A NEW PROOF FOR THE GLOBAL CONVERGENCE OF THE BFGS METHOD FOR NONCONVEX UNCONSTRAINED MINIMIZATION PROBLEMS

Hakima Degaichia
Department of Mathematics
University of Laarbi Tebessi
Box, 12000 Tebessa
Algeria
hakima_deg@yahoo.com

Salah Boulaaras*
Department of Mathematics
College of Science and Arts
Ar-Ras, Qassim University
Kingdom Of Saudi Arabia
and
Laboratory of Fundamental and Applied Mathematics of Oran (LMFAO)
University of Oran 1
Ahmed Benbella
Algeria
saleh_boulaares@yahoo.fr and S.Boularas@qu.edu.sa

Abstract. In this paper we give a new proof for the global convergence of the BFGS method for nonconvex unconstrained minimization problems and we prove that the condition of the appropriate method is to satisfy implicitly with inaccurate linear search of Wolfe type. Furthermore, we have checked directly the convergence of the method BFGS with the inaccurate linear search of Wolfe.

Keywords: Wolfe type, global convergence, BFGS method, nonconvex unconstrained minimization problems.

1. Introduction

The BFGS method is a known quasi-Newton method and has been used extensively for solving unconstrained minimization problems in the past two decades [4], [8] and [9]. The global convergence quasi-Newton methods have also established especially for convex unconstrained minimization problems [2], [3], [8], [14], [15], [16] and [17]. However, in [11], the authors studied the nonconvex case. They also proposed a modified BFGS method with global and superliner convergence. Moreover, the global convergence result for nonlinear equations is due to Griewank [14] for Broyden's rank one method. However, a potential

∗. Corresponding author

trouble with the mentioned method is that the line search may not be executed finitely in a certain special situation ([14], pp. 81-82).

On the other hand, little is known concerning global convergence of the BFGS method for nonconvex optimization problems. in fact, the global convergence of the BFGS method for nonconvex minimization problems has not been proved until now by any one or has given a counter example that shows nonconvergence of the BFGS method. Whether the BFGS method converges globally for a nonconvex function remains unanswered.

In recent work [12], the authors proposed a globally convergent Gauss–Newton-based BFGS method for symmetric nonlinear equations which contain unconstrained optimization problems as a special case. The results obtained in [12] and [11] positively support the unsolved problem. However, their question still remains unanswered. Then in [11], the authors studied the last motioned problem of whether the BFGS method with inexact line search converges globally when applied to nonconvex unconstrained minimization problems. In addition, they proposed a cautious BFGS update and proved that the method with either a Wolfe type [18] or an Armijo-type [13] line search converges globally if the function to be minimized has Lipschitz continuous gradients.

The purpose of this paper is to study this problem further which an extension on the work of Li and Fukushima in [13], as we have mentioned above, the authors proved the convergence of an appropriate method for the BFGS, but they did not completely proved that their method converged to the BFGS. On the other hand, in the current paper we are interested to prove that the condition of the appropriate method is satisfied implicitly with inaccurate linear search of Wolfe type. Furthermore, we have checked directly the convergence of the method BFGS with the inaccurate linear search of Wolfe.

The outline of the paper is as follows: In section 2, we introduce some necessary notations and introduce BFGS method with appropriate update. Then in section 3, we propose an algorithm in order to analyze the convergence of the BFGS method. In addition, in sections 3 and 4, we give a new proposed algorithm and its global convergence with the linear search of Wolfe type isproved. Furthermore, the convergence of the BFGS method for noncovex unconstrained minimization problem is given as well. We prove that this method is interpreted according to the BFGS method.

2. Global convergence of the BFGS method in the nonconvex case

We introduce some notation: Consider $M \geq 0$ (resp. $M > 0$) is a symmetric positive semi-definite matrix (resp. positive definite) and define the following sets

$$S_{+}^{n} := \{M \in S^{n} : M \geq 0,\ n \in \mathbb{N}\} \tag{2.1}$$

and

$$S_{++}^{n} := \{M \in S^{n} : M > 0,\ n \in \mathbb{N},\} \tag{2.2}$$

where S^n symmetric matrices of order n such that:

We know that in [1]: to impose on $M_{k+1}\,(k \in \mathbb{N})$ to be close to $M_k\,(k \in \mathbb{N})$ and minimize the gap between M_{k+1} and M_k, still requesting that to M_{k+1} is symmetric and satisfies the equation of quasi-Newton. We are thus led to consider the problem in the following variable matrix:

$$\begin{cases} \min(gap)(M, M_k), \\ y_k = Ms_k,\ M \in \mathbb{R}^{n\times n}, \\ M = M^\top. \end{cases} \tag{2.3}$$

