EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 16, No. 1, 2023, 233-242
ISSN 1307-5543 – ejpam.com
Published by New York Business Global

A New Scalar of Conjugate Gradient Methods for Solving Unconstrained Minimization

Saja O. Mohammad T.[1], Hamsa Th. Saeed Chilmeran[2], Rana Z. Al-Kawaz[3,*]

[1] *Department of Mathematics, College of Basic Education, University of Mosul, Mosul, Iraq*
[2] *Department of Mathematics, College of Computer Sciences and Mathematics, University of Mosul, Iraq*
[3] *Department of Mathematics, College of Basic Education, University of Telafer, Mosul, Iraq*

Abstract. In this paper, we derive a search direction for the conjugate-gradient method based on the use of the self-scaling Quasi Newton-method, and the usefulness of the new method is to solve unconstrained optimization problems with large dimensions. To clarify the importance of the proposed method, we have shown its characteristics in terms of the sufficient descent condition and the theoretically global convergence condition. Numerically, we applied the proposed method to a variety of known test functions to prove its effectiveness. When compared with some previous methods in the same direction, the proposed method proved to be superior to them that the tools used for this purpose.

2020 Mathematics Subject Classifications: 65K05, 90C53, 65K10, 49M37

Key Words and Phrases: Conjugate-gradient, self-scaling, Quasi Newton-method, sufficient descent, global convergence

1. Introduction

Optimization methods are divided into two types: methods that depend on the existence of a target function and a set of restrictions, regardless of the number of variables used in the issue, and they have much wide life and engineering applications to this day called (constrained optimization methods) and a second type that depends only on the presence of an objective function without restrictions is called (unconstrained optimization methods) also, regardless of the number of the issue variables, and it has wider applications than the first type. The conjugate-gradient methods were influential in the field of finding minimization for quadratic and convex functions because they do not need to store

*Corresponding author.
DOI: https://doi.org/10.29020/nybg.ejpam.v16i1.4619

Email addresses: saja-othmanmt@uomosul.edu.iq (S. Mohammad T.), hamsathrot@uomosul.edu.iq (H. Chilmeran), rana.alkawaz@yahoo.com, ranazidan@uotelafer.edu.iq (R. Al-Kawaz)

any matrix during their implementation. An unconstrained optimization problem known as minimizing a function is defined as:

$$\text{minimize}_{x\in R^n} f(x) \tag{1}$$

As the function in 1 is for the real variable, which can use iterative methods to solve it, such as the conjugate-gradient method (CG) or Newton's methods or Quasi-Newton (QN), which use the update new-point through:

$$x_{k+1} = x_k + \tau_k p_k \tag{2}$$

Most of the iterative methods focus on the length of the step τ_k specified for the necessary distance to the search direction p_{k+1} which is obtained from (CG) i.e.:

$$p_{k+1} = -\mu_{k+1} + \gamma_k p_k \tag{3}$$

Where $\mu_{k+1} = \nabla f(x_{k+1})$. In the past several decades, γ_k has been defined in several scientific papers such as [15, 17, 18, 21]. To ensure the convergence and descent direction of this method when solving different types of general functions, we need to calculate the step length with one of the inexact lines search such as the Wolfe line search (WLS), from which two main types emerge:

(i) Weak type called Wolfe's weak line search (WWLS) and it defines:

$$f_{k+1} - f_k \leq \delta \tau_k {\mu_k}^T p_k \tag{4}$$

$${\mu_{k+1}}^T p_k \geq \sigma {\mu_k}^T p_k \tag{5}$$

(ii) A strong type is called Wolfe's strong line search (WSLS) and it defines as:

$$f_{k+1} - f_k \leq \delta \tau_k {\mu_k}^T p_k \tag{6}$$

$$|{\mu_{k+1}}^T p_k| \leq -\sigma {\mu_k}^T p_k \tag{7}$$

Where $0 < \delta < \sigma < 1$. This is now a very mature field to be now converted into many updated and expanded formats and with different applications as in [1–3, 8, 11–14, 19]. Later, the field of application gradually expanded to include conjugate gradient directions for solving non-linear systems as in [5–7].

