A generalized conditional gradient method for multiobjective composite optimization problems

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

September 6, 2023

Abstract: This article deals with multiobjective composite optimization problems that consist of simultaneously minimizing several objective functions, each of which is composed of a combination of smooth and non-smooth functions. To tackle these problems, we propose a generalized version of the conditional gradient method, also known as Frank-Wolfe method. The method is analyzed with three step size strategies, including Armijo-type, adaptive, and diminishing step sizes. We establish asymptotic convergence properties and iteration-complexity bounds, with and without convexity assumptions on the objective functions. Numerical experiments illustrating the practical behavior of the methods are presented.

Keywords: Conditional gradient method; Frank-Wolfe method; multiobjective optimization; Pareto optimality; constrained optimization problem.

1 Introduction

Multiobjective optimization problems typically involve the simultaneous minimization of multiple and conflicting objectives. A solution to the problem leads to a set of alternatives with different trade-offs between the objectives. In fact, in this scenario, we use the concept of *Pareto optimality* to caracterize a solution. In summary, a point is called *Pareto optimal* if, with respect to this point, none of the objective functions can be improved without degrading another. One strategy for computing Pareto points that has become very popular consists of extending methods for scalar-valued optimization to vector-value optimization, rather than using scalarization approaches [26]. To the best of our knowledge, this strategy was coined in the work [19] that proposed the steepest descent methods for unconstrained 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 [7, 21, 24, 25, 28, 29, 43]. 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 [1, 9, 13, 14, 18, 20, 27, 40, 48, 52, 58, 60].

In the present paper, we consider *multiobjective composite optimization problems*, where the objective function $F : \mathbb{R}^n \to (\mathbb{R} \cup \{+\infty\})^m$, given by $F(x) := (f_1(x), \ldots, f_m(x))$, has the following special separable structure:

$$f_j(x) := g_j(x) + h_j(x), \quad \forall j = 1, \ldots, m,$$

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

where, for each $j = 1, \dots, m$, $g_j : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is proper, convex, and lower semicontinuous, and $h_j : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. We define $G : \mathbb{R}^n \to (\mathbb{R} \cup \{+\infty\})^m$ by $G(x) := (g_1(x), \dots, g_m(x))$ and $H : \mathbb{R}^n \to \mathbb{R}^m$ by $H(x) := (h_1(x), \dots, h_m(x))$, and denote this problem by

$$\min_{x \in \mathbb{R}^n} F(x) := G(x) + H(x). \tag{1.1}$$

We also assume that the domain of G given by $\{x \in \mathbb{R}^n \mid g_j(x) < +\infty,\ j = 1, 2, \dots, m\}$ is a convex and compact set. An important instance of (1.1) is obtained when G is the indicator function (in a vector sense) of a given set $\mathcal{C} \subset \mathbb{R}^n$, i.e., for all $j = 1, \dots, m$, $g_j(x) = 0$ for all $x \in \mathcal{C}$ and $g_j(x) = +\infty$ otherwise. In this case, (1.1) merges into the following constrained multiobjective optimization problem

$$\min_{x \in \mathcal{C}} H(x). \tag{1.2}$$

Furthermore, as discussed in [53], the separable structure in (1.1) can be used to model robust multiobjective optimization problems, which are problems that include uncertain parameters and the optimization process is considered under the worst scenario.

As far as we know, [11] was the first paper to deal with problem (1.1), where a forward–backward proximal point type algorithm was studied. In [53], a proximal gradient method to solve problem (1.1) was proposed, see also [8, 56]. We should also mention that [55, 57] presented accelerated versions of the proximal gradient method to solve (1.1) with both functions G and H convex. More recently, some Newton-type approaches were considered in [2, 45]. It is worth mentioning that a version of the *conditional gradient method* also known as *Frank-Wolfe algorithm*, see [23, 38], to solve (1.2) was proposed and analyzed in [3]. However, a generalized version of this method to solve (1.1) has not yet been considered.

In this paper, a multiobjective version of the scalar generalized conditional gradient method [4, 12, 49] to solve problem (1.1) is proposed. The method is analyzed with three step size strategies, including Armijo-type, adaptive, and diminishing step sizes. Asymptotic convergence properties and iteration-complexity bounds with and without convexity assumptions on the objective functions are established, which are similar to the corresponding scalar case, see [4]. Numerical experiments on some robust multiobjective optimization problems illustrating the practical behavior of the method are presented, and comparisons with the proximal gradient method [53] are discussed.

Subsequently, we provide a comprehensive summary of the obtained results by considering the different step sizes strategies.

- Armijo-type step sizes: We prove that every limit point of the sequence generated by the algorithm is a *Pareto critical point* of (1.1). Additionally, using Lipschitz assumptions, we show that the proposed method achieves a convergence rate of $\mathcal{O}(1/\sqrt{k})$ for a criticality measure. In convex cases, a convergence rate of $\mathcal{O}(1/k)$ is attained concerning the objective function values.

- Adaptive step sizes: Under Lipschitz assumptions, the proposed method also has a convergence rate $\mathcal{O}(1/\sqrt{k})$ for the criticality measure. In convex cases, the convergence rate is improved to $\mathcal{O}(1/k)$ for both the criticality measure and the objective function values.

- Diminishing step sizes: Under Lipschitz and convex assumptions, we show that the proposed algorithm finds a *weakly Pareto optimal point* of (1.1). Using an additional suitable assumption, we obtain a convergence rate of $\mathcal{O}(1/k)$ for both the criticality measure and the objective function values.

We observe that all theoretical results obtained in [3] for the conditional gradient method applied to the particular case (1.2) were fully replicated here. Furthermore, the rates $\mathcal{O}(1/\sqrt{k})$ and $\mathcal{O}(1/k)$ for non-convex and convex problems, respectively, are similar to those obtained for the proximal gradient method in [56].

The organization of this paper is as follows. In Section 2, some notations, definitions, and auxiliary results used throughout of the paper are presented. Section 3 presents the assumptions on the considered multiobjective composite optimization problem need to our analysis. Moreover, we introduce the gap function associated to problem (1.1) and study its main properties. In Section 4, we introduce a generalization of the conditional gradient method for solving problem (1.1). We will also study asymptotic convergence properties and iteration-complexity bounds for the generated sequence by the proposed method. Numerical experiments are presented in Section 5. Finally, some conclusions are given in Section 6.

2 Preliminaries

In this section, we present some notations, definitions, and results used throughout the paper. We denote $\mathbb{N} = \{0, 1, 2, \dots\}$ and $\mathbb{N}^* = \{1, 2, 3, \dots\}$. $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{R}_{++}$ are the set of real numbers, the set of nonnegative real numbers, and the set of positive real numbers, respectively. We define $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$. Likewise, $\overline{\mathbb{R}}^m := (\mathbb{R} \cup \{+\infty\})^m$. Let $\mathcal{J} := \{1, \dots, m\}$, $\mathbb{R}^m_+ := \{u \in \mathbb{R}^m \mid u_j \geq 0, \forall j \in \mathcal{J}\}$, and $\mathbb{R}^m_{++} = \{u \in \mathbb{R}^m \mid u_j > 0, \forall j \in \mathcal{J}\}$. 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 $v - u \in \mathbb{R}^m_{++}$. The symbol $\langle \cdot, \cdot \rangle$ is the usual inner product in $\mathbb{R}^n$ and $\|\cdot\|$ denotes the Euclidean norm. Let $\mathcal{C} \subset \mathbb{R}^n$ be a convex set. If $\mathcal{C}$ is compact, its *diameter* is the finite number $\mathrm{diam}(\mathcal{C}) := \max\big\{\|x - y\| \mid \forall x, y \in \mathcal{C}\big\}$. If $\mathbb{K} = \{k_1, k_2, \dots\} \subseteq \mathbb{N}$, with $k_j < k_{j+1}$ for all $j \in \mathbb{N}$, then we denote $\mathbb{K} \underset{\infty}{\subset} \mathbb{N}$. The notation $\sigma(t) := o(t)$ for $t \in \mathbb{R}/\{0\}$ means that $\lim_{t \to 0} \sigma(t)/t = 0$.

The effective domain of $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is defined as $\mathrm{dom}(\psi) := \big\{x \in \mathbb{R}^n \mid \psi(x) < +\infty\big\}$. The function $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be *Lipschitz continuous* with constants $L > 0$ on $\mathcal{C} \subset \mathrm{dom}(\psi)$ whenever $|\psi(x) - \psi(y)| \leq L\|x - y\|$, for all $x, y \in \mathcal{C}$. Let $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function. The *directional derivative* of ψ at $x \in \mathrm{dom}(\psi)$ in the direction $d \in \mathbb{R}^n$ is given by $\psi'(x; d) := \lim_{\alpha \to 0^+}(\psi(x + \alpha d) - \psi(x))/\alpha$. When ψ is differentiable at $x \in \mathrm{int}(\mathrm{dom}(\psi))$, we can show that $\psi'(x; d) = \big\langle \nabla\psi(x), d\big\rangle$. The next lemma is a well-known result in convex analysis whose proof can be found in [10, Section 4.1].

Lemma 1. *Let $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function. Then, the function $\lambda \mapsto (\psi(x + \lambda d) - \psi(x))/\lambda$, is non-decreasing in $(0, +\infty)$. In particular, for all $\lambda \in (0, 1]$, we have $(\psi(x + \lambda d) - \psi(x))/\lambda \leq \psi(x + d) - \psi(x)$. Consequently, $\psi'(x; d) \leq \psi(x + d) - \psi(x)$.*

Definition 1. *A function $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is upper semicontinuous at $x \in \mathbb{R}^n$ if, for any sequence $(x^k)_{k \in \mathbb{N}}$ converging to x, $\limsup_{k \to \infty} \psi(x^k) \leq \psi(x)$. Likewise, ψ is said to be lower semicontinuous at $x \in \mathbb{R}^n$ whenever $-\psi$ is upper semicontinuous at $x \in \mathbb{R}^n$ or, equivalently, if $\liminf_{k \to \infty} \psi(x^k) \geq \psi(x)$. We say that ψ is upper semicontinuous (resp. lower semicontinuous) if it is upper (resp. lower) semicontinuous at every point of its domain.*

Let $\Psi : \mathbb{R}^n \to \overline{\mathbb{R}}^m$ be a vector-valued function with $\Psi(x) := (\psi_1(x), \dots, \psi_m(x))$ and consider the problem

$$\min\nolimits_{x \in \mathbb{R}^n} \Psi(x). \tag{2.1}$$

The *effective domain* of Ψ is denoted by $\mathrm{dom}(\Psi) := \big\{x \in \mathbb{R}^n \mid \psi_j(x) < +\infty,\ \forall j \in \mathcal{J}\big\}$. A point $x^* \in \mathbb{R}^n$ is called a *Pareto optimal point* of (2.1) if there exists no other $x \in \mathbb{R}^n$ such that

$F(x) \preceq F(x^*)$ and $F(x) \neq F(x^*)$. In turn, $x^* \in \mathbb{R}^n$ is called a *weakly Pareto optimal point* of (2.1), if there exists no other $x \in \mathbb{R}^n$ such that $F(x) \prec F(x^*)$. A *necessary optimality condition* for problem (2.1) at a point $\bar{x} \in \mathbb{R}^n$ is given by

$$\max_{j\in\mathcal{J}} \psi_j'(\bar{x}, d) \geq 0, \quad \forall\, d \in \mathbb{R}^n. \tag{2.2}$$

A point $\bar{x} \in \mathrm{dom}(\Psi)$ satisfying (2.2) is called a *Pareto critical point* or a *stationary point* of problem (2.1). The function Ψ is said to be *convex* on $\mathcal{C}$ if $\Psi\left(\lambda x + (1-\lambda)y\right) \preceq \lambda\Psi(x) + (1-\lambda)\psi(y)$, for all $x, y \in \mathcal{C}$, and all $\lambda \in [0,1]$, or equivalently, if each component ψ_j of Ψ, $j \in \mathcal{J}$, is a convex function on $\mathcal{C}$. We recall that for a differentiable convex function Ψ on $\mathcal{C}$ we have $\psi_j(y) - \psi_j(x) \geq \left\langle \nabla\psi_j(x),\ y - x\right\rangle$, for all $x, y \in \mathcal{C}$ with $x \in \mathrm{int}(\mathrm{dom}(\Psi))$, for all $j \in \mathcal{J}$. Next lemma shows that, in the convex case, the concepts of stationarity and weak Pareto optimality are equivalent, see [28].

