

A New Separable Logarithmic Algorithm for Non-Linear Optimization

¹Abbas Y. Al-Bayati and ²Hawraz N. Jabbar

¹Department of Mathematics, College of Computers Sciences and Mathematics,
University of Mosul, Iraq

²Department of Mathematics, College of Education,
Suleim University, Iraq

Abstract: Problem statement: The idea of this study stemmed from the fact that most of the currently used optimization algorithms use a local quadratic representation of the objective function. It also arisen from the fact that the objective function may not be represented adequately by quadratic functions and the global minimizer may be obtained for objective functions. So, in this study, we generalized the field of quadratic model into the field of the non-quadratic model. **Approach:** A new non-quadratic model was suggested for solving unconstrained optimization problems, which modified the classical Conjugate Gradient (CG) algorithm by scaling the standard quadratic model. **Results:** The new algorithm was derived and evaluated theoretically and numerically for some standard well-known and effective test functions. The results, in general, indicated that the new algorithm had improvements on different well-known algorithms used in this study. **Conclusion:** The new proposed algorithm would be generic and easy to implement in all gradient based optimization process. Its simulation results showed that it was robust and had a potential significantly enhance the computational efficiency of iterations and function evaluations.

Key words: Unconstrained minimization, conjugate-gradient algorithm, rational models, global convergence, exact line searches

INTRODUCTION

This study considers the calculation of a local minimizer x^* say, for the problem

$$\text{Min}f(x), \quad x \in R_n, \quad f \in C^2 \quad (1)$$

It is assumed that expressions for the elements of the gradient vector $g(x) = \text{gradient of } f(x)$ are available are calculated but the Hessian matrix is not available. Standard algorithms for solving this problem include CG-algorithms which are iterative algorithms and generate a sequence of approximations to minimize a function $f(x)$. This type of algorithms requires $4n$ locations of computer storage to implement. However, this study considers a more general model than the usual quadratic function

$$F(x) = \frac{1}{2}x^T Gx + b^T x \quad (2)$$

Definition 1: A set of vectors d_i^* in E^n is said to be conjugate with respect to the symmetric positive definite matrix G if and only if $d_i^T G d_j = 0$ for $i \neq j$.

Hestenes and Stiefel (1952) published the first CG-algorithm, for solving a system of linear algebraic equations. Fletcher and Reeves, (1964) were the first, among other scholars, to use this technique to minimize a non-linear function of several variables. Since then this algorithm has been used successfully to tackle many problems.

We can list outline of the CG-algorithm as follows:

Outline of the standard CG-algorithm: for given TOL and $x_0 \in R^n$ an initial estimate of the minimizer x^* :

- Step 1: Set $d_0 = -g_0$
- Step 2: For $i = 1, 2, \dots$, compute $x_i = x_{i-1} + \lambda_{i-1} d_{i-1}$ where λ_{i-1} is the optimal step-size obtained by a line search procedure.
- Step 3: If $\|g_i\| \leq \text{Tol}$ stop; otherwise continue.
- Step 4: Calculate the new direction $d_i = -g_i + \beta_i d_{i-1}$ where β_i is the conjugacy coefficient.

It is considered as one of the following formulas:

Corresponding Author: Abbas Y. Al Bayati, Department of Mathematics, College of Computers Sciences and Mathematics, Mosul University, Iraq

$$\beta_i = g_i^T (g_i - g_{i-1}) / g_{i-1}^T (g_i - g_{i-1}) \quad (3a)$$

Called: Hestenes and Stiefel (HS):

$$\beta_i = \|g_i\|^2 / \|g_{i-1}\|^2 \quad (3b)$$

Called: Fletcher and Reeves (FR):

$$\beta_i = g_i^T (g_i - g_{i-1}) / (g_{i-1}^T g_{i-1}) \quad (3c)$$

Called: Polak and Ribiere (PR):

$$\beta_i = \|g_i\|^2 / (d_{i-1}^T g_{i-1}) \quad (3d)$$

Called: Descent Dixon (DD):

$$\beta_i = -g_i^T (g_i - g_{i-1}) / d_{i-1}^T g_{i-1} \quad (3e)$$

Called: Al-Bayati and Al-Assady (BA).

For the details of these formulas (Fletcher, 1987; Al-Bayati and Al-Assady, 1994).

When, quadratic functions and exact line searches are used, all the above formulas for β_i 's are equivalent.

However, these formulas vary according to general functions.

