Conditional gradient method for multiobjective optimization

P. B. Assunção * O. P. Ferreira * L. F. Prudente *

April 7, 2020

Abstract: We analyze the *conditional gradient method*, also known as *Frank-Wolfe method*, for constrained multiobjective optimization. The constraint set is assumed to be convex and compact, and the objectives functions are assumed to be continuously differentiable. The method is considered with different strategies for obtaining the step sizes. Asymptotic convergence properties and iteration-complexity bounds with and without convexity assumptions on the objective functions are stablished. Numerical experiments are provided to illustrate the effectiveness of the method and certify the obtained theoretical results.

Keywords: Conditional gradient method; multiobjective optimization; Pareto optimality; constrained optimization problem

1 Introduction

In the constrained multiobjective optimization, we seek to simultaneously minimize several objective functions on a set $\mathcal{C}$. One strategy for solving multiobjective problems that has become very popular consists in the extension of methods for scalar-valued to multiobjective-valued optimization, instead of using scalarization approaches [23]. As far as we know, this strategy originated from the work [14] in which the authors proposed the steepest descent methods for multiobjective optimization. Since of then, new properties related to this method have been discovered and several variants of it have been considered, see for example [4, 5, 16, 19, 20, 26, 27]. In recent years, there has been a significant increase in the number of papers addressing concepts, techniques, and methods for multiobjective optimization, see for example [6, 9, 13, 15, 25, 41, 42, 45, 46, 48, 53, 54]. Following this trend, the goal of present paper is analyze the conditional gradient method in this setting.

The *conditional gradient method* also known as *Frank-Wolfe optimization algorithm* is one of the oldest iterative methods for finding minimizers of differentiable functions onto compact convex sets. Its long history began in 1956 with the work of Frank and Wolfe for minimizing convex quadratic functions over compact polyhedral sets, see [17]. Ten years later, the method was extended to minimize differentiable convex functions with Lipschitz gradients over compact convex feasible sets, see [39]. Since then, this method has attracted the attention of the scientific community working on this subject. One of the factors that explains this interest is its simplicity and ease of implementation: at each iteration, the method requires only access to a linear

*Instituto de Matemática e Estatística, Universidade Federal de Goiás, CEP 74001-970 - Goiânia, GO, Brazil, E-mails: `pedro.ufg.mat@gmail.com`, `orizon@ufg.br`, `lfprudente@ufg.br`. The authors was supported in part by CNPq grants 305158/2014-7 and 302473/2017-3, FAPEG/PRONEM- 201710267000532 and CAPES.

minimization oracle over a compact convex set. In particular, allowing a low cost of storage and ready exploitation of separability and sparsity, it makes the application of the conditional gradient method in large scale problems very attractive. It is worth mentioning that, in recent years, there has been an increase in the popularity of this method due to the emergence of machine learning applications, see [31, 35, 36]. For these reasons, several variants of this method have emerged and properties of it have been discovered throughout the years, resulting in a wide literature on the subject. Papers that address this method include, for example, [3, 8, 18, 24, 28, 34, 37, 43].

The aim in this paper is twofold. First, asymptotic analysis to multiobjective conditional gradient method will be done for Armijo and adaptative step sizes. Second, iteration-complexity bounds will be stablished for Armijo, adaptative, and diminishing step sizes. Numerical experiments will be provided to illustrate the effectiveness of the method in this new setting and certify the obtained theoretical results.

The organization of this paper is as follows. In Section 2, some notations and auxiliary results, used throughout of the paper, are presented. In Section 3, we present the algorithm and the step size strategies that will be considered. The results related to Armijo's step sizes will be presented in Section 4 and the ones related to adaptative, and diminishing step sizes in Section 5. In Section 6, we present some numerical experiments. Finally, some conclusions are given in Section 7.

Notation. The symbol $\langle \cdot, \cdot \rangle$ is the usual inner product in $\mathbb{R}^n$ and $\|\cdot\|$ denotes the Euclidean norm. If $\mathbb{K} = \{k_1, k_2, \ldots\} \subseteq \mathbb{N}$, with $k_j < k_{j+1}$ for all $j \in \mathbb{N}$, then we denote $\mathbb{K} \underset{\infty}{\subset} \mathbb{N}$.

2 Preliminaries

In this section, we present the multicriteria problem studied in the present work, the first order optimality condition for it, and some notations and definitions. Let $\mathcal{J} := \{1, \ldots, m\}$, $\mathbb{R}^m_+ := \{u \in \mathbb{R}^m : \ u_j \geq 0, j \in \mathcal{J}\}$, and $\mathbb{R}^m_{++} = \{u \in \mathbb{R}^m : u_j > 0, j \in \mathcal{J}\}$. For u, $v \in \mathbb{R}^m_+$, $v \succeq u$ (or $u \preceq v$) means that $v - u \in \mathbb{R}^m_+$ and $v \succ u$ (or $u \prec v$) means that $v - u \in \mathbb{R}^m_{++}$. Consider the problem of finding a *optimum Pareto point* of $F : \mathbb{R}^n \to \mathbb{R}^m$ in a compact convex set $\mathcal{C} \subset \mathbb{R}^n$, i.e., a point $x^* \in \mathcal{C}$ such that there exists no other $x \in \mathcal{C}$ with $F(x) \preceq F(x^*)$ and $F(x) \neq F(x^*)$. We denote this constrained problem as

$$\min_{x \in \mathcal{C}} F(x). \tag{1}$$

A point $x^* \in \mathcal{C}$ is called *weak Pareto optimal point* for Problem (1), if there exists no other $x \in \mathcal{C}$ such that $F(x) \prec F(x^*)$. *Throughout of the paper we assume that* $F : \mathbb{R}^n \to \mathbb{R}^m$ *is a continuously differentiable function given by* $F := (f_1, \ldots, f_m)$. The Jacobian of F at $x \in \mathbb{R}^n$ denoted by JF is the $m \times n$ matrix with entries

$$[JF(x)]_{i,j} = \frac{\partial F_i}{\partial x_j}(x),$$

and $\mathrm{Im}(JF(x))$ stands for the image on $\mathbb{R}^m$ by $JF(x)$.

Definition 1. *The function* $F := (f_1, \ldots, f_m)$ *has Jacobian* JF *componentwise Lipschitz continuous, if there exist constants* $L_1, L_2, \ldots, L_m > 0$ *such that*

$$\|\nabla f_j(x) - \nabla f_j(y)\| \leq L_j \|x - y\|, \qquad \forall\ x, y \in \mathcal{C}, \qquad \forall\ j \in \mathcal{J}.$$

For future reference we set $L := \max\{L_j : \ j \in \mathcal{J}\}$.

Next lemma presents a result related to vector functions satisfying Definition 1. Hereafter, we denote:

$$e := (1, \dots, 1)^T \in \mathbb{R}^m.$$

Lemma 1. *Assume that $F := (f_1, \dots, f_m)$ satisfies Definition 1. Then, for all $x, p \in C$ and $\lambda \in (0, 1]$, there holds*

$$F\big(x + \lambda(p - x)\big) \preceq F(x) + \left(\theta\lambda + \frac{L}{2}\|p - x\|^2\lambda^2\right) e. \tag{2}$$

where $\theta := \max_{j\in\mathcal{J}} \big\langle \nabla f_j(x), p - x \big\rangle$.

Proof. Since F satisfies Definition 1, by using the same idea in the proof of [11, Lemma 2.4.2], we conclude that

$$F\big(x + \lambda(p - x)\big) \preceq F(x) + \lambda JF(x)(p - x) + \frac{L}{2}\|p - x\|^2\lambda^2 e.$$

On the other hand, $\theta \geq \langle \nabla f_j(x), p - x\rangle$, for all $j \in \mathcal{J}$. Hence,

$$JF(x)(p - x) = \big(\langle \nabla f_1(x), p - x\rangle, \dots, \langle \nabla f_m(x), p - x\rangle\big)^T \preceq \theta e.$$

Therefore, (2) follows by combining the two previous vector inequalities. $\square$

The first order optimality condition for Problem (1) of a point $\bar{x} \in C$ is

$$-\mathbb{R}^m_{++} \cap JF(\bar{x})(C - \bar{x}) = \varnothing, \tag{3}$$

where $\mathcal{C} - x := \{y - x : \ y \in \mathcal{C}\}$. The geometric optimality condition (3) is also equivalently stated as

$$\max_{j\in\mathcal{J}} \big\langle \nabla f_j(\bar{x}), p - \bar{x} \big\rangle \geq 0, \qquad \forall\, p \in \mathcal{C}. \tag{4}$$

In general, condition (3) is necessary, but not sufficient for the optimality. Thus, a point $\bar{x} \in \mathbb{R}^n$ satisfying (3) is called a *critical Pareto point* or a *stationary point* of Problem (1). The function F is said to be *convex* on $\mathcal{C}$ if

$$F\big(\lambda x + (1 - \lambda)y\big) \preceq \lambda F(x) + (1 - \lambda)F(y), \quad \forall\, x, y \in \mathcal{C}, \quad \forall\, \lambda \in [0, 1],$$

or equivalently, for each $j \in \mathcal{J}$, the component function $f_j : \mathbb{R}^n \to \mathbb{R}$ of F is convex. If F is convex on C, then

$$f_j(y) - f_j(x) \geq \big\langle \nabla f_j(x),\ y - x \big\rangle, \quad \forall\, x, y \in \mathcal{C}, \quad \forall\, j \in \mathcal{J}.$$

Next lemma shows that, in the convex case, the concepts of stationarity and weak Pareto optimality are equivalent. Since this is a well known result (see, for example, [26]), we will omit its proof here.

Lemma 2. *If F is convex and $\bar{x}$ is a critical Pareto point, then $\bar{x}$ is also a weak Pareto optimal point of Problem* (1).

Since we are assuming that $\mathcal{C} \subset \mathbb{R}^n$ is a compact set, its *diameter* is a finite number defined by

$$\operatorname{diam}(\mathcal{C}) := \max\big\{\|x - y\| :\ x, y \in \mathcal{C}\big\}.$$

We end this section stating two results for sequences of real numbers, which will be useful for our study on iteration complexity bounds for the conditional gradient method. Their proofs can be found in [49, Lemma 6] and [2, Lemma 13.13], respectively.

