文章编号:1673-5005(2013)02-0191-06 doi:10.3969/j.issn.1673-5005.2013.02.032

基于信赖域技术和修正拟牛顿方程的非单调超记忆梯度算法

宫恩龙[1], 陈双双[2], 孙清滢[2], 陈颖梅[2]

(1. 青岛酒店管理职业技术学院,山东 青岛 266100; 2. 中国石油大学 理学院,山东 青岛 266580)

摘要:基于信赖域技术和修正拟牛顿方程,结合 Neng-Zhu Gu 非单调策略,设计新的求解无约束最优化问题的非单调超记忆梯度算法,分析算法的收敛性和收敛速度。新算法每次迭代节约了矩阵的存储量和计算量,算法稳定, 适于求解大规模问题。数值试验结果表明新算法是有效的。

关键词: 超记忆梯度算法; 非单调规则; 收敛性; 收敛速度; 数值试验

中图分类号:O 221.2 **文献标志码**:A

A non-monotone super-memory gradient method based on trust region technique and modified quasi-Newton equation

GONG En-long[1], CHEN Shuang-shuang[2], SUN Qing-ying[2], CHEN Ying-mei[2]

(1. *Qingdao Hotel Management College*, *Qingdao* 266100, *China*;
2. *College of Science in China University of Petroleum*, *Qingdao* 266580, *China*)

Abstract: Based on trust region technique and modified quasi-Newton equation, by combining with Neng-Zhu Gu non-monotone strategy, a new super-memory gradient method for unconstrained optimization problem was presented. The global and convergence properties of the new method were proved. It saves the storage and computation of some matrixes in its iteration, and is suitable for solving large scale optimization problems. The numerical results show that the new method is effective.

Key words: super-memory gradient method; non-monotone step rule; convergence; convergence rate; numerical experiment

1 问题的提出

考虑无约束最优化问题

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

其中 $f(\boldsymbol{x}):\boldsymbol{R}^n\to\boldsymbol{R}^1$ 是一阶连续可微函数。

求解问题(1)的共轭梯度算法,结构简单,收敛速度快,存储量小,适于求解大规模问题[1-2]。记忆梯度法[3]是共轭梯度法的一种变形和改进,它具有比共轭梯度法更快的收敛速度。文献[4-6]利用记忆梯度法的基本思想,增加记忆项的项数,引入了超记忆梯度算法。由于这类方法在迭代中较多地利用了已经得到的目标函数的某些信息,因而具有比梯度法更快的收敛速度[7]。但是记忆梯度法在每一次迭代中需要作一次二维精确搜索,超记忆梯度法在每一次迭代中需要作一次多维精确搜索,这一点很不理想,为此很多文献[8-13]对算法进行了改进。最近,Shi 等[14]结合信赖域技术提出了一类具有全局收敛性和数值效果比 FR、PRP、HS、CD、DY 共轭梯度算法好的新超记忆梯度单调下降算法。

非单调技术由于其有利于求解全局最优解和算法的快速收敛而受到关注[15-19]。Gu 等[16]改进传统非单调技术中的参考值的选取,提出了一种新的非单调线搜索技术:

$$f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)\leqslant D_k+\delta\alpha\nabla f(\boldsymbol{x}_k)^{\mathrm T}\boldsymbol{d}_k, \tag{2}$$

其中

$$D_k=\begin{cases} f(\boldsymbol{x}_k),\ k=0,\\ \eta D_{k-1}+(1-\eta)f(\boldsymbol{x}_k),\ k\geqslant 1.\end{cases} \tag{3}$$

这里 $\delta\in(0,1)$, $\eta\in(0,1)$ 是两个参数。

收稿日期:2012-09-05

基金项目:国家自然科学基金项目(61201455);中央高校基本科研业务费专项(10CX04044A;11 CX06087A)

作者简介:宫恩龙(1969-),男,副教授,硕士,从事数学规划研究。E-mail:GongEnlong@163.com。

本文将结合 Gu 的非单调技术,对时贞军提出的超记忆梯度算法改进,设计非单调超记忆梯度算法,同时为了进一步减少算法的存贮量和计算量,提出基于修正拟牛顿方程的 $\boldsymbol{B}_k$ 的一种稀疏对角修正形式,即基于 Wei 等[20] 提出的新的修正拟牛顿方程

$$\boldsymbol{B}_{k+1}\boldsymbol{s}_k=\bar{\boldsymbol{y}}_k,\text{或者 }\boldsymbol{H}_{k+1}\bar{\boldsymbol{y}}_k=\boldsymbol{s}_k. \tag{4}$$