So, we say that the matrix is obtained by the variational approach. It is often useful to impose also the positive definition of matrices M_k. Because, for $d_k = -M_k^{-1}g_k$ is a descent direction, M_k is positive definite must be needed, symmetric matrix, for more detail, in fact, we know that any real symmetric and positive definite matrix is invertible, and its inverse is also positive definite. Therefore, it can be written:

$$g_k^\top d_k = -g_k^\top M_k^{-1} g_k < 0, \tag{2.4}$$

so,d_k is a descent direction. This condition defining an open set that cannot directly be used as a constrained in defining the problem M_{k+1} and for this purpose, we first introduce the following function:

$$\psi : S^n \to \mathbb{R}, \tag{2.5}$$

whose domain is S^n_{++} and forms a "barrier" to the edge of the cone S^n_{++} (it devolves to infinity when its argument approaches the edge of S^n_{++}) and infinity:

$$\psi\left(\Upsilon\right) = tr\Upsilon + ld\Upsilon, \tag{2.6}$$

where the function log-determinant $ld : S^n \to \mathbb{R} \cup \{+\infty\}$ is defined $\Upsilon \in S^n$ by

$$ld(\Upsilon) = \begin{cases} -\log\det\Upsilon, & \text{if } \Upsilon \in S^n_{++}, \\ +\infty, & \text{otherwise.} \end{cases}$$

It can be given the following properties of ψ defined in (2.6):

If we denote $\{\alpha_i\}_{i=1,\ldots,n}$ eigenvalues of Υ, it should be got

$$tr\Upsilon = \sum_{i=1}^{n}\alpha_i \text{ and } \det\Upsilon = \prod_{i=1}^{n}\alpha_i \tag{2.7}$$

and so

$$\psi\left(\Upsilon\right) = \sum_{i=1}^{n}\left(\alpha_i - \log\alpha_i\right),\ \text{if } \Upsilon \in S^n_{++}. \tag{2.8}$$

Being given the shape

$$t \in \mathbb{R}_{++} \mapsto t - \log t, \psi(\Upsilon) \tag{2.9}$$

tends to infinity if one of the eigenvalues of Υ tends to zero or to infinity, i.e.,

$$\exists j \in \{1, ..., n\}\,; \lim_{\alpha j \to 0 \text{ or } \infty} \psi(\Upsilon) = \infty. \tag{2.10}$$

Formula (2.8) also shows that the only minimizer of ψ is $\Upsilon = I$ the identity matrix.

If M_k is areal symmetric matrix then the matrix M_k is positive definite if and only if there exists a positive definite matrix A_k as: $A_k^2 = M_k$ such that, the positive definite matrix A_k and it can be put that $A_k = M_k^{\frac{1}{2}}$ is a unique.

In fact, if M_k is a real symmetric matrix, then we can write:

$$U^\top M_k U = \Lambda \tag{2.11}$$

where U satisfies

$$U^\top U = UU^\top = I \tag{2.12}$$

and Λ is a diagonal matrix where the diagonal elements are the eingenvalues of M_k which are strictly positive. Because M_k is a positive definite matrix. Thus we can write

$$M_k = U\Lambda U^\top = U\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}U^\top = \left(U\Lambda^{\frac{1}{2}}U^\top\right)\left(U\Lambda^{\frac{1}{2}}U^\top\right) = M_k^{\frac{1}{2}}M_k^{\frac{1}{2}}. \tag{2.13}$$

It means that we must find a matrix M that is symmetric and positive definite and be close to M_k. Therefore,

$$M \approxeq M_k^{\frac{1}{2}}M_k^{\frac{1}{2}} \tag{2.14}$$

implies

$$M_k^{-\frac{1}{2}}MM_k^{-\frac{1}{2}} \approxeq I. \tag{2.15}$$

In order to minimize the gap between M and M_k, we seek that $M_k^{-\frac{1}{2}}MM_k^{-\frac{1}{2}}$ is close to I; and this can be get by minimizing the term $\psi(M_k^{-\frac{1}{2}}MM_k^{-\frac{1}{2}})$, so that we shall get M_{k+1} close to M_k by solving :

$$\begin{cases} \min \psi(M_k^{-\frac{1}{2}}MM_k^{-\frac{1}{2}}), \\ y_k = Ms_k, \\ M \in S_{++}^n \ \text{(implicit constraint)}. \end{cases} \tag{2.16}$$

If $s_k = 0$ and $y_k \neq 0$, then (2.16) has no solution or if $s_k = 0$ and $y_k = 0$ so the solution of (2.16) is $M = M_k$, otherwise; the non-trivial case where $s_k \neq 0$ is discussed in the following proposition.

Proposition 1 ([1]). *We assume that M_k is symmetric positive definite and that $s_k \neq 0$. Then, the problem (2.16) has a solution if and only if $y_k^\top s_k > 0$. Under this condition the solution M_{k+1} of (2.16) is unique and is only given by one of the following formulas :*

$$M_{k+1} = M_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{M_k s_k s_k^T M_k}{s_k^T M_k s_k} \tag{2.17}$$

and

$$B_{k+1} = \left(I - \frac{s_k y_k^T}{y_k^T s_k} \right) B_k \left(I - \frac{y_k s_k^T}{y_k^T s_k} \right) + \left(I - \frac{s_k s_k^T}{y_k^T s_k} \right), \tag{2.18}$$

where $B_k := M_k^{-1}$ and $B_{k+1} := M_{k+1}^{-1}$.

2.1 Algorithm 1 for BFGS Method

We give the following algorithm

Initial step:

Let $\varepsilon > 0$ be a determined criterion of stopping. Choose $\varkappa_1$ be an initial point and M_1 be any positive definite (e.g. : $M_1 = I$).

Put $k = 1$ and go to the main stages

Main stages.