Lemma 3. *Let $\{a_k\}$ be a nonnegative sequence of real numbers, if $\Gamma a_k^2 \leq a_k - a_{k+1}$ for some $\Gamma > 0$ and for any $k = 1, ..., \ell$, then*

$$a_\ell \leq \frac{a_0}{1 + \ell \Gamma a_0} < \frac{1}{\Gamma \ell}.$$

Lemma 4. *Let p be a positive integer, and let $\{a_k\}$ and $\{b_k\}$ be nonnegative sequences of real numbers satisfying*

$$a_{k+1} \leq a_k - b_k \beta_k + \frac{A}{2}\beta_k^2, \qquad k = 0, 1, 2, \dots,$$

where $\beta_k = 2/(k+2)$ and A is a positive number. Suppose that $a_k \leq b_k$, for all k. Then

(i) $a_k \leq \dfrac{2A}{k}$*, for all $k = 1, 2, \dots$.*

(ii) $\min_{\ell \in \{\lfloor \frac{k}{2} \rfloor + 2, \cdots, k\}} b_\ell \leq \dfrac{8A}{k-2}$*, for all $k = 3, 4, \dots,$ where, $\lfloor k/2 \rfloor = \max\left\{n \in \mathbb{N} : \; n \leq k/2\right\}$.*

3 The conditional gradient method

The search direction of the conditional gradient (CondG) method at a given $x \in \mathcal{C}$ is defined to be $d(x) := p(x) - x$, where $p(x)$ an optimal solution of the scalar-valued problem

$$\min_{u \in \mathcal{C}} \max_{j \in \mathcal{J}} \left\langle \nabla f_j(x), u - x \right\rangle, \tag{5}$$

i.e.,

$$p(x) \in \operatorname{argmin}_{u \in \mathcal{C}} \max_{j \in \mathcal{J}} \left\langle \nabla f_j(x), u - x \right\rangle. \tag{6}$$

Note that, since the minimand of (5) is a convex function and $\mathcal{C}$ is a compact convex set, this problem has an optimal solution (possibly not unique) and, as consequence, $p(x)$ is well defined. Although (5) is a constrained convex non-differentiable problem, a solution of it can be calculated by solving for $\tau \in \mathbb{R}$ and $u \in \mathcal{C}$ the following constrained linear problem

$$\begin{array}{ll} \min_{u,\tau} & \tau \\ \text{s.t.} & \left\langle \nabla f_j(x), u - x \right\rangle \leq \tau, \quad j \in \mathcal{J}, \\ & u \in \mathcal{C}. \end{array} \tag{7}$$

Whenever $\mathcal{C}$ is compact and has a simples structure, the solution of problem (7) can be obtained by using a linear optimization oracle. Denote by $\theta(x)$ the optimal value of (5) given by

$$\theta(x) := \; \max_{j \in \mathcal{J}} \left\langle \nabla f_j(x), p(x) - x \right\rangle, \tag{8}$$

where $p(x) \in \mathcal{C}$ is as in (6). Following, we characterize $\theta(x)$ with respect to stationarity. Next proposition is a variation of the results in [26, Propositions 3 and 4] (see also [21, Proposition 4.1]).

Proposition 5. *Let $\theta : \mathcal{C} \to \mathbb{R}$ be as in* (8)*. Then, there hold:*

(i) $\theta(x) \leq 0$ *for all* $x \in \mathcal{C}$*;*

(ii) $\theta(\cdot)$ *is continuous;*

(iii) $x \in \mathcal{C}$ *is stationary if and only if* $\theta(x) = 0$*.*

Proof. (*i*) Since $x \in \mathcal{C}$, it follows from (6) and (8) that $\theta(x) \leq \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x), x - x \right\rangle = 0$.
(*ii*) Let $x \in \mathcal{C}$ and consider a sequence $\{x^k\} \subset \mathcal{C}$ such that $\lim_{k\to\infty} x^k = x$. On one hand, since $p(x) \in \mathcal{C}$, using (6) and (8) we have $\theta(x^k) \leq \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x^k), p(x) - x^k \right\rangle$ which implies that

$$\limsup_{k\to\infty} \theta(x^k) \leq \theta(x), \tag{9}$$

because F is continuously differentiable and $\mathbb{R}^m \ni u \mapsto \max_{j\in\mathcal{J}} u_j$ is continuous. On the other hand, since $p(x^k) \in \mathcal{C}$, we obtain

$$\begin{aligned}
\theta(x) &\leq \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x), p(x^k) - x \right\rangle \\
&= \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x), p(x^k) - x^k + x^k - x \right\rangle \\
&\leq \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x), p(x^k) - x^k \right\rangle + \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x), x^k - x \right\rangle.
\end{aligned}$$

Therefore, taking $\liminf_{k\to\infty}$ on both sides of the above inequality and using continuity arguments, we have

$$\begin{aligned}
\theta(x) &\leq \liminf_{k\to\infty} \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x), p(x^k) - x^k \right\rangle \\
&= \liminf_{k\to\infty} \left(\theta(x^k) + \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x), p(x^k) - x^k \right\rangle - \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x^k), p(x^k) - x^k \right\rangle \right) \\
&\leq \liminf_{k\to\infty} \left(\theta(x^k) + \left\| JF(x^k) - JF(x) \right\| \left\| p(x^k) - x^k \right\| \right),
\end{aligned}$$

where the second inequality holds because $u \mapsto \max_{j\in\mathcal{J}} u_j$ is Lipschitz continuos with constant 1. Since F is continuously differentiable, $\mathcal{C}$ is compact, and $\left\| p(x^k) - x^k \right\| \leq \operatorname{diam}(\mathcal{C})$, we obtain

$$\theta(x) \leq \liminf_{k\to\infty} \theta(x^k). \tag{10}$$

Combining (9) and (10) yields $\lim_{k\to\infty} \theta(x^k) = \theta(x)$, concluding the proof of the second statement.
(*iii*) Assume that $x \in \mathcal{C}$ is stationary point of Problem (1), i.e., $\max_{j\in\mathcal{J}} \left\langle \nabla f_j(x), u - x \right\rangle \geq 0$ for all $u \in \mathcal{C}$. Since $p(x) \in \mathcal{C}$, we obtain $\theta(x) = \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x), p(x) - x \right\rangle \geq 0$ which, together with item (i), implies $\theta(x) = 0$. Reciprocally, suppose that $\theta(x) = 0$. It follows from the definition of $\theta(\cdot)$ in (8) that

$$0 = \theta(x) \leq \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x), u - x \right\rangle, \qquad \forall\, u \in \mathcal{C},$$

which implies that x satisfies the optimality condition (4). □

A direct consequence of Proposition 5 is that if $x \in \mathcal{C}$ is a nonstationary point of Problem (1), then $\theta(x) < 0$ and $p(x) \neq x$. This means that, in this case, $d(x) := p(x) - x$ is nonnull and is a descent direction for F at x in the sense that $\left\langle \nabla f_j(x), d(x) \right\rangle \leq \theta(x) < 0$ for all $j \in \mathcal{J}$. In the following, we present formally the conditional gradient method for multiobjective optimization problems.

Algorithm 1: CondG method for multiobjective optimization

Step 0. *Initialization*
Choose $x^0 \in \mathcal{C}$ and initialize $k \leftarrow 0$.

Step 1. *Compute the search direction*
Compute an optimal solution $p(x^k)$ and the optimal value $\theta(x^k)$ as

$$p(x^k) \in \arg\min_{u\in\mathcal{C}} \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x^k), u - x^k \right\rangle$$

and

$$\theta(x^k) := \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x^k), p(x^k) - x^k \right\rangle. \tag{11}$$

Define the search direction by $d(x^k) := p(x^k) - x^k$.

Step 2. *Stopping criteria*
If $\theta(x^k) = 0$, then **stop**.

Step 3. *Compute the step size and iterate*
Compute $\lambda_k \in (0, 1]$ and set

$$x^{k+1} := x^k + \lambda_k d(x^k). \tag{12}$$

Step 4. *Beginning a new iteration*
Set $k \leftarrow k + 1$ and go to **Step 1**.

In view of Proposition 5, Algorithm 1 successfully stops if a stationary point of Problem 1 is found. Thus, hereafter, *we assume that* $\theta(x^k) < 0$ *for all* $k = 0, 1, \ldots$, meaning that Algorithm 1 generates an infinite sequence $\{x^k\}$. Since $x^0 \in \mathcal{C}$, $p(x^k) \in \mathcal{C}$ and $\lambda_k \in (0, 1]$ for all $k = 0, 1, \ldots$, and C is a convex set, it follows from inductive arguments that $\{x^k\} \subset \mathcal{C}$. Therefore, the proposed CondG method is a projection-free algorithm that generates feasible iterates. The choice for computing the step size λ_k at Step 2 remains deliberately open. In the next sections, we study convergence properties of the sequence generated by Algorithm 1 with three different well defined strategies for the step sizes shown below.

Armijo step size. *Let* $\zeta \in (0, 1)$ *and* $0 < \omega_1 < \omega_2 < 1$. *The step size* λ_k *is chosen according the following line search algorithm:*

Step LS0. *Set* $\lambda_{\text{trial}} = 1$.

Step LS1. *If*

$$F\left(x^k + \lambda_{\text{trial}}[p(x^k) - x^k]\right) \preceq F(x^k) + \zeta\lambda_{\text{trial}}\theta(x^k)e,$$

then set $\lambda_k := \lambda_{trial}$ *and return to the main algorithm.*

Step LS2. *Find* $\lambda_{\text{new}} \in [\omega_1\lambda_{\text{trial}}, \omega_2\lambda_{\text{trial}}]$, *set* $\lambda_{\text{trial}} \leftarrow \lambda_{\text{new}}$, *and go to Step LS1.*

Next proposition shows that the line search algorithm of Armijo's step size is well defined.

Proposition 6. *Let* $\zeta \in (0, 1)$, $x \in \mathcal{C}$ *be a nonstationary point, and* $p(x)$ *and* $\theta(x)$ *as in* (6) *and* (8), *respectively. Then, there exists* $0 < \overline{\eta} \leq 1$ *such that*

$$F\left(x + \eta[p(x) - x]\right) \prec F(x) + \zeta\eta\theta(x)e, \qquad \forall\ \eta \in (0, \overline{\eta}].$$

As a consequence, the line search algorithm of the Armijo step size is well-defined.

Proof. Since $\langle \nabla f_j(x), p(x) - x \rangle \leq \theta(x) < 0$ for all $j \in \mathcal{J}$, the proof follows directly from [14, Lemma 4]. $\square$

Adaptative step size. *Assume that $F := (f_1, \ldots, f_m)$ satisfies Definition 1. Define the step size as*

$$\lambda_k := \min \left\{ 1, \frac{-\theta(x^k)}{L\|p(x^k) - x^k\|^2} \right\} = \operatorname{argmin}_{\lambda \in (0,1]} \left\{ \theta(x^k)\lambda + \frac{L}{2}\|p(x^k) - x^k\|^2 \lambda^2 \right\}. \tag{13}$$

Since $\theta(x) < 0$ and $p(x) \neq x$ for nonstationary points, the adaptative step size for Algorithm 1 is well defined. We end this section showing the third step size strategy.

Diminishing step size. *Define the step size as*

$$\lambda_k := \frac{2}{k+2}.$$

4 Analysis of CondG with Armijo's step size

Throughout this section we assume that $\{x^k\}$ is generated by Algorithm 1 with the Armijo step size strategy. Next result is a partial asymptotic convergence property that requires neither convexity nor Lipschitz assumptions on F.

Theorem 7. *Let $\{x^k\}$ be generated by Algorithm 1 with the Armijo step size strategy. Then, every limit point $\bar{x}$ of $\{x^k\}$ is a Pareto critical point for the Problem 1.*

Proof. Let $\bar{x} \in \mathcal{C}$ be a limit point of $\{x^k\}$ and consider $\mathbb{K} \underset{\infty}{\subset} \mathbb{N}$ such that $\lim_{k \in \mathbb{K}} x^k = \bar{x}$. According to the Armijo step size strategy and (12), $\{F(x^k)\}$ satisfies

$$0 \prec -\zeta \lambda_k \theta(x^k) e \preceq F(x^k) - F(x^{k+1}), \qquad k = 0, 1, \ldots. \tag{14}$$

Considering that F is continuous, we have $\lim_{k \in \mathbb{K}} F(x^k) = F(\bar{x})$. Thus, due to $\{F(x^k)\}$ be monotone decreasing, it follows that $\lim_{k \to \infty} F(x^k) = F(\bar{x})$. Then, taking limits in (14), we obtain $0 \preceq \lim_{k \to \infty} -\zeta \lambda_k \theta(x^k) e \preceq \lim_{k \to \infty}[F(x^k) - F(x^{k+1})] = 0$. Hence, $\lim_{k \to \infty} \lambda_k \theta(x^k) = 0$ which, in particular, implies $\lim_{k \in \mathbb{K}} \lambda_k \theta(x^k) = 0$. Therefore, there exists $\mathbb{K}_1 \underset{\infty}{\subset} \mathbb{K}$ such that at least one of the two following possibilities holds:

(a) $\lim_{k \in \mathbb{K}_1} \theta(x^k) = 0$;

(b) $\lim_{k \in \mathbb{K}_1} \lambda_k = 0$.

