A Quadratic Model based Conjugate Gradient Optimization Method

ISAM H. HALIL[1], ISSAM A.R. MOGHRABI[2], AHMED A. FAWZE[3], BASIM A. HASSAN[3], HISHAM M. KHUDHUR[3]

[1]Department of Mathematics, College of Science,
Kirkuk University, Kirkuk,
IRAQ

[2]Department of Computer Science, College of Arts and Sciences,
University Central Asia, Naryn,
KYRGYZ REPUBLIC

[3]Department of Mathematics, College of Computer Science and Mathematics,
University of Mosul, Mosul,
IRAQ

Abstract: - In this paper, we introduce a nonlinear scaled conjugate gradient method, operating on the premise of a descent and conjugacy relationship. The proposed algorithm employs a conjugacy parameter that is determined to ensure that the method generates conjugate directions. It also utilizes a parameter that scales the gradient to enhance the convergence behavior of the method. The derived method not only exhibits the crucial feature of global convergence but also maintains the generation of descent directions. The efficiency of the method is established through numerical tests conducted on a variety of high-dimensional nonlinear test functions. The obtained results attest to the improved behavior of the derived algorithm and support the presented theory.

Key-Words: - unconstrained optimization, conjugate gradient methods, line search methods, global convergence, quadratic modelling, non-linear programming.

Received: March 13, 2023. Revised: October 29, 2023. Accepted: November 13, 2023. Published: December 6, 2023.

1 Introduction

Our concern in this paper is problems of the form:

$$min\,(x), \text{ where } x \in R^n \quad , \qquad (1)$$

for which f is a differentiable convex function. For large dimension *n*, conjugate gradient (CG) methods are among the most sophisticated and straightforward approaches proposed to solve (1), due to storage considerations. Starting with a guess $x_0 \in R^n$, the CG method produces the sequence $\{x_k\}$ using the recurrence:

$$x_{k+1} = x_k + \alpha_k d_k, \text{ for } k = 0,1, \ldots \qquad (2)$$

where α_k is a positive scalar representing the step length along the search direction d_k, which is calculated using some line search method. The search vectors are derived using:

$$d_{k+1} = \begin{cases} -g_{k+1} & \text{if } \mathrm{k} = 0 \\ -g_{k+1} + \beta_k d_k & \text{if } \mathrm{k} > 0. \end{cases} \qquad (3)$$

The parameter β_k in (3) is referred to as the conjugacy parameter and g_{k+1} is the gradient vector evaluated at x_{k+1}. The primary factor in CG strategies is that they generate search directions d_k, which are downhill. The particular choice of the scalar parameter β_k defines different and new conjugate gradient algorithms. For convergence purposes and for aiding in ensuring that the generated search directions are downhill, the step length α_k in (2) is computed subject to satisfying the strict conditions, [1,2], given by:

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k \qquad (4)$$

$$d_k^T g(x_k + \alpha_k d_k) \ge \sigma d_k^T g_k \,, \qquad (5)$$

where $0 < \delta < \sigma$ to guarantee that d_k is a downward direction.

For a quadratic convex problem with positive definite *A* of the form:

$$f(x) = \tfrac{1}{2} x^T A x + b^T x + c \qquad (6)$$

and for accurate line searches used to solve:

$$\alpha_k = arg\min_{\alpha>0} f(x_k + \alpha_k d_k)\ , \tag{7}$$

the following conjugacy condition holds:

$$d_i^T A d_j = 0, \tag{8}$$

for any $i \neq j$. The CG techniques, typically used to solve systems of linear equations, are derived with satisfying conditions (8), as is the case with the original algorithm in, [6]. Almost all CG algorithms, such as those in, [3], [4], [5], [6], [7], [8], [9], satisfy (8) for quadratic objective functions and positive definite Hessian matrix.