Several algorithms have been suggested as alternative ways of modifying the classical CG-algorithms (Al-Bayati, 1993; 2001; 2007; Al-Bayati and Al-Assady, 1994; 1997; Al-Bayati and Al-Naemi, 1995; Al-Bayati and Rassam, 2009).

CG-algorithms have, in general, the following:

BASIC properties:

- The conjugacy condition
- The orthogonally condition
- The descent direction
- The quadratic termination condition with Exact Line Search (ELS)

MATERIALS AND METHODS

Extended CG-algorithms for non-quadratic models:

In this study, a more general model than quadratic one is suggested as a basis for a CG-algorithm. If $q(x)$ is a quadratic function, then a function $F(q(x))$ is defined as a non-linear scaling of $q(x)$ if the following condition holds:

$$f(x) = F(q(x)), \quad \frac{df}{dq} = f' > 0 \quad \text{and} \quad q(x) > 0 \quad (4)$$

where, x^* is the minimizer of $q(x)$ with respect to x , (Spedicato, 1976).

The following SCALING properties for $f(x)$ are immediately derived from the above condition:

- Every contour line of $q(x)$ is a contour line of $f(x)$
- If x^* is a minimizer of $q(x)$, then it is a minimizer of $f(x)$
- If x^* is a local minimizer of $q(x)$, then it is a local minimizer of $f(x)$

Boland *et al.* (1979) was the first who observed that $q(x)$ and $f(q(x))$ have determined the same search directions so that the finite termination property for their algorithm was satisfied. A CG-algorithm which minimizes the function

$$f(x) = q(q(x))^p, \quad p > 0 \quad \text{and} \quad x \in \mathbb{R}^n \quad (5a)$$

in at most n step was described by (Fried, 1971).

The special case:

$$F(q(x)) = \epsilon_1 q(x) + \frac{1}{2} \epsilon_2 q^2(x) \quad (5b)$$

where, ϵ_1 and ϵ_2 are scalars, was investigated by (Boland *et al.*, 1979).

(Tassopoulos and Storey, 1984) proposed two different rational models defined by:

$$F(q(x)) = \frac{(\epsilon_1 q(x) + 1)}{\epsilon_2 q(x)}, \quad \epsilon_2 < 0 \quad \text{and} \quad q(x) > 0 \quad (5c)$$

where, ϵ_1 and ϵ_2 are scalars and:

$$F(q(x)) = \frac{\epsilon q(x)}{(1 + q(x))}, \quad \epsilon > 0 \quad \text{and} \quad q(x) > 0 \quad (5d)$$

Al-Bayati (1993) proposed another rational model defined by:

$$F(q(x)) = \frac{\epsilon_1 q(x)}{1 - \epsilon_2 q_2(x)}, \quad \epsilon_1 > 0 \quad \text{and} \quad \epsilon_2 < 0 \quad (5e)$$

In this study, a new logarithmic model is investigated and tested on a set of several standard test functions, on the assumption that condition (1) holds. An extended CG-algorithm is developed which is based on this new model and scales $q(x)$ by the natural log function for the rational $q(x)$ functions:

$$F(q(x)) = \log \left(\frac{\varepsilon_1 q(x)}{1 - \varepsilon_2 q(x)} \right), \quad \varepsilon_2 < 0 \quad (6)$$

We first observe that $q(x)$ and $F(q(x))$ given by (6) have identical contours, though with different function values and they have the same unique minimum point denoted by x^* .

New ECG-algorithm for non-quadratic model: For any f satisfying the condition (4), it is shown in (Boland *et al.*, 1979) that the updating process given below generates identical conjugate directions and the same sequence of approximations x_i to the minimizer x^* , as does the original algorithm of (Fletcher and Reeves, 1964) when applied to $f(x) = q(x)$. Now let us modify the last scaling property of the function $f(x)$ by:

If x^* is a global minimizer of $q(x)$ implies that it is a global minimizer of $f(x)$, in the new extended logarithmic model defined in (6). Recall the basic theorem of (Renpu, 1989) which was illustrated below.

Suppose that $F(x)$ has the form:

$$F(x) = \frac{f_1(x_1)}{g_2(x_2)} \text{ where } x^T = (x_1^T, x_2^T) \quad (7)$$

and:

$$f_1(x_1) > 0 \text{ and } g_2(x_2) > 0 \quad (8)$$

It follows from (7) that:

$$\log F(x) = \log f_1(x_1) - \log g_2(x_2) \quad (9)$$

Equation 9 is a separable function according to the following theorem.

