Advances in Applied Mathematics 应用数学进展, 2019, 8(2), 250-257
Published Online February 2019 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2019.82029

The Perturbed BFGS Method for the Unconstrained Optimization with the Armijo Line Search

Jiaojiao Yan

Department of Mathematics and Statistics, Changsha University of Science and Technology, Changsha Hunan
Email: 453969709@qq.com

Received: Jan. 24th, 2019; accepted: Feb. 8th, 2019; published: Feb. 15th, 2019

Abstract

In [1], a perturbed BFGS method was proposed to solve the unconstrained optimization and was proved to be globally convergent when the Wolfe line search was used. In this paper, we show that the perturbed BFGS method also possesses global convergence for nonconvex problems with the relatively weaker Armijo line search. Numerical results show that this method with the Armijo search is also promising.

Keywords

BGFS Method, Armijo Linear Search, Global Convergence

Armijo搜索下求解无约束优化问题的扰动BFGS方法

严娇娇

长沙理工大学数学与统计学院，湖南 长沙
Email: 453969709@qq.com

收稿日期：2019年1月24日；录用日期：2019年2月8日；发布日期：2019年2月15日

摘 要

文献[1]提出了一种求解无约束优化问题的扰动BFGS方法，并在Wolfe搜索下证明了其全局收敛性。本文证明了该扰动BFGS方法在较弱的Armijo线性搜索下求解非凸问题也具有全局收敛性。数值结果表明在

文章引用: 严娇娇. Armijo 搜索下求解无约束优化问题的扰动 BFGS 方法[J]. 应用数学进展, 2019, 8(2): 250-257.
DOI: 10.12677/aam.2019.82029

Armijo搜索下该方法也具有较好的数值效果。

关键词

BGFS方法，Armijo线性搜索，全局收敛性

1. 引言

本文主要研究无约束优化问题：

$$\min f(x),\ \ x \in R^n, \tag{1.1}$$

其中 $f: R^n \to R$ 是一个连续可微的函数。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[2]，该方法具有如下迭代格式：

$$x_{k+1} = x_k + \alpha_k d_k,\ \ k \ge 1, \tag{1.2}$$

其中步长 α_k 可通过某种线性搜索得到，搜索方向 d_k 是以下线性方程的解，

$$B_k d_k + \nabla f(x_k) = 0, \tag{1.3}$$

这里 $\nabla f(x)$ 是 f 在 x 处的梯度， B_k 是拟牛顿矩阵，由下面 BFGS 公式进行校正：

$$B_{k+1} = B_k - \frac{B_k s_k s_k^{\mathrm{T}} B_k}{s_k^{\mathrm{T}} B_k s_k} + \frac{y_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k}, \tag{1.4}$$

其中 $s_k = x_{k+1} - x_k$， $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$。BFGS 校正公式的一个重要性质当 $y_k^{\mathrm{T}} s_k > 0$，且 B_k 对称正定时， B_{k+1} 也对称正定[3]。

Powell 在文献[4]中证明了求解凸优化问题的 BFGS 方法只有全局收敛性。但对非凸优化问题，Dai [5]和 Mascarenhas [6]给出例子说明采用 Wolfe 型非精确线性搜索和精确线性搜索的 BFGS 方法不一定具有全局收敛性。在文献[1]中，Liu 提出了一种扰动的 BFGS 方法，简称 PBFGS，在该算法中，搜索方向 d_k 由以下线性方程组确定。

$$(B_k + \mu_k Q) d_k + \nabla f(x_k) = 0, \tag{1.5}$$

这里的 Q 是一个给定的对称正定阵，μ_k 是扰动因子，满足 $\mu_k > 0$，Liu [1]证明了该扰动方法在 Wolfe 搜索下求解非凸问题具有全局收敛性。对 Armijo 搜索，该文未讨论其收敛性，原因是在 Armijo 搜索下 $y_k^{\mathrm{T}} s_k$ 不一定大于 0。