In this paper, we provide new variants of the CG relations in (2) and (3) that are similar conceptually to those in, [10], [11], [12], that always satisfy:

$$g_{k+1}^T d_{k+1} < 0. \tag{9}$$

By the mean value theorem, for a nonlinear function f there is a value ξ such that:

$$d_{k+1}^T y_k = \alpha_k d_{k+1}^T \nabla^2 f(x_k + \xi\alpha_k d_k) d_k. \tag{10}$$

As a result of this, the following conjugacy criterion appears to be an appropriate substitute for (8):

$$d_{k+1}^T y_k = 0. \tag{11}$$

Using (3) and (11) leads to the relation (upon pre-multiplying both sides by α_k):

$$\beta_k v_k^T y_k - g_{k+1}^T y_k = 0, \tag{12}$$

for $v_k = x_{k+1} - x_k$ or, equivalently, the conjugacy parameter, [13]:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{v_k^T y_k}. \tag{13}$$

This paper has the following outline. We provide a novel scaled conjugate gradient approach in subsection 2. In Section 3, the proof that the suggested algorithm generates a search direction that complies with the descent requirements at each iteration is provided. The new CG-methods' global convergence property is proven in Section 4. The effectiveness of the suggested CG method is established with some numerical findings in Section 5. The final Section offers some conclusions and future research suggestions.

2 New Scaled CG Method

We focus attention in this paper on solving unconstrained minimization problems using the recurrence:

$$x_{k+1} = x_k + \alpha_k d_k\ , \tag{14}$$

where $\alpha_k > 0$ is acquired to satisfy the conditions for line search in (4) and (5), and d_k is some computed search direction using:

$$d_{k+1} = -\phi_k g_{k+1} + \beta_k v_k, \tag{15}$$

for some scalars ϕ_k and β_k, whose derivation specifies the fingerprints of the CG method.

The scalar parameters ϕ_k and β_k in Equation (15) are obtained in our approach for all $k \geq 0$ based on the descent condition:

$$g_{k+1}^T d_{k+1} = -\phi_k g_{k+1}^T g_{k+1} + \beta_k g_{k+1}^T v_k < 0. \tag{16}$$

From (15) and (16), we get:

$$g_{k+1}^T d_{k+1} = -\phi_k g_{k+1}^T g_{k+1} + \beta_k g_{k+1}^T v_k + \delta = 0,$$

which gives:

$$g_{k+1}^T d_{k+1} = -\phi_k g_{k+1}^T g_{k+1} + \beta_k g_{k+1}^T v_k + \delta. \tag{17}$$

Now, using the conjugacy condition (11) also gives:

$$y_k^T d_{k+1} = -\phi_k y_k^T g_{k+1} + \beta_k y_k^T v_k = 0. \tag{18}$$

Let us define:

$$\omega_k = -\|g_{k+1}\|^2 (y_k^T v_k). \tag{19}$$

Assuming that $\omega_k \neq 0$, if using (17) and (18) leads to:

$$\phi_k = -\frac{\delta(y_k^T v_k)}{\omega_k}$$

and

$$\beta_k = -\frac{\delta(y_k^T g_{k+1})}{\omega_k}. \tag{20}$$

Therefore, from (15) and (20) we obtain:

$$d_{k+1}^{New} = -\frac{\delta}{\|g_{k+1}\|^2} g_{k+1} + \frac{\delta(y_k^T g_{k+1})}{\|g_{k+1}\|^2 (y_k^T v_k)} v_k,$$

or, equivalently,

$$d_{k+1}^{New} = \frac{\delta}{\|g_{k+1}\|^2}\left[-g_{k+1} + \frac{y_k^T g_{k+1}}{(y_k^T v_k)} v_k\right] = \gamma_k \left[-g_{k+1} + \frac{y_k^T g_{k+1}}{(y_k^T v_k)} v_k\right]. \quad (21)$$

3 The Descent Feature of the Algorithm

The following is a presentation of the scaled conjugate gradient (CG) approach.