Lemma 2. *If Ψ is convex and x^* is a Pareto critical point, then x^* is a weakly Pareto optimal point of problem* (1.1)*.*

The next two lemma will be important for the convergence rate results.

Lemma 3. *[46, Lemma 6] Let $(a_k)_{k\in\mathbb{N}}$ 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 $k = 1, \ldots, \ell$, then $a_\ell \leq a_0/(1 + \ell\Gamma a_0) < 1/(\Gamma\ell)$.*

Lemma 4. *[5, Lemma 13.13] Let $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ be nonnegative sequences of real numbers satisfying $a_{k+1} \leq a_k - b_k\beta_k + (A/2)\beta_k^2$ for all $k \in \mathbb{N}$, where $\beta_k = 2/(k+2)$ and A is a positive number. Suppose that $a_k \leq b_k$ for all k. Then*

(i) $a_k \leq (2A)/k$ for all $k \in \mathbb{N}^*$;

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

3 The multiobjetive composite optimization problem

Throughout our presentation, we assume that $F := (f_1, \ldots, f_m)$, where $f_j := g_j + h_j$ for all $j \in \mathcal{J} := \{1, 2, \ldots, m\}$, satisfies the following three conditions:

(A1) The function h_j is continuously differentiable, for all $j \in \mathcal{J}$;

(A2) The function g_j is proper, convex, and lower semicontinuous, for all $j \in \mathcal{J}$;

(A3) $\mathrm{dom}(G) := \left\{x \in \mathbb{R}^n \mid g_j(x) < +\infty,\ j = 1, 2, \ldots, m\right\}$ is convex and compact.

Since we are assuming that $\mathrm{dom}(G)$ is compact, for future reference we take $\Omega > 0$ satisfying

$$\Omega \geq \mathrm{diam}\left(\mathrm{dom}(G)\right). \tag{3.1}$$

We also consider the following three additional assumptions, which will be considered only when explicitly stated.

(A4) The function g_j is Lipschitz continuous with constant $Lg_j > 0$ in $\mathrm{dom}(g_j)$, for all $j \in \mathcal{J}$, and

$$L_G := \max\left\{Lg_j \mid j \in \mathcal{J}\right\} > 0;$$

(A5) The gradient ∇h_j is Lipschitz continuous with constants $L_j > 0$, for all $j \in \mathcal{J}$, and

$$L := \max\{L_j \mid j \in \mathcal{J}\};$$

(A6) The function h_j, for all $j \in \mathcal{J}$, is convex.

Before presenting the method to solve problem (1.1), we first need to study a gap function associated with this problem, which will play an important role in this work. This study will be made in next section.

3.1 The gap function

This section is devoted to study the *gap function* $\theta : \mathrm{dom}(G) \to \mathbb{R}$ associated to problem (1.1), defined by

$$\theta(x) := \min_{u\in\mathbb{R}^n} \max_{j\in\mathcal{J}} \Big(g_j(u) - g_j(x) + \big\langle \nabla h_j(x), u - x\big\rangle\Big). \tag{3.2}$$

As we prove below, the gap function $\theta(\cdot)$ will serve as a stopping criterion for the algorithm presented in the next section. We observe that, if the components of function G are the indicator function of a set $\mathcal{C}$, the gap function $\theta(\cdot)$ in (3.2) becomes the one presented in [3]. We refer the reader to [54] for other merit (gap) functions in the multiobjective optimization setting.

Clearly, for each $x \in \mathrm{dom}(G)$, the gap function $\theta(x)$ is the optimum value of the optimization problem

$$\min_{u\in\mathbb{R}^n} \max_{j\in\mathcal{J}} \Big(g_j(u) - g_j(x) + \big\langle \nabla h_j(x), u - x\big\rangle\Big). \tag{3.3}$$

4 Thus, we use the notation $p(x) \in \mathrm{dom}(G)$ when referring to a solution of problem (3.3), i.e.,

$$p(x) \in \arg\min_{u\in\mathbb{R}^n} \max_{j\in\mathcal{J}} \Big(g_j(u) - g_j(x) + \big\langle \nabla h_j(x), u - x\big\rangle\Big). \tag{3.4}$$

Therefore, combining (3.2) and (3.4), we conclude that

$$\theta(x) = \max_{j\in\mathcal{J}} \Big(g_j(p(x)) - g_j(x) + \big\langle \nabla h_j(x), p(x) - x\big\rangle\Big), \quad \forall\, x \in \mathrm{dom}(G). \tag{3.5}$$

To simplify the notations, for each $x \in \mathrm{dom}(G)$ and $p(x)$ as in (3.4), we set

$$d(x) := p(x) - x.$$

In the following lemma, we show that $\theta(\cdot)$ can in fact be seen as a gap function for problem (1.1).

Lemma 5. *Let $\theta : \mathrm{dom}(G) \to \mathbb{R}$ be defined as in* (3.2). *Then*

(i) $\theta(x) \le 0$, *for all* $x \in \mathrm{dom}(G)$*;*

(ii) $\theta(x) = 0$ *if, and only if,* x *is a Pareto critical point of problem* (1.1)*;*

(iii) $\theta(x)$ *is upper semicontinuous.*

Proof. Consider (i) and let $x \in \mathrm{dom}(G)$. The definition of $\theta(\cdot)$ in (3.2) implies

$$\theta(x) \le \max_{j\in\mathcal{J}} \Big(g_j(u) - g_j(x) + \big\langle \nabla h_j(x), u - x\big\rangle\Big), \quad \forall u \in \mathbb{R}^n. \tag{3.6}$$

Thus, letting $u = x$ in the previous inequality, we conclude that $\theta(x) \le 0$, which proves (i). To prove item (ii), we first assume that x is a Pareto critical point of problem (1.1). Therefore, by (2.2), we obtain

$$\max_{j\in\mathcal{J}} f_j'(x;d) \ge 0, \quad \forall d \in \mathbb{R}^n. \tag{3.7}$$

Let $d \in \mathbb{R}^n$ be arbitrary. Using (A1) and (A2), we have $f_j'(x;d) = g_j'(x;d) + \langle \nabla h_j(x), d\rangle$. Thus, it follows from (3.7) that $\max_{j\in\mathcal{J}}\{g_j'(x;d)+\langle \nabla h_j(x), d\rangle\} \ge 0$, for all $d \in \mathbb{R}^n$. Hence, by Lemma 1, we conclude that $\max_{j\in\mathcal{J}} \Big\{g_j(x+d) - g_j(x) + \langle \nabla h_j(x), d\rangle\Big\} \ge 0$. In particular, letting $d = p(x) - x$, we have

$$\max_{j\in\mathcal{J}} \Big(g_j(p(x)) - g_j(x) + \langle \nabla h_j(x), p(x) - x\rangle\Big) \ge 0.$$

Thus, using (3.5), we conclude that $\theta(x) \ge 0$ which, together with item *(i)*, gives $\theta(x) = 0$. Reciprocally, now we assume that $\theta(x) = 0$. Thus, as in (3.6), we obtain

$$\max_{j\in\mathcal{J}} \Big\{g_j(u) - g_j(x) + \langle \nabla h_j(x), u - x\rangle\Big\} \ge 0, \quad \forall u \in \mathbb{R}^n.$$

In particular, letting $u = x + \alpha d$, for $\alpha > 0$ and $d \in \mathbb{R}^n$, we conclude that

$$\max_{j\in\mathcal{J}} \Big(\frac{g_j(x+\alpha d) - g_j(x)}{\alpha} + \langle \nabla h_j(x), d\rangle\Big) \ge 0, \quad \forall \alpha > 0,\ \forall d \in \mathbb{R}^n.$$

Since the maximum function is continuous and g_j has directional derivative at $x \in \mathrm{dom}(G)$, we can take limit as α goes to 0 in the last inequality to conclude that $\max_{j\in\mathcal{J}}\{g_j'(x,d)+\langle \nabla h_j(x), d\rangle\} \ge 0$, for all $d \in \mathbb{R}^n$. Therefore, (2.2) holds and thus x is a Pareto critical point of problem (1.1). We proceed to prove item (iii). Let $x \in \mathrm{dom}(G)$ and consider a sequence $(x^k)_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} x^k = x$. Since $p(x) \in \mathrm{dom}(G)$, by (3.2), we have

$$\theta(x^k) \le \max_{j\in\mathcal{J}} \big(g_j(p(x)) - g_j(x^k) + \langle \nabla h_j(x^k), p(x) - x^k\rangle\big).$$

Using the continuity of the maximum function and taking the upper limit in the last inequality, we have

$$\limsup_{k\to\infty} \theta(x^k) \le \max_{j\in\mathcal{J}} \big(g_j(p(x)) + \limsup_{k\to\infty} \big(-g_j(x^k)\big) + \langle \nabla h_j(x), p(x) - x\rangle\big). \tag{3.8}$$

On the other hand, considering that g_j is lower semicontinuous in its effective domain, we obtain $\limsup_{k\to\infty} \Big[-g_j(x^k)\Big] \le -g_j(x)$. Therefore, combining this inequality with (3.8) and (3.5), we have $\limsup_{k\to\infty} \theta(x^k) \le \theta(x)$, which concludes the proof. □

Hereafter, we denote:

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

When there is no confusion, we will also use letter e to denote the column vector of ones with an alternative dimension. In the following lemma, we present the counterpart of [3, Lemma 1] for F being a composite function as defined in (1.1). Note that we assume that only the second component of F has coordinates with Lipschitz gradients.

Lemma 6. *Assume that F satisfies (A5). Let $x \in \mathrm{dom}(G)$ and $\lambda \in [0,1]$. Then*

$$F(x + \lambda[p(x) - x]) \preceq F(x) + \Big(\lambda\theta(x) + \frac{L}{2}\|p(x) - x\|^2\lambda^2\Big)e. \tag{3.9}$$

Proof. Let $j \in \mathcal{J}$. Since h_j has gradient Lipschitz continuous with constant L_j, $x \in \mathrm{dom}(G)$ and $\lambda \in [0,1]$, we have

$$f_j(x+\lambda[p(x)-x]) \leq g_j((1-\lambda)x+\lambda g_j(p(x))+h_j(x)+\lambda\langle\nabla h_j(x),(p(x)-x)\rangle+\frac{L_j}{2}\|p(x)-x\|^2\lambda^2.$$

Considering that g_j is convex, we have $g_j((1-\lambda)x+\lambda p) \leq (1-\lambda)g_j(x)+\lambda g_j(p)$. Thus, combining this two previous inequalities, after some algebraic manipulations, we obtain

$$f_j(x+\lambda[p(x)-x]) \leq f_j(x)+\lambda\left[\langle\nabla h_j(x),(p(x)-x)\rangle-g_j(x)+g_j(p(x))\right]+\frac{L_j}{2}\|p(x)-x\|^2\lambda^2.$$

Therefore, by (3.5) and due to $L=\max\{L_j:\ j=1,\dots m\}$, we have

$$f_j(x+\lambda[p(x)-x]) \leq f_j(x)+\lambda\theta(x)+\frac{L}{2}\|p(x)-x\|^2\lambda^2.$$

Since the last inequality holds for all $j=1,\dots,m$, then (3.9) follows. □

4 The generalized conditional gradient method

In this section, we introduce a generalization of the conditional gradient method, also known as Frank-Wolfe algorithm, to solve multiobjective composite optimization problems. We will also study asymptotic convergence properties and iteration-complexity bounds for the sequence generated by this method. The analysis is carried out with three different step size strategies, namely, Armijo type, adaptive and diminishing step sizes. The conceptual method is described in Algorithm 1 below.

