第 28 卷　第 2 期　　华侨大学学报(自然科学版)　　Vol. 28　No. 2
2007 年 4 月　　Journal of Huaqiao University (Natural Science)　　Apr. 2007

文章编号　1000-5013(2007)02-0124-03

最速下降法和共轭梯度的混合算法及全局收敛

汤仪平, 金福江

(华侨大学 信息科学与工程学院, 福建 泉州 362021)

摘要:　将最速下降法与共轭梯度法有机结合起来,构造出一种混合优化算法,并证明其全局收敛性. 这种混合优化算法结合了共轭梯度法和最速下降法产生搜索方向,既提高了共轭梯度算法的收敛速度, 又解决了目标函数的等值线是扁长椭球时,最速下降法下降缓慢的问题,具有收敛速度快、收敛范围大、适应面广等特点. 文中的算法实例表明,混合算法与单纯的共轭梯度法相比,效果更优.

关键词:　最速下降法; 共轭梯度法; 混合算法; 全局收敛性

中图分类号:　O 224　　　　**文献标识码**:　A

共轭梯度法最初由 Fletcher 和 Reeves 为求解线性方程组而提出[1] ,简称 FR 法. FR 算法具有二次终止性、编程简单、收敛速度较快等优点,但当目标函数在参数迭代点的梯度的范数值较小时,其收敛速度较慢. 最速下降法在接近极小点的地方,由于锯齿现象使收敛速度变慢. 文[2,3]曾提出了一种混合算法,本文在此基础上对参数选取进行改进,提出了一种混合优化算法,并证明了该算法的全局收敛性.

1　混合优化算法的构造

1.1　基本想法

混合优化算法结合共轭梯度法和最速下降法,产生搜索方向,避免了两者的缺点. 一般来说,最速下降方向反映了目标函数的一种局部性质. 从全局看, 由于锯齿现象的影响,使收敛率大为减慢. 即目标函数 $f(x)$ 的 Hesse 矩阵的条件数越大,收敛越慢. 共轭梯度法在 $\|\nabla f(x^{(k)})\|$ 较小时, 一维搜索会带来一定的困难,使收敛速度降低. 因此, 构造出一种好的搜索方向,使它具有朝最速下降方向靠拢的性质,并且具有共轭性,从而优化计算效果.

1.2　算法步骤

混合算法的步骤如下,给出初始点 $x^{(1)}\in\mathbf{R}^n$,容限 $\varepsilon>0$,令 $k=1$. (1) 计算 $g_k=g(x^{(k)})=\nabla f(x^{(k)})$. (2) 如果 $\|g_k\|\le\varepsilon$,则停止计算;否则,令 $p^{(k)}=-g_k+\beta_{k-1}p^{(k-1)}$. 其中,当 $k=1$,$\beta_{k-1}=0$;而当 $k<1$,$\beta_{k-1}=\|g_k\|^2/\|g_{k-1}\|^2$. (3) 计算 $f(x)$ 的 Hesse 矩阵的条件数 γ 和 $\|\nabla f(x^{(k)})\|$. 若 $\gamma\to1$,取 $\alpha=1$;若 $\gamma\gg1$ 且 $\|\nabla f(x^{(k)})\|$ 较大,取 $\alpha\in[0,1/2]$;否则,取 $\alpha\in[1/2,1]$. (4) $q^{(k)}=-\alpha g_k+(1-\alpha)p^{(k)}$. (5) 一维搜索,求 λ_k,使得 $f(x^{(k)}+\lambda_k q^{(k)})=\min\limits_{k\to\infty}f(x^{(k)}+\lambda q^{(k)})$. (6) 置 $x^{(k+1)}=x^{(k)}+\lambda_k q^{(k)}$,$k:=k+1$,返回步骤(2). 参数 α 的选取对混合算法的收敛速度具有一定的影响. 当 $\alpha=0$ 时,混合算法就是共轭梯度法;当 $\alpha=1$ 时,混合算法就是最速下降法. 我们可以根据具体情况选取参数 α 值,以达到最佳的计算效果.

