Journal of Multidisciplinary Modeling and Optimization 3(1) (2020), 1-11.

Parameterized Three-Term Conjugate Gradient Method

Khalil K. ABBO [1] and Nehal H. HAMEED[2]

[1]Department of Mathematics,
Telafer University, Mosul, Iraq
kh_196538@yahoo.com

[2]Department of Mathematics,
Mosul University, Mosul, Iraq
Nehalh926@gmail.com

Received: 28.12.2019, Accepted: 21.08.2020, Published: 17.09.2020

Abstract — *In this paper a new parameterized three-term conjugate gradient algorithm is suggested, the descent property and global convergence are proved for the new suggested method. Numerical experiments are employed to demonstrate the efficiency of the algorithm for solving large scale benchmark test problems, particularly in comparison with the existent state of the art algorithms available in the literature..*

Keywords: Three-term conjugate gradient, Global convergence, Large scale benchmark test.
Mathematics Subject Classification: 65K10, 90C30, 90C26.

1 Introduction

Consider the non-linear unconstrained optimization problem

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

where $f: R^n \to R$ is continuously differentiable function and bounded from below. There are many different methods for solving the problem (1.1) see [8, 10, 12, 15]. We are interested in conjugate gradient (CG) methods, which have low memory requirements and strong local and global convergence properties [2,13].

For solving the problem (1.1), we consider the CG method, which starts from an initial point $x_1 \in R^n$ and generates a sequence $\{x_k\} \subset R^n$ as follows

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

where $\alpha_k > 0$ is a step size, received from the line search, and directions d_k are given [2,17] by, $d_1 = -g_1$ and

$$d_{k+1} = -g_{k+1} + \beta_k s_k \tag{1.3}$$

Cite as: *K. K. Abbo and N. H. Hameed, Parameterized Three-Term Conjugate Gradient Method, Journal of Multidisciplinary Modeling and Optimization 3(1) (2020), 1-11.*

In the equation (1.3) $g_k = \nabla f(x_k)$, $s_k = x_{k+1} - x_k$ and β_k is the conjugate gradient parameter. Various choices of the scalar β_k exist which give different performance on non-quadratic functions, yet they are equivalent for quadratic functions. In order to choose the parameter β_k for the method in present paper, we mention the following choices:

$\beta^{FR} = \dfrac{g_{k+1}^T g_{k+1}}{g_k^T g_k}$ Fletcher-Reeves [9]

$\beta^{HS} = \dfrac{g_{k+1}^T y_k}{d_k^T y_k}$ Hestenses-Stiefel [11]

$\beta^{PR} = \dfrac{g_{k+1}^T y_k}{g_k^T g_k}$ Polak-Ribiere[16]

$\beta^{LS} = -\dfrac{y_k^T g_{k+1}}{g_k^T d_k}$ Liu-Story[14]

$\beta^{DX} = -\dfrac{g_{k+1}^T g_{k+1}}{g_k^T d_k}$ Dixon [6]

where $y_k = g_{k+1} - g_k$.

Zhang (Zhang et al., 2006) have proposed the three – term FR, PR and HS conjugate gradient methods. Their methods always satisfy the descent condition $d_k^T g_k < 0$ or sufficient descent condition $d_k^T g_k = -c\|g_k\|$, where c positive constant, these methods have the following search directions:

1- FR three - term is

$$d_{k+1} = -g_{k+1} + \beta^{FR} d_k - \theta_k^{(1)} g_{k+1}, \quad \theta_k^{(1)} = \frac{d_k^T g_{k+1}}{g_k^T g_k} \tag{1.4}$$

2- PR three - term is

$$d_{k+1} = -g_{k+1} + \beta^{PR} d_k - \theta_k^{(2)} y_k, \qquad \theta_k^{(2)} = \frac{g_{k+1}^T d_k}{g_k^T g_k} \tag{1.5}$$

3-HS three - term is

$$d_{k+1} = -g_{k+1} + \beta^{HS} d_k - \theta_k^{(3)} y_k, \qquad \theta_k^{(3)} = \frac{g_{k+1}^T d_k}{d_k^T y_k} \tag{1.5}$$

we note that these methods always satisfy:

$$d_k^T g_k = -c\|g_k\|^2 < 0 \qquad \forall k \tag{1.6}$$

which implies the sufficient descent condition with $c = 1$.

