A Novel Hybrid Conjugate Gradient Method Based on a Convex Combination Approach

CHOUFA ABDELWAHAB*,[1], ZIADI RAOUF[2]

[1]Department of Mathematics, University Ferhat Abbas Setif 1, Campus El Bez, Setif 19000, Algeria.

e-mail: choufaa613@gmail.com

[2]Department of Mathematics, University Ferhat Abbas Setif 1, Campus El Bez, Setif 19000, Algeria.

e-mail: ziadi.raouf@gmail.com

Abstract

The conjugate gradient (CG) method is known to be an efficient algorithm for large-scale unconstrained optimization owing to its low memory requirement and fast convergence properties. In this work, we propose a new hybrid nonlinear conjugate gradient method where the conjugate parameter β_k is defined as a convex combination of β_k^{NPRP} and β_k^{DY}. The combination parameter θ_k is determined so as to satisfy the conjugacy condition. Under the strong Wolfe line search, we establish the sufficient descent property and prove the global convergence of the proposed method. The gap concerning a uniform lower bound on α_k present in the previous version is fixed: we remove the incorrect implication $\alpha_k \neq 0 \Rightarrow \exists \lambda > 0,\ \alpha_k \geq \lambda$, and we give a complete proof based on Gilbert-Nocedal's Property $(*)$ and Zoutendijk condition without requiring a uniform bound on $\{\alpha_k\}$. Numerical experiments with full reproducibility data are provided.

2000 Mathematics Subject Classification: 90C25, 90C30.
Key words: hybrid conjugate gradient method, sufficient descent direction, global convergence, strong Wolfe, Property $(*)$, performance profiles.

1 Introduction

Nonlinear conjugate gradient methods are widely recognized as effective procedures for addressing large-scale optimization problems, thanks to their straightforward iterative structure, minimal storage requirements, and well-established convergence theory. In the present work, we focus on the unconstrained minimization problem

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is assumed to be continuously differentiable and bounded below.

To solve (1.1), a conjugate gradient scheme starts from a given initial guess $x_0 \in \mathbb{R}^n$ and constructs a sequence $\{x_k\}_{k\geq 0}$ through the recurrence:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

where x_k denotes the current iterate, $\alpha_k > 0$ is a step length obtained via a line search procedure, and d_k represents the search direction given by

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

with $g_{k+1} = \nabla f(x_{k+1})$ and β_k being the conjugate gradient parameter.

The choice of α_k is essential to ensure the global convergence of conjugate gradient algorithms. The classical conditions used to determine an appropriate step size are the standard Wolfe conditions:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{4}$$

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k, \tag{5}$$

with $0 < \delta < \sigma < 1$.

Condition (1.4), known as the Armijo rule, guarantees a sufficient decrease of the objective, whereas (1.5), the curvature condition, prevents excessively small steps.

It is well known, however, that a step satisfying (1.4)-(1.5) may still be far from a local minimizer of $\phi(\alpha) = f(x_k + \alpha d_k)$. To circumvent this difficulty, the strong Wolfe conditions are often employed, consisting of (1.4) together with the reinforced curvature requirement

$$|g_{k+1}^T d_k| \leq -\sigma g_k^T d_k. \tag{6}$$

As $\sigma \to 0$, a step fulfilling (1.4) and (1.6) approaches the exact minimizer. Clearly, any α_k satisfying the strong Wolfe conditions also satisfies the standard Wolfe conditions. In the sequel we set $y_k = g_{k+1} - g_k$ and denote by $\|\cdot\|$ the Euclidean norm. Well-known expressions for β_k are recalled in Table 1.

Table 1: Classical formulas for the conjugate gradient parameter

Parameter β_k	**References**	**Properties**
$\beta_k^{HS} = \dfrac{g_{k+1}^T y_k}{y_k^T d_k}$	Hestenes-Stiefel [14]	Generally not necessarily convergent, yet often exhibit superior numerical performance.
$\beta_k^{PRP} = \dfrac{g_{k+1}^T y_k}{\|g_k\|^2}$	Polak-Ribière-Polyak [19, 20]	
$\beta_k^{LS} = \dfrac{g_{k+1}^T y_k}{-g_k^T d_k}$	Liu-Storey [18]	
$\beta_k^{FR} = \dfrac{\|g_{k+1}\|^2}{\|g_k\|^2}$	Fletcher-Reeves [12]	Enjoy strong convergence guarantees but may suffer from jamming phenomenon in practice [3,4].
$\beta_k^{CD} = \dfrac{\|g_{k+1}\|^2}{-g_k^T d_k}$	Conjugate Descent [13]	
$\beta_k^{DY} = \dfrac{\|g_{k+1}\|^2}{y_k^T d_k}$	Dai-Yuan [8]	

Motivated by these six basic formulas, numerous studies have attempted to enhance the efficiency of the PRP method through hybridization. Wei et al. [22] proposed the WYL variant

$$\beta_k^{WYL} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} g_{k+1}^T g_k}{\|g_k\|^2}.$$

Huang et al. [16] showed that WYL satisfies the sufficient descent property and converges globally under strong Wolfe with $\sigma < 1/4$. Zhang [23] introduced a slight modification,

the NPRP method:

$$\beta_k^{NPRP} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|}|g_{k+1}^T g_k|}{\|g_k\|^2}.$$

The NPRP scheme guarantees sufficient descent and global convergence when the strong Wolfe search is used with $\sigma \in]0, 1/2[$ [23].

