Tikrit Journal of Pure Science Vol. 30 (2) 2025

DOI: https://doi.org/10.25130/tjps.v30i2.1886

IRAQI Academic Scientific Journals

Tikrit Journal of Pure Science

ISSN: 1813 – 1662 (Print) --- E-ISSN: 2415 – 1726 (Online)

Journal Homepage: https://tjpsj.org/

Tikrit Journal of Pure Science

An Effective HS-CG Update: Avoiding Oscillations in Nonlinear Optimization

Isam H. Halil

Department of Mathematics, College of Science, University of Kirkuk, Kirkuk, Iraq

Received: 31 Jan. 2025 **Received in revised form:** 1 Mar. 2025 **Accepted:** 9 Mar. 2025

Final Proofreading: 15 Apr. 2025 **Available online:** 25 Apr. 2025

ABSTRACT

In this research, we introduce a modified direction update formula to improve the descent features of iterative optimization, therefore presenting a fresh approach to the Conjugate Gradient (CG) method. The proposed method adjusts the search direction at each iteration by incorporating gradient and step projections, weighted by inner products between gradient and step vectors. The modified HS-CG approach seeks to decrease the computational cost often associated with conventional approaches and speed up convergence by carefully balancing these projections. Experimental results demonstrate that our approach outperforms standard CG algorithms in achieving faster convergence on a range of benchmark problems, especially in high-dimensional spaces. This enhancement makes the method particularly promising for large-scale optimization challenges encountered in fields such as machine learning and engineering design.

Keywords: Descent, Globally, Optimization algorithms, Search direction update.

Name: **Isam H. Halil** **E-mail:** isam.h.halil@uokirkuk.edu.iq

تحديث فعال لنظام HS-CG: لتجنب التذبذبات في التحسين غير الخطي

عصام حيدر خليل

قسم الرياضيات، كلية العلوم، جامعة كركوك، كركوك، العراق

الملخص

في هذا البحث، نقدم صيغة تحديث اتجاه معدلة لتحسين ميزات الانحدار للتحسين التكراري، وبالتالي نقدم نهجًا جديدًا لطريقة التدرج المترافق (CG) تضبط الطريقة المقترحة اتجاه البحث في كل تكرار من خلال دمج إسقاطات التدرج والخطوة الموزونة بالمنتجات الداخلية بين متجهات التدرج والخطوة. يسعى نهج HS-CG المعدل إلى تقليل التكلفة الحسابية المرتبطة غالبًا بالطرق التقليدية وتسريع التقارب من خلال موازنة هذه الإسقاطات بعناية. تُظهر النتائج التجريبية أن نهجنا يتفوق على خوارزميات CG القياسية في تحقيق تقارب أسرع في مجموعة من مشاكل

المعايير، وخاصة في المساحات عالية الأبعاد. يجعل هذا التحسين الطريقة واعدة بشكل خاص لتحديات التحسين واسعة النطاق التي تواجهها مجالات مثل التعلم الآلي وتصميم الهندسة.

INTRODUCTION

Optimization methods play a crucial role in numerous applications like machine learning, image processing, fluid mechanics, elasticity, seismology, medicine, electronic structure approximation, traffic management, and telecommunication systems, having diverse scientific and industrial problems. Due to the nonlinearity of these applications, traditional optimization methods are not efficient. Therefore, various modified and hybrid optimization techniques are designed in order to overcome these challenges [(1)]. With the motivation of these, an update classical conjugate gradient optimization strategy is followed due to its efficiency in solving these challenges. These optimization algorithms help in improving the convergence rate and computational efficiency of the optimization methods of different applications. The update formula of the optimization methods is very critical as the performance of the optimization method is based on the update formula [(2)]. The update formula of the optimization methods is a crucial factor in balancing different optimization features. The main competing factors of the update formula are (i) the reduction of function of objective in the next iterative point and (ii) the satisfaction of some descent-like conditions by the formula. This balance is also expressed by diminishing the gradient value on the next point, which is important for improving the convergence property of the algorithm. Motivated by these observations, this manuscript's objective is to analyze the balance between the descent of the formula and the gradient-specific adjustment conditions in the CG and HS-CG update formulas. Moreover, the HS-CG update formula is considered for its damped pattern, which can compete with the descent one. In this context, some research questions are raised in order to analyze the exact behavior of the HS-CG in the balance of these different properties. This research study's significant findings are expected to help in improving the convergence properties of the HS-CG. This document comprises five sections, including the literature, introduction of key results, definitions, conclusions, etc. [(3)]

