DOI:10.13656/j.cnki.gxkx.2012.04.020

广西科学 Guangxi Sciences 2012,19(4):319～322

一个修改的三项共轭梯度算法

A Modified Three-Terms Conjugate Gradient Formula

陈 海

CHEN Hai

(广西大学数学与信息科学学院,广西南宁 530004)

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

摘要:给出一个修改的三项共轭梯度算法,证明其具有充分下降性和全局收敛性,其搜索方向拥有梯度值信息和函数值信息,并用数值算例检验算法是可行的.

关键词:共轭梯度法 充分下降 全局收敛

中图法分类号:O224 **文献标识码**:A **文章编号**:1005-9164(2012)04-0319-04

Abstract: A three-terms conjugate gradient algorithm is proposed. The search direction possesses automatically the sufficient descent property. Moreover, the search direction has not only the gradient information but also function information. The global convergence was established. Numerical results about engineering problems are reported.

Key words: conjugate gradient method, sufficient descent, global convergence

对于无约束最优化问题

$$\min f(x)$$

$$x \in R^n, \tag{0.1}$$

其中 $f(x)$ 是连续可微函数,有很多求解方法,其中共轭梯度方法结构简单,存储量小,求解效果明显. 其迭代公式如下

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

其中 x_k 是第 k 迭代点,$\alpha_k > 0$ 是步长,d_k 定义为下述形式的搜索方向

$$d_k = \begin{cases} -g_k + \beta_k d_{k-1}, \text{if} \quad k \geqslant 1, \\ -g_k, \text{if} \quad k = 0, \end{cases} \tag{0.2}$$

其中 $\beta_k \in R$ 是一个参数. 由于 β_k 的选取不同共轭梯度法被称为不同的方法[1~5],PRP 方法[3,5] 是其中最为著名的一个共轭梯度算法,其公式定义如下

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

其中 g_k 和 g_{k+1} 是函数 $f(x)$ 在 x_k 和 x_{k+1} 的梯度值,分别表示成 $\nabla f(x_k)$ 和 $\nabla f(x_{k+1})$. $\|\cdot\|$ 表示欧氏向量范数. PRP 方法数值表现优越,常被人们用于解决实际问题,虽然该方法收敛性不够理想,但仍然是众多学者研究的热点. 人们希望能找到数值表现与 PRP 相当但性质更优越的方法[6~14]. 研究已发现,PRP 方法收敛性不好的主要原因是其不具有充分下降性. 所谓充分下降性,是指对所有的 k,使得下式成立

$$d_k^T g_k \leqslant -c\|g_k\|^2. \tag{0.3}$$

优化算法中步长 α_k 的求解方法很多,其中如下的线搜索技术,比较常见.

$$f(x_k + \alpha_k d_k) \leqslant f_k + \delta\alpha_k g_k^T d_k, \tag{0.4}$$

$$g(x_k + \alpha_k d_k)^T d_k \geqslant \sigma g_k^T d_k, \tag{0.5}$$

式中 $\delta \in (0,1/2)$,$\sigma \in (\delta,1)$,被称为 WWP 线搜索. 迄今为止,对一般函数而言,PRP 方法在 WWP 线搜索下的全局收敛性还没有得到证明.

2006 年 Zhang 等[15] 给出一个三项共轭梯度公式

$$d_{k+1} = \begin{cases} -g_{k+1} + \beta_k^{\text{PRP}} d_k - \vartheta_k y_k, \text{if} \quad k \geqslant 1, \\ -g_{k+1}, \text{if} \quad k = 0, \end{cases} \tag{0.6}$$