Algorithm 1. Generalized CondG method

Step 0. *Choose* $x^0 \in \mathrm{dom}(G)$ *and initialize* $k \leftarrow 0$.

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

$$p(x^k) \in \arg\min_{u\in\mathbb{R}^n}\max_{j\in\mathcal{J}}\left(g_j(u)-g_j(x^k)+\left\langle\nabla h_j(x^k),u-x^k\right\rangle\right), \tag{4.1}$$

$$\theta(x^k)=\max_{j\in\mathcal{J}}\left(g_j(p(x^k))-g_j(x^k)+\left\langle\nabla h_j(x^k),p(x^k)-x^k\right\rangle\right). \tag{4.2}$$

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

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

$$x^{k+1}:=x^k+\lambda_k(p(x^k)-x^k). \tag{4.3}$$

Step 4. *Set* $k \leftarrow k+1$ *and go to* **Step 1**.

Remark 1. *Let* $G:\mathbb{R}^n\to\overline{\mathbb{R}}^m$ *be the indicator function of the set* $\mathcal{C}\subset\mathbb{R}^n$ *in the multiobjective context, i.e., for all* $j\in\mathcal{J}$, *we have* $g_j(x)=0$ *for all* $x\in\mathcal{C}$, *and* $g_j(x)=+\infty$ *for all* $x\notin\mathcal{C}$. *Then, assumptions (A2)–(A4) concerning* G *are satisfied. Furthermore, Algorithm 1 merges into [3, Algorithm 1].*

As a consequence of Lemma 5, Algorithm 1 successfully stops if a Pareto critical point is found. Thus, from now on, we assume, without loss of generality, that $\theta(x^k) < 0$ for all $k \geq 0$ and therefore an infinite sequence $(x^k)_{k\in\mathbb{N}}$ is generated by Algorithm 1. We will analyze the generated sequence with three step size strategies. We begin by presenting the Armijo-type step size.

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 $\lambda_{k_0} = 1$ and initialize $\ell \leftarrow 0$.*

Step LS1. *If $F(x^k + \lambda_{k_\ell}[p(x^k) - x^k]) \preceq F(x^k) - \zeta\lambda_{k_\ell}|\theta(x^k)|e$, then set $\lambda_k := \lambda_{k_\ell}$ and return to the main algorithm.*

Step LS2. *Find $\lambda_{k_{\ell+1}} \in [\omega_1\lambda_{k_\ell}, \omega_2\lambda_{k_\ell}]$, set $\ell \leftarrow \ell + 1$, and go to Step LS1.*

The second step size strategy is classical in the analysis of the scalar conditional gradient method, see for example [6].

Adaptive step size. *Assume that $F := (f_1, \ldots, f_m)$ satisfies (A5) (and thus (3.9) in Lemma 6). Define the step size as*

$$\lambda_k := \min\left\{1, \frac{|\theta(x^k)|}{L\|p(x^k) - x^k\|^2}\right\} = \operatorname{argmin}_{\lambda\in(0,1]}\Big(-|\theta(x^k)|\lambda + \frac{L}{2}\|p(x^k) - x^k\|^2\lambda^2\Big). \tag{4.4}$$

Since $\theta(x) < 0$ and $p(x) \neq x$ for non-stationary points, the adaptive step size is well defined. Next we present the third step size, which is well known in the study of scalar conditional gradient method, see for example [34].

Diminishing step size. *Define the step size as*

$$\lambda_k := \frac{2}{k+2}. \tag{4.5}$$

4.1 Convergence analysis using Armijo step sizes

In this section, we analyze the sequence $(x^k)_{k\in\mathbb{N}}$ generated by Algorithm 1 with Armijo step sizes. We begin by showing that the Armijo step size strategy is well defined. First, notice that assumptions (A2)–(A3) imply that $p(x^k) \in \operatorname{dom}(G)$ and $\theta(x^k)$ in (4.1) and (4.2), respectively, are well defined.

Proposition 7. *Let $\zeta \in (0,1)$, $x^k \in \operatorname{dom}(G)$, $p(x^k)$ and $\theta(x^k)$ as in (4.1) and (4.2), respectively. Then, there exists $0 < \bar{\eta} \leq 1$ such that*

$$F(x^k + \eta[p(x^k) - x^k]) \preceq F(x^k) - \zeta\eta|\theta(x^k)|e, \quad \forall \eta \in (0, \bar{\eta}]. \tag{4.6}$$

Proof. Since H is continuously differentiable, G is convex, $x^k \in \operatorname{dom}(G)$, and $p(x^k) \in \operatorname{dom}(G)$, we conclude, for all $\eta \in (0,1)$, that

$$\begin{aligned} F(x^k + \eta[p(x^k) - x^k]) &= G(x^k + \eta[p(x^k) - x^k]) + H(x^k + \eta[p(x^k) - x^k]) \\ &\preceq (1-\eta)G(x^k) + \eta G(p(x^k)) + H(x^k) + \eta JH(x^k)(p(x^k) - x^k) + o(\eta)e. \end{aligned}$$

where $JH(x^k)$ denotes the Jacobian of H at x^k. After some arrangement in the right hand side of the last inequality, we obtain

$$F(x^k + \eta[p(x^k) - x^k]) \preceq F(x^k) + \eta\Big(JH(x^k)(p(x^k) - x^k) + G(p(x^k)) - G(x^k)\Big) + o(\eta)e.$$

Using (4.2), we have

$$F(x^k + \eta[p(x^k) - x^k]) \preceq F(x^k) + \zeta\eta\theta(x^k)e + \eta\Big((1-\zeta)\theta(x^k) + \frac{o(\eta)}{\eta}\Big)e.$$

Terefore, considering that $\theta(x^k) < 0$, $\zeta \in (0,1)$, and $\lim_{\eta\to 0} o(\eta)/\eta = 0$, there exists $\bar{\eta} > 0$ such that (4.6) holds for all $\eta \in (0, \bar{\eta}]$, concluding the proof. □

In the following, we present our first asymptotic convergence result. It is worth noting that we only assume (A1)–(A3).

Theorem 8. *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 1 with Armijo step sizes. Then, every limit point $\bar{x}$ of $(x^k)_{k\in\mathbb{N}}$ is a Pareto critical point for problem* (1.1).

Proof. Let $\bar{x} \in \operatorname{dom}(G)$ be a limit point of the sequence $(x^k)_{k\in\mathbb{N}}$ generated by Algorithm 1 and $\mathbb{K} \underset{\infty}{\subset} \mathbb{N}$ such that $\lim_{k\in\mathbb{K}} x^k = \bar{x}$. It follows from the Armijo step size strategy that

$$0 \prec -\zeta\lambda_k\theta(x^k)e \preceq F(x^k) - F(x^{k+1}), \quad \forall k \in \mathbb{N}, \tag{4.7}$$

because $\theta(x^k) < 0$ for all $k \in \mathbb{N}$. Consequently, the sequence of functional values $(F(x^k))_{k\in\mathbb{N}}$ is monotone decreasing. Moreover, since F is continuous, we have $\lim_{k\in\mathbb{K}} F(x^k) = F(\bar{x})$. Hence, $(F(x^k))_{k\in\mathbb{N}}$ converges and $\lim_{k\in\mathbb{N}}[F(x^k) - F(x^{k+1})] = 0$. Thus, (4.7) implies $\lim_{k\in\mathbb{N}} \lambda_k\theta(x^k) = 0$ and, *a fortiori*, $\lim_{k\in\mathbb{K}} \lambda_k\theta(x^k) = 0$. Therefore, there exists $\mathbb{K}_1 \underset{\infty}{\subset} \mathbb{K}$ such that at least one of the two following possibilities holds: $\lim_{k\in\mathbb{K}_1} \theta(x^k) = 0$ or $\lim_{k\in\mathbb{K}_1} \lambda_k = 0$. In case $\lim_{k\in\mathbb{K}_1} \theta(x^k) = 0$, using Lemma 5, we obtain $\theta(\bar{x}) = 0$, which implies that $\bar{x}$ is a Pareto critical point. Now consider the case $\lim_{k\in\mathbb{K}_1} \lambda_k = 0$. Suppose by contradiction that $\theta(\bar{x}) < 0$. Since $\theta(\cdot)$ is upper semicontinuous, $\lim_{k\in\mathbb{K}_1} x^k = \bar{x}$, $\theta(\bar{x}) < 0$, and $\lim_{k\in\mathbb{K}_1} \lambda_k = 0$, there exist $\delta > 0$ and $\mathbb{K}_2 \underset{\infty}{\subset} \mathbb{K}_1$ such that

$$\theta(x^k) < -\delta, \quad \forall k \in \mathbb{K}_2, \tag{4.8}$$

and also $\lambda_k < 1$ for all $k \in \mathbb{K}_2$. Moreover, since $(p(x^k))_{k\in\mathbb{N}} \subset \operatorname{dom}(G)$ and $\operatorname{dom}(G)$ is compact, we assume, without loss of generality, that there exists $\bar{p} \in \operatorname{dom}(G)$ such that

$$\lim_{k\in\mathbb{K}_2} p(x^k) = \bar{p}. \tag{4.9}$$

Since $\lambda_k < 1$ for all $k \in \mathbb{K}_2$, by the Armijo step size strategy, there exists

$$\bar{\lambda}_k \in (0, \lambda_k/\omega_1] \tag{4.10}$$

such that

$$F(x^k + \bar{\lambda}_k[p(x^k) - x^k]) \npreceq F(x^k) + \zeta\bar{\lambda}_k\theta(x^k)e, \quad \forall k \in \mathbb{K}_2,$$

which means that

$$f_{j_k}(x^k + \bar{\lambda}_k[p(x^k) - x^k]) > f_{j_k}(x^k) + \zeta\bar{\lambda}_k\theta(x^k), \quad \forall k \in \mathbb{K}_2,$$

for at least one $j_k \in \mathcal{J}$. Since $\mathcal{J}$ is finite set of indexes and $\mathbb{K}_2$ is infinite, there exist $j^* \in \mathcal{J}$ and $\mathbb{K}_3 \underset{\infty}{\subset} \mathbb{K}_2$ such that

$$\frac{f_{j^*}(x^k + \bar{\lambda}_k[p(x^k) - x^k]) - f_{j^*}(x^k)}{\bar{\lambda}_k} > \zeta\theta(x^k), \quad \forall k \in \mathbb{K}_3. \tag{4.11}$$

On the other hand, owing to $0 < \bar{\lambda}_k \leq 1$ and g_{j^*} be convex, we can apply Lemma 1 to obtain

$$\frac{g_{j^*}(x^k + \bar{\lambda}_k[p(x^k) - x^k]) - g_{j^*}(x^k)}{\bar{\lambda}_k} \leq g_{j^*}(p(x^k)) - g_{j^*}(x^k), \quad \forall k \in \mathbb{K}_3. \tag{4.12}$$

Moreover, due to h_{j^*} be continuously differentiable and $\lim_{k\in\mathbb{K}_3} \bar{\lambda}_k = 0$, we have, for all $k \in \mathbb{K}_3$,

$$\bar{\lambda}_k\langle \nabla h_{j^*}(x^k), p(x^k) - x^k\rangle = h_{j^*}\left(x^k + \bar{\lambda}_k[p(x^k) - x^k]\right) - h_{j^*}(x^k) - o(\bar{\lambda}_k\|p(x^k) - x^k\|). \tag{4.13}$$