Theorem 1: The point $x^{*T} = (x_1^{*T}, x_2^{*T}, \dots, x_n^{*T})$ is a global minimizer of the rational separable function $F(x)$ if and only if every x_i^* ($i = 1, 2, \dots, n$) is a global minimizer of $f_i(x_i)$.

Proof: We can conclude that x^* is a global minimizer of $\log F(x)$ if and only if x_1^* and x_2^* are respectively global minimizer of $\log f_1(x_1)$ and $-\log g_2(x_2)$. Furthermore, the monotonicity of the log function implies that x^* is a global minimizer of $F(x)$ if and only if x_1^* and x_2^* are respectively global minimizer of $f_1(x_1)$ and $g_2(x_2)$ (Renpu, 1989).

Before listing the main outline of the new proposed algorithm, let us consider first the main outline of the

(Tassopoulos and Storey, 1984) and (Al-Bayati, 1993) rational model extended CG-algorithms.

ECG-algorithm (Tassopoulos and Storey, 1984):

Step 1: Define:

$$n = -(\lambda_{i-1} g_{i-1}^T d_{i-1}) / 2, \quad w = F_i - F_{i-1}, \\ F_i = F(q(x_i)) \text{ and compute } (n + w), \\ ((n + w)F_{i-1} - nw).$$

Step 2: If $|n + w| \leq EPS$, then set $\rho_i = 1$

and go to Step 5, where *EPS* is a small number (i.e., 0.1E-10).

Step 3: If $|(n + w)F_{i-1} - nw| \leq EPS$ then set

$$\rho_i = \sqrt{(F_{i-1} / F_i)} \text{ and go to Step 5}$$

Step 4: Compute $\rho_i = [n / (2n + w)]^2$

Step 5: Calculate the new direction

$$d_i = -g_i + \rho_i (\|g_i\|^2 / \|g_{i-1}\|^2) d_{i-1}$$

ECG-algorithm (Al-Bayati, 1993):

Step 1: Define:

$$a = (\lambda_{i-1} g_{i-1}^T d_{i-1}) / 2, \\ b = w - a, \\ c = wa - (w-a) * f_{i-1}$$

Step 2: If $|b| \leq Tol$ or $|c| \leq Tol$ then set $\rho_i = 1$

and go to Step 4, where *TOL* is a small number (i.e., 0.1E-10).

Step 3: Set $\rho_i = (a / w)^2$

Step 4: Calculate the new direction

$$d_i = -g_i + \rho_i (\|g_i\|^2 / \|g_{i-1}\|^2) d_{i-1}$$

Outline of the new ECG-algorithm:

Step 1: Set $d_0 = -g_0$

Step 2: For $i = 1, 2 \dots$

compute $x_i = x_{i-1} + \lambda_{i-1} d_{i-1}$ where λ_{i-1} is the optimal step-size obtained by Armijo linear search procedure: set scalars s_k, β and σ with:

$$s_k = -\frac{g_k^T d_k}{\|d_k\|}, \quad \beta \in (0, 1) \quad (9a)$$

and $\sigma \in (0, \frac{1}{2})$ and we set $\lambda_k = \beta^{m_k} s_k$, where m_k is the 1st non-negative integer m for which:

$$w = \exp(f_i) - \exp(f_{i-1}) \quad (9b)$$

i.e., $m= 0,1$ are tried successively until the inequality above is satisfied for $m = m_k$.

Step 3: Calculate f which is a function value of an actual iterate as:

$$\exp(f) = 1 + f + f^2/2! + \dots$$

and define:

$$n = \lambda_{i-1} g_{i-1}^T d_{i-1} / 2.$$

$$w = \exp(f_i) - \exp(f_{i-1}).$$

$$c = w - n \exp(f_{i-1})$$

Step 4: If $|w| < 0.1 \text{ E-}5$ or $|c| < 0.1 \text{ E-}5$, then set $\rho_i = 1.0$ and go to Step 6. Else continue

Step 5: Compute: $\rho_i^{NEW} = \frac{n^2 \cdot \exp(f_{i-1}) \cdot \exp(f_i)}{w^2}$

where, the derivation of scaling ρ_i will be presented later.

