

# 针对非凸向量优化问题的两种新型混合共轭梯度方法

罗雅欣<sup>1</sup>, 彭伟光<sup>1</sup>

<sup>1</sup> 西南大学数学与统计学院, 重庆 400715

**摘要:** 本文针对非凸向量优化问题, 基于Fletcher-Reeves (FR)、Conjugate Descent (CD) 和Polak-Ribière-Polyak (PRP) 共轭梯度法, 提出了三种改进的共轭参数。这些改进方案是其原始向量形式的直接推广。其中, 改进的FR型和CD型共轭参数生成的搜索方向无需依赖特定线搜索即可自然满足充分下降条件, 其全局收敛性在标准Wolfe线搜索条件下得以建立。鉴于改进FR/CD型方法在理论性质上的优势与PRP方法突出的数值表现具有互补性, 我们通过三种改进参数的策略性组合, 进一步构建了两两种新型混合共轭梯度法。在强Wolfe线搜索策略下, 这些混合方法无需对目标函数施加限制性假设即可保证全局收敛性。数值实验验证了所提五种方法的可行性与有效性。

**关键词:** 向量优化; 混合共轭梯度方法; 充分下降性; 全局收敛

**中图分类号:** O221.2; O224

## Two Novel Hybrid Conjugate Gradient Methods for Nonconvex Vector Optimization

LUO Ya-Xin<sup>1</sup>, PENG Wei-Guang<sup>1</sup>

<sup>1</sup> School of Mathematics and statistics, Southwest University, Chongqing 400715

**Abstract:** In this paper, for nonconvex vector optimization problems, we propose three improved conjugate parameters derived from the Fletcher-Reeves (FR), Conjugate Descent (CD), and Polak-Ribière-Polyak (PRP) conjugate gradient methods. These modified variants are direct adaptations of their original vector-form counterparts. The search directions generated by the improved FR-type and CD-type conjugate parameters inherently satisfy the sufficient descent condition without requiring specific line searches, while their global convergence properties are established under standard Wolfe line search conditions. Recognizing the complementary strengths of the improved FR/CD-type methods in theoretical properties and PRP's superior numerical performance, we develop two novel hybrid conjugate gradient methods through strategic combinations of the three improved parameters. The global convergence of these hybrid methods is proved under the strong Wolfe line search strategy, without imposing restrictive assumptions on the objective function. Numerical experiments demonstrate the feasibility and effectiveness of all five proposed methods.

**Key words:** Vector optimization; Hybrid conjugate gradient; Sufficiently descent; Global convergence

## 0 Introduction

Conjugate gradient method is an iterative algorithm that efficiently addresses both linear and non-linear optimization problems and has been widely applied across various fields, such as image processing, feature extraction, and machine learning [1, 2, 3]. Due to their efficiency and minimal memory

Foundations:

Author Introduction: Luo Yaxin, female, major research direction: optimization theory and intelligent decision-making, E-mail: lyaxin0725@163.com. Correspondence author: Peng Weiguang, male, major research direction: mathematical logic and computational complexity theory, E-mail: pwgedu@swu.edu.cn

demands, conjugate gradient (CG) methods are a prominent class of first-order algorithms for solving unconstrained optimization problem:

$$\min f(x) \text{ s.t. } x \in \mathbb{R}^n,$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is continuously differentiable. The nonlinear conjugate gradient method has the following basic iteration form:

$$x^{k+1} = x^k + \alpha_k d^k, \quad (0.1)$$

where  $x^k$  is the current iterate,  $\alpha_k$  is a step size, and  $d^k$  is a search direction defined by

$$d^k = \begin{cases} -\nabla f(x^k), & k = 1, \\ -\nabla f(x^k) + \beta_k d^{k-1}, & k \geq 2, \end{cases} \quad (0.2)$$

where  $\beta_k$  is the conjugate parameter. The conjugate parameter  $\beta_k$  is very vital to ensure the descent property and sufficient descent property of  $d^k$  in (0.2). The different choices of  $\beta_k$  lead to different conjugate gradient methods. In [4], Fletcher and Reeves proposed the first nonlinear CG method in the 1960s, called the Fletcher-Reeves (FR) method. Other recognized conjugate gradient methods include the Polak-Ribière-Polyak (PRP, 1969) method [5], the Hestenes-Stiefel (HS, 1952) method [6], the Dai-Yuan (DY, 1999) method [7], the conjugate descent (CD, 1987) method [8], the Liu-Storey (LS, 1991) method [9], and the Hager-Zhang (HZ, 2005) method [10]. In these conjugate gradient methods, the conjugate parameters  $\beta_k$  are defined as the follows:

$$\begin{aligned} \beta_k^{FR} &:= \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \beta_k^{PRP} := \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \quad \beta_k^{HS} := \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \\ \beta_k^{DY} &:= \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, \quad \beta_k^{CD} := -\frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}}, \quad \beta_k^{LS} := -\frac{g_k^T y_{k-1}}{g_{k-1}^T d_{k-1}}, \\ \beta_k^{HZ} &:= \frac{1}{d_{k-1}^T} \left( y_{k-1}^T g_k - 2 \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \|y_{k-1}\|^2 \right), \end{aligned}$$

where  $g_k := \nabla f(x^k)$ ,  $y_{k-1} := g_k - g_{k-1}$  and  $\|\cdot\|$  is the Euclidean norm. A remarkable feature of the FR, CD, and DY conjugate gradient methods is that if a line search such as the Wolfe conditions is employed, their search directions satisfy the sufficient descent condition, i.e.,

$$\langle \nabla f(x^k), d^k \rangle \leq -c \|\nabla f(x^k)\|^2, \quad \forall k \geq 1, \quad (0.3)$$

where  $c$  is a positive constant, see [8, 11, 12]. As demonstrated in [11] that the PRP method may not generate a descent direction even under the strong Wolfe conditions with arbitrary choice of  $\sigma \in (0, 1)$ . We can also conclude from [6, 9] that neither the HS method nor the LS method can guarantee a descent direction even under the Wolfe line search conditions. To address these limitations, truncation techniques were introduced in [9, 13]. They considered the convergence issues of PRP, HS, and LS, thereby achieving global convergence for these three methods.

In recent years, numerous modified conjugate parameters were developed based on existing conjugate parameters to enhance the descent properties and numerical performance of the conjugate gradient method. For example, regarding the PRP method, many scholars focus on the global convergence analysis of the PRP method. Notably, Wei et al. [14] proposed a PRP-type parameter that possesses sufficient descent properties and ensures global convergence of the PRP type conjugate gradient method. This parameter was defined as follows:

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

Chen et al. [1] proposed a modified PRP type method with the conjugate gradient parameter:

$$\beta_k^{PRP-T} := \frac{g_k^\top y_{k-1}}{\|g_{k-1}\|^2 + \mu \|y_{k-1}\| \|d_{k-1}\|},$$

where  $\mu > 1$ . This method inherently satisfied the sufficient descent property without requiring any line search. Additional modified variants can be found in [15, 16, 17, 18, 19]. Generally, CG methods have undergone significant advancements with recent progress documented in [1, 20, 21, 22, 23, 24, 25, 26].

Vector optimization is a fundamental component of optimization and control theory, which has been extensive applications, such as engineering systems, space exploration, financial modeling, resource management, location science, and statistics; see [27, 28, 29, 30]. One of the important research areas is how to develop efficient algorithms for vector optimization. Current scalarization methods are very vital to design algorithms for solving vector optimization problems, which convert vector problems into single-objective formulations by the scalarization functions, subsequently addressed through standard optimization techniques for dealing with vector optimization problems. As a second-order optimization technique, Newton's method [31] exhibited superior convergence properties at the expense of demanding Hessian matrix inversion, which imposed significant computational burdens, particularly in high-dimensional parameter spaces. It is well-established that the conjugate gradient method outperforms the steepest descent method in numerical performance. Moreover, while being computationally simpler than Newton's method, the conjugate gradient method achieves comparable convergence rates. Thus, it is meaningful to develop the conjugate gradient method for vector optimization.

Recently, several conjugate gradient methods originally designed for scalar optimization have been extended to the vector case. The pioneering work in this field was conducted by [32], which generalized the FR, CD, DY, PRP, and HS conjugate gradient methods to vector optimization. Subsequently, the vector counterparts of the HZ and LS methods were proposed in [33, 34] proposed, respectively. Their respective vector forms are defined as follows:

$$\begin{aligned} \beta_k^{FR} &:= \frac{\psi(x^k, v(x^k))}{\psi(x^{k-1}, v(x^{k-1}))}, \quad \beta_k^{CD} := \frac{\psi(x^k, v(x^k))}{\psi(x^{k-1}, d^{k-1})}, \quad \beta_k^{DY} := \frac{-\psi(x^k, v(x^k))}{\psi(x^k, d^{k-1}) - \psi(x^{k-1}, d^{k-1})}, \\ \beta_k^{HS} &:= \frac{-\psi(x^k, v(x^k)) + \psi(x^{k-1}, v(x^k))}{\psi(x^k, d^{k-1}) - \psi(x^{k-1}, d^{k-1})}, \quad \beta_k^{PRP} := \frac{-\psi(x^k, v(x^k)) + \psi(x^{k-1}, v(x^k))}{-\psi(x^{k-1}, v(x^{k-1}))}, \\ \beta_k^{LS} &:= \frac{-\psi(x^k, v(x^k)) + \psi(x^{k-1}, v(x^k))}{-\psi(x^{k-1}, d^{k-1})}, \quad \beta_k^{HZ} := \frac{1}{\tau_2 - \tau_4} \left( \tau_1 - \mu \tau_2 \frac{\tau_1 + \tau_3}{\tau_2 - \tau_4} \right), \end{aligned}$$

where  $\mu > \frac{1}{4}$  and

$$\begin{aligned} \tau_1 &:= -\psi(x^{k+1}, v(x^{k+1})) + \psi(x^k, v(x^{k+1})), \quad \tau_2 := \psi(x^{k+1}, d^k), \\ \tau_3 &:= \psi(x^{k+1}, v(x^k)) - \psi(x^k, v(x^k)), \quad \tau_4 := \psi(x^k, d^k). \end{aligned}$$

where the functions  $\psi(\cdot, \cdot)$  and  $v(\cdot)$  are defined by (1.5) and (1.7), respectively. Lucambio Pérez and Prudente [32] proved that if  $\beta_k := \max\{\beta_k^{PRP}, 0\}$  or  $\beta_k := \max\{\beta_k^{HS}, 0\}$ , the global convergence of conjugate gradient algorithms related to the PRP and HS parameters can be achieved under strong Wolfe conditions and the commonly used assumption. Similarly, the global convergence of conjugate gradient method can also be attained when  $\beta_k = \max\{\beta_k^{LS}, 0\}$  in [33]. However, the search directions  $d^k$  generated by the PRP, HS, and LS conjugate gradient methods fail to satisfy the sufficient descent condition despite this truncation technique. As pointed out in [32], whether the FR method with inexact

line search can automatically generate descent directions remains an open problem, while the global convergence of the CD method can only be guaranteed when its conjugate parameter maintains non-negativity and bounded above by an appropriate fraction of the CD parameter. In terms of numerical performance, the PRP method exhibits significantly superior performance compared to both the FR and CD methods.