本节，我们给出具体的 Armijo 搜索下扰动的 BFGS 方法，我们采用 Li 在文献[7]中给出的 BFGS 公式校正迭代矩阵 B_k。

$$B_{k+1} = \begin{cases} B_k - \dfrac{B_k s_k s_k^{\mathrm{T}} B_k}{s_k^{\mathrm{T}} B_k s_k} + \dfrac{y_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k}, & \text{如果}\ y_k^{\mathrm{T}} s_k > 0 \\ B_k, & \text{其他} \end{cases} \tag{1.6}$$

其中 s_k 与 y_k 的取值与(1.4)式一致。值得指出的是，(1.6)式中只要 B_0 对称正定，则得到的迭代矩阵序列

$\{B_k\}$总是对称正定的。在第二节中，我们给出具体算法并分析其全局收敛性。在第三节中，我们进行数值实验，验证在 Armijo 搜索下该扰动方法的计算效果。

2. 算法与收敛性分析

本文考虑该扰动的 BFGS 方法在 Armijo 搜索下的全局收敛性。下面我们给出算法的具体步骤。

算法 1 (Armijo 搜索下的 PBFGS 算法):

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 $Q = I$。选取常数 $0<\sigma_1<\sigma_2<1$，$\rho,\tau,\eta \in (0,1), \varepsilon_1 > 0$ 以及 $M_B > 1$。令 $\mu_1 = \varepsilon_1$，$\delta = \|\nabla f(x_1)\|$ 以及 $k=1$。

步 1：解线性方程组

$$(B_k + \mu_k Q)d_k + \nabla f(x_k) = 0, \tag{2.1}$$

解得 d_k。转步 2。

步 2：由 Armijo 线性搜索[8]

$$f(x_k + \alpha_k d_k) - f(x_k) \le \sigma_1 \alpha_k \nabla f(x_k)^{\mathrm{T}} d_k, \tag{2.2}$$

来确定 α_k，其中 α_k 的取值为集合 $\{\rho,\rho^2,\rho^3,\cdots\}$ 中使上面不等式成立的最大者。然后令 $x_{k+1} = x_k + \alpha_k d_k$。

步 3：由修正公式

$$B_{k+1} = \begin{cases} B_k - \dfrac{B_k s_k s_k^{\mathrm{T}} B_k}{s_k^{\mathrm{T}} B_k s_k} + \dfrac{y_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k}, & \text{如果 } y_k^{\mathrm{T}} s_k > 0 \\ B_k, & \text{其他} \end{cases} \tag{2.3}$$

确定 B_{k+1}。

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

$$\mu_{k+1} = \begin{cases} \varepsilon_{k+1}\|B_{k+1}\|_F, & \text{如果 } \|B_{k+1}\|_F \ge M_{k+1} \\ \varepsilon_{k+1}, & \text{其它} \end{cases} \tag{2.4}$$

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

步 5：令 $k = k+1$ 然后转步 1。

在以上的算法中，ε_k 按照如下规则逐渐减小

$$\varepsilon_{k+1} = \begin{cases} \tau\varepsilon_k, & \text{如果 } \|\nabla f(x_{k+1})\|/\delta \le \eta \\ \varepsilon_k, & \text{否则} \end{cases} \tag{2.5}$$

其中 $\tau,\eta \in (0,1)$，并且 $\delta = \|\nabla f(x_i)\|$，$i$ 取值是 $[1,k]$ 中的一个指标。所以 $\{\|\nabla f(x_k)\|\}$ 中必存在一个公比不超过 η 的几何子序列，如果

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0 \tag{2.6}$$

成立，那么 $\{\varepsilon_k\}$ 为单调非增序列，并且当 $k\to\infty$ 时，有 $\varepsilon_k \to 0$。由此可得，当 $\|B_k\|_F$ 有界时，$\varepsilon_k\|B_k\|_F \to 0$，从而 $\mu_k \to 0$，(2.1)式逐渐逼近于(1.3)式。

下面证明算法的全局收敛性，在证明之前需作如下假设：

假设 1： 存在一个水平集 $\Omega = \{x \in R^n \mid f(x) \le f(x_1)\}$ 是一个有界集，函数 $f(x)$ 在 Ω 上有 Lipschitz 连续的梯度，即存在常数 $L > 0$ 使得

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x-y\|, \ (\forall x, y \in \Omega). \tag{2.7}$$

