UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
PROGRAMA DE PÓS GRADUAÇÃO EM MATEMÁTICA

PEDRO BONFIM DE ASSUNÇÃO FILHO

CONDITIONAL GRADIENT METHOD FOR MULTIOBJECTIVE OPTIMIZATION

Goiânia
Agosto 2021

UFG

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

[ ] Dissertação [X] Tese

2. Nome completo do autor

Pedro Bonfim de Assunção Filho

3. Título do trabalho

Conditional Gradient Methods for Multiobjective optimization

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento [ ] SIM [X] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

Documento assinado eletronicamente por **Orizon Pereira Ferreira**, **Professora do Magistério Superior**, em 26/08/2021, às 09:37, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **PEDRO BONFIM DE ASSUNCAO FILHO**, **Discente**, em 26/08/2021, às 12:03, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2305663** e o código CRC **246B9616**.

Referência: Processo nº 23070.033559/2021-21 SEI nº 2305663

UFG

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

[ ] Dissertação [X] Tese [ ] Outro*:_____________

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Pedro Bonfim de Assunção Filho

3. Título do trabalho

"Conditional Gradient Methods for Multiobjective optimization".

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento [X] SIM [ ] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

Documento assinado eletronicamente por **PEDRO BONFIM DE ASSUNCAO FILHO**, **Usuário Externo**, em 18/04/2023, às 20:09, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3683598** e o código CRC **1E4CCD6E**.

Referência: Processo nº 23070.033559/2021-21 SEI nº 3683598

Pedro Bonfim de Assunção Filho

CONDITIONAL GRADIENT METHOD FOR MULTIOBJECTIVE OPTIMIZATION

Tese apresentada ao Programa de Pós-Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Doutor em Matemática.

Área de concentração: Otimização
Orientador: Prof. Dr. Orizon Pereira Ferreira
Coorientador: Prof. Dr. Leandro da Fonseca Prudente

Goiânia
2021

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Assunção Filho, Pedro Bonfim de
Conditional Gradient Methods for Multiobjective optimization [manuscrito] / Pedro Bonfim de Assunção Filho. - 2021.
LXVI, 66 f.

Orientador: Prof. Dr. Orizon Pereira Ferreira; co-orientador Dr. Leandro Da Fonseca Prudente.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Matemática e Estatística (IME), Programa de Pós-Graduação em Matemática, Goiânia, 2021.
Bibliografia.
Inclui gráfico, tabelas, algoritmos.

1. Conditional gradient method. 2. generalized conditional gradient method. 3. multiobjective optimization. 4. pareto optimality. 5. constrained optimization problem. I. Ferreira, Orizon Pereira, orient. II. Título.

CDU 51

UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

ATA DE DEFESA DE TESE

Ata nº 06 da sessão de Defesa de Tese de **Pedro Bonfim de Assunção Filho**, que confere o título de Doutor em Matemática, **na área de Otimização.**

Ao sexto dia do mês de agosto do ano de dois mil e vinte um, a partir das dez horas da manhã, através de web-vídeo-conferência, realizou-se a sessão pública de Defesa de Tese intitulada **"Conditional Gradient Methods for Multiobjective optimization"**. Os trabalhos foram instalados pelo presidente da banca, Professor Doutor Orizon Pereira Ferreira - IME/UFG com a participação dos demais membros da Banca Examinadora: Co-orientador Professor Doutor Leandro da Fonseca Prudente - IME/UFG membro titular interno, Professor Doutor Jefferson Divino Gonçalves de Melo - IME/UFG membro titular interno, Professor Doutor Glaydston de Carvalho Bento - IME/UFG membro titular interno, Professor Doutor João Carlos de Oliveira Souza DMAT/UFPI membro titular externo, Professor Doutor Luis Felipe Cesar da Rocha Bueno - ICT/UNIFESP membro titular externo. Durante a arguição os membros da banca **não fizeram sugestão de alteração do título do trabalho**. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Tese, tendo sido o candidato **aprovado** pelos seus membros. Proclamados os resultados pelo Professor Doutor Orizon Pereira Ferreira - IME/UFG, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, Ao sexto dia do mês de agosto do ano de dois mil e vinte um.

TÍTULO SUGERIDO PELA BANCA

Conditional Gradient Methods for Multiobjective optimization

Documento assinado eletronicamente por **Luis Felipe Cesar da Rocha Bueno**, **Usuário Externo**, em 04/08/2021, às 23:58, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Leandro Da Fonseca Prudente**, **Professor do Magistério Superior**, em 06/08/2021, às 12:30, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **JOÃO CARLOS DE OLIVEIRA SOUZA**, **Usuário Externo**, em 06/08/2021, às 12:30, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Orizon Pereira Ferreira**, **Professora do Magistério**

Superior, em 06/08/2021, às 12:30, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Jefferson Divino Gonçalves De Melo**, **Professor do Magistério Superior**, em 06/08/2021, às 12:41, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Glaydston De Carvalho Bento**, **Professor do Magistério Superior**, em 25/08/2021, às 21:12, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2170749** e o código CRC **EEC29456**.

Referência: Processo nº 23070.033559/2021-21 SEI nº 2170749

Aos meus pais, Pedro Bonfim e
Maria José.

Agradecimentos

À Deus, por ter me concedido vida e saúde para ver mais um sonho sendo realizado, pois sem Ele não seria possível a conclusão deste projeto.

Aos meus pais, pelo amor, compreenção e confiança, durante a realização dessa tese. O apoio deles foi indispensável para realização desta conquista.

Ao meu orientador, professor Dr. Orizon Pereira Ferreira, pela orientação, confiança, competência, paciência, motivação e dedicação que foi essencial para concretização deste trabalho.

Ao meu coorientador professor Dr. Leandro da Fonseca Prudente pela motivação, amizade, competência e pela grande contribuição dada no artigo que compõe o capítulo 3 desta tese, principalmente nos experimentos numéricos.

Aos professores: Dr. João Carlos de Oliveira Souza, Dr. Luiz Felipe, Dr. Jefferson D. G Melo e Dr. Glaydston de Carvalho Bento, por aceitarem o convite para participar da banca examinadora, e pelas correções sugeridas para melhoria da tese.

Aos professores do Instituto de Matemática e Estatística da UFG, que passaram pela minha trajetória durante todo período acadêmico. Em especial o professor Dr. Ole Peter Smith meu grande amigo e o qual fez a revisão do inglês da tese.

Aos amigos do doutorado, Tiago Menezes, Ademir Aguiar, Fernando Lima, Ray Vitor Guimarães, Lucas vidal, Maurício Louzero, Flavio Pinto, pela amizade e força durante essa fase da minha vida.

À FAPEG pela bolsa de estudos concedida.

Abstract

In this work, we analyze the *conditional gradient method* also known as *Frank-Wolfe method* to solve constrained multiobjective optimization problems. We also, propose and analyze a generalized version of this method to solve multiobjective composite optimization problems which consist of simultaneously minimize several objective functions. Each objetive function is sum of two functions, one is assumed to be continuously differentiable and other one is non necessarily differentiable. Both methods are analyzed with three strategies for obtaining the step sizes, namely, Armijo-type, adaptative and diminishing step sizes. Asymptotic convergence properties and iteration-complexity bounds with and without convexity assumptions on the objective function are stablished. Numerical experiments for the conditional gradient method are provided to illustrate the effectiveness of the method and certify the obtained theoretical results.

Keywords: Conditional gradient method, generalized conditional gradient method, multiobjective optimization, pareto optimality, constrained optimization problem.

Resumo

Neste trabalho, analisamos o método do gradiente condicional, também conhecido como método de Frank-Wolfe, para resolver problemas de otimização multiobjetivo restrita. Também propomos e analisamos uma versão generalizada deste método para resolver problemas de otimização composta multiobjetivo que consistem em minimizar simultaneamente várias funções objetivo. Cada função objetiva é a soma de duas funções, uma é considerada continuamente diferenciável e a outra não é necessariamente diferenciável. Ambos os métodos são analisados com três estratégias de obtenção dos tamanhos dos passos, a saber: tipo Armijo, adaptativos e tamanhos decrescentes dos passos. Propriedades de convergência assintótica e limites de complexidade de iteração com e sem suposições de convexidade na função objetivo são estabelecidas. Experimentos numéricos para o método do gradiente condicional são fornecidos para ilustrar a eficácia do método e certificar os resultados teóricos obtidos.

Palavras-chave: Método do gradiente condicional, método do gradiente condicional generalizado, otimização multiobjetivo, pareto ótimo, problema de otimização restrita.

Contents

Chapter 1

Introduction

Consider the *multiobjective composite optimization problem* stated as follows

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

where the function $G : \mathbb{R}^n \to \overline{\mathbb{R}}^m := \mathbb{R}^m \cup \{+\infty\}$. We denote $\{+\infty\}$ the vector with all components $+\infty$. The components functions g_j will be considered proper, convex, inferiorly semi-continuous, possibly nonsmooth, and $H : \mathbb{R}^n \to \mathbb{R}^m$, with continuous differentiable h_j. This way our objective function will be written $F : \mathbb{R}^n \to \overline{\mathbb{R}}^m$ given by $F(x) := (f_1(x), \ldots, f_m(x))$ with $f_j := g_j + h_j$, for all $j \in \mathcal{J} := \{1, 2, \ldots, m\}$. An important instance of this problem (1) is obtained when the function $G : \mathbb{R}^n \to \overline{\mathbb{R}}^m$ is the indicator function of the set $\mathcal{C}$ in a multiobjective context, i.e., $g_j(x) = 0$ for all $x \in \mathcal{C}$ and $g_j(x) = +\infty$ otherwise. In this case, problem (1.1) becomes the following constrained multiobjective optimization problem

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

It is also important to note that the problem (1.1) in particular can be used to model robust multiobjective optimization problem, references dealing with this classe of problem includes [20, 25, 41, 66], and for a scalar version see [7].

In multiobjective optimization problems, a solution to the problem leads to a set of alternatives with different trade-offs between the objectives. In fact, in this new setting, a new concept of optimality becomes fundamental, namely, the concepts of Pareto optimal point and Pareto critical point, see [22]. One strategy for computing a Pareto points that has become very popular consists in the extension of methods for scalar-valued to multiobjective-valued optimization, instead of using scalarization approaches [32]. To the best of our knowledge, this strategy was coined in the work [22] that proposed a steepest descent method 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 [6, 9, 24, 28, 29, 34, 35]. 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 [8, 14, 17, 21, 23, 53, 54, 65, 67]. In particular, in [66] a proximal gradient methods to solve the problem (1.1) has been proposed. It is worth mentioning that a version of the *conditional gradient method* also known as *Frank-Wolfe optimization algorithm*, see [26,51], to solve (1.2) was proposed and analyzed in [2] (Chapter 3 of this work). And a generalized version of this method to solve (1.1) will be proposed and analyzed in Chapter 4.

The conditional gradient method 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 [26]. Ten years later, the method was extended to minimize differentiable convex functions with Lipschitz gradient over compact convex feasible sets, see [51]. 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 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 [42,46,47]. 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, [5,12,27,33,37,45,48,55].

The organization of this work is as follows. In Chapter 2 some notations and auxiliary results, used throughout of the work, are presented. In Chapter 3, we present the conditional gradient method to solve problem (1.2) and the step size strategies that will be considered. The results related to Armijo's step sizes will be presented in Section 3.2 and the ones related to adaptative, and diminishing step sizes in Section 3.3. In Section 3.4, we present some numerical experiments. In Chapter 4 we present the generalized conditional gradient method to solve problem (1.1). In Section 4.1 we present the assumptions on the problem (1.1) need to our analysis. Moreover, we introduce the gap function to solve the sub-problem associated to problem (1.1) and present its main properties. In Section 4.2 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 this method. In Section 4.3 we present some example of multiobjetive composite optimization problems (1.1) satisfying the assumptions stated in Section 4.1. Finally, some conclusions are given in Chapter 5.

Chapter 2

Preliminaries

In this chapter, we introduce some notations, definitions and basic results used throughout this thesis. We will divide into three sections. In the first section we present important results of real analysis. In second section we introduce some notations and recall results of differential calculus and convex analysis. and the chapter is ended with some basic results of multiobjective optimization are recalled in the third section. From now on we denote $\mathbb{N} = \{0, 1, 2, \ldots\}$, $\mathbb{N}^* = \{1, 2, 3, \ldots\}$.

2.1 Basics result of real analysis

In this section we start with two results for sequences of real numbers, which will be useful for our study on iteration complexity bounds for the conditional gradient method for multobjective optimization. Their proofs can be found in [59, Lemma 6, Ch.2, p.48] and [4, Lemma 13.13, Ch.13, p.387], respectively.

Lemma 2.1.1 *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*

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

Lemma 2.1.2 *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) $a_k \leq \dfrac{2A}{k}$, $\forall k \in \mathbb{N}^*$;

(ii) $\displaystyle\min_{\ell=\lfloor \frac{k}{2}+2,\cdots,k\rfloor} b_\ell \leq \frac{8A}{(k-2)}$, $\forall k \in \mathbb{N}^* - \{1, 2\}$ *where,* $\lfloor k/2 \rfloor = \max\{n \in \mathbb{N} : n \leq k/2\}$.

Next, we will present an analysis result which may be found in [62, Theorem 4.16, Ch. 4, p.89]. This guarantees a suitable definition of the subproblem that we will present in Chapter 3.

Theorem 2.1.3 *Let $f : \mathcal{C} \subset \mathbb{R}^n \to \mathbb{R}$ a continuous real function on a compact $\mathcal{C} \subset \mathbb{R}^n$. Then f reaches maximum value and minimum value in $\mathcal{C}$.*

2.2 Notations, definitions and basic results

In this section we present some notations, definitions and results used throughout our presentation. We denote the symbol $\langle \cdot, \cdot \rangle$ is the usual inner product in $\mathbb{R}^n$ and $\| \cdot \|$ 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 $\mathbb{K} \underset{\infty}{\subset} \mathbb{N}$ whenever is conveniente. The notation $\sigma(t) := o(t)$ for $t \in \mathbb{R} - \{0\}$ means that $\lim_{t \to 0} \sigma(t)/t = 0$. For $\mathcal{C} \subset \mathbb{R}^n$ compact, its *diameter* is the finite number

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

The effective domain of $\psi : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is defined as $\operatorname{dom}(\psi) := \{ x \in \mathbb{R}^n : \; \psi(x) < +\infty \}$.

Definition 2.2.1 *A function $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is upper semicontinuous (u.s.c) if, for any sequence $(x^k)_{k \in \mathbb{N}}$ converging to x^*,*

$$\limsup_{k \to \infty} \psi(x^k) \leq \psi(x^*).$$

And ψ is said to be lower semicontinuous (l.s.c) whenever $-\psi$ is (u.s.c) or, equivalently, if $\liminf_{k \to \infty} \psi(x^k) \geq \psi(x^*)$.

The function $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to *Lipschitz continuous* with constants $L > 0$ on $\mathcal{C} \subset \operatorname{dom}(\psi)$ whenever

$$|\psi(x) - \psi(y)| \leq L\|x - y\|, \qquad \forall x, y \in \mathcal{C}.$$

The function ψ is said to be *convex* on $\mathcal{C}$ if

$$\psi(\lambda x + (1 - \lambda) y) \leq \lambda \psi(x) + (1 - \lambda)\psi(y), \quad \forall \; x, y \in \mathcal{C}, \quad \forall \; \lambda \in [0, 1].$$

Next proposition is a well-known result, for a proof see [39, Proposition 2.1.2, pag. 88].

Proposition 2.2.2 *Let $Z \neq \varnothing$ an arbitrary set and for each $\beta \in Z$ define $\psi_\beta : \mathbb{R}^n \to \mathbb{R}$. It ψ_β is convex for all $\beta \in Z$, then their pointwise maximum $\psi(x) := \max\{\psi_\beta(x) : \; \beta \in Z\}$ is also convex.*

To study the nonsmooth case in Chapter 4, we need a generalized derivative in the case where the function is convex. A widely used derivative for this context in the literature will be defined below. Let $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function. The *directional derivative* of the function ψ at $x \in \operatorname{dom}(g)$ in the direction $d \in \mathbb{R}^n$, is given by

$$\psi'(x; d) := \lim_{\alpha \to 0^+} \frac{\psi(x + \alpha d) - \psi(x)}{\alpha}.$$

When ψ is differentiable at $x \in int(\mathrm{dom}(\psi))$, we can show that $\psi'(x;d) = \langle \nabla\psi(x), d\rangle$. Moreover, we have

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

The next lemma, is well known in convex analysis its proof can be found, for example in [10, Section 4.3].