Based on the above analyses of conjugate gradient methods in both single-objective and vector optimization contexts, this analysis reveals distinct advantages across different conjugate gradient methods: some exhibit superior theoretical properties such as sufficient descent property and the global convergence, others demonstrate competitive numerical performance in practice. Naturally, an intriguing question is how can we leverage the advantages of different conjugate gradient methods to develop a more highly efficient algorithm for solving vector optimization. This answer has been extensively validated in scalar optimization. This method was called a hybrid conjugate gradient method, which combined various conjugate gradient approaches. The foundation of hybrid conjugate gradient methods in scalar optimization dates back to Touati-Ahmed and Storey [35], who pioneered the hybridization of the FR and PRP methods to enhance overall performance. The hybrid conjugate gradient method demonstrated significant overall performance improvement by integrating the advantages of different conjugate gradient techniques. This hybrid approach not only inherits the low-memory characteristics of classical conjugate gradient methods but also effectively reduces algorithm restart frequency through dynamically adjusting the conjugate parameter  $\beta_k$  calculation mechanism, thereby substantially enhancing computational efficiency. This balanced characteristic between computational efficiency and convergent theoretical guarantees makes the hybrid conjugate gradient method an essential tool in modern optimization fields. For the hybrid methods, one can refer to [2, 36, 37, 38]. Analogously, hybrid conjugate gradient methods hold significant promise for vector optimization. In [39], Yahaya et al. introduced the first hybrid conjugate gradient framework for vector optimization problems, combining three vector-case variants: NCD (modified CD), VFR (modified FR), and VPRP (modified PRP). Numerical experiments demonstrate that their hybrid method outperforms the non-negative PRP and HS methods proposed in [32].

Inspired by the above works, we introduce three modified conjugate parameters denoted as MVFR, MVCD, and MVPRP for the vector case by applying a correction method given by the PRP-T parameter in [1]. Both the MVFR and MVCD conjugate parameters possess the non-negativity property. The search directions generated by MVFR and MVCD conjugate parameters inherently satisfy the sufficient descent property without relying on any line search. Then two novel hybrid conjugate gradient algorithms for solving vector optimization are proposed by combining the MVPRP conjugate parameter with the MVFR, and MVCD conjugate parameters. Under the strong Wolfe line search conditions and general assumptions, we establish the global convergence of the two proposed hybrid conjugate gradient methods. To evaluate the highly efficiency of the proposed algorithms, we test both hybrid algorithms and three novel conjugate gradient methods (based on MVFR, MVCD, and MVPRP) using benchmark problems from multiobjective optimization literature by comparing the non-negative PRP, FR, and CD methods [32] as well as the hybrid conjugate gradient method [39]. The ultimate experimental outcomes demonstrate that the two hybrid conjugate gradient methods exhibit not only enhanced robustness but also promising performance characteristics.

This paper is organized as follows. Section 1 introduces essential notations and existing theoretical results. In Section 2, we give the three proposed modified conjugate parameters and present two novel hybrid conjugate gradient methods. The global convergence properties of the proposed methods are established in Section 3. Numerical experiments are reported in Section 4. Finally, we arrive the conclusion in Section 5.

# 1 Preliminaries

Let  $K$  be a closed, convex and pointed cone of  $\mathbb{R}^m$  with non-empty interior, and  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a continuously differentiable vector-valued function. For a subset  $H$  of  $\mathbb{R}^m$  with inner product  $\langle \cdot, \cdot \rangle$  and norm  $\|\cdot\|$ , the conic hull and convex hull of  $H$  are denoted by  $\text{cone}(H)$  and  $\text{conv}(H)$ , respectively. The partial order  $\preceq_K$  in  $\mathbb{R}^m$  induced by  $K$  is defined by

$$u \preceq_K v \Leftrightarrow v - u \in K, u, v \in \mathbb{R}^m,$$

and the partial order  $\prec_K$  in  $\mathbb{R}^m$  induced by  $\text{int}(K)$  is defined by

$$u \prec_K v \Leftrightarrow v - u \in \text{int} K, u, v \in \mathbb{R}^m.$$

Let  $K^*$  be the dual cone of  $K$  defined as

$$K^* := \{w \in \mathbb{R}^m \mid \langle w, z \rangle \geq 0, \forall z \in K\}.$$

It is well-known that  $K = K^{**}$  and  $\text{int}(K) = \{z \in \mathbb{R}^m \mid \langle w, z \rangle > 0, \forall w \in K^* \setminus \{0\}\}$  because  $K$  is a closed and convex set.

A compact set  $C \subset K^* \setminus \{0\}$  is called a base of  $K^*$  iff  $K^* = \text{cone}(\text{conv} C)$ .

If  $K$  is polyhedral,  $C$  can be taken as the finite set of extremal rays of  $K^*$ . In particular, if  $K = \mathbb{R}_+^m$ ,  $C$  can be taken as the canonical basis of  $\mathbb{R}^m$ . if  $K = \mathbb{R}_+$ , we take  $C = \{1\}$ .

Following [32], this paper adopts:

$$C := \{w \in K^* \mid \|w\| = 1\}. \quad (1.1)$$

In this paper, we investigate the following unconstrained vector optimization problem:

$$\min_K F(x) \quad \text{s.t. } x \in \mathbb{R}^n. \quad (1.2)$$

We now recall some basic definitions and well-known facts.

**Definition 1.1.** A point  $x^*$  is said to be a weakly efficient solution of problem (1.2) iff, there exists no other point  $x \in \mathbb{R}^n$  such that  $F(x) \prec_K F(x^*)$ .

Observed from [32] that if  $x^*$  is a weakly efficient solution of problem (1.2), then

$$-\text{int} K \cap \text{Image}(JF(x^*)) = \emptyset, \quad (1.3)$$

where  $JF(x^*)$  is the Jacobian of  $F$  at  $x^*$ , and  $\text{Image}(JF(x^*))$  represents the image of the linear operator  $JF(x^*)$  on  $\mathbb{R}^n$ . Meanwhile, a point  $x \in \mathbb{R}^n$  is called a  $K$ -critical point (stationary point) of problem (1.2) iff, (1.3) holds. It follows from (1.3) that if  $x$  is not  $K$ -critical point, there exists a vector  $v \in \mathbb{R}^n$  for  $F$  at  $x$  such that  $JF(x)v \in -\text{int} K$ . This implies that  $v$  can be regarded as a descent direction of  $F$  at  $x$

**Definition 1.2.** [40] A nonzero vector  $v \in \mathbb{R}^n$  is a  $K$ -descent direction of  $F$  at  $x$  iff, there exist  $\varepsilon > 0$  such that  $F(x + \alpha v) \prec_K F(x)$ ,  $\forall \alpha \in (0, \varepsilon]$ .

We next define two functions  $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}$  and  $\psi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  by

$$\varphi(z) = \sup \{\langle z, w \rangle \mid w \in \text{conv}(C)\}, \quad (1.4)$$

and

$$\psi(x, d) = \varphi(JF(x)d) = \sup \{\langle JF(x)d, w \rangle \mid w \in C\}. \quad (1.5)$$

By the compactness of  $C$ , we conclude that  $\varphi$  is well-defined. It follows from [22] that

$$-K = \{z \in \mathbb{R}^m \mid \varphi(z) \leq 0\} \text{ and } -\text{int}K = \{z \in \mathbb{R}^m \mid \varphi(z) < 0\}.$$

It is easy to see that  $\psi$  is closed and convex with respect to the second argument. Besides, a vector  $d$  is a  $K$ -descent direction of  $F$  at  $x$  if and only if  $\psi(x, d) < 0$ .  $x$  is a  $K$ -critical point of  $F$  if and only if  $\psi(x, d) \geq 0$  for all  $d \in \mathbb{R}^n$  hold, see [41].

