文章编号:1673-5005(2009)01-0162-05

结合广义 Armijo 步长搜索的带误差项的记忆梯度算法

孙清滢, 桑兆阳, 吕 炜

(中国石油大学 数学与计算科学学院,山东 东营 257061)

摘要:对非线性无约束规划提出了结合广义 Armijo 步长搜索规则的一类带误差项的记忆梯度求解算法,在目标函数梯度一致连续的条件下,证明了算法的全局收敛性,同时给出带误差项的结合拟-Newton 方程的记忆梯度算法。数值结果表明算法是有效的。

关键词:无约束最优化; 带误差项的记忆梯度法; 广义 Armijo 步长搜索规则; 全局收敛; 数值试验

中图分类号:O 221.2　　**文献标识码**:A

Memory gradient method with errors and generalized Armijo step

SUN Qing-ying, SANG Zhao-yang, LÜ Wei

(*College of Mathematics and Computational Science in China University of Petroleum, Dongying 257061, China*)

Abstract: A new class of memory gradient methods with errors and generalized Armijo step size rule were proposed for nonlinear unconstrained optimization assuming that the gradient of the function is uniformly continuous. Its global convergence property was proved. And a novel memory gradient method with quasi-Newton method and errors was given. Numerical results show that the new methods are efficient.

Key words: unconstrained optimization; memory gradient method with errors; generalized Armijo step size rule; global convergence; numerical experiment

1 问题的提出

考虑无约束优化问题:$(p)\ \min_{x\in R^n} f(x)$,其中 $f(x):\mathbf{R}^n \to \mathbf{R}^1$ 是一阶连续可微函数。求解问题(p)的共轭梯度法收敛速度快,存储量小,适于求解大规模问题[1]。记 $g_k = \nabla f(x_k)$,它具有如下迭代形式:

$$x_{k+1} = x_k + \lambda_k s_k,$$

其中

$$s_k = \begin{cases} -g_k, & k = 1, \\ -g_k + \beta_k s_{k-1}, & k \geq 2. \end{cases}$$

式中,λ_k 为步长,可通过某种策略确定;β_k 为参数,不同算法的 β_k 不同,如 FR, PR, HS 共轭梯度法的 β_k 值分别为

$$\beta_k^{\mathrm{FR}} = \|g_k\|^2 / \|g_{k-1}\|^2,$$

$$\beta_k^{\mathrm{PR}} = (g_k^{\mathrm{T}}(g_k - g_{k-1}))/\|g_{k-1}\|^2,$$

$$\beta_k^{\mathrm{HS}} = (g_k^{\mathrm{T}}(g_k - g_{k-1}))/(d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})).$$

三者比较,FR 算法具有较好的收敛性质,而后两者数值表现差不多,都优于 FR 算法。

若不采取重开始技巧,标准共轭梯度法最多线性收敛,然而重开始策略抛弃了算法沿前一搜索方向得到的二阶信息,且在数值上重开始时函数的即刻下降量常小于不重开始时函数的即刻下降量。故文献[2-4]中研究了三项共轭梯度法,三项共轭梯度法是共轭梯度法的推广。记忆梯度法是共轭梯度法的变形与改进,三项记忆梯度法是记忆梯度法的推广,文献[5-6]中研究了一类数值效果有效的三项记忆梯度下降算法,算法搜索方向中的参数取值是一个区间。然而,由于计算机的舍入误差和随机产生的误差常常破坏搜索方向的下降性,因此研究带误差项的算法具有实际应用价值。对于此问题的

收稿日期:2008-01-05

基金项目:国家自然科学基金项目(10571106);中国石油大学博士基金项目(Y040804)

