第 10 卷　第 19 期　2010 年 7 月
1671-1815(2010)19-4604-04
科　学　技　术　与　工　程
Scie nce Technology and Engineering
Vol. 10　No. 19　July 2010

一类新的混合共轭梯度算法

刘玉建　黄炳家
(中国石油大学(华东)数学与计算科学学院,东营 257061)

摘　要　共轭梯度算法在无约束最优化问题中有着广泛应用。现给出的一类新的共轭梯度算法,在迭代过程中保持了下降性质;在一般 Wolfe 线搜索条件下,新算法是全局收敛的。

关键词　无约束最优化　共轭梯度法　Wolfe 线搜索　下降方向　全局收敛性
中图法分类号　O224;　　**文献标志码**　A

考虑无约束最优化问题

$$\min f(x), x \in \boldsymbol{R}^n \tag{1}$$

式(1)中 $f: R \rightarrow R^n$ 是是连续可微的函数。共轭梯度算法是求解问题(1)的非常有效的算法,特别是对于大规模的无约束优化问题。由于存储量小,算法简便,共轭梯度算法的优势更加明显。

共轭梯度算法的一般迭代公式为

$$x_{k+1} = x_k + \alpha_k d_k \tag{2}$$

$$d_k = \begin{cases} -g_k, k=1 \\ -g_k + \beta_k d_{k-1}, k \geqslant 2 \end{cases} \tag{3}$$

其中,$g_k = \nabla f(x_k)$ 为函数 f 的梯度,α_k 为搜索步长,通过一维搜索得到,β_k 为一标量,不同的 β_k 构成不同的共轭梯度算法。著名的算法有 ${\beta_k}^{HS}$,${\beta_k}^{FR}$,${\beta_k}^{PRP}$,${\beta_k}^{LS}$,${\beta_k}^{DY}$等。这些算法的总体收敛性已经得到了广泛的研究,这其中包括 Zoutendijk[1], Al - Baali[2], Gilbert 和 Nocedal[3],戴彧虹和袁亚湘[4]以及 Powell[5]等。为确定这些法的收敛性结果,一般要求步长 α_k 满足强 Wolfe 线搜索条件,即

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) + \rho \alpha_k g_k^{\mathrm{T}} d_k \tag{4}$$

$$|g(x_k + \alpha_k d_k)^{\mathrm{T}} d_k| \leqslant -\sigma g_k^{\mathrm{T}} d_k \tag{5}$$

这里 $0 < \rho < \sigma < 1$。有时甚至需要 α_k 满足精确线搜索条件,即

$$f(x_k + \alpha_k d_k) = \min_{\alpha > 0} f(x_k + \alpha d_k) \tag{6}$$

1995 年戴彧虹和袁亚湘[6]提出了 DY 方法,即公式 $\beta_k^{DY} = \dfrac{\| g_k \|^2}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})}$,证明了在一般 Wolfe 线搜索条件即式(4)和

$$g(x_k + \alpha_k d_k)^{\mathrm{T}} d_k > \sigma g_k^{\mathrm{T}} d_k \tag{7}$$

下,DY 方法每一步均产生一个下降方向,并证明了算法的全局收敛性。

众所周知,共轭梯度算法的良好的收敛性与数值表现之间并没有必然的联系,例如 FR 虽然有很好的收敛性质,但数值表现往往较差。Powell[7]证明了在某些情况下,即使采用精确步长搜索,FR 也会产生非常小的步长,除非采用重开始策略,否则将很难摆脱这种困境;另一方面,例如 PRP 和 HS 等方法,虽然对一般函数不具备收敛性,但在实际计算当中却往往有好的数值表现。这启发我们,对不同的共轭梯度算法取长补短,采用混合策略结合研究,或许能得到令人满意的结果,这也是近些年共轭梯度领域的一个研究热点。Gilbert 和 Nocedal[3], Touati-ahmed 和 Storey[8]以及 Hu 和 Storey[9]在这方面做出了突出贡献。郑希峰[10]提出了一种新的共轭梯度算法

$$\beta_k = \frac{g_k^{\mathrm{T}}(g_k - d_{k-1})}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})}。$$