2. A New Parameter for γ_k

In this experiment, we introduce a new parameter, γ_k, to improve the CG orientation based on a Quasi-Newton update of matrix approximation A_k known as a modified (DFP-QN) consisting of:

$$A_{k+1} = A_k - \frac{A_k \omega_k \omega_k^T A_k}{\omega_k^T A_k \omega_k} + \frac{s_k s_k^T}{s_k^T \omega_k} \tag{8}$$

where $s_k = x_{k+1} - x_k$, $y_k = \mu_{k+1} - \mu_k$ and

$$\omega_k = y_k + \frac{\rho_k}{\|s_k\|^2} s_k \tag{9}$$

$$\rho_k = 2[f_k - f_{k+1}] + (\mu_{k+1} + \mu_k)^T s_k \tag{10}$$

To learn more about equations (9) and (10) you can see [22]. We suggest multiplying the updated DFB matrix by self-scaling $\eta_{k+1} = \frac{y_k^T s_k}{\mu_k^T A_k \mu_k}$ which is defined in [4] to get:

$$A_{k+1} = \frac{y_k^T s_k}{\mu_k^T A_k \mu_k} \left(A_k - \frac{A_k \omega_k \omega_k^T A_k}{\omega_k^T A_k \omega_k} + \frac{s_k s_k^T}{s_k^T \omega_k} \right) \tag{11}$$

Also,replace y_k in the η_{k+1} formula to the ω_k. As for the search direction for a method QN, it is:

$$p_{k+1} = -A_{k+1} \mu_{k+1} \tag{12}$$

Now we substitute equation (11) into the equation (12), that is:

$$p_{k+1} = -\left(\frac{\omega_k^T s_k}{\mu_k^T A_k \mu_k} \left(A_k - \frac{A_k \omega_k \omega_k^T A_k}{\omega_k^T A_k \omega_k} + \frac{s_k s_k^T}{s_k^T \omega_k} \right) \right) \mu_{k+1} \tag{13}$$

Then we make the search direction memoryless, i.e. by replacing each $A_k = I$, we get:

$$p_{k+1} = -\frac{\omega_k^T s_k}{\|\mu_k\|^2} \left(\mu_{k+1} - \frac{\omega_k^T \mu_{k+1}}{\|\omega_k\|^2} \omega_k + \frac{s_k^T \mu_{k+1}}{s_k^T \omega_k} s_k \right) \tag{14}$$

To increase the efficiency of the search direction for CG methods, we equate it with the new search direction for QN methods, and we deduce from the equation a new γ_k parameter as in the following steps:

$$-\mu_{k+1} + \gamma_k s_k = -\frac{\omega_k^T s_k}{\|\mu_k\|^2} \left(\mu_{k+1} - \frac{\omega_k^T \mu_{k+1}}{\|\omega_k\|^2} \omega_k + \frac{s_k^T \mu_{k+1}}{s_k^T \omega_k} s_k \right) \tag{15}$$

We multiply both sides of the equation (15) by ω_k to get:

$$-\omega_k^T \mu_{k+1} + \gamma_k \omega_k^T s_k = -\frac{\omega_k^T s_k}{\|\mu_k\|^2} \omega_k^T \mu_{k+1} + \frac{\omega_k^T s_k}{\|\mu_k\|^2} \frac{\omega_k^T \mu_{k+1}}{\|\omega_k\|^2} \omega_k^T \omega_k - \frac{\omega_k^T s_k}{\|\mu_k\|^2} \frac{s_k^T \mu_{k+1}}{s_k^T \omega_k} \omega_k^T s_k \tag{16}$$

By simplifying equation (16):

$$\gamma_k = \frac{\omega_k^T \mu_{k+1}}{\omega_k^T s_k} - \frac{\omega_k^T s_k}{\|\mu_k\|^2} \frac{\omega_k^T \mu_{k+1}}{\omega_k^T s_k} + \frac{\omega_k^T s_k}{\|\mu_k\|^2} \frac{\omega_k^T \mu_{k+1}}{\|\omega_k\|^2} \frac{\|\omega_k\|^2}{\omega_k^T s_k} - \frac{\omega_k^T s_k}{\|\mu_k\|^2} \frac{s_k^T \mu_{k+1}}{s_k^T \omega_k} \frac{\omega_k^T s_k}{\omega_k^T s_k} \tag{17}$$

then

$$\gamma_k^{SHR} = \frac{\omega_k^T \mu_{k+1}}{\omega_k^T s_k} - \frac{s_k^T \mu_{k+1}}{\|\mu_k\|^2} \tag{18}$$

