International Journal of Mathematics and
Computer Science, **17**(2022), no. 3, 1241–1250

A New Investigation of the Conjugate Gradient Method for Solving Unconstrained Optimization Problems

Huda Y. Najm[1], **Huda I. Ahmed**[2]

[1]Department of Mathematics
College of Science
University of Duhok
Duhok-Kurdistan Region, Iraq

[2]Department of Operation Research and Intelligent Techniques
College of Computers Sciences and Mathematics
University of Mosul
Mosul, Iraq

email: huda.najm@uod.ac, hudaea72@gmail.com

(Received January 19, 2022, Accepted April 7, 2022)

Abstract

The Conjugate gradient (CG) method has played a valuable and influential role in solving large-scale unconstrained optimization problems due to its simplicity and low memory needs. Based on this, we propose a new parameter that enhances the Dai-Liao formula. The presented parameter has the following characteristics: Sufficient descent condition and global convergence are achieved under a robust search for the Wolfe-Powell line. The numerical computational results indicate that the proposed algorithm is more effective and efficient than other formulas.

Key words and phrases: Conjugate Gradient method, Qusai-Newton method, Search Direction, Sufficient descent condition, Global convergence.
AMS (MOS) Subject Classifications: 65k10.
ISSN 1814-0432, 2022, http://ijmcs.future-in-tech.net

1 Introduction

The numerical solution of nonlinear unconstrained problems is often obtained using optimization techniques. For large-scale unconstrained optimization problems, the conjugate gradient technique may be helpful. The benefit of this technique over Newton's method is that it does not need the second derivative or its approximation. In addition, the conjugate gradient technique is easy and basic. Let $f(x): R^n \to R$ be continuously differentiable and bounded below. Consider

$$\min_{x\in R^n} f(x). \quad (1.1)$$

We use the iteration approach expressed as:

$$x_{k+1} = x_k + \alpha_k d_k \quad (1.2)$$

The following equations describe a strong Wolfe-Powell (SWP) line search that may be used to determine the step length α_k:

$$f(x_{k+1}) - f(x_k) \le c_1 \alpha_k \nabla f(x_k)^T d_k \quad (1.3a)$$

$$\left|g_{k+1}{}^T d_k\right| \le -c_2 g_k{}^T d_k, \quad (1.3b)$$

where $g_k = g(x_k) = \nabla f(x_k)$ and $0 < c_1 < c_2 < 1$. The CG technique defines d_k as a search direction:

$$d_{k+1} = \left\{ \begin{array}{ccc} -g_{k+1} & ; & for \quad k = 0 \\ -g_{k+1} + \beta_k d_k & ; & for \quad k \ge 1 \end{array} \right\} \quad (1.4)$$

The conjugate gradient update parameter is referred to as $\beta_k \in R$ Many evaluations of β_k have been proposed which would lead to different classes of CG methods. The following are the most famous of CG methods: Hestenes-Steifel (HS)[1], Fletcher Reeves (FR)[2], Polak-Ribeire (PR)[3][4], Dai Yuan (DY)[5], and Dai and Liao (DL)[6], CG technique included a nonnegative parameter t into Perry's condition [7].

$$d_{k+1}^T y_k = -t\ g_{k+1}^T s_k, \quad (1.5)$$

where $s_k = x_{k+1} - x_k$ and $t \ge 0$. The CG parameter called β_k^{DL} [6].

$$\beta_k^{DL} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - t\frac{g_{k+1}^T s_k}{d_k^T y_k} = \beta_k^{HS} - t\frac{g_{k+1}^T s_k}{d_k^T y_k} \quad (1.6)$$

The remaining article is organized as follows:
In section 2, we derive a new parameter β_k representing a modification (improvement) of the DL-CG method. In section 3, we discuss the sufficient

descent and global convergence analysis of a new formula using the robust Wolfe line search. In section 4, we present some numerical results. Finally, in section 5 we conclude our paper.