其中 $\bar{\boldsymbol{y}}_k=\boldsymbol{y}_k+\boldsymbol{A}_k\boldsymbol{s}_k$,$\boldsymbol{A}_k$ 是一简单矩阵含有目标函数的二阶信息,满足:$\boldsymbol{A}_k\boldsymbol{s}_k=\boldsymbol{w}_k$,$\boldsymbol{w}_k=\dfrac{\boldsymbol{v}_k}{\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{u}_k}\boldsymbol{u}_k$,$\boldsymbol{v}_k=2(f_k-f_{k+1})+(\boldsymbol{g}_{k+1}+\boldsymbol{g}_k)^{\mathrm{T}}\boldsymbol{s}_k$,$\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{u}_k\neq 0$。给出 $\boldsymbol{A}_k$ 的两种取法:

$$(\mathrm{i})\boldsymbol{A}_k=\frac{\boldsymbol{v}_k}{\|\boldsymbol{s}_k\|^2}\boldsymbol{I}. \tag{5}$$

$$(\mathrm{ii})\boldsymbol{A}_k=\frac{\boldsymbol{v}_k}{(\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{y}_k)^2}(\boldsymbol{y}_k\boldsymbol{y}_k^{\mathrm{T}}). \tag{6}$$

特别地,当 $\boldsymbol{A}_k=0$ 时,则对应于拟牛顿方程。

基于式(4),给出 $\boldsymbol{B}_k$ 的修正形式如下:设 $\boldsymbol{B}_{k-1}$ 为对角稀疏正定矩阵,令 $\boldsymbol{B}_k=\boldsymbol{B}_{k-1}+\Delta\boldsymbol{B}_{k-1}$,其中 $\Delta\boldsymbol{B}_{k-1}$ 为对角矩阵,为保证 $\boldsymbol{B}_k$ 的正定性,限制 $\boldsymbol{B}_k=\mathrm{diag}(b_k^1,b_k^2,\cdots,b_k^n)$ 取值,即满足

$$0<\underline{b}\leqslant b_k^i\leqslant\bar{b},i=1,2,\cdots,n. \tag{7}$$

$\boldsymbol{B}_k$ 近似满足修正拟牛顿方程:

$$\min_{\underline{b}\leqslant b_k^i\leqslant\bar{b},i=1,2,\cdots,n}\sum_{i=1}^{n}\left((b_{k-1}^i+\Delta b_{k-1}^i)s_{k-1}^i-\bar{y}_{k-1}^i\right)^2. \tag{8}$$

由式(8)知,对 $i=1,2,\cdots,n$ 有:

当 $s_{k-1}^i\neq 0$ 时,如果 $\underline{b}\leqslant\bar{y}_{k-1}^i/s_{k-1}^i\leqslant\bar{b}$,则取 $\Delta b_{k-1}^i=\bar{y}_{k-1}^i/s_{k-1}^i-b_{k-1}^i$;如果 $\bar{y}_{k-1}^i/s_{k-1}^i<\underline{b}$,则取 $\Delta b_{k-1}^i=\underline{b}-b_{k-1}^i$;如果 $\bar{y}_{k-1}^i/s_{k-1}^i>\bar{b}$,则取 $\Delta b_{k-1}^i=\bar{b}-b_{k-1}^i$;当 $s_{k-1}^i=0$ 时,则取 $\Delta b_{k-1}^i=0$。

2 算 法

基于信赖域技术和修正拟牛顿方程的非单调超记忆梯度算法(SM):

步骤 1 取 $\boldsymbol{x}_0\in R^n$,$\rho\in(0,1)$,$\mu\in(0,1)$,$m\geqslant 2$,$\boldsymbol{B}_0=\boldsymbol{I}_n$,$0<\underline{b}<\bar{b}$ 为常数,令 $k:=0$。

步骤 2 如果 $\|\boldsymbol{g}_k\|=0$,则停止迭代;否则,转步骤 2。

步骤 3 定义 m_k 满足 $0\leqslant m_k\leqslant\min\{k,m\}$。

步骤 4 $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\boldsymbol{P}_k(\alpha_k)$,其中 α_k 是 $\{1,\rho,\rho^2,\cdots\}$ 中满足下式的最大者:

$$\frac{D_k-f(\boldsymbol{x}_k+\boldsymbol{P}_k(\alpha))}{q_k(0)-q_k(\boldsymbol{y}_k(\alpha))}\geqslant\mu, \tag{9}$$