Lemma 2.2.3 *Let $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function. Then, the following function*

$$(0, +\infty) \ni \lambda \mapsto \frac{\psi(x + \lambda d) - \psi(x)}{\lambda},$$

is non-decreasing. In particular, for all $\lambda \in (0, 1]$, one has

$$\frac{\psi(x + \lambda d) - \psi(x)}{\lambda} \leq \psi(x + d) - \psi(x),$$

Consequently, $\psi'(x;d) \leq \psi(x + d) - \psi(x)$.

2.3 Basics result of multiobjective optimization

Consider the problem

$$\min_{x \in \mathbb{R}^n} \Psi(x), \tag{2.1}$$

where $\Psi : \mathbb{R}^n \to \overline{\mathbb{R}}^m$ is given by $\Psi(x) = (\psi_1(x), \ldots, \psi_m(x))$. Then, we want to find a point that simultaneously minimizes these functions ψ's. To introduce the concept of solution to the problem (2.1), let us first to recall that 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 $v - u \in \mathbb{R}^m_{++}$. Therefore, a point $x^* \in \mathbb{R}^n$ is called *Pareto optimal point* for (2.1) it there exists no other $x \in \mathbb{R}^n$ with $\Psi(x) \preceq \Psi(x^*)$ and $\Psi(x) \neq \Psi(x^*)$. And $x^* \in \mathbb{R}^n$ is called *weak Pareto optimal point* for (2.1), if there exists no other $x \in \mathbb{R}^n$ such that $\Psi(x) \prec \Psi(x^*)$. When Ψ is convex the *optimality condition* for problem (2.1) at a point $\bar{x} \in \mathbb{R}^n$ is given by

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

and $\bar{x} \in \mathrm{dom}(\Psi)$ satisfying (2.2) is called a *critical Pareto point* or a *stationary point* of problem (2.1), for more details in general case see [15]. Whenever $\mathrm{dom}(\Psi) = \mathcal{C}$ and Ψ is differentiable on $\mathcal{C}$ the problem (2.1) becomes

$$\min_{x \in \mathcal{C}} \Psi(x). \tag{2.3}$$

In this case, the condition (2.2) is equivalently to

$$-\mathbb{R}^m_{++} \cap J\Psi(\bar{x})(\mathcal{C} - \bar{x}) = \varnothing, \tag{2.4}$$

where $\mathcal{C}-x := \{y-x : \ y \in \mathcal{C}\}$. The geometric optimality condition (2.4) is also equivalently stated as

$$\max_{j\in\mathcal{J}} \langle \nabla\psi_j(\bar{x}), p-\bar{x}\rangle \geq 0, \qquad \forall\, p \in \mathcal{C}. \tag{2.5}$$

In general, condition (2.5) is necessary, but not sufficient for the optimality. Let us introduce three sets, namely, the index set

$$\mathcal{J} := \{1, \ldots, m\},$$

and the *nonnegative octant* $\mathbb{R}^m_+$ and the *positive octant* $\mathbb{R}^m_{++}$ of the Euclidean space $\mathbb{R}^m$ as

$$\mathbb{R}^m_+ := \{u \in \mathbb{R}^m : \ u_j \geq 0, j \in \mathcal{J}\}, \qquad \mathbb{R}^m_{++} := \{u \in \mathbb{R}^m : u_j > 0, j \in \mathcal{J}\}.$$

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 $v-u \in \mathbb{R}^m_{++}$.

Let $\Psi : \mathbb{R}^n \to \overline{\mathbb{R}}^m$ be given by $\Psi(x) = (\psi_1(x), \ldots, \psi_m(x))$, where $\overline{\mathbb{R}}^m = \mathbb{R}^m \cup \{+\infty\}$ and $\{+\infty\}$ denotes the vector with all components $+\infty$. The *effective domain* of the function Ψ is denoted by

$$\mathrm{dom}(\Psi) := \{x \in \mathbb{R}^n : \ \psi_j(x) < +\infty, \ j \in \mathcal{J}\}.$$

Furthermore, hereafter denotes

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

Lemma 2.3.1 *Let Ψ be a differentiable function on $\mathcal{C} \subset dom(\Psi)$. Then, for all $x, p \in \mathcal{C}$, there holds*

$$\Psi\left(x + \lambda(p-x)\right) \preceq \Psi(x) + \left(\lambda \max_{j\in\mathcal{J}} \langle \nabla\psi_j(x), p-x\rangle + \frac{o(\lambda)}{\lambda}\right) e.$$

For Ψ differentiable and $x \in \mathbb{R}^n$, the Jacobian of Ψ at $x \in \mathbb{R}^n$ denoted by $J\Psi$ is the $m \times n$ matrix with entries

$$[J\Psi(x)]_{i,j} = \frac{\partial\psi_i}{\partial x_j}(x),$$

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

Definition 2.3.2 *The function Ψ has Jacobian $J\Psi$ row-wise Lipschitz continuous on $\mathcal{C} \subset dom(\Psi)$, if there exist constants $L_1, L_2, \ldots, L_m > 0$ such that*

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

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

The following lemma presents a result related to functions satisfying Definition 2.3.2.

Lemma 2.3.3 *Let Ψ be a differentiable function on $\mathcal{C} \subset dom(\Psi)$. Assume that $J\Psi$ has row-wise Lipschitz continuous with constant $L > 0$. Then, for all $x, p \in \mathcal{C}$ and $\lambda \in (0, 1]$, there holds*

$$\Psi\left(x + \lambda(p-x)\right) \preceq \Psi(x) + \big(\lambda \max_{j\in\mathcal{J}} \langle \nabla\psi_j(x), p-x\rangle + \frac{L}{2}\|p-x\|^2\lambda^2\big)e. \tag{2.6}$$

Proof. Since F satisfies Definition 2.3.2, by using the same idea in the proof of [18, Lemma 2.4.2], we conclude that

$$\Psi\left(x+\lambda(p-x)\right) \preceq \Psi(x) + \lambda J\Psi(x)(p-x) + \frac{L}{2}\|p-x\|^2\lambda^2 e.$$

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

$$J\Psi(x)(p-x) = \left(\langle\nabla\psi_1(x), p-x\rangle, \ldots, \langle\nabla\psi_m(x), p-x\rangle\right)^T \preceq \max_{j\in\mathcal{J}}\langle\nabla\psi_j(x), p-x\rangle\, e.$$

Therefore, (2.6) follows by combining the two previous vector inequalities. ■

The function Ψ is said to be *convex* on $\mathcal{C}$ if

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

or equivalently, for each $j\in\mathcal{J}$, the component function $\psi_j : \mathbb{R}^n \to \mathbb{R}\cup+\infty$ of Ψ is convex.

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, [34] for differential case and [66] for nondifferential case), we will omit its proof here.

Lemma 2.3.4 *If Ψ is convex and $\bar{x}$ is a critical Pareto point, then $\bar{x}$ is also a weak Pareto optimal point of problem* (2.1).

Chapter 3

Conditional gradient method for multiobjective optimization

In the chapter we analyze the *conditional gradient method*, for solving the problem (1.2), with $H = F : R^n \to \mathbb{R}^m$, a differentiable function. So the problem can be rewritten as follows

$$\min_{x \in \mathcal{C}} F(x). \tag{3.1}$$

The constraint set $\mathcal{C}$ 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

3.1 The conditional gradient method

In this section we state the conditional gradient method multiobjective for solving the constrained multiobjective optimization (3.1) and the strategies for the step sizes used to define the iterations of the 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 $p(x)$ is an optimal solution of the scalar-valued problem

$$\min_{u \in \mathcal{C}} \max_{j \in \mathcal{J}} \langle \nabla f_j(x), u - x \rangle , \tag{3.2}$$

i.e.,

$$p(x) \in \operatorname*{argmin}_{u \in \mathcal{C}} \max_{j \in \mathcal{J}} \langle \nabla f_j(x), u - x \rangle . \tag{3.3}$$

Note that, since the minimand of (3.2) is a convex function and $\mathcal{C}$ is a compact convex set, this problem has an optimal solution due to the Theorem 2.1.3 (possibly solution not unique) and, as consequence, $p(x)$ is well defined. Although (3.2) is a 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

$$\begin{aligned} \min_{u,\tau} \quad & \tau \\ \text{s.t.} \quad & \langle \nabla f_j(x), u - x \rangle \le \tau, \quad j \in \mathcal{J}. \\ & u \in \mathcal{C}. \end{aligned} \tag{3.4}$$

For instance, the solution of problem (3.4) can be obtained by using a linear optimization "oracle" when the convex set $\mathcal{C}$ is compact and polyhedral. Denote by $\theta(x)$ the optimal value of (3.2) given by

$$\theta(x) := \max_{j\in\mathcal{J}} \langle \nabla f_j(x), p(x) - x \rangle, \tag{3.5}$$

where $p(x) \in \mathcal{C}$ is as in (3.3). Following, we characterize $\theta(x)$ with respect to stationarity. Next proposition is a variation of the results in [34, Propositions 3 and 4] (see also [30, Proposition 4.1]).

Proposition 3.1.1 *Let $\theta : \mathcal{C} \to \mathbb{R}$ be as in* (3.5)*. Then, there hold:*

(i) $\theta(x) \le 0$ 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. Since $x \in \mathcal{C}$, it follows from (3.3) and (3.5) that $\theta(x) \le \max_{j\in\mathcal{J}} \langle \nabla f_j(x), x - x \rangle = 0$, which proof (i). To proceed to prove (ii), let $x \in \mathcal{C}$ and consider a sequence $(x^k)_{k\in\mathbb{N}} \subset \mathcal{C}$ such that $\lim_{k\to\infty} x^k = x$. On one hand, since $p(x) \in \mathcal{C}$, using (3.3) and (3.5) we have $\theta(x^k) \le \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x^k), p(x) - x^k \right\rangle$ which implies that

$$\limsup_{k\to\infty} \theta(x^k) \le \theta(x), \tag{3.6}$$

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

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

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

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

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 $\left\|p(x^k) - x^k\right\| \le \operatorname{diam}(\mathcal{C})$, we obtain, $\theta(x) \le \liminf_{k\to\infty} \theta(x^k)$. Combining (3.6) and last inequality yields $\lim_{k\to\infty} \theta(x^k) = \theta(x)$, concluding the proof of the second statement. Finally, to prove *(iii)* we assume that $x \in \mathcal{C}$ is stationary point of problem (3.1), i.e., $\max_{j\in\mathcal{J}} \langle \nabla f_j(x), u - x\rangle \ge 0$ for all $u \in \mathcal{C}$. Since $p(x) \in \mathcal{C}$, we obtain $\theta(x) = \max_{j\in\mathcal{J}} \langle \nabla f_j(x), p(x) - x\rangle \ge 0$ which, together with item (i), implies $\theta(x) = 0$. Reciprocally, suppose that $\theta(x) = 0$. It follows from the definition of $\theta(\cdot)$ in (3.5) that $0 = \theta(x) \le \max_{j\in\mathcal{J}} \langle \nabla f_j(x), u - x\rangle$, for all $u \in \mathcal{C}$, which implies that x satisfies the optimality condition (2.5). ■

A direct consequence of Proposition 3.1.1 is that if $x \in \mathcal{C}$ is a nonstationary point of problem (3.1), 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 $\langle \nabla f_j(x), d(x)\rangle \le \theta(x) < 0$ 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

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

$$
\theta(x^k) := \max_{j\in\mathcal{J}} \left\langle \nabla f_j(x^k), p(x^k) - x^k \right\rangle. \tag{3.7}
$$

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

$$
x^{k+1} := x^k + \lambda_k d(x^k). \tag{3.8}
$$

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

In view of Proposition 3.1.1, Algorithm 1 successfully stops if a stationary point of problem (3.1) is found. Thus, hereafter, *we assume that* $\theta(x^k) < 0$ *for all* $k \in \mathbb{N}$, meaning that Algorithm 1 generates an infinite sequence $(x^k)_{k\in\mathbb{N}}$. Since $x^0 \in \mathcal{C}$, $p(x^k) \in \mathcal{C}$ and $\lambda_k \in (0,1]$ for all $k \in \mathbb{N}$, and $\mathcal{C}$ is a convex set, it follows from inductive arguments that $(x^k)_{k\in\mathbb{N}} \subset \mathcal{C}$. The choice for computing the step size λ_k at Step 2 remains deliberately open. We point out that CondG method can be seen as a projection-free method whenever $\mathcal{C}$ is compact and has a simple structure, see [42]. 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 1 *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*

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

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.*

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

Proposition 3.1.2 *Let* $\zeta \in (0,1)$, $x \in \mathcal{C}$ *be a nonstationary point, and* $p(x)$ *and* $\theta(x)$ *as in* (3.3) *and* (3.5), *respectively. Then, there exists* $0 < \overline{\eta} \le 1$ *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 $\langle \nabla f_j(x), p(x) - x\rangle \le \theta(x) < 0$ for all $j \in \mathcal{J}$, the proof follows directly from [22, Lemma 4]. ∎

Adaptative step size 1 *Assume that* $F := (f_1, \ldots, f_m)^T$ *satisfies Definition 2.3.2. Define the step size as*

$$\lambda_k := \min\left\{1, \frac{-\theta(x^k)}{L\|p(x^k) - x^k\|^2}\right\} = \operatorname*{argmin}_{\lambda\in(0,1]}\left\{\theta(x^k)\lambda + \frac{L}{2}\|p(x^k) - x^k\|^2\lambda^2\right\}. \tag{3.9}$$

Since $\theta(x) < 0$ and $p(x) \neq x$ for nonstationary points, the adaptative step size for Algorithm 1 is well defined. It is worth noting that if the Lipschitz constant L can not be easily estimated, then (3.9) can not be computed and the adaptative step size is merely theoretical. Despite this limitation, we note that this step size has often been considered in the scalar setting, see, for exemple, [4,5,31]. We end this section showing the third step size strategy.

