SIAM J. OPTIM.
Vol. 28, No. 3, pp. 2690–2720

NONLINEAR CONJUGATE GRADIENT METHODS FOR VECTOR OPTIMIZATION*

L. R. LUCAMBIO PÉREZ† AND L. F. PRUDENTE†

Abstract. In this work, we propose nonlinear conjugate gradient methods for finding critical points of vector-valued functions with respect to the partial order induced by a closed, convex, and pointed cone with nonempty interior. No convexity assumption is made on the objectives. The concepts of Wolfe and Zoutendjik conditions are extended for the vector-valued optimization. In particular, we show that there exist intervals of step sizes satisfying the Wolfe-type conditions. The convergence analysis covers the vector extensions of the Fletcher–Reeves, conjugate descent, Dai–Yuan, Polak–Ribière–Polyak, and Hestenes–Stiefel parameters that retrieve the classical ones in the scalar minimization case. Under inexact line searches and without regular restarts, we prove that the sequences generated by the proposed methods find points that satisfy the first-order necessary condition for Pareto-optimality. Numerical experiments illustrating the practical behavior of the methods are presented.

Key words. vector optimization, Pareto-optimality, conjugate gradient method, unconstrained optimization, line search algorithm, Wolfe conditions

AMS subject classifications. 90C29, 90C52

DOI. 10.1137/17M1126588

1. Introduction.

In vector optimization, one considers a vector-valued function $F: \mathbb{R}^n \to \mathbb{R}^m$, and the partial order defined by a closed, convex, and pointed cone $K \subset \mathbb{R}^m$ with nonempty interior, denoted by $\mathrm{int}(K)$. The partial order $\preceq_K$ ($\prec_K$) is given by $u \preceq_K v$ ($u \prec_K v$) if and only if $v - u \in K$ ($v - u \in \mathrm{int}(K)$). We are interested in the unconstrained vector problem

$$\text{minimize}_K \quad F(x), \quad x \in \mathbb{R}^n, \tag{1}$$

where F is continuously differentiable. The concept of optimality has to be replaced by the concept of *Pareto-optimality* or *efficiency*. A point x is called Pareto-optimal if there is no y such that $y \neq x$ and $F(y) \prec_K F(x)$. A particular case, very important in practical applications, is when $K = \mathbb{R}^m_+$. This case corresponds to the multicriteria or multiobjective optimization. In multicriteria optimization, several objective functions have to be minimized simultaneously. In proper problems of this class, no single point minimizes all objectives at once. When $K = \mathbb{R}^m_+$, a point is Pareto-optimal if there does not exist a different point with smaller or equal objective function values such that there is a decrease in at least one objective. In other words, it is impossible to improve one objective without another becoming worse.

Some applications of multicriteria and vectorial optimization can be found in engineering design [38], space exploration [53], antenna design [37, 38, 39], management science [23, 30, 54, 56], environmental analysis, cancer treatment planning [35], bilevel programming [22], location science [40], statistics [17], etc.

*Received by the editors April 21, 2017; accepted for publication (in revised form) May 31, 2018; published electronically September 27, 2018.
http://www.siam.org/journals/siopt/28-3/M112658.html
Funding: This work was partially supported by CNPq (Grants 406975/2016-7) and FAPEG (Grants FAPEG/CNPq/PRONEM-201710267000532).
†Instituto de Matemática e Estatística, Universidade Federal de Goiás, Avenida Esperança, s/n, Campus Samambaia, Goiânia, GO-74690-900, Brazil (lrlp@ufg.br, lfprudente@ufg.br).

In recent papers, descent, steepest descent, Newton, quasi-Newton, projected gradient, subgradient, interior point, and proximal methods were proposed for vector optimization; see [2, 3, 9, 20, 21, 24, 25, 27, 28, 29, 42, 50, 55]. The vector optimization problem in infinite dimension was considered in [5, 6, 7, 8, 10, 11]. In this work, we propose a conjugate gradient method for the unconstrained vector optimization problem (1). Conjugate gradient methods, originally proposed in [19], constitute an important class of first-order algorithms for solving the unconstrained optimization problem

$$\text{minimize} \quad \mathtt{f}(x), \quad x \in \mathbb{R}^n, \tag{2}$$

where $\mathtt{f}\colon \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. Owing to the efficiency of the algorithms, particularly in large dimensions, considering the extension to the vector case appears natural.

The conjugate gradient algorithm for the scalar problem (2) generates a sequence of iterates according to

$$x^{k+1} = x^k + \alpha_k d^k, \quad k = 0, 1, 2, \ldots, \tag{3}$$

where $d^k \in \mathbb{R}^n$ is the line search direction, and $\alpha_k > 0$ is the step size. The direction is defined by

$$d^k = \begin{cases} -\nabla \mathtt{f}(x^k) & \text{if } k = 0, \\ -\nabla \mathtt{f}(x^k) + \beta_k d^{k-1} & \text{if } k \geq 1, \end{cases}$$

where β_k is a scalar algorithmic parameter. When $\mathtt{f}$ is strictly convex quadratic with Hessian Q, parameters β_k can be chosen such that the d^k are Q-conjugate, which means that $(d^k)^T Q d^{k+1} = 0$ for all $k \geq 0$, and the step sizes can be defined by

$$\alpha_k = \arg\min \left\{ \mathtt{f}(x^k + \alpha d^k) \mid \alpha > 0 \right\},$$

i.e., by exact minimizations in the line searches. In such a case the algorithm finds the minimizer of $\mathtt{f}$ over $\mathbb{R}^n$ after no more than n iterations, and it is called the *linear conjugate gradient algorithm*. For nonquadratic functions, different formulae for the parameters β_k result in different algorithms, known as *nonlinear conjugate gradient methods*. Some notable choices for β_k are as follows.

Fletcher–Reeves (FR) [19]: $\beta_k = \langle g^k, g^k \rangle / \langle g^{k-1}, g^{k-1} \rangle$.

Conjugate descent (CD) [18]: $\beta_k = -\langle g^k, g^k \rangle / \langle d^{k-1}, g^{k-1} \rangle$.

Dai–Yuan (DY) [16]: $\beta_k = \langle g^k, g^k \rangle / \langle d^{k-1}, g^k - g^{k-1} \rangle$.

Polak–Ribière–Polyak (PRP) [47, 48]: $\beta_k = \langle g^k, g^k - g^{k-1} \rangle / \langle g^{k-1}, g^{k-1} \rangle$.

Hestenes–Stiefel (HS) [33]: $\beta_k = \langle g^k, g^k - g^{k-1} \rangle / \langle d^{k-1}, g^k - g^{k-1} \rangle$.

Here $g^k = \nabla \mathtt{f}(x^k)$ and $\langle \cdot, \cdot \rangle$ denotes the usual inner product. A desirable property that the parameters β_k should have is that the directions d^k should be descent directions in the sense that $\langle \nabla \mathtt{f}(x^k), d^k \rangle < 0$ for all $k \geq 0$. In some convergence analyses, the more stringent *sufficient descent condition* is required, namely,

$$\langle \nabla \mathtt{f}(x^k), d^k \rangle \leq -c \|\nabla \mathtt{f}(x^k)\|^2 \quad \text{for some} \quad c > 0 \quad \text{and} \quad k \geq 0,$$

where $\|\cdot\|$ denotes the Euclidian norm. For general functions, the exact line search means that the step size α_k is a stationary point of the problem of minimizing $\mathtt{f}$ at x^k along direction d^k, implying that

$$\langle \nabla \mathtt{f}(x^k + \alpha_k d^k), d^k \rangle = 0.$$

Since the exact line search is usually expensive and impractical, one may be satisfied with step sizes satisfying, for example, the standard Wolfe conditions

$$\begin{aligned} \mathtt{f}(x^k + \alpha_k d^k) &\leq \mathtt{f}(x^k) + \rho\alpha_k \langle \nabla \mathtt{f}(x^k), d^k \rangle, \\ \langle \nabla \mathtt{f}(x^k + \alpha_k d^k), d^k \rangle &\geq \sigma \langle \nabla \mathtt{f}(x^k), d^k \rangle \end{aligned}$$

or the strong Wolfe conditions

$$\begin{aligned} \mathtt{f}(x^k + \alpha_k d^k) &\leq \mathtt{f}(x^k) + \rho\alpha_k \langle \nabla \mathtt{f}(x^k), d^k \rangle, \\ \left| \langle \nabla \mathtt{f}(x^k + \alpha_k d^k), d^k \rangle \right| &\leq \sigma \left| \langle \nabla \mathtt{f}(x^k), d^k \rangle \right|, \end{aligned}$$

where $0 < \rho < \sigma < 1$. Under mild assumptions the FR, CD, and DY methods with a line search satisfying the Wolfe conditions generate descent directions. This is not the case of PRP and HS methods. Furthermore, Powell showed in [49] that the PRP and HS methods with exact line searches can cycle without approaching a solution point. Throughout this paper, a method is called *globally convergent* if

$$\liminf_{k\to\infty} \|\nabla \mathtt{f}(x^k)\| = 0. \tag{4}$$

In the convergence analysis, an important step is to prove that the method satisfies the Zoutendjik condition

$$\sum_{k=0}^{\infty} \frac{\langle \nabla \mathtt{f}(x^k), d^k \rangle^2}{\|d^k\|^2} < \infty.$$

The concepts of Wolfe and Zoutendjik conditions are widely used to derive global convergence results of several line search algorithms for the scalar problem (2), particularly for nonlinear conjugate gradient methods; see, for example, [46]. The convergence, in the sense of (4), of the FR and DY methods under the Wolfe conditions were established by Al-Baali [1] and Dai and Yuan [16], respectively. Dai and Yuan also studied the behavior of the CD method in [15]. In [26], Gilbert and Nocedal established the convergence of the conjugate gradient method with $\beta_k = \max\{\beta_k^{PRP}, 0\}$ and with $\beta_k = \max\{\beta_k^{HS}, 0\}$. Other choices for β_k and some hybrid methods can be found in the literature. We suggest [12, 13, 14, 26, 32] for interesting reviews of these subjects.

In the context of vector optimization, we introduce the standard and strong Wolfe conditions. We say that d is a K-descent direction for F at x if there exists $T > 0$ such that $0 < t < T$ implies that $F(x + td) \prec_K F(x)$. When d is a K-descent direction, we show that there exist intervals of step sizes satisfying the Wolfe conditions. This new theoretical result sheds light on algorithmic properties and suggests the implementation of a Wolfe-type line search procedure. We also introduce the Zoutendjik condition for vector optimization and prove that it is satisfied for a general descent line search method. The considered assumptions are natural extensions of those made for the scalar case (2).

We present the general scheme of a nonlinear conjugate gradient method for vector optimization, and study its convergence for different choices of parameter β_k. The analysis covers the vector extensions of all the aforementioned parameters: FR, CD, DY, PRP, and HS. We point out that the proposed vector extensions retrieve the classical parameters in the scalar minimization case (2). Under a quite reasonable hypothesis, we show that the sequences produced by the proposed methods find points that satisfy the first-order necessary condition for Pareto-optimality. We emphasize that the extended Wolfe and Zoutendjik conditions are essential tools to prove the convergence results. Numerical experiments on convex and nonconvex multiobjective problem instances illustrating the practical behavior of the methods are presented.

This paper is organized as follows. In section 2, we summarize notions, notation, and results related to unconstrained vector optimization that we use. In section 3, we introduce the Wolfe and Zoutendjik conditions. We prove that there exist intervals of step sizes satisfying Wolfe conditions, and that the Zoutendjik condition is satisfied for a general descent line search method. In section 4, we present the general schemes for nonlinear conjugate gradient methods for vector optimization, and study its convergence without imposing explicit restrictions on β_k. In section 5, we analyze the convergence of the conjugate gradient algorithm related to the vector extensions of FR, CD, DY, PRP, and HS parameters. Numerical experiments are presented in section 6. Finally, we provide some final remarks in section 7.

2. General concepts in vector optimization. In this section we summarize the definitions, results, and notation of vector optimization. We refer the reader to [21, 27, 29, 43] and references therein. In [29], Graña Drummond and Svaiter have developed quite useful tools for the vector optimization. In particular, they characterized the directional derivative and the steepest descent direction, which are concepts widely used in the present work.

Let $K \subset \mathbb{R}^m$ be a closed, convex, and pointed cone with nonempty interior. The partial order in $\mathbb{R}^m$ induced by K, $\preceq_K$, is defined by

$$u \preceq_K v \Leftrightarrow v - u \in K,$$

and the partial order defined by $\operatorname{int}(K)$, the interior of K, $\prec_K$, is defined by

$$u \prec_K v \Leftrightarrow v - u \in \operatorname{int}(K).$$

Given a continuously differentiable function $F \colon \mathbb{R}^n \to \mathbb{R}^m$, we are concerned with the problem of finding an unconstrained K-minimizer, K-optimum, K-Pareto-optimal, or K-efficient point of F, i.e., a point $x^* \in \mathbb{R}^n$ such that there exists no other $x \in \mathbb{R}^n$ with $F(x) \preceq_K F(x^*)$ and $F(x) \neq F(x^*)$. In other words, we are seeking unconstrained minimizers for F with respect to the partial order induced by K. We denoted this problem by

$$\text{minimize}_K \quad F(x), \quad x \in \mathbb{R}^n.$$

The first-order derivative of F at x, the Jacobian of F at x, will be denoted by $JF(x)$, and the image on $\mathbb{R}^m$ by $JF(x)$ will be denoted by $\operatorname{Image}(JF(x))$. A necessary condition for K-optimality of x^* is

$$-\operatorname{int}(K) \cap \operatorname{Image}(JF(x^*)) = \emptyset.$$

A point x^* of $\mathbb{R}^n$ is called K-critical for F when it satisfies this condition. Therefore, if x is not K-critical, there exists $v \in \mathbb{R}^n$ such that $JF(x)v \in -\operatorname{int}(K)$. Every such vector v is a K-descent direction for F at x, i.e., there exists $T > 0$ such that $0 < t < T$ implies that $F(x + tv) \prec_K F(x)$; see [43].