并与 DY 方法结合,提出了一种混合的共轭梯度算法

2010 年 4 月 9 日收到
第一作者简介:刘玉建(1980—),男,山东滨州人,硕士研究生,研究方向:最优化理论和方法。E-mail:liuyujianqufu@163.com。

$$\beta_k^{\text{NEW}} = \begin{cases} \dfrac{g_k^{\mathrm{T}}(g_k - d_{k-1})}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})}, \\ \qquad 0 < G_k^{\mathrm{T}} d_{k-1} < \min\left(2, \dfrac{1}{\sigma}\right) \| g_k \|^2; \\ \beta_k^{\text{NEW}}, \quad 其他。 \end{cases}$$

并在一般 Wolfe 线搜索条件下证明了新算法的全局收敛性。本文在此基础上,给出一种新的公式

$$\beta_k^{NF} = \frac{g_k^{\mathrm{T}}(g_k - \lambda d_{k-1})}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})} \tag{8}$$

这里, $0 < \lambda < 1$。

1　新公式的的两个简单性质

定理 1　考虑方法式(2)~式(3),步长 α_k 由 Wolfe 线搜索式(4)、式(7)得到,当 β_k 取为式(8)时,对所有的 $k \geqslant 1$,我们有 $\mathrm{g}_k^{\mathrm{T}}\mathrm{d}_k < 0$。

证明　用数学归纳法来证明

当 $k=1$ 时,有 $\mathrm{g}_1^{\mathrm{T}}\mathrm{d}_1 = -\|\mathrm{g}_1\|^2 < 0$,显然结论成立。

假设结论对 $k-1$ 成立,即有 $g_{k-1}^{\mathrm{T}} d_{k-1} < 0$,那么由式(3)和式(8),有

$$\begin{aligned} g_k^{\mathrm{T}} d_k &= g_k^{\mathrm{T}}(-g_k + \beta_k d_{k-1}) = \\ &-\|g_k\|^2 + \frac{g_k^{\mathrm{T}}(g_k - \lambda d_{k-1})}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})} g_k^{\mathrm{T}} d_{k-1} = \\ &\frac{\|g_k\|^2 g_{k-1}^{\mathrm{T}} d_{k-1} - \lambda \|g_k^{\mathrm{T}} d_{k-1}\|^2}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})} \end{aligned} \tag{9}$$

由式(7)可知,

$$\begin{aligned} d_{k-1}^{\mathrm{T}}(g_k - g_{k-1}) &= g_k^{\mathrm{T}} d_{k-1} - g_{k-1}^{\mathrm{T}} d_{k-1} > \\ &(\sigma - 1) g_{k-1}^{\mathrm{T}} d_{k-1} \end{aligned} \tag{10}$$

由归纳假设及 $0 < \sigma < 1$ 可得

$$d_{k-1}^{\mathrm{T}}(g_k - g_{k-1}) > 0 \tag{11}$$

联系式(10),可知 $g_k^{\mathrm{T}} d_k < 0$,从而定理成立。

定理 2　如果目标函数 $f(x)$ 是严格凸的,考虑方法式(2)~式(3)当 β_k

取为式(8)时,对所有的 $k \geqslant 1$,我们有 $g_k^{\mathrm{T}} d_k < 0$。

证明　由于函数 $f(x)$ 是严格凸的,我们有 $(g_k - g_{k-1})^{\mathrm{T}} \alpha_{k-1} d_{k-1} > 0$,又 $\alpha_{k-1} > 0$,由定理 1 的证明过程,结论成立。

