*Pure Mathematics* 理论数学, 2011, 1, 41-45
http://dx.doi.org/10.12677/pm.2011.11009 Published Online April 2011 http://www.hanspub.org/journal/pm/

# A Three-Terms Conjugate Gradient Algorithm and Its Convergence

**Hai Chen**
College of Mathematic and Information Science, Guangxi University, Nanning
Email: hchengx@126.com
Received: Mar. 18th, 2011; revised: Mar. 30th, 2011; accepted: Apr. 2nd, 2011.

**Abstract:** In this paper, a three-terms conjugate gradient method is proposed .The search direction possesses the sufficient descent property without any line search. Moreover, the search direction has not only the gradient value but also function value. The global convergence will be established and the numerical results are reported.

**Keywords:** Conjugate Gradient; Sufficient Descent Property; Convergence

# 一个三项共轭梯度算法及其收敛性

**陈　海**
广西大学数学与信息科学学院，南宁
Email: hchengx@126.com
收稿日期：2011 年 3 月 18 日；修回日期：2011 年 3 月 30 日；录用日期：2011 年 4 月 2 日

**摘　要：**本文给出一个三项共轭梯度算法，搜索方向在不需要任何线搜索的条件下，拥有充分下降性条件，在此方向的定义中，不但拥有梯度值信息还拥有函数值信息，证明了全局收敛性并给出数值检验结果。

**关键词：**共轭梯度；充分下降；收敛性

## 1. 引言

本文考虑求解下面的无约束最优化问题

$$\min_{x\in\Re^n} f(x) \tag{1.1}$$

其中 $f(x)$ 是连续可微函数。关于上述问题的求解方法有很多种，如牛顿法、拟牛顿法、信赖域方法和共轭梯度法等。其中共轭梯度方法因为结构简单，计算机存储量小，而且效果明显被广泛应用。下面的迭代公式经常被使用

$$x_{k+1} = x_k + \alpha_k d_k, \qquad k = 0, 1, 2, \cdots$$

其中 $x_k$ 是第 $k$ 次迭代点，$\alpha_k > 0$ 是步长，$d_k$ 是搜索方向，定义为

$$d_k = \begin{cases} -g_k + \beta_k d_{k-1}, & \text{if } k \ge 1 \\ -g_k, & \text{if } k = 0 \end{cases} \tag{1.2}$$

其中 $\beta_k \in \Re$ 是一个参数，由于 $\beta_k$ 的选取不同而称为不同的共轭梯度方法（见文献[1-7]等），在这其中，PRP 方法是最为著名的一个，其公式为

$$\beta_k^{PRP} = \frac{g_{k+1}^T(g_{k+1} - g_k)}{\|g_k\|^2}, [6, 7]$$

其中 $g_k$ 和 $g_{k+1}$ 表示 $\nabla f(x_k)$ 和 $\nabla f(x_{k+1})$，为函数 $f(x)$ 在 $x_k$ 和 $x_{k+1}$ 的梯度值，$\|.\|$ 表示欧式向量范数。此方法的数值表现优越但收敛性不理想，但也常常被人们用于实际的问题中，是研究最为热门的公式之一。许多学者都希望找到数值表现可与PRP 相媲美同时性质又比其好的方法，其中有许多成果可见文献[8-15]等。经过研究发现，该方法收敛性不好的主要原因在于它不具有充分下降性，所谓充分下降性，即对所有 $k$，下式成立：

$$d_k^T g_k \le -c\|g_k\|^2, \tag{1.3}$$

研究发现，此性质在收敛性分析中起着非常重要的作用。在优化研究中，下面的线搜索技术常被用来求解步长 $\alpha_k$

$$f(x_k+\alpha_k d_k) \le f_k + \delta\alpha_k g_k^T d_k \tag{1.4}$$

和

$$g(x_k+\alpha_k d_k)^T d_k \ge \sigma g_k^T d_k, \tag{1.5}$$

其中 $\delta\in(0,1/2), \sigma\in(\delta,1)$，被称作 WWP 线搜索，迄今为止，对一般函数而言，PRP 方法在上述线搜索技术下的全局收敛性仍没得到证明。在文章[16]，Zhang 等给出了一个三项共轭梯度方法，其方向定义为

$$d_{k+1}=\begin{cases}-g_{k+1}+\beta_k^{PRP}d_k-\theta_k y_k, & \text{if } k\ge 1\\ -g_{k+1}, & \text{if } k=0\end{cases} \tag{1.6}$$

