Advances in Applied Mathematics 应用数学进展, **2019, 8(2), 258-264**
Published Online February 2019 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2019.82030

A New Perturbed BFGS Method for Unconstrained Optimization Problems

Fei Chen

Department of Mathematics and Statistics, Changsha University of Science and Technology, Changsha Hunan
Email: 2431864206@qq.com

Received: Jan. 27th, 2019; accepted: Feb. 11th, 2019; published: Feb. 19th, 2019

Abstract

Based on the idea of [1] and the BFGS method in [2], this paper presents a new perturbed BFGS method for the unconstrained optimization. We prove that the proposed method has global convergence for nonconvex optimization problems. Numerical results show that this method is efficient.

Keywords

Perturbation, BFGS Method, Global Convergence

求解无约束问题的一种新的扰动BFGS方法

陈　飞

长沙理工大学数学与统计学院，湖南　长沙
Email: 2431864206@qq.com

收稿日期：2019年1月27日；录用日期：2019年2月11日；发布日期：2019年2月19日

摘　要

基于文献[1]的扰动思想和文献[2]中的BFGS型方法，本文提出了一种新的扰动BFGS方法并证明了其在Wolfe搜索下求解非凸优化问题具有全局收敛性。数值结果表明该方法比较有效。

关键词

扰动，BFGS方法，全局收敛

文章引用: 陈飞. 求解无约束问题的一种新的扰动 BFGS 方法[J]. 应用数学进展, 2019, 8(2): 258-264.
DOI: 10.12677/aam.2019.82030

1. 引言

本文考虑如下的无约束问题：

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

其中 $f: R \to R^n$ 是连续可微函数。

BFGS 方法是求解无约束优化问题的一种非常有效的拟牛顿方法，BFGS 方法在 Wolfe 搜索下对求解凸优化问题具有全局收敛性，但对非凸问题不一定收敛。在文献[1]中刘陶文提出了一种扰动的 BFGS 方法，其中对标准的 BFGS 方法的迭代矩阵进行了扰动，该方法证明了在 Wolfe 搜索下对求解非凸问题具有全局收敛性。Zhang，Deng 和 Chen 在文献[2]中基于新的拟牛顿矩阵提出了一种新的拟牛顿算法，该方法满足更好的割线方程。本文的目的就是基于文献[1]的扰动思想，同时结合文献[2]中的 BFGS 迭代公式，提出了一种新的扰动的 BFGS 方法。

本文安排如下，在第二节中我们提出新的扰动 BFGS 方法的算法，在第三节中我们分析了该算法的全局收敛性，第四节中我们进行了数值实验，验证了该算法的有效性，在第五节中我们对论文进行了总结。

2. 算法

求解问题(1)新的扰动 BFGS 方法基本的迭代格式为

$$x_{k+1} = x_k + \alpha_k d_k, k \ge 1, \tag{2}$$

其中 x_k 是当前的近似解，d_k 为搜索方向，α_k 是步长。

基于文献[1]的扰动思想，通过求解下列线性方程组得到搜索方向 d_k

$$(B_k + \mu_k Q) d_k + \nabla f(x_k) = 0, \tag{3}$$

其中 $\nabla f(x)$ 表示其梯度，B_k 为拟牛顿矩阵，也就是 $f(x)$ 在点 x 处的 Hessian 矩阵的近似，并且它要求是正定的。Q 是某个给定的对称正定矩阵，$u_k > 0$ 且它的值依赖于 k，依照某种规则取值为 ε_k 或者 $\varepsilon_k \|B_k\|_F$，其中 $\varepsilon_k > 0$ 是某一个参数且它的值与 k 有关，当 $k \to +\infty$ 时，$\varepsilon_k \to 0$。

步长 α_k 是采用 Wolfe 型非精确搜索确定的，即通过求解

$$\begin{cases} f(x_k + \alpha_k d_k) - f(x_k) \le \sigma_1 \alpha_k \nabla f(x_k)^{\mathrm{T}} d_k \\ \nabla f(x_{k+1})^{\mathrm{T}} d_k \ge \sigma_2 \nabla f(x_k)^{\mathrm{T}} d_k \end{cases} \tag{4}$$