Step one. Choose $x_1 \in R^n$ and the line search parameters δ_1 and δ_2 such that $0 < \delta_1 < \delta_2 < 1$. Compute $f(x_1)$ and g_1. Take $d_1 = -g_1$ and $\alpha_1 = 1/\|g_1\|$.
Step two. Check if the looping can be continued. If $\|g_{k+1}\| \leq 10^{-6}$, we will cease.
Step three. Update the variables $x_{k+1} = x_k + \alpha_k d$ and compute $\alpha_{k+1} > 0$ that satisfies the line search criteria (4) and (5).
Step four. Compute the parameter β_k using equation (13).
Step Five. Compute $d_{k+1} = -\gamma_k(g_{k+1} + \beta_k d_k)$. If the restart strategy in, [5], namely $|g_{k+1}^T g_k| \geq 0.2\|g_{k+1}\|^2$, is met, then use $d_{k+1} = -\gamma_k g_{k+1}$. Else, define d_{k+1} using (21). Compute $\alpha_k = \frac{\alpha_{k-1}\|d_{k-1}\|}{\|d_k\|}$. Set $k = k+1$, and continue to step 2.

The following next theorem establishes that the search vectors generated by the derived scaled CG technique are guaranteed to meet the descent property, for Wolfe conditions-satisfied line searches.

Theorem 1
Assuming α_k meets conditions (3) and (4), hence, d_{k+1} given by(21) is a downhill direction.
Proof: Since $d_0 = -g_0$, we have $g_0^T d_0 = -\|g_0\|^2 \leq 0$. Multiplying (21) by g_{k+1}^T, we have

$$g_{k+1}^T d_{k+1} = \frac{\delta}{\|g_{k+1}\|^2}\left[-g_{k+1}^T g_{k+1} + \frac{y_k^T g_{k+1}}{(y_k^T v_k)} g_{k+1}^T v_k\right] = \frac{\delta}{\|g_{k+1}\|^2 (y_k^T v_k)^2}\begin{bmatrix} -g_{k+1}^T g_{k+1}(y_k^T v_k)^2 \\ +(y_k^T g_{k+1})(y_k^T v_k) g_{k+1}^T v_k \end{bmatrix}. \quad (22)$$

Applying ($u^T v \leq \frac{1}{2}(\|u\|^2 + \|v\|^2)$ to (22) with $u = (y_k^T v_k) g_{k+1}$ and $v = (g_{k+1}^T v_k) y_k$, we get:

$$g_k^T d_{k+1} \leq \frac{\delta}{|g_{k+1}|^2 (y_k^T v_k)^2}\left[-|g_{k+1}|^2 (y_k^T v_k)^2 + \frac{1}{2}\left[|g_{k+1}|^2 (y_k^T v_k)^2 + (g_{k+1}^T v_k)^2 (|y_k|^2)\right]\right] \quad (23)$$

or

$$g_{k+1}^T d_{k+1} \leq \frac{\delta}{\|g_{k+1}\|^2 (y_k^T v_k)^2}\left[-\frac{1}{2}\| g_{k+1}|^2 (y_k^T v_k)^2 + \frac{1}{2}(g_{k+1}^T v_k)^2 (| y_k \|^2)\right]. \quad (24)$$

The last term in equation (24) tends to zero closer to the minimum, so (24) can be expressed as:

$$g_{k+1}^T d_{k+1} \leq \frac{\delta}{\|g_{k+1}|^2 (y_k^T v_k)^2}\left[-\frac{1}{2}|g_{k+1}|^2 (y_k^T v_k)^2\right] \leq -\frac{\delta}{2}.$$

As a result, the direction d_{k+1} meets the criteria for a descent:

$$g_{k+1}^T d_{k+1} \leq 0.$$

4 Convergence Analysis

Using similar approaches to those in, [5], [14], [15], we assume that the objective function f satisfies the Lipschitz continuity criterion on the level set L and is also substantially convex. So,

$$L_0 = \{x \in \mathrm{R}^n : f(x) \leq f(x_0)\}.$$

Assumption A: There are values of $\mu > 0$ and L such that, [11]:

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \mu \| x - y|^2$$

and

$$(\nabla f(x) - \nabla f(y)) \leq L \| x - y \|^2$$

from L_0, for all x and y.

Lemma 1:
Let us assume that Assumption A is true. Then consider the general conjugate gradient approach in (2) and (3), where d_k is a downhill direction and α_k is computed to satisfy the line search conditions (4) and (5). If

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

Then it follows that

$$\lim_{k\to\infty} inf \| g_k |= 0. \quad (31)$$

Similar details can be found in [16], [17].