In case (a), by the continuity of $\theta(\cdot)$, we obtain $\theta(\bar{x}) = 0$. Thus, Proposition 5 (iii) implies that $\bar{x}$ is Pareto critical. Now consider case (b). Without loss of generality, assume that $\lambda_k < 1$ for all $k \in \mathbb{K}_1$ and that there exists $\bar{p} \in \mathcal{C}$ such that $\lim_{k \in \mathbb{K}_1} p(x^k) = \bar{p}$ (recall that $\mathcal{C}$ is compact and $\{p(x^k)\} \subset \mathcal{C}$). Therefore, by the Armijo step size strategy, for all $k \in \mathbb{K}_1$, there exists $\hat{\lambda}_k \in (0, \lambda_k/\omega_1]$ such that

$$F\left(x^k + \hat{\lambda}_k[p(x^k) - x^k]\right) \not\preceq F(x^k) + \zeta \hat{\lambda}_k \theta(x^k) e,$$

which means that

$$f_{j_k}\left(x^k + \hat{\lambda}_k[p(x^k) - x^k]\right) > f_{j_k}(x^k) + \zeta\hat{\lambda}_k\theta(x^k),$$

for at least one $j_k \in \mathcal{J}$. Since $\mathcal{J}$ is a finite set of indexes, there exist $\mathbb{K}_2 \underset{\infty}{\subset} \mathbb{K}_1$ and $j_* \in \mathcal{J}$ such that, for all $k \in \mathbb{K}_2$, we have

$$f_{j_*}\left(x^k + \hat{\lambda}_k[p(x^k) - x^k]\right) > f_{j_*}(x^k) + \zeta\hat{\lambda}_k\theta(x^k). \tag{15}$$

On the other hand, by the mean value theorem, for all $k \in \mathbb{K}_2$, there exists $\xi_k \in [0,1]$ such that

$$\left\langle \nabla f_{j_*}\left(x^k + \xi_k\hat{\lambda}_k[p(x^k) - x^k]\right), \hat{\lambda}_k[p(x^k) - x^k]\right\rangle = f_{j_*}\left(x^k + \hat{\lambda}_k[p(x^k) - x^k]\right) - f_{j_*}(x^k).$$

Therefore, by (11) and (15), for all $k \in \mathbb{K}_2$, we have

$$\left\langle \nabla f_{j_*}\left(x^k + \xi_k\hat{\lambda}_k[p(x^k) - x^k]\right), \hat{\lambda}_k[p(x^k) - x^k]\right\rangle > \zeta\hat{\lambda}_k\left\langle \nabla f_{j_*}(x^k), p(x^k) - x^k\right\rangle.$$

Since $\hat{\lambda}_k \in (0, \lambda_k/\omega_1]$, it follows that $\lim_{k\in\mathbb{K}_2}\hat{\lambda}_k\|p(x^k) - x^k\| = 0$. Thus, dividing both sides of the above inequality by $\hat{\lambda}_k > 0$ and taking limits for $k \in \mathbb{K}_2$, we obtain

$$\left\langle \nabla f_{j_*}(\bar{x}), \bar{p} - \bar{x}\right\rangle \geq \zeta\left\langle \nabla f_{j_*}(\bar{x}), \bar{p} - \bar{x}\right\rangle.$$

Owing to $\zeta \in (0,1)$, this implies that

$$\left\langle \nabla f_{j_*}(\bar{x}), \bar{p} - \bar{x}\right\rangle \geq 0. \tag{16}$$

On the other hand, since $\theta(x^k) < 0$ for all k, we have $\lim_{k\in\mathbb{K}_2}\theta(x^k) = \max_{j\in\mathcal{J}}\left\langle \nabla f_j(\bar{x}), \bar{p} - \bar{x}\right\rangle \leq 0$. Therefore, using (16), we conclude that $\lim_{k\in\mathbb{K}_2}\theta(x^k) = 0$ and, by the continuity of $\theta(\cdot)$, we obtain $\theta(\bar{x}) = 0$. Thus, Proposition 5 (iii) implies that $\bar{x}$ is Pareto critical, which concludes the proof. □

Next, we present our first iteration-complexity bounds for Algorithm 1. For simplicity, let us define the following positive constant:

$$0 < \rho := \sup\{\|\nabla f_j(x)\| : \ x \in \mathcal{C}, \ j \in \mathcal{J}\}. \tag{17}$$

Theorem 8. *Assume that $F := (f_1, \ldots, f_m)$ is convex on $\mathcal{C}$ and satisfies Definition 1. Let $\{x^k\}$ be generated by Algorithm 1 with the Armijo step size strategy. Assume that $\lim_{k\to+\infty} F(x^k) = F(x^*)$. Then, the following inequality holds*

$$\min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right) \leq \frac{1}{\gamma\zeta}\frac{1}{k}, \qquad \gamma := \min\left\{\frac{1}{\rho\,\mathrm{diam}(\mathcal{C})}, \ \frac{2\omega_1(1-\zeta)}{L\,\mathrm{diam}(\mathcal{C})^2}\right\}. \tag{18}$$

for all $k = 1, 2, \ldots$.

Proof. We first claim that

$$\lambda_k \geq -\gamma\theta(x^k), \qquad k = 0, 1, \ldots, \tag{19}$$

where $\gamma > 0$ is defined in (18). Since $\lambda_k \in (0,1]$, for all $k = 0, 1, \ldots$, let us consider two possibilities: $\lambda_k = 1$ and $0 < \lambda_k < 1$. Assume that $\lambda_k = 1$. It follows from (11) that

$$0 < -\theta(x^k) \leq \left\langle \nabla f_j(x^k), x^k - p(x^k)\right\rangle \leq \|\nabla f_j(x^k)\|\|x^k - p(x^k)\|,$$

for all $j \in \mathcal{J}$. Thus, using (17), we have $-\theta(x^k) \leq \rho\,\mathrm{diam}(\mathcal{C})$ or, equivalently, $1 \geq -\theta(x^k)/[\rho\,\mathrm{diam}(\mathcal{C})]$. Therefore, from the definition of γ in (18), we have $1 \geq -\gamma\theta(x^k)$, which corresponds to (19) with $\lambda_k = 1$. Now consider $0 < \lambda_k < 1$. Therefore, by the Armijo step size strategy, there exist $0 < \hat{\lambda}_k \leq \min\{1, \lambda_k/\omega_1\}$ and $j_k \in \mathcal{J}$ such that

$$f_{j_k}\left(x^k + \hat{\lambda}_k[p(x^k) - x^k]\right) > f_{j_k}(x^k) + \zeta\hat{\lambda}_k\theta(x^k).$$

On the other hand, applying Lemma 1 with $\lambda = \hat{\lambda}_k$, $x = x^k$, $p = p(x^k)$, and $\theta = \theta(x^k)$, we obtain

$$f_j\left(x^k + \hat{\lambda}_k[p(x^k) - x^k]\right) \leq f_j(x^k) + \theta(x^k)\hat{\lambda}_k + \frac{L}{2}\|p(x^k) - x^k\|^2\hat{\lambda}_k^2,$$

for all $j \in \mathcal{J}$. Thus, combining the two previous inequality, we conclude that

$$\zeta\hat{\lambda}_k\theta(x^k) < \theta(x^k)\hat{\lambda}_k + \frac{L}{2}\|p(x^k) - x^k\|^2\hat{\lambda}_k^2.$$

The last inequality implies that

$$-\theta(x^k)(1-\zeta) < \frac{L}{2}\|p(x^k) - x^k\|^2\hat{\lambda}_k \leq \frac{L\,\mathrm{diam}(\mathcal{C})^2\lambda_k}{2\omega_1},$$

which together with the definition of γ in (18) gives (19). This proves our claim. Now since λ_k is obtained through the Armijo step size strategy, by (19) and taking into account that $\theta(x^k) < 0$, for all $k = 0, 1, \ldots$, we obtain

$$F\left(x^k + \lambda_k[p(x^k) - x^k]\right) - F(x^*) \preceq F(x^k) - F(x^*) + \zeta\lambda_k\theta(x^k)e \preceq F(x^k) - F(x^*) - \zeta\gamma\theta(x^k)^2 e,$$

which implies that

$$\min_{j\in\mathcal{J}}\left(f_j(x^{k+1}) - f_j(x^*)\right) \leq \min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right) - \zeta\gamma\theta(x^k)^2, \qquad k = 0, 1, \ldots. \tag{20}$$

On the other hand, by the convexity of F, we have $f_j(x^*) - f_j(x^k) \geq \left\langle \nabla f_j(x^k), x^* - x^k \right\rangle$, for all $j \in \mathcal{J}$. Since $\{F(x^k)\}$ is monotone decreasing and $\lim_{k\to+\infty} F(x^k) = F(x^*)$, the last inequality together with the optimality of $p(x^k)$ in (11) yields

$$0 \geq \max_{j\in\mathcal{J}}\left(f_j(x^*) - f_j(x^k)\right) \geq \max_{j\in\mathcal{J}}\left\langle \nabla f_j(x^k), x^* - x^k \right\rangle \geq \theta(x^k),$$

which implies $0 \geq -\min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right) \geq \theta(x^k)$. Hence,

$$0 \leq \left[\min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right)\right]^2 \leq \theta(x^k)^2.$$

The combination of the last inequality with (20) yields

$$\zeta\gamma\left[\min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right)\right]^2 \leq \min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right) - \min_{j\in\mathcal{J}}\left(f_j(x^{k+1}) - f_j(x^*)\right),$$

for all, $k = 0, 1, 2, \ldots$. Finally, applying Lemma 3, with $a_k = \min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right)$ and $\Gamma = \zeta\gamma$, we obtain the desired inequality in (18), which concludes the proof. □

5 Analysis of CondG with adaptative and diminishing step sizes

In this section, we analyze Algorithm 1 with adaptative and diminishing step sizes. *Throughout the section, we assume that $F := (f_1, \ldots, f_m)$ satisfies Definition 1 with constant $L > 0$.* We begin by applying Lemma 1 to show that the sequence $\{x^k\}$ generated by Algorithm 1 with the adaptative step size satisfies an important inequality, which is a version of [3, Lemma A.2] for multicriteria optimization.

