

一种新的MBFGS算法

郑轶天

长沙理工大学数学与统计学院, 湖南 长沙

收稿日期: 2022年7月24日; 录用日期: 2022年8月17日; 发布日期: 2022年8月25日

摘 要

本文在MBFGS算法的研究基础上, 对该算法进行推广, 形成了一种新的MBFGS算法。同时, 研究了其全局收敛性并给出相关证明。一些数值实验表明, 这种改进的MBFGS方法对于给定的测试问题是有效的。对于一些无约束优化问题, 新的MBFGS算法平均比单调或传统的非单调格式的BFGS方法使用更少的函数和梯度求值。

关键词

非线性搜索, 拟牛顿法, 全局收敛性

A New MBFGS Algorithm

Yitian Zheng

College of Mathematics and Computational Science, Changsha University and Technology, Changsha Hunan

Received: Jul. 24th, 2022; accepted: Aug. 17th, 2022; published: Aug. 25th, 2022

Abstract

Based on the research of MBFGS algorithm, this algorithm is extended to form a new MBFGS algorithm. At the same time, its convergence is studied and the relevant proof is given. Some numerical experiments show that the improved MBFGS method is effective for a given test problem. For some unconstrained optimization problems, the new MBFGS algorithm uses less function and gradient evaluation than the mono-tone or traditional non monotone BFGS method.

Keywords

Nonlinear Search, Quasi Newton Method, Global Convergence



1. 引言

在本文中，我们考虑以下无约束优化问题：

$\min f(x), x \in R^n$ ，其中 $f: R^n \rightarrow R$ 是一个连续可微函数，其梯度用 g 表示。单调线搜索方法来解决此问题要求每一次的迭代，目标函数值都要有下降，而非单调线搜索就不需要目标函数值连续下降，非单调线搜索最早是由 Grippo [1] 于 1986 年提出。其具体格式如下：

给定常数 $a > 0$ ， $\delta, \rho \in (0, 1)$ 非负整数 M ，设定步长 $\alpha_k = \max\{a\rho^0, a\rho^1, \dots\}$ 满足下列不等式：

$$f(x_k + \rho^m a d_k) \leq \max_{0 \leq j \leq M} f(x_{k-j}) + \delta \rho^m a g_k^T d_k$$

当 $M = 0$ 时此不等式即变为标准的 Armijo 线搜索规则，非单调线搜索的一个优点是步长可以尽可能松散地选择，大量的数值实验表明，这种非单调线搜索技术是非常有效的，但也存在一些缺点，首先，由于上述不等式中的最大值函数，在任何迭代中生成的好的函数值基本上被丢弃。其次，在某些情况下，该方法数值性能非常依赖于 M 的选择，MBFGS [2] 方法中采用的线搜索就是由 Grippo [3] 提出的方法，因此我们想通过更改 MBFGS 方法中的线搜索规则 [4] 来获得一种新的 MBFGS 方法，从而获得更好的数值性能。

本文其余部分组织如下：第二部分详细介绍了新的 MBFGS 方法；第三部分证明该算法的全局收敛性；第四部分给出数值实验结果。

2. 新的 MBFGS 算法

1. 给定 $x_0 \in R^n$ ， $B_0 = I$ ， $\delta_1 \in (0, 1)$ ， $\delta_2 > 0$ ， $0 \leq \eta_{\min} \leq \eta_{\max} < 1 < \rho$ ， $C_0 := f(x_0)$ ， $Q_0 := 1$ ， $k := 0$
 $0 \leq \eta_{\min} \leq \eta_{\max} \leq 1 < \rho, \eta_k \in [\eta_{\min}, \eta_{\max}], \mu > 0$

2) 由 $B_k d + g_k = 0$ ，计算下降方向 d_k ，如果 $f(x_k + d_k) \leq f(x_k) + \delta_1 g_k^T d_k$ ，取步长 $\alpha_k = 1$ ，否则去步 3。

3) 计算步长 $\alpha_k = \bar{\alpha} \rho^{h_k}$ ，其中 h_k 是使得下列不等式成立的最大整数：

$$f(x_k + \alpha_k d_k) \leq C_k + \delta_1 \alpha_k g_k^T d_k - \delta_2 \|\alpha_k d_k\|^2, \tag{1.1}$$

