Abstract :

This work deals with scalar and vector minimax fractional programs whose objective functions are the maximum of the quotients of difference of convex (DC) functions. These problems are generally nonsmooth and nonconvex.

We give optimality conditions and develop algorithms to find a solution to such problems. We begin our study by the particular generalized fractional programming problems with ratios of convex functions, and convex constraints. We then consider the more general case of minimax fractional programs with ratios of DC functions, and DC constraints. Optimality conditions and algorithms are also developed for vector fractional programs with ratios of DC functions, and DC constraints. For such scalar and vector problems, Dinkelbach-type algorithms fail to work since the parametric subproblems may be nonconvex, whereas the latter need a global optimal solution of these subproblems. To overcome this difficulty, we overestimate the objective function in these subproblems by a convex function, and the constraints set by an inner convex subset of the latter, which leads to convex subproblems. We establish optimality conditions of Karush-Kuhn-Tucker type for these various problems, and show that our algorithms can find points that satisfy these necessary optimality conditions.

Finally, we give some numerical tests on various problems to evaluate the efficiency of the proposed algorithms.

keywords:
Fractional programming, Quotient of convex functions, Difference of convex functions, Convex programming, Optimality conditions, Proximal point methods, Bundle methods, Pareto optimality, Multiobjective programming, Dinkelbach algorithms.

N° d'ordre

Difference of Convex Functions Programming Approaches for Minimax Fractional Optimization Problems: Optimality Conditions and Resolution Algorithms

Année Initiale: 2017/2018 Mathématiques, Informatique et Applications

Université Hassan 1er
Centre d'Études Doctorales en Sciences et Techniques & Sciences Médicales

Faculté des Sciences et Techniques
Settat

THÈSE DE DOCTORAT

Pour l'obtention de grade de Docteur en ***Mathématiques Appliquées***

Formation Doctorale: Mathématiques, Informatique et Applications

Spécialité: Mathématiques Appliquées

Sous le thème

Difference of Convex Functions Programming Approaches for Minimax Fractional Optimization Problems: Optimality Conditions and Resolution Algorithms

Présentée par :

Abdelouafi GHAZI

Soutenue le: *28/06/2022.*
A la Faculté des Sciences et Techniques de Settat devant le jury composé de :

Pr. Abdelkarim HAJJAJ	**PES**	**FST Settat**	Président
Pr. Abdelmalek ABOUSSOROR	**PES**	**ENSA Marrakech**	Rapporteur
Pr. Mohammed ALAOUI	**PES**	**FST Settat**	Rapporteur
Pr. Rachid EL JID	**PH**	**FST Settat**	Rapporteur
Pr. Mohamed NAIMI	**PH**	**ENSA Berrechid**	Examinateur
Pr. Ahmed ROUBI	**PES**	**FST Settat**	Directeur de thèse

Année Universitaire: 2021/2022

I would like to dedicate this thesis to my loving parents and all my family.

Acknowledgements

Au terme de ce travail, j'exprime mes plus vifs remerciements à mon directeur de thèse **Pr. Ahmed ROUBI**, pour avoir accepté de diriger ma thèse, qui m'a été d'une aide précieuse et qui m'a consacré du temps. J'ai pu bénificier de son expertise scientifique et de ses éclairages complémentaires, mais également de son écoute et de sa gentillesse, il a fait preuve d'une grande patience et d'un esprit responsable. Ainsi que pour sa disponibilité, les nombreux encouragements qu'il m'a prodiguée et son soutien qui m'ont permis d'effectuer ma thèse dans les bonnes conditions. J'apprécie ses conseils indispensables à la progression de ce travail de recherche.

J'adresse tous mes remerciements à **Pr. Abdelmalek Aboussoror**, **Pr. Mohammed Alaoui** et **Pr. Rachid El Jid** pour l'intérêt qu'ils ont porté à cette thèse me faisant l'honneur d'en être les rapporteurs.

Je tiens à remercier **Pr. Abdelkarim Hajjaj** pour l'honneur qu'il me fait en acceptant de présider le jury de cette thèse. Je voudrais exprimer ma profonde gratitude à **Pr. Mohamed Naimi** d'avoir accepté d'examiner ce travail de recherche. Je remercie également **Pr. Mostafa El Haffari** pour l'honneur qu'il me fait d'être dans mon jury de thèse.

J'adresse mes sincères remerciements à tous les professeurs du département de mathématiques et informatique que j'ai eu le plaisir de côtoyer durant mes recherches. Je remercie particulièrement tous ceux avec lesquels j'ai le plus eu l'occasion de travailler et d'échanger. Je voudrais exprimer ma reconnaissance envers les collègues du laboratoire MISI pour leur sympathie.

Mes derniers remerciements vont à mes parents, ma grande mère, ma soeur, mon frère, mes tantes et mes oncles.

Enfin, je remercie mes amis qui ont toujours été là pour moi. Leur soutien inconditionnel et leurs encouragements ont été d'une grande aide.

Je dédis ce travail à toute ma famille.

Abstract

This work deals with scalar and vector minimax fractional programs whose objective functions are the maximum of the quotients of difference of convex (DC) functions. These problems are generally nonsmooth and nonconvex.

We give optimality conditions and develop algorithms to find a solution to such problems. We begin our study by the particular generalized fractional programming problems with ratios of convex functions, and convex constraints. We then consider the more general case of minimax fractional programs with ratios of DC functions, and DC constraints. Optimality conditions and algorithms are also developed for vector fractional programs with ratios of DC functions, and DC constraints. For such scalar and vector problems, Dinkelbach-type algorithms fail to work since the parametric subproblems may be nonconvex, whereas the latter need a global optimal solution of these subproblems. To overcome this difficulty, we overestimate the objective function in these subproblems by a convex function, and the constraints set by an inner convex subset of the latter, which leads to convex subproblems. We establish optimality conditions of Karush-Kuhn-Tucker type for these various problems, and show that our algorithms can find points that satisfy these necessary optimality conditions. Finally, we give some numerical tests on various problems to evaluate the efficiency of the proposed algorithms.

keywords:
Fractional programming, Quotient of convex functions, Difference of convex functions, Convex programming, Optimality conditions, Proximal point methods, Bundle methods, Pareto optimality, Multiobjective programming, Dinkelbach algorithms.

Résumé

Ce travail traite des programmes fractionnaires minimax, scalaires et vectoriels, dont les fonctions objectifs sont le maximum de quotients de différence de fonctions convexes. Ces problèmes sont généralement non lisses et non convexes.

Nous donnons des conditions d'optimalité et développons des algorithmes pour trouver une solution à ces problèmes. On commence par des problèmes de programmation fractionnaire généralisée particuliers avec des rapports de fonctions convexes, et des contraintes convexes. On considère ensuite le cas général de programmes minimax fractionnaires avec des rapports de différence de fonctions convexes et des contraintes également différence de fonctions convexes. On établit également des conditions d'optimalité, et on propose un algorithme pour les programmes minimax fractionnaires vectoriels. Les algorithmes de type Dinkelbach ne peuvent fonctionner puisque les sous-problèmes paramétriques peuvent être non convexes, alors que ces derniers nécessitent une solution optimale globale de ces sous-problèmes. Pour surmonter cette difficulté, nous surestimons la fonction objective dans ces sous-problèmes par une fonction convexe, et l'ensemble des contraintes par un sous-ensemble convexe de ce dernier, ce qui conduit à des sous-problèmes convexes. Nous établissons des conditions d'optimalité nécessaires de type Karush-Kuhn-Tucker pour ces divers problemes, et montrons que ces algorithmes peuvent trouver des points qui satisfont ces conditions. Finalement, nous donnons quelques tests numériques sur différents problemes pour évaluer l'efficacité des algorithmes proposés.

Mots clés:
Programmation fractionnaire, Quotient de fonctions convexes, Différence de fonctions convexes, Programmation convexe, Conditions d'optimalité, Méthodes du point proximal, Optimalité de Pareto, Programmation multi-objective, Algorithmes de Dinkelbach.

Publications

- Ghazi, A., Roubi, A. (2020). A DC approach for minimax fractional optimization programs with ratios of convex functions. *Optimization Methods and Software*, 1-19.
- Ghazi, A., Roubi, A. (2021). Optimality conditions and DC-Dinkelbach-type algorithm for generalized fractional programs with ratios of difference of convex functions. *Optimization Letters*, 15(7), 2351-2375.
- Ghazi, A., Roubi, A. Optimality conditions and DC-Dinkelbach-type algorithm for vector fractional programming problems with ratios of difference of convex functions. *(in preparation)*.

Contents

General Introduction

A minimax or generalized fractional programming problem (GFP) is an optimization program whose objective function is the maximum of finite ratios of functions to be minimized on a feasible set of $\mathbb{R}^n$. More precisely, GFP takes the form

$$(P) \qquad \bar{\lambda} = \inf_{x \in X} \left\{ \lambda(x) := \max_{i \in I} \frac{f_i(x)}{g_i(x)} \right\},$$

where X is a nonempty subset of $\mathbb{R}^n$, $I = \{1, 2, \ldots, m\}$ is a discrete set of m elements, f_i, g_i, for $i \in I$ are real-valued functions defined and continuous on X. The functions g_i, for $i \in I$, are assumed to be positive on X. For $m = 1$, the problem (P) corresponds to tranditional fractional programming and has been studied by several authors [40, 47, 111, 112].

Problems of this type arise in many fields of applications, for the single ratio problem, i.e. $m = 1$, the problem is then to minimize, for instance, cost-to-time, risk-to-return, investment-on-return, etc. An example in [32], illustrates how to determine the best combination to load the vessel taking into account of the loading time as follows

$$\max_{x \in \mathbb{R}^n_+} \frac{\sum_i (p_i - C_1 k_i) x_i - C_2 T}{\sum_i k_i x_i + T},$$

where x_i is the unit of cargo type i to load the ship, k_i the loading time for cargo i, p_i is the unit profit for cargo type i, C_1 and C_2 is the cost per unit time at respectively ports and sea, and T the journey time. The objective is to optimize the efficiency measure "cost-to-time" rather than to seek the minimum cost only.

For $m > 1$, the earliest application is the von Neumann's model for an expanding economy [97], where the output-input ratios are efficiency measures to be optimized under max-min criterion. Then, the growth rate is determined by

$$\max_x \left\{ \min_{i \in I} \left\{ \frac{\text{output}_i(x)}{\text{input}_i(x)} \right\} \right\},$$

where x denotes a feasible production plan of the economy. A modern example is the congestion control problem in a wireless telecommunication network where the congestion level $Con(l)$ on a particular link l is defined as the loaded flow f_l of the link divided by the capacity C_l of it. In other words

$$\min_{P=(P_l, l \in L)} \left\{ \max_{l \in L} \left\{ Con(l) = \frac{f_l(P)}{C_l(P)} \right\} \right\}$$

where $P = (P_l, l \in L) = \left(P_1, P_2, ..., P_{|L|}\right)$ is the combination of power levels on each link $l \in L$, L is the collection of links in the given wireless network, and $0 \le P_l \le H$ with H being the highest power level allowed. The objective of the congestion control problem is then to determine the best power level for each link such that the highest congestion level in the network is minimized. For more details in the congestion control problem see [29]. More applications can be found in [12, 32, 34, 66, 113, 115].

For solving a GFP, there have been several primal Dinkelbach-type algorithms in the literature [17, 33–36, 107, 108, 122], and dual algorithms and results [1, 2, 13–15, 21, 22, 24–26, 37, 42, 43, 67]. See [118–121] for a detailed bibliography on fractional programming. These algorithms are based on auxiliary parametric problems having simpler structures than the original problem. For the primal algorithms, the auxiliary problems furnish sequences of approximate optimal values converging decreasingly to the optimal value of (P), whereas the sequences of values generated by the dual algorithms converge increasingly towards the optimal value of (P).

Another strategy was proposed in [122], which consists in applying bundle methods for solving a GFP. These methods consist in approximately solving the primal auxiliary problems associated with the GFP by using primal bundle methods. Recently, since the last algorithm is rather intended to solve linearly constrained GFPs, another primal bundle method, based this time on the extended method of centers [107], was proposed in [1] to deal with nonlinearly constrained GFPs. Very recently, a dual bundle method has been proposed in [26], also for solving such problems, this time without convexity assumptions.

This thesis deals with theoretical and numerical analysis for solving the problem (P), with two versions, the first with ratios of convex functions, with convex constraints, the second with ratios of difference of convex

(DC) functions, with DC constraints. Afterward, we develop optimality conditions and propose an algorithm for a vector fractional mathematical program with ratios of difference of convex (DC) functions, and DC constraints.

In Chapter 1, we present the basic notations and notions used in convex analysis [106], and other notions related to the semi-continuity, sub-differentiability and Clarke sub-differentiability [11, 30]. Chapter 2 concerns DC functions and DC programming, where we consider some important definitions and properties on DC functions, the problems in DC programming and we give some real-world applications of DCA (DC Algorithms) from [81].

In Chapter 3, we propose to solve the first version of (P), that we give necessary optimality conditions, of Clarke stationarity type and we describe our DC Dinkelbach-type algorithm and establish its convergence to a Clarke stationary point.

Chapter 4 deals with the second version of (P), i.e. with ratios of DC functions, with DC constraints. We show that the latter is equivalent to a convex problem. By using only convex analysis tools, we obtain necessary optimality conditions, describe our DC-Dinkelbach-type algorithm and establish its convergence.

Another strategy will be proposed in Chapter 5, that is an approximating scheme based on the proximal point algorithm. We take advantage of the convexity property of the associated approximate parametric problem of DC-GFP studied in Chapter 4. The proposed methods generate a sequence of approximate solutions that converges to critical points satisfying necessary optimality conditions of KKT type.

In Chapter 6, necessary conditions of KKT type for (weak) Pareto optimality are derived and DC-Dinkelbach-type algorithm is proposed by first reducing a vector fractional mathematical programming with ratios of DC functions, with DC constraints to a system of scalar parametric problems and then using convex analysis tools. Later, we give an application to vector fractional mathematical programming with ratios of convex functions.

The last chapter is devoted to numerical experiments to evaluate the efficiency of the algorithms described in Chapter 3, Chapter 4, and give comparisons between these algorithms in various cases.

Chapter 1

Generalities

The main purpose of this chapter is to familiarize the reader with the basic notations and notions used in convex analysis [106], and other notions related to the semi-continuity, sub-differentiability and Clarke sub-differentiability [11, 30]. We also recall fundamental useful methods in the rest of the manuscript.

1.1 Preliminaries

Throughout this thesis, we denote by $\mathbb{R}$ the set of all real numbers and by $\mathbb{R}^n$, $n \in \mathbb{N}^*$, the set of all n-tuples of real numbers x. All the vectors x are considered as column vectors and, correspondingly, all the transposed vectors $x^\top$ are considered as row vectors. Given x and y in $\mathbb{R}^n$ and x_i, y_i their i-th components, respectively, the inner product of x and y is defined by

$$\langle x, y \rangle := x^\top y = \sum_{i=1}^{n} x_i y_i.$$

The Euclidean norm of $x \in \mathbb{R}^n$ is defined by

$$\|x\| := \left(x^\top x\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}.$$

In the following, let X be a subset of $\mathbb{R}^n$.

1.1.1 Sets and Affine Sets

The Cartesian product of the sets X and Y is the set $X \times Y$ defined by

$$X \times Y = \{(x, y) \mid x \in X \text{ and } y \in Y\}.$$

We introduce the definitions of balls in $\mathbb{R}^n$ by using the Euclidean norm.

Definition 1.1.1. An open (closed) ball with center $\bar{x} \in \mathbb{R}^n$ and radius $r > 0$ is denoted by $B(\bar{x}, r)$ $\left(\bar{B}(\bar{x}, r)\right)$. That is,

$$B(\bar{x}, r) = \{x \in \mathbb{R}^n \mid \|x - \bar{x}\| < r\} \text{ and } \bar{B}(\bar{x}, r) = \{x \in \mathbb{R}^n \mid \|x - \bar{x}\| \leq r\}.$$

Definition 1.1.2. The closed unit ball of $\mathbb{R}^n$ is the set

$$\mathbb{B} := \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}.$$

It is easy to observe that $\mathbb{B} = \bar{B}(0, 1)$ and $\bar{B}(\bar{x}, r) = \bar{x} + r\mathbb{B}$.

The closure of X, denoted by $\operatorname{cl} X$ or $\overline{X}$, is the intersection of all closed sets containing X, and is the set of limits of sequences of points in X,

$$\overline{X} = \bigcap_{\varepsilon > 0} \{X + \varepsilon \mathbb{B}\}.$$

The set X is said to be closed if $\overline{X} = X$.
The interior $\operatorname{int} X$ of X is the union of all open sets contained in X,

$$\operatorname{int} X = \bigcup \left\{x \mid \exists \varepsilon > 0, \bar{B}(x, \varepsilon) \subset X\right\}.$$

The set X is said to be open if $\operatorname{int} X = X$.
The boundary of X is denoted by $\operatorname{bd} X$. Notice that we have

$$\operatorname{bd} X = \operatorname{cl} X \setminus \operatorname{int} X.$$

If x and y are different points in $\mathbb{R}^n$, the set of points of the form

$$\lambda x + (1 - \lambda) y = y + \lambda (x - y), \ \lambda \in \mathbb{R}$$

is called the line through x and y.
A subset X is called an affine set if

$$\forall (x, y) \in X^2, \forall \lambda \in \mathbb{R}, \ \lambda x + (1 - \lambda) y \in X.$$

The empty set $\emptyset$ and the space $\mathbb{R}^n$ itself are extreme examples of affine sets. Also covered by the definition is the case where X consists of a solitary point. In general, an affine set has to contain, along with any two different points, the entire line through those points.

The affine hull of a set X is the smallest affine set containing X, or equivalently, the intersection of all affine sets containing X.

The affine hull $\operatorname{aff} X$ of X is the set of all affine combinations of elements of X, that is,

$$\operatorname{aff} X = \left\{ \sum_{i=1}^{m} \lambda_i x_i \mid m \in \mathbb{N}^*, x_i \in X, \lambda_i \in \mathbb{R}, \sum_{i=1}^{m} \lambda_i = 1 \right\}.$$

Now, we can define the notion of relative interiors of sets.

We say that an element $x \in X$ belongs to the relative interior of X denoted by $\operatorname{ri} X$ if there exists $\varepsilon > 0$ such that $\bar{B}(x,\varepsilon) \cap \operatorname{aff} X \subset X$. In other words,

$$\operatorname{ri} X = \left\{ x \in \operatorname{aff} X \mid \exists \varepsilon > 0, \bar{B}(x,\varepsilon) \cap \operatorname{aff} X \subset X \right\}.$$

Definition 1.1.3. Let $\{x_k\}$ be a sequence in $\mathbb{R}^n$ and let $\{k_l\}_l$ be a strictly increasing sequence of positive integers. Then the new sequence $\{x_{k_l}\}$ is called a subsequence of $\{x_k\}$. We say that $\{x_k\}$ converges to $\bar{x}$ if $\|x_k - \bar{x}\| \longrightarrow 0$ as $k \longrightarrow \infty$. We write $\lim\limits_{k \longrightarrow \infty} x_k = \bar{x}$.

We say that a set X is bounded if it is contained in a ball centered at the origin with some radius $r > 0$, i.e., $X \subset \bar{B}(0,r)$. Thus a sequence $\{x_k\}$ is bounded if there is $r > 0$ with

$$\|x_k\| \leq r \text{ for all } k \in \mathbb{N}.$$

The following important result is known as the Bolzano-Weierstrass theorem.

Theorem 1.1.1. Any bounded sequence in $\mathbb{R}^n$ contains a convergent subsequence.

The following result is a consequence of the Bolzano-Weierstrass theorem.

Corollary 1.1.1. We say that a set X is compact in $\mathbb{R}^n$ if every sequence in X contains a subsequence converging to some point in X. Moreover, X is compact if and only if it is closed and bounded.

1.2 Convex Analysis

We start this section by the definition of a convex set.

1.2.1 Convex Sets

We denote by $[x,y]$ the closed line-segment joining x and y, that is,

$$[x,y] = \{z \in \mathbb{R}^n \mid z = \lambda x + (1-\lambda)y, \text{ for } \lambda \in [0,1]\}.$$

Definition 1.2.1. Let X be a subset of $\mathbb{R}^n$. The set X is said to be convex if

$$[x,y] \subset X$$

for all $x, y \in X$.

Geometrically this means that the set is convex if the closed line-segment $[x,y]$ is entirely contained in X whenever its endpoints x and y are in X.

We note that all affine sets are convex. Given $x_1, \dots, x_m \in \mathbb{R}^n$, $m \in \mathbb{N}^*$, the element $x = \sum_{i=1}^m \lambda_i x_i$, where $\sum_{i=1}^m \lambda_i = 1$ and $\lambda_i \geq 0$, is called a convex combination of $x_1, \dots, x_m$.

Proposition 1.2.1. A set X is convex if and only if it contains all convex combinations of its elements.

Definition 1.2.2. The intersection of all convex sets containing X is called the convex hull of X, and is denoted by $\operatorname{conv} X$. Equivalently,

$$\operatorname{conv} X := \bigcap \{C \subset \mathbb{R}^n \mid C \text{ is convex and } X \subset C\}.$$

Theorem 1.2.1. The convex hull of X, $\operatorname{conv} X$ consists of all the convex combinations of the elements of X, i.e.,

$$\operatorname{conv} X := \left\{ \sum_{i=1}^m \lambda_i x_i \mid \sum_{i=1}^m \lambda_i = 1, \lambda_i \geq 0, x_i \in X, \text{ and } m \in \mathbb{N}^* \right\}.$$

Proposition 1.2.2.

1. Let X and Y be two convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^p$, respectively. Then, the cartesian product $X \times Y$ is a convex subset of $\mathbb{R}^n \times \mathbb{R}^p$.
2. Let X and Y be two convex subsets of $\mathbb{R}^n$ and $\mu_1, \mu_2 \in \mathbb{R}$. Then the set $\mu_1 X + \mu_2 Y$ is also convex.

Proposition 1.2.3. If X is a nonempty convex set then $\operatorname{ri} X \neq \emptyset$.
Definition 1.2.3. The set Σ defined by

$$\Sigma = \left\{ (\alpha_1, \alpha_2, \dots, \alpha_n)^T \in \mathbb{R}^n \mid \alpha_i \geq 0, \text{ for }, i = 1, 2, \dots, n, \text{ and } \sum_{i=1}^{n} \alpha_i = 1 \right\},$$

is called the unit simplex in $\mathbb{R}^n$ and it is a convex set.
Definition 1.2.4. A non-empty subset C of $\mathbb{R}^n$ is called a cone if for each $x \in C$ and each $\alpha \geq 0$ we have $\alpha x \in C$.

Further, C is a convex cone if it is cone and also is convex.

1.2.2 Convex Functions

In this section, we investigate some fundamental properties of convex functions.

Definition 1.2.5.

1. A mapping $l : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ is said to be a linear if for any two vectors x, $y \in \mathbb{R}^n$ and any scalar $\alpha \in \mathbb{R}$, the following two conditions are satisfied:

$$l(x+y) = l(x) + l(y) \text{ (additivity) and } l(\alpha x) = \alpha l(x) \text{ (homogeneity)}.$$

2. A mapping $f : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ is affine if there exist a linear mapping $l : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ and an element $b \in \mathbb{R}^p$ such that $f(x) = l(x) + b$ for all $x \in \mathbb{R}^n$.

We denote by $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ the extended real line. For functions that map to the extended reals, we define the domain and epigraph as follows.

Definition 1.2.6. Let $f : X \longrightarrow \overline{\mathbb{R}}$ be an extended real-valued function.

1. The effective domain of f is defined by

$$\operatorname{dom} f = \{x \in X \mid f(x) < \infty\}.$$

The function f is said to be proper if $\operatorname{dom} f \neq \emptyset$ and $f(x) > -\infty$, for all $x \in X$.

2. The epigraph of f is defined by

$$\operatorname{epi} f = \{(x,t) \in X \times \mathbb{R} \mid f(x) \leq t\}.$$

Recall the usual definition of a convex function.

Definition 1.2.7. Let $f : X \longrightarrow \overline{\mathbb{R}}$ be an extended real-valued and proper function defined on a convex set X. Then the function f is convex on X if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \tag{1.2.1}$$

for every $\lambda \in [0,1]$ and every $x, y \in X$.

If (1.2.1) holds with strict inequality for any $\lambda \in]0,1[$ and every $x, y \in X$ with $x \neq y$, then f is called strictly convex.

A (strictly) concave function on X is a function whose negative is (strictly) convex. The affine functions on X are convex and concave.

The convexity of a function can be characterized via the convexity of its epigraph.

Proposition 1.2.4. Let $f : X \longrightarrow \overline{\mathbb{R}}$ be a proper function defined on a convex set X. The following assertions are equivalent:

1. f is convex,
2. epi$f$ is a convex set in $X \times \mathbb{R}$.

Another useful characterization of convexity (called Jensen's inequality) is the following.

Theorem 1.2.2. A function $f : X \longrightarrow \overline{\mathbb{R}}$ is convex on a convex set X if and only if

$$f\left(\sum_{i=1}^{m} \lambda_i x_i\right) \leq \sum_{i=1}^{m} \lambda_i f(x_i)$$

for every $x_i \in X$ and $\lambda_i \geq 0$, $i = 1, \ldots, m$, $m \in \mathbb{N}^*$, with $\sum_{i=1}^{m} \lambda_i = 1$.

The indicator function of the subset X is defined by

$$\text{Ind}_X(x) := \begin{cases} 0 & \text{if } x \in X, \\ \infty & \text{otherwise.} \end{cases}$$

This definition gives a correspondence between convex sets and convex functions. The set X is convex if and only if its indicator function $\text{Ind}_X(\cdot)$

is convex.

In the next result we mention some methods for deriving new convex functions from known ones.

Theorem 1.2.3. Let X, Y be two subsets of $\mathbb{R}^n$.

1. If $f_i : X \longrightarrow \overline{\mathbb{R}}$ is convex for every $i \in I$ (I is an index set, $I \neq \emptyset$) then $\sup_{i\in I} f_i$ is convex. Moreover, $\text{epi}(\sup_{i\in I} f_i) = \cap_{i\in I}\text{epi} f_i$.
2. If $f_1, f_2 : X \longrightarrow \overline{\mathbb{R}}$ are proper convex functions and $\lambda \in \mathbb{R}_+$, then $f_1 + f_2$ and λf_1 are convex, where $0.f_1 = \text{Ind}_{\text{dom} f_1}$. Moreover,
$$\text{dom}\,(f_1 + f_2) = \text{dom} f_1 \cap \text{dom} f_2 \text{ and } \text{dom}\,(\lambda f_1) = \text{dom} f_1.$$
3. If $f_n : X \longrightarrow \overline{\mathbb{R}}$ is convex for every $n \in \mathbb{N}$, and $f : X \longrightarrow \overline{\mathbb{R}}$ is such that $f(x) = \limsup_{n\longrightarrow\infty} f_n(x)$ for every $x \in X$, then f is convex.
4. If $F : X \times Y \longrightarrow \overline{\mathbb{R}}$ is convex, then the marginal function h associated to F is convex, where
$$h : Y \longrightarrow \overline{\mathbb{R}},\ h(y) := \inf_{x\in X} F(x, y).$$

Let $C \subset \mathbb{R}^n$ be a non-empty convex set. The support function of C is the function $\delta^*(\cdot \mid C) : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ defined by

$$\delta^*(x \mid C) = \sup_{y\in C} \langle x, y\rangle,\ x \in \mathbb{R}^n.$$

The (Euclidean) distance function $d(\cdot \mid C)$ is defined by

$$d(x \mid C) = \inf\{\|x - y\| \mid y \in C\}.$$

It is clear that, the support and distance functions are convex.

Now, we introduce the following definition of quasi-convexity.

Definition 1.2.8. A function $f : X \longrightarrow \overline{\mathbb{R}}$ is quasi-convex if its domain and all its sublevel sets

$$S_\alpha = \{x \in \text{dom} f \mid f(x) \leq \alpha\},\ \text{for } \alpha \in \mathbb{R}$$

are convex or, which is equivalent, if

$$\forall x, y \in X,\ \forall \lambda \in [0,1]:\ f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}.$$

This means that the value of the function f on a segment does not exceed the maximum of its values at the endpoints. If we have the strict inequality in the previous inequality for all $x \neq y$ and $\lambda \in]0,1[$ we say that f is strictly quasi-convex. A function f is quasi-concave if $-f$ is quasi-convex, i.e., every superlevel set $\{x \mid f(x) \geq \alpha\}$ is convex.

1.2.3 Some properties of quadratic functions

Definition 1.2.9. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix, we say that the matrix A is

1. positive semi-definite (resp. positive-definite) if $x^\top Ax \geq 0$ (resp. $x^\top Ax > 0$) for every $x \in \mathbb{R}^n$ (resp. $x \in \mathbb{R}^n_*$),
2. negative semi-definite (resp. negative-definite) if $x^\top Ax \leq 0$ (resp. $x^\top Ax < 0$) for every $x \in \mathbb{R}^n$ (resp. $x \in \mathbb{R}^n_*$),
3. undefined if it exists $x, y \in \mathbb{R}^n$ such that $x^\top Ax > 0$ and $y^\top Ay < 0$.

Definition 1.2.10. Let A be a symmetric matrix, we say that the matrix A is diagonally dominant (resp. strictly dominant) if $|a_{ii}| \geq \sum_{j\neq i} |a_{ij}|$ (resp. $|a_{ii}| > \sum_{j\neq i} |a_{ij}|$) for all $i = 1, 2, \ldots, n$.

Proposition 1.2.5. Any diagonal dominant matrix whose diagonal elements are positive or zero (resp. strictly dominant whose diagonal elements are positive) is positive semi-definite (resp. positive-definite).

Definition 1.2.11. Let $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. The function defined on $\mathbb{R}^n$ by

$$f(x) = \tfrac{1}{2}x^\top Ax + bx + c,$$

is called a quadratic function.

In any quadratic function, we can transform its matrix into a symmetric matrix, indeed

$$x^\top Ax = \tfrac{1}{2}\left(2x^\top Ax\right) = \tfrac{1}{2}x^\top\left(A + A^\top\right)x.$$

Proposition 1.2.6. Let f be a quadratic function defined by

$$f(x) = \tfrac{1}{2}x^\top Ax + bx + c,$$

then

- the gradient vector of f at x is given by $\nabla f(x) = \frac{1}{2}(A + A^\top)x + b$,
- the Hessian matrix of f at x is given by $\nabla^2 f(x) = \frac{1}{2}(A + A^\top)$.

If moreover the matrix A is symmetric then $\nabla f(x) = Ax+b$, and $\nabla^2 f(x) = A$.

From the above we deduce the following proposition:
Proposition 1.2.7. Let $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. The quadratic function $f(x) = \frac{1}{2}x^\top Ax + bx + c$, is convex (resp. strictly convex) if and only if the matrix A is positive semi-definite (resp. positive-definite).

1.2.4 Continuous and Lower Semi-Continuous Functions

Definition 1.2.12. Let $f : X \longrightarrow \overline{\mathbb{R}}$ be an extended-real-valued function and let $\bar{x} \in X$ with $f(\bar{x}) < \infty$. Then f is continuous at $\bar{x}$ if for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$|f(x) - f(\bar{x})| < \varepsilon \text{ whenever } \|x - \bar{x}\| < \delta,\ x \in X.$$

We say that f is continuous on X if it is continuous at every point of X.

It is obvious from the definition that if $f : X \longrightarrow \overline{\mathbb{R}}$ is continuous at $\bar{x}$ (with $f(\bar{x}) < \infty$), then it is finite on the intersection of X and a ball centered at $\bar{x}$ with some radius $r > 0$. Furthermore, $f : X \longrightarrow \overline{\mathbb{R}}$ is continuous at $\bar{x}$ (with $f(\bar{x}) < \infty$) if and only if for every sequence $\{x_k\}$ in X converging to $\bar{x}$ the sequence $\{f(x_k)\}$ converges to $f(\bar{x})$.

Definition 1.2.13. For a function $f : X \longrightarrow \overline{\mathbb{R}}$, the lower and upper limits of f at $\bar{x}$ are defined by

1. $\liminf_{x\longrightarrow\bar{x}} f(x) = \sup_{r>0} \inf\{f(x) \ :\ \|x - \bar{x}\| < r\ ,\ x \in X\backslash\{\bar{x}\}\}$.
2. $\limsup_{x\longrightarrow\bar{x}} f(x) = \inf_{r>0} \sup\{f(x) \ :\ \|x - \bar{x}\| < r\ ,\ x \in X\backslash\{\bar{x}\}\}$.

Definition 1.2.14. A function $f : X \longrightarrow \overline{\mathbb{R}}$ is lower semi-continuous at $\bar{x} \in X$ if

$$f(\bar{x}) \leq \liminf_{x\longrightarrow\bar{x}} f(x).$$

The function f is lower semi-continuous on X if f is lower semi-continuous at each point of this set.

A function f is upper semi-continuous at point $\bar{x} \in X$ if

$$\limsup_{x\longrightarrow\bar{x}} f(x) \leq f(\bar{x}).$$

A function f is continuous at a point $\bar{x}$ if and only if it is lower semi-continuous and upper semi-continuous at $\bar{x}$.

Proposition 1.2.8. Let $f : X \longrightarrow \overline{\mathbb{R}}$ be an extended-real-valued function and let $\bar{x} \in X$ with $f(\bar{x}) < \infty$. Then the following assertions are equivalent:

1. f is lower semi-continuous at $\bar{x}$,
2. $f(\bar{x}) \leq \liminf_{x \longrightarrow \bar{x}} f(x) = \sup_{r>0} \inf_{x \in \bar{B}(\bar{x},r)} f(x)$,
3. for every sequence $(x_k)_{k \in \mathbb{N}}$ converging to $\bar{x}$ we have

   $f(\bar{x}) \leq \liminf_{k \longrightarrow \infty} f(x_k)$.

Theorem 1.2.4. Let $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be an extended-real-valued function. Then the following assertions are equivalent:

1. f is lower semi-continuous throughout $\mathbb{R}^n$,
2. the level sets $\{x \mid f(x) \leq \alpha\}$ are closed, for all $\alpha \in \mathbb{R}$,
3. epi$f$ is a closed set in $\mathbb{R}^n \times \mathbb{R}$.

Corollary 1.2.1. Let $f : X \longrightarrow \overline{\mathbb{R}}$ be a convex function. Then f is continuous on int(dom$f$) if and only if int(epi$f$) is nonempty in $X \times \mathbb{R}$.

Proposition 1.2.9.

1. The infimum and the supremum of a finite family of lower semi-continuous functions are lower semi-continuous.
2. Let $f, g : X \longrightarrow \overline{\mathbb{R}}$ be two lower semi-continuous functions, then $f+g$ is also lower semi-continuous.

Theorem 1.2.5. Let $C \subset \mathbb{R}^n$ be a convex set. Then every convex function $f : C \longrightarrow \overline{\mathbb{R}}$ is continuous on ri$C$.

Definition 1.2.15. A function $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ is locally Lipschitz continuous at a point $x \in \mathbb{R}^n$ if there exists a scalar $L > 0$ and $\varepsilon > 0$ such that

$$|f(y) - f(z)| \leq L\|y - z\| \text{ for all } y, z \in B(x, \varepsilon).$$

A function $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ is said to be locally Lipschitz continuous on a set $X \subset \mathbb{R}^n$ if it is locally Lipschitz continuous at every point belonging to the set X. Furthermore, if $X = \mathbb{R}^n$ the function is called locally Lipschitz continuous.

Definition 1.2.16. A function $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ is said to be Lipschitz continuous on a set $X \subset \mathbb{R}^n$ if there exists a scalar L such that

$$|f(x) - f(y)| \leq L\|x - y\| \text{ for all } x, y \in X.$$

If $X = \mathbb{R}^n$, then f is said to be Lipschitz continuous.

1.2.5 Subdifferential and Directional Derivatives

This section generalizes the classical notion of gradient for convex but not necessarily differentiable functions, by introducing the notions of (subdifferential and subgradient), defined for convex functions on $\mathbb{R}^n$.
Definition 1.2.17. Let $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be a proper convex function. The subdifferential of f at point x^* is the set

$$\partial f(x^*) := \{\xi^* \in \mathbb{R}^n \mid f(x) \geq f(x^*) + \langle \xi^*, x - x^* \rangle \text{ for all } x \in \mathbb{R}^n\}.$$

Each vector $\xi^* \in \partial f(x^*)$ is called a subgradient of f at x^*, and we say that f is subdifferentiable at x^* if there is at least one subgradient at this point.
Proposition 1.2.10. Let $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be a convex function and $x \in$ri(dom$f$) such that $f(x) \in \mathbb{R}$. Then f is proper and $\partial f(x)$ is nonempty.
Definition 1.2.18. Let $\varepsilon \geq 0$, then the ε-subdifferential of the function $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ at $x \in \mathbb{R}^n$ is the set

$$\partial_\varepsilon f(x) := \{\xi \in \mathbb{R}^n \mid f(y) \geq f(x) + \langle \xi, y - x \rangle - \varepsilon \text{ for all } y \in \mathbb{R}^n\}.$$

Each element $\xi \in \partial_\varepsilon f(x)$ is called an ε-subgradient of f at x.

We consider some properties related to the directional derivative of convex functions.
Definition 1.2.19. Let $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be an extended-real-valued function and let $x \in \text{dom} f$. The directional derivative of the function f at the point x in the direction $d \in \mathbb{R}^n$ is the following limit if it exists as either a real number or ∞:

$$f'(x;d) = \lim_{t \longrightarrow 0^+} \frac{f(x+td) - f(x)}{t}.$$

Note that $f'(x;d)$ is sometimes called the right directional derivative of f at x in the direction d. Its left counterpart is defined by

$$f'_-(x;d) = \lim_{t \longrightarrow 0^-} \frac{f(x+td) - f(x)}{t}.$$

We have that

$$f'_-(x;d) = -f'(x;-d),$$

so that the one-sided directional derivative $f'(x;d)$ is two-sided if and only if $f'(x;-d)$ exists and

$$f'(x;d) = -f'(x;-d).$$

Theorem 1.2.6. Let $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be convex with $x^* \in \mathrm{dom} f$. The following assertions are equivalent:

1. $\xi^* \in \partial f(x^*)$,
2. $f'(x^*;d) \geq \langle \xi^*, d\rangle$, for all $d \in \mathbb{R}^n$,
3. $f'(x^*;d) \geq \langle \xi^*, d\rangle \geq f'_-(x^*;d)$, for all $d \in \mathbb{R}^n$.

We describe below the subdifferential of different form of the convex function f.

Theorem 1.2.7. 1. For $\alpha > 0$, $\partial(\alpha f)(x) = \alpha \partial f(x)$.

2. Suppose $f = f_1 + \cdots + f_m$, where for all $i \in \{1, \cdots, m\}$, $f_i : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ are proper convex functions satisfying the relative interior qualification condition

$$\mathrm{ri}(\mathrm{dom} f_1) \cap \mathrm{ri}(\mathrm{dom} f_2) \cap \cdots \cap \mathrm{ri}(\mathrm{dom} f_m) \neq \emptyset.$$

Then for $x \in \cap_{i=1}^m \mathrm{dom} f_i$ we have

$$\partial f(x) = \partial f_1(x) + \cdots + \partial f_m(x).$$

3. Suppose $f(x) = \max_{1\leq i\leq m} f_i(x)$, where for all $i \in \{1, \cdots, m\}$, $f_i : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ are proper convex subdifferentiable functions, and let $x \in \cap_{i=1}^m \mathrm{dom} f_i$ be such that each f_i is continuous at x. Then we have

$$\partial f(x) = \mathrm{conv}\Big\{\bigcup \{\partial f_i(x) \mid f(x) = f_i(x)\}\Big\}.$$

4. Let X be a convex set. Then

$$\partial(\mathrm{Ind}_X(x)) = \begin{cases} N_X(x) & \text{if } x \in X, \\ \emptyset & \text{otherwise.} \end{cases}$$

where $N_X(x)$ is called the normal cone of X at x and is defined as

$$N_X(x) := \Big\{\eta \in \mathbb{R}^n \mid \eta^\top (y - x) \leq 0, \forall y \in X\Big\}.$$

1.3 Generalized Subdifferential and Directional Derivatives

This section, generalizes the convex concepts defined in the previous section for nonconvex locally Lipschitz continuous functions. Since the classical directional derivative does not necessarily exist for locally Lipschitz continuous functions, we first define a generalized directional derivative. Then we generalize the subdifferential analogously. We use the approach of Clarke in a finite dimensional case.

We start by generalizing the ordinary directional derivative. Note that this generalized derivative always exists for locally Lipschitz continuous functions.

Definition 1.3.1 : ***(Clarke).*** Let $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be a locally Lipschitz continuous function at $x \in \mathbb{R}^n$. The generalized directional derivative of f at x in the direction of $d \in \mathbb{R}^n$ is defined by

$$f^\circ(x;d) = \limsup_{y \longrightarrow x, t \searrow 0} \frac{f(y+td) - f(y)}{t}.$$

The following summarizes some basic properties of the generalized directional derivative.

Theorem 1.3.1. 1. Let f be locally Lipschitz continuous at x with constant L. Then the function $d \mapsto f^\circ(x;d)$ is positively homogeneous and subadditive on $\mathbb{R}^n$ with

$$|f^\circ(x;d)| \leq L\|d\|.$$

2. If $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ is locally Lipschitz continuous at x, then the function $d \mapsto f^\circ(x;d)$ is convex, its epigraph epi$f^\circ(x;.)$ is a convex cone and we have
$$f^\circ(x;-d) = (-f)^\circ(x;d).$$

3. If f is locally Lipschitz continuous at x with constant L, then the function $(x,d) \mapsto f^\circ(x;d)$ is upper semicontinuous.

Definition 1.3.2 : ***(Clarke).*** Let $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be a locally Lipschitz continuous function at a point $x \in \mathbb{R}^n$. Then the Clarke subdifferential of f at x is the set $\partial^c f(x)$ of vectors $\xi \in \mathbb{R}^n$ such that

$$\partial^c f(x) := \left\{\xi \in \mathbb{R}^n \mid f^\circ(x;d) \geq \xi^\top d \text{ for all } x \in \mathbb{R}^n\right\}.$$

Each vector $\xi \in \partial^c f(x)$ is again called a Clarke subgradient of f at x.

Theorem 1.3.2. Let $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be a locally Lipschitz continuous function at $x \in \mathbb{R}^n$. Then

$$f^\circ(x;d) = \max\left\{\xi^\top d \mid \xi \in \partial^c f(x)\right\} \text{ for all } d \in \mathbb{R}^n.$$

The following theorem shows that the subdifferential for Lipschitz continuous functions is a generalization of the subdifferential for convex functions.

Theorem 1.3.3. If the function $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ is convex, then

1. $f'(x;d) = f^\circ(x;d)$, for all $d \in \mathbb{R}^n$,
2. $\partial f(x) = \partial^c f(x)$.

Next we go through classical derivation rules for locally Lipschitz continuous functions.

Proposition 1.3.1. If the function $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ is locally Lipschitz continuous at x, then for all $\alpha \in \mathbb{R}$

$$\partial^c (\alpha f)(x) = \alpha \partial^c f(x).$$

Now we are going to state a theorem that we will use in the next chapters to deduce the Clarke optimality conditions for a Lipschitzian function, but before we need the following regularity property.

Definition 1.3.3. The function $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ is said to be subdifferentially regular at $x \in \mathbb{R}^n$ if it is locally Lipschitz continuous at x and for all $d \in \mathbb{R}^n$ the classical directional derivative $f'(x;d)$ exists and we have

$$f'(x;d) = f^\circ(x;d).$$

Theorem 1.3.4. Let f_1 and f_2 be locally Lipschitz continuous at $x \in \mathbb{R}^n$ and $f_2(x) \neq 0$. Then the function f_1/f_2 is locally Lipschitz continuous at x and

$$\partial^c \left(\frac{f_1}{f_2} \right)(x) \subseteq \frac{f_2(x)\partial^c f_1(x) - f_1(x)\partial^c f_2(x)}{f_2^2(x)}. \tag{1.3.1}$$

If in addition $f_1(x) \geq 0$, $f_2(x) > 0$ and f_1, $-f_2$ are both subdifferentially regular at x, then the function f_1/f_2 is subdifferentially regular at x and equality holds in (1.3.1).

Theorem 1.3.5. Let $f_i : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be locally Lipschitz continuous at x for all $i = 1, \ldots, m$. Then the function

$$f(x) := \max\{f_i(x) \mid i = 1, \ldots, m\},$$

is locally Lipschitz continuous at x and

$$\partial^c f(x) \subseteq \text{co}\{\partial^c f_i(x) \mid i \in I(x)\}, \tag{1.3.2}$$

where

$$I(x) := \{i = 1, \ldots, m \mid f_i(x) = f(x)\}.$$

In addition, if f_i is subdifferentially regular at x for all $i = 1, \ldots, m$, then f is also subdifferentially regular at x and equality holds in (1.3.2).

1.4 Optimization problems

We consider a nonsmooth optimization problem of the form:

$$(P) \qquad \min_{x \in X} f(x),$$

where the objective function $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ is supposed to be locally Lipschitz continuous and the feasible region $X \subset \mathbb{R}^n$ is nonempty. If f is a convex function and X is a convex set, then the problem (P) is called convex.

1.4.1 Minimizers

Definition 1.4.1. A point $\bar{x} \in X$ is a global optimum of the problem (P) if it satisfies

$$f(\bar{x}) \leq f(x) \quad \text{for all } x \in X.$$

Definition 1.4.2. A point $\bar{x} \in X$ is a local optimum of the problem (P) if there exists an $\epsilon > 0$ such that

$$f(\bar{x}) \leq f(x) \text{ for all } x \in X \cap \bar{B}(\bar{x}, \epsilon).$$

A local minimum of a convex function f on a convex set X is also a global minimum.

If X is closed and f is l.s.c and coercive, then f admits at least one minimizer. Furthermore, if X is convex and f is a strictly convex function, then f admits at most one minimizer.

1.4.2 Unconstrained Optimization

We consider first the unconstrained version of the problem (P), in other words the case $X = \mathbb{R}^n$. Then we are actually looking for local and global minima of a locally Lipschitz continuous function.

Necessary conditions for a locally Lipschitz continuous function to attain its local minimum are given in the next theorem. For convex functions these conditions are also sufficient and the minimum is global.

Theorem 1.4.1. Let $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be a locally Lipschitz continuous function at $\bar{x} \in \mathbb{R}^n$. If f attains its local minimum at $\bar{x}$, then

$$0 \in \partial^c f(\bar{x}) \text{or} f^\circ(\bar{x}; d) \geq 0 \quad \text{for all } d \in \mathbb{R}^n.$$

Definition 1.4.3. A point $x \in \mathbb{R}^n$ satisfying $0 \in \partial^c f(x)$ is called a stationary point of f.

Theorem 1.4.2. If the function $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ is convex, then f attains its global minimum at $\bar{x}$ if and only if

$$0 \in \partial f(\bar{x}) \text{ or } f'(\bar{x}; d) \geq 0 \quad \text{for all } d \in \mathbb{R}^n.$$

Definition 1.4.4. If $\varepsilon \geq 0$, then a point $\bar{x} \in X$ is a global $\varepsilon-$optimum of the problem (P) if it satisfies

$$f(\bar{x}) \leq f(x) + \varepsilon \text{ for all } x \in X.$$

Note, that similarly we can define also local $\varepsilon-$optimality.

Theorem 1.4.3. If the function $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ is convex and $\varepsilon \geq 0$, then f attains its global $\varepsilon-$minimum at $\bar{x}$ if and only if

$$0 \in \partial_\varepsilon f(\bar{x}).$$

1.4.3 Constrained Optimization

Next we consider the problem (P) when the feasible set is not the whole space $\mathbb{R}^n$, in other words $X \subsetneq \mathbb{R}^n$. In this subsection we do not assume any special structure of X, but consider it as a general set. We get the same optimality conditions than in unconstrained case.

Theorem 1.4.4. Let $\bar{x}$ be a local optimum of problem (P), where $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ is locally Lipschitz continuous at $\bar{x} \in X \neq \emptyset$. Then

$$0 \in \partial^c f(\bar{x}) + N_X(\bar{x}).$$

Now we formulate the following necessary and sufficient optimality condition utilizing convexity.

Theorem 1.4.5. If the problem (P) is convex, then $\bar{x} \in X$ is a global optimum of the problem (P) if and only if

$$0 \in \partial f(\bar{x}) + N_X(\bar{x}).$$

1.4.4 Karush-Kuhn-Tucker Optimality Conditions

Consider a nonsmooth optimization problem (P) with a special structure of X determined with inequality constraints

$$X := \{x \in S \mid g_i(x) \le 0, \text{ for } i = 1, 2, \ldots, m\},$$

where the constraint functions $g_i : S \longrightarrow \overline{\mathbb{R}}$ are supposed to be locally Lipschitz continuous for all $i = 1, \ldots, m$ and $S \subset \mathbb{R}^n$ is a nonempty closed set. Without loosing generality, we can scalarize the multiple constraints by introducing the total constraint function $g : S \longrightarrow \overline{\mathbb{R}}$ in the form

$$g(x) := \max\{g_i(x) \mid i = 1, \ldots, m\}.$$

We need some regularization assumptions, called constraint qualifications.

Definition 1.4.5. The problem (P) satisfies the Slater constraint qualification if there exists $\tilde{x} \in S$ such that $g(\tilde{x}) < 0$.
Definition 1.4.6. The problem (P) satisfies the Cottle constraint qualification at $\tilde{x} \in S$ if either $g(\tilde{x}) < 0$ or $0 \in \partial^c g(\tilde{x})$.

The relationship between those two qualifications is given by

Lemma 1.4.1. If the problem (P) satisfies the Cottle constraint qualification at some $\tilde{x} \in S$, then it satisfies also the Slater constraint qualification. If the functions g_i are convex for all $i = 1,\ldots,m$ and the problem (P) satisfies the Slater constraint qualification, then it satisfies also the Cottle constraint qualification at every $\tilde{x} \in S$.

Now we are ready to generalize Karush-Kuhn-Tucker (KKT) optimality conditions for (P).

Theorem 1.4.6 : ***(KKT Necessary Conditions).*** Let the problem (P) satisfy the Cottle constraint qualification at a local optimum $\bar{x}$, where $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ and $g_i : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ for $i = 1,\ldots,m$ are supposed to be locally Lipschitz continuous at $\bar{x} \in S$. Then there exist multipliers $\lambda_i \geq 0$ for $i = 1,\ldots,m$ such that $\lambda_i g_i(\bar{x}) = 0$ and

$$0 \in \partial^c f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \partial^c g_i(\bar{x}) + N_S(\bar{x}).$$

Now we formulate sufficient KKT optimality conditions utilizing convexity.

