k\right)^{\mathrm{T}}\left(x_k-w_k\right)\right|}{\left\|g\left(w_k\right)\right\|} = \frac{-\alpha_k g\left(w_k\right)^{\mathrm{T}} d_k}{\left\|g\left(w_k\right)\right\|} \geq \sigma\alpha_k^2\left\|d_k\right\|^2$$

由(10)式可知，当 $k\to+\infty$ 时：

$$\left\|x_{k+1}-x_k\right\| \to 0$$

综上可得：

$$\lim_{k \to +\infty} \alpha_k d_k = 0 \tag{11}$$

定理 1 若假设 1，假设 2 均成立，序列 $\{\alpha_k, d_k, x_{k+1}, g_{k+1}\}$ 由算法 JG 产生，则有

$$\liminf_{k \to \infty} \|g_k\| = 0 \tag{12}$$

证明：若(12)式不成立，则存在 $\delta > 0$ 使得对任意 $K \geq 0$ 都有 $\|g_k\| \geq \delta$。结合式(8)，有：

$$c_3 \delta \leq c_3 \|g_k\| \leq \|d_k\| \leq c_4 \|g_k\| \leq c_4 \zeta$$

这里 $0 < c_3 < 1$，$c_4 = 1 + \dfrac{2}{\mu}$。由引理 2 可得：

$$\begin{aligned} \alpha_k \|d_k\| &\geq \min\left\{ s, \frac{\rho}{L + \sigma \|g(\hat{w}_k)\|} \frac{\|g(x_k)\|^2}{\|d_k\|^2} \right\} \|d_k\| \\ &\geq \min\left\{ c_3 \delta s, \frac{\rho \delta}{c_4 \zeta (L + \sigma \varsigma)} \right\} > 0 \end{aligned}$$

与(11)式矛盾，故(12)式成立，得证。 □

4. 数值实验

下面我们将 JG 算法与 PRP 算法作比较，选取九个测试函数[9]进行测试，结果见表 1：

Table 1. Test functions
表 1. 测试函数

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

数值实验的参数设置：$\mu = 0.0001$，$v = 0.0001$，$\eta = 0.0001$，$r = 0.0001$，$\varepsilon = 10^{-5}$，终止条件为 $\|g(x)\| \leq 10^{-5}$，其中 CI 为循环次数；CT 为计算次数；GN 为梯度值的范数。结果见表 2：

Table 2. Numerical results
表 2. 数值结果

	Dim	算法 JG			算法 PRP		
		CI/CT	GN	Cputimes/s	CI/CT	GN	Cputimes/s
	3000	121/122	9.984983e−06	1.218683	129/130	9.969876e−06	1.062489
1	5000	102/103	9.951562e−06	1.406245	109/110	9.912214e−06	1.421314
	10000	80/81	9.993524e−06	2.453154	86/87	9.942864e−06	2.281245
	3000	53/203	9.445482e−06	1.201245	58/220	9.947542e−06	1.123458
2	5000	25/101	9.413541e−06	0.826587	24/97	9.752145e−06	0.907852
	10000	100/463	9.551023e−06	7.859952	105/479	9.810142e−06	6.513145
	3000	43/86	8.560854e−06	0.686589	48/95	8.646852e−06	0.671754
3	5000	42/84	8.282587e−06	1.000121	46/91	9.738878e−06	1.048721
	10000	40/80	9.126475e−06	1.842445	44/87	9.983354e−06	1.877724
	3000	299/658	3.719134e−02	5.705445	299/584	3.218877e−02	4.793553
4	5000	299/721	6.924588e−01	9.172574	299/584	2.866659e−02	7.671957
	10000	299/760	5.063289e−01	17.592875	299/637	2.730917e−02	15.250000
	3000	5/6	1.009850e−08	0.046765	11/12	1.012350e−08	0.078226
5	5000	5/6	6.263368e−09	0.093875	11/12	8.539635e−09	0.156789
	10000	5/6	3.618125e−09	0.156364	11/12	7.621875e−09	0.359330
	3000	95/190	9.498546e−06	1.140075	104/208	9.243558e−06	1.250000
6	5000	97/194	9.049356e−06	2.078486	106/212	9.130544e−06	1.936852
	10000	99/198	9.441758e−06	3.703354	108/216	9.878274e−06	3.843549
	3000	1/2	0.000000e+00	0.031250	1/2	0.000000e+00	0.000000
7	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
	3000	35/71	9.252585e−06	0.564574	40/80	9.908979e−06	0.593585
8	5000	34/69	8.639254e−06	0.859358	39/78	9.198357e−06	0.938896
	10000	32/65	9.299685e−06	1.671241	37/74	9.416458e−06	1.822575
	3000	0/1	0.000000e+00	0.000000	0/1	0.000000e+00	0.031250
9	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