其中 $\theta_k=\dfrac{g_{k+1}^T d_k}{\|g_k\|^2}$，$\beta_k^{PRP}=\dfrac{g_{k+1}^T y_k}{\|g_k\|^2}$，$y_k=g_{k+1}-g_k$，容易得到 $d_k^T g_k=-\|g_k\|^2$。但是此方法，仅仅利用了梯度值信息，忽略了函数值信息。在文章[17]，Zhang 等给出了一个新的拟牛顿方程

$$B_{k+1}S_k=\delta_k^1, \tag{1.7}$$

其中 $\delta_k^1=y_k+\gamma_k^1 s_k$ 和

$$\gamma_k^1=\frac{3(g_{k+1}+g_k)^T s_k+6(f(x_k)-f(x_{k+1}))}{\|s_k\|^2}$$

$s_k=x_{k+1}-x_k$，Yuan 和 Wei[18]给出了下面的方程

$$B_{k+1}S_k=\delta_k^2 \tag{1.8}$$

其中 $\delta_k^2=y_k+\gamma_k^2 s_k$ 和

$$\gamma_k^2=\max\left\{0,\frac{(g_{k+1}+g_k)^T s_k+3(f(x_k)-f(x_{k+1}))}{\|s_k\|^2}\right\},$$ $B_{k+1}$ 是拟牛顿矩阵。上述两个拟牛顿方程中均含有函数值信息，受他们方法的启发，我们给出下面的三项共轭梯度方法

$$d_{k+1}=\begin{cases}-g_{k+1}+\beta_k^{MPRP}d_k-\theta_k y_k^*, & \text{if } k\ge 1\\ -g_{k+1}, & \text{if } k=0\end{cases} \tag{1.9}$$

其中 $\theta_k=\dfrac{g_{k+1}^T d_k}{\|g_k\|^2}$，$\beta_k^{MPRP}=\dfrac{g_{k+1}^T y_k^*}{\|g_k\|^2}$，$y_k=g_{k+1}-g_k$，

$$y_k^*=y_k+\max\left\{0,\frac{3(g_{k+1}+g_k)^T s_k+6(f(x_k)-f(x_{k+1}))}{\|s_k\|^2}\right\},$$

从此公式可见，新公式不但拥有梯度值，还拥有函数值信息。本文将对其充分下降性和收敛性进行分析，下面一部分给出算法步骤，第三部分给出收敛性分析。

## 2. 算法

**算法 1**（修改的三项共轭梯度算法）

**步骤 0**：给定 $x_0\in\Re^n$，$\delta\in(0,1/2),\sigma\in(\delta,1)$ 和终止参数 $\varepsilon>0$。令 $d_0=-g_0=-\nabla f(x_0)$，置 $k:=0$。

**步骤 1**：若 $\|g_k\|\le\varepsilon$，停止。

**步骤 2**：寻找满足(1.4)和(1.5)的步长 $\alpha_k$。

**步骤 3**：令 $x_{k+1}=x_k+\alpha_k d_k$，如果 $\|g_{k+1}\|\le\varepsilon$，停止。

**步骤 4**：利用公式(1.9)计算搜索方向。

**步骤 5**：置 $k:=k+1$，转步骤 2。

## 3. 充分下降性和全局收敛性分析

下面我们给出修改的三项公式方向具有充分下降性。

**引理 3.1** 对 $k\ge 0$，修改的三项公式的搜索方向满足

$$d_{k+1}^T g_{k+1}=-\|g_{k+1}\|^2 \tag{3.1}$$

**证明**：如果 $k=0$，则 $g_0^T d_0=-\|g_0\|^2$，则(1.3)成立。当 $k\ge 1$ 时，利用新方向的定义可得

$$d_{k+1}^T g_{k+1}=-\|g_{k+1}\|^2+\frac{g_{k+1}^T y_k^*}{\|g_k\|^2}d_k^T g_{k+1}$$

