

# A New Conjugate Gradient Method

Yong Li<sup>1</sup>, Gonglin Yuan<sup>2</sup>

<sup>1</sup>School of Mathematics and Statistics, Baise University, Baise Guangxi

<sup>2</sup>Department of Mathematics and Information Science, Guangxi University, Nanning Guangxi

Email: liyong3922@163.com

Received: Oct. 27<sup>th</sup>, 2016; accepted: Nov. 9<sup>th</sup>, 2016; published: Nov. 18<sup>th</sup>, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

This paper has designed a new parameter formula. The conjugate gradient algorithm which based on the parameter formula is global convergence with WWP line search under appropriate conditions. Preliminary numerical results turn out that this new method is effective.

## Keywords

Unconstrained Optimization, Conjugate Gradient Method, Global Convergence

---

# 一种新的共轭梯度法

黎 勇<sup>1</sup>, 袁功林<sup>2</sup>

<sup>1</sup>百色学院数学与统计学院, 广西 百色

<sup>2</sup>广西大学数学与信息科学学院, 广西 南宁

Email: liyong3922@163.com

收稿日期: 2016年10月27日; 录用日期: 2016年11月9日; 发布日期: 2016年11月18日

---

## 摘 要

本文设计了一个新的参数公式, 在适当条件下, 建立在此参数公式上的共轭梯度算法在WWP线搜索下全局收敛。初步的数值实验结果表明新算法是有效的。

## 关键词

无约束优化, 共轭梯度法, 全局收敛性

## 1. 引言

考虑无约束优化问题  $\min\{f(\mathbf{x})|\mathbf{x}\in\mathfrak{R}^n\}$ , 共轭梯度法是解决此类问题的一种非常有效方法。共轭梯度法是一种迭代算法, 因为简便、存储需求小的优势而往往被用来求解大规模的优化问题, 它的迭代公式通常如下:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k \quad (1)$$

其中  $t_k$  是搜索步长,  $\mathbf{d}_k$  是搜索方向, 通常定义如下:

$$\mathbf{d}_k = \begin{cases} -\mathbf{g}_k & \text{若 } k=1, \\ -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1} & \text{若 } k \geq 2. \end{cases} \quad (2)$$

这里的  $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$ , 表示目标函数  $f(\mathbf{x})$  在点  $\mathbf{x}_k$  处的梯度, 其中  $f(\mathbf{x})$  在  $\mathfrak{R}^n \rightarrow \mathfrak{R}$  上连续可微。  $\beta_k \in \mathfrak{R}$  则是一个参数。通过改变这个参数的不同选取, 可以获得不同的带参数共轭梯度算法, 人们经常讨论的参数公式有以下这些[1]:

$$\beta_k^{HS} = \frac{\mathbf{g}_k^T (\mathbf{g}_k - \mathbf{g}_{k-1})}{(\mathbf{g}_k - \mathbf{g}_{k-1})^T \mathbf{d}_{k-1}}, \quad \beta_k^{FR} = \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}}, \quad \beta_k^{PRP} = \frac{\mathbf{g}_k^T (\mathbf{g}_k - \mathbf{g}_{k-1})}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}}, \quad \beta_k^{CD} = -\frac{\|\mathbf{g}_k\|^2}{\mathbf{g}_{k-1}^T \mathbf{d}_{k-1}},$$

研究者普遍认为 PRP 方法是目前数值表现最好的共轭梯度法之一, 但收敛性质比较一般。文献[2]举例指出: 即使按照 Curry 原则进行选取步长, PRP 共轭梯度法对一般的非凸函数也未必收敛; 因此, 很多研究者都对 PRP 共轭梯度法进行了修正改进, 文献[3] [4] [5] [6] [7]均是对  $\beta_k^{PRP}$  进行非负修正, 其中文献[5]提出的参数公式

$$\beta_k^{WYL} = \frac{\mathbf{g}_k^T \left( \mathbf{g}_k - \frac{\|\mathbf{g}_k\|}{\|\mathbf{g}_{k-1}\|} \mathbf{g}_{k-1} \right)}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}}$$

