文章编号：1000－5471(2007)06－0033－04

# 一种新参数下的记忆梯度算法①

公锦凤，　于宪伟

渤海大学 数学系，辽宁 锦州 121000

**摘要：**提出了一种记忆梯度法的主要参数的新形式，分析了该算法在 Wolfe-Powell 搜索下的全局收敛性和线性收敛速度.

**关　键　词：**记忆梯度法；Wolfe-Powell 线性搜索；收敛性

**中图分类号：O224**　　　　**文献标识码：A**

共轭梯度法是求解无约束优化问题$\min\limits_{x\in R^n} f(x)$的一种重要方法. 标准共轭梯度法的迭代格式如下：

$$x_{k+1}=x_k+\alpha_k d_k$$

$$d_k=\begin{cases}-g_k & k=1,\\ -g_k+\beta_k d_{k-1} & k\geqslant 2\end{cases}$$

其中：$\alpha_k$ 是由某种线性搜索获得的步长因子，$d_k$ 是第 $k$ 次迭代时的搜索方向，$g_k$ 表示 $f(x)$ 在 $x_k$ 处的梯度 $\nabla f(x)$，参数 $\beta_k$ 的常见形式有 $\beta_k^{\mathrm{PRP}}$，$\beta_k^{\mathrm{HS}}$，$\beta_k^{\mathrm{DY}}$，$\beta_k^{\mathrm{LS}}$ 等[1]. 但是很多共轭梯度法对一般非线性函数来说不具有全局收敛性，为了保证算法具有全局收敛性，人们提出了记忆梯度算法[2].

记忆梯度算法的每步迭代中不需要计算和存储矩阵，因而适于解决大型优化问题. 步长 $\alpha_k$ 的选择也有很多种，大体可分为精确搜索和非精确搜索规则. 在非精确搜索步长规则中，具有代表性的是 Armijo-Goldstein 准则和 Wolfe-Powell 准则[3,4].

Wolfe-Powell 步长规则为：在第 $k$ 次迭代中选择步长 $\alpha_k$ 满足

$$\begin{cases}f(x_k+\alpha_k d_k)\leqslant f(x_k)+\mu\alpha_k g_k^{\mathrm{T}} d_k\\ g_{k+1}^{\mathrm{T}} d_k\geqslant \sigma g_k^{\mathrm{T}} d_k\end{cases}$$

其中 $\mu\in\left(0,\dfrac{1}{2}\right)\sigma\in(\mu,1)$.

为了改进算法的性能，以便求解大型无约束优化问题，提出一种新参数下的记忆梯度算法，并对 Wolfe-Powell 线性搜索下该算法的全局收敛性和线性收敛速度进行了分析.
为了简便记 $f(x_k)=f_k$，$g(x_k)=g_k$.

## 1　新参数下的记忆梯度算法描述

算法中各参数设置为：$\rho\in(0,1)$，$\mu\in\left(0,\dfrac{1}{2}\right)$，$\sigma\in(\mu,1)$

Step 1　令 $k:=1$，若 $\|g(k)\|=0$，则终止，否则转下一步；

Step 2　令 $x_{k+1}=x_k+\alpha_k d_k(\beta_k)$

①　收稿日期：2007－04－28
基金项目：辽宁省教育厅基金资助项目(2006018).
作者简介：公锦凤(1982－)，女，山东蒙阴人，硕士研究生，主要从事最优化理论与算法研究.

其中：

$$d_k(\beta_k)=\begin{cases}-g_k & k=1\\ -[(1-\beta_k)g_k+\beta_k g_{k-1}] & k\geqslant 2\end{cases},\beta_k\in[0,s_k],\ s_k=\begin{cases}\rho/2 & k=1\\ \rho\|g_k\|^2/[\|g_k\|^2+|g_k^Tg_{k-1}|] & k\geqslant 2\end{cases},$$

$\alpha_k$ 由 Wolfe-Powell 搜索规则确定；

Step 3 $k:=k+1$，转 Step 1.

## 2 算法收敛性研究

为了证明算法的收敛性，需要如下几个引理.

