VESTNIK UDMURTSKOGO UNIVERSITETA. MATEMATIKA. MEKHANIKA. KOMP'YUTERNYE NAUKI
MATHEMATICS 2023. Vol. 33. Issue 2. Pp. 348–364.

MSC2020: 90C30, 90C53

 I. Hafaidia, H. Guebbai, M. Al-Baali, M. Ghiat

A NEW HYBRID CONJUGATE GRADIENT ALGORITHM FOR UNCONSTRAINED OPTIMIZATION

It is well known that conjugate gradient methods are useful for solving large-scale unconstrained nonlinear optimization problems. In this paper, we consider combining the best features of two conjugate gradient methods. In particular, we give a new conjugate gradient method, based on the hybridization of the useful DY (Dai–Yuan), and HZ (Hager–Zhang) methods. The hybrid parameters are chosen such that the proposed method satisfies the conjugacy and sufficient descent conditions. It is shown that the new method maintains the global convergence property of the above two methods. The numerical results are described for a set of standard test problems. It is shown that the performance of the proposed method is better than that of the DY and HZ methods in most cases.

Keywords: unconstrained optimization, conjugate gradient methods, conjugacy conditions and sufficient descent conditions.

DOI: 10.35634/vm230211

Introduction

The class of nonlinear conjugate gradient methods is one of the useful and practical techniques that require building and developing algorithms to solve the unconstrained optimization problem [1–7]

$$\min_{x\in\mathbb{R}^n} f(x), \tag{0.1}$$

where f is a smooth function. The class generates a sequence of points $\{x_k\}$ iteratively by

$$x_{k+1} = x_k + \alpha_k d_k, \tag{0.2}$$

where α_k is a steplength and d_k is a search direction. Here α_k is chosen such that the following strong Wolfe conditions are satisfied:

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k g_k^\top d_k, \tag{0.3}$$

$$\sigma g_k^\top d_k \le g(x_k + \alpha_k d_k)^\top d_k \le -\sigma g_k^\top d_k, \tag{0.4}$$

where $0 < \delta < \frac{1}{2}$ and $\delta < \sigma < 1$. The search directions d_k are defined by

$$\begin{cases} d_1 = -g_1, & k = 1, \\ d_{k+1} = -g_{k+1} + \beta_k d_k, & k \ge 1, \end{cases}$$

where $g_k = g(x_k) = \nabla f(x_k)$ and β_k is a conjugate gradient parameter.

Different conjugate gradient algorithms correspond to different choices of the parameter β_k that have been proposed. In particular, the choices of

$$\beta_k^{HS} = \frac{g_{k+1}^\top y_k}{d_k^\top y_k}, \qquad \beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \qquad \beta_k^{PRP} = \frac{g_{k+1}^\top y_k}{\|g_k\|^2},$$

$$\beta_k^{LS} = \frac{g_{k+1}^\top y_k}{-d_k^\top g_k}, \quad \beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^\top y_k}, \qquad \beta_k^{HZ} = \frac{g_{k+1}^\top y_k}{d_k^\top y_k} - 2\|y_k\|^2 \frac{d_k^\top g_{k+1}}{(d_k^\top y_k)^2},$$

where $y_k = g_{k+1} - g_k$, are proposed in [8–14].

In order to guarantee the global convergence property of the conjugate gradient methods, the sufficient descent property

$$d_{k+1}^{\top} g_{k+1} \leq -c\|g_{k+1}\|^2, \quad c > 0, \tag{0.5}$$

is enforced with certain conditions on β_k (see for example [6, 18, 19]).

If the line search is exact, $d_k^{\top} g_{k+1} = 0$, then the descent property (0.5) holds with equality and $c = 1$, DY is reduced to FR, and HS and HZ are reduced to PRP. If, in addition, the objective function is quadratic, all conjugate gradient parameters are reduced to FR. In this case, the conjugacy condition $y_k^{\top} d_{k+1} = 0$ holds so that obtaining the solution of problem (0.1) requires at most n iterations (see for example Fletcher [20]). For a general function, the convergence result depends on the choice of β_k and the line search technique. The first practical global convergence result is obtained for the FR method by [19], based on showing that the sufficient descent property (0.5) holds, for a certain value of c, if the above strong Wolfe conditions are employed with $\sigma < 1/2$. This result is extended by [18] to the choice of Powell $\beta_k^{PRP+} = \max(\beta_k^{PRP}, 0)$.

However, the convergence result for the DY method requires the Wolfe conditions, given by (0.3) and the left inequality of (0.4), while for the HZ method convergence is obtained for any line search technique. In practice, the DY and HZ methods seem to work better than many other conjugate gradient methods (for further details, see e. g. [14]). Therefore, we consider the possibility of combining the best features of both methods by defining the search direction as a linear combination of the form

$$d_{k+1} = \theta_k d_{k+1}^{DY} + \vartheta_k d_{k+1}^{HZ}, \tag{0.6}$$

for some choices of the parameters θ_k and ϑ_k. They are derived in Section 1 such that condition (0.5) holds with equality as well as the above conjugacy condition. We focus here on the above combination, which can be extended to other conjugate gradient methods. Section 2 shows that the proposed method converges globally if the Wolfe conditions hold. In Section 3, we describe some numerical results to show that the performance of the new method is better than that of both the DY and HZ methods. Section 4 concludes the paper.

§ 1. A hybrid conjugate gradient algorithm

In this section, we suggest a hybrid conjugate gradient method that defines the search direction by the linear combination (0.6), which can be written as follows

$$d_{k+1} = -(\theta_k + \vartheta_k) g_{k+1} + (\theta_k \beta_k^{DY} + \vartheta_k \beta_k^{HZ}) d_k. \tag{1.1}$$

Because the exact line search technique implies the equality

$$d_{k+1}^{\top} g_{k+1} = -(\theta_k + \vartheta_k)\|g_{k+1}\|^2,$$

we assume

$$\theta_k + \vartheta_k > 0$$

to guarantee the descent property. In addition, these parameters will be chosen so that the sufficient descent and conjugacy conditions are satisfied, given respectively by

$$d_{k+1}^{\top} g_{k+1} = -c\|g_{k+1}\|^2, \tag{1.2}$$

where $c > 0$ is a constant and

$$y_k^{\top} d_{k+1} = 0. \tag{1.3}$$

Premultiplying (1.1) by $g_{k+1}^\top$ and $y_k^\top$ and using (1.2) and (1.3), we respectively obtain

$$(\|g_{k+1}\|^2 - \beta_k^{DY} d_k^\top g_{k+1})\theta_k + (\|g_{k+1}\|^2 - \beta_k^{HZ} d_k^\top g_{k+1})\vartheta_k = c\|g_{k+1}\|^2 \tag{1.4}$$

and

$$-\vartheta_k g_{k+1}^\top y_k - \theta_k g_{k+1}^\top y_k + \vartheta_k \beta_k^{HZ}(d_k^\top y_k) + \theta_k \beta_k^{DY}(d_k^\top y_k) = 0. \tag{1.5}$$

To consider all possible solutions of this system of two linear equations, we first rearrange equation (1.5), using the definitions of β_k^{DY} and β_k^{HZ}, as follows

$$g_k^\top g_{k+1}\theta_k - 2\frac{||y_k||^2}{d_k^\top y_k} d_k^\top g_{k+1}\vartheta_k = 0$$

or, equivalently,

$$d_k^\top y_k g_k^\top g_{k+1}\theta_k - 2||y_k||^2 d_k^\top g_{k+1}\vartheta_k = 0. \tag{1.6}$$

We consider solving the system (1.4) and (1.6) based on the following four possible cases:

(i) If both values of $d_k^\top g_{k+1} = 0$ and $g_k^\top g_{k+1} = 0$, which hold if the line search is exact and the function is quadratic (e. g., Fletcher [20]), then equation (1.6) holds for any values of θ_k and ϑ_k which satisfy (1.4) for $\theta_k + \vartheta_k = c$ (e. g., let $\theta_k = 0$ and $\vartheta_k = c = 1$).

(ii) If $d_k^\top g_{k+1} = 0$ and $g_k^\top g_{k+1} \neq 0$, (1.6) implies $\theta_k = 0$ and hence (1.4) yields $\vartheta_k = c$ (let $c = 1$, as for (i)).

(iii) If $d_k^\top g_{k+1} \neq 0$ and $g_k^\top g_{k+1} = 0$, (1.6) implies $\vartheta_k = 0$ and hence by (1.4), it follows that $\theta_k = \tilde{\theta}_k$, where

$$\tilde{\theta}_k = -c\frac{d_k^T y_k}{d_k^T g_k}$$

which is positive if the Wolfe line search conditions hold.

(iv) For the remaining case, $d_k^\top g_{k+1} \neq 0$ and $g_k^\top g_{k+1} \neq 0$, we solve the system to obtain $\theta_k = \hat{\theta}_k$ and $\vartheta_k = \hat{\vartheta}_k$, where

$$\hat{\theta}_k = \frac{c\|g_{k+1}\|^2(y_k^\top g_{k+1} - \beta_k^{HZ} d_k^\top y_k)}{\Delta_k(\beta_k^{DY} - \beta_k^{HZ})}, \tag{1.7}$$

$$\hat{\vartheta}_k = \frac{-c\|g_{k+1}\|^2(y_k^\top g_{k+1} - \beta_k^{DY} d_k^\top y_k)}{\Delta_k(\beta_k^{DY} - \beta_k^{HZ})}, \tag{1.8}$$

where

$$\Delta_k = \|g_{k+1}\|^2(d_k^\top y_k) - (d_k^\top g_{k+1})(y_k^\top g_{k+1}). \tag{1.9}$$

The above solution exists if

$$\Delta_k(\beta_k^{DY} - \beta_k^{HZ}) \neq 0 \tag{1.10}$$

which will be guaranteed below if a certain condition holds. However, if (1.10) does not hold, the above system does not have a solution. In this case, we enforce only the sufficient descent condition (1.4) by choosing, as in case (iii), that $\vartheta_k = 0$ and $\theta_k = \tilde{\theta}_k$. (In fact, if $\beta_k^{DY} = \beta_k^{HZ}$ then $d_k^\top y_k g_k^\top g_{k+1} = -2||y_k||^2 d_k^\top g_{k+1}$ and hence (1.4) holds for $\theta_k + \vartheta_k = \tilde{\theta}_k$.)

To choose one formula for c for all cases, we let $c = \max\left(\frac{1}{1+|d_k^\top g_{k+1}|}, \hat{c}\right)$, where $0 < \hat{c} \leq 1$ (e. g., $\hat{c} = 0.9$), which is reduced to 1, as required for the cases (i) and (ii), when the exact line search equation is satisfied.

We now summarize the above analysis by the following result.

Lemma 1. *Letting*

$$(\theta_k, \vartheta_k) = \begin{cases} (\hat{\theta}_k, \hat{\vartheta}_k), & \text{if } (d_k^\top g_{k+1})(g_k^\top g_{k+1})(\beta_k^{DY} - \beta_k^{HZ})\Delta_k \neq 0, \\ (0, c), & \text{if } d_k^\top g_{k+1} = 0, \\ (\tilde{\theta}_k, 0), & \text{otherwise}, \end{cases} \tag{1.11}$$

and assume that the Wolfe line search conditions are employed. Then the sufficient descent condition (1.2) holds for some $c > 0$. Moreover, the conjugacy condition (1.3) holds for some $c > 0$, except when $\Delta_k(\beta_k^{DY} - \beta_k^{HZ}) = 0$.

For convenience, we let

$$\eta_k = \theta_k + \vartheta_k, \qquad \beta_k = \theta_k \beta_k^{DY} + \vartheta_k \beta_k^{HZ}, \tag{1.12}$$

so that the search direction (1.1) can be written as follows

$$d_{k+1} = -\eta_k g_{k+1} + \beta_k d_k. \tag{1.13}$$

From (1.12) and (1.11), we have

$$(\eta_k, \beta_k) = \begin{cases} (\hat{\eta}_k, \hat{\beta}_k), & \text{if } (d_k^\top g_{k+1})(g_k^\top g_{k+1})\Delta_k \neq 0, \\ (c, c\beta_k^{PRP}), & \text{if } d_k^\top g_{k+1} = 0, \\ (\tilde{\theta}_k, \beta_k^{FR}), & \text{otherwise}, \end{cases} \tag{1.14}$$

obviously, using (1.7) and (1.8), where

$$\begin{aligned} \hat{\eta}_k &= \hat{\theta} + \hat{\vartheta} \\ &= \frac{c\|g_{k+1}\|^2(y_k^\top g_{k+1} - \beta_k^{HZ} d_k^\top y_k) - c\|g_{k+1}\|^2(y_k^\top g_{k+1} - \beta_k^{DY} d_k^\top y_k)}{\Delta_k(\beta_k^{DY} - \beta_k^{HZ})} \\ &= \frac{c\|g_{k+1}\|^2(d_k^\top y_k)}{\Delta_k}, \end{aligned} \tag{1.15}$$

$$\begin{aligned} \hat{\beta}_k &= \hat{\theta}\beta_k^{DY} + \hat{\vartheta}\beta_k^{HZ} \\ &= \frac{c\|g_{k+1}\|^2\left(y_k^\top g_{k+1} - \beta_k^{HZ} d_k^\top y_k\right)\beta_k^{DY} - c\|g_{k+1}\|^2\left(y_k^\top g_{k+1} - \beta_k^{DY} d_k^\top y_k\right)\beta_k^{HZ}}{\Delta_k(\beta_k^{DY} - \beta_k^{HZ})} \\ &= \frac{c\|g_{k+1}\|^2(y_k^\top g_{k+1})}{\Delta_k}. \end{aligned} \tag{1.16}$$

We note that $\beta_k^{FR} = \tilde{\theta}_k \beta_k^{DY}$ is used in the third case of (1.14). We also note that the condition $\beta_k^{DY} \neq \beta_k^{HZ}$, which appears in the first case of (1.11), is not required in the first case of (1.14) due to resolving the system (1.4) and (1.6) with respect to η_k and θ_k as follows. On substituting $\vartheta_k = \eta_k - \theta_k$ in (1.4) and (1.5), we obtain the following equivalent system:

$$\begin{aligned} -(\|g_{k+1}\|^2 - \beta_k^{HZ} d_k^\top g_{k+1})\eta_k + (\beta_k^{DY} - \beta_k^{HZ}) d_k^\top g_{k+1}\theta_k &= -c\|g_{k+1}\|^2, \\ (y_k^\top g_{k+1} - \beta_k^{HZ} d_k^\top y_k)\eta_k - (\beta_k^{DY} - \beta_k^{HZ}) d_k^\top y_k \theta_k &= 0. \end{aligned} \tag{1.17}$$

Eliminating θ_k, we simply obtain

$$(-\|g_{k+1}\|^2 d_k^\top y_k + y_k^\top g_{k+1} d_k^\top g_{k+1})\eta_k = -c\|g_{k+1}\|^2 d_k^\top y_k,$$

which implies (1.15), assuming $\Delta_k \neq 0$, whether $\beta_k^{DY} = \beta_k^{HZ}$ holds or not. Similarly, using (1.12) and (1.17), it follows that

$$\beta_k = \beta_k^{HZ}\eta_k + (\beta_k^{DY} - \beta_k^{HZ})\theta_k = \frac{y_k^\top g_{k+1}}{d_k^\top y_k}\eta_k, \tag{1.18}$$

which, by (1.15), yields (1.16).

We now state the following lemma which shows the possibility of rewriting expressions (1.11) and (1.14) without switching among three cases, noting that the case $g_k^\top g_{k+1} = 0$ belongs to the "otherwise".

Lemma 2. *If either condition $d_k^\top g_{k+1} = 0$ or $g_k^\top g_{k+1} = 0$ holds, then the first choice in (1.11) is reduced to the second and third choices, respectively. Hence, similarly for (1.14).*

P r o o f. Using the definitions of β_k^{HZ} and β_k^{DY}, expressions (1.7), (1.8) and (1.9) can be rearranged respectively as follows:

$$\begin{aligned} \hat\theta_k &= \frac{2c\|g_{k+1}\|^2\|y_k\|^2 d_k^\top g_{k+1}}{\Delta_k(\beta_k^{DY} - \beta_k^{HZ}) d_k^\top y_k}, \\ \hat\vartheta_k &= \frac{c\|g_{k+1}\|^2 g_k^\top g_{k+1}}{\Delta_k(\beta_k^{DY} - \beta_k^{HZ})}, \\ \Delta_k &= (d_k^\top g_{k+1})(g_k^\top g_{k+1}) - d_k^\top g_k\|g_{k+1}\|^2. \end{aligned} \tag{1.19}$$

In addition,

$$\beta_k^{DY} - \beta_k^{HZ} = \frac{g_k^\top g_{k+1}}{d_k^\top y_k} + 2\|y_k\|^2\frac{d_k^\top g_{k+1}}{(d_k^\top y_k)^2}.$$

Hence, the condition $d_k^\top g_{k+1} = 0$ reduces the pair $(\hat\theta_k, \hat\vartheta_k)$ to $(0, c)$, while the equation $g_k^\top g_{k+1} = 0$ reduces $(\hat\theta_k, \hat\vartheta_k)$ to $(\tilde\theta_k, 0)$.

Similarly, for choice (1.14), using the above rearrangements, we observe from (1.15) and (1.16) that the conditions $d_k^\top g_{k+1} = 0$ and $g_k^\top g_{k+1} = 0$ reduce the pair $(\hat\eta_k, \hat\beta_k)$ to (c, β_k^{PRP}) and $(\tilde\theta_k, \beta_k^{FR})$, respectively. Note that this result is still valid even when the above two conditions are satisfied at the same time, although choice (1.11) is undefined in this case. □

This result suggests rewriting the first two cases and part of the third case in (1.11) and (1.14) as one case, except when $(\beta_k^{DY} - \beta_k^{HZ})\Delta_k = 0$ and $\Delta_k = 0$, respectively.

If these conditions hold, we suggest using a step of any conjugate gradient globally convergent method. In particular, we let (θ_k, ϑ_k) be equal to $(0, 1)$ which corresponds to the value of $(\eta_k, \beta_k) = (1, \beta_k^{HZ})$. The corresponding search direction satisfies the sufficient descent condition (0.5) for any line search technique.

Therefore, we replace choices (1.11) and (1.14) by the following two expressions:

$$(\theta_k, \vartheta_k) = \begin{cases} (\hat\theta_k, \hat\vartheta_k), & \text{if } \beta_k^{DY} \neq \beta_k^{HZ},\ \Delta_k \neq 0, \\ (0, 1), & \text{otherwise}, \end{cases} \tag{1.20}$$

and

$$(\eta_k, \beta_k) = \begin{cases} (\hat\eta_k, \hat\beta_k), & \text{if } \Delta_k \neq 0, \\ (1, \beta_k^{HZ}), & \text{otherwise}. \end{cases} \tag{1.21}$$

For convenience, we now show that the inequality $\Delta_k > 0$ holds in many cases.

Lemma 3. *The inequality* $\Delta_k > 0$ *holds if either* (i) $(d_k^\top g_{k+1})(g_k^\top g_{k+1}) \geq 0$ *or* (ii) *the strong Wolfe conditions* (0.3) *and* (0.4) *are employed with sufficiently small values of* σ.

P r o o f. Rearranging (1.19) as

$$\Delta_k = (d_k^\top g_{k+1})(g_k^\top g_{k+1}) - d_k^\top g_k \|g_{k+1}\|^2,$$

we observe that $\Delta_k > 0$ if condition (i) holds, since the descent property $d_k^\top g_k < 0$ holds. Now, noting that

$$\Delta_k \geq -|d_k^\top g_{k+1}||g_k^\top g_{k+1}| - d_k^\top g_k \|g_{k+1}\|^2$$

and using condition (0.4), it follows that

$$\Delta_k \geq -d_k^\top g_k (\|g_{k+1}\|^2 - \sigma |g_k^\top g_{k+1}|) \tag{1.22}$$

which is positive for sufficiently small values of σ. □

This result shows that $\Delta_k > 0$ if

$$\|g_{k+1}\|^2 - \sigma |g_k^\top g_{k+1}| \geq \hat{\sigma}, \tag{1.23}$$

which holds for sufficiently small values of the parameters σ and $\hat{\sigma} > 0$. However, this condition cannot be ensured when inexact line search is employed, because these parameters are defined prior to calculating g_{k+1} (e. g., $\sigma = 0.1$, $\hat{\sigma} = 10^{-4}$). Therefore, we modify choices (1.20) and (1.21) respectively as follows:

$$(\theta_k, \vartheta_k) = \begin{cases} (\hat{\theta}_k, \hat{\vartheta}_k), & \text{if } \beta_k^{DY} \neq \beta_k^{HZ},\ \|g_{k+1}\|^2 \geq \sigma |g_k^\top g_{k+1}| + \hat{\sigma}, \\ (0, 1), & \text{otherwise}, \end{cases}$$

and

$$(\eta_k, \beta_k) = \begin{cases} (\hat{\eta}_k, \hat{\beta}_k), & \text{if } \|g_{k+1}\|^2 \geq \sigma |g_k^\top g_{k+1}| + \hat{\sigma}, \\ (1, \beta_k^{HZ}), & \text{otherwise}. \end{cases} \tag{1.24}$$

Therefore, we consider the following outline of our method.

Algorithm 1.

Step 1. Give an initial point x_1, $\varepsilon \geq 0$, δ, σ and $\hat{\sigma} > 0$. Set $k = 1$ and let $d_1 = -g_1$.
Step 2. If $\|g_k\| \leq \varepsilon$, then stop.
Step 3. Calculate a steplength α_k such that the strong Wolfe conditions (0.3) and (0.4) hold.
Step 4. Set $x_{k+1} = x_k + \alpha_k d_k$.
Step 5. Compute η_k and β_k by (1.24) and hence d_{k+1} by (1.13).
Step 6. Set $k = k + 1$ and go to step 2.

§ 2. Global convergence result

Here, we analyse the convergence result for the proposed algorithm on general nonlinear functions, which is mainly based on Zoutendijk condition and the satisfaction of the sufficient descent condition.

We first introduce the following hypotheses on the objective function $f(x)$.

H1. The level set $S = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$ is bounded.

H2. In a neighbourhood N of S, the function f is bounded below and continuously differentiable, and its gradient $\nabla f(x)$ is Lipschitz continuous in N, i.e., there exists a constant $L > 0$ such that

$$\|\nabla f(\tilde{x}) - \nabla f(x)\| \leq L\|\tilde{x} - x\| \tag{2.1}$$

for all $x, \tilde{x} \in N$.

Under these assumptions, the norm $\|g(x)\|$ is bounded for all $x \in S$.

Lemma 4. *Supposing that the Hypotheses* **H1** *and* **H2** *are satisfied and the sequence* $\{x_k\}$ *is generated by* (0.2) *such that the descent property holds and* α_k *is determined such that the Wolfe condition* (0.3) *and the left inequality of* (0.4) *hold. Then, we obtain the Zoutendijk condition*

$$\sum_{k=0}^{\infty} \frac{(g_k^\top d_k)^2}{\|d_k\|^2} < \infty. \tag{2.2}$$

P r o o f. See for example [21] and essentially [24]. □

Theorem 1. *Let* x_0 *be given such that Assumptions* **H1** *and* **H2** *hold and the sequence* $\{x_k\}$ *be generated by Algorithm* 1 *with* $\epsilon = 0$. *Then*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{2.3}$$

P r o o f. Supposing by contradiction that there exists a positive constant γ_1 such that

$$\|g_k\| \geq \gamma_1, \qquad \forall\, k \geq 1.$$

Using this assumption and the sufficient descent condition (1.2), it follows that $(g_k^\top d_k)^2 \geq$ $\geq c^2\|g_k\|^4 \geq c^2\gamma_1^4$ and hence from (2.2) that $\sum_{k=0}^{\infty} \dfrac{1}{\|d_k\|^2} < \infty$. Thus, to contradict the latter inequality, we will show that $\|d_k\|$ is bounded in the following way.

We assume the first case of (1.24) is used and condition (1.23) holds infinitely many times. From (1.13) and (1.18), we obtain

$$d_{k+1} = -\eta_k g_{k+1} + \frac{y_k^\top g_{k+1}}{d_k^\top y_k}\eta_k d_k,$$

Using the strong Wolfe condition (0.4) and the sufficient descent condition (1.2), we obtain the following bounds on the curvature:

$$c(1-\sigma)\|g_k\|^2 \leq -(1-\sigma)d_k^\top g_k \leq d_k^\top y_k \leq -(1+\sigma)d_k^\top g_k. \tag{2.4}$$

Hence, from the Assumptions **H1** and **H2**, which yield $\|g_{k+1}\| < \gamma_2$ and the Lipschitz condition (2.1), we obtain

$$\begin{aligned}\|d_{k+1}\| &\leq |\eta_k|\|g_{k+1}\| + \frac{\|y_k\|\|g_{k+1}\|}{d_k^\top y_k}|\eta_k|\|d_k\| \\ &\leq |\eta_k|\Big(\gamma_2 + \frac{L\alpha_k\|d_k\|\gamma_2}{c(1-\sigma)\gamma_1^2}\|d_k\|\Big).\end{aligned}$$

We first show that $|\eta_k|$ is bounded. Assuming the strong Wolfe conditions are employed and using condition (1.23), the definition of η_k, as in (1.24) and (1.15), (1.22), the curvature condition (2.4), and the bound γ_2, we obtain

$$\begin{aligned}|\eta_k| &= \frac{c\|g_{k+1}\|^2(d_k^\top y_k)}{\Delta_k} \leq \frac{c\|g_{k+1}\|^2(d_k^\top y_k)}{-d_k^\top g_k(\|g_{k+1}\|^2 - \sigma|g_k^\top g_{k+1}|)} \\ &\leq \frac{c\|g_{k+1}\|^2(d_k^\top y_k)}{-d_k^\top g_k\hat{\sigma}} \leq \frac{c\|g_{k+1}\|^2(1+\sigma)}{\hat{\sigma}} \leq M,\end{aligned}$$

where $M = \frac{c\gamma_2^2(1+\sigma)}{\hat{\sigma}}$. Substituting this result in (2), it follows that

$$\|d_{k+1}\| \leq M\gamma_2 + M\frac{L\gamma_2\alpha_k\|d_k\|}{c(1-\sigma)\gamma_1^2}\|d_k\|. \tag{2.5}$$

By $\alpha_k\|d_k\| \to 0$, which follows from the first Wolfe condition (0.3), for a constant $m \in (0,1)$ there exists an integer k_1, such that

$$M\frac{L\gamma_2\alpha_k\|d_k\|}{c(1-\sigma)\gamma_1} \leq m < 1, \quad \forall\, k \geq k_1. \tag{2.6}$$

Therefore, by (2.5) and (2.6) we get

$$\|d_{k+1}\| \leq M\gamma_2 + m\|d_k\|, \quad \forall\, k \geq k_1$$

which implies

$$\|d_{k+1}\| \leq \frac{M\gamma_2}{1-m} + \|d_{k_1}\|m^{k+1-k_1}, \quad \forall\, k \geq k_1,$$

i. e., $\|d_{k+1}\| \leq \zeta$, where $\zeta = \frac{M\gamma_2}{1-m} + \|d_{k_1}\|m^{k+1-k_1}$.

Now assuming the second case of (1.24) is used infinitely many times.

The Theorem 3.2 in [14], the proof of global convergence shows that $\|d_{k+1}\|$ is bounded, and the bound is independent of $k > k_0$, i. e., $\|d_{k+1}\| \leq \kappa$.

Therefore $\|d_{k+1}\| \leq \max(\zeta, \kappa)$, it follows that $\|d_{k+1}\|$ is bounded and the global convergence result (2.3) is obtained, which completes the proof. □

Note that this global convergence result is still valid if the choice β_k^{HZ} in the second case of (1.24) is replaced by other choices (e. g., β_k^{DY}, β_k^{PR+}, etc.) which define globally convergent methods for certain values of σ. The latter choice is usually referred to as a restart of the method.

§ 3. Numerical results

In this section, we present a comparison between the performance of the new hybrid conjugate gradient method (1.24) and those of the HZ and DY methods, to determine the performance of all algorithms on a set of unconstrained optimization test problems [22]. Each problem is tested for a number of variables 1000, 1500, 2000, 5000, and 10000 so that the total number of test problems is the 80 unconstrained problems. We run them on a PC, Intel(R) core (TM) i5 CPU 650 @ 3.20 GHz, 3.00 Go RAM. With the parameter $c = 7/8$, using the strong Wolfe line search conditions (0.3) and (0.4) with $\delta = 0.0001$, $\sigma = 0.1$, $\sigma = 0.9$ (the methods are referred to as DYHZ/HZ.1 and DYHZ/HZ.9, respectively). The termination criterion for all algorithms is that $\|g_k\|^2 \leq 10^{-6}$.

We adopt the performance profiles proposed by Dolan and Moré [23] in order to obtain Figures 1, 2 and 3, showing CPU time, the number of iterations and the number of functions and gradient evaluations respectively, required to solve the problems. The figures clearly show that the proposed hybrid method performs substantially better than that of the efficient DY and HZ methods. We also observe that using $\sigma = 0.1$ gives a better performance than $\sigma = 0.9$, thus we use the former value for the following experiments. We repeated the run for the (1.24), but with β_k^{HZ} in the second case be replaced by β_k^{DY}, β_k^{PR+} and zero (the methods are referred to as DYHZ/DY.1, DYHZ/PR+.1 and DYHZ/SD.1, respectively). The results are represented by the three Figures 4, 5 and 6. We observed that the performance of the DYHZ/HZ.1 method is a little better than the other methods.

In addition, we have also done a numerical performance comparison of our method (1.24) with CG-DESCENT [15], HCG+ [16] and HCG [17] methods, which is presented by the Figures 7, 8 and 9, which showed that the DYHZ/HZ.1 method performs better than the CG-DESCENT, HCG and HCG+ methods.