The positive polar cone of K is the set $K^* = \{w \in \mathbb{R}^m \mid \langle w, y \rangle \geq 0 \ \forall y \in K\}$. Since K is closed and convex, $K = K^{**}$, $-K = \{y \in \mathbb{R}^m \mid \langle y, w \rangle \leq 0 \ \forall w \in K^*\}$ and $-\operatorname{int}(K) = \{y \in \mathbb{R}^m \mid \langle y, w \rangle < 0 \ \forall w \in K^* - \{0\}\}$. Let $C \subset K^* - \{0\}$ be compact and such that

$$K^* = \operatorname{cone}(\operatorname{conv}(C)),$$

i.e., K^* is the conic hull of the convex hull of C. In classical optimization $K = \mathbb{R}_+$, so $K^* = \mathbb{R}_+$ and we can take $C = \{1\}$. For multiobjective optimization $K = \mathbb{R}^m_+$, so $K^* = K$ and we may take C as the canonical basis of $\mathbb{R}^m$. If K is a polyhedral cone, K^* is also polyhedral and C can be taken as the finite set of extremal rays of K^*. For generic K, the set

$$C = \big\{w \in K^* \mid \|w\| = 1\big\}$$

has the desired properties, and, in this paper, C will denote exactly this set.

Define the function $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ by

$$\varphi(y) = \sup\big\{\langle y, w\rangle \mid w \in C\big\}.$$

In view of the compactness of C, φ is well defined. The function φ has some useful properties.

Lemma 2.1. *Let y and $y' \in \mathbb{R}^m$. Then*
(a) *$\varphi(y + y') \leq \varphi(y) + \varphi(y')$ and $\varphi(y) - \varphi(y') \leq \varphi(y - y')$;*
(b) *if $y \preceq_K y'$, then $\varphi(y) \leq \varphi(y')$; if $y \prec_K y'$, then $\varphi(y) < \varphi(y')$;*
(c) *φ is Lipschitz continuous with constant 1.*

Proof. See [29, Lemma 3.1]. ◻

The function φ gives characterizations of $-K$ and $-\mathrm{int}(K)$:

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

Note that $\varphi(x) > 0$ does not imply that $x \in K$, but $x \in K$ implies that $\varphi(x) \geq 0$ and $x \in \mathrm{int}(K)$ implies that $\varphi(x) > 0$.

Now define the function $f\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

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

The function f gives a characterization of K-descent directions and of K-critical points:

- d is K-descent direction for F at x if and only if $f(x, d) < 0$,
- x is K-critical if and only if $f(x, d) \geq 0$ for all d.

The next function allows us to extend the notion of steepest descent direction to the vector case. Define $v\colon \mathbb{R}^n \to \mathbb{R}^n$ by

$$v(x) = \arg\min\left\{ f(x, d) + \frac{\|d\|^2}{2} \;\middle|\; d \in \mathbb{R}^n \right\},$$

and $\theta\colon \mathbb{R}^n \to \mathbb{R}$ by $\theta(x) = f(x, v(x)) + \|v(x)\|^2/2$. Since $f(x, \cdot)$ is a real closed convex function, $v(x)$ exists and is unique. Observe that in the scalar minimization case, where $F\colon \mathbb{R}^n \to \mathbb{R}$ and $K = \mathbb{R}_+$, taking $C = \{1\}$, we obtain $f(x, d) = \langle \nabla F(x), d\rangle$, $v(x) = -\nabla F(x)$, and $\theta(x) = -\|\nabla F(x)\|^2/2$. The following lemma shows that $v(x)$ can be considered the vector extension of the steepest descent direction of the scalar case.

Lemma 2.2.
(a) *If x is K-critical, then $v(x) = 0$ and $\theta(x) = 0$.*
(b) *If x is not K-critical, then $v(x) \neq 0$, $\theta(x) < 0$, $f(x, v(x)) < -\frac{\|v(x)\|^2}{2} < 0$, and $v(x)$ is a K-descent direction for F at x.*
(c) *The mappings v and θ are continuous.*

Proof. See [29, Lemma 3.3]. ∎

For multiobjective optimization, where $K = \mathbb{R}^m_+$, with C given by the canonical basis of $\mathbb{R}^m$, $v(x)$ can be computed by solving

$$\text{(5)} \qquad \begin{array}{ll} \text{minimize} & \alpha + \frac{1}{2}\|d\|^2 \\ \text{subject to} & [JF(x)d]_i \leq \alpha, \quad i = 1, \ldots, m, \end{array}$$

which is a convex quadratic problem with linear inequality constraints; see [21]. For general problems, we refer the reader to [29] for a detailed discussion regarding this issue.

The next result is a direct consequence of Lemma 2.1(c).

LEMMA 2.3. *Let $F\colon \mathbb{R}^n \to \mathbb{R}^m$ be differentiable. If the Jacobian of F is Lipschitz continuous with constant L, then $f(\cdot, d)$ is Lipschitz continuous with constant $L\|d\|$.*

We finish this section with the following observations.

LEMMA 2.4. *For any scalars a, b, and $\alpha \neq 0$, we have*
(a) $ab \leq a^2/2 + b^2/2$,
(b) $2ab \leq 2\alpha^2 a^2 + b^2/(2\alpha^2)$,
(c) $(a + b)^2 \leq 2a^2 + 2b^2$,
(d) $(a + b)^2 \leq (1 + 2\alpha^2)a^2 + [1 + 1/(2\alpha^2)]b^2$.

Proof. (a) We have $0 < (a - b)^2 = a^2 - 2ab + b^2$, which implies $ab \leq a^2/2 + b^2/2$. (b) The result follows from (a), replacing a by $2\alpha a$ and b by b/α. Items (c) and (d) follow directly from (a) and (b), respectively. ∎

3. On line search in vector optimization methods. Line search procedures are crucial tools in methods for optimization. Concepts such as Armijo and Wolfe conditions are widely used in algorithms for minimization. In the classical optimization (2), for an iterative method such as (3), the Armijo condition requires the step size α_k to satisfy

$$\mathtt{f}(x^k + \alpha_k d^k) \leq \mathtt{f}(x^k) + \rho\alpha_k \langle \nabla \mathtt{f}(x^k), d^k \rangle.$$

In [29], the authors extended the Armijo condition for vector optimization, and used it to study the convergence of the steepest descent method. If d is a K-descent direction for F at x, it is said that $\alpha > 0$ satisfies the Armijo condition for $0 < \rho < 1$ if

$$\text{(6)} \qquad F(x + \alpha d) \preceq_K F(x) + \rho\alpha JF(x)d.$$

Other vector optimization methods, including Newton, quasi-Newton, and projected gradient methods, that use the Armijo rule can be found in [9, 20, 21, 24, 25, 27, 28, 42, 50]. In connection with the scalar case, we say that $\alpha > 0$ is obtained by means of an exact line search if

$$f(x + \alpha d, d) = 0.$$

Now we introduce the Wolfe-like conditions for vector optimization.

DEFINITION 3.1. *Let $d \in \mathbb{R}^n$ be a K-descent direction for F at x, and $e \in K$ a vector such that*

$$0 < \langle w, e \rangle \leq 1 \quad \textit{for all} \quad w \in C.$$

Consider $0 < \rho < \sigma < 1$. We say that $\alpha > 0$ satisfies the standard Wolfe conditions if

$$\text{(7a)} \qquad F(x + \alpha d) \preceq_K F(x) + \rho\alpha f(x, d)e,$$

$$\text{(7b)} \qquad f(x + \alpha d, d) \geq \sigma f(x, d),$$

and we say that $\alpha > 0$ *satisfies the strong Wolfe conditions if*

$$F(x + \alpha d) \preceq_K F(x) + \rho\alpha f(x, d)e, \tag{8a}$$

$$\left|f(x + \alpha d, d)\right| \leq \sigma\left|f(x, d)\right|. \tag{8b}$$

Note that the vector e actually exists. Indeed, take any $\tilde{e} \in \operatorname{int}(K)$ and let $0 < \gamma_- < \gamma_+$ be the minimum and maximum values of the continuous linear function $\langle w, \tilde{e}\rangle$ over the compact set C. Defining $e = \tilde{e}/\gamma_+$, we have $0 < \gamma_-/\gamma_+ \leq \langle w, e\rangle \leq 1$ for all $w \in C$. In multiobjective optimization, where $K = \mathbb{R}^m_+$ and C is the canonical basis of $\mathbb{R}^m$, we may take $e = [1, \dots, 1]^T \in \mathbb{R}^m$.

Next, we show that given F, K, x, and d, where d is a K-descent direction for F at x, F is continuously differentiable and bounded below along the direction d beginning at x, and K is a finitely generated cone, there exist intervals of step sizes satisfying the standard and the strong Wolfe conditions.

PROPOSITION 3.2. *Assume that* F *is of class* $\mathcal{C}^1$*,* d *is a* K*-descent direction for* F *at* x*, and there exists* $\mathcal{A} \in \mathbb{R}^m$ *such that*

$$F(x + \alpha d) \succeq_K \mathcal{A} \tag{9}$$

for all $\alpha > 0$*. If* C*, the generator of* K*, is finite, then there exist intervals of positive step sizes satisfying the standard Wolfe conditions* (7) *and the strong Wolfe conditions* (8)*.*

Proof. Given $w \in C$, define ϕ_w and $l_w : \mathbb{R} \to \mathbb{R}$ by

$$\phi_w(\alpha) = \langle w, F(x + \alpha d)\rangle$$

and

$$l_w(\alpha) = \langle w, F(x)\rangle + \alpha\rho f(x, d).$$

Observe that $\phi_w(0) = l_w(0)$. Considering (9), we have that $\phi_w(\alpha)$ is bounded below for all $\alpha > 0$. By the definition of f, we have $f(x, d) < 0$, because d is a K-descent direction for F at x. Since $\rho \in (0, 1)$, as in [46, Lemma 3.1], the line l_w is unbounded below and must intersect the graph of ϕ_w at least once for a positive α. The function ϕ_w is continuously differentiable because F is so. Hence, there exists $T_w > 0$ such that

$$\langle w, F(x + T_w d)\rangle = \langle w, F(x)\rangle + T_w\rho f(x, d),$$

and

$$\langle w, F(x + \alpha d)\rangle < \langle w, F(x)\rangle + \alpha\rho f(x, d)$$

for $\alpha \in (0, T_w)$. Define $\tilde{T} = \min\{T_w \mid w \in C\}$ and let $\tilde{w} \in C$ be such that $T_{\tilde{w}} = \tilde{T}$. Since $\tilde{T} \leq T_w$ for all $w \in C$, we obtain

$$\begin{aligned}\langle w, F(x + \alpha d)\rangle &\leq \langle w, F(x)\rangle + \alpha\rho f(x, d)\\ &\leq \langle w, F(x)\rangle + \alpha\rho f(x, d)\langle w, e\rangle\end{aligned}$$

for all $w \in C$ and $\alpha \in [0, \tilde{T}]$, because $f(x, d) < 0$ and $0 < \langle w, e\rangle \leq 1$. Therefore, (7a) and (8a) hold in $[0, \tilde{T}]$.

Consider the function $\psi : \mathbb{R} \to \mathbb{R}$ defined by

$$\psi(\alpha) = \phi_{\tilde{w}}(\alpha) - \phi_{\tilde{w}}(0) - \alpha\rho f(x, d).$$

By definition of $\tilde{T}$, we have $\psi(0) = \psi(\tilde{T}) = 0$. Then, by the mean value theorem, there exists $\tilde{\alpha} \in (0, \tilde{T})$ such that $\psi'(\tilde{\alpha}) = 0$. Hence, $\phi'_{\tilde{w}}(\tilde{\alpha}) = \rho f(x,d)$, and thus $f(x + \tilde{\alpha} d, d) \geq \rho f(x,d)$. Lemma 2.3 implies that $f(x + \alpha d, d)$, as a function of α, is continuous. So, by the intermediate value theorem, there exists $\alpha^* \in (0, \tilde{\alpha}]$ such that

$$f(x + \alpha^* d, d) = \rho f(x,d).$$

Since $0 < \rho < \sigma$ and $f(x,d) < 0$ we have

$$\sigma f(x,d) < f(x + \alpha^* d, d) < 0.$$

Hence, there is a neighborhood of α^* contained in $[0, \tilde{T}]$ for which inequalities (7b) and (8b) hold. Therefore, the standard and strong Wolfe conditions hold in this neighborhood, completing the proof. ◻

Remark. In the classical optimization, where $F : \mathbb{R}^n \to \mathbb{R}$, $K = \mathbb{R}_+$, $C = \{1\}$, and $e = 1$, we have $f(x,d)e = \langle \nabla F(x), d \rangle$. Thus, with (7) and (8), we retrieve the known Wolfe conditions for the scalar case. We claim that, in the general vector case, it is not possible to replace $f(x,d)e$ by $JF(x)d$ in (7a) and (8a). Consider the following multiobjective problem: $K = \mathbb{R}^2_+$, C is the canonical basis of $\mathbb{R}^2$, $e = [1\ 1]^T$ and $F : \mathbb{R} \to \mathbb{R}^2$, $F(x) = (F_1(x), F_2(x))$, where

$$F_1(x) = \begin{cases} -100x + 10^4 x^2 & \text{if } x < 0, \\ -\log(1 + 100x) & \text{if } 0 \leq x \leq 1, \\ -\log(101) - \frac{100}{101}(x-1) + \frac{100^2}{101^2}(x-1)^2 & \text{if } x > 1, \end{cases}$$

$$F_2(x) = 0.1x^2 - x.$$

Note that $F_1(x)$ and $F_2(x)$ are continuously differentiable and bounded below. Assume that $x = 0$ and $d = 1$. Thus, $f(x,d) = -1$. For $i = 1, 2$, define ϕ_i and $l_i : \mathbb{R} \to \mathbb{R}$ by

$$\phi_i(\alpha) = F_i(x + \alpha d) \ = F_i(\alpha)$$

and

$$l_i(\alpha) = F_i(x) + \alpha\rho \langle \nabla F_i(x), d \rangle = \alpha \rho F'_i(0).$$

Take the exogenous constants as $\rho = 0.1$ and $\sigma = 0.9$. For $\alpha > 0$, the line $l_1(\alpha) = -10\alpha$ intercepts the graph of $\phi_1(\alpha)$ only at $T_1 \approx 0.36$, and the line $l_2(\alpha) = -\alpha/10$ intercepts the graph of $\phi_2(\alpha)$ only at $T_2 = 9$. Therefore, condition (6) holds just for α in $(0, T_1)$. In this interval, we have $\phi'_1(\alpha) = -100/(1 + 100\alpha) < -1$, $\phi'_2(\alpha) > -1$, and hence

$$f(x + \alpha d, d) = \max\{\phi'_1(\alpha), \phi'_2(\alpha)\} = \phi'_2(\alpha) < \phi'_2(0.37) = -0.926 < -0.9 = \sigma f(x,d).$$

Thus, there is not an $\alpha > 0$ satisfying (6) and $f(x + \alpha d) \geq \sigma f(x,d)$ simultaneously.