作者简介:孙清滢(1966-),男(汉族),山东青州人,教授,博士,从事非线性规划理论和算法研究。

研究, Bertsekas 和 Tsitsiklis[7] 对带误差项的梯度算法在发散级数步长搜索下给出算法的全局收敛性分析, 但算法收敛分析过程需要一个较强的假设条件, 即 $f(\boldsymbol{x})$ 的梯度 $\nabla f(\boldsymbol{x})$ 满足 Lipschitz 条件, 即存在常数 $L>0$ 使得

$$\|\nabla f(\boldsymbol{x})-\nabla f(\bar{\boldsymbol{x}})\|\leqslant L\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|,\ \forall\ \boldsymbol{x},\bar{\boldsymbol{x}}\in\mathbf{R}^n.$$

本文中研究带误差项的一类新的记忆梯度算法, 算法中参数的取值范围比文献[5-6] 中的更大, 并在 $\nabla f(\boldsymbol{x})$ 一致连续和结合广义 Armijo 步长搜索条件下讨论算法的全局收敛性, 同时给出带误差项的结合拟 -Newton 方程的记忆梯度算法。

2　带误差项的记忆梯度算法

带误差项的记忆梯度算法(MGME) 的迭代形式为

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\lambda_k(\boldsymbol{s}_k+\boldsymbol{w}_k).\tag{1}$$

(a) 主方向:

$$\boldsymbol{s}_k=-\nabla f(\boldsymbol{x}_k)+\beta_k\boldsymbol{s}_{k-1},\tag{2}$$

其中 β_k 满足

$$\begin{cases}\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\nabla f(\boldsymbol{x}_k)>|\beta_k\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{s}_{k-1}|,\\ |\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{d}_k|\geqslant(|\cos\theta_k|+\Delta)|\beta_k|\|\nabla f(\boldsymbol{x}_k)\|\|\boldsymbol{s}_{k-1}\|.\end{cases}\tag{3}$$

式中, $\Delta>0$ 为参数; θ_k 为 $\nabla f(\boldsymbol{x}_k)$ 和 $\boldsymbol{s}_{k-1}$ 的夹角。

(b) 误差项 $\boldsymbol{w}_k$ 满足

$$\|\boldsymbol{w}_k\|\leqslant\gamma_k(q+p\|\nabla f(\boldsymbol{x}_k)\|),\tag{4}$$

其中 $p,q>0$ 为常数, γ_k 满足

$$\sum_{k=1}^{\infty}\gamma_k^2<+\infty,\ \gamma_k>0,k=1,2,\cdots.\tag{5}$$

由式(3), 仿文献[5-6] 可取

$$\beta_k\in[-\underline{\beta}_k(\Delta),\bar{\beta}_k(\Delta)],\tag{6}$$

其中

$$\underline{\beta}_k(\Delta)=\frac{1}{|\cos\theta_k|+\Delta-\cos\theta_k}\frac{\|\nabla f(\boldsymbol{x}_k)\|}{\|\boldsymbol{s}_{k-1}\|},\tag{7}$$

$$\bar{\beta}_k(\Delta)=\frac{1}{|\cos\theta_k|+\Delta+\cos\theta_k}\frac{\|\nabla f(\boldsymbol{x}_k)\|}{\|\boldsymbol{s}_{k-1}\|}.\tag{8}$$

(c) λ_k 为步长, 由下列规则确定:

如果 $\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}(\boldsymbol{s}_k+\boldsymbol{w}_k)\geqslant0$, 取 $\lambda_k=\gamma_k$; 否则, 取 λ_k 满足

$$\begin{cases}f(\boldsymbol{x}_k+\lambda_k(\boldsymbol{s}_k+\boldsymbol{w}_k))\leqslant f(\boldsymbol{x}_k)+\\ \lambda_k\mu_1\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}(\boldsymbol{s}_k+\boldsymbol{w}_k),\\ \text{且 }\lambda_k\geqslant\bar{\gamma}_1\text{ 或者 }\ \lambda_k\geqslant\bar{\gamma}_2\lambda_k^*>0.\end{cases}\tag{9}$$

