ON A MODIFIED CONJUGATE GRADIENT METHOD FOR SOLVING NONLINEAR UNCONSTRAINED OPTIMIZATION PROBLEMS

[1]Oke, M.O., [2]Oyelade, T.A. and [3]Raji, R.A.
[1,2]Department of Mathematics, Ekiti State University, Ado-Ekiti, **NIGERIA**
[3]Department of Mathematics and Statistics, Osun State Polytechnic, Iree, **NIGERIA**

ABSTRACT

Conjugate gradient method is an important technique for solving unconstrained optimization problems. In this paper, we modified the conjugate gradient method by introducing some parameters to the value of conjugate gradient coefficient (β_i) in the descent direction. The modified conjugate gradient algorithm was then applied to solve some nonlinear unconstrained optimization problems after establishing the convergence criteria. The results obtained compares favourably with existing results.

Keywords: Modified Conjugate Gradient Method, Unconstrained optimization problem, Descent direction, Convergence criteria, Non-linear problem.

INTRODUCTION

Conjugate gradient method is an efficient and organized tool for solving nonlinear optimization problems. The method is popularly used by mathematician, engineers and those who are interested in solving large scale optimization problems, Mital (1976), Liu and Storey (1991) and Wei et al. (2006).

A lot of researchers have worked on conjugate gradient methods for solving nonlinear unconstrained optimization problems. Can (2013) considered a modified conjugate gradient method for unconstrained optimization, Jie and Zhong (2019) looked at a modified spectral PRP conjugate gradient projection method for solving large-scale monotone equations and its application in compressed sensing. Li et al. (2006) worked on global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search. Waziri and Adeboye (2007) considered conjugate gradient method with Ritz method for the solution of boundary value problems. Raji and Oke (2015) looked at higher order conjugate gradient method for solving continuous optimal control problems, to mention a few. None of these researchers had considered the kind of modification we made to the value of conjugate gradient coefficient (β_i) in the descent direction of the newly modified conjugate gradient method.
In this paper, we considered an unconstraint optimization problem of the form:

$$\min f(x), \ x \in \mathbb{R}^n, \tag{1}$$

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is continuously differentiable with available gradient $g(x)$, Fletcher and Reeves (1964), Hager and Zhang (2003) and Oke (2014). Generally, conjugate gradient method (CGM) generates an iterative sequence $\{x_i\}$ defined by

$$x_{i+1} = x_i + \alpha_i p_i, \qquad i = 0, 1, 2, \dots, \tag{2}$$

where $\alpha_i > 0$ is a line search and p_i is a search descent direction defined by

$$p_i = \begin{cases} -g_i & if\ i = 0 \\ -g_i + \beta_i p_i & if\ i \geq 1 \end{cases} \tag{3}$$

where the term $g(x_i) = g_i$ is the gradient at i-th iteration and β_i is a scalar, Fletcher and Reeves (1964) and Sun and Yuan (2006). β_i is given by

$$\beta_i = \frac{g_{i+1}{}^T g_{i+1}}{g_i{}^T g_i} \qquad i = 0, 1, 2, \ldots, \tag{4}$$

An effective method for the generation of conjugate directions proposed by Hestenes and Stiefel (1952) is the so-called conjugate-gradient method. In this method, directions are generated sequentially, one per iteration. For iteration $i + 1$, a new point $\boldsymbol{x}_{i+1}$ is generated by using the previous direction p_i. Then a new direction $\boldsymbol{p}_{i+1}$ is generated by adding a vector $\beta_i p_i$ to $-g_{i+1}$, the negative of the gradient at the new point. Restarting and preconditioning are very important in this case in order to improve the CGM. Some of well-known CGMs are:

