UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

NON-SMOOTH AND VARIATIONAL ANALYSIS OF OPTIMIZATION PROBLEMS AND MULTI-LEADER-FOLLOWER GAMES

TESIS PARA OPTAR AL GRADO DE DOCTOR EN
CIENCIAS DE LA INGENIERÍA MENCIÓN MODELACIÓN MATEMÁTICA
EN COTUTELA CON UNIVERSIDAD DE PERPIGNAN VIA DOMITIA

ANTON KRISTOFFER SVENSSON GRAAN

PROFESOR GUÍA:
DIDIER AUSSEL
PROFESOR GUÍA:
RAFAEL CORREA FONTECILLA
PROFESOR CO-GUÍA:
ABDERRAHIM HANTOUTE

MIEMBROS DE LA COMISIÓN:
STEPHAN DEMPE
ALEJANDRO JOFRÉ
MARCO ANTONIO LÓPEZ
ARIS DANIILIDIS
RALPH TYRREL ROCKAFELLAR

Este trabajo ha sido parcialmente financiado por CMM-Conicyt PIA AFB170001

SANTIAGO DE CHILE
2020

Resumen

La presente tesis se enmarca en la teoría de la optimización no suave y trata con problemas de la teoría de juegos. La tesis está dividida en cuatro partes, de las cuales en la primera se presenta la introducción, la notación y algunos resultados preliminares. En la segunda parte se discute sobre reglas de cálculo subdiferencial en espacios generales y se dan fórmulas nuevas tanto en el caso convexo como no convexo. El enfoque consiste en dar reglas de cálculo o de optimalidad aproximadas evitando así la necesidad de condiciones de calificación. En la tercera parte se discute sobre los juegos Multi-Leader-Follower. Se prueba un resultado de existencia de equilibrios en el caso de un solo leader optimista y se extienden resultados respecto a la relación del problema original con la reformulación obtenida de reemplazar el nivel inferior por la concatenación de sus condiciones tipo KKT. Por último, en la cuarta parte se estudian los problemas de quasi-equilibrio, los cuales permiten estudiar la existencia de soluciones de problemas de equilibrio de Nash y desigualdades quasi-variacionales de manera abstracta. Se prueban nuevos resultados de existencia que relajan algunas de las hipótesis estándares.

Abstract

This thesis is within the framework of optimization and deals with non-smooth optimization and with some problems of game theory. It is divided into four parts. In the first introductory part, give the context and some preliminary results. In the second part we discuss about subdifferential calculus rules in general spaces providing of some improved formulas in both the convex and the non-convex cases. Here the focus is on approximate or fuzzy calculus rules and optimality conditions, for which no qualification conditions are required. In the third part, we discuss about the so-called Multi-Leader-Follower Games. We give an existence result for the case of a single optimistic leader and multiple followers, and extend some results concerning the relation between the original problem with the reformulation obtained by replacing the followers' problem by the concatenation of their KKT conditions. Finally, in the fourth part we study quasi-equilibrium problems which are a general formulation for studying Nash equilibrium problems and quasi-variational inequalities. We provide some new existence results that relax some of the standard hypotheses.

Résumé

Cette thèse, dont le cadre général est l'optimisation, traite de problèmes d'optimisation non-lisse et de problèmes de théorie des jeux. Elle est constituée de quatre parties. Dans la première, nous présentons le contexte et l'introduction. Dans la deuxième partie, nous discutons quelques règles de calcul sous-différentiel dans des espaces généraux, et présentons notamment certaines formules plus fortes que l'état de l'art, autant dans le cas convexe que dans le cas non convexe. L'accent est mis sur les règles de calcul et conditions d'optimalité approchées et "fuzzy", pour lesquelles aucune condition de qualification n'est requise. Dans la troisième partie, nous considérons des jeux bi-niveaux à plusieurs meneurs et plusieurs suiveurs. Après quelques résultats d'existence dans le cas d'un seul meneur optimiste et dans le cas de plusieurs meneurs, nous étendons des résultats existants concernant la relation entre le problème bi-niveau original et sa reformulation obtenue grâce au remplacement des problèmes des suiveurs par la concaténation de leurs conditions d'optimalité (KKT). Finalement, dans la quatrième partie, nous abordons quelques problèmes de quasi-équilibres, qui sont une généralisation des problèmes d'équilibre de Nash et des inégalités quasi-variationnelles. Nous prouvons ainsi de nouveaux résultats d'existence qui permettent de relâcher les hypothèses standard.

Preface

The present thesis gathers the different results obtained during the development of the PhD of Anton Svensson, since August 2015, in the mode of a cotutelle between Universidad de Chile and Université de Perpignan.

The thesis has been supported by grant CONICYT-PFCHA/Doctorado Nacional/2018 N21180645, Fondecyt Regular N° 1190012, Proyecto/Grant PIA AFB-170001, and FMJH Program Gaspard Monge in Optimisation and operations research by EDF.

The thesis covers both theoretic and applied results, which deal with the quantitative and qualitative study of the solutions to structured optimization problems, including subdifferential calculus, optimization with or without constraint qualification, equilibrium problems, variational inequalities, bilevel programming problems and multi-leader-follower games. We provide new results as well as improvements of some existing ones. The underlying framework of our theoretic results is mainly infinite-dimensional, covering general frameworks such as locally convex spaces, Banach spaces, or Asplund spaces, and bringing novelty even in finite dimensions. The study of bilevel programming problems and multi-leader-follower games is developed in finite-dimensional setting for the aim of simplifying the presentation of the results.

Acknowledgements

Hereby, I would like to express my deepest gratitude to all who contributed to the development of this thesis.

Firstly, thanks to my brother Pierre Svensson who, at the age of thirteen, and myself being only eight years old, attracted my attention to mathematics for the first time, telling me about the amazing unknown number x and some techniques for solving equations.

My sincere acknowledgement to my high school math teacher Gloria Lizama, who continued inspiring me on this interesting path, providing challenging and long lists of problems to solve.

At the university, Roberto Cominetti, Jaime San Martin and many other professors finished the work of turning/leading me into maths; great thanks to them, too.

Already/Once inside the math world, on/during my master thesis, my supervisor Abderrahim Hantoute encouraged me to do a PhD, which seemed unachievable to me at first, but after careful consideration I accepted the challenge and agreed on doing it in the mode of a cotutelle, supervised also by Didier Aussel from Université de Perpignan.

Many thanks to both of my supervisors for all time disposed for discussion about math and life and for all help provided with administrative stuff and in particular, for welcoming me and my wife in Perpignan and taking care of our accommodation issues the four stays. Thanks also for their aid with the revision of the manuscript and thanks to the referees for their reading and suggestions which improved the quality of the manuscript.

Thanks to my wife Tamara Venegas for her constant support on/during the whole process, travelling and staying with me in France. Thanks also to my family.

Thanks also to my friends and colleagues Hugo Maturana, Hugo Carrillo, Javier Ramirez, Roberto Morales, John Cotrina, Evelyn Cuevas, Rodrigo Quezada, Sebastián Tapia, Christopher Maulén, Ariel Perez, Miroslav Pistek, Antoine Haddon, Léonard von Niederhäusern, David Salas, Nouha Dkhili, Matthieu Marechal, Pedro Pérez, Kien van Cao, just to name a few.

I am also grateful to Laboratoire PROMES, University of Perpignan and the department of mathematical engineering of University of Chile for providing me of nice environments to work on my thesis and lots of people there who helped with different things along the way.

Finally, I want to thank God, who provided me of health, intelligence, happiness, peace, and of my beloved child to be born more or less a week after the date of my defense.

List of publications

As a result of the present thesis the following publications were possible.

- A. Hantoute, A. Svensson (2017) *A general representation of delta-normal sets to sublevels of convex functions,* Set Valued and Variational Analysis, 25(4): 651-678,

  doi.org/10.1007/s11228-017-0460-5

- D. Aussel, A. Svensson (2018) *Is pessimistic bilevel programming a special case of a mathematical program with complementarity constraints?*, Journal of Optimization Theory and Applications, 181(2): 504-520,

  doi.org/10.1007/s10957-018-01467-7

- D. Aussel, A. Svensson (2018) *Some remarks about existence of equilibria and the validity of the EPCC reformulation for multi-leader-follower games*, Journal of Nonlinear and Convex Analysis, Volume 19(7): 1141-1162,

  www.ybook.co.jp/online2/opjnca/vol19/p1141.html

- D. Aussel, A. Svensson (2019) *Towards tractable constraint qualifications for parametric optimisation problems and applications to Generalised Nash Equilibrium Problems*, Journal of Optimization Theory and Applications, 182(1): 404-416,

  doi.org/10.1007/s10957-019-01529-4

- D. Aussel, A. Svensson, A short state of the art on Multi-Leader-Follower Games, in *Bilevel optimization: advances and next challenges*, Springer, accepted

The following three works are in the final phase of submission:

- A. Hantoute, A. Svensson, Weak Fuzzy Calculus for Non-smooth Optimization and Applications to Bilevel Problems
- J. Cotrina, A. Hantoute, A. Svensson, A Coerciveness Condition for Quasi-Equilibrium Problems
- J. Cotrina, A. Svensson, Finite Intersection Property for Bi-functions and Existence for Quasi-Equilibrium Problems

Contents

Part I

Introduction and Preliminaries

I.1 Contents and Structure

Non-smooth problems are now present everyday and everywhere in the life of an optimizer and, in general, in the life of researchers. A non-smooth problem is simply one in which some of the data defining the problem are not necessarily smooth. The non-smooth data can also appear for instance as a result of some operations, like for example when taking the minimum or maximum of finitely and infinitely many smooth functions, the value or marginal function of parametric optimization problems, and so on. One also face non-smooth data when dealing with the value functions in some formulations of Multi-Leader-Follower games undertaken in the current thesis, and, in particular, in bilevel programming problems. Another area in which non-smooth data naturally arise is in semi-infinite programming; that is, in optimization problems which have an infinite number of constraints.

Non-smooth analysis is a theory that includes the smooth case, based on the consideration of relaxed notions of differentiation, since classical differentials might not exist or be adapted. This will lead us to use notions of subdifferentials, which are one-sided relaxations of usual differentials as the Fréchet and the limiting subdifferentials, including the so-called Fenchel and approximate subdifferentials in the convex case.

The present thesis is organized by grouping the chapters in four parts, each with its own short introduction.

While Part I is dedicated to introduce the main notation and definitions used in the manuscript, Part II is focused on fuzzy calculus rules in both the convex and the non-convex frameworks. This part is developed in a general setting, possibly infinite-dimensional and under possibly weak conditions, which do not require qualification conditions. In Chapter II.1, we discuss about the convex case, which corresponds to the analysis of convex and lower semi-continuous functions. Namely, we develop formulas for the normal cone to sub-level sets in locally convex and Banach spaces. In Chapter II.2, we discuss different calculus rules, namely for the pointwise supremum of possibly non-convex functions defined in an appropriate class of Banach spaces, including Asplund spaces.

The contribution of Part II is twofold. We firstly give approximate and fuzzy subdifferential calculus rules, which are next applied to get optimality conditions for optimization problems without qualification conditions. We use proofs that are in many cases shorter than previously known ones, though most of the techniques used in our proofs are not new. Most of the formulas are improvements even when restricted to $\mathbb{R}^n$. Secondly, we provide in a finite dimensional setting fuzzy optimality conditions for bilevel programming problems without any qualification condition.

More precisely, in the convex case we provide a formula for the approximate normal set to a general sub-level set of a convex function, as well as a formula for the approximate subdifferential of the supremum of an arbitrary family of convex functions. Similarly, in the non-convex case, weak fuzzy optimality conditions, a rule for the supremum and a formula for the normal cone to a sublevel sets are provided, with approximation on values and with complementarity conditions, without the use of any qualification condition.

Part III deals with the analysis of non-cooperative games, and particularly, with Multi-Leader-Follower Games, which are bilevel games within each of its levels a Generalized Nash equilibrium is played. We start in Chapter III.1 by analyzing Generalized Nash Equilibrium Problems and the structure of the solutions set, and discussing constraint qualifications for the equivalence of the problem with the concatenation of first order conditions of each of the players. In Chapter III.2, we study the simple bilevel single-leader-single-follower structure, that is, a bilevel programming problem. Here we focus on the so-called pessimistic formulation of the bilevel problem and we aim at comparing the initial problem with the reformulation obtained by replacing the lower level by its parametric first order conditions. In Chapter III.3, a kind of state of the art for multi-leader-follower games is discussed and a few new results are provided. Special attention is given to the single-leader multi-follower case.

The contribution of this part, is again twofold. Firstly, we give an overview of existence results for multi-leader-follower-games and show how constraint qualifications are involved when explicit constraints are present, which are of particular importance for non-linear data. Secondly, we discuss about equivalence between problems and reformulations that involve first order (KKT) conditions. We recognized[1] the quite obvious fact that, in general, for the equivalence of a GNEP with the concatenation of KKT conditions, a possibly infinite number of constraint qualification should be verified. Therefore, we proposed a way of reducing the number of constraint qualifications in some cases yielding just a finite number of them. This analysis starts in fact in Section I.5, and continues with its consequences in Part III. Our analysis here is mainly focused on finite dimensional spaces, for simplicity.

Part IV, deals with the theory of quasi-equilibrium problems, which is a framework for studying in a unifying manner different problems such as generalized Nash equilibrium problems and quasi-variational inequalities. Here we consider several properties for bi-functions that are useful in the study of existence results for quasi-equilibrium problems.

Our contribution in this last part is that we give some new existence results for quasi-equilibrium problems. Some of them relax the continuity properties of the constraint map and other are based on the finite intersection property for bi-functions. Next, we apply our general results to the cases of quasi-variational inequalities and generalized Nash equilibrium problems.

The rest of this introductory part states our framework, notation, definitions and some preliminary results, that are common to either the whole thesis or at least some of the chapters.

I.2 General Notation

We will use almost always the convention that upper case means sets or set-valued maps, while lower case means elements of certain sets or single-valued functions. The set of natural numbers will be denoted by $\mathbb{N}$. Let $\mathbb{R}$ be the *real line*, and $\mathbb{R}_+$ the subset of non-negative real numbers. Given $x \in \mathbb{R}$ the absolute value of x is denoted by $|x|$ and the sign of x is denoted by $sgn(x) = x/|x|$ if $x \neq 0$ and $sgn(0) = 0$.

We consider the extension of the usual ordering $\leq$ by adding a maximal element ∞ and a minimal $-\infty$ and we write $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty, -\infty\}$, which we call it the *extended real line*. The symbol $:=$ is used to express equality by definition. We also write $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$.

We extend also the domain of the sum operation by

$$\infty + x := x + \infty := \infty, \text{ if } x \neq -\infty$$
$$-\infty + x := x + (-\infty) := -\infty, \text{ if } x \neq \infty$$

[1]We are aware that this fact was also recognized by Ehrenmann in his PhD thesis [46], but has been underestimated by several authors.

and the multiplication operation by

$$\infty \cdot x := x \cdot \infty := \begin{cases} \infty, & \text{if } x \geq 0 \\ -\infty, & \text{if } x < 0 \end{cases}$$

Observe that we do not define the sum $\infty + (-\infty)$ nor the multiplications $x \cdot (-\infty)$, and thus we will always avoid them. Also note that the multiplication $0 \cdot \infty$ is defined as ∞. This choice is convenient for our minimization purposes, since ∞ is thought as a penalization. See [103] for further discussion about this convention.

For two vectors a and b in the euclidean space $\mathbb{R}^d$, we write $a \leq b$ when $a_k \leq b_k$ for all $k = 1, ..., d$, and we write $a < b$ when $a_k < b_k$ for all $k = 1, .., d$. We also write $a \perp b$ whenever their product $a^T b$ is equal to 0.

By X we usually denote a space, that is, a given non-empty set. Whenever X is endowed with a topology τ and $\bar{x} \in X$, we write $\mathcal{N}_\tau(\bar{x})$ (or simply $\mathcal{N}(\bar{x})$ if the topology is understood) to denote the set of neighborhoods of $\bar{x}$. Given a subset $S \subset X$ the *interior*, *closure* and *boundary* of S are denoted by $int(S)$, $cl(S)$ and $bd(S)$, respectively.

In Chapter II.1 we assume X to be (at least) a locally convex space, that is, a Hausdorff separated topological vector space whose topology admits a local base of convex neighborhoods of the origin. Two important cases of locally convex spaces that we consider along this thesis are Banach spaces $(X, \|\cdot\|)$, where $\|\cdot\|$ is the norm of the space, and the topological dual X^* of a locally convex space endowed with the weak star topology w^*. The duality product between a locally convex space and its dual is denoted by $\langle x^*, x \rangle := x^*(x)$.

In the case of a Banach space X we write $B(x, r)$ and $\overline{B}(x, r)$ to denote an the open ball and the closed ball, both centered at $x \in X$ and with radius $r > 0$. We also write $B(r) := B(0, r)$, and $\mathbb{B} := \overline{B}(0, 1)$, and sometimes we put X as a subindex to emphasize the space. If X^* is considered as a Banach space with a dual norm we write $\mathbb{B}^* := \mathbb{B}_{X^*}$.

Asplund spaces, which are considered in Chapter II.2, are defined as Banach space with the property that any convex and continuous function defined in an open and convex domain is actually Fréchet differentiable in a dense subset of the domain.

The *convex hull* and the *closed convex hull* of $S \subset X$ (for X a topological vector space) are denoted by $co(S)$ and $\overline{co}(S)$. The *sum* of two sets $A, B \subset X$ is defined by

$$A + B := \{x \in X \mid x = a + b,\ a \in A, b \in B\}$$

and the *multiplication* of A with a set of scalars $R \subset \mathbb{R}$ is

$$RA := \{x \in X \mid x = ra,\ r \in R, a \in A\}.$$

In particular for $R = \mathbb{R}_+$, the set $\mathbb{R}_+ A$ is the *conic hull* of (or the cone generated by) A. Whenever one of the sets is a singleton, for instance $A = \{a\}$ we simplify the notation and write $a + B = A + B$. Similarly for the multiplication, if $R = \{r\}$ then we write $rA = RA$.

For a function $f : X \to \overline{\mathbb{R}}$ we define its *graph*, its *epigraph*, and its *hypograph*, respectively, as the sets

$$\operatorname{gph} f := \{(x, \lambda) \in X \times \mathbb{R} \ : \ f(x) = \lambda\},$$
$$\operatorname{epi} f := \{(x, \lambda) \in X \times \mathbb{R} \ : \ f(x) \leq \lambda\},$$
$$\operatorname{hyp} f := \{(x, \lambda) \in X \times \mathbb{R} \ : \ f(x) \geq \lambda\}.$$

Given $\lambda \in \mathbb{R}$, we write $[f \leq \lambda] := \{x \in X : f(x) \leq \lambda\}$ and $[f = \lambda] := \{x \in X : f(x) = \lambda\}$ for the *sublevel sets* and *level sets* of f at level λ. We define the *(effective) domain* of f as the set $\operatorname{dom} f := \{x \in X : f(x) \in \mathbb{R}\}$. We say that f is *proper* if $\operatorname{dom} f \neq \emptyset$ and $f(x) > -\infty$ for all $x \in X$.

Given a set $A \subset X$, the *indicator* of A is the function $\chi_A : X \to \overline{\mathbb{R}}$ defined by $\chi_A(x) = 0$ if $x \in A$ and $\chi_A(x) = \infty$ otherwise.

We write $x \to_f \bar{x}$, whenever $x \to \bar{x}$ and $f(x) \to f(\bar{x})$. This of course carries additional information only if f is not necessarily continuous.

I.3 Generalized Convexity

Assume that X is a vector space and let us recall some classical definitions of generalized convexity. A extended real valued function $f : X \to \overline{\mathbb{R}}$ is said to be

- *convex* if, for any $x, y \in X$ and $t \in [0,1]$, we have
$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y);$$
- *quasi-convex* if, for any $x, y \in X$ and $t \in [0,1]$, we have
$$f(tx + (1-t)y) \leq \max\{f(x), f(y)\};$$
- *semi-strictly quasi-convex at level* $\alpha \in \mathbb{R}$ if, for any $x, y \in X$ such that $f(x) \leq \alpha$ and $f(y) < \alpha$, the following holds
$$f(tx + (1-t)y) < \alpha \text{ for all } t \in]0,1[.$$
- *semi-strictly quasi-convex* if it is quasi-convex and, for any $x, y \in X$ such that $f(x) \neq f(y)$, the following holds
$$f(tx + (1-t)y) < \max\{f(x), f(y)\} \text{ for all } t \in]0,1[.$$

The convexity of a function f is equivalent to the convexity of its epigraph, while the quasi-convexity of f is equivalent to the convexity of the (strict) sublevel sets of f. The semi-strictly quasi-convexity of a function has the following characterization.

Proposition 1. *A function f is semi-strictly quasi-convex if and only if, f is semi-strictly quasi-convex at every level $\alpha \in \mathbb{R}$.*

Proof. First assume that f is semi-strictly quasi-convex and let $\alpha \in \mathbb{R}$ be such that $\alpha \geq f(x)$ and $\alpha > f(y)$. Then either $\alpha > \max\{f(x), f(y)\}$ or $f(x) = \alpha > f(y)$ so that in any case we have the inequalities
$$f(tx + (1-t)y) \leq \max\{f(x), f(y)\} \leq \alpha$$
and at least one of them being strict. This proves that f is semi-strictly quasi-convex at level α.

For the converse let us first prove that f is quasi-convex. Fix $x, y \in X$ and $t \in]0,1[$. For any $\alpha > \max\{f(x), f(y)\}$ we know that $f(tx + (1-t)y) < \alpha$, and thus by taking infimum over those α we obtain that $f(tx + (1-t)y) \leq \max\{f(x), f(y)\}$. Second, let us assume that $f(x) \neq f(y)$ and $t \in]0,1[$, without loss of generality $f(x) > f(y)$. If we take $\alpha = f(x)$ then obviously $\alpha > f(y)$ and thus $f(tx + (1-t)y) < \alpha = \max\{f(x), f(y)\}$. □

A semi-strictly quasi-convex function share with convex functions the useful property that any local minimum on a given set is actually a global minimum on that set.

I.4 Optimality Conditions and Constraint Qualifications

Let $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}^d$, and consider an optimization problem in the following form

$$\begin{aligned} \min_{x\in X} \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0 \end{aligned} \tag{1}$$

We do not consider equality constraints here for simplicity. Let us write $C := [g \leq 0]$. A *feasible solution* of the above problem is a point $\bar{x} \in C$. If $\bar{x} \in C$ is such that for any $x \in C$ it holds $f(\bar{x}) \leq f(x)$, then we call $\bar{x}$ an *optimal solution* or simply a *solution* of the problem. A point $\bar{x} \in X$ is said to be a *local optimal solution* or simply a *local solution* if $\bar{x} \in C$ and there exists $U \in \mathcal{N}(\bar{x})$ such that $f(\bar{x}) \leq f(x)$ for all $x \in C \cap U$.

I.4.1 Optimality Conditions

By optimality conditions we refer to certain conditions over a given point $\bar{x}$ usually written as equations, inequalities or even inclusions, that are comparable with the condition of $\bar{x}$ being a local optimal solution of problem (1). Thus they can be necessary optimality conditions, sufficient optimality conditions or both necessary and sufficient optimality conditions.

Let us here assume for simplicity that X is finite dimensional. It is clear that if $\bar{x}$ is a local solution of the above problem and $\tilde{v} \in T_C(\bar{x}) := \{v = \lim v_n : t_n \to 0^+, x + t_n v_n \in C\}$ then $\langle -\nabla f(\bar{x}), \tilde{v}\rangle \leq 0$. This is by definition of the normal cone to C that $-\nabla f(\bar{x}) \in N_C(\bar{x})$. So in other words

$$0 \in \nabla f(\bar{x}) + N_C(\bar{x}).$$

We will consider the so-called KKT optimality conditions due to Karush Kuhn and Tucker. We say that $\bar{x}$ satisfy the KKT optimality conditions if there exists $\mu \in \mathbb{R}^d$ such that

$$\begin{cases} \nabla f(\bar{x}) + \mu \nabla g(\bar{x}) = 0 \\ 0 \leq \mu \perp -g(\bar{x}) \geq 0 \end{cases} \tag{2}$$

We will prove in Chapters II.1 and II.2 that some fuzzy optimality conditions for non-smooth problems do not require any constraint qualification and they are in the smooth case of the form

$$\begin{cases} \lim_n \nabla f(x_n) + \mu \nabla g(x_n) = 0 \\ 0 \leq \mu_n,\ -g(\bar{x}) \geq 0 \\ \lim_n \mu_n^T g(x_n) = 0 \end{cases} \tag{3}$$

for some sequences $(\mu_n)_n \subset \mathbb{R}^d$ and $x_n \to \bar{x}$.

I.4.2 Constraint Qualifications

Given a smooth optimization problem with explicit constraints in the form of inequalities as (1) and a point $\bar{x}$ in the space, a *Constraint Qualification* (CQ for short) is a condition on the constraint functions (not dependent on the objective) at $\bar{x}$, or more precisely on the values and the gradients of the constraint functions at $\bar{x}$, guaranteeing that the KKT conditions are necessary optimality conditions (see for instance [109]).

There are several CQs for smooth optimization in the literature. Some of them are weaker or stronger than other in the sense of implication, but not all of them are comparable. There is one which is the weakest: Guignard's CQ, nevertheless other CQs are easier to verify in some cases, and have additional important properties related to the set of KKT/Lagrange multipliers.

We will recall some of the most classical CQs, for simplicity stated in $\mathbb{R}^n$. Let $C = \{x \in \mathbb{R}^n : g(x) \leq 0\}$, $\bar{x} \in C$ and let $A(\bar{x})$ denote the set of *active indexes* i, that is, $i = 1, ..., d$ such that $g_i(\bar{x}) = 0$.

Guignard Constraint Qualification (GCQ) The normal cone to the feasible set at $\bar{x}$ is equal to the convex cone generated by the gradients of the active inequality constraints: $N_C(\bar{x}) = \mathbb{R}_+ \text{co}\,\{\nabla g_i(\bar{x}) \mid i \in A(\bar{x})\}$.

Abadie Constraint Qualification (ACQ) The linearized cone $L(\bar{x}) = \{v \in X : \langle \nabla g_i(\bar{x}), v\rangle \geq 0, \forall i \in A(\bar{x})\}$ is equal to the tangent cone to the feasible set $T_C(\bar{x})$.

Mangazarian Fromovitz Constraint Qualification (MFCQ) There exists a direction $v \in \mathbb{R}^n$ such that $\nabla g_i(\bar{x}; v) < 0$ for all $i \in A(\bar{x})$. An equivalent dual condition is that if $\mu \in \mathbb{R}^d_+$ is such that $\mu \perp g(\bar{x})$ and $\sum_{i=1}^d \mu_i \cdot \nabla g_i(\bar{x}) = 0$, then $\mu = 0$.

Linear Independence Constraint Qualification (LICQ) The set of gradients of the inequality constraints that are active at the point $\bar{x}$ are linearly independent. The vectors $\nabla g_i(\bar{x})$ with $i \in A(\bar{x})$ is linearly independent.

It is well-known that at any point, LICQ $\Rightarrow$ MFCQ $\Rightarrow$ ACQ $\Rightarrow$ GCQ. Finally, in case that each g_i is convex, MFCQ is equivalent to the following.

Slater Constraint Qualification (SCQ) There exists a point $\tilde{x} \in X$ such that $g_i(\tilde{x}) < 0$ for all $i = 1, ..., d$.

In the case of non-smooth problems there are also CQs but we do not consider them on this thesis.

I.5 Parametric Optimization and KKT Conditions

This section is extracted from [19].

Let X be a real Banach space and P a real vector space of parameters. We consider a parametric optimization problem of the form

$$\begin{array}{ll} \min_x & f(x,p) \\ s.t. & g(x,p) \leq 0, \end{array} \qquad (\mathcal{P}(p))$$

where $f : X \times P \to \mathbb{R}$ is the *objective*, $g : X \times P \to \mathbb{R}^d$ is the *joint constraint function*, and p is a *parameter* in a non-empty set $U \subset P$. We do not consider equality constraints for simplicity. Given $p \in P$ we denote by $\mathcal{F}(p)$ the *feasible set* for problem $\mathcal{P}(p)$, that is, the set of $x \in X$ such that $g(x,p) \leq 0$. We will consider the following parametric assumptions:

(H_1) (Parametric differentiability) For every $p \in U$, $f(\cdot, p)$ and $g(\cdot, p)$ are differentiable.

(H_2) (Parametric convexity) For every $p \in U$, $f(\cdot, p)$ is convex and the components of $g(\cdot, p)$ are quasi-convex.

We can associate to each problem $\mathcal{P}(p)$ the KKT first order optimality conditions. For a point $x \in X$, the KKT(p) conditions are that there exists $\mu \in \mathbb{R}^d$ such that

$$\begin{cases} \nabla_x f(x,p) + \nabla_x g(x,p)\mu = 0, \\ 0 \leq \mu \perp -g(x,p) \geq 0, \end{cases} \tag{KKT(p)}$$

or more explicitly given as

$$\begin{cases} \nabla_x f(x,p) + \sum_{k=1}^{d} \mu_k \nabla_x g_k(x,p) = 0, \\ 0 \leq \mu_k,\ \mu_k g_k(x,p) = 0,\ -g_k(x,p) \geq 0, \quad \forall k = 1,...,d. \end{cases}$$

Let $p \in U$ be fixed for the moment. Thanks to the parametric convexity (H_2), the KKT(p) conditions are sufficient optimality conditions for problem $\mathcal{P}(p)$. If the constraint function $g(\cdot,p)$ satisfies some CQs (at any $x \in \mathcal{F}(p)$, or at least on a set including the $\mathcal{P}(p)$-optimal solutions), then the KKT(p) conditions are also necessary optimality conditions for $\mathcal{P}(p)$.

Thus, if we want to prove that the KKT(p) conditions are both necessary and sufficient for every parameter $p \in U$ simultaneously, one straightforward possibility is to check convexity and qualification conditions for every $p \in U$, but this could be quite demanding. Instead, we are looking for simpler constraint qualifications (and “reduced in number”) on the joint constraint function g (possibly also on the partial constraint functions $g(\cdot,p)$, for some $p \in U$) which ensures that the KKT(p) conditions are necessary and sufficient optimality conditions for problem $\mathcal{P}(p)$, for all $p \in U$.

Definition 1. *A parameter $\hat{p} \in P$ is called*

- admissible *if $\hat{p} \in \mathcal{A} := \operatorname{dom} \mathcal{F}$, that is, $\exists x \in X$ with $g(x,\hat{p}) \leq 0$;*
- interior *if it is an element of $int(\mathcal{A})$;*
- boundary *if it is an element of $bd(\mathcal{A})$.*

Remark 1. *(a) The interior parameters are defined only in terms of the joint constraint function g and not in terms of the set U.*

(b) Whenever $(\hat{x},\hat{p}) \in int\{(x,p) \in X \times P \mid g(x,p) \leq 0\}$, then $\hat{p}$ is an interior parameter.

(c) If the constraint function g is upper semi-continuous, and $\hat{p}$ is a boundary parameter, then for any $x \in X$ there exists some coordinate $k = 1,...,d$ with $g_k(x,\hat{p}) \geq 0$. Thus, on boundary parameters Slater's CQ cannot be fulfilled.

Definition 2. *A parametrized function $h : X \times P \to \mathbb{R}^d$ is said to be* jointly convex *on the product space $X \times P$ if, for any $k = 1,\dots,d$, h_k is jointly convex on $X \times P$, that is, convex with regard to the joint variable (x,p).*

Joint convexity clearly implies the parametric convexity (H_2), that is, $h_k(\cdot,p)$ is convex, for each k and for each p. Nevertheless, in the forthcoming proposition and theorem, this joint convexity will be used to reduce the “number of parameters” for which the qualification conditions needs to be verified.

Proposition 2. *Assume that g is jointly convex on $X \times P$ and that the following joint qualification condition holds*

Joint Slater's CQ*:* $\quad \exists\, (\tilde{x},\tilde{p}) \in X \times P$ *such that* $g(\tilde{x},\tilde{p}) < 0.$

Then for every interior parameter $p \in int(\mathcal{A})$ Slater's CQ holds, that is,
$\exists\, x \in X$ *such that* $g(x,p) < 0$.

Proof. Define on $X \times P$ the real-valued function $\bar{g}$ by $\bar{g}(x,p) := \max_{k=1}^{d} g_k(x,p)$ and let $(\tilde{x},\tilde{p})$ be given by joint Slater's CQ, so that $\bar{g}(\tilde{x},\tilde{p}) < 0$. From the hypothesis, $\bar{g}$ is a jointly convex function.

Assume by contradiction that there exists a $p \in int(\mathcal{A})$ for which Slater's CQ does not hold, that is, $\bar{g}(x,p) \geq 0$ for all $x \in X$. We clearly see that $p \neq \tilde{p}$. Since $p \in int(\mathcal{A})$, then one can find $t > 0$ such that $\hat{p} := p + t(p - \tilde{p}) \in \mathcal{A}$. Now take $\hat{x}$ such that $g(\hat{x},\hat{p}) \leq 0$, and thus $\bar{g}(\hat{x},\hat{p}) \leq 0$. Note that $p = \frac{t}{1+t}\tilde{p} + \frac{1}{1+t}\hat{p} \in [\tilde{p},\hat{p}]$, and take $x := \frac{t}{1+t}\tilde{x} + \frac{1}{1+t}\hat{x}$, which clearly lies in $[\tilde{x},\hat{x}]$. Finally, the joint convexity of $\bar{g}$ yields

$$0 \leq \bar{g}(x,p) \leq \frac{1}{1+t}\bar{g}(\tilde{x},\tilde{p}) + \frac{t}{1+t}\bar{g}(\hat{x},\hat{p}) \leq \frac{1}{1+t}\bar{g}(\tilde{x},\tilde{p}) < 0,$$

a contradiction. □

In Proposition 2, it has been proved that under joint convexity of the constraint functions, Slater's CQ for one parameter implies Slater's CQ for all interior parameters. A natural question is whether a weaker CQ for a single parameter also imply that this CQ persist for all interior parameters.

The following example provides a negative answer for the case of Guignard's CQ.

Example 1. *Let D be the unit closed ball of $\mathbb{R}^2$ and $g : \mathbb{R}^2 \to \mathbb{R}^3$ be the constraint mapping given by*

$$g(x,p) := (d_D^2(x,p), x - 1, -1 - x),$$

where $d_D(x,p) := \inf\{\|(x,p) - (x',p')\| : (x',p') \in D\}$. The set of interior parameters is clearly the interval $]-1,1[$, while the set of boundary parameters is $\{-1,1\}$. Guignard's CQ is satisfied for the parameter $p = 0$ at every feasible point ($x \in [-1,1]$). But for the parameters $0 \neq p \in [-1,-1]$ no CQ holds at the points x in the boundary of the feasible set, since the unique active function at the boundary is $d_D^2(\cdot)$, whose derivative is 0 at these points and does not represent the normal directions to the feasible set at the boundary points.

This example also provides a negative answer to another related question concerning joint CQs. Indeed, one could wonder if a weaker joint CQ for the parametric optimization problem, under the joint convexity assumption, should persist as a parametric CQ along all interior parameters. We observe that in Example 1, the joint Guignard's CQ holds for all feasible points except those (x,p) with $x \neq 0$ and $x^2 + p^2 = 1$. Thus, for all interior parameters in $]-1,0[\, \cup\,]0,1[$ the joint Guignard's CQ does not hold at the boundary of the parametric feasible set.

Theorem 1. *Assume (H_1), (H_2), that g is jointly convex on $X \times P$, and that the following two conditions hold:*

1. *(Joint Slater's CQ) There exists a pair $(\tilde{x},\tilde{p})$ such that $g(\tilde{x},\tilde{p}) < 0$,*
2. *(Guignard's CQ for boundary parameters) For each $\hat{p} \in U \cap bd(\mathcal{A})$, Guignard's CQ holds at each feasible point $x \in \mathcal{F}(\hat{p})$.*

Then, for any $p \in U$, the KKT(p) *conditions are necessary and sufficient optimality conditions for problem $\mathcal{P}(p)$.*

Proof. From the assumptions and Proposition 2 we obtain that for each $p \in U$ the problem $\mathcal{P}(p)$ is convex and either Slater's CQ holds or Guignard's CQ holds for each feasible point. Thus we conclude that $\mathcal{P}(p)$ is equivalent to the KKT(p) system for each $p \in U$. □

Remark 2. *The less the number of boundary parameters in U, the less conditions have to be verified to apply Theorem 1. But boundary parameters usually exist. If the set $R := \{(x,p) \in X \times P \mid g(x,p) \leq 0\}$ is non-empty closed and bounded, then there exist at least one boundary parameter. Even for a two dimensional parameter an infinite number of boundary parameter could arise. Take, for example, $X = \mathbb{R}$ and $P = \mathbb{R}^2$ and B the closed unit ball in $X \times P$, that is $B := \{(x,p) \in X \times P \mid x^2 + p_1^2 + p_2^2 \leq 1\}$. It is clear that the boundary parameters are $bd(\mathcal{A}) = \{p \in P \mid p_1^2 + p_2^2 = 1\}$.*

I.6 Continuity of Set-Valued Maps

Let X and Y be two non-empty sets. A *set-valued map*, which we denote by $T : X \rightrightarrows Y$, is a function that assigns to each point $x \in X$ a (possibly empty) subset $T(x)$ of Y. We consider set-valued maps as extensions of usual functions. These latter being the particular case when for all $x \in X$ the value $T(x)$ is a set of exactly one element. With this idea in mind we consider the graph of T as a subset of $X \times Y$ instead of a subset of $X \times 2^Y$.

The *graph* of a set-valued map $T : X \rightrightarrows Y$ is the set

$$\operatorname{gph} T := \{(x,y) \in X \times Y : y \in T(x)\},$$

and its *domain* is

$$\operatorname{dom} T := \{x \in X : T(x) \neq \emptyset\}.$$

Assume now that X and Y are topological spaces. Appropriate continuity notions will be related directly to the topology of Y, and not to a topology on 2^Y.

Definition 3. *Let $\bar{x} \in X$. We say that a set-valued map $T : X \rightrightarrows Y$ is*

- lower semi-continuous *(lsc, for short) at $\bar{x}$ if for each open set V in Y satisfying $T(\bar{x}) \cap V \neq \emptyset$, there exists $U \in \mathcal{N}_X(\bar{x})$ such that*

$$T(x) \cap V \neq \emptyset, \quad \forall x \in U;$$

- upper semi-continuous *(usc, for short) at $\bar{x}$ if for each open set V in Y satisfying $T(\bar{x}) \subset V$, there exists $U \in \mathcal{N}_X(\bar{x})$ such that*

$$T(x) \subset V, \quad \forall x \in U;$$

- continuous *at $\bar{x}$ if it is both lsc and usc at $\bar{x}$.*

We say that T is lsc (usc, continuous, respectively) in a set $A \subset X$, if it is so at every $\bar{x} \in A$. In the case $A = X$ we omit the reference to the set.

Both lsc and usc for set-valued maps coincide independently with the usual continuity of functions when the set-valued map is single-valued. Nevertheless, they are quite different notions of continuity in general. As shown by the following example, there are set-valued maps that are usc but not lsc, and vice-versa.

Example 2. *Consider an open and non-empty subset A of X such that $A \neq X$ and two arbitrary sets $B, C \subset Y$. Let $T : X \rightrightarrows Y$ be the set-valued map given by*

$$T(x) = \begin{cases} B & \text{if } x \in A \\ C & \text{if } x \notin A. \end{cases}$$

Then T is lsc if and only if $B \subset C$, and T is usc if and only if $C \subset B$. In particular, taking $B = C$ we observe that any constant set-valued map is continuous.

The domain of a set-valued map that is lsc is an open set in X. We now explain the link of these continuity notion with some topological properties of their graphs (see [2,21] for the proofs).

Definition 4. *We say that a set-valued map $T : X \rightrightarrows Y$ is*

- closed *at $\bar{x}$, if $(\bar{x}, \bar{y}) \in cl(\operatorname{gph} T)$ implies that $\bar{y} \in T(\bar{x})$;*
- open *at $\bar{x}$, if for any $\bar{y} \in T(\bar{x})$ it holds $(\bar{x}, \bar{y}) \in int(\operatorname{gph} T)$.*

We simply say that T is closed *(*open*, respectively) if T is closed (open, respectively) at $\bar{x}$, for every point $\bar{x} \in X$.*

We observe that whenever T is closed (open, respectively) at $\bar{x}$ then $T(\bar{x})$ is also closed (open, respectively) in Y. Moreover, T being closed (open, respectively) is equivalent to that the $\operatorname{gph} T$ is a closed (open, respectively) set in $X \times Y$.

Proposition 3. *A set-valued map T is open at $\bar{x}$ if and only if the complement T^c is closed at $\bar{x}$.*

Proof. Given $\bar{y} \in Y$, since $(int(\operatorname{gph} T))^c = cl(\operatorname{gph} T^c)$, the statement "$\bar{y} \in T(\bar{x})$ implies $(\bar{x}, \bar{y}) \in int(\operatorname{gph} T)$" is equivalent to "$(\bar{x}, \bar{y}) \in cl(\operatorname{gph} T^c)$ implies $\bar{y} \in T^c(\bar{x})$". Thus, T open at $\bar{x}$ is equivalent to T^c closed at $\bar{x}$. □

In Example 2 the set-valued map T is closed if and only if B and C are closed and $B \subset C$. Similarly, T is open if and only if B and C are open and $C \subset B$.

Obviously, a set valued map $T : X \rightrightarrows Y$ is lsc and open at each point $\bar{x} \in X$ with $T(\bar{x}) = \emptyset$ (in other words $\bar{x} \notin \operatorname{dom}(T)$), while it is usc and closed at each point with full image, that is, $\bar{x} \in \operatorname{dom} T$ such that $T(\bar{x}) = Y$.

Remark 3. *The definition of a map being open at a point is equivalent to its strongly lsc at that point, a notion that has been used in [21]. It is also an appropriate name since if a map T is strongly lsc/open at $\bar{x}$, then T is also lsc at $\bar{x}$.*

It is easy to see that a set-valued map T that is the union of functions which are continuous at $\bar{x}$ is lsc at $\bar{x}$. We present the following conjecture.

Conjecture 1. *Any set-valued map that is lsc at $\bar{x}$ can be expressed around $\bar{x}$ as the union of a family of functions which are continuous at $\bar{x}$. This would be somehow related to Michael's selection theorems.*

There is also a link between the usc of a set-valued map at a point and its closedness at that point.

Proposition 4. *Let $T : X \rightrightarrows Y$ and $\bar{x} \in X$. Assume that the topological space Y satisfies the separation axiom T_3.*

1. *If T is usc at $\bar{x}$ and $T(\bar{x})$ is a closed set, then T is closed at $\bar{x}$.*

2. *If Y is compact and T has closed graph, then T is usc.*

Without the compactness of Y in the second part of Theorem 4, the graph of T being closed is not enough to guarantee the usc of T. In fact, we have found the following result (valid in normed vector spaces) that seems to limit the possibilities for usc set-valued maps at points where the value is closed but non-compact.

Lemma 1. *Assume Y is a normed space and X is a metric space. Let $T : X \rightrightarrows Y$ be a set-valued map such that it is usc at a point $\bar{x}$ and $T(\bar{x})$ is closed. Then there exist $r_1 > 0$ and $r_2 > 0$ such that*

$$T(x) \backslash B_Y(r_2) \subset T(\bar{x}), \quad \forall x \in B_X(\bar{x}, r_1).$$

Proof. Assume that the conclusion is not true. Then for every $n \in \mathbb{N}$ there exist $x_n \in B_X(\bar{x}, \frac{1}{n})$ and $y_n \in T(x_n)$ such that $y_n \notin T(\bar{x})$ and $y_n \notin B_Y(n)$. Note that $|y_n| \geq n$ and so $\lim |y_n| = +\infty$. Then we can assume without loss of generality that $|y_n|$ is increasing, and moreover, that $|y_n| + 1 < |y_{n+1}|$. We will find an open set V that contains $T(\bar{x})$ while the condition $T(x_n) \subset V$ is false for all $n \in \mathbb{N}$, thus contradicting the usc of T at $\bar{x}$.

Since $T(\bar{x})$ is closed and $y_n \notin T(\bar{x})$ there exist $\varepsilon_n > 0$ such that

$$B_Y(y_n, \varepsilon_n) \cap T(\bar{x}) = \emptyset, \tag{4}$$

which we can chose such that $\varepsilon_n < \frac{1}{2}$. Now consider the set

$$V := \bigcup_{\substack{n \in \mathbb{N} \\ y \in T(\bar{x}) \\ |y_n| < |y| + \frac{1}{2} \leq |y_{n+1}|}} B_Y(y, \varepsilon_n),$$

where y_0 is set to be 0. It is clear that V is an open set, as a union of open balls, and it contains $T(\bar{x})$. But as we shall see $y_n \notin V$ for all $n \in \mathbb{N}$, and thus it is not true that $T(x_n) \subset V$.

Given $n \in \mathbb{N}$, let us see that $y_n \notin V$. If $y_n \in V$, then there exist $m \in \mathbb{N}$ and $y \in T(\bar{x})$ such that $|y_m| < |y| + \frac{1}{2} \leq |y_{m+1}|$ for which $y_n \in B_Y(y, \varepsilon_m)$. If $n < m$ then

$$|y_m| < |y| + \frac{1}{2} \leq |y_n| + \varepsilon_m + \frac{1}{2} < |y_n| + 1 < |y_m|,$$

a contradiction. Also note that

$$|y_n| < |y| + \varepsilon_m < |y_{m+1}| - \frac{1}{2} + \varepsilon_m < |y_{m+1}|,$$

so we have $n \leq m$. Thus $n = m$ and $y_n \in B_Y(y, \varepsilon_n)$, with $y \in T(\bar{x})$. We can write then that $y \in B_Y(y_n, \varepsilon_n) \cap T(\bar{x})$ which is a contradiction with (4). □

The previous result is related with Lemma 2.2.2 in [21]. Inspired on these results we have the following conjecture.

Conjecture 2. *Assume X and Y are metric spaces. Let $T : X \rightrightarrows Y$ be a set-valued map that is usc at a point $\bar{x}$ and $T(\bar{x})$ is closed. Then there exist $r > 0$ and C compact in Y such that*

$$T(x)\backslash C \subset T(\bar{x}), \quad \forall x \in B_X(\bar{x}, r).$$

The following theorem, which is based on Lemma 1, corresponds to the first part of [21, Theorem 4.2.3], when restricted to finite dimensions.

Theorem 2. *Assume X is a metric spaces. Let $K : X \rightrightarrows \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}$ and consider the value function $\varphi(x) := \inf\{f(y) \mid y \in K(x)\}$. If K is usc with closed values, and f is lsc, then φ is lsc.*

Proof. If φ is not lsc at x, then there exist $(x_n)_n \subset X$ converging to x and $\varepsilon > 0$ such that $\lim \varphi(x_n) < \varphi(x) - \varepsilon$. In particular, there exist $y_n \in K(x_n)$ such that $f(y_n) < \varphi(x) - \varepsilon$ for all $n \geq n_0$. Since K is usc and $K(x)$ is closed, from Lemma 1 we see that $K(x_n)\backslash B(r_2) \subset K(x)$ for each $n \geq n_1$. Thus we deduce that $y_n \in B(r_2)$ for all $n \geq n_1$. Without loss of generality assume that $y_n \to y \in Y$. Since K has closed graph (closed valued + usc) then $y \in K(x)$ and we obtain that

$$\liminf f(y_n) \geq f(y) \geq \varphi(x) \geq \liminf f(y_n) + \varepsilon,$$

a contradiction. □

Theorem 3. *Assume X and Y are metric spaces. Let $K : X \rightrightarrows Y$ and $f : X \times Y \to \mathbb{R}$. We consider the value function $\varphi(x) := \inf\{f(x, y) \mid y \in K(x)\}$. If K is lsc (relative to its domain) and f is usc then φ is usc.*

Now we want to analyze the properties of the solution mapping defined by $S(x) := \operatorname{argmin}_y \{f(x, y) \mid y \in K(x)\}$, assuming conditions over $f : X \times Y \to \mathbb{R}$ and a set-valued map $K : X \to Y$. The following result is related to [40, Theorem 4.3].

Proposition 5. *Assume X and Y are metric spaces. Let $K : X \rightrightarrows Y$ and $f : X \times Y \to \mathbb{R}$. If K is lsc and closed, and f is continuous, then the solution set-valued map*

$$S(x) := \operatorname*{argmin}_y \{f(x, y) \mid y \in K(x)\}$$

is closed too.

