文章编号:1671-6833(2009)03-0139-02

Richardson 迭代法的一个常数步长

董云达[1], 尤燕飞[2]

(1. 郑州大学 数学系,河南 郑州 450001; 2. 南京大学 数学系,江苏 南京 210093)

摘 要: 对于求解对称正定线性方程组的 Richardson 迭代法,给出一个新的常数步长. 它仅依赖于系数矩阵的对角线上的最小元素和最大特征值. 而且,还证明了在该步长下 Richardson 迭代法产生的梯度模序列线性地趋于 0. 初步的数值试验表明了新步长的某些优势.

关键词: 正定线性方程组; Richardson 迭代法; 步长; 收敛

中图分类号: O 221.2 **文献标识码**: A

0 引言

考察下列无约束最优化问题:

$$\min f(x)=\frac{1}{2}x^TAx-b^Tx \tag{1}$$

式中:$x\in R^n$ 为自变量;系数矩阵 A 为 n 阶对称正定矩阵;$b\in R^n$.

对于问题(1),可用如下的 Richardson 迭代法[1]进行求解:

$$x^{k+1}=x^k-\alpha_k\ \nabla f(x^k) \tag{2}$$

式中:步长 $\alpha_k>0$; $\nabla f(x^k)=Ax^k-b$. 经典的常数步长为

$$\alpha^{OPT}=\frac{2}{\lambda_1+\lambda_n} \tag{3}$$

式中:λ_1,λ_n 分别为系数矩阵 A 最小、最大特征值.

一般来说,用反幂法等算法来估计 λ_1 并非易事,当 λ_1,λ_2 相等或差别不大时尤为如此. 基于这样的考虑,笔者研究 Richardson 迭代法的一个新的常数步长

$$\alpha^{NEW}=\frac{2}{a+\lambda_n} \tag{4}$$

式中:a 表示系数矩阵 A 主对角线上的最小元素. 我们不仅证明了由 Richardson 迭代法(2)、(4)所产生的梯度模序列线性收剑于 0,而且还构造了一类例子进行了数值试验以表明新步长的某些优势.

1 收剑性分析

定理 1 设$\{x^k\}$是由 Richardson 迭代法(2)、(4)产生的序列,则

$$\|\nabla f(x^{k+1})\|\leqslant\frac{a+\lambda_n-2\lambda_1}{a+\lambda_n}\|\nabla f(x^k)\|$$

证明:因为 A 为对称正定阵,所以存在正交阵 U 使得

$$A=U^TDU$$

式中:$D=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$, $0<\lambda_1\leqslant\cdots\leqslant\lambda_n$ 为 A 的 n 个特征值.

由于

$$\nabla f(x^{k+1})=Ax^{k+1}-b$$

从而由(2)和(4)可知:

$$\begin{aligned}\nabla f(x^{k+1})&=Ax^{k+1}-b\\&=A(x^k-\alpha^{NEW}\ \nabla f(x^k))-b\\&=\nabla f(x^k)-\alpha^{NEW}A\ \nabla f(x^k)\end{aligned}$$

即

$$\nabla f(x^{k+1})=(I-\alpha^{NEW}A)\ \nabla f(x^k)$$

式中:I 为 n 阶单位阵. 由此可得

$$\begin{aligned}&\|\nabla f(x^{k+1})\|^2\\=&\nabla f(x^k)^T(I-\alpha^{NEW}A)^2\ \nabla f(x^k)\\=&(U\ \nabla f(x^k))^T(I-\alpha^{NEW}D)^2(U\ \nabla f(x^k))\\\leqslant&\|(I-\alpha^{NEW}D)^2\|\ \|(U\ \nabla f(x^k))\|^2\\=&\|(I-\alpha^{NEW}D)^2\|\ \|(\ \nabla f(x^k))\|^2\\\leqslant&\max\{(1-\alpha^{NEW}\lambda_1)^2,(1-\alpha^{NEW}\lambda_n)^2\}\cdot\\&\|\nabla f(x^k)\|^2\end{aligned}$$

收稿日期:2009-01-14;**修订日期**:2009-03-13

基金项目:国家自然科学基金资助项目(10702065)

作者简介:董云达(1969-),男,河南新乡人,郑州大学副教授,博士,主要从事最优化研究,E-mail:ydong@zzu.edu.cn.

由于 $\lambda_1 \leqslant a \leqslant \lambda_n$,所以上面关系式中的最大值为

$$\left(\frac{a+\lambda_n-2\lambda_1}{a+\lambda_n}\right)^2$$

从而

$$\|\nabla f(x^{k+1})\| \leqslant \frac{a+\lambda_n-2\lambda_1}{a+\lambda_n}\|\nabla f(x^k)\|$$

由定理1可以看出,Richardson迭代法(2)、(4)所产生的梯度模序列线性地收敛于0.

2 数值试验

为了凸显新步长的优势,我们选取具有以下特点的系数矩阵 $\boldsymbol{A}$:(1)λ_n,λ_{n-1} 相差较大;(2)λ_1,λ_2 相差很小;(3)$\lambda_n+\lambda_1$ 与 λ_n+a 之比接近于1.

这样的例子是不难构造的;笔者所选取的是一个五对角矩阵

$$\boldsymbol{A}=(a_{ij}):\ a_{11}=100;\ a_{ii}=4,\ i\neq 1;$$
$$a_{ij}=1,\ |i-j|=1,2;\ a_{ij}=0,\ |i-j|\geqslant 3$$

对于任意 n,初始点 x^0 恒取为零向量,且令 $b=(1,1,1\cdots1,1,1)^T$. 停止准则为:

$$\|\nabla f(x^k)\| \leqslant 10^{-6}\|\nabla f(x^0)\|$$

或者超过最大迭代次数10 000次. 所用电脑操作系统是微软Vista Home Basic,处理器是Inter奔腾双核T2330,频率为1.60 GHZ,编程软件为Matlab7.1.需特别声明的是NEW法的CPU时间是迭代所花时间与乘幂法求最大特征值所花时间的和;OPT法的CPU时间是迭代所花时间与乘幂法(反幂法)求最大(最小)特征值所花时间的和. NEW法与OPT法的数值计算结果如表1所示.

表1 NEW法与OPT法的数值计算结果

Tab.1 Numerical results of NEW and OPT

	100维		500维		1 000维	
	迭代步数	CPU时间/s	迭代步数	CPU时间/s	迭代步数	CPU时间/s
OPT	316	0.005 4	294	0.202 7	285	1.473 0
NEW	240	0.001 7	218	0.006 0	209	0.011 6

3 结论

由表1可知,同样是常数步长,对于不同的维数 n,NEW法比OPT法的迭代次数少一些;NEW法的CPU时间却比OPT法少得多. 并且,维数越大,NEW法的优势就越为明显.

参考文献:

[1] VARGA R S. Matrix Iterative Analysis[M]. Second Edition. 北京:科学出版社,2006.

Constant Step Length in Richardson Iterative Method

DONG Yun-da[1], YOU Yan-fei[2]

(1. Department of Mathematics, Zhengzhou University, Zhengzhou 450001, China; 2. Department of Mathematics, Nanjing University, Nanjing 210093, China)

Abstract: Richardson iterative method for solving system of linear equations with the positive definite coefficient matrix is considered, and a new constant step length rule only depending on the minimal diagonal element and the maximum eigenvalue is proposed. Furthermore, the linear convergence of the generated gradient norms to zero is proved. Preliminary numerical tests show that the new rule is competitive for certain problems.

Key words: system of linear equations; Richardson iterative method; step length rule; convergence