doi:10.16055/j.issn.1672-058X.2015.0009.001

一个带有干扰因子修正 PRP 共轭梯度法的全局收敛性*

陈 洪 敏
(重庆师范大学 数学学院,重庆 401331)

摘　要:非线性共轭梯度法是解决大规模优化问题的一种非常有效的方法,提出了一个修正的 PRP 方法,该参数带有干扰因子,并证明了这一新的参数具有非负性,且在适当条件下,采用 Wolfe 线搜索,证明该算法具有全局收敛性.

关键词:共轭梯度法;强 Wolfe 条件;Wolfe 条件;干扰因子;全局收敛性

中图分类号:O224　　**文献标志码**:A　　**文章编号**:1672-058X(2015)09-0001-04

1　基础知识

考虑如下的无约束极小化问题:

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

其中,$f:\mathbf{R}^n \to \mathbf{R}$ 是连续可微函数,且其梯度可获得.

非线性共轭梯度法是求解此类问题的一种常用的有效方法,具有算法简便、存储需求小等优点,适用于求解大规模无约束优化问题.共轭梯度法的迭代格式为

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

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

其中 d_k 为搜索方向,$\nabla f(x_k)$ 简记为 g_k,β_k 是一个参数,不同的 β_k 对应着不同的共轭梯度法,α_k 是通过适当的线搜索获得的步长,通常使用 Wolfe 线搜索、强 Wolfe 线搜索.

Wolfe 线搜索条件为

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

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

强 Wolfe 线搜索条件为

$$|g(x_k + \alpha_k d_k)^{\mathrm{T}} d_k| \geqslant -\sigma g_k^{\mathrm{T}} d_k \tag{6}$$

其中 $0<\delta<\sigma<1$.

传统共轭梯度法的参数公式 β_k 的计算公式有 Hestenes-Stiefel(HS)[1],Fletcher-Reeves(FR)[2],Polak-

收稿日期:2014-12-11;**修回日期**:2015-02-24.
* **基金项目**:国家自然科学基金(10771003).
作者简介:陈洪敏(1988-),女,安徽阜阳人,硕士研究生,从事最优化理论与算法研究.

Ribiere(PRP)[3],Dai-Yuan(DY)[4],它们的表达式分别为

$$\beta_k^{\mathrm{HS}}=\frac{g_k^{\mathrm{T}}(g_k-g_{k-1})}{d_k^{\mathrm{T}}(g_k-g_{k-1})},\beta_k^{\mathrm{FR}}=\frac{\|g_k\|^2}{\|g_{k-1}\|^2}$$

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

其中$\|\cdot\|$表示欧几里德范数.FR 方法具有很好的收敛性,且在强 Wolfe 线搜索下具有充分下降性,但因其在一些情况下可能会产生小步长情况,且这种情况会持续很多次迭代,除非重新开始,这就导致 FR 方法的数值表现不佳.PRP 和 HS 方法的数值表现非常相似,且都都优于 FR 方法,但对于非凸的目标函数,即使采用精确线性搜索,它们仍有可能不收敛.这问题启发人们找到新的共轭梯度法,使得这种方法不仅具有好的收敛性,而且具有很好的数值表现.

最近,文献[5]提出了关于β_k^{PRP}的一种修正方案,具体参数公式如式(7):

$$\beta_k=\frac{g_k^{\mathrm{T}}\left(\frac{\|g_{k-1}\|}{\|g_k\|}g_k-g_{k-1}\right)}{\|g_{k-1}\|^2}\tag{7}$$

这种方法在适当条件下,分别证明算法在 Wolfe 和 Grippo-Lucidi 线搜索下是全局收敛的.文献[6]通过对参数β_k加入干扰因子得到两个新的β_k,公式如式(8):

$$\beta_k^{\mathrm{MVPRP}}=\frac{g_k^{\mathrm{T}}\left(g_k-\frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{\eta_k\|g_{k-1}\|^2},\beta_k^{\mathrm{MN}}=\frac{g_k^{\mathrm{T}}\left(g_k-\frac{\|g_k\|}{\|d_{k-1}\|}d_{k-1}\right)}{\tau_k\|g_{k-1}\|^2}\tag{8}$$

其中$\eta_k=1-\nu\min\left\{0,\frac{g_k^{\mathrm{T}}g_{k-1}}{\|g_k\|\|g_{k-1}\|}\right\}(\nu\geqslant1)$,$\tau_k=1-\mu\min\left\{0,\frac{g_k^{\mathrm{T}}d_{k-1}}{\|g_k\|\|d_{k-1}\|}\right\}(\mu\geqslant1)$,这两类方法在强 Wolfe 线搜索下不仅产生充分下降方向,而且对于目标函数是一般的非凸函数,都是全局收敛的.受文献[5,6]的启发,此处提出一个新的参数β_k,公式如式(9):

$$\beta_k^{\mathrm{new}}=\frac{g_k^{\mathrm{T}}\left(\frac{\|g_{k-1}\|}{\|g_k\|}g_k-g_{k-1}\right)}{\eta_k\|g_{k-1}\|^2}\tag{9}$$