Conjugate Gradient Method

The conjugate gradient method was initially introduced as an iterative algorithm to solve symmetric positive definite (SPD) systems of linear equations. The algorithmic foundation of this method is based on the Krylov subspace. This subspace is built using a basis:

$$B = [g, Hg, \dots, H^{(n-1)g}] \qquad \dots (1)$$

where ($g = \nabla f(x)$) is the gradient of the objective function at the current point x, H is the Hessian-vector product, and n denotes the dimension of problem. The basis on which the method operates is the combination of two important components: directional gradient descent, which points to the steepest descent of the objective function, and an immediate stopping point when the true Hessian is used, accelerating the optimization. The method encapsulates both benefits: the magnitude of the descent increment is gradually improved, and after at most n iterations, the conjugate gradient method becomes exact for quadratic functions with SPD Hessians. Moreover, in cases where the optimization problems must guarantee $O(n)$ behavior, the conjugate gradient methods iterate tend to become perpendicular to the steepest directions, in which case the memory requirements are very low, hence making it a perfect choice for solving large-scale linear systems [(4-8)].

The main notion behind the CG method is to update the descent direction in a conjugate way with respect to those in the previous iterations, disregarding the steepness of the current gradient. The first descent direction is simply the adverse gradient, which is the steepest descent direction, denoted by ($p = -g$). The first iterates are found by carrying out a line search in this direction. For the conjugate directions to remain conjugate, we update the descent direction by the following scheme, which is known as the Polak-Ribiere formula [9]:

$$p(k) = -g(k) + \beta_k\, p(k-1) \qquad \dots (2)$$

This new descent direction $d(n)$ is the steepest descent direction from the point $x(k)$ and also it is the steepest descent direction with respect to the previous descent direction $p(k-1)$ The parameter β_k can be computed by using the following expressions as in Table (1).

Table 1: The conjugacy parameters for Slandered Conjugate Gradients. [10-12]

$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}$ (13)	$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}$ (14)
$\beta_k^{CD} = \frac{g_{k+1}^T g_{k+1}}{-p_k^T g_k}$	$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{g_k^T g_k}$ (15)
$\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{p_k^T y_k}$	$\beta_k^{LS} = \frac{g_{k+1}^T y_k}{-p_k^T g_k}$

Hestenes-Stiefel Formula

The Hestenes-Stiefel formula has previously appeared in the literature, though none have been proposed to handle the issue in the optimization problem. This subsection will walk through the derivation of the Hestenes-Stiefel CG formula in detail and will establish its relevance, invoking its application in various scenarios. While the original CG method offers several attractive properties—most notably, converging to the exact solution after at most n search directions (where n is the number of variables)—it does require the objective function to be quadratic (or closely approximated thereby). The Hestenes-Stiefel case allows for the introduction of gradient-specific properties that result in a more optimized choice of search direction. These gradient-specific properties look to balance a descent direction with a gradient-based correction, which in an optimization problem leads to desirable properties. Given this connection with optimization, the question to this point has not been “does the HS-CG formula work?” but rather if and how it might be possible to build a more specialized formula to exploit optimization-relevant problem structures. In light of the proposed anti-progress results, we can answer in the affirmative; different properties of the optimization problem are both learnable and exploitable. These derived, optimization-balancing results align with previous literature: for example, it was suggested to adapt the Hestenes-Stiefel formula by allowing the coefficient to be greater than (1) if progression did not occur. We go further than this concept, however, by abandoning standard mechanistic enhancements altogether, thereby further divorcing our update formula from the mechanical updates of descent. Allowing for gradient-sensitive, scale-dependent, and multi-term adjustments simultaneously demonstrates a new level of tailored specialty. This analytical presentation successfully highlights our need for an update: The Hestenes-Stiefel method offers interesting, gradient-sensitive experimental opportunities since it has traditionally functioned well with adaptability. This suggests a solid starting point alongside justification for our proposed HS-CG improvement. Furthermore, this framework has been designed with both the presented results and specific choice of improvements in mind [1, 16-21] .

