EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 15, No. 4, 2022, 1683-1693
ISSN 1307-5543 – ejpam.com
Published by New York Business Global

The Convergent Properties of a New Parameter for Unconstrained Optimization

Ghada M. Al-Naemi

Department of Mathematics, Faculty of Computer Science and Mathematics, University of Mosul, Mosul, Nineveh, Iraq

Abstract. Because of its simplicity, low memory requirement, low computational cost, and global convergence properties, the Conjugate Gradient (CG) method is the most popular iterative mathematical technique for optimizing both linear and nonlinear systems. Some classical CG methods, however, have drawbacks such as poor global convergence and numerical performance in terms of iterations and function evaluations. To address these shortcomings, researchers proposed new CG parameter variants with efficient numerical results and good convergence properties. We present a new conjugate gradient formula β_k^{Gh} based on the memoryless self-scale DFP quasi-Newton (QN) method in this paper. The proposed new formula fulfills the sufficient descent property and the global convergent condition with any proposed line research. When the exact line search is used, the proposed formula is reduced to the classical HS formula. Finally, we conclude that our proposed method is effective.

2020 Mathematics Subject Classifications: 90C30, 65K05, 90C25, 90C26, 90C27

Key Words and Phrases: Conjugate gradient, self-scale DFP, strong Wolfe-Powell line search, sufficient descent property

1. Introduction

For solving the unconstrained minimization problem, the quasi-Newton methods are extremely useful and efficient to solve.

$$Min.z(x), x \in R^n \tag{1}$$

where $z : R^n \to R$ is twice continuously differentiable [5]. Broyden [2] introduced the QN family of variable metric formulas in 1970, which is the most efficient technique for minimizing a non-linear function $z(x)$. The following quasi-Newton equation has traditionally been used to update the iterate matrix:

$$\beta_{k+1} v_k = y_k \tag{2}$$

DOI: https://doi.org/10.29020/nybg.ejpam.v15i4.4538

Email address: drghadaalnaemi@uomosul.edu.iq (G. M. Al-Naemi)

We have

$$v_k = \gamma_k d_k = x_{k+1} - x_k, \quad \text{and} \quad y_k = g_{k+1} - g_k \tag{3}$$

where $\gamma_k > 0$ is determined by a suitable line search. Iterative methods are commonly used to solve (1), and the iterative formula is as follows:

$$x_{k+1} = x_k + \gamma_k d_k, \qquad k = 0, 1, 2, 3, \ldots, \tag{4}$$

If H_k is to be regarded as a close approximation to B_k^{-1}, it follows that:

$$H_{k+1} y_k = v_k \tag{5}$$

The direction d_k is obtained by solving the equation the updating matrix B_k is required to satisfy the equation (2) and the usual quasi-Newton equation (5).

$$d_k = -H_k g_k \tag{6}$$

The nonlinear conjugate gradient (CG) method is one of the most well-known methods for solving the unconstrained optimization problem (1), which is especially useful when the dimension n of $z(x)$ is large [7]. This is because the iteration is simple and requires little memory. The search direction is typically defined as:

$$d_k = \begin{cases} -g_k, & k = 0 \\ -g_k + \beta_k d_{k-1}, & k \geq 1 \end{cases} \tag{7}$$

$\beta_k \in R$, characterized the CG-method. If $f(x)$ is a strictly convex quadratic function with exact line search, the parameter β_k is typically chosen to reduce the linear CG-method in (4) and (7). [12][11][18][19][10][15][9] define the six pioneering forms of β_k.

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{y_{k-1}^T d_{k-1}}; \quad \beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}; \quad \beta_k^{PRP} = \frac{g_k^T y_{k-1}}{g_{k-1}^T g_{k-1}};$$

$$\beta_k^{CD} = \frac{g_k^T g_k}{y_{k-1}^T d_{k-1}}; \quad \beta_k^{LS} = \frac{g_k^T y_{k-1}}{-g_{k-1}^T d_{k-1}}; \quad \beta_k^{DY} = \frac{g_k^T g_k}{y_{k-1}^T d_{k-1}};$$

Many of the classic parameters β_k mentioned above have been modified by a group of researchers, and another group has derived or imposed new parameters β_k; however, not all of them can be included in a research, for example, see [13][21][16][1][22].

