Fuzzy Optimization and Modelling 3(2) (2022) 15-22

E-ISNN: 2676-7007

Contents lists available at FOMJ

Fuzzy Optimization and Modelling

Journal homepage: http://fomj.qaemshahriau.ir/

Paper Type: Research Paper

A Three-Term Extension of a Descent Conjugate Gradient Method

Zohre Aminifard*

Department of Mathematics, Semnan University, Semnan, Iran

ARTICLE INFO

Article history:
Received 21 February 2022
Revised 3 July 2022
Accepted 17 July 2022
Available online 18 July 2022

Keywords:
Unconstrained Optimization
Conjugate Gradient Method
Sufficient Descent Condition
Global Convergence

ABSTRACT

In an effort to make modification on the classical Hestenes--Stiefel method, Shengwei et al. proposed an efficient conjugate gradient method which possesses the sufficient descent condition when the line search fulfils the strong Wolfe conditions (by restricting the line search parameters). Here, we develop a three--term extension of the method which guarantees the sufficient descent condition independent to the line search. Also, we establish global convergence of the method using convexity assumption. At last, practical merits of the proposed method are investigated by numerical experiments on a set of CUTEr test functions. The results show numerical efficiency of the method.

1. Introduction

Conjugate gradient (CG) algorithms are among the efficient and popular tools for solving optimization problems which appear in practical applications. As typical cases, Lin et al. [33] used the method to deal with the statistical model of mineral potential prediction. Bouter et al. [18] and Esmaeili et al. [24] employed the method in the well-known image reconstruction problem. Efficiency of the CG algorithm in signal recovery has been investigated by Abubakar et al. [1] and Wan et al. [40]. Li et al. [31] applied the method to improve training of the neural networks. Liu et al. [34] applied the CG method for solving the optimization problems which appear in the four-dimensional variational data assimilation systems, to be used in the numerical weather prediction. Li et al. [32] proposed a class of scaled CG methods which can be applied to non-negative matrix factorization. CG methods have been also well-developed for solving nonlinear equations [9, 41, 43] as well as nonlinear least squares problems [19].

Generally, it has been believed that among the main reasons of popularity of the CG methods, there are the explicit usages of second order information of the objective function, strong convergence properties [17], simplicity of the implementation as well as low memory requirement. As known, for the n-dimensional

* Correspondig author
E-mail address: aminisor@semnan.ac.ir (Zohre Aminifard)

DOI: 10.30495/fomj.2022.1953254.1063

unconstrained optimization problem, $\min_{x\in\mathbb{R}^n} f(x)$, with the smooth objective function. The iterations of a CG method are in the form of $x_{k+1} = x_k + s_k, k \geq 0$, starting from an initial point $x_0 \in \mathbb{R}^n$ where $s_k = \alpha_k d_k$ in which $\alpha_k > 0$ is a step length determined by line search along the CG direction, d_k, often defined by

$$d_0 = -g_0, \;\; d_{k+1} = -g_{k+1} + \beta_k \, d_k, \;\; k \geq 0, \tag{1}$$

where $g_k = \nabla f(x_k)$ and β_k is a scalar called the CG parameter. A review of the literature reveals important role of the CG parameter on performance of the methods; see for example [2-7, 13-15, 22, 23, 26, 29, 35].

To achieve the descent property which may not necessarily hold for some of the classical CG methods [29], several efforts have been made mainly in the following lines: conducting eigenvalue analyses on the search direction matrix of some extended versions of the classical methods [10, 11, 36], developing spectral/scaled (preconditioned) versions of the methods [8, 12, 25] and extending three-term versions of the classical CG methods [16, 2, 38, 44].

As an extension of the Hestenes-Stiefel (HS) [30] method, one of the well-known three-term CG algorithms has been suggested by Zhang et al. [44] in which

$$d_0^{MHS} = -g_0, d_{k+1}^{MHS} = -g_{k+1} + \frac{g_{k+1}^T y_k}{d_k^T y_k} \, d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} \, y_k, k \geq 0, \tag{2}$$

Here, we suggest a three--term extension of a recent CG method proposed by Shengwei et al. [42] and show that the method satisfies the sufficient descent condition independent to the line search, leading to global convergence. These are discussed in Section 2. Other parts of this study consist of investigating computational efficiency of the given method in Section 3, and reporting concluding remarks in Section 4.