2　新的混合共轭梯度算法及其全局收敛性

下面我们结合 DY 方法并利用上面给出的新公式(8),讨论一种新的混合共轭梯度算法。公式如式(2)所示。

$$\beta_k^{\text{NCG}} = \begin{cases} \dfrac{g_k^{\mathrm{T}}(g_k - \lambda d_{k-1})}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})}, 0 < \lambda g_k^{\mathrm{T}} d_{k-1} < \|g_k\|^2 \\ \beta_k^{\text{DY}}, \qquad 其他 \end{cases} \tag{12}$$

式(1)中 $0 < \lambda < 1$。新的算法步骤如下:

步 1　给定初始值 $x_1 \in R^n$, $\varepsilon > 0$, $d_1 = -g_1$,置 $k := 1$;

步 2　如果 $\|g_k\| < \varepsilon$,则停止;否则,由 Wolfe 准则式(4)、式(7),计算 α_k,令 $x_{k+1} = x_k + \alpha_k d_k$;

步 3　计算 g_{k+1},若 $\|g_{k+1}\| < \varepsilon$,则停止;否则,令

$$d_{k+1} = -g_{k+1} + \beta_{k+1} d_k。$$

其中,β_{k+1} 由式(12)计算所得. 令 $k := k+1$,转步 2。

在讨论算法的全局收敛性之前,我们需要做出假设(H):

(i) 水平集 $L = \{x \in R^n : f(x) < f(x_1)\}$ 有界,其中 x_1 为初始点。

(ii) 在水平集 L 的某个领域 N 内,f 连续可微,其导函数 g 满足 Lipschitz 条件,即存在常数 $M > 0$,使得:

$$\|g(x) - g(y)\| \leqslant M \|x - y\|, \ \forall x, y \in \mathbf{N}。$$

引理 1[11]　假设目标函数满足假设 H,考虑方法式(2)~式(3),其中 d_k 满足 $g_k^{\mathrm{T}} d_k < 0$,步长 α_k 由 Wolfe 线搜索式(4)、式(7)得到,则有

$$\sum_{k \geqslant 1} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} < +\infty。$$

引理 2　假设目标函数满足假设 H,考虑方法式(2)~式(3),步长 α_k 由 Wolfe 线搜索式(4)、式(7)得到,β_k 取为式(11),则对所有的 $k \geqslant 1$,我们有 $g_k^{\mathrm{T}} d_k < 0$。

证明　用数学归纳法来证明:

当 $k=1$ 时,有 $g_1^{\mathrm{T}} d_1 = -\|g_1\|^2 < 0$,显然结论

成立。

假设结论对 $k-1$ 成立,即有 $g_{k-1}^{\mathrm{T}}d_{k-1}<0$,下面分两种情况来证明:

1) 若 $\beta_k^{NCG}=\beta_k^{DY}=\dfrac{\|g_k\|^2}{d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})}$,由式(3),有 $g_k^{\mathrm{T}}d_k=g_k^{\mathrm{T}}(-g_k+\beta_k d_{k-1})=$

$$-\|g_k\|^2+\frac{\|g_k\|^2}{d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})}g_k^{\mathrm{T}}d_{k-1}=\frac{\|g_k\|^2}{d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})}g_{k-1}^{\mathrm{T}}d_{k-1} \tag{13}$$

由式(7)可知,

$$d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})=g_k^{\mathrm{T}}d_{k-1}-g_{k-1}^{\mathrm{T}}d_{k-1}>(\sigma-1)g_{k-1}^{\mathrm{T}}d_{k-1} \tag{14}$$

联系式(9),可知 $g_k^{\mathrm{T}}d_k<0$,从而定理成立。

2) 若 $\beta_k^{NCG}=\dfrac{g_k^{\mathrm{T}}(g_k-\lambda d_{k-1})}{d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})}$,则由定理1,可知 $g_k^{\mathrm{T}}d_k<0$ 成立。

