Ilorin Journal of Science
Volume 6, Number 1, 2019, pp. 25 – 41 (Printed in Nigeria)
ISSN: 2408 – 4840 © 2019 Faculty of Physical Sciences, University of Ilorin
https://doi.org/10.54908/iljs.2019.06.01.003

ILORIN
JOURNAL OF SCEINCE

ILJS-19-003

A Nonlinear Conjugate Gradient Algorithm under Strong Wolfe-Powell Line Search for Large Scale Unconstrained Optimization Problems

Ejieji*, C. N., Daniel, M. S., Aderinto, Y. O. and Akinsunmade, A. E.

Department of Mathematics, University of Ilorin, Ilorin, Nigeria.

Abstract

The conventional conjugate gradient method solves linear and quadratic optimization problems but most real life problems consist of nonquadratic functions of several variables. In this work a nonlinear conjugate gradient algorithm for solving large scale optimization problems is presented. The new algorithm is a modification of the Fletcher-Reeves conjugate gradient method and it is proved to achieve global convergence under the strong Wolfe-Powell inexact line search technique. Computational experiments show that the new algorithm presented performs better than the Fletcher-Reeves exact line search algorithm in solving high dimensional nonlinear optimization problems.

Keyword: unconstrained optimization, conjugate gradient, inexact line search, sufficient descent condition.

1. Introduction

An unconstrained nonlinear optimization problem is of the form:

$$\underset{x \in R^n}{minimize} \quad f(x), \tag{1}$$

where $f : \mathrm{R}^n \to \mathrm{R}$ is a smooth and real valued nonlinear objective function of the vector $x \in \mathrm{R}^n$.Various optimization techniques have been developed for the solution of problems of the form (1), but the search for new and more efficient ones is an unending one. The most popular among these methods is the Conjugate Gradient Method (CGM). The linear conjugate gradient method was originally proposed by Hestenes and Stiefel in 1952, for solving symmetric positive definite systems of equations while in 1964, based on the idea of the linear

*Corresponding Author: Ejieji, C. N.
Email: ejieji.cn@unilorin.edu.ng

conjugate gradient method, Fletcher and Reeves gave a nonlinear conjugate gradient method for solving unconstrained optimization problems (Yarushi and Hiroshi, 2014). Nonlinear conjugate gradient methods are currently considered to be the most important techniques for solving large scale unconstrained optimization problems. The popularity of these methods are mainly due to their efficiency in solving large scale problems and their simplicity both in their algebraic expression and in their ease of implementation in computer codes (Ejieji and Bamigbola, 2006; Neculai, 2011).

Consider the general unconstrained optimization problem:

$\underset{x\in R^n}{minimize}\ \ f(x)$, where $f:\mathrm{R}^n \to \mathrm{R}$ is continuously differentiable and its gradient is denoted by $g(x) = \nabla f(x)$.

The conjugate gradient algorithm generates a sequence of iterates according to

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

for values of $k = 0,\ 1,\ 2,\ \ldots$, where x_k is the current iteration point and $\alpha_k > 0$ is the step length obtained by some line search, d_k is the search direction and is defined by :

$$d_k = \begin{cases} -g_k, & k = 0; \\ -g_k + \beta_k d_{k-1}, & k \geq 1. \end{cases}, \tag{3}$$

where $g_k = g(x_k)$ and β_k is a scalar parameter known as the conjugate gradient coefficient. Several conjugate gradient methods have been proposed , and they mainly differ in the choice of the scalar parameter β_k. Some of the formulae for β_k as reported in (Ahmad and Zabidin, 2017) are: the Hestenes-Stiefel (HS), the Fletcher-Reeves (FR), the Polak-Ribiere-polyak (PRP), the conjugate descent (CD), the Liu-Storey (LS), and the Dai-Yuan (DY) conjugate gradient coefficients. These choices of β_k , where $y_{k-1} = g_k - g_{k-1}$ are expressed as follows:

- $\beta_k^{(HS)} = \dfrac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}$ by Hestenes and Stiefel,
- $\beta_k^{(FR)} = \dfrac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$ by Fletcher and Reeves ,

- $\beta_k^{(PRP)} = \dfrac{g_k^T y_{k-1}}{g_{k-1}^T g_{k-1}}$ by Polak, Ribeere and Polyak,

- $\beta_k^{(CD)} = -\dfrac{g_k^T g_k}{d_{k-1}^T g_{k-1}}$ by Fletcher,

- $\beta_k^{LS} = -\dfrac{g_k^T y_{k-1}}{d_{k-1}^T g_{k-1}}$ by Liu and Storey,

- $\beta_k^{(DY)} = \dfrac{g_k^T g_k}{d_{k-1}^T y_{k-1}}$ by Dai and Yuan .