2 Derivation of a new parameter

There are many studies of known minimization issues: Ibrahim et al. presented the following research strategy [10]:

$$d_{k+1} = -H_{k+1}g_{k+1} + \lambda_k d_k, \tag{2.7}$$

where H_{k+1} is the BFGS matrix and $\lambda_k = \eta \frac{g_{k+1}^T g_{k+1}}{g_{k+1}^T d_k}$ with $\eta \in (0, 1]$. In order to derive a new parameter of our method, we recall the well-known formula of Al-Bayati (QN- method)[11] with the updating matrix H_{k+1}.

$$H_{k+1} = H_k + [\frac{2y_k^T H_k y_k}{(s_k^T y_k)^2}]s_k s_k^T - [\frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k}] \tag{2.8}$$

when using the memoryless in (2.8). Then the matrix is defined by:

$$H_{k+1} = I + [\frac{2y_k^T y_k}{(s_k^T y_k)^2}]s_k s_k^T - [\frac{y_k s_k^T + s_k y_k^T}{s_k^T y_k}] \tag{2.9}$$

Substituting (2.9) into (2.7):

$$d_{k+1} = -g_{k+1} + [\frac{y_k^T g_{k+1}}{s_k^T y_k} - 2\frac{\|y_k\|^2}{(s_k^T y_k)^2} s_k^T g_{k+1}]s_k + \frac{s_k^T g_{k+1}}{s_k^T y_k} y_k + \lambda_k d_k \tag{2.10}$$

Equating equations (2.10) and (1.4), we get

$$-g_{k+1} + \beta_k d_k = -g_{k+1} + [\frac{y_k^T g_{k+1}}{s_k^T y_k} - 2\frac{\|y_k\|^2}{(s_k^T y_k)^2} s_k^T g_{k+1}]s_k + \frac{s_k^T g_{k+1}}{s_k^T y_k} y_k + \lambda_k d_k \tag{2.11}$$

Multiplying both sides of (2.11) by y_k^T, we get:

$$-y_k^T g_{k+1} + \beta_k y_k^T d_k = -y_k^T g_{k+1} + y_k^T g_{k+1} - 2\frac{\|y_k\|^2}{s_k^T y_k} s_k^T g_{k+1} + \frac{\|y_k\|^2}{s_k^T y_k} s_k^T g_{k+1} + \lambda_k y_k^T d_k \tag{2.12}$$

$$-y_k^T g_{k+1} + \beta_k y_k^T d_k = -\frac{\|y_k\|^2}{s_k^T y_k} s_k^T g_{k+1} + \lambda_k y_k^T d_k$$

Then the proposed conjugacy coefficient β_k is:

$$\beta_k^{NH} = \frac{y_k^T g_{k+1}}{y_k^T d_k} - \frac{\|y_k\|^2}{s_k^T y_k} \frac{s_k^T g_{k+1}}{y_k^T d_k} + \lambda_k, \tag{2.13}$$

where λ_k is a scalar [10] such that

$$\lambda_k = \eta \frac{\|g_{k+1}\|^2}{d_k^T g_{k+1}} \tag{2.14}$$

and $0 < \eta < 0.75$.

$$\beta_k^{NH} = \frac{y_k^T g_{k+1}}{y_k^T d_k} - \frac{\|y_k\|^2}{s_k^T y_k} \frac{s_k^T g_{k+1}}{y_k^T d_k} + \eta \frac{\|g_{k+1}\|^2}{d_k^T g_{k+1}} \tag{2.15}$$

Comparing the version β_k^{NH} with DL-CG defined by (1.6), we find that the parameter t is replaced by $\frac{\|y_k\|^2}{s_k^T y_k}$ as well as another limit defined in the form λ_k ((2.14)) which has an influential role in improving the performance of the conjugate gradient method.

3 Global convergence properties

Any technique must satisfy the descent requirement and the convergence criterion to be deemed effective and robust. The suggested scheme's convergence will be examined in this section.
The following assumptions must be made to investigate the convergence of these algorithms.

3.1 Assumptions (3.1)

a) The level set Ω is bounded.
b) The function is convex and continuously differentiable and its gradient $g(x)$ is Lipchitz continuous(i.e., there exists a constant $L > 0$ such that:

$$\|g(x) - g(y)\| \leq L\|x - y\|, \forall x, y \in \Omega. \tag{3.16}$$

In addition to this assumption, we may deduce that a positive constant B exists such that

$$\|g(x)\| \leq B, \forall x \in \Omega. \tag{3.17}$$

and when the function is uniformly convex, it will satisfy the relationship $y_k^T s_k \geq \mu |s_k|^2$ [12].

