

无约束优化问题的一个下降方法

董 丽¹, 周金川²

(1. 信阳师范学院数学与信息科学学院, 河南 信阳 464000)

(2. 山东理工大学理学院数学系, 山东 淄博 255049)

摘 要: 本文研究了无约束优化问题. 利用当前和前面迭代点的信息以及曲线搜索技巧产生新的迭代点, 得到了一个新的求解无约束优化问题的下降方法. 在较弱条件下证明了算法具有全局收敛性. 当目标函数为一致凸函数时, 证明了算法具有线性收敛速率. 初步的数值试验表明算法是有效的.

关键词: 无约束优化; 记忆梯度法; 曲线搜索; 收敛性

MR(2010) 主题分类号: 49M37

中图分类号: O221.2

文献标识码: A

文章编号: 0255-7797(2015)01-0173-07

1 引言

考虑无约束优化问题:

$$\min f(x), \quad x \in R^n, \quad (1)$$

其中 $f(x): R^n \rightarrow R$ 为连续可微函数. 求解问题 (1) 的算法主要是迭代法, 其基本格式为

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

其中 d_k 为 $f(x)$ 在 x_k 点的搜索方向, α_k 为 $f(x)$ 沿此方向上的搜索步长.

共轭梯度法是求解无约束优化问题最常用的方法之一^[1,2]. 此类算法有效避免了计算和存储矩阵, 是求解大规模优化问题的有效方法之一, 但很多共轭梯度法不具有全局收敛性^[2]. 记忆梯度法是共轭梯度法的推广和改进, 此类算法充分利用前面迭代点的信息来产生新的迭代点, 不用计算和存储矩阵, 算法简单, 适于求解大规模优化问题, 并且与共轭梯度法相比, 记忆梯度法增加了参数选择的自由度, 更有利于构造稳定的快速收敛算法^[3-6].

另一方面, 文献 [7] 提出一种曲线搜索算法, 这种算法沿一条曲线搜索新的迭代点, 每次迭代时同时确定下降方向和步长, 这样可以使算法更容易找到合适的步长. 利用曲线搜索技巧可以保证算法具有全局收敛性和较快的收敛速度^[7-9].

本文在文献 [3-9] 的基础上给出一个新的求解无约束优化问题的下降方法, 在较弱条件下证明了其全局收敛性和线性收敛速率. 算法利用当前和前面迭代点的信息以及曲线搜索技巧产生下一个迭代点, 并且不需计算和存储矩阵, 适于求解大规模优化问题. 初步的数值试验表明算法是有效的.

2 算法及其性质

*收稿日期: 2012-09-12

接收日期: 2012-11-18

基金项目: 国家自然科学基金项目 (11101248); 山东省自然科学基金 (ZR2010AQ026).

作者简介: 董丽 (1981-), 女, 河南信阳, 讲师, 主要研究方向: 优化理论与算法.

在本文中, 设 $g(x)$ 为 $f(x)$ 的梯度, 并且简记 $g(x_k)$ 为 g_k , $f(x_k)$ 为 f_k , $f(x^*)$ 为 f^* .

本文采用如下曲线搜索准则:

曲线搜索准则 取 $0 < \mu_1 < \frac{1}{2} < \mu_2 < 1$, $s_k > 0$, 要求 α_k 满足

$$f(x_k) - f(x_k + \alpha d_k(\alpha)) \geq -\mu_1 \alpha g_k^T d_k(\alpha), \quad (2)$$

$$g(x_k + \alpha d_k(\alpha))^T d_k(\alpha) \geq \mu_2 g_k^T d_k(\alpha), \quad (3)$$

这里

$$d_k(\alpha) = \begin{cases} -g_k, & k = 1, \\ -\alpha g_k + \alpha s_k g_{k-1}, & k \geq 2. \end{cases} \quad (4)$$

算法 (A) 取 $0 < \mu_1 < \frac{1}{2} < \mu_2 < 1$, $\rho \in (0, 1)$, $x_1 \in R^n$. 令 $k := 1$.

步骤 1 若 $\|g_k\| = 0$, 则停止迭代; 否则, 转步骤 2.