$\beta_i^{HS} = \frac{g_{i+1}^T y_i}{p_i^T y_i}$, Hestenes and Stiefel (1952),

$\beta_i^{FR} = \frac{\|g_{i+1}\|^2}{\|g_i\|^2}$, Fletcher and Reeves (1964),

$\beta_i^{LS} = \frac{g_{i+1}^T y_i}{-p_i^T g_i}$, Liu and Storey (1991),

$\beta_i^{DY} = \frac{\|g_{i+1}\|^2}{p_i^T y_i}$, Dai and Yuan (1999),

$\beta_K^{WYL} = \frac{g_{i+1}^T\left(g_i - \frac{\|g_{i+1}\|}{\|g_i\|} g_i\right)}{\|g_i\|^2}$, Wei et al. (2006)

where $y_i = g_{i+1} - g_i$

The CGM is a variant of the gradient method. In its simplest form, the gradient method uses the iterative scheme to generate a sequence $\{x_i\}_{i=1}^{n}$ of vectors which converge to the minimum of $f(x)$. The parameter α in the algorithm is the step length of the descent direction sequence. In particular, if f is a function on a Hilbert space $\mathcal{H}$ such that in $\mathcal{H}$, f admits a Taylor series expansion

$$f(x) = f_0 + \langle a, x\rangle_H + \frac{1}{2}\langle x, Ax\rangle_H \tag{5}$$

where $a, x \in \mathcal{H}$ and A is a symmetric, positive definite, linear operator. It can be shown that f possesses a unique minimum $x^*\ in\ \mathcal{H}$ and that $\nabla f(x^*) = 0$, Dai and Yuan (1999). The CGM algorithm for iteratively locating the minimum x^* of $f(x)$ in $\mathcal{H}$ is described in Dai and Yuan (1999), Fletcher and Powell (1963), Fletcher and Reeves (1964) and Fletcher (1987).

MATERIALS AND METHODS

In this research work, a new value of conjugate gradient coefficient (β_i) in the descent direction was constructed. The algorithm of the proposed newly Modified Conjugate Gradient Method (MCGM) for solving nonlinear unconstrained optimization problems is as follows:

$Step\ 1$: Guess the initial element, x_0

$Step\ 2$: Compute the gradient of the function at initial guess, g_0

$Step\ 3$: Compute the descent direction, $p_0 = -g_0$

$Step\ 4$: Set $x_{i+1} = x_i + \alpha_i p_i$, $\quad \forall\, i = 0, 1, 2, \ldots, n$

where $\alpha_i = \frac{g_i{}^T g_i}{p_i{}^T A p_i}$, $\quad \forall\, i = 0, 1, 2, \ldots, n$

Step 5: Compute $g_{i+1} = g_i + \alpha_i A p_i, \quad \forall\, i = 0, 1, 2, \ldots, n$

Step 6: Update the descent direction, $p_{i+1} = -g_{i+1} + \beta_i p_i$,

where $\beta_i = \frac{-(g_{i+1}+g_i)^T g_{i+1}}{p_i^T g_i}, \qquad \forall\, i = 0, 1, 2, \ldots, n$

Step 7: If $g_i = 0$ for some i then, terminate the sequence; else set $i = i + 1$ and go to Step 4

In the iterative steps 2 through 7 above, g_i denotes the gradient of the function f at x_i, p_i denotes the descent direction at i-th step of the algorithm and α_i denotes the step length of the descent sequence $\{x_i\}$. Steps 4, 5, 6 and 7 of the algorithm reveal the crucial role of the linear operator A in determining the step length of the descent sequence and also in generating a conjugate direction of search.

In the newly modified conjugate gradient method (MCGM), our β_i in the descent direction is given by $\beta_i = \frac{-(g_{i+1}+g_i)^T g_{i+1}}{p_i^T g_i}$ while in CGM, we have the value of β_i as $\beta_i = \frac{g_{i+1}^T g_{i+1}}{g_i^T g_i}$.

Based on the seventh step of MCGM in solving problems, we can use any of the following to set the stopping creteria.