Diminishing step size 1 *Define the step size as*

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

3.2 Analysis of CondG with Armijo's step size

Throughout this section we assume that $(x^k)_{k\in\mathbb{N}}$ 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 3.2.1 *Every limit point $\bar{x}$ of $(x^k)_{k\in\mathbb{N}}$ is a Pareto critical point for the Problem 3.1.*

Proof. Let $\bar{x} \in \mathcal{C}$ be a limit point of $(x^k)_{k\in\mathbb{N}}$ 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 (3.8), $(F(x^k))_{k\in\mathbb{N}}$ satisfies

$$0 \prec -\zeta\lambda_k\theta(x^k)e \preceq F(x^k) - F(x^{k+1}), \qquad k \in \mathbb{N}. \tag{3.10}$$

Considering that F is continuous, we have $\lim_{k\in\mathbb{K}} F(x^k) = F(\bar{x})$. Thus, due to $(F(x^k))_{k\in\mathbb{N}}$ be monotone decreasing, it follows that $\lim_{k\to\infty} F(x^k) = F(\bar{x})$. Then, taking limits in (3.10), 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) $\lim\limits_{k\in\mathbb{K}_1} \theta(x^k) = 0$;

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

In case (a), by the continuity of $\theta(\cdot)$, we obtain $\theta(\bar{x}) = 0$. Thus, Proposition 3.1.1 (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 $(p(x^k))_{k\in\mathbb{N}} \subset \mathcal{C}$). 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

$$F\left(x^k + \hat{\lambda}_k[p(x^k) - x^k]\right) \npreceq F(x^k) + \zeta\hat{\lambda}_k\theta(x^k)e,$$

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

$$f_{j_*}\left(x^k + \hat{\lambda}_k[p(x^k) - x^k]\right) > f_{j_*}(x^k) + \zeta\hat{\lambda}_k\theta(x^k). \tag{3.11}$$

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 (3.7) and (3.11), 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

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

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

$$\langle \nabla f_{j_*}(\bar{x}), \bar{p} - \bar{x} \rangle \geq 0. \tag{3.12}$$

On the other hand, since $\theta(x^k) < 0$ for all $k \in \mathbb{N}$, we have $\lim_{k \in \mathbb{K}_2} \theta(x^k) = \max_{j \in \mathcal{J}} \langle \nabla f_j(\bar{x}), \bar{p} - \bar{x} \rangle \leq 0$. Therefore, using (3.12), 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 3.1.1 (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 constants:

$$\gamma := \min\left\{\frac{1}{\rho \operatorname{diam}(\mathcal{C})}, \frac{2\omega_1(1-\zeta)}{L \operatorname{diam}(\mathcal{C})^2}\right\}, \qquad 0 < \rho := \sup\{\|\nabla f_j(x)\| : \ x \in \mathcal{C}, \ j \in \mathcal{J}\}. \tag{3.13}$$

Lemma 3.2.2 *Assume that* $F := (f_1, \ldots, f_m)^T$ *satisfies Definition 2.3.2. Then,* $\lambda_k \geq \gamma |\theta(x^k)|$*, 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 $0 < \lambda_k < 1$. Assume that $\lambda_k = 1$. It follows from (3.7) 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 (3.13), we have $|\theta(x^k)| \leq \rho \operatorname{diam}(\mathcal{C})$ or, equivalently, $1 \geq |\theta(x^k)|/[\rho \operatorname{diam}(\mathcal{C})]$. Hence, from the definition of γ in (3.13), we have $1 \geq |\gamma \theta(x^k)|$, which corresponds to $\lambda_k \geq |\gamma \theta(x^k)|$ 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 2.3.3 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 inequalities, 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

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

which, together with the definition of γ in (3.13), gives the desired inequality. ■

To state the next results, define

$$F^* = (f_1^*,\ldots,f_m^*) := \inf\{F(x^k):\ k=0,1,\ldots\}, \qquad f_{j_*}^* := \min\{f_j^*:\ j\in\mathcal{J}\}.$$

Theorem 3.2.3 *Assume that $F := (f_1,\ldots,f_m)^T$ satisfies Definition 2.3.2. Then, $\lim_{k\to\infty} F(x^k) = F^*$. Moreover, there holds:*

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

(ii) for every $N\in\mathbb{N}$, one has

$$\min\left\{|\theta(x^k)|:\ k=0,1,\ldots,N-1\right\} \le \sqrt{\frac{[f_{j_*}(x^0)-f_{j_*}^*]}{\zeta\gamma N}}.$$

Proof. By the Armijo step size strategy and taking into account that $\theta(x^k)<0$, for all $k\in\mathbb{N}$, we have

$$\zeta\lambda_k|\theta(x^k)|e \preceq F(x^k)-F(x^{k+1}), \qquad k\in\mathbb{N}.$$

Thus, by Lemma 3.2.2, we obtain

$$0 \prec \zeta\gamma\theta(x^k)^2 e \preceq F(x^k)-F(x^{k+1}). \tag{3.14}$$

Hence, $(F(x^k))_{k\in\mathbb{N}}$ is monotone decreasing. On the other hand, since $(x^k)_{k\in\mathbb{N}}\subset\mathcal{C}$ and $\mathcal{C}$ is compact, there exists a limit point $x^*\in\mathcal{C}$ of $(x^k)_{k\in\mathbb{N}}$. Let $(x^{k_j})_{k\in\mathbb{N}}$ be a subsequence of $(x^k)_{k\in\mathbb{N}}$ such that $\lim_{j\to+\infty}x^{k_j}=x^*$. Considering that F is continuous, we have $\lim_{j\to+\infty}F(x^{k_j})=F(x^*)$. Thus, due to $(F(x^k))_{k\in\mathbb{N}}$ be monotone decreasing, it follows that $\lim_{k\to\infty}F(x^k)=F(x^*)$ and $F(x^*)=F^*$. Therefore, the first statement is proved. We will now prove item (i). Taking limits in (3.14), we obtain that $\lim_{k\to\infty}\theta(x^k)^2=0$, which implies item (i). To prove the item (ii), we perform a sum on both sides of the second inequality in (3.14), for $k=0,1,\ldots,N-1$, to obtain

$$\sum_{k=0}^{N-1}\theta(x^k)^2 \le \frac{1}{\zeta\gamma}[f_{j_*}(x^0)-f_{j_*}^*].$$

Therefore, $\min\left\{\theta(x^k)^2:\ k=0,1,\ldots,N-1\right\} \le [f_{j_*}(x^0)-f_{j_*}^*]/[\zeta\gamma N]$, which proves item (ii). ■

Corollary 3.2.4 *Assume that $F := (f_1,\ldots,f_m)^T$ satisfies Definition 2.3.2 and $\epsilon>0$. Define the set $K(\epsilon) := \{k: |\theta(x^k)|>\epsilon,\ k=0,1,\ldots\}$. Then,*

$$|K(\epsilon)| \le \frac{f_{j_*}(x^0)-f_{j_*}^*}{\zeta\gamma}\frac{1}{\epsilon^2},$$

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

Proof. The proof follows straightforwardly from item (ii) of Theorem 3.2.3. ■

Corollary 3.2.5 *Assume that $F := (f_1, \ldots, f_m)^T$ satisfies Definition 2.3.2 and $\epsilon > 0$. Consider an iteration k and let $F(x^k)$ be given. If $|\theta(x^k)| > \epsilon$, then the Armijo line search algorithm performs, at most, $1 + \ln(\gamma\epsilon)/\ln(\omega_2)$ evaluations of F to compute the step size λ_k.*

Proof. Let ℓ_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 3.2.2, it follows that $\omega_2^{\ell_k} \geq \gamma|\theta(x^k)|$. Since $|\theta(x^k)| > \epsilon$, we have $\omega_2^{\ell_k} \geq \gamma\epsilon$. Therefore, due to $0 < w_2 < 1$, we obtain $\ell_k \leq \ln(\gamma\epsilon)/\ln(\omega_2)$, concluding the proof. ■

Theorem 3.2.6 *Assume that $F := (f_1, \ldots, f_m)^T$ satisfies Definition 2.3.2 and $\epsilon > 0$. Then, Algorithm 1 generates a point x^k such that $\theta(x^k) \leq \epsilon$, performing, at most,*

$$m\left[\left(1 + \frac{\ln(\gamma\epsilon)}{\ln(\omega_2)}\right)\frac{f_{j_*}(x^0) - f_{j_*}^*}{\zeta\gamma}\frac{1}{\epsilon^2} + 1\right] = \mathcal{O}(|\ln(\epsilon)|\epsilon^{-2})$$

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

$$m\left[\frac{f_{j_*}(x^0) - f_{j_*}^*}{\zeta\gamma}\frac{1}{\epsilon^2} + 1\right] = \mathcal{O}(\epsilon^{-2})$$

evaluations of gradients $\nabla f_1, \ldots, \nabla f_m$.

Proof. The proof follows from the combination of Corollaries 3.2.4 and 3.2.5. ■
Similar results of Corollaries (3.2.4) and (3.2.5) and Theorem 3.2.6, with respect to the scalar gradient method, were obtained in [36].

Theorem 3.2.7 *Assume that $F := (f_1, \ldots, f_m)^T$ is convex on $\mathcal{C}$ and satisfies Definition 2.3.2. Let $(x^k)_{k\in\mathbb{N}}$ be generated by Algorithm 1 with the Armijo step size strategy. Suppose that $\lim_{k\to+\infty} x^k = x^*$. Then,*

$$\min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right) \leq \frac{1}{\gamma\zeta}\frac{1}{k}, \qquad k \in \mathbb{N}. \tag{3.15}$$

Proof. By the Armijo step size strategy, item (i) of Proposition 3.1.1, and Lemma 3.2.2 we have $F(x^{k+1}) \preceq F(x^k) - \zeta\gamma\theta(x^k)^2 e$, which implies that

$$\min_{j\in\mathcal{J}}\left(f_j(x^{k+1}) - f_j(x^*)\right) \leq \min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right) - \zeta\gamma\theta(x^k)^2, \qquad k \in \mathbb{N}. \tag{3.16}$$

On the other hand, by the convexity of F, we have $f_j(x^*) - f_j(x^k) \geq \langle\nabla f_j(x^k), x^* - x^k\rangle$, for all $j \in \mathcal{J}$. Since $(F(x^k))_{k\in\mathbb{N}}$ 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 (3.7) 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 (3.16) 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, $k\in\mathbb{N}$. Finally, applying Lemma 2.1.1, 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 bound in (3.15). However, it would be interesting to go further by presenting a bound for $\max_{j\in J}(f_j(x^k)-f_j(x^*))$ instead of the bound in (3.15). This is still a challenge to be overcome. ■

3.3 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* $F := (f_1,\ldots,f_m)^T$ *satisfies Definition 2.3.2 with constant* $L > 0$. We begin by applying Lemma 2.3.3 to show that the sequence $(x^k)_{k\in\mathbb{N}}$ generated by Algorithm 1 with the adaptative step size satisfies an important inequality, which is a version of [5, Lemma A.2] for multicriteria optimization.

Proposition 3.3.1 *Let* $(x^k)_{k\in\mathbb{N}}$ *be generated by Algorithm 1 with the adaptative step size. Then,*

$$F\left(x^k+\lambda_k[p(x^k)-x^k]\right) \preceq F(x^k) - \frac{1}{2}\min\left\{-\theta(x^k), \frac{\theta(x^k)^2}{L\,\mathrm{diam}(\mathcal{C})^2}\right\}e, \qquad k\in\mathbb{N}. \quad (3.17)$$

Proof. Lemma 2.3.3 with $\lambda=\lambda_k$, $x=x^k$, $p=p(x^k)$, and $\theta=\theta(x^k)$ yields

$$F\left(x^k+\lambda_k[p(x^k)-x^k]\right) \preceq F(x^k) + \left(\theta(x^k)\lambda_k + \frac{L}{2}\|p(x^k)-x^k\|^2\lambda_k^2\right)e, \qquad k\in\mathbb{N}. \quad (3.18)$$

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 (3.9), we have $L\|p(x^k)-x^k\|^2 \leq -\theta(x^k)$. Thus inequality (3.18) becomes

$$F\left(x^k+\lambda_k[p(x^k)-x^k]\right) \preceq F(x^k) + \frac{1}{2}\theta(x^k)e. \quad (3.19)$$

Now, we assume that $\lambda_k = -\theta(x^k)/(L\|p(x^k)-x^k\|^2)$. In this case, inequality (3.18) 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 (3.19) with the last vector inequality yields

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

Since $\operatorname{diam}(\mathcal{C}) \geq \|p(x^k) - x^k\|$, the above inequality implies (3.17) and the proof is concluded. ■

Remind that we are assuming that $(x^k)_{k\in\mathbb{N}}$ genkerated by the Algorithm 1 is infinite and then $\theta(x^k) < 0$, for all $k \in \mathbb{N}$. As an application of Proposition 3.3.1, 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 are a multicriteria version of [3, Theorem 13.9]. Let us define:

$$-\infty < f_j^* := \inf\{f_j(x) :\ x \in \mathcal{C}\}, \qquad j \in \mathcal{J}. \tag{3.20}$$

Corollary 3.3.2 *Assume that $(x^k)_{k\in\mathbb{N}}$ is generated by Algorithm 1 with the adaptative step size. Let $j_* \in \mathcal{J}$ be an index such that $f_{j_*}(x^0) - f_{j_*}^* := \min\left\{f_j(x^0) - f_j^* :\ j \in \mathcal{J}\right\}$. Then,*

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

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

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

Proof. By Proposition 3.3.1, $(f_{j_*}(x^k))_{k\in\mathbb{N}}$ is nonincreasing because $\theta(x^k) < 0$, for all $k \in \mathbb{N}$. Since this sequence is also bounded from below by $f_{j_*}^*$ defined in (3.20), it turns out that $(f_{j_*}(x^k))_{k\in\mathbb{N}}$ converges. Moreover, by (3.8), Proposition 3.3.1 also implies

$$0 < \min\left\{-\theta(x^k), \frac{\theta(x^k)^2}{L\operatorname{diam}(\mathcal{C})^2}\right\} \leq 2[f_{j_*}(x^k) - f_{j_*}(x^{k+1})], \qquad k \in \mathbb{N}. \tag{3.21}$$

Since $(f_{j_*}(x^k))_{k\in\mathbb{N}}$ converges, we have $\lim_{k\to+\infty}[f_{j_*}(x^k) - f_{j_*}(x^{k+1})] = 0$. Thus, taking limits in (3.21), we have

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

which proves (i). By summing both sides of the second inequality in (3.21) for $k = 0, 1, \ldots, N-1$ and taking into accont that $f_{j_*}^* := \inf\{f_{j_*}(x) :\ x \in \mathcal{C}\}$, we obtain

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

* Therefore,

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

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

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

which gives the statement of item (ii). ■

Remark 3.3.3 A direct consequence of Corollary 3.3.2 (i) and Proposition 3.1.1 (ii) and (iii) is that every limit point $\bar{x}$ of $(x^k)_{k\in\mathbb{N}}$ generated by Algorithm 1 with adaptative step sizes is a Pareto critical point. Moreover, by Proposition 3.3.1, $f_j(\bar{x}) = \inf\{f_j(x^k) : \ k \in \mathbb{N}\}$, for all $j \in \mathcal{J}$. In addition, if F is convex on $\mathcal{C}$, by Lemma 2.3.4, then $\bar{x}$ is a weak Pareto optimal point of problem (3.1).

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 3.3.4 *Consider $(x^k)_{k\in\mathbb{N}}$ generated by Algorithm 1 with adaptative or diminishing step sizes. Then,*

$$F\left(x^k + \lambda_k[p(x^k) - x^k]\right) \preccurlyeq F(x^k) + \left(\theta(x^k)\beta_k + \frac{L}{2}\|p(x^k) - x^k\|^2\beta_k^2\right)e, \tag{3.22}$$

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

Proof. By Lemma 2.3.3 with $\lambda = \lambda_k$, $x = x^k$, $p = p(x^k)$, and $\theta = \theta(x^k)$, we obtain

$$F\left(x^k + \lambda_k[p(x^k) - x^k]\right) \preceq F(x^k) + \left(\theta(x^k)\lambda_k + \frac{L}{2}\|p(x^k) - x^k\|^2\lambda_k^2\right)e. \tag{3.23}$$

For the diminishing step size where $\lambda_k = \beta_k$, inequality (3.22) follows trivially from (3.23). Now, consider that λ_k is obtained using the adaptative step size strategy. By (3.9), 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 (3.23), we obtain (3.22). ■

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

Theorem 3.3.5 *Assume that $F := (f_1, \ldots, f_m)^T$ is convex on $\mathcal{C}$. Let $(x^k)_{k\in\mathbb{N}}$ be generated by Algorithm 1 with adaptative or diminishing step sizes. Assume that $\lim_{k\to+\infty} F(x^k) = F(x^*)$. Then,, there holds*

$$\min_{\ell\in\left\{\lfloor\frac{k}{2}\rfloor+2,\cdots,k\right\}} |\theta(x^\ell)| \leq \frac{8L\operatorname{diam}(\mathcal{C})^2}{k-2}, \qquad \forall\ k \in \mathbb{N}^* - \{1, 2\},$$

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

Proof. By Lemma 3.3.4, 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 $\|p(x^k) - x^k\| \le \operatorname{diam}(\mathcal{C})$, for all $k = 0, 1, \ldots$, we have

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

which implies that

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

On the other hand, since F is convex we have $f_j(x^k) - f_j(x^*) \le -\left\langle \nabla f_j(x^k), x^* - x^k\right\rangle$, for all $j \in \mathcal{J}$. Hence, by the optimality of $p(x^k)$ in (3.7), it follows that

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

Thus, we can applying item *(ii)* of Lemma 2.1.2 with $a_k = \min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right) \le b_k = -\theta(x^k)$ and $A = L\operatorname{diam}(\mathcal{C})^2$ to obtain the desired results. ■

Corollary 3.3.6 *Assume that $F := (f_1, \ldots, f_m)^T$ is convex on $\mathcal{C}$. Let $(x^k)_{k\in\mathbb{N}}$ be generated by Algorithm 1 with adaptative or diminishing step sizes. Suppose that $\lim_{k\to+\infty} x^k = x^*$. In this case, the following bound holds*

$$\min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right) \le \frac{2L\operatorname{diam}(\mathcal{C})^2}{k}, \qquad \forall\, k \in \mathbb{N}. \tag{3.24}$$

Proof. Indeed, by using the same augment of the proof of Theorem 3.3.5, we managed to get the expression

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

using item *(ii)* of Lemma 2.1.2 with $a_k = \min_{j\in\mathcal{J}}(f_j(x^k) - f_j(x^*))$, $b_k = -\theta(x^k)$ and $A = L\operatorname{diam}(\mathcal{C})^2$ the inequality (3.24) follows. However, as in Theorem 3.2.7, it would be interesting to present a bound for $\max_{j\in\mathcal{J}}(f_j(x^k) - f_j(x^*))$. ■

3.4 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 whether it is able to generate Pareto frontiers 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/`.