The standard Wolfe (WC) line search conditions are frequently used in the conjugate gradient methods, these conditions are given in [18]

$$f(x_k + \alpha_k d_k) \le f(x_k) + \rho \alpha_k g_k^T d_k \tag{1.7}$$

$$g_{k+1}^T d_k \ge \sigma\, g_k^T d_k \tag{1.8}$$

where d_k is descent direction ie $g_k^T d_k < 0$ and $0 < \rho < \sigma < 1$. Strong Wolfe (SWC) conditions consist of (1.7) and the next stronger version of (1.8)

$$\left| g_{k+1}^T d_k \right| \le -\sigma\, g_k^T d_k \tag{1.9}$$

2 A New Modified Three–Term Conjugate Gradient (KN2).

In this section we develop a new three term conjugate gradient method (KN2), our idea is based on the following well-known Zhang's three terms CG-method:

$$x_{k+1} = x_k + \alpha_k d_k \tag{2.1}$$

$$d_{k+1} = \begin{cases} -g_1, & if\ k = 0, \\ -g_{k+1} + \beta_k^{DL} d_k - \xi_k (\mathrm{y}_k - \mathrm{t}\mathrm{s}_k), & \text{if } \mathrm{k} \ge 1, \end{cases} \tag{2.2}$$

where

$$\beta_k^{DL} = \frac{g_{k+1}^T (y_k - t s_k)}{d_k^T y_k}, \quad t \ge 0 \quad \text{and } \xi_k = \frac{g_{k+1}^T d_k}{d_k^T y_k} \tag{2.3}$$

If exact line search is used then $g_{k+1}^T d_k = 0$, therefore the method (2.2) reduces to the classical Liu and Story [14] conjugate gradient method, furthermore if the objective function is convex quadratic and line search is exact then (2.2) reduces to the DX method[6]. The method defined in (2.2)-(2.3) has some disadvantages for example, the value of t is unknown, which is crucial for β_K^{DL}.

To overcome to this disadvantage, we suggest the following search direction:

$$d_{k+1} = -g_{k+1} + \beta_k^{FR} s_k - \xi(\mathrm{y_k} - \mathrm{s_k}) \tag{2.4}$$

To find the value of ξ, we multiply (2.4) by y_k^T we get:

$$d_{k+1}^T \mathrm{y_k} = -g_{k+1}^T \mathrm{y_k} + \beta_k^{FR} s_k^T \mathrm{y_k} - \xi(\mathrm{y_k} - \mathrm{s_k})^T \mathrm{y_k} = 0$$

$$= -g_{k+1}^T \mathrm{y_k} + \beta_k^{FR} s_k^T \mathrm{y_k} - \xi\, y_k^T \mathrm{y_k} + \xi\, s_k^T \mathrm{s_k} = 0$$

$$= -g_{k+1}^T \mathrm{y_k} + \beta_k^{FR} s_k^T \mathrm{y_k} - \xi(y_k^T \mathrm{y_k} - s_k^T \mathrm{s_k}) = 0$$

then

$$\xi = \frac{\beta^{FR} s_k^T y_k - g_{k+1}^T y_k}{y_k^T y_k - y_k^T s_k} \tag{2.5}$$

hence we obtain the following new(KN2) search direction:

$$d_{k+1} = -g_{k+1} + \beta_k^{FR} s_k - \frac{\beta^{FR} s_k^T y_k - g_{k+1}^T y_k}{y_k^T y_k - y_k^T s_k}(y_k - s_k) \tag{2.6}$$

Some Remarks on The New(KN2) Method:

1- If the line search is exact i.e. $g_{K+1}^T s_k = 0$,then the search direction in (2.6) reduces to the following one .

$$d_{k+1} = -g_{k+1} + \beta_k^{FR} s_k + \frac{\beta^{FR} s_k^T g_k + g_{k+1}^T y_k}{y_k^T y_k + s_k^T g_k}(y_k - s_k) \tag{2.7}$$

