doi:10.16055/j.issn.1672-058X.2015.0011.007

一个新的 PRP 共轭梯度法的全局收敛性

周雪琴

(贵州师范大学 数学与计算机科学学院,贵阳 550001)

摘　要:提出了一个不依赖线搜索且具有充分下降性的新的共轭梯度法(ZPRP 法),并证明了 ZPRP 方法在强 Wolfe 搜索条件下全局收敛.

关键词:共轭梯度法;充分下降性;强 Wolfe 线搜索;全局收敛性

中图分类号:O224　　**文献标志码**:A　　**文章编号**:1672-058X(2015)11-0031-03

1 基础知识

非线性共轭梯度法常用来解决下面的无限制优化问题:

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

其中,$f(x):\mathbf{R}^n\to\mathbf{R}$ 是连续可微函数,$g(x)$ 表示目标函数 $f(x)$ 的梯度.求解问题的共轭梯度法的一般迭代格式为

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

其中,α_k 是迭代步长且 $\alpha_k>0$,d_k 是搜索方向,x_{k+1} 是当前迭代点.搜索方向 d_k 由公式(3)确定:

$$d_k=\begin{cases}-g_0, & k=0\\ -g_k+\beta_k d_{k-1}, & k\geqslant 1\end{cases} \tag{3}$$

其中,β_k 是一个标量,g_k 表示 $g(x_k)$.

根据 β_k 的选取方式不同,将其分为不同的共轭梯度法.这里有一些比较著名的 β_k 公式,比如

$$\beta_k^{\mathrm{FR}}=\frac{\|g_k\|^2}{\|g_{k-1}\|^2};\quad \beta_k^{\mathrm{PRP}}=\frac{g_k^{\mathrm{T}}(g_k-g_{k-1})}{\|g_{k-1}\|^2}$$

$$\beta_k^{\mathrm{HS}}=\frac{g_k^{\mathrm{T}}(g_k-g_{k-1})}{d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})};\quad \beta_k^{\mathrm{LS}}=-\frac{g_k^{\mathrm{T}}(g_k-g_{k-1})}{d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})}$$

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

其中,$\|\cdot\|$表示欧式范数.它们分别称为 FR 算法、PRP 算法、HS 算法、LS 算法、CD 算法和 DY 算法.

搜索步长 α_k 可以通过两种不同的线搜索来计算,即精确线搜索和非精确线搜索.但由于精确线搜索的计算量过大,所以在实际计算过程中通常采取非精确线搜索来求 α_k.步长的选取满足一定的充分下降条件 $g_k^{\mathrm{T}}d_k\leqslant -c\|g_k\|^2$.此处采用强 Wolfe 搜索准则:

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

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

其中,$0<\delta\leqslant\sigma<1$.

收稿日期:2015-05-08;**修回日期**:2015-06-18.

作者简介:周雪琴(1990-),女,贵州铜仁人,硕士研究生,从事数值线性代数研究.

Dai 在文献[8]中提出了一个修正的 PRP 方法(记作 DPRP 方法),参数 β_k 定义如下:

$$\beta_k^{\text{DPRP}} = \frac{\|g_k\|^2 - \dfrac{\|g_k\|}{\|g_{k-1}\|}|g_k^{\mathrm{T}}g_{k-1}|}{\|g_{k-1}\|^2 + \mu|g_k^{\mathrm{T}}d_{k-1}|} \tag{6}$$

并且证明了 DPRP 方法不依赖线搜索具有充分下降性,且对 Armijo 线搜索和 Wolfe 线搜索具有全局收敛性.

Huang 在文献[9]中也提出了一种修正的共轭梯度法(记作 new 方法),参数 β_k 定义如下:

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

并且证明了这个新的方法在强 Wolfe 搜索条件下是是全局收敛的.

受到 Dai 和 Huang 二人的启发,现提出一种新的共轭梯度法(记作 ZPRP 方法),参数 β_k 定义如下:

$$\beta_k^{\text{ZPRP}} = \frac{\|g_k\|^2 - \dfrac{(g_k^{\mathrm{T}}g_{k-1})^2}{\|g_{k-1}\|^2}}{\|g_{k-1}\|^2 + \mu|g_k^{\mathrm{T}}d_{k-1}|} \tag{8}$$

其中 $\mu>1$.下面将证明新的 ZPRP 方法具有不依赖于线搜索的充分下降性,并在强 Wolfe 搜索准则下是全局收敛的.

2 算 法

算法 1(采用强 Wolfe 搜索准则的 ZPRP 算法)

Step 0 给定初始值 $x_0 \in \mathbf{R}^n$,$d_0=-g_0$,$k=0$,如果$\|g\|=0$ 则停止;

Step 1 计算满足强 Wolfe 条件的 α_k;

Step 2 计算 $x_{k+1}=x_k+\alpha_k d_k$,如果$\|g_{k+1}\|=0$ 则停止;

Step 3 用式(8)和式(3)分别计算 β_k 和 d_{k+1};

Step 4 置 $k=k+1$,转 Step1.

