UNIVERSITI TEKNOLOGI MARA
MATHEMATICS LETTERS
https://sites.google.com/tmsk.uitm.edu.my/mathematicsletters

Performance Analysis of Selected Spectral Conjugate Gradient Methods for Large-Scale Unconstrained Optimization Problems with Strong-Wolfe Line Search

[1]Wan Khadijah Wan Sulaiman, [2,*]Nur Hamizah Abdul Ghani, [3]Adlina Suraya Bahruddin & [4]Muhanizah Abdul Hamid

Centre of Mathematical Sciences, Faculty of Computer Science and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.
[1]wankhadijah@uitm.edu.my
[*2]hamizahghani@uitm.edu.my
[3]adlinasuraya12@gmail.com
[4]muhanizah@uitm.edu.my
*Corresponding author

Received: 18/3/2026 Revised: 31/3/2026 Accepted: 5/4/2026 Published: 30/4/2026

ABSTRACT

The conjugate gradient (CG) method is crucial for solving unconstrained optimization especially large-scale unconstrained optimization problems. Although many studies have proposed new spectral CG method and demonstrated their superior performance, identifying the most efficient method among them remains challenging. The spectral CG method improves the convergence behaviour of the classical CG method through the incorporation of the spectral parameters, which ensures better descent properties while remaining computationally efficient for large-scale optimization problems. Therefore, this study is done by investigating the performance of four selected spectral CG methods, which are spectral Hestenes-Steifel (SHS), spectral Conjugate Descent-Fletcher (SCD), spectral Rivaie, Mustafa, Ismail, and Leong (SRMIL), and spectral Sri-Mustafa-2 (SSM-2), under inexact strong-Wolfe line search. The main objective of this study is to analyse and compare the numerical performance and evaluate efficiency of these methods. Numerical experiments are conducted on 20 standard unconstrained optimization test functions with five different problem dimensions and various initial points, implemented using MATLAB programming. The performance is analysed based on the number of iterations and Central Processing Unit (CPU) time. In addition, a mathematical proof is provided to verify that all four spectral CG methods satisfy the sufficient descent condition. The results indicate that the SHS method gives the best performance compared to the other methods, which shows that it is the fastest method and able to solve all test problems.

Keywords: Performance Profile, Spectral Conjugate Gradient Method, Strong-Wolfe Line Search, Sufficient Descent Condition, Unconstrained Optimization Problem.

INTRODUCTION

Optimization problems are formulated with the aim of finding the best solution from a set of feasible solutions. The goal of optimization problems is either to minimize or maximize the objective function. There are a few common types of optimization problems, including linear programming, nonlinear programming, constrained optimization, unconstrained optimization, discrete optimization, continuous optimization, and multiobjective optimization (Nocedal & Wright, 1999).

An unconstrained optimization is a specific and important model of optimization problems. It focuses on finding the best solution that minimizes or maximizes an objective function, without any restrictions or constraints on the values of the decision variables. The general form of the unconstrained minimization problem is described as follows,

$$\min_{x \epsilon \mathbb{R}^n} f(x) \tag{1}$$

where $f(x)$ is an objective function, and x represents a vector of decision variables (Andrei, 2020). According to Andrei (2020), there are several approaches for solving unconstrained optimization problems, such as the Steepest Descent Method, Conjugate Gradient (CG) Method, Newton's Method, and Quasi-Newton Method. The unconstrained minimization problem (1) can be solved by an iteration method,

$$x_{k+1} = x_k + \alpha_k d_k \ and \ k = 0,1,2,\ldots \tag{2}$$

where x_{k+1} is the next iteration point, x_k is the current point, α_k is the step size, and d_k is the search direction.

The search direction, d_k of the CG methods is termed as:

$$d_k = \begin{cases} -g_k & if\ k = 0 \\ -g_k + \beta_k d_{k-1} & if\ k \geq 1 \end{cases} \tag{3}$$

where $g_k = \nabla f(x)$ and β_k is denoted for the CG coefficients.

The value of the step size, α_k can be generated either with exact line search or inexact line search. The exact line search is used to find the exact value of the α_k, and defined by,

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha_k d_k) \tag{4}$$

While inexact line search is used to find the approximate value of the α_k. There are a few common types of inexact line search, such as Wolfe (Wolfe, 1969), Armijo (Armijo, 1966), Goldstein (Goldstein, 1965), and Backtracking (Dennis & Schnabel, 1996).

