On Novel Conjugate Gradient Method for Solving Minimization Problems

Sarah Abdulla Salih[a,*],

Abbas Hasan Taqi[b]

[a,b]Department of Mathematics, University of Kirkuk, College of Science, Kirkuk, Iraq.

***Corresponding author email: scmm24001@uokirkuk.edu.iq**

altaqi@uokirkuk.edu.iq

On Novel Conjugate Gradient Method for Solving Minimization Problems

Sarah Abdulla Salih[a,*],

Abbas Hasan Taqi[b]

[a,b]Department of Mathematics, University of Kirkuk, College of Science, Kirkuk, Iraq.

***Corresponding author email: scmm24001@uokirkuk.edu.iq**

altaqi@uokirkuk.edu.iq

Abstract:

The secant condition constitutes a fundamental aspect of deriving the novel coefficient associated with the conjugate gradient method. The formulas proposed herein incorporate a golden ratio value of 0.618; and they demonstrate higher efficiency compared to conventional conjugate gradient methods. Concurrently, a comparative analysis was undertaken between the proposed formulas and the Pollack- Ribière (PR) method, which is broadly regarded as a robust and efficient strategy for addressing the challenges of unconstrained optimization. It delineates the performance metrics of the newly formulated simulated annealing (SA) algorithm as applied to two distinct scenarios (NI and NF), alongside the reference PR algorithm, which serves as a benchmark for performance evaluation. From the examination of these results, it can be inferred that the SA algorithm demonstrates consistent numerical behavior and attains favorable outcomes under specific conditions. However, the **PR** method persists as the most proficient overall and continues to function as a benchmark for performance evaluation.

Keywords: Conjugate gradient method, Novel conjugate gradient, Convergence property.

" دراسة حول طريقة جديدة للتدرج المترافق لحل مسائل التصغير "

المستخلص:

في هذا البحث، تم اشتقاق صيغة جديدة لطرائق التدرج المترافق بالاعتماد على شرط القاطع، الذي يُعد عنصرًا أساسيًا في الاشتقاق. تعتمد الصيغ المُقترحة على النسبة الذهبية (0.618)، وتُظهر كفاءة أعلى مقارنةً بطرائق التدرج المترافق التقليدية. كما أُجريت دراسة مقارنة بين الطريقة المُقترحة وطريقة بولاك–ريبيير (PR)، التي تُعد معيارًا قويًا وموثوقًا في معالجة مسائل الأمثلية غير المُقيدة. ويعرض البحث مؤشرات الأداء الخاصة بخوارزمية التلدين المُحاكى (SA) المُطوّرة حديثًا، كما طُبقت على حالتين مختلفتين (NI وNF)، إضافةً إلى اعتماد خوارزمية PR كخط أساس لتقييم الأداء. تشير النتائج إلى أن خوارزمية SA تُظهر سلوكًا عدديًا مستقرًا وتحقق نتائج جيدة مع شروط مناسبة. وبشكل عام، تظل طريقة PR الأكثر كفاءة وتستمر في العمل كمعيار لتقييم الأداء.

1. Introduction:

One of the primary benefits of the conjugate gradient (CG) approach is its capacity to efficiently solve a wide range of unconstrained optimization problems in a short period of time and with fewer iterations. CG approaches offer reduced processing costs and storage needs since they do not rely on the Hessian matrix or its approximation. Moreover, the CG approach shows quick global convergence and meets the descent criterion. The distinctive feature of this approach is its simplicity, which facilitates the understanding of algebraic operations and computer code creation. Consequently, the method demonstrates competence and potential in solving large-scale unconstrained minimization issues [2]. The following model will be examined in this study:

$$Min f(x) \ , \ \mathrm{x} \in R^n \qquad \dots\dots\dots (1)$$

where f is a smooth function. The iterative formula below is how the technique suggested in [4] generates a sequence:

$$x_{k+1} = x_k + \alpha_k d_k \qquad \dots\dots\dots (2)$$

where α_k is a step length, and $s_k = x_{k+1} - x_k$ and d_k is the search direction defined by:

$$d_{k+1} = -g_{k+1} + \beta_k s_k \qquad \dots\dots\dots (3)$$

where β_k is a coefficient. Currently, the researchers are modifying the present approach in order to enhance its computing efficiency [10]. The conjugate gradient approach differs mostly in the choice of β_k coefficient. In the nonlinear conjugate gradient approach we presented in [9], β_k has the following form:

$$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{g_k^T g_k} \qquad \dots\dots\dots (4)$$