Proposition 9. *Let $\{x^k\}$ be generated by Algorithm 1 with the adaptative step size. Then,*

$$F\left(x^k + \lambda_k[p(x^k) - x^k]\right) \preceq F(x^k) - \frac{1}{2}\min\left\{-\theta(x^k), \frac{\theta(x^k)^2}{L\,\mathrm{diam}(\mathcal{C})^2}\right\} e, \qquad k = 0, 1, \ldots. \tag{21}$$

Proof. Lemma 1 with $\lambda = \lambda_k$, $x = x^k$, $p = p(x^k)$, and $\theta = \theta(x^k)$ yields

$$F\left(x^k + \lambda_k[p(x^k) - x^k]\right) \preceq F(x^k) + \left(\theta(x^k)\lambda_k + \frac{L}{2}\|p(x^k) - x^k\|^2\lambda_k^2\right) e, \qquad k = 0, 1, \ldots. \tag{22}$$

We will consider two separate cases: $\lambda_k = 1$ and $\lambda_k = -\theta(x^k)/(L\|p(x^k) - x^k\|^2)$. First, assume that $\lambda_k = 1$. By the definition of λ_k in (13), we have $L\|p(x^k) - x^k\|^2 \leq -\theta(x^k)$. Thus inequality (22) becomes

$$F\left(x^k + \lambda_k[p(x^k) - x^k]\right) \preceq F(x^k) + \frac{1}{2}\theta(x^k)e. \tag{23}$$

Now, we assume that $\lambda_k = -\theta(x^k)/(L\|p(x^k) - x^k\|^2)$. In this case, inequality (22) becomes

$$F\left(x^k + \lambda_k[p(x^k) - x^k]\right) \preceq F(x^k) - \frac{\theta(x^k)^2}{2L\|p(x^k) - x^k\|^2}e.$$

The combination of (23) with the last vector inequality yields

$$F\left(x^k + \lambda_k[p(x^k) - x^k]\right) \preceq F(x^k) - \frac{1}{2}\min\left\{-\theta(x^k), \frac{\theta(x^k)^2}{L\|p(x^k) - x^k\|^2}\right\} e, \qquad k = 0, 1, \ldots.$$

Since $\mathrm{diam}(\mathcal{C}) \geq \|p(x^k) - x^k\|$, the above inequality implies (21) and the proof is concluded. ☐

Remind that we are assuming that $\{x^k\}$ generated by the Algorithm 1 is infinite and then $\theta(x^k) < 0$, for all $k = 0, 1, \ldots$. As an application of Proposition 9, without convexity assumptions on the objectives, we establish below convergence properties of the CondG algorithm with adaptative step sizes. We point out that these results is a multicriteria version of [2, Theorem 13.9]. Let us define:

$$-\infty < f_j^{\inf} := \inf\{f_j(x) : \ x \in \mathcal{C}\}, \qquad j \in \mathcal{J}. \tag{24}$$

Corollary 10. *Assume that $\{x^k\}$ is generated by Algorithm 1 with the adaptative step size. Let $j_* \in \mathcal{J}$ be an index such that $f_{j_*}(x^0) - f_{j_*}^{\inf} := \min\left\{f_j(x^0) - f_j^{\inf} : \ j \in \mathcal{J}\right\}$. Then,*

(i) $\lim_{k\to+\infty} \theta(x^k) = 0$;

(ii) for every $N \in \mathbb{N}$, there holds

$$\min\left\{|\theta(x^k)| \ ; \ k = 0, 1, \ldots, N-1\right\} \leq \max\left\{\frac{2[f_{j_*}(x^0) - f_{j_*}^{\inf}]}{N}, \mathrm{diam}(\mathcal{C})\sqrt{\frac{2L[f_{j_*}(x^0) - f_{j_*}^{\inf}]}{N}}\right\}.$$

Proof. By Proposition 9, $\{f_{j_*}(x^k)\}$ is nonincreasing because $\theta(x^k) < 0$, for all $k = 0, 1, \ldots$. Since this sequence is also bounded from below by $f_{j_*}^{\inf}$ defined in (24), it turns out that $\{f_{j_*}(x^k)\}$ converges. Moreover, by (12), Proposition 9 also implies

$$0 < \min\left\{-\theta(x^k), \frac{\theta(x^k)^2}{L\,\mathrm{diam}(\mathcal{C})^2}\right\} \leq 2[f_{j_*}(x^k) - f_{j_*}(x^{k+1})], \qquad k = 0, 1, \ldots. \tag{25}$$

Since $\{f_{j_*}(x^k)\}$ converges, we have $\lim_{k\to+\infty}[f_{j_*}(x^k) - f_{j_*}(x^{k+1})] = 0$. Thus, taking limits in (25), we have

$$\lim_{k\to+\infty} \min\left\{-\theta(x^k), \frac{\theta(x^k)^2}{L\,\mathrm{diam}(\mathcal{C})^2}\right\} = 0,$$

which proves (i). By summing both sides of the second inequality in (25) for $k = 0, 1, \ldots, N-1$ and taking into accont that $f_{j_*}^{\inf} := \inf\{f_{j_*}(x) : \ x \in \mathcal{C}\}$, we obtain

$$\sum_{k=0}^{N-1} \min\left\{-\theta(x^k), \frac{\theta(x^k)^2}{L\,\mathrm{diam}(\mathcal{C})^2}\right\} \leq 2[f_{j_*}(x^0) - f_{j_*}^{\inf}].$$

Therefore,

$$\min\left\{\min\left\{-\theta(x^k), \frac{\theta(x^k)^2}{L\,\mathrm{diam}(\mathcal{C})^2}\right\} \ ; \ k = 0, 1, \ldots, N-1\right\} \leq \frac{2[f_{j_*}(x^0) - f_{j_*}^{\inf}]}{N}.$$

In particular, the last inequality implies that there exists $\bar{k} \in \{0, 1, \ldots, N-1\}$ such that

$$-\theta(x^{\bar{k}}) \leq \frac{2[f_{j_*}(x^0) - f_{j_*}^{\inf}]}{N}, \qquad \text{or} \qquad -\theta(x^{\bar{k}}) \leq \mathrm{diam}(\mathcal{C})\sqrt{\frac{2L[f_{j_*}(x^0) - f_{j_*}^{\inf}]}{N}},$$

which gives the statement of item (ii). □

Remark. *A direct consequence of Corollary 10 (i) and Proposition 5 (ii) and (iii) is that every limit point $\bar{x}$ of $\{x^k\}$ generated by Algorithm 1 with adaptative step sizes is a Pareto critical point. Moreover, by Proposition 9, $f_j(\bar{x}) = \inf\{f_j(x^k) : \ k = 0, 1, \ldots\}$, for all $j \in \mathcal{J}$. In addition, if F is convex on $\mathcal{C}$, by Lemma 2, then $\bar{x}$ is a weak Pareto optimal point of Problem* (1).

Following, we establish two iteration-complexity bounds for Algorithm 1 with adaptative or diminishing step sizes for the cases where F is convex on $\mathcal{C}$. We begin by proving a useful auxiliary result.

Lemma 11. *Consider $\{x^k\}$ generated by Algorithm 1 with adaptative or diminishing step sizes. Then,*

$$F\left(x^k + \lambda_k[p(x^k) - x^k]\right) \preccurlyeq F(x^k) + \left(\theta(x^k)\beta_k + \frac{L}{2}\|p(x^k) - x^k\|^2\beta_k^2\right) e, \tag{26}$$

where $\beta_k := 2/(k+2)$.

Proof. By Lemma 1 with $\lambda = \lambda_k$, $x = x^k$, $p = p(x^k)$, and $\theta = \theta(x^k)$, we obtain

$$F\left(x^k + \lambda_k[p(x^k) - x^k]\right) \preceq F(x^k) + \left(\theta(x^k)\lambda_k + \frac{L}{2}\|p(x^k) - x^k\|^2\lambda_k^2\right) e. \tag{27}$$

For the diminishing step size where $\lambda_k = \beta_k$, inequality (26) follows trivially from (27). Now, consider that λ_k is obtained using the adaptative step size strategy. By (13), we have $\theta(x^k)\lambda_k + L\|p(x^k) - x^k\|\lambda_k^2/2 \leq \theta(x^k)\beta_k + L\|p(x^k) - x^k\|\beta_k^2/2$. Combining this inequality with (27), we obtain (26). $\square$

Now we are able to present the iteration-complexity bounds.

Theorem 12. *Assume that $F := (f_1, \ldots, f_m)$ is convex on $\mathcal{C}$. Let $\{x^k\}$ be generated by Algorithm 1 with adaptative or diminishing step sizes. Assume that $\lim_{k\to+\infty} F(x^k) = F(x^*)$. Then, the following inequalities holds:*

(i) $\min_{j\in\mathcal{J}} \Big(f_j(x^k) - f_j(x^*)\Big) \leq \dfrac{2L\mathrm{diam}(\mathcal{C})^2}{k}, \qquad \forall\ k = 1, 2, \ldots;$

(ii) for every $\ell \in \mathbb{N}$, there holds

$$\min_{\ell\in\left\{\lfloor \frac{k}{2}\rfloor+2,\cdots,k\right\}} \left|\theta(x^\ell)\right| \leq \frac{8L\mathrm{diam}(\mathcal{C})^2}{k-2}, \qquad \forall\ k = 3, 4, \ldots,$$

where $\lfloor k/2\rfloor = \max\left\{n \in \mathbb{N} :\ n \leq k/2\right\}$.

Proof. By Lemma 11, we obtain

$$F\left(x^k + \lambda_k[p(x^k) - x^k]\right) - F(x^*) \preccurlyeq F(x^k) - F(x^*) + \left(\theta(x^k)\beta_k + \frac{L}{2}\|p(x^k) - x^k\|^2\beta_k^2\right)e,$$

where $\beta_k := 2/(k+2)$. Therefore, since $\|p(x^k) - x^k\| \leq \mathrm{diam}(\mathcal{C})$, for all $k = 0, 1, \ldots$, we have

$$F\left(x^k + \lambda_k[p(x^k) - x^k]\right) - F(x^*) \preccurlyeq F(x^k) - F(x^*) + \left(\theta(x^k)\beta_k + \frac{L}{2}\,\mathrm{diam}(\mathcal{C})^2\beta_k^2\right)e,$$

which implies that

$$\min_{j\in\mathcal{J}}\left(f_j(x^{k+1}) - f_j(x^*)\right) \leq \min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right) + \theta(x^k)\beta_k + \frac{L}{2}\,\mathrm{diam}(\mathcal{C})^2\beta_k^2.$$

On the other hand, since F is convex we have $f_j(x^k) - f_j(x^*) \leq -\left\langle \nabla f_j(x^k), x^* - x^k\right\rangle$, for all $j \in \mathcal{J}$. Hence, by the optimality of $p(x^k)$ in (11), it follows that

$$0 \leq \min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right) \leq -\max_{j\in\mathcal{J}}\left\langle \nabla f_j(x^k), x^* - x^k\right\rangle \leq -\theta(x^k).$$

Thus, we can applying Lemma 4 with $a_k = \min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right) \leq b_k = -\theta(x^k)$ and $A = L\mathrm{diam}(\mathcal{C})^2$ to obtain the desired results. $\square$

5.1 Iteration-complexity bound under strong convexity of constraint set

Under strong convexity assumptions of constraint set $\mathcal{C}$, we are able to improve the convergence rate of the first item of Theorem 12 for the sequence generated by Algorithm 1 with adaptative step sizes. Let us assume that $\mathcal{C} \subset \mathbb{R}^n$ is a $\sigma-$*strongly convex set* for some $\sigma > 0$, i.e., for any $x, y \in \mathcal{C}$ and $\lambda \in [0, 1]$, the following inclusion holds

$$\lambda x + (1-\lambda)y + \frac{\sigma}{2}\lambda(1-\lambda)\|x - y\|^2 z \in \mathcal{C}, \qquad \forall\ z \in \mathbb{R}^n,\ \|z\| = 1.$$

For simplicity, for each $x \in \mathcal{C}$, let us define

$$\mathcal{K}(x) := \min_{j\in\mathcal{J}}\|\nabla f_j(x)\|, \qquad \kappa := \min_{x\in\mathcal{C}}\mathcal{K}(x).$$

Theorem 13. *Assume that $\mathcal{C}$ is a $\sigma-$strongly convex set and $F := (f_1, \ldots, f_m)$ is convex on $\mathcal{C}$. Let $\{x^k\}$ be the sequence generated by Algorithm 1 with the adaptative step size. Assume that $\lim_{k\to+\infty} x^k = x^*$ and $\kappa > 0$. Then,*

$$\min_{j\in\mathcal{J}} \Big(f_j(x^{k+1}) - f_j(x^*)\Big) \le q^k \min_{j\in\mathcal{J}} \Big(f_j(x^0) - f_j(x^*)\Big), \qquad k = 0, 1, \ldots,$$

where $q := \max\Big\{1/2, 1 - \big(\sigma\kappa/8L\big)\Big\}$.