Proof. Let $(\bar{x}_n, \bar{y}_n) \in S$ such that $(\bar{x}_n, \bar{y}_n) \to (\bar{x}, \bar{y})$. Since $S \subset K$ and K has closed graph, then $(\bar{x}, \bar{y}) \in K$. Take $y \in K(\bar{x})$ and we want to prove that $f(\bar{x}, \bar{y}) \leq f(\bar{x}, y)$. Assume by contradiction that $f(\bar{x}, \bar{y}) > f(\bar{x}, y)$ and take $y_n \in K(\bar{x}_n)$ (lsc of K) such that $y_n \to y$ and without loss of generality $f(\bar{x}_n, y_n) < f(\bar{x}, y) + \varepsilon \leq f(\bar{x}, \bar{y}) - \varepsilon$ for some $\varepsilon > 0$, since f is usc. But lsc of f ensures that $f(\bar{x}_n, \bar{y}_n) > f(\bar{x}, \bar{y}) - \varepsilon$ for n large enough. We obtain

$$f(\bar{x}_n, y_n) < f(\bar{x}, \bar{y}) - \varepsilon < f(\bar{x}_n, \bar{y}_n),$$

which is clearly a contradiction since $\bar{y}_n \in S(\bar{x}_n)$. □

Example 3. *The following four examples show that we cannot drop the hypothesis of lsc of K, nor closedness of K, nor usc of f nor lsc of f. In all of them $K : [0, 1] \rightrightarrows [0, 1]$ and $f : [0, 1] \to [0, 1]$ (more precisely $\tilde{f} : [0, 1] \times [0, 1] \to [0, 1]$ by writing $\tilde{f}(x, y) := f(y)$ for all $x \in [0, 1]$).*

1. *Let $K(0) := [0,1]$ and $K(x) := \{0\}$ for $x > 0$. Let $f(y) := -y$ for $y \in [0,1]$. The solution mapping is $S(x) = \{1\}$ for $x > 0$ and $S(0) = \{0\}$ which is not graph closed. Proposition 5 does not apply because K is not lsc.*

2. *Let $K(0) := \{0\}$ and $K(x) = [0,1]$ for $x > 0$. Let $f \equiv 0$. The solution mapping S is equal to K which has not closed graph. Proposition 5 does not apply because K is not closed.*

3. *Let $K(x) := [0,x]$ for $x \in [0,1]$ and let $f(y) = 1$ for $y > 0$ and $f(1) = 0$. The solution mapping is $S(x) = [0,x]$ for $x > 0$ and $S(0) = \{1\}$ which is not graph closed. Proposition 5 does not apply because f is not usc.*

4. *Let $K(x) := [0,x]$ for $x \in [0,1]$ and let $f(0) = 1$ and $f(y) = 0$ for $y > 0$. The solution mapping is $S(x) = (0,x]$ for $x > 0$ and $S(0) = \{1\}$ which is not graph closed. Proposition 5 does not apply because f is not lsc.*

The intersection of lsc set-valued maps needs not to be lsc. An error concerning this intersection rule has been made in [112, Proposition 5]. But if one of the set valued maps is also open then the intersection is lsc (see [21, Lemma 2.2.5]), in particular we have the following result.

Lemma 2. *Let X, Y be two topological spaces, $T : X \rightrightarrows Y$ a set-valued map, and V an open subset of Y. If T is lsc at $x_0 \in X$, then the set-valued map $T_V : X \rightrightarrows Y$ defined by*

$$T_V(x) := T(x) \cap V, \tag{5}$$

is also lsc at x_0.

Proof. Let V_1 be an open subset of Y such that $T_V(x_0) \cap V_1 \neq \emptyset$. We put $V_2 := V_1 \cap V$, which is open. Since $T_V(x_0) \cap V_1 = T(x_0) \cap V_2$, by lower semi-continuity of T, there exists a neighborhood U of x_0 such that $T(x) \cap V_2 \neq \emptyset$ for all $x \in U$, or equivalently $T_V(x_0) \cap V_1 \neq \emptyset$ for all $x \in U$, so that T_V is lower semi-continuous at x_0. □

Lemma 3. *Let X, Y and T be as in Lemma 2. Assume that T is lsc at $x_0 \in X$, and let a set-valued map $S : X \rightrightarrows Y$ such that $S(x_0) \subset \overline{T(x_0)}$ and*

$$T(x) \subset S(x), \quad \forall x \in X.$$

Then S is lsc at x_0.

Proof. Let V be an open subset of Y such that $S(x_0) \cap V \neq \emptyset$. Clearly, $\overline{T(x_0)} \cap V \neq \emptyset$, and we deduce that $T(x_0) \cap V \neq \emptyset$. Thus, by the lower semi-continuity of T there exists a neighborhood U of x_0 such that $\emptyset \neq T(x) \cap V \subset S(x) \cap V$, for all $x \in U$. □

Lemma 4. *Let X and T be as in Lemma 2, Y a topological vector space, and V an open convex subset of Y. Let $x_0 \in X$ such that $T(x_0) \cap V \neq \emptyset$. If T is lsc at x_0 and $T(x_0)$ is convex, then the set-valued map $T_{\overline{V}}$, defined similarly as in (5), is lsc at x_0.*

Proof. The set-valued map T_V, which is lsc at x_0 by Lemma 2, satisfies $T_V(x) \subset T_{\overline{V}}(x)$ for all $x \in X$ and, due to the accessibility lemma [106],

$$T_{\overline{V}}(x_0) \subset \overline{T(x_0)} \cap \overline{V} = \overline{T_V(x_0)}.$$

Thus, $T_{\overline{V}}$ is lsc at x_0 thanks to Lemma 3. □

The following lemma can be easily proved (see Lemma 2.3 in [85]).

Lemma 5. *Let X, Y and T be as in Lemma 2, A a closed subset of X, and $S : A \rightrightarrows Y$. We define the set-valued map $M : X \rightrightarrows Y$ as*

$$M(x) := \begin{cases} T(x) & \text{if } x \in X \setminus A \\ S(x) & \text{if } x \in A. \end{cases}$$

If S, T are lsc and $S(x) \subset T(x)$, for all $x \in A$, then M is lsc.

The following result is Theorem 5.9(c) in [103].

Lemma 6. *If $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is lsc at $x_0 \in \mathbb{R}^n$, then so is the set-valued map $\mathrm{co}(T) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ defined as*

$$\mathrm{co}(T)(x) := \mathrm{co}(T(x)).$$

The following is a consequence of Himmelberg's fixed point and Michael's selection theorems. Recall that for a set-valued map $T : C \subset \mathbb{R}^n \rightrightarrows C$, $\mathrm{Fix}(T)$ is the set of fixed points of T; that is, $x \in C$ with $x \in T(x)$.

Proposition 6 (Corollary 1 in [35])**.** *Given a non-empty, convex and closed subset C of $\mathbb{R}^n$, if $T : C \rightrightarrows C$ is lsc with non-empty and convex values and $T(C)$ is bounded, then $\mathrm{Fix}(T) \neq \emptyset$.*

Given a set-valued map $T : X \rightrightarrows Y$, between two sets X, Y, the *fibre* of T at $y \in Y$ is the set

$$T^{-1}(y) := \{x \in X : \; y \in T(x)\}.$$

The following result is a particular case of [64, Theorem 5] (see, also [74, Theorem 4 of §5]).

Proposition 7. *Let $C \subset \mathbb{R}^n$ be a compact, convex and non-empty set, and let $S, T : C \rightrightarrows C$ be two set-valued maps such that*

1. *S is usc with convex, compact and non-empty values,*
2. *T is convex-valued with open fibers and $\mathrm{Fix}(T) = \emptyset$,*
3. *the set $V := \{x \in C : \; S(x) \cap T(x) \neq \emptyset\}$ is open in C.*

Then there exists $x \in \mathrm{Fix}(S)$ such that $S(x) \cap T(x) = \emptyset$.

Remark 4. *The previous result includes, in one hand, Kakutani Fixed Point Theorem (taking $\mathrm{gph}\, T = \emptyset$), and on the other hand, the Fan-Browder Fixed Point Theorem (taking $\mathrm{gph}\, S = C \times C$).*

Part II

Non-Smooth Calculus Rules

Context and Abstract

In this part we will discuss about some extensions of the very well-known smooth calculus rules and optimality conditions for non-linear problems to the non-smooth framework by considering subdifferentials of the functions. Our approach concerns fuzzy subdifferential calculus rules, which are approximate rules that considers subdifferentials at points that are close to the reference point. In the convex case we also consider approximate subdifferentials.

In Chapter II.1 we restrict our attention to the convex case, that is, assuming convexity of the involved functions. We start recalling some known results as a small survey but we also present some recent improvements based on our work [54], particularly in supremum rules and normal cone rules. This survey will also serve for as introducing the next chapter.

In Chapter II.2, we consider the non-smooth and non-convex case in the framework of Asplund spaces. We improve some known fuzzy optimality conditions and give estimations of the subdifferential of the supremum of an arbitrary family of lower semi-continuous functions and of the normal cone to the sublevel set of a lower semi-continuous function in terms of the subdifferentials of the data functions.

Chapter II.1

Calculus Rules in Convex Analysis

II.1.1 Introduction

In this chapter, we will discuss about some extensions of well-known smooth calculus rules and optimality conditions to a non-smooth convex framework.

We start by recalling some known results, including some recent improvements that we have made in [54], which rely on subdifferential calculus for the supremum function, the approximate normals to sub-level sets, and optimality conditions. This survey will also serve for as introducing the next chapter where the general non-smooth and non-convex case is considered.

Let X be a locally convex space and consider a function $f : X \to \overline{\mathbb{R}}$. We recall that the *effective domain* of f is the set $\operatorname{dom} f := \{x \in X : f(x) < \infty\}$. We say that f is *convex* if for any $x_0, x_1 \in X$ and $\lambda \in [0,1]$,

$$f(x_\lambda) \leq \lambda f(x_1) + (1-\lambda) f(x_0), \tag{II.1.1}$$

where $x_\lambda := \lambda x_1 + (1-\lambda)x_0$. Note that since the values are allowed to be $+\infty$ we use the convention described in the general Introduction in Chapter 1.

Geometrically, the convexity of the function f is equivalent to the convexity of its *epigraph*,

$$\operatorname{epi} f := \{(x, \alpha) \in X \times \mathbb{R} \mid f(x) \leq \alpha\}. \tag{II.1.2}$$

Recall that a set $C \subset X$ is convex if for all $x_0, x_1 \in X$, the whole segment $[x_0, x_1] := \{\lambda x_1 + (1-\lambda)x_0 \mid \lambda \in [0,1]\}$ is included in C. If f is convex, then also $\operatorname{dom} f$ is convex, as well as the *sublevel* sets $[f \leq \alpha] := \{x \in X : f(x) \leq \alpha\}$ and the *strict sublevel sets* $[f < \alpha] := \{x \in X : f(x) < \alpha\}$ for $\alpha \in \mathbb{R}$.

An important observation is that the tangential approximation of a smooth and convex function lies below the function itself. This fact led to the first ideas of the so-called subgradients for non-smooth convex functions.

We say that $x^* \in X^*$ is a *subgradient* of f at x if

$$\langle x^*, x - y\rangle + f(x) \leq f(y), \quad \forall x \in X. \tag{II.1.3}$$

A subgradient does not always exist and might not be unique. Thus, the set of subgradient of a convex function f at a given point x, which is called the *subdifferential* of f at x and is denoted by $\partial f(x)$ could be empty or possibly contain more than one point.

Given $\varepsilon \geq 0$ the *ε-subdifferential* of f at $x \in \operatorname{dom} f$ is the set of $x^* \in X^*$ such that

$$\langle x^*, x - y\rangle + f(x) \leq f(y) + \varepsilon, \quad \forall x \in X. \tag{II.1.4}$$

It is clear that when $\varepsilon = 0$ the ε-subdifferential coincide with the (exact) subdifferential of a function.

The advantage of considering a positive $\varepsilon > 0$ is that $\partial_\varepsilon f(x)$ is non-empty for all $x \in \operatorname{dom} f$ and as we shall see, it provides of useful information. Obviously, $\partial_\varepsilon f(x) \subset \partial_\delta f(x)$ whenever $\varepsilon < \delta$, and further we also have the following simple approximation properties.

Proposition 8. *For any $f : X \to \overline{\mathbb{R}}$, $x \in \operatorname{dom} f$ and $\varepsilon \geq 0$ we have*

$$\partial_\varepsilon f(x) = \bigcap_{\delta > \varepsilon} \partial_\delta f(x) \tag{II.1.5}$$

Additionally, if $\partial_{\varepsilon_0} f(x) \neq \emptyset$ for some $\varepsilon_0 > 0$ (as in the case of f convex and lsc) then for any $\varepsilon > \varepsilon_0$ we have

$$\partial_\varepsilon f(x) = cl\left(\bigcup_{\delta \in\ ]0,\varepsilon[} \partial_\delta f(x)\right) \tag{II.1.6}$$

Proof. The first relation is obvious and well-known, so we only prove the second. First, $\partial_\varepsilon f(x)$ is closed so that the inclusion to the left follows. Second, let $x^* \in \partial_\varepsilon f(x)$ and consider $x_0^* \in \partial_{\varepsilon_0} f(x)$. For $\lambda \in [0, 1]$ we define $x_\lambda^* := \lambda x^* + (1 - \lambda)x_0^*$, which obviously lies in $\partial_{\delta_\lambda} f(x)$, with $\delta_\lambda := \lambda\varepsilon + (1 - \lambda)\varepsilon_0$. Since $x_\lambda^* \to x^*$ as $\lambda \to 1$ with $\lambda < 1$, then x^* belong to the left hand side of (II.1.6). □

The directional ε-derivative of f at x in a direction $v \in X$ is defined as

$$f'_\varepsilon(x, v) := \inf_{t>0} \frac{f(x + tv) - f(x) + \varepsilon}{t};$$

again, if $\varepsilon = 0$, we just call it directional derivative and write $f'(x, v)$.

Theorem 4. *Given $f \in \Gamma(X)$ and $\varepsilon > 0$ we have*

$$f'_\varepsilon(x, v) = \sup_{x^* \in \partial_\varepsilon f(x)} \langle x^*, v\rangle .$$

The relation in Proposition 8 is related to the following lemma concerning approximate directional derivatives. The proof we present is much shorter than the one presented in [54, Lemma 24] where the analysis was made directly over the function.

Lemma 7. *Let $f \in \Gamma_0(X)$, $\bar{x} \in \operatorname{dom} f$ and $v \in X$. Then the function $\varepsilon \to R(\varepsilon) := f'_\varepsilon(\bar{x}; v)$ is non-decreasing and continuous on $\mathbb{R}_+$.*

Proof. Given $\varepsilon > 0$, and thanks to Theorem 4, the relation (II.1.5) ensures that $\lim_{\delta\to\varepsilon^+} R(\delta) = R(\varepsilon)$ and the relation (II.1.6) that $\lim_{\delta\to\varepsilon^-} R(\delta) = R(\varepsilon)$. The case $\varepsilon = 0$ can be argued by simply interchanging the infimum over $\delta > 0$ and the one defining the directional derivative. This proves the continuity of R. □

We recall now the Ekeland variational principle.

Theorem 5. *Let (X,d) be a complete metric space, let $f : X \to \overline{\mathbb{R}}$ a lsc function that is bounded from below. Let $z \in X$ and $\varepsilon > 0$ satisfy*

$$f(z) < \inf f + \varepsilon.$$

Then, for any $\lambda > 0$ there exists $y \in X$ such that

1. *$d(z,y) \le \lambda$,*
2. *$f(y) + \varepsilon\lambda^{-1}d(y,z) \le f(z)$,*
3. *$f(x) + \varepsilon\lambda^{-1}d(y,x) > f(y)$, for all $x \in X\backslash\{y\}$.*

The following theorem corresponds to the Lemma in the seminal paper [28] in 1964. It consists basically in applying the Ekeland variational principle.

Theorem 6. *Let X be a Banach space, $f : X \to \mathbb{R} \cup \{\infty\}$ be a convex and lsc function and $x_0 \in \operatorname{dom} f$. Let $x_0^* \in \partial_\varepsilon f(x_0)$ for $\varepsilon, \lambda > 0$. Then there exists $x_\varepsilon \in X$ and $x_\varepsilon^* \in \partial f(x_\varepsilon)$.*

$$\|x_\varepsilon - x_0\| \le \lambda, \tag{II.1.7}$$
$$\|x_\varepsilon^* - x_0^*\| \le \varepsilon\lambda^{-1}. \tag{II.1.8}$$

Proof. Consider the function $g = f - x^*$, which is convex and lsc, and satisfy

$$g(x_0) \le \inf g + \varepsilon.$$

From the Ekeland variational principle we know there exists $x_\varepsilon \in X$ which is a minimum point of the function $h(x) := g(x) + \varepsilon/\lambda\|x - x_\varepsilon\|$ and that $g(x_\varepsilon) + \varepsilon/\lambda\|x_\varepsilon - x_0\| \le g(x_0)$. Noting also that $g(x_0) \le g(x_\varepsilon) + \varepsilon$, we deduce that $\|x_\varepsilon - x_0\| \le \lambda$. Now since h is a sum of convex functions, only one of them being possibly not continuous, we can use the sum rule of Theorem 7 and we have that

$$0 \in \partial(g + \varepsilon/\lambda\| \cdot - x_\varepsilon\|)(x_\varepsilon) = \partial f(x_\varepsilon) - x_0^* + \varepsilon/\lambda\mathbb{B}^*.$$

We conclude thus that there exists $x_\varepsilon^* \in \partial f(x_\varepsilon)$ such that $\|x_0^* - x_\varepsilon^*\| \le \varepsilon/\lambda$. □

Remark 5. *It follows from the above theorem that the obtained points x_ε and x_ε^* satisfy also*

$$\|x_\varepsilon - x_0\|\|x_\varepsilon^*\| \le \varepsilon + \lambda\|x_0^*\|, \tag{II.1.9}$$
$$|f(x_\varepsilon) - f(x_0)| \le \varepsilon + \lambda\|x_0^*\|, \tag{II.1.10}$$
$$x_\varepsilon^* \in \partial_{2\varepsilon} f(x_0), \tag{II.1.11}$$

and in particular

$$|\langle x_\varepsilon - x_0, x_\varepsilon^*\rangle| \le \varepsilon + \lambda\|x_0^*\|. \tag{II.1.12}$$

We observe thus that it is not necessary to 'prove again' theorems like [124, Theorem 3.1.1] or [25, Theorem 1], which simply add some of the above conclusions to the ones in Theorem 6. Further we obtained this way a stronger complementarity condition that we will use later on.

Proof. (of remark 5). We see that by the triangular inequality

$$\begin{aligned}\|x_\varepsilon - x_0\|\|x_\varepsilon^*\| &\leq \lambda(\|x_\varepsilon^* - x_0^*\| + \|x_0^*\|)\\ &\leq \lambda(\varepsilon\lambda^{-1} + \|x_0^*\|)\\ &= \varepsilon + \lambda\|x_0^*\|,\end{aligned}$$

which proves the first inequality and also (II.1.12). Thus, since $x_\varepsilon^* \in \partial f(x_\varepsilon)$ we obtain

$$f(x_\varepsilon) - f(x_0) \leq \langle x_\varepsilon^*, x_\varepsilon - x_0\rangle \leq \varepsilon + \lambda\|x_0^*\|.$$

Moreover, since $x_0^* \in \partial_\varepsilon f(x_0)$ then

$$-\lambda\|x_0^*\| - \varepsilon \leq \langle x_0^*, x_\varepsilon - x_0\rangle - \varepsilon \leq f(x_\varepsilon) - f(x_0)$$

and thus we have proved the second inequality of the remark. Finally, since obviously $\langle x_\varepsilon^* - x_0^*, x_\varepsilon - x_0\rangle \leq \varepsilon$, then for any $x \in X$ we have

$$\begin{aligned}\langle x_\varepsilon^*, x - x_0\rangle &= \langle x_\varepsilon^*, x - x_\varepsilon\rangle + \langle x_\varepsilon^*, x_\varepsilon - x_0\rangle\\ &\leq f(x) - f(x_\varepsilon) + \langle x_0^*, x_\varepsilon - x_0\rangle + \varepsilon\\ &\leq f(x) - f(x_0) + \varepsilon + \varepsilon,\end{aligned}$$

which proves that $x_\varepsilon^* \in \partial_{2\varepsilon} f(x_0)$. □

II.1.2 Sum Rule

Under a weak qualification condition we have the following subdifferential sum rule, the Moreau-Rockafellar Theorem.

Theorem 7. *Assume X is a Banach space, let $f, g : X \to \mathbb{R}$ be convex and lower semi-continuous and let $x \in X$. If $x \in \operatorname{dom} f \cap int(\operatorname{dom} g)$, then*

$$\partial(f+g)(x) = \partial f(x) + \partial g(x). \tag{II.1.13}$$

This powerful and quite simple sum rule holds at points x in $\operatorname{dom} f \cap int(\operatorname{dom} g)$ (or symmetrically in $int(\operatorname{dom} f) \cap \operatorname{dom} g$), but not necessarily if x belongs to $\operatorname{dom} f \cap \operatorname{dom} g = \operatorname{dom}(f+g)$. For instance, if we take $f, g : \mathbb{R} \to \overline{\mathbb{R}}$ given by $f(x) := -\sqrt{x}$ for $x \geq 0$, $f(x) := +\infty$ otherwise, and $g(x) := f(-x)$, then $\partial(f+g)(0) = \mathbb{R}$ while $\partial f(0) + \partial g(0) = \emptyset$, since both sets in the sum are empty.

In order to avoid a qualification condition like $x \in \operatorname{dom} f \cap int(\operatorname{dom} g)$ it is possible to consider approximate subdifferentials, which are non-empty in all the domain of a convex and lower semi-continuous function (see [59]).

Theorem 8. *Assume X is a l.c.t.v.s., let $f_i : X \to \overline{\mathbb{R}}$ be convex and lower semi-continuous and let $x \in X$. If $x \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$, then*

$$\partial(f_1 + f_2)(x) = \bigcap_{\varepsilon>0} cl(\partial_\varepsilon f_1(x) + \partial_\varepsilon f_2(x)). \tag{II.1.14}$$

In the above theorem we can apply directly Theorem 6 and obtain a fuzzy sum rule for convex functions. The following result corresponds to [113, Theorem 3].

Theorem 9. *Let X be a Banach space and $f_i : X \to \overline{\mathbb{R}}$ be convex and lower semi-continuous functions and $\bar{x} \in \mathrm{dom}(f_1 + f_2)$. Then*

$$\partial(f_1 + f_2)(\bar{x}) = \limsup_{\substack{x_i \to \bar{x} \\ f_i(x_i) - \langle x_i^*, x_i - \bar{x} \rangle \to f_i(\bar{x})}} \partial f_1(x_1) + \partial f_2(x_2) \tag{II.1.15}$$

Proof. We will only prove the inclusion to the right, since the reverse is direct. Take $\bar{x}^* \in \partial(f_1 + f_2)(\bar{x})$ and let $\varepsilon > 0$ and L a finite dimensional subspace of X. Then by Theorem 8 there exist $\bar{x}_i^* \in \partial_\varepsilon f_i(\bar{x})$ such that

$$\bar{x}^* \in \bar{x}_1^* + \bar{x}_2^* + V_{L,\varepsilon}^*.$$

We can apply Theorem 6 with $\lambda = \sqrt{\varepsilon}$ to obtain that there exist x_i and $x_i^* \in \partial f(x_i)$ such that $\|x_i - \bar{x}\| \leq \sqrt{\varepsilon}$ and $\|x_i^* - \bar{x}_i^*\| \leq \sqrt{\varepsilon}$ and $x_i^* \in \partial_{2\varepsilon} f_i(\bar{x})$. Then obviously $\bar{x}^* \in x_1^* + x_2^* + V_{L,\varepsilon+\sqrt{\varepsilon}}^*$ and moreover

$$0 \leq \langle x_i^*, x_i - \bar{x} \rangle - f_i(x_i) + f_i(\bar{x}) \leq 2\varepsilon,$$

which proves that $\bar{x}^*$ belongs to the right hand side of (II.1.15). □

Remark 6. *Under the same assumption of Theorem 9, but using a different technique (which is also useful in the non-convex case) it is possible to show the following formula given in [114]*

$$\partial(f_1 + f_2)(\bar{x}) = \limsup_{\substack{x_i \to_{f_i} \bar{x} \\ \langle x_i^*, x_i - \bar{x} \rangle \to 0 \\ \|x_i^*\| \|x_1 - x_2\| \to 0}} \partial f_1(x_1) + \partial f_2(x_2) \tag{II.1.16}$$

II.1.3 Supremum Rules

Convexity of functions is a property that is preserved under pointwise maximum or supremum, while smoothness is not.

A first supremum rule is related with a specific kind of supremum function, the conjugate. Given a proper function $f : X \to \overline{\mathbb{R}}$ we define the conjugate of f as the function $f^* : X^* \to \overline{\mathbb{R}}$ defined by

$$f^*(x^*) := \sup_x \langle x^*, x \rangle - f(x),$$

which is a convex function on X^* even if $f : X \to \overline{\mathbb{R}}$ is not necessarily convex. The biconjugate $f^{**} : X \to \overline{\mathbb{R}}$ is defined as

$$f^{**}(x) := \sup_{x^*} \langle x^*, x \rangle - f^*(x^*),$$

and coincides with $\overline{\mathrm{co}} f$, which can be defined as the function whose epigraph is $\overline{\mathrm{co}}(\mathrm{epi}\, f)$, provided the conjugate is proper, for instance.

The following result related with the subdifferential of a conjugate function will be useful in what follows.

Lemma 8. *Given a function $f : X \to \overline{\mathbb{R}}$ such that $\mathrm{dom}\, f^* \neq \emptyset$, we have that for all $\varepsilon > 0$ and $x^* \in Y$*

$$\partial_\varepsilon f^*(x^*) = \mathrm{cl}\left(\sum_{i=1}^k \lambda_i (\partial_{\varepsilon_i} f)^{-1}(x^*) \mid \lambda \in \Delta_k,\ \sum_{i=1}^k \lambda_i \varepsilon_i \leq \varepsilon,\ \varepsilon_i \geq 0,\ k \geq 1 \right)$$

where $\Delta_k := \{\lambda \in \mathbb{R}_+^k : \sum_{i=1}^k \lambda_i = 1\}$.

Proof. Assume first that $x = \sum_{i=1}^k \lambda_i x_i$ with $x_i \in (\partial_{\varepsilon_i} f)^{-1}(x^*)$, $\lambda \in \Delta_k$ and $\sum_{i=1}^k \lambda_i \varepsilon_i \leq \varepsilon$, $\varepsilon_i \geq 0$, and $k \in \mathbb{N}$. Then $x^* \in \partial_{\varepsilon_i} f(x_i)$ and, so,

$$\langle x^*, y - x_i \rangle \leq f(y) - f(x_i) + \varepsilon_i, \ \forall y \in X, \ \forall i = 1, \cdots, k.$$

Multiplying this inequality by λ_i and summing up over i, and using the fact that $f^{**} = \overline{\text{co}} f$ we obtain

$$\langle x^*, y - x \rangle \leq f(y) + \sum_{i=1}^k \lambda_i(-f(x_i) + \varepsilon_i) \leq f(y) - f^{**}(x) + \varepsilon.$$

Hence, $f^*(x^*) + f^{**}(x) \leq \langle x^*, x \rangle + \varepsilon$, we get $x \in \partial_\varepsilon f^*(x^*)$ and we obtain the inclusion "$\supset$".

To establish the converse inclusion let us take $x \in \partial_\varepsilon f^*(x^*)$. Given $\delta > 0$ and $V \in \mathcal{N}_X(0)$, then

$$f^*(x^*) + \overline{\text{co}} f(x) \leq \langle x^*, x \rangle + \varepsilon, \tag{II.1.17}$$

Thus there are elements $x_1, \cdots, x_k \in \text{dom}\, f$, $\lambda \in \Delta_k$, and $k \in \mathbb{N}$, such that $x - \sum_{i=1}^k \lambda_i x_i \in V$ and $\overline{\text{co}} f(x) \geq \sum_{i=1}^k \lambda_i f(x_i) - \delta$. Thus, taking $\varepsilon_i := f(x_i) - \overline{\text{co}} f(x) + \langle x^*, x - x_i \rangle + \varepsilon$, we obtain $\sum_{i=1}^k \lambda_i \varepsilon_i \leq \varepsilon + \delta$ and $x^* \in \partial_{\varepsilon_i} f(x_i)$. Hence, $x \in \sum_{i=1}^k \lambda_i (\partial_{\varepsilon_i} f)^{-1}(x^*) + V$ and we conclude due to the arbitrariness of $\delta > 0$ and V, using Proposition (8). □

We give now a formula for the ε-subdifferential of the supremum function that extends and improves [78, Theorem 1].

Theorem 10. *Given a set T and $f_t \in \Gamma_0(X)$, $t \in T$, we assume that $f = \sup_{t \in T} f_t$. Then for every $x \in X$ and $\varepsilon > 0$ we have*

$$\partial_\varepsilon f(x) = \text{cl} \left(\bigcup_{\substack{\lambda \in \Delta_k,\ t_i \in T,\ \beta_i \geq 0,\ k \geq 1 \\ \sum_{i=1}^k \lambda_i (\beta_i - f_{t_i}(x) + f(x)) \leq \varepsilon}} \sum_{i=1}^k \lambda_i \partial_{\beta_i} f_{t_i}(x) \right).$$

Proof. Let us first set $g := \inf_{t \in T} f_t^*$, so that

$$g^* = \sup_{t \in T} f_t^{**} = \sup_{t \in T} f_t = f$$

and $\partial_\varepsilon f(x) = \partial_\varepsilon g^*(x)$. Then, according to Lemma 8, for every $x \in X$ and $\varepsilon > 0$

$$\partial_\varepsilon f(x) = \text{cl} \left(\sum_{i=1}^k \lambda_i (\partial_{\varepsilon_i} g)^{-1}(x) \mid \lambda \in \Delta_k, \ \sum_{i=1}^k \lambda_i \varepsilon_i \leq \varepsilon, \ \varepsilon_i \geq 0, \ k \geq 1 \right). \tag{II.1.18}$$

To establish the inclusion "$\subset$" of the theorem we pick $x_i^* \in (\partial_{\varepsilon_i} g)^{-1}(x)$, $i \leq k$, where $\varepsilon_i \geq 0$ and $k \in \mathbb{N}$ are such that $\sum_{i=1}^k \lambda_i \varepsilon_i \leq \varepsilon$ for some $\lambda \in \Delta_k$. Then $x \in \partial_{\varepsilon_i} g(x_i^*)$ and we get

$$\inf_{t \in T} f_t^*(x_i^*) + f(x) = g(x_i^*) + g^*(x) \leq \langle x_i^*, x \rangle + \varepsilon_i.$$

Next, for $\gamma > 0$ by choosing $t_i \in T$ such that $\inf_{t \in T} f_t^*(x_i^*) \geq f_{t_i}^*(x_i^*) - \gamma$ we obtain

$$f_{t_i}^*(x_i^*) + f(x) \leq \inf_{t \in T} f_t^*(x_i^*) + f(x) + \gamma \leq \langle x_i^*, x \rangle + \varepsilon_i + \gamma, \tag{II.1.19}$$

Therefore, by defining $\beta_i := f_{t_i}^*(x_i^*) + f_{t_i}(x) - \langle x_i^*, x\rangle \geq 0$ we obviously have $x_i^* \in \partial_{\beta_i} f_{t_i}(x)$ and moreover

$$\begin{aligned}\sum_{i=1}^{k} \lambda_i(\beta_i - f_{t_i}(x) + f(x)) &\leq \sum_{i=1}^{k} \lambda_i \varepsilon_i + \gamma \\ &\leq \varepsilon + \gamma,\end{aligned}$$

and the inclusion "$\subset$" follows thanks to (II.1.18) and relation (II.1.6).

Conversely, to prove the inclusion "$\supset$"we take $x^* = \sum_{i=1}^{k} \lambda_i x_i^*$, where $x_i^* \in \partial_{\beta_i} f_{t_i}(x)$, $\delta > \varepsilon$, $k \in \mathbb{N}$, and $\lambda \in \Delta_k$, with $\sum_{i=1}^{k} \lambda_i(\beta_i - f_{t_i}(x)) + f(x) \leq \delta$. Then

$$\begin{aligned}f^*(x^*) &\leq \sum_{i=1}^{k} \lambda_i f^*(x_i^*) \leq \sum_{i=1}^{k} \lambda_i f_{t_i}^*(x_i^*) \\ &\leq \sum_{i=1}^{k} \lambda_i(\langle x, x_i^*\rangle + \beta_i - f_{t_i}(x)) \\ &\leq \langle x, x^*\rangle + \delta - f(x).\end{aligned}$$

so that $x^* \in \partial_\delta f(x)$. Thus, the desired inclusion follows form the closedness of $\partial_\delta f(x)$ and the arbitrariness of $\delta > \varepsilon$. □

Corollary 1. *Assume that $f = \sup_{n\geq 1} f_n$, with $(f_n)_n$ being a non-decreasing sequence of proper, convex and lsc functions. Then for every $x \in X$ and $\varepsilon > 0$ we have*

$$\partial_\varepsilon f(x) = \limsup_{n\to+\infty} \partial_\varepsilon f_n(x)$$

Proof. Take $\xi \in \partial_\varepsilon f(x)$ and fix $\delta > 0$. According to Theorem 10, for any $V \in \mathcal{N}_{w^*}(0)$ we have that

$$\xi \in \sum_{i=1}^{k} \lambda_i \xi_i + V$$

for some $\xi_i \in \partial_{\beta_i} f_{n_i}(x)$ and $\lambda \in \Delta_k$ ($k \in \mathbb{N}$), where $\beta_i \geq 0$ and $n_i \in \mathbb{N}$ ($i = 1, ..., k$) are such that $\sum_{i=1}^{k} \lambda_i(\beta_i - f_{n_i}(x) + f(x)) \leq \varepsilon + \frac{\delta}{2}$. Put $m_0 := \max_{i=1}^{k} n_i \geq 1$. By writing the relation $\xi_i \in \partial_{\beta_i} f_{n_i}(x)$ into an inequality form and, next, summing over i, we obtain, for all $y \in x + \frac{\delta}{2} V^\circ$ (hence, $\sigma_V(y - x) \leq \frac{\delta}{2}$),

$$\begin{aligned}\langle \xi, y - x\rangle &\leq \sum_{i=1}^{k} \lambda_i \langle \xi_i, y - x\rangle + \frac{\delta}{2} \\ &\leq \sum_{i=\overline{1,k}} \lambda_i(f_{n_i}(y) - f_{n_i}(x) + \beta_i) + \frac{\delta}{2} \\ &\leq f_{m_0}(y) - f(x) + \varepsilon + \delta.\end{aligned}$$

Then,

$$\langle \xi, y - x\rangle \leq f_n(y) - f_n(x) + \varepsilon + \delta \quad \text{for all } n \geq m_0,$$

so that $\xi \in \partial_{\varepsilon+\delta}(f_n + \chi_{x+\frac{\delta}{2}V^\circ})(x)$, for all $n \geq m_0$. Taking into account the sum rule of ε-subdifferentials (e.g., [58])

$$\partial_{\varepsilon+\delta}(f_n + \chi_{x+\frac{\delta}{2}V^\circ})(x) \subset \partial_{\varepsilon+\delta} f_n(x) + \frac{2(\varepsilon+\delta)}{\delta}V.$$

Hence, as V was arbitrarily chosen, we deduce that

$$\xi \in \limsup_{n\to+\infty} \partial_\delta f_n(x).$$

which finishes the proof, since the opposite inclusion "$\supset$" always holds. □

We now recover the case of the maximum of finitely many convex functions; see, e.g., [124].

Corollary 2. *Consider a family of finitely many proper lsc convex functions $f_1, \cdots, f_n$ and $f = \max_{i=1}^n f_i$. Then for every $x \in X$ and $\varepsilon \geq 0$*

$$\partial_\varepsilon f(x) = \bigcup \left\{ \partial_\eta \left(\sum_{i=1}^n \lambda_i f_i(x) \right)(x) \mid \lambda \in \Delta_n,\ \eta = \sum_{i=1}^n \lambda_i f_i(x) - f(x) + \varepsilon \geq 0 \right\},$$

and in particular

$$\partial f(x) = \bigcup \left\{ \partial \left(\sum_{i=1}^n \lambda_i f_i(x) \right)(x) \mid \lambda \in \Delta_n, \lambda_i = 0,\ \forall i \notin A(x) \right\},$$

where $A(x) = \{i = 1, ..., n : f_i(x) = f(x)\}$.

Proof. According to Theorem 10, we have that

$$\begin{aligned}\partial_\varepsilon f(x) &\subset \mathrm{cl}\left\{ \sum_{i=1}^n \lambda_i \partial_{\beta_i} f_i(x) \mid \lambda \in \Delta_n,\ \sum_{i=1}^n \lambda_i(\beta_i - f_i(x)) + f(x) \leq \varepsilon \right\} \\ &\subset \mathrm{cl}\left\{ \partial_\varepsilon (\sum_{i=1}^n \lambda_i f_i)(x) \mid \lambda \in \Delta_n,\ \sum_{i=1}^n \lambda_i f_i(x) \geq f(x) - \varepsilon \right\}. \end{aligned} \quad \text{(II.1.20)}$$

The conclusion follows by using the compactness of the set Δ_n. □

At this point we can consider the case $\varepsilon = 0$ and think of the question if we can estimate the exact subdifferential of a supremum function in terms of the exact subdifferential (maybe not only at the reference point). This question can be addressed in a similar way as in Theorem 9 by applying Theorems 10 and 6, and Remark 5.

Theorem 11. *Assume X is a Banach space. Then*

$$\partial f(\bar{x}) = \limsup_{\substack{k\in\mathbb{N}, t_i \in T, \lambda \in \Delta_k \\ x_i \to_{f_{t_i}} \bar{x}, \\ \lambda_i(f_{t_i}(\bar{x}) - f(\bar{x})) \to 0 \\ \lambda_i \langle x_i^*, x_i - \bar{x}\rangle \to 0}} \sum_{i=1}^k \lambda_i \partial f_{t_i}(x_i) \quad \text{(II.1.21)}$$

II.1.4 Normal Cone and Approximate Normal Sets

Let $C \subset X$ be a non-empty convex set and $x \in C$. A functional $x^* \in X^*$ is said to be normal to C if $\langle x^*, y - x\rangle \leq 0$ for all $y \in C$. The set of all normals to C is denoted by $N_C(x)$. More generally, given $\varepsilon \geq 0$ the set of all $x^* \in X^*$ satisfying $\langle x^*, y - x\rangle \leq \varepsilon$, for all $y \in C$ is the ε-normal set to C and is denoted by $N_C^\varepsilon(x)$. If $\varepsilon = 0$ then of course $N_C^0(x)$ coincide with $N_C(x)$ for every $x \in C$.

The normals to a convex set C can be viewed as subdifferentials of the indicator of the set C. Recall that the indicator of C is the function $\chi_C : X \to \overline{\mathbb{R}}$ defined as $\chi_C(x) = 0$ for $x \in C$ and $\chi_C(x) := +\infty$ otherwise. We observe then that $\partial_\varepsilon \chi_C(x) = N_C^\varepsilon(x)$ for any $x \in X$. From this fact we deduce the following useful result.

Proposition 9. *Given $f \in \Gamma_0(X)$, $\lambda \in \mathbb{R}$, and $x \in [f \leq \lambda]$, for every $\varepsilon > 0$ we have that*

$$N^\varepsilon_{[f\leq\lambda]}(x) = \operatorname{cl}\left(\bigcup\nolimits_{\alpha>0} N^\varepsilon_{[f\leq\lambda+\alpha]}(x)\right).$$

Proof. Let us consider the functions $f_n := \chi_{[f\leq\lambda+1/n]}$ for $n \in \mathbb{N}$ and $f := \chi_{[f\leq\lambda]}$, which are all in $\Gamma_0(X)$, and satisfy that f_n is an increasing family that converges pointwise to f. Observe that since $f_n(x) = f(x) = 0$ for all $n \geq 1$, the sequence of sets $(\partial_\varepsilon f_n(x))_n$ is non-decreasing. We apply the supremum rule of Corollary 1 and obtain

$$\partial_\varepsilon f(x) = \limsup_{n\to+\infty} \partial_\varepsilon f_n(x) = \operatorname{cl}\left(\bigcup\nolimits_{n\geq 1} \partial_\varepsilon f_n(x)\right).$$

which corresponds to

$$N^\varepsilon_{[f\leq\lambda]}(x) = \operatorname{cl}\left(\bigcup\nolimits_{n\geq 1} N^\varepsilon_{[f\leq\lambda+\frac{1}{n}]}(x)\right),$$

as we wanted to prove. □

Lemma 9. *If $f(\bar{x}) \leq \lambda < +\infty$ and the Slater condition holds at λ ($\exists x_0 : f(x_0) < \lambda$), then*

$$N^\delta_{[f\leq\lambda]}(\bar{x}) \subset \bigcup_{\mu\geq 0} \partial_{\delta+\mu(f(\bar{x})-\lambda)}\,(\mu f)\,(\bar{x}).$$

Proof. Given $\xi \in N^\delta_{[f\leq\lambda]}(\bar{x})$, we define the proper lsc convex function $\varphi : X \to \overline{\mathbb{R}}$ as

$$\varphi(x) := \max\left\{f(x) - \lambda, \delta - \langle\xi, x - \bar{x}\rangle\right\}.$$

From the definition of the δ-normal set we have that $\varphi(x) \geq 0$ for all $x \in X$. Thus, since $\varphi(\bar{x}) = \delta$, it follows that $0 \in \partial_\delta\varphi(\bar{x})$. According to Corollary 2,

$$\partial_\delta\varphi(\bar{x}) \subset \bigcup_{\substack{\alpha\in[0,1],\eta\in[0,\delta]\\ \eta\leq\alpha\delta+(1-\alpha)(f(\bar{x})-\lambda)}} \partial_\eta((1-\alpha)(f-\lambda) + \alpha(\delta - \xi + \langle\xi,\bar{x}\rangle))(\bar{x}),$$

and so there exist $\alpha \in [0,1]$ and $\eta \in [0, \alpha\delta+(1-\alpha)(f(\bar{x})-\lambda)]$ such that $\alpha\xi \in \partial_\eta(1-\alpha)f(\bar{x})$. If $\alpha = 0$, then $\eta = 0$ (as $f(\bar{x}) \leq \lambda$) and we get $0 \in \partial f(\bar{x})$, which contradicts the Slater condition. So, $\alpha > 0$ and the number $\mu := \frac{1-\alpha}{\alpha}$ (≥ 0) is well-defined and satisfies $\frac{\eta}{\alpha} \leq \delta + \mu(f(\bar{x}) - \lambda)$ together with

$$\xi \in \frac{1}{\alpha}\partial_\eta(1-\alpha)f(\bar{x}) \subset \partial_{\frac{\eta}{\alpha}}\frac{1-\alpha}{\alpha}f(\bar{x}) \subset \partial_{\delta+\mu(f(\bar{x})-\lambda)}\,(\mu f)\,(\bar{x}).$$

□

Theorem 12. *If $f(\bar{x}) \leq \lambda < +\infty$ then for $\delta > 0$ we have*

$$N^{\delta}_{[f\leq\lambda]}(\bar{x}) = \mathrm{cl}\left(\bigcup_{\mu\geq 0} \partial_{\delta+\mu(f(\bar{x})-\lambda)}\left(\mu f\right)(\bar{x})\right) \tag{II.1.22}$$

and consequently for all $\lambda \in \mathbb{R}$

$$N_{[f\leq\lambda]}(\bar{x}) = \bigcap_{\varepsilon>0} \mathrm{cl}\left(\bigcup_{\mu\geq 0} \partial_{\varepsilon+\mu(f(\bar{x})-\lambda)}\left(\mu f\right)(\bar{x})\right). \tag{II.1.23}$$

Proof. First it is obvious that for any $\alpha > 0$, f satisfies the Slater condition at $\lambda+\alpha$, so that by Lemma 9 we get

$$\begin{aligned} N^{\delta}_{[f\leq\lambda+\alpha]}(x) &\subset \bigcup_{\mu\geq 0} \partial_{\delta+\mu(f(\bar{x})-\lambda-\alpha)}\left(\mu f\right)(\bar{x}) \\ &\subset \bigcup_{\mu\geq 0} \partial_{\delta+\mu(f(\bar{x})-\lambda)}\left(\mu f\right)(\bar{x}). \end{aligned}$$

Thus, using Proposition 9 we conclude that

$$\begin{aligned} N^{\delta}_{[f\leq\lambda]}(x) &= \mathrm{cl}\left(\bigcup\nolimits_{\alpha>0} N^{\delta}_{[f\leq\lambda+\alpha]}(x)\right) \\ &\subset \mathrm{cl}\left(\bigcup_{\mu\geq 0} \partial_{\delta+\mu(f(\bar{x})-\lambda)}\left(\mu f\right)(\bar{x})\right). \end{aligned}$$

The converse implication is trivial. □

Remark 7. *In Theorem 12, if the space X is a reflexive Banach space then the above closure can be taken with respect to the strong topology in X^*.*

We can obtain a formula for a normal cone to the intersection of an arbitrary family of sublevel sets of convex functions by simply combining the supremum formula in Theorem 10 and the normal cone formula in Theorem 12,

Theorem 13. *Let $f_t \in \Gamma_0(X)$ and $\lambda_t \in \mathbb{R}$, $t \in T$ and let $S_t := [f_t \leq \lambda_t]$. Then for any $\delta > 0$,*

$$N^{\delta}_{\bigcap\limits_{t\in T} S_t}(\bar{x}) = \mathrm{cl}\left(\sum_{i=1}^{k} \mu_i \partial_{\varepsilon_i} f_{t_i}(\bar{x}),\ k \in \mathbb{N},\ t_i \in T,\ \mu_i > 0, \sum_{i=1}^{k} \mu_i(\lambda - f_{t_i}(x) + \varepsilon_i) \leq \delta\right) \tag{II.1.24}$$

Proof. We observe first that $\bigcap_{t\in T} S_t = [f \leq 0]$ for $f := \sup_{t\in T} f_t - \lambda_t$. Then by Theorem 12 we have

$$N^{\delta}_{\bigcap\limits_{t\in T} S_t}(\bar{x}) = N^{\delta}_{[f\leq 0]}(\bar{x}) = \overline{\bigcup_{\mu>0} \mu\partial_{\delta/\mu+f(\bar{x})} f(\bar{x})}.$$

Second, we apply Theorem 10 to estimate $\partial_{\varepsilon} f(\bar{x})$ with $\varepsilon = \delta/\mu + f(\bar{x})$. We have that

$$\partial_{\delta/\mu+f(\bar{x})} f(\bar{x}) = \mathrm{cl}\left(\sum_{i=1}^{k} \lambda_i \partial_{\beta_i} f_{t_i}(x) :\ \lambda \in \Delta_k, \sum_{i=1}^{k} \lambda_i(\beta_i - f_{t_i}(x) + \lambda_{t_i}) \leq \delta/\mu\right)$$

and the proof easily follows. □

Our formula takes an algebraic form in the following corollary, giving rise to the result given in [58] and originated in [73]. It follows from Lemma 9.

Corollary 3. *If $f(\bar{x}) \leq \lambda < +\infty$ and the Slater condition holds at λ ($\exists x_0 : f(x_0) < \lambda$), then*

$$N_{[f\leq\lambda]}(\bar{x}) = \bigcup_{\mu\geq 0} \partial_{\mu(f(\bar{x})-\lambda)} (\mu f) (\bar{x}).$$

II.1.4.1 A General Formula

We are going to prove a general formula that extends Theorem 12 including also the case $\bar{x} \notin [f \leq \lambda]$. For this purpose we recall the notion of recession cone of a nonempty convex set $C \subset X$, which is defined as the set $C_\infty := \{v \in X : \exists x \in C,\ x + tv \in C,\ \forall t > 0\}$.

Theorem 14. *Let $f \in \Gamma_0(X)$, $\bar{x} \in \operatorname{dom} f$, $\delta \geq 0$ and $\lambda \in \mathbb{R}$. Assume that either $\delta > 0$ or $\lambda < f(\bar{x})$. Then*

$$N^{\delta}_{[f\leq\lambda]\cup(\bar{x}+[f\leq f(\bar{x})]_\infty)}(\bar{x}) = \overline{\bigcup_{\mu>0} \partial_{\delta+\mu(f(\bar{x})-\lambda)}(\mu f)(\bar{x})}, \tag{II.1.25}$$

and consequently

$$N_{[f\leq\lambda]\cup(\bar{x}+[f\leq f(\bar{x})]_\infty)}(\bar{x}) = \bigcap_{\delta'>0} \overline{\bigcup_{\mu>0} \partial_{\delta'+\mu(f(\bar{x})-\lambda)}(\mu f)(\bar{x})}. \tag{II.1.26}$$

Proof. We will only prove the inclusion "$\subset$ " because the converse is straight forward. We pick a $\xi \in N^{\delta}_{[f\leq\lambda]\cup(\bar{x}+[f\leq f(\bar{x})]_\infty)}(\bar{x})$ such that

$$\xi \notin \overline{\bigcup_{\mu>0} \partial_{\delta+\mu(f(\bar{x})-\lambda)}(\mu f)(\bar{x})}. \tag{II.1.27}$$

Since this last set is convex and obviously closed, by the Hahn-Banach Theorem there exist $v \in X$ and $\alpha \in \mathbb{R}$ such that

$$\langle \xi, v\rangle > \alpha \geq \langle x^*, v\rangle, \quad \text{for all } x^* \in \bigcup_{\mu>0} \partial_{\delta+\mu(f(\bar{x})-\lambda)}(\mu f)(\bar{x});$$

moreover, because $\partial_{\delta+\mu(f(\bar{x})-\lambda)}(\mu f)(\bar{x}) = \mu \partial_{\delta/\mu+f(\bar{x})-\lambda} f(\bar{x})$ taking $\mu \to 0$ we get $\alpha \geq 0$ and, so, one may suppose that $\alpha = \delta$. Hence, the inequalities above read: for all $\mu > 0$

$$\langle \xi, v\rangle > \delta \geq \langle x^*, v\rangle, \quad \text{for all } x^* \in \partial_{\delta+\mu(f(\bar{x})-\lambda)}(\mu f)(\bar{x}); \tag{II.1.28}$$

that is $\mu f'_{\delta/\mu+f(\bar{x})-\lambda}(\bar{x}, v) \leq \delta$, or, equivalently, for all $\varepsilon \geq 0$ (setting $\varepsilon = \delta/\mu$)

$$\inf_{t>0} \frac{f(\bar{x}+tv) - \lambda + \varepsilon}{t} \leq \varepsilon. \tag{II.1.29}$$

Let S be the set-valued map of Lemma 10. First, we assume that $S(f(\bar{x}) - \lambda) \cap [-1, 0] \neq \emptyset$. If $S(f(\bar{x}) - \lambda)$ ($\neq \emptyset$ by Lemma 10) contains a point $s_0 \in [-1, 0[$, then $t_0 := \frac{-1}{s_0} \geq 1$ satisfies

$f(\bar{x}+t_0v)-\lambda \leq 0$, by (II.1.29), and this leads us to the following contradiction (recall (II.1.28)),

$$\langle \xi, v\rangle = t_0^{-1}\langle \xi, \bar{x}+t_0v-\bar{x}\rangle \leq t_0^{-1}\delta \leq \delta < \langle \xi, v\rangle. \tag{II.1.30}$$

If $S(f(\bar{x})-\lambda)$ contains 0, there would exist $t_n \to +\infty$ such that $\lim_{n\to+\infty} t_n^{-1}(f(\bar{x}+t_nv)-\lambda) = R(f(\bar{x})-\lambda) \leq 0$, which shows that

$$\sup_{t>0} t^{-1}(f(\bar{x}+tv)-f(\bar{x})) = \lim_{n\to+\infty} t_n^{-1}(f(\bar{x}+t_nv)-f(\bar{x})) = \lim_{n\to+\infty} t_n^{-1}(f(\bar{x}+t_nv)-\lambda) \leq 0;$$

hence, $v \in [f \leq f(\bar{x})]_\infty$ and we get a contradiction along of (II.1.30).