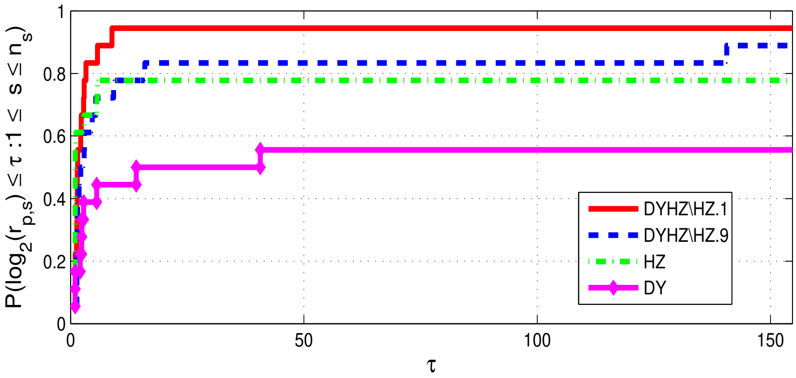

Fig. 1. Performance profile for CPU time

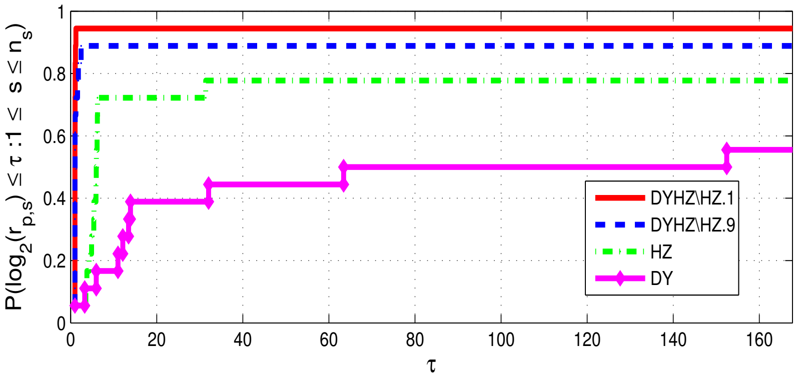

Fig. 2. Performance profile for the number of iterations

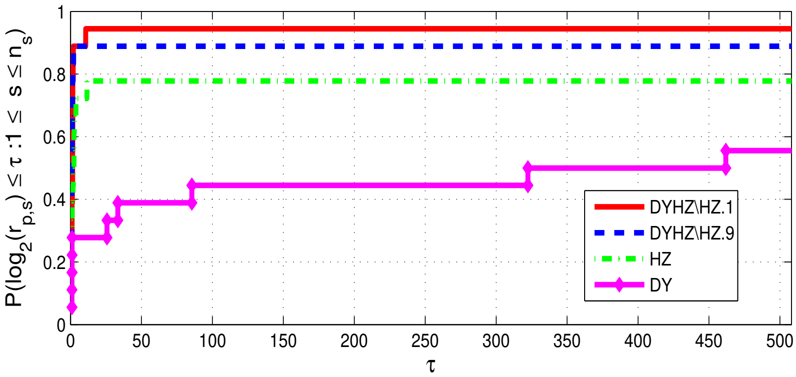

Fig. 3. Performance profile for the number of functions and gradient evaluations

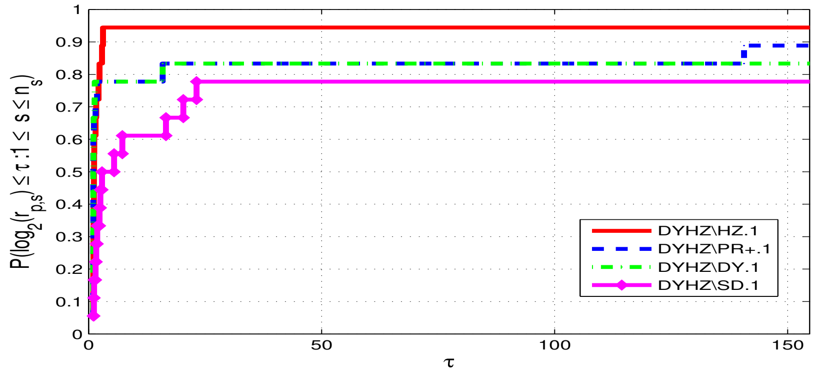

Fig. 4. Performance profile for CPU time

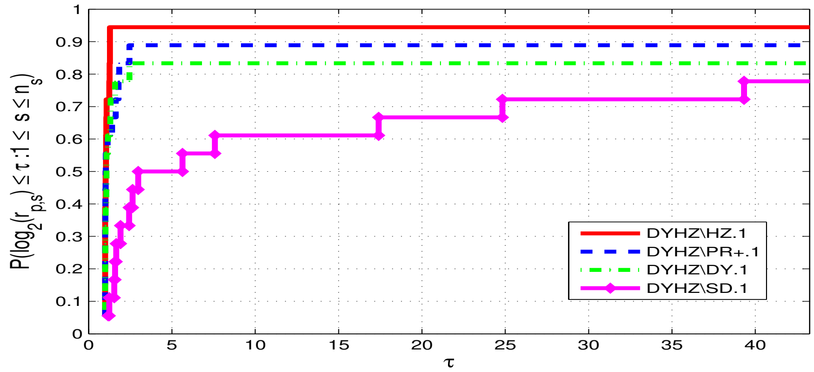

Fig. 5. Performance profile for the number of iterations

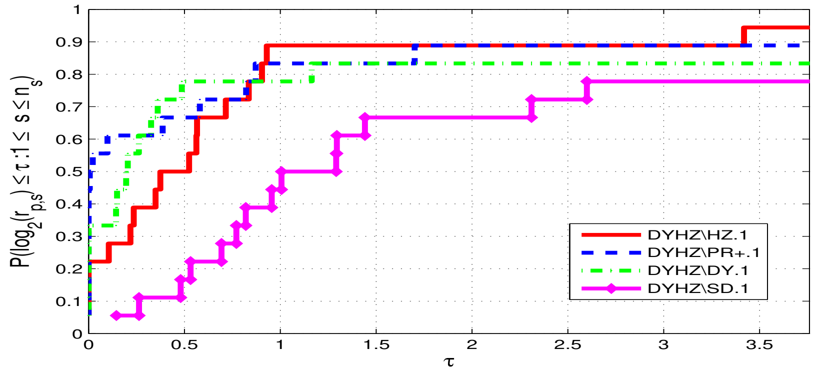

Fig. 6. Performance profile for the number of functions and gradient evaluations

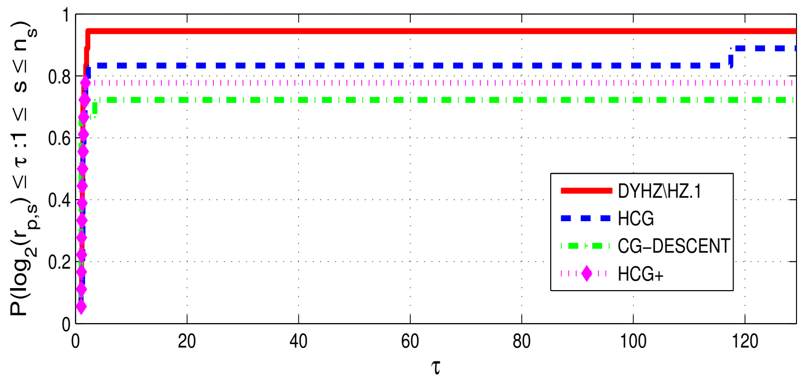

Fig. 7. Performance profile for CPU time

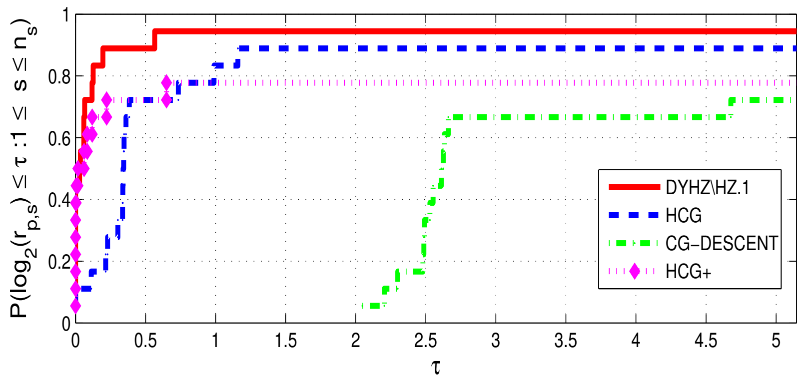

Fig. 8. Performance profile for the number of iterations

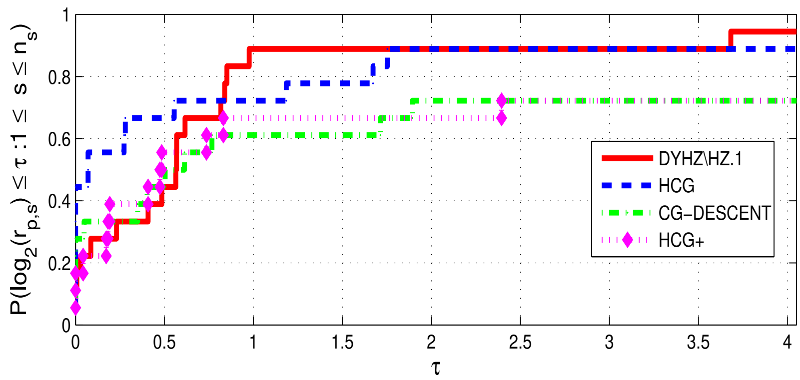

Fig. 9. Performance profile for the number of functions and gradient evaluations

§ 4. Conclusion

In this article, we have presented a new conjugate gradient method, based on the linear combination of DY and HZ conjugate gradient methods such that the sufficient descent and conjugacy conditions are satisfied. We also reported some numerical results for a set of standard test problems which show that the performance of the proposed hybrid DY/HZ method is substantially better than that of both DY and HZ methods in most cases.

Acknowledgements

The authors are very grateful to the anonymous referees for their useful suggestions and comments, which improved the quality of this paper.