来计算步长 α_k，则条件 $y_k^{\mathrm{T}} s_k > 0$ 被满足，其中 σ_1 和 σ_2 是正的常数，且满足 $\sigma_1 < \sigma_2 < 1$。

新的拟牛顿矩阵由 Zhang，Deng 和 Chen 文献[2]提出，由下列 BFGS 公式得到

$$B_{k+1} = B_k - \frac{B_k s_k s_k^{\mathrm{T}} B_k}{s_k^{\mathrm{T}} B_k s_k} + \frac{\overline{y}_k \overline{y}_k^{\mathrm{T}}}{\overline{y}_k^{\mathrm{T}} s_k}, \tag{5}$$

其中

$$\overline{y}_k = \nabla f(x_{k+1}) - \nabla f(x_k) + r_k s_k, \tag{6}$$

$$s_k = x_{k+1} - x_k, \tag{7}$$

$$r_k = \frac{3\left(\nabla f\left(x_{k+1}\right)+\nabla f\left(x_k\right)\right)^{\mathrm{T}} s_k - 6\left(f\left(x_{k+1}\right)-f\left(x_k\right)\right)}{\left\|s_k\right\|^2}. \tag{8}$$

再由 Wolfe 搜索下则条件 $y_k^{\mathrm{T}} s_k > 0$ 被满足，而由 Zhang，Deng 和 Chen 文献[2]中的引理 4 可知当 $y_k^{\mathrm{T}} s_k > 0$， $\overline{y}_k^{\mathrm{T}} s_k > 0$。

下面我们给出求解问题(1)的新的扰动的 BFGS 方法的具体步骤：

算法 1.1：(新的扰动的 BFGS 方法)

步 0：选择初始点 $x_1 \in R^n$ 和初始的对称正定矩阵 $B_1 \in R^{n\times n}$ 以及 $Q = I$。选用常数 $0<\sigma_1<\sigma_2<1$，ρ，τ， $\eta \in (0,1)$， $\varepsilon_1 > 0$ 以及 $M_B > 1$。令 $u_1 = \varepsilon_1$ 以及 $\delta = \left\|\nabla f\left(x_1\right)\right\|$。令 $k=1$。

步 1：解线性方程组(3)得到 d_k， 转步 2。

步 2：由条件(4)来计算 α_k， 令 $x_{k+1} = x_k + \alpha_k d_k$。

步 3：由修正公式(5)，(6)，(7)，(8)确定 B_{k+1}。

步 4：如果 $\left\|\nabla f\left(x_{k+1}\right)\right\|/\delta \le \eta$， 令 $\varepsilon_{k+1} = \tau\varepsilon_k$， $u_{k+1} = \varepsilon_{k+1}$， $\delta = \left\|\nabla f\left(x_{k+1}\right)\right\|$。

否则令 $\varepsilon_{k+1} = \varepsilon_k$ 以及

$$u_{k+1} = \begin{cases} \varepsilon_{k+1}\left\|B_{k+1}\right\|_F, & \text{如果} \left\|B_{k+1}\right\|_F \ge M_{k+1} \\ \varepsilon_{k+1}, & \text{其他} \end{cases}$$

其中 $M_k = \max\left\{M_B, 1/\left\|\nabla f\left(x_k\right)\right\|\right\}$。

步 5：令 $k = k+1$，然后转步 1。

在算法 1.1 中 ε_k 按下面的规则逐步被减少：

$$\varepsilon_{k+1} = \begin{cases} \tau\varepsilon_k, & \text{如果} \left\|\nabla f\left(x_{k+1}\right)\right\|/\delta \le \eta \\ \varepsilon_k, & \text{否则} \end{cases}$$