其中 $\boldsymbol{D}_k$ 由式(3)确定,$\boldsymbol{P}_k(\alpha)=\boldsymbol{V}_k\boldsymbol{y}_k(\alpha)$,且 $\boldsymbol{y}_k(\alpha)$ 是下列子问题的解:

$$\begin{cases}\min q_k(\boldsymbol{y})=f(\boldsymbol{x}_k)+\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{V}_k\boldsymbol{y}+\dfrac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{V}_k\boldsymbol{y},\\ \mathrm{s.t.}\ \|\boldsymbol{V}_k\boldsymbol{y}\|\leqslant-\dfrac{\alpha\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k}\|\boldsymbol{d}_k\|.\end{cases} \tag{10}$$

这里 $\boldsymbol{y}\in\boldsymbol{R}^{m_k+1}$,$\boldsymbol{d}_k=-\boldsymbol{B}_k^{-1}\boldsymbol{g}_k$,$V_k=(\boldsymbol{d}_k,\Delta\boldsymbol{x}_{k-1},\cdots,\Delta\boldsymbol{x}_{k-m_k})\in\boldsymbol{R}^{n\times(m_k+1)}$,其中 $\Delta\boldsymbol{x}_{k-1}=\boldsymbol{x}_k-\boldsymbol{x}_{k-1},\cdots,\Delta\boldsymbol{x}_{k-m_k}=\boldsymbol{x}_{k-m_k+1}-\boldsymbol{x}_{k-m_k}$。

步骤 5 利用式(8)修正 $\boldsymbol{B}_k$ 得到 $\boldsymbol{B}_{k+1}$,令 $k:=k+1$,转步骤 2。

注 1 由于 $\boldsymbol{A}_k$ 有三种选择(5)、(6)及 $\boldsymbol{A}_k=0$,因此有三种不同的校正形式,对应三种不同的算法。当 $\boldsymbol{\eta}=0$ 时,则 $D_k=f_k$,说明算法是单调下降算法。

注 2 新算法的主要工作量集中在求解子问题(10),而由于子问题(10)维数一般不高(一般取 $m=2,3$ 即可)且 $\boldsymbol{B}_k$ 是实对称正定对角矩阵,因此子问题(10)的求解相对变得简便。

引理 1 如果 $\boldsymbol{y}_k(\alpha)$ 是式(10)的解,则 $q_k(0)-q_k(\boldsymbol{y}_k(\alpha))\geqslant\dfrac{\alpha}{2}\cdot\dfrac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k}\geqslant\dfrac{\alpha\bar{b}}{2\underline{b}^2}\|\boldsymbol{g}_k\|^2$。

证明 由算法知,$\{\boldsymbol{B}_k\}$ 是稀疏对角正定矩阵,因此 $\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k>0$。如果 $\bar{\boldsymbol{y}}_k=(y'_k,0,\cdots,0)\in\boldsymbol{R}^{m_k+1}$ 且 $y'_k=\dfrac{-\alpha\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k}$,则 $\boldsymbol{y}_k$ 是式(10)的一个可行解,注意到 $0<\alpha\leqslant 1$ 知

$$q_k(0)-q_k(\boldsymbol{y}_k(\alpha))\geqslant q_k(0)-q_k(\bar{\boldsymbol{y}}_k)=\alpha\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k}-\frac{1}{2}\left(\frac{\alpha\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k}\right)^2\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k\geqslant\alpha\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k}-$$

$$\frac{1}{2}\alpha\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k}=\frac{1}{2}\alpha\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k}\geqslant\frac{\alpha}{2}\cdot\frac{\left(\frac{1}{\underline{b}}\|\boldsymbol{g}_k\|^2\right)^2}{\bar{b}\|\boldsymbol{d}_k\|^2}\geqslant\frac{\alpha}{2}\cdot\frac{\left(\frac{1}{\underline{b}}\|\boldsymbol{g}_k\|^2\right)^2}{\bar{b}\left(\frac{1}{\bar{b}}\|\boldsymbol{g}_k\|^2\right)}=\frac{\alpha\bar{b}}{2\underline{b}^2}\|\boldsymbol{g}_k\|^2.$$

3 全局收敛性

首先对目标函数 $f(\boldsymbol{x})$ 做如下假设:

(H1)目标函数 $f(\boldsymbol{x})$ 在 R^n 上有下界。

(H2)目标函数的梯度 $g(\boldsymbol{x})=\nabla f(\boldsymbol{x})$ 在包含水平集 $L(x_0)=\{\boldsymbol{x}\in R^n\mid f(\boldsymbol{x})\leqslant f(\boldsymbol{x}_0)\}$ 的开凸集 B 上 Lipschitz 连续,即存在 $L>0$ 满足:$\|g(\boldsymbol{x})-g(\boldsymbol{y})\|$

$\leqslant L\|\boldsymbol{x}-\boldsymbol{y}\|,\ \forall \boldsymbol{x},\boldsymbol{y}\in B$。