Step 1:

If $\|\nabla f(\varkappa_k)\| < \varepsilon$ STOP; otherwise, put $d_k = -M_k g_k$ and determine the optimal step λ_k optimal solution of problem

$$\min f(\varkappa_k + \lambda d_k), \lambda \geq 0 \tag{2.19}$$

and putting

$$\varkappa_{k+1} = \varkappa_k + \lambda_k d_k. \tag{2.20}$$

Step 2:

Do M_{k+1} as follows:

$$M_{k+1} = M_k + \frac{y_k y_k^\top}{y_k^\top s_k} - \frac{M_k s_k s_k^\top M_k}{s_k^\top M_k s_k} \tag{2.21}$$

with

$$\begin{cases} s_k = \varkappa_{k+1} - \varkappa_k, \\ y_k = \nabla f(\varkappa_{k+1}) - \nabla f(\varkappa_k). \end{cases} \tag{2.22}$$

Replace k by $k+1$ and go to step 1.

2.2 BFGS method with appropriate update

It has been seen that the properties of the BFGS formula is that the matrix M_{k+1} inherits the positive definiteness of M_k if the condition $y_k^\top s_k > 0$ is checked. It can be noted if one uses an exact linear search or inexact search of Wolf, then the condition $y_k^\top s_k > 0$ is checked. where as, linear search of Armijo [13] does not guarantee this condition, and therefore M_{k+1} is not necessarily positive definite even if M_k is positive definite. To ensure the positive definiteness of M_{k+1}, the condition $y_k^\top s_k > 0$ is sometimes used to decide whether M_{k+1} is an update or not, i.e. we set

$$M_{k+1} = \begin{cases} M_k + \dfrac{y_k y_k^\top}{y_k^\top s_k} - \dfrac{M_k s_k s_k^\top M_k}{s_k^\top M_k s_k}, & \text{if } y_k^\top s_k > 0, \\ M_k, & \text{otherwise.} \end{cases} \tag{2.23}$$

The condition $y_k^\top s_k > 0$ is often replaced by $y_k^\top s_k > \eta$ where $\eta > 0$ is a small constant. Li and Fukushima [11] and [12] often appropriate update to the BFGS method similar to what is mentioned before and stating from this they establish a global convergence theorem for nonconvex problems. Before describes the appropriate update, first, we shall need the following important lemma due to Powell [16] which will be useful later.

Lemma 1. (Powell [16]) *If the BFGS method with Wolfe linear research (wolfe1)-(wolfe2) [18] is applied to a function f which is continuously differentiable; and if there exists a constant $c > 0$ such as:*

$$\frac{\|y_k\|^2}{y_k^\top s_k} \le c, \quad \textit{for all } k \in \mathbb{N}. \tag{2.24}$$

Then we have

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \tag{2.25}$$

Remark 1. If f is two times continuously differentiable and strictly convex, so we shall always get the inequality (2.24), but in the case where f is not convex, it is difficult to guarantee (2.24). May be it is one of the reasons why the global convergence of the BFGS method has not been proven.

Now, we shall present the BFGS method with appropriate update and show later that is globally convergent without the economic function be convex. To be more precise, we determine M_{k+1} depending on M_k function :

$$M_{k+1} = \begin{cases} M_k + \dfrac{y_k y_k^\top}{y_k^\top s_k} - \dfrac{M_k s_k s_k^\top M_k}{s_k^\top M_k s_k}, & \text{if } \frac{y_k^\top s_k}{\|s_k\|^2} \ge \varepsilon \|g_k\|^\alpha, \\ M_k, & \text{otherwise,} \end{cases} \tag{2.26}$$

where ε and α are positive constants.

2.3 Algorithm 2 of BFGS method with appropriate update

We give the following algorithm:

Step 0. Choose an initial point $\varkappa_0 \in IR^n$ with an initial matrix $M_0 \in \mathbb{R}^{n\times n}$ which is symmetric and positive definite choose the constants

$$0 < \sigma_1 < \sigma_2 < 1, \alpha > 0 \text{ and } \varepsilon > 0.$$

Let $k = 0$

Step 1: Solve the linear equation $M_k d_k + g_k = 0$ to have d_k.

Step 2: Determine the domain $\lambda_k > 0$ by the inexact linear search of Wolfe or Armijo[18].

Step 3: Calculate $\varkappa_{k+1} := \varkappa_k + \lambda_k d_k$.

Step 4: Determine M_{k+1} by

$$M_{k+1} = \begin{cases} M_k + \dfrac{y_k y_k^\top}{y_k^\top s_k} - \dfrac{M_k s_k s_k^\top M_k}{s_k^\top M_k s_k}, & \text{if } \frac{y_k^\top s_k}{\|s_k\|^2} \geq \varepsilon \|g_k\|^\alpha \\ M_k, & \text{otherwise} \end{cases} \tag{2.27}$$

with

$$\begin{cases} s_k = \varkappa_{k+1} - \varkappa_k, \\ y_k = \nabla f(\varkappa_{k+1}) - \nabla f(\varkappa_k). \end{cases} \tag{2.28}$$

Step 5: Replace k by $k+1$ and go to step1.