受到了广泛关注。建立在这一参数公式基础上的共轭梯度算法不仅收敛效果好, 而且数值结果也比较理想[4] [5]。在此基础上文献[6]提出了一种修正的 WYL 共轭梯度法

$$\beta_k^{WYL^*} = \frac{\mathbf{g}_k^T \left( \mathbf{g}_k - \frac{\|\mathbf{g}_k\|}{\|\mathbf{g}_{k-1}\|} \mathbf{g}_{k-1} \right)}{\|\mathbf{g}_{k-1}\|^2 + \mu |\mathbf{g}_k^T \mathbf{d}_{k-1}|},$$

该方法能够自动保证参数公式的非负性, 而且不依赖任何线搜索, 可以自动保证充分下降性, 因而收敛效果也比较好。我们在文献[7]中提出的修正的 PRP 共轭梯度法

$$\beta_k^{MPRP} = \frac{\mathbf{g}_k^T \left( \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} \mathbf{g}_k - \mathbf{g}_{k-1} \right)}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}} \quad (3)$$

借鉴了文献[6]的思路。若对参数公式(3)的分母加一个非负的量  $\mu \|\mathbf{g}_{k-1}\| \|\mathbf{g}_k^T \mathbf{d}_{k-1}\|$ , 选择参数为

$$\beta_k^* = \frac{\mathbf{g}_k^T \left( \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} \mathbf{g}_k - \mathbf{g}_{k-1} \right)}{\|\mathbf{g}_{k-1}\|^2 + \mu \|\mathbf{g}_{k-1}\| \|\mathbf{g}_k^T \mathbf{d}_{k-1}\|}, \quad \mu \geq 0 \quad (4)$$

则可以得到一个新的共轭梯度法。

下面将建立并讨论新算法和算法的收敛性。新算法采用弱 Wolfe-Powell(WWP)线搜索, 步长  $t_k$  满足:

$$f(\mathbf{x}_k + t_k \mathbf{d}_k) \leq f(\mathbf{x}_k) + \delta t_k \mathbf{g}_k^T \mathbf{d}_k \quad (5)$$

和

$$\mathbf{g}(\mathbf{x}_k + t_k \mathbf{d}_k)^T \mathbf{d}_k \geq \sigma \mathbf{g}_k^T \mathbf{d}_k \quad (6)$$

其中  $\sigma$  和  $\delta$  是满足  $0 < \delta < \sigma < 1$  的常数。

## 2. 算法与假设

算法 1

步 1: 给出  $\mathbf{x}_1 \in \mathfrak{R}^n$ ,  $\varepsilon \geq 0$ 。令  $\mathbf{d}_1 = -\mathbf{g}_1$ ,  $k=1$ 。若  $\|\mathbf{g}_1\| \leq \varepsilon$ , 则停止。

步 2: 计算步长因子  $t_k$ , 使  $t_k$  满足 WWP 线搜索(5)、(6)。

步 3: 令  $\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k$ ,  $\mathbf{g}_{k+1} = \mathbf{g}(\mathbf{x}_{k+1})$ 。若  $\|\mathbf{g}_{k+1}\| \leq \varepsilon$ , 则停止。

步 4: 按(4)式计算公式  $\beta_k^*$ , 其中  $\mu > 2$ , 利用(3)式计算  $\mathbf{d}_{k+1}$ 。

步 5: 令  $k := k+1$ , 转步 2。

假设:

**假设(i)** 水平集  $\Omega = \{\mathbf{x} \in \mathfrak{R}^n : f(\mathbf{x}) \leq f(\mathbf{x}_1)\}$  有界。

**假设(ii)** 函数  $f$  在  $\Omega$  的某邻域  $N$  内连续可微, 并且它的梯度函数  $\mathbf{g}(\mathbf{x})$  Lipschitz 连续。即存在一个正的常数  $L > 0$ , 对任意  $\mathbf{x}, \mathbf{y} \in N$ ,

$$\|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\|. \quad (7)$$

## 3. 全局收敛性

引理 1. 若  $\beta_k^* = \frac{\mathbf{g}_k^T \left( \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} \mathbf{g}_k - \mathbf{g}_{k-1} \right)}{\|\mathbf{g}_{k-1}\|^2 + \mu \|\mathbf{g}_{k-1}\| \|\mathbf{g}_k^T \mathbf{d}_{k-1}\|}$ , 其中  $\mu \geq 0$ , 则  $0 \leq \beta_k^* \leq \beta_k^{MPRP}$ 。