引理 2　假设条件(H1)和(H2)成立,$\forall \boldsymbol{x},\boldsymbol{x}+\boldsymbol{p}\in B$,则 $f(\boldsymbol{x}+\boldsymbol{p})-f(\boldsymbol{x})\leqslant g(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{p}+\frac{1}{2}L\|\boldsymbol{p}\|^2$。

引理 3　假设条件(H1)和(H2)成立,$\{\boldsymbol{x}_k\}$ 是由算法产生的无穷点列,则有

(i) $f(\boldsymbol{x}_{k+1})\leqslant D_k,\ \forall k$;

(ii) $f(\boldsymbol{x}_k)\leqslant D_k,\ \forall k$;

(iii) $\{D_k\}$ 是单调不增序列。

证明　由式(9)和引理 1 知

$$D_k-f(\boldsymbol{x}_{k+1})\geqslant\mu(q_k(\boldsymbol{0})-q_k(\boldsymbol{y}_k(\alpha)))\geqslant\frac{1}{2}\mu\alpha_k\cdot\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k}.$$

即

$$f(\boldsymbol{x}_{k+1})\leqslant D_k-\frac{1}{2}\mu\alpha_k\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k}\leqslant D_k-\frac{1}{2}\mu\alpha_k\frac{\bar{b}}{\underline{b}^2}\|\boldsymbol{g}_k\|^2.\tag{11}$$

因此 $f(\boldsymbol{x}_{k+1})\leqslant D_k,\ \forall k$。由式(3)和式(11)知

$$D_k=\eta D_{k-1}+(1-\eta)f(\boldsymbol{x}_k)\geqslant\eta f(\boldsymbol{x}_k)+(1-\eta)f(\boldsymbol{x}_k)=f(\boldsymbol{x}_k).$$

再由 $D_0=f(\boldsymbol{x}_0)$ 知 $f(\boldsymbol{x}_k)\leqslant D_k,\ \forall k$。由式(3)和式(11)知

$$D_k=\eta D_{k-1}+(1-\eta)f(\boldsymbol{x}_k)\leqslant\eta D_{k-1}+(1-\eta)D_{k-1}=D_{k-1},\ \forall k.$$

故 $\{D_k\}$ 是单调不增序列。

定理 1　假设(H1)和(H2)成立,$\{\boldsymbol{x}_k\}$ 是由算法产生的无穷迭代点列,则 $\lim\limits_{k\to\infty}\|\boldsymbol{g}_k\|=0$。

证明　(反证法)假设结论不成立,则存在 $\varepsilon_0>0$ 和无穷子集 $K\subset\{1,2,\cdots\}$ 满足:

$$\|\boldsymbol{g}_k\|\geqslant\varepsilon_0,\ \forall k\in K.\tag{12}$$

由式(3)和式(11)知

$$D_{k+1}=\eta D_k+(1-\eta)f(\boldsymbol{x}_{k+1})\leqslant\eta D_k+(1-\eta)\left(D_k-\frac{1}{2}\mu\alpha_k\frac{\bar{b}}{\underline{b}^2}\cdot\|\boldsymbol{x}_k\|^2\right)=D_k-\frac{1}{2}(1-\eta)\mu\alpha_k\frac{\bar{b}}{\underline{b}^2}\|\boldsymbol{g}_k\|^2.$$

即

$$D_{k+1}\leqslant D_k-\frac{1}{2}(1-\eta)\mu\alpha_k\frac{\bar{b}}{\underline{b}^2}\|\boldsymbol{g}_k\|^2.\tag{13}$$

由 $f(\boldsymbol{x})$ 有下界,$f_k\leqslant D_k$ 和式(13)知,$\{D_k\}$ 单调下降且有下界,因而收敛,因而有

$$\lim_{k\to\infty,k\in K}\alpha_k=0.\tag{14}$$

由式(9)知

$$\frac{D_k-f(\boldsymbol{x}_k+\boldsymbol{P}_k(\alpha_k/\rho))}{q_k(\boldsymbol{0})-q_k(\boldsymbol{y}_k(\alpha_k/\rho))}<\mu,\ \forall k\in K.\tag{15}$$