* The function is said to have converged when the gradient value is zero, Polyak (1969).
* The Gradient Norm can also be used as the stopping criterion to determine the convergence of the function as the gradient norm tends towards zero, Fletcher and Powell (1963) and Polyak (1969).
* The solution can be compared with the analytical results or existing results using other methods.

We applied the theorems in Polyak (1969) and Horn and Johnson (1985) to show that the descent direction given in our newly modified conjugate gradient algorithm is indeed a conjugate direction by showing that the set $\{p_i\}$ are conjugate with respect to the linear operator A and that α_i is chosen at each step to minimize $f(x_i + \alpha_i p_i)$, so as to have a descent sequence.

Convergence Analysis

We will now present the convergence properties of our new β_i by first establishing the sufficient condition to be satisfied by the search direction.

Theorem:

Suppose that the sequence g_i and p_i are generated by the MCGM algorithm, and the step length is determined by a symmetric and positive definite operator A in step 5 above. If we choose $\sigma < {}^1\!/_{1000}$, then the sequence p_i satisfy the sufficient descent condition $g_i^T p_i \le -C\|g_i\|^2$ where C is a constant, Jie and Zhong (2019) and Li et al. (2006).

Proof

From step 6, we have $\beta_i = \frac{-(g_{i+1}+g_i)^T g_{i+1}}{p_i^T g_i}$ (6)

Applying the above condition in β_i, we have

$$= \frac{-\|g_{i+1}\|^2 - g_i^T g_{i+1}}{p_i^T g_i} \le \frac{-\|g_{i+1}\|^2}{\|p_i\|^2} \qquad (7)$$

with $\|g_{i+1}\|$ as the Euclidean norm of vectors. By using (3), we have:

$$\frac{g_{i+1}^T p_i}{\|g_{i+1}\|^2} = -1 + \beta_i \frac{g_{i+1}^T p_i}{\|g_{i+1}\|^2} \qquad (8)$$

The descent property is shown by induction, since $g_0^T p_0 = -\|g_0\|^2 < 0 \;\; and \;\; g_0 \neq 0$
Suppose $p_i, i = 1, 2, \dots, k$ are all descent direction. By using step 6 and the inequalities in (7), we obtain the following:

$$\beta_i |g_{i+1}{}^T p_i| \leq \frac{\|g_{i+1}\|^2}{\|p_i\|^2} \sigma(g_i{}^T p_i)$$

$$\sigma \frac{\|g_{i+1}\|^2}{\|g_i\|^2}(g_i{}^T p_i) \leq \beta_i \, g_{i+1}{}^T p_i \leq -\sigma \frac{\|g_{i+1}\|^2}{\|p_i\|^2}(g_i{}^T p_i) \quad (9)$$

From (8) and (9), we can deduce that:

$$-1 + \sigma \frac{g_{i+1}{}^T p_{i+1}}{\|p_i\|^2} \leq \frac{g_{i+1}{}^T p_i}{\|g_{i+1}\|^2} \leq -1 - \sigma \frac{g_{i+1}{}^T p_i}{\|p_i\|^2} \quad (10)$$

Since $g_0^T p_0 = -\|g_0\|^2$, we can repeat the process to get

$$-1 + \sigma(1) \leq \frac{g_0{}^T p_0}{\|g_0\|^2} \leq\leq -1 - \sigma(-1) \quad (11)$$

$$-1 - \sum_{i=1}^{k} \sigma^i \leq \frac{g_{i+1}{}^T p_{i+1}}{\|g_{i+1}\|^2} \leq -1 - \sum_{i=1}^{k} \sigma^i \quad (12)$$

$$-1 - \sum_{i=1}^{k} \sigma^i \leq \frac{g_{i+1}{}^T p_{i+1}}{\|g_{i+1}\|^2} \leq -2 - 1 + \sum_{i=1}^{k} \sigma^i \quad (13)$$

$$-\sum_{i=1}^{k} \sigma^i \leq \frac{g_{i+1}{}^T p_{i+1}}{\|g_{i+1}\|^2} \leq -2 + \sum_{i=1}^{k} \sigma^i \quad (14)$$