Combining (4.2) with (4.12) and (4.13), after some algebraic manipulations, we obtain

$$\begin{aligned} \theta(x^k) &\geq g_{j^*}(p(x^k)) - g_{j^*}(x^k) + \langle \nabla h_{j^*}(x^k), p(x^k) - x^k\rangle \\ &\geq \frac{f_{j^*}\left(x^k + \bar{\lambda}_k[p(x^k) - x^k]\right) - f_{j^*}(x^k)}{\bar{\lambda}_k} - \frac{o(\bar{\lambda}_k\|p(x^k) - x^k\|)}{\bar{\lambda}_k}, \quad \forall k \in \mathbb{K}_3. \end{aligned} \tag{4.14}$$

Hence, (4.11) and (4.14) imply that

$$\frac{f_{j^*}\left(x^k + \bar{\lambda}_k[p(x^k) - x^k]\right) - f_{j^*}(x^k)}{\bar{\lambda}_k} > \left(\frac{-\zeta}{1-\zeta}\right)\frac{o(\bar{\lambda}_k\|p(x^k) - x^k\|)}{\bar{\lambda}_k}, \quad \forall k \in \mathbb{K}_3. \tag{4.15}$$

On the other hand, it follows from (4.8) and (4.14) that

$$-\delta + \frac{o(\bar{\lambda}_k\|p(x^k) - x^k\|)}{\bar{\lambda}_k} > \frac{f_{j^*}\left(x^k + \bar{\lambda}_k[p(x^k) - x^k]\right) - f_{j^*}(x^k)}{\bar{\lambda}_k}, \quad \forall k \in \mathbb{K}_3.$$

Combining the last inequality with (4.15), we have

$$-\delta - \frac{o(\bar{\lambda}_k\|p(x^k) - x^k\|)}{\bar{\lambda}_k} > \left(\frac{-\zeta}{1-\zeta}\right)\frac{o(\bar{\lambda}_k\|p(x^k) - x^k\|)}{\bar{\lambda}_k}, \quad \forall k \in \mathbb{K}_3,$$

and hence

$$-\delta - \frac{o(\bar{\lambda}_k\|p(x^k) - x^k\|)}{\bar{\lambda}_k\|p(x^k) - x^k\|}\|p(x^k) - x^k\| > \left(\frac{-\zeta}{1-\zeta}\right)\frac{o(\bar{\lambda}_k\|p(x^k) - x^k\|)}{\bar{\lambda}_k\|p(x^k) - x^k\|}\|p(x^k) - x^k\|, \quad \forall k \in \mathbb{K}_3.$$

It follows by (4.10) that $\lim_{k\in\mathbb{K}_3} \bar{\lambda}_k = 0$. Thus, considering (4.9) and taking limits for $k \in \mathbb{K}_3$ on both sides of the latter inequality, we obtain $-\delta \geq 0$, which is a contradiction with $\delta > 0$. Therefore, $\theta(\bar{x}) = 0$ and, from Lemma 5 (*ii*), we conclude that $\bar{x}$ is a Pareto critical point. □

To state the following results, we introduce some notations. Since $\mathrm{dom}(G)$ is a compact set and ∇h_j is continuous, we set

$$\rho := \sup\{\|\nabla h_j(x)\| \mid x \in \mathrm{dom}(G),\ j \in \mathcal{J}\}. \tag{4.16}$$

Moreover, assuming (A4) and (A5), we define

$$\gamma := \min\left\{\frac{1}{(\rho + L_G)\Omega},\ \frac{2\omega_1(1-\zeta)}{L\Omega^2}\right\}. \tag{4.17}$$

Lemma 9. *Assume that F satisfies (A4)–(A5). Then $\lambda_k \geq \gamma|\theta(x^k)| > 0$, for all $k \in \mathbb{N}$.*

Proof. Since $\lambda_k \in (0, 1]$ for all $k \in \mathbb{N}$, let us consider two possibilities: $\lambda_k = 1$ and $0 < \lambda_k < 1$. First we assume that $\lambda_k = 1$. It follows from (4.2) and Lemma 5 that

$$\theta(x^k) = \max_{j\in\mathcal{J}} \left\{ g_j(p(x^k)) - g_j(x^k) + \langle \nabla h_j(x^k), p(x^k) - x^k \rangle \right\} < 0,$$

which implies that $0 < -\theta(x^k) \leq g_j(x^k) - g_j(p(x^k)) + \langle \nabla h_j(x^k), x^k - p(x^k) \rangle$ for all $j \in \mathcal{J}$. Thus, the Cauchy inequality together with (A4) and (4.17) imply that

$$0 < -\theta(x^k) \leq \Big(L_G + \|\nabla h_j(x^k)\| \Big) \|p(x^k) - x^k\|.$$

Using (4.16), we have $0 < -\theta(x^k) \leq (\rho + L_G)\Omega$. Hence, the definition of γ in (4.17) implies that

$$0 < -\gamma\theta(x^k) \leq \frac{-\theta(x^k)}{(\rho + L_G)\Omega} \leq 1,$$

which shows that the desired equality holds for $\lambda_k = 1$. Now, we assume $0 < \lambda_k < 1$. Thus, from the Armijo step size strategy, we conclude that there exist $0 < \bar{\lambda}_k \leq \min\{1, \lambda_k/\omega_1\}$ and $j_k \in \mathcal{J}$, such that

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

On the other hand, by using Lemma 6, we have

$$f_j(x^k + \bar{\lambda}_k[p(x^k) - x^k]) \leq f_j(x^k) + \bar{\lambda}_k\theta(x^k) + \frac{L}{2}\|p(x^k) - x^k\|^2\bar{\lambda}_k^2, \quad \forall j \in \mathcal{J}.$$

Thus, combining the two previous inequalities with $0 < \bar{\lambda}_k \leq \min\{1, \lambda_k/\omega_1\}$, we conclude that

$$-\theta(x^k)(1 - \zeta) < \frac{L}{2}\|p(x^k) - x^k\|^2\bar{\lambda}_k \leq \frac{L}{2}\|p(x^k) - x^k\|^2\frac{\lambda_k}{\omega_1}.$$

Therefore, using the definition of Ω in (3.1) together with the definition of γ in (4.17), we obtain

$$0 < -\gamma\theta(x^k) = -\frac{2\omega_1(1 - \zeta)}{L\Omega^2}\theta(x^k) < \lambda_k,$$

which implies that desired inequality also holds for $0 < \lambda_k < 1$. □

In the following theorem, under Lipschitz assumptions, we obtain our first iteration-complexity bound by showing that the gap function (3.2) converges to zero with rate $\mathcal{O}(1/\sqrt{k})$. For that, we define

$$f_0^{\max} := \max\{f_j(x^0) \mid j \in \mathcal{J}\} \quad \text{and} \quad f^{\inf} := \min\{f_j^* \mid j \in \mathcal{J}\}, \tag{4.18}$$

where $f_j^* := \inf\{f_j(x) \mid x \in \mathrm{dom}(G)\}$ for all $j \in \mathcal{J}$.

Theorem 10. *Assume that F satisfies (A4)–(A5). Then $\lim_{k\to\infty} F(x^k) = F(x^*)$, for some $x^* \in \mathrm{dom}(G)$. Moreover, there hold:*

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

ii) $\min\{|\theta(x^k)| \mid k = 0, 1, \ldots, N - 1\} \leq \sqrt{(f_0^{\max} - f^{\inf})/(\zeta\gamma N)}$.

Proof. By the Armijo step size strategy and considering that $\theta(x^k) < 0$ for all $k \in \mathbb{N}$, we have $F(x^k + \lambda_k[p(x^k) - x^k]) \preceq F(x^k) + \zeta\lambda_k\theta(x^k)e$ or, equivalently, $\zeta\lambda_k|\theta(x^k)|e \preceq F(x^k) - F(x^{k+1})$. Hence, due to $\theta(x^k) < 0$, using Lemma 9, we obtain

$$0 \prec \zeta\gamma|\theta(x^k)|^2 e \preceq F(x^k) - F(x^{k+1}), \tag{4.19}$$

which implies that the sequence $(F(x^k))_{k\in\mathbb{N}}$ is monotone decreasing. On the other hand, since $(x^k)_{k\in\mathbb{N}} \subset \mathrm{dom}(G)$ and $\mathrm{dom}(G)$ is compact, there exists $x^* \in \mathrm{dom}(G)$ a limit point of $(x^k)_{k\in\mathbb{N}}$. Let $\mathbb{K} \underset{\infty}{\subset} \mathbb{N}$ be such that $\lim_{k\in\mathbb{K}} x^k = x^*$. Since, by (A4)–(A5), F is continuous in $\mathrm{dom}(G)$, it follows that $\lim_{k\in\mathbb{K}} F(x^k) = F(x^*)$. Thus, due to the monotonicity of the sequence $(F(x^k))_{k\in\mathbb{N}}$, we obtain that $\lim_{k\to\infty} F(x^k) = F(x^*)$. Hence, taking limits on (4.19), we obtain $\lim_{k\to\infty}|\theta(x^k)|^2 = 0$, which implies item *(i)*. By summing both sides of the second inequality in (4.19) for $k = 0, 1, \ldots, N-1$ and using (4.18), we obtain

$$\sum_{k=0}^{N-1} |\theta(x^k)|^2 \leq \frac{1}{\zeta\gamma}(f_0^{\max} - f^{\inf}).$$

Thus, $\min\{|\theta(x^k)|^2 \mid k = 0, 1, \ldots, N-1\} \leq (f_0^{\max} - f^{\inf})/(\zeta\gamma N)$, which implies the item *(ii)*. □

Corollary 11. *Assume that F satisfies (A4)–(A5) and $\varepsilon > 0$. Define the set $K(\varepsilon) := \{k \in \mathbb{N} \mid |\theta(x^k)| > \varepsilon\}$. Then,*

$$|K(\varepsilon)| \leq \frac{f_0^{\max} - f^{\inf}}{\zeta\gamma}\frac{1}{\varepsilon^2},$$

where $|K(\varepsilon)|$ denotes the number of elements of $K(\varepsilon)$.

Proof. The proof follows straightforwardly from item (ii) of Theorem 10. □

We next estimate the total number of evaluations of functions and gradients for Algorithm 1 to find an approximate Pareto critical point. We mention that similar results, with respect to the scalar gradient method, were obtained in [30].

Corollary 12. *Assume that F satisfies (A4)–(A5) and $\varepsilon > 0$. Consider an iteration k and let $F(x^k)$ be given. If $|\theta(x^k)| > \varepsilon$, then the Armijo line search algorithm performs, at most, $1 + \ln(\gamma\varepsilon)/\ln(\omega_2)$ evaluations of F to compute the step size λ_k.*

Proof. Let $\ell(k)$ and $e(k)$ be, respectively, the number of inner iterations and the number of evaluations of F in the Armijo line search algorithm to compute λ_k. Then, by the definition of the algorithm, we have $e(k) = \ell(k) + 1$ and $\omega_2^{\ell(k)} \geq \lambda_k$. Hence, using Lemma 9, it follows that $\omega_2^{\ell(k)} \geq \gamma|\theta(x^k)|$. Since $|\theta(x^k)| > \varepsilon$, we have $\omega_2^{\ell(k)} \geq \gamma\varepsilon$. Therefore, due to $0 < w_2 < 1$, we obtain $\ell(k) \leq \ln(\gamma\varepsilon)/\ln(\omega_2)$, concluding the proof. □

Corollary 13. *Assume that F satisfies (A4)–(A5) and $\varepsilon > 0$. Then, Algorithm 1 generates a point x^k such that $|\theta(x^k)| \leq \varepsilon$, performing, at most,*

$$m\left[\left(1 + \frac{\ln(\gamma\varepsilon)}{\ln(\omega_2)}\right)\frac{f_0^{\max} - f^{\inf}}{\zeta\gamma}\frac{1}{\varepsilon^2} + 1\right] = \mathcal{O}(|\ln(\varepsilon)|\varepsilon^{-2})$$

evaluations of functions $f_1, \ldots f_m$, and

$$m\left[\frac{f_0^{\max} - f^{\inf}}{\zeta\gamma}\frac{1}{\varepsilon^2} + 1\right] = \mathcal{O}(\varepsilon^{-2})$$

evaluations of gradients $\nabla h_1, \ldots, \nabla h_m$.