我们根据计算次数(CT)给出图 1，可以直观看到在相同条件下，算法 JG 所需迭代次数更少，图 1 如下所示：

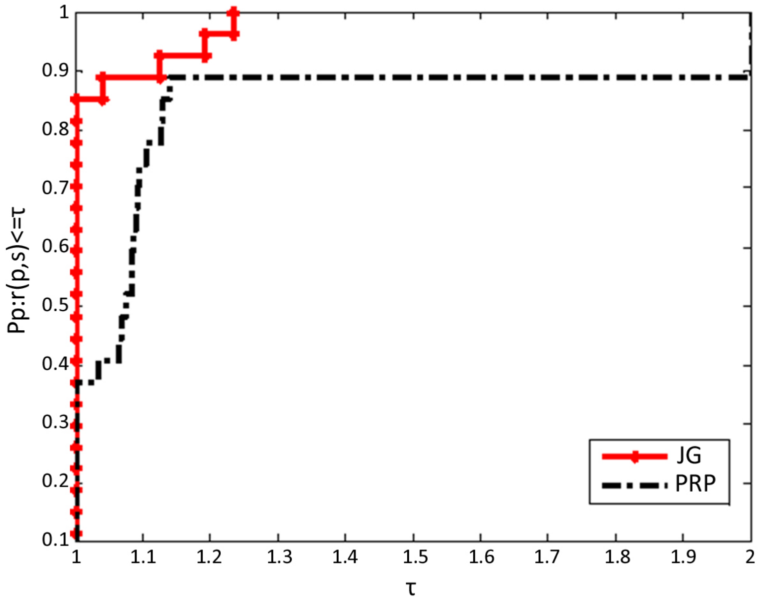

Figure 1. Algorithm JG and algorithm PRP performance chart (CT)
图 1. JG 算法与 PRP 算法的性能比较(CT)

5. 结论

针对求解非线性单调方程组，本文采用文献 6 的线搜索，给出了一个三项共轭梯度算法 JG，并且在一定的假设条件下得到了充分下降性和全局收敛性，从数值实验的结果可以看到，JG 算法在与 PRP 算法比较起来具有更好的性质。

参考文献

[1] Polyak, B.T. (1969) The Conjugate Gradient Method in Extremal Problems. *Ussr Computational Mathematics & Mathematical Physics*, **9**, 94-112. https://doi.org/10.1016/0041-5553(69)90035-4

[2] Dai, Y. and Yuan, Y. (1999) A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property. *Siam Journal on Optimization*, **10**, 177-182. https://doi.org/10.1137/S1052623497318992

[3] Fletcher, R. and Reeves, C.M. (1964) Function Minimization by Conjugate Gradients. *Computer Journal*, **7**, 149-154. https://doi.org/10.1093/comjnl/7.2.149

[4] Polak, E. and Ribiere, G. (1968) Note sur la convergence de methodes de directions conjuguees. *Revue Française D'-informatique et de Recherche Opérationnelle*, **16**, 35-43. https://doi.org/10.1051/m2an/196903R100351

[5] Wei, Z., Yao, S. and Liu, L. (2006) The Convergence Properties of Some New Conjugate Gradient Methods. *Applied Mathematics & Computation*, **183**, 1341-1350. https://doi.org/10.1016/j.amc.2006.05.150

[6] Yuan, G. and Lu, X. (2008) A New Backtracking Inexact BFGS Method for Symmetric Nonlinear Equations. *Computers & Mathematics with Applications*, **55**, 116-129. https://doi.org/10.1016/j.camwa.2006.12.081

[7] Zhang, L., Zhou, W. and Li, D.H. (2006) A Descent Modified Polak-Ribiere-Polyak Conjugate Gradient Method and Its Global Convergence. *IMA Journal of Numerical Analysis*, **26**, 629-640. https://doi.org/10.1093/imanum/drl016

[8] Solodov, M.V. and Svaiter, B.F. (1998) A Globally Convergent Inexact Newton Method for Systems of Monotone Equations. Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods. Springer, US, 355-369. https://doi.org/10.1007/978-1-4757-6388-1_18

[9] Yuan, G., Wei, Z. and Lu, X. (2011) A BFGS Trust-Region Method for Nonlinear Equations. *Computing*, **92**, 317-333. https://doi.org/10.1007/s00607-011-0146-z

知网检索的两种方式：

1. 打开知网页面 http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD
   下拉列表框选择：[ISSN]，输入期刊 ISSN：2324-7991，即可查询
2. 打开知网首页 http://cnki.net/
   左侧“国际文献总库”进入，输入文章标题，即可查询