Remark. When F is K-convex, we can drop the hypothesis of C being finite in Proposition 3.2. Indeed, when F is K-convex it holds that

$$\frac{\partial}{\partial T} \left[\langle w, F(x + Td) - F(x) \rangle - T\rho f(x,d) \right] |_{T = T_w} \neq 0$$

for all $w \in C$ and T_w as in the proof of Proposition 3.2. Then, by the implicit function theorem, for each $w \in C$, there exist an open set $U_w \subset \mathbb{R}^m$ and continuously differentiable function $\o_w$ such that $w \in U_w$, $T_w = \o_w(w)$, and the solution set of

$$\langle \omega, F(x + Td) - F(x) \rangle - T\rho f(x,d) = 0, \quad \omega \in U_w,$$

is $\{(\omega, \o_w(\omega)) \mid \omega \in U_w\}$. Let $U = \cup_{w\in C} U_w$ and define $\o\colon U \to \mathbb{R}$ by $\o(\omega) = \o_w(\omega)$ when $\omega \in U_w$. The function $\o$ is continuous. Since C is compact, we obtain

$$\tilde{T} = \min_{\omega\in C} \o(\omega) > 0.$$

The proof follows as in the demonstration of Proposition 3.2. When F is not K-convex it is possible to build examples where there would exist $w \in C$ such that

$$\langle w, F(x + T_w d)\rangle = \langle w, F(x)\rangle + T_w \rho f(x, d)$$

and

$$\langle w, JF(x + T_w d)d\rangle = \rho f(x, d).$$

So, for the general case, this matter remains an open problem.

Now, consider a general line search method for vector optimization

$$x^{k+1} = x^k + \alpha_k d^k, \quad k \geq 0, \tag{10}$$

where d^k is a K-descent direction for F at x^k and α_k is the step size. Let us suppose that the following assumptions are satisfied.

Assumptions.

(A1) The cone K is finitely generated and there exists an open set $\mathcal{N}$ such that $\mathcal{L} = \{x \in \mathbb{R}^n \mid F(x) \preceq_K F(x^0)\} \subset \mathcal{N}$ and the Jacobian JF is Lipschitz continuous on $\mathcal{N}$ with constant $L > 0$, i.e., x and $y \in \mathcal{N}$ implies that $\|JF(x) - JF(y)\| \leq L\|x - y\|$.

(A2) All monotonically nonincreasing sequences in $F(\mathcal{L})$ are bounded below, i.e., if the sequence $\{G_k\}_{k\in\mathbb{N}} \subset F(\mathcal{L})$ and $G_{k+1} \preceq_K G_k$, for all k, then there exists $\mathcal{G} \in \mathbb{R}^m$ such that $\mathcal{G} \preceq_K G_k$ for all k.

Under Assumptions (A1) and (A2), we establish that the general method (10) satisfies a condition of Zoutendijk's type. We emphasize that the considered assumptions are natural extensions of those made for the scalar case.

PROPOSITION 3.3. *Assume that Assumptions* (A1) *and* (A2) *hold. Consider an iteration of the form* (10)*, where d^k is a K-descent direction for F at x^k and α_k satisfies the standard Wolfe conditions* (7)*. Then*

$$\sum_{k\geq 0} \frac{f^2(x^k, d^k)}{\|d^k\|^2} < \infty. \tag{11}$$

Proof. By (7b), Assumption (A1), and Lemma 2.3, we obtain

$$(\sigma - 1)f(x^k, d^k) \leq f(x^{k+1}, d^k) - f(x^k, d^k) \leq L\alpha_k \|d^k\|^2.$$

Since $f(x^k, d^k) < 0$ and $\|d^k\| \neq 0$, we obtain

$$\frac{f^2(x^k, d^k)}{\|d^k\|^2} \leq L\alpha_k \frac{f(x^k, d^k)}{\sigma - 1}. \tag{12}$$

By (7a), we have that $\{F(x^k)\}_{k\geq 0}$ is monotone decreasing and that

$$F(x^{k+1}) - F(x^0) \preceq_K \rho \sum_{j=0}^{k} \alpha_j f(x^j, d^j) e$$

for all $k \geq 0$. Then, by Assumption (A2), since $\{F(x^k)\}_{k\geq 0} \subset \mathcal{L}$, there exists $\mathcal{F} \in \mathbb{R}^m$ such that

$$\mathcal{F} - F(x^0) \preceq_K \rho \sum_{j=0}^{k} \alpha_j f(x^j, d^j) e$$

for all $k \geq 0$. Therefore, for all $k \geq 0$ and $w \in C$, we obtain

$$\langle \mathcal{F} - F(x^0), w \rangle \leq \rho \langle e, w \rangle \sum_{j=0}^{k} \alpha_j f(x^j, d^j),$$

and hence

$$\frac{1}{\sigma - 1} \min \left\{ \langle \mathcal{F} - F(x^0), \bar{w} \rangle \mid \bar{w} \in C \right\} \geq \rho \langle e, w \rangle \sum_{j=0}^{k} \alpha_j \frac{f(x^j, d^j)}{\sigma - 1} > 0.$$

We conclude that

$$\sum_{k\geq 0} \alpha_k \frac{f(x^k, d^k)}{\sigma - 1} < \infty,$$

which together with (12) gives

$$\sum_{k\geq 0} \frac{f^2(x^k, d^k)}{\|d^k\|^2} < \infty. \qquad \square$$

Remark. The Zoutendijk condition (11) can be used to derive convergence results for several line search algorithms. In particular, the method of steepest descent converges in the sense that

$$\lim_{k\to\infty} \|v(x^k)\| = 0,$$

provided that it uses a line search satisfying the standard Wolfe conditions. In fact, consider the iterative method (10) with $d^k = v(x^k)$, and assume that α_k satisfies (7) for all $k \geq 0$. If $v(x^k) = 0$ for some k, the result trivially holds. Suppose now that $v(x^k) \neq 0$ for all $k \geq 0$. Since $\theta(x^k) = f(x^k, v(x^k)) + \|v(x^k)\|^2/2 < 0$, it follows that

$$\frac{f^2(x^k, d^k)}{\|d^k\|^2} = \frac{f^2(x^k, v(x^k))}{\|v(x^k)\|^2} \geq \frac{1}{4} \|v(x^k)\|^2.$$

Summing this inequality over all k, we obtain from the Zoutendijk condition that

$$\sum_{k\geq 0} \|v(x^k)\|^2 < \infty,$$

which implies that $\|v(x^k)\| \to 0$.

4. The nonlinear conjugate gradient algorithm. In the following we define the nonlinear conjugate gradient algorithm (NLCG algorithm) for the vector optimization problem (1).

NLCG algorithm.

Let $x^0 \in \mathbb{R}^n$. Compute $v(x^0)$ and initialize $k \leftarrow 0$.
Step 1 If $v(x^k) = 0$, then STOP.

Step 2 Define

$$d^k = \begin{cases} v(x^k) & \text{if } k = 0, \\ v(x^k) + \beta_k d^{k-1} & \text{if } k \geq 1, \end{cases} \tag{13}$$

where β_k is an algorithmic parameter.

Step 3 Compute a step size $\alpha_k > 0$ by a line search procedure and set $x^{k+1} = x^k + \alpha_k d^k$.

Step 4 Compute $v(x^{k+1})$, set $k \leftarrow k + 1$, and go to Step 1.

The choice for updating the parameter β_k and the adopted line search procedure remain deliberately open. In the next section, we consider several choices of β_k combined with an appropriate line search strategy that result in globally convergent methods. The NLCG algorithm successfully stops if a K-critical point of F is found. Hence, from now on, let us consider that $v(x^k) \neq 0$ for all $k \geq 0$.

Throughout the paper the line search procedure must be such that the step size satisfies the standard or strong Wolfe conditions. Thus, for the well-definedness of the NLCG algorithm, Proposition 3.2 requires d^k to be a K-descent direction of F at x^k, which is equivalent to $f(x^k, d^k) < 0$. In some convergence analyses we will require the more stringent condition

$$f(x^k, d^k) \leq c f(x^k, v(x^k)) \tag{14}$$

for some $c > 0$ and for all $k \geq 0$. In connection with the scalar case, we call (14) the *sufficient descent condition*. Condition (14) can be obtained for an adequate line search procedure provided that d^{k-1} is a K-descent direction of F at x^{k-1}. In this case, by Proposition 3.2 a line search along d^{k-1} can be performed enforcing the (standard or strong) Wolfe conditions to obtain x^k. Moreover, the line search procedure can be implemented in such a way that it gives, in the limit, a point x^k with $f(x^k, d^{k-1}) = 0$. Now, from (13), and the definition of f, if $\beta_k \geq 0$, we have

$$f(x^k, d^k) \leq f(x^k, v(x^k)) + \beta_k f(x^k, d^{k-1}).$$

Then, if β_k is bounded, the line search procedure can be applied to reduce $\left|f(x^k, d^{k-1})\right|$ sufficiently and obtain (14). A line search procedure with these characteristics may be coded based on the work of Moré and Thuente [45]. We refer the reader to [26] for a careful discussion about this issue in the classical optimization approach.

In the next lemma, we give a sufficient condition on β_k for ensuring the descent property on d^k.

LEMMA 4.1. *Assume that in the NLCG algorithm, the sequence β_k is defined so that it has the following property:*

$$\beta_k \in \begin{cases} [0, \infty) & \textit{if } f(x^k, d^{k-1}) \leq 0, \\ \left[0, -f(x^k, v(x^k))/f(x^k, d^{k-1})\right) & \textit{if } f(x^k, d^{k-1}) > 0, \end{cases} \tag{15}$$

or

$$\beta_k \in \begin{cases} [0, \infty) & \textit{if } f(x^k, d^{k-1}) \leq 0, \\ \left[0, -\mu f(x^k, v(x^k))/f(x^k, d^{k-1})\right] & \textit{if } f(x^k, d^{k-1}) > 0 \end{cases} \tag{16}$$

for some $\mu \in [0, 1)$. If property (15) *holds, then d^k is a K-descent direction for all k. If property* (16) *holds, then d^k satisfies the sufficient descent condition with $c = 1 - \mu$ for all k.*

Proof. Let us show the second statement. Inequality (14) with $c = 1-\mu$ clearly holds for $k = 0$. For $k \geq 1$, by the definition of d^k, we have

$$\langle JF(x^k)d^k, w\rangle = \langle JF(x^k)v(x^k), w\rangle + \beta_k\langle JF(x^k)d^{k-1}, w\rangle \quad \forall w \in \mathbb{R}^m.$$

First suppose that $f(x^k, d^{k-1}) \leq 0$. Since $\beta_k \geq 0$, for all $w \in C$, we obtain

$$\langle JF(x^k)d^k, w\rangle \leq \langle JF(x^k)v(x^k), w\rangle \leq f(x^k, v(x^k)) \leq (1-\mu)f(x^k, v(x^k)).$$

Now assume that $f(x^k, d^{k-1}) > 0$. By the definition of f, and taking into account (16), we verify for all $w \in C$ that

$$\langle JF(x^k)d^k, w\rangle \leq f(x^k, v(x^k)) + \beta_k f(x^k, d^{k-1}) \leq (1-\mu)f(x^k, v(x^k)).$$

In both cases $f(x^k, d^k) = \varphi(JF(x^k)d^k) \leq (1-\mu)f(x^k, v(x^k))$. The proof of the first statement can be similarly obtained. ◻

Without imposing an explicit restriction on β_k, the main convergence result of the NLCG algorithm is as follows. Note that Theorem 4.2(ii) is the vector extension of [14, Theorem 2.3].

THEOREM 4.2. *Consider the NLCG algorithm. Assume that Assumptions* (A1) *and* (A2) *hold and that*

$$\sum_{k\geq 0} \frac{1}{\|d^k\|^2} = \infty. \tag{17}$$

(i) *If d^k satisfies the sufficient descent condition* (14) *and α_k satisfies the standard Wolfe conditions* (7)*, then* $\liminf_{k\to\infty} \|v(x^k)\| = 0$.

(ii) *If $\beta_k \geq 0$, d^k is K-descent direction of F at x^k, and α_k satisfies the strong Wolfe conditions* (8)*, then* $\liminf_{k\to\infty} \|v(x^k)\| = 0$.

Proof. Assume by contradiction that there exists a constant γ such that

$$\|v(x^k)\| \geq \gamma \quad \text{for all} \quad k \geq 0.$$

Let us show item (i). Observe that, by $f(x^k, v(x^k)) + \|v(x^k)\|^2/2 < 0$ and (14), we obtain

$$\frac{c^2\gamma^4}{4\|d^k\|^2} \leq \frac{c^2\|v(x^k)\|^4}{4\|d^k\|^2} \leq \frac{c^2 f^2(x^k, v(x^k))}{\|d^k\|^2} \leq \frac{f^2(x^k, d^k)}{\|d^k\|^2}.$$

Since, under the hypotheses of item (i), the Zoutendijk condition (11) holds, we have a contradiction to (17). Then, item (i) is demonstrated.