3.4.1 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 [34]. The projected steepest descent direction $d_{\rm sd}(x)$ for F at $x \in \mathcal{C}$ is defined as

$$d_{\rm sd}(x) := p_{\rm sd}(x) - x, \tag{3.25}$$

where

$$p_{\rm sd}(x) := \operatorname*{argmin}_{u \in \mathcal{C}} \max_{j \in \mathcal{J}} \langle \nabla f_j(x), u - x \rangle + \frac{1}{2}\|u - x\|^2. \tag{3.26}$$

Since the minimand in (3.26) is proper, closed, and strongly convex, this problem has a unique minimizer. The optimal value of (3.26) will be denoted by $\theta_{\rm sd}(x)$, i.e.,

$$\theta_{\rm sd}(x) := \max_{j \in \mathcal{J}} \langle \nabla f_j(x), d_{\rm sd}(x) \rangle + \frac{1}{2}\|d_{\rm sd}(x)\|^2. \tag{3.27}$$

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

$$\begin{aligned} \min_{u,\tau} \quad & \tau + \tfrac{1}{2}\|u - x\|^2 \\ \text{s.t.} \quad & \langle \nabla f_j(x), u - x \rangle \leq \tau, \quad j \in \mathcal{J}, \\ & u \in \mathcal{C}, \end{aligned} \tag{3.28}$$

which is a convex quadratic problem with linear inequality constraints. In connection to Preposition 3.1.1, it is possible to show that $\theta_{\rm sd}(x) \leq 0$, $\theta_{\rm sd}(\cdot)$ is continuous, and $x \in \mathcal{C}$ is stationary if and only if $\theta_{\rm sd}(x) = 0$. All properties mentioned can be found in [34]. Essentially, the projected steepest descent method corresponds to Algorithm 1 with the search direction given as in (3.25) and using $\theta_{\rm sd}(x^k)$ 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 [54] for line search strategies in the vector optimization setting. For computing the search directions, we used the free software Algencan [11] (an augmented Lagrangian code for general nonlinear programming) to solve problems (3.4) and (3.28) for the conditional gradient and the projected steepest descent methods, respectively. All runs were stopped at an iterate x^k declaring convergence if

$$\frac{\|x^k - x^{k-1}\|_\infty}{\|x^{k-1}\|_\infty} \leq 10^{-5} \quad \text{and} \quad |\theta_{\rm sd}(x^k)| \leq 5 \times \texttt{eps}^{1/2}, \tag{3.29}$$

where $\texttt{eps} = 2^{-52} \approx 2.22 \times 10^{-16}$ is the machine precision, the first criterion in (3.29) when the denominator $\|x^{k-1}\|_\infty$ is very small (close to zero) we consider it equal to 1. Some words about this stopping criterion are in order. First, given $x \in \mathcal{C}$, the values of $\theta(x)$ in (3.5) and $\theta_{\rm sd}(x)$ in (3.27) are different, so we prefer to use only $\theta_{\rm sd}(x)$ to standardize the stopping criteria for both methods. Second, the first criterion in (3.29) seeks to detect the convergence of the sequence $(x^k)_{k\in\mathbb{N}}$, while the second guarantees to stop at an *approximately* stationary point. Third, for the projected steepest descent method, we only calculate $\theta_{\rm sd}(x^k)$ when the first criterion in (3.29) 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., $\mathcal{C} = \{x \in \mathbb{R}^n : l \leq x \leq u\}$. Table 3.2 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 3.1 are shown using performance profiles [19].

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	[40]	2	2	Y	$(-5, -5)$	$(10, 10)$
DD1[a]	[16]	5	2	N	$(-20, \dots, -20)$	$(20, \dots, 20)$
DGO1	[40]	1	2	N	-10	13
DGO2	[40]	1	2	Y	-9	9
FA1	[40]	3	3	N	$(0, 0, 0)$	$(1, 1, 1)$
Far1	[40]	2	2	N	$(-1, -1)$	$(1, 1)$
FDS	[21]	5	3	Y	$(-2, \dots, -2)$	$(2, \dots, 2)$
FF1	[40]	2	2	N	$(-1, -1)$	$(1, 1)$
Hil1	[38]	2	2	N	$(0, 0)$	$(1, 1)$
IKK1	[40]	2	3	Y	$(-50, -50)$	$(50, 50)$
IM1	[40]	2	2	N	$(1, 1)$	$(4, 2)$
JOS1	[43]	100	2	Y	$(-100, \dots, -100)$	$(100, \dots, 100)$
JOS4	[43]	100	2	N	$(-100, \dots, -100)$	$(100, \dots, 100)$
KW2	[44]	2	2	N	$(-3, -3)$	$(3, 3)$
LE1	[40]	2	2	N	$(-5, -5)$	$(10, 10)$
Lov1	[52]	2	2	Y	$(-10, -10)$	$(10, 10)$
Lov2	[52]	2	2	N	$(-0.75, -0.75)$	$(0.75, 0.75)$
Lov3	[52]	2	2	N	$(-20, -20)$	$(20, 20)$
Lov4	[52]	2	2	N	$(-20, -20)$	$(20, 20)$
Lov5	[52]	3	2	N	$(-2, -2, -2)$	$(2, 2, 2)$
Lov6	[52]	6	2	N	$(0.1, -0.16, \dots, -0.16)$	$(0.425, 0.16, \dots, 0.16)$
LTDZ	[50]	3	3	N	$(0, 0, 0)$	$(1, 1, 1)$
MGH9[b]	[58]	3	15	N	$(-2, -2, -2)$	$(2, 2, 2)$
MGH16[b]	[58]	4	5	N	$(-25, -5, -5, -1)$	$(25, 5, 5, 1)$
MGH26[b]	[58]	4	4	N	$(-1, -1, -1 - 1)$	$(1, 1, 1, 1)$
MGH33[b]	[58]	10	10	Y	$(-1, \dots, -1)$	$(1, \dots, 1)$
MHHM2	[40]	2	3	Y	$(0, 0)$	$(1, 1)$
MLF1	[40]	1	2	N	0	20
MLF2	[40]	2	2	N	$(-100, -100)$	$(100, 100)$

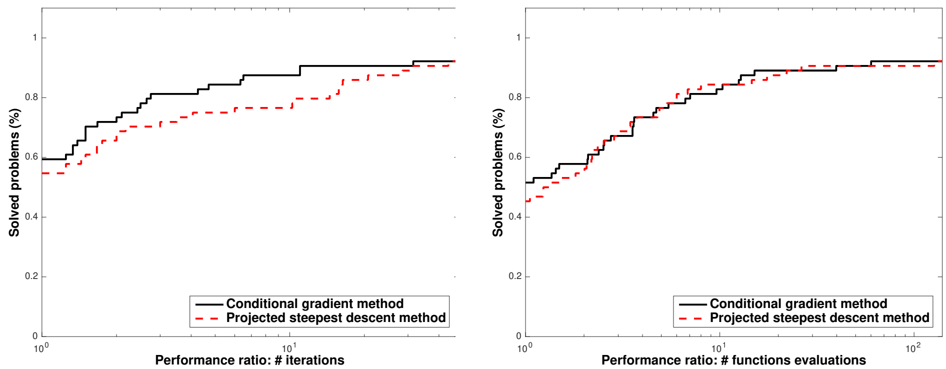
[a] This is a modified version of DD1 problem that can be found in [57].
[b] This is an adaptation of a single-objective optimization problem to the multiobjective setting that can be found in [57].

Table 3.1: List of test problems.

Problem	Source	n	m	Convex	l^T	u^T
MMR1	[56]	2	2	N	$(0.1, 0)$	$(1, 1)$
MMR3	[56]	2	2	N	$(-1, -1)$	$(1, 1)$
MMR4	[56]	3	2	N	$(0, 0, 0)$	$(4, 4, 4)$
MOP2	[40]	2	2	N	$(-4, -4)$	$(4, 4)$
MOP3	[40]	2	2	N	$(-\pi, -\pi)$	(π, π)
MOP5	[40]	2	3	N	$(-30, -30)$	$(30, 30)$
MOP6	[40]	2	2	N	$(0, 0)$	$(1, 1)$
MOP7	[40]	2	3	Y	$(-400, -400)$	$(400, 400)$
PNR	[60]	2	2	Y	$(-2, -2)$	$(2, 2)$
QV1	[40]	10	2	N	$(-5.12, \dots, -5.12)$	$(5.12, \dots, 5.12)$
SD	[64]	4	2	Y	$(1, \sqrt{2}, \sqrt{2}, 1)$	$(3, 3, 3, 3)$
SK1	[40]	1	2	N	-100	100
SK2	[40]	4	2	N	$(-10, -10, -10, -10)$	$(10, 10, 10, 10)$
SLCDT1	[63]	2	2	N	$(-1.5, -1.5)$	$(1.5, 1.5)$
SLCDT2	[63]	10	3	Y	$(-1, \dots, -1)$	$(1, \dots, 1)$
SP1	[40]	2	2	Y	$(-100, -100)$	$(100, 100)$
SSFYY2	[40]	1	2	N	-100	100
TKLY1	[40]	4	2	N	$(0.1, 0, 0, 0)$	$(1, 1, 1, 1)$
Toi4[b]	[68]	4	2	Y	$(-2, -2, -2, -2)$	$(5, 5, 5, 5)$
Toi8[b]	[68]	3	3	Y	$(-1, -1, -1, -1)$	$(1, 1, 1, 1)$
Toi9[b]	[68]	4	4	N	$(-1, -1, -1, -1)$	$(1, 1, 1, 1)$
Toi10[b]	[68]	4	3	N	$(-2, -2, -2, -2)$	$(2, 2, 2, 2)$
VU1	[40]	2	2	N	$(-3, -3)$	$(3, 3)$
VU2	[40]	2	2	Y	$(-3, -3)$	$(3, 3)$
ZDT1	[69]	30	2	Y	$(0, \dots, 0)$	$(1, \dots, 1)$
ZDT2	[69]	30	2	N	$(0.01, \dots, 0.01)$	$(1, \dots, 1)$
ZDT3	[69]	30	2	N	$(0.01, \dots, 0.01)$	$(1, \dots, 1)$
ZDT4	[69]	30	2	N	$(0.01, -5, \dots, -5)$	$(1, 5, \dots, 5)$
ZDT6	[69]	10	2	N	$(0, \dots, 0)$	$(1, \dots, 1)$
ZLT1	[40]	10	5	Y	$(-1000, \dots, -1000)$	$(1000, \dots, 1000)$

[a] This is a modified version of DD1 problem that can be found in [57].
[b] This is an adaptation of a single-objective optimization problem to the multiobjective setting that can be found in [57].

Table 3.2: List of test problems.



(a) (b)

Figure 3.1: 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 (3.29) was sufficient to detect a stationary point. This means that in these cases, $\theta_{\text{sd}}(x^k)$ was calculated only once. For the MMR1 problem, $\theta_{\text{sd}}(x^k)$ 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 (4.45) is a linear programming problem while (3.28) 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.

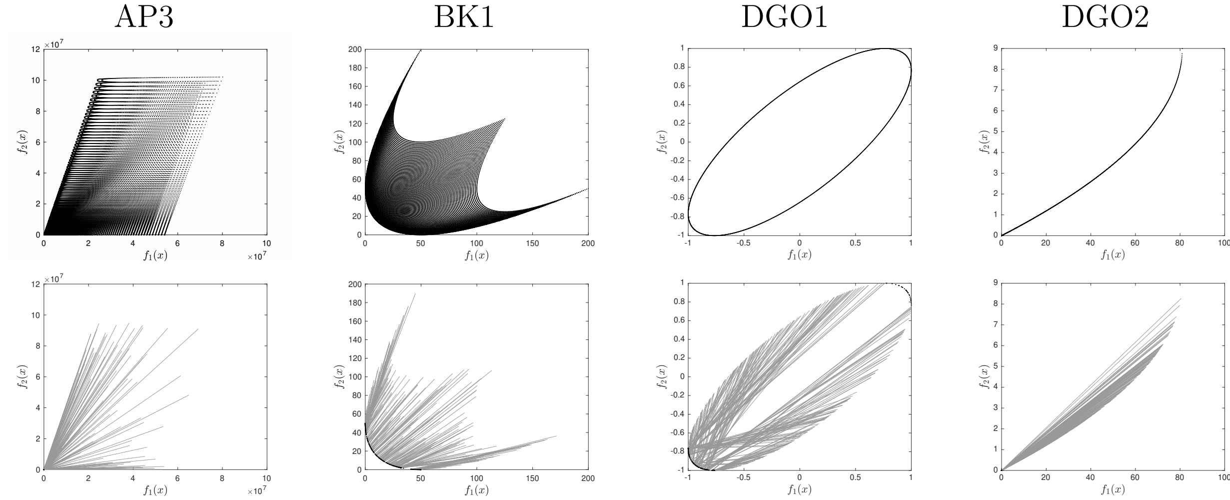
3.4.2 Pareto frontiers

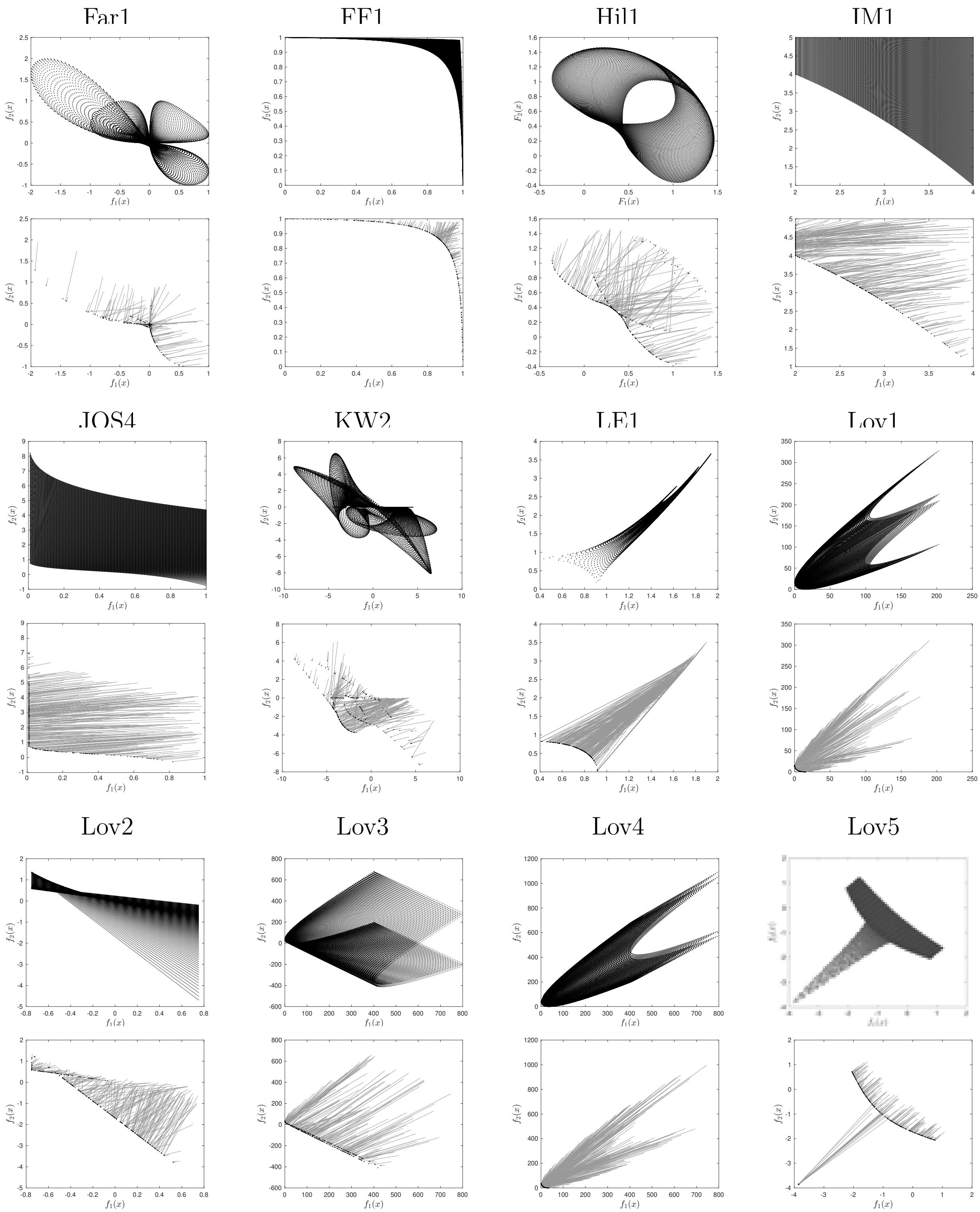
Following, we check if the conditional gradient method is able to generate Pareto frontiers properly. In the present section, we stopped the execution of the conditional gradient method algorithm at x^k declaring convergence if