由(2.3)式所产生的矩阵 B_k 总是正定的，从而序列 $\{f(x_k)\}$ 是单调递减的，由假设可知 $\Omega=\{x\in R^n \mid f(x)\le f(x_1)\}$ 是有界的，因而 $f(x)$ 有下界。由 Armijo 线性搜索不等式(2.2)可得

$$f(x_1+\alpha_1 d_1)-f(x_1)\le \sigma_1\alpha_1\nabla f(x_1)^{\mathrm{T}} d_1$$
$$\vdots$$
$$f(x_k+\alpha_k d_k)-f(x_k)\le \sigma_1\alpha_k\nabla f(x_k)^{\mathrm{T}} d_k$$

将上面 k 个式子累加

$$f(x_k+\alpha_k d_k)-f(x_1)\le \sigma_1\sum_{k=1}^{\infty}\alpha_k\nabla f(x_k)^{\mathrm{T}} d_k.$$

因此

$$-\sum_{k=1}^{\infty}\alpha_k\nabla f(x_k)^{\mathrm{T}} d_k\le -\frac{C_1}{\delta_1}<+\infty.$$

其中常数 $C_1=f(x_k+\alpha_k d_k)-f(x_1)<0$，由此可推出

$$-\lim_{k\to+\infty}\alpha_k\nabla f(x_k)^{\mathrm{T}} d_k=-\lim_{k\to+\infty}\nabla f(x_k)^{\mathrm{T}} s_k=0. \tag{2.8}$$

为了方便证明收敛性，根据步 4 可定义指标集合

$$K_\tau=\{k:\varepsilon_{k+1}=\tau\varepsilon_k\},\ K_B=\{k:\|B_k\|_F\ge M_k\}. \tag{2.9}$$

下面为了证明算法的全局收敛性，我们先证明几个重要的引理。

引理 1：假定集合 K_τ 是有限集，则算法所产生的步长序列 $\{\alpha_k\}$ 有非零的下界，并且当 $k\to\infty$ 时，$d_k\to 0$。

证明：假定 K_τ 是有限集，即 ε_k 被减小有限次，也就是说存在一个整数 k_1 使得对任意的 $k\ge k_1$ 都有 $\varepsilon_k=\varepsilon_{k_1}$。由于 B_k 是正定的，那么对于所有的 k，有

$$-\nabla f(x_k)^{\mathrm{T}} d_k=d_k^{\mathrm{T}}(B_k+\mu_k I)d_k\ge\varepsilon_k\|B_k\|_F\|d_k\|^2\ge\varepsilon_k\min\{1,M_B\}\|d_k\|^2. \tag{2.10}$$

由 Armijo 型线性搜索准则知，当 $\alpha_k\ne 1$ 时，$\rho^{-1}\alpha_k$ 不满足(2.2)式[5] [9]，那么我们有[9]

$$f(x_k+\rho^{-1}\alpha_k d_k)-f(x_k)>\sigma_1\rho^{-1}\alpha_k\nabla f(x_k)^{\mathrm{T}} d_k. \tag{2.11}$$

通过中值定理以及 $\nabla f(x)$ 的 Lipschitz 连续性，存在一个 $\theta_k\in(0,1)$，可得

$$\begin{aligned}&f(x_k+\rho^{-1}\alpha_k d_k)-f(x_k)\\&=\rho^{-1}\alpha_k\nabla f(x_k+\theta_k\rho^{-1}\alpha_k d_k)^{\mathrm{T}} d_k\\&=\rho^{-1}\alpha_k\nabla f(x_k)^{\mathrm{T}} d_k+\rho^{-1}\alpha_k\left[\nabla f(x_k+\theta_k\rho^{-1}\alpha_k d_k)^{\mathrm{T}}-\nabla f(x_k)\right]d_k\\&\le\rho^{-1}\alpha_k\nabla f(x_k)^{\mathrm{T}} d_k+L\rho^{-2}\alpha_k^2\|d_k\|^2\end{aligned}, \tag{2.12}$$