Theorem 2:
Assume that Assumption A holds, then $\{x_k\}$ is computed by the Algorithm (3.1) with $\gamma_k > 0$. Then

$$\lim_{k\to\infty} inf \|g_k\| = 0. \quad (32)$$

Proof:

We assume the conclusion is false because of the seeming conflict. Suppose that $g_k \neq 0$ for all k. By then (28), we get

$$y_k^T d_k = (g_{k+1} - g_k)^T d_k \geq \mu \alpha_k \|d_k\|^2. \quad (33)$$

In contrast, $\|y_k\| = \|g_{k+1} - g_k\| \leq L\|v_k\|^2$. Hence,

$$\left|g_{k+1}^T y_k\right| \leq \|g_{k+1}\| \|y_k\| \leq L\|g_{k+1}\| \|v_k\|. \quad (34)$$

Therefore, from (21), we have

$$\|d_{k+1}\| \leq \| g_{k+1} \left[\gamma_k + \gamma_k \frac{L\|g_{k+1}\|v_k\|}{\mu \alpha_k \|d_k\|^2}\right. \|\iota$$

$$\leq \|g_{k+1}\| \left[\gamma_k + \gamma_k \frac{L \mid g_{k+1} \|}{\mu \mid d_k \|}\right. \|v_k \quad \ldots\ldots (35)$$

$$\leq \|g_{k+1}\| \left[\gamma_k + \gamma_k \frac{L\varepsilon\alpha_k}{\mu}\right]$$

$$\leq \sqrt{\omega} \ \| \ g_{k+1} \mid$$

where $\sqrt{\omega} = \gamma_k + \gamma_k \frac{L\varepsilon\alpha_k}{\mu}$. This relation implies that

$$\sum_{k=1}^{2} \frac{1}{\|d_{k+1}\|^2} \geq \frac{1}{\omega E^2} \sum_{k=1} 1 = \infty. \quad (36)$$

Therefore, $\lim_{k\to\infty} inf \|g_k\| = 0$.

4 Numerical Outcomes

This section presents the outcomes of the computational experiments. A comparison has been made between the standard HS algorithm and the new scaled conjugate gradient method. Both methods were tested using a Fortran implementation. The test problems can be found in references, [1], [16], [17], [18], [19]. In the numerical experiments, the dimensions evaluated were n=1000 and 10000 for each test function. We selected fifteen large-scale, generalized, and unconstrained problems. To determine the termination of the process, we utilized the inequality $\|g_{k+1}\| \leq 10^{-6}$ as the criterion. Table 1 provides a view of the results derived from testing both methods using the line parameters σ=0.001 and $\delta \ = 0.9$ in equations (4) and (5). The columns in Table 1 indicate the following items

Problem**:** function name;
Dim**:** the dimension of the problem;
TNOI**:** total iterations number;
TIRS**:** total restarts number.

Table 1. Numerical comparison between the new algorithm and HS Algorithm Comparison of algorithms for $n = 1000$

Test Problems	**New**		β_k^{HS}	
	NOI	**IRS**	**NOI**	**IRS**
Freudenstein and Roth	137	125	838	810
Perturbed Quadratic	328	94	392	116
Diagonal 2	199	56	200	64
Diagonal 3	1745	1591	F*	F*
Extended Three Expo Terms	28	21	25	18
Extended Powell	75	20	75	20
Extended Maratos	64	29	85	44
Extended Cliff	11	9	13	10
Extended Wood	28	11	28	11
Quadratic QF2	395	122	403	121
EG2(CUTE)	113	48	F*	F*
Tridiagonal Perturbed Quad.	348	101	356	108
Drench (CUTE)	48	32	84	69
Diagonal 6	20	12	20	12
Sinquad (CUTE)	167	75	172	76
	1848	707	2693	1479

Table 2. Numerical comparison between the new algorithm and HS Algorithm for $n = 10000$