Now, we suppose that $S(f(\bar{x})-\lambda)\cap[-1,0]=\emptyset$; that is, $s<-1$ for all $s \in S(f(\bar{x})-\lambda)$. Then two cases may occur:

(a) For every $\varepsilon > f(\bar{x})-\lambda$ and $s \in S(\varepsilon)$ we have $s<-1$. In this case we pick an $s_\varepsilon \in S(\varepsilon)$ and put $t_\varepsilon := \frac{-1}{s}$; hence, $t_\varepsilon < 1$, so that

$$\frac{\varepsilon^{-1}(f(\bar{x}+t_\varepsilon v)-\lambda)+1}{t_\varepsilon} \leq 1.$$

Since $f(\bar{x}+\cdot v)$ is bounded from below in $[0,1]$, this last inequality implies that $t_\varepsilon \to 1$ as $\varepsilon \to +\infty$, as well as $\varepsilon^{-1}(f(\bar{x}+t_\varepsilon v)-\lambda) \leq 0$ for ε large enough (because $t_\varepsilon < 1$). Then $f(\bar{x}+v) = \lim_{\varepsilon\to+\infty} f(\bar{x}+t_\varepsilon v) \leq \lambda$ and we get a contradiction as in (II.1.30).

(b) There exist some $\varepsilon_0 > f(\bar{x})-\lambda$ and $s_0 \in S(\varepsilon_0)$ such that $s_0 \geq -1$. Since S is a maximal monotone operator (Lemma 10), it has a convex range and, so, because $s<-1$ for all $s \in S(f(\bar{x})-\lambda)$ while $s_0 \geq -1$, there must exist some $\varepsilon_1 > 0$ such that $-1 \in S(\varepsilon_1)$; that is,

$$f(\bar{x}+v)-\lambda+\varepsilon_1 \leq \varepsilon_1,$$

and we get $\bar{x}+v \in [f \leq \lambda]$, which leads us to a contradiction similar to the one in (II.1.30). Consequently, (II.1.27) is not true and we must have that $\xi \in \overline{\bigcup_{\mu>0} \partial_{\delta+\mu(f(\bar{x})-\lambda)}(\mu f)(\bar{x})}$. □

Then next lemma is somehow related with the sets of minimizers of

$$t \to \frac{f(x+tv)-f(x)+\varepsilon}{t} \tag{II.1.31}$$

as a function of ε. But in fact this set could be empty. This can be handled by considering instead the following sets

$$S(\varepsilon) := \{-\lim_{n\to+\infty} t_n^{-1} \mid \lim_{n\to+\infty} t_n^{-1}(f(\bar{x}+t_nv)-f(\bar{x})+\varepsilon) = f'_\varepsilon(\bar{x};v)\},$$

for $\varepsilon \geq 0$. It is simple to see that if $t \in S(\varepsilon)$ with $-t>0$ then $-t$ is a minimizer of the function (II.1.31).

Lemma 10. *Given $\bar{x} \in \operatorname{dom} f$ and $v \in X$, we define the set-valued mapping $S:\mathbb{R} \rightrightarrows \mathbb{R}$ as $S(\varepsilon)=\emptyset$ for $\varepsilon<0$, and*

$$S(\varepsilon) := \{-\lim_{n\to+\infty} t_n^{-1} \mid \lim_{n\to+\infty} t_n^{-1}(f(\bar{x}+t_nv)-f(\bar{x})+\varepsilon) = f'_\varepsilon(\bar{x};v)\}, \text{ for } \varepsilon \geq 0.$$

Then the following assertions hold:

(i) $S(\varepsilon) \neq \emptyset$ for all $\varepsilon > 0$

(ii) The set $S(0)$ is a possibly empty closed interval of $\mathbb{R}_-$

(iii) When $S(0) = \emptyset$ there exists $s_\varepsilon \in S(\varepsilon)$ such that $s_\varepsilon \to -\infty$ as $\varepsilon \downarrow 0$

(iv) For every $\varepsilon \geq 0$ the set $S(\varepsilon)$ is convex and closed.

Moreover S is a maximal monotone set-valued map.

Proof. (i) If $\varepsilon > 0$, then each sequence (t_n) realizing the infimum in the definition of $R(\varepsilon) := f'_\varepsilon(\bar{x}; v)$ must converge (up to a subsequence) to some $t > 0$ (possibly $t = +\infty$), so that $-\frac{1}{t} \in S(\varepsilon)$ (with the convention that $\frac{1}{+\infty} = 0$).

(ii) If $S(0)$ is a non-empty subset of $\mathbb{R}_-$, then it is closed, by the continuity of function $g = f(\bar{x} + \cdot v) - f(\bar{x})$. If s_0 ($\in S(0) \subset \mathbb{R}_-$) is the maximum element in $S(0)$, then there is a sequence (t_n) of positive numbers such that $s_0 = -\lim_{n\to+\infty} t_n^{-1}$ and $\lim_{n\to+\infty} t_n^{-1}(f(\bar{x} + t_n v) - f(\bar{x})) = f'(\bar{x}; v)$. If $s_0 < 0$, then for $t_0 := \frac{-1}{s_0}$ we get

$$f'(\bar{x}; v) \leq t^{-1}(f(\bar{x} + tv) - f(\bar{x})) \leq t_0^{-1}(f(\bar{x} + t_0 v) - f(\bar{x})) = f'(\bar{x}; v) \quad \text{for all } t \in \left]0, t_0\right];$$

hence, $\left]-\infty, s_0\right] \subset S(0)$. If $s_0 = 0$, then $t_n \to \infty$ and we obtain

$$\begin{aligned} f'(\bar{x}; v) &\leq \inf_{t>0} t^{-1}(f(\bar{x} + tv) - f(\bar{x}) \\ &\leq \sup_{t>0} t^{-1}(f(\bar{x} + tv) - f(\bar{x})) \\ &= \lim_{n\to+\infty} t_n^{-1}(f(\bar{x} + t_n v) - f(\bar{x})) = f'(\bar{x}; v); \end{aligned}$$

that is, $\left]-\infty, 0\right] \subset S(0)$. Thus, in both cases we have $\left]-\infty, s_0\right] \subset S(0)$.

(iii) Assume that $S(0)$ is empty, Then, as $\varepsilon \downarrow 0$, there always exist positive numbers $s_\varepsilon \in S(\varepsilon)$ such that $s_\varepsilon \to -\infty$. In fact, given an $\varepsilon > 0$, we choose $t_\varepsilon > 0$ such that

$$R(\varepsilon) + \varepsilon = \frac{g(t_\varepsilon) + \varepsilon}{t_\varepsilon} \geq \frac{g(t_\varepsilon)}{t_\varepsilon} > R(0); \tag{II.1.32}$$

such an t_ε always exists, because, for otherwise, we would have $\varepsilon_i \downarrow 0$ and $t_i^n \to_n +\infty$ such that $\lim_{n\to+\infty}(t_i^n)^{-1}(f(\bar{x} + t_i^n v) - f(\bar{x}) + \varepsilon_i) = f'_{\varepsilon_i}(\bar{x}; v)$ for all i. This would yield $\sup_{t>0} t^{-1}(f(\bar{x} + tv) - f(\bar{x})) \leq \lim_{n\to+\infty}(t_i^n)^{-1}(f(\bar{x} + t_i^n v) - f(\bar{x}) + \varepsilon_i) = f'_{\varepsilon_i}(\bar{x}; v)$ for all i and, so, we get $\sup_{t>0} t^{-1}(f(\bar{x} + tv) - f(\bar{x})) \leq f'(\bar{x}; v)$, which gives rise to $S(0) = \left]-\infty, 0\right]$, a contradiction. Consequently, (II.1.32) makes sense, so that the vacuity of $S(0)$ together with the continuity of f leads us to $t_\varepsilon \to 0^+$ (recall Lemma 7). In other words, $s_\varepsilon = -{t_\varepsilon}^{-1}$ goes to $-\infty$ as ε goes to 0.

(iv) Since the function $t \to t^{-1}(g(t) + \varepsilon)$ (for $\varepsilon > 0$) is quasi-convex (has convex sublevel sets) and continuous, the set $A \subset [0, +\infty]$ defined as

$$A := \{t \geq 0 \mid \exists t_n \to t \text{ s.t. } \lim_{n\to+\infty} t_n^{-1}(f(\bar{x} + t_n v) - f(\bar{x}) + \varepsilon) = R(\varepsilon)\}$$

is convex and closed. Moreover, $0 \notin A$ and the image of A by the function $\rho(t) := -\frac{1}{t}$ $(t > 0)$ coincides with $S(\varepsilon)$. Hence, since function ρ is monotone and continuous we conclude that $S(\varepsilon)$ is convex and closed.

Now let us show that S is monotone. We pick $(\varepsilon_i, s_i) \in S$ (the graph of S), $i = 0, 1$, with $0 < \varepsilon_0 < \varepsilon_1$. Then for each $i = 0, 1$ there is a sequence $(t_i^n)^{-1} \to -s_i$ such that $\lim_{n\to+\infty}(t_i^n)^{-1}(f(\bar{x} + t_i^n v) - f(\bar{x}) + \varepsilon_i) = R(\varepsilon_i)$ (recall Lemma 7); hence $t_i^n > 0$. Writing

$$\begin{aligned}
R(\varepsilon_1) &= \lim_{n\to\infty}(t_{\varepsilon_1}^n)^{-1}(f(t_{\varepsilon_1}^n) + \varepsilon_1) \\
&= \lim_{n\to\infty}((t_{\varepsilon_1}^n)^{-1}(f(t_{\varepsilon_1}^n) + \varepsilon_0) + (t_{\varepsilon_1}^n)^{-1}(\varepsilon_1 - \varepsilon_0)) \\
&\geq R(\varepsilon_0) + \liminf_{n\to\infty}(t_{\varepsilon_1}^n)^{-1}(\varepsilon_1 - \varepsilon_0) \\
&= \lim_{n\to\infty}((t_{\varepsilon_0}^n)^{-1}(f(t_{\varepsilon_0}^n) + \varepsilon_0) + \liminf_{n\to\infty}(t_{\varepsilon_1}^n)^{-1}(\varepsilon_1 - \varepsilon_0) \\
&\geq \lim_{n\to\infty}((t_{\varepsilon_0}^n)^{-1}(f(t_{\varepsilon_0}^n) + \varepsilon_1) + \liminf_{n\to\infty}(t_{\varepsilon_0}^n)^{-1}(\varepsilon_0 - \varepsilon_1) + \liminf_{n\to\infty}(t_{\varepsilon_1}^n)^{-1}(\varepsilon_1 - \varepsilon_0) \\
&\geq R(\varepsilon_1) + \liminf_{n\to\infty}((t_{\varepsilon_0}^n)^{-1} - (t_{\varepsilon_1}^n)^{-1})(\varepsilon_0 - \varepsilon_1),
\end{aligned}$$

we deduce that $(\varepsilon_0 - \varepsilon_1)(s_0 - s_1) \geq 0$, and the monotonicity of S follows. To check the maximality of S, we observe that the function $\psi : \mathbb{R} \to \overline{\mathbb{R}}$, de fined as $\psi(\varepsilon) := \inf\{s \mid s \in S(\varepsilon)\}$ for $\varepsilon \geq 0$ and $-\infty$ otherwise, is non-decreasing (and satisfies $\lim_{\varepsilon\downarrow 0}\psi(\varepsilon) = -\infty$); so, it possesses left and right-limits ψ_- and ψ_+ everywhere in $\mathbb{R}_+$. Then, given an $\varepsilon_0 > 0$, by using [124, Theorem 2.1.7] the function φ defined on $\mathbb{R}$ as $\varphi(\tau) := \int_{\varepsilon_0}^{\tau}\psi(s)ds$ is a proper, lsc convex function with $\mathbb{R}_+ \subset \operatorname{dom}\varphi \subset \overline{\mathbb{R}}$, and $\partial\varphi(\tau) = [\psi_-(\tau), \psi_+(\tau)]$ for every $\tau > 0$, while $\partial\varphi(0) =]-\infty, \psi_+(0)]$, and $\partial\varphi(\tau) = \emptyset$ for all $\tau < 0$. Since $S(\varepsilon)$ is convex and closed for every $\varepsilon \geq 0$, by Lemma 10, we infer that $\partial\varphi \subset S$ and, so, by Rockafellar's theorem [102] we infer that $S = \partial\varphi$ and, in particular, S is maximal monotone. □

Corollary 4. *Given $f \in \Gamma_0(X)$ and $\bar{x} \in \operatorname{dom} f$, we assume that $[f \leq \lambda] \neq \emptyset$ and $f(\bar{x}) > \lambda$. Then we have*

$$N_{[f\leq\lambda]\cup(\bar{x}+[f\leq f(\bar{x})]_\infty)}(\bar{x}) = \overline{\mathbb{R}_+\partial_{f(\bar{x})-\lambda}f(\bar{x})}.$$

Proof. The result is an immediate consequence of Theorem 14. □

In the following corollary we consider the case in which the sublevel set $[f \leq \lambda]$ is empty.

Corollary 5. *Given $f \in \Gamma_0(X)$, $\bar{x} \in \operatorname{dom} f$, and $\lambda \in \mathbb{R}$, we assume that $[f \leq \lambda] = \emptyset$. Then, for every $\varepsilon \geq f(\bar{x}) - \lambda$,*

$$N_{\bar{x}+[f\leq f(\bar{x})]_\infty}(\bar{x}) = \overline{\mathbb{R}_+\partial_\varepsilon f(\bar{x})}.$$

Proof. The inclusion $\subset$ follows from Theorem 14, since

$$\bigcup_{\mu>0}\partial_{\mu(f(\bar{x})-\lambda)}(\mu f)(\bar{x}) \subset \mathbb{R}_+\partial_{f(\bar{x})-\lambda}f(\bar{x}) \subset \mathbb{R}_+\partial_\varepsilon f(\bar{x}).$$

For the converse inclusion it is enough to prove that for any $\xi \in \partial_\varepsilon f(\bar{x})$ and $v \in [f \leq f(\bar{x})]_\infty$ we have $\langle\xi, v\rangle \leq 0$. We have

$$\langle\xi, v\rangle \leq f(\bar{x} + v) - f(\bar{x}) + \varepsilon \leq \varepsilon.$$

Now since $[f \leq f(\bar{x})]_\infty$ is a cone we deduce that $\langle\xi, v\rangle \leq 0$, which finishes the proof. □

Chapter II.2

Fuzzy Calculus Rules in Non-smooth Analysis

II.2.1 Introduction

Weak fuzzy subdifferential calculus rules have the advantage over the exact ones, because they do not require any qualification condition, this being true for lower semi-continuous functions defined on Asplund spaces. Maybe one of the most old and known rules is the fuzzy sum rule, which was first proved by Ioffe in 1984 in the case of finite dimension, and subsequently extended to infinite dimensional Asplund spaces.

But it was not until recent years (see [22, 89]) that KKT-type/non-degenerate fuzzy necessary (sub)optimality conditions for non-smooth programs were proved to be valid without any constraint qualification. Before this and even in the finite dimensional case, only FJ-type/degenerate fuzzy optimality conditions were known to be valid without constraint qualifications (see [84, 128]). The delay might have been related to the lack of a representation of the normal cone to a (sub)level set of a (lower semi)continuous function in terms of the subgradients of the function, under the premise of no qualification condition.

In [22], using what we here call the *epigraph approach* for constrained optimization, the authors proved the validity of some non-degenerate fuzzy optimality conditions for a non-smooth program without any constraint qualification. In [89], the authors proved, first, that the representation of the normal cone to a (sub)level was valid without any qualification condition, and second, the validity of non-degenerate fuzzy optimality condition without any constraint qualifications. Furthermore, they were able to include an approximate complementarity slackness condition, both in the representation of the normal cone and in the fuzzy optimality condition.

In this chapter we use the epigraph approach to prove a stronger version of the non-degenerate weak fuzzy (sub)optimality condition for a non-smooth program without any constraint qualification. Our version is in fact stronger since it includes the approximate complementarity slackness condition for inequality constraints (proved in [89]) and at the same time the convergence of the values of the inequality constraint functions. These two conditions together allow us to distinguish between active and non-active inequality constraints. In fact, for non-active inequality constraints we deduce that the corresponding Lagrange multiplier must go to zero. Our improvement is greatly important since it is desirable that necessary optimality conditions are as tight as possible.

Additionally, we use our results to prove some other fuzzy calculus rules. Firstly, we give a quite shorter proof of the representation of the normal cone to sublevel sets. Secondly, we give a formula for the subdifferential of the supremum of an arbitrary family of lower semi-continuous functions defined in an Asplund space. This formula is new even in the finite dimensional setting with a finite family of functions. We are aware that Pérez-Aros in a recent work (see [95]) has obtained a similar formula that is equivalent to ours in the case of finitely many functions, while it relaxes the assumptions of uniform Lipschitz continuity of the functions as required in [83, Theorem 3.1].

We apply all the developed machinery to give fuzzy optimality conditions for a bilevel programming problem without any qualification condition. These fuzzy optimality conditions correspond to those of [43], where several constraint qualifications were used.

Along the chapter we give several examples in order to show the improvements of our formulas with respect to previously known ones.

II.2.2 Notation and Preliminaries

We follow the notation and definitions of the book [84], while some of the definitions are in the book consequences of the fact that we restrict ourselves to Asplund spaces.

Let X be an Asplund space with norm $\|\cdot\|$, and X^* its topological dual endowed with the weak* topology (no other topology will be considered in this chapter). Let $\mathbb{B}$ and $\mathbb{B}^*$ denote the unit closed ball in X and in X^*, respectively, let $\mathcal{N}_{w^*}(0)$ be the set of weak* neighborhoods of the origin in X^*, and $\langle , \rangle$ be the duality product in $X \times X^*$.

The weak* topology in X^* is defined in most text books as the weakest topology generated by the sub-base of neighborhoods of 0 given by the family of sets $V^* = \{x^* \in X^* : \langle x^*, v\rangle \leq \varepsilon\}$ where $\varepsilon > 0$ and $v \in X$. It is also the weakest topology in X^* that the preserves pointwise convergence.

We provide another similar family that conforms a local base that somehow measures the size of the neighborhoods with a positive parameter and a finite dimensional space, which is new to the best of our knowledge[1].

Consider the family of sets in X^* of the form

$$V^*_{L,\varepsilon} := L^\perp + \varepsilon\mathbb{B}^*,$$

with $\varepsilon > 0$ and L a finite dimensional subspace of X, where $L^\perp := \{x^* \in X^* \mid x^*(x) = 0, \forall x \in L\}$ is the orthogonal to L. This family enjoys the property that for any $L, M \subset X$ finite dimensional subspaces of X and $\varepsilon, \delta > 0$ it holds

$$V^*_{L,\varepsilon} + V^*_{M,\delta} = V^*_{L+M,\varepsilon+\delta},$$

so that

$$2V^*_{L,\varepsilon} = V^*_{L,2\varepsilon},$$

and if X is finite dimensional obviously $V^*_{X,\varepsilon} = \mathbb{B}^*_\varepsilon$. Thus, this family can be seen as a 'canonical' base of neighborhoods of the weak star topology in X^*.

Lemma 11. *The family of sets $V^*_{L,\varepsilon} := L^\perp + \varepsilon\mathbb{B}^*$ is a base of neighborhoods of the origin in X^* for the topology w^*.*

[1] The same construction can be done more generally in Banach spaces.

Proof. We need to prove the following assertions

1. For any $V^* \in \mathcal{N}_{w^*}(0)$, there exist a finite dimensional subspace L of X and $\varepsilon > 0$ such that $V^*_{L,\varepsilon} \subset V^*$.

2. For any finite dimensional subspace L of X and $\varepsilon > 0$, we have $V^*_{L,\varepsilon} \in \mathcal{N}_{w^*}(0)$.

Let us prove first 1. Take an arbitrary $V^* \in \mathcal{N}_{w^*}(0)$. We know that there exists a finite set of vectors $v_1, ..., v_k$ in X and $\delta > 0$ such that

$$\bigcap_{j=1}^{k} \{x^* \in X^* : \langle x^*, v_j\rangle < \delta\} \subset V^*. \tag{II.2.1}$$

Define $R := \max_{j=1,...,k} \{\|v_j\|\}$ and take $L := \operatorname{span}(v_1, ..., v_k)$ and $\varepsilon > 0$ such that $\varepsilon R < \delta$. It is easy to verify that for each $j \in \{1, ..., k\}$

$$L^{\perp} + \varepsilon\mathbb{B}^* \subset \{x^* \in X^* : \langle x^*, v_j\rangle < \delta\}\,. \tag{II.2.2}$$

Putting (II.2.1) and (II.2.2) together we obtain that $V^*_{L,\varepsilon} \subset V^*$.

Now let us prove 2. Take a base $\{v_1, ..., v_k\}$ of the finite dimensional subspace L. We are going to find $\delta > 0$ such that

$$\bigcap_{j=1}^{k} \{x^* \in X^* : \langle x^*, v_j\rangle < \delta\} \subset V^*_{L,\varepsilon}.$$

Every $x \in L$ has a unique representation of the form $x = \sum_{j=1}^{k} \lambda_j v_j$, and so the projections $\lambda_j : L \to \mathbb{R}$ are well defined for $j = 1, ..., k$. Moreover, each λ_j is linear in a finite dimensional space so there exist $c_j > 0$ such that

$$\lambda_j(x) \leq c_j \|x\|, \quad \forall x \in L.$$

Take $c := \sum c_j > 0$ and $\delta := \varepsilon/c > 0$. We observe that if $\langle y^*, v_j\rangle < \delta$ for all j and $x \in \mathbb{B} \cap L$ then

$$\langle x^*, x\rangle = \sum_{j=1}^{k} \lambda_j(x) \langle x^*, v_j\rangle \leq c\delta = \varepsilon,$$

and so $x^* \in \{y^* : y^*(\mathbb{B}_X \cap L) \leq \varepsilon\} = L^{\perp} + \varepsilon\mathbb{B}^*$, where the last equality is due to the bipolar theorem. □

Definition 5. [84, Theorem 2.35] *Given a non-empty set $\Omega \subset X$ and a point $x \in \Omega$, we define the Fréchet normal cone as*

$$\hat{\mathrm{N}}_\Omega(x) := \left\{ x^* \in X^* \mid \limsup_{y \to_\Omega x} \frac{\langle x^*, y - x\rangle}{\|y - x\|} \leq 0 \right\}$$

while for $x \notin \Omega$, we write $\hat{\mathrm{N}}_\Omega(x) := \emptyset$. The notation $y \to_\Omega x$ means that $y \to x$ and $y \in \Omega$. The limiting/basic normal cone to Ω at $\bar{x} \in X$ is the set

$$\mathrm{N}_\Omega(\bar{x}) := \sigma^* -^{seq} \limsup_{x \to \bar{x}} \hat{\mathrm{N}}(\Omega; x).$$

Definition 6. [84, Definitions 1.77 and 1.83] *Given* $\varphi : X \to \overline{\mathbb{R}}$, *a point* $x \in \operatorname{dom}\varphi$. *We define the* Fréchet subdifferential *of* φ *at* x *as the set*

$$\begin{aligned}\hat{\partial}\varphi(x) &:= \left\{x^* \in X^* \mid (x^*, -1) \in \hat{\mathrm{N}}_{\operatorname{epi}\varphi}(x, \varphi(x))\right\}\\ &= \left\{x^* \in X^* \mid \liminf_{y\to x} \frac{\varphi(x) - \varphi(y) - \langle x^*, y - x\rangle}{\|y - x\|} \geq 0\right\}.\end{aligned}$$

The limiting subdifferential *of* φ *at* $x \in X$ *is the set*

$$\partial\varphi(x) := \{x^* \in X^* \mid (x^*, -1) \in \mathrm{N}_{\operatorname{epi}\varphi}(x, \varphi(x))\}.$$

The singular subdifferential *of* φ *at* $\bar{x}$ *is the set*

$$\partial^\infty\varphi(x) := \{x^* \in X^* \mid (x^*, 0) \in \mathrm{N}_{\operatorname{epi}\varphi}(x, \varphi(x))\}.$$

We have the following characterizations of the limiting and the singular subdifferentials in terms of the Fréchet subdifferential.

Theorem 15. [84, Theorems 2.34 and 2.38] *Assume* X *is an Asplund space, and* $\varphi : X \to \overline{\mathbb{R}}$ *is lower semi-continuous around* $\bar{x} \in \operatorname{dom}\varphi$. *Then we have*

$$\partial\varphi(\bar{x}) = \sigma^* -^{seq} \limsup_{x\to_\varphi \bar{x}} \hat{\partial}\varphi(x),$$

and

$$\partial^\infty\varphi(\bar{x}) = \|\cdot\|_* - \limsup_{\substack{x\to_\varphi\bar{x},\mu>0\\ \mu\to 0}} \mu\hat{\partial}\varphi(x),$$

where $y \to_\varphi x$ *means that* $y \to x$ *and* $\varphi(y) \to \varphi(x)$.

The weak fuzzy sum rule, that we present next, is the one of [128, Theorem 3.3.3] but extended to the case of functions defined on Asplund spaces. We show next how it can be extended using the separable reduction technique, though the result was in fact already known (see [66–68]).

Lemma 12. *Let* X *be an Asplund space,* $\bar{x} \in X$, *and let* $\varphi_i : X \to \overline{\mathbb{R}}$, $i = 1, ..., m$, *be lower semi-continuous functions around* $\bar{x}$. *Then for any* $\varepsilon > 0$, $x^* \in \hat{\partial}(\varphi_1 + ... + \varphi_m)(\bar{x})$, *and any weak* neighborhood* V^* *of the origin in* X^* *there are* $(x_i, x_i^*) \in \hat{\partial}\varphi_i$ *such that*

$$(x_i, \varphi_i(x_i)) \in (\bar{x}, \varphi_i(\bar{x})) + \varepsilon\mathbb{B}_{X\times\mathbb{R}}, \tag{II.2.3}$$

$$\operatorname{diam}(x_1, ..., x_m) \cdot \|x_i^*\| < \varepsilon, \tag{II.2.4}$$

$$x^* \in \sum_{i=1}^m x_i^* + V^*. \tag{II.2.5}$$

Proof. The result is well-known if we assume X to be Fréchet smooth (see [128, Theorem 3.3.3]), and the general case can be deduced from the separable reduction theorem (see [94, Theorem 6]) as follows.

We can assume without loss of generality (see Lemma 11) that $V^* = L^{\perp} + 2\varepsilon\mathbb{B}_{X^*}$ for a finite dimensional subspace L of X and $\varepsilon > 0$. Take W a separable subspace of X that contains $\bar{x}$ and L with some further hypothesis that will be imposed later. Since $x^* \in \hat{\partial}(\varphi_1 + ... + \varphi_m)(\bar{x})$, then also $x^*|_W \in \hat{\partial}(\varphi_1 + ... + \varphi_m|_W)(\bar{x})$ or equivalently

$$0 \in \hat{\partial}(\varphi_1 + ... + \varphi_m - x^*|_W)(\bar{x}).$$

But separable subspaces of an Asplund space can be renormed with a Fréchet smooth norm, then we can apply Theorem 3.3.3 from [128] (see also [26, Theorem 2.10]) in W. Then for $V^*|_W := L^{\perp_{W^*}} + \varepsilon\mathbb{B}_{W^*}$ ($L^{\perp_{W^*}}$ is the annihilator of L in W^*) there exist $(x_i, y^*_{i,W}) \in \hat{\partial}(\varphi_i|_W)$, $i = 1, ..., m$, such that

$$(x_i, \varphi_i(x_i)) \in (\bar{x}, \varphi_i(\bar{x})) + \varepsilon\mathbb{B}_{W\times\mathbb{R}},$$

$$\operatorname{diam}(x_1, ..., x_m) \cdot \|y^*_{i,W}\|_{W^*} < \varepsilon,$$

$$x^*|_W \in \sum_{i=1}^{m} y^*_{i,W} + V^*|_W.$$

Take extensions $y^*_i \in X^*$ of $y^*_{i,W}$, that is, $y^*_i|_W = y^*_{i,W}$, with $\|y^*_i\|_* = \|y^*_{i,W}\|_{W^*}$.

Now consider W given by the separable reduction theorem and chose $\eta > 0$ such that $\eta \max(1, \operatorname{diam}(x_1, ..., x_m)) \leq \varepsilon m^{-1}$. Since $y^*_i|_W \in \hat{\partial}\varphi_i(x_i)$, then $\hat{\partial}(\varphi_i - y^*_i|_W)(x_i) \cap \eta\mathbb{B}_*$ is non-empty. We deduce that also $\hat{\partial}(\varphi_i - y^*_i|_W)(x_i) \cap \eta\mathbb{B}$ is non-empty, so that there exist $z^*_i \in \hat{\partial}(\varphi_i - y^*_i)(x_i) = \hat{\partial}\varphi_i(x_i) - y^*_i$ with $\|z^*_i\|_{X^*} \leq \eta$. Finally, $x^*_i := y^*_i + z^*_i \in \hat{\partial}\varphi_i(x_i)$ satisfy the desired properties:

$$\operatorname{diam}(x_1, ..., x_m)\|x^*_i\|_* \leq \operatorname{diam}(x_1, ..., x_m)(\|y^*_i\|_* + \|z^*_i\|_*) \leq 2\varepsilon$$

and

$$x^* - \sum_{i=1}^{m} x^*_i = \left(x^* - \sum_{i=1}^{m} y^*_i\right) - \sum_{i=1}^{m} z^*_i \in \frac{1}{2}V^* + \varepsilon\mathbb{B}_{X^*} \subset V^*.$$

□

Remark 8. *In the literature the result of Lemma 12 is usually stated in the general context of Asplund spaces but with a weaker conclusion which does not include estimation* (II.2.4) *(see for instance [84, Lemma 5.27], [22, Theorem 2.2]). It turns out that the additional condition* (II.2.4) *is very important, as we shall see in our analysis in the next sections.*

The following two lemmas are valid in normed vector spaces, but we will only make use of them in the context of Asplund spaces.

Lemma 13. *Let X be a normed vector space, $\varphi : X \to \overline{\mathbb{R}}$ and $x \in \operatorname{dom}\varphi$. If $y \geq z \geq \varphi(x)$ then $\hat{\mathrm{N}}_{\operatorname{epi}\varphi}(x, y) \subset \hat{\mathrm{N}}_{\operatorname{epi}\varphi}(x, z)$. Moreover, if $(x^*, -y^*) \in \hat{\mathrm{N}}_{\operatorname{epi}\varphi}(x, y)$ then $y^* \geq 0$, and if $y > \varphi(x)$ then $y^* = 0$.*

Proof. Let $y > z$ and $(x^*, y^*) \in \hat{\mathrm{N}}_{\operatorname{epi}\varphi}(x, y)$. Then for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\langle x^*, x' - x\rangle + \langle y^*, y' - y\rangle \leq \varepsilon(\|x' - x\| + |y' - y|)$$

whenever $\|x'-x\|<\delta$, $|y'-y|<\delta$ and $y'\geq\varphi(x')$ (i.e. $(x'.y')\in\operatorname{epi}\varphi$). By defining the variable $y'':=y'-y+z$ we see that

$$\langle x^*,x'-x\rangle+\langle y^*,y''-z\rangle\leq\varepsilon(\|x'-x\|+|y''-z|)$$

whenever $\|x'-x\|<\delta$, $|y''-z|<\delta$ and $y''\geq\varphi(x')-y+z$. The last inequality holds in particular when $y''\geq\varphi(x')$, since $y>z$. But this is by definition that $(x^*,y^*)\in\hat{\mathrm{N}}_{\operatorname{epi}\varphi}(x,z)$.

The two final statements are trivial. □

Lemma 14. *Let X be a normed vector space and let $\varphi:X\to\overline{\mathbb{R}}$ be continuous around $\bar{x}$. Then*

$$\hat{\mathrm{N}}_{[\varphi=\varphi(\bar{x})]}(\bar{x})=\hat{\mathrm{N}}_{[\varphi\leq\varphi(\bar{x})]}(\bar{x})\cup\hat{\mathrm{N}}_{[\varphi\geq\varphi(\bar{x})]}(\bar{x}),\tag{II.2.6}$$

and

$$\hat{\mathrm{N}}_{\operatorname{gph}\varphi}(\bar{x},\varphi(\bar{x}))=\hat{\mathrm{N}}_{\operatorname{epi}\varphi}(\bar{x},\varphi(\bar{x}))\cup\hat{\mathrm{N}}_{\operatorname{hyp}\varphi}(\bar{x},\varphi(\bar{x})).\tag{II.2.7}$$

Moreover, if $(x^*,y^*)\in\hat{\mathrm{N}}_{\operatorname{gph}\varphi}(\bar{x},\varphi(\bar{x}))$ with $y^*<0$, then $(x^*,y^*)\in\hat{\mathrm{N}}_{\operatorname{epi}\varphi}(\bar{x},\varphi(\bar{x}))$.

Proof. Let us assume without loss of generality that $\bar{x}=0$ and $\varphi(\bar{x})=0$. Since $[\varphi=0]$ is a subset of both $[\varphi\leq 0]$ and $[\varphi\geq 0]$, then the inclusion to the left in (II.2.6) is trivial. Let us prove now the reverse inclusion.

Assume that x^* is not in $\hat{\mathrm{N}}_{[\varphi\leq 0]}(0)$ nor in $\hat{\mathrm{N}}_{[\varphi\geq 0]}(0)$. Then there exist $\varepsilon>0$, $(x_n)_n\subset[\varphi\leq 0]$ and $(y_n)_n\subset[\varphi\geq 0]$ such that $x_n,y_n\to 0$ and

$$\lim_n\varepsilon\|x_n\|-\langle x^*,x_n\rangle<-\varepsilon,\ \text{ and }\lim_n\varepsilon\|y_n\|-\langle x^*,y_n\rangle<-\varepsilon.$$

By the continuity of φ, for each n we can find $z_n\in[\varphi=0]\cap[x_n,y_n]$, so that $z_n\to 0$ too. Then, the convexity of $x\mapsto\varepsilon\|x\|-\langle x^*,x\rangle$ gives us that

$$\lim_n\varepsilon\|z_n\|-\langle x^*,z_n\rangle\leq-\varepsilon,$$

which implies that $x^*\notin\hat{\mathrm{N}}_{[\varphi=0]}(0)$.

The validity of (II.2.7) follows from considering the lower semi-continuous function $\phi(x,y):=\varphi(x)-y$, for which $[\phi\leq 0]=\operatorname{epi}\varphi$, $[\phi\geq 0]=\operatorname{hyp}\varphi$, and $[\phi=0]=\operatorname{gph}\varphi$.

The last conclusion follows from Lemma 13.

□

Remark 9. *The idea of* (II.2.6) *of Lemma 14 can be found in the proof of [88, Theorem 4.4]. The second part of Lemma 14 allows us to avoid repeating some arguments as done in [128, Theorem 3.3.5] and [84, Theorem 2.40].*

II.2.3 Non-Degenerate Weak Fuzzy Suboptimality Conditions

We consider a constrained optimization problem of the form:

$$\begin{array}{ll}\min & \varphi_0(x)\\ \text{s.t.} & \varphi_i(x)\leq 0,\ i=1,...,m,\\ & \varphi_i(x)=0,\ i=m+1,...,m+p,\end{array}\tag{II.2.8}$$

where X is an Asplund space, and $\varphi_i : X \to \overline{\mathbb{R}}$. We say that problem (II.2.8) is *closed* around a reference point $\bar{x} \in \bigcap_{i=0}^{m+p} \operatorname{dom} \varphi_i$, if φ_i is lower semi-continuous (continuous, respectively) for $i = 0, ..., m$, (for $i = m+1, ..., m+p$, respectively) around $\bar{x}$. This is a basic assumption that we will consider always.

A classical approach to obtain (sub)optimality conditions is (see [84,89,128]) to transform the constrained minimization problem (II.2.8) into the unconstrained minimization of the following 'penalized' function

$$\varphi := \varphi_0 + \sum_{i=1}^{m} \chi_{[\varphi_i \leq 0]} + \sum_{i=m+1}^{m+p} \chi_{[\varphi_i = 0]},$$

which is lower semi-continuous around $\bar{x}$, whenever (II.2.8) is closed around $\bar{x}$, and both problems have the same set of solutions whenever φ is proper (or equivalently, (II.2.8) admits a feasible solution). This approach has the disadvantage that using the sum rule we loose control over the values of the inequality constraints at the approximate points as shown in the following example.

Example 4. *Consider a non-smooth optimization problem of the form*

$$\begin{array}{ll} \min & \varphi_0(x) \\ s.t. & \varphi_1(x) \leq 0, \end{array}$$

where $\varphi_0, \varphi_1 : \mathbb{R}^2 \to \mathbb{R}$ *are defined by*

$$\varphi_0(x, y) := 2|x| - y$$

$$\varphi_1(x, y) := \begin{cases} -1, & \text{if } x = y = 0, \\ 0, & \text{if } y \leq |x|,\ (x, y) \neq (0, 0), \\ 1, & \text{otherwise.} \end{cases}$$

The pair $(\bar{x}, \bar{y}) := (0, 0)$ *is a solution of* (4), *and thus* $(0, 0) \in \hat{\partial}(\varphi_0 + \chi_{[\varphi_1 \leq 0]})(\bar{x}, \bar{y})$. *We observe that there is no point* $(x, y) \neq (\bar{x}, \bar{y})$ *that is close in the value of* φ_1 *to* $(\bar{x}, \bar{y})$. *From the other side, at* $(\bar{x}, \bar{y})$ *the Fréchet subdifferential of the function* $\chi_{[\varphi_1 \leq 0]}$ *is empty. Thus, the point* (x_1, y_1) *associated to* $\chi_{[\varphi_1 \leq 0]}$ *given by the fuzzy sum rule (Lemma 12) cannot have* $\varphi_1(x_1, y_1)$ *arbitrarily close to* $\varphi_1(\bar{x}, \bar{y})$.

A variant of this approach is to replace the indicators $\chi_{[\varphi_i \leq 0]}$, in the sum that defines φ, by $\chi_{[\varphi_i - \varphi_i(\bar{x}) \leq 0]}$. By doing so, we would gain control on the values of the inequality constraint function but we would loose information about its activeness (whether $\varphi_i(\bar{x}) = 0$ or $\varphi_i(\bar{x}) < 0$, for $i = 1, ..., m$).

A less frequent approach that was considered in [22] is what we here call the *epigraph approach*. It is based on the following observations

$$\chi_{[\varphi_i \leq 0]}(x) = \min_{y \in \mathbb{R}} \chi_{\mathbb{R}_-}(y) + \chi_{\operatorname{epi} \varphi_i}(x, y),$$

and

$$\chi_{[\varphi_i = 0]}(x) = \min_{y \in \mathbb{R}} \chi_{\{0\}}(y) + \chi_{\operatorname{gph} \varphi_i}(x, y),$$

for any $x \in X$. We observe that $\bar{x}$ is a (sub)optimal solution of (II.2.8) if and only if $(\bar{x}, \varphi_1(\bar{x}), ..., \varphi_{m+p}(\bar{x}))$ is a (sub)optimal minimum of

$$\phi(x,y) := \varphi_0(x) + \chi_{\mathbb{R}^m_- \times \{0\}^p}(y) + \sum_{i=1}^{m} \chi_{\operatorname{epi}\varphi_i}(x, y_i) + \sum_{i=m+1}^{m+p} \chi_{\operatorname{gph}\varphi_i}(x, y_i). \tag{II.2.9}$$

The combination between the convergence of the values end the complementarity condition allows us to distinguish between active and non-active inequality constraints. For non-active inequality constraints the multiplier always goes to 0. With the previous result of Nghia (see [89]) this was not possible unless the corresponding function were continuous.

The usual complementarity condition ensures that the multiplier associated to non-active inequality constraints is 0. In our case, we have this in the limit.

Before going further, let us precisely recall what is suboptimality (ε-optimality) for a minimization problem, in the sense of [84, Subsection 5.1.4].

Definition 7. *Given $\varepsilon > 0$ we say that $\bar{x}$ is a* ε-optimal solution *of* (II.2.8) *if it is feasible: ($\varphi_0(\bar{x}) < +\infty$, $\varphi_i(\bar{x}) \leq 0$, $\forall i = 1, ..., m$, $\varphi_i(\bar{x}) = 0, \forall i = m+1, ..., m+p$), and $\varphi_0(\bar{x}) < \varepsilon + \inf\{\varphi_0(x) : x \text{ feasible}\}$. In terms of the function ϕ defined in* (II.2.9)*, $\bar{x}$ is ε-optimal if*

$$\phi(\bar{x}) < \varepsilon + \inf \phi < +\infty.$$

Theorem 16. *Assume X is an Asplund space, $L \subset X$ a finite dimensional subspace, $\varepsilon > 0$ and $\lambda > 0$. If $\bar{x}$ is a ε-optimal solution of the minimization problem* (II.2.8)*, then for each $i = 0, ..., m+p$ there exist $x_i \in X$ with $\|x_i - \bar{x}\| < \lambda$, $\varphi_0(x_0) < \varphi_0(\bar{x}) + \varepsilon$, $\varphi_i(x_i) - \varphi_i(\bar{x}) < \lambda$, for $i = 1, ..., m$, $|\varphi_i(x_i) - \varphi_i(\bar{x})| < \lambda$, for $i = m+1, ..., m+p$, and $(\mu_i)_{i=1}^{m+p} \in \mathbb{R}_+^{m+p}$ such that*

$$\mu_i \varphi_i(x_i) < \frac{\varepsilon}{\lambda}|\varphi_i(\bar{x})| + \varepsilon, \quad \forall i = 1, ..., m, \tag{II.2.10}$$

$$\mu_i |\varphi_i(x_i)| < \varepsilon, \quad \forall i = m+1, ..., m+p, \tag{II.2.11}$$

$$0 \in x_0^* + \sum_{i=1}^{m} \mu_i \hat{\partial}\varphi_i(x_i) + \sum_{i=m+1}^{m+p} \mu_i \left(\hat{\partial}\varphi_i(x_i) \cup \hat{\partial}(-\varphi_i)(x_i)\right) + V^*_{L,\frac{\varepsilon}{\lambda}}. \tag{II.2.12}$$

Proof. Take $\bar{x}$ a ε-optimal solution of the constrained minimization problem (II.2.8). Since the inequality defining ε-optimality is strict, we can find $\varepsilon_1 \in]0, \varepsilon[$ such that $\bar{x}$ is still a ε_1-optimal solution. Then $(\bar{x}, \bar{y})$ with $\bar{y} := (\varphi_1(\bar{x}), ..., \varphi_{m+p}(\bar{x})) \in \mathbb{R}^{m+p}$ is a ε_1-minimum of ϕ defined in (II.2.9). For the sake of simplicity of the notation, let us consider the case of only one ($m = 1$) inequality constraint function and no ($p = 0$) equality constraints. For the consideration of equality constraints the proof is almost the same considering (II.2.7), see also [22].

Step 1: By Ekeland's Variational Principle (see e.g. [128, Theorem 2.1.2]), we know there exists $(x, y) \in X \times \mathbb{R}$ such that

$$\|(x,y) - (\bar{x}, \bar{y})\| < \lambda, \tag{II.2.13}$$

and

$$\phi(x,y) \leq \phi(u,v) + \frac{\varepsilon_1}{\lambda}\|(x,y) - (u,v)\|, \quad \forall (u,v) \in X \times \mathbb{R}. \tag{II.2.14}$$

It follows from (II.2.13) and (II.2.14) that $\varphi_0(x) < \varphi_0(\bar{x}) + \varepsilon_1$.

Step 2: We employ the weak fuzzy sum rule for $(0,0) \in \hat{\partial}(\phi + \frac{\varepsilon_1}{\lambda}\| \cdot -(x,y)\|)(x,y)$ (by (II.2.14)) with threshold $\eta > 0$ (small enough), in order to find points $(x_i, y_i) \in (x,y) + \eta\mathbb{B}_{X\times\mathbb{R}}$, $i = 0,1,2$, along with $(x_0^*, y_0^*) \in \hat{\partial}(\varphi_0 + \chi_{\mathbb{R}_-})(x_0, y_0) = \hat{\partial}\varphi_0(x_0) \times \hat{\mathrm{N}}_{\mathbb{R}_-}(y_0)$, $(x_1^*, y_1^*) \in \hat{\mathrm{N}}_{\operatorname{epi}\varphi_1}(x_1, y_1)$, $(x_2^*, y_2^*) \in \frac{\varepsilon_1}{\lambda}\mathbb{B}_{X\times\mathbb{R}}$ such that $|\varphi_0(x) - \varphi_0(x_0)| < \eta$ and

$$\|(x_i^*, y_i^*)\| \cdot \operatorname{diam}\{(x_j, y_j) : j = 0,1,2\} < \eta, \tag{II.2.15}$$

$$\sum_{i=0}^{2}(x_i^*, y_i^*) \in V^*_{L,\eta} \times]-\eta, \eta[. \tag{II.2.16}$$

Since $(x_1, y_1) \in \operatorname{epi}\varphi_1$ and by (II.2.13) we have that

$$\varphi_1(x_1) \le y_1 \le y + \eta < \bar{y} + \lambda = \varphi_1(\bar{x}) + \lambda. \tag{II.2.17}$$

Step 3a: Let us first consider the case when $y_1^* < 0$. By Lemma 13 we have that $y_1 = \varphi_1(x_1)$. If $y_0 = 0$, then (II.2.15) implies

$$-y_1^*|\varphi_1(x_1)| \le \|(x_1^*, y_1^*)\| \cdot \|(x_0, y_0) - (x_1, y_1)\| < \eta.$$

If $y_0 < 0$, then $y_0^* = 0$ and (II.2.16) implies that $-y_1^* < \eta + y_2^* \le \eta + \frac{\varepsilon_1}{\lambda}$ so that with (II.2.17) we have

$$-y_1^*\varphi_1(x_1) \le (\eta + \frac{\varepsilon_1}{\lambda})(|\varphi_1(\bar{x})| + \lambda) = \frac{\varepsilon}{\lambda}|\varphi_1(\bar{x})| + \varepsilon.$$

In this case we have that $z_1^* := -x_1^*/y_1^* \in \hat{\partial}\varphi_1(x_1)$ and $\mu_1 := -y_1^* > 0$ satisfy the desired properties.

Step 3b: Now consider the case when $y_1^* = 0$. By Lemma 13 we have

$$(x_1^*, 0) \in \hat{\mathrm{N}}_{\operatorname{epi}\varphi_1}(x_1, y_1) \subset \hat{\mathrm{N}}_{\operatorname{epi}\varphi_1}(x_1, \varphi_1(x_1)),$$

and then $x_1^* \in \partial^\infty\varphi_1(x_1)$. Now we use Theorem 15 to find $z_1 \in \eta\mathbb{B}(x_1)$, $z_1^* \in \hat{\partial}\varphi_1(z_1)$ and $\mu_1 \in]0, \eta[$ such that $\|\mu_1 z_1^* - x_1^*\| \le \eta$, and $|\varphi_1(z_1) - \varphi_1(x_1)| \le \eta$. Then $x_0^* + \mu_1 z_1^* \in V^*_{L,\varepsilon}$. We observe also that

$$\mu_1|\varphi_1(z_1)| \le \eta(|\varphi_1(x_1)| + \eta) \le \eta(|\varphi_1(x)| + 2\eta) < \beta,$$

since $\eta > 0$ can be taken arbitrarily small. □

We can state a corollary of Theorem 16 in the way suboptimality conditions are usually stated (see for instance [84, Theorem 5.29] for the degenerate case without convergence of values for inequality constraints, nor complementarity condition) as follows.

Corollary 6. *Let X be an Asplund space and $V^* \in \mathcal{N}_{w^*}(0)$. There exist $\bar{\varepsilon} > 0$ such that for every $0 < \varepsilon < \bar{\varepsilon}$ and $\bar{x} \in X$ a ε^2-optimal solution of* (II.2.8)*, there are $(x_i, x_i^*) \in \hat{\partial}\varphi_i$, $i = 0, ..., m+p$, with $(x_i, \varphi_i(x_i)) \in \varepsilon\mathbb{B}(\bar{x}, \varphi_i(\bar{x}))$ and $(\mu_i)_{i=1}^{m+p} \in \mathbb{R}_+^{m+p}$ such that*

$$\mu_i|\varphi_i(x_i)| < \varepsilon, \quad \forall i = 1, ..., m,$$
$$0 \in x_0^* + \sum_{i=1}^{m}\mu_i x_i^* + \sum_{i=m+1}^{m+p}\mu_i\left(\hat{\partial}\varphi_i(x_i) \cup \hat{\partial}(-\varphi_i)(x_i)\right) + V^*.$$

Proof. Assume without loss of generality that $V^* = L^\perp + \bar{\varepsilon}\mathbb{B}^*$ for a $\bar{\varepsilon} \in]0,1[$ and L finite dimensional subspace of X. For any $\varepsilon \in]0, \bar{\varepsilon}[$ and $\bar{x}$ a ε^2-optimal solution we just apply Theorem 16 with $\lambda = \varepsilon$. □

Remark 10. *In Section 5.1.4 of [84], suboptimality conditions (subOC) for constrained problems (like ours) are studied. Its first result, Theorem 5.29, gives weak fuzzy subOC in a degenerate form, that is, a multiplier corresponding to the objective function may be equal or close to 0 (like Fritz-John OC). The subsequent results of that section are either about the non-degenerate case with CQs or the degenerate case without CQs, and in both cases assuming the uniform Lipschitz continuity of the constraints functions and using the basic limiting subdifferentials.*

Next we give an estimation of the normal cone to sublevel sets of lower semi-continuous functions.