其中 τ， $\eta \in (0,1)$ 是正的常数，并且 $\delta = \left\|\nabla f\left(x_i\right)\right\|$， i 取 $[1,k]$ 中的某个指标，它的值随着迭代过程被不断的修改，这样得到一个序列 $\left\{\left\|\nabla f\left(x_k\right)\right\|\right\}$ 中的一个公比不超过 $\eta\,(0<\eta<1)$ 的几何子列。如果

$$\liminf_{k\to\infty}\left\|\nabla f\left(x_k\right)\right\| = 0$$

成立，则 $\{\varepsilon_k\}$ 单调非增并且当 $k\to\infty$ 时 $\varepsilon_k \to 0$。由此可以得到当 $\left\|B_k\right\|_F$ 有界时， $\varepsilon_k\left\|B_k\right\|_F \to 0$。

3. 算法全局收敛性证明

为了证明新的扰动 BFGS 方法的全局收敛性，我们做出如下的假设

假设 3.1：水平集

$$\Omega = \left\{x \in R^n : f(x) \le f\left(x_1\right)\right\}$$

有界，函数 $f(x)$ 在 Ω 上 Lipschitz 连续的梯度，即存在常数 $L>0$ 满足

$$\left\|\nabla f(x) - \nabla f(y)\right\| \le L\left\|x-y\right\|, \forall x, y \in \Omega. \tag{3.1}$$

需要注意的是采用线性搜索(4)，由(5)产生的矩阵总是正定的，从而序列 $\left\{f\left(x_k\right)\right\}$ 是单调递减的，而且由线性搜索(4)的第一个不等式以及假设 3.1 可以得到

$$-\sum_{k=1}^{\infty}\alpha_k \nabla f\left(x_k\right)^{\mathrm{T}} d_k < +\infty,$$

由此可推出

$$-\lim_{k\to\infty}\alpha_k\nabla f\left(x_k\right)^{\mathrm{T}}d_k=-\lim_{k\to\infty}\nabla f\left(x_k\right)^{\mathrm{T}}s_k=0. \tag{9}$$

为了表示的方便，根据算法 1.1 中的步 4 我们定义指标集合

$$K_\tau=\left\{k:\varepsilon_{k+1}=\tau\varepsilon_k\right\}, \tag{10}$$

以及

$$K_B=\left\{k:\left\|B_k\right\|_F\geq M_k\right\}. \tag{11}$$

为了证明算法 1.1 的全局收敛性我们需要先证明如下的引理：

引理 3.1： 假设 K_τ 是有限集，则算法 1.1 产生的序列 $\left\{\alpha_k\right\}$ 有非零的下界，并且当 $k\to\infty$ 时，$d_k\to 0$。

证明：如果 K_τ 是有限集，因此 ε_k 被减少有限次，也就是说存在一个整数 k_1 使得对所有的 $k\geq k_1$ 有 $\varepsilon_k=\varepsilon_{k_1}$，由于 B_k 是正定的，对于所有的 k 有

$$-\nabla f\left(x_k\right)^{\mathrm{T}}d_k=d_k^{\mathrm{T}}\left(B_k+u_kI\right)d_k\geq\varepsilon_{k_1}\min\left\{1,M_B\right\}\left\|d_k\right\|^2. \tag{12}$$

由(4)中的第二个不等式以及假设 3.1，我们得到对所有的 $k\geq 1$，不等式

$$L\alpha_k\left\|d_k\right\|^2\geq\left(\nabla f\left(x_k+\alpha_kd_k\right)-\nabla f\left(x_k\right)\right)^{\mathrm{T}}d_k\geq-\left(1-\sigma_2\right)\nabla f\left(x_k\right)^{\mathrm{T}}d_k$$

成立。因此，由(12)，我们可以推出

$$\begin{aligned}\alpha_k&\geq\frac{-\left(1-\sigma_2\right)\nabla f\left(x_k\right)^{\mathrm{T}}d_k}{L\left\|d_k\right\|^2}=\frac{\left(1-\sigma_2\right)d_k^{\mathrm{T}}\left(B_k+u_kI\right)d_k}{L\left\|d_k\right\|^2}\\&\geq\left(1-\sigma_2\right)L^{-1}\varepsilon_{k_1}\min\left\{1,M_B\right\}>0\end{aligned} \tag{13}$$

