Confluence Journal of Education and Knowledge Management 2025, Vol 2 No 2
ISSN: 2992-2518 (Print) 2992-2526 (Online) ©2025 Federal University Lokoja
DOI: https://doi.org/10.5281/zenodo. 19097814

COMBINATORICS HYBRID FORMULATION OF GRADIENT METHOD FOR NONLINEAR OPTIMIZATION PROBLEMS

OMOLEHIN, J. O., OLUWAFEMI, O. J., MOMOH, S.O.,
KEHINDE, R., GYEGWE, J. M.
joseph.omolehin@fulokoja.edu.ng, myseyi123@gmail.com,
sheidu.momoh@fulokoja.edu.ng, Jessica.gyegwe@fulokoja.edu.ng
Department of Mathematics
Faculty of Science
Federal University Lokoja

Abstract

Conjugate gradient method (CGM) is a widely used method for solving nonlinear optimization problems. Various authors have proposed different classical algorithms based on the formulation of the parameter $\beta. Here$, we propose a new hybrid conjugate gradient method by incorporating the convex combination of the algorithm due to Bamigbola et al (β^{BAN}) $andFletcherReeves$ (β^{FR}). The new method is proven to be descent and Convergent.

Keywords: Hybrid, Convergence, CGM

Citation: Omolehin, J. O., Oluwafemi, O. J., Momoh, S. O., Kehinde, R. and Gyegwe, J. M. (2025). Combinatorics Hybrid Formulation of Gradient Method for Nonlinear Optimization Problems. *Confluence Journal of Education and Knowledge Management,* 2992-2518 (Print) 2992-2526 (Online), Volume 2 (2), pages: 53 - 64. DOI: https://doi.org/10.5281/zenodo.19097814

Introduction

Solving large-scale unconstrained optimization problems arises in many fields such as machine learning and engineering design. Nonlinear conjugate gradient (CG) methods are widely used due to their low memory requirements and strong convergence properties.

Given a general unconstrained optimization problem min $f(x)$, $x \in R^n$ ………………. (1)

There exists many solution techniques in literature for solution of (1). notable among these techniques is the conjugate gradient method (CGM). This is due to low memory requirement and strong local and global convergence properties Fletcher, R. (1970).

Originally designed for solving system of linear equations Hestenes, M.R. &Stiefel, E. (1952), the CGM is an iterative scheme which starts with an initial guess $x_0 \epsilon R^n$ and generates a sequence of approximate solutions $\{x_k\} \epsilon R^n$ as follows:

From an initial guess x_0 , $using\ these\ quence$

$x_{k+1} = x_k + \alpha_k d_k$ .k = 0,1,2,3, … … …. (2)

Where $\alpha_k > 0$is the step size determined by the line search along the direction d_k given by $d_0 = -g_0$

$d_{k+1} = -g_{k+1} + \beta_k d_k$, k = 0,1,2,3, … … (3)

Where β_k is the CG parameter and $g_k = \nabla f(x)$. The strong Wolfe condition Wolfe, P. (1969) is the most popular inexact line search criteria used in the conjugate gradient method and is given by

$f(x + \alpha_k d_k) \leq f(x) + \rho \alpha_k g_k^T d_k$ (4) $|g_{k+1}^T d_k| \leq -\sigma g_{k+1}^T d_k$ (5)

With d_k being the descent direction, i.e$g_k^T d_k < 0$ and $0 < \rho < \sigma < 1$

Various formulation of the parameter β leads to various conjugate gradient methods Andrei, N. (2011).

Some of the well-known formulas in literature are:

HS gT

ISSN: 2992-2518 (Print) 2992-2526 (Online) ©2025 Federal University Lokoja
DOI: https://doi.org/10.5281/zenodo. 19097814

βk= $dkk\pm T1yykk$,βkPR= $ggkTk\pm T1gykk$,$\beta kBAN$= $ggkTk\pm T1yykk$ (6)
βkFR= $|gdkkT+y1k||2$, βkDX= $|-ggk+kT1d|k|2$ βkDY= $|gdkkT+y1k||2$
(7)
Where $y_k = g_{k+1} - g_k$. The CG parameter above can be broadly grouped into two sets based on the term in the numerator viz:
Algorithms with $g_k^T{}_{+1} y_k$ in the numerator and those with $||g_{k+1}||^2$ in their numerator Andrei, N., et al. (2020). It was observed, by numerical experiments, that the once with $g_k^T{}_{+1} y_k$ in their numerator has poor convergence theory but with strong computational capabilities while those with $||g_{k+1}||^2$in their numerator have strong convergence theory but poor computational capability, Ibrahim, A. H (2022). Our contributions include derivation of the hybrid algorithm, proof of descent property, global convergence analysis, and numerical experiments showing superior performance.