The conjugate gradient method's numerical performance is mostly influenced by the step size. Determine a step size α_k that satisfies the standard Wolfe rule, a well-known line search criterion:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \qquad \dots\dots\dots (5)$$

$$d_k^T g(x_k + \alpha_k d_k) \geq \sigma \, d_k^T g_k \qquad \dots\dots\dots (6)$$

where $0 < \delta < \sigma < 1$, see [11 and 12]. For general non-convex functions, it is interesting to investigate the possibility of a conjugate gradient method that converges under the standard Wolfe line search conditions. A useful source for research on the most recent conjugate gradient coefficients with significant results and different variations is [3].

Recently, there has been a noticeable tendency among academics to develop the best and fastest optimization techniques due to the increasing number of real-world applications that rely on optimization techniques [12]. These applications span various fields, including. Several scholars have contributed to the discovery of methods to enhance the speed of solution strategies for these new kinds of CG Algorithms by finding CG Parameters of special significance and applying them in the construction of New CG Algorithms. By demonstrating that their improved approach satisfies the Global Convergence and acceptable Descent Conditions, the validity of a globally convergent acceptable descent PR Method is established.

We introduce a novel modified nonlinear conjugate gradient technique based on conjugacy condition, motivated by formula nonnegative and improved descent property. We also develop global convergence findings for related methodology. Included are the initial numerical findings.

2. Novel parameter conjugate gradient:

One of the most amazing mathematical ratios is the Golden Ratio (0.618), which combines geometric harmony and mathematical elegance. Since ancient times, it has garnered a lot of attention because of its innate harmony and visual attractiveness in nature, art, and architecture. Euclidean geometry, the Fibonacci sequence, ancient art, and contemporary architecture are just a few of the domains in which it has an impact, [10].

To derive the new conjugate formulation, we rely on the improved secant condition proposed by Li and Fukushima [7], as:

$$\bar{G}_{k+1}s_k = (G_{k+1} + \eta I)s_k \cong B_{k+1}s_k = y_k + \eta\, s_k \qquad \ldots\ldots\ldots(7)$$

where G_{k+1} is Hessian matrix, B_{k+1} is an approximation to the Hessian, $y_k = g_{k+1} - g_k$ and $s_k = x_{k+1} - x_k$ and I is identity matrix and $\eta\ is\ constant$. Now, we can present the novel secant condition:

$$B_{k+1}s_k = y_k + 0.618\, s_k \qquad \ldots\ldots\ldots(8)$$

However, from (8), we obtain:

$$y_k^T B_{k+1} s_k = y_k^T y_k + 0.618\, y_k^T s_k \qquad \dots\dots\dots (9)$$

Thus, the Newton direction $d_{k+1} = -B_{k+1}^{-1} g_{k+1}$ may be written as follows:

$$d_{k+1}^N = -\left(\frac{y_k^T s_k}{\|y_k\|^2 + 0.618\, y_k^T s_k}\right) g_{k+1} \qquad \dots\dots\dots (10)$$

These methods aim to accelerate the Newton method's convergence. According to the quasi-Newton direction theory, d_{k+1} in $d_{k+1} = -g_{k+1} + \beta_k s_k$ approximates the quasi-Newton approach. Hence, a β_k parameter that:

$$-B_{k+1}^{-1} g_{k+1} = -g_{k+1} + \beta_k s_k \qquad \dots\dots\dots (11)$$

see Nazareth [8].

$$-\frac{y_k^T s_k}{\|y_k\|^2 + 0.618\, y_k^T s_k} g_{k+1}^T y_k = -g_{k+1}^T y_k + \beta_k s_k^T y_k \qquad \dots\dots\dots (12)$$

Then the result is:

$$\beta_k = \left(1 - \frac{y_k^T s_k}{\|y_k\|^2 + 0.618\, y_k^T s_k}\right) \frac{g_{k+1}^T y_k}{s_k^T y_k} \qquad \dots\dots\dots (13)$$

For the new formula, it is called SA method.

Algorithm SA.

Initialization. Given $x_0 \in R^n$,

set $k = 0$, $d_0 = -g_0$.

Stage 1. Stop if $\|g_k\| \le \varepsilon$.