The step length $\alpha_k > 0$ is computed by performing a line search along the search direction d_k. However, for non linear problems, it is more appropriate and cost efficient to adopt some inexact line search procedure.

Abdelrahman *et al.* (2017) proposed a new conjugate gradient method for unconstrained optimization problems with a parameter derived from the Fletcher Reeves conjugate parameter. The new method called 'A Modified Fletcher Reeves Conjugate Gradient Method' (AMFR CGM) was shown to possess the sufficient descent condition and to achieve global convergence properties under exact line search. The conjugate gradient coefficient of the method is given by:

$$\beta_k^{AMFR} = \frac{\| g_k \|^2 - \dfrac{\| g_k \| \left| g_k^T d_{k-1} \right|}{\| d_{k-1} \|}}{\mu \left| g_k^T d_k \right| + \| g_{k-1} \|^2} \quad , \tag{4}$$

where $\mu > 0$ and $\| . \|$ is the euclidean norm. Using equation (4), the following algorithm was developed:

Algorithm 1.1 (AMFR CGM with exact line search) (Abdelrahman *et al.* , 2017)

- **Step 1:** Given $x_0 \in \mathbb{R}^n$, $\varepsilon = 10^{-6}$, set $d_0 = -g_0$, if $\| g_k \| \le \varepsilon$ then stop.

- **Step 2:** Compute α_k by applying exact line search, that is by using

$$\alpha_k = \alpha^* = \operatorname{argmin} f(x_k + \alpha d_k), \qquad \alpha_k > 0 .$$

- **Step 3:** Set $x_{k+1} = x_k + \alpha_k d_k$, if $\| g_{k+1} \| \le \varepsilon$ then stop.
- **Step 4:** Compute β_k^{AMHF} by (4) and generate d_{k+1} by (3).
- **Step 5:** Set $k = k+1$ and go to step 2.

Due to the fact that exact line search procedure is always very expensive and sometimes impossible especially for large dimensional problems, we apply the strong Wolfe-Powell inexact line search to the conjugate gradient parameter β_k proposed by Abdelrahman's algorithm to generate a new algorithm for solving large scale unconstrained optimization problems.

The Strong Wolfe-Powell inexact line search conditions are :

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k \qquad (5)$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \le \sigma \left| g_k^T d_k \right|, \qquad (6)$$

where $0 < \delta < \sigma < 1$ and d_k is a search direction (Mohamed *et al.* , 2016). We also investigated the convergence properties of our algorithm. The new method is proved to possess global convergence property. Numerical experiments with some standard test problems showed that the new algorithm performs better than the Fletcher-Reeves method.

2. Materials and Methods

The New Nonlinear Conjugate Gradient Algorithm with inexact line search

We consider the nonlinear unconstrained optimization problem given in (1), where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and its gradient is denoted by $g(x) = \nabla f(x)$. Using an iterative scheme of the form $x_{k+1} = x_k + \alpha_k d_k$, where the search direction d_k is defined by:

$$d_{k+1} = \begin{cases} -g_{k+1}, & k = 0; \\ -g_{k+1} + \left(\dfrac{\| g_{k+1} \|^2 - \dfrac{\| g_{k+1} \| \left| g_{k+1}^T d_k \right|}{\| d_k \|}}{\mu \left| g_{k+1}^T d_{k+1} \right| + \| g_k \|^2} \right) d_k, & k \ge 1. \end{cases} \qquad (7)$$

where $\mu > 0$ and $\|.\|$ is the euclidean norm , we propose a nonlinear conjugate gradient algorithm developed under the Strong Wolfe-Powell line search as follows :

Algorithm 2.1

A new nonlinear Conjugate Gradient Method (With Strong Wolfe-Powell Inexact line search)

• **Step 1:** Choose an initial point $x_0 \in R^n$, $\varepsilon \in (0,1)$, $\mu > 0$, $\delta \in (0, \frac{1}{2})$, $\sigma \in (\delta, 1)$ set $k = 0$ and $d_0 = -g_0$.