New Hybrid Conjugate Gradient Algorithm.

This research proposes an algorithm that generates iterates
$x_{k+1} = x_k + \alpha_k d_k$.$k = 0,1,2,3, \ldots \ldots \ldots.$ Along the direction d_k given by
$d_0 = -g_0$
$d_{k+1} = -g_{k+1} + \beta_k d_k, \quad k = 0,1,2,3, \ldots \ldots$
Where β_k is the convex combination of Fletcher Reeves and Bamigbola et al methods. Research has equally shown that CGM algorithms with $||g_{k+1}||^2$ in the numerator have strong convergence while those with $g_k^T{}_{+1} y_k$ in their numerator has good computational capabilities with poor convergence. Classical CGM are often faced with challenges related to convergence speed and getting trapped in local minima Andrei, N., et al. (2020), To address these challenges, hybrid approaches integrating multiple optimization techniques have gained prominence . Among these, the Nonlinear Conjugate Gradient (NCG) method has shown promise due to its ability to navigate complex optimization landscapes efficiently Abdelhamid, M., Bechouat, T., & Chaib, Y. (2025). By combining the strengths of different classical algorithms, hybrid NCG methods offer enhanced convergence properties and improved robustness.
Here we propose a new hybrid β_K^{NEW} Which is a convex combination of the method due to Fletcher Reeves β_k^{FR} Fletcher, R. 1970 and the method due to Bamigbola et al β_k^{BAN} . As follows let $\beta_K^{NEW} = \theta\beta_k^{FR} + (1-\theta)\,\beta_k^{BAN}$ (8) Where $\theta \epsilon R$ is a scalar to be determined.
From $d_{k+1} = -g_{k+1} + \beta_k^{NEW} d_k(s_k)$ (9)
Using the pure conjugacy condition, i.e. $y_k^T d_{k+1} = 0$ (10)
By pre-multiplying (9) by $y_k$from the right we have
$ykTdk+1 = -ykTgk+1 + ykT\beta kNEWdk(sk) = 0$
Using (8) and (9), the above equation becomes
$ykTdk+1 = -ykTgk+1 + [\theta|gdkkT+y1k||2 + (1-\theta)\; ggkTk\pm T1yykk]ykTsk$
(11)
By expansion and simplification (11) can be expressed as
gT
θ ($gk+T1yykk-$ $|||ggkk+|1||2|2$) $ykTsk=$
$(gkT+1yk-|$ $||g|gk+k|1||2|2)(ykTsk)\; k$
By taking L.C.M
$||gk+1||2(gkT+1yk) - (gkT+1yk)$ $||gk||2$ $||g$
θ ($(gkTyk)||gk||2$) $(ykTsk)$ =
$k||2(gkT+1yk||)g-k||||2gk+1||2(ykTsk)$
making θ the subject of the equation gives

Confluence Journal of Education and Knowledge Management 2025, Vol 2 No 2
ISSN: 2992-2518 (Print) 2992-2526 (Online) ©2025 Federal University Lokoja
DOI: https://doi.org/10.5281/zenodo. 19097814

$(g_{k}^{T}{}_{+}1y_{k})(\quad \|g_{k}\|^{2})-(\|g_{k}+1\|^{2}(y_{k}^{T}s_{k}))$

$\theta = [(\quad \|g_k\|2)(g_kT_{+1}yk)-(gkTyk)(\ \|gk+1\|2)](ykTsk)$ (12) where θ is restricted by $0 \le \theta \le 1$, $for\ \ 0 < \theta or \theta > 1\ \ weset \theta = 1$

Next we consider the algorithm that is associated with our proposed method.

Step 1.