3 算法的收敛性

假设 1 水平集 $\Omega=\{x\in\mathbf{R}^n \mid f(x)\leqslant f(x_0)\}$ 有界,x_0 由算法 1 给出.

假设 2 f 在 Ω 的一个邻域 N 内是连续可微的,且它的梯度 g 是 Lipschitz 连续的,即存在一个常数 $L>0$,使得

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

引理 1 β_k^{ZPRP} 满足 $0\leqslant\beta_k^{\text{ZPRP}}\leqslant\beta_k^{\text{FR}}$.

证明 由 β_k^{ZPRP} 的表达式(8)可知,

$$\beta_k^{\text{ZPRP}} = \frac{\|g_k\|^2 - \dfrac{(g_k^{\mathrm{T}}g_{k-1})^2}{\|g_{k-1}\|^2}}{\|g_{k-1}\|^2 + \mu|g_k^{\mathrm{T}}d_{k-1}|} \geqslant \frac{\|g_k\|^2 - \dfrac{\|g_k\|^2\|g_{k-1}\|^2}{\|g_{k-1}\|^2}}{\|g_{k-1}\|^2 + \mu|g_k^{\mathrm{T}}d_{k-1}|}\|g_{k-1}\|^2 = 0$$

$$\beta_k^{\text{ZPRP}} \leqslant \frac{\|g_k\|^2}{\|g_{k-1}\|^2} = \beta_k^{FR}$$

所以引理 1 结论得证.

引理 2 考虑具有式(2)和式(3)形式的迭代格式,其中 $\beta_k=\beta_k^{\text{ZPRP}}$,则对任意的线搜索,有

$$g_k^{\mathrm{T}}d_k \leqslant -\left(1-\frac{1}{\mu}\right)\|g_k\|^2,\mu>1$$

证明 由 d_k 的表达式(3)可知

$$g_k^{\mathrm{T}}d_k = -\|g_k\|^2 + \frac{\|g_k\|^2 - \dfrac{(g_k^{\mathrm{T}}g_{k-1})^2}{\|g_{k-1}\|^2}}{\|g_{k-1}\|^2 + \mu|g_k^{\mathrm{T}}d_{k-1}|}g_k^{\mathrm{T}}d_{k-1} \leqslant$$

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

由此引理 2 结论得证.

由文献中[10]的定理 2 可以直接得到定理 1.

定理 1 若假设 1,2 成立,且 ZPRP 算法中的步长满足强 Wolfe 搜索条件式(4)和式(5),$0<\delta<\sigma<\frac{1}{2}$,则 $\lim\limits_{k\to\infty}\inf\|g_k\|=0$.

参考文献:

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

[2] POLAK E,RIBIÈRE G.Note Sur La Convergence De MéThodes De Directions Conjuguées[J]. Rev Fr Inform Rech Oper,1969(15):35-43

[3] POLYAK T B.The Conjugate Gradient Method in Extremal Problems * 1[J]. USSR Computational Mathematics and Mathematical Physics,1969(4):94-112

[4] HESTENES M R,STIEFEL E L. Methods of Conjugate Gradients for Solving Linear Systems[J]. Journal of Research of the National Bureau of Standards,1952,49(6):409-436

[5] LIU Y,STOREY C. Efficient Generalized Conjugate Gradient Algorithms,Part 1: Theory[J]. Journal of Optimization Theory & Applications,1991,69(1):129-137

[6] FLETCHER R. Practical Methods of Optimization[M]. A Wiley Interscience Publication,1987

[7] DAI H Y,YUAN Y. A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property[J]. Siam J Optim,1999,10(1):177-182

[8] DAI Z,WEN F. Another Improved Wei-yao-liu Nonlinear Conjugate Gradient Method with Sufficient Descent Property[J]. Applied Mathematics & Computation,2012(14):7421-7430

[9] ZHANG Y,ZHENG H,ZHANG C. Global Convergence of a Modified PRP Conjugate Gradient Method[J]. 数学研究与评论,2010,30(1):141-148

[10] GILBERT J C,NOCEDAL J. Global Convergence Properties of Conjugate Gradient Methods for Optimization[J]. Siam Journal on Optimization,1992(2):387-395

A New PRP Conjugate Gradient Method with Global Convergence

ZHOU Xue-qin

(College of Mathematics and Computer Science,Guizhou Normal University,Guizhou Guiyang 550001,China)

Abstract: In this paper, a new PRP conjugate gradient method (ZPRP) satisfying the sufficient decent condition without any line searches is proposed, and the global convergence of ZPRP with the strong Wolfe line search is proved.

Key words: conjugate gradient method; sufficient decent property; the strong Wolfe line search; global convergence