其中 λ_k^* 满足

$$f(\boldsymbol{x}_k+\lambda_k^*(\boldsymbol{s}_k+\boldsymbol{w}_k))>f(\boldsymbol{x}_k)+\mu_2\lambda_k^*\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}(\boldsymbol{s}_k+\boldsymbol{w}_k).\tag{10}$$

式中, $\mu_1,\mu_2\in(0,1)$ 且 $\mu_1\leqslant\mu_2$, $\bar{\gamma}_1,\bar{\gamma}_2>0$ 为常数。

注　Y. H. Dai, L. Z. Liao 在文献[9] 中结合共轭梯度算法与拟 -Newton 方程, 得到结合拟 -Newton 方程的参数 $\beta_k^{\text{Quasi-Newton}}$ 的取法为

$$\beta_k^{\text{Quasi-Newton}}=\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{y}_{k-1}-\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\boldsymbol{s}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}},$$

其中

$$\boldsymbol{d}_{k-1}=\boldsymbol{x}_k-\boldsymbol{x}_{k-1},\ \boldsymbol{y}_{k-1}=\boldsymbol{g}_k-\boldsymbol{g}_{k-1}.$$

结合以上参数, 在 MGME 中可选取参数 $\beta_k=\operatorname{argmin}\{|\beta-\beta_k^{\text{Quasi-Newton}}|:\beta\in[-\underline{\beta}_k(\Delta),\bar{\beta}_k(\Delta)]\}$, 从而得到结合拟 -Newton 方程的带误差项的记忆梯度法, 记为 QGME。

引理 1　若 $\boldsymbol{x}_k$ 不是问题(p) 的稳定点, β_k 满足式(6)~(8), 则有 $\|\boldsymbol{s}_k\|\leqslant c_1\|\nabla f(\boldsymbol{x}_k)\|$, 其中 $c_1=1+\frac{1}{\Delta}$。

证明　由 $\boldsymbol{s}_k$ 的定义易证。

引理 2　若 $\boldsymbol{x}_k$ 不是问题(p) 的稳定点, β_k 满足式(6)~(8), 则有 $\nabla f(\boldsymbol{x}_k)_k^{\mathrm{T}}\leqslant-c_2\|\nabla f(\boldsymbol{x}_k)\|^2$, 其中 $c_2=\frac{\Delta}{1+\Delta}$。

证明　分两种情况证明。

(a) $k=1$, 显然成立。

(b) $k\geqslant2$, 由 $\boldsymbol{s}_k$ 的定义和式(3) 知

$$\begin{aligned}&\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{s}_k=-\|\nabla f(\boldsymbol{x}_k)\|^2+\beta_k\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{s}_{k-1}\leqslant\\&-\|\nabla f(\boldsymbol{x}_k)\|^2+|\beta_k\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{s}_{k-1}|\leqslant\\&-\|\nabla f(\boldsymbol{x}_k)\|^2+|\beta_k|\|\nabla f(\boldsymbol{x}_k)\|\|\boldsymbol{s}_{k-1}\|\leqslant\\&-\|\nabla f(\boldsymbol{x}_k)\|^2+\frac{|\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{s}_k|}{|\cos\theta_k|+\Delta}\leqslant\\&-\|\nabla f(\boldsymbol{x}_k)\|^2-\frac{\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{s}_k}{|\cos\theta_k|+\Delta}.\end{aligned}$$

因此

$$\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{s}_k\leqslant-\frac{|\cos\theta_k|+\Delta}{1+|\cos\theta_k|+\Delta}\|\nabla f(\boldsymbol{x}_k)\|^2.$$

再由 $\frac{\Delta}{1+\Delta}\leqslant\frac{|\cos\theta_k|+\Delta}{1+|\cos\theta_k|+\Delta}$ 知

$$\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{s}_k\leqslant-\frac{\Delta}{1+\Delta}\|\nabla f(\boldsymbol{x}_k)\|^2.$$ 证毕。