Remark 2. It is not difficult to see that the matrix M_k generated by algorithm 2 are symmetric and positive definite for all $k \in \mathbb{N}$. This implies that only with the use of inexact linear search of Wolfe or Armijo, we can obtain that the sequence $\{f(\varkappa_k)\}_{k\in\mathbb{N}}$ is decreasing. Also, we have the considerations of the following: (wolfe1) or (wolfe1)-(wolfe2) and if f is an inferiorly bounded:

$$-\sum_{k=0}^{\infty} g_k^\top s_k < \infty. \tag{2.29}$$

This implies that

$$-\lim_{k\to\infty} (-g_k^\top s_k) = 0 \tag{2.30}$$

and since

$$s_k = \varkappa_{k+1} - \varkappa_k = \lambda_k d_k. \tag{2.31}$$

Thus, we have

$$-\lim_{k\to\infty} (\lambda_k g_k^\top d_k) = 0. \tag{2.32}$$

3. Global convergence of BFGS method with an appropriate update

In this section, we shall prove the global convergence of algorithm 1 under the following hypothesis:

Hypothesis 1. Consider the following set

$$\Omega = \{\varkappa \in IR^n / f(\varkappa) \leq f(\varkappa_0)\}. \tag{3.1}$$

We assume that Ω is contained in a bounded convex set D and that the economic function f is continuously differentiable on D and there exists a constant $l > 0$ such as:

$$\|g(\varkappa) - g(y)\| \leq l \|\varkappa - y\|, \text{for all } \varkappa, y \in D \text{ i.e., } f \in C^{1,1}(D). \tag{3.2}$$

Since the sequence$\{f(\varkappa_k)\}_{k\in\mathbb{N}}$ is decreasing, it is clear that the sequence $\{\varkappa_k\}_{k\in\mathbb{N}}$ generated by the algorithm 2 is contained in Ω. Because

$$f(\varkappa_{k+1}) < f(\varkappa_k) < .... < f(\varkappa_1) < f(\varkappa_0).$$

Define the following sets of indices:

$$\overline{K} = \left\{ i \setminus \frac{y_i^\top s_i}{\|s_i\|^2} \geq \varepsilon \|g_i\|^\alpha \right\} \tag{3.3}$$

and

$$\bar{K}_k = \left\{ i \in \overline{K} \setminus i \leq k \right\}.$$

We note through i_k, the set of indices $i \in \overline{K}_k$.

We can rewrite (2.26) of the form:

$$M_{k+1} = \begin{cases} M_k + \dfrac{y_k y_k^\top}{y_k^\top s_k} - \dfrac{M_k s_k s_k^\top M_k}{s_k^\top M_k s_k}, & \text{if } k \in \overline{K}, \\ M_k, & \text{otherwise} \end{cases} \tag{3.4}$$

and considering the trace of both sides of (3.4), we can write for any $k \in \mathbb{N}$

$$Tr(M_{k+1}) = Tr(M_1) + \sum_{i\in\bar{K}_k} \frac{\|y_i\|^2}{y_i^\top s_i} - \sum_{i\in\overline{K}_k} \frac{\|M_i s_i\|^2}{s_i^\top M_i s_i}. \tag{3.5}$$

Theorem 1. *We assume that the hypothesis 1 is true and be $\{\varkappa_k\}_{k\in\mathbb{N}}$ the sequence generated by algorithm 2. If $\bar{K}$ is a finite set, so*

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{3.6}$$

Proof. If $\overline{K}$ *is a finite set, so, there exists an index* k_0 *such as*

$$M_k = M_{k_0} \triangleq M \quad \textit{for all} \quad k \geq k_0.$$

By the positive definiteness of M*, there exists positive constants* $c_1 \leq C_1$ *such as*

$$\begin{cases} c_1 \|d\|^2 \leq d^\top M d \leq C_1 \|d\|^2, \\ c_1 \|d\|^2 \leq d^\top M^{-1} d \leq C_1 \|d\|^2, \text{ for all } d \in \mathbb{R}^n. \end{cases} \tag{3.7}$$

1-If the inexact linear search is of Wolfe, *by using* (3.2) *we have*

$$l \|s_k\|^2 \geq \left\| \left(g^\top (\varkappa_{k+1}) - g^\top (\varkappa_k) \right) s_k \right\| \geq y_k^\top s_k$$

and by using (3.7), we have

$$\begin{aligned} \|s_k\|^2 &\geq \sigma_2 g_k^\top s_k - g_k^\top s_k \\ &\geq (1-\sigma_2)\lambda_k^\top s_k^\top M^{-1} s_k \\ &\geq (1-\sigma_2)\lambda_k^\top c_1 \|s_k\|^2, \text{ for all } k \geq k_0, \end{aligned}$$

where

$$s_k = \lambda_k d_k$$

and

$$d_k^\top = \lambda_k^{-1} s_k^\top \text{ and } g_k^\top = -d_k^\top M^{-1}.$$

Therefore, it can be deduced

$$\lambda_k \geq (1-\sigma_2) c_1 l^{-1}, \text{ for all } k \geq k_0.$$

Thus, we get from (3.2)

$$\lim_{k \to \infty} \lambda_k g^\top d_k = \lim_{k \to \infty} \left(-g_k^\top s_k \right) = 0$$

with

$$\lambda_k \geq (1-\sigma_2) c_1 l^{-1} > 0, \text{ for all } k \geq k_0.$$

We can write

$$g_k^\top M^{-1} g_k = -g_k^\top d_k \to 0$$

and from (3.7) *we have :*

$$c_1 \|g_k\|^2 \leq g_k^\top M^{-1} g_k \to 0,$$

then

$$c_1 \|g_k\|^2 \to 0$$

with $c_1 \neq 0$*, so*

$$\lim_{k \to \infty} \|g_k\| = 0. \qquad \square$$

Now we shall prove the global convergence of algorithm 2 in the case where $\overline{K}$ is an infinite set. If by way of contradiction, there exists a constant $\delta > 0$ such as

$$\|g_k\| \geq \delta, \text{for all } k \in \mathbb{N} \tag{3.8}$$

and we shall see that this produces is a contradiction. Before establishing the global convergence theorem of the algorithm 2, we first show some useful lemmas.