2. A three--term conjugate gradient method

In order to improve the theoretical and numerical performance of the classical Hestenes-Stiefel method, Shengwei et al. [37] proposed the following CG parameter inspired by the modification approach of [42]:

$$\beta_k^{MHS} = \frac{g_{k+1}^T z_k}{d_k^T y_k}, \tag{3}$$

where $z_k = g_{k+1} - \frac{||g_{k+1}||}{||g_k||} g_k$ and $||.||$ stands for the Euclidean norm. They established descent property as well as global convergence of the corresponding CG method under the strong Wolfe line search conditions [39], i.e.

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{4}$$

$$|\nabla f(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k, \tag{5}$$

with $0 < \delta < \sigma < 1$. Besides, the MHS method satisfies the sufficient descent condition, that is

$$g_k^T d_k \leq -\varrho ||g_k||^2, \tag{6}$$

where $\varrho > 0$ is a constant, provided that $\sigma < 0.25$.

Here, we suggest a three--term extension of the MHS method based on the approach of [44], namely TTMHS, with the following search directions:

$$d_0^{\text{TTMHS}} = -g_0, d_{k+1}^{\text{TTMHS}} = -g_{k+1} + \beta_k^{\text{MHS}} \, d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} \, z_k, k \geq 0. \tag{7}$$

Note that by exact line search we have $g_{k+1}^T d_k = 0$, and consequently, in such situation TTMHS reduces to MHS.

Also, it can be seen that $g_{k+1}^T d_{k+1}^{\text{TTMHS}} \leq -||g_{k+1}||^2$, for all $k \geq 0$. So, in contrast to MHS which satisfies the sufficient descent condition (6) when the strong Wolfe line search conditions hold with $\sigma < 0.25$, TTMHS satisfies the sufficient descent condition independent to the line search.

Now, we are in a position to spell out our algorithm as follows:

The TTMHS algorithm:

Step 0: Choose an $x_0 \in \mathbb{R}^n$, the line search parameters $0 < \delta < \sigma < 1$, and the tolerance $\epsilon > 0$. Set $d_0 = -g_0$ and $k = 0$.

Step 1: **If** $||g_k|| < \epsilon$, **then stop**.

Step 2: Compute the search direction d_k by (7).

Step 3: Compute the step length α_k satisfying (4) and (5).

Step 4: Set $x_{k+1} = x_k + \alpha_k d_k$.

Step 5: Set $k = k + 1$ and **go to** Step 1.

Here, we establish convergence of the TTMHS method. In this regard, the following classic assumptions are needed [44].

Assumption 1. (***i***) The level set $\Omega = \{x: f(x) \leq f(x_0)\}$ is bounded. (***ii***) In some neighbourhood $\mathcal{S}$ of Ω, the function f is continuously differentiable and its gradient is Lipschitz continuous; that is, there exists a positive constant L>0 such that

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||, \ \forall x, y \in \mathcal{S}. \tag{8}$$

Based on Assumption 1, it can be seen that there exists positive constant τ such that $||\nabla f(x_k)|| \leq \tau$, for all $x \in \mathcal{S}$. Also, since $\{f(x_k)\}_{k\geq 0}$ is a decreasing sequence, we have $\{x_k\}_{k\geq 0} \subseteq \Omega$. Hereafter, we suppose that Assumption 1 holds.

Moreover, without specification, the sequences $\{x_k\}_{k\geq 0}$ and $\{d_k\}_{k\geq 0}$ are respectively generated by (1) and (7). Now, we need the following result to establish the convergence of the TTMHS method.

Theorem 1. For TTMHS method, if f is uniformly convex on the neighborhood $\mathcal{S}$ of Ω, and the step length α_k is determined using the Wolfe line search (4) and (5), then, $\lim\inf_{k\to\infty} ||g_k|| = 0$.