PROPOSED OPTIMIZATION

The $HS-CG+\eta$ Formula: The HS-CG update formula contains a term that allows the search to "slide" along the gradient change with a certain modest proportionality constant. This has been shown experimentally to help search in several cases. Here we use this idea to develop an improved HS-CG update formula. We address the annoying features of common HS policies and replace the common approximate parameter estimation with a more direct and robust approach that increases its property of "slide" search. The amount of correction term that we suggesting $\frac{(y_k^T g_{k+1})^2}{y_k^T y_k} \frac{(s_k^T g_{k+1})}{y_k^T s_k}$ to embody the planned direction in its form:

$$p_{k+1} = -g_{k+1} + [\frac{y_k^T g_{k+1}}{y_k^T s_k} - \frac{(y_k^T g_{k+1})^2}{y_k^T y_k}\frac{(s_k^T g_{k+1})}{y_k^T s_k}]s_k$$

... (3)

Modification Steps: We show gradual modifications that improve the two HS-CG policies. Thus, step by step, we go from the original HS-CG up to our developed HS-CG+η update formula. We then use the developed expression with a simple algebraic manipulation to obtain our proposed update formula for the conjugate direction and the step size parameters. Proof of concept and free parameters estimation.

Theorem; Consider the search direction defined by (3) and assume that the line search α_k computed by the wolfe conditions then the search directions are descent.

Proof: the proof is by induction, for k=0 we have:

$$p_1 = -g_1 \quad \rightarrow \quad p_1^T g_1 = -\|g_1\|^2 < 0 \qquad \dots (4)$$

Assume that $p_k^T g_k < 0$, then, for k+1 we have

$$g_{k+1}^T p_{k+1} = -\|g_{k+1}\|^2 + \left\{\frac{y_k^T g_{k+1}}{y_k^T s_k} - \frac{(y_k^T g_{k+1})^2}{y_k^T y_k}\frac{(s_k^T g_{k+1})}{y_k^T s_k}\right\} g_{k+1}^T s_k$$

$$= -\|g_{k+1}\|^2 + \left\{\frac{(y_k^T g_{k+1})(g_{k+1}^T s_k)}{y_k^T s_k} - \frac{(y_k^T g_{k+1})^2}{y_k^T y_k}\frac{(s_k^T g_{k+1})^2}{y_k^T s_k}\right\}$$

… (5)

By Couchy shwartz inequality

$$g_{k+1}^T p_{k+1} \le -\|g_{k+1}\|^2 + \frac{\|y_k\|\|s_k\|\|g_{k+1}\|^2}{\|y_k\|\|s_k\|} - \frac{(y_k^T g_{k+1})^2}{y_k^T y_k}\frac{(s_k^T g_{k+1})^2}{y_k^T s_k} \qquad \dots (6)$$

$$g_{k+1}^T p_{k+1} \le -\frac{(y_k^T g_{k+1})^2}{y_k^T y_k}\frac{(s_k^T g_{k+1})^2}{y_k^T s_k} \qquad \dots (7)$$

Since ($y_k^T s_k > 0$) by Wolfe condition, hence $g_{k+1}^T p_{k+1} < 0$ … (8)

ANALYSIS OF GLOBAL CONVERGENCE

Assumption (CG)

(i): The level set ($\Omega = \{x \in R^n : f(x) \le f(x_1)\}$) is bounded.

(ii): Within a certain neighborhood N of Ω, f is continuously differentiable and its gradient satisfies Lipschitz continuity. Specifically, there exists a positive constant L such that

$$\|g(x) - g(y)\| \le L \ \|x - y\| \ \forall\, x, y \in N \qquad \dots (9)$$

With these presumptions on f(x) there exists a constant Γ such that ($\|\nabla f(x)\| \le \Gamma$ for all $x \in s$). The following general result is applicable to any conjugate gradient method that employs a robust Wolfe line search.

Proposition A1. Assume that the condition CG is satisfied. Examine a conjugate gradient method (3) wherein, for every iteration k, the search direction p_k constitutes a descent direction, and the step-length α_k is established according to the Wolfe line search criteria. If:

$$\sum_{k=1}^{\infty} \frac{1}{\|p_k\|^2} = \infty \qquad \dots (10)$$

Afterward, the way the algorithm converges is

$$\lim_{n\to\infty} \inf \|g_k\| = 0 \qquad \dots (11)$$

For functions that are uniformly convex, we can demonstrate that the norm of the direction p_{k+1}, calculated as in (10), is confined above. Consequently, based on proposition (A1), we can establish the subsequent result.