而且由(13)，(9)和(12)易知当 $k\to\infty$ 时，$d_k\to 0$。

引理 3.2： 如果 K_B 是无限集，而 K_τ 是有限集，则 K_τ 必有

$$\liminf_{k\to\infty}\left\|\nabla f\left(x_k\right)\right\|=0. \tag{14}$$

证明：令 $\widehat{d_k}=\left\|B_k\right\|_F d_k$，$\widehat{\alpha_k}=\alpha_k/\left\|B_k\right\|_F$。则 $\widehat{\alpha_k}\widehat{d_k}=\alpha_kd_k$。当 K_τ 是有限集时，我们从集合 K_B 的定义(11)以及算法 1.1 的步 4 知，对于任何的 $k\geq k_1$ 且 $k\in k_B$ 有，$u_k=\varepsilon_{k_1}\left\|B_k\right\|_F$。即 d_k 满足：

$$B_kd_k+\varepsilon_{k_1}\left\|B_k\right\|_F d_k+\nabla f\left(x_k\right)=0.$$

所以

$$-\nabla f\left(x_k\right)^{\mathrm{T}}\widehat{d_k}=d_k^{\mathrm{T}}B_k\widehat{d_k}+\varepsilon_{k_1}\left\|\widehat{d_k}\right\|^2. \tag{15}$$

再由(4)的第二个不等式和 $\nabla f\left(x\right)$ 的 Lipschitz 连续性得

$$L\widehat{\alpha_k}\left\|\widehat{d_k}\right\|^2=L\alpha_kd_k^{\mathrm{T}}\widehat{d_k}\geq\left(\nabla f\left(x_k+\alpha_kd_k\right)-\nabla f\left(x_k\right)\right)^{\mathrm{T}}\widehat{d_k}\geq-\left(1-\sigma_2\right)\nabla f\left(x_k\right)^{\mathrm{T}}\widehat{d_k}.$$

由上面不等式以及(15)可推出

$$\widehat{\alpha_k}\geq\left(1-\sigma_2\right)\varepsilon_{k_1}L^{-1}>0.$$

进而由上面的关系以及(15)和(9)，我们得

$$\lim_{k\to\infty,k\in K_B}\widehat{d_k}=0. \tag{16}$$

而且对所有的 $k\in K_B$，有

$$\left\|\nabla f\left(x_k\right)\right\|\le\left\|B_k\right\|\left\|\widehat{d_k}\right\|\Big/\left\|B_k\right\|_F+\varepsilon_{k_1}\left\|\widehat{d_k}\right\|\le\left(1+\varepsilon_{k_1}\right)\left\|\widehat{d_k}\right\|.$$

因而由(16)可得(14)。

引理 3.3：如果 K_B 是有限集且 K_τ 也是有限的，那么(14)成立。

证明：我们用反证法假设存在一个某个正数 γ 使得对所有 k 有 $\left\|\nabla f\left(x_k\right)\right\|\ge\gamma$ 成立。由算法的 1.1 的步 4 可得，当 $k\notin K_B$ 时，$u_k=\varepsilon_k$，并且

$$\left\|B_k\right\|_F<\max\left\{M_B,1\big/\left\|\nabla f\left(x_k\right)\right\|\right\}\le\max\left\{M_B,1/\gamma\right\}. \tag{17}$$

当 K_τ 是有限集时，则(12)对所有的 $k\ge k_1$ 成立。由引理 3.1 以及(9)可得 $d_k\to0\left(k\to\infty\right)$。由(3)和(17)可推出，对所有 $k\notin K_B$ 有下面的不等式成立

$$\left\|\nabla f\left(x_k\right)\right\|\le\left(\left\|B_k\right\|+\varepsilon_k\right)\left\|d_k\right\|\le\left(\left\|B_k\right\|_F+\varepsilon_1\right)\left\|d_k\right\|\le\left(\max\left\{M_B,1/\gamma\right\}+\varepsilon_1\right)\left\|d_k\right\|.$$

