# A study of Liu-Storey conjugate gradient methods for vector optimization

M. L. N. Gonçalves[a,*], F. S. Lima[a], L. F. Prudente[a]

[a] *Instituto de Matemática e Estatística, Universidade Federal de Goiás, Campus Samambaia - R. Jacarandá - Chácaras Califórnia, Goiânia, 74001-970, Goiás, Brazil*

---

**Abstract**

This work presents a study of Liu-Storey (LS) nonlinear conjugate gradient (CG) methods to solve vector optimization problems. Three variants of the LS-CG method originally designed to solve single-objective problems are extended to the vector setting. The first algorithm restricts the LS conjugate parameter to be nonnegative and use a sufficiently accurate line search satisfying the (vector) standard Wolfe conditions. The second algorithm combines a modification in the LS conjugate parameter with a line search satisfying the (vector) strong Wolfe conditions. The third algorithm consists of a combination of the LS conjugate parameter with a new Armijo-type line search (to be proposed here for the vector setting). Global convergence results and numerical experiments are presented.

*Keywords:* Vector optimization, conjugate gradient methods, global convergence, Pareto efficiency

---

## 1. Introduction

Let us first consider the single-objective problem

$$\text{Minimize } f(x), \quad x \in \mathbb{R}^n, \tag{1}$$

---

⋆This work was supported by CNPq (424860/2018-0 and 309628/2020-2) and FAPEG (PPP03/15-201810267001725).

*Corresponding author

*Email addresses:* maxlng@ufg.br (M. L. N. Gonçalves), fernandoslima@discente.ufg.br (F. S. Lima), lfprudente@ufg.br (L. F. Prudente)

where $f\colon \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. One of the most important methods to solve (1), belonging the class of the nonlinear conjugate gradient (CG) algorithms, is the Liu-Storey (LS) CG method proposed in [36]. Formally, the method generates a sequence $\{x^k\}$ given by

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

where the step size $\alpha_k > 0$ is obtained by a line search procedure and the search direction $d^k$ is defined by

$$d^k := \begin{cases} -\nabla f(x^k), & \text{if } k = 0 \\ -\nabla f(x^k) + \beta_k d^{k-1}, & \text{if } k \geq 1, \end{cases} \tag{2}$$

with

$$\beta_k = \beta_k^{LS} := -\frac{\langle \nabla f(x^k), \nabla f(x^k) - \nabla f(x^{k-1})\rangle}{\langle \nabla f(x^{k-1}), d^{k-1}\rangle}, \tag{3}$$

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product. Different choices for the conjugate parameter $\beta_k$ in (2) result in other CG methods. Although too many to name, some notable choices would include: Fletcher-Reeves (FR) [17]; Conjugate Descent (CD) [16]; Dai-Yuan (DY) [13]; Polak-Ribière-Polyak (PRP) [45]; Hestenes-Stiefel (HS) [30].

A distinguish feature of the FR, CD, and DY algorithms is that if a line search satisfying the Wolfe conditions is used, then the corresponding search directions satisfy the so-called *sufficient descent condition*, i.e,

$$\langle \nabla f(x^k), d^k \rangle \leq -c\|\nabla f(x^k)\|^2, \quad \forall k \geq 0,$$

where $c$ is a positive constant and $\|\cdot\|$ denotes the Euclidian norm. However, this is not the case for the PRP, HS, and LS algorithms, that is, these algorithms do not necessarily generate descent directions even when Wolfe line searches are employed. Even more, as shown in [46], the PRP, HS, and LS methods using exact line searches can cycle without approaching a solution of (1). In [23], by assuming that the search directions yield descent, the authors established the convergence of the CG method with $\beta_k := \max\{\beta_k^{PRP}, 0\}$, $\beta_k := \max\{\beta_k^{HS}, 0\}$, and $\beta_k := \max\{\beta_k^{LS}, 0\}$. It is noteworthy that the descent property can be obtained by performing a sufficiently accurate line search; see [23, Section 4] for a careful discussion of this issue. On the other hand, some modifications in the PRP, HS, and LS parameters have been proposed in order to build methods that automatically satisfy the descent property when a usual inexact line search is employed; see, for example, [44, 29, 35]. The algorithms obtained in [29] and [35] use the Wolfe line search conditions and are called Hager-Zhang (HZ-CG) and modified Liu-Storey, respectively. We also mentioned that a LS-CG method with an Armijo-type line search satisfying the sufficient descent condition was proposed in [52].

Consider now the unconstrained vector problem

$$\text{Minimize}_K \; F(x), \quad x \in \mathbb{R}^n, \tag{4}$$

where $F\colon \mathbb{R}^n \to \mathbb{R}^m$ is a continuously differentiable function, $K \subset \mathbb{R}^m$ is a closed, convex, and pointed cone with nonempty interior and the partial order in $\mathbb{R}^m$ induced by $K$ (resp. $\operatorname{int}(K)$), $\preceq_K$ (resp. $\prec_K$), is defined by

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

In vector optimization, we seek to find a *$K$-Pareto optimum* or *$K$-efficient* point of $F$, i.e., a point $x^* \in \mathbb{R}^n$ such that there is no $x \in \mathbb{R}^n$ with $F(x) \neq F(x^*)$ and $F(x) \preceq_K F(x^*)$. When $K = \mathbb{R}^m_+$, (4) corresponds to the multiobjective optimization problem.

Recently, some conjugate gradient methods of the scalar case were extended to the vector setting. The first work in this line was [40], which extended the FR, CD, DY, PRP, and HS conjugate gradient algorithms. Later, [25] proposed a HZ-CG method for vector optimization problems, whereas [20] studied a CG method for unconstrained quadratic multiobjective problems. Following this research topic, the aim of the present work is to propose and analyze some Liu-Storey conjugate gradient algorithms for solving (4). Basically, we extend the methods considered in [36, 35, 52] to the vector context. The first algorithm restricts the LS conjugate parameter to be nonnegative and use a line search satisfying the (vector) standard Wolfe conditions. Under the assumption that the directions yield descent, the global convergence of the method is established. We note that the descent property for this particular algorithm can be achieved by performing a sufficiently accurate line search. In the second algorithm, by considering a modification in the LS conjugate parameter, the search directions automatically yield descent regardless of the line search used. The global convergence of the algorithm is provided assuming that the step sizes satisfy the (vector) strong Wolfe conditions. Finally, the third algorithm consists of a combination of the LS parameter with a new Armijo-type line search (to be proposed here for the vector setting).

The search directions of the latter algorithm also satisfy the descent condition and its global convergence is proven. From the applicability viewpoint, we test our algorithms on a set of problems taken from the multiobjective optimization literature and comparisons with the PRP [40] and the steepest descent [19] methods are provided. It is worth mentioning that in the last two decades, the extension of methods of single-objective to vector optimization has been the subject of intense research. Some works on the subject can be found in [1, 19, 3, 9, 18, 21, 22, 26, 27, 28, 38, 48, 54, 5, 6, 7, 8, 10, 11, 2, 24].

This work is organized as follows. Section 2 presents some notation and preliminary results. In section 3, we describe the three proposed LS-CG methods and present their convergence results, whose proofs are postponed to the Appendix. Numerical results are presented in section 4 and some final remarks are given in section 5.

**Notation.** If $\mathbb{K} = \{k_1, k_2, \ldots\} \subseteq \mathbb{N}$ ($k_{j+1} > k_j$, $\forall j \in \mathbb{N}$), we denote $\mathbb{K} \underset{\infty}{\subset} \mathbb{N}$. The signum function is denoted by $\operatorname{sgn}(\cdot)$, i.e, given a real number $x$, then $\operatorname{sgn}(x) = -1, 0$ or $1$ if $x < 0$, $x = 0$ or $x > 0$, respectively. The cardinality of a set $\Omega$ is denoted by $|\Omega|$. If $S \subset \mathbb{R}^m$, then the conic hull and the convex hull of $S$ are denoted by $\operatorname{cone}(S)$ and $\operatorname{conv}(S)$, respectively.

## 2. Preliminaries

This sections summarizes some definitions, notation, and basic results used in this paper. More details about the main tools for vector optimization can be found, for example, in [39, 19, 28, 26, 40].

Let $K^*$ be the positive polar cone of $K$ defined as

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

Using the fact that $K$ is closed and convex, we have $K = K^{**}$,

$$-K = \left\{ z \in \mathbb{R}^m \mid \langle z, w \rangle \leq 0, \forall w \in K^* \right\}$$

and

$$-\operatorname{int}(K) = \left\{ z \in \mathbb{R}^m \mid \langle z, w \rangle < 0, \forall w \in K^* \setminus \{0\} \right\}.$$

Let $C \subset K^* \setminus \{0\}$ be a compact set such that

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

Assuming that $K$ is polyhedral, we can choose $C$ as the finite set of extremal rays of $K^*$. In particular, when $K = \mathbb{R}^m_+$ (multiobjective optimization case), since $K^* = K$, $C$ can be taken as the canonical basis of $\mathbb{R}^m$.

It is said that $x \in \mathbb{R}^n$ is a $K$-*critical Pareto* or *stationary* for $F$ iff

$$-\mathrm{int}(K) \cap \mathrm{Image}(JF(x)) = \emptyset,$$

where $JF(x)$ is the Jacobian of $F$ at $x$ and $\mathrm{Image}(A)$ is the image of a linear operator $A$. Note that if $x \in \mathbb{R}^n$ is not $K$-critical, then there exists a $K$-descent direction $v$ for $F$ at $x$. Here, we say that $v$ is a $K$-descent direction for $F$ at $x$ if there exists $\varepsilon > 0$ such that $F(x+tv) \prec_K F(x)$ for all $0 < t < \varepsilon$; see [39].

In view of the compactness of $C$, define $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ by

$$\varphi(z) := \sup \left\{ \langle z, w \rangle \mid w \in C \right\}.$$

Thus, we have the following characterizations of $-K$ and $-\mathrm{int}(K)$:

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

We next present some properties of the function $\varphi$.

**Lemma 1.** *([28, Lemma 3.1])*

*a) For every $z$ and $z' \in \mathbb{R}^m$, we have $\varphi(z+z') \leq \varphi(z) + \varphi(z')$. Moreover, if $z \preceq_K z'$ (resp. $z \prec_K z'$) then $\varphi(z) \leq \varphi(z')$ (resp. $\varphi(z) < \varphi(z')$).*

*b) $\varphi$ is 1-Lipschitz continuous and positive homogeneous.*

Now define the mapping $\mathcal{D}\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ as

$$\mathcal{D}(x,d) := \varphi(JF(x)d) = \sup \left\{ \langle JF(x)d, w \rangle \mid w \in C \right\}.$$

Note that, if $\mathcal{D}(x,d) < 0$ (resp. $\mathcal{D}(x,d) \geq 0$ for all $d$), then $d$ is a $K$-descent direction for $F$ at $x$ (resp. $x$ is $K$-critical point for $F$). It is said that a direction $d \in \mathbb{R}^n$ satisfies the *sufficient descent condition* at $x \in R^n$ iff

$$\mathcal{D}(x,d) \leq c\mathcal{D}(x, d_{SD}(x)), \tag{5}$$

for some $c > 0$. Below, we highlight some other properties of the function $\mathcal{D}$.