Now consider item (ii). Remembering (13), we have $-\beta_k d^{k-1} = -d^k + v(x^k)$. Then

$$0 \leq \beta_k^2\|d^{k-1}\|^2 \leq \Big[\|d^k\| + \|v(x^k)\|\Big]^2 \leq 2\|d^k\|^2 + 2\|v(x^k)\|^2,$$

where the second inequality follows from Lemma 2.4(c). Hence,

$$\|d^k\|^2 \geq -\|v(x^k)\|^2 + \frac{\beta_k^2}{2}\|d^{k-1}\|^2. \tag{18}$$

On the other hand, by the definition of d^k and the positiveness of β_k, we obtain

$$f(x^k, d^k) \leq f(x^k, v(x^k)) + \beta_k f(x^k, d^{k-1}).$$

Then

(19)
$$0 \le -f(x^k, v(x^k)) \le -f(x^k, d^k) + \beta_k f(x^k, d^{k-1}) \le -f(x^k, d^k) - \sigma\beta_k f(x^{k-1}, d^{k-1}),$$

because we are assuming that α_k satisfies the strong Wolfe conditions, i.e., (8b) holds. From (19) and using Lemma 2.4(b) with $\alpha = 1$, we obtain

$$\begin{aligned} f^2(x^k, v(x^k)) &\le f^2(x^k, d^k) + \sigma^2\beta_k^2 f^2(x^{k-1}, d^{k-1}) + 2\sigma f(x^k, d^k)\beta_k f(x^{k-1}, d^{k-1}) \\ &\le (1+2\sigma^2)\left[f^2(x^k, d^k) + \frac{\beta_k^2}{2} f^2(x^{k-1}, d^{k-1})\right]. \end{aligned}$$

Since d^k is a K-descent direction for F at x^k, we obtain $\theta(x^k) = f(x^k, v(x^k)) + \|v(x^k)\|^2/2 < 0$. Hence,

$$f^2(x^k, d^k) + \frac{\beta_k^2}{2} f^2(x^{k-1}, d^{k-1}) \ge c_1\|v(x^k)\|^4, \tag{20}$$

where $c_1 = [4(1+2\sigma^2)]^{-1}$. Note that

$$\frac{f^2(x^k, d^k)}{\|d^k\|^2} + \frac{f^2(x^{k-1}, d^{k-1})}{\|d^{k-1}\|^2} = \frac{1}{\|d^k\|^2}\left[f^2(x^k, d^k) + \frac{\|d^k\|^2}{\|d^{k-1}\|^2} f^2(x^{k-1}, d^{k-1})\right]$$

is, by (18), greater than or equal to

$$\frac{1}{\|d^k\|^2}\left\{f^2(x^k, d^k) + \left[\frac{\beta_k^2}{2} - \frac{\|v(x^k)\|^2}{\|d^{k-1}\|^2}\right] f^2(x^{k-1}, d^{k-1})\right\}$$

and, by (20), this last expression is greater than or equal to

$$\frac{\|v(x^k)\|^2}{\|d^k\|^2}\left[c_1\|v(x^k)\|^2 - \frac{f^2(x^{k-1}, d^{k-1})}{\|d^{k-1}\|^2}\right].$$

Since, under the hypotheses of item (ii), the Zoutendijk condition (11) holds, and it implies that $f(x^k, d^k)/\|d^k\|$ tends to zero, we have

$$\frac{f^2(x^k, d^k)}{\|d^k\|^2} + \frac{f^2(x^{k-1}, d^{k-1})}{\|d^{k-1}\|^2} \ge \frac{c_1}{2}\frac{\|v(x^k)\|^4}{\|d^k\|^2} \ge \frac{c_1\gamma^4}{2}\frac{1}{\|d^k\|^2}$$

for all sufficiently large k. Hence, we have a contradiction to (17) and the proof is complete. ◻

5. Convergence analysis for specific $\boldsymbol{\beta_k}$. In this section, we analyze the convergence properties of the NLCG algorithm related to the vector extensions of FR, CD, DY, PRP, and HS choices for parameter β_k. In all cases, the proposed vector extensions retrieve the classical parameters in the scalar minimization case.

For some results, we need to replace Assumption (A2) by the following stronger hypothesis.

Assumption.

(A3) The level set $\mathcal{L} = \{x \mid F(x) \preceq_K F(x^0)\}$ is bounded.

If d^k is a K-descent direction of F at x^k in the NLCG algorithm, we claim that under Assumption (A3) the sequence $\{f(x^k, v(x^k))\}$ is bounded. Indeed, in this case $\{x^k\} \subset \mathcal{L}$ and, by continuity arguments, there are constants $\bar{\gamma} > 0$ and $\bar{c} > 0$ such that $\|v(x^k)\| \leq \bar{\gamma}$ and $\|JF(x^k)\| \leq \bar{c}$ for all $k \geq 0$. Then, for all $w \in C$ and $k \geq 0$, it turns out that

$$0 > f(x^k, v(x^k)) \geq \langle JF(x^k)v(x^k), w\rangle \geq -\|JF(x^k)\|\|v(x^k)\| \geq -\bar{c}\bar{\gamma}, \tag{21}$$

because $\|w\| = 1$.

5.1. Fletcher–Reeves. The vector extension of the FR parameter is given by

$$\beta_k^{FR} = \frac{f(x^k, v(x^k))}{f(x^{k-1}, v(x^{k-1}))}.$$

In the next theorem, under a suitable hypothesis, we show that it is possible to obtain global convergence if the parameter β_k is bounded in magnitude by any fraction of β_k^{FR}.

THEOREM 5.1. *Consider the NLCG algorithm and let* $0 \leq \delta < 1$.

(i) *Let Assumptions* (A1) *and* (A2) *hold. If*

$$|\beta_k| \leq \delta \beta_k^{FR},$$

d^k *satisfies the sufficient descent condition* (14), *and* α_k *satisfies the standard Wolfe conditions* (7), *then* $\liminf_{k\to\infty} \|v(x^k)\| = 0$.

(ii) *Let Assumptions* (A1) *and* (A3) *hold. If*

$$0 \leq \beta_k \leq \delta \beta_k^{FR},$$

d^k *is a* K*-descent direction of* F *at* x^k, *and* α_k *satisfies the strong Wolfe conditions* (8), *then* $\liminf_{k\to\infty} \|v(x^k)\| = 0$.

Proof. Assume by contradiction that there exists $\gamma > 0$ such that

$$\|v(x^k)\| \geq \gamma \quad \text{for all} \quad k \geq 0. \tag{22}$$

By (13), and Lemma 2.4(d) with $a = \|v(x^k)\|$, $b = |\beta_k|\,\|d^{k-1}\|$, and $\alpha = \delta/\sqrt{2(1-\delta^2)}$, we have

$$\|d^k\|^2 \leq \left[\|v(x^k)\| + |\beta_k|\,\|d^{k-1}\|\right]^2 \leq \frac{1}{1-\delta^2}\|v(x^k)\|^2 + \frac{1}{\delta^2}\beta_k^2\|d^{k-1}\|^2.$$

Thus,

$$\begin{aligned}\frac{\|d^k\|^2}{f^2(x^k, v(x^k))} &\leq \frac{1}{1-\delta^2}\frac{\|v(x^k)\|^2}{f^2(x^k, v(x^k))} + \frac{1}{\delta^2}\frac{\beta_k^2\|d^{k-1}\|^2}{f^2(x^k, v(x^k))} \\ &\leq \frac{1}{1-\delta^2}\frac{\|v(x^k)\|^2}{f^2(x^k, v(x^k))} + \frac{\|d^{k-1}\|^2}{f^2(x^{k-1}, v(x^{k-1}))},\end{aligned}$$

because $|\beta_k| \leq \delta f(x^k, v(x^k))/f(x^{k-1}, v(x^{k-1}))$. Remembering $0 < \gamma^2 \leq \|v(x^k)\|^2 \leq -2f(x^k, v(x^k))$, we obtain

$$\frac{\|d^k\|^2}{f^2(x^k, v(x^k))} \leq \frac{4}{(1-\delta^2)\gamma^2} + \frac{\|d^{k-1}\|^2}{f^2(x^{k-1}, v(x^{k-1}))}.$$

Applying this relation repeatedly, it follows that

$$\begin{aligned}\frac{\|d^k\|^2}{f^2(x^k, v(x^k))} &\leq \frac{4}{(1-\delta^2)\gamma^2}k + \frac{\|d^0\|^2}{f^2(x^0, v(x^0))} \\ &\leq \frac{4}{(1-\delta^2)\gamma^2}k + \frac{4}{\gamma^2} = \frac{4}{\gamma^2}\left(\frac{1}{1-\delta^2}k + 1\right).\end{aligned} \tag{23}$$

Hence,

$$\frac{f^2(x^k, v(x^k))}{\|d^k\|^2} \geq \frac{\gamma^2(1-\delta^2)}{4}\frac{1}{k+1-\delta^2} \geq \frac{\gamma^2(1-\delta^2)}{4}\frac{1}{k+1}. \tag{24}$$

Consider (i). By the sufficient descent condition (14) and (24) we have

$$\sum_{k\geq 0}\frac{f^2(x^k, d^k)}{\|d^k\|^2} \geq \sum_{k\geq 0} c^2\frac{f^2(x^k, v(x^k))}{\|d^k\|^2} \geq \frac{c^2(1-\delta^2)\gamma^2}{4}\sum_{k\geq 0}\frac{1}{k+1} = \infty.$$

We have a contradiction because, under the hypotheses of (i), the Zoutendijk condition (11) holds. Thus, item (i) is demonstrated.

Now let us show item (ii). By (21) we obtain $f^2(x^k, v(x^k)) \leq \bar{c}^2\bar{\gamma}^2$. Therefore, by (23), it follows that

$$\|d^k\|^2 \leq c_1 k + c_2,$$

where $c_1 = 4\bar{c}^2\bar{\gamma}^2/[\gamma^2(1-\delta^2)] > 0$ and $c_2 = 4\bar{c}^2\bar{\gamma}^2/\gamma^2 > 0$. Then

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

which implies by Theorem 4.2(ii) that $\liminf_{k\to\infty}\|v(x^k)\| = 0$, contradicting (22). $\square$

We clearly have the following corollary.

Corollary 5.2. *Assume that Assumptions* (A1) *and* (A3) *hold. Consider an NLCG algorithm with*

$$\beta_k = \delta\beta_k^{FR}, \quad \textit{where} \quad 0 \leq \delta < 1.$$

If α_k satisfies the strong Wolfe conditions (8)*, and d^k is a K-descent direction of F at x^k, then* $\liminf_{k\to\infty}\|v(x^k)\| = 0$.

We point out that the descent condition can be enforced while performing the line search. In the scalar minimization case, if $|\beta_k| \leq \beta_k^{FR}$ and α_k satisfies the strong Wolfe conditions (8) with $\sigma < 1/2$, it is possible to show that the sufficient descent condition (14) always holds; see [1, 26]. For vector optimization, this is an open question.

5.2. Conjugate descent. Next, we derive the convergence result related to the CD parameter, which is given by

$$\beta_k^{CD} = \frac{f(x^k, v(x^k))}{f(x^{k-1}, d^{k-1})}.$$

In the following lemma, we show that the NLCG algorithm generates descent directions if $0 \leq \beta_k \leq \beta_k^{CD}$ and the line search satisfies the strong Wolfe conditions.

LEMMA 5.3. *Consider an NLCG algorithm with* $0 \leq \beta_k \leq \beta_k^{CD}$ *and suppose that* α_k *satisfies the strong Wolfe conditions* (8)*. Then* d^k *satisfies the sufficient descent condition* (14) *with* $c = 1 - \sigma$*.*

Proof. The proof is by induction. For $k = 0$, since $d^0 = v(x^0)$, $f(x^0, v(x^0)) < 0$, and $0 < \sigma < 1$, we see that inequality (14) with $c = 1 - \sigma$ holds. For some $k \geq 1$, assume that

$$f(x^{k-1}, d^{k-1}) \leq (1 - \sigma) f(x^{k-1}, v(x^{k-1})) < 0.$$

Hence, $\beta_k^{CD} > 0$ and β_k is well defined. By the definition of f and the strong Wolfe condition (8b), we have

$$\begin{aligned} f(x^k, d^k) &\leq f(x^k, v(x^k)) + \beta_k f(x^k, d^{k-1}) \leq f(x^k, v(x^k)) - \sigma \beta_k f(x^{k-1}, d^{k-1}) \\ &\leq f(x^k, v(x^k)) - \sigma \beta_k^{CD} f(x^{k-1}, d^{k-1}) = (1 - \sigma) f(x^k, v(x^k)), \end{aligned}$$

concluding the proof. ∎

Now we show that the NLCG algorithm generates a globally convergent sequence if β_k is nonnegative and bounded above by an appropriate fraction of β_k^{CD} and the line search satisfies the strong Wolfe conditions. It is worth mentioning that even for the scalar case the conjugate gradient method with $\beta_k = \beta_k^{CD}$ and strong Wolfe line search may not converge to a stationary point; see [15].

THEOREM 5.4. *Let Assumptions* (A1) *and* (A2) *hold. Consider an NLCG algorithm with*

$$\beta_k = \eta \beta_k^{CD}, \quad \textit{where} \quad 0 \leq \eta < 1 - \sigma.$$

If α_k *satisfies the strong Wolfe conditions* (8)*, then* $\liminf_{k \to \infty} \|v(x^k)\| = 0$*.*

Proof. It follows from Lemma 5.3 that d^k satisfies the sufficient descent condition (14) with $c = 1 - \sigma$ for all $k \geq 0$. Therefore,

$$0 \leq \beta_k = \frac{\eta}{1 - \sigma}(1 - \sigma) \frac{f(x^k, v(x^k))}{f(x^{k-1}, d^{k-1})} \leq \frac{\eta}{1 - \sigma} \frac{f(x^k, v(x^k))}{f(x^{k-1}, v(x^{k-1}))} = \frac{\eta}{1 - \sigma} \beta_k^{FR}.$$

Since $0 \leq \eta/(1 - \sigma) < 1$ the proof follows from Theorem 5.1. ∎

5.3. Dai–Yuan. Consider the DY vector parameter

$$\beta_k^{DY} = \frac{-f(x^k, v(x^k))}{f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1})}.$$

Following the methodology of CD, let us prove that the NLCG algorithm with $0 \leq \beta_k \leq \beta_k^{DY}$ and a line search satisfying the strong Wolfe conditions also generates descent directions.