REFERENCES

1. Al-Baali M., Narushima Y., Yabe H. A family of three-term conjugate gradient methods with sufficient descent property for unconstrained optimization, *Computational Optimization and Applications*, 2015, vol. 60, issue 1, pp. 89–110. https://doi.org/10.1007/s10589-014-9662-z
2. Andrei N. A new three-term conjugate gradient algorithm for unconstrained optimization, *Numerical Algorithms*, 2015, vol. 68, issue 2, pp. 305–321. https://doi.org/10.1007/s11075-014-9845-9
3. Djordjević S. S. New hybrid conjugate gradient method as a convex combination of LS and FR methods, *Acta Mathematica Scientia*, 2019, vol. 39, issue 1, pp. 214–228. https://doi.org/10.1007/s10473-019-0117-6
4. Liu J. K., Feng Y. M., Zou L. M. Some three-term conjugate gradient methods with the inexact line search condition, *Calcolo*, 2018, vol. 55, issue 2, article number: 16. https://doi.org/10.1007/s10092-018-0258-3
5. Chen Yuting, Cao Mingyuan, Yang Yueting. A new accelerated conjugate gradient method for large-scale unconstrained optimization, *Journal of Inequalities and Applications*, 2019, vol. 2019, issue 1, article number: 300. https://doi.org/10.1186/s13660-019-2238-9
6. Al-Baali M., Caliciotti A., Fasano G., Roma M. Exploiting damped techniques for nonlinear conjugate gradient methods, *Mathematical Methods of Operations Research*, 2017, vol. 86, issue 3, pp. 501–522. https://doi.org/10.1007/s00186-017-0593-1
7. Benrabia N., Laskri Y., Guebbai H., Al-Baali M. Applying the Powell's symmetrical technique to conjugate gradient methods with the generalized conjugacy condition, *Numerical Functional Analysis and Optimization*, 2016, vol. 37, issue 7, pp. 839–849. https://doi.org/10.1080/01630563.2016.1178142
8. Hestenes M. R., Stiefel E. Methods of conjugate gradients for solving linear systems, *Journal of Research of the National Bureau of Standards*, 1952, vol. 49, no. 6, pp. 409–436. https://doi.org/10.6028/jres.049.044
9. Fletcher R., Reeves C. M. Function minimization by conjugate gradients, *The Computer Journal*, 1964, vol. 7, issue 2, pp. 149–154. https://doi.org/10.1093/comjnl/7.2.149
10. Polak E., Ribière G. Note sur la convergence de méthodes de directions conjuguées, *Revue française d'informatique et de recherche opérationnelle. Série rouge*, 1969, vol. 3, no. 16, pp. 35–43 (in French). https://doi.org/10.1051/m2an/196903r100351
11. Polyak B. T. The conjugate gradient method in extremal problems, *USSR Computational Mathematics and Mathematical Physics*, 1969, vol. 9, issue 4, pp. 94–112. https://doi.org/10.1016/0041-5553(69)90035-4
12. Liu Y., Storey C. Efficient generalized conjugate gradient algorithms, part 1: Theory, *Journal of Optimization Theory and Applications*, 1991, vol. 69, issue 1, pp. 129–137. https://doi.org/10.1007/BF00940464
13. Dai Y. H., Yuan Y. A nonlinear conjugate gradient method with a strong global convergence property, *SIAM Journal on Optimization*, 1999, vol. 10, issue 1, pp. 177–182. https://doi.org/10.1137/S1052623497318992

14. Hager W. W., Zhang Hongchao. A new conjugate gradient method with guaranteed descent and an efficient line search, *SIAM Journal on Optimization*, 2005, vol. 16, issue 1, pp. 170–192. https://doi.org/10.1137/030601880
15. Hager W. W., Zhang Hongchao. Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent, *ACM Transactions on Mathematical Software*, 2006, vol. 32, no. 1, pp. 113–137. https://doi.org/10.1145/1132973.1132979
16. Babaie-Kafaki S., Ghanbari R. A hybridization of the Hestenes–Stiefel and Dai–Yuan conjugate gradient methods based on a least-squares approach, *Optimization Methods and Software*, 2015, vol. 30, issue 4, pp. 673–681. https://doi.org/10.1080/10556788.2014.966825
17. Babaie-Kafaki S., Ghanbari R. Two hybrid nonlinear conjugate gradient methods based on a modified secant equation, *Optimization*, 2014, vol. 63, issue 7, pp. 1027–1042. https://doi.org/10.1080/02331934.2012.693083
18. Gilbert J. C., Nocedal J. Global convergence properties of conjugate gradient methods for optimization, *SIAM Journal on Optimization*, 1992, vol. 2, issue 1, pp. 21–42. https://doi.org/10.1137/0802003
19. Al-Baali M. Descent property and global convergence of the Fletcher–Reeves method with inexact line search, *IMA Journal of Numerical Analysis*, 1985, vol. 5, issue 1, pp. 121–124. https://doi.org/10.1093/imanum/5.1.121
20. Fletcher R. *Practical methods of optimization*, John Wiley and Sons, 2000. https://doi.org/10.1002/9781118723203
21. Andrei N. Another accelerated conjugate gradient algorithm with guaranteed descent and conjugacy conditions for large-scale unconstrained optimization, *ICI Technical Report*, 2010.
22. Andrei N. An unconstrained optimization test functions collection, *Advanced Modeling and Optimization*, 2008, vol. 10, no. 1, pp. 147–161.
23. Dolan E. D., Moré J. J. Benchmarking optimization software with performance profiles, *Mathematical Programming*, 2002, vol. 91, issue 2, pp. 201–213. https://doi.org/10.1007/s101070100263
24. Zoutendijk G. Nonlinear programming computational methods, *Integer and Nonlinear Programming*, Amsterdam: North-Holland, 1970, pp. 37–86.

Received 02.07.2022

Accepted 17.03.2023

Imane Hafaidia, PhD student, Department of Mathematics, University 8 Mai 1945, BP 401, Guelma, 24000, Algeria.
ORCID: https://orcid.org/0000-0002-5130-1546
E-mail: hafaidia.imane@yahoo.com; E-mail: hafaidia.imane@univ-guelma.dz

Hamza Guebbai, Professor, Department of Mathematics, University 8 Mai 1945, BP 401, Guelma, 24000, Algeria.
ORCID: https://orcid.org/0000-0001-8119-2881
E-mail: guebaihamza@yahoo.fr; E-mail: guebbai.hamza@univ-guelma.dz

Mehiddin Al-Baali, Professor, Sultan Qaboos University, Oman.
ORCID: https://orcid.org/0000-0002-2515-2591
E-mail: albaali@squ.edu.om; E-mail: albaalim@gmail.com

Mourad Ghiat, PhD, Department of Mathematics, University 8 Mai 1945, BP 401, Guelma, 24000, Algeria.
ORCID: https://orcid.org/0000-0002-4484-2504
E-mail: mourad.ghi24@gmail.com; E-mail: ghiat.mourad@univ-guelma.dz

Citation: I. Hafaidia, H. Guebbai, M. Al-Baali, M. Ghiat. A new hybrid conjugate gradient algorithm for unconstrained optimization, *Vestnik Udmurtskogo Universiteta. Matematika. Mekhanika. Komp'yuternye Nauki*, 2023, vol. 33, issue 2, pp. 348–364.

МАТЕМАТИКА 2023. Т. 33. Вып. 2. С. 348–364.

И. Хафаидия, Х. Геббай, М. Аль-Баали, М. Гиат

Новый гибридный алгоритм сопряженного градиента для оптимизации без ограничений

Ключевые слова: оптимизация без ограничений, методы сопряженного градиента, условия сопряженности и достаточные условия спуска.

УДК 519.6

DOI: 10.35634/vm230211

Хорошо известно, что методы сопряженного градиента полезны при решении масштабных задач нелинейной оптимизации без ограничений. В данной работе мы рассматриваем объединение лучших свойств двух методов сопряженного градиента. В частности, мы даем новый метод сопряженного градиента, основанный на гибридизации полезных методов DY (Dai–Yuan) и HZ (Hager–Zhang). Параметры гибрида выбираются таким образом, чтобы предложенный метод удовлетворял условиям сопряженности и достаточного спуска. Показано, что новый метод сохраняет свойство глобальной сходимости двух вышеупомянутых методов. Описаны численные результаты для набора стандартных тестовых задач. Показано, что в большинстве случаев эффективность предложенного метода выше, чем у DY и HZ.