Lemma 2. *We assume that the hypothesis 1 be true and that* $\{\varkappa_k\}_{k\in\mathbb{N}}$ *be a sequence generated by algorithm* 2, *assume also that the relation* (3.8) *be true for all* $k \in \mathbb{N}$ *so, there exists a constant* $c_2 > 0$ *such as:*

$$Tr\,(M_{k+1}) \leq c_2 i_k \tag{3.9}$$

and

$$\sum_{i\in\bar{K}_k} \frac{\|M_i s_i\|^2}{s_i^\top M_i s_i} \leq c_2 i_k, \tag{3.10}$$

for all k *sufficiently dig.*

Proof. By *using* (3.3) *and* (3.8), *we have* for all $i \in \overline{K}$

$$y_i^\top s_i \geq \varepsilon \|g_i\|^\alpha \|s_i\|^2 \geq \varepsilon\delta^\alpha \|s_i\|^2$$

implies

$$y_i^\top s_i \geq \varepsilon\delta^\alpha \|s_i\|^2 . \tag{3.11}$$

Under (3.2) *and* (3.11), *we have for all* $i \in \bar{K}$

$$\|g_{i+1} - g_i\| \leq l \|\varkappa_{i+1} - \varkappa_i\| .$$

Therefore,

$$\|y_i\|^2 \leq l^2 \|s_i\|^2 .$$

Since

$$\frac{1}{y_i^\top s_i} \leq \frac{1}{\varepsilon\delta^\alpha \|s_i\|^2},$$

we have

$$\frac{\|y_i\|^2}{y_i^\top s_i} \leq \frac{l^2}{\varepsilon\delta^\alpha} \triangleq c_2'. \tag{3.12}$$

Since

$$\begin{aligned} Tr\left(M_{k+1}\right) &= Tr\left(M_1\right) + \sum_{i \in \bar{K}} \frac{\|y_i\|^2}{y_i^T s_i} - \underbrace{\sum_{i \in \overline{K}} \frac{\|M_i s_i\|^2}{s_i^T M_i s_i}}_{\text{positive term}} \\ &\leq Tr\left(M_1\right) + \sum_{i \in \overline{K}} \frac{\|y_i\|^2}{y_i^T s_i} \leq i_k \left(c_0 + c_2'\right) = i_k c_2. \end{aligned}$$

Putting

$$c_2 = \max\left(c_0, c_2'\right).$$

Thus, we have:

$$Tr\left(M_{k+1}\right) \leq i_k c_2.$$

Since $Tr\left(M_{k+1}\right) > 0$ *for any* $k \in \mathbb{N}$, *we get from* (3.5) *and* (3.12),

$$0 < Tr\left(M_1\right) + \sum_{i \in \overline{K}_k} \frac{\|y_i\|^2}{y_i^\top s_i} - \sum_{i \in \bar{K}_k} \frac{\|M_i s_i\|^2}{s_i^\top M_i s_i}$$

implies

$$\sum_{i \in \overline{K}_k} \frac{\|M_i s_i\|^2}{s_i^\top M_i s_i} < Tr\left(M_1\right) + i_k c_2' < i_k c_2.$$

Therefore,

$$\sum_{i \in \bar{K}_k} \frac{\|M_i s_i\|^2}{s_i^\top M_i s_i} \leq c_2 i_k.$$

□

4. Global convergence of algorithm 2 with the linear search of Wolfe type

For this purpose, we prove first the following lemma as lemma 2.

Lemma 3. *We assume that the hypothesis 1 be a true. Let* $\{\varkappa_k\}_{k \in \mathbb{N}}$ *be a sequence generated by algorithm 2 with* λ_k *determined by the search linear of Wolfe* (*wolfe1*)-(*wolfe2*) [*18*]. *If we have* (3.5) *for all* $k \in \mathbb{N}$, *so there exists a constant* $c_3 > 0$ *such that for all* k *big enough we have:*

$$\prod_{i \in \widetilde{K}_k} \lambda_i \geq c_3^{i_k}. \tag{4.1}$$

Proof. *The formula* (3.4) *gives the following recurrence relation*

$$\det\left(M_{i+1}\right) = \det\left(M_i\right)\left(\frac{y_i^\top s_i}{s_i^\top M_i s_i}\right), \text{ for all } i \in \overline{K} \tag{4.2}$$

and

$$\det(M_{i+1}) = \det(M_i), \text{ for all } i \notin \overline{K}. \tag{4.3}$$

If we note through n_k *the largest index in the set* $\overline{K}$*, so we can write:*

$$\det(M_{n_k+1}) = \det(M_1) \prod_{i\in\overline{K}_k} \frac{y_i^\top s_i}{s_i^\top M_i s_i}. \tag{4.4}$$