Test Problems	**New**		β_k^{HS}	
	NOI	**IRS**	**NOI**	**IRS**
Freudenstein and Roth	15	8	F*	F*
Perturbed Quadratic	1223	326	1243	332
Diagonal 2	684	207	686	203
Diagonal 3	F*	F*	F*	F*
Extended Three Expo Terms	65	57	215	207
Extended Powell	82	26	97	29
Extended Maratos	64	29	64	28
Extended Cliff	16	11	24	22
Extended Wood	27	9	29	9
Quadratic QF2	1235	361	1285	368
EG2(CUTE)	F*	F*	F*	F*
Tridiagonal Perturbed Quad.	1157	331	1297	376
Drench (CUTE)	51	40	102	89
Diagonal 6	18	9	22	11
Sinquad (CUTE)	437	29	1005	358
	5059	1435	6069	2032

Table 2 above presents a numerical comparison between the new algorithm and the HS Algorithm for n=10000.

5 Conclusion

In this study, a novel method of scaled nonlinear conjugate gradient is developed that, under certain assumptions, boosts its global convergence property for uniformly convex functions. Additionally, it successfully fulfills the essential descent condition, which is commonly observed in standard gradient algorithms. The utility of the suggested new scaled types was shown in the reported numerical experiments. The obtained results reveal, for at least the chosen set of test problems, that the developed algorithm reduces by an overall 61% NOI and 78.82% IRS against the standard Hestenes-Stiefel (HS) algorithm. The relative utility of the new approach $(n = 1000, 10000)$ is presented in Table 3.

Table 3. Relative utility of the new approach $(n = 1000, 10000)$

Tools	NOI	IRS
HS-Algorithm	100 %	100%
New-Algorithm	39%	21.18 %

The new method is promising and deserves further exploration of a wider spectrum of problems by extending the method to constrained optimization, exploring parallel and distributed implementations for scalability, [8], [20], [21], developing adaptive parameter tuning schemes, and analyzing sensitivity to initial conditions. It is also worth integrating the method with machine learning models, [22], assessing robustness to noisy objectives, comparing it with state-of-the-art methods, exploring real-world applications, and developing user-friendly interfaces for wider accessibility. These directions aim to enhance the algorithm's versatility, efficiency, and practical applicability across diverse optimization scenarios.

References:

[1] E. D. Dolan and J. J. Moré, “Benchmarking optimization software with performance profiles,” *Math. Program. Ser. B*, vol. 91, no. 2, pp. 201–213, 2002, doi: 10.1007/s101070100263.

[2] B. A. Hassan and H. A. Alashoor, “On image restoration problems using new conjugate gradient methods,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 3, p. 1438, Mar. 2023, doi: 10.11591/ijeecs.v29.i3.pp1438-1445.

[3] H. M. Khudhur and A. A. M. Fawze, "An improved conjugate gradient method for solving unconstrained optimization and image restoration problems," *Int. J. Mathematical Modeling Numerical Optimization*, vol. 13, no. 3, pp. 313–325, 2023, doi: 10.1504/IJMMNO.2023.132286.

[4] B. A. Hassan and H. Sadiq, “Efficient New Conjugate Gradient Methods for Removing Impulse Noise Images,” *Eur. J. Pure Appl. Math.*, vol. 15, no. 4, pp. 2011–2021, Oct. 2022, doi: 10.29020/nybg.ejpam.v15i4.4568.

[5] S. Aji, A. B. Abubakar, A. I. Kiri, and A. Ishaku, “A Spectral Conjugate Gradient-like Method for Convex Constrained Nonlinear Monotone Equations and Signal Recovery,” *Nonlinear Convex Anal. Optim.*, vol. 1, no. 1, pp. 1–23, 2022.

[6] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems, "*Journal of Research National Bureau Standard (1934)*, vol. 49, no. 6, pp. 409–436, Dec. 1952, doi: 10.6028/jres.049.044.

[7] N. Andrei, “An accelerated conjugate gradient algorithm with guaranteed descent and conjugacy conditions for unconstrained optimization,” *Optimization Methods and Software*, vol. 27, no. 4–5, 2012, doi: 10.1080/10556788.2010.501379.