Proof. The proof follows from the combination of Corollaries 11 and 12. □

We now show that, in convex cases, a convergence rate of $\mathcal{O}(1/k)$ is achieved with respect to the objective function values.

Theorem 14. *Assume that F satisfies (A4)–(A6). Moreover assume that* $\lim_{k\to\infty} F(x^k) = F(x^*)$*, for some $x^* \in \mathrm{dom}(G)$, and take $\Omega > 0$ satisfying* (3.1)*. Then, we have*

$$\min_{j\in\mathcal{J}}\Big(f_j(x^k) - f_j(x^*)\Big) \leq \frac{1}{\zeta\gamma}\frac{1}{k}, \quad \forall k \in \mathbb{N}^*, \tag{4.20}$$

where γ is given in (4.17).

Proof. Since λ_k satisfies the Armijo step size rule, we have $F(x^{k+1}) - F(x^*) \preceq F(x^k) - F(x^*) + \zeta\lambda_k\theta(x^k)e$. Thus, the last inequality, together with Lemma 9, implies

$$\min_{j\in\mathcal{J}}\Big(f_j(x^{k+1}) - f_j(x^*)\Big) \leq \min_{j\in\mathcal{J}}\Big(f_j(x^k) - f_j(x^*)\Big) - \zeta\gamma\theta(x^k)^2, \quad \forall k \in \mathbb{N}. \tag{4.21}$$

On the other hand, using the convexity of h_j, for all $j \in \mathcal{J}$, we conclude that

$$f_j(x^*) - f_j(x^k) = g_j(x^*) - g_j(x^k) + h_j(x^*) - h_j(x^k) \geq g_j(x^*) - g_j(x^k) + \big\langle \nabla h_j(x^k), x^* - x^k\big\rangle,$$

Since $(F(x^k))_{k\in\mathbb{N}}$ is decreasing monotone and $\lim_{k\to\infty} F(x^k) = F(x^*)$, we have $F(x^*) \preceq F(x^k)$, for all $k \in \mathbb{N}$. Thus, the last inequality implies that

$$0 \geq f_j(x^*) - f_j(x^k) \geq g_j(x^*) - g_j(x^k) + \langle \nabla h_j(x^k), x^* - x^k\rangle, \quad \forall j \in \mathcal{J}.$$

Taking maximum in the least inequality and using the definition of $\theta(x^k)$ in (4.2), we conclude that

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

which implies that $0 \geq -\min_{j\in\mathcal{J}}\{f_j(x^k) - f_j(x^*)\} \geq \theta(x^k)$. Therefore, we obtain

$$0 \leq \Big(\min_{j\in\mathcal{J}}\big(f_j(x^k) - f_j(x^*)\big)\Big)^2 \leq \theta(x^k)^2.$$

The combination of the last inequality with (4.21) yields

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

for all $k \in \mathbb{N}$. Finally, applying Lemma 3, with $a_k = \min_{j\in\mathcal{J}}\Big(f_j(x^k) - f_j(x^*)\Big)$ and $\Gamma = \zeta\gamma$, we obtain the desired inequality (4.20). □

4.2 Convergence analysis using adaptive and diminishing step sizes

The purpose of this section is to analyze the sequence $(x^k)_{k\in\mathbb{N}}$ generated by Algorithm 1 with adaptive and diminishing step sizes. We begin by showing that, in particular, if $(x^k)_{k\in\mathbb{N}}$ is generated by Algorithm 1 with the adaptive step size (4.4), then $(F(x^k))_{k\in\mathbb{N}}$ is a nonincreasing sequence.

Lemma 15. *Assume that F satisfies (A5). Let $(x^k)_{k\in\mathbb{N}}$ be generated by Algorithm 1 with the adaptive step size* (4.4)*. Then*

$$F(x^{k+1}) - F(x^k) \preceq -\frac{1}{2}\min\left\{|\theta(x^k)|, \frac{\theta(x^k)^2}{L\Omega^2}\right\}e, \quad \forall k \in \mathbb{N}, \tag{4.22}$$

where $\Omega > 0$ is given in (3.1)*. As a consequence, $(F(x^k))_{k\in\mathbb{N}}$ is a nonincreasing sequence.*

Proof. Let us analyze the two possibilities for λ_k defined in (4.4). First, assume that $\lambda_k = 1$. In this case, using (4.4), we have $L\|p(x^k) - x^k\|^2 \le |\theta(x^k)|$. Thus, taking into account (A5), we apply Lemma 6 with $\lambda = 1$ and $x = x^k$ to obtain

$$F(p(x^k)) \preceq F(x^k) + \left(\theta(x^k) + \frac{L}{2}\|p(x^k) - x^k\|^2\right)e \preceq F(x^k) + \left(\theta(x^k) + \frac{1}{2}|\theta(x^k)|\right)e. \tag{4.23}$$

Due to $\lambda_k = 1$, it follows from (4.3) that $x^{k+1} = p(x^k)$. Therefore, since $|\theta(x^k)| = -\theta(x^k)$, we conclude from (4.23) that

$$F(x^{k+1}) \preceq F(x^k) - \frac{1}{2}|\theta(x^k)|e. \tag{4.24}$$

Now, assume that $\lambda_k = -\theta(x^k)/(L\|p(x^k) - x^k\|^2)$. Thus, applying Lemma 6 with $\lambda = \lambda_k$ and $x = x^k$, and considering (4.3), we obtain

$$F(x^{k+1}) \preceq F(x^k) + \left(\lambda_k\theta(x^k) + \frac{L}{2}\|p(x^k) - x^k\|^2\lambda_k^2\right)e = F(x^k) - \frac{\theta(x^k)^2}{2L\|p(x^k) - x^k\|^2}e. \tag{4.25}$$

Therefore, the combination of (4.24) and (4.25) yields

$$F(x^{k+1}) \preceq F(x^k) - \frac{1}{2}\min\left\{|\theta(x^k)|, \frac{\theta(x^k)^2}{L\|p(x^k) - x^k\|^2}\right\}e.$$

Since $\Omega \ge \operatorname{diam}(\operatorname{dom}(G))$, the last inequality implies that (4.22) holds. □

The next result shows that Algorithm 1 with adaptive steps size generates a sequence such that the gap function (3.2) converges to zero with rate of (at least) $\mathcal{O}(1/\sqrt{k})$. For that, we consider constants $f_0^{\max}$ and $f^{\inf}$ as in (4.18).

Theorem 16. *Assume that F satisfies (A5). Let $(x^k)_{k\in\mathbb{N}}$ be generated by Algorithm 1 with the adaptive step size* (4.4)*. Then*

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

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

$$\min\left\{|\theta(x^k)| \mid k = 0, 1, \dots, N-1\right\} \le \max\left\{\frac{2}{N}(f_0^{\max} - f^{\inf}), \Omega\sqrt{\frac{2L}{N}(f_0^{\max} - f^{\inf})}\right\}.$$

Proof. Since, by Lemma 15, $(F(x^k))_{k\in\mathbb{N}}$ is nonincreasing and $(f_j(x^k))_{k\in\mathbb{N}}$ is bounded from below by $f^{\inf}$ for all $j \in \mathcal{J}$, it follows that $(F(x^k))_{k\in\mathbb{N}}$ converges. Lemma 15 also implies that

$$0 < \min\left\{|\theta(x^k)|, \frac{\theta(x^k)^2}{L\Omega^2}\right\}e \le 2\left(F(x^k) - F(x^{k+1})\right), \quad \forall k \in \mathbb{N}. \tag{4.26}$$

Thus, taking limits as k goes to infinity on (4.26), we obtain *(i)*. Next we proceed to prove *(ii)*. By summing both sides of the second inequality in (4.26) for $k = 0, 1, \ldots, N-1$, and taking into account the definition of $f_0^{\max}$ and $f^{\inf}$, we obtain

$$\sum_{k=0}^{N-1} \min\left\{|\theta(x^k)|, \frac{\theta(x^k)^2}{L\Omega^2}\right\} \leq 2(f_0^{\max} - f^{\inf}).$$

Therefore, we have

$$\min\left\{\min\left\{|\theta(x^k)|, \frac{\theta(x^k)^2}{L\Omega^2}\right\} \mid k = 0, 1, \ldots, N-1\right\} \leq \frac{2}{N}(f_0^{\max} - f^{\inf}),$$

which implies the statement of item *(ii)*. □

A consequence of Theorem 16 is that every limit point of the sequence generated by Algorithm 1 with adaptive steps size is Pareto critical, as stated below.

Theorem 17. *Assume that F satisfies (A5). Let $(x^k)_{k\in\mathbb{N}}$ be generated by Algorithm 1 with the adaptive step size* (4.4)*. Then, every limit point of $(x^k)_{k\in\mathbb{N}}$ is a Pareto critical point of problem* (1.1)*.*

Proof. Let $\bar{x}$ be a limit point of the sequence $(x^k)_{k\in\mathbb{N}}$ and $\mathbb{K} \underset{\infty}{\subset} \mathbb{N}$ such that $\lim_{k\in\mathbb{K}} x^k = \bar{x}$. Since $x^k \in \operatorname{dom}(G)$ for all $k \geq 0$, and due to $\operatorname{dom}(G)$ be compact, we conclude that $\bar{x} \in \operatorname{dom}(G)$. On the other hand, Theorem 16*(i)* implies that $\lim_{k\in\mathbb{K}} \theta(x^k) = 0$. Hence, considering that $\lim_{k\in\mathbb{K}} x^k = \bar{x}$, it follows from Lemma 5*(iii)* that $0 \leq \theta(\bar{x})$. Thus, owing to $\bar{x} \in \operatorname{dom}(G)$, Lemma 5*(i)* implies $\theta(\bar{x}) = 0$. Therefore, applying Lemma 5*(ii)*, we conclude that $\bar{x}$ is a Pareto critical point of problem (1.1). □

In the following, we show that, for convex problems, the convergence rate of Algorithm 1 with the adaptive or the diminishing step sizes is improved to $\mathcal{O}(1/k)$ for both the gap function and with respect to the objective function values.

Theorem 18. *Assume that F satisfies (A5)–(A6). Let $(x^k)_{k\in\mathbb{N}}$ be generated by Algorithm 1 with λ_k satisfying the adaptive or the diminishing step size, i.e.,* (4.4) *or* (4.5)*. Assume that there exists $x^* \in \operatorname{dom}(G)$ such that $F(x^k) \succeq F(x^*)$, for all $k \in \mathbb{N}$. Then*

(i) $\min_{j\in\mathcal{J}} \Big(f_j(x^k) - f_j(x^*)\Big) \leq \dfrac{2L\Omega^2}{k}, \quad \forall k \in \mathbb{N}^*;$

(ii) $\min_{\ell\in\left\{\lfloor\frac{k}{2}\rfloor+2,\cdots,k\right\}} |\theta(x^\ell)| \leq \dfrac{8L\Omega^2}{k-2}, \quad k = 3, 4, \ldots.$

Proof. We first claim that

$$F(x^{k+1}) \preceq F(x^k) + \Big(\beta_k\theta(x^k) + \frac{L}{2}\|p(x^k) - x^k\|^2\beta_k^2\Big)e, \tag{4.27}$$

where $\beta_k := 2/(k+2)$. Indeed, by applying Lemma 6 with $x = x^k$ and $\lambda = \lambda_k$, we have

$$F(x^k + \lambda_k[p(x^k) - x^k]) \preceq F(x^k) + \Big(\lambda_k\theta(x^k) + \frac{L}{2}\|p(x^k) - x^k\|^2\lambda_k^2\Big)e. \tag{4.28}$$

If λ_k is the diminishing step size given in (4.5), then (4.27) and (4.28) trivially coincide. We now assume that λ_k is the adaptive step size given in (4.4). Since $\beta_k \in (0,1]$, it follows from (4.4) that $\lambda_k\theta(x^k)+(L/2)\|p(x^k)-x^k\|^2\lambda_k^2 \leq \beta_k\theta(x^k)+(L/2)\|p(x^k)-x^k\|^2\beta_k^2$. The latter inequality together with (4.28) yields (4.27). Therefore, (4.27) holds for both adaptive and diminishing strategies. Now, by (4.27) and taking into account that $\|p(x^k) - x^k\| \leq \Omega$, we have

$$\min_{j\in\mathcal{J}} \big(f_j(x^{k+1}) - f_j(x^*)\big) \leq \min_{j\in\mathcal{J}} \big(f_j(x^k) - f_j(x^*)\big) + \beta_k\theta(x^k) + \frac{L}{2}\Omega^2\beta_k^2, \quad \forall k \in \mathbb{N}. \quad (4.29)$$

Nevertheless, given that $F(x^k) \succeq F(x^*)$, for all $k \in \mathbb{N}$, by using (A6) we obtain

$$0 \geq f_j(x^*) - f_j(x^k) \geq \big\langle \nabla h_j(x^k), x^* - x^k \big\rangle + g_j(x^*) - g_j(x^k), \quad \forall k \in \mathbb{N}.$$

Thus, taking the maximum and using the optimality of $p(x^k)$ in (4.1), we have

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

which implies $0 \leq \min_{j\in\mathcal{J}}\{f_j(x^k) - f_j(x^*)\} \leq |\theta(x^k)|$. Therefore, using (4.29), we can apply Lemma 4 with $a_k = \min_{j\in\mathcal{J}}\{f_j(x^k) - f_j(x^*)\}$, $b_k = |\theta(x^k)|$, and $A = L\Omega^2$ to obtain the desired inequalities. □

Remark 2. *It is worth nothing that, if $(F(x^k))_{k\in\mathbb{N}}$ is a nonincreasing sequence, then there exists $x^* \in \mathrm{dom}(G)$ such that $F(x^k) \succeq F(x^*)$, for all $k \in \mathbb{N}$. This can be easily seen by noting that $\mathrm{dom}(G)$ is compact and f_j is lower semicontinuous, for all $j \in \mathcal{J}$. Taking Lemma 15 into account, we can conclude that this property holds for Algorithm 1 with adaptive step sizes. Regarding the diminishing step size, we point out that the existence of such a point x^* is guaranteed in the scalar case.*

Finally, we show that Algorithm 1 with the diminishing step size is capable of finding a weakly Pareto optimal point when applied to a convex problem.

Theorem 19. *Assume that F satisfies (A5)–(A6). Let $(x^k)_{k\in\mathbb{N}}$ be generated by Algorithm 1 with λ_k satisfying the diminishing step size* (4.5). *Then*

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

As a consequence, $(x^k)_{k\in\mathbb{N}}$ has a limit point $\bar{x} \in \mathrm{dom}(G)$, which is weakly Pareto optimal for problem (1.1).

Proof. Assume, by contradiction, that there exists a constant $\delta > 0$ such that

$$\theta(x^k) < -\delta, \quad \forall k \in \mathbb{N}. \quad (4.30)$$

As in the proof of Theorem 18, we obtain

$$F(x^{k+1}) \preceq F(x^k) + \Big(\lambda_k\theta(x^k) + \frac{L}{2}\|p(x^k) - x^k\|^2\lambda_k^2\Big)e, \quad \forall k \in \mathbb{N}.$$

Hence, by (4.30) and taking into account that $\|p(x^k) - x^k\| \leq \Omega$, we have

$$\frac{F(x^{k+1}) - F(x^k)}{\lambda_k} \preceq \Big(-\delta + \frac{L}{2}\Omega^2\lambda_k\Big)e, \quad \forall k \in \mathbb{N}.$$

Since $\lim_{k\to\infty}\lambda_k = 0$, there exists $k_0 \in \mathbb{N}$ such that $-\delta + (L/2)\Omega^2\lambda_k < 0$ for all $k \geq k_0$. Thus, the above inequality implies that $(F(x^k))_{k\geq k_0}$ is a nonincreasing sequence. Now, similarly to Remark 2, there exists $x^* \in \mathrm{dom}(G)$ such that $F(x^k) \succeq F(x^*)$, for all $k \geq k_0$. Therefore, by applying Theorem 18 for $k \geq k_0$, we obtain

$$\min_{\ell\in\{\lfloor \frac{k}{2}\rfloor+2,\cdots,k\}} |\theta(x^\ell)| \leq \frac{8L\Omega^2}{k-2}, \quad \forall k \geq k_0+3,$$

which contradicts (4.30). Thus, it turns out that $\liminf_{k\to\infty}|\theta(x^k)| = 0$. Now, since $(x^k)_{k\in\mathbb{N}} \subset \mathrm{dom}(G)$ and $\mathrm{dom}(G)$ is compact, there exist $\bar{x} \in \mathrm{dom}(G)$ and $\mathbb{K} \underset{\infty}{\subset} \mathbb{N}$ such that $\lim_{k\in\mathbb{K}} x^k = \bar{x}$ and $\lim_{k\in\mathbb{K}}\theta(x^k) = 0$. Therefore, by combining Lemmas 2 and 5, we conclude that $\bar{x}$ is a weakly Pareto optimal point. □

5 Numerical experiments

This section presents some numerical experiments in order to illustrate the applicability of our approach. For this aim, we compare:

- the Generalized Conditional Gradient method (Algorithm 1);
- the Proximal Gradient method proposed in [53].

We implemented both methods using the Armijo step size strategy with parameters $\zeta = 10^{-4}$, $\omega_1 = 0.05$, and $\omega_2 = 0.95$. Without attempting to go into details, we remark that the Armijo line search was coded based on quadratic polynomial interpolations of the coordinate functions, see [41] for line search strategies in the vector optimization setting. The main difference between the two considered methods consists of the subproblem to be solved to calculate the search direction. While for the Generalized Conditional Gradient method the subproblem is given in (4.1), in the Proximal Gradient method the search direction in iteration k is defined by $d_{PG}(x^k) := p_{PG}(x^k) - x^k$, where

$$p_{PG}(x^k) = \arg\min_{u\in\mathbb{R}^n}\max_{j\in\mathcal{J}}\left(g_j(u) - g_j(x^k) + \left\langle \nabla h_j(x^k), u - x^k\right\rangle + \frac{\mu}{2}\|u - x^k\|^2\right), \qquad (5.1)$$

and $\mu > 0$ is an algorithmic parameter. In our experiments, we set $\mu := 1$. In this case, when $G(x) \equiv 0$, (5.1) reduces to the classical steepest descent approach proposed in [19]. We denote the optimal value of problem (5.1) by $\theta_{PG}(x^k)$. As in Lemma 5, $\theta_{PG}(\cdot)$ can be used to characterize Pareto critical points, see [53]. In order to standardize the stopping criteria, all runs were stopped at an iterate x^k declaring convergence if

$$\frac{\|x^k - x^{k-1}\|_\infty}{\max\{1, \|x^{k-1}\|_\infty\}} \leq 10^{-4} \quad \text{and} \quad \left|\theta_{\mathrm{PG}}(x^k)\right| \leq 10^{-4}. \qquad (5.2)$$

The first criterion in (5.2) seeks to detect the convergence of the sequence $\{x^k\}$, while the second guarantees to stop at an *approximately* Pareto critical point. For Algorithm 1, we only calculate $\theta_{\mathrm{PG}}(x^k)$ when the first criterion in (5.2) is satisfied. We also consider a stopping criterion related to failures: the maximum number of allowed iterations was set to 200. The codes are written in Matlab and are freely available at `https://github.com/lfprudente/CompositeMOPCondG`.

Set of test problems: The set of test problems is related to robust multiobjective optimization. Robust optimization deals with uncertainty in the data of optimization problems, in such a way that the optimal solution must occur in the worst possible scenario, i.e., for the worst possible value that the uncertain data can assume. Let us discuss how test problems were designed. The differentiable part H that makes up the objective function F comes from some multiobjective problem found in the literature. Table 1 shows the main characteristics of the chosen problems. The first two columns identify the name of the problem and the corresponding reference where its formulation can be found. Columns "n" and "m" inform the numbers of variables and objectives of the problem, respectively. Column "Convex" indicates whether the corresponding function H is convex or not. For each test problem, we denote the uncertainty parameter by $z \in \mathbb{R}^n$ and assume, for each $j \in \mathcal{J}$, that

$$f_j(x) := h_j(x) + \langle x, z \rangle, \quad z \in \mathcal{Z}_j,$$

where $\mathcal{Z}_j \subset \mathbb{R}^n$ is the *uncertainty set*. Minimizing $f_j(x)$ in the worst possible scenario means solving

$$\min_{x \in \mathbb{R}^n} h_j(x) + \max_{z \in \mathcal{Z}_j} \langle x, z \rangle.$$

Thus, we define

$$g_j(x) := \max_{z \in \mathcal{Z}_j} \langle x, z \rangle, \quad \forall j \in \mathcal{J}. \tag{5.3}$$

Actually, in order to fulfill hypothesis (A3), we assume that $\mathrm{dom}(g_j) = \{x \in \mathbb{R}^n \mid lb \preceq x \preceq ub\}$ for all $j \in \mathcal{J}$, where $lb, ub \in \mathbb{R}^n$ are given in the last columns of Table 1. In our tests, we define $\mathcal{Z}_j$ to be a polytope. Let $B_j \in \mathbb{R}^{n \times n}$ a nonsingular matrix and $\delta > 0$ be given. We set

$$\mathcal{Z}_j := \{z \in \mathbb{R}^n \mid -\delta e \preceq B_j z \preceq \delta e\}, \tag{5.4}$$

where $e = (1, \ldots, 1)^T \in \mathbb{R}^n$. Since $\mathcal{Z}_j$ is a nonempty and compact, $g_j(x)$ is well-defined. It is easy to see that g_j satisfies (A2)–(A4). Note that parameter δ controls the uncertainty of the problem. In the reported experiments, the elements of the matrix B_j were randomly chosen between 0 and 1. In turn, given an arbitrary point $\bar{x} \in \mathcal{C}$, parameter δ was taken as

$$\delta := \bar{\delta} \|\bar{x}\|, \tag{5.5}$$

where $0.02 \leq \bar{\delta} \leq 0.10$ was also chosen at random. We mention that the definition of the non-differentiable function G in (5.3)–(5.4) has appeared in [53]. Other works dealing with robust multiobjective optimization problems include [17, 22, 33]. In particular, [22] deals with Markowitz portfolio optimization problems where conflicting objectives concerning revenue and risk and uncertainties in the data (such as expected future returns and covariances of random variables) are present.