2 新方法(β_k^{new})的全局收敛性

文献[7]分析了 PRP 方法,认为限制其参数β_k的非负性不仅对方法的全局收敛性起到至关重要的作用,而且能够很容易地保证算法的下降性;文献[8]证明了$\beta_k=\max\{0,\beta_k^{\mathrm{PRP}}\}$在充分下降条件被满足,且步长$\alpha_k$满足 Wolfe 线搜索条件的前提下,算法具有全局收敛性.在接下来的引理中证明了β_k^{new}的非负性.

假设目标函数满足下面的假设:

1) 水平集$\Omega=\{x\in\mathbf{R}^n:f(x)\leqslant f(x_1)\}$是有界的;

2) 目标函数f在Ω的某个领域N是连续可微的,且梯度函数是 Lipschitz 连续的,即存在常数$L>0$使得

$$\|g(x)-g(\tilde{x})\|\leqslant L\|x-\tilde{x}\|\tag{10}$$

对任意的$x,\tilde{x}\in N$.

显然,由上面的假设可知存在一个常数$\bar{\gamma}$使得

$$\|g_k\|\leqslant\bar{\gamma}\tag{11}$$

对任意的 $x \in N$.

引理 1　对任意的 $k \geqslant 1$,都有 $\beta_k^{\text{new}} \geqslant 0$.

证明　记 θ_k 为 g_k 和 g_{k-1} 的夹角,即 $g_k^{\mathrm{T}} g_{k-1} = \|g_k\| \|g_{k-1}\| \cos \theta_k$.

$$\beta_k^{\text{new}} = \frac{g_k^{\mathrm{T}}\left(\frac{\|g_{k-1}\|}{\|g_k\|} g_k - g_{k-1}\right)}{\eta_k \|g_{k-1}\|^2} = \frac{\|g_k\|\|g_{k-1}\| - \|g_k\|\|g_{k-1}\| \cos \theta_k}{\eta_k \|g_{k-1}\|^2} =$$

$$\frac{1 - \cos \theta_k}{\eta_k} \cdot \frac{\|g_k\|}{\|g_{k-1}\|} = \frac{1 - \cos \theta_k}{1 - \nu \min\left\{0, \frac{g_k^{\mathrm{T}} g_{k-1}}{\|g_k\|\|g_{k-1}\|}\right\}} \cdot \frac{\|g_k\|}{\|g_{k-1}\|}$$

因为 $\nu \geqslant 1$,故

$$0 \leqslant \frac{1 - \cos \theta_k}{1 - \nu \min\{0, \cos \theta_k\}} = \left\{\begin{array}{l} \dfrac{1 - \cos \theta_k}{1 - \nu \cos \theta_k}, \text{若 } \cos \theta_k < 0 \\ 1 - \cos \theta_k, \text{若 } \cos \theta_k \geqslant 0 \end{array}\right\} \leqslant 1$$

所以 $\beta_k^{\text{new}} \geqslant 0$.

引理 2(性质 *)　考虑形如式(2)(3)的方法,并且假设对任意的 $k \geqslant 1$ 有 $0 \leqslant \gamma \leqslant \|g_k\| \leqslant \bar{\gamma}$,则存在常数 $b > 1, \lambda > 0$,使得对任意的 $k \geqslant 1$ 都有 $\|\beta_k^{\text{new}}\| \leqslant b$,且当 $\|s_{k-1}\| \leqslant \lambda$ 时有 $\beta_k^{\text{new}} \leqslant \frac{1}{2b}$,其中 $s_{k-1} = x_k - x_{k-1}$.

证明　取 $b = \frac{2\bar{\gamma}}{\gamma} > 2, \lambda = \frac{\gamma^3}{8L\bar{\gamma}^2} > 0$,则

$$\beta_k^{\text{new}} = \frac{g_k^{\mathrm{T}}\left(\frac{\|g_{k-1}\|}{\|g_k\|} g_k - g_{k-1}\right)}{\eta_k \|g_{k-1}\|^2} = \frac{1 - \cos \theta_k}{\eta_k} \cdot \frac{\|g_k\|}{\|g_{k-1}\|} \leqslant \frac{2\bar{\gamma}}{\gamma} = b$$