In this study, inexact strong-Wolfe line search is used to generate the value of the step size, α_k since the exact line search reduced the spectral CG method to the classical CG method. The inexact strong-Wolfe line search (Wolfe, 1971) is defined by,

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho d_k^T g_k \tag{5}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma d_k^T g_k \tag{6}$$

where $0 < \rho < \sigma < 1$. In order to satisfy the requirement for an optimization method, the descent property (7) must exist in the search direction, d_k.

$$g_k^T d_k \tag{7}$$

Table 1 presents four CG formulas that are applied in this study. The Hestenes-Stiefel (HS) and Conjugate Descent-Fletcher (CD) CG method is the most classical CG coefficients, β_k. The CG coefficient proposed by Rivaie, Mustafa, Ismail, and Leong, which is denoted as RMIL and the coefficients proposed by Japri, Basri, and Mamat that denoted as SM-2, are the motivation from the HS coefficient. The formula in the denominator that distinguishes the three methods.

Table 1

Conjugate Gradient Coefficients

Author	Year	Formula
Hesteness and Steifel (Hestenes & Stiefel, 1952)	1952	$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})}$
Fletcher (Fletcher, 2000)	1987	$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}$
Rivaie, Mustafa, Ismail, and Leong (Rivaie et al., 2012)	2012	$\beta_k^{RMIL} = \frac{g_k^T(g_k - g_{k-1})}{\Vert d_{k-1}\Vert^2}$
Japri, Basri, and Mamat (Japri et al., 2021)	2021	$\beta_k^{SM-2} = \frac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T\left(g_{k-1} + \frac{\Vert g_k\Vert}{\Vert g_{k-1}\Vert}\right)}$

In this study, the focus is on the spectral CG method. It is one of the CG method types. This method is introduced by Barzilai and Borwein (1988) in their research. After that, Raydan (1997) developed it for large-scale unconstrained optimization problems in his research. The difference of the spectral CG method is in the search direction, d_k. The search direction, d_k of the spectral CG method is taken based on (Birgin & Martínez, 2001) and designed as:

$$d_k = \begin{cases} -g_k & if\ k = 0 \\ -\theta_k g_k + \beta_k d_{k-1} & if\ k \geq 1 \end{cases} \tag{8}$$

where θ_k is a spectral parameter cited from (Ghani et al., 2024) defined as:

$$\theta_k = 1 + \beta_k \frac{g_k^T d_{k-1}}{\Vert g_k\Vert^2} \tag{9}$$

The HS, CD, RMIL, and SM-2 are measured using the spectral CG method. Therefore, all four methods are known as SHS, SCD, SRMIL, and SSM-2.

METHODOLOGY

Spectral Conjugate Gradient Method Algorithm

The algorithm below shows the general algorithm used for computing the spectral CG method.

Algorithm 1. Spectral Conjugate Gradient Method Algorithm

Step 1 : Initialization.
Set iteration, $k = 0$.
Choose an initial point, $x_0 \in \mathbb{R}^n$.
Compute:
The function value, $f_0 = f(x_0)$
The gradient, $g_0 = \nabla f(x_0)$
The initial search direction, $d_0 = -g_0$

If the norm gradient has already reached zero, $\|g_0\| = 0$, this implies that a local minimum has been achieved, thus, the algorithm terminates at this step.

Step 2 : Compute the CG coefficients, β_k.
Use formula in Table 1 to compute the β_k.

Step 3 : Compute the search direction of the spectral CG method, d_k based on the formula in (8), and the spectral parameter, θ_k as (9).

Step 4 : Compute the step size, α_k based on inexact strong-Wolfe line search defined as (5) and (6).

Step 5 : Update the iteration point as shown in (2).

Step 6 : Convergent test and stopping criteria.
If $f(x_{k+1}) < f(x_k)$ and $\|g_k\| \leq \varepsilon$, where $\varepsilon = 10^{-6}$ or when $k = 10{,}000$, then terminate the algorithm. It is considered that the solution is good enough.
Otherwise, go back to Step 1 with $k = k + 1$.