Initialization: set $x_n \in R^n andcomputf(x_0) = g_0$, $Alsowesetd_0 = -g_0 and \alpha_0 = \frac{1}{\|g0\|} andk = 1$

Step 2:

Test for continuity of the algorithm: if $\|g_k\| \le 10^{-6}$, then stop

Step 3

Line search: Compute $\alpha_k satisfying the line search criterion$ (6) and (7) $and update the variables$

$xk+1 = xk + \alpha kdk and compute f(xk+1)$, $g(xk+1) = gk+1$, $sk = xk+1 - xk and y_k = g_{k+1} - g_k$ Step 4:

Parameter Computation if the denominator of (12) is zero then set $\theta = 0$ otherwise compute $\theta as given in$ (12) Step 5

Conjugate gradient parameter computation, β_k^{NEW} compute $\beta_k^{NEW} as given in$(8) Step 6

Computation of direction compute $d_k+1 = -g_k+1 + \beta_k^{NEW}(s_k)$

Step 7

Set k= k+1 and go to step 2

3.0 The sufficient Descent Property

The sufficient descent property is an important property b for any iterative scheme to be globally convergent. Hence we need to prove that the search direction satisfies the sufficient descent condition.

Theorem 1

If $0 < \theta < 1$, then the direction d_{k+1} given by (9) satisfies the sufficient descent condition i.e $d_k^T{}_{+1}g_{k+1} \le -c\|g_{k+1}\|^2$, $c > 0$

Proof

By induction:

If k=1, then

$d_1^T g_1 \le -c\|g_1\|^2$, $c > 0$ (13) Assume it is true for k

i.e let $g_k^T d_k \le -c\ \|g_{k+1}\|^2$

To prove for k+1, we have

$dkT+1gk+1 = -\|gk+1\|2 + [\theta\beta kFR + (1-\theta)\beta kBAN](skTgk+1)$ (14)

$dkT+1gk+1 = -\|gk+1\|2[-\theta \frac{gkT+1yk}{\|gk\|} + (1-\theta)\frac{\|gk+1\|2}{2}](skTgk+1)$ (15)

After expansion and simplification the above equation becomes

Confluence Journal of Education and Knowledge Management 2025, Vol 2 No 2
ISSN: 2992-2518 (Print) 2992-2526 (Online) ©2025 Federal University Lokoja
DOI: https://doi.org/10.5281/zenodo. 19097814

$dkT+1gk+1 = -[-1 - \theta$ ||ygk_kT||y|k|sk|| $-$ s|k|$ggkk$|+|21 + |θ|$gskTkg$||k2]||$gk+1$||2

According to Wolfe's condition (1.4) and (1.5) we have

θ||y ||||s || αdTg 2 (16)

$dkT+1gk+1$ $\leq gkyk\ gkdk$ ||$gk+1$||

θ||y ||||s || 2 (17)

$\leq$ $g_k yk$ ||g_{k+1}||

By descent property, the denominator of the second term is negative hence making the second term positive which makes the quantity in the square bracket positive.

i.e $d_k{}^T{}_{+1} g_{k+1} \leq -c||g_{k+1}||^2$ (18)

which completes the proof.

4.0 Convergence Analysis

In this section we consider the convergence behavior of our proposed method thus

Consider the iterative method defined by (3) , (6) and (7) where d_k satisfying the sufficient descent conditions and α_k satisfied the Wolfe Condition. We assume the objective function satisfies the following conditions:

Assumption

1. f is bounded below in $\mathbb{R}^n$ and is continuously differentiable in a neighborhood $\mathbb{N}$ of the level set $\mathscr{L} = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$ (19)
2. the gradient $\nabla f(x) is lipschitz continuous in$ $\mathbb{N}$, that is there exists a constant $L > 0$ such that $||\nabla f(x) - \nabla f(y)|| \leq L||x - y||, for any x, y \in \mathbb{N}$ (20)

   under the above assumptions we give a the state a lemma which was proved by Zoutendijk, G. (1970) and Wolfe , P. (1969)

Lemma: Suppose that x_1 is the starting point for which the assumption holds, any method in the form (2) and where d_k satisfying the sufficient descent conditions and α_k satisfied the Wolfe Condition. Then we have that

Either $k\lim\to\infty \inf ||gk||$, or $k\geq1$ (||dTk||22 . (21)

Proof:

Assume on the contrary, that is the method is not convergent i.e there exists a real number $c > 0\ such that\ ||g_k|| \geq c \forall k from$ (8) we get

$d_k{}^{NEW} = -g_{k+1} + \{\theta\beta_k{}^{BAN} + (1-\theta)\beta_k{}^{FR}\}s_k$ (22)

$= -gk+1 + \theta gk+1 - gk+1 + \theta\beta kBANsk + (1-\theta)\beta kFRsk$

$rearranging we\ have$

$dkNEW = -\theta gk+1 + \theta\beta kBANsk + (1-\theta)gk+1 + (1-\theta)\beta kFRsk$ (23)