Theorem 1.4.7 : ***(KKT Sufficient Conditions).*** Let $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ and $g_i : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ for $i = 1,\ldots,m$ be convex functions. If at $\bar{x} \in S$ there exist multipliers $\lambda_i \geq 0$ for $i = 1,\ldots,m$ such that $\lambda_i g_i(\bar{x}) = 0$ and

$$0 \in \partial f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \partial g_i(\bar{x}) + N_S(\bar{x}),$$

then $\bar{x}$ is a global optimum of the problem (P).

Finally, we can combine the necessary and sufficient conditions.

Theorem 1.4.8 : ***(KKT Necessary and Sufficient Conditions).*** Suppose, that the problem (P) satisfies the Slater constraint qualification and let $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ and $g_i : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ for $i = 1,\ldots,m$ be convex functions. Then $\bar{x}$ is a global optimum of the problem (P) if and only if there exist multipliers $\lambda_i \geq 0$ for $i = 1,\ldots,m$ such that $\lambda_i g_i(\bar{x}) = 0$ and

$$0 \in \partial f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \partial g_i(\bar{x}) + N_S(\bar{x}).$$

1.5 Multiobjective Optimization

If x and y are two vectors in $\mathbb{R}^n$, we write $x \leq y$ if $x_i \leq y_i$, for $i = 1, 2, \cdots, n$, and $x < y$ if $x_i < y_i$, $i = 1, 2, \cdots, n$, where v_i is the i–th component of the

vector v.
Let $f : X \longrightarrow \mathbb{R}^m$, $f(x) = (f_1(x), \cdots, f_m(x))$, be defined on X, and X be a nonempty and convex subset of $\mathbb{R}^n$.

The function f is convex if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

For all $x, y \in \mathbb{R}^n$ and all $\lambda \in [0,1]$, clearly $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is convex if and only if its components $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ are all convex.

A continuous linear map $T : X \longrightarrow \mathbb{R}^m$ is said to be subgradient of $fv : v\mathbb{R}^n \longrightarrow \mathbb{R}^m$ at point $y \in X$ if

$$f(x) \geq f(y) + T(x-y) \quad \text{for all } x \in X.$$

The set of all subgradients of f at y is called the subdifferential of f at y and is defined by $\partial f(y)$, see [106].

Definition 1.5.1. Let $f : X \longrightarrow \mathbb{R}^m$, with $f(x) = (f_1(x), \ldots, f_m(x))$.

1. A vector $\bar{x} \in X$ is said to be a pareto minimum of f if there exists no $x \in X$ such that $f(x) \neq f(\bar{x})$ and $f(x) \leq f(\bar{x})$.
2. A vector $\bar{x} \in X$ is said to be a weak pareto minimum of f if there exists no $x \in X$ such that $f(x) < f(\bar{x})$.

1.6 Minimax Theorems

We give below some theorems that allow to interchange the "inf" and "sup" of a function F defined on $X \times Y$, where X and Y are finite dimensional spaces. More precisely, we look for conditions that guarantee the equality

$$\inf_{x \in X} \sup_{y \in Y} F(x,y) = \sup_{y \in Y} \inf_{x \in X} F(x,y).$$

Theorem 1.6.1 : ***([99]).*** Let X and Y be convex compact sets in topological linear spaces L and M respectively. If $F : X \times Y \longrightarrow \mathbb{R}$ is a continuous function that is quasi-convex-concave, i.e.

1. $F(\cdot, y) : X \longrightarrow \mathbb{R}$ is quasi-convex for fixed y,
2. $F(x, \cdot) : Y \longrightarrow \mathbb{R}$ is quasi-concave for fixed x,

then we have that

$$\min_{x\in X}\max_{y\in Y} F(x,y) = \max_{y\in Y}\min_{x\in X} F(x,y).$$

Theorem 1.6.2 : ***([117], Theorem3.4).*** Let X and Y be convex compact sets. If F is a real-valued function on $X \times Y$ with

1. $F(\cdot, y)$ lower semi-continuous and quasi-convex on X, for all $y \in Y$,
2. $F(x, \cdot)$ upper semi-continuous and quasi-concave on Y, for all $x \in X$,

then we have that

$$\min_{x\in X}\max_{y\in Y} F(x,y) = \max_{y\in Y}\min_{x\in X} F(x,y).$$

A consequence of the last Theorem is derived if X or Y is compact.
Corollary 1.6.1. Let X be a compact convex subset of a linear topological space and Y a convex subset of a linear topological space. If F is a real-valued function on $X \times Y$ with

1. $F(\cdot, y)$ lower semi-continuous and quasi-convex on X, for all $y \in Y$,
2. $F(x, \cdot)$ upper semi-continuous and quasi-concave on Y, for all $x \in X$,

then we have that

$$\inf_{x\in X}\sup_{y\in Y} F(x,y) = \sup_{y\in Y}\inf_{x\in X} F(x,y).$$

Theorem 1.6.3 : ***([104], Theorem1.B).*** Let X be a topological space, $Y \subset \mathbb{R}^m$ a nonempty convex set and $F : X \times Y \longrightarrow \mathbb{R}$ a function satisfying the following conditions:

1. for each $y \in Y$, the function $F(\cdot, y)$ is lower semi-continuous and inf-compact,
2. for each $x \in X$, the function $F(x, \cdot)$ is upper semi-continuous and concave,

then, at least one of the following assertions holds:

(i) there exists $\hat{y} \in Y$ such that the function $F(\cdot, \hat{y})$ has at least two global minima,

(ii) one has,

$$\inf_{x\in X}\sup_{y\in Y} F(x,y) = \sup_{y\in Y}\inf_{x\in X} F(x,y).$$

For each $m \in \mathbb{N}^*$ we put

$$S_m = \left\{ \lambda \in \mathbb{R}^m_+ \mid \sum_{i=1}^{m} \lambda_i = 1 \right\}.$$

Proposition 1.6.1 : ***([104], Lemma2.1).*** Let X be a topological space and let $F : X \times S_m \longrightarrow \mathbb{R}$ be a function satisfying the following conditions:

1. for each $y \in S_m$ the function $F(\cdot, y)$ is lower semi-continuous, inf-compact and has a unique global minimum,
2. for each $x \in X$, the function $F(x, \cdot)$ is continuous and quasi-concave,

then, one has

$$\min_{x \in X} \max_{y \in S_m} F(x, y) = \max_{y \in S_m} \min_{x \in X} F(x, y).$$

1.7 Bundle Methods

Let $f : X \longrightarrow \overline{\mathbb{R}}$ be a proper closed convex function on a Hilbert space X and denote by

$$\bar{f} = \inf_{x \in X} f(x),$$

its infimal value (possibly $-\infty$). We are interested in estimating f, and also in identifying a minimum point, if any.

The sequence $\{\alpha_k\}$ is chosen a priori and $\{x_k\}$ is constructed by the following prox-iteration

$$x_{k+1} = \operatorname*{argmin}_{x \in X} \left\{ f(x) + \frac{1}{2\alpha_k} \|x - x_k\|^2 \right\}.$$

This prox-sequence is characterized by

$$\frac{x_k - x_{k+1}}{\alpha_k} \in \partial f(x_{k+1}).$$

But, computing x_{k+1} is a difficult task, it is usually impossible, unless f has an amenable structure, for example quadratic or piecewise linear. Bundle

methods provide a answer for remedying this problem. For details, werefer to [31]. Bundle methods need the following assumption:
Given $x \in X$, the value $f(x)$ and some $s(x) \in \partial f(x)$ are available. These methods consist of approximating the function f from below by using some simpler function (e.g., a piecewise linear function) ψ. This gives the following algorithm.

Algorithm 1 Prox-form of bundle methods.

0. An initial point $x_1 \in X$ is given, together with a tolerance $c \in]0,1[$ and a positive sequence $\{\alpha_k\}$. Set $k = l = 1$.
1. Choose a convex function $\psi^l : X \longrightarrow \mathbb{R}$. Solve for x

$$\inf_{x \in X} \left\{ \psi^l(x) + \frac{1}{2\alpha_k} \|x - x_k\|^2 \right\},$$

   to obtain the unique optimal solution x_k^l, as well as $\gamma_k^l := (x_k - x_k^l)/\alpha_k \in \partial \psi^l(x_k^l)$.
2. Compute $f(x_k^l)$. If a good decrease is obtained, namely if

$$f(x_k) - f(x_k^l) \geq c\left[f(x_k) - \psi^l(x_k^l)\right], \tag{1.7.1}$$

   then set $x_{k+1} = x_k^l$ and increase k by 1.
3. Increase l by 1 and go to Step1.

If the condition in Step 2 is satisfied, then we say that a descent-step has been made (a serious step), and x_k^l is significantly better than x_k. Indeed, we have that $\gamma_k^l \in \partial \psi^l(x_k^l)$, and then

$$\psi^l(x) \geq \psi^l(x_k^l) + \frac{1}{\alpha_k}(x_k - x_k^l)^\top (x - x_k^l), \quad \text{for all } x \in \mathbb{R}^n.$$

Since $\psi^l \leq f$, and by taking $x = x_k$ in the last inequality we obtain

$$f(x_k) \geq \psi^l(x_k^l) + \frac{1}{\alpha_k}\|x_k - x_k^l\|^2.$$

This implies that $f(x_k) \geq \psi^l(x_k^l)$. Therefore, $f(x_k) \geq f(x_k^l)$ by (1.7.1). Otherwise, a null-step has been made, the new ψ^{l+1} in the next iteration will supposedly improve the approximation of the true x_{k+1}.

A key-object is then the aggregate affine function

$$l^l(x) = \psi^l(x_k^l) + \langle \gamma_k^l, x - x_k^l \rangle.$$

Because $\gamma_k^l \in \partial\psi^l(x_k^l)$, it is easy to see that $l^l \leq \psi^l$.
The convergence of Algorithm 1 requires the following conditions on the functions $\{\psi^l\}$.

(C1) $\psi^l(x) \leq f(x)$ on X for $l = 1, 2, \ldots,$

(C2) $\psi^{l+1}(x) \geq l^l(x)$ on X for $l = 1, 2, \ldots,$

(C3) $\psi^{l+1}(x) \geq f(x_k^l) + \langle s(x_k^l), x - x_k^l \rangle$ on X for $l = 1, 2, \ldots,$

where $s(x_k^l)$ is a subgradient of f at x_k^l.

In the following we give some possible choices of $\psi^l(\cdot)$

Example 1.7.1. Consider the piecewise-affine model, defined for all $k \in \mathbb{N}$ and $l \in \mathbb{N}$, by

$$\psi^{l+1}(x) = \max_{0 \leq q \leq l} \left\{ f(x_k^q) + \langle s(x_k^q), x - x_k^q \rangle \right\}$$

for all $x \in X$ where $x_k^0 = x_k$.
Example 1.7.2. For all $k \in \mathbb{N}$ and $l \in \mathbb{N}$, we can choose, for all $x \in X$,

$$\psi^{l+1}(x) = \max\left\{ \mathcal{L}^l(x), f(x_k^l) + \langle s(x_k^l), x - x_k^l \rangle \right\}$$

where $x_k^0 = x_k$.
Example 1.7.3. For all $k \in \mathbb{N}$ and $l \in \mathbb{N}$, and $x \in X$, let

$$\psi^{l+1}(x) = \max\left\{ \mathcal{L}^l(x), \max_{0 \leq q \leq l} \left\{ f(x_k^q) + \langle s(x_k^q), x - x_k^q \rangle \right\} \right\}$$

where $x_k^0 = x_k$.

Chapter 2

DC Functions and DC programming

In this chapter, we first consider some important definitions and properties on DC functions, then we define the problems in DC programming, specially DC duality, global optimality in DC programming, local optimality in DC programming. At the end of this chapter, we give some real-world applications of DCA from [81].

2.1 DC Functions

A function $f : X \longrightarrow \overline{\mathbb{R}}$ defined on a convex set X will be called a DC function on X if there exists a pair of convex functions h and g on X such that

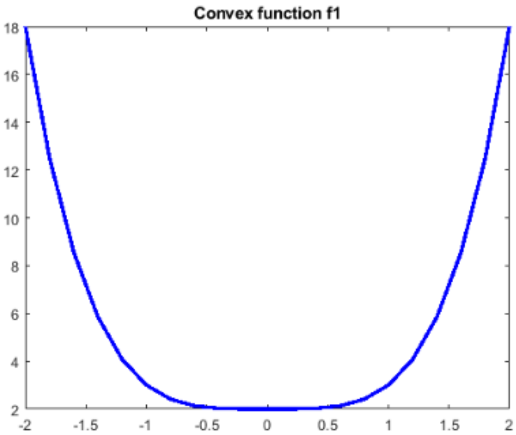
$$f(x) = g(x) - h(x).$$

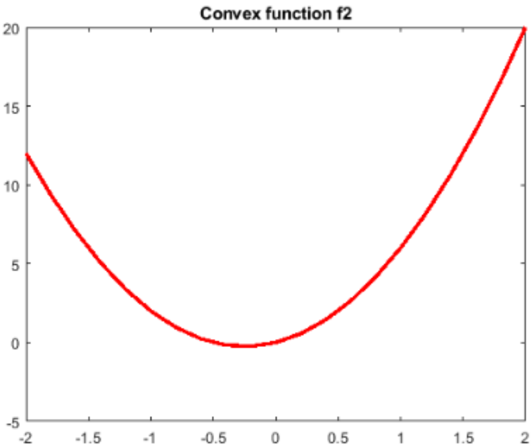
Of course, convex and concave functions are particular examples of DC functions.

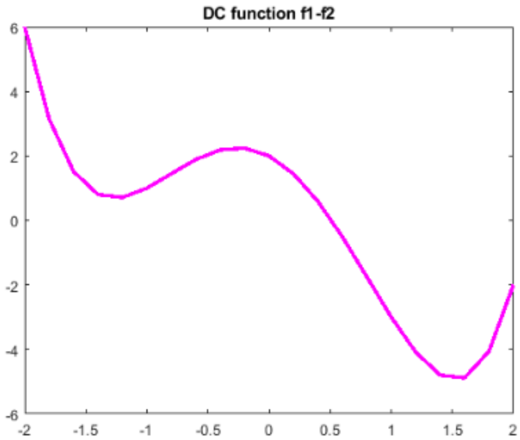
We denote by $DC(X)$ the set of DC functions on X, and by $DC_f(X)$ the case where the functions g and h are finite convex on X.

Example 2.1.1. In this example, we take $X = [-2, 2]$.

For the convex functions $f_1(x) = x^4 + 2$ and $f_2(x) = 4x^2 + 2x$, we get the nonconvex function $f_1(x) - f_2(x) = x^4 - 4x^2 - 2x + 2$.

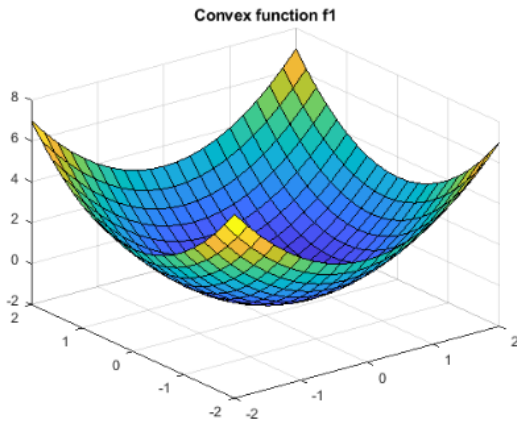


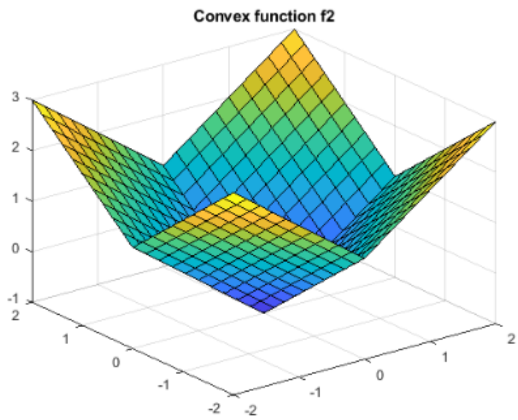


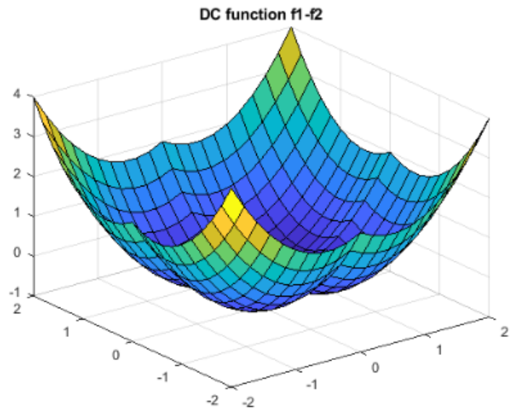


Example 2.1.2. In this example, we take $X = [-2,2] \times [-2,2]$.

For convex functions $f_1(x,y) = x^2 + y^2 - 1$ and $f_2(x,y) = |x| + |y| - 1$, we get the nonconvex function $f_1(x) - f_2(x) = x^2 + y^2 - |x| - |y|$.







DC functions have many important properties which were established from the 1950s by Alexandroff (1949), Landis (1951) and Hartman (1959), see [3, 20, 60]. One of the main properties is their relative stability to operations frequently used in optimization. More precisely we have the following properties. .

Proposition 2.1.1.

1. A linear combination of DC functions on X is DC on X.
2. The upper envelope of a finite set of DC functions with finite value on X is DC on X.
3. The lower envelope of a finite set of DC functions with finite value on X is DC on X.
4. Let $f \in DC_f(X)$, then $|f(x)|$, $f^+(x) = \max\{0, f(x)\}$, and $f^-(x) = \min\{0, f(x)\}$ are DC on X.
5. If $f_i \in DC_f(X)$, $i = 1, \cdots, m$, then $\max_{1 \le i \le m} f_i(x)$ is DC on X.

These results are generalized to the case of functions with values in $\overline{\mathbb{R}}$ [77]. It follows that the set of DC functions on X is a vector space (DC (X)). It is the smallest vector space containing the set of convex functions on X.

Remark 2.1.1. Given a DC function f and its DC representation $f = g - h$, then for any finite convex function ϕ, $f = (g + \phi) - (h + \phi)$ gives another DC representation of f. Thus, a DC function admits an infinity of DC decomposition.

Denote by $C^2(\mathbb{R}^n)$, the class of functions twice continuously differentiable on $\mathbb{R}^n$.

Proposition 2.1.2. Any function $f \in C^2(\mathbb{R}^n)$ is DC on any compact convex set $X \subset \mathbb{R}^n$.

Since the subspace of polynomials on X is dense in the space $C(X)$ of continuous numerical functions on X, we deduce.

Corollary 2.1.1. The space of DC functions on a compact convex set $X \subset \mathbb{R}^n$, is dense in $C(X)$, i.e.,

$$\forall \varepsilon > 0, \exists F \in C(X) : |f(x) - F(x)| \leq \varepsilon \ \forall x \in X.$$

Let us underline that the DC functions intervene very frequently in practice, as well in differentiable as nondifferentiable optimization. An important result established by Hartman [60] makes it possible to identify DC functions in many situations, simply by resorting to a local analysis of convexity (locally convex, locally concave, locally DC).
A function $f : X \longrightarrow \overline{\mathbb{R}}$ defined on an open convex set $X \subset \mathbb{R}^n$ is called locally DC if for all $x \in X$ there exists an open convex neighborhood U of x and a pair of convex functions g, h on U such that $f|_U = g|_U - h|_U$.

Proposition 2.1.3 : ***([60]).*** A function locally DC on a convex set X is DC on X.

A mapping $F = (f_1, f_2, \cdots, f_m) : X \longrightarrow \mathbb{R}^m$ defined on a convex set $X \subset \mathbb{R}^n$ is said to be DC on X: $F \in DC(X)$, if $f_i \in DC(X)$ for every $i = 1, 2, \cdots, m$.
Proposition 2.1.4 : ***([60], Proposition.4).*** Let $\Omega_1 \subset \mathbb{R}^n$, $\Omega_2 \subset \mathbb{R}^m$ be convex sets such that Ω_1 is open or closed and Ω_2 is open. If $F_1 : \Omega_1 \longrightarrow \Omega_2$, $F_2 : \Omega_2 \longrightarrow \mathbb{R}^k$ are DC mappings then $F_2 \circ F_1 : \Omega_1 \longrightarrow \mathbb{R}^k$ is also a DC mapping.
Corollary 2.1.2 : ***([60], Corollary.3).*** Let Ω_1 and Ω_1 be as in Proposition 2.1.4. If $F_1 : \Omega_1 \longrightarrow \Omega_2$ is DC on $\Omega_1 \subset \mathbb{R}^n$ and $F_2 : \Omega_2 \longrightarrow \mathbb{R}^k$ is C^2-smooth, then $F_2 \circ F_1$ is DC on Ω_1. In particular, the product of two DC functions is DC; if f is DC on an open (or closed) convex set X and $f(x) \neq 0$ for a $x \in X$ then $\frac{1}{f}$ and $|f|^{1/m}$ are DC on X.
Remark 2.1.2. A function f is called a factorable function if it results from a finite sequence of compositions of transformed sums and products of simple functions of one variable. In other words, a factorable function is a function which can be obtained as the last in a sequence of functions $f_1, f_2, ...,$ built up as follows:

$$f_i(x) = x_i \qquad (i = 1, 2, ..., n),$$

and for $k > n$, f_k is one of the forms:

$$f_k(x) = f_l(x) + f_j(x) \text{ for some } l, j < k$$
$$f_k(x) = f_l(x) \times f_j(x) \text{ for some } l, j < k$$
$$f_k(x) = F\left(f_j(x)\right) \text{ for some } j < k$$

where $F : \mathbb{R} \longrightarrow \mathbb{R}$ is a simple DC function of one variable, such as $F(t) = t^P$, $F(t) = e^t$, $F(t) = log|t|$, $F(t) = \sin t$ etc.

From the above corollary it also easily follows that a factorable function is a DC function.

A quadratic function $f(x) = x^\top Qx$ where Q is a symmetric matrix, is a DC function which may be neither convex nor concave.

Indeed, Q can be represented as the difference of two symmetric positive definite matrices $Q = Q_1 - Q_2$.

First, we represent the matrix Q via the difference of two matrices with nonnegative components: $Q = D_1 - D_2$, where

$$d_{ij}^{(1)} = \begin{cases} q_{ij} & \text{if } q_{ij} \geq 0, \\ 0 & \text{if } q_{ij} < 0, \end{cases} \qquad d_{ij}^{(2)} = \begin{cases} 0 & \text{if } q_{ij} \geq 0, \\ -q_{ij} & \text{if } q_{ij} < 0. \end{cases}$$

Second, we construct the matrices $\Gamma_1 = D_1 + \Lambda_1$, $\Gamma_2 = D_2 + \Lambda_1$, where Λ_1 is a diagonal matrix:

$$\lambda_{ii}^{(1)} = \begin{cases} 0 & \text{if } d_{ii}^{(1)} > S_i, \\ S_i - d_{ii}^{(1)} + \varepsilon & \text{if } d_{ii}^{(1)} \leq S_i, \end{cases}$$

where $S_i = \sum_{i \neq j} d_{ij}^{(1)}$ is the sum of nondiagonal elements of the row i in the matrix D_1, and the number $\varepsilon > 0$. Thus, Γ_1 is a positive definite matrix.

Similarly, we obtain $Q_1 = \Gamma_1 + \Lambda_2$, $Q_2 = \Gamma_2 + \Lambda_2$, where$\Lambda_2$ is a diagonal matrix:

$$\lambda_{ii}^{(2)} = \begin{cases} 0 & \text{if } d_{ii}^{(2)} > T_i, \\ T_i - d_{ii}^{(2)} + \varepsilon & \text{if } d_{ii}^{(2)} \leq T_i, \end{cases}$$

where $T_i = \sum_{i \neq j} d_{ij}^{(2)}$ is the sum of nondiagonal elements of the row i in the matrix D_2.

Hence, the matrix Q is represented as the difference $Q = Q_1 - Q_2$ of matrices Q_1 and Q_2 with non-negative components and dominant diagonals, and we get the DC representation of f

$$f(x) = x^\top Q_1 x - x^\top Q_2 x = g(x) - h(x),$$

where $g(\cdot)$ and $h(\cdot)$ are strongly convex functions (Q_1, Q_2 are positive definite).

2.2 DC programming

Due to the preponderance and the richness of the properties of the DC functions, the passage of the subspace of convex functions to the vector space DC(X) allows to extend significantly the convex optimization problems to the nonconvexity while maintaining an underlying structure fundamentally related to the convexity. The domain of optimization problems involving DC functions is thus relatively wide and open, covering most of the application problems encountered.
Thus we cannot deal with any non-convex and non-differentiable optimization problem.

The following classification has now become classic.

(1) $\sup\{f(x) : x \in X\}$, where f and X are convex,

(2) $\inf\{g(x) - h(x) : x \in \mathbb{R}^n\}$, where g and h are convex,

(3) $\inf\{g(x) - h(x) : x \in X, f_1(x) - f_2(x) \leq 0\}$, where g, h, f_1, f_2 and X are convex,

Problem (1) is a special case of problem (2) with $g = \text{Ind}_X$, the indicator function of X, and $h = f$. Problem (2) may be modeled in the equivalent form of (1)

$$\inf\{t - h(x) : g(x) - t \leq 0\}.$$

As for the problem (3) it can be transformed in the form (2) via the exact penalty relating to the constraint DC $f_1(x) - f_2(x) \leq 0$. Its resolution can also be reduced, under certain technical conditions, to that of a series of problems (1).

Problem (2) is commonly referred to as DC programming. It is of major interest both from a practical and theoretical point of view. From a theoretical point of view, we can underline that, as we saw above, the class of DC functions is remarkably stable compared to operations frequently used in optimization. In addition, we have an elegant theory of duality [38,39,75–77,125,127] which, as in convex optimization, has profound practical implications for numerical methods.

2.2.1 DC Duality

In convex analysis, the concept of duality (conjugate functions, dual problem, etc.) is a very powerful fundamental notion. For convex and in particular linear problems, a theory of duality has been developed for several decades [105]. More recently, in non-convex analysis important concepts of duality have been proposed and developed, first, for convex maximization problems, before arriving at DC problems. Thus the DC duality introduced by Toland [125] can be considered as a logical generalization of the work of Pham Dinh Tao [38] on convex maximization. We will present below the main results (in DC optimization) concerning the optimality conditions (local and global) and the DC duality. For more details, the reader is referred to the document by Le Thi [77] (see also [75]).

The dual space of $\mathbb{R}^n$, denoted by Y, can be identified with $\mathbb{R}^n$ itself. DC Programming address the problem of minimizing a function f which is a difference of convex functions on the whole space $\mathbb{R}^n$. Generally speaking, a so-called standard DC program takes the form

$$(P) \qquad \inf\{f(x) := g(x) - h(x) : x \in \mathbb{R}^n\},$$

with g and h are convex.
DC duality associates a primal DC program with its dual, which is also a DC program too

$$(D) \qquad \inf\{h^*(y) - g^*(y) : y \in Y\},$$

where ϕ^* defined by $\phi^*(y) := \sup\{\langle x, y\rangle - \phi(x) : x \in \mathbb{R}^n\}$, $\forall y \in Y$ is the conjugate of ϕ. There is so a perfect symmetry between (P) and its dual (D):

the dual of (D) is exactly (P).

2.2.2 Critical and strongly critical point

A point x^* is a critical point of (P) (or of $f = g - h$) if $\partial g(x^*) \cap \partial h(x^*) \neq \emptyset$, or equivalently $0 \in \partial g(x^*) - \partial h(x^*)$, where $\partial g(x^*) - \partial h(x^*) := \left\{ y_1^* - y_2^* \mid y_1^* \in \partial g(x^*), y_2^* \in \partial h(x^*) \right\}$. While it is called strongly critical point of (P) (or of $f = g - h$) if $\emptyset \neq \partial h(x^*) \subset \partial g(x^*)$.

The notion of DC criticality is close to Clarke stationarity critical point in the sense that the Clarke subdifferential $\partial^c f$ of $f = g - h$ verifies $\partial^c f(x) \subset [\partial g(x) - \partial h(x)]$ with equality under technical assumptions. Hence Clarke stationarity of x^* , i.e., $0 \in \partial^c f(x^*)$ implies DC criticality of x^*. We have an equivalence between these notions if the related equality holds in the corresponding inclusion.

This result of challenge DC duality using conjugate functions gives an important relationship in DC optimization [125].
Theorem 2.2.1. Let g and h be convex. Then

(i)
$$\inf_{x \in dom(g)} \{g(x) - h(x)\} = \inf_{y \in dom(h^*)} \{h^*(y) - g^*(y)\}$$

(ii) If $\bar{y}$ is a minimum of $h^* - g^*$ on Y then each $\bar{x} \in \partial g^*(\bar{y})$ is a minimum of $g - h$ over $\mathbb{R}^n$.

The previous theorem shows that solving the primal problem (P) involves solving the dual problem (D) and vice versa.

From the perfect symmetry between the primal problem (P) and the dual problem (D), it clearly appears that the results established for one are directly transposed to the other. However, we choose here not to present them simultaneously in order to simplify the presentation.

2.2.3 Global optimality in DC programming

In convex optimization, $\bar{x}$ minimizes a convex function f on $\mathbb{R}^n$ if and only if $0 \in \partial f(\bar{x})$. In DC optimization, the following global optimality condition [128] is formulated using ε-subdifferentials of g and h. Its proof (based on the study of the behavior of the ϵ-subdifferential of a convex function as a function of the parameter) is complicated. The proof in [77] is simpler and suits the DC optimization framework quite simply: it expresses quite simply that this global optimality condition is a geometric translation of the equality of optimal values in the primal and dual DC programs.

Theorem 2.2.2 : ***(Global optimality DC).*** Let $f = g - h$, such that g and h are convex. Then $\bar{x}$ is a global minimum of $g(x) - h(x)$ over $\mathbb{R}^n$ if and only if,

$$\partial_\epsilon h(\bar{x}) \subset \partial_\varepsilon g(\bar{x}) \quad \forall \varepsilon > 0. \tag{2.2.1}$$

Remark 2.2.1. (i) If f is convex, we can write $f = g - h$ with $f = g$ and $h = 0$. In this case the global optimality in (P) which is identical to the local optimality because (P) is a convex problem is characterized by

$$0 \in \partial f(\bar{x}), \tag{2.2.2}$$

Because $\partial_\varepsilon h(\bar{x}) = \partial h(\bar{x}) = \{0\}$, for all $\varepsilon > 0$ and $x \in \mathbb{R}^n$, and the growth of the ε-subdifferential according to, the relation (2.2.2) is equivalent to (2.2.1).

(ii) In a more general way, let us consider the DC decompositions of a convex function f of the form $f = g - h$ with $g = f + h$ and h convex finite everywhere on $\mathbb{R}^n$. The corresponding DC problem is a "false" DC problem because it is a convex problem. In this case, relation (2.2.2) is equivalent to

$$\partial h(\bar{x}) \subset \partial g(\bar{x}).$$

(iii) We can thus say that (2.2.1) clearly marks the transition from convex optimization to nonconvex optimization. This characteristic of the global optimality of (P) indicates at the same time all the complexity of its practical use because it appeals to all the ϵ-subdifferential in $\bar{x}$.

2.2.4 Local optimality in DC programming

We have seen that the relation $\partial h(\bar{x}) \subset \partial g(\bar{x})$ (using the "exact" subdifferential) is a necessary and sufficient condition of global optimality for a "false" DC problem (convex optimization problem). Now in a global optimization problem, the function to be minimized is locally convex "around" a local minimum, it is then clear that this subdifferential inclusion relation will make it possible to characterize a local minimum of a DC problem.

Definition 2.2.1. Let g and h be convex functions. A point $\bar{x} \in \text{dom}g \cap \text{dom}h$ is a local minimum of $g(x) - h(x)$ on $\mathbb{R}^n$ if and only if

$$g(x) - h(x) \geq g(\bar{x}) - h(\bar{x}) \ \forall x \in U_{\bar{x}},$$

where $U_{\bar{x}}$ is a neighborhood of $\bar{x}$.
Proposition 2.2.1 : ***(Necessary condition of local optimality).*** If $\bar{x}$ is a local minimum of $g - h$, then

$$\partial h(\bar{x}) \subset \partial g(\bar{x}).$$

It is important to formulate sufficient conditions for local optimality.

Theorem 2.2.3 : ***(Sufficient condition of local optimality ([75, 77])).*** if $\bar{x}$ admits a neighborhood U such that

$$\partial h(x) \cap \partial g(\bar{x}) \neq \emptyset, \ \forall x \in U \cap \text{dom}g,$$

then $\bar{x}$ is a local minimum of $g - h$.
Corollary 2.2.1. If $\bar{x} \in \text{int}(\text{dom}h)$ such that

$$\partial h(\bar{x}) \subset \text{int}\left(\partial g(\bar{x})\right),$$

then $\bar{x}$ is a local minimum of $g - h$.

A convex subset C is said to be polyedral convex if

$$C = \bigcap_{i=1}^{m} \{x : \langle a_i, x\rangle - \alpha_i \leq 0\}, \text{ where } a_i \in Y, \alpha_i \in \mathbb{R}, \ \forall i = 1, \cdots, m.$$

A function is said to be polyedral convex if

$$f(x) = \sup\{\langle a_i, x\rangle - \alpha_i : i = 1, \cdots, k\} + \text{Ind}_C$$

where C is a polyhedral convex subset and the symbol Ind_C denotes the indicator function of C, i.e. $\text{Ind}_C(x) = 0$ if $x \in C$ and $+\infty$ otherwise.

Corollary 2.2.2. If h is polyhedral convex then $\partial h(\bar{x}) \subset \partial g(\bar{x})$ is a necessary and sufficient condition for $\bar{x}$ to be a local minimum of $g - h$.

To solve a DC optimization problem, it is sometimes easier to solve the dual problem (D) than the primal problem (P). Theorem 2.2.1 provides the duality transport of global minima. We also establish the transport by duality of the local minima.

Proposition 2.2.2 : ***(Transport by DC duality of local minima ([75,77])).*** Let $\bar{x} \in$dom(∂h) be a local minimum of $g - h$. Let $\bar{y} \in \partial h(\bar{x})$ and $U_{\bar{x}}$ be a neighborhood of $\bar{x}$ such that $g(x) - h(x) \geq g(\bar{x}) - h(\bar{x})$, $\forall x \in U_{\bar{x}} \cap \text{dom}g$. If

$$\bar{x} \in \text{int}(\text{dom}(g^*)) \text{ and } \partial g^*(\bar{y}) \subset U_{\bar{x}},$$

then $\bar{y}$ is a local minimum of $h^* - g^*$.
Remark 2.2.2. Of course, by duality, all the results of this section are transposed to the dual problem D. For example:
if y is a local minimum of $h^* - g^*$, then $\partial g^*(y) \subset \partial h^*(y)$.

2.3 The real-world applications of DCA

DCA (DC Algorithms) was successfully applied to many large-scale DC optimization problems and proved to be more robust and efficient than related standard methods. To get more on applications of DCA, the reader can consult [81].

2.3.1 Applications in biology

DC programming and DCA were extensively developed for several challenging classes of problems in biology such that, Protein fold recognition, Phylogenetic tree reconstruction, and Molecular conformation, for more informations see [78, 82].

Yiming et al. [130] consider the problem of integrating multiple data sources using a kernel based approach. They propose a novel information-theoretic approach based on a Kullback-Leibler (KL) divergence between the output kernel matrix and the input kernel matrix so as to integrate heterogeneous data sources. One of the most appealing properties of this approach is that it can easily cope with multi-class classification and multi-task learning by an appropriate choice of the output kernel matrix. Based on the position of the output and input kernel matrices in the KL-divergence objective, this is a formulation which refer to as MKLdiv-dc. Yiming et al. propose to efficiently solve MKLdiv-dc by DC programming method. The effectiveness of the proposed approach is evaluated on a benchmark dataset for protein fold recognition and a yeast protein function prediction problem.

Kernel matrices (see, [130]) are generally positive semi-definite and thus can be regarded as the covariance matrices of Gaussian distributions. As described in [74], the Kullback-Leibler (KL) divergence (relative entropy) between a Gaussian distribution $\mathcal{N}(0, K_y)$ with the output covariance matrix K_y and a Gaussian distribution $\mathcal{N}(0, K_x)$ with the input kernel covariance matrix K_x is

$$KL\left(\mathcal{N}\left(0, K_y\right) \| \mathcal{N}\left(0, K_x\right)\right) := \frac{1}{2}\mathrm{Tr}\left(K_y K_x^{-1}\right) + \frac{1}{2}\log|K_x| - \frac{1}{2}\log\left|K_y\right| - \frac{n}{2}, \quad (2.3.1)$$

where, for any square matrix B, the notation $\mathrm{Tr}(B)$ denotes its trace. Though $KL\left(\mathcal{N}\left(0, K_y\right) \| \mathcal{N}\left(0, K_x\right)\right)$ is nonconvex, it has a unique minimum at $K_x = K_y$ if K_y is positive definite. If the input kernel matrix K_x is represented by a linear combination of m candidate kernel matrices, i.e. $K_x = K_\lambda = \sum_{l=1}^m \lambda_l K_l$, the above KL-divergence based kernel learning is reduced to the following formulation

$$KL\left(\mathcal{N}\left(0, K_y\right) \| \mathcal{N}\left(0, K_x\right)\right) = \mathrm{Tr}\left(K_y\left(\sum_{l=1}^m \lambda_l K_l + \sigma I_n\right)^{-1}\right) + \log\left|\sum_{l=1}^m \lambda_l K_l + \sigma I_n\right|, \quad (2.3.2)$$

to be minimized Δ, where $\Delta = \left\{\lambda = (\lambda_1, \cdots, \lambda_m) \mid \sum_{l=1}^m \lambda_l = 1, \lambda_l \geq 0\right\}$, I_n denotes the $n \times n$ identity matrix and $\sigma > 0$ is a supplemented small parameter to avoid the singularity of K_λ.

By [130, Theorem 1], the problem (2.3.2) is DC. Then

$$\begin{aligned} KL\left(\mathcal{N}\left(0, K_y\right) \| \mathcal{N}\left(0, K_x\right)\right) &= \mathrm{Tr}\left(K_y\left(\sum_{l=1}^m \lambda_l K_l + \sigma I_n\right)^{-1}\right) + \log\left|\sum_{l=1}^m \lambda_l K_l + \sigma I_n\right|, \\ &= g(\lambda) - h(\lambda), \end{aligned} \quad (2.3.3)$$

where $g(\lambda) = \mathrm{Tr}\left(K_y\left(\sum_{l=1}^m \lambda_l K_l + \sigma I_n\right)^{-1}\right)$, $h(\lambda) = -\log\left|\sum_{l=1}^m \lambda_l K_l + \sigma I_n\right|$, g and h are convex with respect to $\lambda \in \Delta$.

The problem (2.3.3) can be solved by a difference of convex algorithm DCA. This procedure iteratively solves the following convex problem

$$\lambda^{(t+1)} = \underset{\lambda \in \Delta}{\operatorname{argmin}} \left\{ g(\lambda) - h\left(\lambda^{(t)}\right) - \nabla h\left(\lambda^{(t)}\right)\left(\lambda - \lambda^{(t)}\right) \right\}.$$

MKLdiv-dc further improves the fold discrimination accuracy to 75.19% which is a more than 5% improvement over competitive Bayesian probabilistic and SVM margin-based kernel learning methods [130].

2.3.2 Applications in Machine learning and data mining

Machine Learning and Data Mining (MLDM) represent a mine of optimization problems that are almost all DC programs for which appropriate solution methods should use DC programming and DCA. DC programming and DCA have been applied to modeling and solving many problems in MLDM.

Clustering is a fundamental problem in unsupervised learning and has many applications in various domains. DCA is investigated in various works of clustering. The first work was devoted to hard (partitional) clustering via the most popular formulation, the so-called minimum sum-of-squares clustering (MSSC) of the form, see [4],

$$\min \left\{ \sum_{i=1}^{m} \min_{1 \le l \le k} \|x_l - a_i\|^2, x_l \in \mathbb{R}^n, l = 1, 2, \cdots, k \right\}, \tag{2.3.4}$$

where $\|\cdot\|$ denotes the Euclidean norm, $a_i \in \mathbb{R}^n$, for $i = 1, \cdots, m$.

For solving the problem (2.3.4). Le et al. [4], recast this problem in the matrix vector space $\mathbb{R}^{k \times n}$ of $(k \times n)$ real matrices. The variables are then $X \in \mathbb{R}^{k \times n}$ whose ith row X_i is equal to x_i for $i = 1, \cdots, k$. The Euclidean structure of $\mathbb{R}^{k \times n}$ is defined with the help of the usual scalar product

$$\mathbb{R}^{k \times n} \ni X \longleftrightarrow (X_1, X_2, \cdots, X_k) \in (\mathbb{R}^n)^k, \quad X_i \in \mathbb{R}^n, i = 1, \cdots, k,$$

$$\langle X, Y \rangle := \operatorname{Tr}\left(X^\top Y\right) = \sum_{i=1}^{k} \langle X_i, Y_i \rangle,$$

and its Euclidean norm $\|X\|^2 := \sum_{i=1}^{k} \langle X_i, X_i \rangle = \sum_{i=1}^{k} \|X_i\|^2$.

According to the property

$$\min_{1 \le l \le k} \|x_l - a_i\|^2 = \sum_{l=1}^{k} \|x_l - a_i\|^2 - \max_{1 \le r \le k} \sum_{l=1, l \ne r} \|x_l - a_i\|^2,$$

and the convexity of the functions

$$\sum_{l=1}^{k} \|x_l - a_i\|^2, \quad \max_{1\le r\le k} \sum_{l=1, l\neq r} \|x_l - a_i\|^2,$$

we can say that clustering problem (2.3.4) is a DC program. More precisely, it can be expressed in the matrix space $\mathbb{R}^{k\times n}$ as follows

$$\min\left\{G(X) - H(X), \quad X \in \mathbb{R}^{k\times n}\right\},$$

where the DC components G and H are given by

$$G(X) = \sum_{i=1}^{m}\sum_{l=1}^{k} G_{il}(X),\, G_{il}(X) = \tfrac{1}{2}\|X_l - a_i\|^2 \ \text{ for } i = 1,\cdots,m,\, l = 1,\cdots,k,$$

and

$$H(X) = \sum_{i=1}^{m} H_i(X),\, H_i(X) = \max_{1\le r\le k} \sum_{l=1, l\neq r} \tfrac{1}{2}\|X_l - a_i\|^2 \ \text{ for } i = 1,\cdots,m.$$

2.3.3 Applications in Communication system and network optimization

Nonconvex programming becomes an indispensable and powerful tool for design of communication systems (CS) since the last decade. DC programming and DCA are increasingly used in this field. For example network utility maximization (NUM) has many applications in network rate allocation algorithms and internet congestion control protocols. Hoai et al. [79] consider a communication network with L links, each with a fixed capacity of c_l bps, and S sources (i.e., end users), each transmitting at a source rate of x_s bps. Each source s emits one flow, using a fixed set $L(s)$ of links in its path, and has a utility function $U_s(x_s)$. Each link l is shared by a set of sources denoted $S(l)$ (the set of users using link l). NUM, in its basic version, consists of maximizing the total utility of the network $\sum_s U_s(x_s)$ over the source rates x, subject to linear flow constraints for all links l

$$\max_{x\in K}\left\{\sum_{s\in\mathcal{S}} U_s(x_s)\right\}, \tag{2.3.5}$$

where $K = \left\{x = (x_s)_{s\in\mathcal{S}} \in \mathbb{R}^S \mid \sum_{s\in S(l)} x_s \le c_l, \forall l \in \{1,\cdots,L\}, x_s \ge 0, \forall s \in \mathcal{S}\right\}$, $\mathcal{S}$ denotes the set of users. Here the constraint set is a well-defined convex polytope. Hoai et al. consider the NUM problem with Sigmoidal-like utility functions that are used in many multimedia applications and internet congestion control (for example, the utility for voice applications is modeled by a Sigmoidal function with a convex part at low rate and a concave part at high rate). Sigmoidal utilities in a standard form

$$U_s(x_s) = \frac{1}{1+e^{-(a_s x_s + b_s)}},$$

where $a_s > 0$, $b_s < 0$ and a_s, b_s are integers. The Sigmoidal function is neither convex nor concave, but it is DC. Then the resulting NUM problem is a DC program. The DC decomposition for the Sigmoidal function is

$$U_s(x_s) = e^{(a_s x_s + b_s)} - \frac{e^{2(a_s x_s + b_s)}}{1+e^{(a_s x_s + b_s)}} = h_s(x_s) - g_s(x_s),$$

where $h_s(\cdot)$ and $g_s(\cdot)$ are convex (their derivative is increasing). Therefore, U_s is a DC function, and so is $-U_s$.

Denote by $\mathrm{Ind}_K(\cdot)$ the indicator function on K. Then the Sigmoidal NUM problem can be expressed as

$$\begin{aligned}\max_{x\in K}\left\{U(x) := \sum_{s\in\mathcal{S}} U_s(x_s)\right\} &= -\min_{x\in K}\left\{\sum_{s\in\mathcal{S}}\left[\frac{e^{2(a_s x_s + b_s)}}{1+e^{(a_s x_s + b_s)}} - e^{(a_s x_s + b_s)}\right]\right\},\\ &= -\min_{x\in K}\{g(x) - h(x) : x \in K\},\end{aligned}$$

where $g(x) := \sum_{s\in\mathcal{S}} g_s(x_s)$ and $h(x) := \sum_{s\in\mathcal{S}} h_s(x_s)$. Since g_s and h_s are convex functions, the function g and h are convex too (note also that g and h are differentiable). Hence the Sigmoidal NUM problem is a DC program that can be written in the standard form as

$$\min\left\{[\mathrm{Ind}_K(x) + g(x)] - h(x) : x \in \mathbb{R}^S\right\}.$$

We have presented DC programming and DCA for modeling and solving this nonconvex program in communication systems, for more details see [79].

Chapter 3

A DC Approach for Minimax Fractional Optimization Programs with Ratios of Convex Functions

This chapter deals with minimax fractional programs whose objective functions are the maximum of finite ratios of convex functions, with arbitrary convex constraints set. For such problems, Dinkelbach-type algorithms fail to work since the parametric subprolems may be nonconvex, whereas the latters need a global optimal solution of these subproblems. We give necessary optimality conditions for such problems, by means of convex analysis tools. We then propose a method, based on solving approximately a sequence of parametric convex problems, which acts as DC (difference of convex functions) algorithm, if the parameter is positive and as Dinkelbach algorithm if not. We show that every cluster point of the sequence of optimal solutions of these subproblems satisfies necessary optimality conditions of KKT criticality type, that are also of Clarke stationarity type [57].

3.1 Introduction

In this work we analyse optimality conditions and develop an algorithm for finding a solution of minimax or generalized fractional programming problems (GFP) of the form

$$(P) \qquad \bar{\lambda} = \inf_{x \in X} \left\{ \lambda(x) := \max_{i \in I} \frac{f_i(x)}{g_i(x)} \right\},$$

where I is a finite index set, the functions f_i and g_i, for $i \in I$, are convex, and X is a nonempty and convex subset of $\mathbb{R}^n$. The functions g_i, for $i \in I$, are assumed to be positive on X.

For solving a GFP, there have been several primal Dinkelbach-type algorithms in the literature [17, 33, 35–37, 107, 108, 122], and dual algorithms and results [1, 2, 13–15, 21, 22, 24–26, 34, 42, 43, 67]. See also [118–120] for more references on fractional programming. These algorithms are based on auxiliary parametric problems having simpler structures than the original problem. For the primal algorithms, the auxiliary problems furnish sequences of approximate optimal values converging decreasingly to the optimal value of (P), whereas the sequences of values generated by the dual algorithms converge increasingly towards the optimal value of (P). But in all these algorithms and results, convexity of the parametric problems is required. For primal (resp. dual) approaches, this is achieved in particular when the functions $f_i - \mu g_i$, for all $i \in I$, are convex for all $\mu \geq \bar{\lambda}$ (resp. $\mu \leq \bar{\lambda}$). The usual assumptions to get such property is that the function g_i concave and f_i convex and nonnegative, if g_i is not affine. Apart the situation when the parametric subproblems are convex, these approaches fail to work since it is assumed that one can compute their global minimum. In our situation, where both f_i and g_i are convex, the function $f_i - \mu g_i$ is convex when $\mu \leq 0$ and is a difference of convex (DC) functions otherwise. For this reason, we want to modify the last mentioned parametric subproblems in such a way we take into account these two situations. The first situation does not pose special problems, regarding convexity. To deal with the second situation, we resort to DC techniques, see e.g. [80, 124], where it is question to replace the function g_i by a linear approximation, obtained by a subgradient at current point. Doing so we obtain convex parametric subproblems.

The most important difficulty in global optimization, and in particular in DC programming, is how to recognize a global minimum, or even how to recognize local minimum, in contrast with the convex programming where a local minimum is global. The most common necessary optimality condition for the minimization of $f_1 - f_2$, say, over the whole space $\mathbb{R}^n$ is DC criticality, which means that a point x^* is DC critical if $\partial f_1(x^*) \cap \partial f_2(x^*) \neq \emptyset$, or equivalently if $0 \in [\partial f_1(x^*) - \partial f_2(x^*)]$, where $\partial f_i(x^*)$ stands for the subdiferential of the convex function f_i, $i = 1, 2$, at x^*. It is strongly critical if $\partial f_2(x^*) \subset \partial f_1(x^*)$. These notions were introduced since DC algorithms find such points. Generally, criticality does not imply Clarke stationarity, i.e. $0 \in \partial^c[f_1 - f_2](x^*)$, where $\partial^c[f_1 - f_2](x^*)$ is the Clarke subdifferential of $f_1 - f_2$ at x^* [30], since the inclusion $\partial^c[f_1 - f_2](x^*) \subset [\partial f_1(x^*) - \partial f_2(x^*)]$ may be strict. However, equality can occur in some situations, in particular if f_1 or f_2 is differentiable, see e.g. [30, Corollary 2 of Proposition 2.3.1 and 2.3.3]. To get more on optimality conditions for DC unconstrained and DC constrained programs, the reader can consult [5, 62, 80, 124].