Equation (18) is the new formula of the parameter γ_k^{SHR} that can be inserted into the search direction of the CG (3) to give a new search direction formula as in:

$$p_{k+1} = -\mu_{k+1} + \gamma_k^{SHR} s_k \tag{19}$$

The following theorem is to test the property of sufficient descent for the new search direction mentioned in the equations (18) and (19) that is:

Theorem 1. *Suppose x_k is generated by the new algorithm in equations* (18) *and* (19) *with the step τ_k is computed by applying the (WSLS) condition in* (6) *and* (7) *then the direction is sufficiently-descent i.e.:*

$$p_{k+1}^T \mu_{k+1} \leq -(1-u)\left\|\mu_{k+1}\right\|^2 \tag{20}$$

Proof. From the new search direction equation (19) and multiplying it from both directions by $\left(\frac{\mu_{k+1}}{\|\mu_{k+1}\|^2}\right)$ we get:

$$\frac{p_{k+1}^T \mu_{k+1}}{\|\mu_{k+1}\|^2} + 1 = \left(\frac{\omega_k^T \mu_{k+1}}{\omega_k^T s_k} - \frac{s_k^T \mu_{k+1}}{\|\mu_k\|^2}\right) \frac{s_k^T \mu_{k+1}}{\|\mu_{k+1}\|^2} \tag{21}$$

Since $\omega_k^T \mu_{k+1} \leq \|\omega_k\| \|\mu_{k+1}\|$ and $s_k^T \mu_{k+1} \leq s_k^T \omega_k$, that is

$$\frac{p_{k+1}^T \mu_{k+1}}{\|\mu_{k+1}\|^2} + 1 \leq \left(\frac{\|\omega_k\| \|\mu_{k+1}\|}{\omega_k^T s_k} - \frac{s_k^T \omega_k}{\|\mu_k\|^2}\right) \frac{s_k^T \omega_k}{\|\mu_{k+1}\|^2}$$

$$\frac{p_{k+1}^T \mu_{k+1}}{\|\mu_{k+1}\|^2} + 1 \leq \frac{\|\omega_k\| \|\mu_{k+1}\|}{\|\mu_{k+1}\|^2} - \frac{\left(s_k^T \omega_k\right)^2}{\|\mu_k\|^2 \|\mu_{k+1}\|^2}$$

$$\frac{p_{k+1}^T \mu_{k+1}}{\|\mu_{k+1}\|^2} + 1 \leq \frac{\|\omega_k\|}{\|\mu_{k+1}\|} - \frac{\left(s_k^T \omega_k\right)^2}{\|\mu_k\|^2 \|\mu_{k+1}\|^2}$$

$$\frac{p_{k+1}^T \mu_{k+1}}{\|\mu_{k+1}\|^2} + 1 \leq \frac{\|\omega_k\|}{\|\mu_{k+1}\|}$$

let $u = \frac{\|\omega_k\|}{\|\mu_{k+1}\|}$. So we got the equation (20) we need to prove.

3. Global Convergence Property

Before starting this section, we need to make some assumptions necessary to prove the convergence of the new algorithm (SHR-CG), as shown:

3.1. Assumption[20]

(i) Assume $f(x)$ is bound on the level specified by $\varphi = \{x \in \mathbb{R}^n, f(x) \leq f(x_0)\}$ from below. Whereas x_0 is the point of departure.

(ii) In some neighborhoods N of φ, the objective function is continuously-differentiable, and its gradient is continuous-Lipschitz, which is, there is such a constant $L > 0$ such that:

$$\|\mu(x) - \mu(y)\| \leq L\|x - y\|, \forall x, y \in \varphi. \tag{22}$$

We can conclude that $\zeta, \bar{\zeta} > 0$ exists by applying assumptions (i) and (ii) to f.

$$\zeta \leq \|\mu(x)\| \leq \bar{\zeta} \tag{23}$$

(iii) The following requirement is fulfilled:

$$\| \left(\mu(x) - \mu(y)\right)^T (x - y)\| \geq \varsigma\|s_k\|^2, \forall x, y \in \varphi, \varsigma > 0. \tag{24}$$

Lemma 1. *If the assumption is correct and each conjugated gradient method has a descending direction p_{k+1} with a step length τ_k of equations* (6) *and* (7) *if so,*

$$\sum_{k\geq 1} \frac{1}{\|p_{k+1}\|^2} = \infty \tag{25}$$

So the next equation is right:

$$\liminf_{k\to\infty} \|\mu_k\| = 0 \tag{26}$$

Theorem 2. *If the assumption is adhered to and the new direction p_{k+1} described by equation* (19) *is descent, then*

$$\liminf_{k\to\infty} \|\mu_k\| = 0 \tag{27}$$

Proof. From the formula of the new algorithm, we take the absolute value of the proposed parameter γ_k^{SHR} i.e. equation (18) as in:

$$|\gamma_k^{SHR}| = \left| \frac{\omega_k^T \mu_{k+1}}{\omega_k^T s_k} - \frac{s_k^T \mu_{k+1}}{\|\mu_k\|^2} \right|$$

$$|\gamma_k^{SHR}| \leq \left| \frac{\omega_k^T \mu_{k+1}}{\omega_k^T s_k} \right| + \left| \frac{s_k^T \mu_{k+1}}{\|\mu_k\|^2} \right| \tag{28}$$

Since $\omega_k^T \mu_{k+1} \leq \|\omega_k\|\|\mu_{k+1}\|$ and from equation (7) substitute it into equation (28) i.e.:

$$|\gamma_k^{SHR}| \leq \frac{\|\omega_k\|\|\mu_{k+1}\|}{\varsigma\|s_k\|^2} + \left| \frac{\sigma \mu_k^T p_k}{\|\mu_k\|^2} \right| \tag{29}$$

Since $p_k = -\mu_k$ and $\omega_k = y_k + \frac{\rho_k}{\|s_k\|^2} s_k$

$$|\gamma_k^{SHR}| \leq \frac{\|y_k + \frac{\rho_k}{\|s_k\|^2} s_k\| \|\mu_{k+1}\|}{\varsigma \|s_k\|^2} + \frac{\sigma \|\mu_k\| \|s_k\|}{\|\mu_k\|^2}$$

$$|\gamma_k^{SHR}| \leq \frac{\|y_k\| \|\mu_{k+1}\| + \frac{\rho_k}{\|s_k\|^2} \|s_k\| \|\mu_{k+1}\|}{\varsigma \|s_k\|^2} + \frac{\sigma \|s_k\|}{\|\mu_k\|} \tag{30}$$

from the Lipschitz condition:

$$|\gamma_k^{SHR}| \leq \frac{L\|s_k\| \|\mu_{k+1}\| \|s_k\|^2 + \rho_k \|s_k\| \|\mu_{k+1}\|}{\varsigma \|s_k\|^4} + \frac{\sigma \|s_k\|}{\|\mu_k\|}$$

$$|\gamma_k^{SHR}| \leq \frac{L\|\mu_{k+1}\|}{\varsigma \|s_k\|} + \frac{\rho_k \|\mu_{k+1}\|}{\varsigma \|s_k\|^3} + \frac{\sigma \|s_k\|}{\|\mu_k\|} \tag{31}$$

As the assumption is fulfilled

$$|\gamma_k^{SHR}| \leq \frac{L\zeta}{\varsigma \|s_k\|} + \frac{\rho_k \zeta}{\varsigma \|s_k\|^3} + \sigma \tau_k \tag{32}$$

let $\frac{L\zeta}{\varsigma \|s_k\|} + \frac{\rho_k \zeta}{\varsigma \|s_k\|^3} + \sigma \tau_k = \xi$, by taking the absolute value of equation (19) we get:

$$\|p_{k+1}\| \leq \|\mu_{k+1} + \gamma_k^{SHR} s_k\|$$

$$\|p_{k+1}\| \leq \|\mu_{k+1}\| + \|\gamma_k^{SHR}\| \|s_k\|$$

$$\|p_{k+1}\| \leq \bar{\zeta} + \xi \|s_k\| = \breve{D}$$

$$\sum_{k \geq 1} \frac{1}{\|p_{k+1}\|^2} \geq \frac{1}{\breve{D}} \sum_{k \geq 1} 1 = \infty \tag{33}$$

$$\liminf_{k \to \infty} \|\mu_k\| = 0 \tag{34}$$

So the proof is complete.