CONVERGENCE ANALYSIS

For the theoretical phase, usually focuses on a few properties, including the sufficient descent conditions, global convergence, convergence rate, and angle conditions. These properties ensure that the developed algorithm produces valid search directions and converges to a stationary point under certain standard assumptions (Nocedal & Wright, 1999). These theoretical properties are important in assessing the robustness of the algorithm independently of numerical experiments. In order to generate the global convergences, the algorithm needs to satisfy the sufficient descent condition first. Therefore, this study focuses on proving the sufficient descent condition for the selected spectral CG methods. The sufficient descent condition is an important property that needs to be proved in optimization problems. It is because to ensure each iteration moves closer to the minimum of the function. The sufficient descent condition is defined as follows:

$$g_k^T d_k \leq -C\|g_k\|^2 \; for\; k \geq 0, where\; C > 0 \tag{10}$$

The following Theorem 1 is a general theorem used to prove the sufficient descent condition in the spectral CG method.

Theorem 1

Let x_k, d_k, be generated by (2), (8), and β_k as in Table 1, then the condition (10) holds for all $k \geq 0$.

Proof

If $k = 0$, then $g_0^T d_0 = -C\|g_0\|^2$. Therefore, condition (10) holds true. The next step shows for $k \geq 1$, condition (10) must also hold true.

Multiply formula (8) by g_k^T, then

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

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

$$g_k^T d_k = -\|g_k\|^2 - \beta_k \frac{g_k^T d_{k-1}}{\|g_k\|^2} \|g_k\|^2 + \beta_k g_k^T d_{k-1} \tag{11}$$

Then, apply the stong-Wolfe line search to (11).

$$g_k^T d_k \leq -\|g_k\|^2 - \beta_k \frac{\sigma g_k^T d_k}{\|g_k\|^2} \|g_k\|^2 + \beta_k \sigma g_k^T d_k$$

$$g_k^T d_k \leq -\|g_k\|^2 - \beta_k \sigma g_k^T d_k + \beta_k \sigma g_k^T d_k$$

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

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

Based on the above proof, the sufficient descent condition is fulfilled. Thus, the condition (10) holds for all $k \geq 0$, and the proof is completed for the sufficient descent condition.

RESULT AND DISCUSSION

The numerical performance of four spectral CG methods, SHS, SCD, SRMIL, and SSM-2 are computed by solving 20 standard unconstrained optimization test functions by Andrei (2008) and Jamil and Yang (2013). All performances are computed by using MATLAB R2024b programming and compiled with a personal laptop, Intel Core i5 processor, 8 GB RAM, and 64-bit Windows 11 operating system. For every problem and variable, five distinct initial points are randomly selected in a range -20 to 20. The variable started from small-scale towards large-scale, $2 \leq n \leq 5{,}000$. The stopping criteria are set to $\|g_k\| \leq \varepsilon$, and $\varepsilon = 10^{-6}$ for each algorithm. Finally, the number of iterations generated is set to $k = 10{,}000$, otherwise, the CG method is denoted "Fail". The performance of each algorithm is generated based on the number of iterations and CPU time.

Table 2

List of Test Function, Dimension, and Initial Point

No.	Test Function	Dimension	Initial Point
1	Extended Rosenbrock	2	(5,5)
2		4	(12, ...,12)
3		20	(-3, ...,-3)
4		600	(-1.2, 1, ...,-1.2, 1)
5		3,000	(3, ...,3)
6	Extended White & Holst	4	(2,2)
7		6	(14, ...,14)
8		40	(-6, ...,-6)