If the objective function is convex quadratic and line search is exact , $g_{k+1}^T g_k = 0$ and $s_k^T g_{k+1} = 0$, then

$$\begin{aligned} y_k^T y_k &= y_k^T (g_{k+1} - g_k) \\ &= y_k^T g_{k+1} - y_k^T g_k \\ &= (g_{k+1} - g_k)^T g_{k+1} - (g_{k+1} - g_k)^T g_k \\ &= g_{k+1}^T g_{k+1} - g_k^T g_{k+1} - g_{k+1}^T g_k + g_k^T g_k \\ &= g_{k+1}^T g_{k+1} + g_k^T g_k \end{aligned}$$

then the search direction defined in (2.6) reduces to the following

$$d_{k+1} = -g_{k+1} + \beta_k^{FR} s_k + \frac{\beta^{FR} s_k^T g_k + g_{k+1}^T g_{k+1}}{g_{k+1}^T g_{k+1} + g_k^T g_k + s_k^T g_k}(y_k - s_k) \tag{2.8}$$

Note that we denote the search directions defined in (2.7) and (2.8) as KN2-1 and KN2-2, respectively.

Now we give the corresponding algorithms.

Algorithm (KN2):

Step1. Initialization: Select the initial point $x_1 \in R^n$, $\varepsilon > 0$, and select the parameters $0 < \rho < \sigma < 1$. Set $k = 1$. Compute f_k, g_k. Set

$d_k = -g_k$ and set $\alpha_k = \dfrac{1}{\|g_k\|}$.

Step2. Test for continuation of iterations. If $\|g_k\| \le \varepsilon$, then stop

Step3. Line search. Compute $\alpha_k > 0$ satisfying the standard Wolfe conditions (1.7) and(1.8) or strong Wolfe (1.7) and (1.9)

and update the variables $x_{k+1} = x_k + \alpha_k d_k$, compute

f_{k+1}, g_{k+1}, $y_k = g_{k+1} - g_k$ and $s_k = x_{k+1} - x_k$

Step4. Set $\beta_k = \beta_k^{FR}$ and compute ξ. If $y_k^T y_k - y_k^T s_k = 0$, then set $\xi = 0$ else set ξ as in (2.5).

Step5. If the restart criterion is satisfied, then set

$d_{k+1} = -g_{k+1}$ else compute d_{k+1} from (2.6).

Step6. Calculate the initial guess $\alpha_{k+1} = \alpha_k \|d_k\| / \|d_{k+1}\|$

Step7. Set $k = k+1$ go to **step 2**.

To prove the sufficient descent property to the algorithm KN2 we need the following theorem.

Theorem 2.1 [1]. If an α_k is calculated which satisfies strong Wolfe conditions (1.7) and (1.9) with $\sigma \in (0, \frac{1}{2}]$ for all k and $(g_k \ne 0)$, then the descent property for the Fletcher-Reeves method holds for all k.

In the following theorem we will show that the search directions generated by equation (2.6) are descent directions.

Theorem 2.2. Consider the search directions defined by the equation (2.6). Let the step-size α_k satisfies the strong Wolfe conditions (1.7), (1.9) and assume that the condition $y_k^T g_{k+1} \le \beta_{k+1}^{FR} s_k^T y_k$ hold then

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

Proof: If k=0, then $d_1 = -g_1$ and we get $d_1^T g_1 = -\|g_1\|^2 < 0$ where $c = -1$

Suppose $d_k^T g_k \le -c \|g_1\|^2$, to prove for k+1 consider the search direction defined in (2.6)

$$d_{k+1} = -g_{k+1} + \beta^{FR} s_k - \frac{\beta^{FR}_{k+1} s_k^T y_k - g^T_{k+1} y_k}{y_k^T y_k - y_k^T s_k}(y_k - s_k)$$