Proof. Consider $w^k \in \operatorname{argmin}_{\|w\|=1} \max_{j\in\mathcal{J}} \Big\langle w, \nabla f_j(x^k)\Big\rangle$. Since $\mathcal{C}$ is strongly convex and x^k, $p(x^k) \in \mathcal{C}$, we have

$$z^k = \frac{x^k + p(x^k)}{2} + \frac{\sigma\|x^k - p(x^k)\|}{8} w^k \in \mathcal{C}.$$

Thus, the optimality of $p(x^k)$ in (11) yields

$$\theta(x^k) = \max_{j\in\mathcal{J}} \Big\langle \nabla f_j(x^k), p(x^k) - x^k\Big\rangle \le \max_{j\in\mathcal{J}} \Big\langle \nabla f_j(x^k), z^k - x^k\Big\rangle,$$

which gives

$$\theta(x^k) \le \frac{1}{2} \max_{j\in\mathcal{J}} \Big\langle \nabla f_j(x^k), p(x^k) - x^k\Big\rangle + \frac{\sigma\|x^k - p(x^k)\|^2}{8} \max_{j\in\mathcal{J}} \Big\langle \nabla f_j(x^k), w^k\Big\rangle. \tag{28}$$

By the definition of w^k and taking into account that x^k is a nonstationary point, it follows that $\max_{j\in\mathcal{J}} \Big\langle \nabla f_j(x^k), w^k\Big\rangle \le -\mathcal{K}(x^k)$. Also, by using the optimality of $p(x^k)$ in (11) and the convexity of F, we conclude that

$$\max_{j\in\mathcal{J}} \Big\langle \nabla f_j(x^k), p(x^k) - x^k\Big\rangle \le \max_{j\in\mathcal{J}} \Big\langle \nabla f_j(x^k), x^* - x^k\Big\rangle \le \max_{j\in\mathcal{J}} \Big(f_j(x^*) - f_j(x^k)\Big),$$

which implies $\max_{j\in\mathcal{J}} \Big\langle \nabla f_j(x^k), p(x^k) - x^k\Big\rangle \le -\min_{j\in\mathcal{J}} \Big(f_j(x^k) - f_j(x^*)\Big)$. Thus, from (28) we have

$$\theta(x^k) \le -\frac{1}{2} \min_{j\in\mathcal{J}} \Big(f_j(x^k) - f_j(x^*)\Big) - \frac{\sigma\|x^k - p(x^k)\|^2}{8} \mathcal{K}(x^k). \tag{29}$$

On the other hand, by using Lemma (1) with $\lambda = \lambda_k$, $x = x^k$, $p = p(x^k)$, and $\theta = \theta(x^k)$, we obtain

$$F\Big(x^k + \lambda_k[p(x^k) - x^k]\Big) \preceq F(x^k) + \Big(\theta(x^k)\lambda_k + \frac{L}{2}\|p(x^k) - x^k\|^2\lambda_k^2\Big) e,$$

which, subtracting $F(x^*)$ on both sides of this inequality and taking the minimum in j, yields

$$\min_{j\in\mathcal{J}} \Big(f_j(x^{k+1}) - f_j(x^*)\Big) \le \min_{j\in\mathcal{J}} \Big(f_j(x^k) - f_j(x^*)\Big) + \theta(x^k)\lambda_k + \frac{L}{2}\|p(x^k) - x^k\|^2\lambda_k^2. \tag{30}$$

We will analyze two possibilities, considering initially that $\sigma\mathcal{K}(x^k)/(4L) \ge 1$. From the optimality of λ_k in (13), we have $\theta(x^k)\lambda_k + (L/2)\|p(x^k) - x^k\|^2\lambda_k^2 \le \theta(x^k) + (L/2)\|p(x^k) - x^k\|^2$. Thus, by (30), we conclude

$$\min_{j\in\mathcal{J}} \Big(f_j(x^{k+1}) - f_j(x^*)\Big) \le \min_{j\in\mathcal{J}} \Big(f_j(x^k) - f_j(x^*)\Big) + \theta(x^k) + \frac{L}{2}\|p(x^k) - x^k\|^2.$$

Combining the last inequality with (29), and considering that $\sigma\mathcal{K}(x^k)/4L \geq 1$, we have

$$\begin{aligned}\min_{j\in\mathcal{J}}\left(f_j(x^{k+1})-f_j(x^*)\right) &\leq \frac{1}{2}\min_{j\in\mathcal{J}}\left(f_j(x^k)-f_j(x^*)\right)+\frac{\|p(x^k)-x^k\|^2}{2}\left[L-\frac{\sigma\mathcal{K}(x^k)}{4}\right]\\ &\leq \frac{1}{2}\min_{j\in\mathcal{J}}\left(f_j(x^k)-f_j(x^*)\right). \qquad (31)\end{aligned}$$

Now, we consider that $\sigma\mathcal{K}(x^k)/(4L) < 1$. In this case, by the optimality of λ_k in (13), we have

$$\theta(x^k)\lambda_k+\frac{L}{2}\|p(x^k)-x^k\|^2\lambda_k^2 \leq \theta(x^k)\frac{\sigma\mathcal{K}(x^k)}{4L}+\frac{L}{2}\|p(x^k)-x^k\|^2\left[\frac{\sigma\mathcal{K}(x^k)}{4L}\right]^2.$$

Therefore, combining last inequality with (29) and (30), we obtain

$$\min_{j\in\mathcal{J}}\left(f_j(x^{k+1})-f_j(x^*)\right)\leq\left[1-\frac{\sigma\mathcal{K}(x^k)}{8L}\right]\min_{j\in\mathcal{J}}\left(f_j(x^k)-f_j(x^*)\right). \qquad (32)$$

By (31), (32), and considering that $\mathcal{K}(x^k)\geq\kappa\geq 0$, we have

$$\min_{j\in\mathcal{J}}\left(f_j(x^{k+1})-f_j(x^*)\right)\leq\max\left\{\frac{1}{2},1-\frac{\sigma\kappa}{8L}\right\}\min_{j\in\mathcal{J}}\left(f_j(x^k)-f_j(x^*)\right),$$

which implies the desired inequality and the proof is completed. □

Note that if $q < 1$, then Theorem 13 implies that $\{x^k\}$ converges to x^* at the rate of a geometric progression.

6 Numerical experiments

In this section, we present some numerical experiments to verify the applicability of the proposed conditional gradient scheme for multiobjective optimization problems. We are concerned with two objectives: the effectiveness of the method itself and its ability to generate Pareto curves properly. Each of these objectives will be addressed in the following sections. Our experiments were done using Fortran 90. The codes, as well as the formulation of each test problem considered, are freely available at `https://orizon.ime.ufg.br/`.

6.1 Comparisons with the projected steepest descent method

We begin the experiments by showing numerical comparisons between the proposed conditional gradient method with the projected steepest descent method [26]. The projected steepest descent direction $d_{\rm sd}(x)$ for F at $x\in\mathcal{C}$ is defined as

$$d_{\rm sd}(x) := p_{\rm sd}(x)-x, \qquad (33)$$

where

$$p_{\rm sd}(x) := \text{argmin}_{u\in\mathcal{C}}\max_{j\in\mathcal{J}}\left\langle\nabla f_j(x),u-x\right\rangle+\frac{1}{2}\|u-x\|^2. \qquad (34)$$

Since the minimand in (34) is proper, closed, and strongly convex, this problem has a unique minimizer. The optimal value of (34) will be denoted by $\theta_{\mathrm{sd}}(x)$, i.e.,

$$\theta_{\mathrm{sd}}(x) := \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x), d_{\mathrm{sd}}(x)\right\rangle + \frac{1}{2}\|d_{\mathrm{sd}}(x)\|^2. \tag{35}$$

For practical purposes, $p_{\mathrm{sd}}(x)$ can be computed by solving for $\tau \in \mathbb{R}$ and $u \in \mathcal{C}$

$$\begin{array}{ll} \min_{u,\tau} & \tau + \frac{1}{2}\|u - x\|^2 \\ \text{s.t.} & \left\langle \nabla f_j(x), u - x\right\rangle \le \tau, \quad j \in \mathcal{J}, \\ & u \in \mathcal{C}, \end{array} \tag{36}$$

which is a convex quadratic problem with linear inequality constraints. In connection to Preposition 5, it is possible to show that $\theta_{\mathrm{sd}}(x) \le 0$, $\theta_{\mathrm{sd}}(\cdot)$ is continuous, and $x \in \mathcal{C}$ is stationary if and only if $\theta_{\mathrm{sd}}(x) = 0$. All properties mentioned can be found in [26]. Essentially, the projected steepest descent method corresponds to Algorithm 1 with the search direction given as in (33) and using $\theta_{\mathrm{sd}}(x^k)$ at the stopping criterion.

We implemented both the conditional gradient and the projected steepest descent methods using the Armijo step size strategy with parameters $\zeta = 10^{-4}$, $\omega_1 = 0.05$, and $\omega_2 = 0.95$. We remark that the Armijo line search was coded based on quadratic polynomial interpolations of the coordinate functions. We refer the reader to [42] for line search strategies in the vector optimization setting. For computing the search directions, we used the free software Algencan [7] (an augmented Lagrangian code for general nonlinear programming) to solve problems (7) and (36) for the conditional gradient and the projected steepest descent methods, respectively. All runs were stopped at an iterate x^k declaring convergence if

$$\frac{\|x^k - x^{k-1}\|_\infty}{\|x^{k-1}\|_\infty} \le 10^{-5} \quad \text{and} \quad \left|\theta_{\mathrm{sd}}(x^k)\right| \le 5 \times \texttt{eps}^{1/2}, \tag{37}$$

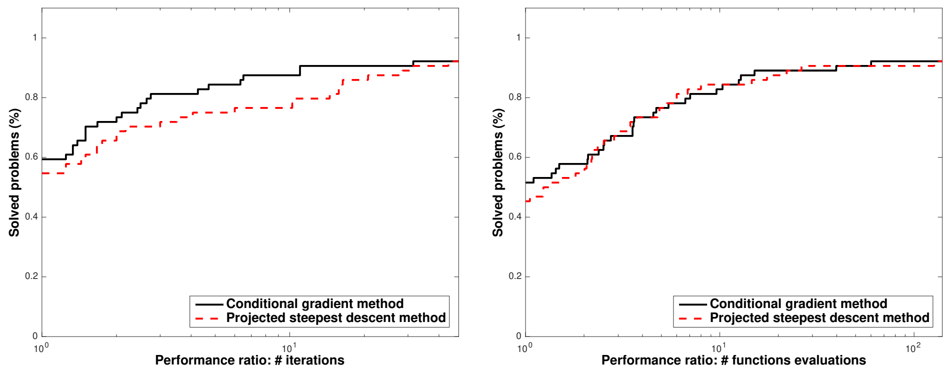
where $\texttt{eps} = 2^{-52} \approx 2.22 \times 10^{-16}$ is the machine precision. Some words about this stopping criterion are in order. First, given $x \in \mathcal{C}$, the values of $\theta(x)$ in (8) and $\theta_{\mathrm{sd}}(x)$ in (35) are different, so we prefer to use only $\theta_{\mathrm{sd}}(x)$ to standardize the stopping criteria for both methods. Second, the first criterion in (37) seeks to detect the convergence of the sequence $\{x^k\}$, while the second guarantees to stop at an *approximately* stationary point. Third, for the projected steepest descent method, we only calculate $\theta_{\mathrm{sd}}(x^k)$ when the first criterion in (37) is satisfied. We will see that the latter condition is, in general, sufficient for detecting stationary points. We also consider a stopping criterion related to failures: the maximum number of allowed iterations was set to 1000.

The set of test problems consists of 63 convex and nonconvex multiobjective problems found in the literature. In all of them, set $\mathcal{C}$ is formed by box constraints, i.e., $\mathcal{C} = \{x \in \mathbb{R}^n : l \le x \le u\}$. Table 1 shows the main characteristics of the problems. The first two columns identify the problem name and the corresponding reference (e.g., AP1 corresponds to the fist problem proposed by Ansary and Panda in [1]). Columns "n" and "m" report the number of variable and objectives, respectively. "Convex" informs whether the problem is convex or not. The last two columns show the bounds of the corresponding set $\mathcal{C}$. For each problem, we random generated a starting point beloging to $\mathcal{C}$ and run both algorithms. We compared the methods with respect to the number of iterations and the number of evaluations of the objectives. We remark that we considered each evaluation of an objective to compute the total number of functions evaluations needed for an algorithm to stop. The results in Figure 1 are shown using performance profiles [12].

Problem	Source	n	m	Convex	l^T	u^T
AP1	[1]	2	3	Y	$(-10, -10)$	$(10, 10)$
AP2	[1]	1	2	Y	-100	100
AP3	[1]	2	2	N	$(-100, -100)$	$(100, 100)$
AP4	[1]	3	3	Y	$(-10, -10, -10)$	$(10, 10, 10)$
BK1	[30]	2	2	Y	$(-5, -5)$	$(10, 10)$
DD1[a]	[10]	5	2	N	$(-20, \ldots, -20)$	$(20, \ldots, 20)$
DGO1	[30]	1	2	N	-10	13
DGO2	[30]	1	2	Y	-9	9
FA1	[30]	3	3	N	$(0, 0, 0)$	$(1, 1, 1)$
Far1	[30]	2	2	N	$(-1, -1)$	$(1, 1)$
FDS	[13]	5	3	Y	$(-2, \ldots, -2)$	$(2, \ldots, 2)$
FF1	[30]	2	2	N	$(-1, -1)$	$(1, 1)$
Hil1	[29]	2	2	N	$(0, 0)$	$(1, 1)$
IKK1	[30]	2	3	Y	$(-50, -50)$	$(50, 50)$
IM1	[30]	2	2	N	$(1, 1)$	$(4, 2)$
JOS1	[32]	100	2	Y	$(-100, \ldots, -100)$	$(100, \ldots, 100)$
JOS4	[32]	100	2	N	$(-100, \ldots, -100)$	$(100, \ldots, 100)$
KW2	[33]	2	2	N	$(-3, -3)$	$(3, 3)$
LE1	[30]	2	2	N	$(-5, -5)$	$(10, 10)$
Lov1	[40]	2	2	Y	$(-10, -10)$	$(10, 10)$
Lov2	[40]	2	2	N	$(-0.75, -0.75)$	$(0.75, 0.75)$
Lov3	[40]	2	2	N	$(-20, -20)$	$(20, 20)$
Lov4	[40]	2	2	N	$(-20, -20)$	$(20, 20)$
Lov5	[40]	3	2	N	$(-2, -2, -2)$	$(2, 2, 2)$
Lov6	[40]	6	2	N	$(0.1, -0.16, \ldots, -0.16)$	$(0.425, 0.16, \ldots, 0.16)$
LTDZ	[38]	3	3	N	$(0, 0, 0)$	$(1, 1, 1)$
MGH9[b]	[47]	3	15	N	$(-2, -2, -2)$	$(2, 2, 2)$
MGH16[b]	[47]	4	5	N	$(-25, -5, -5, -1)$	$(25, 5, 5, 1)$
MGH26[b]	[47]	4	4	N	$(-1, -1, -1 - 1)$	$(1, 1, 1, 1)$
MGH33[b]	[47]	10	10	Y	$(-1, \ldots, -1)$	$(1, \ldots, 1)$
MHHM2	[30]	2	3	Y	$(0, 0)$	$(1, 1)$
MLF1	[30]	1	2	N	0	20
MLF2	[30]	2	2	N	$(-100, -100)$	$(100, 100)$
MMR1	[44]	2	2	N	$(0.1, 0)$	$(1, 1)$
MMR3	[44]	2	2	N	$(-1, -1)$	$(1, 1)$
MMR4	[44]	3	2	N	$(0, 0, 0)$	$(4, 4, 4)$
MOP2	[30]	2	2	N	$(-4, -4)$	$(4, 4)$
MOP3	[30]	2	2	N	$(-\pi, -\pi)$	(π, π)
MOP5	[30]	2	3	N	$(-30, -30)$	$(30, 30)$
MOP6	[30]	2	2	N	$(0, 0)$	$(1, 1)$
MOP7	[30]	2	3	Y	$(-400, -400)$	$(400, 400)$
PNR	[50]	2	2	Y	$(-2, -2)$	$(2, 2)$
QV1	[30]	10	2	N	$(-5.12, \ldots, -5.12)$	$(5.12, \ldots, 5.12)$
SD	[52]	4	2	Y	$(1, \sqrt{2}, \sqrt{2}, 1)$	$(3, 3, 3, 3)$
SK1	[30]	1	2	N	-100	100
SK2	[30]	4	2	N	$(-10, -10, -10, -10)$	$(10, 10, 10, 10)$
SLCDT1	[51]	2	2	N	$(-1.5, -1.5)$	$(1.5, 1.5)$
SLCDT2	[51]	10	3	Y	$(-1, \ldots, -1)$	$(1, \ldots, 1)$
SP1	[30]	2	2	Y	$(-100, -100)$	$(100, 100)$
SSFYY2	[30]	1	2	N	-100	100
TKLY1	[30]	4	2	N	$(0.1, 0, 0, 0)$	$(1, 1, 1, 1)$
Toi4[b]	[55]	4	2	Y	$(-2, -2, -2, -2)$	$(5, 5, 5, 5)$
Toi8[b]	[55]	3	3	Y	$(-1, -1, -1, -1)$	$(1, 1, 1, 1)$
Toi9[b]	[55]	4	4	N	$(-1, -1, -1, -1)$	$(1, 1, 1, 1)$
Toi10[b]	[55]	4	3	N	$(-2, -2, -2, -2)$	$(2, 2, 2, 2)$
VU1	[30]	2	2	N	$(-3, -3)$	$(3, 3)$
VU2	[30]	2	2	Y	$(-3, -3)$	$(3, 3)$
ZDT1	[56]	30	2	Y	$(0, \ldots, 0)$	$(1, \ldots, 1)$
ZDT2	[56]	30	2	N	$(0.01, \ldots, 0.01)$	$(1, \ldots, 1)$
ZDT3	[56]	30	2	N	$(0.01, \ldots, 0.01)$	$(1, \ldots, 1)$
ZDT4	[56]	30	2	N	$(0.01, -5, \ldots, -5)$	$(1, 5, \ldots, 5)$
ZDT6	[56]	10	2	N	$(0, \ldots, 0)$	$(1, \ldots, 1)$
ZLT1	[30]	10	5	Y	$(-1000, \ldots, -1000)$	$(1000, \ldots, 1000)$

[a] This is a modified version of DD1 problem that can be found in [45].
[b] This is an adaptation of a single-objective optimization problem to the multiobjective setting that can be found in [45].

Table 1: List of test problems.



(a) (b)

Figure 1: Performance profile comparing the conditional gradient and the projected steepest descent methods using: (a) number of iterations; (b) number of function evaluations.

As can be seen, the methods are competitive. In general, the conditional gradient method required fewer iterations to find a stationary point than the projected steepest descent method. With respect to the number of functions evaluations, both methods behaved in an equivalent manner. Considering the number of iterations (resp. number of functions evaluations), the conditional gradient method was more efficient in 60.3% (resp. 55.6%) of the problems and the projected steepest descent method in 52.4% (resp. 46.0%) of the problems. Both methods had the same robustness: each of them successfully solved 59 of the 63 problems. The conditional gradient method failed to solve the problems DGO2, SK2, TKLY1, and Toi10, while the projected steepest descent method was unsuccessful for the problems JOS1, QV1, TKLY1, and Toi10.

We observe that for the conditional gradient method, except for problem MMR1, the fulfillment of the first criterion in (37) was sufficient to detect a stationary point. This means that in these cases, $\theta_{\rm sd}(x^k)$ was calculated only once. For the MMR1 problem, $\theta_{\rm sd}(x^k)$ was computed 8 times in 25 iterations required for the conditional gradient method to stop.

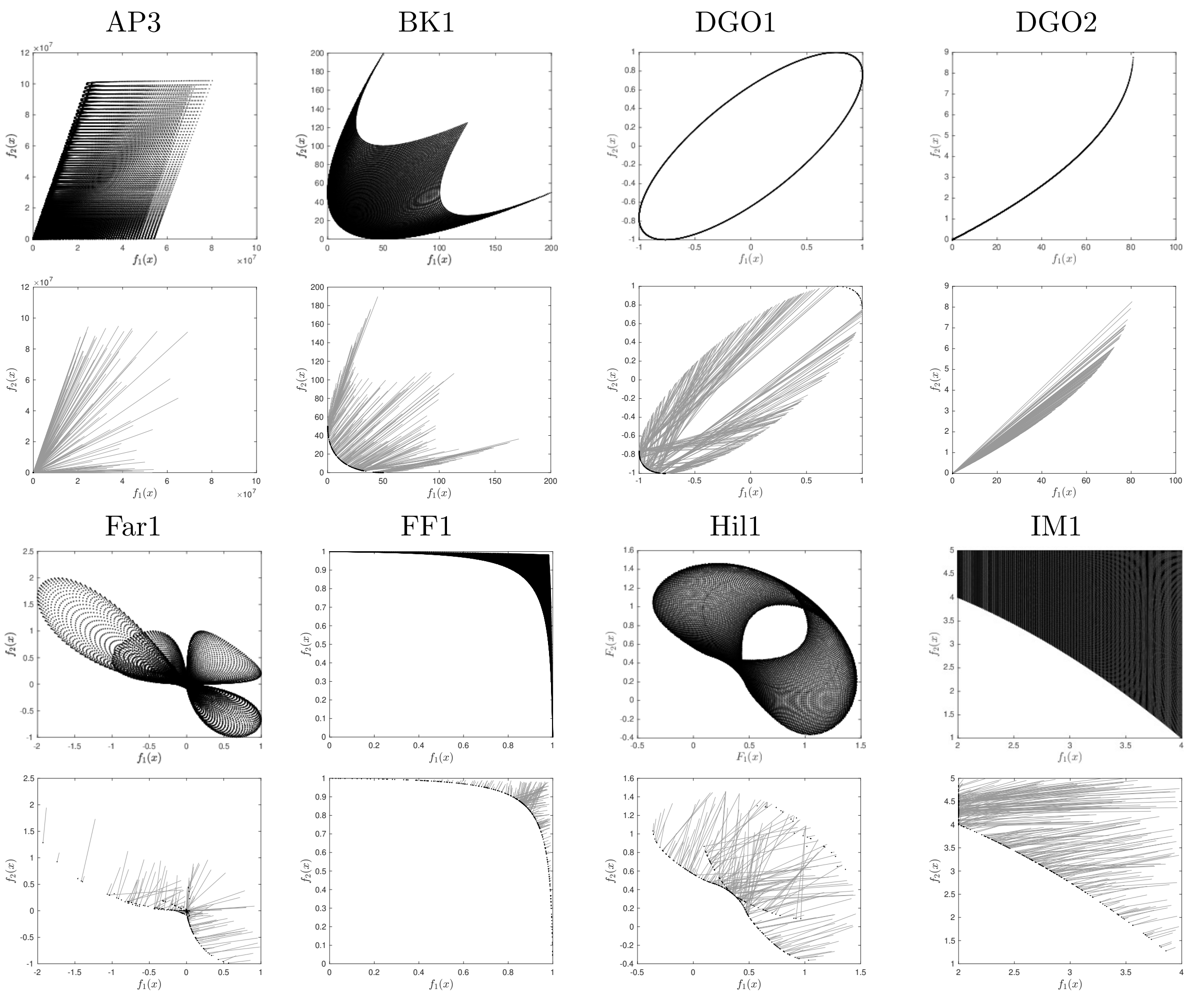
It is worth mentioning that if $\mathcal{C}$ is formed by linear constraints, then (7) is a linear programming problem while (36) is a quadratic programming problem. Taking into account that a linear problem is simpler than a quadratic problem, the present results suggest that in these cases the conditional gradient method is a promising alternative to the projected steepest descent method.

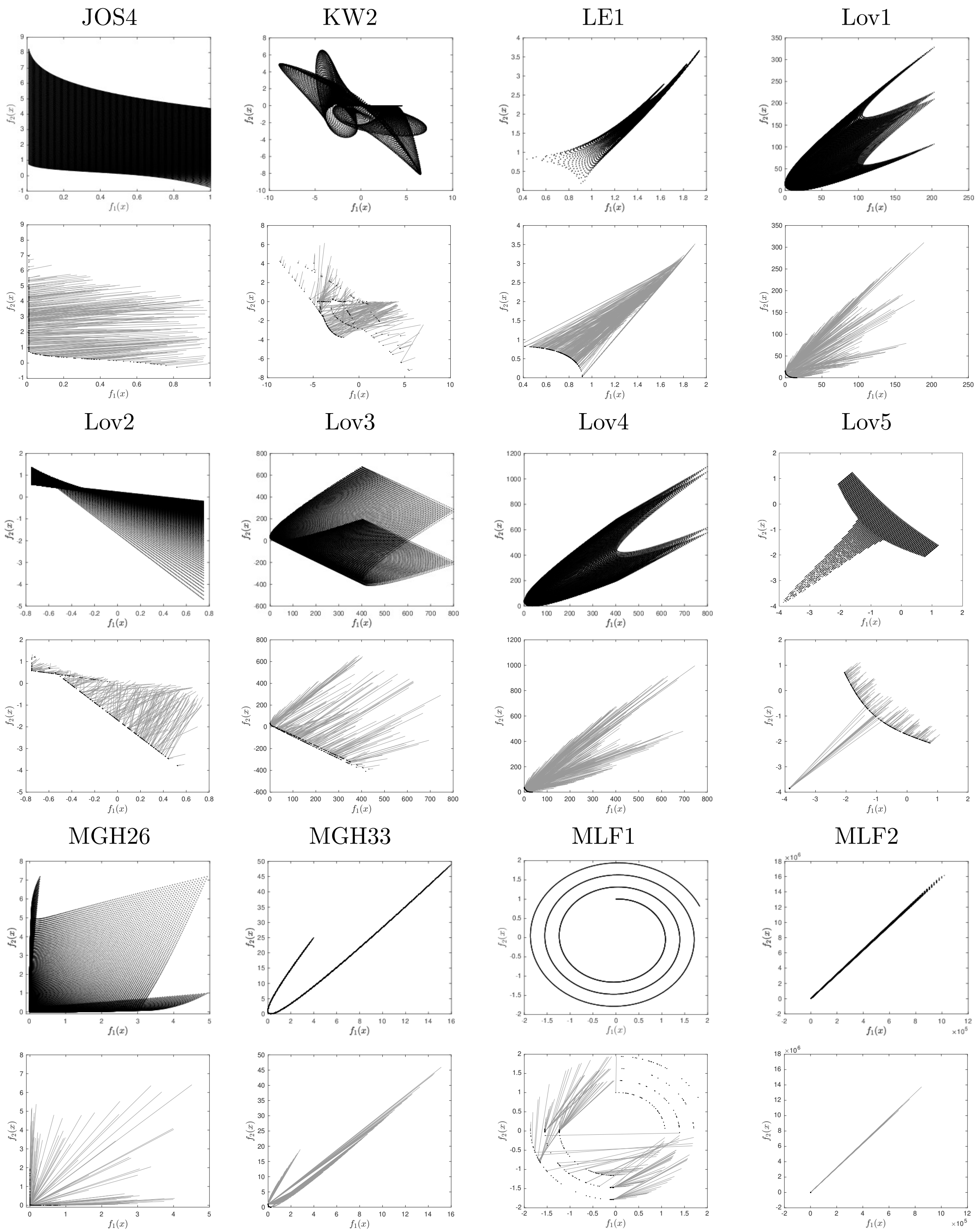
6.2 Pareto frontiers

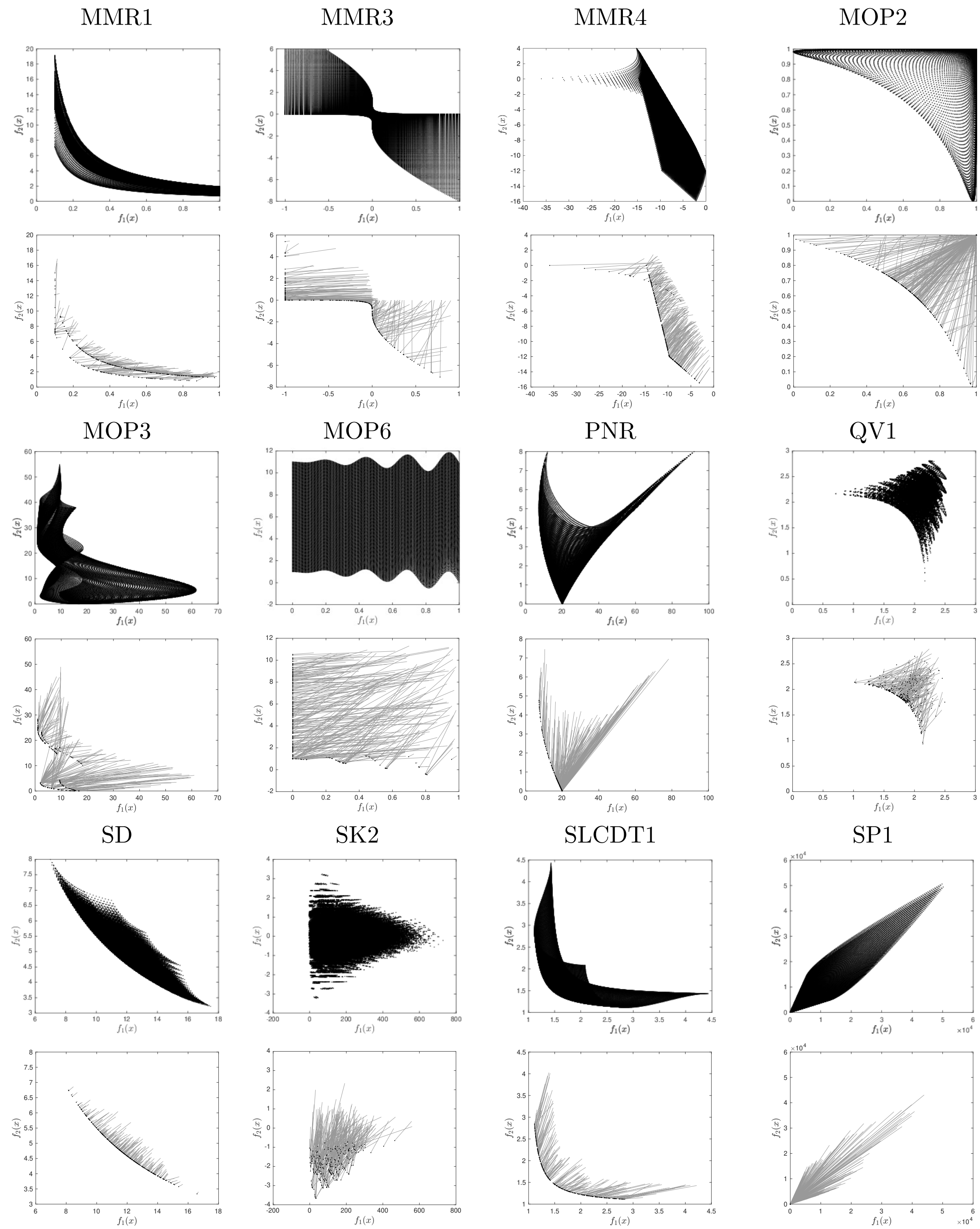
Following, we check the ability of the conditional gradient method to generate Pareto frontiers properly. In the present section, we stopped the execution of the conditional gradient method algorithm at x^k declaring convergence if

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

where $\theta(x^k)$ is given in (11) and `eps` is given as in section 6.1. We considered all bicriteria problems in Table 1 for which $n = 2$, 3 or 4. We also used the problems DGO1, DGO2, and MLF1 for which $n = 1$ and the versions with $n = 2$ and $m = 2$ of the problems JOS4, MGH26, MGH33, QV1, Toi8, Toi9, ZDT1, ZDT2, ZDT3, ZDT4, and ZDT6. The results are in Figure 2. For each problem, there as two graphics. The first ones were obtained by discretizing the corresponding boxes $\mathcal{C}$ by a fine grid and plotting all the image points. These figures provide good representations of the image spaces of F in $\mathcal{C}$ and give us a geometric notion of the Pareto frontiers. The second graphics were obtained by running for each considered problem the conditional gradient method 300 times using randomly generated starting points belonging to the corresponding sets $\mathcal{C}$. In these graphics, a full point represents a final iterate while the beginning of a straight segment represents the associated starting point. Figure 2 shows that for the chosen set of test problems, considering a reasonable number of starting points, the conditional gradient method was able to satisfactorily estimate the Pareto frontiers. We end the numerical experiments observing that, in agreement with theoretical results, the conditional gradient method can converge to *global* Pareto points, *local* (nonglobal) Pareto points (see Far1, Hil1, KW2, Lov2, Lov5, MLF1, MMR1, MMR4, MOP3, QV1, SK2, ZDT4, ZDT6) as well as to weak Pareto points (see IM1, JOS4, Lov2, MGH26, MMR3, MOP6, Toi4, Toi8, and the ZDT family).