设集合 $I=\{k\in N|\ \nabla f(\boldsymbol{x}_k)^{\mathrm{T}}(\boldsymbol{s}_k+\boldsymbol{w}_k)\geqslant0\}$,

集合 $J = N\backslash I$，其中 $N = \{1, 2, \cdots\}$。

引理3 算法中，假设β_k满足式(6) ~ (8)，w_k满足式(4)，(5)，则当 $k \in I$ 充分大时有 $\|\nabla f(x_k)\| \leqslant \sigma\gamma_k$，其中 $\sigma > 0$ 为常数，进而 $\lim\limits_{k\in I, k\to\infty} \nabla f(x_k) = 0$。

证明 $\forall k \in I$，由 I 的定义知

$$-\nabla f(x_k)_k^{\mathrm{T}} s_k \leqslant \nabla f(x_k)^{\mathrm{T}} w_k.$$

由引理2知

$$-\nabla f(x_k)^{\mathrm{T}} s_k \geqslant c_2 \|\nabla f(x_k)\|^2. \tag{11}$$

由式 (4) 知

$$\nabla f(x_k)^{\mathrm{T}} w_k \leqslant \|\nabla f(x_k)\| \|w_k\| \leqslant \gamma_k q \|\nabla f(x_k)\| + \gamma_k p \|\nabla f(x_k)\|^2. \tag{12}$$

由式(10) ~ (12) 得

$$c_2 \|\nabla f(x_k)\|^2 \leqslant -\nabla f(x_k)^{\mathrm{T}} s_k \leqslant \nabla f(x_k)^{\mathrm{T}} w_k \leqslant \gamma_k q \|\nabla f(x_k)\| + \gamma_k p \|\nabla f(x_k)\|^2,$$

即

$$(c_2 - \gamma_k p) \|\nabla f(x_k)\| \leqslant \gamma_k q.$$

由 $\gamma_k \to 0\ (k\to\infty)$ 知，$\exists\ \sigma > 0$，当$k \in I$充分大时有$\|\nabla f(x_k)\| \leqslant \sigma\gamma_k$，再由式(5) 即得

$$\lim_{k\in I, k\to\infty} \nabla f(x_k) = 0.$$ 证毕。

3 全局收敛性

定理1 设$\{x_k\}$是由算法产生的无穷点列，$\{f(x_k)\}$存在有限极限，如果存在一个开凸集 $D \supset \{x_k\}_{k\in N}$ 使得 $\nabla f(x)$ 在 D 上一致连续，则 $\lim\limits_{k\to\infty} \nabla f(x_k) = 0$。

证明 设 $\lim\limits_{k\to\infty} f(x_k) = f^*$，以下用反证法证明 $\lim\limits_{k\to\infty} \nabla f(x_k) = 0$。

事实上，若否，则存在无穷指标集 $K_0 \subset N$ 和 $\varepsilon_0 > 0$ 使得

$$\|\nabla f(x_k)\| \geqslant \varepsilon_0,\ \forall k \in K_0. \tag{13}$$

由 $k_0 = (I \cap K_0) \cup (J \cap K_0)$知，当$k \in I \cap K_0$时，由引理3知

$$\lim_{k\in I\cap K_0, k\to\infty} \|\nabla f(x_k)\| \leqslant \lim_{k\in I\cap K_0, k\to\infty} \sigma\lambda_k = 0,$$

故$I \cap K_0$最多只有无限个元素，不妨设$I \cap K_0 = \Phi$，故 $K_0 \subset J$，$\forall k \in K_0 \subset J$。由引理2及式 (5) 知，当 $k(k \in K_0)$ 充分大时有

$$-\nabla f(x_k)^{\mathrm{T}}(s_k + w_k) \geqslant c_2 \|\nabla f(x_k)\|^2 - \|\nabla f(x_k)\| \cdot \gamma_k(q + p\|\nabla f(x_k)\|) = (c_2 - \gamma_k p)\|\nabla f(x_k)\|^2 - \gamma_k q \|\nabla f(x_k)\| \geqslant [(c_2 - \gamma_k p)\varepsilon_0 - \gamma_k q] \|\nabla f(x_k)\|. \tag{14}$$