再由中值定理、式(14)、引理1和引理2,令 $\alpha=\frac{\alpha_k}{\rho}$ 得

$$\left|\frac{f_k-f(\boldsymbol{x}_k+\boldsymbol{P}_k(\alpha))}{q_k(\boldsymbol{0})-q_k(\boldsymbol{y}_k(\alpha))}-1\right|=\left|\frac{f_k-f(\boldsymbol{x}_k+\boldsymbol{P}_k(\alpha))-\boldsymbol{q}_k(\boldsymbol{0})+q_k(\boldsymbol{y}_k(\alpha))}{q_k(\boldsymbol{0})-q_k(\boldsymbol{y}_k(\alpha))}\right|\leqslant$$

$$\frac{\frac{1}{2}L\|\boldsymbol{P}_k(\alpha)\|^2+\frac{1}{2}\bar{b}\|\boldsymbol{P}_k(\alpha)\|^2}{q_k(\boldsymbol{0})-q_k(\boldsymbol{y}_k(\alpha))}\leqslant\frac{\frac{1}{2}(L+\bar{b})\|\boldsymbol{P}_k(\alpha)\|^2}{\frac{1}{2}\alpha\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k}}\leqslant$$

$$\frac{\frac{1}{2}(L+\bar{b})\alpha^2(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2\|\boldsymbol{d}_k\|^2/(\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k)^2}{\frac{\alpha}{2}\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k}}=$$

$$(L+\bar{b})\alpha\frac{\|\boldsymbol{d}_k\|^2}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}_k}\leqslant\frac{(L+\bar{b})}{\underline{b}}\alpha\to 0,(k\in K,k\to\infty).\tag{16}$$

故当 $\alpha\leqslant\frac{(1-\mu)\underline{b}}{(L+\bar{b})}$ 时有

$$\frac{D_k-f(\boldsymbol{x}_k+\boldsymbol{P}_k(\alpha))}{q_k(\boldsymbol{0})-q_k(\boldsymbol{y}_k(\alpha))}\geqslant\frac{f(\boldsymbol{x}_k)-f(\boldsymbol{x}_k+\boldsymbol{P}_k(\alpha))}{q_k(\boldsymbol{0})-q_k(\boldsymbol{y}_k(\alpha))}\geqslant 1-\frac{L+\bar{b}}{\underline{b}}\alpha\geqslant\mu,\ \forall k\in K.$$

此与式(15)矛盾。

4　线性收敛速度和超线性收敛速度

线性收敛速度分析需要以下假设条件:

(H3) $f(\boldsymbol{x})$ 是强凸函数,即存在常数 $r>0$ 满足

$$f(\boldsymbol{x})>f(\boldsymbol{y})+g(\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y})+\frac{1}{2r}\|\boldsymbol{x}-\boldsymbol{y}\|^2,\ \forall \boldsymbol{x},\boldsymbol{y}\in\boldsymbol{R}^n.\tag{17}$$

定理 2　设 $\{\boldsymbol{x}_k,\alpha_k,\boldsymbol{g}_k\}$ 是由算法产生的序列,假设(H1)～(H3)成立,则存在 $\theta\in(0,1)$ 满足

$$f(\boldsymbol{x}_k)-f(\boldsymbol{x}^*)\leqslant\theta^k(f(\boldsymbol{x}_0)-f(\boldsymbol{x}^*)),\ \forall k.$$

即 $\{f_k\}R-$ 线性收敛于 $f(\boldsymbol{x}^*)$。

超线性收敛速度分析需要以下假设条件:

(H4) 算法产生点列 $\{\boldsymbol{x}_k\}$ 收敛于 $\boldsymbol{x}^*$,$\nabla^2 f(\boldsymbol{x}^*)$ 是一个正定矩阵,且 $f(\boldsymbol{x})$ 在 $N(\boldsymbol{x}^*,\varepsilon_0)=\{\boldsymbol{x}\mid\|\boldsymbol{x}-\boldsymbol{x}^*\|\leqslant\varepsilon_0\}$ 二阶连续可微。

(H5) 算法中 $\boldsymbol{d}_k=-\boldsymbol{B}_k^{-1}\boldsymbol{g}_k$ 满足

$$\lim_{k\to\infty}\frac{\|(\boldsymbol{B}_k-\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d}_k)\|}{\|\boldsymbol{d}_k\|}=0.$$

引理 4 假设条件(H4),(H5)成立,$\{\boldsymbol{x}_k\}$是由算法产生的无穷迭代点列,则存在k'使得$\alpha_k=1$,$\forall k\geqslant k'$。

定理 3 假设条件(H3),(H4),(H5)成立,$\{\boldsymbol{x}_k\}$是由算法产生的无穷迭代点列,则$\{\boldsymbol{x}_k\}$超线性收敛于$\boldsymbol{x}^*$。

5 数值试验

从网站:www. ici. ro/camo/neculai/SCALCG/testuo. paf 选择了 2 个算例,利用 Matlab7. 0 编制程序在 PIII. 933 机器上对本文算法进行数值试验,并与其单调算法进行比较。