LEMMA 5.5. *Consider an NLCG algorithm with* $0 \leq \beta_k \leq \beta_k^{DY}$ *and suppose that* α_k *satisfies the strong Wolfe conditions* (8)*. Then* d^k *satisfies the sufficient descent condition* (14) *with* $c = 1/(1 + \sigma)$*.*

Proof. We proceed by induction. For $k = 0$, since $d^0 = v(x^0)$, $f(x^0, v(x^0)) < 0$, and $0 < \sigma < 1$, we see that inequality (14) with $c = 1/(1 + \sigma)$ is satisfied. For some $k \geq 1$, assume that

$$f(x^{k-1}, d^{k-1}) \leq \frac{1}{1 + \sigma} f(x^{k-1}, v(x^{k-1})) < 0.$$

From the Wolfe condition (7b), we obtain

$$f(x^k, d^{k-1}) \geq \sigma f(x^{k-1}, d^{k-1}) > f(x^{k-1}, d^{k-1}),$$

because $\sigma < 1$ and $f(x^{k-1}, d^{k-1}) < 0$. Hence, $\beta_k^{DY} > 0$ and β_k is well defined. By the definition of d^k and positiveness of β_k we obtain

$$f(x^k, d^k) \leq f(x^k, v(x^k)) + \beta_k f(x^k, d^{k-1}).$$

If $f(x^k, d^{k-1}) \leq 0$ the result holds trivially. Now suppose that $f(x^k, d^{k-1}) > 0$. Then,

$$f(x^k, d^k) \leq f(x^k, v(x^k)) + \beta_k f(x^k, d^{k-1}) \leq f(x^k, v(x^k)) + \beta_k^{DY} f(x^k, d^{k-1}).$$

Define $l_k = f(x^k, d^{k-1})/f(x^{k-1}, d^{k-1})$. By the strong Wolfe condition (8b) we have $l_k \in [-\sigma, \sigma]$. Using the above inequality, direct calculations show that

$$f(x^k, d^k) \leq \frac{1}{1 - l_k} f(x^k, v(x^k)) \leq \frac{1}{1+\sigma} f(x^k, v(x^k)),$$

which concludes the proof. ∎

We observe that, if α_k satisfies the standard Wolfe conditions (7), it is possible to show that the DY method generates simple descent directions. In the following, we state the convergence result related to the DY method.

THEOREM 5.6. *Let Assumptions* (A1) *and* (A2) *hold. Consider an NLCG algorithm with*

$$\beta_k = \eta \beta_k^{DY}, \quad \text{where} \quad 0 \leq \eta < \frac{1-\sigma}{1+\sigma}.$$

If α_k satisfies the strong Wolfe conditions (8)*, then* $\liminf_{k\to\infty} \|v(x^k)\| = 0$.

Proof. It follows from Lemma 5.5 that d^k satisfies the sufficient descent condition (14) with $c = 1/(1+\sigma)$ for all $k \geq 0$. Then, by the Wolfe condition (8b), we have

$$f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1}) \geq (\sigma - 1) f(x^{k-1}, d^{k-1}) \geq \frac{\sigma - 1}{1+\sigma} f(x^{k-1}, v(x^{k-1})) > 0.$$

Then,

$$\frac{-f(x^{k-1}, v(x^{k-1}))}{f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1})} \leq \frac{1+\sigma}{1-\sigma}.$$

Define $\delta = \eta(1+\sigma)/(1-\sigma)$. Thus, by the definition of β_k, we obtain

$$\begin{aligned}
\beta_k &= \delta \frac{1-\sigma}{1+\sigma} \frac{\left[-f(x^k, v(x^k))\right]}{f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1})} \\
&\leq \delta \frac{1-\sigma}{1+\sigma} \left[\frac{-f(x^k, v(x^k))}{-f(x^{k-1}, v(x^{k-1}))}\right] \left[\frac{-f(x^{k-1}, v(x^{k-1}))}{f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1})}\right] \\
&\leq \delta \frac{f(x^k, v(x^k))}{f(x^{k-1}, v(x^{k-1}))} = \delta \beta_k^{FR}.
\end{aligned}$$

Since $0 \leq \delta < 1$ the proof follows from Theorem 5.1. ∎

Theorem 5.6 states the global convergence of an NLCG algorithm with an appropriate fraction of DY parameter. We conclude this section by noting that it is possible to obtain global convergence with a slight modification in the DY parameter. Let us define the modified DY parameter by

$$\beta_k^{mDY} = \frac{-f(x^k, v(x^k))}{f(x^k, d^{k-1}) - \tau f(x^{k-1}, d^{k-1})},$$

where $\tau > 1$. Observe that, if the line searches are exact, we have $\beta_k^{mDY} \leq \frac{1}{\tau}\beta_k^{FR}$ and the convergence can be obtained by using Theorem 5.1.

THEOREM 5.7. *Let Assumptions* (A1) *and* (A2) *hold. Consider an NLCG algorithm with*

$$\beta_k = \beta_k^{mDY}.$$

If α_k satisfies the strong Wolfe conditions (8), *then* $\liminf_{k\to\infty} \|v(x^k)\| = 0$.

Proof. Similarly to Lemma 5.5, we may show that $\beta_k^{mDY} > 0$ and that d^k satisfies the sufficient descent condition (14) with $c = \tau/(\tau+\sigma)$, i.e.,

$$f(x^k, d^k) \leq \frac{\tau}{\tau+\sigma} f(x^k, v(x^k)) \tag{25}$$

for all $k \geq 0$. Note that

$$f(x^k, d^k) \leq f(x^k, v(x^k)) + \beta_k^{mDY} f(x^k, d^{k-1}) = \tau\beta_k^{mDY} f(x^{k-1}, d^{k-1}),$$

which implies that

$$\beta_k^{mDY} \leq \frac{f(x^k, d^k)}{\tau f(x^{k-1}, d^{k-1})}. \tag{26}$$

Assume by contradiction that there exists $\gamma > 0$ such that

$$\|v(x^k)\| \geq \gamma \quad \text{for all} \quad k \geq 0.$$

By the definition of d^k, and Lemma 2.4(d) with $a = \|v(x^k)\|$, $b = \beta_k^{mDY}\|d^{k-1}\|$, and $\alpha = 1/\sqrt{2(\tau^2-1)}$, we have

$$\|d^k\|^2 \leq \left[\|v(x^k)\| + \beta_k^{mDY}\|d^{k-1}\|\right]^2 \leq \frac{\tau^2}{\tau^2-1}\|v(x^k)\|^2 + \tau^2(\beta_k^{mDY})^2\|d^{k-1}\|^2.$$

Thus, from (25) and (26) and observing that $f(x^k, v(x^k)) < 0$, we obtain

$$\begin{aligned}\frac{\|d^k\|^2}{f^2(x^k, d^k)} &\leq \frac{\tau^2}{\tau^2-1}\frac{\|v(x^k)\|^2}{f^2(x^k, d^k)} + \frac{\|d^{k-1}\|^2}{f^2(x^{k-1}, d^{k-1})}\\ &\leq \frac{(\tau+\sigma)^2}{\tau^2-1}\frac{\|v(x^k)\|^2}{f^2(x^k, v(x^k))} + \frac{\|d^{k-1}\|^2}{f^2(x^{k-1}, d^{k-1})}.\end{aligned}$$

Then, remembering that $0 < \gamma^2 \leq \|v(x^k)\|^2 \leq -2f(x^k, v(x^k))$, we obtain

$$\frac{\|d^k\|^2}{f^2(x^k, d^k)} \leq \frac{4(\tau+\sigma)^2}{(\tau^2-1)\gamma^2} + \frac{\|d^{k-1}\|^2}{f^2(x^{k-1}, d^{k-1})}.$$

Applying the above relation repeatedly, it follows that

$$\frac{\|d^k\|^2}{f^2(x^k, d^k)} \leq c_1 k + c_2,$$

where $c_1 = 4(\tau + \sigma)^2/[(\tau^2 - 1)\gamma^2] > 0$ and $c_2 = \|v(x^0)\|^2/f^2(x^0, v(x^0)) > 0$. Then

$$\sum_{k \geq 0} \frac{f^2(x^k, d^k)}{\|d^k\|^2} \geq \sum_{k \geq 0} \frac{1}{c_1 k + c_2} = \infty,$$

contradicting the Zoutendijk condition (11). ∎

5.4. Polak–Ribière–Polyak and Hestenes–Stiefel. In this section, we analyze the convergence of the NLCG algorithm related to the PRP and HS parameters given by

$$\beta_k^{PRP} = \frac{-f(x^k, v(x^k)) + f(x^{k-1}, v(x^k))}{-f(x^{k-1}, v(x^{k-1}))}$$

and

$$\beta_k^{HS} = \frac{-f(x^k, v(x^k)) + f(x^{k-1}, v(x^k))}{f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1})}.$$

In the scalar minimization case, Powell showed in [49] that the PRP and HS methods with exact line searches can cycle without approaching a solution point. Gilbert and Nocedal [26] proved that global convergence can be obtained for $\beta_k = \max\{\beta_k^{PRP}, 0\}$ and for $\beta_k = \max\{\beta_k^{HS}, 0\}$. Dai et al. in [14] showed that the positiveness restriction on β_k cannot be relaxed for the PRP method. We can retrieve these results for the vector optimization context.

Our results in this section are based on the work [26] of Gilbert and Nocedal. They studied the convergence of PRP and HS methods, for the scalar minimization case, introducing the so-called Property $(*)$. The vector extension of this property is as follows.

PROPERTY $(*)$. *Consider an NLCG algorithm and suppose that*

$$0 < \gamma \leq \|v(x^k)\| \leq \bar{\gamma} \tag{27}$$

for all $k \geq 0$. Under this assumption, we say that the method has Property $(*)$ if there exist constants $b > 1$ and $\lambda > 0$ such that, for all k,

$$|\beta_k| \leq b$$

and

$$\|s^{k-1}\| \leq \lambda \Rightarrow |\beta_k| \leq \frac{1}{2b},$$

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

In [26], the authors established global convergence results assuming that the sufficient descent condition holds. On the other hand, Dai et al. showed in [14] that the convergence can be obtained by assuming simple descent. In the present work, we retrieve the results of [14].

We begin by presenting two lemmas. In the first we show that if there exists $\gamma > 0$ such that $\|v(x^k)\| \geq \gamma$ for all $k \geq 0$, then the search directions change slowly, asymptotically. In the second lemma, we demonstrate that, in this case, if the method has Property $(*)$, then a certain fraction of the step sizes cannot be too small.

LEMMA 5.8. *Suppose that Assumptions* (A1) *and* (A3) *hold. Consider an NLCG algorithm where $\beta_k \geq 0$, d^k is a K-descent direction of F at x^k, and α_k satisfies the strong Wolfe conditions* (8). *In addition, suppose that there exists $\gamma > 0$ such that $\|v(x^k)\| \geq \gamma$ for all $k \geq 0$. Then*

(i) $\sum_{k\geq 0} \frac{\|v(x^k)\|^4}{\|d^k\|^2} < \infty$,

(ii) $\sum_{k\geq 1} \|u^k - u^{k-1}\|^2 < \infty$, *where $u^k = d^k/\|d^k\|$.*

Proof. Since d^k is a K-descent direction of F at x^k, it turns out that $d^k \neq 0$. Hence, $\|v(x^k)\|^4/\|d^k\|^2$ and u^k are well defined. The proof of part (i) can be obtained from the proof of Theorem 4.2(ii).

Now consider part (ii). Define $r^k = v(x^k)/\|d^k\|$ and $\delta^k = \beta_k\|d_{k-1}\|/\|d_k\|$. Observe that $u^k = r^k + \delta_k u^{k-1}$. Proceeding as in [26, Lemma 4.1] we obtain $\|u^k - u^{k-1}\| \leq 2\|r^k\|$. Therefore, using part (i), we have

$$\gamma^2 \sum_{k\geq 1} \|u^k - u^{k-1}\|^2 \leq 4\gamma^2 \sum_{k\geq 1} \|r^k\|^2 \leq 4 \sum_{k\geq 1} \|r^k\|^2 \|v(x^k)\|^2 = 4 \sum_{k\geq 1} \frac{\|v(x^k)\|^4}{\|d^k\|^2} < \infty,$$

concluding the proof. ∎

For $\lambda > 0$ and a positive integer Δ, define

$$\mathcal{K}^\lambda_{k,\Delta} = \{i \in \mathbb{N} \mid k \leq i \leq k + \Delta - 1, \|s^{k-1}\| > \lambda\}$$

and denote by $|\mathcal{K}^\lambda_{k,\Delta}|$ the number of elements of $\mathcal{K}^\lambda_{k,\Delta}$.