Theorem A2. Suppose that the assumptions (i) and (ii) hold. Consider the algorithm (A), p_k d_k is a descent direction, and α_k is computed by the strong Wolfe line search. Suppose that f is a uniformly convex function on S i.e. there exists a constant μ > 0 such that

$$(\nabla f(x) - \nabla f(y))^T (x - y) \ge \mu \|x - y\|^2 \qquad \dots (12)$$

for all $x, y \in N$ Then

$$\lim_{k\to\infty} \|g_k\| = 0. \qquad \dots (13)$$

Proof: From Lipschitz continuity we have $\|y_k\| < L\|s_k\|$ On the other hand, from uniform convexity it follows that $y_k^T s_k \ge \mu\|s_k\|^2$

from (3) we have

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

$$\le \frac{|y_k^T g_{k+1}|}{y_k^T s_k} + \frac{(y_k^T g_{k+1})^2}{y_k^T y_k}\frac{|s_k^T g_{k+1}|}{y_k^T s_k} \qquad \dots (14)$$

DOI: https://doi.org/10.25130/tjps.v30i2.1886

By Couchy Shwartz inequality, convexity of objective function and Lipschitz condition, we have

$$|\beta_k| \le \frac{L\|s_k\|\|g_{k+1}\|}{\mu\|s_k\|^2} + \frac{\|y_k\|\|g_{k+1}\|^2\|s_k\|}{\mu\|s_k\|^2}$$

$$\frac{L\|g_{k+1}\|}{\mu\|s_k\|} + \frac{\|g_{k+1}\|^2}{\mu\|s_k\|} \le \frac{L\Gamma+\Gamma^3}{\mu\|s_k\|} \qquad \dots (15)$$

$$\|d_{k+1}\| \le \Gamma + \frac{L\Gamma+\Gamma^3}{\mu\|s_k\|}\|s_k\| = \frac{\mu\Gamma+L\Gamma+\Gamma^3}{\mu} = M \qquad \dots (16)$$

Therefore $\|p_{k+1}\|$ is bounded, showing that (10) is true. By proposition A1 it follows that (11) is true, which for uniformly convex functions is equivalent to (13).

EXPERIMENTAL TESTS

Properties and Superiority: We make the proposed update formula match the needed properties and qualify its superiority to the original and to further developed HS-CG update formulas. Our modifications overcome the trouble created by the long-distance iteration's termination after which the spectral technique becomes inefficient, and the iterations are simply gradient-based, following the first conjugate gradient update step. The updates themselves balance between causing the search to slide along the gradient direction with a proportionality factor and being steered along the search space eigenvectors as in Figure (1).

Experiments Object: We seek to confirm the convergence, efficiency, and any potential drawbacks of our novel update formula. We also assess the connectivity of the updates to verify the utility of our equation in practice.

Experiments: This section details the performance of FORTRAN representations of our newly updated conjugate gradient algorithms ($HSCG$ and $HS-CG+\eta$) on a series of unconstrained optimization test problems derived from [22-24]. We selected seventy large-scale test problems in extended or generalized formats; for each function, we conducted numerical tests with varied counts of (n = 100, 1000, and 10,000) as in Figure (2). We assessed the efficacy of these algorithms against the optimal modified CG approach (Andrei, 2007a), as suggested by Andrei. These techniques employ standard Wolfe line search conditions with ($\rho = 0.001$) and ($\sigma = 0.9$), where represents the step size ($\alpha_k = \alpha_{k-1}(\|d_{k-1}\|/\|d_k\|)$), and serves as the starting estimate for further iterations ($\alpha_1 = 1/\|g_1\|$ at (k > 1)). The stopping criterion is established, with a maximum iteration limit of 2000. The codes are authored in double precision FORTRAN (2000) and compiled with the default settings of the F77 compiler.