To determine the convergence conditions of above methods, it is usually necessary that the step size γ_k verify some properties, one of which is the strong Wolfe-Powel line search (sWP):

$$f(x_k + \gamma_k d_k) \leq f(x_k) + \rho \gamma_k d_k \tag{8}$$

$$|g(x_k + \gamma_k d_k)^T d_k| \leq \sigma |g_k^T d_k| \tag{9}$$

where $0 < \sigma < 0.5 < \rho < 1$ are some fixed parameters. The step-size γ_k plays an essential when investigating the sufficient descent condition

$$g_k^T d_k < -\omega||g_k||^2, \quad \omega > 0 \tag{10}$$

and global convergence properties

$$\lim_{k\to\infty} ||g_k||^2 = 0 \tag{11}$$

2. New Formulas for β's and its Algorithm

The DFP update was first proposed by Davidon, and popularized by Fletcher and Powell. The DFP formula can be expressed as follows:

$$H_k = \left[H_{k-1} + \frac{v_{k-1}v_{k-1}^T}{y_{k-1}^T v_{k-1}} - \frac{H_{k-1}y_{k-1}y_{k-1}^T H_{k-1}}{y_{k-1}^T H_{k-1}y_{k-1}}\right] \tag{12}$$

To scale the Hessian matrix H_k, we will use the self-scaling quasi-Newton method. Oren [17] introduced self-scaling variable metric algorithms, which are defined as

$$\eta_{k-1} = \frac{v_{k-1}^T y_{k-1}}{||v_{k-1}||^2} \tag{13}$$

which is a well-known and effective adaptive formula. Our proposed method's general strategy is to scale all DFP terms, i.e. update the matrix by self-scaling DFP of the form

$$H_k = \eta_{k-1}\left[H_{k-1} + \frac{v_{k-1}v_{k-1}^T}{y_{k-1}^T v_{k-1}} - \frac{H_{k-1}y_{k-1}y_{k-1}^T H_{k-1}}{y_{k-1}^T H_{k-1}y_{k-1}}\right] \tag{14}$$

The preceding self-scaling DFP method is transformed into the memoryless self-scaling DFP method when H_k is substituted for I (i.e. $H_k \equiv I$, where I is the identity matrix). As a result, the memoryless DFP formed by:

$$H_k = \frac{v_{k-1}^T y_{k-1}}{||v_{k-1}||^2}\left[I + \frac{v_{k-1}v_{k-1}^T}{y_{k-1}^T v_{k-1}} - \frac{y_{k-1}y_{k-1}^T}{y_{k-1}^T y_{k-1}}\right] \tag{15}$$

To derive the new formula multiply both sides of equation (15) by $(-g_k)$, we have (6) and in the CG (7), hence

$$-g_k + \beta_k d_{k-1} = \frac{v_{k-1}^T y_{k-1}}{||v_{k-1}||^2}\left[-g_k + \frac{v_{k-1}^T g_k}{y_{k-1}^T v_{k-1}}v_{k-1} - \frac{y_{k-1}^T g_k}{y_{k-1}^T y_{k-1}}y_{k-1}\right] \tag{16}$$

When we multiply both sides of (16) by y_{k-1}, we get

$$-g_k^T y_{k-1} + \beta_k d_{k-1}^T y_{k-1} = \frac{v_{k-1}^T y_{k-1}}{||v_{k-1}||^2}\left[-g_k^T y_{k-1} + \frac{v_{k-1}^T g_k}{y_{k-1}^T v_{k-1}}v_{k-1}^T y_{k-1} - \frac{y_{k-1}^T g_k}{y_{k-1}^T y_{k-1}}y_{k-1}^T y_{k-1}\right]$$

$$\beta_k d_{k-1}^T y_{k-1} = g_k^T y_{k-1} - \frac{v_{k-1}^T y_{k-1}}{||v_{k-1}||^2} v_{k-1}^T g_k \tag{17}$$

Use $v_{k-1} = \gamma_{k-1} d_{k-1}$ in (17), we have

$$\beta_k = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \left(\frac{\gamma_{k-1} d_{k-1}^T y_{k-1}}{\gamma_{k-1}^2 ||d_{k-1}||^2} \right) \left(\frac{\gamma_{k-1} d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right)$$

$$\beta_k^{Gh} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \frac{g_k^T d_{k-1}}{||v_{k-1}||^2} \tag{18}$$

It should be noted that if we used exact line search $\beta_k^{Gh} = \beta_k^{HS}$.