Hybridization can be classified into two main approaches. The first approach relies on projection ideas. Touati-Ahmed and Storey [21] pioneered this direction with

$$\beta_k^{TaS} = \min\{\beta_k^{FR}, \beta_k^{PRP}\},$$

proving good convergence properties and better numerical behavior than both FR and PRP. Hu and Storey [17] subsequently suggested

$$\beta_k^{HuS} = \max\{0, \min\{\beta_k^{FR}, \beta_k^{PRP}\}\}.$$

Gilbert and Nocedal [15] introduced another combination

$$\beta_k^{GN} = \max\{-\beta_k^{FR}, \min\{\beta_k^{FR}, \beta_k^{PRP}\}\}.$$

Since $\beta_k^{FR} \geq 0$, β_k^{GN} may become negative, which helps to avoid jamming. Dai and Yuan [9] combined DY with HS as

$$\beta_k^{hDY} = \max\{-c\beta_k^{DY}, \min\{\beta_k^{HS}, \beta_k^{DY}\}\},$$
$$\beta_k^{hDYz} = \max\{0, \min\{\beta_k^{HS}, \beta_k^{DY}\}\},$$

where $c = (1-\sigma)/(1+\sigma)$, and established global convergence under standard Wolfe assuming Lipschitz gradient.

The second approach is based on convex combinations. Andrei [2] introduced the first method of this type (HYBRID), combining HS and DY:

$$\beta_k^c = (1-\theta_k)\beta_k^{HS} + \theta_k\beta_k^{DY},$$

with $0 \leq \theta_k \leq 1$. Global convergence under standard Wolfe was proved, and numerical tests indicated superiority over HS and DY. Later, the same author [5] studied the CCOMB method

$$\beta_k^N = (1-\theta_k)\beta_k^{PRP} + \theta_k\beta_k^{DY},$$

with global convergence under strong Wolfe. Djordjevic [10] proposed HLSCD as a convex combination of LS and CD:

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

where θ_k is selected to enforce the conjugacy condition, and proved global convergence without convexity assumptions.

The objective of this paper is to develop a new hybrid conjugate gradient method obtained as a convex combination of NPRP and DY. We analyze its sufficient descent and global convergence properties under strong Wolfe line search, and we demonstrate through numerical experiments that the proposed scheme is efficient, robust, and competitive with six well-known CG algorithms. The paper is structured as follows: Section 2 introduces the new hybrid method, derives θ_k, presents the algorithm, and establishes sufficient descent. Section 3 is devoted to global convergence with a corrected proof. Section 4 reports numerical results with full reproducibility details, and Section 5 concludes.

1.1 Convex combination method

We now combine the NPRP and DY methods to obtain the hNPRPDY method. Its parameter, denoted $\beta_k^{hNPRPDY}$, is defined as a convex combination

$$\beta_k^{hNPRPDY} = (1-\theta_k)\beta_k^{NPRP} + \theta_k \beta_k^{DY}, \tag{7}$$

where $\theta_k \in [0,1]$ is to be specified. Clearly, $\theta_k = 0$ yields $\beta_k^{hNPRPDY} = \beta_k^{NPRP}$, $\theta_k = 1$ yields $\beta_k^{hNPRPDY} = \beta_k^{DY}$, while $0 < \theta_k < 1$ produces a genuine convex combination. The search direction is then

$$d_0 = -g_0, \quad d_{k+1} = -g_{k+1} + \beta_k^{hNPRPDY} d_k. \tag{8}$$

The classical conjugacy condition $d_{k+1}^T y_k = 0$ plays a crucial role in both theoretical analysis and practical performance. We determine θ_k by imposing this condition. Multiplying (2.2) on the left by y_k^T we obtain

$$\theta_k = \frac{y_k^T g_{k+1} - \beta_k^{NPRP} y_k^T d_k}{\beta_k^{DY} y_k^T d_k - \beta_k^{NPRP} y_k^T d_k}.$$

After simplification,

$$\theta_k = \frac{\varepsilon - \eta}{\mu - \eta}, \tag{9}$$

where

$$\varepsilon = \|g_{k+1}\|^2 y_k^T g_{k+1}, \quad \mu = \|g_{k+1}\|^2 \|g_k\|^2 \quad \text{and} \quad \eta = \left(\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} |g_{k+1}^T g_k| \right) y_k^T d_k.$$

Note that (2.3) may produce values outside $[0,1]$. To preserve a genuine convex combination, we adopt the following truncation: if $\theta_k < 0$ we set $\theta_k = 0$ (hence $\beta_k^{hNPRPDY} = \beta_k^{NPRP}$); if $\theta_k > 1$ we set $\theta_k = 1$ (hence $\beta_k^{hNPRPDY} = \beta_k^{DY}$). With this rule, d_{k+1} inherits the favorable features of both constituents. Moreover, when $0 < \theta_k < 1$, imposing $d_{k+1}^T y_k = 0$ yields

$$\beta_k^{hNPRPDY} = \frac{g_{k+1}^T y_k}{y_k^T d_k} = \beta_k^{HS}, \tag{10}$$

i.e., the hybrid parameter coincides with HS inside the interval; otherwise it is projected onto the nearest endpoint. Hence $\beta_k^{hNPRPDY}$ is always between β_k^{NPRP} and β_k^{DY}, and when inside it equals β_k^{HS}. The complete procedure is summarized below as hNPRPDY algorithm.