2　混合算法的全局收敛性

引理 1　算法的搜索方向 $q^{(1)},q^{(2)},\cdots,q^{(k)},\cdots$ 是共轭方向.

收稿日期:　2006-06-26

作者简介:　汤仪平(1983-),男,硕士研究生,主要从事最优化控制的研究; 通信作者: 金福江(1965-),男,教授,E-mail:jinfujiang@163.com.

基金项目:　福建省自然科学基金资助项目(A0540002)

证明　已知FR算法得到的 $p^{(1)}, p^{(2)}, \cdots, p^{(k)}, \cdots$ 是共轭方向,满足 $p^{(i)\mathrm{T}}Ap^{(j)}=0, j\neq i$. 因而,可得

$$q^{(i)\mathrm{T}}Aq^{(j)} = (-\alpha g_i + (1-\alpha)p^{(i)})^{\mathrm{T}}A(-\alpha g_i + (1-\alpha)p^{(j)}) = \alpha^2 g_i^{\mathrm{T}}Ag_j + (1-\alpha)^2 p^{(i)\mathrm{T}}Aq^{(j)}, \quad j\neq i.$$

因为, $g_i^{\mathrm{T}}Ag_j=0$, $p^{(i)\mathrm{T}}Ap^{(j)}=0$, $j\neq i$. 所以, $q^{(i)\mathrm{T}}Aq^{(j)}=0, j\neq i$. 即 $q^{(1)}, q^{(2)}, \cdots, q^{(k)}, \cdots$ 是共轭方向.

引理2　算法的搜索方向 $q^{(1)}, q^{(2)}, \cdots, q^{(k)}, \cdots$ 是目标函数的下降方向.

证明　由上述算法定义,有

$$q^{(k)} = -\alpha g_k + (1-\alpha)p^{(k)}. \tag{1}$$

把 $p^{(k)} = -g_k + \beta_{k-1}p^{(k-1)}$ 代入式(1),可得

$$q^{(k)} = -\alpha g_k + (1-\alpha)(-g_k + \beta_{k-1}p^{(k-1)}) = -g_k + (1-\alpha)\beta_{k-1}p^{(k-1)}. \tag{2}$$

式(2)等号两端左乘 g_k^{T},且由一维搜索定义[4]可知, $g_k^{\mathrm{T}}p^{(k-1)}=0$. 因此,可得到

$$g_k^{\mathrm{T}}q^{(k)} = -\|g_k\|^2 < 0, \qquad \alpha>0, \quad g_k\neq 0. \tag{3}$$

由于 $f(x^{(k+1)}) = f(x^{(k)}+\lambda_k q^{(k)}) = \min\limits_{\lambda\geq 0} f(x^{(k)}+\lambda q^{(k)})$, 因此,必有 $f(x^{(k+1)}) < f(x^{(k)})$, 故 $q^{(k)}$ 是在 $x^{(k)}$ 处的下降方向.

引理3　设函数 $f(x)$ 二阶连续可微, $\nabla f(x)$ 满足Lipschitz条件[5]. $\|\nabla f(x^{(k)}) - \nabla f(x^{(k+1)})\| \leq L\|x^{(k)} - x^{(k+1)}\|$, L 为Lipschitz系数,且搜索方向 $d^{(k)}$ 与负梯度 $(-g_k)$ 的夹角 θ_k 满足 $\theta_k\leq \pi/2-\mu, \mu>0$. 即 $d^{(k)}$ 偏离 g_k 的正交方向远一些. 又 $\sum\limits_{k=1}^{\infty}\cos^2\theta_k$ 收敛,那么,以任意的 $x^{(1)}\in \mathbf{R}^n$ 为初始点,依次沿 $d^{(1)}$, $d^{(2)}, \cdots, d^{(k)}$ 共轭的非零向量进行一维搜索,得到的点系列 $\{x^{(k)}\}$ 终止(或收敛)于函数 $f(x)$ 在 $\mathbf{R}^n$ 上的唯一极小点, 则共轭梯度算法全局收敛. 即 $\lim\limits_{k\to\infty}\|\nabla f(x^{(k)})\| = \|\nabla f(\bar{x})\| = 0$.