**Lemma 2.** *(a) Let $x, z, d \in \mathbb{R}^n$ and $\alpha \geq 0$, then $\mathcal{D}(x, z + \alpha d) \leq \mathcal{D}(x, z) + \alpha \mathcal{D}(x, d)$.*

*(b) The mapping $(x, d) \mapsto \mathcal{D}(x, d)$ is continuous.*

*(c) Let $x, z, d \in \mathbb{R}^n$, then $|\mathcal{D}(x, d) - \mathcal{D}(z, d)| \leq \|JF(x) - JF(z)\|\|d\|$. As a consequence, if $JF$ is L-Lipschitz continuous, then $\mathcal{D}(\cdot, d)$ is $L\|d\|$-Lipschitz continuous.*

Proof. Item (a) is a direct consequence of the definition of $\mathcal{D}$. For item (b), see [28]. Item (c) follows directly from Lemma 1 and the definition of $\mathcal{D}$. $\square$

The steepest direction $d_{SD}\colon \mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$d_{SD}(x) := \arg\min \left\{ \mathcal{D}(x, d) + \frac{\|d\|^2}{2} \mid d \in \mathbb{R}^n \right\} \tag{6}$$

and the optimal value of the subproblem in (6) is denoted by

$$\theta(x) := \mathcal{D}(x, d_{SD}(x)) + \frac{\|d_{SD}(x)\|^2}{2}. \tag{7}$$

We trivially obtain that $d_{SD}(x)$ exists and is unique, since $\mathcal{D}(x, \cdot)$ is a real closed convex function. Note that, when $F\colon \mathbb{R}^n \to \mathbb{R}$, $K = \mathbb{R}_+$, and $C = \{1\}$ (single-objective case), it follows that $\mathcal{D}(x, d) = \langle \nabla F(x), d \rangle$, $d_{SD}(x) = -\nabla F(x)$, and $\theta(x) = -\|\nabla F(x)\|^2/2$. Now, if $K = \mathbb{R}^m_+$ and $C$ is the canonical basis of $\mathbb{R}^m$ (multiobjective optimization case), $d_{SD}(x)$ can be computed by solving the following convex quadratic problem:

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

see [19]. The following lemma shows that $d_{SD}(x)$ and $\theta(x)$ can be used to characterize stationary points of (4).

**Lemma 3.** *([28, Lemma 3.3])*

*a) If $x$ is $K$-critical point for $F$, we have $d_{SD}(x) = 0$ and $\theta(x) = 0$.*

*b) If $x$ is not $K$-critical, then $d_{SD}(x) \neq 0$, $\theta(x) < 0$, $\mathcal{D}(x, d_{SD}(x)) < -\|d_{SD}(x)\|^2/2 < 0$. In particular, $d_{SD}(x)$ is a $K$-descent direction for $F$ at $x$.*

*c) The functions $d_{SD}(\cdot)$ and $\theta(\cdot)$ are continuous.*

Let $d \in \mathbb{R}^n$ be a $K$-descent direction for $F$ at $x$ and consider the task of finding a step size $\alpha > 0$ from $x$ along $d$. As an extension of the scalar case, we say that $\alpha > 0$ is obtained satisfying an *exact line search* whenever

$$\mathcal{D}(x + \alpha d, d) = 0.$$

Consider now $e \in K$ such that for every $w \in C$

$$0 < \langle w, e \rangle \leq 1. \tag{8}$$

We say that $\alpha > 0$ satisfies the *standard Wolfe conditions* if

$$\begin{aligned} F(x + \alpha d) &\preceq_K F(x) + \rho \alpha \mathcal{D}(x, d) e, \\ \mathcal{D}(x + \alpha d, d) &\geq \sigma \mathcal{D}(x, d), \end{aligned}$$

where $0 < \rho < \sigma < 1$ are given constants. In turn, it is said that $\alpha > 0$ is obtained satisfying the *strong Wolfe conditions* when

$$\begin{aligned} F(x + \alpha d) &\preceq_K F(x) + \rho \alpha \mathcal{D}(x, d) e, \\ \big|\mathcal{D}(x + \alpha d, d)\big| &\leq \sigma \big|\mathcal{D}(x, d)\big| . \end{aligned}$$

It is easy to see that $e \in K$ as in (8) always exists. In particular, we can take $e = [1, \ldots, 1]^T \in \mathbb{R}^m$ in the multiobjective optimization case.

## 3. Liu-Storey methods for vector optimization

In this section, we propose and analyze three Liu-Storey conjugate gradient (LS-CG) methods to solve (4). We first discuss extensions of the classic and modified LS-CG methods considered in [36, 35] and their global convergence results. We then dedicate to a study of a vector version of the LS-CG method with a Armijo-type line search proposed in [52]. From now on, we assume the following hypotheses.

**(H)** The cone $K$ is polyhedral and $C$ is the finite set of normalized extremal rays of $K^*$ (which, in particular, implies that $\|w\| = 1$ for all $w \in C$). In addition, the level set $\mathcal{L} := \{x \mid F(x) \preceq_K F(x^0)\}$ is bounded, where $x^0 \in \mathbb{R}^n$ is the given starting point, and $JF$ is $L$-Lipschitz continuous on an open set containing $\mathcal{L}$ (i.e., there exists an open set $\mathcal{N} \supset \mathcal{L}$ such that, for all $x, y \in \mathcal{N}$, it holds that $\|JF(x) - JF(y)\| \leq L\|x - y\|$).

We begin by proposing an extension of the classic LS-CG algorithm introduced in [36]. Here, the LS conjugate parameter is restricted to be nonnegative.

---

**Algorithm 1.** ***Nonnegative Liu-Storey CG method.***

**Step 0.** Let $0 < \rho < \sigma < 1$, $e \in K$ as in (8), and $x^0 \in \mathbb{R}^n$ be given. Initialize $k \leftarrow 0$.

**Step 1.** Compute $d_{SD}(x^k)$ and $\theta(x^k)$ as in (6) and (7), respectively. If $\theta(x^k) = 0$, then STOP.

**Step 2.** Define

$$d^k = \begin{cases} d_{SD}(x^k), & \text{if } k = 0 \\ d_{SD}(x^k) + \beta_k d^{k-1}, & \text{if } k \geq 1, \end{cases} \tag{9}$$

where

$$\beta_k := \max\left\{ \beta_k^{LS} := \frac{-\mathcal{D}(x^k, d_{SD}(x^k)) + \mathcal{D}(x^{k-1}, d_{SD}(x^k))}{-\mathcal{D}(x^{k-1}, d^{k-1})}, 0 \right\} \tag{10}$$

**Step 3.** Compute $\alpha_k > 0$ such that

$$F(x^k + \alpha_k d^k) \preceq_K F(x^k) + \rho \alpha_k \mathcal{D}(x^k, d^k) e,$$
$$\mathcal{D}(x^k + \alpha_k d^k, d^k) \geq \sigma \mathcal{D}(x^k, d^k).$$

**Step 4.** Set $x^{k+1} := x^k + \alpha_k d^k$, $k \leftarrow k + 1$, and go to Step 1.

---

**Remark 1.** *(i)* If Algorithm 1 stops at iteration $k$, then Lemma 3 implies that $x^k$ is $K$-critical. *(ii)* $\beta_k^{LS}$ in Step 2 of Algorithm 1 is a direct extension of the classic LS conjugate parameter. Indeed, if $m = 1$, $K = \mathbb{R}_+$, and

$C = \{1\}$, we trivially have that $\beta_k^{LS}$ in (10) corresponds to (3). *(iii)* In Step 3, it is required that the step size $\alpha_k$ satisfies the standard Wolfe conditions. If $d^k$ is a $K$-descent direction of $F$ at $x^k$, under assumption **(H)**, it is possible to guarantee the existence of positive step sizes satisfying such conditions, see [40, Proposition 3.2]. *(iv)* In the convergence analysis of Algorithm 1, we will require that $d^k$ satisfies the sufficient descent condition (5). As in the single-objective optimization, this property is not automatically fulfilled by Algorithm 1, but can be obtained if a sufficiently accurate line search is used. Let us briefly discuss this issue. Since $\beta_k$ in (10) is non-negative, by the definition of $d^k$ and Lemma 2(a), it follows that

$$\mathcal{D}(x^k, d^k) \leq \mathcal{D}(x^k, d_{SD}(x^k)) + \beta_k \mathcal{D}(x^k, d^{k-1}).$$

Remember that *exact line search* means that $\mathcal{D}(x^k, d^{k-1}) = 0$. Therefore, performing a sufficiently accurate line search it is possible to guarantee that the last term in the above inequality is not too large in order to obtain the desired descent property, see [40, 41]. Thus, using inductive arguments, it is possible to show that Algorithm 1 with appropriate line searches is well defined.

We now present the convergence result of Algorithm 1, whose proof is given in Appendix A.1.

**Theorem 4.** *Consider Algorithm 1 with a sufficiently accurate line search such that $d^k$ satisfies the sufficient descent condition* (5) *at $x^k$ for every $k$. Let $\{(x^k, d^k)\}$ be an infinite sequence generated by Algorithm 1. Then,*

$$\liminf_{k\to\infty} \|d_{SD}(x^k)\| = 0.$$

As already mentioned, the Nonnegative LS-CG method does not automatically generates descent directions. In the scalar case, in order to obtain such desired property, a modified LS-CG method was introduced and analyzed in [35]. In the next algorithm, we propose an extension/adaptation of this method for vector optimization, in particular, generating $K$-descent directions regardless of the line search used. This property is especially interesting in the vector case because it can lead to significant computational savings. Indeed, when a trial step size $\alpha$ is calculated in Step 3 of Algorithm 1 but the associated direction at $x^k + \alpha d^k$ does not satisfy the sufficient descent condition, the computation of $d_{SD}(x^k + \alpha d^k)$ is wasted and the algorithm continues performing a more accurate line search. The modified LS-CG is formally defined as follows.

---

**Algorithm 2.** ***Modified Liu-Storey CG method***

**Step 0.** Let $0 < \rho < \sigma < 1$, $t > 1/2$, $\eta > 0$, $e \in K$ as in (8), and $x^0 \in \mathbb{R}^n$ be given. Initialize $k \leftarrow 0$.

**Step 1.** Compute $d_{SD}(x^k)$ and $\theta(x^k)$ as in (6) and (7), respectively. If $\theta(x^k) = 0$, then STOP.

**Step 2.** Define $d^k$ as in (9) with

$$\beta_k := \max\left\{\beta_k^{MLS} := \beta_k^{LS} - t\frac{\|JF(x^k) - JF(x^{k-1})\|^2 \mathcal{D}(x^k, d^{k-1})}{\mathcal{D}^2(x^{k-1}, d^{k-1})}, \eta^k\right\}, \tag{11}$$

where $\beta_k^{LS}$ is as in (10) and

$$\eta^k := \frac{-1}{\|d^{k-1}\| \min\{\eta, \|d_{SD}(x^{k-1})\|\}}.$$

If $\mathcal{D}(x^k, d^k) > [1 - 1/(2t)]\mathcal{D}(x^k, d_{SD}(x^k))$, then $d^k := d_{SD}(x^k)$ and $\beta_k := 0$.