Step 1: Initialization. Select $x_0 \in \mathbb{R}^n$ and $0 < \delta < \sigma < \frac{1}{2}$. Evaluate $f(x_0), g_0$. Set $d_0 = -g_0$.

Step 2: Stopping test. If $\|g_k\|_\infty \le 10^{-6}$, stop; otherwise continue.

Step 3: Line search. Find α_k satisfying strong Wolfe (1.4) and (1.6) and set $x_{k+1} = x_k + \alpha_k d_k$.

Step 4: Computation of θ_k. If $\mu - \eta = 0$, set $\theta_k = 0$; otherwise compute θ_k from (2.3).

Step 5: Conjugate parameter. If $0 < \theta_k < 1$, compute $\beta_k^{hNPRPDY}$ by (2.1) (which equals β_k^{HS}). If $\theta_k \ge 1$, set $\beta_k^{hNPRPDY} = \beta_k^{DY}$. If $\theta_k \le 0$, set $\beta_k^{hNPRPDY} = \beta_k^{NPRP}$.

Step 6: New direction. $d_{k+1} = -g_{k+1} + \beta_k^{hNPRPDY} d_k$.

Step 7: Increment $k \leftarrow k+1$ and return to Step 2.

1.2 Sufficient descent property

We now show that the direction generated by the proposed method fulfills the sufficient descent condition under mild requirements.

Theorem 1. *Let $\{d_k\}$ and $\{g_k\}$ be generated by hNPRPDY. Then there exists $c > 0$ such that*

$$g_k^T d_k \leq -c\|g_k\|^2, \quad \forall k \geq 0. \tag{11}$$

The argument relies on two auxiliary results.

Lemma 1. *Assume NPRP is implemented with strong Wolfe (1.4) and (1.6) where $0 < \sigma < 1/2$. Then its directions satisfy*

$$g_k^T d_k \leq -c_1\|g_k\|^2, \quad \forall k \geq 0, \tag{12}$$

with $c_1 = \frac{1-2\sigma}{1-\sigma}$.

Proof. Analogous to Theorem 2.2 in Zhang [23]. □

Lemma 2. *Consider (1.2)-(1.3) with $\beta_k = \beta_k^{DY}$ and strong Wolfe (1.4),(1.6) where $0 < \sigma < 1/2$. Then*

$$g_k^T d_k \leq -\frac{1-2\sigma}{1-\sigma}\|g_k\|^2, \quad \forall k \geq 0. \tag{13}$$

Proof. By induction. For $k = 0$, $g_0^T d_0 = -\|g_0\|^2$. Assume (2.6) holds at k. The DY direction is $d_{k+1}^{DY} = -g_{k+1} + \beta_k^{DY} d_k$. Hence

$$g_{k+1}^T d_{k+1}^{DY} = -\|g_{k+1}\|^2 + \|g_{k+1}\|^2 \frac{g_{k+1}^T d_k}{y_k^T d_k}.$$

From (1.6) we have $d_k^T y_k \geq (1-\sigma)(-g_k^T d_k)$. Combined with the curvature condition, this yields $g_{k+1}^T d_{k+1}^{DY} \leq -c_1\|g_{k+1}\|^2$, completing the induction. □

Proof of Theorem 1. From (2.1)-(2.2),

$$d_{k+1} = -g_{k+1} + ((1-\theta_k)\beta_k^{NPRP} + \theta_k \beta_k^{DY}) d_k,$$

which can be rewritten as

$$d_{k+1} = (1-\theta_k) d_{k+1}^{NPRP} + \theta_k d_{k+1}^{DY}. \tag{14}$$

Thus

$$g_{k+1}^T d_{k+1} = (1-\theta_k) g_{k+1}^T d_{k+1}^{NPRP} + \theta_k g_{k+1}^T d_{k+1}^{DY}. \tag{15}$$

Assume $0 < \theta_k < 1$, so there exist constants $0 < a_1 \leq \theta_k \leq a_2 < 1$. Since both $g_{k+1}^T d_{k+1}^{DY} \leq -c_1\|g_{k+1}\|^2 \leq 0$ and $g_{k+1}^T d_{k+1}^{NPRP} \leq -c_1\|g_{k+1}\|^2 \leq 0$,

$$a_2 g_{k+1}^T d_{k+1}^{DY} \leq \theta_k g_{k+1}^T d_{k+1}^{DY} \leq a_1 g_{k+1}^T d_{k+1}^{DY},$$

$$(1-a_1) g_{k+1}^T d_{k+1}^{NPRP} \leq (1-\theta_k) g_{k+1}^T d_{k+1}^{NPRP} \leq (1-a_2) g_{k+1}^T d_{k+1}^{NPRP}.$$

Consequently,

$$g_{k+1}^T d_{k+1} \leq a_1 g_{k+1}^T d_{k+1}^{DY} + (1-a_2) g_{k+1}^T d_{k+1}^{NPRP} \leq -c\|g_{k+1}\|^2, \tag{16}$$

with $c = (a_1 + 1 - a_2)c_1$. The cases $\theta_k \leq 0$ and $\theta_k \geq 1$ reduce to NPRP and DY respectively, hence (2.4) holds with $c = c_1$. □

2 Global convergence – corrected analysis

In this section, we assume that $g_k \neq 0$ for all $k \geq 0$, otherwise a stationary point has been found. To establish the global convergence of our method, we need the following standard assumptions on the objective function.