其中 $\vartheta_k = \dfrac{g_{k+1}^T d_k}{\|g_k\|^2}$,$\beta_k^{\text{PRP}} = \dfrac{g_{k+1}^T y_k}{\|g_k\|^2}$,$y_k = g_{k+1} - g_k$. 容易得到 $d_k^T g_k = -\|g_k\|^2$. 从(0.6)式可以看出此方法利用了梯度值信息,却忽略了函数值信息. 1999 年文献[16]给出过一个新的拟牛顿方程

$$B_{k+1} S_k = y_k^1, \tag{0.7}$$

收稿日期:2012-03-07

作者简介:陈 海(1969-),男,讲师,主要从事管理与决策,继续教育与职业教育研究。

其中

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

$s_k = x_{k+1} - x_k$，B_{k+1} 是拟牛顿矩阵. 而我们知道，实际的数值结果和理论性质表明，拥有函数值和梯度值的公式比只有梯度值的公式具有更好的结论. 基于此，本文把该公式中 y_k^1 替换(0.6)式中的 y_k，得到一个修改的三项共轭梯度算法：

$$d_{k+1} = \begin{cases} -g_{k+1} + \beta_k^{\mathrm{MPRP}} d_k - \vartheta_k y_k^1, \text{if} \quad k \geqslant 1, \\ -g_{k+1}, \text{if} \quad k = 0, \end{cases} \tag{0.8}$$

其中 $\vartheta_k = \frac{g_{k+1}^T d_k}{\|g_k\|^2}$，$\beta_k^{\mathrm{MPRP}} = \frac{g_{k+1}^T y_k^1}{\|g_k\|^2}$，$y_k = g_{k+1} - g_k$，并证明了算法具有充分下降性和收敛性.

1 梯度算法步骤

算法 1 （修改的三项共轭梯度法）

初始步 给定 $x_0 \in R^n$，$\delta \in (0, 1/2)$，$\sigma \in (\delta, 1)$ 和参数 $\varepsilon > 0$. 令 $d_0 = -g_0 = -\nabla f(x_0)$，置 $k := 0$.

步骤 1 若 $\|g_k\| \leqslant \varepsilon$，停止，否则转下一步骤.

步骤 2 寻找步长 α_k 满足(0.4)式和(0.5)式.

步骤 3 令 $x_{k+1} = x_k + \alpha_k d_k$. 若 $\|g_{k+1}\| \leqslant \varepsilon$，停止，否则转步骤 4.

步骤 4 利用公式(0.8)计算搜索方向.

步骤 5 令 $k := k+1$，转步骤 2.

2 梯度算法的充分下降性和全局收敛性

引理 2.1 对 $k \geqslant 0$，修改的三项公式具有充分下降性，即满足

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2. \tag{2.1}$$

证明 若 $k = 0$，则 $g_0^T d_0 = -\|g_0\|^2$，(2.1)式满足. 当 $k \geqslant 1$ 时，利用新公式的定义得到

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

所以(2.1)式成立.

再假设下述条件成立.

(i) 水平集 $\Omega = \{x \in R^n \mid , f(x) \leqslant f(x_0)\}$ 有界.

(ii) f 在 Ω 上有下界且连续可微，梯度 g 满足 Lipschitz 条件，即存在常数 $M > 0$ 使得

$$\|g(x) - g(y)\| \leqslant M\|x - y\|, \forall, x, y \in \Omega. \tag{2.2}$$

引理 2.2 设假设条件(i)和(ii)满足，且存在常数 $\gamma > 0$ 使得 $\|d_k\| \leqslant \gamma\|g_k\|$，则 $\lim_{k\to\infty}\|g_k\| = 0$ 成立.

证明 利用(2.2)式和(0.5)式，得到

$$-(1-\sigma)g_k^T d_k \leqslant (g_{k+1} - g_k)^T d_k \leqslant \|g_{k+1} - g_k\| \|d_k\| \leqslant \alpha_k M \|d_k\|^2.$$

再由(0.3)式和 $\|d_k\| \leqslant \gamma\|g_k\|$ 得

$$\alpha_k \geqslant \frac{(1-\sigma)}{M} \frac{|g_k^T d_k|}{\|d_k\|^2} \geqslant \frac{1-\sigma}{M\gamma^2},$$