We multiply by g^T_{k+1} then,

$$g^T_{k+1} d_{k+1} = -g^T_{k+1} g_{k+1} + \beta^{FR} g^T_{k+1} s_k - \frac{\beta^{FR}_{k+1} s_k^T y_k - g^T_{k+1} y_k}{y_k^T y_k - y_k^T s_k} g^T_{k+1}(y_k - s_k)$$

$$\leq -\|g_{k+1}\|^2 + \beta^{FR} g^T_{k+1} s_k - \frac{\beta^{FR}_{k+1} s_k^T y_k - g^T_{k+1} y_k}{y_k^T y_k - y_k^T s_k} g^T_{k+1}(y_k - s_k)$$

$$= -\|g_{k+1}\|^2 + \beta^{FR} g^T_{k+1} s_k - \frac{\beta^{FR}_{k+1} s_k^T y_k}{y_k^T y_k - y_k^T s_k} g^T_{k+1}(y_k - s_k) + \frac{g^T_{k+1} y_k}{y_k^T y_k - y_k^T s_k} g^T_{k+1}(y_k - s_k)$$

Since $y_k^T g_{k+1} \leq \beta^{FR}_{k+1} s_k^T y_k$ therefor $g^T_{k+1} d_{k+1} \leq -g^T_{k+1} g_{k+1} + \beta^{FR}_{k+1} g^T_{k+1} s_k$ which is Fletcher-Reeves search direction, therefor by theorem (2.1), $g^T_{k+1} d_{k+1} \leq -(1-\sigma^k)\|g_{k+1}\|^2$ where $\sigma \leq 1/2$.

On the other hand the search directions generated by the equation (2.6) are conjugate directions for all k.

3 Convergence Analysis

At first, we give the following basic assumption on the objective function Assumption 2.1

i-The level $S = \{x \in R^n \mid f(x) \leq f(x_1)\}$ is bounded, where x_1 is the starting point then there exists a constant $\gamma_1 > 0$. such that:

$$\|x\| \leq \gamma_1 \qquad \textit{for all } x \in S \tag{3.1}$$

ii- In some neighborhood N of continuously differentiable and its gradient is Lipchitz continuous with Lipchitz constant $L > 0$, .i.e

$$\|g(u) - g(w)\| \leq L\|u-w\| \qquad \forall\, u,w \in N \tag{3.2}$$

Assumption (2.1.ii) implies that there exists a positive constant γ_2 such that

$$\|g(x)\| \leq \gamma_2 \qquad \forall\; x \in S \tag{3.3}$$

Proposition 3.1. Suppose that Assumption (2.1) is satisfied, and consider any conjugate gradient method, where d_k is descent direction and α_k is obtained by the Wolfe conditions. If $\sum_{k=1}^{\infty} \frac{1}{\|d_k\|^2} = \infty$ then $\lim_{k \to \infty} \inf \|g_k\| = 0$ for prove see [10].

In the rest of this section, we assume $g_k \neq 0$ *for all* k, otherwise a stationary point has been found. Under Assumption(2.1) , and Zoutendijk condition we can use the following theorem to prove the global convergence of the (KN2).

Theorem 3.1. Suppose that Assumption (2.1) is satisfied. Consider the KN2 method where d_k is descent direction and α_k satisfies Wolfe(standard or strong) conditions then $\lim_{k\to\infty} \inf \|g_k\| = 0$.

Proof: From equations (2.6), (3.1), (3.2) and (3.3)we have

$$\begin{aligned}
\|d_{k+1}\| &= \left\| -g_{k+1} + \beta^{FR} s_k - \frac{\beta^{FR} s_k^T y_k - g_{k+1}^T y_k}{y_k^T y_k - s_k^T y_k}(y_k - s_k \right\| \\
&\le \|g_{k+1}\| + |\beta^{FR}| \|s_k\| + \left| \frac{\beta^{FR} s_k^T y_k - g_{k+1}^T y_k}{y_k^T y_k - s_k^T y_k} \right| (\|y_k - s_k\|) \\
&\le \|g_{k+1}\| + |\beta^{FR}| \|s_k\| + |\beta^{FR}| \frac{\|g_{k+1}\| \|s_k\| L}{2} (L+1) \|s_k\|) \\
&\le \gamma_2 + \gamma_1 + \frac{\gamma_2 \gamma_1 L}{2}(L+1)\gamma_1 = \frac{2(\gamma_2 + \gamma_1) + (L+1)\gamma_2 \gamma_1 L}{2}
\end{aligned}$$