当 $\|s_{k-1}\| \leqslant \lambda$ 时,

$$\beta_k^{\text{new}} = \frac{g_k^{\mathrm{T}}\left(\frac{\|g_{k-1}\|}{\|g_k\|} g_k - g_{k-1}\right)}{\eta_k \|g_{k-1}\|^2} \leqslant \frac{\|g_k\| \left\| \frac{\|g_{k-1}\|}{\|g_k\|} g_k - g_{k-1} \right\|}{\|g_{k-1}\|^2} =$$

$$\frac{\|g_k\| \left\| \frac{\|g_{k-1}\|}{\|g_k\|} g_k - g_k + g_k - g_{k-1} \right\|}{\|g_{k-1}\|^2} \leqslant \frac{\|g_k\| (|\,\|g_{k-1}\| - \|g_k\|\,| + \|g_k - g_{k-1}\|)}{\|g_{k-1}\|^2} \leqslant$$

$$\frac{2\|g_k\|\|g_k - g_{k-1}\|}{\|g_{k-1}\|^2} \leqslant \frac{2\bar{\gamma} L \lambda}{\gamma^2} = \frac{1}{2b}$$

在共轭梯度法中,充分下降条件

$$g_k^{\mathrm{T}} d_k \leqslant - c \|g_k\|^2, c > 0 \tag{12}$$

非常重要,其中假设 new 算法满足式(12).

引理 3[8]**(zoutendijik 条件)**　若假设 1)2)成立,考虑形如式(2)(3)的方法,d_k 是下降方向,α_k 满足式(5)(6),则 $\sum\limits_{k \geqslant 1} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} < +\infty$.

引理 4[8]　若假设 1)2)成立,考虑形如式(2)(3)的方法,α_k 满足式(5)(6),且满足充分下降条件式(12),若 β_k 满足性质(*)且 $0 \leqslant \gamma \leqslant \|g_k\| \leqslant \bar{\gamma}$ 成立,则存在 $\lambda_1 > 0$,使得对任意的 $\Delta \in N^*$ 和下标 k_0,都存在指

标 $k \geqslant k_0$,满足 $|\kappa_{k,\Delta}^{\lambda_1}| > \dfrac{\Delta}{2}$,其中 $|\kappa_{k,\Delta}^{\lambda_1}|$ 表示 $\kappa_{k,\Delta}^{\lambda_1}$ 的个数.

引理 5[8] 若假设 1)2)成立,考虑形如式(2)(3)的方法,满足下面 3 个条件:(i)对任意的 $k \geqslant 1$,都有 $\beta_k \geqslant 0$;(ii)算法采用 Wolfe-power 线搜索,满足 zoutendijik 条件,满足充分下降条件式(12);(iii)满足性质(*),则 $\liminf\limits_{k \to \infty} \|g_k\| = 0$.

定理 1 若假设 1)2)成立,考虑形如式(2)(3)的方法,充分下降条件式(12)成立,α_k 满足式(5)(6),β_k 取 β_k^{new},则 $\liminf\limits_{k \to \infty} \|g_k\| = 0$.即 new 算法具有全局收敛性.

参考文献:

[1] HESTENES M R,STIEFEL E L.Method of Conjugate Gradient for Solving Linear Systems[J].Res Natl Bur Stand,1952(49):409-436

[2] FLETCHER R,REEVES C.Function Minimization by Conjugate Gradients[J].Comput,1964(7):149-154

[3] POLAK B,RIBIEREG.Note Surla Convergence Des Meethodes de Directions Conjugees[J].Rev Fr Inf Rech Oper,1964,3(1):35-43

[4] DAI Y H,YUAN Y X.A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property[J].SIAM Optim,1999(10):177-182

[5] 黎勇.一类修正 PRP 共轭梯度法的全局收敛性及其数值试验结果[J].西南大学学报:自然科学版,2011,33(11):23-28

[6] JIANG X Z, JIAN J B. Two Modified Nonlinear Conjugate Gradient Methods with Disturbance Factors for Unconstrained Optimization[J].Nonlinear Dynamics,2014,77(1-2):387-394

[7] POWELL M J D.NonconVex Minimization Calculations and the Conjugate Method[J].Lecture Notes in Mathematics,1984,10(66):122-141

[8] GLIBERT J C,NOCEDAL J.Global Convergence Properties of Conjugate Gradient Method for Optimization[J].SIAM Optim,1992(2):21-42

Global Convergence Property of a Modified PRP Conjugate Gradient Method with Disturbance Factor

CHEN Hong-min

(School of Mathematical Sciences,Chongqing Normal University,Chongqing 401331,China)

Abstract: The nonlinear conjugate gradient method is a very effective iterative method for solving large-scale optimal problems.In this paper,a modified PRP conjugate gradient method with disturbance factor is proposed and non-negative property of the formula is proved.Under suitable conditions,the global convergence of the algorithm with the Wolfe line search is discussed.

Key words: conjugate gradient method; strong Wolfe condition; Wolfe condition; disturbance factor; global convergence