4. Numerical Result

We present a numerical comparison in this section between the new algorithm proposed in equation (19) concerning (35) standard test functions from [9, 10] with the algorithms (HS) [21] and (HZ) [18]. The tests were based on the following tools:

- The number of times the iteration is calculated (ITC).
- The number of times the function and its derivative are calculated (FDC).
- The time it takes to execute for the processor (TEP).

These comparisons were carried out by code in the FORTRAN program and with a stop scale $\mu_k \leq 10^{-5}$. These results are plotted for better comparison based on the Dolan-More method [16]. We all of the algorithms mentioned, we tried on small (1000) and large dimensions (10000) of the variables with Wolfe's search line (WSLS), and the results were as in the following figures:

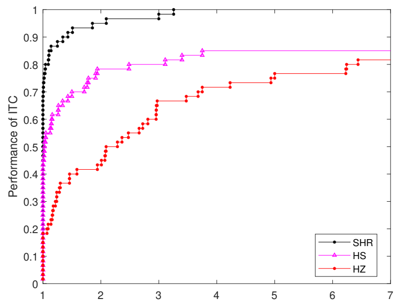

Figure 1: The number of times the iteration is calculated (ITC).

Figure 1 shows the progress of the new algorithm (SHR) over the basic algorithms of Hestenes and Stiefel (HS) and Hager and Zhang (HZ) about the calculated iterations of the test functions during the implementation by Dolan-More method.

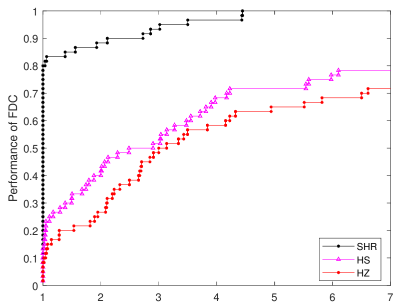

Figure 2: The number of times the function and its derivative are calculated (FDC).

In Figure 2, the new algorithm (SHR) is the best in performance and overcomes the basic algorithms concerning the number of times the function is calculated and it's derivative for each of the test functions computed during implementation through the Dolan-More method.

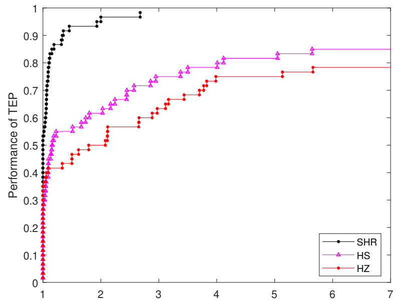

Figure 3: The time it takes to execute for the processor (TEP).

Finally, the Figure 3 confirms that the new algorithm (SHR) is the most efficient of the basic algorithms compared with it within the figure in relation to the time spent during the implementation of the mentioned algorithms.

5. Conclusions

This may be considered a promising aspect of the derivations in this manner used within the paper. The numerical results in the figures presented in the previous section show the distinction and efficiency of the new algorithm (SHR) when compared with the previous basic algorithms and within standard functions for this purpose. Also, for the theoretical side of this proposed algorithm, we have achieved sufficient-descent and its global-convergence within some assumptions.

Acknowledgements

The paper is supported by the College of Computer Science and Mathematics and the College of Basic Education at the University of Mosul and Talafar University in the Republic of Iraq.

References

[1] H I Ahmed, R Z Al-Kawaz, and A Y Al-Bayati. Spectral three-term constrained conjugate gradient algorithm for function minimizations. *Journal of Applied Mathematics*, 2019, 2019.