Corollary 7. *Let X be an Asplund space, $\bar{x} \in X$ and let $\varphi : X \to \overline{\mathbb{R}}$ be lower semi-continuous around $\bar{x}$ with $\varphi(\bar{x}) \leq 0$. If $x^* \in \hat{\mathrm{N}}_{[\varphi \leq 0]}(\bar{x})$ then for any $V^* \in \mathcal{N}_{w^*}(0)$ and $\varepsilon > 0$ there exist $x \in X$ with $(x, \varphi(x)) \in (\bar{x}, \varphi(\bar{x})) + \varepsilon\mathbb{B}_{X\times\mathbb{R}}$, and $\mu > 0$ with $\mu|\varphi(x)| < \varepsilon$ such that*

$$x^* \in \mu\hat{\partial}\varphi(x) + V^*. \tag{II.2.18}$$

Proof. If $x^* \in \hat{\mathrm{N}}_{[\varphi\leq 0]}(\bar{x})$ then for any $\varepsilon > 0$ there exists $0 < \delta$ such that

$$\langle x^*, y - \bar{x}\rangle \leq \varepsilon\|y - \bar{x}\| + \chi_{\delta\mathbb{B}(\bar{x})}(y),\ \forall y \in [\varphi \leq 0]. \tag{II.2.19}$$

Then $\bar{x}$ is an optimal solution of minimizing $\varphi_0(\cdot) := -x^*(\cdot) + \varepsilon\|\cdot - \bar{x}\| + \chi_{\delta\mathbb{B}(\bar{x})}(\cdot)$ subject to $\varphi_1(x) := \varphi(x) \leq 0$. Applying Theorem 16, then there exists $x_0, x_1 \in X$ with $|\varphi(x_1) - \varphi(\bar{x})| < \delta$, $\|x_i - \bar{x}\| < \delta$, $x_0^* \in \hat{\partial}\varphi_0(x_0) \subset -x^* + \varepsilon\mathbb{B}_{X^*}$, $x_1^* \in \hat{\partial}\varphi(x_1)$ and $\mu > 0$ such that $x_0^* + \mu x_1^* \in V^*$ and $\mu|\varphi(x_1)| < \delta$. Finally, we see that $x^* \in \mu x_1^* + V^* + \varepsilon\mathbb{B}_{X^*}$, which finishes the proof. □

Remark 11. *It would also be possible to prove Corollary 7 directly from the Fuzzy sum rule, simply by noting that $x^* \in \hat{\mathrm{N}}_{[\varphi\leq 0]}(\bar{x})$ is equivalent to $(x^*, 0) \in \hat{\partial}(\chi_{\operatorname{epi}\varphi} + \chi_{X\times\mathbb{R}_-})(\bar{x}, \varphi(\bar{x}))$. Though, this would be almost repeating the proof of Theorem 16, and is thus left to the reader.*

II.2.4 Subdifferential of Nonconvex Supremum Functions

We consider in this section the supremum function of an arbitrary family of non-convex lower semi-continuous functions. More precisely, given a family of lower semi-continuous functions $f_t : X \to \overline{\mathbb{R}}$, $t \in T$, with T a non-empty arbitrary (possibly infinite) index set, we consider the supremum function $f : X \to \overline{\mathbb{R}}$ defined by $f(x) := \sup_{t\in T} f_t(x)$ for each $x \in X$.

The following theorem shows an estimation of the subdifferential of the supremum function in terms of the subgradients of the functions of the family that defines the supremum function. It is a refinement of [96, Theorem 3.8, Proposition 3.2] (see also [27, Theorem 3.18]) and it is related to [83, Theorem 3.1]. We will discuss the above mentioned relations after stating the theorem and giving its proof.

Theorem 17. *Assume X is an Asplund space, the functions f_t, $t \in T$ are lower semi continuous, and that $\bar{x} \in X$. Let $x^* \in \hat{\partial} f(\bar{x})$, let $V^* \in \mathcal{N}_{w^*}(0)$, and let $\varepsilon > 0$. Then there exist a finite subset $S \subset T$, $x_s \in X$ with $(x_s, f_s(x_s)) \in \varepsilon\mathbb{B}(\bar{x}, f_s(\bar{x}))$, and $\lambda_s > 0$ for $s \in S$, with $\sum_{s\in S} \lambda_s = 1$, such that*

$$x^* \in \sum_{s\in S} \lambda_s \hat{\partial} f_s(x_s) + V^*, \tag{II.2.20}$$

and such that for each $s \in S$

$$\lambda_s |f_s(\bar{x}) - f(\bar{x})| \leq \varepsilon. \tag{II.2.21}$$

Proof. We can assume without loss of generality that $V^* = L^\perp + \varepsilon\mathbb{B}^*$ with $L \subset X$ a finite dimensional subspace. Since $x^* \in \hat{\partial} f(\bar{x})$, there exists $\delta > 0$ such that

$$\langle x^*, x - \bar{x}\rangle \leq f(x) - f(\bar{x}) + \varepsilon\|x - \bar{x}\|, \quad \forall x \in \delta\mathbb{B}(\bar{x}). \tag{II.2.22}$$

Let us define the function

$$\varphi_0(x, y) := y - \langle x^*, x\rangle + \varepsilon\|x - \bar{x}\| + \chi((x, y); \Omega),$$

where $\Omega := L \cap \delta\mathbb{B}(\bar{x}) \times [f(\bar{x}) - 1, f(\bar{x}) + 1]$, and the family of functions $\varphi_t : X \times \mathbb{R} \to \overline{\mathbb{R}}$ given by $\varphi_t(x, y) := f_t(x) - y$. Consider the sets $K := \{(x, y) \in X \times \mathbb{R} \mid \varphi_0(x, y) + \delta^2 \leq \varphi_0(\bar{x}, f(\bar{x}))\}$ and

$$A_t := \{(x, y) \in X \times \mathbb{R} \mid \varphi_t(x, y) > 0\}, \quad t \in T.$$

By (II.2.22) we observe that $(A_t)_{t\in T}$ is an open cover of K. Since K is compact, then $(A_t)_{t\in T}$ admits a finite subcover $(A_s)_{s\in S}$ of K, with $S \subset T$ finite. It follows that $(\bar{x}, f(\bar{x}))$ is a δ^2-optimal solution of the problem

$$\begin{aligned} &\min\ \varphi_0(x, y) \\ &\text{s.t. } (x, y) \in X \times \mathbb{R}, \\ &\qquad \varphi_s(x, y) \leq 0, \forall s \in S. \end{aligned}$$

We can thus apply the weak fuzzy sub-optimality conditions of Corollary 6 to deduce that for each $s \in S \dot{\cup} \{0\}$ there exist $((x_s, y_s), (x_s^*, y_s^*)) \in \hat{\partial}\varphi_s$ with

$$((x_s, y_s), \varphi_s(x_s, y_s)) \in ((\bar{x}, f(\bar{x})), \varphi_s(\bar{x}, f(\bar{x}))) + \delta\mathbb{B}_{(X\times\mathbb{R})\times\mathbb{R}}, \tag{II.2.23}$$

and $(\mu_s)_{s\in S} \in \mathbb{R}^S_+$ such that

$$\mu_s|\varphi_s(x_s, y_s)| \leq \delta, \quad \forall s \in S, \tag{II.2.24}$$

and

$$0 \in (x_0^*, y_0^*) + \sum_{s\in S} \mu_s(x_s^*, y_s^*) + V^* \times (-\delta, \delta). \tag{II.2.25}$$

We see that

$$x_0^* \in \hat{\partial}_x \varphi_0(x_0, y_0) \subset -x^* + \varepsilon\mathbb{B}^* + L^\perp = -x^* + V^*,$$

$y_0^* = 1$, and $y_s^* = -1$ for $s \in S$. Thus, $\sum_{s\in S} \mu_s \in [1 - \delta, 1 + \delta]$, $x_s^* \in \hat{\partial} f_s(x_s)$ and

$$x^* \in \sum_{s\in S} \mu_s x_s^* + V^*.$$

We also see that $|\varphi_s(x_s, y_s) - \varphi_s(\bar{x}, f(\bar{x}))| \leq \delta$ implies that

$$|f_s(x_s) - f_s(\bar{x})| \leq |y_s - f(\bar{x})| + \delta \leq 2\delta.$$

Finally, define $\lambda_s := \mu_s / \sum_{s' \in S} \mu_{s'}$ so that $\sum_{s \in S} \lambda_s = 1$ and $\lambda_s \in [0,1]$. We conclude using (II.2.24) that

$$\begin{aligned} \lambda_s|f_s(\bar{x}) - f(\bar{x})| &\leq \lambda_s|f_s(x_s) - y_s| + 3\delta, \\ &\leq \frac{\mu_s}{1-\delta}|f_s(x_s) - y_s| + 3\delta, \\ &\leq \frac{\delta}{1-\delta} + 3\delta, \end{aligned}$$

which is less that ε, by taking δ small enough. □

The next example shows that our result is stronger than Theorem 3.8 in [96] since this last result is not exactly an extension of the finite case (Proposition 3.2(v) in [96]), while our result is really an extension.

Example 5. *Let $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ be functions given by $f_1 \equiv 0$ and $f_2(x) := -(n+1)^{-1}$ if $x \in]\frac{1}{n+1}, \frac{1}{n}]$ for some $n \in \mathbb{N}$ and $f(x) = -1$ otherwise. We observe that $f_1 \geq f_2$ so that $f := \sup_{t=1,2} f_t = f_1$. It can be easily seen that our formula in Theorem 17 estimates $\hat{\partial} f(0) = \{0\}$ by $\mathbb{R}_+$, while [96, Theorem 3.8] does estimate it by $\mathbb{R}$. This fact reveals that our formula is stronger that the one proposed in [96, Theorem 3.8]. Moreover, it shows that their formula for the subdifferential of an arbitrary family of functions is not a generalization of the case of finitely many functions [96, Proposition 3.2(v)].*

In [83, Theorem 3.1 (b)] the authors proved a similar formula under the rather strong assumption that the family $(f_t)_{t \in T}$ is uniformly Lipschitzian. In this case, they could restrict the functions used in the representation of $\hat{\partial} f(\bar{x})$ to the set $T_\varepsilon(\bar{x})$ of *ε-active indices*, that is, $t \in T$ such that $f_t(\bar{x}) + \varepsilon \geq f(\bar{x})$.

We will show that we can also get this stronger result with a condition which is weaker than the uniform Lipschitzianity of the family $(f_t)_{t \in T}$. But before this, let us look at the following example which shows that in general we cannot restrict to indexes in the set $T_\varepsilon(\bar{x})$ for ε arbitrarily small.

Example 6. *Let $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ be functions defined by $f_1 \equiv 1$ and $f_2(x) := 2$ for $x \neq 0$ and $f_2(0) := 0$. The supremum function $f := \max(f_1, f_2)$ is attained at each point by only one of the functions. More precisely, if $\varepsilon \in [0, 1[$ then*

$$T_\varepsilon(x) = \begin{cases} \{1\}, & x = 0, \\ \{2\}, & x \neq 0. \end{cases}$$

Consider the point $\bar{x} = 0$. We observe that $\hat{\partial} f(\bar{x}) = \mathbb{R}$. At this point, the only function that is ε-active, for $\varepsilon >$ small (less than 1), is f_1, whose subdifferential is everywhere equal to $\{0\}$. If instead we consider points x such that $|f_t(x) - f(\bar{x})| < \varepsilon$, then we are reduced again to $t = 1$.

Thus, for the representation of the subdifferential of the supremum function in Theorem 17, we cannot only consider points whose values, in some of the functions of the family, are close to the supremum function.

We say that (see Definition 2.17 in [45]) the family $(f_t)_{t\in T}$ is *equi-upper semi-continuous* (equi-usc) at $\bar{x}$ if for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$f_t(\bar{x}) + \varepsilon \geq f_t(x), \quad \forall x \in \delta\mathbb{B}(\bar{x}), \forall t \in T. \tag{II.2.26}$$

Of course, the uniform Lipschitzianity implies (and is in general quite weaker than) our equi-usc. We observe that if $(f_t)_{t\in T}$ is equi-usc at $\bar{x}$, then $f = \sup_{t\in T_\varepsilon(\bar{x})} f_t$ in a ball $\delta\mathbb{B}(\bar{x})$ for some $\delta > 0$. Since the subdifferential is a local object, then applying Theorem 17 to this family we recover [83, Theorem 3.1(ii)].

Remark 12. *Note that the equi-usc of $(f_t)_{t\in T}$ implies that the supremum f is usc. But the usc of f is not enough to deduce $f = \sup_{t\in T_\varepsilon(\bar{x})} f_t$ in any ball around $\bar{x}$. For instance, take real functions $f_0(x) := -|x|$ and $f_1(x) := 0$ for $x \neq 0$ and $f_1(0) = -1$. The max between f_0 and f_1 is the constant $f \equiv 0$ which is obviously continuous, but $T_\varepsilon(0) = 1$ for $\varepsilon \in]0, 1[$ so that the supremum over this subset is not equal to f in any neighborhood of 0. Moreover, we can impose that each function of the family $(f_t)_{t\in T}$ is continuous at $\bar{x}$ and the supremum function f is continuous, but still the representation restricted to the subfamily $T_\varepsilon(\bar{x})$ is not locally the same as f. In fact, take f_0 as before and $f_t(x) := -1$ for $x \in [-t, t]$ and $f_t(x) := 0$ else. Again the supremum is $f \equiv 0$ but the ε-active representation does not hold locally around $\bar{x}$.*

Corollary 8. *Assume X is an Asplund space and let $\bar{x} \in X$. Assume moreover that the functions f_t, $t \in T$, are lower semi continuous, and that the family $(f_t)_{t\in T}$ is equi-usc at $\bar{x}$. Let $x^* \in \hat{\partial} f(\bar{x})$, $V^* \in \mathcal{N}_{w^*}(0)$, and $\varepsilon > 0$. Then there exist a finite subset $S \subset T_\varepsilon(\bar{x})$, $x_s \in \bar{x} + \varepsilon\mathbb{B}$, and $\lambda_s > 0$ for $s \in S$, with $\sum_{s\in S} \lambda_s = 1$, such that*

$$x^* \in \sum_{s\in S} \lambda_s \hat{\partial} f_s(x_s) + V^*. \tag{II.2.27}$$

Corollary 9. *Assume X is an Asplund space, T is non-empty and finite, and the functions f_t, $t \in T$, are continuous. Let $\bar{x} \in X$ and $x^* \in \hat{\partial} f(\bar{x})$. Then for any $V^* \in \mathcal{N}_{w^*}(0)$ and $\varepsilon > 0$, there exist $x_s \in \bar{x} + \varepsilon\mathbb{B}$, and $\lambda_s > 0$ for $s \in T_0(\bar{x})$, with $\sum_{s\in T_0(\bar{x})} \lambda_s = 1$, such that*

$$x^* \in \sum_{s\in T_0(\bar{x})} \lambda_s \hat{\partial} f_s(x_s) + V^*. \tag{II.2.28}$$

In the following examples we discuss some difficulties concerning the use of the singular and limiting subdifferentials under no qualification condition at the reference point. The first example shows that non-active functions cannot be discarded not even by replacing their use by the singular subdifferential of the max function at the reference point. The second example shows that it is not enough to consider only the limiting and singular subdifferentials of the functions at the reference point.

Example 7. *Let $f_1 := \chi_{\mathbb{R}_-}$ and $f_2(x) := -1 - \sqrt{x}$ for $x \geq 0$ and $f_2(x) := -x$ for $x < 0$. Consider the max function $f := \max(f_1, f_2)$. The functions f_1, f_2, and thus also f, are lower semi-continuous. We analyze the Fréchet subdifferential of f at $\bar{x} = 0$, which is $\hat{\partial} f(0) = [-1, +\infty)$. Since $f_2(0) = -1 < 0 = f(0)$ then applying Theorem 17 the multiplier λ_2 goes to 0, but for $x_2^* \in \hat{\partial} f_2(x_2)$ the limit of $x_2^*/|x_2^*| (= -1)$ does not belong to $\partial^\infty f(0) = \mathbb{R}_+$. In the other hand, if we force $\lambda_2 = 0$ then the subdifferential of f_1 would yield mostly $\mathbb{R}_+$.*

Example 8. *Let $f_1, f_2 : \mathbb{R}^2 \to \overline{\mathbb{R}}$ be functions given by*

$$f_1(x,y) := \begin{cases} -\sqrt{x} + y, & \text{if } x \geq 0, \\ \infty, & \text{otherwise,} \end{cases}$$

and $f_2(x,y) := f_1(-x,y)$ for all $(x,y) \in \mathbb{R}^2$. Then $f := \max(f_1, f_2)$ is then equal to $y + \chi_{[x=0]}$. Then we have

$$\hat{\partial} f(0,y) = \mathbb{R} \times \{1\}\,. \tag{II.2.29}$$

In the other hand, the limiting subdifferentials of f_1 and f_2 are empty at $(0,y)$ and their singular subdifferentials are $\partial^\infty f_1(0,y) = \mathbb{R}_+ \times \{0\}$ and $\partial^\infty f_2(0,y) = \mathbb{R}_- \times \{0\}$.

II.2.5 Application to Bilevel Programming

In this section, we obtain fuzzy optimality conditions for a bilevel programming problem without any qualification condition, by using the tools developed in the previous sections. For simplicity, we will restrict ourselves to the case of finite dimensions and smooth data.

Consider an optimistic bilevel programming problem (see Chapter III.2 for more about bilevel problems) of the following form,

$$\begin{array}{ll} \min_x & \varphi_o(x), \\ \text{s.t.} & G(x) \leq 0, \end{array} \tag{II.2.30}$$

where $\varphi_o(x) := \inf_y \{F(x,y) : y \in S(x)\}$, and $S(x)$ stands for the solution of the following lower level problem

$$\begin{array}{ll} \min_y & f(x,y), \\ \text{s.t.} & g(x,y) \leq 0. \end{array} \tag{II.2.31}$$

The functions $F, f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{R}^p$, $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$ are assumed to be smooth along this section.

First we give a lemma that will be used in deriving the fuzzy optimality conditions in Theorem 18. Our Lemma 15 can be compared with Theorem 5.7 in [43], which has a small mistake: Caratheodory's Theorem works for each point, not for a set.

Lemma 15. *Let $x^* \in \hat{\partial}(-\varphi)(x)$. Then, for any $\varepsilon > 0$ there exists a finite set $S \subset Y$ and for each $y \in S$ there exists $x_{f,y}, x_{g,y} \in \bar{x} + \varepsilon\mathbb{B}$, $\lambda_y \geq 0$, and $\mu_y \geq 0$ such that*

$$\sum_{y \in S} \lambda_y = 1, \quad \lambda_y |f(x_{f,y}, y) - \varphi(x)| \leq \varepsilon \tag{II.2.32}$$

$$\mu_y |g(x_{g,y}, y)| \leq \varepsilon, \tag{II.2.33}$$

$$x^* \in -\sum_{y \in S} \lambda_y (\nabla_x f(x_{f,y}, y) + \mu_{g,y} \nabla_x g(x_{g,y}, y)) + \varepsilon\mathbb{B}. \tag{II.2.34}$$

Proof. Note that since f is smooth, then there exists $K > 0$ such that

$$\varphi(x) = \inf_y \left\{ f(x,y) + K d_{[g(x,\cdot) \leq 0]}(y) \right\},$$

for instance $K = \sup_{x,y} \|\nabla_x f(x,y)\|$. Therefore, $-\varphi$ is a supremum of the (lower semi)continuous functions $x \mapsto -f(x,y) - K d_{[g(\cdot,y) \leq 0]}(x)$. So, we can apply Theorem 17 to obtain that there exist a finite set $S \subset Y$ and points x_y close to x and $\lambda_y \geq 0$, $\sum_{y\in S} \lambda_y = 1$ and $x_y^* \in \hat{\partial}(-f(\cdot, y) - Kd_{[g(\cdot,y)\leq 0]})(x_y)$ with $\lambda_y|f(x_y, y) - \varphi(x)| \leq \varepsilon$ such that

$$x^* \in \sum_{y\in S} \lambda_y x_y^* + \varepsilon\mathbb{B}. \tag{II.2.35}$$

We apply now the fuzzy sum rule, the fact that the subdifferential of the distance to a set is included in the normal cone to the set [70, Proposition 1.30], and the fuzzy representation of the normal cone (see Corollary 7) for each $y \in S$ to deduce that there exists $x_{f,y}, x_{g,y} \in \bar{x}+\varepsilon\mathbb{B}$ and $\mu_{g,y} \geq 0$ with $\mu_{g,y}|g(x_{g,y}, y)| \leq \varepsilon$ such that

$$x_y^* \in -\nabla_x f(x_{f,y}, y) - \mu_{g,y}\nabla_x g(x_{g,y}, y) + \varepsilon\mathbb{B}.$$

□

Theorem 18. *Let $\bar{x}$ be a local solution of* (II.2.30) *and assume that S_o is inner semi-compact at $\bar{x}$. Then there exist $\mu > 0$, $x_F, x_G, x_f, x_g \in \varepsilon\mathbb{B}(\bar{x})$, $y_F \in S_o(x_F)$, $y_f, y_g \in \varepsilon\mathbb{B}(y_F)$, $\mu_G \geq 0$, $\mu_g \geq 0$, a finite set $S \subset Y$ and for each $y \in S$ there exist $x_{f,y}, x_{g,y} \in \varepsilon\mathbb{B}(\bar{x})$ and $\mu_y \geq 0$, $\lambda_y \in [0, 1]$ such that $\sum_{y\in S} \lambda_y = 1$ and*

$$\begin{aligned}\nabla_x F(x_F, y_F) + \nabla_x G(x_G)^T\mu_G + \mu\left(\nabla_x f(x_f, y_f) - \textstyle\sum_{y\in S} \lambda_y \nabla_x f(x_{f,y}, y)\right)\\ +\nabla_x g(x_g, y_g)^T\mu_g - \mu \textstyle\sum_{y\in S} \lambda_y \nabla_x g(x_{g,y}, y)^T\mu_y \in \varepsilon\mathbb{B},\end{aligned} \tag{II.2.36}$$

$$\nabla_y F(x_F, y_F) + \mu\nabla_y f(x_f, y_f) + \nabla_y g(x_g, y_g)^T\mu_g \in \varepsilon\mathbb{B}, \tag{II.2.37}$$

$$\lambda_y|f(x_{f,y}, y) - \varphi(x)| \leq \varepsilon, \tag{II.2.38}$$

$$|g(x_{g,y}, y)^T\mu_y| \leq \varepsilon. \tag{II.2.39}$$

Proof. Let $\bar{x}$ be a local solution of (II.2.30). Then there exist x_F, x_G close to $\bar{x}$ along with $x_F^* \in \hat{\partial}\varphi_o(x_F)$ and $x_G^* \in \hat{\partial}G(x_G)$, and $\mu_G \geq 0$ with $|G(x_G)^T\mu_G| \leq \varepsilon$ such that $\|x_F^* + \mu_G x_G^*\| \leq \varepsilon$. Since S_o is inner semi-compact and F is strictly differentiable then we have

$$\hat{\partial}\varphi_o(x_F) \subset \partial\varphi_o(x_F) \subset \bigcup_{y_F\in S_o(x_F)} \{\nabla_x F(x_F, y_F) + D^*S(x_F, y_F)(\nabla_y F(x_F, y_F))\}$$

so that there exists $y_F \in S_o(x_F)$ such that

$$x_F^* - \nabla_x F(x_F, y_F) \in D^*S(x_F, y_F)(\nabla_y F(x_F, y_F)).$$

This is equivalent to

$$(x_F^* - \nabla_x F(x_F, y_F), -\nabla_y F(x_F, y_F)) \in \mathrm{N}_{\mathrm{gph}\, S}(x_F, y_F) = \mathrm{N}_{[\phi\leq 0]}(x_F, y_F)$$

where $\phi(x, y) := f(x, y) - \varphi(x) + \chi_{[g\leq 0]}$ with $\varphi(x) := \inf_y \{f(x, y) : g(x, y) \leq 0\}$. Since φ is upper semi-continuous then ϕ is lower semi-continuous. We can thus apply Corollary 7 in order to find $\mu > 0$, $(\tilde{x}, \tilde{y})$ close to (x_F, y_F) and $(\tilde{x}^*, \tilde{y}^*) \in \hat{\partial}\phi(\tilde{x}, \tilde{y})$ such that $\mu(\tilde{x}^*, \tilde{y}^*)$ is close to the vector

$$(x_F^* - \nabla_x F(x_F, y_F), -\nabla_y F(x_F, y_F)).$$

Then by the fuzzy sum rule there exist (x_i, y_i), $i := f, g, \varphi$ close to $(\tilde{x}, \tilde{y})$ along with $(x_f^*, y_f^*) \in \hat{\partial} f(x_f, y_f)$, $(x_\varphi^*) \in \hat{\partial}(-\varphi)(x_\varphi)$ and $(x_g^*, y_g^*) \in \hat{\mathrm{N}}_{[g \le 0]}(x_g, y_g)$ such that

$$(\tilde{x}^*, \tilde{y}^*) \in (x_f^*, y_f^*) + (x_\varphi^*, 0) + (x_g^*, y_g^*) + \varepsilon\mathbb{B} \tag{II.2.40}$$

By applying Lemma 15 we easily obtain the points that satisfy the desired result. □

Part III

Multi-Leader-Follower Games

Non-Cooperative Game Theory

A game is an interdependent decision problem for several players who have individual objectives. In fact, a game is fully described by a finite set of players, each of them having: (i) a *variable*, (ii) an *objective function* and (iii) a *feasible set*, where both the objective function and the feasible set of a given player could depend on the variable of the other players, that is, of his rivals. This is what makes a game an interdependent problem.

Roughly speaking, the aim of each player is to decide about his variable within his feasible set in order to get the least value of his objective function, that is, minimize his objective (or maximize, see Remark in Section I.4).

In a non-cooperative game each player has control over and decides about his own variable, but he cannot directly control the variables of his rivals. Since we only will focus on non-cooperative games we will usually simply talk about games, and omit the 'non-cooperative' adjective.

Concerning the dependency of objectives and constraints on rivals' strategies there are two extreme cases: independent constraints, but dependent objectives, i.e. a classic game; or independent objectives, but dependent constraints, that is, a special case of a generalized game.

Of course, if both the constraints and the objective of each player are independent of his rivals' strategies, then we simply face a collection of independent optimization problems for which game theory has not much interesting things to say.

We have now reached a crucial point on the discussion, and a natural question arises. What can we expect as an outcome of a game?

The answer in my opinion is not at all obvious. Even though the aim of each player in a game has been already stated in an apparently precise way as the minimization of an objective, it is not clear at all what would be an outcome of the game. In fact, the variables of a player's rivals can be seen as an uncertain parameter, making his problem somehow ambiguous.

Different notions of equilibria have been proposed in the literature as expected possible outcomes, and they are based on the players' behavior facing this uncertainty. A natural idea is that each player makes a *conjecture* about his rivals' decision. A major distinction of the conjecture of a player A about player B's decision is that either

(i) fixed conjecture of B's decision, independently of A's decision, or

(ii) dependent conjecture of B's decision, as dependent on A's decision.

If player A makes a fixed conjecture about each rival B, then it is natural that he will make a best response to that fixed conjectures. Now, if every player behave like A with respect to rivals, and the conjectures are coherent, then we reach a so-called Nash equilibrium (Cournot or Bertrand equilibrium in economy). This is the topic of chapter III.1.

The concept of Nash equilibrium is interesting because in an equilibrium no player has incentives for deviating from his decision, unilaterally, and thus the conjectures should not change. But away from an equilibrium the idea of fixed conjectures is less intuitive. For instance, if two players A and B make decisions with coherent fixed conjectures, A making best response to B but B making a non-optimal response. Then B has incentives to change his decision, but if B changes his decision then A's conjecture about B's decision would not be coherent any more.

On the other hand, some players could make dependent conjectures about some of the rivals based on optimal reaction. This leads to many different possibilities including the so-called bilevel programming problem and multi-leader-follower games, which are the topic of Chapters III.2 and III.3, respectively.

The bilevel programming problem consists of a game with two players called the leader and the follower. The leader conjectures that the behavior of the follower is to best respond to the leader, so the leader makes dependent conjectures about rivals' decision (dependent on the leader's decision, see (ii)). The leader takes into account the dependent conjectures, if single-valued (see discussion in Chapter III.2), and plug them inside his optimization problem. If these conjectures are coherent with the behavior of the follower this would lead to a so-called Stackelberg equilibrium. In this case we say that the leader is the upper level player while the follower is the lower level player.

If two players trying to be leaders make the dependent conjectures that the rival's behavior is to best respond to the player, then both players would solve a bilevel programming problem and the combination of decision is a so-called Bowley equilibrium. A Bowley equilibrium with coherent conjectures will unlikely exist, but if one such exists it would also correspond to a Nash equilibrium. Moreover, in this case we cannot distinguish between upper level or lower level because it would depend on whose perspective we are choosing.

In case of several players, the possibilities are more complex since the dependency of one rival decision might depend on another rival's decision. This leads, even in the case of coherent conjectures, to complex models with more than two levels and/or possibly more than one player in each of the levels. In Chapter III.3 we will discuss about the case of Multi-Leader-Follower games, that is, the case of two levels with possibly more than one player at each of the levels.

Chapter III.1

Generalized Nash Equilibrium Problems

This chapter is based on the paper [17] but includes also some material from [19].

III.1.1 Notation and Basics

A *Generalized Nash Equilibrium Problem* (GNEP for short) consists in several players solving each one a parametric optimization problem, the strategy of each player being a parameter for the others. To be more precise, let J be a finite set of players. Player $j \in J$, controlling his strategy $y_j \in \mathbb{R}^{m_j}$, aims at minimizing his objective f_j, which depends on the joint vector of strategies $y := (y_1, ..., y_{|J|}) \in \mathbb{R}^m$ with $m := \sum_{j \in J} m_j$. The minimization is done by player j within his feasible set

$$Y_j(y_{-j}) := \{y_j \in \mathbb{R}^{m_j} \mid g_j(y_j, y_{-j}) \leq 0\},$$

where $g_j : \mathbb{R}^m \to \mathbb{R}^{d_j}$ is the joint constraint function of player j. Following classical notations, m_{-j} and $y_{-j} \in \mathbb{R}^{m_{-j}}$ stands respectively for $m_{-j} := \sum_{j' \neq j} m_{j'}$ and the joint vector of strategies of all players except player j, so that, up to a reordering of the vector, we have $y = (y_j, y_{-j})$.

Given an opponent strategy y_{-j}, a best response of player j is a solution $y_j \in \mathbb{R}^{m_j}$ of the parametric optimization problem:

$$\begin{aligned} \min_{y_j} \quad & f_j(y_j, y_{-j}) \\ s.t. \quad & g_j(y_j, y_{-j}) \leq 0 \end{aligned} \qquad (\mathcal{P}_j(y_{-j}))$$

and we denote by $S_j(y_{-j})$ the set of its solutions. A joint strategy y is said to be *feasible* for the GNEP if, for all $j \in J$, $y_j \in Y_j(y_{-j})$ and the set of all joint feasible strategies $(y: \; y_j \in Y_j(y_{-j})$ for all $j \in J)$ is denoted by Y. This set will be called the *feasible region* of the GNEP. The GNEP consists in finding a joint strategy y such that for each $j \in J$, $y_j \in S_j(y_{-j})$. We will call such a y an *equilibrium* of the game, and the set of all equilibria will be denoted by GNEP. In the case where the feasible set is constant (not depending of the opponent strategy), one calls the problem a Nash Equilibrium Problem (NEP for short).

III.1.2 Existence for GNEP

The following theorem, which was originally given in [63], gives conditions under which we can ensure the existence of equilibria for the GNEP in a general setting. In the survey [49], some relaxation of the hypothesis are discussed.

Theorem 19 (Ichiishi-Quinzii 1983)**.** *Let a GNEP be given and suppose that*

1. *For each $j \in J$ there exist a non-empty, convex and compact set $K_j \subset \mathbb{R}^{n_j}$ such that the set-valued map $Y_j : K_{-j} \rightrightarrows K_j$, is both upper and lower semi-continuous with non-empty closed and convex values, where $K_{-j} := \prod_{j' \neq j} K_j$.*

2. *For every player j, the function f_j is continuous and $f_j(\cdot, y_{-j})$ is quasi-convex on $Y_j(y_{-j})$.*

Then a generalized Nash equilibrium exists.

Note that in [12] an alternative proof of existence of equilibria has been given, under the assumption of Rosen's law, by using the normal approach technique.

III.1.3 Uniqueness for GNEP

Uniqueness of solutions for GNEPs is usually not possible to guarantee. We first show an example of a NEP whose (joint) objectives are strictly convex, and whose set of equilibria is a closed segment, each equilibrium with different value for both of the players. Compare with Example 1 in [49], where the same conclusions holds for a GNEP (in our case it is a NEP).

Example 9. *Consider the following NEP between two players. Let $x = (x_1, x_2) \in [0,1]^2$ and $f_j(x_1, x_2) := (x_1 - x_2)^2 + (x_{-j})^2$ be the objective of player j. Then the best response (or reaction) maps are $S_j(x_{-j}) = x_{-j}$, for $j = 1, 2$. Thus the solution of the NEP are the pairs (x_1, x_2) with $x_1 = x_2 \in [0, 1]$. Note also that the values of the objectives at the equilibria are all different*

$$f_1(\alpha, \alpha) = f_2(\alpha, \alpha) = \alpha^2, \quad \forall \alpha \in [0, 1].$$

In this example, the objectives are jointly strictly convex.

The following example exhibits the case of a linear GNEP for which multiple solution exists with different optimal values.

Example 10. *Let consider a two players game with objectives $f_1, f_2 : \mathbb{R}^2 \to \mathbb{R}$ defined by $f_1(x) := x_1$ and $f_2(x) := -x_2$ for $x = (x_1, x_2) \in \mathbb{R}^2$. Let the constraint for player 1 be defined by the set $X_1(x_2) := \{x_1 \in \mathbb{R} \mid x_2 - x_1 \leq 0\}$ and for player 2 by the set $X_2(x_1) := \{x_2 \in \mathbb{R} \mid x_2 - x_1 \leq 0\}$. Note that all the involved functions are linear, and the unique optimal response to a given opponent strategy x_{-j} is $x_j = x_{-j}$ so that the equilibria of the game is the set $\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = x_2 = \alpha, \alpha \in \mathbb{R}\}$ and the value for both players at an equilibrium point (α, α) is α, for $\alpha \in \mathbb{R}$.*

III.1.4 Structure of the Set of Solutions of a GNEP

The structure of the set of equilibria of a GNEP (even of a NEP) may be quite complicated. For instance, it is not necessarily convex. An example, with a non-convex set of solutions, can be easily constructed by noting that the graph of the solution of a parametric convex optimization problem (convex on both the optimization variable and the parameter) is not convex in general. Using the above fact, we show a (G)NEP whose solution is not even connected.

Example 11. *Let* $x = (x_1, x_2) \in [0,4]^2$ *and* $f_j(x) := d_{T_j}(x)^2$*, where* T_1 *is the triangle with vertices* $(0,0)$*,* $(0,4)$ *and* $(1,2)$*, and* T_2 *is the triangle whose vertices are* $(0,0)$*,* $(4,0)$ *and* $(2,1)$*. Let* $S_j(x_{-j}) := \operatorname{argmin}_{x_j} \{f_j(x_1,x_2) \mid x_j \in [0,4]\}$*. We see that* $S_1(x_2) = \{x_1 \in [0,4] \mid (x_1,x_2) \in T_1\}$ *for* $x_2 \in [0,1]$ *and* $S_1(x_2) = \{2\}$ *for all* $x_2 \in (1,4])$*. Similarly,* $S_2(x_1) = \{x_2 \in [0,4] \mid (x_1,x_2) \in T_2\}$ *for* $x_1 \in [0,1]$ *and* $S_2(x_1) = \{2\}$ *for all* $x_1 \in (1,4])$*. In Figure III.1.1, the graphs of* S_1 *(in blue) and of* S_2 *(in red) are drawn. The two points of intersection of these graphs (in black) are the equilibrium points of the (G)NEP.*

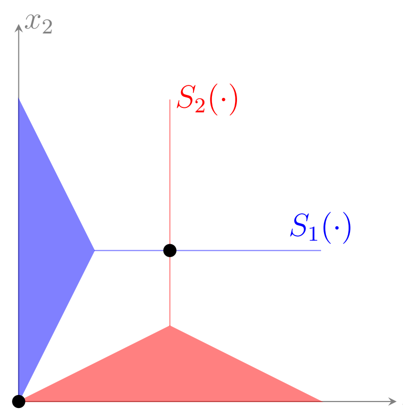


Figure III.1.1: In red the reaction map of player 1 and in blue the reaction map of player 2. The two nodes in black (the intersection of these maps) are the equilibrium points of the NEP of Example 11

.

Let us end this subsection by showing that, even if the GNEP is linear, the set of equilibria can be non-connected, actually composed of two isolated points (see also [111]).

Example 12. *Let there be two players with variables* x_1 *and* x_2*, both in* $\mathbb{R}$*. Let the problem of player 1 be*

$$\min_{x_1} \{x_1 \mid x_1 \in [0,1], x_1 \geq 2x_2 - 1\}$$

and the problem of player 2 be

$$\min_{x_2} \{x_2 \mid x_2 \in [0,1], x_2 \geq 2x_1 - 1\}.$$

It is easy to note that, for a given opponent's strategy, each of these problems has a unique solution, and that the reaction maps are given by $x_1(x_2) = \max\{0, 2x_2 - 1\}$ *and* $x_2(x_1) = \max\{0, 2x_1 - 1\}$*, for player 1 and 2, respectively. Thus, the equilibria are the points* $(0,0)$ *and* $(1,1)$*.*

III.1.5 GNEP and First Order Stationarity Conditions

Following the same lines as in the case of parametric optimization (see Section I.5 in Chapter I), all along this section we will make the following assumptions on the GNEP:

(H_1') (Player differentiability) For every $j \in J$ and every $y_{-j} \in \mathbb{R}^{m_{-j}}$, $f_j(\cdot, y_{-j})$ and $g_j(\cdot, y_{-j})$ are differentiable.

(H_2') (Player convexity) For every $j \in J$ and every $y_{-j} \in \mathbb{R}^{m_{-j}}$, $f_j(\cdot, y_{-j})$ is convex and the components of $g_j(\cdot, y_{-j})$ are quasi-convex functions.

In order to compute an equilibrium of a Generalized Nash game, it is commonly used to consider the complementarity problem composed by the concatenation of the KKT conditions of each of the players. For player j, given its own strategy y_j and an opponent strategy y_{-j}, let $\Lambda_j(y_j, y_{-j})$ denote the set of *Lagrange multipliers* μ_j of the problem $(\mathcal{P}_j(y_{-j}))$ at the point y_j, that is, $\mu_j \in \mathbb{R}^{d_j}$ satisfying

$$\begin{cases} \nabla_{y_j} f_j(y_j, y_{-j}) + \nabla_{y_j} g_j(y_j, y_{-j})\mu_j = 0 \\ 0 \le \mu_j \perp -g_j(y_j, y_{-j}) \ge 0 \end{cases} \qquad (KKT_j(y_{-j}))$$

Given y_{-j}, the set of points y_j such that $\Lambda_j(y_j, y_{-j}) \neq \emptyset$ is denoted by $\mathrm{SP}_j(y_{-j})$ (standing for Stationary Points). Moreover, let cSP denote the solution of the concatenation of stationary point/KKT conditions of all players, that is,

$$\begin{aligned} \mathrm{cSP} \; &:= \{y = (y_1, ..., y_{|J|}) : \quad \forall\, j \in J,\; y_j \in \mathrm{SP}_j(y_{-j})\} \\ &= \{y = (y_1, ..., y_{|J|}) : \quad \forall\, j \in J, \exists\, \mu_j \in \mathbb{R}^{d_j} \text{ with} \\ &\qquad\qquad (y_j, \mu_j) \text{ solution of } \mathrm{KKT}_j(y_{-j})\}. \end{aligned}$$

With assumptions (H_1'),(H_2'), the KKT conditions are sufficient for GNEP, that is, cSP $\subset$ GNEP . Thus, a computed solution of the concatenation of KKT conditions yields a solution of the GNEP, but there could be some GNEP solutions that cannot be obtained from a cSP solution. We are interested in studying constraint qualifications that ensure the concatenation of KKT conditions to be also necessary optimality/equilibrium conditions for the GNEP, that is, cSP $=$ GNEP. In this case, we would be able to find any Nash equilibrium of the game by solving the concatenated KKT conditions.

In a GNEP, a player may face a possibly infinite number of different optimization problems indexed in the opponents strategies. Let us recall that not only the cost function depends on the opponent strategy, but in a GNEP the constraint set of each player depends on the opponent strategy. Thus, at a first glance, one would require that for each player and for each opponent strategy, some constraint qualification should be fulfilled at each optimum strategy of this player. But the optimum strategies are not known in advance and thus one is forced to assume/verify the constraint qualifications at every feasible strategy. Under such strong conditions, one obviously has that GNEP $=$ cSP.

Indeed, we already know that, thanks to the convexity assumption (H_2'), cSP $\subset$ GNEP. On the other hand for any $y \in$ GNEP and for any j, $y_j \in S_j(y_{-j})$, and thus, if a constraint qualification (Guignard's CQ for instance) holds at this point, then $y_j \in \mathrm{SP}_j(y_{-j})$ showing the equivalence between the generalized Nash game and the corresponding cSP.

Proposition 10. *Assume* (H_1'), (H_2') *and that for each player j and each feasible joint strategy $y = (y_j, y_{-j}) \in Y$, Guignard's CQ holds for player j's constraint "$g_j(\cdot, y_{-j}) \le 0$" at y_j. Then* GNEP $=$ cSP.

Note that in Proposition 10, the constraint qualifications are assumed for each feasible joint strategies $y = (y_j, y_{-j})$. A first reduction on the "number" (in the sense of a smaller set) of conditions can be done if the constraint functions $g_{j,k}(\cdot, y_{-j})$ are (not only quasi-convex but) convex, and using Slater's CQs instead of Guignard's CQs.

Corollary 10. *Assume (H_1') and (H_2'). Moreover assume that, for each player j, both following properties hold:*

1. *Each $g_{j,k}(\cdot, y_{-j})$ is convex, for $k = 1, ..., d_j$;*
2. *For any y_{-j} such that $\exists\, y_j \in \mathbb{R}^{m_j}$ with $(y_j, y_{-j}) \in Y$, Slater's CQ holds: there exists $\tilde{y}_j$ such that $g_j(\tilde{y}_j, y_{-j}) < 0$.*

Then $\text{GNEP} = \text{cSP}$.

III.1.5.1 Joint Convexity and Reduction on Number of CQs

As in Section I.5, we would like to reduce as much as possible the "number" of constraint qualifications to be verified in order to obtain the desired equivalence between a GNEP and its associated problem cSP. To this aim, we will assume some joint convexity[1] (in the sense of Definition 2) of the components g_j of the constraint functions.

Remark 13. *Joint convexity is a stronger condition on the constraints than player convexity (H_2'). For instance, a bilinear function $g(y_1, y_2) := ay_1y_2$, for $y_1, y_2 \in \mathbb{R}$, is player convex while it is not jointly convex (unless a is equal to 0).*

We are interested in finding simple qualification conditions over the joint constraint functions g_j which imply Guignard's CQ for any opponent strategy. The following definition is related to Definition 1 for general parametric optimization problems.

Definition 8. *Let $j \in J$. An opponent strategy vector $\hat{y}_{-j} \in \mathbb{R}^{m_{-j}}$ is called*

- admissible *(for player j) if $\hat{y}_{-j} \in \mathcal{A}_j :=$* $\operatorname{dom} Y_j$*, that is, there exists y_j such that $g_j(y_j, \hat{y}_{-j}) \leq 0$;*
- interior *(for player j) if it is an element of $int(\mathcal{A}_j)$;*
- boundary *(for player j) if it is an element of $bd(\mathcal{A}_j)$.*

Combining the joint convexity of the constraint functions, joint Slater's constraint qualifications and constraint qualification only for boundary opponent strategy, one can conclude the equivalence between the GNEP and its associated cSP.

Theorem 20. *Assume (H_1'), (H_2') and that, for each $j \in J$, each the three properties hold:*

1. *(Joint convexity) The function g_j is jointly convex on $\mathbb{R}^m$;*
2. *(Joint Slater's CQ) There exists a joint strategy $\tilde{y}(j) \in \mathbb{R}^m$ such that $g_j(\tilde{y}(j)) < 0$;*
3. *(Guignard's CQ for boundary opponent strategies) For any $\hat{y}_{-j} \in bd(\mathcal{A}_j)$ and y_j such that $(y_j, \hat{y}_{-j}) \in Y$, Guignard's CQ holds for j's problem at y_j.*

[1] This concept is not the joint convexity of a game as defined by Rosen. In general they are not comparable, but our assumption is stronger when all players has the same constraint, i.e. $g_j = g$, for all $j \in J$.

Then GNEP = cSP.

Proof. For each player $j \in J$, the hypotheses of Theorem 1 for player j's parametric optimization problem holds with $U := \{y_{-j} \mid \exists y_j : (y_j, y_{-j}) \in Y\}$, so that $Y \cap \operatorname{gph} S_j = Y \cap \operatorname{gph} \mathrm{cSP}_j$. By taking intersection over $j \in J$, and recalling that GNEP $\subset Y$, we conclude that GNEP = cSP. □

Now we present two examples that will illustrate the use of the assumptions in Theorem 20. In the first one, two players face some boundary opponents strategies but still the hypotheses of Theorem 20 are satisfied. In the second example, we show a shared constraints GNEP for which the first and second assumptions of Theorem 20 are satisfied, but not the third one: at some boundary opponent strategies the constraint qualifications are not fully satisfied. We also exhibit a GNEP solution that is not an element of cSP, thus, showing that it is not possible to drop Assumption 3) in Theorem 20.

Example 13. *Let q_1, q_2 and r be three fixed real numbers and consider a GNEP for which the constraints for player 1 and for player 2 are respectively*

$$\begin{cases} y_1 \geq 0 \\ g(y_1, y_2) \leq 0 \end{cases} \quad \text{and} \quad \begin{cases} y_2 \geq 0 \\ g(y_1, y_2) \leq 0 \end{cases}$$

where the function g is defined as $g(y_1, y_2) := (y_1 - q_1)^2 + (y_2 - q_2)^2 - r^2$. The constraint

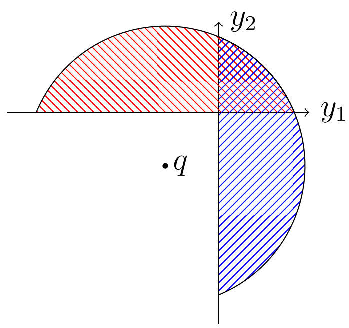


Figure III.1.2: Feasible region of the GNEP of Example 13 with $q_1, q_2 < 0$. In blue the feasible region for player 1, and in red the feasible region for player 2

functions are clearly jointly convex. If $r > \sqrt{2}\max(-q_1, -q_2, 0)$ it is easy to find a Joint Slater strategy/point (the same for both players): $(y_1, y_2) \in \mathbb{R}^2$ such that $y_1, y_2 > 0$ and $g(y_1, y_2) < 0$. For simplicity, let us analyse only the case when $q_1, q_2 \leq 0$, (the other case is left to the reader). A player j has two boundary parameters, one positive and one negative. The negative one is not feasible for player $-j$, and the positive one gives only one feasible point at which the gradients of y_j and g with respect to the variable y_j generates the normal cone $\mathbb{R}$ if and only if $q_j < 0$, i.e. Guignard's CQ holds if and only if $q_j < 0$. Thus, applying Theorem 20 for any GNEP *with differentiable and convex player objectives and whose constraints are the ones just described, we conclude that the set of generalized Nash equilibria coincide with the solution set of the corresponding KKT system if and only if $q_1, q_2 < 0$.*

Example 14. *Consider now two players with variables $y_1 \in \mathbb{R}^2$ and $y_2 \in \mathbb{R}$, respectively. Fix two points $q_1 = (-4, 0, 0), q_2 = (4, 0, 0) \in \mathbb{R}^3$ and let both players share the following two* *constraint functions* $g_k(y) = \|y - q_k\|^2 - 5^2$, $k = 1, 2$, *where* $y = (y_1, y_2)$. *These functions clearly satisfy the two first assumptions of Theorem 20, but not the third. The feasible region is* $Y := B(p, 5) \cap B(q, 5)$. *Consider the joint strategy* $(y_1, y_2) = (0, 0, 3) \in Y$. *The gradients of* g_1, g_2 *w.r.t.* y_1 *at the point* $(0, 0, 3)$ *are parallel to each other so they cannot characterize the normal cone to player 1's feasible set* $Y_1(3) = \{(0, 0)\}$, *which is two dimensional. So, Guignard's CQ does not hold for player 1's problem, and taking for instance, the cost functions as* $f_1(y_1, y_2) = y_{1,2}$ *and* $f_2(y_1, y_2) = -y_2$, *then* $y = (y_1, y_2) = (0, 0, 3)$ *is a generalized Nash equilibrium but it is not a solution of the corresponding concatenated KKT system (see Figure III.1.3).*

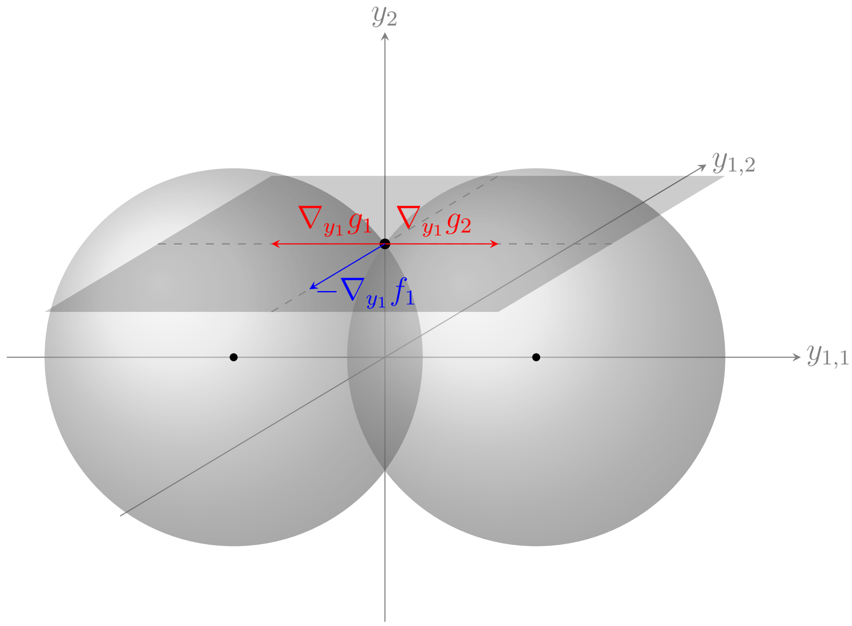


Figure III.1.3: In Example 14, the point $y_1 = (0, 0)$ is the unique feasible point in the y_1-plane with height $y_2 = 3$. The joint strategy $(y_1, y_2) = (0, 0, 3) \in \text{GNEP}$, but no multipliers exist for player 1.