To deal with our problem, we will show first that (P) is equivalent to a dc problem. By writing necessary optimality conditions that are of criticality type, for the latter, we obtain necessary optimality conditions, of Clarke stationarity type, for (P). Hence, we describe our DC Dinkelbach-type algorithm and establish its convergence to a Clarke stationary point.

3.2 New Parametric Approach for convex/convex GFP

In Dinkelbach-type algorithms, the associated parametric problem to (P), with the parameter μ, takes the form

$$(P_\mu) \qquad F(\mu) := \inf_{x \in X} \max_{i \in I} \{f_i(x) - \mu g_i(x)\}.$$

We recall that under mild assumptions, when $\mu = \bar{\lambda}$, the problems (P) and (P_μ) have the same optimal solutions set, see e.g. [35, Proposition 2.1] and [21, Lemma 1].

The problems (P_μ) appear as subproblems in the Dinkelbach-type algorithm [35, 36]. This algorithm acts as follows: from an arbitrary point $x^0 \in X$ a sequence of points x^k is generated by solving a sequence of subproblems (P_{λ_k}), where $\lambda_k = \lambda(x^k)$ and x^{k+1} is an optimal solution of (P_{λ_k}). Habitual assumptions to apply Dinkelbach-type algorithm is that the functions f_i are convex and g_i concave, with f_i positive if g_i not affine. This implies that $f_i - \lambda_k g_i$ is convex. Apart this case where (P_{λ_k}) is convex, it is difficult to find a global optimal solution to (P_{λ_k}), and the method fails to work. This is the case for our problem (P) for which the subproblems (P_{λ_k}) may be nonconvex, since both f_i and g_i are convex.

Our objective in this chapter is to escape the difficulty caused by the non-convexity of the parametric subproblems, and propose a dc Dinkelbach-type algorithm to solve programs whose objective function is the maximum of several ratios of convex functions. The subproblems in this method are convex. The idea behind our proposition is to linearize the functions g_i for $i \in I$, at current points x^k, when $\lambda_k > 0$ and to keep the same subproblem (P_{λ_k}) when $\lambda_k \le 0$. In such a way our procedure coincides with Dinkelbach-type algorithm if $\lambda_k \le 0$ for all k, but acts as a DC algorithm if $\lambda_k > 0$. To develop our algorithm we begin by defining the parametrized subproblems and give some related results.

In all what follows, $\partial f_i(x)$ and $\partial g_i(x)$ stand respectively for the subdifferentials of the functions f_i and g_i at x.

For $y \in \mathbb{R}^n$ and $\gamma_i \in \partial g_i(y)$, for $i \in I$, we define the function parametrized by y,

$$h_y(x) := \begin{cases} \max\limits_{i\in I} [f_i(x) - \lambda(y)(g_i(y) + \langle \gamma_i, x - y\rangle)] & \text{if } \lambda(y) > 0 \\ \max\limits_{i\in I} [f_i(x) - \lambda(y) g_i(x)] & \text{if } \lambda(y) \le 0 \end{cases}$$

and we denote by x_y its global minimum over X, if any.

Observe that with the assumptions made on f_i and g_i, the function h_y is convex for all $y \in \mathbb{R}$. On the other hand, if the functions g_i, $i \in I$, are affine, the two expressions in the definition of $h_y(x)$ are identical.

Before discussing optimality conditions for problem (P), we begin first by some preliminary results.

Lemma 3.2.1. For all $x \in \mathbb{R}^n$ and $y \in X$ we have

$$\max_{i\in I} [f_i(x) - \lambda(y) g_i(x)] \le h_y(x).$$

Proof. If $\lambda(y) \le 0$ there is no thing to show. Assume now that $\lambda(y) > 0$. From the subgradient inequalities $g_i(y) + \langle \gamma_i, x - y\rangle \le g_i(x)$, for all $i \in I$, and the definition of h_y, we get

$$f_i(x) - \lambda(y) g_i(x) \le f_i(x) - \lambda(y)(g_i(y) + \langle \gamma_i, x - y\rangle) \le h_y(x)$$

from which we obtain the desired inequality. □

Proposition 3.2.1. For all $y \in X$ we have

1. $h_y(y) = 0$ and $h_y(x_y) \le 0$.
2. $\lambda(x_y) \le \lambda(y)$.

Proof. 1. From the definition of h_y, we get

$$h_y(y) = \max_{i \in I} [f_i(y) - \lambda(y) g_i(y)].$$

From the definition of $\lambda(y)$ we have $f_i(y)/g_i(y) \leq \lambda(y)$ for all $i \in I$, with equality for some i_0, or equivalently $f_i(y) - \lambda(y) g_i(y) \leq 0$ with equality for $i = i_0$. This shows that $h_y(y) = 0$. On the other hand, the definition of x_y implies that $h_y(x_y) \leq h_y(x)$ for all $x \in X$. In particular, for $x = y$ we get $h_y(x_y) \leq h_y(y) = 0$.

2. From Lemma 3.2.1, with $x = x_y$, and Item 1 we have $f_i(x_y) - \lambda(y) g_i(x_y) \leq h_y(x_y) \leq 0$ for all $i \in I$. This means that $f_i(x_y)/g_i(x_y) \leq \lambda(y)$ for all $i \in I$, implying that $\lambda(x_y) \leq \lambda(y)$. □

Proposition 3.2.2. If the problem (P) has a global optimal solution on X, say $\bar{x}$, then this solution actually globally minimizes $h_{\bar{x}}$ over X, whatever are $\bar{\gamma}_i \in \partial g_i(\bar{x})$, $i \in I$. Conversely, for all global optimal solution $\bar{x}$ of (P), every optimal solution $x_{\bar{x}}$ of $h_{\bar{x}}$ over X also globally solves the problem (P).

Proof. Let $\bar{x} \in X$ be a global optimal solution of (P). Let $x \in X$ and $i \in I$ be such that $\lambda(x) = f_i(x)/g_i(x)$. Then $\lambda(\bar{x}) \leq \lambda(x) = f_i(x)/g_i(x)$. This implies that $f_i(x) - \lambda(\bar{x}) g_i(x) \geq 0$. By using Lemma 3.2.1, with $y = \bar{x}$, we conclude that $h_{\bar{x}}(x) \geq 0$ for all $x \in X$. The conclusion follows since $h_{\bar{x}}(\bar{x}) = 0$, from Proposition 3.2.1, Item 1 with, $y = \bar{x}$. The converse follows directly from Proposition 3.2.1, Item 2, with $y = \bar{x}$. □

3.3 Optimality Conditions

Now we are ready to give optimality conditions for (P).

Theorem 3.3.1. Let $\hat{x} \in X$. Then there exist $\hat{\gamma}_i \in \partial g_i(\hat{x})$, $i \in I$, such that $h_{\hat{x}}(x_{\hat{x}}) = 0$, or equivalently, such that $\hat{x}$ minimizes $h_{\hat{x}}$ over X, where

$$h_{\hat{x}}(x) := \begin{cases} \max_{i \in I} [f_i(x) - \lambda(\hat{x})(g_i(\hat{x}) + \langle \hat{\gamma}_i, x - \hat{x} \rangle)] & \text{if } \lambda(\hat{x}) > 0 \\ \max_{i \in I} [f_i(x) - \lambda(\hat{x}) g_i(x)] & \text{if } \lambda(\hat{x}) \leq 0 \end{cases}$$

if and only if, there exist $\hat{\mu}_i \geq 0$, $i \in I$, with $\sum_{i \in I} \hat{\mu}_i = 1$ such that

$$0 \in \sum_{i \in I} \hat{\mu}_i [\partial f_i(\hat{x}) - \lambda(\hat{x}) \partial g_i(\hat{x})] + N_X(\hat{x}),$$

and $\hat{\mu}_i [f_i(\hat{x}) - \lambda(\hat{x}) g_i(\hat{x})] = 0$. Moreover, if $\lambda(\hat{x}) \leq 0$ then $\hat{x}$ is an optimal solution for (P).

Proof. Let $\hat{\gamma}_i \in \partial g_i(\hat{x})$, $i \in I$, and let $h_{\hat{x}}$ as announced in the theorem. Assume that $h_{\hat{x}}(x_{\hat{x}}) = 0$, where we racall that $x_{\hat{x}}$ is a global minimum of $h_{\hat{x}}$ over X. Then since from Proposition 3.2.1, Item 1 with $y = \hat{x}$, we have $h_{\hat{x}}(\hat{x}) = 0$, we conclude that $\hat{x}$ is also a global minimum over the convex set X, of the convex function $h_{\hat{x}}$. This is equivalent to saying that

$$0 \in \partial h_{\hat{x}}(\hat{x}) + N_X(\hat{x}),$$

see, e.g. [64, Theorem 1.1.1], where $\partial h_{\hat{x}}(\hat{x})$ is the subdifferential of $h_{\hat{x}}$ at $\hat{x}$ and $N_X(\hat{x})$ the normal cone to X at $\hat{x}$. If $\lambda(\hat{x}) > 0$, then from the expression of $h_{\hat{x}}$ and by referring to [64, Corollary 4.3.2], there exist $\hat{\mu}_i \geq 0$, $i \in I$, with $\sum_{i\in I} \hat{\mu}_i = 1$ such that

$$0 \in \sum_{i\in I} \hat{\mu}_i \left[\partial f_i(\hat{x}) - \lambda(\hat{x})\hat{\gamma}_i\right] + N_X(\hat{x}), \tag{3.3.1}$$

and

$$\sum_{i\in I} \hat{\mu}_i \left[f_i(\hat{x}) - \lambda(\hat{x}) g_i(\hat{x})\right] = h_{\hat{x}}(\hat{x}) = 0. \tag{3.3.2}$$

It is clear that (3.3.1) implies that

$$0 \in \sum_{i\in I} \hat{\mu}_i \left[\partial f_i(\hat{x}) - \lambda(\hat{x})\partial g_i(\hat{x})\right] + N_X(\hat{x}). \tag{3.3.3}$$

If $\lambda(\hat{x}) \leq 0$, then again from the expression of $h_{\hat{x}}$ and [64, Corollary 4.3.2], there exist $\hat{\mu}_i \geq 0$, $i \in I$, with $\sum_{i\in I} \hat{\mu}_i = 1$ such that

$$0 \in \sum_{i\in I} \hat{\mu}_i \partial\left[f_i - \lambda(\hat{x}) g_i\right](\hat{x}) + N_X(\hat{x}), \tag{3.3.4}$$

where $\partial\left[f_i - \lambda(\hat{x}) g_i\right](\hat{x})$ is the subdifferential of $f_i - \lambda(\hat{x}) g_i$ at $\hat{x}$, and

$$\sum_{i\in I} \hat{\mu}_i \left[f_i(\hat{x}) - \lambda(\hat{x}) g_i(\hat{x})\right] = h_{\hat{x}}(\hat{x}) = 0.$$

Relation (3.3.4) implies that

$$0 \in \sum_{i\in I} \hat{\mu}_i \left[\partial f_i(\hat{x}) - \lambda(\hat{x})\partial g_i(\hat{x})\right] + N_X(\hat{x}),$$

since $\lambda(\hat{x}) \leq 0$, and thus $\partial\left[f_i - \lambda(\hat{x}) g_i\right](\hat{x}) = \partial f_i(\hat{x}) - \lambda(\hat{x})\partial g_i(\hat{x})$.

To show the converse, assume that we have (3.3.2) and (3.3.3). Then from (3.3.3), for all $i \in I$, there exist $\hat{\gamma}_i^f \in \partial f_i(\hat{x})$, $\hat{\gamma}_i^g \in \partial g_i(\hat{x})$ such that

$$\sum_{i\in I} \hat{\mu}_i \left(\hat{\gamma}_i^f - \lambda(\hat{x})\hat{\gamma}_i^g\right) \in -N_X(\hat{x}),$$

that is

$$\left\langle \sum_{i\in I} \hat{\mu}_i \left(\hat{\gamma}_i^f - \lambda(\hat{x})\hat{\gamma}_i^g\right), x - \hat{x} \right\rangle \geq 0 \quad \text{for all } x \in X. \tag{3.3.5}$$

If $\lambda(\hat{x}) > 0$, then by using the subgradient inequality for f_i, $i \in I$, we get $f_i(x) \geq f_i(\hat{x}) + \langle \hat{\gamma}_i^f, x - \hat{x}\rangle$. It follows that

$$f_i(x) - \lambda(\hat{x})\left(g_i(\hat{x}) + \langle \hat{\gamma}_i^g, x - \hat{x}\rangle\right) \geq f_i(\hat{x}) - \lambda(\hat{x})g_i(\hat{x}) + \langle \hat{\gamma}_i^f - \lambda(\hat{x})\hat{\gamma}_i^g, x - \hat{x}\rangle$$

for all $i \in I$. Now if $\lambda(\hat{x}) \leq 0$, then $\hat{\gamma}_i^f - \lambda(\hat{x})\hat{\gamma}_i^g \in \partial[f_i - \lambda(\hat{x})g_i](\hat{x})$, and thus

$$f_i(x) - \lambda(\hat{x})g_i(x) \geq f_i(\hat{x}) - \lambda(\hat{x})g_i(\hat{x}) + \langle \hat{\gamma}_i^f - \lambda(\hat{x})\hat{\gamma}_i^g, x - \hat{x}\rangle$$

By invoking the definition of $h_{\hat{x}}(x)$ we get

$$h_{\hat{x}}(x) \geq f_i(\hat{x}) - \lambda(\hat{x})g_i(\hat{x}) + \langle \hat{\gamma}_i^f - \lambda(\hat{x})\hat{\gamma}_i^g, x - \hat{x}\rangle. \tag{3.3.6}$$

Multiplying both sides of (3.3.6) by $\hat{\mu}_i$, for all $i \in I$, and summing, we obtain, taking into account (3.3.2) and (3.3.5), that $h_{\hat{x}}(x) \geq 0$ for all $x \in X$. Since $h_{\hat{x}}(\hat{x}) = 0$, we conclude that $h_{\hat{x}}(x_{\hat{x}}) = 0$, where $x_{\hat{x}}$ is a minimum of $h_{\hat{x}}$ over X, which gives the desired result.

Now we will show that $\hat{x}$ is an optimal solution for (P) when $\lambda(\hat{x}) \leq 0$. So, from (3.3.4) we deduce that for all $i \in I$ there exists $\hat{\eta}_i \in \partial[f_i - \lambda(\hat{x})g_i](\hat{x})$ such that $-\sum_{i\in I} \hat{\mu}_i\hat{\eta}_i \in N_X(\hat{x})$. It follows, taking into account (3.3.2), that

$$\begin{aligned}
\sum_{i\in I} \hat{\mu}_i \left[f_i(x) - \lambda(\hat{x})g_i(x)\right] &\geq \sum_{i\in I} \hat{\mu}_i \left[f_i(\hat{x}) - \lambda(\hat{x})g_i(\hat{x})\right] + \left\langle \sum_{i\in I} \hat{\mu}_i\hat{\eta}_i, x - \hat{x} \right\rangle \\
&= \left\langle \sum_{i\in I} \hat{\mu}_i\hat{\eta}_i, x - \hat{x} \right\rangle \\
&\geq 0 \quad \text{for all } x \in X.
\end{aligned}$$

But since for all $x \in X$ we have

$$\max_{i\in I} \left[f_i(x) - \lambda(\hat{x})g_i(x)\right] \geq \sum_{i\in I} \hat{\mu}_i \left[f_i(x) - \lambda(\hat{x})g_i(x)\right],$$

we obtain

$$\max_{i\in I} \left[f_i(x) - \lambda(\hat{x})g_i(x)\right] \geq 0 \quad \text{for all } x \in X.$$

For $x \in X$ there exists $i \in I$ such that $f_i(x) - \lambda(\hat{x})g_i(x) \geq 0$, or equivalently $f_i(x)/g_i(x) \geq \lambda(\hat{x})$, imlpying that $\lambda(x) \geq \lambda(\hat{x})$ for all $x \in X$. This says that $\hat{x}$ is an optimal solution for (P). □

Remark 3.3.1. Another way to show the sufficiency part in the case $\lambda(\hat{x}) \leq 0$, is to remark that in this case we have

$$h_{\hat{x}}(x) = \max_{i\in I}[f_i(x) - \lambda(\hat{x})g_i(x)].$$

The fact that $\hat{x}$ minimizes $h_{\hat{x}}$ over X, implies that

$$\begin{aligned} F(\lambda(\hat{x})) &= \inf_{x\in X}\max_{i\in I}[f_i(x) - \lambda(\hat{x})g_i(x)] \\ &= \max_{i\in I}[f_i(\hat{x}) - \lambda(\hat{x})g_i(\hat{x})] \\ &= h_{\hat{x}}(\hat{x}) = 0. \end{aligned}$$

It is well-known that $F(\lambda(\hat{x})) = 0$ implies that $\hat{x}$ is an optimal solution for (P).

Now we will show that KKT criticality conditions (3.3.2) and (3.3.3) of Theorem 3.3.1 are in fact Clarke stationary ones.

Theorem 3.3.2. Let $\hat{x} \in X$ and assume that $-g_i$ for $i \in I$, are Clarke regular when $\lambda(\hat{x}) > 0$. Then the conditions of Theorem 3.3.1 imply the Clarke stationarity of $\hat{x}$, that is, $0 \in \partial^c\lambda(\hat{x}) + N_X(\hat{x})$, where $\partial^c\lambda(\hat{x})$ is the Clarke subdifferential of λ at $\hat{x}$ and $N_X(\hat{x})$ the normal cone to X at $\hat{x}$.

Proof. We showed in Theorem 3.3.1 that if $\lambda(\hat{x}) \leq 0$ then $\hat{x}$ is an optimal solution for (P). Thus, by using [11, Theorem 4.7] we conclude that $0 \in \partial^c\lambda(\hat{x}) + N_X(\hat{x})$.

Assume now that $\lambda(\hat{x}) > 0$ and let $\hat{\mu}_i$, $i \in I$, be as in Theorem 3.3.1. Observe that since $\hat{\mu}_i\,[f_i(\hat{x}) - \lambda(\hat{x})g_i(\hat{x})] = 0$, for all $i \in I$ it follows that $\hat{\mu}_i = 0$ if $f_i(\hat{x}) - \lambda(\hat{x})g_i(\hat{x}) < 0$ (or equivalently if $f_i(\hat{x})/g_i(\hat{x}) < \lambda(\hat{x})$). Therefore, we can write (3.3.3) as

$$0 \in \sum_{i\in I}\hat{\mu}_i\left[\partial f_i(\hat{x}) - \frac{f_i(\hat{x})}{g_i(\hat{x})}\partial g_i(\hat{x})\right] + N_X(\hat{x}). \tag{3.3.7}$$

It is clear that we can write (3.3.7) as follows

$$0 \in \sum_{i\in I}\hat{\mu}_i g_i(\hat{x})\left[\frac{g_i(\hat{x})\partial f_i(\hat{x}) - f_i(\hat{x})\partial g_i(\hat{x})}{g_i(\hat{x})^2}\right] + N_X(\hat{x}). \tag{3.3.8}$$

Since $\lambda(\hat{x}) > 0$, the previous discussion showed that if $\hat{\mu}_i \neq 0$ then $\lambda(\hat{x}) = f_i(\hat{x})/g_i(\hat{x})$ implying that $f_i(\hat{x}) > 0$, since by our assumption g_i is positive on X. Since $-g_i$ for $i \in I$, are Clarke regular, it follows from [30, Proposition 2.3.14] that for all $i \in I$ such that $\hat{\mu}_i \neq 0$, we have

$$\partial^c\left[\frac{f_i}{g_i}\right](\hat{x}) = \frac{g_i(\hat{x})\partial f_i(\hat{x}) - f_i(\hat{x})\partial g_i(\hat{x})}{g_i(\hat{x})^2}.$$

So (3.3.8) becomes

$$0 \in \sum_{i\in I} \hat{\mu}_i g_i(\hat{x}) \partial^c \left[\frac{f_i}{g_i}\right](\hat{x}) + N_X(\hat{x}). \tag{3.3.9}$$

Let

$$\hat{\alpha}_i = \frac{\hat{\mu}_i g_i(\hat{x})}{\sum_{i\in I} \hat{\mu}_i g_i(\hat{x})}.$$

Then $\hat{\alpha}_i \geq 0$ and $\sum_{i\in I} \hat{\alpha}_i = 1$. Since $N_X(\hat{x})$ is a cone, (3.3.9) entails that

$$0 \in \sum_{i\in I} \hat{\alpha}_i \partial^c \left[\frac{f_i}{g_i}\right](\hat{x}) + N_X(\hat{x}). \tag{3.3.10}$$

On the other hand, from the previous remark, $\hat{\alpha}_i = 0$ if $f_i(\hat{x})/g_i(\hat{x}) < \lambda(\hat{x})$, which gives $\hat{\alpha}_i f_i(\hat{x})/g_i(\hat{x}) = \hat{\alpha}_i \lambda(\hat{x})$. Therefore,

$$\sum_{i\in I} \hat{\alpha}_i \left[\frac{f_i(\hat{x})}{g_i(\hat{x})}\right] = \lambda(\hat{x}).$$

Referring to [30, Proposition 2.3.12] or [11, Theorem 3.23] the last equality, together with (3.3.10), imply that $0 \in \partial^c \lambda(\hat{x}) + N_X(\hat{x})$. □

The next theorem shows that the optimal solutions of (P) satisfy the optimality conditions of Theorems 3.3.1 and 3.3.2.

Theorem 3.3.3. If $\bar{x} \in X$ is an optimal solution of (P), then there exist, $\bar{\mu}_i \geq 0$, $i \in I$, with $\sum_{i\in I} \bar{\mu}_i = 1$ such that

$$0 \in \sum_{i\in I} \bar{\mu}_i \left[\partial f_i(\bar{x}) - \lambda(\bar{x}) \partial g_i(\bar{x})\right] + N_X(\bar{x}),$$

with $\bar{\mu}_i \left[f_i(\bar{x}) - \lambda(\bar{x}) g_i(\bar{x})\right] = 0$, for all $i \in I$. If the assumptions of Theorem 3.3.2 are satisfied, these conditions are equivalent to $0 \in \partial^c \lambda(\bar{x}) + N_X(\bar{x})$, where $\partial^c \lambda(\bar{x})$ is the Clarke subdifferential of λ at $\bar{x}$ and $N_X(\bar{x})$ the normal cone to X at $\bar{x}$. Moreover, if $\lambda(\bar{x}) \leq 0$ the conditions are also sufficient.

Proof. Let $\bar{x} \in X$ be an optimal solution of (P). Then from Proposition 3.2.2, $\bar{x}$ also globally minimizes over the convex set X, the convex function $h_{\bar{x}}$ defined by

$$h_{\bar{x}}(x) := \begin{cases} \max_{i\in I} \left[f_i(x) - \bar{\lambda}\left(g_i(\bar{x}) + \langle \bar{\gamma}_i, x - \bar{x}\rangle\right)\right] & \text{if } \bar{\lambda} > 0 \\ \max_{i\in I} \left[f_i(x) - \bar{\lambda} g_i(x)\right] & \text{if } \bar{\lambda} \leq 0 \end{cases}$$

for all $\bar{\gamma}_i \in \partial g_i(\bar{x})$, $i \in I$, where $\bar{\lambda} = \lambda(\bar{x})$. Since $h_{\bar{x}}(\bar{x}) = 0$, it suffices to use Theorems 3.3.1 and 3.3.2 to conclude. □

Remark 3.3.2.

1. Observe that Theorem 3.3.1 furnishes the necessary optimality condition $h_{\hat{x}}(x_{\hat{x}}) = 0$. This will be used as a natural stopping criterion for our expected algorithm. On the other hand, if the functions g_i, $i \in I$, are differentiable, this condition is verifiable since it requires only solving a convex program.

2. Another interesting indication given by Theorem 3.3.1, is that we have a global minimum at hand when $\lambda(\hat{x}) \leq 0$.

3.4 DC Dinkelbach-type Algorithm

Before describing our algorithm, remark that one can write

$$\lambda(x) = \max_{i \in I} \frac{f_i(x)/\omega_i}{g_i(x)/\omega_i} \tag{3.4.1}$$

for all $\omega_i > 0$, $i \in I$. Even if this transformation has no effect on the function λ, it has a strong effect on the parametric subproblems. This remark has been pointed out in [36] with a special choice of ω_i, $i \in I$, and gave rise to a new Dinkelbach-type algorithm, more efficient than the original Dinkelbach-type one. For computational reasons, we will write λ as in (3.4.1). With this artifice, the parametrized function defined in the begining of the chapter takes the form

$$h_y(x) := \begin{cases} \max_{i \in I} \left[\dfrac{f_i(x) - \lambda(y)\left(g_i(y) + \langle \gamma_i, x - y \rangle\right)}{\omega_i} \right] & \text{if } \lambda(y) > 0 \\[2ex] \max_{i \in I} \left[\dfrac{f_i(x) - \lambda(y) g_i(x)}{\omega_i} \right] & \text{if } \lambda(y) \leq 0 \end{cases}$$

All the previous results remain valid, it suffices to replace f_i and g_i, respectively by f_i/ω_i and g_i/ω_i.

With the insight of the previous section's results, we will develop an algorithm by approximating the function $h_{\bar{x}}$, at each step k by $h_{x^k} =: h_k$, where x^k is a global minimum of the convex function $h_{x^{k-1}}$ over X.

Notice that there is no need to a starting feasible point x^0, i.e. $x^0 \in X$, and one can choose $x \in \mathbb{R}^n$ such that $g_i(x) \neq 0$ for all $i \in I$, set $\mu = \lambda(x)$ or choose any real $\mu < 0$, and then solves the convex subproblem

$$(P_\mu) \qquad \inf_{x\in X} \max_{i\in I} \{f_i(x) - \mu g_i(x)\}$$

to get $x^0 \in X$. Now we are ready to describe our algorithm.

Algorithm 2 DC Dinkelbach-type Algorithm

Let $\{\varepsilon_k\}$ be a sequence of nonnegative reals such that $\sum_{k\geq 0} \varepsilon_k < \infty$. Choose $x^- \in \mathbb{R}^n$ such that $g_i(x^-) \neq 0$ for all $i \in I$, set $\lambda_- = \lambda(x^-)$ or choose $\lambda_- < 0$, and solve (P_{λ_-}) to get a point $x^0 \in X$, and let $k = 0$.

1. At iteration k, we have $x^k \in X$, $\lambda_k = \lambda(x^k)$, $\varepsilon_k \geq 0$ and $\bar{\omega} \geq \omega_i^k \geq \omega > 0$, $i \in I$. If $\lambda_k > 0$ select $\gamma_{k,i} \in \partial g_i(x^k)$ for all $i \in I$. Then find $x^{k+1} \in X$ such that
$$h_k(x^{k+1}) \leq \inf_{x\in X} h_k(x) + \varepsilon_k$$
where
$$h_k(x) := \begin{cases} \max_{i\in I} \left[\dfrac{f_i(x) - \lambda_k \left(g_i(x^k) + \langle \gamma_{k,i}, x - x^k \rangle \right)}{\omega_i^k} \right] & \text{if } \lambda_k > 0, \\ \\ \max_{i\in I} \left[\dfrac{f_i(x) - \lambda_k g_i(x)}{\omega_i^k} \right] & \text{if } \lambda_k \leq 0. \end{cases}$$
If $h_k(x^{k+1}) = 0$, STOP.
2. Set $k = k+1$ and go to step 1.

Remark 3.4.1. A possible choice for the weights is $\omega_i^k = g_i(x^k)$. It has been used first in [33,36] and later in [24,107] where numerical tests confirmed its efficiency. This choice may be used in the previous algorithm.

To establish the convergence of the sequence $\{\lambda_k\}$, we need the following well known lemma.

Lemma 3.4.1. Let $\{\varepsilon_k\}$ be a sequence of positive reals such that $\sum_{k\geq 0} \varepsilon_k < \infty$, and let $\{u_k\}$ be a sequence such that $u_{k+1} \leq u_k + \varepsilon_k$ for all $k \in \mathbb{N}$. Then $\{u_k\}$ converges to some $\hat{u} \in \mathbb{R} \cup \{-\infty\}$.

The next result gives the convergence of the sequence $\{\lambda_k\}$ and a stopping criterion, which is the convergence of $\{h_k(x^{k+1})\}$ towards 0.

We denote and assume

$$\delta := \inf_{x \in X} \min_{i \in I} g_i(x) > 0 \quad \text{and} \quad \Delta := \sup_{x \in X} \max_{i \in I} g_i(x) < \infty.$$

Proposition 3.4.1. If $\sum_{k \geq 0} \varepsilon_k < \infty$, the sequence $\{\lambda_k\}$ converges to some $\hat{\lambda} \geq \bar{\lambda}$, where $\bar{\lambda}$ is the minimum value of (P), and $\{h_k(x^{k+1})\}$ converges to 0. If for an infinite number of iterations k, $\lambda_k \leq 0$, e.g. if $\hat{\lambda} < 0$, then $\hat{\lambda} = \bar{\lambda}$.

Proof. From the definition of x^{k+1} we have

$$\varepsilon_k + h_k(x) \geq h_k(x^{k+1}) \quad \text{for all } x \in X. \tag{3.4.2}$$

For $x = x^k$ we get $h_k(x^{k+1}) \leq h_k(x^k) + \varepsilon_k = \varepsilon_k$, where the equality $h_k(x^k) = 0$ follows from Proposition 3.2.1, Item 1, since $h_k(x^k) = h_{x^k}(x^k)$. From Lemma 3.2.1 we have

$$h_k(x) \geq \max_{i \in I} \left[\frac{f_i(x) - \lambda_k g_i(x)}{\omega_i^k} \right].$$

Therefore,

$$\begin{aligned} \varepsilon_k \geq h_k(x^{k+1}) &\geq \max_{i \in I} \left[\frac{f_i(x^{k+1}) - \lambda_k g_i(x^{k+1})}{\omega_i^k} \right] \\ &\geq \frac{f_{i_k}(x^{k+1}) - \lambda_k g_{i_k}(x^{k+1})}{\omega_{i_k}^k} \\ &= \frac{g_{i_k}(x^{k+1})}{\omega_{i_k}^k} (\lambda_{k+1} - \lambda_k), \end{aligned} \tag{3.4.3}$$

where i_k satisfies $\lambda_{k+1} = f_{i_k}(x^{k+1})/g_{i_k}(x^{k+1})$. This implies that $\lambda_{k+1} \leq \lambda_k + \varepsilon_k \bar{\omega}/\delta$, where we used the assumptions $\delta > 0$ and $0 < \omega_i^k \leq \bar{\omega}$ for all k and $i \in I$. Since $\lambda_k \geq \bar{\lambda}$ and $\sum_{k \geq 0} \varepsilon_k < \infty$, we conclude that the sequence $\{\lambda_k\}$ converges to some $\hat{\lambda} \geq \bar{\lambda}$. Then it follows, taking into account (3.4.3), the facts that $\delta \leq g_{i_k}(x^{k+1}) \leq \Delta$ and $0 < \omega \leq \omega_{i_k}^k \leq \bar{\omega}$, that $\{h_k(x^{k+1})\}$ converges to 0.

Assume now that $\lambda_k \leq 0$ for an infinite number of iterations, say $k \in K \subset \mathbb{N}$. Then for all $x \in X$ and all $k \in K$ we have

$$h_k(x) = \max_{i \in I} \left[\frac{f_i(x) - \lambda_k g_i(x)}{\omega_i^k} \right].$$

By considering a subsequence if necessary, we can assume that for all $i \in I$, $\omega_i^k \to \hat{\omega}_i$. Therefore, by using (3.4.2) and (3.4.3) and passing to limit over $k \in K$, we get

$$\max_{i \in I} \left[\frac{f_i(x) - \hat{\lambda} g_i(x)}{\hat{\omega}_i} \right] \geq 0 \quad \text{for all } x \in X.$$

For all $x \in X$, there exists $i \in I$ such that $f_i(x) - \hat{\lambda} g_i(x) \geq 0$ or equivalently $f_i(x)/g_i(x) \geq \hat{\lambda}$. This entails that $\lambda(x) \geq \hat{\lambda}$ for all $x \in X$, which gives $\bar{\lambda} \geq \hat{\lambda}$. The equality $\bar{\lambda} = \hat{\lambda}$ then follows. □

Remark 3.4.2. If $\sum_{k \geq 0} \varepsilon_k < \infty$, and the set $\{x \in \mathbb{R}^n \mid \lambda(x) \leq \lambda(x^0) + \sum_{k \geq 0} \varepsilon_k \bar{\omega}/\delta\}$ is bounded, for the starting point x^0 (wich is the case if $\lambda(\cdot)$ is inf-compact), then the sequence $\{x^k\}$ is bounded. Indeed, we showed in the last proof that $\lambda_{k+1} \leq \lambda_k + \varepsilon_k \bar{\omega}/\delta$. Therefore, $\lambda_{k+1} \leq \lambda_0 + \sum_{i=0}^{k} \varepsilon_i \bar{\omega}/\delta$ implying that $x^{k+1} \in \{x \in \mathbb{R}^n \mid \lambda(x) \leq \lambda(x^0) + \sum_{k \geq 0} \varepsilon_k \bar{\omega}/\delta\}$.

Now we turn our attention to the convergence of the sequence $\{x^k\}$. If it is bounded, we will show that all its cluster points are Clarke stationary.

Theorem 3.4.1. If $\sum_{k \geq 0} \varepsilon_k < \infty$ and the sequence $\{x^k\}$ is bounded, then for every cluster point $\hat{x}$ of $\{x^k\}$ there exist $\hat{\mu}_i \geq 0$, $i \in I$, with $\sum_{i \in I} \hat{\mu}_i = 1$ such that

$$0 \in \sum_{i \in I} \hat{\mu}_i \left[\partial f_i(\hat{x}) - \hat{\lambda} \partial g_i(\hat{x}) \right] + N_X(\hat{x}),$$

and $\hat{\mu}_i \left[f_i(\hat{x}) - \hat{\lambda} g_i(\hat{x}) \right] = 0$, $i \in I$, where $\hat{\lambda} = \lambda(\hat{x})$. These conditions are equivalent to $0 \in \partial^c \lambda(\hat{x}) + N_X(\hat{x})$, where $\partial^c \lambda(\hat{x})$ is the Clarke subdifferential of λ at $\hat{x}$ and $N_X(\hat{x})$ the normal cone to X at $\hat{x}$. Moreover, if $\hat{\lambda} \leq 0$ then $\hat{\lambda} = \bar{\lambda}$ and $\hat{x}$ is an optimal solution for (P).

Proof. From the definition of x^{k+1} and (3.4.3) we have

$$\begin{aligned} h_k(x) &\geq h_k(x^{k+1}) - \varepsilon_k \\ &\geq \frac{g_{i_k}(x^{k+1})}{\omega_{i_k}^k} (\lambda_{k+1} - \lambda_k) - \varepsilon_k, \end{aligned} \tag{3.4.4}$$

for all $x \in X$. Consider now subsequences of $\{x^k\}$ and $\{\omega_i^k\}$, for all $i \in I$, converging respectively to $\hat{x}$ and $\hat{\omega}_i$, and subsequences of $\{\gamma_{k,i}\}$ that converges to some $\hat{\gamma}_i \in \partial g_i(\hat{x})$, see e.g. [64, Proposition 6.2.2 and 6.2.1]. Let us consider the function

$$h_{\hat{x}}(x) := \begin{cases} \max_{i\in I}\left[\dfrac{f_i(x) - \hat{\lambda}\,(g_i(\hat{x}) + \langle \hat{\gamma}_i, x - \hat{x}\rangle)}{\hat{\omega}_i}\right] & \text{if } \hat{\lambda} > 0, \\[2ex] \max_{i\in I}\left[\dfrac{f_i(x) - \hat{\lambda} g_i(x)}{\hat{\omega}_i}\right] & \text{if } \hat{\lambda} \leq 0. \end{cases}$$

By invoking (3.4.4) and Proposition 4.3.1 we arrive to $h_{\hat{x}}(x) \geq 0$ for all $x \in X$. But since $\hat{\lambda} = \lambda(\hat{x})$, we have $h_{\hat{x}}(\hat{x}) = 0$, from which we deduce that $\hat{x}$ globally minimizes the convex function $h_{\hat{x}}$ over X. Then it suffices to use Theorems 3.3.1 and 3.3.2 to conclude. □

Remark 3.4.3. If $f_i - \hat{\lambda} g_i$, for $i \in I$, are convex and $\partial[f_i - \hat{\lambda} g_i](\hat{x}) = \partial f_i(\hat{x}) - \hat{\lambda}\partial g_i(\hat{x})$, then the conditions of Theorem 3.3.3 are sufficient and $\hat{\lambda} = \bar{\lambda}$. This is the case, for instance, if $\hat{\lambda} \leq 0$. Furtheremore, $\partial[f_i - \hat{\lambda} g_i](\hat{x}) = \partial f_i(\hat{x}) - \hat{\lambda}\partial g_i(\hat{x})$ holds if f_i or g_i are differentiable at $\hat{x}$, for all $i \in I$, see e.g. [30, Corollary 2 of Proposition 2.3.1 and 2.3.3].

Chapter 4

Optimality Conditions and DC-Dinkelbach-type Algorithm for Generalized Fractional Programs with Ratios of Difference of Convex Functions

In this chapter, we develop optimality conditions and propose an algorithm for generalized fractional programming problems whose objective function is the maximum of finite ratios of difference of convex (DC) functions, with DC constraints, that we will call later, DC-GFP. Such problems are generally nonsmooth and nonconvex. We first give in this work, optimality conditions for such problems, by means of convex analysis tools. For solving DC-GFP, the use of Dinkelbach-type algorithms conducts to nonconvex subproblems, which causes the failure of the latter since it requires finding a global minimum for these subprograms. To overcome this difficulty, we propose a DC-Dinkelbach-type algorithm in which we overestimate the objective function in these subproblems by a convex function, and the constraints set by an inner convex subset of the latter, which leads to convex subproblems. We show that every cluster point of the sequence of optimal solutions of these subproblems satisfies necessary optimality conditions of KKT type [58].

4.1 Introduction

In this chapter, we consider fractional programming problems whose objective function is the maximum of finite ratios of difference of convex (DC) functions, with DC constraints. More precisely, we consider problems of the form

$$(P) \qquad \bar{\lambda} = \inf_{x \in X} \left\{ \lambda(x) := \max_{i \in I} \frac{f_i^1(x) - f_i^2(x)}{g_i^1(x) - g_i^2(x)} \right\}$$

where $X = \{x \in C \mid h_j^1(x) - h_j^2(x) \leq 0, \forall j \in J\}$, with $C \subset \mathbb{R}^n$ a nonempty, closed convex set, I and J two finite index sets, and the functions f_i^ℓ, g_i^ℓ, for $i \in I$, and h_j^ℓ, for $j \in J$ and $\ell = 1, 2$ are defined on $\mathbb{R}^n$ and convex, with $g_i^1 - g_i^2$ positive on X for all $i \in I$.

To simplify notations, we will put for all $i \in I$, $j \in J$ and $x \in \mathbb{R}^n$,

$$f_i(x) = f_i^1(x) - f_i^2(x), \quad g_i(x) = g_i^1(x) - g_i^2(x), \quad h_j(x) = h_j^1(x) - h_j^2(x)$$

and

$$h(x) = \max_{j \in J} h_j(x).$$

With the last notation we have, $X = \{x \in C \mid h(x) \leq 0\}$.

Problems of this form have been already studied in [23], where optimality conditions were obtained and a method of resolution was proposed. They include ordinary convex programs, generalized fractional problems (GFP) with: ratios of convex and concave functions, ratios of convex functions, ratios of concave functions, ratios of concave and convex functions, etc. Another important class of such problems is the GFP for which the functions may be expressed as a difference of convex functions (see, e.g., [126] for such functions).

There is a rich literature dealing with GFP, see [118–121] for a detailed bibliography. Also, see [7–9] for recent applications. There are several primal Dinkelbach-type algorithms [17,33,35–37,107,108,122]; and dual algorithms and results [1,2,13–15,21,22,24–26,34,42,43,67]. These algorithms are based on auxiliary parametric problems having simpler structures than the original problem. For the primal algorithms, the auxiliary problems furnish sequences of approximate optimal values converging decreasingly to the optimal value of (P), whereas the sequences of values generated by the dual algorithms converge increasingly towards the optimal value of (P).

Another strategy was proposed in [122], which consists in applying bundle methods for solving a GFP. These methods consist in approximately solving the primal auxiliary problems associated with the GFP by using primal bundle methods. Recently, since the last algorithm is rather intended to solve linearly constrained GFPs, another primal bundle method, based this time on the extended method of centers [107], was proposed in [2] to deal with nonlinearly constrained GFPs. Very recently, a dual bundle method has been proposed in [25], also for solving such problems, this time without convexity assumptions.

But for almost all of these methods, convexity of the parametric problems is required, and apart from the situation when the parametric subproblems are convex, these approaches may fail since it is assumed that one can compute their global minimum, at least approximately. In our situation, where both f_i^ℓ and g_i^ℓ (for $\ell = 1,2$) are convex, the function $(f_i^1 - f_i^2) - \lambda(g_i^1 - g_i^2)$ is nonconvex, but it is a difference of two convex functions, since $f_i^1 + \lambda g_i^2$ and $f_i^2 + \lambda g_i^1$ (resp. $f_i^1 - \lambda g_i^1$ and $f_i^2 - \lambda g_i^2$) are two convex functions when λ is nonnegative (resp. λ is negative). For this reason, we want to modify the last mentioned parametric subproblems in such a way we take into account these two situations. We resort to DC techniques, see e.g [80, 124], where it is question to replace the functions f_i^2 and g_i^ℓ ($\ell = 1$ or $\ell = 2$) by their affine approximations, obtained by a subgradient at current point. Doing so we obtain convex parametric subproblems.

The most important difficulty in global optimization, and in particular in DC programming, is how to recognize a global minimum, or even how to recognize local minimum, in contrast with the convex programming where a local minimum is global. The most common necessary optimality condition for the minimization of $f_1 - f_2$, say, over the whole space $\mathbb{R}^n$ is dc criticality, which means that a point x^* is DC critical if $\partial f_1(x^*) \cap \partial f_2(x^*) \neq \emptyset$, or equivalently if $0 \in \partial f_1(x^*) - \partial f_2(x^*)$, where $\partial f_i(x^*)$ stands for the subdiferential of the convex function f_i, $i = 1,2$, at x^*. It is strongly critical if $\partial f_2(x^*) \subset \partial f_1(x^*)$. To get more on optimality conditions and algorithms for DC unconstrained and DC constrained programs, the reader can consult [6,55,62,80,124], and [10,103] for several extensions and applications.

The chapter is organized as follow. In Section 4.2 we introduce a new parametric approach based on convex parametric subprolems, and show that the problem (P) is equivalent to a convex problem. By writing necessary optimality conditions, for the latter, using only convex analysis tools, we obtain in Section 4.3, necessary optimality conditions for (P). Later, in Section 4.4, we will describe our DC-Dinkelbach-type algorithm and establish its convergence.

4.2 Parametric Approach for DC-GFP

In Dinkelbach-type algorithms, the associated parametric problem to (P), with the parameter μ, takes the form

$$(P_\mu) \qquad F(\mu) := \inf_{x\in X} \max_{i\in I} \{f_i(x) - \mu g_i(x)\}.$$

We recall that under mild assumptions, when $\mu = \bar{\lambda}$, problems (P) and (P_μ) have the same optimal solutions set, see e.g., [35, Proposition 2.1] and [22, Lemma 1].

The problems (P_μ) appear as subproblems in the Dinkelbach-type algorithm [35, 36]. This algorithm acts as follows: from an arbitrary point $x^0 \in X$ a sequence of points x^k is generated by solving a sequence of subproblems (P_{λ_k}), where $\lambda_k = \lambda(x^k)$ and x^{k+1} is an optimal solution of (P_{λ_k}). Habitual assumptions to apply Dinkelbach-type algorithm is that the functions f_i are convex and g_i concave, with f_i positive if g_i is not affine. This implies that $f_i - \lambda_k g_i$ is convex. Apart this case where (P_{λ_k}) is convex, it is difficult to find a global optimal solution to (P_{λ_k}), and the method fails to work. This is the case for our problem (P) for which the subproblems (P_{λ_k}) may be nonconvex.

Our objective in this chapter is to develop optimality conditions for minimax fractional problems whose objective function is the maximum of several ratios of DC functions, to be minimized under constraints set described by DC functions. Our appraoch is based only on convex analysis tools. We overcome the difficulty caused by the nonconvexity of the parametric subproblems in Dinkelbach-type procedures, and propose a DC-Dinkelbach-type algorithm in which the subproblems are convex. The idea behind our proposition is to linearize the functions f_i^2 and the functions g_i^ℓ for $i \in I$, at current points x^k, where $\ell = 1$ when $\lambda_k \geq 0$ and $\ell = 2$ when $\lambda_k < 0$. Since the functions h_j are also DC functions, we follow the same strategy as for f_i and g_i and linearize the functions h_j^2, for all $j \in J$. To develop our algorithm we begin by defining the parametrized subproblems and give some related results.

In all what follows, $\partial f_i^2(x)$, $\partial h_j^2(x)$ and $\partial g_i^\ell(x)$ will designate, respectively, the subdifferentials of the convex functions f_i^2, h_j^2 and g_i^ℓ at x, for $\ell = 1, 2$.

To begin our analysis, we define the function parametrized by $y \in X$,

$$F_y(x) := \max_{i \in I} \Big[f_{i,y}(x) - \lambda(y) g_{i,y}(x) \Big],$$

where

$$f_{i,y}(x) = f_i^1(x) - [f_i^2(y) + \langle x_i^2(y), x - y \rangle], \tag{4.2.1}$$

$$g_{i,y}(x) := \begin{cases} g_i^1(x) - [g_i^2(y) + \langle y_i^2(y), x - y \rangle] & \text{if } \lambda(y) < 0 \\ -g_i^2(x) + [g_i^1(y) + \langle y_i^1(y), x - y \rangle] & \text{if } \lambda(y) \geq 0 \end{cases} \tag{4.2.2}$$

with some $x_i^2(y) \in \partial f_i^2(y)$ and $y_i^\ell(y) \in \partial g_i^\ell(y)$, for $\ell = 1$ or $\ell = 2$. That is, we replace the functions f_i^2 and g_i^ℓ, $\ell = 1$ or $\ell = 2$, by their affine approximations at y, namely, $f_i^2(y) + \langle x_i^2(y), x - y \rangle$ and $g_i^\ell(y) + \langle y_i^\ell(y), x - y \rangle$, respectively.

Also, for all $x \in \mathbb{R}^n$ and $j \in J$, we define the functions $h_{j,y}$, parametrized by $y \in \mathbb{R}^n$,

$$h_{j,y}(x) := h_j^1(x) - [h_j^2(y) + \langle z_j^2(y), x - y \rangle], \tag{4.2.3}$$

with some $z_j^2(y) \in \partial h_j^2(y)$, and consider the set

$$X_y = \Big\{ x \in C \mid h_{j,y}(x) \leq 0, \forall j \in J \Big\}.$$

Notice that by the convexity assumptions made on the functions f_i^ℓ and g_i^ℓ, $\ell = 1, 2$, the functions $f_{i,y}(\cdot)$ and $-\lambda(y) g_{i,y}(\cdot)$ are convex for all $i \in I$ and $y \in X$, and so is the function $F_y(\cdot)$. On the other hand, the convexity of the functions h_j^ℓ, $\ell = 1, 2$, implies the convexity of the functions $h_{j,y}(\cdot)$, for all $j \in J$ and $y \in X$, and thus the convexity of the set X_y.

Remark 4.2.1. From the subgradient inequalities $h_j^2(x) \geq h_j^2(y) + \langle z_j^2(y), x - y \rangle$ for all $x, y \in \mathbb{R}^n$, $j \in J$, we conclude that $X_y \subset X$ for all $y \in \mathbb{R}^n$. On the other hand, $y \in X_y$ if and only if $y \in X$.

Now, for $y \in \mathbb{R}^n$, we associate to (P) the convex approximate parametric problem

$$(P(y)) \qquad \inf_{x \in X_y} \Big\{ F_y(x) := \max_{i \in I} \Big[f_{i,y}(x) - \lambda(y) g_{i,y}(x) \Big] \Big\},$$

and we denote by x_y the global minimum of F_y over X_y, if any.

Before discussing optimality conditions for problem (P), we begin first by some preliminary results.

Lemma 4.2.1. For all $x \in \mathbb{R}^n$ and $y \in X$ we have

$$\max_{i \in I} [f_i(x) - \lambda(y) g_i(x)] \leq F_y(x).$$

Proof. From the definition of f_i, g_i and the subgradient inequalities $f_i^2(y) + \langle x_i^2(y), x - y \rangle \leq f_i^2(x)$, for all $i \in I$ we have

$$\begin{aligned} f_i(x) - \lambda(y) g_i(x) &= f_i^1(x) - f_i^2(x) - \lambda(y) \left[g_i^1(x) - g_i^2(x) \right] \\ &\leq f_i^1(x) - \left[f_i^2(y) + \langle x_i^2(y), x - y \rangle \right] - \lambda(y) \left[g_i^1(x) - g_i^2(x) \right] \\ &= f_{i,y}(x) - \lambda(y) \left[g_i^1(x) - g_i^2(x) \right]. \end{aligned} \tag{4.2.4}$$

If $\lambda(y) < 0$, from the subgradient inequalities $g_i^2(y) + \langle y_i^2(y), x - y \rangle \leq g_i^2(x)$, for all $i \in I$, the definition of F_y, and (6.2.8) we get

$$\begin{aligned} f_i(x) - \lambda(y) g_i(x) &\leq f_{i,y}(x) - \lambda(y) \left[g_i^1(x) - \left(g_i^2(y) + \langle y_i^2(y), x - y \rangle \right) \right] \\ &= f_{i,y}(x) - \lambda(y) g_{i,y}(x) \\ &\leq F_y(x). \end{aligned}$$

Assume now that $\lambda(y) \geq 0$. Then, from the subgradient inequalities $g_i^1(y) + \langle y_i^1(y), x - y \rangle \leq g_i^1(x)$, for all $i \in I$, the definition of F_y, and (6.2.8) we get

$$\begin{aligned} f_i(x) - \lambda(y) g_i(x) &\leq f_{i,y}(x) - \lambda(y) \left[g_i^1(y) + \langle y_i^1(y), x - y \rangle - g_i^2(x) \right] \\ &= f_{i,y}(x) - \lambda(y) \left[-g_i^2(x) + \left(g_i^1(y) + \langle y_i^1(y), x - y \rangle \right) \right] \\ &= f_{i,y}(x) - \lambda(y) g_{i,y}(x) \\ &\leq F_y(x). \end{aligned}$$

Thus, $f_i(x) - \lambda(y) g_i(x) \leq F_y(x)$ for all $i \in I$, from which the desired inequality follows. □

Recall that for all $y \in X$, we designated by x_y a minimum of F_y over X_y. Then we have the following results.