其中 $C_{k+1} = \frac{\eta_k Q_k C_k + f(x_{k+1})}{Q_{k+1}}$ ， $Q_{k+1} = \eta_k Q_k + 1$ ，

4) 计算 $x_{k+1} = x_k + \alpha_k d_k$

5) 迭代更新： $s_k = x_{k+1} - x_k$ ， $y_k = g_{k+1} - g_k$ ，

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{z_k z_k^T}{z_k^T s_k}, \quad z_k = y_k + C \|g(x_k)\|^r s_k + \max\left\{0, -\frac{y_k^T s_k}{\|s_k\|^2}\right\} s_k$$

其中 $C, r > 0$ 为给定常数。

3. 全局收敛性

在这一部分，为了建立起全局收敛性，需要先做以下几个假设 [5]：

1) 水平集 $\Omega = \{x \in R^n \mid f(x) \leq f(x_0)\}$ 是有界的。

2) 在 Ω 的某些邻域 N 里, f 是连续可微的且其导数 Lipschitz 连续[6], 也就是说存在一个常数 L 大于 0 使得 $\|g(x) - g(y)\| \leq L\|x - y\|, \forall x, y \in N$, 也就是说由本算法产生的序列 $\{x_k\}$ 是包含于 Ω 内的, 此外, 由假设得到, 存在常数 γ_1 使得 $\|g(x)\| \leq \gamma_1, \forall x \in \Omega$ 。

3) 假定存在正的常数 c_1, c_2 使得 $g_k^T d_k \leq -c_1 \|g_k\|^2, \|d_k\| \leq c_2 \|g_k\|$ (2.1)

引理 1: 假定假设 2, 3 成立, 如果 α_k 是满足(1.1)的步长, 那么存在一个常数 c 大于 0 使得对每个 k , 都有下列不等式成立:

$$\alpha_k \geq \min \left\{ \frac{\mu}{\rho}, c \frac{|g_k^T d_k|}{\|d_k\|^2} \right\} \quad (2.2)$$

证明: 我们将从两个方面来证明。

情况一: $\rho\alpha_k \geq \mu$, 那么就有 $\alpha_k \geq \frac{\mu}{\rho}$ 则(2.2)成立

情况二: $\rho\alpha_k < \mu$ 在此种情况下, 很明显对于获得的步长 $\rho\alpha_k$ 不满足(1.1), 因此, 则有:

$$f(x_k + \rho\alpha_k d_k) > C_k + \delta_1 \rho\alpha_k g_k^T d_k - \delta_2 \|\rho\alpha_k d_k\|^2 \geq f(x_k) + \delta_1 \rho\alpha_k g_k^T d_k - \delta_2 \|\rho\alpha_k d_k\|^2 \quad (2.3)$$

又由于中值定理, 存在 $\theta_k \in (0,1)$ 使得:

$$\begin{aligned} f(x_k + \rho\alpha_k d_k) - f(x_k) &= \rho\alpha_k g(x_k + \theta_k \rho\alpha_k d_k)^T d_k \\ &= \rho\alpha_k g_k^T d_k + \rho\alpha_k (g(x_k + \theta_k \rho\alpha_k d_k) - g_k)^T d_k \\ &\leq \rho\alpha_k g_k^T d_k + L\rho^2 \alpha_k^2 \|d_k\|^2 \end{aligned} \quad (2.4)$$

联立(2.3), (2.4), 则有:

$$\begin{aligned} \rho\alpha_k g_k^T d_k + L\rho^2 \alpha_k^2 \|d_k\|^2 &\geq \delta_1 \rho\alpha_k g_k^T d_k - \delta_2 \rho^2 \|\alpha_k d_k\|^2 \\ (L + \delta_2) \rho\alpha_k \|d_k\|^2 &\geq (\delta_1 - 1) g_k^T d_k \\ \alpha_k &\geq \left(\frac{\delta_1 - 1}{(L + \delta_2) \rho} \right) \frac{g_k^T d_k}{\|d_k\|^2} \end{aligned} \quad (2.5)$$

其中 $c = \frac{\delta_1 - 1}{(L + \delta_2) \rho}$ 。

引理 2:

$\liminf_{k \rightarrow \infty} \|g_k\| = 0$, 如果 $\eta_{\max} < 1$, 那么 $\lim_{k \rightarrow \infty} g_k = 0$ 。