Proof. At first, considering Theorem 1.3.16 of [39], the uniform convexity of the smooth function f ensures that there exists a constant $\nu > 0$ such that

$$s_k^T y_k \geq \zeta ||s_k||^2, \ \forall k \geq 0. \tag{9}$$

which together with (7) and (8), we have

$$\| d_{k+1} \| \leq \| g_{k+1} \| + \left|\beta_k^{MHS}\right| \ \| d_k \| + \frac{|g_{k+1}^T d_k|}{d_k^T y_k} \| z_k \|$$

$$\leq \| g_{k+1} \| + 2\frac{\left\| g_{k+1} - \frac{\| g_{k+1} \|}{\| g_k \|} g_k \right\|}{d_k^T y_k} \| g_{k+1} \| \| d_k \|$$

$$\leq \| g_{k+1} \| + 2\frac{\| g_{k+1} - g_k \| + \left\| g_k - \frac{\| g_{k+1} \|}{\| g_k \|} g_k \right\|}{d_k^T y_k} \| g_{k+1} \| \| d_k \|$$

$$\leq \| g_{k+1} \| + 2\frac{\| g_{k+1} - g_k \| + \|\| g_k \| - \| g_{k+1} \|\|}{d_k^T y_k} \| g_{k+1} \|\| d_k \|$$

$$\leq \| g_{k+1} \| + 4\frac{\| g_{k+1} - g_k \|}{s_k^T y_k} \| g_{k+1} \|\| s_k \| \leq \tau + 4\tau\frac{L}{\nu^2}$$

Hence, taking Lemma 3.1 of [38] into account, the proof is complete.

3. Numerical experiments

Here, we computationally investigate efficiency of the CG methods TTMHS with the search direction (7), MHS with the parameter (3) and TTHS with the search direction (2). The software and hardware information are given in [3]. Besides, test functions information, including 81 problems of the CUTEr library [27] has been provided in Table 1.

Table 1. Test problems data

Function	n	Function	n	Function	n
ARGLINA	200	DIXMAANK	3000	MANCINO	100
BDEXP	5000	DIXMAANL	3000	MOREBV	5000
BIGGSB1	5000	DIXON3DQ	10000	MSQRTALS	1024
BQPGABIM	50	DMN15103	99	MSQRTBLS	1024
BQPGASIM	50	DQDRTIC	5000	NCB20	5010
BROYDN7D	5000	DQRTIC	5000	NCB20B	5000
BRYBND	5000	DRCAV1LQ	4489	NONCVXU2	5000
CHAINWOO	4000	DRCAV2LQ	4489	NONDQUAR	5000
CHENHARK	5000	DRCAV3LQ	4489	PENALTY2	200
CHNROSNB	50	EDENSCH	2000	POWELLSG	5000
CLPLATEB	5041	EG2	1000	POWER	10000
COSINE	10000	EIGENALS	2550	QUARTC	5000
CRAGGLVY	5000	EIGENBLS	2550	SCHMVETT	5000
CURLY10	10000	EIGENCLS	2652	SENSORS	100
CURLY20	10000	ENGVAL1	5000	SINQUAD	5000
CURLY30	10000	ERRINROS	50	SPARSQUR	10000
DECONVU	63	EXTROSNB	1000	SPMSRTLS	4999
DIXMAANA	3000	FLETCBV2	5000	SROSENBR	5000
DIXMAANB	3000	FLETCBV3	5000	TESTQUAD	5000
DIXMAANC	3000	FLETCHBV	5000	TOINTGOR	50
DIXMAAND	3000	FLETCHCR	1000	TOINTGSS	5000
DIXMAANE	3000	FMINSRF2	5625	TOINTPSP	50
DIXMAANF	3000	FMINSURF	5625	TOINTQOR	50
DIXMAANG	3000	FREUROTH	5000	TRIDIA	5000
DIXMAANH	3000	GENHUMPS	5000	VARDIM	200
DIXMAANI	3000	GENROSE	500	VAREIGVL	50
DIXMAANJ	3000	LIARWHD	5000	WOODS	4000

The approximate Wolfe conditions of [28] have been employed in our implementations with the same values of the parameters. The solution process were ended up when $k > 10000$ or $||g_k|| < 10^{-6}(1 + |f(x_k)|)$ .To assess quality of the outputs, we used the performance profile of [20], with the notations of [5], on TNFGE (the total number of function and gradient evaluations [28]), and the CPUT (CPU time in second). To describe the results in vivid details, Figures 1 and 2 show that with respect to TNFGE, MHS outperforms TTHS and also, TTMHS is superior to both of them. Whereas, with respect to CPUT, TTHS and MHS are competitive and they are more time consuming than TTMHS. So, as affirmed by Figures 1 and 2, TTMHS outperforms the others.