3.2 Lemma 1

Let α_k satisfy conditions (1.3a) and (1.3b). Then the search direction d_{k+1} is generated by:

$$d_{k+1} = -g_{k+1} + \beta_k^{NH} d_k \tag{3.18}$$

satisfying the condition:

$$d_{k+1}^T g_{k+1} \leq -\tau ||g_{k+1}||^2 \tag{3.19}$$

Proof: From (3.18), we get:

$$d_{k+1} = -g_{k+1} + [\frac{y_k^T g_{k+1}}{d_k^T y_k} - (\frac{||y_k||^2}{s_k^T y_k}\frac{s_k^T g_{k+1}}{d_k^T y_k}) + \lambda_k] d_k$$

Since $s_k = \alpha_k d_k$

$$d_{k+1} = -g_{k+1} + [\frac{y_k^T g_{k+1}}{d_k^T y_k} - \frac{||y_k||^2 d_k^T g_{k+1}}{(d_k^T y_k)^2} + \lambda_k] d_k \tag{3.20}$$

Multiplying (3.20) by $\frac{g_{k+1}}{||g_{k+1}||^2}$ gives:

$$\frac{d_{k+1}^T g_{k+1}}{||g_{k+1}||^2} = -1 + [\frac{y_k^T g_{k+1}}{d_k^T y_k} - \frac{||y_k||^2 d_k^T g_{k+1}}{(d_k^T y_k)^2} + \lambda_k] \frac{d_k^T g_{k+1}}{||g_{k+1}||^2}$$

$$\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} = -1 + [\frac{y_k^T g_{k+1}(d_k^T g_{k+1})}{d_k^T y_k \|g_{k+1}\|^2}] - \frac{\|y_k\|^2 (d_k^T g_{k+1})^2}{(d_k^T y_k)^2 \|g_{k+1}\|^2} + \lambda_k \frac{d_k^T g_{k+1}}{\|g_{k+1}\|^2} \tag{3.21}$$

Using $u^T v \leq \frac{1}{2}(||u||^2 + ||v||^2)$ in the second term of (3.21):

$$\frac{d_{k+1}^T g_{k+1}}{||g_{k+1}||^2} \leq -1 + \frac{1}{2}[\frac{1}{2}||g_{k+1}||^2 + \frac{2(d_k^T g_{k+1})^2}{(d_k^T y_k)^2}||y_k||^2] \frac{1}{||g_{k+1}||^2} - \frac{||y_k||^2 (d_k^T g_{k+1})^2}{(d_k^T y_k)^2 ||g_{k+1}||^2} + \lambda_k \frac{d_k^T g_{k+1}}{||g_{k+1}||^2}$$

$$\frac{d_{k+1}^T g_{k+1}}{||g_{k+1}||^2} \leq -1 + \frac{1}{4}\frac{||g_{k+1}||^2}{||g_{k+1}||^2} + \frac{(d_k^T g_{k+1})^2}{(d_k^T y_k)^2}\frac{||y_k||^2}{||g_{k+1}||^2} - \frac{||y_k||^2 (d_k^T g_{k+1})^2}{(d_k^T y_k)^2 ||g_{k+1}||^2} + \lambda_k \frac{d_k^T g_{k+1}}{||g_{k+1}||^2}$$

$$\frac{d_{k+1}^T g_{k+1}}{||g_{k+1}||^2} \leq -1 + \frac{1}{4} + \lambda_k \frac{d_k^T g_{k+1}}{||g_{k+1}||^2} \tag{3.22}$$

when $\lambda_k = \eta \frac{||g_{k+1}||^2}{d_k^T g_{k+1}}$, substitute λ_k into the (3.22):

$$\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} \leq -1 + \frac{1}{4} + \frac{\eta \|g_{k+1}\|^2}{d_k^T g_{k+1}} \frac{d_k^T g_{k+1}}{\|g_{k+1}\|^2}$$

$$\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} \leq -1 + \frac{1}{4} + \eta.$$

Let $\frac{1}{4} + \eta = r_2$,and $0 < r_2 < 0.75$. Then

$$\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} \leq -1 + r_2$$

$$d_{k+1}^T g_{k+1} \leq -(1 - r_2)\|g_{k+1}\|^2 \tag{3.23}$$

Thus, condition (3.19) is fulfilled.