**引理 1** 对 $k\geqslant 1$，$g_k^{\mathrm T}d_k\leqslant-(1-\rho)\|g_k\|^2$.

**证** 若 $k=1$，则 $g_k^{\mathrm T}d_k=g_k^{\mathrm T}\cdot -g_k=-\|g_k\|^2\leqslant-(1-\rho)\|g_k\|^2$；

若 $k\geqslant 2$，则 $-g_k^{\mathrm T}d_k=g_k^{\mathrm T}\cdot[(1-\beta_k)g_k+\beta_k g_{k-1}]\geqslant\|g_k\|^2-\beta_k(\|g_k\|^2-|g_k^{\mathrm T}g_{k-1}|)$，而 $\beta_k\in[0,s_k]\geqslant\|g_k\|^2-\rho\|g_k\|^2=(1-\rho)\|g_k\|^2$. 证毕.

**引理 2** 对 $k\geqslant 2$，$\|d_k\|^2\leqslant\delta_k^2$，其中 $\delta_k=\max\limits_{1\leqslant i\leqslant k}\{\|g_k\|\}$.

**证** 由 $\varphi(x)=\|x\|^2(x\in R^n)$ 为凸函数及 $\beta_k\in[0,1]$ 知 $\|d_k\|^2=\|(1-\beta_k)g_k+\beta_k g_{k-1}\|^2\leqslant(1-\beta_k)\|g_k\|^2+\beta_k\|g_{k-1}\|^2\leqslant\max(\|g_k\|^2,\|g_{k-1}\|^2)=\delta_k^2$.

设定如下假设：

(i) 目标函数 $f(x)$ 在水平集 $L_0=\{x\in R^n\mid f(x)\leqslant f(x_0)\}$ 上有界.

(ii) $f(x)$ 为二次连续可微的凸函数，$g(x)=\nabla f(x)$ 在包含 $L_0$ 的开凸集上 Lipschitz 连续，$f(x)$ 具有唯一极小点 $x^*$，且存在 $0<m\leqslant M$，对下列各式成立：

$$\frac{1}{2}m\|x-x^*\|^2\leqslant f(x)-f(x^*)\leqslant\frac{1}{2}M\|x-x^*\|^2,\ \forall x\in R^n \tag{1}$$

$$m\|x-y\|^2\leqslant(g(x)-g(y))^{\mathrm T}(x-y)\leqslant M\|x-y\|^2,\ \forall x,y\in R^n \tag{2}$$

$$\|g(x)-g(y)\|\leqslant M\|x-y\|,\ \forall x,y\in B;$$

此时 $L=M$. 利用 Cauchy-Schwartz 不等式和 (2) 可得

$$m\|x-x^*\|\leqslant\|g(x)\|\leqslant M\|x-x^*\| \tag{3}$$

**引理 3** 若假设 (ii) 成立，则 $\alpha_k\geqslant\dfrac{-(1-\sigma)g_k^{\mathrm T}d_k}{L\|d_k\|^2}$.

**证** 由 Cauchy-Schwartz 不等式和 Wolfe-Powell 步长规则可得

$$\alpha_k L\|d_k\|^2\geqslant\|g(x_k+\alpha_k d_k)-g_k\|\cdot\|d_k\|\geqslant[g(x_k+\alpha_k d_k)-g_k]^{\mathrm T}\cdot d_k$$

即 $\alpha_k\geqslant\dfrac{-(1-\sigma)g_k^{\mathrm T}d_k}{L\|d_k\|^2}$，证毕.

**定理 1** 若 (i)(ii) 成立，算法产生无穷点列 $\{x_k\}$，则

$$\sum_{k=2}^{\infty}\frac{\|g_k\|^2}{\max\left\{1,\dfrac{\|g_{k-1}\|^2}{\|g_k\|^2}\right\}}<+\infty \tag{4}$$

**证** 由 Wolfe-Powell 搜索规则以及引理 1，2，3 知

$$f_k-f_{k+1}\geqslant-\alpha_k\mu g_k^{\mathrm T}d_k\geqslant\frac{(1-\sigma)g_k^{\mathrm T}d_k}{L\|d_k\|^2}\mu\cdot g_k^{\mathrm T}d_k\geqslant\frac{\mu(1-\sigma)(g_k^{\mathrm T}d_k)^2}{L\cdot\max\{\|g_k\|^2,\|g_{k-1}\|^2\}}$$

$$\geqslant\frac{\mu(1-\sigma)}{L}\cdot\frac{(1-\rho)^2\|g_k\|^4}{\max\{\|g_k\|^2,\|g_{k-1}\|^2\}}\geqslant\frac{\mu(1-\sigma)(1-\rho)^2}{L}\cdot\frac{\|g_k\|^2}{\max\left\{1,\dfrac{\|g_{k-1}\|^2}{\|g_k\|^2}\right\}}$$

因 $\{f_k\}$ 为单调下降数列且有界，则其必有极限，因此 (4) 式成立.