这就意味着

$$\lim_{k\to\infty,k\notin K_B}\left\|\nabla f\left(x_k\right)\right\|=0.$$

这与假设相矛盾，所以必有(14)成立。

在引理 3.2 和 3.3 的基础上我们证明了算法 1.1 的全局收敛性。

定理 3.1：设序列 $\left\{x_k\right\}$ 由算法 1.1 产生并且满足假设 3.1 则(14)成立。

证明：我们考虑两种情形：(i) K_τ 是无限集和(ii) K_τ 是有限集。

情形(i)：由 K_τ 的定义(10)，必存在 $\left\{\left\|\nabla f\left(x_k\right)\right\|\right\}$ 中的一个公比不超过 $\eta\left(0<\eta<1\right)$ 几何子序列，此时必有(14)成立。

情形(ii)：我们需要考虑 K_B 是无限还是有限两种情况。由引理 3.2 和 3.3 可知，上面两种情况的任意一种情况中，(14)总是成立的，也就是说 ε_k 被减少无穷多次，由 K_τ 的定义(10)中可知情形(ii)是不会出现的。

由定理 3.1 的证明可以看出，ε_k 一定被减少无穷多次，所以当 $k\to\infty$ 时，$\varepsilon_k\to0$。如果 B_k 是有界的，那么当 k 充分大时，$u_k=\varepsilon_k$，并且当 $k\to\infty$ 时，$u_k\to0$。

4. 数值实验与结果分析

在 CPU 为 1.6 HZ、内存为 512 MB 的电脑上分别对新的扰动的 BFGS 算法(算法 1.1)和文献[3]中的 CBFGS 等算法进行了数值实验，其中测试问题的函数都来源于文献[4]。程序采用 Matlab 语言编写，在算法中我们使用 $\left\|\nabla f\left(x_k\right)\right\|\le10^{-6}$ 作为算法的终止条件，算法中有关常数与参数的选择如下：$\sigma_1=0.001$，$\sigma_2=0.9$，$\eta=0.5$，$\tau=0.7$，$\varepsilon_1=1.0$，$M_B=10^{10}$，$B_1=I$ (单位矩阵)。

首先选择了 3 个二维问题，1 个三维问题，2 个四维问题，对算法进行了分析，问题分别是：

rose

$$\min f\left(x\right)=100\left(x_1^2-x_2\right)^2+\left(x_1-1\right)^2,\quad x=\left(x_1,x_2\right)^{\mathrm{T}}\in R^2,$$

初始点：$x_0=\left(-1.2,1\right)^T$，精确解：$x^*=\left(1,1\right)^{\mathrm{T}}$，$f\left(x^*\right)=0$。

badscp

$$\min f(x)=\left(10^4 x_1x_2-1\right)^2+\left(\mathrm{e}^{-x_1}+\mathrm{e}^{-x_2}-1.0001\right)^2,\quad x=\left(x_1,x_2\right)^{\mathrm{T}}\in R^2$$

初始点：$x_0=(0,1)^{\mathrm{T}}$，精确解：$x^*=(1.098,9.106)^{\mathrm{T}}$，$f\left(x^*\right)=0$。

badscb

$$\min f(x)=\left(x_1-10^6\right)^2+\left(x_2-2\times10^{-6}\right)^2+\left(x_1x_2-2\right)^2,\ x=\left(x_1,x_2\right)^{\mathrm{T}}\in R^2$$

初始点：$x_0=(1,1)^{\mathrm{T}}$，精确解：$x^*=\left(10^6,2\times10^{-6}\right)^{\mathrm{T}}$，$f\left(x^*\right)=0$。

helix

$$\min f(x)=100\left[x_3-10\theta\left(x_1,x_2\right)\right]^2+100\left[\left(x_1^2+x_2^2\right)^{\frac{1}{2}}-1\right]^2+x_3^2,$$

其中 $\theta\left(x_1,x_2\right)=\begin{cases}\dfrac{1}{2\pi}\arctan\left(\dfrac{x_2}{x_1}\right),x_1>0\\ \dfrac{1}{2\pi}\arctan\left(\dfrac{x_2}{x_1}\right)+0.5,x_1<0\end{cases}$，$x=\left(x_1,x_2,x_3\right)^{\mathrm{T}}\in R^3$，