步骤 2 $x_{k+1} = x_k + \alpha_k d_k(\alpha_k)$, 这里 α_k 由曲线搜索准则确定, 而

$$d_k(\alpha_k) = \begin{cases} -g_k, & k = 1, \\ -\alpha_k g_k + \alpha_k s_k g_{k-1}, & k \geq 2, \end{cases} \quad (5)$$

其中

$$s_k = \frac{\rho \|g_k\|}{\|g_{k-1}\|}. \quad (6)$$

步骤 3 令 $k := k + 1$, 转步骤 1.

算法 (A) 有下面两个性质.

引理 2.1 对任意的 $k \geq 2$, 有 $-g_k^T d_k(\alpha_k) \geq \alpha_k (1 - \rho) \|g_k\|^2$.

引理 2.2 对任意的 $k \geq 2$, 有 $\|d_k(\alpha_k)\| \leq \alpha_k (1 + \rho) \|g_k\|$.

3 算法的全局收敛性

为了下面证明的需要, 本文作如下假设:

(H₁) 目标函数 $f(x)$ 在水平集 $L_0 = \{x \in R^n | f(x) \leq f(x_1)\}$ 上有下界.

(H₂) 梯度函数 $g(x)$ 在包含 L_0 的开凸集 B 上一致连续.

引理 3.1 若 (H₁) 和 (H₂) 成立, $\{x_k\}$ 是由算法 (A) 产生的迭代点列, 则 $\{\|d_k(\alpha_k)\|\}$ 和 $\{\|g_k\|\}$ 都有界.

证 由引理 2.2 可知只需证明 $\{\|g_k\|\}$ 有界. 假设 $\{\|g_k\|\}$ 无界, 则存在无穷子列 $N \subset \{1, 2, \dots\}$ 使

$$\|g_k\| \rightarrow +\infty, (k \in N, k \rightarrow +\infty). \quad (7)$$

由 (2) 式和引理 2.1 知

$$f_k - f_{k+1} \geq -\mu_1 \alpha_k g_k^T d_k(\alpha_k) \geq \mu_1 (1 - \rho) \alpha_k^2 \|g_k\|^2. \quad (8)$$

因为 $\{f_k\}$ 单调不增且有下界, 故知 $\{f_k\}$ 有极限, 从而由 (8) 式知

$$\sum_{k \in N} \alpha_k^2 \|g_k\|^2 < \sum_{k=2}^{\infty} \alpha_k^2 \|g_k\|^2 < +\infty. \quad (9)$$

因此

$$\alpha_k^2 \|g_k\|^2 \rightarrow 0 (k \in N, k \rightarrow +\infty). \quad (10)$$

由 (7) 和 (10) 式可知

$$\alpha_k^2 \|g_k\| \rightarrow 0 (k \in N, k \rightarrow +\infty). \quad (11)$$

由引理 2.2 可得 $\alpha_k \|d_k(\alpha_k)\| \leq \alpha_k^2 (1 + \rho) \|g_k\|$, 从而由 (11) 式可得

$$\alpha_k \|d_k(\alpha_k)\| \rightarrow 0 (k \in N, k \rightarrow +\infty). \quad (12)$$

由 (3) 式, 引理 2.1 及引理 2.2 可得

$$\begin{aligned} \|g(x_{k+1}) - g(x_k)\| &\geq \frac{(g_{k+1} - g_k)^T d_k(\alpha_k)}{\|d_k(\alpha_k)\|} \\ &\geq \frac{(\mu_2 - 1) g_k^T d_k(\alpha_k)}{\|d_k(\alpha_k)\|} \\ &\geq \frac{(1 - \mu_2)(1 - \rho) \alpha_k \|g_k\|^2}{\|d_k(\alpha_k)\|} \\ &\geq \frac{(1 - \mu_2)(1 - \rho) \|g_k\|}{1 + \rho}, \end{aligned}$$

从而由 (H_2) 和 (12) 式可得 $\|g_k\| \rightarrow 0 (k \in N, k \rightarrow +\infty)$, 这与 (7) 式矛盾.

定理 3.1 假设 (H_1) 和 (H_2) 成立, $\{x_k\}$ 是由算法 (A) 产生的迭代点列, 则 $\lim_{k \rightarrow \infty} \|g_k\| = 0$.