Let $\gamma = \max(\gamma_2, \gamma_1)$ therefore

$$\sum_{k=1}^{\infty} \|d_{k+1}\| \le \sum_{k=1}^{\infty} \frac{4\gamma + (L+1)\gamma^2 L}{2}$$

$$\sum_{k=1}^{\infty} \frac{1}{\|d_{k+1}\|} \ge \sum_{k=1}^{\infty} \frac{2}{4\gamma + (L+1)\gamma^2 L} = \infty$$

Hence by proposition(3.1) $\lim_{k\to\infty} \inf \|g_k\| = 0$

4 Numerical Results

This section presents the performance of FORTRAN implementation of our new three-terms conjugate gradient (KN2) algorithm on a set of unconstrained optimization test problems. We selected a number of 72 large-scale unconstrained optimization test problems in extended or generalized form presented in [3]. For each test function, we have considered numerical experiments with the number of variables increasing as n=100, …,1000. The algorithms uses the Wolfe line search conditions (1.7)and (1.8)with cubic interpolation, $\rho = 0.0001$ *and* $\sigma = 0.9$ and the same stopping criterion $\|g_k\|_2 \le 10^{-6}$, where $\|.\|_2$ is the Euclidean norm.

The algorithms we compare in these numerical experiments find the local solutions. Therefore, the comparisons of algorithms are given in the following context. Let f_i^{ALG1} and f_i^{ALG2} be the optimal value found by ALG1 and ALG2 for problem i=1,…,720, respectively. We say that in the particular problem i , the performance of ALG1 was better than the performance of AlG2 if:

$$\left\| f_i^{ALG1} - f_i^{ALG2} \right\| < 10^{-3}$$

and number of iterations (iter), or the number of function-gradient evaluations (fg), or the CPU time corresponding to ALG2 respectively. We have compared our algorithm versus to the following algorithms:

1-Three term Fletcher-Revees (1.4).

2-Three term Hestenes-Stiefel (1.5).

In all these algorithms, the initial step size is $\alpha_1 = 1/\|g_1\|$ and initial guess for other iterations i.e. ($k > 1$) is $\alpha_k = \alpha_{k+1}(\|d_{k+1}\|/\|d_k\|)$.

The codes are written in double precision FORTRAN (2000) and compiled with F77 default compiler setting. This code originally written by Andrei and we modified it.

Figures 1, 2 and 3 show performance of these methods for solving 72 unconstrained optimization test problems for dimensions n=100,..,1000, relative to the iterations (iter), function–gradient evaluations (fg) and CPU time, which are evaluated using the profile of Dolan and More [7]. That is, for each method, we plot the fraction p of problems for which the method is within a factor τ of the best (iter) or (fg) or CPU time. The left side of the figure gives the percentage of the test problems for which a method is the fastest, the right side gives the percentage of the test problems that are successfully solved by each of the methods. The top curve is the method that solved the most problems in a (iter, fg, time) that was within a factor τ of the best (iter, fg, time).