• **Step 2:** If $\| g_k \| \leq \varepsilon$ then stop.

• **Step 3:** Find the step length α_k by Strong Wolfe-Powell line search to satisfy:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k$$

and

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq \sigma \left| g_k^T d_k \right|.$$

• **Step 4:** Set a new iteration point $x_{k+1} = x_k + \alpha_k d_k$.

• **Step 5:** If $\| g_{k+1} \| \leq \varepsilon$ then stop, otherwise go to the next step.

• **Step 6:** Update the search direction by:

$$d_{k+1} = \begin{cases} -g_{k+1}, & k = 0; \\ -g_{k+1} + \left(\dfrac{\| g_{k+1} \|^2 - \dfrac{\| g_{k+1} \| \left| g_{k+1}^T d_k \right|}{\| d_k \|}}{\mu \left| g_{k+1}^T d_{k+1} \right| + \| g_k \|^2} \right) d_k, & k \geq 1. \end{cases}$$

• **Step 7:** Set $k = k + 1$ and go to step 3.

2.2 Some Convergence Properties of Algorithm 2.1

We now investigate the Sufficient descent property, trust region feature and global convergence properties of Algorithm 2.1.

Sufficient Descent Property and Trust Region Feature

We show these by proving the following Lemma:

Lemma 2.1

If the search direction d_{k+1} meets the condition:

$$d_{k+1} = \begin{cases} -g_{k+1}, & k = 0; \\ -g_{k+1} + \left(\dfrac{\| g_{k+1} \|^2 - \dfrac{\| g_{k+1} \| \left| g_{k+1}^T d_k \right|}{\| d_k \|}}{\mu \left| g_{k+1}^T d_{k+1} \right| + \| g_k \|^2} \right) d_k, & k \geq 1. \end{cases}$$

Then:

(i) the Sufficient Descent Property,

$$g_{k+1}^T d_{k+1} \leq -\| g_{k+1} \|^2 \tag{8}$$

and

(ii) the Trust Region Feature

$$\| d_{k+1} \| \quad \leq \quad \| g_{k+1} \| \tag{9}$$

will hold.

Proof:

From (7), if $k = 0$ we have:

$d_{k+1} = -g_{k+1}$, multiplying both sides by g_{k+1}^T, gives: $g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1}$.

Since $k = 0$, we have : $g_1^T d_1 = -g_1^T g_1$, $g_1^T d_1 = -\| g_1 \|^2$.

Hence condition (8) is satisfied when $k = 0$. Also when $k \geq 0$, we have

$$d_{k+1} = -g_{k+1} + \left(\dfrac{\| g_{k+1} \|^2 - \dfrac{\| g_{k+1} \| \left| g_{k+1}^T d_k \right|}{\| d_k \|}}{\mu \left| g_{k+1}^T d_{k+1} \right| + \| g_k \|^2} \right) d_k .$$

Multiplying both sides by g_{k+1}^T, we get:

$$g_{k+1}^T d_{k+1} = g_{k+1}^T\left(-g_{k+1} + \left(\frac{\| g_{k+1}\|^2 - \dfrac{\| g_{k+1}\| \left|g_{k+1}^T d_k\right|}{\| d_k \|}}{\mu\left|g_{k+1}^T d_{k+1}\right| + \| g_k \|^2}\right)d_k\right)$$

which gives: $g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1} + g_{k+1}^T\left(\dfrac{\| g_{k+1}\|^2 - \dfrac{\| g_{k+1}\| \left|g_{k+1}^T d_k\right|}{\| d_k \|}}{\mu\left|g_{k+1}^T d_{k+1}\right| + \| g_k \|^2}\right)d_k$.