证 假设 $\lim_{k \rightarrow \infty} \|g_k\| \neq 0$, 则存在无穷子列 $K \subset \{1, 2, \dots\}$ 及 $\vartheta > 0$ 有

$$\|g_k\| > \epsilon, \forall k \in K. \quad (13)$$

由 (9) 和 (13) 式知

$$\sum_{k \in K} \alpha_k^2 \epsilon^2 < \sum_{k \in K} \alpha_k^2 \|g_k\|^2 < \sum_{k=2}^{\infty} \alpha_k^2 \|g_k\|^2 < +\infty, \quad (14)$$

从而知 $\alpha_k \rightarrow 0 (k \in K, k \rightarrow +\infty)$. 又因为 $\{\|d_k(\alpha_k)\|\}$ 有界, 故知

$$\alpha_k \|d_k(\alpha_k)\| \rightarrow 0, (k \in K, k \rightarrow +\infty). \quad (15)$$

由引理 3.1 的证明过程可知

$$\|g(x_{k+1}) - g(x_k)\| \geq \frac{(1 - \mu_2)(1 - \rho) \|g_k\|}{1 + \rho}, \quad (16)$$

从而由 (H_2) , (15) 和 (16) 式知

$$\|g_k\| \rightarrow 0, (k \in K, k \rightarrow +\infty), \quad (17)$$

这与 (13) 矛盾. 因此 $\lim_{k \rightarrow \infty} \|g_k\| = 0$.

4 算法的线性收敛速率

假设 (H₃): $f(x)$ 是二次连续可微的一致凸函数.

引理 4.1 ^[3] 若 (H₃) 成立, 则 (H₁) 和 (H₂) 成立, $f(x)$ 有唯一的极小点 x^* , 且存在 $M \geq m > 0$, 使

$$\frac{m}{2} \|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{M}{2} \|x - x^*\|^2,$$

$$m \|x - x^*\| \leq \|g(x)\| \leq M \|x - x^*\|.$$

此外, $g(x)$ 在水平集 L_0 上 Lipschitz 连续, 即存在常数 $L > 0$, 使

$$\|g(x) - g(y)\| \leq L \|x - y\|, \quad \forall x, y \in L_0. \quad (18)$$

定理 4.1 若 (H₃) 成立, $\{x_k\}$ 是由算法 (A) 产生的无穷点列, 则 $\{x_k\}$ 至少线性收敛于 x^* .

证 由引理 3.1 的证明过程可知

$$f_k - f_{k+1} \geq -\mu_1 \alpha_k g_k^T d_k(\alpha) \geq \mu_1 (1 - \rho) \alpha_k^2 \|g_k\|^2, \quad (19)$$

并且

$$\|g(x_{k+1}) - g(x_k)\| \geq \frac{(1 - \mu_2)(1 - \rho) \|g_k\|}{1 + \rho}. \quad (20)$$

由 (18) 和 (20) 式可知

$$\alpha_k L \|d_k(\alpha_k)\| \geq \|g(x_k + \alpha_k d_k(\alpha_k)) - g_k\| \geq \frac{(1 - \mu_2)(1 - \rho) \|g_k\|}{1 + \rho},$$

从而由引理 2.2 可得

$$\alpha_k^2 \geq \frac{(1 - \mu_2)(1 - \rho)}{L(1 + \rho)} \frac{\|g_k\| \alpha_k}{\|d_k(\alpha_k)\|} \geq \frac{(1 - \mu_2)(1 - \rho)}{(L + 1)(1 + \rho)^2}. \quad (21)$$

由 (19) 和 (21) 式可得

$$f_k - f_{k+1} \geq \frac{\mu_1 (1 - \mu_2)(1 - \rho)^2}{(L + 1)(1 + \rho)^2} \|g_k\|^2. \quad (22)$$

令 $\lambda = \frac{\mu_1 (1 - \mu_2)(1 - \rho)^2}{(L + 1)(1 + \rho)^2}$, 则 $0 < \lambda < \frac{1}{2}$, 并且有

$$f_k - f_{k+1} \geq \lambda \|g_k\|^2. \quad (23)$$

定理余下的证明类似于文献 [3] 中的定理 3, 在此省略.

5 曲线搜索准则的修正

算法 (A) 采用的是一种类似于 Wolfe 线性搜索的曲线搜索准则. 我们也可以采用下面类似于 Armijo 线性搜索的曲线搜索准则.