这里的 $L>0$ 是 $\nabla f(x_k)$ 的 Lipschitz 常数。再将(2.12)式代入(2.11)式可得

$$\begin{aligned}\alpha_k&\ge\frac{-(1-\sigma_1)\rho\nabla f(x_k)^{\mathrm{T}} d_k}{L\|d_k\|^2}=\frac{(1-\sigma_1)\rho d_k^{\mathrm{T}}(B_k+\mu_k I)d_k}{L\|d_k\|^2}\\&\ge\frac{(1-\sigma_1)\rho\varepsilon_{k_1}}{L}\min\{1,M_B\}>0.\end{aligned} \tag{2.13}$$

当 $\alpha_k = 1$ 时，显然可得 $\alpha_k > 0$。又由(2.8)，(2.10)可得

$$0 = -\lim_{k\to+\infty} \alpha_k \nabla f(x_k)^{\mathrm{T}} d_k = \lim_{k\to+\infty} \alpha_k d_k^{\mathrm{T}} (B_k + \mu_k I) d_k \ge \lim_{k\to+\infty} \alpha_k \varepsilon_k \min\{1, M_B\} \|d_k\|^2 > 0,$$

因此，$k \to \infty$ 时，$d_k \to 0$。

引理 2. 如果 K_B 是无限集，而 K_τ 是有限集，则必有

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{2.14}$$

证明： 令 $\hat{\alpha}_k = \alpha_k / \|B_k\|_F$ 和 $\hat{d}_k = \|B_k\|_F d_k$，那么有 $\hat{\alpha}_k \hat{d}_k = \alpha_k d_k$。当 K_τ 为有限集时，可从集合 K_B 的定义 $K_B = \{k : \|B_k\|_F \ge M_k\}$，这里 $M_k = \max\{M_B, 1/\|\nabla f(x_k)\|\}$ 以及算法的步 4 知，对任意的 $k \ge k_1$，并且 $k \in K_B$，有 $\mu_k = \varepsilon_{k_1} \|B_k\|_F$。又由修正的拟牛顿方法 $(B_k + \mu_k I) d_k + \nabla f(x_k) = 0$，那么 d_k 也满足

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

所以有

$$-\nabla f(x_k)^{\mathrm{T}} \hat{d}_k = d_k^{\mathrm{T}} B_k \hat{d}_k + \varepsilon_{k_1} \|\hat{d}_k\|^2. \tag{2.15}$$

由(2.11)式和(2.12)式可得

$$L\rho^{-2}\alpha_k^2 \|d_k\|^2 \ge (\sigma - 1)\rho^{-1}\alpha_k \nabla f(x_k)^{\mathrm{T}} d_k.$$

由上面不等式及 $\hat{\alpha}_k \hat{d}_k = \alpha_k d_k$，我们可推出

$$L\rho^{-2}\hat{\alpha}_k^2 \|\hat{d}_k\|^2 \ge (\sigma - 1)\rho^{-1}\hat{\alpha}_k \nabla f(x_k)^{\mathrm{T}} \hat{d}_k,$$

将(2.15)式代入上式，从而

$$\hat{\alpha}_k \ge \frac{-(1-\sigma_1)\rho \nabla f(x_k)^{\mathrm{T}} \hat{d}_k}{L\|\hat{d}_k\|^2} = \frac{(1-\sigma_1)\rho\left(d_k^{\mathrm{T}} B_k \hat{d}_k + \varepsilon_{k_1}\|\hat{d}_k\|^2\right)}{L\|\hat{d}_k\|^2} \ge \frac{(1-\sigma_1)\rho\varepsilon_{k_1}}{L} > 0.$$

进一步由(2.8)式及上面的结论有

$$0 = -\lim_{k\to+\infty} \hat{\alpha}_k \nabla f(x_k)^{\mathrm{T}} \hat{d}_k = \lim_{k\to+\infty} \hat{\alpha}\left(d_k^{\mathrm{T}} B_k \hat{d}_k + \varepsilon_{k_1}\|\hat{d}_k\|^2\right) \ge \lim_{k\to+\infty} \hat{\alpha}_k \varepsilon_{k_1} \|\hat{d}_k\|^2 > 0.$$