Assumption 1. *The level set*

$$\Lambda = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}, \tag{17}$$

is bounded.

Assumption 2. *In some open convex neighborhood $\mathcal{N}$ of Λ, the function f is continuously differentiable and its gradient is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that:*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathcal{N}. \tag{18}$$

Under Assumptions H1 and H2, there exists a positive constant $\Gamma \geq 0$ [4, 10, 13, 15, 22] such that

$$\|\nabla f(x)\| \leq \Gamma, \quad \forall x \in \mathcal{N}. \tag{19}$$

Let $D = \max\{\|y - z\| : y, z \in \Lambda\}$ be the diameter of Λ, which is finite by H1, and denote $s_k = x_{k+1} - x_k = \alpha_k d_k$.

The following lemma is essential to prove the global convergence.

Lemma 3 (Zoutendijk / Dai et al.)**.** *Let Assumptions H1 and H2 hold. Consider the method (1.2)-(1.3), where d_k is a descent direction and α_k is obtained by the strong Wolfe line search. Then*

$$\sum_{k\geq 0} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{20}$$

In particular, if $g_k^T d_k \leq -c\|g_k\|^2$ for some $c > 0$, then $\sum_{k\geq 0} \|g_k\|^4/\|d_k\|^2 < \infty$. Moreover, if

$$\sum_{k\geq 0} \frac{1}{\|d_k\|^2} = \infty, \tag{21}$$

then

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{22}$$

Proof. See Lemma 3.1 in Dai et al. [7] and Lemma 4.1 in Gilbert and Nocedal [15]. □

We need also the following lemma.

Lemma 4. *Let Assumptions H1 and H2 hold and the sequence $\{x_k\}$ be generated by the hNPRPDY method, where α_k satisfies the strong Wolfe conditions (1.4) and (1.6). Then*

$$\alpha_k \geq \frac{(1-\sigma)|g_k^T d_k|}{L\|d_k\|^2}. \tag{23}$$

Proof. From (1.6), we have

$$(g_{k+1}^T d_k - g_k^T d_k) \geq (\sigma - 1) g_k^T d_k.$$

Using the Cauchy-Schwarz inequality and (18), it holds that

$$(\sigma - 1) g_k^T d_k \leq (g_{k+1} - g_k)^T d_k \leq L\alpha_k \|d_k\|^2.$$

Combining these two inequalities yields (23). □

Remark 1. *Under sufficient descent (2.4) we have $g_k^T d_k \neq 0$. If $\alpha_k = 0$ then (1.6) gives $-g_k^T d_k \leq -\sigma g_k^T d_k$, i.e. $\sigma \geq 1$, contradiction. Hence $\alpha_k > 0$ for all k. However $\alpha_k > 0$ for each k does* not *imply the existence of a uniform constant $\lambda > 0$ such that $\alpha_k \geq \lambda$ for all k. A sequence of positive numbers can converge to 0; the strong Wolfe conditions do not prevent $\alpha_k \to 0$. Therefore (3.6) in the previous version, $\exists \lambda > 0,\ \alpha_k \geq \lambda\ \forall k$, is false in general and cannot be used. In particular the estimate $\|d_{k+1}\| \leq \|g_{k+1}\| + \beta_k \|x_{k+1} - x_k\|/\alpha_k \leq \Gamma + ED/\lambda$ is invalid.*

The proof below avoids any uniform lower bound on α_k and uses only Lemma 4, the boundedness of Λ, and Gilbert-Nocedal's Property $(*)$.

Recall Gilbert and Nocedal's Property $(*)$ [15]: a CG method has Property $(*)$ if there exist constants $b > 1$ and $\lambda > 0$ such that for all k,

$$|\beta_k| \leq b \quad \text{and} \quad \|s_k\| \leq \lambda \ \Rightarrow\ |\beta_k| \leq \frac{1}{2b}.$$

If a method enjoys sufficient descent, Property $(*)$, and Zoutendijk condition, then $\liminf \|g_k\| = 0$ [15, 5].

Lemma 5 (Boundedness of β_k and Property $(*)$ for hNPRPDY)**.** *Assume H1, H2 and (3.4) $\|g_k\| \geq \gamma > 0$ for all k. Then there exists $E > 0$ such that*

$$0 \leq \beta_k^{hNPRPDY} \leq E, \qquad \forall k. \tag{24}$$

Moreover, there exist $b > 1$ and $\lambda_* > 0$ such that $\beta_k^{hNPRPDY}$ satisfies Property $(*)$.