新的曲线搜索准则 选取 α_k 为 $\{1, \beta^2, \beta^3, \dots\}$ 中满足下式的最大者

$$f(x_k) - f(x_k + \alpha d_k(\alpha)) \geq -\sigma \alpha g_k^T d_k(\alpha),$$

这里 $\beta \in (0, 1)$, $\sigma \in (0, \frac{1}{2})$, 而 $d_k(\alpha)$ 由 (4) 式给出. 类似于文献 [5] 中的定理 3.1 和定理 4.1 的证明过程, 我们可以得到下面两个结论.

定理 5.1 假设 (H_1) 和 (H_2) 成立且算法 (A) 采用新的曲线搜索准则, 则算法 (A) 或有限步终止于问题 (1) 的稳定点; 或产生无穷点列 $\{x_k\}$, 其任意聚点都是问题 (1) 的稳定点.

定理 5.2 假设 (H_3) 成立且算法 (A) 采用新的曲线搜索准则, 如果 $\{x_k\}$ 是由算法 (A) 产生的无穷点列, 则 $\{x_k\}$ 至少线性收敛于 x^* .

此外, 算法 (A) 只利用了当前和前面一步迭代点的信息, 这在算法的设计中似乎是一种信息浪费. 因此, 我们可以利用前面多步迭代点的信息来产生新的迭代点, 即采用下多步曲线搜索准则.

多步曲线搜索准则 取 $0 < \mu_1 < \frac{1}{2} < \mu_2 < 1$, $m \geq 3$ 是一个正整数, 要求 α_k 满足

$$f(x_k) - f(x_k + \alpha d_k(\alpha)) \geq -\mu_1 \alpha g_k^T d_k(\alpha),$$

$$g(x_k + \alpha d_k(\alpha))^T d_k(\alpha) \geq \mu_2 g_k^T d_k(\alpha),$$

这里

$$d_k(\alpha) = \begin{cases} -g_k, & k \leq m-1, \\ -\alpha g_k + \alpha s_k \sum_{i=2}^m g_{k-i+1}, & k \geq m, \end{cases}$$

其中

$$s_k = \begin{cases} 1, & k \leq m-1, \\ \frac{\rho \|g_k\|}{\sum_{i=2}^m \|g_{k-i+1}\|}, & k \geq m. \end{cases}$$

类似于定理 3.1 和定理 4.1 的证明过程, 我们可以证明当算法 (A) 采用多步曲线搜索准则时仍具有全局收敛性和线性收敛速率.

6 数值试验

为检验算法 (A) 的实算效果, 我们选取几个算例对本文算法进行数值试验, 并与共轭梯度法和最速下降法进行比较.

例 1 $f(x) = (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^4$, $x_0 = (-3, 1, 2)^T$, $x^* = (1, 1, 1)^T$, $f^* = 0$.

例 2 $f(x) = (1 - x_1)^2 + (1 - x_{10})^2 + \sum_{i=1}^9 (x_i^2 - x_{i+1})^2$, $x_0 = (3, 2, \dots, 3, 2)^T$, $x^* = (1, \dots, 1)^T$, $f^* = 0$.

例 3 $f(x) = (x_1 + 10x_2)^4 + 5(x_3 - x_4)^4 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$, $x_0 = (-2, 4, -2, 4)^T$, $x^* = (0, 0, 0, 0)^T$, $f^* = 0$.

例 4 $f(x) = (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_3 - 1)^2 + (x_4 - 1)^4 + (x_5 - 1)^6$, $x_0 = (0, 1, 0, 1, 0)^T$, $x^* = (1, 1, 1, 1, 1)^T$, $f^* = 0$.

例 5 $f(x) = (x_1 - x_2)^2 + (x_2 + x_3 - 2)^2 + (x_4 - 1)^2 + (x_5 - 1)^2$, $x_0 = (-3, 3, -3, 3, -3)^T$, $x^* = (1, 1, 1, 1, 1)^T$, $f^* = 0$.

我们用 EX 表示算例, NA 表示算法 (A), 用 FR, PRP 和 LS 分别表示 FR, PRP 和 LS 共轭梯度法 [2], SM 表示最速下降法. 在试验中, 共轭梯度法及最速下降法均采用如下 Wolfe 线性搜索准则.

