doi:10.16055/j.issn.1672-058X.2015.0008.002

带有最优参数选择的修正 DL 共轭梯度法

吴 双 江
(重庆师范大学 数学科学学院,重庆 401331)

摘　要:利用矩阵条件数的求解方法,求解基于 MBFGS 割线条件的修正 DL 共轭梯度法中的参数 t,提出带有优选参数的修正 DL 共轭梯度法;假设搜索方向有下降性,并通过强 Wolfe 线搜索求解步长,证明了新的共轭梯度法对一般函数有全局收敛性;最后比较了新的共轭梯度法的数值有效性。

关键词:共轭梯度法;强 Wolfe 线搜索;条件数;全局收敛性

中图分类号:O224　　**文献标识码**:A　　**文章编号**:1672-058X(2015)08-0006-04

0 引　言

考虑无约束优化问题

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

其中 $f:\boldsymbol{R}^n\rightarrow\boldsymbol{R}$ 是连续可微的。非线性共轭梯度法通常有如下格式:

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

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

其中 $g_k=\nabla f(x_k)$,$\alpha_k>0$ 是由某线搜索获得的步长,β_k 是一个参数。

为了建立共轭梯度法的全局收敛性,选取强 Wolfe 线搜索:

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

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

在文献[1]中,Dai 和 Liao 利用修正共轭条件的方法提出新共轭梯度法,其参数 β_k 的形式:

$$\beta^{\mathrm{DL}}=\frac{g_k^{\mathrm{T}}y_{k-1}}{d_{k-1}^{\mathrm{T}}y_{k-1}}-t\frac{g_k^{\mathrm{T}}s_{k-1}}{d_{k-1}^{\mathrm{T}}y_{k-1}} \tag{6}$$

其中 $s_{k-1}=x_k-x_{k-1}$。

在文献[2]中,Saman Babaie-Kafaki 和 Reza Ghanbari 利用条件数,求解了 DL 法中参数 t,获得两种新的共轭梯度法 M1 和 M2,其选取的参数 t 分别为

$$t_k^{\mathrm{M1}}=\frac{s_k^{\mathrm{T}}y_k}{\|s_k\|^2}+\frac{\|y_k\|}{\|s_k\|}\text{和 }t_k^{\mathrm{M2}}=\frac{\|y_k\|}{\|s_k\|}$$

在文献[3]中,Zhou WeiJun 和 Zhang Li 利用割线条件修正 DL 法,获得新的 ZZ 法,其 β_k 的形式:

$$\beta^{\mathrm{ZZ}}=\frac{g_k^{\mathrm{T}}z_{k-1}}{d_{k-1}^{\mathrm{T}}z_{k-1}}-t\frac{g_k^{\mathrm{T}}s_{k-1}}{d_{k-1}^{\mathrm{T}}z_{k-1}} \tag{7}$$

收稿日期:2015-01-10;**修回日期**:2015-02-20.

作者简介:吴双江(1990-),男,重庆梁平人,硕士研究生,从事最优化理论与算法研究.

其中 $z_{k-1}=y_{k-1}+C\|g_{k-1}\|^r s_{k-1}$。$C>0,r>0$。

受以上启发,利用条件数求解 ZZ 法中参数 t,分别获得如下两种 t 的取值:

$$t_k^{\mathrm{MZ1}}=\frac{s_k^{\mathrm{T}}z_k}{\|s_k\|^2}+\frac{\|z_k\|}{\|s_k\|} \tag{8}$$

$$t_k^{\mathrm{MZ2}}=\frac{\|z_k\|}{\|s_k\|} \tag{9}$$

其中 $z_{k-1}=y_{k-1}+C\|g_{k-1}\|^r s_{k-1}$。$C>0,r>0$。把式(7)-(8)称做 MZ1 法,把式(7)-(9)称作 MZ2 法。

1 MZ1 法与 MZ2 法的全局收敛性

假设 1:(1) 水平集 $\Omega=\{x\in\mathbf{R}^n\,|f(x)\leqslant f(x_0)\}$ 有界,其中 $x_0\in\mathbf{R}^n$ 为初始点。