由1)和2)及数学归纳法可知定理结论成立。

引理3 假设目标函数满足假设H,考虑方法式(2)~式(3),步长 α_k 由Wolfe线搜索式(4)、式(7)得到,β_k 取为式(12),则对所有的 $k\geqslant1$,有 $|\beta_k|\leqslant\beta_k^{DY}$。

证明 由式(11),若 $0<\lambda g_k^{\mathrm{T}}d_{k-1}<\|g_k\|^2$

$$\tag{15}$$

则有

$$\beta_k^{NCG}=\frac{g_k^{\mathrm{T}}(g_k-\lambda d_{k-1})}{d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})}=\frac{|g_k|^2}{d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})}-\frac{\lambda g_k^{\mathrm{T}}d_{k-1}}{d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})}=\beta_k^{DY}-\frac{\lambda g_k^{\mathrm{T}}d_{k-1}}{d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})} \tag{16}$$

由式(15),有

$$\frac{-2\|g_k\|^2}{d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})}<-\frac{\lambda g_k^{\mathrm{T}}d_{k-1}}{d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})}<0。$$

联系式(16)和式(12),可得

$$-\beta_k^{DY}<\beta_k^{NCG}\leqslant\beta_k^{DY}。$$

即有 $|\beta_k|\leqslant\beta_k^{DY}$ 成立。

定理3 假设目标函数满足假设H,考虑方法式(2)~式(3),步长 α_k 由Wolfe线搜索式(4)、式(7)得到,β_k 取为式(12),则有 $\liminf\limits_{k\to\infty}\|g_k\|=0$。

证明 用反证法

假设定理结论不成立,则对所有的 $k\geqslant1$,存在常数 $m>0$,使得

$$\|g_k\|\geqslant m \tag{17}$$

成立。

我们把式(3)重新写为 $d_k+g_k=\beta_k d_{k-1}$ (18)

对式(3)两边平方并移项有 $\|d_k\|^2=\beta_k^2\|d_{k-1}\|^2-2g_k^{\mathrm{T}}d_k-\|g_k\|^2$ (19)

由引理3,式(19)可变为

$$\|d_k\|^2\leqslant(\beta_k^{DY})^2\|d_{k-1}\|^2-2g_k^{\mathrm{T}}d_k-\|g_k\|^2 \tag{20}$$

由式(13)可得

$$\beta_k^{DY}=\frac{\|g_k\|^2}{d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})}=\frac{g_k^{\mathrm{T}}d_k}{g_{k-1}^{\mathrm{T}}d_{k-1}} \tag{21}$$

对式(20)两边同除以 $(g_k^{\mathrm{T}}d_k)^2$ 并利用式(21),则式(20)可变为

$$\frac{\|d_k\|^2}{(g_k^{\mathrm{T}}d_k)^2}\leqslant\frac{\|d_{k-1}\|^2}{(g_{k-1}^{\mathrm{T}}d_{k-1})^2}-\frac{2}{g_k^{\mathrm{T}}d_k}-\frac{\|g_k\|^2}{(g_k^{\mathrm{T}}d_k)^2}=\frac{\|d_{k-1}\|^2}{(g_{k-1}^{\mathrm{T}}d_{k-1})^2}-\left(\frac{1}{\|g_k\|}+\frac{\|g_k\|}{g_k^{\mathrm{T}}d_k}\right)^2+\frac{1}{\|g_k\|^2}\leqslant\frac{\|d_{k-1}\|^2}{(g_{k-1}^{\mathrm{T}}d_{k-1})^2}+\frac{1}{\|g_k\|^2} \tag{22}$$

注意到 $\dfrac{\|d_1\|^2}{(g_1^{\mathrm{T}}d_1)^2}=\dfrac{1}{\|g_1\|^2}$,则有

$$\frac{\|d_k\|^2}{(g_k^{\mathrm{T}}d_k)^2}\leqslant\frac{\|d_{k-1}\|^2}{(g_{k-1}^{\mathrm{T}}d_{k-1})^2}+\frac{1}{\|g_k\|^2}\leqslant\frac{\|d_{k-2}\|^2}{(g_{k-2}^{\mathrm{T}}d_{k-2})^2}+\frac{1}{\|g_{k-1}\|^2}+\frac{1}{\|g_k\|^2}\leqslant\cdots\cdots\leqslant\sum_{i=1}^{k}\frac{1}{\|g_i\|^2} \tag{23}$$