**Step 3.** Compute $\alpha_k > 0$ such that

$$F(x^k + \alpha_k d^k) \preceq_K F(x^k) + \rho\alpha_k \mathcal{D}(x^k, d^k)e,$$
$$\left|\mathcal{D}(x^k + \alpha_k d^k, d^k)\right| \leq \sigma \left|\mathcal{D}(x^k, d^k)\right|.$$

**Step 4.** Set $x^{k+1} := x^k + \alpha_k d^k$, $k \leftarrow k + 1$, and go to Step 1.

---

We claim that the search direction $d^k$ generated by Algorithm 2 satisfies, for all $k \geq 0$, the sufficient descent condition

$$\mathcal{D}(x^k, d^k) \leq \left(1 - \frac{1}{2t}\right)\mathcal{D}(x^k, d_{SD}(x^k)). \tag{12}$$

Suppose that this statement holds for $d^{k-1}$ (which is trivially true for $d^0$). If $\beta_k$ in (11) is nonnegative, then $\beta_k = \beta_k^{MLS}$ and, by Lemma 2(a), we obtain

$$\begin{aligned} &\mathcal{D}(x^k, d^k) \leq \mathcal{D}(x^k, d_{SD}(x^k)) + \beta_k^{MLS}\mathcal{D}(x^k, d^{k-1}) \\ &\leq \mathcal{D}(x^k, d_{SD}(x^k)) + \frac{\mathcal{D}(x^{k-1}, d_{SD}(x^k)) - \mathcal{D}(x^k, d_{SD}(x^k))}{-\mathcal{D}(x^{k-1}, d^{k-1})}\mathcal{D}(x^k, d^{k-1}) - T_k, \end{aligned} \tag{13}$$

where

$$T_k := t\frac{\Lambda_k^2\mathcal{D}^2(x^k, d^{k-1})}{\mathcal{D}^2(x^{k-1}, d^{k-1})} \quad \text{and} \quad \Lambda_k := \|JF(x^k) - JF(x^{k-1})\|.$$

It follows from (13) and Lemma 2(c) that

$$\mathcal{D}(x^k, d^k) \leq \mathcal{D}(x^k, d_{SD}(x^k)) + \frac{|\mathcal{D}(x^k, d^{k-1})|\|d_{SD}(x^k)\|\Lambda_k}{-\mathcal{D}(x^{k-1}, d^{k-1})} - T_k.$$

Thus,

$$\mathcal{D}(x^k, d^k) \leq \mathcal{D}(x^k, d_{SD}(x^k)) + \frac{|\mathcal{D}(x^{k-1}, d^{k-1})||\mathcal{D}(x^k, d^{k-1})|\|d_{SD}(x^k)\|\Lambda_k}{\mathcal{D}^2(x^{k-1}, d^{k-1})} - T_k.$$

Taking $a := |\mathcal{D}(x^{k-1}, d^{k-1})|\|d_{SD}(x^k)\|/\sqrt{2t}$ and $b := \sqrt{2t}\Lambda_k\mathcal{D}(x^k, d^{k-1})$, and using that $ab \leq a^2/2 + b^2/2$, we have

$$\begin{aligned} &\mathcal{D}(x^k, d^k) \leq \mathcal{D}(x^k, d_{SD}(x^k)) \\ &+ \frac{1}{\mathcal{D}^2(x^{k-1}, d^{k-1})}\left[\frac{\|d_{SD}(x^k)\|^2\mathcal{D}^2(x^{k-1}, d^{k-1})}{4t} + t\Lambda_k^2\mathcal{D}^2(x^k, d^{k-1})\right] - T_k. \end{aligned}$$

Therefore, by the definition of $T_k$ and Lemma 3(b), it follows that

$$\mathcal{D}(x^k, d^k) \leq \mathcal{D}(x^k, d_{SD}(x^k)) + \frac{1}{4t}\|d_{SD}(x^k)\|^2 \leq \left(1 - \frac{1}{2t}\right)\mathcal{D}(x^k, d_{SD}(x^k)),$$

as desired. Now, if $\beta_k$ in (11) is negative and $d^k$ as in (9) does not satisfy (12), we redefine $d^k$ to be the steepest descent direction, which trivially satisfies the required descent property, proving our claim. The strategy of redefining $d^k$ is supported by the following fact: if $\beta_k$ in (11) is negative, even employing an exact line search, the descent condition for $d^k$ as in (9) may not hold. This case is illustrated in the following example.

**Example 1.** *Consider the multiobjective problem* (4) *with:* $K = \mathbb{R}^2_+$, $C$ *equal to the canonical basis of* $\mathbb{R}^2$, *and* $F : \mathbb{R}^2 \longrightarrow \mathbb{R}$ *given by*

$$F_1(x) := \frac{x_1^2 + x_2^2}{2} \quad \text{and} \quad F_2(x) := \frac{(x_1 - 1)^2 - (x_2 - 1)^2}{2}.$$

*Define* $t = 0.51$, $\eta = 0.1$, $\sigma = 0.1$, $\rho = 10^{-4}$, *and* $e = (1, 1)^T$. *Choose* $x^0 = (1, 1)^T$. *Solving* (6) *and* (7), *we obtain* $d_{SD}(x^0) = (-0.6, 0.2)^T$ *and* $\theta(x^0) = 0.2$. *Hence,* $d^0 = (-0.6, 0.2)^T$ *and* $\alpha_0 = 1$ *satisfies the exact line search condition, i.e.,* $\mathcal{D}(x^0 + \alpha_0 d^0, d^0) = 0$. *So,* $x^1 = x^0 + d^0 = (0.4, 1.2)^T$, *for which* $d_{SD}(x^1) = (-34, 10)^T/785$ *and* $\theta(x^1) = 628/785$. *Using* $\mathcal{D}(x^1, d^0) = 0$, *we obtain*

$$\beta_1 = \max\left\{\beta_1^{MLS} = -\frac{18.4}{314}, \eta_1 = -\frac{500}{\sqrt{10}}\right\} = -\frac{18.4}{314},$$

*and hence* $d^1 := d_{SD}(x^1) + \beta_1 d^0 = (-6.4, 0.8)^T/785$. *Since* $\mathcal{D}(x^1, d^1) = 2.08/785 > 0$, *we conclude that* $d^1$ *fails to be a* $K$*-descent direction for* $F$ *at* $x^1$.

Under Assumption **(H)**, since (12) holds for all $k$, we obtain from [40, Proposition 3.2] that the line search in Step 3 is well-defined and, as a consequence, Algorithm 2 is also well-defined. Although the strong Wolfe conditions required in Step 3 are not used to show the well-definedness of Algorithm 2, they will be important to ensure convergence. Next theorem presents the convergence result related to Algorithm 2, whose proof is given in Appendix A.2.

**Theorem 5.** *Suppose that Algorithm 2 generates an infinite sequence* $\{(x^k, d^k)\}$. *Then,*

$$\liminf_{k \to \infty} \|d_{SD}(x^k)\| = 0.$$

The remaining part of this section is dedicated to propose a vector version of the LS-CG method with an Armijo-type line search proposed in [52]. The potential advantage of this approach is the ease of implementation of the line search, where a simple backtracking procedure can be considered. In [52], in order to get the sufficient descent condition in the scalar minimization, a modification in the classical Armijo condition was proposed in which, for each iteration, the initial trial step size depend on an estimate $L_k$ of the

Lipschitz constant $L$ of the gradient of the objective function. We refer the reader to [50, 51] for some estimation techniques for the Lipschitz constant in the scalar case. Now, taking into account the following estimate of the Lipschitz constant $L$ of the Jacobian of $F$:

$$L \cong \frac{|\mathcal{D}(x^k, d_{SD}(x^k)) - \mathcal{D}(x^{k-1}, d_{SD}(x^k))|}{\|s^{k-1}\|},$$

where $s^{k-1} := x^k - x^{k-1}$, we consider, for each $k$, the following estimate $L_k$ of $L$:

$$L_k = \max\left(L_{k-1}, \min\left(\frac{|\mathcal{D}(x^k, d_{SD}(x^k)) - \mathcal{D}(x^{k-1}, d_{SD}(x^k))|}{\|s^{k-1}\|}, \bar{M}\right)\right), \quad (14)$$

where $L_0$ and $\bar{M}$ are given positive constants such that $L_0 < \bar{M}$. We observe that (14) extends to the vector case a Lipschitz estimate considered in [52].

---

**Algorithm 3.** ***Liu-Storey CG method with Armijo-type line search***

**Step 0.** Let $\rho \in (0,1)$, $\mu \in (0,1)$, $c \in (0,1)$, $L_0 > 0$, $\bar{M} > 0$, $e \in K$ as in (8), and $x^0 \in \mathbb{R}^n$ be given. Compute $d_{SD}(x^0)$ as in (6) and define $d^0 := d_{SD}(x^0)$. Initialize $k \leftarrow 0$.

**Step 1.** Compute $\theta(x^k)$ as in (7). If $\theta(x^k) = 0$, then STOP.

**Step 2.** Define $\tau_k := -\frac{(1-c)}{L_k}\frac{\mathcal{D}(x^k,d^k)}{\|d^k\|^2}$, where $L_k$ is as in (14) . Let $\alpha_k$ be the largest value in $\{\tau_k, \tau_k\mu, \tau_k\mu^2, ...\}$ such that

$$F(x^+) \preceq_K F(x^k) + \alpha_k\rho\mathcal{D}(x^k, d^k)e, \quad (15)$$

$$\mathcal{D}(x^+, d(x^+)) \leq c\mathcal{D}(x^+, d_{SD}(x^+)), \quad (16)$$

where $x^+ := x^k + \alpha_k d^k$ and

$$d(x^+) := d_{SD}(x^+) + \left(\frac{-\mathcal{D}(x^+, d_{SD}(x^+)) + \mathcal{D}(x^k, d_{SD}(x^+))}{-\mathcal{D}(x^k, d^k)}\right) d^k. \quad (17)$$

**Step 3.** Define $x^{k+1} := x^+$, $d_{SD}(x^{k+1}) := d_{SD}(x^+)$, and $d^{k+1} := d(x^+)$. Set $k \leftarrow k+1$, and go to Step 1.

---

**Remark 2.** Note that $\tau_k$ is inversely proportional to $L_k$. Thus, *large* values of $L_k$ can produce *small* step sizes causing slow convergence. This is undesirable behavior of this type of approach even when $L_k = L$ for large values of $L$. Therefore, in order not to deteriorate the performance of the method, it is important to obtain estimations $L_k$ of $L$ such that they are not so greater than L.

We now present the global convergence result related to Algorithm 3, whose proof is given in Appendix A.3.