Proposition 4.2.1. For all $y \in X$ we have

1. $F_y(y) = 0$ and $F_y(x_y) \leq 0$,
2. $\lambda(x_y) \leq \lambda(y)$.

Proof. 1. From the definition of F_y, we get

$$F_y(y) = \max_{i\in I} [f_i(y) - \lambda(y)g_i(y)],$$

where $f_i(y) = f_i^1(y) - f_i^2(y)$ and $g_i(y) = g_i^1(y) - g_i^2(y)$.

From the definition of $\lambda(y)$ we have $f_i(y)/g_i(y) \le \lambda(y)$ for all $i \in I$, with equality for some i_0, or equivalently $f_i(y) - \lambda(y)g_i(y) \le 0$ with equality for $i = i_0$. This shows that $F_y(y) = 0$. On the other hand, the definition of x_y implies that $F_y(x_y) \le F_y(x)$ for all $x \in X_y$. In particular, for $x = y$ we get $F_y(x_y) \le F_y(y) = 0$.

2. From Lemma 4.2.1, with $x = x_y$, we have $f_i(x_y) - \lambda(y)g_i(x_y) \le F_y(x_y) \le 0$ for all $i \in I$. This means that $f_i(x_y)/g_i(x_y) \le \lambda(y)$ for all $i \in I$, implying that $\lambda(x_y) \le \lambda(y)$. □

Notice that the Item 2 of Proposition 4.2.1 says that having a point $y \in X$, one obtain a better point for λ, which is x_y, by solving the convex program $(P(y))$.

In the next proposition we investigate relations between the problem (P) and the problem $(P(y))$, of minimizing F_y over X_y.
Proposition 4.2.2. If the problem (P) has a global optimal solution on X, say $\bar{x}$, then this solution actually globally minimizes $F_{\bar{x}}$ over $X_{\bar{x}}$, whatever are $x_i^2(\bar{x}) \in \partial f_i^2(\bar{x})$, $y_i^l(\bar{x}) \in \partial g_i^l(\bar{x})$, where $\ell = 1$ if $\lambda(\bar{x}) \ge 0$ and $\ell = 2$ otherwise; and $z_j^2(\bar{x}) \in \partial h_j^2(\bar{x})$ for $i \in I$ and $j \in J$. Conversely, for all global optimal solution $\bar{x}$ of (P), every optimal solution $x_{\bar{x}}$ of $F_{\bar{x}}$ over $X_{\bar{x}}$ also globally solves the problem (P).

Proof. Let $\bar{x} \in X$ be a global optimal solution of (P). Let $x \in X$ and $i \in I$ be such that $\lambda(x) = f_i(x)/g_i(x)$. Then $\lambda(\bar{x}) \le \lambda(x) = f_i(x)/g_i(x)$. This implies that $f_i(x) - \lambda(\bar{x})g_i(x) \ge 0$. By using Lemma 4.2.1, with $y = \bar{x}$, we conclude that $F_{\bar{x}}(x) \ge 0$ for all $x \in X$. Since $F_{\bar{x}}(\bar{x}) = 0$ from Proposition 4.2.1, Item 1, with $y = \bar{x}$, the conclusion follows taking into acount that $\bar{x} \in X_{\bar{x}}$. The converse follows directly from Proposition 4.2.1, Item 2, with $y = \bar{x}$. □

4.3 Optimality Conditions for DC-GFP

We are ready to give optimality conditions for (P). With the previous results, we now give optimality conditions for (P), using only convex analysis tools.

Theorem 4.3.1. Let $\hat{x} \in X$ and $\hat{\lambda} = \lambda(\hat{x})$. If for all $i \in I$, there exist $x_i^2(\hat{x}) \in \partial f_i^2(\hat{x})$, $y_i^\ell(\hat{x}) \in \partial g_i^\ell(\hat{x})$, where $\ell = 1$ if $\hat{\lambda} \geq 0$ and $\ell = 2$ otherwise; and for all $j \in J$, there exist $z_j^2(\hat{x}) \in \partial h_j^2(\hat{x})$, such that $F_{\hat{x}}(x_{\hat{x}}) = 0$, where $x_{\hat{x}}$ is a global minimum of $F_{\hat{x}}$ on $X_{\hat{x}} := \{x \in C \mid h_{j,\hat{x}}(x) \leq 0, \forall j \in J\}$ then, for all $i \in I$, there exist $\hat{\mu}_i \geq 0$, and for all $j \in J$, there exist $\hat{\nu}_j \geq 0$, with $\sum_{i\in I} \hat{\mu}_i + \sum_{j \in J} \hat{\nu}_j = 1$, such that

$$0 \in \sum_{i\in I} \hat{\mu}_i \left[\partial f_i^1(\hat{x}) - \partial f_i^2(\hat{x}) - \hat{\lambda}\left(\partial g_i^1(\hat{x}) - \partial g_i^2(\hat{x})\right)\right] + \sum_{j\in J} \hat{\nu}_j \left[\partial h_j^1(\hat{x}) - \partial h_j^2(\hat{x})\right] + N_C(\hat{x}), \tag{4.3.1}$$

with the equalities

$$\hat{\mu}_i \left[f_i(\hat{x}) - \hat{\lambda} g_i(\hat{x})\right] = 0 \text{ and } \hat{\nu}_j h_j(\hat{x}) = 0 \tag{4.3.2}$$

for all $i \in I$ and $j \in J$, where $N_C(\hat{x})$ is the normal cone to C at $\hat{x}$.

The converse is true, that is $F_{\hat{x}}(x_{\hat{x}}) = 0$, for some $x_i^2(\hat{x}) \in \partial f_i^2(\hat{x})$, $y_i^\ell(\hat{x}) \in \partial g_i^l(\hat{x})$, for $l = 1, 2$ and $i \in I$, if in addition to (6.2.9) and (6.2.10), one has

$$\max_{j\in J} h_{j,\hat{x}}(x) < 0 \qquad \text{for all } z_j^2(\hat{x}) \in \partial h_j^2(\hat{x}) \text{ and } j \in J, \tag{4.3.3}$$

for some $x \in C$.

Proof. Let $\hat{x} \in X$, $\hat{\lambda} = \lambda(\hat{x})$, $x_i^2(\hat{x}) \in \partial f_i^2(\hat{x})$, $y_i^\ell(\hat{x}) \in \partial g_i^\ell(\hat{x})$, where $\ell = 1$ if $\hat{\lambda} \geq 0$ and $\ell = 2$ if $\hat{\lambda} < 0$, for all $i \in I$; and let $z_j^2(\hat{x}) \in \partial h_j^2(\hat{x})$ for all $j \in J$. Define with these elements the functions stated in Eqs. (6.2.4) to (6.2.6). Let

$$F_{\hat{x}}(x) := \max_{i\in I} \left[f_{i,\hat{x}}(x) - \hat{\lambda} g_{i,\hat{x}}(x)\right],$$

and $X_{\hat{x}} := \{x \in C \mid h_{j,\hat{x}}(x) \leq 0, \forall j \in J\}$. Assume that $F_{\hat{x}}(x_{\hat{x}}) = 0$. Then since from Proposition 4.2.1, Item 1, with $y = \hat{x}$, we have $F_{\hat{x}}(\hat{x}) = 0$, we conclude that $\hat{x}$ is also a global minimum over the convex set $X_{\hat{x}}$, of the convex function $F_{\hat{x}}$. Observe that $\hat{x}$ also minimizes on C the function $\hat{F}_{\hat{x}}$ defined by

$$\hat{F}_{\hat{x}}(x) := \max\left[F_{\hat{x}}(x), \max_{j\in J} h_{j,\hat{x}}(x)\right].$$

Indeed, if $x \in X_{\hat{x}}$ then from the previous discussion, $F_{\hat{x}}(x) \geq 0$, implying that $\hat{F}_{\hat{x}}(x) \geq 0$. If $x \notin X_{\hat{x}}$, but $x \in C$ then $h_{j,\hat{x}}(x) > 0$ for some $j \in J$ and this again implies that $\hat{F}_{\hat{x}}(x) \geq 0$. It follows that $\hat{F}_{\hat{x}}(x) \geq 0$ for all $x \in C$. On the other hand, $\hat{F}_{\hat{x}}(\hat{x}) = 0$ since $F_{\hat{x}}(\hat{x}) = 0$ and $h_{j,\hat{x}}(x) \leq 0$ for all $j \in J$. This gives the conclusion. Therefore, from [64, Theorem 1.1.1] we conclude that

$$0 \in \partial \hat{F}_{\hat{x}}(\hat{x}) + N_C(\hat{x}),$$

where $\partial \hat{F}_{\hat{x}}(\hat{x})$ and $N_C(\hat{x})$ are, respectively, the subdifferential of $\hat{F}_{\hat{x}}$ and the normal cone of C at $\hat{x}$. By invoking [64, Corollary 4.3.2] to express $\partial \hat{F}_{\hat{x}}(\hat{x})$, we conclude that there exist $\hat{\alpha}_0, \hat{\alpha}_j \geq 0$ for $j \in J$, such that $\hat{\alpha}_0 + \sum_{j \in J} \hat{\alpha}_j = 1$,

$$0 \in \hat{\alpha}_0 \partial F_{\hat{x}}(\hat{x}) + \sum_{j \in J} \hat{\alpha}_j \partial h_{j,\hat{x}}(\hat{x}) + N_C(\hat{x})$$

and $\hat{\alpha}_j h_{j,\hat{x}}(\hat{x}) = 0$ for all $j \in J$, where $\partial F_{\hat{x}}(\hat{x})$ and $\partial h_{j,\hat{x}}(\hat{x})$ are, respectively, the subdifferential of $F_{\hat{x}}$ and the subdifferential of $h_{j,\hat{x}}$, $j \in J$, at $\hat{x}$. Again by referring to [64, Corollary 4.3.2] in the calculus of $\partial F_{\hat{x}}(\hat{x})$, there exist $\hat{\beta}_i \geq 0$, $i \in I$, with $\sum_{i \in I} \hat{\beta}_i = 1$ such that

$$0 \in \hat{\alpha}_0 \sum_{i \in I} \hat{\beta}_i \left[\partial f_{i,\hat{x}}(\hat{x}) - \hat{\lambda} \partial g_{i,\hat{x}}(\hat{x}) \right] + \sum_{j \in J} \hat{\alpha}_j \partial h_{j,\hat{x}}(\hat{x}) + N_C(\hat{x}). \tag{4.3.4}$$

and

$$\sum_{i \in I} \hat{\beta}_i \left[f_{i,\hat{x}}(\hat{x}) - \hat{\lambda} g_{i,\hat{x}}(\hat{x}) \right] = F_{\hat{x}}(\hat{x}) = 0. \tag{4.3.5}$$

If $\lambda(\hat{x}) < 0$ then from the expression of $f_{i,\hat{x}}$, $g_{i,\hat{x}}$ and $h_{j,\hat{x}}$ it is clear that (4.3.4) and (4.3.5), respectively imply that

$$0 \in \hat{\alpha}_0 \sum_{i \in I} \hat{\beta}_i \left[\partial f_i^1(\hat{x}) - x_i^2(\hat{x}) - \hat{\lambda} \left(\partial g_i^1(\hat{x}) - y_i^2(\hat{x}) \right) \right] + \sum_{j \in J} \hat{\alpha}_j \left[\partial h_j^1(\hat{x}) - z_j^2(\hat{x}) \right] + N_C(\hat{x}) \tag{4.3.6}$$

and

$$\sum_{i \in I} \hat{\beta}_i \left[f_i^1(\hat{x}) - f_i^2(\hat{x}) - \hat{\lambda} \left(g_i^1(\hat{x}) - g_i^2(\hat{x}) \right) \right] := \sum_{i \in I} \hat{\beta}_i \left[f_i(\hat{x}) - \hat{\lambda} g_i(\hat{x}) \right] = 0.$$

Finally, (4.3.6) implies that

$$0 \in \hat{\alpha}_0 \sum_{i \in I} \hat{\beta}_i \left[\partial f_i^1(\hat{x}) - \partial f_i^2(\hat{x}) - \hat{\lambda} \left(\partial g_i^1(\hat{x}) - \partial g_i^2(\hat{x}) \right) \right] + \sum_{j \in J} \hat{\alpha}_j \left[\partial h_j^1(\hat{x}) - \partial h_j^2(\hat{x}) \right] + N_C(\hat{x}). \tag{4.3.7}$$

If $\lambda(\hat{x}) \geq 0$, then again from the expressions of $f_{i,\hat{x}}$, $g_{i,\hat{x}}$ and $h_{j,\hat{x}}$, relations (4.3.4) and (4.3.5) become, respectively,

$$0 \in \hat{\alpha}_0 \sum_{i\in I} \hat{\beta}_i \left[\partial f_i^1(\hat{x}) - x_i^2(\hat{x}) - \hat{\lambda}\left(y_i^1(\hat{x}) - \partial g_i^2(\hat{x})\right)\right] + \sum_{j\in J} \hat{\alpha}_j \left[\partial h_j^1(\hat{x}) - z_j^2(\hat{x})\right] + N_C(\hat{x}) \qquad (4.3.8)$$

and

$$\sum_{i\in I} \hat{\beta}_i \left[f_i^1(\hat{x}) - f_i^2(\hat{x}) - \hat{\lambda}\left(g_i^1(\hat{x}) - g_i^2(\hat{x})\right)\right] := \sum_{i\in I} \hat{\beta}_i \left[f_i(\hat{x}) - \hat{\lambda} g_i(\hat{x})\right] = 0.$$

Clearly, (4.3.8) implies (4.3.7). It suffices to set $\hat{\mu}_i = \hat{\alpha}_0 \hat{\beta}_i$ and $\hat{\nu}_j = \hat{\alpha}_j$ to get the results (6.2.9) and (6.2.10), since $\sum_{i\in I} \hat{\mu}_i + \sum_{j\in J} \hat{\nu}_j = \hat{\alpha}_0 \sum_{i\in I} \hat{\beta}_i + \sum_{j\in J} \hat{\alpha}_j = 1$.

To show the converse, that is (6.2.9) and (6.2.10) imply that for some $x_i^2(\hat{x}) \in \partial f_i^2(\hat{x})$ and $y_i^\ell(\hat{x}) \in \partial g_i^l(\hat{x})$, for $l = 1, 2$ and $i \in I$, we have $F_{\hat{x}}(x_{\hat{x}}) = 0$, assume that (6.2.9) and (6.2.10) hold. Then from (6.2.9), for all $i \in I$, $j \in J$, there exist $\hat{x}_i^l \in \partial f_i^l(\hat{x})$, $\hat{y}_i^l \in \partial g_i^l(\hat{x})$ and $\hat{z}_j^l \in \partial h_j^l(\hat{x})$, for $l = 1, 2$, such that

$$\sum_{i\in I} \hat{\mu}_i \left[\hat{x}_i^1 - \hat{x}_i^2 - \hat{\lambda}\left(\hat{y}_i^1 - \hat{y}_i^2\right)\right] + \sum_{j\in J} \hat{\nu}_j \left[\hat{z}_j^1 - \hat{z}_j^2\right] \in -N_C(\hat{x}).$$

That is, for all $x \in C$ we have

$$\left\langle \sum_{i\in I} \hat{\mu}_i \left[\hat{x}_i^1 - \hat{x}_i^2 - \hat{\lambda}\left(\hat{y}_i^1 - \hat{y}_i^2\right)\right] + \sum_{j\in J} \hat{\nu}_j \left[\hat{z}_j^1 - \hat{z}_j^2\right], x - \hat{x} \right\rangle \geq 0. \qquad (4.3.9)$$

For all $x \in X$, by using the subgradient inequality $f_i^1(x) \geq f_i^1(\hat{x}) + \langle \hat{x}_i^1, x - \hat{x}\rangle$, for $i \in I$, we get

$$f_i^1(x) - \left(f_i^2(\hat{x}) + \langle \hat{x}_i^2, x - \hat{x}\rangle\right) \geq f_i^1(\hat{x}) - f_i^2(\hat{x}) + \langle \hat{x}_i^1 - \hat{x}_i^2, x - \hat{x}\rangle.$$

To comply with the notation of the definition of the functions in Eqs. (6.2.4) to (6.2.6), we set $x_i^2(\hat{x}) = \hat{x}_i^2$, $y_i^\ell(\hat{x}) = \hat{y}_i^\ell$, for $l = 1, 2$, $i \in I$, and $z_j^2(\hat{x}) = \hat{z}_j^2$, $j \in J$. With these notations, the last inequality becomes

$$f_{i,\hat{x}}(x) \geq f_i(\hat{x}) + \langle \hat{x}_i^1 - \hat{x}_i^2, x - \hat{x}\rangle \qquad (4.3.10)$$

Assume first that $\hat{\lambda} < 0$. The subgradient inequality $g_i^1(x) \geq g_i^1(\hat{x}) + \langle \hat{y}_i^1, x - \hat{x}\rangle$, for $i \in I$, implies that

$$g_i^1(x) - \left(g_i^2(\hat{x}) + \langle \hat{y}_i^2, x - \hat{x}\rangle\right) \geq g_i^1(\hat{x}) - g_i^2(\hat{x}) + \langle \hat{y}_i^1 - \hat{y}_i^2, x - \hat{x}\rangle.$$

Taking into account the definition of $g_{i,\hat{x}}$, we obtain

$$-\hat{\lambda} g_{i,\hat{x}}(x) \geq -\hat{\lambda}\left(g_i(\hat{x}) + \langle \hat{y}_i^1 - \hat{y}_i^2, x - \hat{x}\rangle\right). \tag{4.3.11}$$

For the case $\hat{\lambda} \geq 0$, we consider the subgradient inequality $g_i^2(x) \geq g_i^2(\hat{x}) + \langle \hat{y}_i^2, x - \hat{x}\rangle$, for $i \in I$, to get

$$-g_i^2(x) + \left(g_i^1(\hat{x}) + \langle \hat{y}_i^1, x - \hat{x}\rangle\right) \leq -g_i^2(\hat{x}) + g_i^1(\hat{x}) + \langle \hat{y}_i^1 - \hat{y}_i^2, x - \hat{x}\rangle.$$

By again referring to the definition of $g_{i,\hat{x}}$, we also obtain (6.2.19) from the previous inequality by multiplying it by $-\hat{\lambda}$. So, adding (6.2.18) to (6.2.19) we arrive to the inequality

$$f_{i,\hat{x}}(x) - \hat{\lambda} g_{i,\hat{x}}(x) \geq f_i(\hat{x}) - \hat{\lambda} g_i(\hat{x}) + \left\langle \hat{x}_i^1 - \hat{x}_i^2 - \hat{\lambda}\left(\hat{y}_i^1 - \hat{y}_i^2\right), x - \hat{x}\right\rangle.$$

By invoking the definition of $F_{\hat{x}}(x)$ we get

$$F_{\hat{x}}(x) \geq f_i(\hat{x}) - \hat{\lambda} g_i(\hat{x}) + \left\langle \hat{x}_i^1 - \hat{x}_i^2 - \hat{\lambda}\left(\hat{y}_i^1 - \hat{y}_i^2\right), x - \hat{x}\right\rangle. \tag{4.3.12}$$

On the other hand, the subgradient inequality $h_j^1(x) \geq h_j^1(\hat{x}) + \langle \hat{z}_j^1, x - \hat{x}\rangle$, for $j \in J$, gives

$$h_j^1(x) - \left(h_j^2(\hat{x}) + \langle \hat{z}_j^2, x - \hat{x}\rangle\right) \geq h_j^1(\hat{x}) - h_j^2(\hat{x}) + \langle \hat{z}_j^1 - \hat{z}_j^2, x - \hat{x}\rangle,$$

which means that

$$h_{j,\hat{x}}(x) \geq h_j(\hat{x}) + \langle \hat{z}_j^1 - \hat{z}_j^2, x - \hat{x}\rangle. \tag{4.3.13}$$

Multiplying both sides of (6.2.20) by $\hat{\mu}_i$, for all $i \in I$, and (6.2.21) by $\hat{\nu}_j$, for all $j \in J$, summing the resulting inequalities, and taking into account (6.2.10) and (6.2.17), we get

$$F_{\hat{x}}(x) \sum_{i \in I} \hat{\mu}_i + \sum_{j \in J} \hat{\nu}_j h_{j,\hat{x}}(x) \geq 0 \quad \text{for all } x \in C. \tag{4.3.14}$$

In particular, for $x \in X_{\hat{x}}$ we have $h_{j,\hat{x}}(x) \leq 0$ for all $j \in J$, which implies that $F_{\hat{x}}(x) \sum_{i \in I} \hat{\mu}_i \geq 0$. It is clear that (6.2.11) and (6.2.23) imply that $\sum_{i \in I} \hat{\mu}_i \neq 0$, since othrwise we should have

$$\max_{j \in J} h_{j,\hat{x}}(x) \geq \sum_{j \in J} \hat{\nu}_j h_{j,\hat{x}}(x) \geq 0 \quad \text{for all } x \in C,$$

thereby contradicting (6.2.11) with $z_j^2(\hat{x}) = \hat{z}_j^2 \in \partial h_j^2(\hat{x})$ for all $j \in J$. Therefore, $F_{\hat{x}}(x) \geq 0$. Let $x_{\hat{x}}$ be a global minimum of $F_{\hat{x}}$ over $X_{\hat{x}}$. Then $x_{\hat{x}} \in X_{\hat{x}}$, and since by Proposition 4.2.1, Item 1, $F_{\hat{x}}(x_{\hat{x}}) \leq 0$, we conclude that $F_{\hat{x}}(x_{\hat{x}}) = 0$. This achieves the proof. □

Proposition 4.3.1. Let $\hat{x} \in X$ and assume that for all $i \in I$ and $j \in J$, $\partial f_i^1(\hat{x}) - \partial f_i^2(\hat{x}) = \partial^c[f_i^1 - f_i^2](\hat{x})$, $\partial g_i^1(\hat{x}) - \partial g_i^2(\hat{x}) = \partial^c[g_i^1 - g_i^2](\hat{x})$ and $\partial h_j^1(\hat{x}) - \partial h_j^2(\hat{x}) = \partial^c[h_j^1 - h_j^2](\hat{x})$, where ∂^c stands for the Clarke subdifferential. If $\lambda(\hat{x}) \geq 0$, the functions f_i and $-g_i$, for $i \in I$, are regular at $\hat{x}$, in the sense of Clarke, and (6.2.11) holds, then (6.2.9) and (6.2.10) imply Clarke stationary conditions, that is, for all $j \in J$, there exist $\hat{\beta}_j \geq 0$ such that

$$0 \in \partial^c \lambda(\hat{x}) + \sum_{j \in J} \hat{\beta}_j \partial^c h_j(\hat{x}) + N_C(\hat{x}),$$

with the equalities $\hat{\beta}_j h_j(\hat{x}) = 0$ for all $j \in J$.

Proof. Let $\hat{x} \in X$, $\hat{\lambda} = \lambda(\hat{x})$ and assume that we have (6.2.9) and (6.2.10). Remark first that from the first equality in (6.2.10) we have $\hat{\mu}_i f_i(\hat{x})/g_i(\hat{x}) = \hat{\mu}_i \hat{\lambda}$ for all $i \in I$. Therefore, taking also into account the assumptions made in the beginning of the proposition, we can write (6.2.9) as

$$0 \in \sum_{i \in I} \hat{\mu}_i \left[\partial^c f_i(\hat{x}) - \frac{f_i(\hat{x})}{g_i(\hat{x})} \partial^c g_i(\hat{x}) \right] + \sum_{j \in J} \hat{\nu}_j \partial^c h_j(\hat{x}) + N_C(\hat{x}).$$

Thus,

$$0 \in \sum_{i \in I} \hat{\mu}_i g_i(\hat{x}) \left[\frac{g_i(\hat{x}) \partial^c f_i(\hat{x}) - f_i(\hat{x}) \partial^c g_i(\hat{x})}{g_i(\hat{x})^2} \right] + \sum_{j \in J} \hat{\nu}_j \partial^c h_j(\hat{x}) + N_C(\hat{x}). \quad (4.3.15)$$

From the the equality $\hat{\mu}_i f_i(\hat{x})/g_i(\hat{x}) = \hat{\mu}_i \hat{\lambda}$ we see that $\hat{\lambda} = f_i(\hat{x})/g_i(\hat{x})$ if $\hat{\mu}_i \neq 0$. For such $i \in I$, i.e. for which $\hat{\mu}_i \neq 0$, we have $f_i(\hat{x}) \geq 0$ since by assumption $\hat{\lambda} \geq 0$. On the other hand, the functions f_i and g_i, for $i \in I$, are locally Lipschitz as they are difference of convex functions. With the assumption that f_i and $-g_i$ are regular at $\hat{x}$, in the sense of Clarke, it follows from [30, Proposition 2.3.14] that

$$\partial^c \left[\frac{f_i}{g_i} \right](\hat{x}) = \frac{g_i(\hat{x}) \partial f_i(\hat{x}) - f_i(\hat{x}) \partial g_i(\hat{x})}{g_i(\hat{x})^2}.$$

From this equality, we write (6.2.27) as

$$0 \in \sum_{i \in I} \hat{\mu}_i g_i(\hat{x}) \partial^c \left[\frac{f_i}{g_i} \right](\hat{x}) + \sum_{j \in J} \hat{\nu}_j \partial^c h_j(\hat{x}) + N_C(\hat{x}). \quad (4.3.16)$$

Notice that $\sum_{i\in I}\hat{\mu}_i \neq 0$, due to assumption (6.2.11), as justified by (6.2.23) in the proof of the reciprocal assertion in Theorem 4.3.1. Define

$$\hat{\alpha}_i = \frac{\hat{\mu}_i g_i(\hat{x})}{\sum_{i\in I}\hat{\mu}_i g_i(\hat{x})} \text{ and } \hat{\beta}_j = \frac{\hat{\nu}_j}{\sum_{i\in I}\hat{\mu}_i g_i(\hat{x})}.$$

Since $N_C(\hat{x})$ is a cone, (4.3.16) entails that

$$0 \in \sum_{i\in I}\hat{\alpha}_i \partial^c \left[\frac{f_i}{g_i}\right](\hat{x}) + \sum_{j\in J}\hat{\beta}_j \partial^c h_j(\hat{x}) + N_C(\hat{x}). \tag{4.3.17}$$

On the other hand, the equality $\hat{\mu}_i f_i(\hat{x})/g_i(\hat{x}) = \hat{\mu}_i\hat{\lambda}$ entails that $\hat{\mu}_i = \hat{\alpha}_i = 0$ if $f_i(\hat{x})/g_i(\hat{x}) < \lambda(\hat{x})$. Therefore,

$$\sum_{i\in I}\hat{\alpha}_i \partial^c \left[\frac{f_i}{g_i}\right](\hat{x}) = \sum_{i\in I(\hat{x})}\hat{\alpha}_i \partial^c \left[\frac{f_i}{g_i}\right](\hat{x}),$$

where $I(\hat{x}) = \{i \in I \mid f_i(\hat{x})/g_i(\hat{x}) = \lambda(\hat{x})\}$. By our regularity assumption and [30, Proposition 2.3.14], f_i/g_i is regular. Remembering the definition of $\lambda(\cdot)$ and referring to [30, Proposition 2.3.12] or [11, Theorem 3.23] the last equality, together with (4.3.17), imply that

$$0 \in \partial^c\lambda(\hat{x}) + \sum_{j\in J}\hat{\beta}_j \partial^c h_j(\hat{x}) + N_C(\hat{x}).$$

The second equality in (6.2.10) implies that $\hat{\nu}_j h_j(\hat{x}) = 0$, which gives $\hat{\beta}_j h_j(\hat{x}) = 0$. This achieves the proof. □

Remark 4.3.1. Notice that the assumptions made in the beginning of Proposition 4.3.1 are fulfilled, in particular, if one of the two components of the dc decomposition is smooth, see e.g. [55, Propositions 1 and 2].

The next result shows that the solutions of (P) satisfy the optimality conditions (6.2.9) and (6.2.10).

Theorem 4.3.2. If $\bar{x} \in X$ is an optimal solution of (P) and $\bar{\lambda} = \lambda(\bar{x})$, then for all $i \in I$ and $j \in J$, there exist $\bar{\mu}_i, \bar{\nu}_j \geq 0$, with $\sum_{i\in I}\bar{\mu}_i + \sum_{j\in J}\bar{\nu}_j = 1$, such that

$$0 \in \sum_{i\in I}\bar{\mu}_i\left[\partial f_i^1(\bar{x}) - \partial f_i^2(\bar{x}) - \bar{\lambda}\left(\partial g_i^1(\bar{x}) - \partial g_i^2(\bar{x})\right)\right] + \sum_{j\in J}\bar{\nu}_j\left[\partial h_j^1(\bar{x}) - \partial h_j^2(\bar{x})\right] + N_C(\bar{x})$$

with $\bar{\mu}_i[f_i(\bar{x}) - \bar{\lambda}g_i(\bar{x})] = 0$, for all $i \in I$ and $\bar{\nu}_j h_j(\bar{x}) = 0$, for all $j \in J$.

Proof. Let $\bar{x} \in X$ be an optimal solution of (P). Remember that we showed in the proof of Proposition 4.2.2 that $F_{\bar{x}}(x) \geq 0$ for all $x \in X$. Then we also have $F_{\bar{x}}(x) \geq 0$ for all $x \in X_{\bar{x}}$, since $X_{\bar{x}} \subset X$, see Remark 6.2.1. On the other hand $\bar{x} \in X_{\bar{x}}$, see Remark 6.2.1. Therefore, $\bar{x}$ globally minimizes the convex function $F_{\bar{x}}$ over the convex set $X_{\bar{x}}$, since $F_{\bar{x}}(\bar{x}) = 0$. Then it suffices to use Theorem 4.3.1, with $\hat{x} = \bar{x}$ and $x_{\hat{x}} = \bar{x}$, to conclude. □ □

4.4 DC-Dinkelbach-type Algorithm for DC-GFP

Before describing our algorithm, remark that one can write

$$\lambda(x) := \max_{i \in I} \frac{\left(f_i^1(x) - f_i^2(x)\right)/\omega_i}{\left(g_i^1(x) - g_i^2(x)\right)/\omega_i} \tag{4.4.1}$$

for all $\omega_i > 0$, $i \in I$. Even if this transformation has no effect on λ, it has a strong effect on the parametric subproblems. This remark has been pointed out in [36] with a special choice of $\omega_i > 0$, $i \in I$, and gave rise to a new Dinkelbach-type algorithm, more efficient than the original Dinkelbach-type one. It also has been used in [107] and gave significant improvement to the rate of convergence. For computational reasons, we will write λ as in (6.2.28). With this artifice, the parametrized function defined in the begining of the chapter takes the form

$$F_y(x) = \max_{i \in I} \left[\frac{f_{i,y}(x) - \lambda(y) g_{i,y}(x)}{\omega_i} \right]$$

where $f_{i,y}$ and $g_{i,y}$, for $i \in I$, are as defined in (6.2.4) and (6.2.5) with arbitrary subgradients $x_i^2(y) \in \partial f_i^2(y)$ and $y_i^l(y) \in \partial g_i^l(y)$, where $l = 1$ if $\lambda(y) \geq 0$ and $l = 2$ otherwise. All the previous results remain valid, it suffices to replace $f_{i,y}$ and $g_{i,y}$, respectively by $f_{i,y}/\omega_i$ and $g_{i,y}/\omega_i$.

With the insight of the previous section's results, we will develop an algorithm by approximating the unknown function $F_{\bar{x}}$ (since $\bar{x}$ is unknown), at each step k by $F_{x^k} =: F_k$, where x^k is a global minimum of the convex function $F_{x^{k-1}} =: F_{k-1}$ over the convex set $X_{x^{k-1}} =: X_{k-1}$, where the latter is an inner approximating convex set of the nonconvex set X.

Now we are ready to describe our algorithm.

Algorithm 3 DC-Dinkelbach-type Algorithm for DC-GFP

0. Let $\{\varepsilon_k\}$ be a sequence of nonnegative reals such that $\sum_{k\geq 0}\varepsilon_k < \infty$. Choose $x^0 \in X$ and let $k = 0$.
1. At step k we have x^k, $\lambda_k = \lambda(x^k)$, $\varepsilon_k \geq 0$, $\bar{\omega} \geq \omega_i^k \geq \underline{\omega} > 0$, $x_{i,k}^2 \in \partial f_i^2(x^k)$, $y_{i,k}^\ell \in \partial g_i^\ell(x^k)$, for all $i \in I$, with $\ell = 1$ if $\lambda_k \geq 0$, and $\ell = 2$ if $\lambda_k < 0$; and $z_{j,k}^2 \in \partial h_j^2(x^k)$ for all $j \in J$. Then find $x^{k+1} \in X_k$ such that
$$F_k(x^{k+1}) \leq \inf_{x\in X_k} F_k(x) + \varepsilon_k,$$
where
$$F_k(x) = \max_{i\in I}\left[\frac{f_{i,k}(x) - \lambda_k g_{i,k}(x)}{\omega_i^k}\right],$$
and $X_k = \{x \in C \mid h_{j,k}(x) \leq 0, \forall j \in J\}$, with $f_{i,k} = f_{i,x^k}$, $g_{i,k} = g_{i,x^k}$ and $h_{j,k} = h_{j,x^k}$ are defined as in Eqs. (6.2.4) to (6.2.6) with $x_i^2(x^k) = x_{i,k}^2$, $y_i^\ell(x^k) = y_{i,k}^\ell$, $\ell = 1,2$, and $z_j^2(x^k) = z_{j,k}^2$, for $i \in I$ and $j \in J$.
2. If $F_k(x^{k+1}) = 0$ stop, else for all $i \in I$ and $j \in J$, choose $z_{j,k+1}^2 \in \partial h_j^2(x^{k+1})$, $x_{i,k+1}^2 \in \partial f_i^2(x^{k+1})$ and $y_{i,k+1}^\ell \in \partial g_i^\ell(x^{k+1})$, with $\ell = 1$ if $\lambda(x^{k+1}) \geq 0$, and $\ell = 2$ if $\lambda(x^{k+1}) < 0$. Set $\lambda_{k+1} = \lambda(x^{k+1})$, $k = k+1$ and return to step 1.

Remark 4.4.1.

1. A possible choice for the weights is $\omega_i^k = g_i(x^k)$. It has been used first in [36] and later in [107] where numerical tests confirmed its efficiency. This choice may be used in the previous algorithm.

2. Notice that there is no need to a starting feasible point x^0, i.e. $x^0 \in X$, and one can choose $x^- \in \mathbb{R}^n$ such that $g_i^1(x^-) \neq g_i^2(x^-)$ for all $i \in I$, set $\lambda = \lambda(x^-)$ and minimize the convex function F_{x^-} on the convex set X_{x^-} to get a point $x^0 \in X$, since $X_{x^-} \subset X$.

To establish the convergence of the sequence $\{\lambda_k\}$, we need the following well known lemma.

Lemma 4.4.1. Let $\{\varepsilon_k\}$ be a sequence of positive reals such that $\sum_{k\geq 0}\varepsilon_k < \infty$, and let $\{u_k\}$ be a sequence such that $u_{k+1} \leq u_k + \varepsilon_k$ for all $k \in \mathbb{N}$. Then $\{u_k\}$ converges to some $\hat{u} \in \mathbb{R}\cup\{-\infty\}$.

Proof. See, e.g., [101, §2.2.1, Lemma 2] and [108, Lemma 2.1] for a more general form of this lemma. □

Recall that we used the notation $f_i(x) = f_i^1(x) - f_i^2(x)$ and $g_i(x) = g_i^1(x) -$

$g_i^2(x)$, for all $i \in I$.

The next result gives the convergence of the sequence $\{\lambda_k\}$ and a stopping criterion, which is the convergence of $\{F_k(x^{k+1})\}$ towards 0.

We denote and assume

$$\delta := \inf_{x\in X} \min_{i\in I} g_i(x) > 0 \text{ and } \Delta := \sup_{x\in X} \max_{i\in I} g_i(x) < \infty.$$

Proposition 4.4.1. If $\sum_{k\geq 0} \varepsilon_k < \infty$, the sequence $\{\lambda_k\}$ converges to some $\hat{\lambda} \geq \bar{\lambda}$, where $\bar{\lambda}$ is the minimum value of (P), and $\{F_k(x^{k+1})\}$ converges to 0.

Proof. From the definition of x^{k+1} we have

$$\varepsilon_k + F_k(x) \geq F_k(x^{k+1}) \quad \text{for all } x \in X_k. \tag{4.4.2}$$

For $x = x^k$ we get $F_k(x^{k+1}) \leq F_k(x^k) + \varepsilon_k = \varepsilon_k$, where the equality $F_k(x^k) = 0$ follows from Proposition 4.2.1, Item 1, since $F_k(x^k) = F_{x^k}(x^k)$. From Lemma 4.2.1, for all $x \in \mathbb{R}^n$, we have

$$F_k(x) \geq \max_{i\in I} \left[\frac{f_i(x) - \lambda_k g_i(x)}{\omega_i^k} \right].$$

Therefore,

$$\begin{aligned} \varepsilon_k \geq F_k(x^{k+1}) &\geq \max_{i\in I} \left[\frac{f_i(x^{k+1}) - \lambda_k g_i(x^{k+1})}{\omega_i^k} \right] \\ &\geq \frac{f_{i_k}(x^{k+1}) - \lambda_k g_{i_k}(x^{k+1})}{\omega_{i_k}^k} \\ &= \frac{g_{i_k}(x^{k+1})}{\omega_{i_k}^k} (\lambda_{k+1} - \lambda_k) \end{aligned} \tag{4.4.3}$$

where i_k satisfies $\lambda_{k+1} = f_{i_k}(x^{k+1})/g_{i_k}(x^{k+1})$. This implies that $\lambda_{k+1} \leq \lambda_k + \varepsilon_k \bar{\omega}/\delta$, where we used the assumptions $g_{i_k}(x^{k+1}) \geq \delta > 0$ and $0 < \omega_i^k \leq \bar{\omega}$ for all $k \in \mathbb{N}$ and $i \in I$. Since $\lambda_k \geq \bar{\lambda}$ and $\sum_{k\geq 0} \varepsilon_k < \infty$, we conclude from Lemma 6.2.4 that the sequence $\{\lambda_k\}$ converges to some $\hat{\lambda} \geq \bar{\lambda}$.

On the other hand, from (6.2.29) we get

$$\begin{aligned}\varepsilon_k \geq F_k(x^{k+1}) &\geq \frac{g_{i_k}(x^{k+1})}{\omega_{i_k}^k}(\lambda_{k+1}-\lambda_k)\\ &= \frac{g_{i_k}(x^{k+1})}{\omega_{i_k}^k}\left(\lambda_{k+1}-\hat{\lambda}\right)+\frac{g_{i_k}(x^{k+1})}{\omega_{i_k}^k}\left(\hat{\lambda}-\lambda_k\right)\\ &\geq \frac{\delta}{\bar{\omega}}\left(\lambda_{k+1}-\hat{\lambda}\right)+\frac{\Delta}{\underline{\omega}}\left(\hat{\lambda}-\lambda_k\right)\end{aligned}$$

from which we conclude that $\{F_k(x^{k+1})\}$ converges to 0. □

Remark 4.4.2. If $\sum_{k\geq 0}\varepsilon_k < \infty$, and the set $\{x \in \mathbb{R}^n \mid \lambda(x) \leq \lambda(x^0)+\sum_{k\geq 0}\varepsilon_k\bar{\omega}/\delta\}$ is bounded, for the starting point x^0 (wich is the case, e.g., if $\lambda(.)$ is inf-compact), then the sequence $\{x^k\}$ is bounded. Indeed, we showed in the last proof that $\lambda_{k+1} \leq \lambda_k + \varepsilon_k\bar{\omega}/\delta$. Therefore, $\lambda_{k+1} \leq \lambda_0 + \sum_{i=0}^{k}\varepsilon_i\bar{\omega}/\delta$ implying that $x^{k+1} \in \{x \in \mathbb{R}^n \mid \lambda(x) \leq \lambda(x^0) + \sum_{k\geq 0}\varepsilon_k\bar{\omega}/\delta\}$.

Proposition 4.4.2. For all $k \in \mathbb{N}$, there exists $\alpha^k \in \Sigma$ such that

$$\alpha_0^k\left(F_k(x)-F_k(x^{k+1})+\varepsilon_k\right)+\sum_{j\in J}\alpha_j^k h_{j,k}(x) \geq 0 \quad \text{for all } x \in C,$$

where $\Sigma = \left\{(\alpha_j)_{j\in J\cup\{0\}} \geq 0 \mid \alpha_0 + \sum_{j\in J}\alpha_j = 1\right\}$. Moreover,

$$0 \geq \sum_{j\in J}\alpha_j^k h_{j,k}(x^{k+1}) \geq -\alpha_0^k\varepsilon_k \text{ and } 0 \geq \sum_{j\in J}\alpha_j^k h_j(x^k) \geq \alpha_0^k\left(F_k(x^{k+1})-\varepsilon_k\right).$$

So, if $\sum_{k\geq 0}\varepsilon_k < \infty$, the sequences $\left\{\sum_{j\in J}\alpha_j^k h_{j,k}(x^{k+1})\right\}$ and $\left\{\sum_{j\in J}\alpha_j^k h_j(x^k)\right\}$ converge to 0 as k tends to ∞.

Proof. Recall that from the definition of x^{k+1} we have

$$F_k(x)-F_k(x^{k+1})+\varepsilon_k \geq 0 \quad \text{for all } x \in X_k, \tag{4.4.4}$$

and that $X_k := \{x \in C \mid h_{j,k}(x) \leq 0, \forall j \in J\}$. Define the function

$$h^k(x) = \max_{j\in J} h_{j,k}(x).$$

Obviously, $h^k(x) \leq 0$ if and only if $h_{j,k}(x) \leq 0$ for all $j \in J$. Define also the function

$$\mathcal{F}_k(x) = \max\left[F_k(x)-(F_k(x^{k+1})-\varepsilon_k), h^k(x)\right].$$

It is straightforward to show that $\mathcal{F}_k(x) \geq 0$ for all $x \in C$. Indeed, if $x \in X_k$, i.e. $x \in C$ and $h^k(x) \leq 0$, then (4.4.4) entails that $\mathcal{F}_k(x) \geq 0$, and if $x \in C$ but $x \notin X_k$, i.e. $h_k(x) > 0$ then we have $\mathcal{F}_k(x) > 0$. On the other hand, the function $\mathcal{F}_k$ may be expressed by

$$\mathcal{F}_k(x) = \max_{\alpha\in\Sigma}\left[\alpha_0\left(F_k(x) - F_k(x^{k+1}) + \varepsilon_k\right) + \sum_{j\in J}\alpha_j h_{j,k}(x)\right],$$

where Σ is as defined in the proposition. Now, the function defined on $C \times \Sigma$ by

$$(x,\alpha) \longmapsto \alpha_0\left(F_k(x) - F_k(x^{k+1}) + \varepsilon_k\right) + \sum_{j\in J}\alpha_j h_{j,k}(x)$$

is convex with respect to $x \in C$ and linear with respect to $\alpha \in \Sigma$, with C being convex and Σ being convex and compact. Thus, by the minimax theory, see e.g. [45, Theorem 2] or [117, Corollary 3.3], we have

$$\begin{aligned}\inf_{x\in C}\mathcal{F}_k(x) &= \inf_{x\in C}\max_{\alpha\in\Sigma}\left[\alpha_0\left(F_k(x) - F_k(x^{k+1}) + \varepsilon_k\right) + \sum_{j\in J}\alpha_j h_{j,k}(x)\right] \\ &= \max_{\alpha\in\Sigma}\inf_{x\in C}\left[\alpha_0\left(F_k(x) - F_k(x^{k+1}) + \varepsilon_k\right) + \sum_{j\in J}\alpha_j h_{j,k}(x)\right]. \qquad (4.4.5)\end{aligned}$$

Therefore, for all $k \in \mathbb{N}$, there exists $\alpha^k \in \Sigma$ such that

$$\alpha_0^k\left(F_k(x) - F_k(x^{k+1}) + \varepsilon_k\right) + \sum_{j\in J}\alpha_j^k h_{j,k}(x) \geq 0 \quad \text{for all } x \in C.$$

Indeed, the function

$$\Sigma \ni \alpha \longmapsto \inf_{x\in C}\left[\alpha_0\left(F_k(x) - F_k(x^{k+1}) + \varepsilon_k\right) + \sum_{j\in J}\alpha_j h_{j,k}(x)\right]$$

is upper semicontinuous on Σ, since it is the pointwise infimum of a family of linear functions, a fortiori continuous, and then achieves its maximum on the compact set Σ. Therefore, for all $k \in \mathbb{N}$, there exist $\alpha^k \in \Sigma$ such that

$$\begin{aligned}&\max_{\alpha\in\Sigma}\inf_{x\in C}\left[\alpha_0\left(F_k(x) - F_k(x^{k+1}) + \varepsilon_k\right) + \sum_{j\in J}\alpha_j h_{j,k}(x)\right] \\ &= \inf_{x\in C}\left[\alpha_0^k\left(F_k(x) - F_k(x^{k+1}) + \varepsilon_k\right) + \sum_{j\in J}\alpha_j^k h_{j,k}(x)\right].\end{aligned}$$

Since we showed that $\mathcal{F}_k(x) \geq 0$ for all $x \in C$, then from (4.4.5) we obtain

$$\alpha_0^k \left(F_k(x) - F_k(x^{k+1}) + \varepsilon_k \right) + \sum_{j \in J} \alpha_j^k h_{j,k}(x) \geq 0 \quad \text{for all } x \in C. \tag{4.4.6}$$

The rest follows by letting, once $x = x^{k+1}$ and once $x = x^k$, in the previous inequality. □

Now we turn our attention to the convergence of the sequence $\{x^k\}$.

Theorem 4.4.1. Assume that $\sum_{k \geq 0} \varepsilon_k < \infty$ and that the sequence $\{x^k\}$ is bounded. Let $\hat{x}$ be a cluster point of $\{x^k\}$ and $\hat{\lambda} = \lambda(\hat{x})$. Then for all $i \in I$ and $j \in J$, there exists $\hat{\omega}_i$ a cluster point of $\{\omega_i^k\}$, there exist $\hat{\mu}_i, \hat{\nu}_j \geq 0$, with $\sum_{i \in I} \hat{\mu}_i + \sum_{j \in J} \hat{\nu}_j = 1$ such that

$$0 \in \sum_{i \in I} \frac{\hat{\mu}_i}{\hat{\omega}_i} \left[\partial f_i^1(\hat{x}) - \partial f_i^2(\hat{x}) - \hat{\lambda} \left(\partial g_i^1(\hat{x}) - \partial g_i^2(\hat{x}) \right) \right] + \sum_{j \in J} \hat{\nu}_j \left[\partial h_j^1(\hat{x}) - \partial h_j^2(\hat{x}) \right] + N_C(\hat{x}),$$

with $\hat{\mu}_i [f_i(\hat{x}) - \hat{\lambda} g_i(\hat{x})] = 0$ and $\hat{\nu}_j h_j(\hat{x}) = 0$, for all $i \in I$ and $j \in J$.

Proof. Let $\hat{x}$ be a cluster point of the sequence $\{x^k\}$, and let K be an infinite subset of $\mathbb{N}$ such that the subsequence $\{x^k\}_{k \in K}$ converges to $\hat{x}$. Since $x^k \in X_k \subset X$ and X is closed, we have $\hat{x} \in X$. For each $k \in \mathbb{N}$, let $\{\alpha^k\} \subset \Sigma$ as stated in Proposition 4.4.2. Let the sequences $\{x_{i,k}^2\}$, $\{y_{i,k}^\ell\}$, for $i \in I$, $\ell = 1, 2$, and $\{z_{j,k}^2\}$, for $j \in J$, be as defined in Algorithm 3. We recall here that these sequences are bounded by the boundedness of $\{x^k\}$, see e.g. [64, Propositions 6.2.2].

Next, for $k \in K$, we consider subsequences of $\{\alpha^k\}$, $\{z_{j,k}^2\}$, for $j \in J$, $\{\omega_i^k\}$, $\{x_{i,k}^2\}$, $\{y_{i,k}^\ell\}$, for $i \in I$, say for $k \in K'$ an infinite subset of $\mathbb{N}$, converging respectively to $\hat{\alpha}$, $\hat{z}_j^2$, $\hat{\omega}_i$, $\hat{x}_i^2$, $\hat{y}_i^\ell$, where $\ell = 1$ if $\hat{\lambda} > 0$ and $\ell = 2$ if $\hat{\lambda} < 0$. (Notice that $\lambda_k > 0$ and $\ell = 1$ (resp. $\lambda_k < 0$ and $\ell = 2$) for k large, when $\hat{\lambda} > 0$ (resp. $\hat{\lambda} < 0$)). Therefore, $\hat{z}_j^2 \in \partial h_j^2(\hat{x})$, $\hat{x}_i^2 \in \partial f_i^2(\hat{x})$ and $\hat{y}_i^\ell \in \partial g_i^\ell(\hat{x})$, see e.g. [64, Propositions 6.2.1]. With these elements we define the function

$$F_{\hat{x}}(x) = \max_{i \in I} \left[\frac{f_{i,\hat{x}}(x) - \hat{\lambda} g_{i,\hat{x}}(x)}{\hat{\omega}_i} \right],$$

where the functions $f_{i,\hat{x}}$, $g_{i,\hat{x}}$ and $h_{j,\hat{x}}$ are defined as in Eqs. (6.2.4) to (6.2.6) with $x_i^2(\hat{x}) = \hat{x}_i^2$, $y_i^\ell(\hat{x}) = \hat{y}_i^\ell$, for $l = 1, 2$, $i \in I$, and $z_j^2(\hat{x}) = \hat{z}_j^2$, $j \in J$. In the case $\hat{\lambda} = 0$, we ignore the term $\hat{\lambda} g_{i,\hat{x}}(x)$ in the definition of $F_{\hat{x}}(x)$.