**推论** 由定理 1 可得 $\sum\limits_{k=2}^{\infty}\dfrac{\|g_k\|^4}{\delta_k^2}<+\infty$，其中 $\delta_k=\max\limits_{1\leqslant i\leqslant k}\{\|g_i\|\}$.

**定理 2** 若 (i)(ii) 成立，算法产生无穷点列 $\{x_k\}$，则 $\lim\limits_{k\to\infty}\|g_k\|=0$.

**证**　首先证明$\{\|g_k\|\}$有界.

反证，若$\{\|g_k\|\}$无界，则$\delta_k=\max\limits_{1\leqslant i\leqslant k}\{\|g_i\|\}\to\infty$，$k\to\infty$，因而必存在无穷子集$N$，对$\forall k\in N$满足

$$\|g_k\|=\delta_k \tag{5}$$

且

$$\|g_k\|\to\infty,\ k\in N,\ k\to\infty \tag{6}$$

由(5)式知$\sum\limits_{k\in N}\|g_k\|^2=\sum\limits_{k\in N}\frac{\|g_k\|^4}{\delta_k^2}\leqslant\sum\limits_{k=2}\frac{\|g_k\|^4}{\delta_k^2}<+\infty$，即$\sum\limits_{k\in N}\|g_k\|^2<+\infty$，其中$N$为无穷集. 从而$\|g_k\|\to 0$，$k\in N$，$k\to\infty$. 这与(6)式矛盾，因此$\{\|g_k\|\}$必有界，不妨设为$M'$，即$\delta_k\leqslant M'$. 由定理1的推论可得$\sum\limits_{k=2}^{\infty}\frac{\|g_k\|^4}{\delta_k^2}<+\infty$，其中$\delta_k=\max\limits_{1\leqslant i\leqslant k}\{\|g_i\|\}$，故$\sum\limits_{k=2}^{\infty}\frac{\|g_k\|^4}{M'}\leqslant\sum\limits_{k=2}^{\infty}\frac{\|g_k\|^4}{\delta_k^2}$，所以$\lim\limits_{k\to\infty}\|g_k\|=0$.

证毕.

**定理 3**　若假设(ⅱ)成立，算法产生无穷点列$\{x_k\}$，则$\{x_k\}\to x^*$，其中$x^*$是$f(x)$的唯一极小点，而且

$$R_1=\lim_{k\to\infty}\sup\|x_{k+1}-x^*\|^{\frac{1}{k}}<1.$$

**证**　由定理 1 的证明可知

$$f_k-f_{k+1}\geqslant\frac{\mu(1-\sigma)(1-\rho)^2}{L}\cdot\frac{\|g_k\|^2}{\max\left\{1,\ \dfrac{\|g_{k-1}\|^2}{\|g_k\|^2}\right\}}$$

记$\xi=\dfrac{\mu(1-\sigma)(1-\rho)^2}{L}\cdot\dfrac{\|g_k\|^2}{\max\left\{1,\ \dfrac{\|g_{k-1}\|^2}{\|g_k\|^2}\right\}}$，则$f_k-f_{k+1}\geqslant\xi\|g_k\|^2$.

由式(1)，(3)可得$f_k-f_{k+1}\geqslant\xi\|g_k\|^2\geqslant\xi m^2\|x_k-x^*\|^2\geqslant\dfrac{2\xi m^2(f_k-f^*)}{M}$. 记$\theta=\dfrac{2\xi m^2}{M}$，则

$$f_k-f_{k+1}\geqslant\theta(f_k-f^*) \tag{7}$$

由$\theta=\dfrac{2\xi m^2}{M}=\dfrac{2m^2\mu}{ML\max\left\{1,\ \dfrac{\|g_{k-1}\|^2}{\|g_k\|^2}\right\}}\cdot(1-\sigma)(1-\rho)^2\leqslant(1-\sigma)(1-\rho)^2$可知$\theta<1$.