[2] H I Ahmed and Gh M Al-Naemi. A modified dai-yuan conjugate gradient method and its global convergence. *Iraqi Journal of Science*, 53(3):620–628, 2012.

[3] H I Ahmed, E T Hamed, and H Th Saeed Chilmeran. A modified bat algorithm with conjugate gradient method for global optimization. *International Journal of Mathematics and Mathematical Sciences*, 2020, 2020.

[4] A Y Al-Bayati and M S Al-Salih. New vm-methods for nonlinear unconstrained optimization. *Journal of Education and Science, Mosul University*, 21:169–182, 1994.

[5] R Z Al-Kawaz and A Y Al-Bayati. An adaptive search direction algorithm for the modified projection minimization optimization. In *IOP Conference Series: Materials Science and Engineering*, volume 1152, page 012019. IOP Publishing, 2021.

[6] R Z Al-Kawaz and A Y Al-Bayati. An effective new iterative cg-method to solve unconstrained non-linear optimization issues. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 19(6):1847–1856, 2021.

[7] R Z Al-Kawaz and A Y Al-Bayati. Show off the efficiency of the dai-liao method in merging technology for monotonous non-linear problems. *Indonesian Journal of Electrical Engineering and Computer Sciences*, 21(1):505–515, 2021.

[8] R Z Al-Kawaz, A Y Al-Bayati, and M S Jameel. Interaction between un updated fr-cg algorithms with an optimal cuckoo algorithm. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(3):1497–1504, 2020.

[9] N Andrei. An unconstrained optimization test functions collection. *Adv. Model. Optim*, 10(1):147–161, 2008.

[10] N Andrei. An adaptive conjugate gradient algorithm for large-scale unconstrained optimization. *Journal of Computational and Applied Mathematics*, 292:83–91, 2016.

[11] H Th Saeed Chilmeran, H I Ahmed, E T Hamed, and A Y Al-Bayati. A new modification of the quasi-newton method for unconstrained optimization. *Indonesian Journal of Electrical Engineering and Computer Science*, 21(3):1683–1691, 2021.

[12] H Th Saeed Chilmeran, E T Hamed, H I Ahmed, and A Y Al-Bayati. A method of two new augmented lagrange multiplier versions for solving constrained problems. *International Journal of Mathematics and Mathematical Sciences*, 2022, 2022.

[13] H Th Saeed Chilmeran, E S Khaleel, H I Ahmed, and E T Hamed. A modified sufficient descent spectral conjugate gradient method for minimization. In *2021 7th International Conference on Contemporary Information Technology and Mathematics (ICCITM)*, pages 239–242. IEEE, 2021.

[14] H Th Saeed Chilmeran and H Y Najem. New conjugacy coefficient for conjugate gradient method for unconstrained optimization. *AL-Rafidain Journal of Computer Sciences and Mathematics*, 10(2):33–46, 2013.

[15] Y H Dai and Y Yuan. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on optimization*, 10(1):177–182, 1999.

[16] E D Dolan and J J Moré. Benchmarking optimization software with performance profiles. *Mathematical programming*, 91(2):201–213, 2002.

[17] R Fletcher and C M Reeves. Function minimization by conjugate gradients. *The computer journal*, 7(2):149–154, 1964.

[18] W W Hager and H Zhang. A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM Journal on optimization*, 16(1):170–192, 2005.

[19] E T Hamed, R Z Al-Kawaz, and A Y Al-Bayati. New investigation for the liu-story scaled conjugate gradient method for nonlinear optimization. *Journal of Mathematics*, 2020, 2020.

[20] M Hamoda, M Mamat, M Rivaie, and Z Salleh. A conjugate gradient method with strong wolfe-powell line search for unconstrained optimization. *Appl. Math. Sci*, 10:721–734, 2016.

[21] M R Hestenes and E Stiefel. Methods of conjugate gradients for solving. *Journal of research of the National Bureau of Standards*, 49(6):409, 1952.

[22] Z Wei, G Yu, G Yuan, and Z Lian. The superlinear convergence of a modified bfgs-type method for unconstrained optimization. *Computational optimization and applications*, 29(3):315–332, 2004.