By invoking Proposition 4.4.1 and passing to the limit in (4.4.6), as k tends to ∞, $k \in K'$, we arrive to

$$\hat{\alpha}_0 F_{\hat{x}}(x) + \sum_{j\in J} \hat{\alpha}_j h_{j,\hat{x}}(x) \geq 0 \quad \text{for all } x \in C.$$

Therefore, for all $x \in X_{\hat{x}}$ we have $\hat{\alpha}_0 F_{\hat{x}}(x) \geq 0$. So, if $\hat{\alpha}_0 \neq 0$ then since $F_{\hat{x}}(\hat{x}) = 0$, we deduce that $\hat{x}$ globally minimizes the convex function $F_{\hat{x}}$ over $X_{\hat{x}}$. Then it suffices to use Theorem 4.3.1 to conclude. Now if $\hat{\alpha}_0 = 0$, we get

$$\sum_{j\in J} \hat{\alpha}_j h_{j,\hat{x}}(x) \geq 0 \quad \text{for all } x \in C. \tag{4.4.7}$$

Since $h_{j,\hat{x}}(\hat{x}) = h_j(\hat{x}) \leq 0$ we obtain $\hat{\alpha}_j h_j(\hat{x}) = 0$ for all $j \in J$. Therefore, $\hat{x}$ minimizes, on C, the function $x \mapsto \sum_{j\in J} \hat{\alpha}_j h_{j,\hat{x}}(x)$. It follows that

$$0 \in \sum_{j\in J} \hat{\alpha}_j \partial h_{j,\hat{x}}(\hat{x}) + N_C(\hat{x}) \subset \sum_{j\in J} \hat{\alpha}_j \left[\partial h_j^1(\hat{x}) - \partial h_j^2(\hat{x})\right] + N_C(\hat{x}).$$

The desired result is fulfilled with $\hat{\mu}_i = 0$, for all $i \in I$, and $\hat{\nu}_j = \hat{\alpha}_j$, for all $j \in J$. □

Remark 4.4.3. If (6.2.11) is satisfied, i.e., for some $x \in C$ we have

$$\max_{j\in J} h_{j,\hat{x}}(x) < 0 \quad \text{for all } z_j^2(\hat{x}) \in \partial h_j^2(\hat{x}) \text{ and } j \in J,$$

then $\hat{\alpha}_0 \neq 0$, since (6.2.31) becomes impossible.

Chapter 5

Proximal bundle methods for generalized fractional programs with ratios of difference of convex functions

In this chapter, we propose an approximating scheme based on the proximal point algorithm, for solving generalized fractional programs with ratios of difference of convex (DC) functions, with DC constraints, wich we have already called, DC-GFP. Such problems are generally nonsmooth and nonconvex. We take advantage of the convexity property of the associated approximate parametric problem of DC-GFP studied in Chapter 4. The proposed methods generate a sequence of approximate solutions that converges to critical points satisfying necessary optimality conditions of KKT type.

5.1 Inroduction

In this chapter, we consider fractional programming problems whose objective function is the maximum of finite ratios of difference of convex (DC) functions, with DC constraints. More precisely, we consider problem of the form

$$(P) \qquad \bar{\lambda} = \min_{x \in X} \left\{ \lambda(x) := \max_{i \in I} \frac{f_i^1(x) - f_i^2(x)}{g_i^1(x) - g_i^2(x)} \right\},$$

where $X = \{x \in C \mid h_j^1(x) - h_j^2(x) \leq 0, \, j \in J\}$, with $C \subset \mathbb{R}^n$ a nonempty, closed convex set, I and J two finite index sets, and the functions f_i^ℓ, g_i^ℓ, for $i \in I$, and h_j^ℓ, for $j \in J$ and $\ell = 1,2$ are defined on $\mathbb{R}^n$ and convex, with $g_i^1 - g_i^2$ positive on X for all $i \in I$.

To simplify notations, we will put for all $i \in I$, $j \in J$ and $x \in \mathbb{R}^n$,

$$f_i(x) = f_i^1(x) - f_i^2(x), \quad g_i(x) = g_i^1(x) - g_i^2(x), \quad h_j(x) = h_j^1(x) - h_j^2(x)$$

and

$$h(x) = \max_{j \in J} h_j(x).$$

With the last notation we have, $X = \{x \in C \mid h(x) \leq 0\}$.

Problems of this form have been already studied in [23], where optimality conditions were obtained and amethod of resolution was proposed. They include ordinary convex programs, generalized fractional problems (GFP) with ratios of: convex functions, concave functions, convex and concave functions, concave and convex functions, etc.

For solving a GFP, there have been several primal Dinkelbach-type algorithms in the literature [17,33–36,107,108,122], and dual algorithms and results [1,2,13–15,21,22,24–26,37,42,43,67]. See [118–121] for a detailed bibliography. Also, see [7–9] for recent applications. These algorithms are based on auxiliary parametric problems having simpler structures than the original problem. For the primal algorithms, the auxiliary problems furnish sequences of approximate optimal values converging decreasingly to the optimal value of (P), whereas the sequences of values generated by the dual algorithms converge increasingly towards the optimal value of (P).

Apart from the situation when the parametric subproblems are convex, the primal approaches fail to work since it is assumed that one can compute their global minimum. In our situation, where both the numerators and the denominators are difference of convex (DC) functions, the objective functions of the auxiliary problems are also DC functions. For this reason, we will resort to DC techniques, see e.g., [80,124].

Another strategy was proposed in [122], which consists in applying bundle methods for solving a GFP. These methods consist in approximately solving the primal auxiliary problems associated with the GFP by using primal bundle methods. Recently, since the last algorithm is rather intended to solve linearly constrained GFPs, another primal bundle method, based this time on the extended method of centers [107], was proposed in [1] to deal with nonlinearly constrained GFPs. Very recently, a dual bundle method has been proposed in [26], also for solving such problems, this time without convexity assumptions.

The idea behind these works is the successful use of bundle methods in the context of the nonsmooth convex optimization, see e.g. [48, 63, 70, 71, 83, 84, 90–92, 116]. Recall that the bundle strategy is based on collecting information about the objective and constraints functions, by their values and subgradients from previous steps. This information is then used to construct models that approximate the original program. These models are usually polyhedral functions. Stabilized by a quadratic term, a type of prox-regularization, quadratic programs must be solved several times to approximate the original problem.

In this chapter, we use the prox-regularization principle to solve the convex approximate parametric problem associated to DC-GFP studied in Chapter 4, and we approximate from below the convex nonsmooth objective function of this problem in order to make it easier to solve. Our lower approximation is a piecewise linear convex function built piece by piece until a criterion measuring the quality of the approximation is satisfied. This criterion is related to the serious steps used in the bundle methods. With this criterion, the method can be viewed as a classical bundle method where after each serious step the value of the parameter λ is updated. We refer the reader to [31], for more details on the bundle method in convex programming and to [65, 98, 110] for the bundle method in the framework of variational inequalities.

In the first step we stated the results demonstrated in Chapter 4. Then we present a general approximation proximal method, based on the notion of (strong) c-approximation functions and we study the convergence of the sequences generated by Algorithm 4. Next, we construct of the piecewise linear convex approximations. Later, we present results to show the finite terminition of Algorithm 5.

5.2 Parametric approach for DC-GFP

In Dinkelbach-type algorithms, the associated parametric problem to (P), with the parameter μ, takes the form

$$(P_\mu) \qquad F(\mu) := \inf_{x\in X} \max_{i\in I} \{f_i(x) - \mu g_i(x)\}.$$

We recall that under mild assumptions, when $\mu = \bar{\lambda}$, the problems (P) and (P_μ) have the same optimal solutions set, see e.g. [35, Proposition 2.1] and [22, Lemma 1]. The problems (P_μ) appear as subproblems in the Dinkelbach-type algorithm [35, 36].

In all what follows, $\partial f_i^2(x)$, $\partial h_j^2(x)$ and $\partial g_i^\ell(x)$ will designate, respectively, the subdifferentials of the convex functions f_i^2, h_j^2 and g_i^ℓ at x, for $\ell = 1,2$. To begin our analysis, we define the function parametrized by $y \in X$,

$$F_y(x) := \max_{i\in I}\Big[f_{i,y}(x) - \lambda(y)g_{i,y}(x)\Big],$$

where

$$f_{i,y}(x) = f_i^1(x) - [f_i^2(y) + \langle x_i^2(y), x-y\rangle], \tag{5.2.1}$$

$$g_{i,y}(x) := \begin{cases} g_i^1(x) - [g_i^2(y) + \langle y_i^2(y), x-y\rangle] & \text{if } \lambda(y) < 0 \\ -g_i^2(x) + [g_i^1(y) + \langle y_i^1(y), x-y\rangle] & \text{if } \lambda(y) \geq 0 \end{cases} \tag{5.2.2}$$

with some $x_i^2(y) \in \partial f_i^2(y)$ and $y_i^\ell(y) \in \partial g_i^\ell(y)$, for $\ell = 1$ or $\ell = 2$. That is, we replace the functions f_i^2 and g_i^ℓ, $\ell = 1$ or $\ell = 2$, by their affine approximations at y, namely, $f_i^2(y)+\langle x_i^2(y), x-y\rangle$ and $g_i^\ell(y)+\langle y_i^\ell(y), x-y\rangle$, respectively.

Also, for all $x \in \mathbb{R}^n$ and $j \in J$, we define the functions $h_{j,y}$, parametrized by $y \in \mathbb{R}^n$, by

$$h_{j,y}(x) := h_j^1(x) - [h_j^2(y) + \langle z_j^2(y), x-y\rangle], \tag{5.2.3}$$

with some $z_j^2(y) \in \partial h_j^2(y)$, and consider the set

$$X_y = \Big\{x \in C \mid h_{j,y}(x) \leq 0, \forall j \in J\Big\}.$$

Notice that by the convexity assumptions made on the functions f_i^ℓ and g_i^ℓ, $\ell = 1,2$, the functions $f_{i,y}(\cdot)$ and $-\lambda(y)g_{i,y}(\cdot)$ are convex for all $i \in I$ and $y \in X$, and so is the function $F_y(\cdot)$. On the other hand, the convexity of the functions h_j^ℓ, $\ell = 1,2$, implies the convexity of the functions $h_{j,y}(\cdot)$, for all $j \in J$ and $y \in X$, and thus the convexity of the set X_y.

Remark 5.2.1. From the subgradient inequalities $h_j^2(x) \geq h_j^2(y) + \langle z_j^2(y), x - y\rangle$ for all $x, y \in \mathbb{R}^n$, $j \in J$, we conclude that $X_y \subset X$ for all $y \in \mathbb{R}^n$. On the other hand, $y \in X_y$ if and only if $y \in X$.

Now, for $y \in X$, we associate to (P) the convex approximate parametric problem

$$(P(y)) \qquad \inf_{x\in X_y}\Big\{F_y(x) := \max_{i\in I}\Big[f_{i,y}(x) - \lambda(y)g_{i,y}(x)\Big]\Big\},$$

and we denote by x_y the global minimum of F_y over X_y, if any.

Before starting to discuss our method of solving parametric problems, we begin first by some preliminary results demonstrated in Chapter 4.

Lemma 5.2.1. For all $x \in \mathbb{R}^n$ and $y \in X$ we have

$$\max_{i \in I} [f_i(x) - \lambda(y) g_i(x)] \leq F_y(x).$$

Proposition 5.2.1. For all $y \in X$ we have

1. $F_y(y) = 0$ and $F_y(x_y) \leq 0$,
2. $\lambda(x_y) \leq \lambda(y)$.

Notice that the Item 2 of Proposition 5.2.1 says that having a point $y \in X$, one obtain a better point for λ, which is x_y, by solving the convex program $(P(y))$.

In the next proposition we investigate relations between the problem (P) and the problem $(P(y))$, of minimizing F_y over X_y.

Proposition 5.2.2. If the problem (P) has a global optimal solution on X, say $\bar{x}$, then this solution actually globally minimizes $F_{\bar{x}}$ over $X_{\bar{x}}$, whatever are $x_i^2(\bar{x}) \in \partial f_i^2(\bar{x})$, $y_i^l(\bar{x}) \in \partial g_i^l(\bar{x})$, where $\ell = 1$ if $\lambda(\bar{x}) \geq 0$ and $\ell = 2$ otherwise; and $z_j^2(\bar{x}) \in \partial h_j^2(\bar{x})$ for $i \in I$ and $j \in J$. Conversely, for all global optimal solution $\bar{x}$ of (P), every optimal solution $x_{\bar{x}}$ of $F_{\bar{x}}$ over $X_{\bar{x}}$ also globally solves the problem (P).

With the previous results, the following theorem gives optimality conditions for (P), using only convex analysis tools and furnishes the necessary optimality condition $F_{\hat{x}}(x_{\hat{x}}) = 0$. This will be used as a natural stopping criterion for our expected algorithm.

Theorem 5.2.1. Let $\hat{x} \in X$ and $\hat{\lambda} = \lambda(\hat{x})$. If for all $i \in I$, there exist $x_i^2(\hat{x}) \in \partial f_i^2(\hat{x})$, $y_i^\ell(\hat{x}) \in \partial g_i^\ell(\hat{x})$, where $\ell = 1$ if $\hat{\lambda} \geq 0$ and $\ell = 2$ otherwise; and for all $j \in J$, there exist $z_j^2(\hat{x}) \in \partial h_j^2(\hat{x})$, such that $F_{\hat{x}}(x_{\hat{x}}) = 0$, where $x_{\hat{x}}$ is a global minimum of $F_{\hat{x}}$ on $X_{\hat{x}} := \{x \in C \mid h_{j,\hat{x}}(x) \leq 0, \forall j \in J\}$ then, for all $i \in I$, there exist $\hat{\mu}_i \geq 0$, and for all $j \in J$, there exist $\hat{\nu}_j \geq 0$, with $\sum_{i \in I} \hat{\mu}_i + \sum_{j \in J} \hat{\nu}_j = 1$, such that

$$0 \in \sum_{i \in I} \hat{\mu}_i \left[\partial f_i^1(\hat{x}) - \partial f_i^2(\hat{x}) - \hat{\lambda}\left(\partial g_i^1(\hat{x}) - \partial g_i^2(\hat{x})\right)\right] + \sum_{j \in J} \hat{\nu}_j \left[\partial h_j^1(\hat{x}) - \partial h_j^2(\hat{x})\right] + N_C(\hat{x}), \tag{5.2.4}$$

with the equalities

$$\hat{\mu}_i \left[f_i(\hat{x}) - \hat{\lambda} g_i(\hat{x})\right] = 0 \text{ and } \hat{\nu}_j h_j(\hat{x}) = 0 \tag{5.2.5}$$

for all $i \in I$ and $j \in J$, where $N_C(\hat{x})$ is the normal cone to C at $\hat{x}$.

The converse is true, that is $F_{\hat{x}}(x_{\hat{x}}) = 0$, for some $x_i^2(\hat{x}) \in \partial f_i^2(\hat{x})$, $y_i^\ell(\hat{x}) \in \partial g_i^l(\hat{x})$, for $l = 1, 2$ and $i \in I$, if in addition to (5.2.4) and (5.2.5), one has

$$\max_{j\in J} h_{j,\hat{x}}(x) < 0 \quad \text{for all } z_j^2(\hat{x}) \in \partial h_j^2(\hat{x}) \text{ and } j \in J, \tag{5.2.6}$$

for some $x \in C$.

Remark that one can write

$$\lambda(x) := \max_{i\in I} \frac{\left(f_i^1(x) - f_i^2(x)\right)/\omega_i}{\left(g_i^1(x) - g_i^2(x)\right)/\omega_i} \tag{5.2.7}$$

for all $\omega_i > 0$, $i \in I$. Even if this transformation has no effect on λ, it has a strong effect on the parametric subproblems. This remark has been pointed out in [36] with a special choice of $\omega_i > 0$, $i \in I$, and gave rise to a new Dinkelbach-type algorithm, more efficient than the original Dinkelbach type one. For computational reasons, we will write λ as in (5.2.7). With this artifice, the parametrized function defined in the begining of the chapter takes the form

$$F_y(x) = \max_{i\in I} \left[\frac{f_{i,y}(x) - \lambda(y) g_{i,y}(x)}{\omega_i} \right],$$

where $f_{i,y}$ and $g_{i,y}$, for $i \in I$, are as defined in (5.2.1) and (5.2.2) with arbitrary subgradients $x_i^2(y) \in \partial f_i^2(y)$ and $y_i^l(y) \in \partial g_i^l(y)$, where $l = 1$ if $\lambda(y) \geq 0$ and $l = 2$ otherwise. All the previous results remain valid, it suffices to replace $f_{i,y}$ and $g_{i,y}$, respectively by $f_{i,y}/\omega_i$ and $g_{i,y}/\omega_i$.

We will approximating the unknown function F_y (since y is unknown), at each step k by $F_{x^k} =: F_k$, where x^k is a global minimum of the convex function $F_{x^{k-1}} =: F_{k-1}$ over the convex set $X_{x^{k-1}} =: X_{k-1}$, where the latter is an approximating convex set of the nonconvex set X.

So with each step k we solve the following problem

$$(P(k)) \qquad \inf_{x\in X_k} F_k(x),$$

where

$$F_k(x) = F_{x^k}(x) := \max_{i\in I} \left[\frac{f_{i,k}(x) - \lambda_k g_{i,k}(x)}{\omega_i^k} \right],$$

and

$$f_{i,k}(x) = f_i^1(x) - [f_i^2(x^k) + \langle x_{i,k}^2, x - x^k\rangle], \tag{5.2.8}$$

$$g_{i,k}(x) := \begin{cases} g_i^1(x) - [g_i^2(x^k) + \langle y_{i,k}^2, x - x^k\rangle] & \text{if } \lambda_k < 0 \\ -g_i^2(x) + [g_i^1(x^k) + \langle y_{i,k}^1, x - x^k\rangle] & \text{if } \lambda_k \geq 0 \end{cases} \tag{5.2.9}$$

and $X_k = \left\{x \in \mathbb{R}^n \mid h_{j,k}(x) \leq 0, \forall j \in J\right\}$, with $h_{j,k}(x) = h_j^1(x) - [h_j^2(x^k) + \langle z_{j,k}^2, x - x^k\rangle]$ and $z_{i,k}^2 \in \partial h_i^2(x^k)$, $x_{i,k}^2 \in \partial f_i^2(x^k)$, $y_{i,k}^l \in \partial g_i^l(x^k)$, where $l = 1$ if $\lambda_k \geq 0$ and $l = 2$ otherwise, with $\lambda_k = \lambda(x^k)$ and $\omega_i^k > 0$, for all $i \in I$, $k \in \mathbb{N}$.

5.3 An inexact proximal point method

Before formaly stating our general approximating proximal algorithm for solving the problem $(P(k))$, we first give some helpful results which we will use later to interpret the algorithm.

The prox-regularization method consists in replacing the problem $(P(k))$ by the problem

$$(P_{\alpha_k}) \qquad \min_{x \in X_k} F_k(x) + \frac{1}{2\alpha_k}\|x - x^k\|^2 \tag{5.3.1}$$

where

$$F_k(x) = F_{x^k}(x) := \max_{i \in I}\left[\frac{f_{i,k}(x) - \lambda_k g_{i,k}(x)}{\omega_i^k}\right],$$

and $f_{i,k}$, $g_{i,k}$ for $i \in I$, are as defined in (5.2.8) and (5.2.9) with arbitrary subgradients $x_{i,k}^2 \in \partial f_i^2(x^k)$ and $y_{i,k}^l \in \partial g_i^l(x^k)$, where $l = 1$ if $\lambda_k \geq 0$ and $l = 2$ otherwise, with $\lambda_k = \lambda(x^k)$, $\omega_i^k > 0$ and $\alpha_k > 0$, for all $i \in I$, $k \in \mathbb{N}$.

In order to obtain an implementable algorithm, we only compute an approximate solution of this problem. Practically this will be done by approximating in problem (P_{α_k}) the nonsmooth convex function $F_k(x)$ by a convex function $\psi_k(x)$ in such a way that the problem

$$(AP_{\alpha_k}) \qquad \min_{x \in X_k} \psi_k(x) + \frac{1}{2\alpha_k}\|x - x^k\|^2 \tag{5.3.2}$$

is easier to solve exactly. The form of this function and how to construct it will be the subject of the next section. Here we only define the properties that the approximation $\psi_k(x)$ of $F_k(x)$ must satisfy so that the sequence $\{\lambda_k\}$ converges to $\hat{\lambda} \geq \bar{\lambda}$, where $\bar{\lambda}$ the optimal value of (P), and every cluster point of the sequence $\left\{x^k\right\}$ is stationary point. The following approximation is classical.

Definition 5.3.1. Let $c \in (0,1)$ and let $\lambda_k \geq \bar{\lambda}$ and $x^k \in X$. A convex function $\psi_k(x)$ is a c-approximation of $F_k(x)$ if $\psi_k(x) \leq F_k(x)$ for all $x \in X_k$, and if

$$\psi_k(x^{k+1}) \geq \frac{1}{c} F_k(x^{k+1}) \tag{5.3.3}$$

where x^{k+1} is the solution of problem (AP_{α_k}).

Observe that if $\psi_k(.)$ is a c-approximation of $F_k(.)$ at x^k, then at x^{k+1}, we can write

$$\frac{1}{c} F_k(x^{k+1}) \leq \psi_k(x^{k+1}) \leq F_k(x^{k+1}). \tag{5.3.4}$$

In particular, since $c \in (0,1)$, we have

$$\psi_k(x^{k+1}) \leq F_k(x^{k+1}) \leq 0. \tag{5.3.5}$$

Now we are ready to describe our algorithm.

Algorithm 4 General approximating algorithm.

0. Choose $x^0 \in X$, $\alpha_0 > 0$, $c \in (0,1)$ and set $\lambda_0 = \lambda(x^0)$.
1. At step k, we have x^k, α_k, λ_k. Then, construct a c-approximation of $F_k(x)$ and find $x^{k+1} \in X_k$ the unique solution of problem (AP_{α_k}).
2. Set $\lambda_{k+1} = \lambda(x^{k+1})$, choose α_{k+1}, set $k \leftarrow k+1$, and go back to 1.

5.3.1 Convergence of Algorithm 4

In order to study the convergence of the sequence λ_k, we introduce the following notations. Recall that we used $f_i = f_i^1 - f_i^2$ and $g_i = g_i^1 - g_i^2$ for all $i \in I$. We will use and assume that $\Delta = \sup_{x \in X} \max_{i \in I} g_i(x) < \infty$.

For $x \in X$, and λ, we define the set:

$$I(x) = \left\{ i \mid \frac{f_i(x)}{g_i(x)} = \lambda(x) \right\}. \tag{5.3.6}$$

Proposition 5.3.1. Assume $c \in (0,1)$. Then the following results hold:
1. the sequence $\{\lambda_k\}$ is nonincreasing and converges to some $\hat{\lambda} \geq \bar{\lambda}$,
2. if $\bar{\lambda} > -\infty$ and $\omega_i^k \geq \underline{\omega} > 0$ for all k and $i \in I$, then $\psi_k(x^{k+1}) \longrightarrow 0$, $F_k(x^{k+1}) \longrightarrow 0$, and $\frac{1}{2\alpha_k}\|x^{k+1} - x^k\|^2 \longrightarrow 0$ as k tends to ∞.

Proof. 1. From (5.3.5) and the definition of $F_k(.)$, we have for all $i \in I$ that

$$\begin{aligned} 0 \geq F_k(x^{k+1}) &\geq \left(1/\omega_i^k\right)\left[f_{i,k}(x^{k+1}) - \lambda_k g_{i,k}(x^{k+1})\right] \\ &= \left(1/\omega_i^k\right)\left[f_i^1(x^{k+1}) - \left(f_i^2(x^k) + \langle x_{i,k}^2, x^{k+1} - x^k\rangle\right) - \lambda_k g_{i,k}(x^{k+1})\right] \\ &\geq \left(1/\omega_i^k\right)\left[f_i^1(x^{k+1}) - f_i^2(x^{k+1}) - \lambda_k g_{i,k}(x^{k+1})\right] \end{aligned} \tag{5.3.7}$$

where the last inequality follows from the subgradient inequality $f_i^2(x^{k+1}) \geq f_i^2(x^k) + \langle x_{i,k}^2, x^{k+1} - x^k\rangle$.
If $\lambda_k < 0$, we have $g_{i,k}(x^{k+1}) = g_i^1(x^{k+1}) - \left[g_i^2(x^k) + \langle y_{i,k}^2, x^{k+1} - x^k\rangle\right]$ and the subgradient inequality $g_i^2(x^{k+1}) \geq g_i^2(x^k) + \langle y_{i,k}^2, x^{k+1} - x^k\rangle$ gives

$$\begin{aligned} 0 \geq F_k(x^{k+1}) &\geq \left(1/\omega_i^k\right)\left[f_i^1(x^{k+1}) - f_i^2(x^{k+1}) - \lambda_k g_i^1(x^{k+1}) + \lambda_k g_i^2(x^{k+1})\right] \\ &= \left(1/\omega_i^k\right)\left(f_i(x^{k+1}) - \lambda_k g_i(x^{k+1})\right) \\ &\geq \left(g_i(x^{k+1})/\omega_i^k\right)\left(\frac{f_i(x^{k+1})}{g_i(x^{k+1})} - \lambda_k\right) \end{aligned}$$

For $i \in I(x^{k+1})$ we get

$$0 \geq F_k(x^{k+1}) \geq \left(g_i(x^{k+1})/\omega_i^k\right)(\lambda_{k+1} - \lambda_k) \tag{5.3.8}$$

If $\lambda_k \geq 0$, we have $g_{i,k}(x^{k+1}) = -g_i^2(x^{k+1}) + \left[g_i^1(x^k) + \langle y_{i,k}^1, x^{k+1} - x^k\rangle\right]$ and the subgradient inequality $g_i^1(x^{k+1}) \geq g_i^1(x^k) + \langle y_{i,k}^1, x^{k+1} - x^k\rangle$ gives

$$\begin{aligned} 0 \geq F_k(x^{k+1}) &\geq \left(1/\omega_i^k\right)\left[f_i^1(x^{k+1}) - f_i^2(x^{k+1}) - \lambda_k g_i^1(x^{k+1}) + \lambda_k g_i^2(x^{k+1})\right] \\ &= \left(1/\omega_i^k\right)\left(f_i(x^{k+1}) - \lambda_k g_i(x^{k+1})\right) \\ &\geq \left(g_i(x^{k+1})/\omega_i^k\right)\left(\frac{f_i(x^{k+1})}{g_i(x^{k+1})} - \lambda_k\right) \end{aligned}$$

For $i \in I(x^{k+1})$ we get

$$0 \geq F_k(x^{k+1}) \geq \left(g_i(x^{k+1})/\omega_i^k\right)(\lambda_{k+1} - \lambda_k) \tag{5.3.9}$$

Since $g_i(x^{k+1}) > 0$, by (5.3.8) and (5.3.9) it follows that $\lambda_{k+1} \leq \lambda_k$. So $\lambda_k \longrightarrow \hat{\lambda} \geq \bar{\lambda}$ because $\lambda_k \geq \bar{\lambda}$ for all k.
2. If $\bar{\lambda} > -\infty$, then $\hat{\lambda} > -\infty$ and $\lambda_{k+1} - \lambda_k \longrightarrow 0$. Since for all $i \in I$, all $k \in \mathbb{N}$, and $x \in X$ $\left(g_i^1(x) - g_i^2(x)\right) \leq \Delta$ and $\omega_i^k \geq \underline{\omega} > 0$, and since $\lambda_{k+1} - \lambda_k \leq 0$, it follows from (5.3.8) and (5.3.9) that

$$\frac{\Delta}{\underline{\omega}}(\lambda_{k+1} - \lambda_k) \leq F_k(x^{k+1}) \leq 0 \tag{5.3.10}$$

and from (5.3.4) we have

$$\frac{\Delta}{c\underline{\omega}}(\lambda_{k+1} - \lambda_k) \leq \psi_k(x^{k+1}) \leq 0. \tag{5.3.11}$$

So, from (5.3.10) and (5.3.11) it follows that

$$\psi_k(x^{k+1}) \longrightarrow 0 \text{ and } F_k(x^{k+1}) \longrightarrow 0 \quad \text{when} \quad k \longrightarrow \infty. \tag{5.3.12}$$

On the other hand, from the definition of x^{k+1}, we have

$$\psi_k(x) + \frac{1}{2\alpha_k}\|x - x^k\|^2 \geq \psi_k(x^{k+1}) + \frac{1}{2\alpha_k}\|x^{k+1} - x^k\|^2 \quad \text{for all } x \in X_k. \tag{5.3.13}$$

Using Definition 5.3.1 and letting $x = x^k$ in the last inequality, we get

$$F_k(x^k) \geq \frac{1}{c}F_k(x^{k+1}) + \frac{1}{2\alpha_k}\|x^{k+1} - x^k\|^2 \geq \frac{1}{c}F_k(x^{k+1}) \quad \text{for all } x \in X_k, \tag{5.3.14}$$

Since $F_k(x^k) = 0$, then passing to the limit in (5.3.14) when k tends to ∞, we deduce the desired result. □

Now we are going to state a proposition that we will use to the convergence of the sequence $\{x^k\}$.
Proposition 5.3.2. Assume $c \in (0,1)$ and $\alpha_k > 0$, for all $k \in \mathbb{N}$, there exists $\eta^k \in \Sigma$ such that

$$\eta_0^k\left(F_k(x) - \frac{1}{c}F_k(x^{k+1}) + \frac{1}{2\alpha_k}\|x - x^k\|^2\right) + \sum_{j \in J}\eta_j^k h_{j,k}(x) \geq 0 \quad \text{for all } x \in C,$$

where $\Sigma = \left\{(\eta_j)_{j \in J \cup \{0\}} \geq 0 \mid \eta_0 + \sum_{j \in J}\eta_j = 1\right\}$. Moreover,

$$0 \geq \sum_{j \in J}\eta_j^k h_{j,k}(x^{k+1}) \geq -\eta_0^k\left(\frac{c-1}{c}F_k(x^{k+1}) + \frac{1}{2\alpha_k}\|x^{k+1} - x^k\|^2\right)$$

$$\text{and } 0 \geq \sum_{j\in J} \eta_j^k h_j(x^k) \geq \frac{\eta_0^k}{c} F_k(x^{k+1}).$$

So, the sequences $\left\{\sum_{j\in J} \eta_j^k h_{j,k}(x^{k+1})\right\}$ and $\left\{\sum_{j\in J} \eta_j^k h_j(x^k)\right\}$ converge to 0 as k tends to ∞.

Proof. Recall that from the definition of x^{k+1} we have

$$\psi_k(x) + \frac{1}{2\alpha_k}\|x - x^k\|^2 \geq \psi_k(x^{k+1}) + \frac{1}{2\alpha_k}\|x^{k+1} - x^k\|^2 \quad \text{for all } x \in X_k. \tag{5.3.15}$$

From the definition of c-approximation, we obtain

$$F_k(x) + \frac{1}{2\alpha_k}\|x - x^k\|^2 \geq \frac{1}{c} F_k(x^{k+1}) \quad \text{for all } x \in X_k, \tag{5.3.16}$$

and that $X_k := \{x \in C \mid h_{j,k}(x) \leq 0, \forall j \in J\}$. Define the function

$$h^k(x) = \max_{j\in J} h_{j,k}(x).$$

Obviously, $h^k(x) \leq 0$ if and only if $h_{j,k}(x) \leq 0$ for all $j \in J$. Define also the function

$$\mathcal{F}_k(x) = \max\left[F_k(x) + \frac{1}{2\alpha_k}\|x - x^k\|^2 - \frac{1}{c}F_k(x^{k+1}), h^k(x)\right].$$

It is straightforward to show that $\mathcal{F}_k(x) \geq 0$ for all $x \in C$. Indeed, if $x \in X_k$, i.e. $x \in C$ and $h^k(x) \leq 0$, then (5.3.16) entails that $\mathcal{F}_k(x) \geq 0$, and if $x \in C$ but $x \notin X_k$, i.e. $h_k(x) > 0$ then we have $\mathcal{F}_k(x) > 0$. On the other hand, the function $\mathcal{F}_k$ may be expressed by

$$\mathcal{F}_k(x) = \max_{\eta\in\Sigma}\left[\eta_0\left(F_k(x) + \frac{1}{2\alpha_k}\|x - x^k\|^2 - \frac{1}{c}F_k(x^{k+1})\right) + \sum_{j\in J}\eta_j h_{j,k}(x)\right],$$

where Σ is as defined in the proposition. Now, the function defined on $C \times \Sigma$ by

$$(x, \eta) \longmapsto \eta_0\left(F_k(x) + \frac{1}{2\alpha_k}\|x - x^k\|^2 - \frac{1}{c}F_k(x^{k+1})\right) + \sum_{j\in J}\eta_j h_{j,k}(x)$$

is convex with respect to $x \in C$ and linear with respect to $\eta \in \Sigma$, with C being convex and Σ being convex and compact. Thus, by the minimax theory, see e.g. [45, Theorem 2] or [117, Corollary 3.3], we have

$$\begin{aligned}\inf_{x\in C} \mathcal{F}_k(x) &= \inf_{x\in C}\max_{\eta\in\Sigma}\left[\eta_0\left(F_k(x)+\frac{1}{2\alpha_k}\|x-x^k\|^2-\frac{1}{c}F_k(x^{k+1})\right)+\sum_{j\in J}\eta_j h_{j,k}(x)\right]\\ &= \max_{\eta\in\Sigma}\inf_{x\in C}\left[\eta_0\left(F_k(x)+\frac{1}{2\alpha_k}\|x-x^k\|^2-\frac{1}{c}F_k(x^{k+1})\right)+\sum_{j\in J}\eta_j h_{j,k}(x)\right].\end{aligned} \tag{5.3.17}$$

Therefore, for all $k \in \mathbb{N}$, there exist $\eta^k \in \Sigma$ such that

$$\eta_0^k\left(F_k(x)+\frac{1}{2\alpha_k}\|x-x^k\|^2-\frac{1}{c}F_k(x^{k+1})\right)+\sum_{j\in J}\eta_j^k h_{j,k}(x)\geq 0 \quad \text{for all } x\in C.$$

Indeed, the function

$$\Sigma \ni \eta \longmapsto \inf_{x\in C}\left[\eta_0\left(F_k(x)+\frac{1}{2\alpha_k}\|x-x^k\|^2-\frac{1}{c}F_k(x^{k+1})\right)+\sum_{j\in J}\eta_j h_{j,k}(x)\right]$$

is upper semicontinuous on Σ, since it is the pointwise infimum of a family of linear functions, a fortiori continuous, and then achieves its maximum on the compact set Σ. Therefore, for all $k \in \mathbb{N}$, there exist $\eta^k \in \Sigma$ such that

$$\begin{aligned}&\max_{\eta\in\Sigma}\inf_{x\in C}\left[\eta_0\left(F_k(x)+\frac{1}{2\alpha_k}\|x-x^k\|^2-\frac{1}{c}F_k(x^{k+1})\right)+\sum_{j\in J}\eta_j h_{j,k}(x)\right]\\ &=\inf_{x\in C}\left[\eta_0^k\left(F_k(x)+\frac{1}{2\alpha_k}\|x-x^k\|^2-\frac{1}{c}F_k(x^{k+1})\right)+\sum_{j\in J}\eta_j^k h_{j,k}(x)\right].\end{aligned}$$

Since we showed that $\mathcal{F}_k(x) \geq 0$ for all $x \in C$, then from (5.3.17) we obtain

$$\eta_0^k\left(F_k(x)+\frac{1}{2\alpha_k}\|x-x^k\|^2-\frac{1}{c}F_k(x^{k+1})\right)+\sum_{j\in J}\eta_j^k h_{j,k}(x)\geq 0 \quad \text{for all } x\in C. \tag{5.3.18}$$

The rest follows by letting, once $x = x^{k+1}$ and once $x = x^k$, in the previous inequality. □

Theorem 5.3.1. Assume $c \in (0,1)$ and $\alpha_k > 0$, for all $k \in \mathbb{N}$ and that the sequence $\{x^k\}$ is bounded. Let $\hat{x}$ be a cluster point of $\{x^k\}$ and $\hat{\lambda} = \lambda(\hat{x})$. Then for all $i \in I$ and $j \in J$, there exists $\hat{\omega}_i$ a cluster point of $\{\omega_i^k\}$, there exist $\hat{\mu}_i, \hat{\nu}_j \geq 0$, with $\sum_{i\in I} \hat{\mu}_i + \sum_{j\in J} \hat{\nu}_j = 1$ such that

$$0 \in \sum_{i\in I} \frac{\hat{\mu}_i}{\hat{\omega}_i} \left[\partial f_i^1(\hat{x}) - \partial f_i^2(\hat{x}) - \hat{\lambda}\left(\partial g_i^1(\hat{x}) - \partial g_i^2(\hat{x})\right)\right] + \sum_{j\in J} \hat{\nu}_j \left[\partial h_j^1(\hat{x}) - \partial h_j^2(\hat{x})\right] + N_C(\hat{x}),$$

with $\hat{\mu}_i[f_i(\hat{x}) - \hat{\lambda} g_i(\hat{x})] = 0$ and $\hat{\nu}_j h_j(\hat{x}) = 0$, for all $i \in I$ and $j \in J$.

Proof. Let $\hat{x}$ be a cluster point of the sequence $\{x^k\}$, and let K be an infinite subset of $\mathbb{N}$ such that the subsequence $\{x^k\}_{k\in K}$ converges to $\hat{x}$. Since $x^k \in X_k \subset X$ and X is closed, we have $\hat{x} \in X$. For each $k \in \mathbb{N}$, let $\{\eta^k\} \subset \Sigma$ as stated in Proposition 5.3.2. Let the sequences $\{x_{i,k}^2\}$, $\{y_{i,k}^\ell\}$, for $i \in I$, $\ell = 1,2$, and $\{z_{j,k}^2\}$, for $j \in J$, be as defined in (5.2.8) and (5.2.9). We recall here that these sequences are bounded by the boundedness of $\{x^k\}$, see e.g. [64, Propositions 6.2.2].

Next, for $k \in K$, we consider subsequences of $\{\alpha^k\}$, $\{\eta^k\}$, $\{z_{j,k}^2\}$, for $j \in J$, $\{\omega_i^k\}$, $\{x_{i,k}^2\}$, $\{y_{i,k}^\ell\}$, for $i \in I$, say for $k \in K'$ an infinite subset of $\mathbb{N}$, converging respectively to $\hat{\alpha}$, $\hat{\eta}$, $\hat{z}_j^2$, $\hat{\omega}_i$, $\hat{x}_i^2$, $\hat{y}_i^\ell$, where $\ell = 1$ if $\hat{\lambda} > 0$ and $\ell = 2$ if $\hat{\lambda} < 0$. (Notice that $\lambda_k > 0$ and $\ell = 1$ (resp. $\lambda_k < 0$ and $\ell = 2$) for k large, when $\hat{\lambda} > 0$ (resp. $\hat{\lambda} < 0$)). Therefore, $\hat{z}_j^2 \in \partial h_j^2(\hat{x})$, $\hat{x}_i^2 \in \partial f_i^2(\hat{x})$ and $\hat{y}_i^\ell \in \partial g_i^\ell(\hat{x})$, see e.g. [64, Propositions 6.2.1]. With these elements we define the function

$$F_{\hat{x}}(x) = \max_{i\in I} \left[\frac{f_{i,\hat{x}}(x) - \hat{\lambda} g_{i,\hat{x}}(x)}{\hat{\omega}_i}\right],$$

where the functions $f_{i,\hat{x}}$, $g_{i,\hat{x}}$ and $h_{j,\hat{x}}$ are defined as in Eqs. (5.2.1) to (5.2.3) with $x_i^2(\hat{x}) = \hat{x}_i^2$, $y_i^\ell(\hat{x}) = \hat{y}_i^\ell$, for $l = 1,2$, $i \in I$, and $z_j^2(\hat{x}) = \hat{z}_j^2$, $j \in J$. In the case $\hat{\lambda} = 0$, we ignore the term $\hat{\lambda} g_{i,\hat{x}}(x)$ in the definition of $F_{\hat{x}}(x)$.

By invoking Proposition 5.3.1 and passing to the limit in (5.3.18), as k tends to ∞, $k \in K'$, we arrive to

$$\hat{\eta}_0 \left(F_{\hat{x}}(x) + \frac{1}{2\hat{\alpha}}\|x - \hat{x}\|^2\right) + \sum_{j\in J} \hat{\eta}_j h_{j,\hat{x}}(x) \geq 0 \quad \text{for all } x \in C.$$

Therefore, for all $x \in X_{\hat{x}}$ we have $\hat{\eta}_0\left(F_{\hat{x}}(x) + \frac{1}{2\hat{\alpha}}\|x - \hat{x}\|^2\right) \geq 0$. So, if $\hat{\eta}_0 \neq 0$ then since $F_{\hat{x}}(\hat{x}) = 0$, we deduce that $\hat{x}$ globally minimizes the convex function $F_{\hat{x}}$ over $X_{\hat{x}}$. Then it suffices to use Theorem 5.2.1 to conclude. Now if $\hat{\eta}_0 = 0$, we get

$$\sum_{j\in J} \hat{\eta}_j h_{j,\hat{x}}(x) \geq 0 \quad \text{for all } x \in C. \tag{5.3.19}$$

Since $h_{j,\hat{x}}(\hat{x}) = h_j(\hat{x}) \leq 0$ we obtain $\hat{\eta}_j h_j(\hat{x}) = 0$ for all $j \in J$. Therefore, $\hat{x}$ minimizes, on C, the function $x \mapsto \sum_{j\in J} \hat{\eta}_j h_{j,\hat{x}}(x)$. It follows that

$$0 \in \sum_{j\in J} \hat{\eta}_j \partial h_{j,\hat{x}}(\hat{x}) + N_C(\hat{x}) \subset \sum_{j\in J} \hat{\eta}_j \left[\partial h_j^1(\hat{x}) - \partial h_j^2(\hat{x})\right] + N_C(\hat{x}).$$

The desired result is fulfilled with $\hat{\mu}_i = 0$, for all $i \in I$, and $\hat{\nu}_j = \hat{\eta}_j$, for all $j \in J$. □

5.3.2 Construction of the c-approximations.

Before describing the method of constructing c-approximations functions, we begin by defining, as in [122], strong c-approximations.

Definition 5.3.2. Letting $c \in]0,1[$ be a given parameter, a convex function $\psi_k(\cdot)$ is a strong c-approximation of $F_k(\cdot)$ at $x^k \in X$ if $\psi_k(x) \leq F_k(x)$ for all $x \in X_k$ and if

$$F_k(x^{k+1}) - \psi_k(x^{k+1}) \leq \frac{(1-c)}{\alpha_k}\|x^{k+1} - x^k\|^2,$$

where x^{k+1} is the solution of problem $\left(AP_{\alpha_k}\right)$.

Remark 5.3.1. A strong c-approximation of $F_k(\cdot)$ at x^k is also c-approximation of $F_k(\cdot)$ at x^k.

Instead of directly solving the problem $\left(P_{\alpha_k}\right)$, a set of approximating and easier problems, indexed by $l = 1, \cdots, l(k)$,

$$\min_{x\in X_k} \left\{ \psi_k^l(x) + \frac{1}{2\alpha_k}\|x - x^k\|^2 \right\}, \tag{5.3.20}$$

will be solved until an approximate solution $x^k_{l(k)} \in X_k$ of the subproblem $\left(P_{\alpha_k}\right)$ is reached. Then iteration $k+1$ will be performed by approximately solving problem $\left(P_{\alpha_{k+1}}\right)$ with $x^{k+1} = x^k_{l(k)}$.

To obtain the approximate solution $x^k_{l(k)}$ of $\left(P_{\alpha_k}\right)$ one may construct, at each iteration l, an approximation $\psi_k^l(\cdot)$ of $F_k(\cdot)$ and solve the approximating

problem (5.3.20) to obtain the solution $x^k_{l(k)}$. The procedure then stops with l such that $\psi^l_k(\cdot)$ is a strong c-approximation of $F_k(\cdot)$ at x^k, and we set $l(k) = l$.
For $l = 1, 2, \cdots$, we define the affine function $\mathcal{L}^l_k$ on X_k, by

$$\mathcal{L}^l_k(x) = \psi^l_k(x^k_l) + \alpha_k^{-1}\langle x^k - x^k_l, x - x^k_l\rangle \quad \forall x \in X_k, \tag{5.3.21}$$

we recall that

$$\alpha_k^{-1}(x^k - x^k_l) \in \left[\psi^l_k(\cdot) + Ind_{X_k}\right](x^k_l),$$

where Ind_{X_k} is the indicator function associated with X_k. By definition of $\mathcal{L}^l_k(\cdot)$ we see that, for all $x \in X_k$,

$$\mathcal{L}^l_k(x^k_l) = \psi^l_k(x^k_l) \text{ and } \psi^l_k(x) \geq \mathcal{L}^l_k(x). \tag{5.3.22}$$

Next, we discuss the construction of c-approximation functions of $F_k(\cdot)$. To guarantee thier existence, these functions must satisfy the following classical properties:
(C1) $\psi^l_k(x) \leq F_k(x)$ for $l = 1, 2, \cdots, l(k)$ and all $x \in X_k$,
(C2) $\psi^{l+1}_k(x) \geq \mathcal{L}^l_k(x)$ for $l = 1, 2, \cdots, l(k)$ and all $x \in X_k$,
(C3) $\psi^{l+1}_k(x) \geq F_k(x^k_l) + \langle s^l_k, x - x^k_l\rangle$ on X_k for $l = 1, 2, \cdots, l(k)$ and all $x \in X_k$, where s^l_k is any subgradient of $F_k(.)$ at x^k_l.

In the following we give some possible choices of $\psi^l_k(\cdot)$

Example 5.3.1. Consider the piecewise-affine model, defined for all $k \in \mathbb{N}$ and $l \in \mathbb{N}$, by

$$\psi^{l+1}_k(x) = \max_{1\leq q\leq l(k)} \left\{F_k(x^k_q) + \langle s^q_k, x - x^k_q\rangle\right\} \tag{5.3.23}$$

for all $x \in X_k$ where $x^k_0 = x^k$.
Example 5.3.2. For all $k \in \mathbb{N}$ and $l \in \mathbb{N}$, we can choose, for all $x \in X_k$,

$$\psi^{l+1}_k(x) = \max\left\{\mathcal{L}^l_k(x), F_k(x^k_q) + \langle s^l_k, x - x^k_l\rangle\right\} \tag{5.3.24}$$

where $x^k_0 = x^k$.
Example 5.3.3. For all $k \in \mathbb{N}$ and $l \in \mathbb{N}$, and $x \in X_k$, let

$$\psi^{l+1}_k(x) = \max\left\{\mathcal{L}^l_k(x), \max_{1\leq q\leq l(k)}\left\{F_k(x^k_q) + \langle s^q_k, x - x^k_q\rangle\right\}\right\} \tag{5.3.25}$$

where $x^k_0 = x^k$.

Below we describe a procedure to construct a strong c-approximation, at an iteration k, of the function $F_k(\cdot)$ at x^k. this also permits the construction of a c-approximation, following Remark 5.3.1.

Algorithm 5 Construction of the c-approximation

0. Let $c \in]0,1[$ be a given parameter, $x^k \in X$, and $l = 1$.
1. Choose a piecewise linear convex function $\psi_k^l(\cdot)$ satisfying (C1)-(C3). Determine $x_l^k \in X_k$ as the minimizer of

$$\min_{x \in X_k} \left\{ \psi_k^l(x) + (1/2\alpha_k)\|x - x^k\|^2 \right\},$$

to get x_l^k.
2. If

$$F_k(x_l^k) - \psi_k^l(x_l^k) \le \frac{(1-c)}{\alpha_k} \|x_l^k - x^k\|^2,$$

then stop, set $l(k) = l$, and $\psi_k^{l(k)}(\cdot)$ is a strong c-approximation of $F_k(\cdot)$ at x^k. Otherwise, increase l by 1 and go back to step 1.

Now present results to show the finite terminition of Algorithm 5.

Theorem 5.3.2. Suppose that models $\psi_k^l(\cdot)$ satisfy conditions (C1)-(C3). Let

$$x_l^k := \operatorname*{argmin}_{x \in X_k} \left\{ \psi_k^l(x) + (1/2\alpha_k)\|x - x^k\|^2 \right\}$$

and

$$\bar{x}^k := \operatorname*{argmin}_{x \in X_k} \left\{ F_k(x) + (1/2\alpha_k)\|x - x^k\|^2 \right\}.$$

1. If $l(k) = \infty$ and for all $k \in \mathbb{N}$. Then $F_k(x_l^k) - \psi_k^l(x_l^k) \longrightarrow 0$ and $x_l^k \longrightarrow \bar{x}^k$ when $l \longrightarrow \infty$.
2. If $x^k \neq \bar{x}^k$, then Algorithm 5 stops after finitely many iterations $l(k)$ with $\psi_k^{l(k)}(\cdot)$ a strong c-approximation of $F_k(\cdot)$ at x^k, and with, $x^{k+1} = x_{l(k)}^k$.
3. If $x^k = \bar{x}^k$, then x^k verify optimality conditions in the Theorem 5.3.1.