In the case of (non generalized) Nash games, boundary parameters do not exist so that we can deduce from Theorem 20 the following simpler form:

Corollary 11 (Nash game version)**.** *Assume* $(H_1'), (H_2')$ *and that, for each player* $j \in J$, *the constraint function* g_j *does not depend on the other players' variable* y_{-j}*Moreover assume that, for each player* $j \in J$,

1. *(Joint convexity) The constraint function* g_j *is jointly convex on* $\mathbb{R}^m$;

2. *(Joint Slater's CQ) There exists a (joint) strategy* $\tilde{y}(j) \in \mathbb{R}^{m_j}$ *such that* $g_j(\tilde{y}(j)) < 0$.

Then $\text{GNEP} = \text{cSP}$.

III.1.5.2 Finite Number of CQs

Our approach in Theorem 20 is to verify only a reduced set of constraint qualifications and obtain the same conclusion as in Proposition 10. But Assumption 3) of Theorem 20 still requires this verification of constraint qualifications for a possibly infinite set of boundary parameters. In this subsection, we will consider two feasibility conditions that makes this third assumption trivially satisfied, by avoiding the existence of boundary opponent strategies. The first condition we define is a strong one but very simple to express.

Definition 9. *A GNEP is said to be* fully inter-feasible *if:*

$$\forall j \in J: \ \mathcal{A}_j = \mathbb{R}^{m_{-j}}. \tag{III.1.1}$$

In other words, a GNEP is fully inter-feasible if for all $y_{-j} \in \mathbb{R}^{m_{-j}}$, there exist a feasible strategy $y_j \in Y_j(y_{-j}) = \{y_j \mid g_j(y_j, y_{-j}) \leq 0\}$. Clearly, the feasibility condition (III.1.1) implies the unboundedness of $\operatorname{gph} Y_j$, but it really makes sense. It could be read as follows: for each player j, the strategies of the other players may affect his feasible set (even reduce it to a singleton), but cannot make his problem infeasible.

The second condition we consider is weaker than condition (III.1.1), but still avoids the existence of boundary opponent strategies in the game.

Definition 10. *We say that a GNEP* strictly inter-feasible *if:*

$$\forall j \in J: \ P_{-j}(Y) \subset int(\mathcal{A}_j). \tag{III.1.2}$$

where $P_{-j}(Y) := \{y_{-j} \mid \exists\, y_j : \ (y_j, y_{-j}) \in Y\}$.

The following corollary is then a direct consequence of Theorem 20.

Corollary 12. *Assume* (H_1'), (H_2') *and that the* GNEP *is either fully inter-feasible or strictly inter-feasible. Assume moreover that for each* $j \in J$ *both of the following properties holds:*

1. *(Joint Convexity) The constraint function* g_j *is jointly convex with respect to* y*;*

2. *(Joint Slater's CQ) There exists a joint strategy* $\tilde{y}(j) \in \mathbb{R}^m$ *such that* $g_j(\tilde{y}(j)) < 0$.

Then GNEP = cSP.

To give an insight of what condition (III.1.2) means, let us see the following GNEP which is strictly inter-feasible but not fully inter-feasible.

Example 15. *Let us consider a game composed of two players with real variables* y_1 *and* y_2*, respectively, and constraint functions* $g_1, g_2 : \mathbb{R}^2 \to \mathbb{R}$ *given by* $g_1(y_1, y_2) := 4(y_1 - 3)^2 + (y_2 - 3)^2 - 4$ *and* $g_2(y_1, y_2) := (y_1 - 3)^2 + 4(y_2 - 3)^2 - 4$ *(see the feasible set in Figure III.1.4).*

Let us analyze the feasibility conditions for player 1's problem. First, condition (III.1.1) *is not satisfied, since for instance,* $y_2 = 0 \notin \mathcal{A}_1$*. Second, the condition* (III.1.2) *is satisfied. In fact, we have that* $P_2(Y) = [2, 4]$ *is obviously included in the interior of* $\mathcal{A}_1 = [1, 5]$*. Given the symmetry of the problem we conclude that condition* (III.1.2) *is also fulfilled for player 2.*

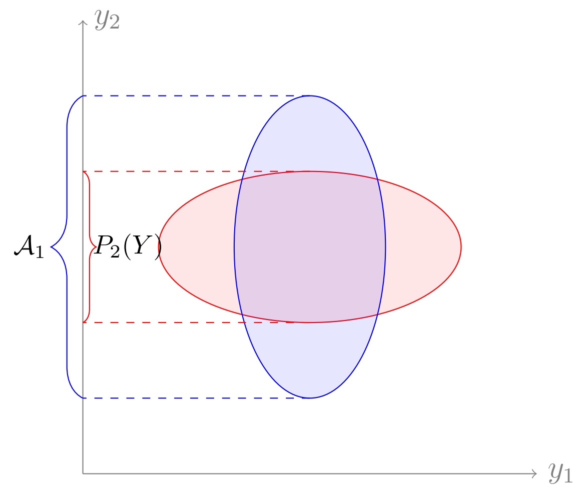


Figure III.1.4: In Example 15, all opponent-feasible strategies $y_2 \in P_2(Y)$ are interior parameters for player 1 (and similarly for player 2).

Chapter III.2

Bilevel Programming Problems

This chapter corresponds to the paper [18].

III.2.1 Introduction

A bilevel programming problem, or simply a bilevel problem, is a hierarchical optimization problem which models a game between two players: In economics it is called a Stackelberg game. One of the players is the leader and the second one is the follower. Each player tries to minimize its own cost function which depends on the decision of both players, but the leader decides first knowing that then the follower will react in an optimal way given the decision of the leader. Thus, the leader optimizes his objective based on a conjectured reaction of the follower.

There is a certain ambiguity, as noticed in [40], in the formulation of the upper level problem; an ambiguity that occurs in the case of non-uniqueness of the lower level optimal solutions. To handle this difficulty it has been proposed in the literature to consider an optimistic and a pessimistic approach, also known as strong Stackelberg game and weak Stackelberg game, respectively. The *optimistic approach* occurs when the leader can convince the follower to cooperate that is, the leader can select in between all the optimal reactions of the follower, the one with least cost for him. Alternatively, if the cooperation of the follower may not be possible, for instance if cooperation is not allowed by law, or if, for some reasons, the follower systematically chooses the worst case for the leader then the leader would need to bound the damages resulting from undesirable decisions of the follower. This kind of interactions is perfectly modeled by the *pessimistic approach* of the bilevel problem.

The pessimistic approach has often been put aside while the optimistic one has received much more attention [41, 43, 122]. However, some work has been done for the pessimistic approach: existence of solutions has been studied in [75, 77]; optimality conditions were studied in [37, 44]; it has been proposed approximating (a certain class of) pessimistic bilevel problems by sequences of optimistic ones in [77]; a solution procedure based on semi-infinite programming was presented in [122] and for the linear case some procedures have been developed in [126]. The pessimistic approach has also been useful in a more general setting of several followers (playing a non-cooperative game between them, see Chapter III.3) to model a certain water resources optimal allocation problem in [125]. Despite the simplicity of the optimistic approach it has the drawback of non-realistic assumptions in the model: the cooperation of the follower with the leader but without any reward considered in the objective. By the contrary, the pessimistic approach has a simple interpretation of bounding damages and minimizing the risk worst case.

Let us now discuss the main problem/question studied in this chapter. In order to compute some solutions for the bilevel programming problem, it is commonly used to reformulate it as a mathematical program with complementarity constraints (MPCC for short): this is done by replacing the lower level problem, (which we assume to be convex) by its KKT optimality conditions. This reformulation plays a central role in the numerical treatment of the difficult bilevel problem since it is usually the MPCC reformulation that is (locally) solved by the use of any of the available algorithms. It is thus very important to know if the bilevel programming problem and its reformulation are equivalent. In particular, one needs to know if the solution obtained by solving the MPCC may generate a solution of the initial bilevel model. As it will be shown later on, the answer to this questionis "no" for the important case of local solutions. Actually, as it will be observed (see Example 16 and Remark 20), this bad situation is not a pathological (or exceptional) one. Indeed, this example shows a local solution of the MPCC reformulation of an elementary bilevel problem which does not correspond to any local solution of the (pessimistic) bilevel programming problem. The concept of local solutions used for the example and for the relation of the bilevel and the MPCC problems will be precised in the following sections.

The interrelation between solutions of the optimistic bilevel problem and solutions of its associated MPCC, in the optimistic case, has been fully addressed in [41]. More precisely, Dempe and Dutta have shown that, even in the optimistic case, this interrelation is not so direct and that, for a point to be local solution of the optimistic bilevel problem, some Slater type constraint qualifications of the lower level problem are needed and local optimality of the associated MPCC must be satisfied for all Lagrange multipliers (see [41, Theorem 3.2]).

Our aim in this chapter is to provide a counterpart, for the pessimistic bilevel problems, of the pioneering analysis driven by Dempe and Dutta in [41] for the optimistic approach. In Section III.2.2, we explain the notation we use, we define the pessimistic bilevel problem and the associated (pessimistic) MPCC, and we {analyze what we call the basic assumption, that is somehow inevitable for considering the pessimistic approach. In Section III.2.3, we focus on global solution concepts for both the pessimistic bilevel problem and its associated MPCC and investigate the interrelation of the solution sets of those problems. Section III.2.4 is dedicated to the important case of local solutions of both the pessimistic bilevel programming problem and its MPCC reformulation. For such minimax problems, we have defined notions of local solutions in both variables of the leader and the follower. Let us emphasize that, as far as we know, no local concept of solutions for those pessimistic problems has been previously defined in the literature. We show that the interrelations between the solution sets of the two problems are far from trivial and depend (in the same spirit as for the optimistic problem) on some Slater type constraint qualification of the lower level problem. The subtlety of the situation is illustrated through two simple examples (See Example 16 and Remark 20). In Section III.2.5, we give some final comments to summarise the work.

III.2.2 Preliminaries and Problem Statement

In this section, we describe more precisely the bilevel problem, and the MPCC reformulation, in their pessimistic forms.

Let the vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ be the decisions of the leader and the follower respectively. The problem of the leader, that we also refer to as the upper level problem, is

$$\text{“}\min_{x\in X}\text{”}\{F(x,y) \mid y \in S(x)\}, \tag{III.2.1}$$

where X is a non-empty closed subset of $\mathbb{R}^n$, $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is the leader's cost function and $S(x)$ stands for the solution set of the following lower level problem, also called the follower's problem,

$$\min_{y\in\mathbb{R}^m} \{f(x,y) \mid g(x,y) \leq 0\}, \tag{III.2.2}$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is the follower's cost function and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^d$ is the follower's constraint function. Through all the text, f and the components g_i, $i = 1, ..., d$ are assumed to be convex and differentiable with respect to the lower level variable y, and F to be continuous on y.

The quotation marks in (III.2.1) have been used to emphasize the ambiguity in the formulation of the leader's problem: as it has been commented in the introduction, it occurs when the follower has possibly more than one optimal reaction (that is $x \mapsto S(x)$ is not single-valued) and it is handled in the pessimistic approach by minimizing the highest value of its cost, giving rise to the following minmax problem

$$\min_{x\in X} \max_{y\in S(x)} F(x,y). \tag{PB}$$

The optimistic approach is just as (PB) but with a “min-min” instead of the “min-max”.

We assume through all this text that the lower level problem is everywhere solvable, that is,

$$S(x) \neq \emptyset, \quad \text{for every } x \in X. \tag{III.2.3}$$

This is a very basic assumption, which ensures that the max in (PB) is taken over a non-empty set, and is implied, for instance, by the compactness of the constraint region $\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m : g(x,y) \leq 0\}$ and the non-emptiness of $\{y \in \mathbb{R}^m : g(x,y) \leq 0\}$ for each $x \in X$.

For any given $x \in X$, let us define the *(partial) pessimistic value function* by

$$\varphi_p(x) := \max_{y\in S(x)} F(x,y)$$

and $S_p(x)$ the solution set of this partial optimization problem that is the set of $y \in S(x)$ which attain the maximum in the definition of $\varphi_p(x)$. This set $S_p(x)$ actually describes the optimal reactions of the follower that are the worst for the leader and, for any decision x of the leader, $\varphi_p(x)$ is the worst value that the leader could face. Observe that assumption (III.2.3) is needed to have $\varphi_p(x) > -\infty$. Note that in the definitions of (PB), and thus of φ_p, a maximum is involved and not a supremum. We will therefore assume the following general assumption:

$$S_p(x) \neq \emptyset, \quad \text{for every } x \in X. \tag{III.2.4}$$

This can be obtained, for instance, under the same conditions as for (III.2.3).

In practice, to solve problem (PB) the constraint $y \in S(x)$ is commonly replaced by the KKT system of the lower level problem, giving place to a Mathematical program with complementarity constraints (MPCC). Recall that a point y satisfies the KKT conditions of the parametrized optimization problem (III.2.2), if there exists $\mu \in \mathbb{R}^d$, called Lagrange-multiplier, for which the couple $(y, \mu) \in \mathbb{R}^m \times \mathbb{R}^d$ satisfies the KKT system

$$\begin{cases} \nabla_y f(x, y) + \mu^T \nabla_y g(x, y) = 0, \\ \quad 0 \le \mu \perp -g(x, y) \ge 0. \end{cases} \tag{III.2.5}$$

For fixed $x \in X$ the set of pairs (y, μ) which solve the system (III.2.5) is denoted by $KKT(x)$, and we denote by $SP(x)$ (for "Stationary Points") the set of those y such that for some μ, $(y, \mu) \in KKT(x)$. Fixing also y, let $\Lambda(x, y)$ be the set of Lagrange multipliers, that is, $\mu \in \mathbb{R}^d$ satisfying (III.2.5).

Therefore, the pessimistic (MPCC) associated with (PB) is

$$\min_{x \in X} \max_{(y,\mu) \in KKT(x)} F(x, y). \tag{MPCC}$$

As for the (PB) we make the following assumption

$$KKT(x) \neq \emptyset, \quad \forall x \in X, \tag{BA}$$

that we call the *Basic Assumption.* Note that this assumption is equivalent to $SP(x) \neq \emptyset$ for all $x \in X$. Assumption (BA) is more delicate to verify than (III.2.3), but it is an inevitable assumption for considering problem (MPCC) so that the max is taken over a non-empty set. An easily verifiable condition that implies the (BA) is the following:

Slater's CQ: for any $x \in X$, there exists $y(x)$ such that $g_i(x, y(x)) < 0$, for all $i = 1, ..., d$.

Similarly to the first problem, given $x \in X$, let $\psi_p(x) := \max_{(y,\mu) \in KKT(x)} F(x, y)$ and $KKT_p(x)$ the couples (y, μ) that attain the maximum in the definition of $\psi_p(x)$. Similarly, let $SP_p(x)$ be the set of y in $\mathbb{R}^n$ for which there exists $\mu \in \mathbb{R}^d$ satisfying $(y, \mu) \in KKT_p(x)$.

Let us summarize the different set-valued maps that we defined and that will play a (central) role in the sequel. The maps $S, S_p, SP, SP_p : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and $KKT, KKT_p : \mathbb{R}^n \rightrightarrows \mathbb{R}^{m+p}$ are respectively defined by

- $S(x) = \{y \in \mathbb{R}^m \; : \; y$ solution of the lower level problem$\}$;
- $S_p(x) = \{y \in S(x) \; : \; y$ global solution of $\max_{y' \in S(x)} F(x, y')\}$;
- $SP(x) = \{y \in \mathbb{R}^m \; : \; \Lambda(x, y) \neq \emptyset\}$;
- $SP_p(x) = \{y \in SP(x) \; : \; y$ global solution of $\max_{y' \in SP(x)} F(x, y')\}$;
- $KKT(x) = \{(y, u) \in \mathbb{R}^{m+d} \; : \; u \in \Lambda(x, y)\}$;
- $KKT_p(x) = \{(y, \mu) \in KKT(x) \; : \; (y, \mu)$ global solution of $\max_{(y',\mu') \in KKT(x)} F(x, y')\}$.

III.2.3 Global Solution Concepts

This section is dedicated to the analysis of interrelations between the (global) solution sets of the pessimistic bilevel problem and the associated MPCC. Different concepts of global solution are considered.

Let us first describe two different, but at the same time natural, definitions of global solution to (PB). The first one has been considered in [44] and corresponds to saying that to solve (PB), the leader has to choose an x that minimizes the worst value $\varphi_p(x)$.

Definition 11. *A point $\bar{x} \in \mathbb{R}^n$ is an* original solution *of* (PB)*, if $\bar{x} \in X$ and for all $x \in X$, $\varphi_p(\bar{x}) \leq \varphi_p(x)$.*

A second definition was considered in the reference monograph [40], which involves at the same time both the decision vectors of the leader and of the follower, and we call it here "conventional solution". The terms "original" and "conventional" are taken as the names given to the corresponding optimistic problems in [43].

Definition 12. *A pair $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ is a* conventional solution *of* (PB)*, if $\bar{x} \in X$, $\bar{y} \in S_p(\bar{x})$ and $\varphi_p(\bar{x}) \leq \varphi_p(x)$, for all $x \in X$.*

Equivalently, one can say that a pair $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ is a conventional solution of (PB) if $(\bar{x}, \bar{y})$ is in the graph of the set-valued map S_p and $\bar{x}$ minimizes φ_p over X.

Remark 14. *It is clear from the definition that if $(\bar{x}, \bar{y})$ is a conventional solution, then the first coordinate $\bar{x}$ is an original solution of* (PB)*. Conversely, if $\bar{x}$ is an original solution, then for any $\bar{y} \in S_p(\bar{x})$ the couple $(\bar{x}, \bar{y})$ is a conventional solution of* (PB)*.*

While for the pessimistic bilevel problem (PB) two concepts of solutions were defined above, for its associated (MPCC), three definitions of global solutions naturally arise.

A first type of solution is defined which considers only the variable of the leader x, and is, in essence, analogous to Definition 11.

Definition 13. *A point $\bar{x}$ is an* original solution *of* (MPCC) *if $\bar{x} \in X$ and $\psi_p(\bar{x}) \leq \psi_p(x)$, for all $x \in X$.*

Now considering both variables of leader and follower, another definition of global solution can be considered.

Definition 14. *A couple $(\bar{x}, \bar{y})$ is a* conventional solution *of* (MPCC) *if $\bar{x} \in X$, $\bar{y} \in SP_p(\bar{x})$ and $\psi_p(\bar{x}) \leq \psi_p(x)$, for all $x \in X$.*

Remark 15. *As for the pessimistic bilevel problem, it is direct that if $(\bar{x}, \bar{y})$ is a conventional solution of* (MPCC)*, then $\bar{x}$ is an original solution of* (MPCC)*. Conversely, if $\bar{x}$ is an original solution of* (MPCC)*, then for any $\bar{y} \in SP_p(\bar{x})$ the pair $(\bar{x}, \bar{y})$ is a conventional solution of* (MPCC)*.*

Now, since the (MPCC) involves the Lagrange multipliers μ, a third concept of global solution can be defined.

Definition 15. *A triplet $(\bar{x}, \bar{y}, \bar{\mu})$ is a* full solution *of* (MPCC) *if $\bar{x} \in X$, $(\bar{y}, \bar{\mu}) \in KKT_p(\bar{x})$ and $\psi_p(\bar{x}) \leq \psi_p(x)$, for all $x \in X$.*

We call it full because it considers all the variables including μ, even though μ is not a variable of the function.

Remark 16. *The relation between full and conventional solutions for* (MPCC) *is very simple. If $(\bar{x}, \bar{y})$ is a conventional solution, then $\Lambda(\bar{x}, \bar{y})$ is non-empty and $(\bar{x}, \bar{y}, \bar{\mu})$ is a full solution for each $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$. Conversely, if $(\bar{x}, \bar{y}, \bar{\mu})$ is a full solution, then $(\bar{x}, \bar{y})$ is a conventional solution.*

Now we have come to the main point in this section of comparing the global solutions of the pessimistic bilevel problem (PB) with those of (MPCC). Recall that (MPCC) appears by replacing the lower level problem by its associated KKT system. Then, to answer the question posed in the title it is convenient to recall the relation between the KKT system with the lower level problem. Under the general convexity assumption of the lower level, it is well known that the KKT conditions are sufficient, that is

$$SP(x) \subset S(x), \quad \forall x \in X.$$

The KKT conditions to be necessary depends on some regularity condition known as constraint qualification. For instance, it is well known that if for a given x, Slater's CQ holds for the lower level (III.2.2), then $S(x)$ and $SP(x)$ coincide. Now we prove that these sets still coincide under a weaker condition.

Proposition 11. *Assume that the lower level problem* (III.2.2) *is convex. Let $x \in X$ be such that $KKT(x) \neq \emptyset$. Then $SP(x) = S(x)$.*

Proof. Let $x \in X$ be fixed. We only need to prove that $S(x) \subset SP(x)$. Take $y \in S(x)$. Since $KKT(x) \neq \emptyset$, $SP(x)$ is non-empty, so take $y_0 \in SP(x)$. Then the set of Lagrange multipliers $\Lambda(x, y_0)$ is non-empty. We know that $SP(x) \subset S(x)$ so that $y_0 \in S(x)$. But in the convex case the set of Lagrange multipliers is the same for all solutions of the lower level problem (see [57, Proposition 3.1.1, VII]), so then $\Lambda(x, y) = \Lambda(x, y_0) \neq \emptyset$. Thus $y \in SP(x)$ and the proof is completed. □

Remark 17. *Consider a fixed $x \in X$. Since F does not depend on μ, and $\Lambda(x, y)$ is a constant (non-empty) set for all $y \in SP(x)$, we observe that the maximization in the definition of $KKT_p(x)$ can be seen as only in the variable $y \in SP(x)$. In fact, we have the following representation*

$$KKT_p(x) = S_p(x) \times \Lambda(x, y), \quad \forall y \in SP(x).$$

As a direct consequence of the above proposition, the convexity assumption added to our (BA) gives us that $SP(x) = S(x)$ for each $x \in X$. Taking this into account we adapt this result for the setting of bilevel programming.

Theorem 21. *Assume that the lower level problem* (III.2.2) *is convex.*

i) Assume that $KKT(\bar{x}) \neq \emptyset$ and that $\bar{x}$ is an original (resp. $(\bar{x}, \bar{y})$ is a conventional) solution of (MPCC). *Then $\bar{x}$ is an original (resp. $(\bar{x}, \bar{y})$ is a conventional) solution of* (PB).

ii) Assume that (BA) *holds and that $\bar{x}$ is an original (resp. $(\bar{x}, \bar{y})$ is a conventional) solution of* (PB). *Then $\bar{x}$ is an original (resp. $(\bar{x}, \bar{y})$ is a conventional) solution of* (MPCC).

Proof. i) The convexity assumption gives us that $SP(x) \subset S(x)$, and so taking supremum over these two sets we get that $\psi_p(x) \leq \varphi_p(x)$, for all $x \in X$. But since $SP(\bar{x})$ is non-empty, by Proposition 11 we have that $SP(\bar{x}) = S(\bar{x})$, and thus $\varphi_p(\bar{x}) = \psi_p(\bar{x})$. Since $\bar{x}$ is an original solution of (MPCC), then

$$\varphi_p(\bar{x}) = \psi_p(\bar{x}) \leq \psi_p(x) \leq \varphi_p(x), \quad \forall x \in X$$

and therefore $\bar{x}$ is also an original solution of (PB). The proof is similar for the conventional concepts.

ii) The equivalence between problem (PB) and (MPCC) in terms of the concepts of original solution is a direct consequence of the combination of Proposition 11 with the assumption (BA) since in this case $\varphi_p(x) = \psi_p(x)$, for any $x \in X$. For the equivalence in terms of conventional solutions just note that since $SP(\bar{x}) = S(\bar{x})$, then also $SP_p(\bar{x}) = S_p(\bar{x})$. □

Remark 18. *In Theorem 21, it seems that the assumption for* (*i*) *is weaker than the one for* (*ii*)*, but this is not the case. The hypothesis* $SP(\bar{x}) \neq \emptyset$ *with* $\bar{x}$ *being an original solution of* (MPCC)*, implies that* $-\infty < \psi_p(\bar{x}) \leq \psi_p(x)$ *for all* $x \in X$ *so that* $SP(x)$ *is non-empty for each* $x \in X$*, that is,* (BA) *holds.*

Finally, we state the connection between conventional solutions of (PB) and full solutions of (MPCC) under (BA), which follows from Remark 16 and Theorem 21.

Corollary 13. *Assume that the lower level problem* (III.2.2) *is convex.*

i) If $KKT(\bar{x}) \neq \emptyset$ *then:* $(\bar{x}, \bar{y}, \bar{\mu})$ *is a full solution of* (MPCC) *implies that* $(\bar{x}, \bar{y})$ *is a conventional solution of* (PB).

ii) If (BA) *holds then:* $(\bar{x}, \bar{y})$ *is a conventional solution of* (PB) *implies that* $\Lambda(\bar{x}, \bar{y}) \neq \emptyset$ *and* $(\bar{x}, \bar{y}, \bar{\mu})$ *is a full solution of* (MPCC)*, for each* $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$.

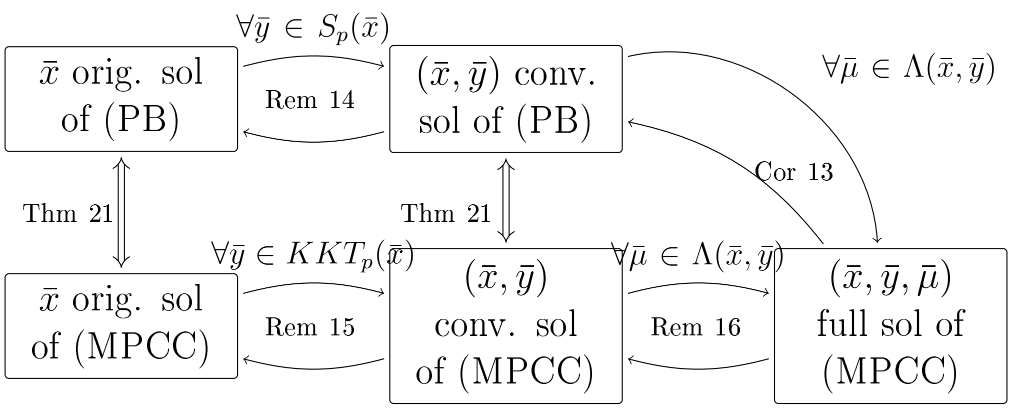


Figure III.2.1: Global solution comparison graph.

To close this section we show in Figure III.2.1 a graph with relations between all the presented notions of global solution for the bilevel problem and for the associated MPCC, under the basic assumptions of convexity of the lower level problem and the (BA) for the lower level problem.

III.2.4 Local Solution Concepts

In this section, we are concerned with the relationship between local solutions to the pessimistic bilevel programming problem with those of the associated MPCC. Since few local concepts of solution of pessimistic for bilevel or MPCC problems are defined in the literature, we propose four different notions for (PB) and three for (MPCC) and establish some interrelations between them.

All along this section, we will make the additional assumption that the gradients $\nabla_y f(x,y)$ and $\nabla_y g(x,y)$ are continuous with respect to the vector (x,y). This ensures (see e.g. [101]) that the set-valued map $(x,y) \mapsto \Lambda(x,y)$ is closed graph.

III.2.4.1 Globally Feasible Local Solutions

Corresponding to the original global solution, defined in the previous section, we recall the following definition of local solution (which was considered in [44]).

Definition 16. *We say that a point $\bar{x}$ is an* original local solution *for* (PB) *if $\bar{x} \in X$ and there exists a neighborhood U of $\bar{x}$ such that*

$$\varphi_p(\bar{x}) \leq \varphi_p(x), \quad \forall x \in X \cap U.$$

Clearly, if $U = \mathbb{R}^n$ then we recover the definition of (global) original solution (see Definition 11). When we regard both variables x and y, as in the conventional global solution, we can define some more types of solution.

Remark 19. *A notion of local solution to problem* (PB) *considered in [40] is the following: a pair $(\bar{x},\bar{y})$ is said to be a* local solution *for* (PB), *if $(\bar{x},\bar{y}) \in \operatorname{gph} S_p$ and there exists a neighborhood U of $\bar{x}$ such that*

$$\varphi_p(\bar{x}) \leq \varphi_p(x), \quad \forall x \in X \cap U.$$

It is interesting to notice that, in this definition, the "locality" is just for the variable x, and nothing is asked for the variable y more than to be in $S_p(\bar{x})$. Clearly, we have the following relation between conventional concept and this new one: for any $\bar{y} \in S_p(\bar{x})$, $\bar{x}$ is an original local solution of (PB), *if and only if, $(\bar{x},\bar{y})$ is a local solution of* (PB). *Given its proximity to the original local solution we do not refer to this definition in the sequel.*

With our general assumptions we have that $S_p(x) \neq \emptyset$ for all $x \in X$, and $F(x,y) = \varphi_p(x)$ whenever $y \in S_p(x)$. Then problem (PB) can be stated equivalently as the following minimization problem

$$\min_{(x,y)\in \operatorname{gph} S_p} F(x,y), \tag{III.2.6}$$

and thus we will refer to the elements of $\operatorname{gph} S_p$ as *feasible* pairs for problem (PB). Taking into account the above consideration we can define a new type of local solution for (PB).

Definition 17. *We say that a pair $(\bar{x},\bar{y})$ is a* conventional type I local solution *to problem* (PB), *if $(\bar{x},\bar{y}) \in \operatorname{gph} S_p$ and there exists a neighborhood V of $(\bar{x},\bar{y})$ such that*

$$F(\bar{x},\bar{y}) \leq F(x,y), \quad \forall (x,y) \in \operatorname{gph} S_p \cap V.$$

One can now consider to define some notions of conventional local solutions to problem (MPCC) that extend the global ones defined in the previous section. But they would lead to totally trivial links between them and the corresponding definitions for the problem (PB) since, under our assumptions, one has $SP(x) = S(x)$, for all $x \in X$ (see Proposition 11 and the tight similarity between the definition of conventional global solution for (MPCC) and the conventional global solution for (PB)).

The same would occur with the definition of original local solutions. For instance, we can define $\bar{x}$ to be an *original local solution* of (MPCC), whenever $\bar{x} \in X$ and there exists a neighborhood U of $\bar{x}$ such that

$$\psi_p(\bar{x}) \leq \psi_p(x), \quad \forall x \in X \cap U.$$

Again since $SP(x) = S(x)$, then $\psi_p(x) = \varphi_p(x)$, for all $x \in X$, and then the definition of original local solution for (MPCC) and (PB) coincide.

Taking these arguments into account we will no longer consider the local concept of original or conventional solutions for (MPCC) and we will thus only concentrate on three notions of, so-called, *full local solutions* of (MPCC) and their interrelations with the conventional concepts of local solutions for (PB). The chosen terminology expresses the fact that, as in the (global) Definition 15, triplets (x, y, μ) will be considered.

Definition 18. *We say that a triplet $(\bar{x}, \bar{y}, \bar{\mu})$ is a* full type I local solution *of* (MPCC) *, if $(\bar{x}, \bar{y}, \bar{\mu}) \in \operatorname{gph} KKT_p$ and there exists a neighborhood W of $(\bar{x}, \bar{y}, \bar{\mu})$ such that*

$$F(\bar{x}, \bar{y}) \leq F(x, y), \quad \forall (x, y, \mu) \in \operatorname{gph} KKT_p \cap W.$$

Let us start by investigating the link between conventional type I local solutions for (PB) and the full type I local solutions for (MPCC).

Theorem 22. *Assume that the lower level problem is convex.*

i) If $(\bar{x}, \bar{y})$ is a conventional type I local solution for (PB)*, then for each $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$, $(\bar{x}, \bar{y}, \bar{\mu})$ is a full type I local solution for* (MPCC)*.*

ii) Conversely, assume that Slater's CQ holds on a neighborhood of $\bar{x}$, $\bar{y} \in SP(\bar{x})$, and for all $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$, $(\bar{x}, \bar{y}, \bar{\mu})$ is a full type I local solution of (MPCC)*. Then, $(\bar{x}, \bar{y})$ is a conventional type I solution of* (PB)*.*

Proof. We follow here the same lines as in the proof of [41, Theorem 3.2]. Let us first observe that whenever Slater's CQ holds for x, then $(x, y) \in \operatorname{gph} S_p$ is equivalent to the fact that there exists μ such that $(x, y, \mu) \in \operatorname{gph} KKT_p$.

For assertion *i*), if $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$ then by the initial comment $(\bar{x}, \bar{y}, \bar{\mu}) \in \operatorname{gph} KKT_p$. If $(\bar{x}, \bar{y}, \bar{\mu})$ is not a full type I local solution for (MPCC), then there exists a sequence $(x_k, y_k, \mu_k)_{k \in \mathbb{N}} \subset \operatorname{gph} KKT_p$ converging to $(\bar{x}, \bar{y}, \bar{\mu})$ such that, for any k, $F(x_k, y_k) < F(\bar{x}, \bar{y})$. A contradiction thus occurs with the fact that $(\bar{x}, \bar{y})$ is a conventional type I local solution for (PB) since the sequence $(x_k, y_k)_{k \in \mathbb{N}} \subset \operatorname{gph} S_p$ converges to $(\bar{x}, \bar{y})$.

Now, let us prove *ii*). Clearly, since $(\bar{x}, \bar{y}, \bar{\mu}) \in \operatorname{gph} KKT_p$ then one has $(\bar{x}, \bar{y}) \in \operatorname{gph} S_p$. If $(\bar{x}, \bar{y})$ is not a conventional type I local solution for (PB), then there exists a sequence $(x_k, y_k)_{k \in \mathbb{N}} \subset \operatorname{gph} S_p$ converging to $(\bar{x}, \bar{y})$ and such that, for any k, $F(x_k, y_k) < F(\bar{x}, \bar{y})$. By the initial comments we know that for each k there exist μ_k such that $(x_k, y_k, \mu_k) \in \operatorname{gph} KKT_p$, and in particular $\mu_k \in \Lambda(x_k, y_k)$. Moreover, Slater's qualification condition guarantees that Λ is locally bounded, and it has closed graph [101, Theorem 2.3], so that one can assume without loss of generality that the sequence $(u_k)_{k \in \mathbb{N}}$ is bounded and converges to some $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$. Then since $(x_k, y_k, \mu_k) \to (\bar{x}, \bar{y}, \bar{\mu})$, the triplet $(\bar{x}, \bar{y}, \bar{\mu})$ cannot be a full type I local solution of (MPCC), raising a contradiction. □

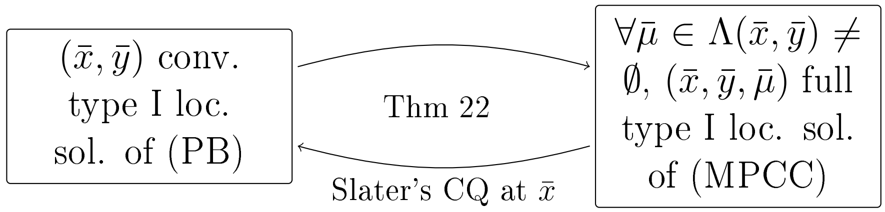


Figure III.2.2: Type I local solutions comparison graph

The interrelations obtained in Theorem 22 are shown in Figure III.2.2.

Let us now show that, even in a very simple case, one cannot drop Slater's CQ in Theorem 22.

Example 16. *Consider a bilevel problem for which the lower level problem is to minimize over $y \in \mathbb{R}$ the function $f(x,y) = xy$ subject to $g(x,y) = x^2(y^2-1) \leq 0$, and the upper level problem is to minimize $F(x,y) = x$, subject to $x \in [-1,1]$. The solution map of the parametric convex lower level program is*

$$S(x) := \begin{cases} \{1\}, & \text{if } x < 0, \\ \{-1\}, & \text{if } x > 0, \\ \mathbb{R}, & \text{if } x = 0. \end{cases}$$

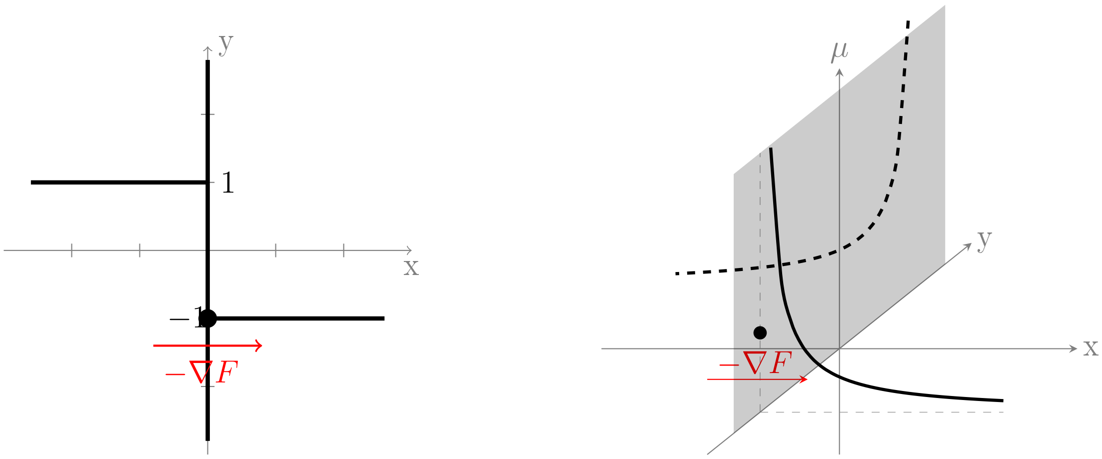
Clearly, for each $x \neq 0$ and y (the associated unique solution of the lower level problem), there is only one Lagrange multiplier, namely $\mu = \frac{-1}{2xy}$, and it diverges when x goes to 0. By the contrary, for $x = 0$ and for any y in $S(0) = \mathbb{R}$, the set of Lagrange multipliers $\Lambda(x,y)$ is the whole half line $\mathbb{R}_+$. Slater's CQ is not fulfilled for $x = 0$, but Abadie's CQ does (see Chapter I Section I.4.2). Since multipliers exist for any optimal reaction, then we see that (BA) *is fulfilled in this example. We can easily see that $(\bar{x},\bar{y}) = (0,-1)$ is not a conventional type I local solution of* (PB)*, while for any $\mu \in \Lambda(0,-1) = \mathbb{R}_+$ the triplet $(0,-1,\mu)$ is a full type I local solution of* (MPCC) *(see Figure III.2.3).*

Remark 20. *On the other hand, Example 3.4 in [41] provides a linear bilevel programming problem with unique lower level solution (thus also working for pessimistic case) for which a point $(\bar{x},\bar{y},\bar{\mu})$ is a full type I local solution of* (MPCC) *at which Slater's CQ is satisfied, but $(\bar{x},\bar{y})$ is not a conventional type I local solution of* (PB)*. In this case, again Theorem 22 cannot be applied since there exists another multiplier $\mu \in \Lambda(\bar{x},\bar{y})\backslash\{\bar{\mu}\}$ such that $(\bar{x},\bar{y},\mu)$ is not a full type I local solution of* (MPCC)*, thus showing that the hypotheses of ii) of Theorem 22 cannot be easily weakened.*

The following example ([18] Example 3.2) shows an optimistic BLP (one leader, one follower) and the associated MPCC, for which points $(\bar{x},\bar{y},\bar{\mu})$ are local solutions of MPCC for all $\bar{\mu} \in \Lambda(\bar{x},\bar{y})\backslash\{\bar{\mu}_0\}$, except one $\bar{\mu}_0 \in \Lambda(\bar{x},\bar{y})$, but the pair $(\bar{x},\bar{y})$ is not a local solution of BLP. This example highlights that it is not enough to check that for some $\bar{\mu} \in \Lambda(\bar{x},\bar{y})$ the triplet $(\bar{x},\bar{y},\bar{\mu})$ is a local solution of the MPCC to guarantee that $(\bar{x},\bar{y})$ is a local solution of the optimistic BLP.

Let the lower level problem be

$$\min_y \; -y$$
$$s.t. \begin{cases} x + y \leq 0 \\ y \leq 0 \end{cases}$$



(a) The point $(0,1)$ is not a conventional type I local solution of (PB), because one can move continuously along $\operatorname{gph} S_p (= \operatorname{gph} S)$ in the direction of $-\nabla F$ decreasing the value of F.

(b) But the points $(0,1,\mu)$ with $\mu \in \mathbb{R}_+$ (in the $(y\mu)$ half plane), are full type I local solution of (MPCC) since one cannot move continuously along $\operatorname{gph} KKT_p (= \operatorname{gph} KKT)$ in the direction of -∇F.

Figure III.2.3: A bilevel problem (PB) and a local solution $(0,1,\mu)$ of the associated (MPCC) such that $(0,-1)$ is not a local solution of (PB) (see Example 16)

The lower level solution map is single-valued and it is given by $S(x) = \{\min\{0,-x\}\}$ *for* $x \in X = [-1,1]$*. Consider as an upper level problem to minimize* $F(x,y) = x$ *subject to* $x \in [-1,1]$ *and* $y \in S(x)$ *(*$\{y\} = S(x)$*).*

Consider now the associated MPCC. First, note that Slater's CQ holds for all $x \in [-1,1]$*. The set of Lagrange multipliers for this problem is given by*

$$\Lambda(x, S(x)) = \begin{cases} \{(\mu_1,\mu_2) = (0,1)\}, & x < 0 \\ \{(\mu_1,\mu_2) = (1,0)\}, & x > 0 \\ \{(\mu_1,\mu_2) \mid \mu_1 + \mu_2 = 1,\ \mu_1,\mu_2 \geq 0\}, & x = 0 \end{cases}$$

Then it can be noticed that $(\bar{x},\bar{y},\mu_1,\mu_2) = (0,0,\mu_1,\mu_2)$*, with* $(\mu_1,\mu_2) \in \Lambda(0,0)$*, is a local solution of MPCC, if and only if* $\mu_1 > 0$*. And it is clear that there is no local solution for BLP except the unique global solution* $(\bar{x},\bar{y}) = (-1,0)$ *of BLP.*

III.2.4.2 Locally Feasible Local Solutions

Even though, in the definitions of local solution given in Subsection III.2.4.1, neighborhoods are considered in terms of both x and y, the maximum that defines $S_p(x)$ is still actually a global maximum. This can appear to be quite artificial, especially if one considers that each of the three optimization problems composing (PB) are solved through local solvers. Note that since the lower level problem is actually convex, there is no need to consider local solution concepts for it.

Let us thus define the set $S_p^{loc}(x)$ of local maxima of

$$\max_{y \in S(x)} F(x,y), \tag{III.2.7}$$

that is, the set of $y \in S(x)$ such that there is a neighborhood N of y with $F(x,y) = \max_{y' \in S(x) \cap N} F(x,y')$. A pair (x,y) is said to be *locally feasible* for (PB), whenever $(x,y) \in$

$\operatorname{gph} S_p^{loc}$. We clearly have $\operatorname{gph} S_p \subset \operatorname{gph} S_p^{loc}$, while the reverse inclusion does not hold in general. In other words, there could be some locally feasible pairs that are not feasible for (PB).

A local feasible pair (x, y) is basically a strategy x along with a locally-worst best-response of the follower. The leader may consider this as a possible outcome since it might be that he is only able to compute a locally worst optimal reaction and not a (real/global) worst optimal reaction of the follower.

Definition 19. *We say that a pair* $(\bar{x}, \bar{y})$ *is a* conventional type II local solution *of* (PB) *if* $\bar{y} \in S_p^{loc}(x)$ *and there exists a neighborhood* U *of* $\bar{x}$ *such that, for each* $x \in X \cap U$*, there exists an* $y_x \in S_p^{loc}(x)$ *with*

$$F(\bar{x}, \bar{y}) \leq F(x, y_x).$$

In other words, the idea of the above definition is to guarantee that, whenever $(\bar{x}, \bar{y})$ is a conventional type II local solution of (PB), then:

a) The follower's strategy $\bar{y}$ is a locally-worst best-response for the leader's strategy $\bar{x}$;

b) For any leader's strategy x sufficiently close to $\bar{x}$, the leader is able to compute at least one locally-worst best-response y_x of the follower such that $F(\bar{x}, \bar{y}) \leq F(x, y_x)$.

This means that the leader is convinced that he has no incentive to change his strategy $\bar{x}$ by another one close to $\bar{x}$. Let us note that using the pessimistic value function one has that, if $(\bar{x}, \bar{y})$ is a conventional type II local solution of (PB), then $(\bar{x}, \bar{y}) \in \operatorname{gph} S_p^{loc}$ and there exists a neighborhood U of $\bar{x}$ such that

$$F(\bar{x}, \bar{y}) \leq \varphi_p(x), \quad \forall\, x \in X \cap U.$$

In general, the definitions of conventional local solutions are independent. In the case when the function $y \mapsto F(x, y)$ is concave for each $x \in X$, then $\operatorname{gph} S_p = \operatorname{gph} S_p^{loc}$ and thus any conventional type II local solution is a conventional type I local solution of (PB). In the next remark, we show that a conventional type I local solution need not to be a conventional type II local solution.

Remark 21. *Consider the bilevel problem described in Example 16. One can observe that the pair* $(\bar{x}, \bar{y}) = (0, 0)$ *is a conventional type I local solution of* (PB)*, while it is not a conventional type II local solution of* (PB).

Finally, based on the definition of S_p^{loc}, one can define a last, and quite natural, definition of local solution of (PB).

Definition 20. *We say that a pair* $(\bar{x}, \bar{y})$ *is a* conventional type III local solution *of* (PB) *if* $(\bar{x}, \bar{y}) \in \operatorname{gph} S_p^{loc}$ *and there exists a neighborhood* V *of* $(\bar{x}, \bar{y})$ *such that*

$$F(\bar{x}, \bar{y}) \leq F(x, y), \quad \forall (x, y) \in \operatorname{gph} S_p^{loc} \cap V.$$

Now, mimicking the definition of the local solution set $S_p^{loc}(x)$, for any $x \in X$, the set of local solutions of the optimization problem

$$\max_{(y,\mu)\in KKT(x)} F(x,y), \tag{III.2.8}$$

is denoted by $KKT_p^{loc}(x)$, that is, the set of $(y,\mu) \in KKT(x)$ such that there is a neighborhood M of (y,μ) with $F(x,y) = \max_{(y',\mu')\in KKT(x)\cap M} F(x,y')$. The elements of $\operatorname{gph} KKT_p^{loc}$ will further be called *locally feasible triplet* of (MPCC).

Now, based on this notation, we will give two new definitions of local solution for (MPCC), each of them being analogous to Definition 19 and Definition 20, respectively.

Definition 21. *We say that a triplet $(\bar{x},\bar{y},\bar{\mu})$ is a* full type II local solution *of* (MPCC) *if $(\bar{y},\bar{\mu}) \in KKT_p^{loc}(\bar{x})$ and there exists a neighborhood U of $\bar{x}$ such that for each $x \in X\cap U$ there exists an $(y_x,\mu_x) \in KKT_p^{loc}(x)$ satisfying*

$$F(\bar{x},\bar{y}) \leq F(x,y_x).$$

Definition 22. *We say that a triplet $(\bar{x},\bar{y},\bar{\mu})$ is a* full type III local solution *of* (MPCC) *if $(\bar{x},\bar{y},\bar{\mu}) \in \operatorname{gph} KKT_p^{loc}$ and there exists a neighborhood W of $(\bar{x},\bar{y},\bar{\mu})$ such that*

$$F(\bar{x},\bar{y}) \leq F(x,y), \quad \forall (x,y,\mu) \in \operatorname{gph} KKT_p^{loc} \cap W.$$

Finally, in order to explore the interrelation between the locally feasible local solution concepts, let us first state the following proposition.

Proposition 12. *Assume that the lower level problem* (III.2.2) *is convex. If $(x,y) \in \operatorname{gph} S_p^{loc}$, then for each $\mu \in \Lambda(x,y)$, $(x,y,\mu) \in \operatorname{gph} KKT_p^{loc}$. Conversely, if $(x,y,\mu) \in \operatorname{gph} KKT_p^{loc}$, then $(x,y) \in \operatorname{gph} S_p^{loc}$.*

Proof. Take $(x,y) \in \operatorname{gph} S_p^{loc}$, that is, $y \in S(x)$ such that there is a neighborhood N of y with

$$F(x,y) = \max_{y'\in S(x)\cap N} F(x,y').$$

Take also $\mu \in \Lambda(x,y)$ so that $(y,\mu) \in KKT(x)$. Recall again that Proposition 11 gives that $S(x) = SP(x)$. Now, we can consider the neighborhood $M := N \times \mathbb{R}^d$ of (y,μ). Then

$$F(x,y) = \max_{(y',\mu')\in KKT(x)\cap M} F(x,y').$$

and thus the triplet (x,y,μ) is an element of $\operatorname{gph} KKT_p^{loc}$.

For the converse, assume that (x,y,μ) is an element of $\operatorname{gph} KKT_p^{loc}$. If y is not in $S_p^{loc}(x)$ this means that there exists a sequence $(y_k)_{k\in\mathbb{N}}$ converging to y such that, for any k, $y_k \in S(x)$ and $F(x,y_k) > F(x,y)$. Since y and y_k are in $S(x) = SP(x)$ (by Proposition 11), then they have the same set of Lagrange multipliers (see [57, Proposition 3.1.1, VII]), and thus $\mu \in \Lambda(x,y) = \Lambda(x,y_k)$, for each k. The sequence $(y_k,\mu)_{k\in\mathbb{N}} \subset KKT(x)$ converges to (y,μ) and $F(x,y_k) > F(x,y)$, so that (x,y,μ) is not an element of $\operatorname{gph} KKT_p^{loc}$, thus providing a contradiction. □

As a consequence of Proposition 12, we have the following interrelations between conventional (type II/type III) and the full (type II/type III) local solution of the associated (MPCC).

Theorem 23. *Assume the convexity condition of the lower level problem and that* (BA) *holds. Then*

i) If $(\bar{x}, \bar{y})$ *is a conventional type II local solution of* (PB)*, then* $\Lambda(\bar{x}, \bar{y})$ *is non empty and for all* $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$ *the triplet* $(\bar{x}, \bar{y}, \bar{\mu})$ *is a full type II local solution of* (MPCC).

Conversely, if there exists at least one $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$ *such that* $(\bar{x}, \bar{y}, \bar{\mu})$ *is a full type II local solution of* (MPCC)*, then* $(\bar{x}, \bar{y})$ *is a type II local solution of* (PB).

ii) If $(\bar{x}, \bar{y})$ *is a conventional type III local solution for* (PB)*, then for each* $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$*,* $(\bar{x}, \bar{y}, \bar{\mu})$ *is a full type III local solution of* (MPCC).

Conversely, assume that Slater's CQ holds on a neighborhood of $\bar{x}$*, that* $\bar{y} \in SP(\bar{x})$*, and for all* $\mu \in \Lambda(\bar{x}, \bar{y})$*,* $(\bar{x}, \bar{y}, \mu)$ *is a full type III local solution of* (MPCC)*. Then,* $(\bar{x}, \bar{y})$ *is a conventional type III local solution of* (PB).

Proof. Let's prove assertion i). Assume $(\bar{x}, \bar{y})$ is a conventional type II local solution of (PB) with U given as in Definition 19. In particular, $(\bar{x}, \bar{y})$ is locally feasible for (PB). From Proposition 11 and BA, we see that $\Lambda(\bar{x}, \bar{y}) \neq \emptyset$. If we take $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$, by Proposition 12, we deduce that $(\bar{x}, \bar{y}, \bar{\mu})$ is locally feasible for (MPCC). If we take another (x, y, μ) that is locally feasible for (MPCC) with $x \in X \cap U$, then (again by Proposition 12) (x, y) is locally feasible for (PB), so that $F(\bar{x}, \bar{y}) \leq F(x, y)$. Thus, we conclude that $(\bar{x}, \bar{y}\bar{\mu})$ is a full type II local solution of (MPCC). The converse is similar.

For assertion ii), the proof follows a similar line as the one for Theorem 22, and is thus left to the reader. □

Figures III.2.4 and III.2.5 summarize the interrelations obtained in Theorem 23.

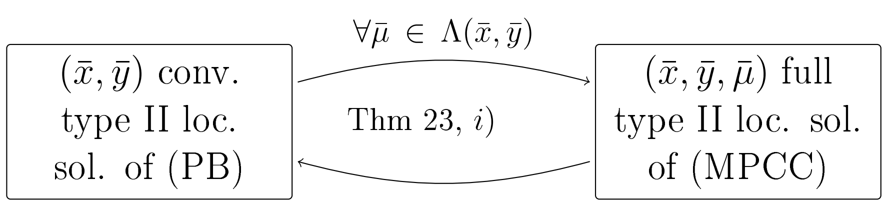


Figure III.2.4: Type II local solutions comparison graph

III.2.5 Conclusions

Under some basic conditions that ensures that the lower level of the bilevel problem is equivalent to its parametrized KKT conditions, the global solutions of the pessimistic bilevel problem have been compared with those of the associated MPCC, giving place to very simple relationships. These relationships are represented in the graph of Figure 1.

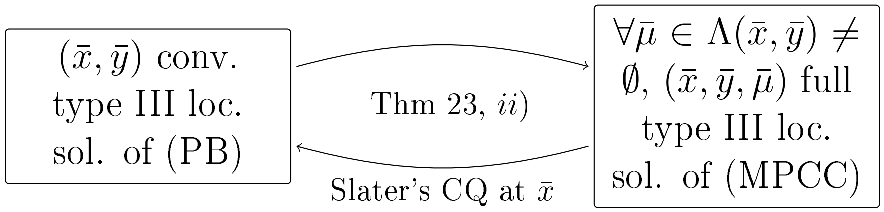


Figure III.2.5: Type III local solutions comparison graph

As shown by the very simple bilevel problem developed as Example 16, the interrelations are less simple when dealing with local concepts of solutions. Thus, we have first defined adapted notions of local solutions, both for the pessimistic bilevel problem (PB) and the associated (MPCC). Looking to Theorem 22 and Theorem 23 *ii*) (see figures III.2.2 and III.2.5), some quite strong assumptions must be assumed to ensure that a local solution of (MPCC) is a local solution of the bilevel problem (PB); the only interrelation between local concepts that can be established under mild hypotheses is the one described in Theorem 23 *i*) (see Figure III.2.4) between conventional type II local solutions of (PB) and full type II local solutions of (MPCC).

Thus, and as a final conclusion, the general answer to the question posed in the title is negative: pessimistic bilevel problems cannot be considered as special cases of pessimistic mathematical programs with complementarity constraints. Even if the pessimistic formulation is clearly more difficult to handle than the optimistic formulation, the conclusion is by the way similar to the one obtained in [41] for optimistic bilevel problems.

Chapter III.3

Multi-Leader-Follower Games

This chapter is based on the paper [17] and the chapter [20].

Multi-Leader-Follower games are bilevel models mixing the (Nash-) equilibrium structure of usual non-cooperative game theory within each level, and a hierarchical feature between the two levels, see Chapters III.1 and III.2.

III.3.1 Notations and Examples of Applications

For a game with N leaders and M followers, with their respective variables $x_1, \dots x_N$ and $y_1, \dots y_M$, the MLMFG can be expressed as:

$$\begin{array}{ccc} \begin{array}{ll} \text{“min”}_{x_1} & F_1(x,y) \\ \text{s.t.} & \left\{ \begin{array}{l} x_1 \in X_1(x_{-1}) \\ y \in GNEP(x) \end{array} \right. \end{array} & \dots & \begin{array}{ll} \text{“min”}_{x_n} & F_N(x,y) \\ \text{s.t.} & \left\{ \begin{array}{l} x_N \in X_N(x_{-N}) \\ y \in GNEP(x) \end{array} \right. \end{array} \end{array}$$

$$\downarrow\uparrow \qquad \qquad \downarrow\uparrow$$

$$\begin{array}{ccc} \begin{array}{ll} \min_{y_1} & f_1(y_1, x, y_{-1}) \\ \text{s.t.} & y_1 \in Y_1(x, y_{-1}) \end{array} & \dots & \begin{array}{ll} \min_{y_M} & f_M(y_M, x, y_{-M}) \\ \text{s.t.} & y_M \in Y_M(x, y_{-M}) \end{array} \end{array}$$

Figure III.3.1: MLMFG

But let us first concentrate on the important case where there is only one leader, that is, Single-Leader-Multi-Follower games SLMFG. This can be represented by the following diagram:

$$\begin{array}{|ll|}\hline \text{“min”}_{x} & F(x,y) \\ \text{s.t.} & \begin{cases} x \in X \\ y \in GNEP(x) \end{cases} \\ \hline \end{array}$$

$$\downarrow\uparrow$$

$$\begin{array}{|ccc|}\hline \begin{array}{|ll|}\hline \min_{y_1} & f_1(y_1, x, y_{-1}) \\ \text{s.t.} & y_1 \in Y_1(x, y_{-1}) \\ \hline \end{array} & \dots & \begin{array}{|ll|}\hline \min_{y_M} & f_M(y_M, x, y_{-M}) \\ \text{s.t.} & y_M \in Y_M(x, y_{-M}) \\ \hline \end{array} \\ \hline \end{array}$$

Figure III.3.2: SLFMG

where for any $j = 1, \dots, M$, the set-valued map $Y_j(x, y_{-j})$ expresses the constraints, parametrized by x and y_{-j} that the decision variable of player j must satisfy and where $GNEP(x)$ stands for the set of (generalized) Nash equilibria of the non-cooperative game between the followers.

The notation "min" is used here to highlight that this is simply a first rough definition that is not free of ambiguities. In particular, if the reaction of the followers is not uniquely determined, the leader cannot anticipate which (GNEP)-reaction will take place and thus the upper level problem becomes ill-posed/ambiguous. This kind of question will be addressed in Section III.3.2.

One application for which the model SLMFG has proved its efficiency is the optimal design of industrial eco-parks. This new way to design industrial parks is based on the sharing of fluids (water, vapor, etc) or of energy between companies in order to reduce their production costs and, at the same time, the ecological impact of the production of the participating companies (measured by the total amount of wastes and/or of the incoming raw materials: water, energy, vapor, etc.). This problem, already considered in the 60's, has been treated in the literature using the *multi-objective optimization approach*.

However, this technique has shown its limits in particular because it requires a selection process between the obtained Pareto points which almost always is based on a prioritization scheme between the companies (through weighted sum, goal programming, etc.). Recently in [98], a Multi-Leader-Follower approach has been proposed with success. In such a model, the followers are the companies, interacting in a non-cooperative way (GNEP), each of them aiming at minimizing their production cost. The unique leader is the designer/manager of the industrial park whose target is to minimize the ecological impact (total water consumption, waste volumes, etc). The designer will also ensure the clearance of the process. Thus for example in the case of the design of the water network developed in [98] the variable of the designer is the vector x of flows of clear water coming to each process of each company while the variable of each company j is the vector y_j of shared flows between the processes of the company j and the processes of the concurrent companies. The resulting SLMFG model is as follows

$$\begin{array}{ll} \min_{x,y} & \sum_i x_i \\ s.t. & \begin{cases} x_i \geq 0, \quad \forall i \\ \forall j,\ y_j \text{ solution of:} \quad \min_{z_j} \quad cost_j(z_j, x, y_{-j}) \\ \qquad s.t. \qquad \begin{cases} \text{water balance equation} \\ \text{contamination bounds} \\ \text{mass balance of contaminants} \\ \text{other technical assumptions} \end{cases} \end{cases} \end{array}$$

Thanks to the use of this approach, an important reduction of the global water consumption has been obtained while ensuring the reduction of the production cost of all of the participating companies. Other recent developments of this approach can be found in [30, 99, 105]. Note that even if historically industrial eco-parks have been focusing on water exchanges, several other things can also be shared between the companies, see e.g. [100] for an optimal design of the energy and water exchange in an eco-park.

Symmetrically to the SLMFG, whenever the set of followers is reduced to only one player, then the "bilevel model" leads to the so-called Multi-Leader-Single-Follower game:

$$\boxed{\begin{array}{ll} \text{“}\min\text{”}_{x_1, y} & F_1(x,y) \\ \text{s.t.} & \left\{\begin{array}{l} x_1 \in X_1(x_{-1}) \\ y \in S(x) \end{array}\right. \end{array}} \quad \cdots \quad \boxed{\begin{array}{ll} \text{“}\min\text{”}_{x_N, y} & F_N(x,y) \\ \text{s.t.} & \left\{\begin{array}{l} x_N \in X_N(x_{-N}) \\ y \in S(x) \end{array}\right. \end{array}}$$

$$\downarrow\uparrow$$

$$\boxed{\begin{array}{ll} \min_y & f(y,x) \\ \text{s.t.} & y \in Y(x) \end{array}}$$

Figure III.3.3: MLSFG

where $S(x)$ denotes the set of global optima of the (unique) follower's problem, which depends on the decision variable of the leaders.

Those difficult models cover a very large class of application in different real-life fields and in particular in the management of energy. For example the MLMFG provides a perfect model for the description of so-called *day-ahead electricity markets*. The leaders are the members of the market (suppliers and/or retailers) whose decision variables x_i are market offers (usually energy/price blocks or affine bid curves) while the unique and common follower is the regulator of the market (often called Independent System Operator - ISO) who, reacting to these offers, fixes the price of electricity and the decision concerning the offers of the leaders (see [1, 4, 5, 8, 9, 47, 48, 56, 62] and references therein).

The main tasks of the regulator is to ensure clearance of the decision process and to maximize the total welfare of the market or equivalently minimizing the total cost of production if the total demand is assumed to be fixed (assumption of no elasticity on the market). The regulator/follower variable is the vector y of decisions (acceptances/rejections) of the bids of the producers. As a by product of the resolution of the follower problem, the Lagrange multiplier associated to the balance constraint will be the unit marginal price of electricity on the market. The corresponding MLMFG, in a simplified form, is thus as follows:

$$\text{For any i,} \quad \min_{x_i,y} \quad profit(x_i, y, x_{-i})$$
$$s.t. \left\{\begin{array}{l} x_i \text{ admissible bid} \\ y \text{ solution of: } \min_z \quad Total\_welfare(z,x) \\ \qquad s.t. \left\{\begin{array}{l} \forall k, z_k \text{ decision concerning} \\ \quad \text{bids of producer } k \\ \text{demand/offer balance} \end{array}\right. \end{array}\right.$$

It can be clearly noticed here that if the regulator/follower's problem admits possibly more than one solution for a given leader strategy x, then the overall MLMFG problem is ill-posed, carrying some ambiguity; see beginning of Section III.3.3. In electricity market modeling, the uniqueness of the solution of the regulator/follower's problem is guaranteed by some strict convexity of the "total_welfare" function with regard to variable z thanks to specific assumptions on the bid structure (strictly convex quadratic bid curves - see e.g. [1,47,56,62]) or some *equity property* on the decision process (see [4,5]).

III.3.2 Single-Leader-Multi-Follower Games

In this section we consider the case where there is a single leader and multiple followers, which we refer to as a SLMFG and we use the notations of the corresponding diagram of Section III.3.1.

As observed in [29], if none of the constraint maps Y_j nor the objectives f_j depend on the decision variable of the other followers then the SLMFG admits an equivalent reformulation as a classical bilevel problem with only one follower. This can be seen by defining a (unique) follower's variable as $y := (y_1, \ldots, y_M)$, the objective $f(x, y) := \sum_{j=1}^{M} f_j(y_j, x)$ and the aggregated constraint map $Y : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ defined by $Y(x) := \{y \mid y_j \in Y_j(x), \forall j\}$. Thus, under this particular structure, several analyzes on the single-leader-single-follower case (see Chapter III.2) can be directly extended to multiple followers, while in general having multiple followers does bring new difficulties.

Now if for any decision x of the leader, there exists (implicitly or explicitly) a unique equilibrium $y(x) = (y_1(x), \ldots, y_M(x))$ among followers then, the SLMFG can be treated as a classical mathematical programming problem

$$\min_x F(x, y(x)), \quad \text{with } x \in X$$

where of course some good properties (semi-continuity, differentiability, convexity, etc) of the response function $y(x)$ must be satisfied for the reformulation to be useful. But in the general case the formulation of SLMFG carries some ambiguities.

The ambiguity coming from the possible non-uniqueness of the lower level equilibrium problem, which is already present in the case of one leader and one follower, is in our setting of several followers an even more inevitable situation. Indeed, since the lower level is an equilibrium problem (GNEP), the uniqueness of an equilibrium can rarely be ensured, and it cannot be avoided simply by assuming strict convexity, see for instance the examples in Chapter III.1. Despite this argument for general problems, there are some cases where the lower level problem might have unique responses as in [110] and others.

The most common approach to tackle this ambiguity is the *optimistic* approach, which consists in considering the best equilibrium reaction of the followers with regard to the leader's objective. It can be argued as a kind of cooperation of the followers with the leader. In fact, it is often the case in applications that the leader is assumed to take his decision before the followers, and thus he can after having computed his optimal decision suggest the followers to take certain equilibrium reaction that is convenient to him. Each of the followers will then have no incentive to unilateral deviate from the proposed equilibrium strategy, because of the nature of equilibria.

Definition 23. *We say that* $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ *is an* optimistic equilibrium *of the SLMFG if* *it is a solution of the following optimization problem*

$$\min_{x,y} \quad F(x,y)$$
$$\begin{cases} x \in X, \\ y \in \mathrm{GNEP}(x) \end{cases}$$

An opposite approach is the *pessimistic* one, which consists for the leader in minimizing the worst possible equilibrium reaction with regard to the leader objective. Thus, it is based on a minmax problem.

Definition 24. *We say that* $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ *is an* pessimistic equilibrium *of the SLMFG if it is a solution of the following minmax problem*

$$\min_{x} \max_{y} \quad F(x,y)$$
$$\begin{cases} x \in X, \\ y \in \mathrm{GNEP}(x) \end{cases}$$

Apart from these two approaches, there are other possibilities based on selections of the lower level problem (see e.g. [47]) and on set-valued optimization but we do not discuss them here. Note also that an alternative approach has been developed in [116] in a specific context.

III.3.2.1 Existence of Optimistic Equilibria in SLMFG

Here we discuss conditions under which a SLMFG admits at least one equilibrium. We present a positive result for the case of optimistic equilibrium. Nevertheless, for the pessimistic case it has been shown an example of an apparently very well behaved problem (linear and compact) which admits no equilibria (see [17, Example 3.2]).

The mathematical tools that are often used in this analysis when the lower level reaction is not unique (which is mostly the case in our setting, see previous paragraphs) are part the so-called set-valued maps theory (see Chapter I.6).

We present now a slight refinement of [17, Theorem 3.1] (see also [76, Corollary 4.4] for analysis in the case of strategy sets that are subsets of reflexive Banach spaces). It is based on continuity properties of both functions and set-valued maps defining the SLMFG. In particular, it assumes the lower semi-continuity of the set-valued maps that defines the feasible set of the followers, that is, for $j = 1, ..., M$ the set-valued map $Y_j : \mathbb{R}^n \times \mathbb{R}^{m-j} \rightrightarrows \mathbb{R}^{m_j}$.

Theorem 24. *Assume for the SLMFG that*

(1) F *is lower semi-continuous and* X *is closed,*

(2) for each $j = 1, ..., M$*,* f_j *is continuous,*

(3) for each $j = 1, ..., M$*,* Y_j *is lower semi-continuous relative to its non-empty domain and has closed graph, and*

(4) either F *is coercive or,* X *is compact and at least for one* j*, the images of* Y_j *are uniformly bounded.*

If the graph of the lower level GNEP *is non-empty, then the SLMFG admits an optimistic equilibrium.*

Proof. As in the proof of [17, Theorem 3.1], assumptions (2) and (3) ensure the closedness of the constraints of the leader's problem in both variables x and y, see Proposition 5. Thus, the classical Weierstrass theorem can be applied to prove the existence of a minimum of the leaders' optimization problem, which constitutes an optimistic equilibrium of the SLMFG. □

Usually the constraints of the followers are described as level sets of certain functions: $Y_j(x, y_{-j}) := \{y_j \in \mathbb{R}^{m_j} : g_j(x, y) \leq 0\}$, with $g_j : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{d_j}$. We recall what is the MFCQ for the parametric optimization problems of the followers (see, e.g. the appendix in [21]).

Definition 25. *The MFCQ for the followers' problems is satisfied at* $(\bar{x}, \bar{y})$ *if for each* j *the family of gradients there exist* $a_j \in \mathbb{R}^{m_j}$ *such that*

$$\nabla_{y_j} g_{jk}(\bar{x}, \bar{y}) a_j < 0 \quad \forall k \in A_j(\bar{x}, \bar{y}).$$

where $A_j(\bar{x}, \bar{y}) := \{k = 1, ., , , d_j : g_{jk}(\bar{x}, \bar{y}) = 0\}$.

We provide conditions on these data functions that ensure the existence of optimistic equilibrium of the SLMFG. As a particular case we recover Theorem 5.2 in [40]. The forthcoming corollary is just a consequence of our previous result, since the conditions on the data functions imply the continuity properties of the constraint set-valued maps (see [40, Theorem 4.3]).

Corollary 14. *Let us assume that*

(1) F *is lower semi-continuous and* X *is closed,*

(2) for each $j = 1, ..., M$, f_j *is continuous,*

(3) for each $j = 1, ..., M$, $Dom\, Y_j$ *is non-empty,* g_j *is continuous on* $\mathbb{R}^n \times \mathbb{R}^m$ *and satisfy MFCQ at each feasible point, and*

(4) either F *is coercive or,* X *is compact and at least for one* j*, the images of* Y_j *are uniformly bounded.*

If the graph of the lower level GNEP *is non-empty, then the SLMFG admits an optimistic equilibrium.*

Remark 22. *Condition (4) in Theorem 24 and Corollary 14 is assumed to obtain the compactness of the graph of the set-valued map* GNEP *which assigns to a leader strategy* x *the set of solutions of the lower level* GNEP(x).

III.3.2.2 Example of SLMFG with no Pessimistic Equilibrium

It is known that in the case of one leader and one follower with fully linear objectives and constraint, the pessimistic bilevel programming problems admits a solution under mild assumption: compactness and feasibility (see [42, 79]). But when there are two followers, the existence cannot be guaranteed as shown in the following example.

Example 17. *Given $x \in \mathbb{R}$, let the problem of two followers be given by*

$$\min_{y_1} \; y_1 \quad s.t. \begin{cases} y_1 \geq 0 \\ 2y_2 - y_1 \leq 2 \\ y_1 + y_2 \geq x \end{cases}$$

$$\min_{y_2} \; y_2 \quad s.t. \begin{cases} y_2 \geq 0 \\ 2y_1 - y_2 \leq 2 \\ y_1 + y_2 \geq x \end{cases}$$

Follower 1's reaction is $S_1(x, y_2) = \{\max(0, 2(y_2 - 1))\}$ if $x \geq \max(y_2, 3y_2 - 2)$, and empty otherwise, while follower 2's reaction is $S_2(x, y_1) = \{\max(0, 2(y_1 - 1))\}$ if $x \geq \max(y_1, 3y_1 - 2)$, and empty otherwise. Thus, the solution of the followers' parametric GNEP is given by

$$\text{GNEP}(x) = \begin{cases} \{(0,0), (2,2)\} & \text{if } x \geq 4, \\ \{(0,0)\} & \text{if } x \in [0,4[\\ \emptyset & \text{otherwise.} \end{cases} \tag{III.3.1}$$

Now, consider the objective of one leader given by $F(x, y) := -x + y_1 + y_2$, and the constraint $x \in [0, 4]$. The pessimistic problem of the leader is of the form

$$\min_{x \in [0,4]} \max_{y \in \text{GNEP}(x)} -x + (y_1 + y_2).$$

Notice that the function

$$\varphi_p(x) := \max_{y \in \text{GNEP}(x)} -x + (y_1 + y_2) = \begin{cases} 0 & \text{if } x = 4 \\ -x & \text{if } x \in [0,4[\end{cases}$$

is not lower semi-continuous, so that Weierstrass theorem argument cannot be applied. And in fact, the value of the problem of the leader is -4, while there does not exist a point $x \in [0, 4]$ with that value. The pessimistic linear single-leader-two-follower problem has no optimal solution.

III.3.2.3 Reformulations

Reformulating a SLMFG is a way of considering it within a framework where a well-developed theory exists for either finding an equilibrium or better understanding the properties of the problem.

We will restrict our discussion here to reformulations of the optimistic approach of the SLMFG, though for the pessimistic approach corresponding reformulations can also be considered.

Two reformulations of the SLMFG are the most classical and can be considered as particular cases of Mathematical Programs with Equilibrium Constraints (MPECs) in certain references. In both cases the reformulation is based on the replacement of the lower level (generalized) Nash equilibrium problem by a related problem.

A first possibility is to replace the lower level problem by the (quasi-)Variational Inequality problem (VI) associated to the gradients of the objectives (assuming the objective functions are differentiable), or the normal operator (see [3,12,14] for quasi-convex objective functions), while keeping the constraints. The resulting problem is the so-called Optimization Problem with Variational Inequality Constraints (OPVIC).

OPQVIC reformulation

An OPQVIC (see for instance [76,117,118]) is a problem of the form

$$\min_{x,y} \quad F(x,y)$$
$$\begin{cases} x \in X, \\ y \in \mathrm{QVI}(T(x,\cdot), K(x,\cdot)) \end{cases}$$

where $\mathrm{QVI}(T(x,\cdot), K(x,\cdot))$ stands for the solutions set of the following parametric Quasi Variational Inequality problem: find $y \in K(x,y)$ such that

$$\langle T(x,y), y - z\rangle \geq 0, \quad \forall z \in K(x,y). \tag{III.3.2}$$

The OPQVIC reformulation of an optimistic SLMFG consists in considering a parametric QVI defined by $T(x,y) := (\nabla_{y_j} f_j(x,y))_{j=1}^M$ and $K(x,y) := \prod_{j=1}^M Y_j(x, y_{-j})$.

In the case where the lower level is a parametric (non-generalized) Nash equilibrium problem, the resulting reformulation reduces to an OPVIC (Variational Inequality Constraints). This specific problem has received much more attention since it is a more tractable case. See for instance [92,121].

It is easy to see that the OPQVIC reformulation is equivalent to the (optimistic) SLMFG whenever the problems of the followers satisfy the following parametric convexity assumption:

for each follower j the objective $f_j(x,\cdot,y_{-j})$ is pseudo-convex with respect to y_j and the constraint sets $Y_j(x,y_{-j})$ are convex.

Thus, under these convexity assumptions the existence of optimistic equilibria for the SLMFG could be deduced also from [55]. Note that Theorem 24 does not require any such convexity assumption.

In some cases it is possible to write the OPVIC as a nonlinear program (see [123]), but some usual constraint qualifications like the MFCQ, are in general not satisfied for that nonlinear program. Therefore, some well adapted constraint qualification for this class of optimization problems have been developed in the literature [120].

MPCC reformulation

Another classical technique consists in replacing the lower level GNEP by the concatenation of the associated parametric KKT conditions of each of the followers and obtaining a so-called Mathematical Program with Complementarity Constraints (MPCC).

In fact, to each follower's problem we can associate its KKT optimality conditions, that is (y, μ_j) satisfying

$$\begin{cases} \nabla_{y_j} f_j(x,y) + \sum_{k=1}^{d_j} \mu_{jk} \nabla_{y_j} g_{jk}(x,y) = 0 \\ 0 \leq \mu_j \perp -g_j(x,y) \geq 0 \end{cases}$$

We denote by $cKKT(x)$ the set of solutions of the concatenation of KKT conditions of all the followers, that is, (y, μ) such that, for each $j = 1, ..., M$, (y_j, μ_j) solves the KKT system given the parameters (x, y_{-j}).

Thus the MPCC reformulation of the SLMFG consists of the following optimization problem

$$\min_{x,y,\mu} \quad F(x,y)$$
$$\begin{cases} x \in X, \\ \quad (y,\mu) \in cKKT(x) \end{cases}$$

Numerical methods for such a reformulation can be found for instance in [53]. See also [51, 80].

An important difference between the OPVIC and the MPCC reformulations is that in the latter a new variable, the Lagrange multipliers μ, appears as part of the definition of the optimization problem. Moreover, to consider the MPCC reformulation it is important to analyze constraint qualifications of the lower level problem for the existence of Lagrange multipliers and their well-behavior.

To be more precise, in order to have a notion of equivalence between the global solutions of the initial SLMFG and those of its MPCC reformulation an (in general) infinite number of constraint qualifications have to be verified. This fact was first noticed in [46].

We make the following basic hypotheses:

(H_1) (Follower's differentiability) For any $j \in J$ and any $(x, y_{-j}) \in X \times \mathbb{R}^{m_{-j}}$, $f_j(x, \cdot, y_{-j})$ and $g_j(x, \cdot, y_{-j})$ are differentiable;

(H_2) (Follower's player convexity) For any $j \in J$ and any $(x, y_{-j}) \in X \times \mathbb{R}^{m_{-j}}$, $f_j(x, \cdot, y_{-j})$ is convex and the components of $g_j(x, \cdot, y_{-j})$ are quasi-convex functions.

Since the lower level equilibrium problem is player convex the concatenated KKT optimality conditions are sufficient. If we somehow knew that the KKT conditions were also necessary, then it is quite simple to deduce that global solutions of SLMFG yields solutions of the MPCC reformulation, and vice versa.

Theorem 25. *Assume (H_1) and (H_2). The relation between solutions of the SLMFG and its MPCC reformulation are as follows.*

1. *If $(\bar{x}, \bar{y}) \in$ SLMFG and $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$, then $(\bar{x}, \bar{y}, \bar{\mu}) \in$ (MPCC).*

2. *Assume that for each $x \in X$, for each $j \in J$, and for each joint strategy $y = (y_j, y_{-j})$ which is feasible for all followers the Guignard's CQ holds for the constraint "$g_j(x, \cdot, y_{-j}) \leq 0$" at the point y_j. If $(\bar{x}, \bar{y}, \bar{\mu}) \in$ (MPCC), then $(\bar{x}, \bar{y}) \in$ SLMFG.*

Remark 23. *Let us observe that the assumptions of Theorem 25 are not really tractable i.e. quite hard to verify. Indeed, the Guignard's constraint qualification should hold true for each joint strategy being feasible for all follower, that is such that $g_l(x, y) \leq 0$, for every follower $l \in J$. On the other hand these assumptions (of Theorem 25) are in some sense minimal. Indeed, the weakest condition that makes SLMFG equivalent to its MPCC reformulation independently of the objective of the leader is that*

$$\text{GNEP}(x) = cKKT(x), \quad \forall x \in X. \tag{III.3.3}$$

Moreover, for a given $x \in X$, the weakest condition that makes $\text{GNEP}(x) = cKKT(x)$, independently of the objectives of the followers, is in fact the huge set of Guignard's CQs described in the assumptions of Theorem 25.

However, using the techniques developed in [19], we can reduce significantly the conditions to be verified in order to have the desired equivalence, as we explain now.

Assume that the followers' constraint functions g_{jk} are jointly convex with respect to the vector (x, y).

Definition 26. *Let $j \in J$. An opponent strategy $(\hat{x}, \hat{y}_{-j}) \in \mathbb{R}^n \times \mathbb{R}^{m_j}$ is said to be*

- *an* admissible opponent strategy *(for player j) if $(\hat{x}, \hat{y}_{-j}) \in \mathcal{A}_j := \operatorname{dom} Y_j$, that is, such that there exists $y_j \in X$ with $g_j(\hat{x}, y_j, \hat{y}_{-j}) \leq 0$;*
- *an* interior opponent strategy *if it is in $int(\mathcal{A}_j)$;*
- *a* boundary opponent strategy *if it is in $bd(\mathcal{A}_j)$.*

Theorem 26. *Assume (H_1), (H_2) and that for each $j \in J$, the three following properties hold:*

(1) (Joint Convexity) Each g_{jk} is jointly convex with respect to (x, y);

(2) (Joint Slater's CQ) There exists a joint strategy $(\tilde{x}(j), \tilde{y}(j))$ such that $g_j(\tilde{x}(j), \tilde{y}(j)) < 0$;

(3) (Guignard's CQs for boundary opponent strategies) For any boundary opponent strategy $(\hat{x}, \hat{y}_{-j}) \in bd(\mathcal{A}_j)$ Guignard's CQ is satisfied at any feasible point $y_j \in Y_j(\hat{x}, \hat{y}_{-j})$.

If $(\bar{x}, \bar{y}, \bar{\mu}) \in$ (MPCC), then $(\bar{x}, \bar{y}) \in$ SLMFG.

Proof. Let $j \in J$ and $x \in X$, and take a $y = (y_j, y_{-j})$ that is feasible for all followers, so that (x, y_{-j}) is an admissible parameter. Let us now verify that Guignard's CQ holds for the constraint "$g_j(x, \cdot, y_{-j}) \leq 0$" at the point y_j. If (x, y_{-j}) is a boundary opponent strategy we know from Assumption (26) that Guignard's CQ is satisfied at y_j. Otherwise, (x, y_{-j}) is an interior opponent strategy. Then by Proposition 2.1 in [19], Slater's CQ holds for this parameter, which itself imply Guignard's CQ at y_j. Thus the conclusion follows by applying Theorem 25. □

Definition 27. *We say that the lower level of a SLMFG is* fully feasible *if for any follower $j \in J$, $\mathcal{A}_j = \mathbb{R}^n \times \mathbb{R}^{m_{-j}}$, that is, for any opponent strategy $(x, y_{-j}) \in \mathbb{R}^n \times \mathbb{R}^{m_{-j}}$, there exists $y_j \in Y_j(x, y_{-j})$.*

The above definition does not allow boundary opponent strategies to exist. Thus, the third assumption of Theorem 26 is trivially satisfied, leading to the following corollary.

Corollary 15. *Assume (H_1), (H_2) and that the lower level is fully feasible (in the sense of Definition 27). For each $j \in J$ we make the following assumptions:*

1. *(Joint Convexity) Each g_j is jointly convex with respect to (x, y);*

2. *(Joint Slater's CQ) There exists a joint strategy $(\tilde{x}(j), \tilde{y}(j))$ such that $g_j(\tilde{x}(j), \tilde{y}(j)) < 0$.*

If $(\bar{x}, \bar{y}, \bar{\mu}) \in$ (MPCC), then $(\bar{x}, \bar{y}) \in$ SLMFG.

III.3.2.4 Algorithms

There exist actually very few algorithms tackling directly the SLMFG model. In the seminal paper [110] where the case of an oligopoly was studied, a first simple algorithm was proposed. The idea of the algorithm was first to divide the interval of strategies of the leader into finite subinterval, in each of them a linearization of the lower level reaction function is considered, and to minimize the leader's objective composed with the linearization of the lower level problem in the subinterval. The new points are added to the grid. When a termination criterion is satisfied, the best point of the grid is the proposed approximate solution. This idea was then adapted to the case where there is an uncertainty in the problem of the leader [38].

Apart from this direct algorithm most of the papers first start with a reformulation and then use algorithms for solving the corresponding reformulation. In [119] an MPCC reformulation was considered and then the problem was solved using a smoothing approach of the complementarity constraints.

The MPCC reformulation is commonly preferred (see discussion in [17]) since it benefits from a more explicit expression. On the other hand, the OPVIC reformulation, being a more direct one, is preferred whenever the constraint qualifications of the lower level problem cannot be established or are too difficult to be proven.

Numerical approaches for the OPQVIC reformulation have been considered in [117, 118] for the general case, while OPVIC have been considered in [120].

On the other hand algorithms developed for the resolution of the MPCC reformulation face the difficulty of the treatment of the complementarity constraints involving the Lagrange multipliers. The main numerical techniques are the smoothing, the decomposition, the penalization, and the relaxation approaches.

Simply to illustrate one of these approaches we give below the main steps of the application of the relaxation method to SLMFG. In the KKT system, the constraints of the form $0 \leq \mu_j \perp -g_j(x, y) \geq 0$ can be described by the nonlinear system

$$\begin{cases} -\mu_j g_j(x, y) \leq 0, & \forall\, j = 1, ..., M \\ \mu_j \geq 0,\ -g_j(x, y) \geq 0, & \forall\, j = 1, ..., M. \end{cases}$$

The source of main difficulties is the product constraint. One approach due to Scholtes is to enlarge the feasible set by imposing instead $-\mu_j g_j(x, y) \leq \varepsilon$. By doing this the relaxed problem might now satisfy some CQs and some usual methods (like interior point method used in [39]) for solving the new nonlinear problem can be applied.

The new family of problems would be

$$\min_{x,y,\mu} \quad F(x,y)$$
$$\begin{cases} x \in X, \\ \nabla_{y_j} f_j(x,y) + \sum_{k=1}^{d_j} \mu_{jk} \nabla_{y_j} g_{jk}(x,y) = 0 \\ 0 \leq \mu_j, 0 \leq -g_j(x,y) \\ -\mu_j g_j(x,y) \leq \varepsilon \end{cases}$$

with $\varepsilon > 0$ tending to 0.

The limit of a solution of such problems as ε tends to 0 is a C-stationary solution of the usual MPCC, under suitable constraint qualifications, see e.g. [107].

III.3.3 Multi-Leader-Multi-Follower Games

Let us now focus on Multi-Leader-Follower games in which there are several leaders.

In the case of multiple leaders, some of the ideas of the single-leader case can also be used. In fact, the selection approach can be directly applied since each leaders' function is determined by the leaders' strategies and the selection. The set-valued optimization approach can also be extended (See [52]). In both approaches it is clear what is the value for leaders' objectives a single value for the selection approach, and a set of values in the other.

By the contrary, the idea of the optimistic approach as a cooperation of the followers with the leader now arise the question of with which of the leaders the followers will cooperate. The cooperation with leader 1 could be opposite to the cooperation with leader 2. Anyway, we can consider the conjectures made by the leaders about which optimal reaction of the followers will take place. Of course, the conjectures made by different leaders need not to be equal. We will denote $y_i := (y_{1,i}, ..., y_{N,i}) \in \text{GNEP}(x) \subset \mathbb{R}^m$ the conjecture made by leader i about the followers' optimal reaction, given $x = (x_i, x_{-i})$.

As a simple example one can consider the following MLMFG: let us define a game with two leaders and a follower for which respective variables and objective functions are x_1, x_2, y and $F_1(x_1, x_2, y) = (x_1 - 2)^2 - y$, $F_2(x_2, x_1, y) = (x_2 - 2)^2 + y$, $f(y, x) = x_1 x_2 - (y-1)^2 + 1$. Let us assume that the only constraint on the variables is that the three of them are non negative. Then $\bar{x} = (2,2)$ while $S_1(2,2) = \{2\}$ and $S_2(2,2) = \{0\}$. This difficulty/ambiguity which is fundamental and intrinsically associated to MLMFG with possibly several optimal responses for the follower's problem, is unfortunately often neglected in the literature, in particular in works dedicated to applications.

We can define what we call a multi-optimistic solution of the MLFG as the equilibrium of the upper level GNEP each leader taking as his conjecture that the followers will cooperate with him.

Definition 28. *We define a* multi-optimistic equilibrium *of the MLFG to be a vector* $(\bar{x}, \bar{y}_1, ..., \bar{y}_N)$*, where* $(\bar{x}_i, \bar{y}_i)$ *is a solution of the problem*

$$\min_{x_i,y} F_i(x_i, \bar{x}_{-i}, y)$$
$$s.t. \begin{cases} x_i \in X_i(\bar{x}_{-i}) \\ y \in \text{GNEP}(x_i, \bar{x}_{-i}) \end{cases}$$

with $\bar{y}_i \in \text{GNEP}(\bar{x})$ *the* conjecture *of leader i. If the game has only one leader, we just call* $(\bar{x}, \bar{y})$ *a* optimistic equilibrium.

Note that the multi-optimistic equilibrium notion does define a GNEP between the leaders. Nevertheless, the question about existence of solutions cannot be deduced directly from the existence result for GNEPs (Theorem 19), because the constraint set-valued map for the leaders (including the followers' reaction map) fail in general to be lower semi-continuous.

We know that the GNEP of the lower level could have non-unique solutions, and we can overcome this with the multi-optimistic approach. Nevertheless, the constraint of each leader involves the graph of a parametric GNEP so that, even with very nice data, this constraint set is not necessarily convex, not even connected (see examples in Chapter III.1). Thus, the problem of each leader is in general a non-convex program.

III.3.3.1 Existence for MLFG

In the literature, existence results for MLMFG games are scarce and most of them (if not all, see for instance [60,108]) are based on a technique that we present now. The technique is basically to reduce the MLMFG to a Nash equilibrium problem by 'plugging' the unique lower level response into the leaders' objectives, and then trying to prove some good properties of the resulting Nash equilibrium problem:

$$\boxed{\begin{array}{ll} \min_{x_1} & F_1(x, y(x)) \\ \text{s.t.} & x_1 \in X_1(x_{-1}) \end{array}} \quad \cdots \quad \boxed{\begin{array}{ll} \min_{x_N} & F_N(x, y(x)) \\ \text{s.t.} & x_N \in X_N(x_{-N}) \end{array}}$$

The general assumptions for this technique are:

(A1) for each leaders' profile of strategies x there exists a unique lower level response $y(x)$,

(A2) for any i, the leaders' objectives F_i and the best response function y are continuous,

(A3) for any i, there exist a non-empty, convex and compact set $K_i \in \mathbb{R}^{n_i}$ such that the set-valued map $X_i : K_i \rightrightarrows K_{-i}$ is both upper and lower semi-continuous with non-empty closed and convex values, where $K_{-i} := \prod_{k \neq i} K_k$,

(A4) for any i, the composition functions $\tilde{F}_i(x) := F_i(x, y(x))$ are quasi-convex with respect to x_i.

Proposition 13. *Assume the above conditions (A1) to (A4). Then the MLMFG admits a solution.*

Proof. According to [63] the Nash equilibrium problem defined by the objectives $\tilde{F}_i$, $i = 1, ..., N$ admits an equilibrium. Thus, $\bar{x}$ along with the corresponding reaction of the followers $\bar{y} := y(\bar{x})$ yield an equilibrium $(\bar{x}, \bar{y})$ of the MLMFG. □

The most intricate condition is (A4). In fact, since usually $y(x)$ is only described implicitly, verifying the quasi-convexity of that composition is very difficult in general, but in some cases it is though possible as has been shown by some researchers.

Sherali in [108, Theorem 2] provided, to the best of our knowledge, the first existence result for a particular class of MLFG, by somehow using this technique. In the context of an oligopolistic Stackelberg-Nash-Cournot competition, a group of firms (the leaders) have objectives $F_i(x,y) := x_i p(\sum_k x_k + \sum_j y_j) - c_i(x_i)$ while the rest of the firms (the followers) have objectives $f_j(x,y) := y_j p(\sum_i x_i + \sum_l y_l) - c_j(y_j)$, where p is the inverse demand function and the c_i, c_j are cost functions.

It is proved in [108, Lemma 1, Theorem 3] that under some reasonable assumption on the inverse demand function p and on the cost functions, the $\tilde{F}_i$'s are convex in x_i. The corresponding existence result ([108, Theorem 2]) can be then expressed as follows:

Theorem 27. *Assume that p is strictly decreasing, twice differentiable and $p'(z) + zp''(z) \leq 0$ for each $z \geq 0$, and that c_i and c_j are non-negative, non-decreasing, convex and twice differentiable and there exists $z_u > 0$ such that $c_i'(z) \geq p(z)$ and $c_j'(z) \geq p(z)$ for all $z \in [0, z_u]$. If the map $x \mapsto \sum_j y_j(x)$ is convex (if for instance, p is linear), where $y(x)$ is the unique equilibrium response of the followers, then the MLMFG has at least one equilibrium.*

Fukushima and Hu's existence results (Theorems 4.3 and 4.4 in [61]) are also obtained using the same technique but in a more general setting that considers uncertainty in both levels and a robust approach.

A different technique has been proposed in [72] which is based on the ideas of potential game theory, see [82]. A first possibility is again based on the uniqueness of the lower level responses, that is, condition (A1). A MLMFG is *implicitly potential* if there exists a so-called *potential* function π for the game defined by the functions $\tilde{F}_i$, that is, for all i and for all $x = (x_i, x_{-i})$ and x_i' it holds

$$\tilde{F}_i(x_i, x_{-i}) - \tilde{F}_i(x_i', x_{-i}) = \pi(x_i, x_{-i}) - \pi(x_i', x_{-i}). \tag{III.3.4}$$

Let us notice that, as in the previous technique, the existence of the potential for the implicit description of the functions $\tilde{F}_i$ is also an intricate condition. A variant of this approach was proposed also in [72], where it is not assumed that the lower level responses are unique. The game is said to be a quasi-potential game if there exist functions h and π such that the functions F_i have the following structure

$$F_i(x,y) := \phi_i(x) + h(x,y) \tag{III.3.5}$$

and the family of functions ϕ_i, $i = 1, ..., N$, admit π as a potential function, that is

$$\phi_i(x_i, x_{-i}) - \phi_i(x_i', x_{-i}) = \pi(x_i, x_{-i}) - \pi(x_i', x_{-i}). \tag{III.3.6}$$

The existence of equilibria for the MLMFG can be deduced, in the first case from the existence of a global minimizer of the potential function, as usual in potential games. In the second case, the existence of equilibria for MLFG can be deduced from the minimization of $\pi + h$, which is not strictly speaking a potential function for the F_i. In fact, $\pi + h$ is defined in the space $X_1 \times ... \times X_N \times Y$, while a potential function for the quasi-potential game should be defined on the product of the strategy spaces $(X_1 \times Y) \times ... \times (X_N \times Y)$, for instance as $\psi(x_1, y_1, ..., x_N, y_N) := \pi(x) + \sum_i h(x, y_i)$. We can thus call the function $\pi + h$ a quasi-potential function for the game.

The following theorem [72] shows a way of computing an equlibrium in a pseudo potential MLMFG.

Theorem 28. *Assume that the MLMFG is a pseudo-potential game, and that the constraints set of player i is a constant set equal to a non-empty compact and convex K_i. Then any minimizer of the pseudo potential function $\pi + h$ corresponds to a solution of the MLMFG.*

III.3.3.2 Non-existence for MLFG and a remedial model

In [93], Pang and Fukushima give a nice example of a simple MLFG, which we recall here, which admits no solution. This is related to the non-convexity of the values of the best response map of one of the leaders (and assumption (A4) does not hold), which does not allow to apply Proposition 13.

Example 18. *Consider a game between two leaders and one follower. Given leaders' strategies $x_1 \in X_1$ and $x_2 \in X_2$ with $X_1, X_2 := [0,1]$, the follower reacts by solving the optimization problem*

$$\min_{y \geq 0} \left\{ y(x_1 + x_2 - 1) + y^2/2 \right\},$$

whose unique solution is given by $y = \max\{0, 1 - x_1 - x_2\}$. Taking into account the optimal reaction of the follower, and the opponent leader's strategy as a parameter, leader 1 solves the optimization problem

$$\begin{array}{rll} L_1(x_2) : & \displaystyle\min_{x_1, y} & \dfrac{1}{2}x_1 + y \\ & s.t. & \left\{ \begin{array}{l} x_1 \in [0,1] \\ y = \max\{0, 1 - x_1 - x_2\} \end{array} \right. \end{array}$$

and leader 2 solves

$$\begin{array}{rll} L_2(x_1) : & \displaystyle\min_{x_2, y} & -\dfrac{1}{2}x_2 - y \\ & s.t. & \left\{ \begin{array}{l} x_2 \in [0,1] \\ y = \max\{0, 1 - x_1 - x_2\} \end{array} \right. \end{array}$$

The reaction maps $\mathcal{R}_1 : X_2 \to X_1$, $\mathcal{R}_2 : X_1 \to X_2$ that capture the best response for leaders 1 and 2, respectively, are given by:

$$\mathcal{R}_1(x_2) = \{1 - x_2\}, \quad x_2 \in [0,1]$$

and

$$\mathcal{R}_2(x_1) = \left\{ \begin{array}{ll} \{0\}, & x_1 \in [0, 1/2) \\ \{0, 1\}, & x_1 = 1/2, \\ \{1\}, & x_1 \in (1/2, 1] \end{array} \right.$$

It is easy to see that $\mathcal{R} := \mathcal{R}_1 \times \mathcal{R}_2$ has no fixed points and thus the game has no equilibrium.

In [71], Kulkarni and Shanbhag propose a remedial model for MLFG that consists in including in each leader's constraint also the opponents' equilibrium constraints. In the above example, the modified problem of leader 1 would be

$$\begin{array}{rll} L_1(x_2, y_2) : & \displaystyle\min_{x_1, y_1} & \dfrac{1}{2}x_1 + y_1 \\ & \text{s.t.} & \left\{ \begin{array}{l} x_1 \in [0,1] \\ y_1 = \max\{0, 1 - x_1 - x_2\} \\ y_2 = \max\{0, 1 - x_1 - x_2\} \end{array} \right. \end{array}$$

and for leader 2

$$L_2(x_1, y_1) : \min_{x_2, y_2} \quad -\frac{1}{2}x_2 - y_2$$
$$\text{s.t.} \quad \begin{cases} x_2 \in [0,1] \\ y_1 = \max\{0, 1 - x_1 - x_2\} \\ y_2 = \max\{0, 1 - x_1 - x_2\} \end{cases}$$

Each leader's problem has now two parameters: the opponent's variable and the opponent's conjecture about the follower. For this new model a solution does exist and is given by $(x_1, y_1) = (0, 0)$ and $(x_2, y_2) = (1, 0)$. In fact, for leader 2 the unique feasible solution $(1, 0)$ is optimal, and for leader 1 the feasible set is $[0, 1] \times \{0\}$ for which we clearly deduce that $(0, 0)$ is optimal.

The last example is not a particular one for which the modified formulation admits solutions. In fact, we will see that for any game which admits a potential for the leaders, at least one solution exists in the modified formulation. The reformulation of the MLFG, proposed in [71], is the so-called *All Equilibrium* formulation. Actually, in [71], the authors consider an EPCC formulation of the problem instead of a MLFG, that is the lower level problem is described by a Variational Inequality in place of the lower level GNEP. Both approaches coincide whenever each of the followers' problem is convex (that is $f_j(x, y_j, y_{-j})$ and $g_{j,k}(x, y_j)$ are convex with respect to y_j) and with a feasible set not depending on the other followers' variables. It is a modified game, (or equilibrium problem) denoted by $\mathcal{E}^{\text{ae}}$, that extends the initial MLFG $\mathcal{E}$ in the sense that the set of solutions of $\mathcal{E}^{\text{ae}}$ includes all the solutions of the initial game $\mathcal{E}$, but for which proving existence is easier.

Let us define what is the All Equilibrium formulation $\mathcal{E}^{\text{ea}}$. A leader i has the same objective and the constraints of the initial game defined by the set-valued map

$$\Omega_i(x_{-i}) := \{(x_i, y_i) \mid x_i \in X_i(x_{-i}),\, y_i \in \text{GNEP}(x)\}$$

given opponent strategies x_{-i}, but also the constraints $y_{i'} \in \text{GNEP}(x)$ for all $i' \neq i$, that actually depend on the conjecture y_{-i} made by other leaders. Thus, the constraints for player i in $\mathcal{E}^{\text{ea}}$ is defined by the set-valued map

$$\Omega_i^{\text{ae}}(x_{-i}, y_{-i}) := \{(x_i, y_i) \mid x_i \in X_i(x_{-i}),\, y_{i'} \in \text{GNEP}(x), \forall i' = 1, ..., N\},$$

which is contained in the previous one, and now does depend on both the opponents' strategy and conjecture (x_{-i}, y_{-i}). So, given x_{-i} the feasible set for player i does also depend on y_{-i} and is a subset of the initial feasible set: $\Omega_i^{\text{ea}}(x_{-i}, y_{-i}) \subset \Omega_i(x_{-i})$ for any y_{-i}. We use also the notation $\mathcal{F} := \{(x, y) \mid (x_i, y_i) \in \Omega_i(x_{-i}),\ \forall i\}$. We also note that in $\mathcal{E}^{\text{ae}}$, a (upper level) GNEP with the special structure of shared constraints is played among the leaders.

We now give a definition of solutions of the modified game $\mathcal{E}^{\text{ea}}$ that corresponds to Definition 28 for $\mathcal{E}$. It is basically the same definition, but replacing the set-valued map Ω_i by Ω_i^{ae} for each leader $i = 1, ..., N$.

Definition 29. *We define a (multi-optimistic) solution/equilibrium of $\mathcal{E}^{ae}$ to be a joint strategy $(\bar{x}, \bar{y}_1, ..., \bar{y}_N)$ where for each $i = 1, ..., N$, the pair $(\bar{x}_i, \bar{y}_i)$ is a solution of the problem*

$$\min_{x_i, y_i} F_i(x_i, \bar{x}_{-i}, y_i)$$
$$s.t. \begin{cases} x_i \in X_i(\bar{x}_{-i}) \\ y_{i'} \in \text{GNEP}(x_i, \bar{x}_{-i}),\ \forall i' = 1, ..., N \end{cases}$$

with $\bar{y}_i \in \text{GNEP}(\bar{x})$ the conjecture made by leader i about the followers' reaction.

In this reformulation, given the upper level strategies x, the selection y_i has no additional constraints, but x is constrained implicitly by the other players conjecture in $y_{i'} \in \text{GNEP}(x)$, for $i' \neq i$. In [71, Proposition 3.1, (iii)], it was proved that $\mathcal{E}^{ae}$ is an extension of $\mathcal{E}$, in the following sense.

Proposition 14. *Any solution of the game $\mathcal{E}$ is a solution of the game $\mathcal{E}^{ea}$.*

Now, for the enhanced game it is proved in [71, Theorem 3.3] that solutions do exist under mild assumptions. For instance, if the leaders admits a potential: there exists π continuous such that

$$F_i(x_i, x_{-i}, y_i) - F_i(\tilde{x}_i, x_{-i}, \tilde{y}_i) = \pi(x_i, x_{-i}, y_i) - \pi(\tilde{x}_i, x_{-i}, \tilde{y}_i), \tag{III.3.7}$$

for all $x_i, \tilde{x}_i, y_i, \tilde{y}_i, x_{-i}, y_{-i}$, and the set of strategies is contained in a compact set, then an equilibrium of $\mathcal{E}^{\text{ae}}$ exists.

Theorem 29. *Assume that there is a potential π for the leaders, that all the cost functions F_i and f_j are continuous, and that $f_j(x, y_j, y_{-j})$ and $g_{j,k}(x, y_j)$ are convex with respect to y_j. If there exists a minimizer of π over $\mathcal{F}$ (for instance, if either π is coercive on $\mathcal{F}$, or if $\mathcal{F}$ is compact), then $\mathcal{E}^{ae}$ admits an equilibrium.*

The All-Equilibrium formulation is an extension of the initial game which guarantees existence of solutions under reasonable assumptions. It is possible that some joint strategies that were not solution of the initial game $\mathcal{E}$ are now solution of $\mathcal{E}^{\text{ae}}$, even if there exist solutions for the initial game. The question now is how far do we get with this extension. The simple example below shows that the All-Equilibrium concept is somehow a “too big extension”.

Example 19. *Let there be two leaders and one follower with cost functions given by $F_1(x_1, x_2, y) = x_1$, $F_2(x_1, x_2, y) = x_2$ and $f(x_1, x_2, y) := |x_1 + x_2 - y|$, for $x_1, x_2 \in [0, 1], y \in [0, 2]$. It is direct to see that, given $x_1, x_2 \in [0, 1]$, the unique optimal reaction of the follower is $y = x_1+x_2$, so that the conjectures of both players are the actual reaction of the follower. Note that in this example, the reaction of the follower does not affect the leaders' costs, and thus it is very easy to check that there is only one equilibrium which is $(\bar{x}_1, \bar{x}_2, \bar{y}_1, \bar{y}_2) = (0, 0, 0, 0)$. By the contrary, in the All Equilibrium formulation any combination of leaders' feasible strategies $(x_1, x_2) \in [0, 1]^2$, with the corresponding conjectures $y_1 = y_2 = x_1 + x_2$, yield a solution $(x_1, x_2, x_1 + x_2, x_1 + x_2)$ of $\mathcal{E}^{ae}$.*

Actually, the above example is only one item of a more general family of counterexamples.

Proposition 15. *Consider any multi-leader-follower game with at least two leaders, with the following assumption for each leader i: for any $x = (x_i, x_{-i})$ that is a fixed point of X and any $\tilde{x}_i \in X_i(x_{-i}) \setminus \{x_i\}$ it holds that $\text{GNEP}(\tilde{x}_i, x_{-i}) \cap \text{GNEP}(x_i, x_{-i}) = \emptyset$. Then, every strategy x of the leaders which is a fixed point of K is automatically a solution of the All Equilibrium formulation of the game.*

Proof. Let $\bar{x}$ be any joint strategy of the leaders that is a fixed point of X, and let $\bar{y}_i \in \text{GNEP}(\bar{x})$ be an optimistic conjecture made by leader i, that is,

$$F_i(\bar{x}, \bar{y}_i) \leq F_i(\bar{x}, y), \quad \forall y \in \text{GNEP}(\bar{x}). \tag{III.3.8}$$

Now, consider a leader $i' \neq i$ and his optimistic conjecture $\bar{y}_{i'}$. In the All Equilibrium formulation, leader i has the constraint $\bar{y}_{i'} \in \text{GNEP}(x_i, \bar{x}_{-i})$ (where $\bar{y}_{i'}$ and $\bar{x}_{-i}$ are parameters for player i) which, under the hypothesis, does not admit other x_i than $\bar{x}_i$. We deduce from (III.3.8) that $(\bar{x}_i, \bar{y}_i)$ is a best response of player i to $(\bar{x}_{-i}, \bar{y}_{-i})$. Since the same analysis can be done for all the players, we deduce that $(\bar{x}, \bar{y})$ is an equilibrium of $\mathcal{E}^{ae}$. □

Thus, looking to the above examples, it is clear that the All Equilibrium approach can generate in some cases a very large set of equilibria, which can be quite difficult to interpret in applications.

III.3.3.3 Reformulations

As explained in the previous section, the analysis of MLMFG in the literature is mostly focused on the case of a unique lower level response. Under this assumption, the lower level response can be plugged into the leaders objectives transforming the initial MLMFG "simply" into a Nash equilibrium problem, though with quite complicated objective functions, in general non-smooth and non-convex. Then, the usual techniques used for solving Nash equilibrium problems can be used for this formulation.

In particular, in [60, 61] the function of (unique) lower level responses are linear with respect to the leaders variables and can be somehow plugged into the leaders objective because of the specific structure that is considered (there is a term in the leaders' objective that is also present in the followers' objective but with negative sign). The resulting Nash equilibrium problem is reformulated as a variational inequality and a forward-backward splitting method is applied to solve the variational inequality.

In the case of non-uniqueness of the lower level problem, and considering the (possibly inconsistent) multi-optimistic MLMFG, we can extend the approach of the SLMFG case by replacing the lower level equilibrium problem by the concatenation of KKT conditions of the followers. The resulting reformulation is a so-called Equilibrium Problem with Complementarity Constraints (EPCC, for short).

A first question to address concerns the equivalence between the initial MLMFG and its EPCC reformulation, facing similar arguments as for SLMFG, that is requiring an infinite number of CQs to be verified. The analysis made for the case of one leader in the previous section can be easily extended to the case of multiple leaders as done in [17, Theorem 4.9], by considering the joint convexity of the followers' constraint functions and the joint Slater's CQ.

Let us precise what is the corresponding EPCC for the case of the multi-optimistic approach. In leader i's problem, for any given x_{-i}, we replace the condition $y \in \text{GNEP}(x_i, x_{-i})$ by the concatenation of KKT conditions of the followers, that is,

$$\begin{cases} \nabla_{y_j} f_j(x,y) + \mu_j \nabla_{y_j} g_j(x,y) = 0 \\ 0 \leq \mu_j \perp -g_j(x,y) \geq 0 \end{cases} \qquad \forall j = 1, ..., M. \tag{III.3.9}$$

Let us denote by $cKKT(x)$ for the set of vectors $(y, u) = (y_1, ..., y_M, \mu_1, ..., \mu_M)$ that satisfy the concatenated KKT conditions (III.3.9). We write $\Lambda(x, y)$ for the set of vectors μ such that $(y, \mu) \in cKKT(x)$.

The EPCC reformulation of a MLFG (in the multi-optimistic approach) is a GNEP played by the leaders where the constraint $y \in \text{GNEP}(x)$ is replaced by $(y, \mu) \in cKKT(x)$. More precisely,

Definition 30. *We define an* equilibrium *for the EPCC reformulation of the (multi-optimistic) MLFG to be a vector* $(\bar{x}, \bar{y}_1, ..., \bar{y}_N, \bar{\mu}_1, ..., \bar{\mu}_N)$ *where for each* $i = 1, ..., N$, $(\bar{x}_i, \bar{y}_i, \bar{\mu}_i)$ *is a solution of the problem*

$$\min_{x_i, y, \mu} F_i(x_i, \bar{x}_{-i}, y)$$
$$s.t. \quad \begin{cases} x_i \in X_i(\bar{x}_{-i}) \\ (y, \mu) \in cKKT(x_i, \bar{x}_{-i}) \end{cases}$$

In the case of only one leader, an equilibrium is called a global solution of the MPCC. We define a global optimal solution *of the MPCC reformulation of a SLMF game to be a vector* $(\bar{x}, \bar{y})$ *that solves the problem*

$$\min_{x, y, \mu} F(x, y)$$
$$s.t. \quad \begin{cases} x \in X \\ (y, \mu) \in cKKT(x) \end{cases}$$

Note that even though the upper level cost function F is not dependent on the multiplier variable μ, it is considered as a variable, because it appears in the constraints of each leaders' problem. This will be of special importance when studying local solutions.

We will discuss here the relation between the MLFG and the EPCC reformulation. First, in Subsection III.3.3.3 we focus on global solutions of both the problem and the reformulation, and in Subsection III.3.3.3 we will make the analysis for local solutions. Both of these cases were first investigated in [41] in the case of one leader and one follower for the optimistic approach while in [18] the pessimistic counterpart was studied.

One reason for considering this reformulation is that it allows us to use the machinery/theory of MPCCs to solve the problem. But an important question is what is the relation between solutions of the optimistic BLP (respectively MLFG) and solutions to the MPCC-reformulation (respectively EPCC). Of course, some assumptions on the lower level problem have to be made. To ensure that the concatenation of KKT conditions of the followers are necessary and sufficient equilibrium conditions for the lower level GNEP, we will assume some basic convexity conditions, and some constraint qualifications for the followers' problems.

Global solutions relation

We show in the next theorem some relations between solution of a MLFG and the corresponding EPCC, where the lower level is replaced by the concatenation of KKT conditions of the followers.

Theorem 30 (Global solution relations for MLFG)**.** *Assume that the followers' problems are player convex, that is,* $f_j(x, \cdot, y_{-j})$ *and the components* $g_{j,k}(x, \cdot, y_{-j})$ *are convex for all followers* j.

(1) Let $(\bar{x}, \bar{y})$ *be a multi-optimistic equilibrium of the MLFG, for which Slater's CQ is satisfied for each follower. Then* $\Lambda(\bar{x}, \bar{y}) \neq \emptyset$ *and for each* $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$*, the point* $(\bar{x}, \bar{y}, \bar{\mu})$ *is an equilibrium of the EPCC-reformulation in the sense of Definition 30.*

(2) Let $(\bar{x}, \bar{y}, \bar{\mu})$ be an equilibrium of the EPCC-reformulation of the MLFG, and assume that for each follower j's problem for each $x \in X$ and each y_j, Slater's CQ is fulfilled. Then $(\bar{x}, \bar{y})$ is a multi-optimistic equilibrium of MLFG.

Proof. For *(1)*, since $\bar{y} \in \text{GNEP}(\bar{x})$ and thanks to Slater's CQ for each follower, we deduce that $\Lambda(\bar{x}, \bar{y}) \neq \emptyset$ and that for any $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$ the vector $(\bar{x}, \bar{y}, \bar{\mu})$ is feasible for EPCC. In the other hand, and thanks to the convexity condition of the followers, the projection of the solution of the KKT system into the variable y is included into the solution of the GNEP, so that the optimality of $(\bar{x}, \bar{y}, \bar{\mu})$ directly comes from the optimality of $(\bar{x}, \bar{y})$ as a subset of inequalities.

For *(2)*, the stronger CQ assumptions in this case guarantee that the projection of the KKT system of the lower level into the y variable coincide with the lower level GNEP, so that the equivalence between MLFG and its EPEC reformulation follows. □

The main difficulty of part (2) (the interesting part) of Theorem 30 is the possibly infinite number of qualification conditions that have to be satisfied. In [19], it was proposed to consider not only player convexity of the lower level GNEP but joint convexity of the lower level constraint functions in order to reduce significantly the number of CQ to be verified for the desired equivalence. Also for the case of SLMF (and BLP) the convexity of the constraint functions of the follower in the full vector of strategies of the leader and the followers, allows us to reduce the number of CQs. Here we make a simple extension of this result for the case of several leaders. In the following theorem we only show the "difficult and interesting" subset-relation of solutions.

Let us consider the feasible set of follower j defined by $Y_j(x, y_{-j}) := \{y_j \mid g_j(x, y_j, y_{-j}) \leq 0\}$ and the following feasibility assumption

$$\forall j = 1, ..., , M, \; \forall x \in \mathbb{R}^n, \; \forall y_{-j} \in \mathbb{R}^{m-j}, Y_j(x, y_{-j}) \neq \emptyset, \tag{III.3.10}$$

in other words, $\operatorname{dom} Y_j = \mathbb{R}^n \times \mathbb{R}^{m_j}$ for all $j = 1, .., M$.

Theorem 31 (Global solution relations for MLFG)**.** *Assume that the followers' objectives f_j are convex on y_j and that the component of the constraint functions $g_{j,k}$ are jointly convex (on (x, y)). Assume that the feasibility assumption* (III.3.10) *holds and that Joint Slater's CQ is fulfilled: For each follower j there exist a joint strategy (x, y) such that $g_{j,k}(x, y) < 0$ for all $k = 1, ..., d_j$. If $(\bar{x}, \bar{y}_1, ..., \bar{y}_N, \bar{\mu}_1, ..., \bar{\mu}_N)$ is an equilibrium of the EPCC reformulation of a multi-optimistic MLFG, then $(\bar{x}, \bar{y}_1, ..., \bar{y}_N)$ is an equilibrium of the multi-optimistic MLFG.*

Proof. We just verify the hypothesis of Theorem 30 *(2)*. These hypotheses follow from Theorem 1 applied for a follower j whose constraint function g_j is parametrized on the vector (x, y_{-j}). □

Local solutions relation

We define here the local concept of solution that corresponds to Definition 28.

Definition 31. *We define a* multi-optimistic local equilibrium *of the MLFG to be a vector $(\bar{x}, \bar{y}_1, ..., \bar{y}_N)$, where $(\bar{x}_i, \bar{y}_i)$ is a local solution of the optimization problem*

$$\min_{x_i, y} F_i(x_i, \bar{x}_{-i}, y)$$
$$s.t. \quad \begin{cases} x_i \in X_i(\bar{x}_{-i}) \\ y \in \text{GNEP}(x_i, \bar{x}_{-i}) \end{cases}$$

Definition 32. *We define an* local equilibrium *for the EPCC reformulation of the (multi-optimistic) MLFG to be a vector* $(\bar{x}, \bar{y}_1, ..., \bar{y}_N, \bar{\mu}_1, ..., \bar{\mu}_N)$ *where for each* $i = 1, ..., N$, $(\bar{x}_i, \bar{y}_i, \bar{\mu}_i)$ *is a local solution of the optimization problem*

$$\min_{x_i, y, \mu} F_i(x_i, \bar{x}_{-i}, y)$$
$$s.t. \quad \begin{cases} x_i \in X_i(\bar{x}_{-i}) \\ (y, \mu) \in cKKT(x_i, \bar{x}_{-i}) \end{cases}$$

The following theorem explains how local equilibria of the MLFG and local equilibria of the EPCC reformulation are related: this relation is more complicated than for global equilibrium.

Theorem 32. *Assume that the objectives* f_j *and the components of the constraints* $g_{j,k}$ *of the followers are convex on their variable* y_j*, and for each follower j's problem Slater's CQ is satisfied for the parameter* $(\bar{x}, \bar{y}_{-j})$.

1. *If* $(\bar{x}, \bar{y}_1, ..., \bar{y}_N, \bar{\mu}_1, ..., \bar{\mu}_N)$ *is a local equilibrium of the EPCC reformulation of a MLFG for all* $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$*, then the point* $(\bar{x}, \bar{y}_1, ..., \bar{y}_N)$ *is a local equilibrium of the MLFG.*

2. *If* $(\bar{x}, \bar{y}_1, ..., \bar{y}_N)$ *is a local equilibrium of the MLFG, then for all* $\bar{\mu} \in \Lambda(\bar{x}, \bar{y})$, $(\bar{x}, \bar{y}, \bar{\mu})$ *is a local equilibrium of the EPCC reformulation of the MLFG.*

Proof. Similarly to Theorem 31, the proof follows from applying Theorem 1 to each follower's problem now parametrized on (x, y_{-j}). □

Another example [18, Example 1.1] shows that even with $(\bar{x}, \bar{y}, \bar{\mu})$ a local solution of MPCC, it can occur that $(\bar{x}, \bar{y})$ is NOT a local solution of BLP.

III.3.3.4 Algorithms

The above mentioned EPEC reformulation of the MLMFG is an equilibrium problem (among leaders) so that we could be tempted to use the machinery for equilibrium problems to solve the MLFG. Nevertheless, considering that the specific type of constraints (equilibrium constraints) are very ill-behaved non-convex and non-smooth, solving the equilibrium problem is in fact extremely challenging.

If we are facing a pseudo-potential game (see previous subsection), then the problem can be solved by minimizing the pseudo-potential function, constrained by the equilibrium problem, thus going back to the SLMF case (see Theorem 28).

III.3.4 Conclusion and Future Challenges

The aim of the present chapter was to present the recent advances for different kinds of Multi-Leader-Follower games. The cases of a single leader game SLMFG and of a single follower MLMFG play particular roles in applications and it is one of the reasons why the general case MLMFG has been actually less investigated in the literature. However we have seen that, for all the models, a special attention must be addressed to avoid an ill-posed problem and to fix possible ambiguities. Let us add that, as observed for MLMFG in [8], those ambiguities are even more tricky when one deals with reformulation involving Lagrange multipliers. Applications of MLFG are numerous and have been well explored (energy or water management, economics, pollution control, telecommunications, metro pricing [97], etc.) but from a theoretical point of view a lot of questions are still open concerning SLMFG, MLMFG and of course even more for MLMFG. For example, to our knowledge, very few papers (see e.g. [65]) consider sensitivity/stability analysis for MLFG. In the same vein, gap functions has not been studied for this class of problems.

We restricted ourself to deterministic versions of MLFG because considering stochastic models would have been beyond the scope of this chapter. But it is important to mention that some models and results in settings with uncertainties or random variables have been recently studied, see e.g. [81, 116].

Models with more than two levels were also not considered here. Some preliminary studies appeared (see e.g. [7, 15]) but applications are calling for more analysis of such models.

Finally we would like to emphasize that one keystone to push further the analysis of MLFG could be to consider, at least as a first step, some specific structures/models like the concept of Multi-Leader-Disjoint-Follower problem presented in [6]. Indeed in those particular interactions between leaders and followers could intrinsically carry properties that allow to obtain more powerful results.

Part IV

Quasi-Equilibrium Problems and Bi-Functions

This part is based on the two unpublished papers [33] and [34].

The theory of equilibrium problems is a quite general one which presents in a unified manner several problems of optimization, game theory, complementarity problems and quasi-variational inequalities.

We will start defining a standard version of the quasi-equilibrium problem and its dual version, the Minty quasi-equilibrium problem. We will be primarily interested on the question of existence of equilibria, that is, solutions of the problems. In the next sections we will discuss some of the special cases of equilibrium problems. In each case, we also show how the results of the general theory are applied, recovering this way some known results in the literature and even improving some other.

IV.1 Quasi-Equilibrium Problems

Let X be a non-empty set. A function of the form $f : X \times X \to \mathbb{R}$ will be called here a *bi-function*. Consider also a set-valued map $K : X \rightrightarrows X$, which we refer here to as the *constraint map*.

The standard version of the *quasi-equilibrium problem* is to find

$$x \in \operatorname{Fix} K \text{ such that } f(x, y) \geq 0 \text{ for all } y \in K(x). \tag{1}$$

This problem is sometimes referred to as the *Stampacchia* quasi-equilibrium problem. In a dual way, the *Minty quasi-equilibrium problem* is to find

$$x \in \operatorname{Fix} K \text{ such that } f(y, x) \leq 0 \text{ for all } y \in K(x). \tag{2}$$

We denote by $\mathrm{QEP}(f, K)$ and $\mathrm{MQEP}(f, K)$.

In case of a constant constraint map $K(x) := C$ for all $x \in X$, the above problems simply reduce to the *equilibrium problem*, which is to find

$$x \in C \text{ such that } f(x, y) \geq 0 \text{ for all } y \in C. \tag{3}$$

and the *Minty equilibrium problem*, which is to find

$$x \in C \text{ such that } f(x, y) \geq 0 \text{ for all } y \in C. \tag{4}$$

There is of course a symmetry in the definitions of Stampacchia and Minty quasi-equilibrium problems. In fact, changing the roles of the two variables and the sign of the bi-function. Therefore, we will focus on existence results for the Stampacchia-type, though the Minty-type solutions will serve sometimes as a tool to prove existence for Stampacchia. The assumptions are thus usually one-sided.

One of the most simple examples we will have in mind is the case where the bi-function is the difference of a univariate function evaluated in both variables (further examples will be discussed in Section IV.4).

Example 20. *Consider the problem of minimizing a function $g : X \to \mathbb{R}$ over a non-empty set $C \subset X$. If we define $f(x, y) := g(y) - g(x)$, then the minimization problem has the same set of solutions as the equilibrium problem and the Minty equilibrium problem (associated to f and C):*

$$\operatorname*{argmin}_C g = \mathrm{EP}(f, C) = \mathrm{MEP}(f, C).$$

In the general (quasi-) setting, both the quasi-equilibrium problem and the Minty quasi-equilibrium problem associated with this f and a constraint map K coincide with the so-called quasi-equilibrium problem defined as follows

$$x \in \operatorname{Fix} K \text{ such that } g(x) \leq g(y) \text{ for all } y \in K(x). \tag{5}$$

Remark 24. *In Example 20, by swapping the roles of the variables (or changing the sign of f) we define $\tilde{f}(x,y) := f(y,x) = g(x) - g(y)$, and we obtain*

$$\operatorname*{argmax}_C g = \mathrm{EP}(\tilde{f}, C) = \mathrm{MEP}(\tilde{f}, C).$$

In applications (see Example 20 and next sections) it is often the case that $f_D = 0$, where $D := \{(x,x) \mid x \in X\}$, because of the structure of f as a difference of functions. We do not make this a blanket assumption but in some cases it is implicitly required by other assumptions as we shall see.

In order to study the existence of equilibria, we shall use properties of set-valued maps and also some generalized convexity properties, discussed in other chapters. Additionally, we will define some properties that are specific for bi-functions.

IV.2 Properties of Bi-functions

IV.2.1 Generalized Monotonicity

Firstly, we recall some of the so-called generalized monotonicity properties for bi-functions.

Definition 33. *We say that $f : X \times X \to \mathbb{R}$ is:*

- cyclically monotone *if $\sum_{i=1}^n f(x_i, x_{i+1}) \leq 0$ for all $n \in \mathbb{N}$ and $x_1, ..., x_{n+1} \in X$ such that $x_{n+1} = x_1$;*
- cyclically quasi-monotone *if $\min_{i=1}^n f(x_i, x_{i+1}) \leq 0$ for all $n \in \mathbb{N}$ and $x_1, ..., x_{n+1} \in X$ such that $x_{n+1} = x_1$;*
- monotone *if $f(x,y) + f(y,x) \leq 0$ for all $x, y \in X$;*
- pseudo-monotone *if $f(x,y) \geq 0 \Rightarrow f(y,x) \leq 0$ for all $x, y \in X$;*
- quasi-monotone *if $f(x,y) > 0 \Rightarrow f(y,x) \leq 0$ for all $x, y \in X$.*

It is clear that cyclic (quasi-)monotonicity implies (quasi-)monotonicity, simply by taking $n = 2$, $x_1 = x$, and $x_2 = y$. We observe also that monotonicity implies pseudo-monotonicity, and that pseudo-monotonicity implies quasi-monotonicity (see Figure 1).

Moreover, the pseudo-monotonicity of a bi-function has the following characterization.

Proposition 16. *A bi-function f is pseudo-monotone if and only if*

$$f(x,y) > 0 \Rightarrow f(y,x) < 0, \quad \forall x, y \in X. \tag{6}$$

Proof. Let $x, y \in X$. Each implication $f(x,y) > 0 \Rightarrow f(y,x) < 0$ and $f(x,y) \geq 0 \Rightarrow f(y,x) \leq 0$ can be obtained from the other by simply swapping the roles of x and y. □

cyclic monotonicity	$\Rightarrow$	monotonicity
		$\Downarrow$
$\Downarrow$		pseudo-monotonicity
		$\Downarrow$
cyclic quasi-monotonicity	$\Rightarrow$	quasi-monotonicity

Figure 1: Generalized monotonicity properties

The bi-function f in Example 20 is cyclic monotone. Thus, this f satisfies all of the generalized monotonicity properties. The following proposition gives a characterization of cyclic monotonicity.

Proposition 17. *A bi-function $f : X \times X \to \mathbb{R}$ is cyclically monotone if and only if there exists $h : X \to \mathbb{R}$ such that*

$$f(x, y) \leq h(y) - h(x).$$

Proposition 18. *A bi-function $f : X \times X \to \mathbb{R}$ is cyclically quasi-monotone if and only if for any finite and non-empty subset A of X there exists $x \in A$ such that*

$$\max_{a \in A} f(a, x) \leq 0.$$

Proof. Assume first the condition of the proposition. Let $x_1, .., x_{n+1} \in X$ with $x_{n+1} = x_1$. Then for the finite set $A = \{x_i \mid i = 1, ..., n\}$ there exists $x = x_i$ for some $i \in \{1, ..., n\}$ such that $f(a, x_i) \leq 0$ for all $a \in A$. In particular, $f(x_i, x_{i+1}) \leq 0$. So we deduce that f is cyclically quasi-monotone.

Let us now prove the direct implication. Reasoning by contradiction, suppose that there exists $A = \{x_1, x_2, \ldots, x_n\} \subset C$ such that $(\bigcap_{i=1}^n F_{x_i}) \cap A = \emptyset$, where $F_{x_i} = \{y \in C : \ f(x_i, y) \leq 0\}$. This is equivalent to

$$\left(\bigcup_{i=1}^{n} F_{x_i}^c \right) \cup A^c = C. \tag{7}$$

Set $x_{i(1)} = x_1$, equality (7) implies that there exists x_j with $x_j \neq x_1$ such that $x_1 \in F_j^c$, that means $f(x_j, x_1) > 0$. We set $x_{i(2)} = x_j$ and apply the equality (7) again. Continuing in this way, we define a sequence $(x_{i(n)})_{n \in \mathbb{N}}$ such that

$$f(x_{i(k+1)}, x_{i(k)}) > 0 \tag{8}$$

for all $k \in \mathbb{N}$.

Since the set $\{x_1, x_2, \ldots, x_n\}$ is finite, there exist $m, k \in \mathbb{N}$ with $m < k$ such that $x_{i(k+1)} = x_{i(m)}$. We now consider the points

$$\hat{x}_1 = x_{i(m)}, \ \hat{x}_2 = x_{i(k)}, \ \hat{x}_3 = x_{i(k-1)}, \ldots, \hat{x}_{k+1-m} = x_{i(m+1)}$$

which, due to the inequality (8), satisfy

$$f(\hat{x}_j, \hat{x}_{j+1}) > 0$$

for all $j = 1, \ldots, k+1-m$, with $\hat{x}_{k+2-m} = \hat{x}_1$. This means that f is not cyclic quasi-monotone and we get a contradiction.

□

Proposition 18 states that the cyclic quasi-monotonicity of f is equivalent to that $\mathrm{EP}(f, A)$ is non-empty, for any finite and non-empty $A \subset C$. It was recently proved in [69, Theorem 2.7] that cyclic quasi-monotonicity allows to give an existence result for quasi-equilibrium problem that extends Weierstrass theorem.

Theorem 33. *Assume C is non-empty and compact, that $-f$ is cyclically quasi-monotone on C, and that the sublevel set $[f \leq 0]$ is closed. Then $\mathrm{EP}(f, C)$ is non-empty.*

We observe that the generalized monotonicity properties are preserved under the product with a non-negative scalar. The (cyclic) monotonicity is also preserved under addition, while the (cyclic) quasi-monotonicity is preserved under the maximum operation.

Example 21. *We give some example in order to show that all the implications in the graph of Figure 1 are strict, in the sense that the converses are not valid in general. Consider $X = [0, 1]$*

1. *$f(x, y) := \max(0, y - x)$ is quasi-monotone but not pseudo-monotone;*
2. *$f(x, y) := \max(y - x, 2(y - x))$ is pseudo-monotone but not monotone;*
3. *$f(x, y) := sgn(y - x)$ is (quasi-)monotone but not cyclic (quasi-)monotone.*

The quasi-monotonicity of a bi-function f, and thus also any of the generalized monotonicity properties, implies that the bi-function has non-positive values on the diagonal, that is, $f_D \leq 0$.

Now let us assume that X is a vector space. Another property that implies non-positive values in the diagonal is the KKM property. A bi-function f is said to have the *KKM property* (in some places called *proper quasi-monotonicity*) on a convex subset C of X if for any non-empty and finite set $A \subset C$ and $x \in \mathrm{co}(A)$, we have

$$\min_{y \in A} f(y, x) \leq 0.$$

If f has the KKM property on C, then we have $f(x, x) \leq 0$ for all $x \in C$. In particular, the bi-function f in Example 20 has the KKM property.

A property that implies that the bi-function has non-negative values is the upper sign property. A bi-function f is said to have the *upper sign property* if for every $x, y \in X$ the following implication holds

$$\{f(tx + (1 - t)y, y) \leq 0, \forall t \in]0, 1[\} \Rightarrow f(y, x) \geq 0$$

As it can be readily seen, the upper sign property has a strong link with pseudo-monotonicity. In fact, if f has the upper sign property with $f_D \leq 0$ and f is quasi-convex in its first argument, then $-f$ is pseudo-monotone. Conversely, if $-f$ is pseudo-monotone and is lower semi-continuous in its first argument, then f has the upper sign property.

Next it is shown that some of these genralized monotonicity properties allows us to link the solution of the equilibrium problem with those of the Minty equilibrium problem.

Proposition 19. *Assume that at least one of the following conditions hold:*

1. *The bi-function $-f$ is pseudo-monotone on C,*

2. C is a convex subset of a vector space, and f has the upper sign property on C.

Then $\mathrm{EP}(f, C) \subset \mathrm{MEP}(f, C)$.

Proof. The first case was proved in [11, Proposition 3.1] and the second in [35, Proposition 2]. □

Proposition 20. *If f is upper sign continuous and quasi-monotone (or properly quasi-monotone) then $f_D = 0$.*

The following result states that cyclic quasi-monotonicity implies proper quasi-monotonicity, under quasi-convexity assumption.

Proposition 21. *Let C be a convex subset of X and $f : X \times X \to \mathbb{R}$ be a bi-function such that f is quasi-convex in its second argument. If f is cyclic quasi-monotone on C, then f is properly quasi-monotone on C.*

Proof. It is a simple and straightforward adaptation of Proposition 4.4 in [36]. □

Note that the quasi-convexity of f in its second argument cannot be dropped from the assumptions. For instance consider the bi-function f defined by (10) which is always cyclically quasi-monotone but it is properly quasi-monotone if and only if the function h is quasi-convex (see part 2 of Proposition 6.2 in [32]).

IV.2.2 The Finite Intersection Properties

In this section we introduce the notion of finite intersection property and one variant, for bi-functions. We discuss their relation with the generalized monotonicity properties, namely proper quasi-monotonicity, quasi-monotonicity and cyclic quasi-monotonicity, in Propositions 23 and 24, and Remark 25, respectively.

Definition 34. *The bi-function $f : X \times X \to \mathbb{R}$ is said to have:*

- *The* finite intersection property *(fip) on C a subset of X if, for any finite and non-empty subset A of C, there exists $x \in C$ such that*

$$\max_{a \in A} f(a, x) \leq 0.$$

- *The* star finite intersection property *(fip*) on C a convex subset of X if, for any finite and non-empty subset A of C, there exists $x \in \mathrm{co}(A)$ such that*

$$\max_{a \in A} f(a, x) \leq 0.$$

Nessah and Tian, in [87], introduced a condition called the α-locally dominatedness of a bi-function, which corresponds in the case of $\alpha = 0$ to a bi-function with the fip by switching the roles of the variables. They discussed the relation of this property with the finite intersection property for families of sets. In fact, for each $x \in C$ we define the set

$$F_x := \{y \in C : \; f(x, y) \leq 0\}. \tag{9}$$

Clearly, f has the fip on C if and only if, the family of sets $\{F_x\}_{x\in C}$ has the finite intersection property. Similarly, f has the fip* on C if and only if for any non-empty and finite subset A of C it holds that

$$\left(\bigcap_{a\in A} F_a\right) \cap \operatorname{co}(A) \neq \emptyset.$$

It is also possible given a family of subsets of X to construct a natural bi-function that has the fip if and only if the family of sets has the finite intersection property, as follows.

Proposition 22. *Let Λ be a subset of X and $\mathcal{F} = \{C_x\}_{x\in\Lambda}$ be a family of subsets of X. If $\mathcal{F}$ has the finite intersection property, then there exists a bi-function $f : X \times X \to \mathbb{R}$ with fip on X such that set F_x, defined as in (9), coincides with C_x for all $x \in \Lambda$.*

Proof. Consider us the bi-function $f : X \times X \to \mathbb{R}$ defined as

$$f(x,y) := \begin{cases} 0, & x \notin \Lambda \\ 0, & x \in \Lambda \wedge y \in C_x \\ 1, & x \in \lambda \wedge y \notin C_x \end{cases}$$

which satisfies

$$F_x = \{y \in X : \ f(x,y) \leq 0\} = \begin{cases} X, & x \notin \Lambda \\ C_x, & x \in \Lambda \end{cases}.$$

Now, it is not difficult to see that the family of sets $\{F_x\}_{x\in X}$ has the finite intersection property if and only if the family $\{F_x\}_{x\in\Lambda}$ also has it. Therefore, f has the fip on X. □

Observe that under fip* by taking $A_x = \{x\}$ we have $f(x,x) \leq 0$ for every $x \in X$, while fip does not guarantee this in general.

Remark 25. *From Proposition 18 it is clear that cyclic quasi-monotonicity implies fip, and moreover, if C is a convex set then cyclic quasi-monotonicity implies fip* and fip* implies fip. The converses to these implication are in general not true, as shown by the following two simple examples.*

Example 22. *The bi-function $f(x,y) := xy$, for $x,y \in [0,1]$ has the fip, which can be observed since $x = 0 \in \operatorname{MEP}(f,C)$. However, f does not have the fip* on $[0,1]$. Indeed, for $A = \{1\}$ we have $\max_{a\in A} f(a,x) = f(1,1) = 1 > 0$, for all $x \in \operatorname{co}(\{1\})$.*

Example 23. *Let $f : [0,1] \times [0,1] \to \mathbb{R}$ be defined as*

$$f(x,y) := \begin{cases} 0, & \textit{if } \ |x-y| \leq 1/2 \\ 1, & \textit{otherwise.} \end{cases}$$

Let us see that f has the fip*. Consider a non-empty and finite set $A \subset [0,1]$. If $\operatorname{diam} A = \max_{a,b\in A} |a-b| \leq 1/2$, then by taking any point $x \in A$ we obtain $\max_{a\in A} f(a,x) = 0$. Otherwise, if $\operatorname{diam} A = \max_{a,b\in A} |a-b| > 1/2$, then there exist $a_0, a_1 \in A$ such that $a_0 < 1/2 + a_1$ and therefore $1/2 \in [a_0, a_1] \subset \operatorname{co} A$. So taking $x = 1/2$ we have again that $\max_{a\in A} f(a,x) = 0$. Thus f has the fip* on $[0,1]$. But we observe that f is not cyclic quasi-monotone on $[0,1]$, in fact, not even quasi-monotone, since $f(1,0) = f(0,1) > 0$.

Now we present a simple case of bi-functions which have the fip*.

Example 24. *Let $h : X \to \mathbb{R}$ be a function and C be a subset of X. Consider the bi-function $f : X \times X \to \mathbb{R}$ defined by*

$$f(x,y) := h(y) - h(x). \tag{10}$$

It is clear that f is cyclic monotone, thus cyclic quasi-monotone and due to Proposition 18 it satisfies the finite intersection property. Moreover, if C is convex, then again by Proposition 18 we deduce that f satisfies fip* on C.

Following the proof of Proposition 2.1 in [87] we will show that a properly quasi-monotone bi-function has the fip* whenever it is lower semi-continuous on its second argument.

Proposition 23. *Let C be a convex and non-empty subset of X (normed space) and $f : X \times X \to \mathbb{R}$ be a bi-function such that for each $x \in C$ the function $f(x, \cdot)$ is lower semi-continuous. If f is properly quasi-monotone on C, then it has the fip* on C.*

Proof. Let us assume by contradiction that f does not have the fip*. So, there exists $\{x_1, ..., x_m\} \subset C$ such that for any $x \in K := \text{co}(\{x_1, ..., x_m\})$, we have

$$\max_{i=1,...,m} f(x_i, x) > 0.$$

By means of the sets $F_{x_i} := \{y \in K : f(x_i, y) \leq 0\}$, this can be stated equivalently as $\bigcap_{i=1}^m F_{x_i} = \emptyset$. Thus, since the sets F_{x_i} are closed (due to the lower semi-continuity of f in its second argument) then the function $g : K \to \mathbb{R}_+$ defined by

$$g(x) := \sum_{i=1}^m d(x, F_{x_i}),$$

satisfies $g(x) > 0$ for all $x \in K$, and is continuous. Further, the function $h : K \to K$ defined as

$$h(x) := \sum_{i=1}^m \frac{d(x, F_{x_i})}{g(x)} x_i,$$

is continuous too. By Schauder-Tychonoff Fixed Point Theorem we deduce that there exists $\bar{x} \in K$ such that $h(\bar{x}) = \bar{x}$. Consider the set of indices

$$J := \{i = 1, ..., m : d(\bar{x}, F_{x_i}) > 0\}$$

which is non-empty by a simple argument similar to the one used to prove that $g(x) > 0$. Then, $\bar{x} \in \text{co}(\{x_i : i \in J\})$ we have that $\min_{i \in J} f(x_i, \bar{x}) > 0$, but this contradicts the proper quasi-monotonicity of f applied for the finite set of points $\{x_i\}_{i \in J}$ and its convex combination $\bar{x}$. □

The previous result is also true if we replace the lsc of $f(x, \cdot)$ by the condition that the sublevel sets $[f(x, \cdot) \leq 0]$ is closed for each x.

Analogous to Proposition 1.2 in [23], we will show that fip* implies quasi-monotonicity under suitable assumptions.

Proposition 24. *Let $f : X \times X \to \mathbb{R}$ be a bi-function such that $-f$ is semi-strictly quasi-convex in its second argument and $f_D \geq 0$. If f has the fip* on X, then it is quasi-monotone on X.*

Proof. Let $x, y \in X$ such that $f(x,y) > 0$. Since $f(x,x) \geq 0$ then by semi-strict quasi-convexity of $-f(x,\cdot)$ we obtain that

$$f(x, tx + (1-t)y) > 0,$$

for all $t \in]0,1[$. Thus, from the fip* we deduce $f(y,x) \leq 0$. □

Example 23 shows that the semi-strict quasi-convexity of f in its second argument is essential in Proposition 24. In fact, the example proposes a bi-function that has the fip* and vanishes on the diagonal, while it is not quasi-monotone.

It is direct, that whenever $\mathrm{MEP}(f,C)$ is non-empty, then f has the fip on C. Moreover, we have the following result.

Lemma 16. *Let C be a topological space, and $f : C \times C \to \mathbb{R}$ be a bifunction and consider the sets F_x as defined in (9). Assume that for each $x \in C$, the set F_x is closed and that there exists $x \in C$ such that the set F_x is compact, and that f has the fip on C. Then,* $\mathrm{MEP}(f,C)$ *is non-empty.*

Proof. It is clear that $\mathrm{MEP}(f,C) = \bigcap_{x\in C} F_x$. Since $\{F_x\}_{x\in C}$ has the finite intersection property due to the fact that f has the fip on C, and some F_x is compact we deduce that the set $\mathrm{MEP}(f,C)$ is non-empty. □

IV.3 Existence for Quasi-Equilibrium Problems

We present here three general existence results for quasi-equilibrium problems that extend several results from the literature. For simplicity, we will assume that $X = \mathbb{R}^n$ and we consider the following hypothesis.

$$(\mathcal{H}) \quad \begin{cases} C & \text{is non-empty, closed and convex,} \\ K & \text{has non-empty convex values.} \end{cases}$$

Definition 35. *We say that f and K satisfy the* uniform coercivity condition *(UCC, for short) at $\rho > 0$ if:*

1. *$K(x) \cap B_\rho \neq \emptyset$, for all $x \in C \cap \overline{B}_\rho$;*
2. *for each $x \in \mathrm{Fix}(K)$ with $\|x\| = \rho$, there exists $y \in K(x)$ such that $\|y\| < \rho$ and $f(x,y) \leq 0$.*

Given $\rho > 0$, we define the set $C_\rho := C \cap \overline{B}_\rho$ and the set-valued map $K_\rho : C_\rho \rightrightarrows C_\rho$ as

$$K_\rho(x) := K(x) \cap \overline{B}_\rho. \tag{11}$$

The following proposition, which is an extension of Lemma 2.2 in [24], provides conditions under which we have $\mathrm{QEP}(f, K_\rho) \subset \mathrm{QEP}(f,K)$ for an appropriate $\rho > 0$.

Proposition 25. *We assume that f and K satisfy $(\mathcal{H})$ and (UCC) at some $\rho > 0$. If $x_0 \in \mathrm{Fix}(K_\rho)$ is such that $f(x_0,x_0) \leq 0$, $f(x_0,\cdot)$ is semi-strictly quasi-convex at level 0, and*

$$f(x_0, y) \geq 0 \;\; \textit{for all } y \in K(x_0) \cap B_\rho,$$

then $x_0 \in \mathrm{QEP}(f,K)$.

Proof. If $x_0 \notin \mathrm{QEP}(f, K)$, then there would exist $y_0 \in K(x_0)$ such that $f(x_0, y_0) < 0$. Since $f(x_0, x_0) \leq 0$, by the semi-strictly quasi-convexity of $f(x_0, \cdot)$ at level 0 we have that

$$f(x_0, y_t) < 0, \ \text{ for all } t \in]0, 1[,$$

where $y_t := (1 - t)x_0 + ty_0$. If $\|x_0\| < \rho$, then for t closed enough to 0, we would have that $y_t \in K(x_0) \cap B_\rho$ and $f(x_0, y_t) < 0$, which is a contradiction. Otherwise, if $\|x_0\| = \rho$, then by (UCC) there exists $y_1 \in K(x_0) \cap B_\rho$ such that $f(x_0, y_1) \leq 0$. Then, by proceeding as above we find an element $z_t := (1 - t)y_1 + ty_0$, for small $t \in]0, 1[$, which yields the contradiction $f(x_0, z_t) < 0$. □

Theorem 34. *We assume that f and K satisfy $(\mathcal{H})$ and (UCC) at some large $\rho > 0$, K is lsc,* $\mathrm{Fix}(K)$ *is closed, and $f(x, \cdot)$ is semi-strictly quasi-convex at level* 0 *for every $x \in \mathrm{Fix}(K)$. Moreover, assume that one of the following assertions hold*

1. *f is properly quasi-monotone, has the upper sign property on C, and the set-valued map $G : \mathrm{Fix}(K) \rightrightarrows C$ defined as*

$$G(x) := \{y \in K_\rho(x) : \ f(y, x) > 0\}$$

   is lsc;

2. *$f(x, x) = 0$, for all $x \in \mathrm{Fix}(K)$, and the set-valued map $R : \mathrm{Fix}(K) \rightrightarrows C$ defined as*

$$R(x) := \{y \in K(x) : \ f(x, y) < 0\}$$

   is lsc.

Then $\mathrm{QEP}(f, K)$ *is non-empty.*

Proof. We may assume that ρ is sufficiently large so that $C_\rho \neq \emptyset$. Then, by (UCC),

$$K(x) \cap B_\rho \neq \emptyset, \ \text{ for all } x \in C_\rho,$$

and so the mapping K_ρ defined in (11) has non-empty and convex values. Moreover, due to Lemma 4, the relation above also ensures that K_ρ is lsc.

In case 1, we define the set-valued map $M_1 : C_\rho \rightrightarrows C_\rho$ by

$$M_1(x) := \begin{cases} K_\rho(x), & x \in C_\rho \setminus \mathrm{Fix}(K_\rho) \\ \mathrm{co}(G(x)), & x \in \mathrm{Fix}(K_\rho), \end{cases}$$

which is lsc due to Lemmas 5 and 6. The map M_1 does not have any fixed point. In fact, every fixed point x of M_1 is also a fixed point of K_ρ, and hence a fixed point of $\mathrm{co}(G)$; that is, $x \in \mathrm{co}\{x_i,\ i = 1, \cdots, k\}$ for some $x_i \in G(x)$. Hence, $\min_{i=1\cdots k} f(x_i, x) > 0$ and this contradicts the proper quasi-monotonicity of f.

Now, since the lsc mapping M_1 has convex values and $M_1(C_\rho) \subset \overline{B}_\rho$, by Proposition 6 there exists $x_0 \in C_\rho$ such that $M_1(x_0) = \emptyset$. Thus, $x_0 \in \mathrm{Fix}(K_\rho)$ and $G(x_0) = \emptyset$. To show that $x_0 \in \mathrm{QEP}(f, K_\rho)$, we suppose by contradiction that $f(x_0, y) < 0$ for some $y \in K_\rho(x_0)$. Then the upper sign property yields some $t \in]0, 1[$ such that

$$f(ty + (1 - t)x_0, x_0) > 0;$$

that is, $ty + (1-t)x_0 \in G(x_0)$, a contradiction.

In the second case we proceed similarly. We consider $R_\rho : \mathrm{Fix}(K_\rho) \rightrightarrows C$ defined as $R_\rho(x) := R(x) \cap B_\rho$, which is lsc (Lemma 2) with convex values. Thus, the set-valued map $M_2 : C \rightrightarrows C$ defined as

$$M_2(x) := \begin{cases} K_\rho(x) & x \in C \setminus \mathrm{Fix}(K_\rho) \\ R_\rho(x) & x \in \mathrm{Fix}(K_\rho) \end{cases}$$

is lsc with convex values. If M_2 is non-empty valued, then again by Proposition 6 there exists $x_0 \in M_2(x_0)$, this means $x_0 \in \mathrm{Fix}(K_\rho)$ and $x_0 \in R_\rho(x_0)$, which in turn implies $f(x_0, x_0) < 0$. So, we get a contradiction. Hence, there exists $x_0 \in C$ such that $M_2(x_0) = \emptyset$. Thus, $x_0 \in \mathrm{Fix}(K_\rho)$ and $R_\rho(x_0) = \emptyset$, i.e.

$$f(x_0, y) \geq 0, \text{ for all } y \in K(x_0) \cap B_\rho.$$

Finally, and in both cases, since $f(x_0, x_0) \leq 0$, by semi-strictly quasi-convexity of f at level 0, we infer using Proposition 25 that $x_0 \in \mathrm{QEP}(f, K)$. □

Since (UCC) holds at a sufficiently large ρ when C is compact, we obtain the following result.