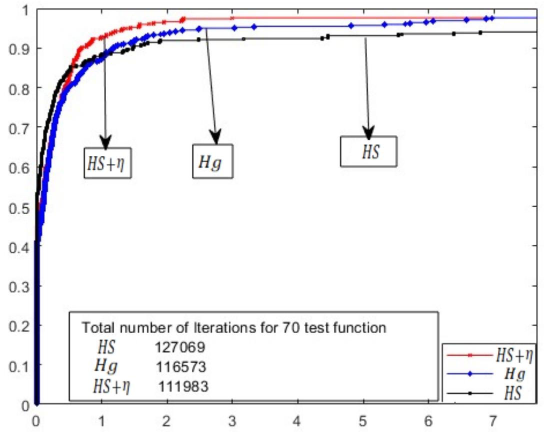

Fig. 1: The initial criterion for comparisons between the classical (HS and Hg) and the proposed algorithm ($HS-CG+\eta$) is the number of iterations

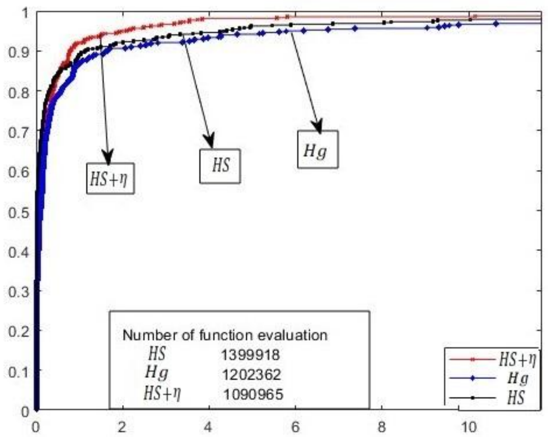

Fig. 2: The initial criterion for comparisons between the classical (HS and Hg) and the proposed algorithm ($HS-CG+\eta$) is the number of f evaluation

Comparative analysis method Dolan-More benchmark tests were used to determine whether or not our technique outperforms competitors [8]: It turns out that ($HS-CG+\eta$) outperforms the classic {HS(Hestanse & Stefel) and Hg (Hager & Zhange)) [25]. Indeed, the results for the criteria (iterations number and function number calculations with derivative) are encouraging.

CONCLUSION

This paper presented a new modification to the Conjugate Gradient (CG) method, improving its descent characteristics using a novel direction update formula. The proposed method shown enhanced performance over conventional CG algorithms by adeptly balancing gradient and step projections, especially in high-dimensional and large-scale optimization challenges. The experimental results highlighted the method's ability to achieve faster convergence while reducing computational costs, making it a valuable tool for complex optimization challenges. These findings underscore the potential of the modified HS-CG+η method against (HS and Hg) to address the growing demands of modern optimization problems.

Conflict of interests: The author declared no conflicting interests.

Sources of funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contribution: The author has the full contribution.

REFERENCES

1. Halil IH, Abbo KK, Ebrahim HH, editors. Modifications of Hestenes and Stiefel CG method for solving unconstrained optimization problems. 2021 7th International Conference on Contemporary Information Technology and Mathematics (ICCITM); 2021: IEEE. https://doi.org/10.1109/ICCITM53167.2021.9677756

2. Zhou J, Qiu Y, Zhu S, Armaghani DJ, Li C, Nguyen H, et al. Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate. Engineering Applications of Artificial Intelligence. 2021;97:104015. https://doi.org/10.1016/j.engappai.2020.104015

3. Abdollahzadeh B, Gharehchopogh FS. A multi-objective optimization algorithm for feature selection problems. Engineering with Computers. 2022;38 (Suppl 3):1845-63. https://doi.org/10.1007/s00366-021-01369-9

4. Andrei N. Nonlinear conjugate gradient methods for unconstrained optimization: Springer; 2020.

5. Hassan Ibrahim A, Kumam P, Abubakar AB, Abubakar J, Muhammad AB. Least-square-based three-term conjugate gradient projection method for ℓ 1-norm problems with application to compressed sensing. Mathematics. 2020;8(4):602. https://doi.org/10.3390/math8040602

6. Yuan G, Lu J, Wang Z. The PRP conjugate gradient algorithm with a modified WWP line search and its application in the image restoration problems. Applied Numerical Mathematics. 2020;152:1-11. https://doi.org/10.1016/j.apnum.2020.01.019