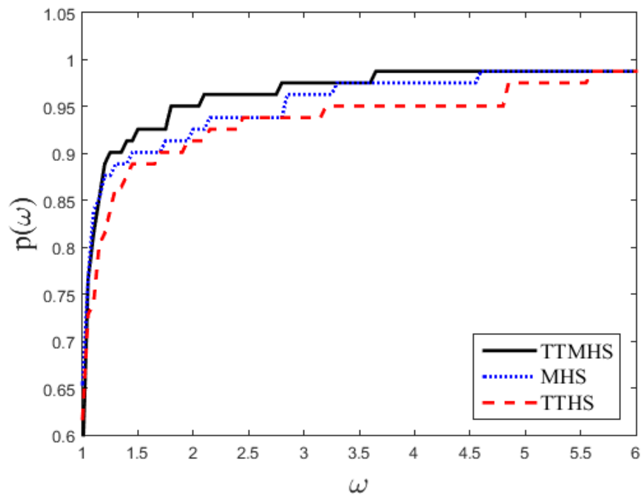

Figure 1. Results of comparisons based on TNFGE

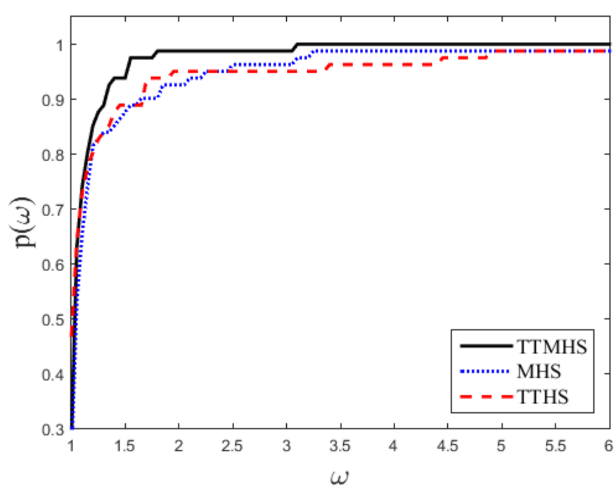

Figure 2. Results of comparisons based on CPUT

4. Conclusions and future works

As known, developing three-term extensions of the classical conjugate gradient methods in order to achieve the sufficient descent property has attracted special attentions [38]. Here, based on the insight gained by the approach of [42], a three-term extension of the nonlinear conjugate gradient method proposed by Shengwei et al. [37] has been suggested. It has been shown that the method fulfils the effective sufficient descent condition independent to the line search. Moreover, it has been established that the proposed method is globally convergent for uniformly convex functions. Using a set of standard test functions of the CUTEr library [27], numerical experiments have been implemented to investigate efficiency of the given method. The results have been compared using the Dolan-Moré [20] performance profile. It has been observed that the proposed method is computationally promising.

As a final note, the convergence of the proposed method relies on convexity of the objective function as well as employing the Wolfe line search strategy. As a future work, one can establish the convergence of the proposed method for general functions regardless of line search technique.