Corollary 16. *Let C be a non-empty, compact and convex subset of $\mathbb{R}^n$ and assume that f is properly quasi-monotone, semi-strictly quasi-convex at level* 0 *in the second argument, and has the upper sign property. If the set*

$$\{y \in C : \ f(x, y) \leq 0\}$$

is closed, for each $x \in C$, then $\mathrm{EP}(f, C) \neq \emptyset$.

Proof. First, the constant set-valued map $K(x) := C$, $x \in C$, is obviously lsc and has convex and non-empty values. Also, we have that $\mathrm{Fix}(K) = C$, which is obviously closed. Then condition (UCC) trivially holds, as well as hypothesis ($\mathcal{H}$). According to Theorem 34, it suffices to show that the mapping G, defined in Theorem 34, is lsc. Indeed, by the current assumption, for each $y \in C$, the fiber

$$G^{-1}(y) = \{x \in C : \ f(y, x) > 0\},$$

is open, and this easily implies the lower semi-continuity of G. □ □

Corollary 16 is given in Proposition 2.1 in [24], where instead of the upper sign property of f, the authors assume that f is quasi-convex in the second argument and $f_D = 0$, as well as the *upper sign continuity* (see [24]) of f; that is,

$$\inf_{t \in]0,1[} f(tx + (1-t)y, y) \geq 0 \ \Rightarrow \ f(x, y) \geq 0, \quad \forall x, \, y \in C.$$

It is known that the last three conditions ensure the upper sign property of f (see Lemma 3 in [31]).

Remark 26. *It is worth recalling that, instead of the semi-strict quasi-convexity at level* 0 *of the function* f *in Corollary 16, Proposition 2.1 in [24] uses the so-called* sign preserving property*; that is, for all* $x,\ y,\ z \in C$,

$$(f(x,y) = 0 \ \wedge \ f(x,z) < 0) \ \Rightarrow \ f(x, ty + (1-t)z) < 0, \ \text{for all } t \in]0,1[\ .$$

We observe that, under the quasi-convexity of the functions $f(x,\cdot)$, $x \in C$, *both the sign preserving property and the semi-strict quasi-convexity at level* 0 *are equivalent.*

Corollary 17 (Theorem 4.5 in [11])**.** *Let* $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *be a function,* C *be a convex, compact and non-empty subset of* $\mathbb{R}^n$, *and* $K : C \rightrightarrows C$ *be a set-valued map. Suppose that the following properties hold*

1. K *is closed and lsc with convex values, and* $int(K(x)) \neq \emptyset$, *for all* $x \in C$;

2. f *is properly quasi-monotone;*

3. f *is semi-strictly quasi-convex and lower semi-continuous with respect to its second argument;*

4. *for all* $x, y \in \mathbb{R}^n$ *and all sequence* $(y_k)_k \subset \mathbb{R}^n$ *converging to* y, *the following implication holds*
$$\liminf_{k\to+\infty} f(y_k, x) \leq 0 \ \Rightarrow \ f(y,x) \leq 0,$$

5. f *has the upper sign property.*

Then, $\mathrm{QEP}(f,K)$ *is non-empty.*

Proof. Since C is compact, the set-valued map G in Theorem 34 can be described by $G(x) = \{y \in K(x) : \ f(y,x) > 0\}$ for every $x \in \mathrm{Fix}(K)$. We can prove the lower semi-continuity of G following the same steps of the proof of Corollary 7 in [35], and thus the conclusion follows applying Theorem 34. □

Theorem 35. *We assume that* f *and* K *satisfy* $(\mathcal{H})$ *and (UCC) at some large* $\rho > 0$, K *is closed and*

(i) $f(\cdot, y)$ *is upper semi-continuous, for all* $y \in C$,

(ii) $f(x,\cdot)$ *is quasi-convex, for all* $x \in C$,

(iii) the set $V = \{x \in C_\rho : \ \inf_{y\in K_\rho(x)} f(x,y) < 0\}$ *is open in* C_ρ,

(iv) $f_D = 0$,

(v) for each $x \in \mathrm{Fix}(K)$, $f(x,\cdot)$ *is semi-strictly quasi-convex at level 0.*

Then $\mathrm{QEP}(f,K)$ *is non-empty.*

Proof. Consider the set-valued map $T : C_\rho \rightrightarrows C_\rho$ defined as

$$T(x) := \{y \in C_\rho : \ f(x,y) < 0\}.$$

Clearly, $V = \{x \in C_\rho : \ T(x) \cap K_\rho(x) \neq \emptyset\}$ and the set-valued map K_ρ is closed since $\mathrm{gph}(K_\rho) = \mathrm{gph}(K) \cap (C_\rho \times C_\rho)$. The (UCC) at ρ and (i) imply that K_ρ is upper semi-continuous with convex, compact and non-empty values. Moreover, from (ii), (iii) and (iv), we deduce that T is convex-valued with open fibers and $\text{Fix}(T) = \emptyset$. Hence, by Theorem 7 there exists $x \in \text{Fix}(K_\rho)$ such that $K_\rho(x) \cap T(x) = \emptyset$, that means $x \in \text{QEP}(f, K_\rho)$. The conclusion follows from applying Proposition 25. □

Our Theorem 35 has some similarities with Theorem 3 in [115], but the set of assumptions in both results differ in two important aspects. Firstly, in [115] it was assumed that f is *0-diagonally convex* on the second variable, while in our case we assume that f is quasi-convex in its second argument and that f vanishes on the diagonal of $C \times C$. Examples in [127] show that these assumptions are not comparable in general. Secondly, there is a difference on the coerciveness conditions. In [115] the authors considered a quite restrictive coerciveness condition, which in particular imply that in a non-empty set the images of K are compact.

The following corollary is related to [91, Theorem 3], where a slightly less general kind of "quasi-equilibrium problem" was considered. Our condition 3 in the corollary is a consequence of this restriction.

Corollary 18. *Let C be a compact, convex and non-empty subset of $\mathbb{R}^n$, let $K, K_C : C \rightrightarrows \mathbb{R}^n$ be set-valued maps such that $K_C(x) = K(x) \cap C$, and $f : C \times C \to \mathbb{R}$ be a function. If the following assumptions hold*

1. *K_C is usc and lsc with convex, compact and non-empty values,*
2. *f is continuous and $f(x, \cdot)$ is convex, for all $x \in C$,*
3. *$f(x, x) = 0$, for all $x \in C$,*
4. *for each $x \in \text{Fix}(K_C)$ there exists $y \in K_C(x)$ such that $f(x, y) \leq 0$ and $]y, z] \cap K_C(x) \neq \emptyset$, for all $z \in K(x) \setminus K_C(x)$;*

then $\text{QEP}(f, K)$ *is non-empty.*

Proof. The set $\text{QEP}(f, K_C)$ is non-empty, due to Theorem 35. The result follows since Assumption 4 implies $\text{QEP}(f, K_C) \subset \text{QEP}(f, K)$. □

To end this section we now provide an existence result for quasi-equilibrium problems that is based on the notion of fip*.

Theorem 36. *Let $f : X \times X \to \mathbb{R}$ be a bi-function, C be a non-empty, convex and compact subset of X and $K : C \rightrightarrows C$ be a set-valued map. If the following assumptions hold:*

1. *the map K is closed and lsc, with convex and non-empty values,*
2. *f has both the upper sign property and the fip* on C,*
3. *the set $M = \{(x, y) \in C \times C : \; f(x, y) \leq 0\}$ is closed,*
4. *for each $x \in C$, the set F_x (defined as in (9)) is convex;*

then the quasi-equilibrium problem admits at least one solution.

Proof. We define $g : X \times X \to \mathbb{R} \cup \{+\infty\}$ as

$$g(x,y) := \chi_K(x,y) = \begin{cases} 0, & y \in K(x) \\ +\infty, & \text{otherwise} \end{cases} .$$

Since K is closed, we deduce that g is lower semi-continuous. Moreover, as K is convex valued, the bi-function g is convex with respect to its second argument. So, for each $x, w \in C$, we define the set

$$G_x(w) := \{y \in C : \; f(w,y) + g(x,y) \leq g(x,w)\}.$$

On the one hand if $w \notin K(x)$, then $G_x(w) = C$. On the other hand, if $w \in K(x)$ we have $G_x(w) = F_w \cap K(x)$. Thus, $G_x(w)$ is a compact, convex and non-empty subset of C. Since f has the fip* on C, for any $w_1, \ldots, w_n \in C$, we have

$$\bigcap_{i=1}^{m} G_x(w_i) \neq \emptyset.$$

Indeed, put $J := \{i \in \{1, \ldots, m\} : \; w_i \in K(x)\}$. If $J = \emptyset$, then $\bigcap_{i=1}^{m} G_x(w_i) = C$. Else,

$$\bigcap_{i=1}^{m} G_x(w_i) = \bigcap_{i \in J} G_x(w_i).$$

Thus, there exists $z \in \mathrm{co}(\{w_i\}_{i \in J}) \subset K(x)$ such that

$$\max_{i \in J} f(w_i, z) \leq 0.$$

Hence $z \in \bigcap_{i \in J} G_x(w_i)$.

So, for each $x \in C$, the family of sets $\{G_x(w)\}_{w \in C}$ has the finite intersection property. Since each $G_x(w)$ is compact, we have $\bigcap_{w \in C} G_x(w) \neq \emptyset$. Thus, the set-valued map $S : C \rightrightarrows C$ defined by

$$S(x) := \bigcap_{w \in C} G_x(w)$$

is compact, convex and non-empty valued. We will show now that S is closed. Indeed, let $(x_i, y_i)_{i \in I}$ be a net in the graph of S such that it converges at (x,y). For all $i \in I$

$$f(w, y_i) + g(x_i, y_i) \leq g(x_i, w) \text{ for all } w \in C.$$

Taking $w \in K(x_i)$ we deduce $y_i \in K(x_i)$, which in turn implies $y \in K(x)$. As K is lower semi-continuous, for all $w \in K(x)$ there exists a subnet $(x_{\varphi(j)})_{j \in J}$ of $(x_i)_{i \in I}$ and a net $(w_j)_{j \in J}$ converging to w such that $w_j \in K(x_{\varphi(j)})$ for all $j \in J$. So $f(w_j, y_{\varphi(j)}) \leq 0$ for all $j \in J$. By the closeness of set M, one has $f(w,y) \leq 0$. So, it holds

$$f(w,y) + g(x,y) \leq g(x,w) \text{ for all } w \in C.$$

Thus, $y \in S(x)$. Additionally, as $S(C)$ is relatively compact, S is upper semi-continuous. Thus, S admits at least a fixed point, due to Theorem 7 and Remark 4, that means there exists $x_0 \in C$ such that

$$f(w, x_0) + g(x_0, x_0) \leq g(x_0, w) \text{ for all } w \in C.$$

Taking $w \in K(x_0)$ in the previous inequality we have $x_0 \in K(x_0)$. Therefore, $x_0 \in \mathrm{MQEP}(f, K)$. Thus, by Proposition 3.1 in [11], x_0 is a solution of the quasi-equilibrium problem. □

As a consequence of Theorem 36 we recover the following result.

Corollary 19 (Proposition 4.5 in [10])**.** *Let $h : X \to \mathbb{R}$ be a continuous and quasi-convex function, C be a convex, compact and non-empty subset of X and $K : C \rightrightarrows C$ be a closed and lower semi-continuous set-valued map with convex and non-empty values. Then there exists $x \in \mathrm{Fix}(K)$ such that*

$$h(x) \leq h(y), \text{ for all } y \in K(x).$$

Proof. Clearly the bi-function f defined as in Example 24 has the fip* on C and it is continuous and quasi-convex in its second argument. Moreover, by the part 2 of [32, Proposition 6.2], it has the upper sign property. Thus, Theorem 36 guarantees the existence of a point $x \in \mathrm{QEP}(f, K)$, which is equivalent to $x \in K(x)$ and $h(y) \geq h(x)$, for all $y \in K(x)$. □

The problem associated to the previous corollary is well-known in the literature as *quasi-optimization*.

Remark 27. *Theorem 36 is strongly related with Theorem 4.5 in [11] and Theorem 4.3 in [50]. However these results are established under generalized monotonicity and quasi-convexity, which are stronger than the finite intersection property.*

IV.4 Applications

In this section, we consider applications on the study of existence of solutions for two well-known problems: (i) the quasi-variational inequality problem, and (ii) the generalized Nash equilibrium problem.

IV.4.1 Quasi-Variational Inequality Problem

Given a subset C of $\mathbb{R}^n$ and two set-valued maps $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and $K : C \rightrightarrows C$, the set $\mathrm{QVI}(T, K)$ denotes the solution set of the *quasi-variational inequality problem* associated to T and K,

$$\{x \in C : \; x \in K(x), \exists x^* \in T(x) \text{ such that } \langle x^*, y - x\rangle \geq 0, \; \forall y \in K(x)\}.$$

We say that T and K satisfy the *uniform coerciveness condition* at ρ if the following two conditions hold:

1. $K(x) \cap B_\rho \neq \emptyset$, for all $x \in C \cap \overline{B}_\rho$,
2. for each $x \in \mathrm{Fix}\, K$ such that $\|x\| = \rho$ there exists $y \in K(x)$ with $\|y\| < \rho$ such that $\langle x^*, y - x\rangle \leq 0$ for every $x^* \in T(x)$.

Now, we consider the bi-function $f_T : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$ defined as

$$f_T(x, y) := \sup_{x^* \in T(x)} \langle x^*, y - x\rangle. \tag{12}$$

The next Lemma relates the quasi-variational inequality with the quasi-equilibrium problem whose bi-function is f_T. We observe though that a converse construction, from a quasi-equilibrium problem to a quasi-variational inequality problem, has been done in [13] under quite general conditions.

Lemma 17. *Given any $\rho > 0$, T and K satisfy the uniform coerciveness condition at ρ if and only if f_T and K satisfies the (UCC) at ρ. Moreover, if T has non-empty and compact values then* $\mathrm{QEP}(f_T, K) = \mathrm{QVI}(T, K)$.

Proof. Direct from the definition of f_T. □

As a direct consequence of Lemma 17 and Theorem 34 we obtain the following existence result for quasi-variational inequality problems.

Theorem 37. *Let C be a closed, convex and non-empty subset of $\mathbb{R}^n$, and $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, $K : C \rightrightarrows C$ be two set-valued maps. Assume that (T, K) satisfies the uniform coerciveness condition at $\rho > 0$ sufficiently large, and that the following conditions are satisfied:*

1. *T has compact and non-empty values,*
2. *T is properly quasi-monotone on C i.e., for all $x_1, \ldots, x_m \in C$ and any $x \in \mathrm{co}(\{x_1, \ldots, x_m\})$, there exists i such that*
$$\langle x_i^*, x - x_i \rangle \leq 0, \text{ for all } x_i^* \in T(x_i),$$
3. *T is upper sign-continuous on C, that means for all $x, y \in C$, the following implication holds*
$$\left(\forall t \in]0,1[, \inf_{x_t^* \in T(x_t)} \langle x_t^*, y - x \rangle \geq 0\right) \Rightarrow \sup_{x^* \in T(x)} \langle x^*, y - x \rangle \geq 0,$$
where $x_t = tx + (1-t)y$.,
4. *K is lsc with convex and non-empty values*
5. *The set $\mathrm{Fix}(K)$ is closed and the set-valued map $G : \mathrm{Fix}(K) \rightrightarrows C$ defined as*
$$G(x) := \left\{ y \in K(x) \cap \overline{B}_\rho : \sup_{x^* \in T(x)} \langle x^*, y - x \rangle > 0 \right\}$$
is lsc.

Then, $\mathrm{QVI}(T, K)$ *is non-empty.*

Proof. Clearly f_T is properly quasi-monotone and has the upper sign property. Therefore, the result follows from the fact $\mathrm{QVI}(T, K) = \mathrm{QEP}(f_T, K)$ and Theorem 34. □

Remark 28. *A few remarks about Theorem 37:*

1. *The previous result is not a consequence of Theorem 1 in [16], because T here is properly quasi-monotone (not pseudo-monotone) and the closedness of K is relaxed to the closedness of* $\mathrm{Fix}(K)$. *Theorem 3 in [16] proposes an existence result under quasi-monotonicity, that means for all (x, x^*) and (y, y^*) in the graph of T the following implication holds*
$$\langle x^*, y - x \rangle > 0 \Rightarrow \langle y^*, y - x \rangle \geq 0,$$
but in this case it needs more regularity assumptions on the constraint map.
2. *The fourth assumption in Theorem 37 holds, for instance, when the map K is closed and the set*
$$\left\{ (x, y) \in C \times C : \sup_{x^* \in T(x)} \langle x^*, y - x \rangle \leq 0 \right\}$$
is closed.

IV.4.2 Generalized Nash Equilibrium Problem

Here we use the notation of Chapter III.1. Recall that given a rival strategy x_{-j}, player j chooses a strategy x_j such that it solves the following optimization problem

$$\min_{x_j} f_j(x_j, x_{-j}), \text{ subject to } x_j \in K_j(x_{-j}), \tag{13}$$

for any given strategy vector x_{-j} of the rival players. If we denote by $S_j(x_{-j})$ the solution set of problem (13) a *generalized Nash equilibrium* is a vector $\hat{x}$ such that $\hat{x}_j \in S_j(\hat{x}_{-j})$, for any j.

We can associate to a GNEP, the following bi-function $f^{NI} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, defined by

$$f^{NI}(x, y) := \sum_{j=1}^{p} \{f_j(y_j, x_{-j}) - f_j(x_j, x_{-j})\},$$

which is called Nikaidô-Isoda function and was introduced in [90]. Additionally, we consider the set-valued map $K : C \rightrightarrows C$ defined as

$$K(x) := \prod_{j=1}^{p} K_j(x_{-j}).$$

Lemma 18. *A vector $\hat{x}$ is a solution of the GNEP if and only if, $\hat{x} \in \mathrm{QEP}(f^{NI}, K)$.*

A GNEP satisfies the *coerciveness condition* at $\rho > 0$ if

1. $K(x) \cap B_\rho \neq \emptyset$, for all $x \in C \cap \overline{B}_\rho$;
2. for each $x \in \mathrm{Fix}(K)$, such that $\|x\| = \rho$ there exists $y \in K(x)$ with $\|y\| < \rho$ such that $f_j(y_j, x_{-j}) \leq f_j(x)$ for each j.

If we consider in $\mathbb{R}^n$ the product norm given by the maximum of the norms of all the $\mathbb{R}^{n_j}$, then the above condition is equivalent to that for each j

1. $K_j(x_{-j}) \cap B_{\mathbb{R}^{n_j},\rho} \neq \emptyset$, for all $x \in C \cap \overline{B}_\rho$;
2. for each $x \in \mathrm{Fix}(K_\rho)$, if $\|x_j\|_{\mathbb{R}^{n_j}} = \rho$ then there exists $y_j \in K(x_{-j})$ with $\|y_j\|_{\mathbb{R}^{n_j}} < \rho$ such that $f_j(y_j, x_{-j}) \leq f_j(x)$.

Lemma 19. *If the GNEP satisfies the coerciveness condition at $\rho > 0$, then the pair f^{NI} and K satisfies the (UCC) at ρ.*

Proof. It is enough to see that if for each j we have $f_j(y_j, x_{-j}) \leq f_j(x)$, then

$$f^{NI}(x, y) = \sum_{j=1}^{p} f_j(y_j, x_{-j}) - f_j(x) \leq 0.$$

□

Thanks to Lemmas 18 and 19, we have the following result on the existence of solutions of a GNEP, which is a direct consequence of Theorems 34 and 35.

Theorem 38. *For any $j \in \{1, 2, ..., p\}$, let C_j be a non-empty, closed and convex subset of $\mathbb{R}^{n_j}$, $f_j : \mathbb{R}^n \to \mathbb{R}$ be a continuous function and $K_j : C_{-j} \rightrightarrows C_j$ be a set-valued map. Assume that the GNEP satisfies the coerciveness condition at ρ, that for each j, f_j is convex with respect to the x_j variable, and at least one of the following assumptions hold:*

A1 (a) *The set* $\mathrm{Fix}(K)$ *is closed,*

(b) *for each j, the map K_j is lower semi-continuous with non-empty and convex values.*

A2 (a) *for each j, the map K_j is closed with non-empty convex values,*

(b) *the set*

$$N = \left\{ x \in C_\rho : \inf_{y \in K_\rho(x)} \sum_j f_j(y_j, x_{-j}) < \sum_j f_j(x) \right\}$$

is open in C_ρ.

Then the GNEP admits a solution.

Proof. It is clear that f^{NI} is continuous and convex in its second argument and the map K is closed with convex and non-empty values. By Lemma 19, we have that f^{NI} and K satisfy the UCC at ρ. In case $A1$, the map K is lsc with convex and non-empty values. Hence, the set-valued map R defined in the second case of Theorem 34 is also lsc with convex values. So, the result follows from Theorem 34 and Lemma 18.

In case $A2$ it holds

$$N = \left\{ x \in C_\rho : \inf_{y \in K_\rho(x)} f^{NI}(x, y) < 0 \right\}.$$

Hence, the result follows from Theorem 35 and Lemma 18. □

The previous result is related to Theorem 5 in [16]. However, we notice that in assumption $A1$ the constraint set-valued maps K_j are not necessarily closed, while for $A2$ the maps K_j are not necessarily lsc. Moreover, none of the cases assume any differentiability, and the images of the constraint maps K_j are allowed to have empty interior. Finally, their 'coerciveness condition' is somehow weaker than ours. In fact, $f_j(y_j, x_{-j}) \leq f_j(x_j, x_{-j})$ clearly implies their condition $\langle \nabla_{x_j} f_j(x), x_j - y_j \rangle \geq 0$, due to the convexity assumption, while the converse implication is not true in general.

We can also use the concept of finite intersection property in this context. Les us consider the bi-function $f_0 : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ given by

$$f_0(x, y) = -f^{NI}(y, x) \tag{14}$$

We consider also the important situation of joint constraints, introduced by Rosen in 1965 (see [104]) and has recently been considered in [12, 49, 86]. This case is described with a non-empty subset C of X by letting the constraint set-valued maps be defined as

$$K_j(x_{-j}) := \{z_j \in X_j : \ (z_j, x_{-j}) \in C\}, \tag{15}$$

for any j and $x = (x_j, x_{-j}) \in C$. The following result (similar to Lemma 18) states that every solution of the Minty equilibrium problem is a solution of the generalized Nash equilibrium problem in the joint case.

Lemma 20. *Let us assume, for any j the subset $K_j(x_{-j})$ is defined as in* (15)*. Then every solution of* $\mathrm{MEP}(f_0, C)$ *is a generalized Nash equilibrium.*

Proof. Let $\hat{x}$ be an element of $\mathrm{MEP}(f_0, C)$. For each j and any $y_j \in K_j(\hat{x}_{-j})$ we have

$$f_j(\hat{x}) - f_j(y_j, \hat{x}_{-j}) = f_0(y, \hat{x}) \leq 0,$$

where $y = (y_j, \hat{x}_{-j}) \in C$, which in turn implies $f_j(\hat{x}) \leq f_j(y_j, \hat{x}_{-j})$. The result follows. □

Corollary 20. *Assume that C is compact and non-empty and for any j the subset $K_j(x_{-j})$ is defined as in* (15)*. If f_0 defined as in* (14) *has the fip on C and the set $F_x = \{y \in C : f_0(x, y) \leq 0\}$ is closed for all $x \in C$, then there exists a generalized Nash equilibrium.*

Proof. It follows from Lemma 16 and Lemma 20. □

We have that [69, Corollary 4.6] is a direct consequence of Corollary 20, thanks to Proposition 18 . The next result establishes sufficient conditions to guarantee the fip* of f_0.

Proposition 26. *Assume that each X_j is a topological vector space and the set C is convex. If each objective function is continuous and convex with respect to the variable of its player, then the bifunction f_0 defined as in* (14) *has fip* on C.*

Proof. It is clear $-f_0(\cdot, y)$ is (quasi-) convex and f_0 vanishes on the diagonal of $X \times X$. By Proposition 1.1 in [23], we deduce f_0 is properly quasi-monotone. Since f_0 is continuous, the result follows from Proposition 23. □

Remark 29. *An important instance (see [69]) where f_0 is cyclically quasi-monotone is when each objective function f_j has separable variables, that is, it can be written as $f_j(x_j, x_{-j}) := g_j(x_j) + h_j(x_{-j})$. Indeed, this follows from writing*

$$f_0(x, y) = \sum_{j=1}^{p} g_j(y_j) - g_j(x_j) = \varphi(y) - \varphi(x),$$

where $\varphi(z) = \sum_{j=1}^{p} g_j(z_j)$, and Example 24.

Bibliography

[1] E. Allevi, D. Aussel, and R. Riccardi. On an equilibrium problem with complementarity constraints formulation of pay-as-clear electricity market with demand elasticity. *J. Global Optim.*, 70(2):329–346, 2018. (Cited on pages 83 and 84.)

[2] J.-P. Aubin and H. Frankowska. *Set-valued analysis.* Springer Science & Business Media, 2009. (Cited on page 12.)

[3] D. Aussel. New developments in quasiconvex optimization. In *Fixed point theory, variational analysis, and optimization*, pages 171–205. CRC Press, Boca Raton, FL, 2014. (Cited on page 88.)

[4] D. Aussel, P. Bendotti, and M. Pištěk. Nash equilibrium in pay-as-bid electricity market : Part 1 - existence and characterisation. *Optimization*, 66(6):1013–1025, 2017. (Cited on pages 83 and 84.)

[5] D. Aussel, P. Bendotti, and M. Pištěk. Nash equilibrium in pay-as-bid electricity market : Part 2 - best response of producer. *Optimization*, 66(6):1027–1053, 2017. (Cited on pages 83 and 84.)

[6] D. Aussel, G. Bouza, S. Dempe, and S. Lepaul. Genericity analysis of multi-leader-follower games. 2019. preprint. (Cited on page 102.)

[7] D. Aussel, L. Brotcorne, S. Lepaul, and L. von Niederhäusern. A trilevel model for best response in energy demand-side management. *Eur. J. Oper. Res. Games*, 281(2):299–315, 2020. (Cited on page 102.)

[8] D. Aussel, M. Červinka, and M. Marechal. Deregulated electricity markets with thermal losses and production bounds: models and optimality conditions. *RAIRO-Oper. Res.*, 50(1):19–38, 2016. (Cited on pages 83 and 101.)

[9] D. Aussel, R. Correa, and M. Marechal. Electricity spot market with transmission losses. *J. Ind. Manag. Optim.*, 9(2):275–290, 2013. (Cited on page 83.)

[10] D. Aussel and J. Cotrina. Quasimonotone quasivariational inequalities: existence results and applications. *J. Optim. Theory Appl.*, 158:637–652, 2013. (Cited on page 118.)

[11] D. Aussel, J. Cotrina, and A. Iusem. An existence result for quasi-equilibrium problems. *J. Convex Anal.*, 24:55–66, 2017. (Cited on pages 109, 115, 117 and 118.)

[12] D. Aussel and J. Dutta. Generalized nash equilibrium problem, variational inequality and quasiconvexity. *Oper. Res. Lett.*, 36(4):461–464, 2008. [Addendum: Oper. Res. Lett. 42 (2014), 398]. (Cited on pages 56, 88 and 121.)

[13] D. Aussel, J. Dutta, and T. Pandit. About the links between equilibrium problems and variational inequalities. In *Mathematical Programming and Game Theory*, pages 115–130. Springer, 2018. (Cited on page 118.)

[14] D. Aussel and N. Hadjisavvas. Adjusted sublevel sets, normal operator, and quasi-convex programming. *SIAM J. Optim.*, 16(2):358–367, 2005. (Cited on page 88.)

[15] D. Aussel, S. Lepaul, and L. von Niederhäusern. A multi-leader-follower game for energy demand-side management. 2019. under revision. (Cited on page 102.)

[16] D. Aussel and A. Sultana. Quasi-variational inequality problems with non-compact valued constraint maps. *J. Math. Anal. Appl.*, 456(2):1482–1494, 2017. (Cited on pages 119 and 121.)

[17] D. Aussel and A. Svensson. Some remarks about existence of equilibria, and the validity of the epcc-reformulation for multi-leader-follower games. *J. Nonlinear Convex Anal.*, 19(7):1141–1162, 2018. (Cited on pages 55, 81, 85, 86, 91 and 98.)

[18] D. Aussel and A. Svensson. Is pessimistic bilevel programming a special case of a mathematical program with complementarity constraints? *J. Optim. Theory Appl.*, 181(2):504–520, 2019. (Cited on pages 65, 74, 99 and 101.)

[19] D. Aussel and A. Svensson. Towards tractable constraint qualifications for parametric optimisation problems and applications to generalised Nash games. *J. Optim. Theory Appl.*, 182(1):404–416, 2019. (Cited on pages 8, 55, 90 and 100.)

[20] D. Aussel and A. Svensson. A short state-of-the-art on multi-leader-follower games. In *Bilevel optimization: advances and next challenges*. chapter in forthcoming book by Springer, to appear. (Cited on page 81.)

[21] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer. *Non-linear parametric optimization*. Springer, 1982. (Cited on pages 12, 13, 14, 15 and 86.)

[22] T. Q. Bao, P. Gupta, and B. S. Mordukhovich. Suboptimality conditions for mathematical programs with equilibrium constraints. *Taiwanese J. Math.*, 12(9):2569–2592, 2008. (Cited on pages 35, 39, 41 and 42.)

[23] M. Bianchi and R. Pini. A note on equilibrium problems with properly quasimonotone bifunctions. *J. Global Optim.*, 20:67–76, 2001. (Cited on pages 111 and 122.)

[24] M. Bianchi and R. Pini. Coercivity Conditions for Equilibrium Problems. *J. Optim. Theory Appl.*, 124(1):79–92, Jan 2005. (Cited on pages 112, 114 and 115.)

[25] J. M. Borwein. A note on ε-subgradients and maximal monotonicity. *Pacific J. Math.*, 103(2):307–314, 1982. (Cited on page 23.)

[26] J. M. Borwein and Q. J. Zhu. Viscosity solutions and viscosity subderivatives in smooth banach spaces with applications to metric regularity. *SIAM J. Control Optim.*, 34(5):1568–1591, 1996. (Cited on page 39.)

[27] J. M. Borwein and Q. J. Zhu. A survey of subdifferential calculus with applications. *Nonlinear Anal.*, 38:687–773, 1999. (Cited on page 44.)

[28] A. Brøndsted and R. T. Rockafellar. On the subdifferentiability of convex functions. *Proc. Amer. Math. Soc.*, pages 605–611, 1965. (Cited on page 23.)

[29] H. I. Calvete and C. Galé. Linear bilevel multi-follower programming with independent followers. *J. Global Optim.*, 39(3):409–417, 2007. (Cited on page 84.)

[30] K. Cao Van, D. Salas, L. Montastruc, and D. Aussel. Optimal design of exchange networks with blind inputs - part 2 : applications to eco-industrial parks. *preprint*, 2019. (Cited on page 83.)

[31] M. Castellani and M. Giuli. Refinements of existence results for relaxed quasimonotone equilibrium problems. *J. Global Optim.*, 57:1213–1227, 2013. (Cited on page 114.)

[32] J. Cotrina and Y. García. Equilibrium problems: Existence results and Applications. *Set-Valued Var. Anal*, 26:159–177, 2018. (Cited on pages 109 and 118.)

[33] J. Cotrina, A. Hantoute, and A. Svensson. Existence for quasi-equilibrium problems on unbounded sets. preprint. (Cited on page 105.)

[34] J. Cotrina and A. Svensson. Finite intersection property for bifunctions and existence for quasi-equilibrium problems. preprint. (Cited on page 105.)

[35] J. Cotrina and J. Zúñiga. Quasi-equilibrium problems with non-self constraint map. *J. Global Optim.*, 2019. (Cited on pages 16, 109 and 115.)

[36] A. Daniilidis and N. Hadjisavvas. On the subdifferentials of quasiconvex and pseudoconvex functions and cyclic monotonicity. *J. Math. Anal. Appl.*, 237:30–42, 1999. (Cited on page 109.)

[37] S. M. Dassanayaka. *Methods of variational analysis in pessimistic bilevel programming.* PhD thesis, Wayne State University, 2010. (Cited on page 65.)

[38] D. De Wolf and Y. Smeers. A stochastic version of a stackelberg-nash-cournot equilibrium model. *Management Science*, 43(2):190–197, 1997. (Cited on page 91.)

[39] V. DeMiguel, M. P. Friedlander, F. J. Nogales, and S. Scholtes. A two-sided relaxation scheme for mathematical programs with equilibrium constraints. *SIAM J. Optim.*, 16(2):587–609, 2005. (Cited on page 91.)

[40] S. Dempe. *Foundations of bilevel programming.* Springer Science & Business Media, 2002. (Cited on pages 14, 65, 69, 72 and 86.)

[41] S. Dempe and J. Dutta. Is bilevel programming a special case of a mathematical program with complementarity constraints? *Math. Program.*, 131(1-2):37–48, 2012. (Cited on pages 65, 66, 73, 74, 79 and 99.)

[42] S. Dempe, G. Luo, and S. Franke. Pessimistic bilevel linear optimization. *preprint*, 2016. (Cited on page 86.)

[43] S. Dempe, B. S. Mordukhovich, and A. B. Zemkoho. Sensitivity analysis for two-level value functions with applications to bilevel programming. *SIAM J. Optim.*, 22(4):1309–1343, 2012. (Cited on pages 36, 48, 65 and 69.)

[44] S. Dempe, B. S. Mordukhovich, and A. B. Zemkoho. Necessary optimality conditions in pessimistic bilevel programming. *Optimization*, 63(4):505–533, 2014. (Cited on pages 65, 69 and 72.)

[45] S. Dolecki, G. Salinetti, and R. J.-B. Wets. Convergence of functions: equi-semicontinuity. *Trans. Amer. Math. Soc.*, 276(1):409–430, 1983. (Cited on page 47.)

[46] A. Ehrenmann. *Equilibrium problems with equilibrium constraints and their application to electricity markets*. PhD thesis, Fitzwilliam College, 2004. (Cited on pages 4 and 89.)

[47] J. F. Escobar and A. Jofré. Monopolistic competition in electricity networks with resistance losses. *Econom. Theory*, 44(1):101–121, 2010. (Cited on pages 83, 84 and 85.)

[48] Juan F Escobar and Alejandro Jofré. Equilibrium analysis of electricity auctions. *Department of Economics Stanford University*, 2008. preprint. (Cited on page 83.)

[49] F. Facchinei and C. Kanzow. Generalized nash equilibrium problems. *4OR*, 5(3):173–210, 2007. (Cited on pages 55, 56 and 121.)

[50] F. Flores-Bazán. Existence Theorems for Generalized Noncoercive Equilibrium Problems: The Quasi-Convex Case. *SIAM J. Optim.*, 11(3):675–690, 2001. (Cited on page 118.)

[51] M. Fukushima and G.-H. Lin. Smoothing methods for mathematical programs with equilibrium constraints. In *International Conference on Informatics Research for Development of Knowledge Society Infrastructure*, pages 206–213. IEEE, 2004. (Cited on page 89.)

[52] J. Guillerme. Nash equilibrium for set-valued maps. *J. Math. Anal. Appl.*, 187(3):705–715, 1994. (Cited on page 92.)

[53] L. Guo, G.-H. Lin, and J. J. Ye. Solving mathematical programs with equilibrium constraints. *J. Optim. Theory Appl.*, 166(1):234–256, 2015. (Cited on page 89.)

[54] A. Hantoute and A. Svensson. A general representation of δ-normal sets to sublevels of convex functions. *Set-Valued Var. Anal.*, 25(4):651–678, 2017. (Cited on pages 19, 21 and 22.)

[55] P. T. Harker and J.-S. Pang. Existence of optimal solutions to mathematical programs with equilibrium constraints. *Oper. Res. Lett.*, 7(2):61–64, 1988. (Cited on page 88.)

[56] R. Henrion, J. Outrata, and T. Surowiec. Analysis of M-stationary points to an EPEC modeling oligopolistic competition in an electricity spot market. *ESAIM Control Optim. Calc. Var.*, 18(2):295–317, 2012. (Cited on pages 83 and 84.)

[57] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms I: Fundamentals*, volume 305. Springer science & business media, 2013. (Cited on pages 70 and 77.)

[58] J.-B. Hiriart-Urruty, M. Moussaoui, A. Seeger, and M. Volle. Subdifferential calculus without qualification conditions, using approximate subdifferentials: a survey. *Nonlinear Anal.*, 24(12):1727–1754, 1995. (Cited on pages 28 and 31.)

[59] J.-B. Hiriart-Urruty and R. R. Phelps. Subdifferential calculus using ϵ-subdifferentials. *J. Funct. Anal.*, 118(1):154–166, 1993. (Cited on page 24.)

[60] M. Hu and M. Fukushima. Existence, uniqueness, and computation of robust nash equilibria in a class of multi-leader-follower games. *SIAM J. Optim*, 23(2):894–916, 2013. (Cited on pages 93 and 98.)

[61] M. Hu and M. Fukushima. Multi-leader-follower games: models, methods and applications. *J. Oper. Res. Soc. Japan*, 58(1):1–23, 2015. (Cited on pages 94 and 98.)

[62] X. Hu and D. Ralph. Using EPECs to model bilevel games in restructured electricity markets with locational prices. *Oper. Res.*, 55(5):809–827, 2007. (Cited on pages 83 and 84.)

[63] T. Ichiishi and M. Quinzii. Decentralization for the core of a production economy with increasing return. *Internat. Econom. Rev.*, 24(2):397–412, 1983. (Cited on pages 55 and 93.)

[64] C. T. Ionescu. On the approximation of upper semi-continuous correspondences and the equilibriums of generalized games. *J. Math. Anal. Appl.*, 136:267–289, 1988. (Cited on page 16.)

[65] W. Jia, S. Xiang, J. He, and Y. Yang. Existence and stability of weakly Pareto-Nash equilibrium for generalized multiobjective multi-leader–follower games. *J. Global Optim.*, 61(2):397–405, 2015. (Cited on page 102.)

[66] F. Jules and M. Lassonde. Sommes approchées de sous-différentiels. partie i: Théorie. *Prépubl. no. 00/03, Dept Math. Info., Univ. Antilles-Guyane*, 2000. (Cited on page 38.)

[67] F. Jules and M. Lassonde. Sommes approchées de sous-différentiels. partie ii: Spécialisations. *Prépubl. no. 00/03, Dept Math. Info., Univ. Antilles-Guyane*, 2000. (Cited on page 38.)

[68] F. Jules and M. Lassonde. Formulas for subdifferentials of sums of convex functions. *J. Convex Anal.*, 9(2):519–534, 2002. (Cited on page 38.)

[69] P. Q. Khanh and N. H. Quan. Versions of the Weierstrass theorem for bifunctions and solution existence in optimization. *SIAM J. Optim.*, 29(2):1502–1523, 2019. (Cited on pages 108 and 122.)

[70] A. Kruger. On fréchet subdifferentials. *J. Math. Sci. (N.Y.)*, 116(3):3325–3358, 2003. (Cited on page 49.)

[71] A. A. Kulkarni and U. V. Shanbhag. A shared-constraint approach to multi-leader multi-follower games. *Set-valued and variational analysis*, 22(4):691–720, 2014. (Cited on pages 95, 96 and 97.)

[72] A. A. Kulkarni and U. V. Shanbhag. An existence result for hierarchical stackelberg v/s stackelberg games. *IEEE Trans. Automat. Control*, 60(12):3379–3384, 2015. (Cited on page 94.)

[73] S. S. Kutateladze. Convex e-programming. *Soviet Math. Dokl.*, 20:391–393, 1979. (Cited on page 31.)

[74] M. Lassonde. Minimax et Point Fixes. 2005. unpublished. (Cited on page 16.)

[75] M. B. Lignola and J. Morgan. Topological existence and stability for stackelberg problems. *J. Optim. Theory Appl.*, 84(1):145–169, 1995. (Cited on page 65.)

[76] M. B. Lignola and J. Morgan. Existence of solutions to generalized bilevel programming problem. In *Multilevel Optimization: Algorithms and Applications*, pages 315–332. Springer, 1998. (Cited on pages 85 and 88.)

[77] P. Loridan and J. Morgan. Weak via strong stackelberg problem: new results. *J. Global Optim.*, 8(3):263–287, 1996. (Cited on page 65.)

[78] M. A. López and M. Volle. On the subdifferential of the supremum of an arbitrary family of extended real-valued functions. *Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. RACSAM*, 105(1):3–21, 2011. (Cited on page 26.)

[79] R. Lucchetti, F. Mignanego, and G. Pieri. Existence theorems of equilibrium points in stackelberg. *Optimization*, 18(6):857–866, 1987. (Cited on page 86.)

[80] Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical programs with equilibrium constraints.* Cambridge University Press, 1996. (Cited on page 89.)

[81] L. Mallozzi and R. Messalli. Multi-leader multi-follower model with aggregative uncertainty. *Games*, 8(3):1–14, 2017. (Cited on page 102.)

[82] D. Monderer and L. S. Shapley. Potential games. *Games Econom. Behav.*, 14(1):124–143, 1996. (Cited on page 94.)

[83] B. S. Mordukhovich and T. T. A. Nghia. Subdifferentials of nonconvex supremum functions and their applications to semi-infinite and infinite programs with lipschitzian data. *SIAM J. Optim.*, 23(1):406–431, 2013. (Cited on pages 36, 44, 46 and 47.)

[84] Boris S Mordukhovich. *Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Applications*, volume 330. Springer Science & Business Media, 2006. (Cited on pages 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44.)

[85] O. Naselli Ricceri. On the covering dimension of the fixed point set of certain multifunctions. *Comment. Math. Univ. Carolin.*, 32(2):281–286, 1991. (Cited on page 16.)

[86] M. Nasri and W. Sosa. Equilibrium problems and generalized Nash games. *Optimization*, 60:1161–1170, 2011. (Cited on page 121.)

[87] R. Nessah and G. Tian. Existence of solution of minimax inequalities, equilibria in games and fixed points without convexity and compactness assumptions. *J. Optim. Theory Appl.*, 157(1):75–95, 2013. (Cited on pages 109 and 111.)

[88] H. V. Ngai and M. Théra. A fuzzy necessary optimality condition for non-lipschitz optimization in asplund spaces. *SIAM J. Optim.*, 12(3):656–668, 2002. (Cited on page 40.)

[89] T. T. A. Nghia. A nondegenerate fuzzy optimality condition for constrained optimization problems without qualification conditions. *Nonlinear Anal.*, 75(18):6379–6390, 2012. (Cited on pages 35, 41 and 42.)

[90] H. Nikaidô and K. Isoda. Note on noncooperative convex games. *Pacific J. Math.*, 52:807–815, 1955. (Cited on page 120.)

[91] M. A. Noor and W. Oettli. On general nonlinear complementary problems and quasi-equilibria. *Le Mathematiche*, 49(2):313–331, 1994. (Cited on page 116.)

[92] J. Outrata and J. Zowe. A numerical approach to optimization problems with variational inequality constraints. *Math. Program.*, 68(1-3):105–130, 1995. (Cited on page 88.)

[93] J.-S. Pang and M. Fukushima. Quasi-variational inequalities, generalized nash equilibria, and multi-leader-follower games. *Comput. Manag. Sci.*, 2:21–56, 2005. (Cited on page 95.)

[94] J.-P. Penot. A short proof of the separable reduction theorem. *Demonstr. Math.*, 43(3):653–664, 2010. (Cited on page 38.)

[95] P. Pérez-Aros. Formulae for the conjugate and the subdifferential of the supremum function. *J. Optim. Theory Appl.*, pages 1–31, 2019. (Cited on page 36.)

[96] P. Pérez-Aros. Subdifferential formulae for the supremum of an arbitrary family of functions. *SIAM J. Optim.*, 29(2):1714–1743, 2019. (Cited on pages 44 and 46.)

[97] A. Printezis, A. Burnetas, and G. Mohan. Pricing and capacity allocation under asymmetric information using Paris metro pricing. *Int. J. Oper. Res.*, 5(3):265–279, 2009. (Cited on page 102.)

[98] M. Ramos, M. Boix, D. Aussel, L. Montastruc, and S. Domenech. Water integration in eco-industrial parks using a multi-leader-follower approach. *Comput. & Chem. Eng.*, 26:190–207, 2016. (Cited on page 82.)

[99] M. Ramos, M. Boix, D. Aussel, L. Montastruc, P. Vilamajo, and S. Domenech. Water exchanges in eco-industrial parks through multiobjective optimization and game theory. In *Computer Aided Chemical Engineering*, volume 37, pages 1997–2002. Elsevier, 2015. (Cited on page 83.)

[100] M. A. Ramos, M. Rocafull, M. Boix, D. Aussel, L. Montastruc, and S. Domenech. Utility network optimization in eco-industrial parks by a multi-leader follower game methodology. *Comput. & Chem. Eng.*, 112:132–153, 2018. (Cited on page 83.)

[101] S. M. Robinson. Generalized equations and their solutions, part ii: Applications to nonlinear programming. In *Optimality and Stability in Mathematical Programming*, pages 200–221. Springer, 1982. (Cited on pages 72 and 73.)

[102] R. T. Rockafellar. *Convex analysis*. Princeton university press, 2015. (Cited on page 34.)

[103] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009. (Cited on pages 5 and 16.)

[104] J. B. Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33:520–534, 1965. (Cited on page 121.)

[105] D. Salas, K. Cao Van, L. Montastruc, and D. Aussel. Optimal design of exchange networks with blind inputs - part 1 : theoretical analysis. *preprint*, 2019. (Cited on page 83.)

[106] H. H. Schaefer. *Topological vector spaces.* Springer-Verlag, Berlin, 1971. (Cited on page 15.)

[107] S. Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM J. Optim.*, 11(4):918–936, 2001. (Cited on page 92.)

[108] H. D. Sherali. A multiple leader stackelberg model and analysis. *Oper. Res.*, 32(2):390–404, 1984. (Cited on pages 93 and 94.)

[109] H. D. Sherali, C. M. Shetty, and M. S. Bazaraa. *Nonlinear programming: Theory and algorithms.* Wiley-Interscience, 2006. (Cited on page 7.)

[110] H. D. Sherali, A. L. Soyster, and F. H. Murphy. Stackelberg-nash-cournot equilibria: characterizations and computations. *Oper. Res.*, 31(2):253–276, 1983. (Cited on pages 84 and 91.)

[111] N. G. Sudermann-Merx. *Linear Generalized Nash Equilibrium Problems.* PhD thesis, Karlsruher Instituts fur Technologie, 2016. (Cited on page 57.)

[112] N. X. Tan. Quasi-equilibrium problems and fixed point theorems of separately lsc and usc mappings. *Numer. Funct. Anal. Optim.*, 39(2):233–255, 2018. (Cited on page 15.)

[113] L. Thibault. A generalized sequential formula for subdifferentials of sums of convex functions defined on banach spaces. *Recent Developments in Optimization, R. Durier and C. Mich- elot (eds.), Lect. Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin*, 429:340–345, 1995. (Cited on page 24.)

[114] L. Thibault. Limiting convex subdifferential calculus with applications to integration and maximal monotonicity of subdifferential. In *Constructive, Experimental and Nonlinear Analysis. CMS Conf. Proc*, volume 27, pages 279–289, 2000. (Cited on page 25.)

[115] G. Tian and J. X. Zhou. Quasi-variational inequalities without the concavity assumption. *J. Math. Anal. Appl.*, pages 289–299, 1993. (Cited on page 116.)

[116] W. van Ackooij, J. De Boeck, B. Detienne, S. Pan, and M. Poss. Optimizing power generation in the presence of micro-grids. *European J. Oper. Res.*, 271(2):450–461, 2018. (Cited on pages 85 and 102.)

[117] J. Wu and L. Zhang. A smoothing newton method for mathematical programs constrained by parametrized quasi-variational inequalities. *Sci. China Math.*, 54(6):1269–1286, 2011. (Cited on pages 88 and 91.)

[118] J. Wu, L. Zhang, and Y. Zhang. An inexact newton method for stationary points of mathematical programs constrained by parameterized quasi-variational inequalities. *Numer. Algorithms*, 69(4):713–735, 2015. (Cited on pages 88 and 91.)

[119] H. Xu. An mpcc approach for stochastic stackelberg–nash–cournot equilibrium. *Optimization*, 54(1):27–57, 2005. (Cited on page 91.)

[120] J. J. Ye. Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints. *SIAM J. Optim.*, 10(4):943–962, 2000. (Cited on pages 88 and 91.)

[121] J. J. Ye and X. Y. Ye. Necessary optimality conditions for optimization problems with variational inequality constraints. *Math. Oper. Res.*, 22(4):977–997, 1997. (Cited on page 88.)

[122] J. J. Ye and D. L. Zhu. New necessary optimality conditions for bilevel programs by combining the mpec and value function approaches. *SIAM J. Optim.*, 20(4):1885–1905, 2010. (Cited on page 65.)

[123] J. J. Ye, D. L. Zhu, and Q. J. Zhu. Exact penalization and necessary optimality conditions for generalized bilevel programming problems. *SIAM J. Optim.*, 7(2):481–507, 1997. (Cited on page 88.)

[124] C. Zălinescu. *Convex analysis in general vector spaces.* World Scientific, 2002. (Cited on pages 23, 28 and 34.)

[125] Y. Zheng, Y. Fan, X. Zhuo, and J. Chen. Pessimistic referential-uncooperative linear bilevel multi-follower decision making with an application to water resources optimal allocation. *preprint*, 2016. (Cited on page 65.)

[126] Y. Zheng, D. Fang, and Z. Wan. A solution approach to the weak linear bilevel programming problems. *Optimization*, 65(7):1437–1449, 2016. (Cited on page 65.)

[127] J. X. Zhou and G. Chen. Diagonal convexity conditions for problems in convex analysis and quasi-variational inequalities. *J. Math. Anal. Appl.*, 132:213–225, 1988. (Cited on page 116.)

[128] Q. J. Zhu and J. M. Borwein. *Techniques of Variational Analysis.* Springer, 2005. (Cited on pages 35, 38, 39, 40, 41 and 42.)