Step 6: Calculate the new direction $d_i = -g_i + \beta_i d_{i-1}$ where, β_i is defined by different modified formulas as follows:

$$\beta_i = \rho_i (\|g_i\|^2 / \|g_{i-1}\|^2) \quad (10a)$$

$$\beta_i = (g_i^T (\rho_i g_i - g_{i-1}) / g_{i-1}^T (g_i - g_{i-1})) \quad (10b)$$

$$\beta_i = (g_i^T (\rho_i g_i - g_{i-1}) / g_{i-1}^T g_{i-1}) \quad (10c)$$

$$\beta_i = \rho_i \|g_i\|^2 / d_{i-1}^T g_{i-1} \quad (10d)$$

$$\beta_i = -g_i^T (\rho_i g_i - g_{i-1}) / d_{i-1}^T g_{i-1} \quad (10e)$$

where, β_i in (10) are modified FR, HS, PR, DD and BA respectively. CG-algorithms are usually implemented by restarts in order to avoid an accumulation of errors affecting the search directions. It is therefore generally agreed that restarting is very helpful in practice, so we have used the following restarting criterion in our practical investigations. If the new direction satisfies:

$$d_i^T g_i \geq -0.8 \|g_i\|^2 \quad (11)$$

then a restart is also initiated and the new direction is sufficiently downhill.

The derivation of the new ECG-algorithm: The implementation of the extended CG-algorithm has been performed for general functions $F(q(x))$ of the form of (6). The unknown quantities ρ_i were expressed in terms of available quantities of the algorithm (i.e., function and gradient values of the objective function).

It is first assumed that neither ε_1 nor ε_2 is zero in Eq. 6. Solving (6) for $q(x)$, yields:

$$q(x) = \frac{\exp(f)}{\varepsilon_2 (\exp(f) + \varepsilon_1 / \varepsilon_2)} \quad (12)$$

and using the expression for ρ_i (henceforth ρ_i^{NEW})

$$\rho_i^{NEW} = \frac{f'_{i-1}}{f'_i} = \left(\frac{\exp(f_i)}{\exp(f_{i-1})} \right) \left(\frac{\exp(f_{i-1}) + \varepsilon_1 / \varepsilon_2}{\exp(f_i) + \varepsilon_1 / \varepsilon_2} \right)^2 \quad (13)$$

the quantity that has to be determined explicitly is $(\varepsilon_1 / \varepsilon_2)$.

During every interaction $(\varepsilon_1 / \varepsilon_2)$ must be evaluated as a function of known available quantities from the relation:

$$g_{i-1}^T = f'_{i-1} G(x_{i-1} - x^*) \quad (14)$$

$$g_{i-1}^T = f'_{i-1} G(x_{i-1} - x^*) \quad (15)$$

where, G is the Hessian Matrix and x^* is the minimum point. Now, we have:

$$\begin{aligned} g_{i-1}^T (x_i - x^*) &= g_{i-1}^T (x_{i-1} + \lambda_{i-1} d_{i-1} - x^*) \\ &= g_{i-1}^T (x_{i-1} - x^*) + \lambda_{i-1} g_{i-1}^T d_{i-1} \end{aligned}$$

$$g_i^T (x_{i-1} - x^*) = g_i^T (x_i - \lambda_{i-1} d_{i-1} - x^*) = g_i^T (x_i - x^*)$$

Since $g_i^T d_{i-1} = 0$, therefore, we can express ρ_i^{NEW} as follows:

$$\rho_i^{NEW} = (g_{i-1}^T (x_{i-1} - x^*) + \lambda_{i-1} g_{i-1}^T d_{i-1}) / (g_i^T (x_i - x^*)) \quad (16)$$

From (14) and (15), we get:

$$\rho_i^{NEW} = \frac{f'_{i-1} (x_{i-1} - x^*)^T G (x_{i-1} - x^*) + \lambda_{i-1} g_{i-1}^T d_{i-1}}{f'_i (x_i - x^*)^T G (x_i - x^*)}$$

Therefore:

$$\begin{aligned} \rho_i^{NEW} &= (2f'_{i-1} q_{i-1} + \lambda_{i-1} g_{i-1}^T d_{i-1}) / 2f'_i q_i \\ &= \rho_i^{NEW} (q_{i-1} / q_i) + \lambda_{i-1} g_{i-1}^T d_{i-1} / (2f'_i q_i) \end{aligned} \quad (17)$$

The quantities (q_{i-1} / q_i) and $f'_i q_i$ can be rewritten as:

$$(q_{i-1} / q_i) (1 / \sqrt{\rho_i^{NEW}}) \left(\sqrt{\exp(f_{i-1}) / \exp(f_i)} \right) \quad (18)$$

$$f_i' q_i = \frac{\exp(f_i) + \varepsilon_1 / \varepsilon_2}{\varepsilon_1 / \varepsilon_2} \quad (19)$$