Solving the subproblems: We first note that a solution of the subproblem in (4.1) can be calculated by solving for $\tau \in \mathbb{R}$ and $u \in \mathbb{R}^n$ the following constrained problem

$$\begin{array}{ll} \min_{\tau,u} & \tau \\ \text{s.t.} & g_j(u) - g_j(x^k) + \langle \nabla h_j(x^k), u - x^k \rangle \leq \tau, \quad \forall j \in \mathcal{J}, \\ & lb \preceq u \preceq ub. \end{array} \tag{5.6}$$

However, since $g_j(\cdot)$ in (5.3)–(5.4) is non-differentiable, the inequalities in (5.6) are difficult to deal with. On the other hand, if we define $A_j := [B_j; -B_j] \in \mathbb{R}^{2n \times n}$ and $b_j := \delta e \in \mathbb{R}^{2n}$, then (5.3)–(5.4) can be rewritten as

$$\begin{array}{ll} \max_z & \langle x, z \rangle \\ \text{s.t.} & A_j z \preceq b_j, \end{array} \tag{5.7}$$

Problem	Ref.	n	m	Convex	lb	ub
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	[32]	2	2	Y	$(-5, -5)$	$(10, 10)$
DD1	[15]	5	2	N	$(-20, \ldots, -20)$	$(20, \ldots, 20)$
DGO1	[32]	1	2	N	-10	13
DGO2	[32]	1	2	Y	-9	9
FA1	[32]	3	3	N	$(0.01, 0.01, 0.01)$	$(1, 1, 1)$
Far1	[32]	2	2	N	$(-1, -1)$	$(1, 1)$
FDS	[18]	5	3	Y	$(-2, \ldots, -2)$	$(2, \ldots, 2)$
FF1	[32]	2	2	N	$(-1, -1)$	$(1, 1)$
Hil1	[31]	2	2	N	$(0, 0)$	$(1, 1)$
IKK1	[32]	2	3	Y	$(-50, -50)$	$(50, 50)$
IM1	[32]	2	2	N	$(1, 1)$	$(4, 2)$
JOS1	[35]	100	2	Y	$(-100, \ldots, -100)$	$(100, \ldots, 100)$
JOS4	[35]	100	2	N	$(-100, \ldots, -100)$	$(100, \ldots, 100)$
KW2	[36]	2	2	N	$(-3, -3)$	$(3, 3)$
LE1	[32]	2	2	N	$(1, 1)$	$(10, 10)$
Lov1	[39]	2	2	Y	$(-10, -10)$	$(10, 10)$
Lov2	[39]	2	2	N	$(-0.75, -0.75)$	$(0.75, 0.75)$
Lov3	[39]	2	2	N	$(-20, -20)$	$(20, 20)$
Lov4	[39]	2	2	N	$(-20, -20)$	$(20, 20)$
Lov5	[39]	3	2	N	$(-2, -2, -2)$	$(2, 2, 2)$
Lov6	[39]	6	2	N	$(0.1, -0.16, \ldots, -0.16)$	$(0.425, 0.16, \ldots, 0.16)$
LTDZ	[37]	3	3	N	$(0, 0, 0)$	$(1, 1, 1)$
MGH9[a]	[44]	3	15	N	$(-2, -2, -2)$	$(2, 2, 2)$
MGH16[a]	[44]	4	5	N	$(-25, -5, -5, -1)$	$(25, 5, 5, 1)$
MGH26[a]	[44]	4	4	N	$(-1, -1, -1 - 1)$	$(1, 1, 1, 1)$
MGH33[a]	[44]	10	10	Y	$(-1, \ldots, -1)$	$(1, \ldots, 1)$
MHHM2	[32]	2	3	Y	$(0, 0)$	$(1, 1)$
MLF1	[32]	1	2	N	0	20
MLF2	[32]	2	2	N	$(-100, -100)$	$(100, 100)$
MMR1	[42]	2	2	N	$(0.1, 0)$	$(1, 1)$
MMR2	[42]	2	2	N	$(0, 0)$	$(1, 1)$
MMR3	[42]	2	2	N	$(-1, -1)$	$(1, 1)$
MMR4	[42]	3	2	N	$(0, 0, 0)$	$(4, 4, 4)$
MOP2	[32]	2	2	N	$(-4, -4)$	$(4, 4)$
MOP3	[32]	2	2	N	$(-\pi, -\pi)$	(π, π)
MOP5	[32]	2	3	N	$(-30, -30)$	$(30, 30)$
MOP6	[32]	2	2	N	$(0, 0)$	$(1, 1)$
MOP7	[32]	2	3	Y	$(-400, -400)$	$(400, 400)$
PNR	[47]	2	2	Y	$(-2, -2)$	$(2, 2)$
QV1	[32]	10	2	N	$(0.01, \ldots, 0.01)$	$(5, \ldots, 5)$
SD	[51]	4	2	Y	$(1, \sqrt{2}, \sqrt{2}, 1)$	$(3, 3, 3, 3)$
SK1	[32]	1	2	N	-100	100
SK2	[32]	4	2	N	$(-10, -10, -10, -10)$	$(10, 10, 10, 10)$
SLCDT1	[50]	2	2	N	$(-1.5, -1.5)$	$(1.5, 1.5)$
SLCDT2	[50]	10	3	Y	$(-1, \ldots, -1)$	$(1, \ldots, 1)$
SP1	[32]	2	2	Y	$(-100, -100)$	$(100, 100)$
SSFYY2	[32]	1	2	N	-100	100
TKLY1	[32]	4	2	N	$(0.1, 0, 0, 0)$	$(1, 1, 1, 1)$
Toi4[a]	[59]	4	2	Y	$(-2, -2, -2, -2)$	$(5, 5, 5, 5)$
Toi8[a]	[59]	3	3	Y	$(-1, -1, -1, -1)$	$(1, 1, 1, 1)$
Toi9[a]	[59]	4	4	N	$(-1, -1, -1, -1)$	$(1, 1, 1, 1)$
Toi10[a]	[59]	4	3	N	$(-2, -2, -2, -2)$	$(2, 2, 2, 2)$
VU1	[32]	2	2	N	$(-3, -3)$	$(3, 3)$
VU2	[32]	2	2	Y	$(-3, -3)$	$(3, 3)$
ZDT1	[61]	30	2	Y	$(0, \ldots, 0)$	$(1, \ldots, 1)$
ZDT2	[61]	30	2	N	$(0.01, \ldots, 0.01)$	$(1, \ldots, 1)$
ZDT3	[61]	30	2	N	$(0.01, \ldots, 0.01)$	$(1, \ldots, 1)$
ZDT4	[61]	30	2	N	$(0.01, -5, \ldots, -5)$	$(1, 5, \ldots, 5)$
ZDT6	[61]	10	2	N	$(0.01, \ldots, 0.01)$	$(1, \ldots, 1)$
ZLT1	[32]	10	5	Y	$(-1000, \ldots, -1000)$	$(1000, \ldots, 1000)$

[a] This is an adaptation of a single-objective optimization problem to the multiobjective setting that can be found in [43].

Table 1: List of test problems.

for which the dual problem is given by

$$\begin{array}{ll} \min_w & \langle b_j, w \rangle \\ \text{s.t.} & A_j^\top w = x, \\ & w \succeq 0. \end{array}$$

By using duality theory, it follows that (5.6) (and thus (4.1)) is equivalent to the following linear programming problem

$$
\begin{array}{ll}
\min_{\tau,u,w_j} & \tau \\
\text{s.t.} & \langle b_j, w\rangle - g_j(x^k) + \langle \nabla h_j(x^k), u - x^k\rangle \leq \tau, \\
& A_j^\top w_j = u, \\
& w_j \succeq 0, \quad \forall j \in \mathcal{J}, \\
& lb \preceq u \preceq ub.
\end{array}
\tag{5.8}
$$

Likewise, the subproblem (5.1) of the Proximal Gradient method can be reformulated as the following quadratic programming problem

$$
\begin{array}{ll}
\min_{\tau,u,w_j} & \tau + \dfrac{\mu}{2}\|u - x^k\|^2 \\
\text{s.t.} & \langle b_j, w\rangle - g_j(x^k) + \langle \nabla h_j(x^k), u - x^k\rangle \leq \tau, \\
& A_j^\top w_j = u, \\
& w_j \succeq 0, \quad \forall j \in \mathcal{J}, \\
& lb \preceq u \preceq ub,
\end{array}
\tag{5.9}
$$

for details see [53, Section 5.2 (a)]. In our codes, we use a simplex-dual method (*linprog* routine) to solve (5.7) and (5.8), and an interior point method (*quadprog* routine) to solve (5.9).

Performance profiles: The numerical results will be shown using performance profiles graphics [16], which are useful tools for comparing several methods on a large set of test problems. Let $\mathcal{S}$ be the set of solvers, $\mathcal{P}$ be the set of problems, and $t_{p,s} > 0$ be the performance of the solver $s \in \mathcal{S}$ on the problem $p \in \mathcal{P}$, where lower values of $t_{p,s}$ mean better performances. Define the performance ratio $r_{p,s} := t_{p,s}/\min\{t_{p,s} \mid s \in \mathcal{S}\}$. Then, the performance profile is obtained by plotting, for all $s \in \mathcal{S}$, the cumulative distribution function $\rho_s : [1, \infty[\to [0, 1]$ for the performance ratio $r_{p,s}$ given by $\rho_s(\mathcal{T}) := (1/|\mathcal{P}|)\left|\{p \in \mathcal{P} \mid r_{p,s} \leq \mathcal{T}\}\right|$, where $|\cdot|$ denotes the cardinality of the set. In a performance profile graphic, $\rho_s(\mathcal{T} = 1)$ is the the fraction of problems for which solver s was the most efficient over all the methods. On the other hand, $\rho_s(\mathcal{T} \equiv \infty)$ represents the fraction of problems for which solver s was able to find a solution, independently of the required effort. Therefore, the fractions $\rho_s(\mathcal{T} = 1)$ and $\rho_s(\mathcal{T} \equiv \infty)$, which can be accessed on the extreme left and right of the graph, are usually associated with the *efficiency* and *robustness* of solver s, respectively.

5.1 Efficiency and robustness

For each test problem, we considered 100 starting points randomly generated at the corresponding $\operatorname{dom}(G) = \{x \in \mathbb{R}^n \mid lb \preceq x \preceq ub\}$. In this phase, each problem/starting point was considered an independent instance and solved by both algorithms. If an approximate critical point is found, a run is considered successful regardless of the objective function value. Figure 1 shows the results using performance profiles, comparing the algorithms with respect to: (a) CPU time; (b) number of iterations. We emphasize that the results are similar if we consider the number of function evaluations. As can be seen, the Conditional Gradient method was more efficient than the Proximal Gradient method considering both performance measures. Regarding CPU time (resp. number of iterations), the efficiencies of the algorithms were 69.1% and 29.9% (resp. 76.3% and 38.4%) for Algorithm 1 and the Proximal Gradient method, respectively. The slightly larger difference with respect to CPU time can be explained by the fact that subproblem (5.8) is simpler than subproblem (5.9), making an iteration of Algorithm 1 cheaper than an iteration of the Proximal Gradient method. Both algorithms proved to be robust on the chosen set of test problems, which is in agreement with their convergence theories. Algorithm 1 and the Proximal Gradient method successfully solved 98.4% and 96.8% of the problem instances.

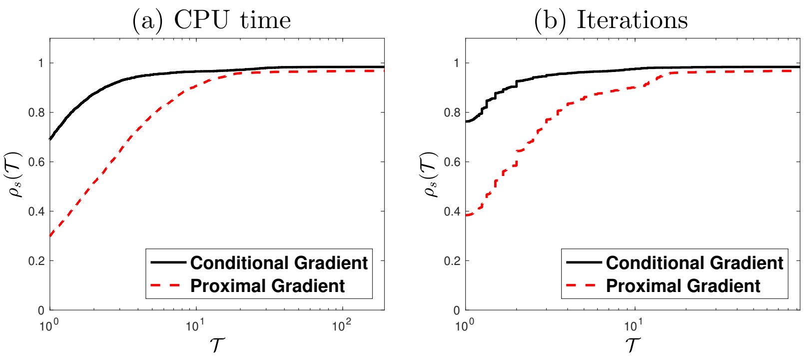

Figure 1: Performance profiles considering 100 starting points for each test problem using as the performance measurement: (a) CPU time; (b) number of iterations.