Stage 2. Using (5) and (6) to compute α_k.

Stage 3. Utilize (13) to calculate β_k and $x_{k+1} = x_k + \alpha_k d_k$.

Stage 4. Estimate $d_{k+1} = -g_{k+1} + \beta_k d_k$.

Stage 5. Put $k = k + 1$ and go to stage 2.

3. Convergence analysis

To give the convergence result, the following assumptions are given.

Assumptions:

1. Let Ω be the level set at the initial point x_0 :

$$\Omega = \{x \in R^n | f(x) \leq f(x_0)\} \qquad \qquad \ldots\ldots\ldots(14)$$

is bounded.

2. f is smooth and the same order between its gradient and the variables holds, i.e.,

$$\|g(o) - g(c)\| \leq L\|o - c\|, \forall o, c \in \Omega. \qquad \qquad \ldots\ldots\ldots(15)$$

See [6].

Examine the theorem to ascertain the potential utility of the descent condition.

Theorem 1.

If condition $s_k^T y_k \neq 0$ is satisfied, the search directions generated by equations (14) and (6) constitute valid descent directions.

Proof:

Since $d_0 = -g_0$, , we take $g_0^T d_0 = -\|g_0\|^2 < 0$. Let $d_k^T g_k \leq 0$ be assumed as true. The product of (6) by g_{k+1}, is:

$$d_{k+1}^T g_{k+1} = -g_{k+1}^T g_{k+1} + \left(1 - \frac{y_k^T s_k}{\|y_k\|^2 + 0.618\, y_k^T s_k}\right) \frac{y_k^T g_{k+1}}{s_k^T y_k}\, s_k^T g_{k+1} \qquad \ldots\ldots\ldots(16)$$

Additionally, by combining equations (8) and (9), we obtain the following relation:

$$y_k^T y_k \cong y_k^T y_k + 0.618\, y_k^T s_k \qquad \qquad \ldots\ldots\ldots(17)$$

The Lipschitz condition yields: $y_k^T g_{k+1} \leq L s_k^T g_{k+1}$ and $s_k^T y_k \leq L\, s_k^T s_k$. Consequently, we may express:

$$d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + \left(\frac{L y_k^T s_k - y_k^T s_k}{y_k^T y_k}\right) \frac{L (s_k^T g_{k+1})^2}{s_k^T y_k} \qquad \ldots\ldots\ldots(18)$$

Given that L and $\alpha_k^2$are exceedingly minimal, it is crucial to note that:

$$d_{k+1}^T g_{k+1} \leq 0 \qquad \text{..........(19)}$$

The conclusive proof has been accomplished. Thus, we may use Lemma 1 to show the following conclusion, see [13].

Lemma 1.

If assumptions (1) and (2) hold, consider any conjugate gradient technique as a descent direction and α_k selected by the strong Wolfe line search. Should:

$$\sum_{k=0}^{\infty} \frac{1}{\|d_{k+1}\|^2} = \infty \qquad \text{..........(20)}$$

Then

$$\lim_{k\to\infty} (inf\|g_{k+1}\|) = 0. \qquad \text{..........(21)}$$

We may use the lemma 1 condition to show the following result.

Theorem 2.

If the following conditions are met by a constant $\mu > 0$:

$$(\nabla f(u) - \nabla f(w))^T \geq \mu\|u - w\|^2, \quad \forall u, w \in R^2. \qquad \text{..........(22)}$$

Under the assumptions of Lemma 1, we have:

$$\lim_{k\to\infty} (inf\|g_{k+1}\|) = 0. \qquad \text{..........(23)}$$

Proof:

from (3)and (13) we get:

$$\|d_{k+1}\| = \|g_{k+1}\| + \left|(1-\omega)\frac{g_{k+1}^T y_k}{s_k^T y_k}\right| \|s_k\| \qquad \text{..........(24)}$$

where $\omega = y_k^T s_k / \|y_k\|^2 + 0.618\ y_k^T s_k$. Using Cauchy's inequality makes it leads to:

$$\|d_{k+1}\| \leq \|g_{k+1}\| \qquad \text{..........(25)}$$

$$+|(1-\omega)|\frac{\|g_{k+1}\|\|y_k\|}{\|s_k\|\|y_k\|}\|s_k\|$$
$$\leq (2-\omega)\|g_{k+1}\|$$

Thus, $\|\nabla f(u)\| \leq \Gamma$ implies that:

$$\sum_{k\geq 1}\frac{1}{\|d_k\|^2}\left(\frac{1}{2-\omega}\right)\frac{1}{\Gamma}\sum_{k\geq 1}1=\infty \qquad \ldots\ldots\ldots(26)$$

by applying Lemma 1, implies that $\lim_{k\to\infty} \inf\|g_k\| = 0$.