本文的算法记为 SM。当$\boldsymbol{A}_k$取(5)、(6)及$\boldsymbol{A}_k=0$时对应的算法为 SM(1)、SM(2)及 SM(0),算法 SM(1)、SM(2)及 SM(0)中取$\eta_k\equiv 0$分别对应其单调算法,分别记为 MSM(1)、MSM(2)及 MSM(0)。当$\boldsymbol{B}_k\equiv\boldsymbol{I}_n$时,记算法为 SGM,算法 SGM 中取$\eta\equiv 0$对应其单调算法,记为 MSGM。当$\boldsymbol{B}_k$分别用 DFP,BFGS 公式修正时,算法分别记为 DSM、BSM,算法 DSM、BSM 中取$\eta\equiv 0$分别对应其单调算法即为文献[14]中算法,分别记为 MDSM、MBSM。

算法中取$\eta=0.36$,$\mu=0.38$,$\rho=0.5$,$m=3$,$\boldsymbol{B}_0=\boldsymbol{I}_n$,和$\bar{b}$取为变化形式,即

$$\underline{b_k}=\max\{0.8|(\boldsymbol{g}_{k+1}-\boldsymbol{g}_k)^{\mathrm{T}}(\boldsymbol{x}_{k+1}-\boldsymbol{x}_k)|/\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|^2,0.000001\},$$

$$\overline{b_k}=\max\{2.13|(\boldsymbol{g}_{k+1}-\boldsymbol{g}_k)^{\mathrm{T}}(\boldsymbol{x}_{k+1}-\boldsymbol{x}_k)|/\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|^2,100000\},$$

算法终止准则采用$\|\boldsymbol{g}_k\|\leqslant 10^{-3}$,给出$n=100$,1 000,10 000,20 000 的计算结果。

例 1 $f(x)=\sum_{i=1}^{n}\left[\left(n-\sum_{j=1}^{n}\cos x_j\right)+\mathrm{i}(1-\cos x_i)-\sin x_i\right]^2$,

初始点$\boldsymbol{x}_0=(0.2,0.2,\cdots,0.2)^{\mathrm{T}}$,$f_{\mathrm{opt}}=0$。计算结果见表 1(其中,“* * *”表示迭代时间大于 300 s 或者迭代次数超过 3 000 仍未达到停机标准)。

表 1 例 1 的数值结果

Table 1 Numerical results of example 1

算法	维数	迭代次数	迭代时间(s)	最优值
SM(1)/MSM(1)	100	23/20	0.040/0.032	$7.7036\times10^{-6}/8.0663\times10^{-6}$
	1 000	8/8	0.037/0.036	$4.3318\times10^{-6}/4.8779\times10^{-6}$
	10 000	2/2	0.087/0.082	$1.4345\times10^{-6}/1.4807\times10^{-6}$
	20 000	3/2	0.267/0.163	$7.1344\times10^{-7}/7.4032\times10^{-7}$
SM(2)/MSM(2)	100	21/18	0.020/0.016	$7.8208\times10^{-6}/7.0327\times10^{-6}$
	1 000	8/8	0.039/0.035	$4.3720\times10^{-6}/4.8138\times10^{-6}$
	10 000	2/2	0.086/0.081	$1.4406\times10^{-6}/1.4873\times10^{-6}$
	20 000	2/2	0.172/0.162	$7.2027\times10^{-7}/7.4363\times10^{-7}$
SM(0)/MSM(0)	100	24/21	0.024/0.014	$4.5619\times10^{-6}/3.9341\times10^{-6}$
	1 000	12/11	0.054/0.045	$3.2139\times10^{-6}/3.3293\times10^{-6}$
	10 000	3/3	0.120/0.113	$1.1890\times10^{-6}/1.2702\times10^{-6}$
	20 000	2/2	0.163/0.155	$7.7081\times10^{-7}/7.9490\times10^{-7}$
SGM/MSGM	100	2 136/2 114	1.124/1.050	$4.6044\times10^{-6}/4.6000\times10^{-6}$
	1 000	2 374/2 343	4.720/4.799	$1.2390\times10^{-6}/1.2381\times10^{-6}$
	10 000	687/684	10.893/11.191	$7.6263\times10^{-7}/7.6244\times10^{-7}$
	20 000	314/314	10.240/10.581	$6.0823\times10^{-7}/6.0823\times10^{-7}$
DSM/MDSM	100	1 379/* * *	11.422/* * *	2.1592×10^{-6}/* * *
	1 000	* * */* * *	* * */* * *	* * */* * *
	10 000	* * */* * *	* * */* * *	* * */* * *
	20 000	* * */* * *	* * */* * *	* * */* * *
BSM/MBSM	100	107/105	0.821/0.702	$4.4236\times10^{-6}/4.3086\times10^{-6}$
	1 000	36/40	103.811/116.315	$3.9860\times10^{-6}/3.4997\times10^{-6}$
	10 000	* * */* * *	* * */* * *	* * */* * *
	20 000	* * */* * *	* * */* * *	* * */* * *