(2) f 在水平集 Ω 的一个领域 N 内连续可微,且其梯度 g 满足 Lipschitz 连续,即存在常数 $L>0$,使得 $\|g(x)-g(y)\|\leqslant L\|x-y\|,\forall x,y\in N$。

由假设 1 中(2)可知 $\|g_k\|\leqslant\gamma$。

假设 MZ1 法与 MZ2 法的搜索方向满足下降条件:$g_k^{\mathrm{T}}d_k<0$

引理 1[4] 若假设 A 成立。考虑迭代格式为(2)-(3)的共轭梯度法,其中 d_k 满足下降条件,α_k 满足强 Wolfe 线搜索。如果

$$\sum_{k=0}^{\infty}\frac{1}{\|d_k\|^2}<+\infty \tag{10}$$

则

$$\liminf_{k\to\infty}\|g_k\|=0 \tag{11}$$

定理 1 若假设 1 成立。分别考虑共轭梯度法 MZ1 法与 MZ2 法。两种方法中满足下降条件,满足强 Wolfe 线搜索。则 MZ1 法与 MZ2 法均对一般函数有全局收敛性。

证明:根据引理 1,只需要证明 MZ1 法与 MZ2 法中 d_k 有界,那么则 MZ1 法与 MZ2 法均对一般函数有全局收敛性。因为 MZ1 法和 MZ2 法中满足下降条件,因此 $d_k\neq0$。下面运用反正法证明 MZ1 法与 MZ2 法中有界。

假设 MZ1 法与 MZ2 法均对一般函数不具有全局收敛性,则存在常数 $\varepsilon>0$,使得 $\|g_k\|\geqslant\varepsilon$ 对任意 k 成立。根据式(5)、(7),有

$$|s_{k-1}^{\mathrm{T}}z_{k-1}|>C\varepsilon^r\|s_{k-1}\|^2 \tag{12}$$

又根据式(7)和假设 1,有

$$\|z_{k-1}\|\leqslant(L+C\gamma^r)\|s_{k-1}\| \tag{13}$$

在 MZ1 法中,根据(2)、(3)、(7)、(12)、(13)和假设 1:

$$\begin{aligned}\|d_k\|\leqslant\|-g_k+\beta_k^{MZ1}d_{k-1}\|&=\left\|-g_k+\left(\frac{g_k^{\mathrm{T}}z_{k-1}}{s_{k-1}^{\mathrm{T}}z_{k-1}}-t_k^{MZ1}\frac{g_{k-1}^{\mathrm{T}}z_{k-1}}{s_{k-1}^{\mathrm{T}}z_{k-1}}\right)s_{k-1}\right\|\leqslant\\ &\gamma+\left(\frac{\gamma(L+C\gamma^r)\|s_{k-1}\|}{C\varepsilon^r\|s_{k-1}\|^2}+2\gamma(L+C\gamma^r)\frac{\gamma(L+C\gamma^r)\|s_{k-1}\|}{C\varepsilon^r\|s_{k-1}\|^2}\right)\|s_{k-1}\|\leqslant\\ &\gamma+(1+2\gamma(L+C\gamma^r))\frac{\gamma(L+C\gamma^r)}{C\varepsilon^r}\end{aligned}$$