证明: 因为  $\mu \|\mathbf{g}_{k-1}\| \|\mathbf{g}_k^T \mathbf{d}_{k-1}\| \geq 0$ , 所以  $\beta_k^* \leq \beta_k^{MPRP}$ 。

由 Cauchy-Schwarz 不等式可得

$$\begin{aligned} \beta_k^* &= \frac{\mathbf{g}_k^T \left( \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} \mathbf{g}_k - \mathbf{g}_{k-1} \right)}{\|\mathbf{g}_{k-1}\|^2 + \mu \|\mathbf{g}_{k-1}\| \|\mathbf{g}_k^T \mathbf{d}_{k-1}\|} = \frac{\frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} \|\mathbf{g}_k\|^2 - \mathbf{g}_k^T \mathbf{g}_{k-1}}{\|\mathbf{g}_{k-1}\|^2 + \mu \|\mathbf{g}_{k-1}\| \|\mathbf{g}_k^T \mathbf{d}_{k-1}\|} \\ &\geq \frac{\|\mathbf{g}_{k-1}\| \|\mathbf{g}_k\| - \|\mathbf{g}_k\| \|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_{k-1}\|^2 + \mu \|\mathbf{g}_{k-1}\| \|\mathbf{g}_k^T \mathbf{d}_{k-1}\|} = 0 \end{aligned}$$

引理 2 (性质\*)。考虑算法 1, 假设(i)、(ii)成立, 若

$$0 \leq \gamma \leq \|\mathbf{g}_k\| \leq \bar{\gamma} \quad (8)$$

则存在常数  $b > 1$  和  $\lambda > 0$ , 使对所有  $k$ , 有

$$|\beta_k^*| \leq b,$$

以及

$$\|s_{k-1}\| \leq \lambda \Rightarrow |\beta_k^*| \leq \frac{1}{2b}.$$

证明: 取  $b = \frac{2\bar{\gamma}}{\gamma} > 1$ ,  $\lambda = \frac{\gamma^2}{8L(\bar{\gamma})^2}$ , 则

$$|\beta_k^*| = \frac{\mathbf{g}_k^T \left( \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} \mathbf{g}_k - \mathbf{g}_{k-1} \right)}{\|\mathbf{g}_{k-1}\|^2} \leq \frac{\|\mathbf{g}_k\| \left( \frac{\gamma}{\bar{\gamma}} \|\mathbf{g}_k\| + \|\mathbf{g}_{k-1}\| \right)}{\|\mathbf{g}_{k-1}\|^2} \leq \frac{\bar{\gamma} \left( \frac{\gamma}{\bar{\gamma}} \bar{\gamma} + \gamma \right)}{\gamma^2} = \frac{2\bar{\gamma}}{\gamma} = b.$$

而若  $\|s_{k-1}\| \leq \lambda$ , 则由 Lipschitz 连续性知

$$\begin{aligned} |\beta_k^*| &= \frac{\mathbf{g}_k^T \left( \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} \mathbf{g}_k - \mathbf{g}_k + \mathbf{g}_k - \mathbf{g}_{k-1} \right)}{\|\mathbf{g}_{k-1}\|^2} \leq \frac{\|\mathbf{g}_k\| \left( \left\| \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} \mathbf{g}_k - \mathbf{g}_k \right\| + \|\mathbf{g}_k - \mathbf{g}_{k-1}\| \right)}{\|\mathbf{g}_{k-1}\|^2} \\ &\leq \frac{\|\mathbf{g}_k\| (\|\mathbf{g}_{k-1}\| - \|\mathbf{g}_k\| + L\lambda)}{\|\mathbf{g}_{k-1}\|^2} \leq \frac{\|\mathbf{g}_k\| (\|\mathbf{g}_k - \mathbf{g}_{k-1}\| + L\lambda)}{\|\mathbf{g}_{k-1}\|^2} \\ &\leq \frac{2\|\mathbf{g}_k\| L\lambda}{\|\mathbf{g}_{k-1}\|} \leq \frac{2L\lambda\bar{\gamma}}{\gamma} = \frac{1}{2b}. \end{aligned}$$