由式(23)及式(17),有

$$\frac{\|d_k\|^2}{(g_k^{\mathrm{T}}d_k)^2}\leqslant\sum_{i=1}^{k}\frac{1}{\|g_i\|^2}\leqslant\frac{k}{m}。$$

从而有 $\dfrac{(g_k^{\mathrm{T}}d_k)^2}{\|d_k\|^2}\geqslant\dfrac{m}{k}$。

因此有 $\sum\limits_{k=1}^{\infty}\dfrac{(g_k^{\mathrm{T}}d_k)^2}{\|d_k\|^2}=\infty$,与引理1矛盾,从而定理成立。

3 结论

研究了一类混合共轭梯度算法,对原算法进行了改进,扩大了搜索方向的范围,并证明了新算法的下降性质和收敛性。我们知道 FR 与 CD 方法也具有比较良好的收敛性质,把新算法中的 DY 方法部分换为上述两种方法或是与他们进行结合研究,能够得到何种结果还有待于验证和计算。

参 考 文 献

1 Zoutendijk G. Nonliner programming, computation methods. In: Integer and Nonlinear Programming(Abadie, J. ed.), 1970: 37—86

2 Al-Baali M. Descent property and global convergence of the Fletcher-Reeves method with in-exact line search. IMA J Numer Anal, 1985; 5: 121—124

3 Gilbert J C, Nocedal J. Global convergence properties of conjugate gradient methods for optimization. SIAM J Optimization, 1992; 2 (1): 21—42

4 Dai Y H, Yuan Y X. Convergence properties of the Fletcher-Reeves method. IMA Journal Numerical Analysis, 1996; 16(2): 155—164

5 Powell M J D. Nonconvex minimization calculations and the conjugate gradient method. In: David F G. Numerical Analysis: Proceedings of the 10th Biennial Conference held at Dundee, scotland, June 28-July 1, 1983, Berlin: Springer Verlag, 1984: 122—141

6 Dai Y H, Yuan Y X. A nonlinear conjugate gradient method with a nice global convergence property. Research report ICM95-038, Institute of Computational Mathematics and Scientifiac Engineering Computing, Chinese Academy of Sciences, Beijing: 1995

7 Powell M J D. Restart procedures for the conjugate gradient method, Math Programming, 1977; 12: 241—254

8 Touati-ahmed D , Storey C. Efficient hybrid conjugate gradient techniques. Journal of Optimization Theory and Applications, 1990; 64 (2): 379—397

9 Hu Y F , Storey C. Global convergence result for conjugate gradient methods. Journal of Optimization Theory and Applications , 1991; 71 (2): 399—405

10 Zhen X F, Tian Z Y, Song L W. The global convergence of a mixed conjugate gradient method with the Wolfe line search. Operations Research Transactions, 2009; 13(2): 18—24

11 Dai Y H, Yuan Y X. An effcient hybrid conjugate gradient methodfor unconstrained optimization. Annals of Operations Research, 2001; 103: 33—47

A New Hybrid Congjugate Gradient Method

LIU Yu-jian, HUANG Bing-jia

(College of Mathematics and Computation Science , China University of Petroleum , Dongying 257061, P. R. China)

[**Abstract**] Conjugate gradient methods are widely used for unconstrained optimization. A class of new conjugate gradient method is presented. The new method produces a decent search direction at every iteration and converges globally provided that the line search satisfies the weak Wolfe conditions.

[**Key words**] unconstrained optimization conjugate gradient method Wlofe line search decent direction global convergence