**Theorem 6.** *Suppose that Algorithm 3 generates an infinite sequence $\{(x^k, d^k)\}$. Then, every limit point of $\{x^k\}$ is a $K$-critical Pareto point of $F$.*

Note that the convergence result for Algorithm 3 given above is stronger than those for Algorithms 1 and 2 presented in their respective convergence theorems. Indeed, Theorems 4 and 5 only imply the existence of a $K$-critical Pareto limit point of the sequences generated by Algorithms 1 and 2.

## 4. Numerical experiments

This section reports the numerical experiments carried out to verify the practical behavior of the proposed algorithms. This will be done by comparing the performance of the LS-CG algorithms with the steepest descent (SD) and the PRP-CG methods introduced in [19] and [40], respectively. Algorithms 1, 2, and 3 are identified as Nonnegative LS, Modified LS, and Armijo LS methods, respectively. We also consider a version of Algorithm 3 where the conjugate parameter in (17) (the term in parentheses) is taken as

$$\max\left\{\frac{-\mathcal{D}(x^+, d_{SD}(x^+)) + \mathcal{D}(x^k, d_{SD}(x^+))}{-\mathcal{D}(x^k, d^k)}, 0\right\}.$$

This method is identified as Armijo LS+. We mention that, by following the proofs in Appendix A.3, it can be proven that Theorem 6 also holds for this LS variant.

We implemented the algorithms in Fortran 90. For computing the steepest decent direction $d_{SD}(x)$ and the optimal value $\theta(x)$ in (6) and (7), respectively, we used the software Algencan [4]. Regarding the line searches, for computing a step size satisfying the Wolfe conditions, we used the algorithm proposed in [41]. In turn, the Armijo-type scheme considered in Algorithm 3 was implemented as a simple backtracking procedure. Basically, we first get a point satisfying (15) and then we check if (16) holds. We used the following algorithmic parameters: $\rho = 10^{-4}$, $\sigma = 10^{-1}$, $t = 0.75$, $\eta = 10^{-2}$, $c = 10^{-2}$, $\mu = 0.75$, $L_0 = 10^{-2}$, and $\bar{M} = 10^4$. The norm of a matrix $A \in M^{m\times n}$ (see (11)) was computed as

$$\|A\|_{\infty,2} := \max_{x\neq 0} \frac{\|A\|_\infty}{\|x\|} = \max_{i=1,\ldots,m} \|A_{i,\cdot}\| = \max_{i=1,\ldots,m} \left(\sum_{j=1}^{n} A_{i,j}^2\right)^{1/2}.$$

Since Lemma 3 implies that $x \in R^n$ is a $K$-critical Pareto point of $F$ iff $\theta(x) = 0$, we stopped the algorithms at $x^k$ reporting convergence when

$$\left|\theta(x^k)\right| \leq 5 \times \texttt{eps}^{1/2},$$

where $\texttt{eps} = 2^{-52} \approx 2.22 \times 10^{-16}$.

The test problems were chosen from the multiobjective optimization literature. Table 1 describes the informations for each test problem, where first and second columns contain the name of the problem and the reference where it can be found. Third and fourth columns contain the number of variables and the number of objectives of the corresponding problem, respectively. The starting points were generated belonging to a box $\{x \in \mathbb{R}^n \mid \ell \leq x \leq u\}$ whose bounds $\ell$ and $u$ are informed in the last two columns. We emphasize that the boxes reported in Table 1 were used only to define the starting points, but were not considered by the algorithms themselves.

Due to numerical reasons, we considered a scaled version of problem (4) given by

$$\text{Minimize}_K \ \left(\gamma_1 F_1(x), \ldots, \gamma_m F_m(x)\right), \quad x \in \mathbb{R}^n, \tag{18}$$

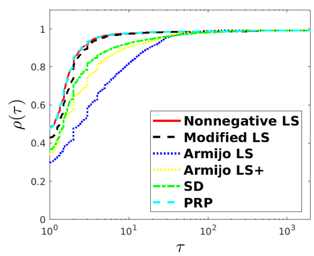
where the scaling factors are computed as $\gamma_j := 1/\max\{1, \|\nabla F_j(x^0)\|_\infty\}$, $j = 1, \ldots, m$, where $x^0 \in \mathbb{R}^n$ is the given starting point. It is important to mention that problems (4) and (18) are equivalent, in the sense that they have the same $K$-critical Pareto points. The numerical comparisons will be shown using performance profiles graphics [15], which we briefly explain here for the sake of completeness. Let $\mathcal{S}$ be the set of solvers, $\mathcal{P}$ be the set of problems, and $t_{p,s} > 0$ be the performance (for example, number of iterations) of the solver $s \in \mathcal{S}$ on the problem $p \in \mathcal{P}$. Here, lower values of $t_{p,s}$ mean better performances. Define the performance ratio $r_{p,s} := t_{p,s}/\min\{t_{p,s} \mid s \in \mathcal{S}\}$

| Problem | Reference | $n$ | $m$ | $\ell^T$ | $u^T$ |
|---|---|---|---|---|---|
| AP1 | [1] | 2 | 3 | $(-10, -10)$ | $(10, 10)$ |
| AP2 | [1] | 1 | 2 | $-100$ | $100$ |
| AP3 | [1] | 2 | 2 | $(-100, -100)$ | $(100, 100)$ |
| AP4 | [1] | 3 | 3 | $(-10, -10, -10)$ | $(10, 10, 10)$ |
| DD1 | [14] | 5 | 2 | $(-20, \dots, -20)$ | $(20, \dots, 20)$ |
| DGO1 | [32] | 1 | 2 | $-10$ | $13$ |
| Far1 | [32] | 2 | 2 | $(-1, -1)$ | $(1, 1)$ |
| FDS | [18] | 5 | 3 | $(-2, \dots, -2)$ | $(2, \dots, 2)$ |
| | | 50 | 3 | $(-2, \dots, -2)$ | $(2, \dots, 2)$ |
| | | 100 | 3 | $(-2, \dots, -2)$ | $(2, \dots, 2)$ |
| | | 200 | 3 | $(-2, \dots, -2)$ | $(2, \dots, 2)$ |
| FF1 | [32] | 2 | 2 | $(-1, -1)$ | $(1, 1)$ |
| Hil1 | [31] | 2 | 2 | $(0, 0)$ | $(1, 1)$ |
| JOS1 | [33] | 2 | 2 | $(-100, \dots, -100)$ | $(100, \dots, 100)$ |
| | | 50 | 2 | $(-100, \dots, -100)$ | $(100, \dots, 100)$ |
| | | 100 | 2 | $(-100, \dots, -100)$ | $(100, \dots, 100)$ |
| | | 200 | 2 | $(-100, \dots, -100)$ | $(100, \dots, 100)$ |
| KW2 | [34] | 2 | 2 | $(-3, -3)$ | $(3, 3)$ |
| Lov1 | [37] | 2 | 2 | $(-10, -10)$ | $(10, 10)$ |
| Lov3 | [37] | 2 | 2 | $(-20, -20)$ | $(20, 20)$ |
| Lov4 | [37] | 2 | 2 | $(-20, -20)$ | $(20, 20)$ |
| Lov5 | [37] | 3 | 2 | $(-2, -2, -2)$ | $(2, 2, 2)$ |
| MGH16 | [43] | 4 | 5 | $(-25, -5, -5, -1)$ | $(25, 5, 5, 1)$ |
| | | 4 | 20 | $(-25, -5, -5, -1)$ | $(25, 5, 5, 1)$ |
| | | 4 | 50 | $(-25, -5, -5, -1)$ | $(25, 5, 5, 1)$ |
| MGH26 | [43] | 4 | 4 | $(-1, -1, -1 - 1)$ | $(1, 1, 1, 1)$ |
| MGH33 | [43] | 10 | 10 | $(-1, \dots, -1)$ | $(1, \dots, 1)$ |
| MLF2 | [32] | 2 | 2 | $(-100, -100)$ | $(100, 100)$ |
| MMR1 | [42] | 2 | 2 | $(0.1, 0)$ | $(1, 1)$ |
| MOP2 | [32] | 2 | 2 | $(-1, -1)$ | $(1, 1)$ |
| MOP3 | [32] | 2 | 2 | $(-\pi, -\pi)$ | $(\pi, \pi)$ |
| MOP5 | [32] | 2 | 3 | $(-1, -1)$ | $(1, 1)$ |
| MOP7 | [32] | 2 | 3 | $(-400, -400)$ | $(400, 400)$ |
| PNR | [47] | 2 | 2 | $(-2, -2)$ | $(2, 2)$ |
| QV1 | [32] | 10 | 2 | $(-5, \dots, -5)$ | $(5, \dots, 5)$ |
| SK1 | [32] | 1 | 2 | $-100$ | $100$ |
| SK2 | [32] | 4 | 2 | $(-10, -10, -10, -10)$ | $(10, 10, 10, 10)$ |
| SLCDT1 | [49] | 2 | 2 | $(-1.5, -1.5)$ | $(1.5, 1.5)$ |
| SLCDT2 | [49] | 10 | 3 | $(-1, \dots, -1)$ | $(1, \dots, 1)$ |
| SP1 | [32] | 2 | 2 | $(-100, -100)$ | $(100, 100)$ |
| SSFYY2 | [32] | 1 | 2 | $-100$ | $100$ |
| Toi4 | [53] | 4 | 2 | $(-2, -2, -2, -2)$ | $(5, 5, 5, 5)$ |
| Toi8 | [53] | 3 | 3 | $(-1, -1, -1, -1)$ | $(1, 1, 1, 1)$ |
| Toi9 | [53] | 4 | 4 | $(-1, -1, -1, -1)$ | $(1, 1, 1, 1)$ |
| | | 50 | 50 | $(-1, -1, -1, -1)$ | $(1, 1, 1, 1)$ |
| | | 100 | 100 | $(-1, -1, -1, -1)$ | $(1, 1, 1, 1)$ |
| Toi10 | [53] | 4 | 3 | $(-2, -2, -2, -2)$ | $(2, 2, 2, 2)$ |
| | | 10 | 9 | $(-2, -2, -2, -2)$ | $(2, 2, 2, 2)$ |
| | | 30 | 29 | $(-2, -2, -2, -2)$ | $(2, 2, 2, 2)$ |
| VU1 | [32] | 2 | 2 | $(-3, -3)$ | $(3, 3)$ |

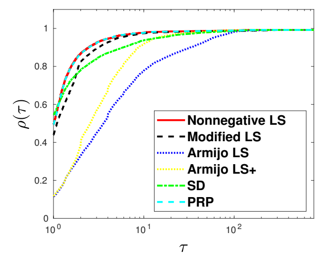
Table 1: Test problems.

and the cumulative distribution function $\rho_s : [1, \infty) \to [0, 1]$ by $\rho_s(\tau) := (1/|\mathcal{P}|) \left|\{p \in \mathcal{P} \mid r_{p,s} \leq \tau\}\right|$. The performance profile is obtained by plotting, for all $s \in \mathcal{S}$, the graph of $\rho_s$. Note that $\rho_s(1)$ corresponds to the fraction of problems for which solver $s$ was the most efficient over all the considered algorithms. In turn, the *robustness* related to a solver $s$ can be accessed on the extreme right of the graph of $\rho_s$.