Given the definition of the function  $\psi$ , we can define the sufficient descent condition in the vector setting, which corresponds to (0.3). A vector  $d \in \mathbb{R}^n$  is said to satisfy the sufficient descent condition at  $x \in \mathbb{R}^n$  if and only if there exists a constant  $c > 0$  such that

$$\psi(x, d) \leq c\psi(x, v(x)), \quad (1.6)$$

where  $v(x)$  is the proximal mapping of  $\psi$  with respect to the second argument at 0 defined by

$$v(x) = \arg \min \left\{ \psi(x, d) + \frac{\|d\|^2}{2} \mid d \in \mathbb{R}^n \right\}. \quad (1.7)$$

Since  $\psi(x, \cdot)$  is a closed convex function,  $v(x)$  exists uniquely.  $v(x)$  can be regarded as the extension of steepest descent direction to the vector case. Meanwhile, the optimal value

$$\theta(x) := \psi(x, v(x)) + \frac{\|v(x)\|^2}{2}. \quad (1.8)$$

In particular, if  $m := 1$ ,  $K := \mathbb{R}_+$ ,  $C := \{1\}$ , we have

$$\psi(x, d) = \langle \nabla F(x), d \rangle, \quad v(x) = -\nabla F(x)$$

and  $\theta(x) = -\frac{\|\nabla F(x)\|^2}{2}$ . If  $K := \mathbb{R}_+^m$ ,  $C$  is taken as the canonical basis of  $\mathbb{R}^m$ ,  $v(x)$  can be computed by solving convex quadratic optimization problem with linear inequality constraints [42]:

$$\begin{aligned} \min \quad & \alpha + \frac{1}{2} \|d\|^2, \\ \text{s.t.} \quad & [JF(x)d]_i \leq \alpha, \quad i = 1, \dots, m. \end{aligned} \quad (1.9)$$

With the steepest descent direction  $v(x)$  in the vector setting, one can define the search direction  $d^k$  in the conjugate gradient method (0.1) for the vector optimization problems as follows:

$$d^k := \begin{cases} v(x^k), & k = 1, \\ v(x^k) + \beta_k d^{k-1}, & k \geq 2. \end{cases} \quad (1.10)$$

In the following, we recall fundamental properties of the functions  $\varphi$  and  $\psi$ .

**Lemma 1.1.** [22, 41] *Let  $w^1, w^2 \in \mathbb{R}^m$ . The following results hold true.*

- (i)  $\varphi(w^1 + w^2) \leq \varphi(w^1) + \varphi(w^2)$  and  $\varphi(w^1 - w^2) \geq \varphi(w^1) - \varphi(w^2)$ .
- (ii) If  $w^1 \prec_K w^2$ , then  $\varphi(w^1) < \varphi(w^2)$ .
- (iii) If  $w^1 \preceq_K w^2$ , then  $\varphi(w^1) \leq \varphi(w^2)$ .
- (iv)  $\varphi(\cdot)$  is 1-Lipschitz continuous.

**Lemma 1.2.** [22, 41] Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a continuously differentiable function. Then, the following statements hold:

- (i) For any  $x, d^1, d^2 \in \mathbb{R}^n$  and  $\alpha \geq 0$ , we can get  $\psi(x, d^1 + \alpha d^2) \leq \psi(x, d^1) + \alpha \psi(x, d^2)$ .
- (ii) The mapping  $(x, d) \rightarrow \psi(x, d)$  is continuous.
- (iii) Let  $x^1, x^2, d \in \mathbb{R}^n$ . Then  $|\psi(x^1, d) - \psi(x^2, d)| \leq \|JF(x^1) - JF(x^2)\| \|d\|$ . If  $JF$  is Lipschitz continuous with constant  $L$ , we have  $\psi(\cdot, d)$  is  $L\|d\|$ -Lipschitz continuous.

**Lemma 1.3.** [41] Let the functions  $\theta(\cdot)$  and  $v(\cdot)$  be defined by (1.8) and (1.7). The following assertions hold.

- (i) If  $x$  is  $K$ -critical point of problem (1.2), then  $v(x) = 0$  and  $\theta(x) = 0$ .
- (ii) If  $x$  is not  $K$ -critical point of problem (1.2), then

$$v(x) \neq 0, \theta(x) < 0, \psi(x, v(x)) < -\frac{\|v(x)\|^2}{2} < 0.$$

- (iii) The mapping  $v(\cdot)$  and  $\theta(\cdot)$  are continuous.

Observed from Lemma 1.3 (ii) that  $v(x)$  is a  $K$ -descent direction of  $F$  at  $x$  when  $x$  is not  $K$ -critical point of problem (1.2). If  $d \in \mathbb{R}^n$  satisfy the sufficient descent condition at  $x \in \mathbb{R}^n$  and  $x$  is not  $K$ -critical point of problem (1.2), then using (1.6) implies that  $d$  is a  $K$ -descent direction of  $F$  at  $x$ . So, the sufficient descent condition is stronger than the descent direction condition. We next review the exact line search. Let  $e \in \text{int } K$  satisfy

$$\langle w, e \rangle \leq 1, \forall w \in C, \quad (1.11)$$

and  $d \in \mathbb{R}^n$  be a  $K$ -descent direction of  $F$  at  $x$ . If  $K = \mathbb{R}_+^m$ , we can take  $e = (1, 1, \dots, 1)^\top \in \mathbb{R}^m$ . If  $\psi(x + \alpha d, d) = 0$  holds for some  $\alpha > 0$ , then  $\alpha$  is said to be obtained via exact line search.

**Definition 1.3.** [41] Let  $\alpha > 0$  and  $\rho, \sigma$  are two positive constants with  $\rho < \sigma < 1$ .  $\alpha$  is said to satisfy:

- (i) the standard Wolfe conditions iff,

$$\begin{aligned} F(x + \alpha d) &\preceq_K F(x) + \rho \alpha \psi(x, d) e, \\ \psi(x + \alpha d, d) &\geq \sigma \psi(x, d). \end{aligned} \quad (1.12)$$

- (ii) the strong Wolfe conditions iff,

$$\begin{aligned} F(x + \alpha d) &\preceq_K F(x) + \rho \alpha \psi(x, d) e, \\ |\psi(x + \alpha d, d)| &\leq \sigma |\psi(x, d)|. \end{aligned} \quad (1.13)$$

**Lemma 1.4.** [32, lemma 2.4] Let  $p, q \in \mathbb{R}$  be any scalars and  $t \neq 0$ . Then the following inequality hold:  $(p + q)^2 \leq (1 + 2t^2)p^2 + (1 + \frac{1}{2t^2})q^2$ .

## 2 Two novel hybrid conjugate gradient algorithms for vector optimization

In this section, we propose three new modified conjugate parameters such as MVFR, MVCD, and MVPRP conjugate parameters and two novel hybrid conjugate gradient methods for solving unconstrained vector optimization problems (1.2). The search directions generated by MVFR and MVCD



conjugate parameters is proved to satisfy the sufficient descent property without relying on any line search.

Let  $u, t_1 > 0$  and  $t_2 > t_1$ . The MVFR, MVCD and MVPRP conjugate parameters are defined as follows:

$$\beta_k^{MVFR} := \frac{-t_1 \psi(x^k, v(x^k))}{-\psi(x^{k-1}, v(x^{k-1})) + t_2 |\psi(x^k, d^{k-1}) - \psi(x^{k-1}, d^{k-1})|}, \quad (2.1)$$

$$\beta_k^{MVCD} := \frac{-t_1 \psi(x^k, v(x^k))}{-\psi(x^{k-1}, d^{k-1}) + t_2 |\psi(x^k, d^{k-1}) - \psi(x^{k-1}, d^{k-1})|}, \quad (2.2)$$

and

$$\beta_k^{MVPRP} := \frac{-\psi(x^k, v(x^k)) + \psi(x^{k-1}, v(x^k))}{-\psi(x^{k-1}, v(x^{k-1})) + u |\psi(x^k, d^{k-1}) - \psi(x^{k-1}, d^{k-1})|}. \quad (2.3)$$

**Remark 2.1.** MVFR is a modified version of the FR method, whereas MVCD and MVPRP extend respectively the CD and PRP methods from the scalar case to the vector case. From (2.1)-(2.3), the three proposed parameters differ fundamentally from those in [39] since the original three parameters in [39] are reduced to two parameters  $t_1$  and  $t_2$ , and the term added in the denominator is modified by replacing  $|\psi(x^k, d^{k-1})|$  with  $|\psi(x^k, d^{k-1}) - \psi(x^{k-1}, d^{k-1})|$ .

The following lemmas demonstrate that the search directions  $d^k$  defined by (1.10) with the MVFR (2.1) and MVCD (2.2) conjugate parameters satisfy the sufficient descent condition (1.6) with  $c = 1 - \frac{t_1}{t_2}$ .

**Lemma 2.1.** Let  $\beta_k^{MVFR}$  be defined by (2.1) and  $d^k$  defined by (1.10) be  $K$ -descent direction of  $F$  at  $x^k$  for all  $k$ . Then  $d^k$  satisfy the sufficient descent condition (1.6) at  $x^k$  with  $c = 1 - \frac{t_1}{t_2}$ , i.e.,

$$\psi(x^k, d^k) \leq \left(1 - \frac{t_1}{t_2}\right) \psi(x^k, v(x^k)), \quad \forall k \geq 1. \quad (2.4)$$

*Proof.* From (1.7), we have  $\psi(x^k, v(x^k)) \leq 0$  for all  $k$  and so,  $-t_1 \psi(x^k, v(x^k)) \geq 0$  for all  $k$ . If  $k = 1$ , we get  $d^1 = v(x^1)$ . Then  $\psi(x^1, d^1) \leq \left(1 - \frac{t_1}{t_2}\right) \psi(x^1, v(x^1))$  due to  $t_2 > t_1 > 0$ .

For  $k \geq 2$ , assume that

$$\psi(x^{k-1}, d^{k-1}) \leq \left(1 - \frac{t_1}{t_2}\right) \psi(x^{k-1}, v(x^{k-1})) \leq 0. \quad (2.5)$$

Note that  $d^k = v(x^k) + \beta_k^{MVFR} d^{k-1}$ . By Lemma 1.2 and  $\beta_k^{MVFR} \geq 0$ , we have

$$\psi(x^k, d^k) \leq \psi(x^k, v(x^k)) + \beta_k^{MVFR} \psi(x^k, d^{k-1}). \quad (2.6)$$

Since  $d^k$  is  $K$ -descent direction of  $F$  at  $x^k$  for each  $k$ , then  $x^k$  is not  $K$ -critical point for each  $k$ . It follows from Lemma 1.3 (ii) that  $\psi(x^k, v(x^k)) < 0$  for each  $k$ . Then

$$-\psi(x^{k-1}, v(x^{k-1})) + t_2 |\psi(x^k, d^{k-1}) - \psi(x^{k-1}, d^{k-1})| > 0.$$

Hence  $\beta_k^{MVFR}$  is well-defined and  $\beta_k^{MVFR} \geq 0$ .