证明: 首先需证明:

$$f_{k+1} \leq C_k - \beta \|g_k^2\|, \quad (2.5.2)$$

情况 1: $\alpha_k \geq \frac{\mu}{\rho}$

由(2.1)有:

$$\begin{aligned} f_{k+1} &\leq C_k + \delta_1 \alpha_k g_k^T d_k - \delta_2 \|\alpha_k d_k\|^2 \leq C_k - \delta_1 \alpha_k c_1 \|g_k\|^2 - \delta_2 c_2^2 \|\alpha_k g_k\|^2 \\ &\leq C_k - \frac{\delta_1 \mu c_1 \|g_k\|^2}{\rho} - \frac{\delta_2 c_2^2 \mu^2 \|g_k\|^2}{\rho} = C_k - \frac{\mu (\delta_1 c_1 \rho + \delta_2 c_2^2 \mu)}{\rho^2} \|g_k\|^2 \end{aligned}$$

于是得证。

情况 2:

$$\rho\alpha_k < \mu \text{ 则由(2.5)有 } \alpha_k \geq \left(\frac{\delta_1 - 1}{L + \delta_2} \right) \frac{\mathbf{g}_k^T d_k}{\|d_k\|^2}。$$

则由(2.1)和(1.1)

$$f_{k+1} \leq C_k + \delta_1 \alpha_k \mathbf{g}_k^T d_k - \delta_2 \|\alpha_k d_k\|^2 \text{ 有:}$$

$$\begin{aligned} f_{k+1} &\leq C_k + \frac{\delta_1(\delta_1 - 1)}{(L + \delta_2)\rho} \left(\frac{\mathbf{g}_k^T d_k}{\|d_k\|} \right)^2 - \delta_2 \left(\frac{\delta_1 - 1}{(L + \delta_2)\rho} \right)^2 \left(\frac{\|\mathbf{g}_k^T d_k\|}{\|d_k\|} \right)^2 \\ &\leq C_k + \frac{\delta_1(\delta_1 - 1)}{(L + \delta_2)\rho} \frac{c_1^2}{c_2^2} \|\mathbf{g}_k\|^2 - \delta_2 \left(\frac{\delta_1 - 1}{(L + \delta_2)\rho} \right)^2 \frac{c_1^2}{c_2^2} \|\mathbf{g}_k\|^2 \\ &= C_k + \left[\frac{\delta_1(\delta_1 - 1)}{(L + \delta_2)\rho} - \delta_2 \left(\frac{\delta_1 - 1}{(L + \delta_2)\rho} \right)^2 \right] \frac{c_1^2}{c_2^2} \|\mathbf{g}_k\|^2 \end{aligned}$$

于是(2.5.2)得证。

由(1.2):

$$C_{k+1} = \frac{\eta_k Q_k C_k + f(x_{k+1})}{Q_{k+1}}, \quad Q_{k+1} = \eta_k Q_k + 1$$

又 $f_{k+1} \leq C_k - \beta \|\mathbf{g}_k\|^2$ 有:

$$C_{k+1} = \frac{\eta_k Q_k C_k + f(x_{k+1})}{Q_{k+1}} \leq \frac{\eta_k Q_k C_k + C_k - \beta \|\mathbf{g}_k\|^2}{Q_{k+1}} = C_k - \frac{\beta \|\mathbf{g}_k\|^2}{Q_{k+1}} \tag{2.6}$$

由于 f 有下界且由[2]中引理(1.1)式,

$f_k \leq C_k \leq A_k$, 我们可以得出 C_k 有下界的结论,

由(2.6)可得 $\sum_{k=0}^{\infty} \frac{\|\mathbf{g}_k\|^2}{Q_{k+1}} < \infty$ 。如果 $\|\mathbf{g}_k\|$ 在 0 附近则将违背此式,

因为由[7]中(2.14),

$$Q_{k+1} \leq k + 2, \text{ 于是 } \liminf_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0 \text{ 成立。}$$

如果 $\eta_{\max} < 1$, 那么由[8]中(1.8):

$$Q_{k+1} = 1 + \sum_{j=0}^k \prod_{i=0}^j \eta_{k-i} \leq 1 + \sum_{j=0}^k \eta_{\max}^{j+1} \leq \sum_{j=0}^k \eta_{\max}^j = \frac{1}{1 - \eta_{\max}}$$