[8] M. M. Abed, U. Öztürk, and H. M. Khudhur, “Spectral CG Algorithm for Solving Fuzzy Nonlinear Equations,” *Iraqi J. Computer Science and Math.*, vol. 3, no. 1, pp. 1–10, Jan. 2022, doi: 10.52866/ijcsm.2022.01.01.001.

[9] Y. A. Laylani, B. A. Hassan, and H. M. Khudhur, “Enhanced spectral conjugate gradient methods for unconstrained optimization,” *Computer Science*, vol. 18, no. 2, pp. 163–172, 2023.

[10] N. Andrei, “Numerical comparison of conjugate gradient algorithms for unconstrained optimization,” *Stud. Informatics Control*, vol. 16, pp. 333–352, 2007.

[11] H. N. Jabbar and B. A. Hassan, “Two-versions of descent conjugate gradient methods for large-scale unconstrained optimization,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 3, p. 1643, Jun. 2021,

doi: 10.11591/ijeecs.v22.i3.pp1643-1649.

[12] Y. Ismail Ibrahim and H. Mohammed Khudhur, "Modified three-term conjugate gradient algorithm and its applications in image restoration," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 3, pp. 1510–1517, Dec. 2022, doi: 10.11591/ijeecs.v28.i3.pp1510-1517.

[13] I. A. R. Moghrabi, "A New Formulation for Quasi-Newton Methods," WSEAS Transactions on Mathematics, vol. 18, pp. 79-84, 2019, DOI:

[14] E. G. Birgin and J. M. Martínez, "A Spectral Conjugate Gradient Method for Unconstrained Optimization," *Appl. Math. Optim.*, vol. 43, no. 2, pp. 117–128, 2001, doi: 10.1007/s00245-001-0003-0.

[15] Y. Laylani, B. A. Hassan, and H. M. Khudhur, "A New Class of Optimization Methods Based on Coefficient Conjugate Gradient," *Eur. J. Pure Appl. Math.*, vol. 15, no. 4, pp. 1908–1916, 2022.

[16] Y. H. Dai, "New Conjugacy Conditions and Related Nonlinear Conjugate Gradient Methods," *Appl. Math. Optim.*, vol. 43, no. 1, pp. 87–101, Jan. 2001, doi: 10.1007/s002450010019.

[17] Z.-F. Dai, "Two modified HS type conjugate gradient methods for unconstrained optimization problems," *Nonlinear Anal. Theory, Methods Appl.*, vol. 74, no. 3, pp. 927–936, 2011.

[18] N. Andrei, "An Unconstrained Optimization Test Functions Collection," *Adv. Model. Optim.*, vol. 10, no. 1, pp. 147–161, 2008.

[19] Z. Aminifard and S. Babaie-Kafaki, "Dai–Liao extensions of a descent hybrid nonlinear conjugate gradient method with application in signal processing," *Numer. Algorithms*, vol. 89, no. 3, pp. 1369–1387, Mar. 2022, doi: 10.1007/s11075-021-01157-y.

[20] A. Olawale, I. Osinuga, R. Raji, "A Globally Convergent Hybrid FR-PRP Conjugate Gradient Method for Unconstrained Optimization Problems", WSEAS Transactions on Mathematics, vol. 20, pp. 736-744, 2021.

[21] S. Singh, S. Gutta, A. Hadaegh, "Stock Prediction Using Machine Learning", WSEAS Transactions on Computer Research, vol. 9, pp. 152-158, 2021.

[22] M. Malik, I. Mohammed Sulaiman, A. Bala Abubakar, G. Ardaneswari, S. Sukono. A new family of hybrid three-term conjugate gradient method for unconstrained optimization with application to image restoration and portfolio selection. AIMS Mathematics, 2023, 8(1):1-28. doi: 10.3934/math.2023001.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

- Isam H. Halil contributed to the mathematical derivations.
- Issam Moghrabi drafted the document and contributed to the derivation.
- Ahmed Fawze carried out the numerical tests on the new method.
- Basim Hassan did the coding necessary for carrying out the tests.
- Hisham Khudhur summarized the results and did proofreading.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare.