Mathematical Foundations of Computing
doi:10.3934/mfc.2023028

A NEW HYBRID CG METHOD AS CONVEX COMBINATION

Amina Hallal*, Mohammed Belloufi
and Badreddine Sellami

Department of Informatics and Mathematics
Mohamed Cherif Messaadia University, 41000 Souk Ahras, Algeria

(Communicated by Yunwen Lei)

Abstract. Conjugate gradient methods are among the most efficient methods for solving optimization models. In this paper, a newly proposed conjugate gradient method is proposed for solving optimization problems as a convex combination of the Harger-Zhan and Dai-Yaun nonlinear conjugate gradient methods, which is capable of producing the sufficient descent condition with global convergence properties under the strong Wolfe conditions. The numerical results demonstrate the efficiency of the proposed method with some benchmark problems.

1. **Introduction.** In this paper, we consider the unconstrained minimization problem of the following form:

$$\min \{f(x), x \in \mathbb{R}^n\}. \tag{1}$$

Let the function $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, we defined the gradient as $g_k = \nabla f(x_k)$.

The conjugate gradient method is used to solve the iterative procedure (1) proposed by

$$x_0 \epsilon \mathbb{R}^n, \quad x_{k+1} = x_k + \alpha_k d_k, \quad k \epsilon \mathbb{N}. \tag{2}$$

$\alpha_k > 0$ is the step size.

The d_k search direction is stated by

$$d_0 = -g_0, \; d_{k+1} = -g_{k+1} + \beta_k d_k. \tag{3}$$

$\beta_k \epsilon \mathbb{R}$ is a scalar parameter known as the conjugate gradient coefficient.

Many formulas have been proposed to calculate the β_k. The most well-known are Hestenes-Stiefel (HS) (Hestenes and Stiefel, 1952), Fletcher-Reeves (FR) (Fletcher and Reeves, 1964), Polak-Ribiere-Polyak (PRP) (Polak and Ribiere, 1969), Conjugate Descent (1987), Liu-Storey (LS) (Liu and Storey, 1991), Dai-Yaun (DY) (Dai and Yaun, 1999) and Hager-Zhan (HZ) (Hager and Zhan, 2005), see in ([15],[10],[19, 20],[9],[18],[8],[3],[11]). It is all given as follows, respectively:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}, \tag{4}$$

2020 *Mathematics Subject Classification.* Primary: 65k05, 90C26, 90C30.

Key words and phrases. Unconstrained optimization, hybrid conjugate gradient method, globale convergence, numerical results.

*Corresponding author: Amina Hallal.

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \tag{5}$$

$$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}, \tag{6}$$

$$\beta_k^{CD} = -\frac{\|g_{k+1}\|^2}{d_k^T g_k}, \tag{7}$$

$$\beta_k^{LS} = -\frac{g_{k+1}^T y_k}{d_k^T g_k}, \tag{8}$$

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k}, \tag{9}$$

$$\beta_k^{HZ} = \frac{1}{d_k^T y_k}\left(y_k^T - 2d_k^T \frac{\|y_k\|^2}{d_k^T y_k}\right)^T g_{k+1}, \tag{10}$$

where $y_k = g_{k+1} - g_k$, $\|.\|$ denotes the Euclidean norm.

The conjugate gradient method can be classified into: classical, hybrid, scaled, modified, parametrized, and accelerated according to formula. The methods mentioned above are called "classical CG" due to their simple approaches. Recently, some hybrid conjugate gradients as the hybrid conjugate gradient methods, which are a combination of several gradient conjugate algorithms, are more efficient than the classical conjugate gradient methods. The first hybrid conjugate algorithm was given by Touati-Ahmed and Storey [22], which is computes β_k as

$$\beta_k^{TS} = \begin{cases} \beta_k^{PRP} & \text{if } 0 \le \beta_k^{PRP} \le \beta_k^{FR}, \\ \beta_k^{FR} & \text{else.} \end{cases} \tag{11}$$

Djordjevic proposed a hybrid conjugate gradient algorithm as convex combinations [6], in which

$$\beta_k^{LSCDCC} = (1-\theta_k)\,\beta_k^{LS} + \theta_k \beta_k^{CD}, \tag{12}$$

and proposed another hybrid conjugate gradient algorithm as convex combinations [5], in which

$$\beta_k^{HLSFR} = (1-\theta_k)\,\beta_k^{LS} + \theta_k \beta_k^{FR}. \tag{13}$$

In this paper, we combine the DY and HZ conjugate gradient algorithms in this research to create a new hybrid nonlinear CG method. The algorithm of the suggested method was detailed in section 2. In Section 3, we show that, utilizing the inexact line search, the proposed approach satisfies the sufficient descent condition and is globally convergent. In Section 4, we provide some numerical tests compared with classical formulas of the DY and HZ. Finally, we will give a conclusion.

2. **A new CG algorithm.** In this paper, we introduce a new conjugate gradient scalar know as β_k^{hHZDY} (h it mean hybrid) is defined by the next equation

$$\beta_k^{hHZDY} = (1-\theta_k)\,\beta_k^{HZ} + \theta_k \beta_k^{DY},\ 0 \le \theta_k \le 1. \tag{14}$$

The search direction d_k is obtained by

$$d_0 = -g_0,\ d_{k+1} = -g_{k+1} + \beta_k^{hHZDY} d_k. \tag{15}$$

The step size α_k is determinated according to the following strong Wolfe conditions

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k \nabla f(x_k)^T d_k. \tag{16}$$

$$\sigma \nabla f(x_k)^T d_k \leq \nabla f(x+\alpha_k d_k)^T d_k \leq -\sigma \nabla f(x_k)^T d_k. \tag{17}$$

Here θ_k is determined in such a way that the search direction satisfies the conjugacy condition

$$d_{k+1}^{hHZDY} y_k = 0. \tag{18}$$

Substituting (15) in (18) we get

$$-g_{k+1} + (1-\theta_k)\left(\frac{1}{d_k^T y_k}\left(y_k^T - 2d_k^T \frac{\|y_k\|^2}{d_k^T y_k}\right) g_{k+1}\right) d_k + \theta_k \frac{\|g_{k+1}\|^2}{d_k^T y_k} d_k = 0.$$

Multiplying both sides of the above equation by y_k, we obtain

$$-g_{k+1}^T y_k + (1-\theta_k)\left(y_k^T g_{k+1} - 2d_k^T g_{k+1} \frac{\|y_k\|^2}{d_k^T y_k}\right) + \theta_k \|g_{k+1}\|^2 = 0. \tag{19}$$

Solving (19) implies that

$$\theta_k = \frac{2d_k^T g_{k+1} \frac{\|y_k\|^2}{d_k^T y_k}}{\left(\|g_{k+1}\|^2 - y_k^T g_{k+1} + 2d_k^T g_{k+1} \frac{\|y_k\|^2}{d_k^T y_k}\right)}. \tag{20}$$

We can conclude that

$$\theta_k = \begin{cases} 0 & \text{if } \theta_k \leq 0, \\ 1 & \text{if } \theta_k \geq 1, \\ \dfrac{2d_k^T g_{k+1} \frac{\|y_k\|^2}{d_k^T y_k}}{\left(\|g_{k+1}\|^2 - y_k^T g_{k+1} + 2d_k^T g_{k+1} \frac{\|y_k\|^2}{d_k^T y_k}\right)} & \text{if } 0 < \theta_k < 1, \end{cases} \tag{21}$$

and

$$\beta_k = \begin{cases} \beta_k^{HZ} & \text{if } \theta_k = 0, \\ \beta_k^{DY} & \text{if } \theta_k = 1, \\ (1-\theta_k)\beta_k^{HZ} + \theta_k \beta_k^{DY} & \text{if } 0 < \theta_k < 1. \end{cases}$$

Next, we give the algorithm of (14) as follows

Algorithm (hHZDY method)

Step 1. Starting Given $x_0 \in \mathbb{R}^n$. Set $d_0 = -g_0$, $\alpha_0 = \frac{1}{\|g_0\|}$.
Step 2. Calculate $x_{k+1} = x_k + \alpha_k d_k$, compute $y_k = g_{k+1} - g_k$.
Step 3. Stopping criteria
Compute $\|g_k\|$, if $\|g_k\| < 10^{-4}$ then Stop else go to the next step.
Step 4. Compute α_k by (16) and (17).
Step 5. If $\left(\|g_{k+1}\|^2 - y_k^T g_{k+1} + 2d_k^T g_{k+1} \frac{\|y_k\|^2}{d_k^T y_k}\right) = 0$ then put $\theta_k = 0$ or compute θ_k as in (21).
Step 6. Compute β_k by (14).
Step 7. Compute d_k by equation (15) .
Step 8. Put $k = k+1$ and go to step 2.

3. **Convergent analysis.**

3.1. **Sufficient descent condition.** To be converged, an algorithm must have sufficient descent conditions as well as global convergent properties. The sufficient descent condition is satisfied by the hHZDY algorithm, as shown by the following theorem.

Theorem 3.1. *Consider a CG method (1) and (3), β_k given by (14) and α_k be generated with (16) and (17) then*

$$g_k^T d_k \leq -c\|g_k\|^2 \ , \ \ c > 0, \ \textit{for all } k. \tag{22}$$

Proof. We use mathematical induction.

It obvious that (22) hold for $k = 0$.

Assume that $g_k^T d_k \leq -c \|g_k\|^2$.

Now for $k = k + 1$.

From the idea in [11]

$$\begin{aligned} d_{k+1}^{hHZDY} &= -g_{k+1} + \beta_k^{hHZDY} d_k \\ &= -(1-\theta_k) g_{k+1} - \theta_k g_{k+1} + (1-\theta_k) \beta_k^{HZ} d_k + \theta_k \beta_k^{DY} d_k. \end{aligned}$$

Implies

$$d_{k+1}^{hHZDY} = (1-\theta_k) d_{k+1}^{HZ} + \theta_k d_{k+1}^{DY}. \tag{23}$$

With multiplying (23) by g_{k+1}^T we get

$$g_{k+1}^T d_{k+1}^{hHZDY} = (1-\theta_k) g_{k+1}^T d_{k+1}^{HZ} + \theta_k g_{k+1}^T d_{k+1}^{DY}. \tag{24}$$

We have 3 cases:

If $\theta_k = 0$ then

$$g_{k+1}^T d_{k+1}^{hHZDY} = g_{k+1}^T d_{k+1}^{HZ}. \tag{25}$$

W.W. Hager and H. Zhang proved in [18] that d_{k+1}^{HZ} satisfies the suffcient descent condition i.e, $\exists m_1 = \frac{7}{8} > 0$

$$g_{k+1}^T d_{k+1}^{HZ} \leq -m_1 \|g_{k+1}\|^2. \tag{26}$$

If $\theta_k = 1$ then

$$g_{k+1}^T d_{k+1}^{hHZDY} = g_{k+1}^T d_{k+1}^{DY}. \tag{27}$$

Clearly from (17) that

$$\begin{aligned} (-\sigma - 1) d_k^T g_k \geq d_k^T y_k = d_k^T (g_{k+1} - g_k) \\ \geq (\sigma - 1) d_k^T g_k. \end{aligned} \tag{28}$$

With substiting we have

$$\begin{aligned} g_{k+1}^T d_{k+1}^{DY} &= -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{d_k^T y_k} (g_{k+1}^T d_k) \\ &\leq -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{(\sigma - 1) d_k^T g_k} \left| g_{k+1}^T d_k \right|. \end{aligned}$$

Now by using (17) again and (28) it results that

$$\begin{aligned} g_{k+1}^T d_{k+1}^{DY} &\leq -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{(\sigma - 1) d_k^T g_k} \left(-\sigma g_k^T d_k \right) \\ &\leq -\left(1 - \frac{\sigma}{1-\sigma} \right) \|g_{k+1}\|^2. \end{aligned}$$

Hence

$$g_{k+1}^T d_{k+1}^{DY} \leq -m_2 \|g_{k+1}\|^2. \tag{29}$$

Where $\sigma < \frac{1}{2}$.

Finaly for $0 < \theta_k < 1$

From (24), (25) and (29)

$$\begin{aligned} g_{k+1}^T d_{k+1}^{hHZDY} &= (1-\theta_k) g_{k+1}^T d_{k+1}^{HZ} + \theta_k g_{k+1}^T d_{k+1}^{DY} \\ &\leq -(m_1 (1-\theta_k) + m_2 \theta_k) \|g_{k+1}\|^2. \end{aligned}$$

[5] There exists $\mu, \lambda \in \mathbb{R}$ where $0 < \mu < \theta_k < \lambda < 1$ that give

$$g_{k+1}^T d_{k+1}^{hHZDY} \leq -\left(m_1\left(1-\lambda\right)+m_2\mu\right)\left\|g_{k+1}\right\|^2,$$

where $m = \left(m_1\left(1-\lambda\right)+m_2\mu\right) > 0$. □

3.2. **Global convergence properties.** The following basic assumptions about the objective function are very important in the convergence analysis of the CG method.

Assumption 1. *The level set $\mathcal{S} = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$ is bounded where x_0 is the starting point.*

Assumption 2. *In some neighborhood $\mathcal{N}$ of $\mathcal{S}$ the function f is continuously differentiable and its gradient is Lipschitz continuous i.e. $\exists L > 0$*

$$\left\|\nabla f(x_{k+1}) - \nabla f(x_k)\right\| \leq L\left\|x_{k+1} - x_k\right\|, \tag{30}$$

which can results

$$\left\|\nabla f(x)\right\| \leq r, \text{ for all} x \in \mathcal{S}. \tag{31}$$

[2].

Lemma 3.2. [17] *Assume that d_k is descent, and assumption 2 satisfies, α_k is determinated by the strong Wolfe inexact line search, then*

$$\alpha_k \geq c\frac{(1-\sigma)\parallel g_k \parallel^2}{L\parallel d_k\parallel^2}. \tag{32}$$

Proof. With appliying (17) and the second assumption, we get

$$\begin{aligned}(\sigma-1)g_k^T d_k &\leq (g_{k+1}-g_k)^T d_k \\ &\leq L\alpha_k\parallel d_k\parallel^2,\end{aligned}$$

where d_k is a descent direction $\sigma < 1$, then we have the next

$$\alpha_k \geq c\frac{(1-\sigma)\parallel g_k \parallel^2}{L\parallel d_k\parallel^2}.$$

□

Lemma 3.3. [15] *Assume that Assumptions 1 and 2 hold. Consider the algorithm of the new method with d_k satisfying the suffiicient descent condition and α_k being obtained by the strong Wolfe line search then*

If

$$\sum_{k\geq 1}\frac{\left\|g_k\right\|^4}{\left\|d_k\right\|^2} = \infty. \tag{33}$$

Then

$$\lim_{k\to\infty}\inf\left\|g_k\right\| = 0. \tag{34}$$

Now we give our convergence theorem

Theorem 3.4. *Assume that the Assumptions 1and 2 satisfies. Consider the Algorithm with d_k descent and α_k is determinated by (16), (17) where $0 < \sigma < \frac{1}{2}$ then*

$$\lim_{k\to\infty}\inf\left\|g_k\right\| = 0. \tag{35}$$

Proof. The proof would be by contradiction.

Suppose that $g_k \neq 0 \; \forall \; k$, then there exists $\exists \eta > 0$ such that

$$\|g_k\| \geq \eta \quad , \; k \geq 0, \tag{36}$$

where D the diameter of $\mathcal{S}$, and $x_{k+1} - x_k = s_k$.

Hence

$$\|y_k\| = \|g_{k+1} - g_k\| \leq L \|s_k\| \leq LD.$$

From(15) we obtain

$$\|d_{k+1}\| \leq \| g_{k+1} \| + \left|\beta_k^{\mathrm{hHZDY}}\right| \| d_k \| \, .$$

After some algebra

$$\begin{aligned}
\left|\beta_k^{hHZDY}\right| &\leq \left|\beta_k^{HZ}\right| + \left|\beta_k^{DY}\right| \\
&\leq \frac{1}{-(1-\sigma) d_k^T g_k}\left(\left|y_k^T g_{k+1}\right| + 2\left|\frac{d_k^T g_{k+1}}{d_k^T y_k}\right| \|y_k\|^2 + \|g_{k+1}\|^2\right) \\
&\leq \frac{1}{(1-\sigma) c \|g_k\|^2}\left(\|y_k\| \|g_{k+1}\| + 2\frac{\sigma}{(1-\sigma)} \|y_k\|^2 + \|g_{k+1}\|^2\right) \\
&\leq \frac{1}{(1-\sigma) c\eta^2}\left(\left(rLD + 2\frac{\sigma}{(1-\sigma)} (LD)^2\right) + r^2\right) \leq \vartheta.
\end{aligned}$$

From Lemma 3.2, we conclude that $\exists \nu > 0 : \alpha_k > \nu$.

Finally, we results

$$\begin{aligned}
\|d_{k+1}\| &\leq \|g_{k+1}\| + \left|\beta_k^{hHZDY}\right| \|d_k\| \leq \|g_{k+1}\| + \left|\beta_k^{hHZDY}\right| \frac{\|s_k\|}{\alpha_k}, \\
&\leq r + \vartheta \frac{D}{\nu},
\end{aligned}$$

and we can conclude that

$$\frac{1}{\|d_{k+1}\|^2} \geq \frac{1}{\left(r + \vartheta \frac{D}{\nu}\right)^2},$$

and

$$\frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \geq \frac{\eta^4}{\left(r + \vartheta \frac{D}{\nu}\right)^2}.$$

Implies

$$\sum_{k\geq 0} \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \geq \frac{\eta^4}{\left(r + \vartheta \frac{LD}{\nu}\right)^2} \sum_{k\geq 0} 1 = \infty.$$

It implies

$$\sum_{k\geq 0} \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} = \infty,$$

with Lemma 3.3 applied, we have a contradiction. □

4. **Numerical result.** In this section, we present numerical experiment results obtained by testing our new algorithm with HZ and DY CG algorithms on a set of unconstrained optimization test problems picked from [1], [16]. We use different starting points and different dimensions. The code for the proposed method was written using Matlab R2013a and run on a personal computer with a 1.60 GHz CPU processor and 2GB of RAM memory. The used line search conditions are the strong Wolfe conditions with $\alpha = 1$, $\sigma = 0.01$, $\delta = 0.0001$ and we set $\varepsilon = 10^{-4}$. This is done based on the number of iterations and CPU time. The iterations fail if the number of iterations exceeds 2000. The comparison of the performance profile curve results between the methods was presented using the profiles of Dolan and Moré [8], CPU denotes the total CPU time (seconds).

It is seen from Figures 1,2 that the "hHZDY" method is better than the "HZ" and "DY" methods.

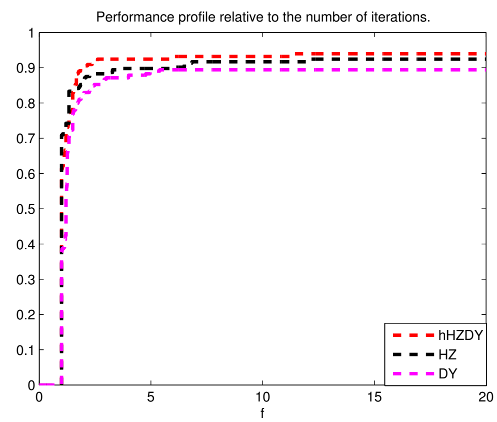

FIGURE 1. Performance profile relative to the CPU time.

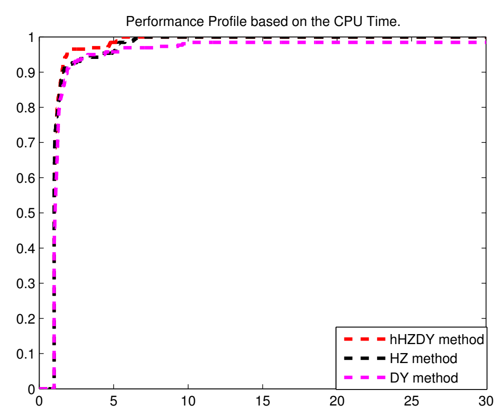

FIGURE 2. Performance profile relative to the number of iterations.

5. **Conclusion.** The hybrid gradient conjugate method is an iterative method, and this method can be used to search for solutions for optimization problems without constraints in large-scale cases. This paper presents a new hybrid CG method, which is the convex combination of HZ and DY, named the hHZDY method. We demonstrated that the proposed method guarantees the sufficient descent and global convergence conditions under the strong Wolfe inexact line search. The numerical results are accurate.

Acknowledgments. The authors would like to thank the editor and the referees for their useful comments, which greatly improved this paper.

REFERENCES

[1] N. Andrei, An unconstrained optimization test functions collection, *Adv. Model. Optim.*, **5** (2008), 147-161.
[2] N. Andrei, Another hybrid conjugate gradient algorithm for unconstrained optimization, *Numerical Algorithms*, **47** (2008), 143-156.
[3] Y. H. Dai and Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, *SIAM J. Optim.*, **10** (1999), 177-182.
[4] S. Delladji, M. Belloufi and B. Sellami, Behavior of the combination of PRP and HZ for unconstrained optimization, *Numerical Algebra Control Optim.*, **11** (2021), 377-389.
[5] S. S. Djordjevic, New hybrid conjugate gradientMethod as a convex combination of FR and PRP methods, *Comput. J.*, **30** (2016), 3083-3100.
[6] S. S. Djordjevic, New hybrid conjugate gradient method as a convex combination of LS and CD methods, *Filomat*, **31** (2017), 1813-1825.
[7] S. S. Djordjevic, New hybrid conjugate gardient method as a convex combination od LLS and FR methods, *Acta Mathematica Scientia*, **39** (2019), 214-228.
[8] E. D. Dolan and J. J. Moré, Benchmarking optimization software with performance profiles, *Mathematical Programming*, **91** (2002), 201-213.
[9] R. Fletcher, A survey of nonlinear conjugate gradient methods, *Pacific journal of Optimization*, **2** (2006), 35-58.
[10] R. Fletcher and C. M. Reeves, Function minimization by conjugate gradients, *Comput. J.*, **7** (1964), 149-154.
[11] W. W. Hager and H. Zhan, 2nd ed A Wiley-Interscience Publication, *John Wiley, Sons, Inc, NY, USA*, (1987).
[12] W. W. Hager and H. Zhan, A survey of nonlinear conjugate gradient methods, *Pacific Journal of Optimization*, **2** (2006), 35-58.
[13] A. Hallal, M. Belloufi and B. Sellami, An efficient new hybrid CG-method as convex combination of DY and CD and HS algorithms, *RAIRO Operations Research*, **56** (2022), 4047-4056.
[14] A. Hallal, M. Belloufi and B. Sellami, Using a new hybrid conjugate gradient method with descent property, *Journal of Information & Optimization Sciences*, (2023), 2169-0103.
[15] M. R. Hestenes and E. L. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Research Nat Bur Standards*, **49** (1952), 409-436.
[16] M. Jamil and X. S. Yang, A literature survey of benchmark functions for global optimization problems, *Journal of Mathematical Modelling and Numerical Optimisation*, **4** (2013), 150-194.
[17] J. K. Liu and S. J. Li, New hybrid conjugate gradient method for unconstrained optimization, *Applied Mathematics and Computation*, **245** (2014), 36-43.
[18] Y. Liu and C. Storey, Efficient generalized conjugate gradient algorithms. I. Theory, *J. Optim. Theory Appl.*, **69** (1991), 129-137.
[19] E. Polak, G. Ribiere, Note sur la convergence de mé thodes de directions conjugués, *Revue Francaise d'Informatique et de Recherche Opérationnelle*, **13** (1969), 35-43.
[20] B. T. Polyak, The conjugate gradient method in extreme problems, *USSR Comp Math. Math. Phys.*, **9** (1969), 94-112.
[21] D. Touati-Ahmed and C. Storev, Efficient hybrid conjugate gradient techniques, *J. Optim. Theory Appl.*, **64** (1990), 379-397.

[22] P. Wolfe, Convergence conditions for ascent methods. II. Some corrections, *SIAM Review*, **13** (1971), 185-188.

Received October 2022; revised April 2023; early access July 2023.