因此 $\lim_{k \rightarrow \infty} \mathbf{g}_k = 0$ 成立

于是得证。

4. 数值实验

实验环境:

硬件: R5 4600CPU、1650ti 图形卡;

软件: VMware、Ubuntu 系统、Matlab2016 软件;

所有代码通过 Matlab 编程[9], 本文算法以 NMBFGS 为代号。

相关参数取值如下:

$$\delta_1 = \delta_2 = 0.1, \quad \rho = 0.4, \quad C = 10^{-6}, \quad r = 2$$

通过选取一些问题[8]进行求解, 和标准的 BFGS [10]方法、MBFGS 方法进行对比, 得出一些数值结果如下:

Problem	n	NMBFGS (iter/fn/gn)	MBFGS (iter/fn/gn)	BFGS (iter/fn/gn)
Rose	2	15/30/16	18/33/18	24/50/25
Waston	2	9/19/10	9/19/10	9/19/10
Wood	4	53/122/54	66/120/67	59/174/67

通过以上实验数据可以看出, 相同的无约束优化问题在使用我们的新算法进行计算时, 取得了较好的数值结果, 和 BFGS [10]以及 MBFGS [11]方法相比, 明显减少了迭代次数, 对于计算函数值次数和求导次数也有明显减少。

5. 结论

在适当条件下我们证明了这个新的算法对于一些非凸函数具有全局收敛性一些得出的数值结果也表明, 相较于传统的单调线搜索, 我们这种新的非单调线搜索在数值性能上具有更好的表现。

参考文献

- [1] Grippo, L., Lampariello, F. and Lucidi, S. (1986) A Nonmonotone Line Search Technique for Newton's Method. *SIAM Journal on Numerical Analysis*, **23**, 707-716. <https://doi.org/10.1137/0723046>
- [2] Zhang, H. and Hager, W.W. (2004) A Nonmonotone Line Search Technique and Its Application to Unconstrained Optimization. *SIAM journal on Optimization*, **14**, 1043-1056. <https://doi.org/10.1137/S1052623403428208>
- [3] Yu, Z. and Pu, D. (2008) A New Nonmonotone Line Search Technique for Unconstrained Optimization. *Journal of Computational and Applied Mathematics*, **219**, 134-144. <https://doi.org/10.1016/j.cam.2007.07.008>
- [4] Zhou, W. and Zhang, L. (2009) Global Convergence of the Nonmonotone MBFGS Method for Nonconvex Unconstrained Minimization. *Journal of Computational and Applied Mathematics*, **223**, 40-47. <https://doi.org/10.1016/j.cam.2007.12.011>
- [5] Moré, J.J., Garbow, B.S. and Hillstom, K.E. (1981) Testing Unconstrained Optimization Software. *ACM Transactions on Mathematical Software*, **7**, 17-41. <https://doi.org/10.1145/355934.355936>
- [6] Byrd, R.H. and Nocedal, J. (1989) A Tool for the Analysis of Quasi-Newton Methods with Application to Unconstrained Minimization. *SIAM Journal on Numerical Analysis*, **26**, 727-739. <https://doi.org/10.1137/0726042>
- [7] Dai, Y.H. (2002) On the Nonmonotone Line Search. *Journal of Optimization Theory and Applications*, **112**, 315-330. <https://doi.org/10.1023/A:1013653923062>
- [8] Li, D. and Fukushima, M. (1999) A Globally and Superlinearly Convergent Gauss—Newton-Based BFGS Method for Symmetric Nonlinear Equations. *SIAM Journal on numerical Analysis*, **37**, 152-172. <https://doi.org/10.1137/S0036142998335704>
- [9] Dai, Y.H. (2002) Convergence Properties of the BFGS Algorithm. *SIAM Journal on Optimization*, **13**, 693-701. <https://doi.org/10.1137/S1052623401383455>
- [10] Dai, Y.H. (2002) A Nonmonotone Conjugate Gradient Algorithm for Unconstrained Optimization. *Journal of System Science and Complexity*, **15**, 139-145.
- [11] Grippo, L., Lampariello, F. and Lucidi, S. (1989) A Truncated Newton Method with Nonmonotone Line Search for Unconstrained Optimization. *Journal of Optimization Theory and Applications*, **60**, 401-419. <https://doi.org/10.1007/BF00940345>