СПИСОК ЛИТЕРАТУРЫ

1. Al-Baali M., Narushima Y., Yabe H. A family of three-term conjugate gradient methods with sufficient descent property for unconstrained optimization // Computational Optimization and Applications. 2015. Vol. 60. Issue 1. P. 89–110. https://doi.org/10.1007/s10589-014-9662-z
2. Andrei N. A new three-term conjugate gradient algorithm for unconstrained optimization // Numerical Algorithms. 2015. Vol. 68. Issue 2. P. 305–321. https://doi.org/10.1007/s11075-014-9845-9
3. Djordjević S. S. New hybrid conjugate gradient method as a convex combination of LS and FR methods // Acta Mathematica Scientia. 2019. Vol. 39. Issue 1. P. 214–228. https://doi.org/10.1007/s10473-019-0117-6
4. Liu J. K., Feng Y. M., Zou L. M. Some three-term conjugate gradient methods with the inexact line search condition // Calcolo. 2018. Vol. 55. Issue 2. Article number: 16. https://doi.org/10.1007/s10092-018-0258-3
5. Chen Yuting, Cao Mingyuan, Yang Yueting. A new accelerated conjugate gradient method for large-scale unconstrained optimization // Journal of Inequalities and Applications. 2019. Vol. 2019. Issue 1. Article number: 300. https://doi.org/10.1186/s13660-019-2238-9
6. Al-Baali M., Caliciotti A., Fasano G., Roma M. Exploiting damped techniques for nonlinear conjugate gradient methods // Mathematical Methods of Operations Research. 2017. Vol. 86. Issue 3. P. 501–522. https://doi.org/10.1007/s00186-017-0593-1
7. Benrabia N., Laskri Y., Guebbai H., Al-Baali M. Applying the Powell's symmetrical technique to conjugate gradient methods with the generalized conjugacy condition // Numerical Functional Analysis and Optimization. 2016. Vol. 37. Issue 7. P. 839–849. https://doi.org/10.1080/01630563.2016.1178142
8. Hestenes M. R., Stiefel E. Methods of conjugate gradients for solving linear systems // Journal of Research of the National Bureau of Standards. 1952. Vol. 49. No. 6. P. 409–436. https://doi.org/10.6028/jres.049.044
9. Fletcher R., Reeves C. M. Function minimization by conjugate gradients // The Computer Journal. 1964. Vol. 7. Issue 2. P. 149–154. https://doi.org/10.1093/comjnl/7.2.149
10. Polak E., Ribière G. Note sur la convergence de méthodes de directions conjuguées // Revue française d'informatique et de recherche opérationnelle. Série rouge. 1969. Vol. 3. No. 16. P. 35–43 (in French). https://doi.org/10.1051/m2an/196903r100351
11. Поляк Б. Т. Метод сопряженных градиентов в задачах на экстремум // Журнал вычислительной математики и математической физики. 1969. Т. 9. № 4. С. 807–821. https://www.mathnet.ru/rus/zvmmf7115

12. Liu Y., Storey C. Efficient generalized conjugate gradient algorithms, part 1: Theory // Journal of Optimization Theory and Applications. 1991. Vol. 69. Issue 1. P. 129–137. https://doi.org/10.1007/BF00940464
13. Dai Y. H., Yuan Y. A nonlinear conjugate gradient method with a strong global convergence property // SIAM Journal on Optimization. 1999. Vol. 10. Issue 1. P. 177–182. https://doi.org/10.1137/S1052623497318992
14. Hager W. W., Zhang Hongchao. A new conjugate gradient method with guaranteed descent and an efficient line search // SIAM Journal on Optimization. 2005. Vol. 16. Issue 1. P. 170–192. https://doi.org/10.1137/030601880
15. Hager W. W., Zhang Hongchao. Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent // ACM Transactions on Mathematical Software. 2006. Vol. 32. No. 1. P. 113–137. https://doi.org/10.1145/1132973.1132979
16. Babaie-Kafaki S., Ghanbari R. A hybridization of the Hestenes–Stiefel and Dai–Yuan conjugate gradient methods based on a least-squares approach // Optimization Methods and Software. 2015. Vol. 30. Issue 4. P. 673–681. https://doi.org/10.1080/10556788.2014.966825
17. Babaie-Kafaki S., Ghanbari R. Two hybrid nonlinear conjugate gradient methods based on a modified secant equation // Optimization. 2014. Vol. 63. Issue 7. P. 1027–1042. https://doi.org/10.1080/02331934.2012.693083
18. Gilbert J. C., Nocedal J. Global convergence properties of conjugate gradient methods for optimization // SIAM Journal on Optimization. 1992. Vol. 2. Issue 1. P. 21–42. https://doi.org/10.1137/0802003
19. Al-Baali M. Descent property and global convergence of the Fletcher–Reeves method with inexact line search // IMA Journal of Numerical Analysis. 1985. Vol. 5. Issue 1. P. 121–124. https://doi.org/10.1093/imanum/5.1.121
20. Fletcher R. Practical methods of optimization. John Wiley and Sons, 2000. https://doi.org/10.1002/9781118723203
21. Andrei N. Another accelerated conjugate gradient algorithm with guaranteed descent and conjugacy conditions for large-scale unconstrained optimization // ICI Technical Report. 2010.
22. Andrei N. An unconstrained optimization test functions collection // Advanced Modeling and Optimization. 2008. Vol. 10. No. 1. P. 147–161.
23. Dolan E. D., Moré J. J. Benchmarking optimization software with performance profiles // Mathematical Programming. 2002. Vol. 91. Issue 2. P. 201–213. https://doi.org/10.1007/s101070100263
24. Zoutendijk G. Nonlinear programming computational methods // Integer and Nonlinear Programming. Amsterdam: North-Holland, 1970. P. 37–86.

Поступила в редакцию 02.07.2022

Принята к публикации 17.03.2023

Хафаидиа Имане, аспирант, кафедра математики, Университет 8 мая 1945 года, BP 401, Гельма, 24000, Алжир.
ORCID: https://orcid.org/0000-0002-5130-1546
E-mail: hafaidia.imane@yahoo.com; E-mail: hafaidia.imane@univ-guelma.dz

Геббай Хамза, профессор, кафедра математики, Университет 8 мая 1945 года, BP 401, Гельма, 24000, Алжир.
ORCID: https://orcid.org/0000-0001-8119-2881
E-mail: guebaihamza@yahoo.fr; E-mail: guebbai.hamza@univ-guelma.dz

Аль-Баали Мехиддин, профессор, Университет Султана Кабуса, Оман.
ORCID: https://orcid.org/0000-0002-2515-2591
E-mail: albaali@squ.edu.om; E-mail: albaalim@gmail.com

Гиат Мурад, доктор философии, кафедра математики, Университет 8 мая 1945 года, BP 401, Гельма, 24000, Алжир.
ORCID: https://orcid.org/0000-0002-4484-2504
E-mail: mourad.ghi24@gmail.com; E-mail: ghiat.mourad@univ-guelma.dz

Цитирование: И. Хафаидия, Х. Геббай, М. Аль-Баали, М. Гиат. Новый гибридный алгоритм сопряженного градиента для оптимизации без ограничений // Вестник Удмуртского университета. Математика. Механика. Компьютерные науки. 2023. Т. 33. Вып. 2. С. 348–364.