$$-\frac{g_{k+1}^T d_k}{\|g_k\|^2}g_{k+1}^T y_k^*=-\|g_{k+1}\|^2$$

命题得证，证毕。

为了证明算法1的收敛性，我们需要下面的假设条件：

(A)：1) 水平集 $\Omega=\{x\in\Re^n: f(x)\le f(x_0)\}$ 有界。

2) $f$ 在 $\Omega$ 上有下界且连续可微，它的梯度 $g$ 满足 Lipschitz 条件，即存在常数 $L>0$ 满足

$$\|g(x)-g(y)\|\le L\|x-y\|, \forall x,y\in\Omega \tag{3.2}$$

**引理 3.2** 设假设(A)满足，且存在常数 $c>0$ 使得 $\|d_k\|\le c\|g_k\|$ 成立，则由算法 1 产生的序列 $\{g_k\}$ 和 $\{d_k\}$ 满足 $\lim_{k\to\infty}\inf\|g_k\|=0$。

**证明**：利用(1.5)和 Lipschitz 条件(3.2)，有

$$-(1-\sigma)g_k^T d_k \le (g_{k+1}-g_k)^T d_k \le \alpha_k L\|d_k\|^2,$$

 

利用(1.3)，获得

$$\alpha_k \geq \frac{(1-\sigma)}{L}\frac{\left|g_k^T d_k\right|}{\|d_k\|^2},$$

将其代入(1.4)，移向得：

$$\frac{\delta(1-\sigma)}{L}\frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} \leq f_k - f_{k+1}$$

将上式从 $k=0$ 到 $\infty$ 相加，有

$$\sum_{k=0}^{\infty}\frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} \leq \frac{L}{\delta(1-\sigma)}\left(f_0 - f_\infty\right)$$

并利用假设(A)中函数 $f$ 有下界，下式

$$\sum_{k=0}^{\infty}\frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} \leq +\infty \tag{3.3}$$

成立，利用充分下降性(3.1)和 $\|d_k\| \leq c\|g_k\|$，上式可变为

$$\sum_{k=0}^{\infty}\|g_k\|^2 \leq +\infty$$

则有 $\|g_k\| \to 0$ 成立，证毕。

## 4. 数值结果

为了验证所给算法的有效性，下面给出数值检验结果，检验函数是在工程领域经常用到的 Benchmark 问题，列举如下。

1. Sphere function

$$f_{Sph}(x) = \sum_{i=1}^{n} x_i^2, x_i \in [-5.12, 5.12]\ x^* = (0,0,\cdots,0), f_{Sph}(x^*) = 0$$

2. Schwefel's function

$$f_{SchDS}(x) = \sum_{i=1}^{n}(\sum_{j=1}^{i} x_j)^2, x_i \in [-65.536, 65.536]\ x^* = (0,0,\ldots,0), f_{SchDS}(x^*) = 0$$

3. Rastrigin function

$$f_{Ras}(x) = 10n + \sum_{i=1}^{n}(x_i^2 - 10\cos(2\pi x_i)), x_i \in [-5.12, 5.12]\ x^* = (0,0,\cdots,0), f_{Ras}(x^*) = 0$$

4. Schwefel function

$$f_{Sch}(x) = 418.9829n + \sum_{i=1}^{n} x_i \sin\sqrt{|x_i|}, x_i \in [-512.03, 511.97]\ x^* = (-420.9678, -420.9678, \cdots, -420.9678), f_{Sch}(x^*) = 0$$

5. Griewank function

$$f_{Gri}(x) = 1 + \sum_{i=1}^{n}\frac{x_i^2}{4000} - \prod_{i=1}^{n}\cos\frac{x_i}{i}, x_i \in [-600, 600]\ x^* = (0,0,\cdots,0), f_{Gri}(x^*) = 0$$

这些 Benchmark 问题也可以从下述的网页找到：http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume24/ortizboyer05a-html/node6.html。

在我们实际计算中，参数选取如下：$\varepsilon = 10^{-5}, \delta = 0.1, \sigma = 0.9$，停止准则采用 Himmeblau 准则：