9		800	(-1.2, 1, ...,-1.2, 1)
10		4,000	(5, ...,5)
11	Extended Himmelblau	2	(3,3)
12		8	(15, ...,15)
13		60	(-9, ...,-9)
14		1,000	(1, ..., 1)
15		5,000	(10, ...,10)
16	Raydan 1	2	(1,1)
17		3	(4,4,4)
18		9	(16, ...,16)
19		80	(-12, ...,-12)
20		300	(5, ...,5)
21	Hager	2	(5,5)
22		5	(1, ...,1)
23		10	(18, ...,18)
24		100	(-15, ...,-15)
25		200	(2, ...,2)
26	Diagonal 4	2	(6,6)
27		6	(20, ...,20)
28		200	(-18, ...,-18)
29		800	(1, ...,1)
30		5,000	(2, ...,2)
31	Extended Maratos	4	(-1, ...,-1)
32		12	(5, ...,5)
33		100	(12, ...,12)
34		1000	(3, ...,3)
35		4000	(1, ...,1)
36	FLETCHCR	3	(10,10,10)
37		9	(14, ...,14)
38		400	(-6, ...,-6)
39		2,000	(0.1, ...,0.1)
40		4,000	(-2, ...,-2)
41	Generalized Quartic	2	(5,5)
42		8	(16, ...,16)
43		500	(9, ...,9)
44		600	(1, ...,1)
45		5,000	(3, ...,3)
46	Extended Tridiagonal 1	2	(1,1)
47		6	(18, ...,18)
48		20	(-12, ...,-12)
49		800	(2, ...,2)
50		3,000	(-5, ...,-5)
51	Extended Penalty	2	(2,2)
52		10	(20, ...,20)
53		30	(-15, ...,-15)
54		1,000	(5, ...,5)
55		4,000	(0, ...,0)
56	Extended Freudenstein & Roth	4	(3, ...,3)
57		10	(15, ...,15)
58		40	(-18, ...,-18)
59		2,000	(-10, ...,-10)
60		5,000	(12, ...,12)
61	Quadratic QF1	2	(4,4)
62		4	(12, ...,12)
63		60	(-3, ...,-3)
64		600	(1, ...,1)
65		3,000	(10, ...,10)
66	POWER	3	(5,5,5)

67		6	(14, ...,14)
68		20	(2, ...,2)
69		80	(-6, ...,-6)
70		200	(1.2, ...,1.2)
71	Camel Three Hump	2	(-1, 2)
72		2	(-2, 1)
73		2	(5, 5)
74		2	(10,10)
75		2	(-15, -15)
76	Camel Six Hump	2	(-1, 2)
77		2	(-5, 10)
78		2	(-8, -8)
79		2	(12,12)
80		2	(2, 2)
81	Extended Beale	4	(6, ...,6)
82		10	(1.2, ...,1.2)
83		100	(9, ...,9)
84		1,000	(1,0.8, ...,1,0.8)
85		5,000	(1, ...,1)
86	NONSCOMP	2	(8,8)
87		10	(1.1, ...,1.1)
88		200	(-5, ...,-5)
89		600	(-3, ...,-3)
90		3,000	(-1.1, ...,-1.1)
91	Extended DENSCHNB	4	(10, ...,10)
92		6	(18, ...,18)
93		300	(-15, ...,-15)
94		800	(1, ...,1)
95		4,000	(2, ...,2)
96	Extended Quadratic Penalty QP1	2	(-1.2,1)
97		10	(20, ...,20)
98		50	(5, ...,5)
99		400	(-18, ...,-18)
100		1,000	(-10, ...,-10)

Table 3

Performance of Four Spectral CG Methods Based on Number of Iterations and CPU Time.

CG Method	Total Iteration Number	Total CPU Time	Successful Rate	Failure Rate
SHS	2,898	2.7035	100%	0%
SCD	17,667	6.0402	86%	14%
SRMIL	13,516	5.4226	91%	9%
SSM-2	38,882	28.1675	100%	0%

Table 2 above shows the list of 20 unconstrained optimization test function used in this study with the dimension and initial point of each test function. Furthermore, the Table 3 shows the summary of the numerical performance in percentage form for each method in terms of the number of iterations and CPU time. These results are generated using the same spectral parameters, unconstrained optimization test function, dimensions, and initial points. The numerical results are analyzed by using a performance profile, which are introduced by Dolan and Moré (2002).