4. Numerical Results:

The performance of the novel parameter in the CG-method is compared to that of other classical CG-method (Polak-Ribière (PR) algorithm) in this section. For each test problem, we have chosen a few large-scale unconstrained optimization tasks from (Andrie, 2008) [1]. We have taken into consideration numerical experiments with the number of variables $n = 100$, 1000 for each test function.

The polak - Ribière (PR) method, a well-known CG algorithm, was compared to the novel parameter. Standard Wolfe line search criteria with $\delta = 0.001$ and $\sigma = 0.9$ are used to implement each of these methods. The $\|g_{k+1}\| = 10^{-6}$ is the halting criterion in each of these situations.

Every code is developed using the F90 default compiler settings and the double precision FORTRAN language. For the sake of our comparisons, we keep track of the number of restart (IRS), number of iterations (NOI) and number of function evaluation (NIF).

Table 1- Numerical results of SA and HS algorithms

		PR			SA		
P.No.	**N**	**NI**	**NR**	**NF**	**NI**	**NR**	**NF**
Trigonometric	100	19	10	35	20	13	37
	1000	39	24	69	35	20	62
Extended Rosenbrock	100	50	19	102	34	18	72
	1000	93	65	143	35	19	81

Extended White & Holst	100	38	16	92	34	18	73
	1000	348	317	403	33	17	71
Extended Beale	100	47	26	74	15	9	29
	1000	38	19	61	15	9	29
Perturbed Quadratic	100	106	40	162	94	23	141
	1000	358	115	561	350	98	553
Extended Tridiagonal 1	100	21	9	42	27	10	55
	1000	45	21	82	29	11	58
Extended PSC1	100	15	10	31	10	7	21
	1000	8	6	17	7	5	15
Extended Maratos	100	94	34	172	68	31	161
	1000	98	36	185	78	40	198
Extended Wood	100	81	37	126	33	14	65
	1000	73	35	111	32	12	62
Extended Hiebert	100	87	34	199	80	50	176
	1000	87	35	196	80	50	173
Extended Quadratic Penalty QP2	100	35	13	74	24	13	56
	1000	51	20	106	34	19	84
ARWHEAD (CUTE)	100	10	5	20	9	5	17
	1000	39	23	642	18	13	153
Partial Perturbed Quadratic	100	85	28	136	80	21	124
	1000	506	264	713	264	51	449
LIARWHD (CUTE)	100	25	13	45	23	13	43
	1000	48	33	74	21	12	47
ENGVAL1 (CUTE)	100	25	10	44	27	11	48
	1000	67	52	1403	44	30	728

DENSCHNA (CUTE)	100	23	13	39	16	9	29
	1000	25	14	43	12	7	23
DENSCHNC (CUTE)	100	17	8	32	17	10	32
	1000	128	66	165	15	10	29
DENSCHNF (CUTE)	100	22	19	37	22	19	38
	1000	23	20	40	20	17	35
Extended Block-Diagonal BD2	100	122	62	156	16	10	29
	1000	130	66	166	15	9	27
SINCOS	100	15	10	31	10	7	21
	1000	8	6	17	7	5	15
Total		**3149**	**1653**	**6846**	**1803**	**765**	**4159**

We display the numerical comparative findings graphically using Dolan and Morse performance profiles [5]. In terms of number of iterations (NI), restart (NR), and function evaluation (NF), Figures 1, 2, and 3, respectively, show that SA performs well when compared to alternative approaches. Tables 1 and Figures 1, 2, and 3 demonstrate that the SA approach outperforms the PR method in terms of numerical performance. In comparison to the other four approaches, the SA is more efficient since it requires fewer iterations (NI), restart (NR), and function evaluation (NF).