Now, we will go over the main steps of the algorithm that was used to create the new formula of β_k^{Gh}

2.1. Algorithm A

Given $x_0 \in R^n$, and $\varepsilon > 0$, set $k = 0$.

S1: Put $d_k = -g_k$, if $||g_k|| < \varepsilon$, stop, otherwise continue.

S2: Calculate γ_k by using (8) and (9).

S3: Calculate x_{k+1} by (4), and g_{k+1}, if $||g_{k+1}|| < \varepsilon$, then stop; Otherwise continue.

S4: Calculate β_k^{Gh} by (19) and d_{k+1} by (7).

S5: If the restarting criteria $|g_k^T g_{k-1}| \geq 0.2||g_k||^2$ is satisfy, proceed to S1, else put $k = k + 1$ go to S2.

2.2. Convergence Property

The following basic assumptions on the objective function are required to determine the global convergence property for algorithm (A).

2.3. Assumption B

i. $f(x)$ is constrained by the level set from below $\psi = \{x \in R^n, f(x) \leq f(x_0)\}$, where x_0 represents the starting point. Namely, there exists $\tau > 0$ which implies $||x_k|| \leq \tau \ \ \forall x \in \psi$ [3].

ii. $f(x)$ it is a smooth in a specific neighborhood $\mathcal{N}$ of ψ, and its gradient is Lipschitz continuous, i.e, there is a constant $\mathcal{L}$ greater than zero, such that

$$||\nabla \mathrm{f}\,(x) - \nabla \mathrm{f}\,(y)\,|| \leq \mathcal{L}||x - y||, \ \forall x, y \in \mathcal{N} \tag{19}$$

Using algorithm (A), there is now a positive constant (ω), resulting in $0 < ||g_k|| \leq \omega, \ \forall x \in \psi$ [14] To attain global convergence, algorithms (A) must be globally converged. To begin, we will look into the new proposed method's descent property.

Theorem 1. *Impose that Assumptions (B) hold. Suppose the method of the form (4) and (7) with β_k^{Gh} satisfy (18), and the step size γ_k satisfy sWP line search (8) and (9), then there exists a constant $\vartheta > 0$, s.t.*

$$g_k^T d_k \leq -\vartheta ||g_k||^2, \vartheta > 0, \ \forall k \geq 0 \tag{20}$$

Proof. To begin, the proof is trivial for $k = 0$, i.e.

$$d_0 = g_0 \Rightarrow g_0^T d_0 = -||g_0||^2$$

Multiplying both sides of (7) by g_k^T, we get

$$g_k^T d_k = -g_k^T g_k + \left[\frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \frac{g_k^T d_{k-1}}{||d_{k-1}||^2} \right] g_k^T d_{k-1}$$

$$= -g_k^T g_k + \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} g_k^T d_{k-1} - \frac{\left(g_k^T d_{k-1}\right)^2}{||d_{k-1}||^2}$$

We know that $g_k^T d_{k-1} \leq d_{k-1}^T y_{k-1}$, and $g_k^T d_{k-1} \leq ||g_k||.||d_{k-1}||$

$$g_k^T d_k \leq -||g_k||^2 + \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} d_{k-1}^T y_{k-1} - \frac{(||g_k||.||d_{k-1}||)^2}{||d_{k-1}||^2}$$

$$= -||g_k||^2 + g_k^T y_{k-1} - ||g_k||^2$$

$$g_k^T y_{k-1} = ||g_k||^2 - g_k^T g_{k-1}$$

Using the restarting criteria, i.e. $g_k^T g_{k-1} \leq -0.2||g_k||^2$, yield

$$g_k^T d_k \leq -2||g_k||^2 + ||g_k|| + 0.2||g_k||^2$$

$$g_k^T d_k \leq -0.8||g_k||^2$$

As a result, (20) holds true for all k. After demonstrating that Algorithm (A) satisfies the descent property, we must demonstrate Algorithm (A) global convergence under assumption (B). We require the following lemmas, which are frequently used to demonstrate global convergence and Zoutendijk [23] provides them.