如果 $|f(x_k)| > e_1$，令 $stop1 = \frac{|f(x_k) - f(x_{k+1})|}{|f(x_k)|}$；否则令 $stop1 = |f(x_k) - f(x_{k+1})|$。如果 $\|g(x\ )\| < \varepsilon$ 或者 $stop1 < e_2$ 满足，算法终止，$e_1 = e_2 = 10^{-5}$。如果总的迭代次数超过 1000 次，我们也终止算法。下表中各代码含义是：

$x_0$：初始点；Dim：问题的维数；NI：迭代次数；NFG：函数值与梯度值迭代次数和；$f(\bar{x})$：算法终止时的函数值。

从表 1 的数值结果可以看出，算法 1 对上述 5 个问题还是能有效地求解。

**Table 1. The numerical results of algorithm 1**
**表 1. 算法 1 的数值结果**

| | | NI/NFG/ $f(\bar{x})$ | NI/NFG/ $f(\bar{x})$ | NI/NFG/ $f(\bar{x})$ | NI/NFG/ $f(\bar{x})$ | NI/NFG/ $f(\bar{x})$ |
|---|---|---|---|---|---|---|
| Sphere | $x_0$ | (–5, –5, ⋯, –5) | (–3, –3, ⋯, –3) | (–1, –1, ⋯, –1) | (2,2, ⋯,2) | (4,4, ⋯,4) |
| | 10 | 2/6/7.888609e–030 | 3/7/1.972152e–030 | 3/7/1.232595e–031 | 3/7/1.972152e–030 | 2/6/7.888609e–030 |
| Dim | 100 | 2/6/3.478877e–026 | 2/6/4.930381e–028 | 2/6/4.930381e–030 | 2/6/3.993608e–028 | 2/6/1.262177e–027 |
| | 300 | 2/6/1.990296e–025 | 2/6/2.366583e–026 | 2/6/2.366583e–028 | 2/6/7.158913e–027 | 2/6/2.366583e–026 |
| Sphere | $x_0$ | (–5,0, –5,0, ⋯) | (–3,0, –3,0, ⋯) | (–1,0, –1,0, ⋯) | (2,0,2,0, ⋯) | (4,0,4,0, ⋯) |
| | 10 | 2/6/6.310887e–029 | 3/7/9.860761e–031 | 3/7/0.000000e+000 | 3/7/9.860761e–031 | 3/7/8.874685e–030 |
| Dim | 100 | 2/6/5.679799e–027 | 2/6/7.987217e–028 | 3/7/2.465190e–030 | 2/6/0.000000e+000 | 2/6/4.831773e–028 |
| | 300 | 2/6/2.319251e–026 | 2/6/1.848893e–026 | 2/6/1.664003e–029 | 2/6/2.958228e–029 | 2/6/1.848893e–026 |
| Schwfel's | $x_0$ | (–0.001, ⋯, –0.001) | (–0.0002, ⋯,–0.0002) | (0.0006, ⋯,0.0006) | (0.0007, ⋯,0.0007) | (0.0008, ⋯,0.0008) |
| | 10 | 3/8/3.539977e–007 | 3/8/1.415991e–008 | 3/8/1.274392e–007 | 3/8/1.734589e–007 | 3/8/2.265585e–007 |
| Dim | 50 | 5/14/2.297452e–006 | 4/11/3.471788e–007 | 5/14/8.270826e–007 | 5/14/1.125751e–006 | 5/14/1.470369e–006 |
| | 100 | 7/20/2.554559e–006 | 5/14/7.557030e–007 | 6/17/2.317420e–006 | 6/17/3.154267e–006 | 7/20/1.634918e–006 |
| Schwfel's | $x_0$ | (–0.001,0, ⋯) | (–0.0002,0, ⋯) | (0.0006,0, ⋯) | (0.0007,0, ⋯) | (0.0008,0, ⋯) |
| | 10 | 3/8/6.244149e–007 | 2/5/5.186065e–008 | 3/8/2.247894e–007 | 3/8/3.059633e–007 | 3/8/3.996256e–007 |
| Dim | 50 | 5/14/3.464888e–006 | 3/8/6.064742e–007 | 4/11/1.848353e–006 | 4/11/2.515814e–006 | 4/11/3.285961e–006 |
| | 100 | 7/20/6.523845e–006 | 4/11/9.360044e–007 | 5/14/3.845945e–006 | 6/17/3.745873e–006 | 6/17/4.892569e–006 |
| Rastrigin | $x_0$ | (–5, –5, ⋯, –5) | (–4, –4, ⋯, –4) | (1,1, ⋯,1) | (2,2, ⋯,2) | (3,3, ⋯,3) |
| | 10 | 4/13/2.487372e+002 | 4/13/1.591924e+002 | 5/16/9.949591e+000 | 4/13/3.979831e+001 | 4/13/8.954601e+001 |