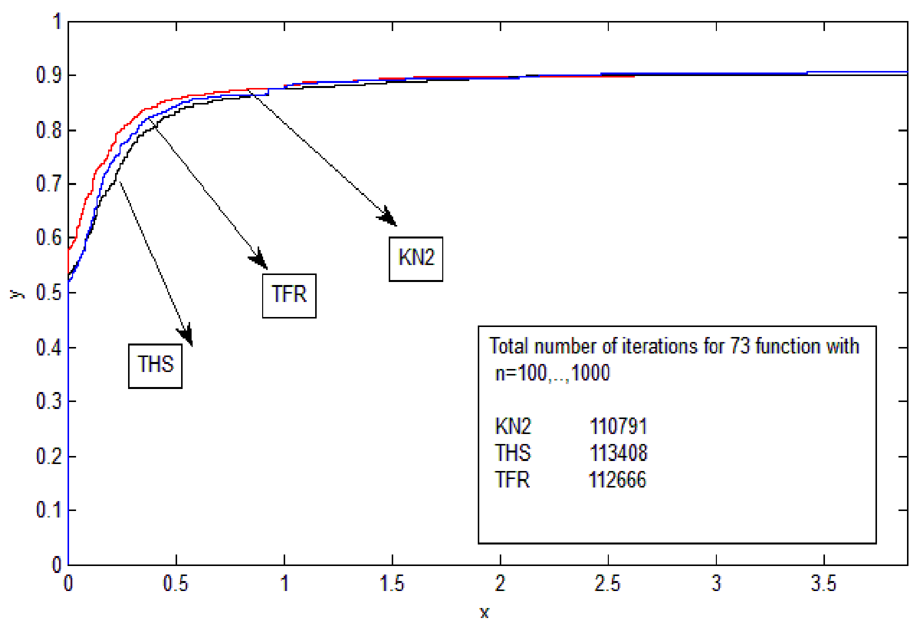

Figure 1: Performance Based on Number of Iterations

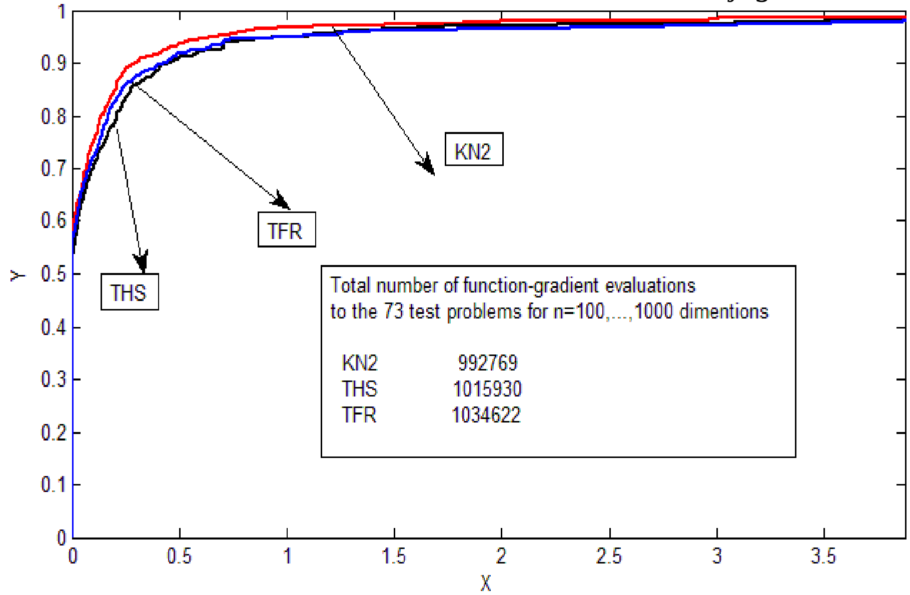

Figure 2: Performance Based on Number of Function-Gradient Evaluations

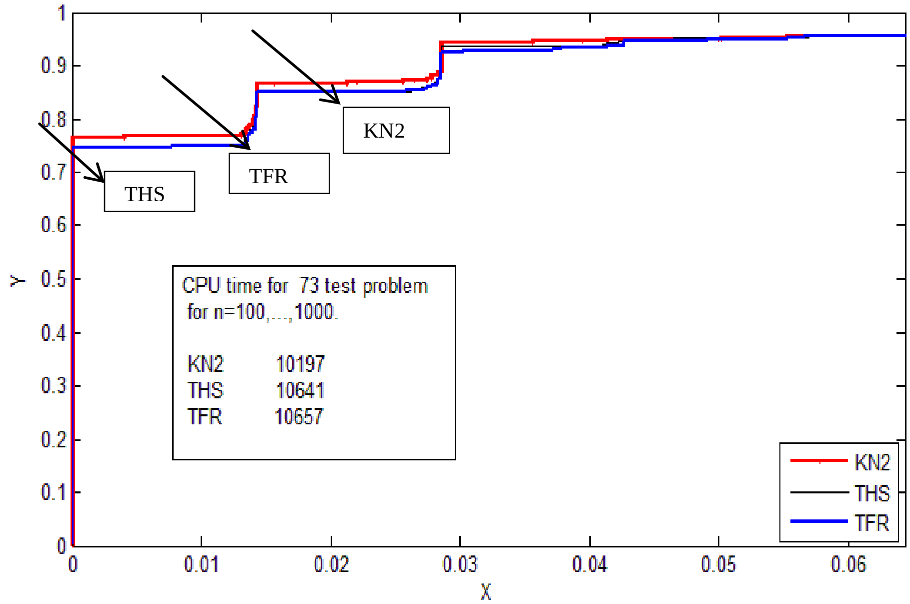

Figure 3: Performance Based on CPU Time in Seconds

Acknowledgment

The authors declare that this work is completely private effort and it has not been supported by any governmental, financial, educational or otherwise establishment.

Conflict of Interest Declaration

The authors declare that there is no conflict of interest statement.

Ethics Committee Approval and Informed Consent

The authors declare that declare that that there is no ethics committee approval and/or informed consent statement.

References

[1] M. Al-Baali, Descent property and global convergence of the Fletcher Reeves method with inexact line search, IMA J. Numer. Anal., 5(1) 1985, 121-124.

[2] N. Andrei, New Hybrid Conjugate Gradient Algorithms for Unconstrained Optimization . In: Floudas C., Pardalos P. (eds) Encyclopedia of Optimization, Springer, Boston, 2008, 2560-2571.

[3] N. Andrei, An unconstrained optimization test functions collection, Adv. Model. Optim, 10(1) 2008, 147-161.

[4] I. Bongartz, A. Conn, N. Gould and P. Toint, Constrained and unconstrained testing environment, J. Optim. Theory Appl., 21(1) 1993, 123-160.

[5] Y. Dai, Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, SIAM J. Optimiz., 10(1) 1999, 177-182.

[6] L. C.W. Dixon, Nonlinear optimization: A survey of the state of the art, Hatfield Polytechnic. Numerical Optimization Centre, 1973.