由引理1及式(5) 得

$$\|s_k + w_k\| \leqslant c_1 \|\nabla f(x_k)\| + \gamma_k(q + p\|\nabla f(x_k)\|) \leqslant (c_1 + (\gamma_k q)/\varepsilon_0 + \gamma_k p) \|\nabla f(x_k)\|. \tag{15}$$

由式(9)，(14) 得

$$f(x_k) - f(x_{k+1}) \geqslant -\mu_1 \lambda_k \nabla f(x_k)^{\mathrm{T}}(s_k + w_k) \geqslant \mu_1 \lambda_k [(c_2 - \gamma_k p)\varepsilon_0 - \gamma_k q] \|\nabla f(x_k)\|, \tag{16}$$

令 $k \xrightarrow{k\in K_0} \infty$ 得

$$\lim_{k\in K_0, k\to\infty} \lambda_k \|\nabla f(x_k)\| = 0. \tag{17}$$

由式(13)，(16) 得

$$f(x_k) - f(x_{k+1}) \geqslant -\mu_1 \lambda_k \varepsilon_0 [(c_2 - \gamma_k p)\varepsilon_0 - \gamma_k q],$$

令 $k \xrightarrow{k\in K_0} \infty$ 得

$$\lim_{k\in K_0, k\to\infty} \lambda_k = 0. \tag{18}$$

由式(15)，(17) 得

$$\lim_{k\in K_0, k\to\infty} \lambda_k \|s_k + w_k\| \leqslant \lim_{k\in K_0, k\to\infty} \lambda_k (c_1 + \gamma_k q/\varepsilon_0 + \gamma_k p) \|\nabla f(x_k)\| = 0.$$

因此

$$\lim_{k\in K_0, k\to\infty} \lambda_k \|s_k + w_k\| = 0. \tag{19}$$

由式(18) 知，当$k(k \in K_0)$充分大时有$\lambda_k < \bar{\gamma}_1$，从而$\lambda_k \geqslant \bar{\gamma}_2 \lambda_k^*$，而$\lambda_k^*$满足式(10)，设$x_{k+1}^* = x_k + \lambda_k^*(s_k + w_k)$。由 式(18)，(19) 及 $\lambda_k \geqslant \bar{\gamma}_2 \lambda_k^*$ 知，$k \in K_0$ 充分大时有

$$\lim_{k\in K_0, k\to\infty} \lambda_k^* = 0,\ \text{且} \lim_{k\in K_0, k\to\infty} \lambda_k^* \|s_k + w_k\| = 0.$$

因而

$$\lim_{k\in K_0, k\to\infty} \|x_{k+1}^* - x_k\| = 0.$$

令 $\rho_k^* = \dfrac{f(x_{k+1}^*) - f(x_k)}{\lambda_k^*\ \nabla f(x_k)^{\mathrm{T}}(s_k + w_k)}$，$k \in K_0$，由式(10) 得

$$\rho_k^* < \mu_2 < 1. \tag{20}$$

由式(5)，(13) ~ (15) 得

$$\limsup_{k\in K_0, k\to\infty} |\rho_k^* - 1| = \limsup_{k\in K_0, k\to\infty} \left| \frac{\nabla f(\xi_k^*)^{\mathrm{T}}(\lambda_k^*(s_k + w_k))}{\lambda_k^*\ \nabla f(x_k)^{\mathrm{T}}(s_k + w_k)} - 1 \right| = \limsup_{k\in K_0, k\to\infty} \left| \frac{(\nabla f(\xi_k) - \nabla f(x_k))^{\mathrm{T}}(s_k + w_k)}{\nabla f(x_k)^{\mathrm{T}}(s_k + w_k)} \right| \leqslant \limsup_{k\in K_0, k\to\infty} \frac{\|\nabla f(\xi_k) - \nabla f(x_k)\| (c_1 + \gamma_k p + \gamma_k q/\varepsilon_0)}{(c_2 - \gamma_k p)\varepsilon_0 - \gamma_k q} = 0. \tag{21}$$