$dkNEW = \theta\beta kBAN + (1-\theta)\beta kFR$

$\therefore ||d_k{}^{NEW}| |\leq ||\beta_k{}^{BAN}|| + || d_k{}^{FR}|||$ (24)

Confluence Journal of Education and Knowledge Management 2025, Vol 2 No 2
ISSN: 2992-2518 (Print) 2992-2526 (Online) ©2025 Federal University Lokoja
DOI: https://doi.org/10.5281/zenodo. 19097814

Furthermore,

$g_k^{T+1} y_k$ —— $L||s$

$||\beta^{kBAN}|| = (g_k^T y_k) \leq$ $g_k y^k{}_k||$ $\leq_{||} Lg|_k|_+ s_1|||^2 L||\omega s\ ||LD — \omega$ (25) similarly,

$||\beta kFR||= |$ —— $||g|gk+k|1||2|2$ $||g\omega k||2gk\omega dk\omega — \varphi$ (26)

From (24), (25) and (26) we have that

$$||dkNEW+1\ || \leq LD\bar{\omega} + \omega\varphi \qquad (27)$$

$$||dk \qquad \omega \quad \varphi \qquad (28)$$

$||g_k+1||^4$ ω^4

$$\Rightarrow \frac{}{k} \quad \frac{}{\omega \quad \varphi} ||d\ ||2^{\geq} (LD - \omega)_2 \qquad (29)$$

$$\sum \infty k=0| —— ||g|dk+k|1||2|4 = \infty \qquad (30)$$

Which gives a contradiction to the assumption. Hence the proof.

4.0 Numerical Results

This section focuses on the numerical implementation of our proposed methods β^{OJO} as compared numerically with its classical method β^{FR}[9] $and \beta^{BAN}$ [10] based on some

indicators. Numerical tests were coded with a Dell intel® core(TM)-i53340M CPU @ 2.70 GHz with 8GB of RAM. The efficiency of an algorithm is studied using standard functions with standard starting points x_0 . The total number of functions evaluations required to find the minimum x^* is usually taken as a measure of the efficiency of the algorithm. Some classical test functions [10] are listed below.

4.1 Extended Rosenbrock Function

Table 1 shows computational results of the extended Rosenbrock function, a widely used nonconvex test problem. All three methods successfully minimized the objective function, with substantial differences in efficiency.

OJO consistently achieved the best performance in terms of iteration count, CPU time, and final accuracy. Notably, the OJO algorithm required less than half the number of iterations needed by FR and BAN across all tested dimensions. Final function values achieved by all methods were close to machine precision, indicating successful convergence. However, the gradient norm attained by OJO was uniformly lower, highlighting superior solution accuracy.

ISSN: 2992-2518 (Print) 2992-2526 (Online) ©2025 Federal University Lokoja
DOI: https://doi.org/10.5281/zenodo. 19097814

Problem 1: Extended Rosenbrock Function

Table 1

N	CIP		CUP Time	F(x)	Norm
100	BAN	70	0.170	9.385220 E-14	6.29e-07
	FR	100	0.235	4.411197e-13	9.42e-07
	OJO	33	0.091	1.181410 E-15	4.94 E-07
1000	BAN	74	0.193	1.365530e-13	7.58e-07
	FR	74	0.339	2.601159e-13	9.68e-07
	OJO	34	0.092	8.489418e-15	1.52e-07
5000	BAN	78	0.414	9.934236e-14	6.47e-07
	FR	72	0.299	7.923557e-14	8.26e-07
	OJO	34	0.167	4.328567e-14	3.43e-07
10000	BAN	78	0.446	1.986806e-13	9.15e-07
	FR	103	0.742	5.503745e-14	3.84e-07
	OJO	34	0.364	8.648529e-14	4.85e-07

4.2 TRIDIA Function

Results in table 2 equally shows the computational capability of the hybrid algorithm (OJO) when tested on the TRIDIA function; which is diagonally dominant and convex but numerically challenging due to its structure. The result shows that both FR and BAN methods failed to produce feasible results, with outputs denoted as 'NAN' (Not a Number), indicating divergence or instability. In stark contrast, the OJO method exhibited strong robustness, converging successfully in all dimensions. The function values attained matched theoretical expectations, and the method maintained computational efficiency.