Proof. 1. For the seek simplicity, we define the functions $\tilde{\psi}_k^l(\cdot)$, $\tilde{\mathcal{L}}_k^l(\cdot)$ and $\tilde{F}_k(\cdot)$ for all $x \in X_k$, $k \in \mathbb{N}$, by

$$\tilde{\psi}_k^l(x) = \psi_k^l(x) + (1/2\alpha_k)\|x - x^k\|^2, \tag{5.3.26}$$

$$\tilde{\mathcal{L}}_k^l(x) = \mathcal{L}_k^l(x) + (1/2\alpha_k)\|x - x^k\|^2 \tag{5.3.27}$$

and

$$\tilde{F}_k(x) = F_k(x) + (1/2\alpha_k)\|x - x^k\|^2. \tag{5.3.28}$$

Recall that by definition we have, for all $x \in X_k$

$$\mathcal{L}_k^l(x) = \psi_k^l(x_l^k) + \alpha_k^{-1}\langle x^k - x_l^k, x - x_l^k\rangle \quad \forall x \in X_k. \tag{5.3.29}$$

Therefore

$$\begin{aligned}\mathcal{L}_k^l(x) + (1/2\alpha_k)\|x - x^k\|^2 = \psi_k^l(x_l^k) + (1/2\alpha_k)\|x - x^k\|^2 \\ + \alpha_k^{-1}\langle x^k - x_l^k, x - x_l^k\rangle \quad \forall x \in X_k.\end{aligned} \tag{5.3.30}$$

Then for all $x \in X_k$

$$\tilde{\mathcal{L}}_k^l(x) = \psi_k^l(x_l^k) + (1/2\alpha_k)\|x - x^k\|^2 + \alpha_k^{-1}\langle x^k - x_l^k, x - x_l^k\rangle. \tag{5.3.31}$$

On the other hand, for all $x \in X_k$, we have

$$\begin{aligned}\tilde{\mathcal{L}}_k^l(x) &= \mathcal{L}_k^l(x) + (1/2\alpha_k)\|x - x^k\|^2 \\ &= \mathcal{L}_k^l(x) + (1/2\alpha_k)\|x - x_l^k + x_l^k - x^k\|^2 \\ &= \mathcal{L}_k^l(x) - \alpha_k^{-1}\langle x^k - x_l^k, x - x_l^k\rangle + (1/2\alpha_k)\|x - x_l^k\|^2 + (1/2\alpha_k)\|x_l^k - x^k\|^2 \\ &= \psi_k^l(x_l^k) + (1/2\alpha_k)\|x - x_l^k\|^2 + (1/2\alpha_k)\|x_l^k - x^k\|^2 \\ &= \tilde{\psi}_k^l(x_l^k) + (1/2\alpha_k)\|x - x_l^k\|^2\end{aligned}$$

where the penultimate equality follows from the definition of $\mathcal{L}_k^l(\cdot)$ and the last equality from the definition of $\tilde{\psi}_k^l(\cdot)$.
Now by the relation (5.3.31), we have

$$\tilde{\mathcal{L}}_k^l(x_l^k) = \tilde{\psi}_k^l(x_l^k), \tag{5.3.32}$$

so that

$$\tilde{\mathcal{L}}_k^l(x) = \tilde{\mathcal{L}}_k^l(x_l^k) + (1/2\alpha_k)\|x - x^k\|^2. \tag{5.3.33}$$

From the condition (C1) and the definitions (5.3.26)-(5.3.28) we get

$$\begin{aligned}F_k(x^k) &\geq \psi_k^{l+1}(x^k) = \tilde{\psi}_k^{l+1}(x^k) \\ &\geq \tilde{\psi}_k^{l+1}(x_{l+1}^k) = \tilde{\mathcal{L}}_k^{l+1}(x_{l+1}^k) \\ &\geq \tilde{\mathcal{L}}_k^l(x_{l+1}^k) = \tilde{\mathcal{L}}_k^l(x_l^k) + (1/2\alpha_k)\|x_{l+1}^k - x_l^k\|^2 \\ &\geq \tilde{\mathcal{L}}_k^l(x_l^k)\end{aligned} \tag{5.3.34}$$

where the second inequality folows from the definition of x_{l+1}^k, the second equality from (5.3.32) with $l+1$, the third inequality follows from (C2)

and the equality from (5.3.33).
Because $\left\{x^k\right\}$ is fixed, the relations two and last of (5.3.34) show that $\left\{\tilde{\mathcal{L}}_k^l(x_l^k)\right\}_l$ is convergent, and (5.3.34) imply that

$$x_{l+1}^k - x_l^k \longrightarrow 0 \text{ when } l \longrightarrow +\infty. \tag{5.3.35}$$

Furthermore, for all $x \in X_k$, condition (C1) and relations (5.3.22) and (5.3.33) entail

$$\begin{aligned} F_k(x^k) \geq \tilde{\psi}_k^l(x) &\geq \tilde{\mathcal{L}}_k^l(x) \\ &= \tilde{\mathcal{L}}_k^l(x_l^k) + (1/2\alpha_k)\|x - x_l^k\|^2. \end{aligned}$$

This

$$F_k(x^k) - \tilde{\mathcal{L}}_k^l(x_l^k) \geq (1/2\alpha_k)\|x - x_l^k\|^2.$$

Since by a previous conclusion, the sequence $\left\{\tilde{\mathcal{L}}_k^l(x_l^k)\right\}_l$ is convergent, then $\left\{x_l^k\right\}$ is bounded .
Now from (C1) and (C3) we get

$$\begin{aligned} F_k(x_{l+1}^k) - F_k(x_l^k) &\geq \psi_k^{l+1}(x_{l+1}^k) - F_k(x_l^k) \\ &\geq \langle s_k^l, x - x_l^k \rangle \end{aligned} \tag{5.3.36}$$

and from the local Lipschitz property of a convex funtion $F_k(\cdot)$ [106, theorem 10.4], there exists $L_k > 0$ such that

$$L_k\|x_{l+1}^k - x_l^k\| \geq F_k(x_{l+1}^k) - F_k(x_l^k).$$

Using the previous inequality and (5.3.36), we obtain

$$L_k\|x_{l+1}^k - x_l^k\| \geq \psi_k^{l+1}(x_{l+1}^k) - F_k(x_l^k) \geq \langle s_k^l, x - x_l^k \rangle. \tag{5.3.37}$$

Since $\left\{x_l^k\right\}_l$ and $\left\{s_k^l\right\}_l$ are bounded ([64, Proposition 4.1.2]) thus, the last inequality with (5.3.33), as l goes to ∞, give

$$\psi_k^{l+1}(x_{l+1}^k) - F_k(x_l^k) \longrightarrow 0 \text{ when } l \longrightarrow +\infty. \tag{5.3.38}$$

On the other hand, we have

$$\begin{aligned} \tilde{\mathcal{L}}_k^l(x_l^k) - \tilde{\mathcal{L}}_k^{l+1}(x_{l+1}^k) &= \psi_k^l(x_l^k) + (1/2\alpha_k)\|x_l^k - x^k\|^2 - \Big(\psi_k^{l+1}(x_{l+1}^k) \\ &\quad + 1/2\alpha_k\|x_{l+1}^k - x^k\|^2\Big) \\ &= \psi_k^l(x_l^k) - \psi_k^{l+1}(x_{l+1}^k) - (1/2\alpha_k)\Big(\|x_{l+1}^k - x^k\|^2 - \|x_l^k - x^k\|^2\Big). \end{aligned} \tag{5.3.39}$$

Notice that we have following equality

$$\|x_{l+1}^k - x^k\|^2 = \|x_{l+1}^k - x_l^k\|^2 + \|x_l^k - x^k\|^2 + 2\langle x_{l+1}^k - x_l^k, x_l^k - x^k\rangle$$

and then

$$\|x_{l+1}^k - x^k\|^2 - \|x_l^k - x^k\|^2 = \|x_{l+1}^k - x_l^k\|^2 + 2\langle x_{l+1}^k - x_l^k, x_l^k - x^k\rangle.$$

We introduce the last equality in (5.3.39) to obtain

$$\begin{aligned} \tilde{\mathcal{L}}_k^l(x_l^k) - \tilde{\mathcal{L}}_k^{l+1}(x_{l+1}^k) = \psi_k^l(x_l^k) - \psi_k^{l+1}(x_{l+1}^k) - (1/2\alpha_k)\|x_{l+1}^k - x_l^k\|^2 \\ - (1/\alpha_k)\langle x_{l+1}^k - x_l^k, x_l^k - x^k\rangle. \end{aligned} \tag{5.3.40}$$

By passing to the limit, as $l \longrightarrow \infty$ in the equality (5.3.40), we deduce, by using (5.3.35) and the fact that $\left\{\tilde{\mathcal{L}}_k^l(x_l^k)\right\}_l$ converges, that

$$\psi_k^l(x_l^k) - \psi_k^{l+1}(x_{l+1}^k) \longrightarrow 0 \text{ when } l \longrightarrow +\infty.$$

By writing

$$F_k(x_l^k) - \psi_k^{l+1}(x_{l+1}^k) = F_k(x_l^k) - \psi_k^l(x_l^k) + \psi_k^l(x_l^k) - \psi_k^{l+1}(x_{l+1}^k)$$

and by using (5.3.38) we deduce that

$$F_k(x_l^k) - \psi_k^l(x_l^k) \longrightarrow 0 \text{ when } l \longrightarrow +\infty. \tag{5.3.41}$$

Since $\alpha_k^{-1}(x^k - x_l^k) \in \left[\psi_k^l(\cdot) + Ind_{X_k}\right](x_l^k)$ then for all $x \in X_k$, and $l = 1, 2, \cdots$, we get

$$\psi_k^l(x) \geq \psi_k^l(x_l^k) + \alpha_k^{-1}\langle x^k - x_l^k, x - x_l^k\rangle. \tag{5.3.42}$$

Let $\hat{x}^k$ be an accumulation point of the bounded sequence $\left\{x_l^k\right\}_l$. By passing to the limit, on a subsequence, in (5.3.42), while keeping in mind the continuity of $F_k(\cdot)$, the limit in (5.3.41) and the inequality $F_k(\cdot) \geq \psi_k^l(\cdot)$, by condition (C1), we obtain for all $x \in X_k$

$$F_k(x) \geq F_k(\hat{x}^k) + \alpha_k^{-1}\langle x^k - \hat{x}^k, x - \hat{x}^k\rangle.$$

This implies that $\alpha_k^{-1}(x^k - \hat{x}^k) \in \left[\psi_k^l(\cdot) + Ind_{X_k}\right](\hat{x}^k)$ and so $\hat{x}^k$ is a solution of P_{α_k}. The unicity of $\bar{x}^k$ implies that $\hat{x}^k = \bar{x}^k$, this concludes that the whole sequence $\left\{x_l^k\right\}_l$ converges to $\bar{x}^k$.

2. If $\bar{x}^k \neq x^k$, by Item 1, $\psi_k^l(x_l^k) - F_k(x_l^k)$ tends to 0 and $\frac{(1-c)}{\alpha_k}\|x_l^k - x^k\|^2$ converges to the positive number $\frac{(1-c)}{\alpha_k}\|\bar{x}^k - x^k\|^2$ when l tends ∞, then there

exists a finite $l(k)$ such that for all $l \geq l(k)$, we have $\psi_k^l(x_l^k) - F_k(x_l^k) < \frac{(1-c)}{\alpha_k}\|x_l^k - x^k\|^2$. Then the strong c-approximation condition holds.
3. Since $\bar{x}^k$ is the unique solution of

$$(P_{\alpha_k}) \qquad \min_{x \in X_k} F_k(x) + \frac{1}{2\alpha_k}\|x - x^k\|^2$$

then $\bar{x}^k = x^k$ implies that $\bar{x}^k$ also solves

$$\min_{x \in X_k} F_k(x)$$

Therefore

$$\begin{aligned}\min_{x \in X_k} F_k(x) &= F_k(\bar{x}^k)\\ &= F_k(x^k)\\ &= 0\end{aligned}$$

where the last equality from Item 1 of Proposition 5.2.1, then x^k verify optimality conditions of the Theorem 5.3.1. □

At this stage, we can describe the complete proximal bundle algorithm by inserting the procedure to construct (strong) c-approximations, detailed in Algorithm 5 in the general scheme described in Algorithm 4.
Now we summarize our proximal bundle algorithm.

Algorithm 6 Proximal bundle algorithm

0. Let $x^0 \in X$, and compute $\lambda_0 = \lambda(x_0)$. Let $k = 0$, and $l = 1$.
1. At step k, we have $\alpha_k > 0$, x_k and λ_k. Choose $\psi_k^l(\cdot)$ a convex piecewise linear function that satisfies (C1)-(C3) and solve the problem

$$\min_{x \in X_k}\left\{\psi_k^l(x) + (1/2\alpha_k)\|x - x^k\|^2\right\},$$

   to obtain the unique solution x_l^k.
2. If

$$\psi_k^l(x_l^k) \geq \frac{1}{c}F_k(x_l^k),$$

   then set $x^{k+1} = x_l^k$, $l(k) = l$, and $x_0^{k+1} = x^{k+1}$. Compute $\lambda_{k+1} = \lambda(x^{k+1})$, increase k by 1 and set $l = 1$.
3. Increase l by 1 and go to step 1.

Chapter 6

Optimality conditions and DC-Dinkelbach-type algorithm for vector fractional programming problems with ratios of difference of convex functions

In this chapter, necessary conditions of KKT type for (weak) Pareto optimality are derived and a DC-Dinkelbach-type algorithm is proposed by first reducing the vector fractional mathematical program with ratios of difference of convex (DC) functions, and DC constraints, to a system of scalar parametric problems and then using convex analysis tools. An application to vector fractional mathematical programming with ratios of convex functions is also given.

6.1 Introduction

Consider a vector fractional mathematical programming problem with ratios of difference of convex functions,

$$(P) \qquad \inf_{x\in X}\left[v(x):=\left(\frac{f_1^1(x)-f_1^2(x)}{g_1^1(x)-g_1^2(x)},\frac{f_2^1(x)-f_2^2(x)}{g_2^1(x)-g_2^2(x)},\cdots,\frac{f_m^1(x)-f_m^2(x)}{g_m^1(x)-g_m^2(x)}\right)\right],$$

where $X = \{x \in C \mid h_j^1(x) - h_j^2(x) \leq 0, j \in J\}$, with $C \subset \mathbb{R}^n$ a nonempty, closed convex set, J a finite index set, and the functions f_i^ℓ, g_i^ℓ, for $i \in I := \{1, 2, \ldots, m\}$, and h_j^ℓ, for $j \in J$, and $\ell = 1, 2$ are defined on $\mathbb{R}^n$ and convex, with $g_i^1 - g_i^2$ positive on X for all $i \in I$.

To simplify notations, we will put for all $i \in I$, $j \in J$ and $x \in \mathbb{R}^n$,

$$f_i(x) = f_i^1(x) - f_i^2(x), \quad g_i(x) = g_i^1(x) - g_i^2(x), \quad h_j(x) = h_j^1(x) - h_j^2(x),$$

and

$$h(x) = \max_{j \in J} h_j(x).$$

With these notations we have, $v(x) := (v_1(x), v_2(x), \ldots, v_m(x))$, with $v_i(x) = f_i(x)/g_i(x)$, and $X = \{x \in C \mid h(x) \leq 0\}$.

Multiobjective programming (also known as vector optimization problems, multi-objective optimization, multi-criteria optimization, multiattribute optimization or Pareto optimization) have received extensive attention from mathematicians. The problem under consideration consists in finding a solution conforming to the following notions. A vector $\bar{x} \in X$ is said to be a Pareto minimum of (P) if there exists no $x \in X$ such that $v(x) \neq v(\bar{x})$ and $v(x) \leq v(\bar{x})$. It is said to be a weak Pareto minimum of (P) if there exists no $x \in X$ such that $v(x) < v(\bar{x})$. The inequalities are to be taken in the sense of component by component.

Many authors have developed the necessary and/or sufficient conditions for Pareto optimality, see [19, 41, 46, 54, 56, 59, 68, 73, 109, 123]. Multiobjective fractional programming (MFP) refers to a multiobjective problem where the objective functions are quotients. Fractional optimization problems arise in many fields of applications such as economics, management applications of goal programming, multi-objective programming involving ratios of functions, data bases, physics, telecommunications and numerical analysis [44, 53, 93, 96, 114, 131, 132].

Multiobjective fractional programming problems have been studied by many authors in recent years. In particular, Bector et al. [16] obtained Fritz John and Karush-Kuhn-Tucker necessary and sufficient optimality conditions for a class of nondifferentiable convex MFP problems and established some duality theorems and saddle-point results for such problems. Liu [85, 88] derived some necessary and sufficient optimality conditions and duality theorems for a class of nonsmooth MFP problems involving pseudoinvex functions or (F,ρ)-convex functions. Kuk et al. [72] established generalized Karush-Kuhn-Tucker necessary and sufficient optimality conditions and derive duality theorems for nonsmooth MFP problems containing V-ρ-invex functions. Liang et al. [86, 87] introduced the concept of (F,α,ρ,d)-convexity and obtained some optimality conditions and duality results for MFP with the (F,α,ρ,d)-convex functions. Xiuhong [28] gave the definitions of the generalized (F,ρ)-convex class about the Clarke subgradient, under the above generalized convexity assumption, the alternative theorem is obtained, and some sufficient and necessary conditions for optimality are also given related to the properly efficient solution for the problems. In [51], Gadhi et al. established sufficient optimality conditions for a weak Pareto minimum for MFP. Gadhi et al. [50] derived necessary and sufficient optimality conditions for a weak Pareto minimum of MFP under reverse convex constraints space. Taa [123] has developed the necessary optimality conditions for a weak Pareto minimum of MFP. In [69] Kim et al. established necessary and sufficient optimality conditions and duality results for weakly efficient solutions of nondifferentiable MFP problems. Recently in [59], Guo et al. gave necessary and sufficient optimality conditions for an ε-weak Pareto minimum and an ε-proper Pareto minimum for MFP. Moreover, when $\varepsilon = 0$, these optimality conditions become the optimality conditions for a weak Pareto minimum and Pareto minimum for the respective MFP. Very recently, the idea of convexificators is used to derive the Karush-Kuhn-Tucker necessary optimality conditions for local weak efficient solutions of MFP [61]. Gadhi et al. [49] used the extremal principle developed by Mordukhovich [94] to get necessary optimality conditions for MFP. See [18,27,52,89,95,100,102,129] for a detailed bibliography. Another important class of such problems is the MFP for which the functions may be expressed as a difference of convex functions (see e.g., [126] for such functions).

In our method, we first reduce a multiobjective fractional programming with ratios of DC functions, with DC constraints to a series of parametric scalar problems, having simpler structures than the original problem. For solving these problems, the use of Dinkelbach-type algorithms conducts to nonconvex subproblems. We resort to DC techniques, see e.g [80, 124], to overestimate the objective function in these subproblems by a convex function, and the constraints set by an inner convex subset of the latter. Doing so we obtain convex parametric subproblems.

The most important difficulty in global optimization, and in particular in DC programming, is how to recognize a global minimum, or even how to recognize local minimum, in contrast with the convex programming where a local minimum is global. The most common necessary optimality condition for the minimization of $f_1 - f_2$, say, over the whole space $\mathbb{R}^n$ is DC criticality, which means that a point x^* is DC critical if $\partial f_1(x^*) \cap \partial f_2(x^*) \neq \emptyset$, or equivalently if $0 \in \partial f_1(x^*) - \partial f_2(x^*)$, where $\partial f_i(x^*)$ stands for the subdifferential of the convex function f_i, $i = 1, 2$, at x^*. It is strongly critical if $\in \partial f_2(x^*) \subset \partial f_1(x^*)$. To get more on optimality conditions and algorithms for DC unconstrained and DC constrained programs, the reader can consult [6,55,62,80,124], and [10,103] for several extensions and applications.

The chapter is organized as follow. In Section 6.2 we introduce a new parametric approach based on convex parametric subprolems, and show that the problem (P) is equivalent to a scalar convex problem. By writing necessary optimality conditions, for the latter, using only convex analysis tools, we obtain in Subsection 6.2.1, necessary optimality conditions for (P). In Subsection 6.2.2, we will describe our vector DC-Dinkelbach-type algorithm and establish its convergence. Later, in Section 6.3, we give an application to vector fractional mathematical programming with ratios of convex functions.

6.2 Parametric approach for (P)

In the scalar minimization of generalized fractionl programs (GFP), Dinkelbach type algorithms [33, 35, 36, 40], replace the resolution of the original program by a series of parametric ones, having simpler structures than the original problem. Using a similar technique to multiobjective setting, we propose to associate to (P), the following parametric problem, with the parameter $\mu = (\mu_1, \mu_2, \dots, \mu_m) \in \mathbb{R}^m$,

$$(P_\mu) \qquad F(\mu) = \inf_{x \in X} \max_{i \in I} \{f_i(x) - \mu_i g_i(x)\}.$$

The following lemma gives a link between (P) and (P_μ) for particular parameters $\mu \in \mathbb{R}^m$.

Lemma 6.2.1. A point $\bar{x} \in X$ is a weak Pareto minimum for (P) if and only if $\bar{x}$ is a global minimum for $(P_{\bar{v}})$, where $\bar{v} = v(\bar{x})$.

Proof. Let $\bar{x} \in X$ be a weak Pareto minimum of (P) and $\bar{v} = v(\bar{x})$. Suppose the contrary, that is, $\bar{x}$ is not a global minimum for $(P_{\bar{v}})$. Then, there exists $y \in X$ such that

$$f_i(y) - \bar{v}_i g_i(y) < \max_{i \in I} \{f_i(\bar{x}) - \bar{v}_i g_i(\bar{x})\} \text{ for all } i \in I.$$

Since $f_i(\bar{x}) - \bar{v}_i g_i(\bar{x}) = 0$ and $g_i(y) > 0$, for all $i \in I$, it follows that

$$\frac{f_i(y)}{g_i(y)} - \frac{f_i(\bar{x})}{g_i(\bar{x})} < 0 \text{ for all } i \in I,$$

i.e. $v(y) < v(\bar{x})$, wich contradicts the fact that $\bar{x}$ is a weak Pareto minimum of (P). So $\bar{x}$ is a global minimum of $P_{\bar{v}}$.

Conversely, we suppose that $\bar{x}$ is a global minimum for $(P_{\bar{v}})$, and $\bar{x}$ is not a weak Pereto minimum of (P). Then there exists $z \in X$ such that $v(z) < v(\bar{x})$, that is,

$$\frac{f_i(z)}{g_i(z)} < \frac{f_i(\bar{x})}{g_i(\bar{x})} =: \bar{v}_i \text{ for all } i \in I.$$

Since $g_i(z) > 0$, for all $i \in I$, it yields that

$$\max_{i\in I} \{f_i(z) - \bar{v}_i g_i(z)\} < 0 = \max_{i\in I} \{f_i(\bar{x}) - \bar{v}_i g_i(\bar{x})\}.$$

This contradicts the fact that $\bar{x}$ is a global minimum of $(P_{\bar{v}})$. □

Lemma 6.2.2. If $\bar{x} \in X$ is a Pareto minimum for (P) then it is a global minimum for $(P_{\bar{v}})$, where $\bar{v} = v(\bar{x})$. Conversely, if $\bar{x}$ is the unique global minimum of $P_{\bar{v}}$, then it is a Pareto minimum for (P).

Proof. The first assertion follows from the previous lemma since a Pareto minimum of (P) is also a weak Pareto minimum of (P).

Conversely, assume that $\bar{x}$ is a global minimum for $(P_{\bar{v}})$, and $\bar{x}$ is not a Pereto minimum of (P). Then there exists $z \in X$ such that

$$\frac{f_i(z)}{g_i(z)} \leq \frac{f_i(\bar{x})}{g_i(\bar{x})} \quad \text{for all } i \in I \text{ and } \frac{f_j(z)}{g_j(z)} < \frac{f_j(\bar{x})}{g_j(\bar{x})} \quad \text{for some } j \in I. \tag{6.2.1}$$

Note that since $\bar{v}_i = f_i(\bar{x})/g_i(\bar{x})$, and $g_i(z) > 0$, for all $i \in I$, we have

$$\max_{i\in I} \{f_i(z) - \bar{v}_i g_i(z)\} \leq 0. \tag{6.2.2}$$

On the other hand, the fact that $\bar{x}$ is a global minimum for $(P_{\bar{v}})$ entails that

$$\max_{i\in I} \{f_i(x) - \bar{v}_i g_i(x)\} \geq \max_{i\in I} \{f_i(\bar{x}) - \bar{v}_i g_i(\bar{x})\} = 0 \quad \text{for all } x \in X.$$

With $x = z$ in that last inequality and (6.2.2) we get

$$\max_{i\in I} \{f_i(z) - \bar{v}_i g_i(z)\} = 0.$$

This means that z is also a global minimum of $(P_{\bar{v}})$. The unicity of such a minimum entails that $z = \bar{x}$. But this cannot hold by the second inequality in (6.2.1). Hence, $\bar{x}$ is a Pareto minimum of (P). □

At this step, it seems reasonable to envision the use of the previous results to solve (P). This conducts us to follow Dinkelbach's procedure. This procedure suggests us to start from an arbitrary point $x^0 \in X$ and then construct a sequence of points x^k by solving a sequence of parametric subproblems (P_{v_k}), where $v_k = v(x^k)$ and x^{k+1} is an optimal solution of (P_{v_k}). Unfortunately, this approach comes up against the nonconvexity of the subproblems (P_{v_k}), nevertheless, as we will see, their objective functions are maximum of difference of convex functions.

Thinking about the DCA algorithm [80, 124], we suggest to linearize the second convex component in the decomposition (in difference of convex functions) of each function $f_i - v_{i,k} g_i$, where $v_{i,k}$ for $i \in I$ are the components of v_k. Obviously, this decomposition, and consequently the linearization, depends on the sign of $v_{i,k}$. On the other hand, the set X is also generally nonconvex, so we will aproximate it by an inner convex subset of X by linearizing the second components of the constraint functions h_j, i.e. h_j^2, for $j \in J$. In doing so, we obtain approximate convex subproblems for (P_{v_k}). Similar idea was already developped in Chapters 3 and 4 for scalar generalized fractional programs and gave rise to DC-Dinkelbach algorithms

Firstly we start by gathering the pieces of the approximate problem. For this, and in all what follows, $\partial f_i^2(x)$, $\partial h_j^2(x)$ and $\partial g_i^\ell(x)$ will designate, respectively, the subdifferentials of the convex functions f_i^2, h_j^2 and g_i^ℓ at x, for $\ell = 1, 2$.

To concretize our previous discussion, consider for $y \in X$, arbitraries $x_i^2(y) \in \partial f_i^2(y)$ and $y_i^\ell(y) \in \partial g_i^\ell(y)$, for $\ell = 1$ or $\ell = 2$. Then, we define the objective function of the approximate subproblem $(P_{v(y)})$ by replacing, in its objective function, the functions f_i^2 and g_i^ℓ, $\ell = 1$ or $\ell = 2$, by their affine approximations at y, that is to say, by $f_i^2(y) + \langle x_i^2(y), x - y\rangle$ and $g_i^\ell(y) + \langle y_i^\ell(y), x - y\rangle$, respectively. More precisely, this function, parametrized by $y \in X$, is given by

$$F_y(x) := \max_{i \in I} \Big[f_{i,y}(x) - v_i(y) g_{i,y}(x) \Big], \tag{6.2.3}$$

where

$$f_{i,y}(x) = f_i^1(x) - [f_i^2(y) + \langle x_i^2(y), x - y\rangle], \tag{6.2.4}$$

$$g_{i,y}(x) := \begin{cases} g_i^1(x) - [g_i^2(y) + \langle y_i^2(y), x - y\rangle] & \text{if } v_i(y) < 0 \\ -g_i^2(x) + [g_i^1(y) + \langle y_i^1(y), x - y\rangle] & \text{if } v_i(y) \geq 0 \end{cases} \tag{6.2.5}$$

where $v_i(y)$ is the i–th component of the vector $v(y)$. Notice that the functions $f_{i,y}(\cdot)$ and $-v_i(y) g_{i,y}(\cdot)$ are convex for all $i \in I$ and $y \in X$, by the convexity assumptions made on the functions f_i^ℓ and g_i^ℓ, $\ell = 1,2$, and so the function $F_y(\cdot)$ is also convex.

In approximating the constraints of $(P_{v(y)})$, we will limite the constraints to a convex region of X. For this, consider arbitraries $z_j^2(y) \in \partial h_j^2(y)$ for $j \in J$, and define the functions $h_{j,y}$, parametrized by $y \in \mathbb{R}^n$, by

$$h_{j,y}(x) := h_j^1(x) - [h_j^2(y) + \langle z_j^2(y), x - y\rangle]. \tag{6.2.6}$$

The constraint set of the approximating convex subproblem is then defined by

$$X_y = \Big\{x \in C \mid h_{j,y}(x) \leq 0, \forall j \in J\Big\}. \tag{6.2.7}$$

Note that the convexity of the functions h_j^ℓ, $\ell = 1,2$, implies the convexity of the functions $h_{j,y}(\cdot)$, for all $j \in J$ and $y \in X$, and thus the convexity of the set X_y.

Remark 6.2.1. From the subgradient inequalities $h_j^2(x) \geq h_j^2(y) + \langle z_j^2(y), x - y\rangle$ for all $x, y \in \mathbb{R}^n$, $j \in J$, we conclude that $X_y \subset X$ for all $y \in \mathbb{R}^n$. On the other hand, $y \in X_y$ if and only if $y \in X$.

Now, for $y \in \mathbb{R}^n$, instead of the subproblem $(P_{v(y)})$, we associate to (P) its approximating convex problem

$$(P(y)) \qquad \inf_{x \in X_y} \Big\{F_y(x) := \max_{i \in I}\Big[f_{i,y}(x) - v_i(y) g_{i,y}(x)\Big]\Big\},$$

and we denote by x_y the global minimum of F_y over X_y.

We will see in the next development and analysis that the function F_y will play the important role of optimality function, by recognizing (weak) Pareto solutions by the minimization of $(P(y))$, and will be the central piece to exprime optimality conditions to vectoriel fractional programming.

The results we present below are close to those of Chapters 3 and 4, but written in the vectorial framework of fractional programming. Their proofs are similar to those of the above-mentioned results, we rewrite them for the sake of completeness.

Lemma 6.2.3. For all $x \in \mathbb{R}^n$ and $y \in X$ we have

$$\max_{i \in I}[f_i(x) - v_i(y) g_i(x)] \leq F_y(x)$$

Proof. Since $f_i(x) = f_i^1(x) - f_i^2(x)$ and $g_i(x) = g_i^1(x) - g_i^2(x)$ the subgradient inequalities $f_i^2(y) + \langle x_i^2(y), x-y\rangle \le f_i^2(x)$, for all $i \in I$ we have

$$\begin{aligned} f_i(x) - v_i(y)g_i(x) &= f_i^1(x) - f_i^2(x) - v_i(y)\left[g_i^1(x) - g_i^2(x)\right] \\ &\le f_i^1(x) - \left[f_i^2(y) + \langle x_i^2(y), x-y\rangle\right] - v_i(y)\left[g_i^1(x) - g_i^2(x)\right] \\ &= f_{i,y}(x) - v_i(y)\left[g_i^1(x) - g_i^2(x)\right]. \end{aligned} \tag{6.2.8}$$

If $v_i(y) < 0$, for $i \in I$, from the subgradient inequalities $g_i^2(y) + \langle y_i^2(y), x-y\rangle \le g_i^2(x)$, for $i \in I$, the definition of F_y, and (6.2.8) we get

$$\begin{aligned} f_i(x) - v_i(y)g_i(x) &\le f_{i,y}(x) - v_i(y)\left[g_i^1(x) - \left(g_i^2(y) + \langle y_i^2(y), x-y\rangle\right)\right] \\ &= f_{i,y}(x) - v_i(y)g_{i,y}(x) \\ &\le F_y(x). \end{aligned}$$

Assume now that $v_i(y) \ge 0$, for $i \in I$. Then, from the subgradient inequalities $g_i^1(y) + \langle y_i^1(y), x-y\rangle \le g_i^1(x)$, for $i \in I$, the definition of F_y, and (6.2.8) we get

$$\begin{aligned} f_i(x) - v_i(y)g_i(x) &\le f_{i,y}(x) - v_i(y)\left[g_i^1(y) + \langle y_i^1(y), x-y\rangle - g_i^2(x)\right] \\ &= f_{i,y}(x) - v_i(y)\left[-g_i^2(x) + \left(g_i^1(y) + \langle y_i^1(y), x-y\rangle\right)\right] \\ &= f_{i,y}(x) - v_i(y)g_{i,y}(x) \\ &\le F_y(x). \end{aligned}$$

Thus, $f_i(x) - v_i(y)g_i(x) \le F_y(x)$ for all $i \in I$, from which the desired inequality follows. □

Recall that for all $y \in X$, we designated by x_y a minimum of F_y over X_y. Then we have the following results.

Proposition 6.2.1. For all $y \in X$, we have

1. $F_y(y) = 0$ and $F_y(x_y) \le 0$,
2. $v(x_y) \le v(y)$.

Proof. 1. The equality $v_i(y) = f_i(y)/g_i(y)$ for all $i \in I$, and the definition of $F_y(y)$,

$$F_y(y) = \max_{i\in I}\left[f_i(y) - v_i(y)g_i(y)\right],$$

give the equality $F_y(y) = 0$.

The definition of x_y implies that $F_y(x_y) \le F_y(x)$ for all $x \in X_y$. With $x = y$ we get $F_y(x_y) \le F_y(y) = 0$.

2. From Lemma 6.2.3, with $x = x_y$, we have $f_i(x_y) - v_i(y)g_i(x_y) \le F_y(x_y) \le 0$, for all $i \in I$. This means that $f_i(x_y)/g_i(x_y) \le v_i(y)$ for all $i \in I$, implying that $v_i(x_y) \le v_i(y)$, for all $i \in I$. □

In the scalar framework, relations between the GFP and particular parametric subproblems was already pointed out in Chapters 3 and 4. We analyze in the next results relations between problem (P) and problem $(P(y))$, for particular y, in the multiobjective setting.

Proposition 6.2.2. Assume that problem (P) has a weak Pareto minimum $\bar{x} \in X$. Then $\bar{x}$ is a global minimizer of $F_{\bar{x}}$ on the set $X_{\bar{x}}$, whatever are $x_i^2(\bar{x}) \in \partial f_i^2(\bar{x})$, $y_i^\ell(\bar{x}) \in \partial g_i^\ell(\bar{x})$, where $\ell = 1$ if $v_i(\bar{x}) \ge 0$ and $\ell = 2$ otherwise, for $i \in I$; and $z_j^2(\bar{x}) \in \partial h_j^2(\bar{x})$ for $j \in J$. Conversely, for all weak Pareto minimum $\bar{x}$ of (P), every optimal solution $x_{\bar{x}}$ of $F_{\bar{x}}$ over $X_{\bar{x}}$ is also a weak Pareto minimum of problem (P).

Proof. Let $\bar{x} \in X$ be a weak Pareto minimum of (P), $\bar{v} = v(\bar{x})$, $x_i^2(\bar{x}) \in \partial f_i^2(\bar{x})$, $y_i^\ell(\bar{x}) \in \partial g_i^\ell(\bar{x})$, where $\ell = 1$ if $\bar{v}_i \ge 0$ and $\ell = 2$ otherwise, for $i \in I$; and $z_j^2(\bar{x}) \in \partial h_j^2(\bar{x})$, for $j \in J$, and let $F_{\bar{x}}$ and $X_{\bar{x}}$ as defined in (6.2.3) and (6.2.7) respectively. Suppose, for contradiction, that $\bar{x}$ is not a global minimum of $F_{\bar{x}}$ on the set $X_{\bar{x}}$. Then there exists some $z \in X_{\bar{x}}$ such that $F_{\bar{x}}(z) < F_{\bar{x}}(\bar{x})$. By using Proposition 6.2.1, Item 1, with $y = \bar{x}$, we see that $F_{\bar{x}}(\bar{x}) = 0$. It follows that $F_{\bar{x}}(z) < 0$. On the other hand, Lemma 6.2.3, with $y = \bar{x}$, together with the last inequality, entails that $f_i(z) - \bar{v}_i g_i(z) < 0$, for all $i \in I$, since $z \in X_{\bar{x}} \subset X$, see Remark 6.2.1, giving rise to $v_i(z) < \bar{v}_i := v_i(\bar{x})$, for all $i \in I$. This contradicts the fact that $\bar{x}$ is a weak Pareto minimum of (P). To prove the converse assertion, let $\bar{x} \in X$ be a weak Pareto minimum of (P) and let $F_{\bar{x}}$ be as defined in (6.2.3) with arbitraries $x_i^2(\bar{x}) \in \partial f_i^2(\bar{x})$, $y_i^l(\bar{x}) \in \partial g_i^l(\bar{x})$, where $\ell = 1$ if $\bar{v}_i \ge 0$ and $\ell = 2$ otherwise, for $i \in I$; and construct $X_{\bar{x}}$ conforming to (6.2.7) with some $z_j^2(\bar{x}) \in \partial h_j^2(\bar{x})$ for $j \in J$. Assume that there is a global minimum $x_{\bar{x}}$ of $F_{\bar{x}}$ over $X_{\bar{x}}$ which is not a weak Pareto minimum of (P). Then there exists $z \in X$ such that $v(z) < v(x_{\bar{x}})$. From Proposition 6.2.1, Item 2, with $y = \bar{x}$, we have $v(x_{\bar{x}}) \le v(\bar{x})$. It follows that $v(z) < v(\bar{x})$, thereby contradicting the fact that $\bar{x}$ is a weak Pareto minimum of (P). □

The same results hold for Pareto minimum. This is detailed in the next proposition.

Proposition 6.2.3. Assume that problem (P) has a Pareto minimum $\bar{x} \in X$. Then $\bar{x}$ is a global minimizer of $F_{\bar{x}}$ on the set $X_{\bar{x}}$, whatever are $x_i^2(\bar{x}) \in \partial f_i^2(\bar{x})$, $y_i^l(\bar{x}) \in \partial g_i^l(\bar{x})$, where $\ell = 1$ if $v_i(\bar{x}) \geq 0$ and $\ell = 2$ otherwise, for $i \in I$; and $z_j^2(\bar{x}) \in \partial h_j^2(\bar{x})$ for $j \in J$. Conversely, for all Pareto minimum $\bar{x}$ of (P), every optimal solution $x_{\bar{x}}$ of $F_{\bar{x}}$ over $X_{\bar{x}}$ is also a Pareto minimum of problem (P).

Proof. Since Pareto minimum of (P) is also weak Pareto minimum, the first implication follows from Proposition 6.2.2. To prove the converse assertion, let $\bar{x} \in X$ be a Pareto minimum of (P) and let $F_{\bar{x}}$ and $X_{\bar{x}}$ as defined in the proof of Proposition 6.2.2. Assume, for contradiction, that there is a global minimum $x_{\bar{x}}$ of $F_{\bar{x}}$ over $X_{\bar{x}}$ which is not a Pareto minimum of (P). Then there exists $z \in X$ such that $v(z) \leq v(x_{\bar{x}})$ and $v_j(z) < v_j(x_{\bar{x}})$, for some $j \in I$. From Proposition 6.2.1, Item 2, with $y = \bar{x}$, we have $v(x_{\bar{x}}) \leq v(\bar{x})$. It follows that $v(z) \leq v(\bar{x})$, and $v_j(z) < v_j(\bar{x})$, for some $j \in I$, which is absurd since $\bar{x}$ is a Pareto minimum of (P). □

6.2.1 Optimality Conditions

Based on the previous results, we give now optimality conditions for the multiobjective fractional program (P), using only convex analysis tools in the scalar setting.

Theorem 6.2.1. Let $\hat{x} \in X$ and $\hat{v} = v(\hat{x})$. Assume that for every $i \in I$, there exist $x_i^2(\hat{x}) \in \partial f_i^2(\hat{x})$, $y_i^\ell(\hat{x}) \in \partial g_i^\ell(\hat{x})$, where $\ell = 1$ if $\hat{v}_i \geq 0$ and $\ell = 2$ otherwise; and for all $j \in J$, there exist $z_j^2(\hat{x}) \in \partial h_j^2(\hat{x})$, such that $F_{\hat{x}}(x_{\hat{x}}) = 0$, where we recall that $x_{\hat{x}}$ is a global minimum of $F_{\hat{x}}$ on $X_{\hat{x}} := \{x \in C \mid h_{j,\hat{x}}(x) \leq 0, \forall j \in J\}$. Then, for all $i \in I$, there exist $\hat{\mu}_i \geq 0$, and for all $j \in J$, there exist $\hat{\nu}_j \geq 0$, with $\sum_{i \in I} \hat{\mu}_i + \sum_{j \in J} \hat{\nu}_j = 1$, such that

$$\begin{aligned} 0 \in &\sum_{i \in I} \hat{\mu}_i \left[\partial f_i^1(\hat{x}) - \partial f_i^2(\hat{x}) - \hat{v}_i \left(\partial g_i^1(\hat{x}) - \partial g_i^2(\hat{x})\right)\right] \\ &+ \sum_{j \in J} \hat{\nu}_j \left[\partial h_j^1(\hat{x}) - \partial h_j^2(\hat{x})\right] + N_C(\hat{x}), \end{aligned} \tag{6.2.9}$$

where $N_C(\hat{x})$ is the normal cone to C at $\hat{x}$, with the equality

$$\hat{\nu}_j h_j(\hat{x}) = 0 \quad \text{for all } j \in J. \tag{6.2.10}$$

The converse is true, that is $F_{\hat{x}}(x_{\hat{x}}) = 0$, for some $x_i^2(\hat{x}) \in \partial f_i^2(\hat{x})$, $y_i^\ell(\hat{x}) \in$

$\partial g_i^\ell(\hat{x})$, for $\ell = 1,2$ and $i \in I$, if in addition to (6.2.9) and (6.2.10), one has

$$\max_{j\in J} h_{j,\hat{x}}(x) < 0 \text{ for all } z_j^2(\hat{x}) \in \partial h_j^2(\hat{x}) \text{ and } j \in J, \tag{6.2.11}$$

for some $x \in C$. This condition also implies that $\sum_{i\in I} \hat{\mu}_i \neq 0$.

Proof. Let $\hat{x} \in X$ and $\hat{v} = v(\hat{x})$. Assume that there exist subgradients $x_i^2(\hat{x}) \in \partial f_i^2(\hat{x})$, $y_i^\ell(\hat{x}) \in \partial g_i^\ell(\hat{x})$, where $\ell = 1$ if $\hat{v}_i \geq 0$ and $\ell = 2$ if $\hat{v}_i < 0$, for all $i \in I$; and there exists $z_j^2(\hat{x}) \in \partial h_j^2(\hat{x})$ for all $j \in J$, such that $F_{\hat{x}}(x_{\hat{x}}) = 0$, where we recall that the function $F_{\hat{x}}$ is defined conforming to Eqs. (6.2.3) to (6.2.5), and $x_{\hat{x}}$ is its global minimum on the set $X_{\hat{x}}$ defined in accordance to (6.2.6) and (6.2.7). Then since from Proposition 6.2.1, Item 1, with $y = \hat{x}$, we have $F_{\hat{x}}(\hat{x}) = 0$, we conclude from the assumption that $F_{\hat{x}}(x_{\hat{x}}) = 0$, that $\hat{x}$ is also a global minimum over the convex set $X_{\hat{x}}$, of the convex function $F_{\hat{x}}$. To establish optimality conditions, we claim that $\hat{x}$ also minimizes over the convex set C, the convex function $\hat{F}_{\hat{x}}$ defined by

$$\hat{F}_{\hat{x}}(x) := \max\left[F_{\hat{x}}(x), \max_{j\in J} h_{j,\hat{x}}(x)\right].$$

Obviously, for all $x \in X_{\hat{x}}$ we have $F_{\hat{x}}(x) \geq 0$, since the minimum value of $F_{\hat{x}}(x) \geq 0$ on $X_{\hat{x}}$ is zero from our assumption. This implies that $\hat{F}_{\hat{x}}(x) \geq 0$ for all $x \in X_{\hat{x}}$. For $x \notin X_{\hat{x}}$, but $x \in C$ it holds that $h_{j,\hat{x}}(x) > 0$ for some $j \in J$ and this again implies that $\hat{F}_{\hat{x}}(x) \geq 0$. In conclusion $\hat{F}_{\hat{x}}(x) \geq 0$ for all $x \in C$. On the other hand, $\hat{F}_{\hat{x}}(\hat{x}) = 0$ since $F_{\hat{x}}(\hat{x}) = 0$ and $h_{j,\hat{x}}(\hat{x}) \leq 0$ for all $j \in J$. This gives the conclusion. Therefore, from [64, Theorem 1.1.1] we conclude that

$$0 \in \partial\hat{F}_{\hat{x}}(\hat{x}) + N_C(\hat{x}),$$

where $\partial\hat{F}_{\hat{x}}(\hat{x})$ and $N_C(\hat{x})$ are, respectively, the subdifferential of $\hat{F}_{\hat{x}}$ and the normal cone of C at $\hat{x}$. By invoking [64, Corollary 4.3.2] to express $\partial\hat{F}_{\hat{x}}(\hat{x})$, we conclude that there exist $\hat{\alpha}_0, \hat{\alpha}_j \geq 0$ for $j \in J$, such that $\hat{\alpha}_0 + \sum_{j\in J} \hat{\alpha}_j = 1$,

$$0 \in \hat{\alpha}_0 \partial F_{\hat{x}}(\hat{x}) + \sum_{j\in J} \hat{\alpha}_j \partial h_{j,\hat{x}}(\hat{x}) + N_C(\hat{x})$$

and $\hat{\alpha}_j h_{j,\hat{x}}(\hat{x}) = 0$, for all $j \in J$, where $\partial F_{\hat{x}}(\hat{x})$, $\partial h_{j,\hat{x}}(\hat{x})$ and $N_C(\hat{x})$ are, respectively, the subdifferential of $F_{\hat{x}}$, the subdifferential of $h_{j,\hat{x}}$, $j \in J$, and the normal cone of C at $\hat{x}$. By referring to [64, Corollary 4.3.2] in the calculus of $\partial F_{\hat{x}}(\hat{x})$, there exist $\hat{\beta}_i \geq 0$, $i \in I$, with $\sum_{i\in I} \hat{\beta}_i = 1$ such that

$$0 \in \hat{\alpha}_0 \sum_{i\in I} \hat{\beta}_i \partial\left[f_{i,\hat{x}}(\hat{x}) - \hat{v}_i g_{i,\hat{x}}(\hat{x})\right] + \sum_{j\in J} \hat{\alpha}_j \partial h_{j,\hat{x}}(\hat{x}) + N_C(\hat{x}). \tag{6.2.12}$$

and

$$\sum_{i\in I}\hat{\beta}_i\left[f_{i,\hat{x}}(\hat{x})-\hat{v}_i g_{i,\hat{x}}(\hat{x})\right]=F_{\hat{x}}(\hat{x})=0. \tag{6.2.13}$$

Let us define the index sets $\hat{I}_1 := \{i \in I \mid \hat{v}_i < 0\}$ and $\hat{I}_2 := \{i \in I \mid \hat{v}_i \geq 0\}$. Then from the expression of $f_{i,\hat{x}}$, $g_{i,\hat{x}}$ and $h_{j,\hat{x}}$ it is clear that (6.2.12) and (6.2.13), respectively imply that

$$\begin{aligned}0 \in \hat{\alpha}_0 &\sum_{i\in\hat{I}_1}\hat{\beta}_i\left[\partial f_i^1(\hat{x})-x_i^2(\hat{x})-\hat{v}_i\left(\partial g_i^1(\hat{x})-y_i^2(\hat{x})\right)\right]\\ &+\hat{\alpha}_0\sum_{i\in\hat{I}_2}\hat{\beta}_i\left[\partial f_i^1(\hat{x})-x_i^2(\hat{x})-\hat{v}_i\left(y_i^1(\hat{x})-\partial g_i^2(\hat{x})\right)\right]\\ &+\sum_{j\in J}\hat{\alpha}_j\left[\partial h_j^1(\hat{x})-z_j^2(\hat{x})\right]+N_C(\hat{x}),\end{aligned} \tag{6.2.14}$$

and

$$\sum_{i\in I}\hat{\beta}_i\left[f_i^1(\hat{x})-f_i^2(\hat{x})-\hat{v}_i\left(g_i^1(\hat{x})-g_i^2(\hat{x})\right)\right]:=\sum_{i\in I}\hat{\beta}_i\left[f_i(\hat{x})-\hat{v}_i g_i(\hat{x})\right]=0.$$

Clearly, (6.2.14) implies that

$$\begin{aligned}0 \in \hat{\alpha}_0 &\sum_{i\in\hat{I}_1}\hat{\beta}_i\left[\partial f_i^1(\hat{x})-\partial f_i^2(\hat{x})-\hat{v}_i\left(\partial g_i^1(\hat{x})-\partial g_i^2(\hat{x})\right)\right]\\ &+\hat{\alpha}_0\sum_{i\in\hat{I}_2}\hat{\beta}_i\left[\partial f_i^1(\hat{x})-\partial f_i^2(\hat{x})-\hat{v}_i\left(\partial g_i^1(\hat{x})-\partial g_i^2(\hat{x})\right)\right]\\ &+\sum_{j\in J}\hat{\alpha}_j\left[\partial h_j^1(\hat{x})-\partial h_j^2(\hat{x})\right]+N_C(\hat{x}).\end{aligned} \tag{6.2.15}$$

Finally, (6.2.15) implies that

$$\begin{aligned}0 \in \hat{\alpha}_0 &\sum_{i\in I}\hat{\beta}_i\left[\partial f_i^1(\hat{x})-\partial f_i^2(\hat{x})-\hat{v}_i\left(\partial g_i^1(\hat{x})-\partial g_i^2(\hat{x})\right)\right]\\ &+\sum_{j\in J}\hat{\alpha}_j\left[\partial h_j^1(\hat{x})-\partial h_j^2(\hat{x})\right]+N_C(\hat{x}).\end{aligned} \tag{6.2.16}$$

It suffices to set $\hat{\mu}_i = \hat{\alpha}_0\hat{\beta}_i$ and $\hat{v}_j = \hat{\alpha}_j$ to get the results (6.2.9) and (6.2.10), since $\sum_{i\in I}\hat{\mu}_i + \sum_{j\in J}\hat{v}_j = \hat{\alpha}_0\sum_{i\in I}\hat{\beta}_i + \sum_{j\in J}\hat{\alpha}_j = 1$.

To show the converse, that is (6.2.9) and (6.2.10) imply that for some $x_i^2(\hat{x}) \in \partial f_i^2(\hat{x})$ and $y_i^\ell(\hat{x}) \in \partial g_i^\ell(\hat{x})$, for $\ell = 1,2$ and $i \in I$, we have $F_{\hat{x}}(x_{\hat{x}}) = 0$, assume that (6.2.9) and (6.2.10) hold. Then from (6.2.9), for all $i \in I$, $j \in J$, there exist $\hat{x}_i^\ell \in \partial f_i^\ell(\hat{x})$, $\hat{y}_i^\ell \in \partial g_i^\ell(\hat{x})$ and $\hat{z}_j^\ell \in \partial h_j^\ell(\hat{x})$, for $\ell = 1,2$, such that

$$\sum_{i\in I} \hat{\mu}_i \left[\hat{x}_i^1 - \hat{x}_i^2 - \hat{v}_i \left(\hat{y}_i^1 - \hat{y}_i^2\right)\right] + \sum_{j\in J} \hat{\nu}_j \left[\hat{z}_j^1 - \hat{z}_j^2\right] \in -N_C(\hat{x}).$$

That is, for all $x \in C$ we have

$$\left\langle \sum_{i\in I} \hat{\mu}_i \left[\hat{x}_i^1 - \hat{x}_i^2 - \hat{v}_i \left(\hat{y}_i^1 - \hat{y}_i^2\right)\right] + \sum_{j\in J} \hat{\nu}_j \left[\hat{z}_j^1 - \hat{z}_j^2\right], x - \hat{x} \right\rangle \geq 0. \qquad (6.2.17)$$

For all $x \in X$, by using the subgradient inequality $f_i^1(x) \geq f_i^1(\hat{x}) + \langle \hat{x}_i^1, x - \hat{x}\rangle$, for $i \in I$, we get

$$f_i^1(x) - \left(f_i^2(\hat{x}) + \langle \hat{x}_i^2, x - \hat{x}\rangle\right) \geq f_i^1(\hat{x}) - f_i^2(\hat{x}) + \langle \hat{x}_i^1 - \hat{x}_i^2, x - \hat{x}\rangle.$$

To comply with the notation of the definition of the functions in Eqs. (6.2.4) to (6.2.6), we set $x_i^2(\hat{x}) = \hat{x}_i^2$, $y_i^\ell(\hat{x}) = \hat{y}_i^\ell$, for $\ell = 1,2$, $i \in I$, and $z_j^2(\hat{x}) = \hat{z}_j^2$, $j \in J$. With these notations, the last inequality becomes

$$f_{i,\hat{x}}(x) \geq f_i(\hat{x}) + \langle \hat{x}_i^1 - \hat{x}_i^2, x - \hat{x}\rangle, \text{ for } i \in I. \qquad (6.2.18)$$

Assume first that $\hat{v}_i < 0$, i.e. $i \in \hat{I}_1$. The subgradient inequality $g_i^1(x) \geq g_i^1(\hat{x}) + \langle \hat{y}_i^1, x - \hat{x}\rangle$, implies that

$$g_i^1(x) - \left(g_i^2(\hat{x}) + \langle \hat{y}_i^2, x - \hat{x}\rangle\right) \geq g_i^1(\hat{x}) - g_i^2(\hat{x}) + \langle \hat{y}_i^1 - \hat{y}_i^2, x - \hat{x}\rangle.$$

Taking into account the definition of $g_{i,\hat{x}}$, we obtain

$$-\hat{v}_i g_{i,\hat{x}}(x) \geq -\hat{v}_i \left(g_i(\hat{x}) + \langle \hat{y}_i^1 - \hat{y}_i^2, x - \hat{x}\rangle\right). \qquad (6.2.19)$$

For the case $\hat{v}_i \geq 0$, i.e. $i \in \hat{I}_2$, we consider the subgradient inequality $g_i^2(x) \geq g_i^2(\hat{x}) + \langle \hat{y}_i^2, x - \hat{x}\rangle$, to get

$$-g_i^2(x) + \left(g_i^1(\hat{x}) + \langle \hat{y}_i^1, x - \hat{x}\rangle\right) \leq -g_i^2(\hat{x}) + g_i^1(\hat{x}) + \langle \hat{y}_i^1 - \hat{y}_i^2, x - \hat{x}\rangle.$$

By again referring to the definition of $g_{i,\hat{x}}$, we also obtain (6.2.19) with $i \in \hat{I}_2$, from the previous inequality by multiplying it by $-\hat{v}_i$. So, adding (6.2.18) to (6.2.19) in both cases, we arrive to the inequalities

$$f_{i,\hat{x}}(x) - \hat{v}_i g_{i,\hat{x}}(x) \geq f_i(\hat{x}) - \hat{v}_i g_i(\hat{x}) + \left\langle \hat{x}_i^1 - \hat{x}_i^2 - \hat{v}_i \left(\hat{y}_i^1 - \hat{y}_i^2\right), x - \hat{x} \right\rangle \quad \text{for all } i \in I,$$

By invoking the definition of $F_{\hat{x}}(x)$ and $\hat{v}_i$ we get

$$F_{\hat{x}}(x) \geq \left\langle \hat{x}_i^1 - \hat{x}_i^2 - \hat{v}_i \left(\hat{y}_i^1 - \hat{y}_i^2 \right), x - \hat{x} \right\rangle \quad \text{for all } i \in I. \tag{6.2.20}$$

On the other hand, the subgradient inequality $h_j^1(x) \geq h_j^1(\hat{x}) + \langle \hat{z}_j^1, x - \hat{x} \rangle$, for $j \in J$, gives

$$h_j^1(x) - \left(h_j^2(\hat{x}) + \langle \hat{z}_j^2, x - \hat{x} \rangle \right) \geq h_j^1(\hat{x}) - h_j^2(\hat{x}) + \langle \hat{z}_j^1 - \hat{z}_j^2, x - \hat{x} \rangle,$$

which means that

$$h_{j,\hat{x}}(x) \geq h_j(\hat{x}) + \langle \hat{z}_j^1 - \hat{z}_j^2, x - \hat{x} \rangle. \tag{6.2.21}$$

Multiplying both sides of the inequalities in (6.2.20) by $\hat{\mu}_i$, for all $i \in I$ and (6.2.21) by $\hat{v}_j$, for all $j \in J$, summing the resulting inequalities, and taking into account (6.2.10), we get

$$\sum_{i \in I} \hat{\mu}_i F_{\hat{x}}(x) + \sum_{j \in J} \hat{v}_j h_{j,\hat{x}}(x) \geq \Bigg\langle \sum_{i \in I} \hat{\mu}_i \left[\hat{x}_i^1 - \hat{x}_i^2 - \hat{v}_i \left(\hat{y}_i^1 - \hat{y}_i^2 \right) \right] + \sum_{j \in J} \hat{v}_j \left[\hat{z}_j^1 - \hat{z}_j^2 \right], x - \hat{x} \Bigg\rangle. \tag{6.2.22}$$

By using (6.2.17), then (6.2.22) implies

$$F_{\hat{x}}(x) \sum_{i \in I} \hat{\mu}_i + \sum_{j \in J} \hat{v}_j h_{j,\hat{x}}(x) \geq 0 \text{ for all } x \in C. \tag{6.2.23}$$

In particular, for $x \in X_{\hat{x}}$ we have $h_{j,\hat{x}}(x) \leq 0$ for all $j \in J$, which implies that $F_{\hat{x}}(x) \sum_{i \in I} \hat{\mu}_i \geq 0$. It is clear that (6.2.11) and (6.2.23) imply that $\sum_{i \in I} \hat{\mu}_i \neq 0$, since othrwise we should have

$$\max_{j \in J} h_{j,\hat{x}}(x) \geq \sum_{j \in J} \hat{v}_j h_{j,\hat{x}}(x) \geq 0 \text{ for all } x \in C,$$

thereby contradicting (6.2.11) with $z_j^2(\hat{x}) = \hat{z}_j^2 \in \partial h_j^2(\hat{x})$ for all $j \in J$. Therefore, $F_{\hat{x}}(x) \geq 0$. Let $x_{\hat{x}}$ be a global minimum of $F_{\hat{x}}$ over $X_{\hat{x}}$. Then $x_{\hat{x}} \in X_{\hat{x}}$, and since by Proposition 6.2.1, Item 1, $F_{\hat{x}}(x_{\hat{x}}) \leq 0$, we conclude that $F_{\hat{x}}(x_{\hat{x}}) = 0$. This achieves the proof. □

In the next corollary we specify Theorem 6.2.1 to the case of continuously differentiable functions.

Corollary 6.2.1. Let $\hat{x} \in X$ and $\hat{v} = v(\hat{x})$. Assume that the functions f_i^ℓ, g_i^ℓ, for $i \in I$ and h_j^ℓ, for $j \in J$, and $\ell = 1, 2$, are continuously differentiable. If $F_{\hat{x}}(x_{\hat{x}}) = 0$, where $x_{\hat{x}}$ is a global minimum of $F_{\hat{x}}$ on $X_{\hat{x}}$, then, for all $i \in I$, there exist $\hat{\mu}_i \geq 0$, and for all $j \in J$, there exist $\hat{\nu}_j \geq 0$, with $\sum_{i\in I} \hat{\mu}_i + \sum_{j\in J} \hat{\nu}_j = 1$, such that

$$\sum_{i\in I} \hat{\mu}_i \left[\nabla f_i^1(\hat{x}) - \nabla f_i^2(\hat{x}) - \hat{v}_i \left(\nabla g_i^1(\hat{x}) - \nabla g_i^2(\hat{x})\right)\right] + \sum_{j\in J} \hat{\nu}_j \left[\nabla h_j^1(\hat{x}) - \nabla h_j^2(\hat{x})\right] \in -N_C(\hat{x}), \tag{6.2.24}$$

where $\nabla f_i^\ell(\hat{x})$, $\nabla g_i^\ell(\hat{x})$, $\nabla h_j^\ell(\hat{x})$, for $\ell = 1, 2$ and $N_C(\hat{x})$ are respectively, the gradient of f_i^ℓ, g_i^ℓ, h_j^ℓ and the normal cone to C at $\hat{x}$; with the equality

$$\hat{\nu}_j h_j(\hat{x}) = 0 \quad \text{for all } j \in J. \tag{6.2.25}$$

The converse is true, that is $F_{\hat{x}}(x_{\hat{x}}) = 0$, if in addition to (6.2.24) and (6.2.25), one has

$$\max_{j\in J} h_{j,\hat{x}}(x) < 0, \tag{6.2.26}$$

for some $x \in C$.

Proof. It suffices to notice that for these convex continuously differentiable functions we have $\partial f_i^\ell(\hat{x}) = \left\{\nabla f_i^\ell(\hat{x})\right\}$, $\partial g_i^\ell(\hat{x}) = \left\{\nabla g_i^\ell(\hat{x})\right\}$, for $i \in I$ and $\partial h_j^\ell(\hat{x}) = \left\{\nabla h_j^\ell(\hat{x})\right\}$, for $j \in J$ and $\ell = 1, 2$, and then to use Theorem 6.2.1. □

In the next proposition, we will show that under some additional assumptions, the KKT criticality conditions stated in (6.2.9) and (6.2.10) of Theorem 6.2.1 conduct to Clarke stationary ones.

Proposition 6.2.4. Let $\hat{x} \in X$ and assume that we have $\partial f_i^1(\hat{x}) - \partial f_i^2(\hat{x}) = \partial^c[f_i^1 - f_i^2](\hat{x})$, $\partial g_i^1(\hat{x}) - \partial g_i^2(\hat{x}) = \partial^c[g_i^1 - g_i^2](\hat{x})$, for all $i \in I$, and that $\partial h_j^1(\hat{x}) - \partial h_j^2(\hat{x}) = \partial^c[h_j^1 - h_j^2](\hat{x})$, for all $j \in J$, where ∂^c stands for the Clarke subdifferential. Assume on the other hand that $v_i(\hat{x}) \geq 0$ and that the functions f_i and $-g_i$, for all $i \in I$, are regular at $\hat{x}$, in the sense of Clarke. If (6.2.11) holds, then (6.2.9) and (6.2.10) imply Clarke stationary conditions, that is to say, for all $i \in I$ there exist $\hat{\alpha}_i \geq 0$, and for all $j \in J$, there exist $\hat{\beta}_j \geq 0$ such that

$$0 \in \sum_{i\in I} \hat{\alpha}_i \partial^c v_i(\hat{x}) + \sum_{j\in J} \hat{\beta}_j \partial^c h_j(\hat{x}) + N_C(\hat{x}),$$

with the equalities $\hat{\beta}_j h_j(\hat{x}) = 0$ for $j \in J$ and $\sum_{i\in I} \hat{\alpha}_i + \sum_{j\in J} \hat{\beta}_j \neq 0$. If in addition condition (6.2.11) is fulfilled, then $\sum_{i\in I} \hat{\alpha}_i \neq 0$.

Proof. Let $\hat{x} \in X$, $\hat{v}_i = v_i(\hat{x})$, $i \in I$ and assume that we have (6.2.9) and (6.2.10). Firstly, recall that we used the notations $f_i = f_i^1 - f_i^2$, $g_i = g_i^1 - g_i^2$ and $h_j = h_j^1 - h_j^2$, and that these functions are locally Lipschitz (see e.g., [30, page 9] for definition) as they are difference of convex functions, and these latter are locally Lipschitz [30, Proposition 2.2.6]. Now by the assumptions made in the beginning of the proposition, we rewrite (6.2.9) as

$$0 \in \sum_{i\in I} \hat{\mu}_i \left[\partial^c f_i(\hat{x}) - \frac{f_i(\hat{x})}{g_i(\hat{x})}\partial^c g_i(\hat{x})\right] + \sum_{j\in J} \hat{v}_j \partial^c h_j(\hat{x}) + N_C(\hat{x}).$$

Thus,

$$0 \in \sum_{i\in I} \hat{\mu}_i g_i(\hat{x}) \left[\frac{g_i(\hat{x})\partial^c f_i(\hat{x}) - f_i(\hat{x})\partial^c g_i(\hat{x})}{g_i(\hat{x})^2}\right] + \sum_{j\in J} \hat{v}_j \partial^c h_j(\hat{x}) + N_C(\hat{x}). \quad (6.2.27)$$

Since by our assumptions the functions f_i and $-g_i$ are regular at $\hat{x}$, in the sense of Clarke [30, Definition 2.3.4], and $f_i(\hat{x}) \geq 0$, [30, Proposition 2.3.14] entails that

$$\partial^c \left[\frac{f_i}{g_i}\right](\hat{x}) = \frac{g_i(\hat{x})\partial^c f_i(\hat{x}) - f_i(\hat{x})\partial^c g_i(\hat{x})}{g_i(\hat{x})^2}.$$

Using this equality in (6.2.27) we get

$$0 \in \sum_{i\in I} \hat{\mu}_i g_i(\hat{x}) \partial^c \left[\frac{f_i}{g_i}\right](\hat{x}) + \sum_{j\in J} \hat{v}_j \partial^c h_j(\hat{x}) + N_C(\hat{x}),$$

Then

$$0 \in \sum_{i\in I} \hat{\mu}_i g_i(\hat{x}) \partial^c v_i(\hat{x}) + \sum_{j\in J} \hat{v}_j \partial^c h_j(\hat{x}) + N_C(\hat{x}),$$

It suffices to set $\hat{\alpha}_i = \hat{\mu}_i g_i(\hat{x})$, for all $i \in I$, and $\hat{\beta}_j = \hat{v}_j$, for all $j \in J$, to get

$$0 \in \sum_{i\in I} \hat{\alpha}_i \partial^c v_i(\hat{x}) + \sum_{j\in J} \hat{\beta}_j \partial^c h_j(\hat{x}) + N_C(\hat{x}).$$

On the other hand, (6.2.10) implies that $\hat{v}_j h_j(\hat{x}) = 0$, which is exactly $\hat{\beta}_j h_j(\hat{x}) = 0$. Finally, note that since $\sum_{i\in I} \hat{\mu}_i + \sum_{j\in J} \hat{v}_j = 1$ we have $\sum_{i\in I} \hat{\alpha}_i + \sum_{j\in J} \hat{\beta}_j \neq 0$. By following the proof of Theorem 6.2.1, we see that $\sum_{i\in I} \hat{\alpha}_i \neq 0$ since $\sum_{i\in I} \hat{\mu}_i \neq 0$ and $g_i(\hat{x}) > 0$ for all $i \in I$. □

The following theorem shows that weak Pareto, and a fortiori Pareto, minimums of (P) satisfy the optimality conditions (6.2.9) and (6.2.10).

Theorem 6.2.2. Let $\bar{x} \in X$ be a weak Pareto minimum of (P) and let $\bar{v} = v(\bar{x})$. Then for all $i \in I$ and $j \in J$, there exist $\bar{\mu}_i, \bar{\nu}_j \geq 0$, with $\sum_{i\in I} \bar{\mu}_i + \sum_{j\in J} \bar{\nu}_j = 1$, such that

$$0 \in \sum_{i\in I} \bar{\mu}_i \left[\partial f_i^1(\bar{x}) - \partial f_i^2(\bar{x}) - \bar{v}_i \left(\partial g_i^1(\bar{x}) - \partial g_i^2(\bar{x})\right)\right] + \sum_{j\in J} \bar{\nu}_j \left[\partial h_j^1(\bar{x}) - \partial h_j^2(\bar{x})\right] + N_C(\bar{x})$$

with $\bar{\nu}_j h_j(\bar{x}) = 0$, for all $j \in J$.

Proof. Let $\bar{x} \in X$ be a weak Pareto minimum of (P). Then from Proposition 6.2.2, $\bar{x}$ is also a global minimizer on the constraint set $X_{\bar{x}}$ of the function $F_{\bar{x}}$ defined conforming to Eqs. (6.2.3) to (6.2.6). Since $F_{\bar{x}}(\bar{x}) = 0$ (see Proposition 6.2.1, Item 1 with $y = \bar{x}$), then it suffices to use Theorem 6.2.1, with $\hat{x} = \bar{x}$ and $x_{\hat{x}} = \bar{x}$, to conclude. □

6.2.2 Dinkelbach-type Algorithm for (P)

In this section we will describe an algorithm that imitates Dinkelbach-type procedure, but this time in the multiobjective framework, which requires some special adaptations. For computational purposes, we write the components of the function v as

$$v_i(x) = \frac{f_i(x)/\omega_i}{g_i(x))/\omega_i} = \frac{\left(f_i^1(x) - f_i^2(x)\right)/\omega_i}{\left(g_i^1(x) - g_i^2(x)\right)/\omega_i} \quad \text{for all } i \in I, \tag{6.2.28}$$

for all $\omega_i > 0$, $i \in I$. Notice that this way of writing v_i has no effect on the function itself, but gives rise to different parametric subproblems since the parametrized functions defined conforming to Eqs. (6.2.3) to (6.2.5) takes the form

$$F_y(x) = \max_{i\in I} \left[\frac{f_{i,y}(x) - v_i(y) g_{i,y}(x)}{\omega_i}\right].$$

Obviously, the previous results may be obtained by directly replacing $f_{i,y}$ and $g_{i,y}$, respectively by $f_{i,y}/\omega_i$ and $g_{i,y}/\omega_i$.

Thinking about Theorem 6.2.1 while trying to solve the multiobjective problem (P), we try to find $\bar{x} \in X$ such that $F_{\bar{x}}(x_{\bar{x}}) = 0$. This idea comes up against the fact that the function $F_{\bar{x}}$ is unknown. Therefore we will get around this problem by approximating this function, iteratively, at each step k by a function $F_k := F_{x^k}$, where x^k is a global minimum of the convex function $F_{k-1} := F_{x^{k-1}}$ on the convex set $X_{k-1} := X_{x^{k-1}}$, the latter being considered as an inner approximating convex set of the nonconvex constraint set X. This scheme leads us straight to the following algorithm.

Algorithm 7 Vector DC-Dinkelbach-type Algorithm

0. Let $\{\varepsilon_k\}$ be a sequence of nonnegative reals such that $\sum_{k\geq 0} \varepsilon_k < \infty$. Choose $x^0 \in X$ and let $k = 0$.
1. At step k we have x^k, $v_{i,k} = v_i(x^k)$, $\varepsilon_k \geq 0$, $\bar{\omega} \geq \omega_i^k \geq \underline{\omega} > 0$, $x_{i,k}^2 \in \partial f_i^2(x^k)$, $y_{i,k}^\ell \in \partial g_i^\ell(x^k)$, for all $i \in I$, with $\ell = 1$ if $v_{i,k} \geq 0$, and $\ell = 2$ if $v_{i,k} < 0$; and $z_{j,k}^2 \in \partial h_j^2(x^k)$ for all $j \in J$. Then find $x^{k+1} \in X_k$ such that

$$F_k(x^{k+1}) \leq \inf_{x \in X_k} F_k(x) + \varepsilon_k,$$

   where

$$F_k(x) = \max_{i \in I} \left[\frac{f_{i,k}(x) - v_{i,k} g_{i,k}(x)}{\omega_i^k} \right],$$

   and $X_k = \{x \in C \mid h_{j,k}(x) \leq 0, \forall j \in J\}$, with $f_{i,k} = f_{i,x^k}$, $g_{i,k} = g_{i,x^k}$ and $h_{j,k} = h_{j,x^k}$ are defined as in Eqs. (6.2.4) to (6.2.6) with $x_i^2(x^k) = x_{i,k}^2$, $y_i^\ell(x^k) = y_{i,k}^\ell$, $\ell = 1, 2$, and $z_j^2(x^k) = z_{j,k}^2$, for $i \in I$ and $j \in J$.
2. If $F_k(x^{k+1}) = 0$ stop, else for all $i \in I$ and $j \in J$, choose $z_{j,k+1}^2 \in \partial h_j^2(x^{k+1})$, $x_{i,k+1}^2 \in \partial f_i^2(x^{k+1})$ and $y_{i,k+1}^\ell \in \partial g_i^\ell(x^{k+1})$, with $\ell = 1$ if $v_i(x^{k+1}) \geq 0$, and $\ell = 2$ if $v_i(x^{k+1}) < 0$. Set $v_{i,k+1} = v_i(x^{k+1})$, $k = k + 1$ and return to step 1.

To establish the convergence of the sequence $\{v_k\}$, we need the following well known lemma.

Lemma 6.2.4. Let $\{\varepsilon_k\}$ be a sequence of positive reals such that $\sum_{k\geq 0} \varepsilon_k < \infty$, and let $\{u_k\}$ be a sequence such that $u_{k+1} \leq u_k + \varepsilon_k$ for all $k \in \mathbb{N}$. Then $\{u_k\}$ converges to some $\hat{u} \in \mathbb{R} \cup \{-\infty\}$.

Proof. See, e.g., [101, §2.2.1, Lemma 2] and [108, Lemma 2.1] for a more general form of this lemma. □

Under habitual the assumptions used for fractional programming, we will establish the convergence of the sequence $\{v_k\}$. With regard to Theorem 6.2.1 we will show that $\{F_k(x^{k+1})\}$ converges to 0. This will be used as a stopping criterion for Algorithm 7.

We denote and assume

$$\delta := \inf_{x\in X}\min_{i\in I} g_i(x) > 0 \text{ and } \Delta := \sup_{x\in X}\max_{i\in I} g_i(x) < \infty.$$

Proposition 6.2.5. If $\sum_{k\geq 0}\varepsilon_k < \infty$, the sequence $\{v_k\}$ converges to some $\hat{v}$. If $\hat{v}_i = -\infty$ for each $i = 1,\cdots,m$, problem (P) has no weak Pareto optimal solution. Otherwise $\{F_k(x^{k+1})\}$ converges to 0 as k tends to ∞.

Proof. From the definition of x^{k+1} we have

$$\varepsilon_k + F_k(x) \geq F_k(x^{k+1}) \text{ for all } x\in X_k.$$

For $x = x^k$ we get $F_k(x^{k+1}) \leq F_k(x^k) + \varepsilon_k = \varepsilon_k$, where the equality $F_k(x^k) = 0$ follows from Proposition 6.2.1, Item 1, since $F_k(x^k) = F_{x^k}(x^k)$. From Lemma 6.2.3, for all $i\in I$ and all $x\in\mathbb{R}^n$, we have

$$F_k(x) \geq \frac{f_i(x) - v_{i,k}g_i(x)}{\omega_i^k}.$$

Therefore,

$$\begin{aligned}\varepsilon_k \geq F_k(x^{k+1}) &\geq \frac{f_i(x^{k+1}) - v_{i,k}g_i(x^{k+1})}{\omega_i^k}\\ &= \frac{g_i(x^{k+1})}{\omega_i^k}(v_{i,k+1} - v_{i,k}),\end{aligned} \tag{6.2.29}$$

where $v_{i,k+1} = f_i(x^{k+1})/g_i(x^{k+1})$. This implies that $v_{i,k+1} \leq v_{i,k} + \varepsilon_k\bar{\omega}/\delta$, for all $i\in I$, where we used the assumptions $g_i(x^{k+1}) \geq \delta > 0$ and $0 < \omega_i^k \leq \bar{\omega}$ for all $k\in\mathbb{N}$ and $i\in I$. Since $v_{i,k} \geq \bar{v}_i$ and $\sum_{k\geq 0}\varepsilon_k < \infty$, for all $i\in I$, we conclude from Lemma 6.2.4 that the sequence $\{v_{i,k}\}$ converges to some $\hat{v}_i \in \mathbb{R}\cup\{-\infty\}$, for all $i\in I$.

Clearly, if $\hat{v}_i = -\infty$ for each $i = 1,\cdots,m$, then for all $x\in X$ there exists $\hat{k}\in\mathbb{N}$ such that $v(x^k) < v(x)$ for all $k\geq\hat{k}$, so that problem (P) cannot have weak Pareto optimal solution.

Assume now that $\hat{v}_i > -\infty$ for some i. Then from (6.2.29) and the fact that

$v_{i,k} \geq \hat{v}_i$, for $i = 1, \cdots, m$, we get

$$\begin{aligned}\varepsilon_k \geq F_k(x^{k+1}) &\geq \frac{g_i(x^{k+1})}{\omega_i^k}(v_{i,k+1} - v_{i,k}) \\ &= \frac{g_i(x^{k+1})}{\omega_i^k}(v_{i,k+1} - \hat{v}_i) + \frac{g_i(x^{k+1})}{\omega_i^k}(\hat{v}_i - v_{i,k}) \\ &\geq \frac{\delta}{\bar{\omega}}(v_{i,k+1} - \hat{v}_i) + \frac{\Delta}{\underline{\omega}}(\hat{v}_i - v_{i,k}),\end{aligned}$$

from which we conclude that $\{F_k(x^{k+1})\}$ converges to 0. □

Remark 6.2.2. Without the assumptions on δ and Δ we can show from (6.2.29) that it $\hat{v}_i > -\infty$ then the sequence $\{F_k(x^{k+1})/g_i(x^{k+1})\}$ converges to 0 as k tends to ∞.

Remark 6.2.3. If $\sum_{k\geq 0} \varepsilon_k < \infty$, and the set $\{x \in \mathbb{R}^n \mid v_i(x) \leq v_i(x^0) + \sum_{k\geq 0} \varepsilon_k \bar{\omega}/\delta, \, i \in I\}$, is bounded, for the starting point x^0 (wich is the case, e.g., if $v_i(.)$, is inf-compact, for some $i \in I$), then the sequence $\{x^k\}$ is bounded. Indeed, we showed in the last proof that $v_{i,k+1} \leq v_{i,k} + \varepsilon_k \bar{\omega}/\delta$, for all $i \in I$. Therefore, $v_{i,k+1} \leq v_{i,0} + \sum_{k\geq 0} \varepsilon_k \bar{\omega}/\delta$ implying that $x^{k+1} \in \{x \in \mathbb{R}^n \mid v_i(x) \leq v_i(x^0) + \sum_{k\geq 0} \varepsilon_k \bar{\omega}/\delta, \, i \in I\}$.

Now we turn our attention to the convergence of the sequence $\{x^k\}$. Before we are going to state a proposition (see, Proposition 4.4.2, Chapter 4), that we will use to show the convergence of the sequence $\{x^k\}$.

Proposition 6.2.6. For all $k \in \mathbb{N}$, there exists $\alpha^k \in \Sigma$ such that

$$\alpha_0^k \Big(F_k(x) - F_k(x^{k+1}) + \varepsilon_k \Big) + \sum_{j\in J} \alpha_j^k h_{j,k}(x) \geq 0 \quad \text{for all } x \in C. \tag{6.2.30}$$

where $\Sigma = \Big\{ (\alpha_j)_{j\in J\cup\{0\}} \geq 0 \mid \alpha_0 + \sum_{j\in J} \alpha_j = 1 \Big\}$. Moreover,

$$0 \geq \sum_{j\in J} \alpha_j^k h_{j,k}(x^{k+1}) \geq -\alpha_0^k \varepsilon_k \text{ and } 0 \geq \sum_{j\in J} \alpha_j^k h_j(x^k) \geq \alpha_0^k \Big(F_k(x^{k+1}) - \varepsilon_k \Big).$$

So, if $\sum_{k\geq 0} \varepsilon_k < \infty$, the sequences $\Big\{ \sum_{j\in J} \alpha_j^k h_{j,k}(x^{k+1}) \Big\}$ and $\Big\{ \sum_{j\in J} \alpha_j^k h_j(x^k) \Big\}$ converge to 0 as k tends to ∞.

Theorem 6.2.3. Assume that $\sum_{k\geq 0} \varepsilon_k < \infty$ and that the sequence $\{x^k\}$ is bounded. Let $\hat{x}$ be a cluster point of $\{x^k\}$ and $\hat{v}_i = v_i(\hat{x})$. Then for all $i \in I$ and $j \in J$, there exists $\hat{\omega}_i$ a cluster point of $\{\omega_i^k\}$, there exist $\hat{\mu}_i, \hat{\nu}_j \geq 0$, with

$\sum_{i\in I}\hat{\mu}_i + \sum_{j\in J}\hat{\nu}_j = 1$ such that

$$0 \in \sum_{i\in I}\frac{\hat{\mu}_i}{\hat{\omega}_i}\left[\partial f_i^1(\hat{x}) - \partial f_i^2(\hat{x}) - \hat{v}_i\left(\partial g_i^1(\hat{x}) - \partial g_i^2(\hat{x})\right)\right] + \sum_{j\in J}\hat{\nu}_j\left[\partial h_j^1(\hat{x}) - \partial h_j^2(\hat{x})\right] + N_C(\hat{x}),$$

with $\hat{\nu}_j h_j(\hat{x}) = 0$, for all $j \in J$.

Proof. Let $\hat{x}$ be a cluster point of the sequence $\{x^k\}$, and let K be an infinite subset of $\mathbb{N}$ such that the subsequence $\{x^k\}_{k\in K}$ converges to $\hat{x}$. Since $x^k \in X_k \subset X$ and X is closed, we have $\hat{x} \in X$. By the continuity of the functions f_i and g_i we have that $\hat{v}_i = v_i(\hat{x})$. For each $k \in \mathbb{N}$, let $\{\alpha^k\} \subset \Sigma$ as stated in Proposition 6.2.6. Let the sequences $\{x_{i,k}^2\}$, $\{y_{i,k}^\ell\}$, for $i \in I$, $\ell = 1,2$, and $\{z_{j,k}^2\}$, for $j \in J$, be as defined in Algorithm 7. We recall here that these sequences are bounded by the boundedness of $\{x^k\}$, see e.g. [64, Proposition 6.2.2].

Next, for $k \in K$, we consider subsequences of $\{\alpha^k\}$, $\{z_{j,k}^2\}$, for $j \in J$, $\{\omega_i^k\}$, $\{x_{i,k}^2\}$, $\{y_{i,k}^\ell\}$, for $i \in I$, say for $k \in K'$ an infinite subset of $\mathbb{N}$, converging respectively to $\hat{\alpha}$, $\hat{z}_j^2$, $\hat{\omega}_i$, $\hat{x}_i^2$, $\hat{y}_i^\ell$, where $\ell = 1$ if $\hat{v}_i > 0$ and $\ell = 2$ if $\hat{v}_i < 0$. (Notice that $v_{i,k} > 0$ and $\ell = 1$ (resp. $v_{i,k} < 0$ and $\ell = 2$) for k large, when $\hat{v}_i > 0$ (resp. $\hat{v}_i < 0$)). Therefore, $\hat{z}_j^2 \in \partial h_j^2(\hat{x})$, $\hat{x}_i^2 \in \partial f_i^2(\hat{x})$ and $\hat{y}_i^\ell \in \partial g_i^\ell(\hat{x})$, see e.g. [64, Proposition 6.2.1]. With these elements we define the function

$$F_{\hat{x}}(x) = \max_{i\in I}\left[\frac{f_{i,\hat{x}}(x) - \hat{v}_i g_{i,\hat{x}}(x)}{\hat{\omega}_i}\right],$$

where the functions $f_{i,\hat{x}}$, $g_{i,\hat{x}}$ and $h_{j,\hat{x}}$ are defined as in Eqs. (6.2.4) to (6.2.6) with $x_i^2(\hat{x}) = \hat{x}_i^2$, $y_i^\ell(\hat{x}) = \hat{y}_i^\ell$, for $l = 1,2$, $i \in I$, and $z_j^2(\hat{x}) = \hat{z}_j^2$, $j \in J$. In the case $\hat{v}_i = 0$, we ignore the term $\hat{v}_i g_{i,\hat{x}}(x)$ in the definition of $F_{\hat{x}}(x)$.

By invoking Proposition 6.2.6 and passing to the limit in (6.2.30), as k tends to ∞, $k \in K'$, we arrive to

$$\hat{\alpha}_0 F_{\hat{x}}(x) + \sum_{j\in J}\hat{\alpha}_j h_{j,\hat{x}}(x) \geq 0 \quad \text{for all } x \in C.$$

Therefore, for all $x \in X_{\hat{x}}$ we have $\hat{\alpha}_0 F_{\hat{x}}(x) \geq 0$. So, if $\hat{\alpha}_0 \neq 0$ then since $F_{\hat{x}}(\hat{x}) = 0$, we deduce that $\hat{x}$ globally minimizes the convex function $F_{\hat{x}}$ over $X_{\hat{x}}$. Then it suffices to use Theorem 6.2.1 to conclude. Now if $\hat{\alpha}_0 = 0$, we get

$$\sum_{j\in J}\hat{\alpha}_j h_{j,\hat{x}}(x) \geq 0 \quad \text{for all } x \in C. \tag{6.2.31}$$

Since $h_{j,\hat{x}}(\hat{x}) = h_j(\hat{x}) \leq 0$ we obtain $\hat{\alpha}_j h_j(\hat{x}) = 0$ for all $j \in J$. Therefore, $\hat{x}$ minimizes, on C, the function $x \mapsto \sum_{j\in J} \hat{\alpha}_j h_{j,\hat{x}}(x)$. It follows that

$$0 \in \sum_{j\in J} \hat{\alpha}_j \partial h_{j,\hat{x}}(\hat{x}) + N_C(\hat{x}) \subset \sum_{j\in J} \hat{\alpha}_j \left[\partial h_j^1(\hat{x}) - \partial h_j^2(\hat{x})\right] + N_C(\hat{x}).$$

The desired result is fulfilled with $\hat{\mu}_i = 0$, for all $i \in I$, and $\hat{\nu}_j = \hat{\alpha}_j$, for all $j \in J$. □

6.3 Special case of ratios of convex functions

In this section, we give application to vector fractional mathematical programs with ratios of convex functions. For this, consider the problem

$$(FP) \qquad \min_{x\in X} \left[v(x) := \left(\frac{f_1(x)}{g_1(x)}, \frac{f_2(x)}{g_2(x)}, \cdots, \frac{f_m(x)}{g_m(x)} \right) \right],$$

where the functions f_i, g_i, for $i \in I := \{1, 2, ..., m\}$ are convex, defined on $\mathbb{R}^n$, and X is a nonempty and convex subset of $\mathbb{R}^n$. The functions g_i, for all $i \in I$, are assumed to be positive on X.

The parametrized functions defined in the beginning of Section 6.2 takes the form

$$F_y(x) := \max_{i\in I} \left[f_i(x) - v_i(y) g_{i,y}(x) \right], \tag{6.3.1}$$

where

$$g_{i,y}(x) := \begin{cases} g_i(y) + \langle \gamma_i, x - y\rangle & \text{if } v_i(y) > 0 \\ g_i(x) & \text{if } v_i(y) \leq 0 \end{cases} \tag{6.3.2}$$

with some $\gamma_i \in \partial g_i(y)$ and $v_i(y)$ is the i−th component of the vector $v(y)$.

All the previous results remain valid, it suffices to replace f_i^1, g_i^1 respectively by f_i, g_i and f_i^2, g_i^2, h_j^l, for $i \in I$, $j \in J$, $l = 1, 2$ by null functions and C by X, in Section 6.2.

6.3.1 Optimality Conditions

Now we are ready to give optimality conditions for (FP).

Theorem 6.3.1. Let $\hat{x} \in X$ and $\hat{v} = v(\hat{x})$. Then there exist $\hat{\gamma}_i \in \partial g_i(\hat{x})$, $i \in I$, such that $F_{\hat{x}}(x_{\hat{x}}) = 0$, or equivalently, such that $\hat{x}$ minimizes $F_{\hat{x}}$ over X if and only if, there exist $\hat{\mu}_i \geq 0$, $i \in I$, with $\sum_{i \in I} \hat{\mu}_i = 1$ such that

$$0 \in \sum_{i \in I} \hat{\mu}_i \left[\partial f_i(\hat{x}) - \hat{v}_i \partial g_i(\hat{x})\right] + N_X(\hat{x}), \tag{6.3.3}$$

Moreover, if $\hat{v}_i \leq 0$ for all $i \in I$, then $\hat{x}$ is a weak Pareto minimum for (FP).

Proof. For the first assertion, it suffices to use the demonstration of Theorem 6.2.1.

To show the converse, assume that we have (6.3.3). Then, for all $i \in I$, there exist $\hat{\gamma}_i^f \in \partial f_i(\hat{x})$, $\hat{\gamma}_i^g \in \partial g_i(\hat{x})$ such that

$$\sum_{i \in I} \hat{\mu}_i \left(\hat{\gamma}_i^f - \hat{v}_i \hat{\gamma}_i^g\right) \in -N_X(\hat{x}),$$

that is

$$\left\langle \sum_{i \in I} \hat{\mu}_i \left(\hat{\gamma}_i^f - \hat{v}_i \hat{\gamma}_i^g\right), x - \hat{x} \right\rangle \geq 0 \quad \text{for all } x \in X. \tag{6.3.4}$$

Let us define the index sets $\hat{I}_1 := \{i \in I \mid \hat{v}_i > 0\}$ and $\hat{I}_2 := \{i \in I \mid \hat{v}_i \leq 0\}$. If $\hat{v}_i > 0$, i.e. $i \in \hat{I}_1$, then by using the subgradient inequality for f_i, we get $f_i(x) \geq f_i(\hat{x}) + \langle \hat{\gamma}_i^f, x - \hat{x} \rangle$. It follows that

$$f_i(x) - \hat{v}_i \left(g_i(\hat{x}) + \langle \hat{\gamma}_i^g, x - \hat{x} \rangle\right) \geq f_i(\hat{x}) - \hat{v}_i g_i(\hat{x}) + \langle \hat{\gamma}_i^f - \hat{v}_i \hat{\gamma}_i^g, x - \hat{x} \rangle$$

Now if $\hat{v}_i \leq 0$, i.e. $i \in \hat{I}_2$, then $\hat{\gamma}_i^f - \hat{v}_i \hat{\gamma}_i^g \in \partial[f_i - \hat{v}_i g_i](\hat{x})$, and thus

$$f_i(x) - \hat{v}_i g_i(x) \geq f_i(\hat{x}) - \hat{v}_i g_i(\hat{x}) + \langle \hat{\gamma}_i^f - \hat{v}_i \hat{\gamma}_i^g, x - \hat{x} \rangle.$$

By invoking the definition of $F_{\hat{x}}(x)$ and $\hat{v}_i$, we get

$$F_{\hat{x}}(x) \geq \langle \hat{\gamma}_i^f - \hat{v}_i \hat{\gamma}_i^g, x - \hat{x} \rangle \text{ for all } i \in I. \tag{6.3.5}$$

Multiplying both sides of (6.3.5) by $\hat{\mu}_i$, for all $i \in I$, and summing, we obtain, taking into account (6.3.4), that $F_{\hat{x}}(x) \geq 0$ for all $x \in X$. Since $F_{\hat{x}}(\hat{x}) = 0$, we conclude that $F_{\hat{x}}(x_{\hat{x}}) = 0$, where $x_{\hat{x}}$ is a global minimum of $F_{\hat{x}}$ over X, which gives the desired result.

Now we will show that $\hat{x}$ is a weak Pareto minimum for (FP) when $\hat{v}_i \leq 0$, for all $i \in I$. Remark that in this case we have

$$F_{\hat{x}}(x) = \max_{i \in I}[f_i(x) - \hat{v}_i g_i(x)].$$

The fact that $\hat{x}$ minimizes $F_{\hat{x}}$ over X, implies that

$$\begin{aligned} F(\hat{v}) &= \inf_{x\in X}\max_{i\in I}[f_i(x) - \hat{v}_i g_i(x)] \\ &= \max_{i\in I}[f_i(\hat{x}) - \hat{v}_i g_i(\hat{x})] \\ &= F_{\hat{x}}(\hat{x}) \\ &= 0. \end{aligned}$$

By Lemma 6.2.1, $F(\hat{v}) = 0$ implies that $\hat{x}$ is a weak Pareto minimum for (FP). □

To show the next proposition it suffices to use Theorem 6.3.1 and Lemma 6.2.2.
Proposition 6.3.1. If $\hat{v}_i \leq 0$ for all $i \in I$ and $\hat{x}$ is the unique global minimum of $F_{\hat{x}}$ over X, then $\hat{x}$ is a Pareto minimum for (FP).
Corollary 6.3.1. Let $\hat{x} \in X$ and $\hat{v} = v(\hat{x})$. Assume that the functions f_i, g_i, for $i \in I$, are continuously differentiable. Then $F_{\hat{x}}(x_{\hat{x}}) = 0$, or equivalently, such that $\hat{x}$ minimizes $F_{\hat{x}}$ over X, if and only if, there exist $\hat{\mu}_i \geq 0$, $i \in I$, with $\sum_{i\in I} \hat{\mu}_i = 1$, such that

$$\sum_{i\in I} \hat{\mu}_i \left[\nabla f_i(\hat{x}) - \hat{v}_i \nabla g_i(\hat{x})\right] \in -N_X(\hat{x}), \tag{6.3.6}$$

where $\nabla f_i(\hat{x})$, $\nabla g_i(\hat{x})$ and $N_X(\hat{x})$ are respectively, the gradient of f_i, g_i and the normal cone to X at $\hat{x}$. Moreover, if $\hat{v}_i \leq 0$ for all $i \in I$, then $\hat{x}$ is a weak Pareto minimum for (FP). Otherwise, if $\hat{v}_i \leq 0$ for all $i \in I$ and $\hat{x}$ is the unique global minimum of $F_{\hat{x}}$ over X, then $\hat{x}$ is a Pareto minimum for (FP).

Proof. It suffices to notice that for these convex continuously differentiable functions we have $\partial f_i(\hat{x}) = \{\nabla f_i(\hat{x})\}$, $\partial g_i(\hat{x}) = \{\nabla g_i(\hat{x})\}$, for $i \in I$ and then to use Theorem 6.3.1, Proposition 6.3.1. □

Recall that we used the notation $\hat{I}_1 := \{i \in I \mid \hat{v}_i > 0\}$ and $\hat{I}_2 := \{i \in I \mid \hat{v}_i \leq 0\}$.

Under certain assumptions, we will show that KKT criticality conditions in Theorem 6.3.1 are in fact Clarke stationary ones.
Theorem 6.3.2. Let $\hat{x} \in X$ and assume that the functions $-g_i$, for all $i \in \hat{I}_1$, are regular at $\hat{x}$, in the sence of Clarke. Then the conditions of Theorem 6.3.1 imply that, for all $i \in \hat{I}_1$, there exist $\hat{\alpha}_i \geq 0$, and for all $i \in \hat{I}_2$, there exist $\hat{\beta}_i \geq 0$, such that

$$0 \in \sum_{i \in \hat{I}_1} \alpha_i \partial^c v_i(\hat{x}) + \sum_{i \in \hat{I}_2} \hat{\beta}_i \left[\partial f_i(\hat{x}) - \hat{v}_i \partial g_i(\hat{x})\right] + N_X(\hat{x}), \tag{6.3.7}$$

with $\sum_{i \in \hat{I}_1} \alpha_i + \sum_{i \in \hat{I}_2} \hat{\beta}_i \neq 0$, where $\partial^c v_i(\hat{x})$ is the Clarke subdifferential of v_i, $i \in \hat{I}_1$ at $\hat{x}$ and $N_X(\hat{x})$ the normal cone to X at $\hat{x}$.

Furthermore, if $v_i(\hat{x}) > 0$, for all $i \in I$. Then the conditions (6.3.7) imply the Clarke stationarity of $\hat{x}$, that is

$$0 \in \sum_{i \in I} \alpha_i \partial^c v_i(\hat{x}) + N_X(\hat{x}),$$

with $\sum_{i \in I} \alpha_i \neq 0$.

Proof. Let $\hat{x} \in X$, $\hat{v}_i = v_i(\hat{x})$, $i \in I$. We can write (6.3.3) as

$$0 \in \sum_{i \in \hat{I}_1} \hat{\mu}_i \left[\partial f_i(\hat{x}) - \hat{v}_i \partial g_i(\hat{x})\right] + \sum_{i \in \hat{I}_2} \hat{\mu}_i \left[\partial f_i(\hat{x}) - \hat{v}_i \partial g_i(\hat{x})\right] + N_X(\hat{x}).$$

Thus,

$$0 \in \sum_{i \in \hat{I}_1} \hat{\mu}_i \left[\partial f_i(\hat{x}) - \frac{f_i(\hat{x})}{g_i(\hat{x})} \partial g_i(\hat{x})\right] + \sum_{i \in \hat{I}_2} \hat{\mu}_i \left[\partial f_i(\hat{x}) - \hat{v}_i \partial g_i(\hat{x})\right] + N_X(\hat{x}). \tag{6.3.8}$$

It is clear that we can write (6.3.8) as follows

$$0 \in \sum_{i \in \hat{I}_1} \hat{\mu}_i g_i(\hat{x}) \left[\frac{g_i(\hat{x}) \partial f_i(\hat{x}) - f_i(\hat{x}) \partial g_i(\hat{x})}{g_i^2(\hat{x})}\right] + \sum_{i \in \hat{I}_2} \hat{\mu}_i \left[\partial f_i(\hat{x}) - \hat{v}_i \partial g_i(\hat{x})\right] + N_X(\hat{x}). \tag{6.3.9}$$

Since by our assumptions the functions $-g_i$ are regular at $\hat{x}$, $i \in \hat{I}_1$, in the sense of Clarke [30, Definition 2.3.4], and $f_i(\hat{x}) \geq 0$, [30, Proposition 2.3.14] entails that

$$\partial^c \left[\frac{f_i}{g_i}\right](\hat{x}) = \frac{g_i(\hat{x}) \partial f_i(\hat{x}) - f_i(\hat{x}) \partial g_i(\hat{x})}{g_i(\hat{x})^2}.$$

Using this equality in (6.3.9) we get

$$0 \in \sum_{i \in \hat{I}_1} \hat{\mu}_i g_i(\hat{x}) \partial^c \left[\frac{f_i}{g_i}\right](\hat{x}) + \sum_{i \in \hat{I}_2} \hat{\mu}_i \left[\partial f_i(\hat{x}) - \hat{v}_i \partial g_i(\hat{x})\right] + N_X(\hat{x}).$$

Then,

$$0 \in \sum_{i\in\hat{I}_1} \hat{\mu}_i g_i(\hat{x})\partial^c v_i(\hat{x}) + \sum_{i\in\hat{I}_2} \hat{\mu}_i \left[\partial f_i(\hat{x}) - \hat{v}_i \partial g_i(\hat{x})\right] + N_X(\hat{x}).$$

It suffices to set $\hat{\alpha}_i = \hat{\mu}_i g_i(\hat{x})$, for all $i \in \hat{I}_1$, and $\hat{\beta}_i = \hat{\mu}_i$, for all $i \in \hat{I}_2$, to get

$$0 \in \sum_{i\in\hat{I}_1} \hat{\alpha}_i \partial^c v_i(\hat{x}) + \sum_{i\in\hat{I}_2} \hat{\beta}_i \left[\partial f_i(\hat{x}) - \hat{v}_i \partial g_i(\hat{x})\right] + N_X(\hat{x}).$$

Finally, note that since $\sum_{i\in I} \hat{\mu}_i = \sum_{i\in\hat{I}_1} \hat{\mu}_i + \sum_{i\in\hat{I}_2} \hat{\mu}_i = 1$ we have $\sum_{i\in\hat{I}_1} \hat{\alpha}_i + \sum_{i\in\hat{I}_2} \hat{\beta}_i \neq 0$.

If $v_i(\hat{x}) > 0$, for all $i \in I$. It suffices to set $\hat{I}_1 = I$ and $\hat{I}_1 = \emptyset$ in (6.3.7) to conclude. □

The next theorem shows that the weak Pareto, and a fortiori Pareto minimum of (FP) satisfies the optimality conditions of Theorem 6.3.1.

Theorem 6.3.3. Let $\bar{x} \in X$ is a weak Pareto minimum of (FP) and $\bar{v} = v(\bar{x})$. Then there exist, $\bar{\mu}_i \geq 0$, $i \in I$, with $\sum_{i\in I} \bar{\mu}_i = 1$, such that

$$0 \in \sum_{i\in I} \bar{\mu}_i \left[\partial f_i(\bar{x}) - v_i(\bar{x})\partial g_i(\bar{x})\right] + N_X(\bar{x}),$$

where $N_X(\bar{x})$ the normal cone to X at $\bar{x}$. Moreover, if $\bar{v}_i \leq 0$ for all $i \in I$, the conditions are also sufficient for a weak Pareto optimality.

Proof. Let $\bar{x} \in X$ be a weak Pareto minimum of (FP). Then from Proposition 6.2.2, $\bar{x}$ is also a global minimizer on the convex set X of the function $F_{\bar{x}}$ defined conforming to (6.3.1) and (6.3.2). Since $F_{\bar{x}}(\bar{x}) = 0$ (see Proposition 6.2.1, Item 1 with $y = \bar{x}$), then it suffices to use Theorem 6.3.1, with $\hat{x} = \bar{x}$ and $x_{\hat{x}} = \bar{x}$, to conclude. □

Chapter 7

Numerical Tests

This chapter is devoted to numerical experiments to evaluate the efficiency of the algorithms described in Chapter 3, Chapter 4, and give the comparisons between these algorithms in various cases.

7.1 Generalized Fractional Programs

In this section, we present numerical tests to evaluate the efficiency of the algorithm described in Chapter 3. Recall first that we are interested in solving the minimax fractional programming problem

$$(P) \qquad \bar{\lambda} = \inf_{x \in X} \left\{ \lambda(x) := \max_{i \in I} \frac{f_i(x)}{g_i(x)} \right\},$$

We will test our algorithm with linear constraints, for the three situations below:

1. For this case the functions f_i are quadratic convex and the functions g_i are affine, for all $i \in I$. Algorithm 2 coincides in this case with Dinkelbach-type algorithm [35, 36].
2. For this case the functions f_i are affine and the functions g_i are quadratic convex, for all $i \in I$.
3. For this case the functions f_i and g_i are quadratic convex, for all $i \in I$.

We will see throught these tests the effect of the parameters ω_i^k (see Section 3.4) on the speed of convergence and on the best found optimal value.

For this, we will compare two versions of Algorithm 2, which will be referred to as Variant 1 and Variant 2, namely with $\omega_i^k = 1$ and with $\omega_i^k = g_i(x^k)$, respectively. We will also investigate the behavior of these two variants of Algorithm 2 with respect to the starting point.

For this, we test the two variants of our algorithm on randomly generated problems with ten randomly generated starting points for each problem. Then we report the best found values of the objective function, the number of iterations and the total execution times, respectively desinated by λ_∞, #Iter and Time, in the next tables. We will distinguish two sets of problems. The first with $\bar{\lambda}$, the minimal value, positive and the second with $\bar{\lambda}$ negative. For these two cases, Algorithm 2 acts, respectively, as pure dc algorithm and as pure Dinkelbach-type algorithms.