由(7)式知$f_{k+1}-f^*\leqslant(1-\theta)(f_k-f^*)\leqslant\cdots\leqslant(1-\theta)^k(f_1-f^*)$，再由(1)式可得$\|x_{k+1}-x^*\|^2\leqslant\dfrac{2}{m}(f_{k+1}-f^*)\leqslant(1-\theta)^k\dfrac{2(f_1-f^*)}{m}$，从而

$$\|x_{k+1}-x^*\|\leqslant(1-\theta)^{\frac{k}{2}}\cdot\sqrt{\frac{2(f_1-f^*)}{m}},$$

$$\lim_{k\to\infty}\sup\|x_{k+1}-x^*\|^{\frac{1}{k}}\leqslant\lim_{k\to\infty}(1-\theta)^{\frac{1}{2}}\cdot\sqrt[2k]{\frac{2(f_1-f^*)}{m}}<1$$

即$R_1<1$，从而$\{x_k\}R-$线性收敛于$x^*$.

证毕.

## 3　数值实验

在共轭梯度中，当$\beta_k$分别取

$$\beta_k^{\mathrm{PRP}}=\frac{g_k^{\mathrm{T}}(g_k-g_{k-1})}{\|g_{k-1}\|^2},\ \beta_k^{\mathrm{HS}}=\frac{g_k^{\mathrm{T}}(g_k-g_{k-1})}{d_{k-1}^{\mathrm{T}}g_{k-1}},\ \beta_k^{\mathrm{LS}}=-\frac{g_k^{\mathrm{T}}(g_k-g_{k-1})}{g_{k-1}^{\mathrm{T}}d_{k-1}}$$

时，相应的算法分别被称为 PRP、HS、LS 共轭梯度算法[1]，本文的算法简记为 GM.

**实验 1**

$$f(x)=(x_1+10x_2)^4+5(x_3-x_4)^4+(x_2-2x_3)^4+10(x_1-x_4)^4$$

$$x_0 = (2, 2, -2, -2)^T, x^* = (0, 0, 0, 0)^T, f^* = 0$$

**实验2** 扩展 Powell 函数

$$f(x) = \sum_{i=1}^{n-3}[(x_i + 10x_{i+1})^2 + 5(x_{i+2} - x_{i+3})^2 + (x_{i+1} - 2x_{i+2})^4 + 10(x_i - x_{i+3})^4]$$

$$x_0 = (3, -1, 0, 1, \cdots, 3, -1, 0, 1)^{\mathrm{T}}, x^* = (0, 0, \cdots, 0)^T, f^* = 0, n = 200$$

**实验3** 设置扩展 Powell 函数中参数 $n = 1\ 000$，重新实验.

数值实验结果结果如表1所示.

**表1 各种算法的迭代次数，函数值和梯度值的计算次数比较**

| 实验 | PRP | HS | LS | GM |
|---|---|---|---|---|
| | *NI*/(*NF*+*NG*) | *NI*/(*NF*+*NG*) | *NI*/(*NF*+*NG*) | *NI*/(*NF*+*NG*) |
| 1 | 24/521 | 46/309 | 60/1101 | 22/160 |
| 2 | 136/1251 | 121/1618 | 78/1548 | 68/306 |
| 3 | 145/4765 | 143/2656 | 181/4050 | 84/1678 |

在表1中，*NI*、*NF*、*NG* 分别是迭代次数、函数值的计算次数、梯度值的计算次数，终止条件是精度 $\leqslant 10^{-8}$，计算结果表明本文算法是比较有效的，即当达到同样精度时的迭代次数，函数值和梯度值的计算次数，都比其它算法要少，这在解决大型优化问题方面是具有较大优越性.

**参考文献：**

[1] 戴彧虹，袁亚湘. 非线性共扼梯度法[M]. 上海：上海科学技术出版社，2001.

[2] Miele a, Cantrell J W. Study on a Memory Gradient Method for the Minimization of Functions [J]. J Optim Thory Apple, 1969, 3(6): 459－470.

[3] 袁亚湘，孙文瑜. 最优化理论与方法[M]. 北京：科学出版社，2003.

[4] 谢 政，李建平，汤泽滢. 非线性最优化[M]. 长沙：国防科技大学出版社，2003：9.

# A New Parameter of Memory Gradient Method

GONG Jin-feng, YU Xian-wei

*Department of Mathematics, Bohai University, Liaoning, Jinzhou 121000*

**Abstract:** In this paper, a new formula of main parameter of memory gradient method is proposed, and analyzed the global convergence and convergence rate of this method with Wolfe-Powell line search rule. Numerical results show that the new method is efficient, and it can be used in large scale optimization problems.

**Key words:** memory gradient method; wolfe-powell line search; convergence

责任编辑 张 枸