其中$\xi_k^* = x_k + \theta_k(x_{k+1}^* - x_k)$，$0 < \theta_k < 1$。由式(21) 知$\rho_k^* = 1$，此与式(20) 矛盾。故 $\lim\limits_{k\to\infty} \|\nabla f(x_k)\| = 0.$ 证毕。

定理2 设$\{x_k\}$是由算法产生的无穷点列,如果存在一个开凸集$D \supseteq \{x_k\}_{k\in N}$,使得$\nabla f(x)$在$D$上一致连续,则或者$\lim\limits_{k\to\infty} f(x_k) = -\infty$或者$\lim\limits_{k\to\infty} \nabla f(x_k) = 0$。

证明 当$k \in I \cap N$充分大时,由引理1,引理3及式(4),(5)知

$$\gamma_k \|s_k + w_k\| \leqslant \gamma_k [c_1 \|\nabla f(x_k))\| + \gamma_k (p\|\nabla f(x_k))\| + q)] \leqslant \gamma_k^2 [\sigma c_1 + \sigma p \gamma_k + q]. \tag{22}$$

由式(22)及中值定理得

$$f(x_{k+1}) - f(x_k) = \gamma_k (\nabla f(x_k + \theta_k \gamma_k (s_k + w_k)) - \nabla f(x_k))^{\mathrm{T}} (s_k + w_k) + \gamma_k \nabla f(x_k)^{\mathrm{T}} (s_k + w_k) \leqslant \gamma_k \|s_k + w_k\| (\|\nabla f(x_k + \theta_k \gamma_k (s_k + w_k)) - \nabla f(x_k)\| + \|\nabla f(x_k)\|) \leqslant \gamma_k^2 (\sigma c_1 + \sigma p \gamma_k + q) \times (\|\nabla f(x_k + \theta_k \gamma_k (s_k + w_k)) - \nabla f(x_k)\| + \sigma \gamma_k). \tag{23}$$

其中$0 < \theta_k < 1$。由式(22),(23)及$\nabla f(x)$的一致连续性知,当k充分大时,存在$\sigma_1 > 0$使得

$$f(x_{k+1}) - f(x_k) \leqslant \sigma_1 \gamma_k^2. \tag{24}$$

当$k \in J \cap N$时,由式(9)易得,当k充分大时$\exists\ \sigma_2 > 0$使得

$$f(x_{k+1}) - f(x_k) \leqslant 0 \leqslant \sigma_2 \gamma_k^2. \tag{25}$$

取$\sigma_3 = \max\{\sigma_1, \sigma_2\}$,由式(24),(25)易得

$$f(x_{k+1}) - f(x_k) \leqslant \sigma_3 \gamma_k^2,\ \forall k \in N. \tag{26}$$

令$T_k = \sigma_3 \sum\limits_{j\geqslant k}^{\infty} \gamma_j^2$,则当$k$充分大时有

$$f(x_{k+1}) + T_{k+1} \leqslant f(x_k) + T_k.$$

由式(26)知,当k充分大时,$\{f(x_k) + T_k\}$单调下降,再由$T_k \to 0 (k\to\infty)$知,或者$\lim\limits_{k\to\infty} f(x_k) = -\infty$,或者$\{f(x_k)\}$存在有限极限,此时由定理1知$\lim\limits_{k\to\infty} \nabla f(x_k) = 0$。证毕。