Substituting (18) and (19) in (17), gives:

$$\rho_i^{NEW} = \left(\sqrt{\rho_i^{NEW}} \right) \left(\sqrt{\exp(f_{i-1}) / \exp(f_i)} \right) + \left(\lambda_{i-1} g_{i-1}^T d_{i-1} / 2 \right) \left(\frac{\varepsilon_1 / \varepsilon_2}{\exp(f_i) + \varepsilon_1 / \varepsilon_2} \right) \quad (20)$$

Using the transformation:

$$\lambda_{i-1} g_{i-1}^T d_{i-1} = 2n \quad (21)$$

and from RHS of (13) ; (20) and (21), we have:

$$\begin{aligned} & \left(\frac{\exp(f_i)}{\exp(f_{i-1})} \right) \left(\frac{\exp(f_{i-1}) + \varepsilon_1 / \varepsilon_2}{\exp(f_i) + \varepsilon_1 / \varepsilon_2} \right)^2 = \left(\sqrt{\frac{\exp(f_{i-1})}{\exp(f_i)}} \right) \\ & \left(\sqrt{\frac{\exp(f_i)}{\exp(f_{i-1})}} \right) \left(\frac{\exp(f_{i-1}) + \varepsilon_1 / \varepsilon_2}{\exp(f_i) + \varepsilon_1 / \varepsilon_2} \right) + n \left(\frac{\varepsilon_1}{\varepsilon_2} \right) \\ & \left(\frac{1}{\exp(f_i) + \varepsilon_1 / \varepsilon_2} \right) \\ & \left(\exp(f_{i-1}) + \varepsilon_1 / \varepsilon_2 \right)^2 = \left(\frac{\exp(f_{i-1})}{\exp(f_i)} \right) \quad (22) \\ & \left(\exp(f_{i-1}) + \varepsilon_1 / \varepsilon_2 \right) \left(\exp(f_i) + \varepsilon_1 / \varepsilon_2 \right) + \\ & \left(\exp(f_i) + \varepsilon_1 / \varepsilon_2 \right) + n \left(\varepsilon_1 / \varepsilon_2 \right) \\ & \frac{\exp(f_{i-1})}{\exp(f_i)} \left(\exp(f_i) + \varepsilon_1 / \varepsilon_2 \right) \\ & \frac{\varepsilon_1}{\varepsilon_2} = \frac{n \left(\exp(f_{i-1}) \right)^2 - \exp(f_{i-1}) w + n w \left(\exp(f_{i-1}) \right)}{w - n \exp(f_{i-1})} \end{aligned}$$

Where:

$$w = \exp(f_i) - \exp(f_{i-1}) \quad (23)$$

Substituting (22) in (13), yields:

$$\rho_i^{NEW} = \left(\frac{\exp(f_i)}{\exp(f_{i-1})} \right) \left(\frac{n \cdot \exp(f_{i-1})}{w} \right)^2 \quad (24a)$$

or:

$$\rho_i^{NEW} = \frac{n^2 \cdot \exp(f_{i-1}) \cdot \exp(f_i)}{w^2} \quad (24b)$$

Now we are prove the global convergence of the new extended CG-algorithm as follows: we first assume that:

Assumption 1: The function f has lower bound on $L_0 = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}$, where $x_0 \in \mathbb{R}^n$ is available.

Assumption 2: The gradient $g = \nabla f(x)$ is Lipschitz continuous in an open convex set B which contains L_0 , i.e., $\exists L > 0 \ni \|g(x) - g(y)\| \leq L \|x - y\|, \forall x, y \in B$.

Theorem 2: If Assumptions 1; 2 and $g_k^T d_k < 0$ holds, then for Armijo line search procedure the new extended CG-algorithm generates an infinite sequence $\{x_k\}$ that satisfy $\lim_{k \rightarrow \infty} \left(\frac{-g_k^T d_k}{\|d_k\|} \right)^2 = 0$.