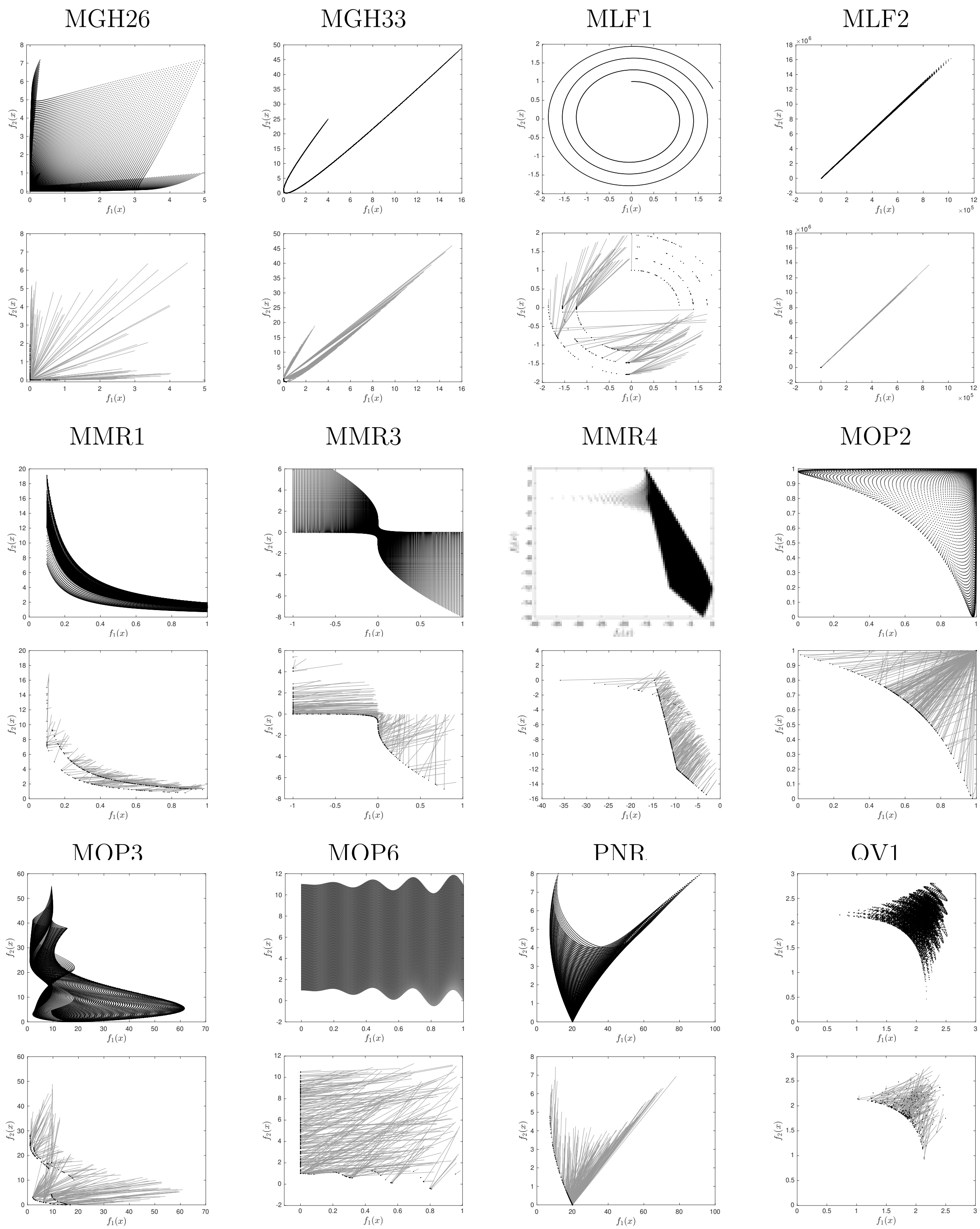
$$|\theta(x^k)| \leq 5 \times \texttt{eps}^{1/2},$$

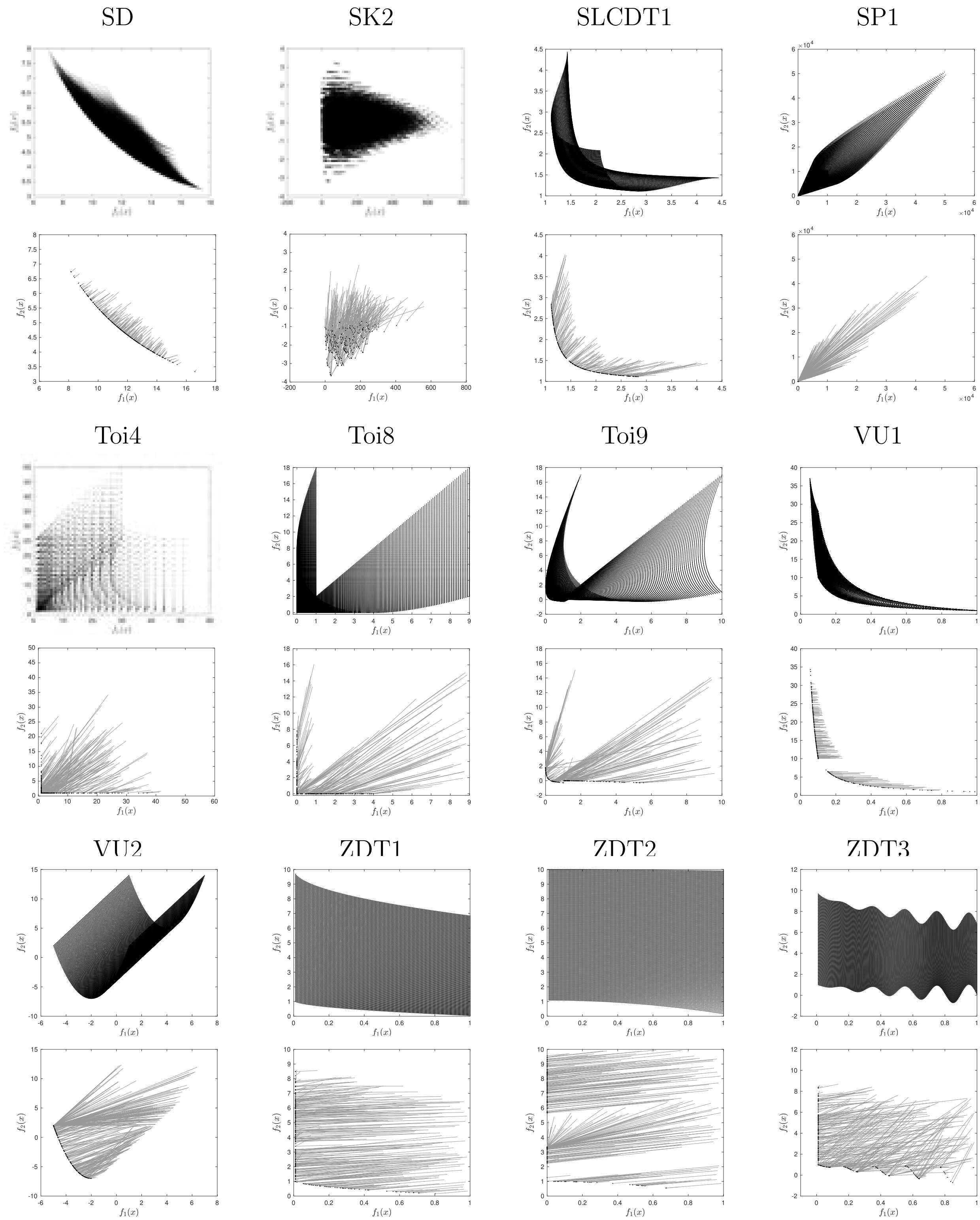
where $\theta(x^k)$ is given in (3.7) and $\texttt{eps}$ is given as in section 3.4.1. We considered all bicriteria problems in Table 3.2 for which $n = 2$, 3 or 4. We also used the problems DGO1, DGO2, and MLF1 for which $n = 1$ and the versions with $n = 2$ and $m = 2$ of the problems JOS4, MGH26, MGH33, QV1, Toi8, Toi9, ZDT1, ZDT2, ZDT3, ZDT4, and ZDT6. The results are in Figure 3.3. For each problem, there are 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 3.3 shows that for the chosen set of test problems, considering a reasonable number of starting points, the conditional gradient method was able to generate a satisfactory outline of 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).

AP3 BK1 DGO1 DGO2




Far1
FF1
Hil1
IM1
JOS4
KW2
LE1
Lov1
Lov2
Lov3
Lov4
Lov5


MGH26
MGH33
MLF1
MLF2
MMR1
MMR3
MMR4
MOP2
MOP3
MOP6
PNR
QV1
$f_1(x)$
$f_2(x)$


SD
SK2
SLCDT1
SP1
Toi4
Toi8
Toi9
VU1
VU2
ZDT1
ZDT2
ZDT3
$f_1(x)$
$f_2(x)$

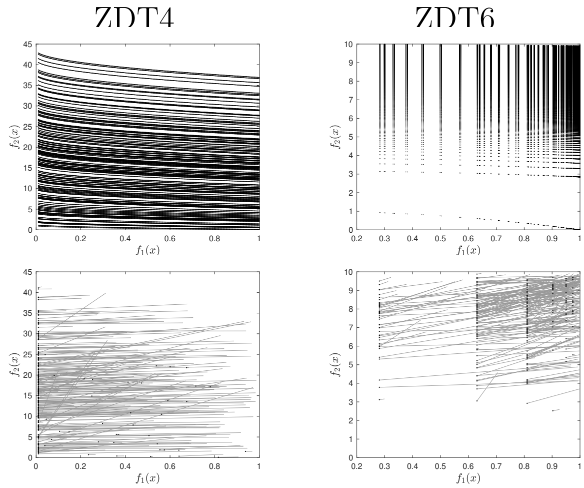


Table 3.3: Image sets and value spaces generated by the conditional gradient method using 300 starting points for each considered problem.

Chapter 4

Conditional gradient method for multiobjective composite optimization problems

This chapter is concerned with *multiobjective composite optimization problems* which is stated as follows

$$\min_{x\in\mathbb{R}^n} F(x) := G(x) + H(x). \tag{4.1}$$

Throughout the chapter we assume that H is a continuously differentiable and G is non necessarily differentiable, we will also consider six hypotheses that will be necessary to retrieve results that were studied in Chapter 3. We propose a *generalized version of the conditional gradient method* studied in the previous chapter, for solving problems (4.1). Indeed, letting $G = (g_1, \ldots, g_m) : \mathbb{R}^n \to \overline{\mathbb{R}}^m$ the indicator function of the set $\mathcal{C}$ in the multiobjective context, i.e., $g_j(x) = 0$ for all $x \in \mathcal{C}$ and $g_j(x) = +\infty$ otherwise, for all $j \in \mathcal{J}$, problem (4.1) merges into the following constrained multiobjective optimization problem

$$\min_{x\in\mathcal{C}} H(x), \tag{4.2}$$

which was studied in the previous chapter. The proposed method is also analyzed with three strategies for obtaining the step sizes, namely, Armijo-type, adaptative and diminishing step sizes. Asymptotic convergence properties and iteration-complexity bounds with and without convexity assumptions on the objective function are stablished, at the end of the chapter we present applications that were extracted from the paper [66] that satisfy the hypotheses considered for the results presented.

4.1 The multiobjetive composite optimization problem

In this section we present the assumptions on the problem (4.1), introduce the gap function associated for it, and present its main properties. Throughout our presentation we assume that $F := (f_1, \ldots, f_m)$, where $f_j := h_j + g_j$ for all $j \in \mathcal{J} := \{1, 2, \ldots, m\}$, satisfies the following three conditions:

(**A1**) The function h_j is 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**) $\mathrm{dom}(G) := \{x \in \mathbb{R}^n : \ g_j(x) < +\infty, \ j = 1, 2, \ldots, m\}$ is convex and compact.

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

$$\Omega \geq \max_{x,y \in \mathrm{dom}(G)} \|x - y\|. \tag{4.3}$$

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 dom(g_j), for all $j \in \mathcal{J}$;

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

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

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

Before presenting the method to solve problem (4.1) we need first to study a gap function associated to this problem, which will play an important role in this work. This study will be made in next section.

4.1.1 The gap function

This section is devoted to study a *gap function* $\theta(\cdot)$ associated to problem (4.1), defined by

$$\theta(x) := \min_{u \in \mathbb{R}^n} \max_{j \in \mathcal{J}} \left(g_j(u) - g_j(x) + \langle \nabla h_j(x), u - x \rangle \right), \qquad \forall\ x \in \mathrm{dom}(G). \tag{4.4}$$

As we will prove below, the gap function θ will serve as a stop criterion for the algorithm presented in next section.

Remark 4.1.1 It is worth pointing out that, in the case the components of function G are the indicator function of set $\mathcal{C}$, the gap function θ becomes the one presented in [2].

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

$$\min_{u \in \mathbb{R}^n} \max_{j \in \mathcal{J}} \left(g_j(u) - g_j(x) + \langle \nabla h_j(x), u - x \rangle \right). \tag{4.5}$$

It follows from assumptions **A1**-**A3** that problem (4.5) has solution and it belongs to dom(G). Thus, we use the notation $p(x) \in \text{dom}(G)$ when referring to a solution of problem (4.5), i.e.,

$$p(x) \in \arg\min_{u\in\mathbb{R}^n} \max_{j\in\mathcal{J}} \left(g_j(u) - g_j(x) + \langle \nabla h_j(x), u - x\rangle\right). \tag{4.6}$$

Therefore, combining (4.4) with (4.6) we conclude that

$$\theta(x) = \max_{j\in\mathcal{J}} \left(g_j(p(x)) - g_j(x) + \langle \nabla h_j(x), p(x) - x\rangle\right), \qquad \forall\, x \in \text{dom}(G). \tag{4.7}$$

To simplify the notations, for each $x \in \text{dom}(G)$ and $p(x)$ as defined (4.6), we set

$$d(x) := p(x) - x. \tag{4.8}$$

In the following proposition, which is a generalization of Proposition 3.1.1, we show that $\theta(\cdot)$ can be seeing in fact as gap function to problem (4.1).

Proposition 4.1.2 *Let $\theta : dom(G) \to \mathbb{R}$ be defined as in* (4.4). *Then the following statements hold:*

(i) $\theta(x) \leq 0$, for all $x \in dom(G)$;

(ii) $\theta(x) = 0$ if, and only if, x is a critical Pareto point of problem (4.5).

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

Proof. To prove (i), let $x \in \text{dom}(G)$. The definition of $\theta(\cdot)$ in (4.4) implies

$$\theta(x) \leq \max_{j\in\mathcal{J}} \left(g_j(u) - g_j(x) + \langle \nabla h_j(x), u - x\rangle\right), \quad \forall u \in \mathbb{R}^n. \tag{4.9}$$

Thus, letting $u = x$ in the previous inequality we conclude that $\theta(x) \leq 0$, which prove (i). To prove item (ii), we first assume that x is a critical Pareto point of problem (4.1) . If x is critical Pareto point, then

$$\max_{j\in\mathcal{J}} f_j'(x; d) \geq 0, \qquad \forall d \in \mathbb{R}^n. \tag{4.10}$$

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 (4.10) that $\max_{j\in\mathcal{J}}\{g_j'(x; d) + \langle \nabla h_j(x), d\rangle\} \geq 0$, for all $d \in \mathbb{R}^n$. Hence, using Lemma 2.2.3 together with last inequality we conclude that $\max_{j\in\mathcal{J}} \{g_j(d + x) - g_j(x) + \langle \nabla h_j(x), d\rangle\} \geq 0$, for all $d \in \mathbb{R}^n$. In particular, letting $d = p(x) - x$, we have

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

Thus, using (4.7) we conclude that $\theta(x) \geq 0$. On the other hand, we have shown in (i) that always $\theta(x) \leq 0$. Therefore, $\theta(x) = 0$. Reciprocally, now we assumed that $\theta(x) = 0$.