因此根据引理 1,MZ1 法对一般函数有全局收敛性。MZ2 法,同理可证得对一般函数的全局收敛性。因此省略对 MZ2 法的全局收敛性证明过程。定理证明完毕。

2 数值试验

现比较 MZ1 法,MZ2 法,ZZ 法与 DL 法的数值效果,测试问题取自于文献[5]。测试问题的维数为 2~5 000 维。在所有的共轭梯度法计算中,步长 α_k 通过强 Wolfe 线搜索获得,其中强 Wolfe 线搜索的参数 $\delta=0.01$,$\sigma=0.1$。DL 法中参数 $t=0.1$。MZ1 法,MZ2 法,和 ZZ 法中参数 $C=0.001$,并且如果 $\|g_k\|\geqslant 1$ 时 $r=1$,否则 $r=3$。这些方法在配置为 1.86 GHz CPU ,2.5 GB RAM,Windows 7 操作系统的联想 Z460 笔记本电脑上用 MATLAB 7.0.1 软件测试数值有效性。算法中的终止条件有两个:如果 $f(x_{k-1})>10^{-6}$,第一个终止条件为 $\left|\dfrac{f(x_{k-1})-f(x_k)}{f(x_{k-1})}\right|\leqslant 10^{-6}$,否则第一个终止条件为 $|f(x_{k-1})-f(x_k)|\leqslant 10^{-6}$;如果迭代次数大于 1 000 次。绘制图 1、2、3、4 来显示 MZ1 法、MZ2 法、ZZ 法、DL 法的数值结果。同时绘制表一显示 MZ1 法、MZ2 法、ZZ 法、DL 法的相对有效性。

通过图 1、2、3、4 和表 1 可知,MZ1 法在函数计算次数,梯度计算次数,迭代次数,时间上均好于其他几种方法。

表 1 MZ1 法、MZ2 法、ZZ 法、DL 法的相对有效性

DL	ZZ	MZ1	MZ2
1.000 0	0.922 3	0.847 7	0.866 0

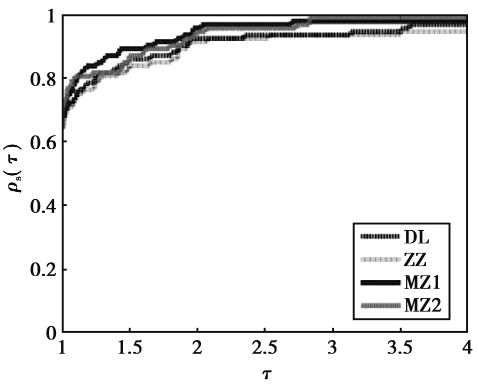

图 1 函数计算次数

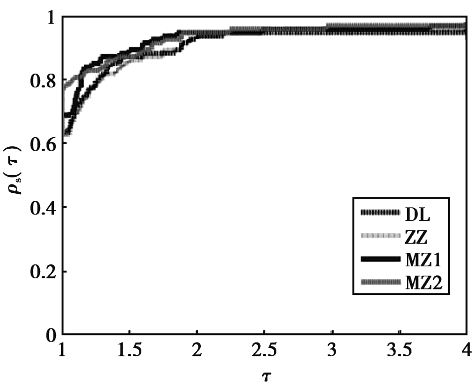

图 2 梯度计算次数

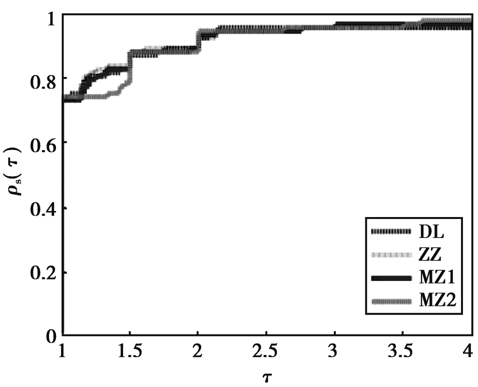

图 3 迭代次数

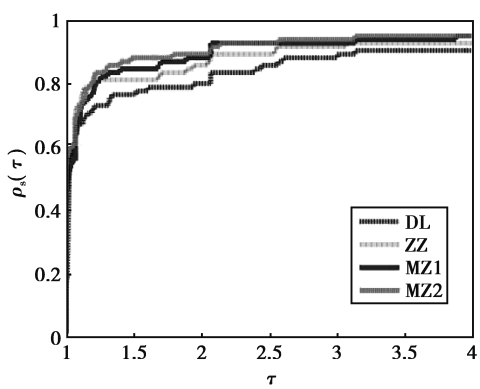

图 4 CPU 时间

(下转第 **18** 页)