初始点：$(-1,0,0)^{\mathrm{T}}$，精确解：$x^*=(1,0,0)^{\mathrm{T}}$，$f\left(x^*\right)=0$。

sing

$$\min f(x)=\left(x_1+10x_2\right)^2+5\left(x_3-x_4\right)^2+\left(x_2-2x_3\right)^4+10\left(x_1-x_4\right)^4,\ x=\left(x_1,x_2,x_3,x_4\right)^{\mathrm{T}}\in R^4,$$

初始点：$x_0=(3,-1,0,1)^{\mathrm{T}}$，精确解：$x^*=(0,0,0,0)^{\mathrm{T}}$，$f\left(x^*\right)=0$。

wood

$$\begin{aligned}\min f(x)=&100\left(x_2-x_1^2\right)^2+\left(1-x_1\right)^2+90\left(x_4-x_3^2\right)^2+\left(1-x_3\right)^2\\&+10\left(x_2+x_4-2\right)^2+\frac{1}{10}\left(x_2-x_4\right)^2\end{aligned},\ x=\left(x_1,x_2,x_3,x_4\right)^{\mathrm{T}}\in R^4,$$

初始点：$x_0=(-3,-1,-3,-1)^{\mathrm{T}}$，精确解 $x^*=(1,1,1,1)^{\mathrm{T}}$，$f\left(x^*\right)=0$。

通过编程计算得到表 1。

Table 1. Result analysis of solving several kinds of problems
表 1. 求解几类问题的结果分析

问题	维度	初始点	迭代步骤	最优值
rose	2	$(-1.2,1)^{\mathrm{T}}$	40	1.2045×10^{-15}
badscp	2	$(0,1)^{\mathrm{T}}$	188	1.0004×10^{-8}
badscb	2	$(1,1)^{\mathrm{T}}$	29	5.4211×10^{-20}
helix	3	$(-1,0,1)^{\mathrm{T}}$	33	1.1439×10^{-16}
sing	4	$(3,-1,0,1)^{\mathrm{T}}$	64	1.0250×10^{-9}
wood	4	$(-3,-1,-3,-1)^{\mathrm{T}}$	59	1.1293×10^{-8}

通过表 1 的结果，得出我们的算法能有效地解决二维和三维问题，以及四维问题，说明我们的算法是有效的。

Table 2. Comparisons with other methods
表 2. 与其他几类方法的比较分析

问题 \ 方法的迭代步	新的扰动 BFGS 迭代步	CBFGS 迭代步	MBFGS 迭代步	BFGS 迭代步
rose	40	27	31	21
badscp	188	441	482	441
badscb	29	20	31	21
helix	33	33	35	34

通过表 2 对不同的 BFGS 方法的迭代步进行分析，我们可以得出我们扰动的算法在求解 badscp 问题时需要 188 步，比其他三种方法的迭代步要少，同时在求解其他问题时的迭代步骤相差不大。

5. 小结

通过数值实验结果和全局收敛性的证明，说明我们提出的新扰动的 BFGS 方法对求解非凸优化问题是有效的。

参考文献

[1] 刘陶文. BFGS 方法及其在求解约束优化问题中的应用[D]: [博士学位论文]. 长沙: 湖南大学, 2006.

[2] Zhang, J.Z., Deng, N.Y. and Chen, L.H. (1999) New Quasi-Newton Equation and Related Method for Unconstrained Optimization. *Journal of Optimization Theory and Applications*, **102**, 147-167. https://doi.org/10.1023/A:1021898630001

[3] 张继伟. 修正 Broyden 族拟牛顿算法及其应用[D]: [博士学位论文]. 长沙: 湖南大学, 2006.

[4] More, J.J., Garbow, B.S. and Hillstrom, K.E. (1981) Testing Unconstrained Optimization Software. *ACM Transactions on Mathematical Software*, **7**, 17-41. https://doi.org/10.1145/355934.355936