[7] D. Dolan, J. Mor´e, Benchmarking optimization software with performance profiles, Math. Program., 91(2) 2002, 201-213.

[8] K. Edwin, H. Stanilow, An introduction to optimization, Second Edition, Wiley and Sons, 2001.

[9] R. Fletcher, C. M. Reeves, Function minimization by Conjugate gradients, Comput. J., 7(2) 1964, 149-154.

[10] W. Hager, H. Zhang, A survey of nonlinear conjugate gradient methods, Pac. J. Optim., 2(1), 2006, 35-58.

[11] M. Hestenes, E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bur. Stand., 49(1) 1952.

[12] K. K. Abbo, L. A. Abdulwahid, Generalized Dai-Yuan non-linear conjugate gradient method for unconstrained optimization,

Int. J. Sci. Math. Educ., 8(6) 2017, 17993-17999,

[13] X. Li, X. Zhao, A hybrid conjugate gradient method for optimization problems, Nat. Sci., 3(1) 2011, 85.

[14] Y. Liu, C. Storey, Efficient generalized conjugate gradient algorithms, part 1: theory, J. Optimiz. Theory App., 69(1) 1991, 129-137.

[15] J. Nocedal, J. Wright, Numerical Optimization, Springer Series in Operations Research, Springer Verlag, New York, 2006.

[16] E. Polak and G. Ribiere, Note sur la convergence de m´ethodes de directions conjugu´ees, Rev. Fr. Inform. Rech. O., 3(16) 1969, 35-43,

[17] S. S. Djordjevi, New hybrid conjugate gradient method as a convex combination of FR and PRP Methods. Filomat, 30(11) 2016, 3083-3100,

[18] P. Wolfe, Convergence conditions for ascent methods, SIAM Rev., 11(2) 1969, 226-235,

Khalil K. Abbo, ORCID: https://orcid.org/0000-0001-5858-625X
Nehal H. Hameed, ORCID: https://orcid.org/0000-0002-4548-3091