可得

$$\lim_{k\to\infty, k\in K_B} \hat{d}_k = 0. \tag{2.16}$$

并且对所有的 $k \in K_B$，我们有

$$\|\nabla f(x_k)\| \le \|B_k\| \|\hat{d}_k\| / \|B_k\|_F + \varepsilon_{k_1} \|\hat{d}_k\| \le (1 + \varepsilon_{k_1}) \|\hat{d}_k\|.$$

因此由(2.16)式的结论可得

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

引理 3： 如果 K_B 是有限集并且 K_τ 也是有限集，那么有

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

证明：反设存在一个正数γ使得对所有的k有$\left\|\nabla f\left(x_k\right)\right\|\geq\gamma$成立。由算法的步4可知，当$k\notin K_B$时，有$\mu_k=\varepsilon_k$，且

$$\left\|B_k\right\|_F<\max\left\{M_B,1/\left\|\nabla f\left(x_k\right)\right\|\right\}\leq\max\left\{M_B,1/\gamma\right\}. \tag{2.17}$$

当K_τ为有限集时，对所有的$k\geq k_1$有(2.10)式成立。由引理1和(2.8)可知，当$k\to\infty$时，$d_k\to0$。进一步由修正后的拟牛顿公式(2.1)以及(2.17)式可得，对所有的$k\notin K_B$有下面不等式

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

以上结论说明

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

但是这与假设矛盾，所以必定有(2.14)成立。

在引理2和引理3的基础上，下面我们证明算法1的全局收敛性。

定理1：设序列$\left\{x_k\right\}$由算法产生且满足假设1，那么

$$\lim_{k\to\infty}\inf\left\|\nabla f\left(x_k\right)\right\|=0. \tag{2.18}$$

证明：这里需要考虑两种情形：(i) K_τ是无限集和(ii) K_τ是有限集。

对情形(i)：由K_τ的定义$K_\tau=\left\{k:\varepsilon_{k+1}=\tau\varepsilon_k\right\}$及(2.4)式可知$\left\{\left\|\nabla f\left(x_k\right)\right\|\right\}$中必存在一个公比不超过$0<\eta<1$的几何子序列，那么此时必有(2.18)式成立。

对情形(ii)：这时需要考虑K_B是无限集还是有限集这两种情况。在引理2和引理3中，我们已证明在这两种情况下，结论(2.18)总是成立的，即ε_k被减小无穷多次。因此，由K_τ的定义可知情形(ii)是不可能出现的。

由定理1可以得到，ε_k一定无限次的减小，即$k\to+\infty$时，$\varepsilon_k\to0$。当B_k有界，且k充分大时，有$\mu_k=\varepsilon_k$，显然当$k\to+\infty$时，也可得到$\mu_k\to0$。

3. 数值试验及结果分析

为了验证本文提出的算法对实际问题的可行性，我们进行了数值实验及结果分析。数值实验是通过MATLAB7.0编程，程序在3.2 GHZ处理器，2 GB内存的电脑上实现。本文的所有测试问题以及问题的初始点均选自参考文献[10]。这些问题也被文献[1]采用来测试扰动的BFGS方法。为了体现算法的有效性，本文将计算结果与文献[1]提出的扰动的BFGS方法(PBFGS)和原始的BFGS方法的计算结果作比较。同时规定$\left\|\nabla f\left(x_k\right)\right\|\leq10^{-6}$为算法的终止条件，并规定算法中的常数和参数的取值为：$\sigma_1=0.001,\sigma_2=0.9,\eta=0.5,\tau=0.7,\varepsilon_1=1.0,M_B=10^{10},B=I$。测试问题如下：

问题1：$\min f\left(x\right)=100\left(x_1^2-x_2\right)^2+\left(x_1-1\right)^2$。

问题2：$\min f\left(x\right)=\left(10^4x_1x_2-1\right)^2+\left(\mathrm{e}^{-x_1}+\mathrm{e}^{-x_2}-1.0001\right)^2$。

问题3：$\min f\left(x\right)=100\left(x_1^2-x_2\right)^2+\left(x_1-1\right)^2$。

问题4：$\min f\left(x\right)=100\left(x_3-10\theta\left(x_1,x_2\right)\right)^2+100\left(\left(x_1^2+x_2^2\right)^{1/2}-1\right)^2+x_3^2$，其中

$$\theta(x_1,x_2)=\begin{cases}\dfrac{1}{2\pi}\arctan\left(\dfrac{x_2}{x_1}\right), & x_1>0\\[2ex] \dfrac{1}{2\pi}\arctan\left(\dfrac{x_2}{x_1}\right)+0.5, & x_1<0\end{cases}\text{。}$$