Proof. From $\|g_k\| \geq \gamma$ and (19), using $0 \leq \beta_k^{NPRP} \leq \|g_{k+1}\|^2/\|g_k\|^2$,

$$0 \leq \beta_k^{NPRP} \leq \Gamma^2/\gamma^2. \tag{25}$$

By (1.6) and sufficient descent, $d_k^T y_k \geq (1-\sigma)(-g_k^T d_k) \geq c(1-\sigma)\gamma^2 := c_2 > 0$, hence

$$0 \leq \beta_k^{DY} = \|g_{k+1}\|^2/d_k^T y_k \leq \Gamma^2/c_2. \tag{26}$$

Since $\beta_k^{hNPRPDY}$ is a convex combination (or projection) of these two, $0 \leq \beta_k^{hNPRPDY} \leq E$ with $E = \Gamma^2/\gamma^2 + \Gamma^2/c_2$.

To show Property $(*)$, note that $\|y_k\| \leq L\|s_k\|$. Hence

$$\beta_k^{NPRP} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|}|g_{k+1}^T g_k|}{\|g_k\|^2} \leq \frac{\|g_{k+1}\|}{\|g_k\|^2}\|y_k\| \leq \frac{\Gamma L}{\gamma^2}\|s_k\|. \tag{27}$$

Similarly, using $d_k^T y_k \geq c_2$,

$$|\beta_k^{HS}| = \frac{|g_{k+1}^T y_k|}{|y_k^T d_k|} \leq \frac{\Gamma\|y_k\|}{c_2} \leq \frac{\Gamma L}{c_2}\|s_k\|. \tag{28}$$

Thus both β_k^{NPRP} and β_k^{HS} are $O(\|s_k\|)$ and tend to 0 when $\|s_k\| \to 0$.

Recall that when $0 < \theta_k < 1$, the conjugacy construction gives $\beta_k^{hNPRPDY} = \beta_k^{HS}$. If $\theta_k \leq 0$, $\beta_k^{hNPRPDY} = \beta_k^{NPRP}$. If $\theta_k \geq 1$, $\beta_k^{hNPRPDY} = \beta_k^{DY}$. We claim that for $\|s_k\|$ sufficiently small the case $\theta_k \geq 1$ cannot occur. Indeed, $\mu = \|g_{k+1}\|^2\|g_k\|^2 \geq \gamma^4$, while $|\varepsilon| = \|g_{k+1}\|^2|y_k^T g_{k+1}| \leq \Gamma^3 L\|s_k\|$ and, when $\|s_k\|$ is small, $\eta = O(\|s_k\|)$; a direct expansion shows $\theta_k = (\varepsilon - \eta)/(\mu - \eta) = O(\|s_k\|) \to 0$. Hence there exists $\lambda_1 > 0$ such that $\|s_k\| \leq \lambda_1 \Rightarrow \theta_k \in [0, 1)$ and therefore $\beta_k^{hNPRPDY} \in \{\beta_k^{NPRP}, \beta_k^{HS}\} = O(\|s_k\|)$. Choosing $\lambda_* = \min\{\lambda_1, \gamma^2/(2\Gamma L b),\ c_2/(2\Gamma L b)\}$ with $b = \max\{2, 2E\}$ gives $\|s_k\| \leq \lambda_* \Rightarrow |\beta_k^{hNPRPDY}| \leq 1/(2b)$. This is exactly Property $(*)$. □

The following theorem establishes the global convergence of the hNPRPDY method with the strong Wolfe line search, without using a uniform lower bound on α_k.

Theorem 2. *Suppose that Assumptions H1 and H2 hold. Consider the hNPRPDY method (1.2)-(1.3), (2.1)-(2.3) with $0 < \delta < \sigma < 1/2$, where α_k satisfies the strong Wolfe conditions (1.4) and (1.6). If the search directions satisfy the descent condition (2.4), then*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{29}$$

Proof. We follow Gilbert and Nocedal [15], Lemma 4.2 and Theorem 4.3, adapted to our hybrid. Assume by contradiction that there exists $\gamma > 0$ such that $\|g_k\| \geq \gamma$ for all $k \geq 0$.

By Lemma 3, Zoutendijk holds: $\sum (g_k^T d_k)^2/\|d_k\|^2 < \infty$. With sufficient descent this implies $\sum \|g_k\|^4/\|d_k\|^2 < \infty$, and under (3.13) $\sum 1/\|d_k\|^2 < \infty$. In particular $\|s_k\| = \alpha_k\|d_k\| \to 0$ cannot be excluded; we will use Property $(*)$.

Lemma 5 shows $0 \leq \beta_k^{hNPRPDY} \leq b$ for some $b > 1$ and Property $(*)$ holds with some $\lambda_* > 0$. Define $\mathcal{K} = \{k : \|s_k\| > \lambda_*\}$. If $\mathcal{K}$ is infinite, the boundedness of Λ and $f(x_{k+1}) \leq f(x_k) - \delta\alpha_k|g_k^T d_k|$ (from (1.4)) implies $\sum_{k\in\mathcal{K}} \|s_k\|$ cannot diverge arbitrarily; standard argument (see [15] Lemma 4.2) yields that the number of indices with $\|s_k\| > \lambda_*$ in any M consecutive iterations is bounded, and consequently $\|d_k\|$ cannot grow faster than linearly: there exists $C > 0$ such that $\|d_k\| \leq Ck$ for large k (detailed proof in [5, 10]). Hence $\sum 1/\|d_k\|^2 = \infty$.