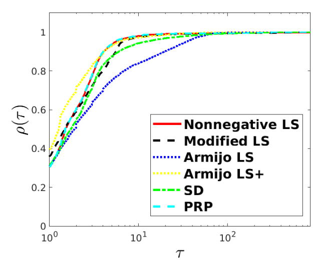
In multiobjective optimization, we are especially interested in estimating the Pareto frontier of a given problem. For this, a strategy is often applied: we run the algorithm at hand from several different starting points and collect the efficient points found. In view of this application, for each problem in Table 1, we considered 300 random initial points belonging to the respective boxes. Each instance was considered an independent problem and was solved by all the considered solvers. Figure 1 shows the results comparing the algorithms using as the performance measurement: (a) number of iterations; (b) number of functions evaluations; (c) number of gradient evaluations; (d) number of steepest descent direction evaluations.



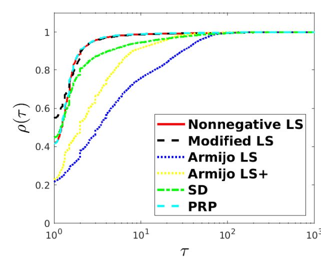

(a) Iterations




(b) Function evaluations




(c) Gradient evaluations




(d) SD evaluations

Figure 1: Performance profiles considering 300 initial points for each test problem using as the performance measurement: (a) number of iterations; (b) number of functions evaluations; (c) number of gradient evaluations; (d) number of steepest descent direction evaluations.

As can be seen in Figure 1, the Nonnegative LS and the PRP methods performed almost identically. Regarding the number of iterations, they were the most efficient algorithms followed by the Modified LS, the SD, and the Armijo LS algorithms. With respect to the number of function evaluations, although the SD algorithm was slightly the most efficient, it was quickly outperformed by the CG methods (except the Armijo LS algorithms). In term of gradient evaluations, the Armijo LS+ was the most efficient algorithm. Such behavior is justified by the fact that it generally requires a moderate amount of iterations and the implementation of its backtracking procedure does not use new derivative information until a trial step size satisfying (15) is obtained. Considering the number of steepest descent direction evaluations, the most efficient method was the Modified LS algorithm. This can be explained because the Modified LS algorithm automatically generates descent directions, performing only one steepest descent direction evaluation per iteration. This suggests that the Modified LS algorithm is an attractive alternative for large scale problems for which the computational cost is mainly associated with solving the subproblem (6). Finally, according to the theoretical results, the methods found stationary points in all runs.

Next, by using the so-called *Purity* and ($\Gamma$ and $\Delta$) *Spread* metrics, we compare the algorithms with respect to their ability to properly generate Pareto frontiers. We refer the reader to [12] for a meticulous explanation of these metrics. Let $PF_{p,s}$ be an approximation to the Pareto front obtained by solver $s$ for problem $p$. Here, $PF_{p,s}$ was obtained by running a particular solver $s$ from the same 300 starting points considered in the first part of the experiments and then removing the dominated points (given $x, y \in \mathbb{R}^n$, it is said that $x$ dominates $y$ if $F(y) - F(x) \in \mathbb{R}^m_+ \backslash \{0\}$). Now, let $PF_p$ be an approximation to the Pareto front obtained by first considering $\cup_{s \in \mathcal{S}} PF_{p,s}$ and then removing the dominated points.

- Purity metric: The Purity metric measures the number of nondominated points belonging to $PF_p$ that a solver is able to compute. Given a problem $p$ and a solver $s$, it is defined by the ratio

$$\bar{t}_{p,s} := \left| PF_{p,s} \cap PF_p \right| / \left| PF_p \right|.$$

  To analyze the Purity metric using the performance profile, we set $t_{p,s} := 1/\bar{t}_{p,s}$ and, thus, lower values of $t_{p,s}$ indicate better performances. If $\bar{t}_{p,s} = 0$, then we define $t_{p,s} := \infty$. As recommended in [12], for the Purity metric, we compared the algorithms in pairs.

- Spread metrics: A Spread metric seeks to measure the capacity of a given solver to find *well-distributed* points along the Pareto frontier. Given a problem $p$ and a solver $s$, consider that $PF_{p,s} \cap PF_p$ is formed by $x_1, \ldots, x_N$ and assume that these points are conveniently sorted by each objective function $j$ such that $F_j(x_i) \leq F_j(x_{i+1})$, $i = 1, ..., N$. Let

$x_0$ and $x_{N+1}$ be the the points corresponding to the lowest and highest values, respectively, of $F_j$ obtained from $PF_p$. The metrics $\Gamma$ and $\Delta$ are defined by

$$\Gamma_{p,s} := \max_{j \in \{1,\ldots,m\}} \max_{i \in \{0,\ldots,N\}} \delta_{i,j}$$

and

$$\Delta_{p,s} := \max_{j \in \{1,\ldots,m\}} \left( \frac{\delta_{0,j} + \delta_{N,j} + \sum_{i=1}^{N} |\delta_{i,j} - \bar{\delta}_j|}{\delta_{0,j} + \delta_{N,j} + (N-1)\bar{\delta}_j} \right),$$

where $\delta_{i,j} := \left|F_j(x_{i+1}) - F_j(x_i)\right|$ and $\bar{\delta}_j$, $j = 1, \ldots, m$, is the average of the distances $\delta_{i,j}$, $i = 1, ..., N$. In the performance profile, we set $t_{p,s} := \Gamma_{p,s}$ or $t_{p,s} := \Delta_{p,s}$, depending on the chosen metric.

The performance profiles related to the Purity metric are shown in Figure 2 comparing the algorithms in pairs. In the Armijo LS class, we only consider the Armijo LS+ since it outperformed the Armijo LS in the plots of Figure 1. As we can see, the LS methods slightly outperformed the SD method. On the other hand, no significant difference was noted among the CGs methods.

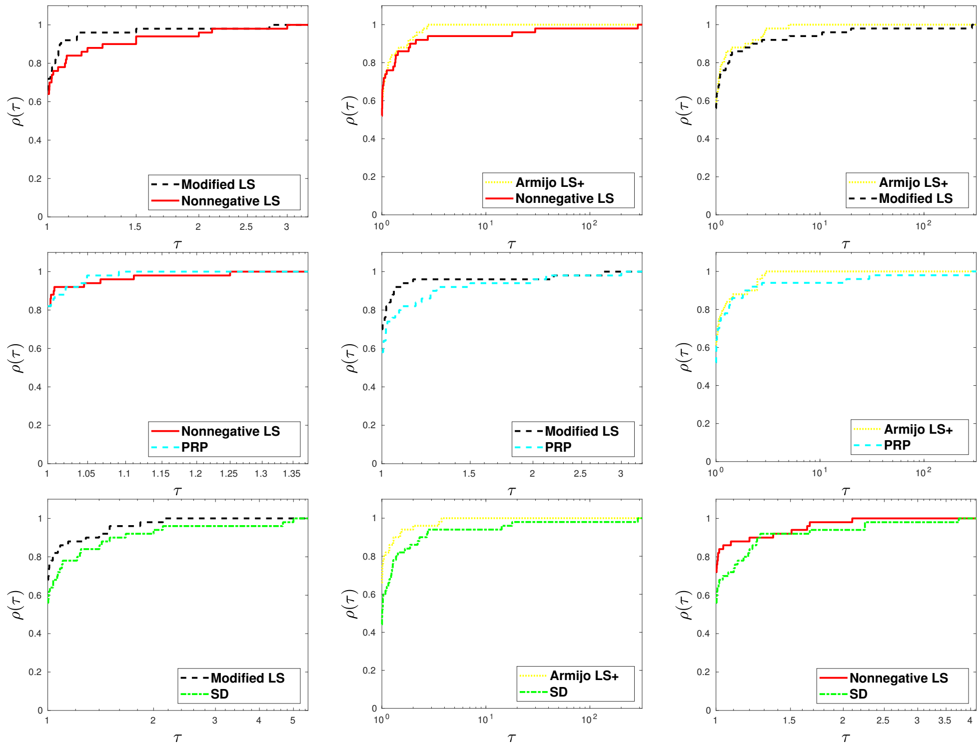
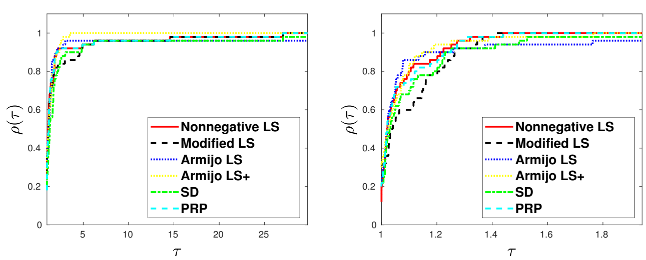



Figure 2: Purity metric performance profiles.

Figure 3 contains the performance profiles of the $\Gamma$ and $\Delta$ Spread metrics. Again, no significant difference among the considered algorithms is noticed for these metrics.




(a) $\Gamma$ metric (b) $\Delta$ metric

Figure 3: Spread metric performance profiles: (a) $\Gamma$ metric; (b) $\Delta$ metric.

## 5. Final remarks

This paper proposed and analyzed three variants of Liu-Storey (LS) nonlinear conjugate gradient (CG) methods for vector optimization. These variants are nontrivial extensions of LS-CG methods of the scalar case to the vector setting. The global convergences of the new methods were discussed in this general context. Numerical experiments on a set of test problems showed that the LS methods are competitive with the well-known PRP method in terms of numerical efficiency, as well as to obtain a good representation of Pareto frontiers.

## References


[1] M. A. Ansary and G. Panda. A modified Quasi-Newton method for vector optimization problem. *Optimization*, 64(11):2289–2306, 2015.

[2] P. B. Assunção, O. P. Ferreira, and L. F. Prudente. Conditional gradient method for multiobjective optimization. *Comput. Optim. Appl.*, 78(3):741–768, 2021.

[3] J. Y. Bello Cruz. A subgradient method for vector optimization problems. *SIAM J. Optim.*, 23(4):2169–2182, 2013.

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

[5] H. Bonnel, A. N. Iusem, and B. F. Svaiter. Proximal methods in vector optimization. *SIAM J. Optim.*, 15(4):953–970, 2005.

[6] L. C. Ceng, B. S. Mordukhovich, and J. C. Yao. Hybrid approximate proximal method with auxiliary variational inequality for vector optimization. *J. Optimiz. Theory App.*, 146(2):267–303, 2010.

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

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

[9] T. D. Chuong. Newton-like methods for efficient solutions in vector optimization. *Comput. Optim. Appl.*, 54(3):495–516, 2013.

[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(2):257–285, 8 2011.

[11] T. D. Chuong and J. C. Yao. Steepest descent methods for critical points in vector optimization problems. *Appl. Anal.*, 91(10):1811–1829, 2012.