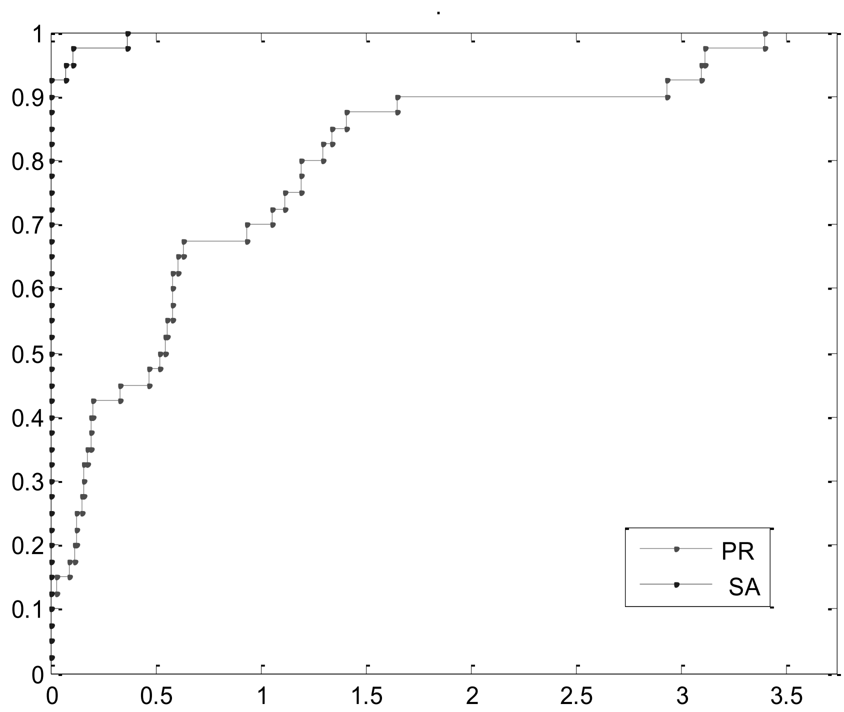

Figure 1 Number of Iteration

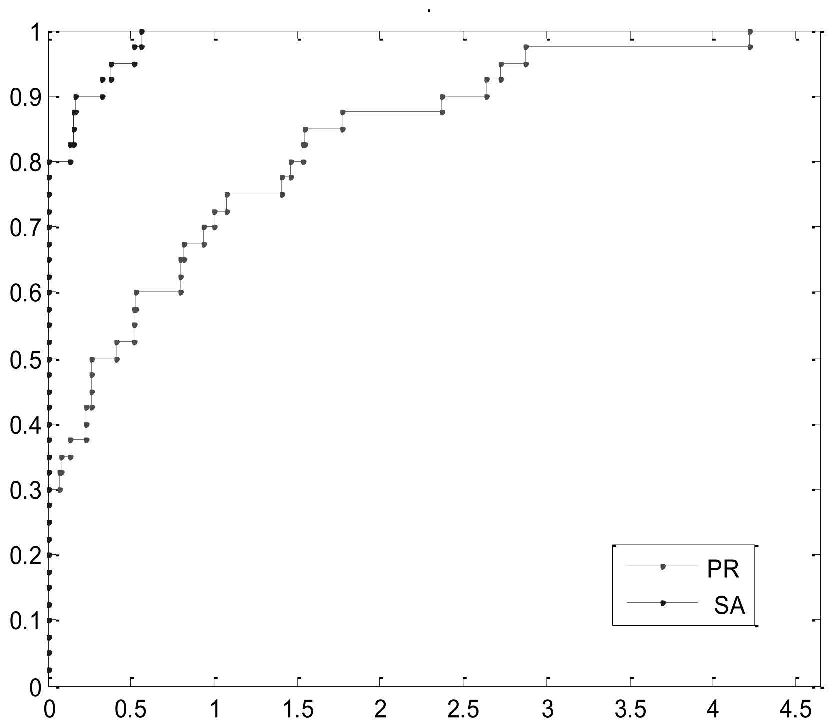

Figure2. Number of Function Evaluation

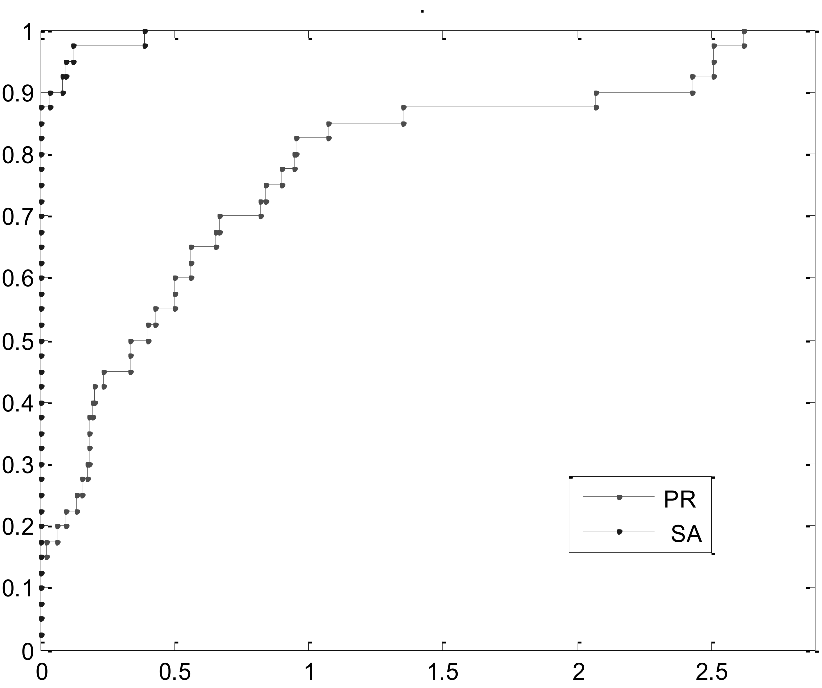

Figure3. Number of restart

5. Findings

This study presents a Secant Algorithm based on the secant condition. The algorithm shows good numerical stability and performs well in specific situations. However, its overall effectiveness is lower than that of traditional conjugate gradient methods. Still, the proposed method remains competitive for certain types of problems. The comparison in this paper only includes traditional conjugate gradient methods. No further improvements or other benchmarks were tested. Therefore, the reported performance is not general but applies only to particular problem cases. The Polak–Ribière method continues to be the most reliable and efficient benchmark for unconstrained optimization problems.

References:

[1] Andrei N. (2008), An Unconstrained Optimization test function collection. Adv. Model. Optimization. 10. pp.147-161.

[2] Basim, A. Hassan, Zeyad, M. A., Hawraz, N. J. (2019). A descent extension of the Dai– Yuan conjugate gradient technique. Indonesian Journal of Electrical Engineering and Computer Science, 661–668.

[3] Basim, A. Hassan, Hussein, O. D., Azzam, S. Y. (2020). A new kind of parameter conjugate gradient for unconstrained optimization. Indonesian Journal of Electrical Engineering and Computer Science, 404–411.