Conflict of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Abubakar, A.B., Kumam, P., & Awwal, A.M. (2019). Global convergence via descent modified three-term conjugate gradient projection algorithm with applications to signal recovery. *Results in Applied Mathematics*, 4,100069.
2. Amini, K., Faramarzi, P., & Pirfalah, N. (2019). A modified Hestenes–Stiefel conjugate gradient method with an optimal property. *Optimization Methods and Software*, 34(4), 770–782.
3. Aminifard, Z., & Babaie–Kafaki, S., (2019). Matrix analyses on the Dai–Liao conjugate gradient method. *The ANZIAM Journal*, 61(2), 195–203.
4. Aminifard, Z. & Babaie–Kafaki, S., (2019). A modified descent Polak–Ribière–Polyak conjugate gradient method with global convergence property for nonconvex functions. *Calcolo*, 56(2), 1-11.
5. Aminifard, Z. & Babaie–Kafaki, S., (2019). An optimal parameter choice for the Dai–Liao family of conjugate gradient methods by avoiding a direction of the maximum magnification by the search direction matrix. *4OR*, 17, 317–330.
6. Aminifard, Z. & Babaie–Kafaki, S., (2020). Modified spectral conjugate gradient methods based on the quasi–Newton aspects. *Pacific Journal of Optimization*, 16(4), 581–594.
7. Andrei, N., (2008). Another hybrid conjugate gradient algorithm for unconstrained optimization. *Numerical Algorithms*, 47(2), 143–156.
8. Andrei, N., (2010). Accelerated scaled memoryless BFGS preconditioned conjugate gradient algorithm for unconstrained optimization. *The European Journal of Operational Research*, 204(3), 410–420.
9. Awwal, A.M., Kumam, P., & Abubakar, A.B., (2019). A modified conjugate gradient method for monotone nonlinear equations with convex constraints. *Applied Numerical Mathematics*, 145, 507–520.
10. Babaie–Kafaki, S., & Ghanbari, R., (2014). A descent extension of the Polak–Ribière–Polyak conjugate gradient method. *Computers & Mathematics with Applications*, 68(12), 2005–2011.
11. Babaie–Kafaki, S., & Ghanbari, R., (2014). A descent family of Dai–Liao conjugate gradient methods. *Optimization Methods and Software*, 29(3), 583–591.
12. Babaie–Kafaki, S., & Ghanbari, R., (2014). A modified scaled conjugate gradient method with global convergence for nonconvex functions. The *Bulletin* of the *Belgian Mathematical Society*, 21(3), 465–477.
13. Babaie–Kafaki, S., & Ghanbari, R., (2017). Extensions of the Hestenes–Stiefel and Polak–Ribière–Polyak conjugate gradient methods with sufficient descent property. The *Bulletin* of the *Iranian Mathematical Society*, 43(7), 2437-2448.
14. Babaie–Kafaki, S., & Ghanbari, R., (2014). Two modified three-term conjugate gradient methods with sufficient descent property. *Optimization Letters*, 8(8), 2285–2297.
15. Babaie–Kafaki, S., & Ghanbari, R., (2016). An adaptive Hager–Zhang conjugate gradient method. *Filomat*, 30(14), 3715–3723.
16. Bojari S., & Eslahchi, M.R., (2020). Two families of scaled three–term conjugate gradient methods with sufficient descent property for nonconvex optimization. *Numerical Algorithms*, 83(11), 901–933.
17. Dai, Y.H., Han, J.Y., Liu, G.H., Sun, D.F., Yin, H.X., & Yuan, Y.X., (1999). Convergence properties of nonlinear conjugate gradient methods. *SIAM Journal on Optimization*., 10(2), 348–358.
18. De Leeuw den Bouter, M.L., van Gijzen, M.B., & Remis, R.F., (2019). Conjugate gradient variants for ℓp–regularized image reconstruction in low–field MRI. *SN Applied Sciences*, 1, 1736.
19. Dehghani, R., & Mahdavi–Amiri, N., (2019). Scaled nonlinear conjugate gradient methods for nonlinear least squares problems. *Numerical Algorithms*, 82, 1–20.
20. Dolan, E.D. & Moré, J.J., (2002). Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2, Ser. A), 201–213.
21. Dong, X.L., Han, D.R., Ghanbari, R., Li, X.L., & Dai, Z.F., (2017). Some new three–term Hestenes–Stiefel conjugate gradient methods with affine combination. *Optimization*, 66(5), 759–776.
22. Dong, X.L., Liu, H.W., He, Y.B., Babaie-Kafaki, S., & Ghanbari, R., (2016). A new three–term conjugate gradient method with descent direction for unconstrained optimization. *Mathematical Modelling and Analysis*, 21(3), 399–411.
23. Dong, X.L., Liu, H.W., He, Y.B., & Yang, X.M., (2015). A modified Hestenes–Stiefel conjugate gradient method with sufficient descent condition and conjugacy condition. *Journal of Computational and Applied Mathematics*, 28, 239–249.
24. Esmaeili, H., Shabani, S., & Kimiaei, M., (2019). A new generalized shrinkage conjugate gradient method for sparse recovery. *Calcolo*, 56(1), 1–38.