例 2 $f(\boldsymbol{x})=\sum_{i=1}^{n}f_i(\boldsymbol{x})^2$,$f_i(\boldsymbol{x})=(3-2x_i)x_i-x_{i-1}-2x_{i+1}+1$,$i=1,2,\cdots,n$ 初始值$\boldsymbol{x}_0=(-1,-1,\cdots,-1)^{\mathrm{T}}$,$f_{\mathrm{opt}}=0$。数值结果见表 2。

表 2 例 2 的数值结果

Table 1 Numerical results of example 2

算法	维数	迭代次数	迭代时间(s)	最优值
SM(1)/MSM(1)	100	39/48	0.147/0.061	$1.4508\times10^{-8}/4.7257\times10^{-9}$
	1 000	35/37	0.160/0.173	$1.6149\times10^{-8}/2.4634\times10^{-8}$
	10 000	36/38	1.402/1.505	$7.9357\times10^{-9}/1.5054\times10^{-8}$
	20 000	32/40	2.580/3.192	$5.2163\times10^{-9}/1.8635\times10^{-8}$
SM(2)/MSM(2)	100	37/44	0.047/0.045	$2.8977\times10^{-9}/8.0134\times10^{-9}$
	1 000	39/52	0.183/0.235	$1.7915\times10^{-9}/2.8030\times10^{-8}$
	10 000	32/50	1.253/1.967	$1.1918\times10^{-8}/9.3147\times10^{-9}$
	20 000	35/55	2.773/4.445	$1.5838\times10^{-8}/3.5264\times10^{-11}$
SM(0)/MSM(0)	100	46/43	0.048/0.062	$1.6613\times10^{-9}/1.5850\times10^{-8}$
	1 000	35/41	0.183/0.208	$1.9999\times10^{-8}/1.9878\times10^{-8}$
	10000	36/45	1.528/1.949	$1.5281\times10^{-8}/2.6040\times10^{-8}$
	20 000	33/41	2.877/3.688	$2.3957\times10^{-8}/2.2439\times10^{-8}$
SGM/MSGM	100	154/140	0.140/0.149	$3.1727\times10^{-8}/3.1297\times10^{-8}$
	1 000	1368/143	2.549/0.269	$3.9707\times10^{-1}/3.1606\times10^{-8}$
	10 000	1291/788	16.103/9.281	$3.9707\times10^{-1}/7.1253\times10^{-1}$
	20 000	986/146	26.561/3.832	2.3259e+000/3.1995×10^{-8}
DSM/MDSM	100	91/79	0.874/0.833	$4.7728\times10^{-9}/7.7943\times10^{-9}$
	1 000	***/***	***/***	***/***
	10 000	***/***	***/***	***/***
	20 000	***/***	***/***	***/***
BSM/MBSM	100	43/33	0.509/0.250	$6.2410\times10^{-9}/1.5792\times10^{-8}$
	1 000	55/39	170.334/117.095	$3.9707\times10^{-1}/8.6952\times10^{-9}$
	10 000	***/***	***/***	***/***
	20 000	***/***	***/***	***/***

6 结 论

本文基于信赖域技术和修正的拟牛顿方程,结合 Neng-Zhu Gu 非单调策略,设计了新的求解无约束最优化问题的非单调超记忆梯度算法,分析了算法的收敛性和收敛速度。新算法 $\boldsymbol{B}_k$ 用公式(8)校正比用 BFGS 公式和 DFP 公式校正在计算大型例子需要更少的存储量、计算量、迭代时间和迭代次数,目标函数值更接近于最优值,问题规模越大,维数越高,新算法优势越明显,因此新算法更适合于大规模问题的计算。另外,非单调算法的数值表现优于单调算法。

参考文献:

[1] 袁亚湘,孙文渝. 最优化理论与方法[M]. 北京:科学出版社,1997.

[2] 戴彧虹,袁亚湘. 非线性规划共轭梯度算法[M]. 上海:上海科学技术出版社,2000.

[3] MIELE A, CANTRELL J W. Memory gradient method for the minimization of functions[J]. JOTA, 1969,3:459-470.

[4] CRAGG E E, LEVY A V. Study on a memory gradient method for the minimization of functions[J]. JOTA, 1969,4(3):191-205.

[5] WOLFE M A. A quasi-Newton method with memory for unconstrained function minimization[J]. J Inst Maths Applics, 1975,15:85-94.

[6] WOLFE M A, VIAZMINSKY C. Super-memory descent methods for unconstrained function minimization[J]. JOTA, 1976,18:455-468.