将上式代入(0.4)式并移项整理得

$$\frac{\delta(1-\sigma)}{M\gamma^2} | g_k^T d_k | \text{leq } f_k - f_{k+1}.$$

从 $k = 0$ 到 ∞ 相加，有

$$\sum_{k=0}^{\infty} | g_k^T d_k | \leqslant \frac{M\gamma^2}{\delta(1-\sigma)}(f_0 - f_\infty).$$

利用假设条件(函数 f 有下界)，得

$$\sum_{k=0}^{\infty} | g_k^T d_k | < +\infty.$$

再次利用(2.1)式，可得

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

则 $\|g_k\| \to 0$ 成立.

3 数值检验

检验函数选择工程领域经常用到的 Benchmark 问题(http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume24/ortizboyer05a-html/node6.html.)：

(1) Sphere function

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

(2) Schwefel's function

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

(3) Rastrigin function

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

(4) Schwefel function

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

(5) Griewank function

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

表 1　算法数值结果

Table 1　Numerical results

Function	x_0	dim	$NI/NFG/f(\bar{x})$
Sphere	(−6,−6,…,−6)	10	2/6/7.888609e−030
		100	2/6/3.818087e−026
		300	2/6/2.735770e−025
	(−4,−4,…,−4)	10	2/6/7.888609e−030
		100	2/6/1.262177e−027
		300	2/6/2.366583e−026
	(−2,−2,…,−2)	10	3/7/1.972152e−030
		100	2/6/3.993608e−028
		300	2/6/7.158913e−027
	(3,3,…,3)	10	3/7/1.972152e−030
		100	2/6/4.930381e−028
		300	2/6/2.366583e−026
	(5,5,…,5)	10	2/6/7.888609e−030
		100	2/6/3.478877e−026
		300	2/6/1.990296e−025
Schwefel′s	(−0.0005,…,−0.0005)	10	3/8/8.849943e−008
		50	5/14/5.743629e−007
		100	6/17/1.609320e−006
	(−0.0003,…,−0.0003)	10	3/8/3.185979e−008
		50	4/11/7.811524e−007
		100	5/14/1.700332e−006
	(0.0005,…,0.0005)	10	3/8/8.849943e−008
		50	5/14/5.743629e−007
		100	6/17/1.609320e−006
	(0.0009,…,0.0009)	10	3/8/2.867381e−007
		50	5/14/1.860936e−006
		100	7/20/2.069193e−006
	(0.001,…,0.001)	10	3/8/3.539977e−007
		50	5/14/2.297452e−006
		100	7/20/2.554559e−006
Rastrigin	(−7,−7,…,−7)	10	14/86/3.979831e+001
		100	3/9/4.875199e+003
		300	9/54/4.775773e+003
	(−6,−6,…,−6)	10	4/13/3.581799e+002
		100	3/9/3.581799e+003
		300	3/9/1.074540e+004
	(2,2,…,2)	10	4/13/3.979831e+001
		100	3/9/3.979831e+002
		300	3/8/1.193950e+003
	(3,3,…,3)	10	4/13/8.954601e+001
		100	3/9/8.954601e+002
		300	3/9/2.686380e+003
	(5,5,…,5)	10	4/13/2.487372e+002
		100	3/9/2.487372e+003
		300	3/9/7.462117e+003
Schwefel	(−200,−200,…,−200)	10	4/12/1.529646e+002
		100	4/12/1.529646e+003
		300	4/12/4.588939e+003
	(−100,−100,…,−100)	10	8/23/1.529646e+002
		100	8/23/1.529647e+003
		300	8/23/4.588940e+003
	(100,100,…,100)	10	2/19/3.101787e+003
		100	2/19/3.101787e+004
		300	2/19/9.305360e+004