证明　由条件可知 $\{f(x^{(k)})\}$ 单调下降,因此极限存在. 所以, $\lim\limits_{k\to\infty}[f(x^{(k)}) - f(x^{(k+1)})]=0$. 令 $g_k = \nabla f(x^{(k)})$,假设对所有 k, $\lim\limits_{k\to\infty}\|\nabla f(x^{(k)})\| = \|\nabla f(\bar{x})\| = 0$ 不成立,则存在常数 $\varepsilon>0$,使得 $\|g_k\|\geq \varepsilon$. 从而有

$$\frac{-g_k^{\mathrm{T}}d^{(k)}}{\|d^{(k)}\|} = \|g_k\|\cdot\cos\theta_k \geq \varepsilon\cos\theta_k, \tag{4}$$

令 $\varepsilon_1 = \varepsilon\cos\theta_k$,又由于有

$$f(x^{(k)}+\lambda_k d^{(k)}) = f(x^{(k)}) + \lambda_k g(\zeta_k)^{\mathrm{T}}d^{(k)} = f(x^{(k)}) + \lambda_k g_k^{\mathrm{T}}d^{(k)} + \lambda_k[g(\zeta_k) - g_k]^{\mathrm{T}}d^{(k)} \leq f(x^{(k)}) + \lambda_k\|d^{(k)}\|\left[\frac{g_k^{\mathrm{T}}d^{(k)}}{\|d^{(k)}\|} + \|g(\zeta_k) - g_k\|\right]. \tag{5}$$

其中, ζ_k 在 $x^{(k)}$ 和 $x^{(k)}+\lambda_k d^{(k)}$ 之间. 由于 g_k 满足Lipschitz条件,故存在 $\bar{\lambda}$,使得当 $0\leq\lambda_k\|d_k\|\leq\bar{\lambda}$ 时,

$$\|g(\zeta_k) - g_k\| \leq \frac{1}{2}\varepsilon_1. \tag{6}$$

所以,由式(4)～(6)可得

$$f\left(x^{(k)} + \bar{\lambda}\frac{d^{(k)}}{\|d^{(k)}\|}\right) \leq f(x^{(k)}) + \bar{\lambda}\left(\frac{g_k^{\mathrm{T}}d^{(k)}}{\|d^{(k)}\|} + \frac{1}{2}\varepsilon_1\right) \leq f(x^{(k)}) - \frac{1}{2}\bar{\lambda}\varepsilon_1,$$

从而有

$$f(x^{(k+1)}) \leq f\left(x^{(k)} + \bar{\lambda}\frac{d^{(k)}}{\|d^{(k)}\|}\right) \leq f(x^{(k)}) - \frac{1}{2}\bar{\lambda}\varepsilon_1.$$

这与 $\lim\limits_{k\to\infty}\|\nabla f(x^{(k)})\| = \|\nabla f(\bar{x})\| = 0$ 矛盾,从而 $g_k\to 0$,即引理3得证.

定理1　假定引理1,2,3满足,则混合算法产生的迭代点 $\{x^{(k)}\}$ 终止(或收敛)于函数 $f(x)$ 在 $\mathbf{R}^n$ 上的唯一极小点, 算法全局收敛. 即 $\lim\limits_{k\to\infty}\|\nabla f(x^{(k)})\| = \|\nabla f(\bar{x})\| = 0$.