5.2 Pareto frontiers

In multiobjective optimization, we are mainly interested in estimating the Pareto frontier of a given problem. A commonly used strategy for this task is to run an algorithm from several starting points and collect the efficient points found. Thus, given a test problem, we run each algorithm for 2 minutes obtaining an approximation of the Pareto frontier. We compare the results using the well-known *Purity* and (Γ and Δ) *Spread* metrics. In summary, given a problem, the Purity metric measures the ability of an algorithm to find points on the Pareto frontier, while a Spread metric measures the ability to obtain well-distributed points along the Pareto frontier. For a careful discussion of these metrics and their uses along with performance profiles, see [14]. Figure 2(a) shows that the Conditional Gradient method was slightly more efficient than the Proximal Gradient method in terms of the Purity metric. This is not surprising, considering that both methods were run for an identical amount of time and the Conditional Gradient method is generally faster, as detailed in Section 5.1. Consequently, the Conditional Gradient method was run from a larger number of starting points, resulting in a more extensive exploration. In contrast, no significant differences are notice for the Spread metrics, as can be seen is Figures 2(b)–(c). This suggests that the Conditional Gradient method is competitive with the Proximal Gradient method in terms of obtaining *good* approximations of the Pareto frontier.

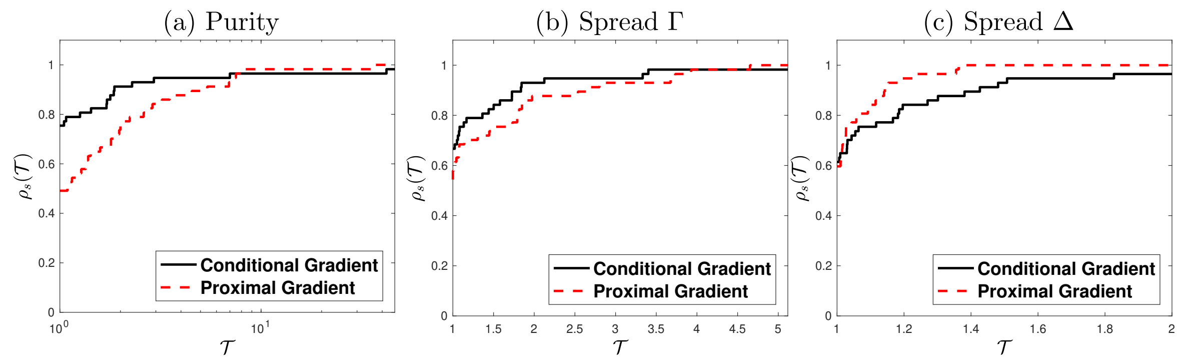

Figure 2: Metric performance profiles considering 2 minutes for each test problem: (a) Purity; (b) Spread Γ; (c) Spread Δ.

We conclude the numerical experiments by illustrating the influence of the uncertainty parameter. Figure 3 shows the image of the Pareto critical points found by Algorithm 1 using 200 random starting points for problems BK1, IM1, MOP2, SD, SLCDT1, and VU2, considering the following values for the uncertainty parameter: δ given by (5.5) with $\bar{\delta} = 0.02$, 0.05, and 0.10. As can be seen in Figure 3, as expected, smaller values of the uncertainty parameter are associated with better objective function values.

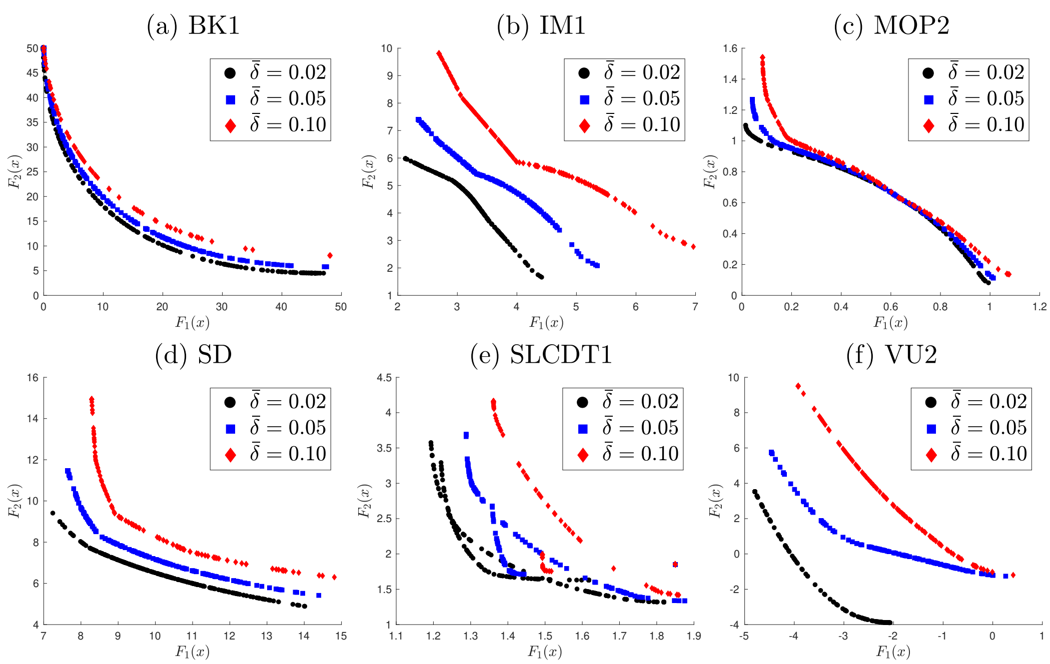

Figure 3: Image of the Pareto critical points found by Algorithm 1 with different values for the uncertainty parameter for problems: (a) BK1; (b) IM1; (c) MOP2; (d) SD; (e) SLCDT1; (f) VU2.

6 Conclusions

This paper extends the generalized conditional gradient method for multiobjective composite optimization problems. Our analysis was carried out with and without convexity and Lipschitz assumptions on the smooth component of the objective functions and considering different step size strategies. The numerical results suggests that the proposed method is competitive with the Proximal Gradient method recently introduced in [53], in terms of computational efficiency and ability to generate Pareto frontiers properly. It is noteworthy that the gap function studied here is very close to the gap function used in [53], differing only in the inclusion of a quadratic term. Therefore, it is not surprising that the two algorithms perform similarly in terms of Purity and Spread metrics. Moreover, owing to the presence of the quadratic term, the proximal gradient method is relatively slower in terms of computation effort; however, it ensures a unique solution at each iteration. Finally, it would be interesting to extend the results of the present paper for composite vector optimization problem, i.e., when the partial order is induced by other underlying cones instead of the non-negative orthant.

Data availability

The codes supporting the numerical experiments are freely available in the Github repository, `https://github.com/lfprudente/CompositeMOPCondG`.

References

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

[2] M. A. T. Ansary. A Newton-type proximal gradient method for nonlinear multi-objective optimization problems. *Optim. Methods Softw.*, 0(0):1–21, 2023.

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

[4] 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.

[5] A. Beck. *First-order methods in optimization*, volume 25 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2017.

[6] 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.

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

[8] Y. Bello-Cruz, J. G. Melo, and R. V. Serra. A proximal gradient splitting method for solving convex vector optimization problems. *Optimization*, 71(1):33–53, 2022.

[9] 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.

[10] D. Bertsekas, A. Nedić, and A. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific optimization and computation series. Athena Scientific, 2003.

[11] R. I. Boţ and S.-M. Grad. Inertial forward–backward methods for solving vector optimization problems. *Optimization*, 67(7):959–974, 2018. PMID: 30008539.

[12] K. Bredies, D. A. Lorenz, and P. Maass. A generalized conditional gradient method and its connection to an iterative shrinkage method. *Comput. Optim. Appl.*, 42(2):173–193, 2009.

[13] 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.

[14] 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.

[15] I. Das and J. E. 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.

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

[17] M. Ehrgott, J. Ide, and A. Schöbel. Minmax robustness for multi-objective optimization problems. *European J. Oper. Res.*, 239(1):17–31, 2014.

[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. Methods Oper. Res.*, 51(3):479–494, 2000.

[20] 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.

[21] J. Fliege, A. I. F. Vaz, and L. N. Vicente. Complexity of gradient descent for multiobjective optimization. *Optim. Methods Softw.*, 34(5):949–959, 2019.

[22] J. Fliege and R. Werner. Robust multiobjective optimization & applications in portfolio optimization. *European J. Oper. Res.*, 234(2):422–433, 2014.

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

[24] 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.

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

[26] A. M. Geoffrion. Proper efficiency and the theory of vector maximization. *J. Math. Anal. Appl.*, 22(3):618–630, 1968.

[27] M. L. N. Gonçalves, F. S. Lima, and L. F. Prudente. Globally convergent Newton-type methods for multiobjective optimization. *Comput. Optim. Appl.*, 83(2):403–434, Nov 2022.

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

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

[30] G. N. Grapiglia and E. W. Sachs. On the worst-case evaluation complexity of non-monotone line search algorithms. *Comput. Optim. Appl.*, 68(3):555–577, 2017.

[31] C. Hillermeier. Generalized homotopy approach to multiobjective optimization. *J. Optim. Theory Appl.*, 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 Trans. Evol. Comput.*, 10(5):477–506, 2006.

[33] J. Ide and A. Schöbel. Robustness for uncertain multi-objective optimization: a survey and analysis of different concepts. *OR Spectrum*, 38(1):235–271, 2016.

[34] 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.

[35] 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.

[36] 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.

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

[38] E. Levitin and B. Polyak. Constrained minimization methods. *USSR Comput. Math. Math. Phys.*, 6(5):1–50, 1966.

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

[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):37:1–37:23, Dec. 2019.

[42] 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.

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

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

[45] J.-W. Peng and J. Ren. Proximal quasi-Newton methods for multiobjective optimization problems. *arXiv preprint:2108.00125*, 2021.

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

[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] L. F. Prudente and D. R. Souza. A quasi-Newton method with Wolfe line searches for multiobjective optimization. *J. Optim. Theory Appl.*, 194(3):1107–1140, Sep 2022.

[49] A. Rakotomamonjy, R. Flamary, and N. Courty. Generalized conditional gradient: analysis of convergence and applications. *arXiv preprint:1510.06567*, 2015.

[50] 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.

[51] W. Stadler and J. Dauer. Multicriteria optimization in engineering: A tutorial and survey. *Progr. Astronaut. Aero.*, 150:209–209, 1993.

[52] 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.

[53] H. Tanabe, E. H. Fukuda, and N. Yamashita. Proximal gradient methods for multiobjective optimization and their applications. *Comput. Optim. Appl.*, 72(2):339–361, 2019.

[54] H. Tanabe, E. H. Fukuda, and N. Yamashita. New merit functions and error bounds for non-convex multiobjective optimization. *arXiv preprint arXiv:2010.09333*, 2020.

[55] H. Tanabe, E. H. Fukuda, and N. Yamashita. An accelerated proximal gradient method for multiobjective optimization. *arXiv preprint arXiv:2202.10994*, 2022.

[56] H. Tanabe, E. H. Fukuda, and N. Yamashita. Convergence rates analysis of a multiobjective proximal gradient method. *Optim. Lett.*, 2022.

[57] H. Tanabe, E. H. Fukuda, and N. Yamashita. A globally convergent fast iterative shrinkage-thresholding algorithm with a new momentum factor for single and multi-objective convex optimization. *arXiv preprint arXiv:2205.05262*, 2022.

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

[59] 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.

[60] J. Wang, Y. Hu, C. K. Wai Yu, C. Li, and X. Yang. Extended Newton methods for multiobjective optimization: majorizing function technique and convergence analysis. *SIAM J. Optim.*, 29(3):2388–2421, 2019.

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