续表 1

Continue table 1

Function	x_0	dim	$NI/NFG/f(\bar{x})$
	(250,250,…,250)	10	2/19/3.673625e+003
		100	2/19/3.673625e+004
		300	2/19/1.102088e+005
	(300,300,…,300)	10	2/21/−1.815299e+003
		100	2/21/−1.819761e+004
		300	2/21/−5.460676e+004
Griewank	(−60,−60,…,−60)	10	5/50/2.274836e+000
		100	4/31/3.621758e−002
		300	60/168/6.396639e−006
	(−20,−20,…,−20)	10	12/52/1.490157e−006
		100	2/6/0.000000e+000
		300	2/6/0.000000e+000
	(2,2,…,2)	10	2/19/1.012130e+000
		100	2/6/0.000000e+000
		300	2/6/0.000000e+000
	(25,25,…,25)	10	2/19/2.562209e+000
		100	5/30/1.101344e+000
		300	4/14/0.000000e+000
	(35,35,…,35)	10	9/52/3.233641e−001
		100	5/16/1.030436e−008
		300	2/6/1.433298e−013

参数选取：$\varepsilon = 10^{-5}$，$\delta = 0.1$，$\sigma = 0.9$. 采用 Himmeblau 停止准则：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，$e_1 = e_2 = 10^{-5}$. 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 能有效地求解 Benchmark 问题，且随着维数增加迭代次数和函数值次数没有太大变化，随初始点不同，结果变化也不太大，说明算法的稳定性较好.

致谢：

感谢广西大学数学与信息科学学院韦增欣教授和袁功林教授为本文提出修改意见，同时感谢他们在数值实验方面给予的帮助！

参考文献：

[1] Fletcher R，Reeves C. Function minimization buconjugate gradients[J]. Compute J，1964，7：149-154.

[2] Liu Y，Storey C. Effcient generalized conjugate gradient algorithms，part 1：theory[J]. Journal of Optimization Theory and Application，1992，69：17-41.

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

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

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

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

[7] Hager W W, Zhang H. Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent[J]. ACM Transactions on Mathematical Software, 2006, 32: 113-137.

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

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

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

[11] Wei Z, Yao S, Liu L. The convergence properties of some new conjugate gradient methods[J]. Applied Mathematics and Computation, 2006, 183: 1341-1350.

[12] Yuan G. Modified nonlinear conjugate gradient methods withsufficient descent property for large-scale optimization problems[J]. Optimization Letters, 2009, 3: 11-21.

[13] Yuan G, Lu X. A modified PRP conjugate gradient method[J]. Annals of Operations Research, 2009, 166: 73-90.

[14] Yuan G, Lu X, Wei Z. A conjugate gradient method with descent direction for unconstrained optimization[J]. Journal of Computational and Applied Mathematics, 2009, 233: 519-530.

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

[16] Zhang J, Deng N, Chen L. New quasi-Newton equation andrelated methods for unconstrained optimization[J]. Journal of Optimization Theory and Application, 1999, 102: 147-167.

（责任编辑：尹　闯）

（上接第 315 页 Continue from page 315）

参考文献：

[1] 张忠志，胡锡炎，周福照. D 对称矩阵反问题的最小二乘解[J]. 湖南大学学报：自然科学版，2001，28(5)：6-10.

[2] 周立平. 几类线性矩阵方程的加权最小二乘解及其最佳逼近问题[D]. 长沙：湖南大学，2007：1-48.

[3] 王明辉，魏木生. 关于矩阵方程 $AXB=E$ 的加权最小二乘 Hermite 解[J]. 计算数学，2004，26(2)：129-136.

[4] 彭向阳，胡锡炎，张磊. 矩阵方程的正交对称与正交反对称最小二乘解[J]. 工程数学学报，2006(6)：1048-1052.

[5] 郁金祥. 一类线性流形上矩阵方程 $X^{T}AX=B$ 的反问题[J]. 数学理论与应用，2008，28(2)：38-42.

[6] 彭向阳，张磊，胡锡炎. 矩阵方程 $A^{T}XA=B$ 的反对称正交反对称最小二乘解[J]. 工程数学学报，2004，21(8)：93-97.

（责任编辑：尹　闯）