7. Hassan BA, Dahawi HO, Younus AS. A new kind of parameter conjugate gradient for unconstrained optimization. Indonesian Journal of Electrical Engineering and Computer Science. 2020;17(1):404. https://doi.org/10.11591/ijeecs.v17.i1.pp404-411

8. Khudhur HM, Fawze AAM. An improved conjugate gradient method for solving unconstrained optimisation and image restoration problems. International Journal of Mathematical Modelling and Numerical Optimisation. 2023;13(3):313-25. https://doi.org/10.1504/IJMMNO.2023.132286

9. Ji Z, Dudík M, Schapire RE, Telgarsky M, editors. Gradient descent follows the regularization path for general losses. Conference on Learning Theory; 2020: PMLR. https://doi.org/10.48550/arXiv.2006.11226

10. Fletcher R. Practical methods of optimization: John Wiley & Sons; 2000. https://doi.org/10.1093/comjnl/7.2.149

11. Dai Y-H, Yuan Y. A nonlinear conjugate gradient method with a strong global convergence property. SIAM Journal on optimization. 1999;10(1):177-82. https://doi.org/10.6028/jres.049.044

12. Hu Y, Storey C. Global convergence result for conjugate gradient methods. Journal of Optimization Theory and Applications. 1991;71(2):399-405. https://doi.org/10.1002/9781118723203

13. Fletcher R, Reeves CM. Function minimization by conjugate gradients. The computer journal. 1964;7(2):149-54. https://doi.org/10.1051/m2an/196903r100351

14. Hestenes MR, Stiefel E. Methods of conjugate gradients for solving linear systems. Journal of research of the National Bureau of Standards. 1952;49(6):409-36. https://doi.org/10.1137/S1052623497318992

15. Polak E, Ribiere G. Note sur la convergence de méthodes de directions conjuguées. Revue française d'informatique et de recherche opérationnelle Série rouge. 1969;3(16):35-43. https://doi.org/10.1007/BF00939927

16. Hu W, Wu J, Yuan G. Some modified Hestenes-Stiefel conjugate gradient algorithms with application in image restoration. Applied Numerical Mathematics. 2020;158:360-76. https://doi.org/10.1016/j.apnum.2020.08.009

17. Ibrahem SA, Halil IH, Abdullah SM. Two Algorithms for Solving Unconstrained Global Optimization by Auxiliary Function Method. Tikrit Journal of Pure Science. 2024;29(3):84-9. https://doi.org/10.25130/tjps.v29i3.1604

18. Abdullah ZM, Asker SA. A New Scaled Three-Term Conjugate Gradient Algorithms for Unconstrained Optimization. Tikrit Journal of Pure Science. 2023;28(4):103-10. https://doi.org/10.25130/tjps.v28i4.1534

19. Awwal AM, Wang L, Kumam P, Mohammad H, Watthayu W. A projection Hestenes–Stiefel method with spectral parameter for nonlinear monotone equations and signal processing. Mathematical and Computational Applications. 2020;25(2):27. https://doi.org/10.3390/mca25020027

20. Taqi AH, Shaker AN. Development Modified Conjugate Gradiente Algorithm. Tikrit Journal of Pure Science. 2015;20(5):185-92. https://doi.org/10.25130/tjps.v20i5.1254

21. Khudhur HM, Halil IH. Noise removal from images using the proposed three-term conjugate gradient algorithm. Компьютерные исследования и моделирование. 2024;16(4):841-53. https://doi.org/10.20537/2076-7633-2024-16-4-841-853

22. Gould NI, Orban D, Toint PL. CUTEr and SifDec: A constrained and unconstrained testing environment, revisited. ACM Transactions on Mathematical Software (TOMS). 2003;29(4):373-94. https://doi.org/10.1145/962437.962439

23. Andrei N. An unconstrained optimization test functions collection. Adv Model Optim. 2008;10(1):147-61. https://camo.ici.ro/journal/vol10/v10a10.pdf

24. Bongartz I, Conn AR, Gould N, Toint PL. CUTE: Constrained and unconstrained testing environment. ACM Transactions on Mathematical Software (TOMS). 1995;21(1):123-60. https://doi.org/10.1145/200979.201043

25. Hager WW, Zhang H. A new conjugate gradient method with guaranteed descent and an efficient line search. SIAM Journal on optimization. 2005;16(1):170-92. https://doi.org/10.1137/030601880