Next, we analyze the term  $\psi(x^k, d^{k-1})$ . If  $\psi(x^k, d^{k-1}) \leq 0$ , then  $\psi(x^k, d^k) \leq \psi(x^k, v(x^k)) \leq 0$ . It therefore follows from  $\psi(x^k, v(x^k)) < 0$  and  $t_2 > t_1 > 0$  that

$$\psi(x^k, d^k) \leq \left(1 - \frac{t_1}{t_2}\right) \psi(x^k, v(x^k)).$$



If  $\psi(x^k, d^{k-1}) > 0$ , using (2.1) and (2.6) yields

$$\begin{aligned}
 & \psi(x^k, d^k) \\
 & \leq \psi(x^k, v(x^k)) + \frac{-t_1 \psi(x^k, v(x^k))}{-\psi(x^{k-1}, v(x^{k-1})) + t_2 |\psi(x^k, d^{k-1}) - \psi(x^{k-1}, d^{k-1})|} \psi(x^k, d^{k-1}) \\
 & = \psi(x^k, v(x^k)) + \frac{-t_1 \psi(x^k, v(x^k))}{-\psi(x^{k-1}, v(x^{k-1})) + t_2 (\psi(x^k, d^{k-1}) - \psi(x^{k-1}, d^{k-1}))} \psi(x^k, d^{k-1}) \\
 & \leq \psi(x^k, v(x^k)) + \frac{-t_1 \psi(x^k, v(x^k))}{t_2 \psi(x^k, d^{k-1})} \psi(x^k, d^{k-1}) \\
 & = \left(1 - \frac{t_1}{t_2}\right) \psi(x^k, v(x^k)),
 \end{aligned}$$

where the second equality is derived from the assumption  $\psi(x^k, d^{k-1}) > 0$  and (2.5), and the second inequality results from  $\psi(x^{k-1}, v(x^{k-1})) \leq 0$  and (2.5). These together with (2.5) show that (2.4) hold. The proof is completed.  $\square$

**Lemma 2.2.** Let  $\beta_k^{MVCD}$  be defined by (2.2) and  $d^k$  defined by (1.10) be  $K$ -descent direction of  $F$  at  $x^k$  for all  $k$ . Then  $d^k$  satisfies the sufficient descent condition (1.6) at  $x^k$  with  $c = 1 - \frac{t_1}{t_2}$ , i.e.,

$$\psi(x^k, d^k) \leq \left(1 - \frac{t_1}{t_2}\right) \psi(x^k, v(x^k)), \quad \forall k \geq 1. \quad (2.7)$$

*Proof.* The proof is similar to that of Lemma 2.1 and so is omitted. Thereby completing the proof.  $\square$

To conclude this section, we present two novel hybrid conjugate gradient methods for solving problem (1.2) by the combination of MVFR and MVCD conjugate parameters with MVPRP conjugate parameter. This combination leverages the global convergence properties of MVFR method and the adaptive restart capabilities of MVPRP method, aiming to balance stability and efficiency. Besides, another combination leverages that MVCD ensures the descent direction consistency, and that MVPRP mitigates stagnation risks in non-convex landscapes and improve the convergence speed.

**Algorithm 2.1. [The first new hybrid conjugate gradient method (Nhyb1)]**

Step 1. Choose starting point  $x^1 \in \mathbb{R}^n$ , and  $0 < t_1 < \min\left\{t_2, \frac{\sqrt{6}}{3}\right\}$ ,  $u > 0$ ,  $0 < \rho < \sigma < 1$ ,  $e \in \text{int}K$  as in (1.11). Set  $k := 1$ .

Step 2. Compute  $v(x^k)$  and  $\theta(x^k)$  by (1.7) and (1.8), respectively. If  $\theta(x^k) = 0$ , then STOP.

Step 3. Compute the search direction  $d^k$  by (1.10), where

$$\beta_k := \max\{0, \min\{\beta_k^{MVFR}, \beta_k^{MVPRP}\}\}. \quad (2.8)$$

Step 4. Compute a step size  $\alpha_k > 0$  by (1.13).

Step 5. Set  $x^{k+1} := x^k + \alpha_k d^k$ , for  $k \leftarrow k + 1$ , and go to Step 2.

**Algorithm 2.2.** [The second new hybrid conjugate gradient method (Nhyb2)]

Step 1. Choose starting point  $x^1 \in \mathbb{R}^n$ , and  $0 < t_1 < \min \left\{ \frac{-\omega + \sqrt{\omega^2 + \frac{4\sqrt{6}}{3}\omega}}{2}, t_2 \right\}$ ,  $\omega = t_2 - t_1$ ,  $u > 0$ ,  $0 < \rho < \sigma < 1$ ,  $e \in \text{int}K$  as in (1.11). Set  $k := 1$ .

Step 2. Compute  $v(x^k)$  and  $\theta(x^k)$  by (1.7) and (1.8), respectively. If  $\theta(x^k) = 0$ , then STOP.

Step 3. Compute the search direction  $d^k$  by (1.10), where

$$\beta_k := \max \{0, \min \{\beta_k^{MVCD}, \beta_k^{MVPRP}\}\}. \quad (2.9)$$

Step 4. Compute a step size  $\alpha_k > 0$  by (1.13).

Step 5. Set  $x^{k+1} := x^k + \alpha_k d^k$ , for  $k \leftarrow k + 1$ , and go to Step 2.

**Remark 2.2.** The only difference between Algorithms 2.1 and 2.2 lies in the conjugate parameter  $\beta_k$ . The conjugate parameter  $\beta_k$  in Algorithms 2.1 and 2.2 is nonnegative, which involves the truncation technique. In Algorithms 2.1 and 2.2, the stopping condition at Step 2 is  $\theta(x^k) = 0$ , which implies that  $x^k$  is a  $K$ -critical point by Lemma 1.1.

### 3 Convergence analysis

In this section, we study the global convergence of the two proposed hybrid conjugate gradient methods under the strong Wolfe line search conditions and general assumptions.

In the convergence analysis, we will need the following assumptions.

**Assumption 3.1.** (A1) Suppose that the cone  $K$  is finitely generated and there exists an open set  $\Omega$  such that  $\mathcal{L} := \{x \mid F(x) \preceq_K F(x^1)\} \subset \Omega$ , where  $x^1 \in \mathbb{R}^n$ .

(A2) The Jacobian  $JF$  is Lipschitz continuous on  $\mathcal{L}$ , i.e., there exists an  $L > 0$  such that satisfies  $\|JF(x^1) - JF(x^2)\| \leq L\|x^1 - x^2\|$  for all  $x^1, x^2 \in \mathcal{L}$ .

(A3) All monotone nonincreasing sequences in  $F(\mathcal{L})$  are bounded below, i.e., a sequence  $\{F^k\}_{k \in \mathbb{N}} \subset F(\mathcal{L})$  and  $F^{k+1} \preceq_K F^k$  hold for all  $k$ , then there exists  $\bar{F} \in \mathbb{R}^m$  for which  $\bar{F} \preceq_K F^k$ .

**Assumption 3.2.** The level set  $\mathcal{L} := \{x \mid F(x) \preceq_K F(x^1)\} \subset \Omega$  is bounded.

**Remark 3.1.** If  $\{x^k\}$  is generated by Algorithm 2.1 or Algorithm 2.2, then the Assumption 3.2 that  $\{x^k\}$  is bounded. It follows from the continuity of  $JF(\cdot)$  and  $v(\cdot)$  that there exist  $\bar{\gamma} > 0$  and  $\bar{\zeta} > 0$  such that

$$\|v(x^k)\| \leq \bar{\gamma}, \quad \forall k \geq 1, \quad (3.1)$$

and

$$\|JF(x^k)\| \leq \bar{\zeta}, \quad \forall k \geq 1. \quad (3.2)$$

By the definition of  $\psi$ , it can be inferred that

$$-\bar{\zeta}\bar{\gamma} \leq -\|JF(x^k)\| \|v(x^k)\| \leq \langle JF(x^k) v(x^k), w \rangle \leq \psi(x^k, v(x^k)) < 0, \quad (3.3)$$

holds for each  $k \geq 1$ , where the second inequality follows from  $\|w\| = 1$ .

The following Property (\*) and the Zoutendijk condition will be useful to the convergence analysis of conjugate gradient algorithms.

**Property (\*)** [32, Property (\*)] Let  $\{x^k\}$  be generated by a conjugate gradient method (0.1). Assume that there exists  $\gamma > 0$  such that

$$0 < \gamma \leq \|v(x^k)\|, \forall k \geq 1. \quad (3.4)$$

Under this assumption, we say that the method (0.1) has Property (\*) if there exist constants  $p > 1$  and  $q > 0$  such that for each  $k$ ,

$$|\beta_k| \leq p, \quad (3.5)$$

and

$$\|s^{k-1}\| \leq q \Rightarrow |\beta_k| \leq \frac{1}{2p}, \quad (3.6)$$

where  $s^{k-1} = x^k - x^{k-1}$ .

**Lemma 3.1.** [32, Proposition 3.3] Assume that Assumption 3.1 holds. Let  $\{x^k\}$  be generated by the iteration (0.1), where  $d^k$  is a  $K$ -descent direction of  $F$  at  $x^k$  and  $\alpha^k$  satisfies the standard Wolfe conditions (1.12). Then

$$\sum_{k=1}^{\infty} \frac{\psi^2(x^k, d^k)}{\|d^k\|^2} < \infty. \quad (3.7)$$

The following lemma states that when the standard assumptions are met, a conjugate gradient algorithm converges if it satisfies Property (\*).

**Lemma 3.2.** [32, Theorem 5.10] Suppose that Assumption 3.1(A1), (A2) and Assumption 3.2 hold. Consider a CG method where  $\beta_k \geq 0$ ,  $d^k$  is a  $K$ -descent direction of  $F$  at  $x^k$ , and the step size  $\alpha_k$  satisfies the strong Wolfe conditions (1.13). If the method satisfies Property (\*), then  $\liminf_{k \rightarrow \infty} \|v(x^k)\| = 0$ .

Theorem 3.1 establishes the . Our proof primarily employ a proof-by-contradiction approach while incorporating some established properties.

Now we show the global convergence of Algorithm 2.1 under the commonly used assumptions.

**Theorem 3.1.** Assume that Assumptions 3.1 and 3.2 hold. Let  $\{x^k\}$  be generated by Algorithm 2.1, and  $d^k$  be a  $K$ -descent direction of  $F$  at  $x^k$ . Then

$$\liminf_{k \rightarrow \infty} \|v(x^k)\| = 0. \quad (3.8)$$

*Proof.* Since  $\beta_k := \max\{0, \min\{\beta_k^{MVFR}, \beta_k^{MVPRP}\}\}$ ,  $\beta_k$  can only have three possible scenarios. Let  $I_1 := \{k \in \mathbb{N} \mid \beta_k = 0\}$ ,  $I_2 := \{k \in \mathbb{N} \mid \beta_k = \beta_k^{MVFR}\}$ , and  $I_3 := \{k \in \mathbb{N} \mid \beta_k = \beta_k^{MVPRP}\}$ . If  $\{x^k\}$  is a finite sequence, then Step 2 of Algorithm 2.1 implies (3.8).

Let us consider the case that  $\{x^k\}$  is a infinite sequence. In this case,  $I_1, I_2$ , and  $I_3$  at least one must be an infinite set. If  $I_1$  is an infinite set, then  $\beta_k = 0$  and so,  $d^k = v(x^k)$  for all  $k \in I_1$  from (1.10). So, Algorithm 2.1 becomes the steepest descent method. Therefore, the desired result (3.8) can be obtained from [41, Theorem 4.2].

If  $I_2$  is an infinite set, then  $\beta_k = \beta_k^{MVFR}$  and  $0 < \beta_k^{MVFR} \leq \beta_k^{MVPRP}$  for all  $k \in I_2$ . In this case, the proof will proceed by contradiction. Suppose that there exists  $\gamma > 0$  such that

$$\|v(x^k)\| \geq \gamma, \forall k \in I_2.$$

Then  $\psi(x^k, v(x^k)) < 0$  for all  $k \in I_2$ . For  $k \in I_2$ , substituting (2.1) into (1.10), we have

$$d^k = v(x^k) + \beta_k^{MVFR} d^{k-1}. \quad (3.9)$$

Then, we obtain

$$\|d^k\| = \|v(x^k) + \beta_k^{MVFR} d^{k-1}\| \leq \|v(x^k)\| + \beta_k^{MVFR} \|d^{k-1}\|. \quad (3.10)$$

Using Lemma 1.4 with  $p = \|v(x^k)\|$ ,  $q = \beta_k^{MVFR} \|d^{k-1}\|$  and  $t = 1$  in (3.10) and substituting (2.1) into (3.10), we have

$$\begin{aligned} & \|d^k\|^2 \\ & \leq 3\|v(x^k)\|^2 + \frac{3}{2} (\beta_k^{MVFR})^2 \|d^{k-1}\|^2 \\ & = 3\|v(x^k)\|^2 + \frac{3}{2} \left( \frac{-t_1 \psi(x^k, v(x^k))}{-\psi(x^{k-1}, v(x^{k-1})) + t_2 |\psi(x^k, d^{k-1}) - \psi(x^{k-1}, d^{k-1})|} \right)^2 \|d^{k-1}\|^2 \\ & = 3\|v(x^k)\|^2 + \frac{3}{2} \frac{t_1^2 \psi^2(x^k, v(x^k))}{(-\psi(x^{k-1}, v(x^{k-1})) + t_2 |\psi(x^k, d^{k-1}) - \psi(x^{k-1}, d^{k-1})|)^2} \|d^{k-1}\|^2 \\ & \leq 3\|v(x^k)\|^2 + \frac{3}{2} \frac{t_1^2 \psi^2(x^k, v(x^k))}{\psi^2(x^{k-1}, v(x^{k-1}))} \|d^{k-1}\|^2. \end{aligned}$$

Dividing by  $\psi^2(x^k, v(x^k))$  in the above inequality, we get

$$\frac{\|d^k\|^2}{\psi^2(x^k, v(x^k))} \leq \frac{3\|v(x^k)\|^2}{\psi^2(x^k, v(x^k))} + \frac{3}{2} t_1^2 \frac{\|d^{k-1}\|^2}{\psi^2(x^{k-1}, v(x^{k-1}))}.$$

Applying Lemma 1.3(ii), we have

$$\frac{\|d^k\|^2}{\psi^2(x^k, v(x^k))} \leq \frac{12}{\|v(x^k)\|^2} + \frac{3}{2} t_1^2 \frac{\|d^{k-1}\|^2}{\psi^2(x^{k-1}, v(x^{k-1}))}.$$

Again, from  $\|v(x^k)\| \geq \gamma > 0$  and  $0 < t_1 < \min\left\{t_2, \frac{\sqrt{6}}{3}\right\}$ , we get

$$\frac{\|d^k\|^2}{\psi^2(x^k, v(x^k))} \leq \frac{12}{\gamma^2} + \frac{\|d^{k-1}\|^2}{\psi^2(x^{k-1}, v(x^{k-1}))}. \quad (3.11)$$

If  $k-1 \in I_1$ , we have  $d^{k-1} = v(x^{k-1})$ . Then, one has

$$\frac{\|d^{k-1}\|^2}{\psi^2(x^{k-1}, v(x^{k-1}))} = \frac{\|v(x^{k-1})\|^2}{\psi^2(x^{k-1}, v(x^{k-1}))} < \frac{4}{\|v(x^{k-1})\|^2} \leq \frac{4}{\gamma^2}.$$

Consequently, we obtain

$$\frac{\|d^{k-1}\|^2}{\psi^2(x^{k-1}, v(x^{k-1}))} \leq \frac{4}{\gamma^2} < \frac{12}{\gamma^2} + \frac{\|d^{k-2}\|^2}{\psi^2(x^{k-2}, v(x^{k-2}))}. \quad (3.12)$$

If  $k-1 \in I_3$ , then  $\beta_{k-1} = \beta_{k-1}^{MVPRP}$ ,  $0 < \beta_{k-1}^{MVPRP} \leq \beta_{k-1}^{MVFR}$  and

$$d^{k-1} = v(x^{k-1}) + \beta_{k-1}^{MVPRP} d^{k-2}.$$

Note that

$$\begin{aligned}
 & \|d^{k-1}\|^2 \\
 & \leq 3 \|v(x^{k-1})\|^2 + \frac{3}{2} (\beta_{k-1}^{MVPRP})^2 \|d^{k-2}\|^2 \\
 & \leq 3 \|v(x^{k-1})\|^2 + \frac{3}{2} (\beta_{k-1}^{MVFR})^2 \|d^{k-2}\|^2 \\
 & \leq 3 \|v(x^{k-1})\|^2 + \frac{3 t_1^2 \psi^2(x^{k-1}, v(x^{k-1}))}{2 \psi^2(x^{k-2}, v(x^{k-2}))} \|d^{k-2}\|^2.
 \end{aligned}$$

Similar to the case  $k \in I_2$ , we obtain

$$\frac{\|d^{k-1}\|^2}{\psi^2(x^{k-1}, v(x^{k-1}))} \leq \frac{12}{\gamma^2} + \frac{\|d^{k-2}\|^2}{\psi^2(x^{k-2}, v(x^{k-2}))}. \quad (3.13)$$

Consequently, (3.13) holds for all  $k \in \mathbb{N}$ . Repeating (3.11), we have

$$\frac{\|d^k\|^2}{\psi^2(x^k, v(x^k))} \leq \frac{12}{\gamma^2} + \frac{\|d^{k-1}\|^2}{\psi^2(x^{k-1}, v(x^{k-1}))} \leq \dots \leq \frac{12(k-1)}{\gamma^2} + \frac{\|d^1\|^2}{\psi^2(x^1, v(x^1))} \leq \frac{4(3k-2)}{\gamma^2}.$$

Moreover, we obtain

$$\frac{\psi^2(x^k, v(x^k))}{\|d^k\|^2} \geq \frac{\gamma^2}{4} \frac{1}{3k-2}. \quad (3.14)$$

Utilizing sufficient descent conditions of MVFR (2.4) and (3.14), we have

$$\psi^2(x^k, d^k) \geq \left(1 - \frac{t_1}{t_2}\right)^2 \psi^2(x^k, v(x^k))$$

and

$$\begin{aligned}
 \sum_{k=1}^{\infty} \frac{\psi^2(x^k, d^k)}{\|d^k\|^2} & \geq \sum_{k=1}^{\infty} \frac{\left(1 - \frac{t_1}{t_2}\right)^2 \psi^2(x^k, v(x^k))}{\|d^k\|^2} \\
 & \geq \left(1 - \frac{t_1}{t_2}\right)^2 \sum_{k=1}^{\infty} \frac{\gamma^2}{4} \frac{1}{3k-2} \\
 & = \left(1 - \frac{t_1}{t_2}\right)^2 \frac{\gamma^2}{4} \sum_{k=1}^{\infty} \frac{1}{3k-2} \\
 & = \infty,
 \end{aligned}$$

which contradicts the Zoutendijk condition (3.7). Therefore, (3.8) holds for  $k \in I_2$ .