4 数值试验

选择文献[8]中的几个算例,在PIII.933计算机上对本文中算法进行数值试验。利用Matlab编制程序,并与FR,PR,HS共轭梯度法进行比较。

取$\sum\limits_{k=1}^{\infty} \gamma_k^2 = \sum\limits_{k=1}^{\infty} \frac{1}{k^2}$,$\mu_1 = \mu_2 = 0.25$,$\lambda_k$取为$\{1, \rho, \rho^2, \rho^3, \cdots\}$满足式(9)的最大者,其中$\rho = \frac{1}{2.9}$,$\Delta = 0.067$;用$T_{i1}$表示$\beta_k^{\text{Quasi-Newton}} \in [-\beta_k(\Delta), \beta_k(\Delta)]$的迭代次数,$T_{i2}$表示算法的迭代次数,$T$表示迭代所用时间,$f_{\mathrm{opt}}$表示目标函数的最优值,在精度要求$\|\nabla f(x_k)\| \leqslant 10^{-2}, 10^{-3}$下给出计算结果。

例1 $f(x) = 10(x_1^2 - x_2)^2 + (1 - x_1)^2 + 9(x_4 - x_3^2)^2 + (1 - x_3)^2 + 10.1((x_2 - 1)^2 + (x_4 - 1)^2) + 19.8(x_2 - 1)(x_4 - 1)$.

初时点$x_1 = (-3, -1, -3, -1)^{\mathrm{T}}$,最优点$x_{\mathrm{opt}} = (1,1,1,1)$,$f(x_{\mathrm{opt}}) = 0$。计算结果见表1。

表1 例1的计算结果

Table 1 Calculated results of case 1

算法	T_{i1}	T_{i2}	T/s	$f_{\mathrm{opt}}/10^{-6}$
FR		195,280	0.2700,0.6000	10.670,0.13785
HS		246,385	0.1000,0.4610	46.981,0.49446
PR		83,91	0.3610,0.1100	2.8806,0.19703
MGME		42,65	0.0470,0.0310	20.092,0.34139
QGME	18,25	31, 38	0.0320,0.0460	14.923,0.028131

例2 $f(x) = \sum\limits_{i=1}^{N/2} [(x_{2i} - x_{2i-1}^2)^2 + (1 - x_{2i-1})^2]$.

维数$N = 120$,初时点$x_1 = (-1.2, 1, \cdots, -1.2, 1)^{\mathrm{T}}$,最优点$x_{\mathrm{opt}} = (1, 1, \cdots, 1)$,$f(x_{\mathrm{opt}}) = 0$。计算结果见表2。

表2 例2的计算结果

Table 2 Calculated results of case 2

算法	T_{i1}	T_{i2}	T/s	$f_{\mathrm{opt}}/10^{-6}$
FR		134,202	0.4710,0.3100	142.86,0.31323
HS		52,79	0.1100,0.1410	59.272,0.96394
PR		11,22	0.0500,0.1100	70.032,0.20066
MGME		16,20	0.0470,0.0470	4.2883,0.15343
QGME	5, 10	8,14	0.0470,0.0470	18.375,0.13503

例3 $f(x) = \sum\limits_{i=1}^{N/4} [(x_{4i-1} + 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2x_{4i-1})^2 + 10(x_{4i-3} - x_{4i})^4]$.

维数$N = 60$,初时点$x_1 = (3, -1, 0, 3, \cdots, 3, -1, 0, 3)^{\mathrm{T}}$,最优点$x_{\mathrm{opt}} = (0, 0, \cdots, 0)$,$f(x_{\mathrm{opt}}) = 0$。计算结果见表3。

表3 例3的计算结果

Table 3 Calculated results of case 3

算法	T_{i1}	T_{i2}	T/s	$f_{\mathrm{opt}}/10^{-6}$
FR		47,51	0.1100,0.1200	7.9996,7.7398
HS		73,235	0.1700,0.8210	160.80,18.033
PR		99,157	0.3900,0.4010	316.45,5.7153
MGME		75,268	0.0940,0.3600	281.13,13.121
QGME	88,867	134,940	0.1870,1.2970	150.76,12.537

例4 $f(x) = x^{\mathrm{T}} Q x - x^{\mathrm{T}} b$,其中,

$$Q = \begin{pmatrix} 96.45 & 53.23 & 78.98 & 61.33 \\ 53.23 & 45.93 & 62.14 & 45.11 \\ 78.98 & 62.14 & 89.14 & 62.45 \\ 61.33 & 45.11 & 62.45 & 47.05 \end{pmatrix},$$

$\boldsymbol{b} = (1, 4, 2, 3)^T$.