[12] A. L. Custódio, J. F. A. Madeira, A. I. F. Vaz, and L. N. Vicente. Direct multisearch for multiobjective optimization. *SIAM J. Optim.*, 21(3):1109–1140, 2011.

[13] Y. H. Dai and Y. Yuan. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM J. Optim.*, 10(1):177–182, 1999.

[14] I. Das and J. Dennis. Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. *SIAM J. Optim.*, 8(3):631–657, 1998.

[15] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.*, 91(2):201–213, 2002.

[16] R. Fletcher. Practical Method of Optimization, Unconstrained Optimization, vol. 1, 1980.

[17] R. Fletcher and C. M. Reeves. Function minimization by conjugate gradients. *Comput. J.*, 7(2):149–154, 1964.

[18] J. Fliege, L. M. Graña Drummond, and B. F. Svaiter. Newton's method for multiobjective optimization. *SIAM J. Optim.*, 20(2):602–626, 2009.

[19] J. Fliege and B. F. Svaiter. Steepest descent methods for multicriteria optimization. *Math. Method. Oper. Res.*, 51(3):479–494, 2000.

[20] E. H. Fukuda, L. M. G. Drummond, and A. M. Masuda. A conjugate directions-type procedure for quadratic multiobjective optimization. *Optimization, https://doi.org/10.1080/02331934.2021.1914034.*

[21] E. H. Fukuda and L. M. Graña Drummond. On the convergence of the projected gradient method for vector optimization. *Optimization*, 60(8-9):1009–1021, 2011.

[22] E. H. Fukuda and L. M. Graña Drummond. Inexact projected gradient method for vector optimization. *Comput. Optim. Appl.*, 54(3):473–493, 2013.

[23] J. C. Gilbert and J. Nocedal. Global convergence properties of conjugate gradient methods for optimization. *SIAM J. Optim.*, 2(1):21–42, 1992.

[24] M. L. N. Gonçalves, F. S. Lima, and L. F. Prudente. Globally convergent Newton-type methods for multiobjective optimization. *http://www.optimization-online.org/DB_FILE/2020/08/7955.pdf*, 2021.

[25] M. L. N. Gonçalves and L. F. Prudente. On the extension of the Hager–Zhang conjugate gradient method for vector optimization. *Comput. Optim. Appl.*, 76(3):889–916, 2020.

[26] L. M. Graña Drummond and A. N. Iusem. A projected gradient method for vector optimization problems. *Comput. Optim. Appl.*, 28(1):5–29, 2004.

[27] L. M. Graña Drummond, F. M. P. Raupp, and B. F. Svaiter. A quadratically convergent Newton method for vector optimization. *Optimization*, 63(5):661–677, 2014.

[28] L. M. Graña Drummond and B. F. Svaiter. A steepest descent method for vector optimization. *J. Comput. Appl. Math.*, 175(2):395 – 414, 2005.

[29] W. W. Hager and H. Zhang. A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM J. Optim.*, 16(1):170–192, 2005.

[30] M. R. Hestenes and E. Stiefel. *Methods of conjugate gradients for solving linear systems*, volume 49. Journal of Research of the National Bureau of Standards, 1952.

[31] C. Hillermeier. Generalized homotopy approach to multiobjective optimization. *J. Optimiz. Theory App.*, 110(3):557–583, 2001.

[32] S. Huband, P. Hingston, L. Barone, and L. While. A review of multiobjective test problems and a scalable test problem toolkit. *IEEE T. Evolut. Comput.*, 10(5):477–506, 2006.

[33] Y. Jin, M. Olhofer, and B. Sendhoff. Dynamic weighted aggregation for evolutionary multi-objective optimization: Why does it work and how? In *Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation*, GECCO'01, page 1042–1049, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[34] I. Kim and O. de Weck. Adaptive weighted-sum method for bi-objective optimization: Pareto front generation. *Struct. Multidiscip. O.*, 29(2):149–158, Feb 2005.

[35] M. Li and H. Feng. A sufficient descent LS conjugate gradient method for unconstrained optimization problems. *Appl. Math. Comput.*, 218(5):1577–1586, 2011.

[36] Y. Liu and C. Storey. Efficient generalized conjugate gradient algorithms, part 1: Theory. *J. Optim. Theory Appl.*, 69(1):129–137, Apr. 1991.

[37] A. Lovison. Singular continuation: Generating piecewise linear approximations to Pareto sets via global analysis. *SIAM J. Optim.*, 21(2):463–490, 2011.

[38] F. Lu and C.-R. Chen. Newton-like methods for solving vector optimization problems. *Appl. Anal.*, 93(8):1567–1586, 2014.

[39] D. T. Luc. *Theory of vector optimization. Lectures Notes in economics and mathematical systems, vol. 319.* Springer, Berlin, 1989.

[40] L. R. Lucambio Pérez and L. F. Prudente. Nonlinear conjugate gradient methods for vector optimization. *SIAM J. Optim.*, 28(3):2690–2720, 2018.

[41] L. R. Lucambio Pérez and L. F. Prudente. A Wolfe line search algorithm for vector optimization. *ACM Trans. Math. Softw.*, 45(4), Dec. 2019.

[42] E. Miglierina, E. Molho, and M. Recchioni. Box-constrained multi-objective optimization: A gradient-like method without a priori scalarization. *Eur. J. Oper. Res.*, 188(3):662–682, 2008.

[43] J. J. Moré, B. S. Garbow, and K. E. Hillstrom. Testing unconstrained optimization software. *ACM Trans. Math. Softw.*, 7(1):17–41, Mar. 1981.

[44] Y. Narushima and H. Yabe. A survey of sufficient descent conjugate gradient methods for unconstrained optimization. *SUT J. Math.*, 50(2):167–203, 2014.

[45] E. Polak and G. Ribière. Note sur la convergence de méthodes de directions conjuguées. *Revue française d'informatique et de recherche opérationnelle, série rouge*, 3(1):35–43, 1969.

[46] M. J. D. Powell. Nonconvex minimization calculations and the conjugate gradient method. In *Numerical analysis*, pages 122–141. Springer, 1984.

[47] M. Preuss, B. Naujoks, and G. Rudolph. Pareto set and EMOA behavior for simple multimodal multiobjective functions. In T. P. Runarsson, H.-G. Beyer, E. Burke, J. J. Merelo-Guervós, L. D. Whitley, and X. Yao, editors, *Parallel Problem Solving from Nature - PPSN IX*, pages 513–522, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[48] S. Qu, M. Goh, and F. T. Chan. Quasi-Newton methods for solving multiobjective optimization. *Oper. Res. Lett.*, 39(5):397–399, 2011.

[49] O. Schütze, M. Laumanns, C. A. Coello Coello, M. Dellnitz, and E.-G. Talbi. Convergence of stochastic search algorithms to finite size Pareto set approximations. *J. Global Optim.*, 41(4):559–577, Aug 2008.

[50] Z.-J. Shi and J. Shen. Convergence of descent method without line search. *Appl. Math. Comput.*, 167(1):94–107, 2005.

[51] Z.-J. Shi and J. Shen. Step-size estimation for unconstrained optimization methods. *Comput. Appl. Math.*, 24(3):399–416, 2005.

[52] Z.-J. Shi and J. Shen. Convergence of Liu–Storey conjugate gradient method. *Eur. J. Oper. Res.*, 182(2):552–560, 2007.

[53] P. L. Toint. Test problems for partially separable optimization and results for the routine PSPMIN. *The University of Namur, Department of Mathematics, Belgium, Tech. Rep.*, 1983.

[54] K. D. Villacorta and P. R. Oliveira. An interior proximal method in vector optimization. *Eur. J. Oper. Res.*, 214(3):485–492, 2011.

## Appendix A. Proofs of Theorems 4, 5, and 6

Assume that $\{x^k\}$ is a sequence generated by one of the proposed algorithms. Since $F(x^k) \preceq_K F(x^0)$ for all $k$, we obtain $\{x^k\} \subset \mathcal{L}$. Thus, taking into account assumption **(H)**, we conclude that $\{x^k\}$ is bounded. So, it follows from Lemma 3(c) and the continuity of $JF$ that there exist $\bar{\gamma} > 0$ and $\bar{c} > 0$ such that

$$\|d_{SD}(x^k)\| \leq \bar{\gamma}, \quad \|JF(x^k)\| \leq \bar{c}, \quad \forall k \geq 0. \tag{A.1}$$

Moreover, by the definition of $\mathcal{D}$, for every $k \geq 0$, it turns out that

$$0 > \mathcal{D}(x^k, d_{SD}(x^k)) \geq \langle JF(x^k)d_{SD}(x^k), w\rangle \geq -\|JF(x^k)\|\|d_{SD}(x^k)\| \geq -\bar{c}\bar{\gamma}, \tag{A.2}$$

because $\|w\| = 1$.

### *Appendix A.1. Proof of Theorem 4*

In the single-objective case, the global convergence of the (non-negative) PRP, HS and LS conjugate gradient methods were obtained by means of the so-called Property $(*)$ introduced in [23]; see also [44]. This property was extended to the vector setting in [40]. We will use Property $(*)$, which is formally stated below, to prove the convergence of Algorithm 1.

**Property $(*)$** *Consider a CG method and assume that there exist $\gamma$ and $\bar{\gamma}$ such that*

$$0 < \gamma \leq \|d_{SD}(x^k)\| \leq \bar{\gamma}, \quad \forall k \geq 0. \tag{A.3}$$

*Property $(*)$ holds if there exist constants $b > 1$ and $\lambda > 0$ such that*

$$|\beta_k| \leq b,$$

*and*

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

*for all $k \geq 1$, where $s^{k-1} := x^k - x^{k-1}$.*

The next result shows that, under mild assumptions, a (vector) CG method that has Property $(*)$ is globally convergent. Since its proof is similar to [40, Theorem 5.10], it is omitted here.

**Theorem 7.** *Consider a CG method with the following characteristics: (i) $\beta_k \geq 0$ for all $k$; (ii) $d^k$ satisfies the sufficient descent condition* (5) *at $x^k$ for all $k$; (iii) $\alpha_k$ satisfies the standard Wolfe conditions for all $k$; (iv) Property $(*)$ holds. Then,*

$$\liminf_{k\to\infty} \|d_{SD}(x^k)\| = 0.$$

We are now ready to prove Theorem 4.

Proof of Theorem 4. In view of Theorem 7, it suffices to show that Algorithm 1 has Property $(*)$. As observed in [44, pp.174], in order to show that a CG method has Property $(*)$, it is suffices to prove that, under assumption (A.3), there is a positive constant $M$ such that

$$|\beta_k| \leq M\|s^{k-1}\|, \quad \forall k \geq 1. \tag{A.4}$$

Assume that (A.3) holds. Then, by Lemma 3(b) and (A.2), it follows that

$$\frac{\gamma^2}{2} \leq -\mathcal{D}(x^k, d_{SD}(x^k)) \leq \bar{c}\bar{\gamma}, \quad \forall k \geq 0. \tag{A.5}$$

By **(H)**, Lemma 2(d) and (A.3), we obtain, for all $k \geq 1$,

$$\left|-\mathcal{D}(x^k, d_{SD}(x^k)) + \mathcal{D}(x^{k-1}, d_{SD}(x^k))\right| \leq L\|d_{SD}(x^k)\|\|x^k - x^{k-1}\| = L\bar{\gamma}\|s^{k-1}\|. \tag{A.6}$$

Since $(x^k, d^k)$ satisfies the sufficient descent condition (5), it follows from (A.6) and the first inequality in (A.5) that

$$\begin{aligned}\left|\beta_k^{LS}\right| \leq \frac{|-\mathcal{D}(x^k, d_{SD}(x^k)) + \mathcal{D}(x^{k-1}, d_{SD}(x^k))|}{-\mathcal{D}(x^{k-1}, d^{k-1})} &\leq \frac{L\bar{\gamma}\|s^{k-1}\|}{-c\mathcal{D}(x^{k-1}, d_{SD}(x^{k-1}))} \\ &\leq \frac{2L\bar{\gamma}}{c\gamma^2}\|s^{k-1}\|, \quad \forall k \geq 1.\end{aligned}$$

By defining $M = 2L\bar{\gamma}/c\gamma^2$, we conclude that (A.4) holds. Therefore, Algorithm 1 has Property $(*)$, completing the proof. $\square$

*Appendix A.2. Proof of Theorem 5*

Let us first establish some estimates on the parameter $\beta_k$ of Algorithm 2 which will help us in the convergence analysis.

**Lemma 8.** *Consider that Algorithm 2 generates an infinite sequence* $\{(x^k, d^k)\}$*. Assume that there exists* $\gamma > 0$ *such that* $\|d_{SD}(x^k)\| \geq \gamma$*, for all* $k \geq 0$*. Then, there exists* $M > 0$ *such that*

$$\eta^k \leq \beta_k, \qquad |\beta_k| \leq M\|s^{k-1}\|, \quad \forall k \geq 1.$$

PROOF. The first inequality follows from the definition of $\beta_k$ in (11). Now, as in Appendix A.1, taking into account (A.1) and (A.2), we obtain that (A.3) and (A.5) hold for all $k \geq 0$. Also, by (A.1), we have

$$\Lambda_k := \|JF(x^k) - JF(x^{k-1})\| \leq 2\bar{c}, \quad \forall k \geq 1.$$

Therefore, from the definition of $\beta_k$, Lemma 2(d), the last inequality, and the fact that $JF$ is Lipschitz continuous, we have

$$\begin{aligned} |\beta_k| \leq |\beta_k^{MLS}| &\leq \frac{|\mathcal{D}(x^{k-1}, d_{SD}(x^k)) - \mathcal{D}(x^k, d_{SD}(x^k)|}{-\mathcal{D}(x^{k-1}, d^{k-1})} + t\Lambda_k^2 \frac{|\mathcal{D}(x^k, d^{k-1})|}{\mathcal{D}^2(x^{k-1}, d^{k-1})} \\ &\leq \frac{L\|d_{SD}(x^k)\|\|s^{k-1}\|}{-\mathcal{D}(x^{k-1}, d^{k-1})} + t\frac{2\bar{c}L\|s^{k-1}\||\mathcal{D}(x^k, d^{k-1})|}{\mathcal{D}^2(x^{k-1}, d^{k-1})}, \quad \forall k \geq 1. \end{aligned}$$

Thus, using the strong Wolfe conditions at Step 3 of Algorithm 2 and the sufficient descent conditions (12) with $c := 1 - 1/(2t)$, it follows that

$$\begin{aligned} |\beta_k| &\leq \frac{L\|d_{SD}(x^k)\|\|s^{k-1}\|}{-\mathcal{D}(x^{k-1}, d^{k-1})} + t\frac{2\bar{c}L\|s^{k-1}\|\sigma}{-\mathcal{D}(x^{k-1}, d^{k-1})} \\ &\leq \frac{L\|d_{SD}(x^k)\|\|s^{k-1}\|}{-c\mathcal{D}(x^{k-1}, d_{SD}(x^{k-1}))} + t\frac{2\bar{c}L\|s^{k-1}\|\sigma}{-c\mathcal{D}(x^{k-1}, d_{SD}(x^{k-1}))}, \quad \forall k \geq 1. \end{aligned}$$

Hence, from (A.3) and (A.5), we obtain

$$|\beta_k| \leq \|s^{k-1}\| \left(2\frac{L\bar{\gamma} + 2t\bar{c}\sigma L}{c\gamma^2}\right), \quad \forall k \geq 1.$$

By defining $M := 2(L\bar{\gamma} + 2t\bar{c}\sigma L)/c\gamma^2$, we obtain the second inequality of the lemma. □

The next result will be useful to establish the convergence of Algorithm 2.

**Lemma 9.** *Consider that Algorithm 2 generates an infinite sequence* $\{(x^k, d^k)\}$*. Assume that there exists* $\gamma > 0$ *such that* $\|d_{SD}(x^k)\| \geq \gamma$*, for all* $k \geq 0$*. Then,* $d^k \neq 0$ *for all* $k \geq 0$*, and*

$$\sum_{k=1}^{\infty} \frac{1}{\|d^k\|^2} < \infty, \qquad \sum_{k=1}^{\infty} \|u^k - u^{k-1}\|^2 < \infty, \quad \textit{where} \quad u^k := \frac{d^k}{\|d^k\|}.$$

Proof. Using Lemma 8, the proof follows similarly to the one presented in [25, Lemma 6] . □

Proof of Theorem 5. Using Lemmas 8 and 9, the proof follows by contradiction similarly to the one presented in [25, Theorem 2]. □

*Appendix A.3. Proof of Theorem 6*

In order to prove Theorem 6, we first present some technical results.

**Lemma 10.** *Assume that $d \in \mathbb{R}^n$ is a $K$-descent direction for $F$ at $x$. Let $\rho \in (0,1)$, $c \in (0,1)$, and $e \in K$ as in (8) be given. Then, there exists $\bar{\alpha} > 0$ such that, for all $\alpha \in (0, \bar{\alpha})$,*

$$F(x + \alpha d) \preceq_K F(x) + \alpha\rho\mathcal{D}(x, d)e, \tag{A.7}$$

$$\mathcal{D}(x + \alpha d, d(x + \alpha d)) \leq c\mathcal{D}(x + \alpha d, d_{SD}(x + \alpha d)), \tag{A.8}$$

*where*

$$d(x+\alpha d) := d_{SD}(x+\alpha d) + \left( \frac{-\mathcal{D}(x + \alpha d, d_{SD}(x + \alpha d)) + \mathcal{D}(x, d_{SD}(x + \alpha d))}{-\mathcal{D}(x, d)} \right) d.$$

Proof. Since $\mathcal{D}(x, d) < 0$, it is easy to see that there exists $\alpha' > 0$ such that (A.7) holds for all $\alpha \in (0, \alpha')$; see, for example, [41, Lemma 1]. Now, define $\phi, \psi : \mathbb{R}_+ \longrightarrow \mathbb{R}$ by

$$\phi(\alpha) := \mathcal{D}(x + \alpha d, d(x + \alpha d)) \quad \text{and} \quad \psi(\alpha) := c\mathcal{D}(x + \alpha d, d_{SD}(x + \alpha d)).$$

Note that $\phi(0) < \psi(0)$. Thus, by continuity arguments, there exists $\alpha'' > 0$ such that $\phi(\alpha) \leq \psi(\alpha)$ for all $\alpha \in (0, \alpha'')$, i.e., (A.8) holds for all $\alpha \in (0, \alpha'')$. Therefore, by taking $\bar{\alpha} := \min(\alpha', \alpha'')$, we conclude that (A.7) and (A.8) hold for all $\alpha \in (0, \bar{\alpha})$. □

**Lemma 11.** *Algorithm 3 is well-defined.*

Proof. By using induction on $k$, the proof follows straightforwardly from Lemma 10. □

**Lemma 12.** *There holds:*

$$\|d^k\| \leq \left(1 + \frac{L(1-c)}{L_0}\right) \|d_{SD}(x^k)\|, \quad \forall k \geq 0. \tag{A.9}$$

PROOF. For $k = 0$, we have $d^0 = d_{SD}(x^0)$ and hence (A.9) is trivially satisfied. For $k \geq 1$, since $L_{k-1} \geq L_0$, it follows from the definition of $\alpha_{k-1}$ in Step 2 of Algorithm 3 that

$$\alpha_{k-1} \leq -\frac{(1-c)}{L_{k-1}} \frac{\mathcal{D}(x^{k-1}, d^{k-1})}{\|d^{k-1}\|^2} \leq -\frac{(1-c)}{L_0} \frac{\mathcal{D}(x^{k-1}, d^{k-1})}{\|d^{k-1}\|^2}, \tag{A.10}$$

because $\mathcal{D}(x^{k-1}, d^{k-1}) < 0$. Now, by the definition of $d^k$, assumption **(H)**, Lemma 2(d), and the fact that $x^k - x^{k-1} = \alpha_{k-1} d^{k-1}$, we have

$$\begin{aligned}
\|d^k\| &\leq \|d_{SD}(x^k)\| + \frac{|-\mathcal{D}(x^k, d_{SD}(x^k)) + \mathcal{D}(x^{k-1}, d_{SD}(x^k))|}{-\mathcal{D}(x^{k-1}, d^{k-1})} \|d^{k-1}\| \\
&\leq \|d_{SD}(x^k)\| + \frac{L\|d_{SD}(x^k)\|\|x^k - x^{k-1}\|}{-\mathcal{D}(x^{k-1}, d^{k-1})} \|d^{k-1}\| \\
&= \|d_{SD}(x^k)\| \left(1 + \frac{L\alpha_{k-1}\|d^{k-1}\|^2}{-\mathcal{D}(x^{k-1}, d^{k-1})}\right),
\end{aligned}$$

which, combined with (A.10), yields the desired inequality (A.9). $\square$

**Lemma 13.** *There exists $\tau_{\min} > 0$ such that $\tau_k > \tau_{\min}$ for all $k \geq 0$.*

PROOF. First, note that $L_k \leq \bar{M}$ for every $k \geq 0$. Thus, from the definition of $\tau_k$, (16), and Lemma 3(b), we have

$$\tau_k = -\frac{(1-c)}{L_k} \frac{\mathcal{D}(x^k, d^k)}{\|d^k\|^2} \geq -\frac{c(1-c)}{\bar{M}} \frac{\mathcal{D}(x^k, d_{SD}(x^k))}{\|d^k\|^2} > \frac{c(1-c)}{2\bar{M}} \frac{\|d_{SD}(x^k)\|^2}{\|d^k\|^2}.$$

Then, using (A.9), it follows that

$$\tau_k > \frac{c(1-c)}{2\bar{M}} \left(1 + \frac{L(1-c)}{L_0}\right)^{-2}.$$

By defining $\tau_{\min}$ as the right hand side of the above inequality, the proof is concluded. $\square$

We are now able to prove Theorem 6.