Proof: For the new algorithm set $K_1 = \{k | \lambda_k = s_k\}, K_2 = \{k | \lambda_k < s_k\}$ by (9b) we have:

$$f_k - f_{k+1} \geq -s_k \sigma g_k^T d_k, \forall k \in K_1 \quad (25)$$

$$f_k - f_{k+1} \geq -\alpha_k \sigma g_k^T d_k, \forall k \in K_2 \quad (26)$$

By line search rule (9), since $\lambda_k / \beta \leq s_k, \forall k \in K_2$, we have:

$$f_k - f(x_k + \alpha_k / \beta d_k) < -\lambda_k \sigma g_k^T d_k / \beta, \forall k \in K_2$$

Using the mean value theorem on the LHS of the above inequality, there exists $\theta_k \in [0, 1]$ such that:

$$-\lambda_k g(x_k + \lambda_k \theta_k d_k / \beta)^T d_k / \beta < -\lambda_k \sigma g_k^T d_k / \beta, \forall k \in K_2$$

Therefore:

$$g(x_k + \theta_k \lambda_k d_k / \beta) > \sigma g_k^T d_k, k \in K_2 \quad (27)$$

By Assumption 2, Cauchy-Schwartz inequality and (27), we have:

$$\begin{aligned} & \lambda_k L \|d_k\|^2 / \beta \geq \|g(x_k + \lambda_k \theta_k d_k / \beta) - g_k\| \cdot \|d_k\| \\ & \geq (g(x_k + \lambda_k \theta_k d_k / \beta) - g_k)^T d_k \geq -(1 - \sigma) g_k^T d_k, k \in K_2 \end{aligned}$$

Thus:

$$\lambda_k \geq -\frac{\beta(1 - \sigma) g_k^T d_k}{L \|d_k\|^2}, k \in K_2 \quad (28)$$

It follows from (26) and (28):

$$f_k - f_{k+1} > \frac{\beta\sigma(1-\sigma)}{L} \left(\frac{-g_k^T d_k}{\|d_k\|} \right)^2, k \in K_2 \tag{29}$$

By (25) and the definition of s_k , we have:

$$f_k - f_{k+1} \geq \sigma \left(\frac{-g_k^T d_k}{\|d_k\|} \right)^2, k \in K_1 \tag{30}$$

Set $\eta = \min\{\sigma, \beta\sigma(1-\sigma)/L\}$, by (29) and (30), we have:

$$f_k - f_{k+1} \geq \eta \left(\frac{-g_k^T d_k}{\|d_k\|} \right)^2 \tag{31}$$

By the first scaling property and $d_k^T g_k < 0$, it follows that:

$$\lim_{k \rightarrow \infty} \left(\frac{-g_k^T d_k}{\|d_k\|} \right)^2 = 0$$

Theorem 3: Suppose that x_1 is arbitrary point for which Assumptions 1 and 2 hold. For the new extended CG-algorithm assume that the line search satisfies the conditions (9a) and (9b). Then if $g_k^T d_k < 0 \forall k \geq 1$ we have:

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0 \tag{32}$$

Proof: Suppose that the theorem is not true, then \exists a constant $\alpha > 0$ such that:

$$\|g_k\| \geq \alpha \quad \forall k \geq 1 \tag{33}$$

Now, for our new search direction defined in Step 6 of the new extended algorithm we have:

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{1}{\|g_k\|^2} \tag{34}$$

Now, for any $k \geq 1$:

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{1}{\alpha^2} = \infty \tag{35}$$

$$\Rightarrow \sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \infty \tag{36}$$

which is a contradiction. Hence, the theorem must be true and the algorithm has super-linear convergence

RESULTS

In order to test the effectiveness of the new algorithm which has been used to extend the standard CG- algorithm, the comparative tests involve several standard; effective and well-known test functions. These test functions has been chosen and solved numerically by utilizing the new and established algorithms.

Table 1-4 utilize the comparisons between our new extended CG-algorithm which is corresponding to the new non-quadratic model represented in (6), denoted by (NEW), the classical Fletcher and Reeves CG-algorithm, denoted by (FRCG), the rational model of Tassopoulos and Storey, denoted by (TS); Al-Bayati rational model, denoted by (B) for low and high dimensions. The cubic line search (Bunday, 1984), Armijo line search (Al-Bayati and Rassam, 2009) and the well-known convergence criterion $\|g_i\| < 1 \times 10^{-5}$ were used.