3.3 Theorem 1

Assuming assumption(3.1) is valid, α_k satisfies conditions (1.3a) and (1.3b)and Lemma 1 holds, Then

$$\lim_{k\to\infty} \inf \|g_k\| = 0 \tag{3.24}$$

Proof: Rewriting the search direction equation in (1.4) by:

$$d_{k+1} = -g_{k+1} + \beta_k^{NH} d_k \tag{3.25}$$

where $\beta_k^{NH} = \beta_k + \lambda_k$.

$$\beta_k = [\frac{g_{k+1}^T y_k}{d_k^T y_k} - \frac{||y_k||^2}{s_k^T y_k}\frac{s_k^T g_{k+1}}{d_k^T y_k}] \tag{3.26}$$

$$|\beta_k| = |\frac{g_{k+1}^T y_k}{d_k^T y_k} - \frac{||y_k||^2}{s_k^T y_k}\frac{s_k^T g_{k+1}}{d_k^T y_k}|$$

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

since $s_k^T g_{k+1} \leq s_k^T y_k$, $\|y_k\| \leq L\|s_k\|$.

$$|\beta_k| \leq \frac{||g_{k+1}||L||s_k||}{|d_k^T y_k|} + \frac{L^2||s_k^2}{|d_k^T y_k|}$$

from condition (1.3b)

$$d_k^T y_k \geq -(1 - c_2)g_k^T d_k \geq (1 - c_2)c\|g_k\|^2$$

and from eq. (3.17)

$$|\beta_k| \leq \frac{BL||s_k||}{(1-c_2)c||g_k||^2} + \frac{L^2||s_k||^2}{(1-c_2)c||g_k||^2} = M_1$$

$$|\beta_k| \leq M_1 \tag{3.27}$$

when $\lambda_k = \eta \frac{\|g_{k+1}\|^2}{g_{k+1}^T d_k}$, we have:

$$|\lambda_k| = |\eta \frac{\|g_{k+1}\|^2}{g_{k+1}^T d_k}|$$

$$|\lambda_k| \leq |\eta| \frac{||g_{k+1}||^2}{g_{k+1}^T d_k|}$$

$$|\lambda_k| = |\eta| \frac{B^2}{\sigma||g_k||^2} = E$$

$$|\lambda_k| \leq E \tag{3.28}$$

$$||d_{k+1}|| \leq ||g_{k+1}|| + |\beta_k|||d_k|| + |\lambda_k|||d_k||$$

from (3.27) and (3.28)

$$\|d_{k+1}\| \leq B + M_1\|d_k\| + E\|d_k\| = D_1$$

$$\sum_{k\geq 1} \frac{1}{\|d_{k+1}\|^2} \geq \frac{1}{D_1^2} \sum_{k\geq 1} 1 = \infty.$$

Then eq. (3.24) is satisfied with λ_k.

4 Numerical results

To analyze the efficiency of a new technique β_k^{NH}, we selected some of the test functions. Some numerical findings are shown in this section by comparing our suggested CG-technique new1 with well-known techniques like Hestenes-Stiefel (HS) [1] and Dai- Liao (DL) [6], Hager and Zhang (HZ) [8], and Andrei (DLE)[9]. This metric was created to compare the performance of a group of solvers on a collection of 45 unconstrained nonlinear problems with dimensions $n = 100, 200, .., 1000$ [13] [9]. The stopping criteria are used $\|g_k\| \leq 10^{-6}$. All of the programs were written in Fortran and used double precision math. Figure 1 (a) illustrates that the new method achieves the