| Dim | 100 | 3/9/2.487372e+003 | 3/9/1.591924e+003 | 3/9/9.949591e+001 | 3/9/3.979831e+002 | 3/9/8.954601e+002 |
| | 100 | 3/9/7.462117e+003 | 3/9/4.775773e+003 | 3/8/2.984877e+002 | 3/8/1.193950e+003 | 3/9/2.686380e+003 |
| Rastrigin | $x_0$ | (–5,0, –5,0, ⋯) | (–4,0, –4,0, ⋯) | (1,0,1,0, ⋯) | (2,0,2,0, ⋯) | (3,0,3,0, ⋯) |
| | 10 | 3/9/1.243686e+002 | 3/9/7.959622e+001 | 3/9/4.974795e+000 | 3/9/1.989916e+001 | 3/9/4.477301e+001 |
| Dim | 100 | 4/12/7.959622e+002 | 4/12/4.974795e+001 | 4/12/4.974795e+001 | 4/12/1.989916e+002 | 4/12/4.477301e+002 |
| | 300 | 3/9/3.731058e+003 | 3/9/2.387887e+003 | 3/9/1.492439e+002 | 3/9/5.969747e+002 | 3/9/1.343190e+003 |
| Schwefel | $x_0$ | (–500,–500, ⋯,–500) | (–400,–400, ⋯,–400) | (100,100, ⋯,100) | (400,400, ⋯,400) | (500,500, ⋯,500) |
| | 10 | 3/9/–1.006797e+004 | 2/6/–4.174987e+003 | 2/19/3.101787e+003 | 4/38/4.705971e+003 | 2/19/5.780458e+002 |
| Dim | 100 | 3/9/–1.006616e+005 | 2/6/–4.174488e+004 | 2/19/3.101787e+004 | 2/19/1.490427e+002 | 2/19/5.780458e+003 |
| | 300 | 3/9/–3.019763e+005 | 2/6/–1.252317e+005 | 2/19/9.305360e+004 | 4/39/5.048285e+005 | 2/19/1.734137e+004 |
| Schwefel | $x_0$ | (–500,0,–500,0, ⋯) | (–400,0,–400,0, ⋯) | (100,0,100,0, ⋯) | (400,0,400,0, ⋯) | (500,0,500,0, ⋯) |
| | 10 | 1/2/5.995721e+003 | 1/2/5.380480e+002 | 1/2/3.645808e+003 | 1/2/7.841610e+003 | 1/2/2.383937e+003 |
| Dim | 100 | 1/2/5.995721e+004 | 1/2/5.380480e+003 | 1/2/3.645808e+004 | 1/2/7.841610e+004 | 1/2/2.383937e+004 |
| | 300 | 1/2/1.798716e+005 | 1/2/1.614144e+004 | 1/2/1.093742e+005 | 1/2/2.352483e+005 | 1/2/7.151812e+004 |
| Griewank | $x_0$ | (–50, –50, ⋯, –50) | (–10,–10, ⋯,–10) | (1,1, ⋯,1) | (20,20, ⋯,20) | (30,30, ⋯,30) |
| | 10 | 2/19/7.250909e+000 | 2/19/1.264953e+000 | 20/60/8.438297e–007 | 13/54/1.115419e–006 | 3/9/5.618497e–006 |
| Dim | 100 | 4/32/2.276173e–002 | 15/45/7.810149e–006 | 3/8/0.000000e+000 | 2/6/0.000000e+000 | 2/6/0.000000e+000 |
| | 300 | 60/164/9.630521e-006 | 2/6/0.000000e+000 | 2/6/0.000000e+000 | 2/6/0.000000e+000 | 2/6/0.000000e+000 |
| Griewank | $x_0$ | (–50,0,–50,0, ⋯) | (–10,0,–10,0, ⋯) | (1,0,1,0, ⋯) | (20,0,20,0, ⋯) | (30,0,30,0, ⋯) |
| | 10 | 4/25/9.540580e–001 | 5/44/1.103014e+000 | 2/19/1.551065e+000 | 10/29/1.549244e–006 | 10/29/1.549244e–006 |
| Dim | 100 | 4/15/1.432188e–014 | 29/77/7.390224e–006 | 3/8/0.000000e+000 | 2/6/0.000000e+000 | 2/6/0.000000e+000 |
| | 300 | 51/133/5.541337e–007 | 2/6/0.000000e+000 | 2/6/0.000000e+000 | 2/6/0.000000e+000 | 2/6/0.000000e+000 |