Figure 1

Performance Profile Based on Number of Iterations

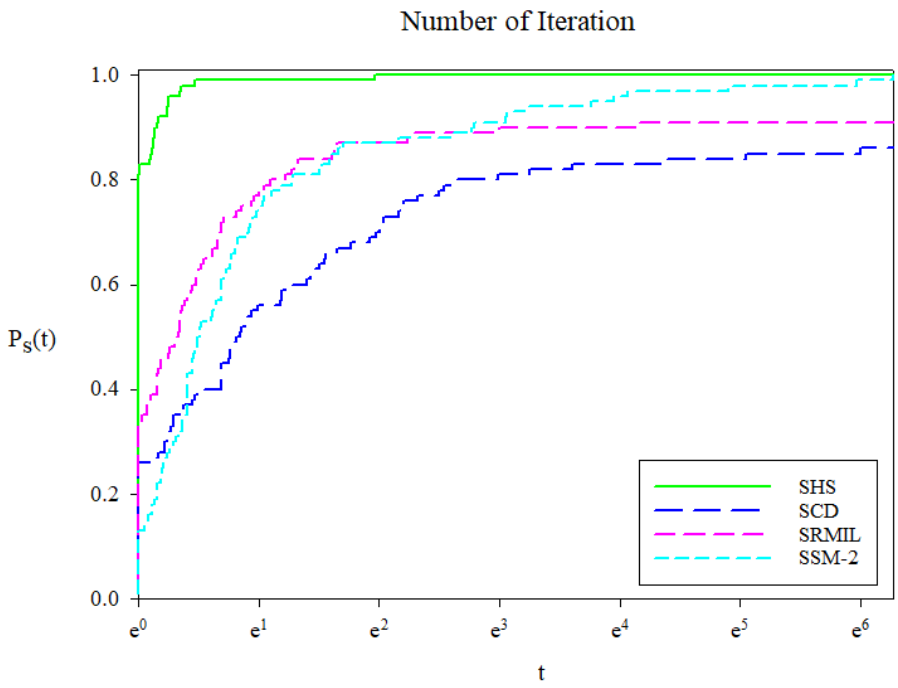

Figure 2

Performanace Profile Based on CPU Time

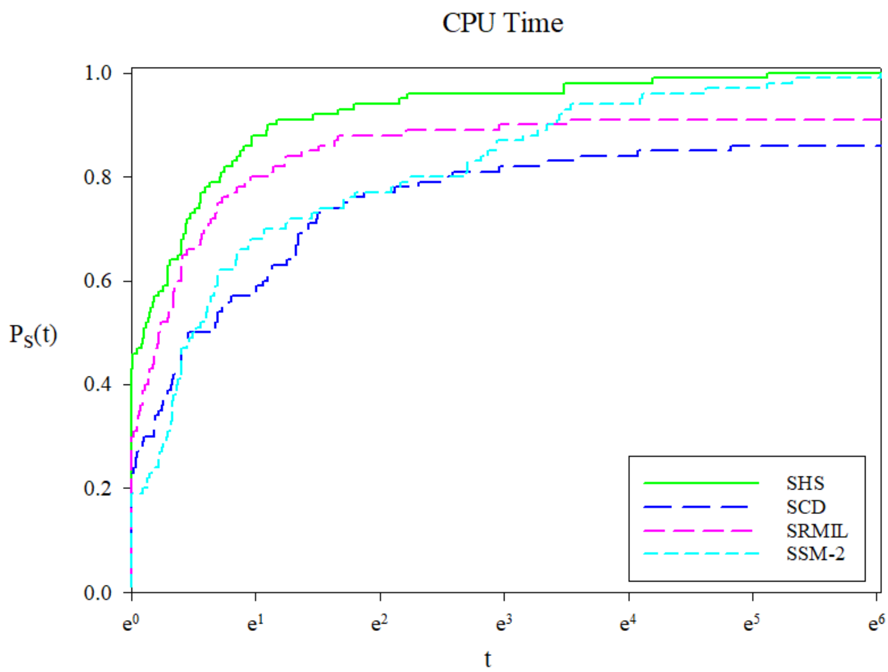

Figure 1 and 2 shows graphs of performance profile, evaluating four algorithms, SHS, SCD, SRMIL, and SSM-2 based on number of iterations and CPU time. In the performance profile, the most successful solver s is generally visible as a higher curve of a graph, meanwhile, the lower curve indicates the slowest performance of the solver s. Based on both graphs, the performance in terms of the fastest solver preceded by the SHS method, where in both performances, it has percentages of 80% and 43%, respectively. Furthermore, the SHS and the SSM-2 methods can solve the problem 100%, which indicates that both methods are robust. However, a close-up of the figure shows that the SSM-2 method has the lowest performance compared to the SHS. Overall, the performance profile demonstrated that the SHS method is the best performing method in both the number of iterations and CPU time, followed by the SSM-2 and the SRMIL methods. Meanwhile, the SCD method shows the lowest performance, which means that the method is less efficient and less robust.