On the other hand, from (*wolfe2*) *we get:*

$$g^\top(\varkappa_i + \lambda_i d_i) d_i \geq \sigma_2 g^\top(\varkappa_i) d_i$$

implies

$$\begin{aligned} y_i^\top s_i &= (g_{i+1} - g_i)^\top s_i \geq \sigma_2 g_i^\top s_i - g_i^\top s_i \\ &\geq -(1-\sigma_2) g_i^\top s_i = (1-\sigma_2)\lambda_i^{-1} s_i^\top M_i s_i, \end{aligned} \tag{4.5}$$

where

$$g^\top = -d_i^\top M_i = -\lambda_i^{-1} s_i^\top M_i.$$

Similarly to the proof of Lemma 2, we obtain (4.1) *by using the last inequality* (4.5)*,* (3.9) *up to* (4.4)

Indeed: from the last inequality (4.5), we can write

$$\prod_{i\in\overline{K}_k} \frac{y_i^\top s_i}{s_i^\top M_i s_i} \geq \prod_{i\in\overline{K}_k} \frac{1-\sigma_2}{\lambda_i}$$

with (4.4)*, we can deduce*

$$\det(M_{n_k+1}) \geq \det(M_1) \prod_{i\in\overline{K}_k} \frac{1-\sigma_2}{\lambda_i} \tag{4.6}$$

or

$$\det(M_{n_k+1}) \leq \left[\frac{Tr(M_{n_k+1})}{n}\right]^n. \tag{4.7}$$

Using (4.6), (4.7) *and from* (3.9)

$$\begin{aligned} \prod_{i\in\overline{K}_k} \frac{1-\sigma_2}{\lambda_i} &\leq \frac{\det(M_{n_k+1})}{\det(M_1)} \leq \frac{1}{\det(M_1)}\left[\frac{Tr(M_{n_k+1})}{n}\right]^n \leq \frac{1}{\det(M_1)}\left[\frac{c_2 i_k}{n}\right]^n \\ &\leq \frac{1}{\det(M_1)}\left[\frac{c_2 i_k}{n}\right]^n \\ &\leq \frac{1}{\det(M_1)\, n^n}\left[c_2^{i_k}\right]^n \\ &\leq \frac{1}{\det(M_1)\, n^n}\left[c_2^n\right]^{i_k}, \end{aligned}$$

so, there exists constant c_3 such as

$$\prod_{i\in\widetilde{K}_k} \lambda_i \geq c_3^{i_k}.$$

□

Now we are able to prove the global convergence of algorithm 2 with the linear search of Wolfe which given by the following theorem:

Theorem 2. *Assume that the hypothesis 1 is true. Be $\{\varkappa_k\}_{k\in\mathbb{N}}$ a sequence generated by algorithm 2 with λ_k, $k \in \mathbb{N}$ determined by the linear search of Wolfe (wolfe1)-(wolfe2). So, we have*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{4.8}$$

Proof. *Taking into consideration the theorem 1, It is sufficient to check* (4.8) *in the case where $\overline{K}$ is infinite. We note $\overline{K}$ by*

$$\overline{K} = \{k_1 < k_2 < ... < k_n.\}$$

We observe that (2.29) *gives that:*

$$-\sum_{j=0}^{\infty} g_{k_j}^\top s_{k_j} < \infty$$

and as, $M_{k_j}s_{k_j} = -\lambda_{k_j}g_{k_j}$, then

$$\sum_{j=1}^{\infty} \left\|g_{k_j}\right\|^2 \lambda_{k_j} \frac{s_{k_j}^\top M_{k_j}s_{k_j}}{\left\|M_{k_j}s_{k_j}\right\|^2} = -\sum_{j=0}^{\infty} g_{k_j}^\top s_{k_j} < \infty. \tag{4.9}$$

Since

$$\begin{aligned} -g_{k_j}^\top s_{k_j} &= -\left\|g_{k_j}\right\|^2 \lambda_{k_j}^2 \frac{g_{k_j}^\top s_{k_j}}{\left\|g_{k_j}\right\|^2 \lambda_{k_j}^2} = \left\|g_{k_j}\right\|^2 \lambda_{k_j} \frac{-\lambda_{k_j} g_{k_j}^\top s_{k_j}}{\left\|\lambda_{k_j} g_{k_j}\right\|^2} \\ &= \left\|g_{k_j}\right\|^2 \lambda_{k_j} \frac{s_{k_j}^\top M_{k_j}s_{k_j}}{\left\|M_{k_j}s_{k_j}\right\|^2}. \end{aligned}$$

If (2.25) *is not satisfied, then, there exists a constant $\delta > 0$ such as $\|g_k\| \geq \delta$, for all k. Also* (4.9) *implies*

$$\sum_{j=1}^{\infty} \lambda_{k_j} \frac{s_{k_j}^\top M_{k_j}s_{k_j}}{\left\|M_{k_j}s_{k_j}\right\|^2} < \infty.$$