[7] 孙麟平. 无约束极小化自适应多信息下降算法[J]. 高校计算数学学报, 1982,14(2):107-114.
SUN Lin-ping. Adaptive supermemory descent method for unconstrained minimization [J]. Numerical Mathematics: A Journal of Chinese Universities, 1982,14(2):107-114.

[8] 赵庆祯. 一个改进的超记忆梯度算法及其敛速估计[J]. 应用数学学报, 1983,6(3):376-385.
ZHAO Qing-zhen. Convergence and rate of convergence of an improved supermemory gradient method[J]. Acta Mathematicae Applicatae Sinica, 1983,6(3):376-385.

[9] 胡宗英. 一个改进的记忆梯度算法[J]. 高等学校计算数学学报, 1989,2:173-179.
HU Zong-ying. A modified memory gradient algorithm [J]. Numerical Mathematics: A Journal of Chinese Universities, 1989,2:173-179.

[10] 时贞军. 无约束优化的超记忆梯度算法[J]. 工程数学学报, 2000,17(2):99-104.
SHI Zhen-jun. A supermemory gradient method for un-

constrained optimization problem [J]. Chinese Journal of Engineering Mathematics, 2000,17(2):99-104.

[11] 时贞军. 改进 HS 共轭梯度算法及其全局收敛性[J]. 计算数学, 2001,23(4):393-406.
SHI Zhen-jun. Modified HS conjugate gradient method and its global convergence [J]. Mathematica Numerica Sinica, 2001,23(4):393-406.

[12] 孙清滢, 刘新海. 结合广义 Armijo 步长搜索的一类新的三项共轭梯度算法及其收敛特征[J]. 计算数学, 2004,26(1):25-36.
SUN Qing-ying, LIU Xin-hai. Global convergence results of a new three terms conjugate gradient method with generalized Armijo step size rule[J]. Mathematica Numerica Sinica, 2004,26(1):25-36.

[13] SUN Qingying. Global convergence results of a Thtee term memory gradient method with non-monotone line search technique [J]. ACTA Mathematics Scientis, 2005,25B(1):170-178.

[14] SHI Zhen-jun, SHEN Jie. A new class of super-memory gradient methods[J]. Applied Mathematicd and Computation, 2006,183:748-760.

[15] GRIPPO L, LAMPARIELLO F, LUCIDI S. A nonmonotone line search technique for Newton's method[J]. SIAM J, NUMER ANAL, 1986,23(4):707-716.

[16] GU Neng-zhu, MO Jiang-tao. Incorporating nonmonotone strategies inti the trust region method for unconstrained optimization [J]. Computers & Mathematics with Applications, 2007, doi: 10.1016/j. camwa. 2007.08.038.

[17] 孙清滢, 崔彬, 王长钰. 新非单调线搜索规则的 Lampariello 修正对角稀疏拟牛顿算法[J]. 计算数学, 2008,30(3):255-268.
SUN Qing-ying, CUI Bin, WANG Chang-yu. Global convergence of a Lampariello modified diagonal-sparse quasi-Newton method with new non-monotone step size rule[J]. Mathematica Numerica Sinica, 2008,30(3):255-268.

[18] 孙清滢, 郑艳梅. 大步长非单调线搜索规则的 Lampariello 修正对角稀疏拟牛顿算法的全局收敛性[J]. 数学进展, 2008,37(3):311-320.
SUN Qing-ying, ZHENG Yan-mei. Global convergence results of Lampariello modified diagonal-sparse quasi-Newton method with larger non-monotone-step size rule [J]. Advances in Mathematics, 2008,37(3):311-320.

[19] ZHANG H C, HAGER William W. A nonmonotone line search technique and its application to unconstrained optimization[J]. SIAM J Optim, 2004, 14(4):1043-1056.

[20] WEI Z, LI G, QI L. New quasi-Newton methords for unconstrained optimization problems [J]. Applied Mathematics and Computation, 2006,175:1156-1188.

(编辑　修荣荣)

(上接第 185 页)

[22] FENG Hui, YAO Shan, LI Lin-yuan, et al. Spline-based nonparametric estimation of the altimeter sea-state bias correction [J]. IEEE Geoscience and Remote Sensing Letters, 2010,7(3):577-581.

[23] TRAN N, LABROUE S, PHILIPPS S, et al. Overview and update of the sea state bias corrections for the Jason-2, Jason-1 and TOPEX missions [J]. Marine Geodesy, 2010,33(S1):348-362.

[24] WITTER D L, CHELTON D B. An apparent wave height dependence in the sea-state bias in GEOSAT altimeter range measurements [J]. Journal of Geophysical Research, 1991b,96:8861-8867.

(编辑　修荣荣)