Table 1: Comparisons of different algorithms for non-quadratic models $2 \leq n \leq 10$ (low dimensions)

Test fun	N	FRCG		TS		B		NEW	
		NOI	NOF	NOI	NOF	NOI	NOF	NOI	NOF
1	2	31	73	31	73	31	81	29	68
2	10	22	46	18	44	20	43	18	44
3	4	28	61	36	78	25	58	30	64
4	4	50	114	38	96	26	77	34	80
5	4	57	178	46	139	162	516	44	124
Total		188	472	169	430	264	775	155	389

Table 2: Comparison of different algorithms for non-quadratic models $20 \leq n \leq 80$ (low dimensions)

Test fun	N	FRCG		TS		B		NEW	
		NOI	NOF	NOI	NOF	NOI	NOF	NOI	NOF
1	60	23	56	23	57	30	79	27	63
2	20	15	37	15	37	15	37	15	37
3	40	51	105	43	90	25	58	34	71
3	80	69	140	43	90	25	58	31	67
4	60	67	136	43	90	30	90	28	59
4	80	112	239	75	167	30	90	47	106
6	20	24	61	24	61	21	55	24	60
7	40	25	62	18	51	34	68	23	65
Total		386	836	284	643	210	535	229	528

Table 3: Comparisons of different algorithm for non-quadratic model $100 \leq n \leq 400$ (high dimensions)

Test fun	N	FRCG		TS		B		NEW	
		NOI	NOF	NOI	NOF	NOI	NOF	NOI	NOF
1	100	23	56	23	57	8	21	25	57
1	400	23	56	23	57	8	21	23	65
3	200	69	140	47	98	28	58	37	80
3	400	69	140	47	98	26	60	29	61
4	100	129	263	78	183	30	90	69	151
5	200	209	472	154	351	65	255	47	128
6	100	25	62	25	62	26	66	23	59
7	100	25	59	28	68	51	115	35	43
Total		572	1248	425	974	242	686	288	644

Table 4: Percentage ranging of NOI and NOF improvements

NEW	TS	FRCG	B
NOI = 100	147	198	-15
NOF = 100	151	193	106

Specifically, we record the Number Of Function calls (NOF), and the Number Of Iteration calls (NOI) for the purpose of our comparisons. For overcoming the overflow in our computer programs we have scaled the function values by a parameter equal to 1.E-10.

It is obvious that the new algorithm, for this set of low dimensionality test functions, improves:

- FRCG in (18%) NOI and (20%) NOF
- TS in (9%) NOI and (12%) NOF
- B in (40%) NOI and (50%) NOF

Also, the new algorithm, for the high dimensions, beats:

- FRCG in (41%) NOI and (37%) NOF
- TS in (20%) NOI and (18%) NOF
- B in (0%) NOI and (2%) NOF

In Table 4, taking the NOI and NOF 100%, we can determine the performance range of the new extended CG-algorithm according to others used in this study. Clearly it has (-15-98%) improvements in the NOI and it has about (6-93%) improvements in NOF.

We conclude that our new proposed rational logarithmic model is superior to some models in both quadratic and non-quadratic models.

The details of used test functions in this study are given in (Al-Bayati, 2007) and they are:

Rosen-Brock function:

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2, x_0 = [-1.2, 1]^T$$

Dixon function:

$$f(x) = \sum_{i=1}^n \left[(1 - x_i)^2 + (1 - x_n)^2 + \sum_{i=1}^{n-1} (x_i^2 - x_{i+1})^2 \right], x_0 = [-1, \dots, -1]^T$$

Generalized Wood function:

$$f(x) = \sum_{i=1}^{n/4} 100 \left[(x_{4i-2} - x_{4i-3}^2)^2 \right] + (1 - x_{4i-3})^2 + 90(x_{4i} - x_{4i-1}^2)^2 + (1 - x_{4i1})^2 + 10.1 \left[(x_{4i-2} - 1)^2 + (x_{4i} - 1)^2 \right] + 19.8(x_{4i-2} - 1)(x_{4i} - 1), x_0 = [-3, -1, -3, -1, \dots, -3, -1, -3, -1]^T$$

Generalized Powell function:

$$f(x) = \sum_{i=1}^{n/4} \left[(x_{4i-3} - 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-3} - x_{4i})^4 \right], x_0 = [3, -1, 0, 1, \dots, 3, -1, 0, 1]^T$$

Generalized Miele Function:

$$f(x) = \sum_{i=1}^{n/4} \left(\exp(x_{4i-3}) + 10x_{4i-2} \right)^2 + 100(x_{4i-2} + x_{4i-1})^6 + (\tan(x_{4i-1} - x_{4i}))^4 + (x_{4i-3})^8 + (x_{4i} - 1)^2, x_0 = [1, 2, 2, 2, \dots, 1, 2, 2, 2]^T$$

Non-diagonal function:

$$f(x) = \sum_{i=1}^n \left[100(x_i - x_i^2)^2 + (1 - x_i)^2 \right], x_0 = [-1, \dots, -1]^T$$