Therefore*, for any $\xi > 0$, there exists an integer $j_0 > 0$, such as for all positive integer* q, *we get*

$$\left(\prod_{j=j_0+1}^{j_0+q} \lambda_{k_j} \frac{s_{k_j}^\top M_{k_j} s_{k_j}}{\|M_{k_j} s_{k_j}\|^2}\right)^{\frac{1}{q}} \leq \frac{1}{q} \sum_{j=j_0+1}^{j_0+q} \lambda_{k_j} \frac{s_{k_j}^\top M_{k_j} s_{k_j}}{\|M_{k_j} s_{k_j}\|^2} \leq \frac{\xi}{q}$$

implies

$$\left(\prod_{j=j_0+1}^{j_0+q} \lambda_{k_j}\right)^{\frac{1}{q}} \leq \frac{\xi}{q} \left(\prod_{j=j_0+1}^{j_0+q} \frac{\|M_{k_j} s_{k_j}\|^2}{s_{k_j}^\top M_{k_j} s_{k_j}}\right)^{\frac{1}{q}} \leq \frac{\xi}{q^2} \sum_{j=j_0+1}^{j_0+q} \frac{\|M_{k_j} s_{k_j}\|^2}{s_{k_j}^\top M_{k_j} s_{k_j}} \leq \frac{\xi}{q^2} \sum_{j=0}^{j_0+q} \frac{\|M_{k_j} s_{k_j}\|^2}{s_{k_j}^\top M_{k_j} s_{k_j}}$$

Using (3.10), it cab be easily deduced

$$\left(\prod_{j=j_0+1}^{j_0+q} \lambda_{k_j}\right)^{\frac{1}{q}} \leq \frac{\xi\,(j_0 + q + 1)}{q^2} c_2,$$

where

$$i_k = j_0 + q + 1.$$

If $q \to \infty$, then we obtain a contradiction. Because, Lemma 3 certifies that the left term of the above inequality is larger than a positive constant. □

Remark 3. To show the global convergence of the BFGS method and the inexact linear search of Wolfe, it is sufficient to show implicitly the existence of the condition $\frac{y_k^\top s_k}{\|s_k\|^2} \geq \varepsilon \|g_k\|^\alpha$.That is to say, the BFGSA is devoted to the BFGS method with inexact linear search of Wolfe.

First step: *for all* $k \geq 1$

$$\frac{y_k^\top s_k}{\|s_k\|^2} \geq (1 - \sigma_2) c_1 \lambda_k^{-\frac{1}{k}} \tag{4.10}$$

indeed*, using* (*Wolfe 2*)*,* we have

$$y_k^\top s_k \geq \sigma_2 g_k^\top s_k - g_k^\top s_k \geq (1 - \sigma_2) \lambda_k^{-1} s_k^\top M_k^{-1} s_k,$$

where

$$s_k = \lambda_k d_k$$

and

$$d_k^\top = \lambda_k^{-1} s_k^\top and g_k^\top = -d_k^\top M_k^{-1} \tag{4.11}$$

and if the inexact linear search used is Wolfe's, then we have the condition $y_k^\top s_k \geq 0$ i.e. the positive definiteness of M_k is preserved, so, there exists positive constants $c_1 \leq C_1$ such as

$$c_1 \|z\|^2 \leq z^\top M_k z \leq C_1 \|z\|^2,$$
$$c_1 \|z\|^2 \leq z^\top M_k^{-1} z \leq C_1 \|z\|^2, \text{for all } z \in \mathbb{R}^n,$$

by (4.11), we have

$$y_k^\top s_k \geq (1-\sigma_2)\lambda_k^{-1} c_1 \|s_k\|^2$$

implies

$$\frac{y_k^\top s_k}{\|s_k\|^2} \geq (1-\sigma_2) c_1 \lambda_k^{-1} \geq (1-\sigma_2) c_1 \lambda_k^{-\frac{1}{k}}.$$

Second step: *For all* $k \geq 1$

$$\left(\prod_{i=1}^{k} \lambda_i \|g_i\|_2^2\right)^{\frac{1}{k}} \leq \frac{c}{k}, \quad c > 0. \tag{4.12}$$

Indeed, we have:

$$\sum_{i=1}^{k} \frac{\|M_i s_i\|_2^2}{s_i^\top M_i s_i} = \sum_{i=1}^{k} \frac{\lambda_i \|g_i\|_2^2}{-g_i^\top s_i} \leq c_1.$$

Then, we use the inequality of the averages twice, and the first condition of Wolfe so

$$\lambda_i \|g_i\|_2^2 \leq c_1 \left(-g_i^\top s_i\right)$$

implies

$$\begin{aligned}
\prod_{i=1}^{k} \lambda_i \|g_i\|^2 &\leq c_1^k \prod_{i=1}^{k} \left(-g_i^\top s_i\right) \\
&\leq \left(\frac{c_1}{k} \sum_{i=1}^{k} \left(-g_i^\top s_i\right)\right)^k \\
&\leq \left(\frac{c_1}{\sigma_1 k} \sum_{i=1}^{k} \left(f(\varkappa_i) - f(\varkappa_{i+1})\right)\right)^k \\
&\leq \left(\frac{c_1}{\sigma_1 k} \left(f(\varkappa_1) - f(\varkappa_{k+1})\right)\right)^k \\
&\leq \frac{c}{k}.
\end{aligned}$$

We deduce (4.12)*, with* $c = c_1 \left(f \left(\varkappa_1 \right) - f_{\min} \right) / \sigma_1$ *where* $f_{\min} \in \mathbb{R}$ *is a lower bound of* $\left\{ f \left(\varkappa_k \right) \right\}_{k \in \mathbb{N}}$.