引理 3 [8]. 考虑共轭梯度方法(1)和(2), 步长因子  $t_k$  满足线搜索 WWP (5)、(6), 若:

- 1)  $\beta_k \geq 0$ ;
- 2) 充分下降条件成立;
- 3) 性质(\*)成立;
- 4) 假设(i)(ii)成立;

则算法全局收敛。

在共轭梯度法的讨论中, 充分下降条件

$$\mathbf{g}_k^T \mathbf{d}_k \leq -c \|\mathbf{g}_k\|^2, \quad c > 0, \quad (9)$$

是一个很重要的性质。

根据以上 3 个引理, 在假设充分下降条件可以满足的前提下, 容易得出本文所给的共轭梯度法在 WWP 线搜索下全局收敛的结论。

定理 1. 若假设(i)(ii)成立, 且(9)式成立, 序列  $\{\mathbf{x}_k\}$  由算法 1 生成, 则

$$\liminf_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0.$$

#### 4. 数值试验

为了考查新算法的数值表现, 本文选取 26 个函数进行数值实验, 部分测试函数来自 CUTE 函数库。数值试验程序我们利用 Fortran 语言编写。表 1 列出的计算结果是在参数  $\mu = 3$ , 线搜索参数  $\delta = 0.01$ ,

**Table 1.** The results of numerical experiment**表 1.** 数值实验结果

测试函数	维数	$f^*$	$g^*$	$N_i/N_f$
Extended Trigonometric	3000	0.3333498E-06	0.9931361E-05	45/212
	6000	0.1153415E-06	0.9473102E-05	47/239
Extended Penalty	3000	0.275597E+04	0.4997212E-05	10/69
	6000	0.5611677E+04	0.8933257E-05	11/73
Raydan 2	3000	0.300000E+04	0.7822789E-05	4/18
	6000	0.6000000E+04	0.7820519E-05	4/18
Hager	3000	-0.2925501E+06	0.3476614E-04	112/550
	6000	-0.9347349E+06	0.7880606E-04	140/602
Generalized Tridiagonal 1	3000	0.2997210E+04	0.7700144E-05	27/83
	6000	0.5997210E+04	0.7715256E-05	27/83
Extended Three Exponential Terms	3000	0.3838900E+04	0.8373884E-05	14/52
	6000	0.7677800E+04	0.3013287E-05	15/59
Generalized Tridiagonal 2	3000	0.1114854E+01	0.9450020E-05	123/371
	6000	0.1114854E+01	0.9573045E-05	123/379
Diagonal 4	3000	0.2804463E-08	0.1933722E-05	4/15
	6000	0.5251179E-08	0.1871038E-05	4/15
Diagonal 5	3000	0.2079442E+04	0.4251189E-07	2/10
	6000	0.4158883E+04	0.4243535E-07	2/10
Extended Himmelblau	3000	0.9804727E-09	0.7983281E-05	14/57
	6000	0.2062873E-08	0.8169256E-05	14/57
Extended PSC1	3000	0.1159799E+04	0.5617920E-05	11/49
	6000	0.2319597E+04	0.5786743E-05	11/49
Extended Block Diagonal BD1	3000	0.4274553E-07	0.8234179E-05	25/120
	6000	0.8402421E-07	0.8160427E-05	25/120
Extended Quadratic Penalty QP1	3000	0.1199000E+05	0.9924719E-05	8/177
	6000	0.2399000E+05	0.8802386E-05	9/242
Extended EP1	3000	0.2379528E+05	0.4934875E-05	2/8
	6000	0.4759058E+05	0.4882155E-05	2/8
Extended Tridiagonal-2	3000	0.1168798E+04	0.9850863E-05	46/150
	6000	0.2337985E+04	0.9371683E-05	47/182
DIXMAANA (CUTE)	3000	0.1000000E+01	0.2617340E-05	5/25
	6000	0.1000000E+01	0.2617859E-05	5/25
DIXMAANB (CUTE)	3000	0.1000000E+01	0.1674320E-05	5/28
	6000	0.1000000E+01	0.1670881E-05	5/28
DIXMAANC (CUTE)	3000	0.1000000E+01	0.8383319E-05	5/30
	6000	0.1000000E+01	0.6797974E-05	5/30
Broyden Tridiagonal	3000	0.8949051E-11	0.9586283E-05	50/156
	6000	0.3970671E+00	0.9702865E-05	286/882