Our goal is to show x is a critical Pareto point of problem (4.1), or equivalently, (4.10). Using (i), $\theta(x) = 0$ if, and only if, $\theta(x) \geq 0$. Thus, using (4.7) we have

$$\begin{aligned}\theta(x) &= \min_{u\in\mathbb{R}^n} \max_{j\in\mathcal{J}} \left(g_j(u) - g_j(x) + \langle \nabla h_j(x), u - x\rangle\right)\\ &= \max_{j\in\mathcal{J}} \left(g_j(p(x)) - g_j(x) + \langle \nabla h_j(x), p(x) - x\rangle\right) \geq 0.\end{aligned}$$

Hence, we conclude that $\max_{j\in\mathcal{J}} \{g_j(u) - g_j(x) + \langle \nabla h_j(x), u - x\rangle\} \geq 0$, for all $u \in \mathbb{R}^n$. In particular, letting $u = x + \alpha d$, for $\alpha > 0$ and $d \in \mathbb{R}^n$, we conclude that

$$\max_{j\in\mathcal{J}} \Big(\frac{g_j(x+\alpha d) - g_j(x)}{\alpha} + \langle \nabla h_j(x), d\rangle\Big) \geq 0 \qquad \forall \alpha > 0,\ \forall d \in \mathbb{R}^n. \tag{4.11}$$

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 $\max_{j\in\mathcal{J}}\{g_j'(x, d) + \langle \nabla h_j(x), d\rangle\} \geq 0$, for all $d \in \mathbb{R}^n$. Therefore, (2.2) holds and x is a critical Pareto point of problem (4.1) . We proceed to prove the 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 $p(x) \in \mathrm{dom}(G)$, using (4.4) we have

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

Due to maximum function is continuous, taking the upper limit in the last inequality, we have

$$\limsup_{k\to\infty} \theta(x^k) \leq \max_{j\in\mathcal{J}} \left(g_j(p(x)) + \limsup_{k\to\infty} \left(- g_j(x^k)\right) + \left\langle \nabla h_j(x), p(x) - x\right\rangle\right). \tag{4.12}$$

On the other hand, considering that g_j lower semicontinuous in its effective domain we obtain $\limsup_{k\to\infty} \left[-g_j(x^k)\right] \leq -g_j(x)$. Therefore, combining this inequality with (4.12) and (4.7) we obtain $\limsup_{k\to\infty} \theta(x^k) \leq \theta(x)$, which concludes the proof. ■

In the following lemma we present the counterpart of [2, Lemma1] for F being a composite function as defined (4.1) . Note that we are assuming just the second component of F has coordinates with Lipschitz gradients.

Lemma 4.1.3 *Assume that h_j satisfies* **(A5)***, for all $j \in \mathcal{J}$. Let $x \in dom(G)$ and $\lambda \in [0, 1]$. Then, the following inequality holds*

$$F(x + \lambda(p(x) - x)) \preceq F(x) + \Big(\lambda\theta(x) + \frac{L}{2}\|p(x) - x\|^2\lambda^2\Big)e. \tag{4.13}$$

Proof. Since h_j has gradient Lipschitz continuous with constant L_j, $x \in \mathrm{dom}(G)$ and $\lambda \in [0, 1]$, Lemma 2.3.3 implies that

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

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 inequality, after some algebraic manipulations we obtain

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

Using (4.7) and due to $L = \max\{L_j : \ j = 1, \ldots m\}$, the last inequality becomes

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

Since the last inequality holds for all $j = 1, \ldots m$, (4.13) follows. ■

4.2 The generalized conditional gradient method

In this section, we introduce a generalization of conditional gradient method also known as Frank-Wolf algorithm for solving multiobjective composite optimization problems stated as problem (4.1), see in [3, 13, 61] the scalar version of this method. We will also study asymptotic convergence properties and iteration-complexity bounds for the generated sequence by this method. The analysis of the proposed method is made with three different step sizes, namely, Armijo-type, adaptative and diminishing step sizes. The statement of the conceptual method is stated in Algorithm 2 below.

Algorithm 2: Generalized CondG method for multiobjective optimization

---

Step 0. *Initialization*: Choose $x^0 \in \text{dom}(G)$ 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 follows

$$p(x^k) \in \arg\min_{u \in \mathbb{R}^n} \max_{j \in \mathcal{J}} \left(g_j(u) - g_j(x^k) + \left\langle \nabla h_j(x^k), u - x^k \right\rangle \right), \tag{4.14}$$

$$\theta(x^k) = \max_{j \in \mathcal{J}} \left(g_j(p(x^k)) - g_j(x^k) + \left\langle \nabla h_j(x^k), p(x^k) - x^k \right\rangle \right). \tag{4.15}$$

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

$$x^{k+1} := x^k + \lambda_k d(x^k). \tag{4.16}$$

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

Due to Proposition 4.1.2, we can assume without loss of generality that $\theta(x^k) < 0$, for all $k \in \mathbb{N}$. Because when $\theta(x^k) = 0$, the method will stop and, as a consequence of Proposition 4.1.2, a critical Pareto point of problem (4.1) is founded. In this way, from now on, we will assume that $(x^k)_{k \in \mathbb{N}}$ generated by the Algorithm 2 is an infinite sequence. We will analyze the sequence $(x^k)_{k \in \mathbb{N}}$ generated by Algorithm 2 with three step sizes. We begin presenting the Armijo-type step size. The steps below have already been presented but for the sake of completeness we re-present them

Armijo step size 2 *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* $F\left(x^k + \lambda_{k_\ell}[p(x^k) - x^k]\right) \preceq F(x^k) + \zeta\lambda_{k_\ell}\theta(x^k)e$, *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 is classical one in the analysis of conditional gradiente method, see for example [5].

Adaptative step size 2 *Assume that* $F := (f_1, \dots, f_m)$ *satisfies* (4.13) *in Lemma 4.1.3. Define the step size as*

$$\lambda_k := \min\left\{1, \frac{|\theta(x^k)|}{L\|p(x^k) - x^k\|^2}\right\} = \underset{\lambda\in(0,1]}{\operatorname{argmin}} \left(\theta(x^k)\lambda + \frac{L}{2}\|p(x^k) - x^k\|^2\lambda^2\right). \tag{4.17}$$

Since $\theta(x) < 0$ and $p(x) \neq x$ for nonstationary points, the adaptative step size is well defined. We end this section presenting the third step size, see [42].

Diminishing step size 2 *Define the step size as*

$$\lambda_k := \frac{2}{k+2}. \tag{4.18}$$

Remark 4.2.1 Let $G = (g_1, \dots, g_m) : \mathbb{R}^n \to \overline{\mathbb{R}}^m := \mathbb{R}^m \cup \{+\infty\}$ be the indicator function of the set $\mathcal{C}$ in the multiobjective context defined by $g_j(x) = 0$ for all $x \in \mathcal{C}$ and $g_j(x) = +\infty$ otherwise. Note that, for all H differentiable with JH row-wise Lipschitz continuous and G the indicator function of the set $\mathcal{C}$, we have the hypotheses (**A1**)-(**A6**) satisfied. And even more, Algorithm 2 merges into [2, Algorithm1].

4.2.1 Convergence analysis to Armijo step size

In this section we analyze the sequence $(x^k)_{k\in\mathbb{N}}$ generated by Algorithm 2 with the Armijo step size. We begin showing that the Armijo step size is well defined. First note that, assumptions (**A2**)-(**A3**) implies that $p(x^k) \in \operatorname{dom}(G)$ and $\theta(x^k)$ in (4.14) and (4.15), respectively, are well defined.

Proposition 4.2.2 *Let* $\zeta \in (0,1)$, $x^k \in dom(G)$, $p(x^k)$ *and* $\theta(x^k)$ *as in* (4.14) *and* (4.15), *respectively. Then, there exists* $0 < \bar{\eta} \leq 1$ *such that*

$$F(x^k + \eta[p(x^k) - x^k]) \preceq F(x^k) + \zeta\eta\theta(x^k)e, \qquad \forall \eta \in (0, \bar{\eta}]. \tag{4.19}$$

Proof. Since H is differentiable, G is convex and x^k, $p(x^k) \in \mathrm{dom}(G)$ we conclude that

$$F(x^k + \eta[p(x^k) - x^k]) = G(x^k + \eta[p(x^k) - x^k]) + H(x^k + \eta[p(x^k) - x^k]),$$

i.e,

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

for all $\eta \in (0, 1)$. After some arrangement in the right hand side of the last inequality we obtain

$$F(x^k + \eta[p(x^k) - x^k]) \;=\; F(x^k) + \eta\Big(JH(x^k)(p(x^k) - x^k) + G(p(x^k)) - G(x^k)\Big) + \frac{o(\eta)}{\eta}e.$$

Since $JH(x^k)(p(x^k) - x^k) \preceq \max_{j\in\mathcal{J}} \left\langle \nabla h_j(^k), p(x^k) - x^k \right\rangle e$, using (4.15) we obtain that

$$F(x^k + \eta[p(x^k) - x^k] \preceq F(x^k) + \eta\zeta\theta(x^k)e + \eta\Big((1-\zeta)\theta(x^k) + \frac{o(\eta)}{\eta}\Big)e. \tag{4.20}$$

Terefore, considering that $\theta(x^k) < 0$, $\zeta \in (0, 1)$ and the limit $\lim_{\eta\to 0} o(\eta)/\eta = 0$, there exist $\bar{\eta} > 0$ such that (4.19) holds for all $\eta \in (0, \bar{\eta}]$, and the proof is concluded. ■

In the following we present our first asymptotic convergence result for Algorithm 2 with the Armijo step size. It is worth to notting that to proof it we just need assumptions (**A1**)-(**A3**).

Theorem 4.2.3 *Every limit point $\bar{x}$ of $(x^k)_{k\in\mathbb{N}}$ is a Pareto critical point for the problem* (4.1) .

Proof. The sequence $(F(x^k))_{k\in\mathbb{N}}$ of the functional values, due to be chosen Armijo step-size and $\theta(x^k) < 0$ for all $k \in \mathbb{N}$, will satisfy the following vectorial inequality

$$0 \prec -\zeta\lambda_k\theta(x^k)e \preceq F(x^k) - F(x^{k+1}), \quad \forall k \in \mathbb{N}. \tag{4.21}$$

Consequently, $(F(x^k))_{k\in\mathbb{N}}$ is monotone decreasing. Moreover, due to $(x^k)_{k\in\mathbb{N}} \subset \mathrm{dom}(G)$ and $\mathrm{dom}(G)$ is compact, we conclude that $(F(x^k))_{k\in\mathbb{N}}$ is bounded from below. Hence $(F(x^k))_{k\in\mathbb{N}}$ converges, and we have $\lim_{k\in\mathbb{N}}[F(x^k) - F(x^{k+1})] = 0$. Thus, (4.21) implies $\lim_{k\in\mathbb{N}} \lambda_k\theta(x^k) = 0$. Let $\bar{x}$ be a limit point of the sequence $\left(x^k\right)_{k\in\mathbb{N}}$ generated by the Algorithm 2, and $\mathbb{K} \subset \mathbb{N}$ a infinite subset of indices such that $\lim_{k\in\mathbb{K}} x^k = \bar{x}$. Therefore, there exists $\mathbb{K}_1 \subset \mathbb{K}$ 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 Proposition 4.1.2 items (i), (ii) and (iii), we obtain $\theta(\bar{x}) = 0$. Thus, item item *(ii)* of Proposition 4.1.2 implies that $\bar{x}$ is Pareto critical. Now consider case $\lim_{k\in\mathbb{K}_1} \lambda_k = 0$. Suppose by contradiction that $\theta(\bar{x}) \neq 0$. Without loss of generality, assume that, $\theta(\bar{x}) < 0$. Since $\theta(\cdot)$ is upper semi-continuous, $\lim_{k\in\mathbb{K}_1} x^k = \bar{x}$, $\theta(\bar{x}) < 0$ and $\lim_{k\in\mathbb{K}_2} \lambda_k = 0$, there exists $\mathbb{K}_2 \subset \mathbb{K}_1$ a infinite subset of indices such that

$$\theta(x^k) < 0, \qquad \forall k \in \mathbb{K}_2, \tag{4.22}$$

and also $\lambda_k < 1$ for all $k \in \mathbb{K}_2$. Moreover, $(p(x^k))_{k\in\mathbb{N}} \subset \mathrm{dom}(G)$ and $\mathrm{dom}(G)$ is compact. There exist, $\mathbb{K}_3 \subset \mathbb{K}_2$ and $\bar{p} \in \mathrm{dom}(G)$ such that $\lim_{k\in\mathbb{K}_3} p(x^k) = \bar{p}$. Therefore, by the Armijo step size strategy, for each $k \in \mathbb{K}_3$, there exists $\bar{\lambda}_k \in (0, \lambda_k/\omega_2]$ such that

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

which means that

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

for at least one $j_k \in \mathcal{J}$. Since $\mathcal{J}$ is finite set of indexes and $\mathbb{K}_3$ is infinite, there exist $\mathbb{K}_4 \subset \mathbb{K}_3$ an infinite subset of indices and $j^* \in \mathcal{J}$ such that, for all $k \in \mathbb{K}_4$, we have

$$\frac{f_{j^*}(x^k + \bar{\lambda}_k[p(x^k) - x^k]t) - f_{j^*}(x^k)}{\bar{\lambda}_k} > \zeta\theta(x^k). \tag{4.23}$$

On the other hand, owing to $0 < \bar{\lambda}_k < 1$ and g_{j^*} be convex, we can apply Lemma 2.2.3 to obtain

$$\frac{g_{j^*}(x^k + \bar{\lambda}_k[p(x^k) - x^k]) - g_{j^*}(x^k)}{\bar{\lambda}_k} \leq g_{j^*}(p(x^k)) - g_{j^*}(x^k), \qquad \forall k \in \mathbb{K}_3. \tag{4.24}$$

Moreover, due to h be differentiable and $\lim_{k\in\mathbb{K}_4} \bar{\lambda}_k = 0$, we have

$$\bar{\lambda}_k\langle \nabla h_{j^*}(x^k), p(x^k) - x^k\rangle = h_{j^*}\left(x^k + \bar{\lambda}_k[p(x^k) - x^k]\right) - h_{j^*}(x^k) - o(\bar{\lambda}_k\|p(x^k) - x^k\|). \tag{4.25}$$

Combining (4.15) with (4.24) and (4.25), we obtain after some algebraic manipulation that

$$\begin{aligned} \theta(x^k) &\geq g_{j^*}(p(x^k)) - g_{j^*}(x^k) + \langle \nabla h_{j^*}(x^k), p(x^k) - x^k\rangle \\ &\geq \frac{f_{j^*}\left(x^k + \bar{\lambda}_k[p(x^k) - x^k]\right) - f_{j^*}(x^k)}{\bar{\lambda}_k} - \frac{o(\bar{\lambda}_k\|p(x^k) - x^k\|)}{\bar{\lambda}_k}. \end{aligned} \tag{4.26}$$

Hence, some calculations shows that (4.23) and (4.26) imply that

$$\frac{f_{j^*}\left(x^k + \bar{\lambda}_k[p(x^k) - x^k]\right) - f_{j^*}(x^k)}{\bar{\lambda}_k} > \left(\frac{-\zeta}{1-\zeta}\right)\frac{o(\bar{\lambda}_k\|p(x^k) - x^k\|)}{\bar{\lambda}_k}. \tag{4.27}$$

We also know from (4.22) that $\theta(x^k) < 0$, for all $k \in \mathbb{K}_4$. Therefore, considering that $\theta(\bar{x}) < 0$ and $\theta(\cdot)$ is upper semicontinuous, there exists $\delta > 0$ and $\mathbb{K}_5 \subset \mathbb{K}_4$ such that $-\delta > \theta(x^k)$, for all $k \in \mathbb{K}_5$. Thus, it follows from (4.26) that

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

Combining the last inequality with (4.27) we have

$$-\delta - \frac{o(\bar{\lambda}_k\|p(x^k) - x^k\|)}{\bar{\lambda}_k} > \left(\frac{-\zeta}{1-\zeta}\right)\frac{o(\bar{\lambda}_k\|p(x^k) - x^k\|)}{\bar{\lambda}_k}.$$