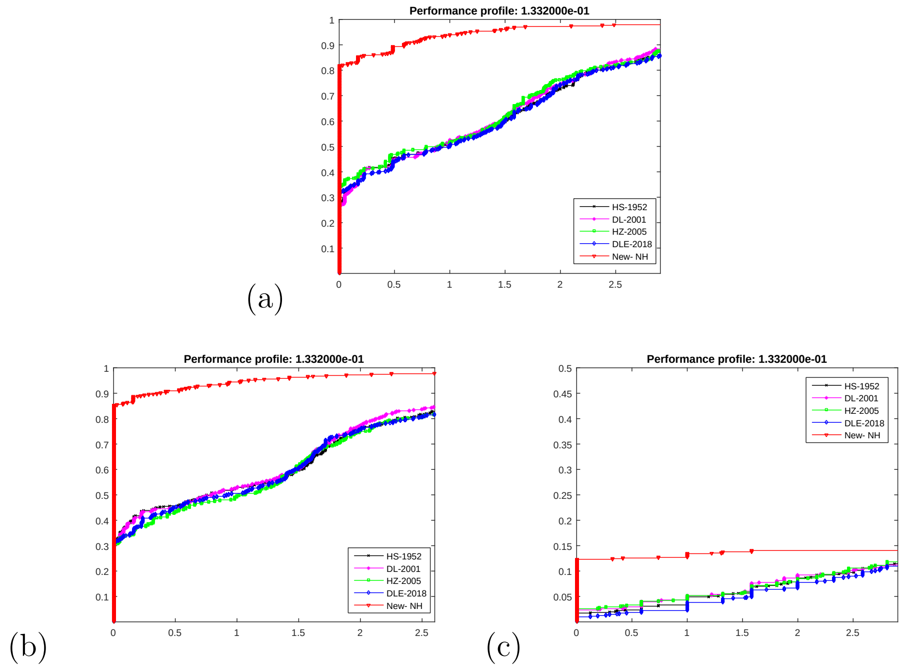

Figure 1: (a) Performance profile based on NOI. (b) Performance profile based on NOF. (c) Performance profile based on CPU time.

best result in the Number of iterations (NOI), as seen by the top curve in the graph Figure 1 (a). Also, Figures 1 (b) and 1 (c) illustrate the performance results, calculated Number of function evaluations (NOF), and running of CPU time, respectively. As a result, the new technique is more efficient than previous classical CG methods.

5 Conclusion

We gave a new conjugate gradient technique for unconstrained optimization problems improving the Dai-Liao type method where the search direction always meets the requirements of appropriate proportions and is globally convergent under strong Wolfe line search conditions. Our proposed approaches are better than many current technologies according to numerical results. The CG algorithm's numerical results have shown that our suggested approach is competitive compared to the other existing algorithms.

References

[1] M. R. Hestenes, E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bur. Standards, **49,** no. 6, (1952), 409–435.

[2] R. Fletcher, M. Colin, Function minimization by conjugate gradients. The computer Journal, Oxford University Press, **7,** no. 2, (1964), 149–154.

[3] Elijah Polak, Gerard Ribiere, Note sur la convergence de méthodes de directions conjuguées, The computer Journal, **3,** no. R1, (1969), 35–43.

[4] Boris Teodorovich Polyak,, The conjugate gradient method in extremal problems, USSR Computational Mathematics and Mathematical Physics, Elsevier, **9,** no. 4, (1969), 94–112.

[5] Yuhong Dai, Jiye Han, Guanghui Liu, Defeng Sun, Hongxia Yin, Yaxiang Yuan, Convergence properties of nonlinear conjugate gradient methods, SIAM Journal on Optimization, **10,** no. 2 ,(2000), 345–358.

[6] Y. H. Dai et al., New conjugacy conditions and related nonlinear conjugate gradient methods, Applied Mathematics and optimization, **43,** no. 1, 87–101.

[7] Avinoam Perry, A class of conjugate gradient algorithms with a two-step variable metric memory, (1977).

[8] William W. Hager, Hongchao Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, SIAM Journal on optimization, **16,** no.1 , (2005), 170–192.

[9] Neculai Andrei, A Dai-Liao conjugate gradient algorithm with clustering of eigenvalues, Numerical Algorithms **77,** no.4 , (2018), 1273-1282.

[10] Mohd Asrul Hery, Mustafa Ibrahim, L. W. June, BFGS method: a new search direction, Sains Malaysiana, **43,** no. 10, (2014), 1591–1597.

[11] A. Y. Alias, D. Hutchinson, A New Family of Self-scaling Variable Metric Algorithms for Unconstrained Optimisation, School of Computer Studies University of Leeds, (1989).

[12] J. C. Gilber, J. Nocedal, Global convergence properties of conjugate gradient, SIAM Journal of Optimization, **2,** (1992), 21–42.

[13] Elizabeth Dolan, D. Moré, J. Jorge, Benchmarking optimization software with performance profiles, Mathematical Programming, **91,** no. 2, (2002), 201–213.