Continued

EDENSCH (CUTE)	3000	0.1800328E+05	0.6594267E-05	26/84
	6000	0.3600328E+05	0.6080001E-05	26/84
DIAGONAL 6	3000	0.9180457E-07	0.7823246E-05	4/18
	6000	0.1836105E-06	0.7823289E-05	4/18
DIXMAAND (CUTE)	3000	0.3328432E+04	0.8480679E-05	21/99
	6000	0.6658787E+04	0.8407817E-05	20/88
FLETCHCR (CUTE)	3000	-0.2999000E+04	0.1190158E-04	14/295
	6000	-0.5999000E+04	0.6698629E-05	14/225
COSINE (CUTE)	3000	0.6352702E-10	0.3965849E-06	7/29
	6000	0.1283959E-09	0.3986312E-06	7/29
Extended DENSCHNB (CUTE)	3000	0.5952919E-10	0.2392949E-05	16/69
	6000	0.1004371E-08	0.9853385E-05	15/66
BIGGSB1 (CUTE)	3000	0	0	0/2
	6000	0	0	0/2

$\sigma = 0.1$ ,  $\varepsilon = 10^{-5}$ ; 终止条件是迭代次数  $N_i > 10^4$ , 或者  $\|g_k\| \leq \varepsilon$  的时候分别对 3000 维和 6000 维进行测试时取得的。表 1 中的  $f^*$  是算法终止时的函数值,  $g^*$  表示算法终止时梯度范数值,  $N_i$  表示迭代次数,  $N_f$  表示计算次数。

上表中的测试结果表明新算法是有效的, 特别是对于高维问题, 新算法表现出了较好的处理能力, 所以对于大规模优化问题的求解, 新算法有能力解决此类问题。

## 基金项目

国家自然科学基金项目(No.11661001, No.11661009); 广西自然科学基金(No.2014GXNSFAA118030); 广西教育厅科研项目(No.YB2014389)。

## 参考文献 (References)

- [1] 戴彧虹, 袁亚湘. 非线性共轭梯度法[M]. 上海: 上海科技出版社, 1999.
- [2] Powell, M.J.D. (1984) Nonconvex Minimization Calculations and the Conjugate Gradient Method. Springer-Verlag, Berlin, 122-141. <http://dx.doi.org/10.1007/bfb0099521>
- [3] Wei, Z., Yao, S. and Liu, L. (2006) The Convergence Properties of Some New Conjugate Gradient Methods. *Applied Mathematics and Computation*, **183**, 1341-1350. <http://dx.doi.org/10.1016/j.amc.2006.05.150>
- [4] Huang, H., Wei, Z. and Yao, S. (2007) The Proof of the Sufficient Descent Condition of the Wei-Yao-Liu Conjugate Gradient Method under the Strong Wolfe-Powell Line Search. *Applied Mathematics and Computation*, **189**, 1241-1245. <http://dx.doi.org/10.1016/j.amc.2006.12.006>
- [5] Lu, S., Wei, Z. and Mo, L. (2011) Some Global Convergence Properties of the Wei-Yao-Liu Conjugate Gradient Method with Inexact Line Search. *Applied Mathematics and Computation*, **217**, 7132-7137. <http://dx.doi.org/10.1016/j.amc.2011.01.097>
- [6] Huang, H. and Lin, S. (2014) A Modified Wei-Yao-Liu Conjugate Gradient Method for Unconstrained Optimization. *Applied Mathematics and Computation*, **231**, 179-186. <http://dx.doi.org/10.1016/j.amc.2014.01.012>
- [7] 黎勇. 一类新的修正 PRP 共轭梯度法[J]. 武汉理工大学学报(交通科学与工程版), 2012, 36(2): 437-440.
- [8] Hager, W.W. and Zhang, H. (2006) A Survey of Nonlinear Conjugate Gradient Methods. *Pacific journal of Optimization*, **2**, 35-58.

**期刊投稿者将享受如下服务：**

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：[aam@hanspub.org](mailto:aam@hanspub.org)