LEMMA 5.9. *Suppose that Assumptions* (A1) *and* (A3) *hold. Consider an NLCG algorithm where d^k is a K-descent direction of F at x^k, α_k satisfies the strong Wolfe conditions* (8), *and assume that the method has Property* $(*)$. *If there exists $\gamma > 0$ such that $\|v(x^k)\| \geq \gamma$, for all $k \geq 0$, then there exists $\lambda > 0$ such that, for any $\Delta \in \mathbb{N}$ and any index k_0, there is a greater index $k \geq k_0$ such that*

$$|\mathcal{K}^\lambda_{k,\Delta}| > \frac{\Delta}{2}.$$

Proof. Using Lemma 2.4(c), we have

$$\|d^l\|^2 \leq (\|v(x^l)\| + |\beta_l|\|d^{l-1}\|)^2 \leq 2\|v(x^l)\|^2 + 2\beta_l^2\|d^{l-1}\|^2 \leq 2\bar{\gamma}^2 + 2\beta_l^2\|d^{l-1}\|^2$$

for all $l \in \mathbb{N}$. Then, we can proceed by contradiction as in [26, Lemma 4.2] to show that there exists k_0 such that

$$\|d^l\|^2 \leq \bar{c}(l - k_0 + 2) \tag{28}$$

for any index $l \geq k_0 + 1$, where $\bar{c}$ is a certain positive constant independent of l. On the other hand, from Lemma 5.8(i), we have

$$\gamma^4 \sum_{k\geq 0} \frac{1}{\|d^k\|^2} \leq \sum_{k\geq 0} \frac{\|v(x^k)\|^4}{\|d^k\|^2} < \infty,$$

contradicting (28) and concluding the proof. ∎

THEOREM 5.10. *Suppose that Assumptions* (A1) *and* (A3) *hold. Consider an NLCG algorithm where $\beta_k \geq 0$, d^k is a K-descent direction of F at x^k, α_k satisfies the strong Wolfe conditions* (8), *and assume that the method has Property* $(*)$. *Then* $\liminf_{k\to\infty} \|v(x^k)\| = 0$.

Proof. Noting that $\{x^k\}$ belongs to the bounded set $\mathcal{L}$, and using Lemmas 5.8 and 5.9, the proof is by contradiction exactly as in [26, Theorem 4.3]. ∎

Now we can prove the convergence result related to the PRP and HS parameters.

THEOREM 5.11. *Suppose that Assumptions* (A1) *and* (A3) *hold. Consider an NLCG algorithm with*

$$\beta_k = \max\{\beta_k^{PRP}, 0\} \quad \text{or} \quad \beta_k = \max\{\beta_k^{HS}, 0\}.$$

If α_k satisfies the strong Wolfe conditions (8), *and d^k is a K-descent direction of F at x^k, then* $\liminf_{k\to\infty} \|v(x^k)\| = 0$.

Proof. First, observe that if an NLCG algorithm with β_k has Property (∗), so does an NLCG algorithm with $\max\{\beta_k, 0\}$. Therefore, by Theorem 5.10, it is sufficient to prove that PRP and HS methods have Property (∗). Assume that (27) holds for all $k \geq 0$. Then

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

Since d^k satisfies the sufficient descent condition (14), the relation (21) holds. Hence,

$$\frac{1}{2}\gamma^2 \leq -f(x^k, v(x^k)) \leq \bar{c}\bar{\gamma}, \tag{29}$$

where $\bar{c}$ is such that $\|JF(x^k)\| \leq \bar{c}$ for all $k \geq 0$. Observe that there exists $\bar{w} \in C$ such that

$$\left|f(x^{k-1}, v(x^k))\right| = \left|\langle JF(x^{k-1})v(x^k), \bar{w}\rangle\right| \leq \|JF(x^{k-1})\|\|v(x^k)\| \leq \bar{c}\bar{\gamma}, \tag{30}$$

because $\|\bar{w}\| = 1$. Moreover, from Lemma 2.3 together with Assumption (A1), we have

$$\left|-f(x^k, v(x^k)) + f(x^{k-1}, v(x^k))\right| \leq L\|v(x^k)\|\|x^k - x^{k-1}\| \leq L\lambda\bar{\gamma}, \tag{31}$$

when $\|s^{k-1}\| \leq \lambda$.

For the PRP method, define $b = 4\bar{c}\bar{\gamma}/\gamma^2$ and $\lambda = \gamma^2/(4L\bar{\gamma}b)$. By (29) and (30) we have

$$\left|\beta_k^{PRP}\right| \leq \frac{-f(x^k, v(x^k)) + \left|f(x^{k-1}, v(x^k))\right|}{-f(x^{k-1}, v(x^{k-1}))} \leq \frac{4\bar{c}\bar{\gamma}}{\gamma^2} = b,$$

and, when $\|s^{k-1}\| \leq \lambda$, it follows from (29) and (31) that

$$\left|\beta_k^{PRP}\right| \leq \frac{2L\lambda\bar{\gamma}}{\gamma^2} = \frac{1}{2b}.$$

Therefore, the PRP method has Property (∗).

Now consider the HS method. By the sufficient descent condition (14) and the Wolfe condition (8b) we obtain

$$\begin{aligned} f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1}) &\geq \sigma f(x^{k-1}, d^{k-1}) - f(x^{k-1}, d^{k-1}) \\ &= -(1-\sigma) f(x^{k-1}, d^{k-1}) \\ &\geq -c(1-\sigma) f(x^{k-1}, v(x^{k-1})) \\ &\geq \frac{c(1-\sigma)\gamma^2}{2} > 0. \end{aligned} \tag{32}$$

Define $b = 4\bar{c}\bar{\gamma}/[c(1-\sigma)\gamma^2]$ and $\lambda = c(1-\sigma)\gamma^2/(4L\bar{\gamma}b)$. By (29), (30), and (32),

$$\left|\beta_k^{HS}\right| \leq \frac{-f(x^k, v(x^k)) + \left|f(x^{k-1}, v(x^k))\right|}{f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1})} \leq \frac{4\bar{c}\bar{\gamma}}{c(1-\sigma)\gamma^2} = b.$$

Now, if $\|s^{k-1}\| \leq \lambda$, (31) and (32) imply

$$\left|\beta_k^{HS}\right| \leq \frac{2L\lambda\bar{\gamma}}{c(1-\sigma)\gamma^2} = \frac{1}{2b},$$

and we conclude that the HS method has Property $(*)$. ◻

6. Numerical results. Now we present some numerical results to illustrate the NLCG algorithm for the different choices of parameter β_k described in section 5. All considered problems are from the multiobjective optimization literature. Thus, in this section, we assume that $K = \mathbb{R}^m_+$, C is the canonical basis of $\mathbb{R}^m$, and $e = [1, \ldots, 1]^T \in \mathbb{R}^m$ in Definition 3.1.

For computing the steepest descent direction $v(x)$ we solve subproblem (5) using Algencan [4], an augmented Lagrangian code for general nonlinear programming.

We implemented a line search procedure that finds step sizes satisfying the strong Wolfe conditions (8). Let us briefly describe this algorithm without attempting to go into details. Let d be a K-descent direction for F at x. Since $f(x, d) < 0$, by the definition of f, we have that direction d is a descent direction for each scalar function $F_i(x)$, where $i = 1, \ldots, m$. The line search algorithm works on a single scalar function separately. Using the algorithm of Moré and Thuente [45], we first find a step size $\alpha > 0$ satisfying

$$\begin{aligned} F_i(x + \alpha d) &\leq F_i(x) + \rho\alpha f(x, d), \\ \left|\langle \nabla F_i(x + \alpha d), d\rangle\right| &\leq \sigma\left|f(x, d)\right| \end{aligned} \tag{33}$$

for an index $i \in \{1, \ldots, m\}$. If α also satisfies the vector-strong Wolfe conditions (8), we terminate the line search. Otherwise, we identify another scalar function $F_j(x)$, $j \in \{1, \ldots, m\}$, for which the interval $(0\ \alpha)$ contains a step size satisfying (33) for $F_j(x)$. Again, we use the algorithm of Moré and Thuente on $(0\ \alpha)$ for $F_j(x)$. This process is repeated until a step size satisfying the strong Wolfe conditions (8) is found. In each iteration k, the choice of the initial trial value for the step size is important for the performance of the line search. For $k = 0$ we used $1/\|v(x^0)\|$, and for the subsequent iterations we set it to

$$\alpha_{k-1}\frac{f(x^{k-1}, d^{k-1})}{f(x^k, d^k)}.$$

This is the vector extension of the choice recommended by Shanno and Phua [52] for conjugate gradient methods. We also observe that the line search procedure gives, in the limit, a point x^k with $f(x^k, d^{k-1}) = 0$ provided that d^{k-1} is a K-descent direction of F at x^{k-1}. Thus, as discussed in section 4, it can be used to enforce the sufficient descent condition (14). In our implementation we set $\rho = 10^{-4}$, $\sigma = 0.1$ in Definition 3.1, and $c = 0.1$ in (14). The values of the constants related to the Wolfe conditions are equal to those adopted in [26] for testing conjugate gradient methods in the scalar minimization case.

We stopped the execution of the NLCG algorithm at x^k, declaring convergence if

$$\theta(x^k) \geq -5 \times \texttt{eps}^{1/2},$$

where $\theta(x^k) = f(x^k, v(x^k)) + \|v(x^k)\|^2/2$ and `eps` denotes the machine precision given. This is the convergence criterion considered in the numerical tests of [20]. We set $\texttt{eps} = 2^{-52} \approx 2.22 \times 10^{-16}$ in our experiments. Since, by Lemma 2.2, $v(x) = 0$ if and only if $\theta(x) = 0$, this is a reasonable stopping criterion. The maximum number of allowed iterations was set to 10000. Codes are written in double precision Fortran 90 and are freely available at https://lfprudente.mat.ufg.br/.

6.1. Influence of endogenous parameters. In the previous section we proved that it is possible to obtain global convergence of the NLCG algorithm with

- $\beta_k = \delta\beta_k^{FR}$, where $0 \leq \delta < 1$ (Corollary 5.2);
- $\beta_k = \eta\beta_k^{CD}$, where $0 \leq \eta < 1 - \sigma$ (Theorem 5.4);
- $\beta_k = \eta\beta_k^{DY}$, where $0 \leq \eta < (1-\sigma)/(1+\sigma)$ (Theorem 5.6);
- $\beta_k = \beta_k^{mDY}$ with $\tau > 1$ (Theorem 5.7).

We start the numerical experiments by checking the influence of the endogenous parameters δ, η, and τ in the robustness of the FR, CD, DY, and mDY methods. Especially in relation to FR and mDY methods, it is natural to question whether we can set $\delta = 1$ and $\tau = 1$, respectively, in order to achieve convergence. These cases correspond to the "pure" methods of FR and DY, where $\beta_k = \beta_k^{FR}$ and $\beta_k = \beta_k^{DY}$.

We considered the following bicriteria problem $F\colon \mathbb{R}^n \to \mathbb{R}^2$ given by

$$F_1(x) = (x_1 - 1)^4 + \sum_{i=2}^{n} (x_i - 1)^2,$$
$$F_2(x) = (x_2 + 1)^4 + \sum_{i=1, i\neq 2}^{n} (x_i + 1)^2.$$

This problem corresponds to the second example of [51] and we will call it SLC2. SLC2 is a simple convex problem that should not pose major challenges to a global convergent NLCG algorithm. We combined each parameter β_k with different values of δ, η, and τ to generate different methods. In all cases we set $n = 100$ and run the NLCG algorithm 200 times using starting points from a uniform random distribution between $[-100\ 100]^n$. Table 1 reports the percentage of runs that reached a critical point for each combination.

TABLE 1

Percentages of SLC2 problem instances solved for each NLCG algorithm. For each combination we set $n = 100$ and run the algorithm 200 times using starting points from a uniform random distribution between $[-100\ 100]^n$.

FR	
δ	%
1.00	64.5
0.99	97.0
0.98	100.0

CD	
η	%
1.00	72.0
0.99	95.0
0.98	97.0
0.97	99.5
0.96	100.0
$1-\sigma$	100.0

DY	
η	%
1.00	70.0
0.99	96.0
0.98	99.5
0.97	100.0
$\frac{1-\sigma}{1+\sigma}$	100.0

mDY	
τ	%
1.00	70.0
1.01	97.0
1.02	99.0
1.03	100.0

The FR method solved 64.5% of the problems when $\delta = 1$, and 97.0% when $\delta = 0.99$. All runs reached a critical point when $\delta = 0.98$. Similar behavior was observed for the mDY method: while 97.0% of the problems were solved with $\tau = 1.01$, only 70% were solved with $\tau = 1$. From a theoretical point of view, despite the practical performance of the FR and mDY methods in this example, it remains an open question whether it is possible to ensure global convergence for the "pure" FR and DY methods. With respect to the CD method, 100% of the problems were solved when $\eta = 0.96$. The same robustness was achieved when $\eta = 1 - \sigma = 0.90$. The DY method with $\eta = 1$ corresponds to the mDY method with $\tau = 1$. Thus, in this case, the robustness of the method was 70%. For $\eta = (1-\sigma)/(1+\sigma) \approx 0.82$, the DY method reached a critical point in all runs.

Based on the numerical results presented in Table 1 and in agreement with the convergence theory, in the following numerical experiments we fixed

- $\beta_k = 0.98\beta_k^{FR}$ for the FR method,
- $\beta_k = 0.99(1-\sigma)\beta_k^{CD}$ for the CD method,
- $\beta_k = 0.99(1-\sigma)/(1+\sigma)\beta_k^{DY}$ for the DY method,
- $\beta_k = \beta_k^{mDY}$ with $\tau = 1.02$ for the mDY method.

We also refer by PRP+ and HS+ to the NLCG algorithm with $\beta_k = \max\{\beta_k^{PRP}, 0\}$ and $\beta_k = \max\{\beta_k^{HS}, 0\}$, respectively.

6.2. Multiobjective problem instances. We tested the NLCG algorithm with different choices of parameter β_k in several multiobjective problem instances found in the literature. The results are in Tables 2 and 3. All problems in Table 2 are convex whereas the problems in Table 3 are nonconvex. The names of the problems are those used in the corresponding references. Alternatively, in order to identify a problem, we use the authors names followed by a number to indicate the problem in the corresponding reference (e.g., AP1 corresponds to the first problem of the work [2] of Ansary and Panda). The first column also gives the number of variables n and the number of objectives m of the problem. Problems AP1, AP4, FDS, MOP5, and MOP7 are tricriteria whereas the other ones are bicriteria. Many problems have box constraints in their definitions. In these cases, we simply take the starting point inside the corresponding box and ignore the constraints.