Proof of Theorem 6. Let $x^*$ be a limit point of the sequence $\{x^k\}$. Thus, there exists $\mathbb{K} := \{k_0, k_1, k_2, ...\} \underset{\infty}{\subset} \mathbb{N}$ such that $\lim_{k \in K} x^k = x^*$. Hence, by continuity of $F$, we have

$$\lim_{k \in \mathbb{K}} F(x^k) = F(x^*).$$

Noting that (8) implies $\varphi(-e) < 0$, by (15) and Lemma 1, we obtain, for all $k \geq 0$,

$$\begin{aligned}\varphi(F(x^{k+1})) \leq \varphi(F(x^k)) + \rho\alpha_k\varphi(\mathcal{D}(x^k, d^k)e) &= \varphi(F(x^k)) - \mathcal{D}(x^k, d^k)\rho\alpha_k\varphi(-e) \\ &\leq \varphi(F(x^k)),\end{aligned}$$

which, in particular, implies that $\{\varphi(F(x^k))\}$ is a nonincreasing sequence. Therefore, since $k_{j+1} \geq k_j + 1$, we have

$$\varphi(F(x^{k_{j+1}})) \leq \varphi(F(x^{k_j+1})) \leq \varphi(F(x^{k_j})) - \mathcal{D}(x^{k_j}, d^{k_j})\rho\alpha_{k_j}\varphi(-e) \leq \varphi(F(x^{k_j})),$$

for all $j \geq 0$. Then, by using continuity arguments, we have

$$\lim_{j \to \infty} \alpha_{k_j}\mathcal{D}(x^{k_j}, d^{k_j}) = 0.$$

As a consequence, there exists $\mathbb{K}_1 \underset{\infty}{\subset} \mathbb{K}$ such that: *(i)* $\lim_{k \in \mathbb{K}_1} \mathcal{D}(x^k, d^k) = 0$; or *(ii)* $\lim_{k \in \mathbb{K}_1} \alpha_k = 0$.
Case *(i)*: By (16) and Lemma 3(b), it follows that

$$\mathcal{D}(x^k, d^k) \leq c\mathcal{D}(x^k, d_{SD}(x^k)) \leq -\frac{c}{2}\|d_{SD}(x^k)\|^2 \leq 0.$$

Thus, taking limits for $k \in \mathbb{K}_1$, we conclude that $\lim_{k \in \mathbb{K}_1} d_{SD}(x^k) = 0$. Then, $d_{SD}(x^*) = 0$ and, hence, $x^*$ is a $K$-critical Pareto point of $F$.
Case *(ii)*: Assume, without loss of generality, that $\alpha_k < \tau_{\min}$ for all $k \in \mathbb{K}_1$, where $\tau_{\min} > 0$ is given as in Lemma 13. Thus, by Step 2 of the algorithm and Lemma 13, defining $\bar{\alpha}_k := \alpha_k/\mu$ for each $k \in \mathbb{K}_1$, it turns out that:

$$F(\bar{x}^k) \succ_K F(x^k) + \bar{\alpha}_k\rho\mathcal{D}(x^k, d^k)e, \tag{A.11}$$

or

$$\mathcal{D}(\bar{x}^k, d(\bar{x}^k)) > c\mathcal{D}(\bar{x}^k, d_{SD}(\bar{x}^k)), \tag{A.12}$$

where $\bar{x}^k := x^k + \bar{\alpha}_k d^k$. Assume that (A.11) holds infinitely many times, say for $\mathbb{K}_2 \underset{\infty}{\subset} \mathbb{K}_1$. Then, for all $k \in \mathbb{K}_2$, there exists $w_k \in C$ such that

$$\langle F(\bar{x}^k) - F(x^k), w_k \rangle > \bar{\alpha}_k \rho \mathcal{D}(x^k, d^k) \langle e, w_k \rangle \geq \bar{\alpha}_k \rho \mathcal{D}(x^k, d^k), \tag{A.13}$$

where the last inequality follows from (8). By the mean value theorem, there exists $\epsilon_k \in (0,1)$ such that

$$F(\bar{x}^k) - F(x^k) = \bar{\alpha}_k JF(x^k + \epsilon_k \bar{\alpha}_k d^k) d^k. \tag{A.14}$$

So, combining (A.13) and (A.14), and using the definition of $\mathcal{D}(\cdot,\cdot)$, we obtain

$$\rho \mathcal{D}(x^k, d^k) < \langle JF(x^k + \epsilon_k \bar{\alpha}_k d^k) d^k, w_k \rangle \leq \mathcal{D}(x^k + \epsilon_k \bar{\alpha}_k d^k, d^k), \tag{A.15}$$

for all $k \in \mathbb{K}_2$. It follows from Lemma 12 and (A.1) that $\{d^k\}$ is bounded and, as a consequence, there exist $\mathbb{K}_3 \underset{\infty}{\subset} \mathbb{K}_2$ and $d^* \in \mathbb{R}^n$ such that $\lim_{k \in \mathbb{K}_3} d^k = d^*$. Moreover, by the definition of $\bar{\alpha}_k$, we obtain $\lim_{k \in \mathbb{K}_3} \bar{\alpha}_k = 0$. Therefore, taking limits in (A.15) for $k \in \mathbb{K}_3$ and using Lemma 2(b), we get

$$\rho \mathcal{D}(x^*, d^*) \leq \mathcal{D}(x^*, d^*).$$

Hence, since $\rho < 1$ and $\mathcal{D}(x^k, d^k) < 0$ for all $k$, we obtain $\mathcal{D}(x^*, d^*) = 0$. Now, by (16) and Lemma 3(b), we have

$$\mathcal{D}(x^k, d^k) \leq c\mathcal{D}(x^k, d_{SD}(x^k)) \leq -\frac{c}{2}\|d_{SD}(x^k)\|^2 \leq 0.$$

Therefore, taking limits in the above inequality for $k \in \mathbb{K}_3$, we conclude that $d_{SD}(x^*) = 0$, proving that $x^*$ is a $K$-critical Pareto point of $F$.

Assume now, without loss of generality, that (A.12) holds and (A.11) is not true for all $k \in \mathbb{K}_1$. Thus, $\{\bar{x}^k\}_{k \in \mathbb{K}_1} \subset \mathcal{L}$, which implies, by using continuity arguments, that there exists $\bar{c} > 0$ such that

$$\|JF(\bar{x}^k)\| \leq \bar{c}, \quad \forall k \in \mathbb{K}_1. \tag{A.16}$$

Define $\chi := \operatorname{sgn}\left(-\mathcal{D}(\bar{x}^k, d_{SD}(\bar{x}^k)) + \mathcal{D}(x^k, d_{SD}(\bar{x}^k))\right)$. From the definitions of $\bar{x}^k$ and $d(\bar{x}^k)$, and Lemma 2, we have

$$\begin{aligned}
\mathcal{D}(\bar{x}^k, d(\bar{x}^k)) &\leq \mathcal{D}(\bar{x}^k, d_{SD}(\bar{x}^k)) \\
&\quad + \frac{|-\mathcal{D}(\bar{x}^k, d_{SD}(\bar{x}^k)) + \mathcal{D}(x^k, d_{SD}(\bar{x}^k))|}{-\mathcal{D}(x^k, d^k)} \mathcal{D}(\bar{x}^k, \chi d^k) \\
&\leq \mathcal{D}(\bar{x}^k, d_{SD}(\bar{x}^k)) + \frac{L\|d_{SD}(\bar{x}^k)\|\|\bar{x}^k - x^k\|}{-\mathcal{D}(x^k, d^k)} |\mathcal{D}(\bar{x}^k, \chi d^k)| \\
&\leq \mathcal{D}(\bar{x}^k, d_{SD}(\bar{x}^k)) + \frac{L\bar{\alpha}_k\|d_{SD}(\bar{x}^k)\|\|d^k\|}{-\mathcal{D}(x^k, d^k)} |\mathcal{D}(\bar{x}^k, \chi d^k)|.
\end{aligned} \tag{A.17}$$

Therefore, combining (A.12) and (A.17), we obtain

$$-(1-c)\mathcal{D}(\bar{x}^k, d_{SD}(\bar{x}^k)) < \frac{L\bar{\alpha}_k \|d_{SD}(\bar{x}^k)\|\|d_k\|}{-\mathcal{D}(x^k, d^k)} |\mathcal{D}(\bar{x}^k, \chi d^k)|, \forall k \in \mathbb{K}_1. \quad \text{(A.18)}$$

By the definition of $\mathcal{D}(\cdot, \cdot)$, there exists $\bar{w}^k \in C$ such that

$$|\mathcal{D}(\bar{x}^k, \chi d^k)| = |\langle JF(\bar{x}^k)(\chi d^k), \bar{w}^k\rangle| \leq \|JF(\bar{x}^k)\|\|d^k\| \leq \bar{c}\|d^k\|, \quad \text{(A.19)}$$

where the inequalities is due to $\|\bar{w}\| = 1$ and (A.16). By (A.18), (A.19) and Lemma 3(b), we find

$$(1-c)\frac{\|d_{SD}(\bar{x}^k)\|^2}{2} < \frac{\bar{c}L\bar{\alpha}_k\|d_{SD}(\bar{x}^k)\|\|d^k\|^2}{-\mathcal{D}(x^k, d^k)}, \quad \forall k \in \mathbb{K}_1,$$

or, equivalently,

$$\bar{\alpha}_k > -\frac{(1-c)}{2\bar{c}L}\frac{\mathcal{D}(x^k, d^k)}{\|d^k\|^2}\|d_{SD}(\bar{x}^k)\|, \quad \forall k \in \mathbb{K}_1.$$

Hence, from the definition of $\bar{\alpha}_k$ and (16), we obtain

$$\alpha_k > -\frac{\mu(1-c)}{2\bar{c}L}\frac{c\mathcal{D}(x^k, d_{SD}(x^k))}{\|d^k\|^2}\|d_{SD}(\bar{x}^k)\|, \quad \forall k \in \mathbb{K}_1,$$

which, combined with Lemmas 3(b) and 12, yields

$$\begin{aligned} \alpha_k &> \frac{\mu c(1-c)}{4\bar{c}L}\frac{\|d_{SD}(x^k)\|^2}{\|d^k\|^2}\|d_{SD}(\bar{x}^k)\| \\ &\geq \frac{\mu c(1-c)}{4\bar{c}L}\left(1 + \frac{L(1-c)}{L_0}\right)^{-2}\|d_{SD}(\bar{x}^k)\|, \end{aligned} \quad \text{(A.20)}$$

for all $k \in \mathbb{K}_1$. From Lemma 12 and (A.1), it turns out that $\{d^k\}$ is bounded. So, since $\lim_{k\in\mathbb{K}_1} \bar{\alpha}_k = 0$, it follows from the definition of $\bar{x}^k$ that $\lim_{k\in\mathbb{K}_1} \bar{x}^k = x^*$. Therefore, taking limits in (A.20) for $k \in \mathbb{K}_1$, we conclude that $d_{SD}(x^*) = 0$. Thus, $x^*$ is a $K$-critical Pareto point of $F$ and the proof is complete. $\square$