If  $I_3$  is an infinite set, then  $\beta_k = \beta_k^{MVPRP}$  and  $0 < \beta_k^{MVPRP} \leq \beta_k^{MVFR}$ . Since  $\beta_k$  satisfies the non-negativity condition and  $d^k$  is a  $K$ -descent direction of  $F$  at  $x^k$ , we next show the desired result (3.8) by Lemma 3.2. For this, we only need to prove that the Property (\*) holds under the condition  $\|x^k - x^{k-1}\| = \|s^{k-1}\| \leq q$ . Suppose that there exists  $\gamma > 0$  such that  $0 < \gamma \leq \|v(x^k)\|$  for all  $k \in I_3$ . From (3.3) and Lemma 1.3(ii), we can get

$$\frac{\gamma^2}{2} < -\psi(x^k, v(x^k)) \leq \bar{\xi} \bar{\gamma}. \quad (3.15)$$

From  $\|x^k - x^{k-1}\| = \|s^{k-1}\| \leq q$ , assumption (A2) and Lemma 1.2, it follows that

$$|-\psi(x^k, v(x^k)) + \psi(x^{k-1}, v(x^k))| \leq Lq\tilde{\gamma}. \quad (3.16)$$

Note that

$$\begin{aligned} \beta_k^{MVPRP} &= \frac{-\psi(x^k, v(x^k)) + \psi(x^{k-1}, v(x^k))}{-\psi(x^{k-1}, v(x^{k-1})) + u|\psi(x^k, d^{k-1}) - \psi(x^{k-1}, d^{k-1})|} \\ &\leq \frac{-\psi(x^k, v(x^k)) + |\psi(x^{k-1}, v(x^k))|}{-\psi(x^{k-1}, v(x^{k-1})) + u|\psi(x^k, d^{k-1}) - \psi(x^{k-1}, d^{k-1})|} \\ &\leq \frac{-\psi(x^k, v(x^k)) + |\psi(x^{k-1}, v(x^k))|}{-\psi(x^{k-1}, v(x^{k-1}))}. \end{aligned}$$

This together with (3.1), (3.2) and (3.15) yields that

$$|\beta_k^{MVPRP}| \leq \frac{2(\tilde{\xi}\tilde{\gamma} + \tilde{\xi}\tilde{\gamma})}{\gamma^2} = \frac{4\tilde{\xi}\tilde{\gamma}}{\gamma^2}. \quad (3.17)$$

Besides, from (3.16) we have

$$\begin{aligned} |\beta_k^{MVPRP}| &= \frac{-\psi(x^k, v(x^k)) + \psi(x^{k-1}, v(x^k))}{-\psi(x^{k-1}, v(x^{k-1})) + u|\psi(x^k, d^{k-1}) - \psi(x^{k-1}, d^{k-1})|} \\ &\leq \frac{-\psi(x^k, v(x^k)) + \psi(x^{k-1}, v(x^k))}{-\psi(x^{k-1}, v(x^{k-1}))} \\ &\leq \frac{2Lq\tilde{\gamma}}{\gamma^2}. \end{aligned} \quad (3.18)$$

Consequently, we conclude from (3.17) and (3.18) that if  $\frac{4\tilde{\xi}\tilde{\gamma}}{\gamma^2} \geq 1$ , there exist  $p := \frac{4\tilde{\xi}\tilde{\gamma}}{\gamma^2}$  and  $q := \frac{\gamma^4}{16L\tilde{\gamma}^2\tilde{\xi}}$  such that Property (\*) holds; if  $0 < \frac{4\tilde{\xi}\tilde{\gamma}}{\gamma^2} < 1$ , there exist  $p := \frac{4\tilde{\xi}\tilde{\gamma}}{\gamma^2} + 1$  and  $q := \frac{\gamma^4}{4L\tilde{\gamma}(4\tilde{\xi}\tilde{\gamma} + \gamma^2)}$  such that Property (\*) holds. Thus the desired result (3.8) is true by Lemma 3.2 when  $k \in I_3$ . Altogether, (3.8) holds for all  $k \in \mathbb{N}$ . The proof is completed.  $\square$

**Theorem 3.2.** Assume that Assumptions 3.1 and 3.2 hold. Let  $\{x^k\}$  be generated by Algorithm 2.2, and  $d^k$  be a  $K$ -descent direction of  $F$  at  $x^k$ . Then

$$\liminf_{k \rightarrow \infty} \|v(x^k)\| = 0. \quad (3.19)$$

*Proof.* The proof is similar to that of Theorem 3.1 and so is omitted.  $\square$

**Remark 3.2.** The proof of Theorem 3.1 demonstrates that if the conjugate parameter  $\beta_k$  takes either  $\beta_k^{MVFR}$  or  $\max\{0, \beta_k^{MVPRP}\}$ , then the conjugate gradient algorithm with the conjugate parameter  $\beta_k$  admits the global convergence. Notably, the MVFR method does not involve the conditions (1.13), and requires only the standard Wolfe line search, whereas the truncated MVPRP method necessitates the strong Wolfe line search to ensure convergence. Besides, the proof of Theorem 3.2 shows that if the conjugate parameter  $\beta_k$  takes  $\beta_k^{MVCD}$ , the conjugate gradient algorithm with standard Wolfe line search also possesses the global convergence.

## 4 Numerical experiments

In this section, we present numerical experiments to demonstrate the effectiveness of the two new hybrid methods we proposed, namely Nhyb1 and Nhyb2. All algorithms are implemented in MATLAB R2024a and run on a Intel(R) Core(TM) i5-8300H, CPU 2.30GHz with 8GB RAM. We evaluate the methods on 27 benchmark problems from the multiobjective optimization literature. In this scenario, we have  $K = \mathbb{R}_+^m$ ,  $C$  represents the canonical basis of  $\mathbb{R}^m$ , and  $e = [1, \dots, 1]^T \in \mathbb{R}^m$ . Table 1 summarizes the test problems: columns 1-2 represent the problem name and the corresponding reference. Symbols  $\sharp$  and  $\dagger$  denote modified versions of the DD1 problem and adaptations from single-objective problems, respectively (see [43] for details); columns 3-4 indicate the number of objectives ( $m$ ) and variables ( $n$ ); columns 5-6 report the bounds  $lb$  and  $ub$  defining the initial point domain  $\{x \in \mathbb{R}^n | lb \leq x \leq ub\}$ . Notably, the box constraints are used solely to generate initial point; the algorithm itself operates without explicit constraint handling.

Below, we summarize the evaluated methods, including their initial parameter settings, alongside our proposed methods and benchmark algorithms:

1. Nhyb1: the first hybrid conjugate gradient method we propose, i.e., Algorithm ?? with  $u = 1.2$ ,  $t_1 = 0.5$ ,  $t_2 = 6.6$ .
2. Nhyb2: the second hybrid conjugate gradient method we propose, i.e., Algorithm ?? with  $u = 1.2$ ,  $t_1 = 0.5$ ,  $t_2 = 6.6$ ;
3. MVFR: the conjugate gradient algorithm with  $\beta_k = \beta_k^{MVFR}$  with  $t_1 = 0.5$ ,  $t_2 = 6.6$ .
4. MVCD: the conjugate gradient algorithm with  $\beta_k = \beta_k^{MVCD}$  with  $t_1 = 0.5$ ,  $t_2 = 6.6$ .
5. MVPRP+: the conjugate gradient algorithm with  $\beta_k = \max\{\beta_k^{MVPRP}, 0\}$  with  $u = 1.2$ .
6. Hyb: A hybrid conjugate gradient algorithm with superior performance proposed in [39]. i.e., the CG algorithm with  $\beta_k := \max\{0, \min\{\beta_k^{VPRP}, \beta_k^{VFR}\}\}$ , where

$$\beta_k^{VPRP} = \frac{-\psi(x^k, v(x^k)) + \psi(x^{k-1}, v(x^k))}{u_1 |\psi(x^k, d^{k-1})| - \psi(x^{k-1}, v(x^{k-1}))},$$

$$\beta_k^{VFR} = \frac{-\lambda_1 \psi(x^k, v(x^k))}{\lambda_2 |\psi(x^k, d^{k-1})| - \lambda_3 \psi(x^{k-1}, v(x^{k-1}))}.$$

The values of  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  and  $u_1$  correspond to the optimal parameters selected from [39], i.e.,  $\lambda_1 = 0.1$ ,  $\lambda_2 = 0.8$ ,  $\lambda_3 = 0.12$ , and  $u_1 = 1.1$ ;

7. PRP+: the nonnegative Polak-Ribière-Polyak conjugate gradient method proposed in [32], i.e., the CG algorithm with  $\beta_k = \max\{\beta_k^{PRP}, 0\}$  under strong Wolfe conditions.
8. FR: the FR conjugate gradient method proposed in [32], i.e., the CG algorithm with  $\beta_k = 0.98\beta_k^{FR}$  under strong Wolfe conditions.
9. CD: the CD conjugate gradient method proposed in [32], i.e., the CG algorithm with  $\beta_k = 0.99(1 - \sigma)\beta_k^{CD}$  under strong Wolfe conditions.