Oren and Spedicato OSP Function:

$$f(x) = \left(\sum_{i=1}^n i(x_i)^2 \right)^2, x_0 = [1, \dots, 1]^T$$

DISCUSSION

In this study, we have introduced a new non-quadratic CG-algorithm which is based on log of two rational functions. The new algorithm is compared with three well-known algorithms; namely FRCG; TS, (Tassopoulos and Storey, 1984) and B; (Al-Bayati, 1993) algorithms using seven well-known non-linear test functions with different dimensions taken from (Al-Bayati, 2007). Our numerical results indicate that the new technique has an improvements of about (-15-98%) in NOI against the others, while it saves about (6-93%) in NOF against the others.

CONCLUSION

In this research, a new fast extended CG algorithm is introduced. The proposed algorithm improved the standard FRCG; TS and B algorithms by adaptively modifying the search direction. The new proposed algorithm is generic and easy to implement in all

gradient based optimization process. The simulation results showed that it is robust and has a potential significantly enhance the computational efficiency of iterations and function evaluations.

REFERENCES

- Al-Bayati, A.Y., 1993. A new non-quadratic model for unconstrained non-linear optimization. *Mu'tah J. Res. Stud. Nat. Applied Sci. Ser.*, 8: 131-155.
- Al-Bayati, A.Y., 2001. New generalized CG-methods for the non-quadratic model in unconstrained optimizations. *J. Al-Yarmouk*, 10: 9-25.
- Al-Bayati, A.Y., 2007. New extended CG-methods for combined non-quadratic models in nonlinear optimization. *J. Al-Yarmouk*, 16: 347-366.
- Al-Bayati, A.Y. and N.H. Al-Assady, 1994. Minimization of extended quadratic functions with inexact line Searches. *JOTA*, 82: 139-147. DOI: 10.1007/BF02191784
- Al-Bayati, A.Y. and N.H. Al-Assady, 1997. PCG-method for nonlinear unconstrained optimization, *Mu'tah J. Res. Stud. Nat. Applied Science. Ser.*, 12:13-29.
- Al-Bayati, A.Y. and G.M. Al-Naemi, 1995. New extended CG-methods for non-linear optimization, *Mu'tah J. Res. Stud. Nat. Applied Sci. Ser.*, 10: 69-87.
- Al-Bayati A. Y. and S.Y. Rassam, 2009. A modified globally convergence CG-algorithm with a hybrid line search procedure. *Aust. J. Basic Applied Sci.*, 3: 2188-2206.
<http://www.insipub.com/ajbas/2009/2188-2206.pdf>
- Boland, W.R. E.R. Kamgnia and J.S. Kowalik, 1979. A conjugate gradient method invariant to non-linear scaling. *J. Optimiz. Theory Appl.*, 27: 221-230. DOI: 10.1007/BF00933228
- Bunday, B., 1984. *Basic Optimization Methods*. 1st Edn., Edward Arnold, Bedford Square, London, ISBN: 0713135069, pp: 24-28.
- Fletcher, R. and C.M. Reeves, 1964. Function minimization by conjugate gradients. *Computer. J.*, 7: 149-154. DOI: 10.1093/comjnl/7.2.149
- Fletcher, R., 1987. *Practical Methods of Optimization*. 2nd Edn., John Wiley and Son, Toronto and Singapore, New York, ISBN: 10: 0471494631, pp: 80-87.
- Fried, I., 1971. N-step conjugate gradient minimization scheme for non-quadratic functions. *AIAA J.*, 9: 2286-2287. DOI: 10.2514/3.6507
- Hestenes, M.R. and E. Stiefel, 1952. Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bureau of Stand.*, 5: 409-436. <http://nvl.nist.gov/pub/nistpubs/jres/049/6/V49.N06.A08.pdf>
- Renpu, G., 1989. A Parallel global optimization algorithm for rational separable factorable function. *J. Applied Math. Computer.*, 32: 61-72. DOI: 10.1016/0096-3003(89)90047-7
- Spedicato, E., 1976. A variable metric method for function minimization derived from invariancy to non-linear scaling. *J. Optimiz. Theory Appl.*, 20: 315-329. DOI: 10.1007/BF00933626
- Tassopoulos, A. and C. Storey, 1984. A conjugate direction method based on a non-quadratic model. *J. Optimiz. Theory Appl.*, 43: 371-381. DOI: 10.1007/BF00934461