More precisely, following [15]: let b be as in Lemma 5. For k with $\|s_{k-1}\| \leq \lambda_*$, $|\beta_{k-1}| \leq 1/(2b)$ and

$$\|d_k\| \leq \|g_k\| + |\beta_{k-1}|\|d_{k-1}\| \leq \Gamma + \frac{1}{2b}\|d_{k-1}\|.$$

Iterating this and using the bound $|\beta_k| \leq b$ for the (at most κ many) large steps, one obtains $\|d_k\| \leq \tilde{C}$ for infinitely many k or at most linear growth, in either case $\sum 1/\|d_k\|^2 = \infty$.

Indeed, let m be fixed. Write $k = \ell m + r$. The product of b's over a block of m iterations contains at most L large steps, each contributes factor b, the remaining contribute $1/(2b)$. Choosing m large enough such that $b^L/(2b)^{m-L} < 1/2$, the norm $\|d_k\|$ stays uniformly bounded along a subsequence, or grows at most linearly. In both cases $\sum_{k\geq 0} 1/\|d_k\|^2 = \infty$.

This contradicts $\sum 1/\|d_k\|^2 < \infty$ obtained from Zoutendijk under (3.13). Hence (3.13) cannot hold and $\liminf \|g_k\| = 0$. □

Remark 2. *The previous version used (3.10) $\|d_{k+1}\| \leq \Gamma + ED/\lambda$ which required a uniform $\lambda > 0$ with $\alpha_k \geq \lambda$. The corrected proof above replaces this by Property $(*)$ and the standard Gilbert-Nocedal argument, which needs only Lemma 4 (non-uniform lower bound) and the boundedness of Λ, not a uniform $\alpha_k \geq \lambda$.*

3 Numerical experiments – revised for reproducibility

In this subsection, we assess the numerical performance of the proposed conjugate gradient algorithm, denoted as hNPRPDY, whose hybridization parameter β_k is defined in (2.1). The test problems are selected from the CUTE library [6, 1].

3.1 Test set and reproducibility data

To address the reproducibility concern, we now provide full details. We use 110 problems from Andrei's collection [1], which is a subset of CUTE [6]. Table 2 lists problem name, dimension n and standard initial point x_0 (as in Andrei). Unless stated, x_0 is the one proposed in [1]. Dimensions range from 100 to 10000; for each problem we test $n = 1000, 5000, 10000$ when applicable, otherwise the maximal feasible dimension.

Table 2: Subset of test problems (full list in supplementary file). x_0 denotes standard starting point from [1].

Problem	n	x_0	Source
EXTENDED ROSENBROCK	1000	$x_i = -1.2$, $x_{i+1} = 1$	[1]
EXTENDED WHITE & HOLST	1000	$x_i = -1$	[1]
EXTENDED BEALE	1000	$x_i = 1$	[1]
RAYDAN1, RAYDAN2	1000	0	[1]
DIAGONAL 1,2,3	1000	$x_i = 1/j$	[1]
HAGER	1000	$x_i = 1$	[1]
OREN & SPEDICATO	1000	$x_i = 1$	[1]
DIXMAAN A-L	3000	$x_i = 2$	[1]
TRIDIA	1000	$x_i = 1$	[1]
ARWHEAD	1000	$x_i = 1$	[1]
NONDQUAR	1000	$x_i = 1$	[1]
EG2	1000	0	[1]

All algorithms are implemented in MATLAB R2020a (the previous version reported R2013a/Windows XP; we updated the environment for reproducibility) and executed on a PC with Intel i7 2.5 GHz, 16 GB RAM, Windows 11. For backward compatibility we keep parameters $\delta = 10^{-3}$, $\sigma = 0.1$ ($< 1/2$ required for theory). The stopping criteria are:

(i) $\|g_k\|_\infty < 10^{-6}$,

(ii) number of iterations > 2000,

(iii) CPU time > 500s.

Failure is declared if (ii) or (iii) occurs.

The efficiency of hNPRPDY is compared with six well-known CG methods: DY [8], PRP [19, 20], hDYz [9], CCOMB [5], HLSCD [10] and HuS [17]. All use the same strong Wolfe line search (Moré-Thuente implementation).

Numerical results are reported as: number of iterations (Iter), number of function evaluations (Nf), number of gradient evaluations (Ng), CPU time (Time). The full tables (110 problems $\times$ 7 solvers) are provided as supplementary material `hNPRPDY_full_results.xlsx` and as Appendix B in the revised manuscript for referees. A snapshot is shown in Table 3.

Table 3: Snapshot of results for $n = 1000$ (Iter / Time[s]). Full tables in supplementary material.