Recall that

$$\sum_{i=1}^{k} [\sigma]^i = \sum_{i=1}^{\infty} [\sigma]^i = \frac{1}{1-\sigma} \quad (15)$$

Then, (14) can written as

$$-\frac{1}{1-\sigma} \leq \frac{g_{i+1}{}^T p_{i+1}}{\|g_{i+1}\|^2} \leq -2 + \frac{1}{1-\sigma} \quad (16)$$

Therefore, by applying the method of induction $g_i{}^T p_i < 0$ hold for all $i \geq 0$

Let $C = 2 - \frac{1}{1-\sigma}$ (17)

where $C \in (0, 1)$, then

$$C - 2 \leq \frac{g_{i+1}{}^T p_{i+1}}{\|g_{i+1}\|^2} \leq -C \quad (18)$$

Thus, this completes the proof.

RESULTS AND DISCUSSION

We now test the performance of our newly modified Conjugate gradient method on a number of unconstrained optimization problems.
The following problems were evaluated and the results are shown in tables 1 to 5 using the superliner and gradient norm values of the newly Modified Conjugate Gradient Method as the bases of convergence.

Problem 1

$$\min_x f(x) = 4(x_1 - 5)^2 + (x_2 - 6)^2 \;\; with \; x_0 = [2, -1]^T$$

Problem 2

$$\min_x f(x) = (x_1 - 3)^2 + 9(x_2 - 5)^2 \;\; with \; x_0 = [1, 1]^T$$

Problem 3

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

Problem 4

$$\min_x f(x) = \frac{1}{2}x_1^2 + x_1 x_2 + x_2^2 \;\; with \; x_0 = [10, -5]^T$$

Problem 5

$$\min_{x} f(x) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2 \quad with\ x_0 = [0,0]^T$$

Table 1: Solution to Problem 1

Iteration	x_1	x_2	MCGM Function Value(s)	CGM Function Value(s)	MCGM Gradient Norm	CGM Gradient Norm
0	2	-1	85	85	772	772
1	5.7056	1.1616	25.4016	25.4016	125.5042253	125.5042253
2	5.00000000026	6.0000000003	$2.8e^{-21}$	$8.32e^{-14}$	$8.49e^{-22}$	$5.67e^{-14}$

Table 2: Solution to Problem 2

Iteration	x_1	x_2	MCGM Function Value(s)	CGM Function Value(s)	MCGM Gradient Norm	CGM Gradient Norm
0	1	1	148	148	5200	5200
1	1.222831676	5.0109701752	3.1594103545	3.1594103	12.67230065	12.6723007
2	3.000000001	5.00000000001	$-1.00296e^{-22}$	$-2.842e^{-14}$	$7.7791e^{-12}$	2.4802298

Table 3: Solution to Problem 3

Iteration	x_1	x_2	MCGM Function Value(s)	CGM Function Value(s)	MCGM Gradient Norm	CGM Gradient Norm
0	2.2	-0.44	24.2	24.2	7763.36	7763.36
1	2.177945552	0.0010889	4.743565413	4.74356541	19.02122177	19.021222
2	$5.00000217e^{-10}$	$-1.36e^{-11}$	$2.68714e^{-28}$	$7.572e^{-29}$	$3.31369e^{-29}$	$0.302e^{-25}$

Table 4: Solution to Problem 4

Iteration	x_1	x_2	MCGM Function Value(s)	CGM Function Value(s)	MCGM Gradient Norm	CGM Gradient Norm
0	10	-5	25	25	25	25
1	5	-5	12.5	12.5	25	25
2	0	0	0	0	0	0

Table 5: Solution to Problem 5

Iteration	x_1	x_2	MCGM Function Value(s)	CGM Function Value(s)	MCGM Gradient Norm	CGM Gradient Norm

0	0	0	0	0	2	2
1	-1	1	-1	-1	2	2
2	-1	1.5	-1.25	-1.25	0	0

From table 1, we can easily see that the gradient norm for MCGM is less than that of CGM. The same is also true for problems 2 and 3 as shown in tables 2 and 3. This clearly shows that the MCGM performs better than the CGM. In problems 4 and 5, as we can see in tables 4 and 5, the MCGM has the same gradient norm as the CGM.