表 1: Test problems

Problem	References	m	n	$lb^T$	$ub^T$
AP3	[44]	2	2	$(-100, -100)$	$(100, 100)$
BK1	[45]	2	2	$(-5, -5)$	$(10, 10)$
DD1 <sup>#</sup>	[46]	2	5	$(-20, \dots, -20)$	$(20, \dots, 20)$
DGO1	[45]	2	1	$-10$	$13$
DGO2	[45]	2	1	$-10$	$13$
FF1	[45]	2	2	$(-1, -1)$	$(1, 1)$
Hi11	[47]	2	2	$(0, 0)$	$(1, 1)$
JOS1(n=100)	[48]	2	100	$(0, \dots, 0)$	$(1, \dots, 1)$
Lov5	[49]	2	3	$(-2, -2, -2)$	$(2, 2, 2)$
MLF2	[45]	2	2	$(-100, -100)$	$(100, 100)$
MOP1	[45]	2	1	$-10^5$	$10^5$
MOP2	[45]	2	2	$(-4, -4)$	$(4, 4)$
PNR	[50]	2	2	$(-2, -2)$	$(2, 2)$
SK1	[45]	2	1	$-100$	$100$
SK2	[45]	2	4	$(-10, -10, -10, -10)$	$(10, 10, 10, 10)$
SSFYY2	[45]	2	1	$-100$	$100$
TE1	[51]	2	2	$(-1, 1)$	$(1, 1)$
TE7	[51]	3	3	$(0, 0, 0)$	$(30, 30, 30)$
TE2	[51]	2	2	$(-2, -2)$	$(2, 2)$
Toi4 <sup>†</sup>	[52]	2	4	$(-2, -2, -2, -2)$	$(5, 5, 5, 5)$
VU1	[45]	2	2	$(-3, -3)$	$(3, 3)$
FDS(n=10)	[53]	3	10	$(-2, \dots, -2)$	$(2, \dots, 2)$
FDS(n=100)	[53]	3	100	$(-2, \dots, -2)$	$(2, \dots, 2)$
IKK1	[45]	3	2	$(-50, -50)$	$(50, 50)$
SLCDT2	[54]	3	10	$(-100, \dots, -100)$	$(100, \dots, 100)$
TE8-2	[51]	3	30	$(0, \dots, 0)$	$(1, \dots, 1)$
TE8-3	[51]	3	50	$(0, \dots, 0)$	$(1, \dots, 1)$

In our numerical experiments, all nine conjugate gradient algorithms compute the step size  $\alpha > 0$  via the strong Wolfe line search (1.13) with parameters  $\rho = 10^{-4}$  and  $\sigma = 0.1$ . To calculate the steepest descent direction  $v(x)$  and the optimal value  $\theta(x)$ , we solve the subproblem (1.9) using MATLAB's built-in quadratic programming solver quadprog. Finally, we set two termination conditions for the algorithms, one being  $\theta(x) \geq -5 \times eps^{\frac{1}{2}}$  ( $eps = 2^{-52} \approx 2.22 \times 10^{-16}$ ) and the other is the count of iterations exceeds 5000.

It is worth noting that, for numerical stability considerations, we apply a scaling technique to problem (1.2) with  $K = \mathbb{R}_+^m$ , as proposed in references[33, 55]. Specifically, we solve the scaled problem:

$$\min_{x \in \mathbb{R}^n} (\gamma_1 F_1(x), \dots, \gamma_m F_m(x)), \quad (4.1)$$

where the scaling factors are defined as  $\gamma_j = \frac{1}{\max(1, \|\nabla F_j(x^0)\|_\infty)}$ ,  $j = 1, \dots, m$ , with  $x_0 \in \mathbb{R}^n$  being the starting point. Crucially, this scaling preserves the set of  $K$ -critical Pareto points of the original problem (1.2)

with  $K = \mathbb{R}_+^m$ , ensuring equivalence between the two formulations.

In our experiments, each method is executed 100 times with starting points randomly generated within the box constraints (last two columns of Table 1) using a uniform distribution. A run is deemed successful if an approximate critical point is identified, irrespective of the objective function value.

To enhance clarity in comparing the performance of the examined methods, we utilize the performance profiles suggested by Dolan and Moré [56]:

$$\rho(\tau) = \frac{1}{|P|} |\{p \in P : r_{p,z} \leq \tau\}|, \quad r_{p,z} = \frac{t_{p,z}}{\min\{t_{p,z}, z \in Z\}}.$$

In this context, let  $P$  denotes the set of problems and  $Z$  refers to the set of algorithms, where  $t_{p,z}$  represents the performance metric when applying algorithm  $z \in Z$  to problem  $p \in P$ . In the performance profile plots,  $\rho(1)$  indicates the proportion of problems for which algorithm  $z$  achieves the best efficiency among all compared methods. Figure 1 and Figure 2 present performance profiles for iteration counts (IT), function evaluations (NF), gradient evaluations (NG), steepest descent direction evaluations (NV), and CPU time in seconds across the nine algorithms. As observed from Figure 1, in IT, NF, NG, NV, both Nhyb1 and Nhyb2 align closely, even overlapping entirely. Nhyb1 and Nhyb2 are slightly better than Hyb, and MVFR, MVCD, MVPRP+ are comparable to Hyb. However, in terms of CPU time, it is obvious that the five methods we proposed are better than Hyb, and it can be seen that Nhyb2 and MVCD perform best. It is evident from Figure 2 that the performance of Nhyb1 and Nhyb2 is almost identical and significantly better than FR, CD and PRP+ in terms of IT, NF and NG. In terms of NV and CPU time, it can be seen that Nhyb2 is better than Nhyb1, and both Nhyb1 and Nhyb2 are also significantly better than FR, CD, PRP+. From the summary of Figure 1 and Figure 2, it can be concluded that the methods we proposed have good performance and has certain advantages compared with the existing CG algorithms.

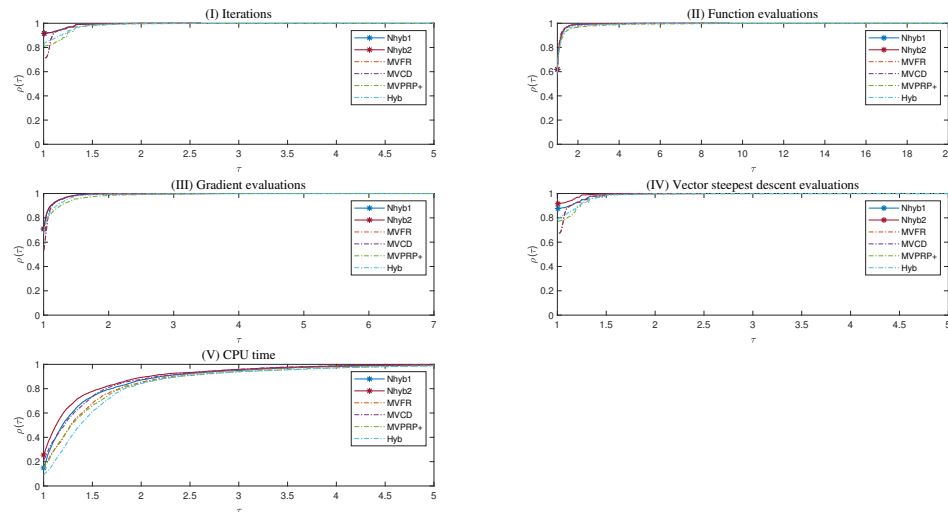


图 1: Performance profiles of Nhyb1, Nhyb2, MVFR, MVCD, MVPRP+, Hyb algorithms with 100 starting points for test problem

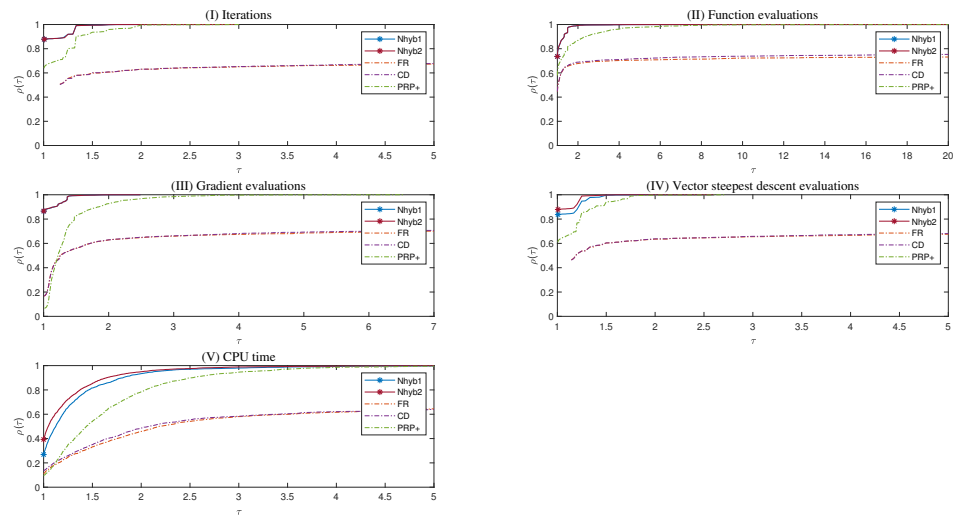


图 2: Performance profiles of Nhyb1, Nhyb2, FR, CD, PRP+ with 100 starting points for test problem

## 5 Conclusions

By directly modifying three existing conjugate parameters in vector optimization, we propose three novel conjugate gradient methods. We demonstrate that the search directions of the MVFR and MVCD methods satisfy the sufficient descent condition without requiring any line search strategy. Under mild assumptions, global convergence of the MVFR and MVCD methods is established using standard Wolfe line search conditions without convexity assumptions. With the incorporation of truncation techniques and under strong Wolfe line search conditions, we also prove global convergence for the MVPRP method. Furthermore, to enhance overall numerical performance, we develop two new hybrid conjugate gradient methods by strategically combining the three modified conjugate parameters. These hybrid methods are shown to achieve global convergence under general assumptions using strong Wolfe line search conditions, again without requiring convexity. Numerical experiments demonstrate the effectiveness of all proposed methods, with particular emphasis on the superior robustness of the two hybrid approaches.

**Acknowledgements.** The authors are deeply grateful to the associate editor and referees for the constructive suggestions, which is helpful to improve the paper greatly. This article was supported by the NSFC (12071379, 12471299), the Natural Science Foundation of Chongqing (cstc2022ycjh-bgzxm0097), and the Southwest University Graduate Research Innovation Program (No. SWUS24064).