Therefore

$$g_{k+1}^T d_{k+1} = -\| g_{k+1}\|^2 + \left(\frac{g_{k+1}^T\| g_{k+1}\|^2 - \dfrac{g_{k+1}^T\| g_{k+1}\| \left|g_{k+1}^T d_k\right|}{\| d_k \|}}{\mu\left|g_{k+1}^T d_{k+1}\right| + \| g_k \|^2}\right)d_k,$$

from which we get $g_{k+1}^T d_{k+1} = -\| g_{k+1}\|^2 + \dfrac{g_{k+1}^T\| g_{k+1}\|^2\, d_k - \dfrac{g_{k+1}^T\| g_{k+1}\| \left|g_{k+1}^T d_k\right| d_k}{\| d_k \|}}{\mu\left|g_{k+1}^T d_{k+1}\right| + \| g_k \|^2}$

or

$$g_{k+1}^T d_{k+1} \le -\| g_{k+1}\|^2 + \frac{\| g_{k+1}\| \;\; \| g_{k+1}\|^2 \| d_k \| - \dfrac{\| g_{k+1}\| \;\; \| g_{k+1}\| \;\; \| g_{k+1} d_k \| \;\; \| d_k \|}{\| d_k \|}}{\mu \| g_{k+1} d_{k+1} \| + \| g_k \|^2}$$

$$\le -\| g_{k+1}\|^2 + \frac{\| g_{k+1}\| \| g_{k+1}\|^2 \| d_k \| - \dfrac{\| g_{k+1}\| \| g_{k+1}\| \| g_{k+1}\| \| d_k \| \| d_k \|}{\| d_k \|}}{\mu \| g_{k+1}\| \| d_{k+1}\| + \| g_k \|^2}$$

$$\le -\| g_{k+1}\|^2 + \frac{\| g_{k+1}\| \| g_{k+1}\|^2 \| d_k \| - \| g_{k+1}\| \| g_{k+1}\| \| g_{k+1}\| \| d_k \|}{\mu \| g_{k+1}\| \| d_{k+1}\| + \| g_k \|^2}.$$

This simplifies to

$$g_{k+1}^T d_{k+1} \le -\| g_{k+1}\|^2 + \frac{\| g_{k+1}\|^3 \| d_k \| - \| g_{k+1}\|^3 \| d_k \|}{\mu \| g_{k+1}\| \| d_{k+1}\| + \| g_k \|^2}$$

which implies that

$$g_{k+1}^T d_{k+1} \le -\| g_{k+1} \|^2 .$$

Therefore, condition (8) is satisfied for $k \ge 1$

To prove (9) we have: when $k = 0$:

$$d_{k+1} = -g_{k+1}$$

$$d_{0+1} = -g_{0+1}$$

$$d_1 = -g_1$$

$$\| d_1 \| = \| g_1 \|$$

When $k \ge 1$, we have: $d_{k+1} = -g_{k+1} + \left(\dfrac{\| g_{k+1} \|^2 - \dfrac{\| g_{x+1} \| \left| g_{k+1}^T d_k \right|}{\| d_k \|}}{\mu \left| g_{k+1}^T d_{k+1} \right| + \| g_k \|^2} \right) d_k$

$$d_{k+1} = -g_{k+1} + \left(\frac{\| g_{k+1} \|^2 \| d_k \| d_k - \| g_{k+1} \| \left| g_{k+1}^T d_k \right| d_k}{\| d_k \| \left(\mu \left| g_{k+1}^T d_{k+1} \right| + \| g_k \|^2 \right)} \right),$$

$$\| d_{k+1} \| = \| -g_{k+1} + \left(\frac{\| g_{k+1} \|^2 \| d_k \| d_k - \| g_{k+1} \| \left| g_{k+1}^T d_k \right| d_k}{\| d_k \| \left(\mu \left| g_{k+1}^T d_{k+1} \right| + \| g_k \|^2 \right)} \right) \| .$$

Therefore

$$\| d_{k+1} \| \le \| g_{k+1} \| + \| \left(\frac{\| g_{k+1} \|^2 \| d_k \| d_k - \| g_{k+1} \| \left| g_{k+1}^T d_k \right| d_k}{\| d_k \| \left(\mu \left| g_{k+1}^T d_{k+1} \right| + \| g_k \|^2 \right)} \right) \|,$$

$$\le \| g_{k+1} \| + \| \frac{\| g_{k+1} \|^2 \| d_k \| d_k}{\| d_k \| \left(\mu \left| g_{k+1}^T d_{k+1} \right| + \| g_k \|^2 \right)} \| - \| \frac{\| g_{k+1} \| \left| g_{k+1}^T d_k \right| d_k}{\| d_k \| \left(\mu \left| g_{k+1}^T d_{k+1} \right| + \| g_k \|^2 \right)} \|,$$