Lemma 1. *Let the Assumption (B) be correct. Assume any iteration method (4) and (7), and γ_k obtained by the sWP (8) and (9). If*

$$\sum_{k \geq 1} \frac{1}{||d_k||^2} = \infty \tag{21}$$

Then

$$\liminf_{k \to \infty} ||g_k|| = 0 \tag{22}$$

Theorem 2. *Consider that Assumption (B) established. Suppose that the algorithm (A), and* γ_k *is obtained through the sWP and* d_k *is the descent direction. Then* $\liminf_{k\to\infty} ||g_k|| = 0$.

Proof. Because the descending property holds, we now have $d_k \neq 0$. Therefore, lemma (2) is sufficient to show that $||d_k||$ is constrained above. From (2) and (8),

$$||d_k|| = \left|\left| -g_k + \left[\frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \frac{d_{k-1}^T g_k}{||d_{k-1}||^2} \right] d_{k-1} \right|\right|$$

Since $\left| d_{k-1}^T y_{k-1} \right| \geq m||d_{k-1}||.||y_{k-1}||$, where $m > 0$ [20], so

$$||d_k|| \leq ||g_k|| + \frac{||g_k||.||y_{k-1}||}{m||d_{k-1}||.||y_{k-1}||} - ||g_k||$$

$$\leq \left(\frac{1}{m||d_{k-1}||} \right) ||g_k|| \leq \left(\frac{1}{m.v} \right) \omega = \mu$$

$$\Rightarrow \sum_{k\geq 1} \frac{1}{||d_k||^2} \geq \frac{1}{\mu^2} \sum_{k\geq 1} 1 = \infty$$

As a result, (24) applies to all k.

3. Numerical Results

The main task in this section is to report the Algorithm (A performance)'s on a set of test functions. All codes are written with double precision in FORTRAN. We chose twenty unconstrained large-scale optimization test problems. We considered two experiments with different numbers of variables (n=100 and 1000) for each test function. The test problems are from the CUTE [6] library, as well as other large-scale optimization test problems from [4]. To assess the reliability of our algorithms, we used the same test functions to compare them to the well-known routines HS, DY, PRP and LS. All of these algorithms use sWP (8) and (9) line searches with $\delta = 0.001$ and $\sigma = 0.5$, respectively. When the following stopping criterion is met $||g_{k+1}|| \leq 10^{-5}$, all of these methods terminate. Dolan and Mor'e created performance profiling software. [8] was also used to analyze the execution Figures 1 and 2.

Table 1: Comparison between $\beta^{HS}, \beta^{PRP}, \beta^{LS}$, and β^{Gh} with $n = 100$

Test functions	β_k^{Gh}		β_k^{HS}		β_k^{PRP}		β_k^{LS}	
	NOI	NOF	NOI	NOF	NOI	NOF	NOI	NOF
Wood	27	64	33	73	29	67	30	69
Wolfe	54	101	55	103	57	114	55	115
Rosen	38	101	34	94	35	96	35	96
NON	28	67	30	78	32	81	31	80
Shallow	10	25	10	25	10	25	10	25
ENGVAL 1	21	44	22	46	24	46	22	46
Diagonal 2	54	207	60	213	58	214	60	221
Ex. BD1	20	40	22	46	19	39	23	48
Ex. Wood	27	61	29	65	32	67	29	65
Powell	31	92	37	104	39	112	41	118
Dixnaanc	5	15	5	15	5	15	5	15
DENSCHNF	21	44	24	51	28	60	26	56
Dixmaanb	29	66	33	73	26	60	27	67
Ex. Rosen	29	66	30	69	31	72	30	70
Cubic	14	39	17	46	21	51	16	44
Ex. Beal U63	10	27	10	27	10	27	10	27
Ex. TET	39	81	45	98	51	117	53	117
Gen.Tridiagonal-2	116	251	120	255	123	261	125	268
Diagonal 6	3	9	3	9	3	9	3	9
SUM	16	77	18	82	21	110	20	103
Total	592	1477	637	1542	653	1643	651	1600