## 参考文献 (References)

- [1] Chen, Y., Kuang, K., Yan, X.: *A modified PRP-type conjugate gradient algorithm with complexity analysis and its application to image restoration problems*. J. Comput. Appl. Math. 427, 115105 (2023)
- [2] Wu, X.D., Ye, X.M., Han, D.L.: *A family of accelerated hybrid conjugate gradient method for unconstrained optimization and image restoration*. J. Appl. Math. Comput. 70(4), 2677-2699 (2024)
- [3] Wang, Y.J., Ouyang, C., Lv, L., Yuan, G.L.: *Analysis of a new BFGS algorithm and conjugate gradient algorithms and their applications in image restoration and machine learning*. Appl. Numer. Math. 210, 199-221 (2025)
- [4] Fletcher, R., Reeves, C.M.: *Function minimization by conjugate gradients*. Comput. J. 7, 149-154 (1964)
- [5] Polak, E., Ribière, G.: *Note sur la convergence de méthodes de directions conjuguées*. Revue française d'inform. et de Recherche Opérationnelle 3(16), 35-43 (1969)
- [6] Hestenes, M.R., Stiefel, E.: *Methods of conjugate gradients for solving linear systems* J. Res. Natl. Bur. Stand. 49, 409-436 (1952)
- [7] Dai, Y.H., Yuan, Y.X.: *A nonlinear conjugate gradient method with a strong global convergence property*. SIAM J. Optim. 10(1), 177-182 (1999)
- [8] Fletcher, R.: *Practical Methods of Optimization*. Unconstrained Optimization. Wiley, New York (1987)
- [9] Liu, Y., Storey, C.: *Efficient generalized conjugate gradient algorithms, part 1: Theory*. J. Optim. Theory Appl. 69(1), 129-137 (1991)

- [10] Hager, W.W., Zhang, H.C.: *A new conjugate gradient method with guaranteed descent and an efficient line search*. SIAM J. Optim. 16(1), 170-192 (2005)
- [11] Dai, Y.H., Han, J.Y., Liu, G.H., Sun, D.F., Yin, H.X., Yuan, Y.X.: *Convergence properties of nonlinear conjugate gradient methods*. SIAM J. Optim. 10(2), 345-358 (2000)
- [12] Dai, Y.H.: *New properties of a nonlinear conjugate gradient method*. Numer. Math. 89(1), 83-98 (2001)
- [13] Gilbert, J.C., Nocedal, J.: *Global convergence properties of conjugate gradient methods for optimization*. SIAM J. Optim. 2(1), 21-42 (1992)
- [14] Wei, Z.X., Yao, S.W., Liu, L.Y.: *The convergence properties of some new conjugate gradient methods*. Appl. Math. Comput. 183(2), 1341-1350 (2006)
- [15] Aminifard, Z., Babaie-Kafaki, S.: *A modified descent Polak-Ribière-Polyak conjugate gradient method with global convergence property for nonconvex functions*. Calcolo 56(2), 1-11 (2019)
- [16] Dong, X.L.: *A modified nonlinear Polak-Ribière-Polyak conjugate gradient method with sufficient descent property*. Calcolo 57(3), 30 (2020)
- [17] Jiang, X.Z., Jian, J.B.: *A sufficient descent Dai-Yuan type nonlinear conjugate gradient method for unconstrained optimization problems*. Nonlinear Dynam. 72, 101-112 (2013)
- [18] Jiang, X.Z., Jian, J.B.: *Two modified nonlinear conjugate gradient methods with disturbance factors for unconstrained optimization*. Nonlinear Dynam. 77, 387-397 (2014)
- [19] Yuan, G.L., Zhang, M.J.: *A three-terms Polak-Ribière-Polyak conjugate gradient algorithm for large-scale nonlinear equations*. J. Comput. Appl. Math. 286, 186-195 (2015)
- [20] Andrei, N.: *Another conjugate gradient algorithm with guaranteed descent and conjugacy conditions for large-scale unconstrained optimization*. J. Optim. Theory Appl. 159(1), 159-182 (2013)
- [21] Babaie-Kafaki, S., Ghanbari, R.: *The Dai-Liao nonlinear conjugate gradient method with optimal parameter choices*. Eur. J. Oper. Res. 234(3), 625-630 (2014)
- [22] Chen, J., Bai, Y., Yu, G., Ou, X.Q., Qin, X.L.: *A PRP type conjugate gradient method without truncation for nonconvex vector optimization*. J. Optim. Theory Appl. 204(1), 13 (2025)
- [23] Dai, Y.H., Kou, C.X.: *A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search*. SIAM J. Optim. 23(1), 296-320 (2013)
- [24] Dong, X.L., Liu, H.W., He, Y.B.: *A self-adjusting conjugate gradient method with sufficient descent condition and conjugacy condition*. J. Optim. Theory Appl. 165(1), 225-241 (2015)
- [25] Grippo, L., Lucidi, S.: *A globally convergent version of the Polak-Ribière conjugate gradient method*. Math. Programming 78(3), 375-391 (1997)
- [26] Hager, W.W., Zhang, H.C.: *A survey of nonlinear conjugate gradient methods*. Pac. J. Optim. 2(1), 35-58 (2006)
- [27] Fliege, J., Vicente, L.N.: *Multicriteria approach to bilevel optimization*. J. Optim. Theory Appl. 131(2), 209-225 (2006)

- [28] Gravel, M., Martel, J.M., Nadeau, R., Price, W., Tremblay, R.: *A multicriterion view of optimal resource allocation in job-shop production*. Eur. J. Oper. Res. 61, 230-244 (1992)
- [29] Jahn, J., Kirsch, A., Wagner, C.: *Optimization of rod antennas of mobile phones*. Math. Methods Oper. Res. 59(1), 37-51 (2004)
- [30] White, D.J.: *Epsilon-dominating solutions in mean-variance portfolio analysis*. Eur. J. Oper. Res. 105(3), 457-466 (1998)
- [31] Drummond, L.G., Raupp, F.M.P., Svaiter, B.F.: *A quadratically convergent Newton method for vector optimization*. Optimization 63(5), 661-677 (2014)
- [32] Lucambio Pérez, L.R., Prudente, L.F.: *Nonlinear conjugate gradient methods for vector optimization*. SIAM J. Optim. 28(3), 2690-2720 (2018)
- [33] Gonçalves, M.L., Lima, F.S., Prudente, L.F.: *A study of Liu-Storey conjugate gradient methods for vector optimization*. Appl. Math. Comput. 425, 127099 (2022)
- [34] Gonçalves, M.L., Prudente, L.F.: *On the extension of the Hager-Zhang conjugate gradient method for vector optimization*. Comput. Optim. Appl. 76(3), 889-916 (2020)
- [35] Touati-Ahmed, D., Storey, C.: *Efficient hybrid conjugate gradient techniques*. J. Optim. Theory Appl. 64(2), 379-397 (1990)
- [36] Andrei, N.: *Hybrid conjugate gradient algorithm for unconstrained optimization*. J. Optim. Theory Appl. 141(2), 249-264 (2009)
- [37] Dai, Y.H., Yuan, Y.X.: *An efficient hybrid conjugate gradient method for unconstrained optimization*. Ann. Oper. Res. 103(1), 33-47 (2001)
- [38] Oumertem, L., Chaib, Y.: *A new hybrid HS-DY conjugate gradient algorithm with application in mode function*. Palest. J. Math. 13(3), 808-822 (2024)
- [39] Yahaya, J., Kumam, P.: *Efficient hybrid conjugate gradient techniques for vector optimization*. Res. Control Optim. 14, 100348 (2024)
- [40] Luc, D.T.: *Theory of Vector Optimization*. Springer (1989)
- [41] Drummond, L.G., Svaiter, B.F.: *A steepest descent method for vector optimization*. J. Comput. Appl. Math. 175(2), 395-414 (2005)
- [42] Fliege, J., Svaiter, B.F.: *Steepest descent methods for multicriteria optimization*. Math. Methods Oper. Res. 51(3), 479-494 (2000)
- [43] Mita, K., Fukuda, E.H., Yamashita, N.: *Nonmonotone line searches for unconstrained multiobjective optimization problems*. J. Global Optim. 75(1), 63-90 (2019)
- [44] Ansary, M.A., Panda, G.: *A modified quasi-Newton method for vector optimization problem*. Optimization 64(11), 2289-2306 (2015)
- [45] Huband, S., Hingston, P., Barone, L., While, L.: *A review of multiobjective test problems and a scalable test problem toolkit*. IEEE Trans. Evol. Comput. 10(5), 477-506 (2006)
- [46] Das, I., Dennis, J.E.: *Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems*. SIAM J. Optim. 8(3), 631-657 (1998)

- [47] Hillermeier, C.: *Generalized homotopy approach to multiobjective optimization*. J. Optim. Theory Appl. 110(3), 557-583 (2001)
- [48] Jin, Y., Olhofer, M., Sendhoff, B.: *Dynamic weighted aggregation for evolutionary multi-objective optimization: why does it work and how?* In: Proceedings of the genetic and evolutionary computation conference, pp. 1042-1049 (2001)
- [49] Lovison, A.: *Singular continuation: Generating piecewise linear approximations to Pareto sets via global analysis*. SIAM J. Optim. 21(2), 463-490 (2011)
- [50] Moré, J.J., Garbow, B.S., Hillstom, K.E.: *Testing unconstrained optimization software*. ACM Trans. Math. Software 7(1), 17-41 (1981)
- [51] Thomann, J., Eichfelder, G.: *Numerical results for the multiobjective trust region algorithm MHT*. Data Brief 25, 104103 (2019)
- [52] Toint, P.L.: *Test problems for partially separable optimization and results for the routine PSPMIN*. The University of Namur, Department of Mathematics, Belgium, Tech. Rep(1983)
- [53] Fliege, J., Drummond, L.G., Svaiter, B.F.: *Newton's method for multiobjective optimization*. SIAM J. Optim. 20(2), 602-626 (2009)
- [54] Schütze, O., Laumanns, M., Coello Coello, C.A., Dellnitz, M., Talbi, E.G.: *Convergence of stochastic search algorithms to finite size Pareto set approximations*. J. Global Optim. 41(4), 559-577 (2008)
- [55] Gonçalves, M.L., Lima, F.S., Prudente, L.F.: *Globally convergent Newton-type methods for multi-objective optimization*. Comput. Optim. Appl. 83(2), 403-434 (2022)
- [56] Dolan, E.D., Moré, J.J.: *Benchmarking optimization software with performance profiles*. Math. Program. 91(2), 201-213 (2002)