Taking the limit on the last inequality we obtain $-\delta \geq 0$, which is a contradiction with $\delta > 0$. Therefore, $\theta(\bar{x}) = 0$ and from Proposition 4.1.2 item (ii) we conclude that the point $\bar{x}$ is a critical Pareto, which concludes the proof. ■

For state the following theorem we need some notations. Since $\mathrm{dom}(G)$ is a compact set and ∇h is continuous, we set

$$\rho := \sup\{\|\nabla h_j(x)\| : \; x \in \mathrm{dom}(G), \; j = 1, \ldots, m\}. \tag{4.28}$$

Moreover, considering g_i satisfies (**A4**) we define

$$L_G := \max\{Lg_j : \; j = 1, \ldots, m\} > 0, \qquad \gamma := \min\left\{\frac{1}{(\rho + L_G)\Omega}, \; \frac{2\omega_1(1-\zeta)}{L\Omega^2}\right\}. \tag{4.29}$$

Lemma 4.2.4 *Assume that F satisfies (**A4**)-(**A5**). Then $\lambda_k \geq \gamma|\theta(x^k)| > 0$, for all $\in \mathbb{N}$.*

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 (4.15) and Proposition 4.1.2 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 $0 < -\theta(x^k) \leq g_j(x^k) - g_j(p(x^k)) + \langle \nabla h_j(x^k), x^k - p(x^k)\rangle$. Thus, Cauchy inequality together (**A4**) and (4.29) imply that

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

Using (4.28), we have $0 < -\theta(x^k) \leq (\rho + L_G)\Omega$. Hence, the definition of γ in (4.29) 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 definition of the Armijo step size λ_k, 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 4.1.3 we have

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

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 (4.3) together with the definition of γ in (4.29) we obtain that

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

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

In the following theorem we obtain our first iteration-complexity bound. For that we define

$$F^* := (f_1^*, \dots, f_m^*) := \inf\{F(x^k) : k \in \mathbb{N}\}, \qquad f_{j^*}^* := \min\{f_j^* : j \in \mathcal{J}\}. \tag{4.30}$$

Theorem 4.2.5 *Assume that* $F := (f_1, \dots, f_m)^T$ *satisfies* **(A4)**. *Then* $\lim_{k\to\infty} F(x^k) = F(x^*)$. *Moreover, there holds:*

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

ii) $\min\{|\theta(x^k)| : \; k = 0, 1, \dots, N-1\} \le \sqrt{\dfrac{f_{j^*}(x^0) - f_{j^*}^*}{\zeta\gamma N}}.$

Proof. By 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 4.2.4 we obtain

$$0 \prec \zeta\gamma|\theta(x^k)|^2 e \preceq F(x^k) - F(x^{k+1}), \tag{4.31}$$

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 exist $x^* \in \mathrm{dom}(G)$ a limit point of $(x^k)_{k\in\mathbb{N}}$. Let $(x^{k_j})_{j\in\mathbb{N}}$ be a subsequence of $(x^k)_{k\in\mathbb{N}}$ such that $\lim_{j\to\infty} x^{k_j} = x^*$. Since $(x^{k_j})_{j\in\mathbb{N}} \subset \mathrm{dom}(G)$ and G satisfies **(A4)**, we have

$$F(x^{k_j}) - F(x^*) = G(x^{k_j}) - G(x^*) + H(x^{k_j}) - H(x^*),$$

or,

$$F(x^{k_j}) - F(x^*) \le L_G\|x^{k_j} - x^*\| + \|H(x^{k_j}) - H(x^*)\|, \quad \forall j \in \mathbb{N}.$$

Considering that H is continuous and $\lim_{j\to\infty} x^{k_j} = x^*$ we conclude from the last inequality that $\lim_{j\to\infty} F(x^{k_j}) = F(x^*)$. 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 the limit on (4.31), we obtain $\lim_{k\to\infty} \theta(x^k)^2 = 0$, which implies item (i). Taking the sum of (4.31) from $k = 0$ to $k = N-1$ and using (4.30), we obtain

$$\sum_{k=0}^{N-1} |\theta(x^k)|^2 \le \frac{1}{\zeta\gamma}[f_{j^*}(x^0) - f_{j^*}^*].$$

Thus, $\min\{|\theta(x^k)|^2 : k = 0, 1, \dots, N-1\} \le [f_{j^*}(x^0) - f_{j^*}^*]/(\zeta\gamma N)$, which implies the item (ii). ■

Corollary 4.2.6 *Assume that* $F := (f_1, \dots, f_m)^T$ *satisfies* **(A4)** *and* $\varepsilon > 0$. *Define the set* $K(\varepsilon) := \{k : |\theta(x^k)| > \varepsilon, \; k \in \mathbb{N}\}$. *Then,*

$$|K(\varepsilon)| \le \frac{f_{j_*}(x^0) - f_{j_*}^*}{\zeta\gamma}\frac{1}{\varepsilon^2},$$

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

Proof. The proof follows straightforwardly from item (ii) of Theorem 4.2.5. ■

Corollary 4.2.7 *Assume that F satisfies* (**A4**)*-*(**A5**) *and $\epsilon > 0$. Consider an iteration k and let $F(x^k)$ be given. If $|\theta(x^k)| > \epsilon$, then the Armijo line search algorithm performs, at most, $1 + \ln(\gamma\epsilon)/\ln(\omega_2)$ evaluations of F to compute the step size λ_k.*

Proof. Let ℓ_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 4.2.4, it follows that $\omega_2^{\ell_k} \geq \gamma|\theta(x^k)|$. Since $|\theta(x^k)| > \epsilon$, we have $\omega_2^{\ell_k} \geq \gamma\epsilon$. Therefore, due to $0 < w_2 < 1$, we obtain $\ell_k \leq \ln(\gamma\epsilon)/\ln(\omega_2)$, concluding the proof. ■

Theorem 4.2.8 *Assume that F satisfies* (**A4**)*-*(**A5**) *and $\epsilon > 0$. Then, Algorithm 2 generates a point x^k such that $\theta(x^k) \leq \epsilon$, performing, at most,*

$$m\left[\left(1 + \frac{\ln(\gamma\epsilon)}{\ln(\omega_2)}\right)\frac{f_{j_*}(x^0) - f_{j_*}^*}{\zeta\gamma}\frac{1}{\epsilon^2} + 1\right] = \mathcal{O}(|\ln(\epsilon)|\epsilon^{-2})$$

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

$$m\left[\frac{f_{j_*}(x^0) - f_{j_*}^*}{\zeta\gamma}\frac{1}{\epsilon^2} + 1\right] = \mathcal{O}(\epsilon^{-2})$$

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

Proof. The proof follows from the combination of Corollaries 4.2.6 and 4.2.7. ■
Similar results of Corollaries 4.2.6 and 4.2.7 and Theorem 4.2.8, with respect to the scalar gradient method, were obtained in [36].

Theorem 4.2.9 *Assume that h satisfies* **(A5)-(A6)**. *And set $L := \max\{L_j : j \in \mathcal{J}\}$. Moreover assume that* $\lim_{k\to\infty} F(x^k) = F(x^*)$ *and take $\Omega > 0$ satisfying* (4.3). *Then, for all $k \in \mathbb{N}^*$, the following inequality holds*

$$\min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right) \leq \frac{1}{\zeta\gamma}\frac{1}{k}, \qquad \gamma := \min\left\{\frac{1}{(\rho + L_g)\Omega}, \frac{2\omega_1(1-\zeta)}{L\Omega^2}\right\}. \tag{4.32}$$

Proof. Indeed, since λ_k is the Armijo step size, we have

$$F(x^{k+1}) - F(x^*) \preceq F(x^k) - F(x^*) + \zeta\lambda_k\theta(x^k)e.$$

Thus, the least inequality together with Lemma 4.2.4 imply

$$\min_{j\in\mathcal{J}}\left(f_j(x^{k+1}) - f_j(x^*)\right) \leq \min_{j\in\mathcal{J}}\left(f_j(x^k) - f_j(x^*)\right) - \zeta\gamma\theta(x^k)^2. \tag{4.33}$$

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

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

for all $j \in \mathcal{J}$. Since $\big(F(x^k)\big)_{k\in\mathbb{N}}$ 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

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

Take maximum in the least inequality and using the definition of $\theta(x^k)$ in (4.15) 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 $0 \geq -\min_{j\in\mathcal{J}}\{f_j(x^k) - f_j(x^*)\} \geq \theta(x^k)$. Therefore, we obtain that

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

The combination of the last inequality with (4.33) 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 2.1.1, 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 (4.32). ■

4.2.2 Convergence analysis for adaptive and diminishing step sizes

The aim of this section is to analyze the sequence $(x^k)_{k\in\mathbb{N}}$ generated by Algorithm 2 with adaptative and diminishing step sizes. We begin showing, in particular, that if $(x^k)_{k\in\mathbb{N}}$ is generated by Algorithm 2 with adaptative step size (4.17), then $(F(x^k))_{k\in\mathbb{N}}$ is a decreasing sequence.

Lemma 4.2.10 *Take a constant $\Omega > 0$ satisfying* (4.3)*, and assume that h_j satisfies* **(A5)***, for all $j \in \mathcal{J}$. Let $\big(x^k\big)_{k\in\mathbb{N}}$ be generated by Algorithm 2 with the adaptive step size* (4.17)*. Then,*

$$F(x^{k+1}) - F(x^k) \preceq -\frac{1}{2}\min\left\{|\theta(x^k)|, \frac{\theta(x^k)^2}{L\Omega^2}\right\} e, \qquad \forall k \in \mathbb{N}. \tag{4.34}$$

As a consequence, $(F(x^k))_{k\in\mathbb{N}}$ is a nonincreasing sequence.

Proof. We will analise two possibilities to λ_k defined in (4.17). First, we assume that $\lambda_k = 1$. In this case, using (4.17), we have $L\|p(x^k) - x^k\|^2 \leq |\theta(x^k)|$.Thus, considering that $\lambda_k = 1$ and h_j satisfies **(A5)**, for all $j \in \mathcal{J}$, we apply Lemma 4.1.3 with $\lambda = 1$ and $x = x^k$ to obtain

$$F(p(x^k)) \preceq F(x^k) + \Big(\theta(x^k) + \frac{L}{2}\|p(x^k) - x^k\|^2\Big)e \preceq F(x^k) + \Big(\theta(x^k) + \frac{1}{2}|\theta(x^k)|\Big)e. \tag{4.35}$$

Due to $\lambda_k = 1$, it follows from (4.16) that $x^{k+1} = p(x^k)$. Therefore, taking into account that $|\theta(x^k)| = -\theta(x^k)$, we conclude from (4.35) that

$$F(x^{k+1}) \preceq F(x^k) - \frac{1}{2}|\theta(x^k)|. \tag{4.36}$$

Now, we assume that $\lambda_k = -\theta(x^k)/(L\|p(x^k) - x^k\|^2)$. Thus, apply Lemma 4.1.3 with $\lambda = \lambda_k$ and $x = x^k$, and considering (4.16) we obtain

$$F(x^{k+1}) \preceq F(x^k) + \left(\theta(x^k)\lambda_k + \frac{L}{2}\|p(x^k) - x^k\|^2\lambda_k^2\right) e = F(x^k) - \frac{\theta(x^k)^2}{2L\|p(x^k) - x^k\|^2} e. \tag{4.37}$$

Therefore, the combination of (4.36) with the inequality (4.37) 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 $\Omega \geq \operatorname{diam}(\operatorname{dom}(G))$, the last inequalit implies (4.34), and (4.34) is proved. The second statement of the lemma is an immediate consequence of the first one, which concludes the proof. ■

For state the next result define the following two numbers

$$f_j^* := \inf\left\{f_j(x) : x \in \operatorname{dom}(G),\ j \in \mathcal{J}\right\}, \qquad f_{j^*}(x^0) - f_{j^*}^* := \min_{j \in \mathcal{J}}\left\{f_j(x^0) - f_j^*\right\}. \tag{4.38}$$

Proposition 4.2.11 *Assume that h_j satisfies* **(A5)**, *for all $j \in \mathcal{J}$. Let $\left(x^k\right)_{k\in\mathbb{N}}$ be generated by Algorithm 2 with the adaptive step size* (4.17). *Then,*

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

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

$$\min_{k\in\{0,1,\ldots,N-1\}}\left\{|\theta(x^k)|\right\} \leq \max\left\{\frac{2}{N}\left(f_{j^*}(x^0) - f_{j^*}^*\right), \Omega\sqrt{\frac{2L}{N}\left(f_{j^*}(x^0) - f_{j^*}^*\right)}\right\}.$$

Proof. Since Lemma 4.2.10 implies that $\left(F(x^k)\right)_{k\in\mathbb{N}}$ is a decreasing sequence. Thus, for $j^* \in \mathcal{J}$ defined in (4.38), the sequence $\left(f_{j^*}(x^k)\right)_{k\in\mathbb{N}}$ is decreasing and bounded from below by f^*. Consequently, $\left(f_{j^*}(x^k)\right)_{k\in\mathbb{N}}$ converges. Moreover, Lemma 4.2.10 also implies that

$$0 < \min\left\{|\theta(x^k)|, \frac{\theta(x^k)^2}{L\Omega^2}\right\} \leq 2\left\{f_{j^*}(x^k) - f_{j^*}(x^{k+1})\right\}, \qquad \forall k \in \mathbb{N}. \tag{4.39}$$

Since $\left(f_{j^*}(x^k)\right)_{k\in\mathbb{N}}$ converges, we have $\lim_{k\to+\infty}(f_{j^*}(x^k) - f_{j^*}(x^{k+1})) = 0$. Thus, taking limits in (4.39) we obtain (a). Next we proceed to prove (b). By summing both sides in (4.39), for $k = 0, 1, \ldots, N-1$, and taking into account (4.38) we obtain

$$\sum_{k=0}^{N-1}\min\left\{|\theta(x^k)|, \frac{\theta(x^k)^2}{L\Omega^2}\right\} \leq 2\left(f_{j^*}(x^0) - f_{j^*}^*\right).$$

Therefore, we have

$$\min\left\{\min\left\{|\theta(x^k)|, \frac{\theta(x^k)^2}{L\Omega^2}\right\}:\ k=0,1,\ldots,N-1\right\} \leq 2\left(f_{j^*}(x^0)-f^*_{j^*}\right),$$

which implies the statement of item (b). ■

Theorem 4.2.12 *Assume that h_j satisfies* **(A5)**, *for all $j \in \mathcal{J}$. Let $\left(x^k\right)_{k\in\mathbb{N}}$ be generated by Algorithm 2 with the adaptive step. Then, every limit point of $\left(x^k\right)_{k\in\mathbb{N}}$ is a critical Pareto point of the problem* (4.1) .