 

## 参考文献 (References)


[1] Y. H. Dai, Y. Yuan. A nonlinear conjugate gradient method with a strong global convergence property. SIAM Journal on Optimization, 2000, 10(177): 177-182.

[2] R. Fletcher. Practical Method of Optimization. New York: Wiley, 1997.

[3] R. Fletcher, C. Reeves. Function minimization by conjugate gradients. Computer Journal, 1964, 7(2): 149-154.

[4] M. R. Hestenes, E. Stiefel. Method of conjugate gradient for solving linear equations. Res. Nat. Bur. Stand., 1952, 49(6): 409-436.

[5] Y. Liu, C. Storey. Efficient generalized conjugate gradient algorithms, part 1: theory. Journal of optimization theory and Application, 1992, 69(1):129-137.

[6] E. Polak, G. Ribiere. Note sur la xonvergence de directions conjugees. Rev. Francaise informat Recherche Operatinelle 3e Annee, 1969, 16: 35-43.

[7] B. T. Polyak. The conjugate gradient method in extreme problems. USSR Comp Math Math Phys., 1969, 9(4): 94-112.

[8] W. W. Hager, H. Zhang. A new conjugate gradient method with guaranteed descent and an efficient line search. SIAM Journal on Optimization, 2005, 16(1): 170-192.

[9] W. W. Hager, H. Zhang. Algorithm 851: CGDESENT, A conjugate gradient method with guaranteed descent. ACM Transactions on Mathematical Software, 2006, 32(1): 113-137.

[10] G. Li, C. Tang, Z. Wei. New conjugacy condition and related new conjugate gradient methods for unconstrained optimization problems. Journal of Computational and Applied Mathemathics, 2007, 202(2): 532-539.

[11] Z. Wei, G. Li, L. Qi. New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems. Applied Mathematics and Computation, 2006, 179(2): 407-430.

[12] Z. Wei, G. Li, L. Qi. Global convergence of the PRP conjugate gradient methods with inexact line search for nonconvex unconstrained optimization problems. Mathematics of Computation, 2008, 77(264): 2173-2193.

[13] Z. Wei, S. Yao, L. Lin. The convergence properties of some new conjugate gradient methods. Applied Mathematics and Computation, 2006, 183(2): 1341-1350.

[14] G. L. Yuan. Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems. Optimization Letters, 2009, 3(1): 11-21.

[15] G. L. Yuan, X. W. Lu. A modified PRP conjugate gradient method. Annals of Operations Research, 2009, 166(1): 73-90.

[16] L. Zhang, W. Zhou, D. Li. A descent modified Polak-Ribiere-Polyak conjugate method and its global convergence. IMA Journal on Numerical Analysis, 2006, 26(4): 629-649.

[17] J. Z. Zhang, N. Y. Deng, L. H. Chen. New quasi-Newton equation and related methods for unconstrained optimization. Journal of Optimization Theory and Application, 1999, 102(1): 147-167.

[18] G. Yuan, Z. Wei. Convergence analysis of a modified BFGS method on convex minimizations. Computational Optimization and Applications, 2010, 47(2): 237-255.


 