Third step: To conclude that

$$\frac{y_k^\top s_k}{\left\| s_k \right\|^2} \geq \varepsilon \left\| g_k \right\|^\alpha .$$

Indeed, from (4.12) we get *for all* $k \geq 1$

$$\left(\lambda_k \left\| g_k \right\|^2 \right)^{\frac{1}{k}} \leq \frac{c}{k}$$

implies for all $k \geq 1$

$$\frac{k}{c} \left\| g_k \right\|^{\frac{2}{k}} \leq \lambda_k^{-\frac{1}{k}}. \tag{4.13}$$

From (4.13), (4.10), it implies

$$\begin{aligned} \frac{y_k^\top s_k}{\left\| s_k \right\|^2} &\geq (1 - \sigma_2) c_1 \lambda_k^{-\frac{1}{k}} \\ &\geq (1 - \sigma_2) c_1 \frac{c}{k} \left\| g_k \right\|^{\frac{2}{k}} \end{aligned}$$

and if we put $\alpha = \frac{2}{k}$ and $\varepsilon = (1 - \sigma_2) c_1 \frac{c}{k}$, so we obtain

$$\frac{y_k^\top s_k}{\left\| s_k \right\|^2} \geq \varepsilon \left\| g_k \right\|^\alpha .$$

5. Conclusion

This paper proposes a new proof for the global convergence of the BFGS method for nonconvex unconstrained minimization problem which an extension of the work of Li and Fukushima in [13]. The authors proved the convergence of an appropriate procedures for the BFGS method, but they did not completely proved that their method converged to the BFGS method. However, in the current paper we are interested in proving that the condition of the appropriate method is to satisfy implicitly with inaccurate linear search of Wolfe type [18]. Furthermore, we have checked directly the convergence of the method BFGS with the inaccurate linear search of Wolfe.

Acknowledgments

The authors would like to thank the handling editor and anonymous referees for their careful reading and for relevant remarks/suggestions which helped them to improve the paper. The second author gratefully acknowledge Qassim University in Kingdom of Saudi Arabia.

References

[1] A. Antoniou, W-S. Lu, *Practical Optimization. Algorithms and Engineering Applications*, Springer Science Business Media, LLC, 2007.

[2] R. Byrd, J. Nocedal, *A tool for the analysis of quasi-Newton methods with application to unconstrained minimization*, SIAM J. Numer. Anal., 26 (1989), 727–739.

[3] R. Byrd, J. Nocedal, Y. Yuan, *Global convergence of a class of quasi-Newton methods on convex problems*, SIAM J. Numer. Anal., 24 (1987), 1171–1190.

[4] J.E. Dennis Jr., R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1983.

[5] J.E. Dennis, J.J. More, *Quasi-Newton methods, motivation and theory*, SIAM Review, 19 (1977), 46–89.

[6] R. Fletcher, *Practical Methods of Optimization*, Second Edition, John Wiley & Sons, Chichester, 1987.

[7] R. Fletcher, *An overview of unconstrained optimization*, in Algorithms for Continuous Optimization: The State of the Art", (E. Spedicato ed.), Kluwer Academic Publishers, Boston, 1994, 109-143.

[8] L.C.W. Dixon, *Variable metric algorithms: Necessary and sufficient conditions for identical behavior on nonquadratic functions*, J. Optim. Theory Appl. 10 (1972), 34–40.

[9] R. Fletcher, *Practical Methods of Optimization*, 2nd ed., John Wiley & Sons, Chichester, 1987.

[10] J. Nocedal, *Theory of algorithms for unconstrained optimization*, Acta Numerica, 1 (1992), 199–242.

[11] D.H. Li, M. Fukushima, *A modified BFGS method and its global convergence in nonconvex minimization*, J. Comput. Appl. Math., 129 (2001), 15–35.

[12] D.H. Li, M. Fukushima, *A globally and superlinearly convergent Gauss-Newton based BFGS method for symmetric nonlinear equations*, SIAM J. Numer. Anal., 37(1999), 152–172.

[13] D.H. Li, M. Fukushima, *On the global convergence of the BFGS method for nonconvex unconstrained optimization problem*, SIAM J. Optim. 11 (2001), 1054–1064.

[14] A. Griewank, *The global convergence of partitioned BFGS on problems with convex decompositions and Lipschitzian gradients*, Math. Program., 50 (1991), 141–175.

[15] M.J.D. Powell, *On the convergence of the variable metric algorithm*, (Journal of the Institute of Mathematics and its Applications, 7 (1971), 21–36.

[16] M.J.D. Powell, *Some global convergence properties of a variable metric algorithm for minimization without exact line searches*, in Nonlinear Programming, SIAM-AMS Proc. IX, R. W. Cottle and C. E. Lemke, eds., AMS, Providence, RI, 1976, 53–72.

[17] Ph.L. Toint, *Global convergence of the partitioned BFGS algorithm for convex partially separable optimization*, Math. Program., 36 (1986), 290–306.

[18] P. Wolfe, *Methods of nonlinear programming*, (Nonlinear Programming, ed. J. Abadie, Interscience, Wiley, New York, 1967, 97–131.

Accepted: 29.05.2017