[4] Dai, Y. H., Han, J. Y., Liu, G. H., Sun, D. F., Yin, X., Yuan, Y. (1999). Convergence properties of nonlinear conjugate gradient methods. SIAM Journal on Optimization, 10, 348–358.

[5] Dolan, E. D., Moré, J. J. (2002). Benchmarking optimization software with performance profiles. Mathematical Programming, Series A, 91, 201–213. https://doi.org/10.48550/arXiv.cs/0102001

[6] Hawraz N. Jabba and Basim A. Hassan, (2021), Two-versions of descent conjugate gradient methods for large-scale unconstrained optimization, Indonesian Journal of Electrical Engineering and Computer Science Vol. 22, No. 3, pp. 1643-1649

[7] Li, H., Fukushima, M. (2001). A modified BFGS method and its global convergence in nonconvex minimization. Journal of Computational and Applied Mathematics, 129, 15–35.

[8] Nazareth, L. (1979). A relationship between BFGS and conjugate gradient algorithms. SIAM Journal on Numerical Analysis, 61, 794–800.

[9] Polak, E., Ribière, G. (1969). Note sur la convergence de directions conjuguées. Revue Française d'Informatique et de Recherche Opérationnelle, 16, 35–43.

[10] Rao, S. S. (2019). Engineering Optimization: Theory and Practice. Wiley. DOI: 10.1002/9781119454816

[11] Wolfe, P. (1969). Convergence conditions for ascent methods. SIAM Review, 11, 226–235.

[12] Wolfe, P. (1971). Convergence conditions for ascent methods II: Some corrections. SIAM Review, 3, 185–188.

[13] Zoutendijk, G. (1970). Nonlinear programming computational methods. In J. Abadie (Ed.), Integer and Nonlinear Programming, 37–86. North-Holland, Amsterdam.