All tests will be performed with $n = 100$ variables, $m = 50$ constraints and $p = 20$ ratios. The stopping criterion in all tests is $|h_k(x^{k+1})| \leq 1.e-8$. In these tests, the same problem is solved with ten starting points x^0 randomly taken in $[0,1]$. The starting point x^0 may be infeasible, provided that $g_i(x^0) \neq 0$ for all $i \in I$.

7.1.1 Problem with positive objective function

We test our algorithm on the problem (P) with the objective functions,

$$f_i(x) = \tfrac{1}{2}x^\top P_i x + a_i^\top x + b_i, \quad g_i(x) = \tfrac{1}{2}x^\top Q_i x + c_i^\top x + d_i,$$

and the constraints set

$$X = \{x \in \mathbb{R}^n \mid Cx \leq b,\, x \geq 0\}.$$

For this first numerical example, the functions f_i are positive on X. Thus, Algorithm 2 acts as a DC algorithm.

The data P_i, a_i, b_i and Q_i, c_i, d_i, $i = 1, \cdots, m$, C and b and are constructed as follows:

- the matrices P_i (resp. Q_i), $i = 1 \cdots, m$, are defined by $P_i := L_i^\top L_i$ (resp. $Q_i := M_i^\top M_i$), where L_i (resp. M_i) are $1 \times n$ matrices with components uniformly drawn from $[-10, 10]$ (resp. $[-1, 1]$),
- each element of the vectors a_i (resp. c_i), for $i = 1, \cdots, m$, is uniformly drawn from $[0, 10]$ (resp. $[0, 1]$). Similarly b_i (resp. d_i), for $i = 1, \cdots, m$, are uniformly drawn from $[10, 100]$ (resp. $[1, 10]$),

- the elements of the $p \times n$ matrix C are uniformly drawn from $[0,1]$. The elments of the $p \times 1$ vector b are uniformly drawn from $[10,100]$.

Type 1 problems

For this type of problems, we put

$$f_i(x) = \frac{1}{2} x^\top P_i x + a_i^\top x + b_i, \quad g_i(x) = c_i^\top x + d_i,$$

and the constraints set

$$X = \{x \in \mathbb{R}^n \mid Cx \leq b, x \geq 0\}.$$

For this type of problems Algorithm 2 under its two variants, coincides with Dinkelbach-type algorithms, resp. given in [35] and [36], since the functions g_i are affine, and gives a global optimal solution. The results of this test are reported in Table 7.1.1.

Table 7.1.1: Type 1 problem ($\bar{\lambda} > 0$)

	Variant 1			Variant 2		
x^0	λ_∞	#Iter	Time	λ_∞	#Iter	Time
1	12.9393	26	20.50	12.9393	16	8.50
2	12.9393	27	18.90	12.9393	98	22.25
3	12.9393	80	42.88	12.9393	14	7.67
4	12.9393	19	17.53	12.9393	12	7.63
5	12.9393	42	26.83	12.9393	38	12.27
6	12.9393	19	18.51	12.9393	28	9.64
7	12.9393	34	23.66	12.9393	6	5.71
8	12.9393	23	20.71	12.9393	19	9.12
9	12.9393	33	20.19	12.9393	46	12.54
10	12.9393	31	23.50	12.9393	12	6.15

Type 2 problems

For this type of problems, we put

$$f_i(x) = a_i^\top x + b_i, \quad g_i(x) = \frac{1}{2}x^\top M_i x + c_i^\top x + d_i,$$

and the constraints set

$$X = \{x \in \mathbb{R}^n \mid Cx \leq b, \, x \geq 0\}.$$

The results of this test are reported in Table 7.1.2.

Table 7.1.2: Type 2 problem ($\bar{\lambda} > 0$)

	Variant 1			Variant 2		
x^0	λ_∞	#Iter	Time	λ_∞	#Iter	Time
1	7.1994	39	62.22	6.9828	17	17.08
2	7.2762	29	51.58	7.0454	14	14.48
3	6.9654	20	33.94	7.6304	8	9.57
4	7.6919	21	39.33	7.2190	11	13.29
5	7.7306	21	40.91	7.5056	13	14.07
6	7.8186	35	78.90	7.9830	9	12.58
7	6.9873	23	59.18	7.2684	11	13.71
8	7.8665	21	40.29	7.8287	15	11.94
9	7.3755	32	80.65	7.7859	7	4.51
10	7.2116	47	116.55	7.4411	11	14.77

Type 3 problems

For this type of problems, we put

$$f_i(x) = \frac{1}{2}x^\top Q_i x + a_i^\top x + b_i, \quad g_i(x) = \frac{1}{2}x^\top M_i x + c_i^\top x + d_i,$$

and the constraints set

$$X = \{x \in \mathbb{R}^n \mid Cx \leq b, \, x \geq 0\}.$$

The results of this test are reported in Table 7.1.3.

Table 7.1.3: Type 3 problem ($\bar{\lambda} > 0$)

	Variant 1			Variant 2		
x^0	λ_∞	#Iter	Time	λ_∞	#Iter	Time
1	9.3384	56	54.24	9.3075	55	14.47
2	9.1983	33	51.71	9.6047	26	12.09
3	9.1026	89	59.59	9.3638	27	10.99
4	9.1026	83	60.34	9.3374	20	10.76
5	10.0617	65	64.00	9.4288	13	7.75
6	9.1026	132	70.75	9.1796	13	7.76
7	9.1884	38	43.90	9.4187	77	19.22
8	10.0617	57	58.01	9.1796	13	9.98
9	9.1884	117	80.41	9.2406	73	18.68
10	9.2673	43	47.24	9.1796	47	15.00

7.1.2 Problem with nonpositive objective function

For this second numerical example, the functions f_i are nonpositive on X. Thus, Algorithm 2 acts as a Dinkelbach-type algorithm. We conserve the same data as in Subsection 7.2.1 except that each element of the vectors a_i, for $i = 1, \cdots, m$, is uniformly drawn from $[-10, 0]$, and the elements b_i, for $i = 1, \cdots, m$, are uniformly drawn from $[-100, -10]$.

This set of tests illustrates the conclusion of Proposition 4.3.1 which says that Algorithm 2 gives global solution when $\hat{\lambda} \leq 0$. We consider only Type 2 and 3 problems, since for Type 1 problems the functions g_i are affine and we have seen that Algorithm 2 gives a global optimal solution.

Type 2 problems

For this type of problems, we put

$$f_i(x) = a_i^\top x + b_i, \quad g_i(x) = \frac{1}{2} x^\top M_i x + c_i^\top x + d_i,$$

and the constraints set

$$X = \{x \in \mathbb{R}^n \mid Cx \leq b, x \geq 0\}.$$

For this type of problems Algorithm 2 under its two variants, coincides with Dinkelbach-type algorithms, resp. given in [35] and [36], and gives a global optimal solution. The results of this test are reported in Table 7.1.4.

Table 7.1.4: Type 2 problem ($\bar{\lambda} < 0$)

	Variant 1			Variant 2		
x^0	λ_∞	#Iter	Time	λ_∞	#Iter	Time
1	-8.9750	22	8.15	-8.9750	14	6.90
2	-8.9750	194	40.23	-8.9750	24	8.99
3	-8.9750	47	13.69	-8.9750	17	8.37
4	-8.9750	21	7.99	-8.9750	25	8.65
5	-8.9750	68	17.54	-8.9750	12	7.50
6	-8.9750	85	19.72	-8.9750	18	8.32
7	-8.9750	24	8.95	-8.9750	30	10.09
6	-8.9750	152	33.31	-8.9750	37	11.13
9	-8.9750	109	24.37	-8.9750	11	6.21
10	-8.9750	33	10.61	-8.9750	25	9.53

Type 3 problems

For this type of problems, we put

$$f_i(x) = \frac{1}{2} x^\top Q_i x + a_i^\top x + b_i, \quad g_i(x) = \frac{1}{2} x^\top M_i x + c_i^\top x + d_i,$$

and the constraints set

$$X = \{x \in \mathbb{R}^n \mid Cx \leq b, x \geq 0\}.$$

For this type of problems Algorithm 2 under its two variants, coincides with Dinkelbach-type algorithms, resp. given in [35] and [36], and give a global optimal solution. The results of this test are reported in Table 7.1.5.

Table 7.1.5: Type 3 problem ($\bar{\lambda} < 0$)

	Variant 1			Variant 2		
x^0	λ_∞	#Iter	Time	λ_∞	#Iter	Time
1	-7.1195	33	40.41	-7.1195	34	18.88
2	-7.1195	38	46.77	-7.1195	37	16.81
3	-7.1195	249	163.50	-7.1195	76	21.61
4	-7.1195	18	41.13	-7.1195	79	27.42
5	-7.1195	46	54.30	-7.1195	40	18.98
6	-7.1195	62	67.04	-7.1195	79	25.74
7	-7.1195	127	93.13	-7.1195	25	11.75
8	-7.1195	78	63.22	-7.1195	52	18.53
9	-7.1195	24	33.53	-7.1195	19	10.95
10	-7.1195	31	38.75	-7.1195	19	11.87

7.1.3 Comparisons on Problems with positive objective function

To end these numerical tests, we will now compare our DC algorithm, denoted by DC-Variant 2 in the next tables, when it acts as a pur DC algorithm, which is the case when $\bar{\lambda} > 0$, with Dinkelbach-type algorithm, denoted by DT-Variant 2 in the next tables, keeping in mind that we have no theoretical results on the convergence of the latter one, which require globally solving the auxiliary programs. We will consider in this comparison, the second variant of each one of them on problems given in Subsection 7.1.1.

Type 2 problems

For this type of problems, we put

$$f_i(x) = a_i^\top x + b_i, \quad g_i(x) = \tfrac{1}{2} x^\top M_i x + c_i^\top x + d_i,$$

and the constraints set

$$X = \{x \in \mathbb{R}^n \mid Cx \leq b,\, x \geq 0\}.$$

The results of this test are reported in Table 7.1.6.

Table 7.1.6: Type 2 problem ($\bar{\lambda} > 0$)

	DT-Variant 2			DC-Variant 2		
x^0	λ_∞	#Iter	Time	λ_∞	#Iter	Time
1	6.2145	8	11.21	6.5939	12	11.65
2	6.3072	8	11.49	6.2270	8	13.11
3	6.7085	8	10.48	6.0536	9	10.54
4	6.5495	9	14.30	6.2841	9	9.29
5	6.1577	7	10.63	6.1553	10	20.49
6	6.6880	12	13.52	5.7755	9	12.82
7	6.9916	7	17.10	6.0276	11	20.85
8	6.1732	7	9.94	7.0904	55	21.84
9	6.6605	11	12.56	5.8670	10	15.24
10	6.7824	30	17.68	5.9070	14	18.89

Type 3 problems

For this type of problems, we put

$$f_i(x) = \tfrac{1}{2} x^\top Q_i x + a_i^\top x + b_i, \quad g_i(x) = \tfrac{1}{2} x^\top M_i x + c_i^\top x + d_i,$$

and the constraints set

$$X = \{x \in \mathbb{R}^n \mid Cx \leq b,\, x \geq 0\}.$$

The results of this test are reported in Table 7.1.7.

Table 7.1.7: Type 3 problem ($\bar{\lambda} > 0$)

	DT-Variant 2			DC-Variant 2		
x^0	λ_∞	#Iter	Time	λ_∞	#Iter	Time
1	10.7729	60	15.90	9.4788	15	10.71
2	10.3339	78	20.48	10.2146	42	14.38
3	10.2691	20	9.10	10.0846	121	26.49
4	10.2242	193	39.44	10.2344	28	9.53
5	10.3911	88	20.36	10.1926	65	17.95
6	10.2450	33	12.03	10.2068	82	19.78
7	10.1306	41	12.32	10.8875	32	10.14
8	10.6128	16	8.34	10.2691	89	20.67
9	10.5886	25	11.36	10.1926	49	14.47
10	10.2213	11	7.78	9.8866	50	14.63

Our algorithm, intended to solve minimax fractional programs whose objective function is the maximum of a finite number of ratios of two convex functions, has been tested on problem with ratios of quadratic convex functions, and on particular instances of this problem, where the numerators or the denominators are affine functions. Our algorithm has two variants (see Section 3.4) and can behave either as a pure DC algorithm or as a pure Dinkelbach-type algorithm, or both. In all tests, we used indifferently feasible or infeasible staring points. In analyzing the results reported in Tables 7.1.1 to 7.1.7, we pointed the following conclusions.

1. The introduction of the parameters ω_i^k has a positive effect on the number of iterations and on the total execution time. Variant 2 appears to be advantageous relatively to variant 1. On the other hand, from the same starting point, each variant may give a different solution. This conlusion is also available for the starting point, i.e., different starting point may give different solutions, which is habitual in dealing with nonconvex problems. These facts are visible from Tables 7.1.2 and 7.1.3.

2. Tables 7.1.1, 7.1.4 and 7.1.5 show that Algorithm 2 find the same value with the ten different points for type 1 problem, and types 2 and 3 when λ is nonpositive on X (Subsection 7.1.2), whereas when λ is negative on X (Subsection 7.1.1), it finds different critical points for types 2 and 3 problem. These last results, reported in Tables 7.1.2 and 7.1.3, confirm the difficulty to solve such problems. The strategy to use different starting points x^0, not necessarily feasible provided that $g_i(x^0) \neq 0$ for all $i \in I$, randomly generated, may be used to find satisfactory solutions.

3. In Tables 7.1.6 and 7.1.7, we reported the results of the comparison of the variant 2 of our algorithm with the efficient variant of the Dinkelbach-type algorithm given in [36]. By analyzing these results, we see that there is no clear tendency on the algorithm which prevails on the other, except that our algorithm finds the smallest value in these tests. The results of Dinkelbach-type algorithm are to be taken with some precaution since there is no theoretical results on the convergence of this algorithm when the auxiliary problems are not globally solvable.

7.2 Generalized Fractional Programs with ratios of Difference of Convex Functions

In this section, we present some numerical tests to evaluate the efficiency of the method described in Chapter 4. Recall first that we are interested in solving the minimax fractional programming problem

$$(P) \qquad \bar{\lambda} = \inf_{x \in X} \left\{ \lambda(x) := \max_{i \in I} \frac{f_i^1(x) - f_i^2(x)}{g_i^1(x) - g_i^2(x)} \right\},$$

where $X = \{x \in C : h_j^1(x) - h_j^2(x) \leq 0,\ j \in J\}$, with $C \subset \mathbb{R}^n$ a nonempty, closed and convex set, I and J two finite index sets, and the functions f_i^ℓ, g_i^ℓ, for $i \in I$, and h_j^ℓ, for $j \in J$ and $\ell = 1, 2$ are defined on $\mathbb{R}^n$ and convex, with $g_i^1 - g_i^2$ positive on X for all $i \in I$.

Remember that in Algorithm 3 we have to find a minimum of the convex program

$$\min_{x\in C}\left[F_k(x) := \max_{i\in I}\frac{f_{i,k}(x)-\lambda_k g_{i,k}(x)}{\omega_i^k}\right]$$
$$h_{j,k}(x)\leq 0,\quad j\in J,$$

where ω_i^k are weights to be adjusted by the user. We will test our algorithm to evaluate its efficiency and to see the effect of the weights and of the starting point on the speed of convergence. We begin by testing our algorithm on the general form of problem (P), once with $\omega_i^k = 1$ and once $\omega_i^k = g_i(x^k)$, then we perform tests with different starting points. We will consider two sets of problems to cover the cases $\lambda_k < 0$ and $\lambda_k \geq 0$.

We will designate by Variant 1 and Variant 2, Algorithm 3 with $\omega_i^k = 1$ and with $\omega_i^k = g_i(x^k)$ for all $k\in\mathbb{N}$, respectively. We will analyse the effect of the parameters ω_i^k on the speed of convergence and on the best found value, and investigate the behavior of these two variants of Algorithm 2 with respect to the starting point. To do so, we test our algorithm on randomly generated problems with ten randomly generated starting points for each problem. Then we report the best found values of the objective function, the number of iterations and the total execution times, respectively designated by λ_∞, #Iter and Time, in the next tables.

All tests will be performed with $n = 50$ variables, $m = 20$ ratios and $p = 30$ constraints. The stopping criterion in all tests is $|F_k(x^{k+1})| \leq 1.\text{e}{-6}$. In these tests, the problems are solved with starting points x^0 randomly taken in $[0,1]$.

7.2.1 Problem Statement

We test our algorithm on the problem (P) with the objective function defined from the functions,

$$f_i^\ell(x) = \tfrac{1}{2}x^\top P_i^\ell x + a_{i,\ell}^\top x + b_i^\ell,\quad g_i^\ell(x) = \tfrac{1}{2}x^\top Q_i^\ell x + c_{i,\ell}^\top x + d_i^\ell,$$

for $i\in I$, $\ell = 1,2$, and the constraints set

$$X = \left\{x\in C \mid h_j^1(x) - h_j^2(x)\leq 0,\ j\in J\right\},$$

where $C = \{x\in\mathbb{R}^n \mid 0\leq x_i\leq 10,\ i = 1,\cdots,n\}$, and the functions h_j^ℓ are given by $h_j^\ell(x) = \frac{1}{2}x^\top C_j^\ell x + \alpha_{j,\ell}^\top x + \beta_j^\ell$ for $j\in J$ and $\ell = 1,2$. The data P_i^ℓ, $a_{i,\ell}$, b_i^ℓ, Q_i^ℓ,

$c_{i,\ell}$, d_i^ℓ for $i \in I$, and C_j^ℓ, $\alpha_{j,\ell}$ and β_j^ℓ for $j \in J$ and $\ell = 1,2$ are constructed as follows:

- the matrices P_i^ℓ for $i \in I$ and $\ell = 1,2$, are defined by $P_i^\ell := L_{i,\ell}^\top L_{i,\ell}$, where $L_{i,\ell}$ are $1\times n$ matrices with components uniformly drawn from $[-10,10]$;
- the matrices Q_i^ℓ for $i \in I$ and $\ell = 1,2$, are defined by $Q_i^\ell := M_{i,\ell}^\top M_{i,\ell}$, where $M_{i,\ell}$ are $1\times n$ matrices with components uniformly drawn from $[-1,1]$;
- each element of the vectors $a_{i,1}$ (resp. $a_{i,2}$) for $i \in I$, is uniformly drawn from $[10,100]$ (resp. $[0,10]$). Similarly, the components of the vectors $c_{i,1}$ (resp. $c_{i,2}$) for $i \in I$, are uniformly drawn from $[10,100]$ (resp. $[-100,-10]$);
- the elements b_i^1 and d_i^1 for $i \in I$, are uniformly drawn from $[10,100]$. The elements b_i^2 (resp. d_i^2) for $i \in I$, are uniformly drawn from $[0,10]$ (resp. $[-100,-10]$);
- the matrices C_j^ℓ for $j \in J$, are defined by $C_j^\ell := N_{j,\ell}^\top N_{j,\ell}$, where $N_{j,\ell}$ are $1\times n$ matrices with components uniformly drawn from $[-1,1]$;
- each element of the vectors $\alpha_{j,1}$ (resp. $\alpha_{j,2}$) for $j \in J$, is uniformly drawn from $[-1,0]$ (resp. $[0,1]$);
- the elements β_j^1 (resp. β_j^2) for $j \in J$, are uniformly drawn from $[-1,0]$ (resp. $[0,1]$).

We will test our algorithm once with $\lambda_k \geq 0$, and once with $\lambda_k < 0$. To cover the first situation we will select the elements of $a_{i,2}$ and b_i^2 for $i \in I$, uniformly from $[-1000,-100]$.

7.2.2 The General Problem

In this case, we test Algorithm 3, under the two variants Variant 1 and Variant 2, on the general form of problem (P). The results of these tests are reported in Table 7.2.1 for the case $\lambda_k < 0$, and in Table 7.2.2 for the case $\lambda_k \geq 0$.

Table 7.2.1: The general problem: case $\lambda_k < 0$.

	Variant 1 ($\omega_i^k = 1$)			Variant 2 ($\omega_i^k = g_i(x^k)$)		
x^0	λ_∞	#Iter	Time	λ_∞	#Iter	Time
1	-0.4980	16	32.2638	-0.4980	9	59.0065
2	-0.6599	17	25.1353	-0.6599	8	50.6853
3	-0.8877	38	67.2881	-0.6192	10	71.7734
4	-0.8877	36	80.4399	-0.6192	10	101.5514
5	-0.4253	23	48.2722	-0.4252	16	115.1117
6	-0.8877	34	71.2498	-0.8875	10	115.0658
7	-0.8877	36	83.9529	-0.8875	9	77.0359
8	-0.5865	18	36.4709	-0.5970	11	90.0671
9	-0.5217	28	52.0755	-0.7176	12	71.8332
10	-0.6875	16	29.7506	-0.0677	10	80.0775

Table 7.2.2: The general problem: case $\lambda_k \geq 0$.

	Variant 1 ($\omega_i^k = 1$)			Variant 2 ($\omega_i^k = g_i(x^k)$)		
x^0	λ_∞	#Iter	Time	λ_∞	#Iter	Time
1	4.1372	16	26.3124	4.3050	13	90.5476
2	3.9755	20	26.6169	4.2021	9	58.1333
3	4.2900	15	24.4941	4.2901	8	55.7973
4	3.9331	29	53.3704	3.9635	9	71.2483
5	4.2900	16	30.0393	3.9310	11	72.7462
6	3.9635	15	19.8178	3.9635	10	66.7486
7	3.9635	15	18.6734	3.9635	9	64.5894
8	3.9635	15	25.7025	3.9635	10	68.2991
9	3.9635	14	18.7335	3.9635	10	67.6574
10	4.2900	15	27.3789	4.2900	8	64.7850

7.2.3 Particular cases

Ratios of convex functions

In this case, we consider problem (P) with several ratios of convex functions, under convex constraints. For this, we will set $P_i^2 = Q_i^2 = 0$ for all $i \in I$, and $C_j^2 = 0$ for all $j \in J$. This is a nonconvex program which was detailled in Chapter 3 and Section 7.1. Notice that in the case $\lambda_k < 0$, Algorithm 3 coincides with Dinkelbach-type algorithm [35, 36]. The results of these tests are reported in Table 7.2.3 for the case $\lambda_k < 0$, and in Table 7.2.4 for the case $\lambda_k \geq 0$.

Table 7.2.3: Multiple ratios convex/convex problem: case $\lambda_k < 0$.

	Variant 1 ($\omega_i^k = 1$)			Variant 2 ($\omega_i^k = g_i(x^k)$)		
x^0	λ_∞	#Iter	Time	λ_∞	#Iter	Time
1	-0.5878	14	32.0848	-0.5878	10	85.4385
2	-0.5878	14	37.5558	-0.5878	10	82.5909
3	-0.5878	15	34.5162	-0.5878	10	84.7764
4	-0.5878	14	32.7398	-0.5878	10	82.9646
5	-0.5878	14	31.2695	-0.5878	10	82.2149
6	-0.5878	14	37.0830	-0.5878	10	92.0408
7	-0.5878	14	33.8775	-0.5878	10	84.3101
8	-0.5878	14	34.5462	-0.5878	10	89.7178
9	-0.5878	14	33.3782	-0.5878	10	84.7858
10	-0.5878	14	32.6360	-0.5878	10	76.5248

Table 7.2.4: Multiple ratios convex/convex problem: case $\lambda_k \geq 0$.

	Variant 1 ($\omega_i^k = 1$)			Variant 2 ($\omega_i^k = g_i(x^k)$)		
x^0	λ_∞	#Iter	Time	λ_∞	#Iter	Time
1	0.4318	14	39.6409	0.4319	9	84.5310
2	0.4318	15	64.8648	0.4319	9	103.9713
3	0.4318	15	51.1473	0.4319	10	113.8662
4	0.4318	15	69.9690	0.4319	10	100.5765
5	0.4318	14	48.8452	0.4319	10	105.7144
6	0.4318	14	53.0442	0.4319	8	91.0156
7	0.4318	14	39.4566	0.4319	9	119.7876
8	0.4318	14	68.2918	0.4319	9	136.2001
9	0.4318	15	54.9799	0.4319	9	110.2204
10	0.4318	15	51.0755	0.4319	9	108.3388

Ratios of convex and concave functions

In this case, we consider problem (P) with several ratios of convex and concave functions, under convex constraints. For this, we will set $P_i^2 = Q_i^1 = 0$ for all $i \in I$, and $C_j^2 = 0$ for all $j \in J$. This a nonconvex program. Remark that in the case $\lambda_k \geq 0$, Algorithm 3 coincides with Dinkelbach-type algorithm [35, 36]. The results of these tests are reported in Table 7.2.5 for the case $\lambda_k < 0$, and in Table 7.2.6 for the case $\lambda_k \geq 0$.

Table 7.2.5: Multiple ratios convex/concave problem: case $\lambda_k < 0$.

	Variant 1 ($\omega_i^k = 1$)			Variant 2 ($\omega_i^k = g_i(x^k)$)		
x^0	λ_∞	#Iter	Time	λ_∞	#Iter	Time
1	-0.5344	17	40.3661	-0.5344	13	115.8067
2	-0.5344	17	40.2339	-0.5344	12	112.7121
3	-0.5344	17	37.4753	-0.5344	13	116.1631
4	-0.5344	17	39.9116	-0.5344	13	119.7059
5	-0.5344	17	38.0231	-0.5344	12	112.9291
6	-0.5344	17	40.6269	-0.5344	12	114.6368
7	-0.5344	17	37.5942	-0.5344	13	116.3712
8	-0.5344	17	39.0809	-0.5344	13	114.6634
9	-0.5344	17	40.3759	-0.5344	13	110.9185
10	-0.5344	17	38.9561	-0.5344	12	108.8777

Table 7.2.6: Multiple ratios convex/concave problem: case $\lambda_k \geq 0$.

	Variant 1 ($\omega_i^k = 1$)			Variant 2 ($\omega_i^k = g_i(x^k)$)		
x^0	λ_∞	#Iter	Time	λ_∞	#Iter	Time
1	0.4344	14	45.0089	0.4344	11	117.1216
2	0.4344	13	42.8282	0.4344	10	102.1632
3	0.4344	13	46.8173	0.4344	11	106.6785
4	0.4344	13	42.3264	0.4344	11	97.6629
5	0.4344	14	42.3255	0.4344	12	106.9286
6	0.4344	13	43.0104	0.4344	11	104.9350
7	0.4344	13	43.0778	0.4344	11	114.0767
8	0.4344	13	47.4300	0.4344	11	121.0118
9	0.4344	14	50.6739	0.4344	12	133.8938
10	0.4344	13	50.4779	0.4344	12	136.3761

Our algorithm was developed to solve minimax fractional programs whose objective function is the maximum of a finite number of ratios of difference of convex (DC) functions. It has been tested on problems with ratios of difference of quadratic convex functions under DC quadratic constraints, and on particular problems, with ratios of convex quadratic functions, and ratios of convex and concave quadratic functions, under quadratic convex constraints. We report these result in Tables 7.2.1 to 7.2.6, from the analysis of which we point out the following remarks.

1. Contrary to the situation of [23], we see from these results that the parameters ω_i^k have a slight positive effect on the number of iterations, but increase the total execution time, thus slowing down the method. This may be due to the evaluation of the weights $g_i(x^k)$.

2. We can see from Tables 7.2.1 and 7.2.2 that from the same starting point, each variant may give a different solution.

3. Tables 7.2.3 to 7.2.6 show that Algorithm 3 find the same value with the ten different points for the minimization of ratios of convex func-

tions, and ratios of convex and concave functions under convex constraints.

Conclusion

In this work, we gave optimality conditions and developed algorithms to find a solution to scalar and vector minimax fractional programs whose objective functions are the maximum of the quotients of difference of convex (DC) functions. We firstly proposed to solve the particular generalized fractional programming problems with ratios of convex functions, and convex constraints, for which we gave necessary optimality conditions, where one of the most important results is Clarke stationarity involving the objective and constraint functions. We then proposed a DC Dinkelbach-type algorithm and established its convergence to a Clarke stationary point. We then considered the more general case of minimax fractional programs with ratios of DC functions, and DC constraints (DC-GFP). We also gave optimality conditions of KKT type and proposed a DC Dinkelbach-type algorithm for these problems. By taking advantage of the convexity property of the associated approximate parametric problems of DC-GFP, another strategy based on the proximal bundle algorithm has been proposed. The proposed methods generate a sequence of approximate solutions that converge to critical points satisfying necessary optimality conditions of KKT type.

Optimality conditions and algorithms are also developed for vector fractional programs with ratios of DC functions, and DC constraints. These results were particularized to vector fractional mathematical programming with ratios of convex functions.

At the end we supported this work with numerical tests which showed the practicability of our algorithms.

Bibliography

[1] S. Addoune, K. Boufi, and A. Roubi. Proximal bundle algorithms for nonlinearly constrained convex minimax fractional programs. *Journal of Optimization Theory and Applications*, 179(1):212–239, 2018.

[2] S. Addoune, M. El Haffari, and A. Roubi. A proximal point algorithm for generalized fractional programs. *Optimization*, 66(9):1495–1517, 2017.

[3] A. Alexandroff. Surfaces represented by the difference of convex functions. In *Doklady Akademii Nauk SSSR (NS)*, volume 72, pages 613–616, 1950.

[4] L. T. H. An, M. T. Belghiti, and P. D. Tao. A new efficient algorithm based on dc programming and dca for clustering. *Journal of Global Optimization*, 37(4):593–608, 2007.

[5] L. T. H. An and P. D. Tao. The dc (difference of convex functions) programming and dca revisited with dc models of real world nonconvex optimization problems. *Annals of operations research*, 133(1):23–46, 2005.

[6] L. T. H. An and P. D. Tao. The dc (difference of convex functions) programming and dca revisited with dc models of real world nonconvex optimization problems. *Annals of operations research*, 133(1):23–46, 2005.

[7] A. Aubry, V. Carotenuto, and A. De Maio. New results on generalized fractional programming problems with toeplitz quadratics. *IEEE Signal Processing Letters*, 23(6):848–852, 2016.

[8] A. Aubry, A. De Maio, Y. Huang, and M. Piezzo. Robust design of radar doppler filters. *IEEE Transactions on Signal Processing*, 64(22):5848–5860, 2016.

[9] A. Aubry, A. De Maio, and M. M. Naghsh. Optimizing radar waveform and doppler filter bank via generalized fractional programming. *IEEE Journal of Selected Topics in Signal Processing*, 9(8):1387–1399, 2015.

[10] A. Aubry, A. De Maio, A. Zappone, M. Razaviyayn, and Z. Q. Luo. A new sequential optimization procedure and its applications to resource allocation for wireless systems. *IEEE Transactions on Signal Processing*, 66(24):6518–6533, 2018.

[11] A. Bagirov, N. Karmitsa, and M. M. Mäkelä. *Introduction to Nonsmooth Optimization: theory, practice and software*, volume 12. Springer, 2014.

[12] I. Barrodale, M. J. D. Powell, and F. D. K. Roberts. The differential correction algorithm for rational _-approximation. *SIAM Journal on Numerical Analysis*, 9(3):493–504, 1972.

[13] A. I. Barros, J. B. G. Frenk, S. Schaible, and S. Zhang. A new algorithm for generalized fractional programs. *Mathematical Programming*, 72(2):147–175, 1996.

[14] A. I. Barros, J. B. G. Frenk, S. Schaible, and S. Zhang. Using duality to solve generalized fractional programming problems. *Journal of Global Optimization*, 8(2):139–170, 1996.

[15] C. R. Bector, S. Chandra, and M. K. Bector. Generalized fractional programming duality: a parametric approach. *Journal of Optimization Theory and Applications*, 60(2):243–260, 1989.

[16] C. R. Bector, S. Chandra, and I. Husain. Optimality conditions and duality in subdifferentiable multiobjective fractional programming. *Journal of Optimization theory and Applications*, 79(1):105–125, 1993.

[17] J. C. Bernard and J. A. Ferland. Convergence of interval-type algorithms for generalized fractional programming. *Mathematical Programming*, 43(1):349–363, 1989.

[18] D. Bhati, P. Singh, and R. Arya. A taxonomy and review of the multiobjective fractional programming (mofp) problems. *International Journal of Applied and Computational Mathematics*, 3(3):2695–2717, 2017.

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

[20] T. Bonnesen and W. Fenchel. Konvexe körper. *New York*, 1948.

[21] H. Boualam and A. Roubi. Dual algorithms based on the proximal bundle method for solving convex minimax fractional programs. *Journal of Industrial & Management Optimization*, 15(4):1897–1920, 2019.

[22] H. Boualam and A. Roubi. Proximal bundle methods based on approximate subgradients for solving lagrangian duals of minimax fractional programs. *Journal of Global Optimization*, 74(2):255–284, 2019.

[23] K. Boufi, M. El Haffari, and A. Roubi. Optimality conditions and a method of centers for minimax fractional programs with difference of convex functions. *Journal of Optimization Theory and Applications*, 187(1):105–132, 2020.

[24] K. Boufi and A. Roubi. Dual method of centers for solving generalized fractional programs. *Journal of Global Optimization*, 69(2):387–426, 2017.

[25] K. Boufi and A. Roubi. Duality results and dual bundle methods based on the dual method of centers for minimax fractional programs. *SIAM Journal on Optimization*, 29(2):1578–1602, 2019.

[26] K. Boufi and A. Roubi. Prox-regularization of the dual method of centers for generalized fractional programs. *Optimization Methods and Software*, 34(3):515–545, 2019.

[27] S. Chandra, B. D. Craven, and B. Mond. Vector-valued lagrangian and multiobjective fractional programming duality. *Numerical Functional Analysis and Optimization*, 11(3-4):239–254, 1990.

[28] X. Chen. Optimality and duality for the multiobjective fractional programming with the generalized (f, ρ) convexity. *Journal of mathematical analysis and applications*, 273(1):190–205, 2002.

[29] S. H. Chu. *Optimal Resources Allocation for a Cognitive Network*. PhD thesis, National Cheng Kung University, Taiwan, 2009.

[30] F. H Clarke. *Optimization and nonsmooth analysis*. Wiley-Interscience, New York, 1983.

[31] R. Correa and C. Lemaréchal. Convergence of some algorithms for convex minimization. *Mathematical Programming*, 62(1):261–275, 1993.

[32] B. D. Craven. *Fractional programming*, volume 4. Heldermann, 1988.

[33] J. P. Crouzeix and J. A. Ferland. Algorithms for generalized fractional programming. *Mathematical Programming*, 52(1):191–207, 1991.

[34] J. P. Crouzeix, J. A. Ferland, and S. Schaible. Duality in generalized linear fractional programming. *Mathematical Programming*, 27(3):342–354, 1983.

[35] J. P. Crouzeix, J. A. Ferland, and S. Schaible. An algorithm for generalized fractional programs. *Journal of Optimization Theory and Applications*, 47(1):35–49, 1985.

[36] J. P. Crouzeix, J. A. Ferland, and S. Schaible. A note on an algorithm for generalized fractional programs. *Journal of Optimization Theory and Applications*, 50(1):183–187, 1986.

[37] J. P. Crouzeix, J. A. Ferland, and H. Van Nguyen. Revisiting dinkelbach-type algorithms for generalized fractional programs. *Opsearch*, 45(2):97–110, 2008.

[38] T. P. Dinh. Eléments homoduaux relatifs à un couple de normes (ϕ, ψ). Applications au calcul de S$\phi\psi$(A). Technical report, 1975.

[39] T. P. Dinh. Calcul du maximum d'une forme quadratique définie positive sur la boule unitée de la norme du max. Technical report, 1976.

[40] W. Dinkelbach. On nonlinear fractional programming. *Management science*, 13(7):492–498, 1967.

[41] J. Dutta and C. Y. Kaya. A new scalarization and numerical method for constructing the weak pareto front of multi-objective optimization problems. *Optimization*, 60(8-9):1091–1104, 2011.

[42] M. El Haffari and A. Roubi. Convergence of a proximal algorithm for solving the dual of a generalized fractional program. *RAIRO-Operations Research*, 51(4):985–1004, 2017.

[43] M. El Haffari and A. Roubi. Prox-dual regularization algorithm for generalized fractional programs. *Journal of Industrial & Management Optimization*, 13(4):1991–2013, 2017.

[44] J. E. Falk. Maximization of signal-to-noise ratio in an optical filter. *SIAM Journal on Applied Mathematics*, 17(3):582–592, 1969.

[45] K. Fan. Minimax theorems. *Proceedings of the National Academy of Sciences of the United States of America*, 39(1):42, 1953.

[46] N. S. Fazzio and M. L. Schuverdt. Convergence analysis of a non-monotone projected gradient method for multiobjective optimization problems. *Optimization Letters*, 13(6):1365–1379, 2019.

[47] J. B. G Frenk and S. Schaible. Fractional programming. erim report series reference no. In *Fractional Programming. ERIM Report Series Reference No*. ERS-2004-074-LIS. http://ssrn. com/abstract=595012, 2004.

[48] M. Fukushima. A descent algorithm for nonsmooth convex optimization. *Mathematical Programming*, 30(2):163–175, 1984.

[49] N. Gadhi and A. Jawhar. Necessary optimality conditions for a set-valued fractional extremal programming problem under inclusion constraints. *Journal of Global Optimization*, 56(2):489–501, 2013.

[50] N. Gadhi, M. Laghdir, and A. Metrane. Optimality conditions for dc vector optimization problems under reverse convex constraints. *Journal of Global Optimization*, 33(4):527–540, 2005.

[51] N. Gadhi and A. Metrane. Sufficient optimality condition for vector optimization problems under dc data. *Journal of Global Optimization*, 28(1):55–66, 2004.

[52] N. A. Gadhi, K. Hamdaoui, M. El Idrissi, and F. Z. Rahou. Necessary optimality conditions for a fractional multiobjective optimization problem. *RAIRO-Operations Research*, 55:S1037–S1049, 2021.

[53] M. Ganji and M. Saraj. Solving multi objective linear fractional programming problem under uncertainty via robust optimization approach. *Journal of Informatics and Mathematical Sciences*, 11(2):115–123, 2019.

[54] Y. Gao, X. Yang, and K. L. Teo. Optimality conditions for approximate solutions of vector optimization problems. *Journal of Industrial & Management Optimization*, 7(2):483–496, 2011.

[55] M. Gaudioso, G. Giallombardo, G. Miglionico, and A. M. Bagirov. Minimizing nonsmooth dc functions via successive dc piecewise-affine approximations. *Journal of Global Optimization*, 71(1):37–55, 2018.

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

[57] A. Ghazi and A. Roubi. A dc approach for minimax fractional optimization programs with ratios of convex functions. *Optimization Methods and Software*, pages 1–19, 2020.

[58] A. Ghazi and A. Roubi. Optimality conditions and dc-dinkelbach-type algorithm for generalized fractional programs with ratios of difference of convex functions. *Optimization Letters*, 15(7):2351–2375, 2021.

[59] X. L. Guo and S. J. Li. Optimality conditions for vector optimization problems with difference of convex maps. *Journal of Optimization Theory and Applications*, 162(3):821–844, 2014.

[60] P. Hartman. On functions representable as a difference of convex functions. *Pacific Journal of Mathematics*, 9(3):707–713, 1959.

[61] M. A. Hejazi and S. Nobakhtian. Optimality conditions for multiobjective fractional programming, via convexificators. *Journal of Industrial & Management Optimization*, 16(2):623–631, 2020.

[62] J. B. Hiriart-Urruty. Generalized differentiability/duality and optimization for problems dealing with differences of convex functions. In *Convexity and duality in optimization*, pages 37–70. Springer, 1985.

[63] J. B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms II*. Springer-Verlag, 1993.

[64] J.B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms*. Springer-Verlag, 1993.

[65] T. T. Hue, J. J. Strodiot, and V. H. Nguyen. Convergence of the approximate auxiliary problem method for solving generalized variational inequalities. *Journal of optimization theory and applications*, 121(1):119–145, 2004.

[66] R. Jagannathan. On projective representations of finite abelian groups. In *Number Theory*, pages 130–139. Springer, 1985.

[67] R. Jagannathan and S. Schaible. Duality in generalized fractional programming via farkas' lemma. *Journal of optimization theory and applications*, 41(3):417–424, 1983.

[68] P. Kanniappan. Necessary conditions for optimality of nondifferentiable convex multiobjective programming. *Journal of Optimization Theory and Applications*, 40(2):167–174, 1983.

[69] D. S. Kim, S. J. Kim, and M. H. Kim. Optimality and duality for a class of nondifferentiable multiobjective fractional programming problems. *Journal of Optimization Theory and Applications*, 129(1):131–146, 2006.

[70] K. C. Kiwiel. *Methods of Descent for Nondifferentiable Optimization, Lecture Notes in Mathematics*. Springer-Verlag, B., 1985.

[71] K. C. Kiwiel. Proximity Control in Bundle Methods for Convex Nondifferentiable Minimization. *Math. Program.*, 46:105–122, 1990.

[72] H. Kuk, G. M. Lee, and T. Tanino. Optimality and duality for nonsmooth multiobjective fractional programming with generalized invexity. *Journal of Mathematical Analysis and Applications*, 262(1):365–375, 2001.

[73] M. Laghdir. Optimality conditions in dc-constrained optimization. *Acta Math. Vietnam*, 30(2):169–179, 2005.

[74] N. D. Lawrence and G. Sanguinetti. Matching kernels through kullback-leibler divergence minimization. *Technical Report CS-04-12, The University of Sheffield, Department of Computer Science*, 2004.

[75] T. H. Le An and P. Dinh Tao. Solving a class of linearly constrained indefinite quadratic problems by dc algorithms. *Journal of global optimization*, 11(3):253–285, 1997.

[76] H. Lê Thi. *Analyse numérique des algorithmes de l'optimisation DC. Approches locale et globale. Codes et simulations numériques en grande dimension. Applications*. PhD thesis, Rouen, 1994.

[77] H. Le Thi. Contribution a l'optimisation non convexe et l'optimisation globale: Théorie, algorithmes et applications. *Habilitation à Diriger des recherches, Université de Rouen*, 1997.

[78] H. Le Thi and T. Pham Dinh. Minimizing the morse potential energy function by a DC programming approach. Technical report, Tech. rep., LITA, University of Lorraine, 2012.

[79] H. A. Le Thi and T. Pham Dinh. DC programming in communication systems: challenging problems and methods. *Vietnam journal of computer science*, 1(1):15–28, 2014.

[80] H. A. Le Thi and T. Pham Dinh. DC programming and DCA: thirty years of developments. *Math. Program., Ser. B*, 15:137–161, 2015.

[81] H. A. Le Thi and T. Pham Dinh. DC programming and DCA: thirty years of developments. *Mathematical Programming*, 169(1):5–68, 2018.

[82] T. Le Thi, H.and Pham Dinh. A two phases DCA based algorithm for solving the lennard–jones problem. Technical report, Tech. rep., LITA, University of Metz, 2011.

[83] C. Lemaréchal. Bundle Methods in Nonsmooth Optimization. *C. Lemaréchal and R. Mifflin, eds., Nonsmooth Optimization (Pergamon Press, Oxford)*, 1978.

[84] C. Lemaréchal. Constructing bundle methods for convex optimization. In *North-Holland Mathematics Studies*, volume 129, pages 201–240. Elsevier, 1986.

[85] Z. A. Liang, H. X. Huang, and P. M. Pardalos. Optimality conditions and duality for a class of nonlinear fractional programming problems. *Journal of Optimization Theory and Applications*, 110(3):611–619, 2001.

[86] Z. A. Liang, H. X. Huang, and P. M. Pardalos. Optimality conditions and duality for a class of nonlinear fractional programming problems. *Journal of Optimization Theory and Applications*, 110(3):611–619, 2001.

[87] Z. A. Liang, H. X. Huang, and P. M. Pardalos. Efficiency conditions and duality for a class of multiobjective fractional programming problems. *Journal of Global Optimization*, 27(4):447–471, 2003.

[88] J. C. Liu. Optimality and duality for multiobjective fractional programming involving nonsmooth pseudoinvex functions. *Optimization*, 37(1):27–39, 1996.

[89] X. J. Long, N. J. Huang, and Z. B. Liu. Optimality conditions, duality and saddle points for nondifferentiable multiobjective fractional programs. *Journal of Industrial & Management Optimization*, 4(2):287–298, 2008.

[90] M. Mäkelä. Survey of bundle methods for nonsmooth optimization. *Optimization methods and software*, 17(1):1–29, 2002.

[91] R. Mifflin. An algorithm for constrained optimization with semismooth functions. *Mathematics of Operations Research*, 2(2):191–207, 1977.

[92] R. Mifflin. A modification and an extension of lemaréchal's algorithm for nonsmooth minimization. In *Nondifferential and variational techniques in optimization*, pages 77–90. Springer, 1982.

[93] B. Mishra and S.R. Singh. Linear fractional programming procedure for multi objective linear programming problem in agricultural system. *International Journal of Computer Applications*, 61(20), 2013.

[94] B. S. Mordukhovich. The extremal principle and its applications to optimization and economics. In *Optimization and related topics*, pages 343–369. Springer, 2001.

[95] R. N. Mukherjee. Generalized convex duality for multiobjective fractional programs. *Journal of mathematical analysis and applications*, 162(2):309–316, 1991.

[96] A. Nagih and G. Plateau. Problèmes fractionnaires: tour d'horizon sur les applications et méthodes de résolution. *RAIRO-Operations Research*, 33(4):383–419, 1999.

[97] J. Neumann. A model of general economic equilibrium. *Review of Economic Studies*, 13:1–9, 1945.

[98] T. P. Nguyen. *Techniques d'optimisation en traitement d'image et vision par ordinateur et en transport logistique*. PhD thesis, Université Paul Verlaine-Metz, 2007.

[99] H. Nikaidô et al. On von neumann's minimax theorem. *Pacific J. Math*, 4(1):65–72, 1954.

[100] S. Nobakhtian. Optimality and duality for nonsmooth multiobjective fractional programming with mixed constraints. *Journal of Global Optimization*, 41(1):103–115, 2008.

[101] B. T. Polyak. Introduction to optimization. translations series in mathematics and engineering. optimization software. *Inc., Publications Division, New York*, 1987.

[102] F. A. Pramy, M. A. Islam, et al. Determining efficient solutions of multi-objective linear fractional programming problems and application. *Open Journal of Optimization*, 6(04):164–175, 2017.

[103] M. Razaviyayn, M. Hong, and Z. Q. Luo. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM Journal on Optimization*, 23(2):1126–1153, 2013.

[104] B. Ricceri. On a minimax theorem: an improvement, a new proof and an overview of its applications. *Minimax Theory and its Applications*, 1:099–152, 2017.

[105] R. T. Rockafellar. Convex analysis. In *Convex analysis*. Princeton University Press, Princeton, N. J., 1970.

[106] R. T. Rockafellar. Convex analysis. In *Convex analysis*. Princeton University Press, Princeton, N. J., 1971.

[107] A. Roubi. Method of centers for generalized fractional programming. *Journal of Optimization Theory and Applications*, 107(1):123–143, 2000.

[108] A. Roubi. Convergence of prox-regularization methods for generalized fractional programming. *RAIRO-Operations Research*, 36(1):73–94, 2002.

[109] P. Ruíz-Canales and A. Rufián-Lizana. A characterization of weakly efficient points. *Mathematical Programming*, 68(1):205–212, 1995.

[110] G. Salmon, J. J. Strodiot, and V. H. Nguyen. A bundle method for solving variational inequalities. *SIAM Journal on Optimization*, 14(3):869–893, 2004.

[111] S. Schaible. Fractional programming: applications and algorithms. *European Journal of Operational Research*, 7(2):111–120, 1981.

[112] S. Schaible. Bibliography in fractional programming. *Zeitschrift für Operations Research*, 26(1):211–241, 1982.

[113] S. Schaible. Multi-ratio fractional programing—a survey. In *Optimization, Parallel Processing and Applications*, pages 57–66. Springer, 1988.

[114] S. Schaible. Fractional programming. In *Handbook of global optimization*, pages 495–608. Springer, 1995.

[115] S. Schaible and J. Shi. Recent developments in fractional programming: single-ratio and max-min case. *Nonlinear analysis and convex analysis*, pages 493–506, 2004.

[116] H. Schramm and J. Zowe. A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results. *SIAM journal on optimization*, 2(1):121–152, 1992.

[117] M. Sion. On general minimax theorems. *Pacific Journal of mathematics*, 8(1):171–176, 1958.

[118] I. M. Stancu-Minasian. A sixth bibliography of fractional programming. *Optimization*, 55(4):405–428, 2006.

[119] I. M. Stancu-Minasian. A seventh bibliography of fractional programming. *Adv. Model. Optim*, 15(2):309–386, 2013.

[120] I. M. Stancu-Minasian. A eighth bibliography of fractional programming. *Optimization*, 66(3):439–470, 2017.

[121] I. M. Stancu-Minasian. A ninth bibliography of fractional programming, 2019.

[122] J. J. Strodiot, J. P. Crouzeix, J. A. Ferland, and V. H. Nguyen. An inexact proximal point method for solving generalized fractional programs. *Journal of Global Optimization*, 42(1):121–138, 2008.

[123] A. Taa. Optimality conditions for vector optimization problems of a difference of convex mappings. *Journal of Global optimization*, 31(3):421–436, 2005.

[124] P. D. Tao and S. Bernoussi. Algorithms for solving a class of nonconvex optimization problems. methods of subgradients. In *North-Holland Mathematics Studies*, volume 129, pages 249–271. Elsevier, 1986.

[125] J. F. Toland. Duality in nonconvex optimization. *Journal of Mathematical Analysis and Applications*, 66(2):399–415, 1978.

[126] H. Tuy. *Convex Analysis and Global Optimization*. Kluwer Academic Publishers, Dordrescht, first edition, 1998.

[127] J. B. H. Urruty. Generalized differentiability, duality and optimization for problem dealing with differences of convex functions. lecture notes in economics and mathematical systems. *Heidelberg, Springer-Verlag*, 256:260–277, 1985.

[128] J. B. H. Urruty. Conditions nécessaires et suffisantes d'optimalité globale en optimisation de différences de deux fonctions convexes. *CR Acad. Sci. Paris Ser. I Math*, 309:459–462, 1989.

[129] T. Weir. A dual for a multiple objective fractional programming problem. *Journal of Information and Optimization Sciences*, 7(3):261–269, 1986.

[130] Y. Ying, K. Huang, and C. Campbell. Enhanced protein fold recognition through a novel data integration approach. *BMC bioinformatics*, 10(1):1–18, 2009.

[131] Q. Zhang and T. J. Chin. Coresets for triangulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(9):2095–2108, 2017.

[132] Q. Zhang, T. J. Chin, and H. M. Le. A fast resection-intersection method for the known rotation problem. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3012–3021, 2018.