All problems were solved 200 times using starting points from a uniform random distribution inside a box specified in the first column of the tables. Tables 2 and 3 inform for each method the percentage of runs that reached a critical point (%), and for the successful runs the median of number of iterations (it), the median of functions evaluations (evalf), and the median of gradient evaluations (evalg). Thus, for a given method/problem combination, the reported data represent a typical successful run. We point out that we considered each evaluation of an objective (objective gradient) in the calculation of evalf (evalg).

JOS1 is a simple convex quadratic test problem that was solved with only one steepest descent iteration regardless of the starting point. Other problems that did not present major challenges were Lov1, Lov3, Lov4, MOP5, MOP7, SLC1, and SP1. On the other hand, the NLCG algorithms presented a poor performance in the VU1 problem. In several runs of this problem the line search procedure generated excessively small step sizes impairing the convergence of the NLCG methods.

In almost all instances, the methods have reached a critical point in all runs, which is consistent with the theoretical results. Overall, the PRP+ and HS+ methods were equivalent and presented superior performance. The FR and mDY methods are clearly the least efficient, requiring an exceedingly large numbers of iterations and function/gradient evaluations in various problems. In intermediate positions, the DY

Table 2
Performance of the NLCG algorithms in a collection of convex multiobjective problem instances.

		FR	CD	DY	mDY	PRP+	HS+
AP1, [2]	%	96.0	97.0	98.0	96.5	97.5	97.5
$n = 2$	it	165.0	48.5	33.5	167.0	11.0	11.0
$m = 3$	evalf	1672.5	537.5	372.0	1693.0	119.0	117.0
$x^0 \in [-100\ 100]^n$	evalg	1358.5	486.0	335.0	1375.0	99.0	99.0
AP4, [2]	%	89.0	90.5	91.0	91.0	93.5	93.5
$n = 3$	it	797.0	166.0	101.0	803.0	19.0	18.0
$m = 3$	evalf	7210.0	1513.0	1106.0	7242.5	203.0	203.0
$x^0 \in [-100\ 100]^n$	evalg	6394.0	1350.0	912.0	6433.0	181.0	181.0
FDS, [20]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 50$	it	1959.0	415.0	215.0	1997.0	46.0	46.0
$m = 3$	evalf	19624.5	3195.5	1879.5	20004.5	507.0	507.0
$x^0 \in [-2\ 2]^n$	evalg	15741.5	3060.5	1761.5	16045.5	462.5	462.5
JOS1, [36]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 1000$	it	1.0	1.0	1.0	1.0	1.0	1.0
$m = 2$	evalf	18.0	18.0	18.0	18.0	18.0	18.0
$x^0 \in [-10^4\ 10^4]^n$	evalg	20.0	20.0	20.0	20.0	20.0	20.0
Lov1, [41]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 2$	it	3.0	3.0	3.0	3.0	3.0	3.0
$m = 2$	evalf	22.0	22.0	22.0	22.0	29.0	29.0
$x^0 \in [-100\ 100]^n$	evalg	21.0	21.0	21.0	21.0	26.5	26.5
MOP7, [36]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 2$	it	8.0	8.0	8.0	8.0	3.0	3.0
$m = 3$	evalf	105.5	110.5	111.0	107.0	52.0	52.0
$x^0 \in [-400\ 400]^n$	evalg	94.0	99.0	100.0	94.0	51.0	51.0
SLC2, [51]	%	100.0	100.0	100.0	99.0	100.0	100.0
$n = 100$	it	128.0	34.0	29.5	96.0	20.0	21.0
$m = 2$	evalf	1000.5	296.5	260.0	720.0	200.5	204.5
$x^0 \in [-100\ 100]^n$	evalg	830.0	267.0	230.5	650.5	178.5	185.5
SP1, [36]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 2$	it	7.0	7.0	8.0	7.0	3.0	3.0
$m = 2$	evalf	62.0	62.0	63.5	62.0	28.0	28.0
$x^0 \in [-100\ 100]^n$	evalg	56.0	56.0	57.0	56.0	30.0	30.0

method appears to be better than the CD method.

Let us take a detailed look at some test problems. Hil1 [34] is given by $F\colon \mathbb{R}^2 \to \mathbb{R}^2$, where

$$F_1(x) = \cos(a(x)) \times b(x),$$
$$F_2(x) = \sin(a(x)) \times b(x),$$

with

$$a(x) = (2\pi/360)\left[45 + 40\sin(2\pi x_1) + 25\sin(2\pi x_2)\right],$$
$$b(x) = 1 + 0.5\cos(2\pi x_1).$$

Each objective is a periodic function of period 1 with respect to both arguments. Figure 1(a) was obtained by discretizing the domain $[0\ 1] \times [0\ 1]$ to a fine grid and plotting all the image points. Thus, it provides a good representation of the image space of F, and gives us a geometric notion of the Pareto front. The value space generated by the PRP+ method using 200 starting points randomly generated at $[0\ 1] \times [0\ 1]$ can be seen in Figure 1(b). A full point represents a final iterate while the beginning of a straight segment represents the corresponding starting point. As can be seen from Figure 1(b), given a reasonable number of starting points, the NLCG algorithm was able to estimate the Pareto front of problem Hil1.

Table 3

Performance of the NLCG algorithms in a collection of nonconvex multiobjective problem instances.

		FR	CD	DY	mDY	PRP+	HS+
AP3, [2]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 2$	it	55.5	27.5	30.0	24.0	11.0	9.5
$m = 2$	evalf	477.0	256.5	264.0	230.5	136.5	119.0
$x^0 \in [-100\ 100]^n$	evalg	406.5	225.5	236.5	198.0	123.5	107.0
Far1, [36]	%	97.5	100.0	100.0	97.5	100.0	100.0
$n = 2$	it	1184.0	258.0	158.5	1205.0	49.0	48.5
$m = 2$	evalf	9297.0	1878.5	1146.0	9473.0	375.0	366.5
$x^0 \in [-1\ 1]^n$	evalg	7504.0	1600.5	964.5	7648.0	347.5	339.0
FF1, [36]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 2$	it	292.5	74.5	46.0	296.5	13.0	13.0
$m = 2$	evalf	2212.5	556.0	360.5	2242.0	103.5	103.5
$x^0 \in [-1\ 1]^n$	evalg	1814.0	524.0	339.0	1836.5	87.0	87.0
Hil1, [34]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 2$	it	186.5	53.0	35.0	189.5	11.5	11.5
$m = 2$	evalf	1498.5	424.0	272.5	1522.5	95.0	96.5
$x^0 \in [0\ 1]^n$	evalg	1133.0	358.0	270.0	1151.0	80.5	81.0
Lov3, [41]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 2$	it	3.0	3.0	3.0	3.0	3.0	3.0
$m = 2$	evalf	21.0	21.0	21.0	21.0	21.0	21.0
$x^0 \in [-100\ 100]^n$	evalg	20.0	20.0	20.0	20.0	20.0	20.0
Lov4, [41]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 2$	it	2.0	2.0	2.0	2.0	2.0	2.0
$m = 2$	evalf	14.0	14.0	14.0	14.0	15.0	15.0
$x^0 \in [-100\ 100]^n$	evalg	14.0	14.0	14.0	14.0	14.0	14.0
MLF2, [36]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 2$	it	1321.0	266.5	146.5	1315.5	37.0	37.0
$m = 2$	evalf	10580.5	2064.0	1033.5	10546.5	313.5	311.5
$x^0 \in [-100\ 100]^n$	evalg	7970.5	1891.5	1027.0	7955.5	308.0	306.5
MMR1[1], [44]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 2$	it	49.5	25.0	18.0	49.5	8.0	8.0
$m = 2$	evalf	400.0	202.5	148.0	400.0	65.0	65.0
$x^0 \in [0\ 1]^n$	evalg	310.0	176.0	130.5	310.0	51.0	51.0
MMR5, [44]	%	88.0	100.0	100.0	87.0	100.0	100.0
$n = 100$	it	6501.0	1363.0	809.5	6639.5	282.0	281.0
$m = 2$	evalf	47595.0	7399.5	3571.5	47373.5	1920.5	1789.5
$x^0 \in [-5\ 5]^n$	evalg	43742.5	7168.0	3538.5	43274.5	1825.0	1718.0
MOP2, [36]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 2$	it	86.0	33.0	23.0	86.0	9.0	9.0
$m = 2$	evalf	688.5	265.0	185.0	688.5	71.5	71.5
$x^0 \in [-1\ 1]^n$	evalg	596.0	240.0	165.5	594.0	59.5	59.5
MOP3, [36]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 2$	it	17.0	13.0	12.0	17.0	7.0	7.0
$m = 2$	evalf	134.0	100.0	91.0	140.0	68.0	68.0
$x^0 \in [-\pi\ \pi]^n$	evalg	121.0	89.5	80.0	122.0	61.0	61.0
MOP5, [36]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 2$	it	2.0	2.0	2.0	2.0	2.0	2.0
$m = 3$	evalf	21.0	21.0	21.0	21.0	21.0	21.0
$x^0 \in [-1\ 1]^n$	evalg	22.0	22.0	22.0	22.0	22.0	22.0
SK2, [36]	%	99.5	100.0	100.0	99.5	100.0	100.0
$n = 4$	it	1306.0	224.5	125.0	1082.0	34.5	34.0
$m = 2$	evalf	9151.0	1729.0	849.5	7588.0	271.0	269.0
$x^0 \in [-10\ 10]^n$	evalg	7847.0	1523.5	774.0	6509.0	239.5	237.5
SLC1, [51]	%	100.0	100.0	100.0	100.0	100.0	100.0
$n = 2$	it	2.0	2.0	2.0	2.0	2.0	2.0
$m = 2$	evalf	20.0	20.0	20.0	20.0	20.0	20.0
$x^0 \in [-5\ 5]^n$	evalg	21.0	21.0	21.0	21.0	22.0	22.0
VU1, [36]	%	16.0	50.5	67.0	16.0	100.0	100.0
$n = 2$	it	4304.5	3813.0	2683.0	4388.0	951.5	951.5
$m = 2$	evalf	34436.0	30504.0	20735.0	35104.0	4246.0	4246.0
$x^0 \in [-3\ 3]^n$	evalg	25841.0	23380.0	20716.5	26342.0	4242.0	4242.0

[1] We consider a modified version of this problem. We set $F_1(x) = 1 + x_1^2$ and $F_2(x) = \Psi(x_2)/F_1(x)$, where $\Psi(x_2)$ is defined in [44].

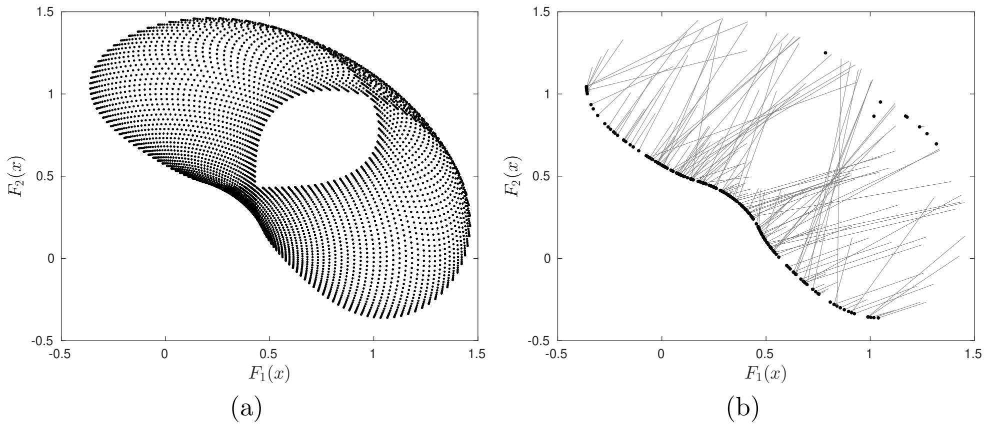

FIG. 1. (a) *Image set of problem Hil*1. (b) *Value space for problem Hil*1 *generated by the PRP+ method.* 200 *starting points were used. A full point represents a final iterate while the beginning of a straight segment represents the corresponding starting point.*

Consider problem MMR5 [44] given by

$$F_1(x) = \left(\frac{1}{n} \sum_{k=1}^{n} \left[x_i^2 - 10\cos(2\pi x_i) + 10 \right] \right)^{1/4},$$

$$F_2(x) = \left(\frac{1}{n} \sum_{k=1}^{n} \left[(x_i - 1.5)^2 - 10\cos(2\pi (x_i - 1.5)) + 10 \right] \right)^{1/4}.$$

We tested the FR and HS+ methods, varying the domain of the starting points. We set $n = 100$ and chose the starting points from a uniform random distribution belonging to the box indicated in the first column of Table 4. Each instance was run 200 times.

TABLE 4
*Performance of the FR and HS+ methods on problem MMR*5*, varying the domain of the starting points.*

	FR				HS+			
x^0	%	it	evalf	evalg	%	it	evalf	evalg
$[-5\ 5]^n$	88.0	6501.0	47557.0	43742.5	100.0	281.0	1789.5	1718.0
$[-50\ 50]^n$	100.0	2865.5	17306.0	16732.0	100.0	135.0	959.5	903.5
$[-500\ 500]^n$	100.0	3545.5	24254.5	22819.5	100.0	159.0	1161.0	1106.0
$[-1000\ 1000]^n$	100.0	3018.0	19442.5	18590.0	100.0	161.5	1182.5	1121.5
$[-2000\ 2000]^n$	100.0	2858.0	18109.0	17477.0	100.0	162.5	1160.5	1104.0

For this problem, Table 4 indicates that the methods are robust with respect to the choice of the starting points. Observe that, especially for the FR method, in some instances the NLCG algorithm presented superior performance when the box was extended.