Wolfe 线性搜索准则 取 $0 < \mu_1 < \frac{1}{2} < \mu_2 < 1$, 计算 α_k 使其满足

$$f(x_k) - f(x_k + \alpha d_k) \geq -\mu_1 \alpha g_k^T d_k, \quad g(x_k + \alpha d_k)^T d_k \geq \mu_2 g_k^T d_k.$$

参数取值为 $\mu_1 = 0.15, \mu_2 = 0.88, \rho = 0.25$. 精度设为 $|f_k - f^*| \leq \text{eps}$. 表格 1 中的数字为达到相应精度时算法的迭代次数, 表格 2 中的数字为算法迭代所需的时间 (单位: 秒). 计算结果如下.

表 1: 算法迭代次数的比较

EX	NA	FR	PRP	LS	SM	eps
1	18	26	21	16	41	$\leq 10^{-6}$
2	25	71	36	31	42	$\leq 10^{-6}$
3	18	19	25	21	26	$\leq 10^{-7}$
4	22	65	26	57	372	$\leq 10^{-7}$
5	8	19	11	13	14	$\leq 10^{-8}$

表 2: 算法迭代所需时间的比较

EX	NA	FR	PRP	LS	SM	eps
1	0.012s	0.025s	0.027s	0.021s	0.039s	$\leq 10^{-6}$
2	0.025s	0.073s	0.051s	0.043s	0.038s	$\leq 10^{-6}$
3	0.013s	0.037s	0.027s	0.036s	0.024s	$\leq 10^{-7}$
4	0.012s	0.053s	0.032s	0.059s	0.257s	$\leq 10^{-7}$
5	0.003s	0.019s	0.014s	0.019s	0.016s	$\leq 10^{-8}$

从以上数值试验及比较的结果可以看出, 当达到同一精度时, 本文提出的新算法所需的迭代次数及时间都比 FR, PRP 和 LS 共轭梯度法及最速下降法少, 表明新算法是有效的.

参 考 文 献

- [1] 袁亚湘, 孙文瑜. 最优化理论与方法 [M]. 北京: 科学出版社, 1997.
- [2] 戴戡虹, 袁亚湘. 非线性共轭梯度法 [M]. 上海: 上海科学技术出版社, 2000.
- [3] 汤京永, 时贞军. 一类全局收敛的记忆梯度法及其线性收敛性 [J]. 数学进展, 2007, 36 (1): 67-75.
- [4] 汤京永, 贺国平, 董丽. 一类新的多步曲线搜索下的超记忆梯度法 [J]. 应用数学学报, 2011, 34 (2): 353-362.
- [5] 汤京永, 贺国平, 董丽. 一类新的求解无约束优化问题的记忆梯度法 [J]. 数学杂志, 2011, 31 (2): 362-368.
- [6] Yu Zhensheng. Global convergence of a memory gradient method without line search[J]. Journal of Applied Mathematics and Computing, 2008, 26: 545-553.
- [7] Shi Zhenjun, Shen Jie. A new descent algorithm with curve search rule[J]. Applied Mathematics and Computation, 2005, 161: 753-768.
- [8] Shi Zhenjun. Convergence of multi-step curve search method for unconstrained optimization[J]. Journal of Numerical Mathematics, 2004, 12: 297-309.
- [9] 汤京永, 董丽, 李学志. 一类新的曲线搜索下的多步下降算法 [J]. 应用数学, 2009, 22 (4): 815-820.

A DESCENT METHOD FOR UNCONSTRAINED OPTIMIZATION PROBLEMS

DONG Li¹, ZHOU Jin-chuan²

(1.College of Mathematics and Information Science, Xinyang Normal University,
Xinyang 464000, China)

(2.Department of Mathematics, School of Science, Shandong University of Technology,
Zibo 255049, China)

Abstract: This paper studies the unconstrained optimization problem. By using the current and previous iterative information and the curve search rule to generate a new iterative point, a new descent algorithm is proposed for solving the unconstrained optimization problem. We prove its global convergence under some mild conditions. The linear convergence rate is also proved when the objective function is uniformly convex. Numerical results show that the new method is efficient in practical computation.

Keywords: unconstrained optimization; memory gradient method; curve search; convergence

2010 MR Subject Classification: 49M37