CONCLUSION

From the table of results for the problems considered, we can easily see that the newly modified conjugate gradient method (MCGM) is better than the classical conjugate gradient method (CGM) on the convergence profile by considering the gradient norm of the problems. It is only in problems 4 and 5 that we have the same gradient norms.

REFERENCES

Can, L. (2013). A Modified Conjugate Gradient Method for Unconstrained Optimization, Telkomnika (II), 11, 6373-6380.

Dai, Y. H. and Yuan, Y. A. (1999). Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property, SIAM Journal on Optimization, 10, 177-182.

Fletcher, R. and Powell, M. J. D. (1963). A Rapidly Convergent Descent Method for Minimization, Computer Journal, 6, 163-168.

Fletcher, R. and Reeves, C. M. (1964). Function Minimization by Conjugate Gradients, Computer Journal, 7, 149–154.

Fletcher, R. (1987). Practical Methods of Unconstrained Optimization, John Wiley & Sons, New York

Hager, W. W. and Zhang, H. (2003). A new Conjugate Gradient Method with Guaranteed Descent and an Efficient Line Search, SIAM Journal on Optimization, 16(1), 170-192

Hestenes, M. R. and Stiefel, E. L. (1952). Methods of Conjugate Gradients for Solving Linear Systems, J. Res. Natl. Bureau Standards, 49, 409–436.

Horn, R. A. and Johnson, C. R. (1985). Matrix Analysis, Cambridge University Press, New York.

Jie, G. and Zhong, W. (2019). A Modified Spectral PRP Conjugate Gradient Projection Method for Solving Large-scale Monotone Equation and its Applications in Compressed Sensing, Mathematical Problems in Engineering, Hindawi.

Li, Z., Weijin, Z. and Donghui, L. (2006). Global Convergence of a Modified Fetcher-Reeves Conjugate Gradient Method with Armijo-type Line Search, Numerische Mathematix, DOI 10.10071500211-006-0028z.

Liu, Y. and Storey, C. (1991). Efficient Generalized Conjugate Gradient Algorithms, Part 1: Theory Journal of Optimization Theory and Applications, 69, 129-137.

Mital, K. V. (1976). Optimization Methods in Operations Research and System Analysis, Wiley Eastern Limited.

Oke, M. O. (2014). A Second Order Method for Minimizing Unconstrained Optimization Problems, Journal of Applied Mathematics and Bioinformatics, International Scientific Press. 4(3), 65 – 73, U.K.

Polyak, B. T. (1969). The Conjugate Gradient Method in Extreme Problems, USSR

Computational Mathematics and Mathematical Physics, 9, 94-112.

Raji, R. A. and Oke, M. O. (2015). Higher order Conjugate Gradient Method (HCGM) for Solving Continuous Optimal Control Problems, International Organization of Scientific Research Journal of Mathematics, 11(1), 88-90.

Sun, W. and Yuan, Y. X. (2006). Optimization Theory and Methods: Nonlinear programming. Springer Optimization and Its Applications.

Waziri, V. O. and Adeboye, K. R. (2007). Conjugate Gradient Method with Ritz Method for the Solution of Boundary Valued Problems, Leonardo Electronic Journal of Prac. and Tech., 10, 67-78.

Wei, Z., Yao, S. and Liu, L. (2006). The Convergence Properties of some new Conjugate Gradient Methods, Applied Mathematics and Computation, 183(2), 1341-1350