Table 2: The percentage between PRP, LS, HS, and Gh for $n = 100$

Measurement	PRP method	LS method	HS method	Gh method
NOI	100%	99.69%	97.55%	90.66%
NOF	100%	97.38%	93.85%	89.9%

Table 3: Comparison between $\beta^{HS}, \beta^{PRP}, \beta^{LS}$ and β^{Gh} with $n = 1000$

Test functions	β_k^{Gh}		β_k^{HS}		β_k^{PRP}		β_k^{LS}	
	NOI	NOF	NOI	NOF	NOI	NOF	NOI	NOF
Wood	32	74	36	84	35	81	37	87
Wolfe	58	108	59	119	58	116	59	118
Rosen	38	101	34	94	35	96	35	96
NON	30	76	33	84	39	94	32	83
Shallow	10	25	10	25	10	25	10	25
ENGVAL 1	20	44	22	46	23	46	22	46
Diagonal 2	54	207	62	225	58	214	59	219
Ex. BD1	21	44	23	48	24	50	23	48
Ex. Wood	27	61	30	67	30	67	29	65
Powell	39	110	41	109	56	129	52	121
Dixnaanc	5	15	5	15	5	15	5	15
DENSCHNF	27	62	29	67	31	73	30	69
Dixmaanb	35	75	39	89	32	67	34	72
Ex. Rosen	31	69	34	75	37	86	39	90
Cubic	16	41	18	50	21	59	20	55
Ex. Beal U63	12	29	12	29	12	29	12	29
Ex. TET	48	95	54	109	59	118	58	116
Gen.Tridiagonal-2	121	232	125	234	127	239	130	241
Diagonal 6	3	9	3	9	3	9	3	9
SUM	19	83	20	87	23	115	22	107
Total	647	1571	689	1663	722	1759	711	1691

Table 2 shows that the proposed β_k^{Gh} formula outperforms the classic $\beta_k^{PRP}, \beta_k^{LS}$, and β_k^{HS} formulas in terms of percentage performance. We discovered that the proposed Gh algorithm saves $(NOI, 9.34\%), (NOF, 10.1\%)$, the LS algorithm saves $(NOI, 0.31\%), (NOF, 2.62\%)$ and the HS algorithm saves $(NOI, 2.45\%), (NOF, 6.15\%)$, for $n = 100$.

Table 4 shows that the proposed β_k^{Gh} formula outperforms the classic $\beta_k^{PRP}, \beta_k^{LS}$, and β_k^{HS} formulas in terms of percentage performance. We discovered that the proposed Gh algorithm saves $(NOI, 9.0\%), (NOF, 7.1\%)$, the LS algorithm saves $(NOI, 1.52\%), (NOF, 3.86\%)$ and the HS algorithm saves $(NOI, 4.57\%), (NOF, 5.46\%)$, for $n = 1000$.

Figures (1) (a) and (b) depict the efficiency of the proposed method in terms of NOI for $n = 100$ and 1000, respectively. Figures (2: (c) and (d)) demonstrate the effectiveness of the suggested method in terms of NOF for $n = 100$ and 1000, respectively.