Proof. Let $\bar{x}$ be a limit point of the sequence $\left(x^k\right)_{k\in\mathbb{N}}$, and $\mathbb{K}\subset\mathbb{N}$ a infinite subset of indices such that $\lim_{k\in\mathbb{K}} x^k = \bar{x}$. Since $x^k \in \mathrm{dom}(G)$, for all $k\in\mathbb{N}$, and due to $\mathrm{dom}(G)$ be compact, we conclude that $\bar{x}\in\mathrm{dom}(G)$. On the other hand, item (a) of Proposition 4.2.11 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 item (iii) of Proposition 4.1.2 that $0\leq\theta(\bar{x})$. Thus, owing to $\bar{x}\in\mathrm{dom}(G)$, item (i) of Proposition 4.1.2 implies $\theta(\bar{x})=0$. Therefore, applying item (ii) of Proposition 4.1.2 we obtain that $\bar{x}$ is a critical Pareto point of the problem (4.1) . ■

Theorem 4.2.13 *Assume that h_j satisfies***(A5)**-**(A6)**, *for all $j \in \mathcal{J}$. Let $\left(x^k\right)_{k\in\mathbb{N}}$ generated by Algorithm 2 with λ_k satisfying adaptive or diminishing step size, i.e.,* (4.17) *or* (4.18). *Assume that $\lim_{k\to\infty} F(x^k)=F(x^*)$. Then,*

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

Proof. Applying Lemma 4.1.3 with $x=x^k$ we have

$$F(x^k+\lambda(p(x^k)-x^k)) \preceq F(x^k)+\Big(\lambda\theta(x^k)+\frac{L}{2}\|p(x^k)-x^k\|^2\lambda^2\Big)e, \qquad \forall\ \lambda\in[0,1]. \tag{4.40}$$

Considering that the diminishing step size defined (4.18) belongs to $[0,1]$, it follows from (4.40) that for λ_k satisfying (4.17) or (4.18) there holds

$$F(x^{k+1}) \preceq F(x^k)+\Big(\theta(x^k)\lambda_k+\frac{L}{2}\|p(x^k)-x^k\|^2\lambda_k^2\Big)e, \qquad \forall n\in\mathbb{N}. \tag{4.41}$$

Hence, taking into account that $\|p(x^k)-x^k\|\leq\Omega$ we conclude that

$$\min_{j\in\mathcal{J}}\left(f_j(x^{k+1})-f_j(x^*)\right) \leq \min_{j\in\mathcal{J}}\left(f_j(x^k)-f_j(x^*)\right)+\theta(x^k)\lambda_k+\frac{L}{2}\Omega^2\lambda_k^2, \qquad \forall n\in\mathbb{N}. \tag{4.42}$$

On the other hand, from Lemma 4.2.10 we obtain that $\left(F(x^k)\right)_{k\in\mathbb{N}}$ is nonincreasing. Thus, due to $\lim_{k\to\infty} F(x^k)=F(x^*)$, using **(A6)** we have

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

Thus, taking the maximum and using the optimality of $p(x^k)$ in (4.14) yields

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

which implies $0 \leq \min_{j\in\mathcal{J}}\{f_j(x^k) - f_j(x^*)\} \leq |\theta(x^k)|$. Therefore, using (4.42), we can apply the item (ii) of Lemma 2.1.2 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 inequality. ■

Corollary 4.2.14 *Assume that h satisfies* **(A5)**-**(A6)**. *Let $\left(x^k\right)_{k\in\mathbb{N}}$ generated by Algorithm 2 with λ_k satisfying adaptive or diminishing step size, i.e.,* (4.17) *or* (4.18). *Suppose that* $\lim_{k\to+\infty} x^k = x^*$. *Then, the following bound holds*

$$\min_{j\in\mathcal{J}} \left(f_j(x^k) - f_j(x^*)\right) \leq \frac{2L\Omega^2}{k}, \qquad k \in \mathbb{N}. \tag{4.43}$$

Proof. Indeed, by using the same augment of the proof of Theorem 4.2.13 and the item (i) of Lemma 2.1.2 the inequality (4.43) follows. ■

4.3 Applications

In this section we present some example of multiobjetive composite optimization problems (4.1) satisfying the assumptions (**A1**)-(**A6**). It is worth to noting that to implement the generalized conditional gradient method presented in Algorithm 2 we need to solve in each iteration a subproblem in the following form

$$\min_{u\in\mathbb{R}^n} \max_{j\in\mathcal{J}} \left(g_j(u) - g_j(\bar{x}) + \langle \nabla h_j(\bar{x}), u - \bar{x}\rangle\right), \tag{4.44}$$

which is a non-differentiable problem. Below we will present some special class of problem (4.44) that can be rewritten as a differentiable problem whose solutions are also solutions (4.44) and can be solved efficiently. In order to present the examples, we first note that a solution of (4.44) can be calculated by solving for $\tau \in \mathbb{R}^n$ and $u \in \mathbb{R}^n$ the following constrained problem

$$\begin{aligned} \min_{u,\tau} \quad & \tau \\ \text{s.t.} \quad & g_j(u) - g_j(\bar{x}) + \langle \nabla h_j(\bar{x}), u - \bar{x}\rangle \leq \tau, \quad j \in \mathcal{J}, \\ & u \in \mathbb{R}^n. \end{aligned} \tag{4.45}$$

Because g_j is generally not differentiable, the inequality in (4.45) is difficult to deal with. Hence, to address the examples, for each $j \in \mathcal{J}$, we consider functions $g_j : \mathbb{R}^n \to \overline{\mathbb{R}}^m$ given by

$$g_j(x) := \max_{z\in\mathcal{Z}_j} \hat{g}_j(x, z), \tag{4.46}$$

where $\mathcal{Z}_j \subset \mathbb{R}^n$ and $\hat{g}_j : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is convex with respect to the first coordinate, for all $j \in \mathcal{J}$. Multiobjetive composite optimization problems (4.1) with g_j given by (4.46) include as a particular instance robust multiobjective optimization problem, see for example [20, 25, 41].

Remark 4.3.1 It is worth to notting that, even though the function $\hat{g}_j(\cdot, z)$ in (4.46) is differentiable for each $z \in \mathcal{Z}_j$, we cannot guarantee that g_j will be a differentiable function. On the other hand, due to the function $\hat{g}_j(\cdot, z)$ be convex for each $z \in \mathcal{Z}_j$, Proposition 2.2.2 implies that g_j is also convex.

In the following we present some special examples of (4.45) and (4.46) that has appeared [66].

Example 4.3.2 Let $\mathcal{C} \subset \mathbb{R}^n$ be a convex and compact set and $\mathcal{Z}_j = \{z \in \mathbb{R}^n : \ A_j z \leq \delta\}$, where A_j is a $p \times n$ nonsingular matrix and $j \in \mathcal{J}$. Define $\hat{g}_j : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by $\hat{g}_j(x, z) = \langle x, z\rangle$ and assume that $\mathcal{Z}_j$ is nonempty and bounded. In this case, $g_j : \mathcal{C} \to \mathbb{R}$ is given by $g_j(x) = \max_{z \in \mathcal{Z}_j} \langle x, z\rangle$. Note that g_j satisfies (**A2**)-(**A4**). Indeed, since $\hat{g}_j(\cdot, z)$ is convex for each $z \in \mathcal{Z}_j$ it follows from Proposition 2.2.2 that g_j is also convex, which implies that it satisfies (**A2**). Considering that $\mathcal{C} \subset \mathbb{R}^n$ is a convex and compact set, we obtain that g_j satisfies (**A3**). We also knows that

$$|g_j(x) - g_j(y)| = |\max_{z \in \mathcal{Z}_j} \langle x, z\rangle - \max_{z \in \mathcal{Z}_j} \langle y, z\rangle| \leq \max_{z \in \mathcal{Z}_j} |\langle x, z\rangle - \langle y, z\rangle| \leq (\max_{z \in \mathcal{Z}_j} \|z\|)\|x - y\|,$$

for all $x, y \in \mathcal{C}$. Thus, due to $\mathcal{Z}_j$ be a compact set we conclude that $L_{g_j} := \max_{z \in \mathcal{Z}_j} \|z\| < +\infty$. Hence, g_j satisfies (**A4**). Finally, let $h_j : \mathbb{R}^n \to \mathbb{R}$ be a convex and differentiable function with ∇h_j being Lipschitz continuous, for instance $h_j(x) := \langle Q_j x, x\rangle$ with Q_j a positive definite $n \times n$ matrix. Therefore, we conclude that g_j and h_j satisfy (**A1**)-(**A6**), for all $j \in \mathcal{J}$. In this case, by using duality theory a solution of the (4.45) can be computed by solving the following equivalent linear problem

$$\begin{aligned} \min_{w_j, u, \tau} \ & \tau \\ \text{s.t.} \quad & \langle \delta_j, w_j\rangle - g_j(\bar{x}) + \langle \nabla h_j(\bar{x}), u - \bar{x})\rangle \leq \tau, \quad j \in \mathcal{J}, \\ & A_j^T w_j - u = 0, \quad j \in \mathcal{J}, \\ & w_j \geq 0, \quad j \in \mathcal{J}. \end{aligned}$$

for details see [66, Section 5.2 (a)].

Example 4.3.3 Let $\mathcal{C} \subset \mathbb{R}^n$ be a convex and compact set and $\mathcal{Z}_j = \{a_j + P_j z \in \mathbb{R}^n : \ \|z\| \leq 1, \ z \in \mathbb{R}^n\}$, where $a_j \in \mathbb{R}^n$, P_j is a $n \times n$ nonsingular matrix and $j \in \mathcal{J}$. Define $\hat{g}_j : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by $\hat{g}_j(x, z) = \langle x, z\rangle$. In this case, the function $g_j : \mathcal{C} \to \mathbb{R}$ is given by $g_j(x) := \langle a_j, x\rangle + \|P_j^T x\|$, for detains see [66, 5.2 (b)]. Note that g_j satisfies (**A2**)-(**A4**). Indeed, it easy to see g_j is convex, which implies that it satisfies (**A2**). Considering that $\mathcal{C} \subset \mathbb{R}^n$ is a convex and compact set, we obtain that g_j satisfies (**A3**). Now note that

$$|g_j(x) - g_j(y)| = |\langle a_j, x\rangle + \|P_j^T x\| - \langle a_j, y\rangle - \|P_j^T y\|| \leq |\langle a_j, x - y\rangle| + |\|P_j^T x\| - \|P_j^T y\||,$$

for all $x, y \in \mathcal{C}$. On the other hand, considering that $|\|P_j^T x\| - \|P_j^T y\|| \leq \|P_j^T x - P_j^T y\|$, we have $|\|P_j^T x\| - \|P_j^T y\|| \leq \|P_j^T\|\|x - y\|$. Thus

$$|g_j(x) - g_j(y)| \leq (\|a_j\| + \|P_j^T\|)\|x - y\|, \qquad \forall x, y \in \mathcal{C}.$$

Hence, the function g_j satisfies (**A4**) with $L_{g_j} := \|a_j\| + \|P_j^T\|$. Finally, let $h_j : \mathbb{R}^n \to \mathbb{R}$ be a convex and differentiable function with ∇h_j being Lipschitz continuous, for instance $h_j(x) := \langle Q_j x, x\rangle$ with Q_j a positive definite $n \times n$ matrix. Therefore, we conclude that the functions g_j and h_j satisfy (**A1**)-(**A6**), for all $j \in \mathcal{J}$. Finally, due to $g_j(x) := \langle a_j, x\rangle + \|P_j^T x\|$ for all $j \in \mathcal{J}$, the problem (4.45) becomes to the following problem

$$\begin{aligned} &\min_{u,\tau} \tau \\ &\text{s.t.} \quad \langle a_j, u - \bar{x}\rangle + \|P_j^T u\| - \|P_j^T \bar{x}\| + \langle \nabla h_j(\bar{x}), u - \bar{x})\rangle \leq \tau, \quad j \in \mathcal{J} \\ &\qquad\qquad u \in \mathbb{R}^n. \end{aligned}$$

for details see [66, Section 5.2 (a)].

Example 4.3.4 Let $\mathcal{C} \subset \mathbb{R}^n$ be a convex and compact set and $\mathcal{Z}_j = \{a_j + P_j z \in \mathbb{R}^n : \ \|z\| \leq 1, \ z \in \mathbb{R}^n\}$, where $a_j \in \mathbb{R}^n$, P_j is a $n \times n$ nonsingular matrix and $j \in \mathcal{J}$. Define $\hat{g}_j : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by $\hat{g}_j(x, z) = \langle x + z, A_j(x + z)\rangle$, where $A_j := M_j^T M_j$ and M_j is a $n \times n$ nonsingular matrix. In this case, the function $g_j : \mathcal{C} \to \mathbb{R}$ is given by

$$g_j(x) := \max_{\|z\|\leq 1} \big\langle x + a_j + P_j z, A_j(x + a_j + P_j z)\big\rangle, \tag{4.47}$$

for detains see [66, 5.2 (c)]. Note that g_j satisfies (**A2**)-(**A4**). Indeed, first note that the definition of g_j in (4.47) and Proposition 2.2.2 implies that g_j is convex. Considering that $\mathcal{C} \subset \mathbb{R}^n$ is a convex and compact set, we obtain that g_j satisfies (**A3**). Since $A_j := M_j M_j^T$, it follows from (4.47) that $g_j(x) := \max_{\|z\|\leq 1} \|M_j(x + a_j + P_j z)\|^2$. Thus, we have

$$|g_j(x) - g_j(y)| = |\max_{\|z\|\leq 1} \|M_j(x + a_j + P_j z)\|^2 - \max_{\|z\|\leq 1} \|M_j(y + a_j + P_j z)\|^2|,$$

for all $x, y \in \mathcal{C}$. Hence, we conclude that

$$|g_j(x) - g_j(y)| = \max_{\|z\|\leq 1} \left|\|M_j(x + a_j + P_j z)\|^2 - \|M_j(y + a_j + P_j z)\|^2\right|.$$

for all $x, y \in \mathcal{C}$. On the other hand, by using the mean value theorem, see for example [49, Corollary4.3 pag.379], we conclude that

$$\left|\|M_j(x + a_j + P_j z)\|^2 - \|M_j(y + a_j + P_j z)\|^2\right| \leq (\max_{x\in\mathcal{C}} \|M_j(x + a_j + P_j z)\|)\|x - y\|,$$

for all $x, y \in \mathcal{C}$. Therefore, the combination of the two previous inequalities yields

$$|g_j(x) - g_j(y)| = (\max_{\|z\|\leq 1} \max_{x\in\mathcal{C}} \|M_j(x + a_j + P_j z)\|)\|x - y\|, \qquad \forall x, y \in \mathcal{C}.$$

Hence, g_j satisfies (**A4**) with $L_{g_j} := \max_{\|z\|\leq 1} \max_{x\in\mathcal{C}} \|M_j(x+a_j+P_j z)\| < +\infty$. Finally, let $h_j : \mathbb{R}^n \to \mathbb{R}$ be a convex and differentiable function with ∇h_j being Lipschitz continuous, for instance $h_j(x) := \langle Q_j x, x\rangle$ with Q_j a positive definite $n \times n$ matrix. Therefore, we conclude that g_j and h_j satisfy (**A1**)-(**A6**), for all $j \in \mathcal{J}$. The subproblem (4.14) for this example follows the same idea of the one in [66, Section 5.2 (c)].

We end this chapter remarking that the development of this thesis, occurred somewhat non-chronological. The differentiable case from chapter 3 is actually a special case of the composite case presented in this chapter. During the development of the former, we did not achieve much progress due to lack of maturity. However, as the differentiable case is working harmoniously, we decided to leave it in its original form, postponing the composite part. Now, with the composite results established, it becomes clear, that the former is merely a special case of the latter.

Chapter 5

Final remarks

This work 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 generate a satisfactory outline of the Pareto frontiers of several convex and nonconvex problems. In the scalar-valued optimization, it is well known that under convexity of the objective function, the functional values of the sequence generated by the conditional gradient method converges with rate rate of $1/k^2$, see for example [31]. It would be interesting to extend this result to multiobjective conditional gradient method. It was also generalized conditional gradient method for multiobjective composite optimization problems. And established results on the asymptotic behavior and iteration-complexity bounds for the sequence generated by it. Our analysis was also carried out with and without convexity and Lipschitz assumptions on the first function defining the objective functions and considering different strategies for the step sizes. Finally, a natural question is whether it is possible to analyze the CondG method for vector optimization problems as well as to vector composite optimization problems, i.e., when the partial order in $\mathbb{R}^m$ is induced by other underlying cones instead of the non-negative orthant.

Bibliography

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

[2] P. Assunção, O. Ferreira, and L. Prudente. Conditional gradient method for multiobjective optimization. *Computational Optimization and Applications*, 78(3):741–768, 2021.

[3] 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.

[4] A. Beck. *First-Order Methods in Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017.

[5] 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.

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

[7] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2009.

[8] 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.

[9] G. C. Bento, J. C. Neto, P. R. Oliveira, and A. Soubeyran. The self regulation problem as an inexact steepest descent method for multicriteria optimization. *European Journal of Operational Research*, 235(3):494–502, 2014.

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

[11] E. G. Birgin and J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization.* Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

[17] G. de Carvalho Bento, G. B. Allende, and Y. R. L. Pereira. A newton-like method for variable order vector optimization problems. *Journal of Optimization Theory and Applications*, 177(1):201–221, 2018.

[18] 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.

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

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

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

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

[23] 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.

[24] 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.

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

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

[27] 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.

[28] 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.

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

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

[31] 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.

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

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

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

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

[36] 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.

[37] 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.

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

[39] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis*. Grundlehren Text Editions. Springer-Verlag, Berlin, 2001. Abridged version of ıt Convex analysis and minimization algorithms. I [Springer, Berlin, 1993; MR1261420 (95m:90001)] and ıt II [ibid.; MR1295240 (95m:90002)].

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

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

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

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

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

[49] S. Lang. *Undergraduate analysis*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1983.

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

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

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

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

[54] 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.

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

[56] 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.

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

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

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

[60] 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.

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

[62] W. Rudin et al. *Principles of mathematical analysis*, volume 3. McGraw-hill New York, 1964.

[63] 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.

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

[65] 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.

[66] 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.

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

[68] 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.

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