初时点 $\boldsymbol{x}_1 = (0, 0, 0, 0)^T$, 最优点 $\boldsymbol{x}_{opt} = (0.1304, 0.8245, -0.4068, -0.3886)^T$, $f(\boldsymbol{x}_{opt}) = -0.7245$。计算结果见表4。

表4 例4的计算结果

Table 4 Calculated results of case 4

算法	T_{i1}	T_{i2}	T/s	f_{opt}
FR		148,265	0.1560,0.2660	-0.7245, -0.7245
HS		698,1251	0.6400,1.1400	-0.7244, -0.7245
PR		137,211	0.1400,0.2500	-0.7245, -0.7245
MGME		154,249	0.1560,0.2340	-0.7244, -0.7245
QGME	48,72	57,82	0.0620,0.0780	-0.7245, -0.7245

例5 $f(\boldsymbol{x}) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$.

初时点 $\boldsymbol{x}_1 = (3, -1, 0, 1)^T$, 最优点 $\boldsymbol{x}_{opt} = (0, 0, 0, 0)^T$, $f(\boldsymbol{x}_{opt}) = 0$。计算结果见表5。

表5 例5的计算结果

Table 5 Calculated results of case 5

算法	T_{i1}	T_{i2}	T/s	$f_{opt}/10^{-6}$
FR		145,336	0.1090,0.2500	27.117,1.2039
HS		85,266	0.0780,0.1720	295.83,15.890
PR		28,73	0.0310,0.0470	205.46,10.337
MGME		59,180	0.0470,0.1400	177.47,14.008
QGME	38,42	49,51	0.0630,0.0460	0.19668,0.010713

计算结果表明新算法是有效的,在广义 Armijo 步长搜索下,MGME 和 QGME 算法在迭代次数和 CPU 时间上都比较少,计算精度较高。另外,区间$[-\underline{\beta}_k(\Delta), \bar{\beta}_k(\Delta)]$的范围较大,$\beta_k^{\text{Quasi-Newton}}$大部分落在区间$[-\underline{\beta}_k(\Delta), \bar{\beta}_k(\Delta)]$内。

参考文献:

[1] 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.

[2] BEALE E M L. A derivative of conjugate gradients[C]// LOOTSMA F A. Numerical methods for nonlinear optimization. London: Academic Press, 1972:39-43.

[3] DENG N Y, LI Z F. Global convergence of three terms conjugate gradient methods[J]. Optimization Method and Software, 1995(4):273-282.

[4] DAI Y H, YUAN Y X. Convergence of three terms conjugate gradient methods[J]. Mathematica Numerica Sinica, 1999(3):355-362.

[5] 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.

[6] SUN Qing-ying. Global convergence results for a new three terms memory gradient method with Curry-Altmans step size rules [J]. Soochow Journal of Mathmatics, 2004,30(1):55-66.

[7] BERTSEKAS Dimitri P, TSITSIKLIS John N. Gradient convergence in gradient methods with errors [J]. SIAM J Optim, 2000(3):627-642.

[8] TOUATI-AHMED D, STOREY C. Efficient hybrid conjugate gradient techniques [J]. Journal of Optimization Theory and Applications, 1990,64(2):379-397.

[9] DAI Y H, LIAO L Z. New conjugate conditions and related nonlinear conjugate gradient methods[J]. Applied Mathematics and Optimization, 2001,43(1):87-101.

(编辑 修荣荣)