CONCLUSION AND RECOMMENDATIONS

In conclusion, the SHS method has been proven to be the best method among other spectral CG methods, SCD, SRMIL, and SSM-2. The SHS method can solve the highest percentage of test problems with 80% of the fewest iterations and also achieves the shortest CPU time on 43% of the test problems, representing the largest proportion among all methods. Therefore, the numerical experiment showed that the SHS method is the fastest and most efficient method under the inexact strong-Wolfe line search in terms of iteration number and CPU time. Furthermore, it can also solve the standard optimization test problem on a variety of scales, from small-scale to large-scale. As a result, the SHS method outperformed the other spectral CG methods in both terms, number of iterations and CPU time, since it has the highest top curve, which indicates its efficiency, effectiveness, and consistency when compared to the other spectral CG methods. For further study, it is hoped that the analysis could be extended to prove the global convergence, since this study just focuses on the sufficient descent property. In addition, future studies may focus on developing hybrid spectral CG methods by combining the strengths of different spectral CG variants to further improve convergence speed and robustness. Furthermore, the studied spectral CG methods for real-world problems such as image reconstruction, neural network training, or parameter estimation could further validate their practical effectiveness.

ACKNOWLEDGEMENT

The authors would like to thank Universiti Teknologi MARA for supporting this research.

Conflict of Interest
The authors have no conflicts of interest to declare.

Non-funded
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Andrei, N. (2008). An Unconstrained Optimization Test Functions Collection. *Adv. Model. Optim*, *10*(1), 147-161.

Andrei, N. (2020). *Nonlinear Conjugate Gradient Methods for Unconstrained Optimization*. https://doi.org/10.1007/978-3-030-42950-8

Armijo, L. (1966). Minimization of Functions Having Lipschitz Continuous First Partial Derivatives. *Pacific Journal of mathematics*, *16*(1), 1-3. https://msp.org/pjm/1966/16-1/pjm-v16-n1-p01-p.pdf

Barzilai, J. & Borwein, J. M. (1988). Two-Point Step Size Gradient Methods. *IMA journal of numerical analysis*, *8*(1), 141-148. https://doi.org/10.1093/imanum/8.1.141

Birgin, E. G. & Martínez, J. M. (2001). A Spectral Conjugate Gradient Method for Unconstrained Optimization. *Applied Mathematics & Optimization*, *43*(2), 117-128. https://doi.org/10.1007/s00245-001-0003-0

Dennis, E. J. & Schnabel, B. R. (1996). Numerical Methods for Unconstrained Optimization and Nonlinear Equations. In: Society for Industrial and Applied Mathematics.

Dolan, E. D. & Moré, J. J. (2002). Benchmarking optimization software with performance profiles. *Mathematical programming*, *91*, 201-213.

Fletcher, R. (2000). *Practical Methods of Optimization*. John Wiley & Sons.

Ghani, N. H. A., Khadijah, W., & Japri, N. A. (2024). A Modification of Spectral Conjugate Gradient Method (MSM-2) for Unconstrained Optimization. AIP Conference Proceedings,

Goldstein, A. A. (1965). On Steepest Descent. *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, *3*(1), 147-151. https://doi.org/10.1137/0303013

Hestenes, M. R. & Stiefel, E. (1952). Methods of Conjugate Gradients for Solving Linear Systems. In *Journal of Research of the National Bureau of Standards* (Vol. 49, pp. 409-436).

Jamil, M. & Yang, X.-S. (2013). A literature survey of benchmark functions for global optimisation problems. *International Journal of Mathematical Modelling and Numerical Optimisation*, *4*(2), 150-194.

Japri, N. A., Basri, S., & Mamat, M. (2021). New Modification of the Hestenes-Stiefel with Strong Wolfe Line Search. AIP Conference Proceedings,

Nocedal, J. & Wright, S. J. (1999). *Numerical Optimization*. Springer.

Raydan, M. (1997). The Barzilai and Borwein Gradient Method for the Large Scale Unconstrained Minimization Problem. *SIAM Journal on Optimization*, *7*(1), 26-33. https://doi.org/10.1137/S1052623494266365

Rivaie, M., Mamat, M., June, L. W., & Mohd, I. (2012). A New Class of Nonlinear Conjugate Gradient Coefficients with Global Convergence Properties. *Applied Mathematics and Computation*, *218*(22), 11323-11332. https://doi.org/10.1016/j.amc.2012.05.030

Wolfe, P. (1969). Convergence Conditions for Ascent Methods. *SIAM review*, *11*(2), 226-235. https://doi.org/10.1137/1011036

Wolfe, P. (1971). Convergence Conditions for Acent Methods. II: Some Corrections. *SIAM review*, *13*(2), 185-188. https://doi.org/10.1137/1013035