Table 4: The percentage between PRP, LS, HS, and Gh for $n = 1000$

Measurement	PRP method	LS method	HS method	Gh method
NOI	100%	98.48%	95.43%	91.0%
NOF	100%	96.14%	94.54%	92.9%

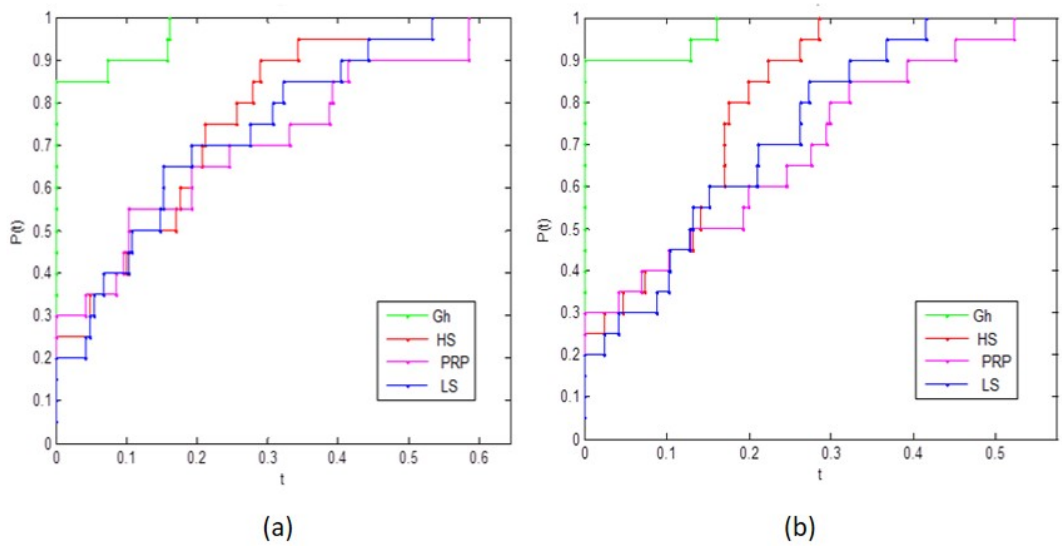

Figure 1: The comparison based (NOI) between $\beta^{HS}, \beta^{PRP}, \beta^{LS}$, and new β^{Gh} (a) n=100 and (b) n=1000

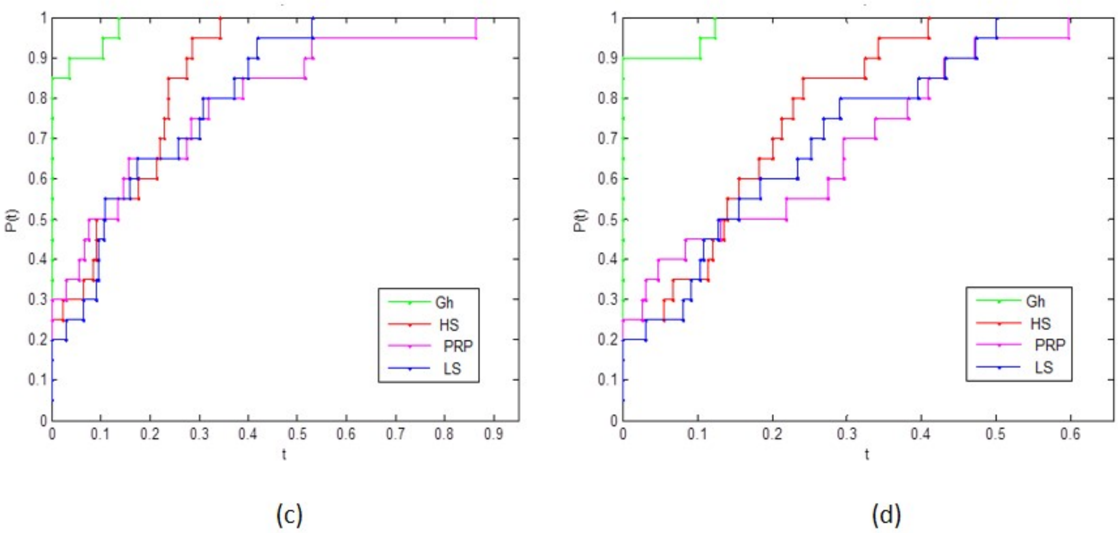

Figure 2: The comparison based (NOI) between $\beta^{HS}, \beta^{PRP}, \beta^{LS}$, and new β^{Gh} (c) n=100 and (d) n=1000

4. Conclusion

We present a new parameter β_k^{Gh} based on the memoryless self-scale DFP QN method in this article. Any line search will suffice to ensure adequate descent property. We also demonstrated that the Zoutendijk condition holds and that the method is globally convergent by using some step-length technique. The numerical results demonstrated the proposed algorithm's efficiency when compared to some standard formulas.