$$\le \| g_{k+1} \| + \frac{\| g_{k+1} \|^2 \| d_k \| \| d_k \|}{\| d_k \| \left(\mu \| g_{k+1} d_{k+1} \| + \| g_k \|^2 \right)} - \frac{\| g_{k+1} \| \| g_{k+1} d_k \| \| d_k \|}{\| d_k \| \left(\mu \| g_{k+1} d_{k+1} \| + \| g_k \|^2 \right)},$$

$$\le \| g_{k+1} \| + \frac{\| g_{k+1} \|^2 \| d_k \|}{\left(\mu \| g_{k+1} d_{k+1} \| + \| g_k \|^2 \right)} - \frac{\| g_{k+1} \| \| g_{k+1} \| \| d_k \|}{\left(\mu \| g_{k+1} d_{k+1} \| + \| g_k \|^2 \right)},$$

$$\le \| g_{k+1} \| + \frac{\| g_{k+1} \|^2 \| d_k \|}{\left(\mu \| g_{k+1} d_{k+1} \| + \| g_k \|^2 \right)} - \frac{\| g_{k+1} \|^2 \| d_k \|}{\left(\mu \| g_{k+1} d_{k+1} \| + \| g_k \|^2 \right)}.$$

Hence

$$\| d_{k+1} \| \quad \leq \quad \| g_{k+1} \| \tag{10}$$

Global Convergence Properties

The following assumptions and lemma are needed to investigate the global convergence of Algorithm 2.1

Assumption A (Mohamed *et al*., 2016)

The level set $\omega = \{ x \in \mathbb{R}^n \mid f(x) \leq f(x_0) \}$ is bounded, where x_0 is the starting point.

Assumption B (Mohamed *et al*., 2016)

In some neighborhood N of ω, the objective function is continuously differentiable and its gradient is lipschitz continuous, namely there exist a constant $L > 0$ such that $\| g(x) - g(y) \| \leq L \| x - y \|$ for any $x, y \in N$.

We now conclude the proof of global convergence by showing that Algorithm 2.1 satisfies the Zoutendijk condition (Zoutendijk, 1970) and that $\lim_{k \to \infty} \| g_k \| = 0$. We achieve these by proving Lemma 2.2 and Theorem 2.1

Lemma 2.2 (Zoutendijk condition)

Suppose Assumptions A and B hold. Consider any conjugate gradient method of the form

$$x_{k+1} = x_k + \alpha_k d_k \text{ and } d_k = \begin{cases} -g_k, & k = 0; \\ -g_k + \beta_k d_{k-1}, & k \geq 1. \end{cases},$$

where d_k satisfies $g_k^T d_k < 0$ for all k and α_k is obtained by the strong Wolfe-Powell line search $f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k$ and $\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq \sigma \left| g_k^T d_k \right|$ then:

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\| d_k \|^2} < +\infty$$

Proof:

From (6) , $\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq \sigma \left| g_k^T d_k \right|$, substracting $g_k^T d_k$ from both sides we obtain:

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| - g_k^T d_k \leq \sigma \left| g_k^T d_k \right| - g_k^T d_k .$$

Therefore:

$$\| g(x_k + \alpha_k d_k) - g(x_k) \| \quad \| d_k \| \leq -\sigma g_k^T d_k - g_k^T d_k ,$$

$$\| g(x_k + \alpha_k d_k) - g(x_k) \| \quad \| d_k \| \leq -(\sigma + 1) g_k^T d_k .$$

By Assumption A, we have:

$$L \| \alpha_k d_k \| \quad \| d_k \| \leq -(\sigma + 1) g_k^T d_k ,$$

$$L\alpha_k \| d_k \|^2 \leq -(\sigma + 1) g_k^T d_k ,$$

$$\alpha_k \leq \frac{-(\sigma + 1) g_k^T d_k}{L \| d_k \|^2} \tag{11}$$

Similarly, from (5)

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k ,$$

therefore

$$-\delta \alpha_k g_k^T d_k \leq f(x_k) - f(x_k + \alpha_k d_k) \tag{12}$$

And substituting (11) for α_k in (12) we have:

$$\frac{\delta(\sigma + 1)(g_k^T d_k)^2}{L \| d_k \|^2} \leq f(x_k) - f(x_k + \alpha_k d_k) . \tag{13}$$