Problem 2: TRIDIA Function

N	CIP		CUP Time	F(x)	Norm
100	BAN	12	0.034	NAN	NAN
	FR	-	-	-	-
	OJO	515	1.456	6.663075e-01	9.68e-07
1000	BAN	348	1.372	NAN	NAN
	FR	-	-	-	-
	OJO	23	0.123	NAN	NAN
5000	BAN	11978	120.104	NAN	NAN
	FR	- 9	-	-	-
	OJO		0.137	NAN	NAN
10000	BAN	11134	241.805	NAN	NAN
	FR	-	-	-	-
	OJO	19	0.669	NAN	NAN

4.3 Extended McCormick Function

It could also be seen from table 3 that the BAN method consistently failed across all dimensions, and the FR method did not return valid results. Conversely, the OJO algorithm consistently reached acceptable local minima, evidenced by function values with large negative magnitudes and gradient norms below standard convergence tolerances. Importantly, the iteration count and CPU time remained within practical limits even at the largest problem scale.

Confluence Journal of Education and Knowledge Management 2025, Vol 2 No 2
ISSN: 2992-2518 (Print) 2992-2526 (Online) ©2025 Federal University Lokoja
DOI: https://doi.org/10.5281/zenodo. 19097814

N	CIP		CUP Time	F(x)	Norm
100	BAN	29760	75.974	NAN	NAN
	FR	-	-	-	
	OJO	33	0.016	-7.239797e+02	-
					7.84e-07
1000	BAN	29850	119.116	NAN	NAN
	FR	-	-	-	
	OJO	69	0.230	-7.239797e+03	-
					2.84e-07
5000	BAN	29740	420.548	NAN	NAN
	FR	-	-	-	
	OJO	198	0.919	-3.619898e+04	-
					6.79e-07
10000	BAN	29840	729.091	NAN	NAN
	FR	-	-	-	
	OJO	69	0.166	-7.239797e+03	-
					2.84e-07

Problem 3: Extended MCCORMCK Function

4.4 Extended Cliff Function

The research equally saw all three methods converged on the Extended Cliff function, which features abrupt gradients and non-smooth behavior near the minimizer. The final function values were identical across methods, reflecting convergence to the same critical point. Nevertheless, OJO significantly outperformed both BAN and FR in computational cost and accuracy. Across all dimensions, OJO required fewer iterations and less CPU time, and attained smaller final gradient norms.