问题 5： $\min f(x)=(x_1+10x_2)^2+5(x_3-x_4)^2+(x_2-2x_3)^4+10(x_1-x_4)^4$。

问题 6： $\min f(x)=100(x_2-x_1^2)^2+(1-x_1)^2+90(x_4-x_3^2)^2+(1-x_3)^2+10(x_2+x_4-2)^2+10(x_2-x_4)^2$。

Table 1. Experimental results of algorithm
表 1. 算法的数值实验结果

	BFGS		PBFGS		PBFGS (Armijo)	
pro	ite	gn	ite	gn	ite	gn
1	25	26	28	29	23	24
2	441	448	14	15	187	188
3	21	41	13	14	14	15
4	28	31	29	30	23	24
5	32	33	33	40	45	46
6	92	96	69	72	17	18

表 1 中表头的各个变量的具体意义如下：pro 表示被测问题序号，ite 表示迭代次数，gn 表示目标函数梯度的计算次数，BFGS 表示使用 Wolfe 搜索下的 BFGS 方法，PBFGS 表示使用 Wolfe 搜索下的扰动的 BFGS 方法，PBFGS (Armijo)表示本文提出的 Armijo 搜索下的扰动的 BFGS 方法。

从表 1 的各项数据可以看出，本论文提出的算法能有效地解决各类问题并且能保持 BFGS 的仿射不变性。虽然在求解其中几个问题时，Armijo 搜索下扰动的 BFGS 方法的迭代次数相对于 PBFGS 方法多几步，是因为 PBFGS 方法采用的是 Wolfe 搜索，对于步长的搜索优于 Armijo 搜索。与 BFGS 方法的迭代次数相比，本文研究的算法数值效果会优于 Wolfe 搜索下的 BFGS 方法。总的来说，我们的 Armijo 搜索下扰动的 BGFS 方法效果较好。

致　谢

本篇论文，我首先要感谢我的导师周伟军教授及我的同班同学，从论文选题、论文的撰写、修改到最终完稿，他们都给予了我很大的帮助。其次，我要感谢我父母对我的支持，是父母的辛勤培育，才造就我现在的学习成果。

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] 李董辉, 童小娇, 万中. 数值最优化算法与理论[M]. 北京: 科学出版社, 2010: 51-63.

[3] 袁亚湘, 孙文瑜. 最优化理论与方法[M]. 北京: 中国科学出版社, 2001: 1-627.

[4] Powell, M.J.D., Cottle, R.W. and Lemke, C.E. (1976) Some Convergence Properties of a Variable Metric Algorithm for Minimization without Exact Line Search. SIAM Publications, 53-72.

[5] Dai, Y.H. (2002) Convergence Properties of the BFGS Algorithm. *SIAM Journal on Optimization*, **13**, 693-701. https://doi.org/10.1137/S1052623401383455

[6] Mascarenhas, W.F. (2004) The BFGS Method with Exact Line Searches Fails for Non-Convex Objective Functions. *Mathematical Programming*, **99**, 49-61. https://doi.org/10.1007/s10107-003-0421-7

[7] Li, D.H. and Fukushima, M. (2001) On the Global Convergence of the BFGS Method for Nonconvex Unconstrained Problems. *SIAM Journal on Optimization*, **11**, 1054-1064. https://doi.org/10.1137/S1052623499354242

[8] Armijo, L. (1966) Minimization of Functions Having Lipschitz Continuous Partial Derivatives. *Pacific Journal of Mathematics*, **16**, 1-3. https://doi.org/10.2140/pjm.1966.16.1

[9] Li, D.H. and Fukushima, M. (2001) A Modified BFGS Method and Its Global Convergence in Nonconvex Minimization. *Journal of Computational and Applied Mathematics*, **192**, 15-35. https://doi.org/10.1016/S0377-0427(00)00540-9

[10] 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