Finally, we tested the PRP+ method in larger instances of problems FDS, MMR5, and SLC2, varying the dimensions. For each problem and value of n we ran the methods 200 times using starting points from a uniform random distribution belonging to the corresponding boxes, as reported in Tables 2 and 3. The results are in Table 5. As we can see, the PRP+ method solved large instances of these problems. FDS is given by

$$F_1(x) = \frac{1}{n^2}\sum_{i=1}^{n} i(x_i - i)^4,$$

$$F_2(x) = \exp\left(\sum_{i=1}^{n}\frac{x_i}{n}\right) + \|x\|_2^2,$$

$$F_3(x) = \frac{1}{n(n+1)}\sum_{i=1}^{n} i(n-i+1)e^{-x_i}.$$

In [20], the authors said that the numerical difficulty of this problem sharply increases with the dimension n. For comparative purposes, the authors reported in [20] that the proposed Newton method reached the maximum number of iterations allowed (500 iterations) in 166 of the 200 runs when $n = 200$. This suggests that the NLCG algorithm is potentially able to solve large and difficult problems.

TABLE 5

Performance of PRP+ method on problems FDS, MMR5, and SLC2, varying the dimensions.

	FDS			
n	%	it	evalf	evalg
200	100.0	69.0	760.0	693.0
500	100.0	79.0	870.0	793.0
1000	100.0	85.0	936.0	853.0
2000	100.0	91.0	1002.0	913.0
4000	100.0	96.0	1057.0	963.0
5000	100.0	98.0	1079.0	983.0

	MMR5			
n	%	it	evalf	evalg
200	100.0	263.0	1775.5	1694.0
500	100.0	140.5	928.5	888.0
1000	100.0	118.0	811.0	770.0
2000	100.0	26.0	279.5	262.5
4000	100.0	33.0	349.5	347.0
5000	100.0	30.0	324.0	320.5

	SLC2			
n	%	it	evalf	evalg
200	100.0	24.0	227.0	205.0
500	100.0	28.0	248.5	225.0
1000	100.0	33.0	283.0	263.0
2000	100.0	38.0	338.0	306.5
4000	100.0	52.0	425.5	404.5
5000	100.0	39.5	360.0	321.0

7. Final remarks. In this work, we have proposed nonlinear conjugate gradient methods for solving unconstrained vector optimization problems. Throughout the paper all the assumptions are natural extensions of those made for the scalar minimization case. The vector extensions of the Fletcher–Reeves, conjugate descent, Dai–Yuan, Polak–Ribière–Polyak, and Hestenes–Stiefel parameters were considered. In particular, we showed that it is possible to obtain global convergence of the NLCG algorithm with any fraction of the FR parameter. However, whether the FR method, with inexact line searches, automatically generates descent directions remains an open question. We also showed that if the parameter β_k is nonnegative and bounded above by an appropriate fraction of the CD parameter, global convergence can be achieved.

With respect to the PRP and HS parameters, all convergence results present in [26] for the classical optimization were retrieved. In addition, we proposed a slight modification in DY parameter for which it was possible to show global convergence. The convergence analyses were made assuming inexact line searches and without regular restarts.

We implemented and tested the NLCG algorithm with the different choices of parameter β_k in several multiobjective problem instances found in the literature. Convex and nonconvex problems were considered. In agreement with the theoretical results, the numerical experiments indicate that the considered methods are robust on the chosen set of test problems. In a particular instance, we also investigated the ability of the method to estimate the Pareto front. The numerical results presented are the first steps towards verifying the practical reliability of the NLCG algorithm. We intend to study the NLCG algorithm from the point of view of actual implementations and provide comparisons with alternative approaches.

Regarding the line search procedure, we have introduced the standard and strong Wolfe conditions for vector optimization, and showed the existence of intervals of step sizes satisfying them. Moreover, Proposition 3.2 sheds light on algorithmic properties and suggests the implementation of a Wolfe-type line search procedure. We also introduced the Zoutendjik condition for vector optimization and proved that it is satisfied for a general descent line search method. As far as we know, this is the first paper in which this type of result has been presented in the vector optimization context. We expect that the proposed extensions of Wolfe and Zoutendjik conditions can be used in the convergence analysis of other methods.

It is worth mentioning that the positiveness of parameter β_k seems to be essential for obtaining methods that generate descent directions. Lemma 4.1 and the technique used in its proof corroborate this intuition. In [31], Hager and Zhang proposed an efficient conjugate gradient method for scalar minimization for which the parameter β_k can be negative. A challenging problem is to extend their work for the vector-valued optimization.

REFERENCES

[1] M. Al-Baali, *Descent property and global convergence of the Fletcher–Reeves method with inexact line search*, IMA J. Numer. Anal., 5 (1985), pp. 121–124.

[2] M. A. Ansary and G. Panda, *A modified quasi-Newton method for vector optimization problem*, Optimization, 64 (2015), pp. 2289–2306.

[3] J. Y. Bello Cruz, *A subgradient method for vector optimization problems*, SIAM J. Optim., 23 (2013), pp. 2169–2182.

[4] E. G. Birgin and J. M. Martínez, *Practical Augmented Lagrangian Methods for Constrained Optimization*, Fundam. Algorithms 10, SIAM, Philadelphia, PA, 2014.

[5] H. Bonnel, A. N. Iusem, and B. F. Svaiter, *Proximal methods in vector optimization*, SIAM J. Optim., 15 (2005), pp. 953–970.

[6] L. C. Ceng, B. S. Mordukhovich, and J. C. Yao, *Hybrid approximate proximal method with auxiliary variational inequality for vector optimization*, J. Optim. Theory Appl., 146 (2010), pp. 267–303.

[7] L. C. Ceng and J. C. Yao, *Approximate proximal methods in vector optimization*, European J. Oper. Res., 183 (2007), pp. 1–19.

[8] T. D. Chuong, *Generalized proximal method for efficient solutions in vector optimization*, Numer. Funct. Anal. Optim., 32 (2011), pp. 843–857.

[9] T. D. Chuong, *Newton-like methods for efficient solutions in vector optimization*, Comput. Optim. Appl., 54 (2013), pp. 495–516.

[10] T. D. Chuong, B. S. Mordukhovich, and J. C. Yao, *Hybrid approximate proximal algorithms for efficient solutions in vector optimization*, J. Nonlinear Convex Anal., 12 (2011), pp. 257–285.

[11] T. D. Chuong and J. C. Yao, *Steepest descent methods for critical points in vector optimization problems*, Appl. Anal., 91 (2012), pp. 1811–1829.

[12] Y. H. Dai, *Convergence analysis of nonlinear conjugate gradient methods*, in Optimization and Regularization for Computational Inverse Problems and Applications, Y. Wang, A. G. Yagola, and C. Yang, eds., Springer, Berlin, 2011, pp. 157–181.

[13] Y. H. Dai, *Nonlinear Conjugate Gradient Methods*, John Wiley, New York, 2010.

[14] Y. H. Dai, J. Han, G. Liu, D. Sun, H. Yin, and Y.-X. Yuan, *Convergence properties of nonlinear conjugate gradient methods*, SIAM J. Optim., 10 (1999), pp. 345–358.

[15] Y. H. Dai and Y.-X. Yuan, *Convergence properties of the conjugate descent method*, Adv. Math., 25 (1996), pp. 552–562.

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

[17] P. De, J. B. Ghosh, and C. E. Wells, *On the minimization of completion time variance with a bicriteria extension*, Oper. Res., 40 (1992), pp. 1148–1155.

[18] R. Fletcher, *Unconstrained Optimization*, Pract. Methods Optim. 1, John Wiley, New York, 1980.

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

[20] J. Fliege, L. M. Graña Drummond, and B. F. Svaiter, *Newton's method for multiobjective optimization*, SIAM J. Optim., 20 (2009), pp. 602–626.

[21] J. Fliege and B. F. Svaiter, *Steepest descent methods for multicriteria optimization*, Math. Methods Oper. Res., 51 (2000), pp. 479–494.

[22] J. Fliege and L. N. Vicente, *Multicriteria approach to bilevel optimization*, J. Optim. Theory Appl., 131 (2006), pp. 209–225.

[23] J. Fliege and R. Werner, *Robust multiobjective optimization & applications in portfolio optimization*, European J. Oper. Res., 234 (2014), pp. 422 – 433.

[24] E. H. Fukuda and L. M. Graña Drummond, *On the convergence of the projected gradient method for vector optimization*, Optimization, 60 (2011), pp. 1009–1021.

[25] E. H. Fukuda and L. M. Graña Drummond, *Inexact projected gradient method for vector optimization*, Comput. Optim. Appl., 54 (2013), pp. 473–493.

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

[27] L. M. Graña Drummond and A. N. Iusem, *A projected gradient method for vector optimization problems*, Comput. Optim. Appl., 28 (2004), pp. 5–29.

[28] L. M. Graña Drummond, F. M. P. Raupp, and B. F. Svaiter, *A quadratically convergent Newton method for vector optimization*, Optimization, 63 (2014), pp. 661–677.

[29] L. M. Graña Drummond and B. F. Svaiter, *A steepest descent method for vector optimization*, J. Comput. Appl. Math., 175 (2005), pp. 395–414.

[30] M. Gravel, J. M. Martel, R. Nadeau, W. Price, and R. Tremblay, *A multicriterion view of optimal resource allocation in job-shop production*, European J. Oper. Res., 61 (1992), pp. 230–244.

[31] W. W. Hager and H. Zhang, *A new conjugate gradient method with guaranteed descent and an efficient line search*, SIAM J. Optim., 16 (2005), pp. 170–192.

[32] W. W. Hager and H. Zhang, *A survey of nonlinear conjugate gradient methods*, Pac. J. Optim., 2 (2006), pp. 35–58.

[33] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bureau Standards, 49 (1952), pp. 409–436.

[34] C. Hillermeier, *Generalized homotopy approach to multiobjective optimization*, J. Optim. Theory Appl., 110 (2001), pp. 557–583.

[35] T. S. Hong, D. L. Craft, F. Carlsson, and T. R. Bortfeld, *Multicriteria optimization in intensity-modulated radiation therapy treatment planning for locally advanced cancer of the pancreatic head*, Internat. J. Radiation Oncology Biol. Phys., 72 (2008), pp. 1208–1214.

[36] S. Huband, P. Hingston, L. Barone, and L. While, *A review of multiobjective test problems and a scalable test problem toolkit*, IEEE Trans. Evol. Comput., 10 (2006), pp. 477–506.

[37] A. Hutterer and J. Jahn, *On the location of antennas for treatment planning in hyperthermia*, OR Spectrum, 25 (2003), pp. 397–412.

[38] J. Jahn, A. Kirsch, and C. Wagner, *Optimization of rod antennas of mobile phones*, Math. Methods Oper. Res., 59 (2004), pp. 37–51.

[39] A. Jüschke, J. Jahn, and A. Kirsch, *A bicriterial optimization problem of antenna design*, Comput. Optim. Appl., 7 (1997), pp. 261–276.

[40] T. M. Leschine, H. Wallenius, and W. A. Verdini, *Interactive multiobjective analysis and assimilative capacity-based ocean disposal decisions*, European J. Oper. Res., 56 (1992),

pp. 278–289.
[41] A. Lovison, *Singular continuation: Generating piecewise linear approximations to Pareto sets via global analysis*, SIAM J. Optim., 21 (2011), pp. 463–490.
[42] F. Lu and C.-R. Chen, *Newton-like methods for solving vector optimization problems*, Appl. Anal., 93 (2014), pp. 1567–1586.
[43] D. T. Luc, *Theory of Vector Optimization*, Lecture Notes in Econom. and Math. Systems 319, Springer, Berlin, 1989.
[44] E. Miglierina, E. Molho, and M. Recchioni, *Box-constrained multi-objective optimization: A gradient-like method without "a priori" scalarization*, European J. Oper. Res., 188 (2008), pp. 662–682.
[45] J. J. Moré and D. J. Thuente, *Line search algorithms with guaranteed sufficient decrease*, ACM Trans. Math. Softw., 20 (1994), pp. 286–307.
[46] J. Nocedal and S. Wright, *Numerical Optimization*, Springer Ser. Oper. Res. Financ. Eng., Springer, New York, 2006.
[47] E. Polak and G. Ribière, *Note sur la convergence de méthodes de directions conjuguées*, Rev. Française Inform. Rech. Opér. Sér. Rouge, 3 (1969), pp. 35–43.
[48] B. T. Polyak, *The conjugate gradient method in extremal problems*, USSR Comput. Math. Math. Phys., 9 (1969), pp. 94–112.
[49] M. J. D. Powell, *Nonconvex minimization calculations and the conjugate gradient method*, in Numerical Analysis, Lect. Notes Math. 1066, Springer, Berlin, 1984, pp. 122–141.
[50] S. Qu, M. Goh, and F. T. Chan, *Quasi-Newton methods for solving multiobjective optimization*, Oper. Res. Letters, 39 (2011), pp. 397–399.
[51] O. Schütze, A. Lara, and C. A. Coello, *The directed search method for unconstrained multi-objective optimization problems*, Technical report TR-OS-2010-01, http://delta.cs.cinvestav.mx/~schuetze/technical_reports/TR-OS-2010-01.pdf.gz, 2010.
[52] D. F. Shanno and K.-H. Phua, *Remark on algorithm* 500*: Minimization of unconstrained multivariate functions*, ACM Trans. Math. Software, 6 (1980), pp. 618–622.
[53] M. Tavana, *A subjective assessment of alternative mission architectures for the human exploration of Mars at NASA using multicriteria decision making*, Comput. & Oper. Res., 31 (2004), pp. 1147–1164.
[54] M. Tavana, M. A. Sodenkamp, and L. Suhl, *A soft multi-criteria decision analysis model with application to the European Union enlargement*, Ann. Oper. Res., 181 (2010), pp. 393–421.
[55] K. D. Villacorta and P. R. Oliveira, *An interior proximal method in vector optimization*, European J. Oper. Res., 214 (2011), pp. 485–492.
[56] D. White, *Epsilon-dominating solutions in mean-variance portfolio analysis*, European J. Oper. Res., 105 (1998), pp. 457–466.