JOS4
KW2
LE1
Lov1
Lov2
Lov3
Lov4
Lov5
MGH26
MGH33
MLF1
MLF2


MMR1
MMR3
MMR4
MOP2
MOP3
MOP6
PNR
QV1
SD
SK2
SLCDT1
SP1

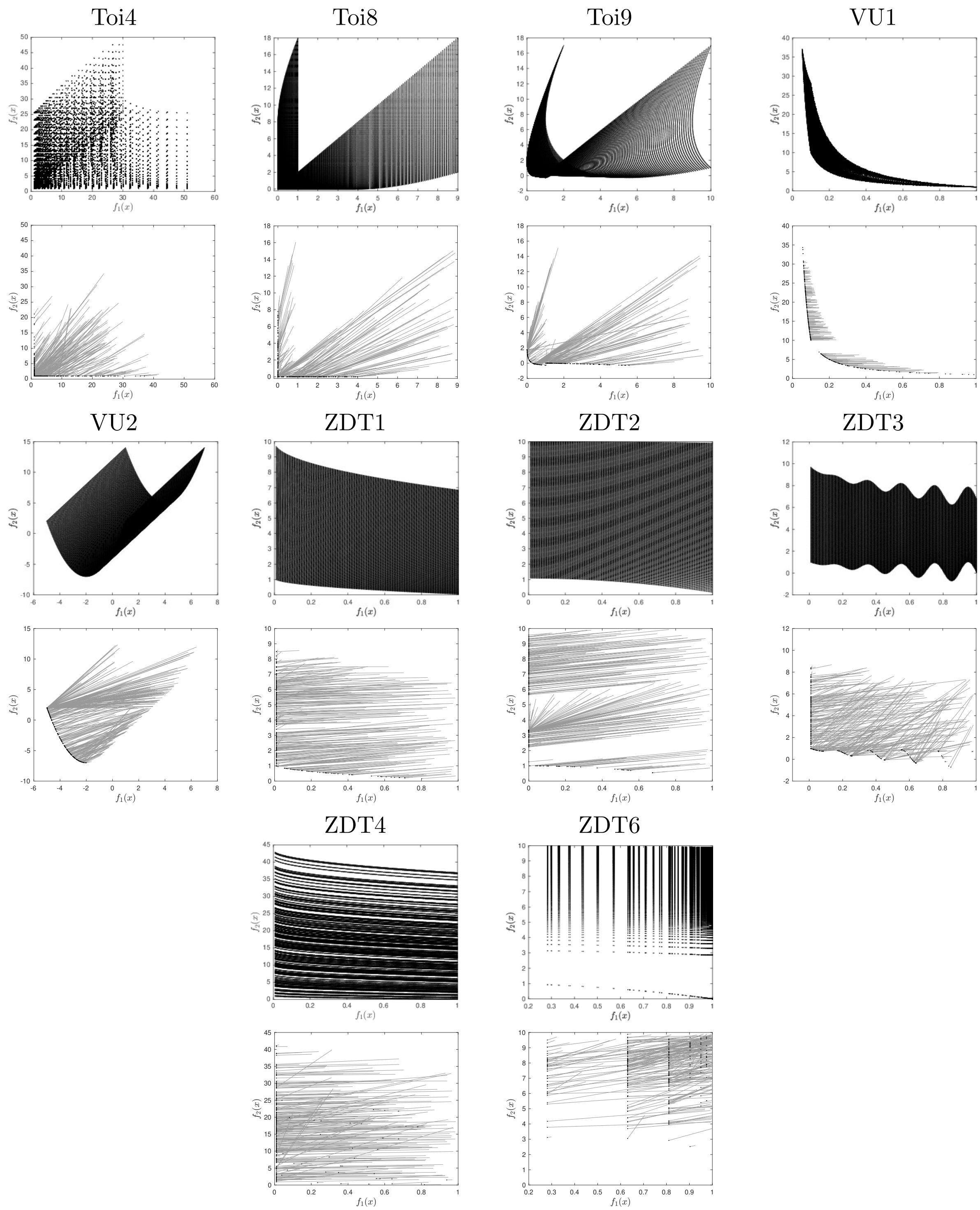


Table 2: Image sets and value spaces generated by the conditional gradient method using 300 starting points for each considered problem.

7 Conclusions

This paper extends the conditional gradient method for constrained multiobjective problems, contributing to the understanding of the connections between iteratives methods for scalar-valued and multiobjective-valued optimization. We established results on the asymptotic behavior and iteration-complexity bounds for the sequence generated by the conditional gradient method. Our analysis was carried out with and without convexity and Lipschitz assumptions on the objective functions and considering different strategies for the step sizes. The numerical experiments indicate that the conditional gradient method is competitive with the the projected steepest descent method on the chosen set of test problems. Moreover, it was able to satisfactorily estimate the Pareto frontiers of several convex and nonconvex problems.

In the scalar-valued optimization, it is well known that under strong convexity of the objective function and constraint set, the functional values of the sequence generated by the conditional gradient method converges with exponential convergence rate, see for example [39]. In particular, Theorem 13 extends this result to the multiobjective context. It would be also interesting to show that the sequence of functional values $\{f(x^k)\}$, where $\{x^k\}$ is generated by the multiobjective conditional gradient method, converges with the rate of $1/k^2$, as in the scalar context, see [22].

References

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

[2] A. Beck. *Introduction to nonlinear optimization*, volume 19 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2014. Theory, algorithms, and applications with MATLAB.

[3] A. Beck and M. Teboulle. A conditional gradient method with linear rate of convergence for solving convex linear systems. *Math. Methods Oper. Res.*, 59(2):235–247, 2004.

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

[5] J. Y. Bello Cruz and G. Bouza Allende. A steepest descent-like method for variable order vector optimization problems. *J. Optim. Theory Appl.*, 162(2):371–391, 2014.

[6] G. C. Bento, J. X. Cruz Neto, G. López, A. Soubeyran, and J. C. O. Souza. The proximal point method for locally Lipschitz functions in multiobjective optimization with application to the compromise problem. *SIAM J. Optim.*, 28(2):1104–1120, 2018.

[7] E. G. Birgin and J. M. Martínez. *Practical augmented Lagrangian methods for constrained optimization*. SIAM, 2014.

[8] N. Boyd, G. Schiebinger, and B. Recht. The alternating descent conditional gradient method for sparse inverse problems. *SIAM J. Optim.*, 27(2):616–639, 2017.

[9] G. A. Carrizo, P. A. Lotito, and M. C. Maciel. Trust region globalization strategy for the nonconvex unconstrained multiobjective optimization problem. *Math. Program.*, 159(1-2, Ser. A):339–369, 2016.

[10] 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.

[11] J. E. Dennis, Jr. and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*, volume 16 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996. Corrected reprint of the 1983 original.

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

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

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

[15] J. Fliege and A. I. F. Vaz. A method for constrained multiobjective optimization based on SQP techniques. *SIAM J. Optim.*, 26(4):2091–2119, 2016.

[16] J. Fliege, A. I. F. Vaz, and L. N. Vicente. Complexity of gradient descent for multiobjective optimization. *Optimization Methods and Software*, 0:1–11, 2018.

[17] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Nav. Res. Log.*, pages 95–110, 1956.

[18] R. M. Freund, P. Grigas, and R. Mazumder. An extended Frank-Wolfe method with "in-face" directions, and its application to low-rank matrix completion. *SIAM J. Optim.*, 27(1):319–346, 2017.

[19] 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.

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

[21] E. H. Fukuda and L. M. Graña Drummond. A survey on multiobjective descent methods. *Pesquisa Operacional*, 34:585–620, 2014.

[22] D. Garber and E. Hazan. Faster rates for the frank-wolfe method over strongly-convex sets. *32nd International Conference on Machine Learning, ICML 2015*, pages 1–12, 2015.

[23] A. M. Geoffrion. Proper efficiency and the theory of vector maximization. *Journal of mathematical analysis and applications*, 22(3):618–630, 1968.

[24] S. Ghadimi. Conditional gradient type methods for composite nonlinear and stochastic optimization. *Math. Program.*, 173(1-2, Ser. A):431–464, 2019.

[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.*, 2019.

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

[28] Z. Harchaoui, A. Juditsky, and A. Nemirovski. Conditional gradient algorithms for norm-regularized smooth. convex optimization. *Math. Program.*, 152(1-2, Ser. A):75–112, 2015.

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

[30] 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(5):477–506, 2006.

[31] M. Jaggi. Revisiting frank-wolfe: Projection-free sparse convex optimization. *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13:I–427–I–435, 2013.

[32] 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.

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

[34] I. V. Konnov. Simplified versions of the conditional gradient method. *Optimization*, 67(12):2275–2290, 2018.

[35] S. Lacoste-Julien and M. Jaggi. On the global linear convergence of frank-wolfe optimization variants. *arXiv e-prints*, arXiv:1511.05932, 2015.

[36] G. Lan. The complexity of large-scale convex programming under a linear optimization oracle. *arXiv e-prints*, page arXiv:1309.5550, 2013.

[37] G. Lan and Y. Zhou. Conditional gradient sliding for convex optimization. *SIAM J. Optim.*, 26(2):1379–1409, 2016.

[38] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler. Combining convergence and diversity in evolutionary multiobjective optimization. *Evolutionary computation*, 10(3):263–282, 2002.

[39] E. Levitin and B. Polyak. Constrained minimization methods. *USSR Computational Mathematics and Mathematical Physics*, 6(5):1–50, 1966.

[40] A. Lovison. Singular continuation: Generating piecewise linear approximations to pareto sets via global analysis. *SIAM Journal on Optimization*, 21(2):463–490, 2011.

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

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

[43] R. Luss and M. Teboulle. Conditional gradient algorithms for rank-one matrix approximations with a sparsity constraint. *SIAM Rev.*, 55(1):65–98, 2013.

[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(3):662—-682, 2008.

[45] K. Mita, E. H. Fukuda, and N. Yamashita. Nonmonotone line searches for unconstrained multiobjective optimization problems. *Journal of Global Optimization*, 75(1):63–90, 2019.

[46] O. Montonen, N. Karmitsa, and M. M. Mäkelä. Multiple subgradient descent bundle method for convex nonsmooth multiobjective optimization. *Optimization*, 67(1):139–158, 2018.

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

[48] V. Morovati, L. Pourkarimi, and H. Basirzadeh. Barzilai and Borwein's method for multiobjective optimization problems. *Numer. Algorithms*, 72(3):539–604, 2016.

[49] B. T. Polyak. *Introduction to Optimization.* Translations Series in Mathematics and Engineering. Optimization Software, New York, 1987.

[50] 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.

[51] 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.

[52] W. Stadler and J. Dauer. Multicriteria optimization in engineering: A tutorial and survey. *Progress in Astronautics and Aeronautics*, 150:209–209, 1993.

[53] M. Tabatabaei, A. Lovison, M. Tan, M. Hartikainen, and K. Miettinen. ANOVA-MOP: ANOVA decomposition for multiobjective optimization. *SIAM J. Optim.*, 28(4):3260–3289, 2018.

[54] J. Thomann and G. Eichfelder. A trust-region algorithm for heterogeneous multiobjective optimization. *SIAM J. Optim.*, 29(2):1017–1047, 2019.

[55] 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.

[56] E. Zitzler, K. Deb, and L. Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8(2):173–195, 2000.