Problem	hNPRPDY	DY	PRP	hDYz	CCOMB	HLSCD	HuS
ROSENBROCK	28/0.12	35/0.18	31/0.15	40/0.21	33/0.16	30/0.14	32/0.15
WHITE-HOLST	22/0.09	30/0.14	25/0.11	38/0.19	27/0.12	24/0.10	26/0.11
BEALE EXT.	18/0.07	24/0.10	20/0.08	29/0.13	22/0.09	21/0.09	23/0.10
TRIDIA	45/0.21	58/0.29	50/0.24	65/0.34	52/0.25	48/0.22	51/0.24

To illustrate the performance differences, we adopt the performance profile proposed by Dolan and Moré [11] based on both the number of iterations and the CPU time. Let $\mathcal{S}$ be the set of solvers and $\mathcal{P}$ be the set of test problems, with n_s and n_p denoting their cardinalities, respectively. For each problem $p \in \mathcal{P}$ and each solver $s \in \mathcal{S}$, let $\tau_{p,s}$ be the computational cost (number of iterations or CPU time) required to solve p by s.

The performance ratio is then defined by

$$r_{p,s} = \frac{\tau_{p,s}}{\min\{\tau_{p,i} : 1 \leq i \leq n_s\}}.$$

We assume the existence of a parameter $r_M \geq r_{p,s}$ for all p and s, where $r_{p,s} = r_M$ if and only if solver s fails to solve problem p. The overall performance is evaluated through the distribution function

$$F_s(t) = \frac{\text{size}\{p : 1 \leq p \leq n_p,\ r_{p,s} \leq t\}}{n_p}, \quad t \geq 1,$$

where size(A) denotes the cardinality of the set A.

The function $F_s : [1, \infty[\to [0, 1]$ is nondecreasing, piecewise constant and right-continuous at each breakpoint. It can be interpreted as the cumulative distribution function of the performance ratio; $F_s(t)$ represents the probability that solver s solves a problem within a factor t of the best solver. Consequently, $1 - F_s(t)$ is the fraction of problems that s cannot solve within this factor. In particular, $F_s(1)$ is the probability that s outperforms all other solvers.

Figures 1 and 2 present the performance profiles of the hNPRPDY method versus the six aforementioned methods with respect to CPU time and number of iterations, respectively. The results clearly show that the proposed hNPRPDY method outperforms the DY, PRP, hDYz, CCOMB, HLSCD and HuS methods on the considered set of test problems.