Acknowledgements

The author wishes to thank the College of Comp. Sci. and Math. at the University of Mosul for their encouragement.

References

[1] Abdelrahman A, Osman O O Yousif, and Khalid I Osman. A robust modification of hestenes–stiefel conjugate gradient method with strong wolfe line search. *Journal of Mathematics and Statistics*, 16(1):54–61, 2020.

[2] M Al-Baali. Descent property and global convergence of the fletcher—reeves method with inexact line search. *IMA Journal of Numerical Analysis*, 5(1):121–124, 1985.

[3] G M Al-Naemi. A new modified hs algorithm with strong powell-wolfe line search for unconstrained optimization. *Eastern-European Journal of Enterprise Technologies*, 2(4):116, 2022.

[4] N Andrei. An unconstrained optimization test functions collection. *Advance Modeling and Optimization*, 10(1):147–161, 2008.

[5] N Andrei. *Nonlinear conjugate gradient methods for unconstrained optimization*. Springer International Publishing, 2020.

[6] I Bongartz, A R Conn, N Gould, and Ph L Toint. Cute: Constrained and unconstrained testing environment. *ACM Transactions on Mathematical Software (TOMS)*, 21(1):123–160, 1995.

[7] Y Chen. Global convergence of a new conjugate gradient method with wolfe type line search. *Journal of Information and Computing Science*, 7(1):67–71, 2012.

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

[9] 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.

[10] R Fletcher. *Practical methods of optimization*. John Wiley & Sons, 1987.

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

[12] 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.

[13] Ishak M I, Marjugi S M, and June L. A new modified conjugate gradient method under the strong wolfe line search for solving unconstrained optimization problems. *Mathematical Modeling and Computing*, 9(1):111–118, 2022.

[14] F N Jardow and G M Al-Naemi. A new hybrid conjugate gradient algorithm as a convex combination of mmwu and rmil nonlinear problems. *Journal of Interdisciplinary Mathematics*, 24(3):637–655, 2021.

[15] Y Liu and C Storey. Efficient generalized conjugate gradient algorithms, part 1: theory. *Journal of optimization theory and applications*, 69(1):129–137, 1991.

[16] Oke M O, Oyelade T A, and Raji R A. On a modified conjugate gradient method for solving nonlinear unconstrained optimization problems. *European Journal of Basic and Applied Sciences*, 6(2):17–23, 2019.

[17] S S Oren and D G Luenberger. Self-scaling variable metric (ssvm) algorithms: Part i: Criteria and sufficient conditions for scaling a class of algorithms. *Management Science*, 20(5):845–862, 1974.

[18] E Polak and G Ribiere. Note sur la convergence de méthodes de directions conjuguées. *Revue française d'informatique et de recherche opérationnelle. Série rouge*, 3(16):35–43, 1969.

[19] B T Polyak. The conjugate gradient method in extremal problems. *USSR Computational Mathematics and Mathematical Physics*, 9(4):94–112, 1969.

[20] H Yabe and M Takano. Global convergence properties of new nonlinear conjugate gradient methods for unconstrained optimization. *Computational Optimization and Application*, 28:203–225, 2004.

[21] Salleh Zabidin, Alhamzi Ghaliah, Masmali Ibitsam, and Alhawarat Ahmad. A modified liu and storey conjugate gradient method for large scale unconstrained optimization problems. *Algorithms*, 14(8):227, 2021.

[22] Yutao Zheng. A new family of dai-liao conjugate gradient methods with modified secant equation for unconstrained optimization. *RAIRO-Operations Research*, 55(6):3281–3291, 2021.

[23] G Zoutendijk. Nonlinear programming, computational methods. *Integer and nonlinear programming*, pages 37–86, 1970.