Summing both sides of (13) over k, we have:

$$\sum_{k=0}^{\infty} \frac{\delta(\sigma + 1)(g_k^T d_k)^2}{L \| d_k \|^2} \leq \sum_{k=0}^{\infty} (f(x_k) - f(x_k + \alpha_k d_k)) < +\infty$$

and

$$\frac{\delta(\sigma + 1)}{L} \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\| d_k \|^2} \leq f(0) - f_\infty < +\infty \tag{14}$$

which Implies that : $\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\| d_k \|^2} < +\infty$.

This completes the proof.

Theorem 2.1

If assumptions A and B are satisfied and the relative sequences of x_k, d_k, g_k and α_k are generated by Algorithm 2.1 , then:

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

Proof:

Applying the sufficient descent condition (8) on (14) we obtain:

$$\frac{\delta(\sigma+1)}{L}\sum_{k=0}^{\infty}\frac{\| g_k \|^4}{\| d_k \|^2} \le f(0) - f_\infty < +\infty, \tag{15}$$

using (9) , (15) becomes

$$\frac{\delta(\sigma+1)}{L}\sum_{k=0}^{\infty}\frac{\| g_k \|^4}{\| g_k \|^2} \le f(0) - f_\infty < +\infty.$$

Therefore

$$\frac{\delta(\sigma+1)}{L}\sum_{k=0}^{\infty}\| g_k \|^2 \le f(0) - f_\infty < +\infty$$

which implies that

$$\lim_{k\to\infty} \| g_k \|^2 = 0.$$

Hence

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

This ends the proof.

3. Result and Discussion

Computational Consideration

The results obtained from the solution of some test problems are presented in this section. Codes are written in MATLAB *R(2007b)* and are run on a Windows 10 Operating System with Intel(R) Celeron(R) CPU N3060 $@1.60GHz$ and $4.00GB$ RAM.

Parameters used are: $\delta = 0.0001$, $\sigma = 0.9$ and $\mu = 0.5$. Large Scale dimensions of 5000 and 10000were used. The algorithm stops if either $\| g(x_i) \| < \varepsilon$ or the number of iteration is greater than 2000, where $\varepsilon = 10^{-6}$. The numerical results are presented in tables 3.1 and 3.2, where

- "ALGORITHM 2.1" refers to A Nonlinear Conjugate Gradient Method (With Strong Wolfe-Powell inexact line search)
- " FR ALGORITHM" means Fletcher Reeves Algorithm
- "PROBLEM NO." is computational experiment problem number
- "DIM" (n) is the problem dimension
- "NI"(k) is the number of iteration
- "F(x)" is the function value at the optimum point
- "Alpha" (α) is the step length at the optimum point
- "TIME"(t) is the system computational time in seconds
- "GNORM" ($\| g_i \|$) means the norm of g at any point i

The following standard problems from (Neculai, 2008), were used to test the performance of our new algorithm (Algorithm 2.1) :

Problem 1 (Extended Block Diagonal BDI Function) (Neculai, 2008)

$$f(x) = \sum_{i=1}^{\frac{n}{2}} \left\{ (x_{2i-1}^2 - x_{2i}^2 - 2)^2 + (\exp(x_{2i-1} - 1) - x_{2i})^2 \right\}$$

$$x_0 = [0.1, 0.1, ...0.1, 0.1]^T$$

Problem 2 (DiagonaL 4 Function) (Neculai, 2008)

$$f(x) = \sum_{i=1}^{\frac{n}{2}} \frac{1}{2} (x_{2i-1}^2 - c x_{2i}^2)$$

$$x_0 = [1, 1, ..., 1, 1]^T$$

$$c = 100$$

Problem 3 (Generalize Rosebrock Function) (Neculai, 2008)

$$f(x) = \sum_{i=1}^{n-1}\left\{ \; c(x_{i+1} - x_i^2) + (1 - x_i)^2 \; \right\}$$

$$x_0 = [-1.2, 1, ..., -1.2, 1]^T$$

$$c = 100$$

Problem 4 (Extended Himmelbau Function) (Neculai, 2008)

$$f(x) = \sum_{i=1}^{\frac{n}{2}}\left\{ \; (x_{2i}^2 + x_{2i} - 11)^2 + (x_{2i-1} + x_{2i}^2 - 7)^2 \; \right\}$$

$$x_0 = [1, 1, ..., 1, 1]^T$$

Problem 5 (Diagonal 5 Function) (Neculai, 2008)

$$f(x) = \sum_{i=1}^{n}\left\{ \; \log\big(\exp(x_i) + \exp(-x_i)\big) \; \right\}$$

$$x_0 = [1.1, 1.1, ..., 1.1, 1.1]^T$$

Problem 6 (Extended Rosebrock Function) (Neculai, 2008)

$$f(x) = \sum_{i=1}^{\frac{n}{2}}\left\{ \; c(x_{2i} - x_{2i-1}^2)^2 + (1 - x_{2i-1})^2 \; \right\}$$

$$x_0 = [-1.2, 1, ..., -1.2, 1]^T$$

$$c = 100$$

Problem 7 (Generalize White and Holst Function) (Neculai, 2008)

$$f(x) = \sum_{i=1}^{n-1}\left\{ \; c(x_{i+1} - x_i^3)^2 + (1 - x_i)^2 \; \right\}$$

$$x_0 = [-1.2, 1, ..., -1.2, 1]^T$$

$$c = 100$$

Problem 8 (Extended Quadratic Penalty QPI Function) (Neculai, 2008)

$$f(x) = \sum_{i=1}^{n-1}(x_i^2 - 2)^2 + \sum_{i=1}^{n}(x_i - 0.5)^2$$

$$x_0 = [1,1,...,1,1]^T$$

Problem 9 (Extended Beale Function) (Neculai, 2008)

$$f(x) = \sum_{i=1}^{\frac{n}{2}} (1.5 - x_{2i-1}(1 - x_{2i}))^2 + (2.25 - x_{2i-1}(1 - x_{2i}^2))^2 + (2.625 - x_{2i-1}(1 - x_{2i}^3))^2$$

$$x_0 = [1,0.8,...,1,0.8]^T$$

3.2 Computational Results

Table 3.1: Results obtained with the new algorithm (Algorithm 2.1).

	ALGORITHM 2.1					
PROBLEM NO.	**DIM (n)**	**No. of Iterations N I**	**Function Value $F(x)$**	**Alpha (α)**	**TIME (t)**	**GNORM ($\Vert g_i \Vert$)**
1	5000	60	1.670471e-13	6.250000e-02	1. 816	7.13e-07
	10000	57	6.506764e-15	6.250000e-02	2. 200	2.28e-07
2	5000	15	6.658698e-16	3.906250e-03	1.676	7.57e-07
	10000	25	6.198532e-16	3.906250e-03	2.570	7.31e-07
3	**5000**	**36**	**3.986624e+00**	9.765625e-04	1.710	9.03e-07
	10000	**36**	**3.986624e+00**	9.765625e-04	2.296	9.03e-07
4	5000	27	1.166950e-15	1.562500e-02	1.162	4.34e-07
	10000	27	2.333900e-15	1.562500e-02	1.398	6.13e-07
5	**5000**	**13**	**5.000000e+03**	**1.000000e+00**	**0.546**	**6.91e-07**
	10000	**13**	**1.000000e+04**	**1.000000e+00**	**0.595**	**9.77e-07**
6	5000	99	2.923226e-16	1.953125e-03	3.826	7.52e-07
	10000	97	3.584847e-17	1.953125e-03	4.626	5.24e-07
7	5000	215	4.041508e-16	4.882813e-04	10.694	4.49e-07
	10000	215	4.041508e-16	4.882813e-04	17.915	4.49e-07
8	**5000**	**2000**	**2.060031e+04**	6.938894e-18	241.029	3.15e+00
	10000	**2000**	**4.000010e+04**	6.938894e-18	390.136	1.41e+00
9	**5000**	102	**1.709004e-13**	3.125000e-02	9.169	7.29e-07
	10000	130	**3.620310e-13**	3.125000e-02	20.406	6.54e-07

Table 3.2: Results obtained with Fletcher Reeves Algorithm.

	Fletcher Reeves (FR) **Algorithm**					
PROBLEM NO.	**DIM (n)**	**No. of Iterations N I**	**Function Value $F(x)$**	**Alpha (α)**	**TIME (t)**	**GNORM ($\|g_i\|$)**
1	5000	2000	5.240000e-02	7.630000e-06	50.062	8.35e-01
	10000	97	5.190021e-14	6.250000e-02	4.157	6.28e-07
2	5000	25	1.201032e-15	3.906250e-03	1.237	9.73e-07
	10000	27	4.010644e-16	7.806250e-03	0.700	5.64e-07
3	**5000**	82	**3.986624e+00**	2.440025e-04	1.826	5.47e-07
	10000	82	**3.986624e+00**	2.440025e-04	2.930	5.47e-07
4	5000	80	1.140343e-14	7.806250e-03	1.657	8.96e-07
	10000	89	8.270451e-15	1.561250e-02	2.675	6.98e-07
5	**5000**	97	**5.000000e+03**	1.000000e+00	6.486	2.50e-07
	10000	97	**1.000000e+04**	1.000000e+00	10.117	3.53e-07
6	5000	2000	3.440204e+04	4.241054e-22	71.097	4.04e+04
	10000	2000	6.895641e+04	1.685527e-21	115.703	5.70e+04
7	5000	2000	7.380303e-05	1.530960e-05	63.883	4.80e-01
	10000	2000	7.380353e-05	1.530960e-05	90.060	4.80e-01
8	**5000**	**2000**	**2.023659e+04**	6.780418e-21	271.726	2.18e+00
	10000	**2000**	**4.038028e+04**	6.780418e-21	418.847	6.07e+00
9	**5000**	**129**	**1.709445e-13**	1.562500e-02	4.587	3.23e-07
	10000	**141**	**3.670684e-13**	6.250000e-02	5.220	7.80e-07

3.3 Discussion on Numerical Results

From the results presented in tables 3.1 and 3.2, it can be observed that the new nonlinear conjugate gradient algorithm 2.1 proposed in this work is more effective than the Fletcher-Reeves algorithm in terms of lower number of iterations and reduced computational time. Also table 3.2 shows that the Fletcher-Reeves algorithm did not solve problems 6, 7, and 8 completely but stopped because it has performed the maximum number of iterations.

4. **Conclusion**

A nonlinear conjugate gradient algorithm with the strong Wolfe-Powell inexact line search technique, for solving large scale optimization problems was presented. The algorithm was shown to possess the sufficient descent property and to be globally convergent. Computational experiments illustrated that algorithm 2.1 presented in this work is efficient and performs better than the Fletcher Reeves algorithm. Hence the new algorithm is recommended for the solution of nonlinear large scale optimization problems.

References

Abdelrahman, A., Mustafa, M., Ahmad, A., Fatma, S., Zabindin, S. and Zahrahtul, A. Z. (2017): A modified Fletcher-Reeves conjugate gradient method for unconstrained optimization. *International Journal of Science Engineering and Technology*. **5** (1), 1-4.

Ahmad, A. and Zabidin, S. (2017): Modification of nonlinear conjugate gradient method with weak Wolfe-powell line search. *Abstract and Applied Analysis*. **2017**, 1-6.

Ejieji, C. N. and Bamigbola, O. M. (2006): A conjugate gradient method for solving discrete optimal control problems. *Journal of Mathematical Association of Nigeria (ABACUS)*. **33(2B)**, 318 – 329.

Hamoda, M., Mamat, M., Rivaie, M. and Salleh, Z. (2016): A Conjugate gradient method with strong Wolfe-Powell line search for unconstrained optimization. *Hikari: Applied Mathematical Sciences*, **10** (15), 721 – 734.

Neculai, A. (2008): An unconstrained optimization test functions collection, *Advanced Modeling and Optimization*. **10**, 147 – 161.

Neculai, A. (2011): Open problems in nonlinear conjugate gradient algorithm for unconstrained optimization. *Bulletin of the Malaysian Mathematical Society*. **2**, 319 – 330.

Yarushi, N. and Hirishi, Y. (2014): A survey of sufficient descent conjugate gradient methods for unconstrained optimization. *SUT journal of mathematics*. **50** (2), 167 – 203.

Zoutendijk, G. (1970): Nonlinear programming, computational methods. In Integer and Nonlinear Programming. Abadie, J., Ed., North-Holland, Amsterdam, 37 – 86.