证明　由引理1,2可知, $q^{(k)}$ 是在 $x^{(k)}$ 处的下降方向且共轭,即混合算法是严格下降算法,这里只需证明搜索方向 $d^{(k)}$ 与负梯度 $(-g_k)$ 的夹角 θ_k 满足 $\theta_k\leq\pi/2-\mu, \mu>0$. 即 $\theta_k\neq\pi/2$,则算法全局收敛. 由式(3)可知, $(-g_k, q^{(k)}) = \|g_k\|^2$, $g_k\neq 0$,又因为 $(-g_k, q^{(k)}) = \|-g_k\|\cdot\|q^{(k)}\|\cos\theta_k$. 所以, $\cos\theta_k = \dfrac{\|g_k\|^2}{\|-g_k\|\cdot\|q^{(k)}\|}\neq 0$, $0<\alpha<1$, $g_k\neq 0$. 即 $\theta_k\neq\pi/2$. 因此,由引理3可知,定理1成立.

3 算法实例

为了验证本文提出的混合优化算法的有效性,用它求解常用函数的极值.现有目标函数 $f(x)$ 如下

$$f(x) = \frac{1}{100}(1 - x_1)^2 + \frac{1}{80}(1 + x_2)^2,$$

取初值 $x^{(1)} = (2,3)^T$. 在点 x 处,目标函数的梯度是

$$g(x) = \nabla f(x) = [-\frac{1}{50}(1 - x_1), \frac{1}{40}(1 + x_2)]^T,$$

而 Hesse 矩阵 $A = \begin{bmatrix} \frac{1}{50} & 0 \\ 0 & \frac{1}{40} \end{bmatrix}$,其条件数 $\gamma = 5/4$,则数值结果如表 1 所示.由表 1 可知,混合算法是可行的,其计算量和 FR 算法大致相同,但前者的计算效果优于后者.

表 1 两种算法迭代收敛性比较

Tab.1 The comparison between the two algorithms convergence

算法名称	迭代次数	最优解
共轭梯度法	2	$(0.9862497, -0.9789286)^T$
本文的混合算法	2	$(1.0001314, -0.9999247)^T$

4 结束语

本文提出的算法是共轭梯度法的推广,它结合共轭梯度法和最速下降法产生搜索方向,避免两者的缺点,证明了新算法在一定条件下具有全局收敛性.数值实验说明,算法是有效的.为进一步研究,可将算法推广到非精确搜索下,从而探讨非精确搜索下的算法.

参考文献:

[1] 陈宝林.最优化理论与算法[M].北京:清华大学出版社,2005:334-368.
[2] 欧志英,严克明,王柏岩.共轭梯度法和最速下降法的混合算法[J].甘肃工业大学学报,1999,25(1):89-91.
[3] 宁 伟,卿熙宏,陶华学.基于共轭梯度法和最速下降法的非线性测量数据处理[J].山东科技大学学报:自然科学版,2004,23(4):5-7.
[4] 袁亚湘,孙文瑜.最优化理论与方法[M].北京:科学出版社,1997:108-199.
[5] 薛嘉庆.最优化原理与方法[M].北京:冶金工业出版社,1983:66-100.

A Hybrid Algorithm of the Steepest Descent Method and the Conjugate Gradient Method and Its Global Convergence

TANG Yi-ping, JIN Fu-jiang

(College of Information Science and Engineering, Huaqiao University, Quanzhou 362021, China)

Abstract: Based on the steepest descent method and the conjugate gradient method, a hybrid algorithm is proposed in this paper, and its global convergence is proved. The hybrid algorithm raises the convergence rate of the conjugate gradient method and solves the problem for which the convergence rate of the steepest descent method get slower when the isopleth of goal function is oblong. In conclusion, the method has features with quick convergence rate, large convergence range and wide accommodation compared with the conjugate gradient method, the hybrid algorithm method has a better result in the example.

Keywords: the steepest descent method; conjugate gradient method; hybrid algorithm; global convergence

(责任编辑: 黄仲一)