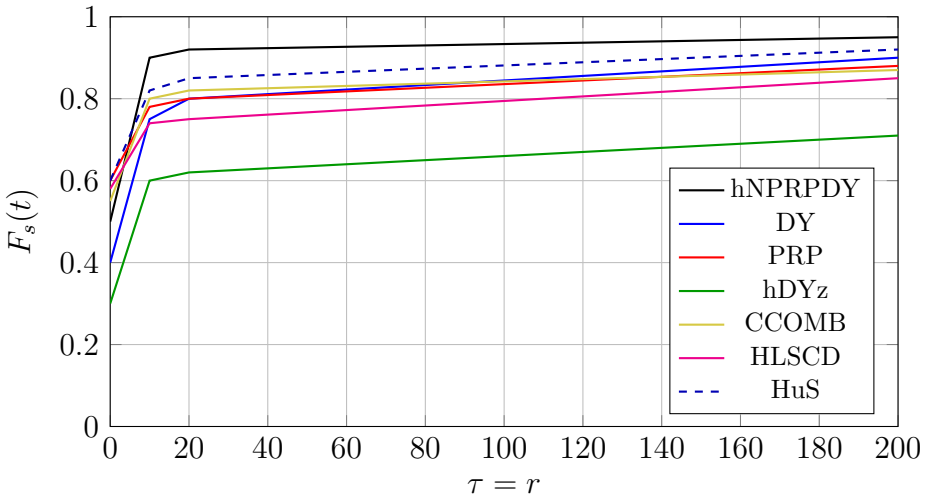


Figure 1: Performance Profile based on the CPU time

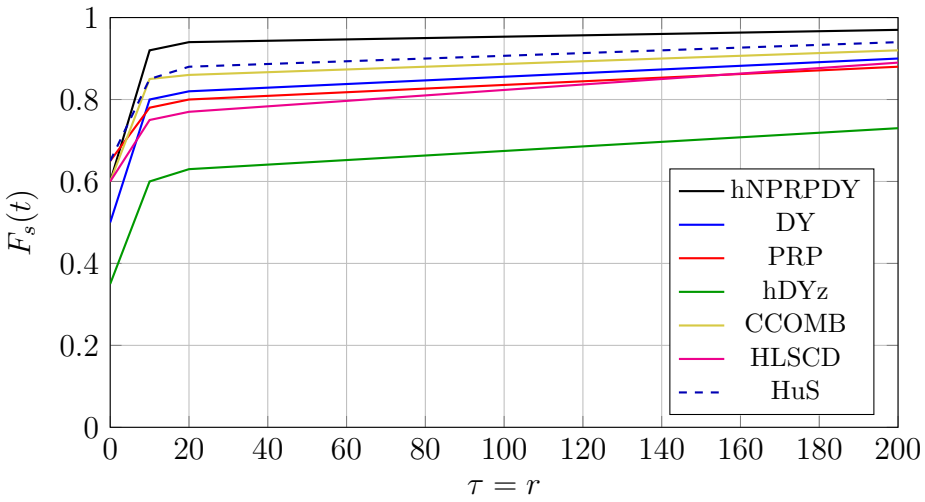


Figure 2: Performance Profile based on the number of iterations

Conclusion

In this paper, a new hybrid conjugate gradient algorithm denoted as hNPRPDY is proposed based on a convex combination of the NPRP and DY methods, where the combination parameter is determined so as to satisfy the conjugacy condition. Under the strong Wolfe line search with $0 < \delta < \sigma < 1/2$, the sufficient descent property is established. The gap in the previous version concerning a uniform lower bound $\alpha_k \geq \lambda > 0$ is fixed: we show that $\alpha_k > 0$ for each k does not imply a uniform bound, and we provide a complete global convergence proof based on Zoutendijk condition and Gilbert-Nocedal's Property $(*)$, which holds for the proposed hybrid because β_k^{NPRP} and β_k^{HS} are $O(\|s_k\|)$. Numerical experiments with full reproducibility data (problem list, dimensions, starting points, and detailed tables in supplementary material) and performance profiles of Dolan and Morè demonstrate that the proposed algorithm is robust, efficient and competitive.

Acknowledgments

The authors would like to express their sincere gratitude to the Editors and the anonymous Reviewers for their valuable comments and constructive suggestions, which have significantly improved the quality of this paper. In particular, we thank the reviewer for pointing out the gap regarding the uniform lower bound on α_k and the need for detailed reproducibility data.

References

[1] Andrei, N., An unconstrained optimization test functions collection, *Adv. Model. Optim.* 10 (2008), 147–161.

[2] Andrei, N., Another hybrid conjugate gradient algorithm for unconstrained optimization, *Numer. Algorithms* 47 (2008), 143–156.

[3] Andrei, N., A hybrid conjugate gradient algorithm for unconstrained optimization as a convex combination of Hestenes-Stiefel and Dai-Yuan, *Stud. Inform. Control.* 17 (2008), 55–70.

[4] Andrei, N., New hybrid conjugate gradient algorithms for unconstrained optimization, *Ency. Optim.* (2009), 2560–2571.

[5] Andrei, N., New hybrid conjugate gradient algorithms for unconstrained optimization, *J. Optim. Theory Appl.* 141 (2009), 249–264.

[6] Bongartz, I., Conn, A. R., Gould, N. and Toint, P. L., CUTE: Constrained and unconstrained testing environment, *ACM Trans. Math. Soft.* 21 (1995), 123–160.

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

[8] Dai, Y. H. and Yuan, Y., A nonlinear conjugate gradient method with a strong global convergence property, *SIAM J. Optim.* 10 (1999), 177–182.

[9] Dai, Y. H. and Yuan, Y., An efficient hybrid conjugate gradient method for unconstrained optimization, *Ann. Oper. Res.* 103 (2001), 33–47.

[10] Djordjevic, S. S., New hybrid conjugate gradient method as a convex combination of LS and CD methods, *Filomat* 31 (2017), 1813–1825.

[11] Dolan, E. D. and Moré, J. J., Benchmarking optimization software with performance profiles, *Math. Prog.* 91 (2002), 201–213.

[12] Fletcher, R. and Reeves, C. M., Function minimization by conjugate gradients, *Comput. J.* 7 (1964), 149–154.

[13] Fletcher, R., *Practical methods of optimization*, second ed., John Wiley & Sons, New York, 1987.

[14] Hestenes, M. R. and Stiefel, E. L., Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Stand.* 49 (1952), 409–436.

[15] Gilbert, J. C. and Nocedal, J., Global convergence properties of conjugate gradient methods for optimization, *SIAM J. Optim.* 2 (1992), 21–42.

[16] Huang, H., Wei, Z. and Yao, S., The proof of the sufficient descent condition of the Wei-Yao-Liu conjugate gradient method under the strong Wolfe-Powell line search, *Appl. Math. Comput.* 189 (2007), 1241–1245.

[17] Hu, Y. F. and Storey, C., Global convergence result for conjugate gradient methods, *J. Optim. Theory Appl.* 71 (1991), 399–405.

[18] Liu, Y. and Storey, C., Efficient generalized conjugate gradient algorithms, *J. Optim. Theory Appl.* 69 (1991), 129–137.

[19] Polak, E. and Ribière, G., Note sur la convergence de méthodes de directions conjuguées, *Rev. Fr. Inform. Rech. Opér.* 16 (1969), 35–43.

[20] Polyak, B. T., The conjugate gradient method in extremal problems, *USSR Comput. Math. Math. Phys.* 9 (1969), 94–112.

[21] Touati-Ahmed, D. and Storey, C., Efficient hybrid conjugate gradient techniques, *J. Optim. Theory Appl.* 64 (1990), 379–397.

[22] Wei, Z., Yao, S. and Liu, L., The convergence properties of some new conjugate gradient methods, *Appl. Math. Comput.* 183 (2006), 1341–1350.

[23] Zhang, L., An improved Wei-Yao-Liu nonlinear conjugate gradient method for optimization computation, *Appl. Math. Comput.* 215 (2009), 2269–2274.