25. Faramarzi, P., & Amini, K., (2019). A modified spectral conjugate gradient method with global convergence. *Journal* of *Optimization Theory and Applications*, 182, 667–690.
26. Gilbert, J.C., & Nocedal, J., (1992). Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on Optimization*, 2(1), 21–42.
27. Gould, N.I.M., Orban, D., & Toint, Ph.L., (2003). CUTEr: a constrained and unconstrained testing environment, revisited. *ACM Transactions* on *Mathematical Software*, 29(4), 373–394.
28. Hager W.W., & Zhang, H., (2006). Algorithm 851: CG−Descent, a conjugate gradient method with guaranteed descent. *ACM Transactions* on *Mathematical Software*, 32(1), 113–137.
29. Hager W.W., & Zhang, H., (2006). A survey of nonlinear conjugate gradient methods. *Pacific Journal of Optimization*, 2(1), 35–58.
30. Hestenes, M.R. & Stiefel, E., (1952). Methods of conjugate gradients for solving linear systems. *Journal* of *Research* of the *National Bureau* of *Standards*, 49(6), 409–436.
31. Li, W., Liu, Y., Yang, J., & Wu, W., (2018). A new conjugate gradient method with smoothing L1/2 regularization based on a modified secant equation for training neural networks. *Neural Processing Letters*, 48:955–978.
32. Li, X., Zhang, W., & Dong, X., (2017). A class of modified FR conjugate gradient method and applications to non–negative matrix factorization. *Computers & Mathematics with Applications*, 73, 270–276.
33. Lin, N., Chen, Y., & Lu, L., (2020). Mineral potential mapping using a conjugate gradient logistic regression model. *Natural Resources Research*, 29, 173–188.
34. Liu, Y., Zhang, L., & Lian, Z., (2018). Conjugate gradient algorithm in the four–dimensional variational data assimilation system in *GRAPES Journal of Meteorol. Research*, 32, 974–984.
35. Mirhoseini, N., Babaie-Kafaki, S., & Aminifard, Z., (2022). A nonmonotone scaled Fletcher–Reeves conjugate gradient method with application in image reconstruction. *Bulletin of the Malaysian Mathematical Sciences Society*, 1–20.
36. Nezhadhosein, S., (2020). New nonlinear conjugate gradient methods based on optimal Dai–Liao parameters. *Journal of Mathematical Modeling*, 8(1), 21–39.
37. Shengwei, Y., Wei, Z., & Huang, H., (2007). A note about WYL's conjugate gradient method and its applications. *Applied Mathematics and Computation*, 191(2), 381–388.
38. Sugiki, K., Narushima, Y., & Yabe, H., (2012). Globally convergent three–term conjugate gradient methods that use secant conditions and generate descent search directions for unconstrained optimization. *Journal* of *Optimization Theory* and *Applications*, 153(3), 733–757.
39. Sun, W., & Yuan, Y.X., (2006). *Optimization Theory and Methods: Nonlinear Programming*. Springer, New York.
40. Wan, Z., Guo, J., Liu, J., & Liu, W., (2018). A modified spectral conjugate gradient projection method for signal recovery. *Signal, Image and Video Processing*, 12, 1455–1462.
41. Wang, X.Y., Li, S.J., & Kou, X.P., (2016). A self–adaptive three–term conjugate gradient method for monotone nonlinear equations with convex constraints. *Calcolo*, 53, 133–145.
42. Wei, Z.X., Yao, S.W., & Liu, L.Y., (2006). The convergence properties of some new conjugate gradient methods. *Applied Mathematics and Computation*, 183:1341–1350.
43. Yuan, G., Li, T., & Hu, W., (2020). A conjugate gradient algorithm for large–scale nonlinear equations and image restoration problems. *Applied Numerical Mathematics*, 147, 129–141.
44. Zhang, L., Zhou, W., & Li, D.H., (2007). Some descent three–term conjugate gradient methods and their global convergence. *Optimization Methods and Software*, 22(4), 697–711.

Aminifard, Z. (2022). A Three-Term Extension of a Descent Conjugate Gradient Method. *Fuzzy Optimization and Modelling Journal*, 3(2), 15-22.

https://doi.org/10.30495/fomj.2022.1953254.1063

Received: 21 February 2022 Revised: 3 July 2022 Accepted: 17 July 2022