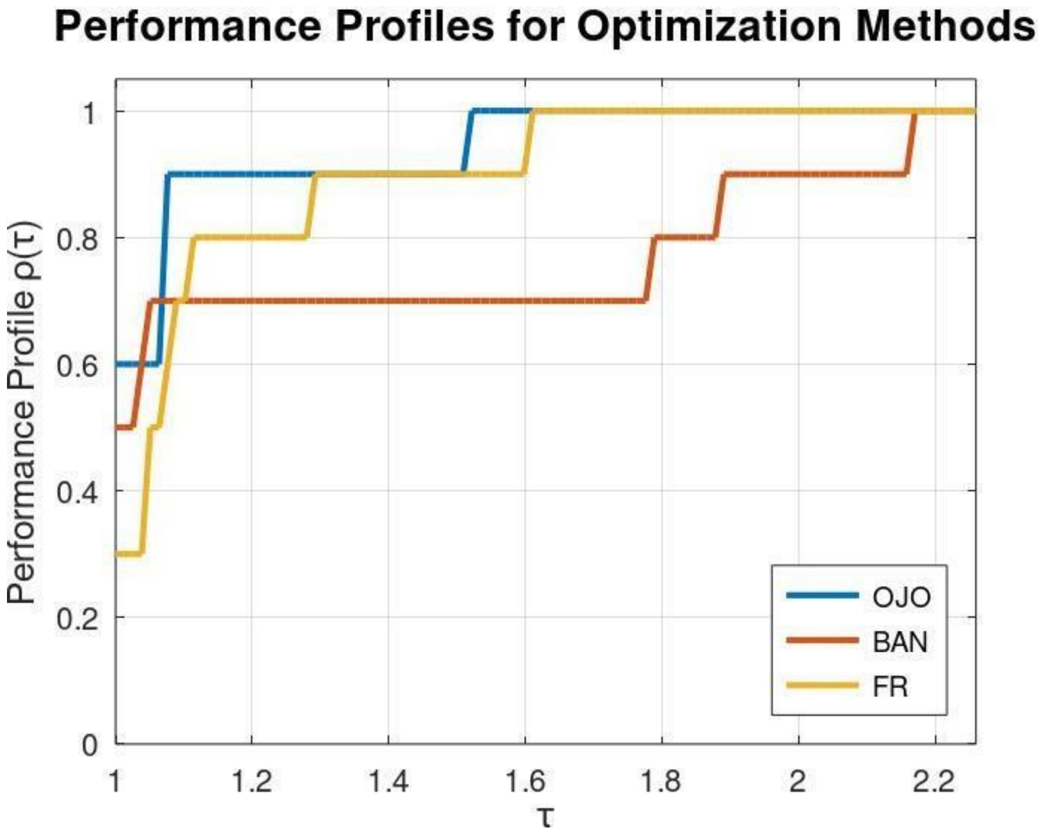
Problem 4: Extended Cliff Function

N	CIP		CUP Time	F(x)	Norm
100	BAN	1090	2.526	NAN	NAN
	FR	518	1.221	9.989331e+00	7.79e-07
	OJO	208	0.458	9.989331e+00	4.13e-07
1000	BAN	1351	3.675	9.989331e+01	5.41e-07 9.43e-07
	FR	488	1.339	9.989331e+01	4.81e-07
	OJO	216	0.577	9.989331e+01	
5000	BAN	1391	8.873	4.994665e+02	5.18e-07 7.54e-07
	FR	613	3.650	4.994665e+02	6.66e-07
	OJO	297	1.540	4.994665e+02	
10000	BAN	1410	14.918	9.989331e+02	8.19e-07 1.98e-07
	FR	316	2.864	9.989331e+02	3.74e-07
	OJO	571	4.215	9.989331e+02	

4.5 Summary of Numerical Observations

The numerical evidence demonstrates the clear superiority of the proposed OJO hybrid method across all considered test problems. In problems where FR and BAN failed to converge or produced unstable results, OJO maintained consistent robustness. Even in cases where all methods succeeded, OJO required markedly fewer iterations and lower computational time. The final gradient norms and objective values confirmed the effectiveness of the method in obtaining accurate solutions.

The figure below summarizes the comparative performance profile for each method:



5.0 Conclusion

This research confirms the advantages of hybrid CG methods. Here we combined Fletcher Reeves (1970) and Bamigbola et al. (2010) updates, the proposed the proposed OJO algorithm leverages theoretical guarantees of FR while achieving the efficiency of practical schemes. This aligns with recent studies emphasizing the value of hybrid CG methods. The strengths of OJO are its robustness, efficiency, and low overhead. Limitations include reliance on line search and restriction to unconstrained smooth problems. Future work includes extensions to constrained problems, adaptive weighting, and applications in large-scale machine learning.

References

Abdelhamid, M., Bechouat, T., & Chaib, Y. (2025). A new hybrid conjugate gradient method as a convex combination methods. *Bulletin of the Transilvania University of Brasov. Series III: Mathematics and Computer Science*, 179-192.

Andrei, N. (2011). Open problems in nonlinear conjugate gradient algorithms for unconstrained optimization. Bulletin of the Malaysian Mathematical Sciences Society. Second Series ,34 (2), 319_330.

Andrei, N., et al. (2020). Nonlinear conjugate gradient methods for unconstrained optimization. Springer.

Bamigbola, O., Ali, M., & Nwaeze, E. (2010). An efficient and convergent method for unconstrained nonlinear optimization. In Proceedings of international congress of mathematicians. hyderabad, India

Fletcher, R. (1970). A new approach to variable metric algorithms, *Computer J.*, Vol. 13, 317–322.

Hager, W. W. & Zhang, H. (2006). A survey of nonlinear conjugate gradient methods, *System modeling and optimization*, eds. F. Ceragioli et al, Proceedings of the 22nd IFIP Conference, Turin, 2006, Springer–Verlag, 67–82.

Hestenes, M.R. &Stiefel, E. 1952 Methods of conjugate gradients for solving linear systems, *Journal of Research of the National Bureau of Standards*, Vol. 29, 409–439.

Ibrahim, A, H., PoomKumam, A. K., & Auwal, B. A. (2022)"An efficient hybrid conjugate gradient method for unconstrained optimization." *Optimization Methods and Software* 37, no. 4 (2022): 1370-1383.

Polak, E., & Ribiere, G. (1969). Note sur la convergence de méthodes de directions conjuguées. Revue françaised' informatique et de recherche opérationnelle. Série rouge, 3 (16), 35_43.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

Wolfe, P. 1969 Convergence conditions for ascent methods, II some corrections, *SIAM Review*, Vol. 13, 185–188.

Zoutendijk, G. 1970 Nonlinear programming, computational methods, in *Integer and Nonlinear Programming*, (J. Abadie, ed.), North–Holland, Amsterdam, 37–86.